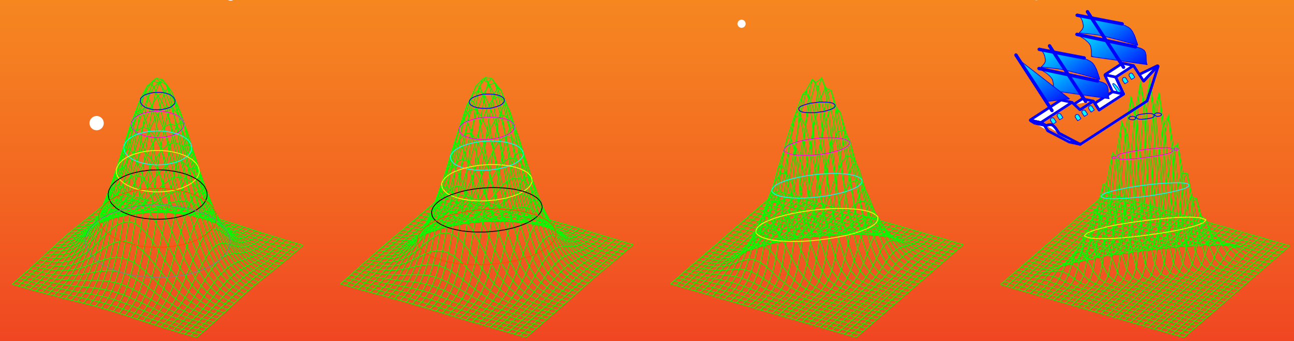
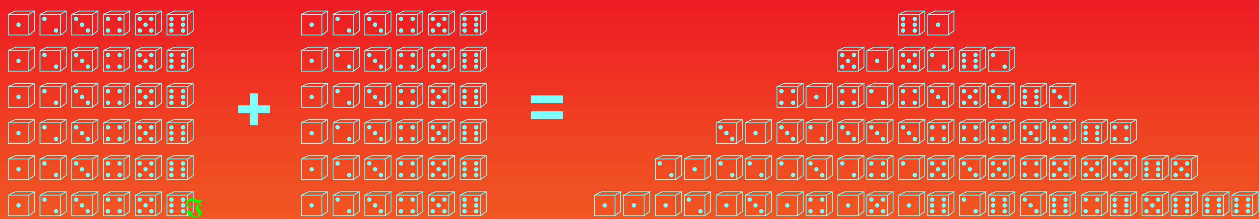
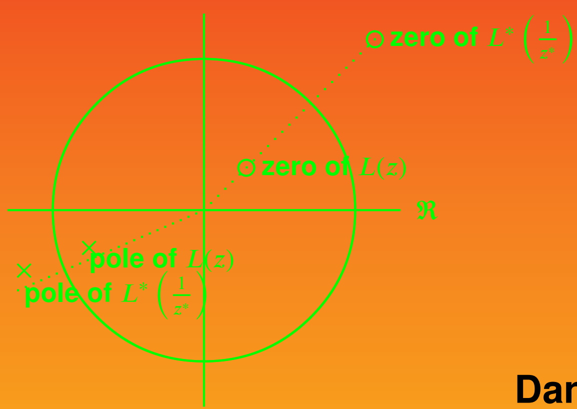
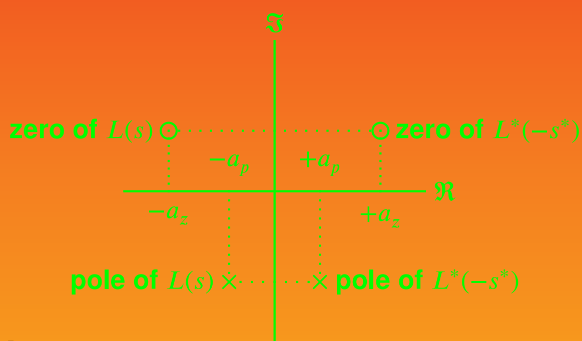
A Book Concerning Statistical Signal Processing

VERSION 0.21

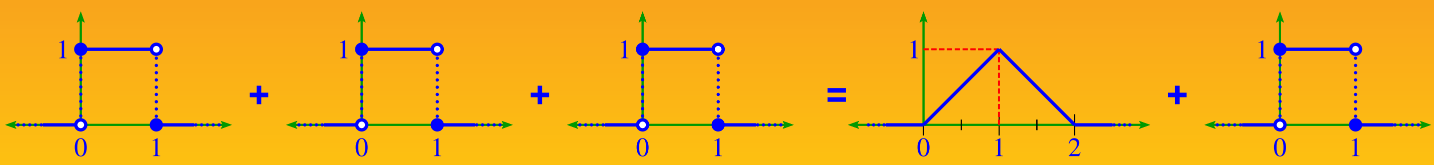




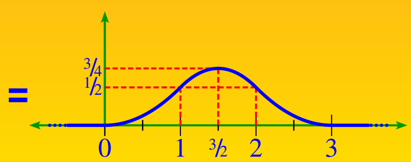




Daniel J. Greenhoe



Signal Processing ABCs series
volume $\boxed{2}$

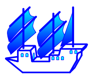


title:	*A Book Concerning Statistical Signal Processing*
document type:	`book`
series:	*Signal Processing ABCs*
volume:	$\boxed{2}$
author:	Daniel J. Greenhoe
version:	VERSION 0.21
time stamp:	2023 March 02 (Thursday) 06:06am EST
copyright:	Copyright © 2023 Daniel J. Greenhoe
license:	Creative Commons license CC BY-NC-ND 4.0
typesetting engine:	XƎLaTeX
document url:	https://github.com/dgreenhoe/pdfs/blob/master/abcstat.pdf https://www.researchgate.net/project/Signal-Processing-ABCs

This text was typeset using XƎLaTeX, which is part of the TeXfamily of typesetting engines, which is arguably the greatest development since the Gutenberg Press. Graphics were rendered using the *pstricks* and related packages, and LaTeX graphics support.

The main roman, *italic*, and **bold** font typefaces used are all from the *Heuristica* family of typefaces (based on the *Utopia* typeface, released by *Adobe Systems Incorporated*). The math font is XITS from the XITS font project. The font used in quotation boxes is adapted from *Zapf Chancery Medium Italic*, originally from URW++ Design and Development Incorporated. The font used for the text in the title is Adventor (similar to *Avant-Garde*) from the *TeX-Gyre Project*. The font used for the ISBN in the footer of individual pages is LIQUID CRYSTAL (*Liquid Crystal*) from *FontLab Studio*. The Latin handwriting font is Lavi from the *Free Software Foundation*.

The ship appearing throughout this text is loosely based on the *Golden Hind*, a sixteenth century English galleon famous for circumnavigating the globe.[1]



[1] Paine (2000) page 63 ⟨Golden Hind⟩

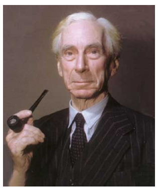
❝*Here, on the level sand,*
Between the sea and land,
What shall I build or write
Against the fall of night?❞

❝*Tell me of runes to grave*
That hold the bursting wave,
Or bastions to design
For longer date than mine.❞

Alfred Edward Housman, English poet (1859–1936) [2]

❝*The uninitiated imagine that one must await inspiration in order to create. That is a mistake. I am far from saying that there is no such thing as inspiration; quite the opposite. It is found as a driving force in every kind of human activity, and is in no wise peculiar to artists. But that force is brought into action by an effort, and that effort is work. Just as appetite comes by eating so work brings inspiration, if inspiration is not discernible at the beginning.*❞

Igor Fyodorovich Stravinsky (1882–1971), Russian-born composer [3]



❝*As I think about acts of integrity and grace, I realise that there is nothing in my knowledge to compare with Frege's dedication to truth. His entire life's work was on the verge of completion, much of his work had been ignored to the benefit of men infinitely less capable, his second volume was about to be published, and upon finding that his fundamental assumption was in error, he responded with intellectual pleasure clearly submerging any feelings of personal disappointment. It was almost superhuman and a telling indication of that of which men are capable if their dedication is to creative work and knowledge instead of cruder efforts to dominate and be known.*❞

Bertrand Russell (1872–1970), British mathematician, in a 1962 November 23 letter to Dr. van Heijenoort. [4]

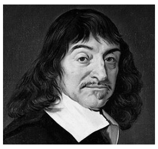
❝*This is all we can do for you now.*❞

Translated note written in Czech found in a World War II Nazi fired shell, one of eleven in the fuel tanks of the B-17 bomber "Tondelayo", that somehow made it safely back to base instead of bursting into flames over Kassel, Germany—all the shells being devoid of explosive charge.[5] In like manner, perhaps this text in some small way may help someone find success in accomplishing a task with significant returns to the well-being of humanity.

[2] quote: Housman (1936) page 64 ⟨"Smooth Between Sea and Land"⟩, Hardy (1940) ⟨section 7⟩
image: http://en.wikipedia.org/wiki/Image:Housman.jpg
[3] quote: Ewen (1961) page 408, Ewen (1950)
image: http://en.wikipedia.org/wiki/Image:Igor_Stravinsky.jpg
[4] quote: Heijenoort (1967) page 127
image: http://www-history.mcs.st-andrews.ac.uk/PictDisplay/Russell.html
[5] Bendiner (1980) page 139, https://www.truthorfiction.com/fall-of-fortresses-bendiner/

SYMBOLS

❝*rugula XVI. Quae vero praesentem mentis attentionem non requirunt, etiamsi ad conclusionem necessaria sint, illa melius est per brevissimas notas designare quam per integras figuras: ita enim memoria non poterit falli, nec tamen interim cogitatio distrahetur ad haec retinenda, dum aliis deducendis incumbit.*❞



❝*Rule XVI. As for things which do not require the immediate attention of the mind, however necessary they may be for the conclusion, it is better to represent them by very concise symbols rather than by complete figures. It will thus be impossible for our memory to go wrong, and our mind will not be distracted by having to retain these while it is taken up with deducing other matters.*❞

René Descartes (1596–1650), French philosopher and mathematician [6]

❝*In signs one observes an advantage in discovery which is greatest when they express the exact nature of a thing briefly and, as it were, picture it; then indeed the labor of thought is wonderfully diminished.*❞

Gottfried Leibniz (1646–1716), German mathematician, [7]

Symbol list

symbol	description	
numbers:		
$\mathbb{Z}$	integers	$\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots$
$\mathbb{W}$	whole numbers	$0, 1, 2, 3, \ldots$

…continued on next page…

[6]quote: Descartes (1684a) ⟨rugula XVI⟩, translation: Descartes (1684b) ⟨rule XVI⟩, image: Frans Hals (circa 1650), http://en.wikipedia.org/wiki/Descartes, public domain

[7]quote: Cajori (1993) ⟨paragraph 540⟩, image: http://en.wikipedia.org/wiki/File:Gottfried_Wilhelm_von_Leibniz.jpg, public domain

symbol	description	
$\mathbb{N}$	natural numbers	$1, 2, 3, \ldots$
$\mathbb{Z}^{\dashv}$	non-positive integers	$\ldots, -3, -2, -1, 0$
$\mathbb{Z}^{-}$	negative integers	$\ldots, -3, -2, -1$
$\mathbb{Z}_{\mathrm{o}}$	odd integers	$\ldots, -3, -1, 1, 3, \ldots$
$\mathbb{Z}_{\mathrm{e}}$	even integers	$\ldots, -4, -2, 0, 2, 4, \ldots$
$\mathbb{Q}$	rational numbers	$\frac{m}{n}$ with $m \in \mathbb{Z}$ and $n \in \mathbb{Z}\backslash 0$
$\mathbb{R}$	real numbers	completion of $\mathbb{Q}$
$\mathbb{R}^{\vdash}$	non-negative real numbers	$[0, \infty)$
$\mathbb{R}^{\dashv}$	non-positive real numbers	$(-\infty, 0]$
$\mathbb{R}^{+}$	positive real numbers	$(0, \infty)$
$\mathbb{R}^{-}$	negative real numbers	$(-\infty, 0)$
$\mathbb{R}^{*}$	extended real numbers	$\mathbb{R}^{*} \triangleq \mathbb{R} \cup \{-\infty, \infty\}$
$\mathbb{C}$	complex numbers	
$\mathbb{F}$	arbitrary field	(often either $\mathbb{R}$ or $\mathbb{C}$)
∞	positive infinity	
$-\infty$	negative infinity	
π	pi	$3.14159265\ldots$
relations:		
®	relation	
ⓐ	relational and	
$X \times Y$	Cartesian product of X and Y	
$(\triangle, \nabla)$	ordered pair	
$\lvert z \rvert$	absolute value of a complex number z	
$=$	equality relation	
$\triangleq$	equality by definition	
$\to$	maps to	
$\in$	is an element of	
$\notin$	is not an element of	
$\mathscr{D}$(®)	domain of a relation ®	
$\mathscr{I}$(®)	image of a relation ®	
$\mathscr{R}$(®)	range of a relation ®	
$\mathscr{N}$(®)	null space of a relation ®	
set relations:		
$\subseteq$	subset	
$\subsetneq$	proper subset	
$\supseteq$	super set	
$\supsetneq$	proper superset	
$\nsubseteq$	is not a subset of	
$\not\subset$	is not a proper subset of	
operations on sets:		
$A \cup B$	set union	
$A \cap B$	set intersection	
$A \triangle B$	set symmetric difference	
$A \backslash B$	set difference	
A^{c}	set complement	
$\lvert \cdot \rvert$	set order	
$\mathbb{1}_A(x)$	set indicator function or characteristic function	
logic:		
1	"true" condition	

...continued on next page...

symbol	description	
0	"false" condition	
$\neg$	logical NOT operation	
$\wedge$	logical AND operation	
$\vee$	logical inclusive OR operation	
$\oplus$	logical exclusive OR operation	
$\Longrightarrow$	"implies";	"only if"
$\Longleftarrow$	"implied by";	"if"
$\Longleftrightarrow$	"if and only if";	"implies and is implied by"
$\forall$	universal quantifier:	"for each"
$\exists$	existential quantifier:	"there exists"
order on sets:		
$\vee$	join or least upper bound	
$\wedge$	meet or greatest lower bound	
$\leq$	reflexive ordering relation	"less than or equal to"
$\geq$	reflexive ordering relation	"greater than or equal to"
$<$	irreflexive ordering relation	"less than"
$>$	irreflexive ordering relation	"greater than"
measures on sets:		
$\lvert X \rvert$	order or counting measure of a set X	
distance spaces:		
d	metric or distance function	
linear spaces:		
$\lVert\cdot\rVert$	vector norm	
$\lvert\lVert\cdot\rVert\rvert$	operator norm	
$\langle \triangle \mid \nabla \rangle$	inner-product	
$\operatorname{span}(\boldsymbol{V})$	span of a linear space $\boldsymbol{V}$	
algebras:		
$\Re$	real part of an element in a $*$-algebra	
$\Im$	imaginary part of an element in a $*$-algebra	
set structures:		
$\boldsymbol{T}$	a topology of sets	
$\boldsymbol{R}$	a ring of sets	
$\boldsymbol{A}$	an algebra of sets	
$\varnothing$	empty set	
2^X	power set on a set X	
sets of set structures:		
$\mathscr{T}(X)$	set of topologies on a set X	
$\mathscr{R}(X)$	set of rings of sets on a set X	
$\mathscr{A}(X)$	set of algebras of sets on a set X	
classes of relations/functions/operators:		
2^{XY}	set of *relations* from X to Y	
Y^X	set of *functions* from X to Y	
$\mathscr{S}_{\mathsf{j}}(X,Y)$	set of *surjective* functions from X to Y	
$\mathscr{I}_{\mathsf{j}}(X,Y)$	set of *injective* functions from X to Y	
$\mathscr{B}_{\mathsf{j}}(X,Y)$	set of *bijective* functions from X to Y	
$\mathscr{B}(\boldsymbol{X},\boldsymbol{Y})$	set of *bounded* functions/operators from $\boldsymbol{X}$ to $\boldsymbol{Y}$	
$\mathscr{L}(\boldsymbol{X},\boldsymbol{Y})$	set of *linear bounded* functions/operators from $\boldsymbol{X}$ to $\boldsymbol{Y}$	
$\mathscr{C}(\boldsymbol{X},\boldsymbol{Y})$	set of *continuous* functions/operators from $\boldsymbol{X}$ to $\boldsymbol{Y}$	
specific transforms/operators:		

...continued on next page...

symbol	description
$\tilde{\mathbf{F}}$	*Fourier Transform* operator (Definition T.2 page 408)
$\hat{\mathbf{F}}$	*Fourier Series* operator
$\breve{\mathbf{F}}$	*Discrete Time Fourier Series* operator (Definition U.1 page 419)
$\mathbf{Z}$	*Z-Transform* operator (Definition V.1 page 429)
$\tilde{\mathsf{f}}(\omega)$	*Fourier Transform* of a function $\mathsf{f}(x) \in \boldsymbol{L}^2_{\mathbb{R}}$
$\breve{\mathsf{x}}(\omega)$	*Discrete Time Fourier Transform* of a sequence $(\!(x_n \in \mathbb{C})\!)_{n\in\mathbb{Z}}$
$\breve{\mathsf{x}}(z)$	*Z-Transform* of a sequence $(\!(x_n \in \mathbb{C})\!)_{n\in\mathbb{Z}}$

SYMBOL INDEX

CONTENTS

Part I

Statistical Analysis

CHAPTER 1

EXPECTATION OPERATOR

❝A likely impossibility is always preferable to an unconvincing possibility.❞
Aristotle (384 BC – 322 BC) [1]

1.1 Definitions

In a *probability space* $(\Omega, \mathbb{E}, \mathsf{P})$ (Definition A.2 page 173), all probability information is contained in the *measure* P. Often times this information is overwhelming and a simpler statistic, which does not offer so much information, is sufficient. Some of the most common statistics can be conveniently expressed in terms of the *expectation operator* $\mathbf{E}$.

Definition 1.1. *Let* $(\Omega, \mathbb{E}, \mathsf{P})$ *be a* PROBABILITY SPACE (Definition A.2 page 173) *and* X *a* RANDOM VARIABLE (Definition B.1 page 184) *on* $(\Omega, \mathbb{E}, \mathsf{P})$ *with* PROBABILITY DENSITY FUNCTION p_x.

DEF *The* ***expectation operator*** $\mathbf{E}_\mathsf{x}$ *on* X *is defined as*

$$\mathbf{E}_\mathsf{x}\mathsf{X} \triangleq \int_{x\in\mathbb{F}} x\mathsf{p}_\mathsf{x}(x)\, \mathsf{d}x.$$

We already said that a *random variable* X is neither random nor a variable, but is rather a function of an underlying process that does appear to be random. However, because it is a function of a process that does appear random, the *random variable* X also appears to be random. That is, if we don't know the outcome of of the underlying experimental process, then we also don't know for sure what X is, and so X does indeed appear to be random. However, eventhough X appears to be random, the expected value $\mathbf{E}_\mathsf{x}\mathsf{X}$ of X is **not random**. Rather it is a fixed value (like 0 or 7.9 or −2.6).

[1] quote: http://en.wikiquote.org/wiki/Aristotle
image: http://en.wikipedia.org/wiki/Aristotle

Two common statistics that are conveniently expressed in terms of the expectation operator are the *mean* and *variance*. The mean is an indicator of the "middle" of a probability distribution and the variance is an indicator of the "spread".

Definition 1.2. *Let* X *be a* RANDOM VARIABLE *on the* PROBABILITY SPACE $(\Omega, \mathbb{E}, \mathsf{P})$.

DEF

(1). *The **mean** μ_{X} of* X *is* $\mu_{\mathsf{X}} \triangleq \mathbf{E}_{\mathsf{x}}\mathsf{X}$

(2). *The **variance** $\mathrm{Var}(\mathsf{X})$ or σ^2_{X} of* X *is* $\mathrm{Var}(\mathsf{X}) \triangleq \mathbf{E}_{\mathsf{x}}\left[(\mathsf{X}-\mathbf{E}_{\mathsf{x}}\mathsf{X})^2\right]$

1.2 Expectation as a linear operator

The next theorem demonstrates that the operator $\mathbf{E}$ is a *linear operator* (Definition R.3 page 360)—which in turn makes $\mathbf{E}$ part of a distinguished club of operators along with fellow member operators differentiation $\frac{\mathsf{d}}{\mathsf{dx}}$, integration $\int \mathsf{dx}$, Laplace $\mathbf{L}$, Fourier $\tilde{\mathbf{F}}$, z-transform $\mathbf{Z}$, etc. Because $\mathbf{E}$ is a linear operator, it immediately inherits all the properties that it's linear operator birthright grants it (Corollary 1.1 page 4).

Theorem 1.1 (Linearity of $\mathbf{E}$). [2] *Let* X *be a* RANDOM VARIABLE *on a* PROBABILITY SPACE $(\Omega, \mathbb{E}, \mathsf{P})$.

THM

$$\mathbf{E}_{\mathsf{x}}(a\mathsf{X}+b\mathsf{Y}+c) = \left(a\mathbf{E}_{\mathsf{x}}\mathsf{X}\right) + \left(b\mathbf{E}_{\mathsf{x}}\mathsf{Y}\right) + c \quad \forall a,b,c\in\mathbb{R} \quad \text{(LINEAR)}$$

✎PROOF:

$$\begin{aligned}
\boxed{\mathbf{E}_{\mathsf{xy}}(a\mathsf{X}+b\mathsf{Y}+c)} &\triangleq \int_{x\in\mathbb{R}}\int_{y\in\mathbb{R}} [ax+by+c]\mathsf{p}_{\mathsf{xy}}(x,y)\;\mathsf{dy}\;\mathsf{dx} && \text{by definition of } \mathbf{E} \text{ (Definition 1.1 page 3)}\\
&= \int_{x\in\mathbb{R}}\int_{y\in\mathbb{R}} ax\mathsf{p}_{\mathsf{xy}}(x,y)\;\mathsf{dy}\;\mathsf{dx} + \int_{x\in\mathbb{R}}\int_{y\in\mathbb{R}} by\mathsf{p}_{\mathsf{xy}}(x,y)\;\mathsf{dy}\;\mathsf{dx} + \int_{x\in\mathbb{R}}\int_{y\in\mathbb{R}} c\mathsf{p}_{\mathsf{xy}}(x,y)\;\mathsf{dy}\;\mathsf{dx}\\
&= \int_{x\in\mathbb{R}} ax \underbrace{\int_{y\in\mathbb{R}} \mathsf{p}_{\mathsf{xy}}(x,y)\;\mathsf{dy}}_{\mathsf{p}_{\mathsf{x}}(x)}\;\mathsf{dx} + \int_{y\in\mathbb{R}} by \underbrace{\int_{x\in\mathbb{R}} \mathsf{p}_{\mathsf{xy}}(x,y)\;\mathsf{dx}}_{\mathsf{p}_{\mathsf{y}}(y)}\;\mathsf{dy} + c\underbrace{\int_{y\in\mathbb{R}}\int_{x\in\mathbb{R}} \mathsf{p}_{\mathsf{xy}}(x,y)\;\mathsf{dx}\;\mathsf{dy}}_{1}\\
&= a\underbrace{\int_{x\in\mathbb{R}} x\mathsf{p}_{\mathsf{x}}(x)\;\mathsf{dx}}_{\mathbf{E}_{\mathsf{x}}\mathsf{X}} + b\underbrace{\int_{y\in\mathbb{R}} y\mathsf{p}_{\mathsf{y}}(y)\;\mathsf{dy}}_{\mathbf{E}_{\mathsf{y}}\mathsf{Y}} + c\\
&\triangleq \boxed{\left(a\mathbf{E}_{\mathsf{x}}\mathsf{X}\right) + \left(b\mathbf{E}_{\mathsf{y}}\mathsf{Y}\right) + c}
\end{aligned}$$

✏

Corollary 1.1. *Let* $\mathbf{E}$ *be the* EXPECTATION OPERATOR *over a* PROBABILITY SPACE $(\Omega, \mathbb{E}, \mathsf{P})$. *Let* $\boldsymbol{L}^2_{\mathbb{F}}$ *be a* VECTOR SPACE OF RANDOM VARIABLES *over* $(\Omega, \mathbb{E}, \mathsf{P})$.

COR

(1).	$\mathbf{E}\mathbb{0}$	$=$	$\mathbb{0}$		*and*
(2).	$\mathbf{E}(-\mathsf{X})$	$=$	$-(\mathbf{E}\mathsf{X})$	$\forall \mathsf{X}\in\boldsymbol{L}^2_{\mathbb{F}}$	*and*
(3).	$\mathbf{E}(\mathsf{X}-\mathsf{Y})$	$=$	$\mathbf{E}\mathsf{X}-\mathbf{E}\mathsf{Y}$	$\forall \mathsf{X},\mathsf{Y}\in\boldsymbol{L}^2_{\mathbb{F}}$	*and*

(4). $$\mathbf{E}\left(\sum_{n=1}^{N}\alpha_n\mathsf{X}_n\right) = \sum_{n=1}^{N}\alpha_n\left(\mathbf{E}\mathsf{X}_n\right) \quad \forall\alpha_n\in\mathbb{F},\quad \forall\mathsf{X}\in\boldsymbol{L}^2_{\mathbb{F}}$$

✎PROOF: These all follow immediately from the fact that $\mathbf{E}$ is a *linear operator* (Theorem 1.1 page 4) and from properties of all linear operators (Theorem R.1 page 360). ✏

[2] Haykin (2014) page 107 ⟨"PROPERTY 1 Linearity"⟩, Wilks (1963b) page 73 ⟨§3.2 "Mean value of a random variable"⟩, Hernandez (2016) page 3

Remark 1.1. Projecting a stochastic process onto a basis often yields valuable insights into the nature of the underlying data. Typical projection operators include the Fourier operator $\tilde{\mathbf{F}}$, Laplace $\mathbf{L}$, and z-transform $\mathbf{Z}$ ...not to mention wavelet operators. But note that any such projection on a random sequence simply produces another random sequence. For example, the Fourier transform $\tilde{\mathbf{F}}\mathsf{x}(n)$ of a random sequence $\mathsf{x}(n)$ is another random sequence.

One way to overcome this difficulty is to simply invoke a *sampling* operator $\mathbf{S}\mathsf{x}(n)$, yielding a deterministic sequence, and then take the Fourier transform of the resulting deterministic sequence. The problem here is that every time you resample the sequence, you will very likely get a different Fourier transform.

Arguably a better approach (and the standard one at that) is to first invoke the expectation operator $\mathbf{E}\mathsf{x}(n)$, also yielding a deterministic sequence.

The good news here is that because $\mathbf{E}$ and all the above mentioned operators are *linear*, we can do all the standard arithmetic acrobatics associated with linear algebra operators (next corollary).

Corollary 1.2. *Let* $\mathbf{M}$ *and* $\mathbf{N}$ *be* LINEAR OPERATORS (Definition R.3 page 360).

COR

1.	$\mathbf{E}(\mathbf{MN})$	$=$	$(\mathbf{EM})\mathbf{N}$	$\forall \mathbf{E}\in\mathscr{L}(\boldsymbol{Z},\boldsymbol{W}), \mathbf{M}\in\mathscr{L}(\boldsymbol{Y},\boldsymbol{Z}), \mathbf{N}\in\mathscr{L}(\boldsymbol{X},\boldsymbol{Y})$	(ASSOCIATIVE)
2.	$\mathbf{E}(\mathbf{M}+\mathbf{N})$	$=$	$(\mathbf{EM})+(\mathbf{EN})$	$\forall \mathbf{E}\in\mathscr{L}(\boldsymbol{Y},\boldsymbol{Z}), \mathbf{M}\in\mathscr{L}(\boldsymbol{X},\boldsymbol{Y}), \mathbf{N}\in\mathscr{L}(\boldsymbol{X},\boldsymbol{Y})$	(LEFT DISTRIBUTIVE)
3.	$(\mathbf{E}+\mathbf{M})\mathbf{N}$	$=$	$(\mathbf{EN})+(\mathbf{MN})$	$\forall \mathbf{E}\in\mathscr{L}(\boldsymbol{Y},\boldsymbol{Z}), \mathbf{M}\in\mathscr{L}(\boldsymbol{Y},\boldsymbol{Z}), \mathbf{N}\in\mathscr{L}(\boldsymbol{X},\boldsymbol{Y})$	(RIGHT DISTRIBUTIVE)
4.	$\alpha(\mathbf{EM})$	$=$	$(\alpha\mathbf{E})\mathbf{M} = \mathbf{E}(\alpha\mathbf{M})$	$\forall \mathbf{E}\in\mathscr{L}(\boldsymbol{Y},\boldsymbol{Z}), \mathbf{M}\in\mathscr{L}(\boldsymbol{X},\boldsymbol{Y}), \alpha\in\mathbb{F}$	(HOMOGENEOUS)

✎PROOF: These all follow immediately from the fact that $\mathbf{E}$ is a *linear operator* (Theorem 1.1 page 4) and from properities of all linear operators (Theorem R.4 page 363). ✏

Corollary 1.3. [3] *Let* X *be a* RANDOM VARIABLE *on a* PROBABILITY SPACE $(\Omega, \mathbb{E}, \mathsf{P})$.

COR

$$\begin{aligned} \mathsf{Var}(\sigma\mathsf{X}+\mu) &= \sigma^2\mathsf{Var}(\mathsf{X}) && \forall\sigma,b\in\mathbb{R} \\ \mathsf{Var}(\mathsf{X}) &= \mathbf{E}_{\mathsf{x}}(\mathsf{X}^2) - (\mathbf{E}_{\mathsf{x}}\mathsf{X})^2 \end{aligned}$$

✎PROOF:

$$\begin{aligned} \mathsf{Var}(\mathsf{X}) &\triangleq \mathbf{E}_{\mathsf{x}}\left[(\mathsf{X}-\mathbf{E}_{\mathsf{x}}\mathsf{X})^2\right] && \text{by definition of } \mathsf{Var} && \text{(Definition 1.2 page 4)} \\ &= \mathbf{E}_{\mathsf{x}}\left[\mathsf{X}^2 - 2\mathsf{X}\mathbf{E}_{\mathsf{x}}\mathsf{X} + (\mathbf{E}_{\mathsf{x}}\mathsf{X})^2\right] && \text{by \textit{Binomial Theorem}} \\ &= \mathbf{E}_{\mathsf{x}}\mathsf{X}^2 - \mathbf{E}_{\mathsf{x}}\left[2\mathsf{X}\mathbf{E}_{\mathsf{x}}\mathsf{X}\right] + \mathbf{E}_{\mathsf{x}}(\mathbf{E}_{\mathsf{x}}\mathsf{X})^2 && \text{by \textit{linearity} of } \mathbf{E} && \text{(Theorem 1.1 page 4)} \\ &= \mathbf{E}_{\mathsf{x}}\mathsf{X}^2 - 2(\mathbf{E}_{\mathsf{x}}\mathsf{X})[\mathbf{E}_{\mathsf{x}}\mathsf{X}] + (\mathbf{E}_{\mathsf{x}}\mathsf{X})^2 \\ &= \mathbf{E}_{\mathsf{x}}(\mathsf{X}^2) - (\mathbf{E}_{\mathsf{x}}\mathsf{X})^2 \\ \mathsf{Var}(\sigma\mathsf{X}+b) &= \mathbf{E}_{\mathsf{x}}(\sigma\mathsf{X}+\mu)^2 - [\mathbf{E}_{\mathsf{x}}(\sigma\mathsf{X}+\mu)]^2 \\ &= \mathbf{E}_{\mathsf{x}}(\sigma^2\mathsf{X}^2 + 2\sigma\mu\mathsf{X} + \mu^2) - [\sigma(\mathbf{E}_{\mathsf{x}}\mathsf{X})+\mu]^2 \\ &= \sigma^2\mathbf{E}_{\mathsf{x}}\mathsf{X}^2 + 2\sigma\mu\mathbf{E}_{\mathsf{x}}\mathsf{X} + \mu^2 - \left[\sigma^2[\mathbf{E}_{\mathsf{x}}\mathsf{X}]^2 + 2\sigma\mu\mathbf{E}_{\mathsf{x}}\mathsf{X} + \mu^2\right] && \text{by \textit{linearity} of } \mathbf{E} && \text{(Theorem 1.1 page 4)} \\ &= \sigma^2\left[\mathbf{E}_{\mathsf{x}}\mathsf{X}^2 - (\mathbf{E}_{\mathsf{x}}\mathsf{X})^2\right] \\ &\triangleq \sigma^2\mathsf{Var}(\mathsf{X}) && \text{by previous result} \end{aligned}$$

✏

Theorem 1.2 (Law of the Unconscious Statistician). [4]

THM

$$\mathbf{E}\left[\mathsf{g}(\mathsf{X})\right] = \int_{x\in\mathbb{R}} \mathsf{g}(x)\mathsf{p}_{\mathsf{x}}(x)\,\mathsf{d}x$$

[3] Hernandez (2016) page 4

[4] Suhov et al. (2005) page 145 ⟨(2.69)⟩, Allen (2018) page 490 ⟨18.3.4 The Law of the Unconscious Statistician⟩, Papoulis (1990) page 124 ⟨Fundamental Theorem⟩, Wasserman (2013) page 48 ⟨“3.6 Theorem (The Rule of the Lazy Statistician).”⟩

1.3 Expectation inner product space

When possible, we like to generalize any given mathematical structure to a more general mathematical structure and then take advantage of the properties of that more general structure. Such a generalization can be done with *random variables*. Random variables can be viewed as vectors in a vector space. Furthermore, the expectation of the product of two *random variables* (e.g. $\mathbf{E}(\mathsf{XY})$) can be viewed as an *inner product* in an *inner product space*. Since we have an *inner product* space, we can then immediately use all the properties of *inner product spaces*, *norm*ed spaces, vector spaces, metric spaces, and topological spaces.

Theorem 1.3. [5] *Let R be a ring, $(\Omega, \mathbb{E}, \mathsf{P})$ be a* PROBABILITY SPACE, $\mathbf{E}$ *the expectation operator, and $\boldsymbol{V} = \{\mathsf{X} | \mathsf{X} : \Omega \to R\}$ be the set of all random vectors in* PROBABILITY SPACE $(\Omega, \mathbb{E}, \mathsf{P})$.

THM

(1).	$\boldsymbol{V} \triangleq \{\mathsf{X} \mid \mathsf{X} : \Omega \to R\}$	*is a* VECTOR SPACE.
(2).	$\langle \mathsf{X} \mid \mathsf{Y} \rangle \triangleq \mathbf{E}(\mathsf{XY}^*)$	*is an* INNER PRODUCT.
(3).	$\|\mathsf{X}\| \triangleq \sqrt{\mathbf{E}(\mathsf{XX}^*)}$	*is a* NORM.
(4).	$(\boldsymbol{V}, \langle \triangle \mid \nabla \rangle)$	*is an* INNER PRODUCT SPACE.

✎PROOF:

1. Proof that $\boldsymbol{V}$ is a vector space:

1)	$\forall \mathsf{X}, \mathsf{Y}, \mathsf{Z} \in \boldsymbol{V}$	$(\mathsf{X} + \mathsf{Y}) + \mathsf{Z} = \mathsf{X} + (\mathsf{Y} + \mathsf{Z})$	(+ is associative)
2)	$\forall \mathsf{X}, \mathsf{Y} \in \boldsymbol{V}$	$\mathsf{X} + \mathsf{Y} = \mathsf{Y} + \mathsf{X}$	(+ is commutative)
3)	$\exists 0 \in \boldsymbol{V}$ such that $\forall \mathsf{X} \in \boldsymbol{V}$	$\mathsf{X} + 0 = \mathsf{X}$	(+ identity)
4)	$\forall \mathsf{X} \in \boldsymbol{V} \exists \mathsf{Y} \in \boldsymbol{V}$ such that	$\mathsf{X} + \mathsf{Y} = 0$	(+ inverse)
5)	$\forall \alpha \in S$ and $\mathsf{X}, \mathsf{Y} \in \boldsymbol{V}$	$\alpha \cdot (\mathsf{X} + \mathsf{Y}) = (\alpha \cdot \mathsf{X}) + (\alpha \cdot \mathsf{Y})$	($\cdot$ distributes over +)
6)	$\forall \alpha, \beta \in S$ and $\mathsf{X} \in \boldsymbol{V}$	$(\alpha + \beta) \cdot \mathsf{X} = (\alpha \cdot \mathsf{X}) + (\beta \cdot \mathsf{X})$	($\cdot$ pseudo-distributes over +)
7)	$\forall \alpha, \beta \in S$ and $\mathsf{X} \in \boldsymbol{V}$	$\alpha(\beta \cdot \mathsf{X}) = (\alpha \cdot \beta) \cdot \mathsf{X}$	($\cdot$ associates with $\cdot$)
8)	$\forall \mathsf{X} \in \boldsymbol{V}$	$1 \cdot \mathsf{X} = \mathsf{X}$	($\cdot$ identity)

2. Proof that $\langle \mathsf{X} \mid \mathsf{Y} \rangle \triangleq \mathbf{E}(\mathsf{XY}^*)$ is an *inner product*.

1)	$\mathbf{E}(\mathsf{XX}^*)$	≥ 0	$\forall \mathsf{X} \in \boldsymbol{V}$	(non-negative)
2)	$\mathbf{E}(\mathsf{XX}^*)$	$= 0 \iff \mathsf{X} = 0$	$\forall \mathsf{X} \in \boldsymbol{V}$	(non-degenerate)
3)	$\mathbf{E}(\alpha\mathsf{XY}^*)$	$= \alpha\mathbf{E}(\mathsf{XY}^*)$	$\forall \mathsf{X}, \mathsf{Y} \in \boldsymbol{V}, \ \forall \alpha \in \mathbb{C}$	(homogeneous)
4)	$\mathbf{E}[(\mathsf{X} + \mathsf{Y})\mathsf{Z}^*]$	$= \mathbf{E}(\mathsf{XZ}^*) + \mathbf{E}(Y\mathsf{Z}^*)$	$\forall \mathsf{X}, \mathsf{Y}, \mathsf{Z} \in \boldsymbol{V}$	(additive)
5)	$\mathbf{E}(\mathsf{XY}^*)$	$= \mathbf{E}(\mathsf{YX}^*)$	$\forall \mathsf{X}, \mathsf{Y} \in \boldsymbol{V}$	(conjugate symmetric).

3. Proof that $\|\mathsf{X}\| \triangleq \sqrt{\mathbf{E}(\mathsf{XX}^*)}$ is a *norm*: This *norm* is simply induced by the above *inner product*.

4. Proof that $(\boldsymbol{V}, \langle \triangle \mid \nabla \rangle)$ is an *inner product space*: Because $\boldsymbol{V}$ is a vector space and $\langle \triangle \mid \nabla \rangle$ is an *inner product*, $(\boldsymbol{V}, \langle \triangle \mid \nabla \rangle)$ is an *inner product space*.

⇒

The next theorem gives some results that follow directly from vector space properties:

[5] Lindquist and Picci (2015) pages 25–26 ⟨2.1 Hilbert Space of Second-Order Random Variables. $\langle \xi \mid \eta \rangle = \mathbf{E}\{\xi\bar{\eta}\}$⟩, Caines (1988) page 21 ⟨$Exy = \int_\Omega x(\omega)y(\omega)dP(\omega)$⟩, Caines (2018) page 21 ⟨$Exy = \int_\Omega x(\omega)y(\omega)dP(\omega)$⟩, Moon and Stirling (2000) pages 105–106

Theorem 1.4 (Parallelogram Law). *Let* $(\Omega, \mathbb{E}, \mathsf{P})$ *be a* PROBABILITY SPACE *with* EXPECTATION *functional* $\mathbf{E}$.

THM
$$2\mathbf{E}(\mathsf{X}\mathsf{X}^*) + 2\mathbf{E}(\mathsf{Y}\mathsf{Y}^*) = \mathbf{E}[(\mathsf{X}+\mathsf{Y})(\mathsf{X}+\mathsf{Y})^*] + \mathbf{E}[(\mathsf{X}-Y)(\mathsf{X}-\mathsf{Y})^*]$$

✎PROOF:

1. $\left(\mathbb{R}^{\Omega}, \mathbf{E}(\mathsf{x},\mathsf{y})\right)$ is an *inner product space*. Proof: Theorem 1.3 (page 6).
2. Because $\left(\mathbb{R}^{\Omega}, \mathbf{E}(\mathsf{x},\mathsf{y})\right)$ is an *inner product space*, the *Parallelogram Law* follows (Theorem N.7 page 317).

✏

1.4 Expectation inequalities

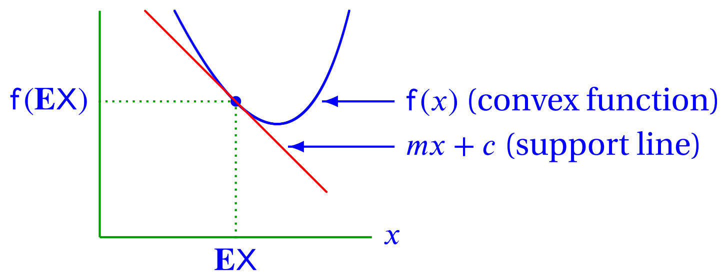


Figure 1.1: *Jensen's Inequality*

Jensen's Inequality is an extremely useful application of *convexity* (Definition P.9 page 340) to the *expectation* operator. Jensen's inequality is stated in Corollary 1.4 (next) and illustrated in Figure 1.1 (page 7).

Corollary 1.4 (Jensen's inequality). [6] *Let* f *be a function in* $\mathbb{R}^{\mathbb{R}}$ *and* X *be a* RANDOM VARIABLE *on* $(\Omega, \mathbb{E}, \mathsf{P})$.

COR
$$\{\mathsf{f} \text{ is CONVEX}\} \implies \{\mathsf{f}(\mathbf{E}\mathsf{X}) \le \mathbf{E}\mathsf{f}(\mathsf{X})\}$$

✎PROOF:

1. Proof 1: Let $mx + c$ be a "support line" under $\mathsf{f}(x)$ (Figure 1.1 page 7) such that

$$\begin{array}{rcll} mx + c & < & \mathsf{f}(x) & \text{for } x \neq \mathbf{E}\mathsf{X} \\ mx + c & = & \mathsf{f}(x) & \text{for } x = \mathbf{E}\mathsf{X}. \end{array}$$

Then

$$\begin{aligned} \mathsf{f}(\mathbf{E}\mathsf{X}) &= m[\mathbf{E}\mathsf{X}] + c \\ &= \mathbf{E}[mX + c] \\ &\le \mathbf{E}\mathsf{f}(\mathsf{X}) \end{aligned}$$

[6] Wasserman (2013) page 66 ⟨"4.9 Theorem (Jensen's inequality)."⟩, Shao (2003) page 31 ⟨"1.3 Distributions and Their Characteristics"⟩, Cover and Thomas (1991) page 25, Jensen (1906) pages 179–180

2. Proof 2 (alternate proof):

$$\begin{aligned}
\mathsf{f}(\mathbf{E}\mathsf{X}) &\triangleq \mathsf{f}\left(\sum_{x\in\mathbb{E}} x\mathsf{P}(x)\right) \\
&\le \sum_{x\in\mathbb{E}} \mathsf{f}(x)\mathsf{P}(x) && \text{by } \textit{Jensen's inequality} \text{ for convex sets} && \text{(Theorem P.1 page 340)}
\end{aligned}$$

✏

Example 1.1. [7] Some examples of *Jensen's Inequality* (Corollary 1.4 page 7) applied to the *expectation operator* are the following:

EX $(\mathbf{E}\mathsf{X})^{-1} < \mathbf{E}(\mathsf{X}^{-1})$ | $\mathbf{E}(\log \mathsf{X}) < \log(\mathbf{E}\mathsf{X})$ | $e^{-\mathbf{E}\mathsf{X}} \le \mathbf{E}\left[e^{-\mathsf{X}}\right]$

Theorem 1.5 (Markov's inequality). [8] *Let* $\mathsf{X} : \Omega \to [0, \infty)$ *be a non-negative valued* RANDOM VARIABLE *and* $a \in (0, \infty)$. *Then*

THM $$\mathsf{P}\{\mathsf{X} \ge a\} \le \frac{1}{a}\mathbf{E}\mathsf{X}$$

✎PROOF:

$$\begin{aligned}
I &\triangleq \begin{cases} 1 & \text{for} \quad \mathsf{X} \ge a \\ 0 & \text{for} \quad \mathsf{X} < a \end{cases} \\
aI &\le \mathsf{X} \\
I &\le \frac{1}{a}\mathsf{X} \\
\mathbf{E}I &\le \mathbf{E}\left(\frac{1}{a}\mathsf{X}\right)
\end{aligned}$$

$$\begin{aligned}
\mathsf{P}\{\mathsf{X} \ge a\} &= 1 \cdot \mathsf{P}\{\mathsf{X} \ge a\} + 0 \cdot \mathsf{P}\{\mathsf{X} < a\} \\
&= \mathbf{E}I \\
&\le \mathbf{E}\left(\frac{1}{a}\mathsf{X}\right) \\
&= \frac{1}{a}\mathbf{E}\mathsf{X}
\end{aligned}$$

✏

Theorem 1.6 (Chebyshev's inequality). [9] *Let* X *be a* RANDOM VARIABLE *with mean* μ *and variance* σ^2.

THM $$\mathsf{P}\{|\mathsf{X} - \mu| \ge a\} \le \frac{\sigma^2}{a^2}$$

✎PROOF:

$$\begin{aligned}
\mathsf{P}\{|\mathsf{X} - \mu| \ge a\} &= \mathsf{P}\left\{(\mathsf{X} - \mu)^2 \ge a^2\right\} \\
&\le \frac{1}{a^2}\mathbf{E}(\mathsf{X} - \mu)^2 && \text{by } \textit{Markov's inequality} && \text{(Theorem 1.5 page 8)} \\
&= \frac{\sigma^2}{a^2}
\end{aligned}$$

✏

[7] Shao (2003) pages 31–32 ⟨"Example 1.18"⟩, Dekking et al. (2006) page 110 ⟨"8.5 Solutions to the quick exercises"⟩

[8] Wasserman (2013) page 63 ⟨"4.1 Theorem (Markov's inequality)."⟩, Ross (1998) page 395

[9] Wasserman (2013) page 64 ⟨"4.2 Theorem (Chebyshev's Inequality)."⟩, Ross (1998) page 396

Theorem 1.7 (Kolmogorov's inequality). [10] *Let* X *be a* RANDOM VARIABLE *with mean* μ *and variance* σ^2.

THM
$$\left\{\begin{array}{lll} \text{(A).} & (\mathsf{x}_n) \text{ are INDEPENDENT} & \text{and} \\ \text{(B).} & \text{Each } \mathsf{x}_n \text{ has ZERO-MEAN} & \text{and} \\ \text{(C).} & \text{Each } \mathsf{x}_n \text{ has variance } \sigma^2 & \end{array}\right\} \implies \mathsf{P}\left[\left|\sum_{n=1}^{N}\mathsf{x}_n\right| < \lambda\sum_{n=1}^{N}\mathsf{x}_n^2\right] \geq 1 - \frac{1}{\lambda^2}$$

Theorem 1.8 (Hoeffding's Inequality). [11] *Let* $(\mathsf{X}_1, \mathsf{X}_2, \ldots, \mathsf{X}_N)$ *be a sequence of* RANDOM VARIABLES.

THM
$$\left\{\begin{array}{llcll} \text{(A).} & \mathbf{E}(\mathsf{X}_n) & = & 0 & \text{and} \\ \text{(B).} & \mathsf{X}_n & \in & [a_n : b_n] & \text{and} \\ \text{(C).} & \epsilon & > & 0 & \text{and} \\ \text{(D).} & t & > & 0 & \end{array}\right\} \implies \left\{\mathsf{P}\left(\sum_{n=1}^{N}\mathsf{X}_n\right) \leq e^{-\epsilon t}\prod_{n=1}^{N}\exp\left[\frac{t^2(b_n - a_n)}{8}\right]\right\}$$

Theorem 1.9. [12] *Let* $(\mathsf{X}_1, \mathsf{X}_2, \ldots, \mathsf{X}_N)$ *be a sequence of* RANDOM VARIABLES.

THM
$$\left\{\begin{array}{llcll} \text{(A).} & \mathsf{X}_n & \sim & \text{Bernoulli}(p) & \text{and} \\ \text{(B).} & \epsilon & > & 0 & \end{array}\right\} \implies \left\{\mathsf{P}\left(\left|\frac{1}{N}\sum_{n=1}^{N}\mathsf{X}_n - p\right| > \epsilon\right) \leq 2e^{-2N\epsilon^2}\right\}$$

Theorem 1.10 (Mill's Inequality). [13] *Let* X *be a* RANDOM VARIABLE.

THM
$$\{\ \mathsf{X}_n \sim N(0,1)\ \} \implies \left\{\mathsf{P}(|\mathsf{X}| > x) \leq \left(\sqrt{\frac{2}{\pi}}\right)\frac{e^{-x^2/2}}{x}\right\}$$

Theorem 1.11 (Generalized Triangle Inequality). *Let* $(\Omega, \mathbb{E}, \mathsf{P})$ *be a* PROBABILITY SPACE *with* EXPECTATION *functional* $\mathbf{E}$.

THM
$$\sqrt{\mathbf{E}\left|\sum_{n=1}^{N}\mathsf{X}_n\right|} \leq \sum_{n=1}^{N}\mathbf{E}(\mathsf{X}_n X_n^*)$$

✎PROOF:

1. $(\mathbb{R}^\Omega, \mathbf{E}(\mathsf{x}, \mathsf{y}))$ is an *inner product space*. Proof: Theorem 1.3 (page 6).
2. Because $(\mathbb{R}^\Omega, \mathbf{E}(\mathsf{x}, \mathsf{y}))$ is an *inner product space*, the *Generalized triangle inequality* follows (Theorem O.1 page 327).

✏

Theorem 1.12 (Cauchy-Schwartz Inequality). [14] *Let* $(\Omega, \mathbb{E}, \mathsf{P})$ *be a* PROBABILITY SPACE *with* EXPECTATION *functional* $\mathbf{E}$.

THM
$$|\mathbf{E}(\mathsf{X}\mathsf{Y}^*)|^2 \leq \mathbf{E}(\mathsf{X}\mathsf{X}^*)\,\mathbf{E}(YY^*)$$

✎PROOF:

1. $(\mathbb{R}^\Omega, \mathbf{E}(\mathsf{x}, \mathsf{y}))$ is an *inner product space*. Proof: Theorem 1.3 (page 6).

[10] Wilks (1963b) page 107 ⟨§4.5 "Kolmogorov's inequality"⟩
[11] Devroye et al. (2013) pages 122–123 ⟨"Theorem 8.1 (Hoeffding (1963))."⟩, Wasserman (2013) page 64 ⟨"4.4 Theorem (Hoeffding's Inequality)."⟩, Hoeffding (1963)
[12] Wasserman (2013) page 65 ⟨"4.5 Theorem."⟩
[13] Wasserman (2013) page 65 ⟨"4.7 Theorem (Mill's Inequality)."⟩
[14] Wasserman (2013) page 66 ⟨"4.8 Theorem (Cauchy-Schwartz Inequality)."⟩

2. Because $(\mathbb{R}^{\Omega}, \mathbf{E}(\mathsf{x}, \mathsf{y}))$ is an *inner product space*, the *Cauchy-Schwartz inequality* follows (Theorem N.2 page 310).

✏

1.5 Conditional expectation

Sometimes the problem of finding the expected value of a *random variable* X can be simplified by "conditioning X on Y". It has already been pointed out in Section 1.1 (page 3) that the expected value $\mathbf{E}_{\mathsf{x}}\mathsf{X}$ of X is **not random**. On the other hand, note that $\mathbf{E}(\mathsf{X}|\mathsf{Y})$ **is random**. This is because $\mathbf{E}(\mathsf{X}|\mathsf{Y})$ is a function of Y. That is, once we know that Y equals some fixed value y (like 0 or 2.7 or -5.1) then $\mathbf{E}(\mathsf{X}|\mathsf{Y}=y)$ is also fixed. However, if we don't know the value of Y, then Y is still a *random variable* and the expression $\mathbf{E}(\mathsf{X}|\mathsf{Y})$ is also random (a function of *random variable* Y).

Theorem 1.13. [15] *Let* X *and* Y *be* RANDOM VARIABLES. *Then*

THM
$$\mathbf{E}_{\mathsf{x}}\mathsf{X} = \mathbf{E}_{\mathsf{y}}\mathbf{E}_{\mathsf{x}|\mathsf{y}}(\mathsf{X}|\mathsf{Y})$$

✎PROOF:

$$\begin{aligned}
\mathbf{E}_{\mathsf{y}}\mathbf{E}_{\mathsf{x}|\mathsf{y}}(\mathsf{X}|\mathsf{Y}) &\triangleq \mathbf{E}_{\mathsf{y}}\left[\int_{x\in\mathbb{R}} x\mathsf{p}(\mathsf{X}=x|\mathsf{Y})\,\mathsf{d}x\right] && \text{by definition of } \mathbf{E} && \text{(Definition 1.1 page 3)}\\
&\triangleq \int_{y\in\mathbb{R}}\left[\int_{x\in\mathbb{R}} x\mathsf{p}(x|\mathsf{Y}=y)\,\mathsf{d}x\right]\mathsf{p}(y)\,\mathsf{d}y && \text{by definition of } \mathbf{E} && \text{(Definition 1.1 page 3)}\\
&= \int_{y\in\mathbb{R}}\int_{x\in\mathbb{R}} x\mathsf{p}(x|y)\mathsf{p}(y)\,\mathsf{d}x\,\mathsf{d}y\\
&= \int_{x\in\mathbb{R}} x\int_{y\in\mathbb{R}} \mathsf{p}(x,y)\,\mathsf{d}y\,\mathsf{d}x && \text{by Theorem B.3 page 185}\\
&= \int_{x\in\mathbb{R}} x\mathsf{p}(x)\,\mathsf{d}x && \text{by Theorem B.3 page 185}\\
&\triangleq \mathbf{E}_{\mathsf{x}}\mathsf{X} && \text{by definition of } \mathbf{E} && \text{(Definition 1.1 page 3)}
\end{aligned}$$

✏

[15] Jazwinski (1970) page 40 ⟨"Theorem 2.9 (Conditional Expectations)"⟩, Jazwinski (2007) page 40

CHAPTER 2

RANDOM SEQUENCES

❝*We are quite in danger of sending highly trained and highly intelligent young men out into the world with tables of erroneous numbers under their arms, and with a dense fog in the place where their brains ought to be. In this century, of course, they will be working on guided missiles and advising the medical profession on the control of disease, and there is no limit to the extent to which they could impede every sort of national effort.*❞

Ronald A. Fisher, (1890–1962), Statistician, at a lecture in 1958 at Michigan State University [1]

2.1 Definitions

Definition 2.1.

DEF *A **random sequence*** $\mathsf{x}(n) \in \Omega$ *is a* SEQUENCE *over a* PROBABILITY SPACE $(\Omega, \mathbb{E}, \mathsf{P})$ *(Definition A.2 page 173)*.

Definition 2.2. [2] *Let* $\mathsf{x}(n)$ *and* $\mathsf{y}(n)$ *be* RANDOM SEQUENCES.

*The **mean***	$\mu_{\mathsf{x}}(n)$	*of* $\mathsf{x}(n)$ *is*	$\mu_{\mathsf{x}}(n)$	$\triangleq$	$\mathbf{E}[\mathsf{x}(n)]$
*The **variance***	$\sigma^2_{\mathsf{x}}(n)$	*of* $\mathsf{x}(n)$ *is*	$\sigma^2_{\mathsf{x}}(n)$	$\triangleq$	$\mathbf{E}\left(\left[\mathsf{x}(n) - \mu_{\mathsf{x}}(n)\right]^2\right)$
*The **cross-correlation***	$\mathsf{R}_{\mathsf{xy}}(n,m)$	*of* $\mathsf{x}(n)$ *and* $\mathsf{y}(n)$ *is*	$\mathsf{R}_{\mathsf{xy}}(n,m)$	$\triangleq$	$\mathbf{E}\left[\mathsf{x}(n+m)\mathsf{y}^*(n)\right]$
*The **auto-correlation***	$\mathsf{R}_{\mathsf{xx}}(n,m)$	*of* $\mathsf{x}(n)$ *is*	$\mathsf{R}_{\mathsf{xx}}(n,m)$	$\triangleq$	$\left.\mathsf{R}_{\mathsf{xy}}(n,m)\right\|_{\mathsf{y}=\mathsf{x}}$

[1] quote: Yates and Mather (1963) page 107. image: http://www.genetics.org/content/154/4/1419

[2] Papoulis (1984) page 263 $\langle R_{xy}(m) = E\left\{\mathsf{x}(m)\mathsf{y}^*(0)\right\}\rangle$, Wilks (1963b) page 77 ⟨§3.4 "Moments of two-dimensional random variables"⟩, Cadzow (1987) page 341 $\langle r_{xy}(m) = E\left[\mathsf{x}(m)\mathsf{y}^*(0)\right]\rangle$, MatLab (2018b) $\langle R_{xy}(m) = E\left\{x_{n+m}y_n^*\right\}\rangle$, MatLab (2018a) $\langle R_{xy}(m) = E\left\{x_{n+m}y_n^*\right\}\rangle$

2.2 Properties

Theorem 2.1.

THM
$$\begin{aligned}\mathsf{R}_{xx}(n,m) &= \mathsf{R}^*_{xx}(n+m,-m)\\ \mathsf{R}_{xy}(n,m) &= \mathsf{R}^*_{yx}(n+m,-m)\end{aligned}$$

✎PROOF:

$$\begin{aligned}
\mathsf{R}_{xy}(n,m) &\triangleq \mathbf{E}\left[\mathsf{x}(n+m)\mathsf{y}^*(n)\right] && \text{by definition of } \mathsf{R}_{xy}(n,m) && \text{(Definition 2.2 page 11)}\\
&= \mathbf{E}\left[\mathsf{y}^*(n)\mathsf{x}(n+m)\right] && \text{by } \textit{commutative} \text{ property of } (\mathbb{C},+,\cdot,0,1) && \text{(Definition G.5 page 222)}\\
&= \left(\mathbf{E}\left[\mathsf{y}(n)\mathsf{x}^*(n+m)\right]\right)^* && \text{by } \textit{distributive} \text{ property of } *\textbf{-algebras} && \text{(Definition M.3 page 304)}\\
&= \left(\mathbf{E}\left[\mathsf{y}(n+m-m)\mathsf{x}^*(n+m)\right]\right)^* && \text{by } \textit{additive identity} \text{ property of } (\mathbb{R},+,\cdot,0,1) && \text{(Definition G.5 page 222)}\\
&\triangleq \mathsf{R}^*_{yx}(n+m,-m) && \text{by definition of } \mathsf{R}_{xy}(n,m) && \text{(Definition 2.2 page 11)}
\end{aligned}$$

$$\begin{aligned}
\mathsf{R}_{xx}(n,m) &= \mathsf{R}_{xy}(n,m)\big|_{y=x} && \text{by } \mathsf{y}=\mathsf{x} \text{ constraint}\\
&= \mathsf{R}^*_{xy}(n+m,-m)\big|_{y=x} && \text{by previous result}\\
&= \mathsf{R}^*_{xx}(n+m,-m) && \text{by } \mathsf{y}=\mathsf{x} \text{ constraint}
\end{aligned}$$

✏

2.3 Wide Sense Stationary processes

Definition 2.3. *Let* $\mathsf{x}(n)$ *be a* RANDOM SEQUENCE *with* MEAN $\mu_{\mathsf{x}}(n)$ *and* VARIANCE $\sigma^2(n)$ *(Definition 2.2 page 11).*

DEF

$\mathsf{x}(n)$ *is* ***wide sense stationary*** *(****WSS****) if*

1. $\mu_{\mathsf{x}}(n)$ *is* CONSTANT *with respect to* n (STATIONARY IN THE 1ST MOMENT) *and*
2. $\sigma^2_{\mathsf{x}}(n)$ *is* CONSTANT *with respect to* n (STATIONARY IN THE 2ND MOMENT)

Definition 2.4. [3] *Let* $\mathsf{x}(n)$ *be a* RANDOM SEQUENCE *with statistics* $\mu_{\mathsf{x}}(n)$, $\sigma^2_{\mathsf{x}}(n)$, $\mathsf{R}_{xx}(n,m)$, *and* $\mathsf{R}_{xy}(n,m)$ *(Definition 2.2 page 11).*

DEF

$\{\ \mathsf{x}$ *and* y *are* WIDE SENSE STATIONARY $\}\implies$

(1).	*The* ***mean***	μ_{x}	*of* $\mathsf{x}(n)$ *is*	μ_{x}	$\triangleq$	$\mathbf{E}[\mathsf{x}(0)]$
(2).	*The* ***variance***	σ^2_{x}	*of* $\mathsf{x}(n)$ *is*	σ^2_{x}	$\triangleq$	$\mathbf{E}\left(\left[\mathsf{x}(0)-\mu_{\mathsf{x}}\right]^2\right)$
(4).	*The* ***cross-correlation***	$\mathsf{R}_{xy}(m)$	*of* $\mathsf{x}(n)$ *and* $\mathsf{y}(n)$ *is*	$\mathsf{R}_{xy}(m)$	$\triangleq$	$\mathbf{E}\left[\mathsf{x}(m)\mathsf{y}^*(0)\right]$
(3).	*The* ***auto-correlation***	$\mathsf{R}_{xx}(m)$	*of* $\mathsf{x}(n)$ *is*	$\mathsf{R}_{xx}(m)$	$\triangleq$	$\mathsf{R}_{xy}(m)\big\|_{y=x}$

Remark 2.1. The $\mathsf{R}_{xy}(n,m)$ of Definition 2.2 (page 11) and the $\mathsf{R}_{xy}(m)$ of Definition 2.4 (page 12) (etc.) are examples of *function overload*—that is, functions that use the same mnemonic but are distinguished by different domains. Perhaps a more common example of function overload is the "+" mnemonic. Traditionally it is used with domain of the natural numbers $\mathbb{N}$ as in $3+2$. Later it was extended for domain real numbers $\mathbb{R}$ as in $\sqrt{3}+\sqrt{2}$, or even complex numbers $\mathbb{C}$ as in

[3] Papoulis (1984) page 263 ⟨"$R_{xy}(\tau) = E\{\mathsf{x}(t+\tau)\mathsf{y}^*(t)\}$"⟩, Cadzow (1987) page 341 ⟨$r_{xy}(n) = E\left[\mathsf{x}(k+n)\mathsf{y}^*(k)\right]$ (10.41)⟩

$\left(\sqrt{3}+i\sqrt{2}\right)+(e+i\pi)$. And it was even more dramatically extended for use with domain $\mathbb{R}^N \times \mathbb{R}^M$ in "linear algebra" as in

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} = \begin{bmatrix} 6 & 8 \\ 10 & 12 \end{bmatrix}$$

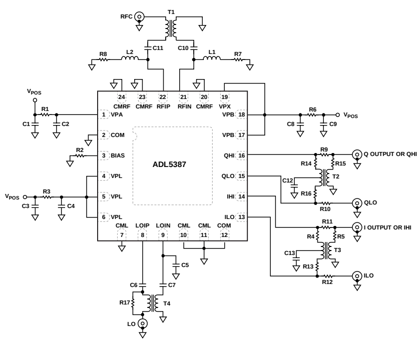
Remark 2.2. [4] The definition for $\mathsf{R}_{xy}(m)$ can be defined with the conjugate $*$ on either x or y, or on neither or both; and moreover x may either lead or lag y. In total, there are $2\times2\times2=8$ different ways to define $\mathsf{R}_{xy}(m)$.[5] and $\mathsf{R}_{xx}(m)$ involve complex numbers. This may seem curious when typical ADCs provide real-valued sequences. Note however that complex-valued sequences often come up in signal processing due to some common system architectures:

1. The presence of an *FFT* operator in the signal processing path
2. The *complex envelope* $\mathsf{x}_l(t)$ of a modulated *narrowband* communications signal $\mathsf{x}(t)$.
3. Communications channel processing involving phase discrimination (e.g. PSK and QAM).

In the case of a narrowband signal $\mathsf{x}(t)$ modulated by a sinusoid at center frequence f_c, we have three canonical forms. These can be shown to be equivalent:

$$\begin{aligned}
\mathsf{x}(t) &\triangleq \underbrace{\boxed{\mathsf{a}(t)\cos\left[2\pi f_c t+\phi(t)\right]}}_{\text{amplitude-phase form}} && \text{amplitude and phase form}\\
&= \underbrace{\mathsf{a}(t)\cos[\phi(t)]}_{\mathsf{p}(t)}\cos[2\pi f_c t] - \underbrace{\mathsf{a}(t)\sin[\phi(t)]}_{\mathsf{q}(t)}\sin[2\pi f_c t] && \text{by \textit{double angle formulas}} \quad \text{(Theorem I.9 page 239)}\\
&= \underbrace{\boxed{\mathsf{p}(t)\cos[2\pi f_c t] - \mathsf{q}(t)\sin[2\pi f_c t]}}_{\text{quadrature form}} && \text{quadrature form}\\
&= \mathbf{R_e}\left([\mathsf{p}(t)+i\mathsf{q}(t)][\cos(2\pi f_c t)+i\sin(2\pi f_c t)]\right) && \text{by definitions of } \mathbf{R_e}\\
&= \underbrace{\boxed{\mathbf{R_e}\left[\mathsf{x}_l(t)e^{i2\pi f_c t}\right]}}_{\text{complex envelope form}} && \text{by \textit{Euler's identity}} \quad \text{(Theorem I.5 page 234)}
\end{aligned}$$

Note that in these equivalent forms, the *complex envelope* $\mathsf{x}_l(t)$ is conveniently represented as a *complex-valued* function in terms of the *quadrature component* $\mathsf{p}(t)$ and the *inphase component* $\mathsf{q}(t)$ such that $\mathsf{x}_l(t)=\mathsf{p}(t)+i\mathsf{q}(t)$.



Example 2.1. In practice (with real hardware), you will likely first have access to the quadrature components $\mathsf{p}(t)$ and $\mathsf{q}(t)$. Take for example the *Analog Devices ADL5387 Quadrature Demodulator* and evaluation board, as illustrated to the right.[6] Note that *quadrature component* $\mathsf{p}(t)$ is available at connector "Q OUTPUT" and *inphase component* $\mathsf{q}(t)$ is available at connector "I OUTPUT".

[4] S. Lawrence Marple (1987) pages 51–53 ⟨"Appendix 2.A Source of Complex-Valued Signals"⟩, S. Lawrence Marple (2019) pages 48–50 ⟨§"2.12 Extra: Source of Complex-Valued Signals"⟩, Greenhoe (2019b) ⟨Chapter 2: Narrowband Signals⟩

[5] Greenhoe (2019a)

[6] Diagram extracted from Devices (2016). Extraction notes: `pdftk ADL5387.pdf cat 24 output page24.pdf pdfcrop --margins "-50 -120 -60 -260" --clip page24.pdf image.pdf gswin32c.exe -sDEVICE=pdfwrite -dNOPAUSE -dBATCH -dSAFER -dCompatibilityLevel=1.5 -sOutputFile=ADL5387_page24_schematic.pdf image.pdf`

Proposition 2.1. *Let* $\mathsf{y}(n)$ *be a* RANDOM SEQUENCE, $\mathsf{x}(n)$ *a* RANDOM SEQUENCE *with* AUTO-CORRELATION $\mathsf{R}_{\mathsf{xx}}(n,m)$, *and* R_{xy} *the* CROSS-CORRELATION *of* x *and* y.

PRP

$$\left\{\begin{array}{l}\mathsf{x} \textit{ and } \mathsf{y} \textit{ are}\\ \text{WIDE SENSE STATIONARY}\\ (\text{WSS}) \text{ (Definition 8.1 page 59)}\end{array}\right\} \implies \left\{\begin{array}{rcll}\mathsf{R}_{\mathsf{xx}}(n,m) &=& \mathsf{R}_{\mathsf{xx}}(m) & \forall n\in\mathbb{Z}\\ \mathsf{R}_{\mathsf{xy}}(n,m) &=& \mathsf{R}_{\mathsf{xy}}(m) & \forall n\in\mathbb{Z}\\ \text{(Definition 2.2 page 11)} && \text{(Definition 2.4 page 12)} &\end{array}\right\}$$

✎PROOF:

$$\begin{aligned}
\mathsf{R}_{\mathsf{xy}}(n,m) &\triangleq \mathbf{E}\left[\mathsf{x}[n+m]\mathsf{y}^*[n]\right] && \text{by definition of } \mathsf{R}_{\mathsf{xy}}(n,m) && \text{(Definition 2.2 page 11)}\\
&= \mathbf{E}\left[\mathsf{x}[n-n+m]\mathsf{y}^*[n-n]\right] && \text{by } \textit{wide sense stationary} \text{ hypothesis}\\
&= \mathbf{E}\left[\mathsf{x}[m]\mathsf{y}^*[0]\right]\\
&\triangleq \mathsf{R}_{\mathsf{xy}}(m) && \text{by definition of } \mathsf{R}_{\mathsf{xy}}(m) && \text{(Definition 2.4 page 12)}\\
\mathsf{R}_{\mathsf{xx}}(n,m) &= \mathsf{R}_{\mathsf{xy}}(n,m)\big|_{\mathsf{y}=\mathsf{x}}\\
&= \mathsf{R}_{\mathsf{xy}}(m)\big|_{\mathsf{y}=\mathsf{x}} && \text{by previous result}\\
&= \mathsf{R}_{\mathsf{xx}}(m)
\end{aligned}$$

✏

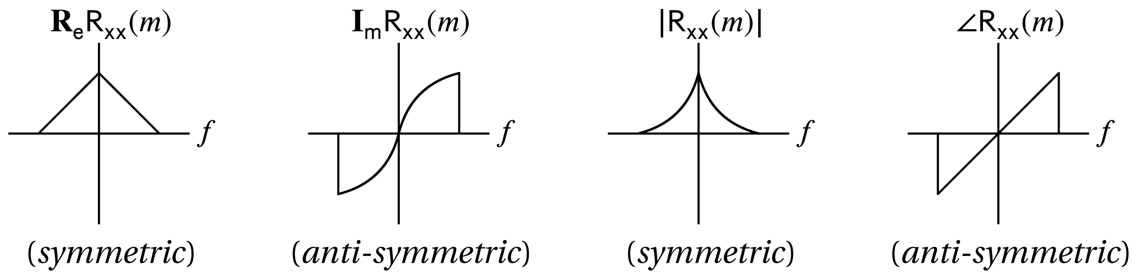


Figure 2.1: *auto-correlation* $\mathsf{R}_{\mathsf{xx}}(m)$

Corollary 2.1. *Let* $\mathsf{x}(n)$ *be a* RANDOM SEQUENCE *with* AUTO-CORRELATION $\mathsf{R}_{\mathsf{xx}}(n,m)$, $\mathsf{y}(n)$ *a* RANDOM SEQUENCE *with* AUTO-CORRELATION $\mathsf{R}_{\mathsf{yy}}(n,m)$, *and* $\mathsf{R}_{\mathsf{xy}}(n,m)$ *the* CROSS-CORRELATION *of* x *and* y. *Let* $\mathbf{S}$ *be a* SYSTEM *with input* $\mathsf{x}(n)$ *and output* $\mathsf{y}(n)$.

COR

$$\left\{\begin{array}{lll}\text{(A).} & \mathsf{x} \textit{ is } \text{WSS} & \textit{and}\\ \text{(B).} & \mathsf{y} \textit{ is } \text{WSS} & \textit{and}\\ \text{(C).} & \mathbf{S} \textit{ is } \text{LTI} &\end{array}\right\} \implies \left\{\begin{array}{lrcll}
(1). & \mathsf{R}_{\mathsf{xy}}(m) &=& \mathsf{R}^*_{\mathsf{yx}}(-m) & & \textit{and}\\
(2). & \mathsf{R}_{\mathsf{xx}}(m) &=& \mathsf{R}^*_{\mathsf{xx}}(-m) & (\text{CONJUGATE SYMMETRIC}) & \textit{and}\\
(3). & \mathbf{R}_{\mathrm{e}}\mathsf{R}_{\mathsf{xx}}(m) &=& \mathbf{R}_{\mathrm{e}}\mathsf{R}_{\mathsf{xx}}(-m) & (\text{SYMMETRIC}) & \textit{and}\\
(4). & \mathbf{I}_{\mathrm{m}}\mathsf{R}_{\mathsf{xx}}(m) &=& -\mathbf{I}_{\mathrm{m}}\mathsf{R}_{\mathsf{xx}}(-m) & (\text{ANTI-SYMMETRIC}) & \textit{and}\\
(5). & \left|\mathsf{R}_{\mathsf{xx}}(m)\right| &=& \left|\mathsf{R}_{\mathsf{xx}}(-m)\right| & (\text{SYMMETRIC}) & \textit{and}\\
(6). & \angle\mathsf{R}_{\mathsf{xx}}(m) &=& -\angle\mathsf{R}_{\mathsf{xx}}(-m) & (\text{ANTI-SYMMETRIC}) &
\end{array}\right\}$$

✎PROOF:

$$\begin{aligned}
\mathsf{R}_{\mathsf{xy}}(m) &= \mathsf{R}_{\mathsf{xy}}(n,m) && \text{by Proposition 2.1 page 14} && \text{and hypotheses (A),(B)}\\
&= \mathsf{R}^*_{\mathsf{yx}}(n+m,-m) && \text{by Theorem 2.1 page 12} && \text{and hypothesis (B)}\\
&= \mathsf{R}^*_{\mathsf{yx}}(-m) && \text{by Proposition 2.1 page 14} && \text{and hypothesis (A)}\\
\mathsf{R}_{\mathsf{xx}}(m) &= \mathsf{R}_{\mathsf{xx}}(n,m) && \text{by Proposition 2.1 page 14} && \text{and hypothesis (A)}\\
&= \mathsf{R}^*_{\mathsf{xx}}(n+m,-m) && \text{by Theorem 2.1 page 12} && \text{and hypothesis (B)}\\
&= \mathsf{R}^*_{\mathsf{xx}}(-m) && \text{by Proposition 2.1 page 14} && \text{and hypothesis (A)}
\end{aligned}$$

✏

2.4 Spectral density

Definition 2.5. *Let* $\mathsf{x}(n)$ *and* $\mathsf{y}(n)$ *be* WIDE SENSE STATIONARY RANDOM SEQUENCES *with auto-correlation* $\mathsf{R}_{\mathsf{xx}}(m)$ *and cross-correlation* $\mathsf{R}_{\mathsf{xy}}(m)$. *Let* $\mathbf{Z}$ *be the* Z-TRANSFORM OPERATOR (Definition V.1 page 429).

DEF

The ***z-domain cross spectral density*** ***(CSD)*** $\check{\mathsf{S}}_{\mathsf{xy}}(z)$ *of* $\mathbf{x}$ *and* $\mathbf{y}$ *is*
$$\check{\mathsf{S}}_{\mathsf{xy}}(z) \triangleq \mathbf{Z}\mathsf{R}_{\mathsf{xy}}(m) \triangleq \sum_{m\in\mathbb{Z}} \mathsf{R}_{\mathsf{xy}}(m) z^{-m}$$
The ***z-domain power spectral density*** ***(PSD)*** $\check{\mathsf{S}}_{\mathsf{xx}}(z)$ *of* $\mathbf{x}$ *is* $\check{\mathsf{S}}_{\mathsf{xx}}(z) \triangleq \check{\mathsf{S}}_{\mathsf{xy}}(z)\big|_{\mathsf{y}(n)=\mathsf{x}(n)}$

Definition 2.6. *Let* $\mathsf{x}(n)$ *and* $\mathsf{y}(n)$ *be* WIDE SENSE STATIONARY RANDOM SEQUENCES *with auto-correlation* $\mathsf{R}_{\mathsf{xx}}(m)$ *and cross-correlation* $\mathsf{R}_{\mathsf{xy}}(m)$. *Let* $\check{\mathbf{F}}$ *be the* DISCRETE TIME FOURIER TRANSFORM (DTFT) *operator* (Definition U.1 page 419).

DEF

The ***auto-spectral density*** $\check{\mathsf{S}}_{\mathsf{xx}}(z)$ *of* $\mathbf{x}$ *is*
$$\tilde{\mathsf{S}}_{\mathsf{xx}}(z) \triangleq \check{\mathbf{F}}\mathsf{R}_{\mathsf{xx}}(m) \triangleq \sum_{m\in\mathbb{Z}} \mathsf{R}_{\mathsf{xx}}(m) e^{-i\omega m}$$
The ***cross spectral density*** ***(CSD)*** $\check{\mathsf{S}}_{\mathsf{xy}}(z)$ *of* $\mathbf{x}$ *and* $\mathbf{y}$ *is*
$$\tilde{\mathsf{S}}_{\mathsf{xy}}(z) \triangleq \check{\mathbf{F}}\mathsf{R}_{\mathsf{xy}}(m) \triangleq \sum_{m\in\mathbb{Z}} \mathsf{R}_{\mathsf{xy}}(m) e^{-i\omega m}$$
The ***auto-spectral density*** *is also called* ***power spectral density*** ***(PSD)***.

Theorem 2.2. *Let* $\mathbf{S}$ *be a system with* IMPULSE RESPONSE $\mathsf{h}(n)$, INPUT $\mathsf{x}(n)$, *and* OUTPUT $\mathsf{y}(n)$.

THM

$$\left\{ \mathsf{x} \text{ and } \mathsf{y} \text{ are WIDE SENSE STATIONARY} \right\} \implies \left\{ \begin{array}{lrcll} (1). & \check{\mathsf{S}}_{\mathsf{xx}}(z) & = & \check{\mathsf{S}}^*_{\mathsf{xx}}\left(\frac{1}{z^*}\right) & \text{and} \\ (2). & \check{\mathsf{S}}_{\mathsf{yx}}(z) & = & \check{\mathsf{S}}^*_{\mathsf{xy}}\left(\frac{1}{z^*}\right) & \end{array} \right\}$$

✎PROOF:

$$\begin{aligned}
\check{\mathsf{S}}_{\mathsf{yx}}(z) &\triangleq \mathbf{Z}\mathsf{R}_{\mathsf{yx}}(m) && \text{by definition of } \check{\mathsf{S}}_{\mathsf{xy}}(z) && \text{(Definition 2.6 page 15)} \\
&\triangleq \sum_{m\in\mathbb{Z}} \mathsf{R}_{\mathsf{yx}}(m) z^{-m} && \text{by definition of } \mathbf{Z} && \text{(Definition V.1 page 429)} \\
&\triangleq \sum_{m\in\mathbb{Z}} \mathsf{R}^*_{\mathsf{xy}}(-m) z^{-m} && \text{by Corollary 2.1 page 14} \\
&= \left[\sum_{m\in\mathbb{Z}} \mathsf{R}_{\mathsf{xy}}(-m)(z^*)^{-m}\right]^* && \text{by \textit{antiautomorphic} property of *-algebras} && \text{(Definition M.3 page 304)} \\
&= \left[\sum_{-p\in\mathbb{Z}} \mathsf{R}_{\mathsf{xy}}(p)(z^*)^{p}\right]^* && \text{where } p \triangleq -m && \implies m = -p \\
&= \left[\sum_{p\in\mathbb{Z}} \mathsf{R}_{\mathsf{xy}}(p)(z^*)^{p}\right]^* && \text{by \textit{absolutely summable} property} && \text{(Definition D.3 page 201)} \\
&= \left[\sum_{p\in\mathbb{Z}} \mathsf{R}_{\mathsf{xy}}(p)\left(\frac{1}{z^*}\right)^{-p}\right]^* \\
&= \check{\mathsf{S}}^*_{\mathsf{xy}}\left(\frac{1}{z^*}\right) && \text{by definition of } \mathbf{Z} && \text{(Definition V.1 page 429)} \\
\check{\mathsf{S}}_{\mathsf{xx}}(z) &= \check{\mathsf{S}}_{\mathsf{xy}}(z)\big|_{\mathsf{y}=\mathsf{x}} \\
&= \check{\mathsf{S}}^*_{\mathsf{yx}}(z)\big|_{\mathsf{y}=\mathsf{x}} \\
&= \check{\mathsf{S}}^*_{\mathsf{xy}}\left(\frac{1}{z^*}\right)\bigg|_{\mathsf{y}=\mathsf{x}} && \text{by (2)—previous result}
\end{aligned}$$

$$= \check{S}^*_{xx}\left(\frac{1}{z^*}\right)$$

✏

Corollary 2.2. *Let* **S** *be a system with* IMPULSE RESPONSE $h(n)$, INPUT $x(n)$, *and* OUTPUT $y(n)$.

COR
$$\left\{\begin{array}{lll}\text{(A).} & h \text{ is LTI} & \text{and}\\ \text{(B).} & x \text{ and } y \text{ are WSS} & \end{array}\right\} \implies \left\{\begin{array}{lllll}\text{(1).} & \tilde{S}^*_{xy}(\omega) & = & \tilde{S}_{yx}(\omega) & \text{(CONJUGATE-SYMMETRIC) and}\\ \text{(2).} & \tilde{S}^*_{xx}(\omega) & = & \tilde{S}_{xx}(\omega) & \text{(CONJUGATE SYMMETRIC) and}\\ \text{(3).} & \tilde{S}_{xx}(\omega) & \in & \mathbb{R} & \text{(REAL-VALUED)}\end{array}\right\}$$

✎PROOF:

$$\begin{aligned}
\tilde{S}^*_{xy}(\omega) &= \check{S}^*_{xy}(z)\Big|_{z=e^{i\omega}} && \text{by definition of } DTFT && \text{(Definition U.1 page 419)}\\
&= \check{S}^{**}_{yx}\left(\frac{1}{z^*}\right)\Big|_{z=e^{i\omega}} && \text{by Theorem 2.2 page 15}\\
&= \check{S}_{yx}\left(\frac{1}{z^*}\right)\Big|_{z=e^{i\omega}} && \text{by } involutory \text{ property of } *\textbf{-algebras} && \text{(Definition M.3 page 304)}\\
&= \check{S}_{yx}\left(\frac{1}{e^{i\omega *}}\right)\\
&= \check{S}_{yx}\left(e^{i\omega}\right)\\
&= \tilde{S}_{yx}(\omega) && \text{by definition of } DTFT && \text{(Definition U.1 page 419)}\\
\tilde{S}^*_{xx}(\omega) &= \check{S}^*_{xx}(z)\Big|_{z=e^{i\omega}} && \text{by definition of } DTFT && \text{(Definition U.1 page 419)}\\
&= \check{S}^{**}_{xx}\left(\frac{1}{z^*}\right)\Big|_{z=e^{i\omega}} && \text{by Theorem 2.2 page 15}\\
&= \check{S}_{xx}\left(\frac{1}{z^*}\right)\Big|_{z=e^{i\omega}} && \text{by } involutory \text{ property of } *\textbf{-algebras} && \text{(Definition M.3 page 304)}\\
&= \check{S}_{xx}\left(\frac{1}{e^{i\omega *}}\right)\\
&= \check{S}_{xx}\left(e^{i\omega}\right)\\
&= \tilde{S}_{xx}(\omega) && \text{by definition of } DTFT && \text{(Definition U.1 page 419)}\\
&\implies \tilde{S}_{xx}(\omega) \text{ is } real\text{-}valued\\
\tilde{S}^*_{xx}(\omega) &= \tilde{S}^*_{xy}(\omega)\Big|_{y=x}\\
&= \tilde{S}_{yx}(\omega)\Big|_{y=x} && \text{by previous result}\\
&= \tilde{S}_{xx}(\omega)
\end{aligned}$$

✏

2.5 Spectral Power

The term "*spectral power*" is a bit of an oxymoron because "spectral" deals with leaving the time-domain for the frequency-domain, howbeit the concept of power is solidly founded on the concept of time in that power = energy per time.

However, the *Plancherel Formula*, or more generally *Parseval's Identity* (Proposition K.2 page 271), demonstrates that power in time can also be calculated in frequency.[7] So, it makes some sense to speak of the term "spectral power". Moreover, one way to estimate this power is to average the Fourier Transforms of the product $|x(n)|^2 = x(n)x^*(n)$…that is, to use an estimate of the auto-spectral density $\tilde{S}_{xx}(\omega)$. Thus, an alternate name for *auto-spectral density* is **power spectral density** (PSD).

[7] https://math.stackexchange.com/questions/3785037/

CHAPTER 3

CONTINUOUS RANDOM PROCESSES

3.1 Definitions

Definition 3.1. [1] *Let* $(\Omega, \mathbb{E}, \mathsf{P})$ *be a* PROBABILITY SPACE.

DEF
The function $\mathsf{x} : \Omega \to \mathbb{R}$ *is a* ***random variable***.   *The function* $\mathsf{y} : \mathbb{R} \times \Omega \to \mathbb{R}$ *is a* ***random process***.

The random process $\mathsf{x}(t, \omega)$, where t commonly represents time and $\omega \in \Omega$ is an outcome of an experiment, can take on more specialized forms depending on whether t and ω are fixed or allowed to vary. These forms are illustrated in Figure 3.1 page 17[2] and Figure 3.2 page 18.

$\mathsf{x}(t, \omega)$	fixed t	variable t
fixed ω	number	time function
variable ω	random variable	random process

Figure 3.1: Specialized forms of a random process $\mathsf{x}(t, \omega)$

Definition 3.2. [3] *Let* $\mathsf{x}(t)$ *and* $\mathsf{y}(t)$ *be random processes.*

DEF					
The ***mean***	$\mu_{\mathsf{x}}(t)$	*of* $\mathsf{x}(t)$ *is*	$\mu_{\mathsf{x}}(t)$	$\triangleq$	$\mathbf{E}[\mathsf{x}(t)]$
The ***cross-correlation***	$\mathsf{R}_{\mathsf{xy}}(t)$	*of* $\mathsf{x}(t)$ *and* $\mathsf{y}(t)$ *is*	$\mathsf{R}_{\mathsf{xy}}(t, u)$	$\triangleq$	$\mathbf{E}\left[\mathsf{x}(t)\mathsf{y}^*(u)\right]$
The ***auto-correlation function***	$\mathsf{R}_{\mathsf{xx}}(t)$	*of* $\mathsf{x}(t)$ *is*	$\mathsf{R}_{\mathsf{xx}}(t, u)$	$\triangleq$	$\mathbf{E}\left[\mathsf{x}(t)\mathsf{x}^*(u)\right]$

Remark 3.1. [4] The equation $\int_{u \in \mathbb{R}} \mathsf{R}_{\mathsf{xx}}(t, u)\mathsf{f}(u)\, \mathsf{d}u$ is a *Fredholm integral equation of the first kind* and $\mathsf{R}_{\mathsf{xx}}(t, u)$ is the *kernel* of the equation.

---

[1] Papoulis (1991) page 63, Papoulis (1991) page 285

[2] Papoulis (1991) pages 285–286

[3] Papoulis (1984) page 216 ⟨$R_{xy}(t_1, t_2) = E\left\{\mathsf{x}(t_1)\mathsf{y}^*(t_2)\right\}$ (9-35)⟩

[4] Fredholm (1900), Fredholm (1903) page 365, Michel and Herget (1993) page 97, Keener (1988) page 101

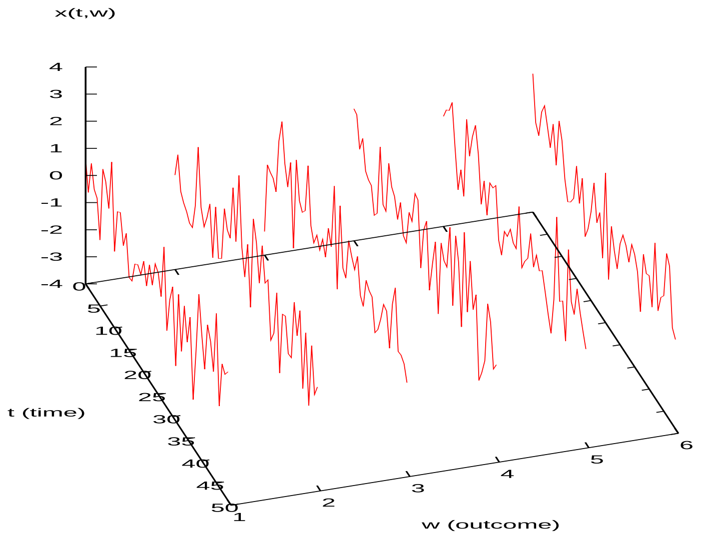


Figure 3.2: Example of a random process $\mathsf{x}(t, \omega)$

3.2 Properties

Theorem 3.1. *Let* $\mathsf{x}(t)$ *and* $\mathsf{y}(t)$ *be random processes with cross-correlation* $\mathsf{R}_{\mathsf{xy}}(t,u)$ *and let* $\mathsf{R}_{\mathsf{xx}}(t,u)$ *be the auto-correlation of* $\mathsf{x}(t)$.

THM

$$\begin{aligned} \mathsf{R}_{\mathsf{xx}}(t,u) &= \mathsf{R}^*_{\mathsf{xx}}(u,t) \quad \text{(CONJUGATE SYMMETRIC)} \\ \mathsf{R}_{\mathsf{xy}}(t,u) &= \mathsf{R}^*_{\mathsf{yx}}(u,t) \end{aligned}$$

✎PROOF:

$$\begin{aligned} \mathsf{R}_{\mathsf{xx}}(t,u) &\triangleq \mathbf{E}\left[\mathsf{x}(t)\mathsf{x}^*(u)\right] = \mathbf{E}\left[\mathsf{x}^*(u)\mathsf{x}(t)\right] = \left(\mathbf{E}\left[\mathsf{x}(u)\mathsf{x}^*(t)\right]\right)^* \triangleq \mathsf{R}^*_{\mathsf{xx}}(u,t) \\ \mathsf{R}_{\mathsf{xy}}(t,u) &\triangleq \mathbf{E}\left[\mathsf{x}(t)\mathsf{y}^*(u)\right] = \mathbf{E}\left[\mathsf{y}^*(u)\mathsf{x}(t)\right] = \left(\mathbf{E}\left[\mathsf{y}(u)\mathsf{x}^*(t)\right]\right)^* \triangleq \mathsf{R}^*_{\mathsf{yx}}(u,t) \end{aligned}$$

⇨

CHAPTER 4

KL EXPANSION—CONTINUOUS CASE

4.1 Definitions

Definition 4.1. *Let* $\mathsf{x}(t)$ *be a* RANDOM PROCESS *with continuous* AUTO-CORRELATION $\mathsf{R_{xx}}(t,u)$ *(Definition 3.2 page 17).*

DEF
The ***auto-correlation operator*** $\mathbf{R}$ *of* $\mathsf{x}(t)$ *is defined as*

$$\mathbf{R}\mathsf{f} \triangleq \int_{u\in\mathbb{R}} \mathsf{R_{xx}}(t,u)\mathsf{f}(u)\,\mathsf{d}u$$

Definition 4.2. *Let* $\mathsf{x}(t)$ *be a* RANDOM PROCESS *with* AUTO-CORRELATION $\mathsf{R_{xx}}(\tau)$ *(Definition 3.2 page 17).*

DEF
A RANDOM PROCESS $\mathsf{x}(t)$ *is* ***white*** *if*

$$\mathsf{R_{xx}}(\tau) = \delta(\tau)$$

If a random process $\mathsf{x}(t)$ is *white* (Definition 4.2 page 19) and the set $\Psi = \{\psi_1(t), \psi_2(t), \ldots, \psi_N(t)\}$ is **any** set of orthonormal basis functions, then the innerproducts $\langle n(t) \,|\, \psi_n(t)\rangle$ and $\langle n(t) \,|\, \psi_m(t)\rangle$ are *uncorrelated* for $m \neq n$. However, if $\mathsf{x}(t)$ is **colored** (not white), then the innerproducts are not in general uncorrelated. But if the elements of Ψ are chosen to be the eigenfunctions of $\mathbf{R}$ such that $\mathbf{R}\psi_n = \lambda_n\psi_n$, then by Theorem 3.1 (page 18), the set $\{\psi_n(t)\}$ are *orthogonal* and the innerproducts **are** *uncorrelated* even though $\mathsf{x}(t)$ is not white. This criterion is called the *Karhunen-Loève criterion* for $\mathsf{x}(t)$.

4.2 Properties

Theorem 4.1. *Let* $\mathbf{R}$ *be an* AUTO-CORRELATION *operator.*

THM

$$\left\{ \langle \mathsf{x} \,|\, \mathsf{y}\rangle \triangleq \int_{t\in\mathbb{R}} \mathsf{x}(t)\mathsf{y}^*(t)\,\mathsf{d}t \right\} \implies \left\{ \begin{array}{lll} (1). & \langle \mathbf{R}\mathsf{x} \,|\, \mathsf{x}\rangle \geq 0 & \text{(NON-NEGATIVE) and} \\ (2). & \langle \mathbf{R}\mathsf{x} \,|\, \mathsf{y}\rangle = \langle \mathsf{x} \,|\, \mathbf{R}\mathsf{y}\rangle & \text{(SELF-ADJOINT)} \end{array} \right\}$$

✎PROOF:

1. Proof that $\mathbf{R}$ is *non-negative* under hypothesis (A):

$$\begin{aligned}
\langle \mathbf{R}\mathsf{y} \,|\, \mathsf{y}\rangle &= \left\langle \int_{u\in\mathbb{R}} \mathsf{R_{xx}}(t,u)\mathsf{y}(u)\,\mathsf{d}u \,\middle|\, \mathsf{y}(t) \right\rangle && \text{by definition of } \mathbf{R} && \text{(Definition 4.1 page 19)}\\
&= \left\langle \int_{u\in\mathbb{R}} \mathbf{E}\left[\mathsf{x}(t)\mathsf{x}^*(u)\right]\mathsf{y}(u)\,\mathsf{d}u \,\middle|\, \mathsf{y}(t) \right\rangle && \text{by definition of } \mathsf{R_{xx}}(t,u) && \text{(Definition 3.2 page 17)}\\
&= \mathbf{E}\left[\left\langle \int_{u\in\mathbb{R}} \mathsf{x}(t)\mathsf{x}^*(u)\mathsf{y}(u)\,\mathsf{d}u \,\middle|\, \mathsf{y}(t) \right\rangle\right] && \text{by } \textit{linearity} \text{ of } \langle\triangle\,|\,\nabla\rangle \text{ and } \int && \text{(Definition N.1 page 309)}\\
&= \mathbf{E}\left[\int_{u\in\mathbb{R}} \mathsf{x}^*(u)\mathsf{y}(u)\,\mathsf{d}u\, \langle \mathsf{x}(t) \,|\, \mathsf{y}(t)\rangle\right] && \text{by } \textit{additivity} \text{ property of } \langle\triangle\,|\,\nabla\rangle && \text{(Definition N.1 page 309)}\\
&= \mathbf{E}\left[\langle \mathsf{y}(u) \,|\, \mathsf{x}(u)\rangle \langle \mathsf{x}(t) \,|\, \mathsf{y}(t)\rangle\right] && \text{by local definition of } \langle\triangle\,|\,\nabla\rangle\\
&= \mathbf{E}\left[\langle \mathsf{x}(u) \,|\, \mathsf{y}(u)\rangle^* \langle \mathsf{x}(t) \,|\, \mathsf{y}(t)\rangle\right] && \text{by } \textit{conjugate symmetry} \text{ prop.} && \text{(Definition N.1 page 309)}\\
&= \mathbf{E}\left|\langle \mathsf{x}(t) \,|\, \mathsf{y}(t)\rangle\right|^2 && \text{by definition of } |\cdot| && \text{(Definition G.4 page 222)}\\
&\ge 0 && \text{by } \textit{strictly positive} \text{ property of norms} && \text{(Definition O.1 page 327)}
\end{aligned}$$

2. Proof that $\mathbf{R}$ is *self-adjoint* under hypothesis (A):

$$\begin{aligned}
\langle [\mathbf{R}\mathsf{x}](t) \,|\, \mathsf{y}\rangle &= \left\langle \int_{u\in\mathbb{R}} \mathsf{R_{xx}}(t,u)\mathsf{x}(u)\,\mathsf{d}u \,\middle|\, \mathsf{y}(t) \right\rangle && \text{by definition of } \mathbf{R} && \text{(Definition 4.1 page 19)}\\
&= \int_{u\in\mathbb{R}} \mathsf{x}(u)\,\langle \mathsf{R_{xx}}(t,u) \,|\, \mathsf{y}(t)\rangle\,\mathsf{d}u && \text{by } \textit{additive} \text{ property of } \langle\triangle\,|\,\nabla\rangle && \text{(Definition N.1 page 309)}\\
&= \int_{u\in\mathbb{R}} \mathsf{x}(u)\,\langle \mathsf{y}(t) \,|\, \mathsf{R_{xx}}(t,u)\rangle^*\,\mathsf{d}u && \text{by } \textit{conjugate symmetry} \text{ prop.} && \text{(Definition N.1 page 309)}\\
&= \langle \mathsf{x}(u) \,|\, \langle \mathsf{y}(t) \,|\, \mathsf{R_{xx}}(t,u)\rangle\rangle && \text{by local definition of } \langle\triangle\,|\,\nabla\rangle\\
&= \left\langle \mathsf{x}(u) \,\middle|\, \int_{t\in\mathbb{R}} \mathsf{y}(t)\mathsf{R^*_{xx}}(t,u)\,\mathsf{d}t \right\rangle && \text{(Definition N.1 page 309)}\\
&= \left\langle \mathsf{x}(u) \,\middle|\, \int_{t\in\mathbb{R}} \mathsf{y}(t)\mathsf{R_{xx}}(u,t)\,\mathsf{d}t \right\rangle && \text{by property of } \mathsf{R_{xx}} && \text{(Theorem 3.1 page 18)}\\
&= \left\langle \mathsf{x}(u) \,\middle|\, \underbrace{\mathbf{R}}_{\mathbf{R}^*}\mathsf{y} \right\rangle && \text{by definition of } \mathbf{R} && \text{(Definition 4.1 page 19)}\\
\Longrightarrow \mathbf{R} &= \mathbf{R}^* && \text{by definition of } \textit{adjoint } \mathbf{R}^* && \text{(Definition R.6 page 369)}\\
\Longrightarrow \mathbf{R} &\text{ is } \textit{self-adjoint} && \text{by definition of } \textit{self-adjoint} && \text{(Definition R.8 page 376)}
\end{aligned}$$

3. Proofs under hypothesis (B): substitute $\sum_{n\in\mathbb{Z}}$ operator for $\int_{t\in\mathbb{R}}\,\mathsf{d}t$ operator in above proofs.

✏

Theorem 4.2. [1] *Let $(\lambda_n)_{n\in\mathbb{Z}}$ be the eigenvalues and $(\psi_n)_{n\in\mathbb{Z}}$ be the eigenfunctions of operator $\mathbf{R}$ such that $\mathbf{R}\psi_n = \lambda_n\psi_n$.*

T H M			
	(1).	$\lambda_n \in \mathbb{R}$	(REAL-VALUED)
	(2).	$\lambda_n \neq \lambda_m \implies \langle \psi_n \,\|\, \psi_m\rangle = 0$	(ORTHOGONAL)
	(3).	$\left\|\psi_n(t)\right\|^2 > 0 \implies \lambda_n \ge 0$	(NON-NEGATIVE)
	(4).	$\left\|\psi_n(t)\right\|^2 > 0, \langle \mathbf{R}\mathsf{f} \,\|\, \mathsf{f}\rangle > 0 \implies \lambda_n > 0$	($\mathbf{R}$ POSITIVE DEFINITE $\implies \lambda_n$ POSITIVE)

✎PROOF:

[1] Keener (1988) pages 114–119

1. Proof that eigenvalues are *real-valued*: Because $\mathbf{R}$ is *self-adjoint*, its eigenvalues are real (Theorem R.18 page 377).

2. Proof that eigenfunctions associated with distinct eigenvalues are orthogonal: Because $\mathbf{R}$ is *self-adjoint*, this property follows (Theorem R.18 page 377).

3. Proof that eigenvalues are *non-negative*:

$$\begin{aligned}
0 &\le \langle \mathbf{R}\psi_n \,|\, \psi_n \rangle && \text{by definition of \textit{non-negative definite}} \\
&= \langle \lambda_n \psi_n \,|\, \psi_n \rangle && \text{by hypothesis} \\
&= \lambda_n \langle \psi_n \,|\, \psi_n \rangle && \text{by definition of inner-products} \\
&= \lambda_n \|\psi_n\|^2 && \text{by definition of norm induced by inner-product}
\end{aligned}$$

4. Proof that eigenvalues are *positive* if $\mathbf{R}$ is *positive definite*:

$$\begin{aligned}
0 &< \langle \mathbf{R}\psi_n \,|\, \psi_n \rangle && \text{by definition of \textit{positive definite}} \\
&= \langle \lambda_n \psi_n \,|\, \psi_n \rangle && \text{by hypothesis} \\
&= \lambda_n \langle \psi_n \,|\, \psi_n \rangle && \text{by \textit{homogeneous} property of } \langle \triangle \,|\, \nabla \rangle && \text{(Definition N.1 page 309)} \\
&= \lambda_n \|\psi_n\|^2 && \text{by \textit{induced norm} theorem} && \text{(Theorem N.4 page 314)}
\end{aligned}$$

✏

Theorem 4.3 (Karhunen-Loève Expansion). [2] *Let* $\mathbf{R}$ *be the* AUTO-CORRELATION OPERATOR (*Definition 4.1 page 19*) *of a* RANDOM PROCESS $\mathsf{x}(t)$. *Let* $(\!(\lambda_n)\!)_{n\in\mathbb{Z}}$ *be the eigenvalues of* $\mathbf{R}$ *and* $(\!(\psi_n)\!)_{n\in\mathbb{Z}}$ *are the eigenfunctions of* $\mathbf{R}$ *such that* $\mathbf{R}\psi_n = \lambda_n\psi_n$.

THM

$$\underbrace{\|\psi_n(t)\| = 1}_{\{\psi_n(t)\} \text{ are NORMALIZED}} \implies \underbrace{\mathbf{E}\left(\left|\mathsf{x}(t) - \sum_{n\in\mathbb{Z}} \langle \mathsf{x}(t) \,|\, \psi_n(t)\rangle \psi_n(t)\right|^2\right) = 0}_{\text{CONVERGENCE IN PROBABILITY}} \qquad (\{\psi_n(t)\} \text{ is a BASIS for } \mathsf{x}(t))$$

✎PROOF:

1. Define $\dot{x}_n \triangleq \langle \mathsf{x}(t) \,|\, \psi_n(t) \rangle$

2. lemma: $\mathbf{E}[\mathsf{x}(t)\mathsf{x}(t)] = \sum_{n\in\mathbb{Z}} \lambda_n |\psi_n(t)|^2$. Proof:

$$\mathbf{E}[\mathsf{x}(t)\mathsf{x}(t)] = \sum_{n\in\mathbb{Z}} \lambda_n |\psi_n(t)|^2 \qquad \begin{array}{lll} \text{by} & \textit{non-negative}\text{ property} & \text{(Theorem 5.1 page 25)} \\ \text{and} & \textit{Mercer's Theorem} & \text{(Theorem E.4 page 208)} \end{array}$$

[2] Keener (1988) pages 114–119

3. lemma:

$$\begin{aligned}
&\mathbf{E}\left[\mathsf{x}(t)\left(\sum_{n\in\mathbb{Z}}\dot{x}_n\psi_n(t)\right)^*\right] \\
&\triangleq \mathbf{E}\left[\mathsf{x}(t)\left(\sum_{n\in\mathbb{Z}}\int_{u\in\mathbb{R}}\mathsf{x}(u)\psi_n^*(u)\;\mathsf{d}u\psi_n(t)\right)^*\right] && \text{by definition of } \dot{x} && \text{(definition 1 page 27)} \\
&= \sum_{n\in\mathbb{Z}}\left(\int_{u\in\mathbb{R}}\mathbf{E}\left[\mathsf{x}(t)\mathsf{x}^*(u)\right]\psi_n(u)\;\mathsf{d}u\right)\psi_n^*(t) && \text{by } \textit{linearity} && \text{(Theorem 1.1 page 4)} \\
&\triangleq \sum_{n\in\mathbb{Z}}\left(\int_{u\in\mathbb{R}}\mathsf{R}_{\mathsf{xx}}(t,u)\psi_n(u)\;\mathsf{d}u\right)\psi_n^*(t) && \text{by definition of } \mathsf{R}_{\mathsf{xx}}(t,u) && \text{(Definition 3.2 page 17)} \\
&\triangleq \sum_{n\in\mathbb{Z}}\left(\mathbf{R}\psi_n(t)\psi_n^*(t)\right) && \text{by definition of } \mathbf{R} && \text{(Definition 4.1 page 19)} \\
&= \sum_{n\in\mathbb{Z}}\lambda_n\psi_n(t)\psi_n^*(t) && \text{by property of } \textit{eigen-system} \\
&= \sum_{n\in\mathbb{Z}}\lambda_n\left|\psi_n(t)\right|^2
\end{aligned}$$

4. lemma:

$$\begin{aligned}
&\mathbf{E}\left[\sum_{n\in\mathbb{Z}}\dot{x}_n\psi_n(t)\left(\sum_{m\in\mathbb{Z}}\dot{x}_m\psi_m(t)\right)^*\right] \\
&\triangleq \mathbf{E}\left[\sum_{n\in\mathbb{Z}}\int_{u\in\mathbb{R}}\mathsf{x}(u)\psi_n^*(u)\;\mathsf{d}u\psi_n(t)\left(\sum_{m\in\mathbb{Z}}\int_{v}\mathsf{x}(v)\psi_m^*(v)\;\mathsf{d}v\psi_m(t)\right)^*\right] && \text{by definition of } \dot{x} \text{ (definition 1 page 27)} \\
&= \sum_{n\in\mathbb{Z}}\sum_{m\in\mathbb{Z}}\int_u\left(\int_v\mathbf{E}\left[\mathsf{x}(u)\mathsf{x}^*(v)\right]\psi_m(v)\;\mathsf{d}v\right)\psi_n^*(u)\;\mathsf{d}u\psi_n(t)\psi_m^*(t) && \text{by } \textit{linearity} \text{ (Theorem 1.1 page 4)} \\
&= \sum_{n\in\mathbb{Z}}\sum_{m\in\mathbb{Z}}\int_u\left(\int_v\mathsf{R}_{\mathsf{xx}}(u,v)\psi_m(v)\;\mathsf{d}v\right)\psi_n^*(u)\;\mathsf{d}u\psi_n(t)\psi_m^*(t) && \text{by definition of } \mathsf{R}_{\mathsf{xx}}(t,u) \text{ (Definition 3.2 page 17)} \\
&= \sum_{n\in\mathbb{Z}}\sum_{m\in\mathbb{Z}}\int_u\left(\mathbf{R}\psi_m(u)\right)\psi_n^*(u)\;\mathsf{d}u\psi_n(t)\psi_m^*(t) && \text{by definition of } \mathbf{R} \text{ (Definition 4.1 page 19)} \\
&= \sum_{n\in\mathbb{Z}}\sum_{m\in\mathbb{Z}}\int_u\left(\lambda_m\psi_m(u)\right)\psi_n^*(u)\;\mathsf{d}u\psi_n(t)\psi_m^*(t) && \text{by property of } \textit{eigen-system} \\
&= \sum_{n\in\mathbb{Z}}\sum_{m\in\mathbb{Z}}\lambda_m\left(\int_{u\in\mathbb{R}}\psi_m(u)\psi_n^*(u)\;\mathsf{d}u\right)\psi_n(t)\psi_m^*(t) && \text{by } \textit{linearity} \\
&= \sum_{n\in\mathbb{Z}}\sum_{m\in\mathbb{Z}}\lambda_m\left\|\psi(t)\right\|^2\bar{\delta}_{mn}\psi_n(t)\psi_m^*(t) && \text{by } \textit{orthogonal} \text{ property (Theorem 4.2 page 20)} \\
&= \sum_{n\in\mathbb{Z}}\sum_{m\in\mathbb{Z}}\lambda_m\bar{\delta}_{mn}\psi_n(t)\psi_m^*(t) && \text{by } \textit{normalized} \text{ hypothesis} \\
&= \sum_{n\in\mathbb{Z}}\lambda_n\psi_n(t)\psi_n^*(t) \qquad \text{by definition of } \textit{Kronecker delta } \bar{\delta} && \text{(Definition N.3 page 323)} \\
&= \sum_{n\in\mathbb{Z}}\lambda_n\left|\psi_n(t)\right|^2
\end{aligned}$$

5. Proof that $\{\psi_n(t)\}$ is a *basis* for $\mathsf{x}(t)$:

$$\mathbf{E}\left(\left|\mathsf{x}(t)-\sum_{n\in\mathbb{Z}}\dot{x}_n\psi_n(t)\right|^2\right)$$

$$
\begin{aligned}
&= \mathbf{E}\left(\left[\mathsf{x}(t) - \sum_{n\in\mathbb{Z}} \dot{x}_n\psi_n(t)\right]\left[\mathsf{x}(t) - \sum_{m\in\mathbb{Z}} \dot{x}_m\psi_m(t)\right]^*\right) \\
&= \mathbf{E}\left(\mathsf{x}(t)\mathsf{x}^*(t) - \mathsf{x}(t)\left[\sum_{n\in\mathbb{Z}} \dot{x}_n\psi_n(t)\right]^* - \mathsf{x}^*(t)\sum_{n\in\mathbb{Z}} \dot{x}_n\psi_n(t) + \left[\sum_{n\in\mathbb{Z}} \dot{x}_n\psi_n(t)\right]\left[\sum_{m\in\mathbb{Z}} \dot{x}_m\psi_m(t)\right]^*\right) \\
&= \mathbf{E}\left(\mathsf{x}(t)\mathsf{x}^*(t)\right) - \mathbf{E}\left[\mathsf{x}(t)\left[\sum_{n\in\mathbb{Z}} \dot{x}_n\psi_n(t)\right]^*\right] - \mathbf{E}\left[\mathsf{x}^*(t)\sum_{n\in\mathbb{Z}} \dot{x}_n\psi_n(t)\right] + \mathbf{E}\left[\sum_{n\in\mathbb{Z}} \dot{x}_n\psi_n(t)\left[\sum_{m\in\mathbb{Z}} \dot{x}_m\psi_m(t)\right]^*\right] \\
&\qquad \text{by } \textit{linearity} \text{ of } \mathbf{E} \text{ (Theorem 1.1 page 4)} \\
&= \underbrace{\sum_{n\in\mathbb{Z}} \lambda_n|\psi_n(t)|^2}_{\text{by (2) lemma}} - \underbrace{\sum_{n\in\mathbb{Z}} \lambda_n|\psi_n(t)|^2}_{\text{by (3) lemma}} - \underbrace{\left[\sum_{n\in\mathbb{Z}} \lambda_n|\psi_n(t)|^2\right]^*}_{\text{by (3) lemma}} + \underbrace{\sum_{n\in\mathbb{Z}} \lambda_n|\psi_n(t)|^2}_{\text{by (4) lemma}} \\
&= 0
\end{aligned}
$$

✎

4.3 Quasi-basis

The *auto-correlation operator* $\mathbf{R}$ (Definition 4.1 page 19) in the discrete case can be approximated using a *correlation matrix*. In the *zero-mean* case, this becomes

$$
\mathbf{R} \triangleq \begin{bmatrix}
\mathbf{E}[\mathsf{y}_1\mathsf{y}_1] & \mathbf{E}[\mathsf{y}_1\mathsf{y}_2] & \cdots & \mathbf{E}[\mathsf{y}_1\mathsf{y}_n] \\
\mathbf{E}[\mathsf{y}_2\mathsf{y}_1] & \mathbf{E}[\mathsf{y}_2\mathsf{y}_2] & & \mathbf{E}[\mathsf{y}_2\mathsf{y}_n] \\
\vdots & \vdots & \ddots & \vdots \\
\mathbf{E}[\mathsf{y}_n\mathsf{y}_1] & \mathbf{E}[\mathsf{y}_n\mathsf{y}_2] & \cdots & \mathbf{E}[\mathsf{y}_n\mathsf{y}_n]
\end{bmatrix}
$$

The eigen-vectors (and hence a quasi-basis) for $\mathbf{R}$ can be found using a *Cholesky Decomposition*.

Proposition 4.1. [3]

PRP *The* AUTO-CORRELATION MATRIX $\mathbf{R}$ *is* ***Toeplitz***.

Remark 4.1. For more information about the properties of **Toeplitz matrices**, see

1. Grenander and Szegö (1958),
2. Widom (1965),
3. Gray (1971),
4. Smylie et al. (1973) page 408 ⟨§"B. Properties of the Toeplitz Matrix"⟩,
5. Grenander and Szegö (1984),
6. Haykin and Kesler (1979),
7. Haykin and Kesler (1983),
8. S. Lawrence Marple (1987) pages 80–92 ⟨§"3.8 The Toeplitz Matrix"⟩,
9. Böttcher and Silbermann (1999) ⟨ISBN:9780387985701⟩,
10. Gray (2006),
11. S. Lawrence Marple (2019) pages 80–93 ⟨§"3.8 The Toeplitz Matrix"⟩.

[3] See Clarkson (1993) page 131 ⟨§"Appendix 3A — Positive Semi-Definite Form of the Autocorrelation Matrix"⟩

CHAPTER 5

KL EXPANSION—DISCRETE CASE

5.1 Definitions

Definition 5.1. *Let* $\mathsf{x}(n)$ *and* $\mathsf{y}(n)$ *be* RANDOM PROCESSS. *Let* $\mathsf{R}_{\mathsf{xx}}(n,m)$ *be the* AUTO-CORRELATION *(Definition 2.2 page 11) of* $\mathsf{x}(n)$.

DEF *The **auto-correlation operator*** $\mathbf{R}_{\mathsf{x}}$ *of* $\mathsf{y}(n)$ *is defined as*

$$\mathbf{R}_{\mathsf{x}}\mathsf{y}(n) \triangleq \sum_{m\in\mathbb{Z}} \mathsf{R}_{\mathsf{xx}}(m,n)\mathsf{y}(m)$$

Definition 5.2. *Let* $\mathsf{x}(n)$ *and* $\mathsf{y}(n)$ *be* RANDOM PROCESSS. *Let* $\mathsf{R}_{\mathsf{xx}}(n,m)$ *be the* AUTO-CORRELATION *of* $\mathsf{x}(n)$.

DEF *A* RANDOM PROCESS $\mathsf{x}(n)$ *is **white** if*

$$\mathsf{R}_{\mathsf{xx}}(m) = K\bar{\delta}(m) \quad \text{for some } K > 0.$$

5.2 Properties

Theorem 5.1. *Let* $\mathbf{R}_{\mathsf{x}}$ *be an* AUTO-CORRELATION *operator.*

THM
$$\left\{ \langle \mathsf{x} \,|\, \mathsf{y} \rangle \triangleq \sum_{n\in\mathbb{Z}} \mathsf{x}(n)\mathsf{y}^*(n) \right\} \implies \left\{ \begin{array}{lll} (1). & \langle \mathbf{R}_{\mathsf{x}}\mathsf{x} \,|\, \mathsf{x} \rangle \geq 0 & (\text{NON-NEGATIVE}) \quad \text{and} \\ (2). & \langle \mathbf{R}_{\mathsf{x}}\mathsf{x} \,|\, \mathsf{y} \rangle = \langle \mathsf{x} \,|\, \mathbf{R}_{\mathsf{x}}\mathsf{y} \rangle & (\text{SELF-ADJOINT}) \end{array} \right\}$$

✎PROOF:

1. Proof that $\mathbf{R}_{\mathsf{x}}$ is *non-negative*:

$$\begin{aligned}
\langle \mathbf{R}_{\mathsf{x}}\mathsf{y} \,|\, \mathsf{y} \rangle &= \left\langle \sum_{m\in\mathbb{Z}} \mathsf{R}_{\mathsf{xx}}(n,m)\mathsf{y}(m) \,\middle|\, \mathsf{y}(n) \right\rangle && \text{by definition of } \mathbf{R}_{\mathsf{x}} && \text{(Definition 5.1 page 25)} \\
&= \left\langle \sum_{m\in\mathbb{Z}} \mathbf{E}\left[\mathsf{x}(n)\mathsf{x}^*(m)\right]\mathsf{y}(m) \,\middle|\, \mathsf{y}(n) \right\rangle && \text{by definition of } \mathsf{R}_{\mathsf{xx}}(n,m) && \text{(Definition 3.2 page 17)} \\
&= \mathbf{E}\left[\left\langle \sum_{m\in\mathbb{Z}} \mathsf{x}(n)\mathsf{x}^*(m)\mathsf{y}(m) \,\middle|\, \mathsf{y}(n) \right\rangle\right] && \text{by } \textit{linearity} \text{ of } \langle \triangle \,|\, \triangledown \rangle \text{ and } \sum && \text{(Definition N.1 page 309)}
\end{aligned}$$

$$= \mathbf{E}\left[\sum_{m\in\mathbb{Z}} \mathsf{x}^*(m)\mathsf{y}(m)\,\langle \mathsf{x}(n)\,|\,\mathsf{y}(n)\rangle\right]$$ by *additivity* property of $\langle \triangle \,|\, \nabla \rangle$ (Definition N.1 page 309)

$$= \mathbf{E}\left[\langle \mathsf{y}(m)\,|\,\mathsf{x}(m)\rangle\,\langle \mathsf{x}(n)\,|\,\mathsf{y}(n)\rangle\right]$$ by local definition of $\langle \triangle \,|\, \nabla \rangle$

$$= \mathbf{E}\left[\langle \mathsf{x}(m)\,|\,\mathsf{y}(m)\rangle^*\,\langle \mathsf{x}(n)\,|\,\mathsf{y}(n)\rangle\right]$$ by *conjugate symmetry* prop. (Definition N.1 page 309)

$$= \mathbf{E}|\langle \mathsf{x}(n)\,|\,\mathsf{y}(n)\rangle|^2$$ by definition of $|\cdot|$ (Definition G.4 page 222)

$$\geq 0$$ by *strictly positive* property of norms (Definition O.1 page 327)

2. Proof that $\mathbf{R}_{\mathsf{x}}$ is *self-adjoint*:

$$\left\langle [\mathbf{R}_{\mathsf{x}}\mathsf{x}](n)\,|\,\mathsf{y}\right\rangle = \left\langle \sum_{m\in\mathbb{Z}} \mathsf{R}_{\mathsf{xx}}(n,m)\mathsf{x}(m)\,\Big|\,\mathsf{y}(n)\right\rangle$$ by definition of $\mathbf{R}_{\mathsf{x}}$ (Definition 5.1 page 25)

$$= \sum_{m\in\mathbb{Z}} \mathsf{x}(m)\,\langle \mathsf{R}_{\mathsf{xx}}(n,m)\,|\,\mathsf{y}(n)\rangle$$ by *additive* property of $\langle \triangle \,|\, \nabla \rangle$ (Definition N.1 page 309)

$$= \sum_{m\in\mathbb{Z}} \mathsf{x}(m)\,\langle \mathsf{y}(n)\,|\,\mathsf{R}_{\mathsf{xx}}(n,m)\rangle^*$$ by *conjugate symmetry* prop. (Definition N.1 page 309)

$$= \langle \mathsf{x}(m)\,|\,\langle \mathsf{y}(n)\,|\,\mathsf{R}_{\mathsf{xx}}(n,m)\rangle\rangle$$ by local definition of $\langle \triangle \,|\, \nabla \rangle$

$$= \left\langle \mathsf{x}(m)\,\Big|\,\sum_{n\in\mathbb{Z}} \mathsf{y}(n)\mathsf{R}^*_{\mathsf{xx}}(n,m)\right\rangle$$ (Definition N.1 page 309)

$$= \left\langle \mathsf{x}(m)\,\Big|\,\sum_{n\in\mathbb{Z}} \mathsf{y}(n)\mathsf{R}_{\mathsf{xx}}(m,n)\right\rangle$$ by property of R_{xx} (Theorem 3.1 page 18)

$$= \left\langle \mathsf{x}(m)\,\Big|\,\underbrace{\mathbf{R}_{\mathsf{x}}}_{\mathbf{R}^*_{\mathsf{x}}}\mathsf{y}\right\rangle$$ by definition of $\mathbf{R}_{\mathsf{x}}$ (Definition 5.1 page 25)

$$\implies \mathbf{R}_{\mathsf{x}} = \mathbf{R}^*_{\mathsf{x}}$$ by definition of *adjoint* $\mathbf{R}^*_{\mathsf{x}}$ (Definition R.6 page 369)

$$\implies \mathbf{R}_{\mathsf{x}} \text{ is } \textit{self-adjoint}$$ by definition of *self-adjoint* (Definition R.8 page 376)

✑

Theorem 5.2. *Let $(\lambda_n)_{n\in\mathbb{Z}}$ be the* EIGENVALUES *and $(\psi_n)_{n\in\mathbb{Z}}$ be the* EIGENFUNCTIONS *of operator $\mathbf{R}_{\mathsf{x}}$ such that $\mathbf{R}_{\mathsf{x}}\psi_n = \lambda_n\psi_n$ for all $n \in \mathbb{Z}$.*

(1).	$\lambda_n \in \mathbb{R}$			(REAL-VALUED)
(2).	$\lambda_n \neq \lambda_m$	$\implies$	$\langle \psi_n \,|\, \psi_m \rangle = 0$	(ORTHOGONAL)
(3).	$\|\psi_n\|^2 > 0$	$\implies$	$\lambda_n \geq 0$	(NON-NEGATIVE)
(4).	$\|\psi_n(t)\|^2 > 0, \langle \mathbf{R}_{\mathsf{x}}\mathsf{f} \,|\, \mathsf{f} \rangle > 0$	$\implies$	$\lambda_n > 0$	($\mathbf{R}_{\mathsf{x}}$ POSITIVE DEFINITE $\implies$ λ_n POSITIVE)

✎PROOF:

1. Proof that eigenvalues are *real-valued*:

$\mathbf{R}_{\mathsf{x}}$ is *self-adjoint* — by Theorem 5.1 page 25

$\implies$ eigenvalues of $\mathbf{R}_{\mathsf{x}}$ are *real* — (Theorem R.18 page 377)

2. Proof that eigenfunctions associated with distinct eigenvalues are orthogonal: Because $\mathbf{R}_{\mathsf{x}}$ is *self-adjoint*, this property follows (Theorem R.18 page 377).

3. Proof that eigenvalues are *non-negative*:

$$0 \leq \langle \mathbf{R}_{\mathsf{x}}\psi_n \,|\, \psi_n \rangle$$ by definition of *non-negative definite*

$$= \langle \lambda_n\psi_n \,|\, \psi_n \rangle$$ by definition of *eigenvalue* ($\mathbf{R}_{\mathsf{x}}\psi_n = \lambda_n\psi_n$)

$$= \lambda_n\,\langle \psi_n \,|\, \psi_n \rangle$$ by *homogeneous* property of *inner products* (Definition N.1 page 309)

$$= \lambda_n\,\|\psi_n\|^2$$ by definition of norm induced by inner-product (Definition O.1 page 327)

4. Proof that eigenvalues are *positive* if $\mathbf{R}_\mathsf{x}$ is *positive definite*:

$$\begin{aligned}
0 &< \langle \mathbf{R}_\mathsf{x}\psi_n \,|\, \psi_n \rangle && \text{by definition of } \textit{positive definite} \\
&= \langle \lambda_n\psi_n \,|\, \psi_n \rangle && \text{by hypothesis} \\
&= \lambda_n \langle \psi_n \,|\, \psi_n \rangle && \text{by } \textit{homogeneous} \text{ property of } \langle \triangle \,|\, \nabla \rangle && \text{(Definition N.1 page 309)} \\
&= \lambda_n \|\psi_n\|^2 && \text{by } \textit{induced norm} \text{ theorem} && \text{(Theorem N.4 page 314)}
\end{aligned}$$

⇰

Theorem 5.3 (Karhunen-Loève Expansion). [1] *Let* $\mathbf{R}_\mathsf{x}$ *be the* AUTO-CORRELATION OPERATOR *(Definition 5.1 page 25) of a* RANDOM PROCESS $\mathsf{x}(n)$. *Let* $(\!(\lambda_n)\!)_{n\in\mathbb{Z}}$ *be the eigenvalues of* $\mathbf{R}_\mathsf{x}$ *and* $(\!(\psi_n)\!)_{n\in\mathbb{Z}}$ *are the eigenfunctions of* $\mathbf{R}_\mathsf{x}$ *such that* $\mathbf{R}_\mathsf{x}\psi_n = \lambda_n\psi_n$.

THM

$$\underbrace{\|\psi_n\| = 1}_{\{\psi_n(p)\} \text{ are NORMALIZED}} \implies \underbrace{\mathbf{E}\left(\left|\mathsf{x}(m) - \sum_{n\in\mathbb{Z}} \langle \mathsf{x}(m) \,|\, \psi_n(m) \rangle\, \psi_n(m)\right|^2\right) = 0}_{\text{CONVERGENCE IN PROBABILITY}} \qquad (\{\psi_n(m)\} \text{ is a BASIS for } \mathsf{x}(m))$$

✎PROOF:

1. Define $\dot{x}_n \triangleq \langle \mathsf{x}(m) \,|\, \psi_n(m) \rangle \triangleq \sum_{m\in\mathbb{Z}} \mathsf{x}(m)\psi_n(m)$

2. lemma: $\mathbf{E}[\mathsf{x}(m)\mathsf{x}(m)] = \sum_{n\in\mathbb{Z}} \lambda_n |\psi_n(m)|^2$. Proof: by *non-negative* property (Theorem 5.1 page 25) and *Mercer's Theorem* (Theorem E.4 page 208)

3. lemma:

$$\begin{aligned}
&\mathbf{E}\left[\mathsf{x}(p)\left(\sum_{n\in\mathbb{Z}} \dot{x}_n\psi_n(p)\right)^*\right] \\
&\triangleq \mathbf{E}\left[\mathsf{x}(p)\left(\sum_{n\in\mathbb{Z}} \int_{u\in\mathbb{R}} \mathsf{x}(u)\psi_n^*(u)\;\mathsf{d}u\psi_n(p)\right)^*\right] && \text{by definition of } \dot{x} && \text{(definition 1 page 27)} \\
&= \sum_{n\in\mathbb{Z}} \left(\int_{u\in\mathbb{R}} \mathbf{E}[\mathsf{x}(p)\mathsf{x}^*(u)]\psi_n(u)\;\mathsf{d}u\right)\psi_n^*(p) && \text{by } \textit{linearity} && \text{(Theorem 1.1 page 4)} \\
&\triangleq \sum_{n\in\mathbb{Z}} \left(\int_{u\in\mathbb{R}} \mathsf{R}_{\mathsf{xx}}(p,u)\psi_n(u)\;\mathsf{d}u\right)\psi_n^*(p) && \text{by definition of } \mathsf{R}_{\mathsf{xx}}(p,u) && \text{(Definition 3.2 page 17)} \\
&\triangleq \sum_{n\in\mathbb{Z}} (\mathbf{R}_\mathsf{x}\psi_n(p)\psi_n^*(p)) && \text{by definition of } \mathbf{R}_\mathsf{x} && \text{(Definition 5.1 page 25)} \\
&= \sum_{n\in\mathbb{Z}} \lambda_n\psi_n(p)\psi_n^*(p) && \text{by property of } \textit{eigen-system} \\
&= \sum_{n\in\mathbb{Z}} \lambda_n|\psi_n(p)|^2
\end{aligned}$$

[1] Keener (1988) pages 114–119

4. lemma:

$$
\begin{aligned}
&\mathbf{E}\left[\sum_{n\in\mathbb{Z}} \dot{x}_n\psi_n(p)\left(\sum_{m\in\mathbb{Z}} \dot{x}_m\psi_m(p)\right)^*\right] \\
&\triangleq \mathbf{E}\left[\sum_{n\in\mathbb{Z}} \int_{u\in\mathbb{R}} \mathsf{x}(u)\psi_n^*(u)\,\mathsf{d}u\psi_n(p)\left(\sum_{m\in\mathbb{Z}} \int_v \mathsf{x}(v)\psi_m^*(v)\,\mathsf{d}v\psi_m(p)\right)^*\right] && \text{by definition of } \dot{x} \text{ (definition 1 page 27)} \\
&= \sum_{n\in\mathbb{Z}}\sum_{m\in\mathbb{Z}} \int_u \left(\int_v \mathbf{E}\left[\mathsf{x}(u)\mathsf{x}^*(v)\right]\psi_m(v)\,\mathsf{d}v\right)\psi_n^*(u)\,\mathsf{d}u\psi_n(p)\psi_m^*(p) && \text{by \textit{linearity} (Theorem 1.1 page 4)} \\
&= \sum_{n\in\mathbb{Z}}\sum_{m\in\mathbb{Z}} \int_u \left(\int_v \mathsf{R}_{\mathsf{xx}}(u,v)\psi_m(v)\,\mathsf{d}v\right)\psi_n^*(u)\,\mathsf{d}u\psi_n(p)\psi_m^*(p) && \text{by definition of } \mathsf{R}_{\mathsf{xx}}(p,u) \text{ (Definition 3.2 page 17)} \\
&= \sum_{n\in\mathbb{Z}}\sum_{m\in\mathbb{Z}} \int_u \left(\mathbf{R}_{\mathsf{x}}\psi_m(u)\right)\psi_n^*(u)\,\mathsf{d}u\psi_n(p)\psi_m^*(p) && \text{by definition of } \mathbf{R}_{\mathsf{x}} \text{ (Definition 5.1 page 25)} \\
&= \sum_{n\in\mathbb{Z}}\sum_{m\in\mathbb{Z}} \int_u \left(\lambda_m\psi_m(u)\right)\psi_n^*(u)\,\mathsf{d}u\psi_n(p)\psi_m^*(p) && \text{by property of \textit{eigen-system}} \\
&= \sum_{n\in\mathbb{Z}}\sum_{m\in\mathbb{Z}} \lambda_m\left(\int_{u\in\mathbb{R}} \psi_m(u)\psi_n^*(u)\,\mathsf{d}u\right)\psi_n(p)\psi_m^*(p) && \text{by \textit{linearity}} \\
&= \sum_{n\in\mathbb{Z}}\sum_{m\in\mathbb{Z}} \lambda_m\left\|\psi(p)\right\|^2\bar{\delta}_{mn}\psi_n(p)\psi_m^*(p) && \text{by \textit{orthogonal} property (Theorem 4.2 page 20)} \\
&= \sum_{n\in\mathbb{Z}}\sum_{m\in\mathbb{Z}} \lambda_m\bar{\delta}_{mn}\psi_n(p)\psi_m^*(p) && \text{by \textit{normalized} hypothesis} \\
&= \sum_{n\in\mathbb{Z}} \lambda_n\psi_n(p)\psi_n^*(p) \qquad \text{by definition of \textit{Kronecker delta} } \bar{\delta} && \text{(Definition N.3 page 323)} \\
&= \sum_{n\in\mathbb{Z}} \lambda_n\left|\psi_n(p)\right|^2
\end{aligned}
$$

5. Proof that $\{\psi_n(p)\}$ is a *basis* for $\mathsf{x}(p)$:

$$
\begin{aligned}
&\mathbf{E}\left(\left|\mathsf{x}(p) - \sum_{n\in\mathbb{Z}} \dot{x}_n\psi_n(p)\right|^2\right) \\
&= \mathbf{E}\left(\left[\mathsf{x}(p) - \sum_{n\in\mathbb{Z}} \dot{x}_n\psi_n(p)\right]\left[\mathsf{x}(p) - \sum_{m\in\mathbb{Z}} \dot{x}_m\psi_m(p)\right]^*\right) \\
&= \mathbf{E}\left(\mathsf{x}(p)\mathsf{x}^*(p) - \mathsf{x}(p)\left[\sum_{n\in\mathbb{Z}} \dot{x}_n\psi_n(p)\right]^* - \mathsf{x}^*(p)\sum_{n\in\mathbb{Z}} \dot{x}_n\psi_n(p) + \left[\sum_{n\in\mathbb{Z}} \dot{x}_n\psi_n(p)\right]\left[\sum_{m\in\mathbb{Z}} \dot{x}_m\psi_m(p)\right]^*\right) \\
&= \mathbf{E}\left(\mathsf{x}(p)\mathsf{x}^*(p)\right) - \mathbf{E}\left[\mathsf{x}(p)\left[\sum_{n\in\mathbb{Z}} \dot{x}_n\psi_n(p)\right]^*\right] - \mathbf{E}\left[\mathsf{x}^*(p)\sum_{n\in\mathbb{Z}} \dot{x}_n\psi_n(p)\right] + \mathbf{E}\left[\sum_{n\in\mathbb{Z}} \dot{x}_n\psi_n(p)\left[\sum_{m\in\mathbb{Z}} \dot{x}_m\psi_m(p)\right]^*\right] \\
&\qquad \text{by \textit{linearity} of } \mathbf{E} \text{ (Theorem 1.1 page 4)} \\
&= \underbrace{\sum_{n\in\mathbb{Z}} \lambda_n\left|\psi_n(p)\right|^2}_{\text{by (2) lemma}} - \underbrace{\sum_{n\in\mathbb{Z}} \lambda_n\left|\psi_n(p)\right|^2}_{\text{by (3) lemma}} - \underbrace{\left[\sum_{n\in\mathbb{Z}} \lambda_n\left|\psi_n(p)\right|^2\right]^*}_{\text{by (3) lemma}} + \underbrace{\sum_{n\in\mathbb{Z}} \lambda_n\left|\psi_n(p)\right|^2}_{\text{by (4) lemma}} \\
&= 0
\end{aligned}
$$

✎

5.3 Quasi-basis

The *auto-correlation operator* $\mathbf{R}_x$ (Definition 5.1 page 25) in the discrete case can be approximated using a *correlation matrix*. In the *zero-mean* case, this becomes

$$\mathbf{R}_x \triangleq \begin{bmatrix} \mathbf{E}[y_1y_1] & \mathbf{E}[y_1y_2] & \cdots & \mathbf{E}[y_1y_n] \\ \mathbf{E}[y_2y_1] & \mathbf{E}[y_2y_2] & & \mathbf{E}[y_2y_n] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{E}[y_ny_1] & \mathbf{E}[y_ny_2] & \cdots & \mathbf{E}[y_ny_n] \end{bmatrix}$$

The eigen-vectors (and hence a quasi-basis) for $\mathbf{R}_x$ can be found using a *Cholesky Decomposition*.

Proposition 5.1. [2]

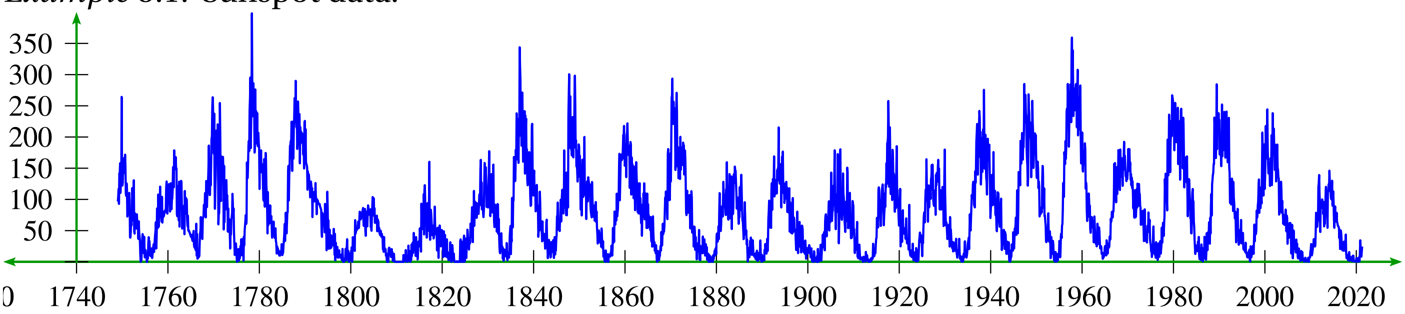
PRP *The* AUTO-CORRELATION MATRIX $\mathbf{R}_x$ *is* ***Toeplitz***.

Remark 5.1. For more information about the properties of **Toeplitz matrices**, see

1. Grenander and Szegö (1958),
2. Widom (1965),
3. Gray (1971),
4. Smylie et al. (1973) page 408 ⟨§"B. Properties of the Toeplitz Matrix"⟩,
5. Grenander and Szegö (1984),
6. Haykin and Kesler (1979),
7. Haykin and Kesler (1983),
8. S. Lawrence Marple (1987) pages 80–92 ⟨§"3.8 The Toeplitz Matrix"⟩,
9. Böttcher and Silbermann (1999) ⟨ISBN:9780387985701⟩,
10. Gray (2006),
11. S. Lawrence Marple (2019) pages 80–93 ⟨§"3.8 The Toeplitz Matrix"⟩.

5.4 Examples

Example 5.1. Sunspot data:[3]



The period the sunspot activity can be estimated using several methods:

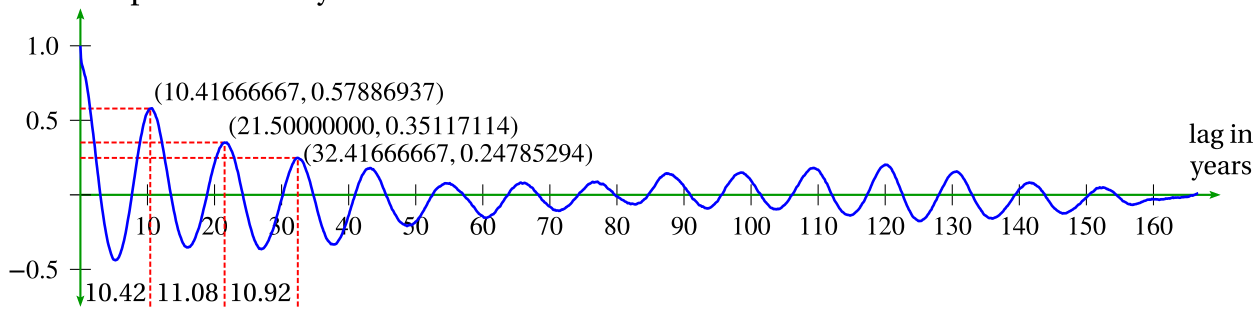
1. Period estimation using an estimated *auto-correlation function* (ACF); Here other than at lag=0, the greatest[4] correlation occurs at 10.42 years …implying a sunspot oscillation with

[2] See Clarkson (1993) page 131 ⟨§"Appendix 3A — Positive Semi-Definite Form of the Autocorrelation Matrix"⟩

[3] SILSO World Data Center (2021)

[4] $20\log_{10}\left(\frac{0.5918556724}{0.3730703718}\right) \approx 4.01\text{dB}$ greater

estimated period 10.42 years.

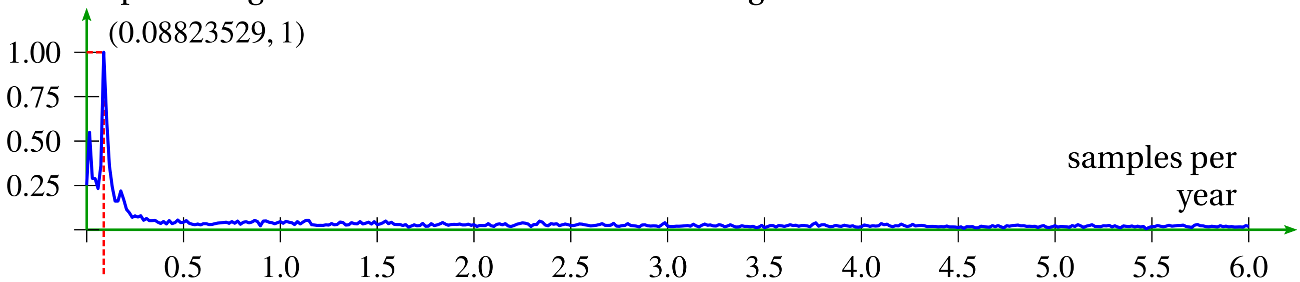


2. Period estimation using the *power spectral density* (PSD) of sunspot data, PSD estimated in turn using the *Welch Method*:

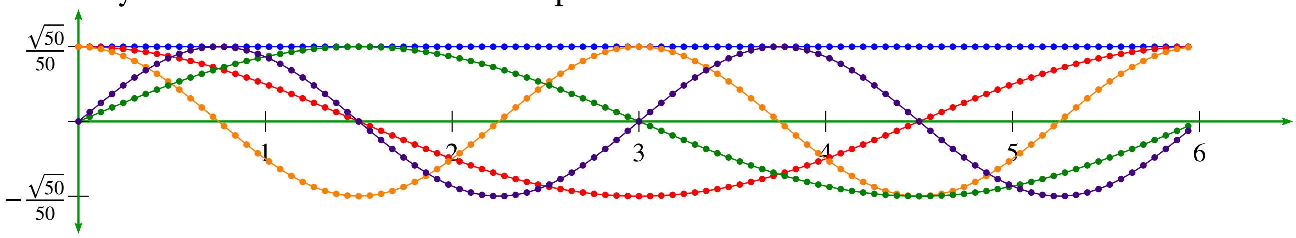
number of segments	peak frequency (samples/year)	period (years)
2	0.08812729	11.34722222
3	0.08823529	11.33333333
4	0.08823529	11.33333333
5	0.09202454	10.86666667
6	0.08823529	11.33333333
7	0.10300429	9.70833333
8	0.08823529	11.33333333
9	0.09944751	10.05555556
10	0.11042945	9.05555556

$$\begin{aligned} \text{mode period} &\approx 11.3 \text{ years} \\ \text{median period} &\approx 11.3 \text{ years} \\ \text{mean period} &\approx 10.7 \text{ years} \end{aligned}$$

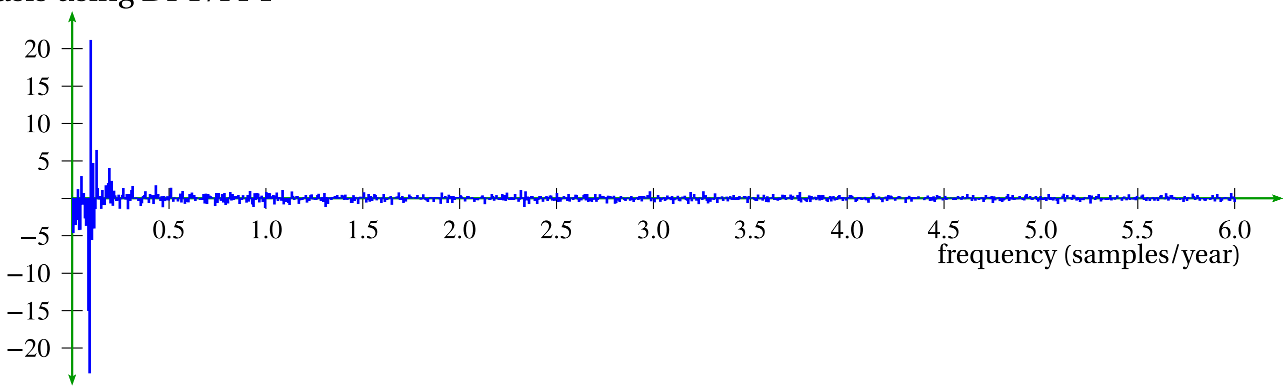
Here is a plot using Welch Method with number of segments = 4:

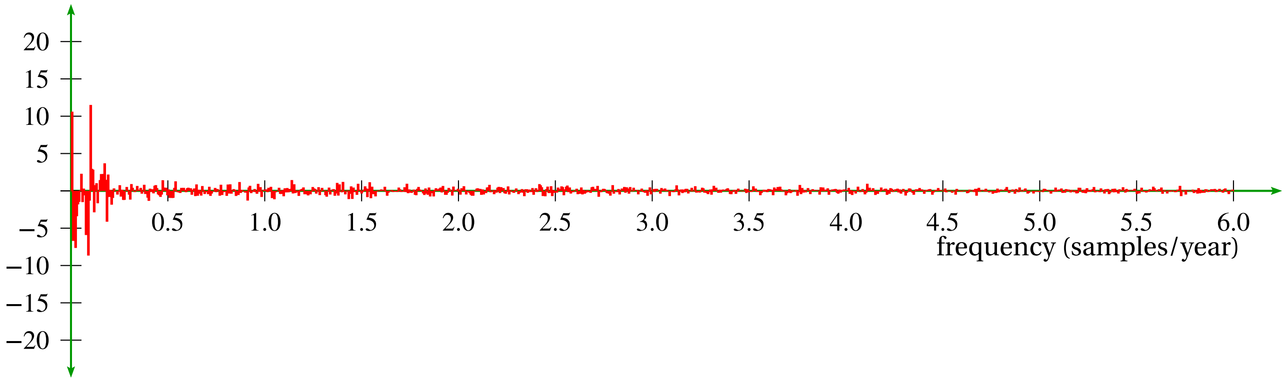


3. DFT analysis: The basis vectors are sampled sinusoids.

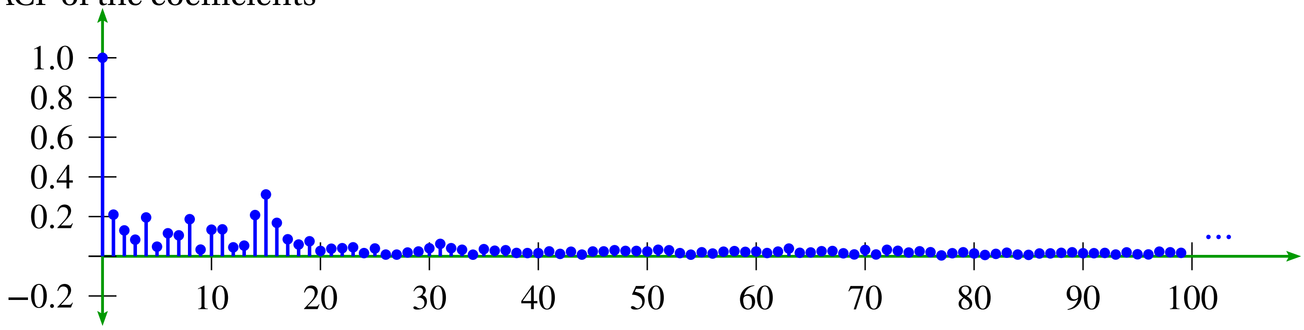


Coefficients can be calculated by projecting scaled zero-mean sunspot data onto sinusoidal basis using DFT/FFT

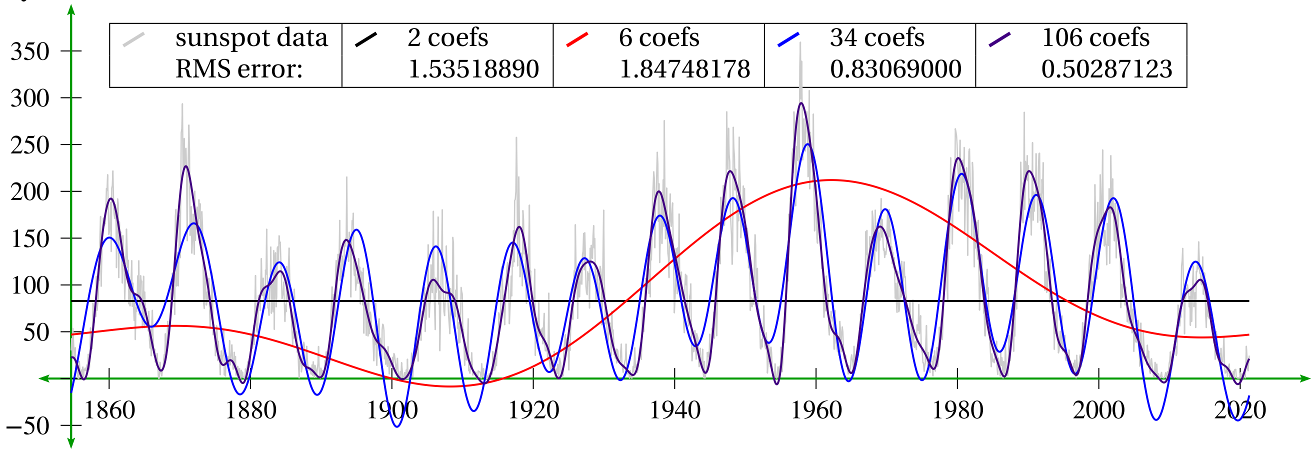




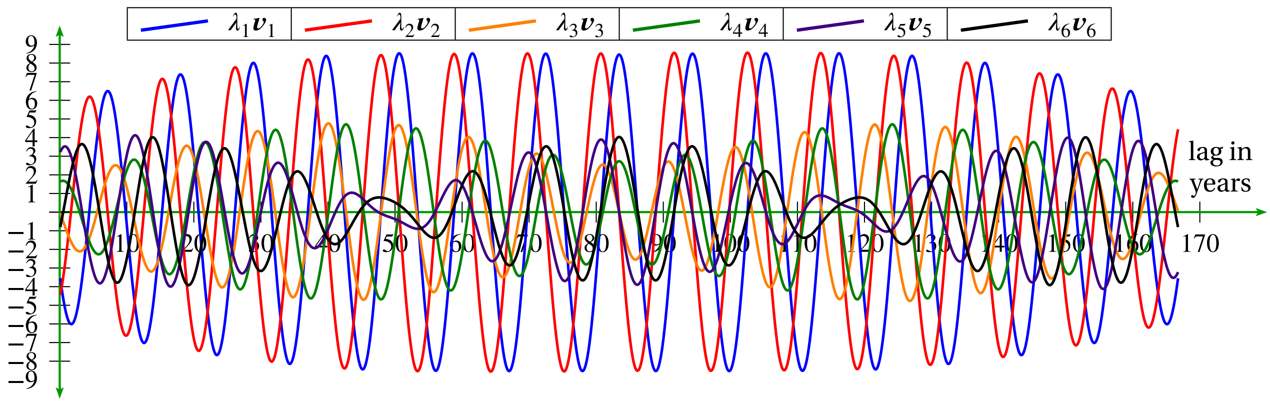
ACF of the coefficients



Synthesis



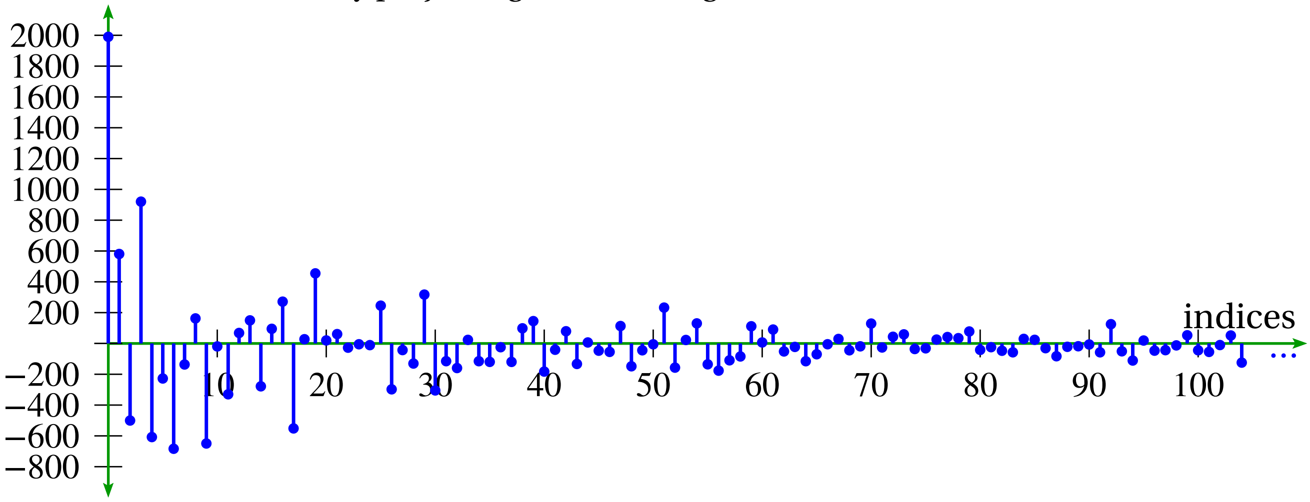
4. Eigen analysis



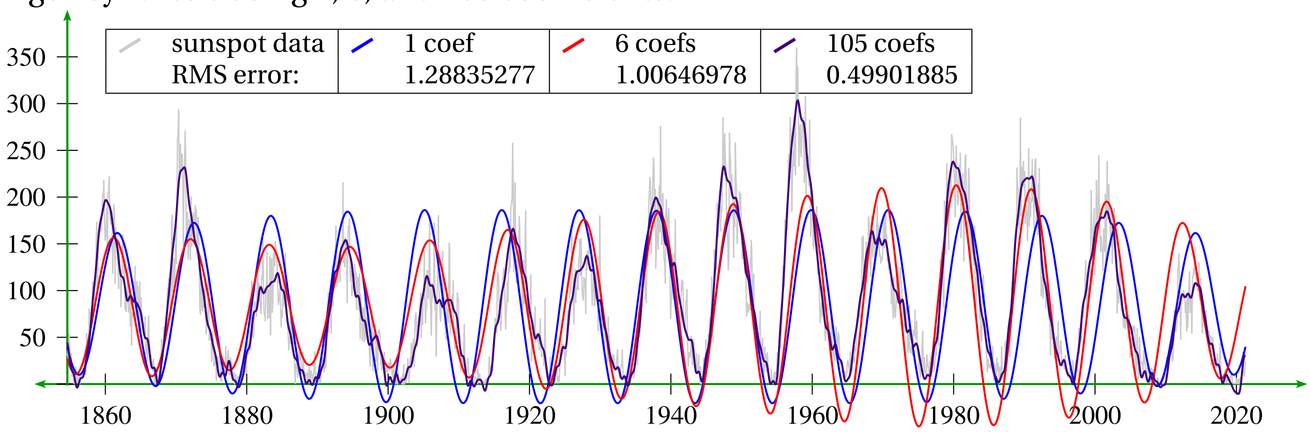
The DFT primary frequencies of the first 10 of these eigen values are

n	λ_n	f (Hz)	T (years)	Weighted period using 1–n (years)		
1	253.543147	0.089955	11.116667	weighted period using 1 :	11.11666667	years
2	250.981384	0.089955	11.116667	weighted period using 1–2 :	11.11666667	years
3	119.265072	0.095952	10.421875	weighted period using 1–3 :	10.98382640	years
4	119.253338	0.095952	10.421875	weighted period using 1–4 :	10.89363702	years
5	96.288166	0.101949	9.808824	weighted period using 1–5 :	10.76918709	years
6	95.069013	0.083958	11.910714	weighted period using 1–6 :	10.88532991	years
7	73.982222	0.011994	83.375000	weighted period using 1–7 :	16.20369641	years
8	67.216471	0.011994	83.375000	weighted period using 1–8 :	20.40137489	years
9	51.990321	0.083958	11.910714	weighted period using 1–9 :	20.00989172	years
10	50.999349	0.083958	11.910714	weighted period using 1–10:	19.65942777	years
11	34.936578	0.017991	55.583333	weighted period using 1–11:	20.69365308	years
12	33.830843	0.017991	55.583333	weighted period using 1–12:	21.63993254	years

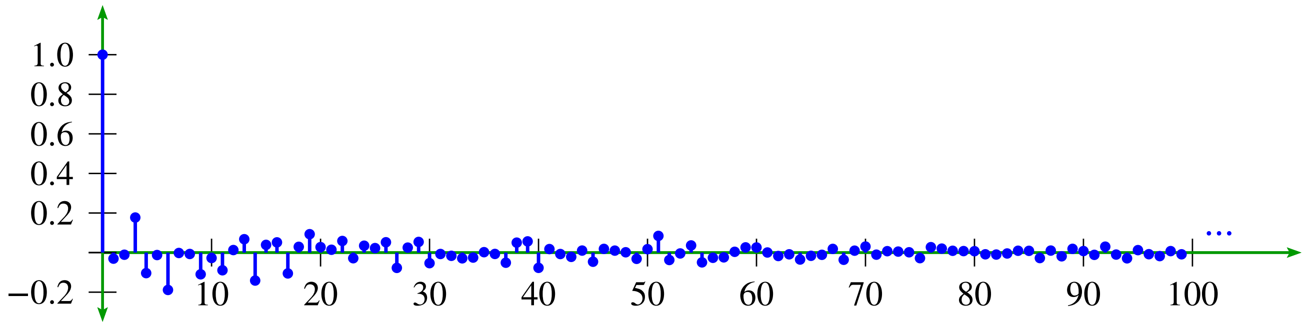
5. Coefficients calculated by projecting data onto eigen vectors:



6. Eigen synthesis using 1, 6, and 105 coefficients:

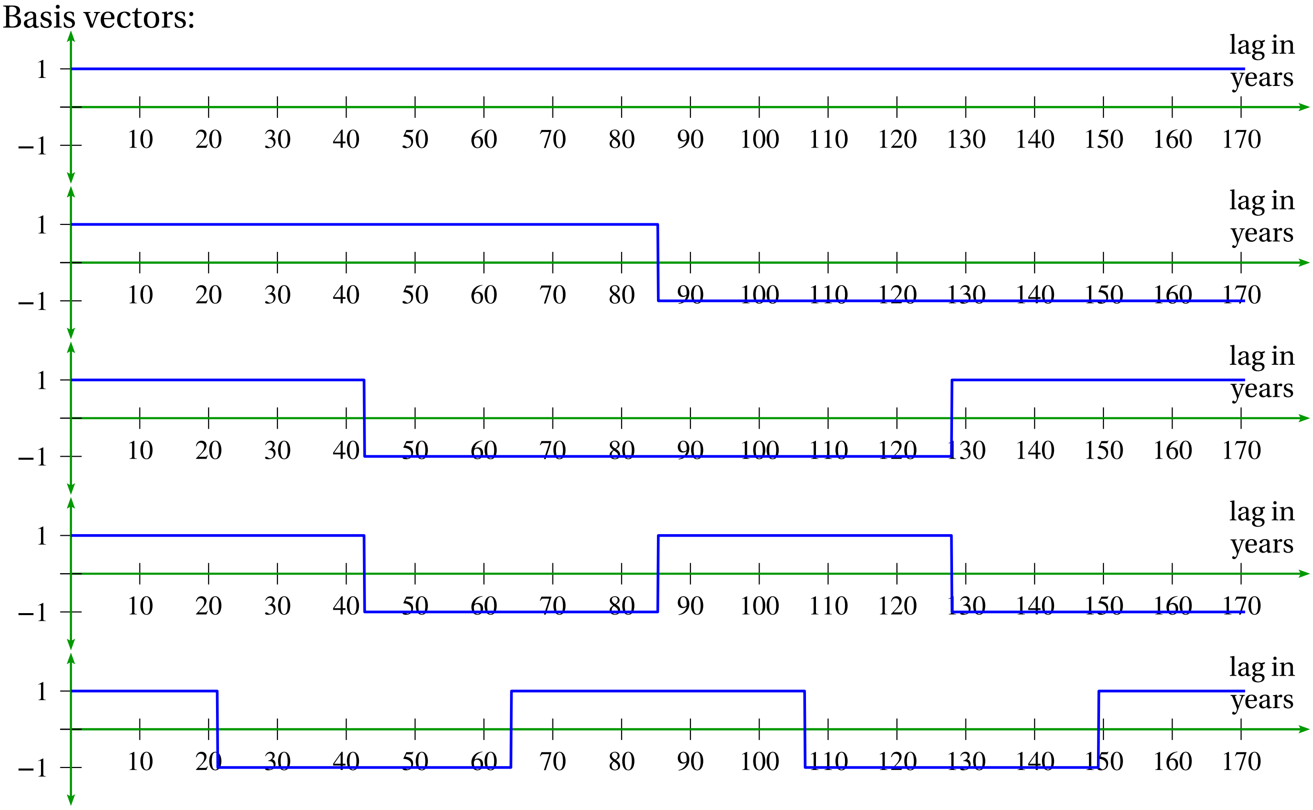


7. The ACF of the eigen coefficient sequence demonstrates that it is much whiter than the original sunspot data sequence:

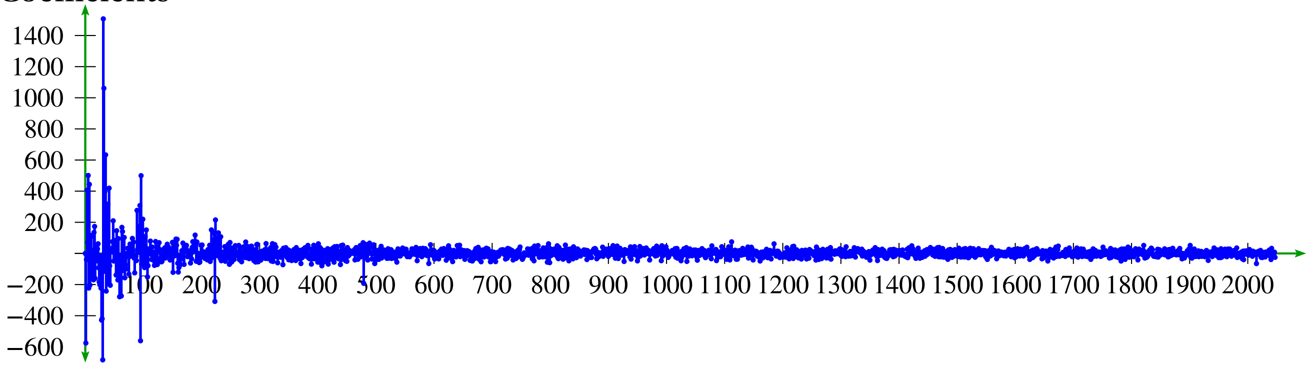


8. Walsh Sequency analysis

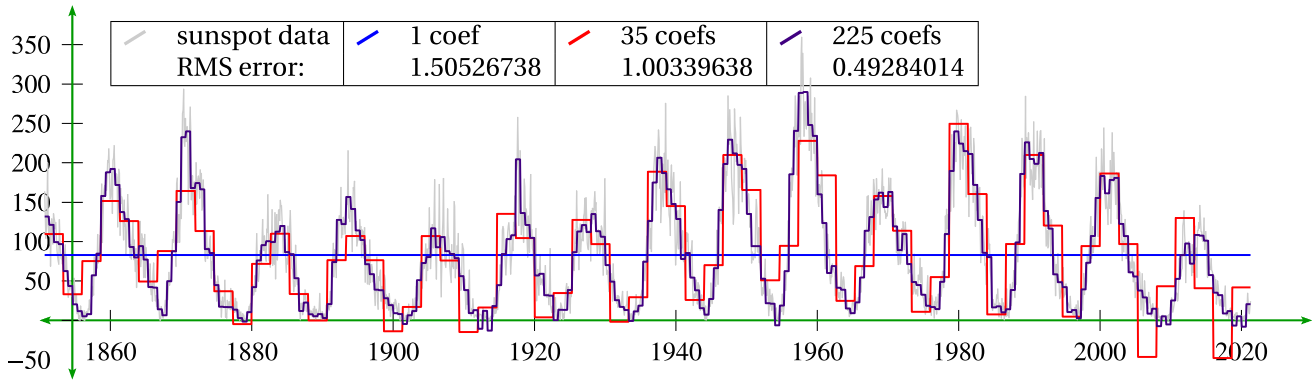
Basis vectors:



Coefficients



Synthesis


sunspot data
RMS error:
1 coef
1.50526738
35 coefs
1.00339638
225 coefs
0.49284014
350
300
250
200
150
100
50
−50
1860
1880
1900
1920
1940
1960
1980
2000
2020

Part II

Statistical Processing

CHAPTER 6

OPERATIONS ON RANDOM VARIABLES

6.1 Functions of one random variable

Proposition 6.1. *Let* $(\Omega, \mathbb{E}, \mathsf{P})$ *be a probability space, and* X *a* RANDOM VARIABLE *with* CUMULATIVE DISTRIBUTION FUNCTION $\mathsf{c_x}(x)$ *(Definition B.2 page 184).*

PRP

$$\left\{\begin{array}{l}\mathsf{X}\ \textit{is}\ \text{UNIFORMLY DISTRIBUTED}\\ \text{(Definition C.1 page 189)}\end{array}\right\} \iff \mathsf{c_x}(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } 0 \le x \le 1 \\ 1 & \text{otherwise}\end{cases}$$

Theorem 6.1 (Probability integral transform). [1] *Let* $(\Omega, \mathbb{E}, \mathsf{P})$ *be a probability space. Let* X *be a* RANDOM VARIABLE *with* PROBABILITY DENSITY FUNCTION $\mathsf{p_x}(x)$ *(Definition B.2 page 184) and* CUMULATIVE DISTRIBUTION FUNCTION $\mathsf{c_x}(x)$. *Let* Y *be a* RANDOM VARIABLE CUMULATIVE DISTRIBUTION FUNCTION $\mathsf{c_Y}(y)$.

THM

$$\left\{\begin{array}{lll}(1). & \mathsf{Y} = \mathsf{c_x}(\mathsf{X}) & \textit{and}\\ (2). & \mathsf{p_x}(x)\ \textit{is}\ \text{CONTINUOUS} & \end{array}\right\} \implies \left\{\begin{array}{l}\mathsf{Y}\ \textit{is}\ \text{UNIFORMLY DISTRIBUTED}\\ \text{(Definition C.1 page 189)}\end{array}\right\}$$

✎PROOF:

$$\begin{aligned}
\mathsf{c_Y}(y) &\triangleq \mathsf{P}\{\mathsf{Y} \le y\} && \text{by definition of } cdf && \text{(Definition B.2 page 184)}\\
&= \mathsf{P}\left\{\mathsf{c_x}(\mathsf{X}) \le y\right\} && \text{by hypothesis (1)} &&\\
&= \mathsf{P}\left\{\mathsf{X} \le \mathsf{c_x}^{-1}(y)\right\} && \text{by hypothesis (2) and} && \text{Proposition A.2 page 176}\\
&\triangleq \mathsf{c_x}\left[\mathsf{c_x}^{-1}(y)\right] && \text{by definition of } cdf && \text{(Definition B.2 page 184)}\\
&= y &&&&\\
\implies &\quad \mathsf{Y} \text{ is } \textit{uniformly distributed} && \text{by} && \text{Proposition 6.1 page 37}
\end{aligned}$$

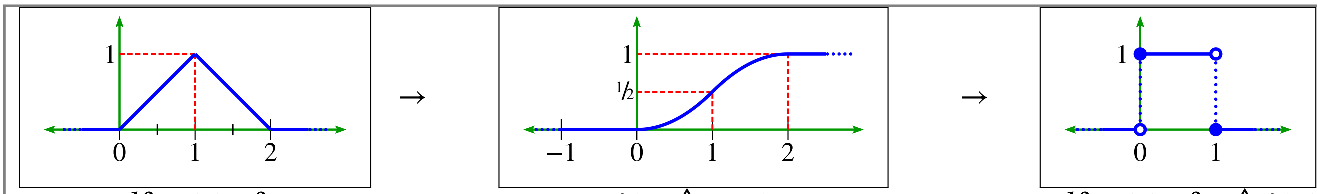
✎

Example 6.1.

EX Let X be a random variable with *cdf* (Definition B.2 page 184) $\mathsf{c_x}(x)$ and *pdf* (Definition B.2 page 184) $\mathsf{p_x}(x)$ where $\mathsf{p_x}(x)$ is the triangle-like *first order B-spline* $\mathsf{N}_1(x)$ (Example W.2 page 446). Then the random variable $\mathsf{Y} \triangleq \mathsf{c_x}(\mathsf{X})$ has *uniform distribution* (Definition C.1 page 189), as illustrated in Figure 6.1 (page 38).

---

[1] Angus (1994), Roussas (2014) page 232 ⟨Theorem 10⟩, Devroye (1986) page 28 ⟨Theorem 2.1⟩



pdf $\mathsf{p}_{\mathsf{X}}(x)$ of X $\qquad$ $\mathsf{f}(x) \triangleq \mathsf{c}_{\mathsf{X}}(x)$ $\qquad$ *pdf* $\mathsf{p}_{\mathsf{Y}}(x)$ of $\mathsf{Y} \triangleq \mathsf{f}(\mathsf{X})$

$$\mathsf{p}_{\mathsf{X}}(x) = \begin{cases} x & \text{for } 0 < x \le 1 \\ -x & \text{for } 1 < x \le 2 \\ 0 & \text{otherwise} \end{cases} \qquad \mathsf{f}(x) = \begin{cases} 0 & \text{for } x \le 0 \\ \frac{1}{2}x^2 & \text{for } 0 < x \le 1 \\ -\frac{1}{2}x^2 + 2x - 1 & \text{for } 1 < x \le 2 \\ 1 & \text{for } x > 2 \end{cases} \qquad \mathsf{p}_{\mathsf{Y}}(x) = \begin{cases} 1 & \text{for } 0 \le x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Figure 6.1: Flow diagram for Example 6.1 (page 37)

Here is some R code if you want to try out the concept empirically:[2]

```
N  = 1e6                                                    # N = number of samples
X1 = runif( n=N, min=0, max=1 )                             # X1 has uniform distribution
X2 = runif( n=N, min=0, max=1 )                             # X2 has uniform distribution
X  = X1 + X2                                                # X  has triangular distribution N1(x)
Y  = ifelse( X<1, X^2/2, -X^2/2+2*Y-1)                      # Y = cdf_x( X ) has uniform distribution
plot( density( Y, bw="SJ" ), lwd=3, xlim=c(0,4), col="blue" ) # plot estimated pdf of Y
```

✎PROOF:

$$\begin{aligned}
\mathsf{c}_{\mathsf{X}}(x) &\triangleq \int_{-\infty}^{x} \mathsf{p}_{\mathsf{X}}(u)\,\mathsf{d}u && \text{by definition of } \mathsf{c}_{\mathsf{X}}(x) \text{ (Definition B.2 page 184)} \\
&\triangleq \int_{-\infty}^{x} \mathsf{N}_1(u)\,\mathsf{d}u && \text{by definition of } \mathsf{p}_{\mathsf{X}}(x) \\
&= \int_{-\infty}^{x} \left\{ \begin{array}{ll} u & \text{for } u \in [0:1] \\ -u+2 & \text{for } u \in (1:2] \\ 0 & \text{otherwise} \end{array} \right\} \mathsf{d}u && \text{by definition of } \mathsf{N}_1(x) \text{ (Example W.2 page 446)} \\
&= \begin{cases} 0 & \text{for } x \le 0 \\ \int_0^x u\,\mathsf{d}u & \text{for } 0 < x \le 1 \\ \int_1^x -u+2\,\mathsf{d}u & \text{for } 1 < x \le 2 \\ 1 & \text{for } x > 2 \end{cases}
= \begin{cases} 0 & \text{for } x \le 0 \\ \left[\frac{1}{2}u^2\right]_0^x & \text{for } 0 < x \le 1 \\ \frac{1}{2} + \left[-\frac{1}{2}u^2 + 2u\right]_1^x & \text{for } 1 < x \le 2 \\ 1 & \text{for } x > 2 \end{cases} \\
&= \begin{cases} 0 & \text{for } x \le 0 \\ \frac{1}{2}x^2 - \frac{1}{2}0^2 & \text{for } 0 < x \le 1 \\ \frac{1}{2} + \left(-\frac{1}{2}x^2 + 2x\right) - \left(-\frac{1}{2}1^2 + 2\right) & \text{for } 1 < x \le 2 \\ 1 & \text{for } x > 2 \end{cases}
= \begin{cases} 0 & \text{for } x \le 0 \\ \frac{1}{2}x^2 & \text{for } 0 < x \le 1 \\ -\frac{1}{2}x^2 + 2x - 1 & \text{for } 1 < x \le 2 \\ 1 & \text{for } x > 2 \end{cases}
\end{aligned}$$

✏

Theorem 6.2 (Inverse probability integral transform). [3] *Let* $(\Omega, \mathbb{E}, \mathsf{P})$ *be a probability space. Let* X *be a* RANDOM VARIABLE *with* PROBABILITY DENSITY FUNCTION $\mathsf{p}_{\mathsf{X}}(x)$ *and* CUMULATIVE DISTRIBUTION FUNCTION $\mathsf{c}_{\mathsf{X}}(x)$. *Let* Y *be a* RANDOM VARIABLE CUMULATIVE DISTRIBUTION FUNCTION $\mathsf{c}_{\mathsf{Y}}(y)$.

THM
$$\left\{ \begin{array}{lll} (1). & \mathsf{Y} = \mathsf{c}_{\mathsf{Z}}^{-1}(\mathsf{X}) & \textit{and} \\ (2). & \mathsf{X} \textit{ is } \text{UNIFORMLY DISTRIBUTED} & \textit{and} \\ (3). & \mathsf{p}_{\mathsf{Z}}(z) \textit{ is } \text{CONTINUOUS} & \end{array} \right\} \implies \left\{ \begin{array}{l} \mathsf{p}_{\mathsf{Y}}(y) = \mathsf{p}_{\mathsf{Z}}(y) \\ (\mathsf{Y} \textit{ has distribution } \mathsf{p}_{\mathsf{Z}}(y)) \end{array} \right\}$$

[2]Note: To understand why $\mathsf{X} \triangleq \mathsf{X1} + \mathsf{X2}$ has triangle distribution $\mathsf{N}_1(x)$, see *Sum of Uniformly Distributed Random Variables* Example 6.6 page 49.

[3] Devroye (1986) page 28 ⟨Theorem 2.1⟩, Balakrishnan and Lai (2009) page 624 ⟨14.2.1 Introduction⟩

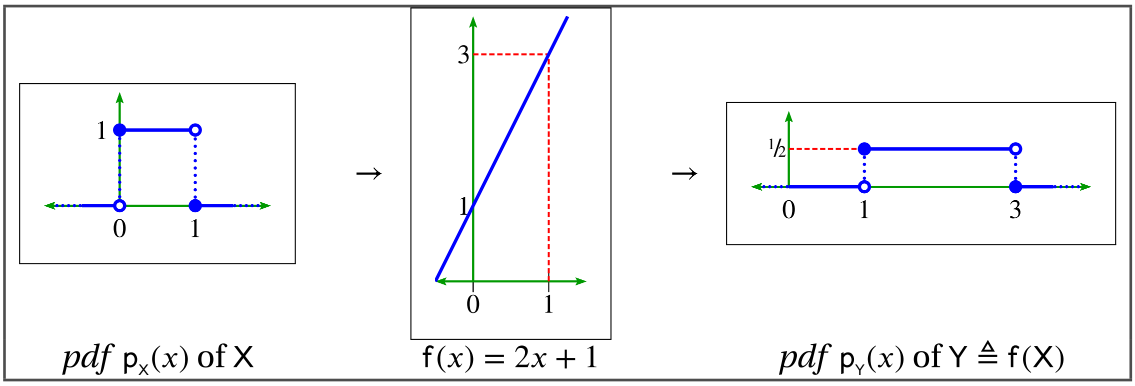


Figure 6.2: Flow diagram for Example 6.2 (page 39)

✎PROOF:

$$\begin{aligned}
\mathsf{c}_{\mathsf{Y}}(y) &\triangleq \mathsf{P}\{\mathsf{Y} \le y\} && \text{by definition of } \mathsf{c}_{\mathsf{Y}} && \text{(Definition B.2 page 184)}\\
&= \mathsf{P}\left\{\mathsf{c}_{\mathsf{Z}}^{-1}(\mathsf{X}) \le y\right\} && \text{by hypothesis (1)}\\
&= \mathsf{P}\left\{\mathsf{X} \le \mathsf{c}_{\mathsf{Z}}(y)\right\} && \text{by hypothesis (3) and} && \text{Proposition A.2 page 176}\\
&\triangleq \mathsf{c}_{\mathsf{X}}\left[\mathsf{c}_{\mathsf{Z}}(y)\right] && \text{by definition of } \mathsf{c}_{\mathsf{X}} && \text{(Definition B.2 page 184)}\\
&= \mathsf{c}_{\mathsf{Z}}(y) && \text{because } 0 \le \mathsf{c}_{\mathsf{Z}}(y) \le 1 \text{ and by} && \text{Proposition 6.1 page 37}\\
\implies \quad \mathsf{p}_{\mathsf{Y}}(y) &= \mathsf{p}_{\mathsf{Z}}(y) && \text{(Y has the distribution of Z)}
\end{aligned}$$

✏

Example 6.2. Let X be a *random variable* with *uniform distribution*. Select a function $\mathsf{f}(x)$ such that $\mathsf{Y} \triangleq \mathsf{f}(\mathsf{X})$ has *pdf* (Definition B.2 page 184) $\mathsf{p}_{\mathsf{Y}}(y) \triangleq \frac{1}{2}$ for $y \in (1:3]$ and 0 otherwise.

EX

$$\left\{\begin{array}{lll} (1) & \mathsf{p}_{\mathsf{X}}(x) \text{ is } \textit{uniform} & \text{and}\\ (2) & \mathsf{f}(x) \triangleq 2x + 1 & \end{array}\right\} \implies \left\{ \mathsf{Y} \triangleq \mathsf{f}(\mathsf{X}) \quad \text{has } \textit{pdf} \quad \mathsf{p}_{\mathsf{Y}}(y) \triangleq \left\{\begin{array}{ll} \frac{1}{2} & \text{for } y \in (1:3]\\ 0 & \text{otherwise} \end{array}\right. \right\}$$

as illustrated in Figure 6.2 (page 39).

Here's some R code demonstrating the concept:

```
N = 1e6                                                     # Number of samples
X = runif( n=N, min=0, max=1 )                              # X = Uniformly distributed RV
Y = 2 * X + 1                                               # Y = f( X )
plot( X, Y, ylim=c(0,3), col="red" )                        # plot X -> Y mapping
plot( density( Y, bw="SJ" ), lwd=3, xlim=c(0,4), col="blue" ) # plot estimated pdf of Y
```

✎PROOF:

1. The *cumulative distribution function* $\mathsf{c}_{\mathsf{Y}}(y)$ of Y with desired pdf $\mathsf{p}_{\mathsf{Y}}(y)$ is

$$\begin{aligned}
\mathsf{c}_{\mathsf{Y}}(y) &\triangleq \int_{-\infty}^{y} \mathsf{p}_{\mathsf{Y}}(u)\,\mathsf{d}u && \text{by definition of cdf} && \text{(Definition B.2 page 184)}\\
&\triangleq \int_{-\infty}^{y} \left\{\begin{array}{ll} \frac{1}{2} & \text{for } u \in [1:3]\\ 0 & \text{otherwise} \end{array}\right\} \mathsf{d}u && \text{by definition of } \mathsf{p}_{\mathsf{Y}}(y)\\
&= \left\{\begin{array}{ll} 0 & \text{for } y \le 1\\ \frac{1}{2}y & \text{for } 1 < y \le 3\\ 1 & \text{otherwise} \end{array}\right\}
\end{aligned}$$

2. The inverse cdf $\mathsf{c}_{\mathsf{Y}}^{-1}(x)$ is $\mathsf{c}_{\mathsf{Y}}^{-1}(x) = \left\{\begin{array}{lll} [0:1] & \text{for } x = 0 & \text{(undefined but in } [0:1])\\ 2x + 1 & \text{for } 0 < x \le 1 & \\ \text{undefined} & \text{otherwise} & \end{array}\right.$

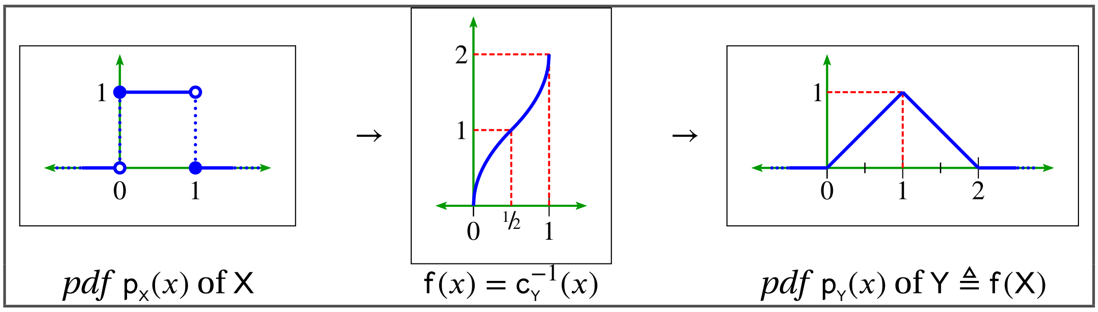


Figure 6.3: Flow diagram for Example 6.3 (page 40)

3. Since we are using $c_Y^{-1}(x)$ with X which only yields values of non-zero probablity in $(0 : 1]$ (Definition C.1 page 189), we can simplify $f(x)$ to only necessarily match c_Y^{-1} in the domain $(0 : 1]$, and be whatever is convenient elsewhere. As such, let $f(x) \triangleq 2x + 1$.

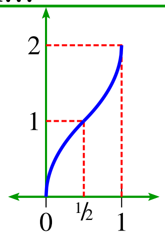
4. By the *Inverse probability integral transform* (Theorem 6.2 page 38), $Y \triangleq f(X) \triangleq 2X + 1$ has the desired pdf $p_Y(y)$.

5. Note: For an alternative proof using Corollary 6.1 (page 42) rather than the *Inverse probability integral transform* (Theorem 6.2 page 38), see Example 6.4 (page 43).

⇨

Example 6.3. Let X be a *random variable* with *uniform distribution*. Select a function $f(x)$ such that $Y \triangleq f(X)$ has distribution $N_1(x)$ as defined in Example 6.1 page 37 (triangle-like 1st order B-spline). Then...

EX



$$\left[f(x) \triangleq \begin{cases} \sqrt{2x} & \text{for } x \in \left[0 : \frac{1}{2}\right] \\ 2 - \sqrt{2(1-x)} & \text{for } x \in \left[\frac{1}{2} : 1\right] \\ \text{undefined} & \text{otherwise} \end{cases} \right] \implies \left[\begin{array}{l} Y \triangleq f(X) \\ \text{has } pdf \\ N_1(x) \end{array} \right]$$

...as illustrated in Figure 6.3 (page 40). Here's some R code demonstrating the concept (the "blue" plot should look approximately like $N_1(x)$):

```
N = 1e6                                                       # Number of samples
X = runif( n=N, min=0, max=1 )                                # X = Uniformly distributed RV
Y = ifelse( X<0.5, sqrt(2*X), (2-sqrt(2*(1-X))) )             # Y = f( X )
plot( X, Y, col="red" )                                       # plot X -> Y mapping
plot( density( Y, bw="SJ" ), lwd=3, xlim=c(0,4), col="blue" ) # plot estimated pdf of Y
```

✎PROOF:

1. $c_X(x) \triangleq y = \frac{1}{2}x^2 \implies x = \pm\sqrt{2y} \implies c_X^{-1}(x) = \pm\sqrt{2x}$

2. $c_X(x) \triangleq y = -\frac{1}{2}x^2 + 2x - 1 \implies c_X^{-1}(x) = 2 \pm \sqrt{2(1-x)}$

$$y = -\frac{1}{2}x^2 + 2x - 1 \implies x = \frac{-2 \pm \sqrt{2^2 - 4\left(-\frac{1}{2}\right)(-y-1)}}{2\left(-\frac{1}{2}\right)} \quad \text{by the } \textit{Quadratic Equation}$$

$$= 2 \pm \sqrt{4 - 4\left(-\frac{1}{2}\right)(-y-1)}$$

$$= 2 \pm \sqrt{4 - 2(y+1)}$$

$$= 2 \pm \sqrt{2(1-y)}$$

3. Taking (1) and (2) above and the fact that the *cdf* is always *non-negative*, we have

$$\mathsf{f}(x) \triangleq \mathsf{c}_X^{-1}(x) = \begin{cases} \sqrt{2x} & \text{for } x \in \left(0 : \frac{1}{2}\right] \\ 2 - \sqrt{2(1-x)} & \text{for } x \in \left[\frac{1}{2} : 1\right] \end{cases}$$

4. By the *Inverse probability integral transform* (Theorem 6.2 page 38), $\mathsf{Y} \triangleq \mathsf{f}(\mathsf{X})$ has the desired pdf $\mathsf{N}_1(x)$.

✏

Theorem 6.3. [4] *Let* X *and* Y *be* RANDOM VARIABLES *in* $\mathbb{R}^{\mathbb{R}}$. *Let* f *be a* DIFFERENTIABLE FUNCTION *in* $\mathbb{R}^{\mathbb{R}}$. *Let N be the number of solutions in x of the equation $y = \mathsf{f}(x)$ and $\{\mathsf{s}_n(y) | n = 1, 2, \ldots, N\}$ be those solutions. Let $\{A_n | n = 1, 2, \ldots, N\}$ be the partitions of the domain of $\mathsf{f}(x)$ such that each A_n is the interval over which $\mathsf{s}_n(y)$ is valid.*

$$\text{THM}\quad \left\{\begin{array}{ll} (1). & \mathsf{Y} = \mathsf{f}(\mathsf{X}) \quad \textit{and} \\ (2). & \mathsf{f} \textit{ is } \text{DIFFERENTIABLE} \end{array}\right\} \implies \left\{\mathsf{p}_Y(y) = \sum_{n=1}^{N+1} \frac{\mathsf{p}_X(\mathsf{s}_n(y))}{\left|\mathsf{f}'(\mathsf{s}_n(y))\right|}\right\}$$

✎PROOF:

1. Note that while a function $\mathsf{f}(x)$ always has an *inverse relation* $\mathsf{f}^{-1}(x)$, this relation is not always a *function*. That is, a function $\mathsf{f}(x)$ may have more than one solution N. Let N be the number of solutions of a function $\mathsf{f}(x)$.

2. Note further that for each of the N solutions in x of $y = \mathsf{f}(x)$, there is a set A_n such that for $x \in A_n$ the function $\mathsf{f}(x)$ is invertible (as a function). Let $x = \mathsf{s}_n(y)$ be the solution in x of the equation $y = \mathsf{f}(x)$ for $x \in A_n$.

3. The remainder of the proof follows ...

$$\begin{aligned}
\mathsf{p}_Y(y) &\triangleq \frac{\mathsf{d}}{\mathsf{dy}}\mathsf{P}\{\mathsf{Y} \le y\} && \text{by definition of } \mathsf{p}_Y \text{ (Definition B.2 page 184)} \\
&= \frac{\mathsf{d}}{\mathsf{dy}}\mathsf{P}\{\mathsf{f}(\mathsf{X}) \le y\} && \text{by hypothesis (1)} \\
&= \frac{\mathsf{d}}{\mathsf{dy}}\sum_{n=1}^{N+1}\mathsf{P}\left\{[\mathsf{f}(\mathsf{X}) \le y] \wedge [\mathsf{X} \in A_n]\right\} && \text{by \textit{sum of products} (Theorem A.3 page 175)} \\
&= \frac{\mathsf{d}}{\mathsf{dy}}\sum_{n=1}^{N+1}\mathsf{P}\left\{\mathsf{f}(\mathsf{X}) \le y \mid \mathsf{X} \in A_n\right\}\mathsf{P}\left\{\mathsf{X} \in A_n\right\} && \text{by definition of } \mathsf{P}\{X|Y\} \text{ (Definition A.4 page 174)} \\
&= \frac{\mathsf{d}}{\mathsf{dy}}\left\{\begin{array}{ll} \displaystyle\sum_{n=1}^{N+1}\mathsf{P}\left\{\mathsf{X} \le \mathsf{s}_n(y) \mid \mathsf{X} \in A_n\right\}\mathsf{P}\left\{\mathsf{X} \in A_n\right\} & \text{for } \mathsf{f}'(x) \ge 0 \\ \displaystyle\sum_{n=1}^{N+1}\mathsf{P}\left\{\mathsf{X} \ge \mathsf{s}_n(y) \mid \mathsf{X} \in A_n\right\}\mathsf{P}\left\{\mathsf{X} \in A_n\right\} & \text{otherwise} \end{array}\right\} && \text{by item (2)} \\
&= \frac{\mathsf{d}}{\mathsf{dy}}\left\{\begin{array}{ll} \displaystyle\sum_{n=1}^{N+1}\mathsf{P}\left\{\mathsf{X} \le \mathsf{s}_n(y) \mid \mathsf{X} \in A_n\right\}\mathsf{P}\left\{\mathsf{X} \in A_n\right\} & \text{for } \mathsf{f}'(x) \ge 0 \\ \displaystyle\sum_{n=1}^{N+1}\left[1 - \mathsf{P}\left\{\mathsf{X} \le \mathsf{s}_n(y) \mid \mathsf{X} \in A_n\right\}\mathsf{P}\left\{\mathsf{X} \in A_n\right\}\right] & \text{otherwise} \end{array}\right\}
\end{aligned}$$

[4] Papoulis (1984) pages 95–96 ⟨"Fundamental Theorem"⟩, Papoulis (1990) page 157 ⟨"Fundamental Theorem"⟩, Papoulis (1991) page 93, Haykin (1994) pages 235–239 ⟨§"4.5 TRANSFORMATIONS OF RANDOM VARIABLES"⟩, Proakis (2001) page 30

$$= \frac{\mathsf{d}}{\mathsf{dy}} \begin{cases} \sum_{n=1}^{N+1} \mathsf{P}\left\{\mathsf{X} \le \mathsf{s}_n(y) | \mathsf{X} \in A_n\right\} & \text{for } \mathsf{f}'(x) \ge 0 \\ \sum_{n=1}^{N+1} \left[1 - \mathsf{P}\left\{\mathsf{X} \le \mathsf{s}_n(y) | \mathsf{X} \in A_n\right\}\right] & \text{otherwise} \end{cases} \quad \text{by definition of } \mathsf{P}\{X|Y\}$$

$$= \frac{\mathsf{d}}{\mathsf{dy}} \begin{cases} \sum_{n=1}^{N+1} \left(\mathsf{P}\{\mathsf{X} \le \mathsf{s}_n(y)\} - \mathsf{P}\{\mathsf{X} < \min A_n\}\right) & \text{for } \mathsf{f}'(x) \ge 0 \\ \sum_{n=1}^{N+1} \left[1 - \left(\mathsf{P}\{\mathsf{X} \le \mathsf{s}_n(y)\} - \mathsf{P}\{\mathsf{X} < \min A_n\}\right)\right] & \text{otherwise} \end{cases}$$

$$= \frac{\mathsf{d}}{\mathsf{dy}} \begin{cases} \sum_{n=1}^{N+1} \mathsf{P}\{\mathsf{X} \le \mathsf{s}_n(y)\} & \text{for } \mathsf{f}'(x) \ge 0 \\ \sum_{n=1}^{N+1} \left[1 - \mathsf{P}\{\mathsf{X} \le \mathsf{s}_n(y)\}\right] & \text{otherwise} \end{cases} \quad \text{because } \frac{\mathsf{d}}{\mathsf{dy}}\mathsf{P}\{\mathsf{X} < \text{constant}\} = 0$$

$$= \begin{cases} \sum_{n=1}^{N+1} \frac{\mathsf{d}}{\mathsf{dy}} \mathsf{c}_{\mathsf{X}}\left[\mathsf{s}_n(y)\right] & \text{for } \mathsf{f}'(x) \ge 0 \\ \sum_{n=1}^{N+1} \frac{\mathsf{d}}{\mathsf{dy}} \left[1 - \mathsf{c}_{\mathsf{X}}\left(\mathsf{s}_n(y)\right)\right] & \text{otherwise} \end{cases} \quad \text{by } \textit{linearity} \text{ of } \frac{\mathsf{d}}{\mathsf{dy}} \text{ operator}$$

$$= \begin{cases} \sum_{n=1}^{N+1} \mathsf{p}_{\mathsf{X}}\left[\mathsf{s}_n(y)\right] \frac{\mathsf{d}}{\mathsf{dy}} \left[\mathsf{s}_n(y)\right] & \text{for } \mathsf{f}'(x) \ge 0 \\ \sum_{n=1}^{N+1} \left[-\mathsf{p}_{\mathsf{X}}\left[\mathsf{s}_n(y)\right] \frac{\mathsf{d}}{\mathsf{dy}} \left[\mathsf{s}_n(y)\right]\right] & \text{otherwise} \end{cases}$$

by definition of p_{X} (Definition B.2 page 184) and the *chain rule*

$$= \sum_{n=1}^{N+1} \mathsf{p}_{\mathsf{X}}\left(\mathsf{s}_n(y)\right) \left| \frac{\mathsf{d}}{\mathsf{dy}} \left[\mathsf{s}_n(y)\right] \right|$$

$$= \sum_{n=1}^{N+1} \frac{\mathsf{p}_{\mathsf{X}}\left(\mathsf{s}_n(y)\right)}{\left|\mathsf{f}'\left(\mathsf{s}_n(y)\right)\right|} \quad \text{by Lemma H.1 page 224}$$

✏

Corollary 6.1. [5] *Let* X *and* Y *be* RANDOM VARIABLES *in* $\mathbb{R}^{\mathbb{R}}$. *Let* $a, b \in \mathbb{R}$.

COR

$$\left\{\begin{array}{ll} \text{(A).} & \mathsf{Y} = \sigma\mathsf{X} + \mu \quad \textit{and} \\ \text{(B).} & \sigma \ne 0 \end{array}\right\} \implies \left\{\begin{array}{llll} (1) & \mathsf{p}_{\mathsf{Y}}(y) & = & \underbrace{\frac{1}{|\sigma|}\mathsf{p}_{\mathsf{X}}\left(\frac{y-\mu}{\sigma}\right)}_{\text{DILATION by } \sigma \text{ and TRANSLATION by } \mu} \quad \textit{and} \\ (2) & \mathbf{E}(\mathsf{Y}) & = & \sigma\mathbf{E}(\mathsf{X}) + \mu \quad \textit{and} \\ (3) & \mathsf{Var}(\mathsf{Y}) & = & \sigma^2\mathsf{Var}(\mathsf{X}) \end{array}\right\}$$

✎PROOF:

1. Proof for (1):

   (a) lemma: $\mathsf{f}(x) = \sigma x + \mu$ is a *differentiable function* and $\mathsf{f}'(x) = \sigma$.

   (b) lemma: The equation $y = \mathsf{f}(x) = \sigma x + b$ has $N = 1$ solution and that solution is $x = \mathsf{s}_1(y) = \frac{y-\mu}{\sigma}$.

[5] Papoulis (1984) page 96 ⟨"Illustrations" 1⟩, Papoulis (1991) page 95, Proakis (2001) page 29

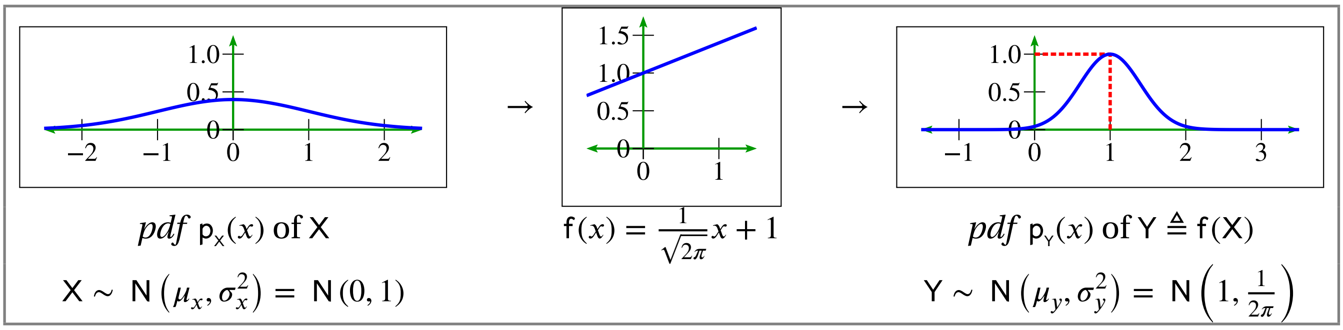


Figure 6.4: Flow diagram for Example 6.5 (page 43)

(c) It follows that

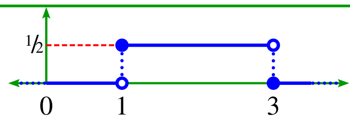
$$\begin{aligned}
p_Y(y) &= \sum_{n=1}^{N} \frac{p_X(s_n(y))}{|f'(s_n(y))|} && \text{by Theorem 6.3 (page 41)} \\
&= \frac{p_X(s_1(y))}{|f'(s_1(y))|} && \text{because } N = 1 \text{ ((1b) lemma page 42)} \\
&= \frac{p_X(s_1(y))}{|\sigma|} && \text{by (1a) lemma} \\
&= \frac{1}{|\sigma|} p_X\left(\frac{y-\mu}{\sigma}\right) && \text{by (1b) lemma}
\end{aligned}$$

2. Proof for (2): By Theorem 1.1 (page 4).
3. Proof for (3): By Corollary 1.3 (page 5).

✎

Example 6.4. Revisit Example 6.2 (page 39), but this time using Corollary 6.1 (page 42) rather than the *Inverse probability integral transform* (Theorem 6.2 page 38): Suppose we have a random generator that yields a value X with *uniform distribution*. Choose a function $f(x)$ such that $Y \triangleq f(X)$ has distribution $p_Y(y) \triangleq \frac{1}{2}$ for $y \in (1:3]$ and 0 otherwise.

EX $\{f(x) \triangleq 2x+1\} \implies Y \triangleq f(X)$ has *pdf* $p_Y(y) \triangleq \begin{cases} \frac{1}{2} & \text{for } y \in (1:3] \\ 0 & \text{otherwise} \end{cases}$



✎PROOF:

$$\begin{aligned}
p_Y(y) &\triangleq \begin{cases} \frac{1}{2} & \text{for } y \in (1:3] \\ 0 & \text{otherwise} \end{cases} \\
&= \frac{1}{2} p_X\left(\frac{x-1}{2}\right) && \text{by definition of \textit{uniform distribution}} \quad \text{(Definition C.1 page 189)} \\
\implies f(x) &= 2x+1 && \text{by Corollary 6.1 page 42}
\end{aligned}$$

✎

Example 6.5.

EX $\left\{ \begin{array}{lrcl} \text{(A)} & f(x) & \triangleq & \frac{1}{\sqrt{2\pi}}x + 1 \quad \text{and} \\ \text{(B)} & X & \sim & N(0,1) \end{array} \right\} \implies \left\{ Y \triangleq f(X) \sim N\left(1, \frac{1}{\sqrt{2\pi}}\right) \right\}$

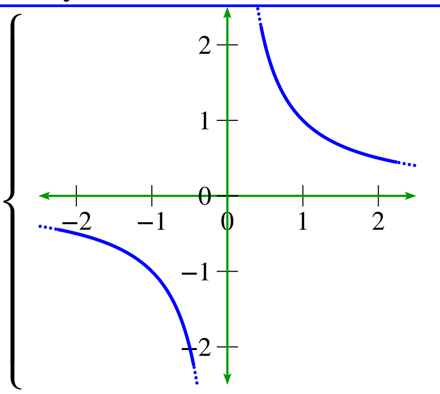
…as illustrated in Figure 6.4 (page 43).

✎PROOF: This follows from Corollary 6.1 (page 42). ✑

Corollary 6.2. [6]

COR

$$\left\{ \text{[graph]} \quad \mathsf{Y} = \frac{1}{\mathsf{X}} \quad \mathsf{X} \neq 0 \right\} \implies \left\{ \mathsf{p}_{\mathsf{Y}}(y) = \frac{1}{y^2}\mathsf{p}_{\mathsf{X}}\left(\frac{1}{y}\right) \quad \textit{for } y \neq 0 \right\}$$



✎PROOF:

1. lemma: $\mathsf{f}(x) = 1/x$ is a *differentiable function* in $x \neq 0$ and $\mathsf{f}'(x) = \frac{-1}{x^2}$.
2. lemma: The equation $y = \mathsf{f}(x) = \frac{1}{x}$ has $N = 1$ solution which is $x = \mathsf{s}_1(y) = \frac{1}{y}$ for $y \neq 0$.
3. It follows that

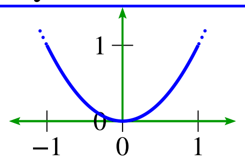
$$\begin{aligned}
\mathsf{p}_{\mathsf{Y}}(y) &= \sum_{n=1}^{N} \frac{\mathsf{p}_{\mathsf{X}}(\mathsf{s}_n(y))}{|\mathsf{f}'(\mathsf{s}_n(y))|} && \text{for } y \neq 0 && \text{by Theorem 6.3 (page 41)} \\
&= \frac{\mathsf{p}_{\mathsf{X}}(\mathsf{s}_1(y))}{|\mathsf{f}'(\mathsf{s}_1(y))|} && \text{for } y \neq 0 && \text{because } N = 1 \text{ ((2) lemma page 44)} \\
&= \frac{\mathsf{p}_{\mathsf{X}}\left(\frac{1}{y}\right)}{\left|\mathsf{f}'\left(\frac{1}{y}\right)\right|} && \text{for } y \neq 0 && \text{by (2) lemma} \\
&= \frac{1}{|-1/(1/y)^2|}\mathsf{p}_{\mathsf{X}}\left(\frac{1}{y}\right) && \text{for } y \neq 0 && \text{by (1) lemma} \\
&= \frac{1}{y^2}\mathsf{p}_{\mathsf{X}}\left(\frac{1}{y}\right) && \text{for } y \neq 0 && \text{by definition of } |\cdot|
\end{aligned}$$

✑

Corollary 6.3. [7] *Let* X *and* Y *be* RANDOM VARIABLES.

COR

$$\left\{ \text{[graph]} \quad \mathsf{Y} = \mathsf{X}^2 \right\} \implies \left\{ \mathsf{p}_{\mathsf{Y}}(y) = \frac{1}{2\sqrt{y}}\left[\mathsf{p}_{\mathsf{X}}\left(-\sqrt{y}\right) + \mathsf{p}_{\mathsf{X}}\left(\sqrt{y}\right)\right] \quad \textit{for } y \geq 0 \right\}$$



✎PROOF:

1. lemma: $\mathsf{f}(x) = x^2$ is a *differentiable function* for $x \in \mathbf{R}_e$ and $\mathsf{f}'(x) = 2x$.

[6] Papoulis (1984) page 97 ⟨Example 5-10⟩, Papoulis (1991) page 94

[7] Papoulis (1984) page 95 ⟨Example 5-9⟩, Devroye (1986) page 27 ⟨Example 4.4⟩, Papoulis (1991) page 95, Proakis (2001) page 29

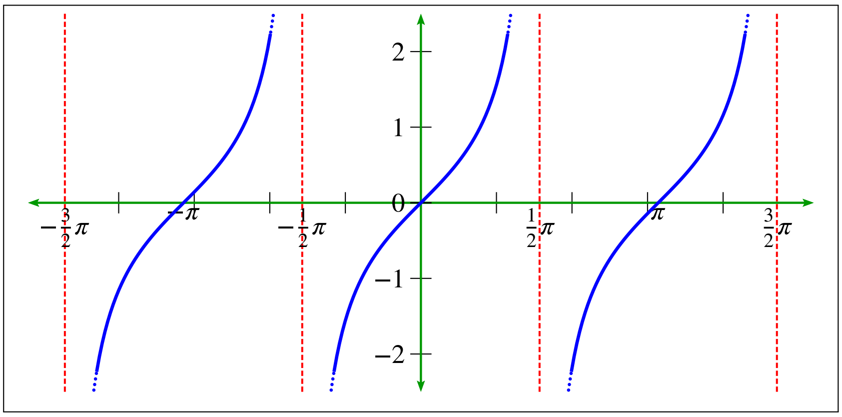


Figure 6.5: $\mathsf{Z} = \tan \Theta$

2. lemma: The equation $y = \mathsf{f}(x) = x^2$ has $N = 2$ solutions which are

$$x = \begin{cases} \mathsf{s}_1(y) &= +\sqrt{y} \quad \text{for } y \geq 0 \\ \mathsf{s}_2(y) &= -\sqrt{y} \quad \text{for } y \geq 0 \end{cases}$$

3. And so it follows that …

$$\begin{aligned}
\mathsf{p}_\mathsf{Y}(y) &= \sum_{n=1}^{N} \frac{\mathsf{p}_\mathsf{X}\big(\mathsf{s}_n(y)\big)}{\big|\mathsf{f}'\big(\mathsf{s}_n(y)\big)\big|} && \text{for } y \geq 0 && \text{by Theorem 6.3 (page 41)} \\
&= \frac{\mathsf{p}_\mathsf{X}\big(\mathsf{s}_1(y)\big)}{\big|\mathsf{f}'\big(\mathsf{s}_1(y)\big)\big|} + \frac{\mathsf{p}_\mathsf{X}\big(\mathsf{s}_2(y)\big)}{\big|\mathsf{f}'\big(\mathsf{s}_2(y)\big)\big|} && \text{for } y \geq 0 && \text{because } N = 2 \text{ ((2) lemma page 45)} \\
&= \frac{\mathsf{p}_\mathsf{X}\big(+\sqrt{y}\big)}{\big|\mathsf{f}'\big(+\sqrt{y}\big)\big|} + \frac{\mathsf{p}_\mathsf{X}\big(-\sqrt{y}\big)}{\big|\mathsf{f}'\big(-\sqrt{y}\big)\big|} && \text{for } y \geq 0 && \text{by (2) lemma} \\
&= \frac{\mathsf{p}_\mathsf{X}\big(+\sqrt{y}\big)}{\big|2\big(+\sqrt{y}\big)\big|} + \frac{\mathsf{p}_\mathsf{X}\big(-\sqrt{y}\big)}{\big|2\big(-\sqrt{y}\big)\big|} && \text{for } y \geq 0 && \text{by (1) lemma} \\
&= \frac{1}{2\sqrt{y}}\Big[\mathsf{p}_\mathsf{X}\Big(-\sqrt{y}\Big) + \mathsf{p}_\mathsf{X}\Big(\sqrt{y}\Big)\Big] && \text{for } y \geq 0 && \text{by definition of } |\cdot|
\end{aligned}$$

✏

Corollary 6.4. [8] *Let* $\mathsf{Z} = \tan\Theta$. *Then*

COR
$$\{\mathsf{Z} = \tan\Theta\} \implies \left\{ \mathsf{p}_\mathsf{z}(z) = \frac{1}{1+z^2} \sum_{n \in \mathbb{Z}} \mathsf{p}_\theta(\operatorname{atan}(z) + n\pi) \right\}$$

✎PROOF:

1. lemma: $\mathsf{f}(\theta) = \tan(\theta)$ is a *differentiable function* for $\theta \in \mathbf{R}_\mathsf{e}$ and $\mathsf{f}'(\theta) = \sec^2(\theta)$.

2. lemma: The equation $y = \mathsf{f}(x) = x^2$ has a countably infinite number of solutions N which are the elements of the sequence $\left(\!\left(\theta_n = \mathsf{s}_n(z) = \operatorname{atan}(z) + n\pi\right)\!\right)_{n \in \mathbb{Z}}$.

[8] Papoulis (1991) pages 99–100

3. And so it follows that ...

$$\begin{aligned}
\mathsf{p}_z(z) &= \sum_{n\in\mathbb{Z}} \frac{\mathsf{p}_\theta(\theta_n)}{|\mathsf{f}'(\theta_n)|} && \text{by Theorem 6.3 page 41}\\
&= \sum_{n\in\mathbb{Z}} \frac{\mathsf{p}_\theta(\operatorname{atan}(z)+n\pi)}{|\mathsf{f}'(\operatorname{atan}(z)+n\pi)|} && \text{by (2) lemma page 45}\\
&= \sum_{n\in\mathbb{Z}} \frac{\mathsf{p}_\theta(\operatorname{atan}(z)+n\pi)}{\left|\sec^2(\operatorname{atan}(z)+n\pi)\right|} && \text{by (1) lemma page 45}\\
&= \sum_{n\in\mathbb{Z}} \cos^2(\operatorname{atan}(z)+n\pi)\mathsf{p}_\theta(\operatorname{atan}(z)+n\pi) && \text{by definitions of } \cos(x) \text{ and } \sec(x)\\
&= \cos^2(\operatorname{atan}(z)) \sum_{n\in\mathbb{Z}} \mathsf{p}_\theta(\operatorname{atan}(z)+n\pi) && \text{by \textit{distributive} property of field } (\mathbb{R},+,\cdot,0,1)\\
&= \frac{1}{1+z^2} \sum_{n\in\mathbb{Z}} \mathsf{p}_\theta(\operatorname{atan}(z)+n\pi) && \text{by Theorem I.14 page 245}
\end{aligned}$$

✏

6.2 Functions of two random variables

Theorem 6.4. [9] *Let* X, Y, *and* Z *be* RANDOM VARIABLES. *Let* $\star$ *be the* CONVOLUTION *operator* (Definition D.1 page 199).

THM $\left\{\begin{array}{lll}(1). & Z \triangleq X+Y & \text{and}\\ (2). & X \text{ and } Y \text{ are INDEPENDENT} & \text{(Definition A.3 page 173)}\end{array}\right\} \implies \left\{\mathsf{p}_z(z) = \mathsf{p}_x(z) \star \mathsf{p}_y(z)\right\}$

✎PROOF:

$$\begin{aligned}
\mathsf{p}_z(z) &\triangleq \frac{\mathsf{d}}{\mathsf{d}z}\mathsf{c}_z(z) && \text{by definition of } \mathsf{p}_z \quad \text{(Definition B.2 page 184)}\\
&\triangleq \frac{\mathsf{d}}{\mathsf{d}z}\mathsf{P}\{Z \le z\} && \text{by definition of } \mathsf{c}_z \quad \text{(Definition B.2 page 184)}\\
&\triangleq \frac{\mathsf{d}}{\mathsf{d}z}\mathsf{P}\{X+Y \le z\} && \text{by hypothesis (1)}\\
&= \frac{\mathsf{d}}{\mathsf{d}z}\lim_{\varepsilon\to 0}\sum_{n\in\mathbb{Z}} \mathsf{P}\{[X+Y\le z] \wedge [y+n\varepsilon < Y \le y+(n+1)\varepsilon]\} && \text{by \textit{sum of products}} \quad \text{(Theorem A.3 page 175)}\\
&= \frac{\mathsf{d}}{\mathsf{d}z}\lim_{\varepsilon\to 0}\sum_{n\in\mathbb{Z}} \mathsf{P}\left\{X+Y\le z | y+n\varepsilon < Y \le y+(n+1)\varepsilon\right\}\mathsf{P}\{y+n\varepsilon < Y \le y+(n+1)\varepsilon\}\\
&= \frac{\mathsf{d}}{\mathsf{d}z}\lim_{\varepsilon\to 0}\sum_{n\in\mathbb{Z}} \mathsf{P}\left\{X+Y\le z | Y=y\right\}\mathsf{P}\{y+n\varepsilon < Y \le y+(n+1)\varepsilon\}\\
&= \frac{\mathsf{d}}{\mathsf{d}z}\int_{y\in\mathbb{R}} \mathsf{P}\left\{X+Y\le z \mid Y=y\right\}\mathsf{p}_y(y)\,\mathsf{d}y\\
&= \frac{\mathsf{d}}{\mathsf{d}z}\int_{y\in\mathbb{R}} \mathsf{P}\left\{X\le z-y \mid Y=y\right\}\mathsf{p}_y(y)\,\mathsf{d}y && Y \text{ is given as } Y=y\\
&= \frac{\mathsf{d}}{\mathsf{d}z}\int_{y\in\mathbb{R}} \mathsf{P}\{X\le z-y\}\mathsf{p}_y(y)\,\mathsf{d}y && \text{by hypothesis (2)}\\
&\triangleq \frac{\mathsf{d}}{\mathsf{d}z}\int_{y\in\mathbb{R}} \mathsf{c}_x(z-y)\mathsf{p}_y(y)\,\mathsf{d}y && \text{by definition of } \mathsf{c}_x \quad \text{(Definition B.2 page 184)}
\end{aligned}$$

[9] Papoulis (1990) page 160 ⟨Example 5.16⟩

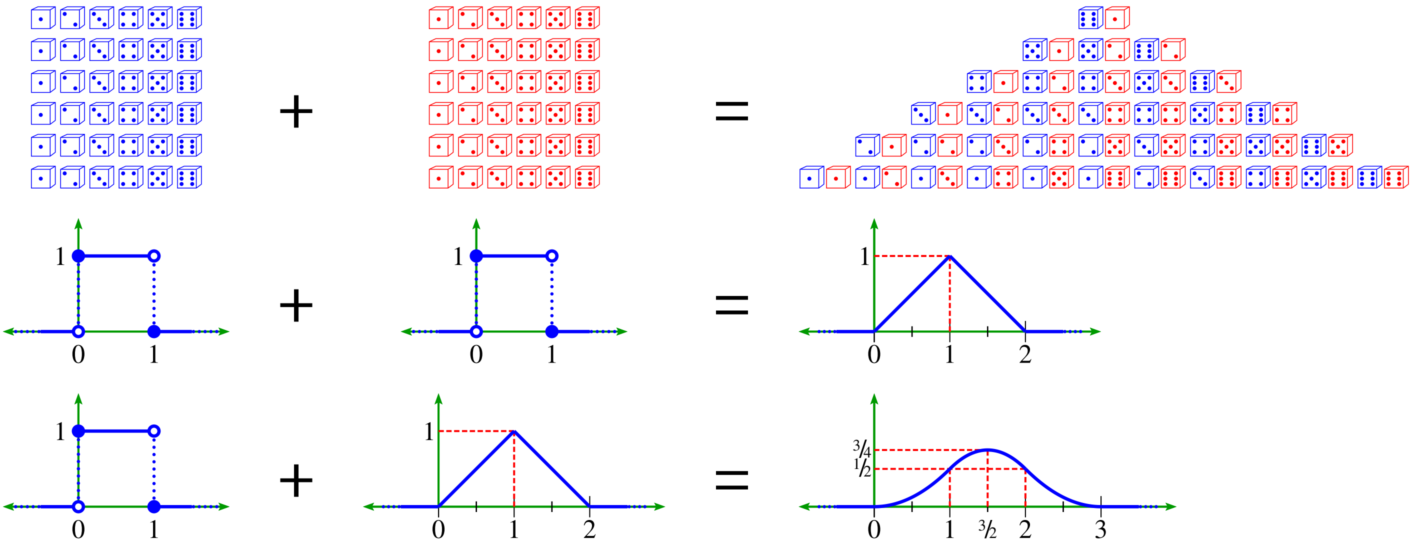


Figure 6.6: **Sum** of random variables yields **convolution** of pdfs (Theorem 6.4 page 46)

$$
\begin{aligned}
&= \int_{y\in\mathbb{R}} \frac{\mathrm{d}}{\mathrm{d}z}\left[\mathsf{c}_{\mathsf{X}}(z-y)\mathsf{p}_{\mathsf{Y}}(y)\right]\,\mathrm{d}y && \text{by } \textit{linearity} \text{ of } \frac{\mathrm{d}}{\mathrm{d}z}\\
&= \int_{y\in\mathbb{R}} \left[\frac{\mathrm{d}}{\mathrm{d}z}\mathsf{c}_{\mathsf{X}}(z-y)\right]\mathsf{p}_{\mathsf{Y}}(y)\,\mathrm{d}y && \text{because } y \text{ is fixed inside the integral}\\
&\triangleq \int_{y} \mathsf{p}_{\mathsf{X}}(z-y)\mathsf{p}_{\mathsf{Y}}(y)\,\mathrm{d}y && \text{by definition of } \mathsf{p}_{\mathsf{X}} \quad \text{(Definition B.2 page 184)}\\
&= \mathsf{p}_{\mathsf{X}}(z) \star \mathsf{p}_{\mathsf{Y}}(z) && \text{by definition of } \star \quad \text{(Definition D.1 page 199)}
\end{aligned}
$$

✏

Theorem 6.5. *Let* X_1 *and* X_2 *be random variables with joint distribution* $\mathsf{p}_{\mathsf{X}_1,\mathsf{X}_2}(x_1,x_2)$. *Let* $\mathsf{Y}_1 \triangleq \mathsf{f}_1(x_1,x_2)$ *and* $\mathsf{Y}_2 \triangleq \mathsf{f}_2(x_1,x_2)$ *with joint distribution* $\mathsf{p}_{\mathsf{Y}_1,\mathsf{Y}_2}(y_1,y_2)$.

THM

$$
\mathsf{p}_{\mathsf{Y}_1,\mathsf{Y}_2}(y_1,y_2) = \frac{\mathsf{p}_{\mathsf{X}_1,\mathsf{X}_2}(x_1,x_2)}{|J(x_1,x_2)|} = \frac{\mathsf{p}_{\mathsf{X}_1,\mathsf{X}_2}(x_1,x_2)}{\begin{vmatrix}\frac{\partial \mathsf{f}_1}{\partial x_1} & \frac{\partial \mathsf{f}_1}{\partial x_2}\\ \frac{\partial \mathsf{f}_2}{\partial x_1} & \frac{\partial \mathsf{f}_2}{\partial x_2}\end{vmatrix}} = \frac{\mathsf{p}_{\mathsf{X}_1,\mathsf{X}_2}(x_1,x_2)}{\frac{\partial \mathsf{f}_1}{\partial x_1}\frac{\partial \mathsf{f}_2}{\partial x_2} - \frac{\partial \mathsf{f}_1}{\partial x_2}\frac{\partial \mathsf{f}_2}{\partial x_1}}
$$

Proposition 6.2. *Let* X *and* Y *be random variables with joint distribution* $\mathsf{p}_{\mathsf{XY}}(x,y)$ *and* $R^2 \triangleq \mathsf{X}^2 + \mathsf{Y}^2$ *and* $\Theta \triangleq \operatorname{atan}\left(\frac{\mathsf{Y}}{\mathsf{X}}\right)$. *Then*

PRP

$$
\mathsf{p}_{\mathsf{R},\Theta}(r,\theta) = r\ \mathsf{p}_{\mathsf{XY}}(r\cos\theta, r\sin\theta)
$$

✎PROOF:

$$
\begin{aligned}
\mathsf{p}_{\mathsf{R},\Theta}(r,\theta) &= \frac{\mathsf{p}_{\mathsf{XY}}(x,y)}{|J(x,y)|} = \frac{\mathsf{p}_{\mathsf{XY}}(x,y)}{\begin{vmatrix}\frac{\partial R}{\partial x} & \frac{\partial R}{\partial y}\\ \frac{\partial \theta}{\partial x} & \frac{\partial \theta}{\partial y}\end{vmatrix}} = \frac{\mathsf{p}_{\mathsf{XY}}(x,y)}{\begin{vmatrix}\frac{x}{\sqrt{x^2+y^2}} & \frac{y}{\sqrt{x^2+y^2}}\\ \frac{-y}{x^2+y^2} & \frac{x}{x^2+y^2}\end{vmatrix}} && \text{by Theorem 6.5 page 47}\\
&= \frac{\mathsf{p}_{\mathsf{XY}}(x,y)}{\frac{x}{\sqrt{x^2+y^2}}\frac{x}{x^2+y^2} - \frac{y}{\sqrt{x^2+y^2}}\frac{-y}{x^2+y^2}}\\
&= \frac{\mathsf{p}_{\mathsf{XY}}(x,y)}{\frac{x^2+y^2}{(x^2+y^2)^{3/2}}}
\end{aligned}
$$

$$= \mathsf{p}_{XY}(x,y)\frac{(x^2+y^2)^{3/2}}{x^2+y^2}$$
$$= \mathsf{p}_{XY}(r\cos\theta, r\sin\theta)\frac{r^3}{r^2}$$
$$= r\ \mathsf{p}_{XY}(r\cos\theta, r\sin\theta)$$

✏

Proposition 6.3. *Let* $\mathsf{X} \sim \mathsf{N}\left(0,\sigma^2\right)$ *and* $\mathsf{Y} \sim \mathsf{N}\left(0,\sigma^2\right)$ *be independent random variables and* $\mathsf{R}^2 \triangleq \mathsf{X}^2 + \mathsf{Y}^2$ *and* $\Theta \triangleq \operatorname{atan}\frac{\mathsf{Y}}{\mathsf{X}}$. *Then*

PRP

1. R *and* Θ *are independent with joint distribution* $\mathsf{p}_{R,\Theta}(r,\theta) = \mathsf{p}_R(r)\mathsf{p}_\theta(\theta)$
2. R *has Rayleigh distribution* $\mathsf{p}_R(r) = \frac{r}{\sigma^2}\exp\frac{r^2}{-2\sigma^2}$
3. Θ *has uniform distribution* $\mathsf{p}_\theta(\theta) = \frac{1}{2\pi}$

✎PROOF:

$$\begin{aligned}
\mathsf{p}_{R,\Theta}(r,\theta) &= r\ \mathsf{p}_{XY}(r\cos\theta, r\sin\theta) && \text{by Proposition 6.2 (page 47)}\\
&= r\ \mathsf{p}_X(r\cos\theta)\ \mathsf{p}_Y(r\sin\theta) && \text{by independence hypothesis}\\
&= r\ \frac{1}{\sqrt{2\pi\sigma^2}}\exp\frac{(r\cos\theta-0)^2}{-2\sigma^2}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\frac{(r\sin\theta-0)^2}{-2\sigma^2}\\
&= \frac{1}{2\pi\sigma^2}\ r\ \exp\frac{r^2(\cos^2\theta+\sin^2\theta)}{-2\sigma^2}\\
&= \frac{1}{2\pi\sigma^2}\ r\ \exp\frac{r^2}{-2\sigma^2}\\
&= \left[\frac{1}{2\pi}\right]\left[\frac{r}{\sigma^2}\exp\frac{r^2}{-2\sigma^2}\right]
\end{aligned}$$

✏

Proposition 6.4. *Let* X *and* Y *be* RANDOM VARIABLES *with covariance* σ_{xy} *on a* PROBABILITY SPACE $(\Omega, \mathbb{E}, \mathsf{P})$.

PRP

$$\left\{\begin{array}{lll}
\text{(A).} & \mathsf{X} \textit{ is Gaussian with } \mathsf{N}\left(\mu_X,\sigma_X^2\right) & \textit{and}\\
\text{(B).} & \mathsf{Y} \textit{ is Gaussian with } \mathsf{N}\left(\mu_Y,\sigma_Y^2\right) & \textit{and}\\
\text{(C).} & \sigma_{xy} = \operatorname{cov}[\mathsf{X},\mathsf{Y}] &
\end{array}\right\} \implies \left\{\mathsf{P}\{\mathsf{X}>\mathsf{Y}\} = \mathsf{Q}\left(\frac{-\mu_X+\mu_Y}{\sigma_X^2+\sigma_Y^2-2\sigma_{xy}}\right)\right\}$$

✎PROOF: Because X and Y are jointly Gaussian, their linear combination Z = X − Y is also Gaussian. A Gaussian distribution is completely defined by its mean and variance. So, to determine the distribution of Z, we just have to determine the mean and variance of Z.

$$\begin{aligned}
\mathbf{E}Z &= \mathbf{E}\mathsf{X} - \mathbf{E}Y\\
&= \mu_X - \mu_Y
\end{aligned}$$

$$\begin{aligned}
\operatorname{var} Z &= \mathbf{E}Z^2 - (\mathbf{E}Z)^2\\
&= \mathbf{E}(\mathsf{X}-\mathsf{Y})^2 - (\mathbf{E}\mathsf{X}-\mathbf{E}Y)^2\\
&= \mathbf{E}(\mathsf{X}^2 - 2XY + Y^2) - [(\mathbf{E}\mathsf{X})^2 - 2\mathbf{E}\mathsf{X}\mathbf{E}Y + (\mathbf{E}Y)^2]\\
&= [\mathbf{E}X^2 - (\mathbf{E}\mathsf{X})^2] + [Y^2 - (\mathbf{E}Y)^2] - 2[\mathbf{E}XY - \mathbf{E}\mathsf{X}\mathbf{E}Y]\\
&= \operatorname{var}\mathsf{X} + \operatorname{var}Y - 2\operatorname{cov}[\mathsf{X},\mathsf{Y}]\\
&\triangleq \sigma_X^2 + \sigma_Y^2 - 2\sigma_{xy}
\end{aligned}$$

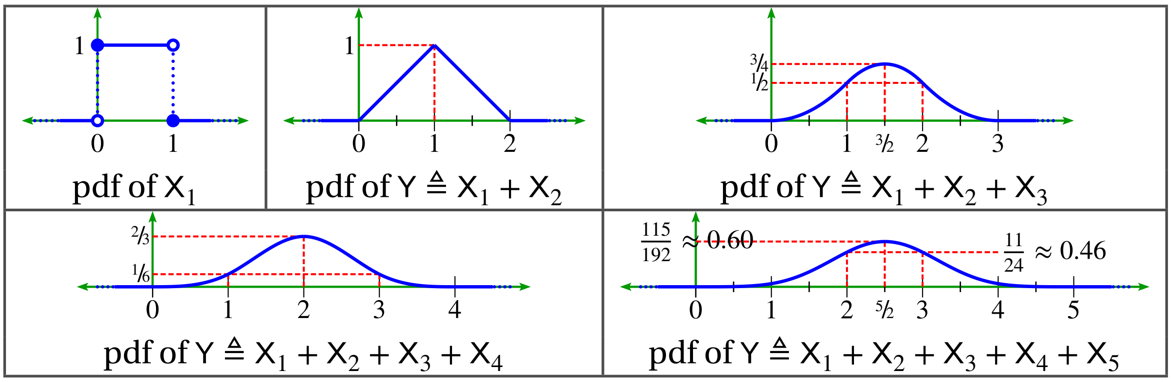


Figure 6.7: The distributions of sums of independent uniformly distributed random variables (Example 6.6 page 49)

$$
\begin{aligned}
\mathsf{P}\{X > Y\} &= \mathsf{P}\{X - Y > 0\} \\
&= \mathsf{P}\{Z > 0\} \\
&= \mathsf{Q}\left(\frac{z - \mathbf{E}Z}{\operatorname{var} Z}\right)\bigg|_{z=0} \\
&= \mathsf{Q}\left(\frac{0 - \mu_X + \mu_Y}{\sigma_X^2 + \sigma_Y^2 - 2\sigma_{xy}}\right)
\end{aligned}
$$

✏

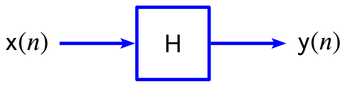
Example 6.6 (Sum of Uniformly Distributed Random Variables). Let $(\!(X_1, X_2, X_3, \ldots)\!)$ be a *sequence* of *independent* (Definition A.3 page 173) *uniformly distributed* random variables. Let $\mathsf{p}_N(x)$ be the *probability density function* of $Y \triangleq \sum_{n=1}^{N} X_n$. Some of these distributions are illustrated in Figure 6.7 (page 49). Note that the distributions of the sequence $(\!(\mathsf{p}_1, \mathsf{p}_2, \mathsf{p}_3, \ldots)\!)$ are all *B-spline*s (Definition W.3 page 444) and all form a *partition of unity*.

CHAPTER 7

OPERATORS ON DISCRETE RANDOM SEQUENCES

7.1 LTI operators on random sequences

Theorem 7.1. [1] *Let* $\mathsf{x}(n)$ *be a* RANDOM SEQUENCE *with* MEAN μ_{X} *and* $\mathsf{y}(n)$ *a* RANDOM SEQUENCE *with* MEAN μ_{Y}. *Let* **S** *be a system with* IMPULSE RESPONSE $\mathsf{h}(n)$, INPUT $\mathsf{x}(n)$, *and* OUTPUT $\mathsf{y}(n)$.



THM

$$\left\{\ \mathbf{S}\text{ is }(LTI)\ \right\} \implies \left\{\begin{array}{lrcll}
(1). & \mu_{\mathsf{Y}}(n) & = & \sum_{k\in\mathbb{Z}} \mathsf{h}(k)\mu_{\mathsf{X}}(n-k) & \triangleq \mathsf{h}(n)\star\mu_{\mathsf{X}}(n) \quad \text{and}\\
(2). & \mathsf{R}_{\mathsf{xy}}(n,m) & = & \sum_{k\in\mathbb{Z}} \mathsf{h}^*(k)\mathsf{R}_{\mathsf{xx}}(n-k,m+k) & \quad \text{and}\\
(3). & \mathsf{R}_{\mathsf{yy}}(n,m) & = & \sum_{k\in\mathbb{Z}} \mathsf{h}^*(k)\mathsf{R}_{\mathsf{yx}}(n-k,m+k) &
\end{array}\right\}$$

✎PROOF:

$\mu_{\mathsf{Y}}(n) \triangleq \mathbf{E}\left[\mathsf{y}(n)\right]$	by definition of μ_{Y}	(Definition 2.2 page 11)
$= \mathbf{E}\left[\sum_{k\in\mathbb{Z}} \mathsf{h}(k)\mathsf{x}(n-k)\right]$	by *LTI* hypothesis	
$= \sum_{k\in\mathbb{Z}} \mathsf{h}(k)\mathbf{E}[\mathsf{x}(n-k)]$	by *linear* property	
$= \sum_{k\in\mathbb{Z}} \mathsf{h}(k)\mu_{\mathsf{X}}(n-k)$	by definition of μ_{X}	(Definition 2.2 page 11)
$\triangleq \mathsf{h}(n)\star\mu_{\mathsf{X}}(n)$	by definition of *convolution*	(Definition D.4 page 201)
$\mathsf{R}_{\mathsf{xy}}(n,m) \triangleq \mathbf{E}\left[\mathsf{x}(n+m)\mathsf{y}^*(n)\right]$	by definition of $\mathsf{R}_{\mathsf{xy}}(n,m)$	(Definition 2.2 page 11)
$= \mathbf{E}\left[\mathsf{x}(n+m)(\mathsf{h}(n)\star\mathsf{x}(n))^*\right]$	by *LTI* hypothesis	
$\triangleq \mathbf{E}\left[\mathsf{x}(n+m)\left(\sum_{k\in\mathbb{Z}} h(k)\mathsf{x}(n-k)\right)^*\right]$	by definition of *convolution* $\star$	(Definition D.4 page 201)
$= \mathbf{E}\left[\mathsf{x}(n+m)\sum_{k\in\mathbb{Z}} \mathsf{h}^*(k)\mathsf{x}^*(n-k)\right]$	by *distributive* property of $*$-**algebras**	(Definition M.3 page 304)

[1] Papoulis (1991) page 310

$$\begin{aligned}
&= \mathbf{E}\left[\sum_{k\in\mathbb{Z}} \mathsf{h}^*(k)\mathsf{x}(n+m)\mathsf{x}^*(n-k)\right] && \text{by \textit{distributive} property of } (\mathbb{C},+,\cdot,0,1) && \text{(Definition G.5 page 222)}\\
&= \sum_{k\in\mathbb{Z}} \mathsf{h}^*(k)\mathbf{E}\left[\mathsf{x}(n-k+k+m)\mathsf{x}^*(n-k)\right] && \text{by \textit{linear} property of } \mathbf{E} && \text{(Theorem 1.1 page 4)}\\
&\triangleq \sum_{k\in\mathbb{Z}} \mathsf{h}^*(k)\mathsf{R}_{\mathsf{xx}}(n-k,m+k) && \text{by definition of } \mathsf{R}_{\mathsf{xx}}(n,m) && \text{(Definition 2.2 page 11)}
\end{aligned}$$

$$\begin{aligned}
\mathsf{R}_{\mathsf{yy}}(n,m) &\triangleq \mathbf{E}\left[\mathsf{y}(n+m)\mathsf{y}^*(n)\right] && \text{by definition of } \mathsf{R}_{\mathsf{xy}}(n,m) && \text{(Definition 2.2 page 11)}\\
&= \mathbf{E}\left[\mathsf{y}(n+m)(\mathsf{h}(n)\star\mathsf{x}(n))^*\right] && \text{by \textit{LTI} hypothesis}\\
&\triangleq \mathbf{E}\left[\mathsf{y}(n+m)\left(\sum_{k\in\mathbb{Z}}\mathsf{h}(k)\mathsf{x}(n-k)\right)^*\right] && \text{by definition of \textit{convolution}} && \text{(Definition D.4 page 201)}\\
&= \mathbf{E}\left[\mathsf{y}(n+m)\sum_{k\in\mathbb{Z}}\mathsf{h}^*(k)\mathsf{x}^*(n-k)\right] && \text{by \textit{distributive} property of } *\text{-\textbf{algebras}} && \text{(Definition M.3 page 304)}\\
&= \mathbf{E}\left[\sum_{k\in\mathbb{Z}}\mathsf{h}^*(k)\mathsf{y}(n+m)\mathsf{x}^*(n-k)\right] && \text{by \textit{distributive} property of } (\mathbb{C},+,\cdot,0,1) && \text{(Definition G.5 page 222)}\\
&= \sum_{k\in\mathbb{Z}}\mathsf{h}^*(k)\mathbf{E}\left[\mathsf{y}(n-k+k+m)\mathsf{x}^*(n-k)\right] && \text{by \textit{linear} property of } \mathbf{E} && \text{(Theorem 1.1 page 4)}\\
&\triangleq \sum_{k\in\mathbb{Z}}\mathsf{h}^*(k)\mathsf{R}_{\mathsf{yx}}(n-k,m+k) && \text{by definition of } \mathsf{R}_{\mathsf{xy}}(n,m) && \text{(Definition 2.2 page 11)}
\end{aligned}$$

✎

7.2 LTI operators on WSS random sequences

Corollary 7.1. *Let* **S** *be the system defined in Theorem 7.1 (page 51).*

COR

$$\left.\begin{array}{lll}
\text{(A).} & \mathbf{S} \text{ is LTI} & \text{and}\\
\text{(B).} & \mathsf{x}(n) \text{ is WSS} &
\end{array}\right\} \implies
\left\{\begin{array}{llllll}
\text{(1).} & \mu_{\mathsf{Y}} &=& \mu_{\mathsf{X}}\sum_{n\in\mathbb{Z}}\mathsf{h}(k) & & \text{and}\\
\text{(2).} & \mathsf{R}_{\mathsf{xy}}(m) &=& \mathsf{R}_{\mathsf{xx}}(m)\star\mathsf{h}^*(-m) &\triangleq \sum_{k\in\mathbb{Z}}\mathsf{h}^*(k)\mathsf{R}_{\mathsf{xx}}(m+k) & \text{and}\\
\text{(3).} & \mathsf{R}_{\mathsf{yy}}(m) &=& \mathsf{R}_{\mathsf{yx}}(m)\star\mathsf{h}^*(-m) &\triangleq \sum_{k\in\mathbb{Z}}\mathsf{h}^*(k)\mathsf{R}_{\mathsf{xy}}(m+k) & \text{and}\\
\text{(4).} & \mathsf{R}_{\mathsf{yy}}(m) &=& \mathsf{R}^*_{\mathsf{xx}}(m)\star\mathsf{h}(-m)\star\mathsf{h}^*(-m) & &
\end{array}\right.$$

✎PROOF:

$$\begin{aligned}
\mu_{\mathsf{Y}} &= \mu_{\mathsf{Y}}(n) && \text{by Proposition 2.1 page 14} && \text{and hypothesis (A)}\\
&= \sum_{n\in\mathbb{Z}}\mathsf{h}(k)\mu_{\mathsf{X}}(n-k) && \text{by Theorem 2.1 page 12} && \text{and hypothesis (B)}\\
&= \sum_{n\in\mathbb{Z}}\mathsf{h}(k)\mu_{\mathsf{X}}(0) && \text{by Definition 8.1 page 59} && \text{and hypothesis (B)}\\
&= \mu_{\mathsf{X}}(0)\sum_{n\in\mathbb{Z}}\mathsf{h}(k) && \text{by \textit{linear} property of } \textstyle\sum\\
&= \mu_{\mathsf{X}}\sum_{n\in\mathbb{Z}}\mathsf{h}(k) && \text{by Proposition 2.1 page 14}
\end{aligned}$$

[1] Papoulis (1991) page 323

$$
\begin{aligned}
\mathsf{R}_{xy}(m) &\triangleq \mathsf{R}_{xy}(0,m) && \text{by Proposition 2.1 page 14} && \text{and hypothesis (A)}\\
&= \sum_{k\in\mathbb{Z}} \mathsf{h}^*(k)\mathsf{R}_{xx}(0-k, m+k) && \text{by Theorem 7.1 page 51} && \text{and hypothesis (B)}\\
&= \sum_{k\in\mathbb{Z}} \mathsf{h}^*(k)\mathsf{R}_{xx}(m+k) && \text{by Proposition 2.1 page 14} && \text{and hypothesis (A)}\\
&= \mathsf{h}^*(-m) \star \mathsf{R}_{xx}(m) && \text{by Proposition D.1 page 202}\\
\mathsf{R}_{yy}(m) &\triangleq \mathsf{R}_{yy}(0,m) && \text{by Proposition 2.1 page 14} && \text{and hypothesis (A)}\\
&= \sum_{k\in\mathbb{Z}} \mathsf{h}^*(k)\mathsf{R}_{yx}(n-k, m+k) && \text{by Theorem 7.1 page 51} && \text{and hypothesis (B)}\\
&= \sum_{k\in\mathbb{Z}} \mathsf{h}^*(k)\mathsf{R}_{yx}(m+k) && \text{by Proposition 2.1 page 14} && \text{and hypothesis (A)}\\
&= \mathsf{h}^*(-m) \star \mathsf{R}_{yx}(m) && \text{by Proposition D.1 page 202}\\
\mathsf{R}_{yy}(m) &= \mathsf{h}^*(-m) \star \mathsf{R}_{yx}(m) && \text{by result (2)}\\
&= \mathsf{h}^*(-m) \star \mathsf{R}^*_{xy}(m) && \text{by Corollary 2.1 page 14}\\
&= \mathsf{h}^*(-m) \star \left[\mathsf{h}^*(-m) \star \mathsf{R}_{xx}(m)\right]^* && \text{by result (1)}\\
&= \mathsf{h}^*(-m) \star \mathsf{h}(-m) \star \mathsf{R}^*_{xx}(m) && \text{by \textit{distributive} property of } *\text{-\textbf{algebras}} && \text{(Definition M.3 page 304)}
\end{aligned}
$$

⇒

Corollary 7.2. [2] *Let* $\mathbf{S}$ *be a system with* IMPULSE RESPONSE $\mathsf{h}(n)$, INPUT $\mathsf{x}(n)$, *and* OUTPUT $\mathsf{y}(n)$.

COR

$$
\left\{\begin{array}{lll}
\text{(A).} & \mathsf{h} \textit{ is } \text{LINEAR TIME INVARIANT} & \textit{and}\\
\text{(B).} & \mathsf{x} \textit{ and } \mathsf{y} \textit{ are } \text{WIDE SENSE STATIONARY} &
\end{array}\right\}
\implies
\left\{\begin{array}{llcll}
\text{(1).} & \check{\mathsf{S}}_{xy}(z) & = & \check{\mathsf{S}}_{xx}(z)\check{\mathsf{H}}^*\left(\frac{1}{z^*}\right) & \textit{and}\\
\text{(2).} & \check{\mathsf{S}}_{yy}(z) & = & \check{\mathsf{S}}_{yx}(z)\check{\mathsf{H}}^*\left(\frac{1}{z^*}\right) & \textit{and}\\
\text{(3).} & \check{\mathsf{S}}_{yy}(z) & = & \check{\mathsf{S}}_{xx}(z)\check{\mathsf{H}}(z)\check{\mathsf{H}}^*\left(\frac{1}{z^*}\right) &
\end{array}\right\}
$$

✎PROOF: The proof is given in Proposition **??** (page **??**) (1). ⇒

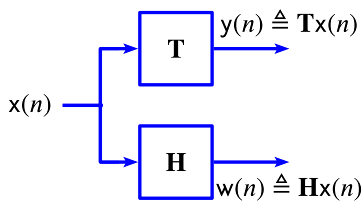
Corollary 7.3. *Let* $\mathbf{S}$ *be a system with* IMPULSE RESPONSE $\mathsf{h}(n)$, INPUT $\mathsf{x}(n)$, *and* OUTPUT $\mathsf{y}(n)$.

COR

$$
\left\{\begin{array}{llll}
\text{(A).} & \mathsf{h} \textit{ is} & \text{LTI} & \textit{and}\\
\text{(B).} & \mathsf{x} \textit{ and } \mathsf{y} \textit{ are} & \text{WSS} &
\end{array}\right\}
\implies
\left\{\begin{array}{llcll}
\text{(1).} & \tilde{\mathsf{S}}_{xy}(\omega) & = & \tilde{\mathsf{S}}_{xx}(\omega)\tilde{\mathsf{H}}^*(\omega) & \textit{and}\\
\text{(2).} & \tilde{\mathsf{S}}_{yy}(\omega) & = & \tilde{\mathsf{S}}_{xy}(\omega)\tilde{\mathsf{H}}(\omega) & \textit{and}\\
\text{(3).} & \tilde{\mathsf{S}}_{yy}(\omega) & = & \tilde{\mathsf{S}}_{xx}(\omega)\left|\tilde{\mathsf{H}}(\omega)\right|^2 &
\end{array}\right\}
$$

✎PROOF: The proof is given in Proposition **??** (page **??**) (1). ⇒

7.3 Parallel operators on WSS random sequences



Theorem 7.2. *Let* $\mathbf{S}$ *be the* SYSTEM *illustrated to the right, where* $\mathbf{T}$ *is* NOT NECESSARILY LINEAR. *Let*

$$(\!(\mathsf{h}(n))\!) \triangleq \mathbf{H}\bar{\delta}(n) \triangleq \sum_{m\in\mathbb{Z}} \mathsf{h}(m)\bar{\delta}(n-m)$$

be the IMPULSE RESPONSE *of* $\mathbf{H}$.

[2] Papoulis (1991) page 323

THM

$$\left\{\begin{array}{lll}\text{(A).} & \mathsf{x}(n) \text{ is} & \textit{WSS} \text{ and} \\ \text{(B).} & \mathbf{H} \text{ is} & \textit{LTI}\end{array}\right\} \implies \left\{\begin{array}{lrcll}\text{(1).} & \mathsf{R}_{\mathsf{wy}}(m) & = & \sum_{n\in\mathbb{Z}} \mathsf{h}(n)\mathsf{R}_{\mathsf{xy}}(m-n) & \text{(convolution)} \\ & & \triangleq & \mathsf{h}(m) \star \mathsf{R}_{\mathsf{xy}}(m) & \text{and} \\ \text{(2).} & \check{\mathsf{S}}_{\mathsf{wy}}(z) & = & \check{\mathsf{H}}(z)\check{\mathsf{S}}_{\mathsf{xy}}(z) & \text{and} \\ \text{(3).} & \tilde{\mathsf{S}}_{\mathsf{wy}}(\omega) & = & \tilde{\mathsf{H}}(\omega)\tilde{\mathsf{S}}_{\mathsf{xy}}(\omega) & \end{array}\right\}$$

✎PROOF:

$\mathsf{R}_{\mathsf{wy}}(m) \triangleq \mathbf{E}\left[\mathsf{w}(m)\mathsf{y}^*(0)\right]$	by (A) and definition of R_{wy}	(Definition 2.4 page 12)
$\triangleq \mathbf{E}\left([\mathbf{H}\mathsf{x}](m)\mathsf{y}^*(0)\right)$	by definition of $\mathbf{S}$	
$= \mathbf{H}\mathbf{E}\left(\mathsf{x}(m)\mathsf{y}^*(0)\right)$	by *LTI* hypothesis	(B)
$\triangleq \mathbf{H}\mathsf{R}_{\mathsf{xy}}(m)$	by definition of R_{xy}	(Definition 2.4 page 12)
$= \sum_{n\in\mathbb{Z}} \mathsf{h}(n)\mathsf{R}_{\mathsf{xy}}(m-n)$	by definition of $\mathbf{H}$ *impulse response* $(\!(\mathsf{h}(n))\!)$	
$= \left[\mathsf{h}(m) \star \mathsf{R}_{\mathsf{xy}}(m)\right]$	by definition of *convolution*	(Definition D.4 page 201)
$\check{\mathsf{S}}_{\mathsf{wy}}(z) \triangleq \mathbf{Z}\mathsf{R}_{\mathsf{wy}}(m)$	by definition of $\check{\mathsf{S}}_{\mathsf{wy}}$	(Definition 2.5 page 15)
$= \left[\mathsf{h}(m) \star \mathsf{R}_{\mathsf{xy}}(m)\right]$	by previous result	
$= \check{\mathsf{H}}(z)\check{\mathsf{S}}_{\mathsf{xy}}(z)$	by *Convolution Theorem*	(Theorem V.2 page 430)
$\tilde{\mathsf{S}}_{\mathsf{wy}}(\omega) \triangleq \check{\mathbf{F}}\mathsf{R}_{\mathsf{wy}}(m)$	by definition of $\tilde{\mathsf{S}}_{\mathsf{wy}}$	(Definition 8.3 page 60)
$= \left[\mathsf{h}(m) \star \mathsf{R}_{\mathsf{xy}}(m)\right]$	by previous result	
$= \tilde{\mathsf{H}}(\omega)\tilde{\mathsf{S}}_{\mathsf{xy}}(\omega)$	by *Convolution Theorem*	(Theorem V.2 page 430)

⇰

7.4 Whitening discrete random sequences

Note that if $\mathsf{L}(z)$ has a root at $z = re^{i\theta}$, then $\mathsf{L}^*(1/z^*)$ has a root at

$$\frac{1}{z^*} = \frac{1}{\left(re^{i\theta}\right)^*} = \frac{1}{re^{-i\theta}} = \frac{1}{r}e^{i\theta}.$$

That is, if $\mathsf{L}(z)$ has a root inside the unit circle, then $\mathsf{L}^*(1/z^*)$ has a root directly opposite across the unit circle boundary (see Figure 7.1 page 55). A causal stable filter $\check{\mathsf{H}}(z)$ must have all of its poles inside the unit circle. A filter has *minimum phase* (Definition V.2 page 432) if both its poles and zeros are inside the unit circle. One advantage of a minimum phase filter is that its inverse (zeros become poles and poles become zeros) is also causal and stable.

The next theorem demonstrates a method for "whitening" a *random sequence* $\mathsf{x}(n)$ with a filter constructed from a decomposition of $\mathsf{R}_{\mathsf{xx}}(m)$. The technique is stated precisely in Theorem 7.3 page 54 and illustrated in Figure 7.2 page 55. Both imply two filters with impulse responses $l(n)$ and $\gamma(n)$. Filter $l(n)$ is referred to as the **innovations filter** (because it generates or "innovates" $\mathsf{x}(n)$ from a white noise process $\mathsf{w}(n)$) and $\gamma(n)$ is referred to as the **whitening filter** because it produces a white noise sequence when the input sequence is $\mathsf{x}(n)$.[3]

Theorem 7.3. *Let* $\mathsf{x}(n)$ *be a WSS* RANDOM SEQUENCE *with auto-correlation* $\mathsf{R}_{\mathsf{xx}}(m)$ *and spectral density* $\check{\mathsf{S}}_{\mathsf{xx}}(z)$. ***If*** $\check{\mathsf{S}}_{\mathsf{xx}}(z)$ *has a* ***rational expression****, then the following are true:*

[3] Papoulis (1991) pages 401–402

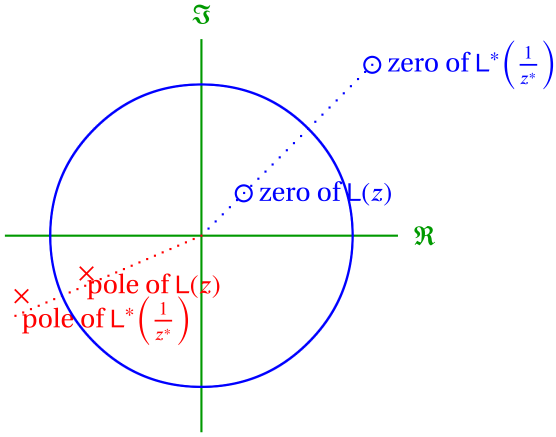


Figure 7.1: Mirrored roots in complex-z plane

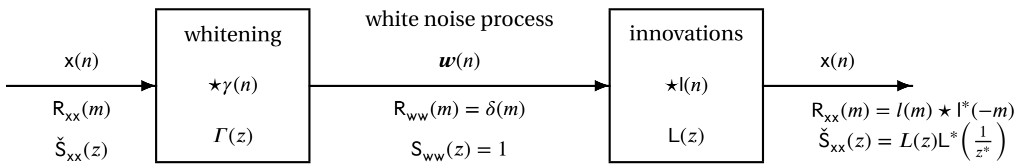


Figure 7.2: Innovations and whitening filters

1. *There exists a rational expression* $\mathsf{L}(z)$ *with minimum phase such that*

$$\check{\mathsf{S}}_{\mathsf{xx}}(z) = \mathsf{L}(z)\mathsf{L}^*\left(\frac{1}{z^*}\right).$$

2. *An LTI filter for which the Laplace transform of the impulse response* $\gamma(n)$ *is*

$$\Gamma(z) = \frac{1}{\mathsf{L}(z)}$$

   is both causal and stable.
3. *If* $\mathsf{x}(n)$ *is the input to the filter* $\gamma(n)$*, the output* $\mathsf{y}(n)$ *is a* ***white noise sequence*** *such that*

$$\mathsf{S_{yy}}(z) = 1 \qquad\qquad \mathsf{R_{yy}}(m) = \bar{\delta}(m).$$

✎PROOF:

$$\begin{aligned}
\mathsf{S_{ww}}(z) &= \Gamma(z)\Gamma^*\left(\frac{1}{z^*}\right)\check{\mathsf{S}}_{\mathsf{xx}}(z) \\
&= \frac{1}{\mathsf{L}(z)}\frac{1}{\mathsf{L}^*\left(\frac{1}{z^*}\right)}\check{\mathsf{S}}_{\mathsf{xx}}(z) \\
&= \frac{1}{\mathsf{L}(z)}\frac{1}{\mathsf{L}^*\left(\frac{1}{z^*}\right)}\mathsf{L}(z)\mathsf{L}^*\left(\frac{1}{z^*}\right) \\
&= 1
\end{aligned}$$

⇛

CHAPTER 8

OPERATORS ON CONTINUOUS RANDOM SEQUENCES

8.1 LTI Operations on non-stationary random processes

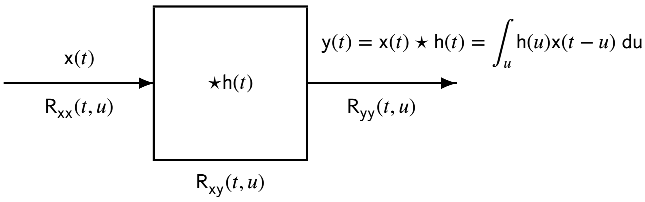


Figure 8.1: Linear system with random process input and output

Theorem 8.1. [1] *Let* $\mathsf{h} : \mathbb{R} \to \mathbb{C}$ *be the impulse response of a linear time-invariant system and Let* $\mathsf{y}(t) = h(t) \star \mathsf{x}(t) \triangleq \int_{u\in\mathbb{R}} h(u)\mathsf{x}(t-u)\,\mathrm{d}u$ *as illustrated in Figure 8.1 page 57. Then*

THM

Correlation functions

$$\begin{aligned}
\mathsf{R}_{xy}(t,u) &= \mathsf{R}_{xx}(t,u) \star h^*(u) &&\triangleq \int_{v\in\mathbb{R}} \mathsf{h}^*(v)\mathsf{R}_{xx}(t,u-v)\,\mathrm{d}v\\
\mathsf{R}_{yy}(t,u) &= \mathsf{R}_{xy}(t,u) \star \mathsf{h}(t) &&\triangleq \int_{v\in\mathbb{R}} \mathsf{h}(v)\mathsf{R}_{xy}(t-v,u)\,\mathrm{d}v\\
\mathsf{R}_{yy}(t,u) &= \mathsf{R}_{xx}(t,u) \star \mathsf{h}(t) \star \mathsf{h}^*(u) &&\triangleq \int_{w\in\mathbb{R}} \mathsf{h}^*(w) \int_{v\in\mathbb{R}} \mathsf{h}(v)\mathsf{R}_{xx}(t-v,u-w)\,\mathrm{d}v\,\mathrm{d}\omega
\end{aligned}$$

Laplace power spectral density functions

$$\begin{aligned}
\check{S}_{xy}(s,r) &= \check{S}_{xx}(s,r)\check{\mathsf{h}}^*(r^*)\\
\check{S}_{yy}(s,r) &= \check{S}_{xy}(s,r)\check{\mathsf{h}}(s)\\
\check{S}_{yy}(s,r) &= \check{S}_{xx}(s,r)\check{\mathsf{h}}(s)\check{\mathsf{h}}^*(r^*)
\end{aligned}$$

Power spectral density functions

$$\begin{aligned}
\mathsf{S}_{xy}(f,g) &= \mathsf{S}_{xx}(f,g)\tilde{\mathsf{h}}^*(-g)\\
\mathsf{S}_{yy}(f,g) &= \mathsf{S}_{xy}(f,g)\tilde{\mathsf{h}}(\omega)\\
\mathsf{S}_{yy}(f,g) &= \mathsf{S}_{xx}(f,g)\tilde{\mathsf{h}}(\omega)\tilde{\mathsf{h}}^*(-g)
\end{aligned}$$

[1] Papoulis (1991) page 312

✎PROOF:

$$
\begin{aligned}
\mathsf{R_{xy}}(t,u) &\triangleq \mathbf{E}\left[\mathsf{x}(t)\mathsf{y}^*(u)\right] && \text{by definition of } \mathsf{R_{xy}}(t,u) && \text{(Definition 3.2 page 17)}\\
&\triangleq \mathbf{E}[\mathsf{x}(t)(\mathsf{h}(t)\star\mathsf{x}(t))] && \text{by definition of } \mathsf{y}(t)\\
&= \mathbf{E}\left[\mathsf{x}(t)\left(\int_{v\in\mathbb{R}}\mathsf{h}(v)\mathsf{x}(u-v)\,\mathsf{d}v\right)^*\right] && \text{by definition of } \star && \text{(Definition D.1 page 199)}\\
&= \mathbf{E}\left[\mathsf{x}(t)\int_{v\in\mathbb{R}}\mathsf{h}^*(v)\mathsf{x}^*(u-v)\,\mathsf{d}v\right]\\
&= \int_{v\in\mathbb{R}}\mathsf{h}^*(v)\mathbf{E}\left[\mathsf{x}(t)\mathsf{x}^*(u-v)\right]\,\mathsf{d}v && \text{by \textit{linear} property of } \mathbf{E} && \text{(Theorem 1.1 page 4)}\\
&= \int_{v\in\mathbb{R}}\mathsf{h}^*(v)\mathsf{R_{xx}}(t,u-v)\,\mathsf{d}v && \text{by definition of } \mathsf{R_{xy}}(t,u) && \text{(Definition 3.2 page 17)}\\
&\triangleq \mathsf{R_{xx}}(t,u)\star h^*(u) && \text{by definition of } \star && \text{(Definition D.1 page 199)}
\end{aligned}
$$

$$
\begin{aligned}
\mathsf{R_{yy}}(t,u) &\triangleq \mathbf{E}\left[\mathsf{y}(t)\mathsf{y}^*(u)\right] && \text{by definition of } \mathsf{R_{xy}}(t,u) && \text{(Definition 3.2 page 17)}\\
&= \mathbf{E}\left[\left(\int_{v\in\mathbb{R}}\mathsf{h}(v)\mathsf{x}(t-v)\,\mathsf{d}v\right)\mathsf{y}^*(u)\right] && \text{by definition of } \star && \text{(Definition D.1 page 199)}\\
&= \int_{v\in\mathbb{R}}\mathsf{h}(v)\mathbf{E}\left[\mathsf{x}(t-v)\mathsf{y}^*(u)\right]\,\mathsf{d}v && \text{by \textit{linear} property of } \mathbf{E} && \text{(Theorem 1.1 page 4)}\\
&= \int_{v\in\mathbb{R}}\mathsf{h}(v)\mathsf{R_{xy}}(t-v,u)\,\mathsf{d}v && \text{by definition of } \mathsf{R_{xy}}(t,u) && \text{(Definition 3.2 page 17)}\\
&\triangleq \mathsf{R_{xy}}(t,u)\star\mathsf{h}(t) && \text{by definition of } \star && \text{(Definition D.1 page 199)}
\end{aligned}
$$

$$
\begin{aligned}
\mathsf{R_{yy}}(t,u) &\triangleq \mathbf{E}\left[\mathsf{y}(t)\mathsf{y}^*(u)\right]\\
&= \mathbf{E}\left[\left(\int_{v\in\mathbb{R}}\mathsf{h}(v)\mathsf{x}(t-v)\,\mathsf{d}v\right)\left(\int_{w\in\mathbb{R}}\mathsf{h}(w)\mathsf{x}(u-w)\,\mathsf{d}\omega\right)^*\right] && \text{by definition of } \star && \text{(Definition D.1 page 199)}\\
&= \int_{w\in\mathbb{R}}\mathsf{h}^*(w)\int_{v\in\mathbb{R}}\mathsf{h}(v)\mathbf{E}\left[\mathsf{x}(t-v)\mathsf{x}^*(u-w)\right]\,\mathsf{d}v\,\mathsf{d}\omega && \text{by \textit{linear} property of } \mathbf{E} && \text{(Theorem 1.1 page 4)}\\
&= \int_{w\in\mathbb{R}}\mathsf{h}^*(w)\int_{v\in\mathbb{R}}\mathsf{h}(v)\mathsf{R_{xx}}(t-v,u-w)\,\mathsf{d}v\,\mathsf{d}\omega && \text{by definition of } \mathsf{R_{xx}}(t,u) && \text{(Definition 3.2 page 17)}\\
&= \int_{w\in\mathbb{R}}\mathsf{h}^*(w)\left[\mathsf{R_{xx}}(t,u-w)\star\mathsf{h}(t)\right]\,\mathsf{d}\omega && \text{by definition of } \star && \text{(Definition D.1 page 199)}\\
&\triangleq \mathsf{R_{xx}}(t,u)\star\mathsf{h}(t)\star\mathsf{h}^*(u) && \text{by definition of } \star && \text{(Definition D.1 page 199)}
\end{aligned}
$$

$$
\begin{aligned}
\check{S}_{\mathsf{xy}}(s,r) &\triangleq \mathbf{L}\mathsf{R_{xy}}(t,u)\\
&= \mathbf{L}[\mathsf{R_{xx}}(t,u)\star\mathsf{h}^*(u)]\\
&= \mathbf{L}[\mathsf{R_{xx}}(t,u)]\mathbf{L}[\mathsf{h}^*(u)]\\
&= \check{S}_{\mathsf{xx}}(s,r)\int_{u\in\mathbb{R}}\mathsf{h}^*(u)e^{-ru}\,\mathsf{d}u\\
&= \check{S}_{\mathsf{xx}}(s,r)\left[\int_{u\in\mathbb{R}}\mathsf{h}(u)e^{-r^*u}\,\mathsf{d}u\right]^*\\
&= \check{S}_{\mathsf{xx}}(s,r)\check{\mathsf{h}}^*(r^*)
\end{aligned}
$$

$$
\begin{aligned}
\check{S}_{\mathsf{yy}}(s,r) &\triangleq \mathbf{L}\mathsf{R_{yy}}(t,u)\\
&= \mathbf{L}[\mathsf{R_{xy}}(t,u)\star\mathsf{h}(t)]\\
&= \mathbf{L}[\mathsf{R_{xy}}(t,u)]\mathbf{L}[\mathsf{h}(t)]\\
&= \check{S}_{\mathsf{xy}}(s,r)\check{\mathsf{h}}(s)
\end{aligned}
$$

$$
\begin{aligned}
&= \check{S}_{xy}(s,r)\check{\mathsf{h}}(s)\\
&= \check{S}_{xx}(s,r)\check{\mathsf{h}}^*(r^*)\check{\mathsf{h}}(s)\\
&= \check{S}_{xx}(s,r)\check{\mathsf{h}}(s)\check{\mathsf{h}}^*(r^*)
\end{aligned}
$$

$$
\begin{aligned}
\mathsf{S}_{xy}(f,g) &\triangleq \tilde{\mathbf{F}}\mathsf{R}_{xy}(t,u)\\
&= \tilde{\mathbf{F}}[\mathsf{R}_{xx}(t,u) \star \mathsf{h}^*(u)]\\
&= \tilde{\mathbf{F}}[\mathsf{R}_{xx}(t,u)]\tilde{\mathbf{F}}[\mathsf{h}^*(u)]\\
&= \mathsf{S}_{xx}(f,g)\int_{u\in\mathbb{R}} \mathsf{h}^*(u)e^{-i2\pi gu}\,\mathsf{d}u\\
&= \mathsf{S}_{xx}(f,g)\left[\int_{u\in\mathbb{R}} \mathsf{h}(u)e^{i2\pi gu}\,\mathsf{d}u\right]^*\\
&= \mathsf{S}_{xx}(f,g)\left[\int_{u\in\mathbb{R}} \mathsf{h}(u)e^{-i2\pi(-g)u}\,\mathsf{d}u\right]^*\\
&= \mathsf{S}_{xx}(f,g)\tilde{\mathsf{h}}^*(-g)
\end{aligned}
$$

$$
\begin{aligned}
\mathsf{S}_{yy}(f,g) &\triangleq \tilde{\mathbf{F}}\mathsf{R}_{yy}(t,u)\\
&= \tilde{\mathbf{F}}[\mathsf{R}_{xy}(t,u) \star \mathsf{h}(t)]\\
&= \tilde{\mathbf{F}}[\mathsf{R}_{xy}(t,u)]\tilde{\mathbf{F}}[\mathsf{h}(t)]\\
&= \mathsf{S}_{xy}(f,g)\tilde{\mathsf{h}}(\omega)\\
\\
&= \mathsf{S}_{xy}(f,g)\tilde{\mathsf{h}}(\omega)\\
&= \mathsf{S}_{xx}(f,g)\tilde{\mathsf{h}}^*(-g)\tilde{\mathsf{h}}(\omega)
\end{aligned}
$$

✏

8.2 LTI Operations on WSS random processes

Definition 8.1.

DEF

A random process $\mathsf{x}(t)$ *is* ***wide sense stationary (WSS)*** *if*

(1).	$\mu_{\mathsf{x}}(t)$	*is* CONSTANT *with respect to* t	(STATIONARY IN THE MEAN)	*and*
(2).	$\mathsf{R}_{xx}(t+\tau,t)$	*is* CONSTANT *with respect to* t	(STATIONARY IN CORRELATION)	

If a process $\mathsf{x}(t)$ is *wide sense stationary*, mean and correlation are often written μ_{x} and $\mathsf{R}_{xx}(\tau)$, respectively. If a pair of processes $\mathsf{x}(t)$ and $\mathsf{y}(t)$ are *WSS*, then their cross-correlation is commonly written $\mathsf{R}_{xy}(\tau)$.

Definition 8.2. *Let* $\mathsf{x}(t)$ *and* $\mathsf{y}(t)$ *be WSS random processes. Let* $\mathbf{L}$ *be the* LAPLACE TRANSFORM *operator.*

DEF

$$\check{S}_{xx}(s) \triangleq \mathbf{L}R_{xx}(\tau) \triangleq \int_{\tau\in\mathbb{R}} R_{xx}(\tau)e^{-s\tau}\,d\tau$$
$$\check{S}_{yy}(s) \triangleq \mathbf{L}R_{yy}(\tau) \triangleq \int_{\tau\in\mathbb{R}} R_{yy}(\tau)e^{-s\tau}\,d\tau$$
$$\check{S}_{xy}(s) \triangleq \mathbf{L}R_{xy}(\tau) \triangleq \int_{\tau\in\mathbb{R}} R_{xy}(\tau)e^{-s\tau}\,d\tau$$
$$\check{S}_{yx}(s) \triangleq \mathbf{L}R_{yx}(\tau) \triangleq \int_{\tau\in\mathbb{R}} R_{yx}(\tau)e^{-s\tau}\,d\tau$$

Definition 8.3. *Let* $x(t)$ *and* $y(t)$ *be WSS random processes.*

DEF

$$\tilde{S}_{xx}(\omega) \triangleq [\tilde{\mathbf{F}}R_{xx}(\tau)](\omega) \triangleq \int_{\tau\in\mathbb{R}} R_{xx}(\tau)e^{-i\omega\tau}\,d\tau$$
$$\tilde{S}_{yy}(\omega) \triangleq [\tilde{\mathbf{F}}R_{yy}(\tau)](\omega) \triangleq \int_{\tau\in\mathbb{R}} R_{yy}(\tau)e^{-i\omega\tau}\,d\tau$$
$$\tilde{S}_{xy}(\omega) \triangleq [\tilde{\mathbf{F}}R_{xy}(\tau)](\omega) \triangleq \int_{\tau\in\mathbb{R}} R_{xy}(\tau)e^{-i\omega\tau}\,d\tau$$
$$\tilde{S}_{yx}(\omega) \triangleq [\tilde{\mathbf{F}}R_{yx}(\tau)](\omega) \triangleq \int_{\tau\in\mathbb{R}} R_{yx}(\tau)e^{-i\omega\tau}\,d\tau$$

Definition 8.4. [2] *Let* $x(t)$ *be a random variable that is* STATIONARY IN THE MEAN *such that* $\mathbf{E}[x(t)]$ *is constant with respect to* t.

DEF

$x(t)$ *is* ***ergodic in the mean*** *if*

$$\underbrace{\mathbf{E}[x(t)]}_{\text{ENSEMBLE AVERAGE}} = \underbrace{\lim_{\tau\to\infty}\frac{1}{2\tau}\int_{-\tau}^{+\tau} x(t)\,dt}_{\text{TIME AVERAGE}}$$

Proposition 8.1.

PRP

$$\left\{ x(t) \textit{ is } \text{NON-STATIONARY} \right\} \implies \left\{ x(t) \textit{ is } \text{NOT ERGODIC IN THE MEAN} \right\}$$

✎PROOF: If $x(t)$ is *non-stationary*, then $\mathbf{E}[x(t)]$ is not constant with time. But $\lim_{\tau\to\infty}\frac{1}{2\tau}\int_{-\tau}^{+\tau} x(t)\,dt$ must be a constant (if it is *convergent*). ⇨

Definition 8.5. [3] *Let* $x(t)$ *be a* WIDE SENSE STATIONARY *random process.*

DEF

(1). *The* ***average power*** P_{avg} *of* $x(t)$ *is* $\quad P_{avg}x(t) \triangleq \lim_{\tau\to\infty}\frac{1}{2\tau}\int_{t\in\mathbb{R}} |x(t)|^2\,dt$

(2). *The* ***energy spectral density*** $|\tilde{x}(\omega)|^2$ *of* $x(t)$ *is* $\quad |\tilde{x}(\omega)|^2 \triangleq \left|\tilde{\mathbf{F}}x(t)\right|^2$

Remark 8.1 (**spectral power**). Why does $\tilde{S}_{xx}(\omega)$ deserve the name *power spectral density*? This is answered by Theorem 8.2 (next). But to elaborate further, note that $\tilde{S}_{xx}$ is the spectral representation of the statistical relationship (the *variance*) between samples of $x(t)$. For example, if there is no relationship, then $\tilde{S}_{xx}(\omega) = 1$. But in the case that $x(t)$ is *ergodic in the mean*, then $\tilde{S}_{xx}$ takes on an additional meaning—it describes the "spectral power" present in $x(t)$. This is demonstrated by the next theorem.

[2] Papoulis (1984) page 246 ⟨Mean-Ergodic processes⟩, Papoulis (2002) page 523 ⟨12-1 ERGODICITY⟩, KAY (1988) PAGE 58 ⟨3.6 ERGODICITY OF THE AUTOCORRELATION FUNCTION⟩, MANOLAKIS ET AL. (2005) PAGE 106 ⟨ERGODIC RANDOM PROCESSES⟩, KOOPMANS (1995) PAGES 53–61, CADZOW (1987) PAGE 378 ⟨11.13 ERGODIC TIME SERIES⟩, HELSTROM (1991) PAGE 336

[3] Bendat and Piersol (2010) page 177

Theorem 8.2. *Let* $\mathsf{x}(t)$ *be a* RANDOM PROCESS.

THM

$$\left\{\begin{array}{lll} \text{(A).} & \mathsf{x}(t) \text{ is ERGODIC IN THE MEAN} & \text{and} \\ \text{(B).} & \tilde{\mathsf{x}}(\omega) \text{ EXISTS} & \end{array}\right\} \implies \left\{\begin{array}{llll} \text{(1).} & \tilde{\mathsf{S}}_{\mathsf{xx}}(\omega) & = \lim_{\tau\to\infty} \frac{1}{2\tau} \underbrace{|\tilde{\mathsf{x}}(\omega)|^2}_{\text{(ESD)}} & \text{and} \\ \text{(2).} & \mathsf{P}_{\text{avg}}[\mathsf{x}(t)] & = \int_{\omega\in\mathbb{R}} \tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)\, \mathrm{d}\omega & \end{array}\right\}$$

✎PROOF:

$$\begin{aligned}
\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega) &\triangleq \int_{\tau\in\mathbb{R}} \mathsf{R}_{\mathsf{xx}}(\tau) e^{-i\omega\tau}\, \mathrm{d}\tau && \text{by definition of } \tilde{\mathsf{S}}_{\mathsf{xx}}(\omega) && \text{(Definition 8.3 page 60)} \\
&= \int_{\tau\in\mathbb{R}} \mathbf{E}\left[\mathsf{x}(t+\tau)\mathsf{x}^*(t)\right] e^{-i\omega\tau}\, \mathrm{d}\tau && \text{by definition of } \mathsf{R}_{\mathsf{xx}}(t) && \text{(Definition 3.2 page 17)} \\
&= \mathbf{E}\left[\mathsf{x}^*(t) \int_{\tau\in\mathbb{R}} \mathsf{x}(t+\tau) e^{-i\omega\tau}\, \mathrm{d}\tau\right] && \text{by \textit{linearity} of } \mathbf{E} \text{ operator} \\
&= \mathbf{E}\left[\mathsf{x}^*(t) \int_{u\in\mathbb{R}} \mathsf{x}(u) e^{-i\omega(u-t)}\, \mathrm{d}u\right] && \text{where } u \triangleq t+\tau \implies \tau = u - t \\
&= \mathbf{E}\left[\mathsf{x}^*(t) e^{i\omega t} \int_{u\in\mathbb{R}} \mathsf{x}(u) e^{-i\omega u}\, \mathrm{d}u\right] \\
&= \mathbf{E}\left[\mathsf{x}^*(t) e^{i\omega t} \tilde{\mathsf{x}}(\omega)\right] && \text{by definition of \textit{Fourier Transform}} && \text{(Definition T.2 page 408)} \\
&= \mathbf{E}\left[\mathsf{x}^*(t) e^{i\omega t}\right] \tilde{\mathsf{x}}(\omega) && \text{by hypothesis (B)} \\
&= \left[\lim_{\tau\to\infty} \frac{1}{2\tau} \int_{-\tau}^{+\tau} \mathsf{x}^*(t) e^{i\omega t}\, \mathrm{d}t\right] \tilde{\mathsf{x}}(\omega) && \text{by \textit{ergodic in the mean} hypothesis} && \text{(Definition 8.4 page 60)} \\
&= \lim_{\tau\to\infty} \frac{1}{2\tau} \left[\int_{t\in\mathbb{R}} \mathsf{x}(t) e^{-i\omega t}\, \mathrm{d}t\right]^* \tilde{\mathsf{x}}(\omega) \\
&= \lim_{\tau\to\infty} \frac{1}{2\tau} \tilde{\mathsf{x}}^*(\omega)\tilde{\mathsf{x}}(\omega) && \text{by hypothesis (B)} \\
&= \lim_{\tau\to\infty} \frac{1}{2\tau} |\tilde{\mathsf{x}}(\omega)|^2
\end{aligned}$$

$$\begin{aligned}
\int_{\omega\in\mathbb{R}} \tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)\, \mathrm{d}\omega &= \int_{\omega\in\mathbb{R}} \lim_{\tau\to\infty} \frac{1}{2\tau} |\tilde{\mathsf{x}}(\omega)|^2\, \mathrm{d}\omega \\
&= \lim_{\tau\to\infty} \frac{1}{2\tau} \int_{\omega\in\mathbb{R}} |\tilde{\mathsf{x}}(\omega)|^2\, \mathrm{d}\omega \\
&= \lim_{\tau\to\infty} \frac{1}{2\tau} \int_{t\in\mathbb{R}} |\mathsf{x}(t)|^2\, \mathrm{d}t && \text{by \textit{Plancheral's formula}} && \text{(Theorem T.3 page 409, Theorem K.10 page 273)} \\
&= \mathsf{P}_{\text{avg}} && \text{by definition of } \mathsf{P}_{\text{avg}} && \text{(Definition 8.5 page 60)}
\end{aligned}$$

Thus, $\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)$ is the power density of $\mathsf{x}(t)$ in the frequency domain. ✏

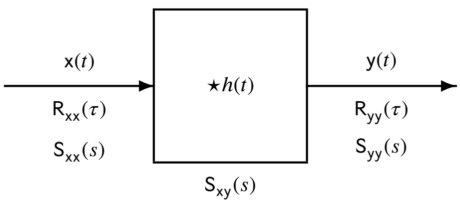


Figure 8.2: Linear system with WSS random process input and output

Theorem 8.3. [4] *Let* $\mathsf{h} : \mathbb{R} \to \mathbb{C}$ *be the impulse response of a linear time-invariant system and let* $\mathsf{y}(t) = h(t) \star \mathsf{x}(t) \triangleq \int_{u\in\mathbb{R}} h(u)\mathsf{x}(t-u)\,\mathsf{d}u$ *as illustrated in Figure 8.1 page 57. Then*

THM

$$\begin{aligned}
\mathsf{R}_{\mathsf{xy}}(\tau) &= \mathsf{R}_{\mathsf{xx}}(\tau) \star h^*(-\tau) &&\triangleq \int_{u\in\mathbb{R}} h^*(-u)\mathsf{R}_{\mathsf{xx}}(\tau-u)\,\mathsf{d}u\\
\mathsf{R}_{\mathsf{yy}}(\tau) &= \mathsf{R}_{\mathsf{xy}}(\tau) \star h(\tau) &&\triangleq \int_{u\in\mathbb{R}} h(u)\mathsf{R}_{\mathsf{xy}}(\tau-u)\,\mathsf{d}u\\
\mathsf{R}_{\mathsf{yy}}(\tau) &= \mathsf{R}_{\mathsf{xx}}(\tau) \star h(\tau) \star h^*(-\tau) &&\triangleq \int_{v\in\mathbb{R}}\int_{u\in\mathbb{R}} h(u-v)h^*(-v)\mathsf{R}_{\mathsf{xx}}(\tau-u-v)\,\mathsf{d}u\,\mathsf{d}v\\
\\
\mathsf{S}_{\mathsf{xy}}(s) &= \mathsf{S}_{\mathsf{xx}}(s)\hat{h}^*(-s^*)\\
\mathsf{S}_{\mathsf{yy}}(s) &= \mathsf{S}_{\mathsf{xy}}(s)\hat{h}(s)\\
\mathsf{S}_{\mathsf{yy}}(s) &= \mathsf{S}_{\mathsf{xx}}(s)\hat{h}(s)\hat{h}^*(-s^*)\\
\\
\tilde{\mathsf{S}}_{\mathsf{xy}}(\omega) &= \tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)\tilde{\mathsf{h}}^*(\omega)\\
\tilde{\mathsf{S}}_{\mathsf{yy}}(\omega) &= \tilde{\mathsf{S}}_{\mathsf{xy}}(\omega)\tilde{\mathsf{h}}(\omega)\\
\tilde{\mathsf{S}}_{\mathsf{yy}}(\omega) &= \tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)|\tilde{\mathsf{h}}(\omega)|^2
\end{aligned}$$

✎PROOF:

$$\begin{aligned}
\mathsf{R}_{\mathsf{xx}}(\tau) \star h^*(-\tau) &\triangleq \int_{u\in\mathbb{R}} h^*(-u)\mathsf{R}_{\mathsf{xx}}(\tau-u)\,\mathsf{d}u\\
&= \int_{u\in\mathbb{R}} h^*(-u)\mathbf{E}\left[\mathsf{x}(t)\mathsf{x}^*(t-\tau+u)\right]\mathsf{d}u\\
&= \mathbf{E}\left[\mathsf{x}(t)\int_{u\in\mathbb{R}} h^*(-u)\mathsf{x}^*(t-\tau+u)\,\mathsf{d}u\right]\\
&= \mathbf{E}\left[\mathsf{x}(t)\int_{u\in\mathbb{R}} h^*(u')\mathsf{x}^*(t-\tau-u')\,\mathsf{d}u'\right]\\
&= \mathbf{E}\left[\mathsf{x}(t)\mathsf{y}^*(t-\tau)\right]\\
&\triangleq \mathsf{R}_{\mathsf{xy}}(\tau)
\end{aligned}$$

$$\begin{aligned}
\mathsf{R}_{\mathsf{xy}}(\tau) \star h(\tau) &\triangleq \int_{u\in\mathbb{R}} h(u)\mathsf{R}_{\mathsf{xy}}(\tau-u)\,\mathsf{d}u\\
&= \int_{u\in\mathbb{R}} h(u)\mathbf{E}\left[\mathsf{x}(t+\tau-u)\mathsf{y}^*(t)\right]\mathsf{d}u\\
&= \mathbf{E}\left[\mathsf{y}^*(t)\int_{u\in\mathbb{R}} h(u)\mathsf{x}(t+\tau-u)\,\mathsf{d}u\right]\\
&= \mathbf{E}\left[\mathsf{y}^*(t)\mathsf{y}(t+\tau)\right]\\
&= \mathbf{E}\left[\mathsf{y}(t+\tau)\mathsf{y}^*(t)\right]\\
&\triangleq \mathsf{R}_{\mathsf{yy}}(\tau)
\end{aligned}$$

$$\begin{aligned}
\mathsf{R}_{\mathsf{yy}}(\tau) &= \mathsf{R}_{\mathsf{xy}}(\tau) \star h(\tau)\\
&= \mathsf{R}_{\mathsf{xx}}(\tau) \star h^*(-\tau) \star h(\tau)\\
&= \mathsf{R}_{\mathsf{xx}}(\tau) \star h(\tau) \star h^*(-\tau)
\end{aligned}$$

$$\begin{aligned}
\mathsf{S}_{\mathsf{xy}}(s) &\triangleq \mathbf{L}\mathsf{R}_{\mathsf{xy}}(\tau)\\
&\triangleq \int_{\tau\in\mathbb{R}} \mathsf{R}_{\mathsf{xy}}(\tau)e^{-s\tau}\,d\tau\\
&= \int_{\tau\in\mathbb{R}} \left[\mathsf{R}_{\mathsf{xx}}(\tau) \star h^*(-\tau)\right]e^{-s\tau}\,d\tau
\end{aligned}$$

[4] Papoulis (1991) pages 323–324

$$\begin{aligned}
&= \int_{\tau\in\mathbb{R}} \left[\int_{u\in\mathbb{R}} h^*(-u)\mathsf{R}_{xx}(\tau-u)\,\mathsf{d}u\right] e^{-s\tau}\,d\tau \\
&= \int_{u\in\mathbb{R}} h^*(-u) \int_{\tau\in\mathbb{R}} \mathsf{R}_{xx}(\tau-u)e^{-s\tau}\,d\tau\,\mathsf{d}u \\
&= \int_{u\in\mathbb{R}} h^*(-u) \int_{v\in\mathbb{R}} \mathsf{R}_{xx}(v)e^{-s(v+u)}\,\mathsf{d}v\,\mathsf{d}u && \text{where } v = \tau - u \iff \tau = v+u \\
&= \int_{u\in\mathbb{R}} h^*(-u)e^{-su}\,\mathsf{d}u \int_{v\in\mathbb{R}} \mathsf{R}_{xx}(v)e^{-sv}\,\mathsf{d}v \\
&= \int_{u\in\mathbb{R}} h^*(u)e^{-s(-u)}\,\mathsf{d}u \int_{v\in\mathbb{R}} \mathsf{R}_{xx}(v)e^{-sv}\,\mathsf{d}v \\
&= \left(\int_{u\in\mathbb{R}} h(u)e^{-(-s^*)u}\,\mathsf{d}u\right)^* \int_{v\in\mathbb{R}} \mathsf{R}_{xx}(v)e^{-sv}\,\mathsf{d}v \\
&\triangleq \hat{h}^*(-s^*)\mathsf{S}_{xx}(s)
\end{aligned}$$

$$\begin{aligned}
\mathsf{S}_{yy}(s) &\triangleq \mathbf{L}\mathsf{R}_{yy}(\tau) \\
&\triangleq \int_{\tau\in\mathbb{R}} \mathsf{R}_{yy}(\tau)e^{-s\tau}\,d\tau \\
&= \int_{\tau\in\mathbb{R}} \left[\mathsf{R}_{xy}(\tau)\star h(\tau)\right]e^{-s\tau}\,d\tau \\
&= \int_{\tau\in\mathbb{R}} \left[\int_{u\in\mathbb{R}} h(u)\mathsf{R}_{xy}(\tau-u)\,\mathsf{d}u\right] e^{-s\tau}\,d\tau \\
&= \int_{u\in\mathbb{R}} h(u) \int_{\tau\in\mathbb{R}} \mathsf{R}_{xy}(\tau-u)e^{-s\tau}\,d\tau\,\mathsf{d}u \\
&= \int_{u\in\mathbb{R}} h(u) \int_{\tau\in\mathbb{R}} \mathsf{R}_{xy}(v)e^{-s(v+u)}\,d\tau\,\mathsf{d}u && \text{where } v = \tau - u \iff \tau = v+u \\
&= \int_{u\in\mathbb{R}} h(u)e^{-su}\,du \int_{\tau\in\mathbb{R}} \mathsf{R}_{xy}(v)e^{-sv}\,d\tau \\
&\triangleq \hat{h}(s)\mathsf{S}_{xy}(s)
\end{aligned}$$

$$\begin{aligned}
\mathsf{S}_{yy}(s) &= \hat{h}(s)\mathsf{S}_{xy}(s) \\
&= \hat{h}(s)\hat{h}^*(-s^*)\mathsf{S}_{xx}(s)
\end{aligned}$$

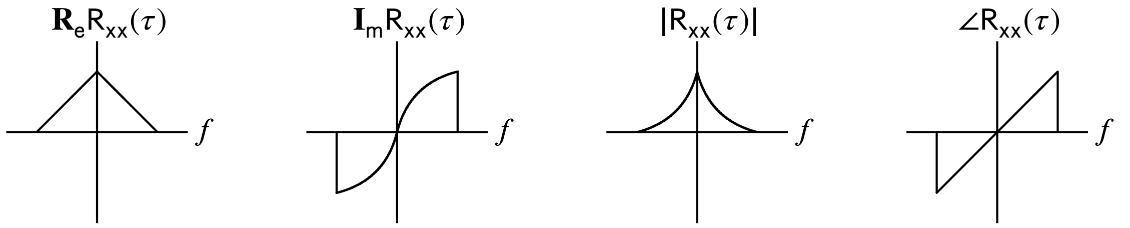
$$\begin{aligned}
\tilde{\mathsf{S}}_{xy}(\omega) &= \mathsf{S}_{xy}(s)\big|_{s=j\omega} \\
&= \hat{h}^*(-s^*)\mathsf{S}_{xx}(s)\big|_{s=j\omega} \\
&= \left(\int_{u\in\mathbb{R}} h(u)e^{-(-s^*)u}\,\mathsf{d}u\right)^* \int_{v\in\mathbb{R}} \mathsf{R}_{xx}(v)e^{-sv}\,\mathsf{d}v\bigg|_{s=j\omega} \\
&= \left(\int_{u\in\mathbb{R}} h(u)e^{-(-j\omega)^*u}\,\mathsf{d}u\right)^* \int_{v\in\mathbb{R}} \mathsf{R}_{xx}(v)e^{-j\omega v}\,\mathsf{d}v \\
&= \left(\int_{u\in\mathbb{R}} h(u)e^{-j\omega u}\,\mathsf{d}u\right)^* \int_{v\in\mathbb{R}} \mathsf{R}_{xx}(v)e^{-j\omega v}\,\mathsf{d}v \\
&\triangleq \tilde{\mathsf{h}}^*(\omega)\tilde{\mathsf{S}}_{xx}(\omega)
\end{aligned}$$

$$\begin{aligned}
\tilde{\mathsf{S}}_{yy}(\omega) &= \mathsf{S}_{yy}(s)\big|_{s=j\omega} \\
&= \hat{h}(s)\mathsf{S}_{xy}(s)\big|_{s=j\omega} \\
&= \int_{u\in\mathbb{R}} h(u)e^{-su}\,du \int_{\tau\in\mathbb{R}} \mathsf{R}_{xy}(v)e^{-sv}\,d\tau\bigg|_{s=j\omega}
\end{aligned}$$

$$
\begin{aligned}
&= \int_{u\in\mathbb{R}} h(u)e^{-j\omega u}\,du \int_{\tau\in\mathbb{R}} \mathsf{R}_{xy}(v)e^{-j\omega v}\,d\tau \\
&= \tilde{\mathsf{h}}(\omega)\tilde{\mathsf{S}}_{xy}(\omega)
\end{aligned}
$$

$$
\begin{aligned}
\tilde{\mathsf{S}}_{yy}(\omega) &= \tilde{\mathsf{h}}(\omega)\tilde{\mathsf{S}}_{xy}(\omega) \\
&= \tilde{\mathsf{h}}(\omega)\tilde{\mathsf{h}}^*(\omega)\tilde{\mathsf{S}}_{xx}(\omega) \\
&= |\tilde{\mathsf{h}}(\omega)|^2\tilde{\mathsf{S}}_{xx}(\omega)
\end{aligned}
$$

⇰



(*symmetric*) (*anti-symmetric*) (*symmetric*) (*anti-symmetric*)

Figure 8.3: *auto-correlation* $\mathsf{R}_{xx}(\tau)$

Theorem 8.4. *Let* $\mathsf{x} : \mathbb{R} \to \mathbb{C}$ *be a WSS random process with auto-correlation* $\mathsf{R}_{xx}(\tau)$*. Then* $\mathsf{R}_{xx}(\tau)$ *is* ***conjugate symmetric*** *such that (see Figure 8.3 page 64)*

THM

$$
\begin{array}{rcll}
\mathsf{R}_{xx}(\tau) &=& \mathsf{R}^*_{xx}(-\tau) & (\text{CONJUGATE SYMMETRIC}) \\
\mathbf{R_e}\left[\mathsf{R}_{xx}(\tau)\right] &=& \mathbf{R_e}\left[\mathsf{R}_{xx}(-\tau)\right] & (\text{SYMMETRIC}) \\
\mathbf{I_m}\left[\mathsf{R}_{xx}(\tau)\right] &=& -\mathbf{I_m}\left[\mathsf{R}_{xx}(-\tau)\right] & (\text{ANTI-SYMMETRIC}) \\
|\mathsf{R}_{xx}(\tau)| &=& |\mathsf{R}_{xx}(-\tau)| & (\text{SYMMETRIC}) \\
\angle\mathsf{R}_{xx}(\tau) &=& \angle\mathsf{R}_{xx}(-\tau) & (\text{ANTI-SYMMETRIC}).
\end{array}
$$

✎PROOF:

$$
\begin{aligned}
\mathsf{R}^*_{xx}(\tau) &\triangleq \left(\mathbf{E}\left[\mathsf{x}(t-\tau)\mathsf{x}^*(t)\right]\right)^* \\
&= \mathbf{E}\left[\mathsf{x}^*(t-\tau)\mathsf{x}(t)\right] \\
&= \mathbf{E}\left[\mathsf{x}(t)\mathsf{x}^*(t-\tau)\right] \\
&= \mathbf{E}\left[\mathsf{x}(u+\tau)\mathsf{x}^*(u)\right] && \text{where } u \triangleq t-\tau \iff t = u+\tau \\
&\triangleq \mathsf{R}_{xx}(\tau)
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{R_e}\left[\mathsf{R}_{xx}(\tau)\right] &= \mathbf{R_e}\left[\mathsf{R}^*_{xx}(-\tau)\right] && = \mathbf{R_e}\left[\mathsf{R}_{xx}(-\tau)\right] \\
\mathbf{I_m}\left[\mathsf{R}_{xx}(\tau)\right] &= \mathbf{I_m}\left[\mathsf{R}^*_{xx}(-\tau)\right] && = -\mathbf{I_m}\left[\mathsf{R}_{xx}(-\tau)\right] \\
abs\mathsf{R}_{xx}(\tau) &= \left|\mathsf{R}^*_{xx}(-\tau)\right| && = \left|\mathsf{R}_{xx}(-\tau)\right| \\
\angle\mathsf{R}_{xx}(\tau) &= \angle\mathsf{R}^*_{xx}(-\tau) && = -\angle\mathsf{R}_{xx}(-\tau)
\end{aligned}
$$

⇰

8.3 Whitening continuous random sequences

Simple algebraic operations on white noise processes (processes with autocorrelation $\mathsf{R}_{xx}(\tau) = \delta(\tau)$) often produce *colored* noise (processes with autocorrelation $\mathsf{R}_{xx}(\tau) \neq \delta(\tau)$). Sometimes we would like to process a colored noise process to produce a white noise process. This operation is known as *whitening*. Reasons for why we may want to whiten a noise process include

1. Samples from a white noise process are uncorrelated. If the noise process is Gaussian, then these samples are also independent which often greatly simplifies analysis.
2. Any orthonormal basis can be used to decompose a white noise process. This is not true of a colored noise process. Karhunen–Loève expansion can be used to decompose colored noise. [5]

Definition 8.6. *A **rational expression** $\mathsf{p}(s)$ is a polynomial divided by a polynomial such that*

DEF
$$\mathsf{p}(s) = \frac{\sum_{n=0}^{N} b_n s^n}{\sum_{n=0}^{M} a_n s^n}.$$
*The **zeros** of a rational expression are the roots of its numerator polynomial. The **poles** of a rational expression are the roots of its denominator polynomial.*

Definition 8.7. *Let $\check{\mathsf{h}}(s)$ be the Laplace transform of the impulse response of a filter. If $\check{\mathsf{h}}(s)$ can be expressed as a rational expression with poles and zeros at $a_n + ib_n$, then the filter is **minimum phase** if each $a_n < 0$ (all roots lie in the left hand side of the complex s-plane).*

Note that if $L(s)$ has a root at $s = -a + ib$, then $L^*(-s^*)$ has a root at

$$-s^* = -(-a+ib)^* = -(-a-ib) = a + ib.$$

That is, if $L(s)$ has a root in the left hand plane, then $L^*(-s^*)$ has a root directly opposite across the imaginary axis in the right hand plane (see Figure 8.4 page 65). A causal stable filter $\hat{h}(s)$ must have all of its poles in the left hand plane. A minimum phase filter is a filter with both its poles and zeros in the left hand plane. One advantage of a minimum phase filter is that its reciprical (zeros become poles and poles become zeros) is also causal and stable.

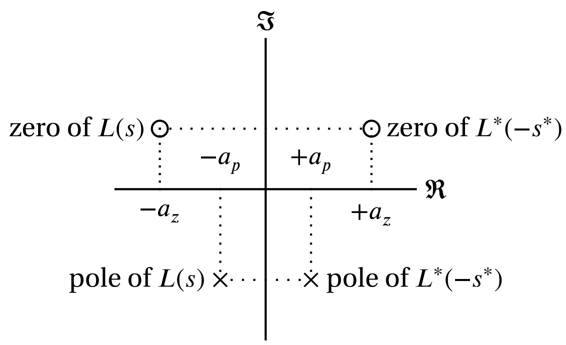


Figure 8.4: Mirrored roots in complex-s plane

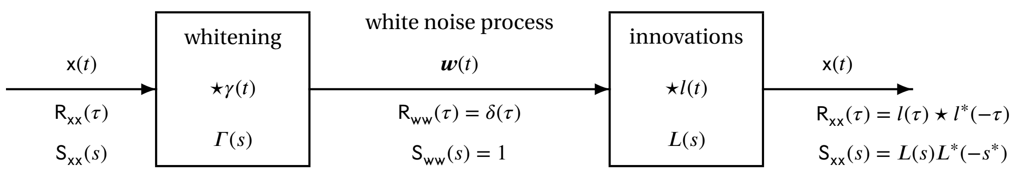


Figure 8.5: Innovations and whitening filters

The next theorem demonstrates a method for "whitening" a random process $\mathsf{x}(t)$ with a filter constructed from a decomposition of $\mathsf{R_{xx}}(\tau)$. The technique is stated precisely in Theorem 8.5 page 66

[5] *Karhunen–Loève expansion*: Section 5.1 page 25

and illustrated in Figure 8.5 page 65. Both imply two filters with impulse responses $l(t)$ and $\gamma(t)$. Filter $l(t)$ is referred to as the **innovations filter** (because it generates or "innovates" $\mathsf{x}(t)$ from a white noise process $\mathsf{w}(t)$) and $\gamma(t)$ is referred to as the **whitening filter** because it produces a white noise sequence when the input sequence is $\mathsf{x}(t)$. [6]

Theorem 8.5. *Let $\mathsf{x}(t)$ be a WSS random process with autocorrelation $\mathsf{R}_{\mathsf{xx}}(\tau)$ and spectral density $\mathsf{S}_{\mathsf{xx}}(s)$. **If** $\mathsf{S}_{\mathsf{xx}}(s)$ has a **rational expression**, then the following are true:*

1. *There exists a rational expression $L(s)$ with minimum phase such that*

$$\mathsf{S}_{\mathsf{xx}}(s) = L(s)L^*(-s^*).$$

2. *An LTI filter for which the Laplace transform of the impulse response $\gamma(t)$ is*

$$\Gamma(s) = \frac{1}{L(s)}$$

is both causal and stable.

3. *If $\mathsf{x}(t)$ is the input to the filter $\gamma(t)$, the output $\mathsf{y}(t)$ is a **white noise sequence** such that*

$$\mathsf{S}_{\mathsf{yy}}(s) = 1 \qquad\qquad \mathsf{R}_{\mathsf{yy}}(\tau) = \delta(\tau).$$

✎Proof:

$$\begin{aligned}
\mathsf{S}_{\mathsf{ww}}(s) &= \Gamma(s)\Gamma^*(-s^*)\mathsf{S}_{\mathsf{xx}}(s) \\
&= \frac{1}{L(s)}\frac{1}{L^*(-s^*)}\mathsf{S}_{\mathsf{xx}}(s) \\
&= \frac{1}{L(s)}\frac{1}{L^*(-s^*)}L(s)L^*(-s^*) \\
&= 1
\end{aligned}$$

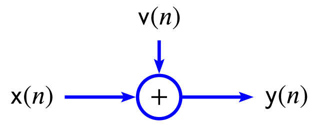
✎

[6] Papoulis (1991) pages 401–402

CHAPTER 9

ADDITIVE NOISE ON RANDOM SEQUENCES

9.1 Additive noise and correlation



Theorem 9.1. *Let* **S** *be the* SYSTEM *illustrated to the right, where* **T** *is* NOT NECESSARILY LINEAR.

THM

$$\left.\begin{array}{llll}(\text{A}). & \mathsf{x}(n) \text{ is} & \textit{WSS} & \text{and}\\ (\text{B}). & \mathsf{x}(n) \text{ and } \mathsf{v}(n) \text{ are} & \textit{uncorrelated} & \text{and}\\ (\text{C}). & \mathsf{v}(n) \text{ is} & \textit{zero-mean} & \end{array}\right\} \implies \left\{\begin{array}{lllll}(1). & \mathsf{R}_{yv}(m) & = & \mathsf{R}_{vv}(m) & \text{and}\\ (2). & \mathsf{R}_{xy}(m) & = & \mathsf{R}_{xx}(m) & \text{and}\\ (3). & \mathsf{R}_{yy}(m) & = & \mathsf{R}_{xx}(m)+\mathsf{R}_{vv}(m) & \text{and}\\ (4). & \mathsf{R}_{xx}(m) & = & \mathsf{R}_{yy}(m)+\mathsf{R}_{vv}(m)-2\mathbf{R_e}\mathsf{R}_{yv}(m) & \end{array}\right.$$

✎PROOF:

$$\begin{aligned}
\mathsf{R}_{yv}(m) &\triangleq \mathbf{E}\left[\mathsf{y}(m)\mathsf{v}^*(0)\right] && \text{by (A) and definition of } \mathsf{R}_{yv} && \text{(Definition 2.4 page 12)}\\
&\triangleq \mathbf{E}\left[(\mathsf{x}(m)+\mathsf{v}(m))\mathsf{v}^*(0)\right] && \text{by definition of } \mathsf{y}\\
&= \mathbf{E}\left[\mathsf{x}(m)\mathsf{v}^*(0)\right]+\mathbf{E}\left[\mathsf{v}(m)\mathsf{v}^*(0)\right] && \text{by \textit{linearity} of } \mathbf{E} && \text{(Theorem 1.1 page 4)}\\
&= \mathbf{E}\mathsf{x}(m)\mathbf{E}\mathsf{v}^*(0)+\mathbf{E}\left[\mathsf{v}(m)\mathsf{v}^*(0)\right] && \text{by \textit{uncorrelated} hypothesis} && \text{(B)}\\
&= \mathbf{E}\mathsf{x}(m)\cancelto{0}{\mathbf{E}\mathsf{v}^*(0)}+\mathbf{E}\left[\mathsf{v}(m)\mathsf{v}^*(0)\right] && \text{by \textit{zero-mean} hypothesis} && \text{(C)}\\
&= \mathsf{R}_{vv}(m) && \text{by definition of } \mathsf{R}_{vv} && \text{(Definition 2.4 page 12)}\\
\mathsf{R}_{xy}(m) &\triangleq \mathbf{E}\left[\mathsf{x}(m)\mathsf{y}^*(0)\right] && \text{by (A) and definition of } \mathsf{R}_{xy} && \text{(Definition 2.4 page 12)}\\
&\triangleq \mathbf{E}\left(\mathsf{x}(m)[\mathsf{x}(0)+\mathsf{v}(0)]^*\right) && \text{by definition of } \mathsf{y}\\
&= \mathbf{E}\left[\mathsf{x}(m)\mathsf{x}^*(0)\right]+\mathbf{E}\left[\mathsf{x}(m)\mathsf{v}^*(0)\right] && \text{by \textit{linearity} of } \mathbf{E} && \text{(Theorem 1.1 page 4)}\\
&= \mathbf{E}\left[\mathsf{x}(m)\mathsf{x}^*(0)\right]+\mathbf{E}[\mathsf{x}(m)]\mathbf{E}\left[\mathsf{v}^*(0)\right] && \text{by \textit{uncorrelated} hypothesis} && \text{(B)}\\
&= \mathbf{E}\left[\mathsf{x}(m)\mathsf{x}^*(0)\right]+\mathbf{E}[\mathsf{x}(m)]\cancelto{0}{\mathbf{E}\left[\mathsf{v}^*(0)\right]} && \text{by \textit{zero-mean} hypothesis} && \text{(C)}\\
&= \mathsf{R}_{xx}(m) && \text{by definition of } \mathsf{R}_{xx} && \text{(Definition 2.4 page 12)}
\end{aligned}$$

$$\begin{aligned}
\mathsf{R}_{yy}(m) &\triangleq \mathbf{E}\left[\mathsf{y}(m)\mathsf{y}^*(0)\right] && \text{by (A) and definition of } \mathsf{R}_{yy}\\
&\triangleq \mathbf{E}\left[(\mathsf{x}(m)+\mathsf{v}(m))(\mathsf{x}(0)+\mathsf{v}(0))^*\right] && \text{by definition of } \mathsf{y}\\
&= \mathbf{E}\left[\mathsf{x}(m)\mathsf{x}^*(0)\right]+\mathbf{E}\left[\mathsf{x}(m)\mathsf{v}^*(0)\right]+\mathbf{E}\left[\mathsf{v}(m)\mathsf{x}^*(0)\right]+\mathbf{E}\left[\mathsf{v}(m)\mathsf{v}^*(0)\right]\\
&= \mathbf{E}\left[\mathsf{x}(m)\mathsf{x}^*(0)\right]+\mathbf{E}\mathsf{x}(m)\mathbf{E}\mathsf{v}^*(0)+\mathbf{E}\mathsf{v}(m)\mathbf{E}\mathsf{x}^*(0)+\mathbf{E}\left[\mathsf{v}(m)\mathsf{v}^*(0)\right] && \text{by \textit{uncorrelated} hypothesis (B)}\\
&= \mathbf{E}\left[\mathsf{x}(m)\mathsf{x}^*(0)\right]+\mathbf{E}\mathsf{x}(m)\cancelto{0}{\mathbf{E}\mathsf{v}^*(0)}+\cancelto{0}{\mathbf{E}\mathsf{v}(m)}\mathbf{E}\mathsf{x}^*(0)+\mathbf{E}\left[\mathsf{v}(m)\mathsf{v}^*(0)\right] && \text{by \textit{zero-mean} hypothesis (C)}
\end{aligned}$$

$$\begin{aligned}
&= R_{xx}(m) + R_{vv}(m) && \text{by definition of } R_{xx} \\
R_{xx}(m) &\triangleq \mathbf{E}\left[x(m)x^*(0)\right] \\
&\triangleq \mathbf{E}\Big(\left[y(m) - v(m)\right]\left[y(0) - v(0)\right]^*\Big) \\
&= \mathbf{E}\left[y(m)y^*(0)\right] - \mathbf{E}\left[y(m)v^*(0)\right] - \mathbf{E}\left[v(m)y^*(0)\right] + \mathbf{E}\left[v(m)v^*(0)\right] \\
&\triangleq R_{yy}(m) - R_{yv}(m) - R_{vy}(m) + R_{vv}(m) \\
&= R_{yy}(m) + R_{vv}(m) - 2\mathbf{R_e}R_{yv}(m)
\end{aligned}$$

✏

Remark 9.1. Because in Theorem 9.1 $y = x + v$ and $R_{yy} = R_{xx} + R_{vv}$, one might assume that R is a kind of *linear operator* (Definition R.3 page 360) and further assume that because $x = y - v$ and $R_{(-v)(-v)} = R_{vv}$, that $R_{xx} = R_{yy} + R_{vv}$. As Theorem 9.1 demonstrates, this is simply **not the case**. The problem here is that y and v are very much *correlated*—in fact y is obviously a *function* of v.

Corollary 9.1. *Let* **S** *be the* SYSTEM *illustrated in Theorem 9.1 (page 67).*

COR

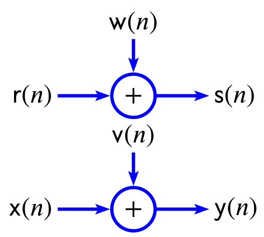
$$\left.\begin{array}{l}\textit{hypotheses of}\\ \textit{Theorem 9.1 (page 67)}\end{array}\right\} \Longrightarrow \left\{\begin{array}{llcll}
(1). & \check{S}_{yy}(z) & = & \check{S}_{xx}(z) + \check{S}_{vv}(z) & \textit{and}\\
(2). & \check{S}_{yv}(z) & = & \check{S}_{vv}(z) & \textit{and}\\
(3). & \check{S}_{yy}(z) & = & \check{S}_{yy}(z) + \check{S}_{vv}(z) + \check{S}_{yv}(z) + \check{S}^*_{yv}(z^*) & \textit{and}\\
(4). & \tilde{S}_{yy}(\omega) & = & \tilde{S}_{xx}(\omega) + \tilde{S}_{vv}(\omega) & \textit{and}\\
(5). & \tilde{S}_{yv}(\omega) & = & \tilde{S}_{vv}(\omega) & \textit{and}\\
(6). & \tilde{S}_{yv}(\omega) & = & \tilde{S}_{yy}(\omega) + \tilde{S}_{vv}(\omega) + \tilde{S}_{yv}(\omega) + \tilde{S}^*_{yv}(-\omega) &
\end{array}\right.$$

✎PROOF:

$$\begin{aligned}
\check{S}_{yy}(z) &\triangleq \mathbf{Z}R_{yy}(m) && \text{by definition of } \check{S}_{xx} && \text{(Definition 2.5 page 15)}\\
&= \mathbf{Z}R_{qq}(m) + \mathbf{Z}R_{vv}(m) && \text{by previous result} && (1)\\
&= \check{S}_{qq}(z) + \check{S}_{vv}(z) && \text{by definition of } \check{S}_{yy} && \text{(Definition 8.3 page 60)}\\
\tilde{S}_{yy}(\omega) &\triangleq \check{\mathbf{F}}R_{yy}(m) && \text{by definition of } \tilde{S}_{yy} && \text{(Definition 8.3 page 60)}\\
&= \check{\mathbf{F}}R_{qq}(m) + \check{\mathbf{F}}R_{vv}(m) && \text{by previous result} && (1)\\
&= \tilde{S}_{qq}(\omega) + \tilde{S}_{vv}(\omega) && \text{by definition of } \tilde{S}_{yy} && \text{(Definition 8.3 page 60)}
\end{aligned}$$

✏

Theorem 9.2. *Let* **S** *be the* SYSTEM *illustrated to the right:*



THM

$$\left\{\begin{array}{lrcll}
\text{(A).} & \multicolumn{3}{l}{x(n) \text{ and } r(n) \text{ are } \textit{wide sense stationary}} & \text{and}\\
\text{(B).} & \mathbf{E}[x(n)w(n)] & = & \mathbf{E}x(n)\mathbf{E}w(n) \quad (\textit{uncorrelated}) & \text{and}\\
\text{(C).} & \mathbf{E}[r(n)v(n)] & = & \mathbf{E}r(n)\mathbf{E}v(n) \quad (\textit{uncorrelated}) & \text{and}\\
\text{(D).} & \mathbf{E}[w(n)v(n)] & = & \mathbf{E}w(n)\mathbf{E}v(n) \quad (\textit{uncorrelated}) & \text{and}\\
\text{(E).} & \mathbf{E}v(n) = \mathbf{E}w(n) & = & 0 \quad (\textit{zero-mean}) &
\end{array}\right\} \Longrightarrow \left\{\begin{array}{rcl}
R_{sy}(m) & = & R_{sx}(m)\\
& = & R_{ry}(m)\\
& = & R_{rx}(m)
\end{array}\right\}$$

✎PROOF:

$$\begin{aligned}
\mathsf{R}_{sy}(m) &\triangleq \mathbf{E}\left[\mathsf{s}(m)\mathsf{y}^*(0)\right] && \text{by definition of } \mathsf{R}_{xy} && \text{(Definition 2.4 page 12)}\\
&\triangleq \mathbf{E}\left([\mathsf{r}(m)+\mathsf{w}(m)][\mathsf{x}(0)+\mathsf{v}(0)]^*\right) && \text{by definition of } \mathbf{S}\\
&= \mathbf{E}\left[\mathsf{r}(m)\mathsf{x}^*(0)\right] + \mathbf{E}\left[\mathsf{r}(m)\mathsf{v}^*(0)\right] + \mathbf{E}\left[\mathsf{w}(m)\mathsf{x}^*(0)\right] + \mathbf{E}\left[\mathsf{w}(m)\mathsf{v}^*(0)\right]\\
&= \mathbf{E}\left[\mathsf{r}(m)\mathsf{x}^*(0)\right] + \mathbf{E}\mathsf{r}(m)\mathbf{E}\mathsf{v}^*(0)\\
&\quad + \mathbf{E}\mathsf{w}(m)\mathbf{E}\mathsf{x}^*(0) + \mathbf{E}\mathsf{w}(m)\mathbf{E}\mathsf{v}^*(0) && \text{by \textit{uncorrelated} hypotheses} && \text{(B), (C), and (D)}\\
&= \mathbf{E}\left[\mathsf{r}(m)\mathsf{x}^*(0)\right] + \mathbf{E}\mathsf{r}(m)\cancelto{0}{\mathbf{E}\mathsf{v}^*(0)}\\
&\quad + \cancelto{0}{\mathbf{E}\mathsf{w}(m)}\mathbf{E}\mathsf{x}^*(0) + \cancelto{0}{\mathbf{E}\mathsf{w}(m)}\mathbf{E}\mathsf{v}^*(0) && \text{by \textit{zero-mean} hypothesis} && \text{(E)}\\
&= \mathsf{R}_{rx}(m) && \text{by definition of } \mathsf{R}_{rx} && \text{(Definition 2.4 page 12)}\\
\mathsf{R}_{sx}(m) &\triangleq \mathbf{E}\left[\mathsf{s}(m)\mathsf{x}^*(0)\right] && \text{by definition of } \mathsf{R}_{xy} && \text{(Definition 2.4 page 12)}\\
&\triangleq \mathbf{E}\left([\mathsf{r}(m)+\mathsf{w}(m)]\mathsf{x}^*(0)\right) && \text{by definition of } \mathbf{S}\\
&= \mathbf{E}\left[\mathsf{r}(m)\mathsf{x}^*(0)\right] + \mathbf{E}\mathsf{w}(m)\mathbf{E}\mathsf{x}^*(0) && \text{by \textit{uncorrelated} hypothesis} && \text{(B)}\\
&= \mathbf{E}\left[\mathsf{r}(m)\mathsf{x}^*(0)\right] + \cancelto{0}{\mathbf{E}\mathsf{w}(m)}\mathbf{E}\mathsf{x}^*(0) && \text{by \textit{zero-mean} hypothesis} && \text{(E)}\\
&= \mathsf{R}_{rx}(m) && \text{by definition of } \mathsf{R}_{rx} && \text{(Definition 2.4 page 12)}\\
\mathsf{R}_{ry}(m) &\triangleq \mathbf{E}\left[\mathsf{r}(m)\mathsf{y}^*(0)\right] && \text{by definition of } \mathsf{R}_{xy} && \text{(Definition 2.4 page 12)}\\
&\triangleq \mathbf{E}\left(\mathsf{r}(m)[\mathsf{x}(0)+\mathsf{v}(0)]^*\right) && \text{by definition of } \mathbf{S}\\
&= \mathbf{E}\left[\mathsf{r}(m)\mathsf{x}^*(0)\right] + \mathbf{E}\mathsf{r}(m)\mathbf{E}\mathsf{v}^*(0) && \text{by \textit{uncorrelated} hypothesis} && \text{(C)}\\
&= \mathbf{E}\left[\mathsf{r}(m)\mathsf{x}^*(0)\right] + \mathbf{E}\mathsf{r}(m)\cancelto{0}{\mathbf{E}\mathsf{v}^*(0)} && \text{by \textit{zero-mean} hypothesis} && \text{(E)}\\
&= \mathsf{R}_{rx}(m) && \text{by definition of } \mathsf{R}_{rx} && \text{(Definition 2.4 page 12)}
\end{aligned}$$

✏

Corollary 9.2. *Let* $\mathbf{S}$ *be the* SYSTEM *illustrated in Theorem 9.2 (page 68).*

COR
$$\left\{\begin{array}{l}\text{hypotheses of}\\ \text{Theorem 9.2 (page 68)}\end{array}\right\} \implies \left\{\begin{array}{lccccccc}\text{(1).} & \check{\mathsf{S}}_{sy}(z) &=& \check{\mathsf{S}}_{sx}(z) &=& \check{\mathsf{S}}_{ry}(z) &=& \check{\mathsf{S}}_{rx}(z) \quad \text{and}\\ \text{(2).} & \tilde{\mathsf{S}}_{sy}(\omega) &=& \tilde{\mathsf{S}}_{sx}(\omega) &=& \tilde{\mathsf{S}}_{ry}(\omega) &=& \tilde{\mathsf{S}}_{rx}(\omega)\end{array}\right\}$$

✎PROOF:

$$\begin{aligned}
\check{\mathsf{S}}_{sy}(\omega) &\triangleq \mathbf{Z}\mathsf{R}_{sy}(m) && \text{by definition of } \check{\mathsf{S}}_{xy} && \text{(Definition 2.5 page 15)}\\
&= \mathbf{Z}\mathsf{R}_{rx}(m) && \text{by previous result} && \text{(1)}\\
&= \check{\mathsf{S}}_{rx}(z) && \text{by definition of } \check{\mathsf{S}}_{xy} && \text{(Definition 2.5 page 15)}\\
\tilde{\mathsf{S}}_{sy}(\omega) &\triangleq \check{\mathbf{F}}\mathsf{R}_{sy}(m) && \text{by definition of } \tilde{\mathsf{S}}_{xy} && \text{(Definition 8.3 page 60)}\\
&= \check{\mathbf{F}}\mathsf{R}_{rx}(m) && \text{by previous result} && \text{(1)}\\
&= \tilde{\mathsf{S}}_{rx}(\omega) && \text{by definition of } \tilde{\mathsf{S}}_{xy} && \text{(Definition 8.3 page 60)}
\end{aligned}$$

✏

9.2 Additive noise and operators

Lemma 9.1. *Let* $\mathbf{S}$ *be the* SYSTEM *illustrated in Figure 9.2 (page 71) (A).*

LEM
$$\left\{\begin{array}{lrcll}\text{(A).} & \mathsf{R}_{xx}(n_1,m) &=& \mathsf{R}_{xx}(n_2,m) & \text{(WSS)} \quad \text{and}\\ \text{(B).} & \mathbf{E}[\mathsf{x}(n)\mathsf{v}(n)] &=& \mathbf{E}\mathsf{x}(n)\mathbf{E}\mathsf{v}(n) & \text{(UNCORRELATED)} \quad \text{and}\\ \text{(E).} & \mathbf{E}\mathsf{v}(n) &=& 0 & \text{(ZERO-MEAN)}\end{array}\right\} \implies \left\{\begin{array}{lrcll}\text{(1).} & \mathsf{R}_{xy}(m) &=& \mathsf{R}_{xq}(m) & \text{and}\\ \text{(2).} & \check{\mathsf{S}}_{xy}(z) &=& \check{\mathsf{S}}_{xq}(z) & \text{and}\\ \text{(3).} & \tilde{\mathsf{S}}_{xy}(\omega) &=& \tilde{\mathsf{S}}_{xq}(\omega)\end{array}\right\}$$

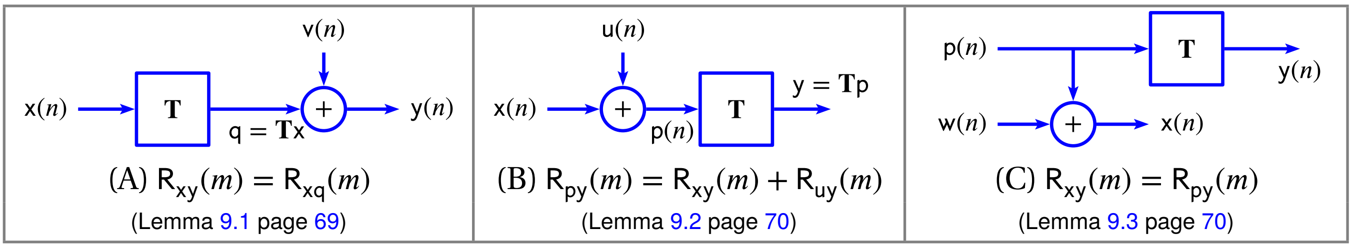


Figure 9.1: Additive noise with *linear*/**non-linear** operator **T**

✎PROOF:

$$
\begin{aligned}
R_{xy}(m) &\triangleq \mathbf{E}\left[x(m)y^*(0)\right] && \text{by definition of } R_{xy} && \text{(Definition 2.4 page 12)}\\
&\triangleq \mathbf{E}\left[x(m)(q(0)+v(0))^*\right] && \text{by definition of } \mathbf{S} && \text{(Figure 9.2 page 71) (A)}\\
&= \mathbf{E}\left[x(m)q^*(0) + p(m)v^*(0)\right] \\
&= \mathbf{E}\left[x(m)q^*(0)\right] + \mathbf{E}\left[x(m)v^*(0)\right] \\
&= \mathbf{E}\left[x(m)q^*(0)\right] + \left[\mathbf{E}x(m)\right]\left[\mathbf{E}v^*(0)\right] && \text{by \textit{uncorrelated} hypothesis} && \text{(B)}\\
&= \mathbf{E}\left[x(m)q^*(0)\right] + \left[\mathbf{E}p(m)\right]\cancelto{0}{\left[\mathbf{E}v^*(0)\right]} && \text{by \textit{zero-mean} hypothesis} && \text{(E)}\\
&= R_{xq}(m) && \text{by definition of } R_{xq} && \text{(Definition 2.4 page 12)}\\
\check{S}_{xy}(z) &\triangleq \mathbf{Z}R_{xy}(m) && \text{by definition of } \check{S}_{xy} && \text{(Definition 2.5 page 15)}\\
&= \mathbf{Z}R_{xq}(m) && \text{by previous result} && (1)\\
&= \check{S}_{xq}(z) && \text{by definition of } \check{S}_{xq} && \text{(Definition 2.5 page 15)}\\
\tilde{S}_{xy}(\omega) &\triangleq \check{\mathbf{F}}R_{xy}(m) && \text{by definition of } \tilde{S}_{xy} && \text{(Definition 8.3 page 60)}\\
&= \check{\mathbf{F}}R_{xq}(m) && \text{by previous result} && (1)\\
&= \tilde{S}_{xq}(\omega) && \text{by definition of } \tilde{S}_{xq} && \text{(Definition 8.3 page 60)}
\end{aligned}
$$

⇰

Lemma 9.2. *Let* **S** *be the* SYSTEM *illustrated in Figure 9.2 (page 71) (B).*

LEM

$$
\left\{\begin{array}{llll}
\text{(A).} & x(n) \text{ is} & \text{(WSS)} & \text{and}\\
\text{(B).} & u(n) \text{ is} & \text{(ZERO-MEAN)} & \text{and}\\
\text{(C).} & x(n) \text{ and } u(n) \text{ are} & \text{(UNCORRELATED)} &
\end{array}\right\}
\implies
\left\{\begin{array}{llcll}
\text{(1).} & R_{pq}(m) & = & R_{xy}(m) + R_{uy}(m) & \text{and}\\
\text{(2).} & \check{S}_{pq}(z) & = & \check{S}_{xy}(z) + \check{S}_{uy}(z) & \text{and}\\
\text{(3).} & \tilde{S}_{pq}(\omega) & = & \tilde{S}_{xy}(\omega) + \tilde{S}_{uy}(\omega) &
\end{array}\right\}
$$

✎PROOF:

$$
\begin{aligned}
R_{xy}(m) &\triangleq \mathbf{E}\left[x(m)y^*(0)\right] && \text{by definition of } R_{xy} && \text{(Definition 2.4 page 12)}\\
&\triangleq \mathbf{E}\left(\left[p(m)-u(m)\right]y^*(0)\right) && \text{by definition of } \mathbf{S}\\
&= \mathbf{E}\left[p(m)y^*(0) - u(m)y^*(0)\right]\\
&= \mathbf{E}\left[p(m)y^*(0)\right] - \mathbf{E}\left[u(m)y^*(0)\right] && \text{because } \mathbf{E} \text{ is a \textit{linear operator}} && \text{(Theorem 1.1 page 4)}\\
&\triangleq R_{py}(m) - R_{uy}(m) && \text{by definition of } R_{xy} && \text{(Definition 2.4 page 12)}
\end{aligned}
$$

⇰

Lemma 9.3. *Let* **S** *be the* SYSTEM *illustrated in Figure 9.2 (page 71) (C).*

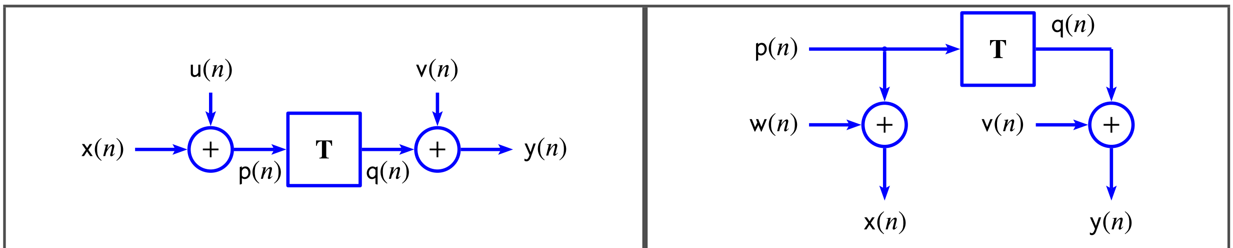
LEM

$$
\left\{\begin{array}{llll}
\text{(A).} & x(n) \text{ is} & \text{(WSS)} & \text{and}\\
\text{(B).} & u(n) \text{ is} & \text{(ZERO-MEAN)} & \text{and}\\
\text{(C).} & x(n) \text{ and } u(n) \text{ are} & \text{(UNCORRELATED)} &
\end{array}\right\}
\implies
\left\{\begin{array}{llcll}
\text{(1).} & R_{xy}(m) & = & R_{py}(m) & \text{and}\\
\text{(2).} & \check{S}_{xy}(z) & = & \check{S}_{py}(z) & \text{and}\\
\text{(3).} & \tilde{S}_{xy}(\omega) & = & \tilde{S}_{py}(\omega) &
\end{array}\right\}
$$

✎PROOF:

$$
\begin{aligned}
\mathsf{R}_{xy}(m) &\triangleq \mathbf{E}\left[\mathsf{x}(m)\mathsf{y}^*(0)\right] && \text{by definition of } \mathsf{R}_{py} && \text{(Definition 2.4 page 12)}\\
&\triangleq \mathbf{E}\left(\left[\mathsf{p}(m)+\mathsf{u}(m)\right]\mathsf{y}^*(0)\right) && \text{by definition of } \mathbf{S}\\
&= \mathbf{E}\left[\mathsf{p}(m)\mathsf{y}^*(0)+\mathsf{u}(m)\mathsf{y}^*(0)\right] && \text{by field properties of } (\mathbb{R},+,\cdot,0,1)\\
&= \mathbf{E}\left[\mathsf{p}(m)\mathsf{y}^*(0)\right]+\mathbf{E}\left[\mathsf{u}(m)\mathsf{y}^*(0)\right] && \text{because } \mathbf{E} \text{ is a } \textit{linear operator} && \text{(Theorem 1.1 page 4)}\\
&= \mathbf{E}\left[\mathsf{p}(m)\mathsf{y}^*(0)\right]+\mathbf{E}[\mathsf{u}(m)]\mathbf{E}\left[\mathsf{y}^*(0)\right] && \text{by } \textit{uncorrelated} \text{ hypothesis} && \text{(C)}\\
&= \mathbf{E}\left[\mathsf{p}(m)\mathsf{y}^*(0)\right]+\mathbf{E}[\mathsf{u}(m)]\cancelto{0}{\mathbf{E}\left[\mathsf{y}^*(0)\right]} && \text{by } \textit{zero-mean} \text{ hypothesis} && \text{(B)}\\
&\triangleq \mathsf{R}_{py}(m) && \text{by definition of } \mathsf{R}_{xy} && \text{(Definition 2.4 page 12)}
\end{aligned}
$$

✏



(A) *Communications additive noise system*
$\tilde{\mathsf{S}}_{pq}(\omega) = \tilde{\mathsf{S}}_{py}(\omega) = \tilde{\mathsf{S}}_{xy}(\omega) + \tilde{\mathsf{S}}_{uy}(\omega)$
(Theorem 9.3 page 71)

(B) *Measurement additive noise system*
$\tilde{\mathsf{S}}_{pq}(\omega) = \tilde{\mathsf{S}}_{py}(\omega) = \tilde{\mathsf{S}}_{xq}(\omega) = \tilde{\mathsf{S}}_{xy}(\omega)$
(Theorem 9.4 page 71)

Figure 9.2: *linear* / **non-linear** additive noise systems

Theorem 9.3 (*communications additive noise cross-correlation*).
Let $\mathbf{S}$ *be the* SYSTEM *illustrated in Figure 9.2 page 71 (A).*

THM

$$
\left.\begin{array}{lll}
\text{(A).} & \mathsf{x}(n) \textit{ is} & \text{WSS} \quad \textit{and}\\
\text{(B).} & \mathsf{u}(n) \textit{ is} & \text{ZERO-MEAN} \quad \textit{and}\\
\text{(C).} & \mathsf{v}(n) \textit{ is} & \text{ZERO-MEAN} \quad \textit{and}\\
\text{(D).} & \mathsf{x}(n), \mathsf{u}(n), \mathsf{v}(n) \textit{ are} & \text{UNCORRELATED}
\end{array}\right\}
\implies
\left\{\begin{array}{lll}
\text{(1).} & \mathsf{R}_{pq}(m)=\mathsf{R}_{py}(m)=\mathsf{R}_{xy}(m)+\mathsf{R}_{uy}(m) & \textit{and}\\
\text{(2).} & \check{\mathsf{S}}_{pq}(z)=\check{\mathsf{S}}_{py}(z)=\check{\mathsf{S}}_{xy}(z)+\check{\mathsf{S}}_{uy}(z) & \textit{and}\\
\text{(3).} & \tilde{\mathsf{S}}_{pq}(\omega)=\tilde{\mathsf{S}}_{py}(\omega)=\tilde{\mathsf{S}}_{xy}(\omega)+\tilde{\mathsf{S}}_{uy}(\omega) &
\end{array}\right.
$$

✎PROOF:

$$
\begin{aligned}
\mathsf{R}_{pq}(m) &= \mathsf{R}_{py}(m) && \text{by Lemma 9.1 page 69}\\
\mathsf{R}_{pq}(m) &= \mathsf{R}_{xq}(m)+\mathsf{R}_{uq}(m) && \text{by Lemma 9.2 page 70}\\
\mathsf{R}_{py}(m) &\triangleq \mathbf{E}\left[\mathsf{p}(m)\mathsf{y}^*(0)\right] && \text{by definition } \mathsf{R}_{py} && \text{(Definition 2.4 page 12)}\\
&\triangleq \mathbf{E}\left([\mathsf{x}(m)+\mathsf{u}(m)]\mathsf{y}^*(0)\right) && \text{by definition } \mathbf{S} && \text{(Figure 9.2 page 71) (A)}\\
&= \mathbf{E}\left[\mathsf{x}(m)\mathsf{y}^*(0)+\mathsf{u}(m)\mathsf{y}^*(0)\right]\\
&= \mathbf{E}\left[\mathsf{x}(m)\mathsf{y}^*(0)\right]+\mathbf{E}\left[\mathsf{u}(m)\mathsf{y}^*(0)\right] && \text{by } \textit{linearity} \text{ of } \mathbf{E} && \text{(Theorem 1.1 page 4)}\\
&= \mathsf{R}_{xy}(m)+\mathsf{R}_{uy}(m) && \text{by definitions } \mathsf{R}_{xy} \text{ and } \mathsf{R}_{uy} && \text{(Definition 2.4 page 12)}
\end{aligned}
$$

✏

Theorem 9.4 (*measurement additive noise cross-correlation*).
Let $\mathbf{S}$ *be the* SYSTEM *illustrated in Figure 9.2 page 71 (B).*

THM

$$
\left.\begin{array}{lll}
\text{(A).} & \mathsf{x}(n) \textit{ is} & \text{WSS} \quad \textit{and}\\
\text{(B).} & \mathsf{u}(n) \textit{ is} & \text{ZERO-MEAN} \quad \textit{and}\\
\text{(C).} & \mathsf{v}(n) \textit{ is} & \text{ZERO-MEAN} \quad \textit{and}\\
\text{(D).} & \mathsf{x}(n), \mathsf{u}(n), \mathsf{v}(n) \textit{ are} & \text{UNCORRELATED}
\end{array}\right\}
\implies
\left\{\begin{array}{lll}
\text{(1).} & \mathsf{R}_{pq}(m)=\mathsf{R}_{py}(m)=\mathsf{R}_{xq}(m)=\mathsf{R}_{xy}(m) & \textit{and}\\
\text{(2).} & \check{\mathsf{S}}_{pq}(z)=\check{\mathsf{S}}_{py}(z)=\check{\mathsf{S}}_{xq}(z)=\check{\mathsf{S}}_{xy}(z) & \textit{and}\\
\text{(3).} & \tilde{\mathsf{S}}_{pq}(\omega)=\tilde{\mathsf{S}}_{py}(\omega)=\tilde{\mathsf{S}}_{xq}(\omega)=\tilde{\mathsf{S}}_{xy}(\omega) &
\end{array}\right.
$$

✎PROOF:

$$
\begin{aligned}
\mathsf{R}_{pq}(m) &= \mathsf{R}_{py}(m) && \text{by Lemma 9.1 page 69} \\
\mathsf{R}_{pq}(m) &= \mathsf{R}_{xq}(m) && \text{by Lemma 9.3 page 70} \\
\mathsf{R}_{xy}(m) &\triangleq \mathbf{E}\left[\mathsf{x}(m)\mathsf{y}^*(0)\right] && \text{by definition } \mathsf{R}_{xy} && \text{(Definition 2.4 page 12)} \\
&\triangleq \mathbf{E}\left(\left[\mathsf{p}(m)+\mathsf{u}(m)\right]\mathsf{y}^*(0)\right) && \text{by definition } \mathbf{S} && \text{(Figure 9.2 page 71) (B)} \\
&= \mathbf{E}\left[\mathsf{p}(m)\mathsf{y}^*(0)+\mathsf{u}(m)\mathsf{y}^*(0)\right] \\
&= \mathbf{E}\left[\mathsf{p}(m)\mathsf{y}^*(0)\right]+\mathbf{E}\left[\mathsf{u}(m)\mathsf{y}^*(0)\right] && \text{by } \textit{linearity} \text{ of } \mathbf{E} && \text{(Theorem 1.1 page 4)} \\
&= \mathbf{E}\left[\mathsf{p}(m)\mathsf{y}^*(0)\right]+\cancelto{0}{\mathbf{E}\left[\mathsf{u}(m)\mathsf{y}^*(0)\right]} && \text{by } \textit{uncorrelated} \text{ hypothesis} && \text{(D)} \\
&= \mathsf{R}_{py}(m) && \text{by definition of } \mathsf{R}_{py} && \text{(Definition 2.4 page 12)}
\end{aligned}
$$

✏

9.3 Additive noise and LTI operators

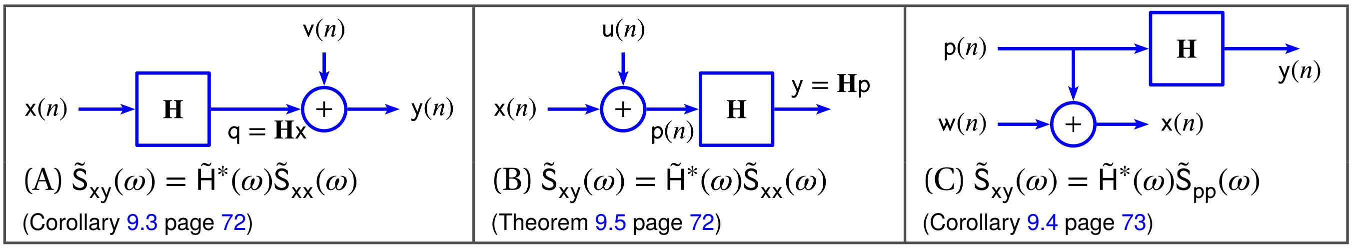


Figure 9.3: Additive noise with **LTI** operator **H**

Corollary 9.3. *Let* **S** *be the* SYSTEM *illustrated in Figure 9.3 (page 72) (A).*

COR

$$
\left\{\begin{array}{llll}
\text{(A).} & \mathsf{x}(n) \textit{ is} & \text{(WSS)} & \textit{and} \\
\text{(B).} & \mathsf{u}(n) \textit{ is} & \text{(ZERO-MEAN)} & \textit{and} \\
\text{(C).} & \mathsf{x}(n) \textit{ and } \mathsf{u}(n) \textit{ are} & \text{(UNCORRELATED)} & \textit{and} \\
\text{(D).} & \mathbf{H} \textit{ is} & \text{(LTI)} &
\end{array}\right\}
\implies
\left\{ \tilde{\mathsf{S}}_{xy}(\omega) = \tilde{\mathsf{H}}^*(\omega)\tilde{\mathsf{S}}_{xx}(\omega) \right\}
$$

✎PROOF:

$$
\begin{aligned}
\tilde{\mathsf{S}}_{xy}(\omega) &= \tilde{\mathsf{S}}_{xq}(\omega) && \text{by Lemma 9.1 page 69} \\
&= \tilde{\mathsf{H}}^*(\omega)\tilde{\mathsf{S}}_{xx}(\omega) && \text{by Corollary 7.3 page 53}
\end{aligned}
$$

✏

Theorem 9.5. *Let* **S** *be the* SYSTEM *illustrated in Figure 9.3 (page 72) (B).*

THM

$$
\left\{\begin{array}{llll}
\text{(A).} & \mathsf{x}(n) \textit{ is} & \text{(WSS)} & \textit{and} \\
\text{(B).} & \mathsf{u}(n) \textit{ is} & \text{(ZERO-MEAN)} & \textit{and} \\
\text{(C).} & \mathsf{x}(n) \textit{ and } \mathsf{u}(n) \textit{ are} & \text{(UNCORRELATED)} & \textit{and} \\
\text{(D).} & \mathbf{H} \textit{ is} & \text{(LTI)} &
\end{array}\right\}
\implies
\left\{\begin{array}{lrcll}
\text{(1).} & \mathsf{R}_{yx}(m) &=& \mathsf{h}(m) \star \mathsf{R}_{xx}(m) & \textit{and} \\
\text{(2).} & \check{\mathsf{S}}_{yx}(z) &=& \check{\mathsf{H}}(z)\check{\mathsf{S}}_{xx}(z) & \textit{and} \\
\text{(3).} & \tilde{\mathsf{S}}_{xy}(\omega) &=& \tilde{\mathsf{H}}^*(\omega)\tilde{\mathsf{S}}_{xx}(\omega) &
\end{array}\right\}
$$

✎PROOF:

1. definition: Let $(\!(\mathsf{h}(n))\!)$ be the *impulse response* of operator **H** such that
$$\mathbf{H}\bar{\delta}(n) \triangleq \sum_{m\in\mathbb{Z}} \mathsf{h}(m)\bar{\delta}(n-m)$$

2. lemma: $\mathbf{H}\boldsymbol{x}(n) = \sum_{m\in\mathbb{Z}} \mathsf{h}(n)\boldsymbol{x}(m-n) = \mathsf{h}(n) \star \mathsf{R}_{\mathsf{xx}}(n)$.
   Proof: by the *linear time-invariant* hypotheses (D) and definition of *convolution* operator $\star$ (Definition D.1 page 199)

3. Proof that $\mathsf{R}_{\mathsf{yx}}(m) = \mathsf{h}(n) \star \mathsf{R}_{\mathsf{xx}}(m)$:

$\mathsf{R}_{\mathsf{yx}}(m) \triangleq \mathbf{E}\left[\mathsf{y}(m)\mathsf{x}^*(0)\right]$	by definition of R_{py}	(Definition 2.4 page 12)
$= \mathbf{E}\left([\mathbf{H}\mathsf{x}(m) + \mathbf{H}\mathsf{u}(m)]\mathsf{x}^*(0)\right)$	by *linear* hypothesis	(D)
$= \mathbf{E}\left(\left[\mathbf{H}\mathsf{x}^*(m)\right]\mathsf{x}^*(0) + [\mathbf{H}\mathsf{u}(0)]\mathsf{x}^*(0)\right)$		
$= \mathbf{E}\left(\left[\mathbf{H}\mathsf{x}^*(m)\right]\mathsf{x}^*(0)\right) + \mathbf{E}\left([\mathbf{H}\mathsf{u}(0)]\mathsf{x}^*(0)\right)$	by *linearity* of $\mathbf{E}$	(Theorem ?? page ??)
$= \mathbf{HE}\left[\mathsf{x}(m)\mathsf{x}^*(0)\right] + \mathbf{HE}\left[\mathsf{u}(m)\mathsf{x}^*(0)\right]$	by *LTI* hypotheses	(D)
$= \mathbf{HE}\left[\mathsf{x}(m)\mathsf{x}^*(0)\right] + \mathbf{HE}\mathsf{u}(m)\mathbf{E}\mathsf{x}^*(0)$	by *uncorrelated* hypothesis	(C)
$= \mathbf{HE}\left[\mathsf{x}(m)\mathsf{x}^*(0)\right] + \mathbf{HE}\overset{0}{\cancel{\mathsf{u}(m)}}\mathbf{E}\mathsf{x}^*(0)$	by *zero-mean* hypothesis	(B)
$= \mathbf{H}\mathsf{R}_{\mathsf{xx}}(m)$	by definition of R_{xx}	(Definition 2.4 page 12)
$= \mathsf{h}(m) \star \mathsf{R}_{\mathsf{xx}}(m)$	by (2) lemma	

✏

When $\mathbf{H}$ is *LTI*, what effect does the additive uncorrelated noise sources have on the cross-statistical properties of x and y? Corollary 9.5 (next) demonstrates that, amazingly, under very general conditions, the noise sources have **no effect**.

Corollary 9.4. *Let* $\mathbf{S}$ *be the* SYSTEM *illustrated in Figure 9.3 (page 72) (C).*

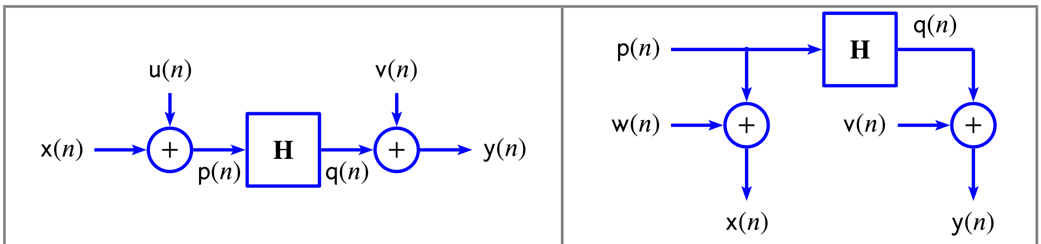
COR

$$\left\{\begin{array}{llll} \text{(A).} & \mathsf{x}(n) \text{ is} & (\text{WSS}) & \text{and}\\ \text{(B).} & \mathsf{u}(n) \text{ is} & (\text{ZERO-MEAN}) & \text{and}\\ \text{(C).} & \mathsf{x}(n) \text{ and } \mathsf{w}(n) \text{ are} & (\text{UNCORRELATED}) & \text{and}\\ \text{(D).} & \mathbf{H} \text{ is} & (\text{LTI}) & \end{array}\right\} \implies \left\{ \tilde{\mathsf{S}}_{\mathsf{xy}}(\omega) = \tilde{\mathsf{H}}^*(\omega)\tilde{\mathsf{S}}_{\mathsf{pp}}(\omega) \right\}$$

✎PROOF:

$\tilde{\mathsf{S}}_{\mathsf{xy}}(\omega) = \tilde{\mathsf{S}}_{\mathsf{py}}(\omega)$	by Lemma 9.3 page 70
$= \tilde{\mathsf{H}}^*(\omega)\tilde{\mathsf{S}}_{\mathsf{pp}}(\omega)$	by Corollary 7.3 page 53

✏



(A) $\tilde{\mathsf{S}}_{\mathsf{xy}}(\omega) = \tilde{\mathsf{H}}^*(\omega)\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)$ (Corollary 9.5 page 73)

(B) $\tilde{\mathsf{S}}_{\mathsf{xy}}(\omega) = \tilde{\mathsf{H}}^*(\omega)\tilde{\mathsf{S}}_{\mathsf{pp}}(\omega)$ (Corollary 9.6 page 74)

Figure 9.4:

Corollary 9.5. *Let* $\mathbf{S}$ *be the* SYSTEM *illustrated in Figure 9.4 page 73 (A).*

COR

$$\left\{\begin{array}{llll} \text{(A).} & \mathsf{x}(n) \text{ is} & (\text{WSS}) & \text{and}\\ \text{(B).} & \mathsf{u}(n) \text{ is} & (\text{ZERO-MEAN}) & \text{and}\\ \text{(C).} & \mathsf{v}(n) \text{ is} & (\text{ZERO-MEAN}) & \text{and}\\ \text{(D).} & \mathsf{x}(n), \mathsf{u}(n), \mathsf{v}(n) \text{ are} & (\text{UNCORRELATED}) & \text{and}\\ \text{(E).} & \mathbf{H} \text{ is} & (\text{LTI}) & \end{array}\right\} \implies \left\{ \tilde{\mathsf{S}}_{\mathsf{xy}}(\omega) = \tilde{\mathsf{H}}^*(\omega)\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega) \right\}$$

✎PROOF:

$$\begin{aligned}
\tilde{S}_{yx}(\omega) &= \tilde{S}_{qx}(\omega) && \text{by Lemma 9.1 page 69}\\
&= \tilde{H}(\omega)\tilde{S}_{xx}(\omega) && \text{by Corollary 7.3 page 53}
\end{aligned}$$

✏

Corollary 9.6. *Let* **S** *be the* SYSTEM *illustrated in Figure 9.4 page 73 (B).*

COR

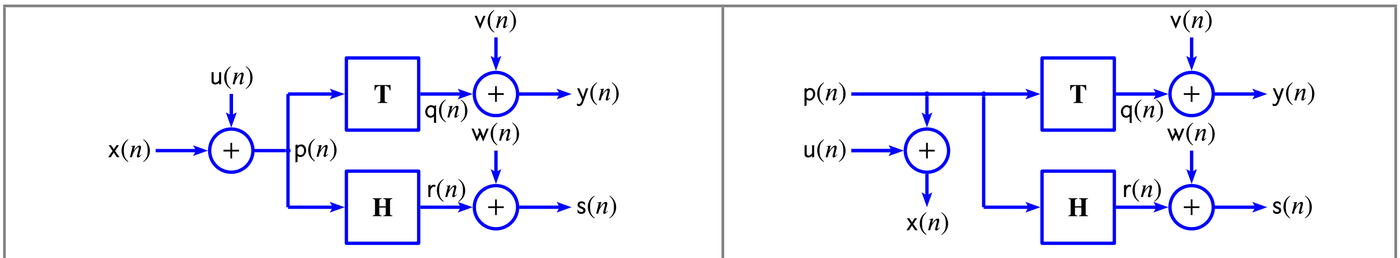
$$\left\{\begin{array}{llll}
\text{(A).} & \mathsf{x}(n) \textit{ is} & \text{WSS} & \textit{and}\\
\text{(B).} & \mathsf{w}(n) \textit{ is} & \text{ZERO-MEAN} & \textit{and}\\
\text{(C).} & \mathsf{v}(n) \textit{ is} & \text{ZERO-MEAN} & \textit{and}\\
\text{(D).} & \mathsf{x}(n), \mathsf{w}(n), \mathsf{v}(n) \textit{ are} & \text{UNCORRELATED} & \textit{and}\\
\text{(E).} & \mathbf{H} \textit{ is} & \text{LTI} &
\end{array}\right\} \implies \left\{ \tilde{S}_{xy}(\omega) = \tilde{H}^*(\omega)\tilde{S}_{pp}(\omega) \right\}$$

✎PROOF:

$$\begin{aligned}
\tilde{S}_{yx}(\omega) &= \tilde{S}_{qx}(\omega) && \text{by Lemma 9.1 page 69}\\
&= \tilde{S}_{qp}(\omega) && \text{by Lemma 9.1 page 69}\\
&= \tilde{H}(\omega)\tilde{S}_{pp}(\omega) && \text{by Corollary 7.3 page 53}
\end{aligned}$$

✏

9.4 Additive noise and dual operators



(A) dual communications additive noise system (Corollary 9.7 page 74)

(B) dual measurement additive noise system (Corollary 9.8 page 75)

Figure 9.5: Dual Additive Noise Systems

Corollary 9.7. *Let* **S** *be the* SYSTEM *illustrated in Figure 9.5 (page 74) (A).*

COR

$$\left\{\begin{array}{llll}
\text{(A).} & \mathbf{H} \textit{ is} & \text{LTI} & \textit{and}\\
\text{(B).} & \mathsf{x}(n) \textit{ is} & \text{WSS} & \textit{and}\\
\text{(C).} & \mathsf{u} \textit{ and } \mathsf{v} \textit{ are} & \text{ZERO-MEAN} & \textit{and}\\
\text{(D).} & \mathsf{x}, \mathsf{u}, \mathsf{v} \textit{ are} & \text{UNCORRELATED} &
\end{array}\right\} \implies \left\{\begin{array}{llll}
\text{(1).} & \check{S}_{sy}(z) & = & \check{H}(z)\left[\check{S}_{xy}(z) + \check{S}_{uy}(z)\right] \quad \textit{and}\\
\text{(2).} & \tilde{S}_{sy}(\omega) & = & \tilde{H}(\omega)\left[\tilde{S}_{xy}(\omega) + \tilde{S}_{uy}(\omega)\right]
\end{array}\right\}$$

✎PROOF:

$$\begin{aligned}
\check{S}_{sy}(z) &= \check{S}_{rq}(z) && \text{by Corollary 9.2 page 69} && \text{and (B), (C) and (D)}\\
&= \check{H}(z)\check{S}_{pq}(z) && \text{by Theorem 7.2 page 53} && \text{and (A)}\\
&= \check{H}(z)\left[\check{S}_{xq}(z) + \check{S}_{uq}(z)\right] && \text{by Lemma 9.2 page 70}\\
&= \check{H}(z)\left[\check{S}_{xy}(z) + \check{S}_{uy}(z)\right] && \text{by Lemma 9.1 page 69}\\
\tilde{S}_{sy}(\omega) &= \check{S}_{sy}(z)\big|_{z=e^{i\omega}}\\
&= \check{H}(z)\left[\check{S}_{xy}(z) + \check{S}_{uy}(z)\right]\big|_{z=e^{i\omega}} && \text{by previous result} && (1)\\
&= \tilde{H}(\omega)\left[\tilde{S}_{xy}(\omega) + \tilde{S}_{uy}(\omega)\right]
\end{aligned}$$

✏

Corollary 9.8. *Let* **S** *be the* SYSTEM *illustrated in Figure 9.5 (page 74) (B).*

COR

$$\left\{\begin{array}{llll} \text{(A).} & \mathbf{H}\ \textit{is} & \text{LTI} & \textit{and} \\ \text{(B).} & \mathsf{x}(n)\ \textit{is} & \text{WSS} & \textit{and} \\ \text{(C).} & \mathsf{u}\ \textit{and}\ \mathsf{v}\ \textit{are} & \text{ZERO-MEAN} & \textit{and} \\ \text{(D).} & \mathsf{p},\mathsf{u},\mathsf{v}\ \textit{are} & \text{UNCORRELATED} & \end{array}\right\} \implies \left\{\begin{array}{lrcll} \text{(1).} & \check{\mathsf{S}}_{\mathsf{sy}}(z) & = & \check{\mathsf{H}}(z)\check{\mathsf{S}}_{\mathsf{xy}}(z) & \textit{and} \\ \text{(2).} & \tilde{\mathsf{S}}_{\mathsf{sy}}(\omega) & = & \tilde{\mathsf{H}}(\omega)\tilde{\mathsf{S}}_{\mathsf{xy}}(\omega) & \end{array}\right\}$$

✎PROOF:

$$\begin{aligned}
\check{\mathsf{S}}_{\mathsf{sy}}(z) &= \check{\mathsf{S}}_{\mathsf{rq}}(z) && \text{by Corollary 9.2 page 69} && \text{and (B), (C) and (D)} \\
&= \check{\mathsf{H}}(z)\check{\mathsf{S}}_{\mathsf{pq}}(z) && \text{by Theorem 7.2 page 53} && \text{and (A)} \\
&= \check{\mathsf{H}}(z)\check{\mathsf{S}}_{\mathsf{xq}}(z) && \text{by Lemma 9.3 page 70} && \\
&= \check{\mathsf{H}}(z)\check{\mathsf{S}}_{\mathsf{xy}}(z) && \text{by Lemma 9.1 page 69} && \\
\tilde{\mathsf{S}}_{\mathsf{sy}}(\omega) &= \check{\mathsf{S}}_{\mathsf{sy}}(z)\big|_{z=e^{i\omega}} && \text{by definition of } \mathbf{Z} && \text{(Definition V.1 page 429)} \\
&= \check{\mathsf{H}}(z)\check{\mathsf{S}}_{\mathsf{xy}}(z)\big|_{z=e^{i\omega}} && \text{by previous result} && \text{(1)} \\
&= \tilde{\mathsf{H}}(\omega)\tilde{\mathsf{S}}_{\mathsf{xy}}(\omega)
\end{aligned}$$

✏

Part III

Statistical Estimation

CHAPTER 10

ESTIMATION OVERVIEW

10.1 Model-based estimation

Starting in the 1980s and continuing into 2019, James V. Candy has proposed the **model-based approach**[1] to estimation. This approach partitions the task of estimation into three parts:

1. The **model**:[2] The presumed system architecture, "the form of which is usually suggested by prior knowledge, physical understanding or guesswork."[3] A very common example is the *additive noise model*.
2. The **algorithm**: The operation used to calculate the estimate. An *algorithm* may also be called a **processor**, **filter**, or **method**.
3. The **criterion function**: A *function* which measures the performance of the algorithm. This function may also be called a **cost function**.

10.1.1 Estimation models

Estimation modeling includes several sub-models:

1. The signal model (parametric, nonparametric)—more on this below.
2. The noise model—this can be assumed (Guassian is a common choice) or estimated (Chapter 17 page 139).
3. The model architecture (Figure 10.1 page 80)—additive noise model is a common choice, with justification from *Wold's Theorem*.

---

[1] Candy (1985) ⟨ISBN:9780070097254⟩, Candy (1988) ⟨ISBN:9780070097513⟩, Candy and Sullivan (1992), Candy (2005) page 5, Candy (2019) page 7. "Dr. Candy received the IEEE Distinguished Technical Achievement Award for the *development of model-based signal processing in ocean acoustics*": Rossing (Springer) page 1234

[2] Box and Jenkins (1976) pages 173–207 ⟨Chapter 6 "Model Identification"⟩

[3] Scargle (1979) page 5

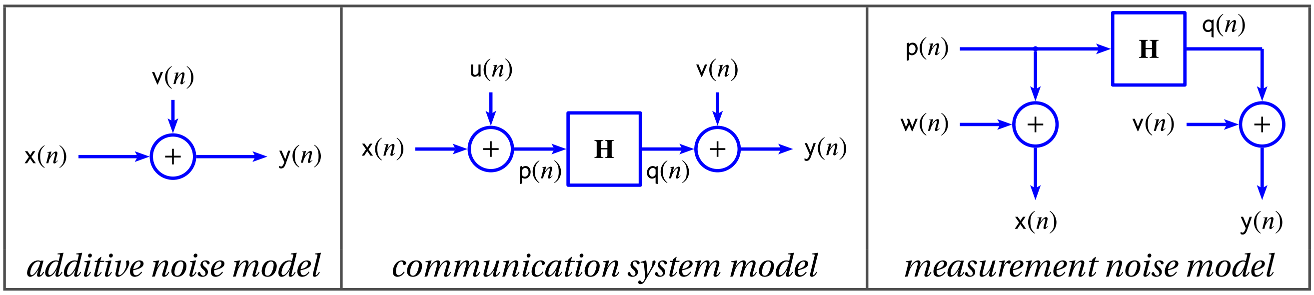


Figure 10.1: Some estimation models

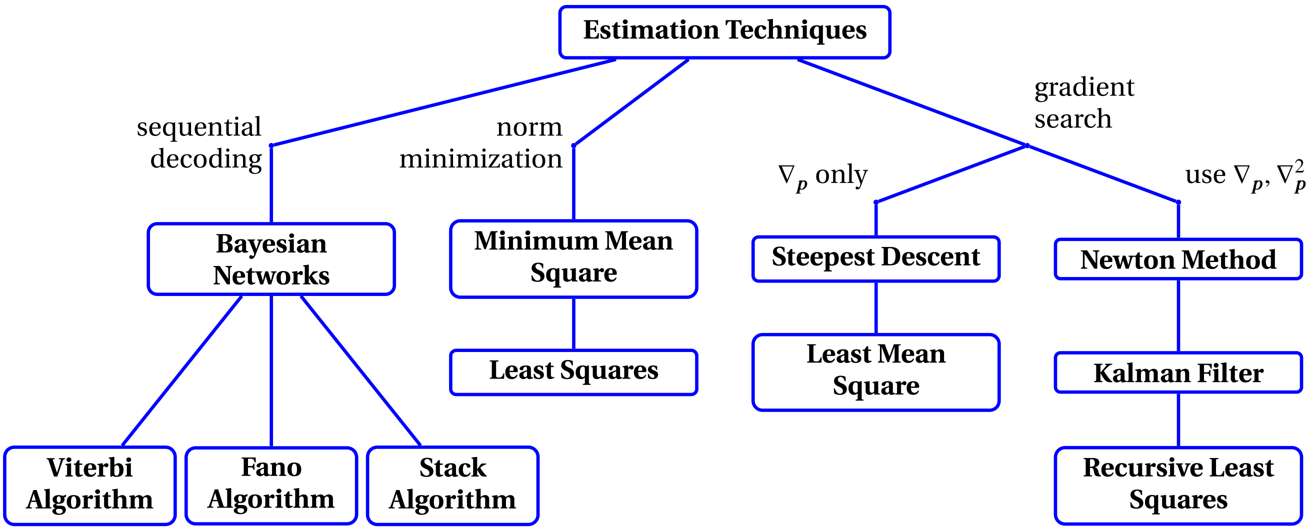


Figure 10.2: Estimation Algorithms

As for the the signal model, let $\mathsf{x}(t;\theta)$ be a signal with parameter θ. There are three basic types of signal models, leading to three fundamental types of estimation:

1. *detection*:
   - The waveform $\mathsf{x}(t;\theta_n)$ is known except for the value of parameter θ_n.
   - The parameter θ_n is one of a finite set of values.
   - Estimate θ_n and thereby also estimate $\mathsf{x}(t;\theta)$.
2. *parametric* estimation:
   - The waveform $\mathsf{x}(t;\theta)$ is known except for the value of parameter θ.
   - The parameter θ is one of an infinite set of values.
   - Estimate θ and thereby also estimate $\mathsf{x}(t;\theta)$.
3. *nonparametric* estimation:
   - The waveform $\mathsf{x}(t)$ is unknown and assumed without any parameter θ.
   - Estimate $\mathsf{x}(t)$.

10.1.2 Estimation algorithms

Estimation algorithms include the following (Figure 10.2 page 80):[4]

[4] Nelles (2001) page 26 ⟨“Fig 2.2 Overview of linear and nonlinear optimization techniques”⟩, Nelles (2001) page 33 ⟨“Fig 2.5 The Bayes method is the most general approach but...”⟩, Nelles (2001) page 63 ⟨“Table 3.3 Relationship between linear recursive and nonlinear optimization techniques”⟩, Nelles (2001) page 66, Clarkson

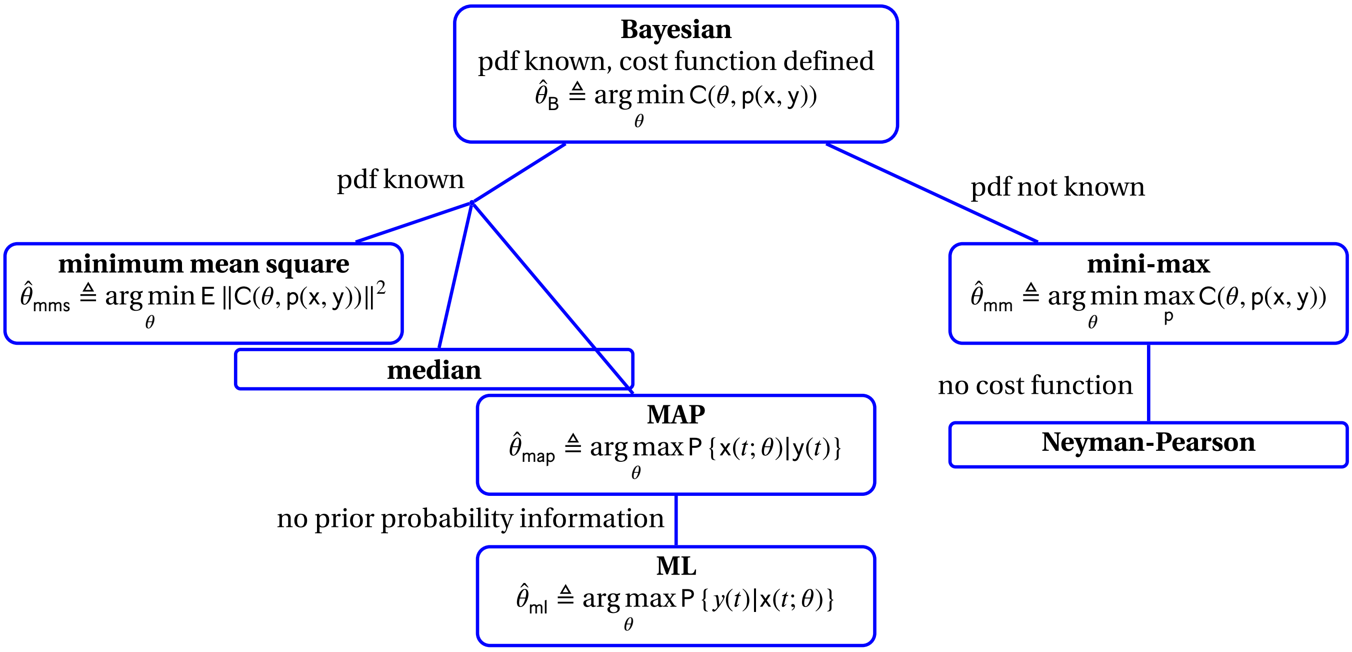


Figure 10.3: Estimation criterion

1. Bayesian Networks (CHAPTER 12 page 97)
2. Gradient Search
3. Bayesian signal processing
4. Neural Networks
5. Direct search

Bayesian signal processing involves estimating the joint-pdf of a process (CHAPTER 17 page 139), and then integrating over a portion of the this (possibly high dimensional) pdf to find an MAP estimate. The integration itself may be estimated using *Monte Carlo Integration*.[5]

10.1.3 Estimation criterion function

Optimization requires a criterion against which the quality of an estimate is measured.[6] The most demanding and general criterion is the *Bayesian* criterion. The Bayesian criterion requires knowledge of the probability distribution functions and the definition of a *cost function*. Other criterion are special cases of the Bayesian criterion such that the cost function is defined in a special way, no cost function is defined, and/or the distribution is not known (Figure 10.2 page 80).

Definition 10.1. *Let*

(A).	$\mathsf{x}(t;\theta)$	*be a random process with unknown parameter* θ
(B).	$\mathsf{y}(t)$	*an observed random process which is statistically dependent on* $\mathsf{x}(t;\theta)$
(C).	$\mathsf{C}(\theta, \mathsf{p}(x, y))$	*be a cost function.*

Then the following ***estimates*** *are defined as follows:*

(1993) page 276 ⟨"Figure 6.1.1 Options for the adaptive algorithm design."⟩, Vaseghi (2000) pages 4–5 ⟨§"1.2 Signal Processing Methods"⟩

[5] Candy (2009) pages 4–7 ⟨§"1.3 SIMULATION-BASED APPROACH TO BAYESIAN PROCESSING"⟩, Liu (2013) page 1 ⟨§1.1 The Need of Monte Carlo Techniques⟩

[6] Srinath et al. (1996) ⟨013125295X⟩.

DEF

(1). ***Bayesian estimate*** $\hat{\theta}_{\mathsf{B}} \triangleq \arg\min_{\theta} C(\theta, \mathsf{p}(x,y))$

(2). ***Mean square estimate*** ***("MS estimate")*** $\hat{\theta}_{\mathsf{mms}} \triangleq \arg\min_{\theta} \mathbf{E}\,\|C(\theta, \mathsf{p}(x,y))\|^2$

(3). ***mini-max estimate*** ***("MM estimate")*** $\hat{\theta}_{\mathsf{mm}} \triangleq \arg\min_{\theta}\max_{\mathsf{p}} C(\theta, \mathsf{p}(x,y))$

(4). ***maximum a-posteriori probability estimate*** ***("MAP estimate")*** $\hat{\theta}_{\mathsf{map}} \triangleq \arg\max_{\theta} \mathsf{P}\{\mathsf{x}(t;\theta)|\mathsf{y}(t)\}$

(5). ***maximum likelihood estimate*** ***("ML estimate")*** $\hat{\theta}_{\mathsf{ml}} \triangleq \arg\max_{\theta} \mathsf{P}\{\mathsf{y}(t)|\mathsf{x}(t;\theta)\}$

Theorem 10.1. *Let* $\mathsf{x}(t;\theta)$ *be a random process with unknown parameter* θ.

THM

$$\{\mathsf{P}\{\theta\} = \text{CONSTANT}\} \implies \left\{\hat{\theta}_{\mathsf{map}} = \hat{\theta}_{\mathsf{ml}}\right\}$$

✎PROOF:

$$\begin{aligned}
\hat{\theta}_{\mathsf{map}} &\triangleq \arg\max_{\theta} \mathsf{P}\{\mathsf{x}(t;\theta)|\mathsf{y}(t)\} && \text{by definition of } \hat{\theta}_{\mathsf{map}} && \text{(Definition 10.1 page 81)}\\
&\triangleq \arg\max_{\theta} \frac{\mathsf{P}\{\mathsf{x}(t;\theta)\wedge \mathsf{y}(t)\}}{\mathsf{P}\{r(t)\}} && \text{by definition of \textit{conditional probability}} && \text{(Definition A.4 page 174)}\\
&\triangleq \arg\max_{\theta} \frac{\mathsf{P}\{r(t)|\mathsf{x}(t;\theta)\}\mathsf{P}\{\mathsf{x}(t;\theta)\}}{\mathsf{P}\{\mathsf{y}(t)\}} && \text{by definition of \textit{conditional probability}} && \text{(Definition A.4 page 174)}\\
&= \arg\max_{\theta} \mathsf{P}\{\mathsf{y}(t)|\mathsf{x}(t;\theta)\}\mathsf{P}\{\mathsf{x}(t;\theta)\} && \text{because } \mathsf{y}(t) \text{ is independent of } \theta\\
&= \arg\max_{\theta} \mathsf{P}\{\mathsf{y}(t)|\mathsf{x}(t;\theta)\}\\
&\triangleq \hat{\theta}_{\mathsf{ml}} && \text{by definition of } \hat{\theta}_{\mathsf{ml}} && \text{(Definition 10.1 page 81)}
\end{aligned}$$

✏

10.2 Estimation applications

Applications of estimation include the following:

1. **prediction**—for speech processing, financial markets, etc.
2. **deconvolution**—estimate a system transfer function $\hat{\mathsf{H}}(z)$ and calculate it's inverse $\hat{\mathsf{H}}^{-1}(z)$.
3. **de-noising**—estimate a signal $\mathsf{x}(t)$ that is buried in noise $\mathsf{v}(t)$.

10.3 Measures of estimator quality

Definition 10.2. [7]

DEF

*The **mean square error*** $\mathsf{mse}(\hat{\theta})$ *of an estimate* $\hat{\theta}$ *of a parameter* θ *is defined as*

$$\mathsf{mse}(\hat{\theta}) \triangleq \mathbf{E}\left[\left(\hat{\theta}-\theta\right)^2\right]$$

[7] Silverman (1986) page 35 ⟨§"1.3.2 Measures of discrepancy…"⟩, Clarkson (1993) page 50 ⟨"c) Mean-Squared Error"⟩, Bendat and Piersol (2010) ⟨§"1.4.3 Error Analysis Criteria"⟩, Bendat and Piersol (1966) page 183§"5.3 Statistical Errors for Parameter Estimates"

Definition 10.3. [8]

DEF: *The **normalized rms error** $\epsilon(\hat{\theta})$ of an estimate $\hat{\theta}$ of a parameter θ is defined as*

$$\epsilon(\hat{\theta}) \triangleq \frac{\sqrt{\mathsf{mse}(\hat{\theta})}}{\theta} \triangleq \frac{\sqrt{\mathbf{E}\left[(\hat{\theta}-\theta)^2\right]}}{\theta}$$

Definition 10.4. [9]

DEF: *The **mean integrated square error** $\mathsf{mse}(\hat{\theta})$ of an estimate $\hat{\theta}$ of a parameter θ is defined as*

$$\mathsf{mse}(\hat{\theta}) \triangleq \mathbf{E}\int_{\theta\in\mathbb{R}}\left[(\hat{\theta}-\theta)^2\right]$$

The *mean square error* of $\hat{\theta}$ can be expressed as the sum of two components: the variance of $\hat{\theta}$ and the bias of $\hat{\theta}$ squared (next Theorem). For an example of Theorem 10.2 in action, see the proof for the $\mathsf{mse}(\hat{\mu})$ of the *arithmetic mean estimate* as provided in Theorem 16.1 (page 131).

Theorem 10.2. [10] *Let* $\mathsf{mse}(\hat{\theta})$ *be the* MEAN SQUARE ERROR *(Definition 10.2 page 82) and* $\epsilon(\hat{\theta})$ *the* NORMALIZED RMS ERROR *(Definition 10.3 page 83) of an estimator* $\hat{\theta}$.

THM:

$$\mathsf{mse}(\hat{\theta}) = \underbrace{\mathbf{E}\left[(\hat{\theta}-\mathbf{E}\hat{\theta})^2\right]}_{\text{variance of }\hat{\theta}} + \underbrace{\left[\mathbf{E}\hat{\theta}-\theta\right]^2}_{\text{bias of }\hat{\theta}\text{ squared}} \qquad \epsilon(\hat{\theta}) = \frac{\sqrt{\mathbf{E}\left[(\hat{\theta}-\mathbf{E}\hat{\theta})^2\right]+\left[\mathbf{E}\hat{\theta}-\theta\right]^2}}{\theta}$$

✎PROOF:

$$\begin{aligned}
\mathsf{mse}(\hat{\theta}) &\triangleq \mathbf{E}\left[(\hat{\theta}-\theta)^2\right] && \text{by definition of mse (Definition 10.2 page 82)}\\
&= \mathbf{E}\left[\left(\underbrace{\hat{\theta}-\mathbf{E}\hat{\theta}+\mathbf{E}\hat{\theta}}_{0}-\theta\right)^2\right] && \text{by \textit{additive identity} property of } (\mathbb{C},+,\cdot,0,1)\\
&= \mathbf{E}\left[(\hat{\theta}-\mathbf{E}\hat{\theta})^2+\underbrace{(\mathbf{E}\hat{\theta}-\theta)^2}_{\text{constant}}-2(\hat{\theta}-\mathbf{E}\hat{\theta})(\mathbf{E}\hat{\theta}-\theta)\right] && \text{by \textit{Binomial Theorem}}\\
&= \mathbf{E}(\hat{\theta}-\mathbf{E}\hat{\theta})^2+(\mathbf{E}\hat{\theta}-\theta)^2-2\mathbf{E}\left[\hat{\theta}\mathbf{E}\hat{\theta}-\hat{\theta}\theta-\mathbf{E}\hat{\theta}\hat{\theta}+\mathbf{E}\hat{\theta}\theta\right] && \text{by \textit{linearity} of } \mathbf{E} \text{ (Theorem 1.1 page 4)}\\
&= \mathbf{E}(\hat{\theta}-\mathbf{E}\hat{\theta})^2+(\mathbf{E}\hat{\theta}-\theta)^2-2\underbrace{\left[\mathbf{E}\hat{\theta}\mathbf{E}\hat{\theta}-\mathbf{E}\hat{\theta}\mathbf{E}\theta-\mathbf{E}\hat{\theta}\mathbf{E}\hat{\theta}+\mathbf{E}\hat{\theta}\mathbf{E}\theta\right]}_{0} && \text{by \textit{linearity} of } \mathbf{E} \text{ (Theorem 1.1 page 4)}\\
&= \mathbf{E}(\hat{\theta}-\mathbf{E}\hat{\theta})^2+(\mathbf{E}\hat{\theta}-\theta)^2
\end{aligned}$$

✎

Definition 10.5. [11]

DEF: *An estimator $\hat{\theta}$ of a parameter θ is a **minimum variance unbiased estimator (MVUE)** if*

1. $\mathbf{E}\hat{\theta} = 0$ (UNBIASED) *and*
2. *no other unbiased estimator $\hat{\phi}$ has smaller variance* $\mathrm{var}(\hat{\phi})$

[8] Bendat and Piersol (2010) ⟨§"1.4.3 Error Analysis Criteria"⟩

[9] Silverman (1986) page 35 ⟨§"1.3.2 Measures of discrepancy…"⟩, Rosenblatt (1956) page 835 ⟨"integrated mean square error"⟩

[10] Choi (1978) page 76, Kay (1988) page 45 ⟨§"3.3 Estimation Theory"⟩, Stuart and Ord (1991) page 629 ⟨"Minium mean-square-error estimation"⟩, Clarkson (1993) page 51 ⟨§"2.6 Estimation of Moments"⟩, Bendat and Piersol (2010) ⟨§"1.4.3 Error Analysis Criteria"⟩, Bendat and Piersol (1966) page 183§"5.3 Statistical Errors for Parameter Estimates", Bendat and Piersol (1980) page 39 ⟨§"2.4.1 Bias versus Random Errors"⟩

[11] Choi (1978) page 76, Shao (2003) page 161 ⟨§"The UMVUE"⟩, Bolstad (2007) page 164 ⟨§"Minimum Variance Unbiased Estimator"⟩

CHAPTER 11

LEAST SQUARES ALGORITHMS WHEN MODEL IS UNKNOWN

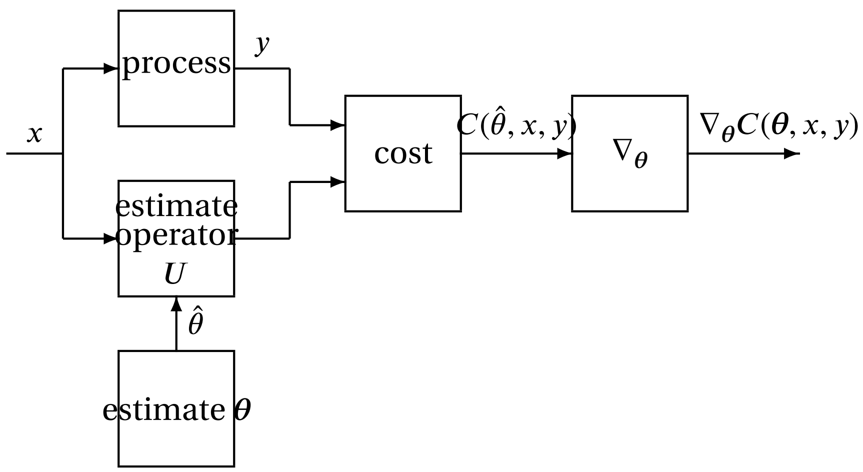


Figure 11.1: Estimation using gradient of cost function

Norm minimization techniques are useful when the system model is unknown.[1] In this section we present two types of norm minimization:[2]

1. minimum mean square estimation (MMSE):
   The MMS estimate is a *stochastic* estimate. To compute the MMS estimate, the we do not need to know the actual data values, but we must know certain system statistics which are the input data autocorrelation and input/output crosscorrelation. The cost function is the expected value of the norm squared error.
2. least square estimation (LSE):[3]
   The LS estimate is a *deterministic* estimate. To compute the LS estimate, we must know the actual data values (although these may be "noisy" measurements). The cost function is the norm squared error.

[1] Clarkson (1993) page 1 ⟨chapter one Introduction⟩

[2] The Least Squares algorithm is nothing new to mathematics. It was first published by Legendre in 1805, but there is a credible claim by Gauss that he had it as far back as 1795. Gauss, by the way, was also the first to discover the FFT. References: Sorenson (1970) page 63, Plackett (1972), Stigler (1981), Dutka (1995)

[3] Scargle (1979) page 5

Solutions to both are given in terms of two matrices:

Y: Autocorrelation matrix
W: Crosscorrelation matrix.

11.1 Minimum mean square estimation

Definition 11.1. *Let the following vectors, matrices, and functions be defined as follows:*

$\boldsymbol{x}$	$\in$	$\mathbb{C}^m$	DATA VECTOR	U	$\in$	C_{mn}	REGRESSION MATRIX
$\boldsymbol{y}$	$\in$	$\mathbb{C}^n$	PROCESSED DATA VECTOR	R	$\in$	C_{mm}	AUTOCORRELATION MATRIX
$\hat{\boldsymbol{y}}$	$\in$	$\mathbb{C}^n$	PROCESSED DATA ESTIMATE VECTOR	W	$\in$	$\mathbb{C}^m$	CROSS-CORRELATION VECTOR
$\boldsymbol{e}$	$\in$	$\mathbb{C}^n$	ERROR VECTOR	C	:	$\mathbb{R}^m \to \mathbb{R}^{\vdash}$	COST FUNCTION
$\boldsymbol{\theta}$	$\in$	$\mathbb{R}^m$	PARAMETER VECTOR				

Theorem 11.1 (Minimum mean square estimation)**.**

Let

$\hat{\boldsymbol{y}}(\boldsymbol{\theta})$	$\triangleq$	$\mathbf{U}^H\boldsymbol{\theta}$	$\hat{\theta}_{\mathsf{mms}}$	$\triangleq$	$\arg\min_{\theta} \mathsf{C}(\boldsymbol{\theta})$
$\boldsymbol{e}(\boldsymbol{\theta})$	$\triangleq$	$\hat{\boldsymbol{y}} - \boldsymbol{y}$	$\mathbf{R}$	$\triangleq$	$\mathbf{E}\left[\mathbf{U}\mathbf{U}^H\right]$
$\mathsf{C}(\boldsymbol{\theta})$	$\triangleq$	$\mathbf{E}\,\lVert\boldsymbol{e}\rVert^2 \triangleq \mathbf{E}\left[\boldsymbol{e}^H\boldsymbol{e}\right]$	$\mathbf{W}$	$\triangleq$	$\mathbf{E}\left[\mathbf{U}\boldsymbol{y}\right]$

, then

$$\begin{aligned}
\hat{\theta}_{\mathsf{mms}} &= (\mathbf{R}_{\mathsf{e}}Y)^{-1}(\mathbf{R}_{\mathsf{e}}W)\\
\mathsf{C}(\boldsymbol{\theta}) &= \boldsymbol{\theta}^H\mathbf{R}\boldsymbol{\theta} - 2\mathbf{R}_{\mathsf{e}}\mathbf{W}^H\boldsymbol{\theta} + \mathbf{E}\left[\boldsymbol{y}^H\boldsymbol{y}\right]\\
\nabla_{\theta}\mathsf{C}(\boldsymbol{\theta}) &= 2\mathbf{R}_{\mathsf{e}}[Y]\boldsymbol{\theta} - 2\mathbf{R}_{\mathsf{e}}W\\
\mathsf{C}(\hat{\theta}_{\mathsf{mms}}) &= \begin{cases} \mathbf{E}\boldsymbol{y}^H\boldsymbol{y} + (\mathbf{R}_{\mathsf{e}}W^H)(\mathbf{R}_{\mathsf{e}}Y)^{-1}R(\mathbf{R}_{\mathsf{e}}Y)^{-1}(\mathbf{R}_{\mathsf{e}}W) - 2(\mathbf{R}_{\mathsf{e}}W^H)(\mathbf{R}_{\mathsf{e}}Y)^{-1}(\mathbf{R}_{\mathsf{e}}W) \\ \mathbf{E}\boldsymbol{y}^H\boldsymbol{y} - (\mathbf{R}_{\mathsf{e}}W^H)R^{-1}(\mathbf{R}_{\mathsf{e}}W) \quad \text{if } \mathbf{R} \text{ is } \textbf{\textit{real-valued}} \end{cases}
\end{aligned}$$

✎PROOF: See APPENDIX F (page 209) for a Matrix Calculus reference.

1. Proof that cost $\mathsf{C}(\boldsymbol{\theta}) = \boldsymbol{\theta}^H\mathbf{R}\boldsymbol{\theta} - 2\mathbf{R}_{\mathsf{e}}\mathbf{W}^H\boldsymbol{\theta} + \mathbf{E}\left[\boldsymbol{y}^H\boldsymbol{y}\right]$:

$$\begin{aligned}
\mathsf{C}(\boldsymbol{\theta}) &\triangleq \mathbf{E}\,\lVert\boldsymbol{e}\rVert^2 && \text{by definition of } \textit{cost function } \mathsf{C}\\
&\triangleq \mathbf{E}\,\lVert\hat{\boldsymbol{y}} - \boldsymbol{y}\rVert^2 && \text{by definition of } \textit{error vector } \boldsymbol{e}\\
&= \mathbf{E}\left[\lVert\hat{\boldsymbol{y}}\rVert^2 - 2\mathbf{R}_{\mathsf{e}}\langle\hat{\boldsymbol{y}}\,|\,\boldsymbol{y}\rangle + \lVert\boldsymbol{y}\rVert^2\right] && \text{by } \textit{Polar Identity} \text{ (Lemma N.1 page 314)}\\
&= \mathbf{E}\left[\langle\hat{\boldsymbol{y}}\,|\,\hat{\boldsymbol{y}}\rangle - 2\mathbf{R}_{\mathsf{e}}\langle\hat{\boldsymbol{y}}\,|\,\boldsymbol{y}\rangle + \langle\boldsymbol{y}\,|\,\boldsymbol{y}\rangle\right] && \text{by definition of } \lVert\cdot\rVert\\
&= \mathbf{E}\left[\left\langle\mathbf{U}^H\boldsymbol{\theta}\,|\,\mathbf{U}^H\boldsymbol{\theta}\right\rangle - 2\mathbf{R}_{\mathsf{e}}\left\langle\mathbf{U}^H\boldsymbol{\theta}\,|\,\boldsymbol{y}\right\rangle + \langle\boldsymbol{y}\,|\,\boldsymbol{y}\rangle\right] && \text{by definition of } \hat{\boldsymbol{y}}\\
&= \mathbf{E}\left[\left(\mathbf{U}^H\boldsymbol{\theta}\right)^H\left(\mathbf{U}^H\boldsymbol{\theta}\right) - 2\mathbf{R}_{\mathsf{e}}\left[\boldsymbol{y}^H\mathbf{U}^H\boldsymbol{\theta}\right] + \boldsymbol{y}^H\boldsymbol{y}\right] && \text{by definition of } \langle\triangle\,|\,\nabla\rangle \text{ in matrix algebra}\\
&= \mathbf{E}\left[\boldsymbol{\theta}^H\mathbf{U}\mathbf{U}^H\boldsymbol{\theta}\right] - 2\mathbf{R}_{\mathsf{e}}\mathbf{E}\left[\boldsymbol{y}^H\mathbf{U}^H\boldsymbol{\theta}\right] + \mathbf{E}\left[\boldsymbol{y}^H\boldsymbol{y}\right] && \text{by } \textit{linearity} \text{ of } \mathbf{E} \text{ (Theorem 1.1 page 4)}\\
&= \boldsymbol{\theta}^H\mathbf{E}\left[\mathbf{U}\mathbf{U}^H\right]\boldsymbol{\theta} - 2\mathbf{R}_{\mathsf{e}}\mathbf{E}\left[\mathbf{U}\boldsymbol{y}\right]^H\boldsymbol{\theta} + \mathbf{E}\left[\boldsymbol{y}^H\boldsymbol{y}\right]\\
&= \boldsymbol{\theta}^H\mathbf{R}\boldsymbol{\theta} - 2\mathbf{R}_{\mathsf{e}}\mathbf{W}^H\boldsymbol{\theta} + \mathbf{E}\left[\boldsymbol{y}^H\boldsymbol{y}\right] && \text{by definitions of } \mathbf{R} \text{ and } \mathbf{W}
\end{aligned}$$

2. Proof that optimal $\boldsymbol{\theta}_{\mathsf{opt}} = (\mathbf{R}_{\mathsf{e}}Y)^{-1}(\mathbf{R}_{\mathsf{e}}W)$:

$$\begin{aligned}
\nabla_{\theta}\mathsf{C}(\boldsymbol{\theta}) &= \nabla_{\theta}\left[\boldsymbol{\theta}^H\mathbf{R}\boldsymbol{\theta} - 2\mathbf{R}_{\mathsf{e}}\mathbf{W}^H\boldsymbol{\theta} + \mathbf{E}\left[\boldsymbol{y}^H\boldsymbol{y}\right]\right] && \text{by item (1)}\\
&= \nabla_{\theta}\left[\boldsymbol{\theta}^H\mathbf{R}\boldsymbol{\theta} - (\mathbf{W}^H)^*\boldsymbol{\theta} - \mathbf{W}^H\boldsymbol{\theta} + \mathbf{E}\boldsymbol{y}^H\boldsymbol{y}\right] && \text{by definition of } \mathbf{R}_{\mathsf{e}} \text{ (Definition M.5 page 305)}\\
&= \mathbf{R}\boldsymbol{\theta} + \mathbf{R}^T\boldsymbol{\theta} - \nabla_{\theta}\left[(\mathbf{W}^H)^*\boldsymbol{\theta} + \mathbf{W}^H\boldsymbol{\theta}\right] + 0 && \text{by } \textit{quadratic form} \text{ result (Theorem F.6 page 214)}\\
&= \mathbf{R}\boldsymbol{\theta} + \mathbf{R}^T\boldsymbol{\theta} - [(\mathbf{W}^H)^*]^T - [\mathbf{W}^H]^T && \text{by } \textit{affine equations} \text{ result (Theorem F.3 page 213)}\\
&= \mathbf{R}\boldsymbol{\theta} + (\mathbf{R}^H)^*\boldsymbol{\theta} - \mathbf{W} - \mathbf{W}^* && \text{by definition of } \textit{Hermitian Transpose}\ ^H
\end{aligned}$$

$$
\begin{aligned}
&= \mathbf{R}\boldsymbol{\theta} + \mathbf{R}^*\boldsymbol{\theta} - \mathbf{W} - \mathbf{W}^* && \text{because } \mathbf{R} \text{ is } \textit{Hermitian symmetric}\\
&= (\mathbf{R} + \mathbf{R}^*)\boldsymbol{\theta} - (\mathbf{W} + \mathbf{W}^*) && \text{by } \textit{ring} \text{ property}\\
&= 2(\mathbf{R}_e Y)\boldsymbol{\theta} - 2\mathbf{R}_e\mathbf{W} && \text{by definition of } \mathbf{R}_e \text{ (Definition M.5 page 305)}\\
\implies \boldsymbol{\theta}_{\text{opt}} &= (\mathbf{R}_e Y)^{-1}(\mathbf{R}_e\mathbf{W}) && \text{by setting } \nabla_{\boldsymbol{\theta}}\mathsf{C}(\boldsymbol{\theta}) = 0
\end{aligned}
$$

3. Cost of optimal $\boldsymbol{\theta}_{\text{opt}}$:

$$
\begin{aligned}
\mathsf{C}(\boldsymbol{\theta}_{\text{opt}}) &= \boldsymbol{\theta}^H_{\text{opt}}\mathbf{R}\boldsymbol{\theta}_{\text{opt}} - 2\mathbf{R}_e\left[\mathbf{W}^H\right]\boldsymbol{\theta}_{\text{opt}} + \mathbf{E}\boldsymbol{y}^H\boldsymbol{y} && \text{by item (1)}\\
&= \left[(\mathbf{R}_e Y)^{-1}(\mathbf{R}_e\mathbf{W})\right]^H\mathbf{R}\left[(\mathbf{R}_e Y)^{-1}(\mathbf{R}_e\mathbf{W})\right] - 2\mathbf{R}_e\left[\mathbf{W}^H\right]\left[(\mathbf{R}_e Y)^{-1}(\mathbf{R}_e\mathbf{W})\right] + \mathbf{E}\boldsymbol{y}^H\boldsymbol{y} && \text{by item (2)}\\
&= (\mathbf{R}_e\mathbf{W}^H)(\mathbf{R}_e Y)^{-H}\mathbf{R}(\mathbf{R}_e Y)^{-1}(\mathbf{R}_e\mathbf{W}) - 2\mathbf{R}_e\left[\mathbf{W}^H\right](\mathbf{R}_e Y)^{-1}(\mathbf{R}_e\mathbf{W}) + \mathbf{E}\boldsymbol{y}^H\boldsymbol{y}\\
&= (\mathbf{R}_e\mathbf{W}^H)(\mathbf{R}_e\mathbf{R}^H)^{-1}\mathbf{R}(\mathbf{R}_e Y)^{-1}(\mathbf{R}_e\mathbf{W}) - 2\mathbf{R}_e\left[\mathbf{W}^H\right](\mathbf{R}_e Y)^{-1}(\mathbf{R}_e\mathbf{W}) + \mathbf{E}\boldsymbol{y}^H\boldsymbol{y}\\
&= (\mathbf{R}_e\mathbf{W}^H)(\mathbf{R}_e Y)^{-1}\mathbf{R}(\mathbf{R}_e Y)^{-1}(\mathbf{R}_e\mathbf{W}) - 2(\mathbf{R}_e\mathbf{W}^H)(\mathbf{R}_e Y)^{-1}(\mathbf{R}_e\mathbf{W}) + \mathbf{E}\boldsymbol{y}^H\boldsymbol{y}
\end{aligned}
$$

$$
\begin{aligned}
\mathsf{C}(\boldsymbol{\theta}_{\text{opt}})|_{\mathbf{R}\text{ real}} &= (\mathbf{R}_e\mathbf{W}^H)(\mathbf{R}_e Y)^{-1}\mathbf{R}(\mathbf{R}_e Y)^{-1}(\mathbf{R}_e\mathbf{W}) - 2(\mathbf{R}_e\mathbf{W}^H)(\mathbf{R}_e Y)^{-1}(\mathbf{R}_e\mathbf{W}) + \mathbf{E}\boldsymbol{y}^H\boldsymbol{y}\\
&= (\mathbf{R}_e\mathbf{W}^H)\mathbf{R}^{-1}\mathbf{R}\mathbf{R}^{-1}(\mathbf{R}_e\mathbf{W}) - 2(\mathbf{R}_e\mathbf{W}^H)\mathbf{R}^{-1}(\mathbf{R}_e\mathbf{W}) + \mathbf{E}\boldsymbol{y}^H\boldsymbol{y}\\
&= (\mathbf{R}_e\mathbf{W}^H)\mathbf{R}^{-1}(\mathbf{R}_e\mathbf{W}) - 2(\mathbf{R}_e\mathbf{W}^H)\mathbf{R}^{-1}(\mathbf{R}_e\mathbf{W}) + \mathbf{E}\boldsymbol{y}^H\boldsymbol{y}\\
&= \mathbf{E}\boldsymbol{y}^H\boldsymbol{y} - (\mathbf{R}_e\mathbf{W}^H)\mathbf{R}^{-1}(\mathbf{R}_e\mathbf{W})
\end{aligned}
$$

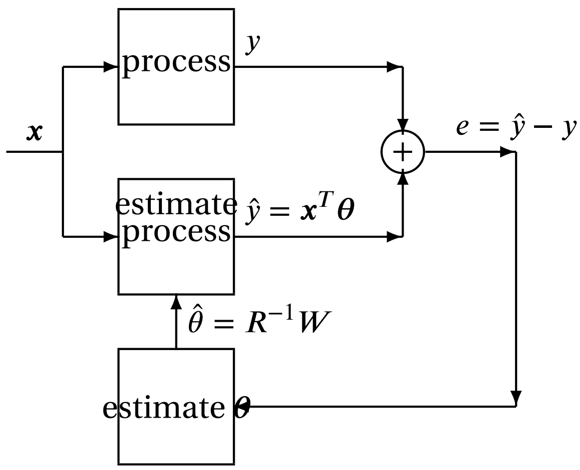


Figure 11.2: Adative filter example

In many adaptive filter and equalization applications, the autocorrelation matrix U is simply the m-element random data vector $\boldsymbol{x}(k)$ at time k, as in the *Wiener-Hopf equations* (next).

Corollary 11.1 (Wiener-Hopf equations). [4]

COR

$$
\left\{ U \triangleq \boldsymbol{x}(k) \triangleq \begin{bmatrix} x(k)\\ x(k-1)\\ x(k-2)\\ \vdots\\ x(k-m+1)\end{bmatrix}\right\} \implies \left\{ \begin{array}{rcl} \hat{\theta}_{\text{mms}} &=& R^{-1}W\\ \mathsf{C}(\hat{\theta}_{\text{mms}}) &=& W^T R^{-1} R R^{-1} W - 2W^T R^{-1} W + \mathbf{E}\boldsymbol{y}^T\boldsymbol{y} \end{array}\right\}
$$

[4] Ifeachor and Jervis (1993) pages 547–549 ⟨§"9.3 Basic Wiener filter theory"⟩, Ifeachor and Jervis (2002) pages 651–654 ⟨§"10.3 Basic Wiener filter theory"⟩, Kay (1988) page 51 ⟨§"3.3.3 Random Parameters"⟩

✎PROOF: This is a special case of the more general case discussed in Theorem 11.1 (page 86). Here, the dimension of U is $m \times 1$ (n=1). As a result, $\boldsymbol{y}$, $\hat{\boldsymbol{y}}$, and $\boldsymbol{e}$ are simply scalar quantities (not vectors). In this special case, we have the following results (Figure 11.2 page 87):

$$\begin{array}{rcl}
\hat{y}(\boldsymbol{\theta}) & \triangleq & \boldsymbol{x}^T\boldsymbol{\theta} \\
e(\boldsymbol{\theta}) & \triangleq & \hat{y} - y \\
\mathsf{C}(\boldsymbol{\theta}) & \triangleq & \mathbf{E}\,\|e\|^2 \triangleq \mathbf{E}\left[e^2\right] \\
\hat{\theta}_{\mathsf{mms}} & \triangleq & \arg\min_\theta \mathsf{C}(\boldsymbol{\theta})
\end{array}
\qquad
\begin{array}{rcl}
R & \triangleq & \mathbf{E}\left[\boldsymbol{x}\boldsymbol{x}^T\right] \\
W & \triangleq & \mathbf{E}[\boldsymbol{x}y] \\
\mathsf{C}(\boldsymbol{\theta}) & = & \boldsymbol{\theta}^T R\boldsymbol{\theta} - 2W^T\boldsymbol{\theta} + \mathbf{E}\left[\boldsymbol{y}^T\boldsymbol{y}\right] \\
\nabla_\theta \mathsf{C}(\boldsymbol{\theta}) & = & 2R\boldsymbol{\theta} - 2W \\
\mathsf{C}(\hat{\theta}_{\mathsf{mms}})|_R\text{ real} & = & \mathbf{E}\boldsymbol{y}^T\boldsymbol{y} - W^T R^{-1} W
\end{array}$$

✏

11.2 Least squares

11.2.1 General results

Theorem 11.2 (Least squares). *Let*

$$\begin{array}{rcl}
\hat{\boldsymbol{y}}(\boldsymbol{\theta}) & \triangleq & U^H\boldsymbol{\theta} \\
\boldsymbol{e}(\boldsymbol{\theta}) & \triangleq & \hat{\boldsymbol{y}} - \boldsymbol{y} \\
\mathsf{C}(\boldsymbol{\theta}) & \triangleq & \|\boldsymbol{e}\|^2 \triangleq \boldsymbol{e}^H\boldsymbol{e}
\end{array}
\qquad
\begin{array}{rcl}
\hat{\theta}_{\mathsf{ls}} & \triangleq & \arg\min_\theta \mathsf{C}(\boldsymbol{\theta}) \\
R & \triangleq & UU^H \\
W & \triangleq & U\boldsymbol{y}.
\end{array}$$

Then

THM

$$\begin{array}{rcl}
\hat{\theta}_{\mathsf{ls}} & = & (\mathbf{R_e}Y)^{-1}(\mathbf{R_e}W) \\
\mathsf{C}(\boldsymbol{\theta}) & = & \boldsymbol{\theta}^H R\boldsymbol{\theta} - (W^H\boldsymbol{\theta})^* - W^H\boldsymbol{\theta} + \mathbf{E}\boldsymbol{y}^H\boldsymbol{y} \\
\nabla_\theta \mathsf{C}(\boldsymbol{\theta}) & = & 2\mathbf{R_e}[Y]\boldsymbol{\theta} - 2\mathbf{R_e}W \\
\mathsf{C}(\hat{\theta}_{\mathsf{ls}}) & = & (\mathbf{R_e}W^H)(\mathbf{R_e}Y)^{-1}R(\mathbf{R_e}Y)^{-1}(\mathbf{R_e}W) - 2(\mathbf{R_e}W^H)(\mathbf{R_e}Y)^{-1}(\mathbf{R_e}W) + \mathbf{E}\boldsymbol{y}^H\boldsymbol{y} \\
\mathsf{C}(\hat{\theta}_{\mathsf{ls}})|_{R\ real} & = & \mathbf{E}\boldsymbol{y}^H\boldsymbol{y} - (\mathbf{R_e}W^H)R^{-1}(\mathbf{R_e}W).
\end{array}$$

✎PROOF: See APPENDIX F (page 209) for a Matrix Calculus reference.

$$\begin{aligned}
\mathsf{C}(\boldsymbol{\theta}) &\triangleq \|\boldsymbol{e}\|^2 \\
&= \boldsymbol{e}^H\boldsymbol{e} \\
&= (\hat{\boldsymbol{y}} - \boldsymbol{y})^H(\hat{\boldsymbol{y}} - \boldsymbol{y}) \\
&= \left(U^H\boldsymbol{\theta} - \boldsymbol{y}\right)^H\left(U^H\boldsymbol{\theta} - \boldsymbol{y}\right) \\
&= \left(\boldsymbol{\theta}^H U - \boldsymbol{y}^H\right)\left(U^H\boldsymbol{\theta} - \boldsymbol{y}\right) \\
&= \boldsymbol{\theta}^H UU^H\boldsymbol{\theta} - \boldsymbol{\theta}^H U\boldsymbol{y} - \boldsymbol{y}^H U^H\boldsymbol{\theta} + \boldsymbol{y}^H\boldsymbol{y} \\
&= \boldsymbol{\theta}^H R\boldsymbol{\theta} - (W^H\boldsymbol{\theta})^H - W^H\boldsymbol{\theta} + \boldsymbol{y}^H\boldsymbol{y} \\
&= \boldsymbol{\theta}^H R\boldsymbol{\theta} - (W^H\boldsymbol{\theta})^* - W^H\boldsymbol{\theta} + \boldsymbol{y}^H\boldsymbol{y} \\
&= \boldsymbol{\theta}^H R\boldsymbol{\theta} - (W^H)^*\boldsymbol{\theta} - W^H\boldsymbol{\theta} + \boldsymbol{y}^H\boldsymbol{y} \\
&= \boldsymbol{\theta}^H R\boldsymbol{\theta} - 2\mathbf{R_e}\left[W^H\right]\boldsymbol{\theta} + \boldsymbol{y}^H\boldsymbol{y}
\end{aligned}$$

$$\begin{aligned}
\nabla_\theta\mathsf{C}(\boldsymbol{\theta}) &= \nabla_\theta\left[\boldsymbol{\theta}^H R\boldsymbol{\theta} - (W^H)^*\boldsymbol{\theta} - W^H\boldsymbol{\theta} + \boldsymbol{y}^H\boldsymbol{y}\right] \\
&= R\boldsymbol{\theta} + R^T\boldsymbol{\theta} - [(W^H)^*]^T - [W^H]^T + 0 \\
&= R\boldsymbol{\theta} + (R^H)^*\boldsymbol{\theta} - W - W^* \\
&= R\boldsymbol{\theta} + R^*\boldsymbol{\theta} - W - W^* \\
&= (R + R^*)\boldsymbol{\theta} - (W + W^*) \\
&= 2(\mathbf{R_e}Y)\boldsymbol{\theta} - 2\mathbf{R_e}W
\end{aligned}$$

$$\theta_{\text{opt}} = (\mathbf{R_e}Y)^{-1}(\mathbf{R_e}W)$$

$$\begin{aligned}
\mathsf{C}(\theta_{\text{opt}}) &= \theta^H_{\text{opt}}R\theta_{\text{opt}} - 2\mathbf{R_e}\left[W^H\right]\theta_{\text{opt}} + \boldsymbol{y}^H\boldsymbol{y} \\
&= [(\mathbf{R_e}Y)^{-1}(\mathbf{R_e}W)]^H R[(\mathbf{R_e}Y)^{-1}(\mathbf{R_e}W)] - 2\mathbf{R_e}\left[W^H\right][(\mathbf{R_e}Y)^{-1}(\mathbf{R_e}W)] + \boldsymbol{y}^H\boldsymbol{y} \\
&= (\mathbf{R_e}W^H)(\mathbf{R_e}Y)^{-H}R(\mathbf{R_e}Y)^{-1}(\mathbf{R_e}W) - 2\mathbf{R_e}\left[W^H\right](\mathbf{R_e}Y)^{-1}(\mathbf{R_e}W) + \boldsymbol{y}^H\boldsymbol{y} \\
&= (\mathbf{R_e}W^H)(\mathbf{R_e}R^H)^{-1}R(\mathbf{R_e}Y)^{-1}(\mathbf{R_e}W) - 2\mathbf{R_e}\left[W^H\right](\mathbf{R_e}Y)^{-1}(\mathbf{R_e}W) + \boldsymbol{y}^H\boldsymbol{y} \\
&= (\mathbf{R_e}W^H)(\mathbf{R_e}Y)^{-1}R(\mathbf{R_e}Y)^{-1}(\mathbf{R_e}W) - 2(\mathbf{R_e}W^H)(\mathbf{R_e}Y)^{-1}(\mathbf{R_e}W) + \boldsymbol{y}^H\boldsymbol{y}
\end{aligned}$$

$$\begin{aligned}
\mathsf{C}(\theta_{\text{opt}})|_{R\text{ real}} &= (\mathbf{R_e}W^H)(\mathbf{R_e}Y)^{-1}R(\mathbf{R_e}Y)^{-1}(\mathbf{R_e}W) - 2(\mathbf{R_e}W^H)(\mathbf{R_e}Y)^{-1}(\mathbf{R_e}W) + \boldsymbol{y}^H\boldsymbol{y} \\
&= (\mathbf{R_e}W^H)R^{-1}RR^{-1}(\mathbf{R_e}W) - 2(\mathbf{R_e}W^H)R^{-1}(\mathbf{R_e}W) + \boldsymbol{y}^H\boldsymbol{y} \\
&= (\mathbf{R_e}W^H)R^{-1}(\mathbf{R_e}W) - 2(\mathbf{R_e}W^H)R^{-1}(\mathbf{R_e}W) + \boldsymbol{y}^H\boldsymbol{y} \\
&= \boldsymbol{y}^H\boldsymbol{y} - (\mathbf{R_e}W^H)R^{-1}(\mathbf{R_e}W)
\end{aligned}$$

✏

11.2.2 Examples using polynomial modeling

Example 11.1 (Polynomial approximation).

Suppose we **know** the locations $\{(x_n, y_n)|n = 1,2,3,4,5\}$ of 5 data points. Let $\boldsymbol{x}$ and $\boldsymbol{y}$ represent the locations of these points such that

$$\boldsymbol{x} \triangleq \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} \qquad \boldsymbol{y} \triangleq \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix}$$

Suppose we want to find a second order polynomial $cx^2 + bx + a$ that best approximates these 5 points in the least squares sense. We define the matrix U (known) and vector $\hat{\theta}$ (to be computed) as follows:[5]

$$U^H \triangleq \underbrace{\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ 1 & x_4 & x_4^2 \\ 1 & x_5 & x_5^2 \end{bmatrix}}_{\text{Vandermonde matrix}} \qquad \hat{\theta} \triangleq \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

Then, using Theorem 11.2 (page 88), the best coefficients $\hat{\theta}$ for the polynomial are

$$\hat{\theta} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \underbrace{(UU^H)^{-1}}_{R^{-1}}\underbrace{(U\boldsymbol{y})}_{W} = \left(\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ 1 & x_4 & x_4^2 \\ 1 & x_5 & x_5^2 \end{bmatrix}^H \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ 1 & x_4 & x_4^2 \\ 1 & x_5 & x_5^2 \end{bmatrix}\right)^{-1}\left(\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ 1 & x_4 & x_4^2 \\ 1 & x_5 & x_5^2 \end{bmatrix}^H \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix}\right)$$

Example 11.2. [6] Find the best fit 3rd order polynomial $\mathsf{p}(x) = dx^3 + cx^2 + bx + a$ for the climate measurements illustrated in Figure 11.3 (page 90).

$$\hat{\theta} = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \left(\mathbf{U}\mathbf{U}^H\right)^{-1}(\mathbf{U}\boldsymbol{y}) = \left(\begin{bmatrix} 1 & 0 & 0^2 & 0^3 \\ 1 & 1 & 1^2 & 1^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 138 & 138^2 & 138^3 \end{bmatrix}^H \begin{bmatrix} 1 & 0 & 0^2 & 0^3 \\ 1 & 1 & 1^2 & 1^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 138 & 138^2 & 138^3 \end{bmatrix}\right)^{-1}\left(\begin{bmatrix} 1 & 0 & 0^2 & 0^3 \\ 1 & 1 & 1^2 & 1^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 138 & 138^2 & 138^3 \end{bmatrix}^H \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{138} \end{bmatrix}\right)$$

[5] Horn and Johnson (1990)29
[6] NASA (2019), UEA (2019)

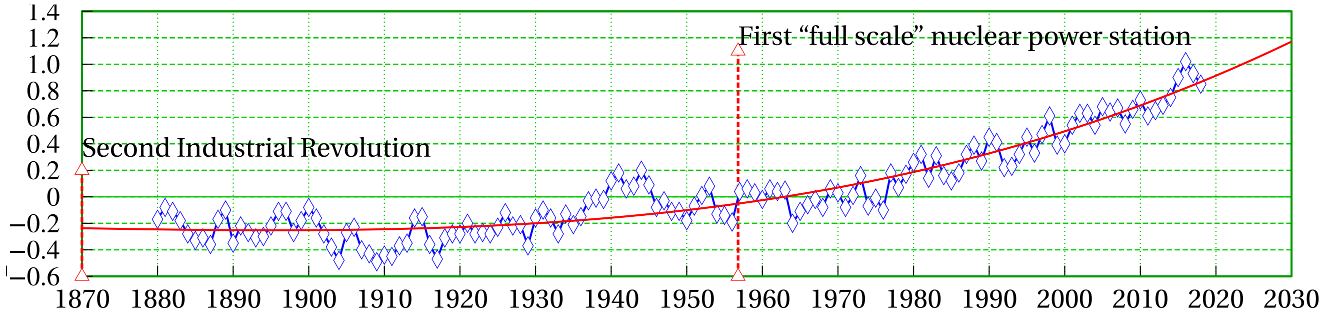


Figure 11.3: Global temperature deviations from average over time (Example 11.2 page 89)

$$= \frac{1}{139}\begin{bmatrix} 1 & 69 & 6371 & 661779 \\ 69 & 6371 & 661779 & 73323839 \\ 6371 & 661779 & 73323839 & 8462609259 \\ 661779 & 73323839 & 8462609259 & 1004606426831 \end{bmatrix}^{-1} \left(\begin{bmatrix} 1 & 0 & 0^2 & 0^3 \\ 1 & 1 & 1^2 & 1^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 138 & 138^2 & 138^3 \end{bmatrix}^H \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{138} \end{bmatrix} \right)$$

$$= \begin{bmatrix} -0.24606376973 \\ -0.00076398238 \\ 0.00001657135 \\ 0.00000034336 \end{bmatrix}$$

Using this polynomial, we could “predict”/estimate (not necessarily very well) what the temperature deviation will be in 2030 (+1.1710) and was in 1870 (-0.23711).

```
Y = [
  -0.17
  -0.08
  -0.11
   ...
   0.93
   0.85
 ];
N = length(Y);
X = [0:2018-1880]';    % years-1880
U = [ones(1,N) ; X', ; X'.^2, ; X'.^3];
coefs = inv(U*U')*(U*Y)
a=coefs(1)
b=coefs(2)
c=coefs(3)
d=coefs(4)
bestFit = d*X.^3 + c*X.^2 + b*X + a;
plot(bestFit)
x = 2030;
predict2030 = d*(x-1880)^3 + c*(x-1880)^2 + b*(x-1880) + a
x = 1870;
predict1870 = d*(x-1880)^3 + c*(x-1880)^2 + b*(x-1880) + a
```

Example 11.3. [7] Using a 3rd order polynomial fitting (Figure 11.4 page 91), we can predict that the number of cagefree chickens in the U.S. will reach 32.07% in 2022. Note that this is also an example of **non-uniform sampling**. That is, there is only one sample per year before 2016, but multiple samples per year from 2016 onward. This is no problem for the LS algorithm. Simply use the data points that are available and ignore the ones that are not:

$$\mathbf{U}^H = \begin{bmatrix} 1 & 07.997260 & 07.997260^2 & 07.997260^3 \\ 1 & 08.997268 & 08.997268^2 & 08.997268^3 \\ 1 & 09.997260 & 09.997260^2 & 09.997260^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 15.997260 & 15.997260^2 & 15.997260^3 \\ 1 & 16.581967 & 16.581967^2 & 16.581967^3 \\ 1 & 16.666667 & 16.666667^2 & 16.666667^3 \\ 1 & 16.748634 & 16.748634^2 & 16.748634^3 \\ 1 & 16.833333 & 16.833333^2 & 16.833333^3 \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

[7] Mendez (2019)

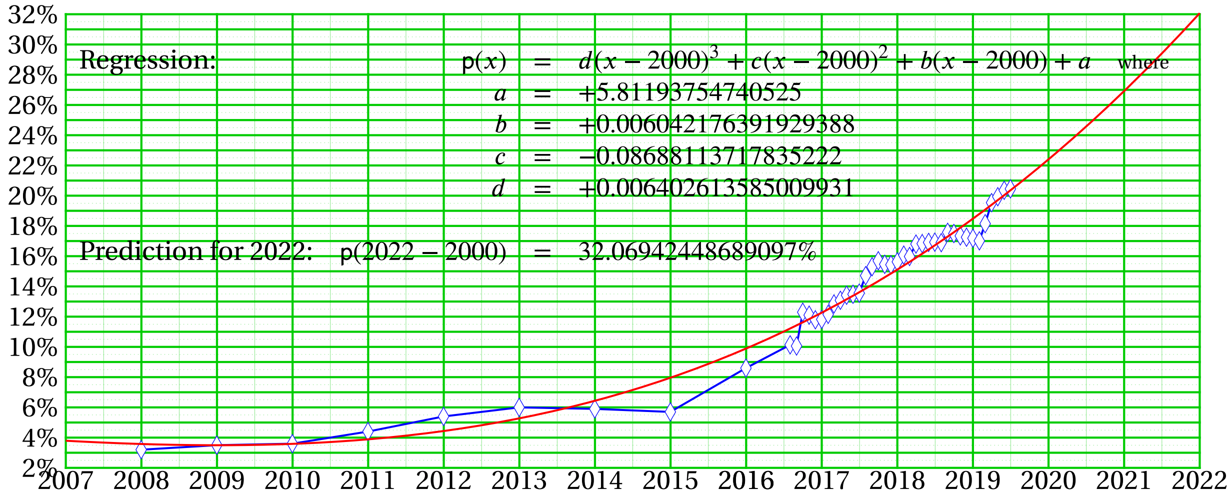


Figure 11.4: Ratio of cagefree chickens in U.S. (Example 11.3 page 90)

```
% years - 2000
X = [
   07.997260  % 2007-12-31 decimal form
   08.997268  % 2008-12-31 decimal form
   09.997260  % 2009-12-31 decimal form
   10.997260  % 2010-12-31 decimal form
   11.997260  % 2011-12-31 decimal form
   12.997268  % 2012-12-31 decimal form
   13.997260  % 2013-12-31 decimal form
   14.997260  % 2014-12-31 decimal form
   15.997260  % 2015-12-31 decimal form
   16.581967  % 2016-08-01 decimal form
   16.666667  % 2016-09-01 decimal form
   16.748634  % 2016-10-01 decimal form
   16.833333  % 2016-11-01 decimal form
   16.915301  % 2016-12-01 decimal form
   ...
 ];
Y = [
   03.2000    % data for 2007-12-31
   03.5000    % data for 2008-12-31
   03.6000    % data for 2009-12-31
   04.4000    % data for 2010-12-31
   05.4000    % data for 2011-12-31
   06.0000    % data for 2012-12-31
   05.9000    % data for 2013-12-31
   05.7000    % data for 2014-12-31
   08.6000    % data for 2015-12-31
   10.1357    % data for 2016-08-01
   10.0570    % data for 2016-09-01
   12.2935    % data for 2016-10-01
   12.1006    % data for 2016-11-01
   11.7935    % data for 2016-12-01
   ...
 ];
format long g
pkg load signal;
N = length(Y);
U = [ones(1,N) ; X', ; X'.^2, ; X'.^3];
coefs = inv(U*U')*(U*Y)
a=coefs(1)
b=coefs(2)
c=coefs(3)
d=coefs(4)
bestFit = d*X.^3 + c*X.^2 + b*X + a;
plot(bestFit)
x = 2022;
predict2022 = d*(x-2000)^3 + c*(x-2000)^2 + b*(x-2000) + a
```

You can convert a date in the yyyy-mm-dd form such as 2007-12-31 to its decimal form 07.997260 using R with the lubridate package like this:

```
install.packages("lubridate");
require(lubridate);
x     = read.csv("../common/datasets/cagefree-ratios_osf-6hty8.csv", comment.char = "#");
ratio = x$ratio_hens;
year  = lubridate::decimal_date(lubridate::ymd(x$observed_month));
plot( year, ratio, col="blue", type='o' );
```

Example 11.4. [8] let $tn = c(0 : 10)$ and $yn = c(18, 33, 56, 90, 130, 170, 203, 225, 239, 247, 251)$ represent the growth of a plant.

You want to find the α that minimizes that cost; that is, you want to find where (with respect to α) the cost function "goes to the lowest point". Suppose you set $N_0 \triangleq N(0) = 18$ and $N_* \triangleq 252$. Let $\hat{\mathsf{h}}(t) \triangleq \dfrac{H_\infty}{1+\left(\frac{H_\infty}{H_0}-1\right)e^{-\alpha t}}$

✎PROOF:

1. definition: Define a cost function (the error cost) as the norm squared of $e(t_n)$ as

$$\mathsf{C}(\alpha) \triangleq \|e\|^2 \triangleq \sum_{n=0}^{n=10} e^2(t_n)$$

2. lemma:

$$\begin{aligned}
\boxed{\frac{\partial}{\partial\alpha}\hat{\mathsf{h}}(t)} &\triangleq \frac{\partial}{\partial\alpha}\left[\frac{H_\infty}{1+\left(\frac{H_\infty}{H_0}-1\right)e^{-\alpha t}}\right] && \text{by definition of } \hat{\mathsf{h}}(t)\\
&= \frac{0-H_\infty\left[\frac{H_\infty}{H_0}-1\right]e^{-\alpha t}(-t)}{\left(1+\left[\frac{H_\infty}{H_0}-1\right]e^{-\alpha t}\right)^2} && \text{by Quotient Rule}\\
&= \frac{H_\infty^2}{\left(1+\left[\frac{H_\infty}{H_0}-1\right]e^{-\alpha t}\right)^2}\left[\frac{1}{H_0}-\frac{1}{H_\infty}\right]te^{-\alpha t}\\
&\triangleq \boxed{\left[\frac{1}{H_0}-\frac{1}{H_\infty}\right]\hat{\mathsf{h}}^2(t)te^{-\alpha t}} && \text{by definition of } \hat{\mathsf{h}}(t)
\end{aligned}$$

3. lemma:

$$\begin{aligned}
\boxed{\frac{\partial}{\partial A}\hat{\mathsf{h}}(t)} &\triangleq \frac{\partial}{\partial A}\left[\frac{A}{1+\left(\frac{A}{B}-1\right)e^{-\alpha t}}\right] && \text{by definition of } \hat{\mathsf{h}}(t)\\
&= \frac{\left[1+\left(\frac{A}{B}-1\right)e^{-\alpha t}\right]-A\left[\frac{1}{B}\right]e^{-\alpha t}}{\left(1+\left[\frac{A}{B}-1\right]e^{-\alpha t}\right)^2} && \text{by Quotient Rule}\\
&= \frac{1-e^{-\alpha t}}{\left(1+\left[\frac{A}{B}-1\right]e^{-\alpha t}\right)^2}\\
&= \left[\frac{1-e^{-\alpha t}}{A^2}\right]\left[\frac{A}{1+\left[\frac{A}{B}-1\right]e^{-\alpha t}}\right]^2
\end{aligned}$$

[8] https://math.stackexchange.com/questions/3990086/

$$\triangleq \boxed{\left[\frac{1-e^{-\alpha t}}{A^2}\right]\hat{\mathsf{h}}^2(t)} \qquad \text{by definition of } \hat{\mathsf{h}}(t)$$

4. lemma:

$$\begin{aligned}
\boxed{\frac{\partial}{\partial N_0}\hat{\mathsf{h}}(t)} &\triangleq \frac{\partial}{\partial N_0}\left[\frac{N_*}{1+\left(\frac{N_*}{N_0}-1\right)e^{-a_0 t}}\right] && \text{by definition of } \hat{\mathsf{h}}(t)\\
&= \frac{-N_*\left(\frac{-1}{N_0^2}\right)e^{-a_0 t}}{\left[1+\left(\frac{N_*}{N_0}-1\right)e^{-a_0 t}\right]^2} && \text{by Quotient Rule}\\
&= \left[\frac{e^{-a_0 t}}{N_* N_0^2}\right]\left[\frac{N_*}{1+\left[\frac{N_*}{N_0}-1\right]e^{-a_0 t}}\right]^2\\
&= \left[\frac{e^{-a_0 t}}{N_* N_0^2}\right]\hat{\mathsf{h}}^2(t) && \text{by definition of } \hat{\mathsf{h}}(t)
\end{aligned}$$

5.

$$\begin{aligned}
\boxed{0} &= \frac{1}{2\left(\frac{1}{H_0}-\frac{1}{H_\infty}\right)}\cdot 0\\
&= \frac{1}{2\left(\frac{1}{H_0}-\frac{1}{H_\infty}\right)}\frac{\partial}{\partial\alpha}\|e\|^2\\
&\triangleq \frac{1}{2\left(\frac{1}{H_0}-\frac{1}{H_\infty}\right)}\frac{\partial}{\partial\alpha}\sum_{n=0}^{n=10} e^2(t_n) && \text{by definition of } \|\cdot\|\\
&\triangleq \frac{1}{2\left(\frac{1}{H_0}-\frac{1}{H_\infty}\right)}\frac{\partial}{\partial\alpha}\sum_{n=0}^{n=10}\left[\hat{\mathsf{h}}(t_n)-y_n\right]^2 && \text{by definition of } e\\
&= \frac{1}{2\left(\frac{1}{H_0}-\frac{1}{H_\infty}\right)}\sum_{n=0}^{n=10} 2\left[\hat{\mathsf{h}}(t_n)-y_n\right]\hat{\mathsf{h}}'(t_n) && \text{by \textit{Chain Rule}}\\
&= \frac{1}{2\left(\frac{1}{H_0}-\frac{1}{H_\infty}\right)}\sum_{n=0}^{n=10} 2\left[\hat{\mathsf{h}}(t_n)-y_n\right]\left[\frac{1}{H_0}-\frac{1}{H_\infty}\right]\hat{\mathsf{h}}^2(t_n)te^{-\alpha t_n} && \text{by (2) lemma}\\
&= \boxed{\sum_{n=0}^{n=10}\hat{\mathsf{h}}^2(t_n)\left[\hat{\mathsf{h}}(t_n)-y_n\right]\left[t_n e^{-\alpha t_n}\right]}\\
&\triangleq DC(\alpha) && \text{(call the sum } DC(\alpha))
\end{aligned}$$

6.

$$\begin{aligned}
\boxed{0} &= \frac{A^2}{2}\cdot 0\\
&= \frac{A^2}{2}\frac{\partial}{\partial A}\|e\|^2\\
&\triangleq \frac{A^2}{2}\frac{\partial}{\partial A}\sum_{n=0}^{n=10} e^2(t_n) && \text{by definition of } \|\cdot\|
\end{aligned}$$

$$\triangleq \frac{A^2}{2}\frac{\partial}{\partial A}\sum_{n=0}^{n=10}\left[\hat{\mathsf{h}}(t_n) - y_n\right]^2 \qquad \text{by definition of } e$$
$$= \frac{A^2}{2}\sum_{n=0}^{n=10} 2\left[\hat{\mathsf{h}}(t_n) - y_n\right]\hat{\mathsf{h}}'(t_n) \qquad \text{by \textit{Chain Rule}}$$
$$= A^2\sum_{n=0}^{n=10}\left[\hat{\mathsf{h}}(t_n) - y_n\right]\left[\frac{1-e^{-\alpha t_n}}{A^2}\right]\hat{\mathsf{h}}^2(t_n) \qquad \text{by (3) lemma}$$
$$= \boxed{\sum_{n=0}^{n=10}\hat{\mathsf{h}}^2(t_n)\left[\hat{\mathsf{h}}(t_n) - y_n\right]\left[1-e^{-\alpha t_n}\right]}$$

7.

$$\boxed{0} = \frac{N_* N_0^2}{2}\cdot 0$$
$$= \frac{N_* N_0^2}{2}\frac{\partial}{\partial N_0}\|e\|^2$$
$$\triangleq \frac{N_* N_0^2}{2}\frac{\partial}{\partial N_0}\sum_{n=0}^{n=10} e^2(t_n) \qquad \text{by definition of } \|\cdot\|$$
$$\triangleq \frac{N_* N_0^2}{2}\frac{\partial}{\partial N_0}\sum_{n=0}^{n=10}\left[\hat{\mathsf{h}}(t_n) - y_n\right]^2 \qquad \text{by definition of } e$$
$$= \frac{N_* N_0^2}{2}\sum_{n=0}^{n=10} 2\left[\hat{\mathsf{h}}(t_n) - y_n\right]\frac{\partial}{\partial N_0}\hat{\mathsf{h}}(t_n) \qquad \text{by Chain Rule}$$
$$= \frac{N_* N_0^2}{2}\sum_{n=0}^{n=10} 2\left[\hat{\mathsf{h}}(t_n) - y_n\right]\left[\frac{e^{-a_0 t_n}}{N_* N_0^2}\right]\hat{\mathsf{h}}^2(t_n) \qquad \text{by lemma}$$
$$= \boxed{\sum_{n=0}^{n=10}\hat{\mathsf{h}}^2(t_n)\left[\hat{\mathsf{h}}(t_n) - y_n\right]e^{-a_0 t_n}}$$

Plotting $Dcost$ with respect to α, it appears that $DC(\alpha)$ crosses 0 at around $\alpha = 0.66$.

The uniroot function from the R stats package indicates that $DC(\alpha)$ crosses 0 at $\alpha = 0.6631183$ with estim.prec=6.103516e-05.

Using $H_0 \triangleq 18$, $H_\infty \triangleq 252$, and $\alpha \triangleq 0.6631183$, $\hat{\mathsf{h}}(t)$ seems to fit the 11 data points fairly well ($cost(0.6631183) = 31.32307$) …

✎

11.3 Recursive forms

One of the biggest advantages of using a recursive form / gradient search technique is that it can be implemented *recursively* as shown in the next equation. The general form of the gradient search parameter estimation techniques is[9]

[9] Nelles (2001) page 90

THM $\boldsymbol{\theta}_n = \boldsymbol{\theta}_{n-1} - \eta_{n-1} R \left[\nabla_{\boldsymbol{\theta}} \mathsf{C}(\boldsymbol{\theta}_n)\right]$ where at time n

$\boldsymbol{\theta}_n$	is the *state*	(vector)
η_n	is the *step size*	(scalar)
Y	is the *direction*	(matrix)
$\nabla_{\boldsymbol{\theta}} \mathsf{C}(\boldsymbol{\theta}_n)$	is the *gradient* of the cost function $\mathsf{C}(\boldsymbol{\theta}_n)$	(vector)

Two major categories of gradient search techniques are

- steepest descent (includes LMS)
- Newton's method (includes RLS and Kalman filters).

The key difference between the two is that ***steepest descent* uses only first derivative information**, while ***Newton's method* uses both first and second derivative information** making it converge much faster but with significantly higher complexity.

First derivative techniques

Steepest descent. In this algorithm, $R = I$ (identity matrix). First derivative information is contained in $\nabla\mathsf{C}$. Second derivative information, if present, is contained in Y. Thus, steepest descent algorithms do not use second derivative information.

THM $\boldsymbol{\theta}_n = \boldsymbol{\theta}_{n-1} - \eta_{n-1} \left[\nabla_{\boldsymbol{\theta}} \mathsf{C}(\boldsymbol{\theta}_n)\right]$

Least Mean Squares (LMS). [10] This is a special case of *steepest descent*. In minimum mean square estimation (Section 11.1 page 86), the cost function $\mathsf{C}(\boldsymbol{\theta})$ is defined as a *statistical average* of the error vector such that $\mathsf{C}(\boldsymbol{\theta}) = \mathbf{E}\left[\boldsymbol{e}^H \boldsymbol{e}\right]$. In this case the gradient $\nabla\mathsf{C}$ is difficult to compute. However, the LMS algorithm greatly simplifies the problem by instead defining the cost function as a function of the *instantaneous error* such that

$$\begin{aligned}
\boldsymbol{y} &= y(n)\\
\hat{\boldsymbol{y}} &= \hat{y}(n)\\
\mathsf{C}(\boldsymbol{\theta}) &= \|e(n)\|^2\\
&= e^2(n)\\
&= (\hat{y}(n) - y(n))^2
\end{aligned}$$

Computing the gradient of this cost function is then just a special case of *least squares estimation* (Section 11.2 page 88). Using LS, we let $U = \boldsymbol{x}^T$ and hence

$$\begin{aligned}
\nabla_{\boldsymbol{\theta}} \mathsf{C}(\boldsymbol{\theta}) &= 2U^T U \boldsymbol{\theta} - 2U^T \boldsymbol{y} && \text{by Theorem 11.2 page 88}\\
&= 2\boldsymbol{x}\boldsymbol{x}^T \boldsymbol{\theta} - 2\boldsymbol{x}y && \text{by above definitions}\\
&= 2\boldsymbol{x}\hat{y} - 2\boldsymbol{x}y\\
&= 2\boldsymbol{x}(\hat{y} - y)\\
&= 2\boldsymbol{x}e(n)
\end{aligned}$$

[10] Manolakis et al. (2000) page 526

The LMS algorithm uses this instantaneous gradient for $\nabla \mathsf{C}$, lets $\boldsymbol{R} = \boldsymbol{I}$, and uses a constant step size η to give

THM
$$\boldsymbol{\theta}_n = \boldsymbol{\theta}_{n-1} - 2\eta \boldsymbol{x}_n e(n)$$

Second derivative techniques

Newton's Method. This algorithm uses the *Hessian* matrix $\boldsymbol{H}$, which is the second derivative of the cost function $\mathsf{C}(\boldsymbol{\theta})$, and lets $\boldsymbol{R} = \boldsymbol{H}^{-1}$.

$$\begin{aligned}
\boldsymbol{H}_n &\triangleq \nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}} \mathsf{C}(\boldsymbol{\theta}_n) \\
\boldsymbol{\theta}_n &= \boldsymbol{\theta}_{n-1} - \eta_{n-1} \boldsymbol{H}_n^{-1} \left[\nabla_{\boldsymbol{\theta}} \mathsf{C}(\boldsymbol{\theta}_n)\right]
\end{aligned}$$

Kalman filtering [11]

$$\begin{aligned}
\gamma(k) &= \frac{1}{x^T(k)\boldsymbol{P}(k-1)\boldsymbol{x}(k)+1} \boldsymbol{P}(k-1)x(k) \\
\boldsymbol{P}(k) &= (\boldsymbol{I} - \gamma(k)x^T(k))\boldsymbol{P}(k-1) + \boldsymbol{V} \\
e(k) &= y(k) - x^T(k)\hat{\boldsymbol{\theta}}(k-1) \\
\hat{\boldsymbol{\theta}}(k) &= \hat{\boldsymbol{\theta}}(k-1) + \gamma(k)e(k)
\end{aligned}$$

Recursive Least Squares (RLS) [12] This algorithm is a special case of either the RLS with forgetting or the Kalman filter.

$$\begin{aligned}
\gamma(k) &= \frac{1}{x^T(k)\boldsymbol{P}(k-1)\boldsymbol{x}(k)+1} \boldsymbol{P}(k-1)x(k) \\
\boldsymbol{P}(k) &= (\boldsymbol{I} - \gamma(k)x^T(k))\boldsymbol{P}(k-1) \\
e(k) &= y(k) - x^T(k)\hat{\boldsymbol{\theta}}(k-1) \\
\hat{\boldsymbol{\theta}}(k) &= \hat{\boldsymbol{\theta}}(k-1) + \gamma(k)e(k)
\end{aligned}$$

11.4 Direct search

A direct search algorithm may be used in cases where the cost function over $\boldsymbol{\theta}$ has several local minimums, making convergence difficult. Furthermore, direct search algorithms can be very computationally demanding.

[11] Nelles (2001) page 66
[12] Nelles (2001) page 66

CHAPTER 12

BAYESIAN NETWORK ALGORITHMS WHEN JOINT-PDF MODEL IS KNOWN

12.1 Sequential decoding

It has been shown that the Viterbi algorithm (trellis) produces an optimal estimate in the maximal likelihood (ML) sense. A Verterbi trellis is shown in Figure 12.1 (page 97).

12.2 References

1. Ben-Gal (2008) ⟨ISBN:9780470018613⟩: Introduction to Bayesian Networks

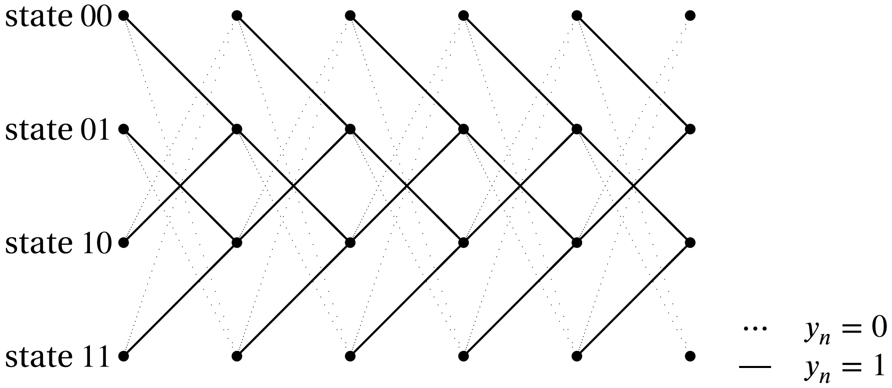


Figure 12.1: Viterbi Algorithm trellis

2. Pourret et al. (2008) ⟨ISBN:9780470994542⟩: A number of real-life examples including
   - Chapter 2: “Medical diagnosis”
   - Chapter 4: “Complex genetic models”
   - Chapter 5: “Crime risk factors analysis”
   - Chapter 7: “Inference problems in forensic science”
   - Chapter 12: “An information retrieval system”
   - Chapter 14: “Terrorism risk management”
   - Chapter 20: “Risk management in robotics”
   - Chapter 21: “Enhancing Human Cognition”
   - Section 22.1: “An artificial intelligence perspective”
   - Section 22.2: “A rational approach of knowledge”

3. Darwiche (2009) ⟨ISBN:9780521884389⟩
   - Chapter 17: “Learning: The Maximum Likelihood Approach”
   - Chapter 18: “Learning: The Bayesian Approach”

4. Fenton and Neil (2018) ⟨ISBN:9781351978965⟩: Risk assessment using Bayesian Networks

5. Jensen (2013) ⟨ISBN:9781475735024⟩: Relationship between Decision Graphs and Bayesian Networks

6. Chapmann (2017) ⟨ISBN:9781978304871⟩: HMM and Bayesian Networks

7. Fano (1963)

Part IV

System Model Estimation

CHAPTER 13

MODEL ESTIMATION

A key point in model building is choosing what to model. Defining a model with too many system properties (too many dimensions) can quickly lead to an intractable solution computationally and require more data than is available (the "curse of dimensionality").

For modeling real-world random processes above the quantum level, here are some possibly useful parameters to estimate and use:

1. Derivatives $0, 1, \ldots n$, where n may be 2 or 3. Brook Taylor showed that for *analytic* functions,[1] knowledge of the derivatives about a point at $x = a$ allows you to determine (predict) arbitrarily closely all the points in the vicinity of $x = a$:[2]
   $$\mathsf{f}(x) = \mathsf{f}(a) + \frac{1}{1!}\mathsf{f}'(a)[x-a] + \frac{1}{2!}\mathsf{f}''(x)[x-a]^2 + \frac{1}{3!}\mathsf{f}'''(x)[x-a]^3 + \cdots$$
   For linear or quasi-linear systems, $n = 2$ may be sufficient. For example, the classical "vibrating string" vertical displacement $\mathsf{u}(x,t)$ wave equation can be described as
   $$\frac{\partial^2 u}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} = 0$$
   Also note that all solutions of homogeneous second order differential equations are linear combinations of sine and cosine (Theorem I.3 page 232):
   $$\left\{\frac{\mathsf{d}^2}{\mathsf{d}x^2}\mathsf{f} + \mathsf{f} = 0\right\} \iff \left\{\mathsf{f}(x) = \mathsf{f}(0)\cos(x) + \mathsf{f}'(0)\sin(x)\right\} \qquad \forall \mathsf{f} \in \boldsymbol{C}, \forall x \in \mathbb{R}$$

2. The *Fourier Transform* is a kind of counter-part of the Taylor expansion demonstrated above:[3]

	Taylor coefficients	Fourier coefficients
🙵	Depend on derivatives $\frac{\mathsf{d}^n}{\mathsf{d}x^n}\mathsf{f}(x)$	Depend on integrals $\int_{x\in\mathbb{R}} \mathsf{f}(x)e^{-i\omega x}\,\mathsf{d}x$
🙵	Behavior in the vicinity of a point.	Behavior over the entire function.
🙵	Demonstrate trends locally.	Demonstrate trends globally.
🙵	Admits *analytic* functions only.	Admits *non-analytic* functions as well.
🙵	Function must be *continuous*.	Function can be *discontinuous*.

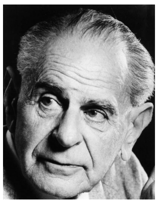
[1] analytic function: a function for which all its derivatives exist

[2] Robinson (1982) page 886

[3] Robinson (1982) page 886

CHAPTER 14

TRANSFER FUNCTION ESTIMATION



❝I can therefore gladly admit that falsificationists like myself much prefer an attempt to solve an interesting problem by a bold conjecture, even (and especially) if it so turns out to be false, to any recital of a sequence of irrelevant truisms. We prefer this because we believe that this is the way in which we can learn from our mistakes and that in finding that our conjecture was false we shall have learned much about the truth, and shall have got nearer to the truth.❞

Karl R. Popper (1902–1994) [1]

14.1 Estimation techniques

Let **S** be a *system* with *impulse response* $\mathsf{h}(n)$ with with *DTFT* $\tilde{\mathsf{H}}(\omega)$, input $\mathsf{x}(n)$, and output $\mathsf{y}(n)$. Often in the field of "digital signal processing" (DSP), **S** is a "filter" with known $\mathsf{h}(n)$ and $\tilde{\mathsf{H}}(\omega)$ because the filter **S** was designed by a designer who had direct control over $\mathsf{h}(n)$.

However in many other practical situations, **S** is some other system for which $\mathsf{h}(n)$ and $\tilde{\mathsf{H}}(\omega)$ are *not* known…but which we may want to *estimate*. Examples of such **S** is a device on an industrial shaker table, a communication channel, or the entire earth.

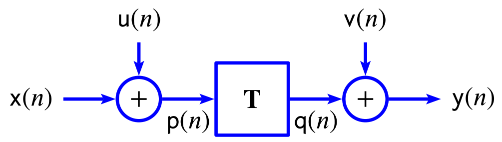
Determining $\mathsf{h}(n)$ and/or $\tilde{\mathsf{H}}(\omega)$ is part of an operation called "*system identication*". Determining $\tilde{\mathsf{H}}(\omega)$ in particular is referred to as "*Frequency Response Identification*"[2] or as "*Frequency Response Function*" ("*FRF*") estimation.[3] *FRF* estimation is a challenging problem and one that many have devoted much effort to. This chapter describes some of that effort.

In the early days, people used a rather obvious technique for determining $\tilde{\mathsf{H}}(\omega)$—the humble *sine sweep*. That is, they drove the input with a sine wave with slowly increasing (or decreasing) fre-

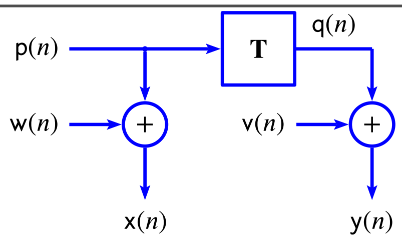
[1] quote: Popper (1962) page 231, Popper (1963) page 313
image: `https://en.wikipedia.org/wiki/File:Karl_Popper.jpg`, "no known copyright restrictions"

[2] Shin and Hammond (2008) page 292

[3] Cobb (1988) page 1 ⟨FRF "measurement"⟩



(A) **communications additive noise model**
The "input signal" is $x(n)$.



(B) **measurement additive noise model**
The "input signal" is $p(n)$.

In each model, $x(n)$ and $y(n)$ are "known", and $u(n)$, $v(n)$, and $w(n)$ are *not*.
In definition, the two models are **equivalent** under the relation $u(n) = -w(n)$.
In practice, they are **different**:
- in (A), x and u would be typically *uncorrelated*;
- in (B), x and $w = -u$ are very much *correlated* (x is a function of u).

Figure 14.1: Additive noise systems with *linear*/**non-linear** operator **T**

quency while measuring the resulting output. This technique, although effective, was "very slow".[4] *And* there is another problem here—we don't always have control over the input signal. Examples of this include earthquake and volcanic activity analysis.

An alternative to the sine-sweep input is *random sequence* input. All the techniques that follow in this chapter are of this type. A problem with using random sequences directly for estimating $\tilde{H}(\omega)$ is that the estimate $\hat{H}(\omega)$ is itself also random. This is not what we want. We want an estimate that we can actually write down on paper or at least plot on paper.

A solution to this is to not use the random sequences directly to estimate $\tilde{H}(\omega)$, but instead to first use the *expectation* operator **E** (Definition 1.1 page 3). The expectation operator takes a quantity X that is inherently "random" (with some probability distribution $p(x)$) and turns it into a deterministic "constant" **E**X.

The operator **E** is also used by the spectral density functions $\tilde{S}_{xx}(\omega)$ and $\tilde{S}_{xy}(\omega)$ (Definition 8.3 page 60). And $\tilde{S}_{xx}(\omega)$ and $\tilde{S}_{xy}(\omega)$ are what are typically used to calculate an estimate $\hat{H}(\omega)$.

14.2 Additive noise system models

Consider the additive noise systems illustrated in Figure 15.1 (page 129).

- The illustration on the left is suitable for modeling a communications system where x is the transmitted signal, y is the received signal, u and v are thermal noise, and the "transfer function" **H** is the communications channel (air, water, wires, etc.) that one wishes to estimate.
- The illustration on the right is suitable for modeling a testing system where p is an input test signal (from an industrial shaker or from a naturally occurring signal originating from geophysical activity), w is measurement noise, x is the measured input contaminated by noise, and **H** is the device under test (a piece of equipment, a building, or the entire earth).

[4] Leuridan et al. (1986)911"Stepped Sine Testing", Cobb (1988) page 1 ⟨Chapter 1—Introduction⟩, Ewins (1986) pages 125–140 ⟨3.7 Use of different excitation types⟩

Note that the two models are an equivalent system **S** under the relation $\mathsf{u} = -\mathsf{w}$. But although one might expect such a sign difference to wreak mathematical havoc in resulting equations, this is simply not the case here because

$$\tilde{\mathsf{S}}_{\mathsf{ww}} = \tilde{\mathbf{F}}\mathbf{E}[\mathsf{w}(m)\mathsf{w}*(0)] = \tilde{\mathbf{F}}\mathbf{E}\big[(-\mathsf{u}(m))\big(-\mathsf{u}^*(0)\big)\big] = \tilde{\mathbf{F}}\mathbf{E}\big[(\mathsf{u}(m))\big(\mathsf{u}^*(0)\big)\big] = \tilde{\mathsf{S}}_{\mathsf{uu}}$$

So the sign difference is not that big of a difference after all. But there are some key differences in practice:

- In the communications model (on the left), the "input signal" is $\mathsf{x}(n)$ and the frequency-domain input *signal-to-noise ratio* (*SNR*) is $\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)/\tilde{\mathsf{S}}_{\mathsf{uu}}(\omega)$. In the measurement model (on the right), the "input signal" is $\mathsf{p}(n)$ and the frequency-domain input *signal-to-noise ratio* (*SNR*) is $\tilde{\mathsf{S}}_{\mathsf{pp}}(\omega)/\tilde{\mathsf{S}}_{\mathsf{ww}}(\omega) = \tilde{\mathsf{S}}_{\mathsf{pp}}(\omega)/\tilde{\mathsf{S}}_{\mathsf{uu}}(\omega)$.
- On the left, x and u would be typically *uncorrelated*; on the right, x and $\mathsf{w} = -\mathsf{u}$ are very much *correlated* (x is a function of u).

14.3 Transfer function estimate definitions and interpretation

As a first attempt at estimating the transfer function **H** of **S**, or at least the magnitude squared of **H**, we might assume **H** to be *LTI*, take a cue from the relation $\tilde{\mathsf{S}}_{\mathsf{yy}} = \tilde{\mathsf{S}}_{\mathsf{xx}}\big|\tilde{\mathsf{H}}\big|^2$ of Corollary 7.3 (page 53), and arrive at a function called "*transmissibility*" (next definition).

Definition 14.1. [5] *Let* **S** *be a* SYSTEM *with input* $\mathsf{x}(n)$ *and output* $\mathsf{y}(n)$.

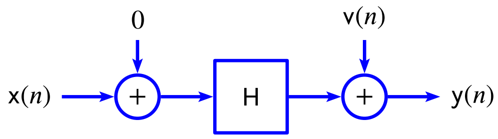
DEF: *The **transmissibility** function* $\tilde{\mathsf{T}}_{\mathsf{xy}}(\omega)$ *is defined as* $\quad \tilde{\mathsf{T}}_{\mathsf{xy}}(\omega) \triangleq \sqrt{\dfrac{\tilde{\mathsf{S}}_{\mathsf{yy}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)}}$

Transmissibility is in essence the ratio of "*spectral power*" (Remark 8.1 page 60) output to *spectral power* input. Note that it is a real-valued function (because $\tilde{\mathsf{S}}_{\mathsf{xx}}$ and $\tilde{\mathsf{S}}_{\mathsf{yy}}$ are real-valued). We might suspect that we could attain better estimates of **H** by allowing the estimates to be complex-valued. And in fact, all the remaining estimates in this section are in general complex-valued.

And so to start (again), and in the very special (a.k.a unrealistic) case of **S** having *zero measurement noise* (*zero measurement error*) ($\mathsf{v} = \mathsf{u} = \mathsf{w} = 0$), $\mathsf{h}(n)$ being *linear time invariant* (*LTI*), and input $\mathsf{x}(n)$ being *wide sense stationary*...then we can determine (a.k.a "identify") $\mathsf{h}(n)$ or $\tilde{\mathsf{H}}(\omega)$ exactly by $\tilde{\mathsf{H}}(\omega) = \tilde{\mathsf{S}}_{\mathsf{yx}}(\omega)/\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)$ (Corollary 7.3 page 53).

However, in practical situations, there is measurement noise/error. Examples may include "road noise" from a test being performed in a moving vehicle or *quantization noise* from an *analog-to-digital converter* (*ADC*).

If the measurement error is at the output only (and under the assumptions of *LTI* and *WSS*) then $\hat{\mathsf{H}}_1$ (next definition) is the ideal estimator in the sense that $\hat{\mathsf{H}}_1 = \tilde{\mathsf{H}}$ (Corollary 14.4 page 123).



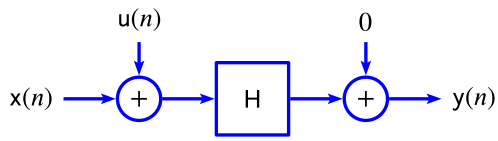
[5] Bendat and Piersol (2010) page 469 ⟨$|H(f)| = \big[G_{yy}(f)/G_{xx}(f)\big]^{1/2}$⟩, Yan and Ren (2012) page 204 ⟨(1) $\big[G_{YY}(s)\big] = [H(s)]\big[G_{FF}(s)\big]\big[H^*(s)\big]^T$⟩, Goldman (1999) page 179 ⟨Transmissibility ...$H'_{ab} = G_{bb}/G_{aa}$ (note: differs by $\sqrt{\cdot}$ from Bendat and Piersol⟩, Zhang et al. (2016), Zhou and Wahab (2018) page 824, https://link.springer.com/chapter/10.1007/978-3-319-54109-9_4

Definition 14.2. [6] *Let* $\mathbf{S}$ *be a* SYSTEM *with input* $\mathsf{x}(n)$ *and output* $\mathsf{y}(n)$.

DEF *The **Least Squares transfer function estimate*** $\hat{\mathsf{H}}_1(\omega)$ *of* $\mathbf{S}$ *is defined as* $\hat{\mathsf{H}}_1(\omega) \triangleq \dfrac{\tilde{\mathsf{S}}_{\mathsf{yx}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)}$

The estimator $\hat{\mathsf{H}}_1$ is a good start. However in the early 1980s, L. D. Mitchell pointed out that in the presence of input noise, $\hat{\mathsf{H}}_1$ is far from ideal in that it is *biased* with respect to $\tilde{\mathsf{H}}$; in fact, $\hat{\mathsf{H}}_1$ *under estimates* $\tilde{\mathsf{H}}$ (Corollary 14.4 page 123). Mitchell proposed a new estimator $\hat{\mathsf{H}}_2$ (next definition).

This estimator has the special property that when there is input noise but no output noise (and under *LTI*, *WSS*, and *uncorrelated* assumptions), then it is ideal in the sense that $\hat{\mathsf{H}}_2(\omega) = \tilde{\mathsf{H}}(\omega)$ (Corollary 14.4 page 123).



Note also that in the case of both no input and no output noise, then $\hat{\mathsf{H}}_1 = \hat{\mathsf{H}}_2$ (Corollary 7.3 page 53).

Definition 14.3. [7] *Let* $\mathbf{S}$ *be a* SYSTEM *with input* $\mathsf{x}(n)$ *and output* $\mathsf{y}(n)$.

DEF *The **Inverse Method transfer function estimate*** $\hat{\mathsf{H}}_2(\omega)$ *of* $\mathbf{S}$ *is defined as* $\hat{\mathsf{H}}_2(\omega) \triangleq \dfrac{\tilde{\mathsf{S}}_{\mathsf{yy}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{xy}}(\omega)}$

Mitchell's $\hat{\mathsf{H}}_2$ contribution "generated a flurry of activity"[8] and soon more $\tilde{\mathsf{H}}$ estimators appeared. So far we have

- $\hat{\mathsf{H}}_1$ which is ideal when there is no input noise but *under estimates* $\tilde{\mathsf{H}}$ when there is (Corollary 14.4 page 123)
- $\hat{\mathsf{H}}_2$ which is ideal when there is no output noise but *over estimates* $\tilde{\mathsf{H}}$ when there is (Corollary 14.4 page 123).

But what about estimators for when there is noise on both input and output? Armed with two estimators that between them account for both input and output noise, an "ad hoc" solution might be to somehow take mean values of $\hat{\mathsf{H}}_1$ and $\hat{\mathsf{H}}_2$ to induce new estimators—this approach summarizes the next three definitions. An arguably more mature approach is to find estimators that are optimal with respect to least squares measures—and this approach summarizes Definition 14.9 – Definition 14.7 (page 109).

Definition 14.4. *Let* $\mathbf{S}$ *be a* SYSTEM *with input* $\mathsf{x}(n)$ *and output* $\mathsf{y}(n)$.

DEF *The **Arithmetic Mean transfer function estimate*** $\hat{\mathsf{H}}_{\mathsf{am}}(\omega)$ *of* $\mathbf{S}$ *is defined as*

$$\hat{\mathsf{H}}_{\mathsf{am}}(\omega) \triangleq \frac{\left|\tilde{\mathsf{S}}_{\mathsf{xy}}(\omega)\right|^2 + \tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)\tilde{\mathsf{S}}_{\mathsf{yy}}(\omega)}{2\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)\tilde{\mathsf{S}}_{\mathsf{xy}}(\omega)}$$

[6] Bendat and Piersol (1993) pages 106–109 ⟨5.1.1 Optimality of Calculations⟩, Bendat and Piersol (2010) page 185 ⟨$H_1(f) = G_{xy}(f)/G_{xx}(f)$ (6.37)⟩, Shin and Hammond (2008) page 293 ⟨ $H_1(f) = \tilde{\mathsf{S}}_{\mathsf{xy}}(f)/\tilde{\mathsf{S}}_{\mathsf{xx}}(f)$ (9.63); which differs from Definition 14.2, but see Greenhoe (2019a) ⟩, Bendat (1978)cited by Cobb(1988)—variance estimate for $\hat{\mathsf{H}}_1$, Allemang et al. (1979) ⟨cited by Shin(2008)⟩, Leuridan et al. (1986) page 910 ⟨*Least Squares Technique*; (8) $[G_{xx}](H) = [G_{xy}]$⟩, Abom (1986)cited by Cobb(1988)—variance estimate for $\hat{\mathsf{H}}_1$, Allemang et al. (1987) pages 54–55 ⟨5.3.1 H_1 Technique; $[H] = [G_{XF}][G_{FF}]^{-1}$ (11)⟩, Cobb (1988) page 2 ⟨$^1\hat{H}(f) = \hat{G}_{yx}(f)/\hat{G}_{xx}(f)$ (1)⟩, Goyder (1984) page 438 ⟨$H(i\omega) = S_{qp}/S_{pp}$ (3)⟩, Pintelon and Schoukens (2012) page 233 ⟨$\hat{G}(\Omega_k) = S_{yu}(j\omega_k)S_{uu}^{-1}(j\omega_k)$ (7-30)⟩, White et al. (2006) page 678 ⟨$H_1(f) = \hat{S}_{x_m y_m}(f)/\hat{S}_{x_m x_m}(f)$ (1) which differs by conjugate, references Bendat and Piersol⟩,

[7] Shin and Hammond (2008) page 293 ⟨$H_2(f) = \tilde{\mathsf{S}}_{\mathsf{yy}}(f)/\tilde{\mathsf{S}}_{\mathsf{yx}}(f)$ (9.65); which differs from Definition 14.3, but see Greenhoe (2019a) ⟩, Bendat and Piersol (2010) page 186 ⟨$H_2(f) = G_{yy}(f)/G_{yx}(f)$ (6.42)⟩, Mitchell (1980) ⟨cited by Cobb(1988)⟩, Mitchell (1982) page 278 ⟨"Define what will be called an inverse method for calculation of a FRF as…"; $H_2(f) = G_{yy}/G_{yx}$ (6); Note this differs with Definition 14.3 by a conjugate, but note that Mitchell seems to follow Bendat (see his [3] and [4]), which would explain this difference⟩, Cobb (1988) page 3 ⟨$^2\hat{H}(f) = \hat{G}_{yy}(f)/\hat{G}_{xy}(f)$ (1)⟩, White et al. (2006) page 678 ⟨$H_2(f) = \hat{S}_{y_m y_m}(f)/\hat{S}_{y_m x_m}(f)$ (2) which differs by conjugate, references Bendat and Piersol⟩

[8] Cobb (1988) page 3

Proposition 14.1. [9] *Let* **S** *be a* SYSTEM *with input* x(n) *and output* y(n).

PRP

$$\hat{H}_{am}(\omega) = \frac{\hat{H}_1(\omega)+\hat{H}_2(\omega)}{2} \qquad \text{(\textbf{arithmetic mean} of } \hat{H}_1 \text{ and } \hat{H}_2\text{)}$$

✎PROOF:

$$\begin{aligned}
\hat{H}_{am}(\omega) &\triangleq \frac{\left|\tilde{S}_{xy}(\omega)\right|^2+\tilde{S}_{xx}(\omega)\tilde{S}_{yy}(\omega)}{2\tilde{S}_{xx}(\omega)\tilde{S}_{xy}(\omega)} && \text{by definition of } \hat{H}_{am} && \text{(Definition 14.4 page 106)}\\
&= \frac{\tilde{S}_{xy}(\omega)\tilde{S}^*_{xy}(\omega)}{2\tilde{S}_{xx}(\omega)\tilde{S}_{xy}(\omega)} + \frac{\tilde{S}_{xx}(\omega)\tilde{S}_{yy}(\omega)}{2\tilde{S}_{xx}(\omega)\tilde{S}_{xy}(\omega)} = \frac{\frac{\tilde{S}^*_{xy}(\omega)}{\tilde{S}_{xx}(\omega)}+\frac{\tilde{S}_{yy}(\omega)}{\tilde{S}_{xy}(\omega)}}{2}\\
&= \frac{\hat{H}_1(\omega)+\hat{H}_2(\omega)}{2} && \text{by definitions of } \hat{H}_1 \text{ and } \hat{H}_2 && \text{(Definition 14.2 page 106, Definition 14.3 page 106)}
\end{aligned}$$

✏

Definition 14.5. *Let* **S** *be a* SYSTEM *with input* x(n) *and output* y(n).

DEF

The ***Geometric mean transfer function estimate*** $\hat{H}_{gm}(\omega)$ *of* **S** *is defined as*

$$\hat{H}_{gm}(\omega) \triangleq \frac{\tilde{S}^*_{xy}(\omega)}{\left|\tilde{S}_{xy}(\omega)\right|}\sqrt{\frac{\tilde{S}_{yy}(\omega)}{\tilde{S}_{xx}(\omega)}}$$

Proposition 14.2. [10] *Let* **S** *be a* SYSTEM *with input* x(n) *and output* y(n).

PRP

$$\pm\hat{H}_{gm}(\omega) = \sqrt{\hat{H}_1(\omega)\hat{H}_2(\omega)} \qquad \text{(\textbf{geometric mean} of } \hat{H}_1 \text{ and } \hat{H}_2\text{)}$$

✎PROOF:

$$\begin{aligned}
\pm\hat{H}_{gm}(\omega) &\triangleq \pm\frac{\tilde{S}^*_{xy}(\omega)}{\left|\tilde{S}_{xy}(\omega)\right|}\sqrt{\frac{\tilde{S}_{yy}(\omega)}{\tilde{S}_{xx}(\omega)}} && \text{by definition of } \hat{H}_{gm} && \text{(Definition 14.5 page 107)}\\
&= \sqrt{\frac{\left[\tilde{S}^*_{xy}(\omega)\right]^2}{\left|\tilde{S}_{xy}(\omega)\right|^2}\frac{\tilde{S}_{yy}(\omega)}{\tilde{S}_{xx}(\omega)}} = \sqrt{\frac{\tilde{S}^*_{xy}(\omega)}{\tilde{S}_{xy}(\omega)}\frac{\tilde{S}_{yy}(\omega)}{\tilde{S}_{xx}(\omega)}} = \sqrt{\frac{\tilde{S}^*_{xy}(\omega)}{\tilde{S}_{xx}(\omega)}\frac{\tilde{S}_{yy}(\omega)}{\tilde{S}_{xy}(\omega)}}\\
&= \sqrt{\hat{H}_1(\omega)\hat{H}_2(\omega)} && \text{by definitions of } \hat{H}_1 \text{ and } \hat{H}_2 && \text{(Definition 14.2 page 106, Definition 14.3 page 106)}\\
&= \textbf{Geometric mean} \text{ of } \hat{H}_1(\omega) \text{ and } \hat{H}_2(\omega)
\end{aligned}$$

Note that for a complex number $z \triangleq |z|e^{i\phi}$, $\sqrt{z}$ has two solutions:[11]

$$\sqrt{z} = \sqrt{|z|e^{i\phi}} = \{z_1, z_2\} = \left\{\sqrt{|z|}e^{i(\phi/2)},\ \sqrt{|z|}e^{i(\phi/2+\pi)}\right\} = \pm\sqrt{|z|}e^{i(\phi/2)}$$

because $z_1^2 = z$ and $z_2^2 = z$. ✏

Note that the *geometric mean estimator* (Definition 14.5 page 107) and *transmissibility* (Definition 14.1 page 105) are closely related (next).

---

[9] Mitchell (1982) page 279 ⟨"Frequency Response Calculation: The Average Method"⟩, Zheng et al. (2002) page 918 ⟨"1.3 Arithmetic Mean Estimator H_3"⟩

[10] Zheng et al. (2002) page 918 ⟨"1.4 Geometric Mean Estimator H_4"⟩

[11] Many many thanks to Ben Cleveland for his help with this!!!

Proposition 14.3. *Let $\phi(\omega)$ be the* PHASE *of* $\tilde{\mathsf{S}}_{xy}(\omega)$ *such that* $\tilde{\mathsf{S}}_{xy}(\omega) \triangleq \left|\tilde{\mathsf{S}}_{xy}(\omega)\right| e^{-i\phi(\omega)}$

PRP
$$\hat{\mathsf{H}}_{gm}(\omega) = \tilde{\mathsf{T}}_{xy}(\omega)\, e^{-i\phi(\omega)} \quad \left(\begin{array}{lcll} \left|\hat{\mathsf{H}}_{gm}(\omega)\right| &=& \tilde{\mathsf{T}}_{xy}(\omega) & \text{is the MAGNITUDE of } \hat{\mathsf{H}}_{gm}(\omega) \quad \text{and} \\ \angle\hat{\mathsf{H}}_{gm}(\omega) &=& -\angle\tilde{\mathsf{S}}_{xy}(\omega) & \text{is the PHASE of } \hat{\mathsf{H}}_{gm}(\omega) \end{array}\right)$$

✎PROOF: Let $\phi(\omega)$ be the *phase* of

$$\begin{aligned}
\hat{\mathsf{H}}_{gm}(\omega) &\triangleq \sqrt{\hat{\mathsf{H}}_1(\omega)\hat{\mathsf{H}}_2(\omega)} && \text{by definition of } \hat{\mathsf{H}}_{gm} && \text{(Definition 14.5 page 107)} \\
&\triangleq \sqrt{\frac{\tilde{\mathsf{S}}^*_{xy}(\omega)}{\tilde{\mathsf{S}}_{xx}(\omega)}\frac{\tilde{\mathsf{S}}_{yy}(\omega)}{\tilde{\mathsf{S}}_{xy}(\omega)}} && \text{by definitions of } \hat{\mathsf{H}}_1 \text{ and } \hat{\mathsf{H}}_2 && \text{(Definition 14.2 page 106, Definition 14.3 page 106)} \\
&= \sqrt{\frac{\tilde{\mathsf{S}}_{yy}(\omega)}{\tilde{\mathsf{S}}_{xx}(\omega)}\frac{\tilde{\mathsf{S}}^*_{xy}(\omega)}{\tilde{\mathsf{S}}_{xy}(\omega)}} \\
&= \tilde{\mathsf{T}}_{xy}(\omega)\sqrt{\frac{\tilde{\mathsf{S}}^*_{xy}(\omega)}{\tilde{\mathsf{S}}_{xy}(\omega)}} && \text{by definition of } \tilde{\mathsf{T}}_{xy} && \text{(Definition 14.1 page 105)} \\
&= \tilde{\mathsf{T}}_{xy}(\omega)\sqrt{\frac{\left|\tilde{\mathsf{S}}_{xy}(\omega)\right|e^{-i\phi(\omega)}}{\left|\tilde{\mathsf{S}}_{xy}(\omega)\right|e^{i\phi(\omega)}}} && \text{where } \tilde{\mathsf{S}}_{xy}(\omega) \triangleq \left|\tilde{\mathsf{S}}_{xy}(\omega)\right|e^{-i\phi(\omega)} \\
&= \tilde{\mathsf{T}}_{xy}(\omega)\sqrt{e^{-i2\phi(\omega)}} \\
&= \tilde{\mathsf{T}}_{xy}(\omega)\, e^{-i\phi(\omega)}
\end{aligned}$$

⇨

Remark 14.1. *Transmissibility* $\tilde{\mathsf{T}}_{xy}$ is a kind of "black sheep" of the system identification function family. All the other members of this family ($\hat{\mathsf{H}}_1$, $\hat{\mathsf{H}}_2$, $\hat{\mathsf{H}}_v$, $\hat{\mathsf{H}}_s$) are *complex-valued*, but $\tilde{\mathsf{T}}_{xy}$ is only *real-valued*—a seemingly ordinary Joe born into a super-hero family. But Proposition 14.3 suggests that $\tilde{\mathsf{T}}_{xy}$ is not simply a "black sheep", but rather a "dark horse" with abilities that can easily be unleashed by slight redefinition. In particular, Proposition 14.3 demonstrates that $\tilde{\mathsf{T}}_{xy}$ is the *magnitude* of the geometric mean of $\hat{\mathsf{H}}_1$ and $\hat{\mathsf{H}}_2$. We can thus justifiably define a **complex transmissibility** function as $\hat{\mathsf{H}}_{gm}$...and the magnitude of this *complex transmissibility* function is the *ordinary transmissibility* function of Definition 14.1 (page 105).

REM
The **complex transmissibility** function $\tilde{\mathsf{T}}'_{xy}(\omega) \triangleq \hat{\mathsf{H}}_{gm}(\omega)$

Definition 14.6. *Let* $\mathbf{S}$ *be a* SYSTEM *with input* $\mathsf{x}(n)$ *and output* $\mathsf{y}(n)$.

DEF
The ***Harmonic mean transfer function estimate*** $\hat{\mathsf{H}}_{hm}(\omega)$ *of* $\mathbf{S}$ *is defined as*
$$\hat{\mathsf{H}}_{hm}(\omega) \triangleq \frac{2\tilde{\mathsf{S}}_{yy}(\omega)\tilde{\mathsf{S}}^*_{xy}(\omega)}{\tilde{\mathsf{S}}_{xx}(\omega)\tilde{\mathsf{S}}_{yy}(\omega) + \left|\tilde{\mathsf{S}}_{xy}(\omega)\right|^2}$$

Proposition 14.4. [12] *Let* $\mathbf{S}$ *be a* SYSTEM *with input* $\mathsf{x}(n)$ *and output* $\mathsf{y}(n)$.

PRP
$$\hat{\mathsf{H}}_{hm}(\omega) = \frac{2}{\frac{1}{\hat{\mathsf{H}}_1(\omega)} + \frac{1}{\hat{\mathsf{H}}_2(\omega)}} \quad \text{(\textbf{Harmonic mean} of } \hat{\mathsf{H}}_1 \text{ and } \hat{\mathsf{H}}_2\text{)}$$

[12] Carne and Dohrmann (2006) $\langle H_C = [H_A^{-1} + H_B^{-1}]^{-1}\rangle$

✎PROOF:

$$\begin{aligned}
\hat{\mathsf{H}}_{hm}(\omega) &\triangleq \frac{2\tilde{\mathsf{S}}_{yy}(\omega)\tilde{\mathsf{S}}^*_{xy}(\omega)}{\tilde{\mathsf{S}}_{xx}(\omega)\tilde{\mathsf{S}}_{yy}(\omega)+\left|\tilde{\mathsf{S}}_{xy}(\omega)\right|^2} && \text{by definition of } \hat{\mathsf{H}}_{hm} && \text{(Definition 14.6 page 108)}\\
&= \frac{2}{\frac{\tilde{\mathsf{S}}_{xx}(\omega)\tilde{\mathsf{S}}_{yy}(\omega)+\left|\tilde{\mathsf{S}}_{xy}(\omega)\right|^2}{\tilde{\mathsf{S}}_{yy}(\omega)\tilde{\mathsf{S}}^*_{xy}(\omega)}} = \frac{2}{\frac{\tilde{\mathsf{S}}_{xx}(\omega)}{\tilde{\mathsf{S}}^*_{xy}(\omega)}+\frac{\tilde{\mathsf{S}}_{xy}(\omega)}{\tilde{\mathsf{S}}_{yy}(\omega)}}\\
&= \frac{2}{\frac{1}{\hat{\mathsf{H}}_1(\omega)}+\frac{1}{\hat{\mathsf{H}}_2(\omega)}} && \text{by definitions of } \hat{\mathsf{H}}_1 \text{ and } \hat{\mathsf{H}}_2 && \text{(Definition 14.2 page 106, Definition 14.3 page 106)}\\
&= \textbf{Harmonic mean} \text{ of } \hat{\mathsf{H}}_1(\omega) \text{ and } \hat{\mathsf{H}}_2(\omega)
\end{aligned}$$

✏

A bit of review reveals $\hat{\mathsf{H}}_1$ at the low end of the estimation problem, $\hat{\mathsf{H}}_2$ at the high end, and $\hat{\mathsf{H}}_{hm}$, $\hat{\mathsf{H}}_{gm}$, and $\hat{\mathsf{H}}_{am}$ somewhere between. But these three "between" estimates are not shown to be optimal in any sense—they are just conceptually interesting. What we might really like is a family of estimators that

- include $\hat{\mathsf{H}}_1$ and $\hat{\mathsf{H}}_2$ as limiting cases
- include the between cases
- are optimal in some sense

The estimator $\hat{\mathsf{H}}_\kappa(\omega;\kappa)$ is one such estimator (next definition) that

- has $\hat{\mathsf{H}}_1$ and $\hat{\mathsf{H}}_2$ as limiting cases (Theorem 14.1 page 111),
- is optimal in the least squares sense (Theorem 14.6 page 124), and
- allows for a system designer to specify an output-input spectral noise ratio $\kappa(\omega)$ that can vary with frequency ω.

Moreover, $\hat{\mathsf{H}}_\kappa(\omega)$ includes some special cases:

- In the case of constant κ, $\hat{\mathsf{H}}_\kappa$ simplifies to the *Scaling transfer function estimate* $\hat{\mathsf{H}}_s$ (Definition 14.8 page 109).
- In the case of $\kappa = 1$, $\hat{\mathsf{H}}_\kappa$ and $\hat{\mathsf{H}}_s$ simplify to the *Total least squares transfer function estimate* $\hat{\mathsf{H}}_v$ (Definition 14.9 page 110).

Definition 14.7. [13] *Let* **S** *be a* SYSTEM *with input* $\mathsf{x}(n)$ *and output* $\mathsf{y}(n)$.

DEF *The **transfer function estimate** $\hat{\mathsf{H}}_\kappa(\omega;\kappa)$ with **scaling function** $\kappa(\omega)$ is defined as*

$$\hat{\mathsf{H}}_\kappa(\omega;\kappa) \triangleq \frac{\tilde{\mathsf{S}}_{yy}(\omega)-\kappa(\omega)\tilde{\mathsf{S}}_{xx}(\omega)+\sqrt{\left[\tilde{\mathsf{S}}_{yy}(\omega)-\kappa(\omega)\tilde{\mathsf{S}}_{xx}(\omega)\right]^2+4\kappa(\omega)\left|\tilde{\mathsf{S}}_{xy}(\omega)\right|^2}}{2\tilde{\mathsf{S}}_{xy}(\omega)}$$

Definition 14.8. [14] *Let* **S** *be a* SYSTEM *with input* $\mathsf{x}(n)$ *and output* $\mathsf{y}(n)$.

DEF *The **Scaling transfer function estimate** $\hat{\mathsf{H}}_s(\omega;s)$ of* **S** *with **scaling parameter** $s\in[0:\infty)$ is defined as* $\quad\hat{\mathsf{H}}_s(\omega;s)\triangleq\hat{\mathsf{H}}_\kappa(\omega;\kappa)\quad$ *with* $\kappa(\omega)\triangleq s^2$

[13] White et al. (2006) page 679 ⟨(6)⟩, Shin and Hammond (2008) page 293 ⟨(9.67)⟩

[14] Shin and Hammond (2008) page 293 ⟨(9.67) with $\kappa(\omega)=s^2$⟩, White et al. (2006) page 679 ⟨(6) with $\kappa(\omega)=s^2$⟩, Leclere et al. (2014) ⟨(10) $\kappa(f)=1/s^2$ and x and y swapped⟩, Wicks and Vold (1986) page 898 ⟨has additional s in denominator⟩, Zheng et al. (2002) page 918 ⟨(10), seems to differ⟩

Definition 14.9. [15] *Let* $\mathbf{S}$ *be a* SYSTEM *with input* $\mathsf{x}(n)$ *and output* $\mathsf{y}(n)$.

DEF: *The* ***Total Least Squares transfer function estimate*** $\hat{\mathsf{H}}_{\mathsf{v}}(\omega)$ *of* $\mathbf{S}$ *is defined as*

$$\hat{\mathsf{H}}_{\mathsf{v}}(\omega) \triangleq \hat{\mathsf{H}}_{\kappa}(\omega;\kappa) \qquad \text{with } \kappa(\omega) = 1$$

The previous estimators all assumed two signals: an input $\mathsf{x}(n)$ and an output $\mathsf{y}(n)$. However, in many practical systems, there is a third signal that is "driving" the system. In 1984 Goyder proposed an estimator (next definition) that is based on three signals.

Definition 14.10 (Three channel estimate)**.** [16] *Let* $\mathbf{S}$ *be a system with input* $\mathsf{x}(n)$, *output* $\mathsf{y}(n)$, *and a driver* $\mathsf{p}(n)$.

DEF: *The* ***transfer function estimate*** $\hat{\mathsf{H}}_{\mathsf{c}}(\omega)$ *is defined as* $\displaystyle \hat{\mathsf{H}}_{\mathsf{c}}(\omega) \triangleq \frac{\tilde{\mathsf{S}}_{\mathsf{py}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{px}}(\omega)}$

14.4 Estimator relationships

Lemma 14.1.

LEM:

$$\frac{\mathsf{d}}{\mathsf{dp}}\left[\tilde{\mathsf{S}}_{\mathsf{yy}} - p\tilde{\mathsf{S}}_{\mathsf{xx}} + \sqrt{\left(\tilde{\mathsf{S}}_{\mathsf{yy}} - p\tilde{\mathsf{S}}_{\mathsf{xx}}\right)^2 + 4p\left|\tilde{\mathsf{S}}_{\mathsf{xy}}\right|^2}\right] = \frac{\tilde{\mathsf{S}}_{\mathsf{xx}}\left(p\tilde{\mathsf{S}}_{\mathsf{xx}} - \tilde{\mathsf{S}}_{\mathsf{yy}}\right) + 2\left|\tilde{\mathsf{S}}_{\mathsf{xy}}\right|^2 - \tilde{\mathsf{S}}_{\mathsf{xx}}\sqrt{\left(p\tilde{\mathsf{S}}_{\mathsf{xx}} - \tilde{\mathsf{S}}_{\mathsf{yy}}\right)^2 + 4p\left|\tilde{\mathsf{S}}_{\mathsf{xy}}\right|^2}}{\sqrt{\left(p\tilde{\mathsf{S}}_{\mathsf{xx}} - \tilde{\mathsf{S}}_{\mathsf{yy}}\right)^2 + 4p\left|\tilde{\mathsf{S}}_{\mathsf{xy}}\right|^2}}$$

$$\frac{\mathsf{d}}{\mathsf{dp}}\left[p\tilde{\mathsf{S}}_{\mathsf{yy}} - \tilde{\mathsf{S}}_{\mathsf{xx}} + \sqrt{\left(p\tilde{\mathsf{S}}_{\mathsf{yy}} - \tilde{\mathsf{S}}_{\mathsf{xx}}\right)^2 + 4p\left|\tilde{\mathsf{S}}_{\mathsf{xy}}\right|^2}\right] = \frac{\tilde{\mathsf{S}}_{\mathsf{yy}}\left(p\tilde{\mathsf{S}}_{\mathsf{yy}} - \tilde{\mathsf{S}}_{\mathsf{xx}}\right) + 2\left|\tilde{\mathsf{S}}_{\mathsf{xy}}\right|^2 + \tilde{\mathsf{S}}_{\mathsf{yy}}\sqrt{\left(p\tilde{\mathsf{S}}_{\mathsf{yy}} - \tilde{\mathsf{S}}_{\mathsf{xx}}\right)^2 + 4p\left|\tilde{\mathsf{S}}_{\mathsf{xy}}\right|^2}}{\sqrt{\left(p\tilde{\mathsf{S}}_{\mathsf{yy}} - \tilde{\mathsf{S}}_{\mathsf{xx}}\right)^2 + 4p\left|\tilde{\mathsf{S}}_{\mathsf{xy}}\right|^2}}$$

✎PROOF:

$$\begin{aligned}
\frac{\mathsf{d}}{\mathsf{dp}}\left[\tilde{\mathsf{S}}_{\mathsf{yy}} - p\tilde{\mathsf{S}}_{\mathsf{xx}} + \sqrt{\left(\tilde{\mathsf{S}}_{\mathsf{yy}} - p\tilde{\mathsf{S}}_{\mathsf{xx}}\right)^2 + 4p\left|\tilde{\mathsf{S}}_{\mathsf{xy}}\right|^2}\right] &= -\tilde{\mathsf{S}}_{\mathsf{xx}} + \frac{-2\tilde{\mathsf{S}}_{\mathsf{xx}}\left(\tilde{\mathsf{S}}_{\mathsf{yy}} - p\tilde{\mathsf{S}}_{\mathsf{xx}}\right) + 4\left|\tilde{\mathsf{S}}_{\mathsf{xy}}\right|^2}{2\sqrt{\left(\tilde{\mathsf{S}}_{\mathsf{yy}} - p\tilde{\mathsf{S}}_{\mathsf{xx}}\right)^2 + 4p\left|\tilde{\mathsf{S}}_{\mathsf{xy}}\right|^2}} \\
&= \frac{4\left|\tilde{\mathsf{S}}_{\mathsf{xy}}\right|^2 - 2\tilde{\mathsf{S}}_{\mathsf{xx}}\left(\tilde{\mathsf{S}}_{\mathsf{yy}} - p\tilde{\mathsf{S}}_{\mathsf{xx}}\right) - 2\tilde{\mathsf{S}}_{\mathsf{xx}}\sqrt{\left(\tilde{\mathsf{S}}_{\mathsf{yy}} - p\tilde{\mathsf{S}}_{\mathsf{xx}}\right)^2 + 4p\left|\tilde{\mathsf{S}}_{\mathsf{xy}}\right|^2}}{2\sqrt{\left(\tilde{\mathsf{S}}_{\mathsf{yy}} - p\tilde{\mathsf{S}}_{\mathsf{xx}}\right)^2 + 4p\left|\tilde{\mathsf{S}}_{\mathsf{xy}}\right|^2}} \\
&= \frac{\tilde{\mathsf{S}}_{\mathsf{xx}}\left(p\tilde{\mathsf{S}}_{\mathsf{xx}} - \tilde{\mathsf{S}}_{\mathsf{yy}}\right) + 2\left|\tilde{\mathsf{S}}_{\mathsf{xy}}\right|^2 - \tilde{\mathsf{S}}_{\mathsf{xx}}\sqrt{\left(p\tilde{\mathsf{S}}_{\mathsf{xx}} - \tilde{\mathsf{S}}_{\mathsf{yy}}\right)^2 + 4p\left|\tilde{\mathsf{S}}_{\mathsf{xy}}\right|^2}}{\sqrt{\left(p\tilde{\mathsf{S}}_{\mathsf{xx}} - \tilde{\mathsf{S}}_{\mathsf{yy}}\right)^2 + 4p\left|\tilde{\mathsf{S}}_{\mathsf{xy}}\right|^2}}
\end{aligned}$$

$$\begin{aligned}
\frac{\mathsf{d}}{\mathsf{dp}}\left[p\tilde{\mathsf{S}}_{\mathsf{yy}} - \tilde{\mathsf{S}}_{\mathsf{xx}} + \sqrt{\left(p\tilde{\mathsf{S}}_{\mathsf{yy}} - \tilde{\mathsf{S}}_{\mathsf{xx}}\right)^2 + 4p\left|\tilde{\mathsf{S}}_{\mathsf{xy}}\right|^2}\right] &= \tilde{\mathsf{S}}_{\mathsf{yy}} + \frac{2\tilde{\mathsf{S}}_{\mathsf{yy}}\left(p\tilde{\mathsf{S}}_{\mathsf{yy}} - \tilde{\mathsf{S}}_{\mathsf{xx}}\right) + 4\left|\tilde{\mathsf{S}}_{\mathsf{xy}}\right|^2}{2\sqrt{\left(p\tilde{\mathsf{S}}_{\mathsf{yy}} - \tilde{\mathsf{S}}_{\mathsf{xx}}\right)^2 + 4p\left|\tilde{\mathsf{S}}_{\mathsf{xy}}\right|^2}} \\
&= \frac{4\left|\tilde{\mathsf{S}}_{\mathsf{xy}}\right|^2 + 2\tilde{\mathsf{S}}_{\mathsf{yy}}\left(p\tilde{\mathsf{S}}_{\mathsf{yy}} - \tilde{\mathsf{S}}_{\mathsf{xx}}\right) + 2\tilde{\mathsf{S}}_{\mathsf{yy}}\sqrt{\left(p\tilde{\mathsf{S}}_{\mathsf{yy}} - \tilde{\mathsf{S}}_{\mathsf{xx}}\right)^2 + 4p\left|\tilde{\mathsf{S}}_{\mathsf{xy}}\right|^2}}{2\sqrt{\left(p\tilde{\mathsf{S}}_{\mathsf{yy}} - \tilde{\mathsf{S}}_{\mathsf{xx}}\right)^2 + 4p\left|\tilde{\mathsf{S}}_{\mathsf{xy}}\right|^2}}
\end{aligned}$$

[15] White et al. (2006) page 679 ⟨(6)⟩, Shin and Hammond (2008) page 294 ⟨(9.69)⟩

[16] Shin and Hammond (2008) page 297 ⟨$H_3(f) = S_{ry}(f)/S_{rx}(f)$ (9.78)⟩, Cobb (1988) page 4 ⟨${}^{c}\hat{H}(f) = \hat{G}_{ys}(f)/\hat{G}_{xs}(f)$ (1.4)⟩, Goyder (1984) page 440 ⟨$H(i\omega) = S_{qz}/S_{pz}$ (5)⟩, Cobb and Mitchell (1990) page 450 ⟨(1)⟩, Pintelon and Schoukens (2012) page 241 ⟨$\hat{G}(\Omega_k) = \hat{G}_{ry}(\Omega_k)\hat{G}_{ru}^{-1}(\Omega_k)$ (7-49)⟩

$$= \frac{\tilde{S}_{yy}\left(p\tilde{S}_{yy} - \tilde{S}_{xx}\right) + 2\left|\tilde{S}_{xy}\right|^2 + \tilde{S}_{yy}\sqrt{\left(p\tilde{S}_{yy} - \tilde{S}_{xx}\right)^2 + 4p\left|\tilde{S}_{xy}\right|^2}}{\sqrt{\left(p\tilde{S}_{yy} - \tilde{S}_{xx}\right)^2 + 4p\left|\tilde{S}_{xy}\right|^2}}$$

✏

Lemma 14.2.

LEM

$$\begin{aligned} \tilde{S}_{yy} - p\tilde{S}_{xx} + \sqrt{\left(\tilde{S}_{yy} - p\tilde{S}_{xx}\right)^2 + 4p\left|\tilde{S}_{xy}\right|^2} &\geq 0 \\ p\tilde{S}_{yy} - \tilde{S}_{xx} + \sqrt{\left(p\tilde{S}_{yy} - \tilde{S}_{xx}\right)^2 + 4p\left|\tilde{S}_{xy}\right|^2} &\geq 0 \end{aligned}$$

✎PROOF:

$$\begin{aligned} & \tilde{S}_{yy} - p\tilde{S}_{xx} + \sqrt{\left(\tilde{S}_{yy} - p\tilde{S}_{xx}\right)^2 + 4p\left|\tilde{S}_{xy}\right|^2} && \geq 0 \\ \iff & \sqrt{\left(p\tilde{S}_{xx} - \tilde{S}_{yy}\right)^2 + 4p\left|\tilde{S}_{xy}\right|^2} && \geq p\tilde{S}_{xx} - \tilde{S}_{yy} \\ \iff & \left(p\tilde{S}_{xx} - \tilde{S}_{yy}\right)^2 + 4p\left|\tilde{S}_{xy}\right|^2 && \geq \left(p\tilde{S}_{xx} - \tilde{S}_{yy}\right)^2 \\ \iff & 4p\left|\tilde{S}_{xy}\right|^2 && \geq 0 \\ \iff & \left|\tilde{S}_{xy}\right| && \geq 0 \end{aligned}$$

$$\begin{aligned} & p\tilde{S}_{yy} - \tilde{S}_{xx} + \sqrt{\left(p\tilde{S}_{yy} - \tilde{S}_{xx}\right)^2 + 4p\left|\tilde{S}_{xy}\right|^2} && \geq 0 \\ \iff & \sqrt{\left(\tilde{S}_{xx} - p\tilde{S}_{yy}\right)^2 + 4p\left|\tilde{S}_{xy}\right|^2} && \geq \tilde{S}_{xx} - p\tilde{S}_{yy} \\ \iff & \left(\tilde{S}_{xx} - p\tilde{S}_{yy}\right)^2 + 4p\left|\tilde{S}_{xy}\right|^2 && \geq \left(\tilde{S}_{xx} - p\tilde{S}_{yy}\right)^2 \\ \iff & 4p\left|\tilde{S}_{xy}\right|^2 && \geq 0 \\ \iff & \left|\tilde{S}_{xy}\right| && \geq 0 \end{aligned}$$

✏

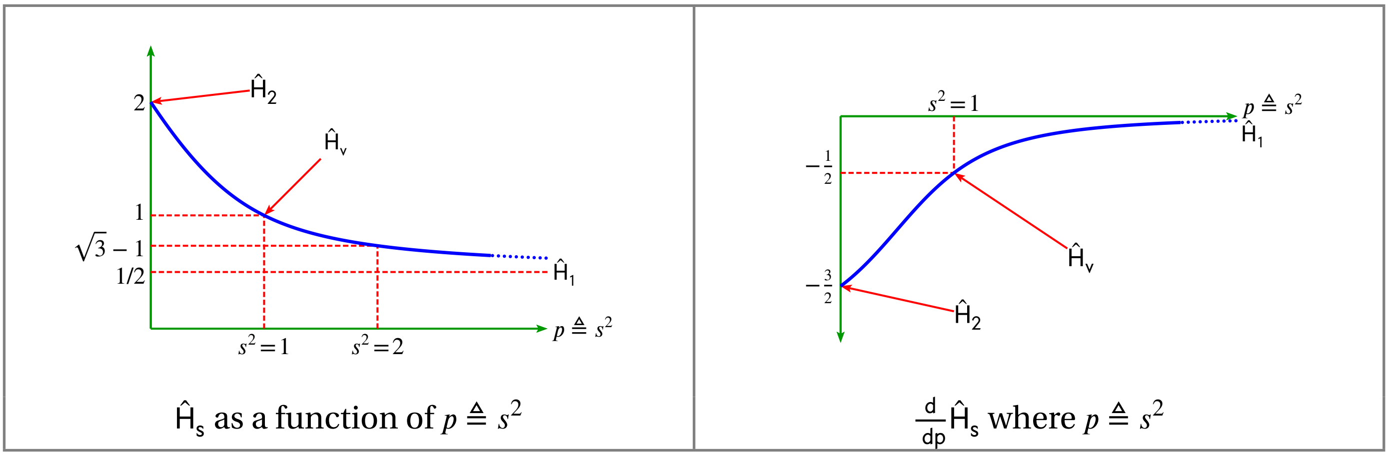


Figure 14.2: $\hat{H}_s$ with $\tilde{S}_{xx} = \tilde{S}_{yy} = 1$ and $\tilde{S}_{xy} = \frac{1}{2}$

Theorem 14.1. *Let $\hat{H}_s$ be defined as in Definition 14.8 (page 109).*

THM

$$\begin{array}{llll} \{s_1 < s_2\} & \implies & \left|\hat{H}_s(\omega; s_2)\right| \leq \hat{H}_s(\omega; s_1) & (\hat{H}_s \textit{ is } \textbf{monotonically decreasing} \textit{ in } s) \\ \left|\hat{H}_1(\omega)\right| & \leq \left|\hat{H}_s(\omega; s)\right| \leq & \left|\hat{H}_2(\omega)\right| & \\ \hat{H}_s(\omega; s)\Big|_{s=0} & = & \hat{H}_2(\omega) & \\ \hat{H}_s(\omega; s)\Big|_{s=1} & = & \hat{H}_v(\omega) & \\ \lim\limits_{s\to\infty} \hat{H}_s(\omega; s) & = & \hat{H}_1(\omega) & \end{array}$$

✎PROOF: I. Proofs for equalities:

$$\begin{aligned}
\hat{\mathsf{H}}_s(\omega;s)\big|_{s=0} &\triangleq \left.\frac{\tilde{S}_{yy} - s^2\tilde{S}_{xx} + \sqrt{\left[\tilde{S}_{yy} - s^2\tilde{S}_{xx}\right]^2 + 4s^2\left|\tilde{S}_{xy}\right|^2}}{2\tilde{S}_{xy}}\right|_{s=0} && \text{by def. of } \hat{\mathsf{H}}_s && \text{(Definition 14.8 page 109)}\\
&= \frac{\tilde{S}_{yy} - 0 + \sqrt{\left[\tilde{S}_{yy} - 0\right]^2 + 0}}{2\tilde{S}_{xy}} = \frac{\tilde{S}_{yy}}{\tilde{S}_{xy}}\\
&\triangleq \hat{\mathsf{H}}_2 && \text{by def. of } \hat{\mathsf{H}}_2 && \text{(Definition 14.3 page 106)}\\
\hat{\mathsf{H}}_s(\omega;s)\big|_{s=1} &\triangleq \left.\frac{\tilde{S}_{yy} - s^2\tilde{S}_{xx} + \sqrt{\left[\tilde{S}_{yy} - s^2\tilde{S}_{xx}\right]^2 + 4s^2\left|\tilde{S}_{xy}\right|^2}}{2\tilde{S}_{xy}}\right|_{s=1} && \text{by def. of } \hat{\mathsf{H}}_s && \text{(Definition 14.8 page 109)}\\
&= \frac{\tilde{S}_{yy} - \tilde{S}_{xx} + \sqrt{\left[\tilde{S}_{yy} - \tilde{S}_{xx}\right]^2 + 4\left|\tilde{S}_{xy}\right|^2}}{2\tilde{S}_{xy}}\\
&\triangleq \hat{\mathsf{H}}_v && \text{by def. of } \hat{\mathsf{H}}_v && \text{(Definition 14.9 page 110)}\\
\lim_{s\to\infty}\hat{\mathsf{H}}_s(\omega;s) &\triangleq \lim_{s\to\infty}\frac{\tilde{S}_{yy} - s^2\tilde{S}_{xx} + \sqrt{\left[\tilde{S}_{yy} - s^2\tilde{S}_{xx}\right]^2 + 4s^2\left|\tilde{S}_{xy}\right|^2}}{2\tilde{S}_{xy}} && \text{by def. of } \hat{\mathsf{H}}_s && \text{(Definition 14.8 page 109)}\\
&\triangleq \lim_{p\to 0}\frac{\tilde{S}_{yy} - {}^1\!/_p\tilde{S}_{xx} + \sqrt{\left[\tilde{S}_{yy} - {}^1\!/_p\tilde{S}_{xx}\right]^2 + 4\,{}^1\!/_p\left|\tilde{S}_{xy}\right|^2}}{2\tilde{S}_{xy}} && \text{where } p \triangleq \frac{1}{s^2}\\
&= \lim_{p\to 0}\frac{p\tilde{S}_{yy} - \tilde{S}_{xx} + \sqrt{\left[p\tilde{S}_{yy} - \tilde{S}_{xx}\right]^2 + 4p\left|\tilde{S}_{xy}\right|^2}}{2p\tilde{S}_{xy}} && \text{by mult. by } 1 = \frac{p}{p}\\
&= \lim_{p\to 0}\frac{\frac{\mathrm{d}}{\mathrm{dp}}\left[p\tilde{S}_{yy} - \tilde{S}_{xx} + \sqrt{\left[p\tilde{S}_{yy} - \tilde{S}_{xx}\right]^2 + 4p\left|\tilde{S}_{xy}\right|^2}\right]}{\frac{\mathrm{d}}{\mathrm{dp}}\left[2p\tilde{S}_{xy}\right]} && \text{by \textit{l'Hôpital's rule}}\\
&= \lim_{p\to 0}\frac{\dfrac{\tilde{S}_{yy}\left(p\tilde{S}_{yy} - \tilde{S}_{xx}\right) + 2\left|\tilde{S}_{xy}\right|^2 + \tilde{S}_{yy}\sqrt{\left(p\tilde{S}_{yy} - \tilde{S}_{xx}\right)^2 + 4p\left|\tilde{S}_{xy}\right|^2}}{\sqrt{\left(p\tilde{S}_{yy} - \tilde{S}_{xx}\right)^2 + 4p\left|\tilde{S}_{xy}\right|^2}}}{2\tilde{S}_{xy}} && \text{by Lemma 14.1 page 110}\\
&= \frac{\tilde{S}_{yy}\left(-\tilde{S}_{xx}\right) + 2\left|\tilde{S}_{xy}\right|^2 + \tilde{S}_{yy}\sqrt{\left(-\tilde{S}_{xx}\right)^2}}{2\tilde{S}_{xy}\sqrt{\left(-\tilde{S}_{xx}\right)^2}}\\
&= \frac{2\left|\tilde{S}_{xy}\right|^2}{2\tilde{S}_{xx}\tilde{S}_{xy}} = \frac{\tilde{S}^*_{xy}}{\tilde{S}_{xx}} \triangleq \hat{\mathsf{H}}_1 && \text{by def. of } \hat{\mathsf{H}}_1 && \text{(Definition 14.2 page 106)}
\end{aligned}$$

II. Proof for monoticity:

1. Let $p \triangleq s^2$
2. lemma:

$$\begin{aligned}
&\boxed{\left[2|\tilde{S}_{xy}|^2 + \tilde{S}_{xx}(p\tilde{S}_{xx} - \tilde{S}_{yy})\right]^2} \\
&= 4|\tilde{S}_{xy}|^4 + 4\tilde{S}_{xx}|\tilde{S}_{xy}|^2(p\tilde{S}_{xx} - \tilde{S}_{yy}) + \tilde{S}_{xx}^2(p\tilde{S}_{xx} - \tilde{S}_{yy})^2 \\
&\boxed{\le} 4|\tilde{S}_{xy}|^2\tilde{S}_{xx}\tilde{S}_{yy} + 4p\tilde{S}_{xx}|\tilde{S}_{xy}|^2\tilde{S}_{xx} - 4\tilde{S}_{xx}|\tilde{S}_{xy}|^2\tilde{S}_{yy} + \tilde{S}_{xx}^2(p\tilde{S}_{xx} - \tilde{S}_{yy})^2 && \left(\text{by } \textit{Cauchy Schwartz Inequality} \text{ (Theorem N.2 page 310)}\right) \\
&= \cancelto{0}{4\tilde{S}_{xx}\tilde{S}_{yy}|\tilde{S}_{xy}|^2} + 4p\tilde{S}_{xx}|\tilde{S}_{xy}|^2\tilde{S}_{xx} - \cancelto{0}{4\tilde{S}_{xx}\tilde{S}_{yy}|\tilde{S}_{xy}|^2} + \tilde{S}_{xx}^2(p\tilde{S}_{xx} - \tilde{S}_{yy})^2 \\
&= \tilde{S}_{xx}^2\left[(p\tilde{S}_{xx} - \tilde{S}_{yy})^2 + 4p|\tilde{S}_{xy}|^2\right] \\
&= \boxed{\left[\tilde{S}_{xx}\sqrt{(p\tilde{S}_{xx} - \tilde{S}_{yy})^2 + 4p|\tilde{S}_{xy}|^2}\right]^2}
\end{aligned}$$

3. lemma: $2|\tilde{S}_{xy}|^2 + \tilde{S}_{xx}(p\tilde{S}_{xx} - \tilde{S}_{yy}) \le \tilde{S}_{xx}\sqrt{(p\tilde{S}_{xx} - \tilde{S}_{yy})^2 + 4p|\tilde{S}_{xy}|^2}$. Proof:

$$\begin{aligned}
&2|\tilde{S}_{xy}|^2 + \tilde{S}_{xx}(p\tilde{S}_{xx} - \tilde{S}_{yy}) \le \tilde{S}_{xx}\sqrt{(p\tilde{S}_{xx} - \tilde{S}_{yy})^2 + 4p|\tilde{S}_{xy}|^2} \\
\iff &\left[2|\tilde{S}_{xy}|^2 + \tilde{S}_{xx}(p\tilde{S}_{xx} - \tilde{S}_{yy})\right]^2 \le \left[\tilde{S}_{xx}\sqrt{(p\tilde{S}_{xx} - \tilde{S}_{yy})^2 + 4p|\tilde{S}_{xy}|^2}\right]^2 && \left(\text{because } \mathsf{f}(x) \triangleq x^2 \text{ is } \textit{strictly monotonic increasing}\right)
\end{aligned}$$

The previous inequality is true by (2) lemma, so (3) lemma also true.

4. Proof that $\frac{\mathsf{d}}{\mathsf{d}p}|\hat{\mathsf{H}}_s| \le 0$:

$$\begin{aligned}
\frac{\mathsf{d}}{\mathsf{d}p}|\hat{\mathsf{H}}_s| &\triangleq \frac{\mathsf{d}}{\mathsf{d}p}\left|\frac{\tilde{S}_{yy} - s^2\tilde{S}_{xx} + \sqrt{(\tilde{S}_{yy} - s^2\tilde{S}_{xx})^2 + 4s^2|\tilde{S}_{xy}|^2}}{2\tilde{S}_{xy}}\right| && \text{by def. of } \hat{\mathsf{H}}_s \text{ (Definition 14.8 page 109)} \\
&\triangleq \frac{\mathsf{d}}{\mathsf{d}p}\left|\frac{\tilde{S}_{yy} - p\tilde{S}_{xx} + \sqrt{(\tilde{S}_{yy} - p\tilde{S}_{xx})^2 + 4p|\tilde{S}_{xy}|^2}}{2\tilde{S}_{xy}}\right| && \text{by definition of } p \text{ (item (1) page 113)} \\
&= \frac{\mathsf{d}}{\mathsf{d}p}\left(\frac{\left|\tilde{S}_{yy} - p\tilde{S}_{xx} + \sqrt{(\tilde{S}_{yy} - p\tilde{S}_{xx})^2 + 4p|\tilde{S}_{xy}|^2}\right|}{|2\tilde{S}_{xy}|}\right) \\
&= \frac{\mathsf{d}}{\mathsf{d}p}\left(\frac{\tilde{S}_{yy} - p\tilde{S}_{xx} + \sqrt{(\tilde{S}_{yy} - p\tilde{S}_{xx})^2 + 4p|\tilde{S}_{xy}|^2}}{2|\tilde{S}_{xy}|}\right) && \text{by Lemma 14.2 page 111} \\
&= \frac{\dfrac{\tilde{S}_{xx}(p\tilde{S}_{xx} - \tilde{S}_{yy}) + 2|\tilde{S}_{xy}|^2 - \tilde{S}_{xx}\sqrt{(p\tilde{S}_{xx} - \tilde{S}_{yy})^2 + 4p|\tilde{S}_{xy}|^2}}{\sqrt{(\tilde{S}_{yy} - p\tilde{S}_{xx})^2 + 4p|\tilde{S}_{xy}|^2}}}{2|\tilde{S}_{xy}|} && \text{by Lemma 14.1 page 110} \\
&= \frac{\tilde{S}_{xx}(p\tilde{S}_{xx} - \tilde{S}_{yy}) + 2|\tilde{S}_{xy}|^2 - \tilde{S}_{xx}\sqrt{(p\tilde{S}_{xx} - \tilde{S}_{yy})^2 + 4p|\tilde{S}_{xy}|^2}}{2|\tilde{S}_{xy}|\sqrt{(\tilde{S}_{yy} - p\tilde{S}_{xx})^2 + 4p|\tilde{S}_{xy}|^2}} \\
&\le 0 && \text{by (3) lemma}
\end{aligned}$$

Theorem 14.2. *Let* S *be a* SYSTEM *with input* x(*n*) *and output* y(*n*).

$$\text{THM}\quad \left|\hat{\mathsf{H}}_1(\omega)\right| \le \left|\hat{\mathsf{H}}_{hm}(\omega)\right| \le \left|\hat{\mathsf{H}}_{gm}(\omega)\right| \le \left|\hat{\mathsf{H}}_{am}(\omega)\right| \le \left|\hat{\mathsf{H}}_2(\omega)\right|$$

✎PROOF:

1. lemma: $\hat{\mathsf{H}}_1(\omega) \le \hat{\mathsf{H}}_2(\omega)$. Proof:

$$\begin{aligned}
\left|\hat{\mathsf{H}}_1\right| &\triangleq \left|\frac{\tilde{\mathsf{S}}_{yx}(\omega)}{\tilde{\mathsf{S}}_{xx}(\omega)}\right| && \text{by definition of } \hat{\mathsf{H}}_1 && \text{(Definition 14.2 page 106)}\\
&= \left|\frac{\langle \mathsf{y} \mid \mathsf{x}\rangle}{\|\mathsf{x}\|^2}\right| = \frac{|\langle \mathsf{y} \mid \mathsf{x}\rangle|}{\|\mathsf{x}\|^2}\\
&\le \frac{|\langle \mathsf{y} \mid \mathsf{x}\rangle|}{\|\mathsf{x}\|^2}\left|\frac{\|\mathsf{x}\|\,\|\mathsf{y}\|}{\langle \mathsf{y} \mid \mathsf{x}\rangle}\right|^2 && \text{by \textit{Cauchy Schwartz Inequality}} && \text{Theorem N.2 page 310}\\
&= \frac{\|\mathsf{y}\|^2}{|\langle \mathsf{y} \mid \mathsf{x}\rangle|} = \left|\frac{\|\mathsf{y}\|^2}{\langle \mathsf{x} \mid \mathsf{y}\rangle}\right| = \left|\frac{\tilde{\mathsf{S}}_{yy}(\omega)}{\tilde{\mathsf{S}}_{xy}(\omega)}\right|\\
&= \left|\hat{\mathsf{H}}_2\right| && \text{by definition of } \hat{\mathsf{H}}_2 && \text{(Definition 14.3 page 106)}
\end{aligned}$$

2. remainder of the proof:

$$\begin{aligned}
\left|\hat{\mathsf{H}}_1(\omega)\right| &= \min\left\{\hat{\mathsf{H}}_1(\omega),\, \hat{\mathsf{H}}_2(\omega)\right\} && \text{by (1) lemma}\\
&\le \left|\hat{\mathsf{H}}_{hm}(\omega)\right| && \text{by Corollary Q.1 page 351} && \text{with } \lambda_1 = \lambda_2 = \tfrac{1}{2}\\
&\le \left|\hat{\mathsf{H}}_{gm}(\omega)\right| && \text{by Corollary Q.1 page 351} && \text{with } \lambda_1 = \lambda_2 = \tfrac{1}{2}\\
&\le \left|\hat{\mathsf{H}}_{am}(\omega)\right| && \text{by Corollary Q.1 page 351} && \text{with } \lambda_1 = \lambda_2 = \tfrac{1}{2}\\
&\le \max\left\{\hat{\mathsf{H}}_1(\omega),\, \hat{\mathsf{H}}_2(\omega)\right\} && \text{by Corollary Q.1 page 351} && \text{with } \lambda_1 = \lambda_2 = \tfrac{1}{2}\\
&= \left|\hat{\mathsf{H}}_2(\omega)\right| && \text{by (1) lemma}
\end{aligned}$$

⇨

Theorem 14.2 (page 114) compared the magnitudes of several transfer function estimates and demonstrated a simple *linear* relationship. What about phase? The phase of those estimates is even simpler than the magnitude, as demonstrated next.

Proposition 14.5 (Estimator phase). *Let* $z \triangleq |z|e^{i\phi}$ *be a* COMPLEX *number in the set of complex numbers* $\mathbb{C}$. *Let* $\angle z \triangleq \phi$ *be the* PHASE *of* z.

$$\text{PRP}\quad \begin{aligned}
\angle\hat{\mathsf{H}}_1(\omega) &= \angle\hat{\mathsf{H}}_{hm}(\omega) = \angle\hat{\mathsf{H}}_{gm}(\omega) = \angle\hat{\mathsf{H}}_{am}(\omega) = \angle\hat{\mathsf{H}}_2(\omega) = \angle\hat{\mathsf{H}}_s(\omega) = \angle\hat{\mathsf{H}}_v(\omega) = \angle\hat{\mathsf{H}}_\kappa(\omega)\\
&= \angle\mathsf{C}_{xy}(\omega)\\
&= -\angle\tilde{\mathsf{S}}_{xy}(\omega)
\end{aligned}$$

✎PROOF:

$$\begin{aligned}
\angle\hat{\mathsf{H}}_1 &\triangleq \angle\frac{\tilde{\mathsf{S}}_{yx}}{\tilde{\mathsf{S}}_{xx}} && = && = \angle\tilde{\mathsf{S}}^*_{xy} = -\angle\tilde{\mathsf{S}}_{xy} \\
&\uparrow \text{(Definition 14.2 page 106)} \\
\angle\hat{\mathsf{H}}_{hm} &\triangleq \angle\frac{2\tilde{\mathsf{S}}_{yy}\tilde{\mathsf{S}}^*_{xy}}{\tilde{\mathsf{S}}_{xx}\tilde{\mathsf{S}}_{yy}+\left|\tilde{\mathsf{S}}_{xy}\right|^2} && = && = \angle\tilde{\mathsf{S}}^*_{xy} = -\angle\tilde{\mathsf{S}}_{xy} \\
&\uparrow \text{(Definition 14.6 page 108)} \\
\angle\hat{\mathsf{H}}_{gm} &\triangleq \angle\frac{\tilde{\mathsf{S}}^*_{xy}}{\left|\tilde{\mathsf{S}}_{xy}\right|}\sqrt{\frac{\tilde{\mathsf{S}}_{yy}}{\tilde{\mathsf{S}}_{xx}}} && = && = \angle\tilde{\mathsf{S}}^*_{xy} = -\angle\tilde{\mathsf{S}}_{xy} \\
&\uparrow \text{(Definition 14.5 page 107)} \\
\angle\hat{\mathsf{H}}_{am} &\triangleq \angle\frac{\left|\tilde{\mathsf{S}}_{xy}\right|^2+\tilde{\mathsf{S}}_{xx}\tilde{\mathsf{S}}_{yy}}{2\tilde{\mathsf{S}}_{xx}\tilde{\mathsf{S}}_{xy}} && = \angle\frac{1}{\tilde{\mathsf{S}}_{xy}} && = \angle\tilde{\mathsf{S}}^*_{xy} = -\angle\tilde{\mathsf{S}}_{xy} \\
&\uparrow \text{(Definition 14.4 page 106)} \\
\angle\hat{\mathsf{H}}_2 &\triangleq \angle\frac{\tilde{\mathsf{S}}_{yy}}{\tilde{\mathsf{S}}_{xy}} && = \angle\frac{1}{\tilde{\mathsf{S}}_{xy}} && = \angle\tilde{\mathsf{S}}^*_{xy} = -\angle\tilde{\mathsf{S}}_{xy} \\
&\uparrow \text{(Definition 14.3 page 106)} \\
\angle\hat{\mathsf{H}}_s &\triangleq \angle\frac{\tilde{\mathsf{S}}_{yy}-s^2\tilde{\mathsf{S}}_{xx}+\sqrt{\left[\tilde{\mathsf{S}}_{yy}-s^2\tilde{\mathsf{S}}_{xx}\right]^2+4s^2\left|\tilde{\mathsf{S}}_{xy}\right|^2}}{2\tilde{\mathsf{S}}_{xy}} && = \angle\frac{1}{\tilde{\mathsf{S}}_{xy}} && = \angle\tilde{\mathsf{S}}^*_{xy} = -\angle\tilde{\mathsf{S}}_{xy} \\
&\uparrow \text{(Definition 14.8 page 109)} \\
\angle\hat{\mathsf{H}}_v &\triangleq \angle\frac{\tilde{\mathsf{S}}_{yy}-\tilde{\mathsf{S}}_{xx}+\sqrt{\left[\tilde{\mathsf{S}}_{yy}-\tilde{\mathsf{S}}_{xx}\right]^2+4\left|\tilde{\mathsf{S}}_{xy}\right|^2}}{2\tilde{\mathsf{S}}_{xy}} && = \angle\frac{1}{\tilde{\mathsf{S}}_{xy}} && = \angle\tilde{\mathsf{S}}^*_{xy} = -\angle\tilde{\mathsf{S}}_{xy} \\
&\uparrow \text{(Definition 14.9 page 110)} \\
\angle\hat{\mathsf{H}}_\kappa &\triangleq \angle\frac{\tilde{\mathsf{S}}_{yy}-\kappa\tilde{\mathsf{S}}_{xx}+\sqrt{\left[\tilde{\mathsf{S}}_{yy}-\kappa\tilde{\mathsf{S}}_{xx}\right]^2+4\kappa\left|\tilde{\mathsf{S}}_{xy}\right|^2}}{2\tilde{\mathsf{S}}_{xy}} && = \angle\frac{1}{\tilde{\mathsf{S}}_{xy}} && = \angle\tilde{\mathsf{S}}^*_{xy} = -\angle\tilde{\mathsf{S}}_{xy} \\
&\uparrow \text{(Definition 14.7 page 109)} \\
\angle\mathsf{C}_{xy} &\triangleq \angle\frac{\tilde{\mathsf{S}}^*_{xy}}{\sqrt{\tilde{\mathsf{S}}_{xx}\tilde{\mathsf{S}}_{yy}}} && = && = \angle\tilde{\mathsf{S}}^*_{xy} = -\angle\tilde{\mathsf{S}}_{xy} \\
&\uparrow \text{(Definition 14.12 page 126)}
\end{aligned}$$

✏

14.5 Alternate forms

Any standard kit of algebraic tricks should arguably always include the ability to swap the location of a square root between numerator and denominator. If you are of this persuasion, after traveling from the definition of $\hat{\mathsf{H}}_s$ on page 109, you won't be dissappointed when arriving at the next proposition (Proposition 14.6 page 115). But it has more use than just allowing you to entertain friends at social occasions. It also makes it very easy to see (using only algebra) what previously employed *l'Hôpital's rule* (using calculus) in the proof of Theorem 14.1—that $\lim_{s\to\infty}\hat{\mathsf{H}}_s = \hat{\mathsf{H}}_1$.

Proposition 14.6. [17] *Let* $\hat{\mathsf{H}}_\kappa(\omega;\kappa)$ *be defined as in Definition 14.7 (page 109).*

[17] Shin and Hammond (2008) page 293 ⟨(9.67)⟩, Leclere et al. (2014) ⟨(10) $\kappa(f) = 1/s^2$ and x and y swapped⟩

PRP

$$\begin{aligned}
\hat{H}_\kappa(\omega; s) &= \frac{2\kappa(\omega)\tilde{S}_{yx}(\omega)}{\kappa(\omega)\tilde{S}_{xx}(\omega) - \tilde{S}_{yy}(\omega) + \sqrt{\left[\kappa(\omega)\tilde{S}_{xx}(\omega) - \tilde{S}_{yy}(\omega)\right]^2 + 4\kappa(\omega)\left|\tilde{S}_{xy}(\omega)\right|^2}} \\
&= \frac{2\tilde{S}^*_{xy}}{\tilde{S}_{xx} - \frac{1}{\kappa(\omega)}\tilde{S}_{yy} + \sqrt{\left[\tilde{S}_{xx} - \frac{1}{\kappa(\omega)}\tilde{S}_{yy}\right]^2 + \frac{4}{\kappa(\omega)}\left|\tilde{S}_{xy}\right|^2}}
\end{aligned}$$

✎PROOF: We can transform $\hat{H}_\kappa$ from that found in Definition 14.8 (page 109) to the forms in this proposition by the technique of "*rationalizing the denominator*"[18]

$$\hat{H}_\kappa \triangleq \frac{\tilde{S}_{yy} - \kappa\tilde{S}_{xx} + \sqrt{\left[\tilde{S}_{yy} - \kappa\tilde{S}_{xx}\right]^2 + 4\kappa\left|\tilde{S}_{xy}\right|^2}}{2\tilde{S}_{xy}} \qquad \text{by definition of } \hat{H}_\kappa \text{ (Definition 14.8 page 109)}$$

$$= \frac{\left[\tilde{S}_{yy} - \kappa\tilde{S}_{xx} + \sqrt{\left[\tilde{S}_{yy} - \kappa\tilde{S}_{xx}\right]^2 + 4\kappa\left|\tilde{S}_{xy}\right|^2}\right]\overbrace{\left[\tilde{S}_{yy} - \kappa\tilde{S}_{xx} - \sqrt{\left[\tilde{S}_{yy} - \kappa\tilde{S}_{xx}\right]^2 + 4\kappa\left|\tilde{S}_{xy}\right|^2}\right]}^{\text{"rationalizing factor"}}}{2\tilde{S}_{xy}\underbrace{\left[\tilde{S}_{yy} - \kappa\tilde{S}_{xx} - \sqrt{\left[\tilde{S}_{yy} - \kappa\tilde{S}_{xx}\right]^2 + 4\kappa\left|\tilde{S}_{xy}\right|^2}\right]}_{\text{"rationalizing factor"}}}$$

$$= \frac{\left[\tilde{S}_{yy} - \kappa\tilde{S}_{xx}\right]^2 - \left[\tilde{S}_{yy} - \kappa\tilde{S}_{xx}\right]^2 - 4\kappa\left|\tilde{S}_{xy}\right|^2}{2\tilde{S}_{xy}\left[\tilde{S}_{yy} - \kappa\tilde{S}_{xx} - \sqrt{\left[\tilde{S}_{yy} - \kappa\tilde{S}_{xx}\right]^2 + 4\kappa\left|\tilde{S}_{xy}\right|^2}\right]} = \frac{-4\kappa\left|\tilde{S}_{xy}\right|^2}{2\tilde{S}_{xy}\left[\tilde{S}_{yy} - \kappa\tilde{S}_{xx} - \sqrt{\left[\tilde{S}_{yy} - \kappa\tilde{S}_{xx}\right]^2 + 4\kappa\left|\tilde{S}_{xy}\right|^2}\right]}$$

$$= \boxed{\frac{2\kappa\tilde{S}^*_{xy}}{\kappa\tilde{S}_{xx} - \tilde{S}_{yy} + \sqrt{\left[\kappa\tilde{S}_{xx} - \tilde{S}_{yy}\right]^2 + 4\kappa\left|\tilde{S}_{xy}\right|^2}}} = \frac{2\frac{\kappa}{\kappa}\tilde{S}^*_{xy}}{\frac{\kappa}{\kappa}\tilde{S}_{xx} - \frac{1}{\kappa}\tilde{S}_{yy} + \sqrt{\frac{1}{s^4}\left[\kappa\tilde{S}_{xx} - \tilde{S}_{yy}\right]^2 + \frac{4\kappa}{s^4}\left|\tilde{S}_{xy}\right|^2}}$$

$$= \frac{2\tilde{S}^*_{xy}}{\tilde{S}_{xx} - \frac{1}{\kappa}\tilde{S}_{yy} + \sqrt{\left[\frac{\kappa}{\kappa}\tilde{S}_{xx} - \frac{1}{\kappa}\tilde{S}_{yy}\right]^2 + \frac{4}{\kappa}\left|\tilde{S}_{xy}\right|^2}} = \boxed{\frac{2\tilde{S}^*_{xy}}{\tilde{S}_{xx} - \frac{1}{\kappa}\tilde{S}_{yy} + \sqrt{\left[\tilde{S}_{xx} - \frac{1}{\kappa}\tilde{S}_{yy}\right]^2 + \frac{4}{\kappa}\left|\tilde{S}_{xy}\right|^2}}}$$

Integrity check for $s = 0$ and $s \to \infty$ cases: Let $p \triangleq \kappa$.

$$\lim_{p\to\infty}\hat{H}_\kappa = \lim_{p\to\infty}\frac{2\tilde{S}_{yx}}{\tilde{S}_{xx} - \frac{1}{p}\tilde{S}_{yy} + \sqrt{\left[\tilde{S}_{xx} - \frac{1}{p}\tilde{S}_{yy}\right]^2 + \frac{4}{p}\left|\tilde{S}_{xy}\right|^2}} = \frac{2\tilde{S}_{yx}}{\tilde{S}_{xx} + \sqrt{\left[\tilde{S}_{xx}\right]^2}}$$

$$= \frac{\tilde{S}_{yx}}{\tilde{S}_{xx}} \triangleq \hat{H}_1 \qquad \text{by def. of } \hat{H}_1 \text{ (Definition 14.2 page 106)}$$

$$\begin{aligned}
\lim_{p\to 0}\hat{H}_\kappa &= \lim_{p\to 0}\frac{2p\tilde{S}_{yx}}{p\tilde{S}_{xx} - \tilde{S}_{yy} + \sqrt{\left[p\tilde{S}_{xx} - \tilde{S}_{yy}\right]^2 + 4p\left|\tilde{S}_{xy}\right|^2}} \\
&= \lim_{p\to 0}\frac{\frac{d}{dp}\left(2p\tilde{S}_{yx}\right)}{\frac{d}{dp}\left(p\tilde{S}_{xx} - \tilde{S}_{yy} + \sqrt{\left[p\tilde{S}_{xx} - \tilde{S}_{yy}\right]^2 + 4p\left|\tilde{S}_{xy}\right|^2}\right)} && \text{by } \textit{l'Hôpital's rule}
\end{aligned}$$

[18] Slaught and Lennes (1915) page 274 ⟨"197. Rationalizing the Denominator."⟩ `https://archive.org/details/elementaryalgebr00slaurich/page/274` Note that the operation in the proof of Proposition 14.6 is being performed essentially in reverse…or rather "rationalizing the numerator".

$$= \lim_{p\to 0} \frac{2\tilde{\mathsf{S}}_{yx}}{\dfrac{\tilde{\mathsf{S}}_{xx}\left(p\tilde{\mathsf{S}}_{xx}-\tilde{\mathsf{S}}_{yy}\right) + 2\left|\tilde{\mathsf{S}}_{xy}\right|^2 + \tilde{\mathsf{S}}_{xx}\sqrt{\left(p\tilde{\mathsf{S}}_{xx}-\tilde{\mathsf{S}}_{yy}\right)^2+4p\left|\tilde{\mathsf{S}}_{xy}\right|^2}}{\sqrt{\left(p\tilde{\mathsf{S}}_{xx}-\tilde{\mathsf{S}}_{yy}\right)^2+4p\left|\tilde{\mathsf{S}}_{xy}\right|^2}}} \qquad \text{by Lemma 14.1 page 110}$$

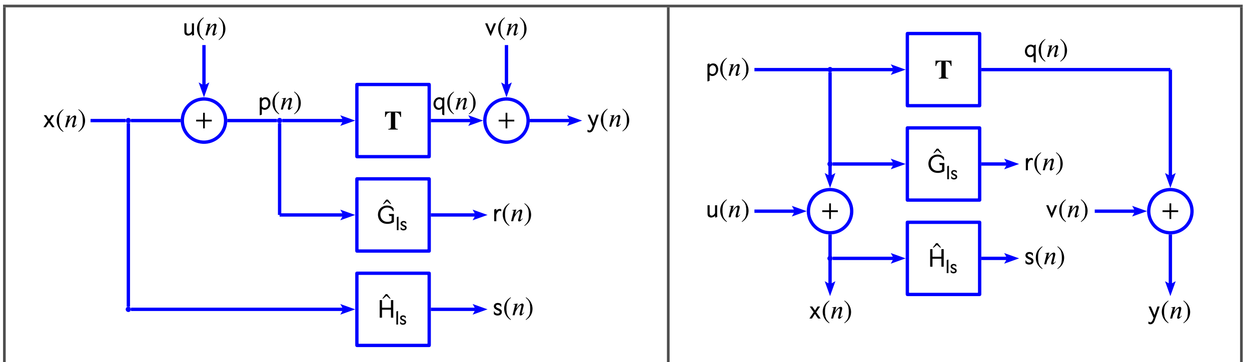
$$= \frac{2\tilde{\mathsf{S}}_{yy}\tilde{\mathsf{S}}_{yx}}{-\tilde{\mathsf{S}}_{xx}\tilde{\mathsf{S}}_{yy} + 2\left|\tilde{\mathsf{S}}_{xy}\right|^2 + \tilde{\mathsf{S}}_{xx}\tilde{\mathsf{S}}_{yy}} = \frac{2\tilde{\mathsf{S}}_{yy}\tilde{\mathsf{S}}_{yx}}{2\left|\tilde{\mathsf{S}}_{xy}\right|^2} = \frac{\tilde{\mathsf{S}}_{yy}}{\tilde{\mathsf{S}}_{xy}}$$

$$\triangleq \hat{\mathsf{H}}_2 \qquad \text{by def. of } \hat{\mathsf{H}}_2 \text{ (Definition 14.3 page 106)}$$

✏

14.6 Least squares estimates of non-linear systems

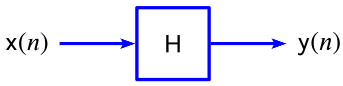
> ❝*The legendary Hungarian mathematician John von Neumann once referred to the the theory of nonequilibrium systems as the "theory of non-elephants," …Nevertheless, such a theory of non-elephants will be attempted here.*❞
>
> Per Bak, in *"how nature works…"* [19]



(A) *Least squares estimation* for *communications additive noise model* (Corollary 14.1 page 120)

(B) *Least squares estimations* for *measurement additive noise model* (Corollary 14.2 page 121)

Figure 14.3: Least Square estimation (Theorem 14.3 page 118)



Let **S** be the *system* illustrated to the right. **If** there is no measurement noise on the input and output and **if** **H** is *linear time invariant*, then $\tilde{\mathsf{H}} = \tilde{\mathsf{S}}_{xy}/\tilde{\mathsf{S}}_{xx}$ (Corollary 7.1 page 52). But what if there is output measurement noise? And what if **H** is not *LTI*? What is the best least-squares estimate of $\tilde{\mathsf{H}}$? The answer depends on how you define "the best".

The defintion of "best" or "optimal" is given by a *cost function* $\mathsf{C}(\hat{\mathsf{H}})$. There are several possible cost functions. Definition 14.11 provides some. Theorem 14.3 then demonstrate optimal solutions with respect to these definitions.

Definition 14.11. *Let* **S** *be a system defined as in Figure 14.3 (page 117) (A) or (B). Define the following* COST FUNCTIONS *for spectral* LEAST-SQUARES *estimates:*

[19] Bak (2013) page 29 ⟨§ Systems in Balance Are Not Complex⟩

DEF

$$\begin{aligned}
\mathsf{C}_{rq}(\hat{\mathsf{G}}) &\triangleq \tilde{\mathbf{F}}\,\|\mathsf{r}(n)-\mathsf{q}(n)\|^2 \triangleq \tilde{\mathbf{F}}\,\langle \mathsf{r}(n)-\mathsf{q}(n)\,|\,\mathsf{r}(0)-\mathsf{q}(0)\rangle \triangleq \tilde{\mathbf{F}}\mathbf{E}\Big(\big[\mathsf{r}(n)-\mathsf{q}(n)\big]\big[\mathsf{r}(0)-\mathsf{q}(0)\big]^*\Big)\\
\mathsf{C}_{sy}(\hat{\mathsf{H}}) &\triangleq \tilde{\mathbf{F}}\,\|\mathsf{s}(n)-\mathsf{y}(n)\|^2 \triangleq \tilde{\mathbf{F}}\,\langle \mathsf{s}(n)-\mathsf{y}(n)\,|\,\mathsf{s}(0)-\mathsf{y}(0)\rangle \triangleq \tilde{\mathbf{F}}\mathbf{E}\Big(\big[\mathsf{s}(n)-\mathsf{y}(n)\big]\big[\mathsf{s}(0)-\mathsf{y}(0)\big]^*\Big)
\end{aligned}$$

Lemma 14.3. *Let* $\mathsf{C}_{rq}(\hat{\mathsf{G}})$ *and* $\mathsf{C}_{sy}(\hat{\mathsf{H}})$ *be defined as in Definition 14.11 (page 117).*

LEM

$$\begin{aligned}
\mathsf{C}_{rq}(\hat{\mathsf{G}}) &= \tilde{\mathsf{S}}_{pp}(\omega)\big|\hat{\mathsf{G}}(\omega)\big|^2 - \tilde{\mathsf{S}}_{py}(\omega)\hat{\mathsf{G}}(\omega) - \tilde{\mathsf{S}}^*_{py}(\omega)\hat{\mathsf{G}}^*(\omega) + \tilde{\mathsf{S}}_{qq}(\omega)\\
\mathsf{C}_{sy}(\hat{\mathsf{H}}) &= \tilde{\mathsf{S}}_{xx}(\omega)\big|\hat{\mathsf{H}}(\omega)\big|^2 - \tilde{\mathsf{S}}_{xy}(\omega)\hat{\mathsf{H}}(\omega) - \tilde{\mathsf{S}}^*_{xy}(\omega)\hat{\mathsf{H}}^*(\omega) + \tilde{\mathsf{S}}_{yy}(\omega)
\end{aligned}$$

✎PROOF:

$$\boxed{\mathsf{C}_{rq}(\hat{\mathsf{G}})}$$

$$\begin{aligned}
&\triangleq \tilde{\mathbf{F}}\mathbf{E}\Big(\big[\mathsf{r}(n)-\mathsf{q}(n)\big]\big[\mathsf{r}(0)-\mathsf{q}(0)\big]^*\Big) && \text{by definition of } \mathsf{C}_{rq} && \text{(Definition 14.11 page 117)}\\
&= \tilde{\mathbf{F}}\Big[\mathbf{E}\big[\mathsf{r}(n)\mathsf{r}^*(0)\big] - \mathbf{E}\big[\mathsf{r}(n)\mathsf{q}^*(0)\big] - \mathbf{E}\big[\mathsf{q}(n)\mathsf{r}^*(0)\big] + \mathbf{E}\big[\mathsf{q}(n)\mathsf{q}^*(0)\big]\Big] && \text{by } \textit{linearity} \text{ of } \mathbf{E} && \text{(Theorem 1.1 page 4)}\\
&\triangleq \tilde{\mathbf{F}}\big[\mathsf{R}_{rr}(m) - \mathsf{R}_{rq}(m) - \mathsf{R}_{qr}(m) + \mathsf{R}_{qq}(m)\big] && \text{by definition of } \mathsf{R}_{xy} && \text{(Definition 2.4 page 12)}\\
&\triangleq \big[\tilde{\mathsf{S}}_{rr}(\omega) - \tilde{\mathsf{S}}_{rq}(\omega) - \tilde{\mathsf{S}}_{qr}(\omega) + \tilde{\mathsf{S}}_{qq}(\omega)\big] && \text{by definition of } \tilde{\mathsf{S}}_{xy} && \text{(Definition 8.3 page 60)}\\
&= \boxed{\tilde{\mathsf{S}}_{pp}(\omega)\big|\hat{\mathsf{G}}(\omega)\big|^2 - \tilde{\mathsf{S}}_{py}(\omega)\hat{\mathsf{G}}(\omega) - \tilde{\mathsf{S}}^*_{py}(\omega)\hat{\mathsf{G}}^*(\omega) + \tilde{\mathsf{S}}_{qq}(\omega)} && \text{by (A)–(D) and} && \text{Corollary 9.8 page 75}
\end{aligned}$$

$$\boxed{\mathsf{C}_{sy}(\hat{\mathsf{H}})}$$

$$\begin{aligned}
&\triangleq \tilde{\mathbf{F}}\mathbf{E}\Big(\big[\mathsf{s}(n)-\mathsf{y}(n)\big]\big[\mathsf{s}(0)-\mathsf{y}(0)\big]^*\Big) && \text{by definition of } \mathsf{C}_{sy} && \text{(Definition 14.11 page 117)}\\
&= \tilde{\mathbf{F}}\Big[\mathbf{E}\big[\mathsf{s}(n)\mathsf{s}^*(0)\big] - \mathbf{E}\big[\mathsf{s}(n)\mathsf{y}^*(0)\big] - \mathbf{E}\big[\mathsf{y}(n)\mathsf{s}^*(0)\big] + \mathbf{E}\big[\mathsf{y}(n)\mathsf{y}^*(0)\big]\Big] && \text{by } \textit{linearity} \text{ of } \mathbf{E} && \text{(Theorem 1.1 page 4)}\\
&\triangleq \tilde{\mathbf{F}}\big[\mathsf{R}_{ss}(m) - \mathsf{R}_{sy}(m) - \mathsf{R}_{ys}(m) + \mathsf{R}_{yy}(m)\big] && \text{by definition of } \mathsf{R}_{xy} && \text{(Definition 2.4 page 12)}\\
&\triangleq \big[\tilde{\mathsf{S}}_{ss}(\omega) - \tilde{\mathsf{S}}_{sy}(\omega) - \tilde{\mathsf{S}}_{ys}(\omega) + \tilde{\mathsf{S}}_{yy}(\omega)\big] && \text{by definition of } \tilde{\mathsf{S}}_{xy} && \text{(Definition 8.3 page 60)}\\
&= \boxed{\tilde{\mathsf{S}}_{xx}(\omega)\big|\hat{\mathsf{H}}(\omega)\big|^2 - \tilde{\mathsf{S}}_{xy}(\omega)\hat{\mathsf{H}}(\omega) - \tilde{\mathsf{S}}^*_{xy}(\omega)\hat{\mathsf{H}}^*(\omega) + \tilde{\mathsf{S}}_{yy}(\omega)} && \text{by (A)–(D) and} && \text{Corollary 9.8 (page 75)}
\end{aligned}$$

✏

Theorem 14.3. *Let* $\mathbf{S}$ *be the* SYSTEM *illustrated in Figure 14.3 page 117 (A) or (B).*

THM

$$\left.\begin{array}{llll}
\text{(A).} & \mathsf{x},\ \mathsf{u},\ \textit{and}\ \mathsf{v}\ \textit{are} & \text{WSS} & \textit{and}\\
\text{(B).} & \mathsf{x},\ \mathsf{u},\ \textit{and}\ \mathsf{v}\ \textit{are} & \text{UNCORRELATED} & \textit{and}\\
\text{(C).} & \mathbf{E}\mathsf{u} = \mathbf{E}\mathsf{v} = 0 & (\text{ZERO-MEAN}) & \textit{and}\\
\text{(D).} & \hat{\mathsf{G}}_{ls}\ \textit{and}\ \hat{\mathsf{H}}_{ls}\ \textit{are} & \text{LTI} &
\end{array}\right\} \implies \left\{\begin{array}{lll}
\text{(1).} & \arg\min\limits_{\hat{\mathsf{G}}} \mathsf{C}_{rq}(\hat{\mathsf{G}}) = \dfrac{\tilde{\mathsf{S}}_{yp}(\omega)}{\tilde{\mathsf{S}}_{pp}(\omega)}\\
\text{(2).} & \arg\min\limits_{\hat{\mathsf{H}}} \mathsf{C}_{sy}(\hat{\mathsf{H}}) = \dfrac{\tilde{\mathsf{S}}_{yx}(\omega)}{\tilde{\mathsf{S}}_{xx}(\omega)}
\end{array}\right.$$

✎PROOF:

1. Define shorthand function names $\hat{\mathsf{G}} \triangleq \hat{\mathsf{G}}_{ls}$ and $\hat{\mathsf{H}} \triangleq \hat{\mathsf{H}}_{ls}$.
2. lemma:

$$\begin{aligned}
0 &= \frac{\partial}{\partial \hat{\mathsf{G}}_R}\mathsf{C}_{rq}(\hat{\mathsf{G}})\\
&= \frac{\partial}{\partial \hat{\mathsf{G}}_R}\Big(\tilde{\mathsf{S}}_{pp}|\hat{\mathsf{G}}|^2 - \hat{\mathsf{G}}\tilde{\mathsf{S}}_{py} - \hat{\mathsf{G}}^*\tilde{\mathsf{S}}^*_{py} + \tilde{\mathsf{S}}_{qq}\Big) && \text{by Lemma 14.3 page 118}\\
&= \frac{\partial}{\partial \hat{\mathsf{G}}_R}\big(\tilde{\mathsf{S}}_{pp}[\hat{\mathsf{G}}_R^2 + \hat{\mathsf{G}}_I^2] - (\hat{\mathsf{G}}_R + i\hat{\mathsf{G}}_I)\tilde{\mathsf{S}}_{py} - (\hat{\mathsf{G}}_R + i\hat{\mathsf{G}}_I)^*\tilde{\mathsf{S}}^*_{py} + \tilde{\mathsf{S}}_{qq}\big)\\
&= 2\hat{\mathsf{G}}_R\tilde{\mathsf{S}}_{pp} - \tilde{\mathsf{S}}_{py} - \tilde{\mathsf{S}}^*_{py} + \cancelto{0}{\frac{\partial}{\partial \hat{\mathsf{G}}_R}\tilde{\mathsf{S}}_{qq}} && \text{because } \mathsf{q} \text{ does not vary with } \hat{\mathsf{G}}
\end{aligned}$$

$$\begin{aligned}
&= 2\hat{G}_R\tilde{S}_{pp} - 2\mathbf{R}_e\tilde{S}_{py} \\
&= 2\hat{G}_R\tilde{S}_{pp} - 2\mathbf{R}_e\tilde{S}_{yp} && \text{by Corollary 2.2 page 16} \\
\implies & \boxed{\hat{G}_R(\omega) = \frac{\mathbf{R}_e\tilde{S}_{yp}(\omega)}{\tilde{S}_{pp}(\omega)}}
\end{aligned}$$

3. lemma:

$$\begin{aligned}
0 &= \frac{\partial}{\partial \hat{G}_I} C_{rq}(\hat{G}) \\
&= \frac{\partial}{\partial \hat{G}_I}\left(\tilde{S}_{pp}|\hat{G}|^2 - \hat{G}\tilde{S}_{py} - \hat{G}^*\tilde{S}^*_{py} + \tilde{S}_{qq}\right) && \text{by Lemma 14.3 page 118} \\
&= \frac{\partial}{\partial \hat{G}_I}[\tilde{S}_{pp}[\hat{G}_R^2 + \hat{G}_I^2] - (\hat{G}_R + i\hat{G}_I)\tilde{S}_{py} - (\hat{G}_R - i\hat{G}_I)\tilde{S}^*_{py} + \tilde{S}_{qq}] \\
&= 2\hat{G}_I\tilde{S}_{pp} - i\tilde{S}_{py} + i\tilde{S}^*_{py} + \cancelto{0}{\frac{\partial}{\partial \hat{G}_I}\tilde{S}_{qq}} && \text{because q does not vary with } \hat{G} \\
&= 2\hat{G}_I\tilde{S}_{pp} - 2i(i\mathbf{I}_m\tilde{S}_{py}) \\
&= 2\hat{G}_I\tilde{S}_{pp} + 2i(i\mathbf{I}_m\tilde{S}_{yp}) && \text{by Corollary 2.2 page 16} \\
\implies & \boxed{\hat{G}_I(\omega) = \frac{\mathbf{I}_m\tilde{S}_{yp}(\omega)}{\tilde{S}_{pp}(\omega)}}
\end{aligned}$$

4. Proof for $\hat{G} \triangleq \hat{G}_{ls}$ expression:

$$\begin{aligned}
\boxed{\hat{G}(\omega)} &= \hat{G}_R(\omega) + i\hat{G}_I(\omega) \\
&= \frac{\mathbf{R}_e\tilde{S}_{yp}(\omega)}{\tilde{S}_{pp}(\omega)} + \frac{i\mathbf{I}_m\tilde{S}_{yp}(\omega)}{\tilde{S}_{pp}(\omega)} && \text{by (2) lemma and (3) lemma} \\
&= \boxed{\frac{\tilde{S}_{yp}(\omega)}{\tilde{S}_{pp}(\omega)}}
\end{aligned}$$

5. lemma:

$$\begin{aligned}
0 &= \frac{\partial}{\partial \hat{H}_R} C_{sy}(\hat{H}) \\
&= \frac{\partial}{\partial \hat{H}_R}\left(\tilde{S}_{xx}|\hat{H}|^2 - \hat{H}\tilde{S}_{xy} - \hat{H}^*\tilde{S}^*_{xy} + \tilde{S}_{yy}\right) && \text{by Lemma 14.3 page 118} \\
&= \frac{\partial}{\partial \hat{H}_R}(\tilde{S}_{xx}[\hat{H}_R^2 + \hat{H}_I^2] - (\hat{H}_R + i\hat{H}_I)\tilde{S}_{xy} - (\hat{H}_R + i\hat{H}_I)^*\tilde{S}^*_{xy} + \tilde{S}_{yy}) \\
&= 2\hat{H}_R\tilde{S}_{xx} - \tilde{S}_{xy} - \tilde{S}^*_{xy} + \cancelto{0}{\frac{\partial}{\partial \hat{H}_R}\tilde{S}_{yy}} && \text{because y does not vary with } \hat{H} \\
&= 2\hat{H}_R\tilde{S}_{xx} - 2\mathbf{R}_e\tilde{S}_{xy} \\
&= 2\hat{H}_R\tilde{S}_{xx} - 2\mathbf{R}_e\tilde{S}_{yx} && \text{by Corollary 2.2 page 16} \\
\implies & \boxed{\hat{H}_R(\omega) = \frac{\mathbf{R}_e\tilde{S}_{yx}(\omega)}{\tilde{S}_{xx}(\omega)}}
\end{aligned}$$

6. lemma:

$$
\begin{aligned}
0 &= \frac{\partial}{\partial \hat{\mathsf{H}}_I} \mathsf{C}_{\mathsf{sy}}(\hat{\mathsf{H}}) \\
&= \frac{\partial}{\partial \hat{\mathsf{H}}_I} \left(\tilde{\mathsf{S}}_{\mathsf{xx}} |\hat{\mathsf{H}}|^2 - \tilde{\mathsf{S}}_{\mathsf{xy}} \hat{\mathsf{H}} - \tilde{\mathsf{S}}^*_{\mathsf{xy}} \hat{\mathsf{H}}^* + \tilde{\mathsf{S}}_{\mathsf{yy}} \right) && \text{by Lemma 14.3 page 118} \\
&= \frac{\partial}{\partial \hat{\mathsf{H}}_I} [\tilde{\mathsf{S}}_{\mathsf{xx}} [\hat{\mathsf{H}}_R^2 + \hat{\mathsf{H}}_I^2] - \tilde{\mathsf{S}}_{\mathsf{xy}}(\hat{\mathsf{H}}_R + i\hat{\mathsf{H}}_I) - \tilde{\mathsf{S}}^*_{\mathsf{xy}}(\hat{\mathsf{H}}_R - i\hat{\mathsf{H}}_I) + \tilde{\mathsf{S}}_{\mathsf{yy}}] \\
&= 2\hat{\mathsf{H}}_I \tilde{\mathsf{S}}_{\mathsf{xx}} - i\tilde{\mathsf{S}}_{\mathsf{xy}} + i\tilde{\mathsf{S}}^*_{\mathsf{xy}} + \cancelto{0}{\frac{\partial}{\partial \hat{\mathsf{H}}_R} \tilde{\mathsf{S}}_{\mathsf{yy}}} && \text{because q does not vary with } \hat{\mathsf{H}} \\
&= 2\hat{\mathsf{H}}_I \tilde{\mathsf{S}}_{\mathsf{xx}} - 2i(i\mathbf{I_m}\tilde{\mathsf{S}}_{\mathsf{xy}}) \\
&= 2\hat{\mathsf{H}}_I \tilde{\mathsf{S}}_{\mathsf{xx}} + 2i(i\mathbf{I_m}\tilde{\mathsf{S}}_{\mathsf{yx}}) && \text{by Corollary 2.2 page 16} \\
&= 2\hat{\mathsf{H}}_I \tilde{\mathsf{S}}_{\mathsf{xx}} - 2\mathbf{I_m}\tilde{\mathsf{S}}_{\mathsf{yx}} \\
\implies & \boxed{\hat{\mathsf{H}}_I(\omega) = \frac{\mathbf{I_m}\tilde{\mathsf{S}}_{\mathsf{yx}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)}}
\end{aligned}
$$

7. Proof for $\hat{\mathsf{H}} \triangleq \hat{\mathsf{H}}_{\mathsf{ls}}$ expression:

$$
\begin{aligned}
\boxed{\hat{\mathsf{H}}(\omega)} &= \hat{\mathsf{H}}_R(\omega) + i\hat{\mathsf{H}}_I(\omega) \\
&= \frac{\mathbf{R_e}\tilde{\mathsf{S}}_{\mathsf{yp}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)} + \frac{i\mathbf{I_m}\tilde{\mathsf{S}}_{\mathsf{yp}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)} && \text{by (5) lemma and (6) lemma} \\
&= \frac{\mathbf{R_e}\tilde{\mathsf{S}}_{\mathsf{yx}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)} + \frac{i\mathbf{I_m}\tilde{\mathsf{S}}_{\mathsf{yx}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)} && \text{by Theorem 9.4 page 71} \\
&= \boxed{\frac{\tilde{\mathsf{S}}_{\mathsf{yx}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)}}
\end{aligned}
$$

✎

Using Theorem 14.3 (previous) we can see that the optimal **least-squares** operators $\hat{\mathsf{G}}_{\mathsf{ls}}$ and $\hat{\mathsf{H}}_{\mathsf{ls}}$ for the **non-linear** operator $\mathbf{T}$ in Figure 14.3 (page 117) (A) and (B) are (next two corollaries)

$$
\boxed{
\begin{array}{llcll}
(1). & \hat{\mathsf{G}}_{\mathsf{ls}}(\omega) & = & \dfrac{\tilde{\mathsf{S}}_{\mathsf{yx}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega) + \tilde{\mathsf{S}}_{\mathsf{uu}}(\omega)} & \text{for (A)---}\textit{communication system} \\
(2). & \hat{\mathsf{G}}_{\mathsf{ls}}(\omega) & = & \dfrac{\tilde{\mathsf{S}}_{\mathsf{yx}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega) - \tilde{\mathsf{S}}_{\mathsf{uu}}(\omega)} & \text{for (B)---}\textit{measurement system} \\
(3). & \hat{\mathsf{H}}_{\mathsf{ls}}(\omega) & = & \dfrac{\tilde{\mathsf{S}}_{\mathsf{yx}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)} & \text{for either (A) or (B)}
\end{array}
}
$$

Corollary 14.1. *Let* $\mathbf{S}$ *be the* SYSTEM *illustrated in Figure 14.3 page 117 (A).*

THM

$$
\left\{ \textit{hypotheses of Theorem 14.3 page 118} \right\} \implies \left\{
\begin{array}{llcl}
(1). & \underset{\hat{\mathsf{G}}_{\mathsf{ls}}}{\arg\min}\, \mathsf{C}_{\mathsf{rq}}(\hat{\mathsf{G}}_{\mathsf{ls}}) & = & \dfrac{\tilde{\mathsf{S}}_{\mathsf{yx}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega) + \tilde{\mathsf{S}}_{\mathsf{uu}}(\omega)} \\
(2). & \underset{\hat{\mathsf{H}}_{\mathsf{ls}}}{\arg\min}\, \mathsf{C}_{\mathsf{sy}}(\hat{\mathsf{H}}_{\mathsf{ls}}) & = & \dfrac{\tilde{\mathsf{S}}_{\mathsf{yx}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)}
\end{array}
\right\}
$$

✎PROOF:

$$\begin{aligned}
\hat{\mathsf{G}}_{\mathsf{ls}} &= \frac{\tilde{\mathsf{S}}_{\mathsf{yx}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{pp}}(\omega)} && \text{by Theorem 14.3 page 118}\\
&= \frac{\tilde{\mathsf{S}}_{\mathsf{yx}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)+\tilde{\mathsf{S}}_{\mathsf{uu}}(\omega)} && \text{by Theorem 9.1 page 67}\\
\hat{\mathsf{H}}_{\mathsf{ls}} &= \frac{\tilde{\mathsf{S}}_{\mathsf{yx}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)} && \text{by Theorem 14.3 page 118}
\end{aligned}$$

✏

Corollary 14.2. *Let* **S** *be the* SYSTEM *illustrated in Figure 14.3 page 117 (B).*

THM

$$\left\{\text{hypotheses of Theorem 14.3 page 118}\right\} \implies \left\{\begin{array}{lllcl}(1). & \arg\min_{\hat{\mathsf{G}}_{\mathsf{ls}}} \mathsf{C}_{\mathsf{rq}}(\hat{\mathsf{G}}_{\mathsf{ls}}) & = & \dfrac{\tilde{\mathsf{S}}_{\mathsf{yx}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)-\tilde{\mathsf{S}}_{\mathsf{uu}}(\omega)}\\ (2). & \arg\min_{\hat{\mathsf{H}}_{\mathsf{ls}}} \mathsf{C}_{\mathsf{sy}}(\hat{\mathsf{H}}_{\mathsf{ls}}) & = & \dfrac{\tilde{\mathsf{S}}_{\mathsf{yx}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)}\end{array}\right\}$$

✎PROOF:

$$\begin{aligned}
\hat{\mathsf{G}} &= \frac{\tilde{\mathsf{S}}_{\mathsf{yx}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{pp}}(\omega)} && \text{by Theorem 14.3 page 118}\\
&= \frac{\tilde{\mathsf{S}}_{\mathsf{yx}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)-\tilde{\mathsf{S}}_{\mathsf{uu}}(\omega)} && \text{by Theorem 9.1 page 67}\\
\hat{\mathsf{H}} &= \frac{\tilde{\mathsf{S}}_{\mathsf{yx}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)} && \text{by Theorem 14.3 page 118}
\end{aligned}$$

✏

It follows immediately from Corollary 14.1 (page 120) and Corollary 14.2 (page 121) that, in the special case of no input noise ($\mathsf{u}(n) = 0$), $\hat{\mathsf{H}}_1$ is the optimal least-squares estimate of $\tilde{\mathsf{H}}$ (next corollary).

Corollary 14.3. [20] *Let* **S** *be the* SYSTEM *illustrated in Figure 14.3 page 117 (A) or (B).*

COR

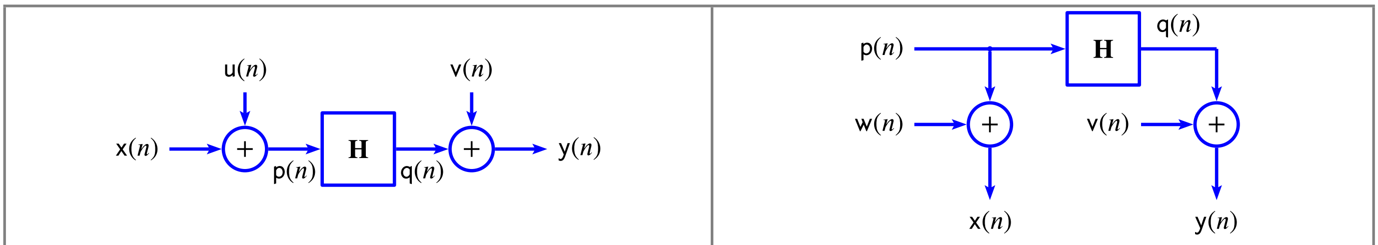
$$\left\{\begin{array}{ll}(1). & \text{hypotheses of Theorem 14.3 and}\\ (2). & \mathsf{u}(n)=0\end{array}\right\} \implies \left\{\hat{\mathsf{G}}_{\mathsf{ls}}(\omega) = \hat{\mathsf{H}}_{\mathsf{ls}}(\omega) = \hat{\mathsf{H}}_1(\omega)\right\}$$

14.7 Least squares estimates of linear systems

The previous section did assume the estimates $\hat{\mathsf{H}}_1$ and $\hat{\mathsf{H}}_2$ to be *linear time invariant* (*LTI*), but it did *not* assume that the system transfer function **T** itself to be *LTI*. But making the LTI assumption on **H** yields some interesting and insightful results, such as those in this section.

Theorem 14.4 (Estimating H in communication additive noise system)**.** *Let* **S** *be the* SYSTEM *illustrated in Figure 14.4 page 122 (A).*

[20] Bendat and Piersol (1980) pages 98–100 ⟨5.1.1 Optimal Character of Calculations; note: proof minimizing $\tilde{\mathsf{S}}_{\mathsf{vv}}$ but yields same result⟩, Bendat and Piersol (1993) pages 106–109 ⟨5.1.1 Optimality of Calculations⟩, Bendat and Piersol (2010) pages 187–190 ⟨6.1.4 Optimum Frequency Response Functions⟩



(A) **communications LTI additive noise model** (B) **measurement LTI additive noise model**

Figure 14.4: Additive noise systems with **LTI** operator **H**

THM

$$\left\{\begin{array}{llll}
\text{(A).} & \mathbf{H} \textit{ is} & \text{LINEAR TIME INVARIANT} & \text{(LTI)} \quad \textit{and} \\
\text{(B).} & \mathsf{x}(n) \textit{ is} & \text{WIDE-SENSE STATIONARY} & \text{(WSS)} \quad \textit{and} \\
\text{(C).} & \mathsf{x}(n), \mathsf{u}(n), \textit{ and } \mathsf{v}(n) \textit{ are} & \text{UNCORRELATED} &
\end{array}\right\}$$

$$\Longrightarrow \left\{\begin{array}{llll}
\text{(1).} & \hat{\mathsf{H}}_1(\omega) & = & \tilde{\mathsf{H}}(\omega) \quad \textit{and} \\
\text{(2).} & \hat{\mathsf{H}}_2(\omega) & = & \dfrac{\tilde{\mathsf{S}}_{\mathsf{vv}}(\omega)}{\tilde{\mathsf{H}}^*(\omega)\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)} + \tilde{\mathsf{H}}(\omega)\left[1 + \dfrac{\tilde{\mathsf{S}}_{\mathsf{uu}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)}\right]
\end{array}\right\}$$

✎PROOF:

$$\begin{aligned}
\hat{\mathsf{H}}_1(\omega) &\triangleq \frac{\tilde{\mathsf{S}}_{\mathsf{yx}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)} && \text{by definition of } \hat{\mathsf{H}}_1 \quad \text{(Definition 14.2 page 106)} \\
&= \frac{\tilde{\mathsf{H}}(\omega)\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)} && \text{by Corollary 9.5 page 73} \\
&= \tilde{\mathsf{H}}(\omega) \\
\hat{\mathsf{H}}_2(\omega) &\triangleq \frac{\tilde{\mathsf{S}}_{\mathsf{yy}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{xy}}(\omega)} && \text{by definition of } \hat{\mathsf{H}}_2 \quad \text{(Definition 14.3 page 106)} \\
&= \frac{\tilde{\mathsf{S}}_{\mathsf{yy}}(\omega)}{\tilde{\mathsf{H}}^*(\omega)\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)} && \text{by Corollary 9.5 page 73} \\
&= \frac{\tilde{\mathsf{S}}_{\mathsf{vv}}(\omega) + \tilde{\mathsf{S}}_{\mathsf{qq}}(\omega)}{\tilde{\mathsf{H}}^*(\omega)\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)} && \text{by Theorem 9.1 page 67} \\
&= \frac{\tilde{\mathsf{S}}_{\mathsf{vv}}(\omega) + \tilde{\mathsf{H}}^*(\omega)\tilde{\mathsf{H}}(\omega)\tilde{\mathsf{S}}_{\mathsf{pp}}(\omega)}{\tilde{\mathsf{H}}^*(\omega)\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)} && \text{by Corollary 7.3 page 53} \\
&= \frac{\tilde{\mathsf{S}}_{\mathsf{vv}}(\omega)}{\tilde{\mathsf{H}}^*(\omega)\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)} + \frac{\tilde{\mathsf{H}}^*(\omega)\tilde{\mathsf{H}}(\omega)\left[\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega) + \tilde{\mathsf{S}}_{\mathsf{uu}}(\omega)\right]}{\tilde{\mathsf{H}}^*(\omega)\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)} \\
&= \frac{\tilde{\mathsf{S}}_{\mathsf{vv}}(\omega)}{\tilde{\mathsf{H}}^*(\omega)\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)} + \tilde{\mathsf{H}}(\omega)\left[1 + \frac{\tilde{\mathsf{S}}_{\mathsf{uu}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)}\right]
\end{aligned}$$

✐

Theorem 14.5 (Estimating H in measurement additive noise system). [21] *Let* **S** *be the* SYSTEM *illustrated in Figure 14.4 page 122 (B).*

[21] Shin and Hammond (2008) page 294 ⟨$H_1(f) = H(f)$ (9.70); $H_2(f) = H(f)\left(1 + S_{n_y n_y}(f)/S_{yy}(f)\right)$ (9.71)⟩, Shin and Hammond (2008) page 294 ⟨$H_1(f) = H(f)/\left(1 + S_{n_x n_x}/S_{xx}(f)\right)$ (9.72); $H_2(f) = H(f)$ (9.73)⟩, Mitchell (1982) page 277 ⟨$H_1(f) = H_0(f)/\left(1 + G_{nn}/G_{uu}\right)$⟩ Mitchell (1982) page 278 ⟨$H_2(f) = H_0(f)\left(1 + G_{mm}/G_{vv}\right)$⟩

THM

$$\left\{\begin{array}{llll}\text{(A).} & \mathbf{H}\textit{ is} & \text{LINEAR TIME INVARIANT} & \textit{and}\\ \text{(B).} & \mathsf{x}(n)\textit{ is} & \text{WIDE-SENSE STATIONARY} & \textit{and}\\ \text{(C).} & \mathsf{x}(n), \mathsf{u}(n), \textit{and } \mathsf{v}(n) \textit{ are} & \text{UNCORRELATED} & \end{array}\right\}$$

$$\Longrightarrow \left\{\begin{array}{lllll}\text{(1).} & \hat{\mathsf{H}}_1(\omega) & = & \tilde{\mathsf{H}}(\omega)\left[\dfrac{1}{1+\dfrac{\tilde{\mathsf{S}}_{\mathsf{ww}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{pp}}(\omega)}}\right] & \text{(UNDER-ESTIMATED)} \quad \textit{and}\\ \text{(2).} & \hat{\mathsf{H}}_2(\omega) & = & \tilde{\mathsf{H}}(\omega)\left[1+\dfrac{\tilde{\mathsf{S}}_{\mathsf{vv}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{qq}}(\omega)}\right] & \text{(OVER-ESTIMATED)}\end{array}\right\}$$

✎PROOF:

$$\begin{aligned}
\hat{\mathsf{H}}_1(\omega) &\triangleq \frac{\tilde{\mathsf{S}}_{\mathsf{yx}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)} && \text{by definition of } \hat{\mathsf{H}}_1 && \text{(Definition 14.2 page 106)}\\
&= \frac{\tilde{\mathsf{S}}_{\mathsf{pp}}(\omega)\tilde{\mathsf{H}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)} && \text{by Corollary 9.4 page 73}\\
&= \frac{\tilde{\mathsf{S}}_{\mathsf{pp}}(\omega)\tilde{\mathsf{H}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{pp}}(\omega)+\tilde{\mathsf{S}}_{\mathsf{ww}}(\omega)} && \text{by hypothesis (A)} && \text{and Corollary 7.3 page 53}\\
&= \tilde{\mathsf{H}}(\omega)\left[\frac{1}{1+\frac{\tilde{\mathsf{S}}_{\mathsf{ww}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{pp}}(\omega)}}\right]\\
\hat{\mathsf{H}}_2(\omega) &\triangleq \frac{\tilde{\mathsf{S}}_{\mathsf{yy}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{xy}}(\omega)} && \text{by definition of } \hat{\mathsf{H}}_2 && \text{(Definition 14.3 page 106)}\\
&= \frac{\tilde{\mathsf{S}}_{\mathsf{qq}}(\omega)+\tilde{\mathsf{S}}_{\mathsf{vv}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{xy}}(\omega)} && \text{by hypothesis (C)} && \text{and Corollary 9.1 page 68}\\
&= \frac{\tilde{\mathsf{S}}_{\mathsf{qq}}(\omega)+\tilde{\mathsf{S}}_{\mathsf{vv}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{xq}}(\omega)} && \text{by hypothesis (C)} && \text{and Theorem 9.4 page 71}\\
&= \frac{\tilde{\mathsf{S}}_{\mathsf{qq}}(\omega)+\tilde{\mathsf{S}}_{\mathsf{vv}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{pq}}(\omega)} && \text{by hypothesis (C)} && \text{and Lemma 9.3 page 70}\\
&= \frac{\tilde{\mathsf{S}}_{\mathsf{qq}}(\omega)+\tilde{\mathsf{S}}_{\mathsf{vv}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{qq}}(\omega)/\tilde{\mathsf{H}}(\omega)} && \text{by \textit{LTI} hypothesis (A)} && \text{and Corollary 7.3 page 53}\\
&= \tilde{\mathsf{H}}(\omega)\left[1+\frac{\tilde{\mathsf{S}}_{\mathsf{vv}}(\omega)}{\tilde{\mathsf{S}}_{\mathsf{qq}}(\omega)}\right] && \text{by hypotheses (A) and (B)} && \text{and Corollary 7.3 page 53}
\end{aligned}$$

✏

Corollary 14.4. *Let* **S** *be the* SYSTEM *illustrated in Figure 14.4 (page 122).*

COR

$$\begin{array}{l}\left\{\begin{array}{ll}\text{(A).} & \textit{hypotheses of Theorem 14.5 and}\\ \text{(B).} & \mathsf{u}(n)=\mathsf{u}(n)=0 \quad \text{(NO INPUT NOISE)}\end{array}\right\} \Longrightarrow \left\{\hat{\mathsf{H}}_1(\omega) = \tilde{\mathsf{H}}(\omega) \quad \text{(UNBIASED)}\right\}\\ \left\{\begin{array}{ll}\text{(A).} & \textit{hypotheses of Theorem 14.5 and}\\ \text{(B).} & \mathsf{v}(n)=0 \quad \text{(NO OUTPUT NOISE)}\end{array}\right\} \Longrightarrow \left\{\hat{\mathsf{H}}_2(\omega) = \tilde{\mathsf{H}}(\omega) \quad \text{(UNBIASED)}\right\}\end{array}$$

Lemma 14.4. *Let* **S** *be the* SYSTEM *illustrated in Figure 14.4 (page 122).*

LEM

$$\left\{\textit{There exists } \kappa(\omega) \textit{ such that } \tilde{\mathsf{S}}_{\mathsf{vv}}(\omega) = \kappa(\omega)\tilde{\mathsf{S}}_{\mathsf{uu}}(\omega)\right\}$$

$$\Longrightarrow \left\{\tilde{\mathsf{S}}_{\mathsf{uu}}(\omega) = \frac{\left|\hat{\mathsf{H}}(\omega)\right|^2\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega) - \hat{\mathsf{H}}(\omega)\tilde{\mathsf{S}}_{\mathsf{xy}}(\omega) - \hat{\mathsf{H}}^*(\omega)\tilde{\mathsf{S}}^*_{\mathsf{xy}}(\omega) + \tilde{\mathsf{S}}_{\mathsf{yy}}(\omega)}{\kappa(\omega) + \left|\hat{\mathsf{H}}(\omega)\right|^2}\right\}$$

✎PROOF:

1. Development based on results of previous chapters:

$$\begin{aligned}
\tilde{S}_{vv} &= \tilde{S}_{yy} - \tilde{S}_{qq} && \text{by Corollary 9.1 page 68}\\
&= \tilde{S}_{yy} - \tilde{S}_{pq}\hat{H} && \text{by Corollary 7.3 page 53}\\
&= \tilde{S}_{yy} - \tilde{S}_{xy}\hat{H} && \text{by Theorem 9.4 page 71}\\
\tilde{S}_{uu} &= \tilde{S}_{xx} - \tilde{S}_{pp} && \text{by Corollary 9.1 page 68}\\
&= \tilde{S}_{xx} - \frac{\tilde{S}_{qp}}{\hat{H}} && \text{by Corollary 7.3 page 53}\\
&= \tilde{S}_{xx} - \frac{\tilde{S}_{yx}}{\hat{H}} && \text{by Theorem 9.4 page 71}\\
\tilde{S}_{uu}\left[\left|\hat{H}\right|^2 + \kappa\right] &= \left|\hat{H}\right|^2\tilde{S}_{uu} + \kappa\tilde{S}_{uu}\\
&\triangleq \tilde{S}_{uu}\left|\hat{H}\right|^2 + \tilde{S}_{vv} && \text{by definition of } \kappa(\omega)\\
&= \left|\hat{H}\right|^2\left[\tilde{S}_{xx} - \frac{\tilde{S}_{yx}}{\hat{H}}\right] + \left[\tilde{S}_{yy} - \tilde{S}_{xy}\hat{H}\right]\\
&= \left|\hat{H}\right|^2\tilde{S}_{xx} - \hat{H}^*\tilde{S}_{yx} - \tilde{S}_{xy}\hat{H} + \tilde{S}_{yy}\\
\implies & \boxed{\tilde{S}_{uu}(\omega) = \frac{\left|\hat{H}(\omega)\right|^2\tilde{S}_{xx}(\omega) - \hat{H}(\omega)\tilde{S}_{xy}(\omega) - \hat{H}^*(\omega)\tilde{S}^*_{xy}(\omega) + \tilde{S}_{yy}(\omega)}{\kappa(\omega) + \left|\hat{H}(\omega)\right|^2}}
\end{aligned}$$

2. Development of Wicks and Vold (Wicks and Vold (1986)):

$$\begin{aligned}
\tilde{Y} - \tilde{V} = \tilde{Q} = \hat{H}\tilde{P} &= \hat{H}(\tilde{X} - \tilde{U}) && \text{by definition of } \hat{H}\\
\hat{H}\tilde{U} - \tilde{V} &= \hat{H}\tilde{X} - \tilde{Y} && \text{by \textit{left distributive} property (Theorem R.4 page 363)}\\
\mathbf{E}\left(\left[\hat{H}\tilde{U} - \tilde{V}\right]\left[\hat{H}\tilde{U} - \tilde{V}\right]^*\right) &= \mathbf{E}\left(\left[\hat{H}\tilde{X} - \tilde{Y}\right]\left[\hat{H}\tilde{X} - \tilde{Y}\right]^*\right)\\
\left|\hat{H}\right|^2\tilde{S}_{uu} - \hat{H}\cancelto{0}{\tilde{S}_{uv}} - \hat{H}^*\cancelto{0}{\tilde{S}_{vu}} + \tilde{S}_{vv} &= \left|\hat{H}\right|^2\tilde{S}_{xx} - \hat{H}\tilde{S}_{xy} - \hat{H}^*\tilde{S}^*_{xy} + \tilde{S}_{yy} && \text{because u and v are \textit{uncorrelated}}\\
\left|\hat{H}\right|^2\tilde{S}_{uu} + \kappa\tilde{S}_{uu} &= \left|\hat{H}\right|^2\tilde{S}_{xx} - \hat{H}\tilde{S}_{xy} - \hat{H}^*\tilde{S}^*_{xy} + \tilde{S}_{yy} && \text{by hypothesis}
\end{aligned}$$

✑

Theorem 14.6. [22] *Let* **S** *be the* SYSTEM *illustrated in Figure 14.4 (page 122). Let* $\hat{H}_\kappa(\omega)$ *be the transfer function estimate defined in Definition 14.7 (page 109).*

THM

$$\left\{\begin{array}{lll} (1). & \text{There exists } \kappa(\omega) \text{ such that} & \\ & \tilde{S}_{vv}(\omega) = \kappa(\omega)\tilde{S}_{uu}(\omega) & \text{and} \\ (2). & C(\hat{H}_{ls}) = \tilde{S}_{uu}(\omega) & \end{array}\right\} \implies \left\{\begin{array}{l} \arg\min_{\hat{H}} C(\hat{H}) = \hat{H}_\kappa(\omega) \\ (\hat{H}_\kappa \text{ is the “optimal” estimator for minimizing system noise}) \end{array}\right\}$$

✎PROOF:

1. Let

$$\begin{aligned}
F &\triangleq \left|\hat{H}(\omega)\right|^2\tilde{S}_{xx}(\omega) - \hat{H}(\omega)\tilde{S}_{xy}(\omega) - \hat{H}^*(\omega)\tilde{S}^*_{xy}(\omega) + \tilde{S}_{yy}(\omega) && \text{(numerator in Lemma 14.4) and}\\
G &\triangleq \kappa(\omega) + \left|\hat{H}(\omega)\right|^2 && \text{(denominator in Lemma 14.4)}
\end{aligned}$$

[22] Wicks and Vold (1986) page 898 ⟨has additional s in denominator⟩, Shin and Hammond (2008) page 293 ⟨(9.67)⟩, White et al. (2006) page 679 ⟨(6)⟩

2. lemma $\left(\frac{\partial}{\partial \hat{H}_R}\tilde{S}_{uu}\right)$:

$$\begin{aligned}
\boxed{0} &= \frac{1}{2}G^2 0 = \frac{1}{2}G^2\frac{\partial}{\partial \hat{H}_R}\tilde{S}_{uu} && \text{set } \frac{\partial}{\partial \hat{H}_R}\tilde{S}_{uu} = 0 \text{ to find optimum } \hat{H}_R\\
&= \frac{1}{2}G^2\frac{\partial}{\partial \hat{H}_R}\frac{F}{G} && \text{by Lemma 14.4 page 123}\\
&= \frac{1}{2}G^2\frac{(F'G - G'F)}{G^2} && \text{by \textit{Quotient Rule}}\\
&= \frac{1}{2}(F'G - G'F)\\
&= \frac{1}{2}\left[2\hat{H}_R\tilde{S}_{xx} - \tilde{S}_{xy} - \tilde{S}^*_{xy}\right]G - \frac{1}{2}2\hat{H}_R F && \text{by definition of } F, G \quad \text{(item (1) page 124)}\\
&= \boxed{\hat{H}_R\tilde{S}_{xx}G - G\mathbf{R}_e\tilde{S}_{xy} - \hat{H}_R F}
\end{aligned}$$

3. lemma $\left(\frac{\partial}{\partial \hat{H}_I}\tilde{S}_{uu}\right)$:

$$\begin{aligned}
\boxed{0} &= \frac{1}{2}G^2 0 = \frac{1}{2}G^2\frac{\partial}{\partial \hat{H}_I}\tilde{S}_{uu} && \text{set } \frac{\partial}{\partial \hat{H}_I}\tilde{S}_{uu} = 0 \text{ to find optimum } \hat{H}_I\\
&= \frac{1}{2}G^2 0 = \frac{1}{2}G^2\frac{\partial}{\partial \hat{H}_I}\frac{F}{G} && \text{by Lemma 14.4 page 123}\\
&= \frac{1}{2}G^2\frac{(F'G - G'F)}{G^2} && \text{by \textit{Quotient Rule}}\\
&= \frac{1}{2}(F'G - G'F)\\
&= \frac{1}{2}\left[2\hat{H}_I\tilde{S}_{xx} - i\tilde{S}_{xy} + i\tilde{S}^*_{xy}\right]G - \frac{1}{2}2\hat{H}_I F && \text{by definition of } F, G \quad \text{(item (1) page 124)}\\
&= \boxed{\hat{H}_I\tilde{S}_{xx}G + G\mathbf{I}_m\tilde{S}_{xy} - \hat{H}_I F}
\end{aligned}$$

4. Solve for $\hat{H}$ ...

$$\begin{aligned}
0 = 0 + i0 &= \frac{1}{2}G^2 0 + \frac{1}{2}G^2 0 = \frac{1}{2}G^2\frac{\partial}{\partial \hat{H}_R}\tilde{S}_{uu} + i\frac{1}{2}G^2\frac{\partial}{\partial \hat{H}_I}\tilde{S}_{uu}\\
&= \left[\hat{H}_R\tilde{S}_{xx}G - G\mathbf{R}_e\tilde{S}_{xy} - \hat{H}_R F\right] + i\left[\hat{H}_I\tilde{S}_{xx}G + G\mathbf{I}_m\tilde{S}_{xy} - \hat{H}_I F\right] && \text{by (2) lemma and (3) lemma}\\
&= \hat{H}\tilde{S}_{xx}G - \tilde{S}^*_{xy}G - \hat{H}F \quad \text{because } \mathbf{R}_e(z) + i\mathbf{I}_m(z) = z \text{ and } \mathbf{R}_e(z) - i\mathbf{I}_m(z) = z^*\\
&= \hat{H}\tilde{S}_{xx}G - \tilde{S}_{yx}G - \hat{H}F && \text{by Corollary 2.2 page 16}\\
&= \hat{H}\tilde{S}_{xx}\left(\kappa + |\hat{H}|^2\right) - \tilde{S}_{yx}\left(\kappa + |\hat{H}|^2\right) - \hat{H}\left(|\hat{H}|^2\tilde{S}_{xx} - \hat{H}\tilde{S}_{xy} - \hat{H}^*\tilde{S}^*_{xy} + \tilde{S}_{yy}\right) && \text{by } F, G \text{ defs.}\\
&= \hat{H}\tilde{S}_{xx}\left(\kappa + \cancel{|\hat{H}|^2}\right) - \tilde{S}_{yx}\left(\kappa + |\hat{H}|^2\right) - \hat{H}\left(\cancel{|\hat{H}|^2\tilde{S}_{xx}} - \hat{H}\tilde{S}_{xy} - \hat{H}^*\tilde{S}^*_{xy} + \tilde{S}_{yy}\right)\\
&= \kappa\hat{H}\tilde{S}_{xx} - \tilde{S}_{yx}\left(\kappa + |\hat{H}|^2\right) + \left(\hat{H}^2\tilde{S}_{xy} + |\hat{H}|^2\tilde{S}^*_{xy} - \hat{H}\tilde{S}_{yy}\right)\\
&= \kappa\hat{H}\tilde{S}_{xx} - \kappa\tilde{S}_{yx} - \cancel{\tilde{S}_{yx}|\hat{H}|^2} + \left(\hat{H}^2\tilde{S}_{xy} + \cancel{|\hat{H}|^2\tilde{S}^*_{xy}} - \hat{H}\tilde{S}_{yy}\right)\\
&= \hat{H}^2\tilde{S}_{xy} + \hat{H}\left[\kappa\tilde{S}_{xx} - \tilde{S}_{yy}\right] - \kappa\tilde{S}^*_{xy}\\
\implies & \boxed{\hat{H} = \frac{(\tilde{S}_{yy} - \kappa\tilde{S}_{xx}) \pm \sqrt{(\tilde{S}_{yy} - \kappa\tilde{S}_{xx})^2 + 4\kappa|\tilde{S}_{xy}|^2}}{2\tilde{S}_{xy}}} && \text{by \textit{Quadratic Equation}}
\end{aligned}$$

14.8 Coherence

14.8.1 Application

Coherence has two basic purposes:

1. The *coherence* of x and y is a measure of how closely x and y are statistically related. That is, it is an indication of how much x and y "cohere" or "stick" together
2. The *coherence* of x and y is a measure of how reliable are the estimates $\hat{\mathsf{H}}_1$ and $\hat{\mathsf{H}}_2$ (Definition 14.2 page 106, Definition 14.3 page 106). If the coherence is 0.70 or above, then we can have high confidence that the estimates $\hat{\mathsf{H}}_1$ and $\hat{\mathsf{H}}_2$ are "good" estimates.[23]

14.8.2 Definitions

Definition 14.12. [24] *Let* **S** *be a* SYSTEM *with input* $\mathsf{x}(n)$ *and output* $\mathsf{y}(n)$.

DEF

The ***complex coherence*** *function* *is defined as* $$\mathsf{C}_{\mathsf{xy}}(\omega) \triangleq \frac{\tilde{\mathsf{S}}^*_{\mathsf{xy}}(\omega)}{\sqrt{\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)\tilde{\mathsf{S}}_{\mathsf{yy}}(\omega)}}$$

The ***ordinary coherence*** *function* *is defined as* $$\gamma^2_{\mathsf{xy}}(\omega) \triangleq \frac{\left|\tilde{\mathsf{S}}_{\mathsf{xy}}(\omega)\right|^2}{\tilde{\mathsf{S}}_{\mathsf{xx}}(\omega)\tilde{\mathsf{S}}_{\mathsf{yy}}(\omega)}$$

Proposition 14.7.

PRP

$$\gamma^2_{\mathsf{xy}}(\omega) = \frac{\hat{\mathsf{H}}_1(\omega)}{\hat{\mathsf{H}}_2(\omega)}$$

✎PROOF:

$$\boxed{\gamma^2_{\mathsf{xy}}(\omega)} \triangleq \frac{\left|\tilde{\mathsf{S}}_{\mathsf{xy}}\right|^2}{\tilde{\mathsf{S}}_{\mathsf{xx}}\tilde{\mathsf{S}}_{\mathsf{yy}}} \qquad \text{by definition of } \gamma^2_{\mathsf{xy}} \qquad \text{(Definition 14.12 page 126)}$$

$$= \frac{\tilde{\mathsf{S}}^*_{\mathsf{xy}}/\tilde{\mathsf{S}}_{\mathsf{xx}}}{\tilde{\mathsf{S}}_{\mathsf{yy}}/\tilde{\mathsf{S}}_{\mathsf{xy}}} \triangleq \boxed{\frac{\hat{\mathsf{H}}_1}{\hat{\mathsf{H}}_2}} \qquad \text{by definitions of } \hat{\mathsf{H}}_1 \text{ and } \hat{\mathsf{H}}_2 \qquad \text{(Definition 14.2 page 106, Definition 14.3 page 106)}$$

⇰

Remark 14.2. Note that the *complex transmissibility* $\tilde{\mathsf{T}}'_{\mathsf{xy}}$ of Remark 14.1 provides a nice mathematical symmetry (always a good sign of good direction) with *coherence* in the system identification family tree. In particular, note that the following:

REM

$$\mathsf{C}_{\mathsf{xy}} \triangleq \sqrt{\frac{\hat{\mathsf{H}}^*_1}{\hat{\mathsf{H}}_2}} \text{ whereas } \tilde{\mathsf{T}}'_{\mathsf{xy}} \triangleq \sqrt{\hat{\mathsf{H}}_1\hat{\mathsf{H}}_2}$$

✎PROOF:

$$\sqrt{\frac{\hat{\mathsf{H}}^*_1(\omega)}{\hat{\mathsf{H}}_2(\omega)}} \qquad \text{by definition of } \hat{\mathsf{H}}_{\mathsf{gm}} \qquad \text{(Definition 14.5 page 107)}$$

⇰

[23] Liang and Lee (2015) pages 363–365 ⟨7.4.2COHERECE FUNCTION⟩

[24] Chen et al. (2012) page 4699(1), (2), Liang and Lee (2015) pages 363–365 ⟨7.4.2 Coherence function⟩, Ewins (1986) page 131 ⟨$\gamma^2 = H_1(\omega)/H_2(\omega)$ (3.8)⟩

14.8.3 A warning

Estimators yield, as the name implies, estimates. These estimates in general contain some error.

Example 14.1 (The K=1 Welch estimate of coherence). Suppose we have two *uncorrelated* stationary sequences $\mathsf{x}(n)$ and $\mathsf{y}(n)$. Then, there CSD $\mathsf{S}_{\mathsf{xy}}(\omega)$ should be 0 because

$$\begin{aligned}
\mathsf{S}_{\mathsf{xy}}(\omega) &\triangleq \check{\mathbf{F}}\mathbf{E}\mathsf{R}_{\mathsf{xy}}(m) \\
&= \check{\mathbf{F}}\mathbf{E}[x(n)y[n+m]] \\
&= \check{\mathbf{F}}\big[\mathbf{E}_{\mathsf{x}}(n)\big]\big[\mathbf{E}_{\mathsf{y}}[n+m]\big] \\
&= \check{\mathbf{F}}[0][0] \\
&= 0
\end{aligned}$$

This will give a coherence of 0 also:

$$C(\omega) = \frac{\mathsf{S}_{\mathsf{xy}}}{\sqrt{\mathsf{S}_{\mathsf{xx}}\mathsf{S}_{\mathsf{yy}}}} = 0$$

However, the Welch estimate with $K = 1$ will yield

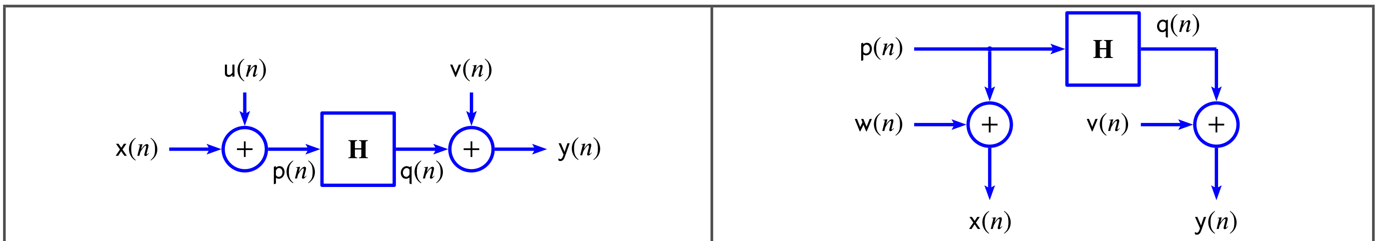
$$\begin{aligned}
|C(\omega)| &= \left|\frac{\mathsf{S}_{\mathsf{xy}}}{\sqrt{\mathsf{S}_{\mathsf{xx}}\mathsf{S}_{\mathsf{yy}}}}\right| \\
&= \left|\frac{(\tilde{\mathbf{F}}x)(\tilde{\mathbf{F}}y)^*}{\sqrt{|\tilde{\mathbf{F}}x|^2|\tilde{\mathbf{F}}y|^2}}\right| \\
&= 1
\end{aligned}$$

CHAPTER 15

NOISE ESTIMATION

Estimating noise in a system is difficult and many estimation methods are possible.

1. Thong et al. (2001)
2. Zheng et al. (2002)
3. Kim and Kamel (2003)
4. Kamel and Sim (2004)



(A) **communications LTI additive noise model** (B) **measurement LTI additive noise model**

Figure 15.1: Additive noise systems with **LTI** operator **H**

CHAPTER 16

MOMENT ESTIMATION

16.1 Mean Estimation

Theorem 16.1. *Let* $\hat{\mu} \triangleq \sum_{n=1}^{N} \lambda_n \mathsf{x}_n$ *with* $\sum_{n=1}^{N} \lambda_n = 1$ *be the* ARITHMETIC MEAN *(Definition Q.4 page 351)*.

THM

$$\left\{\begin{array}{lll} \text{(A).} & (\!(\mathsf{x}_n)\!) \text{ is WIDE SENSE STATIONARY} & \text{and} \\ \text{(B).} & \mu \triangleq \mathbf{E}\mathsf{x}_n & \text{and} \\ \text{(C).} & (\!(\mathsf{x}_n)\!) \text{ is UNCORRELATED} & \text{and} \\ \text{(D).} & \hat{\mu} \triangleq \sum_{n=1}^{N} \lambda_n \mathsf{x}_n \quad \text{(ARITHMETIC MEAN)} & \end{array}\right\} \implies \left\{\begin{array}{lrcll} \text{(1).} & \mathbf{E}\hat{\mu} & = & \mu \quad \text{(UNBIASED)} & \text{and} \\ \text{(2).} & \mathrm{var}(\hat{\mu}) & = & \sigma^2 \sum_{n=1}^{N} \lambda_n^2 & \text{and} \\ \text{(3).} & \mathrm{mse}(\hat{\mu}) & = & \sigma^2 \sum_{n=1}^{N} \lambda_n^2 & \end{array}\right\}$$

✎PROOF:

$\mathbf{E}\hat{\mu} \triangleq \mathbf{E} \sum_{n\in\mathbb{Z}} \lambda_n \mathsf{x}_n$	by definition of *arithmetic mean*	(Definition Q.4 page 351)
$= \sum_{n\in\mathbb{Z}} \lambda_n \mathbf{E}\mathsf{x}_n$	by *linearity* of $\mathbf{E}$	(Theorem 1.1 page 4)
$= \mu \sum_{n\in\mathbb{Z}} \lambda_n$	by *WSS* hypothesis	(A)
$= \mu$	by $\sum \lambda_n = 1$ hypothesis	(Definition Q.4 page 351)
$\mathrm{var}(\hat{\mu}) \triangleq \mathbf{E}(\hat{\mu} - \mathbf{E}\hat{\mu})^2$	by definition of *variance*	
$= \mathbf{E}(\hat{\mu} - \mu)^2$	by previous result	
$= \mathbf{E}\left(\sum_{n=1}^{N} \lambda_n \mathsf{x}_n - \mu\right)^2$	by definition of $\hat{\mu}$	
$= \mathbf{E}\left[\sum_{n=1}^{N} \lambda_n \mathsf{x}_n - \mu \underbrace{\sum_{n=1}^{N} \lambda_n}_{1}\right]^2$	by $\sum \lambda_n = 1$ hypothesis	(Definition Q.4 page 351)

$$
\begin{aligned}
&= \mathbf{E}\left[\sum_{n=1}^{N} \lambda_n(\mathsf{x}_n - \mu)\right]^2 \\
&= \mathbf{E}\left[\sum_{n=1}^{N} \lambda_n(\mathsf{x}_n - \mu) \sum_{m=1}^{N} \lambda_m(\mathsf{x}_m - \mu)\right] \\
&= \sum_{n=1}^{N}\sum_{m=1}^{N} \lambda_n\lambda_m\left(\mathbf{E}\left[(\mathsf{x}_n - \mu)(\mathsf{x}_m - \mu)\right]\right) \\
&= \sum_{n=1}^{N}\sum_{m=1}^{N} \lambda_n\lambda_m\left(\mathbf{E}\left[\mathsf{x}_n\mathsf{x}_m\right] - \mu\mathbf{E}\left[\mathsf{x}_n\right] - \mu\mathbf{E}\left[\mathsf{x}_m\right] + \mu^2\right) \\
&= \sum_{n=1}^{N}\sum_{m=1}^{N} \lambda_n\lambda_m\left(\mathbf{E}\left[\mathsf{x}_n\mathsf{x}_m\right] - \mu^2 - \mu^2 + \mu^2\right) \\
&= \sum_{n=1}^{N}\sum_{m=1}^{N} \lambda_n\lambda_m\left(\mathbf{E}\left[\mathsf{x}_n\mathsf{x}_m\right] - \mu^2\right) \\
&= \sum_{n=1}^{N} \lambda_n^2\left(\mathbf{E}\left[\mathsf{x}_n^2\right] - \mu^2\right) + \sum_{n=1}^{N}\sum_{m\neq n} \lambda_n\lambda_m\left(\mathbf{E}\left[\mathsf{x}_n\mathsf{x}_m\right] - \mu^2\right) \\
&= \sum_{n=1}^{N} \lambda_n^2\left(\mathbf{E}\left[\mathsf{x}_n^2\right] - \mu^2\right) + \sum_{n=1}^{N}\sum_{m\neq n} \lambda_n\lambda_m\left(\mathbf{E}\mathsf{x}_n\mathbf{E}\mathsf{x}_m - \mu^2\right) \\
&= \sum_{n=1}^{N} \lambda_n^2\sigma^2 + \cancelto{0}{\sum_{n=1}^{N}\sum_{m\neq n} \lambda_n\lambda_m\left(\mu\mu - \mu^2\right)} && \text{by } \textit{WSS} \text{ hypothesis} \quad \text{(A)} \\
&= \sigma^2 \sum_{n=1}^{N} \lambda_n^2
\end{aligned}
$$

$$
\begin{aligned}
\mathsf{mse}(\hat{\mu}) &= \mathbf{E}(\hat{\mu} - \mathbf{E}\hat{\mu})^2 + (\mathbf{E}\hat{\mu} - \mu)^2 && \text{by Theorem 10.2 page 83} \\
&= \sigma^2 \sum_{n=1}^{N} \lambda_n^2 + (\mu - \mu)^2 && \text{by previous results} \\
&= \sigma^2 \sum_{n=1}^{N} \lambda_n^2
\end{aligned}
$$

✏

Definition 16.1.

DEF: *The **average** $\hat{\mu}$ of a length N sequence $(\!(\mathsf{x}_n)\!)_1^N$ is defined as* $\hat{\mu} \triangleq \dfrac{1}{N}\sum_{n=1}^{N} \mathsf{x}_n$

Corollary 16.1. [1]

COR:

$$
\left\{\begin{array}{lll}
\text{(A).} & (\!(\mathsf{x}_n)\!) \text{ is WIDE SENSE STATIONARY} & \text{and} \\
\text{(B).} & \mu \triangleq \mathbf{E}\mathsf{x}_n & \text{and} \\
\text{(C).} & (\!(\mathsf{x}_n)\!) \text{ is UNCORRELATED} & \text{and} \\
\text{(D).} & \hat{\mu} \triangleq \dfrac{1}{N}\sum_{n=1}^{N} \mathsf{x}_n \quad \text{(AVERAGE)} &
\end{array}\right\}
\implies
\left\{\begin{array}{llll}
\text{(1).} & \mathbf{E}\hat{\mu} = \mu & \text{(UNBIASED)} & \text{and} \\
\text{(2).} & \mathsf{var}(\hat{\mu}) = \dfrac{\sigma^2}{N} & & \text{and} \\
\text{(3).} & \mathsf{mse}(\hat{\mu}) = \dfrac{\sigma^2}{N} & \text{(CONSISTENT)} &
\end{array}\right\}
$$

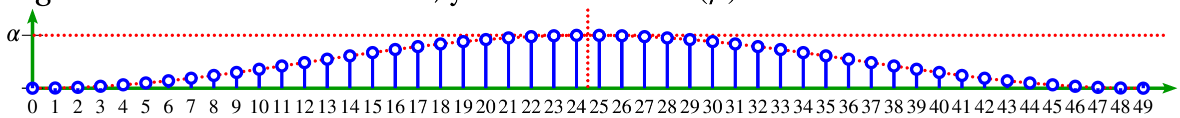
[1] Kay (1988) page 45 ⟨§"3.3 Estimation Theory"⟩, Clarkson (1993) pages 52–54 ⟨§"2.6.2 Estimation of Moments — Time Averages"⟩,

✎PROOF: These results follow from Theorem 16.1 (page 131) with $\lambda_n = \frac{1}{N}$. ✏

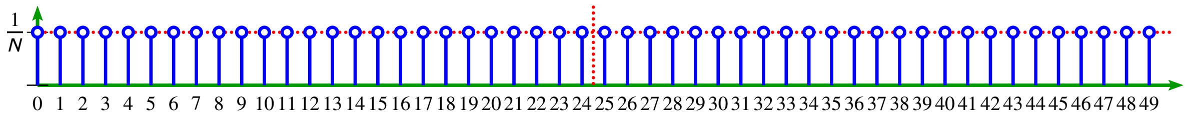
The *arithmetic mean* estimator $\hat{\mu} \triangleq \sum \lambda_n \mathsf{x}_n$ is *unbiased* and *consistent* for any $\sum \lambda_n = 1$ and yields *mean square error* $\mathsf{mse}(\hat{\mu}) = \sigma^2 \sum \lambda_n^2$ (Theorem 16.1 page 131). But…

1. Said qualitatively: "What is the 'best' sequence $(\!(\lambda_n)\!)$ to use?"
2. Said quantitatively: "What sequence $(\!(\lambda_n)\!)$ yields the smallest $\mathsf{mse}(\hat{\mu})$?"

For example, would fashioning $(\!(\lambda_n)\!)$ to be a scaled version of a standard window function, like the *Hanning window*[2] illustrated below, yield the best $\mathsf{mse}(\hat{\mu})$?



Theorem 16.2 (page 134) answers question (2) stating that the best sequence in terms of minimal mse is the sequence $(\!(\lambda_n)\!) \triangleq \frac{1}{N}(\!(\ldots, 1, 1, 1, \ldots)\!)$, which is the *average* estimator, which yields $\mathsf{mse}(\hat{\mu}) = \frac{\sigma^2}{N}$ (Corollary 16.1 page 132).



That is, it turns out that $\frac{1}{N} \le \sum \lambda_n^2$ for all possible sequences $(\!(\lambda_n)\!)$. This fact is demonstrated by Lemma 16.1 (next), which in turn follows more or less directly from the ubiquitous *Cauchy-Schwarz Inequality* (Theorem Q.6 page 354, Theorem N.2 page 310).

Even further strengthening the average as choice estimator is Corollary 17.1 (page 139) which demonstrates that in the case where $(\!(\mathsf{x}_n)\!)$ is *uncorrelated* and *Gaussian*, then the optimal maximum likelihood estimator is the average.

Lemma 16.1.

LEM
$$\left\{\sum_{n=1}^{N} \lambda_n = 1\right\} \quad\Longrightarrow\quad \left\{\frac{1}{N} \le \sum_{n=1}^{N} \lambda_n^2\right\}$$

✎PROOF:

1. Let the sequence $(\!(a_n)\!)$ be defined as $(\!(a_n)\!) \triangleq (\!(\ldots, 1, 1, 1, \ldots)\!)$
2. Let *inner product* $\langle a_n \,|\, b_n \rangle$ be defined as $\langle a_n \,|\, b_n \rangle \triangleq \sum_{n=1}^{N} a_n b_n$
3. Let *norm* $\|a_n\|$ be defined as $\|a_n\| \triangleq \sum_{n=1}^{N} a_n^2$
4. Proof of lemma:

$$\begin{aligned}
\boxed{\frac{1}{N}} &= \frac{1}{N}\left(\sum_{n=1}^{N} \lambda_n\right)^2 && \text{by } \sum_{n=1}^{N} \lambda_n = 1 \text{ hypothesis} \\
&= \frac{1}{N}\left(\sum_{n=1}^{N} a_n \lambda_n\right)^2 && \text{by } (\!(a_n)\!) \triangleq (\!(\ldots, 1, 1, 1, \ldots)\!) \text{ definition} && \text{(definition 1 page 133)} \\
&\boxed{\le} \frac{1}{N}\left(\sum_{n=1}^{N} a_n^2\right)\left(\sum_{n=1}^{N} \lambda_n^2\right) && \text{by \textit{Cauchy-Schwartz inequality}} && \text{(Theorem Q.6 page 354)}
\end{aligned}$$

[2] Abdaheer (2009) page 130

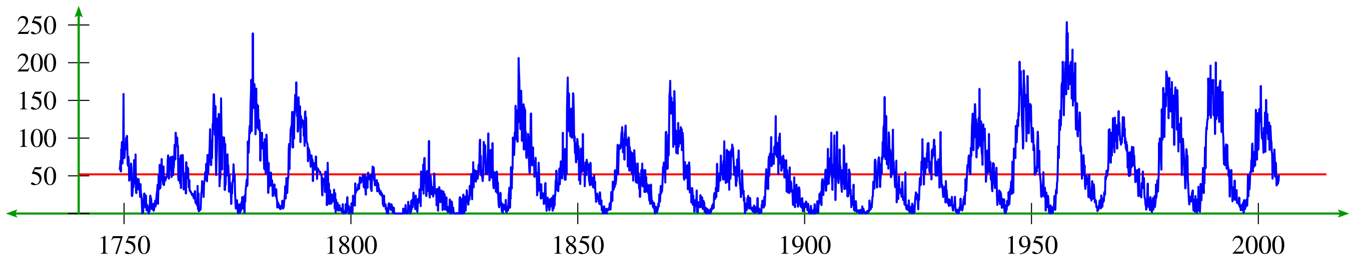


Figure 16.1: Sunspots data from 1749 into 2004

$$\triangleq \frac{1}{N}\left(\sum_{n=1}^{N} 1^2\right)\left(\sum_{n=1}^{N} \lambda_n^2\right) \qquad \text{by definition of } (\!(a_n)\!) \qquad \text{(definition 1 page 133)}$$

$$= \boxed{\sum_{n=1}^{N} \lambda_n^2}$$

✏

Theorem 16.2. *Let* mse(*average mean*) *be the mean square error of the* AVERAGE *estimator* (Corollary 16.1 page 132) *and* mse(*arithmetic mean*) *be the mean square error of the* ARITHMETIC *estimator* (Theorem 16.1 page 131)*.*

PRP	$\text{mse}(\textit{average mean}) \le \text{mse}(\textit{arithmetic mean})$

✎PROOF:

$$\begin{aligned}
\text{mse}(\text{average mean}) &= \sigma^2 \frac{1}{N} && \text{by Corollary 16.1 page 132} \\
&\le \sigma^2 \sum_{n=1}^{N} \lambda_n^2 && \text{by Lemma 16.1 page 133} \\
&= \text{mse}(\text{arithmetic mean}) && \text{by Theorem 16.1 page 131}
\end{aligned}$$

✏

Example 16.1. The R code to the right calculates the *average* (Definition Q.4 page 351) number of sunspots from 1749 into 2013 (Figure 16.1 page 134).

```
data(sunspots);
x    = sunspot.month;
N    = length(x);
avg  = sum(x)  / N;
print(avg);
plot  (x,  col="blue");
lines(c(start(x)[1],end(x)[1]),  c(avg,avg),  col="red");
```

16.2 Variance Estimation

If we know the true *mean* μ of a stationary random process $(\!(x_n)\!)$, then a reasonable estimate of the variance might be $\frac{1}{N}\sum_{n=1}^{N}(x_n - \mu)^2$. This estimate has the highly touted property of being *unbiased*:

$$\mathbf{E}\left[\frac{1}{N}\sum_{n=1}^{N}(x_n-\mu)^2\right] = \frac{1}{N}\sum_{n=1}^{N}\mathbf{E}\left[(x_n-\mu)^2\right] = \frac{1}{N}\sum_{n=1}^{N}\sigma^2 = \sigma^2$$

Very good. However, in many cases we don't know the **true mean** μ, but rather only have an **estimated mean** $\hat{\mu} \triangleq \frac{1}{N} = \sum_{n=1}^{N} \mathsf{x}_n$. In this case, substituting in the estimated mean for the true mean as in $\hat{\text{var}}_B(⦅\mathsf{x}_n⦆)$ (next definition) yields a *biased* variance estimate (next theorem).

Definition 16.2. [3] *Let $\hat{\mu}$ be an estimate of the mean of a random sequence $⦅\mathsf{x}_n⦆$.*

DEF

*The **sample variance** $\hat{\text{var}}(⦅\mathsf{x}_n⦆)$ is defined as*
$$\hat{\text{var}}(⦅\mathsf{x}_n⦆) \triangleq \frac{1}{N-1}\sum_{n=1}^{N}(\mathsf{x}_n - \hat{\mu})^2$$

*The **biased sample variance** $\hat{\text{var}}_B(⦅\mathsf{x}_n⦆)$ is here defined as*
$$\hat{\text{var}}_B(⦅\mathsf{x}_n⦆) \triangleq \frac{1}{N}\sum_{n=1}^{N}(\mathsf{x}_n - \hat{\mu})^2$$

The factor $\frac{N}{N-1}$ such that $\hat{\text{var}}(⦅\mathsf{x}_n⦆) = \frac{N}{N-1}\hat{\text{var}}_B(⦅\mathsf{x}_n⦆)$ is known as "*Bessel's correction*". Why such "correction" would be useful at all is demonstrated by Theorem 16.3 (next). Theorem 16.3 demonstrates that the *biased sample variance* $\hat{\text{var}}_B(⦅\mathsf{x}_n⦆)$ is *biased*, and multiplication by $\frac{N}{N-1}$ makes it *unbiased*.

Theorem 16.3. [4] *Let $\hat{\mu}$ be the* AVERAGE (Definition Q.4 page 351) *of a sequence $⦅\mathsf{x}_n⦆$. Let $\mu^4 \triangleq \mathbf{E}\left[(\mathsf{x}_n - \mu)^4\right]$ be the* 4TH CENTRAL MOMENT *of x_n.*

THM

$$\left.\begin{array}{lll} \text{(A).} & ⦅\mathsf{x}_n⦆ \text{ is WSS} & \text{and} \\ \text{(B).} & \mu \triangleq \mathbf{E}\mathsf{x}_n & \text{and} \\ \text{(C).} & ⦅\mathsf{x}_n⦆ \text{ is UNCORRELATED} & \end{array}\right\} \implies \left\{\begin{array}{lrcl ll} \text{(1).} & \mathbf{E}\hat{\text{var}}_B(⦅\mathsf{x}_n⦆) &=& \frac{N-1}{N}\sigma^2 & \text{(BIASED)} & \text{and} \\ \text{(2).} & \mathbf{E}\hat{\text{var}}(⦅\mathsf{x}_n⦆) &=& \sigma^2 & \text{(UNBIASED)} & \text{and} \\ \text{(3).} & \text{var}\left[\hat{\text{var}}(⦅\mathsf{x}_n⦆)\right] &=& \frac{1}{N}\left[\mu^4 - \left(\frac{N-3}{N-1}\right)\sigma^4\right] & \text{(CONSISTENT)} & \end{array}\right.$$

✎PROOF:

1. lemma: $\mathbf{E}(\mathsf{x}_n\hat{\mu}) = \frac{1}{N}\sigma^2 + \mu^2$. Proof:

$$\begin{aligned}
\mathbf{E}(\mathsf{x}_n\hat{\mu}) &\triangleq \mathbf{E}\left(\mathsf{x}_n\frac{1}{N}\sum_{m=1}^{N}\mathsf{x}_m\right) && \text{by definition of } \textit{average} && \text{(Definition Q.4 page 351)} \\
&= \mathbf{E}\left(\frac{1}{N}\sum_{m=1}^{N}\mathsf{x}_n\mathsf{x}_m\right) \\
&= \frac{1}{N}\sum_{m=1}^{N}\mathbf{E}(\mathsf{x}_n\mathsf{x}_m) && \text{by } \textit{linearity} \text{ of } \mathbf{E} && \text{(Theorem 1.1 page 4)} \\
&= \frac{1}{N}\left[\mathbf{E}\mathsf{x}_n^2 + \sum_{m\neq n}\mathbf{E}(\mathsf{x}_n\mathsf{x}_m)\right] \\
&= \frac{1}{N}\left[\mathbf{E}\mathsf{x}_n^2 + \sum_{m\neq n}(\mathbf{E}\mathsf{x}_n)(\mathbf{E}\mathsf{x}_m)\right] && \text{by } \textit{uncorrelated} \text{ hypothesis} && \text{(C)} \\
&= \frac{1}{N}\left[(\sigma^2+\mu^2) + (N-1)\mu^2\right] && \text{by Corollary 1.3 (page 5)} \\
&= \frac{1}{N}\sigma^2 + \mu^2
\end{aligned}$$

[3] Wilks (1963a) page 199 ⟨§"8.2 MEANS AND VARIANCES OF MEAN, VARIANCE,..."⟩, WILKS (1963B) PAGE 199 ⟨§"(B) MEAN AND VARIANCE OF SAMPLE VARIANCE"⟩, KENNEY (1947) PAGE 125 ⟨"BESSEL'S CORRECTION"⟩, BAJPAI (1967) PAGE 509 ⟨???⟩

[4] Wilks (1963a) page 199 ⟨§"8.2 MEANS AND VARIANCES OF MEAN, VARIANCE,..."⟩, TUCKER (1965) PAGE 111 ⟨§"8.2 UNBIASED AND CONSISTENT ESTIMATES"⟩, STUART AND ORD (1991) PAGE 609 ⟨§"UNBIASED ESTIMATORS"⟩

2. Proof for (1):

$$\begin{aligned}
\mathbf{E}\hat{\text{var}}_B\left(\left(\!\left(\mathsf{x}_n\right)\!\right)\right) &\triangleq \mathbf{E}\left[\frac{1}{N}\sum_{n=1}^{N}\left(\mathsf{x}_n-\hat{\mu}\right)^2\right] && \text{by definition of } \hat{\text{var}}_B && \text{(Definition 16.2 page 135)}\\
&= \frac{1}{N}\mathbf{E}\left[\sum_{n=1}^{N}\left(\mathsf{x}_n\underbrace{-\mu+\mu}_{0}-\hat{\mu}\right)^2\right]\\
&= \frac{1}{N}\sum_{n=1}^{N}\left[\underbrace{\mathbf{E}\left(\mathsf{x}_n-\mu\right)^2}_{\sigma^2}+2\mathbf{E}\left[(\mathsf{x}_n-\mu)(\mu-\hat{\mu})\right]+\mathbf{E}\left(\mu-\underbrace{\hat{\mu}}_{\mathbf{E}\mu}\right)^2\right]\\
&= \frac{1}{N}\sum_{n=1}^{N}\left[\sigma^2+2\mathbf{E}\left[\mathsf{x}_n\mu-\mathsf{x}_n\hat{\mu}-\mu^2+\mu\hat{\mu}\right]+\frac{1}{N}\sigma^2\right]\\
&= \frac{1}{N}\sum_{n=1}^{N}\left[\sigma^2+2\left[\mu^2-\mathbf{E}(\mathsf{x}_n\hat{\mu})-\mu^2+\mu^2\right]+\frac{1}{N}\sigma^2\right] && \text{by Corollary 16.1 page 132}\\
&= \frac{1}{N}\sum_{n=1}^{N}\left[\sigma^2+2\left[\mu^2-\left(\mu^2+\frac{1}{N}\sigma^2\right)\right]+\frac{1}{N}\sigma^2\right] && \text{by \textit{unbiased} prop. of } \hat{\mu} && \text{((1) lemma page 135)}\\
&= \frac{1}{N}\sum_{n=1}^{N}\left[\sigma^2-\frac{1}{N}\sigma^2\right]\\
&= \frac{N-1}{N}\sigma^2
\end{aligned}$$

3. Proof for (2):

$$\begin{aligned}
\mathbf{E}\hat{\text{var}}\left(\left(\!\left(\mathsf{x}_n\right)\!\right)\right) &\triangleq \mathbf{E}\left[\frac{1}{N-1}\sum_{n=1}^{N}\left(\mathsf{x}_n-\hat{\mu}\right)^2\right] && \text{by definition of } \hat{\text{var}} && \text{(Definition 16.2 page 135)}\\
&= \frac{N}{N-1}\mathbf{E}\left[\frac{1}{N}\sum_{n=1}^{N-1}\left(\mathsf{x}_n-\hat{\mu}\right)^2\right] && \text{by \textit{linearity} of } \mathbf{E} && \text{(Theorem 1.1 page 4)}\\
&= \frac{N}{N-1}\left[\frac{N-1}{N}\sigma^2\right] && \text{by } \hat{\text{var}}_B \text{ result} && \text{(item (2) page 136)}\\
&= \sigma^2
\end{aligned}$$

4. lemma: $\mathbf{E}\left[\left(\hat{\text{var}}\right)^2\right] = \frac{\mu^4}{N}+\frac{(N-1)^2+2}{N(N-1)}\sigma^4$. Proof: No proof here at this time. The assertion is made by Wilks (1963a) page 199 who also without there supplying a proof says, "Carrying out similar mean value operations we find after some reduction that" the result follows.

$$\begin{aligned}
\mathbf{E}\left[\left(\hat{\text{var}}\right)^2\right] &\triangleq \mathbf{E}\left[\left(\frac{1}{N-1}\sum_{n=1}^{N}\left(\mathsf{x}_n-\hat{\mu}\right)^2\right)^2\right] && \text{by definition of } \hat{\text{var}} \text{ (Definition 16.2 page 135)}\\
&= \mathbf{E}\left[\left(\frac{1}{N-1}\sum_{n=1}^{N}\left(\mathsf{x}_n-\hat{\mu}\right)^2\right)\left(\frac{1}{N-1}\sum_{m=1}^{N}\left(\mathsf{x}_m-\hat{\mu}\right)^2\right)\right]\\
&= \left(\frac{1}{N-1}\right)^2\mathbf{E}\left[\sum_{n=1}^{N}\sum_{m=1}^{N}\left(\mathsf{x}_n-\hat{\mu}\right)^2\left(\mathsf{x}_m-\hat{\mu}\right)^2\right]\\
&= \left(\frac{1}{N-1}\right)^2\mathbf{E}\left[\sum_{n=1}^{N}\sum_{m=1}^{N}\left(\mathsf{x}_n^2-2\mathsf{x}_n\hat{\mu}+\hat{\mu}^2\right)\left(\mathsf{x}_m^2-2\mathsf{x}_m\hat{\mu}+\hat{\mu}^2\right)\right]
\end{aligned}$$

$$= \left(\frac{1}{N-1}\right)^2 \mathbf{E}\left[\sum_{n=1}^{N}\sum_{m=1}^{N}\left[\begin{array}{l}(\mathsf{x}_n^2\mathsf{x}_m^2 - 2\hat{\mu}\mathsf{x}_n^2\mathsf{x}_m + \hat{\mu}^2\mathsf{x}_n^2) + (-2\hat{\mu}\mathsf{x}_n\mathsf{x}_m^2 + 4\hat{\mu}^2\mathsf{x}_n\mathsf{x}_m - 2\mathsf{x}_n\hat{\mu}^3)\\ +(\hat{\mu}^2\mathsf{x}_m^2 - 2\hat{\mu}^3\mathsf{x}_m + \hat{\mu}^4)\end{array}\right]\right]$$

$$= \left(\frac{1}{N-1}\right)^2 \sum_{n=1}^{N}\sum_{m=1}^{N}\left[\begin{array}{l}\mathbf{E}\left[\mathsf{x}_n^2\mathsf{x}_m^2\right] - 2\mathbf{E}\left[\hat{\mu}\mathsf{x}_n^2\mathsf{x}_m\right] - 2\mathbf{E}\left[\hat{\mu}\mathsf{x}_n\mathsf{x}_m^2\right] + \mathbf{E}\left[\hat{\mu}^2\mathsf{x}_n^2\right] + \mathbf{E}\left[\hat{\mu}^2\mathsf{x}_m^2\right]\\ +4\,\mathbf{E}\left[\hat{\mu}^2\mathsf{x}_n\mathsf{x}_m\right] - 2\mathbf{E}\left[\hat{\mu}^3\mathsf{x}_n\right] - 2\mathbf{E}\left[\hat{\mu}^3\mathsf{x}_m\right] + \mathbf{E}\left[\hat{\mu}^4\right]\end{array}\right]$$

$$= \left(\frac{1}{N-1}\right)^2 \sum_{n=1}^{N}\sum_{m=1}^{N}\left[\begin{array}{l}\mathbf{E}\left[\mathsf{x}_n^2\mathsf{x}_m^2\right] - 4\mathbf{E}\left[\hat{\mu}\mathsf{x}_n^2\mathsf{x}_m\right] + 2\mathbf{E}\left[\hat{\mu}^2\mathsf{x}_n^2\right]\\ +4\,\mathbf{E}\left[\hat{\mu}^2\mathsf{x}_n\mathsf{x}_m\right] - 4\mathbf{E}\left[\hat{\mu}^3\mathsf{x}_n\right] + \mathbf{E}\left[\hat{\mu}^4\right]\end{array}\right]$$

$$\stackrel{?}{=} \frac{\mu^4}{N} + \frac{(N-1)^2+2}{N(N-1)}\sigma^4$$

5. Proof for (3):

$$\begin{aligned}
\operatorname{var}\left[\hat{\operatorname{var}}(\!(\mathsf{x}_n)\!)\right] &\triangleq \mathbf{E}\left[\left(\hat{\operatorname{var}} - \mathbf{E}\hat{\operatorname{var}}\right)^2\right] && \text{by definition of } \hat{\operatorname{var}} && \text{(Definition 16.2 page 135)}\\
&= \mathbf{E}\left[\left(\hat{\operatorname{var}} - \sigma^2\right)^2\right] && \text{by (2)} && \text{(item (3) page 136)}\\
&= \mathbf{E}\left[\left(\hat{\operatorname{var}}\right)^2 - 2\sigma^2\hat{\operatorname{var}} + \left(\sigma^2\right)^2\right] && \text{by } \textit{Binomial Theorem}\\
&= \mathbf{E}\left[\left(\hat{\operatorname{var}}\right)^2\right] - 2\sigma^2\mathbf{E}\left[\hat{\operatorname{var}}\right] + \mathbf{E}\left[\left(\sigma^2\right)^2\right] && \text{by } \textit{linearity} \text{ of } \mathbf{E} && \text{(Theorem 1.1 page 4)}\\
&= \mathbf{E}\left[\left(\hat{\operatorname{var}}\right)^2\right] - 2\left(\sigma^2\right)^2 + \left(\sigma^2\right)^2 && \text{by (2)} && \text{(item (3) page 136)}\\
&= \left[\frac{\mu^4}{N} + \left(\frac{(N-1)^2+2}{N(N-1)}\right)\sigma^4\right] - \sigma^4 && \text{by (4) lemma}\\
&= \frac{1}{N}\left[\mu^4 + \left(\frac{(N^2-N+1)+2-N(N-1)}{N-1}\right)\sigma^4\right]\\
&= \frac{1}{N}\left[\mu^4 - \left(\frac{N-3}{N-1}\right)\sigma^4\right]
\end{aligned}$$

✎

16.3 Estimates in terms of moment estimates

Definition 16.3.

DEF *The **order-k moment estimate** is here defined as*

$$\hat{\mathsf{M}}_k(\!(\mathsf{x}_n)\!) \triangleq \frac{1}{N}\sum_{n=1}^{N}\mathsf{x}_n^k$$

Proposition 16.1.

PRP

$$\begin{aligned}
\hat{\mu}(\!(\mathsf{x}_n)\!) &= \hat{\mathsf{M}}_1\\
\hat{\operatorname{var}}(\!(\mathsf{x}_n)\!) &= \frac{N}{N-1}\hat{\mathsf{M}}_2 - \frac{N}{N-1}\hat{\mathsf{M}}_1^2
\end{aligned}$$

✎PROOF:

$$\begin{aligned}
\hat{\mu}(\mathsf{x}_n) &\triangleq \frac{1}{N}\sum_{n=1}^{N}\mathsf{x}_n && \text{by definition of } \hat{\mu} && \text{(Definition Q.4 page 351)}\\
&\triangleq \hat{\mathsf{M}}_1 && \text{by definition of } \hat{\mathsf{M}}_1 && \text{(Definition 16.3 page 137)}
\end{aligned}$$

$$\begin{aligned}
\hat{\mathrm{var}}(\mathsf{x}_n) &\triangleq \frac{1}{N-1}\sum_{n=1}^{N}\left(\mathsf{x}_n-\hat{\mu}\right)^2 && \text{by definition of } \hat{\mathrm{var}} && \text{(Definition 16.2 page 135)}\\
&= \frac{1}{N-1}\sum_{n=1}^{N}\left(\mathsf{x}_n^2-2\mathsf{x}_n\hat{\mu}+(\hat{\mu})^2\right)\\
&= \frac{1}{N-1}\sum_{n=1}^{N}\mathsf{x}_n^2-2\frac{1}{N-1}\sum_{n=1}^{N}\mathsf{x}_n\hat{\mu}+\frac{1}{N-1}\sum_{n=1}^{N}(\hat{\mu})^2\\
&= \frac{N}{N-1}\underbrace{\frac{1}{N}\sum_{n=1}^{N}\mathsf{x}_n^2}_{\hat{\mathsf{M}}_2}-2\underbrace{\hat{\mu}}_{\hat{\mathsf{M}}_1}\frac{N}{N-1}\underbrace{\frac{1}{N}\sum_{n=1}^{N}\mathsf{x}_n}_{\hat{\mathsf{M}}_1}+\frac{N}{N-1}\underbrace{(\hat{\mu})^2}_{\hat{\mathsf{M}}_1^2}\\
&= \frac{N}{N-1}\hat{\mathsf{M}}_2-\frac{N}{N-1}\hat{\mathsf{M}}_1^2
\end{aligned}$$

✏

16.4 Recursive forms

In software/firmware implementations, recursive forms are very useful and efficient.

Proposition 16.2. [5]

PRP

$$\underbrace{\hat{\mu}_N}_{new}=\underbrace{\hat{\mu}_{N-1}}_{previous}+\underbrace{\frac{1}{N}}_{weight}\underbrace{\left[\mathsf{x}_N-\hat{\mu}_{N-1}\right]}_{error}$$

✎PROOF:

$$\begin{aligned}
\hat{\mu}_N &\triangleq \frac{1}{N}\sum_{n=1}^{N}\mathsf{x}_n && \text{by definition of } \textit{average} && \text{(Definition Q.4 page 351)}\\
&= \frac{1}{N}\mathsf{x}_N+\frac{1}{N}\sum_{n=1}^{N-1}\mathsf{x}_n\\
&= \frac{1}{N}\mathsf{x}_N+\frac{N-1}{N}\left(\frac{1}{N-1}\right)\sum_{n=1}^{N-1}\mathsf{x}_n\\
&\triangleq \frac{1}{N}\mathsf{x}_N+\frac{N-1}{N}\hat{\mu}_{N-1} && \text{by definition of } \textit{average} && \text{(Definition Q.4 page 351)}\\
&= \frac{1}{N}\mathsf{x}_N+\hat{\mu}_{N-1}-\frac{1}{N}\hat{\mu}_{N-1}\\
&= \hat{\mu}_{N-1}+\frac{1}{N}\left[\mathsf{x}_N-\hat{\mu}_{N-1}\right]
\end{aligned}$$

✏

[5] Candy (2009) pages 11–12 ⟨Example 1.3⟩, Candy (2016) pages 12–13 ⟨Example 1.3⟩

CHAPTER 17

DENSITY ESTIMATION

17.1 Introduction

Moment estimation is extremely useful for model-building when a *type* of parameterized *probability density function* (*pdf*) $\mathsf{p}_{\mathsf{x}}(x)$ (Definition B.2 page 184) is known or more likely *assumed*, but the *parameters* themselves are *not known*. Examples of parameterized pdfs include the following:

- The *Uniform Distribution* with parameter mean μ (Definition C.1 page 189).
- The *Gamma Distribution* with parameter b (Definition C.3 page 191).
- The *Chi-square Distribution* with parameter variance σ^2 (Definition C.5 page 192).
- The *Gaussian Distribution* with parameters mean μ and variance σ^2 (Definition C.2 page 190).
- The *Rayleigh Distribution* with parameters radius r and variance σ^2 (Definition C.10 page 195).

In the case of distributions with mean and/or variance parameters, these parameters can be estimated using the techniques in this chapter, and the technique of estimating the pdf in this manner is called **parametric density estimation**.[1] If a distribution type is not known and not assumed, then the distribution itself must be estimated using **nonparametric density estimation** (Chapter 17 page 139)

17.2 Parametric density estimation

Corollary 17.1.

COR

$$\left\{\begin{array}{lll}\text{(A).} & (\!(\mathsf{x}_n)\!) \text{ is UNCORRELATED} & \text{and}\\ \text{(B).} & \mathsf{x}_n \text{ is GAUSSIAN} & \end{array}\right\} \implies \left\{\begin{array}{lll}\text{(1).} & \hat{\mu}_{\mathsf{ml}} = \dfrac{1}{N}\displaystyle\sum_{n=1}^{N}\mathsf{x}_n & \text{and}\\ \text{(2).} & \hat{\mu}_{\mathsf{ml}} \text{ is CONSISTENT} & \text{and}\\ \text{(3).} & \hat{\mu}_{\mathsf{ml}} \text{ is EFFICIENT} & \end{array}\right\}$$

✎PROOF: This result follows directly from Theorem 21.9 (page 161) with

[1] Gramacki (2017) page 2 ⟨§"1.1 Background"⟩

$$\begin{aligned}
\mathsf{y}(t) &\triangleq \mathsf{x}(t;\theta) + \mathsf{v}(t) && \text{where } \mathsf{v}(t) \text{ is a \emph{zero-mean} white Gaussian noise process}\\
\mathsf{x}(t;\theta) &\triangleq \mathsf{g}(\theta)\\
&\triangleq \theta\\
&\triangleq \mu\\
\mathsf{x}_n &\triangleq \dot{y}_n\\
&\triangleq \langle \mathsf{y}(t) \,|\, \psi_n(t)\rangle\\
&\triangleq \langle \mathsf{y}(t) \,|\, \delta(t-n\tau)\rangle\\
&\triangleq \int_{t\in\mathbb{R}} \mathsf{y}(t)\delta(t-n\tau)\,\mathrm{d}t\\
&= \mathsf{y}(n\tau)
\end{aligned}$$

Alternatively, the results follow from Theorem 21.7 (page 157). ⇒

17.3 Nonparametric density estimation

Some techniques of *nonparametric density estimation* include[2]

- Histogram
- Average Shifted Histogram
- *Kernel Density Estimation* (*KDE*)
- Fourier-based
- Wavelet-based
- Fast Gauss Transform

KDE's are conceptually quite simple, but do have some difficulties:

- KDEs are good for 6 dimensions or less—the "curse of dimensionality" (Chapter 20 page 147).
- KDEs have a *bandwidth* parameter and the computation of this parameter can be quite demanding.[3]

For cases where the dimension is 6 or greater, one option is to perform *data dimensionality reduction* (Chapter 20 page 147).

17.4 Smoothness

For a given stochastic process, an infinite number of density estimates exist. Some constraints on inclusion into the estimate family are in order. One contraint is that the density estimate is not "too wiggly".[4] Mathematically, this may expressed as requiring the estimate to be a member of a **Sobelov Space** S

$$S \triangleq \left\{ \mathsf{f} \,\middle|\, \int_{x\in\mathbb{R}} [\mathsf{f}''(x)]^2\,\mathrm{d}x < r^2 \right\}$$

where r is some constant in $\mathbb{R}$ or where r^2 is ∞.

[2] Silverman (1986) ⟨ISBN:9780412246203⟩, Tsybakov (2008) pages 1–27 ⟨§1.1–§1.3⟩, Gramacki (2017) ⟨ISBN:9783319716886⟩, Vidakovic (1999) pages 217–245 ⟨Chapter 7 "Density Estimation"⟩

[3] Gramacki (2017) page 4, 63–83 ⟨Chapter 4 "Bandwidth Selectors for Kernel Density Estimation"⟩

[4] Wasserman (2013) page 88 ⟨"6.5 Example (Nonparametric density estimation)."⟩

CHAPTER 18

CORRELATION ESTIMATION

18.1 Time series method

Definition 18.1. [1]

DEF

*The **windowed auto-correlation estimate** $\hat{\mathsf{R}}_{xx}(m)$ is defined as*

$$\hat{\mathsf{R}}_{xx}(m) \triangleq \frac{1}{N}\sum_{n=0}^{N-|m|} \mathsf{x}(n)\mathsf{x}(n+m)$$

Theorem 18.1. [2]

THM

$$\begin{aligned}
\mathbf{E}\left[\hat{\mathsf{R}}_{xx}(m)\right] &= \left(1-\frac{|m|}{N}\right)\mathsf{R}_{xx}(m) && \text{(ASYMPTOTICALLY UNBIASED)}\\
\operatorname{var}\left[\hat{\mathsf{R}}_{xx}(m)\right] &= \frac{1}{N}\sum_{n\in\mathbb{Z}}\left[\mathsf{R}^2_{xx}(n)+\mathsf{R}_{xx}(n-m)\mathsf{R}_{xx}(n+m)\right] && \text{(CONSISTENT)}
\end{aligned}$$

18.2 Spectral methods

Here are two methods for estimating correlation using spectral methods:

1. Calculate the estimate $\hat{\mathsf{H}}(\omega)$ (CHAPTER 14 page 103) of the true system model $\mathsf{H}(\omega)$ and then calculate the *Inverse Fourier Transform* of $\hat{\mathsf{H}}(\omega)$.

2. Calculate an *AR*, *MA*, or *ARMA* estimate of the true system model $\mathsf{H}(z)$ and then compute the *Inverse Z-Transform* of the estimate.[3]

[1] Jenkins and Watts (1968) pages 180–183 ⟨§“5.3.4 Discrete time autocovariance estimates”⟩, Vaseghi (2000) page 271 ⟨§“9.3.3 Energy-Spectral Density and Power-Spectral Density”⟩

[2] Jenkins and Watts (1968), Clarkson (1993) pages 54–56 ⟨§“2.6.2 Estimation of Moments — Time Averages”⟩, Vaseghi (2000) page 272 ⟨§“9.3.3 Energy-Spectral Density and Power-Spectral Density”⟩

[3] Clarkson (1993) pages 56–59 ⟨§“Appendix 2A — Calculating the Correlation via Contour Integration”⟩

CHAPTER 19

SPECTRAL ESTIMATION

> *"Sometimes we come to life's crossroads*
> *and view what we think is the end.*
> *But God has a much wider vision*
> *and He knows that it's only a bend...."*
>
> Helen Steiner Rice (1900–1981); from *A Bend in the Road*[1]

19.1 Nonparametric spectral estimation

Quality of spectral estimators[2]

THM

Periodogram:	Q	$=$	1
Welch Method 0% overlap:	Q	$=$	$0.78N\Delta f$
Welch Method 50% overlap:	Q	$=$	$1.39N\Delta f$
Bartlett Method:	Q	$=$	$1.11N\Delta f$
Blackman-Tukey Method:	Q	$=$	$2.34N\Delta f$

Example 19.1. [3] Using the PSD, we can estimate the periodicity of sunspots (Figure 19.1 page 144) as

---

[1] Rice (2012) page 116 ⟨Note: not sure about page number⟩. Many many thanks to Mary L. Greenhoe for bringing this poem to the author's attention.

[2] Proakis (2002) pages 452–457 ⟨§"8.2.4 Performance Characteristics of Nonparametric Power Spectrum Estimators"⟩, Proakis and Manolakis (1996) pages 916–919 ⟨§"12.2.4 Performance Characteristics of Nonparametric Power Spectrum Estimators"⟩, Rao and Swamy (2018) page 731 ⟨"Table 12.1 Comparison of performance of classical methods"⟩, Salivahanan and Vallavaraj (2001) page 606 ⟨§"12.5 Power Spectrum Estimation: Non-Parametric Methods"⟩, Ifeachor and Jervis (2002) pages 706–707 ⟨§"11.3.7 Comparison of the power spectral density estimation methods"⟩, J.S.Chitode (2009b) page P-100, Abdaheer (2009) page 204

[3] SILSO World Data Center (2019) ⟨`http://www.sidc.be/silso/DATA/SN_m_tot_V2.0.txt`⟩, rdocumentation.org (2013)

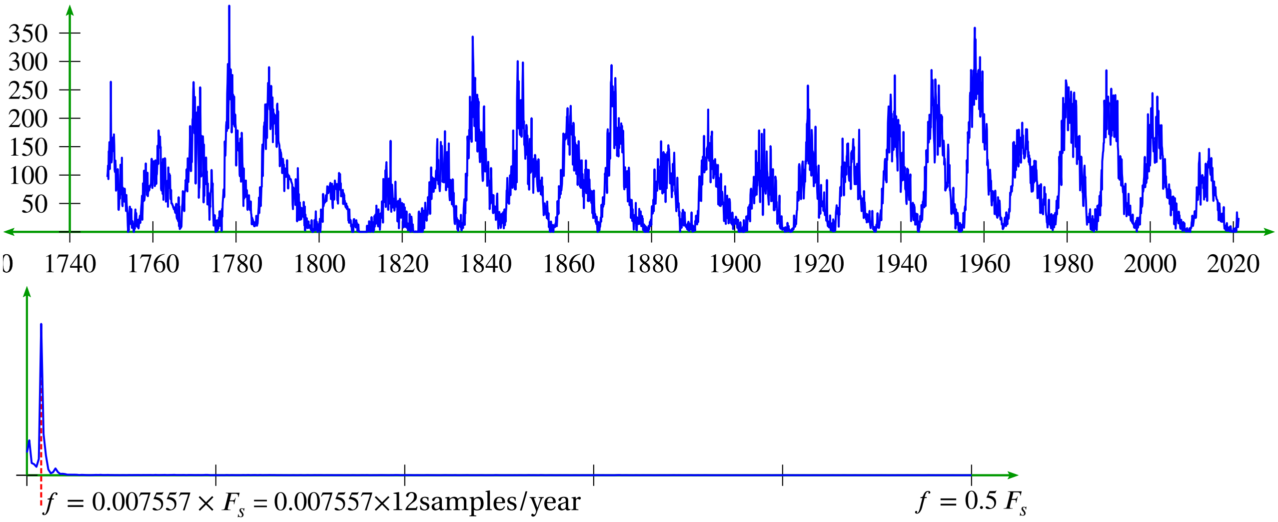


Figure 19.1: Estimation of sunspot periodicity using Welch PSD estimation

EX	number of segments = 2	estimated periodicty = 11.0277777777777786 years
	number of segments = 3	estimated periodicty = 11.0208333333333321 years
	number of segments = 4	estimated periodicty = 11.0277777777777786 years
	number of segments = 8	estimated periodicty = 11.0000000000000000 years

That is, period $= \frac{1}{f} = \frac{1}{0.007557\times 12} \approx 11.0273$ years. Here is some R code that can be used to experiment with:

```
require(stats);
require(graphics);
require(datasets);
require(ramify);
require(bspec);
data(sunspots, package="datasets");
mypsd = function(x, numSegments=4)
  {
  xts       = as.ts(as.vector(x)); # year indices seems to confuse welchPSD
  N         = length(xts);         # length of time series
  Fs        = 12;                  # sample rate = 12 samples per year
  estMean   = mean(xts);           # estimated mean
  segLength = N / numSegments;     # segment length
  xpsd      = bspec::welchPSD(xts - estMean, seglength = segLength);
  psdMax    = max(xpsd$power);     # maximum PSD value
  binMax    = ramify::argmax(as.matrix(xpsd$power), rows = FALSE);
  freqMax   = xpsd$frequency[binMax] * Fs; # dominate non-DC frequency
  periodT   = 1 / freqMax;         # estimated period
  plot(xpsd$power, type="b", xlim=c(1,50), ylim=c(0,psdMax), col="blue");
  periodT;                         # return estimated period
  }
 estT = mypsd(x, numSegments=4);
 print(sprintf("estimated period = %.16f years", estT));
```

19.2 Bandwidth-Time Product

BT-product references:

1. Haykin (2014) pages 25–28 ⟨§"2.4 The Inverse Relationship between Time-Domain and Frequency-Domain Representations"⟩,
2. S. Lawrence Marple (1987) pages 144–146 ⟨§"5.4 Resolution and the Stability-Time-Bandwidth Product"⟩

Part V

Dimensionality Reduction

CHAPTER 20

DIMENSIONALITY REDUCTION

20.1 Introduction

A number of statistical estimation algorithms yield subspace structures with high dimensionality—one key example is *Kernel Density Estimation* (*KDE*)(Chapter 17 page 139). Richard Bellman described this problem as the "curse of dimensionality".[1]

Actually, a *time series* is a point in infinite dimensional space.[2] Statistical analysis is much to do with approximating this point in low dimensional space.

Example 20.1. [3] An example of a one-dimensional representation is the *estimated mean* or average of the time series.

Example 20.2. [4] In the case where the data is (or assumed to be) IID Gaussian, the point can be represented in two dimensional space with the dimensions being mean and variance.

References discussing *dimensionality reduction* include the following:

Cunningham and Ghahramani (2015):	"*Linear Dimensionality Reduction:...*"
Sorzano et al. (2014):	"*A survey of dimensionality reduction techniques*"
Lee and Verleysen (2007):	"*Nonlinear Dimensionality Reduction*"

20.2 Principal Component Analysis

A very common algorithm for dimensionality reduction is *Principal Component Analysis* (*PCA*).[5]

[1] Bellman (1954) page 206, Bellman (1961) page 94 ⟨§"5.16 The Curse of Dimensionality"⟩, Bellman (1971) page 44, Bishop (2006) pages 33–38 ⟨§1.4 "The Curse of Dimensionality"⟩, Gramacki (2017) page 3, 59 ⟨§3.9 "The Curse of Dimensionality"⟩

[2] Wasserman (2006) page 1 ⟨§"1.1 What Is Nonparametric Inference?"⟩

[3] Wasserman (2013) pages 88–89 ⟨"6.5 Example (Nonparametric estimation of functionals)."⟩

[4] Wasserman (2013) page 88 ⟨"6.2 Example (Two-dimensional Parametric Estimation)"⟩

[5] Pearson (1901), Hotelling (1933), Eckart and Young (1936), Kendall (1968) pages 10–36 ⟨"2. Component Analysis"⟩, Jeffers (1967), Jolliffe (2013) ⟨ISBN:9781475719048⟩

Example 20.3. [6] Data is often *multi-variate* and represented in tabular form with each column representing a variable and rows the samples. Such a table is basically a matrix representing high dimensional space. We can perform an *eigen-decomposition* on this matrix. Many eigenvalues may result. However, we can retain the largest eigenvalues (and their associated eigenvectors) and set the remaining eignevalues to 0. Estimation can then be performed using just the non-zero eigen-pairs (assumed to represent the true "signal") while ignoring the zeroed-out eigen-pairs (assumed to represent the "noise"). This is in essence the technique of *Principal Component Analysis*.

Remark 20.1. Despite the description of *PCA* in Example 20.3, it is *not* to say that small eigenvalues are never important. For several counter-examples demonstrating the importance of small eigenvalues, see Jolliffe (1982).

[6] Carmona (2013) pages 171–180 ⟨§"3.5 Principal Component Analysis"⟩

CHAPTER 21

PROJECTION STATISTIC ALGORITHMS FOR ADDITIVE NOISE MODELS

21.1 Projection Statistics

Theorem 21.1 (page 151) (next) shows that the finite set $Y \triangleq \left\{\dot{y}_n | n = 1, 2, \ldots, N\right\}$ (a finite number of values) provides just as good an estimate as having the entire $\mathsf{y}(t;\theta)$ waveform (an uncountably infinite number of values) with respect to the following cases:

1. the conditional probability of $\mathsf{x}(t;\theta)$ given $\mathsf{y}(t;\theta)$
2. the *MAP estimate* of the sequence
3. the *ML estimate* of the sequence.

Thus even with a drastic reduction in the number of statistics from uncountably infinite to finite N, no quality is lost with respect to the estimators listed above—that is, these statistics are *sufficient*, or are "*sufficient statistics*". This amazing result is very useful in practical system implementation and also for proving other theoretical results (notably estimation and detection theorems).

The concept of sufficient statistics in "estimation theory" is very similar to the concept of *feature extraction*[1] in "machine learning". Thus, estimation theory *is* a form of machine learning.

Definition 21.1. *Let* $\Psi \triangleq \left\{\psi_n | n = 1, 2, \ldots, N\right\}$ *be an* ORTHONORMAL BASIS *for a parameterized function* $\mathsf{x}(t;\theta)$ *with parameter* θ. *Let* $\mathsf{y}(t;\theta)$ *be* $\mathsf{x}(t;\theta)$ *plus a* RANDOM PROCESS $\mathsf{v}(t)$ *such that*

$$\mathsf{y}(t;\theta) \triangleq \mathsf{x}(t;\theta) + \mathsf{v}(t)$$

Let $\dot{y}_n$, $\dot{x}_n$, *and* $\dot{v}_n$ *be* PROJECTIONS *(Definition R.7 page 374) onto the* BASIS VECTOR $\psi_n(t)$ *such that*

DEF

$$\begin{aligned}
\dot{y}_n(\theta) &\triangleq \mathbf{P}_n\mathsf{y}(t;\theta) &&\triangleq \langle \mathsf{y}(t;\theta) \,|\, \psi_n(t)\rangle &&\triangleq \int_{t\in\mathbb{R}} \mathsf{y}(t;\theta)\psi_n(t)\,\mathsf{d}t \\
\dot{x}_n(\theta) &\triangleq \mathbf{P}_n\mathsf{x}(t) &&\triangleq \langle \mathsf{x}(t;\theta) \,|\, \psi_n(t)\rangle &&\triangleq \int_{t\in\mathbb{R}} \mathsf{x}(t;\theta)\psi_n(t)\,\mathsf{d}t \\
\dot{v}_n &\triangleq \mathbf{P}_n\mathsf{v}(t) &&\triangleq \langle \mathsf{v}(t) \,|\, \psi_n(t)\rangle &&\triangleq \int_{t\in\mathbb{R}} \mathsf{v}(t)\psi_n(t)\,\mathsf{d}t
\end{aligned}$$

Let the set Y *be defined as* $Y \triangleq \left\{\dot{y}_n(\theta) | 1, 2, \ldots, N\right\}$ *Let* $\hat{\theta}_{\mathsf{map}}$ *be the* MAP ESTIMATE *and* $\hat{\theta}_{\mathsf{ml}}$ *be the* ML ESTIMATE *(Definition 10.1 page 81) of* θ.

[1] Bishop (2006) page 2 ⟨9780387310732⟩"1. INTRODUCTION"

Lemma 21.1. *Let Ψ, $\mathsf{v}(t)$, $\dot{v}_n$, and Y be defined as in Definition 21.1 (page 149).*

LEM
$$\left\{ \mathbf{E}\mathsf{v}(t) \;=\; 0 \quad (\text{ZERO-MEAN}) \right\} \implies \left\{ \mathbf{E}\dot{v}_n \;=\; 0 \quad (\text{ZERO-MEAN}) \right\}$$

✎PROOF:

$$\begin{aligned}
\mathbf{E}\dot{v}_n &= \mathbf{E}\left\langle \mathsf{v}(t) \mid \psi_n(t) \right\rangle && \text{by definition of } \dot{v}_n && \text{(Definition 21.1 page 149)}\\
&= \left\langle \mathbf{E}\mathsf{v}(t) \mid \psi_n(t) \right\rangle && \text{by } \textit{linearity} \text{ of } \langle \triangle \mid \nabla \rangle\\
&= \left\langle 0 \mid \psi_n(t) \right\rangle && \text{by } \textit{zero-mean} \text{ hypothesis}\\
&= 0
\end{aligned}$$

✑

Lemma 21.2. *Let Ψ, $\mathsf{v}(t)$, $\dot{v}_n$, and Y be defined as in Definition 21.1 (page 149).*

LEM
$$\left\{ \mathsf{v}(t) \;\sim\; \mathsf{N}\left(0, \sigma^2\right) \quad (\text{GAUSSIAN}) \right\} \implies \left\{ \dot{v}_n \;\sim\; \mathsf{N}\left(0, \sigma^2\right) \quad (\text{GAUSSIAN}) \right\}$$

✎PROOF: The distribution follows because it is a linear operation on a Gaussian process. ✑

Lemma 21.3. *Let Ψ, $\mathsf{y}(t;\theta)$, $\mathsf{x}(t)$, $\mathsf{v}(t)$, $\dot{y}_n$, $\dot{x}_n$, $\dot{v}_n$, and Y be defined as in Definition 21.1 (page 149).*

LEM
$$\left\{ \begin{array}{lrcll} (A). & \mathbf{E}[\mathsf{v}(t)] & = & 0 & \text{and}\\ (B). & \operatorname{cov}[\mathsf{v}(t), \mathsf{v}(u)] & = & \sigma^2\delta(t-u) & \end{array} \right\} \implies \left\{ \begin{array}{lrcll} (1). & \mathbf{E}\dot{v}_n & = & 0 & (\text{ZERO-MEAN})\\ (2). & \operatorname{cov}\left[\dot{v}_n, \dot{v}_m\right] & = & \sigma^2\bar{\delta}_{n-m} & (\text{UNCORRELATED}) \end{array} \right\}$$

✎PROOF:

1.

$$\mathbf{E}\dot{v}_n = 0 \qquad \text{by } \textit{additive} \text{ property and Theorem 21.2 page 154}$$

2.

$$\begin{aligned}
\operatorname{cov}\left[\dot{v}_m, \dot{v}_n\right] &= \operatorname{cov}\left[\left\langle \mathsf{v}(t) \mid \psi_m(t)\right\rangle, \left\langle \mathsf{v}(t) \mid \psi_n(t)\right\rangle\right] && \text{by def. of } \dot{v}_n && \text{(Definition 21.1 page 149)}\\
&= \operatorname{cov}\left[\left(\int_{t\in\mathbb{R}} \mathsf{v}(t)\psi_m(t)\,\mathrm{d}t\right), \left(\int_{u\in\mathbb{R}} \mathsf{v}(u)\psi_n(u)\,\mathrm{d}u\right)\right] && \text{by def. of } \langle \triangle \mid \nabla \rangle && \text{(Definition 21.1 page 149)}\\
&= \mathbf{E}\left[\left(\int_{t\in\mathbb{R}} \mathsf{v}(t)\psi_m(t)\,\mathrm{d}t\right)\left(\int_{u\in\mathbb{R}} \mathsf{v}(u)\psi_n(u)\,\mathrm{d}u\right)\right] && \text{by def. of } \mathsf{Cov}\\
&= \mathbf{E}\left[\int_{t\in\mathbb{R}}\int_{u\in\mathbb{R}} \mathsf{v}(t)\mathsf{v}(u)\psi_m(t)\psi_n(u)\,\mathrm{d}t\,\mathrm{d}u\right]\\
&= \int_{t\in\mathbb{R}}\int_{u\in\mathbb{R}} \mathbf{E}[\mathsf{v}(t)\mathsf{v}(u)]\psi_m(t)\psi_n(u)\,\mathrm{d}t\,\mathrm{d}u\\
&= \int_{t\in\mathbb{R}}\int_{u\in\mathbb{R}} \sigma^2\delta(t-u)\psi_m(t)\psi_n(u)\,\mathrm{d}t\,\mathrm{d}u && \text{by } \textit{white} \text{ hyp.} && \text{(B)}\\
&= \sigma^2\int_{t\in\mathbb{R}} \psi_m(t)\psi_n(u)\,\mathrm{d}t\\
&= \sigma^2\left\langle \psi_m(t) \mid \psi_n(u)\right\rangle && \text{by def. of } \langle \triangle \mid \nabla \rangle && \text{(Definition 21.1 page 149)}\\
&= \begin{cases} \sigma^2 & \text{for } n = m\\ 0 & \text{for } n \neq m. \end{cases} && \text{by } \textit{orthonormal} \text{ prop.} && \text{(Definition 21.1 page 149)}
\end{aligned}$$

✑

Lemma 21.4. *Let Ψ, $\mathsf{y}(t;\theta)$, $\mathsf{x}(t)$, $\mathsf{v}(t)$, $\dot{y}_n$, $\dot{x}_n$, $\dot{v}_n$, and Y be defined as in Definition 21.1 (page 149).*

LEM

$$\left.\begin{array}{lrcll} \text{(A).} & \operatorname{cov}[\mathsf{v}(t),\mathsf{v}(u)] & = & \sigma^2\delta(t-u) & \text{and} \\ \text{(B).} & \mathsf{v}(t) & \sim & \mathsf{N}\left(0,\sigma^2\right) & \text{and} \\ \text{(C).} & \langle\psi_n\,|\,\psi_m\rangle & = & \bar{\delta}_{mn} & \end{array}\right\} \implies \left\{\begin{array}{lrcll} \text{(1).} & \dot{v}_n & \sim & \mathsf{N}\left(0,\sigma^2\right) & \text{(GAUSSIAN)} \\ \text{(2).} & \operatorname{cov}\left[\dot{v}_n,\dot{v}_m\right] & = & \sigma^2\bar{\delta}_{nm} & \text{(UNCORRELATED)} \\ \text{(3).} & \mathsf{P}\left\{\dot{v}_n\wedge\dot{v}_m\right\} & = & \mathsf{P}\left\{\dot{v}_n\right\}\mathsf{P}\left\{\dot{v}_m\right\} & \text{(INDEPENDENT)} \end{array}\right.$$

✎PROOF:

1. Because the operations are *linear* on processes are *Gaussian* (hypothesis C).

2.

$$\begin{aligned} \mathbf{E}\dot{v}_n = 0 &\qquad \text{by } \textit{AWN} \text{ properties and Theorem 21.4 page 155} \\ \operatorname{cov}\left[\dot{v}_m,\dot{v}_n\right] = \sigma^2\bar{\delta}_{mn} &\qquad \text{by } \textit{AWN} \text{ properties and Lemma 21.3 page 150} \end{aligned}$$

3. Because the processes are *Gaussian*, *uncorrelated* implies *independent*.

✏

21.2 Sufficient Statistics

Theorem 21.1 (Sufficient Statistic Theorem). [2] *Let Ψ, $\mathsf{y}(t;\theta)$, $\mathsf{x}(t)$, $\mathsf{v}(t)$, $\dot{y}_n$, $\dot{x}_n$, $\dot{v}_n$, and Y be defined as in Definition 21.1 (page 149). Let $\hat{\theta}_{\mathsf{map}}$ be the* MAP ESTIMATE *and $\hat{\theta}_{\mathsf{ml}}$ be the* ML ESTIMATE *(Definition 10.1 page 81) of θ.*

THM

$$\left\{\begin{array}{lll} \text{(A).} & \mathsf{v}(t) \textit{ is ZERO-MEAN} & \text{and} \\ \text{(B).} & \mathsf{v}(t) \textit{ is WHITE} & \text{and} \\ \text{(C).} & \mathsf{v}(t) \textit{ is GAUSSIAN} & \end{array}\right\} \implies \underbrace{\left\{\begin{array}{lrcll} \text{(1).} & \mathsf{P}\left\{\mathsf{x}(t;\theta)|\mathsf{y}(t;\theta)\right\} & = & \mathsf{P}\left\{\mathsf{x}(t;\theta)|Y\right\} & \text{and} \\ \text{(2).} & \hat{\theta}_{\mathsf{map}} & = & \arg\max_{\hat{\theta}}\mathsf{P}\left\{\mathsf{x}(t;\theta)|Y\right\} & \text{and} \\ \text{(3).} & \hat{\theta}_{\mathsf{ml}} & = & \arg\max_{\hat{\theta}}\mathsf{P}\left\{Y|\mathsf{x}(t;\theta)\right\} & \end{array}\right\}}_{\textit{the } N \textit{ element set } Y \textit{ is a SUFFICIENT STATISTIC for estimating } \mathsf{x}(t;\theta)}$$

✎PROOF:

1. definition: Let $\mathsf{v}'(t) \triangleq \mathsf{v}(t) - \sum_{n=1}^{N}\dot{v}_n\psi_n(t)$.

[2] Fisher (1922) page 316 ⟨"Criterion of Sufficiency"⟩

2. lemma: The relationship between Y and $\mathsf{v}'(t)$ is given by

$$\begin{aligned}
&\boxed{\mathsf{y}(t;\theta)}\\
&= \sum_{n=1}^{N} \langle \mathsf{y}(t;\theta)\,|\,\psi_n(t)\rangle\,\psi_n(t) + \left[\mathsf{y}(t;\theta) - \sum_{n=1}^{N} \langle \mathsf{y}(t;\theta)\,|\,\psi_n(t)\rangle\,\psi_n(t)\right] && \text{by \textit{additive identity} property of } (\mathbb{C},+,\cdot,0,1)\\
&\triangleq \sum_{n=1}^{N} \langle \mathsf{y}(t;\theta)\,|\,\psi_n(t)\rangle\,\psi_n(t) + \left[\mathsf{y}(t;\theta) - \sum_{n=1}^{N} \langle \mathsf{x}(t)+\mathsf{v}(t)\,|\,\psi_n(t)\rangle\,\psi_n(t)\right] && \text{by definition of } \mathsf{y}(t;\theta)\\
&= \sum_{n=1}^{N} \dot{y}_n\psi_n(t) + \underbrace{\mathsf{x}(t)+\mathsf{v}(t)}_{\mathsf{y}(t;\theta)} - \underbrace{\sum_{n=1}^{N} \langle \mathsf{x}(t)\,|\,\psi_n(t)\rangle\,\psi_n(t)}_{\mathsf{x}(t)} - \underbrace{\sum_{n=1}^{N} \langle \mathsf{v}(t)\,|\,\psi_n(t)\rangle\,\psi_n(t)}_{\mathsf{v}(t)-\mathsf{v}'(t)} && \text{by definition of } \dot{y}_n \text{ and \textit{additive} property of } \langle\triangle\,|\,\triangledown\rangle \text{ (Definition N.1 page 309)}\\
&= \sum_{n=1}^{N} \dot{y}_n\psi_n(t) + \mathsf{x}(t) + \mathsf{v}(t) - \mathsf{x}(t) - \left[\mathsf{v}(t) - \mathsf{v}'(t)\right]\\
&= \boxed{\sum_{n=1}^{N} \dot{y}_n\psi_n(t) + \mathsf{v}'(t)}
\end{aligned}$$

3. lemma: $\mathbf{E}\left[\dot{v}_n\mathsf{v}(t)\right] = N_o\psi_n(t)$. Proof:

$$\begin{aligned}
&\mathbf{E}\left[\dot{v}_n\mathsf{v}(t)\right]\\
&\triangleq \mathbf{E}\left[\left(\int_{t\in\mathbb{R}} \mathsf{v}(u)\psi_n(u)\,\mathsf{d}u\right)\mathsf{v}(t)\right] && \text{by definition of } \dot{v}_n(t) && \text{(Definition 21.1 page 149)}\\
&= \mathbf{E}\left[\int_{t\in\mathbb{R}} \mathsf{v}(u)\mathsf{v}(t)\psi_n(u)\,\mathsf{d}u\right] && \text{by \textit{linearity} of } \int \mathsf{d}u \text{ operator}\\
&= \int_{t\in\mathbb{R}} \mathbf{E}[\mathsf{v}(u)\mathsf{v}(t)]\psi_n(u)\,\mathsf{d}u && \text{by \textit{linearity} of } \mathbf{E} && \text{(Theorem 1.1 page 4)}\\
&= \int_{t\in\mathbb{R}} N_o\delta(u-t)\psi_n(u)\,\mathsf{d}u && \text{by \textit{white} hypothesis}\\
&= N_o\psi_n(t) && \text{by property of \textit{Dirac delta} } \delta(t)
\end{aligned}$$

4. lemma: Y and $\mathsf{v}'(t)$ are *uncorrelated*: Proof:

$$\begin{aligned}
&\mathbf{E}\left[\dot{y}_n\mathsf{v}'(t)\right]\\
&\triangleq \mathbf{E}\left[\langle \mathsf{y}(t;\theta)\,|\,\psi_n(t)\rangle\left(\mathsf{v}(t) - \sum_{n=1}^{N}\langle \mathsf{v}(t)\,|\,\psi_n(t)\rangle\,\psi_n(t)\right)\right] && \text{by definitions of } \dot{y}_n \text{ and } \mathsf{v}'(t)\\
&\triangleq \mathbf{E}\left[\langle \mathsf{x}(t)+\mathsf{v}(t)\,|\,\psi_n(t)\rangle\left(\mathsf{v}(t) - \sum_{n=1}^{N}\langle \mathsf{v}(t)\,|\,\psi_n(t)\rangle\,\psi_n(t)\right)\right] && \text{by definition of } \mathsf{y}(t;\theta)\\
&= \mathbf{E}\left[\left(\langle \mathsf{x}(t)\,|\,\psi_n(t)\rangle + \langle \mathsf{v}(t)\,|\,\psi_n(t)\rangle\right)\left(\mathsf{v}(t) - \sum_{n=1}^{N}\langle \mathsf{v}(t)\,|\,\psi_n(t)\rangle\,\psi_n(t)\right)\right] && \text{by \textit{additive} property of } \langle\triangle\,|\,\triangledown\rangle \text{ (Definition N.1 page 309)}\\
&= \mathbf{E}\left[\left(\dot{x}_n + \dot{v}_n\right)\left(\mathsf{v}(t) - \sum_{n=1}^{N}\dot{v}_n\psi_n(t)\right)\right] && \text{by definitions of } \dot{x}_n \text{ and } \dot{v}_n \text{ (Definition 21.1 page 149)}\\
&= \mathbf{E}\left[\dot{x}_n\mathsf{v}(t) - \dot{x}_n\sum_{n=1}^{N}\dot{v}_n\psi_n(t) + \dot{v}_n\mathsf{v}(t) - \dot{v}_n\sum_{n=1}^{N}\dot{v}_n\psi_n(t)\right]\\
&= \mathbf{E}\left[\dot{x}_n\mathsf{v}(t)\right] - \mathbf{E}\left[\dot{x}_n\sum_{n=1}^{N}\dot{v}_n\psi_n(t)\right] + \mathbf{E}\left[\dot{v}_n\mathsf{v}(t)\right] - \mathbf{E}\left[\sum_{m=1}^{N}\dot{v}_n\dot{v}_m\psi_m(t)\right] && \text{by \textit{linearity} of } \mathbf{E} \text{ (Theorem 1.1 page 4)}
\end{aligned}$$

$$= \dot{x}_n \cancelto{0}{\mathbf{E}\mathsf{v}(t)} - \dot{x}_n \sum_{n=1}^{N} \mathbf{E}\cancelto{0}{[\dot{\upsilon}_n]}\psi_n(t) + \mathbf{E}\left[\dot{\upsilon}_n \mathsf{v}(t)\right] - \sum_{m=1}^{N} \mathbf{E}\left[\dot{\upsilon}_n\dot{\upsilon}_m\right]\psi_m(t)$$ by *linearity* of $\mathbf{E}$ (Theorem 1.1 page 4)

$$= 0 - 0 + \mathbf{E}\left[\dot{\upsilon}_n \mathsf{v}(t)\right] - \sum_{m=1}^{N} N_o \bar{\delta}_{mn}\psi_m(t)$$ by *white* hypothesis

$$= N_o\psi_n(t) - N_o\psi_n(t)$$ by (3) lemma

$$= 0$$

$\Longrightarrow$ **uncorrelated**

5. lemma: Y and $\mathsf{v}'(t)$ are *independent*. Proof: By (4) lemma, $\dot{y}_n$ and $\mathsf{v}'(t)$ are *uncorrelated*. By hypothesis, they are *Gaussian*, and thus are also **independent**.
6. Proof that $\mathsf{P}\{\mathsf{x}(t;\theta)|\mathsf{y}(t;\theta)\} = \mathsf{P}\left\{\mathsf{x}(t;\theta)|\dot{y}_1,\ \dot{y}_2, \ldots, \dot{y}_N\right\}$:

$$\begin{aligned}
\mathsf{P}\{\mathsf{x}(t;\theta)|\mathsf{y}(t;\theta)\} &= \mathsf{P}\left\{\mathsf{x}(t;\theta)|\sum_{n=1}^{N}\dot{y}_n\psi_n(t) + \mathsf{v}'(t)\right\}\\
&= \mathsf{P}\left\{\mathsf{x}(t;\theta)|Y, \mathsf{v}'(t)\right\} && \text{because } Y \text{ and } \mathsf{v}'(t) \text{ can be extracted by } \langle\cdots\mid\psi_n(t)\rangle\\
&= \frac{\mathsf{P}\left\{Y, \mathsf{v}'(t)|\mathsf{x}(t;\theta)\right\} P\{\mathsf{x}(t;\theta)\}}{\mathsf{P}\{Y, \mathsf{v}'(t)\}}\\
&= \frac{\mathsf{P}\{Y|\mathsf{x}(t;\theta)\}\,\mathsf{P}\left\{\mathsf{v}'(t)|\mathsf{x}(t;\theta)\right\}\mathsf{P}\{\mathsf{x}(t;\theta)\}}{\mathsf{P}\{Y\}\mathsf{P}\{\mathsf{v}'(t)\}} && \text{by \textit{independence} of } Y \text{ and } \mathsf{v}'(t) \text{ ((5) lemma page 153)}\\
&= \frac{\mathsf{P}\{Y|\mathsf{x}(t;\theta)\}\,\mathsf{P}\left\{\mathsf{v}'(t)\right\}\mathsf{P}\{\mathsf{x}(t;\theta)\}}{\mathsf{P}\{Y\}\mathsf{P}\{\mathsf{v}'(t)\}} && \text{by \textit{independence} of x and v}\\
&= \frac{\mathsf{P}\{Y|\mathsf{x}(t;\theta)\}\,\mathsf{P}\{\mathsf{x}(t;\theta)\}}{\mathsf{P}\{Y\}}\\
&= \frac{\mathsf{P}\{Y, \mathsf{x}(t;\theta)\}}{\mathsf{P}\{Y\}}\\
&= \mathsf{P}\{\mathsf{x}(t;\theta)|Y\} && \text{by definition of \textit{conditional probability} (Definition A.4 page 174)}
\end{aligned}$$

7. Proof that Y is a *sufficient statistic* for the *MAP estimate*:

$$\begin{aligned}
\hat{\theta}_{\text{map}} &\triangleq \arg\max_{\hat{\theta}} \mathsf{P}\{\mathsf{x}(t;\theta)|\mathsf{y}(t;\theta)\} && \text{by definition of \textit{MAP estimate} (Definition 10.1 page 81)}\\
&= \arg\max_{\hat{\theta}} \mathsf{P}\{\mathsf{x}(t;\theta)|Y\} && \text{by item (6)}
\end{aligned}$$

8. Proof that Y is a *sufficient statistic* for the *ML estimate*:

$$\begin{aligned}
\hat{\theta}_{\text{ml}} &\triangleq \arg\max_{\hat{\theta}} \mathsf{P}\{\mathsf{y}(t;\theta)|\mathsf{x}(t;\theta)\} && \text{by definition of \textit{ML estimate} (Definition 10.1 page 81)}\\
&= \arg\max_{\hat{\theta}} \mathsf{P}\left\{\sum_{n=1}^{N}\dot{y}_n\psi_n(t) + \mathsf{v}'(t)|\mathsf{x}(t;\theta)\right\}\\
&= \arg\max_{\hat{\theta}} \mathsf{P}\left\{Y, \mathsf{v}'(t)|\mathsf{x}(t;\theta)\right\} && \text{because } Y \text{ and } \mathsf{v}'(t) \text{ can be extracted by } \langle\cdots\mid\psi_n(t)\rangle\\
&= \arg\max_{\hat{\theta}} \mathsf{P}\{Y|\mathsf{x}(t;\theta)\}\,\mathsf{P}\left\{\mathsf{v}'(t)\right\}\mathsf{x}(t;\theta) && \text{by \textit{independence} of } Y \text{ and } \mathsf{v}'(t) \text{ ((5) lemma page 153)}\\
&= \arg\max_{\hat{\theta}} \mathsf{P}\{Y|\mathsf{x}(t;\theta)\}\,\mathsf{P}\left\{\mathsf{v}'(t)\right\} && \text{by \textit{independence} of } \mathsf{x}(t) \text{ and } \mathsf{v}'(t)\\
&= \arg\max_{\hat{\theta}} \mathsf{P}\{Y|\mathsf{x}(t;\theta)\} && \text{by \textit{independence} of } \mathsf{v}'(t) \text{ and } \theta
\end{aligned}$$

21.3 Additive noise

Theorem 21.2 (Additive noise projection statistics). *Let Ψ, $\mathsf{y}(t;\theta)$, $\mathsf{x}(t)$, $\mathsf{v}(t)$, $\dot{y}_n$, $\dot{x}_n$, $\dot{v}_n$, and Y be defined as in Definition 21.1 (page 149).*

$$\text{THM}\quad \left\{\begin{array}{llcll}
\text{(A).} & \mathsf{y}(t;\theta) & \triangleq & \mathsf{x}(t;\theta)+\mathsf{v}(t) & (\textit{additive}) \quad \text{and}\\
\text{(B).} & \mathbf{E}[\mathsf{v}(t)] & = & 0 & (\text{ZERO-MEAN}) \quad \text{and}\\
\text{(C).} & \mathsf{x}(t) & \subseteq & \operatorname{span}\Psi & (\Psi \text{ SPANS } \mathsf{x}(t)) \quad \text{and}\\
\text{(D).} & \langle \psi_n \,|\, \psi_m \rangle & = & \bar{\delta}_{mn} & (\text{ORTHONORMAL})
\end{array}\right\} \implies \left\{\ \mathbf{E}\big[\dot{y}_n(\theta)\big] \ = \ \dot{x}_n(\theta)\ \right\}$$

✎PROOF:

$$\begin{aligned}
\mathbf{E}\big[\dot{y}_n(\theta)\big] &\triangleq \mathbf{E}\big[\langle \mathsf{y}(t;\theta)\,|\,\psi_n(t)\rangle\big] && \text{by definition of } \dot{y}_n && \text{(Definition 21.1 page 149)}\\
&= \mathbf{E}\langle \mathsf{x}(t;\theta)+\mathsf{v}(t)\,|\,\psi_n(t)\rangle && \text{by } \textit{additive} \text{ hypothesis} && \text{hypothesis (A)}\\
&= \mathbf{E}\big[\langle \mathsf{x}(t;\theta)\psi_n(t)\,|\,+\rangle\ \langle \mathsf{v}(t)\,|\,\psi_n(t)\rangle\big] && \text{by } \textit{additive} \text{ property of } \langle \triangle \,|\, \triangledown\rangle && \text{(Definition N.1 page 309)}\\
&= \mathbf{E}\left[\left\langle \sum_{k=1}^{N} \dot{x}_k(\theta)\psi_k(t)\,\Big|\,\psi_n(t)\right\rangle + \dot{v}_n\right] && \text{by } \textit{basis} \text{ hypothesis} && \text{(C)}\\
&= \mathbf{E}\left[\sum_{k=1}^{N} \dot{x}_k(\theta)\,\langle \psi_k(t)\,|\,\psi_n(t)\rangle + \dot{v}_n\right] && \text{by } \textit{additive} \text{ property of } \langle \triangle \,|\, \triangledown\rangle && \text{(Definition N.1 page 309)}\\
&= \mathbf{E}\left[\sum_{k=1}^{N} \dot{x}_k(\theta)\bar{\delta}_{k-n}(t) + \dot{v}_n\right] && \text{by } \textit{orthonormal} \text{ hypothesis} && \text{(D)}\\
&= \mathbf{E}\big[\dot{x}_n(\theta)+\dot{v}_n\big] && \text{by definition of } \bar{\delta} && \text{(Definition N.3 page 323)}\\
&= \mathbf{E}\dot{x}_n(\theta) + \cancelto{0}{\mathbf{E}\dot{v}_n} && \text{by } \textit{linearity} \text{ of } \mathbf{E} && \text{(Theorem 1.1 page 4)}\\
&= \dot{x}_n(\theta) && \text{by (B) and Lemma 21.1 page 150}
\end{aligned}$$

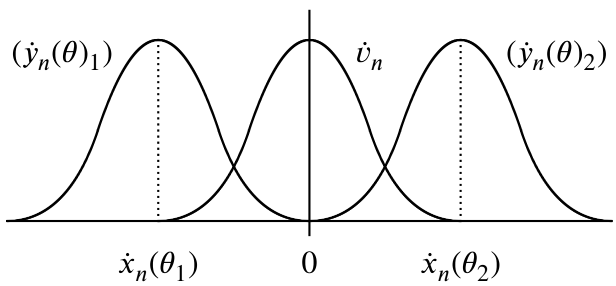


Figure 21.1: Additive *Gaussian* noise channel Statistics

Theorem 21.3 (Additive Gaussian noise projection statistics). *Let Ψ, $\mathsf{y}(t;\theta)$, $\mathsf{x}(t)$, $\mathsf{v}(t)$, $\dot{y}_n$, $\dot{x}_n$, $\dot{v}_n$, and Y be defined as in Definition 21.1 (page 149).*

$$\text{THM}\quad \underbrace{\left\{\begin{array}{llcll}
\text{(A).} & \mathsf{y}(t;\theta) & \triangleq & \mathsf{x}(t)+\mathsf{v}(t) & (\textit{additive}) \quad \text{and}\\
\text{(B).} & \mathsf{v}(t) & \sim & \mathsf{N}\left(0,\sigma^2\right) & (\textit{Gaussian}) \quad \text{and}\\
\text{(C).} & \mathsf{x}(t) & \subseteq & \operatorname{span}\Psi & (\Psi \text{ SPANS } \mathsf{x}(t)) \quad \text{and}\\
\text{(D).} & \langle \psi_n \,|\, \psi_m \rangle & = & \bar{\delta}_{mn} & (\text{ORTHONORMAL})
\end{array}\right\}}_{\text{ADDITIVE GAUSSIAN } \textit{system}} \implies \left\{\ \dot{y}_n(\theta) \ \sim \ \mathsf{N}\left(\dot{x}_n(\theta),\sigma^2\right) \quad (\text{GAUSSIAN})\ \right\}$$

✎PROOF:

1. Proof for (1): By hypothesis (B) and Lemma 21.1 page 150.
2. Proof for (2):

$$\begin{aligned}
\mathbf{E}\left[\dot{y}_n(\theta)\right] &\triangleq \mathbf{E}\left[\langle \mathsf{y}(t;\theta)\,|\,\psi_n(t)\rangle\,|\theta\right] && \text{by definition of } \dot{y}_n && \text{(Definition 21.1 page 149)}\\
&= \mathbf{E}\left[\langle \mathsf{x}(t;\theta)+\mathsf{v}(t)\,|\,\psi_n(t)\rangle\right] && \text{by } \textit{additive} \text{ hypothesis} && \text{hypothesis (A)}\\
&= \mathbf{E}\left[\langle \mathsf{x}(t;\theta)\,|\,\psi_n(t)\rangle\right] + \mathbf{E}\left[\langle \mathsf{v}(t)\,|\,\psi_n(t)\rangle\right] && \text{by } \textit{additive} \text{ property of } \langle\triangle\,|\,\nabla\rangle && \text{(Definition N.1 page 309)}\\
&= \mathbf{E}\left\langle \sum_{k=1}^{N} \dot{x}_k(\theta)\psi_k(t)\,|\,\psi_n(t)\right\rangle + \mathbf{E}\dot{v}_n && \text{by } \textit{basis} \text{ hypothesis} && \text{(C)}\\
&= \sum_{k=1}^{N} \mathbf{E}\left[\dot{x}_k(\theta)\right]\langle \psi_k(t)\,|\,\psi_n(t)\rangle + \mathbf{E}\dot{v}_n && \text{by } \textit{additive} \text{ property of } \langle\triangle\,|\,\nabla\rangle && \text{(Definition N.1 page 309)}\\
&= \sum_{k=1}^{N} \mathbf{E}\left[\dot{x}_k(\theta)\right]\bar{\delta}_{k-n}(t) + \mathbf{E}\dot{v}_n && \text{by } \textit{orthonormal} \text{ hypothesis} && \text{(D)}\\
&= \mathbf{E}\dot{x}_n(\theta) + \mathbf{E}\dot{v}_n && \text{by definition of } \bar{\delta} && \text{(Definition N.3 page 323)}\\
&= \dot{x}_n(\theta) + 0 && \text{by Lemma 21.1 page 150}
\end{aligned}$$

3. Proof for (3): The distribution follows because the process is a linear operations on a Gaussian process.

✏

Theorem 21.4 (Additive white noise projection statistics). *Let Ψ, $\mathsf{y}(t;\theta)$, $\mathsf{x}(t)$, $\mathsf{v}(t)$, $\dot{y}_n$, $\dot{x}_n$, $\dot{v}_n$, and Y be defined as in Definition 21.1 (page 149).*

THM

$$\underbrace{\left.\begin{array}{llcll}
(A). & \mathsf{y}(t;\theta) & \triangleq & \mathsf{x}(t)+\mathsf{v}(t) & \textit{and}\\
(B). & \operatorname{cov}\left[\mathsf{v}(t),\mathsf{v}(u)\right] & = & \sigma^2\delta(t-u) & \textit{and}\\
(C). & \mathbf{E}\left[\mathsf{v}(t)\right] & = & 0 & \textit{and}\\
(E). & \mathsf{x}(t) & \subseteq & \operatorname{span}\Psi & \textit{and}\\
(E). & \langle \psi_n\,|\,\psi_m\rangle & = & \bar{\delta}_{mn} &
\end{array}\right\}}_{\text{ADDITIVE WHITE } \textit{system}} \implies \left\{\begin{array}{llcll}
(1). & \mathbf{E}\dot{v}_n & = & 0 & \text{(ZERO-MEAN)}\\
(2). & \mathbf{E}(\dot{y}_n(\theta)) & = & \dot{x}_n(\theta) & \\
(3). & \operatorname{cov}\left[\dot{v}_n,\dot{v}_m\right] & = & \sigma^2\bar{\delta}_{nm} & \text{(UNCORRELATED)}\\
(4). & \operatorname{cov}\left[\dot{y}_n(\theta),\dot{y}_m|\theta\right] & = & \sigma^2\bar{\delta}_{nm} & \text{(UNCORRELATED)}
\end{array}\right.$$

✎PROOF:

1. Because the noise is *additive* (hypothesis A)...

$$\begin{aligned}
\mathbf{E}\dot{v}_n &= 0 && \text{by } \textit{additive} \text{ property and Theorem 21.2 page 154}\\
(\dot{y}_n(\theta)) &= \dot{x}_n(\theta) + \dot{v}_n && \text{by } \textit{additive} \text{ property and Theorem 21.2 page 154}\\
\mathbf{E}(\dot{y}_n|\theta) &= \dot{x}_n(\theta) && \text{by } \textit{additive} \text{ property and Theorem 21.2 page 154}
\end{aligned}$$

2. Proof for (4):

$$\begin{aligned}
\operatorname{cov}\left[\dot{y}_n(\theta),\dot{y}_m|\theta\right] &= \mathbf{E}\left[\dot{y}_n\dot{y}_m|\theta\right] - [\mathbf{E}\dot{y}_n(\theta)][\mathbf{E}\dot{y}_m|\theta]\\
&= \mathbf{E}\left[(\dot{x}_n(\theta)+\dot{v}_n)(\dot{x}_m(\theta)+\dot{v}_m)\right] - \dot{x}_n(\theta)\dot{x}_m(\theta)\\
&= \mathbf{E}\left[\dot{x}_n(\theta)\dot{x}_m(\theta) + \dot{x}_n(\theta)\dot{v}_m + \dot{v}_n\dot{x}_m(\theta) + \dot{v}_n\dot{v}_m\right] - \dot{x}_n(\theta)\dot{x}_m(\theta)\\
&= \dot{x}_n(\theta)\dot{x}_m(\theta) + \dot{x}_n(\theta)\mathbf{E}\left[\dot{v}_m\right] + \mathbf{E}\left[\dot{v}_n\right]\dot{x}_m(\theta) + \mathbf{E}\left[\dot{v}_n\dot{v}_m\right] - \dot{x}_n(\theta)\dot{x}_m(\theta)\\
&= 0 + \dot{x}_n(\theta)\cdot 0 + 0\cdot\dot{x}_m(\theta) + \operatorname{cov}\left[\dot{v}_n,\dot{v}_m\right] + [\mathbf{E}\dot{v}_n][\mathbf{E}\dot{v}_m]\\
&= \sigma^2\bar{\delta}_{nm} + 0\cdot 0 && \text{by Lemma 21.3}\\
&= \begin{cases} \sigma^2 & \text{for } n=m\\ 0 & \text{for } n\neq m. \end{cases}
\end{aligned}$$

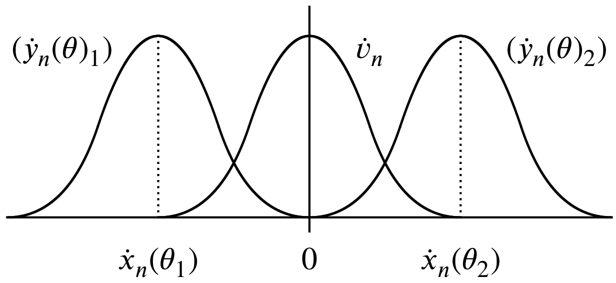


Figure 21.2: Additive white *Gaussian* noise channel statistics

Theorem 21.5 (AWGN projection statistics). *Let* Ψ, $\mathsf{y}(t;\theta)$, $\mathsf{x}(t)$, $\mathsf{v}(t)$, $\dot{y}_n$, $\dot{x}_n$, $\dot{v}_n$, *and* Y *be defined as in Definition 21.1 (page 149).*

THM

$$\underbrace{\left.\begin{array}{llcll} \text{(A).} & \mathsf{y}(t;\theta) & \triangleq & \mathsf{x}(t)+\mathsf{v}(t) & \text{and} \\ \text{(B).} & \operatorname{cov}[\mathsf{v}(t),\mathsf{v}(u)] & = & \sigma^2\delta(t-u) & \text{and} \\ \text{(C).} & \mathsf{v}(t) & \sim & \mathsf{N}\left(0,\sigma^2\right) & \text{and} \\ \text{(D).} & \mathsf{x}(t) & \subseteq & \operatorname{span}\Psi & \text{and} \\ \text{(E).} & \langle \psi_n \,|\, \psi_m \rangle & = & \bar{\delta}_{mn} & \end{array}\right\}}_{\text{additive white Gaussian system}} \implies \left\{\begin{array}{lll} \text{(1).} & \dot{y}_n(\theta) \sim \mathsf{N}\left(\dot{x}_n(\theta),\sigma^2\right) & \text{(Gaussian)} \\ \text{(2).} & \operatorname{cov}\left[\dot{y}_n,\dot{y}_m\right] = \sigma^2\bar{\delta}_{nm} & \text{(uncorrelated)} \\ \text{(3).} & \mathsf{P}\left\{\dot{y}_n \wedge \dot{y}_m\right\} = \mathsf{P}\left\{\dot{y}_n\right\}\mathsf{P}\left\{\dot{y}_m\right\} & \text{(independent)} \end{array}\right.$$

✎PROOF:

1. Proof for (1) follow because the operations are *linear* on processes are *Gaussian* (hypothesis C).

2.

$$\begin{aligned} \mathbf{E}\dot{v}_n &= 0 && \text{by } \textit{AWN} \text{ properties and Theorem 21.4 page 155} \\ \dot{y}_n &= \dot{x}_n + \dot{v}_n && \text{by } \textit{AWN} \text{ properties and Theorem 21.4 page 155} \\ \mathbf{E}\dot{y}_n &= \dot{x}_n && \text{by } \textit{AWN} \text{ properties and Theorem 21.4 page 155} \\ \operatorname{cov}\left[\dot{y}_n,\dot{y}_m\right] &= \sigma^2\bar{\delta}_{mn} && \text{by } \textit{AWN} \text{ properties and Theorem 21.4 page 155} \end{aligned}$$

3. Because the processes are *Gaussian*, *uncorrelated* implies *independent*.

21.4 ML estimates

The AWGN projection statistics provided by Theorem 21.5 (page 156) help generate the optimal ML-estimates for a number of communication systems. These ML-estimates can be expressed in either of two standard forms:

- **Spectral decompostion**: The optimal estimate is expressed in terms of *projections* of signals onto orthonormal basis functions.
- **Matched signal**: The optimal estimate is expressed in terms of the (noisy) received signal correlated with ("matched" with) the (noiseless) transmitted signal.

Theorem 21.6 (page 157) (next) expresses the general optimal *ML estimate* in both of these forms.

Parameter detection is a special case of parameter estimation. In parameter detection, the estimate is a member of an finite set. In parameter estimation, the estimate is a member of an infinite set (Section 21.4 page 156).

Theorem 21.6 (General ML estimation). *Let Ψ, $\mathsf{y}(t;\theta)$, $\mathsf{x}(t)$, $\mathsf{v}(t)$, $\dot{y}_n$, $\dot{x}_n$, $\dot{v}_n$, and Y be defined as in Definition 21.1 (page 149). Let $\hat{\theta}_{\mathsf{ml}}$ be the* ML ESTIMATE *(Definition 10.1 page 81) of θ.*

THM

$$\begin{aligned}
\hat{\theta}_{\mathsf{ml}} &= \arg\min_{\hat{\theta}}\left[\sum_{n=1}^{N}[\dot{y}_n - \dot{x}_n(\hat{\theta})]^2\right] && \textit{(spectral decomposition)}\\
&= \arg\max_{\hat{\theta}}\left[2\langle \mathsf{y}(t;\theta)\,|\,\mathsf{x}(t;\theta)\rangle - \|\mathsf{x}(t;\theta)\|^2\right] && \textit{(matched signal)}
\end{aligned}$$

✎PROOF:

$$\begin{aligned}
\hat{\theta}_{\mathsf{ml}} &= \arg\max_{\hat{\theta}} \mathsf{P}\{\mathsf{y}(t;\theta)|\mathsf{x}(t;\theta)\}\\
&= \arg\max_{\hat{\theta}} \mathsf{P}\left\{\dot{y}_1,\dot{y}_2,\ldots,\dot{y}_n|\mathsf{x}(t;\theta)\right\} && \text{by Theorem 21.1 (page 151)}\\
&= \arg\max_{\hat{\theta}} \prod_{n=1}^{N}\mathsf{P}\left\{\dot{y}_n|\mathsf{x}(t;\theta)\right\}\\
&= \arg\max_{\hat{\theta}} \prod_{n=1}^{N}\mathsf{p}\left[\dot{\mathsf{y}}_n|\mathsf{x}(\mathsf{t};\theta)\right]\\
&= \arg\max_{\hat{\theta}} \prod_{n=1}^{N}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\frac{[\dot{y}_n - \dot{x}_n(\hat{\theta})]^2}{-2\sigma^2} && \text{by Theorem 21.5 (page 156)}\\
&= \arg\max_{\hat{\theta}} \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N \exp\frac{-1}{2\sigma^2}\sum_{n=1}^{N}[\dot{y}_n - \dot{x}_n(\hat{\theta})]^2\\
&= \arg\max_{\hat{\theta}}\left[-\sum_{n=1}^{N}[\dot{y}_n - \dot{x}_n(\hat{\theta})]^2\right]\\
&= \arg\max_{\hat{\theta}}\left[-\lim_{N\to\infty}\sum_{n=1}^{N}[\dot{y}_n - \dot{x}_n(\hat{\theta})]^2\right] && \text{by Theorem 21.1 (page 151)}\\
&= \arg\max_{\hat{\theta}}\left[-\|\mathsf{y}(t;\theta)-\mathsf{x}(t;\theta)\|^2\right] && \text{by \textit{Plancheral's formula} (Theorem K.10 page 273)}\\
&= \arg\max_{\hat{\theta}}\left[-\|\mathsf{y}(t;\theta)\|^2 + 2\mathbf{R_e}\langle \mathsf{y}(t;\theta)\,|\,\mathsf{x}(t;\theta)\rangle - \|\mathsf{x}(t;\theta)\|^2\right]\\
&= \arg\max_{\hat{\theta}}\left[2\langle \mathsf{y}(t;\theta)\,|\,\mathsf{x}(t;\theta)\rangle - \|\mathsf{x}(t;\theta)\|^2\right] && \text{because } \mathsf{y}(t;\theta) \textit{ independent } \text{of } \hat{\theta}
\end{aligned}$$

✏

Theorem 21.7 (ML amplitude estimation). [3] *Let* **S** *be an additive white gaussian noise system.*

[3] Srinath et al. (1996) pages 158–159

THM

$$\left\{\begin{array}{llcll}
\text{(A).} & \mathsf{v}(t) \text{ is AWGN} & & & \text{and}\\
\text{(B).} & \mathsf{y}(t;a) & = & \mathsf{x}(t;a)+\mathsf{v}(t) & \text{and}\\
\text{(C).} & \mathsf{x}(t;a) & \triangleq & a\lambda(t). &
\end{array}\right\}
\implies
\left\{\begin{array}{lrcll}
\text{(1).} & \hat{a}_{\mathsf{ml}} & = & \frac{1}{\|\lambda(t)\|^2}\sum_{n=1}^{N}\dot{y}_n\dot{\lambda}_n & \\
\text{(2).} & \mathbf{E}\hat{a}_{\mathsf{ml}} & = & a & (\text{UNBIASED})\\
\text{(3).} & \operatorname{var}\hat{a}_{\mathsf{ml}} & = & \frac{\sigma^2}{\|\lambda(t)\|^2} & \\
\text{(4).} & \operatorname{var}\hat{a}_{\mathsf{ml}} & = & \text{CR lower bound} & (\text{EFFICIENT})
\end{array}\right\}$$

✎PROOF:

1. *ML estimate* in "matched signal" form:

$$\begin{aligned}
\hat{a}_{\mathsf{ml}} &= \arg\max_a \left[2\langle \mathsf{y}(t;\theta)\,|\,\mathsf{x}(t;\theta)\rangle - \|\mathsf{x}(t;\phi)\|^2\right] && \text{by Theorem 21.6 (page 157)}\\
&= \arg\max_a \left[2\langle \mathsf{y}(t;\theta)\,|\,a\lambda(t)\rangle - \|a\lambda(t)\|^2\right] && \text{by hypothesis}\\
&= \arg_a \left[\frac{\partial}{\partial a}2a\langle \mathsf{y}(t;\theta)\,|\,\lambda(t)\rangle - \frac{\partial}{\partial a}a^2\|\lambda(t)\|^2 = 0\right]\\
&= \arg_a \left[2\langle \mathsf{y}(t;\theta)\,|\,\lambda(t)\rangle - 2a\|\lambda(t)\|^2 = 0\right]\\
&= \arg_a \left[\langle \mathsf{y}(t;\theta)\,|\,\lambda(t)\rangle = a\|\lambda(t)\|^2\right]\\
&= \frac{1}{\|\lambda(t)\|^2}\langle \mathsf{y}(t;\theta)\,|\,\lambda(t)\rangle
\end{aligned}$$

2. *ML estimate* in "spectral decomposition" form:

$$\begin{aligned}
\hat{a}_{\mathsf{ml}} &= \arg\min_a \left(\sum_{n=1}^{N}\left[\dot{y}_n - \dot{x}_n(a)\right]^2\right) && \text{by Theorem 21.6 (page 157)}\\
&= \arg_a \left(\frac{\partial}{\partial a}\sum_{n=1}^{N}\left[\dot{y}_n - \dot{x}_n(a)\right]^2 = 0\right)\\
&= \arg_a \left(2\sum_{n=1}^{N}\left[\dot{y}_n - \dot{x}_n(a)\right]\frac{\partial}{\partial a}\dot{x}_n(a) = 0\right)\\
&= \arg_a \left(\sum_{n=1}^{N}\left[\dot{y}_n - \langle a\lambda(t)\,|\,\psi_n(t)\rangle\right]\frac{\partial}{\partial a}\langle a\lambda(t)\,|\,\psi_n(t)\rangle = 0\right)\\
&= \arg_a \left(\sum_{n=1}^{N}\left[\dot{y}_n - a\langle \lambda(t)\,|\,\psi_n(t)\rangle\right]\frac{\partial}{\partial a}(a\langle \lambda(t)\,|\,\psi_n(t)\rangle) = 0\right)\\
&= \arg_a \left(\sum_{n=1}^{N}\left[\dot{y}_n - a\dot{\lambda}_n\right]\langle \lambda(t)\,|\,\psi_n(t)\rangle = 0\right)\\
&= \arg_a \left(\sum_{n=1}^{N}\left[\dot{y}_n - a\dot{\lambda}_n\right]\dot{\lambda}_n = 0\right)\\
&= \arg_a \left(\sum_{n=1}^{N}\dot{y}_n\dot{\lambda}_n = \sum_{n=1}^{N}a\dot{\lambda}_n^2\right)\\
&= \left(\frac{1}{\sum_{n=1}^{N}\dot{\lambda}_n^2}\right)\sum_{n=1}^{N}\dot{y}_n\dot{\lambda}_n\\
&= \frac{1}{\|\lambda(t)\|^2}\sum_{n=1}^{N}\dot{y}_n\dot{\lambda}_n
\end{aligned}$$

3. Prove that the estimate $\hat{a}_{\mathsf{ml}}$ is **unbiased**:

$$\begin{aligned}
\mathbf{E}\hat{a}_{\mathsf{ml}} &= \mathbf{E}\frac{1}{\|\lambda(t)\|^2}\int_{t\in\mathbb{R}} \mathsf{y}(t;\theta)\lambda(t)\,\mathsf{d}t && \text{by previous result}\\
&= \mathbf{E}\frac{1}{\|\lambda(t)\|^2}\int_{t\in\mathbb{R}} [a\lambda(t)+\mathsf{v}(t)]\lambda(t)\,\mathsf{d}t && \text{by hypothesis}\\
&= \frac{1}{\|\lambda(t)\|^2}\int_{t\in\mathbb{R}} \mathbf{E}[a\lambda(t)+\mathsf{v}(t)]\lambda(t)\,\mathsf{d}t && \text{by linearity of } \int\cdot\,\mathsf{d}t \text{ and } \mathbf{E}\\
&= \frac{1}{\|\lambda(t)\|^2}a\int_{t\in\mathbb{R}} \lambda^2(t)\,\mathsf{d}t && \text{by } \mathbf{E} \text{ operation}\\
&= \frac{1}{\|\lambda(t)\|^2}a\,\|\lambda(t)\|^2 && \text{by definition of } \|\cdot\|^2\\
&= a
\end{aligned}$$

4. Compute the variance of $\hat{a}_{\mathsf{ml}}$:

$$\begin{aligned}
\mathbf{E}\hat{a}^2_{\mathsf{ml}} &= \mathbf{E}\left[\frac{1}{\|\lambda(t)\|^2}\int_{t\in\mathbb{R}} \mathsf{y}(t;\theta)\lambda(t)\,\mathsf{d}t\right]^2\\
&= \mathbf{E}\left[\frac{1}{\|\lambda(t)\|^4}\int_{t\in\mathbb{R}} \mathsf{y}(t;\theta)\lambda(t)\,\mathsf{d}t\int_v \mathsf{y}(v)\lambda(v)\,\mathsf{d}v\right]\\
&= \mathbf{E}\left[\frac{1}{\|\lambda(t)\|^4}\int_{t\in\mathbb{R}}\int_v [a\lambda(t)+\mathsf{v}(t)][a\lambda(v)+\mathsf{v}(v)]\lambda(t)\lambda(v)\,\mathsf{d}v\,\mathsf{d}t\right]\\
&= \mathbf{E}\left[\frac{1}{\|\lambda(t)\|^4}\int_{t\in\mathbb{R}}\int_v [a^2\lambda(t)\lambda(v)+a\lambda(t)\mathsf{v}(v)+a\lambda(v)\mathsf{v}(t)+\mathsf{v}(t)\mathsf{v}(v)]\lambda(t)\lambda(v)\,\mathsf{d}v\,\mathsf{d}t\right]\\
&= \left[\frac{1}{\|\lambda(t)\|^4}\int_{t\in\mathbb{R}}\int_v [a^2\lambda(t)\lambda(v)+0+0+\sigma^2\delta(t-v)]\lambda(t)\lambda(v)\,\mathsf{d}v\,\mathsf{d}t\right]\\
&= \frac{1}{\|\lambda(t)\|^4}\int_{t\in\mathbb{R}}\int_v a^2\lambda^2(t)\lambda^2(v)\,\mathsf{d}v\,\mathsf{d}t + \frac{1}{\|\lambda(t)\|^4}\int_{t\in\mathbb{R}}\int_v \sigma^2\delta(t-v)\lambda(t)\lambda(v)\,\mathsf{d}v\,\mathsf{d}t\\
&= \frac{1}{\|\lambda(t)\|^4}a^2\int_{t\in\mathbb{R}}\lambda^2(t)\,\mathsf{d}t\int_v\lambda^2(v)\,\mathsf{d}v + \frac{1}{\|\lambda(t)\|^4}\sigma^2\int_{t\in\mathbb{R}}\lambda^2(t)\,\mathsf{d}t\\
&= a^2\frac{1}{\|\lambda(t)\|^4}\,\|\lambda(t)\|^2\,\|\lambda(v)\|^2 + \frac{1}{\|\lambda(t)\|^4}\sigma^2\,\|\lambda(t)\|^2\\
&= a^2 + \frac{\sigma^2}{\|\lambda(t)\|^2}
\end{aligned}$$

$$\begin{aligned}
\operatorname{var}\hat{a}_{\mathsf{ml}} &= \mathbf{E}\hat{a}^2_{\mathsf{ml}} - (\mathbf{E}\hat{a}_{\mathsf{ml}})^2\\
&= \left(a^2+\frac{\sigma^2}{\|\lambda(t)\|^2}\right) - \left(a^2\right)\\
&= \frac{\sigma^2}{\|\lambda(t)\|^2}
\end{aligned}$$

5. Compute the Cramér-Rao Bound:

$$\begin{aligned}
\mathsf{p}\left[\mathsf{y}(\mathsf{t};\theta)|\mathsf{x}(\mathsf{t};\mathsf{a})\right] &= \mathsf{p}\left[\dot{\mathsf{y}}_1,\dot{\mathsf{y}}_2,\ldots,\dot{\mathsf{y}}_N|\mathsf{x}(\mathsf{t};\mathsf{a})\right]\\
&= \prod_{n=1}^{N}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\frac{(\dot{y}_n-a\dot{\lambda}_n)^2}{-2\sigma^2}\\
&= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N \exp\frac{1}{-2\sigma^2}\sum_{n=1}^{N}(\dot{y}_n-a\dot{\lambda}_n)^2
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial a}\ln \mathsf{p}\left[\mathsf{y}(t;\theta)|\mathsf{x}(t;a)\right] &= \frac{\partial}{\partial a}\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N \exp\frac{1}{-2\sigma^2}\sum_{n=1}^{N}(\dot{y}_n - a\dot{\lambda}_n)^2 \\
&= \frac{\partial}{\partial a}\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N + \frac{\partial}{\partial a}\ln\exp\frac{1}{-2\sigma^2}\sum_{n=1}^{N}(\dot{y}_n - a\dot{\lambda}_n)^2 \\
&= \frac{\partial}{\partial a}\left[\frac{1}{-2\sigma^2}\sum_{n=1}^{N}(\dot{y}_n - a\dot{\lambda}_n)^2\right] \\
&= \frac{1}{-2\sigma^2}\sum_{n=1}^{N}2(\dot{y}_n - a\dot{\lambda}_n)(-\dot{\lambda}_n) \\
&= \frac{1}{\sigma^2}\sum_{n=1}^{N}\dot{\lambda}_n(\dot{y}_n - a\dot{\lambda}_n)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2}{\partial a^2}\ln \mathsf{p}\left[\mathsf{y}(t;\theta)|\mathsf{x}(t;a)\right] &= \frac{\partial}{\partial a}\frac{\partial}{\partial a}\ln \mathsf{p}\left[\mathsf{y}(t;\theta)|\mathsf{x}(t;a)\right] \\
&= \frac{\partial}{\partial a}\frac{1}{\sigma^2}\sum_{n=1}^{N}\dot{\lambda}_n(\dot{y}_n - a\dot{\lambda}_n) \\
&= \frac{1}{\sigma^2}\sum_{n=1}^{N}\dot{\lambda}_n(-\dot{\lambda}_n) \\
&= \frac{-1}{\sigma^2}\sum_{n=1}^{N}\dot{\lambda}_n^2 \\
&= \frac{-\|\lambda(t)\|^2}{\sigma^2}
\end{aligned}$$

$$\begin{aligned}
\operatorname{var}\hat{a}_{\mathsf{ml}} &\triangleq \mathbf{E}\left[\hat{a}_{\mathsf{ml}} - \mathbf{E}\hat{a}_{\mathsf{ml}}\right]^2 \\
&= \mathbf{E}\left[\hat{a}_{\mathsf{ml}} - a\right]^2 \\
&\geq \frac{-1}{\mathbf{E}\left(\frac{\partial^2}{\partial a^2}\ln \mathsf{p}\left[\mathsf{y}(t;\theta)|\mathsf{x}(t;a)\right]\right)} \\
&= \frac{-1}{\mathbf{E}\left(\frac{-\|\lambda(t)\|^2}{\sigma^2}\right)} \\
&= \frac{\sigma^2}{\|\lambda(t)\|^2} \qquad \text{(Cramér-Rao lower bound of the variance)}
\end{aligned}$$

6. Proof that $\hat{a}_{\mathsf{ml}}$ is an *efficient* estimate:

   An estimate is *efficient* if $\operatorname{var}\hat{a}_{\mathsf{ml}}$ = CR lower bound. We have already proven this, so $\hat{a}_{\mathsf{ml}}$ is an *efficient* estimate.

   Also, even without explicitly computing the variance of $\hat{a}_{\mathsf{ml}}$, the variance equals the *Cramér-Rao lower bound* (and hence $\hat{a}_{\mathsf{ml}}$ is an *efficient* estimate) if and only if

$$\hat{a}_{\mathsf{ml}} - a = \left(\frac{-1}{\mathbf{E}\left[\frac{\partial^2}{\partial a^2}\ln \mathsf{p}\left[\mathsf{y}(t;\theta)|\mathsf{x}(t;a)\right]\right]}\right)\left(\frac{\partial}{\partial a}\ln \mathsf{p}\left[\mathsf{y}(t;\theta)|\mathsf{x}(t;a)\right]\right)$$

$$\begin{aligned}\left(\frac{-1}{\mathbf{E}\left(\frac{\partial^2}{\partial a^2}\ln \mathsf{p}\left[\mathsf{y}(t;\theta)|\mathsf{x}(t;a)\right]\right)}\right)\left(\frac{\partial}{\partial a}\ln \mathsf{p}\left[\mathsf{y}(t;\theta)|\mathsf{x}(t;a)\right]\right) &= \left(\frac{\sigma^2}{\|\lambda(t)\|^2}\right)\left(\frac{1}{\sigma^2}\sum_{n=1}^{N}\dot{\lambda}(\dot{y}-a\dot{\lambda})\right)\\ &= \frac{1}{\|\lambda(t)\|^2}\sum_{n=1}^{N}\dot{\lambda}\dot{y} - \frac{1}{\|\lambda(t)\|^2}\sum_{n=1}^{N}\dot{\lambda}^2\\ &= \hat{a}_{\mathsf{ml}} - a\end{aligned}$$

✏

Theorem 21.8 (ML phase estimation). [4]

THM

$$\left\{\begin{array}{llcll}\text{(A).} & \mathsf{v}(t) \textit{ is AWGN} & & & \textit{and}\\ \text{(B).} & \mathsf{y}(t;\phi) & = & \mathsf{x}(t;\phi)+\mathsf{v}(t) & \textit{and}\\ \text{(C).} & \mathsf{x}(t;\phi) & \triangleq & A\cos(2\pi f_c t+\phi) & \end{array}\right\} \implies \left\{\hat{\phi}_{\mathsf{ml}} = -\operatorname{atan}\left(\frac{\langle \mathsf{y}(t;\theta)\,|\,\sin(2\pi f_c t)\rangle}{\langle \mathsf{y}(t;\theta)\,|\,\cos(2\pi f_c t)\rangle}\right)\right\}$$

✎PROOF:

$$\begin{aligned}
\hat{\phi}_{\mathsf{ml}} &= \arg\max_{\phi}\left[2\langle \mathsf{y}(t;\phi)\,|\,\mathsf{x}(t;\phi)\rangle - \|\mathsf{x}(t;\phi)\|^2\right] && \text{by Theorem 21.6 (page 157)}\\
&= \arg\max_{\phi}\left[2\langle \mathsf{y}(t;\phi)\,|\,\mathsf{x}(t;\phi)\rangle\right] && \text{because } \|\mathsf{x}(t;\phi)\| \text{ does not depend on } \phi\\
&= \arg_{\phi}\left[\frac{\partial}{\partial\phi}\langle \mathsf{y}(t;\phi)\,|\,\mathsf{x}(t;\phi)\rangle = 0\right]\\
&= \arg_{\phi}\left[\left\langle \mathsf{y}(t;\phi)\,|\,\frac{\partial}{\partial\phi}\mathsf{x}(t;\phi)\right\rangle = 0\right] && \text{because } \langle\triangle\,|\,\nabla\rangle \text{ is \textit{linear}}\\
&= \arg_{\phi}\left[\left\langle \mathsf{y}(t;\phi)\,|\,\frac{\partial}{\partial\phi}A\cos(2\pi f_c t+\phi)\right\rangle = 0\right] && \text{by definition of } \mathsf{x}(t;\phi)\\
&= \arg_{\phi}\left[\langle \mathsf{y}(t;\phi)\,|\,-A\sin(2\pi f_c t+\phi)\rangle = 0\right] && \text{because } \frac{\partial}{\partial\phi}\cos(x) = -\sin(x)\\
&= \arg_{\phi}\left[-A\langle \mathsf{y}(t;\phi)\,|\,\cos(2\pi f_c t)\sin\phi + \sin(2\pi f_c t)\cos\phi\rangle = 0\right] && \text{by \textit{double angle formulas}} \quad \text{(Theorem}\\
&= \arg_{\phi}\left[\sin\phi\langle \mathsf{y}(t;\phi)\,|\,\cos(2\pi f_c t)\rangle = -\cos\phi\langle \mathsf{y}(t;\phi)\,|\,\sin(2\pi f_c t)\rangle\right]\\
&= \arg_{\phi}\left[\frac{\sin\phi}{\cos\phi} = -\frac{\langle \mathsf{y}(t;\phi)\,|\,\sin(2\pi f_c t)\rangle}{\langle \mathsf{y}(t;\phi)\,|\,\cos(2\pi f_c t)\rangle}\right]\\
&= \arg_{\phi}\left[\tan\phi = -\frac{\langle \mathsf{y}(t;\phi)\,|\,\sin(2\pi f_c t)\rangle}{\langle \mathsf{y}(t;\phi)\,|\,\cos(2\pi f_c t)\rangle}\right]\\
&= -\operatorname{atan}\left(\frac{\langle \mathsf{y}(t;\phi)\,|\,\sin(2\pi f_c t)\rangle}{\langle \mathsf{y}(t;\phi)\,|\,\cos(2\pi f_c t)\rangle}\right)
\end{aligned}$$

✏

Theorem 21.9 (ML estimation of a function of a parameter). [5] *Let* **S** *be an additive white gaussian noise system such that*

$$\begin{aligned}\mathsf{y}(t;\theta) &= \mathsf{x}(t;\theta)+\mathsf{v}(t)\\ \mathsf{x}(t;\theta) &= \mathsf{g}(\theta)\end{aligned}$$

and g *is* ONE-TO-ONE AND ONTO (INVERTIBLE).

[4] Srinath et al. (1996) pages 159–160

[5] Srinath et al. (1996) pages 142–143

THM

Then the optimal ML-estimate of parameter θ *is* $\hat{\theta}_{\mathsf{ml}} = \mathsf{g}^{-1}\left(\frac{1}{N}\sum_{n=1}^{N}\dot{y}_n\right).$

If an ML ESTIMATE $\hat{\theta}_{\mathsf{ml}}$ *is unbiased* ($\mathbf{E}\hat{\theta}_{\mathsf{ml}} = \theta$) *then* $\operatorname{var}\hat{\theta}_{\mathsf{ml}} \geq \frac{\sigma^2}{N}\frac{1}{\left[\frac{\partial \mathsf{g}(\theta)}{\partial\theta}\right]^2}.$

If $\mathsf{g}(\theta) = \theta$ *then* $\hat{\theta}_{\mathsf{ml}}$ *is an* ***efficient*** *estimate such that* $\operatorname{var}\hat{\theta}_{\mathsf{ml}} = \frac{\sigma^2}{N}.$

✎PROOF:

$$\begin{aligned}
\hat{\theta}_{\mathsf{ml}} &= \arg\min_{\theta}\left[\sum_{n=1}^{N}[\dot{y}_n - \mathsf{g}(\theta)]^2\right] && \text{by Theorem 21.6 page 157}\\
&= \arg_\theta\left[\frac{\partial}{\partial\theta}\sum_{n=1}^{N}[\dot{y}_n - \mathsf{g}(\theta)]^2 = 0\right] && \text{because form is \textit{quadratic}}\\
&= \arg_\theta\left[2\sum_{n=1}^{N}[\dot{y}_n - \mathsf{g}(\theta)]\frac{\partial}{\partial\theta}\mathsf{g}(\theta) = 0\right]\\
&= \arg_\theta\left[2\sum_{n=1}^{N}[\dot{y}_n - \mathsf{g}(\theta)] = 0\right]\\
&= \arg_\theta\left[\sum_{n=1}^{N}\dot{y}_n = N\mathsf{g}(\theta)\right]\\
&= \arg_\theta\left[\mathsf{g}(\theta) = \frac{1}{N}\sum_{n=1}^{N}\dot{y}_n\right]\\
&= \arg_\theta\left[\theta = \mathsf{g}^{-1}\left(\frac{1}{N}\sum_{n=1}^{N}\dot{y}_n\right)\right]\\
&= \mathsf{g}^{-1}\left(\frac{1}{N}\sum_{n=1}^{N}\dot{y}_n\right)
\end{aligned}$$

If $\hat{\theta}_{\mathsf{ml}}$ is unbiased ($\mathbf{E}\hat{\theta}_{\mathsf{ml}} = \theta$), we can use the *Cramér-Rao bound* to find a lower bound on the variance:

$$\begin{aligned}
\operatorname{var}\hat{\theta}_{\mathsf{ml}} &\triangleq \mathbf{E}\left[\hat{\theta}_{\mathsf{ml}} - \mathbf{E}\hat{\theta}_{\mathsf{ml}}\right]^2\\
&= \mathbf{E}\left[\hat{\theta}_{\mathsf{ml}} - \theta\right]^2\\
&\geq \frac{-1}{\mathbf{E}\left(\frac{\partial^2}{\partial\theta^2}\ln \mathsf{p}\left[\mathsf{y}(t;\theta)|\mathsf{x}(t;\theta)\right]\right)} && \text{by \textit{Cramér-Rao Inequality}}\\
&= \frac{-1}{\mathbf{E}\left(\frac{\partial^2}{\partial\theta^2}\ln \mathsf{p}\left[\dot{\mathsf{y}}_1, \dot{\mathsf{y}}_2, \ldots, \dot{\mathsf{y}}_N|\mathsf{x}(t;\theta)\right]\right)} && \text{by \textit{Sufficient Statistic Theorem} (Theorem 21.1 page 151)}\\
&= \frac{-1}{\mathbf{E}\left(\frac{\partial^2}{\partial\theta^2}\ln\left[\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N \exp\left(\frac{-1}{2\sigma^2}\sum_{n=1}^{N}[\dot{y}_n - \mathsf{g}(\theta)]^2\right)\right]\right)} && \text{by \textit{AWGN} hypothesis and Theorem 21.5 page 156}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{-1}{\mathbf{E}\left(\frac{\partial^2}{\partial\theta^2}\ln\left[\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N\right] + \frac{\partial^2}{\partial\theta^2}\ln\left[\exp\frac{-1}{2\sigma^2}\sum_{n=1}^{N}[\dot{y}_n - \mathrm{g}(\theta)]^2\right]\right)} \\
&= \frac{-1}{\mathbf{E}\left(\frac{\partial^2}{\partial\theta^2}\left(\frac{-1}{2\sigma^2}\sum_{n=1}^{N}[\dot{y}_n - \mathrm{g}(\theta)]^2\right)\right)} \\
&= \frac{2\sigma^2}{\mathbf{E}\left(\frac{\partial}{\partial\theta}\frac{\partial}{\partial\theta}\sum_{n=1}^{N}[\dot{y}_n - \mathrm{g}(\theta)]^2\right)} \\
&= \frac{2\sigma^2}{\mathbf{E}\left(-2\frac{\partial}{\partial\theta}\frac{\partial\mathrm{g}(\theta)}{\partial\theta}\sum_{n=1}^{N}[\dot{y}_n - \mathrm{g}(\theta)]\right)} && \text{by \textit{Chain Rule}} \\
&= \frac{-\sigma^2}{\mathbf{E}\left(\frac{\partial\mathrm{g}^2(\theta)}{\partial\theta^2}\sum_{n=1}^{N}[\dot{y}_n - \mathrm{g}(\theta)] + \frac{\partial\mathrm{g}(\theta)}{\partial\theta}\frac{\partial}{\partial\theta}\sum_{n=1}^{N}[\dot{y}_n - \mathrm{g}(\theta)]\right)} && \text{by \textit{Product Rule}} \\
&= \frac{-\sigma^2}{\mathbf{E}\left(\frac{\partial\mathrm{g}^2(\theta)}{\partial\theta^2}\sum_{n=1}^{N}[\dot{y}_n - \mathrm{g}(\theta)] - N\frac{\partial\mathrm{g}(\theta)}{\partial\theta}\frac{\partial\mathrm{g}(\theta)}{\partial\theta}\right)} \\
&= \frac{-\sigma^2}{\frac{\partial\mathrm{g}^2(\theta)}{\partial\theta^2}\sum_{n=1}^{N}\mathbf{E}[\dot{y}_n - \mathrm{g}(\theta)] - N\frac{\partial\mathrm{g}(\theta)}{\partial\theta}\frac{\partial\mathrm{g}(\theta)}{\partial\theta}} \\
&= \frac{-\sigma^2}{-N\frac{\partial\mathrm{g}(\theta)}{\partial\theta}\frac{\partial\mathrm{g}(\theta)}{\partial\theta}} && \text{because derivative of constant} = 0 \\
&= \frac{\sigma^2}{N}\frac{1}{\left[\frac{\partial\mathrm{g}(\theta)}{\partial\theta}\right]^2}
\end{aligned}
$$

The inequality becomes equality (an *efficient* estimate) if and only if

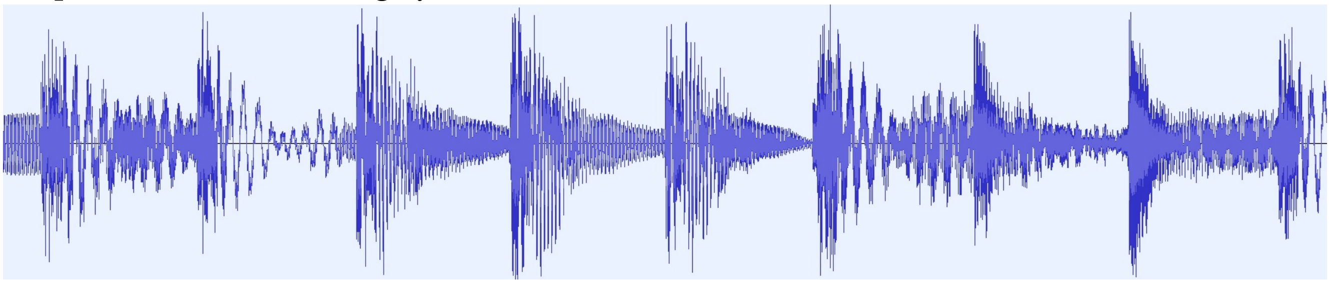
$$
\hat{\theta}_{\mathrm{ml}} - \theta = \left(\frac{-1}{\mathbf{E}\left(\frac{\partial^2}{\partial\theta^2}\ln\mathrm{p}\left[\mathrm{y}(t;\theta)|\mathrm{x}(t;\theta)\right]\right)}\right)\left(\frac{\partial}{\partial\theta}\ln\mathrm{p}\left[\mathrm{y}(t;\theta)|\mathrm{x}(t;\theta)\right]\right).
$$

$$
\begin{aligned}
\left(\frac{-1}{\mathbf{E}\left(\frac{\partial^2}{\partial\theta^2}\ln\mathrm{p}\left[\mathrm{y}(t;\theta)|\mathrm{x}(t;\theta)\right]\right)}\right)\left(\frac{\partial}{\partial\theta}\ln\mathrm{p}\left[\mathrm{y}(t;\theta)|\mathrm{x}(t;\theta)\right]\right) &= \left(\frac{\sigma^2}{N}\frac{1}{\left[\frac{\partial\mathrm{g}(\theta)}{\partial\theta}\right]^2}\right)\left(\frac{-1}{2\sigma^2}(2)\frac{\partial\mathrm{g}(\theta)}{\partial\theta}\sum_{n=1}^{N}[\dot{y}_n - \mathrm{g}(\theta)]\right) \\
&= -\frac{1}{N}\frac{1}{\frac{\partial\mathrm{g}(\theta)}{\partial\theta}}\left(\sum_{n=1}^{N}[\dot{y}_n - \mathrm{g}(\theta)]\right) \\
&= -\frac{1}{\frac{\partial\mathrm{g}(\theta)}{\partial\theta}}\left(\frac{1}{N}\sum_{n=1}^{N}\dot{y}_n - \mathrm{g}(\theta)\right) \\
&= -\frac{1}{\frac{\partial\mathrm{g}(\theta)}{\partial\theta}}\left(\hat{\theta}_{\mathrm{ml}} - \mathrm{g}(\theta)\right)
\end{aligned}
$$

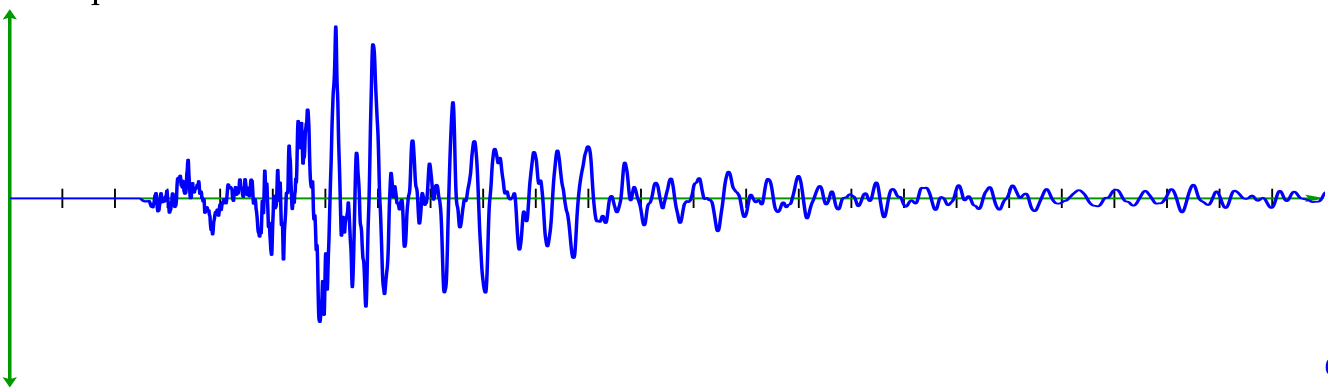
$$= -(\hat{\theta}_{\mathsf{ml}} - \theta)$$

21.5 Example data

"Pop Goes the World" song by Men Without Hats

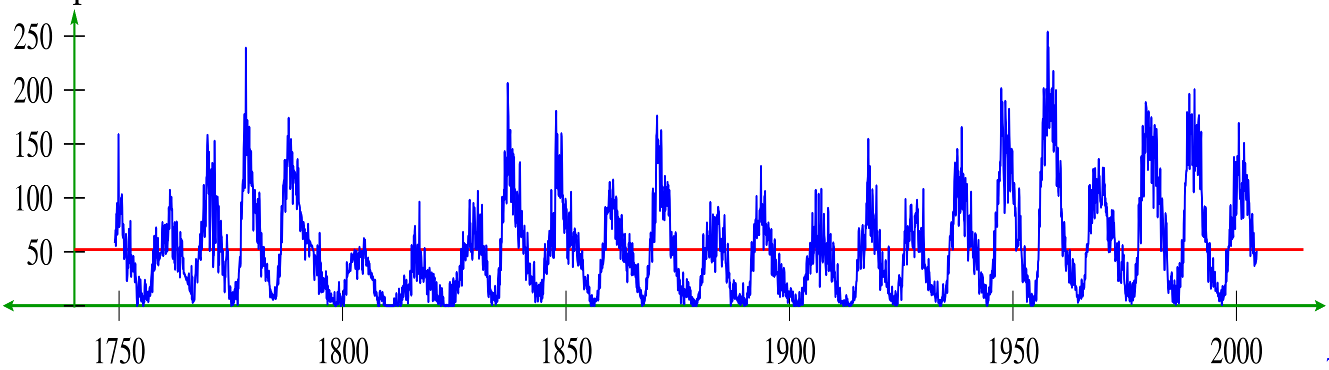


Earthquake data



[6]

Sunspot data

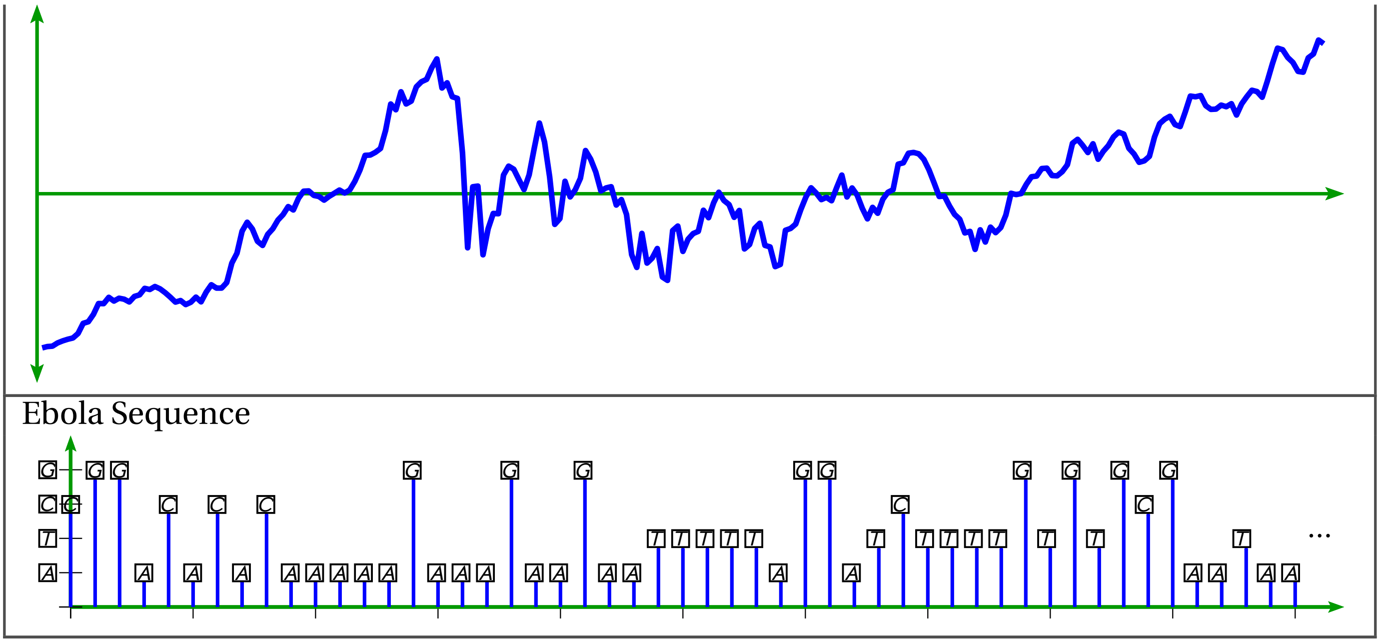


[7]

Dow Jones Industrial Average

[6] https://www.iris.edu/wilber3/find_stations/10953070

[7] https://d32ogoqmya1dw8.cloudfront.net/files/introgeo/teachingwdata/examples/GreenwichSSNvstime.txt



21.6 Matched Filter Algorithm

Let S be the set of transmitted waveforms and Y be a set of orthonormal basis functions that span S. *Signal matching* computes the innerproducts of a received signal $\mathsf{y}(t;\theta)$ with each signal from S. *Orthonormal decomposition* computes the innerproducts of $\mathsf{y}(t;\theta)$ with each signal from the set Y.

In the case where $|S|$ is large, often $|Y| \ll |S|$ making orthonormal decomposition much easier to implement. For example, in a QAM-64 modulation system, signal matching requires $|S| = 64$ innerproduct calculations, while orthonormal decomposition only requires $|Y| = 2$ innerproduct calculations because all 64 signals in S can be spanned by just 2 orthonormal basis functions.

Maximizing SNR. Theorem 21.1 (page 151) shows that the innerproducts of $\mathsf{y}(t;\theta)$ with basis functions of Y is *sufficient* for optimal detection. Theorem 21.10 (page 165) (next) shows that a receiver can maximize the SNR of a received signal when signal matching is used.

Theorem 21.10. *Let* $\mathsf{x}(t)$ *be a transmitted signal,* $\mathsf{v}(t)$ *noise, and* $\mathsf{y}(t;\theta)$ *the received signal in an AWGN channel. Let the* SIGNAL TO NOISE RATIO *SNR be defined as*

$$\mathrm{SNR}[\mathsf{y}(t;\theta)] \triangleq \frac{|\langle \mathsf{x}(t) \,|\, \mathsf{x}(t)\rangle|^2}{\mathbf{E}\left[|\langle \mathsf{v}(t) \,|\, \mathsf{x}(t)\rangle|^2\right]}.$$

THM

$$\mathrm{SNR}[\mathsf{y}(t;\theta)] \le \frac{2\,\|\mathsf{x}(t)\|^2}{N_o} \qquad \textit{and is maximized (equality) when } \mathsf{x}(t) = a\mathsf{x}(t), \textit{ where } a \in \mathbb{R}.$$

✎PROOF:

$$\begin{aligned}
\mathrm{SNR}[\mathsf{y}(t;\theta)] &\triangleq \frac{|\langle \mathsf{x}(t) \,|\, \mathsf{x}(t)\rangle|^2}{\mathbf{E}\left[|\langle \mathsf{v}(t) \,|\, \mathsf{x}(t)\rangle|^2\right]} \\
&= \frac{|\langle \mathsf{x}(t) \,|\, f(t)\rangle|^2}{\mathbf{E}\left[\left[\int_{t\in\mathbb{R}} \mathsf{v}(t)\mathsf{x}^*(t)\,dt\right]\left[\int_{\hat{\theta}} n(\hat{\theta}) f^*(\hat{\theta})\,du\right]^*\right]}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{|\langle \mathsf{x}(t) \mid \mathsf{x}(t)\rangle|^2}{\mathbf{E}\left[\int_{t\in\mathbb{R}}\int_{\hat\theta} \mathsf{v}(t)n^*(\hat\theta)\mathsf{x}^*(t)\mathsf{x}(\hat\theta)\ \mathsf{d}t\ \mathsf{d}u\right]}\\
&= \frac{|\langle \mathsf{x}(t) \mid f(t)\rangle|^2}{\int_{t\in\mathbb{R}}\int_{\hat\theta} \mathbf{E}\left[\mathsf{v}(t)n^*(\hat\theta)\right]\mathsf{x}^*(t)\mathsf{x}(\hat\theta)\ \mathsf{d}t\ \mathsf{d}u}\\
&= \frac{|\langle \mathsf{x}(t) \mid \mathsf{x}(t)\rangle|^2}{\int_{t\in\mathbb{R}}\int_{\hat\theta} \frac{1}{2}N_o\delta(t-\hat\theta)\mathsf{x}^*(t)\mathsf{x}(\hat\theta)\ \mathsf{d}t\ \mathsf{d}u}\\
&= \frac{|\langle \mathsf{x}(t) \mid \mathsf{x}(t)\rangle|^2}{\frac{1}{2}N_o\int_{t\in\mathbb{R}} \mathsf{x}^*(t)\mathsf{x}(t)\ \mathsf{d}t}\\
&= \frac{|\langle \mathsf{x}(t) \mid \mathsf{x}(t)\rangle|^2}{\frac{1}{2}N_o\,\|\mathsf{x}(t)\|^2}\\
&\le \frac{|\|\mathsf{x}(t)\|\ \ \|\mathsf{x}(t)\||^2}{\frac{1}{2}N_o\,\|\mathsf{x}(t)\|^2} && \text{by } \textit{Cauchy-Schwarz Inequality} \quad \text{(Theorem N.2 page 310)}\\
&= \frac{2\,\|\mathsf{x}(t)\|^2}{N_o}
\end{aligned}
$$

The Cauchy-Schwarz Inequality becomes an equality (SNR is maximized) when $\mathsf{x}(t) = a\mathsf{x}(t)$. ✎

Implementation. The innerproduct operations can be implemented using either

1. a correlator or
2. a matched filter.

A correlator is simply an integrator of the form $\langle \mathsf{y}(t;\theta) \mid f(t)\rangle = \int_0^T \mathsf{y}(t;\theta)f(t)\ \mathsf{d}t$.

A matched filter introduces a function $\mathsf{h}(t)$ such that $\mathsf{h}(t) = \mathsf{x}(T-t)$ (which implies $\mathsf{x}(t) = h(T-t)$) giving

$$
\underbrace{\langle \mathsf{y}(t;\theta) \mid \mathsf{x}(t)\rangle = \int_0^T \mathsf{y}(t;\theta)\mathsf{x}(t)\ \mathsf{d}t}_{\text{correlator}} = \underbrace{\int_0^\infty \mathsf{x}(\tau)h(t-\tau)\ \mathsf{d}\tau\bigg|_{t=T} = \mathsf{x}(t)\star\mathsf{h}(t)|_{t=T}}_{\text{matched filter}}.
$$

This shows that $\mathsf{h}(t)$ is the impulse response of a filter operation sampled at time τ. By Theorem 21.10 (page 165), the optimal impulse response is $\mathsf{h}(\tau - t) = \mathsf{f}(t) = \mathsf{x}(t)$. That is, the optimal $\mathsf{h}(t)$ is just a “flipped” and shifted version of $\mathsf{x}(t)$.

21.7 Colored noise

This chapter presented several theorems whose results depended on the noise being white. However if the noise is **colored**, then these results are invalid. But there is still hope for colored noise. Processing colored signals can be accomplished using two techniques:

1. Karhunen-Loève basis functions (Section 5.1 page 25)

2. whitening filter [8]

Karhunen-Loève. If the noise is *white*, the set $\left\{\langle \mathsf{y}(t;\theta)\,|\,\psi_n(t)\rangle\,|n=1,2,\ldots,N\right\}$ is a *sufficient statistic* regardless of which set $\left\{\psi_n(t)\right\}$ of orthonormal basis functions are used. If the noise is *colored*, and if $\{\psi_n(t)\}$ satisfy the Karhunen-Loève criterion

$$\int_{t_2} \mathsf{R}_{\mathsf{xx}}(t,u)\psi_n(u)\,\mathsf{d}u = \lambda_n\psi_n(t)$$

then the set $\left\{\langle \mathsf{y}(t;\theta)\,|\,\psi_n(t)\rangle\right\}$ is still a *sufficient statistic*.

Whitening filter. The whitening filter makes the received signal $\mathsf{y}(t;\theta)$ statistically white (uncorrelated in time). In this case, any orthonormal basis set can be used to generate sufficient statistics.

Wavelets. Wavelets have the property that they tend to whiten data. For more information, see Walter and Shen (2001) pages 329–350 ⟨"Chapter 14 Orthogonal Systems and Stochastic Processes"⟩, Mallat (1999) ⟨ISBN:9780124666061⟩, Johnstone and Silverman (1997), Wornell and Oppenheim (1992), and Vidakovic (1999) pages 10–14 ⟨"Example 1.2.5 Wavelets whiten data"⟩ (first four references cited by B. Vidakovic).

[8] *Continuous data whitening*: Section 8.3 page 64
Discrete data whitening: Section 7.4 page 54

Part VI

Appendices

APPENDIX A
PROBABILITY SPACE

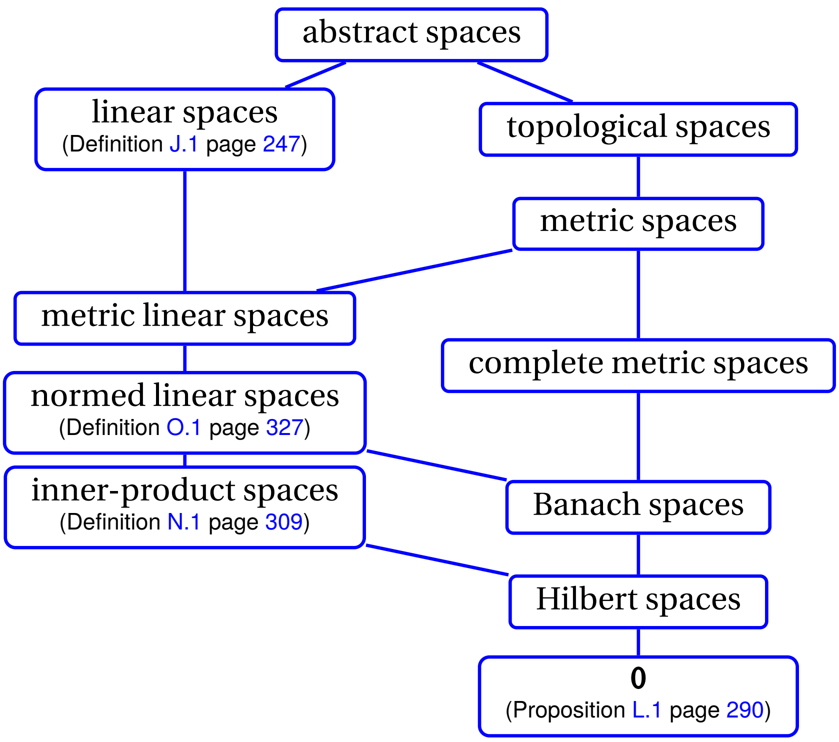


Figure A.1: Lattice of mathematical spaces

"It is not certain that everything is certain."
Blaise Pascal (1623–1662), mathematician [1]

[1] quote: http://en.wikiquote.org/wiki/Blaise_Pascal
image: http://en.wikipedia.org/wiki/Image:Blaise_pascal.jpg

A.1 Probability functions

Definition A.1. [2] *Let* $(X, \vee, \wedge, \neg, 0, 1; \leq)$ *be a* LATTICE WITH NEGATION.

DEF

The function P *is a* ***probability function*** *if*

$$\begin{array}{llcll}
(1). & \mathsf{P}(1) & = & 1 & & (\text{NORMALIZED}) & \text{and}\\
(2). & \mathsf{P}(x) & \geq & 0 & \forall x \in X & (\text{NONNEGATIVE}) & \text{and}\\
(3). \; \bigwedge_{n=1}^{\infty} x_n = 0 \implies & \mathsf{P}\left(\bigvee_{n=1}^{\infty} x_n\right) & = & \sum_{n=1}^{\infty} \mathsf{P}(x_n) & \forall x_n \in X & (\sigma\text{-ADDITIVE}) & .
\end{array}$$

Remark A.1. The advantage of this definition is that P is a *measure*, and hence all the power of measure theory is subsequently at one's disposal in using P. However, it has often been argued that the requirement of *σ-additivity* is unnecessary for a probability function. Even as early as 1930, de Finetti argued against it, in what became a kind of polite running debate with Fréchet.[3] In fact, Kolmogorov himself provided some argument against *σ-additivity* when referring to the closely related *Axiom of Continuity* saying, "Since the new axiom is essential for infinite fields of probability only, it is almost impossible to elucidate its empirical meaning…For, in describing any observable random process we can obtain only finite fields of probability.…" But in its support he added, "This limitation has been found expedient in researches of the most diverse sort."[4]

There are several other definitions of probability that only require *additivity* rather than *σ-additivity*. On a *Boolean lattice*, the **traditional probability** function is defined as[5]

$$\begin{array}{llcll}
(1). & \mathsf{P}(1) & = & 1 & & (\textit{normalized}) & \text{and}\\
(2). & \mathsf{P}(x) & \geq & 0 & \forall x \in X & (\textit{nonnegative}) & \text{and}\\
(3). \; x \wedge y = 0 \implies & \mathsf{P}(x \vee y) & = & \mathsf{P}(x) + \mathsf{P}(y) & \forall x,y \in X & (\textit{additive}) & .
\end{array}$$

This definition implies (on a *Boolean lattice*) that

$$\begin{array}{llcll}
(a). & \mathsf{P}(0) & = & 0 & & (\textit{nondegenerate}) & \text{and}\\
(b). & \mathsf{P}(x) & \leq & 1 & \forall x \in X & (\textit{upper bounded}) & \text{and}\\
(c). & \mathsf{P}(x) & = & 1 - \mathsf{P}(x^{\perp}) & \forall x \in X & & \text{and}\\
(d). & \mathsf{P}(x \vee y) & \leq & \mathsf{P}(x) + \mathsf{P}(y) & \forall x,y \in X & (\textit{subadditive}) & \text{and}\\
(e). & \mathsf{P}(x \vee y) & = & \mathsf{P}(x) + \mathsf{P}(y) - \mathsf{P}(x \wedge y) & \forall x,y \in X & & \text{and}\\
(f). \; x \leq y \implies & \mathsf{P}(x) & \leq & \mathsf{P}(y) & \forall x,y \in X & (\textit{monotone}) & .
\end{array}$$

On a *distributive pseudocomplemented lattice*, the **generalized probability** function has been defined as[6]

$$\begin{array}{llcll}
(1). & \mathsf{P}(0) & = & 0 & (\textit{nondegenerate}) & \text{and}\\
(2). & \mathsf{P}(1) & = & 1 & (\textit{normalized}) & \text{and}\\
(3). \; 0 \leq & \mathsf{P}(1) & \leq & 1 & & \text{and}\\
(4). & \mathsf{P}(x \vee y) & = & \mathsf{P}(x) + \mathsf{P}(y) - \mathsf{P}(x \wedge y) \quad \forall x,y \in X & & .
\end{array}$$

On an *orthomodular lattice*, or a *finite modular lattice*, the **quantum probability** function is defined as[7]

$$\begin{array}{llcll}
(1). & \mathsf{P}(0) & = & 0 & & (\textit{nondegenerate}) & \text{and}\\
(2). & \mathsf{P}(1) & = & 1 & & (\textit{normalized}) & \text{and}\\
(3). \; x \perp y \implies & \mathsf{P}(x \vee y) & = & \mathsf{P}(x) + \mathsf{P}(y) & \forall x,y \in X & (\textit{additive}) & .
\end{array}$$

However, for lattices that are not *distributive*, *modular*, or *orthomodular*, none of these definitions

[2] Billingsley (1995) pages 22–23 ⟨Probability Measures⟩, Kolmogorov (1933a), Kolmogorov (1933b) page 16 ⟨*field of probability*⟩, Pap (1995) pages 8–9 ⟨Definition 2.3(13)⟩, Kalmbach (1986) page 27

[3] de Finetti (1930a), Fréchet (1930a), de Finetti (1930b), Fréchet (1930b), de Finetti (1930c), Cifarelli and Regazzini (1996) pages 258–260

[4] Kolmogorov (1933b) page 15

[5] Papoulis (1991) pages 21–22, Kolmogorov (1933b) page 2 ⟨§1. Axioms I–V⟩

[6] Narens (2014) page 118, Narens (2007) ⟨ISBN:9812708014⟩

[7] Greechie (1971) page 126 ⟨DEFINITIONS⟩, Narens (2014) page 118

work out so well. Take for example the O_6 *lattice* with the "very reasonable" probability function given in Example ?? (page ??). This probability space $(\mathsf{O}_6, \mathsf{P})$ fails to be any of the 4 probability functions defined in this Remark. It fails to be a *measure-theoretic* or *traditional probability* function because

$$a \wedge b = 0 \qquad \text{but} \qquad \mathsf{P}(a \vee b) = \mathsf{P}(1) = 1 \neq \tfrac{1}{3} + \tfrac{1}{2} = \mathsf{P}(a) + \mathsf{P}(b)\,.$$

It fails to be a *generalized probability* function because

$$\mathsf{P}(a \vee b) = \mathsf{P}(1) = 1 \neq \tfrac{1}{3} + \tfrac{1}{2} - 0 = \mathsf{P}(a) + \mathsf{P}(b) - \mathsf{P}(0) = \mathsf{P}(a) + \mathsf{P}(b) - \mathsf{P}(a \wedge b)\,.$$

It fails to be an *quantum probability* function because

$$a \perp b = 0 \qquad \text{but} \qquad \mathsf{P}(a \vee b) = \mathsf{P}(1) = 1 \neq \tfrac{1}{3} + \tfrac{1}{2} = \mathsf{P}(a) + \mathsf{P}(b)\,.$$

In each of these cases, the function P fails to be *additive*. The solution of Definition A.1 (page 172) is simply to "switch off" *additivity* when the lattice is not *distributive*. This method is a little "crude", but at least it allows us to define probability on a very wide class of lattices, while retaining compatibility with the *Boolean* case.

A.2 Probability Space

In mathematics, a *space* is simply a set and in the most general definition, nothing else. However, normally for a space to actually be useful, some additional structure is added. One of the most general additional structures is a *topology*; and a space together with a topology is called a *topological space*. A topological space imposes additional structure on a space in the form of subsets and guarantees that these subsets are closed under such fundamental operations as set *union* and set *intersection*. With the additional structure available in a topological space, we are able to analyze such basic concepts as continuity, convergence, and connectivity.

However for a great number of mathematical applications, we need to *measure* mathematical objects—the most general measurement being measures on subsets of some set. Examples of measurement in mathematics include integration and probability. Before measurement can be effectively performed on a set, the set must be equipped with a subset structure. In analysis, arguably the most fundamental subset structure is the humble *topology* (Definition ?? page ??). However, a simple topology does not provide sufficient structure for effective measurement. For example, often we would not only like to measure some subset A, but also its complement A^c. A topology is not closed under the complement operation. So instead of a topology only, we equip the space with a more powerful (and thus less general) structure called a *σ-algebra* (*sigma-algebra*) (Definition ?? page ??). A *σ-algebra* is a subset structure that is closed under set complement. A set together with a *σ-algebra* is called a *measurable space*. And a set together with a *σ-algebra* and a *measure* on that *σ-algebra* is called a *measure space* (Definition ?? page ??).

The next definition presents a very important measure space—the *probability space*.

Definition A.2.

DEF

The triple $(\Omega, \mathbb{E}, \mathsf{P})$ *is a* ***probability space*** *if*

(1). Ω *is a* SET *and*

(2). $\mathbb{E}$ *is a* σ-ALGEBRA *on* Ω *(Definition ?? page ??) and*

(3). $\mathsf{P} : \mathbb{E} \to [0,1]$ *is a* MEASURE *on* $\mathbb{E}$ *(Definition ?? page ??) .*

If $\boldsymbol{S} \triangleq (\Omega, \mathbb{E}, \mathsf{P})$ *is a* PROBABILITY SPACE *then* x *is an* ***outcome*** *in* $\boldsymbol{S}$ *if* $x \in \Omega$*,* A *is an* ***event*** *in* $\boldsymbol{S}$ *if* $A \in \mathbb{E}$*, and* $\mathsf{P}A$ *is the* ***probability*** *of* A *in* $\boldsymbol{S}$ *if* A *is an* EVENT *in* $\boldsymbol{S}$*.*

Definition A.3. [8] *Let* $(\Omega, \mathbb{E}, \mathsf{P})$ *be a* PROBABILITY SPACE *(Definition A.2 page 173).*

[8] Papoulis (1990) page 52 ⟨Independent Events⟩

DEF: *Two* EVENTS A *and* B *in* $\mathbb{E}$ *are* ***independent*** *if*

$$\mathsf{P}(A \cap B) = \mathsf{P}(A)\mathsf{P}(B)$$

Definition A.4. [9] *Let* $(\Omega, \mathbb{E}, \mathsf{P})$ *be a* PROBABILITY SPACE (Definition A.2 page 173). *Let* x *and* y *be* EVENTS *in* L.

DEF: *The* ***conditional probability of*** x ***given*** y *is defined as*

$$\mathsf{P}(A|B) \triangleq \frac{\mathsf{P}(A \cap B)}{\mathsf{P}(B)}$$

A.3 Properties

Proposition A.1.

PRP: $$\underbrace{(\Omega, \mathbb{E}, \mathsf{P}) \text{ is a PROBABILITY SPACE} \implies (\Omega, \mathbb{E}, \mathsf{P}) \text{ is a MEASURE SPACE}}_{\text{(every probability space is a measure space)}}$$

Theorem A.1. [10] *Let* $(\Omega, \mathbb{E}, \mathsf{P})$ *be a* PROBABILITY SPACE (Definition A.2 page 173).

THM:

(1).	0	$\le$	$\mathsf{P}(x) \le 1$	$\forall x \in X$	(BOUNDED)	and
(2).	$\mathsf{P}(x)$	$=$	$1 - \mathsf{P}(x^\perp)$	$\forall x \in X$	(PARTITION OF UNITY)	and
(3).	$x \le y$	$\implies$	$\mathsf{P}(y^\perp) \le \mathsf{P}(x^\perp)$	$\forall x, y \in X$	(ANTITONE)	

✎PROOF:

1. Proof for $0 \le \mathsf{P}(x) \le 1$:

$$\begin{aligned}
\boxed{0} &= \mathsf{P}(0) && \text{by by } \textit{nondegenerate} \text{ property of } \mathsf{P} \text{ (Definition A.2 page 173)}\\
&\le \boxed{\mathsf{P}(x)} && \text{because } 0 \le x \text{ and } \textit{monotone} \text{ property of } \mathsf{P}\\
&\le \mathsf{P}(1) && \text{because } x \le 1 \text{ and } \textit{monotone} \text{ property of } \mathsf{P}\\
&= \boxed{1} && \text{by } \textit{normalized} \text{ property of } \mathsf{P}
\end{aligned}$$

2. Proof for $\mathsf{P}(x) = 1 - \mathsf{P}(x^\perp)$:
   (a) Proof that P is *additive* (Definition A.2 page 173) over $\{0, x, x^\perp\} \subseteq X$:
      i. $\{0, x, x^\perp\}$ is *distributive*.
      ii. $x \wedge x^\perp = 0$ for all $x \in X$ by the *non-contradiction* property of *orthocomplemented lattices*.
      iii. Therefore, by Definition A.2, P is *additive* over $\{0, x, x^\perp\}$.
   (b) Then ...

$$\begin{aligned}
1 - \mathsf{P}(x^\perp) &= \mathsf{P}(1) - \mathsf{P}(x^\perp) && \text{by } \textit{normalized} \text{ property of } \mathsf{P} && \text{(Definition A.2 page 173)}\\
&= \mathsf{P}(x \vee x^\perp) - \mathsf{P}(x^\perp) && \text{by } \textit{excluded middle} \text{ property of ortho. lat.}\\
&= \mathsf{P}(x) + \mathsf{P}(x^\perp) - \mathsf{P}(x^\perp) && \text{by } \textit{additive} \text{ property of } (\Omega, \mathbb{E}, \mathsf{P}) && \text{(item (2a) page 174)}\\
&= \mathsf{P}(x) && \text{by field property of } (\mathbb{R}, +, \cdot, 0, 1)
\end{aligned}$$

3. Proof for $x \le y \implies \mathsf{P}(y^\perp) \le \mathsf{P}(x^\perp)$:

$$\begin{aligned}
x \le y &\implies y^\perp \le x^\perp && \text{by } \textit{antitone} \text{ property of } \textit{orthocomplemented lattices}\\
&\implies \mathsf{P}(y^\perp) \le \mathsf{P}(x^\perp) && \text{by } \textit{monotone} \text{ property of } \mathsf{P} && \text{(Definition A.2 page 173)}
\end{aligned}$$

[9] Papoulis (1990) page 45 ⟨2-3 Conditional Probability and Indenpendence⟩
[10] property (1): Papoulis (1991) page 21 ⟨(2-11)⟩

✏

Theorem A.2. [11] *Let* $(\Omega, \mathbb{E}, \mathsf{P})$ *be a* PROBABILITY SPACE *(Definition A.2 page 173)*.

THM

$$\left\{\begin{array}{l} \boldsymbol{L} \text{ is BOOLEAN} \\ \text{(Definition ?? page ??)} \end{array}\right\} \implies \left\{\begin{array}{llcll} 1. & \mathsf{P}(x \vee y) & = & \mathsf{P}(x) + \mathsf{P}(y) - \mathsf{P}(x \wedge y) & \forall x,y \in X \quad \text{and} \\ 2. & \mathsf{P}(x \vee y) & \le & \mathsf{P}(x) + \mathsf{P}(y) & \forall x,y \in X \quad \text{(BOOLE'S INEQUALITY)} \end{array}\right.$$

✎PROOF:

1. lemma: Proof that $\mathsf{P}((\neg x) \wedge y) = \mathsf{P}(y) - \mathsf{P}(x \wedge y)$:

$$\begin{aligned}
\mathsf{P}(y) - \mathsf{P}(xy) &= \mathsf{P}(1 \wedge y) - \mathsf{P}(xy) && \text{by definition of 1 and } \wedge \\
&= \mathsf{P}\big[(x \vee x^{\perp})y\big] - \mathsf{P}(xy) && \text{by *excluded middle* property of *Boolean lattices*} \\
&= \mathsf{P}(xy \vee x^{\perp}y) - \mathsf{P}(xy) && \text{by *distributive* property of *Boolean lattices*} \\
&= \mathsf{P}(xy) + \mathsf{P}(x^{\perp}y) - \mathsf{P}(xy) && \text{because } (xy)(x^{\perp}y) = 0 \text{ and by *additive* property} \\
&= \mathsf{P}(x^{\perp}y)
\end{aligned}$$

2. Proof that $\mathsf{P}(x \vee y) = \mathsf{P}(x) + \mathsf{P}(y) - \mathsf{P}(x \wedge y)$:

$$\begin{aligned}
\mathsf{P}(x \vee y) &= \mathsf{P}(x \vee x^{\perp}y) && \text{by property of *Boolean lattices*} \\
&= \mathsf{P}(x) + \mathsf{P}(x^{\perp}y) && \text{because } (x)(x^{\perp}y) = 0 \text{ and by *additive* property} \\
&= \mathsf{P}(x) + \mathsf{P}(y) - \mathsf{P}(x \wedge y) && \text{by item (1) (page 175)}
\end{aligned}$$

✏

Theorem A.3 (sum of products). *Let* $(X, \vee, \wedge, 0, 1; \le)$ *be a* BOUNDED LATTICE, $(\Omega, \mathbb{E}, \mathsf{P})$ *a* PROBABILITY SPACE *(Definition A.2 page 173), and* $\{y, x_1, x_2, x_3, \ldots\}$ *a subset of* X.

THM

$$\left\{\begin{array}{lll} 1. & \boldsymbol{L} \text{ is DISTRIBUTIVE} & \text{and} \\ 2. & \{x_1, x_2, \ldots\} \text{ is a PARTITION of } y & \end{array}\right\} \implies \left\{\begin{array}{lrcll} 1. & \mathsf{P}(y) & = & \sum_n \mathsf{P}(x_n) & \text{and} \\ 2. & \mathsf{P}(y) & = & \sum_n \mathsf{P}(y \wedge x_n) & \text{and} \\ 3. & \mathsf{P}(z \wedge y) & = & \sum_n \mathsf{P}(z \wedge x_n) & \end{array}\right\}$$

✎PROOF:

1. Proof that P is *additive* (Definition A.2 page 173) on $(\Omega, \mathbb{E}, \mathsf{P})$:

   (a) Proof that $(yx_n) \wedge (yx_m) = 0$ for $n \neq m$:

$$\begin{aligned}
(yx_n) \wedge (yx_m) &= y(x_n x_m) && \text{by definition of } \wedge \\
&= y \wedge 0 && \text{by *mutually exclusive* property of *partitions*} \\
&= 0 && \text{by *lower bounded* property of *bounded lattices*}
\end{aligned}$$

   (b) Proof that $\boldsymbol{L}$ is *distributive*: by *distributive* hypothesis

2. Proof that $\mathsf{P}(y) = \sum_n \mathsf{P}(x_n)$

$$\begin{aligned}
\mathsf{P}(y) &= \mathsf{P}\big(yx_1 \vee yx_2 \vee \cdots \vee yx_n\big) \\
&= \sum_n \mathsf{P}\big(yx_n\big) && \text{by item (1) and *additive* property} && \text{(Definition A.2 page 173)} \\
&= \sum_n \mathsf{P}\big(y|x_n\big)\mathsf{P}(x_n) && \text{by *conditional probability*} && \text{(Definition A.4 page 174)}
\end{aligned}$$

[11] Papoulis (1991) page 21 ⟨(2-13)⟩, Feller (1970) pages 22–23 ⟨(7.4),(7.6)⟩

✏

As described in Definition A.2 (page 173), every *probability space* $(\Omega, \mathbb{E}, \mathsf{P})$ contains a probability *measure* $\mathsf{P} : \mathbb{E} \to [0,1]$. This probability *measure* has some basic properties as described in Theorem A.4 (next).

Theorem A.4. *Let* $(\Omega, \mathbb{E}, \mathsf{P})$ *be a* PROBABILITY SPACE. *Let* B *be a set and* $\{B_n | n = 1, 2, \ldots, N\}$ *a set of sets.*

THM

$$\left\{ \begin{array}{l} \{B_n | n = 1, 2, \ldots, N\} \text{ is a} \\ \text{PARTITION of } B. \end{array} \right\} \implies \left\{ \begin{array}{llcll} (1). & \mathsf{P}(B) & = & \sum_{n=1}^{N} \mathsf{P}(B_n) & \forall B \in \mathbb{E} \quad \text{and} \\ (2). & \mathsf{P}(AB) & = & \sum_{n=1}^{N} \mathsf{P}(AB_n) & \forall A, B \in \mathbb{E} \end{array} \right\}$$

✎PROOF: P is a *measure* and by Definition ?? (page ??). ✏

Proposition A.2. *Let* $(\Omega, \mathbb{E}, \mathsf{P})$ *be a probability space, and* X *a* RANDOM VARIABLE *(Definition B.1 page 184) with* PROBABILITY DENSITY FUNCTION $\mathsf{p}_\mathsf{x}(x)$ *and* CUMULATIVE DISTRIBUTION FUNCTION $\mathsf{c}_\mathsf{x}(x)$.

PRP

$$\begin{array}{llcll} (1). & \mathsf{c}_\mathsf{x}(x) \text{ is MONOTONE} & & & \text{and} \\ (2). & \mathsf{p}_\mathsf{x}(x) \text{ is CONTINUOUS} & \implies & \mathsf{c}_\mathsf{x}(x) \text{ is STRICTLY MONOTONE} & \text{and} \\ (3). & \mathsf{p}_\mathsf{x}(x) \text{ is CONTINUOUS} & \implies & \mathsf{c}_\mathsf{x}(x) \text{ is INVERTIBLE} & \end{array}$$

Theorem A.5 (Bayes' Rule). [12]

THM

$$\mathsf{P}(X|Y) = \frac{\mathsf{P}(Y|X)\mathsf{P}(X)}{\mathsf{P}(Y)}$$

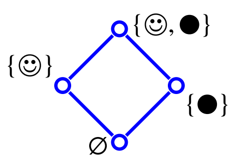
$$\begin{aligned} \mathsf{P}(X|Y) &\triangleq \frac{\mathsf{P}(X \cap Y)}{\mathsf{P}(Y)} && \text{by definition of } \textit{conditional probability} && \text{(Definition A.4 page 174)} \\ &= \frac{\mathsf{P}(Y \cap X)}{\mathsf{P}(Y)} && \text{by } \textit{commutative} \text{ property of } \cap \\ &= \frac{\mathsf{P}(Y|X)\mathsf{P}(X)}{\mathsf{P}(Y)} && \text{by definition of } \textit{conditional probability} && \text{(Definition A.4 page 174)} \end{aligned}$$

A.4 Examples

Example A.1 (single coin toss). Let ☺ represent "heads" and ● represent "tails" in a coin toss. Let $0 < p < 1$ be the probability of a head. A *probability space* $(\Omega, \mathbb{E}, \mathsf{P})$ for a single coin toss is as follows:

EX

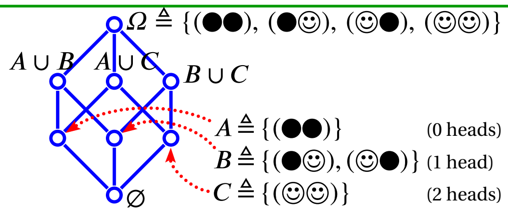
$$\begin{aligned} \Omega &= \{☺, ●\} \\ \mathbb{E} &= \left\{ \underbrace{\varnothing}_{\text{neither}}, \underbrace{\{☺\}}_{\text{heads}}, \underbrace{\{●\}}_{\text{tails}}, \underbrace{\{☺, ●\}}_{\text{heads or tails}} \right\} \\ \mathsf{P}(X) &= \left\{ \begin{array}{llcll} 0 & \text{for } X & = & \varnothing & \text{(neither heads nor tails)} \\ p & \text{for } X & = & \{☺\} & \text{(heads)} \\ 1-p & \text{for } X & = & \{●\} & \text{(tails)} \\ 1 & \text{for } X & = & \{☺, ●\} & \text{(either heads or tails)} \end{array} \right\} \end{aligned}$$



$$\begin{aligned} \mathsf{P}(\{☺, ●\}) &= 1 \\ \mathsf{P}(\{☺\}) &= p \\ \mathsf{P}(\{●\}) &= 1-p \\ \mathsf{P}(\varnothing) &= 0 \end{aligned}$$

[12] Haykin (2014) page 95 ⟨"Bayes' Rule"⟩

EX



$$\begin{aligned} \mathsf{P}(\varnothing) &= 0 \\ \mathsf{P}(A) &= (1-p)^2 \\ \mathsf{P}(B) &= 2p(1-p) \\ \mathsf{P}(C) &= p^2 \end{aligned} \qquad \begin{aligned} \mathsf{P}(\Omega) &= 1 \\ \mathsf{P}(A\cup B) &= (1-p)^2 + 2p(1-p) \\ \mathsf{P}(A\cup C) &= 2p^2 - 2p + 1 \\ \mathsf{P}(B\cup C) &= -p^2 + 2p \end{aligned}$$

Figure A.2: *Double coin toss* (Example A.2 page 177)

Example A.2 (Double coin toss). Let ☺ represent "heads" and ● represent "tails" in a double coin toss in which each toss is *independent* (Definition A.3 page 173) of the other. Let $0 < p < 1$ be the probability of a head. The *probability space* $(\Omega, \mathbb{E}, \mathsf{P})$ is illustrated in Figure A.2 (page 177).

✎PROOF:

$$\begin{aligned}
\mathsf{P}(\Omega) &= 1 && \text{by *normalized* property of } \mathsf{P} && \text{(Definition A.1 page 172)} \\
\mathsf{P}(C) &= \mathsf{P}\{☺☺\} && \text{by definition of } C \\
&= \mathsf{P}(☺)\mathsf{P}(☺) && \text{by definition of *independence*} && \text{(Definition A.3 page 173)} \\
&= p^2 && \text{by definition of } p \\
\mathsf{P}(A) &= \mathsf{P}\{●●\} && \text{by definition of } A \\
&= \mathsf{P}(●)\mathsf{P}(●) && \text{by definition of *independence*} && \text{(Definition A.3 page 173)} \\
&= \{1-\mathsf{P}(☺)\}\{1-\mathsf{P}(☺)\} && \text{by *antitone* property of } \mathsf{P} && \text{(Theorem A.1 page 174)} \\
&= (1-p)^2 && \text{by definition of } p \\
\mathsf{P}(B) &= \mathsf{P}\{(●☺),(☺●)\} && \text{by definition of } B \\
&= \mathsf{P}\{●☺\} + \mathsf{P}\{☺●\} && \text{by *additive* property of } \mathsf{P} && \text{(Definition A.1 page 172)} \\
&= \mathsf{P}\{●\}\mathsf{P}\{☺\} + \mathsf{P}\{☺\}\mathsf{P}\{●\} && \text{by definition of *independence*} && \text{(Definition A.3 page 173)} \\
&= (1-p)p + p(1-p) && \text{by *antitone* property of } \mathsf{P} && \text{(Theorem A.1 page 174) and definition of } p \\
&= -2p^2 + p + 1 \\
\mathsf{P}(A\cup B) &= \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(A\cap B) && \text{by Theorem A.2} \\
&= \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(\varnothing) \\
&= (1-p)^2 + (-2p^2 + p + 1) + 0 && \text{by previous results} \\
&= -p^2 - p + 1 \\
\mathsf{P}(\varnothing) &= 0 && \text{by *nondegenerate* property of } \mathsf{P} && \text{(Definition A.1 page 172)}
\end{aligned}$$

✑

EX

$$\Omega = \{⚀,⚁,⚂,⚃,⚄,⚅\}$$

$$\mathbb{E} = \left\{ \underbrace{\{\ \}}_{\varnothing}, \underbrace{\{⚀,⚂,⚄\}}_{\text{odd}}, \underbrace{\{⚁,⚃,⚅\}}_{\text{even}}, \underbrace{\{⚀,⚁,⚂,⚃,⚄,⚅\}}_{\Omega} \right\}$$

$$\mathsf{P}(X) = \left\{ \begin{array}{lll} 0 & \text{for } e = \{\ \} & (\varnothing) \\ 1 & \text{for } e = \{⚀,⚁,⚂,⚃,⚄,⚅\} & (\Omega) \\ p & \text{for } e = \{⚀,⚂,⚄\} & (\text{odd}) \\ 1-p & \text{for } e = \{⚁,⚃,⚅\} & (\text{even}) \end{array} \right\}$$

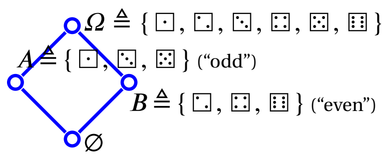


Figure A.3: *even/odd die probability space* (Example A.3 page 177)

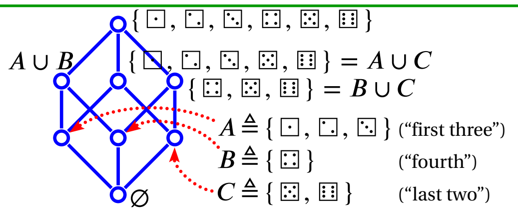
Example A.3 (even/odd die toss). The *probability space* for an **even/odd die toss**, with $0 < p < 1$ being the probability of the die toss being odd, is illustrated in Figure A.3 (page 177).

✎PROOF:

$$
\begin{array}{rcll}
\mathsf{P}(\Omega) &=& 1 & \text{by \textit{normalized} property of } \mathsf{P} \quad \text{(Definition A.1 page 172)}\\
\mathsf{P}(C) &=& \mathsf{P}\{☺☺\} & \text{by definition of } C\\
&=& \mathsf{P}(☺)\mathsf{P}(☺) & \text{by definition of \textit{independence}} \quad \text{(Definition A.3 page 173)}\\
&=& p^2 & \text{by definition of } p\\
\mathsf{P}(A) &=& \mathsf{P}\{ ⚀, ⚂, ⚄ \} & \text{by definition of } A\\
&=& p & \text{by definition of } p\\
\mathsf{P}(B) &=& \mathsf{P}\{ ⚁, ⚃, ⚅ \} & \text{by definition of } B\\
&=& \mathsf{P}\{ ⚀, ⚂, ⚄ \}^{\mathsf{c}} & \text{by definition of set complement } \mathsf{c}\\
&=& \mathsf{P}A^{\mathsf{c}} & \text{by definition of } A\\
&=& \mathsf{P}(\neg A) & \text{by definition of } \neg\\
&=& 1 - \mathsf{P}(A) & \text{by Theorem A.1 page 174}\\
&=& 1 - p & \text{by definition of } p\\
\mathsf{P}(\varnothing) &=& 0 & \text{by \textit{nondegenerate} property of } \mathsf{P} \quad \text{(Definition A.1 page 172)}
\end{array}
$$

⇰

The two previous *even/odd die* example (Example A.5 page 178) is in essence the same as the *single coin toss* (Example A.1 page 176). The next offers a little more complexity.



$$
\begin{array}{rcl@{\qquad}rcl}
\mathsf{P}(\varnothing) &=& 0 & \mathsf{P}(\Omega) &=& 1\\
\mathsf{P}(A) &=& (1-p)^2 & \mathsf{P}(A \cup B) &=& (1-p)^2 + 2p(1-p)\\
\mathsf{P}(B) &=& 2p(1-p) & \mathsf{P}(A \cup C) &=& 2p^2 - 2p + 1\\
\mathsf{P}(C) &=& p^2 & \mathsf{P}(B \cup C) &=& -p^2 + 2p
\end{array}
$$

Figure A.4: 3-4-2 die example (Example A.4 page 178)

Example A.4. Suppose we have a "fair" die and we are primarily interested in the events of the first three { ⚀, ⚁, ⚂ }, the next two, { ⚃, ⚄ } and the final one { ⚅ }. The resulting *probability space* is illustrated in Figure A.4 (page 178).

The two previous examples (Example A.5 page 178, Example A.4 page 178) illustrate a *probabilty spaces* in which the events are *mutually exclusive*. The (next) illustrates one where events are *not*.

Example A.5. Suppose we have a "fair" die and we are primarily interested in the events of the first four ({⚀,⚁,⚂,⚃}) (that is, whether one roll of the die will produce a value in the set { ⚀, ⚁, ⚂, ⚃ }) and the last three ({⚃,⚄,⚅}) However, these events do not by themselves form a *σ-algebra*. Rather under the ∩ and ∪ operations, these two events generate a total of eight possible events that together form a *σ-algebra*. The resulting *probability space* is illustrated in Figure A.5 (page 179).

But why go through all the trouble of requiring a σ-algebra? Having a σ-algebra in place ensures that anything we might possibly want to measure *can* be measured. It makes sure all possible combinations are taken into account. And why go through the additional trouble of requiring a measure space? With a measure space available, expressing the measure over a complex set is often greatly simplified because the measure space provides nice algebraic properties (namely the *σ-additive* property). Example A.6 (next) illustrates how a rather complex σ-algebra (64 elements) can be compactly represented in a measure space.

EX

$$\begin{aligned}
\Omega &= \{⚀,⚁,⚂,⚃,⚄,⚅\}\\
\mathbb{E} &= \left\{ \underbrace{\{\ \}}_{\varnothing},\ \underbrace{\{⚀,⚁,⚂,⚃,⚄,⚅\}}_{\Omega},\ \underbrace{\{⚀,⚁,⚂,⚃\}}_{\text{first four}},\ \underbrace{\{⚃,⚄,⚅\}}_{\text{last three}},\right.\\
&\qquad \left. \underbrace{\{⚃\}}_{\{1234\}\cap\{456\}},\ \underbrace{\{⚀,⚁,⚂,⚄,⚅\}}_{\{4\}^c},\ \underbrace{\{⚄,⚅\}}_{\{4\}^c\cap\{456\}},\ \underbrace{\{⚀,⚁,⚂\}}_{\{1234\}\cap\{4\}^c} \right\}\\
\mathsf{P}(e) &= \begin{cases}
0 & \text{for } e = \{\ \} & (\varnothing)\\
1 & \text{for } e = \{⚀,⚁,⚂,⚃,⚄,⚅\} & (\Omega)\\
\frac{2}{3} & \text{for } e = \{⚀,⚁,⚂,⚃\} & (\text{first four})\\
\frac{1}{2} & \text{for } e = \{⚃,⚄,⚅\} & (\text{last three})\\
\frac{1}{6} & \text{for } e = \{⚃\} & (\{1234\}\cap\{456\})\\
\frac{5}{6} & \text{for } e = \{⚀,⚁,⚂,⚄,⚅\} & (\{4\}^c)\\
\frac{1}{3} & \text{for } e = \{⚄,⚅\} & (\{4\}^c\cap\{456\})\\
\frac{1}{2} & \text{for } e = \{⚀,⚁,⚂\} & (\{1234\}\cap\{4\}^c)
\end{cases}
\end{aligned}$$

Figure A.5: First 4 / last 3 die example (Example A.5 page 178)

Example A.6. Suppose we have a "fair" dice and we are interested in measuring over the power set of events (largest possible algebra—$2^6 = 64$ events). This leads to the probability space $(\Omega, \mathbb{E}, \mathsf{P})$ where

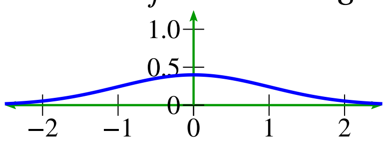
EX

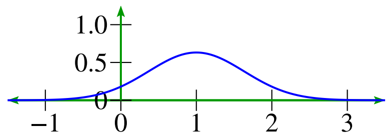
$$\begin{aligned}
\Omega &= \{⚀,⚁,⚂,⚃,⚄,⚅\}\\
\mathbb{E} &= \mathscr{P}(\Omega) && \text{(the power-set of } \Omega)\\
\mathsf{P}(e) &= \tfrac{1}{6}|e| && (\tfrac{1}{6} \text{ times the number of possible outcomes in event } e)
\end{aligned}$$

Example A.7 (Gaussian distribution on $\mathbb{R}$). Let $\boldsymbol{B}$ be the *Borel algebra on* $\mathbb{R}$. Let $\boldsymbol{L} \triangleq (\boldsymbol{B}, \subseteq)$ be the lattice formed by the elements of $\boldsymbol{B}$—this lattice is a *Boolean algebra*. Let

$$\mathsf{P}(A) \triangleq \frac{1}{\sqrt{2\pi\sigma^2}} \int_A e^{\frac{x^2}{2\sigma^2}}\, \mathrm{d}x \text{ for } A \subseteq \boldsymbol{B}$$

and where $\int$ is the *Lebesgue integral* (Definition ?? page ??). Then $(\boldsymbol{L}, \mathsf{P})$ is a **probability space**.





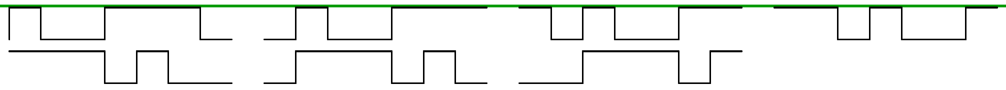
Example A.8 (Gaussian noise). Let $X \sim \mathsf{N}\left(0, \sigma^2\right)$ be a random variable with Gaussian distribution. We can construct the following probability space $(\Omega, \mathbb{E}, \mathsf{P})$:

EX

$$\begin{aligned}
\Omega &= \mathbb{R}\\
\mathbb{E} &= \{\varnothing, \Omega\} \cup \{(a,b) | a, b \in \mathbb{R}, a < b\}\\
\mathsf{P}x &= \begin{cases}
0 & \text{for } x = \varnothing\\
1 & \text{for } x = \Omega\\
\frac{1}{\sqrt{2\pi\sigma^2}} \int_a^b e^{\frac{x^2}{2\sigma^2}}\, \mathrm{d}x & \text{otherwise}
\end{cases}
\end{aligned}$$

Example A.9. The set of outcomes Ω can also be a set of waveforms:

EX

$\Omega = \{$  $\}$

$\mathbb{E} = \mathscr{P}(\Omega)$

$\mathsf{P}(e) = \frac{1}{7}|e|$

A.5 Probability subspaces

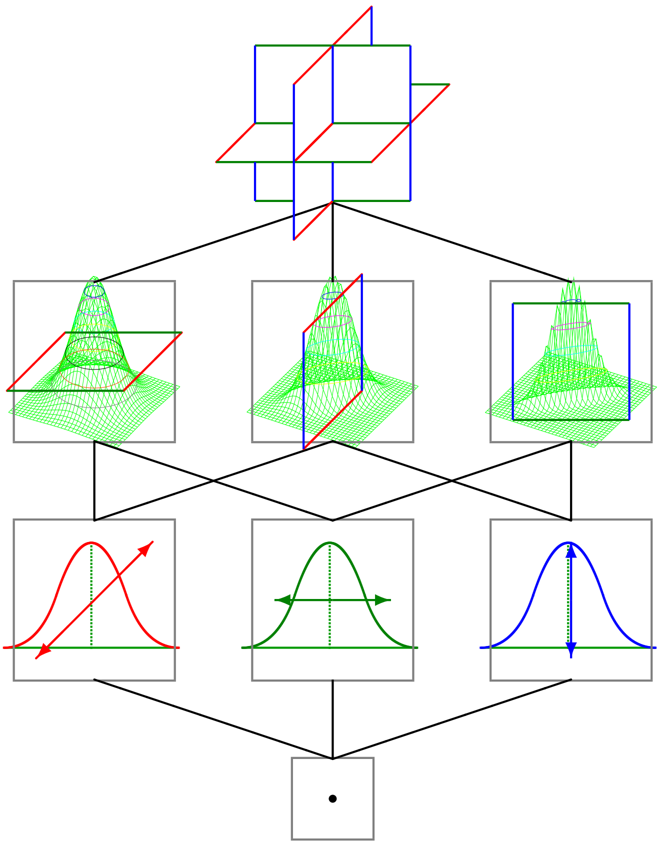
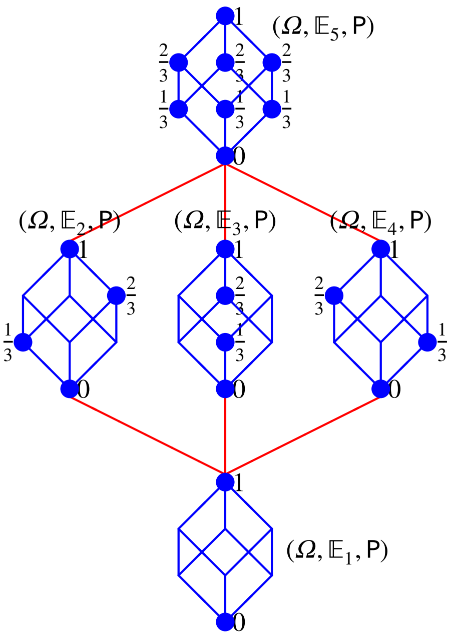


Figure A.6: Euclidean 3-dimensional space partitioned as a power lattice

Example A.10. Suppose a random process is capable of producing three values $\Omega \triangleq \{x, y, z\}$. There are five *algebras of sets* on Ω and therefore five probability spaces $(\Omega, \mathbb{E}_n, \mathsf{P})$ on Ω with the five values of $\mathbb{E}_n$ listed below: [13]

[13] *algebra of sets*: Definition ?? page ??



$$
\begin{array}{lllllllllll}
\mathbb{E}_1 = \{ & \varnothing, & & & & & & & X & \} \\
\mathbb{E}_2 = \{ & \varnothing, & \{x\}, & & & & & \{y,z\}, & X & \} \\
\mathbb{E}_3 = \{ & \varnothing, & & \{y\}, & & & \{x,z\}, & & X & \} \\
\mathbb{E}_4 = \{ & \varnothing, & & & \{z\}, & \{x,y\}, & & & X & \} \\
\mathbb{E}_5 = \{ & \varnothing, & \{x\}, & \{y\}, & \{z\}, & \{x,y\}, & \{x,z\}, & \{y,z\}, & X & \}
\end{array}
$$

Suppose the samples of Ω are generated by a physical process such that they are all equally likely to occur. Then by varying the *σ-algebra* $\mathbb{E}_n$ effectively varies the probability distribution of the probability space $(\Omega, \mathbb{E}_n, \mathsf{P})$. This is illustrated by the figure to the left.

APPENDIX B

PROBABILITY DENSITY FUNCTIONS

❝While writing my book I had an argument with Feller. He asserted that everyone said "random variable" and I asserted that everyone said "chance variable." We obviously had to use the same name in our books, so we decided the issue by a stochastic procedure. That is, we tossed for it and he won.❞

Joseph Leonard Doob (1910–2004), pioneer of and key contributor to mathematical probability[1]

B.1 Random variables

The concept of the *random variable* is widely used in probability and random processes. Before discussing what a *random variable is*, note two things that a *random variable* is *not* (next remark).

Remark B.1. [2] As pointed out by others, the term "random variable" is a "misnomer":

REM
- A *random variable* is **not random**.
- A *random variable* is **not a variable**.

What is it then? It is a *function* (next definition). In particular, it is a function that maps from an underlying stochastic process into $\mathbb{R}$. Any "*randomness*" (whatever that means) it may *appear* to have comes from the stochastic process it is mapping *from*. But the function itself (the *random*

[1] quote: Snell (1997) page 307, Snell (2005) page 251
image: http://www.dartmouth.edu/~chance/Doob/conversation.html

[2] Miller (2006) page 130, Feldman and Valdez-Flores (2010) page 4 ⟨"The name "random variable" is actually a misnomer, since it is not random and not a variable.…the *random variable* simply maps each point (outcome) in the sample space to a number on the real line…Technically, the space into which the *random variable* maps the sample space may be more general than the real line…"⟩, Curry and Feldman (2010) page 4, Trivedi (2016) page 2.1 ⟨"The term "random variable" is actually a misnomer, since a *random variable* X is really a function whose domain is the sample space S, and whose range is the set of all real numbers, $\mathbb{R}$."⟩

variable itself) is very deterministic and well-defined. What gives it the appearance of being random is that the outcome ω of the experiment appears to be random to the observer. So the *random variable* $\mathsf{X}(\omega)$ is simply a function of an underlying mechanism that appears to be random.

Definition B.1. [3] *Let* $(\Omega, \mathbb{E}, \mathsf{P})$ *be a* PROBABILITY SPACE *(Definition A.2 page 173)*.

DEF: *A* ***random variable*** X *is any function in* $\mathbb{R}^{\Omega}$.

B.2 Probability distributions

The probability information about σ-algebra $\mathbb{E}$ in a *probability space* (Definition A.2 page 173) is completely specified by *measure* P. However, sometimes it is more convenient to express this same *measure* information in terms of the *probability density function* or the *cummulative distribution function* of the *probability space*.

Definition B.2. [4] *Let* X *be a* RANDOM VARIABLE *(Definition B.1 page 184) on a* PROBABILITY SPACE $(\Omega, \mathbb{E}, \mathsf{P})$.

DEF: X *has* ***cummulative distribution function*** ***(cdf)*** $\mathsf{c_x}(x)$ *if*

$$\mathsf{c_x}(x) \triangleq \mathsf{P}\{x \in \mathbb{E} | \mathsf{X} < x\}$$

X ***probability density function*** ***(pdf)*** $\mathsf{p_x}(x)$ *if*

$$\mathsf{p_x}(x) \triangleq \frac{\mathsf{d}}{\mathsf{dx}}\mathsf{c_x}(x) \triangleq \frac{\mathsf{d}}{\mathsf{dx}}\mathsf{P}\{x \in \mathbb{E} | \mathsf{X} < x\}$$

Remark B.2. Suppose X be a *random variable* on a *probability space* $(\Omega, \mathbb{E}, \mathsf{P})$. Note that

- Both X and $\mathbb{E}$ are *functions*.
- But X is a function that maps from Ω to $\mathbb{R}$,
- whereas P is a function that maps from $\mathbb{E}$ to $\mathbb{R}$.

Definition B.3. *Let* $(\Omega, \mathbb{E}, \mathsf{P})$ *be a* PROBABILITY SPACE *(Definition A.2 page 173) and* X *and* $\mathsf{Y} : \Omega \to \mathbb{R}$ *random variables. Then a* ***joint probability density function*** $\mathsf{p_x} : \mathbb{E} \times \Omega \to [0 : 1]$ *and a* ***joint cumulative distribution function*** $\mathsf{c_x} : \mathbb{E} \times \Omega \to [0 : 1]$ *are defined as*

DEF:

$$\begin{aligned} \mathsf{c_{XY}}(x, y) &\triangleq \mathsf{P}\{\mathsf{X} \le x | \mathsf{Y} \le y\} && \text{(JOINT CUMULATIVE DISTRIBUTION FUNCTION)} \\ \mathsf{p_{XY}}(x, y) &\triangleq \frac{\mathsf{d}}{\mathsf{dy}}\frac{\mathsf{d}}{\mathsf{dx}}\mathsf{c_{XY}}(x, y) && \text{(JOINT PROBABILITY DENSITY FUNCTION)} \end{aligned}$$

Definition B.4. *Let* $(\Omega, \mathbb{E}, \mathsf{P})$ *be a* PROBABILITY SPACE *(Definition A.2 page 173) and* X *a random variable. Then a* ***conditional probability density function*** $\mathsf{p_x} : \mathbb{E} \times \Omega \to [0 : 1]$ *and a* ***conditional cumulative distribution function*** $\mathsf{c_x} : \mathbb{E} \times \Omega \to [0 : 1]$ *are defined as*

DEF:

$$\begin{aligned} \mathsf{c_x}(x|y) &\triangleq \mathsf{P}\{\mathsf{X} \le x | \mathsf{Y} = y\} && \text{(CONDITIONAL CUMULATIVE DISTRIBUTION FUNCTION—CDF)} \\ \mathsf{p_x}(x|y) &\triangleq \frac{\mathsf{d}}{\mathsf{dx}}\mathsf{c_x}(x|y) && \text{(CONDITIONAL PROBABILITY DENSITY FUNCTION—PDF)} \end{aligned}$$

B.3 Properties

Definition B.2 (page 184) defines the pdf and cdf of a *probability space* $(\Omega, \mathbb{E}, \mathsf{P})$ in terms of *measure* P. Conversely, the probability *measure* $\mathsf{P}\{a \le \mathsf{X} < b\}$ of an event $\{a \le \mathsf{X} < b\}$ can be expressed in terms of either the pdf or cdf.

[3] Papoulis (1991) page 63
[4] von der Linden et al. (2014) page 93 ⟨Definitions 7.1, 7.2⟩

Proposition B.1. *Let* X *a* RANDOM VARIABLE *with* PDF $\mathsf{p_x}$ *and* CDF $\mathsf{c_x}$ *(Definition B.2 page 184) on the* PROBABILITY SPACE $(\Omega, \mathbb{E}, \mathsf{P})$ *(Definition A.2 page 173).*

PRP

$$\left\{\begin{array}{ll}(1). & \mathsf{c_x}(x) \text{ and } \mathsf{c_Y}(y) \text{ are CONTINUOUS} \quad \text{OR} \\ (2). & \mathsf{p_x}(x) \text{ and } \mathsf{p_Y}(y) \text{ are CONTINUOUS}\end{array}\right\} \implies \left\{\begin{array}{rcl}\mathsf{p_x}(x) &=& \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\mathsf{P}\{x \le \mathsf{X} < x+\varepsilon\} \\ \mathsf{p_{XY}}(x,y) &=& \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\mathsf{P}\{x \le \mathsf{X} < x+\varepsilon \wedge y \le \mathsf{Y} < y+\varepsilon\}\end{array}\right\}$$

✎PROOF:

$$\begin{aligned}
\mathsf{p_x}(x) &\triangleq \frac{\mathsf{d}}{\mathsf{dx}}\mathsf{c_x}(x) && \text{by definition of } \mathsf{p_x} && \text{(Definition B.2 page 184)}\\
&= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\mathsf{P}\{x\in\mathbb{R}|x \le \mathsf{X} < x+\varepsilon\} && \text{by definition of } \frac{\mathsf{d}}{\mathsf{dx}} && \text{(Definition H.3 page 223)}
\end{aligned}$$

✏

Theorem B.1. *Let* $(\Omega, \mathbb{E}, \mathsf{P})$ *be a probability space,* X *be a random variable, and* (a,b) *a real interval.*

THM

$$\left\{\begin{array}{ll}(1). & \mathsf{c_x}(x) \text{ is CONTINUOUS} \quad \text{OR} \\ (2). & \mathsf{p_x}(x) \text{ is CONTINUOUS}\end{array}\right\} \implies \left\{\mathsf{P}\{a < \mathsf{X} \le b\} = \mathsf{c_x}(b) - \mathsf{c_x}(a) = \int_a^b \mathsf{p_x}(x)\,\mathsf{dx}\right\}$$

✎PROOF:

$$\begin{aligned}
\mathsf{P}\{a < \mathsf{X} \le b\} &= \mathsf{P}\{\mathsf{X} \le b\} - \mathsf{P}\{\mathsf{X} < a\} && \text{by \textit{sum of products}} && \text{(Theorem A.3 page 175)}\\
&= \mathsf{P}\{\mathsf{X} \le b\} - \mathsf{P}\{\mathsf{X} \le a\} && \text{by \textit{continuity} hypothesis}\\
&\triangleq \mathsf{c_x}(b) - \mathsf{c_x}(a) && \text{by definition of } \mathsf{c_x} && \text{(Definition B.2 page 184)}
\end{aligned}$$

$$\begin{aligned}
\int_a^b \mathsf{p_x}(x)\,\mathsf{dx} &\triangleq \int_a^b \left[\frac{\mathsf{d}}{\mathsf{dx}}\mathsf{c_x}(x)\right]\mathsf{dx} && \text{by definition of } \mathsf{p_x} && \text{(Definition B.2 page 184)}\\
&= \mathsf{c_x}(x)\big|_{x=b} - \mathsf{c_x}(x)\big|_{x=a} && \text{by \textit{Fundamental theorem of calculus}}\\
&= \mathsf{c_x}(b) - \mathsf{c_x}(a)
\end{aligned}$$

✏

Theorem B.2. *Let* $(\Omega, \mathbb{E}, \mathsf{P})$ *be a* PROBABILITY SPACE, X *be a* RANDOM VARIABLE, *and* $(a : b)$ *a* REAL INTERVAL.

THM

$$\mathsf{P}\{a \le \mathsf{X} < b\} = \int_a^b \mathsf{p_x}(x)\,\mathsf{dx} = \int_{-\infty}^b \mathsf{c_x}(x)\,\mathsf{dx} - \int_{-\infty}^a \mathsf{c_x}(x)\,\mathsf{dx}$$

The properties of the pdf follow closely the properties of *measure* P.

Theorem B.3. [5]

THM

$$\left\{\begin{array}{ll}(A). & \mathsf{c_x}(x) \text{ is CONTINUOUS} \quad \text{OR} \\ (B). & \mathsf{p_x}(x) \text{ is CONTINUOUS}\end{array}\right\} \implies \left\{\begin{array}{lrcl}(1). & \mathsf{p_{x|Y}}(x|y) &=& \dfrac{\mathsf{p_{XY}}(x,y)}{\mathsf{p_Y}(y)} \quad \text{and} \\ (2). & \mathsf{p_x}(x) &=& \displaystyle\int_{y\in\mathbb{R}} \mathsf{p_{XY}}(x,y)\,\mathsf{dy}\end{array}\right\}$$

[5] Papoulis (1990) page 158 ⟨Auxiliary Variable⟩, Jazwinski (1970) page 39 ⟨"(2.102)"⟩, Jazwinski (2007) page 39 ⟨"(2.102)"⟩

✎PROOF:

$$\begin{aligned}
\mathsf{p}_{\mathsf{X}|\mathsf{Y}}(x|y) &\triangleq \frac{\mathsf{d}}{\mathsf{d}x}\mathsf{c}_{\mathsf{X}|\mathsf{Y}}(x|y) && \text{by definition of } \mathsf{c}_{\mathsf{X}} && \text{(Definition A.4 page 174)}\\
&\triangleq \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\mathsf{P}\{x \le \mathsf{X} < x+\varepsilon | \mathsf{Y}=y\} && \text{by definition of } \frac{\mathsf{d}}{\mathsf{d}x} && \text{(Definition H.3 page 223)}\\
&\triangleq \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\frac{\mathsf{P}\{(x \le \mathsf{X} < x+\varepsilon)\land(\mathsf{Y}=y)\}}{\mathsf{P}\{\mathsf{Y}=y\}} && \text{by definition of } \mathsf{P}\{A|B\} && \text{(Definition A.4 page 174)}\\
&= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\frac{\mathsf{P}\{(x \le \mathsf{X} < x+\varepsilon)\land(y \le \mathsf{Y} < y+\varepsilon)\}}{\mathsf{P}\{y \le \mathsf{Y} < y+\varepsilon\}} && \text{by \textit{continuity} hypothesis}\\
&= \frac{\lim\limits_{\varepsilon\to 0}\frac{1}{\varepsilon}\mathsf{P}\{(x \le \mathsf{X} < x+\varepsilon)\land(y \le \mathsf{Y} < y+\varepsilon)\}}{\lim\limits_{\varepsilon\to 0}\mathsf{P}\{y \le \mathsf{Y} < y+\varepsilon\}} && \text{by property of } \lim_{\varepsilon\to 0}\\
&= \frac{\mathsf{p}_{\mathsf{XY}}(x,y)}{\mathsf{p}_{\mathsf{Y}}(y)} && \text{by Proposition B.1 page 185}
\end{aligned}$$

$$\begin{aligned}
\int_{y\in\mathbb{R}}\mathsf{p}_{\mathsf{XY}}(x,y)\,\mathsf{d}y &\triangleq \int_{y\in\mathbb{R}}\left[\frac{\mathsf{d}}{\mathsf{d}y}\frac{\mathsf{d}}{\mathsf{d}x}\mathsf{c}_{\mathsf{XY}}(x,y)\right]\mathsf{d}y && \text{by definition of } \mathsf{p}_{\mathsf{X}} && \text{(Definition B.2 page 184)}\\
&= \frac{\mathsf{d}}{\mathsf{d}x}\mathsf{c}_{\mathsf{XY}}(x,y)\\
&\triangleq \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_{y\in\mathbb{R}}\mathsf{P}\{x \le \mathsf{X} < x+\varepsilon, y \le \mathsf{Y} < y+\varepsilon\}\,\mathsf{d}y && \text{by definition of } \frac{\mathsf{d}}{\mathsf{d}x} && \text{(Definition H.3 page 223)}\\
&= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\mathsf{P}\{x \le \mathsf{X} < x+\varepsilon\}\\
&= \mathsf{p}_{\mathsf{X}}(x) && \text{by Proposition B.1 page 185}
\end{aligned}$$

✏

Theorem B.4.

THM

$$\begin{aligned}
\mathsf{c}_{\mathsf{X}}(\sup\mathbb{R}) &= 1\\
\mathsf{c}_{\mathsf{X}}(\inf\mathbb{R}) &= 0
\end{aligned}$$

✎PROOF:

$$\begin{aligned}
\mathsf{c}_{\mathsf{X}}(\sup\mathbb{R}) &\triangleq \mathsf{P}\left\{\mathsf{X} \le \sup\mathbb{R}\right\} && \text{by definition of } \mathsf{c}_{\mathsf{X}} && \text{(Definition B.2 page 184)}\\
&= 1\\
\mathsf{c}_{\mathsf{X}}(\inf\mathbb{R}) &\triangleq \mathsf{P}\left\{\mathsf{X} \le \inf\mathbb{R}\right\} && \text{by definition of } \mathsf{c}_{\mathsf{X}} && \text{(Definition B.2 page 184)}\\
&= 0
\end{aligned}$$

✏

The properties of the pdf follow closely the properties of measure P.

Theorem B.5.

THM

$$\mathsf{c}_{\mathsf{X}|\mathsf{Y}}(x|y) = \frac{\frac{\mathsf{d}}{\mathsf{d}y}\mathsf{c}_{\mathsf{XY}}(x,y)}{\mathsf{p}_{\mathsf{Y}}(y)} \qquad \mathsf{p}_{\mathsf{X}|\mathsf{Y}}(x|y) = \frac{\mathsf{p}_{\mathsf{XY}}(x,y)}{\mathsf{p}_{\mathsf{Y}}(y)}$$

✎PROOF:

$$
\begin{aligned}
\mathsf{c}_{\mathsf{X}|\mathsf{Y}}(x|y) &\triangleq \mathsf{P}\{\mathsf{X} \le x \mid \mathsf{Y} = y\} && \text{by definition of } \mathsf{c}_{\mathsf{X}|\mathsf{Y}} && \text{(Definition B.4 page 184)}\\
&\triangleq \frac{\mathsf{P}\{\mathsf{X} \le x|\mathsf{Y} = y\}}{\mathsf{P}\{\mathsf{Y} = y\}} && \text{by definition of } \mathsf{P}\{\mathsf{X}|\mathsf{Y}\} && \text{(Definition A.4 page 174)}\\
&= \lim_{\varepsilon\to 0} \frac{\mathsf{P}\{\mathsf{X} \le x|y < \mathsf{Y} \le y+\varepsilon\}}{\mathsf{P}\{y < \mathsf{Y} \le y+\varepsilon\}}\\
&= \lim_{\varepsilon\to 0} \frac{[\mathsf{P}\{\mathsf{X} \le x|\mathsf{Y} \le y+\varepsilon\} - \mathsf{P}\{\mathsf{X} \le x|\mathsf{Y} \le y\}]/\varepsilon}{[\mathsf{P}\{\mathsf{Y} \le y+\varepsilon\} - \mathsf{P}\{\mathsf{Y} \le y\}]/\varepsilon}\\
&\triangleq \lim_{\varepsilon\to 0} \frac{\left[\mathsf{c}_{\mathsf{XY}}(x, y+\varepsilon) - \mathsf{c}_{\mathsf{XY}}(x, y)\right]/\varepsilon}{\left[\mathsf{c}_{\mathsf{Y}}(y+\varepsilon) - \mathsf{c}_{\mathsf{Y}}(y)\right]/\varepsilon} && \text{by definition of } \mathsf{c}_{\mathsf{XY}} && \text{(Definition B.3 page 184)}\\
&\triangleq \frac{\frac{\mathsf{d}}{\mathsf{d}y}\mathsf{c}_{\mathsf{XY}}(x, y)}{\frac{\mathsf{d}}{\mathsf{d}y}\mathsf{c}_{\mathsf{Y}}(y)} && \text{by definition of } \frac{\mathsf{d}}{\mathsf{d}y}\mathsf{f}(y)\\
&\triangleq \frac{\frac{\mathsf{d}}{\mathsf{d}y}\mathsf{c}_{\mathsf{XY}}(x, y)}{\mathsf{p}_{\mathsf{Y}}(y)} && \text{by definition of } \mathsf{p}_{\mathsf{Y}} && \text{(Definition B.2 page 184)}\\
&= \frac{\frac{\mathsf{d}}{\mathsf{d}y}\mathsf{c}_{\mathsf{XY}}(x, y)}{\mathsf{p}_{\mathsf{Y}}(y)} && \text{because } y \text{ is fixed}
\end{aligned}
$$

$$
\begin{aligned}
\mathsf{p}_{\mathsf{X}|\mathsf{Y}}(x|y) &\triangleq \frac{\mathsf{d}}{\mathsf{d}x}\mathsf{c}_{\mathsf{X}|\mathsf{Y}}(x|y) && \text{by definition of } \mathsf{p}_{\mathsf{X}|\mathsf{Y}} && \text{(Definition B.4 page 184)}\\
&= \frac{\mathsf{d}}{\mathsf{d}x}\frac{\frac{\mathsf{d}}{\mathsf{d}y}\mathsf{c}_{\mathsf{XY}}(x, y)}{\mathsf{p}_{\mathsf{Y}}(y)} && \text{by previous result}\\
&= \frac{\frac{\mathsf{d}}{\mathsf{d}x}\frac{\mathsf{d}}{\mathsf{d}y}\mathsf{c}_{\mathsf{XY}}(x, y)}{\mathsf{p}_{\mathsf{Y}}(y)} && \text{because } \mathsf{p}_{\mathsf{Y}}(y) \text{ is not a function of } x\\
&\triangleq \frac{\mathsf{p}_{\mathsf{XY}}(x, y)}{\mathsf{p}_{\mathsf{Y}}(y)} && \text{by definition of } \mathsf{p}_{\mathsf{XY}}(x, y) && \text{(Definition B.3 page 184)}
\end{aligned}
$$

✏

Theorem B.6. *Let* $(\Omega, \mathbb{E}, \mathsf{P})$ *be a probability space.*

THM

$$
\begin{aligned}
\int_{x\in\mathbb{R}} \mathsf{p}_{\mathsf{X}}(x)\,\mathsf{d}x &= 1 \\
\int_{y\in\mathbb{R}} \mathsf{p}_{\mathsf{XY}}(x, y)\,\mathsf{d}y &= \mathsf{p}_{\mathsf{X}}(x) \quad \forall x\in\Omega
\end{aligned}
\qquad
\begin{aligned}
\int_{x\in\mathbb{R}} \mathsf{p}_{\mathsf{X}|\mathsf{Y}}(x|y)\,\mathsf{d}x &= 1\\
\int_{x\in\mathbb{R}}\int_{y\in\mathbb{R}} \mathsf{p}_{\mathsf{XY}}(x, y)\,\mathsf{d}y\,\mathsf{d}x &= 1
\end{aligned}
$$

✎PROOF:

$$
\begin{aligned}
\int_{\mathbb{R}} \mathsf{p}_{\mathsf{X}}(x)\,\mathsf{d}x &= \mathsf{c}_{\mathsf{X}}(\sup\mathbb{R}) - \mathsf{c}_{\mathsf{X}}(\inf\mathbb{R}) && \text{by Theorem B.1 page 185}\\
&= 1 - 0\\
&= 1 && \text{because 0 is the additive identity element in } (\mathbb{R}, +, \cdot, 0, 1)\\
\int_{x\in\mathbb{R}} \mathsf{p}_{\mathsf{X}|\mathsf{Y}}(x|y)\,\mathsf{d}x &\triangleq \int_{x\in\mathbb{R}} \frac{\mathsf{d}}{\mathsf{d}x}\mathsf{c}_{\mathsf{X}|\mathsf{Y}}(x|y)\,\mathsf{d}x && \text{by definition of } \mathsf{p}_{\mathsf{X}|\mathsf{Y}}(x|y) \text{ (Definition B.4 page 184)}\\
&= \mathsf{c}_{\mathsf{X}|\mathsf{Y}}(\sup\mathbb{R}|y) - \mathsf{c}_{\mathsf{X}|\mathsf{Y}}(\inf\mathbb{R}|y) && \text{by \textit{Fundamental theorem of calculus}}
\end{aligned}
$$

$$\begin{aligned}
&= 1 - 0 \\
&= 1 && \text{because 0 is the additive identity element in } (\mathbb{R}, +, \cdot, 0, 1) \\
\int_{y\in\mathbb{R}} \mathsf{p}_{\mathsf{XY}}(x,y)\,\mathsf{d}y &= \int_{y\in\mathbb{R}} \mathsf{p}_{\mathsf{YX}}(y,x)\,\mathsf{d}y \\
&= \int_{y\in\mathbb{R}} \mathsf{p}_{\mathsf{Y|X}}(y|x)\mathsf{p}_{\mathsf{X}}(x)\,\mathsf{d}y && \text{by Theorem B.5 page 186} \\
&= \mathsf{p}_{\mathsf{X}}(x)\int_{y\in\mathbb{R}} \mathsf{p}_{\mathsf{Y|X}}(y|x)\,\mathsf{d}y && \text{because } \mathsf{p}_{\mathsf{X}}(x) \text{ is not a function of } y \\
&= \mathsf{p}_{\mathsf{X}}(x)\cdot 1 && \text{by previous result} \\
&= \mathsf{p}_{\mathsf{X}}(x) && \text{because 1 is the multiplicative identity element in } (\mathbb{R}, +, \cdot, 0, 1)
\end{aligned}$$

$$\begin{aligned}
\int_{x\in\mathbb{R}}\int_{y\in\mathbb{R}} \mathsf{p}_{\mathsf{XY}}(x,y)\,\mathsf{d}y\,\mathsf{d}x &= \int_{x\in\mathbb{R}} \mathsf{p}_{\mathsf{X}}(x)\,\mathsf{d}x && \text{by previous result} \\
&= 1 && \text{by previous result}
\end{aligned}$$

✎

APPENDIX C

SOME PROBABILITY DENSITY FUNCTIONS

C.1 Discrete distributions

Example C.1. [1] Suppose we throw two “fair” dice and want to know the probabilities of their sum. Let X represent the sum of the face values of the two dice. The resulting probability distribution is illustrated in Figure C.1 (page 190) and has probability space as follows:

EX

$$\begin{aligned} \Omega &= \{⚀⚀,\ ⚀⚁,\ ⚀⚂,\ \ldots,\ ⚅⚅\} \\ \mathbb{E} &= \left\{ 2^{\{X=n \mid n=2,3,\ldots,10,11, \text{ or } 12\}} \right\} \\ \mathsf{P}(e) &= \frac{1}{36} |\, e \,| \end{aligned}$$

C.2 Continuous distributions

C.2.1 Uniform distribution

Definition C.1. *The **uniform distribution** $\mathsf{p}_{\mathsf{x}}(x)$ is defined as*

DEF

$$\mathsf{p}_{\mathsf{x}}(x) \triangleq \left\{ \begin{array}{ll} 1 & \textit{for } 0 < x \le 1 \\ 0 & \textit{otherwise} \end{array} \right\}$$

Note that although “simple” in form, in light of *Wold's Theorem*,[2] the value of the *uniform distribution* should *not* be taken lightly.

[1] Osgood (2002)

[2] Wold (1938), Wold (1954), Scargle (1979)

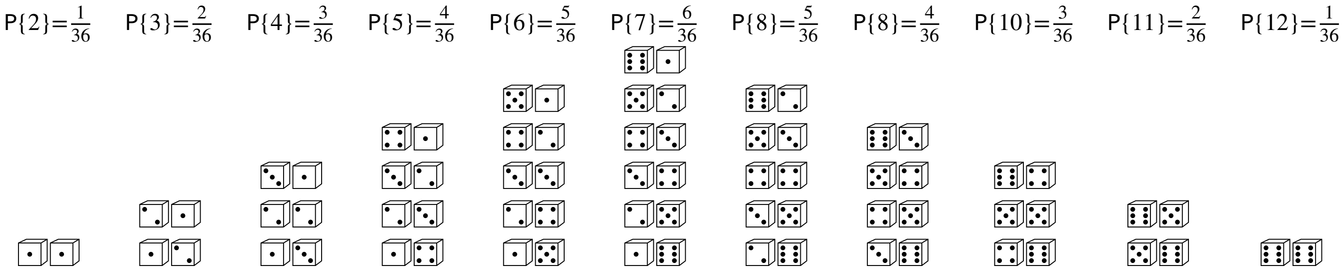
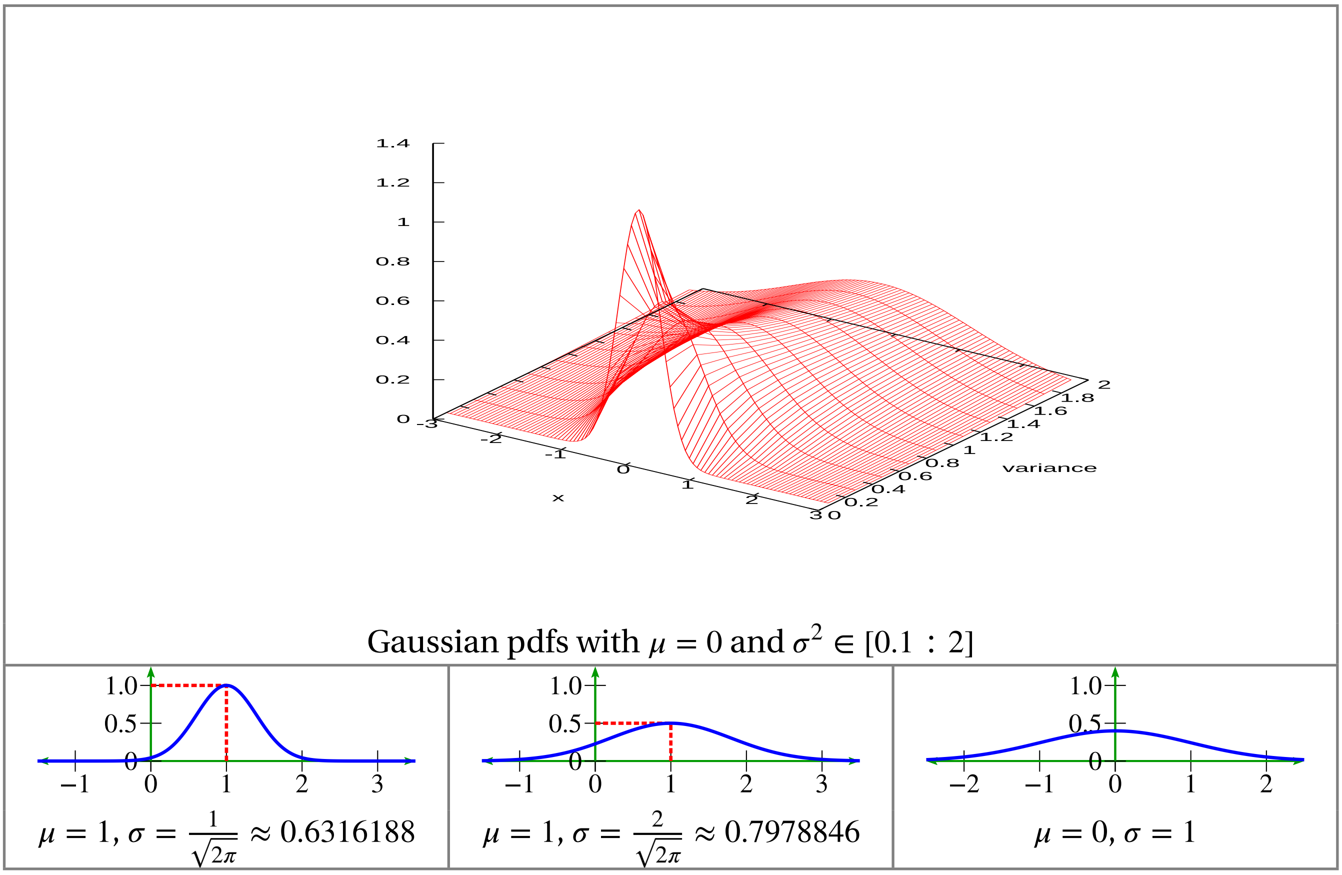


Figure C.1: Probability distribution for two dice (see Example C.1 page 189)



Gaussian pdfs with $\mu = 0$ and $\sigma^2 \in [0.1 : 2]$

$\mu = 1, \sigma = \frac{1}{\sqrt{2\pi}} \approx 0.6316188$

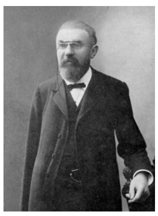
$\mu = 1, \sigma = \frac{2}{\sqrt{2\pi}} \approx 0.7978846$

$\mu = 0, \sigma = 1$

Figure C.2: Gaussian pdfs

C.2.2 Gaussian distribution

❝*Tout le monde y croit cependant, me disait un jour M. Lippmann, car les expérimentateurs s'imaginent que c'est un théorème de mathématiques, et les mathématiciens que c'est un fait expérimental.*❞



❝*Everyone believes in it [(the normal distribution)] however, said to me one day Mr. Lippmann, because the experimenters imagine that it is a theorem of mathematics, and mathematicians that it is an experimental fact.*❞

Bernard A. Lippmann as told by Henri Poincaré [3]

Definition C.2.

[3] quote: Poincaré (1912) page 171
translation: assisted by Google Translate
image:

DEF

*The **Gaussian distribution** (or **normal distribution**) has pdf*

$$\mathsf{p}_{\mathsf{x}}(x) \triangleq \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

A random variable X *with this distribution is denoted*

$$\mathsf{X} \sim \mathsf{N}\left(\mu, \sigma^2\right)$$

The function Q(x) *is defined as the area under a Gaussian PDF with zero mean and variance equal to one from x to infinity such that*

$$\mathsf{Q}(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{\frac{-u^2}{2}} \,\mathsf{d}u$$

C.2.3 Gamma distribution

Definition C.3. [4] *Let $b \in \mathbb{R}$. The **gamma function** $\Gamma(b)$ is*

DEF

$$\Gamma(b) \triangleq \int_0^\infty x^{b-1} e^{-x} \,\mathsf{d}x$$

Proposition C.1. [5] *Let $b \in \mathbb{R}$ and $n \in \mathbb{N}$.*

PRP

$$\begin{aligned}
\Gamma(b) &= (b-1)\Gamma(b-1) \\
\Gamma(n) &= (n-1)!
\end{aligned}$$

✎PROOF: Let

$$\begin{aligned}
u &= x^{b-1} & \mathsf{d}u &= (b-1)x^{b-1}\,\mathsf{d}x \\
\mathsf{d}v &= e^{-x}\,\mathsf{d}x & v &= -e^{-x}
\end{aligned}$$

$$\begin{aligned}
\Gamma(b) &\triangleq \int_0^\infty x^{b-1} e^{-x} \,\mathsf{d}x \\
&= \int_{x=0}^\infty u \,\mathsf{d}v \\
&= uv|_{x=0}^\infty - \int_{x=0}^\infty v \,\mathsf{d}u \\
&= -x^{b-1}e^{-x}|_{x=0}^\infty + (b-1)\int_{x=0}^\infty e^{-x}x^{b-1}\,\mathsf{d}x \\
&= (-0+0) + (b-1)\Gamma(b-1)
\end{aligned}$$

Note that

$$\Gamma(1) = \int_0^\infty x^{1-1}e^{-x}\,\mathsf{d}x = \int_0^\infty e^{-x}\,\mathsf{d}x = -e^{-x}|_0^\infty = -0+1 = 1$$

$$\begin{aligned}
\Gamma(n) &= (n-1)\Gamma(n-1) \\
&= (n-1)(n-2)\Gamma(n-2) \\
&= (n-1)(n-2)(n-3)\Gamma(n-3) \\
&\vdots \\
&= (n-1)(n-2)(n-3)\cdots(1)\Gamma(1) \\
&= (n-1)(n-2)(n-3)\cdots(1) \\
&\triangleq (n-1)!
\end{aligned}$$

[4] Papoulis (1991) page 79, Ross (1998) page 222
[5] Ross (1998) page 223

Definition C.4. *A **Gamma distribution** (b, λ) has pdf*

DEF
$$\mathsf{p}_\mathsf{x}(x) \triangleq \frac{\lambda}{\Gamma(b)} e^{-\lambda x}(\lambda x)^{b-1}$$

Theorem C.1. [6] *Let* X *and* Y *be* RANDOM VARIABLES *on a* PROBABILITY SPACE $(\Omega, \mathbb{E}, \mathsf{P})$.

THM
$$\left\{\begin{array}{lll} \text{(A).} & \mathsf{X} \text{ and } \mathsf{Y} \text{ are INDEPENDENT} & \text{and} \\ \text{(B).} & \mathsf{X} \text{ has GAMMA DISTRIBUTION } (a, \lambda) & \text{and} \\ \text{(C).} & \mathsf{Y} \text{ has GAMMA DISTRIBUTION } (b, \lambda) & \text{and} \\ \text{(D).} & \mathsf{Z} \triangleq \mathsf{X} + \mathsf{Y} & \end{array}\right\} \implies \left\{\begin{array}{l} \mathsf{Z} \text{ has Gamma distribution} \\ (a+b, \lambda). \end{array}\right\}$$

✎PROOF:

$$\begin{aligned}
\mathsf{p}_\mathsf{z}(z) &= \mathsf{p}_\mathsf{x}(z) \star \mathsf{p}_\mathsf{y}(z) \\
&= \int_{u\in\mathbb{R}} \mathsf{p}_\mathsf{x}(u)\mathsf{p}_\mathsf{y}(z-u)\, \mathsf{d}u && \text{by definition of } \textit{convolution} \quad \text{(Definition D.1 page 199)} \\
&= \int_0^z \frac{1}{\Gamma(a)}\lambda e^{-\lambda u}(\lambda u)^{a-1} \frac{1}{\Gamma(b)}\lambda e^{-\lambda(z-u)}(\lambda(z-u))^{b-1}\, \mathsf{d}u \\
&= \frac{1}{\Gamma(a)}\frac{1}{\Gamma(b)} e^{-\lambda z}\lambda^{1+1+a-1+b-1} \int_0^z u^{a-1}(z-u)^{b-1}\, \mathsf{d}u \\
&= \frac{1}{\Gamma(a)}\frac{1}{\Gamma(b)} \lambda e^{-\lambda z}\lambda^{a+b-1} \int_0^1 (vz)^{a-1}(z-vz)^{b-1} z\, \mathsf{d}v \\
&= \frac{1}{\Gamma(a)}\frac{1}{\Gamma(b)} \lambda e^{-\lambda z}\lambda^{a+b-1} z^{a-1+b-1+1} \int_0^1 v^{a-1}(1-v)^{b-1}\, \mathsf{d}v \\
&= \left[\frac{1}{\Gamma(a)}\frac{1}{\Gamma(b)} \int_0^1 v^{a-1}(1-v)^{b-1}\, \mathsf{d}v\right] \lambda e^{-\lambda z}(\lambda z)^{a+b-1} \\
&= C\lambda e^{-\lambda z}(\lambda z)^{a+b-1} && \text{where } C \text{ is some constant} \\
&= \frac{\lambda}{\Gamma(a+b)} e^{-\lambda z}(\lambda z)^{a+b-1} && C \text{ must be the value that makes } \int_z \mathsf{p}_\mathsf{z}(z) = 1
\end{aligned}$$

$\implies \mathsf{p}_\mathsf{z}(z)$ is a $(a+b, \lambda)$ Gamma distribution

C.2.4 Chi-squared distributions

Definition C.5. [7] *Let* $\mathsf{p}(x)$ *be a* PROBABILITY DENSITY FUNCTION *on a* PROBABILITY SPACE $(\Omega, \mathbb{E}, \mathsf{P})$.

DEF
$\mathsf{p}(x)$ *is a **chi-square distribution** if*

$$\mathsf{p}(x) \triangleq \left\{\begin{array}{ll} 0 & \text{if } x < 0 \\ \dfrac{1}{\sqrt{2\pi\sigma^2 x}} \exp\left[-\dfrac{x}{2\sigma^2}\right] & \text{if } x \geq 0 \end{array}\right\} \qquad \text{for } \sigma > 0$$

Theorem C.2. [8]

[6] Ross (1998) page 266
[7] Proakis (2001) page 41, Papoulis (1990) page 219 ⟨7-4 Special Distributions of Statistics, (7-78)⟩
[8] Ross (1998) page 267

THM *The following distributions are equivalent:*
(1). *chi-squared distribution* *and*
(2). *distribution of* X^2 *where* $\mathsf{X} \sim \mathsf{N}\left(0, \sigma^2\right)$ *and*
(3). *Gamma distribution* $\left(\frac{1}{2}, \frac{1}{2\sigma^2}\right)$

✎PROOF:

1. Proof that X^2 has chi-squared distribution:

$$\begin{aligned}
\mathsf{p}_{\mathsf{Y}}(y) &= \frac{1}{2\sqrt{y}}\left[\mathsf{p}_{\mathsf{X}}(-\sqrt{y}) + \mathsf{p}_{\mathsf{X}}(\sqrt{y})\right] && \text{by Corollary 6.3 page 44}\\
&= \frac{1}{2\sqrt{y}}\left[\frac{1}{\sqrt{2\pi\sigma^2}}\exp - \frac{(-\sqrt{y}-0)^2}{2\sigma^2} + \frac{1}{\sqrt{2\pi\sigma^2}}\exp - \frac{(+\sqrt{y}-0)^2}{2\sigma^2}\right]\\
&= \frac{1}{2\sqrt{y}}\left[2\frac{1}{\sqrt{2\pi\sigma^2}}\exp - \frac{y}{2\sigma^2} +\right]\\
&= \frac{1}{\sqrt{2\pi\sigma^2 y}}\exp - \frac{y}{2\sigma^2}
\end{aligned}$$

2. Proof that chi-distribution is a Gamma distribution (b, λ):

$$\begin{aligned}
b &\triangleq && \frac{1}{2}\\
\lambda &\triangleq \frac{1}{2\sigma^2}
\end{aligned}$$

$$\begin{aligned}
\frac{1}{\sqrt{2\pi\sigma^2 y}}\exp - \frac{y}{2\sigma^2} &= \frac{1}{\sqrt{\pi}}\lambda^{1/2}\lambda^{1/2}(\lambda y)^{-1/2}e^{-\lambda y}\\
&= \frac{\lambda}{\sqrt{\pi}}(\lambda y)^{b-1}e^{-\lambda y}
\end{aligned}$$

✏

Definition C.6. [9] *The* ***Chi-squared distribution with*** n ***degrees of freedom*** *has pdf*

DEF
$$\mathsf{p}_{\mathsf{Y}}(y) \triangleq \begin{cases} 0 & : y < 0\\ \dfrac{1}{2\sigma^2\Gamma(n/2)}\left(\dfrac{y}{2\sigma^2}\right)^{\frac{n}{2}-1}\exp - \dfrac{y}{2\sigma^2} & : y \geq 0\end{cases}$$

Theorem C.3. [10] *The following distributions are equivalent:*

1. *chi-squared distribution with* n *degrees of freedom*
2. *the distribution of* $\sum_{k=1}^{n} X_k^2$ *where* $\left\{X_k | X_k \sim \mathsf{N}\left(0, \sigma^2\right), k = 1, 2, \ldots, n\right\}$ *are independent random variables.*
3. *Gamma distribution* $\left(\frac{n}{2}, \frac{1}{2\sigma^2}\right)$.

[9] Proakis (2001) page 41
[10] Ross (1998) page 267

✎PROOF:

1. Prove chi-squared distribution with n degrees of freedom is the Gamma distribution $\left(\frac{n}{2}, \frac{1}{2\sigma^2}\right)$:

$$\begin{aligned}
\lambda &\triangleq \frac{1}{2\sigma^2}\\
b &\triangleq \frac{1}{2}\\
\frac{1}{2\sigma^2\Gamma(n/2)}\left(\frac{y}{2\sigma^2}\right)^{\frac{n}{2}-1}\exp-\frac{y}{2\sigma^2} &= \frac{\lambda}{\Gamma(nb)}(\lambda y)^{nb-1}\exp-\lambda y
\end{aligned}$$

2. Prove $\sum_{k=1}^{n} X^2$ is Gamma $\left(\frac{n}{2}, \frac{1}{2\sigma^2}\right)$:

   (a) By Theorem C.2, X_k has Gamma distribution $\left(\frac{1}{2}, \frac{1}{2\sigma^2}\right)$.

   (b) By Theorem C.1, $\sum_{k=1}^{n} X_k^2$ has distribution $\left(\frac{n}{2}, \frac{1}{2\sigma^2}\right)$.

✏

Definition C.7. [11] *A* ***noncentral chi-square distribution*** (μ, σ^2) *has pdf*

DEF
$$\mathsf{p}_{\mathsf{Y}}(y) = \frac{1}{\sqrt{2\pi\sigma^2 y}}\exp\frac{y+\mu^2}{-2\sigma^2}\cosh\frac{\mu\sqrt{y}}{\sigma^2}$$

Theorem C.4.

THM
The following distributions are equivalent:

(1). NON-CENTRAL CHI-SQUARED DISTRIBUTION (μ, σ^2)

(2). *distribution of* X^2 *where* $\mathsf{X} \sim \mathsf{N}(\mu, \sigma^2)$

✎PROOF:

1. Proof that $\mathsf{Y} = X^2$ has a non-central chi-squared distribution:

$$\begin{aligned}
\mathsf{p}_{\mathsf{Y}}(y) &= \frac{1}{2\sqrt{y}}\left[\mathsf{p}_{\mathsf{X}}(-\sqrt{y}) + \mathsf{p}_{\mathsf{X}}(\sqrt{y})\right] && \text{by Corollary 6.3 page 44}\\
&= \frac{1}{2\sqrt{y}}\left[\frac{1}{\sqrt{2\pi\sigma^2}}\exp-\frac{(-\sqrt{y}-\mu)^2}{2\sigma^2} + \frac{1}{\sqrt{2\pi\sigma^2}}\exp-\frac{(+\sqrt{y}-\mu)^2}{2\sigma^2}\right]\\
&= \frac{1}{2\sqrt{y}}\left[\frac{1}{\sqrt{2\pi\sigma^2}}\exp-\frac{y+\mu^2}{2\sigma^2}\exp\frac{-2\mu\sqrt{y}}{2\sigma^2} + \frac{1}{\sqrt{2\pi\sigma^2}}\exp-\frac{y+\mu^2}{2\sigma^2}\exp\frac{2\mu\sqrt{y}}{2\sigma^2}\right]\\
&= \frac{1}{\sqrt{2\pi\sigma^2 y}}\exp-\frac{y+\mu^2}{2\sigma^2}\frac{1}{2}\left[\exp\frac{2\mu\sqrt{y}}{2\sigma^2} + \exp\frac{-2\mu\sqrt{y}}{2\sigma^2}\right]\\
&= \frac{1}{\sqrt{2\pi\sigma^2 y}}\exp\frac{y+\mu^2}{-2\sigma^2}\cosh\frac{\mu\sqrt{y}}{\sigma^2}
\end{aligned}$$

✏

[11] Proakis (2001) page 42

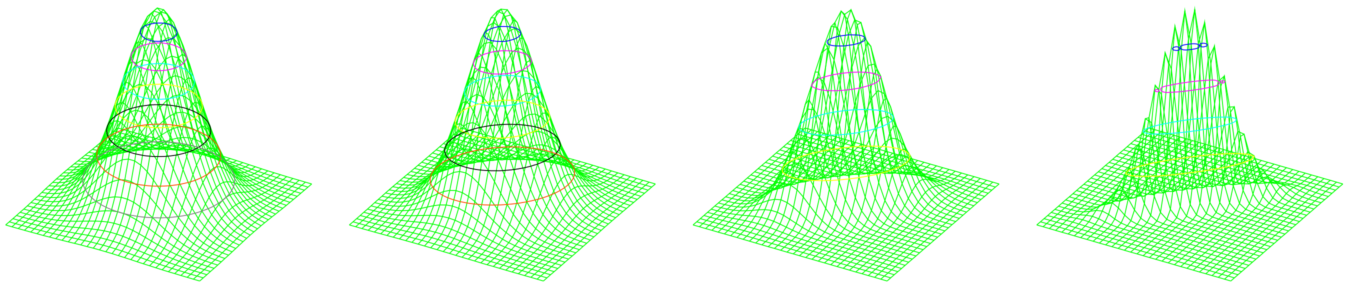


Figure C.3: *Joint Gaussian distributions* $\mathsf{p}_{\mathsf{XY}}(x,y)$ with varying correlations

Definition C.8. [12] *The* α ***th-order modified Bessel function of the first kind*** $I_\alpha(x)$ *is*

DEF
$$I_\alpha(x) = \sum_{k=0}^{\infty} \frac{1}{k!\Gamma(\alpha+k+1)}\left(\frac{x}{2}\right)^{\alpha+2k}$$

Definition C.9. [13] *The* ***noncentral chi-square with*** n***-degrees of freedom*** *distribution has pdf*

DEF
$$\mathsf{p}_{\mathsf{Y}}(y) = \frac{1}{2\sigma^2}\left(\frac{y}{s^2}\right)^{\frac{n-2}{4}} \exp\frac{y+s^2}{-2\sigma^2} I_{n/2-1}\left(\sqrt{y}\frac{s}{\sigma^2}\right) \qquad \text{where } s^2 \triangleq \sum_{k=1}^{n}\mu_k^2$$

C.2.5 Radial distributions

Definition C.10. [14] *The* ***Rayleigh distribution*** *is the pdf*

DEF
$$\mathsf{p}_{\mathsf{R}}(r) = \begin{cases} 0 & \text{for} \quad r<0 \\ \frac{r}{\sigma^2}\exp-\frac{r^2}{2\sigma^2} & \text{for} \quad r\ge 0\end{cases}$$

Note that by Proposition 6.3, this distribution is equivalent to the distribution of $R = \sqrt{X^2+Y^2}$ where X and Y are independent random variables each with distribution $\mathsf{N}(0,\sigma^2)$.

Definition C.11. [15] *The* ***Rice distribution*** *is the pdf*

DEF
$$\mathsf{p}_{\mathsf{R}}(r) = \begin{cases} 0 & \text{for} \quad r<0 \\ \frac{r}{\sigma^2}\exp\frac{r^2+s^2}{-2\sigma^2} I_o\left(\frac{rs}{\sigma^2}\right) & \text{for} \quad r\ge 0\end{cases}$$

C.3 Joint Gaussian distributions

Definition C.12 (Joint Gaussian pdf). [16]

DEF

$$\mathsf{p}(x_1, x_2, \ldots, x_n) \triangleq \frac{1}{\sqrt{(2\pi)^n |\boldsymbol{M}|}} \exp -\frac{1}{2}(\boldsymbol{x} - \mathbf{E}\boldsymbol{x})^T \boldsymbol{M}^{-1}(\boldsymbol{x} - \mathbf{E}\boldsymbol{x}) \qquad \text{(Gaussian joint pdf)}$$

$$\boldsymbol{x} \triangleq \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$\mathsf{Z}_k \triangleq X_k - \mathbf{E}X_k \qquad \text{(zero mean random variables)}$$

$$\boldsymbol{M} \triangleq \begin{bmatrix} \mathbf{E}[\mathsf{Z}_1\mathsf{Z}_1] & \mathbf{E}[\mathsf{Z}_1\mathsf{Z}_2] & \cdots & \mathbf{E}[\mathsf{Z}_1\mathsf{Z}_n] \\ \mathbf{E}[\mathsf{Z}_2\mathsf{Z}_1] & \mathbf{E}[\mathsf{Z}_2\mathsf{Z}_2] & & \mathbf{E}[\mathsf{Z}_2\mathsf{Z}_n] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{E}[\mathsf{Z}_n\mathsf{Z}_1] & \mathbf{E}[\mathsf{Z}_n\mathsf{Z}_2] & \cdots & \mathbf{E}[\mathsf{Z}_n\mathsf{Z}_n] \end{bmatrix} \qquad \text{(correlation matrix)}$$

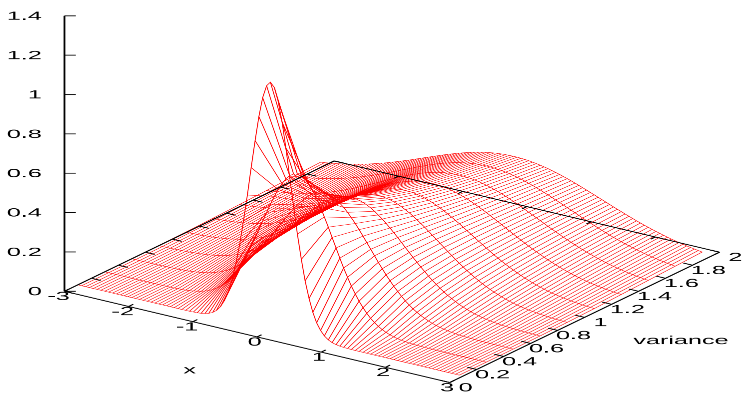


Figure C.4: Gaussian pdf with $\mu = 0$ and $\sigma \in [0.1, 2]$.

Example C.2 (1 variable joint Gaussian pdf). The **Gaussian distribution** (or **normal distribution** has pdf

EX

$$\mathsf{p}_\mathsf{x}(x) \triangleq \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

---

[12] Proakis (2001) page 43

[13] Proakis (2001) page 43

[14] Proakis (2001) page 44

[15] Proakis (2001) page 46

[16] Anderson (1984) page 21 ⟨Theorem 2.3.1⟩, Anderson (1958) page 14 ⟨§"2.3 The Multivariate Normal Distribution"⟩, Proakis (2001) page 49, Moon and Stirling (2000) page 34

$$
\begin{aligned}
t &= \arg_t \min_t \left[\frac{1}{2} \int_t^\infty \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} + \frac{1}{2} \int_{-\infty}^t \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\eta)^2}{2\sigma^2}} \right] \\
&= \arg_t \left\{ \frac{\partial}{\partial t} \left[\frac{1}{2} \int_t^\infty \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} + \frac{1}{2} \int_{-\infty}^t \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\eta)^2}{2\sigma^2}} \right] = 0 \right\} \\
&= \arg_t \left\{ \frac{1}{2\sqrt{2\pi\sigma^2}} \left[\frac{\partial}{\partial t} \int_t^\infty e^{\frac{-(x-\mu)^2}{2\sigma^2}} + \frac{\partial}{\partial t} \int_{-\infty}^t e^{\frac{-(x-\eta)^2}{2\sigma^2}} \right] = 0 \right\} \\
&= \arg_t \left\{ \left[\left(e^{\frac{-(\infty-\mu)^2}{2\sigma^2}} 0 - e^{\frac{-(t-\mu)^2}{2\sigma^2}} 1 \right) + \left(e^{\frac{-(t-\mu)^2}{2\sigma^2}} 1 - e^{\frac{-(\infty-\mu)^2}{2\sigma^2}} 0 \right) \right] = 0 \right\} \\
&= \arg_t \left\{ \left[e^{\frac{-(t-\eta)^2}{2\sigma^2}} - e^{\frac{-(t-\mu)^2}{2\sigma^2}} \right] = 0 \right\} \\
&= \arg_t \left\{ (t-\eta)^2 = (t-\mu)^2 \right\} \\
&= \frac{\mu+\eta}{2}
\end{aligned}
$$

Example C.3 (2 variable joint Gaussian pdf).

EX

$$
\begin{aligned}
z_1 &\triangleq x_1 - \mathbf{E}x_1 \\
z_2 &\triangleq x_2 - \mathbf{E}x_2 \\
|M| &\triangleq \left| \mathbf{E}[z_1 z_1]\mathbf{E}[z_2 z_2] - \mathbf{E}[z_1 z_2]\mathbf{E}[z_1 z_2] \right| \\
\mathsf{p}(x_1, x_2) &\triangleq \frac{1}{2\pi\sqrt{|M|}} \exp\left(\frac{z_1^2 \mathbf{E}[z_2 z_2] - 2 z_1 z_2 \mathbf{E}[z_1 z_2] + z_2^2 \mathbf{E}[z_1 z_1]}{-2|M|} \right)
\end{aligned}
$$

APPENDIX D

CONVOLUTION

D.1 Convolution over continuous domains

D.1.1 Definition

Definition D.1. [1] *Let $\boldsymbol{L}^2_{(\mathbb{R},\mathscr{B},\mu)}$ be the space of all* LEBESGUE SQUARE-INTEGRABLE FUNCTIONS.

DEF *The **convolution** operation $\star$ is defined as*

$$\left[\mathsf{f}\star\mathsf{g}\right](x) \triangleq \mathsf{f}(x)\star\mathsf{g}(x) \triangleq \int_{u\in\mathbb{R}} \mathsf{f}(u)\mathsf{g}(x-u)\,\mathsf{d}u \qquad \forall \mathsf{f},\mathsf{g}\in\boldsymbol{L}^2_{(\mathbb{R},\mathscr{B},\mu)}$$

D.1.2 Properties

Theorem D.1. [2] *Let $\star$ be the* CONVOLUTION *operatation* (Definition D.1 page 199). *Let $\boldsymbol{L}^2_{(\mathbb{R},\mathscr{B},\mu)}$ be the space of all* LEBESGUE SQUARE-INTEGRABLE FUNCTIONS.

THM

$$\begin{array}{rcllr}
\mathsf{f}\star\mathsf{g} &=& \mathsf{g}\star\mathsf{f} & \forall \mathsf{f},\mathsf{g}\in\boldsymbol{L}^2_{(\mathbb{R},\mathscr{B},\mu)} & (\text{COMMUTATIVE})\\
\mathsf{f}\star(\mathsf{g}\star\mathsf{h}) &=& (\mathsf{f}\star\mathsf{g})\star\mathsf{h} & \forall \mathsf{f},\mathsf{g}\mathsf{h}\in\boldsymbol{L}^2_{(\mathbb{R},\mathscr{B},\mu)} & (\text{ASSOCIATIVE})\\
(\alpha\mathsf{f})\star\mathsf{g} &=& \alpha(\mathsf{f}\star\mathsf{g}) = \mathsf{f}\star(\alpha\mathsf{g}) & \forall \mathsf{f},\mathsf{g}\in\boldsymbol{L}^2_{(\mathbb{R},\mathscr{B},\mu)},\ \alpha\in\mathbb{C} & (\text{HOMOGENEOUS})\\
\mathsf{f}\star(\mathsf{g}+\mathsf{h}) &=& (\mathsf{f}\star\mathsf{g})+(\mathsf{f}\star\mathsf{h}) & \forall \mathsf{f},\mathsf{g},\mathsf{h}\in\boldsymbol{L}^2_{(\mathbb{R},\mathscr{B},\mu)} & (\text{DISTRIBUTIVE})
\end{array}$$

✎PROOF:

$$\begin{aligned}
\mathsf{f}\star\mathsf{g} &\triangleq \int_{u=-\infty}^{u=\infty} \mathsf{f}(u)\mathsf{g}(x-u)\,\mathsf{d}u && \text{by definition of } \star \quad \text{(Definition D.1 page 199)}\\
&\triangleq \int_{x-v=-\infty}^{x-v=\infty} \mathsf{f}(x-v)\mathsf{g}(v)(-1)\,\mathsf{d}v && \text{where } v \triangleq x-u \quad \Longrightarrow u = x-v,\ \mathsf{d}u = -\,\mathsf{d}v\\
&= -\int_{-v=-\infty}^{-v=\infty} \mathsf{f}(x-v)\mathsf{g}(v)\,\mathsf{d}v
\end{aligned}$$

[1] Bachman et al. (2002) page 268 ⟨Definition 5.2.1, but with $1/2\pi$ scaling factor⟩, Bachman (1964) page 6, Bracewell (1978) page 224 ⟨Table 11.1 Theorems for the Laplace Transform⟩

[2] Bachman et al. (2002) pages 268–270, Schatzman (2002) page 147 ⟨7.2.1 Convolution⟩

$$
\begin{aligned}
&= -\int_{v=\infty}^{v=-\infty} \mathsf{f}(x-v)\mathsf{g}(v)\,\mathrm{d}v \\
&= \int_{v=-\infty}^{v=\infty} \mathsf{g}(v)\mathsf{f}(x-v)\,\mathrm{d}v \\
&\triangleq \mathsf{g} \star \mathsf{f} && \text{by definition of } \star && \text{(Definition D.1 page 199)}
\end{aligned}
$$

$$
\begin{aligned}
\mathsf{f} \star (\mathsf{g} \star \mathsf{h}) &\triangleq \mathsf{f} \star \int_{u\in\mathbb{R}} \mathsf{g}(u)\mathsf{h}(x-u)\,\mathrm{d}u && \text{by definition of } \star && \text{(Definition D.1 page 199)} \\
&\triangleq \int_{v\in\mathbb{R}} \mathsf{f}(v) \int_{u\in\mathbb{R}} \mathsf{g}(u)\mathsf{h}(x-v-u)\,\mathrm{d}u\,\mathrm{d}v && \text{by definition of } \star && \text{(Definition D.1 page 199)} \\
&= \int_{v\in\mathbb{R}} \int_{(w-v)\in\mathbb{R}} \mathsf{f}(v)\mathsf{g}(w-v)\mathsf{h}(x-w)\,\mathrm{d}\omega\,\mathrm{d}v && \text{where } w \triangleq u+v && \implies u = w-v \\
&= \int_{w\in\mathbb{R}} \underbrace{\left[\int_{v\in\mathbb{R}} \mathsf{f}(v)\mathsf{g}(w-v)\,\mathrm{d}v\right]}_{\text{function of } w} \mathsf{h}(x-w)\,\mathrm{d}\omega \\
&\triangleq \left[\int_{v\in\mathbb{R}} \mathsf{f}(v)\mathsf{g}(w-v)\,\mathrm{d}v\right] \star \mathsf{h} && \text{by definition of } \star && \text{(Definition D.1 page 199)} \\
&\triangleq [\mathsf{f} \star \mathsf{g}] \star \mathsf{h} && \text{by definition of } \star && \text{(Definition D.1 page 199)}
\end{aligned}
$$

$$
\begin{aligned}
[\alpha\mathsf{f}] \star \mathsf{g} &\triangleq \int_{\mathbb{R}} [\alpha\mathsf{f}(u)]\mathsf{g}(x-u)\,\mathrm{d}u && \text{by definition of } \star && \text{(Definition D.1 page 199)} \\
&= \alpha \int_{\mathbb{R}} [\alpha\mathsf{f}(u)]\mathsf{g}(x-u)\,\mathrm{d}u && \text{by } \textit{homogeneous} \text{ property of } \int \text{ dt operator} \\
&= \boxed{\alpha[\mathsf{f} \star \mathsf{g}]} && \text{by definition of } \star && \text{(Definition D.1 page 199)} \\
&\triangleq \alpha\left[\int_{\mathbb{R}} \mathsf{f}(u)\mathsf{g}(x-u)\,\mathrm{d}u\right] && \text{by definition of } \star && \text{(Definition D.1 page 199)} \\
&\triangleq \int_{\mathbb{R}} \mathsf{f}(u)\left[\alpha\mathsf{g}(x-u)\right]\,\mathrm{d}u && \text{by } \textit{homogeneous} \text{ property of } \int \text{ dt operator} \\
&\triangleq \boxed{\mathsf{f} \star [\alpha\mathsf{g}]} && \text{by definition of } \star && \text{(Definition D.1 page 199)}
\end{aligned}
$$

$$
\begin{aligned}
\mathsf{f} \star (\mathsf{g}+\mathsf{h}) &\triangleq \int_{u\in\mathbb{R}} \mathsf{f}(u)\left[\mathsf{g}(x-u)+\mathsf{h}(x-u)\right]\,\mathrm{d}u && \text{by definition of } \star && \text{(Definition D.1 page 199)} \\
&= \int_{u\in\mathbb{R}} \mathsf{f}(u)\mathsf{g}(x-u)\,\mathrm{d}u + \int_{u\in\mathbb{R}} \mathsf{h}(u)\mathsf{g}(x-u)\,\mathrm{d}u && \text{by } \textit{additive} \text{ property of } \int \text{ dt operator} \\
&= (\mathsf{f} \star \mathsf{g}) + (\mathsf{f} \star \mathsf{h}) && \text{by definition of } \star && \text{(Definition D.1 page 199)}
\end{aligned}
$$

✎

D.2 Convolution over discrete domains

Definition D.2. [3] *Let X^Y be the set of all functions from a set Y to a set X. Let $\mathbb{Z}$ be the set of integers.*

DEF

*A function f in X^Y is a **sequence** over X if $Y = \mathbb{Z}$.*
A sequence may be denoted in the form $(\!(x_n)\!)_{n\in\mathbb{Z}}$ or simply as $(\!(x_n)\!)$.

[3] Bromwich (1908) page 1, Thomson et al. (2008) page 23 ⟨Definition 2.1⟩, Joshi (1997) page 31

Definition D.3. [4] *Let* $(\mathbb{F},+,\cdot,0,1)$ *be a* FIELD (Definition G.5 page 222).

DEF

*The **space of all absolutely square summable sequences** $\ell^2_{\mathbb{F}}$ over $\mathbb{F}$ is defined as*

$$\ell^2_{\mathbb{F}} \triangleq \left\{ \left(x_n\right)_{n\in\mathbb{Z}} \mid \sum_{n\in\mathbb{Z}} \left|x_n\right|^2 < \infty \right\}$$

The space $\ell^2_{\mathbb{R}}$ is an example of a *separable Hilbert space*. In fact, $\ell^2_{\mathbb{R}}$ is the *only* separable Hilbert space in the sense that all separable Hilbert spaces are isomorphically equivalent. For example, $\ell^2_{\mathbb{R}}$ is isomorphic to $\boldsymbol{L}^2_{\mathbb{R}}$, the *space of all absolutely square Lebesgue integrable functions*.

Definition D.4. *Let* $\left(x_n\right)_{n\in\mathbb{Z}}$ *and* $\left(y_n\right)_{n\in\mathbb{Z}}$ *be sequences* (Definition D.2 page 200) *in the space* $\ell^2_{\mathbb{R}}$ (Definition D.3 page 201).

DEF

*The **convolution** $\star$ of sequences (x_n) and (y_n) is defined as*

$$\left(\left(x_n\right)\star\left(y_n\right)\right)_{m\in\mathbb{Z}} \triangleq \left(\sum_{n\in\mathbb{Z}} x_n y_{m-n}\right)_{m\in\mathbb{Z}} \qquad \forall (x_n)_{n\in\mathbb{Z}},(y_n)_{n\in\mathbb{Z}}\in\ell^2_{\mathbb{R}}$$

Theorem D.2. *Let* $\star$ *be the discrete* CONVOLUTION *operatation* (Definition D.4 page 201).

THM

$$\begin{array}{rcll}
(x_n)\star(y_n) &=& (y_n)\star(x_n) & \forall (x_n),(y_n)\in\ell^2_{\mathbb{R}} \quad \text{(COMMUTATIVE)}\\
(x_n)\star\left[(y_n)\star(z_n)\right] &=& \left[(x_n)\star(y_n)\right]\star(z_n) & \forall (x_n),(y_n)(z_n)\in\ell^2_{\mathbb{R}} \quad \text{(ASSOCIATIVE)}\\
\left[\alpha\,(x_n)\right]\star(y_n) &=& \alpha\left[(x_n)\star(y_n)\right] = (x_n)\star\left[\alpha\,(y_n)\right] & \forall (x_n),(y_n)\in\ell^2_{\mathbb{R}},\alpha\in\mathbb{C} \quad \text{(HOMOGENEOUS)}\\
(x_n)\star\left[(y_n)+(z_n)\right] &=& \left[(x_n)\star(y_n)\right]+\left[(x_n)\star(z_n)\right] & \forall (x_n),(y_n),(z_n)\in\ell^2_{\mathbb{R}} \quad \text{(DISTRIBUTIVE)}
\end{array}$$

✎PROOF:

$$\begin{aligned}
\left(\left(x_n\right)\star\left(y_n\right)\right)_{m\in\mathbb{Z}} &\triangleq \left(\sum_{n=-\infty}^{n=\infty} (x_n)(y_{m-n})\right)_{m\in\mathbb{Z}} && \text{by definition of } \star \quad \text{(Definition D.4 page 201)}\\
&\triangleq \left(\sum_{m-k=-\infty}^{m-k=\infty} (x_{m-k})(y_k)\right)_{m\in\mathbb{Z}} && \text{where } k \triangleq m-n \quad \implies n = m-k\\
&= \left(\sum_{-k=-\infty}^{-k=\infty} (y_k)(x_{m-k})\right)_{m\in\mathbb{Z}}\\
&= \left(\sum_{k=-\infty}^{k=\infty} (y_k)(x_{m-k})\right)_{m\in\mathbb{Z}}\\
&\triangleq \left(\left(y_n\right)\star\left(x_n\right)\right)_{m\in\mathbb{Z}} && \text{by definition of } \star \quad \text{(Definition D.4 page 201)}
\end{aligned}$$

$$\begin{aligned}
&(x_n)\star\left[(y_n)\star(z_n)\right]\\
&\triangleq (x_n)\star\left(\sum_{n\in\mathbb{Z}}(y_n)(z_{m-n})\right)_{m\in\mathbb{Z}} && \text{by definition of } \star \quad \text{(Definition D.4 page 201)}\\
&\triangleq \left(\sum_{k\in\mathbb{Z}}(x_k)\sum_{n\in\mathbb{Z}}(y_n)(z_{m-k-n})\right)_{m\in\mathbb{Z}} && \text{by definition of } \star \quad \text{(Definition D.4 page 201)}\\
&= \left(\sum_{k\in\mathbb{Z}}(x_k)\sum_{p-k\in\mathbb{Z}}(y_{p-k})(z_{m-p})\right)_{m\in\mathbb{Z}} && \text{where } p \triangleq k+n \quad \implies n = p-k
\end{aligned}$$

[4] Kubrusly (2011) page 347 ⟨Example 5.K⟩

$$
\begin{aligned}
&= \left(\sum_{k\in\mathbb{Z}} \sum_{p-k\in\mathbb{Z}} (\!(x_k)\!) (\!(y_{p-k})\!) (\!(z_{m-p})\!) \right)_{m\in\mathbb{Z}} \\
&= \left(\sum_{p\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} (\!(x_k)\!) (\!(y_{p-k})\!) (\!(z_{m-p})\!) \right)_{m\in\mathbb{Z}} \\
&= \sum_{p\in\mathbb{Z}} \left((\!(x_k)\!) \star (\!(y_k)\!)\right)_p (\!(z_{m-p})\!) && \text{by definition of } \star && \text{(Definition D.4 page 201)} \\
&= \sum_{p\in\mathbb{Z}} \left((\!(x_n)\!) \star (\!(y_n)\!)\right)_p (\!(z_{m-p})\!) && \text{by change of variable} && k \to n \\
&= \left[(\!(x_n)\!) \star (\!(y_n)\!)\right] \star (\!(z_n)\!) && \text{by definition of } \star && \text{(Definition D.4 page 201)}
\end{aligned}
$$

$$
\begin{aligned}
\left[\alpha (\!(x_n)\!)\right] \star (\!(y_n)\!) &\triangleq \left(\sum_{n\in\mathbb{Z}} \left[\alpha (\!(x_n)\!)\right] (\!(y_{m-n})\!) \right)_{m\in\mathbb{Z}} && \text{by definition of } \star && \text{(Definition D.4 page 201)} \\
&= \alpha \left(\sum_{n\in\mathbb{Z}} \left[\alpha (\!(x_n)\!)\right] (\!(y_{m-n})\!) \right)_{m\in\mathbb{Z}} && \text{by } \textit{homogeneous} \text{ property of } \textstyle\sum \text{ operator} \\
&= \boxed{\alpha \left[(\!(x_n)\!) \star (\!(y_n)\!)\right]} && \text{by definition of } \star && \text{(Definition D.4 page 201)} \\
&\triangleq \alpha \left(\sum_{n\in\mathbb{Z}} (\!(x_n)\!) (\!(y_{m-n})\!) \right)_{m\in\mathbb{Z}} && \text{by definition of } \star && \text{(Definition D.4 page 201)} \\
&\triangleq \left(\sum_{n\in\mathbb{Z}} (\!(x_n)\!) \left[\alpha (\!(y_{m-n})\!)\right] \right)_{m\in\mathbb{Z}} && \text{by } \textit{homogeneous} \text{ property of } \textstyle\sum \text{ operator} \\
&\triangleq \boxed{(\!(x_n)\!) \star \left[\alpha (\!(y_n)\!)\right]} && \text{by definition of } \star && \text{(Definition D.4 page 201)}
\end{aligned}
$$

$$
\begin{aligned}
(\!(x_n)\!) \star \left[(\!(y_n)\!) + (\!(z_n)\!)\right] &\triangleq \sum_{n\in\mathbb{Z}} (\!(x_n)\!) \left[(\!(y_{m-n})\!) + (\!(z_{m-n})\!)\right] && \text{by definition of } \star && \text{(Definition D.4 page 201)} \\
&= \sum_{n\in\mathbb{Z}} (\!(x_n)\!) (\!(y_{m-n})\!) + \sum_{n\in\mathbb{Z}} (\!(z_n)\!) (\!(y_{m-n})\!) && \text{by } \textit{additive} \text{ property of } \textstyle\sum \text{ operator} \\
&= \left[(\!(x_n)\!) \star (\!(y_n)\!)\right] + \left[(\!(x_n)\!) \star (\!(z_n)\!)\right] && \text{by definition of } \star && \text{(Definition D.4 page 201)}
\end{aligned}
$$

✏

Proposition D.1. *Let* $\star$ *be the* CONVOLUTION OPERATOR *(Definition D.4 page 201). Let* $\ell^2_{\mathbb{R}}$ *be the set of* ABSOLUTELY SUMMABLE *sequences (Definition D.3 page 201).*

PRP
$$
\left\{ \begin{array}{llll} \text{(A).} & \mathsf{x}(n) & \in & \ell^2_{\mathbb{R}} \quad \textit{and} \\ \text{(B).} & \mathsf{y}(n) & \in & \ell^2_{\mathbb{R}} \end{array} \right\} \implies \left\{ \sum_{k\in\mathbb{Z}} \mathsf{x}[k]\mathsf{y}[n+k] \;=\; \mathsf{x}[-n] \star \mathsf{y}(n) \right\}
$$

✎PROOF:

$$
\begin{aligned}
\sum_{k\in\mathbb{Z}} \mathsf{x}[k]\mathsf{y}[n+k] &= \sum_{-p\in\mathbb{Z}} \mathsf{x}[-p]\mathsf{y}[n-p] && \text{where } p \triangleq -k && \implies k = -p \\
&= \sum_{p\in\mathbb{Z}} \mathsf{x}[-p]\mathsf{y}[n-p] && \text{by } \textit{absolutely summable} \text{ hypothesis} && \text{(Definition D.3 page 201)} \\
&= \sum_{p\in\mathbb{Z}} \mathsf{x}'[p]\mathsf{y}[n-p] && \text{where } \mathsf{x}'[n] \triangleq \mathsf{x}[-n] && \implies \mathsf{x}[-n] = \mathsf{x}'[n] \\
&\triangleq \mathsf{x}'[n] \star \mathsf{y}[n] && \text{by definition of } \textit{convolution} \star && \text{(Definition D.4 page 201)} \\
&\triangleq \mathsf{x}[-n] \star \mathsf{y}[n] && \text{by definition of } \mathsf{x}'[n]
\end{aligned}
$$

✏

D.3 References

1. History of convolution: Dominguez-Torres (2015)
2. Doetsch (1958)
3. Dominguez-Torres (2010)
4. Dimovski (2012)

APPENDIX E

SPECTRAL THEORY

E.1 Operator Spectrum

Definition E.1. [1] *Let* $\mathbf{A} \in \mathscr{B}(\boldsymbol{X}, \boldsymbol{Y})$ *be an operator over the linear spaces* $\boldsymbol{X} = (X, F, \oplus, \otimes)$ *and* $\boldsymbol{Y} \triangleq (Y, F, \oplus, \otimes)$. *Let* $\mathscr{N}(\mathbf{A})$ *be the* NULL SPACE *of* $\mathbf{A}$.

DEF

*An **eigenvalue** of* $\mathbf{A}$ *is any value* λ *such that there exists* $\mathbf{x}$ *such that* $\mathbf{A}\boldsymbol{x} = \lambda\boldsymbol{x}$.
*The **eigenspace*** $\boldsymbol{H}_\lambda$ *of* $\mathbf{A}$ *at eigenvalue* λ *is* $\mathscr{N}(\mathbf{A} - \lambda\mathbf{I})$.
*An **eigenvector** of* $\mathbf{A}$ *associated with eigenvalue* λ *is any element of* $\mathscr{N}(\mathbf{A} - \lambda\mathbf{I})$.

Example E.1. [2] Let $\mathbf{D}$ be the differntial operator.

EX

The set $\{e^{\lambda x} | \lambda \in \mathbb{C}\}$ are the eigenvectors of $\mathbf{D}$.

$$
\begin{array}{rcll}
\rho(\mathbf{D}) & = & \varnothing & \text{($\mathbf{D}$ has no non-spectral points whatsoever)}\\
\sigma_\mathsf{p}(\mathbf{D}) & = & \sigma(\mathbf{D}) & \text{(the spectrum of $\mathbf{D}$ is all eigenvalues)}\\
\sigma_\mathsf{c}(\mathbf{D}) & = & \varnothing & \text{($\mathbf{D}$ has no continuous spectrum)}\\
\sigma_\mathsf{r}(\mathbf{D}) & = & \varnothing & \text{($\mathbf{D}$ has no resolvent spectrum)}
\end{array}
$$

✎PROOF:

$$
\begin{aligned}
(\mathbf{D} - \lambda\mathbf{I})e^{\lambda x} &= \mathbf{D}e^{\lambda x} - \lambda\mathbf{I}e^{\lambda x}\\
&= \lambda e^{\lambda x} - \lambda e^{\lambda x}\\
&= 0 && \forall \lambda \in \mathbb{C}
\end{aligned}
$$

This theorem and proof needs more work and investigation to prove/disprove its claims. ✏

Definition E.2. [3] *Let* $\mathbf{A} \in \mathscr{B}(\boldsymbol{X}, \boldsymbol{Y})$ *be an operator over the linear spaces* $\boldsymbol{X} = (X, F, \oplus, \otimes)$ *and* $\boldsymbol{Y} \triangleq (Y, F, \oplus, \otimes)$.

---

[1] Bollobás (1999) page 168, Descartes (1637), Descartes (1954), Cayley (1858), Hilbert (1904) page 67, Hilbert (1912),

[2] Pedersen (2000) page 79

[3] Michel and Herget (1993) page 439

quantity	$\mathscr{N}(\mathbf{A}-\lambda\mathbf{I}) = \{\mathbb{0}\}$ ($\boldsymbol{x} = \mathbb{0}$ is the only solution)	$\overline{\mathscr{R}(\mathbf{A}-\lambda\mathbf{I})} = \boldsymbol{X}$ (dense)	$(\mathbf{A}-\lambda\mathbf{I})^{-1} \in \mathscr{B}(\boldsymbol{X}, \boldsymbol{Y})$ (continuous/bounded)
$\rho(\mathbf{A})$ (resolvent set)	1	1	1
$\sigma_p(\mathbf{A})$ (point spectrum)	0		
$\sigma_r(\mathbf{A})$ (residual spectrum)	1	0	
$\sigma_c(\mathbf{A})$ (continuous spectrum)	1	1	0

Table E.1: Spectrum of an operator $\mathbf{A}$

DEF

*The **resolvent set** $\rho(\mathbf{A})$ of operator $\mathbf{A}$ is defined as*

$$\rho(\mathbf{A}) \triangleq \left\{ \lambda \in F \left| \begin{array}{lll} 1. & \mathscr{N}(\mathbf{A}-\lambda\mathbf{I}) = \{\mathbb{0}\} & \text{(no non-zero eigenvectors)} \quad \text{and} \\ 2. & \overline{\mathscr{R}(\mathbf{A}-\lambda\mathbf{I})} = \boldsymbol{X} & \text{(the range is dense in } \boldsymbol{X}\text{).} \quad \text{and} \\ 3. & (\mathbf{A}-\lambda\mathbf{I})^{-1} \in \mathscr{B}(\boldsymbol{X}, \boldsymbol{Y}) & \text{(inverse is continuous/bounded).} \end{array} \right. \right\}$$

*The **spectrum** $\sigma(\mathbf{A})$ of operator $\mathbf{A}$ is defined as*

$$\sigma(\mathbf{A}) \triangleq F \backslash \rho(\mathbf{A}).$$

Definition E.3. [4] *Let $\mathbf{A} \in \mathscr{B}(\boldsymbol{X}, \boldsymbol{Y})$ be an operator over the linear spaces $\boldsymbol{X} = (X, F, \oplus, \otimes)$ and $\boldsymbol{Y} \triangleq (Y, F, \oplus, \otimes)$.*

DEF

*The **point spectrum** $\sigma_p(\mathbf{A})$ of operator $\mathbf{A}$ is defined as*

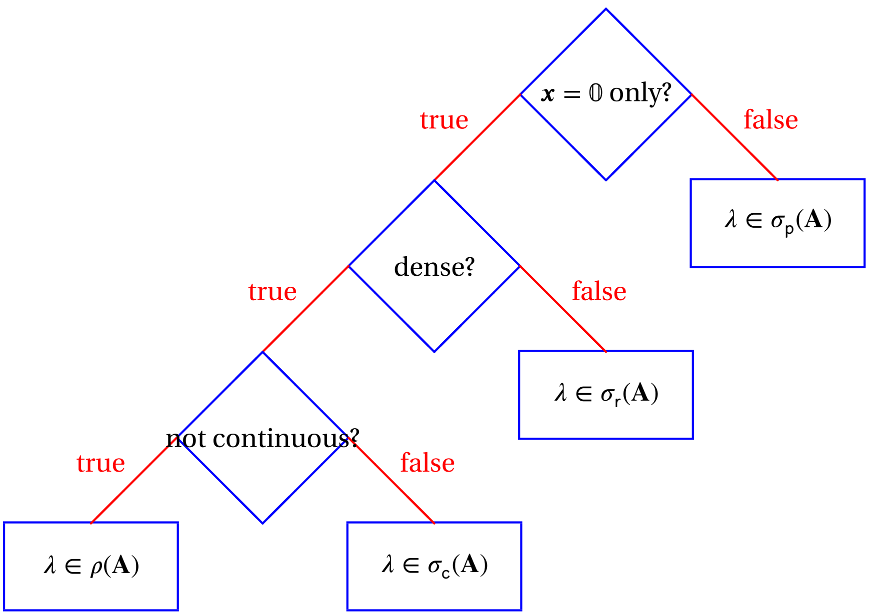
$$\sigma_p(\mathbf{A}) \triangleq \left\{ \lambda \in F \left| \; 1. \;\; \mathscr{N}(\mathbf{A}-\lambda\mathbf{I}) \supsetneq \{\mathbb{0}\} \quad \text{(has non-zero eigenvector)} \right. \right\}$$

*The **residual spectrum** $\sigma_r(\mathbf{A})$ of operator $\mathbf{A}$ is defined as*

$$\sigma_r(\mathbf{A}) \triangleq \left\{ \lambda \in F \left| \begin{array}{lll} 1. & \mathscr{N}(\mathbf{A}-\lambda\mathbf{I}) = \{\mathbb{0}\} & \text{(no non-zero eigenvectors)} \quad \text{and} \\ 2. & \overline{\mathscr{R}(\mathbf{A}-\lambda\mathbf{I})} \neq \boldsymbol{X} & \text{(not dense in } \boldsymbol{X}\text{—has gaps).} \end{array} \right. \right\}$$

*The **continuous spectrum** $\sigma_c(\mathbf{A})$ of operator $\mathbf{A}$ is defined as*

$$\sigma_c(\mathbf{A}) \triangleq \left\{ \lambda \in F \left| \begin{array}{lll} 1. & \mathscr{N}(\mathbf{A}-\lambda\mathbf{I}) = \{\mathbb{0}\} & \text{(no non-zero eigenvectors)} \quad \text{and} \\ 2. & \overline{\mathscr{R}(\mathbf{A}-\lambda\mathbf{I})} = \boldsymbol{X} & \text{(dense in } \boldsymbol{X}\text{).} \quad \text{and} \\ 3. & (\mathbf{A}-\lambda\mathbf{I})^{-1} \notin \mathscr{B}(\boldsymbol{X}, \boldsymbol{Y}) & \text{(not continuous / not bounded)} \end{array} \right. \right\}$$



The spectral components' definitions are illustrated in the figure to the left and summarized in Table E.1 (page 206). Let a family of operators $\mathbf{B}(\lambda)$ be defined with respect to an operator $\mathbf{A}$ such that $\mathbf{B}(\lambda) \triangleq (\mathbf{A} - \lambda\mathbf{I})$. Normally, we might expect a "normal" or "regular" or even "mundane" operator $\mathbf{B}(\lambda)$ to have the properties

1. $\mathbf{B}(\lambda)\boldsymbol{x} = 0$ if and only if $\boldsymbol{x} = \mathbb{0}$
2. $\mathbf{B}(\lambda)\boldsymbol{x}$ spans virtually all of $\boldsymbol{X}$ as we vary $\boldsymbol{x}$
3. $\mathbf{B}^{-1}(\lambda)$ is continuous.

After all, these are the properties that we would have if $\mathbf{B}(\lambda)$ were simply an affine operator in the

[4] Bollobás (1999) page 168, Hilbert (1906) pages 169–172

field of real numbers— such as $[\mathbf{B}(\lambda)](x) \triangleq [\lambda](x) = \lambda x$ which is 0 if and only if $x = 0$, has range $\mathscr{R}(\lambda) = \mathbb{R}$, and its inverse $\lambda^{-1}x$ is continuous.

If for some λ the operator $\mathbf{B}(\lambda)$ does have all these "regular" properties, then that λ part of the *resolvent set* of $\mathbf{A}$ and λ is called *regular*. However if for some λ the operator $\mathbf{B}(\lambda)$ fails any of these conditions, then that λ part of the *spectrum* of $\mathbf{A}$. And which conditions it fails determines which component of the spectrum it is in.

Theorem E.1. [5] *Let $\mathbf{A} \in \mathscr{B}(\boldsymbol{X}, \boldsymbol{Y})$ be an operator.*

THM
$$\sigma(\mathbf{A}) = \sigma_{\mathsf{p}}(\mathbf{A}) \cup \sigma_{\mathsf{c}}(\mathbf{A}) \cup \sigma_{\mathsf{r}}(\mathbf{A})$$

Theorem E.2 (Spectral Theorem). [6] *Let $\mathbf{N} \in Y^X$ be an operator.*

THM
$$\left.\begin{array}{ll} (A). & \underbrace{\mathbf{N}^*\mathbf{N} = \mathbf{N}\mathbf{N}^*}_{\mathbf{N} \text{ is NORMAL}} \\ (B). & \mathbf{N} \text{ is COMPACT} \end{array}\right\} \implies \left\{\begin{array}{lrcl} (1). & \mathbf{N} & = & \sum_n \lambda_n \mathbf{P}_n \\ (2). & \sum_n \mathbf{P}_n & = & \mathbf{I} \\ (3). & \mathbf{P}_n\mathbf{P}_m & = & \bar{\delta}_{n-m}\mathbf{P}_n \\ (4). & \dim(\boldsymbol{H}_n) & < & \infty \\ (5). & \left|\left\{\lambda_n | \lambda_n \neq 0\right\}\right| & \multicolumn{2}{l}{\text{is COUNTABLY INFINITE}} \end{array}\right.$$

where

$$\begin{array}{rcll} (\!(\lambda_n)\!)_{n\in\mathbb{Z}} & \triangleq & \sigma_{\mathsf{p}}(\mathbf{N}) & \text{(eigenvalues of } \mathbf{N}) \\ \boldsymbol{H}_n & \triangleq & \mathscr{N}(\mathbf{N} - \lambda_n\mathbf{I}) & (\lambda_n \text{ is the eigenspace of } \mathbf{N} \text{ at } \lambda_n \text{ in } \boldsymbol{Y}) \\ \boldsymbol{H}_n & = & \mathbf{P}_n\boldsymbol{Y} & (\mathbf{P}_n \text{ is the projection operator that generates } \boldsymbol{H}_n) \end{array}$$

E.2 Fredholm kernels

Definition E.4. [7]

DEF
*A **Fredholm operator** $\mathbf{K}$ is defined as*

$$[\mathbf{K}\mathsf{f}](t) \triangleq \underbrace{\int_a^b \underbrace{\kappa(t,s)}_{\text{kernel}}\mathsf{f}(s)\,\mathsf{d}s}_{\text{Fredholm integral equation of the first kind}^{8}} \qquad \forall \mathsf{f} \in \boldsymbol{L}_2([a,b])$$

Example E.2. Examples of Fredholm operators include

1. Fourier Transform $\quad [\tilde{\mathbf{F}}\mathsf{x}](f) = \int_t \mathsf{x}(t)e^{-i2\pi ft}\,\mathsf{d}t \quad \kappa(t,f) = e^{-i2\pi ft}$
2. Inverse Fourier Transform $\quad [\tilde{\mathbf{F}}^{-1}\tilde{\mathsf{x}}](t) = \int_f \tilde{\mathsf{x}}(f)e^{i2\pi ft}\,\mathsf{d}f \quad \kappa(f,t) = e^{i2\pi ft}$
3. Laplace operator $\quad [\mathbf{L}\mathsf{x}](s) = \int_t \mathsf{x}(t)e^{-st}\,\mathsf{d}t \quad \kappa(t,s) = e^{-st}$
4. autocorrelation operator $\quad [\mathbf{R}\mathsf{x}](t) = \int_s R(t,s)\mathsf{x}(s)\,\mathsf{d}s \quad \kappa(t,s) = R(t,s)$

Theorem E.3. *Let $\mathbf{K}$ be a Fredholm operator with kernel $\kappa(t,s)$ and adjoint $\mathbf{K}^*$.*

THM
$$[\mathbf{K}\mathsf{f}](t) = \int_A \kappa(t,s)\mathsf{f}(s)\,\mathsf{d}s \iff \left[\mathbf{K}^*\mathsf{f}\right](t) = \int_A \kappa^*(s,t)\mathsf{f}(s)\,\mathsf{d}s$$

[5] Michel and Herget (1993) page 440
[6] Michel and Herget (1993) page 457, Bollobás (1999) page 200, Hilbert (1906), Hilbert (1912), von Neumann (1929), de Witt (1659)
[7] Michel and Herget (1993) page 425
[8] The equation $\int_u \kappa(t,s)\mathsf{f}(s)\,\mathsf{d}s$ is a **Fredholm integral equation of the first kind** and $\kappa(t,u)$ is the **kernel** of the equation. References: Fredholm (1900), Fredholm (1903) page 365, Michel and Herget (1993) page 97, Keener (1988) page 101

✎PROOF:

$$\begin{aligned}
[\mathbf{K}f](t) &= \int_A \kappa(t,s)f(s)\,ds \\
\iff \langle [\mathbf{K}f](t) \,|\, g(t)\rangle &= \left\langle \int_s \kappa(t,s)f(s)\,ds \,\Big|\, g(t) \right\rangle && \text{by left hypothesis} \\
&= \int_s f(s)\,\langle \kappa(t,s) \,|\, g(t)\rangle\,ds && \text{by additivity property of } \langle \triangle \,|\, \nabla \rangle \text{ (Definition N.1 page 309)} \\
&= \int_s f(s)\,\langle g(t) \,|\, \kappa(t,s)\rangle^*\,ds && \text{by conjugate symmetry property of } \langle \triangle \,|\, \nabla \rangle \text{ (Definition N.1 page 309)} \\
&= \langle f(s) \,|\, \langle g(t) \,|\, \kappa(t,s)\rangle\rangle && \text{by local definition of } \langle \triangle \,|\, \nabla \rangle \\
&= \left\langle f(s) \,\Big|\, \underbrace{\int_t \kappa^*(t,s)g(t)\,dt}_{[\mathbf{K}^*g](s)} \right\rangle && \text{by local definition of } \langle \triangle \,|\, \nabla \rangle \\
\iff \left[\mathbf{K}^*g\right](s) &= \int_A \kappa^*(t,s)g(t)\,dt && \text{by right hypothesis} \\
\iff \left[\mathbf{K}^*g\right](\sigma) &= \int_A \kappa^*(\tau,\sigma)g(\tau)\,d\tau && \text{by change of variable: } \tau = t,\ \sigma = s \\
\iff \left[\mathbf{K}^*f\right](t) &= \int_A \kappa^*(s,t)f(s)\,ds && \text{by change of variable: } t = \sigma,\ s = \tau,\ f = g
\end{aligned}$$

✏

Corollary E.1. [9] *Let* $\mathbf{K}$ *be an Fredholm operator with kernel* $\kappa(t,s)$ *and adjoint* $\mathbf{K}^*$.

COR

$$\underbrace{\mathbf{K} = \mathbf{K}^*}_{\mathbf{K} \text{ is self-adjoint}} \iff \underbrace{\kappa(t,s) = \kappa^*(s,t)}_{\text{kernel is conjugate symmetric}}$$

✎PROOF:

$$\begin{aligned}
\mathbf{K} = \mathbf{K}^* &\iff \int_A \kappa(t,s)f(s)\,ds = \int_A \kappa^*(s,t)f(s)\,ds && \text{by Theorem E.3 page 207} \\
&\iff \kappa(t,s) = \kappa^*(s,t)
\end{aligned}$$

✏

Theorem E.4 (Mercer's Theorem). [10] *Let* $\mathbf{K}$ *be an Fredholm operator with kernel* $\kappa(t,s)$ *and eigen-system* $\left(\!\left(\lambda_n, \phi_n(t)\right)\!\right)_{n\in\mathbb{Z}}$.

THM

$$\left\{\begin{array}{ll}
(A). & \underbrace{\int_a^b\int_a^b \kappa(t,s)f(t)f^*(s)\,dt \geq 0}_{\text{positive}} \quad \text{and} \\
(B). & \kappa(t,s) \text{ is CONTINUOUS on } [a:b]\times[a:b]
\end{array}\right\} \implies \left\{\begin{array}{ll}
(1). & \kappa(t,s) = \sum_n \lambda_n \phi_n(t)\phi_n^*(s) \quad \text{and} \\
(2). & \kappa(t,s) \text{ CONVERGES ABSOLUTELY and UNIFORMLY on } [a:b]\times[a:b]
\end{array}\right\}$$

[9] Michel and Herget (1993) page 430

[10] Gohberg et al. (2003) page 198, Courant and Hilbert (1930) pages 138–140, Mercer (1909) page 439

APPENDIX F

MATRIX CALCULUS

Optimization problems often require finding the value of some parameter which results in some measure reaching a minimum or maximum value. Often this optimal parameter value can be found by solving the single equation generated by the partial derivative of the measure with respect to the parameter. When there are several parameters, optimization often requires several simultaneous equations generated by the partial derivatives of the measure with respect to each parameter. The need for several partial derivatives and several simultaneous equations leads to a natual union of two branches of mathematics— partial differential equations and linear algebra. In general, we would like to not only be able to take the partial derivative of a scalar with respect to another scalar, but to be able to take the partial derivative of a vector with respect to another vector. This generalization is the problem addressed in this section. Other references are also available.[1]

F.1 First derivative of a vector with respect to a vector

Definition F.1.

DEF

$\boldsymbol{x}$ is a vector with the following properties:

1. $$\boldsymbol{x} \triangleq \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad \text{(n element column vector)}$$
2. $$\frac{\partial}{\partial x_k} x_j = \bar{\delta}_{kj} \qquad ((x_1, x_2, \ldots, x_n) \text{ are mutually independent})$$

Definition F.2 (Jacobian matrix). [2] *The* ***gradient of $\boldsymbol{y}$ with respect to $\boldsymbol{x}$****, as well as the* ***gradient of $\boldsymbol{y}^T$ with respect to $\boldsymbol{x}$****, is defined as*

[1] Graham (1981) ⟨Chapter 4⟩, Haykin (2001) ⟨Appendix B⟩, Moon and Stirling (2000) ⟨Appendix E⟩, Scharf (1991) pages 274–276, Trees (2002) ⟨Section A.7⟩, Felippa (1999)

[2] Graham (1981) page 52, Graham (2018) page 529780486824178§"4.2 The Derivatives of Vectors", Scharf (1991) page 274, Trees (2002) page 1398, Anderson (1984) page 13 ⟨§"2.2.5 Transformation of Variables"⟩, Anderson (1958) page 11 ⟨§"2.2.5 Transformation of Variables"⟩

DEF

$$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} \triangleq \frac{\partial \boldsymbol{y}^T}{\partial \boldsymbol{x}} \triangleq \underbrace{\begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial x_n} & \frac{\partial y_2}{\partial x_n} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}}_{n\times m \text{ matrix}} \qquad \forall \boldsymbol{y} \in \mathbb{C}^m$$

Remark F.1. Depending on whether $\boldsymbol{x}$ and $\boldsymbol{y}$ are scalars or vectors, $\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}$ takes on the following forms:[3]

	y scalar	y vector
x scalar	$\frac{\partial y}{\partial x}$	$\frac{\partial \boldsymbol{y}}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x} & \frac{\partial y_2}{\partial x} & \cdots & \frac{\partial y_m}{\partial x} \end{bmatrix}$
x vector	$\frac{\partial y}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{bmatrix}$	$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} \triangleq \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial x_n} & \frac{\partial y_2}{\partial x_n} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}$

Lemma F.1. *Let $\boldsymbol{x} \in \mathbb{R}^n$ be a vector. Then*

LEM

$$\frac{\partial}{\partial x_k} x_i x_j = \bar{\delta}_{ik} x_j + \bar{\delta}_{jk} x_i = \begin{cases} 2x_k & \text{for } i = j = k \\ x_j & \text{for } i = k \text{ and } j \neq k \\ x_i & \text{for } i \neq k \text{ and } j = k \\ 0 & \text{otherwise} \end{cases}$$

Lemma F.2.

LEM

$$\left(\boldsymbol{x}^H \mathbf{A} \boldsymbol{x}\right) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i^* x_j \qquad \forall \begin{array}{lll} \mathbf{A} \in (\mathbb{C}^n \times \mathbb{C}^n) & (n\times n \text{ array}) & \text{and} \\ \boldsymbol{x} \in \mathbb{C}^n & (n \text{ element column vector}) & \end{array}$$

✎PROOF:

$$\begin{aligned}
\boldsymbol{x}^H \mathbf{A} \boldsymbol{x} &\triangleq \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^* \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} && \text{by definitions of } \mathbf{A} \text{ and } \boldsymbol{x} \\
&= \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^* \sum_{i=1}^{n} x_i \begin{bmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{ni} \end{bmatrix} \\
&= \sum_{i=1}^{n} x_i \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^* \begin{bmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{ni} \end{bmatrix} \\
&= \sum_{i=1}^{n} x_i \sum_{j=1}^{n} a_{ji} x_j^*
\end{aligned}$$

[3]For the generalization of the partial derivative of a matrix with respect to a matrix, see Graham (1981) ⟨chapter 6⟩. Graham uses *kronecker products* to handle the additional dimensions(?)

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} x_i^* x_j$$

Lemma F.3.

LEM

$$\frac{\partial}{\partial \boldsymbol{x}}[\mathrm{a}(\boldsymbol{x})\,\mathrm{b}(\boldsymbol{x})] = \mathrm{a}(\boldsymbol{x})\left[\frac{\partial}{\partial \boldsymbol{x}}\mathrm{b}(\boldsymbol{x})\right] + \left[\frac{\partial}{\partial \boldsymbol{x}}\mathrm{a}(\boldsymbol{x})\right]\mathrm{b}(\boldsymbol{x}) \qquad \underbrace{\forall \mathrm{a}, \mathrm{b} : \mathbb{R}^n \to \mathbb{R}}_{\mathrm{a}(\boldsymbol{x}),\, \mathrm{b}(\boldsymbol{x}) \textit{ are functions from a vector } \boldsymbol{x} \textit{ to a scalar in } \mathbb{R}}$$

✎PROOF:

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{x}}[\mathrm{a}(\boldsymbol{x})\,\mathrm{b}(\boldsymbol{x})] &= \begin{bmatrix} \frac{\partial}{\partial x_1}[\mathrm{a}(\boldsymbol{x})\,\mathrm{b}(\boldsymbol{x})] \\ \frac{\partial}{\partial x_2}[\mathrm{a}(\boldsymbol{x})\,\mathrm{b}(\boldsymbol{x})] \\ \vdots \\ \frac{\partial}{\partial x_n}[\mathrm{a}(\boldsymbol{x})\,\mathrm{b}(\boldsymbol{x})] \end{bmatrix} && \text{by definition of } \frac{\partial}{\partial \boldsymbol{x}} \qquad \text{(Definition F.2 page 209)} \\
&= \begin{bmatrix} \mathrm{a}(\boldsymbol{x})\frac{\partial \mathrm{b}(\boldsymbol{x})}{\partial x_1} + \mathrm{b}(\boldsymbol{x})\frac{\partial \mathrm{a}(\boldsymbol{x})}{\partial x_1} \\ \mathrm{a}(\boldsymbol{x})\frac{\partial \mathrm{b}(\boldsymbol{x})}{\partial x_2} + \mathrm{b}(\boldsymbol{x})\frac{\partial \mathrm{a}(\boldsymbol{x})}{\partial x_2} \\ \vdots \\ \mathrm{a}(\boldsymbol{x})\frac{\partial \mathrm{b}(\boldsymbol{x})}{\partial x_n} + \mathrm{b}(\boldsymbol{x})\frac{\partial \mathrm{a}(\boldsymbol{x})}{\partial x_n} \end{bmatrix} && \text{by } \textit{linearity} \text{ of } \frac{\partial}{\partial \boldsymbol{x}} \\
&= \begin{bmatrix} \mathrm{a}(\boldsymbol{x})\frac{\partial \mathrm{b}(\boldsymbol{x})}{\partial x_1} \\ \mathrm{a}(\boldsymbol{x})\frac{\partial \mathrm{b}(\boldsymbol{x})}{\partial x_2} \\ \mathrm{a}(\boldsymbol{x})\frac{\partial \mathrm{b}(\boldsymbol{x})}{\partial x_n} \end{bmatrix} + \begin{bmatrix} \frac{\partial \mathrm{a}(\boldsymbol{x})}{\partial x_1}\mathrm{b}(\boldsymbol{x}) \\ \frac{\partial \mathrm{a}(\boldsymbol{x})}{\partial x_2}\mathrm{b}(\boldsymbol{x}) \\ \vdots \\ \frac{\partial \mathrm{a}(\boldsymbol{x})}{\partial x_n}\mathrm{b}(\boldsymbol{x}) \end{bmatrix} && \text{by } \textit{linearity} \text{ of } \textit{vector addition} \\
&= \mathrm{a}(\boldsymbol{x})\left[\frac{\partial \mathrm{b}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right] + \left[\frac{\partial \mathrm{a}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right]\mathrm{b}(\boldsymbol{x})
\end{aligned}$$

✏

Theorem F.1. [4]

LEM

$$\frac{\partial}{\partial \boldsymbol{x}}\boldsymbol{x} = \mathbf{I} \qquad \forall \boldsymbol{x} \in \mathbb{R}^n$$

✎PROOF:

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{x}}\boldsymbol{x} &= \begin{bmatrix} \frac{\partial x_1}{\partial x_1} & \frac{\partial x_2}{\partial x_1} & \cdots & \frac{\partial x_n}{\partial x_1} \\ \frac{\partial x_1}{\partial x_2} & \frac{\partial x_2}{\partial x_2} & \cdots & \frac{\partial x_n}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_1}{\partial x_n} & \frac{\partial x_2}{\partial x_n} & \cdots & \frac{\partial x_n}{\partial x_n} \end{bmatrix} && \text{by Definition F.2 page 209} \\
&= \begin{bmatrix} \bar{\delta}_{11} & \bar{\delta}_{21} & \cdots & \bar{\delta}_{n1} \\ \bar{\delta}_{12} & \bar{\delta}_{22} & \cdots & \bar{\delta}_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{\delta}_{1n} & \bar{\delta}_{2n} & \cdots & \bar{\delta}_{nn} \end{bmatrix} && \text{by Definition F.1 page 209 (mutual independence property)}
\end{aligned}$$

[4] Scharf (1991) page 274, Trees (2002) page 1398

$$= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \qquad \text{by definition of kronecker delta function } \bar{\delta}$$

$$= \mathbf{I} \qquad \text{by definition of identity operator } \mathbf{I}$$

✑

Theorem F.2.

THM
$$\frac{\partial}{\partial \boldsymbol{x}}(\mathbf{A}\boldsymbol{x}) = \mathbf{A}^T + \sum_{i=1}^{n}\left(\frac{\partial}{\partial \boldsymbol{x}}\begin{bmatrix} a_{1i} & a_{2i} & \cdots & a_{mi}\end{bmatrix}\right)x_i \qquad \forall \boldsymbol{x}\in\mathbb{R}^n,\ \mathbf{A}\in\mathbb{C}^m\times\mathbb{C}^n$$

✎PROOF: Let

$$\boldsymbol{x} \triangleq \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad \mathbf{A} \triangleq \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

$$\frac{\partial}{\partial \boldsymbol{x}}(\mathbf{A}\boldsymbol{x}) = \frac{\partial}{\partial \boldsymbol{x}}\left(\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right) \qquad \text{by definition of } A \text{ and } \boldsymbol{x}$$

$$= \frac{\partial}{\partial \boldsymbol{x}}\begin{bmatrix} \sum_{i=1}^{n} a_{1i}x_i \\ \sum_{i=1}^{n} a_{2i}x_i \\ \vdots \\ \sum_{i=1}^{n} a_{mi}x_i \end{bmatrix} \qquad \text{by matrix multiplication}$$

$$= \frac{\partial}{\partial \boldsymbol{x}}\sum_{i=1}^{n}\begin{bmatrix} a_{1i}x_i \\ a_{2i}x_i \\ \vdots \\ a_{mi}x_i \end{bmatrix}$$

$$= \sum_{i=1}^{n}\frac{\partial}{\partial \boldsymbol{x}}\begin{bmatrix} a_{1i}x_i \\ a_{2i}x_i \\ \vdots \\ a_{mi}x_i \end{bmatrix}$$

$$= \sum_{i=1}^{n}\begin{bmatrix} \frac{\partial a_{1i}x_i}{\partial x_1} & \frac{\partial a_{2i}x_i}{\partial x_1} & \cdots & \frac{\partial a_{mi}x_i}{\partial x_1} \\ \frac{\partial a_{1i}x_i}{\partial x_2} & \frac{\partial a_{2i}x_i}{\partial x_2} & \cdots & \frac{\partial a_{mi}x_i}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial a_{1i}x_i}{\partial x_n} & \frac{\partial a_{2i}x_i}{\partial x_n} & \cdots & \frac{\partial a_{mi}x_i}{\partial x_n} \end{bmatrix} \qquad \text{by Definition F.2 page 209}$$

$$= \sum_{i=1}^{n}\begin{bmatrix} a_{1i}\frac{\partial x_i}{\partial x_1} + \frac{\partial a_{1i}}{\partial x_1}x_i & a_{2i}\frac{\partial x_i}{\partial x_1} + \frac{\partial a_{2i}}{\partial x_1}x_i & \cdots & a_{mi}\frac{\partial x_i}{\partial x_1} + \frac{\partial a_{mi}}{\partial x_1}x_i \\ a_{1i}\frac{\partial x_i}{\partial x_2} + \frac{\partial a_{1i}}{\partial x_2}x_i & a_{2i}\frac{\partial x_i}{\partial x_2} + \frac{\partial a_{2i}}{\partial x_2}x_i & \cdots & a_{mi}\frac{\partial x_i}{\partial x_2} + \frac{\partial a_{mi}}{\partial x_2}x_i \\ \vdots & \vdots & \ddots & \vdots \\ a_{1i}\frac{\partial x_i}{\partial x_n} + \frac{\partial a_{1i}}{\partial x_n}x_i & a_{2i}\frac{\partial x_i}{\partial x_n} + \frac{\partial a_{2i}}{\partial x_n}x_i & \cdots & a_{mi}\frac{\partial x_i}{\partial x_n} + \frac{\partial a_{mi}}{\partial x_n}x_i \end{bmatrix} \qquad \text{by Lemma F.3 page 211}$$

$$
= \sum_{i=1}^{n} \begin{bmatrix} a_{1i}\frac{\partial x_i}{\partial x_1} & a_{2i}\frac{\partial x_i}{\partial x_1} & \cdots & a_{mi}\frac{\partial x_i}{\partial x_1} \\ a_{1i}\frac{\partial x_i}{\partial x_2} & a_{2i}\frac{\partial x_i}{\partial x_2} & \cdots & a_{mi}\frac{\partial x_i}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1i}\frac{\partial x_i}{\partial x_n} & a_{2i}\frac{\partial x_i}{\partial x_n} & \cdots & a_{mi}\frac{\partial x_i}{\partial x_n} \end{bmatrix} + \sum_{i=1}^{n} \begin{bmatrix} \frac{\partial a_{1i}}{\partial x_1}x_i & \frac{\partial a_{2i}}{\partial x_1}x_i & \cdots & \frac{\partial a_{mi}}{\partial x_1}x_i \\ \frac{\partial a_{1i}}{\partial x_2}x_i & \frac{\partial a_{2i}}{\partial x_2}x_i & \cdots & \frac{\partial a_{mi}}{\partial x_2}x_i \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial a_{1i}}{\partial x_n}x_i & \frac{\partial a_{2i}}{\partial x_n}x_i & \cdots & \frac{\partial a_{mi}}{\partial x_n}x_i \end{bmatrix}
$$

$$
= \sum_{i=1}^{n} \begin{bmatrix} a_{1i}\bar{\delta}_{i1} & a_{2i}\bar{\delta}_{i1} & \cdots & a_{mi}\bar{\delta}_{i1} \\ a_{1i}\bar{\delta}_{i2} & a_{2i}\bar{\delta}_{i2} & \cdots & a_{mi}\bar{\delta}_{i2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1i}\bar{\delta}_{in} & a_{2i}\bar{\delta}_{in} & \cdots & a_{mi}\bar{\delta}_{in} \end{bmatrix} + \sum_{i=1}^{n} \left(\frac{\partial}{\partial \mathbf{x}} \begin{bmatrix} a_{1i} & a_{2i} & \cdots & a_{mi} \end{bmatrix} \right) x_i \qquad \text{by Lemma F.1}
$$

$$
= \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix} + \sum_{i=1}^{n} \left(\frac{\partial}{\partial \mathbf{x}} \begin{bmatrix} a_{1i} & a_{2i} & \cdots & a_{mi} \end{bmatrix} \right) x_i \qquad \text{by definition of } \bar{\delta}
$$

$$
= \mathbf{A}^T + \sum_{i=1}^{n} \left(\frac{\partial}{\partial \mathbf{x}} \begin{bmatrix} a_{1i} & a_{2i} & \cdots & a_{mi} \end{bmatrix} \right) x_i
$$

Theorem F.3 (Affine equations). [5]

THM $\mathbf{A}$ *and* $\mathbf{B}$ *are independent of* $\mathbf{x}$ $\implies$
$$
\left\{ \begin{array}{lcll} \frac{\partial}{\partial \mathbf{x}}(\mathbf{A}\mathbf{x}) & = & \mathbf{A}^T & \forall \mathbf{x}\in\mathbb{R}^n,\ \mathbf{A}\in\mathbb{C}^m\times\mathbb{C}^n \\ \frac{\partial}{\partial \mathbf{x}}(\mathbf{x}^T\mathbf{B}) & = & \mathbf{B} & \forall \mathbf{x}\in\mathbb{R}^n,\ \mathbf{B}\in\mathbb{C}^n\times\mathbb{C}^m \end{array} \right.
$$

✎PROOF: Let $\mathbf{B} \triangleq \mathbf{A}^T$.

1. Proof that $\frac{\partial}{\partial \mathbf{x}}(A\mathbf{x}) = \mathbf{A}^T$:

$$
\begin{aligned}
\frac{\partial}{\partial \mathbf{x}}(\mathbf{A}\mathbf{x}) &= \mathbf{A}^T + \sum_{i=1}^{n} \left(\frac{\partial}{\partial \mathbf{x}} \begin{bmatrix} a_{1i} & a_{2i} & \cdots & a_{mi} \end{bmatrix} \right) x_i && \text{by Theorem F.2 page 212} \\
&= \mathbf{A}^T + \sum_{i=1}^{n} \begin{bmatrix} \frac{\partial}{\partial \mathbf{x}}a_{1i} & \frac{\partial}{\partial \mathbf{x}}a_{2i} & \cdots & \frac{\partial}{\partial \mathbf{x}}a_{mi} \end{bmatrix} x_i \\
&= \mathbf{A}^T + \sum_{i=1}^{n} \begin{bmatrix} 0 & 0 & \cdots & 0 \end{bmatrix} x_i \quad 0 && \text{by left hypothesis} \\
&= \mathbf{A}^T
\end{aligned}
$$

2. Proof that $\frac{\partial}{\partial \mathbf{x}}\left(\mathbf{x}^T\mathbf{B}\right) = \mathbf{B}$:

$$
\begin{aligned}
\frac{\partial}{\partial \mathbf{x}}\left(\mathbf{x}^T\mathbf{B}\right) &= \frac{\partial}{\partial \mathbf{x}}\left(\mathbf{x}^T\mathbf{A}^T\right) && \text{by definition of } \mathbf{B} \\
&= \frac{\partial}{\partial \mathbf{x}}\left[(\mathbf{A}\mathbf{x})^T\right] \\
&= \frac{\partial}{\partial \mathbf{x}}(\mathbf{A}\mathbf{x}) && \text{by Definition F.2 page 209} \\
&= \mathbf{A}^T && \text{by Theorem F.3 page 213} \\
&= \mathbf{B} && \text{by definition of } \mathbf{B}
\end{aligned}
$$

[5] Graham (1981) page 54, Graham (2018) page 549780486824178§"4.2 The Derivatives of Vectors"

Theorem F.4 (Product rule). [6] *Let $\boldsymbol{y}$ and $\boldsymbol{z}$ be functions of $\boldsymbol{x}$ and*

THM
$$\frac{\partial}{\partial \boldsymbol{x}}\boldsymbol{z}^T\boldsymbol{y} = \frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}}\boldsymbol{y} + \frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}\boldsymbol{z} \qquad \forall \boldsymbol{x}\in\mathbb{R}^n,\ \boldsymbol{y}\in\mathbb{R}^m,\ \boldsymbol{z}\in\mathbb{R}^m$$

✎PROOF:

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{x}}\boldsymbol{z}^T\boldsymbol{y} &= \frac{\partial}{\partial \boldsymbol{x}}\sum_{k=1}^{m} z_k y_k \\
&= \sum_{k=1}^{m}\frac{\partial}{\partial \boldsymbol{x}} z_k y_k \\
&= \sum_{k=1}^{m}\frac{\partial z_k}{\partial \boldsymbol{x}} y_k + \sum_{k=1}^{m}\frac{\partial y_k}{\partial \boldsymbol{x}} z_k && \text{by Lemma F.3 page 211} \\
&= \begin{bmatrix}
\frac{\partial z_1}{\partial x_1}y_1 & + & \frac{\partial z_2}{\partial x_1}y_2 & + & \cdots & + & \frac{\partial z_n}{\partial x_1}y_n \\
\frac{\partial z_1}{\partial x_1}y_1 & + & \frac{\partial z_2}{\partial x_1}y_2 & + & \cdots & + & \frac{\partial z_n}{\partial x_1}y_n \\
\vdots & & & & \ddots & & \vdots \\
\frac{\partial z_1}{\partial x_1}y_1 & + & \frac{\partial z_2}{\partial x_1}y_2 & + & \cdots & + & \frac{\partial z_n}{\partial x_1}y_n
\end{bmatrix} + \begin{bmatrix}
\frac{\partial y_1}{\partial x_1}z_1 & + & \frac{\partial y_2}{\partial x_1}z_2 & + & \cdots & + & \frac{\partial y_n}{\partial x_1}z_n \\
\frac{\partial y_1}{\partial x_1}z_1 & + & \frac{\partial y_2}{\partial x_1}z_2 & + & \cdots & + & \frac{\partial y_n}{\partial x_1}z_n \\
\vdots & & & & \ddots & & \vdots \\
\frac{\partial y_1}{\partial x_1}z_1 & + & \frac{\partial y_2}{\partial x_1}z_2 & + & \cdots & + & \frac{\partial y_n}{\partial x_1}z_n
\end{bmatrix} \\
&= \begin{bmatrix}
\frac{\partial z_1}{\partial x_1} & \frac{\partial z_2}{\partial x_1} & \cdots & \frac{\partial z_n}{\partial x_1} \\
\frac{\partial z_1}{\partial x_1} & \frac{\partial z_2}{\partial x_1} & \cdots & \frac{\partial z_n}{\partial x_1} \\
\vdots & & \ddots & \vdots \\
\frac{\partial z_1}{\partial x_1} & \frac{\partial z_2}{\partial x_1} & \cdots & \frac{\partial z_n}{\partial x_1}
\end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} + \begin{bmatrix}
\frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_1} \\
\frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_1} \\
\vdots & & \ddots & \vdots \\
\frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_1}
\end{bmatrix}\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \\
&= \frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}}\boldsymbol{y} + \frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}\boldsymbol{z}
\end{aligned}$$

✎

Theorem F.5.

THM
$$\frac{\partial}{\partial \boldsymbol{x}}\left(\boldsymbol{x}^T A\boldsymbol{x}\right) = \mathbf{A}\boldsymbol{x} + \mathbf{A}^T\boldsymbol{x} + \left[\sum_{i=1}^{n}\left(\frac{\partial}{\partial \boldsymbol{x}}\left[\, a_{1i} \quad a_{2i} \quad \cdots \quad a_{ni}\,\right]\right)x_i\right]\boldsymbol{x} \qquad \forall \boldsymbol{x}\in\mathbb{R}^n, \mathbf{A}\in\mathbb{R}^n\times\mathbb{R}^n$$

✎PROOF:

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{x}}\left(\boldsymbol{x}^T\mathbf{A}\boldsymbol{x}\right) &= \left[\frac{\partial}{\partial \boldsymbol{x}}\boldsymbol{x}\right]\mathbf{A}\boldsymbol{x} + \left[\frac{\partial}{\partial \boldsymbol{x}}\mathbf{A}\boldsymbol{x}\right]\boldsymbol{x} && \text{by Theorem F.4 page 214} \\
&= \mathbf{I}\mathbf{A}\boldsymbol{x} + \left[\mathbf{A}^T + \sum_{i=1}^{n}\left(\frac{\partial}{\partial \boldsymbol{x}}\left[\, a_{1i} \quad a_{2i} \quad \cdots \quad a_{ni}\,\right]\right)x_i\right]\boldsymbol{x} && \text{by Theorem F.1 and Theorem F.2} \\
&= \mathbf{A}\boldsymbol{x} + \mathbf{A}^T\boldsymbol{x} + \left[\sum_{i=1}^{n}\left(\frac{\partial}{\partial \boldsymbol{x}}\left[\, a_{1i} \quad a_{2i} \quad \cdots \quad a_{ni}\,\right]\right)x_i\right]\boldsymbol{x} && \text{by definition of identity operator } \mathbf{I}
\end{aligned}$$

✎

Theorem F.6 (Quadratic form). [7]

THM
$$\mathbf{A} \textit{ is independent of } \boldsymbol{x} \implies \frac{\partial}{\partial \boldsymbol{x}}\left(\boldsymbol{x}^T A\boldsymbol{x}\right) = A\boldsymbol{x} + A^T\boldsymbol{x} \qquad \forall \boldsymbol{x}\in\mathbb{R}^n, \mathbf{A}\in\mathbb{R}^n\times\mathbb{R}^n$$

[6] Scharf (1991) page 274, Trees (2002) page 1398
[7] Graham (1981) page 54

✎PROOF:

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{x}}\left(\boldsymbol{x}^T \mathbf{A}\boldsymbol{x}\right) &= \left[\frac{\partial}{\partial \boldsymbol{x}}\boldsymbol{x}\right]\mathbf{A}\boldsymbol{x} + \left[\frac{\partial}{\partial \boldsymbol{x}}\mathbf{A}\boldsymbol{x}\right]\boldsymbol{x} && \text{by Theorem F.4 page 214}\\ &= \mathbf{I}\mathbf{A}\boldsymbol{x} + \mathbf{A}^T\boldsymbol{x} && \text{by Theorem F.1 page 211 and Theorem F.3 page 213}\end{aligned}$$

✏

Corollary F.1. [8]

COR
$$\frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}^T\boldsymbol{x}) = 2\boldsymbol{x} \qquad \forall \boldsymbol{x} \in \mathbb{R}^n$$

✎PROOF:

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}^T\boldsymbol{x}) &= \frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}^T\mathbf{I}\boldsymbol{x}) && \text{by property of identity operator } \boldsymbol{I}\\ &= \mathbf{I}\boldsymbol{x} + \mathbf{I}^T\boldsymbol{x} && \text{by previous result 3.}\\ &= \boldsymbol{x} + \boldsymbol{x} && \text{by property of identity operator } \boldsymbol{I}\\ &= 2\boldsymbol{x}\end{aligned}$$

✏

Theorem F.7 (Chain rule). [9] *Let* $\boldsymbol{z}$ *be a function of* $\boldsymbol{y}$ *and* $\boldsymbol{y}$ *a function of* $\boldsymbol{x}$ *and*

$$\boldsymbol{x} \triangleq \begin{bmatrix} x_1\\ x_2\\ \vdots\\ x_n\end{bmatrix} \qquad \boldsymbol{y} \triangleq \begin{bmatrix} y_1\\ y_2\\ \vdots\\ y_m\end{bmatrix} \qquad \boldsymbol{z} \triangleq \begin{bmatrix} z_1\\ z_2\\ \vdots\\ z_k\end{bmatrix}$$

THM
$$\frac{\partial}{\partial \boldsymbol{x}}\boldsymbol{z} = \frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}\frac{\partial \boldsymbol{z}}{\partial \boldsymbol{y}}$$

✎PROOF:

$$\begin{aligned}\frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}} &= \begin{bmatrix} \frac{\partial z_1}{\partial x_1} & \frac{\partial z_2}{\partial x_1} & \cdots & \frac{\partial z_k}{\partial x_1}\\ \frac{\partial z_1}{\partial x_2} & \frac{\partial z_2}{\partial x_2} & \cdots & \frac{\partial z_k}{\partial x_2}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{\partial z_1}{\partial x_n} & \frac{\partial z_2}{\partial x_n} & \cdots & \frac{\partial z_k}{\partial x_n}\end{bmatrix}\\
&= \begin{bmatrix} \sum_{j=0}^{m}\frac{\partial z_1}{\partial y_j}\frac{\partial y_j}{\partial x_1} & \sum_{j=0}^{m}\frac{\partial z_2}{\partial y_j}\frac{\partial y_j}{\partial x_1} & \cdots & \sum_{j=0}^{m}\frac{\partial z_k}{\partial y_j}\frac{\partial y_j}{\partial x_1}\\ \sum_{j=0}^{m}\frac{\partial z_1}{\partial y_j}\frac{\partial y_j}{\partial x_2} & \sum_{j=0}^{m}\frac{\partial z_2}{\partial y_j}\frac{\partial y_j}{\partial x_2} & \cdots & \sum_{j=0}^{m}\frac{\partial z_k}{\partial y_j}\frac{\partial y_j}{\partial x_2}\\ \vdots & \vdots & \ddots & \vdots\\ \sum_{j=0}^{m}\frac{\partial z_1}{\partial y_j}\frac{\partial y_j}{\partial x_n} & \sum_{j=0}^{m}\frac{\partial z_2}{\partial y_j}\frac{\partial y_j}{\partial x_n} & \cdots & \sum_{j=0}^{m}\frac{\partial z_k}{\partial y_j}\frac{\partial y_j}{\partial x_n}\end{bmatrix}\\
&= \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_1}\\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_2}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{\partial y_1}{\partial x_n} & \frac{\partial y_2}{\partial x_n} & \cdots & \frac{\partial y_m}{\partial x_n}\end{bmatrix}\begin{bmatrix} \frac{\partial z_1}{\partial y_1} & \frac{\partial z_2}{\partial y_1} & \cdots & \frac{\partial z_k}{\partial y_1}\\ \frac{\partial z_1}{\partial y_2} & \frac{\partial z_2}{\partial y_2} & \cdots & \frac{\partial z_k}{\partial y_2}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{\partial z_1}{\partial y_m} & \frac{\partial z_2}{\partial y_m} & \cdots & \frac{\partial z_k}{\partial y_m}\end{bmatrix}\\
&= \frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}\frac{\partial \boldsymbol{z}}{\partial \boldsymbol{y}}\end{aligned}$$

[8] Graham (1981) page 54
[9] Graham (1981) pages 54–55

F.2 First derivative of a matrix with respect to a scalar

Definition F.3. *Let* $x \in \mathbb{R}$, $\{y_{jk} \in \mathbb{C} | j = 1,2,\ldots,m;\ k = 1,2,\ldots,n\}$ *and*

$$Y = \underbrace{\begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mn} \end{bmatrix}}_{m \times n \text{ matrix}}$$

The ***derivative of*** Y ***with respect to*** x *is*

DEF

$$\frac{\mathrm{d}Y}{\mathrm{d}x} \triangleq \underbrace{\begin{bmatrix} \frac{\mathrm{d}y_{11}}{\mathrm{d}x} & \frac{\mathrm{d}y_{12}}{\mathrm{d}x} & \cdots & \frac{\mathrm{d}y_{1n}}{\mathrm{d}x} \\ \frac{\mathrm{d}y_{21}}{\mathrm{d}x} & \frac{\mathrm{d}y_{22}}{\mathrm{d}x} & \cdots & \frac{\mathrm{d}y_{2n}}{\mathrm{d}x} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\mathrm{d}y_{m1}}{\mathrm{d}x} & \frac{\mathrm{d}y_{m2}}{\mathrm{d}x} & \cdots & \frac{\mathrm{d}y_{mn}}{\mathrm{d}x} \end{bmatrix}}_{m \times n \text{ matrix}}$$

Theorem F.8. [10] *Let* $x \in \mathbb{R}$, $\{y_{jp} \in \mathbb{C} | j = 1,2,\ldots,m;\ p = 1,2,\ldots,n\}$, $\{w_{jp} \in \mathbb{C} | j = 1,2,\ldots,n;\ p = 1,2,\ldots,k\}$, *and*

$$Y = \underbrace{\begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mn} \end{bmatrix}}_{m \times n \text{ matrix}} \qquad W = \underbrace{\begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1k} \\ w_{21} & w_{22} & \cdots & w_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ w_{p1} & w_{p2} & \cdots & w_{pk} \end{bmatrix}}_{p \times k \text{ matrix}}$$

THM

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}x}\left(Y + W\right) &= \frac{\mathrm{d}}{\mathrm{d}x}Y + \frac{\mathrm{d}}{\mathrm{d}x}W && (\text{for } p = m,\ k = n) \\ \frac{\mathrm{d}}{\mathrm{d}x}\left(YW\right) &= \left(\frac{\mathrm{d}}{\mathrm{d}x}Y\right)W + Y\left(\frac{\mathrm{d}}{\mathrm{d}x}W\right) && (\text{for } p = n) \\ \frac{\mathrm{d}}{\mathrm{d}x}\left(Y^T\right) &= \left(\frac{\mathrm{d}}{\mathrm{d}x}Y\right)^T && \\ \frac{\mathrm{d}}{\mathrm{d}x}\left(Y^{-1}\right) &= -Y^{-1}\left(\frac{\mathrm{d}}{\mathrm{d}x}Y\right)Y^{-1} && (\text{for } m = n \text{ and } Y \text{ invertible}) \end{aligned}$$

✎PROOF:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}x}\left(Y + W\right) &= \frac{\mathrm{d}}{\mathrm{d}x}\left(\begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mn} \end{bmatrix} + \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m1} & w_{m2} & \cdots & w_{mn} \end{bmatrix}\right) \\ &= \frac{\mathrm{d}}{\mathrm{d}x}\begin{bmatrix} y_{11} + w_{11} & y_{12} + w_{12} & \cdots & y_{1n} + w_{1n} \\ y_{21} + w_{21} & y_{22} + w_{22} & \cdots & y_{2n} + w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} + w_{m1} & y_{m2} + w_{m2} & \cdots & y_{mn} + w_{mn} \end{bmatrix} \end{aligned}$$

[10] Gradshteyn and Ryzhik (1980) pages 1106–1107

$$
\begin{aligned}
&= \begin{bmatrix} (y_{11}+w_{11})' & (y_{12}+w_{12})' & \cdots & (y_{1n}+w_{1n})' \\ (y_{21}+w_{21})' & (y_{22}+w_{22})' & \cdots & (y_{2n}+w_{2n})' \\ \vdots & \vdots & \ddots & \vdots \\ (y_{m1}+w_{m1})' & (y_{m2}+w_{m2})' & \cdots & (y_{mn}+w_{mn})' \end{bmatrix} \\
&= \begin{bmatrix} y'_{11}+w'_{11} & y'_{12}+w'_{12} & \cdots & y'_{1n}+w'_{1n} \\ y'_{21}+w'_{21} & y'_{22}+w'_{22} & \cdots & y'_{2n}+w'_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y'_{m1}+w'_{m1} & y'_{m2}+w'_{m2} & \cdots & y'_{mn}+w'_{mn} \end{bmatrix} \\
&= \begin{bmatrix} y'_{11} & y'_{12} & \cdots & y'_{1n} \\ y'_{21} & y'_{22} & \cdots & y'_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y'_{m1} & y'_{m2} & \cdots & y'_{mn} \end{bmatrix} + \begin{bmatrix} w'_{11} & w'_{12} & \cdots & w'_{1n} \\ w'_{21} & w'_{22} & \cdots & w'_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w'_{m1} & w'_{m2} & \cdots & w'_{mn} \end{bmatrix} \\
&= \frac{\mathrm{d}}{\mathrm{d}x}\begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mn} \end{bmatrix} + \frac{\mathrm{d}}{\mathrm{d}x}\begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m1} & w_{m2} & \cdots & w_{mn} \end{bmatrix} \\
&= \frac{\mathrm{d}}{\mathrm{d}x}Y + \frac{\mathrm{d}}{\mathrm{d}x}W
\end{aligned}
$$

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}\left(YW\right) &= \frac{\mathrm{d}}{\mathrm{d}x}\left(\begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mn} \end{bmatrix}\begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1k} \\ w_{21} & w_{22} & \cdots & w_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nk} \end{bmatrix}\right) \\
&= \frac{\mathrm{d}}{\mathrm{d}x}\begin{bmatrix} \sum_{j=1}^{n} y_{1j}w_{j1} & \sum_{j=1}^{n} y_{1j}w_{j2} & \cdots & \sum_{j=1}^{n} y_{1j}w_{jk} \\ \sum_{j=1}^{n} y_{2j}w_{j1} & \sum_{j=1}^{n} y_{2j}w_{j2} & \cdots & \sum_{j=1}^{n} y_{2j}w_{jk} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^{n} y_{mj}w_{j1} & \sum_{j=1}^{n} y_{mj}w_{j2} & \cdots & \sum_{j=1}^{n} y_{mj}w_{jk} \end{bmatrix} \\
&= \frac{\mathrm{d}}{\mathrm{d}x}\sum_{j=1}^{n}\begin{bmatrix} y_{1j}w_{j1} & y_{1j}w_{j2} & \cdots & y_{1j}w_{jk} \\ y_{2j}w_{j1} & y_{2j}w_{j2} & \cdots & y_{2j}w_{jk} \\ \vdots & \vdots & \ddots & \vdots \\ y_{mj}w_{j1} & y_{mj}w_{j2} & \cdots & y_{mj}w_{jk} \end{bmatrix} \\
&= \sum_{j=1}^{n}\frac{\mathrm{d}}{\mathrm{d}x}\begin{bmatrix} y_{1j}w_{j1} & y_{1j}w_{j2} & \cdots & y_{1j}w_{jk} \\ y_{2j}w_{j1} & y_{2j}w_{j2} & \cdots & y_{2j}w_{jk} \\ \vdots & \vdots & \ddots & \vdots \\ y_{mj}w_{j1} & y_{mj}w_{j2} & \cdots & y_{mj}w_{jk} \end{bmatrix} \\
&= \sum_{j=1}^{n}\begin{bmatrix} \frac{\mathrm{d}}{\mathrm{d}x}(y_{1j}w_{j1}) & \frac{\mathrm{d}}{\mathrm{d}x}(y_{1j}w_{j2}) & \cdots & \frac{\mathrm{d}}{\mathrm{d}x}(y_{1j}w_{jk}) \\ \frac{\mathrm{d}}{\mathrm{d}x}(y_{2j}w_{j1}) & \frac{\mathrm{d}}{\mathrm{d}x}(y_{2j}w_{j2}) & \cdots & \frac{\mathrm{d}}{\mathrm{d}x}(y_{2j}w_{jk}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\mathrm{d}}{\mathrm{d}x}(y_{mj}w_{j1}) & \frac{\mathrm{d}}{\mathrm{d}x}(y_{mj}w_{j2}) & \cdots & \frac{\mathrm{d}}{\mathrm{d}x}(y_{mj}w_{jk}) \end{bmatrix} \\
&= \sum_{j=1}^{n}\begin{bmatrix} \vdots & \vdots & \ddots & \vdots \\ y'_{1j}w_{j1}+y_{1j}w'_{j1} & y'_{1j}w_{j2}+y_{1j}w'_{j2} & \cdots & y'_{1j}w_{jk}+y_{1j}w'_{jk} \\ y'_{2j}w_{j1}+y_{2j}w'_{j1} & y'_{2j}w_{j2}+y_{2j}w'_{j2} & \cdots & y'_{2j}w_{jk}+y_{2j}w'_{jk} \\ y'_{mj}w_{j1}+y_{mj}w'_{j1} & y'_{mj}w_{j2}+y_{mj}w'_{j2} & \cdots & y'_{mj}w_{jk}+y_{mj}w'_{jk} \end{bmatrix}
\end{aligned}
$$

$$= \sum_{j=1}^{n}\left(\begin{bmatrix} y'_{1j}w_{j1} & y'_{1j}w_{j2} & \cdots & y'_{1j}w_{jk} \\ y'_{2j}w_{j1} & y'_{2j}w_{j2} & \cdots & y'_{2j}w_{jk} \\ \vdots & \vdots & \ddots & \vdots \\ y'_{mj}w_{j1} & y'_{mj}w_{j2} & \cdots & y'_{mj}w_{jk} \end{bmatrix} + \begin{bmatrix} y_{1j}w'_{j1} & y_{1j}w'_{j2} & \cdots & y_{1j}w'_{jk} \\ y_{2j}w'_{j1} & y_{2j}w'_{j2} & \cdots & y_{2j}w'_{jk} \\ \vdots & \vdots & \ddots & \vdots \\ y_{mj}w'_{j1} & y_{mj}w'_{j2} & \cdots & y_{mj}w'_{jk} \end{bmatrix}\right)$$

$$= \left(\frac{\mathrm{d}}{\mathrm{d}x}Y\right)W + Y\left(\frac{\mathrm{d}}{\mathrm{d}x}W\right)$$

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}\left(Y^T\right) &= \frac{\mathrm{d}}{\mathrm{d}x}\left(\begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mn} \end{bmatrix}^T\right) \\
&= \frac{\mathrm{d}}{\mathrm{d}x}\begin{bmatrix} y_{11} & y_{21} & \cdots & y_{n1} \\ y_{12} & y_{22} & \cdots & y_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ y_{1n} & y_{2n} & \cdots & y_{nm} \end{bmatrix} \\
&= \begin{bmatrix} y'_{11} & y'_{21} & \cdots & y'_{n1} \\ y'_{12} & y'_{22} & \cdots & y'_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ y'_{1n} & y'_{2n} & \cdots & y'_{nm} \end{bmatrix} \\
&= \begin{bmatrix} y'_{11} & y'_{12} & \cdots & y'_{1n} \\ y'_{21} & y'_{22} & \cdots & y'_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y'_{m1} & y'_{m2} & \cdots & y'_{mn} \end{bmatrix}^T \\
&= \left(\frac{\mathrm{d}}{\mathrm{d}x}\begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mn} \end{bmatrix}\right)^T
\end{aligned}$$

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}\left(Y^{-1}\right) &= \frac{\mathrm{d}}{\mathrm{d}x}\frac{\operatorname{adj}Y}{|Y|} \\
&\vdots \\
&\text{no proof at this time} \\
&\vdots \\
&= -Y^{-1}\left(\frac{\mathrm{d}}{\mathrm{d}x}Y\right)Y^{-1}
\end{aligned}$$

✏

F.3 Second derivative of a scalar with respect to a vector

Definition F.4. [11] *Let*

$$\boldsymbol{x} \triangleq \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

[11] Lieb and Loss (2001) page 240, Horn and Johnson (1990) page 167

*The **Hessian matrix** of a scalar y with respect to the vector $\boldsymbol{x}$ is*

DEF

$$\frac{\partial^2 y}{\partial \boldsymbol{x}^2} = \frac{\partial}{\partial \boldsymbol{x}}\left(\frac{\partial y}{\partial \boldsymbol{x}}\right) = \frac{\partial}{\partial \boldsymbol{x}}\begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{\partial}{\partial x_1}\frac{\partial y}{\partial x_1} & \frac{\partial}{\partial x_1}\frac{\partial y}{\partial x_2} & \cdots & \frac{\partial}{\partial x_1}\frac{\partial y}{\partial x_n} \\ \frac{\partial}{\partial x_2}\frac{\partial y}{\partial x_1} & \frac{\partial}{\partial x_2}\frac{\partial y}{\partial x_2} & \cdots & \frac{\partial}{\partial x_2}\frac{\partial y}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_n}\frac{\partial y}{\partial x_1} & \frac{\partial}{\partial x_n}\frac{\partial y}{\partial x_2} & \cdots & \frac{\partial}{\partial x_n}\frac{\partial y}{\partial x_n} \end{bmatrix}}_{n \times n \text{ matrix}}$$

F.4 Multiple derivatives of a vector with respect to a scalar

Definition F.5. *Let*

$$\boldsymbol{y} \triangleq \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

The derivative of a vector $\boldsymbol{y}$ with respect to the scalar x is

DEF

$$\begin{bmatrix} \boldsymbol{y} \\ \frac{\mathrm{d}}{\mathrm{d}x}\boldsymbol{y} \\ \frac{\mathrm{d}^2}{\mathrm{d}x^2}\boldsymbol{y} \\ \vdots \\ \frac{\mathrm{d}^n}{\mathrm{d}x^n}\boldsymbol{y} \end{bmatrix} = \underbrace{\begin{bmatrix} y_1 & y_2 & \cdots & y_m \\ \frac{\mathrm{d}}{\mathrm{d}x}y_1 & \frac{\mathrm{d}}{\mathrm{d}x}y_2 & \cdots & \frac{\mathrm{d}}{\mathrm{d}x}y_m \\ \frac{\mathrm{d}^2}{\mathrm{d}x^2}y_1 & \frac{\mathrm{d}^2}{\mathrm{d}x^2}y_2 & \cdots & \frac{\mathrm{d}^2}{\mathrm{d}x^2}y_m \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\mathrm{d}^n}{\mathrm{d}x^n}y_1 & \frac{\mathrm{d}^n}{\mathrm{d}x^n}y_2 & \cdots & \frac{\mathrm{d}^n}{\mathrm{d}x^n}y_m \end{bmatrix}}_{(n+1) \times m \text{ matrix}}$$

APPENDIX G

ALGEBRAIC STRUCTURES

❝*In this book, learned reader, you have the rules of algebra. . . It unties the knot not only where one term is equal to another or two to one but also where two are equal to two or three to one.. . . this most abstruse and unsurpassed treasury of the entire arithmetic being brought to light and, as in a theater, exposed to the sight of all. . .*❞

Gerolamo Cardano (1501–1576), Italian mathematician, physician, and astrologer [1]

A set together with one or more operations forms several standard mathematical structures:

group $\supseteq$ *ring* $\supseteq$ *commutative ring* $\supseteq$ *integral domain* $\supseteq$ *field*

Definition G.1. [2] *Let X be a set and $\diamond : X \times X \to X$ be an operation on X.*

DEF

*The pair $(X, \diamond)$ is a **group** if*

1. $\exists e \in X$ such that $e\diamond x = x\diamond e = x$ $\quad \forall x{\in}X$ *(IDENTITY element)* and
2. $\exists(-x) \in X$ such that $(-x)\diamond x = x\diamond(-x) = e$ $\quad \forall x{\in}X$ *(INVERSE element)* and
3. $x\diamond(y\diamond z) = (x\diamond y)\diamond z$ $\quad \forall x,y,z{\in}X$ *(ASSOCIATIVE)*

Definition G.2. [3] *Let $+ : X \times X \to X$ and $\cdot : X \times X \to X$ be operations on a set X. Furthermore, let the operation $\cdot$ also be represented by juxtapostion as in $a \cdot b \equiv ab$.*

DEF

*The triple $(X, +, \cdot)$ is a **ring** if*

1. $(X, +)$ *is a group.* *(additive group)* and
2. $x(yz) = (xy)z$ $\quad \forall x,y,z{\in}X$ *(ASSOCIATIVE with respect to $\cdot$)* and
3. $x(y+z) = (xy)+(xz)$ $\quad \forall x,y,z{\in}X$ *($\cdot$ is LEFT DISTRIBUTIVE over $+$)* and
4. $(x+y)z = (xz)+(yz)$ $\quad \forall x,y,z{\in}X$ *($\cdot$ is RIGHT DISTRIBUTIVE over $+$).*

Definition G.3. [4]

[1] quote: Cardano (1545) page 1
image: http://en.wikipedia.org/wiki/Image:Cardano.jpg

[2] Durbin (2000) page 29

[3] Durbin (2000) pages 114–115

[4] Durbin (2000) page 118

DEF *A triple $(X, +, \cdot)$ is a **commutative ring** if*

1. $(X, +, \cdot)$ *is a* RING *and*
2. $xy = yx \quad \forall x,y \in X$ (COMMUTATIVE).

Definition G.4. [5] *Let R be a* COMMUTATIVE RING (Definition G.3 page 221).

DEF *A function $|\cdot|$ in $\mathbb{R}^{\mathbb{R}}$ is an **absolute value** (or **modulus**) if*

1. $\lvert x\rvert \geq 0$	$x\in\mathbb{R}$	(NON-NEGATIVE)	*and*
2. $\lvert x\rvert = 0 \iff x = 0$	$x\in\mathbb{R}$	(NONDEGENERATE)	*and*
3. $\lvert xy\rvert = \lvert x\rvert \cdot \lvert y\rvert$	$x,y\in\mathbb{R}$	(HOMOGENEOUS / SUBMULTIPLICATIVE)	*and*
4. $\lvert x+y\rvert \leq \lvert x\rvert + \lvert y\rvert$	$x,y\in\mathbb{R}$	(SUBADDITIVE / TRIANGLE INEQUALITY)	

Definition G.5. [6]

DEF *The structure $F \triangleq (X, +, \cdot, 0, 1)$ is a **field** if*

1. $(X, +, \cdot)$ *is a ring*		(ring)	*and*
2. $xy = yx$	$\forall x,y\in X$	(commutative with respect to $\cdot$)	*and*
3. $(X\backslash\{0\}, \cdot)$ *is a group*		(group with respect to $\cdot$).	

Definition G.6. [7] *Let $\boldsymbol{V} = (F, +, \cdot)$ be a* VECTOR SPACE *and $\otimes : \boldsymbol{V}\times\boldsymbol{V} \to \boldsymbol{V}$ be a vector-vector multiplication operator.*
*An **algebra** is any pair $(\boldsymbol{V}, \otimes)$ that satisfies ($\otimes$ is represented by juxtaposition)*

DEF

1. $(\boldsymbol{ux})\boldsymbol{y} = \boldsymbol{u}(\boldsymbol{xy})$	$\forall \boldsymbol{u},\boldsymbol{x},\boldsymbol{y}\in\boldsymbol{V}$	(ASSOCIATIVE)	*and*
2. $\boldsymbol{u}(\boldsymbol{x}+\boldsymbol{y}) = (\boldsymbol{ux}) + (\boldsymbol{uy})$	$\forall \boldsymbol{u},\boldsymbol{x},\boldsymbol{y}\in\boldsymbol{V}$	(LEFT DISTRIBUTIVE)	*and*
3. $(\boldsymbol{u}+\boldsymbol{x})\boldsymbol{y} = (\boldsymbol{uy}) + (\boldsymbol{xy})$	$\forall \boldsymbol{u},\boldsymbol{x},\boldsymbol{y}\in\boldsymbol{V}$	(RIGHT DISTRIBUTIVE)	*and*
4. $\alpha(\boldsymbol{xy}) = (\alpha\boldsymbol{x})\boldsymbol{y} = \boldsymbol{x}(\alpha\boldsymbol{y})$	$\forall \boldsymbol{x},\boldsymbol{y}\in\boldsymbol{V}$ *and* $\alpha\in F$	(SCALAR COMMUTATIVE)	.

[5] Cohn (2002) page 312
[6] Durbin (2000) page 123, Weber (1893)
[7] Abramovich and Aliprantis (2002) page 3, Michel and Herget (1993) page 56

APPENDIX H

CALCULUS

Definition H.1. *Let X and Y be sets.*

DEF *The space Y^X represents the set of all functions with* DOMAIN *X and* RANGE *Y such that*

$$Y^X \triangleq \{\mathsf{f}(x) | \mathsf{f}(x) : X \to Y\}$$

Definition H.2. *Let $\mathbb{R}$ be the set of real numbers, $\mathscr{B}$ the set of* BOREL SETS *on $\mathbb{R}$, and μ the standard* BOREL MEASURE *on $\mathscr{B}$. Let $\mathbb{R}^{\mathbb{R}}$ be as in Definition H.1 page 223.*

DEF *The **space of Lebesgue square-integrable functions** $\boldsymbol{L}^2_{(\mathbb{R},\mathscr{B},\mu)}$ (or $\boldsymbol{L}^2_{\mathbb{R}}$) is defined as*

$$\boldsymbol{L}^2_{\mathbb{R}} \triangleq \boldsymbol{L}^2_{(\mathbb{R},\mathscr{B},\mu)} \triangleq \left\{\mathsf{f} \in \mathbb{R}^{\mathbb{R}} | \left(\int_{\mathbb{R}} |\mathsf{f}|^2\right)^{\frac{1}{2}} \mathsf{d}\mu < \infty\right\}.$$

*The **standard inner product** $\langle \triangle \mid \nabla \rangle$ on $\boldsymbol{L}^2_{\mathbb{R}}$ is defined as*

$$\langle \mathsf{f}(x) \mid \mathsf{g}(x) \rangle \triangleq \int_{\mathbb{R}} \mathsf{f}(x)\mathsf{g}^*(x)\,\mathsf{d}x.$$

*The **standard norm** $\|\cdot\|$ on $\boldsymbol{L}^2_{\mathbb{R}}$ is defined as $\|\mathsf{f}(x)\| \triangleq \langle \mathsf{f}(x) \mid \mathsf{f}(x) \rangle^{\frac{1}{2}}$*

Definition H.3. *Let $\mathsf{f}(x)$ be a* FUNCTION *in $\mathbb{R}^{\mathbb{R}}$.*

DEF

$$\frac{\mathsf{d}}{\mathsf{d}x}\mathsf{f}(x) \triangleq \mathsf{f}'(x) \triangleq \lim_{\varepsilon \to 0} \frac{\mathsf{f}(x+\varepsilon) - \mathsf{f}(x)}{\varepsilon}$$

Proposition H.1.

PRP

$$\left\{\begin{array}{lll} (1). & \mathsf{f}(x) \text{ is CONTINUOUS} & \textit{and} \\ (2). & \underbrace{\mathsf{f}(a+x) = \mathsf{f}(a-x)}_{\text{SYMMETRIC about a point } a} & \end{array}\right\} \implies \left\{\begin{array}{lllll} (1). & \mathsf{f}'(a+x) & = & -\mathsf{f}'(a-x) & (\text{ANTI-SYMMETRIC about } a) \\ (2). & \mathsf{f}'(a) & = & 0 & \end{array}\right\}$$

✎PROOF:

$$\begin{aligned}
\mathsf{f}'(a+x) &= \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}[\mathsf{f}(a+x+\varepsilon) - \mathsf{f}(a+x-\varepsilon)] \\
&= \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}[\mathsf{f}(a-x-\varepsilon) - \mathsf{f}(a-x+\varepsilon)] && \text{by hpothesis (2)} \\
&= -\lim_{\varepsilon\to 0} \frac{1}{\varepsilon}[\mathsf{f}(a-x+\varepsilon) - \mathsf{f}(a-x-\varepsilon)] \\
&= -\mathsf{f}(a-x)
\end{aligned}$$

$$\begin{aligned}
\mathsf{f}'(a) &= \frac{1}{2}\mathsf{f}'(a+0) + \frac{1}{2}\mathsf{f}'(a-0) \\
&= \frac{1}{2}\left[\mathsf{f}'(a+0) - \mathsf{f}'(a+0)\right] && \text{by previous result} \\
&= 0
\end{aligned}$$

✎

Lemma H.1.

LEM: $\mathsf{f}(x)$ *is* INVERTIBLE $\implies \left\{ \frac{\mathsf{d}}{\mathsf{dy}}\mathsf{f}^{-1}(y) = \frac{1}{\frac{\mathsf{d}}{\mathsf{dx}}\mathsf{f}\left[\mathsf{f}^{-1}(y)\right]} \right\}$

✎PROOF:

$$\begin{aligned}
\frac{\mathsf{d}}{\mathsf{dy}}\mathsf{f}^{-1}(y) &\triangleq \lim_{\varepsilon\to 0}\frac{\mathsf{f}^{-1}(y+\varepsilon) - \mathsf{f}^{-1}(y)}{\varepsilon} && \text{by definition of } \frac{\mathsf{d}}{\mathsf{dy}} && \text{(Definition H.3 page 223)} \\
&= \lim_{\delta\to 0}\frac{1}{\left[\frac{\mathsf{f}(x+\delta)-\mathsf{f}(x)}{\delta}\right]}\Bigg|_{x\triangleq \mathsf{f}^{-1}(y)} && \text{because in the limit, } \frac{\Delta y}{\Delta x} = \left(\frac{\Delta x}{\Delta y}\right)^{-1} \\
&\triangleq \frac{1}{\frac{\mathsf{d}}{\mathsf{dx}}\mathsf{f}(x)}\Bigg|_{x\triangleq \mathsf{f}^{-1}(y)} && \text{by definition of } \frac{\mathsf{d}}{\mathsf{dx}} && \text{(Definition H.3 page 223)} \\
&= \frac{1}{\frac{\mathsf{d}}{\mathsf{dx}}\mathsf{f}\left[\mathsf{f}^{-1}(y)\right]} && \text{because } x \triangleq \mathsf{f}^{-1}(y)
\end{aligned}$$

✎

Theorem H.1. [1] *Let* f *be a continuous function in* $\boldsymbol{L}^2_{\mathbb{R}}$ *and* $\mathsf{f}^{(n)}$ *the* n*th derivative of* f.

THM: $$\int_{[0:1)^n} \mathsf{f}^{(n)}\left(\sum_{k=1}^{n} x_k\right) \mathsf{dx}_1\, \mathsf{dx}_2 \cdots \mathsf{dx}_n = \sum_{k=0}^{n}(-1)^{n-k}\binom{n}{k}\mathsf{f}(k) \qquad \forall n \in \mathbb{N}$$

✎PROOF: Proof by induction:

1. Base case ...proof for $n = 1$ case:

$$\begin{aligned}
\int_{[0:1)} \mathsf{f}^{(1)}(x)\,\mathsf{dx} &= \mathsf{f}(1) - \mathsf{f}(0) && \text{by } \textit{Fundamental theorem of calculus} \\
&= (-1)^{1+1}\binom{1}{1}\mathsf{f}(1) + (-1)^{1+0}\binom{1}{0}\mathsf{f}(0) \\
&= \sum_{k=0}^{1}(-1)^{n-k}\binom{n}{k}\mathsf{f}(k)
\end{aligned}$$

[1] Chui (1992) page 86 ⟨item (ii)⟩, Prasad and Iyengar (1997) pages 145–146 ⟨Theorem 6.2 (b)⟩

2. Induction step ...proof that n case $\implies$ $n+1$ case:

$$\begin{aligned}
&\int_{[0:1)^{n+1}} \mathsf{f}^{(n+1)}\left(\sum_{k=1}^{n+1} x_k\right) \mathrm{d}x_1\, \mathrm{d}x_2 \cdots \mathrm{d}x_{n+1} \\
&= \int_{[0:1)^n} \left[\int_0^1 \mathsf{f}^{(n+1)}\left(x_{n+1} + \sum_{k=1}^{n} x_k\right) \mathrm{d}x_{n+1}\right] \mathrm{d}x_1\, \mathrm{d}x_2 \cdots \mathrm{d}x_n \\
&= \int_{[0:1)^n} \left[\mathsf{f}^{(n)}\left(x_{n+1} + \sum_{k=1}^{n} x_k\right)\Bigg|_{x_{n+1}=0}^{x_{n+1}=1}\right] \mathrm{d}x_1\, \mathrm{d}x_2 \cdots \mathrm{d}x_n && \text{by \textit{Fundamental theorem of calculus}} \\
&= \int_{[0:1)^n} \left[\mathsf{f}^{(n)}\left(1 + \sum_{k=1}^{n} x_k\right) - \mathsf{f}^{(n)}\left(0 + \sum_{k=1}^{n} x_k\right)\right] \mathrm{d}x_1\, \mathrm{d}x_2 \cdots \mathrm{d}x_n \\
&= \sum_{k=0}^{n}(-1)^{n-k}\binom{n}{k}\mathsf{f}(k+1) - \sum_{k=0}^{n}(-1)^{n-k}\binom{n}{k}\mathsf{f}(k) && \text{by induction hypothesis} \\
&= \sum_{m=1}^{m=n+1}(-1)^{n-m+1}\binom{n}{m-1}\mathsf{f}(m) + \sum_{k=0}^{n}(-1)(-1)^{n-k}\binom{n}{k}\mathsf{f}(k) && \text{where } m \triangleq k+1 \implies k = m-1 \\
&= \left[\mathsf{f}(n+1) + \sum_{k=1}^{n}(-1)^{n-k+1}\binom{n}{k-1}\mathsf{f}(k)\right] + \left[(-1)^{n+1}\mathsf{f}(0) + \sum_{k=1}^{n}(-1)^{n-k+1}\binom{n}{k}\mathsf{f}(k)\right] \\
&= \mathsf{f}(n+1) + (-1)^{n+1}\mathsf{f}(0) + \sum_{k=1}^{n}(-1)^{n-k+1}\underbrace{\left[\binom{n}{k-1} + \binom{n}{k}\right]}_{\text{use \textit{Stifel formula}}}\mathsf{f}(k) \\
&= (-1)^0\binom{n+1}{n+1}\mathsf{f}(n+1) + (-1)^{n+1}\binom{n+1}{0}\mathsf{f}(0) + \sum_{k=1}^{n}(-1)^{n-k+1}\binom{n+1}{k}\mathsf{f}(k) && \text{by \textit{Stifel formula}} \\
&= \sum_{k=0}^{n+1}(-1)^{n-k+1}\binom{n+1}{k}\mathsf{f}(k)
\end{aligned}$$

✎

Some proofs invoke differentiation multiple times. This is simplified thanks to the *Leibniz rule*, also called the *generalized product rule* (*GPR*, next lemma). The Leibniz rule is remarkably similar in form to the *binomial theorem*.

Lemma H.2 (Leibniz rule / generalized product rule). [2] *Let* $\mathsf{f}(x), \mathsf{g}(x) \in \boldsymbol{L}^2_{\mathbb{R}}$ *with derivatives* $\mathsf{f}^{(n)}(x) \triangleq \frac{\mathrm{d}^n}{\mathrm{d}x^n}\mathsf{f}(x)$ *and* $\mathsf{g}^{(n)}(x) \triangleq \frac{\mathrm{d}^n}{\mathrm{d}x^n}\mathsf{g}(x)$ *for* $n = 0, 1, 2, \ldots,$ *and* $\binom{n}{k} \triangleq \frac{n!}{(n-k)!k!}$ *(binomial coefficient). Then*

LEM
$$\boxed{\frac{\mathrm{d}^n}{\mathrm{d}x^n}[\mathsf{f}(x)\mathsf{g}(x)] = \sum_{k=0}^{n}\binom{n}{k}\mathsf{f}^{(k)}(x)\mathsf{g}^{(n-k)}(x)}$$

Example H.1.

EX
$$\boxed{\frac{\mathrm{d}^3}{\mathrm{d}x^3}\left[\mathsf{f}(x)\mathsf{g}(x)\right] = \mathsf{f}'''(x)\mathsf{g}(x) + 3\mathsf{f}''(x)\mathsf{g}'(x) + 3\mathsf{f}'(x)\mathsf{g}''(x) + \mathsf{f}(x)\mathsf{g}'''(x)}$$

Theorem H.2 (Leibniz integration rule). [3]

[2] Ben-Israel and Gilbert (2002) page 154, Leibniz (1710)
[3] Flanders (1973) page 615 ⟨(1.1)⟩ Talvila (2001), Knapp (2005b) page 389 ⟨Chapter VII⟩, Protter and Morrey (2012) page 422 ⟨Leibniz Rule. Theorem 1.⟩, http://planetmath.org/encyclopedia/DifferentiationUnderIntegralSign.html

THM
$$\frac{\mathrm{d}}{\mathrm{d}x}\int_{\mathrm{a}(x)}^{\mathrm{b}(x)} \mathrm{g}(t)\,\mathrm{d}t = \mathrm{g}[\mathrm{b}(x)]\mathrm{b}'(x) - \mathrm{g}[\mathrm{a}(x)]\mathrm{a}'(x)$$

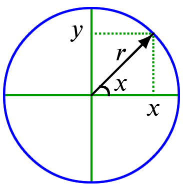
APPENDIX I

TRIGONOMETRIC FUNCTIONS

I.1 Definition Candidates

There are several ways of defining the sine and cosine functions, including the following:[1]

1. **Planar geometry:** Trigonometric functions are traditionally introduced as they have come to us historically—that is, as related to the parameters of triangles.[2]



$$\cos x \triangleq \frac{x}{r} \qquad \sin x \triangleq \frac{y}{r}$$

2. **Complex exponential:** The cosine and sine functions are the real and imaginary parts of the complex exponential such that[3]

$$\cos x \triangleq \mathbf{R}_\mathsf{e} e^{ix} \qquad \sin x \triangleq \mathbf{I}_\mathsf{m}\left(e^{ix}\right)$$

3. **Polynomial:** Let $\sum_{n=0}^{\infty} x_n \triangleq \lim_{N\to\infty} \sum_{n=0}^{N} x_n$ in some topological space. The sine and cosine functions can be defined in terms of *Taylor expansions* such that[4]

$$\begin{aligned}
\cos(x) &\triangleq \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n} &&= 1 - \frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots \\
\sin(x) &\triangleq \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1} &&= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots
\end{aligned}$$

[1] The term *sine* originally came from the Hindu word *jiva* and later adapted to the Arabic word *jiba*. Abrabic-Latin translator Robert of Chester apparently confused this word with the Arabic word *jaib*, which means "bay" or "inlet"—thus resulting in the Latin translation *sinus*, which also means "bay" or "inlet". Reference: Boyer and Merzbach (1991) page 252

[2] Abramowitz and Stegun (1972) page 78

[3] Euler (1748)

[4] Rosenlicht (1968) page 157, Abramowitz and Stegun (1972) page 74

4. **Product of factors:** Let $\prod_{n=0}^{\infty} x_n \triangleq \lim_{N\to\infty} \prod_{n=0}^{N} x_n$ in some topological space. The sine and cosine functions can be defined in terms of a product of factors such that[5]

$$\cos(x) \triangleq \prod_{n=1}^{\infty}\left[1-\left(\frac{x}{(2n-1)\frac{\pi}{2}}\right)^2\right] \qquad \sin(x) \triangleq x\prod_{n=1}^{\infty}\left[1-\left(\frac{x}{n\pi}\right)^2\right]$$

5. **Partial fraction expansion:** The sine function can be defined in terms of a partial fraction expansion such that[6]

$$\sin(x) \triangleq \frac{1}{\frac{1}{x}+2x\sum_{n=1}^{\infty}\frac{(-1)^n}{x^2-(n\pi)^2}} \qquad \cos(x) \triangleq \underbrace{\left(\frac{1}{x}+2x\sum_{n=1}^{\infty}\frac{1}{x^2-(n\pi)^2}\right)}_{\cot(x)}\sin(x)$$

6. **Differential operator:** The sine and cosine functions can be defined as solutions to differential equations expressed in terms of the differential operator $\frac{\mathrm{d}}{\mathrm{d}x}$ such that

$$\begin{array}{llllll}
\cos(x) & \triangleq & \mathsf{f}(x) & \text{where} & \underbrace{\frac{\mathrm{d}^2}{\mathrm{d}x^2}\mathsf{f}+\mathsf{f}=0}_{\text{differential equation}} & \underbrace{\mathsf{f}(0)=1}_{\text{1st initial condition}} & \underbrace{\left[\frac{\mathrm{d}}{\mathrm{d}x}\mathsf{f}\right](0)=0}_{\text{2nd initial condition}} \\
\sin(x) & \triangleq & \mathsf{g}(x) & \text{where} & \underbrace{\frac{\mathrm{d}^2}{\mathrm{d}x^2}\mathsf{g}+\mathsf{g}=0}_{\text{differential equation}} & \underbrace{\mathsf{g}(0)=0}_{\text{1st initial condition}} & \underbrace{\left[\frac{\mathrm{d}}{\mathrm{d}x}\mathsf{g}\right](0)=1}_{\text{2nd initial condition}}
\end{array}$$

7. **Integral operator:** The sine and cosine functions can be defined as inverses of integrals of square roots of rational functions such that[7]

$$\begin{array}{lllll}
\cos(x) & \triangleq & \mathsf{f}^{-1}(x) & \text{where} & \mathsf{f}(x) \triangleq \underbrace{\int_x^1 \sqrt{\frac{1}{1-y^2}}\,\mathrm{d}y}_{\arccos(x)} \\
\sin(x) & \triangleq & \mathsf{g}^{-1}(x) & \text{where} & \mathsf{g}(x) \triangleq \underbrace{\int_0^x \sqrt{\frac{1}{1-y^2}}\,\mathrm{d}y}_{\arcsin(x)}
\end{array}$$

For purposes of analysis, it can be argued that the more natural approach for defining harmonic functions is in terms of the differentiation operator $\frac{\mathrm{d}}{\mathrm{d}x}$ (Definition I.1 page 229). Support for such an approach includes the following:

- Both sine and cosine are very easily represented analytically as polynomials with coefficients involving the operator $\frac{\mathrm{d}}{\mathrm{d}x}$ (Theorem I.1 page 230).
- All solutions of homogeneous second order differential equations are linear combinations of sine and cosine (Theorem I.3 page 232).
- Sine and cosine themselves are related to each other in terms of the differentiation operator (Theorem I.4 page 233).

[5] Abramowitz and Stegun (1972) page 75
[6] Abramowitz and Stegun (1972) page 75
[7] Abramowitz and Stegun (1972) page 79

- The complex exponential function is a solution of a second order homogeneous differential equation (Definition I.4 page 234).
- Sine and cosine are orthogonal with respect to an innerproduct generated by an integral operator—which is a kind of inverse differential operator (Section I.6 page 242).

I.2 Definitions

Definition I.1. [8] *Let $\boldsymbol{C}$ be the* SPACE OF ALL CONTINUOUSLY DIFFERENTIABLE REAL FUNCTIONS *and $\frac{\mathsf{d}}{\mathsf{d}x} \in \boldsymbol{C}^{\boldsymbol{C}}$ the differentiation operator.*

DEF: *The function $\mathsf{f} \in \mathbb{C}^{\mathbb{C}}$ is the **cosine** function $\cos(x) \triangleq \mathsf{f}(x)$ if*

$$\begin{array}{rrcll}
1. & \frac{\mathsf{d}^2}{\mathsf{d}x^2}\mathsf{f} + \mathsf{f} &=& 0 & \text{(second order homogeneous differential equation)} \quad \text{and} \\
2. & \mathsf{f}(0) &=& 1 & \text{(first initial condition)} \quad \text{and} \\
3. & \left[\frac{\mathsf{d}}{\mathsf{d}x}\mathsf{f}\right](0) &=& 0 & \text{(second initial condition).}
\end{array}$$

Definition I.2. [9] *Let $\boldsymbol{C}$ and $\frac{\mathsf{d}}{\mathsf{d}x} \in \boldsymbol{C}^{\boldsymbol{C}}$ be defined as in definition of $\cos(x)$ (Definition I.1 page 229).*

DEF: *The function $\mathsf{f} \in \mathbb{C}^{\mathbb{C}}$ is the **sine** function $\sin(x) \triangleq \mathsf{f}(x)$ if*

$$\begin{array}{rrcll}
1. & \frac{\mathsf{d}^2}{\mathsf{d}x^2}\mathsf{f} + \mathsf{f} &=& 0 & \text{(second order homogeneous differential equation)} \quad \text{and} \\
2. & \mathsf{f}(0) &=& 0 & \text{(first initial condition)} \quad \text{and} \\
3. & \left[\frac{\mathsf{d}}{\mathsf{d}x}\mathsf{f}\right](0) &=& 1 & \text{(second initial condition).}
\end{array}$$

Definition I.3. [10]

DEF: *Let π ("pi") be defined as the element in $\mathbb{R}$ such that*

$$\begin{array}{rrcll}
(1). & \cos\left(\frac{\pi}{2}\right) &=& 0 & \text{and} \\
(2). & \pi &>& 0 & \text{and}
\end{array}$$

(3). *π is the **smallest** of all elements in $\mathbb{R}$ that satisfies (1) and (2).*

I.3 Basic properties

Lemma I.1. [11] *Let $\boldsymbol{C}$ be the* SPACE OF ALL CONTINUOUSLY DIFFERENTIABLE REAL FUNCTIONS *and $\frac{\mathsf{d}}{\mathsf{d}x} \in \boldsymbol{C}^{\boldsymbol{C}}$ the differentiation operator.*

LEM:

$$\left\{\frac{\mathsf{d}^2}{\mathsf{d}x^2}\mathsf{f} + \mathsf{f} = 0\right\} \iff$$

$$\left\{\begin{aligned}
\mathsf{f}(x) &= \underbrace{[\mathsf{f}](0)\sum_{n=0}^{\infty}(-1)^n\frac{x^{2n}}{(2n)!}}_{\text{even terms}} + \underbrace{\left[\frac{\mathsf{d}}{\mathsf{d}x}\mathsf{f}\right](0)\sum_{n=0}^{\infty}(-1)^n\frac{x^{2n+1}}{(2n+1)!}}_{\text{odd terms}} \\
&= \left(\mathsf{f}(0) + \left[\frac{\mathsf{d}}{\mathsf{d}x}\mathsf{f}\right](0)x\right) - \left(\frac{\mathsf{f}(0)}{2!}x^2 + \frac{\left[\frac{\mathsf{d}}{\mathsf{d}x}\mathsf{f}\right](0)}{3!}x^3\right) + \left(\frac{\mathsf{f}(0)}{4!}x^4 + \frac{\left[\frac{\mathsf{d}}{\mathsf{d}x}\mathsf{f}\right](0)}{5!}x^5\right)\cdots
\end{aligned}\right\}$$

[8] Rosenlicht (1968) page 157, Flanigan (1983) pages 228–229
[9] Rosenlicht (1968) page 157, Flanigan (1983) pages 228–229
[10] Rosenlicht (1968) page 158
[11] Rosenlicht (1968) page 156, Liouville (1839)

✎PROOF: Let $\mathsf{f}'(x) \triangleq \frac{d}{dx}\mathsf{f}(x)$.

$$\mathsf{f}'''(x) = -\left[\frac{d}{dx}\mathsf{f}\right](x)$$

$$\mathsf{f}^{(4)}(x) = -\left[\frac{d}{dx}\mathsf{f}\right](x) \qquad = -\left[\frac{d^2}{dx^2}\mathsf{f}\right](x) = \mathsf{f}(x)$$

1. Proof that $\left[\frac{d^2}{dx^2}\mathsf{f}\right](x) + \mathsf{f}(x) = 0 \implies \mathsf{f}(x) = \sum_{n=0}^{\infty}(-1)^n\left[\frac{\mathsf{f}(0)}{(2n)!}x^{2n} + \frac{\left[\frac{d}{dx}\mathsf{f}\right](0)}{(2n+1)!}x^{2n+1}\right]$:

$$\begin{aligned}
\mathsf{f}(x) &= \sum_{n=0}^{\infty} \frac{\mathsf{f}^{(n)}(0)}{n!}x^n && \text{by Taylor expansion}\\
&= \mathsf{f}(0) + \left[\frac{d}{dx}\mathsf{f}\right](0)x - \frac{\left[\frac{d^2}{dx^2}\mathsf{f}\right](0)}{2!}x^2 - \frac{\mathsf{f}^3(0)}{3!}x^3 + \frac{\mathsf{f}^4(0)}{4!}x^4 + \frac{\mathsf{f}^5(0)}{5!}x^5 - \cdots\\
&= \mathsf{f}(0) + \left[\frac{d}{dx}\mathsf{f}\right](0)x - \frac{\mathsf{f}(0)}{2!}x^2 - \frac{\left[\frac{d}{dx}\mathsf{f}\right](0)}{3!}x^3 + \frac{\mathsf{f}(0)}{4!}x^4 + \frac{\left[\frac{d}{dx}\mathsf{f}\right](0)}{5!}x^5 - \cdots\\
&= \mathsf{f}(x) = \sum_{n=0}^{\infty}(-1)^n\left[\frac{\mathsf{f}(0)}{(2n)!}x^{2n} + \frac{\left[\frac{d}{dx}\mathsf{f}\right](0)}{(2n+1)!}x^{2n+1}\right]
\end{aligned}$$

2. Proof that $\left[\frac{d^2}{dx^2}\mathsf{f}\right](x) + \mathsf{f}(x) = 0 \impliedby \mathsf{f}(x) = \sum_{n=0}^{\infty}(-1)^n\left[\frac{\mathsf{f}(0)}{(2n)!}x^{2n} + \frac{\left[\frac{d}{dx}\mathsf{f}\right](0)}{(2n+1)!}x^{2n+1}\right]$:

$$\begin{aligned}
\left[\frac{d^2}{dx^2}\mathsf{f}\right](x) &= \frac{d}{dx}\frac{d}{dx}[\mathsf{f}(x)]\\
&= \frac{d}{dx}\frac{d}{dx}\sum_{n=0}^{\infty}(-1)^n\left[\frac{\mathsf{f}(0)}{(2n)!}x^{2n} + \frac{\left[\frac{d}{dx}\mathsf{f}\right](0)}{(2n+1)!}x^{2n+1}\right] && \text{by right hypothesis}\\
&= \sum_{n=1}^{\infty}(-1)^n\left[\frac{(2n)(2n-1)\mathsf{f}(0)}{(2n)!}x^{2n-2} + \frac{(2n+1)(2n)\left[\frac{d}{dx}\mathsf{f}\right](0)}{(2n+1)!}x^{2n-1}\right]\\
&= \sum_{n=1}^{\infty}(-1)^n\left[\frac{\mathsf{f}(0)}{(2n-2)!}x^{2n-2} + \frac{\left[\frac{d}{dx}\mathsf{f}\right](0)}{(2n-1)!}x^{2n-1}\right]\\
&= \sum_{n=0}^{\infty}(-1)^{n+1}\left[\frac{\mathsf{f}(0)}{(2n)!}x^{2n} + \frac{\left[\frac{d}{dx}\mathsf{f}\right](0)}{(2n+1)!}x^{2n+1}\right]\\
&= -\mathsf{f}(x) && \text{by right hypothesis}
\end{aligned}$$

✏

Theorem I.1 (Taylor series for cosine/sine). [12]

THM

$$\begin{aligned}
\cos(x) &= \sum_{n=0}^{\infty}(-1)^n\frac{x^{2n}}{(2n)!} &&= 1 - \frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots && \forall x\in\mathbb{R}\\
\sin(x) &= \sum_{n=0}^{\infty}(-1)^n\frac{x^{2n+1}}{(2n+1)!} &&= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots && \forall x\in\mathbb{R}
\end{aligned}$$

[12] Rosenlicht (1968) page 157

✎PROOF:

$$\begin{aligned}
\cos(x) &= \underbrace{\mathsf{f}(0)\sum_{n=0}^{\infty}(-1)^n\frac{x^{2n}}{(2n)!}}_{\text{even terms}} + \underbrace{\left[\tfrac{\mathsf{d}}{\mathsf{d}x}\mathsf{f}\right](0)\sum_{n=0}^{\infty}(-1)^n\frac{x^{2n+1}}{(2n+1)!}}_{\text{odd terms}} && \text{by Lemma I.1 page 229}\\
&= 1\sum_{n=0}^{\infty}(-1)^n\frac{x^{2n}}{(2n)!} + 0\sum_{n=0}^{\infty}(-1)^n\frac{x^{2n+1}}{(2n+1)!} && \text{by cos initial conditions (Definition I.1 page 229)}\\
&= \sum_{n=0}^{\infty}(-1)^n\frac{x^{2n}}{(2n)!}\\
\sin(x) &= \underbrace{\mathsf{f}(0)\sum_{n=0}^{\infty}(-1)^n\frac{x^{2n}}{(2n)!}}_{\text{even terms}} + \underbrace{\left[\tfrac{\mathsf{d}}{\mathsf{d}x}\mathsf{f}\right](0)\sum_{n=0}^{\infty}(-1)^n\frac{x^{2n+1}}{(2n+1)!}}_{\text{odd terms}} && \text{by Lemma I.1 page 229}\\
&= 0\sum_{n=0}^{\infty}(-1)^n\frac{x^{2n}}{(2n)!} + 1\sum_{n=0}^{\infty}(-1)^n\frac{x^{2n+1}}{(2n+1)!} && \text{by sin initial conditions (Definition I.2 page 229)}\\
&= \sum_{n=0}^{\infty}(-1)^n\frac{x^{2n+1}}{(2n+1)!}
\end{aligned}$$

✎

Theorem I.2. [13]

THM		
	$\cos(0) = 1$	$\cos(-x) = \cos(x) \quad \forall x\in\mathbb{R}$ (EVEN)
	$\sin(0) = 0$	$\sin(-x) = -\sin(x) \quad \forall x\in\mathbb{R}$ (ODD)

✎PROOF:

$$\begin{aligned}
\cos(0) &= 1 - \frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots\bigg|_{x=0} && \text{by \textit{Taylor series for cosine} (Theorem I.1 page 230)}\\
&= 1\\
\sin(0) &= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots\bigg|_{x=0} && \text{by \textit{Taylor series for sine} (Theorem I.1 page 230)}\\
&= 0\\
\cos(-x) &= 1 - \frac{(-x)^2}{2} + \frac{(-x)^4}{4!} - \frac{(-x)^6}{6!} + \cdots && \text{by \textit{Taylor series for cosine} (Theorem I.1 page 230)}\\
&= 1 - \frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots\\
&= \cos(x) && \text{by \textit{Taylor series for cosine} (Theorem I.1 page 230)}\\
\sin(-x) &= (-x) - \frac{(-x)^3}{3!} + \frac{(-x)^5}{5!} - \frac{(-x)^7}{7!} + \cdots && \text{by \textit{Taylor series for sine} (Theorem I.1 page 230)}\\
&= -\left[x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots\right]\\
&= \sin(x) && \text{by \textit{Taylor series for sine} (Theorem I.1 page 230)}
\end{aligned}$$

✎

Lemma I.2. [14]

LEM		
	$\cos(1) > 0$	$x \in (0:2) \implies \sin(x) > 0$
	$\cos(2) < 0$	

[13] Rosenlicht (1968) page 157
[14] Rosenlicht (1968) page 158

✎PROOF:

$$\begin{aligned}
\cos(1) &= 1 - \frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots \Big|_{x=1} && \text{by \textit{Taylor series for cosine}} && \text{(Theorem I.1 page 230)}\\
&= 1 - \frac{1}{2} + \frac{1}{4!} - \frac{1}{6!} + \cdots\\
&> 0\\
\cos(2) &= 1 - \frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots \Big|_{x=2} && \text{by \textit{Taylor series for cosine}} && \text{(Theorem I.1 page 230)}\\
&= 1 - \frac{4}{2} + \frac{16}{24} - \frac{64}{720} + \cdots\\
&< 0
\end{aligned}$$

$$\begin{aligned}
x \in (0:2) &\implies \text{each term in the sequence } \left(\left(x - \frac{x^3}{3!}\right), \left(\frac{x^5}{5!} - \frac{x^7}{7!}\right), \left(\frac{x^9}{9!} - \frac{x^{11}}{11!}\right), \ldots\right) \text{ is} > 0\\
&\implies \sin(x) > 0
\end{aligned}$$

✏

Proposition I.1. *Let π be defined as in Definition I.3 (page 229).*

PRP
(A). *The value π **exists** in $\mathbb{R}$.*
(B). $2 < \pi < 4$.

✎PROOF:

$$\begin{aligned}
&\cos(1) > 0 && \text{by Lemma I.2 page 231}\\
&\cos(2) < 0 && \text{by Lemma I.2 page 231}\\
&\implies 1 < \frac{\pi}{2} < 2\\
&\implies 2 < \pi < 4
\end{aligned}$$

✏

Theorem I.3. [15] *Let $\boldsymbol{C}$ be the space of all continuously differentiable real functions and $\frac{\mathrm{d}}{\mathrm{d}x} \in \boldsymbol{C}^{\boldsymbol{C}}$ the differentiation operator. Let $\mathrm{f}'(0) \triangleq \left[\frac{\mathrm{d}}{\mathrm{d}x}\mathrm{f}\right](0)$.*

THM
$$\left\{\frac{\mathrm{d}^2}{\mathrm{d}x^2}\mathrm{f} + \mathrm{f} = 0\right\} \iff \left\{\mathrm{f}(x) = \mathrm{f}(0)\cos(x) + \mathrm{f}'(0)\sin(x)\right\} \qquad \forall \mathrm{f}\in\boldsymbol{C}, \forall x\in\mathbb{R}$$

✎PROOF:

1. Proof that $\left[\frac{\mathrm{d}^2}{\mathrm{d}x^2}\mathrm{f}\right](x) = -\mathrm{f}(x) \implies \mathrm{f}(x) = \mathrm{f}(0)\cos(x) + \left[\frac{\mathrm{d}}{\mathrm{d}x}\mathrm{f}\right](0)\sin(x)$:

$$\begin{aligned}
\mathrm{f}(x) &= \mathrm{f}(0)\underbrace{\sum_{n=0}^{\infty}(-1)^n\frac{x^{2n}}{(2n)!}}_{\cos(x)} + \left[\frac{\mathrm{d}}{\mathrm{d}x}\mathrm{f}\right](0)\underbrace{\sum_{n=0}^{\infty}(-1)^n\frac{x^{2n+1}}{(2n+1)!}}_{\sin(x)} && \text{by left hypothesis and Lemma I.1 page 229}\\
&= \mathrm{f}(0)\cos x + \left[\frac{\mathrm{d}}{\mathrm{d}x}\mathrm{f}\right](0)\sin x && \text{by definitions of cos and sin (Definition I.1 page 229, Definition I.2 page 229)}
\end{aligned}$$

[15] Rosenlicht (1968) page 157.

2. Proof that $\frac{d^2}{dx^2}f = -f \iff f(x) = f(0)\cos(x) + \left[\frac{d}{dx}f\right](0)\sin(x)$:

$$\begin{aligned}
f(x) &= f(0)\cos x + \left[\frac{d}{dx}f\right](0)\sin x && \text{by right hypothesis}\\
&= f(0)\underbrace{\sum_{n=0}^{\infty}(-1)^n\frac{x^{2n}}{(2n)!}}_{\cos(x)} + \left[\frac{d}{dx}f\right](0)\underbrace{\sum_{n=0}^{\infty}(-1)^n\frac{x^{2n+1}}{(2n+1)!}}_{\sin(x)}\\
\implies \quad & \frac{d^2}{dx^2}f + f = 0 && \text{by Lemma I.1 page 229}
\end{aligned}$$

✏

Remark I.1. The general solution for the *non-homogeneous* equation $\frac{d^2}{dx^2}f(x)+f(x) = g(x)$ with initial conditions $f(a) = 1$ and $f'(a) = \rho$ is $f(x) = \cos(x) + \rho\sin(x) + \int_a^x g(y)\sin(x-y)\,dy$. This type of equation is called a *Volterra integral equation of the second type*. References: Folland (1992) page 371, Liouville (1839). Volterra equation references: Pedersen (2000) page 99, Lalescu (1908), Lalescu (1911)

Theorem I.4. [16] *Let $\frac{d}{dx} \in C^C$ be the differentiation operator.*

THM
$$\frac{d}{dx}\cos(x) = -\sin(x) \quad \forall x\in\mathbb{R} \qquad \frac{d}{dx}\sin(x) = \cos(x) \quad \forall x\in\mathbb{R} \qquad \cos^2(x)+\sin^2(x) = 1 \quad \forall x\in\mathbb{R}$$

✎PROOF:

$$\begin{aligned}
\frac{d}{dx}\cos(x) &= \frac{d}{dx}\sum_{n=0}^{\infty}(-1)^n\frac{x^{2n}}{(2n)!} && \text{by \textit{Taylor series}} \quad \text{(Theorem I.1 page 230)}\\
&= \sum_{n=1}^{\infty}(-1)^n\frac{2nx^{2n-1}}{(2n)!} = \sum_{n=1}^{\infty}(-1)^n\frac{x^{2n-1}}{(2n-1)!} = \sum_{n=0}^{\infty}(-1)^{n+1}\frac{x^{2n}}{(2n)!}\\
&= -\sin(x) && \text{by \textit{Taylor series}} \quad \text{(Theorem I.1 page 230)}\\
\frac{d}{dx}\sin(x) &= \frac{d}{dx}\sum_{n=0}^{\infty}(-1)^n\frac{x^{2n+1}}{(2n+1)!} && \text{by \textit{Taylor series}} \quad \text{(Theorem I.1 page 230)}\\
&= \sum_{n=0}^{\infty}(-1)^n\frac{(2n+1)x^{2n}}{(2n+1)!} = \sum_{n=0}^{\infty}(-1)^n\frac{x^{2n}}{(2n)!}\\
&= \cos(x) && \text{by \textit{Taylor series}} \quad \text{(Theorem I.1 page 230)}
\end{aligned}$$

$$\begin{aligned}
\frac{d}{dx}\left[\cos^2(x)+\sin^2(x)\right] &= -2\cos(x)\sin(x) + 2\sin(x)\cos(x)\\
&= 0\\
&\implies \cos^2(x)+\sin^2(x) \text{ is \textit{constant}}\\
&\implies \cos^2(x)+\sin^2(x)\\
&= \cos^2(0)+\sin^2(0)\\
&= 1+0 = 1 && \text{by Theorem I.2 page 231}
\end{aligned}$$

✏

Proposition I.2.

PRP
$$\sin\left(\frac{\pi}{2}\right) = 1$$

[16] Rosenlicht (1968) page 157

✎PROOF:

$$\begin{aligned}
\sin(\pi/2) &= \pm\sqrt{\sin^2(\pi/2)+0} \\
&= \pm\sqrt{\sin^2(\pi/2)+\cos^2(\pi/2)} && \text{by definition of } \pi && \text{(Definition I.3 page 229)} \\
&= \pm\sqrt{1} && \text{by Theorem I.4 page 233} \\
&= \pm 1 \\
&= 1 && \text{by Lemma I.2 page 231}
\end{aligned}$$

⇒

I.4 The complex exponential

Definition I.4.

DEF *The function* $\mathsf{f} \in \mathbb{C}^{\mathbb{C}}$ *is the* ***exponential function*** $\exp(ix) \triangleq \mathsf{f}(x)$ *if*

1. $\frac{\mathrm{d}^2}{\mathrm{d}x^2}\mathsf{f} + \mathsf{f} = 0$ *(second order homogeneous differential equation)* *and*
2. $\mathsf{f}(0) = 1$ *(first initial condition)* *and*
3. $\left[\frac{\mathrm{d}}{\mathrm{d}x}\mathsf{f}\right](0) = i$ *(second initial condition).*

Theorem I.5 (Euler's Identity). [17]

THM $$e^{ix} = \cos(x) + i\sin(x) \qquad \forall x\in\mathbb{R}$$

✎PROOF:

$$\begin{aligned}
\exp(ix) &= \mathsf{f}(0)\cos(x) + \left[\frac{\mathrm{d}}{\mathrm{d}x}\mathsf{f}\right](0)\sin(x) && \text{by Theorem I.3 page 232} \\
&= \cos(x) + i\sin(x) && \text{by Definition I.4 page 234}
\end{aligned}$$

⇒

Proposition I.3.

PRP $$e^{-i\pi/2} = -i \quad \Big| \quad e^{i\pi/2} = i$$

✎PROOF:

$$\begin{aligned}
e^{i\pi/2} &= \cos(\pi/2) + i\sin(\pi/2) && \text{by } \textit{Euler's Identity} \text{ (Theorem I.5 page 234)} \\
&= 0 + i && \text{by Theorem I.2 (page 231) and Proposition I.2 (page 233)} \\
e^{-i\pi/2} &= \cos(-\pi/2) + i\sin(-\pi/2) && \text{by } \textit{Euler's Identity} \text{ (Theorem I.5 page 234)} \\
&= \cos(\pi/2) - i\sin(\pi/2) && \text{by Theorem I.2 page 231} \\
&= 0 - i && \text{by Theorem I.2 (page 231) and Proposition I.2 (page 233)}
\end{aligned}$$

⇒

Corollary I.1.

COR $$e^{ix} = \sum_{n\in\mathbb{W}} \frac{(ix)^n}{n!} \qquad \forall x\in\mathbb{R}$$

[17] Euler (1748), Bottazzini (1986) page 12

✎PROOF:

$$\begin{aligned}
\boxed{e^{ix}} &= \cos(x) + i\sin(x) && \text{by } \textit{Euler's Identity} && \text{(Theorem I.5 page 234)}\\
&= \underbrace{\sum_{n\in\mathbb{W}} \frac{(-1)^n x^{2n}}{(2n)!}}_{\cos(x)} + i\underbrace{\sum_{n\in\mathbb{W}} \frac{(-1)^n x^{2n+1}}{(2n+1)!}}_{\sin(x)} && \text{by } \textit{Taylor series} && \text{(Theorem I.1 page 230)}\\
&= \sum_{n\in\mathbb{W}} \frac{(i^2)^n x^{2n}}{(2n)!} + \sum_{n\in\mathbb{W}} \frac{i(i^2)^n x^{2n+1}}{(2n+1)!} && = \sum_{n\in\mathbb{W}} \frac{(ix)^{2n}}{(2n)!} + \sum_{n\in\mathbb{W}} \frac{(ix)^{2n+1}}{(2n+1)!}\\
&= \sum_{n\in\mathbb{W}\cap\mathbb{Z}_e} \frac{(ix)^n}{n!} + \sum_{n\in\mathbb{W}\cap\mathbb{Z}_o} \frac{(ix)^n}{n!} && = \boxed{\sum_{n\in\mathbb{W}} \frac{(ix)^n}{n!}}
\end{aligned}$$

✏

Corollary I.2 (Euler formulas). [18]

COR

$$\cos(x) = \mathbf{R}_e\left(e^{ix}\right) = \frac{e^{ix}+e^{-ix}}{2} \quad \forall x\in\mathbb{R} \qquad \sin(x) = \mathbf{I}_m\left(e^{ix}\right) = \frac{e^{ix}-e^{-ix}}{2i} \quad \forall x\in\mathbb{R}$$

✎PROOF:

$$\begin{aligned}
\boxed{\mathbf{R}_e\left(e^{ix}\right)} &\triangleq \frac{e^{ix}+\left(e^{ix}\right)^*}{2} = \frac{e^{ix}+e^{-ix}}{2} && \text{by definition of } \Re && \text{(Definition M.5 page 305)}\\
&= \frac{\cos(x)+i\sin(x)}{2} + \frac{\cos(-x)+i\sin(-x)}{2} && \text{by } \textit{Euler's Identity} && \text{(Theorem I.5 page 234)}\\
&= \frac{\cos(x)+i\sin(x)}{2} + \frac{\cos(x)-i\sin(x)}{2} && = \frac{\cos(x)}{2} + \frac{\cos(x)}{2} && = \boxed{\cos(x)}\\
\boxed{\mathbf{I}_m\left(e^{ix}\right)} &\triangleq \frac{e^{ix}-\left(e^{ix}\right)^*}{2i} = \frac{e^{ix}-e^{-ix}}{2i} && \text{by definition of } \Im && \text{(Definition M.5 page 305)}\\
&= \frac{\cos(x)+i\sin(x)}{2i} - \frac{\cos(-x)+i\sin(-x)}{2i} && \text{by } \textit{Euler's Identity} && \text{(Theorem I.5 page 234)}\\
&= \frac{\cos(x)+i\sin(x)}{2i} - \frac{\cos(x)-i\sin(x)}{2i} && = \frac{i\sin(x)}{2i} + \frac{i\sin(x)}{2i} && = \boxed{\sin(x)}
\end{aligned}$$

✏

Theorem I.6. [19]

THM

$$e^{(\alpha+\beta)} = e^{\alpha}\, e^{\beta} \qquad \forall \alpha,\beta\in\mathbb{C}$$

✎PROOF:

$$\begin{aligned}
e^{\alpha}\, e^{\beta} &= \left(\sum_{n\in\mathbb{W}} \frac{\alpha^n}{n!}\right)\left(\sum_{m\in\mathbb{W}} \frac{\beta^m}{m!}\right) && \text{by Corollary I.1 page 234}\\
&= \sum_{n\in\mathbb{W}} \sum_{k=0}^{n} \frac{\alpha^k}{k!} \frac{\beta^{n-k}}{(n-k)!}\\
&= \sum_{n\in\mathbb{W}} \sum_{k=0}^{n} \frac{n!\,\alpha^k}{n!\,k!} \frac{\beta^{n-k}}{(n-k)!}
\end{aligned}$$

[18] Euler (1748), Bottazzini (1986) page 12
[19] Rudin (1987) page 1

$$
\begin{aligned}
&= \sum_{n\in\mathbb{W}} \frac{1}{n!} \sum_{k=0}^{n} \frac{n!}{k!\,(n-k)!}\,\alpha^k\,\beta^{n-k} \\
&= \sum_{n\in\mathbb{W}} \frac{1}{n!} \sum_{k=0}^{n} \binom{n}{k}\alpha^k\,\beta^{n-k} \\
&= \sum_{n\in\mathbb{W}} \frac{(\alpha+\beta)^n}{n!} && \text{by the \textit{Binomial Theorem}} \\
&= e^{\alpha+\beta} && \text{by Corollary I.1 page 234}
\end{aligned}
$$

✐

I.5 Trigonometric Identities

Theorem I.7 (shift identities).

THM

$$
\begin{aligned}
\cos\left(x+\frac{\pi}{2}\right) &= -\sin x \quad \forall x\in\mathbb{R} & \sin\left(x+\frac{\pi}{2}\right) &= \cos x \quad \forall x\in\mathbb{R} \\
\cos\left(x-\frac{\pi}{2}\right) &= \sin x \quad \forall x\in\mathbb{R} & \sin\left(x-\frac{\pi}{2}\right) &= -\cos x \quad \forall x\in\mathbb{R}
\end{aligned}
$$

✎PROOF:

$$
\begin{aligned}
\cos\left(x+\frac{\pi}{2}\right) &= \frac{e^{i\left(x+\frac{\pi}{2}\right)}+e^{-i\left(x+\frac{\pi}{2}\right)}}{2} && \text{by \textit{Euler formulas}} && \text{(Corollary I.2 page 235)} \\
&= \frac{e^{ix}e^{i\frac{\pi}{2}}+e^{-ix}e^{-i\frac{\pi}{2}}}{2} && \text{by } e^{\alpha\beta}=e^{\alpha}e^{\beta} \text{ result} && \text{(Theorem I.6 page 235)} \\
&= \frac{e^{ix}(i)+e^{-ix}(-i)}{2} && \text{by Proposition I.3 page 234} \\
&= \frac{e^{ix}-e^{-ix}}{-2i} \\
&= -\sin x && \text{by \textit{Euler formulas}} && \text{(Corollary I.2 page 235)} \\
\cos\left(x-\frac{\pi}{2}\right) &= \frac{e^{i\left(x-\frac{\pi}{2}\right)}+e^{-i\left(x-\frac{\pi}{2}\right)}}{2} && \text{by \textit{Euler formulas}} && \text{(Corollary I.2 page 235)} \\
&= \frac{e^{ix}e^{-i\frac{\pi}{2}}+e^{-ix}e^{+i\frac{\pi}{2}}}{2} && \text{by } e^{\alpha\beta}=e^{\alpha}e^{\beta} \text{ result} && \text{(Theorem I.6 page 235)} \\
&= \frac{e^{ix}(-i)+e^{-ix}(i)}{2} && \text{by Proposition I.3 page 234} \\
&= \frac{e^{ix}-e^{-ix}}{2i} \\
&= \sin x && \text{by \textit{Euler formulas}} && \text{(Corollary I.2 page 235)} \\
\sin\left(x+\frac{\pi}{2}\right) &= \cos\left(\left[x+\frac{\pi}{2}\right]-\frac{\pi}{2}\right) && \text{by previous result} \\
&= \cos(x) \\
\sin\left(x-\frac{\pi}{2}\right) &= -\cos\left(\left[x-\frac{\pi}{2}\right]+\frac{\pi}{2}\right) && \text{by previous result} \\
&= -\cos(x)
\end{aligned}
$$

✐

Theorem I.8 (product identities).

T H M	(A).	$\cos x \cos y$	$=$	$\frac{1}{2}\cos(x-y)$	$+$	$\frac{1}{2}\cos(x+y)$	$\forall x,y\in\mathbb{R}$
	(B).	$\cos x \sin y$	$=$	$-\frac{1}{2}\sin(x-y)$	$+$	$\frac{1}{2}\sin(x+y)$	$\forall x,y\in\mathbb{R}$
	(C).	$\sin x \cos y$	$=$	$\frac{1}{2}\sin(x-y)$	$+$	$\frac{1}{2}\sin(x+y)$	$\forall x,y\in\mathbb{R}$
	(D).	$\sin x \sin y$	$=$	$\frac{1}{2}\cos(x-y)$	$-$	$\frac{1}{2}\cos(x+y)$	$\forall x,y\in\mathbb{R}$

✎PROOF:

1. Proof for (A) using *Euler formulas* (Corollary I.2 page 235)
   (algebraic method requiring *complex number system* $\mathbb{C}$):

$$\begin{aligned}
\cos x\cos y &= \left(\frac{e^{ix}+e^{-ix}}{2}\right)\left(\frac{e^{iy}+e^{-iy}}{2}\right) && \text{by } \textit{Euler formulas} \quad \text{(Corollary I.2 page 235)}\\
&= \frac{e^{i(x+y)}+e^{i(x-y)}+e^{i(-x+y)}+e^{i(-x-y)}}{4}\\
&= \frac{e^{i(x+y)}+e^{-i(x+y)}}{4}+\frac{e^{i(x-y)}+e^{-i(x-y)}}{4}\\
&= \frac{2\cos(x+y)}{4}+\frac{2\cos(x-y)}{4} && \text{by } \textit{Euler formulas} \quad \text{(Corollary I.2 page 235)}\\
&= \frac{1}{2}\cos(x+y)+\frac{1}{2}\cos(x-y)
\end{aligned}$$

2. Proof for (A) using *Volterra integral equation* (Theorem I.3 page 232)
   (differential equation method requiring only *real number system* $\mathbb{R}$):

$$\begin{aligned}
&\mathsf{f}(x) \triangleq \tfrac{1}{2}\cos(x-y)+\tfrac{1}{2}\cos(x+y)\\
&\implies \tfrac{d}{dx}\mathsf{f}(x) = -\tfrac{1}{2}\sin(x-y)-\tfrac{1}{2}\sin(x+y) && \text{by Theorem I.4 page 233}\\
&\implies \tfrac{d^2}{dx^2}\mathsf{f}(x) = -\tfrac{1}{2}\cos(x-y)-\tfrac{1}{2}\cos(x+y) && \text{by Theorem I.4 page 233}\\
&\implies \tfrac{d^2}{dx^2}\mathsf{f}(x)+\mathsf{f}(x) = 0 && \text{by } \textit{additive inverse} \text{ property}\\
&\implies \underbrace{\tfrac{1}{2}\cos(x-y)+\tfrac{1}{2}\cos(x+y)}_{\mathsf{f}(x)} = \underbrace{[\tfrac{1}{2}\cos(0-y)+\tfrac{1}{2}\cos(0+y)]}_{\mathsf{f}''(0)}\cos(x)+\underbrace{[-\tfrac{1}{2}\sin(0-y)-\tfrac{1}{2}\sin(0+y)]}_{\mathsf{f}'(0)}\sin(x)\\
&\implies \tfrac{1}{2}\cos(x-y)+\tfrac{1}{2}\cos(x+y) = \cos y\cos x + 0\sin(x)\\
&\implies \cos x\cos y = \tfrac{1}{2}\cos(x-y)+\tfrac{1}{2}\cos(x+y)
\end{aligned}$$

3. Proof for (B) using *Euler formulas* (Corollary I.2 page 235):

$$\begin{aligned}
\sin x\sin y &= \left(\frac{e^{ix}-e^{-ix}}{2i}\right)\left(\frac{e^{iy}-e^{-iy}}{2i}\right) && \text{by Corollary I.2 page 235}\\
&= \frac{e^{i(x+y)}-e^{i(x-y)}-e^{i(-x+y)}+e^{i(-x-y)}}{-4}\\
&= \frac{e^{i(x+y)}+e^{-i(x+y)}-e^{i(x-y)}-e^{-i(x-y)}}{-4}\\
&= \frac{e^{i(x+y)}+e^{-i(x+y)}}{-4}-\frac{e^{i(x-y)}+e^{-i(x-y)}}{-4}\\
&= \frac{2\cos(x-y)}{4}-\frac{2\cos(x+y)}{4} && \text{by Corollary I.2 page 235}\\
&= \frac{1}{2}\cos(x-y)-\frac{1}{2}\cos(x+y)
\end{aligned}$$

4. Proofs for (C) and (D) using (A) and (B):

$$
\begin{aligned}
\cos x \sin y &= \cos(x)\cos\left(y - \frac{\pi}{2}\right) && \text{by } \textit{shift identities} && \text{(Theorem I.7 page 236)} \\
&= \frac{1}{2}\cos\left(x + y - \frac{\pi}{2}\right) + \frac{1}{2}\cos\left(x - y + \frac{\pi}{2}\right) && \text{by (A)} \\
&= \frac{1}{2}\sin(x+y) - \frac{1}{2}\sin(x-y) && \text{by } \textit{shift identities} && \text{(Theorem I.7 page 236)} \\
\sin x \cos y &= \cos y \sin x \\
&= \frac{1}{2}\sin(y+x) - \frac{1}{2}\sin(y-x) && \text{by (B)} \\
&= \frac{1}{2}\sin(x+y) + \frac{1}{2}\sin(x-y) && \text{by Theorem I.2 page 231}
\end{aligned}
$$

✏

Proposition I.4.

PRP					
	(A). $\cos(\pi) = -1$	(C). $\cos(2\pi) = 1$	(E). $e^{i\pi} = -1$
	(B). $\sin(\pi) = 0$	(D). $\sin(2\pi) = 0$	(F). $e^{i2\pi} = 0$

✎PROOF:

$$
\begin{aligned}
\cos(\pi) &= -1 + 1 + \cos(\pi) \\
&= -1 + 2[{}^1\!/\!_2\cos({}^\pi\!/\!_2 - {}^\pi\!/\!_2) + {}^1\!/\!_2\cos({}^\pi\!/\!_2 + {}^\pi\!/\!_2)] && \text{by } \cos(0) = 1 \text{ result} && \text{(Theorem I.2 page 231)} \\
&= -1 + 2\cos({}^\pi\!/\!_2)\cos({}^\pi\!/\!_2) && \text{by } \textit{product identities} && \text{(Theorem I.8 page 236)} \\
&= -1 + 2(0)(0) && \text{by definition of } \pi && \text{(Definition I.3 page 229)} \\
&= -1 \\
\sin(\pi) &= 0 + \sin(\pi) \\
&= 2[-{}^1\!/\!_2\sin({}^\pi\!/\!_2 - {}^\pi\!/\!_2) + {}^1\!/\!_2\sin({}^\pi\!/\!_2 + {}^\pi\!/\!_2)] && \text{by } \sin(0) = 0 \text{ result} && \text{(Theorem I.2 page 231)} \\
&= 2\cos({}^\pi\!/\!_2)\sin({}^\pi\!/\!_2) && \text{by } \textit{product identities} && \text{(Theorem I.8 page 236)} \\
&= 2(0)\sin({}^\pi\!/\!_2) && \text{by definition of } \pi && \text{(Definition I.3 page 229)} \\
&= 0 \\
\cos(2\pi) &= 1 + \cos(2\pi) - 1 \\
&= 2[{}^1\!/\!_2\cos(\pi - \pi) + {}^1\!/\!_2\cos(\pi + \pi)] - 1 && \text{by } \cos(0) = 1 \text{ result} && \text{(Theorem I.2 page 231)} \\
&= 2\cos(\pi)\cos(\pi) - 1 && \text{by } \textit{product identities} && \text{(Theorem I.8 page 236)} \\
&= 2(-1)(-1) - 1 && \text{by (A)} \\
&= 1 \\
\sin(2\pi) &= 0 + \sin(2\pi) \\
&= 2[{}^1\!/\!_2\sin(\pi - \pi) + {}^1\!/\!_2\sin(\pi + \pi)] && \text{by } \sin(0) = 0 \text{ result} && \text{(Theorem I.2 page 231)} \\
&= 2\sin(\pi)\cos(\pi) && \text{by } \textit{product identities} && \text{(Theorem I.8 page 236)} \\
&= 2(0)(-1) && \text{by (A) and (B)} \\
&= 0 \\
e^{i\pi} &= \cos(\pi) + i\sin(\pi) && \text{by } \textit{Euler's Identity} && \text{(Theorem I.5 page 234)} \\
&= -1 + 0 && \text{by (A) and (B)} \\
&= -1 \\
e^{i2\pi} &= \cos(2\pi) + i\sin(2\pi) && \text{by } \textit{Euler's Identity} && \text{(Theorem I.5 page 234)} \\
&= 1 + 0 && \text{by (C) and (D)} \\
&= 1
\end{aligned}
$$

✏

Theorem I.9 (double angle formulas). [20]

THM

$$\begin{array}{llcll}
(A). & \cos(x+y) & = & \cos x\cos y - \sin x\sin y & \forall x,y\in\mathbb{R}\\
(B). & \sin(x+y) & = & \sin x\cos y + \cos x\sin y & \forall x,y\in\mathbb{R}\\
(C). & \tan(x+y) & = & \dfrac{\tan x + \tan y}{1-\tan x\tan y} & \forall x,y\in\mathbb{R}
\end{array}$$

✎PROOF:

1. Proof for (A) using *product identities* (Theorem I.8 page 236).

$$\begin{aligned}
\cos(x+y) &= \underbrace{\frac{1}{2}\cos(x+y)+\frac{1}{2}\cos(x+y)}_{\cos(x+y)} + \underbrace{\frac{1}{2}\cos(x-y)-\frac{1}{2}\cos(x-y)}_{0}\\
&= \left[\frac{1}{2}\cos(x-y)+\frac{1}{2}\cos(x+y)\right] - \left[\frac{1}{2}\cos(x-y)-\frac{1}{2}\cos(x+y)\right]\\
&= \cos x\cos y - \sin x\sin y && \text{by Theorem I.8 page 236}
\end{aligned}$$

2. Proof for (A) using *Volterra integral equation* (Theorem I.3 page 232):

$$\begin{aligned}
\mathsf{f}(x)\triangleq\cos(x+y) &\implies \frac{\mathsf{d}}{\mathsf{d}x}\mathsf{f}(x) = -\sin(x+y) && \text{by Theorem I.4 page 233}\\
&\implies \frac{\mathsf{d}^2}{\mathsf{d}x^2}\mathsf{f}(x) = -\cos(x+y) && \text{by Theorem I.4 page 233}\\
&\implies \frac{\mathsf{d}^2}{\mathsf{d}x^2}\mathsf{f}(x)+\mathsf{f}(x) = 0 && \text{by \textit{additive inverse} property}\\
&\implies \cos(x+y) = \cos y\cos x - \sin y\sin x && \text{by Theorem I.3 page 232}\\
&\implies \cos(x+y) = \cos x\cos y - \sin x\sin y && \text{by \textit{commutative} property}
\end{aligned}$$

3. Proof for (B) and (C) using (A):

$$\begin{aligned}
\sin(x+y) &= \cos\left(x-\frac{\pi}{2}+y\right) && \text{by \textit{shift identities} (Theorem I.7 page 236)}\\
&= \cos\left(x-\frac{\pi}{2}\right)\cos(y) - \sin\left(x-\frac{\pi}{2}\right)\sin(y) && \text{by (A)}\\
&= \sin(x)\cos(y)+\cos(x)\sin(y) && \text{by \textit{shift identities} (Theorem I.7 page 236)}
\end{aligned}$$

$$\begin{aligned}
\tan(x+y) &= \frac{\sin(x+y)}{\cos(x+y)}\\
&= \frac{\sin x\cos y+\cos x\sin y}{\cos x\cos y-\sin x\sin y} && \text{by (A)}\\
&= \left(\frac{\sin x\cos y+\cos x\sin y}{\cos x\cos y-\sin x\sin y}\right)\left(\frac{\cos x\cos y}{\cos x\cos y}\right)\\
&= \frac{\frac{\sin x\cos y}{\cos x\cos y}+\frac{\cos x\sin y}{\cos x\cos y}}{\frac{\cos x\cos y}{\cos x\cos y}-\frac{\sin x\sin y}{\cos x\cos y}} = \frac{\frac{\sin x}{\cos x}+\frac{\sin y}{\cos y}}{1-\frac{\sin x}{\cos x}\frac{\sin y}{\cos y}} = \frac{\tan x+\tan y}{1-\tan x\tan y}
\end{aligned}$$

✏

Theorem I.10 (trigonometric periodicity).

THM

$$\begin{array}{llcll|llcll}
(A). & \cos(x+M\pi) & = & (-1)^M\cos(x) & \forall x\in\mathbb{R},\ M\in\mathbb{Z} & (D). & \cos(x+2M\pi) & = & \cos(x) & \forall x\in\mathbb{R},\ M\in\mathbb{Z}\\
(B). & \sin(x+M\pi) & = & (-1)^M\sin(x) & \forall x\in\mathbb{R},\ M\in\mathbb{Z} & (E). & \sin(x+2M\pi) & = & \sin(x) & \forall x\in\mathbb{R},\ M\in\mathbb{Z}\\
(C). & e^{i(x+M\pi)} & = & (-1)^M e^{ix} & \forall x\in\mathbb{R},\ M\in\mathbb{Z} & (F). & e^{i(x+2M\pi)} & = & e^{ix} & \forall x\in\mathbb{R},\ M\in\mathbb{Z}
\end{array}$$

[20] Expressions for $\cos(\alpha+\beta)$, $\sin(\alpha+\beta)$, and $\sin^2 x$ appear in works as early as Ptolemy (circa 100AD). Reference: http://en.wikipedia.org/wiki/History_of_trigonometric_functions

✎PROOF:

1. Proof for (A):

   (a) $M = 0$ case: $\cos(x + 0\pi) \;=\; \cos(x) \;=\; (-1)^0 \cos(x)$

   (b) Proof for $M > 0$ cases (by induction):

      i. Base case $M = 1$:

$$\begin{aligned}
\cos(x+\pi) &= \cos x \cos\pi - \sin x \sin\pi && \text{by } \textit{double angle formulas} && \text{(Theorem I.9 page 239)}\\
&= \cos x(-1) - \sin x(0) && \text{by } \cos\pi = -1 \text{ result} && \text{(Proposition I.4 page 238)}\\
&= (-1)^1 \cos x
\end{aligned}$$

      ii. Inductive step…Proof that M case $\implies$ $M+1$ case:

$$\begin{aligned}
\cos(x+[M+1]\pi) &= \cos([x+\pi]+M\pi)\\
&= (-1)^M \cos(x+\pi) && \text{by induction hypothesis } (M \text{ case})\\
&= (-1)^M(-1)\cos(x) && \text{by base case (item (1(b)i) page 240)}\\
&= (-1)^{M+1}\cos(x)\\
&\implies M+1 \text{ case}
\end{aligned}$$

   (c) Proof for $M < 0$ cases: Let $N \triangleq -M \ldots \implies N > 0$.

$$\begin{aligned}
\cos(x+M\pi) &\triangleq \cos(x-N\pi) && \text{by definition of } N\\
&= \cos(x)\cos(-N\pi) - \sin(x)\sin(-N\pi) && \text{by } \textit{double angle formulas} && \text{(Theorem I.9 page 239)}\\
&= \cos(x)\cos(N\pi) + \sin(x)\sin(N\pi) && \text{by Theorem I.2 page 231}\\
&= \cos(x)\cos(0+N\pi) + \sin(x)\sin(0+N\pi)\\
&= \cos(x)(-1)^N\cos(0) + \sin(x)(-1)^N\sin(0) && \text{by } M \ge 0 \text{ results} && \text{(item (1b) page 240)}\\
&= (-1)^N\cos(x) && \text{by } \cos(0)=1,\ \sin(0)=0 \text{ results} && \text{(Theorem I.2 page 231)}\\
&\triangleq (-1)^{-M}\cos(x) && \text{by definition of } N\\
&= (-1)^M\cos(x)
\end{aligned}$$

   (d) Proof using complex exponential:

$$\begin{aligned}
\cos(x+M\pi) &= \frac{e^{i(x+M\pi)} + e^{-i(x+M\pi)}}{2} && \text{by } \textit{Euler formulas} && \text{(Corollary I.2 page 235)}\\
&= e^{iM\pi}\left[\frac{e^{ix}+e^{-ix}}{2}\right] && \text{by } e^{\alpha\beta} = e^{\alpha}e^{\beta} \text{ result} && \text{(Theorem I.6 page 235)}\\
&= \left(e^{i\pi}\right)^M \cos x && \text{by } \textit{Euler formulas} && \text{(Corollary I.2 page 235)}\\
&= (-1)^M \cos x && \text{by } e^{i\pi} = -1 \text{ result} && \text{(Proposition I.4 page 238)}
\end{aligned}$$

2. Proof for (B):

   (a) $M = 0$ case: $\sin(x + 0\pi) \;=\; \sin(x) \;=\; (-1)^0 \sin(x)$

   (b) Proof for $M > 0$ cases (by induction):

      i. Base case $M = 1$:

$$\begin{aligned}
\sin(x+\pi) &= \sin x \cos\pi + \cos x \sin\pi && \text{by } \textit{double angle formulas} && \text{(Theorem I.9 page 239)}\\
&= \sin x(-1) - \cos x(0) && \text{by } \sin\pi = 0 \text{ results} && \text{(Proposition I.4 page 238)}\\
&= (-1)^1 \sin x
\end{aligned}$$

ii. Inductive step…Proof that M case $\implies$ $M+1$ case:

$$\begin{aligned}
\sin(x+[M+1]\pi) &= \sin([x+\pi]+M\pi) \\
&= (-1)^M \sin(x+\pi) && \text{by induction hypothesis (M case)} \\
&= (-1)^M(-1)\sin(x) && \text{by base case (item (2(b)i) page 240)} \\
&= (-1)^{M+1}\sin(x) \\
&\implies M+1 \text{ case}
\end{aligned}$$

(c) Proof for $M<0$ cases: Let $N \triangleq -M \ldots \implies N>0$.

$$\begin{aligned}
\sin(x+M\pi) &\triangleq \sin(x-N\pi) && \text{by definition of } N \\
&= \sin(x)\sin(-N\pi) - \sin(x)\sin(-N\pi) && \text{by \textit{double angle formulas}} && \text{(Theorem I.9 page 239)} \\
&= \sin(x)\sin(N\pi) + \sin(x)\sin(N\pi) && \text{by Theorem I.2 page 231} \\
&= \sin(x)\sin(0+N\pi) + \sin(x)\sin(0+N\pi) \\
&= \sin(x)(-1)^N\sin(0) + \sin(x)(-1)^N\sin(0) && \text{by } M \geq 0 \text{ results} && \text{(item (2b) page 240)} \\
&= (-1)^N \sin(x) && \text{by } \sin(0)=1,\ \sin(0)=0 \text{ results} && \text{(Theorem I.2 page 231)} \\
&\triangleq (-1)^{-M}\sin(x) && \text{by definition of } N \\
&= (-1)^M \sin(x)
\end{aligned}$$

(d) Proof using complex exponential:

$$\begin{aligned}
\sin(x+M\pi) &= \frac{e^{i(x+M\pi)} - e^{-i(x+M\pi)}}{2i} && \text{by \textit{Euler formulas}} && \text{(Corollary I.2 page 235)} \\
&= e^{iM\pi}\left[\frac{e^{ix}-e^{-ix}}{2i}\right] && \text{by } e^{\alpha\beta} = e^{\alpha}e^{\beta} \text{ result} && \text{(Theorem I.6 page 235)} \\
&= \left(e^{i\pi}\right)^M \sin x && \text{by \textit{Euler formulas}} && \text{(Corollary I.2 page 235)} \\
&= (-1)^M \sin x && \text{by } e^{i\pi} = -1 \text{ result} && \text{(Proposition I.4 page 238)}
\end{aligned}$$

3. Proof for (C):

$$\begin{aligned}
e^{i(x+M\pi)} &= e^{iM\pi}e^{ix} && \text{by } e^{\alpha\beta} = e^{\alpha}e^{\beta} \text{ result} && \text{(Theorem I.6 page 235)} \\
&= \left(e^{i\pi}\right)^M\left(e^{ix}\right) \\
&= (-1)^M e^{ix} && \text{by } e^{i\pi} = -1 \text{ result} && \text{(Proposition I.4 page 238)}
\end{aligned}$$

4. Proofs for (D), (E), and (F):

$$\begin{array}{lclcll}
\cos(i(x+2M\pi)) &=& (-1)^{2M}\cos(ix) &=& \cos(ix) & \text{by (A)} \\
\sin(i(x+2M\pi)) &=& (-1)^{2M}\sin(ix) &=& \sin(ix) & \text{by (B)} \\
e^{i(x+2M\pi)} &=& (-1)^{2M}e^{ix} &=& e^{ix} & \text{by (C)}
\end{array}$$

⇒

Theorem I.11 (half-angle formulas/squared identities).

THM		
	(A). $\cos^2 x = \tfrac{1}{2}(1+\cos 2x) \quad \forall x \in \mathbb{R}$	(C). $\cos^2 x + \sin^2 x = 1 \quad \forall x \in \mathbb{R}$
	(B). $\sin^2 x = \tfrac{1}{2}(1-\cos 2x) \quad \forall x \in \mathbb{R}$	

✎PROOF:

$$\begin{aligned}
\cos^2 x &\triangleq (\cos x)(\cos x) = \tfrac{1}{2}\cos(x-x) + \tfrac{1}{2}\cos(x+x) && \text{by \textit{product identities}} && \text{(Theorem I.8 page 236)} \\
&= \tfrac{1}{2}[1+\cos(2x)] && \text{by } \cos(0)=1 \text{ result} && \text{(Theorem I.2 page 231)} \\
\sin^2 x &= (\sin x)(\sin x) = \tfrac{1}{2}\cos(x-x) - \tfrac{1}{2}\cos(x+x) && \text{by \textit{product identities}} && \text{(Theorem I.8 page 236)} \\
&= \tfrac{1}{2}[1-\cos(2x)] && \text{by } \cos(0)=1 \text{ result} && \text{(Theorem I.2 page 231)} \\
\cos^2 x + \sin^2 x &= \tfrac{1}{2}[1+\cos(2x)] + \tfrac{1}{2}[1-\cos(2x)] = 1 && \text{by (A) and (B)} \\
& && \text{note: see also} && \text{Theorem I.4 page 233}
\end{aligned}$$

⇒

I.6 Planar Geometry

The harmonic functions $\cos(x)$ and $\sin(x)$ are *orthogonal* to each other in the sense

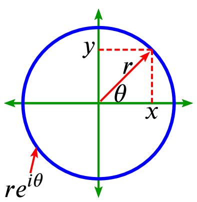
$$\begin{aligned}
\langle \cos(x) \,|\, \sin(x) \rangle &= \int_{-\pi}^{+\pi} \cos(x)\sin(x)\,\mathrm{d}x \\
&= \frac{1}{2}\int_{-\pi}^{+\pi} \sin(x-x)\,\mathrm{d}x + \frac{1}{2}\int_{-\pi}^{+\pi} \sin(x+x)\,\mathrm{d}x && \text{by Theorem I.8 page 236} \\
&= \cancelto{0}{\frac{1}{2}\int_{-\pi}^{+\pi} \sin(0)\,\mathrm{d}x} + \frac{1}{2}\int_{-\pi}^{+\pi} \sin(2x)\,\mathrm{d}x \\
&= -\frac{1}{2}\cdot\frac{1}{2}\cos(2x)\Big|_{-\pi}^{+\pi} \cos(2x) \\
&= -\frac{1}{4}[\cos(2\pi) - \cos(-2\pi)] \\
&= 0
\end{aligned}$$

Because $\cos(x)$ are $\sin(x)$ are orthogonal, they can be conveniently represented by the x and y axes in a plane— because perpendicular axes in a plane are also orthogonal. Vectors in the plane can be represented by linear combinations of $\cos x$ and $\sin x$. Let $\tan x$ be defined as

$$\tan x \triangleq \frac{\sin x}{\cos x}.$$

We can also define a value θ to represent the angle between such a vector and the x-axis such that

$$\theta = \tan^{-1}\left(\frac{\sin\theta}{\cos\theta}\right)$$



$$\begin{aligned}
\cos\theta &\triangleq \frac{x}{r} & \sec\theta &\triangleq \frac{r}{x} \\
\sin\theta &\triangleq \frac{y}{r} & \csc\theta &\triangleq \frac{r}{y} \\
\tan\theta &\triangleq \frac{y}{x} & \cot\theta &\triangleq \frac{x}{y}
\end{aligned}$$

I.7 Trigonometric functions of complex numbers

Definition I.5. [21]

DEF

$$\begin{aligned}
\cosh(z) &\triangleq \frac{e^z + e^{-z}}{2} && \forall z \in \mathbb{C} \\
\sinh(z) &\triangleq \frac{e^z - e^{-z}}{2} && \forall z \in \mathbb{C}
\end{aligned}$$

Theorem I.12. [22]

[21] Saxelby (1920) page 225

[22] https://proofwiki.org/wiki/Cosine_of_Complex_Number, https://proofwiki.org/wiki/Sine_of_Complex_Number, Saxelby (1920) pages 416–417

THM

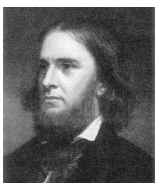
$$
\begin{array}{lcll}
\cosh(ix) &=& \cos(x) & \forall x \in\mathbb{R}\\
\sinh(ix) &=& i\sin(x) & \forall x \in\mathbb{R}\\
\cos(ix) &=& \cosh(x) & \forall x \in\mathbb{R}\\
\sin(ix) &=& i\sinh(x) & \forall x \in\mathbb{R}\\
\cos(x+iy) &=& \cos(x)\cosh(y) - i\sin(x)\sinh(y) & \forall x,y \in\mathbb{R}\\
\sin(x+iy) &=& \sin(x)\cosh(y) + i\cos(x)\sinh(y) & \forall x,y \in\mathbb{R}
\end{array}
$$

✎PROOF:

$$
\begin{aligned}
\cosh(ix) &\triangleq \frac{e^{ix}+e^{-ix}}{2} && \text{by definition of } \cosh(x) && \text{(Definition I.5 page 242)}\\
&= \cos(x) && \text{by \textit{Euler's Identity}} && \text{(Theorem I.5 page 234)}\\
\sinh(ix) &\triangleq \frac{e^{ix}-e^{-ix}}{2} && \text{by definition of } \sinh(x) && \text{(Definition I.5 page 242)}\\
&\triangleq i\left[\frac{e^{ix}-e^{-ix}}{2i}\right] && \text{by definition of } \sinh(x) && \text{(Definition I.5 page 242)}\\
&= i\sin(x) && \text{by \textit{Euler's Identity}} && \text{(Theorem I.5 page 234)}\\
\cos(ix) &\triangleq \frac{e^{iix}+e^{-iix}}{2} && \text{by \textit{Euler's Identity}} && \text{(Theorem I.5 page 234)}\\
&= \frac{e^{-x}+e^{x}}{2}\\
&= \frac{e^{x}+e^{-x}}{2}\\
&\triangleq \cosh(x) && \text{by definition of } \cosh(x) && \text{(Definition I.5 page 242)}\\
\sin(ix) &\triangleq \frac{e^{iix}-e^{-iix}}{2i} && \text{by \textit{Euler's Identity}} && \text{(Theorem I.5 page 234)}\\
&= \frac{e^{-x}-e^{x}}{2i}\\
&= -(-i^2)\left[\frac{e^{x}-e^{-x}}{2i}\right]\\
&= i\left[\frac{e^{x}-e^{-x}}{2}\right]\\
&\triangleq i\sinh(x) && \text{by definition of } \cosh(x) && \text{(Definition I.5 page 242)}\\
\cos(x+iy) &= \cos(x)\cos(iy) - \sin(x)\sin(iy) && \text{by \textit{double angle formulas}} && \text{(Theorem I.9 page 239)}\\
&= \cos(x)\cosh(y) - i\sin(x)\sinh(y) && \text{by previous results}\\
\sin(x+iy) &= \sin(x)\cos(iy) + \cos(x)\sin(iy) && \text{by \textit{double angle formulas}} && \text{(Theorem I.9 page 239)}\\
&= \sin(x)\cosh(y) + i\cos(x)\sinh(y) && \text{by previous results}
\end{aligned}
$$

✏

I.8 The power of the exponential



❝*Gentlemen, that is surely true, it is absolutely paradoxical; we cannot understand it, and we don't know what it means. But we have proved it, and therefore we know it must be the truth.*❞

Benjamin Peirce (1809–1880), American Harvard University mathematician after proving $e^{i\pi} = -1$ in a lecture. [23]

❝*Young man, in mathematics you don't understand things. You just get used to them.*❞
John von Neumann (1903–1957), Hungarian-American mathematician, as allegedly told to Gary Zukav by Felix T. Smith, Head of Molecular Physics at Stanford Research Institute, about a “physicist friend”. [24]

The following corollary presents one of the most amazing relations in all of mathematics. It shows a simple and compact relationship between the transcendental numbers π and e, the imaginary number i, and the additive and multiplicative identity elements 0 and 1. The fact that there is any relationship at all is somewhat amazing; but for there to be such an elegant one is truly one of the wonders of the world of numbers.

Corollary I.3. [25]

COR
$$e^{i\pi} + 1 = 0$$

✎PROOF:

$$\begin{aligned} e^{ix}\Big|_{x=\pi} &= [\cos x + i \sin x]_{x=\pi} && \text{by Euler's Identity (Theorem I.5 page 234)} \\ &= -1 + i \cdot 0 && \text{by Proposition I.4 page 238} \\ &= -1 \end{aligned}$$

⇨

There are many transforms available, several of them integral transforms $[\mathbf{A}f](s) \triangleq \int_t f(s)\kappa(t,s)\,ds$ using different kernels $\kappa(t,s)$. But of all of them, two of the most often used themselves use an exponential kernel:

① The *Laplace Transform* with kernel $\kappa(t, s) \triangleq e^{st}$
② The *Fourier Transform* with kernel $\kappa(t, \omega) \triangleq e^{i\omega t}$.

Of course, the Fourier kernel is just a special case of the Laplace kernel with $s = i\omega$ ($i\omega$ is a unit circle in s if s is depicted as a plane with real and imaginary axes). What is so special about exponential kernels? Is it just that they were discovered sooner than other kernels with other transforms? The answer in general is “no”. The exponential has two properties that makes it extremely special:

- The exponential is an eigenvalue of any *linear time invariant* (LTI) operator (Theorem I.13 page 244).
- The exponential generates a *continuous point spectrum* for the *differential operator*.

Theorem I.13. [26] *Let* $\mathbf{L}$ *be an operator with kernel* $h(t,\omega)$ *and*
$$\check{h}(s) \triangleq \langle h(t,\omega) \mid e^{st} \rangle \qquad \text{(LAPLACE TRANSFORM).}$$

---

[23] quote: Kasner and Newman (1940) page 104
image: http://www-history.mcs.st-andrews.ac.uk/history/PictDisplay/Peirce_Benjamin.html

[24] quote: Zukav (1980) page 208
image: http://en.wikipedia.org/wiki/John_von_Neumann

The quote appears in a footnote in Zukav (1980) that reads like this: Dr. Felix Smith, Head of Molecular Physics, Stanford Research Institute, once related to me the true story of a physicist friend who worked at Los Alamos after World War II. Seeking help on a difficult problem, he went to the great Hungarian mathematician, John von Neumann, who was at Los Alamos as a consultant. “Simple,” said von Neumann. “This can be solved by using the method of characteristics.” After the explanation the physicist said, “I'm afraid I don't understand the method of characteristics.” “Young man,” said von Neumann, “in mathematics you don't understand things, you just get used to them.”

[25] Euler (1748), Euler (1988) ⟨chapter 8?⟩, http://www.daviddarling.info/encyclopedia/E/Eulers_formula.html

[26] Mallat (1999) page 2, ...page 2 online: http://www.cmap.polytechnique.fr/~mallat/WTintro.pdf

THM
$$\left\{\begin{array}{ll}1. & \mathbf{L} \text{ is LINEAR and} \\ 2. & \mathbf{L} \text{ is TIME-INVARIANT}\end{array}\right\} \implies \left\{\mathbf{L}e^{st} = \underbrace{\check{\mathsf{h}}^*(-s)}_{\text{eigenvalue}} \underbrace{e^{st}}_{\text{eigenvector}}\right\}$$

✎PROOF:

$$\begin{aligned}
[\mathbf{L}e^{st}](s) &= \langle e^{su} \,|\, \mathsf{h}((t;u),s)\rangle && \text{by linear hypothesis}\\
&= \langle e^{su} \,|\, \mathsf{h}((t-u),s)\rangle && \text{by time-invariance hypothesis}\\
&= \left\langle e^{s(t-v)} \,|\, \mathsf{h}(v,s)\right\rangle && \text{let } v = t-u \implies u = t-v\\
&= e^{st}\ \langle e^{-sv} \,|\, \mathsf{h}(v,s)\rangle && \text{by additivity of } \langle\triangle\,|\,\triangledown\rangle \text{ (Definition N.1 page 309)}\\
&= \langle \mathsf{h}(v,s) \,|\, e^{-sv}\rangle^*\ e^{st} && \text{by conjugate symmetry of } \langle\triangle\,|\,\triangledown\rangle \text{ (Definition N.1 page 309)}\\
&= \left\langle \mathsf{h}(v,s) \,|\, e^{(-s)v}\right\rangle^*\ e^{st} && \\
&= \check{\mathsf{h}}^*(-s)\, e^{st} && \text{by definition of } \check{\mathsf{h}}(s)
\end{aligned}$$

✏

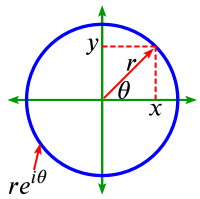
I.9 Composite Trigonometric Functions

Theorem I.14.

THM
$$\cos[\operatorname{atan}(z)] = \frac{1}{\sqrt{1+z^2}} \qquad \forall z \in \mathbb{R}$$

✎PROOF:

1. Let x be the adjacent side length,
y the opposite side length,
r the hypotenuse length, and
θ the angle between side r and side x of a right triangle.



2. Let $z \triangleq \frac{y}{x}$.
3. It then follows that ...

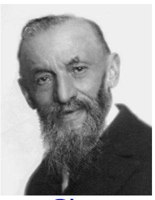
$$\begin{aligned}
\cos[\operatorname{atan}(z)] &= \cos\left[\operatorname{atan}\left(\frac{y}{x}\right)\right] && \text{by definition of } z && \text{item (2)}\\
&= \cos[\theta] && \text{by definitions of } \tan(\phi) \text{ and } \operatorname{atan}(\phi)\\
&= \frac{x}{r} && \text{by definition of } \cos(\phi) && \text{item (1)}\\
&= \frac{x}{\sqrt{x^2+y^2}} && \text{by \textit{Pythagoras' Theorem}}\\
&= \frac{\frac{1}{x}x}{\frac{1}{x}\sqrt{x^2+y^2}}\\
&= \frac{1}{\sqrt{\frac{x^2}{x^2}+\frac{y^2}{x^2}}}
\end{aligned}$$

$$= \frac{1}{\sqrt{1+z^2}} \qquad \text{by definition of } z \qquad \text{item (2)}$$

✏

APPENDIX J

LINEAR SPACES



❝*The geometric calculus, in general, consists in a system of operations on geometric entities, and their consequences, analogous to those that algebra has on the numbers. It permits the expression in formulas of the results of geometric constructions, the representation with equations of propositions of geometry, and the substitution of a transformation of equations for a verbal argument.*❞

Giuseppe Peano (1858–1932), Italian mathematician, credited with being one of the first to introduce the concept of the *linear space* (*vector space*).[1]

J.1 Definition and basic results

A *metric space* is a *set* together with nothing else save a *metric* that gives the space a *topology* (Definition **??** page **??**). A *linear space* (next definition) in general has no topology but does have some additional *algebraic* structure (Appendix G page 221) that is useful in generalizing a number of mathematical concepts. If one wishes to have both algebraic structure and a topology, then this can be accomplished by appending a *topology* to a *linear space* giving a *topological linear space* (Definition **??** page **??**), a *metric* giving a *metric linear space*, an *inner product* giving an *inner product space* (Definition N.1 page 309), or a *norm* giving a *normed linear space* (Definition O.1 page 327).

Definition J.1. [2] *Let* $(\mathbb{F}, +, \cdot, 0, 1)$ *be a* FIELD *(Definition G.5 page 222). Let* X *be a set, let* $+$ *be an* OPERATOR *(Definition R.1 page 359) in* X^{X^2}*, and let* $\otimes$ *be an operator in* $X^{\mathbb{F}\times X}$.

---

[1] quote: Peano (1888b) page ix
image http://en.wikipedia.org/wiki/File:Giuseppe_Peano.jpg, public domain

[2] Kubrusly (2001) pages 40–41 ⟨Definition 2.1 and following remarks⟩, Haaser and Sullivan (1991) page 41, Halmos (1948) pages 1–2, Peano (1888a) ⟨Chapter IX⟩, Peano (1888b) pages 119–120, Banach (1922) pages 134–135

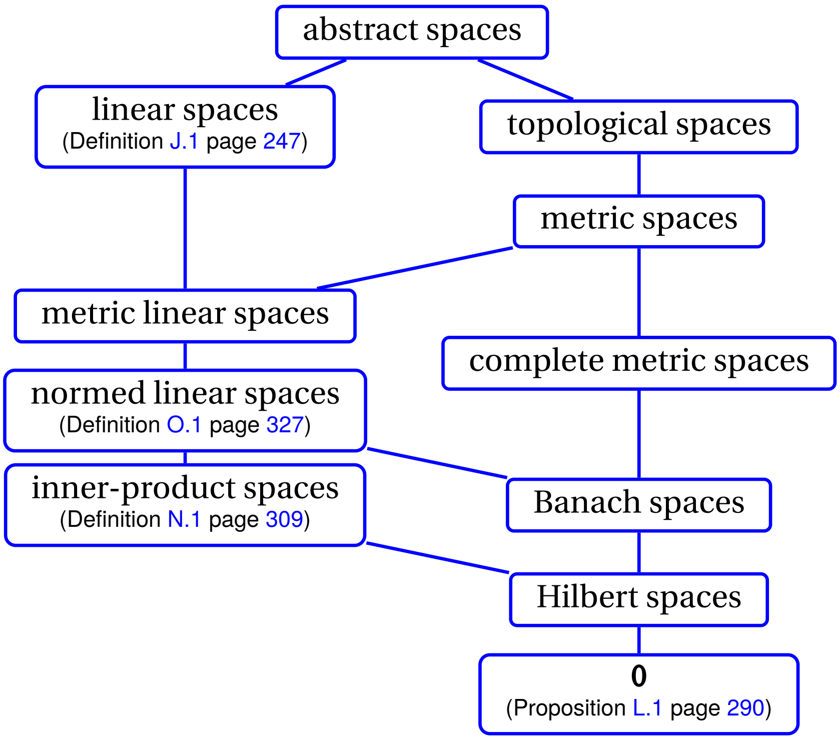


Figure J.1: Lattice of mathematical spaces

DEF

The structure $\boldsymbol{\Omega} \triangleq \big(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times})\big)$ *is a* ***linear space*** *over* $(\mathbb{F}, +, \cdot, 0, 1)$ *if*

1.	$\exists \mathbb{0} \in X$ such that	$\boldsymbol{x} + \mathbb{0} = \boldsymbol{x}$	$\forall \boldsymbol{x} \in X$	(+ IDENTITY)	*
2.	$\exists \boldsymbol{y} \in X$ such that	$\boldsymbol{x} + \boldsymbol{y} = \mathbb{0}$	$\forall \boldsymbol{x} \in X$	(+ INVERSE)	*
3.		$(\boldsymbol{x} + \boldsymbol{y}) + \boldsymbol{z} = \boldsymbol{x} + (\boldsymbol{y} + \boldsymbol{z})$	$\forall \boldsymbol{x},\boldsymbol{y},\boldsymbol{z} \in X$	(+ *is* ASSOCIATIVE)	*
4.		$\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{y} + \boldsymbol{x}$	$\forall \boldsymbol{x},\boldsymbol{y} \in X$	(+ *is* COMMUTATIVE)	*
5.		$1 \cdot \boldsymbol{x} = \boldsymbol{x}$	$\forall \boldsymbol{x} \in X$	($\cdot$ IDENTITY)	
6.		$\alpha \cdot (\beta \cdot \boldsymbol{x}) = (\alpha \cdot \beta) \cdot \boldsymbol{x}$	$\forall \alpha,\beta \in S$ *and* $\boldsymbol{x} \in X$	($\cdot$ ASSOCIATES *with* $\cdot$)	
7.		$\alpha \cdot (\boldsymbol{x} + \boldsymbol{y}) = (\alpha \cdot \boldsymbol{x}) + (\alpha \cdot \boldsymbol{y})$	$\forall \alpha \in S$ *and* $\boldsymbol{x},\boldsymbol{y} \in X$	($\cdot$ DISTRIBUTES *over* +)	
8.		$(\alpha + \beta) \cdot \boldsymbol{x} = (\alpha \cdot \boldsymbol{x}) + (\beta \cdot \boldsymbol{x})$	$\forall \alpha,\beta \in S$ *and* $\boldsymbol{x} \in X$	($\cdot$ PSEUDO-DISTRIBUTES *over* +)	

The set X *is called the* ***underlying set****. The elements of* X *are called* ***vectors****. The elements of* $\mathbb{F}$ *are called* ***scalars****. A* LINEAR SPACE *is also called a* ***vector space****. If* $\mathbb{F} \triangleq \mathbb{R}$*, then* $\boldsymbol{\Omega}$ *is a* ***real linear space****. If* $\mathbb{F} \triangleq \mathbb{C}$*, then* $\boldsymbol{\Omega}$ *is a* ***complex linear space****.*

Definition J.2. *Let* $\boldsymbol{L}_1 \triangleq \big(X_1, +, \cdot, (\mathbb{F}_1, \dot{+}, \dot{\times})\big)$ *and* $\boldsymbol{L}_2 \triangleq \big(X_2, +, \cdot, (\mathbb{F}_2, \dot{+}, \dot{\times})\big)$.

DEF

$\boldsymbol{\Omega}_2$ *is a* ***linear subspace*** *of* $\boldsymbol{\Omega}_1$ *if*

1. $\boldsymbol{L}_1$ *is a* LINEAR SPACE *(Definition J.1 page 247)* *and*
2. $\boldsymbol{L}_2$ *is a* LINEAR SPACE *(Definition J.1 page 247)* *and*
3. $\mathbb{F}_2 \subseteq \mathbb{F}_1$ *and*
4. $X_2 \subseteq X_1$ *and*

Remark J.1. [3] By the first four conditions ($*$]) listed in Definition J.1, $(X, +)$ is a **commutative group** (or **abelian group**).

[3] Akhiezer and Glazman (1993) page 1, Haaser and Sullivan (1991) page 41

Often when discussing a linear space, the operator $\cdot$ is simply expressed with juxtaposition (e.g. $\alpha\boldsymbol{x}$ is equivalent to $\alpha \cdot \boldsymbol{x}$). In doing this, there is no risk of ambiguity between scalar-vector multiplication and scalar-scalar multiplication because the operands uniquely identify the precise operator.[4]

Example J.1 (tuples in $\mathbb{F}^N$). [5] Let $\llparenthesis x_n \rrparenthesis_1^N$ be an N-*tuple* (Definition D.2 page 200) over a *field* (Definition G.5 page 222) $(\mathbb{F}, +, \cdot, 0, 1)$.

EX

$$\begin{array}{rrcll} \text{Let} & X & \triangleq & \left\{ \llparenthesis x_n \rrparenthesis_1^N \mid x_n \in \mathbb{F} \right\} & & \text{and} \\ & \llparenthesis x_n \rrparenthesis_1^N + \llparenthesis y_n \rrparenthesis_1^N & \triangleq & \llparenthesis x_n \dotplus y_n \rrparenthesis_1^N & \forall \llparenthesis x_n \rrparenthesis \in X & \text{and} \\ & \alpha \cdot \llparenthesis x_n \rrparenthesis_1^N & \triangleq & \llparenthesis \alpha \dot{\times} x_n \rrparenthesis_1^N & \forall \llparenthesis x_n \rrparenthesis \in X, \alpha \in \mathbb{F}. & \end{array}$$

Then the structure $\left(X,\ +,\ \cdot,\ (\mathbb{F},\ \dotplus,\ \dot{\times})\right)$ is a *linear space*.

Example J.2 (real numbers). [6] Let $(\mathbb{R}, +, \cdot, 0, 1)$ be the field of real numbers.

EX

The structure $(\mathbb{R},\ +,\ \cdot,\ (\mathbb{R},\ +,\ \cdot))$ is a *linear space*.
That is, the field of real numbers forms a linear space over itself.

Example J.3 (functions). [7] Let $(\mathbb{F}, +, \cdot, 0, 1)$ be a field. Let Y^X be the set of all functions with domain X and range Y.

EX

$$\begin{array}{rrcll} \text{Let} & \left[\mathsf{f} + \mathsf{g}\right](x) & \triangleq & \mathsf{f}(x) + \mathsf{g}(x) & \forall \mathsf{f},\mathsf{g} \in Y^X \quad (\textit{pointwise addition}) \quad \text{and} \\ & [\alpha \cdot \mathsf{f}](x) & \triangleq & \alpha \cdot [\mathsf{f}(x)] & \forall \mathsf{f} \in Y^X, \alpha \in \mathbb{F}. \end{array}$$

Then the structure $\left(Y^X,\ +,\ \cdot,\ (\mathbb{F},\ \dotplus,\ \dot{\times})\right)$ is a *linear space*.

Example J.4 (functions onto $\mathbb{F}$). [8] Let $(\mathbb{F}, +, \cdot, 0, 1)$ be a field. Let $\mathbb{F}^{\boldsymbol{X}}$ be the set of all functions with domain X and range $\mathbb{F}$.

EX

$$\begin{array}{rrcll} \text{Let} & \left[\mathsf{f} + \mathsf{g}\right](x) & \triangleq & \mathsf{f}(x) + \mathsf{g}(x) & \forall \mathsf{f},\mathsf{g} \in \mathbb{F}^{\boldsymbol{X}} \quad (\textit{pointwise addition}) \quad \text{and} \\ & [\alpha \cdot \mathsf{f}](x) & \triangleq & \alpha \cdot [\mathsf{f}(x)] & \forall \mathsf{f} \in \mathbb{F}^{\boldsymbol{X}}, \alpha \in \mathbb{F}. \end{array}$$

Then the structure $\left(\mathbb{F}^{\boldsymbol{X}},\ +,\ \cdot,\ (\mathbb{F},\ +,\ \cdot)\right)$ is a *linear space*.

Theorem J.1 (Additive identity properties). [9] *Let* $\left(X,\ +,\ \cdot,\ (\mathbb{F},\ \dotplus,\ \dot{\times})\right)$ *be a linear space,* 0 *the* ADDITIVE IDENTITY ELEMENT *(Definition G.1 page 221) with respect to* $\dotplus$*, and* $\mathbb{0}$ *the* ADDITIVE IDENTITY ELEMENT *with respect to* $+$.

THM

1.	$0\boldsymbol{x}$	$=$	$\mathbb{0}$		$\forall \boldsymbol{x} \in X$
2.	$\alpha\mathbb{0}$	$=$	$\mathbb{0}$		$\forall \alpha \in \mathbb{F}$
3.	$\alpha\boldsymbol{x}$	$=$	$\mathbb{0}$	$\implies \alpha = 0$ *or* $\boldsymbol{x} = \mathbb{0}$	
4.	$\boldsymbol{x} + \boldsymbol{x}$	$=$	$\boldsymbol{x}$	$\implies \boldsymbol{x} = \mathbb{0}$	
5.	$\alpha \neq 0$ *and* $\boldsymbol{x} \neq \mathbb{0}$			$\implies \alpha\boldsymbol{x} \neq \mathbb{0}$	

✎PROOF:

[4] *Operator overload* is a technique in which two fundamentally different operators or functions share the same symbol or label. It is inherent in the programming language C++ and is therein called *operator overload*. In C++, you can define two (or more) operators or functions that share the same symbol or name, but yet are completely different. Two such operators (or functions) are distinguished from each other by the type of their operands. So for example, in C++, you can define an $m \times n$ matrix *type* and use operator overload to define a + operator that operates on this new matrix type. So if variables x and y are of floating point type and A and B are of the matrix type, you can then add either type using the same syntax style:

x+y (add two floating point numbers)
A+B (add two matrices)

Even though both of these operations "look" the same, they are of course fundamentally different.

[5] Kubrusly (2001) page 41 ⟨Example 2D⟩
[6] Kubrusly (2001) page 41 ⟨Example 2D⟩, Hamel (1905)
[7] Kubrusly (2001) page 42 ⟨Example 2F⟩
[8] Kubrusly (2001) page 41 ⟨Example 2E⟩
[9] Berberian (1961) page 6 ⟨Theorem 1⟩, Michel and Herget (1993) page 77

1. Proof that $0\boldsymbol{x} = \mathbb{0}$:

$$\begin{aligned}
0\boldsymbol{x} &= 0\boldsymbol{x} + 0\mathbb{0} && \text{by definition of } + \text{ additive identity element}\\
&= 0\boldsymbol{x} + 0\boldsymbol{x} + (-0\boldsymbol{x}) && \text{by definition of } + \text{ additive inverse}\\
&= (0+0)\boldsymbol{x} + (-0\cdot\boldsymbol{x}) && \text{by definition of } + \text{ additive identity element}\\
&= 0\boldsymbol{x} + (-0\boldsymbol{x}) && \text{by Definition J.1 property 4}\\
&= \mathbb{0} && \text{by definition of } + \text{ additive identity element}
\end{aligned}$$

2. Proof that $\alpha\mathbb{0} = \mathbb{0}$:

$$\begin{aligned}
\alpha\mathbb{0} &= \alpha(0\boldsymbol{x}) && \text{by item 1}\\
&= (\alpha 0)\boldsymbol{x} && \text{by Definition J.1 property 6}\\
&= 0\boldsymbol{x}\\
&= \mathbb{0} && \text{by item 1}
\end{aligned}$$

3. Proof that $\alpha \neq 0$ and $\boldsymbol{x} \neq \mathbb{0} \implies \alpha\boldsymbol{x} \neq \mathbb{0}$: Suppose $\alpha\boldsymbol{x} = \mathbb{0}$. Then

$$\begin{aligned}
\boldsymbol{x} &= \left(\frac{1}{\alpha}\alpha\right)\boldsymbol{x}\\
&= \frac{1}{\alpha}(\alpha\boldsymbol{x})\\
&= \frac{1}{\alpha}\mathbb{0}\\
&= \mathbb{0} && \text{by item 2}\\
&\implies \boldsymbol{x} = \mathbb{0}
\end{aligned}$$

This is a *contradiction* and so $\alpha\boldsymbol{x} \neq \mathbb{0}$.

4. Proof that $\alpha\boldsymbol{x} = \mathbb{0} \implies \alpha = 0$ or $\boldsymbol{x} = \mathbb{0}$: contrapositive argument of item 3

5. Proof that $\boldsymbol{x} + \boldsymbol{x} = \boldsymbol{x} \implies \boldsymbol{x} = \mathbb{0}$:

$$\begin{aligned}
\boldsymbol{x} &= \boldsymbol{x} + \mathbb{0} && \text{by \textit{additive identity} property (Definition J.1 page 247)}\\
&= \boldsymbol{x} + [\boldsymbol{x} + (-\boldsymbol{x})] && \text{by \textit{additive inverse} property (Definition J.1 page 247)}\\
&= [\boldsymbol{x} + \boldsymbol{x}] + (-\boldsymbol{x}) && \text{by \textit{associative} property (Definition J.1 page 247)}\\
&= \boldsymbol{x} + (-\boldsymbol{x}) && \text{by left hypothesis}\\
&= \mathbb{0} && \text{by \textit{additive inverse} property (Definition J.1 page 247)}
\end{aligned}$$

✏

Definition J.3. [10] *Let $\boldsymbol{\Omega} \triangleq (X, +, \cdot, (\mathbb{F}, \dotplus, \dot{\times}))$ be a linear space with vectors $\boldsymbol{x}, \boldsymbol{y} \in X$. Let $-\boldsymbol{y}$ be the additive inverse of $\boldsymbol{y}$ such that $\boldsymbol{y} + (-\boldsymbol{y}) = \mathbb{0}$.*

DEF *The **difference** of $\boldsymbol{x}$ and $\boldsymbol{y}$ is $\boldsymbol{x} + (-\boldsymbol{y})$ and is denoted $\boldsymbol{x} - \boldsymbol{y}$.*

Theorem J.2 (Additive inverse properties). [11] *Let $\boldsymbol{\Omega} \triangleq (X, +, \cdot, (\mathbb{F}, \dotplus, \dot{\times}))$ be a linear space, $\mathbb{0}$ the* ADDITIVE IDENTITY ELEMENT *with respect to $+$, and $-\boldsymbol{x}$ the* ADDITIVE INVERSE (Definition G.1 page 221) *of $\boldsymbol{x}$ with respect to $+$.*

THM

1.	$\boldsymbol{x} + \boldsymbol{y}$	$=$	$\mathbb{0} \implies \boldsymbol{x} = -\boldsymbol{y}$	$\forall \boldsymbol{x},\boldsymbol{y} \in X$	(*additive inverse is* UNIQUE)
2.	$(-\alpha)\boldsymbol{x}$	$=$	$-(\alpha\boldsymbol{x}) = \alpha(-\boldsymbol{x})$	$\forall \boldsymbol{x} \in X, \alpha \in \mathbb{F}$	
3.	$\alpha(\boldsymbol{x} - \boldsymbol{y})$	$=$	$\alpha\boldsymbol{x} - \alpha\boldsymbol{y}$	$\forall \boldsymbol{x},\boldsymbol{y} \in X, \alpha \in \mathbb{F}$	(DISTRIBUTIVE)
4.	$(\alpha - \beta)\boldsymbol{x}$	$=$	$\alpha\boldsymbol{x} - \beta\boldsymbol{x}$	$\forall \boldsymbol{x} \in X, \alpha,\beta \in \mathbb{F}$	(DISTRIBUTIVE)

[10] Berberian (1961) page 7 ⟨Definition 1⟩

[11] Berberian (1961) page 7 ⟨Corollary 1⟩, Michel and Herget (1993) page 77, Peano (1888a) ⟨Chapter IX⟩, Peano (1888b) pages 119–120, Banach (1922) pages 134–135

✎PROOF:

1. Proof that $x + y = 0 \implies x = -y$:

$$\begin{aligned} x &= x - \mathbb{0} \\ &= x - (x + y) && \text{by left hypothesis} \\ &= (x - x) - y \\ &= \mathbb{0} - y \\ &= -y \end{aligned}$$

2. Proof that $(-\alpha)x = -(\alpha x)$:

$$\begin{aligned} \mathbb{0} &= 0x && \text{by Theorem J.1 page 249} \\ &= (\alpha - \alpha)x && \text{by field property of } \mathbb{F} \\ &= [\alpha + (-\alpha)]x && \text{by field property of } \mathbb{F} \\ &= \alpha x + (-\alpha)x && \text{by Definition J.1 page 247} \\ \implies -(\alpha x) &= (-\alpha)x && \text{by item (1) page 251} \end{aligned}$$

3. Proof that $\alpha(-x) = -(\alpha x)$:

$$\begin{aligned} \mathbb{0} &= \alpha\mathbb{0} && \text{by Theorem J.1 page 249} \\ &= \alpha[x + (-x)] && \text{by definition of additive identity element } -x \\ &= \alpha x + \alpha(-x) && \text{by Definition J.1 page 247} \\ &= \alpha x + \alpha(-x) \\ \implies -(\alpha x) &= \alpha(-x) && \text{by item (1) page 251} \end{aligned}$$

4. Proof that $\alpha(x - y) = \alpha x - \alpha y$:

$$\begin{aligned} \alpha(x - y) &= \alpha\big[x + (-y)\big] && \text{by Definition J.3 page 250} \\ &= \alpha x + \alpha(-y) && \text{by Definition J.1 page 247} \\ &= \alpha x + (-\alpha y) && \text{by item (3) page 251} \\ &= \alpha x - \alpha y && \text{by Definition J.3 page 250} \end{aligned}$$

5. Proof that $(\alpha - \beta)x = \alpha x - \beta x$:

$$\begin{aligned} (\alpha - \beta)x &= [\alpha + (-\beta)]x && \text{by field properties of } \mathbb{F} \\ &= \alpha x + (-\beta)x && \text{by Definition J.1} \\ &= \alpha x + [-(\beta x)] && \text{by item (2) page 251} \\ &= \alpha x - (\beta x) && \text{by Definition J.3 page 250} \end{aligned}$$

✏

Theorem J.3. [12] *Let $\big(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times})\big)$ be a linear space, $\mathbb{0}$ the additive identity element with respect to $+$, and $-x$ additive inverse of x with respect to $+$.*

THM	1. $\alpha x = \alpha y$ and $\alpha \neq 0$	$\implies x = y$	$\forall x,y \in X$
	2. $\alpha x = \beta x$ and $x \neq \mathbb{0}$	$\implies \alpha = \beta$	$\forall x,y \in X,\ \alpha,\beta \in \mathbb{F}$
	3. $z + x = z + y$	$\implies x = y$	$\forall x,y,z \in X$

[12] Michel and Herget (1993) page 77, Peano (1888a) ⟨Chapter IX⟩, Peano (1888b) pages 119–120, Banach (1922) pages 134–135

✎PROOF:

1. Proof that $\alpha\boldsymbol{x} = \alpha\boldsymbol{y}$ and $\alpha \neq 0 \implies \boldsymbol{x} = \boldsymbol{y}$:

$$\begin{aligned}
\mathbb{0} &= \frac{1}{\alpha}(\mathbb{0}) && \text{by left hypothesis } (\alpha \neq 0)\\
&= \frac{1}{\alpha}\Big(\alpha\boldsymbol{x} - \alpha\boldsymbol{y}\Big) && \text{by left hypothesis } (\alpha\boldsymbol{x} = \alpha\boldsymbol{y})\\
&= \frac{1}{\alpha}\alpha\Big(\boldsymbol{x} - \boldsymbol{y}\Big) && \text{by Definition J.1 page 247}\\
&= \boldsymbol{x} - \boldsymbol{y}
\end{aligned}$$

2. Proof that $\alpha\boldsymbol{x} = \beta\boldsymbol{x}$ and $\boldsymbol{x} \neq \mathbb{0} \implies \alpha = \beta$:

$$\begin{aligned}
\mathbb{0} &= \alpha\boldsymbol{x} + (-\alpha\boldsymbol{x}) && \text{by definition of additive inverse}\\
&= \beta\boldsymbol{x} + (-\alpha\boldsymbol{x}) && \text{by left hypothesis}\\
&= \beta\boldsymbol{x} + (-\alpha)\boldsymbol{x} && \text{by Theorem J.2 page 250}\\
&= [\beta + (-\alpha)]\boldsymbol{x} && \text{by Definition J.1 page 247}\\
\implies \beta - \alpha &= 0 && \text{by Theorem J.1 page 249}\\
\implies \alpha &= \beta && \text{by field properties of } \mathbb{F}
\end{aligned}$$

3. Proof that $\boldsymbol{z} + \boldsymbol{x} = \boldsymbol{z} + \boldsymbol{y} \implies \boldsymbol{x} = \boldsymbol{y}$:

$$\begin{aligned}
\mathbb{0} &= (\boldsymbol{z} + \boldsymbol{x}) - (\boldsymbol{z} + \boldsymbol{y})\\
&= (\boldsymbol{x} + \boldsymbol{z}) - (\boldsymbol{z} + \boldsymbol{y}) && \text{by Definition J.1 property 1}\\
&= (\boldsymbol{x} + \boldsymbol{z}) + [(-1)\boldsymbol{z} + (-1)\boldsymbol{y}] && \text{by Definition J.1 property 3}\\
&= (\boldsymbol{x} + \boldsymbol{z}) + (-\boldsymbol{z} - \boldsymbol{y}) && \text{by previous result 2.}\\
&= \boldsymbol{x} + (\boldsymbol{z} - \boldsymbol{z}) - \boldsymbol{y}\\
&= \boldsymbol{x} - \boldsymbol{y}
\end{aligned}$$

⇰

J.2 Order on Linear Spaces

Definition J.4. [13] *Let* $\boldsymbol{\Omega} \triangleq \big(X, +, \cdot, (\mathbb{R}, \dot{+}, \dot{\times})\big)$ *be a real linear space.*

DEF

The pair $(\boldsymbol{\Omega}, \leq)$ *is an* ***ordered linear space*** *if*

1. $\boldsymbol{x} \leq \boldsymbol{y} \implies \boldsymbol{x} + \boldsymbol{z} \leq \boldsymbol{y} + \boldsymbol{z} \quad \forall z \in X$ *and*
2. $\boldsymbol{x} \leq \boldsymbol{y} \implies \alpha\boldsymbol{x} \leq \alpha\boldsymbol{y} \quad \forall \alpha \in \mathbb{F}$

A vector $\boldsymbol{x}$ *is* ***positive*** *if* $\mathbb{0} \leq \boldsymbol{x}$.
The ***positive cone*** $\boldsymbol{X}^+$ *of* $(\boldsymbol{X}, \leq)$ *is the set* $\boldsymbol{X}^+ \triangleq \{\boldsymbol{x} \in \boldsymbol{X} | \mathbb{0} \leq \boldsymbol{x}\}$.

Definition J.5. [14] *Let* $(\boldsymbol{X}, \leq)$ *be an ordered linear space.*

DEF

The tuple $\boldsymbol{L} \triangleq (\boldsymbol{X}, \vee, \wedge; \leq)$ *is a* ***Riesz space*** *if* $\boldsymbol{L}$ *is a* ***lattice***.
A RIESZ SPACE *is also called a* ***vector lattice***.

Theorem J.4. [15] *Let* $(\boldsymbol{X}, \vee, \wedge; \leq)$ *be a Riesz space* (Definition J.5 page 252).

THM

$$\begin{array}{lcl|lcl|l}
\boldsymbol{x} \vee \boldsymbol{y} &=& -\big[(-\boldsymbol{x}) \wedge (-\boldsymbol{y})\big] & \boldsymbol{x} \wedge \boldsymbol{y} &=& -\big[(-\boldsymbol{x}) \vee (-\boldsymbol{y})\big] & \forall \boldsymbol{x},\boldsymbol{y} \in \boldsymbol{X}\\
\boldsymbol{x} + (\boldsymbol{y} \vee \boldsymbol{z}) &=& (\boldsymbol{x} + \boldsymbol{y}) \vee (\boldsymbol{x} + \boldsymbol{z}) & \boldsymbol{x} + (\boldsymbol{y} \wedge \boldsymbol{z}) &=& (\boldsymbol{x} + \boldsymbol{y}) \wedge (\boldsymbol{x} + \boldsymbol{z}) & \forall \boldsymbol{x},\boldsymbol{y},\boldsymbol{z} \in \boldsymbol{X}\\
\alpha(\boldsymbol{x} \vee \boldsymbol{y}) &=& (\alpha\boldsymbol{x}) \vee (\alpha\boldsymbol{y}) & \alpha(\boldsymbol{x} \wedge \boldsymbol{y}) &=& (\alpha\boldsymbol{x}) \wedge (\alpha\boldsymbol{y}) & \forall \boldsymbol{x},\boldsymbol{y} \in \boldsymbol{X}, \alpha \geq 0\\
\boldsymbol{x} + \boldsymbol{y} &=& (\boldsymbol{x} \wedge \boldsymbol{y}) + (\boldsymbol{x} \vee \boldsymbol{y}) & & & & \forall \boldsymbol{x},\boldsymbol{y} \in \boldsymbol{X}, \alpha \in \mathbb{F}
\end{array}$$

[13] Aliprantis and Burkinshaw (2006) pages 1–2
[14] Aliprantis and Burkinshaw (2006) page 2
[15] Aliprantis and Burkinshaw (2006) page 3 ⟨Theorem 1.2⟩

✎PROOF:

1. Proof that $x \vee y = -\big[(-x) \wedge (-y)\big]$:

$$
\begin{array}{|rcl|rcl|}
\hline
(-x) \wedge (-y) & \le & -x & (-x) \wedge (-y) & \le & -y \\
x & \le & -\big[(-x) \wedge (-y)\big] & y & \le & -\big[(-x) \wedge (-y)\big] \\
x \vee y & \le & -\big[(-x) \wedge (-y)\big] & & & \\
\hline
x & \le & x \vee y & y & \le & x \vee y \\
-(x \vee y) & \le & -x & -(x \vee y) & \le & -y \\
-(x \vee y) & \le & (-x) \wedge (-y) & & & \\
-\big[(-x) \wedge (-y)\big] & \le & x \vee y & & & \\
\hline
\end{array}
$$

2. Proof that $x \wedge y = -\big[(-x) \vee (-y)\big]$:

$$
\begin{array}{|rcll|}
\hline
x \vee y & = & -\big[(-x) \wedge (-y)\big] & \text{by item (1)} \\
(-x) \vee (-y) & = & -\big[(-(-x)) \wedge (-(-y))\big] & \text{replace } x \text{ with } -x \text{ and } y \text{ with } y \\
(-x) \vee (-y) & = & -\big[x \wedge y\big] & -(-x) = x \\
-\big[x \wedge y\big] & = & (-x) \vee (-y) & \text{by symmetry of } = \text{ relation} \\
x \wedge y & = & -\big[(-x) \vee (-y)\big] & \text{multipy both sides by } -1 \\
\hline
\end{array}
$$

3. Proof that $x + (y \vee z) = (x + y) \vee (x + z)$:

$$
\begin{array}{|rcl|rcl|}
\hline
x + y & \le & x + (y \vee z) & x + z & \le & x + (y \vee z) \\
(x + y) \vee (x + z) & \le & x + (y \vee z) & & & \\
\hline
y & = & -x + (x + y) & z & = & -x + (x + z) \\
 & \le & -x + \big[(x + y) \vee (x + z)\big] & & \le & -x + \big[(x + y) \vee (x + z)\big] \\
y \vee z & \le & -x + \big[(x + y) \vee (x + z)\big] & & & \\
x + (y \vee z) & \le & (x + y) \vee (x + z) & & & \\
\hline
\end{array}
$$

4. Proof that $x + (y \wedge z) = (x + y) \wedge (x + z)$:

$$
\begin{array}{|rcl|rcl|}
\hline
x + y & \ge & x + (y \wedge z) & x + z & \ge & x + (y \wedge z) \\
(x + y) \wedge (x + z) & \ge & x + (y \wedge z) & & & \\
\hline
y & = & -x + (x + y) & z & = & -x + (x + z) \\
 & \ge & -x + \big[(x + y) \wedge (x + z)\big] & & \ge & -x + \big[(x + y) \wedge (x + z)\big] \\
y \wedge z & \ge & -x + \big[(x + y) \wedge (x + z)\big] & & & \\
x + (y \wedge z) & \ge & (x + y) \wedge (x + z) & & & \\
\hline
\end{array}
$$

5. Proof that $\alpha(x \vee y) = (\alpha x) \vee (\alpha y)$ for $\alpha \ge 0$:

$$
\begin{array}{|rcl|rcl|l|}
\hline
x & \le & x \vee y & y & \le & x \vee y & \\
\alpha x & \le & \alpha(x \vee y) & \alpha y & \le & \alpha(x \vee y) & \text{by Definition J.4 page 252} \\
(\alpha x) \vee (\alpha y) & \le & \alpha(x \vee y) & & & & \\
\hline
\alpha x & \le & (\alpha x) \vee (\alpha y) & \alpha y & \le & (\alpha x) \vee (\alpha y) & \\
x & \le & \alpha^{-1}(\alpha x) \vee (\alpha y) & y & \le & \alpha^{-1}(\alpha x) \vee (\alpha y) & \\
x \vee y & \le & \alpha^{-1}(\alpha x) \vee (\alpha y) & & & & \\
\alpha(x \vee y) & \le & (\alpha x) \vee (\alpha y) & & & & \\
\hline
\end{array}
$$

6. Proof that $\alpha(x \wedge y) = (\alpha x) \wedge (\alpha y)$ for $\alpha \ge 0$:

$$
\begin{array}{|rcl|rcl|l|}
\hline
x & \ge & x \wedge y & y & \ge & x \wedge y & \\
\alpha x & \ge & \alpha(x \wedge y) & \alpha y & \ge & \alpha(x \wedge y) & \text{by Definition J.4 page 252} \\
(\alpha x) \wedge (\alpha y) & \ge & \alpha(x \wedge y) & & & & \\
\hline
\end{array}
$$

$\begin{array}{rcl} \alpha\boldsymbol{x} & \geq & (\alpha\boldsymbol{x}) \wedge (\alpha\boldsymbol{y}) \\ \boldsymbol{x} & \geq & \alpha^{-1}(\alpha\boldsymbol{x}) \wedge (\alpha\boldsymbol{y}) \\ \boldsymbol{x}\wedge\boldsymbol{y} & \geq & \alpha^{-1}(\alpha\boldsymbol{x}) \wedge (\alpha\boldsymbol{y}) \\ \alpha(\boldsymbol{x}\wedge\boldsymbol{y}) & \geq & (\alpha\boldsymbol{x}) \wedge (\alpha\boldsymbol{y}) \end{array}$	$\begin{array}{rcl} \alpha\boldsymbol{y} & \geq & (\alpha\boldsymbol{x}) \wedge (\alpha\boldsymbol{y}) \\ \boldsymbol{y} & \geq & \alpha^{-1}(\alpha\boldsymbol{x}) \wedge (\alpha\boldsymbol{y}) \end{array}$	

7. Proof that $\boldsymbol{x}+\boldsymbol{y} = (\boldsymbol{x}\wedge\boldsymbol{y}) + (\boldsymbol{x}\vee\boldsymbol{y})$:

$\begin{array}{rcl} \boldsymbol{x} & \leq & \boldsymbol{x}\vee\boldsymbol{y} \\ \boldsymbol{x}+\boldsymbol{y} & \leq & (\boldsymbol{x}\vee\boldsymbol{y}) + \boldsymbol{y} \\ \boldsymbol{x}+\boldsymbol{y}-(\boldsymbol{x}\vee\boldsymbol{y}) & \leq & \boldsymbol{y} \\ \boldsymbol{x}+\boldsymbol{y}-(\boldsymbol{x}\vee\boldsymbol{y}) & \leq & \boldsymbol{x}\wedge\boldsymbol{y} \\ \boldsymbol{x}+\boldsymbol{y} & \leq & (\boldsymbol{x}\vee\boldsymbol{y}) + (\boldsymbol{x}\wedge\boldsymbol{y}) \end{array}$	$\begin{array}{rcl} \boldsymbol{y} & \leq & \boldsymbol{x}\vee\boldsymbol{y} \\ \boldsymbol{x}+v\boldsymbol{y} & \leq & \boldsymbol{x} + (\boldsymbol{x}\vee\boldsymbol{y}) \\ \boldsymbol{x}+v\boldsymbol{y}-(\boldsymbol{x}\vee\boldsymbol{y}) & \leq & \boldsymbol{x} \end{array}$
$\begin{array}{rcl} \boldsymbol{x}\wedge\boldsymbol{y} & \leq & \boldsymbol{x} \\ 0 & \leq & \boldsymbol{x}-(\boldsymbol{x}\wedge\boldsymbol{y}) \\ \boldsymbol{y} & \leq & \boldsymbol{y}+\boldsymbol{x}-(\boldsymbol{x}\wedge\boldsymbol{y}) \\ \boldsymbol{y} & \leq & \boldsymbol{x}+\boldsymbol{y}-(\boldsymbol{x}\wedge\boldsymbol{y}) \\ \boldsymbol{x}\vee\boldsymbol{y} & \leq & \boldsymbol{x}+\boldsymbol{y}-(\boldsymbol{x}\wedge\boldsymbol{y}) \\ (\boldsymbol{x}\wedge\boldsymbol{y})+(\boldsymbol{x}\vee\boldsymbol{y}) & \leq & \boldsymbol{x}+\boldsymbol{y} \end{array}$	$\begin{array}{rcl} \boldsymbol{x}\wedge\boldsymbol{y} & \leq & \boldsymbol{y} \\ 0 & \leq & \boldsymbol{y}-(\boldsymbol{x}\wedge\boldsymbol{y}) \\ \boldsymbol{x} & \leq & \boldsymbol{x}+\boldsymbol{y}-(\boldsymbol{x}\wedge\boldsymbol{y}) \\ \boldsymbol{x} & \leq & \boldsymbol{x}+\boldsymbol{y}-(\boldsymbol{x}\wedge\boldsymbol{y}) \end{array}$

✏

Definition J.6. [16] *Let* $(\boldsymbol{X}, \vee, \wedge; \leq)$ *be a* Riesz space (Definition J.5 page 252).

DEF

$\boldsymbol{x}^+$	*is defined as* $\boldsymbol{x}^+ \triangleq \boldsymbol{x}\vee\mathbb{0}$	*and is called the* ***positive part*** *of* $\boldsymbol{x}$.
$\boldsymbol{x}^-$	*is defined as* $\boldsymbol{x}^- \triangleq (-\boldsymbol{x})\vee\mathbb{0}$	*and is called the* ***negative part*** *of* $\boldsymbol{x}$.
$\lvert\boldsymbol{x}\rvert$	*is defined as* $\lvert\boldsymbol{x}\rvert \triangleq \boldsymbol{x}\vee(-\boldsymbol{x})$	*and is called the* ***absolute value*** *of* $\boldsymbol{x}$.

Theorem J.5. [17] *Let* $(\boldsymbol{X}, \vee, \wedge; \leq)$ *be a* Riesz space (Definition J.5 page 252).

THM

$$\left.\begin{array}{rcl} \boldsymbol{y}-\boldsymbol{z} & = & \boldsymbol{x} \quad \text{and} \\ \boldsymbol{y}\wedge\boldsymbol{z} & = & \mathbb{0} \end{array}\right\} \iff \left\{\begin{array}{rcl} \boldsymbol{y} & = & \boldsymbol{x}^+ \quad \text{and} \\ \boldsymbol{z} & = & \boldsymbol{x}^- \end{array}\right.$$

✎Proof:

1. Proof that left hypothesis $\implies$ right hypothesis:

$$\begin{array}{rll}
\boldsymbol{x}^+ & = \boldsymbol{x}\vee\mathbb{0} & \text{by definition of } \boldsymbol{x}^+ \text{ Definition J.6 page 254} \\
& = (\boldsymbol{y}-\boldsymbol{z})\vee\mathbb{0} & \text{by left hypothesis} \\
& = (\boldsymbol{y}-\boldsymbol{z})\vee(\boldsymbol{z}-\boldsymbol{z}) & \\
& = (\boldsymbol{y}\vee\boldsymbol{z})-\boldsymbol{z} & \text{by Theorem J.4 page 252} \\
& = \left[\boldsymbol{y}+\boldsymbol{z}-(\boldsymbol{y}\wedge\boldsymbol{z})\right]-\boldsymbol{z} & \text{by Theorem J.4 page 252} \\
& = \boldsymbol{y}-(\boldsymbol{y}\wedge\boldsymbol{z}) & \\
& = \boldsymbol{y}-\mathbb{0} & \text{by left hypothesis} \\
& = \boldsymbol{y} & \\
\boldsymbol{x}^- & = (-\boldsymbol{x})\vee\mathbb{0} & \text{by definition of } \boldsymbol{x}^- \text{ Definition J.6 page 254} \\
& = (\boldsymbol{z}-\boldsymbol{y})\vee\mathbb{0} & \text{by left hypothesis} \\
& = (\boldsymbol{z}-\boldsymbol{y})\vee(\boldsymbol{y}-\boldsymbol{y}) & \\
& = (\boldsymbol{z}\vee\boldsymbol{y})-\boldsymbol{y} & \text{by Theorem J.4 page 252}
\end{array}$$

[16] Aliprantis and Burkinshaw (2006) page 4, Istrăţescu (1987) page 129
[17] Aliprantis and Burkinshaw (2006) page 4 ⟨Theorem 1.3⟩

$$\begin{aligned}
&= \left[z + y - (z \wedge y)\right] - z && \text{by Theorem J.4 page 252}\\
&= z - (z \wedge y)\\
&= z - \mathbb{0} && \text{by left hypothesis}\\
&= z
\end{aligned}$$

2. Proof that left hypothesis $\Longleftarrow$ right hypothesis:

$$\begin{aligned}
y - z &= x^+ - x^- && \text{by right hypothesis}\\
&= [x \vee \mathbb{0}] - [(-x) \vee \mathbb{0}] && \text{by Definition J.6 page 254}\\
&= (x \vee \mathbb{0}) + (x \wedge \mathbb{0}) && \text{by Theorem J.4 page 252}\\
&= x && \text{by Theorem J.4 page 252}\\
y \wedge z &= x^+ \wedge x^- && \text{by right hypothesis}\\
&= \left[x^- + \left(x^+ - x^-\right)\right] \wedge [x^- + \mathbb{0}]\\
&x^- + \left[\left(x^+ - x^-\right) \wedge \mathbb{0}\right] && \text{by Theorem J.4 page 252}\\
&x^- + \left[(y - z) \wedge \mathbb{0}\right] && \text{by right hypothesis}\\
&x^- + (x \wedge \mathbb{0}) && \text{by previous result}\\
&x^- - [-x \vee \mathbb{0}] && \text{by Theorem J.4 page 252}\\
&x^- - x^- && \text{by definition of } x^- \text{ (Definition J.6 page 254)}\\
&\mathbb{0}
\end{aligned}$$

✏

Theorem J.6. [18] *Let* $(\boldsymbol{X}, \vee, \wedge; \leq)$ *be a* RIESZ SPACE (Definition J.5 page 252). *Let* x^+ *the* POSITIVE PART *of* $x \in \boldsymbol{X}$, x^- *the* NEGATIVE PART *of* $x \in \boldsymbol{X}$, *and* $|x|$ *the* ABSOLUTE VALUE (Definition J.6 page 254) *of* $x \in \boldsymbol{X}$.

THM

$$\begin{aligned}
|x| &= x^+ + x^- && \forall x \in \boldsymbol{X}\\
x &= (x - y)^+ + (x \wedge y) && \forall x \in \boldsymbol{X}
\end{aligned}$$

✎PROOF:

$$\begin{aligned}
|x| &= x \vee (-x) && \text{by definition of } |x| \text{ (Definition J.6 page 254)}\\
&= (2x - x) \vee (\mathbb{0} - x)\\
&= (-x + 2x) \vee (-x + \mathbb{0}) && \text{by \textit{commutative} property (Definition J.1 page 247)}\\
&= (-x) + (2x \vee \mathbb{0}) && \text{by Theorem J.4 page 252}\\
&= (2x \vee 2\mathbb{0}) - x && \text{by the \textit{commutative} property (Definition J.1 page 247)}\\
&= 2(x \vee \mathbb{0}) - x && \text{by Theorem J.4 page 252}\\
&= 2x^+ - x && \text{by definition of } x^+ \text{ (Definition J.6 page 254)}\\
&= 2x^+ - \left(x^+ - x^-\right) && \text{by 1}\\
&= x^+ + x^-\\
x &= x + \mathbb{0} && x + y - y\\
&= (x \vee y) + (x \wedge y) - y && \text{by Theorem J.4 page 252}\\
&= \left[(x - y) \vee (y - y)\right] + (x \wedge y) && \text{by Theorem J.4 page 252}\\
&= \left[(x - y) \vee \mathbb{0}\right] + (x \wedge y)\\
&= (x - y)^+ + (x \wedge y) && \text{by definition of } x^+ \text{ (Definition J.6 page 254)}
\end{aligned}$$

✏

[18] Aliprantis and Burkinshaw (2006) page 4

Theorem J.7. [19] *Let* $(\boldsymbol{X}, \vee, \wedge; \leq)$ *be a* RIESZ SPACE (Definition J.5 page 252). *Let* $\boldsymbol{x}^+$ *the* POSITIVE PART *of* $\boldsymbol{x} \in \boldsymbol{X}$, $\boldsymbol{x}^-$ *the* NEGATIVE PART *of* $\boldsymbol{x} \in \boldsymbol{X}$, *and* $|\boldsymbol{x}|$ *the* ABSOLUTE VALUE (Definition J.6 page 254) *of* $\boldsymbol{x} \in \boldsymbol{X}$.

THM

$$\begin{array}{llcll}
1. & \boldsymbol{x} \vee \boldsymbol{y} & = & \frac{1}{2}(\boldsymbol{x} + \boldsymbol{y} + |\boldsymbol{x} - \boldsymbol{y}|) & \forall \boldsymbol{x},\boldsymbol{y} \in \boldsymbol{X} \\
2. & \boldsymbol{x} \wedge \boldsymbol{y} & = & \frac{1}{2}(\boldsymbol{x} + \boldsymbol{y} - |\boldsymbol{x} - \boldsymbol{y}|) & \forall \boldsymbol{x},\boldsymbol{y} \in \boldsymbol{X} \\
3. & |\boldsymbol{x} - \boldsymbol{y}| & = & (\boldsymbol{x} \vee \boldsymbol{y}) - (\boldsymbol{x} \wedge \boldsymbol{y}) & \forall \boldsymbol{x},\boldsymbol{y} \in \boldsymbol{X} \\
4. & |\boldsymbol{x}| \vee |\boldsymbol{y}| & = & \frac{1}{2}(|\boldsymbol{x} + \boldsymbol{y}| + |\boldsymbol{x} - \boldsymbol{y}|) & \forall \boldsymbol{x},\boldsymbol{y} \in \boldsymbol{X} \\
5. & |\boldsymbol{x}| \vee |\boldsymbol{y}| & = & \frac{1}{2}\big||\boldsymbol{x} + \boldsymbol{y}| - |\boldsymbol{x} - \boldsymbol{y}|\big| & \forall \boldsymbol{x},\boldsymbol{y} \in \boldsymbol{X}
\end{array}$$

✎PROOF:

$$\begin{aligned}
(\boldsymbol{x} + \boldsymbol{y} + |\boldsymbol{x} - \boldsymbol{y}|) &= (\boldsymbol{x} + \boldsymbol{y}) + \big[(\boldsymbol{x} - \boldsymbol{y}) \vee (\boldsymbol{y} - \boldsymbol{x})\big] && \text{by Definition J.6 page 254} \\
&= \big[(\boldsymbol{x} + \boldsymbol{y}) + (\boldsymbol{x} - \boldsymbol{y})\big] \vee \big[(\boldsymbol{x} + \boldsymbol{y}) + (\boldsymbol{y} - \boldsymbol{x})\big] && \text{by Theorem J.4 page 252} \\
&= (2\boldsymbol{x}) \vee (2\boldsymbol{y}) \\
&= 2(\boldsymbol{x} \vee \boldsymbol{y}) && \text{by Theorem J.4 page 252} \\
(\boldsymbol{x} + \boldsymbol{y} - |\boldsymbol{x} - \boldsymbol{y}|) &= (\boldsymbol{x} + \boldsymbol{y}) - \big[(\boldsymbol{x} - \boldsymbol{y}) \vee (\boldsymbol{y} - \boldsymbol{x})\big] && \text{by Definition J.6 page 254} \\
&= (\boldsymbol{x} + \boldsymbol{y}) - \big[(-(\boldsymbol{y} - \boldsymbol{x})) \vee (-(\boldsymbol{x} - \boldsymbol{y}))\big] \\
&= (\boldsymbol{x} + \boldsymbol{y}) + \big[(\boldsymbol{y} - \boldsymbol{x}) \wedge (\boldsymbol{x} - \boldsymbol{y})\big] && \text{by Theorem J.4 page 252} \\
&= \big[(\boldsymbol{x} + \boldsymbol{y}) + (\boldsymbol{y} - \boldsymbol{x})\big] \wedge \big[(\boldsymbol{x} + \boldsymbol{y}) + (\boldsymbol{x} - \boldsymbol{y})\big] && \text{by Theorem J.4 page 252} \\
&= (2\boldsymbol{y}) \wedge (2\boldsymbol{x}) \\
&= 2(\boldsymbol{y} \wedge \boldsymbol{x}) && \text{by Theorem J.4 page 252} \\
&= 2(\boldsymbol{x} \wedge \boldsymbol{y}) \\
|\boldsymbol{x} - \boldsymbol{y}| &= \frac{1}{2}(\boldsymbol{x} + \boldsymbol{y} + |\boldsymbol{x} - \boldsymbol{y}|) - \frac{1}{2}(\boldsymbol{x} + \boldsymbol{y} + |\boldsymbol{x} - \boldsymbol{y}|) \\
&= (\boldsymbol{x} \vee \boldsymbol{y}) - (\boldsymbol{x} \wedge \boldsymbol{y}) && \text{by 1 and 2} \\
|\boldsymbol{x} + \boldsymbol{y}| + |\boldsymbol{x} - \boldsymbol{y}| &= \frac{1}{2}(\mathbb{0} + |2\boldsymbol{x} + 2\boldsymbol{y}|) + |\boldsymbol{x} - \boldsymbol{y}| \\
&= \frac{1}{2}\big[(\boldsymbol{x} + \boldsymbol{y}) + (-\boldsymbol{x} - \boldsymbol{y}) + |(\boldsymbol{x} + \boldsymbol{y}) - (-\boldsymbol{x} - \boldsymbol{y})|\big] + |\boldsymbol{x} - \boldsymbol{y}| \\
&= \big[(\boldsymbol{x} + \boldsymbol{y}) \vee (-\boldsymbol{x} - \boldsymbol{y})\big] + |\boldsymbol{x} - \boldsymbol{y}| && \text{by 1} \\
&= \big[(\boldsymbol{x} + \boldsymbol{y}) + |\boldsymbol{x} - \boldsymbol{y}|\big] \vee \big[(-\boldsymbol{x} - \boldsymbol{y}) + |\boldsymbol{x} - \boldsymbol{y}|\big] && \text{by Theorem J.4 page 252} \\
&= 2(\boldsymbol{x} \vee \boldsymbol{y}) \vee 2\big[(-\boldsymbol{y}) + (-\boldsymbol{x}) + |(-\boldsymbol{y}) - (-\boldsymbol{x})|\big] && \text{by 1} \\
&= 2(\boldsymbol{x} \vee \boldsymbol{y}) \vee 2\big[(-\boldsymbol{y}) \vee (-\boldsymbol{x})\big] && \text{by 1} \\
&= 2([\boldsymbol{x} \vee (-\boldsymbol{x})] \vee (\boldsymbol{y} \vee (-\boldsymbol{y}))) \\
&= 2(|\boldsymbol{x}| \vee |\boldsymbol{y}|) && \text{by Definition J.6 page 254} \\
\big||\boldsymbol{x} + \boldsymbol{y}| - |\boldsymbol{x} - \boldsymbol{y}|\big| &= 2(|\boldsymbol{x} + \boldsymbol{y}| \vee |\boldsymbol{x} - \boldsymbol{y}|) - (|\boldsymbol{x} + \boldsymbol{y}| + |\boldsymbol{x} - \boldsymbol{y}|) && \text{by 1} \\
&= (|\boldsymbol{x} + \boldsymbol{y} + \boldsymbol{x} - \boldsymbol{y}| + |\boldsymbol{x} + \boldsymbol{y} - \boldsymbol{x} + \boldsymbol{y}|) - 2(|\boldsymbol{x}| \vee |\boldsymbol{y}|) && \text{by 3} \\
&= 2(|\boldsymbol{x}| + |\boldsymbol{y}|) - 2(|\boldsymbol{x}| \vee |\boldsymbol{y}|) \\
&= 2(|\boldsymbol{x}| \vee |\boldsymbol{y}|) && \text{by Theorem J.4 page 252}
\end{aligned}$$

✏

Definition J.7. [20] *Let* $(\boldsymbol{X}, \vee, \wedge; \leq)$ *be a* RIESZ SPACE (Definition J.5 page 252). *Let* $\boldsymbol{x}^+$ *the* POSITIVE PART *of* $\boldsymbol{x} \in \boldsymbol{X}$, $\boldsymbol{x}^-$ *the* NEGATIVE PART *of* $\boldsymbol{x} \in \boldsymbol{X}$, *and* $|\boldsymbol{x}|$ *the* ABSOLUTE VALUE (Definition J.6 page 254) *of* $\boldsymbol{x} \in \boldsymbol{X}$.

DEF

$\boldsymbol{x}$ *and* $\boldsymbol{y}$ *are* ***disjoint***, *denoted by* $\boldsymbol{x} \perp \boldsymbol{y}$, *if*

$$|\boldsymbol{x}| \wedge |\boldsymbol{y}| = \mathbb{0}.$$

Two subsets $\boldsymbol{U}$ *and* $\boldsymbol{V}$ *of* $\boldsymbol{X}$ *are* ***disjoint***, *denoted by* $\boldsymbol{U} \perp \boldsymbol{V}$ *if*

$$\boldsymbol{x} \perp \boldsymbol{y} \qquad \forall \boldsymbol{x} \in \boldsymbol{U} \text{ and } \boldsymbol{y} \in \boldsymbol{V}$$

[19] Aliprantis and Burkinshaw (2006) page 5 ⟨Theorem 1.4⟩

[20] Aliprantis and Burkinshaw (2006) page 5

Definition J.8. [21] *Let* $(\boldsymbol{X}, \vee, \wedge; \le)$ *be a* RIESZ SPACE (Definition J.5 page 252). *Let* $\boldsymbol{x}^+$ *the* POSITIVE PART *of* $\boldsymbol{x} \in \boldsymbol{X}$, $\boldsymbol{x}^-$ *the* NEGATIVE PART *of* $\boldsymbol{x} \in \boldsymbol{X}$, *and* $|\boldsymbol{x}|$ *the* ABSOLUTE VALUE (Definition J.6 page 254) *of* $\boldsymbol{x} \in \boldsymbol{X}$. *Let* $\boldsymbol{Y}$ *be a subset of* $\boldsymbol{X}$.

DEF

$\boldsymbol{Y}^d$ *is the* ***disjoint complement*** *of* $\boldsymbol{Y}$ *if* $\quad \boldsymbol{Y}^d \triangleq \{\boldsymbol{x} \in \boldsymbol{X} | \boldsymbol{x} \perp \boldsymbol{y} \quad \forall \boldsymbol{y} \in \boldsymbol{Y}\}$.

The quantity $\boldsymbol{Y}^{dd}$ *is defined as* $\left(\boldsymbol{Y}^d\right)^d$.

Definition J.9. [22] *Let* $(\boldsymbol{X}, \vee, \wedge; \le)$ *be a* RIESZ SPACE (Definition J.5 page 252). *Let* $\boldsymbol{x}^+$ *the* POSITIVE PART *of* $\boldsymbol{x} \in \boldsymbol{X}$, $\boldsymbol{x}^-$ *the* NEGATIVE PART *of* $\boldsymbol{x} \in \boldsymbol{X}$, *and* $|\boldsymbol{x}|$ *the* ABSOLUTE VALUE (Definition J.6 page 254) *of* $\boldsymbol{x} \in \boldsymbol{X}$.

DEF

$$\begin{aligned}
|A| &\triangleq \{|\boldsymbol{a}| | \boldsymbol{a} \in A\} \\
A^+ &\triangleq \left\{\boldsymbol{a}^+ | \boldsymbol{a} \in A\right\} \\
A^- &\triangleq \{\boldsymbol{a}^- | \boldsymbol{a} \in A\} \\
A \vee B &\triangleq \left\{\boldsymbol{a} \vee \boldsymbol{b} | \boldsymbol{a} \in A \text{ and } \boldsymbol{b} \in B\right\} \\
A \wedge B &\triangleq \left\{\boldsymbol{a} \wedge \boldsymbol{b} | \boldsymbol{a} \in A \text{ and } \boldsymbol{b} \in B\right\} \\
\boldsymbol{x} \vee A &\triangleq \{\boldsymbol{x} \vee \boldsymbol{a} | \boldsymbol{a} \in A\} \\
\boldsymbol{x} \wedge A &\triangleq \{\boldsymbol{x} \wedge \boldsymbol{a} | \boldsymbol{a} \in A\}
\end{aligned}$$

[21] Aliprantis and Burkinshaw (2006) page 5
[22] Aliprantis and Burkinshaw (2006) page 7

APPENDIX K

LINEAR COMBINATIONS

K.1 Linear combinations in linear spaces

A *linear space* (Definition J.1 page 247) in general is not equipped with a *topology*. Without a topology, it is not possible to determine whether an *infinite sum* of vectors converges. Therefore in this section (dealing with linear spaces), all definitions related to sums of vectors will be valid for *finite* sums(Definition Q.1 page 345) only (finite "N").

Definition K.1. [1] *Let* $\left\{ \boldsymbol{x}_n \in X \,\middle|\, n=1,2,\ldots,N \right\}$ *be a set of vectors in a* LINEAR SPACE $\left(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times})\right)$.

DEF: *A vector* $\boldsymbol{x} \in X$ *is a* ***linear combination*** *of the vectors in* $\{\boldsymbol{x}_n\}$ *if there exists* $\left\{ \alpha_n \in \mathbb{F} \,\middle|\, n=1,2,\ldots,N \right\}$ *such that*

$$\boldsymbol{x} = \sum_{n=1}^{N} \alpha_n \boldsymbol{x}_n .$$

Definition K.2. [2] *Let* $\left(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times})\right)$ *be a linear space and* Y *be a subset of* X.

DEF: *The* ***linear span*** *of* Y *is defined as*

$$\operatorname{span} Y \triangleq \left\{ \sum_{\gamma \in \Gamma} \alpha_\gamma \boldsymbol{y}_\gamma \,\middle|\, \alpha_\gamma \in \mathbb{F},\ \boldsymbol{y}_\gamma \in Y \right\}.$$

The set Y ***spans*** *a set* A *if* $A \subseteq \operatorname{span} Y$.

Proposition K.1. [3] *Let* $\left\{ \boldsymbol{x}_n \in X \,\middle|\, n=1,2,\ldots,N \right\}$ *be a set of vectors in a* LINEAR SPACE $\boldsymbol{L} \triangleq \left(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times})\right)$.

PRP:
1. $\operatorname{span}\{\boldsymbol{x}_n\}$ *is a* LINEAR SPACE (*Definition J.1 page 247*) *and*
2. $\operatorname{span}\{\boldsymbol{x}_n\}$ *is a* LINEAR SUBSPACE *of* $\boldsymbol{L}$ (*Definition J.2 page 248*).

Definition K.3. [4] *Let* $\boldsymbol{L} \triangleq \left(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times})\right)$ *be a* LINEAR SPACE.

DEF: *The set* $Y \triangleq \left\{ \boldsymbol{x}_n \in X \,\middle|\, n=1,2,\ldots,N \right\}$ *is* ***linearly independent*** *in* $\boldsymbol{L}$ *if*

$$\left\{ \sum_{n=1}^{N} \alpha_n \boldsymbol{x}_n = 0 \right\} \implies \left\{ \alpha_1 = \alpha_2 = \cdots = \alpha_N = 0 \right\}.$$

The set Y *is* ***linearly dependent*** *in* $\boldsymbol{L}$ *if* Y *is not linearly independent in* $\boldsymbol{L}$.

[1] Berberian (1961) page 11 ⟨Definition I.4.1⟩, Kubrusly (2001) page 46

[2] Michel and Herget (1993) page 86 ⟨3.3.7 Definition⟩, Kurdila and Zabarankin (2005) page 44, Searcóid (2002) page 71 ⟨Definition 3.2.5—more general definition⟩

[3] Kubrusly (2001) page 46

[4] Bachman and Narici (1966) pages 3–4, Christensen (2003) page 2, Heil (2011) page 156 ⟨Definition 5.7⟩

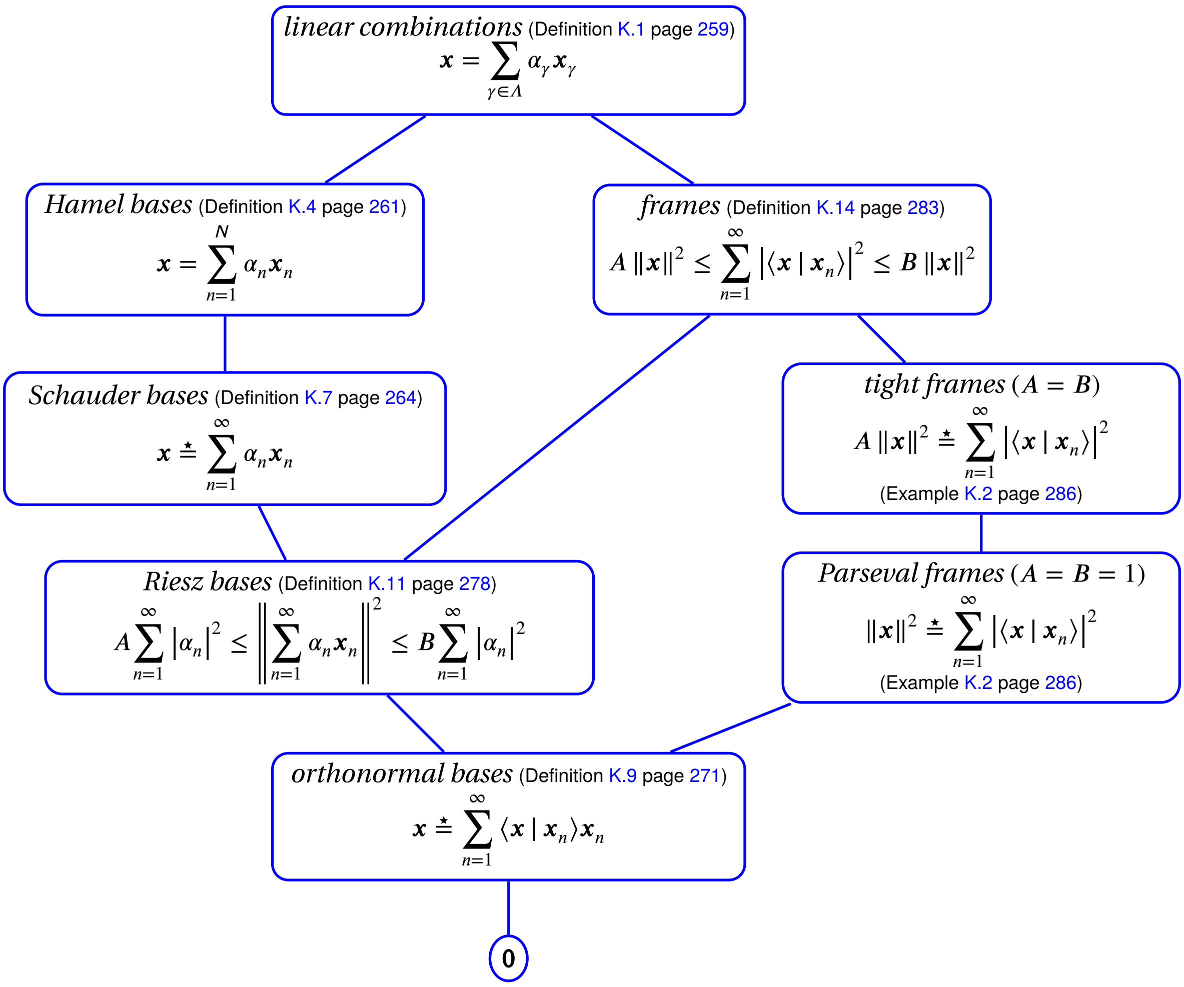


Figure K.1: Lattice of *linear combinations*

Definition K.4. [5] *Let* $\{x_n \in X | n=1,2,\ldots,N\}$ *be a set of vectors in a* LINEAR SPACE $\mathbf{L} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$.

DEF: *The set* $\{x_n\}$ *is a* ***Hamel basis*** *for* $\mathbf{L}$ *if*

1. $\{x_n\}$ SPANS $\mathbf{L}$ (Definition K.2 page 259) *and*
2. $\{x_n\}$ *is* LINEARLY INDEPENDENT *in* $\mathbf{L}$ (Definition K.1 page 259) .

A HAMEL BASIS *is also called a* ***linear basis***.

Definition K.5. [6] *Let* $\mathbf{L} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$ *be a* LINEAR SPACE. *Let* x *be a* VECTOR *in* $\mathbf{L}$ *and* $Y \triangleq \{x_n \in X | n=1,2,\ldots,N\}$ *be a set of vectors in* $\mathbf{L}$.

DEF: *The expression* $\sum_{n=1}^{N} \alpha_n x_n$ *is the* ***expansion*** *of* x *on* Y *in* $\mathbf{L}$ *if* $x = \sum_{n=1}^{N} \alpha_n x_n$.
In this case, the sequence $(\alpha_n)_{n=1}^{N}$ *is the* ***coordinates*** *of* x *with respect to* Y *in* $\mathbf{L}$.
If $\alpha_N \neq 0$, *then* N *is the* ***dimension*** $\dim \mathbf{L}$ *of* $\mathbf{L}$.

Theorem K.1. [7] *Let* $\{x_n | n=1,2,\ldots,N\}$ *be a* HAMEL BASIS (Definition K.4 page 261) *for a* LINEAR SPACE $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$.

THM:
$$\left\{ x = \sum_{n=1}^{N} \alpha_n x_n = \sum_{n=1}^{N} \beta_n x_n \right\} \implies \underbrace{\alpha_n = \beta_n \quad \forall n=1,2,\ldots,N}_{\text{coordinates of } x \text{ are UNIQUE}} \qquad \forall x \in X$$

✎PROOF:

$$\begin{aligned}
\mathbb{0} &= x - x \\
&= \sum_{n=1}^{N} \alpha_n x_n - \sum_{n=1}^{N} \beta_n x_n \\
&= \sum_{n=1}^{N} (\alpha_n - \beta_n) x_n \\
&\implies \{x_n\} \text{ is } \textit{linearly dependent} \text{ if } (\alpha_n - \beta_n) \neq 0 \qquad \forall n = 1,2,\ldots,N \\
&\implies (\alpha_n - \beta_n) = 0 \qquad \forall n = 1,2,\ldots,N \qquad \text{(because } \{x_n\} \text{ is a } \textit{basis} \text{ and therefore must be } \textit{linearly independent}\text{)} \\
&\implies \alpha_n = \beta_n \text{ for } n = 1,2,\ldots,N
\end{aligned}$$

✏

Theorem K.2. [8] *Let* $\mathbf{L} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$ *be a* LINEAR SPACE.

THM:
$$\left\{ \begin{array}{ll} 1. & \{x_n \in X | n=1,2,\ldots,N\} \text{ is a HAMEL BASIS for } \mathbf{L} \quad \text{and} \\ 2. & \{y_n \in X | n=1,2,\ldots,M\} \text{ is a set of LINEARLY INDEPENDENT vectors in } \mathbf{L} \end{array} \right\}$$
$$\implies \left\{ \begin{array}{ll} 1. & M \leq N \quad \text{and} \\ 2. & M = N \implies \{y_n | n=1,2,\ldots,M\} \text{ is a BASIS for } \mathbf{L} \quad \text{and} \\ 3. & M \neq N \implies \{y_n | n=1,2,\ldots,M\} \text{ is NOT a basis for } \mathbf{L} \end{array} \right\}$$

✎PROOF:

[5] Hamel (1905), Bachman and Narici (1966) page 4, Kubrusly (2001) pages 48–49 ⟨Section 2.4⟩, Young (2001) page 1, Carothers (2005) page 25, Heil (2011) page 125 ⟨Definition 4.1⟩
[6] Hamel (1905), Bachman and Narici (1966) page 4, Kubrusly (2001) pages 48–49 ⟨Section 2.4⟩, Young (2001) page 1, Carothers (2005) page 25, Heil (2011) page 125 ⟨Definition 4.1⟩
[7] Michel and Herget (1993) pages 89–90 ⟨Theorem 3.3.25⟩
[8] Michel and Herget (1993) pages 90–91 ⟨Theorem 3.3.26⟩

1. Proof that $\{y_1, x_1, \ldots, x_{N-1}\}$ is a *basis* for L:

   (a) Proof that $\{y_1, x_1, \ldots, x_{N-1}\}$ *spans* L:

      i. Because $\{x_n|_{n=1,2,\ldots,N}\}$ is a *basis* for L, there exists $\beta \in \mathbb{F}$ and $\{\alpha_n \in \mathbb{F}|_{n=1,2,\ldots,N}\}$ such that
      $$\beta y_1 + \sum_{n=1}^{N} \alpha_n x_n = 0.$$

      ii. Select an n such that $\alpha_n \neq 0$ and renumber (if necessary) the above indices such that
      $$x_n = -\frac{\beta}{\alpha_n} y_1 + \sum_{n=1}^{N-1} \frac{\alpha_n}{\alpha_n} x_n.$$

      iii. Then, for any $y \in X$, we can write
      $$\begin{aligned}
      y &= \sum_{n=1}^{N} \gamma_{n\in\mathbb{Z}} x_n \\
      &= \left(\sum_{n=1}^{N-1} \gamma_{n\in\mathbb{Z}} x_n\right) + \gamma_{n\in\mathbb{Z}}\left(-\frac{\beta}{\alpha_n} y_1 - \sum_{n=1}^{N-1} \frac{\alpha_n}{\alpha_n} x_n\right) \\
      &= -\frac{\beta\gamma_n}{\alpha_n} y_1 + \sum_{n=1}^{N-1}\left(\gamma_n - \frac{\alpha_n\gamma_n}{\alpha_n}\right) x_n \\
      &= \delta y_1 + \sum_{n=1}^{N-1} \delta_{n\in\mathbb{Z}} x_n
      \end{aligned}$$

      iv. This implies that $\{y_1, x_1, \ldots, x_{N-1}\}$ *spans* L:

   (b) Proof that $\{y_1, x_1, \ldots, x_{N-1}\}$ is *linearly independent*:

      i. If $\{y_1, x_1, \ldots, x_{N-1}\}$ is *linearly dependent*, then there exists $\{\epsilon, \epsilon_1, \ldots, \epsilon_{N-1}\}$ such that
      $$\epsilon y_1 + \left(\sum_{n=1}^{N-1} \epsilon_{n\in\mathbb{Z}} x_n\right) + 0x_n = 0.$$

      ii. item (1(b)i) implies that the coordinate of y_1 associated with x_n *is* 0.
      $$y_1 = -\left(\sum_{n=1}^{N-1} \frac{\epsilon_n}{\epsilon} x_n\right) + 0x_n = 0.$$

      iii. item (1(a)i) implies that the coordinate of y_1 associated with x_n is *not* 0.
      $$y_1 = -\sum_{n=1}^{N} \frac{\alpha_n}{\beta} x_n.$$

      iv. This implies that item (1(b)i) (that the set is linearly dependent) is *false* because item (1(b)ii) and item (1(b)iii) *contradict* each other.

      v. This implies $\{y_1, x_1, \ldots, x_{N-1}\}$ is *linearly independent*.

2. Proof that $\{y_1, y_2, x_1, \ldots, x_{N-2}\}$ is a *basis*: Repeat item (1).

3. Suppose $m = n$. Proof that $\{y_1, y_2, \ldots, y_M\}$ is a *basis*: Repeat item (1) $M - 1$ times.

4. Proof that $M \not> N$:

   (a) Suppose that $M = N + 1$.

   (b) Then because $\{y_n|_{n=1,2,\ldots,N}\}$ is a *basis*, there exists $\{\zeta_n|_{n=1,2,\ldots,N+1}\}$ such that
   $$\sum_{n=1}^{N+1} \zeta_{n\in\mathbb{Z}} y_{n\in\mathbb{Z}} = 0.$$

   (c) This implies that $\{y_n|_{n=1,2,\ldots,N+1}\}$ is *linearly dependent*.

(d) This implies that $\{y_n|_{n=1,2,\ldots,N+1}\}$ is *not* a basis.

(e) This implies that $M \not> N$.

5. Proof that $M \neq N \implies \{y_n|_{n=1,2,\ldots,M}\}$ is *not* a basis for $\boldsymbol{L}$:

(a) Proof that $M > N \implies \{y_n|_{n=1,2,\ldots,M}\}$ is *not* a basis for $\boldsymbol{L}$: same as in item (4).

(b) Proof that $M < N \implies \{y_n|_{n=1,2,\ldots,M}\}$ is *not* a basis for $\boldsymbol{L}$:

i. Suppose $m = N - 1$.

ii. Then $\{y_n|_{n=1,2,\ldots,N-1}\}$ is a *basis* and there exists λ such that

$$\sum_{n=1}^{N} \lambda_{n\in\mathbb{Z}} y_{n\in\mathbb{Z}} = 0.$$

iii. This implies that $\{y_n|_{n=1,2,\ldots,N}\}$ is *linearly dependent* and is *not* a basis.

iv. But this contradicts item (3), therefore $M \neq N - 1$.

v. Because $M = N$ yields a basis but $M = N - 1$ does not, $M < N - 1$ also does not yield a basis.

⇒

Corollary K.1. [9] *Let* $\boldsymbol{L} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$ *be a linear space.*

COR

$$\underbrace{\left\{\begin{array}{ll} 1. & \{x_n \in X|_{n=1,2,\ldots,N}\} \text{ is a HAMEL BASIS for } \boldsymbol{L} \quad \text{and} \\ 2. & \{y_n \in X|_{n=1,2,\ldots,M}\} \text{ is a HAMEL BASIS for } \boldsymbol{L} \end{array}\right\} \implies \{N = M\}}_{\text{(all Hamel bases for } \boldsymbol{L} \text{ have the same number of vectors)}}$$

✎PROOF: This follows from Theorem K.2 (page 261). ⇒

K.2 Bases in topological linear spaces

A linear space supports the concept of the *span* of a set of vectors (Definition K.2 page 259). In a topological linear space $\boldsymbol{\Omega} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), T)$, a set A is said to be *total* in $\boldsymbol{\Omega}$ if the span of A is *dense* in $\boldsymbol{\Omega}$. In this case, A is said to be a *total set* or a *complete set*. However, this use of "complete" in a "*complete set*" is not equivalent to the use of "complete" in a "*complete metric space*".[10] In this text, except for these comments and Definition K.6, "complete" refers to the metric space definition only.

If a set is both *total* and *linearly independent* (Definition K.3 page 259) in $\boldsymbol{\Omega}$, then that set is a *Hamel basis* (Definition K.4 page 261) for $\boldsymbol{\Omega}$.

Definition K.6. [11] *Let* A^- *be the* CLOSURE *of a* A *in a* TOPOLOGICAL LINEAR SPACE $\boldsymbol{\Omega} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), T)$. *Let* $\operatorname{span} A$ *be the* SPAN *(Definition K.2 page 259) of a set* A.

DEF

A set of vectors A *is* ***total*** *(or* ***complete*** *or* ***fundamental****) in* $\boldsymbol{\Omega}$ *if*

$$(\operatorname{span} A)^- = \boldsymbol{\Omega} \qquad (\text{SPAN of } A \text{ is DENSE in } \boldsymbol{\Omega}).$$

[9] Kubrusly (2001) page 52 ⟨Theorem 2.7⟩, Michel and Herget (1993) page 91 ⟨Theorem 3.3.31⟩

[10] Haaser and Sullivan (1991) pages 296–297 ⟨6·Orthogonal Bases⟩, Rynne and Youngson (2008) page 78 ⟨Remark 3.50⟩, Heil (2011) page 21 ⟨Remark 1.26⟩

[11] Young (2001) page 19 ⟨Definition 1.5.1⟩, Sohrab (2003) page 362 ⟨Definition 9.2.3⟩, Gupta (1998) page 134 ⟨Definition 2.4⟩, Bachman and Narici (1966) pages 149–153 ⟨Definition 9.3, Theorems 9.9 and 9.10⟩

K.3 Schauder bases in Banach spaces

Definition K.7. [12] *Let* $\boldsymbol{B} \triangleq \left(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|\right)$ *be a* BANACH SPACE. *Let* $\stackrel{\star}{=}$ *represent* STRONG CONVERGENCE *in* $\boldsymbol{B}$.

DEF

The countable set $\left\{ \boldsymbol{x}_n \in X \,\middle|\, n\in\mathbb{N} \right\}$ *is a* ***Schauder basis*** *for* $\boldsymbol{B}$ *if for each* $\boldsymbol{x} \in X$

1. $\exists \left(\!\left(\alpha_n \in \mathbb{F}\right)\!\right)_{n\in\mathbb{N}}$ such that $\boldsymbol{x} \stackrel{\star}{=} \sum_{n=1}^{\infty} \alpha_n \boldsymbol{x}_n$ (STRONG CONVERGENCE *in* $\boldsymbol{B}$) *and*
2. $\left\{ \sum_{n=1}^{\infty} \alpha_n \boldsymbol{x}_n \stackrel{\star}{=} \sum_{n=1}^{\infty} \beta_n \boldsymbol{x}_n \right\} \implies \left\{ \left(\!\left(\alpha_n\right)\!\right) = \left(\!\left(\beta_n\right)\!\right) \right\}$ (COEFFICIENT FUNCTIONALS *are* UNIQUE)

In this case, $\sum_{n=1}^{\infty} \alpha_n \boldsymbol{x}_n$ *is the* ***expansion*** *of* $\boldsymbol{x}$ *on* $\left\{ \boldsymbol{x}_n \,\middle|\, n\in\mathbb{N} \right\}$ *and*
the elements of $\left(\!\left(\alpha_n\right)\!\right)$ *are the* ***coefficient functionals*** *associated with the basis* $\left\{ \boldsymbol{x}_n \right\}$. *Coefficient functionals are also called* ***coordinate functionals***.

In a Banach space, the existence of a Schauder basis implies that the space is *separable* (Theorem K.3 page 264). The question of whether the converse is also true was posed by Banach himself in 1932,[13] and became know as "*The basis problem*". This remained an open question for many years. The question was finally answered some 41 years later in 1973 by Per Enflo (University of California at Berkley), with the answer being "no". Enflo constructed a counterexample in which a separable Banach space does *not* have a Schauder basis.[14] Life is simpler in Hilbert spaces where the converse *is* true: a Hilbert space has a Schauder basis *if and only if* it is separable (Theorem K.12 page 277).

Theorem K.3. [15] *Let* $\boldsymbol{B} \triangleq \left(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|\right)$ *be a* BANACH SPACE. *Let* $\mathbb{Q}$ *be the field of rational numbers.*

THM

$$\left\{ \begin{array}{ll} 1. & \boldsymbol{B} \text{ has a SCHAUDER BASIS and} \\ 2. & \mathbb{Q} \text{ is DENSE in } \mathbb{F}. \end{array} \right\} \implies \left\{ \boldsymbol{B} \text{ is SEPARABLE} \right\}$$

✎PROOF:

1. lemma:

$$\begin{aligned}
\left| \left\{ \boldsymbol{x} \,|\, \exists \left(\!\left(\alpha_n \mathbb{Q}\right)\!\right)_{n\in\mathbb{N}} \text{ such that } \lim_{N\to\infty} \left\| \boldsymbol{x} - \sum_{n=1}^{N} \alpha_n \boldsymbol{x}_n \right\| = 0 \right\} \right| &= |\mathbb{Q}\times\mathbb{N}| \\
&= |\mathbb{Z}\times\mathbb{Z}| \\
&= |\mathbb{Z}| \\
&= \textit{countably infinite}
\end{aligned}$$

---

[12] Carothers (2005) pages 24–25, Christensen (2003) pages 46–49 ⟨Definition 3.1.1 and page 49⟩, Young (2001) page 19 ⟨Section 6⟩, Singer (1970) page 17, Schauder (1927), Schauder (1928)

[13] Banach (1932a) page 111

[14] Enflo (1973), Lindenstrauss and Tzafriri (1977) pages 84–95 ⟨Section 2.d⟩

[15] Bachman et al. (2002) page 112 ⟨3.4.8⟩, Giles (2000) page 17, Heil (2011) page 21 ⟨Theorem 1.27⟩

2. remainder of proof:

$$\begin{aligned}
&\boldsymbol{B} \text{ has a } \textit{Schauder basis } (\boldsymbol{x}_n)_{n\in\mathbb{N}}\\
&\implies \text{for every } \boldsymbol{x}\in\boldsymbol{B}, \text{ there exists } (\alpha_n\in\mathbb{F})_{n\in\mathbb{N}} \text{ such that } \boldsymbol{x} \stackrel{\star}{=} \sum_{n=1}^{\infty}\alpha_n\boldsymbol{x}_n && \text{by Definition K.7 page 264}\\
&\implies \text{for every } \boldsymbol{x}\in\boldsymbol{B}, \text{ there exists } (\alpha_n\in\mathbb{F})_{n\in\mathbb{N}} \text{ such that } \lim_{N\to\infty}\left\|\boldsymbol{x}-\sum_{n=1}^{N}\alpha_n\boldsymbol{x}_n\right\| = 0\\
&\implies \text{for every } \boldsymbol{x}\in\boldsymbol{B}, \text{ there exists } (\alpha_n\in\mathbb{Q})_{n\in\mathbb{N}} \text{ such that } \lim_{N\to\infty}\left\|\boldsymbol{x}-\sum_{n=1}^{N}\alpha_n\boldsymbol{x}_n\right\| = 0 && \text{because } \mathbb{Q}^- = \mathbb{F}\\
&\implies \boldsymbol{B} = \left\{\boldsymbol{x} \,|\, \exists\, (\alpha_n\mathbb{Q})_{n\in\mathbb{N}} \text{ such that } \lim_{N\to\infty}\left\|\boldsymbol{x}-\sum_{n=1}^{N}\alpha_n\boldsymbol{x}_n\right\| = 0\right\}\\
&\implies \boldsymbol{B} = \left\{\boldsymbol{x} \,|\, \exists\, (\alpha_n\mathbb{Q})_{n\in\mathbb{N}} \text{ such that } \boldsymbol{x} = \lim_{N\to\infty}\sum_{n=1}^{N}\alpha_n\boldsymbol{x}_n\right\}^-\\
&\implies \boldsymbol{B} \text{ is } \textit{separable} && \text{by (1) lemma page 264}
\end{aligned}$$

✏

Definition K.8. [16] *Let* $\{\boldsymbol{x}_n|\, n\in\mathbb{N}\}$ *and* $\{\boldsymbol{y}_n|\, n\in\mathbb{N}\}$ *be* SCHAUDER BASES *of a* BANACH SPACE $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$.

DEF: $\{\boldsymbol{x}_n\}$ ***is equivalent*** *to* $\{\boldsymbol{y}_n\}$
if there exists a BOUNDED INVERTIBLE *operator* $\mathbf{R}$ *in* $\boldsymbol{X}^{\boldsymbol{X}}$ *such that* $\quad \mathbf{R}\boldsymbol{x}_n = \boldsymbol{y}_n \quad \forall n\in\mathbb{Z}$

Theorem K.4. [17] *Let* $\{\boldsymbol{x}_n|\, n\in\mathbb{N}\}$ *and* $\{\boldsymbol{y}_n|\, n\in\mathbb{N}\}$ *be* SCHAUDER BASES *of a* BANACH SPACE $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$.

THM:
$$\left\{\{\boldsymbol{x}_n\} \text{ is EQUIVALENT to } \{\boldsymbol{y}_n\}\right\} \iff \left\{\sum_{n=1}^{\infty}\alpha_n\boldsymbol{x}_n \text{ is CONVERGENT} \iff \sum_{n=1}^{\infty}\alpha_n\boldsymbol{y}_n \text{ is CONVERGENT}\right\}$$

Lemma K.1. [18] *Let* $\boldsymbol{\Omega} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \boldsymbol{T})$ *be a topological linear space. Let* span A *be the* SPAN *of a set* A (Definition K.2 page 259). *Let* $\tilde{\mathsf{f}}(\omega)$ *and* $\tilde{\mathsf{g}}(\omega)$ *be the* FOURIER TRANSFORMS (Definition T.2 page 408) *of the functions* $\mathsf{f}(x)$ *and* $\mathsf{g}(x)$, *respectively, in* $\boldsymbol{L}^2_{\mathbb{R}}$ (Definition H.2 page 223). *Let* $\breve{a}(\omega)$ *be the* DTFT (Definition U.1 page 419) *of a sequence* $(a_n)_{n\in\mathbb{Z}}$ *in* $\boldsymbol{\ell}^2_{\mathbb{R}}$ (Definition D.3 page 201).

LEM:
$$\left\{\begin{array}{ll}(1). & \{\mathbf{T}^n\mathsf{f}|\, n\in\mathbb{Z}\} \text{ is a SCHAUDER BASIS for } \boldsymbol{\Omega} \quad \text{and}\\ (2). & \{\mathbf{T}^n\mathsf{g}|\, n\in\mathbb{Z}\} \text{ is a SCHAUDER BASIS for } \boldsymbol{\Omega}\end{array}\right\} \implies \left\{\begin{array}{l}\exists\, (a_n)_{n\in\mathbb{Z}} \quad \text{such that}\\ \tilde{\mathsf{f}}(\omega) = \breve{a}(\omega)\tilde{\mathsf{g}}(\omega)\end{array}\right\}$$

✎PROOF: Let V_0' be the space spanned by $\{\mathbf{T}^n\phi|\, n\in\mathbb{Z}\}$.

$$\begin{aligned}
\tilde{\mathsf{f}}(\omega) &\triangleq \tilde{\mathbf{F}}\mathsf{f} && \text{by definition of } \tilde{\mathbf{F}} && \text{(Definition T.2 page 408)}\\
&= \tilde{\mathbf{F}}\sum_{n\in\mathbb{Z}} a_n\mathbf{T}\mathsf{g} && \text{by (2)}\\
&= \sum_{n\in\mathbb{Z}} a_n\tilde{\mathbf{F}}\mathbf{T}\mathsf{g}
\end{aligned}$$

[16] Young (2001) page 25 ⟨Definition 1.8.1, Theorem 1.8.7⟩
[17] Young (2001) page 25 ⟨Definition 1.8.1, Theorem 1.8.7⟩
[18] Daubechies (1992) page 140

$$= \underbrace{\sum_{n\in\mathbb{Z}} a_n e^{-i\omega n}}_{\breve{a}(\omega)} \underbrace{\tilde{\mathbf{F}}\mathsf{g}}_{\tilde{\mathsf{g}}(\omega)}$$

$$= \breve{a}(\omega)\tilde{\mathsf{g}}(\omega)$$ by definition of $\breve{\mathbf{F}}$ and $\tilde{\mathbf{F}}$ (Definition U.1 page 419) (Definition T.2 page 408)

$$\begin{aligned}
V_0 &\triangleq \left\{ \mathsf{f}(x) | \mathsf{f}(x) = \sum_{n\in\mathbb{Z}} b_n \mathbf{T}^n \mathsf{g}(x) \right\} \\
&= \left\{ \mathsf{f}(x) | \tilde{\mathbf{F}}\mathsf{f}(x) = \tilde{\mathbf{F}} \sum_{n\in\mathbb{Z}} b_n \mathbf{T}^n \mathsf{g}(x) \right\} \\
&= \left\{ \mathsf{f}(x) | \tilde{\mathsf{f}}(\omega) = \tilde{\mathsf{b}}(\omega)\tilde{\mathsf{g}}(\omega) \right\} \\
&= \left\{ \mathsf{f}(x) | \tilde{\mathsf{f}}(\omega) = \tilde{\mathsf{b}}(\omega)\breve{\mathsf{a}}(\omega)\tilde{\mathsf{f}}(\omega) \right\} \\
&= \left\{ \mathsf{f}(x) | \tilde{\mathsf{f}}(\omega) = \tilde{\mathsf{c}}(\omega)\tilde{\mathsf{f}}(\omega) \right\} && \text{where } \tilde{\mathsf{c}}(\omega) \triangleq \tilde{\mathsf{b}}(\omega)\breve{\mathsf{a}}(\omega) \\
&= \left\{ \mathsf{f}(x) | \mathsf{f}(x) = \sum_{n\in\mathbb{Z}} \mathsf{c}_n \mathsf{f}(x-n) \right\} \\
&\triangleq V_0'
\end{aligned}$$

⇨

K.4 Linear combinations in inner product spaces

In an *inner product space*, *orthogonality* is a special case of *linear independence*; or alternatively, linear independence is a generalization of orthogonality (next theorem).

Theorem K.5. [19] *Let* $\{ \boldsymbol{x}_n \in X | _{n=1,2,\ldots,N} \}$ *be a set of vectors in an* INNER PRODUCT SPACE *(Definition N.1 page 309)* $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle | \nabla \rangle)$.

THM
$$\left\{ \begin{array}{c} \{\boldsymbol{x}_n\} \text{ is ORTHOGONAL} \\ \text{(Definition N.4 page 323)} \end{array} \right\} \implies \left\{ \begin{array}{c} \{\boldsymbol{x}_n\} \text{ is LINEARLY INDEPENDENT} \\ \text{(Definition K.1 page 259)} \end{array} \right\}$$

✎PROOF:

1. Proof using *Pythagorean theorem* (Theorem N.10 page 324):
Let $(\alpha_n)_{n\in\mathbb{N}}$ be a sequence with at least one nonzero element.

$$\begin{aligned}
\left\| \sum_{n=1}^{N} \alpha_n \boldsymbol{x}_n \right\|^2 &= \sum_{n=1}^{N} \|\alpha_n \boldsymbol{x}_n\|^2 && \text{by left hypoth. and } \textit{Pythagorean Theorem} \text{ (Theorem N.10 page 324)} \\
&= \sum_{n=1}^{N} |\alpha_n|^2 \|\boldsymbol{x}_n\|^2 && \text{by definition of } \|\cdot\| \text{ (Definition O.1 page 327)} \\
&> 0
\end{aligned}$$

$$\implies \sum_{n=1}^{N} \alpha_n \boldsymbol{x}_n \neq 0$$

$$\implies (\boldsymbol{x}_n)_{n\in\mathbb{N}} \text{ is } \textit{linearly independent}$$ by definition of *linear independence* (Definition K.3 page 259)

[19] Aliprantis and Burkinshaw (1998) page 283 ⟨Corollary 32.8⟩, Kubrusly (2001) page 352 ⟨Proposition 5.34⟩

2. Alternative proof:

$$\begin{aligned}
\sum_{n=1}^{N} \alpha_n x_n = \mathbb{0} &\implies \left\langle \sum_{n=1}^{N} \alpha_n x_n \,\middle|\, x_m \right\rangle = \langle \mathbb{0} \,|\, x_m \rangle \\
&\implies \sum_{n=1}^{N} \alpha_n \langle x_n \,|\, x_m \rangle = 0 \\
&\implies \sum_{n=1}^{N} \alpha_n \bar{\delta}(k-m) = 0 \\
&\implies \alpha_m = 0 \qquad \text{for } m = 1, 2, \ldots, N
\end{aligned}$$

✏

Theorem K.6 (Bessel's Equality). [20] *Let* $\{ x_n \in X | _{n=1,2,\ldots,N} \}$ *be a set of vectors in an* INNER PRODUCT SPACE *(Definition N.1 page 309)* $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \,|\, \nabla \rangle)$ *and with* $\|x\| \triangleq \sqrt{\langle x \,|\, x \rangle}$ *(Definition N.2 page 315).*

THM

$$\left\{ \begin{array}{l} \{x_n\} \text{ is ORTHONORMAL} \\ \text{(Definition N.4 page 323)} \end{array} \right\} \implies \left\{ \underbrace{\left\| x - \sum_{n=1}^{N} \langle x \,|\, x_n \rangle x_n \right\|^2}_{\text{approximation error}} = \|x\|^2 - \sum_{n=1}^{N} |\langle x \,|\, x_n \rangle|^2 \quad \forall x \in X \right\}$$

✎PROOF:

$$\begin{aligned}
&\left\| x - \sum_{n=1}^{N} \langle x \,|\, x_n \rangle x_n \right\|^2 \\
&= \|x\|^2 + \left\| \sum_{n=1}^{N} \langle x \,|\, x_n \rangle x_n \right\|^2 - 2\mathbf{R_e} \left\langle x \,\middle|\, \sum_{n=1}^{N} \langle x \,|\, x_n \rangle x_n \right\rangle && \text{by \textit{polar identity}} && \text{(Lemma N.1 page 314)} \\
&= \|x\|^2 + \left\| \sum_{n=1}^{N} \langle x \,|\, x_n \rangle x_n \right\|^2 - 2\mathbf{R_e} \left[\left(\sum_{n=1}^{N} \langle x \,|\, x_n \rangle \right)^* \langle x \,|\, x_n \rangle \right] && \text{by property of } \langle \triangle \,|\, \nabla \rangle && \text{(Definition N.1 page 309)} \\
&= \|x\|^2 + \sum_{n=1}^{N} \| \langle x \,|\, x_n \rangle x_n \|^2 - 2\mathbf{R_e} \left[\left(\sum_{n=1}^{N} \langle x \,|\, x_n \rangle \right)^* \langle x \,|\, x_n \rangle \right] && \text{by \textit{Pythagorean Theorem}} && \text{(Theorem N.10 page 324)} \\
&= \|x\|^2 + \sum_{n=1}^{N} \| \langle x \,|\, x_n \rangle x_n \|^2 - 2\mathbf{R_e} \left(\sum_{n=1}^{N} \langle x \,|\, x_n \rangle^* \langle x \,|\, x_n \rangle \right) \\
&= \|x\|^2 + \sum_{n=1}^{N} |\langle x \,|\, x_n \rangle|^2 \underbrace{\|x_n\|^2}_{1} - 2\mathbf{R_e} \left(\sum_{n=1}^{N} \langle x \,|\, x_n \rangle^* \langle x \,|\, x_n \rangle \right) && \text{by property of } \|\cdot\| && \text{(Definition O.1 page 327)} \\
&= \|x\|^2 + \sum_{n=1}^{N} |\langle x \,|\, x_n \rangle|^2 \cdot 1 - 2\mathbf{R_e} \left(\sum_{n=1}^{N} \langle x \,|\, x_n \rangle^* \langle x \,|\, x_n \rangle \right) && \text{by def. of \textit{orthonormality}} && \text{(Definition N.4 page 323)} \\
&= \|x\|^2 + \sum_{n=1}^{N} |\langle x \,|\, x_n \rangle|^2 - 2\mathbf{R_e} \sum_{n=1}^{N} |\langle x \,|\, x_n \rangle|^2 \\
&= \|x\|^2 + \sum_{n=1}^{N} |\langle x \,|\, x_n \rangle|^2 - 2 \sum_{n=1}^{N} |\langle x \,|\, x_n \rangle|^2 && \text{because } |\cdot| \text{ is real}
\end{aligned}$$

[20] Bachman et al. (2002) page 103, Pedersen (2000) pages 38–39

$$= \|\boldsymbol{x}\|^2 - \sum_{n=1}^{N} |\langle \boldsymbol{x} \,|\, \boldsymbol{x}_n \rangle|^2$$

✏

Theorem K.7 (Bessel's inequality). [21] *Let* $\{ \boldsymbol{x}_n \in X |_{n=1,2,\ldots,N} \}$ *be a set of vectors in an* INNER PRODUCT SPACE *(Definition N.1 page 309)* $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \,|\, \nabla \rangle)$ *and with* $\|\boldsymbol{x}\| \triangleq \sqrt{\langle \boldsymbol{x} \,|\, \boldsymbol{x} \rangle}$ *(Definition N.2 page 315)*.

THM

$$\left\{ \{\boldsymbol{x}_n\} \text{ is ORTHONORMAL (Definition N.4 page 323)} \right\} \implies \left\{ \sum_{n=1}^{N} |\langle \boldsymbol{x} \,|\, \boldsymbol{x}_n \rangle|^2 \le \|\boldsymbol{x}\|^2 \quad \forall x \in X \right\}$$

✎PROOF:

$$\begin{aligned}
0 &\le \left\| \boldsymbol{x} - \sum_{n=1}^{N} \langle \boldsymbol{x} \,|\, \boldsymbol{x}_n \rangle \boldsymbol{x}_n \right\|^2 && \text{by definition of } \|\cdot\| && \text{(Definition O.1 page 327)} \\
&= \|\boldsymbol{x}\|^2 - \sum_{n=1}^{N} |\langle \boldsymbol{x} \,|\, \boldsymbol{x}_n \rangle|^2 && \text{by } \textit{Bessel's Equality} && \text{(Theorem K.6 page 267)}
\end{aligned}$$

✏

Theorem K.8. *Let* $\{ \boldsymbol{x}_n \in X |_{n=1,2,\ldots,N} \}$ *be a set of vectors in an* INNER PRODUCT SPACE *(Definition N.1 page 309)* $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \,|\, \nabla \rangle)$ *and with* $\|\boldsymbol{x}\| \triangleq \sqrt{\langle \boldsymbol{x} \,|\, \boldsymbol{x} \rangle}$ *(Definition N.2 page 315)*.

THM

$$\{(\alpha_k) \text{ is REAL-VALUED}\} \implies \alpha_k \|\boldsymbol{x}_k\|^2 = \mathbf{R_e} \langle \boldsymbol{x} \,|\, \boldsymbol{x}_k \rangle - \sum_{m \ne k} \alpha_m \langle \boldsymbol{x}_k \,|\, \boldsymbol{x}_m \rangle$$

✎PROOF:

1. lemma: $\dfrac{\partial}{\partial \alpha_k} \|\boldsymbol{x}\|^2 = 0$ because $\|\boldsymbol{x}\|^2$ does not vary with varying α_k.

2. lemma:

$$\begin{aligned}
\frac{\partial}{\partial \alpha_k} \left\langle \boldsymbol{x} \,\middle|\, \sum_{n=1}^{N} \alpha_n \boldsymbol{x}_n \right\rangle &= \frac{\partial}{\partial \alpha_k} \sum_{n=1}^{N} \alpha_n^* \langle \boldsymbol{x} \,|\, \boldsymbol{x}_n \rangle \\
&= \left(\frac{\partial}{\partial \alpha_k} \sum_{n=1}^{N} \alpha_n \langle \boldsymbol{x} \,|\, \boldsymbol{x}_n \rangle^* \right)^* \\
&= \left(\langle \boldsymbol{x} \,|\, \boldsymbol{x}_k \rangle^* \right)^* \\
&= \langle \boldsymbol{x} \,|\, \boldsymbol{x}_k \rangle
\end{aligned}$$

[21] Giles (2000) pages 54–55 ⟨3.13 Bessel's inequality⟩, Bollobás (1999) page 147, Aliprantis and Burkinshaw (1998) page 284

3. lemma:

$$\begin{aligned}
\frac{\partial}{\partial \alpha_k}\left\|\sum_{n=1}^{N}\alpha_n \boldsymbol{x}_n\right\|^2 &= \frac{\partial}{\partial \alpha_k}\left\langle \sum_{n=1}^{N}\alpha_n \boldsymbol{x}_n \,\middle|\, \sum_{m=1}^{N}\alpha_m \boldsymbol{x}_m \right\rangle \\
&= \frac{\partial}{\partial \alpha_k}\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_n\alpha_m^* \langle \boldsymbol{x}_n \,|\, \boldsymbol{x}_m\rangle \\
&= \frac{\partial}{\partial \alpha_k}\sum_{n=1}^{N}\alpha_n\alpha^* \left\|\boldsymbol{x}_n\right\|^2 + \frac{\partial}{\partial \alpha_k}\sum_{n=1}^{N}\sum_{m\neq n}\alpha_n\alpha_m^* \langle \boldsymbol{x}_n \,|\, \boldsymbol{x}_m\rangle \\
&= 2\alpha_k \left\|\boldsymbol{x}_k\right\|^2 + \frac{\partial}{\partial \alpha_k}\sum_{m\neq n} 2\alpha_k\alpha_m^* \langle \boldsymbol{x}_n \,|\, \boldsymbol{x}_m\rangle \\
&= 2\alpha_k \left\|\boldsymbol{x}_k\right\|^2 + 2\sum_{m\neq n}\alpha_m^* \langle \boldsymbol{x}_k \,|\, \boldsymbol{x}_m\rangle
\end{aligned}$$

4. Proof that $\alpha_k \left\|\boldsymbol{x}_k\right\|^2 = \mathbf{R}_e \langle \boldsymbol{x} \,|\, \boldsymbol{x}_k\rangle - \sum_{m\neq k}\alpha_m \langle \boldsymbol{x}_k \,|\, \boldsymbol{x}_m\rangle$

$$\begin{aligned}
0 &= \frac{1}{2}\cdot 0 \\
&= \frac{1}{2}\frac{\partial}{\partial \alpha_k}\left\|\boldsymbol{x} - \sum_{n=1}^{N}\alpha_n \boldsymbol{x}_n\right\|^2 \\
&= \cancelto{0}{\frac{1}{2}\frac{\partial}{\partial \alpha_k}\|\boldsymbol{x}\|^2} - \frac{2}{2}\mathbf{R}_e\left[\frac{\partial}{\partial \alpha_n}\left\langle \boldsymbol{x} \,\middle|\, \sum_{n=1}^{N}\alpha_n \boldsymbol{x}_n\right\rangle\right] + \frac{\partial}{\partial \alpha_k}\frac{2}{2}\left\|\sum_{n=1}^{N}\alpha_n \boldsymbol{x}_n\right\|^2 && \text{by } \textit{Polar Identity} \quad \text{(Lemma N.1 page 314)} \\
&= 0 - \mathbf{R}_e\left[\langle \boldsymbol{x} \,|\, \boldsymbol{x}_k\rangle\right] + \alpha_k \langle \boldsymbol{x}_k \,|\, \boldsymbol{x}_k\rangle + \sum_{m\neq k}\alpha_m^* \langle \boldsymbol{x}_k \,|\, \boldsymbol{x}_m\rangle \\
&= -\mathbf{R}_e \langle \boldsymbol{x} \,|\, \boldsymbol{x}_k\rangle + \alpha_k \left\|\boldsymbol{x}_k\right\|^2 + \sum_{m\neq k}\alpha_m \langle \boldsymbol{x}_k \,|\, \boldsymbol{x}_m\rangle
\end{aligned}$$

5. Note in the special case of $\langle\!\langle \boldsymbol{x}_n \rangle\!\rangle$ being *orthonormal*...

$$\begin{aligned}
\alpha_k &= \frac{1}{\cancelto{1}{\left\|\boldsymbol{x}_k\right\|^2}}\left[\mathbf{R}_e\left[\langle \boldsymbol{x} \,|\, \boldsymbol{x}_k\rangle\right] - \sum_{m\neq k}\alpha_m \langle \boldsymbol{x}_k \,|\, \boldsymbol{x}_m\rangle\right] \\
&= \frac{1}{1}\left[\mathbf{R}_e\left[\langle \boldsymbol{x} \,|\, \boldsymbol{x}_k\rangle\right] - \sum_{m\neq k}\alpha_m \cancelto{0}{\langle \boldsymbol{x}_k \,|\, \boldsymbol{x}_m\rangle}\right] \\
&= \mathbf{R}_e \langle \boldsymbol{x} \,|\, \boldsymbol{x}_k\rangle
\end{aligned}$$

✏

Theorem K.8 (previous) demonstrates that in the case where $\langle\!\langle \boldsymbol{x}_n \rangle\!\rangle$ is *orthonormal* and $\langle\!\langle \alpha_k \rangle\!\rangle$ is **real-valued**, then the best α_k in the least square sense is simply equal to the projection $\langle \boldsymbol{x} \,|\, \boldsymbol{x}_k\rangle$. The *Best Approximation Theorem* (next) shows that in the case of the *orthonormal* $\langle\!\langle \boldsymbol{x}_n \rangle\!\rangle$,

- the constraint on $\langle\!\langle \alpha_k \rangle\!\rangle$ can actually be dropped ($\langle\!\langle \alpha_k \rangle\!\rangle$ can be complex)
- the error of the projection is orthogonal to the projection.

Theorem K.9 (Best Approximation Theorem). [22] *Let* $\left\{ \boldsymbol{x}_n \in X \,\middle|\, {}_{n=1,2,\ldots,N} \right\}$ *be a set of vectors in an* INNER PRODUCT SPACE *(Definition N.1 page 309)* $\left(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \,|\, \triangledown \rangle\right)$ *and with* $\|\boldsymbol{x}\| \triangleq \sqrt{\langle \boldsymbol{x} \,|\, \boldsymbol{x}\rangle}$ *(Definition*

[22] Walter and Shen (2001) pages 3–4, Pedersen (2000) page 39, Edwards (1995) pages 94–100, Weyl (1940)

N.2 page 315).

THM

$$\left\{\begin{array}{c}\{\boldsymbol{x}_n\} \textit{ is} \\ \text{ORTHONORMAL} \\ \text{(Definition N.4 page 323)}\end{array}\right\} \implies \left\{\begin{array}{lll}1. & \underset{(\alpha_n)_{n=1}^N}{\arg\min}\left\|\boldsymbol{x}-\sum_{n=1}^N \alpha_n\boldsymbol{x}_n\right\| = \underbrace{\left(\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle\right)_{n=1}^N}_{\textit{best } \alpha_n = \langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle} & \forall\boldsymbol{x}\in X \quad \textit{and} \\ 2. & \underbrace{\left(\sum_{n=1}^N\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle\boldsymbol{x}_n\right)}_{\textit{approximation}} \perp \underbrace{\left(\boldsymbol{x}-\sum_{n=1}^N\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle\boldsymbol{x}_n\right)}_{\textit{approximation error}} & \forall\boldsymbol{x}\in X\end{array}\right\}$$

✎PROOF:

1. Proof that $(\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle)$ is the best sequence:

$$\begin{aligned}
&\left\|\boldsymbol{x}-\sum_{n=1}^N\alpha_n\boldsymbol{x}_n\right\|^2 \\
&= \|\boldsymbol{x}\|^2 - 2\mathbf{R_e}\left\langle\boldsymbol{x}\,\middle|\,\sum_{n=1}^N\alpha_n\boldsymbol{x}_n\right\rangle + \left\|\sum_{n=1}^N\alpha_n\boldsymbol{x}_n\right\|^2 && \text{by \textit{Polar Identity}} && \text{(Lemma N.1 page 314)} \\
&= \|\boldsymbol{x}\|^2 - 2\mathbf{R_e}\left(\sum_{n=1}^N\alpha_n^*\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle\right) + \sum_{n=1}^N\|\alpha_n\boldsymbol{x}_n\|^2 && \text{by \textit{Pythagorean Theorem}} && \text{(Theorem N.10 page 324)} \\
&= \|\boldsymbol{x}\|^2 - 2\mathbf{R_e}\left(\sum_{n=1}^N\alpha_n^*\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle\right) + \sum_{n=1}^N|\alpha_n|^2\cancelto{1}{\|\boldsymbol{x}_n\|^2} + \underbrace{\left[\sum_{n=1}^N|\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle|^2 - \sum_{n=1}^N|\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle|^2\right]}_{0} \\
&= \left[\|\boldsymbol{x}\|^2 - \sum_{n=1}^N|\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle|^2\right] + \sum_{n=1}^N\left[|\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle|^2 - 2\mathbf{R_e}\left[\alpha_n^*\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle\right] + |\alpha_n|^2\right] \\
&= \left[\|\boldsymbol{x}\|^2 - \sum_{n=1}^N|\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle|^2\right] + \sum_{n=1}^N\left[|\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle|^2 - \alpha_n^*\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle - \alpha_n\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle^* + |\alpha_n|^2\right] \\
&= \left\|\boldsymbol{x}-\sum_{n=1}^N\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle\boldsymbol{x}_n\right\|^2 + \sum_{n=1}^N|\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle - \alpha_n|^2 && \text{by \textit{Bessel's Equality}} && \text{(Theorem K.6 page 267)} \\
&\geq \left\|\boldsymbol{x}-\sum_{n=1}^N\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle\boldsymbol{x}_n\right\|^2
\end{aligned}$$

2. Proof that the approximation and approximation error are orthogonal:

$$\begin{aligned}
\left\langle\sum_{n=1}^N\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle\boldsymbol{x}_n\,\middle|\,\boldsymbol{x}-\sum_{n=1}^N\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle\boldsymbol{x}_n\right\rangle &= \left\langle\sum_{n=1}^N\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle\boldsymbol{x}_n\,\middle|\,\boldsymbol{x}\right\rangle - \left\langle\sum_{n=1}^N\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle\boldsymbol{x}_n\,\middle|\,\sum_{n=1}^N\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle\boldsymbol{x}_n\right\rangle \\
&= \sum_{n=1}^N\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle^*\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle - \sum_{n=1}^N\sum_{m=1}^N\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle\langle\boldsymbol{x}\,|\,\boldsymbol{x}_m\rangle^*\langle\boldsymbol{x}_n\,|\,\boldsymbol{x}_m\rangle \\
&= \sum_{n=1}^N|\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle|^2 - \sum_{n=1}^N\sum_{m=1}^N\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle\langle\boldsymbol{x}\,|\,\boldsymbol{x}_m\rangle^*\bar{\delta}_{nm} \\
&= \sum_{n=1}^N|\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle|^2 - \sum_{n=1}^N|\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle|^2
\end{aligned}$$

$$= 0$$

✏

K.5 Orthonormal bases in Hilbert spaces

Definition K.9. *Let* $\{x_n \in X \mid n=1,2,\dots,N\}$ *be a set of vectors in an* INNER PRODUCT SPACE *(Definition N.1 page 309)* $\boldsymbol{\Omega} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \mid \triangledown \rangle)$.

DEF

The set $\{x_n\}$ *is an* ***orthogonal basis*** *for* $\boldsymbol{\Omega}$ *if* $\{x_n\}$ *is* ORTHOGONAL *and is a* SCHAUDER BASIS *for* $\boldsymbol{\Omega}$.

The set $\{x_n\}$ *is an* ***orthonormal basis*** *for* $\boldsymbol{\Omega}$ *if* $\{x_n\}$ *is* ORTHONORMAL *and is a* SCHAUDER BASIS *for* $\boldsymbol{\Omega}$.

Definition K.10. [23] *Let* $\boldsymbol{H} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \mid \triangledown \rangle)$ *be a Hilbert space.*

DEF

Suppose there exists a set $\{x_n \in X \mid n\in\mathbb{N}\}$ *such that* $x \stackrel{\star}{=} \sum_{n=1}^{\infty} \langle x \mid x_n\rangle x_n$.

Then the quantities $\langle x \mid x_n\rangle$ *are called the* ***Fourier coefficients*** *of* x *and the sum* $\sum_{n=1}^{\infty} \langle x \mid x_n\rangle x_n$ *is called the* ***Fourier expansion*** *of* x *or the* ***Fourier series*** *for* x.

Lemma K.2 (Perfect reconstruction). *Let* $\{x_n \in X \mid n\in\mathbb{N}\}$ *be a set of vectors in a* HILBERT SPACE $\boldsymbol{H} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \mid \triangledown \rangle)$.

LEM

$$\left\{\begin{array}{lll} (1). & (\!(x_n)\!) \text{ is a BASIS for } \boldsymbol{H} & \text{and} \\ (2). & (\!(x_n)\!) \text{ is ORTHONORMAL} & \end{array}\right\} \implies \overbrace{x \stackrel{\star}{=} \sum_{n=1}^{\infty} \underbrace{\langle x \mid x_n\rangle}_{\text{Fourier coefficient}} x_n}^{\text{Fourier expansion}} \quad \forall x\in X$$

✎PROOF:

$$\begin{aligned}
\langle x \mid x_n\rangle &= \left\langle \sum_{m\in\mathbb{Z}} \alpha_m x_m \mid x_n \right\rangle && \text{by left hypothesis (1)} \\
&= \sum_{m\in\mathbb{Z}} \alpha_m \langle x_m \mid x_n\rangle && \text{by } \textit{homogeneous} \text{ property of } \langle \triangle \mid \triangledown \rangle && \text{(Definition N.1 page 309)} \\
&= \sum_{m\in\mathbb{Z}} \alpha_m \bar{\delta}_{n-m} && \text{by left hypothesis (2)} && \text{(Definition N.4 page 323)} \\
&= \alpha_n
\end{aligned}$$

✏

Proposition K.2. [24] *Let* $\{x_n \in X \mid n\in\mathbb{N}\}$ *be a set of vectors in a* HILBERT SPACE $\boldsymbol{H} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \mid \triangledown \rangle)$.

PRP

$$\underbrace{\|x\|^2 \stackrel{\star}{=} \sum_{n=1}^{\infty} |\langle x \mid x_n\rangle|^2}_{\text{PARSEVAL FRAME}} \iff \underbrace{x \stackrel{\star}{=} \sum_{n=1}^{\infty} \langle x \mid x_n\rangle x_n}_{\text{FOURIER EXPANSION (Definition K.10 page 271)}} \quad \forall x\in X$$

[23] Fabian et al. (2010) page 27 ⟨Theorem 1.55⟩, Young (2001) page 6, Young (1980) page 6

[24] Han et al. (2007) pages 93–94 ⟨Proposition 3.11⟩

✎PROOF:

1. Proof that *Parseval frame* $\Longleftarrow$ *Fourier expansion*

$$\begin{aligned}
\|x\|^2 &\triangleq \langle x \mid x\rangle && \text{by definition of } \|\cdot\| && \text{(Definition O.1 page 327)}\\
&= \left\langle \sum_{n=1}^{\infty} \langle x \mid x_n\rangle\, x \mid x_n \right\rangle && \text{by right hypothesis}\\
&\stackrel{\star}{=} \sum_{n=1}^{\infty} \langle x \mid x_n\rangle \langle x \mid x_n\rangle && \text{by property of } \langle \triangle \mid \triangledown\rangle && \text{(Definition N.1 page 309)}\\
&\stackrel{\star}{=} \sum_{n=1}^{\infty} \langle x \mid x_n\rangle \langle x \mid x_n\rangle^* && \text{by property of } \langle \triangle \mid \triangledown\rangle && \text{(Definition N.1 page 309)}\\
&\stackrel{\star}{=} \sum_{n=1}^{\infty} \left|\langle x \mid x_n\rangle\right|^2 && \text{by property of } \mathbb{C} && \text{(Definition M.7 page 307)}
\end{aligned}$$

2. Proof that *Parseval frame* $\Longrightarrow$ *Fourier expansion*

   (a) Let $(e_n)_{n\in\mathbb{N}}$ be the *standard othornormal basis* such that the nth element of e_n is 1 and all other elements are 0.

   (b) Let $\mathbf{M}$ be an operator in $\boldsymbol{H}$ such that $\mathbf{M}x \triangleq \sum_{n=1}^{\infty} \langle x \mid x_n\rangle\, e_n$.

   (c) lemma: $\mathbf{M}$ is *isometric*. Proof:

$$\begin{aligned}
\|\mathbf{M}x\|^2 &= \left\| \sum_{n=1}^{\infty} \langle x \mid x_n\rangle\, e_n \right\|^2 && \text{by definition of } \mathbf{M} && \text{(item (2b) page 272)}\\
&= \sum_{n=1}^{\infty} \left\|\langle x \mid x_n\rangle\, e_n\right\|^2 && \text{by \textit{Pythagorean Theorem}} && \text{(Theorem N.10 page 324)}\\
&= \sum_{n=1}^{\infty} \left|\langle x \mid x_n\rangle\right|^2 \|e_n\|^2 && \text{by \textit{homogeneous} property of } \|\cdot\| && \text{(Definition O.1 page 327)}\\
&= \sum_{n=1}^{\infty} \left|\langle x \mid x_n\rangle\right|^2 && \text{by definition of \textit{orthonormal}} && \text{(Definition N.4 page 323)}\\
&= \|x\|^2 && \text{by Parseval frame hypothesis}\\
\Longrightarrow &\quad \mathbf{M} \text{ is \textit{isometric}} && \text{by definition of \textit{isometric}} && \text{(Definition R.10 page 379)}
\end{aligned}$$

   (d) Let $(u_n)_{n\in\mathbb{N}}$ be an *orthornormal basis* for $\boldsymbol{H}$.

   (e) Proof for *Fourier expansion*:

$$\begin{aligned}
x &= \sum_{n=1}^{\infty} \langle x \mid u_n\rangle\, u_n && \text{by \textit{Fourier expansion} (Proposition K.3 page 275)}\\
&= \sum_{n=1}^{\infty} \langle \mathbf{M}x \mid \mathbf{M}u_n\rangle\, u_n && \text{by (2c) lemma page 272 and Theorem R.23 page 380}\\
&= \sum_{n=1}^{\infty} \left\langle \sum_{m=1}^{\infty} \langle x \mid x_m\rangle\, e_m \mid \sum_{k=1}^{\infty} \langle u_n \mid x_k\rangle\, e_k \right\rangle u_n && \text{by item (2b) page 272}\\
&= \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \langle x \mid x_m\rangle \sum_{k=1}^{\infty} \langle u_n \mid x_k\rangle^* \langle e_m \mid e_k\rangle\, u_n && \text{by Definition N.1 page 309}
\end{aligned}$$

$$= \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \langle \boldsymbol{x} \,|\, \boldsymbol{x}_m \rangle \langle \boldsymbol{u}_n \,|\, \boldsymbol{x}_m \rangle^* \boldsymbol{u}_n \qquad \text{by item (2a) page 272 and Definition N.4 page 323}$$

$$= \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \langle \boldsymbol{x} \,|\, \boldsymbol{x}_m \rangle \langle \boldsymbol{x}_m \,|\, \boldsymbol{u}_n \rangle \boldsymbol{u}_n \qquad \text{by Definition N.1 page 309}$$

$$= \sum_{m=1}^{\infty} \langle \boldsymbol{x} \,|\, \boldsymbol{x}_m \rangle \sum_{n=1}^{\infty} \langle \boldsymbol{x}_m \,|\, \boldsymbol{u}_n \rangle \boldsymbol{u}_n$$

$$= \sum_{m=1}^{\infty} \langle \boldsymbol{x} \,|\, \boldsymbol{x}_m \rangle \boldsymbol{x}_m \qquad \text{by item (2d) page 272}$$

⇨

When is a set of orthonormal vectors in a Hilbert space $\boldsymbol{H}$ *total*? Theorem K.10 (next) offers some help.

Theorem K.10 (The Fourier Series Theorem). [25] *Let* $\left\{\boldsymbol{x}_n \in X \,\middle|\, n\in\mathbb{N}\right\}$ *be a set of vectors in a* HILBERT SPACE $\boldsymbol{H} \triangleq \left(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \,|\, \nabla \rangle\right)$ *and let* $\|\boldsymbol{x}\| \triangleq \sqrt{\langle \boldsymbol{x} \,|\, \boldsymbol{x} \rangle}$ *(Definition N.2 page 315).*

THM

(A) $\{\boldsymbol{x}_n\}$ *is* ORTHONORMAL *in* $\boldsymbol{H}$ $\Longrightarrow$

$$\left\{\begin{array}{llrcll} & (1). & \left(\operatorname{span}\{\boldsymbol{x}_n\}\right)^- & = & \boldsymbol{H} & & (\{\boldsymbol{x}_n\} \text{ is TOTAL in } \boldsymbol{H})\\ \iff & (2). & \langle \boldsymbol{x} \,|\, \boldsymbol{y} \rangle & \stackrel{\star}{=} & \displaystyle\sum_{n=1}^{\infty} \langle \boldsymbol{x} \,|\, \boldsymbol{x}_n \rangle \langle \boldsymbol{y} \,|\, \boldsymbol{x}_n \rangle^* & \forall \boldsymbol{x},\boldsymbol{y}\in X & (\text{GENERALIZED PARSEVAL'S IDENTITY})\\ \iff & (3). & \|\boldsymbol{x}\|^2 & \stackrel{\star}{=} & \displaystyle\sum_{n=1}^{\infty} \left|\langle \boldsymbol{x} \,|\, \boldsymbol{x}_n \rangle\right|^2 & \forall \boldsymbol{x}\in X & (\text{PARSEVAL'S IDENTITY})\\ \iff & (4). & \boldsymbol{x} & \stackrel{\star}{=} & \displaystyle\sum_{n=1}^{\infty} \langle \boldsymbol{x} \,|\, \boldsymbol{x}_n \rangle \boldsymbol{x}_n & \forall \boldsymbol{x}\in X & (\text{FOURIER SERIES EXPANSION}) \end{array}\right\}$$

✎PROOF:

1. Proof that (1) $\Longrightarrow$ (2):

$$\langle \boldsymbol{x} \,|\, \boldsymbol{y} \rangle \stackrel{\star}{=} \left\langle \sum_{n=1}^{\infty} \langle \boldsymbol{x} \,|\, \boldsymbol{x}_n \rangle \boldsymbol{x}_n \,\middle|\, \sum_{m=1}^{\infty} \langle \boldsymbol{y} \,|\, \boldsymbol{x}_m \rangle \boldsymbol{x}_m \right\rangle \qquad \text{by (A) and (1)}$$

$$\stackrel{\star}{=} \sum_{n=1}^{\infty} \langle \boldsymbol{x} \,|\, \boldsymbol{x}_n \rangle \left\langle \boldsymbol{x}_n \,\middle|\, \sum_{m=1}^{\infty} \langle \boldsymbol{y} \,|\, \boldsymbol{x}_m \rangle \boldsymbol{x}_m \right\rangle \qquad \text{by property of } \langle\cdot\,|\,\cdot\rangle \quad \text{(Definition N.1 page 309)}$$

$$\stackrel{\star}{=} \sum_{n=1}^{\infty} \langle \boldsymbol{x} \,|\, \boldsymbol{x}_n \rangle \sum_{m=1}^{\infty} \langle \boldsymbol{y} \,|\, \boldsymbol{x}_m \rangle^* \langle \boldsymbol{x}_n \,|\, \boldsymbol{x}_m \rangle \qquad \text{by property of } \langle\cdot\,|\,\cdot\rangle \quad \text{(Definition N.1 page 309)}$$

$$\stackrel{\star}{=} \sum_{n=1}^{\infty} \langle \boldsymbol{x} \,|\, \boldsymbol{x}_n \rangle \sum_{m=1}^{\infty} \langle \boldsymbol{y} \,|\, \boldsymbol{x}_m \rangle^* \bar{\delta}_{mn} \qquad \text{by (A)}$$

$$\stackrel{\star}{=} \sum_{n=1}^{\infty} \langle \boldsymbol{x} \,|\, \boldsymbol{x}_n \rangle \langle \boldsymbol{y} \,|\, \boldsymbol{x}_n \rangle^* \qquad \text{by definition of } \bar{\delta}_n \quad \text{(Definition N.3 page 323)}$$

[25] Bachman and Narici (1966) pages 149–155 ⟨Theorem 9.12⟩, Kubrusly (2001) pages 360–363 ⟨Theorem 5.48⟩, Aliprantis and Burkinshaw (1998) pages 298–299 ⟨Theorem 34.2⟩, Christensen (2003) page 57 ⟨Theorem 3.4.2⟩, Berberian (1961) pages 52–53 ⟨Theorem II§8.3⟩, Heil (2011) pages 34–35 ⟨Theorem 1.50⟩, Bracewell (1978) page 112 ⟨Rayleigh's theorem⟩

2. Proof that (2) $\implies$ (3):

$$\begin{aligned}
\|x\|^2 &\triangleq \langle x \,|\, x\rangle && \text{by definition of } \textit{induced norm} \quad \text{(Theorem N.4 page 314)}\\
&= \sum_{n=1}^{\infty} \langle x \,|\, x_n\rangle \langle x \,|\, x_n\rangle^* && \text{by (2)}\\
&= \sum_{n=1}^{\infty} |\langle x \,|\, x_n\rangle|^2
\end{aligned}$$

3. Proof that (3) $\iff$ (4) *not* using (A): by Proposition K.2 page 271

4. Proof that (3) $\implies$ (1) (proof by contradiction):

   (a) Suppose $\{x_n\}$ is *not total*.

   (b) Then there must exist a vector y in H such that the set $B \triangleq \{x_n\} \cup y$ is *orthonormal*.

   (c) Then $1 = \|y\|^2 \neq \sum_{n=1}^{\infty} |\langle y \,|\, x_n\rangle|^2 = 0$.

   (d) But this contradicts (3), and so $\{x_n\}$ must be *total* and (3) $\implies$ (1).

5. Extraneous proof that (3) $\implies$ (4) (this proof is not really necessary here):

$$\begin{aligned}
\left\| x - \sum_{n=1}^{\infty} \langle x \,|\, x_n\rangle x_n \right\|^2 &= \|x\|^2 - \sum_{n=1}^{\infty} |\langle x \,|\, x_n\rangle|^2 && \text{by } \textit{Bessel's Equality} \quad \text{(Theorem K.6 page 267)}\\
&= 0 && \text{by (3)}\\
\implies \quad x &\doteq \sum_{n=1}^{\infty} \langle x \,|\, x_n\rangle x_n && \text{by definition of } \doteq
\end{aligned}$$

6. Extraneous proof that (A) $\implies$ (4) (this proof is not really necessary here)

   (a) The sequence $\sum_{n=1}^{N} |\langle x \,|\, x_n\rangle|^2$ is *monotonically increasing* in n.

   (b) By Bessel's inequality (page 268), the sequence is upper bounded by $\|x\|^2$:

$$\sum_{n=1}^{N} |\langle x \,|\, x_n\rangle|^2 \leq \|x\|^2$$

   (c) Because this sequence is both monotonically increasing and bounded in n, it must equal its bound in the limit as n approaches infinity:

$$\lim_{N\to\infty} \sum_{n=1}^{N} |\langle x \,|\, x_n\rangle|^2 = \|x\|^2 \qquad \text{(K.1)}$$

   (d) If we combine this result with *Bessel's Equality* (Theorem K.6 page 267) we have

$$\begin{aligned}
\lim_{N\to\infty} \left\| x - \sum_{n=1}^{N} \langle x \,|\, x_n\rangle x_n \right\|^2 &= \|x\|^2 - \lim_{N\to\infty} \sum_{n=1}^{N} |\langle x \,|\, x_n\rangle|^2 && \text{by } \textit{Bessel's Equality} \text{ (Theorem K.6 page 267)}\\
&= \|x\|^2 - \|x\|^2 && \text{by equation (K.1) page 274}\\
&= 0
\end{aligned}$$

Proposition K.3 (Fourier expansion). *Let* $\{ \boldsymbol{x}_n \in X \,|\, n\in\mathbb{N} \}$ *be a set of vectors in a* HILBERT SPACE $\big(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \,|\, \triangledown \rangle\big)$.

PRP
$$\underbrace{\{\boldsymbol{x}_n\} \text{ is an ORTHONORMAL BASIS for } \boldsymbol{H}}_{(A)} \implies \left\{ \underbrace{\boldsymbol{x} \stackrel{\star}{=} \sum_{n=1}^{\infty} \alpha_n \boldsymbol{x}_n}_{(1)} \iff \underbrace{\alpha_n = \langle \boldsymbol{x} \,|\, \boldsymbol{x}_n \rangle}_{(2)} \right\}$$

✎PROOF:

1. Proof that (1) $\implies$ (2): by Lemma K.2 page 271
2. Proof that (1) $\impliedby$ (2):

$$\begin{aligned}
\left\| \boldsymbol{x} - \sum_{n=1}^{\infty} \alpha_n \boldsymbol{x}_{n\in\mathbb{Z}} \right\|^2 &= \left\| \boldsymbol{x} - \sum_{n=1}^{\infty} \langle \boldsymbol{x} \,|\, \boldsymbol{x}_n \rangle \boldsymbol{x}_{n\in\mathbb{Z}} \right\|^2 && \text{by right hypothesis} \\
&= \|\boldsymbol{x}\|^2 - \sum_{n=1}^{\infty} |\langle \boldsymbol{x} \,|\, \boldsymbol{x}_n \rangle|^2 && \text{by Bessel's equality} && \text{(Theorem K.6 page 267)} \\
&= 0 && \text{by } \textit{Parseval's Identity} && \text{(Theorem K.10 page 273)} \\
&\stackrel{\text{def}}{\iff} \quad \boldsymbol{x} \stackrel{\star}{=} \sum_{n=1}^{\infty} \langle \boldsymbol{x} \,|\, \boldsymbol{x}_n \rangle \boldsymbol{x}_n && \text{by definition of } \textit{strong convergence}
\end{aligned}$$

✏

Proposition K.4 (Riesz-Fischer Theorem). [26] *Let* $\{ \boldsymbol{x}_n \in X \,|\, n\in\mathbb{N} \}$ *be a set of vectors in a* HILBERT SPACE $\big(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \,|\, \triangledown \rangle\big)$.

PRP
$$\underbrace{\{\boldsymbol{x}_n\} \text{ is an ORTHONORMAL BASIS for } \boldsymbol{H}}_{(A)} \implies \left\{ \underbrace{\sum_{n=1}^{\infty} |\alpha_n|^2 < \infty}_{(1)} \iff \underbrace{\exists \boldsymbol{x}\in\boldsymbol{H} \text{ such that } \alpha_n = \langle \boldsymbol{x} \,|\, \boldsymbol{x}_n \rangle}_{(2)} \right\}$$

✎PROOF:

1. Proof that (1) $\implies$ (2):
   (a) If (1) is true, then let $\boldsymbol{x} \triangleq \sum_{n\in\mathbb{N}} \alpha_n \boldsymbol{x}_n$.
   (b) Then

$$\begin{aligned}
\langle \boldsymbol{x} \,|\, \boldsymbol{x}_n \rangle &= \left\langle \sum_{m\in\mathbb{N}} \alpha_m \boldsymbol{x}_m \,\middle|\, \boldsymbol{x}_n \right\rangle && \text{by definition of } \boldsymbol{x} \\
&= \sum_{m\in\mathbb{N}} \alpha_m \langle \boldsymbol{x}_m \,|\, \boldsymbol{x}_n \rangle && \text{by } \textit{homogeneous} \text{ property of } \langle \triangle \,|\, \triangledown \rangle && \text{(Definition N.1 page 309)} \\
&= \sum_{m\in\mathbb{N}} \alpha_m \bar{\delta}_{mn} && \text{by (A)} \\
&= \sum_{m\in\mathbb{N}} \alpha_n && \text{by definition of } \bar{\delta} && \text{(Definition N.3 page 323)}
\end{aligned}$$

[26] Young (2001) page 6

2. Proof that (1) $\Longleftarrow$ (2):

$$\begin{aligned}
\sum_{n\in\mathbb{N}} |\alpha_n|^2 &= \sum_{n\in\mathbb{N}} |\langle \boldsymbol{x} \,|\, \boldsymbol{x}_n \rangle|^2 && \text{by (2)} \\
&\le \|\boldsymbol{x}\|^2 && \text{by } \textit{Bessel's Inequality} \quad \text{(Theorem K.7 page 268)} \\
&\le \infty
\end{aligned}$$

✏

Theorem K.11. [27]

THM *All* SEPARABLE HILBERT SPACES *are* ISOMORPHIC. *That is,*

$$\left\{\begin{array}{l} \boldsymbol{X} \textit{ is a separable} \\ \textit{Hilbert space} \quad \textit{and} \\ \boldsymbol{Y} \textit{ is a separable} \\ \textit{Hilbert space} \end{array}\right\} \implies \left\{\begin{array}{lllll} \multicolumn{5}{l}{\textit{there is a } \text{BIJECTIVE} \textit{ operator } \mathbf{M} \in \boldsymbol{Y}^{\boldsymbol{X}} \textit{ such that}} \\ (1). & \boldsymbol{y} & = \mathbf{M}\boldsymbol{x} & \forall x\in\boldsymbol{X}, y\in\boldsymbol{Y} & \textit{and} \\ (2). & \|\mathbf{M}\boldsymbol{x}\| & = \|\boldsymbol{x}\| & \forall x\in\boldsymbol{X} & \textit{and} \\ (3). & \langle \mathbf{M}\boldsymbol{x} \,|\, \mathbf{M}\boldsymbol{y}\rangle & = \langle \boldsymbol{x} \,|\, \boldsymbol{y}\rangle & \forall x\in\boldsymbol{X}, y\in\boldsymbol{Y} & \end{array}\right\}$$

✎PROOF:

1. Let $\boldsymbol{X} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \,|\, \nabla \rangle)$ be a *separable Hilbert space* with *orthonormal basis* $\{\boldsymbol{x}_n | n\in\mathbb{N}\}$.
   Let $\boldsymbol{Y} \triangleq (Y, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \,|\, \nabla \rangle)$ be a *separable Hilbert space* with *orthonormal basis* $\{\boldsymbol{y}_n | n\in\mathbb{N}\}$.

2. Proof that there exists *bijective* operator $\mathbf{M}$ and its inverse $\mathbf{M}^{-1}$ between $\{\boldsymbol{x}_n\}$ and $\{\boldsymbol{y}_n\}$:

   (a) Let $\mathbf{M}$ be defined such that $\boldsymbol{y}_n \triangleq \mathbf{M}\boldsymbol{x}_n$.

   (b) Thus $\mathbf{M}$ is a *bijection* between $\{\boldsymbol{x}_n\}$ and $\{\boldsymbol{y}_n\}$.

   (c) Because $\mathbf{M}$ is a *bijection* between $\{\boldsymbol{x}_n\}$ and $\{\boldsymbol{y}_n\}$, $\mathbf{M}$ has an inverse operator $\mathbf{M}^{-1}$ between $\{\boldsymbol{x}_n\}$ and $\{\boldsymbol{y}_n\}$ such that $\boldsymbol{x}_n = \mathbf{M}^{-1}\boldsymbol{y}_n$.

3. Proof that $\mathbf{M}$ and $\mathbf{M}^{-1}$ are *bijective* operators between $\boldsymbol{X}$ and $\boldsymbol{Y}$:

   (a) Proof that $\mathbf{M}$ maps $\boldsymbol{X}$ *into* $\boldsymbol{Y}$:

$$\begin{aligned}
\boldsymbol{x} \in \boldsymbol{X} &\iff \boldsymbol{x} \doteq \sum_{n\in\mathbb{N}} \langle \boldsymbol{x} \,|\, \boldsymbol{x}_n \rangle \boldsymbol{x}_n && \text{by } \textit{Fourier expansion} \quad \text{(Theorem K.10 page 273)} \\
&\implies \exists \boldsymbol{y} \in \boldsymbol{Y} \quad \text{such that} \quad \langle \boldsymbol{y} \,|\, \boldsymbol{y}_n \rangle = \langle \boldsymbol{x} \,|\, \boldsymbol{x}_n \rangle && \text{by } \textit{Riesz-Fischer Thm.} \quad \text{(Proposition K.4 page 275)} \\
&\implies \\
\boldsymbol{y} &= \sum_{n\in\mathbb{N}} \langle \boldsymbol{y} \,|\, \boldsymbol{y}_n \rangle \boldsymbol{y}_n && \text{by } \textit{Fourier expansion} \quad \text{(Theorem K.10 page 273)} \\
&= \sum_{n\in\mathbb{N}} \langle \boldsymbol{x} \,|\, \boldsymbol{x}_n \rangle \boldsymbol{y}_n && \text{by } \textit{Riesz-Fischer Thm.} \quad \text{(Proposition K.4 page 275)} \\
&= \sum_{n\in\mathbb{N}} \langle \boldsymbol{x} \,|\, \boldsymbol{x}_n \rangle \mathbf{M}\boldsymbol{x}_n && \text{by definition of } \mathbf{M} \quad \text{(item (2a) page 276)} \\
&= \mathbf{M} \sum_{n\in\mathbb{N}} \langle \boldsymbol{x} \,|\, \boldsymbol{x}_n \rangle \boldsymbol{x}_n && \text{by prop. of linear ops.} \quad \text{(Theorem R.1 page 360)} \\
&= \mathbf{M}\boldsymbol{x} && \text{by definition of } \boldsymbol{x}
\end{aligned}$$

[27] Young (2001) page 6

(b) Proof that $\mathbf{M}^{-1}$ maps $\boldsymbol{Y}$ *into* $\boldsymbol{X}$:

$$\begin{aligned}
\boldsymbol{y}\in\boldsymbol{Y} \iff \boldsymbol{y} &\stackrel{*}{=} \sum_{n\in\mathbb{N}}\langle\boldsymbol{y}\,|\,\boldsymbol{y}_n\rangle\,\boldsymbol{y}_n && \text{by } \textit{Fourier expansion} && \text{(Theorem K.10 page 273)}\\
&\implies \exists\boldsymbol{x}\in\boldsymbol{X} \text{ such that } \langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle=\langle\boldsymbol{y}\,|\,\boldsymbol{y}_n\rangle && \text{by } \textit{Riesz-Fischer Thm.} && \text{(Proposition K.4 page 275)}\\
&\implies\\
\boldsymbol{x} &= \sum_{n\in\mathbb{N}}\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle\,\boldsymbol{x}_n && \text{by } \textit{Fourier expansion} && \text{(Theorem K.10 page 273)}\\
&= \sum_{n\in\mathbb{N}}\langle\boldsymbol{y}\,|\,\boldsymbol{y}_n\rangle\,\boldsymbol{x}_n && \text{by } \textit{Riesz-Fischer Thm.} && \text{(Proposition K.4 page 275)}\\
&= \sum_{n\in\mathbb{N}}\langle\boldsymbol{y}\,|\,\boldsymbol{y}_n\rangle\,\mathbf{M}^{-1}\boldsymbol{y}_n && \text{by definition of } \mathbf{M}^{-1} && \text{(item (2c) page 276)}\\
&= \mathbf{M}^{-1}\sum_{n\in\mathbb{N}}\langle\boldsymbol{y}\,|\,\boldsymbol{y}_n\rangle\,\boldsymbol{y}_n && \text{by prop. of } \textit{linear ops.} && \text{(Theorem R.1 page 360)}\\
&= \mathbf{M}^{-1}\boldsymbol{y} && \text{by definition of } \boldsymbol{y}
\end{aligned}$$

4. Proof for (2):

$$\begin{aligned}
\|\mathbf{M}\boldsymbol{x}\|^2 &= \left\|\mathbf{M}\sum_{n\in\mathbb{N}}\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle\,\boldsymbol{x}_n\right\|^2 && \text{by } \textit{Fourier expansion} && \text{(Theorem K.10 page 273)}\\
&= \left\|\sum_{n\in\mathbb{N}}\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle\,\mathbf{M}\boldsymbol{x}_n\right\|^2 && \text{by property of } \textit{linear operators} && \text{(Theorem R.1 page 360)}\\
&= \left\|\sum_{n\in\mathbb{N}}\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle\,\boldsymbol{y}_n\right\|^2 && \text{by definition of } \mathbf{M} && \text{(item (2a) page 276)}\\
&= \sum_{n\in\mathbb{N}}\left|\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle\right|^2 && \text{by } \textit{Parseval's Identity} && \text{(Proposition K.4 page 275)}\\
&= \left\|\sum_{n\in\mathbb{N}}\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle\,\boldsymbol{x}_n\right\|^2 && \text{by } \textit{Parseval's Identity} && \text{(Proposition K.4 page 275)}\\
&= \|\boldsymbol{x}\|^2 && \text{by } \textit{Fourier expansion} && \text{(Theorem K.10 page 273)}
\end{aligned}$$

5. Proof for (3): by (2) and Theorem R.23 page 380

✏

Theorem K.12. [28] *Let* $\boldsymbol{H}$ *be a* HILBERT SPACE.

THM			
	$\boldsymbol{H}$ *has a* SCHAUDER BASIS	$\iff$	$\boldsymbol{H}$ *is* SEPARABLE

Theorem K.13. [29] *Let* $\boldsymbol{H}$ *be a* HILBERT SPACE.

THM			
	$\boldsymbol{H}$ *has an* ORTHONORMAL BASIS	$\iff$	$\boldsymbol{H}$ *is* SEPARABLE

[28] Bachman et al. (2002) page 112 ⟨3.4.8⟩, Berberian (1961) page 53 ⟨Theorem II§8.3⟩
[29] Kubrusly (2001) page 357 ⟨Proposition 5.43⟩

K.6 Riesz bases in Hilbert spaces

Definition K.11. [30] *Let* $\{x_n \in X \,|\, n \in \mathbb{N}\}$ *be a set of vectors in a* SEPARABLE HILBERT SPACE $H \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \,|\, \nabla \rangle)$.

DEF: $\{x_n\}$ *is a* ***Riesz basis*** *for* H *if* $\{x_n\}$ *is* EQUIVALENT (Definition K.8 page 265) *to some* ORTHONORMAL BASIS (Definition K.9 page 271) *in* H.

Definition K.12. [31] *Let* $(x_n \in X)_{n\in\mathbb{N}}$ *be a sequence of vectors in a* SEPARABLE HILBERT SPACE $H \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \,|\, \nabla \rangle)$.

DEF: *The sequence* (x_n) *is a* ***Riesz sequence*** *for* H *if*

$$\exists A, B \in \mathbb{R}^+ \quad \text{such that} \quad A\sum_{n=1}^{\infty}|\alpha_n|^2 \le \left\|\sum_{n=1}^{\infty}\alpha_n x_n\right\|^2 \le B\sum_{n=1}^{\infty}|\alpha_n|^2 \qquad \forall (\alpha_n) \in \ell^2_{\mathbb{F}}.$$

Definition K.13. *Let* $X \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \,|\, \nabla \rangle)$ *be an* INNER PRODUCT SPACE (Definition N.1 page 309).

DEF: *The sequences* $(x_n \in X)_{n\in\mathbb{Z}}$ *and* $(y_n \in X)_{n\in\mathbb{Z}}$ *are* ***biorthogonal*** *with respect to each other in* X *if* $\quad \langle x_n \,|\, y_m \rangle = \bar{\delta}_{nm}$

Lemma K.3. [32] *Let* $\{x_n \,|\, n\in\mathbb{N}\}$ *be a sequence in a* HILBERT SPACE $X \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \,|\, \nabla \rangle)$. *Let* $\{y_n \,|\, n\in\mathbb{N}\}$ *be a sequence in a* HILBERT SPACE $Y \triangleq (Y, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \,|\, \nabla \rangle)$. *Let*

LEM:

$$\left\{\begin{array}{lll}
\text{(i).} & \{x_n\} \text{ is TOTAL in } X & \text{and}\\
\text{(ii).} & \text{There exists } A>0 \text{ such that } A\sum_{n\in C}|a_n|^2 \le \left\|\sum_{n\in C} a_n x_n\right\|^2 \quad \text{for finite } C & \text{and}\\
\text{(iii).} & \text{There exists } B>0 \text{ such that } \left\|\sum_{n=1}^{\infty} b_n y_n\right\|^2 \le B\sum_{n=1}^{\infty}|b_n|^2 \quad \forall (b_n)_{n\in\mathbb{N}} \in \ell^2_{\mathbb{F}} &
\end{array}\right\} \implies$$

$$\left\{\begin{array}{lll}
\text{(1).} & \mathbf{R}^{\circ} \text{ is a linear bounded operator that maps from } \operatorname{span}\{x_n\} \text{ to } \operatorname{span}\{y_n\} & \\
& \quad \text{where } \mathbf{R}^{\circ}\sum_{n\in C} c_n x_n \triangleq \sum_{n\in C} c_n y_n, \text{ for some sequence } (c_n) \text{ and finite set } C & \text{and}\\
\text{(2).} & \mathbf{R} \text{ has a unique extension to a bounded operator } \mathbf{R} \text{ that maps from } X \text{ to } Y & \text{and}\\
\text{(3).} & |||\mathbf{R}^{\circ}||| \le \frac{B}{A} & \text{and}\\
\text{(4).} & |||\mathbf{R}||| \le \frac{B}{A} &
\end{array}\right\}$$

Theorem K.14. [33] *Let* $\{x_n \in X \,|\, n\in\mathbb{N}\}$ *be a set of vectors in a* SEPARABLE HILBERT SPACE $H \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \,|\, \nabla \rangle)$.

THM:

$$\left\{\begin{array}{l}\{x_n\} \text{ is a RIESZ BASIS}\\ \text{for } H\end{array}\right\} \iff \left\{\begin{array}{lll}
\text{(1).} & \{x_n\} \text{ is TOTAL in } H & \text{and}\\
\text{(2).} & \exists A,B\in\mathbb{R}^+ \text{ such that } \forall (\alpha_n)\in\ell^2_{\mathbb{F}}, & \\
& A\sum_{n=1}^{\infty}|\alpha_n|^2 \le \left\|\sum_{n=1}^{\infty}\alpha_n x_n\right\|^2 \le B\sum_{n=1}^{\infty}|\alpha_n|^2 &
\end{array}\right\}$$

✎PROOF:

[30] Young (2001) page 27 ⟨Definition 1.8.2⟩, Christensen (2003) page 63 ⟨Definition 3.6.1⟩, Heil (2011) page 196 ⟨Definition 7.9⟩

[31] Christensen (2003) pages 66–68 ⟨page 68 and (3.24) on page 66⟩, Wojtaszczyk (1997) page 20 ⟨Definition 2.6⟩

[32] Christensen (2003) pages 65–66 ⟨Lemma 3.6.5⟩

[33] Young (2001) page 27 ⟨Theorem 1.8.9⟩, Christensen (2003) page 66 ⟨Theorem 3.6.6⟩, Heil (2011) pages 197–198 ⟨Theorem 7.13⟩, Christensen (2008) pages 61–62 ⟨Theorem 3.3.7⟩

1. Proof for ($\Longrightarrow$) case:
   (a) Proof that *Riesz basis* hypothesis $\Longrightarrow$ (1): all bases for $\boldsymbol{H}$ are *total* in $\boldsymbol{H}$.
   (b) Proof that *Riesz basis* hypothesis $\Longrightarrow$ (2):
      i. Let $(\boldsymbol{u}_n)_{n\in\mathbb{N}}$ be an *orthonormal basis* for $\boldsymbol{H}$.
      ii. Let $\mathbf{R}$ be a *bounded bijective* operator such that $\boldsymbol{x}_n = \mathbf{R}\boldsymbol{u}_n$.
      iii. Proof for upper bound B:

$$
\begin{aligned}
\left\|\sum_{n=1}^{\infty}\alpha_n\boldsymbol{x}_n\right\|^2 &= \left\|\sum_{n=1}^{\infty}\alpha_n\mathbf{R}\boldsymbol{u}_n\right\|^2 && \text{by definition of } \mathbf{R} && \text{(item (1(b)ii))}\\
&= \left\|\mathbf{R}\sum_{n=1}^{\infty}\alpha_n\boldsymbol{u}_n\right\|^2 && \text{by Theorem R.1 page 360}\\
&\le \|\mathbf{R}\|^2\left\|\sum_{n=1}^{\infty}\alpha_n\boldsymbol{u}_n\right\|^2 && \text{by Theorem R.6 page 366}\\
&= \|\mathbf{R}\|^2\sum_{n=1}^{\infty}\left\|\alpha_n\boldsymbol{u}_n\right\|^2 && \text{by \textit{Pythagorean Theorem}} && \text{(Theorem N.10 page 324)}\\
&= \|\mathbf{R}\|^2\sum_{n=1}^{\infty}|\alpha|^2\left\|\boldsymbol{u}_n\right\|^2 && \text{by \textit{homogeneous} property of norms} && \text{(Definition O.1 page 327)}\\
&= \underbrace{\|\mathbf{R}\|^2}_{B}\sum_{n=1}^{\infty}|\alpha|^2 && \text{by definition of \textit{orthonormality}} && \text{(Definition N.4 page 323)}
\end{aligned}
$$

      iv. Proof for lower bound A:

$$
\begin{aligned}
\left\|\sum_{n=1}^{\infty}\alpha_n\boldsymbol{x}_n\right\|^2 &= \frac{\|\mathbf{R}^{-1}\|^2}{\|\mathbf{R}^{-1}\|^2}\left\|\sum_{n=1}^{\infty}\alpha_n\boldsymbol{x}_n\right\|^2 && \text{because } \|\mathbf{R}^{-1}\| > 0 && \text{(Proposition R.1 page 364)}\\
&\ge \frac{1}{\|\mathbf{R}^{-1}\|^2}\left\|\mathbf{R}^{-1}\sum_{n=1}^{\infty}\alpha_n\boldsymbol{x}_n\right\|^2 && \text{by Theorem R.6 page 366}\\
&= \frac{1}{\|\mathbf{R}^{-1}\|^2}\left\|\mathbf{R}^{-1}\sum_{n=1}^{\infty}\alpha_n\mathbf{R}\boldsymbol{u}_n\right\|^2 && \text{by definition of } \mathbf{R} && \text{(item (1(b)ii) page 279)}\\
&= \frac{1}{\|\mathbf{R}^{-1}\|^2}\left\|\mathbf{R}^{-1}\mathbf{R}\sum_{n=1}^{\infty}\alpha_n\boldsymbol{u}_n\right\|^2 && \text{by property of \textit{linear operators}} && \text{(Theorem R.1 page 360)}\\
&= \frac{1}{\|\mathbf{R}^{-1}\|^2}\left\|\sum_{n=1}^{\infty}\alpha_n\boldsymbol{u}_n\right\|^2 && \text{by definition of inverse op.} && \text{(Definition R.2 page 359)}\\
&= \frac{1}{\|\mathbf{R}^{-1}\|^2}\sum_{n=1}^{\infty}\left\|\alpha_n\boldsymbol{u}_n\right\|^2 && \text{by \textit{Pythagorean Theorem}} && \text{(Theorem N.10 page 324)}\\
&= \frac{1}{\|\mathbf{R}^{-1}\|^2}\sum_{n=1}^{\infty}\left|\alpha_n\right|^2\left\|\boldsymbol{u}_n\right\|^2 && \text{by } \|\cdot\| \text{ \textit{homogeneous} prop.} && \text{(Definition O.1 page 327)}\\
&= \underbrace{\frac{1}{\|\mathbf{R}^{-1}\|^2}}_{A}\sum_{n=1}^{\infty}\left|\alpha_n\right|^2 && \text{by def. of \textit{orthonormality}} && \text{(Definition N.4 page 323)}
\end{aligned}
$$

2. Proof for ($\Longrightarrow$) case:

(a) Let $\{u_n | n\in\mathbb{N}\}$ be an *orthonormal basis* for $\boldsymbol{H}$.

(b) Using (2) and Lemma K.3 (page 278), construct an bounded extension operator $\mathbf{R}$ such that $\mathbf{R}u_n = x_n$ for all $n \in \mathbb{N}$.

(c) Using (2) and Lemma K.3 (page 278), construct an bounded extension operator $\mathbf{S}$ such that $\mathbf{S}x_n = u_n$ for all $n \in \mathbb{N}$.

(d) Then, $\mathbf{RV}x = \mathbf{VR}x \implies \mathbf{V} = \mathbf{R}^{-1}$, and so $\mathbf{R}$ is a bounded invertible operator

(e) and $\{x_n\}$ is a *Riesz sequence*.

✏

Theorem K.15. [34] *Let* $\boldsymbol{H} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \mid \nabla \rangle)$ *be a* SEPARABLE HILBERT SPACE.

THM

$$\left\{ \begin{array}{l} (\!(x_n \in \boldsymbol{H})\!)_{n\in\mathbb{Z}} \text{ is a} \\ \text{RIESZ BASIS for } \boldsymbol{H} \end{array} \right\} \implies \left\{ \begin{array}{lll} \multicolumn{3}{l}{\text{There exists } (\!(y_n \in \boldsymbol{H})\!)_{n\in\mathbb{Z}} \text{ such that}} \\ (1). & (\!(x_n)\!) \text{ and } (\!(y_n)\!) \text{ are BIORTHOGONAL} & \text{and} \\ (2). & (\!(y_n)\!) \text{ is also a RIESZ BASIS for } \boldsymbol{H} & \text{and} \\ (3). & \exists B>A>0 \text{ such that} & \\ & \displaystyle A\sum_{n=1}^{\infty}|a_n|^2 \le \left\|\sum_{n=1}^{\infty}a_nx_n\right\|^2 = \left\|\sum_{n=1}^{\infty}a_ny_n\right\|^2 \le B\sum_{n=1}^{\infty}|a_n|^2 & \\ & \forall (\!(a_n)\!)_{n\in\mathbb{N}} \in \ell^2_{\mathbb{F}} & \end{array} \right\}$$

✎PROOF:

1. Proof for (1):

   (a) Let e_n be the *unit vector* in $\boldsymbol{H}$ such that the nth element of e_n is 1 and all other elements are 0.

   (b) Let $\mathbf{M}$ be an operator on $\boldsymbol{H}$ such that $\mathbf{M}e_n = x_n$.

   (c) Note that $\mathbf{M}$ is *isometric*, and as such $\|\mathbf{M}x\| = \|x\| \quad \forall x\in\boldsymbol{H}$.

   (d) Let $y_n \triangleq (\mathbf{M}^{-1})^*$.

   (e) Then,

$$\begin{aligned} \langle y_n \mid x_m \rangle &= \left\langle (\mathbf{M}^{-1})^* e_n \mid \mathbf{M}e_m \right\rangle \\ &= \left\langle e_n \mid \mathbf{M}^{-1}\mathbf{M}e_m \right\rangle \\ &= \langle e_n \mid e_m \rangle \\ &= \bar{\delta}_{nm} \\ &\implies \{x_n\} \text{ and } \{y_n\} \text{ are } \textit{biorthogonal} && \text{by Definition N.4 page 323} \end{aligned}$$

[34] Wojtaszczyk (1997) page 20 ⟨Lemma 2.7(a)⟩

2. Proof for (3):

$$\begin{aligned}
\left\|\sum_{n\in\mathbb{Z}} \alpha_n \boldsymbol{y}_n\right\| &= \left\|\sum_{n\in\mathbb{Z}} \alpha_n \left(\mathbf{M}^{-1}\right)^* \boldsymbol{e}_n\right\| && \text{by definition of } \boldsymbol{y}_n && \text{(Proposition 1d page 280)}\\
&= \left\|\left(\mathbf{M}^{-1}\right)^* \sum_{n\in\mathbb{Z}} \alpha_n \boldsymbol{e}_n\right\| && \text{by property of } \textit{linear ops.}\\
&= \left\|\sum_{n\in\mathbb{Z}} \alpha_n \boldsymbol{e}_n\right\| && \text{because } \left(\mathbf{M}^{-1}\right)^* \text{ is } \textit{isometric} && \text{(Definition R.10 page 379)}\\
&= \left\|\mathbf{M} \sum_{n\in\mathbb{Z}} \alpha_n \boldsymbol{e}_n\right\| && \text{because } \mathbf{M} \text{ is } \textit{isometric} && \text{(Definition R.10 page 379)}\\
&= \left\|\sum_{n\in\mathbb{Z}} \alpha_n \mathbf{M} \boldsymbol{e}_n\right\| && \text{by property of } \textit{linear operators}\\
&= \left\|\sum_{n\in\mathbb{Z}} \alpha_n \boldsymbol{x}_n\right\| && \text{by definition of } \mathbf{M}\\
&\implies \left\{\boldsymbol{y}_n\right\} \text{ is a } \textit{Riesz basis} && \text{by left hypothesis}
\end{aligned}$$

3. Proof for (2): by (3) and definition of *Riesz basis* (Definition K.11 page 278)

✏

Proposition K.5. [35] *Let* $\left\{\boldsymbol{x}_n \middle| n\in\mathbb{N}\right\}$ *be a set of vectors in a* HILBERT SPACE $\boldsymbol{H} \triangleq \left(X,\ +,\ \cdot,\ (\mathbb{F},\ \dot{+},\ \dot{\times}),\ \langle\triangle\,|\,\nabla\rangle\right)$.

PRP

$$\left\{\begin{array}{l} \left\{\boldsymbol{x}_n\right\} \text{ is a } \textsc{Riesz basis} \text{ for } \boldsymbol{H} \text{ with}\\ A\sum_{n=1}^{\infty}\left|a_n\right|^2 \le \left\|\sum_{n=1}^{\infty} a_n \boldsymbol{x}_n\right\|^2 \le B\sum_{n=1}^{\infty}\left|a_n\right|^2\\ \forall\left\{a_n\right\}\in\boldsymbol{\ell}^2_{\mathbb{F}}\end{array}\right\} \implies \left\{\begin{array}{l}\left\{\boldsymbol{x}_n\right\} \text{ is a } \text{FRAME} \text{ for } \boldsymbol{H} \text{ with}\\ \underbrace{\frac{1}{B}\|\boldsymbol{x}\|^2 \le \sum_{n=1}^{\infty}\left|\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle\right|^2 \le \frac{1}{A}\|\boldsymbol{x}\|^2}_{\text{STABILITY CONDITION}}\\ \forall\boldsymbol{x}\in\boldsymbol{H}\end{array}\right\}$$

✎PROOF:

1. Let $\left\{\boldsymbol{y}_n \middle| n\in\mathbb{N}\right\}$ be a *Riesz basis* that is *biorthogonal* to $\left\{\boldsymbol{x}_n \middle| n\in\mathbb{N}\right\}$ (Theorem K.15 page 280).

2. Let $\boldsymbol{x} \triangleq \sum_{n=1}^{\infty} a_n \boldsymbol{y}_n$.

3. lemma:

$$\begin{aligned}
\sum_{n=1}^{\infty}\left|\langle\boldsymbol{x}\,|\,\boldsymbol{x}_n\rangle\right|^2 &= \sum_{n=1}^{\infty}\left|\left\langle\sum_{m=1}^{\infty} a_n \boldsymbol{y}_m \,\middle|\, \boldsymbol{x}_n\right\rangle\right|^2 && \text{by definition of } \boldsymbol{x} && \text{(item (2) page 281)}\\
&= \sum_{n=1}^{\infty}\left|\sum_{m=1}^{\infty} a_n \langle\boldsymbol{y}_m\,|\,\boldsymbol{x}_n\rangle\right|^2 && \text{by } \textit{homogeneous} \text{ property of } \langle\triangle\,|\,\nabla\rangle && \text{(Definition N.1 page 309)}\\
&= \sum_{n=1}^{\infty}\left|\sum_{m=1}^{\infty} a_n \bar{\delta}_{mn}\right|^2 && \text{by definition of } \textit{biorthogonal} && \text{(Definition K.13 page 278)}\\
&= \sum_{n=1}^{\infty}\left|a_n\right|^2 && \text{by definition of } \bar{\delta} && \text{(Definition N.3 page 323)}
\end{aligned}$$

[35] Igari (1996) page 220 ⟨Lemma 9.8⟩, Wojtaszczyk (1997) pages 20–21 ⟨Lemma 2.7(a)⟩

4. Then

$$\begin{array}{rllll}
& A\sum_{n=1}^{\infty}|a_n|^2 \le & \left\|\sum_{n=1}^{\infty}a_n x_n\right\|^2 & \le B\sum_{n=1}^{\infty}|a_n|^2 & \text{by definition of } \{y_n\} \text{ (item (1) page 281)}\\
\implies & A\sum_{n=1}^{\infty}|a_n|^2 \le & \left\|\sum_{n=1}^{\infty}a_n y_n\right\|^2 & \le B\sum_{n=1}^{\infty}|a_n|^2 & \text{by definition of } \{y_n\} \text{ (item (1) page 281)}\\
\implies & A\sum_{n=1}^{\infty}|a_n|^2 \le & \|x\|^2 & \le B\sum_{n=1}^{\infty}|a_n|^2 & \text{by definition of } x \text{ (item (2) page 281)}\\
\implies & A\sum_{n=1}^{\infty}|\langle x\,|\,x_n\rangle|^2 \le & \|x\|^2 & \le B\sum_{n=1}^{\infty}|\langle x\,|\,x_n\rangle|^2 & \text{by (3) lemma}\\
\implies & \frac{1}{B}\|x\|^2 \le & \sum_{n=1}^{\infty}|\langle x\,|\,x_n\rangle|^2 & \le \frac{1}{A}\|x\|^2 &
\end{array}$$

⇨

Theorem K.16 (Battle-Lemarié orthogonalization).[36] *Let* $\tilde{f}(\omega)$ *be the* FOURIER TRANSFORM *(Definition T.2 page 408) of a function* $f \in L^2_{\mathbb{R}}$.

THM

$$\left\{\begin{array}{ll}
1. & \{\mathbf{T}^n g \,|\, n\in\mathbb{Z}\} \text{ is a RIESZ BASIS for } L^2_{\mathbb{R}} \text{ and}\\
2. & \tilde{f}(\omega) \triangleq \dfrac{\tilde{g}(\omega)}{\sqrt{2\pi\sum_{n\in\mathbb{Z}}|\tilde{g}(\omega+2\pi n)|^2}}
\end{array}\right\} \implies \left\{\begin{array}{l}\{\mathbf{T}^n f \,|\, n\in\mathbb{Z}\}\\ \text{is an ORTHONORMAL BASIS for } L^2_{\mathbb{R}}\end{array}\right\}$$

✎PROOF:

1. Proof that $\{\mathbf{T}^n f \,|\, n\in\mathbb{Z}\}$ is orthonormal:

$$\begin{aligned}
\tilde{S}_{\phi\phi}(\omega) &= 2\pi\sum_{n\in\mathbb{Z}}\left|\tilde{f}(\omega+2\pi n)\right|^2 && \text{by Theorem ?? page ??}\\
&= 2\pi\sum_{n\in\mathbb{Z}}\left|\frac{\tilde{g}(\omega+2\pi n)}{\sqrt{2\pi\sum_{m\in\mathbb{Z}}|\tilde{g}(\omega+2\pi(n-m))|^2}}\right|^2 && \text{by left hypothesis}\\
&= \sum_{n\in\mathbb{Z}}\left|\frac{\tilde{g}(\omega+2\pi n)}{\sqrt{\sum_{m\in\mathbb{Z}}|\tilde{g}(\omega+2\pi m)|^2}}\right|^2\\
&= \sum_{n\in\mathbb{Z}}\left|\frac{1}{\sqrt{\sum_{m\in\mathbb{Z}}|\tilde{g}(\omega+2\pi m)|^2}}\right|^2|\tilde{g}(\omega+2\pi n)|^2\\
&= \frac{1}{\sum_{m\in\mathbb{Z}}|\tilde{g}(\omega+2\pi m)|^2}\sum_{n\in\mathbb{Z}}|\tilde{g}(\omega+2\pi n)|^2\\
&= 1\\
\implies & \{f_n \,|\, n\in\mathbb{Z}\} \text{ is orthonormal} && \text{by Theorem ?? page ??}
\end{aligned}$$

[36] Wojtaszczyk (1997) page 25 ⟨Remark 2.4⟩, Vidakovic (1999) page 71, Mallat (1989) page 72, Mallat (1999) page 225, Daubechies (1992) page 140 ⟨(5.3.3)⟩

2. Proof that $\{\mathbf{T}^n\mathsf{f} \mid n\in\mathbb{Z}\}$ is a basis for $\boldsymbol{V}_0$: by Lemma K.1 page 265.

⇨

K.7 Frames in Hilbert spaces

Definition K.14. [37] *Let $\{\boldsymbol{x}_n \in X \mid n\in\mathbb{N}\}$ be a set of vectors in a* HILBERT SPACE $\boldsymbol{H} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \mid \triangledown \rangle)$.

DEF

*The set $\{\boldsymbol{x}_n\}$ is a **frame** for $\boldsymbol{H}$ if* (STABILITY CONDITION)

$$\exists A, B \in \mathbb{R}^+ \quad \text{such that} \qquad A\,\|\boldsymbol{x}\|^2 \le \sum_{n=1}^{\infty} \left|\langle \boldsymbol{x} \mid \boldsymbol{x}_n\rangle\right|^2 \le B\,\|\boldsymbol{x}\|^2 \qquad \forall x \in X.$$

*The quantities A and B are **frame bounds**.*
*The quantity A' is the **optimal lower frame bound** if*
$A' = \sup\{A \in \mathbb{R}^+ \mid A \text{ is a lower frame bound}\}$.
*The quantity B' is the **optimal upper frame bound** if*
$B' = \inf\{B \in \mathbb{R}^+ \mid B \text{ is an upper frame bound}\}$.
*A frame is a **tight frame** if $A = B$.*
*A frame is a **normalized tight frame** (or a **Parseval frame**) if $A = B = 1$.*
*A frame $\{\boldsymbol{x}_n \mid n\in\mathbb{N}\}$ is an **exact frame** if for some $m \in \mathbb{Z}$, $\{\boldsymbol{x}_n \mid n\in\mathbb{N}\} \setminus \{\boldsymbol{x}_m\}$ is* NOT *a frame.*

A frame is a *Parseval frame* (Definition K.14) if it satisfies *Parseval's Identity* (Theorem K.10 page 273). All orthonormal bases are Parseval frames (Theorem K.10 page 273); but not all Parseval frames are orthonormal bases.

Definition K.15. *Let $\{\boldsymbol{x}_n\}$ be a **frame*** (Definition K.14 page 283) *for the* HILBERT SPACE $\boldsymbol{H} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \mid \triangledown \rangle)$. *Let $\mathbf{S}$ be an* OPERATOR *on $\boldsymbol{H}$.*

DEF

*$\mathbf{S}$ is a **frame operator** for $\{\boldsymbol{x}_n\}$ if* $\quad \mathbf{S}\mathsf{f}(x) = \sum_{n\in\mathbb{Z}} \langle \mathsf{f} \mid \boldsymbol{x}_n\rangle\, \boldsymbol{x}_n \qquad \forall \mathsf{f}\in\boldsymbol{H}.$

Theorem K.17. [38] *Let $\mathbf{S}$ be a* FRAME OPERATOR (Definition K.15 page 283) *of a* FRAME $\{\boldsymbol{x}_n\}$ (Definition K.14 page 283) *for the* HILBERT SPACE $\boldsymbol{H} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \mid \triangledown \rangle)$.

THM

(1). $\mathbf{S}$ *is* INVERTIBLE. *and*

(2). $\mathsf{f}(x) = \sum_{n\in\mathbb{Z}} \langle \mathsf{f} \mid \mathbf{S}^{-1}\boldsymbol{x}_n\rangle\, \boldsymbol{x}_n = \sum_{n\in\mathbb{Z}} \langle \mathsf{f} \mid \boldsymbol{x}_n\rangle\, \mathbf{S}^{-1}\boldsymbol{x}_n \quad \forall \mathsf{f}\in\boldsymbol{H}$

Theorem K.18. [39] *Let $\{\boldsymbol{x}_n \in X \mid n=1,2,\ldots,N\}$ be a set of vectors in a* HILBERT SPACE $\boldsymbol{H} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \mid \triangledown \rangle)$.

THM

$\{\boldsymbol{x}_n\}$ *is a* FRAME *for* $\operatorname{span}\{\boldsymbol{x}_n\}$.

✎PROOF:

[37] Young (2001) pages 154–155, Christensen (2003) page 88 ⟨Definitions 5.1.1, 5.1.2⟩, Heil (2011) pages 204–205 ⟨Definition 8.2⟩, Jørgensen et al. (2008) page 267 ⟨Definition 12.22⟩, Duffin and Schaeffer (1952) page 343, Daubechies et al. (1986) page 1272

[38] Christensen (2008) pages 100–102 ⟨Theorem 5.1.7⟩

[39] Christensen (2003) page 3

1. Upper bound: Proof that there exists B such that $\sum_{n=1}^{N}\left|\langle \boldsymbol{x} \mid \boldsymbol{x}_n\rangle\right|^2 \le B\,\|\boldsymbol{x}\|^2 \quad \forall \boldsymbol{x}\in \boldsymbol{H}$:

$$\begin{aligned}
\sum_{n=1}^{N}\left|\langle \boldsymbol{x} \mid \boldsymbol{x}_n\rangle\right|^2 &\le \sum_{n=1}^{N}\langle \boldsymbol{x}_n \mid \boldsymbol{x}_n\rangle\,\langle \boldsymbol{x} \mid \boldsymbol{x}\rangle && \text{by \textit{Cauchy-Schwarz inequality} (Theorem N.2 page 310)}\\
&= \underbrace{\left\{\sum_{n=1}^{N}\|\boldsymbol{x}_n\|^2\right\}}_{B}\|\boldsymbol{x}\|^2
\end{aligned}$$

2. Lower bound: Proof that there exists A such that $A\,\|\boldsymbol{x}\|^2 \le \sum_{n=1}^{N}\left|\langle \boldsymbol{x} \mid \boldsymbol{x}_n\rangle\right|^2 \quad \forall \boldsymbol{x}\in \boldsymbol{H}$:

$$\begin{aligned}
\sum_{n=1}^{N}\left|\langle \boldsymbol{x} \mid \boldsymbol{x}_n\rangle\right|^2 &= \sum_{n=1}^{N}\left|\left\langle \boldsymbol{x}_n \;\middle|\; \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}\right\rangle\right|^2 \|\boldsymbol{x}\|^2\\
&\ge \underbrace{\left(\inf_{\boldsymbol{y}}\left\{\sum_{n=1}^{N}\left|\langle \boldsymbol{x}_n \mid \boldsymbol{y}\rangle\right|^2 \;\middle|\; \|\boldsymbol{y}\|=1\right\}\right)}_{A}\|\boldsymbol{x}\|^2
\end{aligned}$$

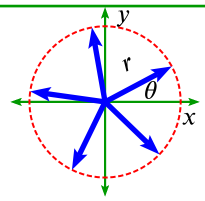
Example K.1. Let $\left(\mathbb{R}^2, +, \cdot, (\mathbb{R}, \dot{+}, \dot{\times}), \langle \triangle \mid \triangledown\rangle\right)$ be an inner product space with $\left\langle \begin{bmatrix} x_1\\ y_1\end{bmatrix} \middle| \begin{bmatrix} x_2\\ y_2\end{bmatrix}\right\rangle \triangleq x_1x_2 + y_1y_2$. Let $\mathbf{S}$ be the *frame operator* (Definition K.15 page 283) with *inverse* $\mathbf{S}^{-1}$.

EX

Let $N \in \{3,4,5,\ldots\}$, $\theta\in\mathbb{R}$, and $r\in\mathbb{R}^+$ ($r>0$).

Let $\boldsymbol{x}_n \triangleq r\begin{bmatrix}\cos(\theta + 2n\pi/N)\\ \sin(\theta+2n\pi/N)\end{bmatrix} \quad \forall n\in\{0,1,\ldots,N-1\}$.

Then, $\left(\!\left(\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{N-1}\right)\!\right)$ is a **tight frame** for $\mathbb{R}^2$ with *frame bound* $A = \frac{Nr^2}{2}$.



Moreover, $\quad \mathbf{S} = \frac{Nr^2}{2}\begin{bmatrix}1&0\\0&1\end{bmatrix}, \quad \mathbf{S}^{-1} = \frac{2}{Nr^2}\begin{bmatrix}1&0\\0&1\end{bmatrix}, \quad$ and $\quad \boldsymbol{v} = \frac{2}{Nr^2}\sum_{n=0}^{N-1}\langle \boldsymbol{v} \mid \boldsymbol{x}_n\rangle\,\boldsymbol{x}_n \quad \forall \boldsymbol{v}\in\mathbb{R}^2$.

✎PROOF:

1. Proof that $\left(\!\left(\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{N-1}\right)\!\right)$ is a *tight frame* with *frame bound* $A = \frac{Nr^2}{2}$: Let $\boldsymbol{v} \triangleq (x,y)\in\mathbb{R}^2$.

$$\begin{aligned}
\sum_{n=0}^{N-1}\left|\langle \boldsymbol{v} \mid \boldsymbol{x}_n\rangle\right|^2 &\triangleq \sum_{n=0}^{N-1}\left\|\boldsymbol{v}^{\mathbf{H}} r\begin{bmatrix}\cos\left(\theta + \frac{2n\pi}{N}\right)\\ \sin\left(\theta + \frac{2n\pi}{N}\right)\end{bmatrix}\right\|^2 && \text{by definitions of } \boldsymbol{v} \text{ of } \langle \boldsymbol{y} \mid \boldsymbol{x}\rangle\\
&\triangleq \sum_{n=0}^{N-1} r^2\left|x\cos\left(\theta+\frac{2n\pi}{N}\right) + y\sin\left(\theta + \frac{2n\pi}{N}\right)\right|^2 && \text{by definition of } \boldsymbol{y}^{\mathbf{H}}\boldsymbol{x} \text{ operation}\\
&= r^2x^2\sum_{n=0}^{N-1}\cos^2\left(\theta+\frac{2n\pi}{N}\right) + r^2y^2\sum_{n=0}^{N-1}\sin^2\left(\theta+\frac{2n\pi}{N}\right) + r^2xy\sum_{n=0}^{N-1}\cos\left(\theta+\frac{2n\pi}{N}\right)\sin\left(\theta+\frac{2n\pi}{N}\right)\\
&= r^2x^2\frac{N}{2} + r^2y^2\frac{N}{2} + r^2xy0 && \text{by Corollary ?? page ??}\\
&= (x^2+y^2)\frac{Nr^2}{2} = \left(\frac{Nr^2}{2}\right)\boldsymbol{v}^{\mathbf{H}}\boldsymbol{v} \triangleq \underbrace{\left(\frac{Nr^2}{2}\right)}_{A}\|\boldsymbol{v}\|^2 && \text{by definition of } \|\boldsymbol{v}\|
\end{aligned}$$

2. Proof that $\mathbf{S} = \frac{Nr^2}{2}\begin{bmatrix}1&0\\0&1\end{bmatrix}$:

(a) Let $\boldsymbol{e}_1 \triangleq \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\boldsymbol{e}_2 \triangleq \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

(b) lemma: $\mathbf{S}\boldsymbol{e}_1 = \frac{Nr^2}{2}\boldsymbol{e}_1$. Proof:

$$\begin{aligned}
\mathbf{S}\boldsymbol{e}_1 &= \sum_{n=0}^{N-1} \langle \boldsymbol{e}_1 \,|\, \boldsymbol{x}_n \rangle \boldsymbol{x}_n \\
&= \sum_{n=0}^{N-1} r\cos\left(\theta + \frac{2n\pi}{N}\right) r \begin{bmatrix} \cos\left(\theta + \frac{2n\pi}{N}\right) \\ \sin\left(\theta + \frac{2n\pi}{N}\right) \end{bmatrix} = r^2 \sum_{n=0}^{N-1} \begin{bmatrix} \cos^2\left(\theta + \frac{2n\pi}{N}\right) \\ \cos\left(\theta + \frac{2n\pi}{N}\right)\sin\left(\theta + \frac{2n\pi}{N}\right) \end{bmatrix} \\
&= r^2 \begin{bmatrix} N/2 \\ 0 \end{bmatrix} = \frac{Nr^2}{2}\boldsymbol{e}_1 && \text{by } \textit{Summation around unit circle} \text{ (Corollary ?? page ??)}
\end{aligned}$$

(c) lemma: $\mathbf{S}\boldsymbol{e}_2 = \frac{Nr^2}{2}\boldsymbol{e}_2$. Proof:

$$\begin{aligned}
\mathbf{S}\boldsymbol{e}_2 &= \sum_{n=0}^{N-1} \langle \boldsymbol{e}_2 \,|\, \boldsymbol{x}_n \rangle \boldsymbol{x}_n \\
&= \sum_{n=0}^{N-1} r\sin\left(\theta + \frac{2n\pi}{N}\right) r \begin{bmatrix} \cos\left(\theta + \frac{2n\pi}{N}\right) \\ \sin\left(\theta + \frac{2n\pi}{N}\right) \end{bmatrix} = r^2 \sum_{n=0}^{N-1} \begin{bmatrix} \sin\left(\theta + \frac{2n\pi}{N}\right)\cos\left(\theta + \frac{2n\pi}{N}\right) \\ \sin^2\left(\theta + \frac{2n\pi}{N}\right) \end{bmatrix} \\
&= r^2 \begin{bmatrix} 0 \\ N/2 \end{bmatrix} = \frac{Nr^2}{2}\boldsymbol{e}_2 && \text{by } \textit{Summation around unit circle} \text{ (Corollary ?? page ??)}
\end{aligned}$$

(d) Complete the proof of item (2) using *Eigendecomposition* $\mathbf{S} = \mathbf{Q}\Lambda\mathbf{Q}^{-1}$:

$$\begin{aligned}
\mathbf{S}\boldsymbol{e}_1 &= \frac{Nr^2}{2}\boldsymbol{e}_1 && \text{by (2c) lemma} \\
&\implies \boldsymbol{e}_1 \text{ is an } \textit{eigenvector} \text{ of } \mathbf{S} \text{ with } \textit{eigenvalue } \frac{Nr^2}{2} \\
\mathbf{S}\boldsymbol{e}_2 &= \frac{Nr^2}{2}\boldsymbol{e}_2 && \text{by (2c) lemma} \\
&\implies \boldsymbol{e}_2 \text{ is an } \textit{eigenvector} \text{ of } \mathbf{S} \text{ with } \textit{eigenvalue } \frac{Nr^2}{2}
\end{aligned}$$

$$\mathbf{S} = \overbrace{\underbrace{\begin{bmatrix} | & | \\ \boldsymbol{e}_1 & \boldsymbol{e}_2 \\ | & | \end{bmatrix}}_{\mathbf{Q}} \underbrace{\begin{bmatrix} \frac{Nr^2}{2} & 0 \\ 0 & \frac{Nr^2}{2} \end{bmatrix}}_{\Lambda} \underbrace{\begin{bmatrix} | & | \\ \boldsymbol{e}_1 & \boldsymbol{e}_2 \\ | & | \end{bmatrix}^{-1}}_{\mathbf{Q}^{-1}}}^{\textit{Eigendecomposition of } \mathbf{S}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{Nr^2}{2} & 0 \\ 0 & \frac{Nr^2}{2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^{-1} = \frac{Nr^2}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

3. Proof that $\mathbf{S}^{-1} = \frac{2}{Nr^2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$:

$$\begin{aligned}
\mathbf{S}\mathbf{S}^{-1} &= \frac{Nr^2}{2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \frac{2}{Nr^2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \triangleq \mathbf{I}_2 && \text{by item (2)} \\
\mathbf{S}^{-1}\mathbf{S} &= \frac{2}{Nr^2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \frac{Nr^2}{2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \triangleq \mathbf{I}_2 && \text{by item (2)}
\end{aligned}$$

4. Proof that $\boldsymbol{v} = \frac{2}{Nr^2}\sum_{n=0}^{N-1} (\boldsymbol{v}^{\mathbf{H}}\boldsymbol{x}_n)\boldsymbol{x}_n$:

$$\begin{aligned}
\boldsymbol{v} &= \sum_{n=0}^{N-1} \left\langle \boldsymbol{v} \,|\, \mathbf{S}^{-1}\boldsymbol{x}_n \right\rangle \boldsymbol{x}_n = \sum_{n=0}^{N-1} \left\langle \boldsymbol{v} \,\middle|\, \frac{2}{Nr^2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\boldsymbol{x}_n \right\rangle \boldsymbol{x}_n && \text{by item (3)} \\
&= \frac{2}{Nr^2}\sum_{n=0}^{N-1} \langle \boldsymbol{v} \,|\, \boldsymbol{x}_n \rangle \boldsymbol{x}_n = \frac{2}{Nr^2}\sum_{n=0}^{N-1} (\boldsymbol{v}^{\mathbf{H}}\boldsymbol{x}_n)\boldsymbol{x}_n && \text{by definition of } \langle \boldsymbol{y} \,|\, \boldsymbol{x} \rangle
\end{aligned}$$

Example K.2 (Peace Frame/Mercedes Frame). [40] Let $\left(\mathbb{R}^2, +, \cdot, (\mathbb{R}, \dot{+}, \dot{\times}), \langle \triangle \,|\, \triangledown \rangle\right)$ be an inner product space with $\left\langle (x_1, y_1) \,|\, (x_2, y_2) \right\rangle \triangleq x_1 y_1 + x_2 y_2$. Let $\mathbf{S}$ be the *frame operator* (Definition K.15 page 283) with *inverse* $\mathbf{S}^{-1}$.

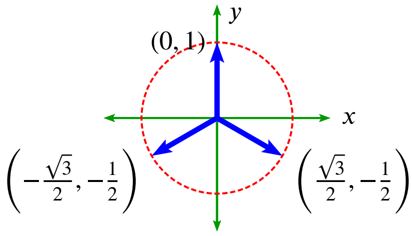
EX

Let $\boldsymbol{x}_1 \triangleq \frac{1}{2}\begin{bmatrix} 0 \\ 2 \end{bmatrix}$, $\boldsymbol{x}_2 \triangleq \frac{1}{2}\begin{bmatrix} -\sqrt{3} \\ -1 \end{bmatrix}$, and $\boldsymbol{x}_3 \triangleq \frac{1}{2}\begin{bmatrix} \sqrt{3} \\ -1 \end{bmatrix}$.

Then, $(\!(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3)\!)$ is a **tight frame** for $\mathbb{R}^2$ with *frame bound* $A = \frac{3}{2}$.

Moreover, $\mathbf{S} = \frac{3}{2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\quad \mathbf{S}^{-1} = \frac{2}{3}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$,

and $\quad \boldsymbol{v} = \frac{2}{3}\sum_{n=1}^{3} \langle \boldsymbol{v} \,|\, \boldsymbol{x}_n \rangle \boldsymbol{x}_n \triangleq \frac{2}{3}\sum_{n=1}^{3} \left(\boldsymbol{v}^{\mathbf{H}} \boldsymbol{x}_n\right) \boldsymbol{x}_n \qquad \forall \boldsymbol{v} \in \mathbb{R}^2.$



✎PROOF:

1. This frame is simply a special case of the frame presented in Example K.1 (page 284) with $r = 1$, $N = 3$, and $\theta = \pi/2$.

2. Let's give it a try! Let $\boldsymbol{v} \triangleq \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

$$\begin{aligned}
\sum_{n=1}^{3} \left\langle \boldsymbol{v} \,|\, \mathbf{S}^{-1}\boldsymbol{x}_n \right\rangle \boldsymbol{x}_n &= \frac{2}{3}\sum_{n=1}^{3} \left(\boldsymbol{v}^{\mathbf{H}}\boldsymbol{x}_n\right)\boldsymbol{x}_n && \text{by Example K.1 page 284} \\
&= \left(\boldsymbol{v}^{\mathbf{H}}\boldsymbol{x}_1\right)\boldsymbol{x}_1 + \left(\boldsymbol{v}^{\mathbf{H}}\boldsymbol{x}_2\right)\boldsymbol{x}_2 + \left(\boldsymbol{v}^{\mathbf{H}}\boldsymbol{x}_3\right)\boldsymbol{x}_3 \\
&= \frac{2}{3}\left(\left(\boldsymbol{v}^{\mathbf{H}}\frac{1}{2}\begin{bmatrix} 0 \\ 2 \end{bmatrix}\right)\boldsymbol{x}_1 + \left(\boldsymbol{v}^{\mathbf{H}}\frac{1}{2}\begin{bmatrix} -\sqrt{3} \\ -1 \end{bmatrix}\right)\boldsymbol{x}_2 + \left(\boldsymbol{v}^{\mathbf{H}}\frac{1}{2}\begin{bmatrix} \sqrt{3} \\ -1 \end{bmatrix}\right)\boldsymbol{x}_3\right) \\
&= \frac{2}{3}\cdot\frac{1}{2}\left(\left(\boldsymbol{v}^{\mathbf{H}}\begin{bmatrix} 0 \\ 2 \end{bmatrix}\right)\boldsymbol{x}_1 + \left(\boldsymbol{v}^{\mathbf{H}}\begin{bmatrix} -\sqrt{3} \\ -1 \end{bmatrix}\right)\boldsymbol{x}_2 + \left(\boldsymbol{v}^{\mathbf{H}}\begin{bmatrix} \sqrt{3} \\ -1 \end{bmatrix}\right)\boldsymbol{x}_3\right) \\
&= \frac{1}{3}\left((2)\frac{1}{2}\begin{bmatrix} 0 \\ 2 \end{bmatrix} + \left(-\sqrt{3}-1\right)\frac{1}{2}\begin{bmatrix} -\sqrt{3} \\ -1 \end{bmatrix} + \left(\sqrt{3}-1\right)\frac{1}{2}\begin{bmatrix} \sqrt{3} \\ -1 \end{bmatrix}\right) \\
&= \frac{1}{6}\begin{bmatrix} 2(0) & + & (-\sqrt{3}-1)(-\sqrt{3}) & + & (\sqrt{3}-1)(\sqrt{3}) \\ 2(2) & + & (-\sqrt{3}-1)(-1) & + & (\sqrt{3}-1)(-1) \end{bmatrix} \\
&= \frac{1}{6}\begin{bmatrix} 0 & + & (3+\sqrt{3}) & + & (3-\sqrt{3}) \\ 4 & + & (1+\sqrt{3}) & + & (1-\sqrt{3}) \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \triangleq \boldsymbol{v}
\end{aligned}$$

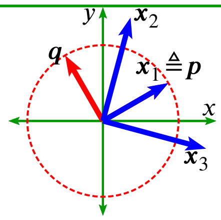
In Example K.1 (page 284) and Example K.2 (page 286), the frame operator $\mathbf{S}$ and its inverse $\mathbf{S}^{-1}$ were computed. In general however, it is not always necessary or even possible to compute these, as illustrated in Example K.3 (next).

Example K.3. [41] Let $\left(\mathbb{R}^2, +, \cdot, (\mathbb{R}, \dot{+}, \dot{\times}), \langle \triangle \,|\, \triangledown \rangle\right)$ be an inner product space with $\left\langle (x_1, y_1) \,|\, (x_2, y_2) \right\rangle \triangleq x_1 y_1 + x_2 y_2$. Let $\mathbf{S}$ be the *frame operator* (Definition K.15 page 283) with *inverse* $\mathbf{S}^{-1}$.

[40] Heil (2011) pages 204–205 ⟨$r = 1$ case⟩, Byrne (2005) page 80 ⟨$r = 1$ case⟩, Han et al. (2007) page 91 ⟨Example 3.9, $r = \sqrt{2/3}$ case⟩

[41] Christensen (2003) pages 7–8 ⟨?⟩

EX Let p and q be *orthonormal* vectors in $X \triangleq \operatorname{span}\{p, q\}$. Let $x_1 \triangleq p$, $x_2 \triangleq p + q$, and $x_3 \triangleq p - q$. Then, $\{x_1, x_2, , x_3\}$ is a **frame** for X with *frame bounds* $A = 0$ and $B = 5$.



Moreover,

$\mathbf{S}^{-1}x_1 = {}^1\!/\!_3 p$ and
$\mathbf{S}^{-1}x_2 = {}^1\!/\!_3 p + {}^1\!/\!_2 q$ and
$\mathbf{S}^{-1}x_3 = {}^1\!/\!_3 p - {}^1\!/\!_2 q$.

✎PROOF:

1. Proof that (x_1, x_2, x_3) is a *frame* with *frame bounds* $A = 0$ and $B = 5$:

$$\begin{aligned}
\sum_{n=1}^{3} |\langle v \mid x_n \rangle|^2 &\triangleq |\langle v \mid p\rangle|^2 + |\langle v \mid p+q\rangle|^2 + |\langle v \mid p-q\rangle|^2 && \text{by definitions of } x_1, x_2, \text{ and } x_3\\
&= |\langle v \mid p\rangle|^2 + |\langle v \mid p\rangle + \langle v \mid q\rangle|^2 + |\langle v \mid p\rangle - \langle v \mid q\rangle|^2 && \text{by \textit{additivity} of } \langle \triangle \mid \bigtriangledown \rangle \text{ (Definition N.1 page 309)}\\
&= |\langle v \mid p\rangle|^2 + \left(|\langle v \mid p\rangle|^2 + |\langle v \mid q\rangle|^2 + \langle v \mid p\rangle \langle v \mid q\rangle^* + \langle v \mid q\rangle \langle v \mid p\rangle^*\right)\\
&\quad + \left(|\langle v \mid p\rangle|^2 + |\langle v \mid q\rangle|^2 - \langle v \mid p\rangle \langle v \mid q\rangle^* - \langle v \mid q\rangle \langle v \mid p\rangle^*\right)\\
&= 3|\langle v \mid p\rangle|^2 + 2|\langle v \mid q\rangle|^2\\
&\boxed{\le} 3\,\|v\|\,\|p\| + 2\,\|v\|\,\|q\| && \text{by \textit{CS Inequality} (Theorem N.2 page 310)}\\
&= \|v\|\,(3\,\|p\| + 2\,\|q\|)\\
&= \boxed{5}\,\|v\| && \text{by \textit{orthonormality} of } p \text{ and } q
\end{aligned}$$

2. lemma: $\mathbf{S}p = 3p$, $\mathbf{S}q = 2q$, $\mathbf{S}^{-1}p = \frac{1}{3}p$, and $\mathbf{S}^{-1}q = \frac{1}{2}q$. Proof:

$$\begin{aligned}
\mathbf{S}p &\triangleq \sum_{n=1}^{3} \langle p \mid x_n \rangle x_n\\
&= \langle p \mid p\rangle p + \langle p \mid p+q\rangle (p+q) + \langle p \mid p-q\rangle (p-q)\\
&= (1)p + (1+0)(p+q) + (1-0)(p-q)\\
&= 3p\\
\implies \mathbf{S}^{-1}p &= \frac{1}{3}p\\
\mathbf{S}q &\triangleq \sum_{n=1}^{3} \langle q \mid x_n \rangle x_n\\
&= \langle q \mid p\rangle p + \langle q \mid p+q\rangle (p+q) + \langle q \mid p-q\rangle (p-q)\\
&= (0)q + (0+1)(p+q) + (0-1)(p-q)\\
&= 2q\\
\implies \mathbf{S}^{-1}q &= \frac{1}{2}q
\end{aligned}$$

3. Remark: Without knowing p and q, from (2) lemma it follows that it is not possible to compute $\mathbf{S}$ or $\mathbf{S}^{-1}$ explicitly.

4. Proof that $\mathbf{S}^{-1}x_1 = {}^1\!/\!_3 p$, $\mathbf{S}^{-1}x_2 = {}^1\!/\!_3 p + {}^1\!/\!_2 q$ and $\mathbf{S}^{-1}x_3 = {}^1\!/\!_3 p - {}^1\!/\!_2 q$:

$$\begin{aligned}
\mathbf{S}^{-1}x_1 &\triangleq \mathbf{S}^{-1}p && \text{by definition of } x_1\\
&= \frac{1}{3}p && \text{by (2) lemma}\\
\mathbf{S}^{-1}x_2 &\triangleq \mathbf{S}^{-1}(p+q) && \text{by definition of } x_2\\
&= \frac{1}{3}p + \frac{1}{2}q && \text{by (2) lemma}
\end{aligned}$$

$$\begin{aligned}
\mathbf{S}^{-1}\boldsymbol{x}_3 &\triangleq \mathbf{S}^{-1}(\boldsymbol{p}-\boldsymbol{q}) && \text{by definition of } \boldsymbol{x}_2 \\
&= \frac{1}{3}\boldsymbol{p} - \frac{1}{2}\boldsymbol{q} && \text{by (2) lemma}
\end{aligned}$$

5. Check that $\boldsymbol{v} = \sum_n \langle \boldsymbol{v} \,|\, \boldsymbol{x}_n \rangle \boldsymbol{x}_n = \langle \boldsymbol{v} \,|\, \boldsymbol{p} \rangle \boldsymbol{p} + \langle \boldsymbol{v} \,|\, \boldsymbol{q} \rangle \boldsymbol{q}$:

$$\begin{aligned}
\boldsymbol{v} &= \sum_{n=1}^{3} \left\langle \boldsymbol{v} \,|\, \mathbf{S}^{-1}\boldsymbol{x}_n \right\rangle \boldsymbol{x}_n \\
&= \left\langle \boldsymbol{v} \,|\, \mathbf{S}^{-1}\boldsymbol{p} \right\rangle \boldsymbol{p} + \left\langle \boldsymbol{v} \,|\, \mathbf{S}^{-1}(\boldsymbol{p}+\boldsymbol{q}) \right\rangle (\boldsymbol{p}+\boldsymbol{q}) + \left\langle \boldsymbol{v} \,|\, \mathbf{S}^{-1}(\boldsymbol{p}-\boldsymbol{q}) \right\rangle (\boldsymbol{p}-\boldsymbol{q}) \\
&= \left\langle \boldsymbol{v} \,|\, \tfrac{1}{3}\boldsymbol{p} \right\rangle \boldsymbol{p} + \left\langle \boldsymbol{v} \,|\, \tfrac{1}{3}\boldsymbol{p} + \tfrac{1}{2}\boldsymbol{q} \right\rangle (\boldsymbol{p}+\boldsymbol{q}) + \left\langle \boldsymbol{v} \,|\, \tfrac{1}{3}\boldsymbol{p} - \tfrac{1}{2}\boldsymbol{q}) \right\rangle (\boldsymbol{p}-\boldsymbol{q}) \\
&= \left(\frac{1}{3}+\frac{1}{3}+\frac{1}{3}\right) \langle \boldsymbol{v} \,|\, \boldsymbol{p} \rangle \boldsymbol{p} + \left(\frac{1}{3}-\frac{1}{3}\right) \langle \boldsymbol{v} \,|\, \boldsymbol{p} \rangle \boldsymbol{q} + \left(\frac{1}{2}-\frac{1}{2}\right) \langle \boldsymbol{v} \,|\, \boldsymbol{q} \rangle \boldsymbol{p} + \left(\frac{1}{2}+\frac{1}{2}\right) \langle \boldsymbol{v} \,|\, \boldsymbol{q} \rangle \boldsymbol{q} \\
&= \langle \boldsymbol{v} \,|\, \boldsymbol{p} \rangle \boldsymbol{p} + \langle \boldsymbol{v} \,|\, \boldsymbol{q} \rangle \boldsymbol{q}
\end{aligned}$$

✏

APPENDIX L

LINEAR SUBSPACES

L.1 Subspaces of a linear space

Linear spaces (Definition J.1 page 247) can be decomposed into a collection of *linear subspaces* (Definition L.1 page 290). Often such a collection along with an *order relation* forms a *lattice*.

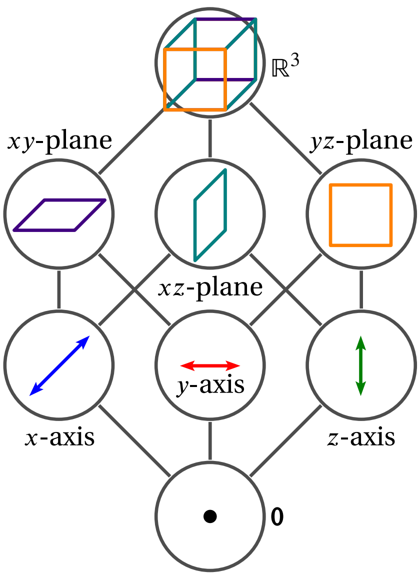
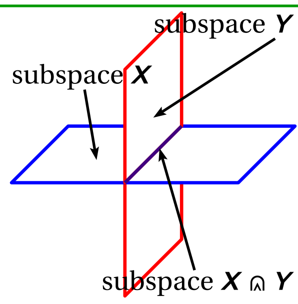


Figure L.1: lattice of subspaces of $\mathbb{R}^3$ (Example L.1 page 289)

EX

Example L.1. The 3-dimensional Euclidean space $\mathbb{R}^3$ contains the 2-dimensional xy-plane and xz-plane subspaces, which in turn both contain the 1-dimensional x-axis subspace. These subspaces are illustrated in the figure to the right and in Figure L.1 (page 289).



Definition L.1. [1] *Let* $\boldsymbol{\Omega} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$ *be a* LINEAR SPACE (Definition J.1 page 247).

DEF

A tupple $(Y, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$ *is a **linear subspace** of* $\boldsymbol{\Omega}$ *if*

1. $Y \neq \varnothing$ *(Y must contain at least one element)* *and*
2. $Y \subseteq X$ *(Y is a subset of X)* *and*
3. $\boldsymbol{x}, \boldsymbol{y} \in Y \implies \boldsymbol{x} + \boldsymbol{y} \in Y$ *(closed under vector addition)* *and*
4. $\boldsymbol{x} \in Y$ *and* $\alpha \in \mathbb{F} \implies \alpha\boldsymbol{x} \in Y$ *(closed under scalar-vector multiplication).*

*A linear subspace is also called a **linear manifold**.*

Every *linear space* (Definition J.1 page 247) $\boldsymbol{X}$ has at least two *linear subspaces*—itself and $\mathbf{0}$ (Proposition L.1 page 290), called the *trivial linear space*. The *linear span* (Definition K.2 page 259) of every subset of a linear linear space is a subspace (Proposition L.2 page 291). Every *linear subspace* contains the "zero" vector $\mathbb{0}$, and is *convex* (Definition P.6 page 338, Proposition L.3 page 291).

Proposition L.1. [2] *Let* $\boldsymbol{X} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$ *and* $\mathbf{0} \triangleq (\{\mathbb{0}\}, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$.

PRP

$$\left\{ \begin{array}{l} \boldsymbol{X} \text{ is a LINEAR SPACE} \\ \text{(Definition J.1 page 247)} \end{array} \right\} \implies \left\{ \begin{array}{lll} 1. & \mathbf{0} & \text{is a LINEAR SUBSPACE of } \boldsymbol{X} \quad \text{and} \\ 2. & \boldsymbol{X} & \text{is a LINEAR SUBSPACE of } \boldsymbol{X} \end{array} \right\}$$

✎PROOF: For a structure to be a linear subspace of $\boldsymbol{X}$, it must satisfy the requirements of Definition L.1 (page 290).

1. Proof that $\{\mathbb{0}\}$ is a linear subspace:

   (a) Note that $\{\mathbb{0}\} \neq \varnothing$.

   (b) Proof that $\boldsymbol{x}, \boldsymbol{y} \in \{\mathbb{0}\} \implies \boldsymbol{x} + \boldsymbol{y} \in \{\mathbb{0}\}$:

   $$\begin{aligned} \boldsymbol{x} + \boldsymbol{y} &= \mathbb{0} + \mathbb{0} && \text{by } \boldsymbol{x}, \boldsymbol{y} \in \{\mathbb{0}\} \text{ hypothesis} \\ &= \mathbb{0} \\ &\in \{\mathbb{0}\} \end{aligned}$$

   (c) Proof that $\boldsymbol{x} \in \{\mathbb{0}\}, \alpha \in \mathbb{F} \implies \alpha\boldsymbol{x} \in \{\mathbb{0}\}$:

   $$\begin{aligned} \alpha\boldsymbol{x} &= \alpha\mathbb{0} && \text{by } \boldsymbol{x} \in \{\mathbb{0}\} \text{ hypothesis} \\ &= \mathbb{0} && \text{by definition of } \mathbb{0} \\ &\in \{\mathbb{0}\} \end{aligned}$$

2. Proof that $\boldsymbol{\Omega}$ is a linear subspace of itself:

   (a) Proof that $X \neq \varnothing$:

   $$X \neq \varnothing$$

   (b) Proof that $\boldsymbol{x}, \boldsymbol{y} \in X \implies \boldsymbol{x} + \boldsymbol{y} \in X$:

   $\boldsymbol{x} + \boldsymbol{y} \in \{\mathbb{0}\}$ because $+ : X \times X \to X$ (X is closed under vector addition)

   (c) Proof that $\boldsymbol{x} \in X, \alpha \in \mathbb{F} \implies \alpha\boldsymbol{x} \in X$:

   $\alpha\boldsymbol{x} \in X$ because $\cdot : \mathbb{F} \times X \to X$ (X is closed under scalar-vector multiplication)

[1] Michel and Herget (1993) page 81 ⟨Definition 3.2.1⟩, Berberian (1961) page 13 ⟨Definition I.5.1⟩, Halmos (1958) page 16

[2] Michel and Herget (1993) pages 81–83, Haaser and Sullivan (1991) page 43

✏

Proposition L.2. [3] *Let* $\boldsymbol{X} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$ *be a* LINEAR SPACE (*Definition J.1 page 247*). *Let* span *be the* LINEAR SPAN *of a set* Y *in* $\boldsymbol{X}$.

PRP
$$\left\{\begin{array}{l} Y \text{ is a SUBSET of the set } X \\ (Y \subseteq X) \end{array}\right\} \implies \left\{ \operatorname{span} Y \quad \text{is a LINEAR SUBSPACE of } \boldsymbol{X}. \right\}$$

Proposition L.3. [4] *Let* $\boldsymbol{X} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$ *be a* LINEAR SPACE *and* $\mathbb{0}$ *the zero vector of* $\boldsymbol{X}$.

PRP
$$\left\{ \boldsymbol{Y} \text{ is a LINEAR SUBSPACE of } \boldsymbol{X} \right\} \implies \left\{\begin{array}{lll} 1. & \mathbb{0} \in \boldsymbol{Y} & \text{and} \\ 2. & \boldsymbol{Y} \text{ is CONVEX in } \boldsymbol{X} & \end{array}\right\}$$

✎PROOF:

$$\begin{aligned}
\boldsymbol{Y} \text{ is a } \textit{subspace} &\implies \exists(\alpha \boldsymbol{y}) \in \boldsymbol{Y} \quad \forall \alpha \in \mathbb{F} && \text{by Definition L.1 page 290} \\
&\implies \exists \mathbb{0} \in \boldsymbol{Y} && \text{because } \alpha = 0 \in \mathbb{F}
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{Y} \text{ is a linear subspace} &\implies \boldsymbol{x} + \boldsymbol{y} \in Y \; \forall \boldsymbol{x}, \boldsymbol{y} \in Y \\
&\implies \lambda \boldsymbol{x} + (1 - \lambda)\boldsymbol{y} \in Y \; \forall \boldsymbol{x}, \boldsymbol{y} \in Y \\
&\implies \boldsymbol{Y} \text{ is } \textit{convex}
\end{aligned}$$

✏

Definition L.2. [5] *Let* $\boldsymbol{X} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$ *and* $\boldsymbol{Y} \triangleq (Y, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$ *be* LINEAR SUBSPACES (*Definition L.1 page 290*) *of a* LINEAR SPACE (*Definition J.1 page 247*) $\boldsymbol{\Omega} \triangleq (\Omega, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$.

DEF
$$\begin{array}{lcll}
\boldsymbol{X} \hat{+} \boldsymbol{Y} & \triangleq & (\{\boldsymbol{x} + \boldsymbol{y} | \boldsymbol{x} \in X \text{ and } \boldsymbol{y} \in Y\}, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times})) & \text{(Minkowski addition)} \\
\boldsymbol{X} \,\hat{\cup}\, \boldsymbol{Y} & \triangleq & (X \cup Y, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times})) & \text{(subspace union)} \\
\boldsymbol{X} \,\hat{\cap}\, \boldsymbol{Y} & \triangleq & (X \cap Y, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times})) & \text{(subspace intersection)}
\end{array}$$

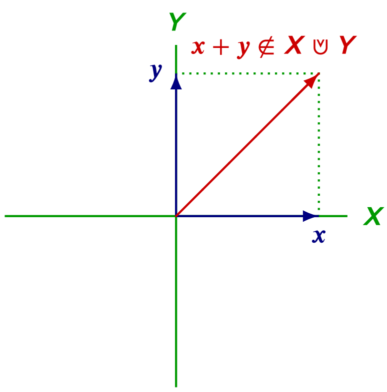
Example L.2. Some examples of operations on subspaces in $\mathbb{R}^3$ are illustrated next:

Remark L.1.

Notice the similarities between the properties of linear subspaces in a linear space (Proposition L.4 page 292) and the properties of closed sets in a topological space:

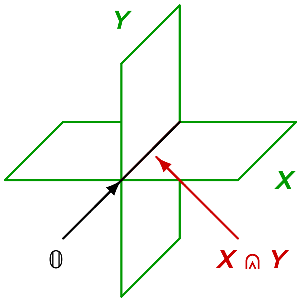
linear subspaces	*closed sets*
$\mathbb{0}$	$\varnothing$
$\boldsymbol{\Omega}$	$\boldsymbol{\Omega}$
$\boldsymbol{X} \hat{+} \boldsymbol{Y}$	$X \cup Y$
$\bigcap_{n=1}^{N} \boldsymbol{X}_n$	$\bigcap_{\gamma \in \Gamma} X_\gamma$

One key difference is that the union of two linear subspaces is not in general a linear subspace. For example, if $\boldsymbol{x}$ is the vector $[1\;0]$ in the x direction linear subspace of $\mathbb{R}^2$ and $\boldsymbol{y}$ is the vector $[0\;1]$ in the y direction linear subspace, then $\boldsymbol{x} + \boldsymbol{y}$ is not in the union of the two linear subspaces (it is not on the x axis or y axis but rather at $(1, 1)$).[6]



[3] Michel and Herget (1993) page 86
[4] Michel and Herget (1993) page 81
[5] Wedderburn (1907) page 79
[6] Michel and Herget (1993) page 82

In general, the set of all linear subspaces of a linear space $\boldsymbol{\Omega}$ is *not* closed under the subspace union ($\Cup$) operation; that is, the union of of two linear subspaces is *not* necessarily a linear subspace. However the set *is* closed under Minkowski sum ($\hat{+}$) and subspace intersection ($\Cap$). Proposition L.4 (next) shows four useful objects are always subspaces. Some of these in Euclidean space $\mathbb{R}^3$ are illustrated to the right.



Proposition L.4. [7] *Let $\boldsymbol{X}$ be a* LINEAR SPACE *(Definition J.1 page 247).*

PRP	
	$\left\{ \boldsymbol{X}_n \middle| {\scriptstyle n=1,2,\ldots,N} \right\}$ *are* LINEAR SUBSPACES *of* $\boldsymbol{X}$ $\Bigg\} \implies \Bigg\{$ 1. $\boldsymbol{X}_1 \hat{+} \boldsymbol{X}_2 \hat{+} \cdots \hat{+} \boldsymbol{X}_N$ *is a* LINEAR SUBSPACE *of* $\boldsymbol{X}$ *and* 2. $\boldsymbol{X}_1 \Cap \boldsymbol{X}_2 \Cap \cdots \Cap \boldsymbol{X}_N$ *is a* LINEAR SUBSPACE *of* $\boldsymbol{X}$

✎PROOF: For a structure to be a linear subspace of $\boldsymbol{X}$, it must satisfy the requirements of Definition L.1 (page 290).

1. Proof that $\boldsymbol{X}_1 \hat{+} \boldsymbol{X}_2 \hat{+} \cdots \hat{+} \boldsymbol{X}_N$ is a *linear subspace* (proof by induction):

   (a) proof for $N = 1$ case: by left hypothesis.

   (b) proof for $N = 2$ case:

   i. proof that $\boldsymbol{X}_1 \hat{+} \boldsymbol{X}_2 \neq \varnothing$:

$$
\begin{aligned}
\boldsymbol{X}_1 \hat{+} \boldsymbol{X}_2 &= \left\{ \boldsymbol{v} + \boldsymbol{w} | \boldsymbol{v} \in \boldsymbol{X}_1 \text{ and } \boldsymbol{w} \in Y \right\} && \text{by Definition L.2 page 291} \\
&\supseteq \left\{ \boldsymbol{v} + \boldsymbol{w} | \boldsymbol{v} \in \{\mathbb{0}\} \subseteq \boldsymbol{X}_1 \text{ and } \boldsymbol{w} \in \{\mathbb{0}\} \subseteq \boldsymbol{X}_2 \right\} \\
&= \{\mathbb{0} + \mathbb{0}\} \\
&= \{\mathbb{0}\} \\
&\neq \varnothing
\end{aligned}
$$

   ii. proof that $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{X}_1 \hat{+} \boldsymbol{X}_2 \implies \boldsymbol{x} + \boldsymbol{y} \in \boldsymbol{X}_1 \hat{+} \boldsymbol{X}_2$:

$$
\begin{aligned}
\boldsymbol{x} + \boldsymbol{y} &= (\boldsymbol{v}_1 + \boldsymbol{w}_1) + (\boldsymbol{v}_2 + \boldsymbol{w}_2) && \text{by } \boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{X}_1 \hat{+} \boldsymbol{X}_2 \text{ hypothesis} \\
&= \underbrace{(\boldsymbol{v}_1 + \boldsymbol{v}_2)}_{\text{in } \boldsymbol{X}_1} + \underbrace{(\boldsymbol{w}_1 + \boldsymbol{w}_2)}_{\text{in } \boldsymbol{X}_2 \text{ because } \boldsymbol{X}_2 \text{ is a linear subspace}} && \text{by Definition J.1 page 247} \\
&\in \left\{ \boldsymbol{v} + \boldsymbol{w} | \boldsymbol{v} \in \boldsymbol{X}_1 \text{ and } \boldsymbol{w} \in Y \right\} \\
&= \boldsymbol{X}_1 \hat{+} \boldsymbol{X}_2 && \text{by Definition L.2 page 291}
\end{aligned}
$$

   iii. proof that $\boldsymbol{v} \in \boldsymbol{X}_1 \hat{+} \boldsymbol{X}_2, \alpha \in F \implies \alpha\boldsymbol{v} \in \boldsymbol{X}_1 \hat{+} \boldsymbol{X}_2$:

$$
\begin{aligned}
\alpha\boldsymbol{x} &= \alpha(\boldsymbol{v}_1 + \boldsymbol{w}_1) && \text{by } \boldsymbol{x} \in \boldsymbol{X}_1 \hat{+} \boldsymbol{X}_2 \text{ hypothesis} \\
&= \underbrace{\alpha\boldsymbol{v}_1}_{\text{in } \boldsymbol{X}_1} + \underbrace{\alpha\boldsymbol{w}_1}_{\text{in } \boldsymbol{X}_2 \text{ because } \boldsymbol{X}_2 \text{ is a linear subspace}} && \text{by Definition J.1 page 247} \\
&\in \left\{ \boldsymbol{v} + \boldsymbol{w} | \boldsymbol{v} \in \boldsymbol{X}_1 \text{ and } \boldsymbol{w} \in Y \right\} \\
&= \boldsymbol{X}_1 \hat{+} \boldsymbol{X}_2 && \text{by Definition L.2 page 291}
\end{aligned}
$$

   (c) Proof that [N case] $\implies$ [$N+1$ case]:

$$
\begin{aligned}
\boldsymbol{X}_1 \hat{+} \boldsymbol{X}_2 \hat{+} \cdots \hat{+} \boldsymbol{X}_{N+1} &= \underbrace{\left(\boldsymbol{X}_1 \hat{+} \boldsymbol{X}_2 \hat{+} \cdots \hat{+} \boldsymbol{X}_N\right)}_{\text{linear subspace by } N \text{ case hypothsis}} \hat{+} \boldsymbol{X}_{N+1} \\
&\implies \text{linear subspace by } N = 2 \text{ case (item (1b) page 292)}
\end{aligned}
$$

[7] Michel and Herget (1993) pages 81–83

2. Proof that $\boldsymbol{X}_1 \Cap \boldsymbol{X}_2 \Cap \cdots \Cap \boldsymbol{X}_N$ is a *linear subspace* (proof by induction):
   (a) proof for $N = 1$ case: $\boldsymbol{X}_1$ is a linear subspace by left hypothesis.
   (b) Proof for $N = 2$ case:
      i. proof that $\boldsymbol{X} \Cap \boldsymbol{Y} \neq \varnothing$:

$$\begin{aligned}\boldsymbol{X} \Cap \boldsymbol{Y} &= \left\{\boldsymbol{x} \in X | \boldsymbol{x} \in X \text{ and } \boldsymbol{w} \in Y\right\}\\ &\supseteq \left\{\boldsymbol{x} \in X | \boldsymbol{x} \in \{\mathbb{0}\} \subseteq \boldsymbol{X} \text{ and } \boldsymbol{x} \in \{\mathbb{0}\} \subseteq \boldsymbol{Y}\right\}\\ &= \{\mathbb{0} + \mathbb{0}\}\\ &= \{\mathbb{0}\}\\ &\neq \varnothing\end{aligned}$$

      ii. proof that $\boldsymbol{x}, \boldsymbol{y} \in X \Cap \boldsymbol{Y} \implies \boldsymbol{x} + \boldsymbol{y} \in X \Cap \boldsymbol{Y}$:

$$\begin{aligned}\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{X} \Cap \boldsymbol{Y} &\implies \boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{X} \text{ and } \boldsymbol{x}, \boldsymbol{y} \in Y\\ &\implies \boldsymbol{x} + \boldsymbol{y} \in \boldsymbol{X} \text{ and } \boldsymbol{x} + \boldsymbol{y} \in Y && \text{because } \boldsymbol{X} \text{ and } \boldsymbol{Y} \text{ are linear subspaces}\\ &\implies \boldsymbol{x} + \boldsymbol{y} \in \boldsymbol{X} \Cap \boldsymbol{Y}\end{aligned}$$

      iii. proof that $\boldsymbol{v} \in X \Cap \boldsymbol{Y}, \alpha \in F \implies \alpha\boldsymbol{v} \in X \Cap \boldsymbol{Y}$:

$$\begin{aligned}\boldsymbol{x} \in \boldsymbol{X} \Cap \boldsymbol{Y} &\implies \boldsymbol{x} \in \boldsymbol{X} \text{ and } \boldsymbol{x} \in Y\\ &\implies \alpha\boldsymbol{x} \in \boldsymbol{X} \text{ and } \alpha\boldsymbol{x} \in Y && \text{because } \boldsymbol{X} \text{ and } \boldsymbol{Y} \text{ are linear subspaces}\\ &\implies \alpha\boldsymbol{x} \in \boldsymbol{X} \Cap \boldsymbol{Y}\end{aligned}$$

   (c) Proof that $[N \text{ case}] \implies [N + 1 \text{ case}]$:

$$\begin{aligned}\boldsymbol{X}_1 \Cap \boldsymbol{X}_2 \Cap \cdots \Cap \boldsymbol{X}_{N+1} &= \underbrace{\left(\boldsymbol{X}_1 \Cap \boldsymbol{X}_2 \Cap \cdots \Cap \boldsymbol{X}_N\right)}_{\text{linear subspace by } N \text{ case hypothesis}} \Cap \boldsymbol{X}_{N+1}\\ &\implies \text{linear subspace by } N = 2 \text{ case} && \text{(item (2b) page 293)}\end{aligned}$$

✏

Every linear subspace contains the zero vector $\mathbb{0}$ (Proposition L.3 page 291). But if a pair of linear subspaces of a linear space $\boldsymbol{X}$ *only* have $\mathbb{0}$ in commmon, then any vector in $\boldsymbol{X}$ can be *uniquely* represented by a single vector from each of the two subspaces (next).

Theorem L.1. [8] *Let* $\boldsymbol{X} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$ *and* $\boldsymbol{Y} \triangleq (Y, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$ *be* LINEAR SUBSPACES *(Definition L.1 page 290) of a* LINEAR SPACE *(Definition J.1 page 247)* $\boldsymbol{\Omega} \triangleq (\Omega, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$.

THM	
	$X \cap Y = \{\mathbb{0}\} \iff \left\{\begin{array}{l}\textit{for every } \boldsymbol{u} \in \boldsymbol{X} \hat{+} \boldsymbol{Y} \textit{ there exist } \boldsymbol{x} \in X \textit{ and } \boldsymbol{y} \in Y \textit{ such that}\\ \quad 1.\ \ \boldsymbol{u} = \boldsymbol{x} + \boldsymbol{y} \qquad\qquad \textit{and}\\ \quad 2.\ \ \boldsymbol{x} \textit{ and } \boldsymbol{y} \textit{ are } \text{UNIQUE}.\end{array}\right\}$

✎PROOF:

1. Proof that $X \cap Y = \{\mathbb{0}\} \implies$ *unique* $\boldsymbol{x}, \boldsymbol{y}$:
   Suppose that $\boldsymbol{x}$ and $\boldsymbol{y}$ are not unique, but rather $\boldsymbol{u} = \boldsymbol{x}_1 + \boldsymbol{y}_1 = \boldsymbol{x}_2 + \boldsymbol{y}_2$ where $\boldsymbol{x}_1, \boldsymbol{x}_2 \in X$ and $\boldsymbol{y}_1, \boldsymbol{y}_2 \in Y$.

$$\begin{aligned}\boldsymbol{u} = \boldsymbol{x}_1 + \boldsymbol{y}_1 = \boldsymbol{x}_2 + \boldsymbol{y}_2 &\implies \underbrace{\boldsymbol{x}_1 - \boldsymbol{x}_2}_{\in X} = \underbrace{\boldsymbol{y}_2 - \boldsymbol{y}_1}_{\in Y}\\ &\implies \boldsymbol{x}_1 - \boldsymbol{x}_2,\ \boldsymbol{y}_2 - \boldsymbol{y}_1 \in X \cap Y\\ &\implies \boldsymbol{x}_1 - \boldsymbol{x}_2 = \boldsymbol{y}_2 - \boldsymbol{y}_1 = \mathbb{0} && \text{by left hypothesis}\\ &\implies \boldsymbol{x}_1 = \boldsymbol{x}_2 \quad \text{and} \quad \boldsymbol{y}_2 = \boldsymbol{y}_1\\ &\implies \boldsymbol{x} \text{ and } \boldsymbol{y} \text{ are } \textit{unique}\end{aligned}$$

[8] Michel and Herget (1993) page 83 ⟨Theorem 3.2.12⟩, Kubrusly (2001) page 67 ⟨Theorem 2.14⟩

2. Proof that $X \cap Y = \{\mathbb{0}\} \impliedby$ *unique* $\boldsymbol{x}, \boldsymbol{y}$:

$$\begin{aligned}
\boldsymbol{u} &= \boldsymbol{x} + \boldsymbol{y} \\
&= \boldsymbol{x} + \boldsymbol{y} + \boldsymbol{y} - \boldsymbol{y} && \text{for some vector } \boldsymbol{y} \in X \cap Y \\
&= \underbrace{(\boldsymbol{x} + \boldsymbol{y})}_{\in X} + \underbrace{(\boldsymbol{y} - \boldsymbol{y})}_{\in Y} && \text{because } \boldsymbol{x} \in X \text{ and } \boldsymbol{y} \in X \cap Y \ldots \\
&\implies \boldsymbol{x} \text{ and } \boldsymbol{y} \text{ are } \textit{not unique} \text{ if } \boldsymbol{y} \neq \mathbb{0} \\
&\implies \boldsymbol{y} = \mathbb{0} && \text{by right hypothesis} \\
&\implies X \cap Y = \{\mathbb{0}\}
\end{aligned}$$

⇨

Theorem L.2. [9] *Let* $\boldsymbol{\Omega}$ *be a linear subspace and* $2^{\boldsymbol{\Omega}}$ *the set of closed linear subspaces of* $\boldsymbol{\Omega}$.

THM $\left(2^{\boldsymbol{\Omega}}, \hat{+}, \Cap, \mathbf{0}, \boldsymbol{\Omega}; \subseteq\right)$ *is a* LATTICE. *In particular*

$$\begin{array}{rclrcll}
\boldsymbol{X} \hat{+} \boldsymbol{X} &=& \boldsymbol{X} & \boldsymbol{X} \Cap \boldsymbol{X} &=& \boldsymbol{X} & \forall \boldsymbol{X} \in 2^{\boldsymbol{\Omega}} \\
\boldsymbol{X} \hat{+} \boldsymbol{Y} &=& \boldsymbol{Y} \hat{+} \boldsymbol{X} & \boldsymbol{X} \Cap \boldsymbol{Y} &=& \boldsymbol{Y} \Cap \boldsymbol{X} & \forall \boldsymbol{X}, \boldsymbol{Y} \in 2^{\boldsymbol{\Omega}} \\
(\boldsymbol{X} \hat{+} \boldsymbol{Y}) \hat{+} \boldsymbol{Z} &=& \boldsymbol{X} \hat{+} (\boldsymbol{Y} \hat{+} \boldsymbol{Z}) & (\boldsymbol{X} \Cap \boldsymbol{Y}) \Cap \boldsymbol{Z} &=& \boldsymbol{X} \Cap (\boldsymbol{Y} \Cap \boldsymbol{Z}) & \forall \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z} \in 2^{\boldsymbol{\Omega}} \\
\boldsymbol{X} \hat{+} (\boldsymbol{X} \Cap \boldsymbol{Y}) &=& \boldsymbol{X} & \boldsymbol{X} \Cap (\boldsymbol{X} \hat{+} \boldsymbol{Y}) &=& \boldsymbol{X} & \forall \boldsymbol{X}, \boldsymbol{Y} \in 2^{\boldsymbol{\Omega}}
\end{array}$$

✎PROOF: These results follow directly from the properties of lattices. ⇨

L.2 Subspaces of an inner product space

Definition L.3. [10] *Let* $\boldsymbol{\Omega} \triangleq \left(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \mid \triangledown \rangle\right)$ *be an* INNER PRODUCT SPACE *(Definition N.1 page 309)*.

DEF *The* ***orthogonal complement*** $A^{\perp}$ *in* $\boldsymbol{\Omega}$ *of a set* $A \subseteq X$ *is*

$$A^{\perp} \triangleq \{\boldsymbol{x} \in X \mid \langle \boldsymbol{x} \mid \boldsymbol{y} \rangle = 0 \quad \forall \boldsymbol{y} \in A\}.$$

The expression $A^{\perp\perp}$ *is defined as* $\left(A^{\perp}\right)^{\perp}$.

Proposition L.5. [11] *Let* $\left(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \mid \triangledown \rangle\right)$ *be an* INNER PRODUCT SPACE *(Definition N.1 page 309)*.

PRP
$$A \subseteq B \implies B^{\perp} \subseteq A^{\perp} \quad \forall A, B \in 2^X \quad \text{(ANTITONE)}$$

✎PROOF:

$$\begin{aligned}
B^{\perp} &\triangleq \{\boldsymbol{x} \in X \mid \langle \boldsymbol{x} \mid \boldsymbol{y} \rangle = 0 \quad \forall \boldsymbol{y} \in B\} && \text{by definition of } B^{\perp} \text{ (Definition L.3 page 294)} \\
&\subseteq \{\boldsymbol{x} \in X \mid \langle \boldsymbol{x} \mid \boldsymbol{y} \rangle = 0 \quad \forall \boldsymbol{y} \in A\} && \text{by } A \subseteq B \text{ hypothesis} \\
&= A^{\perp} && \text{by definition of } A^{\perp} \text{ (Definition L.3 page 294)}
\end{aligned}$$

⇨

Every *linear space* $\boldsymbol{X}$ contains $\mathbf{0}$ and $\boldsymbol{X}$ as *linear subspaces* (Proposition L.1 page 290). If $\boldsymbol{X}$ is also an *inner product space*, then $\mathbf{0}$ and $\boldsymbol{X}$ are *orthogonal complements* of each other (next proposition).

[9] Iturrioz (1985) pages 56–57

[10] Berberian (1961) page 59 ⟨Definition III.2.1⟩, Michel and Herget (1993) page 382, Kubrusly (2001) page 328

[11] Berberian (1961) page 60 ⟨Theorem III.2.2⟩, Kubrusly (2011) page 326

Proposition L.6. [12] *Let* $\boldsymbol{\Omega} \triangleq \left(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle\triangle \mid \triangledown\rangle\right)$ *be an* INNER PRODUCT SPACE (Definition N.1 page 309) *and* $\mathbb{0}$ *the* VECTOR ADDITIVE IDENTITY ELEMENT (Definition J.1 page 247) *in* $\boldsymbol{\Omega}$.

PRP

1. $\{\mathbb{0}\}^\perp = X$
2. $X^\perp = \{\mathbb{0}\}$

✎PROOF:

$$\begin{aligned}
\{\mathbb{0}\}^\perp &= \{\boldsymbol{x} \in X \mid \langle \boldsymbol{x} \mid \boldsymbol{y}\rangle = 0 \quad \forall \boldsymbol{y} \in \{\mathbb{0}\}\} && \text{by definition of } \perp \text{ (Definition L.3 page 294)}\\
&= \{\boldsymbol{x} \in X \mid \langle \boldsymbol{x} \mid \mathbb{0}\rangle = 0\}\\
&= X\\
X^\perp &= \{\boldsymbol{x} \in X \mid \langle \boldsymbol{x} \mid \boldsymbol{y}\rangle = 0 \quad \forall \boldsymbol{y} \in X\} && \text{by definition of } \perp \text{ Definition L.3 page 294}\\
&= \{\boldsymbol{x} \in X \mid \langle \boldsymbol{x} \mid \boldsymbol{x}\rangle = 0\}\\
&= \{\mathbb{0}\}
\end{aligned}$$

✏

For any set A contained in a linear space $\boldsymbol{X}$, $A^{\perp\perp}$ is a *linear subspace*, and it is the smallest linear subspace containing the set A ($A^{\perp\perp} = \operatorname{span} A$, next theorem). In the case that A is a *linear subspace* rather than just a subset, results simplify significantly (next corollary).

Theorem L.3. [13] *Let* $\boldsymbol{\Omega} \triangleq \left(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle\triangle \mid \triangledown\rangle\right)$ *be an* INNER PRODUCT SPACE (Definition N.1 page 309). *Let* $\operatorname{span} A$ *be the span of a set* A (Definition K.2 page 259).

THM

$$\left\{\begin{array}{l} A \text{ is a **subset** of } X \\ (A \subseteq X) \end{array}\right\} \implies \left\{\begin{array}{llcll}
1. & A \cap A^\perp & = & \begin{cases}\{\mathbb{0}\} & \text{if } \mathbb{0} \in A\\ \varnothing & \text{if } \mathbb{0} \notin A\end{cases} & \text{and}\\
2. & A & \subseteq & A^{\perp\perp} = \operatorname{span} A & \text{and}\\
3. & A^\perp & = & A^{\perp\perp\perp} = A^{\perp-} = A^{-\perp} = (\operatorname{span} A)^\perp & \text{and}\\
4. & \multicolumn{3}{l}{A^\perp \text{ is a subspace of } \boldsymbol{\Omega}}
\end{array}\right\}$$

✎PROOF:

1. Proof that $A \cap A^\perp = \cdots$:

$$\begin{aligned}
A \cap A^\perp &= \{\boldsymbol{x} \in X \mid \boldsymbol{x} \in A\} \cap \{\boldsymbol{x} \in X \mid \langle \boldsymbol{x} \mid \boldsymbol{y}\rangle \quad \forall \boldsymbol{y} \in A\} && \text{by definition of } A^\perp\\
&= \left\{\boldsymbol{x} \in X \mid \boldsymbol{x} \in A \quad \text{and} \quad \langle \boldsymbol{x} \mid \boldsymbol{y}\rangle \quad \forall \boldsymbol{y} \in A\right\}\\
&= \begin{cases}\{\mathbb{0}\} & \text{if } \mathbb{0} \in A\\ \varnothing & \text{if } \mathbb{0} \notin A\end{cases}
\end{aligned}$$

2. Proof that $A \subseteq A^{\perp\perp} = \operatorname{span} A$:

$$\begin{aligned}
\boldsymbol{x} \in A &\implies \{\boldsymbol{x}\}^{\perp\perp} \subseteq A^{\perp\perp}\\
&\implies \boldsymbol{x} \in \{\boldsymbol{x}\}^{\perp\perp} \subseteq A^{\perp\perp}\\
&\implies \boldsymbol{x} \in A^{\perp\perp}
\end{aligned}$$

but

$$\boldsymbol{x} \in A^{\perp\perp} \not\Longrightarrow \boldsymbol{x} \in A$$

Here is an example for the $\not\Longrightarrow$ part using the linear space $\mathbb{R}^3$:

[12] Kubrusly (2011) page 326, Michel and Herget (1993) page 383
[13] Michel and Herget (1993) page 383, Kubrusly (2011) page 326

(a) Let $A \triangleq \{i\}$, where i is the unit vector on the x-axis.

(b) Then $A^{\perp} = \{\boldsymbol{x} \in X | \boldsymbol{x} \in \text{yz plane}\}$.

(c) Then $A^{\perp\perp} = \{\boldsymbol{x} \in X | \boldsymbol{x} \in \text{x axis}\}$.

(d) Therefore, $A \subsetneq A^{\perp\perp}$

3. Proof for $A^{\perp}$ equivalent expressions:

(a) Proof that $A^{\perp} = A^{\perp\perp\perp}$:

$$\begin{aligned}
A^{\perp} &\subseteq \left(A^{\perp}\right)^{\perp\perp} && \text{by item (2)}\\
&= \left(A^{\perp\perp}\right)^{\perp}\\
&= A^{\perp\perp\perp} && \text{by Definition L.3 page 294}\\
A^{\perp\perp\perp} &= \left(A^{\perp\perp}\right)^{\perp} && \text{by Definition L.3 page 294}\\
&\subseteq A^{\perp} && \text{by item (2) and Proposition L.5 (page 294)}
\end{aligned}$$

(b) Proof that $A^{\perp\perp\perp} = (\operatorname{span} A)^{\perp}$: follows directly from item (2) ($A^{\perp\perp} = \operatorname{span} A$).

(c) Proof that $A^{\perp} = A^{\perp -}$:

i. Let $(\!(\boldsymbol{x}_n)\!)$ be an $A^{\perp}$-valued sequence that converges to the limit $\boldsymbol{x}$ in X.

ii. The limit point $\boldsymbol{x}$ must be in $A^{\perp}$ because for all $\boldsymbol{y} \in A$

$$\begin{aligned}
\langle \boldsymbol{x} \,|\, \boldsymbol{y} \rangle &= \langle \lim \boldsymbol{x}_n \,|\, \boldsymbol{y} \rangle && \text{by definition of the sequence } (\!(\boldsymbol{x}_n)\!)\\
&= \lim \langle \boldsymbol{x}_n \,|\, \boldsymbol{y} \rangle\\
&= 0 && \text{because } (\!(\boldsymbol{x}_n)\!) \text{ is } A^{\perp}\text{-valued}
\end{aligned}$$

iii. Because $\langle \boldsymbol{x} \,|\, \boldsymbol{y} \rangle = 0 \quad \forall \boldsymbol{y} \in A$, $\boldsymbol{x}$ is in $A^{\perp}$.

iv. Because $A^{\perp}$ contains all its limit points, and by the *Closed Set Theorem* (Theorem ?? page ??), it must be *closed* ($A^{\perp} = A^{\perp -}$)

(d) Proof that $A^{\perp} = A^{-\perp}$:

i. Let $x \in A^{\perp}$ and $y \in A^{-}$.

ii. Let $(\!(\boldsymbol{y}_n)\!)$ be an $A^{\perp}$-valued sequence that converges in X to y.

iii. Thus $A^{\perp} \perp A^{-}$ because

$$\begin{aligned}
\langle \boldsymbol{y} \,|\, \boldsymbol{x} \rangle &= \langle \lim \boldsymbol{y}_n \,|\, \boldsymbol{x} \rangle && \text{by definition of } (\!(\boldsymbol{y}_n)\!)\\
&= \lim \langle \boldsymbol{y}_n \,|\, \boldsymbol{x} \rangle\\
&= 0 && \text{because } (\!(\boldsymbol{y}_n)\!) \text{ is } A^{\perp}\text{-valued}
\end{aligned}$$

iv. Because $A^{\perp} \perp A^{-}$, so $A^{\perp} \subseteq A^{\perp -}$.

v. But $A^{\perp -} \subseteq A^{\perp}$ because

$$A \subseteq A^{-} \implies A^{\perp -} \subseteq A^{\perp} \qquad \text{by } \textit{antitone} \text{ property (Proposition L.5 page 294)}$$

vi. And so $A^{\perp} = A^{\perp -}$.

4. Proof that $A^{\perp}$ is a **subspace** of $\boldsymbol{\Omega}$ (must satisfy the conditions of Definition L.1 page 290):

(a) Proof that $A^{\perp} \neq \varnothing$: $A^{\perp}$ has at least one element, the element $\mathbb{0}$...

$$\begin{aligned}
\langle \mathbb{0} \,|\, \boldsymbol{y} \rangle = 0 \quad \forall \boldsymbol{y} \in A & && \text{by definition of } \mathbb{0}\\
\implies \mathbb{0} \in A^{\perp} & && \text{by definition of } A^{\perp} \text{ (Definition L.3 page 294)}
\end{aligned}$$

(b) Proof that $A^\perp \subseteq X$:

$$\begin{aligned}
\boldsymbol{u} \in A^\perp &\implies \boldsymbol{u} \in \{\boldsymbol{x} \in X \mid \langle \boldsymbol{x} \mid \boldsymbol{y} \rangle = 0 \quad \forall \boldsymbol{y} \in A\} && \text{by definition of } A^\perp \text{ (Definition L.3 page 294)}\\
&\implies \boldsymbol{u} \in X && \text{by definition of sets}
\end{aligned}$$

(c) Proof that $\boldsymbol{u}, \boldsymbol{v} \in A^\perp \implies (\boldsymbol{u}+\boldsymbol{v}) \in A^\perp$:

$$\begin{aligned}
\boldsymbol{u}, \boldsymbol{v} \in A^\perp &\implies \langle \boldsymbol{u} \mid \boldsymbol{y} \rangle = \langle \boldsymbol{v} \mid \boldsymbol{y} \rangle = 0 \quad \forall \boldsymbol{y} \in A && \text{by definition of } A^\perp \text{ (Definition L.3 page 294)}\\
&\implies \langle \boldsymbol{u} \mid \boldsymbol{y} \rangle + \langle \boldsymbol{v} \mid \boldsymbol{y} \rangle = 0 \quad \forall \boldsymbol{y} \in A\\
&\implies \langle \boldsymbol{u}+\boldsymbol{v} \mid \boldsymbol{y} \rangle = 0 \quad \forall \boldsymbol{y} \in A && \text{by \textit{additive} property of } \langle \triangle \mid \triangledown \rangle \text{ (Definition N.1 page 309)}\\
&\implies \boldsymbol{u}+\boldsymbol{v} \in A^\perp && \text{by definition of } A^\perp \text{ (Definition L.3 page 294)}
\end{aligned}$$

(d) Proof that $\boldsymbol{v} \in \Omega \implies \alpha \boldsymbol{v} \in A^\perp$:

$$\begin{aligned}
\boldsymbol{v} \in A^\perp &\implies \langle \boldsymbol{v} \mid \boldsymbol{y} \rangle = 0 \quad \forall \boldsymbol{y} \in A && \text{by definition of } A^\perp \text{ (Definition L.3 page 294)}\\
&\implies \alpha \langle \boldsymbol{v} \mid \boldsymbol{y} \rangle = \alpha \cdot 0 \quad \forall \boldsymbol{y} \in A\\
&\implies \langle \alpha\boldsymbol{v} \mid \boldsymbol{y} \rangle = 0 \quad \forall \boldsymbol{y} \in A && \text{by \textit{homogeneous} property of } \langle \triangle \mid \triangledown \rangle \text{ (Definition N.1 page 309)}\\
&\implies \alpha\boldsymbol{v} \in A^\perp && \text{by definition of } A^\perp \text{ (Definition L.3 page 294)}
\end{aligned}$$

✏

Corollary L.1. *Let* $\boldsymbol{X} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \mid \triangledown \rangle)$ *and* $\boldsymbol{Y} \triangleq (Y, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \mid \triangledown \rangle)$ *be* INNER PRODUCT SPACES. *Let* $\operatorname{span} Y$ *be the span of the set* Y *(Definition K.2 page 259).*

COR

$$\left\{ \boldsymbol{Y} \text{ is a \textbf{linear subspace} of } \boldsymbol{X} \right\} \implies \left\{ \begin{array}{lrcll} 1. & Y \cap Y^\perp & = & \{\mathbb{0}\} & \text{and}\\ 2. & Y & = & Y^{\perp\perp} = \operatorname{span} Y & \text{and}\\ 3. & Y^\perp & = & Y^{\perp\perp\perp} & \text{and}\\ 4. & \multicolumn{4}{l}{Y^\perp \text{ is a \textbf{subspace} of } \boldsymbol{X}} \end{array} \right\}$$

✎PROOF:

1. Proof that $\boldsymbol{Y} \cap Y^\perp = \{\mathbb{0}\}$: This follows from Theorem L.3 (page 295) and the fact that all subspaces contain the zero vector $\mathbb{0}$ (Proposition L.3 page 291).
2. Proof that $\boldsymbol{Y} = \boldsymbol{Y}^{\perp\perp} = \operatorname{span} \boldsymbol{Y}$: This follows directly from Theorem L.3 (page 295).
3. Proof that $\boldsymbol{Y}^\perp = \boldsymbol{Y}^{\perp\perp\perp}$: This follows directly from Theorem L.3 (page 295).
4. Proof that $Y^\perp$ is a **subspace** of $\boldsymbol{X}$: This follows directly from Theorem L.3 (page 295).

✏

Theorem L.4. [14] *Let* $\boldsymbol{Y} \triangleq (Y, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \mid \triangledown \rangle)$ *and* $\boldsymbol{Z} \triangleq (Z, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \mid \triangledown \rangle)$ *be* LINEAR SUBSPACES *of an* INNER PRODUCT SPACE $\boldsymbol{\Omega} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \mid \triangledown \rangle)$.

THM

$$\boldsymbol{Y} \perp \boldsymbol{Z} \implies Y \cap Z = \{\mathbb{0}\}$$

✎PROOF:

$$\begin{aligned}
\boldsymbol{x} \in \boldsymbol{Y} \cap \boldsymbol{Z} &\implies \boldsymbol{x} \in \boldsymbol{Y} \text{ and } \boldsymbol{x} \in \boldsymbol{Z} && \text{by definition of } \cap\\
&\implies \langle \boldsymbol{x} \mid \boldsymbol{x} \rangle = 0 && \text{by hypothesis } \boldsymbol{Y} \perp \boldsymbol{Z}\\
&\implies \boldsymbol{x} = \mathbb{0} && \text{by \textit{non-isotropic} property of } \langle \triangle \mid \triangledown \rangle \text{ (Definition N.1 page 309)}
\end{aligned}$$

✏

[14] Kubrusly (2001) page 324

Theorem L.5. [15] *Let* $\boldsymbol{Y} \triangleq (Y, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \mid \triangledown \rangle)$ *and* $\boldsymbol{Z} \triangleq (Z, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \mid \triangledown \rangle)$ *be linear subspaces of an* INNER PRODUCT SPACE $\boldsymbol{\Omega} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \mid \triangledown \rangle)$.

THM
$$\left\{\begin{array}{ll} 1. & \boldsymbol{Y} \perp \boldsymbol{Z} \text{ and} \\ 2. & \boldsymbol{x} \in \boldsymbol{Y} \hat{+} \boldsymbol{Z} \end{array}\right\} \implies \left\{\begin{array}{ll} 1. & \text{There exists } \boldsymbol{y} \in \boldsymbol{Y} \text{ and } \boldsymbol{z} \in \boldsymbol{Z} \text{ such that } \boldsymbol{x} = \boldsymbol{y} + \boldsymbol{z} \text{ and} \\ 2. & \boldsymbol{y} \text{ and } \boldsymbol{z} \text{ are UNIQUE.} \end{array}\right\}$$

✎PROOF:

1. Proof that $\boldsymbol{y}$ and $\boldsymbol{z}$ exist: by definition of Minkowski addition operator $\hat{+}$ (Definition L.2 page 291).
2. Proof that $\boldsymbol{y}$ and $\boldsymbol{z}$ are *unique*:
   (a) Suppose $\boldsymbol{x} = \boldsymbol{y}_1 + \boldsymbol{z}_1 = \boldsymbol{y}_1 + \boldsymbol{z}_2$ for $\boldsymbol{y}_1, \boldsymbol{y}_2 \in \boldsymbol{Y}$ and $\boldsymbol{z}_1, \boldsymbol{z}_2 \in \boldsymbol{Z}$.
   (b) This implies
$$\begin{aligned} \mathbb{0} &= \boldsymbol{x} - \boldsymbol{x} \\ &= (\boldsymbol{y}_1 + \boldsymbol{z}_1) - (\boldsymbol{y}_1 + \boldsymbol{z}_2) \\ &= \underbrace{(\boldsymbol{y}_1 - \boldsymbol{y}_2)}_{\text{in } \boldsymbol{Y}} + \underbrace{(\boldsymbol{z}_1 - \boldsymbol{z}_2)}_{\text{in } \boldsymbol{Z}} \end{aligned}$$
   (c) Because $\boldsymbol{y}_1 - \boldsymbol{y}_2 \in \boldsymbol{Y}$, $\boldsymbol{z}_1 - \boldsymbol{z}_2 \in \boldsymbol{Z}$, $(\boldsymbol{y}_1 - \boldsymbol{y}_2) + (\boldsymbol{z}_1 + \boldsymbol{z}_2) = \mathbb{0}$, and $\langle \boldsymbol{y}_1 - \boldsymbol{y}_2 \mid \boldsymbol{z}_1 - \boldsymbol{z}_2 \rangle = 0$, then by Theorem N.9 (page 323), $\boldsymbol{y}_1 - \boldsymbol{y}_2 = \mathbb{0}$ and $\boldsymbol{z}_1 - \boldsymbol{z}_2 = \mathbb{0}$.
   (d) This implies $\boldsymbol{y}_1 = \boldsymbol{y}_2$ and $\boldsymbol{z}_1 = \boldsymbol{z}_2$.
   (e) This implies $\boldsymbol{y}$ and $\boldsymbol{z}$ are *unique*.

✏

L.3 Subspaces of a Hilbert Space

Theorem L.6. [16] *Let* $\boldsymbol{H} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \mid \triangledown \rangle)$ *be a* HILBERT SPACE *(Definition ?? page ??). Let* Y *be a* SUBSET *of* X*, and let* $\mathsf{d}(\boldsymbol{x}, Y) \triangleq \inf_{\boldsymbol{y} \in Y} \|\boldsymbol{x} - \boldsymbol{y}\|$.

THM
$$\left\{\begin{array}{lll} 1. & Y \neq \varnothing & \text{and} \\ 2. & Y \text{ is CLOSED} & \text{and} \\ 3. & Y \text{ is CONVEX} \text{ (Definition P.6 page 338)} & \end{array}\right\} \implies \left\{\begin{array}{lll} \multicolumn{3}{l}{\text{There exists } \boldsymbol{p} \in Y \text{ such that}} \\ 1. & \mathsf{d}(\boldsymbol{x}, Y) = \|\boldsymbol{x} - \boldsymbol{p}\| & \text{and} \\ 2. & \boldsymbol{p} \text{ is UNIQUE.} & \end{array}\right\}$$

✎PROOF:

1. Let $\delta \triangleq \inf\{\boldsymbol{x} - \boldsymbol{y} | \boldsymbol{y} \in Y\}$.
2. Let $(\boldsymbol{y}_n)_{n \in \mathbb{Z}}$ be a sequence such that $\|\boldsymbol{x} - \boldsymbol{y}_n\| \to \delta$.

[15] Berberian (1961) page 61 ⟨Theorem III.2.3⟩
[16] Kubrusly (2001) page 330 ⟨Theorem 5.13⟩, Aliprantis and Burkinshaw (1998) page 290 ⟨Theorem 33.6⟩, Berberian (1961) page 68 ⟨Theorem III.5.1⟩

3. Proof that $(\!(\boldsymbol{y}_n)\!)$ is *Cauchy*:

$$\begin{aligned}
&\lim_{m,n\to\infty} \|\boldsymbol{y}_n - \boldsymbol{y}_m\|^2 \\
&= \lim_{m,n\to\infty} \|(\boldsymbol{y}_n - x) + (x - \boldsymbol{y}_m)\|^2 \\
&= \lim_{m,n\to\infty} \left\{ -\|(\boldsymbol{y}_n - x) - (x - \boldsymbol{y}_m)\|^2 + 2\|\boldsymbol{y}_n - x\|^2 + 2\|x - \boldsymbol{y}_m\|^2 \right\} && \text{by } \textit{parallelogram law} \text{ (page 317)} \\
&= \lim_{m,n\to\infty} \left\{ -4 \left\| \underbrace{\left(\frac{1}{2}\boldsymbol{y}_n + \frac{1}{2}\boldsymbol{y}_m\right)}_{\text{in } \boldsymbol{Y} \text{ by } \textit{convexity}} - x \right\|^2 + 2\|\boldsymbol{y}_n - x\|^2 + 2\|x - \boldsymbol{y}_m\|^2 \right\} \\
&\le \lim_{m,n\to\infty} \left\{ -4\delta^2 + 2\|\boldsymbol{y}_n - x\|^2 + 2\|x - \boldsymbol{y}_m\|^2 \right\} && \text{by definition of } \delta \text{ (item (1))} \\
&= -4\delta^2 + \lim_{m,n\to\infty} \left\{ 2\|\boldsymbol{y}_n - x\|^2 \right\} + \lim_{m,n\to\infty} \left\{ 2\|x - \boldsymbol{y}_m\|^2 \right\} \\
&= -4\delta^2 + 2\delta^2 + 2\delta^2 && \text{by definition of } \delta \text{ (item (1))} \\
&= 0
\end{aligned}$$

4. Proof that $\mathsf{d}(\boldsymbol{x}, Y) = \|\boldsymbol{x} - \boldsymbol{y}\|$: because $(\!(\boldsymbol{y}_n)\!)$ is *Cauchy* (item (1)) and by the *closed* hypothesis.

5. Proof that $\boldsymbol{y}$ is *unique*: Because in a metric space, the limit of a convergent sequence is *unique*.

✏

Theorem L.7. [17] *Let* $\boldsymbol{H} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \,|\, \triangledown \rangle)$ *be a* HILBERT SPACE *(Definition ?? page ??). Let* $\mathsf{d}(\boldsymbol{x}, Y) \triangleq \inf_{\boldsymbol{y}\in Y} \|\boldsymbol{x} - \boldsymbol{y}\|$. *Let* $\boldsymbol{Y} \triangleq (Y, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \,|\, \triangledown \rangle)$ *and* $Y^\perp$ *the* ORTHOGONAL COMPLEMENT *of* Y.

THM
$$\left\{ \boldsymbol{Y} \text{ is a SUBSPACE of } \boldsymbol{H} \right\} \implies \left\{ \begin{array}{lll} \multicolumn{3}{l}{\text{There exists } \boldsymbol{p} \in Y \text{ such that}} \\ 1. & \mathsf{d}(\boldsymbol{x}, Y) = \|\boldsymbol{x} - \boldsymbol{p}\| & \text{and} \\ 2. & \boldsymbol{p} \text{ is UNIQUE} & \text{and} \\ 3. & \boldsymbol{x} - \boldsymbol{p} \in Y^\perp. & \end{array} \right\}$$

Theorem L.8 (Projection Theorem). [18] *Let* $\boldsymbol{H} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \,|\, \triangledown \rangle)$ *be a Hilbert space.*

THM
$$\left\{ \boldsymbol{Y} \text{ is a SUBSPACE of } \boldsymbol{H} \right\} \implies \left\{ \boldsymbol{Y} \hat{+} \boldsymbol{Y}^\perp = \boldsymbol{H} \right\}$$

✎PROOF:

$$\begin{aligned}
\boldsymbol{Y} \hat{+} \boldsymbol{Y}^\perp &= \left[\boldsymbol{Y} \hat{+} \boldsymbol{Y}^\perp\right]^{\perp\perp} && \text{by Corollary L.1 page 297} \\
&= \left[\boldsymbol{Y}^\perp \Cap \boldsymbol{Y}^{\perp\perp}\right]^\perp && \text{by Proposition L.5 (page 294)} \\
&= \{\mathbb{0}\}^\perp && \text{by Corollary L.1 page 297} \\
&= \boldsymbol{H} && \text{by Proposition L.6 page 295}
\end{aligned}$$

✏

The inclusion relation $\subseteq$ is an order relation on the set of subspaces of a linear space $\boldsymbol{\Omega}$.

[17] Kubrusly (2001) page 330 ⟨Theorem 5.13⟩
[18] Bachman and Narici (1966) page 172 ⟨Theorem 10.8⟩, Kubrusly (2001) page 339 ⟨Theorem 5.20⟩

Proposition L.7. *Let S be the set of subspaces of a linear space Ω. Let $\subseteq$ be the inclusion relation.*

PRP $(S, \subseteq)$ *is an* ***ordered set***

✎PROOF: $(S, \subseteq)$ is an *ordered set* and because

1. $X \subseteq X$ $\quad \forall X \in S$ (reflexive) and
2. $X \subseteq Y$ and $Y \subseteq Z \implies X \subseteq Z$ $\quad \forall X, Y, Z \in S$ (transitive) and } preorder
3. $X \subseteq Y$ and $Y \subseteq X \implies X = Y$ $\quad \forall X, Y \in S$ (anti-symmetric)

✏

Theorem L.9. [19] *Let H be a Hilbert space and 2^{H} the set of closed linear subspaces of H.*

THM $\left(2^{H}, \hat{+}, \hat{\cap}, \mathbf{0}, H; \subseteq\right)$ *is an* ORTHOMODULAR LATTICE. *In particular*

1.	$X \hat{+} X^{\perp}$	$=$	H	$\forall X \in H$	(COMPLEMENTED)
2.	$X \hat{\cap} X^{\perp}$	$=$	$\mathbf{0}$	$\forall X \in H$	(COMPLEMENTED)
3.	$\left(X^{\perp}\right)^{\perp}$	$=$	X	$\forall X \in H$	(INVOLUTORY)
4.	$X \leq Y$	$\implies$	$Y^{\perp} \leq X^{\perp}$	$\forall X, Y \in H$	(ANTITONE)
5.	$X \leq Y$	$\implies$	$X \hat{+} \left(X^{\perp} \hat{\cap} Y\right) = Y$	$\forall X, Y \in \mathcal{X}$	(ORTHOMODULAR IDENTITY)

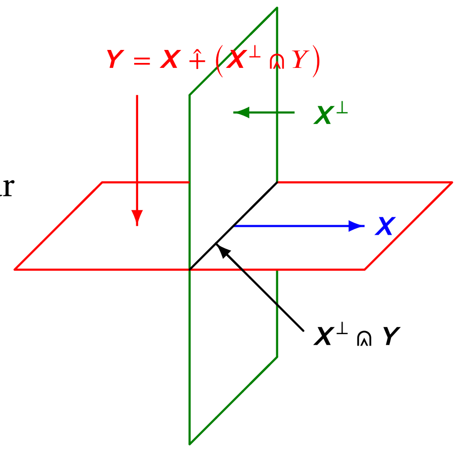
✎PROOF:

1. Proof for *complemented* (1) property: by *Projection Theorem* (Theorem L.8 page 299).
2. Proof for *complemented* (2) property: by Corollary L.1 (page 297).
3. Proof for *involutory* property: by Corollary L.1 (page 297).
4. Proof for *antitone* property: by Proposition L.5 (page 294).
5. Proof for *orthomodular identity* property:
6. Proof that lattice is *orthomodular*: by 5 properties and definition of *orthomodular lattice*.

✏

This concept is illustrated to the right where $X, Y \in 2^{H}$ are linear subspaces of the linear space H and

$$X \subseteq Y \quad \implies \quad Y = X \hat{+} \left(X^{\perp} \hat{\cap} Y\right).$$



Corollary L.2. *Let H be a Hilbert space with orthogonality operation $\perp$. Let $\left(2^{H}, \hat{+}, \hat{\cap}, \mathbf{0}, H; \subseteq\right)$ be the lattice of subspaces of H.*

COR

$$(X \hat{+} Y)^{\perp} = X^{\perp} \hat{\cap} Y^{\perp} \quad \forall X, Y \in 2^{H} \quad \text{(DE MORGAN)} \quad \textit{and}$$
$$(X \hat{\cap} Y)^{\perp} = X^{\perp} \hat{+} Y^{\perp} \quad \forall X, Y \in 2^{H} \quad \text{(DE MORGAN)}$$

✎PROOF: By properties of *orthocomplemented lattices* . ✏

[19] Iturrioz (1985) pages 56–57

L.4 Subspace Metrics

Definition L.4 (Hilbert space gap metric). [20] *Let $\boldsymbol{X}$ be a **Hilbert space** and $\boldsymbol{S}$ the set of subspaces of $\boldsymbol{X}$. Then we define the following metric between subspaces of $\boldsymbol{X}$.*

DEF

$$\begin{aligned} \mathsf{d}(\boldsymbol{V},\boldsymbol{W}) &\triangleq \|\mathbf{P}-\mathbf{Q}\| \quad \forall \boldsymbol{V},\boldsymbol{W}\in\boldsymbol{S} && \text{(the distance between subspaces } \boldsymbol{V} \text{ and } \boldsymbol{W} \text{ is the size of the difference of their projection operators)} \\ \text{where } \boldsymbol{V} &\triangleq \mathbf{P}\boldsymbol{X} && (\mathbf{P} \text{ is the projection operator that generates the subspace } \boldsymbol{V}) \\ \text{and } \boldsymbol{W} &\triangleq \mathbf{Q}\boldsymbol{X} && (\mathbf{Q} \text{ is the projection operator that generates the subspace } \boldsymbol{W}). \end{aligned}$$

Definition L.5 (Banach space gap metric). [21] *Let $\boldsymbol{X}$ be a **Banach space** and $\boldsymbol{S}$ the set of subspaces of $\boldsymbol{X}$. Then we define the following metric between subspaces of $\boldsymbol{X}$.*

DEF

$$\begin{aligned} \mathsf{d}(\boldsymbol{V},\boldsymbol{W}) &\triangleq \max\left\{ \sup_{\boldsymbol{v}\in\boldsymbol{V},\|\boldsymbol{v}\|=1} \mathsf{p}(\boldsymbol{v},\boldsymbol{W}),\; \sup_{\boldsymbol{w}\in\boldsymbol{W},\|\boldsymbol{w}\|=1} \mathsf{p}(\boldsymbol{w},\boldsymbol{V}) \right\} \quad \forall \boldsymbol{V},\boldsymbol{W}\in\boldsymbol{S} \\ \text{where } \mathsf{p}(\boldsymbol{v},\boldsymbol{W}) &\triangleq \inf_{\boldsymbol{w}\in\boldsymbol{W}} \|\boldsymbol{v}-\boldsymbol{w}\| \qquad \text{(metric from the point } \boldsymbol{v} \text{ to the subspace } \boldsymbol{W}) \end{aligned}$$

Definition L.6 (Schäffer's metric). [22]

DEF

$$\begin{aligned} \mathsf{d}(\boldsymbol{V},\boldsymbol{W}) &= \log\left(1+\max\{\mathsf{r}(\boldsymbol{V},\boldsymbol{W}),\mathsf{r}(\boldsymbol{W},\boldsymbol{V})\}\right) \qquad \text{where} \\ \mathsf{r}(\boldsymbol{V},\boldsymbol{W}) &\triangleq \begin{cases} \inf\{\|\mathbf{A}-\mathbf{I}\| \,|\, \mathbf{A}\boldsymbol{V}=\boldsymbol{W}\} & \text{if } \mathbf{A} \text{ and } \mathbf{A}^{-1} \text{ both exist} \\ 1 & \text{otherwise} \end{cases} \end{aligned}$$

L.5 Literature

LITERATURE SURVEY:

1. Lattice of subspaces
   - Birkhoff and Neumann (1936)
   - Husimi (1937)
   - Sasaki (1954)
   - Loomis (1955)
   - von Neumann (1960)
   - Holland (1970)
   - Halmos (1998b)
   - Amemiya and Araki (1966)
   - Gudder (1979)
   - Gudder (2005)
2. Characterizations of lattice of Hilbert subspaces (cf Iturrioz (1985) page 60):
   - Kakutani and Mackey (1946) ⟨using Banach spaces⟩
   - Piron (1964a) ⟨using pre-Hilbert spaces⟩
     - Piron (1964b) ⟨using pre-Hilbert spaces⟩
   - Amemiya and Araki (1966) ⟨using pre-Hilbert spaces⟩
   - Wilbur (1975) ⟨using locally convex spaces⟩
3. Metrics on subspaces:
   - Burago et al. (2001)

[20] Deza and Deza (2006) page 235, Akhiezer and Glazman (1993) page 69, Berkson (1963) page 8, Krein and Krasnoselski (1947)
[21] Akhiezer and Glazman (1993) page 70, Berkson (1963) page 8, Krein et al. (1948)
[22] Massera and Schäffer (1958) pages 562–563, Berkson (1963) pages 7–8

APPENDIX M

NORMED ALGEBRAS

M.1 Algebras

All *linear spaces* (Definition J.1 page 247) are equipped with an operation by which vectors in the spaces can be added together. Linear spaces also have an operation that allows a scalar and a vector to be "multiplied" together. But linear spaces in general have no operation that allows two vectors to be multiplied together. A linear space together with such an operator is an **algebra**.[1]

There are many many possible algebras—many more than one can shake a stick at, as indicated by Michiel Hazewinkel in his book, *Handbook of Algebras*: "Algebra, as we know it today (2005), consists of many different ideas, concepts and results. A reasonable estimate of the number of these different items would be somewhere between 50,000 and 200,000. Many of these have been named and many more could (and perhaps should) have a "name" or other convenient designation."[2]

Definition M.1. [3] *Let* $\boldsymbol{A}$ *be an* ALGEBRA.

DEF *An algebra* $\boldsymbol{A}$ *is* ***unital*** *if* $\exists u \in \boldsymbol{A}$ such that $ux = xu = x \qquad \forall x \in \boldsymbol{A}$

Definition M.2. [4] *Let* $\boldsymbol{A}$ *be an* UNITAL ALGEBRA *(Definition M.1 page 303) with unit* e.

DEF

$$\begin{array}{llcll}
\text{The \textbf{spectrum} of } x \in \boldsymbol{A} \text{ is} & \sigma(x) & \triangleq & \left\{\lambda \in \mathbb{C} | \lambda e - x \text{ is not invertible}\right\}. & \\
\text{The \textbf{resolvent} of } x \in \boldsymbol{A} \text{ is} & \rho_x(\lambda) & \triangleq & (\lambda e - x)^{-1} & \forall \lambda \notin \sigma(x). \\
\text{The \textbf{spectral radius} of } x \in \boldsymbol{A} \text{ is} & r(x) & \triangleq & \sup\left\{|\lambda| \,|\, \lambda \in \sigma(x)\right\}. &
\end{array}$$

[1] Fuchs (1995) page 2
[2] Hazewinkel (2000) page v
[3] Folland (1995) page 1
[4] Folland (1995) pages 3–4

M.2 Star-Algebras

Definition M.3. [5] *Let $\boldsymbol{A}$ be an* ALGEBRA.

DEF

*The pair $(\boldsymbol{A}, *)$ is a **$*$-algebra**, or **star-algebra**, if*

1.	$(x+y)^* = x^* + y^*$	$\forall x,y \in \boldsymbol{A}$	(DISTRIBUTIVE)	and
2.	$(\alpha x)^* = \bar{\alpha} x^*$	$\forall x \in \boldsymbol{A}, \alpha \in \mathbb{C}$	(CONJUGATE LINEAR)	and
3.	$(xy)^* = y^* x^*$	$\forall x,y \in \boldsymbol{A}$	(ANTIAUTOMORPHIC)	and
4.	$x^{**} = x$	$\forall x \in \boldsymbol{A}$	(INVOLUTORY)	

*The operator $*$ is called an **involution** on the algebra $\boldsymbol{A}$.*

Proposition M.1. [6] *Let $(\boldsymbol{A}, *)$ be an* UNITAL $*$-ALGEBRA.

PRP

$$x \text{ is invertible} \implies \begin{cases} 1. & x^* \text{ is INVERTIBLE} & \forall x \in \boldsymbol{A} \quad \text{and} \\ 2. & (x^*)^{-1} = (x^{-1})^* & \forall x \in \boldsymbol{A} \end{cases}$$

✎PROOF: Let e be the unit element of $(\boldsymbol{A}, *)$.

1. Proof that $e^* = e$:

$$\begin{aligned}
x\,e^* &= (x\,e^*)^{**} && \text{by } \textit{involutory} \text{ property of } * && \text{(Definition M.3 page 304)}\\
&= (x^*\,e^{**})^* && \text{by } \textit{antiautomorphic} \text{ property of } * && \text{(Definition M.3 page 304)}\\
&= (x^*\,e)^* && \text{by } \textit{involutory} \text{ property of } * && \text{(Definition M.3 page 304)}\\
&= (x^*)^* && \text{by definition of } e\\
&= x && \text{by } \textit{involutory} \text{ property of } * && \text{(Definition M.3 page 304)}\\
e^*\,x &= (e^*\,x)^{**} && \text{by } \textit{involutory} \text{ property of } * && \text{(Definition M.3 page 304)}\\
&= (e^{**}\,x^*)^* && \text{by } \textit{antiautomorphic} \text{ property of } * && \text{(Definition M.3 page 304)}\\
&= (e\,x^*)^* && \text{by } \textit{involutory} \text{ property of } * && \text{(Definition M.3 page 304)}\\
&= (x^*)^* && \text{by definition of } e\\
&= x && \text{by } \textit{involutory} \text{ property of } * && \text{(Definition M.3 page 304)}
\end{aligned}$$

2. Proof that $(x^*)^{-1} = (x^{-1})^*$:

$$\begin{aligned}
(x^{-1})^*\,(x^*) &= [x\,(x^{-1})]^* && \text{by } \textit{antiautomorphic} \text{ and } \textit{involution} \text{ properties of } * && \text{(Definition M.3 page 304)}\\
&= e^*\\
&= e && \text{by item (1) page 304}\\
(x^*)\,(x^{-1})^* &= [x^{-1}\,x]^* && \text{by } \textit{antiautomorphic} \text{ and } \textit{involution} \text{ properties of } * && \text{(Definition M.3 page 304)}\\
&= e^*\\
&= e && \text{by item (1) page 304}
\end{aligned}$$

⇰

Definition M.4. [7] *Let $(\boldsymbol{A}, \|\cdot\|)$ be a $*$-*ALGEBRA *(Definition M.3 page 304)*.

DEF

- *An element $x \in \boldsymbol{A}$ is **hermitian** or **self-adjoint** if $x^* = x$.*
- *An element $x \in \boldsymbol{A}$ is **normal** if $xx^* = x^*x$.*
- *An element $x \in \boldsymbol{A}$ is a **projection** if $xx = x$ (INVOLUTORY) and $x^* = x$ (HERMITIAN).*

[5] Rickart (1960) page 178, Gelfand and Naimark (1964), page 241
[6] Folland (1995) page 5
[7] Rickart (1960) page 178, Gelfand and Naimark (1964), page 242

Theorem M.1. [8] *Let* $(\boldsymbol{A}, \|\cdot\|)$ *be a* $*$-ALGEBRA *(Definition M.3 page 304).*

THM

$$\underbrace{x = x^* \text{ and } y = y^*}_{x \text{ and } y \text{ are HERMITIAN}} \implies \left\{\begin{array}{lll} x+y &= (x+y)^* & (x+y \text{ is self adjoint}) \\ x^* &= (x^*)^* & (x^* \text{ is self adjoint}) \\ \underbrace{xy = (xy)^*}_{(xy) \text{ is HERMITIAN}} & \iff \underbrace{xy = yx}_{\text{commutative}} & \end{array}\right.$$

✎PROOF:

$$\begin{aligned} (x+y)^* &= x^* + y^* && \text{by } \textit{distributive} \text{ property of } * && \text{(Definition M.3 page 304)} \\ &= x + y && \text{by left hypothesis} \end{aligned}$$

$$\left(x^*\right)^* = x \qquad \text{by } \textit{involutory} \text{ property of } * \qquad \text{(Definition M.3 page 304)}$$

Proof that $xy = (xy)^* \implies xy = yx$

$$\begin{aligned} xy &= (xy)^* && \text{by left hypothesis} \\ &= y^*x^* && \text{by } \textit{antiautomorphic} \text{ property of } * && \text{(Definition M.3 page 304)} \\ &= yx && \text{by left hypothesis} \end{aligned}$$

Proof that $xy = (xy)^* \impliedby xy = yx$

$$\begin{aligned} (xy)^* &= (yx)^* && \text{by left hypothesis} \\ &= x^*y^* && \text{by } \textit{antiautomorphic} \text{ property of } * && \text{(Definition M.3 page 304)} \\ &= xy && \text{by left hypothesis} \end{aligned}$$

✏

Definition M.5 (Hermitian components). [9] *Let* $(\boldsymbol{A}, \|\cdot\|)$ *be a* $*$-ALGEBRA *(Definition M.3 page 304).*

DEF

The ***real part*** *of* x *is defined as* $\mathbf{R}_{\mathrm{e}}x \triangleq \frac{1}{2}\left(x + x^*\right)$

The ***imaginary part*** *of* x *is defined as* $\mathbf{I}_{\mathrm{m}}x \triangleq \frac{1}{2i}\left(x - x^*\right)$

Theorem M.2. [10] *Let* $(\boldsymbol{A}, *)$ *be a* $*$-ALGEBRA *(Definition M.3 page 304).*

THM

$$\begin{array}{llll} \mathbf{R}_{\mathrm{e}}x &= \left(\mathbf{R}_{\mathrm{e}}x\right)^* & \forall x \in \boldsymbol{A} & (\mathbf{R}_{\mathrm{e}}x \text{ is HERMITIAN}) \\ \mathbf{I}_{\mathrm{m}}x &= \left(\mathbf{I}_{\mathrm{m}}x\right)^* & \forall x \in \boldsymbol{A} & (\mathbf{I}_{\mathrm{m}}x \text{ is HERMITIAN}) \end{array}$$

✎PROOF:

$$\begin{aligned} \left(\mathbf{R}_{\mathrm{e}}x\right)^* &= \left(\frac{1}{2}\left(x + x^*\right)\right)^* && \text{by definition of } \Re && \text{(Definition M.5 page 305)} \\ &= \frac{1}{2}\left(x^* + x^{**}\right) && \text{by } \textit{distributive} \text{ property of } * && \text{(Definition M.3 page 304)} \\ &= \frac{1}{2}\left(x^* + x\right) && \text{by } \textit{involutory} \text{ property of } * && \text{(Definition M.3 page 304)} \\ &= \mathbf{R}_{\mathrm{e}}x && \text{by definition of } \Re && \text{(Definition M.5 page 305)} \\ \left(\mathbf{I}_{\mathrm{m}}x\right)^* &= \left(\frac{1}{2i}\left(x - x^*\right)\right)^* && \text{by definition of } \Im && \text{(Definition M.5 page 305)} \end{aligned}$$

[8] Michel and Herget (1993) page 429
[9] Michel and Herget (1993) page 430, Rickart (1960) page 179, Gelfand and Naimark (1964), page 242
[10] Michel and Herget (1993) page 430, Halmos (1998a) page 42

$$
\begin{aligned}
&= \frac{1}{2i}\left(x^* - x^{**}\right) && \text{by } \textit{distributive} \text{ property of } * && \text{(Definition M.3 page 304)}\\
&= \frac{1}{2i}\left(x^* - x\right) && \text{by } \textit{involutory} \text{ property of } * && \text{(Definition M.3 page 304)}\\
&= \mathbf{I}_{\mathsf{m}} x && \text{by definition of } \Im && \text{(Definition M.5 page 305)}
\end{aligned}
$$

✏

Theorem M.3 (Hermitian representation). [11] *Let* $(\boldsymbol{A}, *)$ *be a* $*$-ALGEBRA *(Definition M.3 page 304)*.

THM

$$a = x + iy \quad \iff \quad x = \mathbf{R}_{\mathsf{e}} a \quad \textit{and} \quad y = \mathbf{I}_{\mathsf{m}} a$$

✎PROOF:

Proof that $a = x + iy \implies x = \mathbf{R}_{\mathsf{e}} a$ and $y = \mathbf{I}_{\mathsf{m}} a$:

$$
\begin{aligned}
& & a &= x + iy && \text{by left hypothesis}\\
&\implies & a^* &= (x + iy)^* && \text{by definition of } \textit{adjoint} && \text{(Definition M.4 page 304)}\\
& & &= x^* - iy^* && \text{by } \textit{distributive} \text{ property of } * && \text{(Definition M.3 page 304)}\\
& & &= x - iy && \text{by Theorem M.2 page 305}\\
&\implies & x &= a - iy && \text{by solving for } x \text{ in } a = x + iy \text{ equation}\\
& & x &= a^* + iy && \text{by solving for } x \text{ in } a^* = x - iy \text{ equation}\\
&\implies & x + x &= a + a^* && \text{by adding previous 2 equations}\\
&\implies & 2x &= a + a^* && \text{by solving for } x \text{ in previous equation}\\
&\implies & x &= \frac{1}{2}\left(a + a^*\right)\\
& & &= \mathbf{R}_{\mathsf{e}} a && \text{by definition of } \Re && \text{(Definition M.5 page 305)}
\end{aligned}
$$

$$
\begin{aligned}
& & iy &= a - x && \text{by solving for } iy \text{ in } a = x + iy \text{ equation}\\
& & iy &= -a^* + x && \text{by solving for } iy \text{ in } a = x + iy \text{ equation}\\
&\implies & iy + iy &= a - a^* && \text{by adding previous 2 equations}\\
&\implies & y &= \frac{1}{2i}\left(a - a^*\right) && \text{by solving for } iy \text{ in previous equations}\\
& & &= \mathbf{I}_{\mathsf{m}} a && \text{by definition of } \Im && \text{(Definition M.5 page 305)}
\end{aligned}
$$

Proof that $a = x + iy \impliedby x = \mathbf{R}_{\mathsf{e}} a$ and $y = \mathbf{I}_{\mathsf{m}} a$:

$$
\begin{aligned}
x + iy &= \mathbf{R}_{\mathsf{e}} a + i\,\mathbf{I}_{\mathsf{m}} a && \text{by right hypothesis}\\
&= \underbrace{\frac{1}{2}\left(a + a^*\right)}_{\mathbf{R}_{\mathsf{e}} a} + i\underbrace{\frac{1}{2i}\left(a - a^*\right)}_{\mathbf{I}_{\mathsf{m}} a} && \text{by definition of } \Re \text{ and } \Im && \text{(Definition M.5 page 305)}\\
&= \left(\frac{1}{2}a + \frac{1}{2}a\right) + \left(\cancelto{0}{\frac{1}{2}a^* - \frac{1}{2}a^*}\right)\\
&= a
\end{aligned}
$$

✏

[11] Michel and Herget (1993) page 430, Rickart (1960) page 179, Gelfand and Neumark (1943b) page 7

M.3 Normed Algebras

Definition M.6. [12] *Let $\boldsymbol{A}$ be an algebra.*

DEF

*The pair $(\boldsymbol{A}, \|\cdot\|)$ is a **normed algebra** if*

$$\|xy\| \le \|x\|\,\|y\| \qquad \forall x, y \in \boldsymbol{A} \qquad \text{(multiplicative condition)}$$

*A normed algebra $(\boldsymbol{A}, \|\cdot\|)$ is a **Banach algebra** if $(\boldsymbol{A}, \|\cdot\|)$ is also a Banach space.*

Proposition M.2.

PRP

$(\boldsymbol{A}, \|\cdot\|)$ *is a normed algebra* $\implies$ *multiplication is **continuous** in* $(\boldsymbol{A}, \|\cdot\|)$

✎PROOF:

1. Define $\mathsf{f}(x) \triangleq zx$. That is, the function f represents multiplication of x times some arbitrary value z.
2. Let $\delta \triangleq \|x - y\|$ and $\epsilon \triangleq \|\mathsf{f}(x) - \mathsf{f}(y)\|$.
3. To prove that multiplication (f) is *continuous* with respect to the metric generated by $\|\cdot\|$, we have to show that we can always make ϵ arbitrarily small for some $\delta > 0$.
4. And here is the proof that multiplication is indeed continuous in $(\boldsymbol{A}, \|\cdot\|)$:

$$\begin{aligned}
\|\mathsf{f}(x) - \mathsf{f}(y)\| &\triangleq \|zx - zy\| && \text{by definition of } \mathsf{f} && \text{(item (1) page 307)}\\
&= \|z(x - y)\| \\
&\le \|z\|\;\|x - y\| && \text{by definition of \textit{normed algebra}} && \text{(Definition M.6 page 307)}\\
&\triangleq \|z\|\;\delta && \text{by definition of } \delta && \text{(item (2) page 307)}\\
&\le \epsilon && \text{for some value of } \delta > 0
\end{aligned}$$

✏

Theorem M.4 (Gelfand-Mazur Theorem). [13] *Let $\mathbb{C}$ be the field of complex numbers.*

THM

$$\left.\begin{array}{l}(\boldsymbol{A}, \|\cdot\|) \text{ is a Banach algebra}\\ \text{every nonzero } x \in \boldsymbol{A} \text{ is invertible}\end{array}\right\} \implies \boldsymbol{A} \equiv \mathbb{C} \quad (\boldsymbol{A} \text{ is isomorphic to } \mathbb{C})$$

M.4 C* Algebras

Definition M.7. [14]

DEF

*The triple $(\boldsymbol{A}, \|\cdot\|, *)$ is a C^* **algebra** if*

1. $(\boldsymbol{A}, \|\cdot\|)$ *is a Banach algebra* *and*
2. $(\boldsymbol{A}, *)$ *is a $*$-algebra* *and*
3. $\|x^* x\| = \|x\|^2 \quad \forall x \in \boldsymbol{A}$.

*A C^* **algebra** $(\boldsymbol{A}, \|\cdot\|, *)$ is also called a **C star algebra**.*

[12] Rickart (1960) page 2, Berberian (1961) page 103 ⟨Theorem IV.9.2⟩
[13] Folland (1995) page 4, Mazur (1938) ⟨(statement)⟩, Gelfand (1941) ⟨(proof)⟩
[14] Folland (1995) page 1, Gelfand and Naimark (1964), page 241, Gelfand and Neumark (1943a), Gelfand and Neumark (1943b)

Theorem M.5. [15] *Let $\boldsymbol{A}$ be an algebra.*

THM
$(\boldsymbol{A}, \|\cdot\|, *)$ *is a C^** ***algebra*** $\implies$ $\|x^*\| = \|x\|$

✎PROOF:

$$
\begin{aligned}
\|x\| &= \frac{1}{\|x\|}\ \|x\|^2 \\
&= \frac{1}{\|x\|}\ \|x^* x\| && \text{by definition of } C^*\text{-}\textit{algebra} && \text{(Definition M.7 page 307)} \\
&\le \frac{1}{\|x\|}\ \|x^*\|\ \|x\| && \text{by definition of } \textit{normed algebra} && \text{(Definition M.6 page 307)} \\
&= \|x^*\| \\
\|x^*\| &\le \|x^{**}\| && \text{by previous result} \\
&= \|x\| && \text{by } \textit{involution} \text{ property of } * && \text{(Definition M.3 page 304)}
\end{aligned}
$$

✏

[15] Folland (1995) page 1, Gelfand and Neumark (1943b) page 4, Gelfand and Neumark (1943a)

APPENDIX N

INNER PRODUCT SPACES

N.1 Definition and basic results

Definition N.1. [1] *Let* $\boldsymbol{\Omega} \triangleq \left(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times})\right)$ *be a* LINEAR SPACE (Definition J.1 page 247).

DEF

A FUNCTIONAL $\langle \triangle \mid \triangledown \rangle \in \mathbb{F}^{X \times X}$ *is an* ***inner product*** *on* $\boldsymbol{\Omega}$ *if*

1.	$\langle \alpha \boldsymbol{x} \mid \boldsymbol{y} \rangle = \alpha \langle \boldsymbol{x} \mid \boldsymbol{y} \rangle$	$\forall \boldsymbol{x}, \boldsymbol{y} \in X,\ \forall \alpha \in \mathbb{C}$	(HOMOGENEOUS)	*and*
2.	$\langle \boldsymbol{x} + \boldsymbol{y} \mid \boldsymbol{u} \rangle = \langle \boldsymbol{x} \mid \boldsymbol{u} \rangle + \langle \boldsymbol{y} \mid \boldsymbol{u} \rangle$	$\forall \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u} \in X$	(ADDITIVE)	*and*
3.	$\langle \boldsymbol{x} \mid \boldsymbol{y} \rangle = \langle \boldsymbol{y} \mid \boldsymbol{x} \rangle^*$	$\forall \boldsymbol{x}, \boldsymbol{y} \in X$	(CONJUGATE SYMMETRIC).	*and*
4.	$\langle \boldsymbol{x} \mid \boldsymbol{x} \rangle \geq 0$	$\forall \boldsymbol{x} \in X$	(NON-NEGATIVE)	*and*
5.	$\langle \boldsymbol{x} \mid \boldsymbol{x} \rangle = 0 \iff \boldsymbol{x} = \mathbb{0}$	$\forall \boldsymbol{x} \in X$	(NON-ISOTROPIC)	

An inner product is also called a ***scalar product****.*
The tuple $\left(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \mid \triangledown \rangle\right)$ *is called an* ***inner product space****.*

Theorem N.1. [2] *Let* $\boldsymbol{\Omega} \triangleq \left(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \mid \triangledown \rangle\right)$ *be a* LINEAR SPACE (Definition J.1 page 247).

THM

1.	$\langle \boldsymbol{x} \mid \boldsymbol{y} + \boldsymbol{z} \rangle = \langle \boldsymbol{x} \mid \boldsymbol{y} \rangle + \langle \boldsymbol{x} \mid \boldsymbol{z} \rangle$	$\forall \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in X$		
2.	$\langle \boldsymbol{x} \mid \alpha \boldsymbol{y} \rangle = \alpha^* \langle \boldsymbol{x} \mid \boldsymbol{y} \rangle$	$\forall \boldsymbol{x}, \boldsymbol{y} \in X,\ \alpha \in \mathbb{F}$		
3.	$\langle \boldsymbol{x} \mid \mathbb{0} \rangle = \langle \mathbb{0} \mid \boldsymbol{x} \rangle = 0$	$\forall \boldsymbol{x} \in X$		
4.	$\langle \boldsymbol{x} - \boldsymbol{y} \mid \boldsymbol{z} \rangle = \langle \boldsymbol{x} \mid \boldsymbol{z} \rangle - \langle \boldsymbol{y} \mid \boldsymbol{z} \rangle$	$\forall \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in X$		
5.	$\langle \boldsymbol{x} \mid \boldsymbol{y} - \boldsymbol{z} \rangle = \langle \boldsymbol{x} \mid \boldsymbol{y} \rangle - \langle \boldsymbol{x} \mid \boldsymbol{z} \rangle$	$\forall \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in X$		
6.	$\langle \boldsymbol{x} \mid \boldsymbol{z} \rangle = \langle \boldsymbol{y} \mid \boldsymbol{z} \rangle$	$\forall \boldsymbol{z} \in X \neq \{\mathbb{0}\}$	$\iff$	$\boldsymbol{x} = \boldsymbol{y}$
7.	$\langle \boldsymbol{x} \mid \boldsymbol{y} \rangle = 0$	$\forall \boldsymbol{x} \in X$	$\iff$	$\boldsymbol{y} = \mathbb{0}$

✎PROOF:

$\langle \boldsymbol{x} \mid \boldsymbol{y} + \boldsymbol{z} \rangle = \langle \boldsymbol{y} + \boldsymbol{z} \mid \boldsymbol{x} \rangle^*$	by *conjugate symmetric* property of $\langle \triangle \mid \triangledown \rangle$	(Definition N.1 page 309)
$= (\langle \boldsymbol{y} \mid \boldsymbol{x} \rangle + \langle \boldsymbol{z} \mid \boldsymbol{x} \rangle)^*$	by *additive* property of $\langle \triangle \mid \triangledown \rangle$	(Definition N.1 page 309)
$= \langle \boldsymbol{y} \mid \boldsymbol{x} \rangle^* + \langle \boldsymbol{z} \mid \boldsymbol{x} \rangle^*$	by *distributive* property of $*$	(Definition M.3 page 304)
$= \langle \boldsymbol{x} \mid \boldsymbol{y} \rangle + \langle \boldsymbol{x} \mid \boldsymbol{z} \rangle$	by *conjugate symmetric* property of $\langle \triangle \mid \triangledown \rangle$	(Definition N.1 page 309)
$\langle \boldsymbol{x} \mid \alpha \boldsymbol{y} \rangle = \langle \alpha \boldsymbol{y} \mid \boldsymbol{x} \rangle^*$	by *conjugate symmetric* property of $\langle \triangle \mid \triangledown \rangle$	(Definition N.1 page 309)

[1] Istrăţescu (1987) page 111 ⟨Definition 4.1.1⟩, Bollobás (1999) pages 130–131, Haaser and Sullivan (1991) page 277, Aliprantis and Burkinshaw (1998) page 276, Peano (1888b) page 72
[2] Berberian (1961) page 27, Haaser and Sullivan (1991) page 277

$$
\begin{aligned}
&= (\alpha \langle y \mid x \rangle)^* && \text{by } \textit{homogeneous} \text{ property of } \langle \triangle \mid \triangledown \rangle && \text{(Definition N.1 page 309)}\\
&= \alpha^* \langle y \mid x \rangle^* && \text{by } \textit{antiautomorphic} \text{ property of } * && \text{(Definition M.3 page 304)}\\
&= \alpha^* \langle x \mid y \rangle && \text{by } \textit{conjugate symmetric} \text{ property of } \langle \triangle \mid \triangledown \rangle && \text{(Definition N.1 page 309)}\\
\langle x \mid \mathbb{0} \rangle &= \langle \mathbb{0} \mid x \rangle^* && \text{by } \textit{conjugate symmetric} \text{ property of } \langle \triangle \mid \triangledown \rangle && \text{(Definition N.1 page 309)}\\
&= \langle 0 \cdot y \mid x \rangle^* \\
&= (0 \cdot \langle y \mid x \rangle)^* && \text{by } \textit{homogeneous} \text{ property of } \langle \triangle \mid \triangledown \rangle && \text{(Definition N.1 page 309)}\\
&= 0 \\
\langle \mathbb{0} \mid x \rangle &= \langle 0 \cdot y \mid x \rangle \\
&= (0 \cdot \langle y \mid x \rangle) && \text{by } \textit{homogeneous} \text{ property of } \langle \triangle \mid \triangledown \rangle && \text{(Definition N.1 page 309)}\\
&= 0 \\
\langle x - y \mid z \rangle &= \langle x + (-y) \mid z \rangle && \text{by definition of } + \\
&= \langle x \mid z \rangle + \langle -y \mid z \rangle && \text{by } \textit{additive} \text{ property of } \langle \triangle \mid \triangledown \rangle && \text{(Definition N.1 page 309)}\\
&= \langle x \mid z \rangle - \langle y \mid z \rangle && \text{by } \textit{homogeneous} \text{ property of } \langle \triangle \mid \triangledown \rangle && \text{(Definition N.1 page 309)}\\
\langle x \mid y - z \rangle &= \langle y - z \mid x \rangle^* && \text{by } \textit{conjugate symmetric} \text{ property of } \langle \triangle \mid \triangledown \rangle && \text{(Definition N.1 page 309)}\\
&= (\langle y \mid x \rangle - \langle z \mid x \rangle)^* && \text{by 4.} \\
&= \langle y \mid x \rangle^* - \langle z \mid x \rangle^* && \text{by } \textit{distributive} \text{ property of } * && \text{(Definition M.3 page 304)}\\
&= \langle x \mid y \rangle - \langle x \mid z \rangle && \text{by } \textit{conjugate symmetric} \text{ property of } \langle \triangle \mid \triangledown \rangle && \text{(Definition N.1 page 309)}
\end{aligned}
$$

$$
\begin{aligned}
& & \langle x \mid z \rangle &= \langle y \mid z \rangle && \forall z \\
&\iff & \langle x \mid z \rangle - \langle y \mid z \rangle &= 0 && \forall z && \text{by property of complex numbers} \\
&\iff & \langle x - y \mid z \rangle &= 0 && \forall z && \text{by 4.} \\
&\iff & x - y &= \mathbb{0} && \forall z && \text{by } \textit{non-isotropic} \text{ property of } \langle \triangle \mid \triangledown \rangle \text{ (Definition N.1 page 309)}
\end{aligned}
$$

Proof that $\langle x \mid y \rangle = 0 \implies y = \mathbb{0}$:

1. Suppose $y \neq \mathbb{0}$;
2. Then $\langle y \mid y \rangle \neq \mathbb{0}$ by the *non-isotropic* property of $\langle \triangle \mid \triangledown \rangle$ (Definition N.1 page 309)
3. But because $y \in X$, the left hypothesis implies that $\langle y \mid y \rangle = 0$.
4. This is a *contradiction*.
5. Therefore $y \neq \mathbb{0}$ must be incorrect and $y = \mathbb{0}$ must be correct.

Proof that $\langle x \mid y \rangle = 0 \impliedby y = \mathbb{0}$:

$$
\begin{aligned}
\langle x \mid y \rangle &= \langle x \mid \mathbb{0} \rangle && \text{by right hypothesis} \\
&= 0 && \text{by Theorem N.1 page 309}
\end{aligned}
$$

⇰

One of the most useful and widely used inequalities in analysis is the *Cauchy-Schwarz Inequality* (sometimes also called the *Cauchy-Bunyakovsky-Schwarz Inequality*). In fact, we will use this inequality shortly to prove that every inner product space *has* a norm and therefore every inner product space *is* a normed linear space.

Theorem N.2 (Cauchy-Schwarz Inequality). [3] *Let* $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \mid \triangledown \rangle)$ *be an* INNER PRODUCT SPACE *and* $|\cdot| \in \mathbb{R}^{\mathbb{C}}$ *an* ABSOLUTE VALUE *function* (Definition G.4 page 222). *Let* $\|\cdot\|$ *be a function in* $\mathbb{R}^{\mathbb{F}}$ *such*

[3] Haaser and Sullivan (1991) page 278, Aliprantis and Burkinshaw (1998) page 278, Cauchy (1821) page 455, Bunyakovsky (1859) page 6, Schwarz (1885)

that $\|x\| \triangleq \sqrt{\langle x \mid x\rangle}$.[4]

THM	$\lvert\langle x \mid y\rangle\rvert^2 \leq \langle x \mid x\rangle \langle y \mid y\rangle$					$\forall x,y \in X$
	$\lvert\langle x \mid y\rangle\rvert^2 = \langle x \mid x\rangle \langle y \mid y\rangle$	$\iff$	$\exists \alpha \in \mathbb{F}$	such that	$y = \alpha x$	$\forall x,y \in X$
	$\lvert\langle x \mid y\rangle\rvert \leq \|x\| \|y\|$					$\forall x,y \in X$
	$\lvert\langle x \mid y\rangle\rvert = \|x\| \|y\|$	$\iff$	$\exists \alpha \in \mathbb{F}$	such that	$y = \alpha x$	$\forall x,y \in X$

✎PROOF:

1. Proof that $|\langle x \mid y\rangle| \leq \|x\| \|y\|$: [5]

   (a) $y = \mathbb{0}$ case:

$$\begin{aligned}
|\langle x \mid y\rangle|^2 &= |\langle x \mid \mathbb{0}\rangle|^2 && \text{by } y = \mathbb{0} \text{ hypothesis} \\
&= |\langle \mathbb{0} \mid x\rangle|^2 && \text{by Definition N.1 page 309} \\
&= |\langle 0\mathbb{0} \mid x\rangle|^2 && \text{by Definition J.1 page 247} \\
&= |0 \langle \mathbb{0} \mid x\rangle|^2 && \text{by Definition N.1 page 309} \\
&= 0 \\
&= \langle x \mid x\rangle \langle \mathbb{0} \mid \mathbb{0}\rangle \\
&= \langle x \mid x\rangle \langle y \mid y\rangle && \text{by } y = \mathbb{0} \text{ hypothesis}
\end{aligned}$$

   (b) $y \neq \mathbb{0}$ case: Let $\lambda \triangleq \frac{\langle x \mid y\rangle}{\langle y \mid y\rangle}$.

$$\begin{aligned}
0 &\leq \langle x - \lambda y \mid x - \lambda y\rangle && \text{by Definition N.1} \\
&= \langle x \mid x - \lambda y\rangle + \langle -\lambda y \mid x - \lambda y\rangle && \text{by Definition N.1} \\
&= \langle x - \lambda y \mid x\rangle^* - \lambda \langle y \mid x - \lambda y\rangle && \text{by Definition N.1} \\
&= \langle x \mid x\rangle^* + \langle -\lambda y \mid x\rangle^* - \lambda \langle x - \lambda y \mid y\rangle^* && \text{by Definition N.1} \\
&= \langle x \mid x\rangle^* - \lambda^* \langle y \mid x\rangle^* - \lambda \langle x \mid y\rangle^* - \lambda \langle -\lambda y \mid y\rangle^* && \text{by Definition N.1} \\
&= \langle x \mid x\rangle - \lambda^* \langle x \mid y\rangle - \lambda \langle x \mid y\rangle^* + \lambda\lambda^* \langle y \mid y\rangle^* && \text{by Definition N.1} \\
&= \langle x \mid x\rangle + \cancelto{0}{\left[\frac{\langle x \mid y\rangle}{\langle y \mid y\rangle} \lambda^* \langle y \mid y\rangle - \lambda^* \langle x \mid y\rangle\right]} - \frac{\langle x \mid y\rangle}{\langle y \mid y\rangle} \langle x \mid y\rangle^* && \text{by definition of } \lambda \\
&= \langle x \mid x\rangle - \frac{1}{\langle y \mid y\rangle} |\langle x \mid y\rangle|^2
\end{aligned}$$

$$\implies |\langle x \mid y\rangle|^2 \leq \langle x \mid x\rangle \langle y \mid y\rangle$$

2. Proof that $|\langle x \mid y\rangle|^2 = \langle x \mid x\rangle \langle y \mid y\rangle \iff y = ax$:
   Let $\frac{1}{a} \triangleq \lambda \triangleq \frac{\langle x \mid y\rangle}{\langle y \mid y\rangle}$. Then...

$$\begin{aligned}
& y = ax \\
\iff\quad & x = \lambda y \\
\iff\quad & x - \lambda y = \mathbb{0} \\
\iff\quad & 0 = \langle x - \lambda y \mid x - \lambda y\rangle && \text{by Definition N.1 page 309} \\
& = \langle x \mid x - \lambda y\rangle + \langle -\lambda y \mid x - \lambda y\rangle && \text{by Definition N.1 page 309} \\
& = \langle x - \lambda y \mid x\rangle^* - \lambda \langle y \mid x - \lambda y\rangle && \text{by Definition N.1 page 309} \\
& \qquad \vdots && \text{(same steps as in 1(b))}
\end{aligned}$$

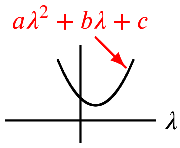
[4] The function $\|\cdot\|$ is a *norm* (Theorem N.4 page 314) and is called the *norm induced by the inner product* $\langle \triangle \mid \nabla\rangle$ (Definition N.2 page 315).

[5] Haaser and Sullivan (1991), page 278

$$
\begin{aligned}
&= \langle \boldsymbol{x} \mid \boldsymbol{x} \rangle - \frac{1}{\langle \boldsymbol{y} \mid \boldsymbol{y} \rangle} \; |\langle \boldsymbol{x} \mid \boldsymbol{y} \rangle|^2 \\
\iff \qquad & |\langle \boldsymbol{x} \mid \boldsymbol{y} \rangle|^2 = \langle \boldsymbol{x} \mid \boldsymbol{x} \rangle \; \langle \boldsymbol{y} \mid \boldsymbol{y} \rangle
\end{aligned}
$$

3. Alternate proof for $|\langle \boldsymbol{x} \mid \boldsymbol{y} \rangle| \le \|\boldsymbol{x}\| \, \|\boldsymbol{y}\|$: (Note: This is essentially the same proof as used by Schwarz). [6]

   (a) Proof that $\left\{ a\lambda^2 + b\lambda + c \ge 0 \quad \forall \lambda \in \mathbb{R} \right\} \implies \left\{ b^2 \le 4ac \right\}$ (quadratic discriminate inequality):



   Let $k \in (0, \infty)$, and $r_1, r_2 \in \mathbb{C}$ be the roots of $a\lambda^2 + b\lambda + c = 0$. Then

$$
\begin{aligned}
& & 0 &\le a\lambda^2 + b\lambda + c && \text{by left hypothesis} \\
& & &= k(\lambda - r_1)(\lambda - r_2) && \text{by definition of } r_1 \text{ and } r_2 \\
& & &= k(\lambda^2 - r_1\lambda - r_2\lambda + r_1 r_2) && \\
\implies \quad & \lambda^2 - r_1\lambda - r_2\lambda + r_1 r_2 \ge 0 & & && \\
& \implies & r_1 &= r_2^* && \text{because } r_1 r_2 \ge 0 \text{ for } \lambda = 0
\end{aligned}
$$

   The *quadratic equation* places another constraint on r_1 and r_2:

$$
\begin{aligned}
\frac{b^2 + \sqrt{b^2 - 4ac}}{2a} &= r_1 && \text{by quadratic equation} \\
&= r_2^* && \text{by previous result} \\
&= \left(\frac{b^2 - \sqrt{b^2 - 4ac}}{2a} \right)^* && \text{by quadratic equation}
\end{aligned}
$$

   The only way for this to be true is if $b^2 \le 4ac$ (the **discriminate** is non-positive).

   (b) Proof that $\langle \boldsymbol{y} \mid \boldsymbol{y} \rangle \, \lambda^2 + 2|\langle \boldsymbol{x} \mid \boldsymbol{y} \rangle|\lambda + \langle \boldsymbol{x} \mid \boldsymbol{x} \rangle \ge 0 \quad \forall \lambda \in \mathbb{R}$:

$$
\begin{aligned}
0 &\le \langle \boldsymbol{x} + \alpha\boldsymbol{y} \mid \boldsymbol{x} + \alpha\boldsymbol{y} \rangle && \text{by Definition N.1 page 309} \\
&= \langle \boldsymbol{x} \mid \boldsymbol{x} + \alpha\boldsymbol{y} \rangle + \langle \alpha\boldsymbol{y} \mid \boldsymbol{x} + \alpha\boldsymbol{y} \rangle && \text{by Definition N.1 page 309} \\
&= \langle \boldsymbol{x} \mid \boldsymbol{x} + \alpha\boldsymbol{y} \rangle + \alpha \, \langle \boldsymbol{y} \mid \boldsymbol{x} + \alpha\boldsymbol{y} \rangle && \text{by Definition N.1 page 309} \\
&= \langle \boldsymbol{x} + \alpha\boldsymbol{y} \mid \boldsymbol{x} \rangle^* + \alpha \, \langle \boldsymbol{x} + \alpha\boldsymbol{y} \mid \boldsymbol{y} \rangle^* && \text{by Definition N.1 page 309} \\
&= \langle \boldsymbol{x} \mid \boldsymbol{x} \rangle^* + \langle \alpha\boldsymbol{y} \mid \boldsymbol{x} \rangle^* + \alpha \, \langle \boldsymbol{x} \mid \boldsymbol{y} \rangle^* + \alpha \, \langle \alpha\boldsymbol{y} \mid \boldsymbol{y} \rangle^* && \text{by Definition N.1 page 309} \\
&= \langle \boldsymbol{x} \mid \boldsymbol{x} \rangle^* + \alpha^* \, \langle \boldsymbol{y} \mid \boldsymbol{x} \rangle^* + \alpha \, \langle \boldsymbol{x} \mid \boldsymbol{y} \rangle^* + \alpha\alpha^* \, \langle \boldsymbol{y} \mid \boldsymbol{y} \rangle^* && \text{by Definition N.1 page 309} \\
&= \langle \boldsymbol{x} \mid \boldsymbol{x} \rangle + \alpha^* \, \langle \boldsymbol{x} \mid \boldsymbol{y} \rangle + \left(\alpha^* \, \langle \boldsymbol{x} \mid \boldsymbol{y} \rangle \right)^* + |\alpha|^2 \, \langle \boldsymbol{y} \mid \boldsymbol{y} \rangle && \text{by Definition N.1 page 309} \\
&= \langle \boldsymbol{x} \mid \boldsymbol{x} \rangle + 2\Re \left(\alpha^* \, \langle \boldsymbol{x} \mid \boldsymbol{y} \rangle \right) + |\alpha|^2 \, \langle \boldsymbol{y} \mid \boldsymbol{y} \rangle && \\
&\le \langle \boldsymbol{x} \mid \boldsymbol{x} \rangle + 2\left|\alpha^* \, \langle \boldsymbol{x} \mid \boldsymbol{y} \rangle\right| + |\alpha|^2 \, \langle \boldsymbol{y} \mid \boldsymbol{y} \rangle && \\
&= \langle \boldsymbol{x} \mid \boldsymbol{x} \rangle + 2|\langle \boldsymbol{x} \mid \boldsymbol{y} \rangle||\alpha| + \langle \boldsymbol{y} \mid \boldsymbol{y} \rangle \, |\alpha|^2 && \\
&= \langle \boldsymbol{y} \mid \boldsymbol{y} \rangle \, |\alpha|^2 + 2 \, |\langle \boldsymbol{x} \mid \boldsymbol{y} \rangle| \, |\alpha| + \langle \boldsymbol{x} \mid \boldsymbol{x} \rangle && \\
&= \underbrace{\langle \boldsymbol{y} \mid \boldsymbol{y} \rangle}_{a} \lambda^2 + \underbrace{2|\langle \boldsymbol{x} \mid \boldsymbol{y} \rangle|}_{b} \lambda + \underbrace{\langle \boldsymbol{x} \mid \boldsymbol{x} \rangle}_{c} && \text{because } \lambda \triangleq |\alpha| \in \mathbb{R}
\end{aligned}
$$

[6] Aliprantis and Burkinshaw (1998) page 278, Steele (2004) page 11

(c) The above equation is in the quadratic form used in the lemma of part (a).

$$\Big(\underbrace{2|\langle \boldsymbol{x} \,|\, \boldsymbol{y}\rangle|}_{b}\Big)^2 \le 4\underbrace{\langle \boldsymbol{y} \,|\, \boldsymbol{y}\rangle}_{a}\ \underbrace{\langle \boldsymbol{x} \,|\, \boldsymbol{x}\rangle}_{c} \qquad \text{by the results of parts (a) and (b)}$$

$$\implies |\langle \boldsymbol{x} \,|\, \boldsymbol{y}\rangle|^2 \le \langle \boldsymbol{x} \,|\, \boldsymbol{x}\rangle\ \langle \boldsymbol{y} \,|\, \boldsymbol{y}\rangle$$

4. Proof that $|\langle \boldsymbol{x} \,|\, \boldsymbol{y}\rangle| \le \|\boldsymbol{x}\|\ \|\boldsymbol{y}\|$:
   This follows directly from the definition $\|\boldsymbol{x}\| \triangleq \sqrt{\langle \boldsymbol{x} \,|\, \boldsymbol{x}\rangle}$.

5. Proof that $|\langle \boldsymbol{x} \,|\, \boldsymbol{y}\rangle| = \|\boldsymbol{x}\|\ \|\boldsymbol{y}\| \iff \exists \alpha \in \mathbb{C}$ such that $\boldsymbol{y} = \alpha\boldsymbol{x}$:
   This follows directly from the definition $\|\boldsymbol{x}\| \triangleq \sqrt{\langle \boldsymbol{x} \,|\, \boldsymbol{x}\rangle}$.

✏

Corollary N.1. [7] *Let* $\big(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \,|\, \nabla\rangle\big)$ *be an* INNER PRODUCT SPACE.

COR: $\langle \boldsymbol{x} \,|\, \boldsymbol{y}\rangle$ *is* CONTINUOUS *in both* $\boldsymbol{x}$ *and* $\boldsymbol{y}$.

✎PROOF:

1. Let $\|\boldsymbol{x}\| \triangleq \sqrt{\langle \boldsymbol{x} \,|\, \boldsymbol{x}\rangle}$.

2. Proof:

$$\begin{aligned}
|\langle \boldsymbol{x}+\boldsymbol{\epsilon} \,|\, \boldsymbol{y}\rangle - \langle \boldsymbol{x} \,|\, \boldsymbol{y}\rangle|^2 &= |\langle \boldsymbol{x}+\boldsymbol{\epsilon}-\boldsymbol{x} \,|\, \boldsymbol{y}\rangle|^2 && \text{by } \textit{additivity} \text{ of } \langle \triangle \,|\, \nabla\rangle && \text{(Definition N.1 page 309)}\\
&= |\langle \boldsymbol{\epsilon} \,|\, \boldsymbol{y}\rangle|^2\\
&\le \|\boldsymbol{\epsilon}\|^2\ \|\boldsymbol{y}\| && \text{by } \textit{Cauchy-Schwarz Inequality} && \text{(Theorem N.2 page 310)}
\end{aligned}$$

3. Alternate proof (see Haaser and Sullivan (1991) page 278)

$$\begin{aligned}
|\langle \boldsymbol{x} \,|\, \boldsymbol{y}\rangle - \langle \boldsymbol{x}_0 \,|\, \boldsymbol{y}_0\rangle| &= \Big|\langle \boldsymbol{x}-\boldsymbol{x}_0 \,|\, \boldsymbol{y}-\boldsymbol{y}_0\rangle \underbrace{-\langle \boldsymbol{x}_0 \,|\, \boldsymbol{y}_0\rangle + \langle \boldsymbol{x} \,|\, \boldsymbol{y}_0\rangle}_{\langle \boldsymbol{x}-\boldsymbol{x}_0 \,|\, \boldsymbol{y}_0\rangle} \underbrace{+\langle \boldsymbol{x}_0 \,|\, \boldsymbol{y}\rangle - \langle \boldsymbol{x}_0 \,|\, \boldsymbol{y}_0\rangle}_{\langle \boldsymbol{x}_0 \,|\, \boldsymbol{y}-\boldsymbol{y}_0\rangle}\Big|\\
&= |\langle \boldsymbol{x}-\boldsymbol{x}_0 \,|\, \boldsymbol{y}-\boldsymbol{y}_0\rangle + \langle \boldsymbol{x}-\boldsymbol{x}_0 \,|\, \boldsymbol{y}_0\rangle + \langle \boldsymbol{x}_0 \,|\, \boldsymbol{y}-\boldsymbol{y}_0\rangle|\\
&\le |\langle \boldsymbol{x}-\boldsymbol{x}_0 \,|\, \boldsymbol{y}-\boldsymbol{y}_0\rangle| + |\langle \boldsymbol{x}-\boldsymbol{x}_0 \,|\, \boldsymbol{y}_0\rangle| + |\langle \boldsymbol{x}_0 \,|\, \boldsymbol{y}-\boldsymbol{y}_0\rangle|\\
&\le \|\boldsymbol{x}-\boldsymbol{x}_0\|\ \|\boldsymbol{y}-\boldsymbol{y}_0\| + \|\boldsymbol{x}-\boldsymbol{x}_0\|\ \|\boldsymbol{y}_0\| + \|\boldsymbol{x}_0\|\ \|\boldsymbol{y}-\boldsymbol{y}_0\| && \text{by Cauchy-Swartz ineq. page 310}\\
&\le \big(\max\{\|\boldsymbol{x}-\boldsymbol{x}_0\|, \|\boldsymbol{y}-\boldsymbol{y}_0\|\}\big)^2 + (\|\boldsymbol{x}_0\| + \|\boldsymbol{y}_0\|)\max\{\|\boldsymbol{x}-\boldsymbol{x}_0\|, \|\boldsymbol{y}-\boldsymbol{y}_0\|\}\\
&= \max\{\|\boldsymbol{x}-\boldsymbol{x}_0\|, \|\boldsymbol{y}-\boldsymbol{y}_0\|\}\big(1+\max\{\|\boldsymbol{x}-\boldsymbol{x}_0\|, \|\boldsymbol{y}-\boldsymbol{y}_0\|\}\big)\\
&= \mathsf{d}\big((\boldsymbol{x},\boldsymbol{y}),(\boldsymbol{x}_0,\boldsymbol{y}_0)\big)\big[1+\mathsf{d}\big((\boldsymbol{x},\boldsymbol{y}),(\boldsymbol{x}_0,\boldsymbol{y}_0)\big)\big]\\
&\implies \langle \boldsymbol{x} \,|\, \boldsymbol{y}\rangle \to \langle \boldsymbol{x}_0 \,|\, \boldsymbol{y}_0\rangle \text{ as } (\boldsymbol{x},\boldsymbol{y}) \to (\boldsymbol{x}_0,\boldsymbol{y}_0)
\end{aligned}$$

✏

[7] Bollobás (1999) page 132, Aliprantis and Burkinshaw (1998) page 279 ⟨Lemma 32.4⟩, Haaser and Sullivan (1991) page 278

N.2 Relationship between norms and inner products

N.2.1 Norms induced by inner products

Lemma N.1 (Polar Identity). [8] *Let* $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \,|\, \triangledown \rangle)$ *be an* INNER PRODUCT SPACE *(Definition N.1 page 309). Let* $\Re z$ *represent the real part of* $z \in \mathbb{C}$*. Let* $\|\cdot\|$ *be a function in* $\mathbb{R}^{\mathbb{F}}$ *such that* $\|x\| \triangleq \sqrt{\langle x \,|\, x \rangle}$.[9]

LEM
$$\|x + y\|^2 = \|x\|^2 + 2\mathbf{R}_e\left[\langle x \,|\, y \rangle\right] + \|y\|^2 \qquad \forall x,y \in X$$

✎PROOF:

$$\begin{aligned}
\|x + y\|^2 &= \langle x + y \,|\, x + y \rangle && \text{by definition of } \textit{induced norm} && \text{(Theorem N.4 page 314)}\\
&= \langle x \,|\, x + y \rangle + \langle y \,|\, x + y \rangle && \text{by Definition N.1 page 309}\\
&= \langle x + y \,|\, x \rangle^* + \langle x + y \,|\, y \rangle^* && \text{by Definition N.1 page 309}\\
&= \langle x \,|\, x \rangle^* + \langle y \,|\, x \rangle^* + \langle x \,|\, y \rangle^* + \langle y \,|\, y \rangle^* && \text{by Definition N.1 page 309}\\
&= \langle x \,|\, x \rangle + \langle x \,|\, y \rangle + \langle x \,|\, y \rangle^* + \langle y \,|\, y \rangle && \text{by definition of } \textit{inner product} && \text{(Definition N.1 page 309)}\\
&= \|x\|^2 + 2\Re \langle x \,|\, y \rangle + \|y\|^2 && \text{by definition of } \textit{induced norm} && \text{(Theorem N.4 page 314)}
\end{aligned}$$

✏

Theorem N.3 (Minkowski's Inequality). [10] *Let* $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \,|\, \triangledown \rangle)$ *be an* INNER PRODUCT SPACE*. Let* $\|\cdot\|$ *be a function in* $\mathbb{R}^{\mathbb{F}}$ *such that* $\|x\| \triangleq \sqrt{\langle x \,|\, x \rangle}$.[11]

THM
$$\|x + y\| \le \|x\| + \|y\| \qquad \forall x,y \in X$$

✎PROOF:

$$\begin{aligned}
\|x + y\|^2 &= \|x\|^2 + 2\Re\langle x \,|\, y \rangle + \|y\|^2 && \text{by } \textit{Polar Identity} && \text{(Lemma N.1 page 314)}\\
&\le \|x\|^2 + 2|\langle x \,|\, y \rangle| + \|y\|^2\\
&\le \|x\|^2 + 2\sqrt{\langle x \,|\, x \rangle}\sqrt{\langle y \,|\, y \rangle} + \|y\|^2 && \text{by } \textit{Cauchy-Schwarz Inequality} && \text{(Theorem N.2 page 310)}\\
&= (\|x\| + \|y\|)^2
\end{aligned}$$

✏

N.2.2 Inner products induced by norms

Theorem N.4 (induced norm). [12] *Let* $\boldsymbol{\Omega} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \,|\, \triangledown \rangle)$ *be an* INNER PRODUCT SPACE *(Definition N.1 page 309).*

---

[8] Conway (1990) page 4, Heil (2011) page 27 ⟨Lemma 1.36(a)⟩

[9] The function $\|\cdot\|$ is a *norm* (Theorem N.4 page 314) and is called the *norm induced by the inner product* $\langle \triangle \,|\, \triangledown \rangle$ (Definition N.2 page 315).

[10] Aliprantis and Burkinshaw (1998) pages 278–279 ⟨Theorem 32.3⟩, Maligranda (1995), Minkowski (1910) page 115

[11] The function $\|\cdot\|$ is a *norm* (Theorem N.4 page 314) and is called the *normed induced by the inner product* $\langle \triangle \,|\, \triangledown \rangle$ (Definition N.2 page 315).

[12] Aliprantis and Burkinshaw (1998) pages 278–279, Haaser and Sullivan (1991) page 278

THM $\|x\| \triangleq \sqrt{\langle x \,|\, x\rangle} \implies \|\cdot\|$ *is a* NORM

✎PROOF: For a function to be a norm, it must satisfy the four properties listed in Definition O.1 (page 327).

1. Proof that $\|\cdot\|$ is a norm:
   (a) Proof that $\|x\| > 0$ for $x \neq 0$ (non-negative):
   By Definition N.1 page 309, all inner products have this property.
   (b) Proof that $\|x\| = 0 \iff x = 0$ (non-isometric):
   By Definition N.1, all inner products have this property.
   (c) Prove $\|ax\| = |a|\,\|x\|$ (homogeneous):
   $$\|ax\| \triangleq \sqrt{\langle ax \,|\, ax\rangle} = \sqrt{aa^*\langle x \,|\, x\rangle} = \sqrt{|a|^2\langle x \,|\, x\rangle} = |a|\,\|x\|$$
   (d) Proof that $\|x + y\| \leq \|x\| + \|y\|$ (subadditive): This is true by *Minkowski's Inequality* (Theorem Q.5 page 353).
2. Proof that every inner product space is a normed linear space:
   Since every inner product induces a norm, so every inner product space has a norm (the norm induced by the inner product) and is therefore a normed linear space.

⇛

Theorem N.4 (previous theorem) demonstrates that in any inner product space $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \,|\, \nabla\rangle)$, the function $\|x\| \triangleq \sqrt{\langle x \,|\, x\rangle}$ is a norm. That is, $\|x\|$ is the *norm induced by the inner product*. This norm is formally defined next.

Definition N.2. *Let* $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \,|\, \nabla\rangle)$ *be an* INNER PRODUCT SPACE *(Definition N.1 page 309).*

DEF *The* ***norm induced by the inner product*** $\langle \triangle \,|\, \nabla\rangle$ *is defined as*
$$\|x\| \triangleq \sqrt{\langle x \,|\, x\rangle}$$

Theorem N.4 (page 314) demonstrates that if a *linear space* (Definition J.1 page 247) has an *inner product* (Definition N.1 page 309), then that inner product always induces a *norm* (Definition O.1 page 327), and the relationship between the two is simply $\|x\| = \sqrt{\langle x \,|\, x\rangle}$ (Definition N.2 page 315). But what about the converse? What if a linear space has a norm—can that norm also induce an inner product? The answer in general is "no": Not all norms can induce an inner product. But a less harsh answer is "sometimes": Some norms **can** induce inner products. This leads to some important and interesting questions:

1. How many different inner products can be induced from a single norm? The answer turns out to be **at most** one, but maybe none (Theorem N.5 page 316).
2. When a norm *can* induce an inner product, what is that (unique) inner product? The inner product expressed in terms of the norm is given by the *Polarization Identity* (Theorem N.6 page 316).
3. Which norms can induce an inner product and which ones cannot? The answer is that norms that satisfy the *parallelogram law* (Theorem N.7 page 317) **can** induce an inner product; and the ones that don't, cannot (Theorem N.7 page 317).

Theorem N.5. [13] *Let* $\left(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|\right)$ *be a* NORMED LINEAR SPACE *(Definition O.1 page 327).*

THM

$$\underbrace{\left.\begin{array}{l} \exists\ \langle\triangle\,|\,\triangledown\rangle \text{ and } (\cdot\,|\,\cdot) \text{ such that} \\ \|x\|^2 = \langle x\,|\,x\rangle = (x\,|\,x) \quad \forall x\in X \end{array}\right\}}_{\text{If a norm is induced by two inner products...}} \implies \underbrace{\langle x\,|\,y\rangle = (x\,|\,y) \quad \forall x,y\in X}_{\text{...then those two inner products are equivalent.}}$$

✎PROOF:

$$\begin{aligned}
2\langle x\,|\,y\rangle &= \left[\langle x\,|\,y\rangle + \langle y\,|\,x\rangle\right] + \left[\langle x\,|\,y\rangle - \langle y\,|\,x\rangle\right] \\
&= \left[\langle x\,|\,y\rangle + \langle y\,|\,x\rangle\right] + i\left[-i\langle x\,|\,y\rangle + i\langle y\,|\,x\rangle\right] \\
&= \left[\langle x\,|\,y\rangle + \langle y\,|\,x\rangle\right] + i\left[-\langle ix\,|\,y\rangle - \langle y\,|\,ix\rangle\right] \\
&= \left(\underbrace{\left[\langle x\,|\,y\rangle + \langle y\,|\,x\rangle + \langle x\,|\,x\rangle + \langle y\,|\,y\rangle\right]}_{\langle x+y\,|\,x+y\rangle} - \underbrace{\left[\langle x\,|\,x\rangle + \langle y\,|\,y\rangle\right]}_{\text{residue}}\right) \\
&\quad - i\left(\underbrace{\left[\langle ix\,|\,y\rangle + \langle y\,|\,ix\rangle + \langle ix\,|\,ix\rangle + \langle y\,|\,y\rangle\right]}_{\langle ix+y\,|\,ix+y\rangle} - \underbrace{\left[\langle ix\,|\,ix\rangle + \langle y\,|\,y\rangle\right]}_{\text{residue}}\right) \\
&= \left(\left[\langle x+y\,|\,x+y\rangle\right] - \left[\langle x\,|\,x\rangle + \langle y\,|\,y\rangle\right]\right) - i\left(\left[\langle ix+y\,|\,ix+y\rangle\right] - \left[\langle ix\,|\,ix\rangle + \langle y\,|\,y\rangle\right]\right) \\
&= \left(\left[(x+y\,|\,x+y)\right] - \left[(x\,|\,x) + (y\,|\,y)\right]\right) - i\left(\left[(ix+y\,|\,ix+y)\right] - \left[(ix\,|\,ix) + (y\,|\,y)\right]\right) \\
&= \left(\underbrace{\left[(x\,|\,y) + (y\,|\,x) + (x\,|\,x) + (y\,|\,y)\right]}_{(x+y\,|\,x+y)} - \underbrace{\left[(x\,|\,x) + (y\,|\,y)\right]}_{\text{residue}}\right) \\
&\quad - i\left(\underbrace{\left[(ix\,|\,y) + (y\,|\,ix) + (ix\,|\,ix) + (y\,|\,y)\right]}_{(ix+y\,|\,ix+y)} - \underbrace{\left[(ix\,|\,ix) + (y\,|\,y)\right]}_{\text{residue}}\right) \\
&= \left[(x\,|\,y) + (y\,|\,x)\right] + i\left[-(ix\,|\,y) - (y\,|\,ix)\right] \\
&= \left[(x\,|\,y) + (y\,|\,x)\right] + i\left[-i(x\,|\,y) + i(y\,|\,x)\right] \\
&= \left[(x\,|\,y) + (y\,|\,x)\right] + \left[(x\,|\,y) - (y\,|\,x)\right] \\
&= 2(x\,|\,y)
\end{aligned}$$

✎

Theorem N.6 (Polarization Identities). [14] *Let* $\boldsymbol{\Omega} \triangleq \left(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times})\right)$ *be a linear space,* $\langle\triangle\,|\,\triangledown\rangle \in \mathbb{F}^{X\times X}$ *a function, and* $\|x\| \triangleq \sqrt{\langle x\,|\,x\rangle}$.

THM

$$\left(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle\triangle\,|\,\triangledown\rangle\right) \text{ is an inner product space} \implies$$

$$\underbrace{4\langle x\,|\,y\rangle = \begin{cases} \|x+y\|^2 - \|x-y\|^2 + i\,\|x+iy\|^2 - i\,\|x-iy\|^2 & \text{for } \mathbb{F} = \mathbb{C} \quad \forall x,y\in X \\ \|x+y\|^2 - \|x-y\|^2 & \text{for } \mathbb{F} = \mathbb{R} \quad \forall x,y\in X \end{cases}}_{\text{inner product induced by norm}}$$

✎PROOF:

[13] Aliprantis and Burkinshaw (1998) page 280, Bollobás (1999) page 132, Jordan and von Neumann (1935) page 721

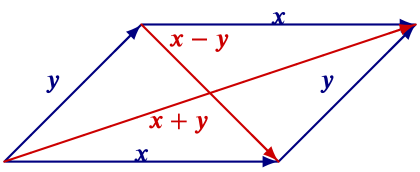
[14] Berberian (1961) pages 29–30 ⟨Theorem II.3.3⟩, Istrăţescu (1987) page 110 ⟨Proposition 4.1.5⟩, Bollobás (1999) page 132, Jordan and von Neumann (1935) page 721

1. These follow directly from properties of *bilinear functionals* (Theorem ?? page ??).

2. Alternative proof for $\mathbb{F} = \mathbb{C}$ case:

$$\begin{aligned}
&\|\boldsymbol{x}+\boldsymbol{y}\|^2 - \|\boldsymbol{x}-\boldsymbol{y}\|^2 + i\,\|\boldsymbol{x}+i\boldsymbol{y}\|^2 - i\,\|\boldsymbol{x}-i\boldsymbol{y}\|^2 \\
&= \underbrace{\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 + 2\Re\langle \boldsymbol{x} \,|\, \boldsymbol{y}\rangle}_{\langle \boldsymbol{x}+\boldsymbol{y} \,|\, \boldsymbol{x}+\boldsymbol{y}\rangle} - \underbrace{\left(\|\boldsymbol{x}\|^2 + \|-\boldsymbol{y}\|^2 + 2\Re\langle \boldsymbol{x} \,|\, -\boldsymbol{y}\rangle\right)}_{\langle \boldsymbol{x}-\boldsymbol{y} \,|\, \boldsymbol{x}-\boldsymbol{y}\rangle} \\
&\quad + \underbrace{i\left(\|\boldsymbol{x}\|^2 + \|i\boldsymbol{y}\|^2 + 2\Re\langle \boldsymbol{x} \,|\, i\boldsymbol{y}\rangle\right)}_{i\langle \boldsymbol{x}+i\boldsymbol{y} \,|\, \boldsymbol{x}+i\boldsymbol{y}\rangle} - \underbrace{i\left(\|\boldsymbol{x}\|^2 + \|-i\boldsymbol{y}\|^2 + 2\Re\langle \boldsymbol{x} \,|\, -i\boldsymbol{y}\rangle\right)}_{i\langle \boldsymbol{x}-i\boldsymbol{y} \,|\, \boldsymbol{x}-i\boldsymbol{y}\rangle} && \text{by Lemma N.1 page 314} \\
&= \underbrace{\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 + 2\Re\langle \boldsymbol{x} \,|\, \boldsymbol{y}\rangle}_{\langle \boldsymbol{x}+\boldsymbol{y} \,|\, \boldsymbol{x}+\boldsymbol{y}\rangle} - \underbrace{\left(\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 - 2\Re\langle \boldsymbol{x} \,|\, \boldsymbol{y}\rangle\right)}_{\langle \boldsymbol{x}-\boldsymbol{y} \,|\, \boldsymbol{x}-\boldsymbol{y}\rangle} \\
&\quad + \underbrace{i\left(\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 + 2\Re\langle \boldsymbol{x} \,|\, i\boldsymbol{y}\rangle\right)}_{i\langle \boldsymbol{x}+i\boldsymbol{y} \,|\, \boldsymbol{x}+i\boldsymbol{y}\rangle} - \underbrace{i\left(\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 - 2\Re\langle \boldsymbol{x} \,|\, i\boldsymbol{y}\rangle\right)}_{i\langle \boldsymbol{x}-i\boldsymbol{y} \,|\, \boldsymbol{x}-i\boldsymbol{y}\rangle} && \text{by Definition N.1 page 309} \\
&= 4\Re\langle \boldsymbol{x} \,|\, \boldsymbol{y}\rangle + 4i\Re\langle \boldsymbol{x} \,|\, i\boldsymbol{y}\rangle \\
&= \underbrace{2\left(\langle \boldsymbol{x} \,|\, \boldsymbol{y}\rangle + \langle \boldsymbol{x} \,|\, \boldsymbol{y}\rangle^*\right)}_{4\Re\langle \boldsymbol{x} \,|\, \boldsymbol{y}\rangle} + \underbrace{2i\left(\langle \boldsymbol{x} \,|\, i\boldsymbol{y}\rangle + \langle \boldsymbol{x} \,|\, i\boldsymbol{y}\rangle^*\right)}_{4i\Re\langle \boldsymbol{x} \,|\, i\boldsymbol{y}\rangle} \\
&= 2\left(\langle \boldsymbol{x} \,|\, \boldsymbol{y}\rangle + \langle \boldsymbol{x} \,|\, \boldsymbol{y}\rangle^*\right) + 2i\left(i^*\langle \boldsymbol{x} \,|\, \boldsymbol{y}\rangle + (i^{**})\langle \boldsymbol{x} \,|\, \boldsymbol{y}\rangle^*\right) \\
&= 2\left(\langle \boldsymbol{x} \,|\, \boldsymbol{y}\rangle + \langle \boldsymbol{x} \,|\, \boldsymbol{y}\rangle^*\right) + 2i\left(-i\langle \boldsymbol{x} \,|\, \boldsymbol{y}\rangle + i\langle \boldsymbol{x} \,|\, \boldsymbol{y}\rangle^*\right) && \text{by Definition N.1 page 309} \\
&= 2\langle \boldsymbol{x} \,|\, \boldsymbol{y}\rangle + 2\langle \boldsymbol{x} \,|\, \boldsymbol{y}\rangle^* + 2\langle \boldsymbol{x} \,|\, \boldsymbol{y}\rangle - 2\langle \boldsymbol{x} \,|\, \boldsymbol{y}\rangle^* \\
&= 4\langle \boldsymbol{x} \,|\, \boldsymbol{y}\rangle
\end{aligned}$$

✏



In plane geometry ($\mathbb{R}^2$), the *parallelogram law* states that the sum of the squares of the lengths of the sides of a parallelogram is equal to the sum of the squares of the lengths of its diagonals. This is illustrated in the figure to the left.

Acutally, the parallelogram law can be generalized to *any inner product space* (not just in the plane). And if the parallelogram law happens to hold true in a normed linear space, then that normed linear space is actually an *inner product space*. The parallelogram law and its relation to inner product spaces is stated in the next theorem.

Theorem N.7 (Parallelogram law). [15] *Let* $\boldsymbol{\Omega} \triangleq \left(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \,|\, \triangledown\rangle\right)$ *and* $\|\boldsymbol{x}\| \triangleq \sqrt{\langle \boldsymbol{x} \,|\, \boldsymbol{x}\rangle}$.

THM

$$\boldsymbol{\Omega} \text{ is an inner product space} \iff \underbrace{2\|\boldsymbol{x}\|^2 + 2\|\boldsymbol{y}\|^2 = \|\boldsymbol{x}+\boldsymbol{y}\|^2 + \|\boldsymbol{x}-\boldsymbol{y}\|^2}_{\text{PARALLELOGRAM LAW / VON NEUMANN-JORDAN CONDITION}} \qquad \forall \boldsymbol{x},\boldsymbol{y} \in \boldsymbol{\Omega}$$

✎PROOF:

1. Proof that $[\exists \langle \boldsymbol{x} \,|\, \boldsymbol{y}\rangle$ such that $\|\boldsymbol{x}\|^2 = \langle \boldsymbol{x} \,|\, \boldsymbol{x}\rangle\,] \implies$ [parallelogram law is true]:

[15] Amir (1986) page 8, Istrăţescu (1987) page 110, Day (1973) page 151, Halmos (1998a) page 14, Aliprantis and Burkinshaw (1998) pages 280–281 ⟨Theorem 32.6⟩, Riesz (1934) page 36?, Jordan and von Neumann (1935) pages 721–722

$$
\begin{aligned}
\|x+y\|^2 + \|x-y\|^2 &= \left[\|x\|^2 + \|y\|^2 + 2\mathbf{R}_e\left[2\langle x \mid y\rangle\right]\right] \qquad + \left[\|x\|^2 + \|-y\|^2 + 2\mathbf{R}_e\left[2\langle x \mid -y\rangle\right]\right] \\
&\qquad \text{by Lemma N.1 page 314} \\
&= \left[\|x\|^2 + \|y\|^2 + 2\mathbf{R}_e\left[2\langle x \mid y\rangle\right]\right] \qquad + \left[\|x\|^2 + \|y\|^2 - 2\mathbf{R}_e\left[2\langle x \mid y\rangle\right]\right] \\
&= 2\|x\|^2 + 2\|y\|^2
\end{aligned}
$$

2. Proof that $[\exists \langle x \mid y\rangle$ such that $\|x\|^2 = \langle x \mid x\rangle\,] \Longleftarrow$ [parallelogram law is true]:
   Note that **if** an inner product exists in the norm linear space $(\boldsymbol{\Omega}, \|\cdot\|)$, **then** that norm linear space is actually an inner product space $\left(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \mid \nabla\rangle\right)$. And if it is an inner product space, then by Theorem N.6 page 316 that inner product must be given by the **Polarization Identity**
   $\langle x \mid y\rangle = \|ax+y\|^2 - \|ax-y\|^2 + i\,\|ax+iy\|^2 - i\,\|ax-iy\|^2.$
   Therefore, here we must use the parallelogram law to show that the bilinear function $\mathsf{f}(x,y) \triangleq \langle x \mid y\rangle$ given on the left hand side of the "=" relation is indeed an inner product— that is, that it satisfies the requirements of Definition N.1 page 309.

   (a) Proof that $\langle x \mid x\rangle \geq 0$ (non-negative):

$$
\begin{aligned}
4\langle x \mid x\rangle &\triangleq \|x+x\|^2 - \cancelto{0}{\|x-x\|^2} + i\,\|x+ix\|^2 - i\,\|x-ix\|^2 && \text{by \textit{Polarization Identity}} \\
&= \|2x\|^2 - \mathbb{0} + i\left(\|x+ix\|^2 - \|x-ix\|^2\right) && \text{by Definition O.1 page 327} \\
&= |2|^2\,\|x\|^2 + i\left(\|x+ix\|^2 - |i|\,\|x-ix\|^2\right) \\
&= 4\|x\|^2 + i\left(\cancelto{0}{\|x+ix\|^2 - \|ix+x\|^2}\right) && \text{by Definition O.1 page 327} \\
&= 4\|x\|^2 \\
&\geq 0 && \text{by Definition O.1 page 327}
\end{aligned}
$$

   (b) Proof that $\langle x \mid x\rangle = 0 \iff x = \mathbb{0}$ (non-isotropic):

$$
\begin{aligned}
4\langle x \mid x\rangle &= 4\|x\|^2 && \text{by result of part (a)} \\
&= 0 \quad\iff\quad x = \mathbb{0} && \text{by Definition O.1 page 327}
\end{aligned}
$$

   (c) Proof that $\langle x+u \mid y\rangle = \langle x \mid y\rangle + \langle u \mid y\rangle$ (additive):[16]

$$
\begin{aligned}
4\langle x+y \mid z\rangle &= 8\left\langle \frac{x+y}{2} \,\middle|\, z\right\rangle && \text{by Definition N.1 page 309} \\
&= 2\left\|\frac{x+y}{2}+z\right\|^2 - 2\left\|\frac{x+y}{2}-z\right\|^2 \\
&\quad + 2i\left\|\frac{x+y}{2}+z\right\|^2 - 2i\left\|\frac{x+y}{2}-iz\right\|^2 && \text{by \textit{Polarization Identity}} \\
&= \left(2\left\|\frac{x+y}{2}+z\right\|^2 + 2\left\|\frac{x-y}{2}\right\|^2\right) \\
&\quad - \left(2\left\|\frac{x+y}{2}-z\right\|^2 + 2\left\|\frac{x-y}{2}\right\|^2\right) \\
&\quad + \left(2i\left\|\frac{x+y}{2}+z\right\|^2 + 2i\left\|\frac{x-y}{2}\right\|^2\right) \\
&\quad - \left(2i\left\|\frac{x+y}{2}-iz\right\|^2 + 2i\left\|\frac{x-y}{2}\right\|^2\right) \\
&= \left(\|x+z\|^2 + \|y+z\|^2\right) - \left(\|x-z\|^2 + \|y-z\|^2\right) \\
&\quad + \left(i\,\|x+z\|^2 + i\,\|y+z\|^2\right) - \left(i\,\|x-iz\|^2 + i\,\|y-iz\|^2\right) && \text{by parallelogram law}
\end{aligned}
$$

[16] Aliprantis and Burkinshaw (1998), page 281

$$\begin{aligned}
&= \left(\|x+z\|^2 - \|x-z\|^2 + i\,\|x+z\|^2 - i\,\|x-iz\|^2\right) \\
&\quad + \left(\|y+z\|^2 - \|y-z\|^2 + i\,\|y+z\|^2 - i\,\|y-iz\|^2\right) \\
&= 4\langle x \mid z\rangle + 4\langle y \mid z\rangle && \text{by Polarization Identity}
\end{aligned}$$

(d) Proof that $\langle x \mid y\rangle = \langle y \mid y\rangle^*$ (*conjugate symmetric*):

$$\begin{aligned}
4\langle x \mid y\rangle &\triangleq \|x+y\|^2 - \|x-y\|^2 + i\,\|x+iy\|^2 - i\,\|x-iy\|^2 && \text{by Polarization Identity} \\
&= \|y+x\|^2 - \|y-x\|^2 + i\,\|i(y-ix)\|^2 - i\,\|-i(y+ix)\|^2 && \text{by Definition J.1 page 247} \\
&= \|y+x\|^2 - \|y-x\|^2 + i\,\|y-ix\|^2 - i\,\|y+ix\|^2 && \text{by Definition O.1 page 327} \\
&= \left(\|y+x\|^2 - \|y-x\|^2 - i\,\|y-ix\|^2 + i\,\|y+ix\|^2\right)^* \\
&= \left(\|y+x\|^2 - \|y-x\|^2 + i\,\|y+ix\|^2 - i\,\|y-ix\|^2\right)^* \\
&\triangleq 4\langle y \mid x\rangle^* && \text{by \textit{Polarization Identity}}
\end{aligned}$$

(e) Proof that $\langle \alpha x \mid y\rangle = \alpha\langle x \mid y\rangle$ (*homogeneous*): [17]

i. Proof that $\langle \alpha x \mid y\rangle$ is linear in α:

$$\begin{aligned}
0 &\le \big|\|\alpha x + y\| - \|\beta x + y\|\big| && \text{by Definition G.4 page 222} \\
&\le \big|\|(\alpha x + y) - (\beta x + y)\|\big| && \text{by Theorem O.2 page 328} \\
&\le \big|\|(\alpha - \beta)x\|\big|
\end{aligned}$$

This implies that as $a \to \beta$, $\|\alpha x + y\| \to \|\beta x + y\|$, which by definition implies that $\|\alpha x + y\|$ linear in α. And by the parallelogram law, $\langle \alpha x \mid y\rangle$ is also linear in α.

ii. Proof that $\langle nx \mid y\rangle = n\langle x \mid y\rangle$ for $n \in \mathbb{Z}$ (integer case):

A. Proof for $n = \pm 1$:

$$\begin{aligned}
\langle nx \mid y\rangle &= \langle \pm 1x \mid y\rangle && \text{by } n = \pm 1 \text{ hypothesis} \\
&= \pm 1\langle x \mid y\rangle && \text{by definition of \textit{inner product}} \quad \text{(Definition N.1 page 309)} \\
&= n\langle x \mid y\rangle && \text{by } n = \pm 1 \text{ hypothesis}
\end{aligned}$$

B. Proof for $n = 0$:

$$\begin{aligned}
\langle nx \mid y\rangle &= \langle 0x \mid y\rangle && \text{by } n = 0 \text{ hypothesis} \\
&= \langle x - x \mid y\rangle \\
&= \langle x \mid y\rangle + \langle -1x \mid y\rangle \\
&= \langle x \mid y\rangle - 1\langle x \mid y\rangle \\
&= \langle x \mid y\rangle - \langle x \mid y\rangle \\
&= 0\langle x \mid y\rangle \\
&= n\langle x \mid y\rangle && \text{by } n = 0 \text{ hypothesis}
\end{aligned}$$

C. Proof for $n = \pm 2$:

$$\begin{aligned}
\langle nx \mid y\rangle &= \langle \pm 2x \mid y\rangle && \text{by } n = \pm 1 \text{ hypothesis} \\
&= \langle \pm(x + x) \mid y\rangle \\
&= \pm\langle x + x \mid y\rangle && \text{by definition of \textit{inner product}} \quad \text{(Definition N.1 page 309)} \\
&= \pm(\langle x \mid y\rangle + \langle x \mid y\rangle) && \text{by additive property} \\
&= \pm 2\langle x \mid y\rangle \\
&= n\langle x \mid y\rangle && \text{by } n = \pm 2 \text{ hypothesis}
\end{aligned}$$

[17] Aliprantis and Burkinshaw (1998), page 138

D. Proof that [n case] $\implies$ [$n \pm 1$ case]:

$$\begin{aligned}
\langle (n \pm 1)\boldsymbol{x} \mid \boldsymbol{y} \rangle &= \langle n\boldsymbol{x} \pm 1\boldsymbol{x} \mid \boldsymbol{y} \rangle \\
&= \langle n\boldsymbol{x} \mid \boldsymbol{y} \rangle + \langle \pm 1\boldsymbol{x} \mid \boldsymbol{y} \rangle && \text{by additive property} \\
&= n \langle \boldsymbol{x} \mid \boldsymbol{y} \rangle \pm 1 \langle \boldsymbol{x} \mid \boldsymbol{y} \rangle && \text{by left hypothesis} \\
&= (n \pm 1) \langle \boldsymbol{x} \mid \boldsymbol{y} \rangle
\end{aligned}$$

iii. Proof that $\langle q\boldsymbol{x} \mid \boldsymbol{y} \rangle = q \langle \boldsymbol{x} \mid \boldsymbol{y} \rangle$ for $q \in \mathbb{Q}$ (rational number case):

$$\begin{aligned}
\frac{n}{m} \langle \boldsymbol{x} \mid \boldsymbol{y} \rangle &= \frac{n}{m} \left\langle \frac{m}{m}\boldsymbol{x} \mid \boldsymbol{y} \right\rangle && \text{where } n, m \in \mathbb{Z} \text{ and } m \neq 0 \\
&= \frac{nm}{m} \left\langle \frac{1}{m}\boldsymbol{x} \mid \boldsymbol{y} \right\rangle && \text{by previous result} \\
&= \frac{m}{m} \left\langle \frac{n}{m}\boldsymbol{x} \mid \boldsymbol{y} \right\rangle && \text{by previous result} \\
&= \left\langle \frac{n}{m}\boldsymbol{x} \mid \boldsymbol{y} \right\rangle
\end{aligned}$$

iv. Proof that $\langle r\boldsymbol{x} \mid \boldsymbol{y} \rangle = r \langle \boldsymbol{x} \mid \boldsymbol{y} \rangle$ for all $r \in \mathbb{R}$ (real number case):
Because $\mathbb{Q}$ is dense in $\mathbb{R}$ and because $\|\alpha\boldsymbol{x} + \boldsymbol{y}\|$ is continuous in α, so $\langle \alpha\boldsymbol{x} \mid \boldsymbol{y} \rangle = \alpha \langle \boldsymbol{x} \mid \boldsymbol{y} \rangle$ for all $\alpha \in \mathbb{R}$.

v. Proof that $\langle c\boldsymbol{x} \mid \boldsymbol{y} \rangle = c \langle \boldsymbol{x} \mid \boldsymbol{y} \rangle$ for all $c \in \mathbb{C}$ (complex number case):
No proof at this time.

⇰

Remark N.1. [18] The inner product has already been defined in Definition N.1 (page 309) as a bilinear function that is *non-negative*, *non-isotropic*, *homogeneous*, *additive*, and *conjugate symmetric*. However, given a normed linear space, we could alternatively define the inner product using the *parallelogram law* (Theorem N.7 page 317) together with the *Polarization Identity* (Theorem N.6 page 316). Under this new definition, an inner product *exists* if the parallelogram law is satisfied, and is *specified*, in terms of the norm, by the Polarization Identity.

Of the uncountably infinite number of $\boldsymbol{\ell}^p_{\mathbb{F}}$ norms, only the norm for $p = 2$ induces an inner product (Proposition N.1, next).

Proposition N.1. [19] *Let $\left\|(x_n)_{n\in\mathbb{Z}}\right\|_p$ be the $\boldsymbol{\ell}^p_{\mathbb{F}}$ norm of the sequence (x_n) in the space $\boldsymbol{\ell}^p_{\mathbb{F}}$.*

PRP: $\left\|(x_n)\right\|_p$ *induces an inner product* $\iff$ $p = 2$

✎PROOF:

[18] Loomis (1953) pages 23–24, Kubrusly (2001) page 317

[19] Kubrusly (2001) pages 318–319 ⟨Example 5B⟩

1. Proof that $\|\cdot\|_p$ induces an inner product $\Longleftarrow$ $p = 2$ (using the *Parallelogram law* page 317):

$$\begin{aligned}
&\|\boldsymbol{x}+\boldsymbol{y}\|_p^2 + \|\boldsymbol{x}-\boldsymbol{y}\|_p^2 \\
&= \|\boldsymbol{x}+\boldsymbol{y}\|_2^2 + \|\boldsymbol{x}-\boldsymbol{y}\|_2^2 && \text{by right hypothesis}\\
&= \left(\sum_{n\in\mathbb{Z}} |x_n + y_n|^2\right)^{\frac{2}{2}} + \left(\sum_{n\in\mathbb{Z}} |x_n - y_n|^2\right)^{\frac{2}{2}} && \text{by definition of } \|\cdot\|_p\\
&= \sum_{n\in\mathbb{Z}} (x_n + y_n)(x_n + y_n)^* + \sum_{n\in\mathbb{Z}} (x_n - y_n)(x_n - y_n)^* \\
&= \sum_{n\in\mathbb{Z}} \left(|x_n|^2 + |y_n|^2 + 2\Re(x_n y_n)\right) + \sum_{n\in\mathbb{Z}} \left(|x_n|^2 + |y_n|^2 - 2\Re(x_n y_n)\right) \\
&= 2\sum_{n\in\mathbb{Z}} |x_n|^2 + 2\sum_{n\in\mathbb{Z}} |y_n|^2 \\
&= 2\,\|\boldsymbol{x}\|_2^2 + 2\,\|\boldsymbol{y}\|_2^2 && \text{by definition of } \|\cdot\|_p\\
&= 2\,\|\boldsymbol{x}\|_p^2 + 2\,\|\boldsymbol{y}\|_p^2 && \text{by right hypothesis}\\
&\Longrightarrow \|\cdot\|_2 \text{ induces an inner product} && \text{by Theorem N.7 page 317}
\end{aligned}$$

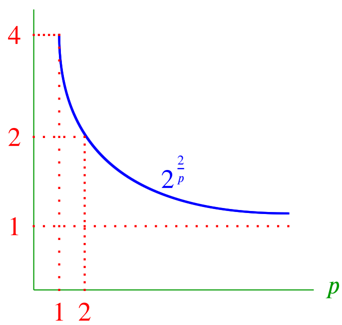
2. Proof that $\|\cdot\|_p$ induces an inner product $\Longrightarrow$ $p = 2$:

   (a) Let $\boldsymbol{x} \triangleq (1, 0)$ and $\boldsymbol{y} \triangleq (0, 1)$. Then [20]

$$\begin{aligned}
\|\boldsymbol{x}+\boldsymbol{y}\|_p^2 + \|\boldsymbol{x}-\boldsymbol{y}\|_p^2 &= \left(\sum_{n\in\mathbb{Z}} |x_n + y_n|^p\right)^{\frac{2}{p}} + \left(\sum_{n\in\mathbb{Z}} |x_n - y_n|^p\right)^{\frac{2}{p}} && \text{by definition of } \|\cdot\|_p\\
&= \left(|1+0|^p + |0+1|^p\right)^{\frac{2}{p}} + \left(|1-0|^p + |0-1|^p\right)^{\frac{2}{p}} && \text{by definitions of } \boldsymbol{x} \text{ and } \boldsymbol{y}\\
&= 2^{\frac{2}{p}} + 2^{\frac{2}{p}}\\
&= 2\cdot 2^{\frac{2}{p}}\\
2\,\|\boldsymbol{x}\|_p^2 + 2\,\|\boldsymbol{y}\|_p^2 &= 2\left(\sum_{n\in\mathbb{Z}} |x_n|^p\right)^{\frac{2}{p}} + 2\left(\sum_{n\in\mathbb{Z}} |y_n|^p\right)^{\frac{2}{p}} && \text{by definition of } \|\cdot\|_p\\
&= 2\left(|1|^p + |0|^p\right)^{\frac{2}{p}} + 2\left(|1|^p + |0|^p\right)^{\frac{2}{p}} && \text{by definitions of } \boldsymbol{x} \text{ and } \boldsymbol{y}\\
&= 2 + 2\\
&= 4\\
2\cdot 2^{\frac{2}{p}} = 4 &= \Longleftrightarrow\ 2^{\frac{2}{p}} = 2\\
&= \Longrightarrow\ p = 2
\end{aligned}$$

   (b) Proof that $2^{2/p}$ is monotonic decreasing in p (and so $p = 2$ is the only solution):

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}p} 2^{\frac{2}{p}} &= \frac{\mathrm{d}}{\mathrm{d}p} e^{\ln 2^{\frac{2}{p}}}\\
&= \left(e^{\ln 2^{\frac{2}{p}}}\right)\frac{\mathrm{d}}{\mathrm{d}p}\ln 2^{\frac{2}{p}}\\
&= \left(2^{\frac{2}{p}}\right)\frac{\mathrm{d}}{\mathrm{d}p}(2\ln 2)\frac{1}{p}\\
&= \left(2^{\frac{2}{p}}\right)2\ln 2\left(-\frac{1}{p^2}\right)\\
&< 0 \qquad \forall p \in (0, \infty)
\end{aligned}$$



[20] http://groups.google.com/group/sci.math/msg/531b1173f08871e9

Example N.1. [21]

EX

The norm $\underbrace{\|\boldsymbol{x}\| \triangleq \sum_{i=1}^{n} |x_i|}_{l_1\text{-norm or taxi cab norm}}$ **does not** induce an innerproduct.

The norm $\underbrace{\|\boldsymbol{x}\| \triangleq \left(\sum_{i=1}^{n} |x_i|^2\right)^{\frac{1}{2}}}_{l_2\text{-norm or Euclidean norm}}$ **does** induce an innerproduct.

The norm $\underbrace{\|\boldsymbol{x}\| \triangleq \max\left\{|x_i|\,\middle|\,i=1,\dots,n\right\}}_{l_\infty\text{-norm or sup norm}}$ **does not** induce an innerproduct.

✎PROOF:

1. Taxi-cab norm case (counter-example): $\boldsymbol{x} = [1,\ 1],\ \ \boldsymbol{y} = [1,\ -1]$

$$\begin{aligned}
\|\boldsymbol{x}+\boldsymbol{y}\|^2 + \|\boldsymbol{x}-\boldsymbol{y}\|^2 &= \|[1,\ 1]+[1,\ -1]\|^2 + \|[1,\ 1]-[1,\ -1]\|^2 && \text{by definition of } \boldsymbol{x} \text{ and } \boldsymbol{y}\\
&= \|[2,\ 0]\|^2 + \|[0,\ 2]\|^2\\
&= (|2|+|0|)^2 + (|0|+|2|)^2 && \text{by definition of } \|\cdot\|\\
&= 8\\
&\neq 16\\
&= 8+8\\
&= 2(|1|+|1|)^2 + 2(|1|+|-1|)^2\\
&= 2\,\|[1,\ 1]\|^2 + 2\,\|[1,\ -1]\|^2 && \text{by definition of } \|\cdot\|\\
&= 2\,\|\boldsymbol{x}\|^2 + 2\,\|\boldsymbol{y}\|^2 && \text{by definition of } \boldsymbol{x} \text{ and } \boldsymbol{y}
\end{aligned}$$

2. Euclidean norm case:

$$\begin{aligned}
\|\boldsymbol{x}+\boldsymbol{y}\|^2 + \|\boldsymbol{x}-\boldsymbol{y}\|^2 &= \left(\sum_{i=1}^{n} |x_i+y_i|^2\right) + \left(\sum_{i=1}^{n} |x_i-y_i|^2\right) && \text{by definition of } \|\cdot\|\\
&= \sum_{i=1}^{n} \left(|x_i+y_i|^2 + |x_i-y_i|^2\right)\\
&= \sum_{i=1}^{n} \left[(x_i+y_i)(x_i+y_i)^* + (x_i-y_i)(x_i-y_i)^*\right]\\
&= \sum_{i=1}^{n} \left[(x_ix_i^* + y_iy_i^* + x_iy_i^* + x_i^*y_i) + (x_ix_i^* + y_iy_i^* - x_iy_i^* - x_i^*y_i)\right]\\
&= \sum_{i=1}^{n} \left[2|x_i|^2 + 2|y_i|^2\right]\\
&= 2\sum_{i=1}^{n} |x_i|^2 + 2\sum_{i=1}^{n} |y_i|^2\\
&= 2\,\|\boldsymbol{x}\|^2 + 2\,\|\boldsymbol{y}\|^2 && \text{by definition of } \|\cdot\|
\end{aligned}$$

[21] Bachman et al. (2002) page 23

3. Sup norm case(counter-example): $\boldsymbol{x} = [1,\ 1]$, $\boldsymbol{y} = [1,\ -1]$

$$
\begin{aligned}
\|\boldsymbol{x}+\boldsymbol{y}\|^2 + \|\boldsymbol{x}-\boldsymbol{y}\|^2 &= \|[1,\ 1]+[1,\ -1]\|^2 + \|[1,\ 1]-[1,\ -1]\|^2 && \text{by definition of } \boldsymbol{x} \text{ and } \boldsymbol{y}\\
&= \|[2,\ 0]\|^2 + \|[0,\ 2]\|^2\\
&= (\max\{|2|,\ |0|\})^2 + (\max\{|0|,\ |2|\})^2 && \text{by definition of } \|\cdot\|\\
&= 2^2 + 2^2\\
&= 8\\
&\neq 4\\
&= 2\cdot 1^2 + 2\cdot 1^2\\
&= 2\,(\max\{|1|,\ |1|\})^2 + 2\,(\max\{|1|,\ |-1|\})^2\\
&= 2\,\|[1,\ 1]\|^2 + 2\,\|[1,\ -1]\|^2 && \text{by definition of } \|\cdot\|\\
&= 2\,\|\boldsymbol{x}\|^2 + 2\,\|\boldsymbol{y}\|^2 && \text{by definition of } \boldsymbol{x} \text{ and } \boldsymbol{y}
\end{aligned}
$$

✏

N.3 Orthogonality

Definition N.3.

DEF *The **Kronecker delta function** $\bar{\delta}_n$ is defined as*

$$\bar{\delta}_n \triangleq \begin{cases} 1 & \text{for } n=0 \quad \text{and} \\ 0 & \text{otherwise} \end{cases} \qquad \forall n\in\mathbb{Z}$$

Definition N.4. *Let* $\left(X,\ +,\ \cdot,\ (\mathbb{F},\ \dot{+},\ \dot{\times}),\ \langle\triangle\,|\,\triangledown\rangle\right)$ *be an* INNER PRODUCT SPACE (Definition N.1 page 309).

DEF *Two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ in X are **orthogonal** if*

$$\langle\boldsymbol{x}\,|\,\boldsymbol{y}\rangle = \begin{cases} 0 & \text{for } \boldsymbol{x}\neq\boldsymbol{y}\\ c\in\mathbb{F}\backslash 0 & \text{for } \boldsymbol{x}=\boldsymbol{y}\end{cases}$$

The notation $\boldsymbol{x}\perp\boldsymbol{y}$ implies $\boldsymbol{x}$ and $\boldsymbol{y}$ are ORTHOGONAL.
*A set $Y\in 2^X$ is **orthogonal** if $\boldsymbol{x}\perp\boldsymbol{y}$ $\quad\forall \boldsymbol{x},\boldsymbol{y}\in Y$.*
*A set Y is **orthonomal** if it is* ORTHOGONAL *and $\langle\boldsymbol{y}\,|\,\boldsymbol{y}\rangle = 1 \quad \forall \boldsymbol{y}\in Y$.*
*A sequence $(\boldsymbol{x}_n\in X)_{n\in\mathbb{Z}}$ is **orthogonal** if $\langle\boldsymbol{x}_n\,|\,\boldsymbol{x}_m\rangle = c\bar{\delta}_{nm}$ for some $c\in\mathbb{R}\backslash 0$.*
*A sequence $(\boldsymbol{x}_n\in X)_{n\in\mathbb{Z}}$ is **orthonormal** if $\langle\boldsymbol{x}_n\,|\,\boldsymbol{x}_m\rangle = \bar{\delta}_{nm}$.*

The definition of the orthogonality relation $\perp$ has several immediate consequences (next theorem):

Theorem N.8. [22] *Let* $\left(X,\ +,\ \cdot,\ (\mathbb{F},\ \dot{+},\ \dot{\times}),\ \langle\triangle\,|\,\triangledown\rangle\right)$ *be an* INNER PRODUCT SPACE.

THM

1. $\boldsymbol{x}\perp\boldsymbol{x} \iff \boldsymbol{x}=\mathbb{0}$ $\quad\forall \boldsymbol{x}\in X$
2. $\boldsymbol{x}\perp\boldsymbol{y} \implies \alpha\boldsymbol{x}\perp\boldsymbol{y}$ $\quad\forall\alpha\in\mathbb{R},\boldsymbol{x},\boldsymbol{y}\in X$ (HOMOGENEOUS)
3. $\boldsymbol{x}\perp\boldsymbol{y} \iff \boldsymbol{y}\perp\boldsymbol{x}$ $\quad\forall \boldsymbol{x},\boldsymbol{y}\in X$ (SYMMETRY)
4. $\boldsymbol{x}\perp\boldsymbol{y}$ and $\boldsymbol{y}\perp\boldsymbol{z} \implies \boldsymbol{x}\perp(\boldsymbol{y}+\boldsymbol{z})$ $\quad\forall \boldsymbol{x},\boldsymbol{y},\boldsymbol{z}\in X$ (ADDITIVE)
5. $\exists\beta\in\mathbb{R}$ such that $\boldsymbol{x}\perp(\beta\boldsymbol{x}+\boldsymbol{y})$ $\quad\forall \boldsymbol{x}\in X\backslash\mathbb{0},\ \boldsymbol{y}\in X$

Theorem N.9. *Let* $\left(X,\ +,\ \cdot,\ (\mathbb{F},\ \dot{+},\ \dot{\times}),\ \langle\triangle\,|\,\triangledown\rangle\right)$ *be an* INNER PRODUCT SPACE.

THM

$$\left.\begin{array}{lrcl} 1. & \langle\boldsymbol{x}\,|\,\boldsymbol{y}\rangle & = & 0 \quad\text{and}\\ 2. & \boldsymbol{x}+\boldsymbol{y} & = & \mathbb{0}\end{array}\right\} \iff \left\{\begin{array}{lrcl} 1. & \boldsymbol{x} & = & \mathbb{0}\quad\text{and}\\ 2. & \boldsymbol{y} & = & \mathbb{0}\end{array}\right. \qquad\forall \boldsymbol{x},\boldsymbol{y}\in X$$

✎PROOF:

[22] James (1945) page 292, Drljević (1989) page 232

1. Proof that $\boldsymbol{x} = \boldsymbol{y} = \mathbb{0}$:

$$\begin{aligned}
0 &= \langle \mathbb{0} \,|\, \mathbb{0} \rangle && \text{by } \textit{non-isotropic} \text{ property of } \langle \triangle \,|\, \triangledown \rangle \text{ (Definition N.1 page 309)}\\
&= \langle \boldsymbol{x}+\boldsymbol{y} \,|\, \boldsymbol{x}+\boldsymbol{y} \rangle && \text{by left hypothesis 2}\\
&= \langle \boldsymbol{x} \,|\, \boldsymbol{x}+\boldsymbol{y} \rangle + \langle \boldsymbol{y} \,|\, \boldsymbol{x}+\boldsymbol{y} \rangle && \text{by } \textit{additive} \text{ property of } \langle \triangle \,|\, \triangledown \rangle \text{ (Definition N.1 page 309)}\\
&= \langle \boldsymbol{x} \,|\, \boldsymbol{x} \rangle + \langle \boldsymbol{x} \,|\, \boldsymbol{y} \rangle + \langle \boldsymbol{x} \,|\, \boldsymbol{y} \rangle^* + \langle \boldsymbol{y} \,|\, \boldsymbol{y} \rangle && \text{by } \textit{conjugate symmetric} \text{ and } \textit{additive} \text{ properties of } \langle \triangle \,|\, \triangledown \rangle\\
&= \underbrace{\langle \boldsymbol{x} \,|\, \boldsymbol{x} \rangle}_{\ge 0} + 0 + 0 + \underbrace{\langle \boldsymbol{y} \,|\, \boldsymbol{y} \rangle}_{\ge 0} && \text{by left hypothesis 1}\\
\implies & \boldsymbol{x} = \mathbb{0} \text{ and } \boldsymbol{y} = \mathbb{0} && \text{by } \textit{non-negative} \text{ and } \textit{non-isotropic} \text{ properties of } \langle \triangle \,|\, \triangledown \rangle
\end{aligned}$$

2. Proof that $\langle \boldsymbol{x} \,|\, \boldsymbol{y} \rangle = 0$:

$$\begin{aligned}
\langle \boldsymbol{x} \,|\, \boldsymbol{y} \rangle &= \langle \mathbb{0} \,|\, \mathbb{0} \rangle && \text{by right hypotheses}\\
&= 0 && \text{by } \textit{non-isotropic} \text{ property of } \langle \triangle \,|\, \triangledown \rangle \text{ (Definition N.1 page 309)}
\end{aligned}$$

3. Proof that $\boldsymbol{x} + \boldsymbol{y} = \mathbb{0}$:

$$\begin{aligned}
\boldsymbol{x} + \boldsymbol{y} &= \mathbb{0} + \mathbb{0} && \text{by right hypotheses}\\
&= \mathbb{0}
\end{aligned}$$

✏

The *triangle inequality for vectors* in a *normed linear space* (Theorem O.1 page 327) demonstrates that $\left\| \sum_{n=1}^{N} \boldsymbol{x}_n \right\| \le \sum_{n=1}^{N} \left\| \boldsymbol{x}_n \right\|$. The *Pythagorean Theorem* (next) demonstrates that this *inequality* becomes *equality* when the set $\{\boldsymbol{x}_n\}$ is *orthogonal*.

Theorem N.10 (Pythagorean Theorem). [23] *Let* $\{\boldsymbol{x}_n \in X \,|\, n=1,2,\ldots,N\}$ *be a set of vectors in an* INNER PRODUCT SPACE *(Definition N.1 page 309)* $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \,|\, \triangledown \rangle)$ *and let* $\|\boldsymbol{x}\| \triangleq \sqrt{\langle \boldsymbol{x} \,|\, \boldsymbol{x} \rangle}$ *(Definition N.2 page 315).*

THM

$$\{\boldsymbol{x}_n\} \text{ is ORTHOGONAL} \iff \left\| \sum_{n=1}^{N} \boldsymbol{x}_n \right\|^2 = \sum_{n=1}^{N} \left\| \boldsymbol{x}_n \right\|^2 \qquad \forall N \in \mathbb{N}$$

✎PROOF: 1. Proof for ($\implies$) case:

$$\begin{aligned}
\left\| \sum_{n=1}^{N} \boldsymbol{x}_n \right\|^2 &= \left\langle \sum_{n=1}^{N} \boldsymbol{x}_n \,\middle|\, \sum_{m=1}^{N} \boldsymbol{x}_m \right\rangle && \text{by def. of } \|\cdot\| && \text{(Definition O.1 page 327)}\\
&= \sum_{n=1}^{N} \sum_{m=1}^{N} \langle \boldsymbol{x}_n \,|\, \boldsymbol{x}_m \rangle && \text{by def. of } \langle \triangle \,|\, \triangledown \rangle && \text{(Definition N.1 page 309)}\\
&= \sum_{n=1}^{N} \sum_{m=1}^{N} \langle \boldsymbol{x}_n \,|\, \boldsymbol{x}_m \rangle \bar{\delta}_{n-m} && \text{by left hypothesis}\\
&= \sum_{n=1}^{N} \langle \boldsymbol{x}_n \,|\, \boldsymbol{x}_n \rangle && \text{by def. of } \bar{\delta} && \text{(Definition N.3 page 323)}\\
&= \sum_{n=1}^{N} \left\| \boldsymbol{x}_n \right\|^2 && \text{by def. of } \|\cdot\| && \text{(Definition O.1 page 327)}
\end{aligned}$$

[23] Aliprantis and Burkinshaw (1998) pages 282–283 ⟨Theorem 32.7⟩, Kubrusly (2001) page 324 ⟨Proposition 5.8⟩, Bollobás (1999) pages 132–133 ⟨Theorem 3⟩

2. Proof for ($\Longleftarrow$) case:

$$
\begin{aligned}
&4\langle \boldsymbol{x} \mid \boldsymbol{y}\rangle \\
&= \|\boldsymbol{x}+\boldsymbol{y}\|^2 - \|\boldsymbol{x}-\boldsymbol{y}\|^2 + i\,\|\boldsymbol{x}+i\boldsymbol{y}\|^2 - i\,\|\boldsymbol{x}-i\boldsymbol{y}\|^2 && \text{by \textit{polarization identity} (Theorem N.6 page 316)} \\
&= \left(\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2\right) - \left(\|\boldsymbol{x}\|^2 + \|-\boldsymbol{y}\|^2\right) + i\left(\|\boldsymbol{x}\|^2 + \|i\boldsymbol{y}\|^2\right) - i\left(\|\boldsymbol{x}\|^2 + \|-i\boldsymbol{y}\|^2\right) && \text{by right hypothesis} \\
&= \left(\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2\right) - \left(\|\boldsymbol{x}\|^2 + |-1|^2\,\|\boldsymbol{y}\|^2\right) + i\left(\|\boldsymbol{x}\|^2 + |i|^2\,\|\boldsymbol{y}\|^2\right) - i\left(\|\boldsymbol{x}\|^2 + |-i|^2\,\|\boldsymbol{y}\|^2\right) && \text{by definition of } \|\cdot\| \\
&= \left(\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2\right) - \left(\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2\right) + i\left(\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2\right) - i\left(\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2\right) && \text{by def. of } |\cdot| \text{ (Definition G.4 page 222)} \\
&= 0
\end{aligned}
$$

✏

N.4 Literature

LITERATURE SURVEY:

1. Inner product spaces: Istrăţescu (1987)
2. Characterizations of inner product spaces
   (a) specific characterizations: Jordan and von Neumann (1935), James (1947a)
   (b) surveys: Day (1947), Amir (1986), Mendoza and Pakhrou (2003)

APPENDIX O

NORMED LINEAR SPACES

O.1 Definition and basic results

Definition O.1. [1] *Let* $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$ *be a* LINEAR SPACE *(Definition J.1 page 247) and* $|\cdot| \in \mathbb{R}^{\mathbb{F}}$ *the* ABSOLUTE VALUE *function (Definition G.4 page 222).*

DEF

A functional $\|\cdot\|$ *in* $\mathbb{R}^X$ *is a* ***norm*** *if*

1.	$\|x\| \geq 0$	$\forall x \in X$	(STRICTLY POSITIVE)	*and*
2.	$\|x\| = 0 \iff x = \mathbb{0}$	$\forall x \in X$	(NONDEGENERATE)	*and*
3.	$\|\alpha x\| = \|\alpha\|\,\|x\|$	$\forall x \in X,\ \alpha \in \mathbb{C}$	(HOMOGENEOUS)	*and*
4.	$\|x + y\| \leq \|x\| + \|y\|$	$\forall x, y \in X$	(SUBADDITIVE/TRIANGLE INQUALITY).	

A ***normed linear space*** *is the tuple* $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$.

The definition of the *norm* (Definition O.1 page 327) requires that any two vectors in a norm space be *subadditive* (they satisfy the *triangle inequality* property) such that $\|x_1 + x_2\| \leq \|x_1\| + \|x_2\|$. Acutally, in **any** normed linear space, this property holds true for **any** finite number of vectors—not just two—such that $\|x_1 + x_2 + \cdots + x_N\| \leq \|x_1\| + \|x_2\| + \cdots + \|x_N\|$ (next theorem).

Theorem O.1 (triangle inequality). [2] *Let* $(\!(x_n \in X)\!)_1^N$ *be an* N-TUPLE *(Definition D.2 page 200) of vectors in a* NORMED LINEAR SPACE $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$.

THM

$$\left\| \sum_{n=1}^{N} x_n \right\| \leq \sum_{n=1}^{N} \|x_n\| \qquad \forall N \in \mathbb{N},\ x_n \in V$$

✎PROOF: Proof is by induction:

[1] Aliprantis and Burkinshaw (1998) pages 217–218, Banach (1932a) page 53, Banach (1932b) page 33, Banach (1922) page 135

[2] Michel and Herget (1993) page 344, Euclid (circa 300BC) ⟨Book I Proposition 20⟩

1. Proof for the $N = 1$ case:

$$\begin{aligned}
\left\|\sum_{n=1}^{1} x_n\right\| &= \|x_1\| \\
&= \sum_{n=1}^{1} \|x_1\|
\end{aligned}$$

2. Proof for the $N = 2$ case:

$$\begin{aligned}
\left\|\sum_{n=1}^{2} x_n\right\| &= \left\|\sum_{n=1}^{2} x_n\right\| \\
&= \|x_1 + x_2\| \\
&\le \|x_1\| + \|x_2\| && \text{by Definition O.1 page 327 (triangle inequality)} \\
&= \sum_{n=1}^{2} \|x_n\|
\end{aligned}$$

3. Proof that [N case] $\implies$ [$N+1$ case]:

$$\begin{aligned}
\left\|\sum_{n=1}^{N+1} x_n\right\| &= \left\|\sum_{n=1}^{N} x_n + x_{n+1}\right\| \\
&\le \left\|\sum_{n=1}^{N} x_n\right\| + \|x_{n+1}\| && \text{by Definition O.1 page 327 (triangle inequality)} \\
&\le \sum_{n=1}^{N} \|x_n\| + \|x_{n+1}\| && \text{by left hypothesis} \\
&= \sum_{n=1}^{N+1} \|x_n\|
\end{aligned}$$

⇰

Theorem O.2 (Reverse Triangle Inequality). [3] *Let* $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$ *be a* NORMED LINEAR SPACE *(Definition O.1 page 327).*

THM

$$\underbrace{\big|\|x\| - \|y\|\big| \le \|x - y\|}_{\text{REVERSE TRIANGLE INEQUALITY}} \le \|x\| + \|y\| \qquad \forall x,y \in X$$

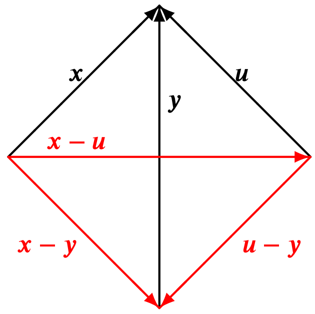
✎PROOF:

$$\begin{aligned}
\big|\|x\| - \|y\|\big| &= \big|\|(x-y) + y\| - \|y\|\big| \\
&\le \big|\|x-y\| + \|y\| - \|y\|\big| && \text{by Definition O.1 page 327} \\
&= \big|\|x-y\|\big| \\
&= \|x-y\| && \text{by Definition O.1 page 327}
\end{aligned}$$

$$\begin{aligned}
\|x-y\| &\le \|x - \mathbb{0}\| + \|\mathbb{0} - y\| && \text{by previous result with } u = 0 \\
&= \|x\| + |-1|\,\|y\| && \text{by Definition O.1 page 327} \\
&= \|x\| + \|y\|
\end{aligned}$$

⇰

[3] Aliprantis and Burkinshaw (1998) page 218, Giles (2000) page 2, Banach (1922) page 136



The shortest distance between two vectors is always the difference of the vectors. This is proven in next and illustrated to the left in the Euclidean space $\mathbb{R}^2$ (the plane)

Proposition O.1. [4] *Let* $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$ *be a* NORMED LINEAR SPACE *(Definition O.1 page 327)*.

PRP

$$\|x-y\| \le \|x-u\| + \|u-y\| \qquad \forall x,y,u \in X$$

✎PROOF:

$$\begin{aligned}\|x-y\| &= \|(x-u)+(u-y)\| \\ &\le \|x-u\| + \|u-y\| && \text{by Definition O.1 page 327}\end{aligned}$$

✎

Example O.1 (The usual norm). [5] Let $\mathbb{R}^{\mathbb{R}}$ be the set of all functions with domain and range the set of *real numbers* $\mathbb{R}$.

EX The **absolute value** (Definition G.4 page 222) $|\cdot| \in \mathbb{R}^{\mathbb{R}}$ is a *norm*.

Example O.2 (l_p norms). Let $(x_n)_{n\in\mathbb{Z}}$ be a *sequence* (Definition D.2 page 200) of real numbers. An uncountably infinite number of norms is provided by the $\ell^p_{\mathbb{F}}$ norms $\|(x_n)\|_p$:

EX

$$\|(x_n)\|_p \triangleq \left(\sum_{n\in\mathbb{Z}} |x_n|^p\right)^{\frac{1}{p}} \qquad \text{is a norm for } p \in [1:\infty]$$

O.2 Relationship between metrics and norms

O.2.1 Metrics generated by norms

The concept of *length* is very closely related to the concept of *distance*. Thus it is not surprising that a *norm* (a "length" function) can be used to define a *metric* (a "distance" function) on any *metric linear space* (Definition **??** page **??**). Another way to say this is that the norm of a normed linear space *induces* a metric on this space. And so every normed linear space also has a metric. And because every normed linear space has a metric, **every normed linear space is also a metric space**. Actually this can be generalized one step further in that every metric space is also a *topological space*. And so **every normed linear space is also a topological space.** In symbols,

$$\text{normed linear space} \implies \text{metric space} \implies \text{topological space.}$$

[4] Aliprantis and Burkinshaw (1998) page 218
[5] Giles (1987) page 3

Theorem O.3. [6] *Let* $\mathsf{d} \in \mathbb{R}^{X \times X}$ *be a function on a* REAL *normed linear space* $(X, +, \cdot, (\mathbb{R}, \dot{+}, \dot{\times}), \|\cdot\|)$. *Let* $\mathsf{B}(\boldsymbol{x}, r) \triangleq \{\boldsymbol{y} \in \boldsymbol{X} | \|\boldsymbol{y} - \boldsymbol{x}\| < r\}$ *be the* OPEN BALL *of radius* r *centered at a point* $\boldsymbol{x}$.

THM: $\mathsf{d}(\boldsymbol{x}, \boldsymbol{y}) \triangleq \|\boldsymbol{x} - \boldsymbol{y}\|$ *is a metric on* $\boldsymbol{X}$

✎PROOF: The proof follows directly from the definition of a metric (not included in this text) the definition of *norm* (Definition O.1 page 327). ✏

The previous theorem defined a metric $\mathsf{d}(\boldsymbol{x}, \boldsymbol{y})$ induced by the norm $\|\boldsymbol{x}\|$. The next definition defines this metric formally.

Definition O.2. [7] *Let* $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$ *be a* NORMED LINEAR SPACE *(Definition O.1 page 327).*

DEF: *The* ***metric induced by the norm*** $\|\cdot\|$ *is the function* $\mathsf{d} \in \mathbb{R}^{X}$ *such that*
$$\mathsf{d}(\boldsymbol{x}, \boldsymbol{y}) \triangleq \|\boldsymbol{x} - \boldsymbol{y}\| \qquad \forall x, y \in X$$

Due to its algebraic structure, every norm is *continuous* (Corollary O.1 page 330). This makes norm spaces very useful in analysis. For a function f be to *continuous*, for every $\epsilon > 0$ there must exist a $\delta > 0$ such that $|\mathsf{f}(x + \delta) - \mathsf{f}(x)| < \epsilon$. The *Reverse Triangle Inequality* (Theorem O.2 page 328) shows this to be true when $\mathsf{f}(\cdot) \triangleq \|\cdot\|$.

Corollary O.1. [8] *Let* $\boldsymbol{\Omega} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$ *be a* NORMED LINEAR SPACE *(Definition O.1 page 327).*

COR: *The norm* $\|\cdot\|$ *is* CONTINUOUS *in* $\boldsymbol{\Omega}$.

✎PROOF: This follows from these concepts:

1. The fact that $\mathsf{d}(\boldsymbol{x}, \boldsymbol{y}) \triangleq \|\boldsymbol{x} - \boldsymbol{y}\|$ is a *metric* (Theorem O.3 page 330).
2. *Continuity* in a metric space.
3. The *Reverse Triangle Inequality* (Theorem O.2 page 328).

✏

Theorem O.4 (next) demonstrates that **all open or closed** balls in **any normed linear space** are *convex*. However, the converse is not true—that is, a metric not generated by a norm may still produce a ball that is convex.

Theorem O.4. [9] *Let* $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \mathsf{d})$ *be a* METRIC LINEAR SPACE *(Definition ?? page ??). Let* B *be the* OPEN BALL $\mathsf{B}(\boldsymbol{p}, r) \triangleq \{\boldsymbol{x} \in \boldsymbol{X} | \mathsf{d}(\boldsymbol{p}, \boldsymbol{x}) < r\}$ *(open ball with respect to metric* d *centered at point* $\boldsymbol{p}$ *and with radius* r*).*

THM:
$$\left.\begin{array}{l} \exists \|\cdot\| \in \mathbb{R}^{X} \quad \text{such that} \\ \underbrace{\mathsf{d}(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{y} - \boldsymbol{x}\|}_{\mathsf{d} \text{ is generated by a norm}} \end{array}\right\} \implies \left\{\begin{array}{lll} 1. & \mathsf{B}(\boldsymbol{x}, r) = \boldsymbol{x} + B(0, r) & \\ 2. & \mathsf{B}(\mathbb{0}, r) = r\,\mathsf{B}(\mathbb{0}, 1) & \\ 3. & \mathsf{B}(\boldsymbol{x}, r) \text{ is CONVEX} & \\ 4. & \boldsymbol{x} \in \mathsf{B}(\mathbb{0}, r) \iff -\boldsymbol{x} \in \mathsf{B}(\mathbb{0}, r) & \text{(SYMMETRIC)} \end{array}\right.$$

[6] Michel and Herget (1993) page 344, Banach (1932a) page 53
[7] Giles (2000) page 1 ⟨1.1 Definition⟩
[8] Giles (2000) page 2
[9] Giles (2000) page 2 ⟨1.2 Remarks⟩, Giles (1987) pages 22–26 ⟨2.4 Theorem, 2.11 Theorem⟩

✎PROOF:

1. Proof that $\mathsf{d}(\boldsymbol{x}+\boldsymbol{z}, \boldsymbol{y}+v\boldsymbol{z}) = \mathsf{d}(\boldsymbol{x}, \boldsymbol{y})$ (invariant):

$$\begin{aligned}
\mathsf{d}(\boldsymbol{x}+\boldsymbol{z}, \boldsymbol{y}+v\boldsymbol{z}) &= \|(\boldsymbol{y}+v\boldsymbol{z}) - (\boldsymbol{x}+\boldsymbol{z})\| && \text{by left hypothesis}\\
&= \|\boldsymbol{y}-\boldsymbol{x}\| \\
&= \mathsf{d}(\boldsymbol{x}, \boldsymbol{y}) && \text{by left hypothesis}
\end{aligned}$$

2. Proof that $\mathsf{B}(\boldsymbol{x}, r) = \boldsymbol{x} + B(0, r)$:

$$\begin{aligned}
\mathsf{B}(\boldsymbol{x}, r) &= \{\boldsymbol{y} \in \boldsymbol{X} | \mathsf{d}(\boldsymbol{x}, \boldsymbol{y}) < r\} && \text{by definition of open ball } \boldsymbol{B}\\
&= \{\boldsymbol{y} \in \boldsymbol{X} | \mathsf{d}(\boldsymbol{y}-\boldsymbol{y}, \boldsymbol{y}-\boldsymbol{x}) < r\} && \text{by right result 1.}\\
&= \{\boldsymbol{y} \in \boldsymbol{X} | \mathsf{d}(\mathbb{0}, \boldsymbol{y}-\boldsymbol{x}) < r\} \\
&= \{\boldsymbol{u}+\boldsymbol{x} \in \boldsymbol{X} | \mathsf{d}(\mathbb{0}, \boldsymbol{u}) < r\} && \text{let } \boldsymbol{u} \triangleq \boldsymbol{y}-\boldsymbol{x}\\
&= \boldsymbol{x} + \{\boldsymbol{u} \in \boldsymbol{X} | \mathsf{d}(\mathbb{0}, \boldsymbol{u}) < r\} \\
&= \boldsymbol{x} + B(0, r) && \text{by definition of open ball } \boldsymbol{B}
\end{aligned}$$

3. Proof that $\mathsf{B}(\mathbb{0}, r) = r\,\mathsf{B}(\mathbb{0}, 1)$:

$$\begin{aligned}
\mathsf{B}(\mathbb{0}, r) &= \{\boldsymbol{y} \in \boldsymbol{X} | \mathsf{d}(\mathbb{0}, \boldsymbol{y}) < r\} && \text{by definition of open ball } \boldsymbol{B}\\
&= \left\{\boldsymbol{y} \in \boldsymbol{X} | \frac{1}{r}\mathsf{d}(\mathbb{0}, \boldsymbol{y}) < 1\right\} \\
&= \left\{\boldsymbol{y} \in \boldsymbol{X} | \frac{1}{r}\|\boldsymbol{y}-\mathbb{0}\| < 1\right\} && \text{by left hypothesis}\\
&= \left\{\boldsymbol{y} \in \boldsymbol{X} | \left\|\frac{1}{r}\boldsymbol{y}-\frac{1}{r}\mathbb{0}\right\| < 1\right\} && \text{by homogeneous property of } \|\cdot\| \text{ page 327}\\
&= \left\{\boldsymbol{y} \in \boldsymbol{X} | \mathsf{d}\left(\frac{1}{r}\mathbb{0}, \frac{1}{r}\boldsymbol{y}\right) < 1\right\} && \text{by left hypothesis}\\
&= \{r\boldsymbol{u} \in \boldsymbol{X} | \mathsf{d}(\mathbb{0}, \boldsymbol{u}) < 1\} && \text{let } \boldsymbol{u} \triangleq \frac{1}{r}\boldsymbol{y}\\
&= r\,\{\boldsymbol{u} \in \boldsymbol{X} | \mathsf{d}(\mathbb{0}, \boldsymbol{u}) < 1\} \\
&= r\,B(\mathbb{0}, 1) && \text{by definition of open ball } \boldsymbol{B}
\end{aligned}$$

4. Proof that $\mathsf{B}(\boldsymbol{p}, r)$ is convex:
We must prove that for any pair of points $\boldsymbol{x}$ and $\boldsymbol{y}$ in the open ball $\mathsf{B}(\boldsymbol{p}, r)$, any point $\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y}$ is also in the open ball. That is, the distance from any point $\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y}$ to the ball's center $\boldsymbol{p}$ must be less than r.

$$\begin{aligned}
\mathsf{d}(\boldsymbol{p}, \lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y}) &= \|\boldsymbol{p} - \lambda\boldsymbol{x} - (1-\lambda)\boldsymbol{y}\| && \text{by left hypothesis}\\
&= \left\|\underbrace{\lambda\boldsymbol{p} + (1-\lambda)\boldsymbol{p}}_{\boldsymbol{p}} - \lambda\boldsymbol{x} - (1-\lambda)\boldsymbol{y}\right\| \\
&= \|\lambda\boldsymbol{p} - \lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{p} - (1-\lambda)\boldsymbol{y}\| \\
&\le \|\lambda\boldsymbol{p} - \lambda\boldsymbol{x}\| + \|(1-\lambda)\boldsymbol{p} - (1-\lambda)\boldsymbol{y}\| && \text{by subadditivity property of } \|\cdot\| \text{ page 327}\\
&= |\lambda|\,\|\boldsymbol{p}-\boldsymbol{x}\| + |1-\lambda|\,\|\boldsymbol{p}-\boldsymbol{y}\| && \text{by homogeneous property of } \|\cdot\| \text{ page 327}\\
&= \lambda\,\|\boldsymbol{p}-\boldsymbol{x}\| + (1-\lambda)\,\|\boldsymbol{p}-\boldsymbol{y}\| && \text{because } 0 \le \lambda \le 1\\
&\le \lambda r + (1-\lambda)r && \text{because } \boldsymbol{x}, \boldsymbol{y} \text{ are in the ball } \mathsf{B}(\boldsymbol{p}, r)\\
&= r
\end{aligned}$$

5. Proof that $\boldsymbol{x} \in \mathsf{B}(\mathbb{0}, r) \iff -\boldsymbol{x} \in \mathsf{B}(\mathbb{0}, r)$ (symmetric):

$$\begin{aligned}
\boldsymbol{x} \in \mathsf{B}(\mathbb{0}, r) &\iff \boldsymbol{x} \in \{\boldsymbol{y} \in \boldsymbol{X} | \mathsf{d}(\mathbb{0}, \boldsymbol{y}) < r\} && \text{by definition of open ball } B \\
&\iff \boldsymbol{x} \in \{\boldsymbol{y} \in \boldsymbol{X} | \, \|\boldsymbol{y} - \mathbb{0}\| < r\} && \text{by left hypothesis} \\
&\iff \boldsymbol{x} \in \{\boldsymbol{y} \in \boldsymbol{X} | \, \|\boldsymbol{y}\| < r\} \\
&\iff \boldsymbol{x} \in \{\boldsymbol{y} \in \boldsymbol{X} | \, \|(-1)(-\boldsymbol{y})\| < r\} \\
&\iff \boldsymbol{x} \in \{\boldsymbol{y} \in \boldsymbol{X} | |-1| \, \|-\boldsymbol{y}\| < r\} && \text{by homogeneous property of } \|\cdot\| \text{ page 327} \\
&\iff \boldsymbol{x} \in \{\boldsymbol{y} \in \boldsymbol{X} | \, \|-\boldsymbol{y} - \mathbb{0}\| < r\} \\
&\iff \boldsymbol{x} \in \{\boldsymbol{y} \in \boldsymbol{X} | \mathsf{d}(\mathbb{0}, -\boldsymbol{y}) < r\} && \text{by left hypothesis} \\
&\iff \boldsymbol{x} \in \{-\boldsymbol{u} \in \boldsymbol{X} | \mathsf{d}(\mathbb{0}, \boldsymbol{u}) < r\} && \text{let } \boldsymbol{u} \triangleq -\boldsymbol{y} \\
&\iff \boldsymbol{x} \in (-\{\boldsymbol{u} \in \boldsymbol{X} | \mathsf{d}(\mathbb{0}, \boldsymbol{u}) < r\}) \\
&\iff \boldsymbol{x} \in (-\mathsf{B}(\mathbb{0}, r)) \\
&\iff -\boldsymbol{x} \in \mathsf{B}(\mathbb{0}, r)
\end{aligned}$$

⇒

Theorem O.4 (page 330) demonstrates that if a metric d in a metric space $\left(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \mathsf{d}\right)$ is generated by a norm, then the ball $B(x, r)$ in that metric linear space is *convex*. However, the converse is not true. That is, it is possible for the balls in a metric space $(\boldsymbol{Y}, \mathsf{p})$ to be *convex*, but yet the metric p not be generated by a norm.

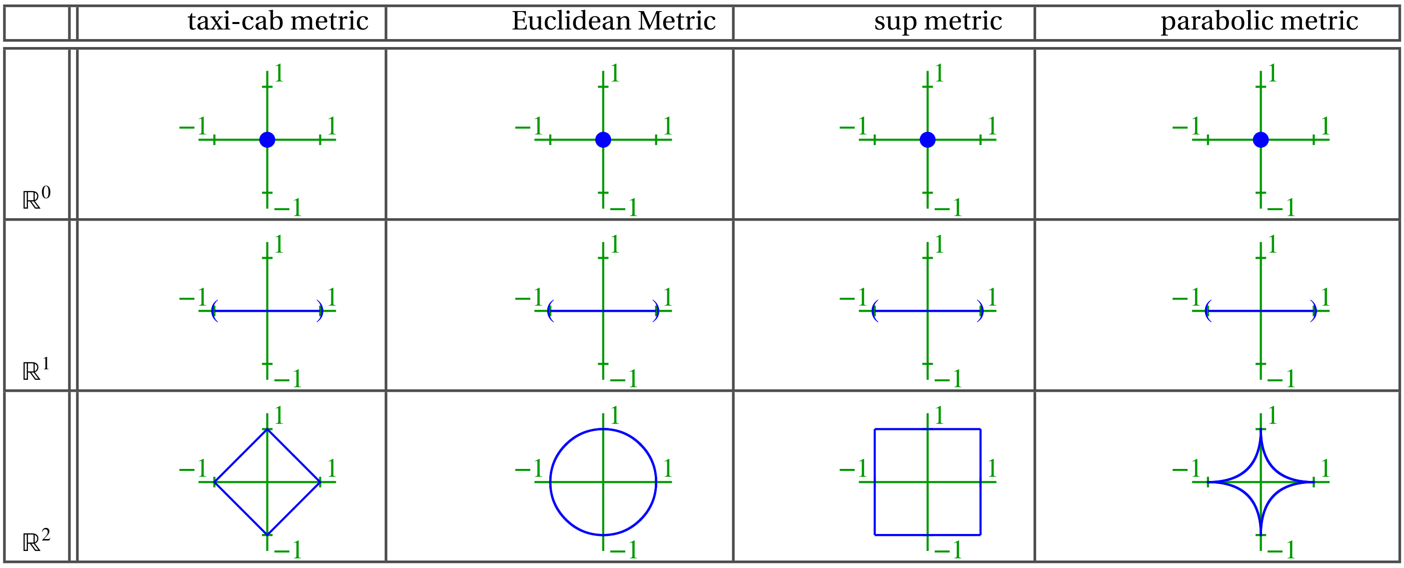


Figure O.1: Open balls in $(\mathbb{R}^0, \mathsf{d}_n)$, $(\mathbb{R}, \mathsf{d}_n)$, $(\mathbb{R}^2, \mathsf{d}_n)$, and $(\mathbb{R}^3, \mathsf{d}_n)$.

O.2.2 Norms generated by metrics

Every normed linear space is also a metric linear space (Theorem O.3 page 330). That is, a metric linear space generates a *normed linear space*. However, the converse is not true—not every metric linear space is a *normed linear space*. A characterization of metric linear spaces that *are* normed linear spaces is given by Theorem O.5 page 333.

Lemma O.1. [10] *Let* $\left(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \mathsf{d}\right)$ *be a* METRIC LINEAR SPACE. *Let* $\|\boldsymbol{x}\| \triangleq \mathsf{d}(\boldsymbol{x}, \mathbb{0}) \; \forall \boldsymbol{x} \in X$.

[10] Oikhberg and Rosenthal (2007) page 599

LEM
$$\underbrace{\mathsf{d}(\boldsymbol{x}+\boldsymbol{z},\boldsymbol{y}+\boldsymbol{z}) = \mathsf{d}(\boldsymbol{x},\boldsymbol{y}) \quad \forall \boldsymbol{x},\boldsymbol{y},\boldsymbol{z}\in X}_{\text{TRANSLATION INVARIANT}} \implies \left\{\begin{array}{lrcll} 1. & \|\boldsymbol{x}\| & = & \|-\boldsymbol{x}\| & \forall \boldsymbol{x}\in X \quad \text{and} \\ 2. & \|\boldsymbol{x}\| & = & 0 \iff \boldsymbol{x}=0 & \forall \boldsymbol{x}\in X \quad \text{and} \\ 3. & \|\boldsymbol{x}+\boldsymbol{y}\| & \le & \|\boldsymbol{x}\|+\|\boldsymbol{y}\| & \forall \boldsymbol{x},\boldsymbol{y}\in X \end{array}\right.$$

✎PROOF:

1. Proof that $\|\boldsymbol{x}\| = \|-\boldsymbol{x}\|$:

$$\begin{aligned} \|\boldsymbol{x}\| &= \mathsf{d}(\boldsymbol{x},\mathbb{0}) && \text{by definition of } \|\cdot\| \\ &= \mathsf{d}(\boldsymbol{x}-\boldsymbol{x},\mathbb{0}-\boldsymbol{x}) && \text{by translation invariance hypothesis} \\ &= \mathsf{d}(\mathbb{0},-\boldsymbol{x}) \\ &= \|-\boldsymbol{x}\| && \text{by definition of } \|\cdot\| \end{aligned}$$

2a. Proof that $\|\boldsymbol{x}\| = 0 \implies \boldsymbol{x} = 0$:

$$\begin{aligned} 0 &= \|\boldsymbol{x}\| && \text{by left hypothesis} \\ &= \mathsf{d}(\boldsymbol{x},\mathbb{0}) && \text{by definition of } \|\cdot\| \\ &= \mathsf{d}(\boldsymbol{x},\mathbb{0}) && \text{by definition of } \|\cdot\| \\ &\implies \boldsymbol{x} = \mathbb{0} && \text{by property of metrics} \end{aligned}$$

2b. Proof that $\|\boldsymbol{x}\| = 0 \impliedby \boldsymbol{x} = 0$:

$$\begin{aligned} \|\boldsymbol{x}\| &= \mathsf{d}(\boldsymbol{x},\mathbb{0}) && \text{by definition of } \|\cdot\| \\ &= \mathsf{d}(\mathbb{0},\mathbb{0}) && \text{by right hypothesis} \\ &= 0 && \text{by property of metrics} \end{aligned}$$

3. Proof that $\|\boldsymbol{x}+\boldsymbol{y}\| \le \|\boldsymbol{x}\|+\|\boldsymbol{y}\|$:

$$\begin{aligned} \|\boldsymbol{x}+\boldsymbol{y}\| &= \mathsf{d}(\boldsymbol{x}+\boldsymbol{y},\mathbb{0}) && \text{by definition of } \|\cdot\| \\ &= \mathsf{d}(\boldsymbol{x}+\boldsymbol{y}-\boldsymbol{y},\mathbb{0}-\boldsymbol{y}) && \text{by translation invariance hypothesis} \\ &= \mathsf{d}(\boldsymbol{x},-\boldsymbol{y}) \\ &\le \mathsf{d}(\boldsymbol{x},\mathbb{0})+\mathsf{d}(\mathbb{0},\boldsymbol{y}) && \text{by property of metrics} \\ &= \mathsf{d}(\boldsymbol{x},\mathbb{0})+\mathsf{d}(\boldsymbol{y},\mathbb{0}) && \text{by property of metrics} \\ &= \|\boldsymbol{x}\|+\|\boldsymbol{y}\| && \text{by definition of } \|\cdot\| \end{aligned}$$

✏

Theorem O.5. [11] *Let* $\left(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times})\right)$ *be a* LINEAR SPACE. *Let* $\mathsf{d}(\boldsymbol{x},\boldsymbol{y}) \triangleq \|\boldsymbol{x}-\boldsymbol{y}\| \; \forall \boldsymbol{x},\boldsymbol{y}\in X$.

THM
$$\left.\begin{array}{lllll} 1. & \mathsf{d}(\boldsymbol{x}+\boldsymbol{z},\boldsymbol{y}+\boldsymbol{z}) = \mathsf{d}(\boldsymbol{x},\boldsymbol{y}) & \forall \boldsymbol{x},\boldsymbol{y},\boldsymbol{z}\in X & (\text{TRANSLATION INVARIANT}) & \text{and} \\ 2. & \mathsf{d}(\alpha\boldsymbol{x},\alpha\boldsymbol{y}) = |\alpha|\mathsf{d}(\boldsymbol{x},\boldsymbol{y}) & \forall \boldsymbol{x},\boldsymbol{y}\in X, \alpha\in\mathbb{F} & (\text{HOMOGENEOUS}) & \end{array}\right\} \iff \|\cdot\| \textit{ is a } \text{NORM}$$

✎PROOF:

1. Proof of $\implies$ assertion:

   (a) Proof that $\|\cdot\|$ is *strictly positive*: This follows directly from the definition of d.

   (b) Proof that $\|\cdot\|$ is *nondegenerate*: This follows directly from Lemma O.1 (page 332).

   (c) Proof that $\|\cdot\|$ is *homogeneous*: This follows from the second left hypothesis.

[11] Bollobás (1999) page 21

(d) Proof that $\|\cdot\|$ satisfies the *triangle-inequality*: This follows directly from Lemma O.1 (page 332).

2. Proof of $\Longleftarrow$ assertion:

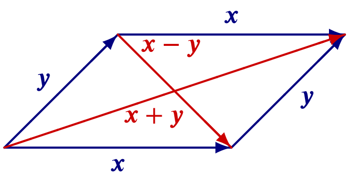
$$\begin{aligned}
\mathsf{d}(\boldsymbol{x}+\boldsymbol{z}, \boldsymbol{y}+\boldsymbol{z}) &= \|(\boldsymbol{x}+\boldsymbol{z}) - (\boldsymbol{y}+\boldsymbol{z})\| && \text{by definition of } \mathsf{d} \\
&= \|\boldsymbol{x}-\boldsymbol{y}\| \\
&= \mathsf{d}(\boldsymbol{x}, \boldsymbol{y}) && \text{by definition of } \mathsf{d} \\
\mathsf{d}(\alpha\boldsymbol{x}, \alpha\boldsymbol{y}) &= \|(\alpha\boldsymbol{x}) - (\alpha\boldsymbol{y})\| && \text{by definition of } \mathsf{d} \\
&= \|\alpha(\boldsymbol{x}-\boldsymbol{y})\| \\
&= |\alpha|\ \|\boldsymbol{x}-\boldsymbol{y}\| && \text{by definition of } \|\cdot\| \text{ page 327} \\
&= |\alpha|\mathsf{d}(\boldsymbol{x}, \boldsymbol{y}) && \text{by definition of } \mathsf{d}
\end{aligned}$$

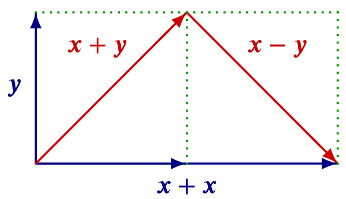
⇰

O.3 Orthogonality on normed linear spaces

Traditionally, *orthogonality* (Definition N.4 page 323) is a property defined in *inner product spaces* (Definition N.1 page 309) However, the concept of orthogonality can be extended to *normed linear spaces* (Definition O.1 page 327). Here are some examples:

① *Isosceles orthogonality*: Definition O.3 page 334
② *Pythagorean orthogonality*: Definition O.4 page 336
③ *Birkhoff orthogonality*: Definition O.5 page 336





Isosceles orthogonality (Definition O.3 page 334) can be illustrated using a *parallelogram*, as illustrated in the figure to the upper left. In this case, orthogonality implies that the parallelogram is a rectangle, which in turn implies that the lengths of the two diagonals are equal ($\|\boldsymbol{x}+\boldsymbol{y}\| = \|\boldsymbol{x}-\boldsymbol{y}\|$). Isosceles orthogonality can also be illustrated with a triangle where the sides are of lengths $\|\boldsymbol{x}+\boldsymbol{y}\|$ and $\|\boldsymbol{x}-\boldsymbol{y}\|$ and base of length $\|\boldsymbol{x}+\boldsymbol{x}\|$. In this case if $\boldsymbol{x}$ and $\boldsymbol{y}$ are orthogonal, then the triangle is *isosceles*. This is illustrated in figure to the lower left. Isosceles orthogonality is formally defined next.

Definition O.3. [12] *Let* $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$ *be a* NORMED LINEAR SPACE *(Definition O.1 page 327).*

DEF

Two vectors $\boldsymbol{x}$ *and* $\boldsymbol{y}$ *are* ***orthogonal in the sense of James*** *if*

$$\|\boldsymbol{x}+\boldsymbol{y}\| = \|\boldsymbol{x}-\boldsymbol{y}\|.$$

This property is also called ***isosceles orthogonality*** *or* ***James orthogonality****.*

Theorem O.6. *Let* $(X, +, \cdot, (\mathbb{R}, \dot{+}, \dot{\times}), \langle \triangle \mid \triangledown \rangle)$ *be an* INNER PRODUCT SPACE *(Definition N.1 page 309) with induced norm* $\|\boldsymbol{x}\| \triangleq \sqrt{\langle \boldsymbol{x} \mid \boldsymbol{x} \rangle}$, ISOSCELES ORTHOGONALITY *(Definition O.3 page 334) relation* ⦶*, and inner product relation* ORTHOGONALITY *(Definition N.4 page 323) relation* $\perp$.

THM

$$\underbrace{\boldsymbol{x} \,⦶\, \boldsymbol{y}}_{\text{orthogonal in the sense of James}} \iff \underbrace{\boldsymbol{x} \perp \boldsymbol{y}}_{\text{orthogonal in the sense of inner product space}}$$

[12] James (1945) page 292 ⟨DEFINITION 2.1⟩, Amir (1986) page 24, Dunford and Schwartz (1957) page 93

✎PROOF:

1. Proof that $\boldsymbol{x} \obot \boldsymbol{y} \quad\Longrightarrow\quad \boldsymbol{x} \perp \boldsymbol{y}$:

$$\begin{aligned}
&4\langle \boldsymbol{x} \mid \boldsymbol{y}\rangle \\
&= \underbrace{\|\boldsymbol{x}+\boldsymbol{y}\|^2 - \|\boldsymbol{x}-\boldsymbol{y}\|^2}_{0 \text{ by } \boldsymbol{x} \obot \boldsymbol{y} \text{ hypothesis}} + i\,\|\boldsymbol{x}+i\boldsymbol{y}\|^2 - i\,\|\boldsymbol{x}-i\boldsymbol{y}\|^2 && \text{by \textit{polarization identity} (Theorem N.6 page 316)} \\
&= 0 + i\,\|\boldsymbol{x}+i\boldsymbol{y}\|^2 - i\,\|\boldsymbol{x}-i\boldsymbol{y}\|^2 && \text{by } \boldsymbol{x} \obot \boldsymbol{y} \text{ hypothesis} \\
&= i\left[\|\boldsymbol{x}\|^2 + \|i\boldsymbol{y}\|^2 + 2\Re\langle \boldsymbol{x} \mid i\boldsymbol{y}\rangle\right] \\
&\quad - i\left[\|\boldsymbol{x}\|^2 + \|-i\boldsymbol{y}\|^2 + 2\Re\langle \boldsymbol{x} \mid -i\boldsymbol{y}\rangle\right] && \text{by \textit{Polar Identity} (Lemma N.1 page 314)} \\
&= i\left[\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 + 2\Re\langle \boldsymbol{x} \mid i\boldsymbol{y}\rangle\right] \\
&\quad - i\left[\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 + 2\Re\langle \boldsymbol{x} \mid -i\boldsymbol{y}\rangle\right] && \text{by Definition O.1 page 327 and Definition N.1 page 309} \\
&= 4i\Re\langle \boldsymbol{x} \mid i\boldsymbol{y}\rangle \\
&= 4i\Re\left[i^*\langle \boldsymbol{x} \mid \boldsymbol{y}\rangle\right] \\
&= 0 && \text{because inner product space is \textit{real} } (\mathbb{F} = \mathbb{R})
\end{aligned}$$

2. Proof that $\boldsymbol{x} \obot \boldsymbol{y} \quad\Longleftarrow\quad \boldsymbol{x} \perp \boldsymbol{y}$:

$$\begin{aligned}
\|\boldsymbol{x}+\boldsymbol{y}\|^2 &= \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 + 2\Re\langle \boldsymbol{x} \mid \boldsymbol{y}\rangle && \text{by \textit{Polar Identity} (Lemma N.1 page 314)} \\
&= \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 + 0 && \text{by } \boldsymbol{x} \perp \boldsymbol{y} \text{ hypothesis} \\
&= \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 - 2\Re\cancel{\langle \boldsymbol{x} \mid \boldsymbol{y}\rangle} \rightarrow 0 \text{ when } \boldsymbol{x} \perp \boldsymbol{y} && \text{by } \boldsymbol{x} \perp \boldsymbol{y} \text{ hypothesis} \\
&= \|\boldsymbol{x}\|^2 + \|-\boldsymbol{y}\|^2 + 2\Re\langle \boldsymbol{x} \mid -\boldsymbol{y}\rangle \\
&= \|\boldsymbol{x}-\boldsymbol{y}\|^2 && \text{by \textit{Polar Identity} (Lemma N.1 page 314)}
\end{aligned}$$

✏

Theorem O.7. [13] *Let* $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$ *be a normed linear space and with* ISOSCELES ORTHOGONALITY *(Definition O.3 page 334) relation* $\obot$.

THM
$$\boldsymbol{x} \obot \boldsymbol{y} \quad\Longleftrightarrow\quad \boldsymbol{y} \obot \boldsymbol{x} \quad\Longleftrightarrow\quad \alpha\boldsymbol{x} \obot \alpha\boldsymbol{y} \qquad \forall\alpha\in\mathbb{F}$$

✎PROOF:

$$\begin{aligned}
\boldsymbol{x} \obot \boldsymbol{y} &\Longrightarrow \|\boldsymbol{x}+\boldsymbol{y}\| &&= \|\boldsymbol{x}-\boldsymbol{y}\| && \text{by Definition O.3 page 334} \\
&\Longrightarrow \|\boldsymbol{x}+\boldsymbol{y}\| &&= |-1|\;\|\boldsymbol{x}-\boldsymbol{y}\| \\
&\Longrightarrow \|\boldsymbol{x}+\boldsymbol{y}\| &&= \|-(\boldsymbol{x}-\boldsymbol{y})\| && \text{by Definition O.1 page 327} \\
&\Longrightarrow \|\boldsymbol{y}+\boldsymbol{x}\| &&= \|\boldsymbol{y}-\boldsymbol{x}\| && \text{by Definition J.1 page 247} \\
&\Longrightarrow \boldsymbol{y} \obot \boldsymbol{x} &&&& \text{by Definition O.3 page 334}
\end{aligned}$$

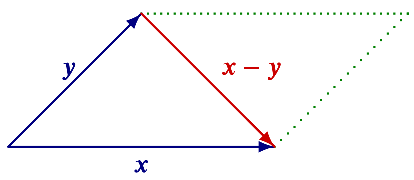
$$\begin{aligned}
\boldsymbol{y} \obot \boldsymbol{x} &\Longrightarrow \|\boldsymbol{y}+\boldsymbol{x}\| &&= \|\boldsymbol{y}-\boldsymbol{x}\| && \text{by Definition O.3 page 334} \\
&\Longrightarrow |\alpha|\;\|\boldsymbol{y}+\boldsymbol{x}\| &&= |\alpha|\;\|\boldsymbol{y}-\boldsymbol{x}\| \\
&\Longrightarrow \|\alpha(\boldsymbol{y}+\boldsymbol{x})\| &&= \|\alpha(\boldsymbol{y}-\boldsymbol{x})\| && \text{by Definition O.1 page 327} \\
&\Longrightarrow \|\alpha\boldsymbol{y}+\alpha\boldsymbol{x}\| &&= \|\alpha\boldsymbol{y}-\alpha\boldsymbol{x}\| \\
&\Longrightarrow \|\alpha\boldsymbol{x}+\alpha\boldsymbol{y}\| &&= \|-(\alpha\boldsymbol{x}-\alpha\boldsymbol{y})\| && \text{by Definition J.1 page 247}
\end{aligned}$$

[13] Amir (1986) page 24

$$\begin{aligned}
&\implies \|\alpha x + \alpha y\| &&= |-1|\ \|\alpha x - \alpha y\| && \text{by Definition O.1 page 327}\\
&\implies \|\alpha x + \alpha y\| &&= \|\alpha x - \alpha y\| && \text{by Definition G.4 page 222}\\
&\implies \alpha x \textcircled{\scriptsize I} \alpha y && && \text{by Definition O.3 page 334}
\end{aligned}$$

$$\begin{aligned}
\alpha x \textcircled{\scriptsize I} \alpha y &\implies \|\alpha x + \alpha y\| &&= \|\alpha x - \alpha y\| && \text{by Definition O.3 page 334}\\
&\implies \|\alpha(x + y)\| &&= \|\alpha(x - y)\| && \text{by Definition J.1 page 247}\\
&\implies |\alpha|\ \|x + y\| &&= |\alpha|\ \|x - y\| && \text{by Definition O.1 page 327}\\
&\implies \|x + y\| &&= \|x - y\| && \text{by Definition O.1 page 327}\\
&\implies x \textcircled{\scriptsize I} y && && \text{by Definition O.3 page 334}
\end{aligned}$$

⇨



If a triangle in a plane has two perpendicular sides of lengths a and b and a hypotenuse of length c, then by the *Pythagorean Theorem* (Theorem N.10 page 324), $a^2 + b^2 = c^2$. This concept of orthogonality can be generalized to normed linear spaces. Two vectors x and y (with lengths $\|x\|$ and $\|y\|$) are orthogonal when $\|x\|^2 + \|y\|^2 = \|x - y\|^2$ ($x - y$ is a kind of "*hypotenuse*"). This kind of orthogonality is defined next and illustrated in the figure to the left.

Definition O.4. [14] *Let* $(X, +, \cdot, (\mathbb{F}, \dot+, \dot\times), \|\cdot\|)$ *be a* NORMED LINEAR SPACE *(Definition O.1 page 327).*

DEF

Two vectors x *and* y *are* ***orthogonal in the Pythagorean sense*** *if*

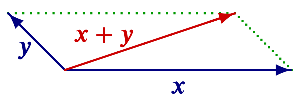
$$\|x - y\|^2 = \|x\|^2 + \|y\|^2.$$

This relationship is also called ***Pythagorean orthogonality****.*

Theorem O.8. [15] *Let* $(X, +, \cdot, (\mathbb{F}, \dot+, \dot\times), \langle\triangle\,|\,\nabla\rangle)$ *be an* INNER PRODUCT SPACE *(Definition N.1 page 309) with induced norm* $\|x\| \triangleq \sqrt{\langle x\,|\,x\rangle}$, PYTHAGOREAN ORTHOGONALITY *(Definition O.4 page 336) relation* ⓘ, *and inner product relation* ORTHOGONALITY *(Definition N.4 page 323) relation* $\perp$.

THM

$$\underbrace{x \textcircled{\scriptsize I} y}_{\text{orthogonal in the Pythagorean sense}} \iff \underbrace{x \perp y}_{\text{orthogonal in the sense of inner product space}}$$



Besides *isosceles orthogonality* (Definition O.3 page 334), orthogonality in normed linear spaces can be defined using *Birkhoff orthogonality*, as defined in Definition O.5 (next) and illustrated to the left.

Definition O.5. [16] *Let* $(X, +, \cdot, (\mathbb{F}, \dot+, \dot\times), \|\cdot\|)$ *be a* NORMED LINEAR SPACE *(Definition O.1 page 327).*

DEF

Two vectors x *and* y *are* ***orthogonal in the sense of Birkhoff*** *if*

$$\|x\| \le \|x + \alpha y\| \qquad \forall \alpha \in \mathbb{F}.$$

This relationship is also called ***Birkhoff orthogonality****.*

Theorem O.9. *Let* $(X, +, \cdot, (\mathbb{F}, \dot+, \dot\times), \langle\triangle\,|\,\nabla\rangle)$ *be an* INNER PRODUCT SPACE *(Definition N.1 page 309) with induced norm* $\|x\| \triangleq \sqrt{\langle x\,|\,x\rangle}$, BIRKHOFF ORTHOGONALITY *relation* ⓘ *(Definition O.5 page 336), and inner product relation* ORTHOGONALITY *relation* $\perp$ *(Definition N.4 page 323).*

THM

$$\underbrace{x \textcircled{\scriptsize I} y}_{\text{orthogonal in the sense of Birkhoff}} \iff \underbrace{x \perp y}_{\text{orthogonal in the sense of inner product space}}$$

---

[14] James (1945) page 292 ⟨DEFINITION 2.2⟩, Amir (1986) page 57, Drljević (1989) page 232
[15] Amir (1986) page 57
[16] Amir (1986) page 33, Dunford and Schwartz (1957) page 93, James (1947b) page 265

APPENDIX P

INTERVALS AND CONVEXITY

P.1 Intervals

In the real number system, for $a \le b$, the *interval* $[a : b]$ is the set a and b and all the numbers inbetween, as in $[a : b] \triangleq \{x \in \mathbb{R} | a \le x \le b\}$. This concept can be easily generalized:

- In an **ordered set**, if two elements x and y are *comparable* and $x \le y$, then we say that x and y and all the elements inbetween, as determined by the ordering relation $\le$, are the interval $[a : b]$.
- In a **lattice**, the concept of the *interval* can be generalized even further. In an arbitrary ordered set, the interval $[x : y]$ of item (P.1) is restricted to the case in which x and y are *comparable*. This restriction can be lifted (Definition P.2 page 337) with the additional structure of upper and lower bounds provided by lattices.
- A **metric space** in general has no *order relation* $\le$. But intervals can still be defined (Definition P.4 page 338) in a metric space in terms of the *triangle inequality*.
- A **linear space** (Definition J.1 page 247) over a real or complex field in general has no *order relation* that compares *vectors* in the space, but the standard order relation $\le$ for real numbers $\mathbb{R}$ can still be used (Definition P.5 page 338) to define an interval in a linear space.

Definition P.1 (intervals on ordered sets). [1] *Let* $(X, \le)$ *be an* ORDERED SET.

DEF	*The set* $[x : y]$	$\triangleq \{z \in X \mid x \le z \le y\}$	*is called a* ***closed interval***	*and*
	The set $(x : y]$	$\triangleq \{z \in X \mid x < z \le y\}$	*is called a* ***half-open interval***	*and*
	The set $[x : y)$	$\triangleq \{z \in X \mid x \le z < y\}$	*is called a* ***half-open interval***	*and*
	The set $(x : y)$	$\triangleq \{z \in X \mid x < z < y\}$	*is called an* ***open interval***.	

Definition P.2 (intervals on lattices). [2] *Let* $(X, \vee, \wedge; \le)$ *be a* LATTICE.

DEF	*The set* $[x : y]$	$\triangleq \{z \in X \mid x \wedge y \le z \le x \vee y\}$	*is called a* ***closed interval***.
	The set $(x : y]$	$\triangleq \{z \in X \mid x \wedge y < z \le x \vee y\}$	*is called a* ***half-open interval***.
	The set $[x : y)$	$\triangleq \{z \in X \mid x \wedge y \le z < x \vee y\}$	*is called a* ***half-open interval***.
	The set $(x : y)$	$\triangleq \{z \in X \mid x \wedge y < z < x \vee y\}$	*is called an* ***open interval***.

When x and y *are* comparable and $x \le y$, then Definition P.2 (previous) simplifies to item (P.1) (page

[1] Apostol (1975) page 4, Ore (1935) page 409
[2] Duthie (1942) page 2, Ore (1935) page 425 ⟨*quotient structures*⟩

337).

Definition P.3. [3] *Let* $\mathbf{L} \triangleq (X, \vee, \wedge; \leq)$ *be a* LATTICE *with dual* $\mathbf{L}^*$. *Let* $[x : y]$ *be a* CLOSED INTERVAL (Definition P.2 page 337) *on set* X. *The sublattices* $\mathbf{L}[x : y]$ *and* $\mathbf{L}^*[x : y]$ *are defined as follows:*

DEF
$$\begin{array}{lcll} \mathbf{L}[x : y] & \triangleq & \{z \in \mathbf{L} | z \in [x : y]\} & \forall x,y \in X \\ \mathbf{L}^*[x : y] & \triangleq & \left\{z \in \mathbf{L}^* | z \in [x : y]\right\} & \forall x,y \in X \end{array}$$

Definition P.4. [4]

DEF
In a METRIC SPACE (X, d),
the set $[a : b]$ *is the* ***closed interval*** *from* x *to* y *and is defined as*
$$[x : y] \triangleq \{z \in X | \mathsf{d}(x, z) + \mathsf{d}(z, y) = \mathsf{d}(x, y)\}.$$
An element $z \in X$ *is* ***geodesically between*** x *and* y *if* $z \in [x : y]$.

Definition P.5. [5]

DEF
In a LINEAR SPACE $\left(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times})\right)$ (Definition J.1 page 247),
$$\begin{array}{lcll} \left[\boldsymbol{x} : \boldsymbol{y}\right] & \triangleq & \{\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y} = \boldsymbol{z} | 0 \leq \lambda \leq 1\} & \text{is called a } \textbf{closed interval} \quad \text{and} \\ (\boldsymbol{x} : \boldsymbol{y}] & \triangleq & \{\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y} = \boldsymbol{z} | 0 < \lambda \leq 1\} & \text{is called a } \textbf{half-open interval} \quad \text{and} \\ \left[\boldsymbol{x} : \boldsymbol{y}\right) & \triangleq & \{\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y} = \boldsymbol{z} | 0 \leq \lambda < 1\} & \text{is called a } \textbf{half-open interval} \quad \text{and} \\ (\boldsymbol{x} : \boldsymbol{y}) & \triangleq & \{\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y} = \boldsymbol{z} | 0 < \lambda < 1\} & \text{is called an } \textbf{open interval}. \end{array}$$

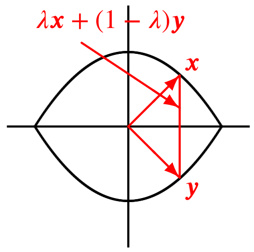
P.2 Convex sets

Using the concept of the *interval* (previous section), we can define the *convex set* (next definition).

Definition P.6. [6] *Let* X *be a* SET *in an* ORDERED SET $(X, \leq)$, *a* LATTICE $(X, \vee, \wedge; \leq)$, *a* METRIC SPACE (X, d), *or a* LINEAR SPACE $\left(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times})\right)$.

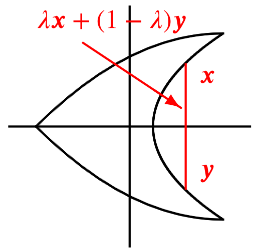
DEF
A subset $D \subseteq X$ *is a* ***convex set*** *in* X *if*
$$x, y \in D \quad\Longrightarrow\quad [x : y] \subseteq D.$$
A set that is ***not*** *convex is* ***concave***.

Example P.1. Consider the *Euclidean space* $\mathbb{R}^2$ (a special case of a *linear space*).



$\Leftarrow$ $\left\{\text{The figure to the left is a } \textit{convex} \text{ set in } \mathbb{R}^2.\right.$

$\left.\text{The figure to the right is a } \textit{concave} \text{ set in } \mathbb{R}^2.\right\}$ $\Rightarrow$



Example P.2. In a *metric space*, examples of *convex sets* are convex *balls*. Examples include those balls generated by the following metrics:

- Taxi-cab metric
- Euclidean metric
- Sup metric
- Tangential metric

[3] Maeda and Maeda (1970) page 1
[4] van de Vel (1993) page 8
[5] Barvinok (2002) page 2
[6] Barvinok (2002) page 5

Examples of metrics generating balls which are *not* convex include the following:

- Parabolic metric
- Exponential metric

P.3 Convex functions

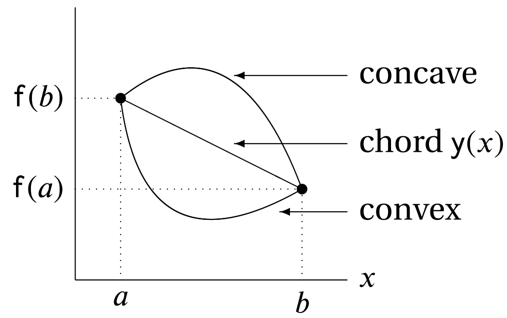


Figure P.1: Convex and concave functions

Definition P.7. [7] *Let* $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$ *be a* LINEAR SPACE (Definition J.1 page 247) *and* D *a* CONVEX SET (Definition P.6 page 338) *in* X.

DEF

A function $f \in F^D$ *is* ***convex*** *if*

$$f\left(\lambda x + [1-\lambda] y\right) \leq \lambda f(x) + (1-\lambda) f(y) \qquad \forall x, y \in D \text{ and } \forall \lambda \in (0,1)$$

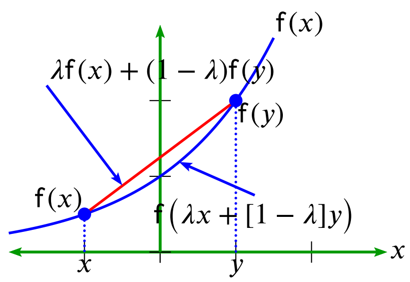
A function $g \in F^D$ *is* ***strictly convex*** *if*

$$g\left(\lambda x + [1-\lambda] y\right) = \lambda g(x) + (1-\lambda) g(y) \qquad \forall x, y \in D,\ x \neq y,\ \text{and } \forall \lambda \in (0,1)$$

A function $f \in F^D$ *is* ***concave*** *if* $-f$ *is* CONVEX.

A function $f \in F^D$ *is* ***affine*** *if* f *is* CONVEX *and* CONCAVE.

Example P.3. The function $f(x) = 2^x$ is a *convex function* (Definition P.7 page 339), as illustrated to the right.



Definition P.8. [8] *Let* $(X, +, \cdot, (\mathbb{R}, \dot{+}, \dot{\times}))$ *be a* LINEAR SPACE (Definition J.1 page 247).

DEF

The ***epigraph*** $\mathrm{epi}(f)$ *and* ***hypograph*** $\mathrm{hyp}(f)$ *of a functional* $f \in \mathbb{R}^X$ *are defined as*

$$\mathrm{epi}(f) \triangleq \{(x,y) \in X \times \mathbb{R} \mid y \geq f(x)\}$$

$$\mathrm{hyp}(f) \triangleq \{(x,y) \in X \times \mathbb{R} \mid y \leq f(x)\}$$

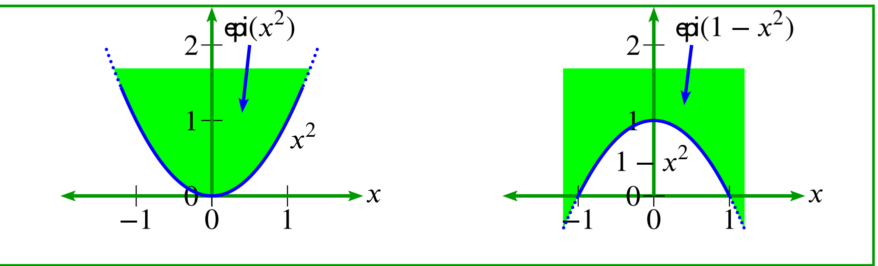
Example P.4.

---

[7] Simon (2011) page 2, Barvinok (2002) page 2, Bollobás (1999) page 3, Jensen (1906) page 176, Clarkson (1936) ⟨strictly convex⟩

[8] Beer (1993) page 13 ⟨§1.3⟩, Aubin and Frankowska (2009) page 222, Aubin (2011) page 223

EX



Proposition P.1. [9] *Let* $(X, +, \cdot, (\mathbb{R}, \dot{+}, \dot{\times}))$ *be a* LINEAR SPACE *(Definition J.1 page 247). Let* f *be a* FUNCTIONAL *in* $\mathbb{R}^X$.

PRP

$$\left\{\begin{array}{l}\mathsf{f} \textit{ is a} \\ \text{CONVEX FUNCTION}\end{array}\right\} \iff \left\{\begin{array}{l}\mathrm{epi}(\mathsf{f}) \textit{ is a} \\ \text{CONVEX SET}\end{array}\right\}$$

Often a function can be proven to be *convex* or *concave*. *Convex* and *concave* functions are defined in Definition P.9 (page 340) (next) and illustrated in Figure P.1 (page 339).

Definition P.9. *Let*

$$\mathsf{y}(x) = \frac{\mathsf{f}(b) - \mathsf{f}(a)}{b-a}(x-a) + \mathsf{f}(a).$$

A function $\mathsf{f} : \mathbb{R} \to \mathbb{R}$ *is*

DEF

(1).	***convex***	*in* $(a:b)$ *if*	$\mathsf{f}(x)$	$\le$	$\mathsf{y}(x)$	*for* $x \in (a:b)$
(2).	***concave***	*in* $(a:b)$ *if*	$\mathsf{f}(x)$	$\ge$	$\mathsf{y}(x)$	*for* $x \in (a:b)$
(3).	***strictly convex***	*in* $(a:b)$ *if*	$\mathsf{f}(x)$	$<$	$\mathsf{y}(x)$	*for* $x \in (a:b)$
(4).	***strictly concave***	*in* $(a:b)$ *if*	$\mathsf{f}(x)$	$>$	$\mathsf{y}(x)$	*for* $x \in (a:b)$

Theorem P.1 (Jensen's Inequality). [10] *Let* $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$ *be a* LINEAR SPACE *(Definition J.1 page 247), D a subset of X, and* f *a functional in* $\mathbb{F}^D$. *Let* $\sum$ *be the* SUMMATION OPERATOR *(Definition Q.1 page 345).*

THM

$$\left\{\begin{array}{lll}1. & D \text{ is CONVEX} & \textit{and} \\ 2. & \mathsf{f} \text{ is CONVEX} & \textit{and} \\ 3. & \sum_{n=1}^{N} \lambda_n = 1 & (\text{WEIGHTS})\end{array}\right\} \implies \mathsf{f}\left(\sum_{n=1}^{N} \lambda_n \boldsymbol{x}_n\right) \le \sum_{n=1}^{N} \lambda_n \mathsf{f}(\boldsymbol{x}_n) \qquad \forall \boldsymbol{x}_n \in D,\ N \in \mathbb{N}$$

✎PROOF: Proof is by induction:

1. Proof that statement is true for $N = 1$:

$$\begin{aligned}\mathsf{f}\left(\sum_{n=1}^{N=1} \lambda_n \boldsymbol{x}_n\right) &= \mathsf{f}(\lambda_1 \boldsymbol{x}_1) \\ &\le \mathsf{f}(\lambda_1 \boldsymbol{x}_1) \\ &= \sum_{n=1}^{N=1} \lambda_n \mathsf{f}(\boldsymbol{x}_n)\end{aligned}$$

[9] Udriste (1994) page 63, Kurdila and Zabarankin (2005) page 178 ⟨Proposition 6.1.1⟩, Rockafellar (1970) page 23 ⟨Section 4 Convex Functions⟩, Çinlar and Vanderbei (2013) page 86 ⟨5.4 Theorem⟩
[10] Mitrinović et al. (2010) page 6, Bollobás (1999) page 3, Lay (1982) page 7, Jensen (1906) pages 179–180

2. Proof that statement is true for $N = 2$:

$$\begin{aligned}
\mathsf{f}\left(\sum_{n=1}^{N=2} \lambda_n \boldsymbol{x}_n\right) &= \mathsf{f}(\lambda_1 \boldsymbol{x}_1 + \lambda_2 \boldsymbol{x}_2) \\
&\le \lambda_1 \mathsf{f}(\boldsymbol{x}_1) + \lambda_2 \mathsf{f}(\boldsymbol{x}_2) && \text{by convexity hypothesis} \\
&= \sum_{n=1}^{N=2} \lambda_n \mathsf{f}(\boldsymbol{x}_n)
\end{aligned}$$

3. Proof that if the statement is true for N, then it is also true for $N + 1$:

$$\begin{aligned}
\mathsf{f}\left(\sum_{n=1}^{N+1} \lambda_n \boldsymbol{x}_n\right) &= \mathsf{f}\left(\sum_{n=1}^{N} \lambda_n \boldsymbol{x}_n + \lambda_{N+1}\boldsymbol{x}_{N+1}\right) \\
&= \mathsf{f}\left([1-\lambda_{N+1}]\sum_{n=1}^{N} \frac{\lambda_n}{1-\lambda_{N+1}} \boldsymbol{x}_n + \lambda_{N+1}\boldsymbol{x}_{N+1}\right) \\
&\le [1-\lambda_{N+1}]\mathsf{f}\left(\sum_{n=1}^{N} \frac{\lambda_n}{1-\lambda_{N+1}} \boldsymbol{x}_n\right) + \lambda_{N+1}\mathsf{f}(\boldsymbol{x}_{N+1}) && \text{by convexity hypothesis} \\
&\le [1-\lambda_{N+1}]\sum_{n=1}^{N} \frac{\lambda_n}{1-\lambda_{N+1}} \mathsf{f}(\boldsymbol{x}_n) + \lambda_{N+1}\mathsf{f}(\boldsymbol{x}_{N+1}) && \text{by “true for } N\text{” hypothesis} \\
&= \sum_{n=1}^{N} \lambda_n \mathsf{f}(\boldsymbol{x}_n) + \lambda_{N+1}\mathsf{f}(\boldsymbol{x}_{N+1}) \\
&= \sum_{n=1}^{N+1} \lambda_n \mathsf{f}(\boldsymbol{x}_n)
\end{aligned}$$

4. Since the statement is true for $N = 1$, $N = 2$, and true for $N \implies$ true for $N + 1$, then it is true for $N = 1, 2, 3, 4, \ldots$

✏

The next theorem gives another form of convex functions that is a little less intuitive but provides powerful analytic results.

Theorem P.2. *Let* $\mathsf{f} : \mathbb{R} \to \mathbb{R}$. *For every* $x_1, x_2 \in (a, b)$ *and* $\lambda \in [0, 1]$

THM

$$\mathsf{f} \text{ is convex in } (a,b) \iff \mathsf{f}(\lambda x_1 + (1-\lambda)x_2) \le \lambda\mathsf{f}(x_1) + (1-\lambda)\mathsf{f}(x_2)$$

✎PROOF:

1. prove f is convex $\implies \mathsf{f}(\lambda x_1 + (1-\lambda)x_2) \le \lambda\mathsf{f}(x_1) + (1-\lambda)\mathsf{f}(x_2)$:

$$\begin{aligned}
\mathsf{f}(\lambda x_1 + (1-\lambda)x_2) &\le \frac{\mathsf{f}(b)-\mathsf{f}(a)}{b-a}[\lambda x_1 + (1-\lambda)x_2 - a] + \mathsf{f}(a) \\
&= \frac{\mathsf{f}(x_2)-\mathsf{f}(x_1)}{x_2-x_1}[\lambda x_1 + (1-\lambda)x_2 - x_1] + \mathsf{f}(x_1) \\
&= \frac{\mathsf{f}(x_2)-\mathsf{f}(x_1)}{x_2-x_1}[(x_2-x_1)(1-\lambda)] + \mathsf{f}(x_1) \\
&= (1-\lambda)\mathsf{f}(x_2) - (1-\lambda)\mathsf{f}(x_1) + \mathsf{f}(x_1) \\
&= \lambda\mathsf{f}(x_1) + (1-\lambda)\mathsf{f}(x_2)
\end{aligned}$$

2. prove f is convex $\Longleftarrow$ $f(\lambda x_1 + (1-\lambda)x_2) \le \lambda \mathsf{f}(x_1) + (1-\lambda)\mathsf{f}(x_2)$:

   Let $x = \lambda(b-a) + a$ Notice that as λ varies from 0 to 1, x varies from b to a. So free variable λ works as a change of variable for free variable x.

$$\lambda = \frac{x-a}{b-a}$$

$$\begin{aligned}
\mathsf{f}(x) &= \mathsf{f}(\lambda(b-a)+a) \\
&\le \lambda \mathsf{f}(b) + (1-\lambda)\mathsf{f}(a) \\
&= \lambda[\mathsf{f}(b) - \mathsf{f}(a)] + \mathsf{f}(a) \\
&= \frac{\mathsf{f}(b) - \mathsf{f}(a)}{b-a}(x-a) + \mathsf{f}(a)
\end{aligned}$$

✏

Taking the second derivative of a function provides a convenient test for whether that function is convex.

Theorem P.3. [11]

THM: $\mathsf{f}''(x) > 0 \implies \mathsf{f}$ *is convex*

✎PROOF:

$$\begin{aligned}
\mathsf{f}(x) &= \mathsf{f}(x_0) + \mathsf{f}'(x_0)(x-x_0) + \mathsf{f}''(c)(x-x_0)^2 \\
&\ge \mathsf{f}(x_0) + \mathsf{f}'(x_0)(x-x_0) \\
&= \mathsf{f}(x_0) + \mathsf{f}'(x_0)(x - \lambda x_1 - (1-\lambda)x_2)
\end{aligned}$$

$$\begin{aligned}
\mathsf{f}(x_1) &\ge \mathsf{f}(x_0) + \mathsf{f}'(x_0)(x_1 - \lambda x_1 - (1-\lambda)x_2) \\
&= \mathsf{f}(x_0) + \mathsf{f}'(x_0)(1-\lambda)(x_1 - x_2) \\
&= \mathsf{f}(x_0) - \mathsf{f}'(x_0)(1-\lambda)(x_2 - x_1)
\end{aligned}$$

$$\begin{aligned}
\mathsf{f}(x_2) &\ge \mathsf{f}(x_0) + \mathsf{f}'(x_0)(x_2 - \lambda x_1 - (1-\lambda)x_2) \\
&= \mathsf{f}(x_0) + \mathsf{f}'(x_0)\lambda(x_2 - x_1)
\end{aligned}$$

$$\begin{aligned}
\lambda \mathsf{f}(x_1) + (1-\lambda)\mathsf{f}(x_2) &\ge \lambda\left[\mathsf{f}(x_0) - \mathsf{f}'(x_0)(1-\lambda)(x_2-x_1)\right] + (1-\lambda)\left[\mathsf{f}(x_0) + \mathsf{f}'(x_0)\lambda(x_2-x_1)\right] \\
&= \lambda\left[\mathsf{f}(x_0) - \mathsf{f}'(x_0)(1-\lambda)(x_2-x_1)\right] + \left[\mathsf{f}(x_0) + \mathsf{f}'(x_0)\lambda(x_2-x_1)\right] - \lambda\left[\mathsf{f}(x_0) + \mathsf{f}'(x_0)\lambda(x_2-x_1)\right] \\
&= \mathsf{f}(x_0) \\
&= \mathsf{f}(\lambda x_1 + (1-\lambda)x_2)
\end{aligned}$$

By Theorem P.2 (page 341), $\mathsf{f}(x)$ is convex. ✏

P.4 Literature

LITERATURE SURVEY:

[11] Cover and Thomas (1991) pages 24–25

1. Abstract convexity:
   Edelman and Jamison (1985)
   van de Vel (1993)
   Hörmander (1994)

2. Order convexity (lattice theory):
   Edelman (1986)

3. Metric convexity:
   Menger (1928)
   Blumenthal (1970) page 41 ⟨?⟩
   Khamsi and Kirk (2001) pages 35–38

APPENDIX Q

FINITE SUMS

❝ *I think that it was Harald Bohr who remarked to me that "all analysts spend half their time hunting through the literature for inequalities which they want to use and cannot prove."* ❞

G.H. Hardy (1877–1947) in his "Presidential Address" to the London Mathematical Society on November 8, 1928, about a remark that he suggested was from Harald Bohr (1887–1951), Danish mathematician pictured to the left. [1]

Q.1 Summation

Definition Q.1. [2] *Let $+$ be an addition operator on a tuple $\llparenthesis x_n \rrparenthesis_m^N$.*

DEF

*The **summation** of $\llparenthesis x_n \rrparenthesis$ from index m to index N with respect to $+$ is*

$$\sum_{n=m}^{N} x_n \triangleq \begin{cases} 0 & \text{for } N < m \\ \left(\displaystyle\sum_{n=m}^{N-1} x_n\right) + x_N & \text{for } N \geq m \end{cases}$$

Theorem Q.1 (Generalized associative property). [3] *Let $+$ be an addition operator on a tuple $\llparenthesis x_n \rrparenthesis_m^N$.*

THM

$+$ *is* ASSOCIATIVE $\implies$

$$\underbrace{\sum_{n=m}^{L} x_n + \left(\sum_{n=L+1}^{M} x_n + \sum_{n=M+1}^{N} x_n\right) = \left(\sum_{n=m}^{L} x_n + \sum_{n=L+1}^{M} x_n\right) + \sum_{n=M+1}^{N} x_n \qquad \text{for } m < L < M \leq N}_{\sum_{n=m}^{N} \text{ is ASSOCIATIVE}}$$

---

[1] quote: Hardy (1929) page 64
image: http://www-history.mcs.st-andrews.ac.uk/PictDisplay/Bohr_Harald.html

[2] reference: Berberian (1961) page 8 ⟨Definition I.3.1⟩
"$\sum$" notation: Fourier (1820) page 280

[3] Berberian (1961) pages 9–10 ⟨Theorem I.3.1⟩

✎PROOF:

1. Proof for $N < m$ case: $\sum_{n=m}^{N} x_n = 0$.

2. Proof for $N = m$ case: $\sum_{n=m}^{m} x_n = \left(\sum_{n=m}^{m-1} x_n\right) + x_m = 0 + x_m = x_m$.

3. Proof for $N = m+1$ case: $\sum_{n=m}^{m+1} x_n = \left(\sum_{n=m}^{m} x_n\right) + x_{m+1} = x_m + x_{m+1}$

4. Proof for $N = m+2$ case:

$$\begin{aligned}
\sum_{n=m}^{m+2} x_n &= \left(\sum_{n=m}^{m+1} x_n\right) + x_{m+2} && \text{by Definition Q.1 page 345} \\
&= (x_m + x_{m+1}) + x_{m+2} && \text{by item (3)} \\
&= x_m + (x_{m+1} + x_{m+2}) && \text{by left hypothesis}
\end{aligned}$$

5. Proof that N case $\implies$ $N+1$ case:

$$\begin{aligned}
\sum_{n=m}^{N+1} x_n &= \underbrace{\left(\sum_{n=m}^{N} x_n\right)}_{\text{associative}} + x_{N+1} && \text{by Definition Q.1 page 345} \\
&= \left(\sum_{n=m}^{L} x_n + \left(\sum_{n=L+1}^{M} x_n + \sum_{n=M+1}^{N} x_n\right)\right) + x_{N+1} && = \left(\left(\sum_{n=m}^{L} x_n + \sum_{n=L+1}^{M} x_n\right) + \sum_{n=M+1}^{N} x_n\right) + x_{N+1} \\
&= \left(\sum_{n=m}^{L} x_n + \sum_{n=L+1}^{M} x_n\right) + \left(\sum_{n=M+1}^{N} x_n + x_{N+1}\right) && = \left(\sum_{n=m}^{L} x_n + \sum_{n=L+1}^{M} x_n\right) + \left(\sum_{n=M+1}^{N+1} x_n\right)
\end{aligned}$$

✏

Q.2 Means

Q.2.1 Weighted ϕ-means

Definition Q.2. [4]

DEF

*The $(\lambda_n)_1^N$ weighted ϕ-**mean** of a tuple $(x_n)_1^N$ is defined as*

$$\mathsf{M}_\phi((x_n)) \triangleq \phi^{-1}\left(\sum_{n=1}^{N} \lambda_n \phi(x_n)\right)$$

where ϕ is a CONTINUOUS *and* STRICTLY MONOTONIC *function in $\mathbb{R}^{\mathbb{R}}$ and $(\lambda_n)_{n=1}^{N}$ is a sequence of weights for which $\sum_{n=1}^{N} \lambda_n = 1$.*

[4] Bollobás (1999) page 5

Lemma Q.1. [5] *Let $\mathsf{M}_\phi(\langle x_n \rangle)$ be the $\langle \lambda_n \rangle_1^N$ weighted ϕ-mean of a tuple $\langle x_n \rangle_1^N$. Let the property* CONVEX *be defined as in Definition P.7 (page 339).*

LEM

$$\begin{array}{llllll}
\phi\psi^{-1} \text{ is CONVEX} & \text{and} & \phi \text{ is INCREASING} & \Longrightarrow & \mathsf{M}_\phi(\langle x_n \rangle) \geq \mathsf{M}_\psi(\langle x_n \rangle) \\
\phi\psi^{-1} \text{ is CONVEX} & \text{and} & \phi \text{ is DECREASING} & \Longrightarrow & \mathsf{M}_\phi(\langle x_n \rangle) \leq \mathsf{M}_\psi(\langle x_n \rangle) \\
\phi\psi^{-1} \text{ is CONCAVE} & \text{and} & \phi \text{ is INCREASING} & \Longrightarrow & \mathsf{M}_\phi(\langle x_n \rangle) \leq \mathsf{M}_\psi(\langle x_n \rangle) \\
\phi\psi^{-1} \text{ is CONCAVE} & \text{and} & \phi \text{ is DECREASING} & \Longrightarrow & \mathsf{M}_\phi(\langle x_n \rangle) \geq \mathsf{M}_\psi(\langle x_n \rangle)
\end{array}$$

✎PROOF:

1. Case where $\phi\psi^{-1}$ is *convex* and ϕ is *increasing*:

$$\begin{aligned}
\mathsf{M}_\phi(\langle x_n \rangle) &\triangleq \phi^{-1}\left(\sum_{n=1}^N \lambda_n \phi(x_n)\right) && \text{by definition of } \mathsf{M}_\phi && \text{(Definition Q.2 page 346)} \\
&= \phi^{-1}\left(\sum_{n=1}^N \lambda_n \phi\psi^{-1}\psi(x_n)\right) && \text{by definition of } \psi^{-1} \\
&\geq \phi^{-1}\left(\phi\psi^{-1}\sum_{n=1}^N \lambda_n \psi(x_n)\right) && \text{by \textit{Jensen's Inequality}} && \text{(Theorem P.1 page 340)} \\
&= \left(\psi^{-1}\sum_{n=1}^N \lambda_n \psi(x_n)\right) && \text{by definition of } \psi^{-1} \\
&\triangleq \mathsf{M}_\psi(\langle x_n \rangle) && \text{by definition of } \mathsf{M}_\psi && \text{(Definition Q.2 page 346)}
\end{aligned}$$

2. Case where $\phi\psi^{-1}$ is *convex* and ϕ is *decreasing*:

$$\begin{aligned}
\mathsf{M}_\phi(\langle x_n \rangle) &\triangleq \phi^{-1}\left(\sum_{n=1}^N \lambda_n \phi(x_n)\right) && \text{by definition of } \mathsf{M}_\phi && \text{(Definition Q.2 page 346)} \\
&= \phi^{-1}\left(\sum_{n=1}^N \lambda_n \phi\psi^{-1}\psi(x_n)\right) && \text{by definition of } \psi^{-1} \\
&\leq \phi^{-1}\left(\phi\psi^{-1}\sum_{n=1}^N \lambda_n \psi(x_n)\right) && \text{by \textit{Jensen's Inequality}} && \text{(Theorem P.1 page 340)} \\
& && \text{and because } \phi^{-1} \text{ is \textit{decreasing}} && \text{(by hypothesis)} \\
&= \left(\psi^{-1}\sum_{n=1}^N \lambda_n \psi(x_n)\right) && \text{by definition of } \psi^{-1} \\
&\triangleq \mathsf{M}_\psi(\langle x_n \rangle) && \text{by definition of } \mathsf{M}_\psi && \text{(Definition Q.2 page 346)}
\end{aligned}$$

✏

One of the most well known inequalities in mathematics is *Minkowski's Inequality* (Theorem Q.5 page 353). In 1946, H.P. Mulholland submitted a result[6] that generalizes Minkowski's Inequality to an equal weighted φ-mean. And Milovanović and Milovanovć (1979) generalized this even further to a *weighted* φ-mean (Theorem Q.2, next).

Theorem Q.2. [7]

[5] Pečarić et al. (1992) page 107, Bollobás (1999) page 5, Hardy et al. (1952) page 75
[6] Mulholland (1950)
[7] Milovanović and Milovanović (1979), Bullen (2003) page 306 ⟨Theorem 9⟩

THM
$$\left\{\begin{array}{lll|lll}\text{(1).} & \phi \textit{ is } \text{CONVEX} & \textit{and} & \text{(3).} & \phi(0)=0 & \textit{and}\\ \text{(2).} & \phi \textit{ is } \text{STRICTLY MONOTONIC} & \textit{and} & \text{(4).} & \log\circ\phi\circ\exp \textit{ is } \text{CONVEX} & \end{array}\right\} \implies \left\{\phi^{-1}\left(\sum_{n=1}^{N}\lambda_n\phi(x_n+y_n)\right) \le \phi^{-1}\left(\sum_{n=1}^{N}\lambda_n\phi(x_n)\right)+\phi^{-1}\left(\sum_{n=1}^{N}\lambda_n\phi(y_n)\right)\right\}$$

Q.2.2 Power means

Definition Q.3. [8] *Let* $\mathsf{M}_{\phi(x;r)}(\langle\!\langle x_n \rangle\!\rangle)$ *be the* $\langle\!\langle \lambda_n \rangle\!\rangle_1^N$ *weighted* ϕ*-mean of a* NON-NEGATIVE *tuple* $\langle\!\langle x_n \rangle\!\rangle_1^N$ *(Definition* Q.2 *page* 346*).*

DEF *A mean* $\mathsf{M}_{\phi(x;r)}(\langle\!\langle x_n \rangle\!\rangle)$ *is a **power mean** with parameter* r *if* $\phi(x) \triangleq x^r$. *That is,*

$$\mathsf{M}_{\phi(x;r)}(\langle\!\langle x_n \rangle\!\rangle) = \left(\sum_{n=1}^{N}\lambda_n(x_n)^r\right)^{\frac{1}{r}}$$

Theorem Q.3. [9] *Let* $\mathsf{M}_{\phi(x;r)}(\langle\!\langle x_n \rangle\!\rangle)$ *be* POWER MEAN *with parameter* r *of an* N*-tuple* $\langle\!\langle x_n \rangle\!\rangle_1^N$. *Let* $\mathbb{R}^*$ *be the set of extended real numbers* $(\mathbb{R}^* \triangleq \mathbb{R}\cup\{-\infty,\ \infty\})$.[10]

THM
$$\mathsf{M}_{\phi(x;r)}(\langle\!\langle x_n \rangle\!\rangle) \triangleq \left(\sum_{n=1}^{N}\lambda_n(x_n)^r\right)^{\frac{1}{r}} \quad \textit{is } \text{CONTINUOUS } \textit{and } \text{STRICTLY INCREASING } \textit{in } \mathbb{R}^*.$$

$$\mathsf{M}_{\phi(x;r)}(\langle\!\langle x_n \rangle\!\rangle) = \begin{cases} \min\limits_{n=1,2,\ldots,N}\langle\!\langle x_n \rangle\!\rangle & \textit{for } r=-\infty\\ \prod\limits_{n=1}^{N} x_n^{\lambda_n} & \textit{for } r=0\\ \max\limits_{n=1,2,\ldots,N}\langle\!\langle x_n \rangle\!\rangle & \textit{for } r=+\infty\end{cases}$$

✎PROOF:

1. Proof that $M_{\phi(x;r)}$ is *strictly increasing* in r:
   (a) Let r and s be such that $-\infty < r < s < \infty$.
   (b) Let $\phi_r \triangleq x^r$ and $\phi_s \triangleq x^s$. Then $\phi_r\phi_s^{-1} = x^{\frac{r}{s}}$.
   (c) The composite function $\phi_r\phi_s^{-1}$ is *convex* or *concave* depending on the values of r and s:

	$r<0$ (ϕ_r decreasing)	$r>0$ (ϕ_r increasing)
$s<0$	*convex*	(not possible)
$s>0$	*convex*	*concave*

   (d) Therefore by Lemma Q.1 (page 347),
   $$-\infty < r < s < \infty \implies \mathsf{M}_{\phi(x;r)}(\langle\!\langle x_n \rangle\!\rangle) < \mathsf{M}_{\phi(x;s)}(\langle\!\langle x_n \rangle\!\rangle).$$

2. Proof that $\mathsf{M}_{\phi(x;r)}$ is continuous in r for $r \in \mathbb{R}\backslash 0$: The sum of continuous functions is continuous. For the cases of $r \in \{-\infty,\ 0,\ \infty\}$, see the items that follow.

[8] Bullen (2003) page 175, Bollobás (1999) page 6
[9] Bullen (2003) pages 175–177 ⟨see also page 203⟩, Bollobás (1999) pages 6–8, Besso (1879), Bienaymé (1840) page 68
[10] Rana (2002) pages 385–388 ⟨Appendix A⟩

3. Lemma: $\mathsf{M}_{\phi(x;-r)}\left(\lParen x_n \rParen\right) = \left\{\mathsf{M}_{\phi(x;r)}\left(\lParen x_n^{-1} \rParen\right)\right\}^{-1}$. Proof:

$$\begin{aligned}
\left\{\mathsf{M}_{\phi(x;r)}\left(\lParen x_n^{-1} \rParen\right)\right\}^{-1} &= \left\{\left(\sum_{n=1}^{N} \lambda_n \left(x_n^{-1}\right)^r\right)^{\frac{1}{r}}\right\}^{-1} && \text{by definition of } \mathsf{M}_\phi \\
&= \left(\sum_{n=1}^{N} \lambda_n \left(x_n\right)^{-r}\right)^{\frac{1}{-r}} \\
&= \mathsf{M}_{\phi(x;-r)}\left(\lParen x_n \rParen\right) && \text{by definition of } \mathsf{M}_\phi
\end{aligned}$$

4. Proof that $\lim_{r\to\infty} \mathsf{M}_\phi\left(\lParen x_n \rParen\right) = \max_{n\in\mathbb{Z}} \lParen x_n \rParen$:

(a) Let $x_m \triangleq \max_{n\in\mathbb{Z}} \lParen x_n \rParen$

(b) Note that $\lim_{r\to\infty} \mathsf{M}_\phi \le \max_{n\in\mathbb{Z}} \lParen x_n \rParen$ because

$$\begin{aligned}
\lim_{r\to\infty} \mathsf{M}_\phi\left(\lParen x_n \rParen\right) &= \lim_{r\to\infty} \left(\sum_{n=1}^{N} \lambda_n x_n^r\right)^{\frac{1}{r}} && \text{by definition of } \mathsf{M}_\phi \\
&\le \lim_{r\to\infty} \left(\sum_{n=1}^{N} \lambda_n x_m^r\right)^{\frac{1}{r}} && \text{by definition of } x_m \text{ in item (4a) and because } \phi(x) \triangleq x^r \text{ and } \phi^{-1} \text{ are both increasing or both decreasing} \\
&= \lim_{r\to\infty} \left(x_m^r \underbrace{\sum_{n=1}^{N} \lambda_n}_{1}\right)^{\frac{1}{r}} && \text{because } x_m \text{ is a constant} \\
&= \lim_{r\to\infty} \left(x_m^r \cdot 1\right)^{\frac{1}{r}} \\
&= x_m \\
&= \max_{n\in\mathbb{Z}} \lParen x_n \rParen && \text{by definition of } x_m \text{ in item (4a)}
\end{aligned}$$

(c) But also note that $\lim_{r\to\infty} \mathsf{M}_\phi \ge \max_{n\in\mathbb{Z}} \lParen x_n \rParen$ because

$$\begin{aligned}
\lim_{r\to\infty} \mathsf{M}_\phi\left(\lParen x_n \rParen\right) &= \lim_{r\to\infty} \left(\sum_{n=1}^{N} \lambda_n x_n^r\right)^{\frac{1}{r}} && \text{by definition of } \mathsf{M}_\phi \\
&\ge \lim_{r\to\infty} \left(w_m x_m^r\right)^{\frac{1}{r}} && \text{by definition of } x_m \text{ in item (4a) and because } \phi(x) \triangleq x^r \text{ and } \phi^{-1} \text{ are both increasing or both decreasing} \\
&= \lim_{r\to\infty} w_m^{\frac{1}{r}} x_m^{\frac{r}{r}} \\
&= x_m \\
&= \max_{n\in\mathbb{Z}} \lParen x_n \rParen && \text{by definition of } x_m \text{ in item (4a)}
\end{aligned}$$

(d) Combining items (b) and (c) we have $\lim_{r\to\infty} \mathsf{M}_\phi = \max_{n\in\mathbb{Z}} \lParen x_n \rParen$.

5. Proof that $\lim_{r\to-\infty} \mathsf{M}_\phi\left(\left(x_n\right)\right) = \min_{n\in\mathbb{Z}} \left(x_n\right)$:

$$\begin{aligned}
\lim_{r\to-\infty} \mathsf{M}_{\phi(x;r)}\left(\left(x_n\right)\right) &= \lim_{r\to\infty} \mathsf{M}_{\phi(x;-r)}\left(\left(x_n\right)\right) && \text{by change of variable } r\\
&= \lim_{r\to\infty} \left\{\mathsf{M}_{\phi(x;r)}\left(\left(x_n^{-1}\right)\right)\right\}^{-1} && \text{by Lemma in item (3) page 349}\\
&= \lim_{r\to\infty} \frac{1}{\mathsf{M}_{\phi(x;r)}\left(\left(x_n^{-1}\right)\right)}\\
&= \frac{\lim_{r\to\infty} 1}{\lim_{r\to\infty} \mathsf{M}_{\phi(x;r)}\left(\left(x_n^{-1}\right)\right)} && \text{by property of lim [11]}\\
&= \frac{1}{\max_{n\in\mathbb{Z}} \left(x_n^{-1}\right)} && \text{by item (4)}\\
&= \frac{1}{\left(\min_{n\in\mathbb{Z}} \left(x_n\right)\right)^{-1}}\\
&= \min_{n\in\mathbb{Z}} \left(x_n\right)
\end{aligned}$$

6. Proof that $\lim_{r\to0} \mathsf{M}_\phi\left(\left(x_n\right)\right) = \prod_{n=1}^{N} x_n^{\lambda_n}$:

$$\begin{aligned}
\lim_{r\to0} \mathsf{M}_\phi\left(\left(x_n\right)\right) &= \lim_{r\to0} \exp\left\{\ln\left\{\mathsf{M}_\phi\left(\left(x_n\right)\right)\right\}\right\}\\
&= \lim_{r\to0} \exp\left\{\ln\left\{\left(\sum_{n=1}^{N} \lambda_n\left(x_n^r\right)\right)^{\frac{1}{r}}\right\}\right\} && \text{by definition of } \mathsf{M}_\phi\\
&= \exp\left\{\frac{\frac{\partial}{\partial r}\ln\left(\sum_{n=1}^{N} \lambda_n\left(x_n^r\right)\right)}{\frac{\partial}{\partial r} r}\right\}_{r=0} && \text{by l'Hôpital's rule [12]}\\
&= \exp\left\{\frac{\sum_{n=1}^{N} \lambda_n \frac{\partial}{\partial r}\left(x_n^r\right)}{\sum_{n=1}^{N} \lambda_n\left(x_n^r\right)}\right\}_{r=0}
= \exp\left\{\frac{\sum_{n=1}^{N} \lambda_n \frac{\partial}{\partial r}\exp\left(\ln\left(x_n^r\right)\right)}{\sum_{n=1}^{N} \lambda_n}\right\}_{r=0}\\
&= \exp\left\{\frac{\sum_{n=1}^{N} \lambda_n \frac{\partial}{\partial r}\exp\left(r\ln\left(x_n\right)\right)}{1}\right\}_{r=0}
= \exp\left\{\sum_{n=1}^{N} \lambda_n \frac{\partial}{\partial r}\exp\left(r\ln\left(x_n\right)\right)\right\}_{r=0}\\
&= \exp\left\{\sum_{n=1}^{N} \lambda_n \exp\left\{r\ln x_n\right\}\ln\left(x_n\right)\right\}_{r=0}
= \exp\left\{\sum_{n=1}^{N} \lambda_n \ln\left(x_n\right)\right\}\\
&= \exp\left\{\sum_{n=1}^{N} \ln\left(x_n^{\lambda_n}\right)\right\}
= \exp\left\{\ln\prod_{n=1}^{N} x_n^{\lambda_n}\right\} = \prod_{n=1}^{N} x_n^{\lambda_n}
\end{aligned}$$

[11] Rudin (1976) page 85 ⟨4.4 Theorem⟩

[12] Rudin (1976) page 109 ⟨5.13 Theorem⟩

Definition Q.4. *Let $⦇x_n⦈_1^N$ be a tuple. Let $⦇\lambda_n⦈_1^N$ be a tuple of weighting values.*

DEF

*The **harmonic mean** of $⦇x_n⦈$ is defined as* $\mu_h \triangleq \left(\sum_{n=1}^{N} \lambda_n \frac{1}{x_n}\right)^{-1}$ *where* $\sum_{n=1}^{N} \lambda_n = 1$

*The **geometric mean** of $⦇x_n⦈$ is defined as* $\mu_g \triangleq \sum_{n=1}^{N} x_n^{\lambda_n}$ *where* $\sum_{n=1}^{N} \lambda_n = 1$

*The **arithmetic mean** of $⦇x_n⦈$ is defined as* $\mu_a \triangleq \sum_{n=1}^{N} \lambda_n x_n$ *where* $\sum_{n=1}^{N} \lambda_n = 1$

*The **average** of $⦇x_n⦈$ is defined as* $\mu_a \triangleq \frac{1}{N}\sum_{n=1}^{N} x_n$

Q.3 Inequalities on power means

Corollary Q.1. [13] *Let $⦇x_n⦈_1^N$ be a tuple. Let $⦇\lambda_n⦈_1^N$ be a tuple of weighting values.*

COR

$$\min ⦇x_n⦈ \le \underbrace{\left(\sum_{n=1}^{N} \lambda_n \frac{1}{x_n}\right)^{-1}}_{\text{harmonic mean}} \le \underbrace{\prod_{n=1}^{N} x_n^{\lambda_n}}_{\text{geometric mean}} \le \underbrace{\sum_{n=1}^{N} \lambda_n x_n}_{\text{arithmetic mean}} \le \max ⦇x_n⦈ \quad \text{where} \quad \sum_{n=1}^{N} \lambda_n = 1$$

✎PROOF:

1. These five means are all special cases of the *power mean* $\mathsf{M}_{\phi(x:r)}$ (Definition Q.3 page 348):
   - $r = \infty$: $\max ⦇x_n⦈$
   - $r = 1$: arithmetic mean
   - $r = 0$: geometric mean
   - $r = -1$: harmonic mean
   - $r = -\infty$: $\min ⦇x_n⦈$
2. The inequalities follow directly from Theorem Q.3 (page 348).
3. Generalized AM-GM inequality: If one is only concerned with the arithmetic mean and geometric mean, their relationship can be established directly using *Jensen's Inequality*:

$$\sum_{n=1}^{N} \lambda_n x_n = b^{\log_b\left(\sum_{n=1}^{N} \lambda_n x_n\right)} \ge b^{\left(\sum_{n=1}^{N} \lambda_n \log_b x_n\right)} \quad \text{by } \textit{Jensen's Inequality} \text{ (Theorem P.1 page 340)}$$

$$= \prod_{n=1}^{N} b^{(\lambda_n \log_b x_n)} = \prod_{n=1}^{N} b^{(\log_b x_n)\lambda_n} = \prod_{n=1}^{N} x_n^{\lambda_n}$$

[13] Bullen (2003) page 71, Bollobás (1999) page 5, Cauchy (1821) pages 457–459 ⟨Note II, theorem 17⟩, Jensen (1906) page 183

Lemma Q.2 (Young's Inequality). [14]

LEM

$$xy < \frac{x^p}{p} + \frac{y^q}{q} \quad \textit{with } \frac{1}{p}+\frac{1}{q}=1 \quad \forall 1<p<\infty,\ x,y\geq 0, \quad \textit{but } y \neq x^{p-1}$$

$$xy = \frac{x^p}{p} + \frac{y^q}{q} \quad \textit{with } \frac{1}{p}+\frac{1}{q}=1 \quad \forall 1<p<\infty,\ x,y\geq 0, \quad \textit{and } y = x^{p-1}$$

✎PROOF:

1. Proof that $\frac{1}{p-1} = q-1$:

$$\begin{aligned}
\frac{1}{p}+\frac{1}{q}=1 &\iff \frac{q}{q}+\frac{q}{p}=q\\
&\iff q\left(1-\frac{1}{p}\right)=1\\
&\iff q=\frac{1}{1-\frac{1}{p}}\\
&\iff q=\frac{p}{p-1}\\
&\iff q-1=\frac{p}{p-1}-\frac{p-1}{p-1}\\
&\iff q-1=\frac{p-(p-1)}{p-1}\\
&\iff q-1=\frac{1}{p-1}
\end{aligned}$$

2. Proof that $v=u^{p-1} \iff u=v^{q-1}$:

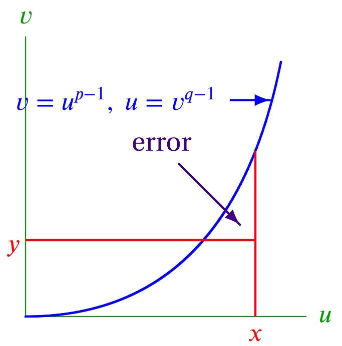
$$\begin{aligned}
u &= v^{\frac{1}{p-1}} && \text{by left hypothesis}\\
&= v^{q-1} && \text{by item (1)}
\end{aligned}$$

3. Proof that $v=u^{p-1}$ is propemonotonically increasing in u and $u=v^{q-1}$ is propemonotonically increasing in v:

$$\begin{aligned}
\frac{\mathrm{d}v}{\mathrm{d}u} &= \frac{\mathrm{d}}{\mathrm{d}u}u^{p-1} && = (p-1)u^{p-2} && > 0\\
\frac{\mathrm{d}u}{\mathrm{d}v} &= \frac{\mathrm{d}}{\mathrm{d}v}v^{q-1} && = (q-1)v^{q-2} && > 0
\end{aligned}$$

4. Proof that $xy \leq \frac{x^p}{p}+\frac{y^q}{q}$:

$$\begin{aligned}
xy &\leq \int_0^x u^{p-1}\,\mathrm{d}u + \int_0^y v^{q-1}\,\mathrm{d}v\\
&= \left.\frac{u^p}{p}\right|_0^x + \left.\frac{v^q}{q}\right|_0^y\\
&= \frac{x^p}{p}+\frac{y^q}{q}
\end{aligned}$$



[14] Carothers (2000) page 43, Tolsted (1964) page 5, Maligranda (1995) page 257, Hardy et al. (1952) ⟨Theorem 24⟩, Young (1912) page 226

Theorem Q.4 (Hölder's Inequality). [15] *Let* $(x_n \in \mathbb{C})_1^N$ *and* $(y_n \in \mathbb{C})_1^N$ *be complex* N*-tuples.*

THM
$$\underbrace{\sum_{n=1}^{N} |x_n y_n|}_{\|x \cdot y\|_1} \le \underbrace{\left(\sum_{n=1}^{N} |x_n|^p\right)^{\frac{1}{p}}}_{\|x\|_p} \underbrace{\left(\sum_{n=1}^{N} |y_n|^q\right)^{\frac{1}{q}}}_{\|y\|_q} \qquad \text{with} \qquad \frac{1}{p} + \frac{1}{q} = 1 \qquad \forall\, 1 < p < \infty$$

✎PROOF: Let $\|x_n\|_p \triangleq \left(\sum_{n=1}^{N} |x_n|^p\right)^{\frac{1}{p}}$.

$$\begin{aligned}
\sum_{n=1}^{N} |x_n y_n| &= \|(x_n)\|_p \|(y_n)\|_q \sum_{n=1}^{N} \frac{|x_n y_n|}{\|(x_n)\|_p \|(y_n)\|_q} \\
&= \|(x_n)\|_p \|(y_n)\|_q \sum_{n=1}^{N} \frac{|x_n|}{\|(x_n)\|_p} \frac{|y_n|}{\|(y_n)\|_q} \\
&\le \|(x_n)\|_p \|(y_n)\|_q \sum_{n=1}^{N} \left(\frac{1}{p}\frac{|x_n|^p}{\|(x_n)\|_p^p} + \frac{1}{q}\frac{|y_n|^q}{\|(y_n)\|_q^q}\right) && \text{by } \textit{Young's Inequality} \quad \text{(Lemma Q.2 page 352)} \\
&= \|(x_n)\|_p \|(y_n)\|_q \left(\frac{1}{p} \cdot \frac{\sum |x_n|^p}{\|(x_n)\|_p^p} + \frac{1}{q} \cdot \frac{\sum |y_n|^q}{\|(y_n)\|_q^q}\right) \\
&= \|(x_n)\|_p \|(y_n)\|_q \left(\frac{1}{p}\frac{\|(x_n)\|_p^p}{\|(x_n)\|_p^p} + \frac{1}{q}\frac{\|(y_n)\|_q^q}{\|(y_n)\|_q^q}\right) && \text{by definition of } \|\cdot\| \\
&= \|(x_n)\|_p \|(y_n)\|_q \underbrace{\left(\frac{1}{p} + \frac{1}{q}\right)}_{1} \\
&= \|(x_n)\|_p \|(y_n)\|_q && \text{by } \frac{1}{p} + \frac{1}{q} = 1 \text{ constraint}
\end{aligned}$$

✏

Theorem Q.5 (Minkowski's Inequality for sequences). [16] *Let* $(x_n \in \mathbb{C})_1^N$ *and* $(y_n \in \mathbb{C})_1^N$ *be complex* N*-tuples.*

THM
$$\left(\sum_{n=1}^{N} |x_n + y_n|^p\right)^{\frac{1}{p}} \le \left(\sum_{n=1}^{N} |x_n|^p\right)^{\frac{1}{p}} + \left(\sum_{n=1}^{N} |y_n|^p\right)^{\frac{1}{p}} \qquad \forall\, 1 < p < \infty$$

✎PROOF:

1. Define $q \triangleq \dfrac{p}{p-1}$

[15] Bullen (2003) page 178 ⟨2.1⟩, Carothers (2000) page 44, Tolsted (1964) page 6, Maligranda (1995) page 257, Hardy et al. (1952) ⟨Theorem 11⟩, Hölder (1889)

[16] Bullen (2003) page 179, Carothers (2000) page 44, Tolsted (1964) page 7, Maligranda (1995) page 258, Hardy et al. (1952) ⟨Theorem 24⟩, Minkowski (1910) page 115

2. Define $\|\boldsymbol{x}\|_p \triangleq \left(\sum_{n=1}^{N} |x_n|^p\right)^{\frac{1}{p}}$.

3. Proof that $\|x_n + y_n\|_p \le \|x_n\|_p + \|y_n\|_p$:

$$
\begin{aligned}
&\boxed{\|x_n + y_n\|_p^p} \\
&= \sum_{n=1}^{N} |x_n + y_n|^p && \text{by definition of } \|\cdot\|_p && \text{(definition 2 page 354)} \\
&= \sum_{n=1}^{N} |x_n + y_n||x_n + y_n|^{p-1} && \text{by \textit{homogeneous} property of } |\cdot| \\
&\boxed{\le} \sum_{n=1}^{N} |x_n||x_n + y_n|^{p-1} + \sum_{n=1}^{N} |y_n||x_n + y_n|^{p-1} && \text{by \textit{subadditive} property of } |\cdot| \\
&= \sum_{n=1}^{N} |x_n(x_n + y_n)^{p-1}| + \sum_{n=1}^{N} |y_n(x_n + y_n)^{p-1}| && \text{by \textit{homogeneous} property of } |\cdot| \\
&\boxed{\le} \|x_n\|_p \|(x_n + y_n)^{p-1}\|_q + \|y_n\|_p \|(x_n + y_n)^{p-1}\|_q && \text{by \textit{Hölder's Inequality}} && \text{(Theorem Q.4 page 353)} \\
&= \left(\|x_n\|_p + \|y_n\|_p\right) \|(x_n + y_n)^{p-1}\|_q \\
&= \left(\|x_n\|_p + \|y_n\|_p\right) \left(\sum_{n=1}^{N} |(x_n + y_n)^{p-1}|^q\right)^{\frac{1}{q}} && \text{by definition of } \|\cdot\|_p && \text{(definition 2 page 354)} \\
&= \left(\|x_n\|_p + \|y_n\|_p\right) \left(\sum_{n=1}^{N} |(x_n + y_n)^{p-1}|^{\frac{p}{p-1}}\right)^{\frac{p-1}{p}} && \text{by definition 1} \\
&= \left(\|x_n\|_p + \|y_n\|_p\right) \left(\sum_{n=1}^{N} |(x_n + y_n)|^p\right)^{\frac{p-1}{p}} \\
&= \left(\|x_n\|_p + \|y_n\|_p\right) \|x_n + y_n\|^{p-1} && \text{by definition of } \|\cdot\|_p && \text{(definition 2 page 354)} \\
&\implies \boxed{\|x_n + y_n\|_p \le \|x_n\|_p + \|y_n\|_p}
\end{aligned}
$$

✏

❝*Cauchy is the only one occupied with pure mathematics: Poisson, Fourier, Ampere, etc., busy themselves exclusively with magnetism and other physical subjects. Mr. Laplace writes nothing now, I believe.*❞
Niels Henrik Abel in an 1826 letter [17]

Theorem Q.6 (Cauchy-Schwarz Inequality for sequences). [18] *Let $(x_n \in \mathbb{C})_1^N$ and $(y_n \in \mathbb{C})_1^N$ be complex N-tuples.*

[17] quote: Bell (1986) page 318 ⟨Chapter 17. "Genius and Poverty" "Abel (1802–1829)"⟩, Boyer and Merzbach (2011) page 462 ⟨without "Mr. Laplace…" portion⟩. image: http://en.wikipedia.org/wiki/File:Augustin-Louis_Cauchy_1901.jpg, public domain

[18] Aliprantis and Burkinshaw (1998) page 278, Scharz (1885), Bouniakowsky (1859), Hardy et al. (1952) page 25 ⟨Theorem 11⟩, Cauchy (1821) page 455 ⟨???⟩

THM

$$\left|\sum_{n=1}^{N} x_n y_n^*\right|^2 \le \left(\sum_{n=1}^{N} |x_n|^2\right)\left(\sum_{n=1}^{N} |y_n|^2\right) \qquad \forall \boldsymbol{x},\boldsymbol{y} \in \boldsymbol{X}$$

$$\left|\sum_{n=1}^{N} x_n y_n^*\right|^2 = \left(\sum_{n=1}^{N} |x_n|^2\right)\left(\sum_{n=1}^{N} |y_n|^2\right) \iff \exists a \in \mathbb{C} \text{ such that } \boldsymbol{y} = a\boldsymbol{x} \qquad \forall \boldsymbol{x},\boldsymbol{y} \in \boldsymbol{X}$$

✎PROOF:

1. The *Cauchy-Schwarz Inequality for sequences* is a special case of the *Hölder inequality* (Theorem Q.4 page 353) for $p = q = 2$.
2. Alternatively, the *Cauchy-Schwarz inequality for sequences* is a special case of the *Cauchy-Schwarz inequality for inner-product spaces*:
   (a) $\langle x_n \mid y_n \rangle \triangleq \sum_{n=1}^{N} x_n y_n$ is an inner-product and $\left(\llparenthesis x_n \rrparenthesis, \langle \triangle \mid \triangledown \rangle \right)$ is an inner-product space.
   (b) By the more general *Cauchy-Schwarz Inequality for inner-product spaces*,

$$\begin{aligned}
\left(\sum_{n=1}^{N} a_n \lambda_n\right)^2 &\triangleq \langle a_n \mid \lambda_n \rangle^2 && \text{by definition of } \langle x_n \mid y_n \rangle \\
&\boxed{\le} \|x_n\|^2 \, \|y_n\|^2 && \text{by } \textit{Cauchy-Schwarz Inequality for inner-product spaces} \text{ (Theorem N.2 page 310)} \\
&\triangleq \left(\sum_{n=1}^{N} x_n^2\right)\left(\sum_{n=1}^{N} y_n^2\right) && \text{by definition of } \|\cdot\|
\end{aligned}$$

3. Not only does the *Hölder inequality* imply the *Cauchy-Schwarz inequality*, but somewhat surprisingly, the converse is also true: The Cauchy-Schwarz inequality implies the Hölder inequality.[19]

✎

Proposition Q.1. [20]

PRP $(x + y)^p \le 2^p(x^p + y^p) \qquad \forall x, y \ge 0, \ 1 < p < \infty$

✎PROOF:

$$\begin{aligned}
(x + y)^p &\le (2 \max\{x, y\})^p \\
&= 2^p (\max\{x, y\})^p \\
&= 2^p (\max\{x^p, y^p\}) \\
&\le 2^p (x^p + y^p)
\end{aligned}$$

✎

[19] Bullen (2003) pages 183–185 ⟨Theorem 5⟩
[20] Carothers (2000) page 43

Q.4 Power Sums

Theorem Q.7 (Geometric Series). [21]

THM
$$\sum_{k=0}^{n-1} r^k = \frac{1-r^n}{1-r} \qquad \forall r \in \mathbb{C}\backslash 0$$

✎PROOF:

$$\boxed{\sum_{k=0}^{n-1} r^k} = \left(\frac{1}{1-r}\right)\left[(1-r)\sum_{k=0}^{n-1} r^k\right] = \left(\frac{1}{1-r}\right)\left[\sum_{k=0}^{n-1} r^k - r\sum_{k=0}^{n-1} r^k\right] = \left(\frac{1}{1-r}\right)\left[\sum_{k=0}^{n-1} r^k - \left(\sum_{k=0}^{n-1} r^k - 1 + r^n\right)\right]$$
$$= \left(\frac{1}{1-r}\right)\left[1-r^n\right] = \boxed{\frac{1-r^n}{1-r}}$$

✏

Lemma Q.3. *Let* $\mathsf{f}(x)$ *be a function.*

LEM
$$\mathsf{S}(x) \triangleq \sum_{n\in\mathbb{Z}} \mathsf{f}(x+n\tau) = \mathsf{S}(x+\tau) \qquad (\mathsf{S}(x) \text{ is PERIODIC with period } \tau)$$

✎PROOF:

$$\mathsf{S}(x+\tau) \triangleq \sum_{n\in\mathbb{Z}} \mathsf{f}(x+\tau+n\tau) = \sum_{n\in\mathbb{Z}} \mathsf{f}(x+(n+1)\tau) = \sum_{m\in\mathbb{Z}} \mathsf{f}(x+m\tau) \quad (\text{where } m \triangleq n+1)$$
$$\triangleq \mathsf{S}(x)$$

✏

Proposition Q.2 (Power Sums). [22]

PRP
$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2} \quad \forall n\in\mathbb{N} \qquad \sum_{k=1}^{n} k^3 = \frac{n^2(n+1)^2}{4} \quad \forall n\in\mathbb{N}$$
$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6} \quad \forall n\in\mathbb{N} \qquad \sum_{k=1}^{n} k^4 = \frac{n(n+1)(2n+1)(3n^2+3n-1)}{30} \quad \forall n\in\mathbb{N}$$

✎PROOF:

1. Proof that $\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$: (proof by induction)

$$\sum_{k=1}^{n=1} k = 1 = \frac{1(1+1)}{2} = \left.\frac{n(n+1)}{2}\right|_{n=1}$$
$$\sum_{k=1}^{n+1} k = \left(\sum_{k=1}^{n} k\right) + \left(n+1\right) = \underbrace{\left(\frac{n(n+1)}{2}\right) + \left(n+1\right)}_{\text{by left hypothesis}} = \left(n+1\right)\left(\frac{n}{2}+1\right)$$
$$= \left(n+1\right)\left(\frac{n+2}{2}\right) = \frac{(n+1)(n+2)}{2}$$

[21] Hall and Knight (1894) page 39 ⟨article 55⟩
[22] Amann and Escher (2008) pages 51–57, Menini and Oystaeyen (2004) page 91 ⟨Exercises 5.36–5.39⟩

2. Proof that $\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$: (proof by induction)

$$\sum_{k=1}^{n=1} k^2 = 1 = \frac{1(1+1)(2+1)}{6} = \left.\frac{n(n+1)(2n+1)}{6}\right|_{n=1}$$

$$\sum_{k=1}^{n+1} k^2 = \left(\sum_{k=1}^{n} k^2\right) + \left(n+1\right)^2 = \underbrace{\left(\frac{n(n+1)(2n+1)}{6}\right) + \left(n+1\right)^2}_{\text{by left hypothesis}} = \left(n+1\right)\left(\frac{n(2n+1)+6(n+1)}{6}\right)$$

$$= \left(n+1\right)\left(\frac{2n^2+7n+6}{6}\right) = \left(n+1\right)\left(\frac{(n+2)(2n+3)}{6}\right) = \frac{(n+1)[(n+1)+1][2(n+1)+1]}{6}$$

✏

APPENDIX R

OPERATORS ON LINEAR SPACES

❝*And I am not afraid to say that there is a way to advance algebra as far beyond what Vieta and Descartes have left us as Vieta and Descartes carried it beyond the ancients.... we need still another analysis which is distinctly geometrical or linear, and which will express situation directly as algebra expresses magnitude directly.*❞
Gottfried Leibniz (1646–1716), German mathematician, in a September 8, 1679 letter to Christian Huygens. [1]

R.1 Operators on linear spaces

R.1.1 Operator Algebra

An operator is simply a function that maps from a linear space to another linear space (or to the same linear space).

Definition R.1. [2]

D E F | *A function* $\mathbf{A}$ *in* $\boldsymbol{Y}^{\boldsymbol{X}}$ *is an* ***operator*** *in* $\boldsymbol{Y}^{\boldsymbol{X}}$ *if* $\boldsymbol{X}$ *and* $\boldsymbol{Y}$ *are both* LINEAR SPACES *(Definition J.1 page 247)*.

Two operators $\mathbf{A}$ and $\mathbf{B}$ in $\boldsymbol{Y}^{\boldsymbol{X}}$ are **equal** if $\mathbf{A}\boldsymbol{x} = \mathbf{B}\boldsymbol{x}$ for all $\boldsymbol{x} \in \boldsymbol{X}$. The inverse relation of an operator $\mathbf{A}$ in $\boldsymbol{Y}^{\boldsymbol{X}}$ always exists as a *relation* in $\mathbb{2}^{\boldsymbol{X}\boldsymbol{Y}}$, but may not always be a *function* (may not always be an operator) in $\boldsymbol{Y}^{\boldsymbol{X}}$.

The operator $\mathbf{I} \in \boldsymbol{X}^{\boldsymbol{X}}$ is the *identity* operator if $\mathbf{I}\boldsymbol{x} = \mathbf{I}$ for all $\boldsymbol{x} \in \boldsymbol{X}$.

Definition R.2. [3] *Let* $\boldsymbol{X}^{\boldsymbol{X}}$ *be the set of all operators with from a* LINEAR SPACE $\boldsymbol{X}$ *to* $\boldsymbol{X}$. *Let* $\mathbf{I}$ *be an*

[1] quote: Leibniz (1679) pages 248–249
image: http://en.wikipedia.org/wiki/File:Gottfried_Wilhelm_von_Leibniz.jpg, public domain
[2] Heil (2011) page 42
[3] Michel and Herget (1993) page 411

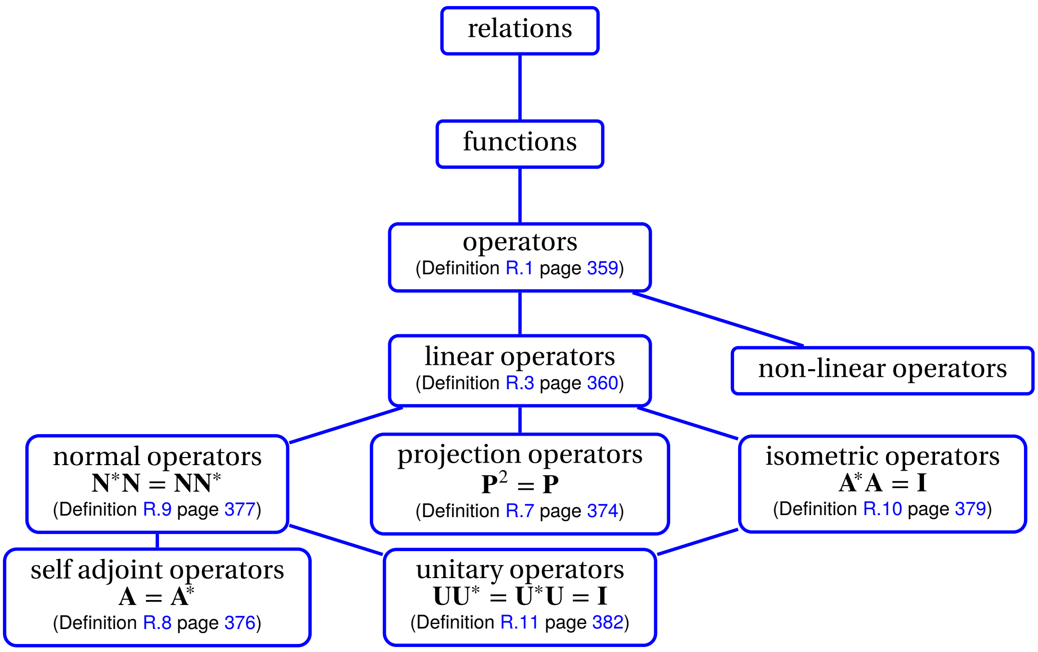


Figure R.1: Some operator types

operator in $\boldsymbol{X}^{\boldsymbol{X}}$. *Let* $\mathbb{1}(\boldsymbol{X})$ *be the* IDENTITY ELEMENT *in* $\boldsymbol{X}^{\boldsymbol{X}}$.

DEF: $\mathbf{I}$ *is the* ***identity operator*** *in* $\boldsymbol{X}^{\boldsymbol{X}}$ *if* $\qquad \mathbf{I} = \mathbb{1}(\boldsymbol{X})$.

R.1.2 Linear operators

Definition R.3. [4] *Let* $\boldsymbol{X} \triangleq \left(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times})\right)$ *and* $\boldsymbol{Y} \triangleq \left(Y, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times})\right)$ *be linear spaces.*

DEF: *An operator* $\mathbf{L} \in \boldsymbol{Y}^{\boldsymbol{X}}$ *is* ***linear*** *if*

1. $\mathbf{L}(\boldsymbol{x}+\boldsymbol{y}) = \mathbf{L}\boldsymbol{x} + \mathbf{L}\boldsymbol{y} \qquad \forall \boldsymbol{x},\boldsymbol{y} \in \boldsymbol{X}$ (ADDITIVE) *and*
2. $\mathbf{L}(\alpha\boldsymbol{x}) = \alpha\mathbf{L}\boldsymbol{x} \qquad \forall \boldsymbol{x} \in \boldsymbol{X},\ \forall \alpha \in \mathbb{F}$ (HOMOGENEOUS).

The set of all linear operators from $\boldsymbol{X}$ *to* $\boldsymbol{Y}$ *is denoted* $\mathscr{L}(\boldsymbol{X}, \boldsymbol{Y})$ *such that*
$\mathscr{L}(\boldsymbol{X}, \boldsymbol{Y}) \triangleq \left\{\mathbf{L} \in \boldsymbol{Y}^{\boldsymbol{X}} \mid \mathbf{L} \text{ is linear}\right\}$.

Theorem R.1. [5] *Let* $\mathbf{L}$ *be an operator from a linear space* $\boldsymbol{X}$ *to a linear space* $\boldsymbol{Y}$, *both over a field* $\mathbb{F}$.

THM:
$$\{\mathbf{L} \text{ is LINEAR}\} \implies \left\{\begin{array}{llllll} 1. & \mathbf{L}\mathbb{0} & = & \mathbb{0} & & \text{and} \\ 2. & \mathbf{L}(-\boldsymbol{x}) & = & -(\mathbf{L}\boldsymbol{x}) & \forall \boldsymbol{x} \in \boldsymbol{X} & \text{and} \\ 3. & \mathbf{L}(\boldsymbol{x}-\boldsymbol{y}) & = & \mathbf{L}\boldsymbol{x} - \mathbf{L}\boldsymbol{y} & \forall \boldsymbol{x},\boldsymbol{y} \in \boldsymbol{X} & \text{and} \\ 4. & \mathbf{L}\left(\sum_{n=1}^{N} \alpha_n \boldsymbol{x}_n\right) & = & \sum_{n=1}^{N} \alpha_n \left(\mathbf{L}\boldsymbol{x}_n\right) & \boldsymbol{x}_n \in \boldsymbol{X}, \alpha_n \in \mathbb{F} & \end{array}\right\}$$

✎PROOF:

[4] Kubrusly (2001) page 55, Aliprantis and Burkinshaw (1998) page 224, Hilbert et al. (1927) page 6, Stone (1932) page 33

[5] Berberian (1961) page 79 ⟨Theorem IV.1.1⟩

1. Proof that $\mathbf{L}\mathbb{0} = \mathbb{0}$:

$$\begin{aligned}
\mathbf{L}\mathbb{0} &= \mathbf{L}(0\cdot\mathbb{0}) && \text{by additive identity property} && \text{(Theorem J.1 page 249)}\\
&= 0\cdot(\mathbf{L}\mathbb{0}) && \text{by } \textit{homogeneous} \text{ property of } \mathbf{L} && \text{(Definition R.3 page 360)}\\
&= \mathbb{0} && \text{by } \textit{additive identity} \text{ property} && \text{(Theorem J.1 page 249)}
\end{aligned}$$

2. Proof that $\mathbf{L}(-\boldsymbol{x}) = -(\mathbf{L}\boldsymbol{x})$:

$$\begin{aligned}
\mathbf{L}(-\boldsymbol{x}) &= \mathbf{L}(-1\cdot\boldsymbol{x}) && \text{by } \textit{additive inverse} \text{ property} && \text{(Theorem J.2 page 250)}\\
&= -1\cdot(\mathbf{L}\boldsymbol{x}) && \text{by } \textit{homogeneous} \text{ property of } \mathbf{L} && \text{(Definition R.3 page 360)}\\
&= -(\mathbf{L}\boldsymbol{x}) && \text{by } \textit{additive inverse} \text{ property} && \text{(Theorem J.2 page 250)}
\end{aligned}$$

3. Proof that $\mathbf{L}(\boldsymbol{x}-\boldsymbol{y}) = \mathbf{L}\boldsymbol{x} - \mathbf{L}\boldsymbol{y}$:

$$\begin{aligned}
\mathbf{L}(\boldsymbol{x}-\boldsymbol{y}) &= \mathbf{L}(\boldsymbol{x}+(-\boldsymbol{y})) && \text{by } \textit{additive inverse} \text{ property} && \text{(Theorem J.2 page 250)}\\
&= \mathbf{L}(\boldsymbol{x}) + \mathbf{L}(-\boldsymbol{y}) && \text{by } \textit{linearity} \text{ property of } \mathbf{L} && \text{(Definition R.3 page 360)}\\
&= \mathbf{L}\boldsymbol{x} - \mathbf{L}\boldsymbol{y} && \text{by item (2)}
\end{aligned}$$

4. Proof that $\mathbf{L}\left(\sum_{n=1}^{N}\alpha_n\boldsymbol{x}_n\right) = \sum_{n=1}^{N}\alpha_n\left(\mathbf{L}\boldsymbol{x}_n\right)$:

   (a) Proof for $N = 1$:

$$\begin{aligned}
\mathbf{L}\left(\sum_{n=1}^{N}\alpha_n\boldsymbol{x}_n\right) &= \mathbf{L}\left(\alpha_1\boldsymbol{x}_1\right) && \text{by } N=1 \text{ hypothesis}\\
&= \alpha_1\left(\mathbf{L}\boldsymbol{x}_1\right) && \text{by } \textit{homogeneous} \text{ property of } \mathbf{L} && \text{(Definition R.3 page 360)}
\end{aligned}$$

   (b) Proof that N case $\implies$ $N+1$ case:

$$\begin{aligned}
\mathbf{L}\left(\sum_{n=1}^{N+1}\alpha_n\boldsymbol{x}_n\right) &= \mathbf{L}\left(\alpha_{N+1}\boldsymbol{x}_{N+1} + \sum_{n=1}^{N}\alpha_n\boldsymbol{x}_n\right)\\
&= \mathbf{L}\left(\alpha_{N+1}\boldsymbol{x}_{N+1}\right) + \mathbf{L}\left(\sum_{n=1}^{N}\alpha_n\boldsymbol{x}_n\right) && \text{by } \textit{linearity} \text{ property of } \mathbf{L} && \text{(Definition R.3 page 360)}\\
&= \alpha_{N+1}\mathbf{L}\left(\boldsymbol{x}_{N+1}\right) + \sum_{n=1}^{N}\mathbf{L}\left(\alpha_n\boldsymbol{x}_n\right) && \text{by left } N+1 \text{ hypothesis}\\
&= \sum_{n=1}^{N+1}\mathbf{L}\left(\alpha_n\boldsymbol{x}_n\right)
\end{aligned}$$

✏

Theorem R.2. [6] *Let $\mathscr{L}(\boldsymbol{X},\boldsymbol{Y})$ be the set of all linear operators from a linear space $\boldsymbol{X}$ to a linear space $\boldsymbol{Y}$. Let $\mathscr{N}(\mathbf{L})$ be the* NULL SPACE *of an operator $\mathbf{L}$ in $\boldsymbol{Y}^{\boldsymbol{X}}$ and $\mathscr{I}(\mathbf{L})$ the* IMAGE SET *of $\mathbf{L}$ in $\boldsymbol{Y}^{\boldsymbol{X}}$.*

THM

$\mathscr{L}(\boldsymbol{X},\boldsymbol{Y})$	*is a linear space*		*(space of linear transforms)*
$\mathscr{N}(\mathbf{L})$	*is a linear subspace of $\boldsymbol{X}$*	$\forall\mathbf{L}\in\boldsymbol{Y}^{\boldsymbol{X}}$	
$\mathscr{I}(\mathbf{L})$	*is a linear subspace of $\boldsymbol{Y}$*	$\forall\mathbf{L}\in\boldsymbol{Y}^{\boldsymbol{X}}$	

✎PROOF:

[6] Michel and Herget (1993) pages 98–104, Berberian (1961) pages 80–85 ⟨Theorem IV.1.4 and Theorem IV.3.1⟩

1. Proof that $\mathscr{N}(\mathbf{L})$ is a linear subspace of $\boldsymbol{X}$:
   (a) $\mathbb{0} \in \mathscr{N}(\mathbf{L}) \implies \mathscr{N}(\mathbf{L}) \neq \emptyset$
   (b) $\mathscr{N}(\mathbf{L}) \triangleq \{\boldsymbol{x} \in \boldsymbol{X} | \mathbf{L}\boldsymbol{x} = \mathbb{0}\} \subseteq \boldsymbol{X}$
   (c) $\boldsymbol{x} + \boldsymbol{y} \in \mathscr{N}(\mathbf{L}) \implies \mathbb{0} = \mathbf{L}(\boldsymbol{x} + \boldsymbol{y}) = \mathbf{L}(\boldsymbol{y} + \boldsymbol{x}) \implies \boldsymbol{y} + \boldsymbol{x} \in \mathscr{N}(\mathbf{L})$
   (d) $\alpha \in \mathbb{F},\ \boldsymbol{x} \in \boldsymbol{X} \implies \mathbb{0} = \mathbf{L}\boldsymbol{x} \implies \mathbb{0} = \alpha\mathbf{L}\boldsymbol{x} \implies \mathbb{0} = \mathbf{L}(\alpha\boldsymbol{x}) \implies \alpha\boldsymbol{x} \in \mathscr{N}(\mathbf{L})$
2. Proof that $\mathscr{I}(\mathbf{L})$ is a linear subspace of $\boldsymbol{Y}$:
   (a) $\mathbb{0} \in \mathscr{I}(\mathbf{L}) \implies \mathscr{I}(\mathbf{L}) \neq \emptyset$
   (b) $\mathscr{I}(\mathbf{L}) \triangleq \{\boldsymbol{y} \in \boldsymbol{Y} | \exists \boldsymbol{x} \in \boldsymbol{X} \text{ such that } \boldsymbol{y} = \mathbf{L}\boldsymbol{x}\} \subseteq \boldsymbol{Y}$
   (c) $\boldsymbol{x} + \boldsymbol{y} \in \mathscr{I}(\mathbf{L}) \implies \exists \boldsymbol{v} \in \boldsymbol{X} \text{ such that } \mathbf{L}\boldsymbol{v} = \boldsymbol{x} + \boldsymbol{y} = \boldsymbol{y} + \boldsymbol{x} \implies \boldsymbol{y} + \boldsymbol{x} \in \mathscr{I}(\mathbf{L})$
   (d) $\alpha \in \mathbb{F},\ \boldsymbol{x} \in \mathscr{I}(\mathbf{L}) \implies \exists \boldsymbol{x} \in \boldsymbol{X} \text{ such that } \boldsymbol{y} = \mathbf{L}\boldsymbol{x} \implies \alpha\boldsymbol{y} = \alpha\mathbf{L}\boldsymbol{x} = \mathbf{L}(\alpha\boldsymbol{x}) \implies \alpha\boldsymbol{x} \in \mathscr{I}(\mathbf{L})$

✏

Example R.1. [7] Let $\mathscr{C}([a : b], \mathbb{R})$ be the set of all *continuous* functions from the closed real interval $[a : b]$ to $\mathbb{R}$.

EX: $\mathscr{C}([a : b], \mathbb{R})$ is a linear space.

Theorem R.3. [8] *Let $\mathscr{L}(\boldsymbol{X}, \boldsymbol{Y})$ be the set of linear operators from a linear space $\boldsymbol{X}$ to a linear space $\boldsymbol{Y}$. Let $\mathscr{N}(\mathbf{L})$ be the* NULL SPACE *of a linear operator $\mathbf{L} \in \mathscr{L}(\boldsymbol{X}, \boldsymbol{Y})$.*

THM:
$$\begin{array}{lcl} \mathbf{L}\boldsymbol{x} = \mathbf{L}\boldsymbol{y} & \iff & \boldsymbol{x} - \boldsymbol{y} \in \mathscr{N}(\mathbf{L}) \\ \mathbf{L} \textit{ is } \text{INJECTIVE} & \iff & \mathscr{N}(\mathbf{L}) = \{\mathbb{0}\} \end{array}$$

✎PROOF:

1. Proof that $\mathbf{L}\boldsymbol{x} = \mathbf{L}\boldsymbol{y} \implies \boldsymbol{x} - \boldsymbol{y} \in \mathscr{N}(\mathbf{L})$:
$$\begin{aligned} \mathbf{L}(\boldsymbol{x} - \boldsymbol{y}) &= \mathbf{L}\boldsymbol{x} - \mathbf{L}\boldsymbol{y} && \text{by Theorem R.1 page 360} \\ &= \mathbb{0} && \text{by left hypothesis} \\ &\implies \boldsymbol{x} - \boldsymbol{y} \in \mathscr{N}(\mathbf{L}) && \text{by definition of } \textit{Null Space} \end{aligned}$$
2. Proof that $\mathbf{L}\boldsymbol{x} = \mathbf{L}\boldsymbol{y} \impliedby \boldsymbol{x} - \boldsymbol{y} \in \mathscr{N}(\mathbf{L})$:
$$\begin{aligned} \mathbf{L}\boldsymbol{y} &= \mathbf{L}\boldsymbol{y} + \mathbb{0} && \text{by definition of linear space (Definition J.1 page 247)} \\ &= \mathbf{L}\boldsymbol{y} + \mathbf{L}(\boldsymbol{x} - \boldsymbol{y}) && \text{by right hypothesis} \\ &= \mathbf{L}\boldsymbol{y} + (\mathbf{L}\boldsymbol{x} - \mathbf{L}\boldsymbol{y}) && \text{by Theorem R.1 page 360} \\ &= (\mathbf{L}\boldsymbol{y} - \mathbf{L}\boldsymbol{y}) + \mathbf{L}\boldsymbol{x} && \text{by } \textit{associative} \text{ and } \textit{commutative} \text{ properties (Definition J.1 page 247)} \\ &= \mathbf{L}\boldsymbol{x} \end{aligned}$$
3. Proof that $\mathbf{L}$ is *injective* $\iff \mathscr{N}(\mathbf{L}) = \{\mathbb{0}\}$:
$$\begin{aligned} \mathbf{L} \text{ is } \textit{injective} &\iff \{(\mathbf{L}\boldsymbol{x} = \mathbf{L}\boldsymbol{y} \iff \boldsymbol{x} = \boldsymbol{y}) \quad \forall \boldsymbol{x}, \boldsymbol{y} \in X\} \\ &\iff \left\{\left[\mathbf{L}\boldsymbol{x} - \mathbf{L}\boldsymbol{y} = \mathbb{0} \iff (\boldsymbol{x} - \boldsymbol{y}) = \mathbb{0}\right] \quad \forall \boldsymbol{x}, \boldsymbol{y} \in X\right\} \\ &\iff \left\{\left[\mathbf{L}(\boldsymbol{x} - \boldsymbol{y}) = \mathbb{0} \iff (\boldsymbol{x} - \boldsymbol{y}) = \mathbb{0}\right] \quad \forall \boldsymbol{x}, \boldsymbol{y} \in X\right\} \\ &\iff \mathscr{N}(\mathbf{L}) = \{\mathbb{0}\} \end{aligned}$$

[7] Eidelman et al. (2004) page 3
[8] Berberian (1961) page 88 ⟨Theorem IV.1.4⟩

✏

Theorem R.4. [9] *Let $\boldsymbol{W}$, $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{Z}$ be linear spaces over a field $\mathbb{F}$.*

1.	$\mathbf{L}(\mathbf{MN})$	$=$	$(\mathbf{LM})\mathbf{N}$	$\forall \mathbf{L}\in\mathscr{L}(\boldsymbol{Z},\boldsymbol{W}), \mathbf{M}\in\mathscr{L}(\boldsymbol{Y},\boldsymbol{Z}), \mathbf{N}\in\mathscr{L}(\boldsymbol{X},\boldsymbol{Y})$	(ASSOCIATIVE)
2.	$\mathbf{L}(\mathbf{M} \dotplus \mathbf{N})$	$=$	$(\mathbf{LM}) \dotplus (\mathbf{LN})$	$\forall \mathbf{L}\in\mathscr{L}(\boldsymbol{Y},\boldsymbol{Z}), \mathbf{M}\in\mathscr{L}(\boldsymbol{X},\boldsymbol{Y}), \mathbf{N}\in\mathscr{L}(\boldsymbol{X},\boldsymbol{Y})$	(LEFT DISTRIBUTIVE)
3.	$(\mathbf{L} \dotplus \mathbf{M})\mathbf{N}$	$=$	$(\mathbf{LN}) \dotplus (\mathbf{MN})$	$\forall \mathbf{L}\in\mathscr{L}(\boldsymbol{Y},\boldsymbol{Z}), \mathbf{M}\in\mathscr{L}(\boldsymbol{Y},\boldsymbol{Z}), \mathbf{N}\in\mathscr{L}(\boldsymbol{X},\boldsymbol{Y})$	(RIGHT DISTRIBUTIVE)
4.	$\alpha(\mathbf{LM})$	$=$	$(\alpha\mathbf{L})\mathbf{M} = \mathbf{L}(\alpha\mathbf{M})$	$\forall \mathbf{L}\in\mathscr{L}(\boldsymbol{Y},\boldsymbol{Z}), \mathbf{M}\in\mathscr{L}(\boldsymbol{X},\boldsymbol{Y}), \alpha\in\mathbb{F}$	(HOMOGENEOUS)

✎PROOF:

1. Proof that $\mathbf{L}(\mathbf{MN}) = (\mathbf{LM})\mathbf{N}$: Follows directly from property of *associative* operators.

2. Proof that $\mathbf{L}(\mathbf{M} \dotplus \mathbf{N}) = (\mathbf{LM}) \dotplus (\mathbf{LN})$:

$$\begin{aligned}
[\mathbf{L}(\mathbf{M} \dotplus \mathbf{N})]\boldsymbol{x} &= \mathbf{L}[(\mathbf{M} \dotplus \mathbf{N})\boldsymbol{x}] \\
&= \mathbf{L}[(\mathbf{M}\boldsymbol{x}) \dotplus (\mathbf{N}\boldsymbol{x})] \\
&= [\mathbf{L}(\mathbf{M}\boldsymbol{x})] \dotplus [\mathbf{L}(\mathbf{N}\boldsymbol{x})] && \text{by \textit{additive} property Definition R.3 page 360} \\
&= [(\mathbf{LM})\boldsymbol{x}] \dotplus [(\mathbf{LN})\boldsymbol{x}]
\end{aligned}$$

3. Proof that $(\mathbf{L} \dotplus \mathbf{M})\mathbf{N} = (\mathbf{LN}) \dotplus (\mathbf{MN})$: Follows directly from property of *associative* operators.

4. Proof that $\alpha(\mathbf{LM}) = (\alpha\mathbf{L})\mathbf{M}$: Follows directly from *associative* property of linear operators.

5. Proof that $\alpha(\mathbf{LM}) = \mathbf{L}(\alpha\mathbf{M})$:

$$\begin{aligned}
[\alpha(\mathbf{LM})]\boldsymbol{x} &= \alpha[(\mathbf{LM})\boldsymbol{x}] \\
&= \mathbf{L}[\alpha(\mathbf{M}\boldsymbol{x})] && \text{by \textit{homogeneous} property Definition R.3 page 360} \\
&= \mathbf{L}[(\alpha\mathbf{M})\boldsymbol{x}] \\
&= [\mathbf{L}(\alpha\mathbf{M})]\boldsymbol{x}
\end{aligned}$$

✏

Theorem R.5 (Fundamental theorem of linear equations)**.** [10] *Let $\boldsymbol{Y}^{\boldsymbol{X}}$ be the set of all operators from a linear space $\boldsymbol{X}$ to a linear space $\boldsymbol{Y}$. Let $\mathscr{N}(\mathbf{L})$ be the* NULL SPACE *of an operator $\mathbf{L}$ in $\boldsymbol{Y}^{\boldsymbol{X}}$ and $\mathscr{I}(\mathbf{L})$ the* IMAGE SET *of $\mathbf{L}$ in $\boldsymbol{Y}^{\boldsymbol{X}}$.*

$$\dim \mathscr{I}(\mathbf{L}) + \dim \mathscr{N}(\mathbf{L}) = \dim \boldsymbol{X} \qquad \forall \mathbf{L}\in\boldsymbol{Y}^{\boldsymbol{X}}$$

✎PROOF: Let $\{\psi_k | k = 1, 2, \dots, p\}$ be a basis for $\boldsymbol{X}$ constructed such that $\{\psi_{p-n+1}, \psi_{p-n+2}, \dots, \psi_p\}$ is a basis for

[9] 📘 Berberian (1961) page 88 ⟨Theorem IV.5.1⟩
[10] 📘 Michel and Herget (1993) page 99

$\mathscr{N}(\mathbf{L})$.

Let $p \triangleq \dim \boldsymbol{X}$.
Let $n \triangleq \dim \mathscr{N}(\mathbf{L})$.

$$\begin{aligned}
\dim \mathscr{I}(\mathbf{L}) &= \dim \{ \boldsymbol{y} \in \boldsymbol{Y} | \exists \boldsymbol{x} \in \boldsymbol{X} \quad \text{such that} \quad \boldsymbol{y} = \mathbf{L}\boldsymbol{x} \} \\
&= \dim \left\{ \boldsymbol{y} \in \boldsymbol{Y} | \exists (\alpha_1, \alpha_2, \dots, \alpha_p) \quad \text{such that} \quad \boldsymbol{y} = \mathbf{L} \sum_{k=1}^{p} \alpha_k \psi_k \right\} \\
&= \dim \left\{ \boldsymbol{y} \in \boldsymbol{Y} | \exists (\alpha_1, \alpha_2, \dots, \alpha_p) \quad \text{such that} \quad \boldsymbol{y} = \sum_{k=1}^{p} \alpha_k \mathbf{L} \psi_k \right\} \\
&= \dim \left\{ \boldsymbol{y} \in \boldsymbol{Y} | \exists (\alpha_1, \alpha_2, \dots, \alpha_p) \quad \text{such that} \quad \boldsymbol{y} = \sum_{k=1}^{p-n} \alpha_k \mathbf{L} \psi_k + \sum_{k=1}^{n} \alpha_k \mathbf{L} \psi_k \right\} \\
&= \dim \left\{ \boldsymbol{y} \in \boldsymbol{Y} | \exists (\alpha_1, \alpha_2, \dots, \alpha_p) \quad \text{such that} \quad \boldsymbol{y} = \sum_{k=1}^{p-n} \alpha_k \mathbf{L} \psi_k + \mathbb{0} \right\} \\
&= p - n \\
&= \dim \boldsymbol{X} - \dim \mathscr{N}(\mathbf{L})
\end{aligned}$$

Note: This "proof" may be missing some necessary detail. ✎

R.2 Operators on Normed linear spaces

R.2.1 Operator norm

Definition R.4. [11] *Let $\mathscr{L}(\boldsymbol{X}, \boldsymbol{Y})$ be the space of linear operators over normed linear spaces $\boldsymbol{X}$ and $\boldsymbol{Y}$.* [12]

DEF

*The **operator norm** $|||\cdot|||$ is defined as*

$$|||\mathbf{A}||| \triangleq \sup_{\boldsymbol{x} \in \boldsymbol{X}} \{ \|\mathbf{A}\boldsymbol{x}\| \mid \|\boldsymbol{x}\| \le 1 \} \qquad \forall \mathbf{A} \in \mathscr{L}(\boldsymbol{X}, \boldsymbol{Y})$$

*The pair $(\mathscr{L}(\boldsymbol{X}, \boldsymbol{Y}), |||\cdot|||)$ is the **normed space of linear operators** on $(\boldsymbol{X}, \boldsymbol{Y})$.*

Proposition R.1 (next) shows that the functional defined in Definition R.4 (previous) is a *norm* (Definition O.1 page 327).

Proposition R.1. [13] *Let $(\mathscr{L}(\boldsymbol{X}, \boldsymbol{Y}), |||\cdot|||)$ be the normed space of linear operators over the normed linear spaces $\boldsymbol{X} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$ and $\boldsymbol{Y} \triangleq (Y, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$.*

PRP

*The functional $|||\cdot|||$ is a **norm** on $\mathscr{L}(\boldsymbol{X}, \boldsymbol{Y})$. In particular,*

1.	$|||\mathbf{A}||| \ge 0$	$\forall \mathbf{A} \in \mathscr{L}(\boldsymbol{X}, \boldsymbol{Y})$	(NON-NEGATIVE)	*and*
2.	$|||\mathbf{A}||| = 0 \iff \mathbf{A} \doteq \mathbb{0}$	$\forall \mathbf{A} \in \mathscr{L}(\boldsymbol{X}, \boldsymbol{Y})$	(NONDEGENERATE)	*and*
3.	$|||\alpha\mathbf{A}||| = |\alpha| \, |||\mathbf{A}|||$	$\forall \mathbf{A} \in \mathscr{L}(\boldsymbol{X}, \boldsymbol{Y}), \alpha \in \mathbb{F}$	(HOMOGENEOUS)	*and*
4.	$\left|\left|\left| \mathbf{A} \dot{+} \mathbf{B} \right|\right|\right| \le |||\mathbf{A}||| + |||\mathbf{B}|||$	$\forall \mathbf{A} \in \mathscr{L}(\boldsymbol{X}, \boldsymbol{Y})$	(SUBADDITIVE).	

*Moreover, $(\mathscr{L}(\boldsymbol{X}, \boldsymbol{Y}), |||\cdot|||)$ is a **normed linear space**.*

[11] Rudin (1991) page 92, Aliprantis and Burkinshaw (1998) page 225

[12] The operator norm notation $|||\cdot|||$ is introduced (as a Matrix norm) in Horn and Johnson (1990) page 290

[13] Rudin (1991) page 93

✎PROOF:

1. Proof that $|||\mathbf{A}||| > 0$ for $\mathbf{A} \neq \mathbb{0}$:

$$\begin{aligned} |||\mathbf{A}||| &\triangleq \sup_{x\in\boldsymbol{X}} \{\|\mathbf{A}\boldsymbol{x}\| \mid \|\boldsymbol{x}\| \le 1\} && \text{by definition of } |||\cdot||| \text{ (Definition R.4 page 364)} \\ &> 0 \end{aligned}$$

2. Proof that $|||\mathbf{A}||| = 0$ for $\mathbf{A} \stackrel{\circ}{=} \mathbb{0}$:

$$\begin{aligned} |||\mathbf{A}||| &\triangleq \sup_{x\in\boldsymbol{X}} \{\|\mathbf{A}\boldsymbol{x}\| \mid \|\boldsymbol{x}\| \le 1\} && \text{by definition of } |||\cdot||| \text{ (Definition R.4 page 364)} \\ &= \sup_{x\in\boldsymbol{X}} \{\|\mathbb{0}\boldsymbol{x}\| \mid \|\boldsymbol{x}\| \le 1\} \\ &= 0 \end{aligned}$$

3. Proof that $|||\alpha\mathbf{A}||| = |\alpha|\; |||\mathbf{A}|||$:

$$\begin{aligned} |||\alpha\mathbf{A}||| &\triangleq \sup_{x\in\boldsymbol{X}} \{\|\alpha\mathbf{A}\boldsymbol{x}\| \mid \|\boldsymbol{x}\| \le 1\} && \text{by definition of } |||\cdot||| \text{ (Definition R.4 page 364)} \\ &= \sup_{x\in\boldsymbol{X}} \{|\alpha|\; \|\mathbf{A}\boldsymbol{x}\| \mid \|\boldsymbol{x}\| \le 1\} && \text{by definition of } |||\cdot||| \text{ (Definition R.4 page 364)} \\ &= |\alpha| \sup_{x\in\boldsymbol{X}} \{\|\mathbf{A}\boldsymbol{x}\| \mid \|\boldsymbol{x}\| \le 1\} && \text{by definition of sup} \\ &= |\alpha|\; |||\mathbf{A}||| && \text{by definition of } |||\cdot||| \text{ (Definition R.4 page 364)} \end{aligned}$$

4. Proof that $\big|\big|\big|\mathbf{A} \stackrel{\circ}{+} \mathbf{B}\big|\big|\big| \le |||\mathbf{A}||| + |||\mathbf{B}|||$:

$$\begin{aligned} \big|\big|\big|\mathbf{A} \stackrel{\circ}{+} \mathbf{B}\big|\big|\big| &\triangleq \sup_{x\in\boldsymbol{X}} \left\{\big\|(\mathbf{A} \stackrel{\circ}{+} \mathbf{B})\boldsymbol{x}\big\| \mid \|\boldsymbol{x}\| \le 1\right\} && \text{by definition of } |||\cdot||| \text{ (Definition R.4 page 364)} \\ &= \sup_{x\in\boldsymbol{X}} \{\|\mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{x}\| \mid \|\boldsymbol{x}\| \le 1\} \\ &\le \sup_{x\in\boldsymbol{X}} \{\|\mathbf{A}\boldsymbol{x}\| + \|\mathbf{B}\boldsymbol{x}\| \mid \|\boldsymbol{x}\| \le 1\} && \text{by definition of } |||\cdot||| \text{ (Definition R.4 page 364)} \\ &\le \sup_{x\in\boldsymbol{X}} \{\|\mathbf{A}\boldsymbol{x}\| \mid \|\boldsymbol{x}\| \le 1\} + \sup_{x\in\boldsymbol{X}} \{\|\mathbf{B}\boldsymbol{x}\| \mid \|\boldsymbol{x}\| \le 1\} \\ &\triangleq |||\mathbf{A}||| + |||\mathbf{B}||| && \text{by definition of } |||\cdot||| \text{ (Definition R.4 page 364)} \end{aligned}$$

⇨

Lemma R.1. *Let* $(\mathscr{L}(\boldsymbol{X}, \boldsymbol{Y}), |||\cdot|||)$ *be the normed space of linear operators over normed linear spaces* $\boldsymbol{X} \triangleq \big(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|\big)$ *and* $\boldsymbol{Y} \triangleq \big(Y, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|\big)$.

LEM

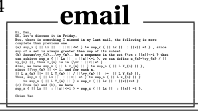
$$|||\mathbf{L}||| = \sup_{x} \{\|\mathbf{L}\boldsymbol{x}\| \mid \|\boldsymbol{x}\| = 1\} \qquad \forall \boldsymbol{x} \in \mathscr{L}(\boldsymbol{X}, \boldsymbol{Y})$$

✎PROOF: [14]

1. Proof that $\sup_x \{\|\mathbf{L}\boldsymbol{x}\| \mid \|\boldsymbol{x}\| \le 1\} \ge \sup_x \{\|\mathbf{L}\boldsymbol{x}\| \mid \|\boldsymbol{x}\| = 1\}$:

$$\sup_{x} \{\|\mathbf{L}\boldsymbol{x}\| \mid \|\boldsymbol{x}\| \le 1\} \ge \sup_{x} \{\|\mathbf{L}\boldsymbol{x}\| \mid \|\boldsymbol{x}\| = 1\} \qquad \text{because } A \subseteq B \implies \sup A \le \sup B$$

[14] email



Many many thanks to former NCTU Ph.D. student Chien Yao (Chinese: 姚建; PinYin: Yáo Jiàn) for his brilliant help with this proof. (If you are viewing this text as a pdf file, zoom in on the figure to the left to see text from Chien Yao's 2007 April 16 email.)

2. Let the subset $Y \subsetneq X$ be defined as

$$Y \triangleq \left\{ y \in X \,\middle|\, \begin{array}{ll} 1. & \|\mathbf{L}y\| = \sup_{x\in X}\{\|\mathbf{L}x\| \mid \|x\| \le 1\} \text{ and} \\ 2. & 0 < \|y\| \le 1 \end{array} \right\}$$

3. Proof that $\sup_x \{\|\mathbf{L}x\| \mid \|x\| \le 1\} \le \sup_x \{\|\mathbf{L}x\| \mid \|x\| = 1\}$:

$$\begin{aligned}
\sup_{x\in X}\{\|\mathbf{L}x\| \mid \|x\| \le 1\} &= \|\mathbf{L}y\| && \text{by definition of set } Y\\
&= \frac{\|y\|}{\|y\|}\|\mathbf{L}y\| \\
&= \|y\| \left\| \frac{1}{\|y\|}\mathbf{L}y \right\| && \text{by homogeneous property (page 327)}\\
&= \|y\| \left\| \mathbf{L}\frac{y}{\|y\|} \right\| && \text{by homogeneous property (page 360)}\\
&\le \|y\| \sup_{y\in Y} \left\{ \left\| \mathbf{L}\frac{y}{\|y\|} \right\| \right\} && \text{by definition of supremum}\\
&= \|y\| \sup_{y\in Y} \left\{ \left\| \mathbf{L}\frac{y}{\|y\|} \right\| \,\middle|\, \left\| \frac{y}{\|y\|} \right\| = 1 \right\} && \text{because } \left\| \frac{y}{\|y\|} \right\| = 1 \text{ for all } y \in Y\\
&\le \sup_{y\in Y} \left\{ \left\| \mathbf{L}\frac{y}{\|y\|} \right\| \,\middle|\, \left\| \frac{y}{\|y\|} \right\| = 1 \right\} && \text{because } 0 < \|y\| \le 1\\
&\le \sup_{x\in X}\{\|\mathbf{L}x\| \mid \|x\| = 1\} && \text{because } \frac{y}{\|y\|} \in X \quad \forall y \in Y
\end{aligned}$$

4. By (1) and (3),

$$\sup_{x\in X}\{\|\mathbf{L}x\| \mid \|x\| \le 1\} = \sup_{x\in X}\{\|\mathbf{L}x\| \mid \|x\| = 1\}$$

⇨

Proposition R.2. [15] *Let* $\mathbf{I}$ *be the identity operator in the normed space of linear operators* $(\mathscr{L}(\boldsymbol{X},\boldsymbol{X}), |||\cdot|||)$.

PRP	$|||\mathbf{I}||| = 1$

✎PROOF:

$$\begin{aligned}
|||\mathbf{I}||| &\triangleq \sup\{\|\mathbf{I}x\| \mid \|x\| \le 1\} && \text{by definition of } |||\cdot||| \text{ (Definition R.4 page 364)}\\
&= \sup\{\|x\| \mid \|x\| \le 1\} && \text{by definition of } \mathbf{I} \text{ (Definition R.2 page 359)}\\
&= 1
\end{aligned}$$

⇨

Theorem R.6. [16] *Let* $(\mathscr{L}(\boldsymbol{X},\boldsymbol{Y}), |||\cdot|||)$ *be the normed space of linear operators over normed linear spaces* $\boldsymbol{X}$ *and* $\boldsymbol{Y}$.

THM			
$\|\mathbf{L}x\|$	$\le$	$|||\mathbf{L}|||\ \|x\|$	$\forall \mathbf{L} \in \mathscr{L}(\boldsymbol{X},\boldsymbol{Y}),\ x \in \boldsymbol{X}$
$|||\mathbf{K}\mathbf{L}|||$	$\le$	$|||\mathbf{K}|||\ |||\mathbf{L}|||$	$\forall \mathbf{K},\mathbf{L} \in \mathscr{L}(\boldsymbol{X},\boldsymbol{Y})$

[15] Michel and Herget (1993) page 410
[16] Rudin (1991) page 103, Aliprantis and Burkinshaw (1998) page 225

✎PROOF:

1. Proof that $\|\mathbf{L}\boldsymbol{x}\| \le |||\mathbf{L}|||\ \|\boldsymbol{x}\|$:

$$\begin{aligned}
\|\mathbf{L}\boldsymbol{x}\| &= \frac{\|\boldsymbol{x}\|}{\|\boldsymbol{x}\|}\|\mathbf{L}\boldsymbol{x}\| \\
&= \|\boldsymbol{x}\|\left\|\frac{1}{\|\boldsymbol{x}\|}\mathbf{L}\boldsymbol{x}\right\| && \text{by property of norms} \\
&= \|\boldsymbol{x}\|\left\|\mathbf{L}\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}\right\| && \text{by property of linear operators} \\
&\triangleq \|\boldsymbol{x}\|\ \|\mathbf{L}\boldsymbol{y}\| && \text{where } y \triangleq \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|} \\
&\le \|\boldsymbol{x}\| \sup_{y} \|\mathbf{L}\boldsymbol{y}\| && \text{by definition of supremum} \\
&= \|\boldsymbol{x}\| \sup_{y} \{\|\mathbf{L}\boldsymbol{y}\| \mid \|\boldsymbol{y}\| = 1\} && \text{because } \|\boldsymbol{y}\| = \left\|\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}\right\| = \frac{\|\boldsymbol{x}\|}{\|\boldsymbol{x}\|} = 1 \\
&\triangleq \|\boldsymbol{x}\|\ \|\mathbf{L}\| && \text{by definition of operator norm}
\end{aligned}$$

2. Proof that $|||\mathbf{KL}||| \le |||\mathbf{K}|||\ |||\mathbf{L}|||$:

$$\begin{aligned}
|||\mathbf{KL}||| &\triangleq \sup_{x\in\boldsymbol{X}} \{\|(\mathbf{KL})\boldsymbol{x}\| \mid \|\boldsymbol{x}\| \le 1\} && \text{by Definition R.4 page 364 } (|||\cdot|||) \\
&= \sup_{x\in\boldsymbol{X}} \{\|\mathbf{K}(\mathbf{L}\boldsymbol{x})\| \mid \|\boldsymbol{x}\| \le 1\} \\
&\le \sup_{x\in\boldsymbol{X}} \{|||\mathbf{K}|||\ \|\mathbf{L}\boldsymbol{x}\| \mid \|\boldsymbol{x}\| \le 1\} && \text{by 1.} \\
&\le \sup_{x\in\boldsymbol{X}} \{|||\mathbf{K}|||\ |||\mathbf{L}|||\ \|\boldsymbol{x}\| \mid \|\boldsymbol{x}\| \le 1\} && \text{by 1.} \\
&= \sup_{x\in\boldsymbol{X}} \{|||\mathbf{K}|||\ |||\mathbf{L}|||\ 1 \mid \|\boldsymbol{x}\| \le 1\} && \text{by definition of sup} \\
&= |||\mathbf{K}|||\ |||\mathbf{L}||| && \text{by definition of sup}
\end{aligned}$$

⇰

R.2.2 Bounded linear operators

Definition R.5. [17] *Let $(\mathscr{L}(\boldsymbol{X},\boldsymbol{Y}), |||\cdot|||)$ be a normed space of linear operators.*

DEF
*An operator $\mathbf{B}$ is **bounded** if $|||\mathbf{B}||| < \infty$.*
*The quantity $\mathscr{B}(\boldsymbol{X},\boldsymbol{Y})$ is the set of all **bounded linear operators** on $(\boldsymbol{X},\boldsymbol{Y})$ such that*
$$\mathscr{B}(\boldsymbol{X},\boldsymbol{Y}) \triangleq \{\mathbf{L} \in \mathscr{L}(\boldsymbol{X},\boldsymbol{Y}) \mid |||\mathbf{L}||| < \infty\}.$$

Theorem R.7. [18] *Let $(\mathscr{L}(\boldsymbol{X},\boldsymbol{Y}), |||\cdot|||)$ be the set of linear operators over normed linear spaces $\boldsymbol{X} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$ and $\boldsymbol{Y} \triangleq (Y, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$.*

THM
The following conditions are all EQUIVALENT*:*

1.	$\mathbf{L}$ *is continuous at* A SINGLE POINT $\boldsymbol{x}_0 \in X$	$\forall \mathbf{L}\in\mathscr{L}(\boldsymbol{X},\boldsymbol{Y})$	$\iff$
2.	$\mathbf{L}$ *is* CONTINUOUS *(at every point $\boldsymbol{x} \in X$)*	$\forall \mathbf{L}\in\mathscr{L}(\boldsymbol{X},\boldsymbol{Y})$	$\iff$
3.	$|||\mathbf{L}||| < \infty$ *($\mathbf{L}$ is* BOUNDED*)*	$\forall \mathbf{L}\in\mathscr{L}(\boldsymbol{X},\boldsymbol{Y})$	$\iff$
4.	$\exists M \in \mathbb{R}$ such that $\|\mathbf{L}\boldsymbol{x}\| \le M\,\|\boldsymbol{x}\|$	$\forall \mathbf{L}\in\mathscr{L}(\boldsymbol{X},\boldsymbol{Y}), x\in X$	

[17] Rudin (1991) pages 92–93
[18] Aliprantis and Burkinshaw (1998) page 227

✎PROOF:

1. Proof that 1 $\implies$ 2:

$$\begin{aligned}
\epsilon &> \left\|\mathbf{L}\boldsymbol{x} - \mathbf{L}\boldsymbol{x}_0\right\| && \text{by hypothesis 1}\\
&= \left\|\mathbf{L}(\boldsymbol{x} - \boldsymbol{x}_0)\right\| && \text{by linearity (Definition R.3 page 360)}\\
&= \left\|\mathbf{L}(\boldsymbol{x} + \boldsymbol{y} - \boldsymbol{x}_0 - \boldsymbol{y})\right\| &&\\
&= \left\|\mathbf{L}(\boldsymbol{x} + \boldsymbol{y}) - \mathbf{L}(\boldsymbol{x}_0 + \boldsymbol{y})\right\| && \text{by linearity (Definition R.3 page 360)}\\
&\implies \mathbf{L} \text{ is continuous at point } \boldsymbol{x}+\boldsymbol{y} &&\\
&\implies \mathbf{L} \text{ is continuous at every point in } X && \text{(hypothesis 2)}
\end{aligned}$$

2. Proof that 2 $\implies$ 1: obvious:

3. Proof that 4 $\implies$ 2:[19]

$$\begin{aligned}
|||\mathbf{L}\boldsymbol{x}||| \le M\,\|\boldsymbol{x}\| &\implies |||\mathbf{L}(\boldsymbol{x}-\boldsymbol{y})||| \le M\,\|\boldsymbol{x}-\boldsymbol{y}\| && \text{by hypothesis 4}\\
&\implies |||\mathbf{L}\boldsymbol{x}-\mathbf{L}\boldsymbol{y}||| \le M\,\|\boldsymbol{x}-\boldsymbol{y}\| && \text{by linearity of } \mathbf{L} \text{ (Definition R.3 page 360)}\\
&\implies |||\mathbf{L}\boldsymbol{x}-\mathbf{L}\boldsymbol{y}||| \le \epsilon \text{ whenever } M\,\|\boldsymbol{x}-\boldsymbol{y}\| < \epsilon &&\\
&\implies |||\mathbf{L}\boldsymbol{x}-\mathbf{L}\boldsymbol{y}||| \le \epsilon \text{ whenever } \|\boldsymbol{x}-\boldsymbol{y}\| < \frac{\epsilon}{M} && \text{(hypothesis 2)}
\end{aligned}$$

4. Proof that 3 $\implies$ 4:

$$\begin{aligned}
\|\mathbf{L}\boldsymbol{x}\| &\le \underbrace{|||\mathbf{L}|||}_{M}\ \|\boldsymbol{x}\| && \text{by Theorem R.6 page 366}\\
&= M\ \|\boldsymbol{x}\| && \text{where } M \triangleq |||\mathbf{L}||| < \infty \text{ (by hypothesis 1)}
\end{aligned}$$

5. Proof that 1 $\implies$ 3:[20]

$$\begin{aligned}
|||\mathbf{L}||| = \infty &\implies \{\|\mathbf{L}\boldsymbol{x}\| \mid \|\boldsymbol{x}\| \le 1\} = \infty\\
&\implies \exists\,(\boldsymbol{x}_n) \quad \text{such that} \quad \left\|\boldsymbol{x}_n\right\| = 1 \text{ and } |||\mathbf{L}||| = \left\{\left\|\mathbf{L}\boldsymbol{x}_n\right\| \mid \left\|\boldsymbol{x}_n\right\| \le 1\right\} = \infty\\
&\implies \left\|\boldsymbol{x}_n\right\| = 1 \text{ and } \infty = |||\mathbf{L}||| = \left\|\mathbf{L}\boldsymbol{x}_n\right\|\\
&\implies \left\|\boldsymbol{x}_n\right\| = 1 \text{ and } \left\|\mathbf{L}\boldsymbol{x}_n\right\| \ge n\\
&\implies \frac{1}{n}\left\|\boldsymbol{x}_n\right\| = \frac{1}{n} \text{ and } \frac{1}{n}\left\|\mathbf{L}\boldsymbol{x}_n\right\| \ge 1\\
&\implies \left\|\frac{\boldsymbol{x}_n}{n}\right\| = \frac{1}{n} \text{ and } \left\|\mathbf{L}\frac{\boldsymbol{x}_n}{n}\right\| \ge 1\\
&\implies \lim_{n\to\infty}\left\|\frac{\boldsymbol{x}_n}{n}\right\| = 0 \text{ and } \lim_{n\to\infty}\left\|\mathbf{L}\frac{\boldsymbol{x}_n}{n}\right\| \ge 1\\
&\implies \mathbf{L} \text{ is not continuous at } 0
\end{aligned}$$

But by hypothesis, $\mathbf{L}$ *is* continuous. So the statement $|||\mathbf{L}||| = \infty$ must be *false* and thus $|||\mathbf{L}||| < \infty$ ($\mathbf{L}$ is *bounded*).

⇰

[19] Bollobás (1999) page 29
[20] Aliprantis and Burkinshaw (1998) page 227

R.2.3 Adjoints on normed linear spaces

Definition R.6. *Let $\mathscr{B}(\boldsymbol{X}, \boldsymbol{Y})$ be the space of* BOUNDED LINEAR OPERATORS *on normed linear spaces $\boldsymbol{X}$ and $\boldsymbol{Y}$. Let $\boldsymbol{X}^*$ be the* TOPOLOGICAL DUAL SPACE *of $\boldsymbol{X}$.*

DEF: $\mathbf{B}^*$ *is the* ***adjoint*** *of an operator $\mathbf{B} \in \mathscr{B}(\boldsymbol{X}, \boldsymbol{Y})$ if*

$$\mathsf{f}(\mathbf{B}\boldsymbol{x}) = \left[\mathbf{B}^*\mathsf{f}\right](\boldsymbol{x}) \qquad \forall \mathsf{f} \in \boldsymbol{X}^*,\ \boldsymbol{x} \in \boldsymbol{X}$$

Theorem R.8. [21] *Let $\mathscr{B}(\boldsymbol{X}, \boldsymbol{Y})$ be the space of* BOUNDED LINEAR OPERATORS *on* NORMED LINEAR SPACES *$\boldsymbol{X}$ and $\boldsymbol{Y}$.*

THM:

$$\begin{array}{rcll}
(\mathbf{A} \dotplus \mathbf{B})^* &=& \mathbf{A}^* \dotplus \mathbf{B}^* & \forall \mathbf{A},\mathbf{B} \in \mathscr{B}(\boldsymbol{X},\boldsymbol{Y}) \\
(\lambda\mathbf{A})^* &=& \lambda\mathbf{A}^* & \forall \mathbf{A},\mathbf{B} \in \mathscr{B}(\boldsymbol{X},\boldsymbol{Y}) \\
(\mathbf{A}\mathbf{B})^* &=& \mathbf{B}^*\mathbf{A}^* & \forall \mathbf{A},\mathbf{B} \in \mathscr{B}(\boldsymbol{X},\boldsymbol{Y})
\end{array}$$

✎PROOF:

$$\begin{aligned}
\left[\mathbf{A} \dotplus \mathbf{B}\right]^*\mathsf{f}(\boldsymbol{x}) &= \mathsf{f}\left(\left[\mathbf{A} \dotplus \mathbf{B}\right]\boldsymbol{x}\right) && \text{by definition of adjoint} && \text{(Definition R.6 page 369)} \\
&= \mathsf{f}(\mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{x}) && \text{by definition of linear operators} && \text{(Definition R.3 page 360)} \\
&= \mathsf{f}(\mathbf{A}\boldsymbol{x}) + \mathsf{f}(\mathbf{B}\boldsymbol{x}) && \text{by definition of } \textit{linear functional} && \\
&= \mathbf{A}^*\mathsf{f}(\boldsymbol{x}) + \mathbf{B}^*\mathsf{f}(\boldsymbol{x}) && \text{by definition of } \textit{adjoint} && \text{(Definition R.6 page 369)} \\
&= \left[\mathbf{A}^* + \mathbf{B}^*\right]\mathsf{f}(\boldsymbol{x}) && \text{by definition of } \textit{linear functional} &&
\end{aligned}$$

$$\begin{aligned}
[\lambda\mathbf{A}]^*\mathsf{f}(\boldsymbol{x}) &= \mathsf{f}([\lambda\mathbf{A}]\boldsymbol{x}) && \text{by definition of } \textit{adjoint} && \text{(Definition R.6 page 369)} \\
&= \lambda\mathsf{f}(\mathbf{A}\boldsymbol{x}) && \text{by definition of } \textit{linear functional} && \\
&= \left[\lambda\mathbf{A}^*\right]\mathsf{f}(\boldsymbol{x}) && \text{by definition of } \textit{adjoint} && \text{(Definition R.6 page 369)}
\end{aligned}$$

$$\begin{aligned}
[\mathbf{A}\mathbf{B}]^*\mathsf{f}(\boldsymbol{x}) &= \mathsf{f}([\mathbf{A}\mathbf{B}]\boldsymbol{x}) && \text{by definition of } \textit{adjoint} && \text{(Definition R.6 page 369)} \\
&= \mathsf{f}(\mathbf{A}[\mathbf{B}\boldsymbol{x}]) && \text{by definition of } \textit{linear operators} && \text{(Definition R.3 page 360)} \\
&= \left[\mathbf{A}^*\mathsf{f}\right](\mathbf{B}\boldsymbol{x}) && \text{by definition of } \textit{adjoint} && \text{(Definition R.6 page 369)} \\
&= \mathbf{B}^*\left[\mathbf{A}^*\mathsf{f}\right](\boldsymbol{x}) && \text{by definition of } \textit{adjoint} && \text{(Definition R.6 page 369)} \\
&= \left[\mathbf{B}^*\mathbf{A}^*\right]\mathsf{f}(\boldsymbol{x}) && \text{by definition of } \textit{adjoint} && \text{(Definition R.6 page 369)}
\end{aligned}$$

✏

Theorem R.9. [22] *Let $\mathscr{B}(\boldsymbol{X}, \boldsymbol{Y})$ be the space of* BOUNDED LINEAR OPERATORS *on normed linear spaces $\boldsymbol{X}$ and $\boldsymbol{Y}$. Let $\mathbf{B}^*$ be the adjoint of an operator $\mathbf{B}$.*

THM:

$$|||\mathbf{B}||| = \left|\left|\left|\mathbf{B}^*\right|\right|\right| \qquad \forall \mathbf{B} \in \mathscr{B}(\boldsymbol{X}, \boldsymbol{Y})$$

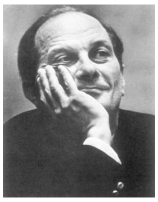
✎PROOF:

$$\begin{aligned}
|||\mathbf{B}||| &\triangleq \sup\{\|\mathbf{B}\boldsymbol{x}\| \mid \|\boldsymbol{x}\| \le 1\} && \text{by Definition R.4 page 364} \\
&\stackrel{?}{=} \sup\left\{|\mathsf{g}(\mathbf{B}\boldsymbol{x}; \boldsymbol{y}^*)|\mid \|\boldsymbol{x}\| \le 1,\ \left\|\boldsymbol{y}^*\right\| \le 1\right\} && \\
&= \sup\left\{|\mathsf{f}(\boldsymbol{x}; \mathbf{B}^*\boldsymbol{y}^*)|\mid \|\boldsymbol{x}\| \le 1,\ \left\|\boldsymbol{y}^*\right\| \le 1\right\} && \\
&\triangleq \sup\left\{\left\|\mathbf{B}^*\boldsymbol{y}^*\right\| \mid \|\boldsymbol{x}\| \le 1,\ \left\|\boldsymbol{y}^*\right\| \le 1\right\} && \\
&= \sup\left\{\left\|\mathbf{B}^*\boldsymbol{y}^*\right\| \mid \left\|\boldsymbol{y}^*\right\| \le 1\right\} && \\
&\triangleq \left|\left|\left|\mathbf{B}^*\right|\right|\right| && \text{by Definition R.4 page 364}
\end{aligned}$$

✏

[21] Bollobás (1999) page 156
[22] Rudin (1991) page 98

R.2.4 More properties



❝Beginning with the third year of studies, most of my mathematical work was really started in conversations with Mazur and Banach. And according to Banach some of my own contributions were characterized by a certain "strangeness" in the formulation of problems and in the outline of possible proofs. As he told me once some years later, he was surprised how often these "strange" approaches really worked.❞
Stanislaus M. Ulam (1909–1984), Polish mathematician [23]

Theorem R.10 (Mazur-Ulam theorem). [24] *Let $\phi \in \mathscr{L}(\boldsymbol{X}, \boldsymbol{Y})$ be a function on normed linear spaces $(\boldsymbol{X}, \|\cdot\|_{\boldsymbol{X}})$ and $(\boldsymbol{Y}, \|\cdot\|_{\boldsymbol{Y}})$. Let $\mathbf{I} \in \mathscr{L}(\boldsymbol{X}, \boldsymbol{X})$ be the identity operator on $(\boldsymbol{X}, \|\cdot\|_{\boldsymbol{X}})$.*

THM

$$\left.\begin{array}{ll} 1. & \underbrace{\phi^{-1}\phi = \phi\phi^{-1} = \mathbf{I}}_{\text{bijective}} \quad \text{and} \\ 2. & \underbrace{\|\phi\boldsymbol{x} - \phi\boldsymbol{y}\|_{\boldsymbol{Y}} = \|\boldsymbol{x} - \boldsymbol{y}\|_{\boldsymbol{X}} \quad \forall \boldsymbol{x},\boldsymbol{y} \in \boldsymbol{X}}_{\text{isometric}} \end{array}\right\} \implies \underbrace{\phi([1-\lambda]\boldsymbol{x} + \lambda\boldsymbol{y}) = [1-\lambda]\phi\boldsymbol{x} + \lambda\phi\boldsymbol{y} \forall\lambda \in \mathbb{R}}_{\text{affine}}$$

✎PROOF: Proof not yet complete.

1. Let ψ be the *reflection* of $\boldsymbol{z}$ in $\boldsymbol{X}$ such that $\psi\boldsymbol{x} = 2\boldsymbol{z} - \boldsymbol{x}$

   (a) $\|\psi\boldsymbol{x} - \boldsymbol{z}\| = \|\boldsymbol{x} - \boldsymbol{z}\|$

2. Let $\lambda \triangleq \sup_{g} \{\|g\boldsymbol{z} - \boldsymbol{z}\|\}$

3. Proof that $g \in W \implies g^{-1} \in W$:

   Let $\hat{\boldsymbol{x}} \triangleq g^{-1}\boldsymbol{x}$ and $\hat{\boldsymbol{y}} \triangleq g^{-1}\boldsymbol{y}$.

$$\begin{aligned} \left\|g^{-1}\boldsymbol{x} - g^{-1}\boldsymbol{y}\right\| &= \|\hat{\boldsymbol{x}} - \hat{\boldsymbol{y}}\| && \text{by definition of } \hat{\boldsymbol{x}} \text{ and } \hat{\boldsymbol{y}} \\ &= \|g\hat{\boldsymbol{x}} - g\hat{\boldsymbol{y}}\| && \text{by left hypothesis} \\ &= \left\|gg^{-1}\boldsymbol{x} - gg^{-1}\boldsymbol{y}\right\| && \text{by definition of } \hat{\boldsymbol{x}} \text{ and } \hat{\boldsymbol{y}} \\ &= \|\boldsymbol{x} - \boldsymbol{y}\| && \text{by definition of } g^{-1} \end{aligned}$$

[23] quote: Ulam (1991) page 33
image: http://www-history.mcs.st-andrews.ac.uk/Biographies/Ulam.html

[24] Oikhberg and Rosenthal (2007) page 598, Väisälä (2003) page 634, Giles (2000) page 11, Dunford and Schwartz (1957) page 91, Mazur and Ulam (1932)

4. Proof that $gz = z$:

$$\begin{aligned}
2\lambda &= 2\sup\{\|gz - z\|\} && \text{by definition of } \lambda \text{ item (2)}\\
&\le 2\|gz - z\| && \text{by definition of sup}\\
&= \|2z - 2gz\|\\
&= \|\psi gz - gz\| && \text{by definition of } \psi \text{ item (1)}\\
&= \left\|g^{-1}\psi gz - g^{-1}gz\right\| && \text{by item (3)}\\
&= \left\|g^{-1}\psi gz - z\right\| && \text{by definition of } g^{-1}\\
&= \left\|\psi g^{-1}\psi gz - z\right\|\\
&= \left\|g^{*}z - z\right\|\\
&\le \lambda && \text{by definition of } \lambda \text{ item (2)}\\
&\implies 2\lambda \le \lambda\\
&\implies \lambda = 0\\
&\implies gz = z
\end{aligned}$$

5. Proof that $\phi\left(\frac{1}{2}x + \frac{1}{2}y\right) = \frac{1}{2}\phi x + \frac{1}{2}\phi y$:

$$\begin{aligned}
\phi\left(\frac{1}{2}x + \frac{1}{2}y\right) &=\\
&= \frac{1}{2}\phi x + \frac{1}{2}\phi y
\end{aligned}$$

6. Proof that $\phi([1-\lambda]x + \lambda y) = [1-\lambda]\phi x + \lambda\phi y$:

$$\begin{aligned}
\phi([1-\lambda]x + \lambda y) &=\\
&= [1-\lambda]\phi x + \lambda\phi y
\end{aligned}$$

✏

Theorem R.11 (Neumann Expansion Theorem). [25] *Let* $\mathbf{A} \in \boldsymbol{X}^{\boldsymbol{X}}$ *be an operator on a linear space* $\boldsymbol{X}$. *Let* $\mathbf{A}^0 \triangleq \mathbf{I}$.

THM

$$\left.\begin{array}{ll}
1. & \mathbf{A} \in \mathscr{B}(\boldsymbol{X}, \boldsymbol{X}) \quad (\mathbf{A} \text{ is bounded})\\
2. & |||\mathbf{A}||| < 1
\end{array}\right\} \implies \left\{\begin{array}{llll}
1. & (\mathbf{I} - \mathbf{A})^{-1} & & \textit{exists}\\
2. & |||(\mathbf{I} - \mathbf{A})^{-1}||| & \le & \dfrac{1}{1 - |||\mathbf{A}|||}\\
3. & (\mathbf{I} - \mathbf{A})^{-1} & = & \displaystyle\sum_{n=0}^{\infty} \mathbf{A}^n\\
 & \multicolumn{3}{l}{\textit{with uniform convergence}}
\end{array}\right.$$

R.3 Operators on Inner product spaces

R.3.1 General Results

Theorem R.12. [26] *Let* $\mathbf{A}, \mathbf{B} \in \mathscr{B}(\boldsymbol{X}, \boldsymbol{X})$ *be* BOUNDED LINEAR OPERATORS *on an inner product space* $\boldsymbol{X} \triangleq \left(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle \mid \triangledown \rangle\right)$.

THM

$$\begin{array}{lllllll}
\langle \mathbf{B}x \mid x \rangle = 0 & \forall x \in X & \iff & \mathbf{B}x & = & \mathbb{0} & \forall x \in X\\
\langle \mathbf{A}x \mid x \rangle = \langle \mathbf{B}x \mid x \rangle & \forall x \in X & \iff & \mathbf{A} & = & \mathbf{B}
\end{array}$$

[25] Michel and Herget (1993) page 415

[26] Rudin (1991) page 310 ⟨Theorem 12.7, Corollary⟩

✎PROOF:

1. Proof that $\langle \mathbf{B}x \mid x\rangle = 0 \implies \mathbf{B}x = \mathbb{0}$:

$$\begin{aligned}
0 &= \langle \mathbf{B}(x+\mathbf{B}x) \mid (x+\mathbf{B}x)\rangle + i\,\langle \mathbf{B}(x+i\mathbf{B}x) \mid (x+i\mathbf{B}x)\rangle && \text{by left hypothesis}\\
&= \left\{\left\langle \mathbf{B}x+\mathbf{B}^2x) \mid x+\mathbf{B}x\right\rangle\right\} + i\left\{\left\langle \mathbf{B}x+i\mathbf{B}^2x) \mid x+i\mathbf{B}x\right\rangle\right\} && \text{by Definition R.3 page 360}\\
&= \left\{\langle \mathbf{B}x \mid x\rangle + \langle \mathbf{B}x \mid \mathbf{B}x\rangle + \left\langle \mathbf{B}^2x \mid x\right\rangle + \left\langle \mathbf{B}^2x \mid \mathbf{B}x\right\rangle\right\} && \text{by Definition N.1 page 309}\\
&\quad + i\left\{\langle \mathbf{B}x \mid x\rangle - i\,\langle \mathbf{B}x \mid \mathbf{B}x\rangle + i\left\langle \mathbf{B}^2x \mid x\right\rangle - i^2\left\langle \mathbf{B}^2x \mid \mathbf{B}x\right\rangle\right\}\\
&= \left\{0 + \langle \mathbf{B}x \mid \mathbf{B}x\rangle + \left\langle \mathbf{B}^2x \mid x\right\rangle + 0\right\} + i\left\{0 - i\,\langle \mathbf{B}x \mid \mathbf{B}x\rangle + i\left\langle \mathbf{B}^2x \mid x\right\rangle - i^2 0\right\} && \text{by left hypothesis}\\
&= \left\{\langle \mathbf{B}x \mid \mathbf{B}x\rangle + \left\langle \mathbf{B}^2x \mid x\right\rangle\right\} + \left\{\langle \mathbf{B}x \mid \mathbf{B}x\rangle - \left\langle \mathbf{B}^2x \mid x\right\rangle\right\}\\
&= 2\,\langle \mathbf{B}x \mid \mathbf{B}x\rangle\\
&= 2\,\|\mathbf{B}x\|^2\\
&\implies \mathbf{B}x = \mathbb{0} && \text{by Definition O.1 page 327}
\end{aligned}$$

2. Proof that $\langle \mathbf{B}x \mid x\rangle = 0 \impliedby \mathbf{B}x = \mathbb{0}$: by property of inner products (Theorem N.1 page 309).

3. Proof that $\langle \mathbf{A}x \mid x\rangle = \langle \mathbf{B}x \mid x\rangle \implies \mathbf{A} \doteq \mathbf{B}$:

$$\begin{aligned}
0 &= \langle \mathbf{A}x \mid x\rangle - \langle \mathbf{B}x \mid x\rangle && \text{by left hypothesis}\\
&= \langle \mathbf{A}x - \mathbf{B}x \mid x\rangle && \text{by } \textit{additivity} \text{ property of } \langle \triangle \mid \nabla\rangle \text{ (Definition N.1 page 309)}\\
&= \langle (\mathbf{A}-\mathbf{B})x \mid x\rangle && \text{by definition of operator addition}\\
\implies (\mathbf{A}-\mathbf{B})x &= \mathbb{0} && \text{by item 1}\\
\implies \mathbf{A} &= \mathbf{B} && \text{by definition of operator subtraction}
\end{aligned}$$

4. Proof that $\langle \mathbf{A}x \mid x\rangle = \langle \mathbf{B}x \mid x\rangle \impliedby \mathbf{A} \doteq \mathbf{B}$:

$$\langle \mathbf{A}x \mid x\rangle = \langle \mathbf{B}x \mid x\rangle \qquad \text{by } \mathbf{A} \doteq \mathbf{B} \text{ hypothesis}$$

✏

R.3.2 Operator adjoint

A fundamental concept of operators on inner product spaces is the *operator adjoint* (Proposition R.3 page 372). The adjoint of an operator is a kind of generalization of the conjugate of a complex number in that

- Both are *star-algebras* (Theorem R.13 page 373).
- Both support decomposition into "real" and "imaginary" parts (Theorem M.3 page 306).

Structurally, the operator adjoint provides a convenient symmetric relationship between the *range space* and *Null Space* of an operator (Theorem R.14 page 374).

Proposition R.3. [27] *Let $\mathscr{B}(\boldsymbol{H},\boldsymbol{H})$ be the space of* BOUNDED LINEAR OPERATORS *(Definition R.5 page 367) on a* HILBERT SPACE $\boldsymbol{H}$.

PRP: *An operator $\mathbf{B}^*$ is the **adjoint** of $\mathbf{B} \in \mathscr{B}(\boldsymbol{H},\boldsymbol{H})$ if*
$$\langle \mathbf{B}x \mid y\rangle = \langle x \mid \mathbf{B}^* y\rangle \qquad \forall x, y \in \boldsymbol{H}.$$

[27] Michel and Herget (1993) page 220, Rudin (1991) page 311, Giles (2000) page 182, von Neumann (1929) page 49, Stone (1932) page 41

✎PROOF:

1. For fixed y, $\mathsf{f}(\boldsymbol{x}) \triangleq \langle \boldsymbol{x} \mid \boldsymbol{y}\rangle$ is a *functional* in $\mathbb{F}^{\boldsymbol{X}}$.
2. $\mathbf{B}^*$ is the *adjoint* of $\mathbf{B}$ because

$$\begin{aligned}
\langle \mathbf{B}\boldsymbol{x} \mid \boldsymbol{y}\rangle &\triangleq \mathsf{f}(\mathbf{B}\boldsymbol{x}) \\
&\triangleq \mathbf{B}^*\mathsf{f}(\boldsymbol{x}) && \text{by definition of } \textit{operator adjoint} && \text{(Definition R.6 page 369)} \\
&= \langle \boldsymbol{x} \mid \mathbf{B}^*\boldsymbol{y}\rangle
\end{aligned}$$

✏

Example R.2.

EX

In matrix algebra ("linear algebra")

- The inner product operation $\langle \boldsymbol{x} \mid \boldsymbol{y}\rangle$ is represented by $\boldsymbol{y}^H\boldsymbol{x}$.
- The linear operator is represented as a matrix $\boldsymbol{A}$.
- The operation of $\boldsymbol{A}$ on a vector $\boldsymbol{x}$ is represented as $\boldsymbol{A}\boldsymbol{x}$.
- The adjoint of matrix $\boldsymbol{A}$ is the Hermitian matrix $\boldsymbol{A}^H$.

✎PROOF:

$$\langle \boldsymbol{A}\boldsymbol{x} \mid \boldsymbol{y}\rangle \triangleq \boldsymbol{y}^H\boldsymbol{A}\boldsymbol{x} = [(\boldsymbol{A}\boldsymbol{x})^H\boldsymbol{y}]^H = [\boldsymbol{x}^H\boldsymbol{A}^H\boldsymbol{y}]^H = (\boldsymbol{A}^H\boldsymbol{y})^H\boldsymbol{x} \triangleq \left\langle \boldsymbol{x} \mid \boldsymbol{A}^H\boldsymbol{y}\right\rangle$$

✏

Structures that satisfy the four conditions of the next theorem are known as **-algebras* ("*star-algebras*" (Definition M.3 page 304). Other structures which are *-algebras include the *field of complex numbers* $\mathbb{C}$ and any *ring of complex square* $n \times n$ *matrices*.[28]

Theorem R.13 (operator star-algebra). [29] *Let* $\boldsymbol{H}$ *be a* HILBERT SPACE *with operators* $\mathbf{A}, \mathbf{B} \in \mathscr{B}(\boldsymbol{H}, \boldsymbol{H})$ *and with adjoints* $\mathbf{A}^*, \mathbf{B}^* \in \mathscr{B}(\boldsymbol{H}, \boldsymbol{H})$. *Let* $\bar{\alpha}$ *be the complex conjugate of some* $\alpha \in \mathbb{C}$.

THM

The pair $(\boldsymbol{H}, *)$ *is a* $*$-ALGEBRA (STAR-ALGEBRA). *In particular,*

1.	$(\mathbf{A} \dotplus \mathbf{B})^*$	$=$	$\mathbf{A}^* + \mathbf{B}^*$	$\forall \mathbf{A},\mathbf{B} \in \boldsymbol{H}$	(DISTRIBUTIVE)	*and*
2.	$(\alpha\mathbf{A})^*$	$=$	$\bar{\alpha}\mathbf{A}^*$	$\forall \mathbf{A},\mathbf{B} \in \boldsymbol{H}$	(CONJUGATE LINEAR)	*and*
3.	$(\mathbf{A}\mathbf{B})^*$	$=$	$\mathbf{B}^*\mathbf{A}^*$	$\forall \mathbf{A},\mathbf{B} \in \boldsymbol{H}$	(ANTIAUTOMORPHIC)	*and*
4.	$\mathbf{A}^{**}$	$=$	$\mathbf{A}$	$\forall \mathbf{A},\mathbf{B} \in \boldsymbol{H}$	(INVOLUTARY)	

✎PROOF:

$$\begin{aligned}
\left\langle \boldsymbol{x} \mid (\mathbf{A} \dotplus \mathbf{B})^*\boldsymbol{y}\right\rangle &= \left\langle (\mathbf{A} \dotplus \mathbf{B})\boldsymbol{x} \mid \boldsymbol{y}\right\rangle && \text{by definition of adjoint} && \text{(Proposition R.3 page 372)} \\
&= \langle \mathbf{A}\boldsymbol{x} \mid \boldsymbol{y}\rangle + \langle \mathbf{B}\boldsymbol{x} \mid \boldsymbol{y}\rangle && \text{by definition of inner product} && \text{(Definition N.1 page 309)} \\
&= \langle \boldsymbol{x} \mid \mathbf{A}^*\boldsymbol{y}\rangle + \langle \boldsymbol{x} \mid \mathbf{B}^*\boldsymbol{y}\rangle && \text{by definition of operator addition} \\
&= \langle \boldsymbol{x} \mid \mathbf{A}^*\boldsymbol{y} + \mathbf{B}^*\boldsymbol{y}\rangle && \text{by definition of inner product} && \text{(Definition N.1 page 309)} \\
&= \langle \boldsymbol{x} \mid (\mathbf{A}^* + \mathbf{B}^*)\boldsymbol{y}\rangle && \text{by definition of operator addition}
\end{aligned}$$

$$\begin{aligned}
\langle \boldsymbol{x} \mid (\alpha\mathbf{A})^*\boldsymbol{y}\rangle &= \langle (\alpha\mathbf{A})\boldsymbol{x} \mid \boldsymbol{y}\rangle && \text{by definition of adjoint} && \text{(Proposition R.3 page 372)} \\
&= \langle \alpha(\mathbf{A}\boldsymbol{x}) \mid \boldsymbol{y}\rangle && \text{by definition of scalar multiplication} \\
&= \alpha\,\langle \mathbf{A}\boldsymbol{x} \mid \boldsymbol{y}\rangle && \text{by definition of inner product} && \text{(Definition N.1 page 309)}
\end{aligned}$$

[28] Sakai (1998) page 1
[29] Halmos (1998a) pages 39–40, Rudin (1991) page 311

$$
\begin{aligned}
&= \alpha \langle \boldsymbol{x} \,|\, \mathbf{A}^* \boldsymbol{y} \rangle && \text{by definition of adjoint} && \text{(Proposition R.3 page 372)}\\
&= \langle \boldsymbol{x} \,|\, \alpha^* \mathbf{A}^* \boldsymbol{y} \rangle && \text{by definition of inner product} && \text{(Definition N.1 page 309)}
\end{aligned}
$$

$$
\begin{aligned}
\langle \boldsymbol{x} \,|\, (\mathbf{AB})^* \boldsymbol{y} \rangle &= \langle (\mathbf{AB})\boldsymbol{x} \,|\, \boldsymbol{y} \rangle && \text{by definition of adjoint} && \text{(Proposition R.3 page 372)}\\
&= \langle \mathbf{A}(\mathbf{B}\boldsymbol{x}) \,|\, \boldsymbol{y} \rangle && \text{by definition of operator multiplication} &&\\
&= \langle (\mathbf{B}\boldsymbol{x}) \,|\, \mathbf{A}^* \boldsymbol{y} \rangle && \text{by definition of adjoint} && \text{(Proposition R.3 page 372)}\\
&= \langle \boldsymbol{x} \,|\, \mathbf{B}^* \mathbf{A}^* \boldsymbol{y} \rangle && \text{by definition of adjoint} && \text{(Proposition R.3 page 372)}
\end{aligned}
$$

$$
\begin{aligned}
\langle \boldsymbol{x} \,|\, \mathbf{A}^{**} \boldsymbol{y} \rangle &= \langle \mathbf{A}^* \boldsymbol{x} \,|\, \boldsymbol{y} \rangle && \text{by definition of adjoint} && \text{(Proposition R.3 page 372)}\\
&= \langle \boldsymbol{y} \,|\, \mathbf{A}^* \boldsymbol{x} \rangle^* && \text{by definition of inner product} && \text{(Definition N.1 page 309)}\\
&= \langle \mathbf{A}\boldsymbol{y} \,|\, \boldsymbol{x} \rangle^* && \text{by definition of adjoint} && \text{(Proposition R.3 page 372)}\\
&= \langle \boldsymbol{x} \,|\, \mathbf{A}\boldsymbol{y} \rangle && \text{by definition of inner product} && \text{(Definition N.1 page 309)}
\end{aligned}
$$

⇰

Theorem R.14. [30] *Let $\boldsymbol{Y}^{\boldsymbol{X}}$ be the set of all operators from a linear space $\boldsymbol{X}$ to a linear space $\boldsymbol{Y}$. Let $\mathscr{N}(\mathbf{L})$ be the* Null Space *of an operator $\mathbf{L}$ in $\boldsymbol{Y}^{\boldsymbol{X}}$ and $\mathscr{I}(\mathbf{L})$ the* image set *of $\mathbf{L}$ in $\boldsymbol{Y}^{\boldsymbol{X}}$.*

THM

$$
\boxed{\begin{aligned}
\mathscr{N}(\mathbf{A}) &= \mathscr{I}(\mathbf{A}^*)^{\perp}\\
\mathscr{N}(\mathbf{A}^*) &= \mathscr{I}(\mathbf{A})^{\perp}
\end{aligned}}
$$

✎Proof:

$$
\begin{aligned}
\mathscr{I}(\mathbf{A}^*)^{\perp} &= \left\{ y \in \boldsymbol{H} | \langle \boldsymbol{y} \,|\, \boldsymbol{u} \rangle = 0 \quad \forall u \in \mathscr{I}(\mathbf{A}^*) \right\} &&&&\\
&= \left\{ y \in \boldsymbol{H} | \langle \boldsymbol{y} \,|\, \mathbf{A}^* \boldsymbol{x} \rangle = 0 \quad \forall \boldsymbol{x} \in \boldsymbol{H} \right\} &&&&\\
&= \{ y \in \boldsymbol{H} | \langle \mathbf{A}\boldsymbol{y} \,|\, \boldsymbol{x} \rangle = 0 \quad \forall \boldsymbol{x} \in \boldsymbol{H} \} && \text{by definition of } \mathbf{A}^* && \text{(Proposition R.3 page 372)}\\
&= \{ y \in \boldsymbol{H} | \mathbf{A}\boldsymbol{y} = 0 \} &&&&\\
&= \mathscr{N}(\mathbf{A}) && \text{by definition of } \mathscr{N}(\mathbf{A}) &&
\end{aligned}
$$

$$
\begin{aligned}
\mathscr{I}(\mathbf{A})^{\perp} &= \{ y \in \boldsymbol{H} | \langle \boldsymbol{y} \,|\, \boldsymbol{u} \rangle = 0 \quad \forall u \in \mathscr{I}(\mathbf{A}) \} &&&&\\
&= \{ y \in \boldsymbol{H} | \langle \boldsymbol{y} \,|\, \mathbf{A}\boldsymbol{x} \rangle = 0 \quad \forall \boldsymbol{x} \in \boldsymbol{H} \} && \text{by definition of } \mathscr{I} &&\\
&= \left\{ y \in \boldsymbol{H} | \langle \mathbf{A}^* \boldsymbol{y} \,|\, \boldsymbol{x} \rangle = 0 \quad \forall \boldsymbol{x} \in \boldsymbol{H} \right\} && \text{by definition of } \mathbf{A}^* && \text{(Proposition R.3 page 372)}\\
&= \left\{ y \in \boldsymbol{H} | \mathbf{A}^* \boldsymbol{y} = 0 \right\} &&&&\\
&= \mathscr{N}(\mathbf{A}^*) && \text{by definition of } \mathscr{N}(\mathbf{A}) &&
\end{aligned}
$$

⇰

R.4 Special Classes of Operators

R.4.1 Projection operators

Definition R.7. [31] *Let $\mathscr{B}(\boldsymbol{X}, \boldsymbol{Y})$ be the space of* bounded linear operators *on normed linear spaces $\boldsymbol{X}$ and $\boldsymbol{Y}$. Let $\mathbf{P}$ be a bounded linear operator in $\mathscr{B}(\boldsymbol{X}, \boldsymbol{Y})$.*

[30] Rudin (1991) page 312

[31] Rudin (1991) page 126 ⟨5.15 Projections⟩, Kubrusly (2001) page 70, Bachman and Narici (1966) page 26, Halmos (1958) page 73 ⟨§41. Projections⟩

DEF $\mathbf{P}$ *is a* ***projection*** *operator if* $\quad \mathbf{P}^2 = \mathbf{P}$.

Theorem R.15. [32] *Let* $\mathscr{B}(\boldsymbol{X}, \boldsymbol{Y})$ *be the space of* BOUNDED LINEAR OPERATORS *on normed linear spaces* $\boldsymbol{X}$ *and* $\boldsymbol{Y}$. *Let* $\mathbf{P}$ *be a bounded linear operator in* $\mathscr{B}(\boldsymbol{X}, \boldsymbol{Y})$ *with* NULL SPACE $\mathscr{N}(\mathbf{P})$ *and* IMAGE SET $\mathscr{I}(\mathbf{P})$.

THM
$$\left.\begin{array}{lrcll} 1. & \mathbf{P}^2 & = & \mathbf{P} & (\mathbf{P}\text{ is a projection operator}) \quad \text{and}\\ 2. & \boldsymbol{\Omega} & = & \boldsymbol{X} \dotplus \boldsymbol{Y} & (\boldsymbol{Y}\text{ compliments }\boldsymbol{X}\text{ in }\boldsymbol{\Omega}) \quad \text{and}\\ 3. & \mathbf{P}\boldsymbol{\Omega} & = & \boldsymbol{X} & (\mathbf{P}\text{ projects onto }\boldsymbol{X}) \end{array}\right\} \implies \left\{\begin{array}{lrcll} 1. & \mathscr{I}(\mathbf{P}) & = & \boldsymbol{X} & \text{and}\\ 2. & \mathscr{N}(\mathbf{P}) & = & \boldsymbol{Y} & \text{and}\\ 3. & \boldsymbol{\Omega} & = & \mathscr{I}(\mathbf{P}) \dotplus \mathscr{N}(\mathbf{P}) & \end{array}\right.$$

✎PROOF:

$$\begin{aligned}
\mathscr{I}(\mathbf{P}) &= \mathbf{P}\boldsymbol{\Omega}\\
&= \mathbf{P}(\boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_2)\\
&= \mathbf{P}\boldsymbol{\Omega}_1 + \mathbf{P}\boldsymbol{\Omega}_2\\
&= \boldsymbol{\Omega}_1 + \{\mathbb{0}\}\\
&= \boldsymbol{\Omega}_1
\end{aligned}$$

$$\begin{aligned}
\mathscr{N}(\mathbf{P}) &= \{x \in \boldsymbol{\Omega} | \mathbf{P}x = \mathbb{0}\}\\
&= \left\{x \in (\boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_2) | \mathbf{P}x = \mathbb{0}\right\}\\
&= \left\{x \in \boldsymbol{\Omega}_1 | \mathbf{P}x = \mathbb{0}\right\} + \left\{x \in \boldsymbol{\Omega}_2 | \mathbf{P}x = \mathbb{0}\right\}\\
&= \{\mathbb{0}\} + \boldsymbol{\Omega}_2\\
&= \boldsymbol{\Omega}_2
\end{aligned}$$

✎

Theorem R.16. [33] *Let* $\mathscr{B}(\boldsymbol{X}, \boldsymbol{Y})$ *be the space of* BOUNDED LINEAR OPERATORS *on normed linear spaces* $\boldsymbol{X}$ *and* $\boldsymbol{Y}$. *Let* $\mathbf{P}$ *be a bounded linear operator in* $\mathscr{B}(\boldsymbol{X}, \boldsymbol{Y})$.

THM
$$\underbrace{\mathbf{P}^2 = \mathbf{P}}_{\mathbf{P}\text{ is a projection operator}} \iff \underbrace{(\mathbf{I}-\mathbf{P})^2 = (\mathbf{I}-\mathbf{P})}_{(\mathbf{I}-\mathbf{P})\text{ is a projection operator}}$$

✎PROOF:

Proof that $\mathbf{P}^2 = \mathbf{P} \implies (\mathbf{I}-\mathbf{P})^2 = (\mathbf{I}-\mathbf{P})$:

$$\begin{aligned}
(\mathbf{I}-\mathbf{P})^2 &= (\mathbf{I}-\mathbf{P})(\mathbf{I}-\mathbf{P})\\
&= \mathbf{I}(\mathbf{I}-\mathbf{P}) + (-\mathbf{P})(\mathbf{I}-\mathbf{P})\\
&= \mathbf{I} - \mathbf{P} - \mathbf{P}\mathbf{I} + \mathbf{P}^2\\
&= \mathbf{I} - \mathbf{P} - \mathbf{P} + \mathbf{P} && \text{by left hypothesis}\\
&= \mathbf{I} - \mathbf{P}
\end{aligned}$$

Proof that $\mathbf{P}^2 = \mathbf{P} \impliedby (\mathbf{I}-\mathbf{P})^2 = (\mathbf{I}-\mathbf{P})$:

$$\begin{aligned}
\mathbf{P}^2 &= \underbrace{\mathbf{I} - \mathbf{P} - \mathbf{P} + \mathbf{P}^2}_{(\mathbf{I}-\mathbf{P})^2} - (\mathbf{I} - \mathbf{P} - \mathbf{P})\\
&= (\mathbf{I}-\mathbf{P})^2 - (\mathbf{I} - \mathbf{P} - \mathbf{P})\\
&= (\mathbf{I}-\mathbf{P}) - (\mathbf{I} - \mathbf{P} - \mathbf{P}) && \text{by right hypothesis}\\
&= \mathbf{P}
\end{aligned}$$

[32] Michel and Herget (1993) pages 120–121
[33] Michel and Herget (1993) page 121

Theorem R.17. [34] *Let $\boldsymbol{H}$ be a* HILBERT SPACE *and* $\mathbf{P}$ *an operator in* $\boldsymbol{H}^{\boldsymbol{H}}$ *with adjoint* $\mathbf{P}^*$, NULL SPACE $\mathcal{N}(\mathbf{P})$, *and* IMAGE SET $\mathcal{I}(\mathbf{P})$.

THM

If $\mathbf{P}$ *is a* PROJECTION OPERATOR, *then the following are equivalent:*

$$\begin{array}{llll}
1. & \mathbf{P}^* = \mathbf{P} & (\mathbf{P} \text{ is SELF-ADJOINT}) & \iff \\
2. & \mathbf{P}^*\mathbf{P} = \mathbf{P}\mathbf{P}^* & (\mathbf{P} \text{ is NORMAL}) & \iff \\
3. & \mathcal{I}(\mathbf{P}) = \mathcal{N}(\mathbf{P})^\perp & & \iff \\
4. & \langle \mathbf{P}\boldsymbol{x} \,|\, \boldsymbol{x} \rangle = \|\mathbf{P}\boldsymbol{x}\|^2 \quad \forall x \in X & &
\end{array}$$

✎PROOF: This proof is incomplete at this time.

Proof that (1) $\implies$ (2):

$$\begin{aligned}
\mathbf{P}^*\mathbf{P} &= \mathbf{P}^{**}\mathbf{P}^* && \text{by (1)} \\
&= \mathbf{P}\mathbf{P}^* && \text{by Theorem R.13 page 373}
\end{aligned}$$

Proof that (1) $\implies$ (3):

$$\begin{aligned}
\mathcal{I}(\mathbf{P}) &= \mathcal{N}(\mathbf{P}^*)^\perp && \text{by Theorem R.14 page 374} \\
&= \mathcal{N}(\mathbf{P})^\perp && \text{by (1)}
\end{aligned}$$

Proof that (3) $\implies$ (4):

Proof that (4) $\implies$ (1):

R.4.2 Self Adjoint Operators

Definition R.8. [35] *Let* $\mathbf{B} \in \mathcal{B}(\boldsymbol{H}, \boldsymbol{H})$ *be a* BOUNDED *operator with adjoint* $\mathbf{B}^*$ *on a* HILBERT SPACE $\boldsymbol{H}$.

DEF

The operator $\mathbf{B}$ *is said to be* ***self-adjoint*** *or* ***hermitian*** *if* $\mathbf{B} \stackrel{\circ}{=} \mathbf{B}^*$.

Example R.3 (Autocorrelation operator). Let $\mathsf{x}(t)$ be a random process with autocorrelation
$$\mathsf{R}_{\mathsf{xx}}(t,u) \triangleq \underbrace{\mathbf{E}[\mathsf{x}(t)\mathsf{x}^*(u)]}_{\text{expectation}}.$$

Let an autocorrelation operator $\mathbf{R}$ be defined as $[\mathbf{R}\mathsf{f}](t) \triangleq \displaystyle\int_{\mathbb{R}} \underbrace{\mathsf{R}_{\mathsf{xx}}(t,u)}_{\text{kernel}}\mathsf{f}(u)\,\mathsf{d}u$.

EX

$\mathbf{R} = \mathbf{R}^*$ (The auto-correlation operator $\mathbf{R}$ is *self-adjoint*)

[34] Rudin (1991) page 314

[35] Historical works regarding self-adjoint operators: von Neumann (1929) page 49, “linearer Operator R selbstadjungiert oder Hermitesch”, Stone (1932) page 50 ⟨“self-adjoint transformations”⟩

Theorem R.18. [36] *Let* $\mathbf{S} : \boldsymbol{H} \to \boldsymbol{H}$ *be an operator over a* HILBERT SPACE $\boldsymbol{H}$ *with eigenvalues* $\{\lambda_n\}$ *and eigenfunctions* $\{\psi_n\}$ *such that* $\mathbf{S}\psi_n = \lambda_n\psi_n$ *and let* $\|\mathbf{x}\| \triangleq \sqrt{\langle \boldsymbol{x} \,|\, \boldsymbol{x}\rangle}$.

THM

$$\left\{\underbrace{\mathbf{S} = \mathbf{S}^*}_{\mathbf{S}\text{ is self adjoint}}\right\} \implies \left\{\begin{array}{lll} 1. & \langle \mathbf{S}\boldsymbol{x} \,|\, \boldsymbol{x}\rangle \in \mathbb{R} & \text{(the hermitian quadratic form of } \mathbf{S} \text{ is REAL-VALUED)} \\ 2. & \lambda_n \in \mathbb{R} & \text{(eigenvalues of } \mathbf{S} \text{ are REAL-VALUED)} \\ 3. & \lambda_n \neq \lambda_m \implies \langle \psi_n \,|\, \psi_m\rangle = 0 & \text{(eigenvectors are ORTHOGONAL)} \end{array}\right\}$$

✎PROOF:

1. Proof that $\mathbf{S} = \mathbf{S}^* \implies \langle \mathbf{S}\boldsymbol{x} \,|\, \boldsymbol{x}\rangle \in \mathbb{R}$:

$$\begin{aligned}
\langle \boldsymbol{x} \,|\, \mathbf{S}\boldsymbol{x}\rangle &= \langle \mathbf{S}\boldsymbol{x} \,|\, \boldsymbol{x}\rangle && \text{by left hypothesis} \\
&= \langle \boldsymbol{x} \,|\, \mathbf{S}\boldsymbol{x}\rangle^* && \text{by definition of } \langle \triangle \,|\, \nabla\rangle \text{ Definition N.1 page 309}
\end{aligned}$$

2. Proof that $\mathbf{S} = \mathbf{S}^* \implies \lambda_n \in \mathbb{R}$:

$$\begin{aligned}
\lambda_n \|\psi_n\|^2 &= \lambda_n \langle \psi_n \,|\, \psi_n\rangle && \text{by definition} \\
&= \langle \lambda_n\psi_n \,|\, \psi_n\rangle && \text{by definition of } \langle \triangle \,|\, \nabla\rangle \text{ Definition N.1 page 309} \\
&= \langle \mathbf{S}\psi_n \,|\, \psi_n\rangle && \text{by definition of eigenpairs} \\
&= \langle \psi_n \,|\, \mathbf{S}\psi_n\rangle && \text{by left hypothesis} \\
&= \langle \psi_n \,|\, \lambda_n\psi_n\rangle && \text{by definition of eigenpairs} \\
&= \lambda_n^* \langle \psi_n \,|\, \psi_n\rangle && \text{by definition of } \langle \triangle \,|\, \nabla\rangle \text{ Definition N.1 page 309} \\
&= \lambda_n^* \|\psi_n\|^2 && \text{by definition}
\end{aligned}$$

3. Proof that $\mathbf{S} = \mathbf{S}^* \implies [\lambda_n \neq \lambda_m \implies \langle \psi_n \,|\, \psi_m\rangle = 0]$:

$$\begin{aligned}
\lambda_n \langle \psi_n \,|\, \psi_m\rangle &= \langle \lambda_n\psi_n \,|\, \psi_m\rangle && \text{by definition of } \langle \triangle \,|\, \nabla\rangle \text{ Definition N.1 page 309} \\
&= \langle \mathbf{S}\psi_n \,|\, \psi_m\rangle && \text{by definition of eigenpairs} \\
&= \langle \psi_n \,|\, \mathbf{S}\psi_m\rangle && \text{by left hypothesis} \\
&= \langle \psi_n \,|\, \lambda_m\psi_m\rangle && \text{by definition of eigenpairs} \\
&= \lambda_m^* \langle \psi_n \,|\, \psi_m\rangle && \text{by definition of } \langle \triangle \,|\, \nabla\rangle \text{ Definition N.1 page 309} \\
&= \lambda_m \langle \psi_n \,|\, \psi_m\rangle && \text{because } \lambda_m \text{ is real}
\end{aligned}$$

This implies for $\lambda_n \neq \lambda_m \neq 0$, $\langle \psi_n \,|\, \psi_m\rangle = 0$.

✏

R.4.3 Normal Operators

Definition R.9. [37] *Let* $\mathscr{B}(\boldsymbol{X}, \boldsymbol{Y})$ *be the space of* BOUNDED LINEAR OPERATORS *on normed linear spaces* $\boldsymbol{X}$ *and* $\boldsymbol{Y}$. *Let* $\mathbf{N}^*$ *be the adjoint of an operator* $\mathbf{N} \in \mathscr{B}(\boldsymbol{X}, \boldsymbol{Y})$.

DEF

$\mathbf{N}$ ***is normal*** *if* $\mathbf{N}^*\mathbf{N} = \mathbf{N}\mathbf{N}^*$.

[36] Lax (2002) pages 315–316, Keener (1988) pages 114–119, Bachman and Narici (1966) page 24 ⟨Theorem 2.1⟩, Bertero and Boccacci (1998) page 225 ⟨§"9.2 SVD of a matrix …If all eigenvectors are normalized…"⟩

[37] Rudin (1991) page 312, Michel and Herget (1993) page 431, Dieudonné (1969) page 167, Frobenius (1878), Frobenius (1968) page 391

Theorem R.19. [38] *Let* $\mathscr{B}(\boldsymbol{H},\boldsymbol{H})$ *be the space of* BOUNDED LINEAR OPERATORS *on a* HILBERT SPACE $\boldsymbol{H}$. *Let* $\mathscr{N}(\mathbf{N})$ *be the* NULL SPACE *of an operator* $\mathbf{N}$ *in* $\mathscr{B}(\boldsymbol{H},\boldsymbol{H})$ *and* $\mathscr{I}(\mathbf{N})$ *the* IMAGE SET *of* $\mathbf{N}$ *in* $\mathscr{B}(\boldsymbol{H},\boldsymbol{H})$.

THM
$$\underbrace{\mathbf{N}^*\mathbf{N} = \mathbf{N}\mathbf{N}^*}_{\mathbf{N} \text{ is normal}} \iff \|\mathbf{N}^*\boldsymbol{x}\| = \|\mathbf{N}\boldsymbol{x}\| \qquad \forall \boldsymbol{x} \in \boldsymbol{H}$$

✎PROOF:

1. Proof that $\mathbf{N}^*\mathbf{N} = \mathbf{N}\mathbf{N}^* \implies \|\mathbf{N}^*\boldsymbol{x}\| = \|\mathbf{N}\boldsymbol{x}\|$:

$$\begin{aligned}
\|\mathbf{N}\boldsymbol{x}\|^2 &= \langle \mathbf{N}\boldsymbol{x} \,|\, \mathbf{N}\boldsymbol{x}\rangle && \text{by definition}\\
&= \langle \boldsymbol{x} \,|\, \mathbf{N}^*\mathbf{N}\boldsymbol{x}\rangle && \text{by Proposition R.3 page 372 (definition of } \mathbf{N}^*)\\
&= \langle \boldsymbol{x} \,|\, \mathbf{N}\mathbf{N}^*\boldsymbol{x}\rangle && \text{by left hypothesis (} \mathbf{N} \text{ is normal)}\\
&= \langle \mathbf{N}\boldsymbol{x} \,|\, \mathbf{N}^*\boldsymbol{x}\rangle && \text{by Proposition R.3 page 372 (definition of } \mathbf{N}^*)\\
&= \|\mathbf{N}^*\boldsymbol{x}\|^2 && \text{by definition}
\end{aligned}$$

2. Proof that $\mathbf{N}^*\mathbf{N} = \mathbf{N}\mathbf{N}^* \impliedby \|\mathbf{N}^*\boldsymbol{x}\| = \|\mathbf{N}\boldsymbol{x}\|$:

$$\begin{aligned}
\langle \mathbf{N}^*\mathbf{N}\boldsymbol{x} \,|\, \boldsymbol{x}\rangle &= \langle \mathbf{N}\boldsymbol{x} \,|\, \mathbf{N}^{**}\boldsymbol{x}\rangle && \text{by Proposition R.3 page 372 (definition of } \mathbf{N}^*)\\
&= \langle \mathbf{N}\boldsymbol{x} \,|\, \mathbf{N}\boldsymbol{x}\rangle && \text{by Theorem R.13 page 373 (property of adjoint)}\\
&= \|\mathbf{N}\boldsymbol{x}\|^2 && \text{by definition}\\
&= \|\mathbf{N}^*\boldsymbol{x}\|^2 && \text{by right hypothesis (} \|\mathbf{N}^*\boldsymbol{x}\| = \|\mathbf{N}\boldsymbol{x}\|)\\
&= \langle \mathbf{N}^*\boldsymbol{x} \,|\, \mathbf{N}^*\boldsymbol{x}\rangle && \text{by definition}\\
&= \langle \mathbf{N}\mathbf{N}^*\boldsymbol{x} \,|\, \boldsymbol{x}\rangle && \text{by Proposition R.3 page 372 (definition of } \mathbf{N}^*)
\end{aligned}$$

✏

Theorem R.20. [39] *Let* $\mathscr{B}(\boldsymbol{H},\boldsymbol{H})$ *be the space of* BOUNDED LINEAR OPERATORS *on a* HILBERT SPACE $\boldsymbol{H}$. *Let* $\mathscr{N}(\mathbf{N})$ *be the* NULL SPACE *of an operator* $\mathbf{N}$ *in* $\mathscr{B}(\boldsymbol{H},\boldsymbol{H})$ *and* $\mathscr{I}(\mathbf{N})$ *the* IMAGE SET *of* $\mathbf{N}$ *in* $\mathscr{B}(\boldsymbol{H},\boldsymbol{H})$.

THM
$$\underbrace{\mathbf{N}^*\mathbf{N} = \mathbf{N}\mathbf{N}^*}_{\mathbf{N} \text{ is normal}} \implies \underbrace{\mathscr{N}(\mathbf{N}^*) = \mathscr{N}(\mathbf{N})}_{\mathbf{N} \text{ and } \mathbf{N}^* \text{ have the same Null Space}}$$

✎PROOF:

$$\begin{aligned}
\mathscr{N}(\mathbf{N}^*) &= \left\{\boldsymbol{x} | \mathbf{N}^*\boldsymbol{x} = 0 \quad \forall \boldsymbol{x} \in \boldsymbol{X}\right\} && \text{by definition of } \textit{Null Space}\\
&= \left\{\boldsymbol{x} | \, \|\mathbf{N}^*\boldsymbol{x}\| = 0 \quad \forall \boldsymbol{x} \in \boldsymbol{X}\right\} && \text{by definition of } \|\cdot\| \text{ (Definition O.1 page 327)}\\
&= \left\{\boldsymbol{x} | \, \|\mathbf{N}\boldsymbol{x}\| = 0 \quad \forall \boldsymbol{x} \in \boldsymbol{X}\right\} && \\
&= \left\{\boldsymbol{x} | \mathbf{N}\boldsymbol{x} = 0 \quad \forall \boldsymbol{x} \in \boldsymbol{X}\right\} && \text{by definition of } \|\cdot\| \text{ (Definition O.1 page 327)}\\
&= \mathscr{N}(\mathbf{N}) && \text{by definition of } \textit{Null Space } \mathscr{N}
\end{aligned}$$

✏

Theorem R.21. [40] *Let* $\mathscr{B}(\boldsymbol{H},\boldsymbol{H})$ *be the space of* BOUNDED LINEAR OPERATORS *on a* HILBERT SPACE $\boldsymbol{H}$. *Let* $\mathscr{N}(\mathbf{N})$ *be the* NULL SPACE *of an operator* $\mathbf{N}$ *in* $\mathscr{B}(\boldsymbol{H},\boldsymbol{H})$ *and* $\mathscr{I}(\mathbf{N})$ *the* IMAGE SET *of* $\mathbf{N}$ *in* $\mathscr{B}(\boldsymbol{H},\boldsymbol{H})$.

THM
$$\left\{\underbrace{\mathbf{N}^*\mathbf{N} = \mathbf{N}\mathbf{N}^*}_{\mathbf{N} \text{ is normal}}\right\} \implies \left\{\underbrace{\lambda_n \neq \lambda_m \implies \langle \psi_n \,|\, \psi_m\rangle = 0}_{\text{eigenfunctions associated with distinct eigenvalues are orthogonal}}\right\}$$

[38] Rudin (1991) pages 312–313
[39] Rudin (1991) pages 312–313
[40] Rudin (1991) pages 312–313

✎PROOF: The proof in (1) is flawed. This implies that (2) is also flawed. Rudin (1991) page 313 claims both to be true.

1. Proof that $\mathbf{N}^*\mathbf{N} = \mathbf{N}\mathbf{N}^* \implies \mathbf{N}^*\psi = \lambda^*\psi$:

$$\begin{aligned}
&\mathbf{N}\psi = \lambda\psi \\
&\Longleftarrow \\
&\quad 0 = \mathcal{N}(\mathbf{N} - \lambda\mathbf{I}) \\
&\quad\quad = \mathcal{N}([\mathbf{N} - \lambda\mathbf{I}]^*) && \text{by } \mathcal{N}(\mathbf{N}) = \mathcal{N}(\mathbf{N}^*) \\
&\quad\quad = \mathcal{N}(\mathbf{N}^* - [\lambda\mathbf{I}]^*) && \text{by Theorem R.13 page 373} \\
&\quad\quad = \mathcal{N}(\mathbf{N}^* - \lambda^*\mathbf{I}^*) && \text{by Theorem R.13 page 373} \\
&\quad\quad = \mathcal{N}(\mathbf{N}^* - \lambda^*\mathbf{I}) \\
&\Longrightarrow \\
&(\mathbf{N}^* - \lambda^*\mathbf{I})\psi = 0 \\
&\iff \mathbf{N}^*\psi = \lambda^*\psi
\end{aligned}$$

2. Proof that $\mathbf{N}^*\mathbf{N} = \mathbf{N}\mathbf{N}^* \implies [\lambda_n \neq \lambda_m \implies \langle \psi_n \,|\, \psi_m \rangle = 0]$:

$$\begin{aligned}
\lambda_n \langle \psi_n \,|\, \psi_m \rangle &= \langle \lambda_n\psi_n \,|\, \psi_m \rangle && \text{by definition of } \langle \triangle \,|\, \nabla \rangle \text{ Definition N.1 page 309} \\
&= \langle \mathbf{N}\psi_n \,|\, \psi_m \rangle && \text{by definition of eigenpairs} \\
&= \langle \psi_n \,|\, \mathbf{N}^*\psi_m \rangle && \text{by Proposition R.3 page 372 (definition of adjoint)} \\
&= \langle \psi_n \,|\, \lambda_m^*\psi_m \rangle && \text{by (4.)} \\
&= \lambda_m \langle \psi_n \,|\, \psi_m \rangle && \text{by definition of } \langle \triangle \,|\, \nabla \rangle \text{ Definition N.1 page 309}
\end{aligned}$$

This implies for $\lambda_n \neq \lambda_m \neq 0$, $\langle \psi_n \,|\, \psi_m \rangle = 0$.

⇨

R.4.4 Isometric operators

An operator on a pair of normed linear spaces is *isometric* (next definition) if it is an *isometry*.

Definition R.10. *Let* $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$ *and* $(Y, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$ *be* NORMED LINEAR SPACES *(Definition O.1 page 327).*

DEF: *An operator* $\mathbf{M} \in \mathcal{L}(\boldsymbol{X}, \boldsymbol{Y})$ *is* ***isometric*** *if*

$$\|\mathbf{M}x\| = \|x\| \qquad \forall x \in X.$$

Theorem R.22. [41] *Let* $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$ *and* $(Y, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$ *be* NORMED LINEAR SPACES. *Let* $\mathbf{M}$ *be a linear operator in* $\mathcal{L}(\boldsymbol{X}, \boldsymbol{Y})$.

THM:
$$\underbrace{\|\mathbf{M}x\| = \|x\| \quad \forall x \in X}_{\textit{isometric in length}} \iff \underbrace{\|\mathbf{M}x - \mathbf{M}y\| = \|x - y\| \quad \forall x,y \in X}_{\textit{isometric in distance}}$$

✎PROOF:

[41] Kubrusly (2001) page 239 ⟨Proposition 4.37⟩, Berberian (1961) page 27 ⟨Theorem IV.7.5⟩

1. Proof that $\|\mathbf{M}x\| = \|x\| \implies \|\mathbf{M}x - \mathbf{M}y\| = \|x - y\|$:

$$\begin{aligned}
\|\mathbf{M}x - \mathbf{M}y\| &= \|\mathbf{M}(x - y)\| && \text{by definition of linear operators (Definition R.3 page 360)}\\
&= \|\mathbf{M}u\| && \text{let } u \triangleq x - y\\
&= \|x - y\| && \text{by left hypothesis}
\end{aligned}$$

2. Proof that $\|\mathbf{M}x\| = \|x\| \impliedby \|\mathbf{M}x - \mathbf{M}y\| = \|x - y\|$:

$$\begin{aligned}
\|\mathbf{M}x\| &= \|\mathbf{M}(x - \mathbb{0})\|\\
&= \|\mathbf{M}x - \mathbf{M}\mathbb{0}\| && \text{by definition of linear operators (Definition R.3 page 360)}\\
&= \|x - \mathbb{0}\| && \text{by right hypothesis}\\
&= \|x\|
\end{aligned}$$

⇨

Isometric operators have already been defined (Definition R.10 page 379) in the more general normed linear spaces, while Theorem R.22 (page 379) demonstrated that in a normed linear space $\boldsymbol{X}$, $\|\mathbf{M}x\| = \|x\| \iff \|\mathbf{M}x - \mathbf{M}y\| = \|x - y\|$ for all $x, y \in \boldsymbol{X}$. Here in the more specialized inner product spaces, Theorem R.23 (next) demonstrates two additional equivalent properties.

Theorem R.23. [42] *Let $\mathscr{B}(\boldsymbol{X}, \boldsymbol{X})$ be the space of* BOUNDED LINEAR OPERATORS *on a normed linear space $\boldsymbol{X} \triangleq \left(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|\right)$. Let $\mathbf{N}$ be a bounded linear operator in $\mathscr{L}(\boldsymbol{X}, \boldsymbol{X})$, and $\mathbf{I}$ the identity operator in $\mathscr{L}(\boldsymbol{X}, \boldsymbol{X})$. Let $\|x\| \triangleq \sqrt{\langle x \,|\, x\rangle}$.*

THM *The following conditions are all* ***equivalent****:*

$$\begin{array}{llcllll}
1. & \mathbf{M}^*\mathbf{M} & = & \mathbf{I} & & & \iff\\
2. & \langle \mathbf{M}x \,|\, \mathbf{M}y\rangle & = & \langle x \,|\, y\rangle & \forall x,y\in X & (\mathbf{M}\text{ is surjective}) & \iff\\
3. & \|\mathbf{M}x - \mathbf{M}y\| & = & \|x - y\| & \forall x,y\in X & (\text{isometric in distance}) & \iff\\
4. & \|\mathbf{M}x\| & = & \|x\| & \forall x\in X & (\text{isometric in length}) &
\end{array}$$

✎PROOF:

1. Proof that (1) $\implies$ (2):

$$\begin{aligned}
\langle \mathbf{M}x \,|\, \mathbf{M}y\rangle &= \langle x \,|\, \mathbf{M}^*\mathbf{M}y\rangle && \text{by Proposition R.3 page 372 (definition of adjoint)}\\
&= \langle x \,|\, \mathbf{I}y\rangle && \text{by (1)}\\
&= \langle x \,|\, y\rangle && \text{by Definition R.2 page 359 (definition of } \mathbf{I})
\end{aligned}$$

2. Proof that (2) $\implies$ (4):

$$\begin{aligned}
\|\mathbf{M}x\| &= \sqrt{\langle \mathbf{M}x \,|\, \mathbf{M}x\rangle} && \text{by definition of } \|\cdot\|\\
&= \sqrt{\langle x \,|\, x\rangle} && \text{by right hypothesis}\\
&= \|x\| && \text{by definition of } \|\cdot\|
\end{aligned}$$

3. Proof that (2) $\impliedby$ (4):

$$\begin{aligned}
4\langle \mathbf{M}x \,|\, \mathbf{M}y\rangle &= \|\mathbf{M}x + \mathbf{M}y\|^2 - \|\mathbf{M}x - \mathbf{M}y\|^2 + i\,\|\mathbf{M}x + i\mathbf{M}y\|^2 - i\,\|\mathbf{M}x - i\mathbf{M}y\|^2 && \text{by polarization id.}\\
&= \|\mathbf{M}(x + y)\|^2 - \|\mathbf{M}(x - y)\|^2 + i\,\|\mathbf{M}(x + iy)\|^2 - i\,\|\mathbf{M}(x - iy)\|^2 && \text{by Definition R.3}\\
&= \|x + y\|^2 - \|x - y\|^2 + i\,\|x + iy\|^2 - i\,\|x - iy\|^2 && \text{by left hypothesis}
\end{aligned}$$

[42] Michel and Herget (1993) page 432 ⟨Theorem 7.5.8⟩, Kubrusly (2001) page 391 ⟨Proposition 5.72⟩

4. Proof that (3) $\iff$ (4): by Theorem R.22 page 379

5. Proof that (4) $\implies$ (1):

$$\begin{aligned}
\langle \mathbf{M}^*\mathbf{M}\boldsymbol{x} \,|\, \boldsymbol{x}\rangle &= \langle \mathbf{M}\boldsymbol{x} \,|\, \mathbf{M}^{**}\boldsymbol{x}\rangle && \text{by Proposition R.3 page 372 (definition of adjoint)}\\
&= \langle \mathbf{M}\boldsymbol{x} \,|\, \mathbf{M}\boldsymbol{x}\rangle && \text{by Theorem R.13 page 373 (property of adjoint)}\\
&= \|\mathbf{M}\boldsymbol{x}\|^2 && \text{by definition}\\
&= \|\boldsymbol{x}\|^2 && \text{by left hypothesis with } y = 0\\
&= \langle \boldsymbol{x} \,|\, \boldsymbol{x}\rangle && \text{by definition}\\
&= \langle \mathbf{I}\boldsymbol{x} \,|\, \boldsymbol{x}\rangle && \text{by Definition R.2 page 359 (definition of } \mathbf{I})\\
\implies \qquad \mathbf{M}^*\mathbf{M} &= \mathbf{I} && \forall \boldsymbol{x} \in X
\end{aligned}$$

✏

Theorem R.24. [43] *Let $\mathscr{B}(\boldsymbol{X}, \boldsymbol{Y})$ be the space of* BOUNDED LINEAR OPERATORS *on normed linear spaces $\boldsymbol{X}$ and $\boldsymbol{Y}$. Let $\mathbf{M}$ be a bounded linear operator in $\mathscr{B}(\boldsymbol{X}, \boldsymbol{Y})$, and $\mathbf{I}$ the identity operator in $\mathscr{L}(\boldsymbol{X}, \boldsymbol{X})$. Let Λ be the set of eigenvalues of $\mathbf{M}$. Let $\|\boldsymbol{x}\| \triangleq \sqrt{\langle \boldsymbol{x} \,|\, \boldsymbol{x}\rangle}$.*

THM

$$\underbrace{\mathbf{M}^*\mathbf{M} = \mathbf{I}}_{\mathbf{M} \text{ is isometric}} \implies \left\{\begin{array}{lcll} |||\mathbf{M}||| &=& 1 & \text{(UNIT LENGTH)} \quad \textit{and}\\ |\lambda| &=& 1 & \forall \lambda \in \Lambda \end{array}\right.$$

✎PROOF:

1. Proof that $\mathbf{M}^*\mathbf{M} = \mathbf{I} \implies |||\mathbf{M}||| = 1$:

$$\begin{aligned}
|||\mathbf{M}||| &= \sup_{\boldsymbol{x}\in X} \{\|\mathbf{M}\boldsymbol{x}\| \,|\, \|\boldsymbol{x}\| = 1\} && \text{by Definition R.4 page 364}\\
&= \sup_{\boldsymbol{x}\in X} \{\|\boldsymbol{x}\| \,|\, \|\boldsymbol{x}\| = 1\} && \text{by Theorem R.23 page 380}\\
&= \sup_{\boldsymbol{x}\in X}\{1\}\\
&= 1
\end{aligned}$$

2. Proof that $|\lambda| = 1$: Let $(\boldsymbol{x}, \lambda)$ be an eigenvector-eigenvalue pair.

$$\begin{aligned}
1 &= \frac{1}{\|\boldsymbol{x}\|} \|\boldsymbol{x}\|\\
&= \frac{1}{\|\boldsymbol{x}\|} \|\mathbf{M}\boldsymbol{x}\| && \text{by Theorem R.23 page 380}\\
&= \frac{1}{\|\boldsymbol{x}\|} \|\lambda\boldsymbol{x}\| && \text{by definition of } \lambda\\
&= \frac{1}{\|\boldsymbol{x}\|} |\lambda| \; \|\boldsymbol{x}\| && \text{by homogeneous property of } \|\cdot\|\\
&= |\lambda|
\end{aligned}$$

✏

Example R.4 (One sided shift operator). [44] Let $\boldsymbol{X}$ be the set of all sequences with range $\mathbb{W}$ $(0, 1, 2, \ldots)$ and shift operators defined as

$$\begin{array}{llcll}
1. & \mathbf{S}_r\,(\!(x_0, x_1, x_2, \ldots)\!) & \triangleq & (\!(0, x_0, x_1, x_2, \ldots)\!) & \text{(right shift operator)}\\
2. & \mathbf{S}_l\,(\!(x_0, x_1, x_2, \ldots)\!) & \triangleq & (\!(x_1, x_2, x_3, \ldots)\!) & \text{(left shift operator)}
\end{array}$$

[43] Michel and Herget (1993) page 432
[44] Michel and Herget (1993) page 441

EX
1. $\mathbf{S}_r$ is an isometric operator.
2. $\mathbf{S}_r^* = \mathbf{S}_l$

✎PROOF:

1. Proof that $\mathbf{S}_r^* = \mathbf{S}_l$:

$$\begin{aligned}
\langle \mathbf{S}_r (\!(x_0, x_1, x_2, \dots)\!) \,|\, (\!(y_0, y_1, y_2, \dots)\!) \rangle &= \langle (\!(0, x_0, x_1, x_2, \dots)\!) \,|\, (\!(y_0, y_1, y_2, \dots)\!) \rangle \\
&= \sum_{n=1}^{\infty} x_{n-1}\, y_n^* \\
&= \sum_{n=0}^{\infty} x_n\, y_{n+1}^* \\
&= \sum_{n=0}^{\infty} x_n\, y_{n+1}^* \\
&= \langle (\!(x_0, x_1, x_2, \dots)\!) \,|\, (\!(y_1, y_2, y_3, \dots)\!) \rangle \\
&= \left\langle (\!(x_0, x_1, x_2, \dots)\!) \,\middle|\, \underbrace{\mathbf{S}_l}_{\mathbf{S}_r^*} (\!(y_0, y_1, y_2, \dots)\!) \right\rangle
\end{aligned}$$

2. Proof that $\mathbf{S}_r$ is isometric ($\mathbf{S}_r^*\mathbf{S}_r = \mathbf{I}$):

$$\begin{aligned}
\mathbf{S}_r^*\mathbf{S}_r &= \mathbf{S}_l\mathbf{S}_r && \text{by 1.} \\
&= \mathbf{I}
\end{aligned}$$

✏

R.4.5 Unitary operators

Definition R.11. [45] *Let $\mathscr{B}(\boldsymbol{X}, \boldsymbol{Y})$ be the space of* BOUNDED LINEAR OPERATORS *on normed linear spaces $\boldsymbol{X}$ and $\boldsymbol{Y}$. Let $\mathbf{U}$ be a bounded linear operator in $\mathscr{B}(\boldsymbol{X}, \boldsymbol{Y})$, and $\boldsymbol{I}$ the identity operator in $\mathscr{B}(\boldsymbol{X}, \boldsymbol{X})$.*

DEF *The operator $\mathbf{U}$ is* ***unitary*** *if* $\mathbf{U}^*\mathbf{U} = \mathbf{U}\mathbf{U}^* = \mathbf{I}$.

Proposition R.4. *Let $\mathscr{B}(\boldsymbol{X}, \boldsymbol{Y})$ be the space of* BOUNDED LINEAR OPERATORS *on normed linear spaces $\boldsymbol{X}$ and $\boldsymbol{Y}$. Let $\mathbf{U}$ and $\mathbf{V}$ be* BOUNDED LINEAR OPERATORS *in $\mathscr{B}(\boldsymbol{X}, \boldsymbol{Y})$.*

PRP $\left.\begin{array}{l} \mathbf{U} \text{ is UNITARY} \quad \text{and} \\ \mathbf{V} \text{ is UNITARY} \end{array}\right\} \implies (\mathbf{UV}) \text{ is UNITARY.}$

[45] Rudin (1991) page 312, Michel and Herget (1993) page 431, Autonne (1901) page 209, Autonne (1902), Schur (1909), Steen (1973)

✎PROOF:

$$\begin{aligned}
(\mathbf{UV})(\mathbf{UV})^* &= (\mathbf{UV})(\mathbf{V}^*\mathbf{U}^*) && \text{by Theorem R.8 page 369} \\
&= \mathbf{U}(\mathbf{VV}^*)\mathbf{U}^* && \text{by associative property} \\
&= \mathbf{UIU}^* && \text{by definition of } \textit{unitary} \text{ operators} && \text{(Definition R.11 page 382)} \\
&= \mathbf{I} && \text{by definition of } \textit{unitary} \text{ operators} && \text{(Definition R.11 page 382)}
\end{aligned}$$

$$\begin{aligned}
(\mathbf{UV})^*(\mathbf{UV}) &= (\mathbf{V}^*\mathbf{U}^*)(\mathbf{UV}) && \text{by Theorem R.8 page 369} \\
&= \mathbf{V}^*(\mathbf{U}^*\mathbf{U})\mathbf{V} && \text{by associative property} \\
&= \mathbf{V}^*\mathbf{IV} && \text{by definition of } \textit{unitary} \text{ operators} && \text{(Definition R.11 page 382)} \\
&= \mathbf{I} && \text{by definition of } \textit{unitary} \text{ operators} && \text{(Definition R.11 page 382)}
\end{aligned}$$

✏

Theorem R.25. [46] *Let $\mathscr{B}(\boldsymbol{H},\boldsymbol{H})$ be the space of* BOUNDED LINEAR OPERATORS *on a* HILBERT SPACE $\boldsymbol{H}$. *Let $\mathscr{I}(\mathbf{U})$ be the* IMAGE SET *of* $\mathbf{U}$.

THM

If $\mathbf{U}$ *is a* ***bounded linear operator*** ($\mathbf{U}\in\mathscr{B}(\boldsymbol{H},\boldsymbol{H})$), *then the following conditions are* ***equivalent***:

1. $\mathbf{UU}^* = \mathbf{U}^*\mathbf{U} = \mathbf{I}$ (UNITARY) $\iff$
2. $\langle \mathbf{U}\boldsymbol{x} \,|\, \mathbf{U}\boldsymbol{y}\rangle = \langle \mathbf{U}^*\boldsymbol{x} \,|\, \mathbf{U}^*\boldsymbol{y}\rangle = \langle \boldsymbol{x} \,|\, \boldsymbol{y}\rangle$ *and* $\mathscr{I}(\mathbf{U}) = X$ (SURJECTIVE) $\iff$
3. $\|\mathbf{U}\boldsymbol{x} - \mathbf{U}\boldsymbol{y}\| = \|\mathbf{U}^*\boldsymbol{x} - \mathbf{U}^*\boldsymbol{y}\| = \|\boldsymbol{x}-\boldsymbol{y}\|$ *and* $\mathscr{I}(\mathbf{U}) = X$ (ISOMETRIC IN DISTANCE) $\iff$
4. $\|\mathbf{U}\boldsymbol{x}\| = \|\boldsymbol{x}\|$ *and* $\mathscr{I}(\mathbf{U}) = X$ (ISOMETRIC IN LENGTH)

✎PROOF:

1. Proof that (1) $\implies$ (2):
   (a) $\langle \mathbf{U}\boldsymbol{x} \,|\, \mathbf{U}\boldsymbol{y}\rangle = \langle \mathbf{U}^*\boldsymbol{x} \,|\, \mathbf{U}^*\boldsymbol{y}\rangle = \langle \boldsymbol{x} \,|\, \boldsymbol{y}\rangle$ by Theorem R.23 (page 380).
   (b) Proof that $\mathscr{I}(\mathbf{U}) = X$:

$$\begin{aligned}
X &\supseteq \mathscr{I}(\mathbf{U}) && \text{because } \mathbf{U}\in X^X \\
&\supseteq \mathscr{I}(\mathbf{UU}^*) \\
&= \mathscr{I}(\mathbf{I}) && \text{by left hypothesis } (\mathbf{U}^*\mathbf{U} = \mathbf{UU}^* = \mathbf{I}) \\
&= X && \text{by Definition R.2 page 359 (definition of } \mathbf{I})
\end{aligned}$$

2. Proof that (2) $\iff$ (3) $\iff$ (4): by Theorem R.23 page 380.
3. Proof that (3) $\implies$ (1):
   (a) Proof that $\|\mathbf{U}\boldsymbol{x}-\mathbf{U}\boldsymbol{y}\| = \|\boldsymbol{x}-\boldsymbol{y}\| \implies \mathbf{U}^*\mathbf{U} = \mathbf{I}$: by Theorem R.23 page 380
   (b) Proof that $\|\mathbf{U}^*\boldsymbol{x}-\mathbf{U}^*\boldsymbol{y}\| = \|\boldsymbol{x}-\boldsymbol{y}\| \implies \mathbf{UU}^* = \mathbf{I}$:

$$\begin{aligned}
\|\mathbf{U}^*\boldsymbol{x}-\mathbf{U}^*\boldsymbol{y}\| = \|\boldsymbol{x}-\boldsymbol{y}\| \implies \quad \mathbf{U}^{**}\mathbf{U}^* &= \mathbf{I} && \text{by Theorem R.23 page 380} \\
\mathbf{UU}^* &= \mathbf{I} && \text{by Theorem R.13 page 373}
\end{aligned}$$

✏

Theorem R.26. *Let $\mathscr{B}(\boldsymbol{H},\boldsymbol{H})$ be the space of* BOUNDED LINEAR OPERATORS *on a* HILBERT SPACE $\boldsymbol{H}$. *Let* $\mathbf{U}$ *be a bounded linear operator in* $\mathscr{B}(\boldsymbol{H},\boldsymbol{H})$, $\mathscr{N}(\mathbf{U})$ *the* NULL SPACE *of* $\mathbf{U}$, *and* $\mathscr{I}(\mathbf{U})$ *the* IMAGE SET

[46] Rudin (1991) pages 313–314 ⟨Theorem 12.13⟩, Knapp (2005a) page 45 ⟨Proposition 2.6⟩

of $\mathbf{U}$.

THM
$$\underbrace{\mathbf{U}\mathbf{U}^* = \mathbf{U}^*\mathbf{U} = \mathbf{I}}_{\mathbf{U} \textit{ is unitary}} \implies \left\{\begin{array}{lllll} \mathbf{U}^{-1} & = & \mathbf{U}^* & & \textit{and} \\ \mathscr{I}(\mathbf{U}) & = & \mathscr{I}(\mathbf{U}^*) & = X & \textit{and} \\ \mathscr{N}(\mathbf{U}) & = & \mathscr{N}(\mathbf{U}^*) & = \{\mathbb{0}\} & \textit{and} \\ |||\mathbf{U}||| & = & |||\mathbf{U}^*||| & = 1 & \text{(UNIT LENGTH)} \end{array}\right\}$$

✎PROOF:

1. Note that $\mathbf{U}$, $\mathbf{U}^*$, and $\mathbf{U}^{-1}$ are all both *isometric* and *normal*:

$$\begin{array}{lllllll} \mathbf{U}^*\mathbf{U} & & & = & \mathbf{I} & \implies & \mathbf{U} \text{ is } \textit{isometric} \\ \mathbf{U}\mathbf{U}^* & = & \mathbf{U}^*\mathbf{U} & = & \mathbf{I} & \implies & \mathbf{U}^* \text{ is } \textit{isometric} \\ \mathbf{U}^{-1} & = & \mathbf{U}^* & & & \implies & \mathbf{U}^{-1} \text{ is } \textit{isometric} \\ & & & & & & \\ \mathbf{U}^*\mathbf{U} & = & \mathbf{U}\mathbf{U}^* & = & & \implies & \mathbf{U} \text{ is } \textit{normal} \\ \mathbf{U}\mathbf{U}^* & = & \mathbf{U}^*\mathbf{U} & = & & \implies & \mathbf{U}^* \text{ is } \textit{normal} \\ \mathbf{U}^{-1} & = & \mathbf{U}^* & & & \implies & \mathbf{U}^{-1} \text{ is } \textit{normal} \end{array}$$

2. Proof that $\mathbf{U}^*\mathbf{U} = \mathbf{U}\mathbf{U}^* = \mathbf{I} \implies \mathscr{I}(\mathbf{U}) = \mathscr{I}(\mathbf{U}^*) = \boldsymbol{H}$: by Theorem R.25 page 383.

3. Proof that $\mathbf{U}^*\mathbf{U} = \mathbf{U}\mathbf{U}^* = \mathbf{I} \implies \mathscr{N}(\mathbf{U}) = \mathscr{N}(\mathbf{U}^*) = \mathscr{N}(\mathbf{U}^{-1})$:

$$\begin{aligned} \mathscr{N}(\mathbf{U}^*) &= \mathscr{N}(\mathbf{U}) && \text{because } \mathbf{U} \text{ and } \mathbf{U}^* \text{ are both } \textit{normal} \text{ and by Theorem R.20 page 378} \\ &= \mathscr{I}(\mathbf{U})^{\perp} && \text{by Theorem R.14 page 374} \\ &= X^{\perp} && \text{by above result} \\ &= \{\mathbb{0}\} && \text{by Proposition L.6 page 295} \end{aligned}$$

4. Proof that $\mathbf{U}^*\mathbf{U} = \mathbf{U}\mathbf{U}^* = \mathbf{I} \implies \left\|\mathbf{U}^{-1}\right\| = \left\|\mathbf{U}^*\right\| = \|\mathbf{U}\| = 1$:
Because $\mathbf{U}$, $\mathbf{U}^*$, and $\mathbf{U}^{-1}$ are all isometric and by Theorem R.24 page 381.

✏

Example R.5 (Rotation matrix). [47]

EX
$$\underbrace{\left\{\mathbf{R}_\theta \triangleq \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\right\}}_{\textbf{rotation matrix } \mathbf{R}_\theta \,:\, \mathbb{R}^2 \to \mathbb{R}^2} \implies \left\{\begin{array}{llcll} (1). & \mathbf{R}^{-1}{}_\theta & = & \mathbf{R}_{-\theta} & \text{and} \\ (2). & \mathbf{R}^*{}_\theta & = & \mathbf{R}^{-1}{}_\theta & (\mathbf{R} \text{ is unitary}) \end{array}\right\}$$

✎PROOF:

$$\begin{aligned} \mathbf{R}^* &= \mathbf{R}^H \\ &= \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}^H && \text{by definition of } \mathbf{R} \\ &= \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} && \text{by definition of } \textit{Hermetian transpose} \text{ operator } H \\ &= \begin{bmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{bmatrix} && \text{(Theorem I.2 page 231)} \\ &= \mathbf{R}_{-\theta} && \text{by definition of } \mathbf{R} \\ &= \mathbf{R}^{-1} && \text{by 1.} \end{aligned}$$

✏

[47] Noble and Daniel (1988) page 311

Example R.6. [48] Let $\mathbf{A}$ and $\mathbf{B}$ be matrix operators.

EX

$$\underbrace{\begin{array}{c|c} \mathbf{A} \triangleq \begin{bmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix} & \mathbf{B} \triangleq \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \\ \mathbf{A} \text{ is a } \textit{rotation operator}. & \mathbf{B} \text{ is a } \textit{reflection operator}. \end{array}}_{\text{Both } \mathbf{A} \text{ and } \mathbf{B} \text{ are } \textbf{unitary}.}$$

Example R.7. Examples of *Fredholm integral operators* include

EX

1. **Fourier Transform** $[\tilde{\mathbf{F}}\mathsf{x}](f) = \int_{t\in\mathbb{R}} \mathsf{x}(t)e^{-i2\pi ft}\,\mathsf{d}t \quad \kappa(t,f) = e^{-i2\pi ft}$
2. **Inverse Fourier Transform** $[\tilde{\mathbf{F}}^{-1}\tilde{\mathsf{x}}](t) = \int_{f\in\mathbb{R}} \tilde{\mathsf{x}}(f)e^{i2\pi ft}\,\mathsf{d}f \quad \kappa(f,t) = e^{i2\pi ft}$
3. **Laplace operator** $[\mathbf{L}\mathsf{x}](s) = \int_{t\in\mathbb{R}} \mathsf{x}(t)e^{-st}\,\mathsf{d}t \quad \kappa(t,s) = e^{-st}$

Example R.8 (Translation operator). Let $\boldsymbol{X} = \boldsymbol{L}^2_{\mathbb{R}}$ and $\mathbf{T} \in \boldsymbol{X}^{\boldsymbol{X}}$ be defined as

$$\mathbf{T}\mathsf{f}(x) \triangleq \mathsf{f}(x-1) \qquad \forall \mathsf{f}\in\boldsymbol{L}^2_{\mathbb{R}} \qquad \text{(translation operator)}$$

EX

1. $\mathbf{T}^{-1}\mathsf{f}(x) = \mathsf{f}(x+1) \qquad \forall \mathsf{f}\in\boldsymbol{L}^2_{\mathbb{R}}$ (inverse translation operator)
2. $\mathbf{T}^* = \mathbf{T}^{-1}$ ($\mathbf{T}$ is invertible)
3. $\mathbf{T}^*\mathbf{T} = \mathbf{T}\mathbf{T}^* = \mathbf{I}$ ($\mathbf{T}$ is unitary)

✎PROOF:

1. Proof that $\mathbf{T}^{-1}\mathsf{f}(x) = \mathsf{f}(x+1)$:

$$\mathbf{T}^{-1}\mathbf{T} = \mathbf{I}$$
$$\mathbf{T}\mathbf{T}^{-1} = \mathbf{I}$$

2. Proof that $\mathbf{T}$ is unitary:

$$\begin{aligned}
\langle \mathbf{T}\mathsf{f}(x) \,|\, \mathsf{g}(x)\rangle &= \langle \mathsf{f}(x-1) \,|\, \mathsf{g}(x)\rangle && \text{by definition of } \mathbf{T} \\
&= \int_x \mathsf{f}(x-1)\mathsf{g}^*(x)\,\mathsf{d}x \\
&= \int_x \mathsf{f}(x)\mathsf{g}^*(x+1)\,\mathsf{d}x \\
&= \langle \mathsf{f}(x) \,|\, \mathsf{g}(x+1)\rangle \\
&= \left\langle \mathsf{f}(x) \,\middle|\, \underbrace{\mathbf{T}^{-1}}_{\mathbf{T}^*}\mathsf{g}(x) \right\rangle && \text{by 1.}
\end{aligned}$$

⇨

Example R.9 (Dilation operator). Let $\boldsymbol{X} = \boldsymbol{L}^2_{\mathbb{R}}$ and $\mathbf{T} \in \boldsymbol{X}^{\boldsymbol{X}}$ be defined as

$$\mathbf{D}\mathsf{f}(x) \triangleq \sqrt{2}\mathsf{f}(2x) \qquad \forall \mathsf{f}\in\boldsymbol{L}^2_{\mathbb{R}} \qquad \text{(dilation operator)}$$

EX

1. $\mathbf{D}^{-1}\mathsf{f}(x) = \frac{1}{\sqrt{2}}\mathsf{f}\left(\frac{1}{2}x\right) \qquad \forall \mathsf{f}\in\boldsymbol{L}^2_{\mathbb{R}}$ (inverse dilation operator)
2. $\mathbf{D}^* = \mathbf{D}^{-1}$ ($\mathbf{D}$ is invertible)
3. $\mathbf{D}^*\mathbf{D} = \mathbf{D}\mathbf{D}^* = \mathbf{I}$ ($\mathbf{D}$ is unitary)

[48] Gel'fand (1963) page 4, Gelfand et al. (2018) page 4

✎PROOF:

1. Proof that $\mathbf{D}^{-1}\mathsf{f}(x) = \frac{1}{\sqrt{2}}\mathsf{f}\left(\frac{1}{2}x\right)$:

$$\begin{aligned}\mathbf{D}^{-1}\mathbf{D} &= \mathbf{I}\\ \mathbf{D}\mathbf{D}^{-1} &= \mathbf{I}\end{aligned}$$

2. Proof that $\mathbf{D}$ is unitary:

$$\begin{aligned}
\langle \mathbf{D}\mathsf{f}(x) \,|\, \mathsf{g}(x)\rangle &= \left\langle \sqrt{2}\mathsf{f}(2x) \,|\, \mathsf{g}(x)\right\rangle && \text{by definition of } \mathbf{D}\\
&= \int_x \sqrt{2}\mathsf{f}(2x)\mathsf{g}^*(x)\,\mathsf{d}x\\
&= \int_{u\in\mathbb{R}} \sqrt{2}\mathsf{f}(u)\mathsf{g}^*\left(\frac{1}{2}u\right)\frac{1}{2}\,\mathsf{d}u && \text{let } u \triangleq 2x \implies \mathsf{d}x = \frac{1}{2}\,\mathsf{d}u\\
&= \int_{u\in\mathbb{R}} \mathsf{f}(u)\left[\frac{1}{\sqrt{2}}\mathsf{g}\left(\frac{1}{2}u\right)\right]^*\,\mathsf{d}u\\
&= \left\langle \mathsf{f}(x) \,|\, \frac{1}{\sqrt{2}}\mathsf{g}\left(\frac{1}{2}x\right)\right\rangle\\
&= \left\langle \mathsf{f}(x) \,|\, \underbrace{\mathbf{D}^{-1}}_{\mathbf{D}^*}\mathsf{g}(x)\right\rangle && \text{by 1.}
\end{aligned}$$

✎

Example R.10 (Delay operator). Let $\boldsymbol{X}$ be the set of all sequences and $\mathbf{D} \in \boldsymbol{X}^{\boldsymbol{X}}$ be a delay operator.

EX The delay operator $\mathbf{D}\,(\!(x_n)\!)_{n\in\mathbb{Z}} \triangleq (\!(x_{n-1})\!)_{n\in\mathbb{Z}}$ is unitary.

✎PROOF: The inverse $\mathbf{D}^{-1}$ of the delay operator $\mathbf{D}$ is

$$\mathbf{D}^{-1}\,(\!(x_n)\!)_{n\in\mathbb{Z}} \triangleq (\!(x_{n+1})\!)_{n\in\mathbb{Z}}.$$

$$\begin{aligned}
\langle \mathbf{D}\,(\!(x_n)\!) \,|\, (\!(y_n)\!)\rangle &= \langle (\!(x_{n-1})\!) \,|\, (\!(y_n)\!)\rangle && \text{by definition of } \mathbf{D}\\
&= \sum_n \boldsymbol{x}_{n-1}\,\boldsymbol{y}_n^*\\
&= \sum_n \boldsymbol{x}_n\,\boldsymbol{y}_{n+1}^*\\
&= \left\langle (\!(x_n)\!) \,|\, (\!(y_{n+1})\!)\right\rangle\\
&= \left\langle (\!(x_n)\!) \,|\, \underbrace{\mathbf{D}^{-1}}_{\mathbf{D}^*}(\!(y_n)\!)\right\rangle
\end{aligned}$$

Therefore, $\mathbf{D}^* = \mathbf{D}^{-1}$. This implies that $\mathbf{D}\mathbf{D}^* = \mathbf{D}^*\mathbf{D} = \mathbf{I}$ which implies that $\mathbf{D}$ is unitary. ✎

Example R.11 (Fourier transform). Let $\tilde{\mathbf{F}}$ be the *Fourier Transform* and $\tilde{\mathbf{F}}^{-1}$ the *inverse Fourier Transform* operator

$$[\tilde{\mathbf{F}}\mathsf{x}](f) \triangleq \int_t \mathsf{x}(t)\underbrace{e^{-i2\pi ft}}_{\kappa(t,f)}\,\mathsf{d}t \qquad \left[\tilde{\mathbf{F}}^{-1}\tilde{\mathsf{x}}\right](t) \triangleq \int_f \tilde{\mathsf{x}}(f)\underbrace{e^{i2\pi ft}}_{\kappa^*(t,f)}\,\mathsf{d}f.$$

EX $\tilde{\mathbf{F}}^* = \tilde{\mathbf{F}}^{-1}$ (the Fourier Transform operator $\tilde{\mathbf{F}}$ is unitary)

✎PROOF:

$$\begin{aligned}
\langle \tilde{\mathbf{F}}\mathsf{x} \,|\, \tilde{\mathsf{y}} \rangle &= \left\langle \int_t \mathsf{x}(t) e^{-i2\pi f t}\, \mathsf{d}t \,\middle|\, \tilde{\mathsf{y}}(f) \right\rangle \\
&= \int_t \mathsf{x}(t) \left\langle e^{-i2\pi f t} \,|\, \tilde{\mathsf{y}}(f) \right\rangle \mathsf{d}t \\
&= \int_t \mathsf{x}(t) \int_f e^{-i2\pi f t} \tilde{\mathsf{y}}^*(f)\, \mathsf{d}f\, \mathsf{d}t \\
&= \int_t \mathsf{x}(t) \left[\int_f e^{i2\pi f t} \tilde{\mathsf{y}}(f)\, \mathsf{d}f \right]^* \mathsf{d}t \\
&= \left\langle \mathsf{x}(t) \,\middle|\, \int_f \tilde{\mathsf{y}}(f) e^{i2\pi f t}\, \mathsf{d}f \right\rangle \\
&= \left\langle \mathsf{x} \,\middle|\, \underbrace{\tilde{\mathbf{F}}^{-1}}_{\tilde{\mathbf{F}}^*} \tilde{\mathsf{y}} \right\rangle
\end{aligned}$$

This implies that $\tilde{\mathbf{F}}$ is unitary ($\tilde{\mathbf{F}}^* = \tilde{\mathbf{F}}^{-1}$). ✏

R.5 Operator order

Definition R.12. [49] *Let* $\mathbf{P} \in \boldsymbol{Y}^{\boldsymbol{X}}$ *be an operator.*

DEF

***P** is positive if* $\langle \mathbf{P}\boldsymbol{x} \,|\, \boldsymbol{x} \rangle \geq 0 \;\forall \boldsymbol{x} \in \boldsymbol{X}$.
This condition is denoted $\mathbf{P} \geq 0$.

Theorem R.27. [50]

THM

$$\underbrace{\mathbf{P} \geq 0 \text{ and } \mathbf{Q} \geq 0}_{\mathbf{P} \text{ and } \mathbf{Q} \text{ are both positive}} \implies \left\{ \begin{array}{lcllr} (\mathbf{P}+\mathbf{Q}) & \geq & 0 & & ((\mathbf{P}+\mathbf{Q}) \text{ is positive}) \\ \mathbf{A}^*\mathbf{P}\mathbf{A} & \geq & 0 & \forall \mathbf{A} \in \mathscr{B}(\boldsymbol{X},\boldsymbol{X}) & (\mathbf{A}^*\mathbf{P}\mathbf{A} \text{ is positive}) \\ \mathbf{A}^*\mathbf{A} & \geq & 0 & \forall \mathbf{A} \in \mathscr{B}(\boldsymbol{X},\boldsymbol{X}) & (\mathbf{A}^*\mathbf{A} \text{ is positive}) \end{array} \right.$$

✎PROOF:

$$\begin{aligned}
\langle (\mathbf{P}+\mathbf{Q})\boldsymbol{x} \,|\, \boldsymbol{x} \rangle &= \langle \mathbf{P}\boldsymbol{x} \,|\, \boldsymbol{x} \rangle + \langle \mathbf{Q}\boldsymbol{x} \,|\, \boldsymbol{x} \rangle && \text{by additive property of } \langle \triangle \,|\, \nabla \rangle \text{ (Definition N.1 page 309)} \\
&\geq \langle \mathbf{P}\boldsymbol{x} \,|\, \boldsymbol{x} \rangle && \text{by left hypothesis} \\
&\geq 0 && \text{by left hypothesis} \\
\langle \mathbf{A}^*\mathbf{P}\mathbf{A}\boldsymbol{x} \,|\, \boldsymbol{x} \rangle &= \langle \mathbf{P}\mathbf{A}\boldsymbol{x} \,|\, \mathbf{A}\boldsymbol{x} \rangle && \text{by definition of adjoint (Proposition R.3 page 372)} \\
&= \langle \mathbf{P}\boldsymbol{y} \,|\, \boldsymbol{y} \rangle && \text{where } \boldsymbol{y} \triangleq \mathbf{A}\boldsymbol{x} \\
&\geq 0 && \text{by left hypothesis} \\
\langle \mathbf{I}\boldsymbol{x} \,|\, \boldsymbol{x} \rangle &= \langle \boldsymbol{x} \,|\, \boldsymbol{x} \rangle && \text{by definition of } \mathbf{I} \text{ (Definition R.2 page 359)} \\
&\geq 0 && \text{by non-negative property of } \langle \triangle \,|\, \nabla \rangle \text{ (Definition N.1 page 309)} \\
&\implies \mathbf{I} \text{ is positive} \\
\langle \mathbf{A}^*\mathbf{A}\boldsymbol{x} \,|\, \boldsymbol{x} \rangle &= \langle \mathbf{A}^*\mathbf{I}\mathbf{A}\boldsymbol{x} \,|\, \boldsymbol{x} \rangle && \text{by definition of } \mathbf{I} \text{ (Definition R.2 page 359)} \\
&\geq 0 && \text{by two previous results}
\end{aligned}$$

✏

[49] Michel and Herget (1993) page 429 ⟨Definition 7.4.12⟩
[50] Michel and Herget (1993) page 429

Definition R.13. [51] *Let* $\mathbf{A}, \mathbf{B} \in \mathscr{B}(\boldsymbol{X}, \boldsymbol{Y})$ *be* BOUNDED *operators.*

DEF

$\mathbf{A} \geq \mathbf{B}$ *("$\mathbf{A}$ is greater than or equal to $\mathbf{B}$") if*
$\mathbf{A} - \mathbf{B} \geq 0$ *("($\mathbf{A} - \mathbf{B}$) is positive")*

[51] Michel and Herget (1993) page 429

APPENDIX S

LAPLACE TRANSFORM

❝La langue de l analyse, la plus parfaite de toutes les langues, tant par elle-même un puissant instrument de découvertes; ses notations, lorsqu'elles sont nécessaires et heureusement imaginées, sont des germes de nouveaux calculs.❞	*❝The language of analysis, most perfect of all, being in itself a powerful instrument of discoveries, its notations, especially when they are necessary and happily imagined, are the seeds of new calculi.❞*

Pierre-Simon Laplace[1]

S.1 Operator Definition

Definition S.1. [2] *Let $\boldsymbol{L}^2_{(\mathbb{R},\mathscr{B},\mu)}$ be the space of all* LEBESGUE SQUARE-INTEGRABLE FUNCTIONS.

DEF *The **Laplace Transform** operator* $\mathbf{L}$ *is here defined as*

$$[\mathbf{L}\mathsf{f}](s) \triangleq \int_{x\in\mathbb{R}} \mathsf{f}(x)e^{-sx}\,\mathsf{d}x \qquad \forall \mathsf{f}\in\boldsymbol{L}^2_{(\mathbb{R},\mathscr{B},\mu)}$$

Such integrals may *converge* for certain values of s and *diverge* for others.

Definition S.2. *Let* $\mathbf{L}\left[\mathsf{g}(x)\right]$ *be the* LAPLACE TRANSFORM (Definition S.1 page 389) *of a function* $\mathsf{g}(x)$.

DEF *The set* $\mathbf{RoC}\mathbf{L}\left[\mathsf{g}(x)\right]$ *of all s for which* $\mathbf{L}\left[\mathsf{g}(x)\right]$ CONVERGES *is the **Region of Convergence** of* $\mathbf{L}\left[\mathsf{g}(x)\right]$.

In this text, the region of convergence may in places be specified using the *open interval* $(A : B)$ and *closed interval* $[A : B]$ (Definition P.1 page 337).

[1] Laplace (1814) page xxxi ⟨Introduction⟩, Laplace (1812), Laplace (1902) pages 48–49, Moritz (1914) page 200 ⟨Quote 1222., but "conceived" not "imagined", and "are so many germs" not "are the seeds"⟩, `https://todayinsci.com/L/Laplace_Pierre/LaplacePierre-Analysis-Quotations.htm`, `https://translate.google.com/`,

[2] Bracewell (1978) page 219 ⟨Chapter 11 The Laplace transform⟩, van der Pol and Bremmer (1959) page 13 ⟨5. Strip of convergence of the Laplace integral⟩, Levy (1958) page 2 ⟨"two-sided transformation"⟩, Betten (2008b) page 295 ⟨(B.1)⟩

Remark S.1. A scaling factor $\frac{1}{\sqrt{2\pi}}$ in front of $\int_{\mathbb{R}}$ in Definition S.1 is not typically found in references offering definitions of the Laplace Transform, and is not included here either. That is not to say, however, that it's not a good idea. Including it would make the operator $\mathbf{L}$ more directly compatible with the *Unitary Fourier Transform* operator $\tilde{\mathbf{F}}$ (Definition T.2 page 408). Note also that a $\frac{1}{2\pi}$ scaling factor is included in [Bachman et al. (2002) page 268] in their definition of *convolution* (Definition D.1 page 199, Section S.8 page 401).

S.2 Operator Inverse

Theorem S.1. [3]

THM
$$\mathrm{g}(x) \;=\; \mathbf{L}^{-1}[\mathrm{G}(s)] \triangleq \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \mathrm{G}(s)e^{sx}\,ds \quad \text{for some } c \in \mathbb{R}^+$$

S.3 Transversal properties

Theorem S.2. [4] *Let* $\mathbf{L}[\mathrm{g}(x)]$ *be the* LAPLACE TRANSFORM *(Definition S.1 page 389) of a function* $\mathrm{g}(x)$*. Let the* REGION OF CONVERGENCE *of* $\mathbf{L}[\mathrm{g}(x)](s)$ *be* $A \le \mathbf{R_e}(s) \le B$ *with* $(A, B) \in \mathbb{R}^2$*.*

THM

Mapping			Region of Convergence	Domain	Property
$\mathbf{L}[\mathrm{g}(x-\alpha)]$	$=$	$e^{-\alpha s}\mathbf{L}[\mathrm{g}(x)](s)$	for $\mathbf{R_e}(s) \in [A : B]$	$\forall x,\alpha\in\mathbb{C}$	(TRANSLATION)
$\mathbf{L}[\mathrm{g}(\alpha x)]$	$=$	$\frac{1}{\lvert\alpha\rvert}\mathbf{L}[\mathrm{g}(x)]\left(\frac{s}{\alpha}\right)$	for $\mathbf{R_e}\left(\frac{s}{\alpha}\right) \in [A : B]$	$\forall x,\alpha\in\mathbb{C}$	(DILATION)

✎PROOF:

$$\begin{aligned}
\mathbf{L}[\mathrm{g}(x-\alpha)] &\triangleq \int_{x=-\infty}^{x=\infty} \mathrm{g}(x-\alpha)e^{-sx}\,\mathrm{d}x && \text{by definition of } \mathbf{L} && \text{(Definition S.1 page 389)}\\
&= \int_{u+\alpha=-\infty}^{u+\alpha=\infty} \mathrm{g}(u)e^{-s(\alpha+u)}\,\mathrm{d}u && \text{where } u \triangleq x-\alpha && \Longrightarrow\; x = \alpha + u\\
&= e^{-\alpha s}\int_{u=-\infty}^{u=\infty} \mathrm{g}(u)e^{-su}\,\mathrm{d}u \quad \forall A \le \mathbf{R_e}(s) \le B && \text{by property of exponents} && b^{x+\alpha} = b^x b^\alpha\\
&\triangleq e^{-\alpha s}\int_{x=-\infty}^{x=\infty} \mathrm{g}(x)e^{-sx}\,\mathrm{d}x \quad \forall A \le \mathbf{R_e}(s) \le B && \text{by change of variable} && u \to x\\
&\triangleq e^{-\alpha s}\,[\mathbf{L}\mathrm{g}(x)] \quad \forall A \le \mathbf{R_e}(s) \le B && \text{by definition of } \mathbf{L} && \text{(Definition S.1 page 389)}
\end{aligned}$$

$$\begin{aligned}
\mathbf{L}[\mathrm{g}(\alpha x)] &\triangleq \int_{x=-\infty}^{x=\infty} \mathrm{g}(\alpha x)e^{-sx}\,\mathrm{d}x && \text{by definition of } \mathbf{L} && \text{(Definition S.1 page 389)}\\
&= \int_{u/\alpha=-\infty}^{u/\alpha=\infty} \mathrm{g}(u)e^{-s(u/\alpha)}\frac{1}{\alpha}\,\mathrm{d}u && \text{where } u \triangleq \alpha x && \Longrightarrow\; x = \frac{u}{\alpha}\\
&= \frac{1}{\alpha}\int_{u/\alpha=-\infty}^{u/\alpha=\infty} \mathrm{g}(u)e^{-(s/\alpha)u}\,\mathrm{d}u
\end{aligned}$$

[3] Bracewell (1978) page 220 ⟨Chapter 11 The Laplace transform⟩
[4] Bracewell (1978) page 224 ⟨Table 11.1: "Shift" and "Similarity" entries⟩, Levy (1958) page 15 ⟨Equation 0.8⟩

$$
\begin{aligned}
&= \begin{cases} \frac{1}{\alpha}\int_{u=-\infty}^{u=\infty} \mathsf{g}(u)e^{-(s/\alpha)u}\,\mathrm{d}u & \text{if } \alpha \ge 0 \quad \forall A \le \mathbf{R_e}\left(\frac{s}{\alpha}\right) \le B \\ \frac{1}{\alpha}\int_{u=\infty}^{u=-\infty} \mathsf{g}(u)e^{-(s/\alpha)u}\,\mathrm{d}u & \text{otherwise} \quad \forall A \le \mathbf{R_e}\left(\frac{s}{\alpha}\right) \le B \end{cases} \\
&= \begin{cases} \frac{1}{\alpha}\int_{u=-\infty}^{u=\infty} \mathsf{g}(u)e^{-(s/\alpha)u}\,\mathrm{d}u & \text{if } \alpha \ge 0 \quad \forall A \le \mathbf{R_e}\left(\frac{s}{\alpha}\right) \le B \\ -\frac{1}{\alpha}\int_{u=-\infty}^{u=\infty} \mathsf{g}(u)e^{-(s/\alpha)u}\,\mathrm{d}u & \text{otherwise} \quad \forall A \le \mathbf{R_e}\left(\frac{s}{\alpha}\right) \le B \end{cases} \\
&= \frac{1}{|\alpha|}\int_{x\in\mathbb{R}} \mathsf{g}(x)e^{-(s/\alpha)x}\,\mathrm{d}x \qquad \forall A \le \mathbf{R_e}\left(\frac{s}{\alpha}\right) \le B \qquad \text{by change of variable} \qquad u \to x \\
&\triangleq \frac{1}{|\alpha|}[\mathbf{L}\mathsf{g}(x)]\left(\frac{s}{\alpha}\right) \qquad \forall A \le \mathbf{R_e}\left(\frac{s}{\alpha}\right) \le B \qquad \text{by definition of } \mathbf{L} \qquad \text{(Definition S.1 page 389)}
\end{aligned}
$$

⇰

Corollary S.1. [5] *Let* $\mathbf{L}$, $\mathsf{G}(s)$, A, *and* B *be defined as in Theorem S.2 (page 390).*

COR	Mapping			Region of Convergence	Domain	Property
	$\mathbf{L}\left[\mathsf{g}(-x)\right]$	$=$	$\mathsf{G}(-s)$	*for* $\mathbf{R_e}(s) \in [-B : -A]$	$\forall x,\alpha\in\mathbb{C}$	(REVERSAL)

✎PROOF:

$$
\begin{aligned}
\mathbf{L}\left[\mathsf{g}(-x)\right] &= \mathbf{L}\left[\mathsf{g}([-1]x)\right] & \mathbf{R_e}(s) \in [A : B] & \quad \text{by definition of unary operator } - \\
&= \mathbf{L}\left[\frac{1}{|-1|}\mathsf{g}\left(\frac{x}{-1}\right)\right] & \mathbf{R_e}\left(\frac{s}{-1}\right) \in [A : B] & \quad \text{by } \textit{dilation} \text{ property (Theorem S.2 page 390)} \\
&= \mathsf{G}(-s) & \mathbf{R_e}(s) \in [-B : -A] &
\end{aligned}
$$

⇰

S.4 Linear properties

Theorem S.3. [6] *Let* $\mathbf{L}$ *be the* LAPLACE TRANSFORM *operator* (Definition S.1 page 389). *Let* $\mathsf{G}(s) \triangleq [\mathbf{L}\mathsf{g}(x)]$ *and* $\mathsf{F}(s) \triangleq [\mathbf{L}\mathsf{f}(x)]$. *Let the* REGION OF CONVERGENCE *of* $\mathsf{G}(s)$ *be* $A \le \mathbf{R_e}(s) \le B$ *and the* REGION OF CONVERGENCE *of* $\mathsf{F}(s)$ *be* $C \le \mathbf{R_e}(s) \le D$.

THM	Mapping			Region of Convergence	Domain	Property
	$\mathbf{L}\left[\mathsf{f}(x)+\mathsf{g}(x)\right]$	$=$	$\mathsf{F}(s)+\mathsf{G}(s)$	*for* $\mathbf{R_e}(s) \in [A : B] \cap [C : D]$	$\forall x\in\mathbb{C}$	(ADDITIVE)
	$\mathbf{L}\left[\alpha\mathsf{g}(x)\right]$	$=$	$\alpha\mathsf{G}(s)$	*for* $\mathbf{R_e}(s) \in [A : B]$	$\forall x,\alpha\in\mathbb{C}$	(HOMOGENEOUS)

Corollary S.2 (Linear Properties). *Let* $\mathbf{L}$ *be the* LAPLACE TRANSFORM *operator* (Definition S.1 page 389). *Let* A *and* B *be real numbers such that* $[A : B]$ *is the* REGION OF CONVERGENCE *of* $\mathbf{L}\left[\mathsf{g}(x)\right]$. *Let* C *and* D *be real numbers such that* $[C : D]$ *is the* REGION OF CONVERGENCE *of* $\mathbf{L}[\mathsf{f}(x)]$. *Let* A_n *and* B_n *be real numbers such that* $\left[A_n : B_n\right]$ *is the* REGION OF CONVERGENCE *of* $\mathbf{L}\left[\mathsf{g}_n(x)\right]$.

---

[5] Bracewell (1978) page 224 ⟨Table 11.1 Theorems for the Laplace Transform⟩

[6] Bracewell (1978) page 224 ⟨Table 11.1 Theorems for the Laplace Transform⟩, Betten (2008a) page 296 ⟨(B.6)⟩, Levy (1958) page 13 ⟨Equation 0.2⟩, van der Pol and Bremmer (1959) page 22 ⟨Introduction⟩, Shafii-Mousavi (2015) page 7 ⟨Theorem 1.4⟩

COR

Mapping			Region of Convergence	Domain
$\mathbf{L}[0]$	$=$	0	*for* $\mathbf{R_e}(s) \in [A : B]$	$\forall x \in \mathbb{C}$
$\mathbf{L}\left[-\mathsf{g}(x)\right]$	$=$	$-\mathbf{L}\left[\mathsf{g}(x)\right]$	*for* $\mathbf{R_e}(s) \in [A : B]$	$\forall x \in \mathbb{C}$
$\mathbf{L}\left[\mathsf{f}(x) - \mathsf{g}(x)\right]$	$=$	$\mathbf{L}\left[\mathsf{g}(x)\right] - \mathbf{L}[\mathsf{f}(x)]$	*for* $\mathbf{R_e}(s) \in [A : B] \cap [C : D]$	$\forall x \in \mathbb{C}$
$\mathbf{L}\left[\sum_{n=1}^{N} \alpha_n \mathsf{g}_n(x)\right]$	$=$	$\sum_{n=1}^{N} \alpha_n \mathbf{L}\left[\mathsf{g}_n(x)\right]$	*for* $\mathbf{R_e}(s) \in \bigcap_{n=1}^{N} \left[A_n : B_n\right]$	$\forall x, \alpha_n \in \mathbb{C}$

✎PROOF:

1. By Theorem S.3 (page 391), the operator *Laplace Transform* operator $\mathbf{L}$ is *additive* and *homogeneous*.
2. By item (1) and Definition R.3 (page 360), $\mathbf{L}$ is *linear*.
3. By item (2) and Theorem R.1 (page 360), the four properties listed follow.

✏

S.5 Modulation properties

Theorem S.4. [7] *Let* $\mathbf{L}$ *be the* LAPLACE TRANSFORM *operator* (Definition S.1 page 389). *Let* $\mathsf{G}(s) \triangleq \left[\mathbf{L}\mathsf{g}(x)\right]$. *Let the* REGION OF CONVERGENCE *of* $\mathsf{G}(s)$ *be* $A \le \mathbf{R_e}(s) \le B$.

THM

Mapping			Region of Convergence	Domain	Property
$\mathbf{L}\left[e^{-\alpha x}\mathsf{g}(x)\right]$	$=$	$\mathsf{G}(s+\alpha)$	*for* $A - \mathbf{R_e}(\alpha) \le \mathbf{R_e}(s) \le B - \mathbf{R_e}(\alpha)$	$\forall x, \alpha \in \mathbb{C}$	(MODULATION)

✎PROOF:

$$
\begin{aligned}
\mathbf{L}\left[e^{-\alpha x}\mathsf{g}(x)\right] &\triangleq \int_{x\in\mathbb{R}} e^{-\alpha x}\mathsf{g}(x) e^{-sx}\,\mathrm{d}x && \text{by definition of } \mathbf{L} && \text{(Definition S.1 page 389)} \\
&= \int_{x\in\mathbb{R}} \mathsf{g}(x) e^{-(s+\alpha)x}\,\mathrm{d}x && A \le \mathbf{R_e}(s+\alpha) \le B && b^{x+y} = b^x b^y \\
&\triangleq \left[\mathbf{L}\mathsf{g}(x)\right](s+\alpha) && A - \mathbf{R_e}(\alpha) \le \mathbf{R_e}(s) \le B - \mathbf{R_e}(\alpha) && \text{(Definition S.1 page 389)} \\
&\triangleq \mathsf{G}(s+\alpha) && A - \mathbf{R_e}(\alpha) \le \mathbf{R_e}(s) \le B - \mathbf{R_e}(\alpha) && \text{by definition of } \mathsf{G}(s)
\end{aligned}
$$

✏

Corollary S.3. [8] *Let* $\mathbf{L}$ *be the* LAPLACE TRANSFORM *operator* (Definition S.1 page 389). *Let* $\mathsf{G}(s) \triangleq \left[\mathbf{L}\mathsf{g}(x)\right]$ *Let the* REGION OF CONVERGENCE *of* $\mathsf{G}(s)$ *be* $A \le \mathbf{R_e}(s) \le B$.

COR

Mapping			Region of Convergence	Domain
$\mathbf{L}\left[\cos(\omega_o x)\mathsf{g}(x)\right]$	$=$	$\frac{1}{2}\mathsf{G}(s - i\omega_o) + \frac{1}{2}\mathsf{G}(s + i\omega_o)$	*for* $\mathbf{R_e}(s) \in [A : B]$	$\forall x\omega_o \in \mathbb{C}$
$\mathbf{L}\left[\sin(\omega_o x)\mathsf{g}(x)\right]$	$=$	$-\frac{i}{2}\mathsf{G}(s - i\omega_o) + \frac{i}{2}\mathsf{G}(s + i\omega_o)$	*for* $\mathbf{R_e}(s) \in [A : B]$	$\forall x\omega_o \in \mathbb{C}$
$\mathbf{L}\left[\cosh(\omega_o x)\mathsf{g}(x)\right]$	$=$	$\frac{1}{2}\mathsf{G}(s - \omega_o) + \frac{1}{2}\mathsf{G}(s + \omega_o)$	*for* $\mathbf{R_e}(s) \in \left[A + \left\lvert\mathbf{R_e}(\omega_o)\right\rvert : B - \left\lvert\mathbf{R_e}(\omega_o)\right\rvert\right]$	$\forall x\omega_o \in \mathbb{C}$
$\mathbf{L}\left[\sinh(\omega_o x)\mathsf{g}(x)\right]$	$=$	$\frac{1}{2}\mathsf{G}(s - \omega_o) - \frac{1}{2}\mathsf{G}(s + \omega_o)$	*for* $\mathbf{R_e}(s) \in \left[A + \left\lvert\mathbf{R_e}(\omega_o)\right\rvert : B - \left\lvert\mathbf{R_e}(\omega_o)\right\rvert\right]$	$\forall x\omega_o \in \mathbb{C}$

[7] Bracewell (1978) page 224 ⟨Table 11.1: “Modulation” entry⟩, Levy (1958) page 19 ⟨Equation 1.2⟩
[8] Bracewell (1978) page 224 ⟨Table 11.1 Theorems for the Laplace Transform⟩

✎PROOF:

1. Mappings:

$$\begin{aligned}
\mathbf{L}\left[\cosh\left(\omega_o x\right)\mathsf{g}(x)\right] &= \mathbf{L}\left[\left(\frac{e^{\omega_o x}+e^{-\omega_o x}}{2}\right)\mathsf{g}(x)\right] && \text{by definition of } \cosh(x) && \text{(Definition I.5 page 242)}\\
&= \frac{1}{2}\mathbf{L}\left[e^{\omega_o x}\mathsf{g}(x)\right](s) + \frac{1}{2}\mathbf{L}\left[e^{-\omega_o x}\mathsf{g}(x)\right](s) && \text{by \textit{additive} property} && \text{(Theorem S.3 page 391)}\\
&= \frac{1}{2}\mathbf{L}\left[\mathsf{g}(x)\right](s-\omega) + \frac{1}{2}\mathbf{L}\left[\mathsf{g}(x)\right](s+\omega) && \text{by \textit{modulation} prop.} && \text{(Theorem S.4 page 392)}\\
&= \frac{1}{2}\mathsf{G}(s-\omega_o) + \frac{1}{2}\mathsf{G}(s+\omega_o) && \text{by definition of } \mathsf{G}(s)
\end{aligned}$$

$$\begin{aligned}
\mathbf{L}\left[\sinh\left(\omega_o x\right)\mathsf{g}(x)\right] &= \mathbf{L}\left[\left(\frac{e^{\omega_o x}-e^{-\omega_o x}}{2}\right)\mathsf{g}(x)\right] && \text{by definition of } \sinh(x) && \text{(Definition I.5 page 242)}\\
&= \frac{1}{2}\mathbf{L}\left[e^{\omega_o x}\mathsf{g}(x)\right](s) - \frac{1}{2}\mathbf{L}\left[e^{-\omega_o x}\mathsf{g}(x)\right](s) && \text{by \textit{additive} property} && \text{(Theorem S.3 page 391)}\\
&= \frac{1}{2}\mathbf{L}\left[\mathsf{g}(x)\right](s-\omega) - \frac{1}{2}\mathbf{L}\left[\mathsf{g}(x)\right](s+\omega) && \text{by \textit{modulation} prop.} && \text{(Theorem S.4 page 392)}\\
&= \frac{1}{2}\mathsf{G}(s-\omega_o) - \frac{1}{2}\mathsf{G}(s+\omega_o) && \text{by definition of } \mathsf{G}(s)
\end{aligned}$$

$$\begin{aligned}
\mathbf{L}\left[\cos\left(\omega_o x\right)\mathsf{g}(x)\right] &= \mathbf{L}\left[\cosh\left(i\omega_o x\right)\mathsf{g}(x)\right] && \text{by Theorem I.12 page 242}\\
&= \frac{1}{2}\mathsf{G}(s-i\omega_o) + \frac{1}{2}\mathsf{G}(s+i\omega_o) && \text{by } \mathbf{L}\left[\cos(\omega_o x)\mathsf{g}(x)\right] \text{ result}
\end{aligned}$$

$$\begin{aligned}
\mathbf{L}\left[\sin\left(\omega_o x\right)\mathsf{g}(x)\right] &= \mathbf{L}\left[-i^2\sin\left(\omega_o x\right)\mathsf{g}(x)\right]\\
&= -i\mathbf{L}\left[i\sin\left(\omega_o x\right)\mathsf{g}(x)\right] && \text{by \textit{homogeneous} property} && \text{(Theorem S.3 page 391)}\\
&= -i\mathbf{L}\left[\sinh\left(i\omega_o x\right)\mathsf{g}(x)\right] && \text{by Theorem I.12 page 242}\\
&= -\frac{i}{2}\mathsf{G}(s-i\omega_o) + \frac{i}{2}\mathsf{G}(s+i\omega_o) && \text{by } \mathbf{L}\left[\sin(\omega_o x)\mathsf{g}(x)\right] \text{ result}
\end{aligned}$$

2. Region of Convergence of $\mathbf{L}\left[\cos\left(\omega_o x\right)\mathsf{g}(x)\right]$ and $\mathbf{L}\left[\sin\left(\omega_o x\right)\mathsf{g}(x)\right]$:

$$\begin{aligned}
&\mathbf{RocL}\left[{}^{\cos}\!/\!{}_{\sin}\left(\omega_o x\right)\mathsf{g}(x)\right]\\
&= \mathbf{RocL}\left[\left(\frac{e^{i\omega_o x}\pm e^{-i\omega_o x}}{2}\right)\mathsf{g}(x)\right] && \text{by \textit{Euler's Identity}} && \text{(Theorem I.5 page 234)}\\
&= \mathbf{Roc}\left(\mathbf{L}\left[\frac{e^{i\omega_o x}}{2}\mathsf{g}(x)\right] \pm \mathbf{L}\left[\frac{e^{-i\omega_o x}}{2}\mathsf{g}(x)\right]\right) && \text{by \textit{additive} property} && \text{(Theorem S.3 page 391)}\\
&= \mathbf{RocL}\left[\left(\frac{e^{-i\omega_o x}}{2}\right)\mathsf{g}(x)\right] \cap \mathbf{RocL}\left[\left(\frac{e^{i\omega_o x}}{2}\right)\mathsf{g}(x)\right]\\
&= \left[A-\mathbf{R_e}(i\omega) \,:\, B-\mathbf{R_e}(i\omega)\right] \cap \left[A-\mathbf{R_e}(-i\omega) \,:\, B-\mathbf{R_e}(-i\omega)\right]\\
&= [A-0 \,:\, B-0] \cap [A-0 \,:\, B-0]\\
&= [A \,:\, B]
\end{aligned}$$

3. Region of Convergence of $\mathbf{L}\left[\cosh\left(\omega_o x\right)\mathsf{g}(x)\right]$ and $\mathbf{L}\left[\sinh\left(\omega_o x\right)\mathsf{g}(x)\right]$:

$$\begin{aligned}
&\mathbf{RocL}\left[{}^{\cosh}\!/\!{}_{\sinh}\left(\omega_o x\right)\mathsf{g}(x)\right]\\
&= \mathbf{RocL}\left[\left(\frac{e^{\omega_o x}\pm e^{-\omega_o x}}{2}\right)\mathsf{g}(x)\right] && \text{by def. } \cosh(x),\ \sinh(x) && \text{(Definition I.5 page 242)}
\end{aligned}$$

$$
\begin{aligned}
&= \mathbf{Roc}\left(\mathbf{L}\left[\frac{e^{\omega_o x}}{2}\mathrm{g}(x)\right] \pm \mathbf{L}\left[\frac{e^{-\omega_o x}}{2}\mathrm{g}(x)\right]\right) && \text{by } \textit{additive} \text{ property} && \text{(Theorem S.3 page 391)}\\
&= \mathbf{RocL}\left[\left(\frac{e^{-\omega_o x}}{2}\right)\mathrm{g}(x)\right] \cap \mathbf{RocL}\left[\left(\frac{e^{\omega_o x}}{2}\right)\mathrm{g}(x)\right]\\
&= \left[A - \mathbf{R_e}(\omega_o) : B - \mathbf{R_e}(\omega_o)\right] \cap \left[A - \mathbf{R_e}(-\omega_o) : B - \mathbf{R_e}(-\omega_o)\right]\\
&= \begin{cases} \left[A + \mathbf{R_e}(\omega_o) : B - \mathbf{R_e}(\omega_o)\right] & \text{for } \omega \geq 0\\ \left[A - \mathbf{R_e}(\omega_o) : B + \mathbf{R_e}(\omega_o)\right] & \text{otherwise}\end{cases}\\
&= \left[A + \left|\mathbf{R_e}(\omega_o)\right| : B - \left|\mathbf{R_e}(\omega_o)\right|\right] && \text{by definition of } |x|
\end{aligned}
$$

✏

S.6 Causality properties

Definition S.3. [9] *The* ***Heaviside step function*** $\mu(x)$ *or* ***unit step function*** *is defined as*

DEF
$$\mu(x) \triangleq \begin{cases} 1 & \forall x \geq 0\\ 0 & \textit{otherwise}\end{cases}$$

Theorem S.5. [10] *Let* $\mathbf{L}$ *be the* LAPLACE TRANSFORM *operator* (Definition S.1 page 389) *and* $\mu(x)$ *the* UNIT STEP *function* (Definition S.3 page 394).

THM

Mapping			*Region of Convergence*	*Domain*
$\mathbf{L}[\mu(x)]$	$=$	$\frac{1}{s}$	*for* $\mathbf{R_e}(s) > 0$	$\forall x \in \mathbb{R}$

✎PROOF:

$$
\begin{aligned}
\mathbf{L}[\mu(x)] &\triangleq \int_{\mathbb{R}} \mu(x)e^{-sx}\,\mathrm{d}x && \text{by definition of } \mathbf{L} && \text{(Definition S.1 page 389)}\\
&= \int_0^\infty e^{-sx}\,\mathrm{d}x && \text{by definition of } \mu(x) && \text{(Definition S.3 page 394)}\\
&= \left.\frac{e^{-sx}}{-s}\right|_0^\infty && \text{by } \textit{Fundamental Theorem of Calculus}\\
&= \lim_{x\to\infty}\left[\frac{e^{-sx}}{-s}\right] - \left(\frac{e^0}{-s}\right)\\
&= 0 + \frac{1}{s} && \forall \mathbf{R_e}(s) > 0\\
&= \frac{1}{s} && \forall \mathbf{R_e}(s) > 0
\end{aligned}
$$

✏

Corollary S.4. *Let* $\mathbf{L}$ *be the* LAPLACE TRANSFORM *operator* (Definition S.1 page 389) *and* $\mu(x)$ *the* UNIT STEP *function* (Definition S.3 page 394).

THM

	Mapping			*Region of Convergence*	*Domain*
(1).	$\mathbf{L}[\mu(-x)]$	$=$	$-\frac{1}{s}$	*for* $\mathbf{R_e}(s) < 0$	$\forall x \in \mathbb{R}$
(2).	$\mathbf{L}[\mu(x-a)]$	$=$	$\frac{e^{-as}}{s}$	*for* $\mathbf{R_e}(s) < 0$	$\forall a,x \in \mathbb{R}$
(2).	$\mathbf{L}[\mu(a-x)]$	$=$	$-\frac{e^{as}}{s}$	*for* $\mathbf{R_e}(s) > 0$	$\forall a,x \in \mathbb{R}$

[9] Betten (2008a) page 285
[10] Bracewell (1978) page 227 ⟨Table 11.2 Some Laplace transforms⟩

✎PROOF:

$$\begin{aligned}
\mathbf{L}[\mu(-x)] &= \mathbf{L}[\mu(x)](-s) && \mathbf{R}_e(s)<0 && \text{by } \textit{reversal} \text{ property} && \text{(Corollary S.1 page 391)}\\
&= \frac{-1}{s} && \mathbf{R}_e(s)<0 && \text{by } \textit{step} \text{ property} && \text{by (Theorem S.5 page 394)}
\end{aligned}$$

$$\begin{aligned}
\mathbf{L}[\mu(x-a)] &= e^{-as}\mathbf{L}[\mu(x)] && \mathbf{R}_e(s)<0 && \text{by } \textit{translation} \text{ property} && \text{(Theorem S.2 page 390)}\\
&= \frac{e^{-as}}{s} && \mathbf{R}_e(s)<0 && \text{by } \textit{step} \text{ property} && \text{by (Theorem S.5 page 394)}
\end{aligned}$$

$$\begin{aligned}
\mathbf{L}[\mu(a-x)] &= e^{as}\mathbf{L}[\mu(-x)] && \mathbf{R}_e(s)>0 && \text{by } \textit{translation} \text{ property} && \text{(Theorem S.2 page 390)}\\
&= -\frac{e^{as}}{s} && \mathbf{R}_e(s)>0 && \text{by previous result}
\end{aligned}$$

✏

Corollary S.5. [11] *Let* $\mathbf{L}$ *be the* LAPLACE TRANSFORM *operator* (Definition S.1 page 389) *and* $\mu(x)$ *the* UNIT STEP *function.*

COR

Mapping			*Region of Convergence*	*Domain*
$\mathbf{L}\left[e^{-\alpha x}\mu(x)\right]$	$=$	$\dfrac{1}{s+\alpha}$	*for* $\mathbf{R}_e(s) > -\mathbf{R}_e(\alpha)$	$\forall x\in\mathbb{R};\alpha\in\mathbb{C}$
$\mathbf{L}\left[e^{-\alpha x}\mu(-x)\right]$	$=$	$\dfrac{1}{s+\alpha}$	*for* $\mathbf{R}_e(s) < \mathbf{R}_e(\alpha)$	$\forall x\in\mathbb{R};\alpha\in\mathbb{C}$

✎PROOF:

$$\begin{aligned}
\mathbf{L}\left[e^{-\alpha x}\mu(x)\right](s) &= \mathbf{L}[\mu(x)](s+\alpha) && && \text{by } \textit{modulation} && \text{(Theorem S.4 page 392)}\\
&= \frac{1}{s+\alpha} && \forall\mathbf{R}_e(s)\in\big(0-\mathbf{R}_e(\alpha)\,:\,\infty-\mathbf{R}_e(\alpha)\big) && \text{by Theorem S.5 page 394}\\
&= \frac{1}{s+\alpha} && \forall\mathbf{R}_e(s) > -\mathbf{R}_e(\alpha)
\end{aligned}$$

$$\begin{aligned}
\mathbf{L}\left[e^{-\alpha x}\mu(-x)\right](s) &= \mathbf{L}[\mu(-x)](s+\alpha) && && \text{by } \textit{modulation} && \text{(Theorem S.4 page 392)}\\
&= \frac{-1}{s+\alpha} && \forall\mathbf{R}_e(s)\in\big(-\infty-\mathbf{R}_e(\alpha)\,:\,0-(-\mathbf{R}_e(\alpha))\big) && \text{by Corollary S.4 page 394}\\
&= \frac{-1}{s+\alpha} && \forall\mathbf{R}_e(s) < \mathbf{R}_e(\alpha) && \text{by } \textit{anti-causality} && \text{(Corollary S.4 page 394)}
\end{aligned}$$

✏

Corollary S.6. [12] *Let* $\mathbf{L}$ *be the* LAPLACE TRANSFORM *operator* (Definition S.1 page 389) *and* $\mu(x)$ *the* UNIT STEP *function.*

COR

	Mapping			*Region of Convergence*	*Domain*
(1).	$\mathbf{L}\left[\cos(\omega_o x)\mu(x)\right]$	$=$	$\dfrac{s}{s^2+\omega_o^2}$	*for* $\mathbf{R}_e(s) > 0$	$x,\omega_o\in\mathbb{R}$
(2).	$\mathbf{L}\left[\sin(\omega_o x)\mu(x)\right]$	$=$	$\dfrac{\omega_o}{s^2+\omega_o^2}$	*for* $\mathbf{R}_e(s) > 0$	$x,\omega_o\in\mathbb{R}$
(3).	$\mathbf{L}\left[\cos(\omega_o x)\mu(-x)\right]$	$=$	$\dfrac{-s}{s^2+\omega_o^2}$	*for* $\mathbf{R}_e(s) < 0$	$x,\omega_o\in\mathbb{R}$
(4).	$\mathbf{L}\left[\sin(\omega_o x)\mu(-x)\right]$	$=$	$\dfrac{-\omega_o}{s^2+\omega_o^2}$	*for* $\mathbf{R}_e(s) < 0$	$x,\omega_o\in\mathbb{R}$

[11] van der Pol and Bremmer (1959) page 22 ⟨Introduction⟩, Shafii-Mousavi (2015) page 3 ⟨Table 1, using One-Sided Laplace Transform⟩, van der Pol and Bremmer (1959) page 26 ⟨(8) seems to have an error: $\frac{s}{s+\alpha}$⟩

[12] Bracewell (1978) page 227 ⟨Table 11.2 Some Laplace transforms⟩, Shafii-Mousavi (2015) page 3 ⟨Table 1, using One-Sided Laplace Transform⟩

✎PROOF:

$$\begin{aligned}
\mathbf{L}\left[\cos(\omega_o x)\mu(x)\right](s) &= \frac{1}{2}\mathbf{L}[\mu(x)](s-i\omega_o) + \frac{1}{2}\mathbf{L}[\mu(x)](s+i\omega_o) && \text{by \textit{modulation}} && \text{(Corollary S.3 page 392)}\\
&= \frac{1}{2}\left[\frac{1}{s-i\omega_o}\right] + \frac{1}{2}\left[\frac{1}{s+i\omega_o}\right] && \mathbf{R_e}(s) > 0 \quad \text{by \textit{causal} prop.} && \text{(Theorem S.5 page 394)}\\
&= \frac{1}{2}\left[\frac{1}{s-i\omega_o}\right]\left[\frac{s+i\omega_o}{s+i\omega_o}\right] + \frac{1}{2}\left[\frac{1}{s+i\omega_o}\right]\left[\frac{s-i\omega_o}{s-i\omega_o}\right] && \text{(\textit{Rationalizing the Denominator})}\\
&= \frac{1}{2}\left[\frac{(s+i\omega_o)+(s-i\omega_o)}{s^2+\omega_o^2}\right] && \mathbf{R_e}(s) > 0\\
&= \frac{s}{s^2+\omega_o^2} && \mathbf{R_e}(s) > 0\\
\mathbf{L}\left[\sin(\omega_o x)\mu(x)\right](s) &= -\frac{i}{2}\mathbf{L}[\mu(x)](s-i\omega_o) + \frac{i}{2}\mathbf{L}[\mu(x)](s+i\omega_o) && \text{by \textit{modulation}} && \text{(Corollary S.3 page 392)}\\
&= -\frac{i}{2}\left[\frac{1}{s-i\omega_o}\right] + \frac{i}{2}\left[\frac{1}{s+i\omega_o}\right] && \mathbf{R_e}(s) > 0 \quad \text{by \textit{causal} prop.} && \text{(Theorem S.5 page 394)}\\
&= -\frac{i}{2}\left[\frac{1}{s-i\omega_o}\right]\left[\frac{s+i\omega_o}{s+i\omega_o}\right] + \frac{i}{2}\left[\frac{1}{s+i\omega_o}\right]\left[\frac{s-i\omega_o}{s-i\omega_o}\right] && \text{(\textit{Rationalizing the Denominator})}\\
&= \frac{i}{2}\left[\frac{-(s+i\omega_o)+(s-i\omega_o)}{s^2+\omega_o^2}\right] && \mathbf{R_e}(s) > 0\\
&= \frac{\omega_o}{s^2+\omega_o^2} && \mathbf{R_e}(s) > 0
\end{aligned}$$

$$\begin{aligned}
\mathbf{L}\left[\mu(-x)\cos(\omega_o x)\right](s) &= \mathbf{L}\left[\mu(-x)\cos(\omega_o(-x))\right](s) && \text{by \textit{even} property of } \cos(x) && \text{(Theorem I.2 page 231)}\\
&= \mathbf{L}\left[\mu(x)\cos(\omega_o x)\right](-s) && \text{by \textit{reversal} property} && \text{(Corollary S.1 page 391)}\\
&= \frac{(-s)}{(-s)^2+\omega_o^2} && \mathbf{R_e}(s) < 0 \quad \text{by (1)}\\
&= \frac{-s}{s^2+\omega_o^2} && \mathbf{R_e}(s) < 0\\
\mathbf{L}\left[\sin(\omega_o x)\mu(-x)\right](s) &= \mathbf{L}\left[-\sin(\omega_o(-x))\mu(-x)\right](s) && \text{by \textit{odd} property of } \sin(x) && \text{(Theorem I.2 page 231)}\\
&= -\mathbf{L}\left[\sin(\omega_o(-x))\mu(-x)\right](s) && \text{by \textit{homogeneous} property} && \text{(Theorem S.3 page 391)}\\
&= -\mathbf{L}\left[\sin(\omega_o x)\mu(x)\right](-s) && \text{by \textit{reversal} property} && \text{(Corollary S.1 page 391)}\\
&= -\left[\frac{\omega_o}{(-s)^2+\omega_o^2}\right] && \mathbf{R_e}(s) < 0 \quad \text{by (2)}\\
&= \frac{-\omega_o}{s^2+\omega_o^2} && \mathbf{R_e}(s) < 0
\end{aligned}$$

✐

Corollary S.7. [13] *Let* $\mathbf{L}$ *be the* LAPLACE TRANSFORM *operator (Definition S.1 page 389) and* $\mu(x)$ *the* UNIT STEP *function.*

COR

	Mapping			*Region of Convergence*	*Domain*
(1).	$\mathbf{L}\left[\cosh(\omega_o x)\mu(x)\right]$	$=$	$\dfrac{s}{s^2-\omega_o^2}$	*for* $\mathbf{R_e}(s) > \lvert\omega_o\rvert$	$x,\omega_o\in\mathbb{R}$
(2).	$\mathbf{L}\left[\sinh(\omega_o x)\mu(x)\right]$	$=$	$\dfrac{\omega_o}{s^2-\omega_o^2}$	*for* $\mathbf{R_e}(s) > \lvert\omega_o\rvert$	$x,\omega_o\in\mathbb{R}$
(3).	$\mathbf{L}\left[\sinh(\omega_o x)\mu(-x)\right]$	$=$	$\dfrac{-s}{s^2-\omega_o^2}$	*for* $\mathbf{R_e}(s) < \lvert\omega_o\rvert$	$x,\omega_o\in\mathbb{R}$
(4).	$\mathbf{L}\left[\sinh(\omega_o x)\mu(-x)\right]$	$=$	$\dfrac{-\omega_o}{s^2-\omega_o^2}$	*for* $\mathbf{R_e}(s) < \lvert\omega_o\rvert$	$x,\omega_o\in\mathbb{R}$

[13] 📘 Shafii-Mousavi (2015) page 3 ⟨Table 1, using One-Sided Laplace Transform⟩

✎PROOF:

1. Mappings for $\mathbf{L}\big[\cosh(\omega_o x)\mu(x)\big]$ and $\mathbf{L}\big[\sinh(\omega_o x)\mu(x)\big]$:

$$
\begin{aligned}
\mathbf{L}\big[\cosh(\omega_o x)\mu(x)\big](s) &= \frac{1}{2}\mathbf{L}[\mu(x)](s-\omega_o) + \frac{1}{2}\mathbf{L}[\mu(x)](s+\omega_o) && \text{by } \textit{modulation} && \text{(Corollary S.3 page 392)}\\
&= \frac{1}{2}\left[\frac{1}{s-\omega_o}\right] + \frac{1}{2}\left[\frac{1}{s+\omega_o}\right] && \text{by } \textit{causal} \text{ property} && \text{(Theorem S.5 page 394)}\\
&= \frac{1}{2}\left[\frac{1}{s-\omega_o}\right]\left[\frac{s+\omega_o}{s+\omega_o}\right] + \frac{1}{2}\left[\frac{1}{s+\omega_o}\right]\left[\frac{s-\omega_o}{s-\omega_o}\right]\\
&= \frac{1}{2}\left[\frac{(s+\omega_o)+(s-\omega_o)}{s^2-\omega_o^2}\right]\\
&= \frac{s}{s^2-\omega_o^2}
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{L}\big[\sinh(\omega_o x)\mu(x)\big](s) &= \frac{1}{2}\mathbf{L}[\mu(x)](s-\omega_o) - \frac{1}{2}\mathbf{L}[\mu(x)](s+\omega_o) && \text{by } \textit{modulation} && \text{(Corollary S.3 page 392)}\\
&= \frac{1}{2}\left[\frac{1}{s-\omega_o}\right] - \frac{1}{2}\left[\frac{1}{s+\omega_o}\right] && \text{by } \textit{causal} \text{ property} && \text{(Theorem S.5 page 394)}\\
&= \frac{1}{2}\left[\frac{1}{s-\omega_o}\right]\left[\frac{s+\omega_o}{s+\omega_o}\right] - \frac{1}{2}\left[\frac{1}{s+\omega_o}\right]\left[\frac{s-\omega_o}{s-\omega_o}\right]\\
&= \frac{1}{2}\left[\frac{(s+\omega_o)-(s-\omega_o)}{s^2-\omega_o^2}\right]\\
&= \frac{\omega_o}{s^2-\omega_o^2}
\end{aligned}
$$

2. Region of Convergence of $\mathbf{L}\big[\cosh(\omega_o x)\mu(x)\big]$ and $\mathbf{L}\big[\sinh(\omega_o x)\mu(x)\big]$:

$$
\begin{aligned}
\mathbf{RoC}\,\mathbf{L}\big[\cosh(\omega_o x)\mu(x)\big] &= \big[A + |\mathbf{R}_e(\omega_o)| : B - |\mathbf{R}_e(\omega_o)|\big] && \text{by Corollary S.3 page 392}\\
&= \big(0 + |\mathbf{R}_e(\omega_o)| : \infty - |\mathbf{R}_e(\omega_o)|\big) && \text{by Theorem S.5 page 394}\\
&= \big(|\mathbf{R}_e(\omega_o)| : \infty\big)\\
&\implies \mathbf{RoC}\,\mathbf{L}\big[\cosh(\omega_o x)\mu(x)\big] > |\mathbf{R}_e(\omega_o)|
\end{aligned}
$$

3. Mappings for $\mathbf{L}\big[\cosh(\omega_o x)\mu(-x)\big]$ and $\mathbf{L}\big[\sinh(\omega_o x)\mu(-x)\big]$:

$$
\begin{aligned}
\mathbf{L}\big[\cosh(\omega_o x)\mu(-x)\big](s) &= \mathbf{L}\big[\cosh(\omega_o(-x))\mu(-x)\big](s)\\
&= \mathbf{L}\big[\cosh(\omega_o x)\mu(x)\big](-s) && \text{by } \textit{reversal} \text{ property} && \text{(Corollary S.1 page 391)}\\
&= \frac{(-s)}{(-s)^2-\omega_o^2} && \text{by previous result}\\
&= \frac{-s}{s^2-\omega_o^2}
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{L}\big[\sinh(\omega_o x)\mu(-x)\big](s) &= \mathbf{L}\big[-\sinh(\omega_o(-x))\mu(-x)\big](s)\\
&= -\mathbf{L}\big[\sinh(\omega_o(-x))\mu(-x)\big](s) && \text{by } \textit{homogeneous} \text{ property} && \text{(Theorem S.3 page 391)}\\
&= -\mathbf{L}\big[\sinh(\omega_o x)\mu(x)\big](-s) && \text{by } \textit{reversal} \text{ property} && \text{(Corollary S.1 page 391)}\\
&= \frac{-\omega_o}{(-s)^2-\omega_o^2} && \text{by previous result}\\
&= \frac{-\omega_o}{s^2-\omega_o^2}
\end{aligned}
$$

4. Region of Convergence of $\mathbf{L}\left[\cosh(\omega_o x)\mu(-x)\right]$ and $\mathbf{L}\left[\sinh(\omega_o x)\mu(-x)\right]$:

$$\begin{aligned}
\mathrm{RoC}\mathbf{L}\left[\cosh(\omega_o x)\mu(-x)\right] &= \left[A + \left|\mathbf{R_e}(\omega_o)\right| : B - \left|\mathbf{R_e}(\omega_o)\right|\right] && \text{by Corollary S.3 page 392}\\
&= \left(-\infty + \left|\mathbf{R_e}(\omega_o)\right| : 0 - \left|\mathbf{R_e}(\omega_o)\right|\right) && \text{by Corollary S.4 page 394}\\
&= \left(-\infty : \left|\mathbf{R_e}(\omega_o)\right|\right)\\
\implies \mathrm{RoC}\mathbf{L}\left[\cosh(\omega_o x)\mu(-x)\right] &< \left|\mathbf{R_e}(\omega_o)\right|
\end{aligned}$$

✏

Corollary S.8. [14] *Let* $\mathbf{L}$ *be the* LAPLACE TRANSFORM *operator* (Definition S.1 page 389) *and* $\mu(x)$ *the* UNIT STEP *function.*

COR

	Mapping			Region of Convergence	Domain
(1).	$\mathbf{L}\left[\cos(\omega_o x)e^{-\alpha x}\mu(x)\right]$	$=$	$\dfrac{s+\alpha}{(s+\alpha)^2+\omega_o^2}$	for $\mathbf{R_e}(s) > -\mathbf{R_e}(\alpha)$	$x,\omega_o\in\mathbb{R};\ \alpha\in\mathbb{C}$
(2).	$\mathbf{L}\left[\sin(\omega_o x)e^{-\alpha x}\mu(x)\right]$	$=$	$\dfrac{\omega_o}{(s+\alpha)^2+\omega_o^2}$	for $\mathbf{R_e}(s) > -\mathbf{R_e}(\alpha)$	$x,\omega_o\in\mathbb{R};\ \alpha\in\mathbb{C}$
(3).	$\mathbf{L}\left[\cos(\omega_o x)e^{\alpha x}\mu(-x)\right]$	$=$	$\dfrac{-s+\alpha}{(s-\alpha)^2+\omega_o^2}$	for $\mathbf{R_e}(s) > -\mathbf{R_e}(\alpha)$	$x,\omega_o\in\mathbb{R};\ \alpha\in\mathbb{C}$
(4).	$\mathbf{L}\left[\sin(\omega_o x)e^{\alpha x}\mu(-x)\right]$	$=$	$\dfrac{-\omega_o}{(s-\alpha)^2+\omega_o^2}$	for $\mathbf{R_e}(s) > -\mathbf{R_e}(\alpha)$	$x,\omega_o\in\mathbb{R};\ \alpha\in\mathbb{C}$

✎PROOF:

$$\begin{aligned}
\mathbf{L}\left[\cos(\omega_o x)e^{-\alpha x}\mu(x)\right](s) &= \mathbf{L}\left[\mu(x)\cos(\omega_o x)\right](s+\alpha) && \text{by } \textit{modulation} \text{ property (Theorem S.4 page 392)}\\
&= \frac{s+\alpha}{(s+\alpha)^2+\omega_o^2} && \forall\mathbf{R_e}(s)\in\left(0-\mathbf{R_e}(\alpha) : \infty-\mathbf{R_e}(\alpha)\right) \text{ by Corollary S.6}\\
&= \frac{s+\alpha}{(s+\alpha)^2+\omega_o^2} && \forall\mathbf{R_e}(s) > -\mathbf{R_e}(\alpha)\\
\mathbf{L}\left[\sin(\omega_o x)e^{-\alpha x}\mu(x)\right](s) &= \mathbf{L}\left[\mu(x)\sin(\omega_o x)\right](s+\alpha) && \text{by } \textit{modulation} \text{ property (Theorem S.4 page 392)}\\
&= \frac{\omega_o}{(s+\alpha)^2+\omega_o^2} && \forall\mathbf{R_e}(s)\in\left(0-\mathbf{R_e}(\alpha) : \infty-\mathbf{R_e}(\alpha)\right) \text{ by Corollary S.6}\\
&= \frac{\omega_o}{(s+\alpha)^2+\omega_o^2} && \forall\mathbf{R_e}(s) > -\mathbf{R_e}(\alpha)\\
\mathbf{L}\left[\cos(\omega_o x)e^{\alpha x}\mu(-x)\right](s) &= \mathbf{L}\left[\mu(-x)\cos(\omega_o x)\right](s-\alpha) && \text{by } \textit{modulation} \text{ property (Theorem S.4 page 392)}\\
&= \frac{-(s-\alpha)}{(s-\alpha)^2+\omega_o^2} && \forall\mathbf{R_e}(s)\in\left(0-\mathbf{R_e}(\alpha) : \infty-\mathbf{R_e}(\alpha)\right) \text{ by Corollary S.6}\\
&= \frac{-s+\alpha}{(s-\alpha)^2+\omega_o^2} && \forall\mathbf{R_e}(s) > -\mathbf{R_e}(\alpha)\\
\mathbf{L}\left[\sin(\omega_o x)e^{\alpha x}\mu(-x)\right](s) &= \mathbf{L}\left[\mu(-x)\sin(\omega_o x)\right](s-\alpha) && \text{by } \textit{modulation} \text{ property (Theorem S.4 page 392)}\\
&= \frac{-\omega_o}{(s-\alpha)^2+\omega_o^2} && \forall\mathbf{R_e}(s)\in\left(0-\mathbf{R_e}(\alpha) : \infty-\mathbf{R_e}(\alpha)\right) \text{ by Corollary S.6}\\
&= \frac{-\omega_o}{(s-\alpha)^2+\omega_o^2} && \forall\mathbf{R_e}(s) > -\mathbf{R_e}(\alpha)
\end{aligned}$$

✏

Corollary S.9. *Let* $\mathbf{L}$ *be the* LAPLACE TRANSFORM *operator* (Definition S.1 page 389).

COR

$\mathbf{L}\left[\cos(\omega_o x)\right]$ ***is divergent*** $\forall s\in\mathbb{C}$ $\quad\forall x,\omega_o\in\mathbb{R}$

$\mathbf{L}\left[\sin(\omega_o x)\right]$ ***is divergent*** $\forall s\in\mathbb{C}$ $\quad\forall x,\omega_o\in\mathbb{R}\setminus\{0\}$

[14] Bracewell (1978) page 227 ⟨Table 11.2 Some Laplace transforms⟩

✎PROOF:

$$\begin{aligned}
\mathbf{L}\big[\cos(\omega_o x)\big] &= \underbrace{\mathbf{L}\big[\mu(x)\cos(\omega_o x)\big]}_{\forall \mathbf{R_e}(s)>0} + \underbrace{\mathbf{L}\big[\mu(-x)\cos(\omega_o x)\big]}_{\forall \mathbf{R_e}(s)<0} && \text{by Corollary S.6 page 395}\\
&= \underbrace{\frac{s}{s^2+\omega_o^2}}_{\forall \mathbf{R_e}(s)>0} + \underbrace{\frac{-s}{s^2+\omega_o^2}}_{\forall \mathbf{R_e}(s)<0} && \text{by Corollary S.6 page 395}\\
&= \begin{cases} 0 & \forall \mathbf{R_e}(s) \in (-\infty : 0) \cap (0 : \infty) = \varnothing \\ \infty & \forall s \in \mathbb{C} \end{cases}\\
&\implies \mathbf{L}\big[\cos(\omega_o x)\big] \text{ is \textbf{divergent} } \forall s \in \mathbb{C}\\
\mathbf{L}\big[\sin(\omega_o x)\big] &= \underbrace{\mathbf{L}\big[\mu(x)\sin(\omega_o x)\big]}_{\forall \mathbf{R_e}(s)>0} + \underbrace{\mathbf{L}\big[\mu(-x)\sin(\omega_o x)\big]}_{\forall \mathbf{R_e}(s)<0} && \text{by Corollary S.6 page 395}\\
&= \underbrace{\frac{\omega_o}{s^2+\omega_o^2}}_{\forall \mathbf{R_e}(s)>0} + \underbrace{\frac{-\omega_o}{s^2+\omega_o^2}}_{\forall \mathbf{R_e}(s)<0} && \text{by Corollary S.6 page 395}\\
&= \begin{cases} 0 & \forall \mathbf{R_e}(s) \in (-\infty : 0) \cap (0 : \infty) = \varnothing \\ \infty & \forall s \in \mathbb{C} \end{cases}\\
&\implies \mathbf{L}\big[\sin(\omega_o x)\big] \text{ is \textbf{divergent} } \forall s \in \mathbb{C}
\end{aligned}$$

✏

S.7 Exponential decay properties

Corollary S.10. [15] *Let* $\mathbf{L}$ *be the* LAPLACE TRANSFORM *operator (Definition S.1 page 389) and* $\mu(x)$ *the* UNIT STEP *function. Let* $A \triangleq \mathbf{R_e}(\alpha)$.

COR	*Mapping*	*Region of Convergence*	*Domain*
	$\mathbf{L}\big[e^{-\alpha\lvert x\rvert}\big] = \dfrac{2\alpha}{\alpha^2 - s^2}$	*for* $\mathbf{R_e}(s) \in (-A : A)$	$x,\alpha \in \mathbb{C}$

✎PROOF:

$$\begin{aligned}
\mathbf{L}\big[e^{-\alpha|x|}\big] &= \mathbf{L}\big[e^{-\alpha|x|}\mu(x) + e^{-\alpha|x|}\mu(-x)\big] && \text{by definition of } \mu(x) && \text{(Definition S.3 page 394)}\\
&= \mathbf{L}\big[e^{-\alpha|x|}\mu(x)\big] + \mathbf{L}\big[e^{-\alpha|x|}\mu(-x)\big] && \text{by \textit{homogeneous} property} && \text{(Theorem S.3 page 391)}\\
&= \underbrace{\mathbf{L}\big[e^{-\alpha x}\mu(x)\big]}_{\mathbf{R_e}(s) > -\mathbf{R_e}(\alpha)} + \underbrace{\mathbf{L}\big[e^{\alpha x}\mu(-x)\big]}_{\mathbf{R_e}(s) < \mathbf{R_e}(\alpha)} && \text{by Definition S.3 page 394} && \text{and Corollary S.5 page 395}\\
&= \left[\frac{1}{s+\alpha}\right] + \left[\frac{-1}{s-\alpha}\right] && \forall \mathbf{R_e}(s) \in \big(-\mathbf{R_e}(\alpha) : \mathbf{R_e}(\alpha)\big) && \text{by Corollary S.5 page 395}\\
&= \frac{(s-\alpha)-(s+\alpha)}{(s+\alpha)(s-\alpha)} && \forall \mathbf{R_e}(s) \in \big(-\mathbf{R_e}(\alpha) : \mathbf{R_e}(\alpha)\big)\\
&= \frac{2\alpha}{\alpha^2 - s^2} && \forall \mathbf{R_e}(s) \in \big(-\mathbf{R_e}(\alpha) : \mathbf{R_e}(\alpha)\big)
\end{aligned}$$

✏

[15] Bracewell (1978) page 227 ⟨Table 11.2 Some Laplace transforms⟩,

Corollary S.11. [16] *Let* $\mathbf{L}$ *be the* LAPLACE TRANSFORM *operator* (Definition S.1 page 389) *and* $\mu(x)$ *the* UNIT STEP *function. Let* $A \triangleq \mathbf{R_e}(\alpha)$.

COR

	Mapping	Region of Convergence	Domain
(1).	$\mathbf{L}\left[\cos(\omega_o x)e^{-\alpha\lvert x\rvert}\mu(x)\right] = \dfrac{s+\alpha}{(s+\alpha)^2+\omega_o^2}$	*for* $\mathbf{R_e}(s) \in (-A : A)$	$x,\alpha\in\mathbb{R}$
(2).	$\mathbf{L}\left[\cos(\omega_o x)e^{-\alpha\lvert x\rvert}\mu(-x)\right] = \dfrac{-s+\alpha}{(s-\alpha)^2+\omega_o^2}$	*for* $\mathbf{R_e}(s) \in (-A : A)$	$x,\alpha\in\mathbb{R}$
(3).	$\mathbf{L}\left[\cos(\omega_o x)e^{-\alpha\lvert x\rvert}\right] = \dfrac{s+\alpha}{(s+\alpha)^2+\omega_o^2} + \dfrac{-s+\alpha}{(s-\alpha)^2+\omega_o^2}$	*for* $\mathbf{R_e}(s) \in (-A : A)$	$x,\alpha\in\mathbb{R}$

✎PROOF:

1. Proof for (1):

$$\begin{aligned}
&\mathbf{L}\left[\cos(\omega_o x)e^{-\alpha|x|}\mu(x)\right](s)\\
&= \mathbf{L}\left[\cos(\omega_o x)e^{-\alpha x}\mu(x)\right](s) && \text{by definition of } \mu(x) && \text{(Definition S.3 page 394)}\\
&= \mathbf{L}\left[\cos(\omega_o x)\mu(x)\right](s+\alpha) \quad \forall \mathbf{R_e}(s) \in \left(0-\mathbf{R_e}(\alpha) : \infty - \mathbf{R_e}(\alpha)\right) && \text{by } \textit{modulation} \text{ prop.} && \text{(Theorem S.4 page 392)}\\
&= \frac{s+\alpha}{(s+\alpha)^2+\omega_o^2} \quad \forall \mathbf{R_e}(s) \in \left(-\mathbf{R_e}(\alpha) : \mathbf{R_e}(\alpha)\right) && \text{by Corollary S.6 page 395}
\end{aligned}$$

2. Proof for (2):

$$\begin{aligned}
&\mathbf{L}\left[\cos(\omega_o x)e^{-\alpha|x|}\mu(-x)\right]\\
&= \mathbf{L}\left[\cos(\omega_o x)e^{\alpha x}\mu(-x)\right] && \text{by definition of } \mu(x) && \text{(Definition S.3 page 394)}\\
&= \mathbf{L}\left[\cos(-\omega_o x)e^{\alpha x}\mu(-x)\right] && \text{by } \textit{even} \text{ property of } \cos(x) && \text{(Theorem I.2 page 231)}\\
&= \mathbf{L}\left[e^{\alpha x}\cos(\omega_o(-x))\mu(-x)\right]\\
&= \mathbf{L}\underbrace{\cos(\omega_o(-x))\mu(-x)}_{g(x)}(s-\alpha) && \text{by } \textit{modulation} \text{ property} && \text{(Theorem S.4 page 392)}\\
&= \mathbf{L}\underbrace{\cos(\omega_o(-x))\mu(-x)}_{g(x)}(s-\alpha) && \text{by } \textit{modulation} \text{ property} && \text{(Theorem S.4 page 392)}\\
&= \frac{-s+\alpha}{(s-\alpha)^2+\omega_o^2} \quad \forall \mathbf{R_e}(s) \in \left(-\mathbf{R_e}(\alpha) : \mathbf{R_e}(\alpha)\right) && \text{by Corollary S.6 and Theorem S.4 page 392}
\end{aligned}$$

3. Proof for (3):

$$\begin{aligned}
\mathbf{L}\left[\cos(\omega_o x)e^{-\alpha|x|}\right] &= \mathbf{L}\left[\cos(\omega_o x)e^{-\alpha|x|}\mu(x)\right] + \mathbf{L}\left[\cos(\omega_o x)e^{-\alpha|x|}\mu(-x)\right]\\
&= \mathbf{L}\left[\cos(\omega_o x)e^{-\alpha x}\mu(x)\right] + \mathbf{L}\left[\cos(-\omega_o x)e^{\alpha x}\mu(-x)\right]\\
&= \frac{s+\alpha}{(s+\alpha)^2+\omega_o^2} + \frac{-s+\alpha}{(s-\alpha)^2+\omega_o^2} \qquad \forall \mathbf{R_e}(s) \in \left(-\mathbf{R_e}(\alpha) : \mathbf{R_e}(\alpha)\right)
\end{aligned}$$

✏

Corollary S.12. [17] *Let* $\mathbf{L}$ *be the* LAPLACE TRANSFORM *operator* (Definition S.1 page 389) *and* $\mu(x)$ *the* UNIT STEP *function. Let* $A \triangleq \mathbf{R_e}(\alpha)$.

[16] Bracewell (1978) page 227 ⟨Table 11.2 Some Laplace transforms⟩, Levy (1958) page 19 ⟨with $\psi = 0$, $\alpha_0 = \alpha$, and $\alpha_1 = 1$⟩, http://ece-research.unm.edu/bsanthan/ece541/table_ME.pdf

[17] Bracewell (1978) page 227 ⟨Table 11.2 Some Laplace transforms⟩, Levy (1958) page 19 ⟨with $\psi = 0$, $\alpha_0 = \alpha$, and $\alpha_1 = 1$⟩, http://ece-research.unm.edu/bsanthan/ece541/table_ME.pdf

COR

	Mapping			Region of Convergence	Domain
(1).	$\mathbf{L}\left[\sin(\omega_o x)e^{-\alpha\lvert x\rvert}\mu(x)\right]$	$=$	$\dfrac{\omega_o}{(s+\alpha)^2+\omega_o^2}$	for $\mathbf{R_e}(s) \in (-A : A)$	$x,\alpha\in\mathbb{R}$
(2).	$\mathbf{L}\left[\sin(\omega_o x)e^{-\alpha\lvert x\rvert}\mu(-x)\right]$	$=$	$\dfrac{-\omega_o}{(s-\alpha)^2+\omega_o^2}$	for $\mathbf{R_e}(s) \in (-A : A)$	$x,\alpha\in\mathbb{R}$
(3).	$\mathbf{L}\left[\sin(\omega_o x)e^{-\alpha\lvert x\rvert}\right]$	$=$	$\dfrac{\omega_o}{(s+\alpha)^2+\omega_o^2} + \dfrac{-\omega_o}{(s-\alpha)^2+\omega_o^2}$	for $\mathbf{R_e}(s) \in (-A : A)$	$x,\alpha\in\mathbb{R}$

✎PROOF:

1. Proof for (1):

$$\begin{aligned}
&\mathbf{L}\left[\sin(\omega_o x)e^{-\alpha|x|}\mu(x)\right] \\
&= \mathbf{L}\left[\sin(\omega_o x)e^{-\alpha x}\mu(x)\right] && \text{by definition of } \mu(x) && \text{(Definition S.3 page 394)} \\
&= \frac{s+\alpha}{(\omega_o)^2+\omega_o^2} && \forall \mathbf{R_e}(s) \in \left(-\mathbf{R_e}(\alpha) : \mathbf{R_e}(\alpha)\right) && \text{by Corollary S.6 page 395 and Theorem S.4 page 392}
\end{aligned}$$

2. Proof for (2):

$$\begin{aligned}
&\mathbf{L}\left[\sin(\omega_o x)e^{-\alpha|x|}\mu(-x)\right] \\
&= \mathbf{L}\left[\sin(-\omega_o x)e^{\alpha x}\mu(-x)\right] && \text{by definition of } \mu(x) && \text{(Definition S.3 page 394)} \\
&= \mathbf{L}\left[-\sin(\omega_o x)e^{\alpha x}\mu(-x)\right] && \text{by \textit{odd} property of } \sin(x) && \text{(Theorem I.2 page 231)} \\
&= -\mathbf{L}\left[\sin(\omega_o x)e^{\alpha x}\mu(-x)\right] && \text{by \textit{homogeneous} property} && \text{(Theorem S.3 page 391)} \\
&= \frac{-\omega_o}{(s-\alpha)^2+\omega_o^2} && \forall \mathbf{R_e}(s) \in \left(-\mathbf{R_e}(\alpha) : \mathbf{R_e}(\alpha)\right) && \text{by Theorem S.4 page 392 and Corollary S.6}
\end{aligned}$$

3. Proof for (3):

$$\begin{aligned}
\mathbf{L}\left[\sin(\omega_o x)e^{-\alpha|x|}\right] &= \mathbf{L}\left[\sin(\omega_o x)e^{-\alpha|x|}\mu(x)\right] + \mathbf{L}\left[\sin(\omega_o x)e^{-\alpha|x|}\mu(-x)\right] \\
&= \mathbf{L}\left[\sin(\omega_o x)e^{-\alpha x}\mu(x)\right] + \mathbf{L}\left[\sin(-\omega_o x)e^{\alpha x}\mu(-x)\right] \\
&= \frac{\omega_o}{(s+\alpha)^2+\omega_o^2} + \frac{-\omega_o}{(s-\alpha)^2+\omega_o^2} && \forall \mathbf{R_e}(s) \in \left(-\mathbf{R_e}(\alpha) : \mathbf{R_e}(\alpha)\right)
\end{aligned}$$

✏

S.8 Product properties

Theorem S.6 (next) demonstrates that multiplication in the "time domain" is equivalent to convolution in the "s domain" and vice-versa.

Theorem S.6 (convolution theorem). [18] *Let* $\mathbf{L}$ *be the* LAPLACE TRANSFORM *operator* (Definition S.1 page 389) *and* $\star$ *the convolution operator* (Definition D.1 page 199). *Let* A, B, C, *and* D *be defined as in Corollary S.2 (page 391).*

THM

$$\begin{aligned}
\mathbf{L}\left[\mathsf{f}(x)\star\mathsf{g}(x)\right](s) &= [\mathbf{L}\mathsf{f}](s)\left[\mathbf{L}\mathsf{g}\right](s) && \forall \mathbf{R_e}(s) \in [A : B] \cap [C : D] && \forall \mathsf{f},\mathsf{g}\in L^2_{(\mathbb{R},\mathscr{B},\mu)} \\
\mathbf{L}\left[\mathsf{f}(x)\mathsf{g}(x)\right](s) &= [\mathbf{L}\mathsf{f}](s)\star\left[\mathbf{L}\mathsf{g}\right](s) && \forall \mathbf{R_e}(s) \in [A+C : B+D],\ c\in(A : B) && \forall \mathsf{f},\mathsf{g}\in L^2_{(\mathbb{R},\mathscr{B},\mu)}.
\end{aligned}$$

[18] Bracewell (1978) page 224, Bachman et al. (2002) pages 268–270, Bachman (1964) page 8

✎PROOF:

$$\begin{aligned}
\mathbf{L}\big[\mathsf{f}(x)\star\mathsf{g}(x)\big](s) &= \mathbf{L}\left[\int_{u\in\mathbb{R}}\mathsf{f}(u)\mathsf{g}(x-u)\,\mathsf{d}u\right](s) && \text{by definition of }\star && \text{(Definition D.1 page 199)}\\
&= \int_{u\in\mathbb{R}}\mathsf{f}(u)\big[\mathbf{L}\mathsf{g}(x-u)\big](s)\,\mathsf{d}u\\
&= \int_{u\in\mathbb{R}}\mathsf{f}(u)e^{-su}\,\big[\mathbf{L}\mathsf{g}(x)\big](s)\,\mathsf{d}u && \text{by \textit{translation} property} && \text{(Theorem S.2 page 390)}\\
&= \underbrace{\left(\int_{u\in\mathbb{R}}\mathsf{f}(u)e^{-su}\,\mathsf{d}u\right)}_{[\mathbf{L}\mathsf{f}](s)}\big[\mathbf{L}\mathsf{g}\big](s)\\
&= [\mathbf{L}\mathsf{f}](s)\,\big[\mathbf{L}\mathsf{g}\big](s) && \mathbf{R_e}(s)\in[A:B]\cap[C:D] && \text{by definition of }\mathbf{L}
\end{aligned}$$

$$\begin{aligned}
\mathbf{L}[\mathsf{f}(x)\mathsf{g}(x)](s) &= \mathbf{L}\big[\big(\mathbf{L}^{-1}\mathbf{L}\mathsf{f}(x)\big)\,\mathsf{g}(x)\big](s) && \text{by def. of operator inverse} && \text{(Definition R.2 page 359)}\\
&= \mathbf{L}\left[\left(\int_{v\in\mathbb{R}}[\mathbf{L}\mathsf{f}(x)](v)e^{sxv}\,\mathsf{d}v\right)\mathsf{g}(x)\right](s) && \text{by Theorem S.1 page 390}\\
&= \int_{v\in\mathbb{R}}[\mathbf{L}\mathsf{f}(x)](v)\big[\mathbf{L}\big(e^{sxv}\,\mathsf{g}(x)\big)\big](s,v)\,\mathsf{d}v\\
&= \int_{v\in\mathbb{R}}[\mathbf{L}\mathsf{f}(x)](v)\big[\mathbf{L}\mathsf{g}(x)\big](s-v)\,\mathsf{d}v && \text{by Theorem S.2 page 390}\\
&= [\mathbf{L}\mathsf{f}](s)\star\big[\mathbf{L}\mathsf{g}\big](s) && \text{by definition of }\star && \text{(Definition D.1 page 199)}
\end{aligned}$$

✏

S.9 Calculus properties

Theorem S.7. [19] *Let* $\mathbf{L}$ *be the* LAPLACE TRANSFORM *operator (Definition S.1 page 389).*

THM

$$\begin{aligned}
\mathbf{L}\left[\frac{\mathsf{d}}{\mathsf{d}x}\mathsf{g}(x)\right](s) &= s\big[\mathbf{L}\mathsf{g}\big](s)+\lim_{x\to\infty}\big[e^{-sx}\mathsf{g}(x)-e^{sx}\mathsf{g}(-x)\big]\\
\mathbf{L}\left[\int_{u=-\infty}^{u=x}\mathsf{g}(u)\,\mathsf{d}u\right](s) &= \frac{1}{s}\big[\mathbf{L}\mathsf{g}\big](s)
\end{aligned}$$

✎PROOF:

$$\begin{aligned}
\mathbf{L}\left[\frac{\mathsf{d}}{\mathsf{d}x}\mathsf{g}(x)\right] &\triangleq \int_{x\in\mathbb{R}}\underbrace{\left[\frac{\mathsf{d}}{\mathsf{d}x}\mathsf{g}(x)\right]}_{\mathsf{d}v}\underbrace{e^{-sx}}_{u}\,\mathsf{d}x && \text{by definition of }\mathbf{L} && \text{(Definition S.1 page 389)}\\
&= \Big[\underbrace{e^{-sx}}_{u}\underbrace{\mathsf{g}(x)}_{v}\Big]_{x=-\infty}^{x=+\infty}-\left[\int_{x\in\mathbb{R}}\underbrace{\mathsf{g}(x)}_{v}\underbrace{(-s)e^{-sx}\,\mathsf{d}x}_{\mathsf{d}u}\right] && \text{by \textit{Integration by Parts}}\\
&= \lim_{x\to\infty}\big[e^{-sx}\mathsf{g}(x)\big]-\lim_{x\to-\infty}\big[e^{-sx}\mathsf{g}(-\infty)\big]+s\int_{x\in\mathbb{R}}\mathsf{g}(x)e^{-sx}\,\mathsf{d}x\\
&= s\big[\mathbf{L}\mathsf{g}\big](s)+\lim_{x\to\infty}\big[e^{-sx}\mathsf{g}(x)-e^{sx}\mathsf{g}(-x)\big] && \text{by definition of }\mathbf{L} && \text{(Definition S.1 page 389)}
\end{aligned}$$

[19] Betten (2008b) page 301 ⟨(B.27)⟩, Levy (1958) page 15 ⟨Equation 0.7⟩

$$\begin{aligned}
\mathbf{L}\int_{v=-\infty}^{v=x} \mathsf{g}(v)\,\mathsf{d}v &\triangleq \int_{x=-\infty}^{x=+\infty}\left[\int_{v=-\infty}^{v=x}\mathsf{g}(v)\,\mathsf{d}v\right]e^{-sx}\,\mathsf{d}x && \text{by definition of } \mathbf{L} && \text{(Definition S.1 page 389)}\\
&= \int_{x=-\infty}^{x=+\infty}\left[\int_{v=-\infty}^{v=+\infty}\mathsf{g}(v)\mu(x-v)\,\mathsf{d}v\right]e^{-sx}\,\mathsf{d}x && \text{by definition of } \mu(x) && \text{(Definition S.3 page 394)}\\
&= \int_{v\in\mathbb{R}}\mathsf{g}(v)\left[\int_{x\in\mathbb{R}}\mu(x-v)e^{-sx}\,\mathsf{d}x\right]\mathsf{d}v\\
&= \int_{v\in\mathbb{R}}\mathsf{g}(v)\left[e^{-vs}\frac{1}{s}\right]\mathsf{d}v && \forall\mathbf{R_e}(s)<0 && \text{by Corollary S.4 page 394}\\
&= \frac{1}{s}\int_{v\in\mathbb{R}}\mathsf{g}(v)e^{-vs}\,\mathsf{d}v && \forall\mathbf{R_e}(s)<0 && \text{by Corollary S.4 page 394}\\
&\triangleq \frac{1}{s}\left[\mathbf{L}\mathsf{g}\right](s) && \text{by definition of } \mathbf{L} && \text{(Definition S.1 page 389)}
\end{aligned}$$

✎

Example S.1.

EX The *Laplace Transform* of the *impedance* $Z(s)$ of an ideal capacitor with *capacitance* C, initial voltage $\mathsf{v}(t)=0$ for $t<0$, and finite voltage over all time, is

$$Z(s)=\frac{1}{sC}.$$

✎PROOF:

1. An ideal capacitor is modeled with the equation $\mathsf{i}(t)=C\frac{\mathsf{d}}{\mathsf{d}t}\mathsf{v}(t)$.
2. lemma: Because $\mathsf{v}(t)=0$ for $t<0$, then $\lim_{t\to\infty}\left[e^{st}\mathsf{v}(-t)\right]=0$.
3. lemma: Because voltage $\mathsf{v}(t)$ is finite, $\lim_{t\to\infty}\left[e^{-st}\mathsf{v}(t)\right]=0$.
4. Then the impedance $\mathsf{Z}(s)$ of the capacitor is ...

$$\begin{aligned}
\mathsf{Z}(s) &\triangleq \frac{\mathsf{V}(s)}{\mathsf{I}(s)} && \text{by definition of \textit{impedance} } \mathsf{Z}(s)\\
&\triangleq \frac{\mathsf{V}(s)}{\mathbf{L}[\mathsf{i}(t)]} && \text{by definition of \textit{Laplace Transform} } \mathsf{I}(s) \text{ of } \mathsf{i}(t)\\
&\triangleq \frac{\mathsf{V}(s)}{\mathbf{L}\left[C\frac{\mathsf{d}}{\mathsf{d}t}\mathsf{v}(t)\right]} && \text{by definition of \textit{ideal Capacitor} with \textit{Capacitance} } C\\
&= \frac{\mathsf{V}(s)}{C\mathbf{L}\left[\frac{\mathsf{d}}{\mathsf{d}t}\mathsf{v}(t)\right]} && \text{by \textit{homogeneous} property of } \mathbf{L} \text{ Theorem S.3 (page 391)}\\
&= \frac{\mathsf{V}(s)}{Cs\left(\mathbf{L}[\mathsf{v}(t)]+\lim_{t\to\infty}\left[e^{-st}\mathsf{v}(t)+e^{st}\mathsf{v}(-t)\right]\right)} && \text{by Theorem S.7 page 402}\\
&= \frac{\mathsf{V}(s)}{Cs\mathbf{L}[\mathsf{v}(t)]} && \text{by (2) lemma and (3) lemma}\\
&\triangleq \frac{\mathsf{V}(s)}{sC\mathsf{V}(s)} && \text{by definition of } \mathsf{V}(s)\\
&= \frac{1}{sC}
\end{aligned}$$

✎

Example S.2.

EX The *Laplace Transform* of the *impedance* $Z(s)$ of an ideal inductor with *inductance* L, initial current $\mathsf{i}(t) = 0$ for $t < 0$, and finite current over all time, is
$$Z(s) = sL.$$

✎PROOF:

1. An ideal inductor is modeled with the equation $\mathsf{v}(t) = L\frac{\mathsf{d}}{\mathsf{dt}}\mathsf{i}(t)$.
2. lemma: Because $\mathsf{i}(t) = 0$ for $t < 0$, then $\lim_{t\to\infty}\left[e^{st}\mathsf{i}(-t)\right] = 0$.
3. lemma: Because $\mathsf{i}(t)$ is finite, $\lim_{t\to\infty}\left[e^{-st}\mathsf{i}(t)\right] = 0$.
4. Then the impedance $\mathsf{Z}(s)$ of the inductor is ...

$$\begin{aligned}
\mathsf{Z}(s) &\triangleq \frac{\mathsf{V}(s)}{\mathtt{I}(s)} && \text{by definition of \textit{impedance} } \mathsf{Z}(s)\\
&\triangleq \frac{\mathbf{L}[\mathsf{v}(t)]}{\mathtt{I}(s)} && \text{by definition of \textit{Laplace Transform} } \mathtt{I}(s) \text{ of } \mathsf{i}(t)\\
&\triangleq \frac{\mathbf{L}\left[L\frac{\mathsf{d}}{\mathsf{dt}}\mathsf{i}(t)\right]}{\mathtt{I}(s)} && \text{by definition of \textit{ideal inductor} with \textit{inductance} } L\\
&= \frac{L\mathbf{L}\left[\frac{\mathsf{d}}{\mathsf{dt}}\mathsf{i}(t)\right]}{\mathtt{I}(s)} && \text{by \textit{homogeneous} property of } \mathbf{L} \text{ Theorem S.3 (page 391)}\\
&= \frac{Ls\left(\mathbf{L}[\mathsf{i}(t)] + \lim_{t\to\infty}\left[e^{-st}\mathsf{i}(t) + e^{st}\mathsf{i}(-t)\right]\right)}{\mathtt{I}(s)} && \text{by Theorem S.7 page 402}\\
&= \frac{Ls\mathbf{L}[\mathsf{i}(t)]}{\mathtt{I}(s)} && \text{by (2) lemma and (3) lemma}\\
&\triangleq \frac{sL\mathtt{I}(s)}{\mathtt{I}(s)} && \text{by definition of } \mathtt{I}(s)\\
&= sL
\end{aligned}$$

✏

S.10 Moment properties

Definition S.4. [20]

DEF *The quantity* M_n *is the* n***th moment*** *of a function* $\mathsf{f}(x) \in L^2_{\mathbb{R}}$ *if*
$$\mathsf{M}_n \triangleq \int_{\mathbb{R}} x^n \mathsf{f}(x)\,\mathsf{d}x \qquad \text{for } n \in \mathbb{W}.$$

Theorem S.8. [21] *Let* M_n *be the* nTH MOMENT *(Definition S.4 page 404) of a function* $\mathsf{f}(x) \in L^2_{\mathbb{R}}$.

THM
$$\mathsf{M}_n \triangleq \int_{\mathbb{R}} x^n \mathsf{f}(x)\,\mathsf{d}x = \left[(-1)^n \left[\left[\frac{\mathsf{d}}{\mathsf{ds}}\right]^n \mathbf{L}\mathsf{f}\right](s)\right]_{s=0} \qquad \forall n\in\mathbb{W}, \mathsf{f}\in L^2_{\mathbb{R}}$$

[20] Jawerth and Sweldens (1994) pages 16–17, Sweldens and Piessens (1993) page 2, Vidakovic (1999) page 83
[21] Goswami and Chan (1999) pages 38–39

✎PROOF:

$$\begin{aligned}
(-1)^n\left[\left[\tfrac{\mathrm{d}}{\mathrm{d}s}\right]^n \mathbf{L}\mathsf{f}\right](s)_{s=0} &= (-1)^n\left[\left[\tfrac{\mathrm{d}}{\mathrm{d}s}\right]^n \int_{\mathbb{R}} \mathsf{f}(x)e^{-sx}\,\mathrm{d}x\right]_{s=0} && \text{by definition of } \mathbf{L} \quad \text{(Definition S.1 page 389)}\\
&= (-1)^n\left[\int_{\mathbb{R}} \mathsf{f}(x)\left[\left[\tfrac{\mathrm{d}}{\mathrm{d}s}\right]^n e^{-sx}\right]\mathrm{d}x\right]_{s=0}\\
&= (-1)^n\left[\int_{\mathbb{R}} \mathsf{f}(x)(-x)^n e^{-sx}\,\mathrm{d}x\right]_{s=0}\\
&= (-1)^n \int_{\mathbb{R}} \mathsf{f}(x)(-x)^n\,\mathrm{d}x\\
&= (-1)^{2n} \int_{\mathbb{R}} \mathsf{f}(x)x^n\,\mathrm{d}x\\
&\triangleq \mathsf{M}_n && \text{by definition of } \mathsf{M}_n \quad \text{(Definition T.3 page 412)}
\end{aligned}$$

✏

Corollary S.13. [22] *Let* M_n *be the* nTH MOMENT *(Definition S.4 page 404) of a function* $\mathsf{f}(x)$ *in* $\boldsymbol{L}^2_{\mathbb{R}}$.

COR
$$\mathsf{M}_n = 0 \quad\iff\quad \left[\left[\tfrac{\mathrm{d}}{\mathrm{d}s}\right]^n \mathbf{L}\mathsf{f}\right]_{s=0} = 0 \qquad \forall n\in\mathbb{W}$$

✎PROOF:

1. Proof for ($\Longrightarrow$) case:

$$\begin{aligned}
0 &= (-1)^n \cdot 0\\
&= (-1)^n \mathsf{M}_n && \text{by left hypothesis}\\
&= (-1)^n \int_{\mathbb{R}} x^n \mathsf{f}(x)\,\mathrm{d}x\\
&= (-1)^n\left[(-1)^n\left[\tfrac{\mathrm{d}}{\mathrm{d}s}\right]^n \mathbf{L}\mathsf{f}\right]_{s=0} && \text{by Theorem S.8 page 404}\\
&= \left[\left[\tfrac{\mathrm{d}}{\mathrm{d}s}\right]^n \mathbf{L}\mathsf{f}\right]_{s=0}
\end{aligned}$$

2. Proof for ($\Longleftarrow$) case:

$$\begin{aligned}
0 &= \left[\left[\tfrac{\mathrm{d}}{\mathrm{d}s}\right]^n \mathbf{L}\mathsf{f}\right]_{s=0} && \text{by right hypothesis}\\
&= \left[\left[\tfrac{\mathrm{d}}{\mathrm{d}s}\right]^n \int_{\mathbb{R}} \mathsf{f}(x)e^{-sx}\,\mathrm{d}x\right]_{s=0} && \text{by definition of } \mathbf{L} \quad \text{(Definition S.1 page 389)}\\
&= \left[\int_{\mathbb{R}} \mathsf{f}(x)\left[\tfrac{\mathrm{d}}{\mathrm{d}s}\right]^n e^{-sx}\,\mathrm{d}x\right]_{s=0}\\
&= \left[\int_{\mathbb{R}} \mathsf{f}(x)\left[(-1)^n x^n e^{-sx}\right]\mathrm{d}x\right]_{s=0}\\
&= (-1)^n \int_{\mathbb{R}} \mathsf{f}(x)x^n\,\mathrm{d}x
\end{aligned}$$

✏

[22] Vidakovic (1999) pages 82–83, Mallat (1999) pages 241–242

APPENDIX T

FOURIER TRANSFORM

❝ *Up to this point we have supposed that the function whose development is required in a series of sines of multiple arcs can be developed in a series arranged according to powers of the variable x, ... We can extend the same results to any functions, even to those which are discontinuous and entirely arbitrary. ... even entirely arbitrary functions may be developed in series of sines of multiple arcs.* ❞

Joseph Fourier (1768–1830) [1]

T.1 Introduction

Historically, before the Fourier Transform was the Taylor Expansion (transform). The Taylor Expansion demonstrates that for **analytic** functions, knowledge of the derivatives of a function at a location $x = a$ allows you to determine (predict) arbitrarily closely all the points $\mathtt{f}(x)$ in the vicinity of $x = a$. But analytic functions are by definition functions for which all their derivatives exist. Thus, if a function is *discontinuous*, it is simply not a candidate for the Taylor Expansion. And some 300 years ago, mathematician giants of the day were fairly content with this.

But then in came an engineer named Joseph Fourier whose day job was working as a governor of lower Egypt under Napolean. He claimed that, rather than expansion based on derivatives, one could expand based on integrals over sinusoids, and that this would work not just for analytic functions, but for **discontinuous** ones as well![2]

Needless to say, this did not go over too well initially in the mathematical community. But over time (on the order of 200 or so years), the Fourier Transform has in many ways won the day.


[3]

[1] quote: Fourier (1878) page 184,186 ⟨§219,220⟩
image: http://en.wikipedia.org/wiki/File:Fourier2.jpg, public domain

[2] Robinson (1982) page 886

[3] Caricature of Legendre (left) and Fourier (right), 1820, by Julien-Léopold Boilly (1796–1874). *“Album de 73*

T.2 Definitions

This chapter deals with the *Fourier Transform* in the space of *Lebesgue square-integrable functions* $\boldsymbol{L}^2_{(\mathbb{R},\mathscr{B},\mu)}$, where $\mathbb{R}$ is the set of real numbers, $\mathscr{B}$ is the set of *Borel sets* on $\mathbb{R}$, μ is the standard *Borel measure* on $\mathscr{B}$, and

$$\boldsymbol{L}^2_{(\mathbb{R},\mathscr{B},\mu)} \triangleq \left\{ \mathsf{f} \in \mathbb{R}^{\mathbb{R}} \,\middle|\, \int_{\mathbb{R}} |\mathsf{f}|^2 \,\mathrm{d}\mu < \infty \right\}.$$

Furthermore, $\langle \triangle \,|\, \triangledown \rangle$ is the *inner product* induced by the operator $\int_{\mathbb{R}} \,\mathrm{d}\mu$ such that

$$\langle \mathsf{f} \,|\, \mathsf{g} \rangle \triangleq \int_{\mathbb{R}} \mathsf{f}(x)\mathsf{g}^*(x)\,\mathrm{d}x,$$

and $\left(\boldsymbol{L}^2_{(\mathbb{R},\mathscr{B},\mu)}, \langle \triangle \,|\, \triangledown \rangle\right)$ is a *Hilbert space*.

Definition T.1. *Let κ be a* FUNCTION *in* $\mathbb{C}^{\mathbb{R}^2}$.

DEF: *The function κ is the **Fourier kernel** if* $\quad \kappa(x,\omega) \triangleq e^{i\omega x} \qquad \forall x,\omega\in\mathbb{R}$

Definition T.2. [4] *Let $\boldsymbol{L}^2_{(\mathbb{R},\mathscr{B},\mu)}$ be the space of all* LEBESGUE SQUARE-INTEGRABLE FUNCTIONS.

DEF: *The **Fourier Transform** operator $\tilde{\mathbf{F}}$ is defined as*

$$\left[\tilde{\mathbf{F}}\mathsf{f}\right](\omega) \triangleq \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathsf{f}(x)\,e^{-i\omega x}\,\mathrm{d}x \qquad \forall \mathsf{f}\in\boldsymbol{L}^2_{(\mathbb{R},\mathscr{B},\mu)}$$

*This definition of the Fourier Transform is also called the **unitary Fourier Transform**.*

Remark T.1 (**Fourier transform scaling factor**). [5] If the Fourier transform operator $\tilde{\mathbf{F}}$ and inverse Fourier transform operator $\tilde{\mathbf{F}}^{-1}$ are defined as

$$\tilde{\mathbf{F}}\mathsf{f}(x) \triangleq \mathsf{F}(\omega) \triangleq A \int_{\mathbb{R}} \mathsf{f}(x)e^{-i\omega x}\,\mathrm{d}x \qquad \text{and} \qquad \tilde{\mathbf{F}}^{-1}\tilde{\mathsf{f}}(\omega) \triangleq B \int_{\mathbb{R}} \mathsf{F}(\omega)e^{i\omega x}\,\mathrm{d}\omega$$

then A and B can be any constants as long as $AB = \frac{1}{2\pi}$. The Fourier transform is often defined with the scaling factor A set equal to 1 such that $\left[\tilde{\mathbf{F}}\mathsf{f}(x)\right](\omega) \triangleq \int_{\mathbb{R}} \mathsf{f}(x)\,e^{-i\omega x}\,\mathrm{d}x$. In this case, the inverse Fourier transform operator $\tilde{\mathbf{F}}^{-1}$ is either defined as

- $\left[\tilde{\mathbf{F}}^{-1}\mathsf{f}(x)\right](f) \triangleq \int_{\mathbb{R}} \mathsf{f}(x)\,e^{i2\pi f x}\,\mathrm{d}x$ (using oscillatory frequency free variable f) or
- $\left[\tilde{\mathbf{F}}^{-1}\mathsf{f}(x)\right](\omega) \triangleq \frac{1}{2\pi}\int_{\mathbb{R}} \mathsf{f}(x)\,e^{i\omega x}\,\mathrm{d}x$ (using angular frequency free variable ω).

In short, the 2π has to show up somewhere, either in the argument of the exponential ($e^{-i2\pi f t}$) or in front of the integral ($\frac{1}{2\pi}\int\cdots$). One could argue that it is unnecessary to burden the exponential argument with the 2π factor ($e^{-i2\pi f t}$), and thus could further argue in favor of using the angular frequency variable ω thus giving the inverse operator definition $\left[\tilde{\mathbf{F}}^{-1}\mathsf{f}(x)\right](\omega) \triangleq \frac{1}{2\pi}\int_{\mathbb{R}} \mathsf{f}(x)\,e^{-i\omega x}\,\mathrm{d}x$. But this causes a new problem. In this case, the Fourier operator $\tilde{\mathbf{F}}$ is not *unitary* (see Theorem T.2 page 409)—in particular, $\tilde{\mathbf{F}}\tilde{\mathbf{F}}^* \neq \mathbf{I}$, where $\tilde{\mathbf{F}}^*$ is the *adjoint* of $\tilde{\mathbf{F}}$; but rather, $\tilde{\mathbf{F}}\left(\frac{1}{2\pi}\tilde{\mathbf{F}}^*\right) = \left(\frac{1}{2\pi}\tilde{\mathbf{F}}^*\right)\tilde{\mathbf{F}} = \mathbf{I}$. But if we define the operators $\tilde{\mathbf{F}}$ and $\tilde{\mathbf{F}}^{-1}$ to both have the scaling factor $\frac{1}{\sqrt{2\pi}}$, then $\tilde{\mathbf{F}}$ and $\tilde{\mathbf{F}}^{-1}$ are inverses *and* $\tilde{\mathbf{F}}$ is *unitary*—that is, $\tilde{\mathbf{F}}\tilde{\mathbf{F}}^* = \tilde{\mathbf{F}}^*\tilde{\mathbf{F}} = \mathbf{I}$.

Portraits-Charge Aquarelle's des Membres de I'Institute (watercolor portrait #29). Biliotheque de l'Institut de France." Public domain. https://en.wikipedia.org/wiki/File:Legendre_and_Fourier_(1820).jpg

[4] Bachman et al. (2002) page 363, Chorin and Hald (2009) page 13, Loomis and Bolker (1965) page 144, Knapp (2005b) pages 374–375, Fourier (1822), Fourier (1878) page 336?

[5] Chorin and Hald (2009) page 13, Jeffrey and Dai (2008) pages xxxi–xxxii, Knapp (2005b) pages 374–375

T.3 Operator properties

Theorem T.1 (Inverse Fourier transform). [6] *Let* $\tilde{\mathbf{F}}$ *be the Fourier Transform operator* (Definition T.2 page 408). *The inverse* $\tilde{\mathbf{F}}^{-1}$ *of* $\tilde{\mathbf{F}}$ *is*

THM
$$\left[\tilde{\mathbf{F}}^{-1}\tilde{\mathsf{f}}\right](x) = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} \tilde{\mathsf{f}}(\omega)e^{i\omega x}\,\mathrm{d}\omega \qquad \forall\tilde{\mathsf{f}}\in L^2_{(\mathbb{R},\mathscr{B},\mu)}$$

Theorem T.2. *Let* $\tilde{\mathbf{F}}$ *be the Fourier Transform operator with inverse* $\tilde{\mathbf{F}}^{-1}$ *and adjoint* $\tilde{\mathbf{F}}^{*}$.

THM
$$\tilde{\mathbf{F}}^{*} = \tilde{\mathbf{F}}^{-1}$$

✎PROOF:

$$\begin{aligned}
\left\langle \tilde{\mathbf{F}}\mathsf{f} \,|\, \mathsf{g}\right\rangle &= \left\langle \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} \mathsf{f}(x)\, e^{-i\omega x}\,\mathrm{d}x \,|\, \mathsf{g}(\omega)\right\rangle && \text{by definition of } \tilde{\mathbf{F}} \text{ page 408}\\
&= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} \mathsf{f}(x)\,\left\langle e^{-i\omega x} \,|\, \mathsf{g}(\omega)\right\rangle\,\mathrm{d}x && \text{by \textit{additive property} of } \langle\triangle\,|\,\nabla\rangle \text{ page 309}\\
&= \int_{\mathbb{R}} \mathsf{f}(x)\,\frac{1}{\sqrt{2\pi}}\left\langle \mathsf{g}(\omega) \,|\, e^{-i\omega x}\right\rangle^{*}\,\mathrm{d}x && \text{by \textit{conjugate symmetric property} of } \langle\triangle\,|\,\nabla\rangle \text{ page 309}\\
&= \left\langle \mathsf{f}(x) \,|\, \frac{1}{\sqrt{2\pi}}\left\langle \mathsf{g}(\omega) \,|\, e^{-i\omega x}\right\rangle\right\rangle && \text{by definition of } \langle\triangle\,|\,\nabla\rangle\\
&= \left\langle \mathsf{f} \,|\, \underbrace{\tilde{\mathbf{F}}^{-1}}_{\tilde{\mathbf{F}}^{*}}\mathsf{g}\right\rangle && \text{by Theorem T.1 page 409}
\end{aligned}$$

✏

The Fourier Transform operator has several nice properties:

- $\tilde{\mathbf{F}}$ is *unitary* [7] (Corollary T.1—next corollary).
- Because $\tilde{\mathbf{F}}$ is unitary, it automatically has several other nice properties (Theorem T.3 page 409).

Corollary T.1. *Let* $\mathbf{I}$ *be the identity operator and let* $\tilde{\mathbf{F}}$ *be the Fourier Transform operator with adjoint* $\tilde{\mathbf{F}}^{*}$ *and inverse* $\tilde{\mathbf{F}}^{-1}$.

COR
$$\underbrace{\tilde{\mathbf{F}}\tilde{\mathbf{F}}^{*} = \tilde{\mathbf{F}}^{*}\tilde{\mathbf{F}} = \mathbf{I}}_{\tilde{\mathbf{F}}^{*} = \tilde{\mathbf{F}}^{-1}} \qquad (\tilde{\mathbf{F}} \textit{ is unitary})$$

✎PROOF: This follows directly from the fact that $\tilde{\mathbf{F}}^{*} = \tilde{\mathbf{F}}^{-1}$ (Theorem T.2 page 409). ✏

Theorem T.3. *Let* $\tilde{\mathbf{F}}$ *be the Fourier transform operator with adjoint* $\tilde{\mathbf{F}}^{*}$ *and inverse* $\tilde{\mathbf{F}}$. *Let* $|||\cdot|||$ *be the operator norm with respect to the vector norm* $\|\cdot\|$ *with respect to the Hilbert space* $\left(\mathbb{C}^{\mathbb{R}}, \langle\triangle\,|\,\nabla\rangle\right)$. *Let* $\mathscr{R}(\mathbf{A})$ *be the range of an operator* $\mathbf{A}$.

THM
$$\begin{array}{rclcll}
\mathscr{R}(\mathbf{F}\tau) &=& \mathscr{R}(\tilde{\mathbf{F}}^{-1}) &=& L^2_{\mathbb{R}} & \\
|||\tilde{\mathbf{F}}||| &=& |||\tilde{\mathbf{F}}^{-1}||| &=& 1 & \text{(UNITARY)}\\
\left\langle \tilde{\mathbf{F}}\mathsf{f} \,|\, \tilde{\mathbf{F}}\mathsf{g}\right\rangle &=& \left\langle \tilde{\mathbf{F}}^{-1}\mathsf{f} \,|\, \tilde{\mathbf{F}}^{-1}\mathsf{g}\right\rangle &=& \langle \mathsf{f} \,|\, \mathsf{g}\rangle & \text{(PARSEVAL'S EQUATION)}\\
\|\tilde{\mathbf{F}}\mathsf{f}\| &=& \|\tilde{\mathbf{F}}^{-1}\mathsf{f}\| &=& \|\mathsf{f}\| & \text{(PLANCHEREL'S FORMULA)}\\
\|\tilde{\mathbf{F}}\mathsf{f} - \tilde{\mathbf{F}}\mathsf{g}\| &=& \|\tilde{\mathbf{F}}^{-1}\mathsf{f} - \tilde{\mathbf{F}}^{-1}\mathsf{g}\| &=& \|\mathsf{f} - \mathsf{g}\| & \text{(ISOMETRIC)}
\end{array}$$

✎PROOF: These results follow directly from the fact that $\tilde{\mathbf{F}}$ is unitary (Corollary T.1 page 409) and from the properties of unitary operators (Theorem R.26 page 383). ✏

[6] Chorin and Hald (2009) page 13

[7] *unitary operators*: Definition R.11 page 382

T.4 Transversal properties

Theorem T.4 (Shift relations). *Let $\tilde{\mathbf{F}}$ be the Fourier Transform operator* (Definition T.2 page 408).

THM
$$\begin{aligned}\tilde{\mathbf{F}}[\mathsf{f}(x-y)](\omega) &= e^{-i\omega y}\left[\tilde{\mathbf{F}}\mathsf{f}(x)\right](\omega)\\ \left[\tilde{\mathbf{F}}\left(e^{irx}\mathsf{g}(x)\right)\right](\omega) &= \left[\tilde{\mathbf{F}}\mathsf{g}(x)\right](\omega-r)\end{aligned}$$

✎PROOF: Let $\mathbf{L}$ be the *Laplace Transform* operator (Definition S.1 page 389).

$$\begin{aligned}
\tilde{\mathbf{F}}[\mathsf{f}(x-y)](\omega) &= \mathbf{L}[\mathsf{f}(x-y)](s)\big|_{s=i\omega} && \text{by definition of } \mathbf{L} && \text{(Definition S.1 page 389)}\\
&= e^{-sy}\,[\mathbf{L}\mathsf{f}(x)](s)\big|_{s=i\omega} && \text{by Laplace \textit{translation} property} && \text{(Theorem S.2 page 390)}\\
&= e^{-i\omega y}\left[\tilde{\mathbf{F}}\mathsf{f}(x)\right](\omega) && \text{by definition of } \tilde{\mathbf{F}} && \text{(Definition T.2 page 408)}\\
\left[\tilde{\mathbf{F}}\left(e^{irx}\mathsf{g}(x)\right)\right](\omega) &= \left[\mathbf{L}\left(e^{irx}\mathsf{g}(x)\right)\right](s)\big|_{s=i\omega} && \text{by definition of } \mathbf{L} && \text{(Definition S.1 page 389)}\\
&= \left[\left[\mathbf{L}\mathsf{g}(x)\right](s-r)\right]\big|_{s=i\omega} && \text{by Laplace \textit{dilation} property} && \text{(Theorem S.2 page 390)}\\
&= \left[\tilde{\mathbf{F}}\mathsf{g}(x)\right](\omega-r) && \text{by definition of } \tilde{\mathbf{F}} && \text{(Definition T.2 page 408)}
\end{aligned}$$

⇒

Theorem T.5 (Complex conjugate). *Let $\tilde{\mathbf{F}}$ be the Fourier Transform operator and $*$ represent the complex conjugate operation on the set of complex numbers.*

THM
$$\tilde{\mathbf{F}}\mathsf{f}^*(-x) = -\left[\tilde{\mathbf{F}}\mathsf{f}(x)\right]^* \qquad \forall \mathsf{f}\in L^2_{(\mathbb{R},\mathscr{B},\mu)}$$

✎PROOF:

$$\begin{aligned}
\left[\tilde{\mathbf{F}}\mathsf{f}^*(-x)\right](\omega) &\triangleq \frac{1}{\sqrt{2\pi}}\int \mathsf{f}^*(-x)e^{-i\omega x}\,\mathsf{d}x && \text{by definition of } \tilde{\mathbf{F}} && \text{(Definition T.2 page 408)}\\
&= \frac{1}{\sqrt{2\pi}}\int \mathsf{f}^*(u)e^{i\omega u}(-1)\,\mathsf{d}u && \text{where } u\triangleq -x \implies \mathsf{d}x = -\,\mathsf{d}u\\
&= -\left[\frac{1}{\sqrt{2\pi}}\int \mathsf{f}(u)e^{-i\omega u}\,\mathsf{d}u\right]^*\\
&\triangleq -\left[\tilde{\mathbf{F}}\mathsf{f}(x)\right]^* && \text{by definition of } \tilde{\mathbf{F}} && \text{(Definition T.2 page 408)}
\end{aligned}$$

⇒

T.5 Convolution relations

Theorem V.2 (next) demonstrates that multiplication in the “time domain” is equivalent to convolution in the “frequency domain” and vice-versa.

Theorem T.6 (convolution theorem). [8] *Let $\tilde{\mathbf{F}}$ be the Fourier Transform operator* (Definition T.2 page 408) *and $\star$ the convolution operator* (Definition D.1 page 199).

[8] Bachman et al. (2002) pages 269–270 ⟨5.2.3 Convolutions to Products⟩, Bachman (1964) page 8, Bracewell (1978) page 110

THM

$$\underbrace{\tilde{\mathbf{F}}\left[\mathsf{f}(x) \star \mathsf{g}(x)\right](\omega)}_{\text{convolution in "time domain"}} = \underbrace{\sqrt{2\pi}\left[\tilde{\mathbf{F}}\mathsf{f}\right](\omega)\left[\tilde{\mathbf{F}}\mathsf{g}\right](\omega)}_{\text{multiplication in "frequency domain"}} \qquad \forall \mathsf{f},\mathsf{g} \in \boldsymbol{L}^2_{(\mathbb{R},\mathscr{B},\mu)}$$

$$\underbrace{\tilde{\mathbf{F}}\left[\mathsf{f}(x)\mathsf{g}(x)\right](\omega)}_{\text{multiplication in "time domain"}} = \underbrace{\frac{1}{\sqrt{2\pi}}\left[\tilde{\mathbf{F}}\mathsf{f}\right](\omega) \star \left[\tilde{\mathbf{F}}\mathsf{g}\right](\omega)}_{\text{convolution in "frequency domain"}} \qquad \forall \mathsf{f},\mathsf{g} \in \boldsymbol{L}^2_{(\mathbb{R},\mathscr{B},\mu)}.$$

✎PROOF: Let $\mathbf{L}$ be the *Laplace Transform* operator (Definition S.1 page 389).

$$\begin{aligned}
\tilde{\mathbf{F}}\left[\mathsf{f}(x) \star \mathsf{g}(x)\right](\omega) &= \mathbf{L}\left[\mathsf{f}(x) \star \mathsf{g}(x)\right](s)\Big|_{s=i\omega} && \text{by definition of } \mathbf{L} && \text{(Definition S.1 page 389)}\\
&= \sqrt{2\pi}[\mathbf{L}\mathsf{f}](s)\left[\mathbf{L}\mathsf{g}\right](s)\Big|_{s=i\omega} && \text{by \textit{Laplace convolution} result} && \text{(Theorem S.6 page 401)}\\
&= \sqrt{2\pi}\left[\tilde{\mathbf{F}}\mathsf{f}\right](\omega)\left[\tilde{\mathbf{F}}\mathsf{g}\right](\omega)\\
\tilde{\mathbf{F}}[\mathsf{f}(x)\mathsf{g}(x)](\omega) &= \mathbf{L}[\mathsf{f}(x)\mathsf{g}(x)](s)|_{s=i\omega}\\
&= \frac{1}{\sqrt{2\pi}}[\mathbf{L}\mathsf{f}](s) \star \left[\mathbf{L}\mathsf{g}\right](s)\Big|_{s=i\omega}\\
&= \frac{1}{\sqrt{2\pi}}\left[\tilde{\mathbf{F}}\mathsf{f}\right](\omega) \star \left[\tilde{\mathbf{F}}\mathsf{g}\right](\omega)
\end{aligned}$$

✏

T.6 Calculus relations

Theorem T.7. *Let $\tilde{\mathbf{F}}$ be the* FOURIER TRANSFORM *operator (Definition T.2 page 408).*

THM

$$\left\{\lim_{t\to-\infty} \mathsf{x}(t) = 0\right\} \implies \left\{\tilde{\mathbf{F}}\left[\frac{\mathsf{d}}{\mathsf{dt}}\mathsf{x}(t)\right] = i\omega\left[\tilde{\mathbf{F}}\mathsf{x}\right](\omega)\right\}$$

✎PROOF: Let $\mathbf{L}$ be the *Laplace Transform* operator (Definition S.1 page 389).

$$\begin{aligned}
\tilde{\mathbf{F}}\left[\frac{\mathsf{d}}{\mathsf{dt}}\mathsf{x}(t)\right] &\triangleq \mathbf{L}\left[\frac{\mathsf{d}}{\mathsf{dt}}\mathsf{x}(t)\right](s)\Big|_{s=i\omega} && \text{by definitions of } \mathbf{L} \text{ and } \tilde{\mathbf{F}} && \text{(Definition S.1 page 389)}\\
&= s[\mathbf{L}\mathsf{x}(t)](s)|_{s=i\omega} && \text{by Theorem S.7 page 402}\\
&= i\omega\left[\tilde{\mathbf{F}}\mathsf{x}\right](\omega)
\end{aligned}$$

✏

Theorem T.8. *Let $\tilde{\mathbf{F}}$ be the* FOURIER TRANSFORM *operator (Definition T.2 page 408).*

THM

$$\tilde{\mathbf{F}} \int_{u=-\infty}^{u=t} \mathsf{x}(u)\,\mathsf{d}u = \frac{1}{i\omega}\left[\tilde{\mathbf{F}}\mathsf{x}\right](\omega)$$

Let $\mathbf{L}$ be the *Laplace Transform* operator (Definition S.1 page 389). ✎PROOF:

$$\begin{aligned}
\tilde{\mathbf{F}} \int_{u=-\infty}^{u=t} \mathsf{x}(u)\,\mathsf{d}u &\triangleq \mathbf{L} \int_{u=-\infty}^{u=t} \mathsf{x}(u)\,\mathsf{d}u\Big|_{s=i\omega}\\
&= \frac{1}{s}[\mathbf{L}\mathsf{x}(t)](s)\Big|_{s=i\omega} && \text{by Theorem S.7 page 402}\\
&= \frac{1}{i\omega}\left[\tilde{\mathbf{F}}\mathsf{x}(t)\right](\omega)
\end{aligned}$$

✏

T.7 Real valued functions

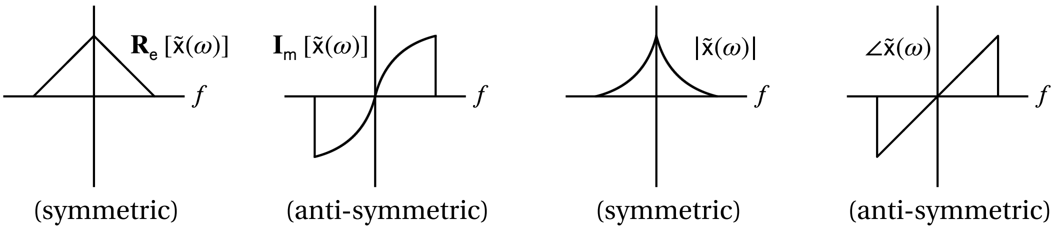


Figure T.1: Fourier transform components of real-valued signal

Theorem T.9. *Let* $\mathsf{f}(x)$ *be a function in* $\boldsymbol{L}^2_{\mathbb{R}}$ *and* $\tilde{\mathsf{f}}(\omega)$ *the* FOURIER TRANSFORM *of* $\mathsf{f}(x)$.

THM

$$\left\{\begin{array}{l}\mathsf{f}(x) \text{ is REAL-VALUED} \\ (\mathsf{f} \in \mathbb{R}^{\mathbb{R}})\end{array}\right\} \implies \left\{\begin{array}{rcll}\tilde{\mathsf{f}}(\omega) &=& \tilde{\mathsf{f}}^*(-\omega) & \text{(HERMITIAN SYMMETRIC)} \\ \mathbf{R_e}\left[\tilde{\mathsf{f}}(\omega)\right] &=& \mathbf{R_e}\left[\tilde{\mathsf{f}}(-\omega)\right] & \text{(SYMMETRIC)} \\ \mathbf{I_m}\left[\tilde{\mathsf{f}}(\omega)\right] &=& -\mathbf{I_m}\left[\tilde{\mathsf{f}}(-\omega)\right] & \text{(ANTI-SYMMETRIC)} \\ |\tilde{\mathsf{f}}(\omega)| &=& |\tilde{\mathsf{f}}(-\omega)| & \text{(SYMMETRIC)} \\ \angle\tilde{\mathsf{f}}(\omega) &=& \angle\tilde{\mathsf{f}}(-\omega) & \text{(ANTI-SYMMETRIC).}\end{array}\right\}$$

✎PROOF:

$$\begin{array}{lclclclcl}\tilde{\mathsf{f}}(\omega) &\triangleq& [\tilde{\mathbf{F}}\mathsf{f}(x)](\omega) &\triangleq& \left\langle \mathsf{f}(x) \,|\, e^{i\omega x}\right\rangle &=& \left\langle \mathsf{f}(x) \,|\, e^{i(-\omega)x}\right\rangle^* &\triangleq& \tilde{\mathsf{f}}^*(-\omega) \\ \mathbf{R_e}\left[\tilde{\mathsf{f}}(\omega)\right] &=& \mathbf{R_e}\left[\tilde{\mathsf{f}}^*(-\omega)\right] &=& \mathbf{R_e}\left[\tilde{\mathsf{f}}(-\omega)\right] \\ \mathbf{I_m}\left[\tilde{\mathsf{f}}(\omega)\right] &=& \mathbf{I_m}\left[\tilde{\mathsf{f}}^*(-\omega)\right] &=& -\mathbf{I_m}\left[\tilde{\mathsf{f}}(-\omega)\right] \\ |\tilde{\mathsf{f}}(\omega)| &=& |\tilde{\mathsf{f}}^*(-\omega)| &=& |\tilde{\mathsf{f}}(-\omega)| \\ \angle\tilde{\mathsf{f}}(\omega) &=& \angle\tilde{\mathsf{f}}^*(-\omega) &=& -\angle\tilde{\mathsf{f}}(-\omega)\end{array}$$

✏

T.8 Moment properties

Definition T.3. [9]

DEF

The quantity M_n *is the* n***th moment*** *of a function* $\mathsf{f}(x) \in \boldsymbol{L}^2_{\mathbb{R}}$ *if*

$$\mathsf{M}_n \triangleq \int_{\mathbb{R}} x^n \mathsf{f}(x)\, \mathrm{d}x \qquad \text{for } n \in \mathbb{W}.$$

Lemma T.1. [10] *Let* M_n *be the* nTH MOMENT *(Definition T.3 page 412) and* $\tilde{\mathsf{f}}(\omega) \triangleq \left[\tilde{\mathbf{F}}\mathsf{f}\right](\omega)$ *the* FOURIER TRANSFORM *(Definition T.2 page 408) of a function* $\mathsf{f}(x)$ *in* $\boldsymbol{L}^2_{\mathbb{R}}$ *(Definition H.2 page 223).*

LEM

$$\mathsf{M}_n \triangleq \int_{\mathbb{R}} x^n \mathsf{f}(x)\, \mathrm{d}x = \sqrt{2\pi}(i)^n \left[\frac{\mathrm{d}}{\mathrm{d}\omega}\right]^n \tilde{\mathsf{f}}(\omega)\bigg|_{\omega=0} \qquad \forall n \in \mathbb{W}, \mathsf{f} \in \boldsymbol{L}^2_{\mathbb{R}}$$

[9] Jawerth and Sweldens (1994) pages 16–17, Sweldens and Piessens (1993) page 2, Vidakovic (1999) page 83
[10] Goswami and Chan (1999) pages 38–39

✎PROOF:

$$
\begin{aligned}
\sqrt{2\pi}(i)^n\left[\left[\tfrac{\mathrm{d}}{\mathrm{d}\omega}\right]^n\tilde{\mathsf{f}}(\omega)\right]_{\omega=0} &= \sqrt{2\pi}(i)^n\left[\left[\tfrac{\mathrm{d}}{\mathrm{d}\omega}\right]^n\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}\mathsf{f}(x)e^{-i\omega x}\,\mathrm{d}x\right]_{\omega=0} && \text{by definition of } \tilde{\mathbf{F}} \quad \text{(Definition T.2 page 408)}\\
&= (i)^n\int_{\mathbb{R}}\mathsf{f}(x)\left[\left[\tfrac{\mathrm{d}}{\mathrm{d}\omega}\right]^n e^{-i\omega x}\right]\mathrm{d}x\Big|_{\omega=0}\\
&= (i)^n\int_{\mathbb{R}}\mathsf{f}(x)\left[(-i)^n x^n e^{-i\omega x}\right]\mathrm{d}x\Big|_{\omega=0}\\
&= (-i^2)^n\int_{\mathbb{R}}\mathsf{f}(x)x^n\,\mathrm{d}x\\
&= \int_{\mathbb{R}}x^n\mathsf{f}(x)\,\mathrm{d}x\\
&\triangleq \mathsf{M}_n && \text{by definition of } \mathsf{M}_n \quad \text{(Definition T.3 page 412)}
\end{aligned}
$$

✏

Lemma T.2. [11] *Let M_n be the* nTH MOMENT *(Definition T.3 page 412) and $\tilde{\mathsf{f}}(\omega) \triangleq [\tilde{\mathbf{F}}\mathsf{f}](\omega)$ the* FOURIER TRANSFORM *(Definition T.2 page 408) of a function $\mathsf{f}(x)$ in $L^2_{\mathbb{R}}$ (Definition H.2 page 223).*

LEM

$$\mathsf{M}_n = 0 \quad\iff\quad \left[\tfrac{\mathrm{d}}{\mathrm{d}\omega}\right]^n\tilde{\mathsf{f}}(\omega)\Big|_{\omega=0} = 0 \qquad \forall n\in\mathbb{W}$$

✎PROOF:

1. Proof for ($\Longrightarrow$) case:

$$
\begin{aligned}
0 &= \langle \mathsf{f}(x) \,|\, x^n\rangle && \text{by left hypothesis}\\
&= \sqrt{2\pi}(-i)^{-n}\left[\tfrac{\mathrm{d}}{\mathrm{d}\omega}\right]^n\tilde{\mathsf{f}}(\omega)\Big|_{\omega=0} && \text{by Lemma T.1 page 412}\\
\Longrightarrow\quad & \left[\tfrac{\mathrm{d}}{\mathrm{d}\omega}\right]^n\tilde{\mathsf{f}}(\omega)\Big|_{\omega=0} = 0
\end{aligned}
$$

2. Proof for ($\Longleftarrow$) case:

$$
\begin{aligned}
0 &= \left[\tfrac{\mathrm{d}}{\mathrm{d}\omega}\right]^n\tilde{\mathsf{f}}(\omega)\Big|_{\omega=0} && \text{by right hypothesis}\\
&= \left[\tfrac{\mathrm{d}}{\mathrm{d}\omega}\right]^n\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}\mathsf{f}(x)e^{-i\omega x}\,\mathrm{d}x\Big|_{\omega=0} && \text{by definition of } \tilde{\mathsf{f}}(\omega)\\
&= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}\mathsf{f}(x)\left[\tfrac{\mathrm{d}}{\mathrm{d}\omega}\right]^n e^{-i\omega x}\,\mathrm{d}x\Big|_{\omega=0}\\
&= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}\mathsf{f}(x)\left[(-i)^n x^n e^{-i\omega x}\right]\mathrm{d}x\Big|_{\omega=0}\\
&= (-i)^n\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}\mathsf{f}(x)x^n\,\mathrm{d}x\\
&= (-i)^n\frac{1}{\sqrt{2\pi}}\langle \mathsf{f}(x) \,|\, x^n\rangle && \text{by definition of } \langle\cdot\,|\,\cdot\rangle \text{ in } L^2_{\mathbb{R}} \quad \text{(Definition H.2 page 223)}
\end{aligned}
$$

✏

[11] Vidakovic (1999) pages 82–83, Mallat (1999) pages 241–242

Lemma T.3 (Strang-Fix condition). [12] *Let* $\mathsf{f}(x)$ *be a function in* $\boldsymbol{L}^2_{\mathbb{R}}$ *and* M_n *the* nTH MOMENT *(Definition T.3 page 412) of* $\mathsf{f}(x)$. *Let* $\mathbf{T}$ *be the* TRANSLATION OPERATOR *(Definition ?? page ??).*

LEM

$$\underbrace{\sum_{k\in\mathbb{Z}} \mathbf{T}^k x^n \mathsf{f}(x) = \mathsf{M}_n}_{\text{STRANG-FIX CONDITION in "time"}} \iff \underbrace{\left[\frac{\mathsf{d}}{\mathsf{d}\omega}\right]^n \tilde{\mathsf{f}}(\omega)\bigg|_{\omega=2\pi k} = \frac{1}{\sqrt{2\pi}}(-i)^n \bar{\delta}_k \mathsf{M}_n}_{\text{STRANG-FIX CONDITION in "frequency"}}$$

✎PROOF:

1. Proof for ($\Longrightarrow$) case:

$$\begin{aligned}
\left[\left[\frac{\mathsf{d}}{\mathsf{d}\omega}\right]^n \tilde{\mathsf{f}}(\omega)\right]_{\omega=2\pi k} &= \sum_{k\in\mathbb{Z}} \left[\left[\frac{\mathsf{d}}{\mathsf{d}\omega}\right]^n \tilde{\mathsf{f}}(\omega)\right]_{\omega=2\pi k} e^{i2\pi kx}\bar{\delta}_k \\
&= \frac{1}{\sqrt{2\pi}} \sum_{k\in\mathbb{Z}} \left[\left[\frac{\mathsf{d}}{\mathsf{d}\omega}\right]^n \int_{\mathbb{R}} \mathsf{f}(x) e^{-i\omega x}\,\mathsf{d}x\right]_{\omega=2\pi k} e^{i2\pi kx}\bar{\delta}_k && \text{by definition of } \tilde{\mathsf{f}}(\omega) \quad \text{(Definition T.2 page 408)} \\
&= \frac{1}{\sqrt{2\pi}} \sum_{k\in\mathbb{Z}} \left[\int_{\mathbb{R}} \mathsf{f}(x)(-ix)^n e^{-i\omega x}\,\mathsf{d}x\right]_{\omega=2\pi k} e^{i2\pi kx}\bar{\delta}_k \\
&= (-i)^n \frac{1}{\sqrt{2\pi}} \sum_{k\in\mathbb{Z}} \left[\int_{\mathbb{R}} x^n \mathsf{f}(x) e^{-i\omega x}\,\mathsf{d}x\right]_{\omega=2\pi k} e^{i2\pi kx}\bar{\delta}_k \\
&= (-i)^n \frac{1}{\sqrt{2\pi}} \sum_{k\in\mathbb{Z}} (x-k)^n \mathsf{f}(x-k)\bar{\delta}_k && \text{by } \textit{PSF} \quad \text{(Theorem ?? page ??)} \\
&= \frac{1}{\sqrt{2\pi}}(-i)^n \bar{\delta}_k \mathsf{M}_n && \text{by left hypothesis}
\end{aligned}$$

2. Proof for ($\Longleftarrow$) case:

$$\begin{aligned}
\frac{1}{\sqrt{2\pi}}(-i)^n \mathsf{M}_n &= \frac{1}{\sqrt{2\pi}} \sum_{k\in\mathbb{Z}} \left[(-i)^n \bar{\delta}_k \mathsf{M}_n\right] e^{-i2\pi kx} && \text{by definition of } \bar{\delta} \quad \text{(Definition N.3 page 323)} \\
&= \sum_{k\in\mathbb{Z}} \left[\left[\frac{\mathsf{d}}{\mathsf{d}\omega}\right]^n \tilde{\mathsf{f}}(\omega)\right]\bigg|_{\omega=2\pi k} e^{-i2\pi kx} && \text{by right hypothesis} \\
&= \sum_{k\in\mathbb{Z}} \left[\left[\frac{\mathsf{d}}{\mathsf{d}\omega}\right]^n \int_{\mathbb{R}} \mathsf{f}(x) e^{-i\omega x}\,\mathsf{d}x\right]\bigg|_{\omega=2\pi k} e^{-i2\pi kx} \\
&= \sum_{k\in\mathbb{Z}} \left[\int_{\mathbb{R}} \mathsf{f}(x)(-ix)^n e^{-i\omega x}\,\mathsf{d}x\right]\bigg|_{\omega=2\pi k} e^{-i2\pi kx} \\
&= (-i)^n \sum_{k\in\mathbb{Z}} \left[\int_{\mathbb{R}} x^n \mathsf{f}(x) e^{-i\omega x}\,\mathsf{d}x\right]\bigg|_{\omega=2\pi k} e^{-i2\pi kx} \\
&= (-i)^n \sum_{k\in\mathbb{Z}} (x-k)^n \mathsf{f}(x-k) && \text{by } \textit{PSF} \quad \text{(Theorem ?? page ??)}
\end{aligned}$$

⇨

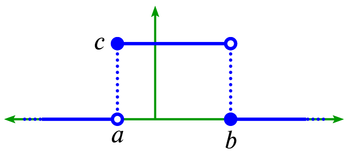
[12] Jawerth and Sweldens (1994) pages 16–17, Sweldens and Piessens (1993) page 2, Vidakovic (1999) page 83, Mallat (1999) pages 241–243, Fix and Strang (1969)

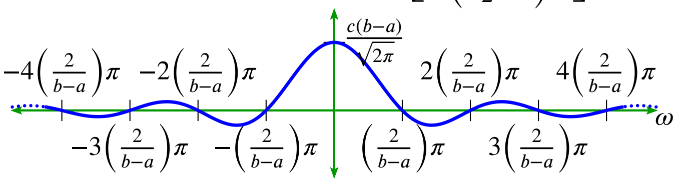
T.9 Examples

Example T.1 (rectangular pulse). Let $\tilde{\mathsf{f}}(\omega)$ be the *Fourier transform* of a function $\mathsf{f}(x) \in \boldsymbol{L}^2_{\mathbb{R}}$.

EX

$$\mathsf{f}(x) = \begin{cases} c & \text{for } x \in [a:b) \\ 0 & \text{otherwise} \end{cases} \qquad \tilde{\mathsf{f}}(\omega) = \frac{c(b-a)}{\sqrt{2\pi}} e^{-i\left(\frac{a+b}{2}\omega\right)} \left[\frac{\sin\left(\frac{b-a}{2}\omega\right)}{\left(\frac{b-a}{2}\omega\right)}\right]$$





✎PROOF:

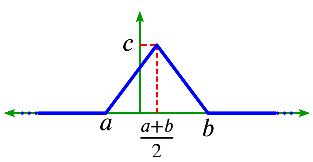
$$\begin{aligned}
\tilde{\mathsf{f}}(\omega) &= \tilde{\mathbf{F}}[\mathsf{f}(x)](\omega) && \text{by definition of } \tilde{\mathsf{f}}(\omega) \\
&= e^{-i\left(\frac{a+b}{2}\right)\omega}\tilde{\mathbf{F}}\left[\mathsf{f}\left(x-\frac{a+b}{2}\right)\right](\omega) && \text{by \textit{shift relation}} \quad \text{(Theorem T.4 page 410)} \\
&= e^{-i\left(\frac{a+b}{2}\right)\omega}\tilde{\mathbf{F}}\left[c\mathbb{1}_{[a:b)}\left(x-\frac{a+b}{2}\right)\right](\omega) && \text{by definition of } \mathsf{f}(x) \\
&= e^{-i\left(\frac{a+b}{2}\right)\omega}\tilde{\mathbf{F}}\left[c\mathbb{1}_{\left[-\frac{b-a}{2}:\frac{b-a}{2}\right)}(x)\right](\omega) && \text{by definition of } \mathbb{1} \quad \text{(Definition W.1 page 443)} \\
&= \frac{1}{\sqrt{2\pi}}e^{-i\left(\frac{a+b}{2}\right)\omega}\int_{\mathbb{R}} c\mathbb{1}_{\left[-\frac{b-a}{2}:\frac{b-a}{2}\right)}(x)e^{-i\omega x}\,\mathrm{d}x && \text{by definition of } \tilde{\mathbf{F}} \quad \text{(Definition T.2 page 408)} \\
&= \frac{1}{\sqrt{2\pi}}e^{-i\left(\frac{a+b}{2}\right)\omega}\int_{-\frac{b-a}{2}}^{\frac{b-a}{2}} ce^{-i\omega x}\,\mathrm{d}x && \text{by definition of } \mathbb{1} \quad \text{(Definition W.1 page 443)} \\
&= \frac{c}{\sqrt{2\pi}}e^{-i\left(\frac{a+b}{2}\right)\omega}\left.\frac{1}{-i\omega}e^{-i\omega x}\right|_{-\frac{b-a}{2}}^{\frac{b-a}{2}} \\
&= \frac{2c}{\sqrt{2\pi}\omega}e^{-i\left(\frac{a+b}{2}\right)\omega}\left[\frac{e^{i\left(\frac{b-a}{2}\omega\right)} - e^{-i\left(\frac{b-a}{2}\omega\right)}}{2i}\right] \\
&= \frac{c(b-a)}{\sqrt{2\pi}}e^{-i\left(\frac{a+b}{2}\omega\right)}\left[\frac{\sin\left(\frac{b-a}{2}\omega\right)}{\left(\frac{b-a}{2}\omega\right)}\right] && \text{by \textit{Euler formulas}} \quad \text{(Corollary I.2 page 235)}
\end{aligned}$$

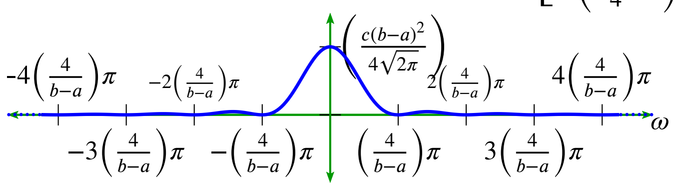
✏

Example T.2 (triangle). Let $\tilde{\mathsf{f}}(\omega)$ be the *Fourier transform* of a function $\mathsf{f}(x) \in \boldsymbol{L}^2_{\mathbb{R}}$.

EX

$$\mathsf{f}(x) = \left\{ \begin{array}{ll} c\left[1 - \frac{|2x-b-a|}{b-a}\right] & \text{for } x \in [a:b) \\ 0 & \text{otherwise} \end{array} \right\} \qquad \tilde{\mathsf{f}}(\omega) = \frac{c(b-a)^2}{4\sqrt{2\pi}} e^{-i\left(\frac{a+b}{2}\omega\right)} \left[\frac{\sin\left(\frac{b-a}{4}\omega\right)}{\left(\frac{b-a}{4}\omega\right)}\right]^2$$





✎PROOF:

$$\tilde{\mathsf{f}}(\omega) = \tilde{\mathbf{F}}[\mathsf{f}(x)](\omega) \qquad \text{by definition of } \tilde{\mathsf{f}}(\omega)$$

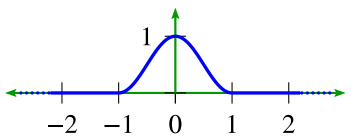
$$
\begin{aligned}
&= e^{-i\left(\frac{a+b}{2}\right)\omega}\tilde{\mathbf{F}}\left[\mathsf{f}\left(x-\frac{a+b}{2}\right)\right](\omega) && \text{by \textit{shift relation}} && \text{(Theorem T.4 page 410)}\\
&= \tilde{\mathbf{F}}\left[c\left(1-\frac{|2x-b-a|}{b-a}\right)\mathbb{1}_{[a:b)}(x)\right](\omega) && \text{by definition of } \mathsf{f}(x)\\
&= c\tilde{\mathbf{F}}\left[\mathbb{1}_{\left[\frac{a}{2}:\frac{b}{2}\right)}(x)\star\mathbb{1}_{\left[\frac{a}{2}:\frac{b}{2}\right)}(x)\right](\omega)\\
&= c\sqrt{2\pi}\tilde{\mathbf{F}}\left[\mathbb{1}_{\left[\frac{a}{2}:\frac{b}{2}\right)}\right]\tilde{\mathbf{F}}\left[\mathbb{1}_{\left[\frac{a}{2}:\frac{b}{2}\right)}\right] && \text{by \textit{convolution theorem}} && \text{(Theorem V.2 page 430)}\\
&= c\sqrt{2\pi}\left(\tilde{\mathbf{F}}\left[\mathbb{1}_{\left[\frac{a}{2}:\frac{b}{2}\right)}\right]\right)^2\\
&= c\sqrt{2\pi}\left(\frac{\left(\frac{b}{2}-\frac{a}{2}\right)}{\sqrt{2\pi}}e^{-i\left(\frac{a+b}{4}\omega\right)}\left[\frac{\sin\left(\frac{b-a}{4}\omega\right)}{\left(\frac{b-a}{4}\omega\right)}\right]\right)^2 && \text{by \textit{Rectangular pulse} ex.} && \text{Example T.1 page 415}\\
&= \frac{c(b-a)^2}{4\sqrt{2\pi}}e^{-i\left(\frac{a+b}{2}\omega\right)}\left[\frac{\sin\left(\frac{b-a}{4}\omega\right)}{\left(\frac{b-a}{4}\omega\right)}\right]^2
\end{aligned}
$$

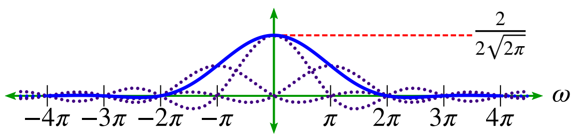
✏

Example T.3. Let a function f be defined in terms of the cosine function (Definition I.1 page 229) as follows:

EX

$$
\mathsf{f}(x) \triangleq \begin{cases} \cos^2\left(\frac{\pi}{2}x\right) & \text{for } |x|\le 1\\ 0 & \text{otherwise}\end{cases}
\qquad
\tilde{\mathsf{f}}(\omega) = \frac{1}{2\sqrt{2\pi}}\left[\underbrace{\frac{2\sin\omega}{\omega}}_{2\,\mathrm{sinc}(\omega)} + \underbrace{\frac{\sin(\omega+\pi)}{(\omega+\pi)}}_{\mathrm{sinc}(\omega+\pi)} + \underbrace{\frac{\sin(\omega-\pi)}{(\omega-\pi)}}_{\mathrm{sinc}(\omega-\pi)}\right]
$$





✎PROOF: Let $\mathbb{1}_A(x)$ be the *set indicator function* (Definition W.1 page 443) on a set A.

$$
\begin{aligned}
\tilde{\mathsf{f}}(\omega) &\triangleq \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}\mathsf{f}(x)e^{-i\omega x}\,\mathrm{d}x && \text{by definition of } \tilde{\mathsf{f}}(\omega) \text{ (Definition T.2)}\\
&= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}\cos^2\left(\frac{\pi}{2}x\right)\mathbb{1}_{[-1:1]}(x)e^{-i\omega x}\,\mathrm{d}x && \text{by definition of } \mathsf{f}(x)\\
&= \frac{1}{\sqrt{2\pi}}\int_{-1}^{1}\cos^2\left(\frac{\pi}{2}x\right)e^{-i\omega x}\,\mathrm{d}x && \text{by definition of } \mathbb{1} \text{ (Definition W.1)}\\
&= \frac{1}{\sqrt{2\pi}}\int_{-1}^{1}\left[\frac{e^{i\frac{\pi}{2}x}+e^{-i\frac{\pi}{2}x}}{2}\right]^2 e^{-i\omega x}\,\mathrm{d}x && \text{by Corollary I.2 page 235}\\
&= \frac{1}{4\sqrt{2\pi}}\int_{-1}^{1}\left[2+e^{i\pi x}+e^{-i\pi x}\right]e^{-i\omega x}\,\mathrm{d}x\\
&= \frac{1}{4\sqrt{2\pi}}\int_{-1}^{1}2e^{-i\omega x}+e^{-i(\omega+\pi)x}+e^{-i(\omega-\pi)x}\,\mathrm{d}x\\
&= \frac{1}{4\sqrt{2\pi}}\left[2\frac{e^{-i\omega x}}{-i\omega}+\frac{e^{-i(\omega+\pi)x}}{-i(\omega+\pi)}+\frac{e^{-i(\omega-\pi)x}}{-i(\omega-\pi)}\right]_{-1}^{1}\\
&= \frac{1}{2\sqrt{2\pi}}\left[2\frac{e^{-i\omega}-e^{+i\omega}}{-2i\omega}+\frac{e^{-i(\omega+\pi)}-e^{+i(\omega+\pi)}}{-2i(\omega+\pi)}+\frac{e^{-i(\omega-\pi)}-e^{+i(\omega-\pi)}}{-2i(\omega-\pi)}\right]_{-1}^{1}
\end{aligned}
$$

$$= \frac{1}{2\sqrt{2\pi}}\left[\underbrace{\frac{2\sin\omega}{\omega}}_{2\,\text{sinc}(\omega)} + \underbrace{\frac{\sin(\omega+\pi)}{(\omega+\pi)}}_{\text{sinc}(\omega+\pi)} + \underbrace{\frac{\sin(\omega-\pi)}{(\omega-\pi)}}_{\text{sinc}(\omega-\pi)}\right]$$

✏

Example T.4. [13]

EX

$$\tilde{\mathbf{F}}\left[e^{-\alpha|x|}\right] = \frac{1}{\sqrt{2\pi}}\left[\frac{2\alpha}{\alpha^2+\omega^2}\right]$$

✎PROOF:

1. Proof using *Laplace Transform*:

$$\begin{aligned}
\sqrt{2\pi}\tilde{\mathbf{F}}\left[e^{-\alpha|x|}\right] &\triangleq \left[\sqrt{2\pi}\right]\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-\alpha|x|}e^{-i\omega x}\,\mathrm{d}x && \text{by definition of } \tilde{\mathbf{F}} && \text{(Definition T.2 page 408)}\\
&= \left[\int_{\mathbb{R}} e^{-\alpha|x|}e^{-sx}\,\mathrm{d}x\right]_{s=i\omega}\\
&= \left[\frac{2\alpha}{\alpha^2-s^2}\right]_{s=i\omega} && \forall \mathbf{R_e}(s) \in (-\alpha : \alpha) && \text{by Corollary S.10 page 399}\\
&= \frac{2\alpha}{\alpha^2+\omega^2} && \text{because } s = i\omega \text{ is in } (-\alpha : \alpha)
\end{aligned}$$

2. Alternate proof:

$$\begin{aligned}
\sqrt{2\pi}\tilde{\mathbf{F}}\left[e^{-\alpha|x|}\right] &\triangleq \left[\sqrt{2\pi}\right]\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\alpha|x|}e^{-i\omega x}\,\mathrm{d}x && \text{by definition of } \tilde{\mathbf{F}} && \text{(Definition T.2 page 408)}\\
&= \int_{-\infty}^{0} e^{-\alpha(-x)}e^{-i\omega x}\,\mathrm{d}x + \int_{0}^{\infty} e^{-\alpha(x)}e^{-i\omega x}\,\mathrm{d}x\\
&= \int_{-\infty}^{0} e^{x(\alpha-i\omega)}\,\mathrm{d}x + \int_{0}^{\infty} e^{x(-\alpha-i\omega)}\,\mathrm{d}x\\
&= \left.\frac{e^{x(\alpha-i\omega)}}{\alpha-i\omega}\right|_{-\infty}^{0} + \left.\frac{e^{x(-\alpha-i\omega)}}{-\alpha-i\omega}\right|_{0}^{\infty} && \text{by Fundamental Theorem of Calculus}\\
&= \left[\frac{1}{\alpha-i\omega}-0\right] + \left[0-\frac{1}{-\alpha-i\omega}\right]\\
&= \left[\frac{1}{\alpha-i\omega}\right]\left[\frac{\alpha-i\omega}{\alpha-i\omega}\right] + \left[\frac{1}{\alpha+i\omega}\right]\left[\frac{\alpha+i\omega}{\alpha+i\omega}\right]\\
&= \frac{\alpha-i\omega}{\alpha^2+\omega^2} + \frac{\alpha+i\omega}{\alpha^2+\omega^2}\\
&= \left[\frac{2\alpha}{\alpha^2+\omega^2}\right]
\end{aligned}$$

✏

[13] https://math.stackexchange.com/questions/4015842/

APPENDIX U

DISCRETE TIME FOURIER TRANSFORM

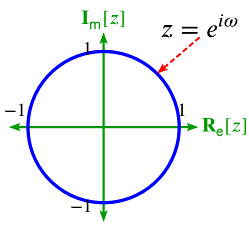
U.1 Definition

Definition U.1.

DEF *The **discrete-time Fourier transform** $\breve{\mathbf{F}}$ of $(\!(x_n)\!)_{n\in\mathbb{Z}}$ is defined as*

$$\left[\breve{\mathbf{F}}\,(\!(x_n)\!)\right](\omega) \triangleq \sum_{n\in\mathbb{Z}} x_n e^{-i\omega n} \qquad \forall (\!(x_n)\!)_{n\in\mathbb{Z}} \in \ell^2_{\mathbb{R}}$$

If we compare the definition of the *Discrete Time Fourier Transform* (Definition U.1 page 419) to the definition of the Z-transform (Definition V.1 page 429), we see that the DTFT is just a special case of the more general Z-Transform, with $z = e^{i\omega}$. If we imagine $z \in \mathbb{C}$ as a complex plane, then $e^{i\omega}$ is a unit circle in this plane. The "frequency" ω in the DTFT is the unit circle in the much larger z-plane, as illustrated to the right.



U.2 Properties

Proposition U.1 (DTFT periodicity)**.** *Let* $\breve{\mathsf{x}}(\omega) \triangleq \breve{\mathbf{F}}\left[(\!(x_n)\!)\right](\omega)$ *be the* DISCRETE-TIME FOURIER TRANSFORM *(Definition U.1 page 419) of a sequence* $(\!(x_n)\!)_{n\in\mathbb{Z}}$ *in* $\ell^2_{\mathbb{R}}$.

PRP
$$\underbrace{\breve{\mathsf{x}}(\omega) = \breve{\mathsf{x}}(\omega + 2\pi n)}_{\text{PERIODIC with period } 2\pi} \qquad \forall n \in \mathbb{Z}$$

✎PROOF:

$$\begin{aligned}
\breve{\mathsf{x}}(\omega + 2\pi n) &= \sum_{m\in\mathbb{Z}} x_m e^{-i(\omega+2\pi n)m} &&= \sum_{m\in\mathbb{Z}} x_m e^{-i\omega m} \cancelto{1}{e^{-i2\pi nm}} \\
&= \sum_{m\in\mathbb{Z}} x_m e^{-i\omega m} &&= \breve{\mathsf{x}}(\omega)
\end{aligned}$$

✏

Theorem U.1. *Let* $\breve{x}(\omega) \triangleq \breve{\mathbf{F}}\left[\left(\mathsf{x}[n]\right)\right](\omega)$ *be the* DISCRETE-TIME FOURIER TRANSFORM *(Definition U.1 page 419) of a sequence* $\left(x_n\right)_{n\in\mathbb{Z}}$ *in* $\boldsymbol{\ell}^2_{\mathbb{R}}$.

THM

$$\left\{ \tilde{\mathsf{x}}(\omega) \triangleq \breve{\mathbf{F}}\left(\mathsf{x}[n]\right) \right\} \implies \left\{ \begin{array}{lrcll} (1). & \breve{\mathbf{F}}\left(\mathsf{x}[-n]\right) &=& \tilde{\mathsf{x}}(-\omega) & \text{and} \\ (2). & \breve{\mathbf{F}}\left(\mathsf{x}^*[n]\right) &=& \tilde{\mathsf{x}}^*(-\omega) & \text{and} \\ (3). & \breve{\mathbf{F}}\left(\mathsf{x}^*[-n]\right) &=& \tilde{\mathsf{x}}^*(\omega) & \end{array} \right\}$$

✎PROOF:

$$\begin{aligned}
\breve{\mathbf{F}}\left(\mathsf{x}[-n]\right) &\triangleq \sum_{n\in\mathbb{Z}} \mathsf{x}[-n]e^{-i\omega n} && \text{by definition of } DTFT && \text{(Definition U.1 page 419)}\\
&= \sum_{m\in\mathbb{Z}} \mathsf{x}[m]e^{i\omega m} && \text{where } m \triangleq -n \implies n = -m\\
&= \sum_{m\in\mathbb{Z}} \mathsf{x}[m]e^{-i(-\omega)m}\\
&\triangleq \tilde{\mathsf{x}}(-\omega) && \text{by left hypothesis}
\end{aligned}$$

$$\begin{aligned}
\breve{\mathbf{F}}\left(\mathsf{x}^*[n]\right) &\triangleq \sum_{n\in\mathbb{Z}} \mathsf{x}^*[n]e^{-i\omega n} && \text{by definition of } DTFT && \text{(Definition U.1 page 419)}\\
&= \left(\sum_{n\in\mathbb{Z}} \mathsf{x}[n]e^{i\omega n}\right)^* && \text{by } \textit{distributive} \text{ property of } *\textbf{-algebras} && \text{(Definition M.3 page 304)}\\
&= \left(\sum_{n\in\mathbb{Z}} \mathsf{x}[n]e^{-i(-\omega)n}\right)^*\\
&\triangleq \tilde{\mathsf{x}}^*(-\omega) && \text{by left hypothesis}
\end{aligned}$$

$$\begin{aligned}
\breve{\mathbf{F}}\left(\mathsf{x}^*[-n]\right) &\triangleq \sum_{n\in\mathbb{Z}} \mathsf{x}^*[-n]e^{-i\omega n} && \text{by definition of } DTFT && \text{(Definition U.1 page 419)}\\
&= \left(\sum_{n\in\mathbb{Z}} \mathsf{x}[-n]e^{i\omega n}\right)^* && \text{by } \textit{distributive} \text{ property of } *\textbf{-algebras} && \text{(Definition M.3 page 304)}\\
&= \left(\sum_{m\in\mathbb{Z}} \mathsf{x}[m]e^{-i\omega m}\right)^* && \text{where } m \triangleq -n \implies n = -m\\
&\triangleq \tilde{\mathsf{x}}^*(\omega) && \text{by left hypothesis}
\end{aligned}$$

⇰

Theorem U.2. *Let* $\breve{x}(\omega) \triangleq \breve{\mathbf{F}}\left[\left(\mathsf{x}[n]\right)\right](\omega)$ *be the* DISCRETE-TIME FOURIER TRANSFORM *(Definition U.1 page 419) of a sequence* $\left(\mathsf{x}[n]\right)_{n\in\mathbb{Z}}$ *in* $\boldsymbol{\ell}^2_{\mathbb{R}}$.

THM

$$\left\{ \begin{array}{lll} (1). & \tilde{\mathsf{x}}(\omega) \triangleq \breve{\mathbf{F}}\left(\mathsf{x}[n]\right) & \text{and} \\ (2). & \left(\mathsf{x}[n]\right) \textit{ is } \text{REAL-VALUED} & \end{array} \right\} \implies \left\{ \begin{array}{lrcllll} (1). & \breve{\mathbf{F}}\left(\mathsf{x}[-n]\right) &=& \tilde{\mathsf{x}}(-\omega) & & & \text{and} \\ (2). & \breve{\mathbf{F}}\left(\mathsf{x}^*[n]\right) &=& \tilde{\mathsf{x}}^*(-\omega) &=& \tilde{\mathsf{x}}(\omega) & \text{and} \\ (3). & \breve{\mathbf{F}}\left(\mathsf{x}^*[-n]\right) &=& \tilde{\mathsf{x}}^*(\omega) &=& \tilde{\mathsf{x}}(-\omega) & \end{array} \right\}$$

✎PROOF:

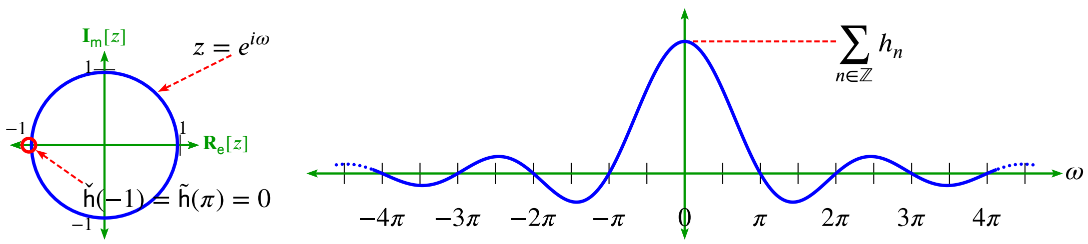
$$\begin{aligned}
\breve{\mathbf{F}}\left(\mathsf{x}[-n]\right) &\triangleq \sum_{n\in\mathbb{Z}} \mathsf{x}[-n]e^{-i\omega n} && \text{by definition of } DTFT && \text{(Definition U.1 page 419)}\\
&= \sum_{m\in\mathbb{Z}} \mathsf{x}[m]e^{i\omega m} && \text{where } m \triangleq -n \implies n = -m\\
&= \sum_{m\in\mathbb{Z}} \mathsf{x}[m]e^{-i(-\omega)m}
\end{aligned}$$

$$\triangleq \breve{x}(-\omega) \qquad \text{by left hypothesis}$$

$$\begin{aligned}
\boxed{\tilde{x}^*(-\omega)} &= \boxed{\breve{\mathbf{F}}\,(\!(x^*[n])\!)} && \text{by Theorem U.1 page 420}\\
&= \breve{\mathbf{F}}\,(\!(x[n])\!) && \text{by } \textit{real-valued} \text{ hypothesis}\\
&= \boxed{\tilde{x}(\omega)} && \text{by definition of } \tilde{x}(\omega) && \text{(Definition U.1 page 419)}
\end{aligned}$$

$$\begin{aligned}
\boxed{\tilde{x}^*(\omega)} &= \boxed{\breve{\mathbf{F}}\,(\!(x^*[-n])\!)} && \text{by Theorem U.1 page 420}\\
&= \breve{\mathbf{F}}\,(\!(x[-n])\!) && \text{by } \textit{real-valued} \text{ hypothesis}\\
&= \boxed{\tilde{x}(-\omega)} && \text{by result (1)}
\end{aligned}$$

✏



Proposition U.2. *Let* $\check{x}(z)$ *be the* Z-TRANSFORM *(Definition V.1 page 429) and* $\check{x}(\omega)$ *the* DISCRETE-TIME FOURIER TRANSFORM *(Definition U.1 page 419) of* $(\!(x_n)\!)$.

PRP

$$\underbrace{\left\{\sum_{n\in\mathbb{Z}} x_n = c\right\}}_{\text{(1) time domain}} \iff \underbrace{\left\{\check{x}(z)\Big|_{z=1} = c\right\}}_{\text{(2) z domain}} \iff \underbrace{\left\{\check{x}(\omega)\Big|_{\omega=0} = c\right\}}_{\text{(3) frequency domain}} \qquad \forall (\!(x_n)\!)_{n\in\mathbb{Z}}\in\ell^2_{\mathbb{R}}, c\in\mathbb{R}$$

✎PROOF:

1. Proof that (1) $\implies$ (2):

$$\begin{aligned}
\check{x}(z)\Big|_{z=1} &= \sum_{n\in\mathbb{Z}} x_n z^{-n}\Big|_{z=1} && \text{by definition of } \check{x}(z) \text{ (Definition V.1 page 429)}\\
&= \sum_{n\in\mathbb{Z}} x_n && \text{because } z^n = 1 \text{ for all } n\in\mathbb{Z}\\
&= c && \text{by hypothesis (1)}
\end{aligned}$$

2. Proof that (2) $\implies$ (3):

$$\begin{aligned}
\check{x}(\omega)\Big|_{\omega=0} &= \sum_{n\in\mathbb{Z}} x_n e^{-i\omega n}\Big|_{\omega=0} && \text{by definition of } \check{x}(\omega) && \text{(Definition U.1 page 419)}\\
&= \sum_{n\in\mathbb{Z}} x_n z^{-n}\Big|_{z=1}\\
&= \check{x}(z)\Big|_{z=1} && \text{by definition of } \check{x}(z) && \text{(Definition V.1 page 429)}\\
&= c && \text{by hypothesis (2)}
\end{aligned}$$

3. Proof that (3) $\implies$ (1):

$$
\begin{aligned}
\sum_{n\in\mathbb{Z}} x_n &= \left.\sum_{n\in\mathbb{Z}} x_n e^{-i\omega n}\right|_{\omega=0} \\
&= \check{\mathsf{x}}(\omega) && \text{by definition of } \check{\mathsf{x}}(\omega) && \text{(Definition U.1 page 419)} \\
&= c && \text{by hypothesis (3)}
\end{aligned}
$$

✎

Proposition U.3. *If the coefficients are **real**, then the magnitude response (MR) is **symmetric**.*

✎PROOF:

$$
\begin{aligned}
\left|\tilde{\mathsf{h}}(-\omega)\right| &\triangleq \left|\check{\mathsf{h}}(z)\right|_{z=e^{-i\omega}} &&\triangleq \left|\sum_{m\in\mathbb{Z}} \mathsf{x}[m] z^{-m}\right|_{z=e^{-i\omega}} \\
&= \left|\sum_{m\in\mathbb{Z}} \mathsf{x}[m] e^{i\omega m}\right| &&= \left|\left(\sum_{m\in\mathbb{Z}} \mathsf{x}^*[m] e^{-i\omega m}\right)^*\right| \\
&= \underbrace{\left|\left(\sum_{m\in\mathbb{Z}} \mathsf{x}[m] e^{-i\omega m}\right)^*\right|}_{\text{if } x[m] \text{ is } real} &&= \left|\sum_{m\in\mathbb{Z}} \mathsf{x}[m] e^{-i\omega m}\right| \\
&\triangleq \left|\check{\mathsf{h}}(z)\right|_{z=e^{-i\omega}} &&\triangleq \left|\tilde{\mathsf{h}}(\omega)\right|
\end{aligned}
$$

✎

Proposition U.4. [1]

PRP

$$
\underbrace{\sum_{n\in\mathbb{Z}} (-1)^n x_n = c}_{\text{(1) in "time"}} \iff \underbrace{\check{\mathsf{x}}(z)|_{z=-1} = c}_{\text{(2) in "z domain"}} \iff \underbrace{\check{\mathsf{x}}(\omega)|_{\omega=\pi} = c}_{\text{(3) in "frequency"}}
$$

$$
\iff \underbrace{\left(\sum_{n\in\mathbb{Z}} h_{2n}, \sum_{n\in\mathbb{Z}} h_{2n+1}\right) = \left(\frac{1}{2}\left(\sum_{n\in\mathbb{Z}} h_n + c\right), \frac{1}{2}\left(\sum_{n\in\mathbb{Z}} h_n - c\right)\right)}_{\text{(4) sum of even, sum of odd}}
$$

$\forall c\in\mathbb{R}, (x_n)_{n\in\mathbb{Z}}, (y_n)_{n\in\mathbb{Z}} \in \ell^2_{\mathbb{R}}$

✎PROOF:

1. Proof that (1) $\implies$ (2):

$$
\begin{aligned}
\check{\mathsf{x}}(z)|_{z=-1} &= \left.\sum_{n\in\mathbb{Z}} x_n z^{-n}\right|_{z=-1} \\
&= \sum_{n\in\mathbb{Z}} (-1)^n x_n \\
&= c && \text{by (1)}
\end{aligned}
$$

[1] Chui (1992) page 123

2. Proof that (2) $\implies$ (3):

$$
\begin{aligned}
\left.\sum_{n\in\mathbb{Z}} x_n e^{-i\omega n}\right|_{\omega=\pi} &= \sum_{n\in\mathbb{Z}}(-1)^n x_n \\
&= \sum_{n\in\mathbb{Z}}(-1)^{-n} x_n && = \left.\sum_{n\in\mathbb{Z}} z^{-n} x_n\right|_{z=-1} \\
&= c && \text{by (2)}
\end{aligned}
$$

3. Proof that (3) $\implies$ (1):

$$
\begin{aligned}
\sum_{n\in\mathbb{Z}}(-1)^n x_n &= \sum_{n\in\mathbb{Z}}(-1)^{-n} x_n \\
&= \left.\sum_{n\in\mathbb{Z}} e^{-i\omega n} x_n\right|_{\omega=\pi} \\
&= c && \text{by (3)}
\end{aligned}
$$

4. Proof that (2) $\implies$ (4):

   (a) Define $A \triangleq \sum_{n\in\mathbb{Z}} h_{2n}$ $\qquad B \triangleq \sum_{n\in\mathbb{Z}} h_{2n+1}$.

   (b) Proof that $A - B = c$:

$$
\begin{aligned}
c &= \sum_{n\in\mathbb{Z}}(-1)^n x_n && \text{by (2)} \\
&= \underbrace{\sum_{n\in\mathbb{Z}_e}(-1)^n x_n}_{\text{even terms}} + \underbrace{\sum_{n\in\mathbb{Z}_o}(-1)^n x_n}_{\text{odd terms}} \\
&= \sum_{n\in\mathbb{Z}}(-1)^{2n} x_{2n} + \sum_{n\in\mathbb{Z}}(-1)^{2n+1} x_{2n+1} \\
&= \underbrace{\sum_{n\in\mathbb{Z}} x_{2n}}_{A} - \underbrace{\sum_{n\in\mathbb{Z}} x_{2n+1}}_{B} \\
&\triangleq A - B && \text{by definitions of } A \text{ and } B
\end{aligned}
$$

   (c) Proof that $A + B = \sum_{n\in\mathbb{Z}} x_n$:

$$
\begin{aligned}
\sum_{n\in\mathbb{Z}} x_n &= \sum_{n \text{ even}} x_n + \sum_{n \text{ odd}} x_n \\
&= \underbrace{\sum_{n\in\mathbb{Z}} x_{2n}}_{A} + \underbrace{\sum_{n\in\mathbb{Z}} x_{2n+1}}_{B} \\
&= A + B && \text{by definitions of } A \text{ and } B
\end{aligned}
$$

   (d) This gives two simultaneous equations:

$$
\begin{aligned}
A - B &= c \\
A + B &= \sum_{n\in\mathbb{Z}} x_n
\end{aligned}
$$

(e) Solutions to these equations give

$$\begin{aligned}
\sum_{n\in\mathbb{Z}} x_{2n} \triangleq A &= \frac{1}{2}\left(\sum_{n\in\mathbb{Z}} x_n + c\right)\\
\sum_{n\in\mathbb{Z}} x_{2n+1} \triangleq B &= \frac{1}{2}\left(\sum_{n\in\mathbb{Z}} x_n - c\right)
\end{aligned}$$

5. Proof that (2) $\Longleftarrow$ (4):

$$\begin{aligned}
\sum_{n\in\mathbb{Z}}(-1)^n x_n &= \underbrace{\sum_{n\in\mathbb{Z}_e}(-1)^n x_n}_{\text{even terms}} + \underbrace{\sum_{n\in\mathbb{Z}_o}(-1)^n x_n}_{\text{odd terms}}\\
&= \sum_{n\in\mathbb{Z}}(-1)^{2n} x_{2n} + \sum_{n\in\mathbb{Z}}(-1)^{2n+1} x_{2n+1}\\
&= \sum_{n\in\mathbb{Z}} x_{2n} - \sum_{n\in\mathbb{Z}} x_{2n+1}\\
&= \frac{1}{2}\left(\sum_{n\in\mathbb{Z}} x_n + c\right) - \frac{1}{2}\left(\sum_{n\in\mathbb{Z}} x_n - c\right) && \text{by (3)}\\
&= c
\end{aligned}$$

⇰

Lemma U.1. *Let* $\tilde{\mathsf{f}}(\omega)$ *be the* DTFT (Definition U.1 page 419) *of a sequence* $(x_n)_{n\in\mathbb{Z}}$.

LEM

$$\underbrace{(x_n\in\mathbb{R})_{n\in\mathbb{Z}}}_{\text{REAL-VALUED sequence}} \implies \underbrace{|\breve{\mathsf{x}}(\omega)|^2 = |\breve{\mathsf{x}}(-\omega)|^2}_{\text{EVEN}} \qquad \forall (x_n)_{n\in\mathbb{Z}}\in\ell^2_{\mathbb{R}}$$

✎PROOF:

$$\begin{aligned}
|\breve{\mathsf{x}}(\omega)|^2 &= |\breve{\mathsf{x}}(z)|^2\Big|_{z=e^{i\omega}}\\
&= \breve{\mathsf{x}}(z)\breve{\mathsf{x}}^*(z)\Big|_{z=e^{i\omega}}\\
&= \left[\sum_{n\in\mathbb{Z}} x_n z^{-n}\right]\left[\sum_{m\in\mathbb{Z}} x_m z^{-n}\right]^*\Bigg|_{z=e^{i\omega}}\\
&= \left[\sum_{n\in\mathbb{Z}} x_n z^{-n}\right]\left[\sum_{m\in\mathbb{Z}} x_m^*(z^*)^{-m}\right]\Bigg|_{z=e^{i\omega}}\\
&= \sum_{n\in\mathbb{Z}}\sum_{m\in\mathbb{Z}} x_n x_m^* z^{-n}(z^*)^{-m}\Bigg|_{z=e^{i\omega}}\\
&= \sum_{n\in\mathbb{Z}}\left[|x_n|^2 + \sum_{m>n} x_n x_m^* z^{-n}(z^*)^{-m} + \sum_{m<n} x_n x_m^* z^{-n}(z^*)^{-m}\right]\Bigg|_{z=e^{i\omega}}\\
&= \sum_{n\in\mathbb{Z}}\left[|x_n|^2 + \sum_{m>n} x_n x_m e^{i\omega(m-n)} + \sum_{m<n} x_n x_m e^{i\omega(m-n)}\right]\\
&= \sum_{n\in\mathbb{Z}}\left[|x_n|^2 + \sum_{m>n} x_n x_m e^{i\omega(m-n)} + \sum_{m>n} x_n x_m e^{-i\omega(m-n)}\right]\\
&= \sum_{n\in\mathbb{Z}}\left[|x_n|^2 + \sum_{m>n} x_n x_m \left(e^{i\omega(m-n)} + e^{-i\omega(m-n)}\right)\right]
\end{aligned}$$

$$\begin{aligned}
&= \sum_{n\in\mathbb{Z}} \left[|x_n|^2 + \sum_{m>n} x_n x_m 2\cos[\omega(m-n)] \right] \\
&= \sum_{n\in\mathbb{Z}} |x_n|^2 + 2\sum_{n\in\mathbb{Z}}\sum_{m>n} x_n x_m \cos[\omega(m-n)]
\end{aligned}$$

Since cos is real and even, then $|\breve{x}(\omega)|^2$ must also be real and even. ⇨

Theorem U.3 (inverse DTFT). [2] *Let $\breve{x}(\omega)$ be the* DISCRETE-TIME FOURIER TRANSFORM *(Definition U.1 page 419) of a sequence $(x_n)_{n\in\mathbb{Z}} \in \ell^2_{\mathbb{R}}$. Let $\breve{x}^{-1}$ be the inverse of $\breve{x}$.*

THM

$$\underbrace{\left\{ \breve{x}(\omega) \triangleq \sum_{n\in\mathbb{Z}} x_n e^{-i\omega n} \right\}}_{\breve{x}(\omega)\triangleq\breve{\mathbf{F}}(x_n)} \implies \underbrace{\left\{ x_n = \frac{1}{2\pi}\int_{\alpha-\pi}^{\alpha+\pi} \breve{x}(\omega)e^{i\omega n}\,d\omega \quad \forall\alpha\in\mathbb{R} \right\}}_{(x_n)=\breve{\mathbf{F}}^{-1}\breve{\mathbf{F}}(x_n)} \qquad \forall (x_n)_{n\in\mathbb{Z}}\in\ell^2_{\mathbb{R}}$$

✎PROOF:

$$\begin{aligned}
\frac{1}{2\pi}\int_{\alpha-\pi}^{\alpha+\pi} \breve{x}(\omega)e^{i\omega n}\,d\omega &= \frac{1}{2\pi}\int_{\alpha-\pi}^{\alpha+\pi} \underbrace{\left[\sum_{m\in\mathbb{Z}} x_m e^{-i\omega m}\right]}_{\breve{x}(\omega)} e^{i\omega n}\,d\omega && \text{by definition of } \breve{x}(\omega) \\
&= \frac{1}{2\pi}\int_{\alpha-\pi}^{\alpha+\pi} \sum_{m\in\mathbb{Z}} x_m e^{-i\omega(m-n)}\,d\omega \\
&= \frac{1}{2\pi}\sum_{m\in\mathbb{Z}} x_m \int_{\alpha-\pi}^{\alpha+\pi} e^{-i\omega(m-n)}\,d\omega \\
&= \frac{1}{2\pi}\sum_{m\in\mathbb{Z}} x_m \left[2\pi\bar{\delta}_{m-n}\right] \\
&= x_n
\end{aligned}$$

⇨

Theorem U.4 (orthonormal quadrature conditions). [3] *Let $\breve{x}(\omega)$ be the* DISCRETE-TIME FOURIER TRANSFORM *(Definition U.1 page 419) of a sequence $(x_n)_{n\in\mathbb{Z}} \in \ell^2_{\mathbb{R}}$. Let $\bar{\delta}_n$ be the* KRONECKER DELTA FUNCTION *at n (Definition N.3 page 323).*

THM

$$\begin{array}{rclclcll}
\sum_{m\in\mathbb{Z}} x_m y^*_{m-2n} &=& 0 &\iff& \breve{x}(\omega)\breve{y}^*(\omega) + \breve{x}(\omega+\pi)\breve{y}^*(\omega+\pi) &=& 0 & \forall n\in\mathbb{Z}, \forall (x_n),(y_n)\in\ell^2_{\mathbb{R}} \\
\sum_{m\in\mathbb{Z}} x_m x^*_{m-2n} &=& \bar{\delta}_n &\iff& |\breve{x}(\omega)|^2 + |\breve{x}(\omega+\pi)|^2 &=& 2 & \forall n\in\mathbb{Z}, \forall (x_n),(y_n)\in\ell^2_{\mathbb{R}}
\end{array}$$

✎PROOF: Let $z \triangleq e^{i\omega}$.

[2] J.S.Chitode (2009a) page 3-95 ⟨(3.6.2)⟩
[3] Daubechies (1992) pages 132–137 ⟨(5.1.20),(5.1.39)⟩

1. Proof that $2\sum_{n\in\mathbb{Z}}\left[\sum_{k\in\mathbb{Z}} x_k y^*_{k-2n}\right] e^{-i2\omega n} = \breve{x}(\omega)\breve{y}^*(\omega) + \breve{x}(\omega+\pi)\breve{y}^*(\omega+\pi)$:

$$
\begin{aligned}
&2\sum_{n\in\mathbb{Z}}\left[\sum_{k\in\mathbb{Z}} x_k y^*_{k-2n}\right] e^{-i2\omega n} \\
&= 2\sum_{k\in\mathbb{Z}} x_k \sum_{n\in\mathbb{Z}} y^*_{k-2n} z^{-2n} \\
&= 2\sum_{k\in\mathbb{Z}} x_k \sum_{n \text{ even}} y^*_{k-n} z^{-n} \\
&= \sum_{k\in\mathbb{Z}} x_k \sum_{n\in\mathbb{Z}} y^*_{k-n} z^{-n}\left(1+e^{i\pi n}\right) \\
&= \sum_{k\in\mathbb{Z}} x_k \sum_{n\in\mathbb{Z}} y^*_{k-n} z^{-n} + \sum_{k\in\mathbb{Z}} x_k \sum_{n\in\mathbb{Z}} y^*_{k-n} z^{-n} e^{i\pi n} \\
&= \sum_{k\in\mathbb{Z}} x_k \sum_{m\in\mathbb{Z}} y^*_{m} z^{-(k-m)} + \sum_{k\in\mathbb{Z}} x_k \sum_{m\in\mathbb{Z}} y^*_{m} e^{-i(\omega+\pi)(k-m)} && \text{where } m \triangleq k-n \\
&= \sum_{k\in\mathbb{Z}} x_k z^{-k} \sum_{m\in\mathbb{Z}} y^*_{m} z^{m} + \sum_{k\in\mathbb{Z}} x_k e^{-i(\omega+\pi)k} \sum_{m\in\mathbb{Z}} y^*_{m} e^{+i(\omega+\pi)m} \\
&= \sum_{k\in\mathbb{Z}} x_k e^{-i\omega k}\left[\sum_{m\in\mathbb{Z}} y_m e^{-i\omega m}\right]^* + \sum_{k\in\mathbb{Z}} x_k e^{-i(\omega+\pi)k}\left[\sum_{m\in\mathbb{Z}} y_m e^{-i(\omega+\pi)m}\right]^* \\
&\triangleq \breve{x}(\omega)\breve{y}^*(\omega) + \breve{x}(\omega+\pi)\breve{y}^*(\omega+\pi)
\end{aligned}
$$

2. Proof that $\sum_{m\in\mathbb{Z}} x_m y^*_{m-2n} = 0 \implies \breve{x}(\omega)\breve{y}^*(\omega) + \breve{x}(\omega+\pi)\breve{y}^*(\omega+\pi) = 0$:

$$
\begin{aligned}
0 &= 2\sum_{n\in\mathbb{Z}}\left[\sum_{k\in\mathbb{Z}} x_k y^*_{k-2n}\right] e^{-i2\omega n} && \text{by left hypothesis} \\
&= \breve{x}(\omega)\breve{y}^*(\omega) + \breve{x}(\omega+\pi)\breve{y}^*(\omega+\pi) && \text{by item (1)}
\end{aligned}
$$

3. Proof that $\sum_{m\in\mathbb{Z}} x_m y^*_{m-2n} = 0 \impliedby \breve{x}(\omega)\breve{y}^*(\omega) + \breve{x}(\omega+\pi)\breve{y}^*(\omega+\pi) = 0$:

$$
\begin{aligned}
2\sum_{n\in\mathbb{Z}}\left[\sum_{k\in\mathbb{Z}} x_k y^*_{k-2n}\right] e^{-i2\omega n} &= \breve{x}(\omega)\breve{y}^*(\omega) + \breve{x}(\omega+\pi)\breve{y}^*(\omega+\pi) && \text{by item (1)} \\
&= 0 && \text{by right hypothesis}
\end{aligned}
$$

Thus by the above equation, $\sum_{n\in\mathbb{Z}}\left[\sum_{k\in\mathbb{Z}} x_k y^*_{k-2n}\right] e^{-i2\omega n} = 0$. The only way for this to be true is if $\sum_{k\in\mathbb{Z}} x_k y^*_{k-2n} = 0$.

4. Proof that $\sum_{m\in\mathbb{Z}} x_m x^*_{m-2n} = \bar{\delta}_n \implies |\breve{x}(\omega)|^2 + |\breve{x}(\omega'+\pi)|^2 = 2$:
Let $g_n \triangleq x_n$.

$$
\begin{aligned}
2 &= 2\sum_{n\in\mathbb{Z}} \bar{\delta}_{n\in\mathbb{Z}} e^{-i2\omega n} \\
&= 2\sum_{n\in\mathbb{Z}}\left[\sum_{k\in\mathbb{Z}} x_k y^*_{k-2n}\right] e^{-i2\omega n} && \text{by left hypothesis} \\
&= \breve{x}(\omega)\breve{y}^*(\omega) + \breve{x}(\omega+\pi)\breve{y}^*(\omega+\pi) && \text{by item (1)}
\end{aligned}
$$

5. Proof that $\sum_{m\in\mathbb{Z}} x_m x^*_{m-2n} = \bar{\delta}_n \impliedby |\breve{x}(\omega)|^2 + |\breve{x}(\omega'+\pi)|^2 = 2$:
Let $g_n \triangleq x_n$.

$$
\begin{aligned}
2\sum_{n\in\mathbb{Z}}\left[\sum_{k\in\mathbb{Z}} x_k y^*_{k-2n}\right] e^{-i2\omega n} &= \breve{x}(\omega)\breve{y}^*(\omega) + \breve{x}(\omega+\pi)\breve{y}^*(\omega+\pi) && \text{by item (1)} \\
&= 2 && \text{by right hypothesis}
\end{aligned}
$$

Thus by the above equation, $\sum_{n\in\mathbb{Z}}\left[\sum_{k\in\mathbb{Z}} x_k y^*_{k-2n}\right] e^{-i2\omega n} = 1$. The only way for this to be true is if $\sum_{k\in\mathbb{Z}} x_k y^*_{k-2n} = \bar{\delta}_n$.

⇨

U.3 Derivatives

Theorem U.5. [4] *Let* $\breve{x}(\omega)$ *be the* DTFT (Definition U.1 page 419) *of a sequence* $(\!(x_n)\!)_{n\in\mathbb{Z}}$.

THM

(A)	$\left[\frac{\mathrm{d}}{\mathrm{d}\omega}\right]^n \breve{x}(\omega)\Big	_{\omega=0} = 0 \iff \sum_{k\in\mathbb{Z}} k^n x_k = 0$	(B)	$\forall n\in\mathbb{W}$
(C)	$\left[\frac{\mathrm{d}}{\mathrm{d}\omega}\right]^n \breve{x}(\omega)\Big	_{\omega=\pi} = 0 \iff \sum_{k\in\mathbb{Z}} (-1)^k k^n x_k = 0$	(D)	$\forall n\in\mathbb{W}$

✎PROOF:

1. Proof that $(A) \implies (B)$:

$$\begin{aligned}
0 &= \left[\frac{\mathrm{d}}{\mathrm{d}\omega}\right]^n \breve{x}(\omega)\Big|_{\omega=0} && \text{by hypothesis (A)}\\
&= \left[\frac{\mathrm{d}}{\mathrm{d}\omega}\right]^n \sum_{k\in\mathbb{Z}} x_k e^{-i\omega k}\Bigg|_{\omega=0} && \text{by definition of } \breve{x}(\omega) \text{ (Definition U.1 page 419)}\\
&= \sum_{k\in\mathbb{Z}} x_k \left[\frac{\mathrm{d}}{\mathrm{d}\omega}\right]^n e^{-i\omega k}\Bigg|_{\omega=0}\\
&= \sum_{k\in\mathbb{Z}} x_k \left[(-i)^n k^n e^{-i\omega k}\right]\Bigg|_{\omega=0}\\
&= (-i)^n \sum_{k\in\mathbb{Z}} k^n x_k
\end{aligned}$$

2. Proof that $(A) \impliedby (B)$:

$$\begin{aligned}
\left[\frac{\mathrm{d}}{\mathrm{d}\omega}\right]^n \breve{x}(\omega)\Big|_{\omega=0} &= \left[\frac{\mathrm{d}}{\mathrm{d}\omega}\right]^n \sum_{k\in\mathbb{Z}} x_k e^{-i\omega k}\Bigg|_{\omega=0} && \text{by definition of } \tilde{g}\\
&= \sum_{k\in\mathbb{Z}} x_k \left[\left[\frac{\mathrm{d}}{\mathrm{d}\omega}\right]^n e^{-i\omega k}\right]\Bigg|_{\omega=0}\\
&= \sum_{k\in\mathbb{Z}} x_k \left[(-i)^n k^n e^{-i\omega k}\right]\Bigg|_{\omega=0}\\
&= (-i)^n \sum_{k\in\mathbb{Z}} k^n x_k\\
&= 0 && \text{by hypothesis (B)}
\end{aligned}$$

[4] Vidakovic (1999) pages 82–83, Mallat (1999) pages 241–242

3. Proof that $(C) \implies (D)$:

$$\begin{aligned}
0 &= \left[\frac{\mathrm{d}}{\mathrm{d}\omega}\right]^n \breve{x}(\omega)\Big|_{\omega=\pi} && \text{by hypothesis (C)} \\
&= \left[\frac{\mathrm{d}}{\mathrm{d}\omega}\right]^n \sum_{k\in\mathbb{Z}} x_k e^{-i\omega k}\Bigg|_{\omega=\pi} && \text{by definition of } \breve{x} \text{ (Definition U.1 page 419)} \\
&= \sum_{k\in\mathbb{Z}} x_k \left[\frac{\mathrm{d}}{\mathrm{d}\omega}\right]^n e^{-i\omega k}\Bigg|_{\omega=\pi} \\
&= \sum_{k\in\mathbb{Z}} x_k \left[(-i)^n k^n e^{-i\omega k}\right]\Bigg|_{\omega=\pi} \\
&= \sum_{k\in\mathbb{Z}} x_k \left[(-i)^n k^n (-1)^k\right] \\
&= (-i)^n \sum_{k\in\mathbb{Z}} (-1)^k k^n x_k
\end{aligned}$$

4. Proof that $(C) \impliedby (D)$:

$$\begin{aligned}
\left[\frac{\mathrm{d}}{\mathrm{d}\omega}\right]^n \breve{x}(\omega)\Big|_{\omega=\pi} &= \left[\frac{\mathrm{d}}{\mathrm{d}\omega}\right]^n \sum_{k\in\mathbb{Z}} x_k e^{-i\omega k}\Bigg|_{\omega=\pi} && \text{by definition of } \breve{x} \text{ (Definition U.1 page 419)} \\
&= \sum_{k\in\mathbb{Z}} x_k \left[\frac{\mathrm{d}}{\mathrm{d}\omega}\right]^n e^{-i\omega k}\Bigg|_{\omega=\pi} \\
&= \sum_{k\in\mathbb{Z}} x_k \left[(-i)^n k^n e^{-i\omega k}\right]\Bigg|_{\omega=\pi} \\
&= \sum_{k\in\mathbb{Z}} x_k \left[(-i)^n k^n (-1)^k\right] \\
&= (-i)^n \sum_{k\in\mathbb{Z}} (-1)^k k^n x_k \\
&= 0 && \text{by hypothesis (D)}
\end{aligned}$$

✏

APPENDIX V

Z TRANSFORM

A digital filter is an operator on a sequence For a wide class of digital filters, this operator is linear. This operation can often be more clearly understood by the use of a special transform called the *Z-transform* (Definition V.1 page 429). The Z-transform represents linear filters by ratios of polynomials (a polynomial divided by a polynomial) in a free variable z. The roots of the numerator polynomial are called *zeros*; the roots of the denominator polynomial are called *poles* (Definition ?? page ??). The location in the z-plane of these poles and zeros determine the behavior of the filter operation.

V.1 Z-transform

Definition V.1. [1]

DEF *The **z-transform** $\mathbf{Z}$ of $(\!(x_n)\!)_{n\in\mathbb{Z}}$ is defined as*

$$\left[\mathbf{Z}\,(\!(x_n)\!)\right](z) \triangleq \underbrace{\sum_{n\in\mathbb{Z}} x_n z^{-n}}_{\textit{Laurent series}} \qquad \forall (\!(x_n)\!)\in\ell^2_{\mathbb{R}}$$

Theorem V.1. *Let* $X(z) \triangleq \mathbf{Z}\mathsf{x}[n]$ *be the* Z-TRANSFORM *of* $\mathsf{x}[n]$.

THM

$$\left\{\ \check{\mathsf{x}}(z)\ \triangleq\ \mathbf{Z}\,(\!(\mathsf{x}[n])\!)\ \right\} \implies \left\{\begin{array}{lrcll} \textit{(1).} & \mathbf{Z}\,(\!(\alpha\mathsf{x}[n])\!) & = & \alpha\check{\mathsf{x}}(z) & \forall (\!(x_n)\!)\in\ell^2_{\mathbb{R}} \quad \textit{and} \\ \textit{(2).} & \mathbf{Z}\,(\!(\mathsf{x}[n-k])\!) & = & z^{-k}\check{\mathsf{x}}(z) & \forall (\!(x_n)\!)\in\ell^2_{\mathbb{R}} \quad \textit{and} \\ \textit{(3).} & \mathbf{Z}\,(\!(\mathsf{x}[-n])\!) & = & \check{\mathsf{x}}\left(\frac{1}{z}\right) & \forall (\!(x_n)\!)\in\ell^2_{\mathbb{R}} \quad \textit{and} \\ \textit{(4).} & \mathbf{Z}\,(\!(\mathsf{x}^*[n])\!) & = & \check{\mathsf{x}}^*\left(z^*\right) & \forall (\!(x_n)\!)\in\ell^2_{\mathbb{R}} \quad \textit{and} \\ \textit{(5).} & \mathbf{Z}\,(\!(\mathsf{x}^*[-n])\!) & = & \check{\mathsf{x}}^*\left(\frac{1}{z^*}\right) & \forall (\!(x_n)\!)\in\ell^2_{\mathbb{R}} \end{array}\right\}$$

✎PROOF:

$$\begin{aligned} \alpha\mathbb{Z}\check{\mathsf{x}}(z) &\triangleq \alpha\mathbf{Z}\,(\!(\mathsf{x}[n])\!) && \text{by definition of } \check{\mathsf{x}}(z) \\ &\triangleq \alpha\sum_{n\in\mathbb{Z}} \mathsf{x}[n]z^{-n} && \text{by definition of } \mathbf{Z} \text{ operator} \end{aligned}$$

[1] *Laurent series*: Abramovich and Aliprantis (2002) page 49

$$
\begin{aligned}
&\triangleq \sum_{n\in\mathbb{Z}} (\alpha \mathsf{x}[n]) z^{-n} && \text{by } \textit{distributive} \text{ property} \\
&\triangleq \mathbf{Z}\,(\!(\alpha \mathsf{x}[n])\!) && \text{by definition of } \mathbf{Z} \text{ operator} \\
z^{-k}\check{\mathsf{x}}(z) &= z^{-k}\mathbf{Z}\,(\!(\mathsf{x}[n])\!) && \text{by definition of } \check{\mathsf{x}}(z) && \text{(left hypothesis)} \\
&\triangleq z^{-k} \sum_{n=-\infty}^{n=+\infty} \mathsf{x}[n] z^{-n} && \text{by definition of } \mathbf{Z} && \text{(Definition V.1 page 429)} \\
&= \sum_{n=-\infty}^{n=+\infty} \mathsf{x}[n] z^{-n-k} \\
&= \sum_{m-k=-\infty}^{m-k=+\infty} \mathsf{x}[m-k] z^{-m} && \text{where } m \triangleq n+k && \implies n = m-k \\
&= \sum_{m=-\infty}^{m=+\infty} \mathsf{x}[m-k] z^{-m} \\
&= \sum_{n=-\infty}^{n=+\infty} \mathsf{x}[n-k] z^{-n} && \text{where } n \triangleq m \\
&\triangleq \mathbf{Z}\,(\!(\mathsf{x}[n-k])\!) && \text{by definition of } \mathbf{Z} && \text{(Definition V.1 page 429)} \\
\mathbf{Z}\,(\!(\mathsf{x}^*[n])\!) &\triangleq \sum_{n\in\mathbb{Z}} \mathsf{x}^*[n] z^{-n} && \text{by definition of } \mathbf{Z} && \text{(Definition V.1 page 429)} \\
&\triangleq \left(\sum_{n\in\mathbb{Z}} \mathsf{x}[n] (z^*)^{-n} \right)^* && \text{by definition of } \mathbf{Z} && \text{(Definition V.1 page 429)} \\
&\triangleq \check{\mathsf{x}}^*(z^*) && \text{by definition of } \mathbf{Z} && \text{(Definition V.1 page 429)} \\
\mathbf{Z}\,(\!(\mathsf{x}[-n])\!) &\triangleq \sum_{n\in\mathbb{Z}} \mathsf{x}[-n] z^{-n} && \text{by definition of } \mathbf{Z} && \text{(Definition V.1 page 429)} \\
&= \sum_{-m\in\mathbb{Z}} \mathsf{x}[m] z^{m} && \text{where } m \triangleq -n && \implies n = -m \\
&= \sum_{m\in\mathbb{Z}} \mathsf{x}[m] z^{m} && \text{by } \textit{absolutely summable} \text{ property} && \text{(Definition D.3 page 201)} \\
&= \sum_{m\in\mathbb{Z}} \mathsf{x}[m] \left(\frac{1}{z}\right)^{-m} && \text{by } \textit{absolutely summable} \text{ property} && \text{(Definition D.3 page 201)} \\
&\triangleq \check{\mathsf{x}}\left(\frac{1}{z}\right) && \text{by definition of } \mathbf{Z} && \text{(Definition V.1 page 429)} \\
\mathbf{Z}\,(\!(\mathsf{x}^*[-n])\!) &\triangleq \sum_{n\in\mathbb{Z}} \mathsf{x}^*[-n] z^{-n} && \text{by definition of } \mathbf{Z} && \text{(Definition V.1 page 429)} \\
&= \sum_{-m\in\mathbb{Z}} \mathsf{x}^*[m] z^{m} && \text{where } m \triangleq -n && \implies n = -m \\
&= \sum_{m\in\mathbb{Z}} \mathsf{x}^*[m] z^{m} && \text{by } \textit{absolutely summable} \text{ property} && \text{(Definition D.3 page 201)} \\
&= \sum_{m\in\mathbb{Z}} \mathsf{x}^*[m] \left(\frac{1}{z}\right)^{-m} && \text{by } \textit{absolutely summable} \text{ property} && \text{(Definition D.3 page 201)} \\
&= \left(\sum_{m\in\mathbb{Z}} \mathsf{x}[m] \left(\frac{1}{z^*}\right)^{-m} \right)^* && \text{by } \textit{absolutely summable} \text{ property} && \text{(Definition D.3 page 201)} \\
&\triangleq \check{\mathsf{x}}^*\left(\frac{1}{z^*}\right) && \text{by definition of } \mathbf{Z} && \text{(Definition V.1 page 429)}
\end{aligned}
$$

✎

Theorem V.2 (convolution theorem). *Let $\star$ be the convolution operator (Definition D.4 page 201).*

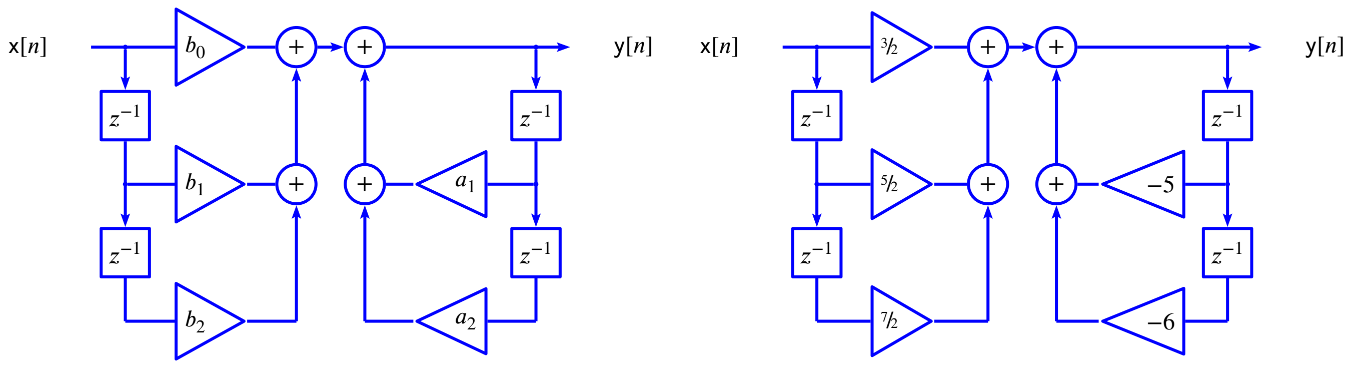


Figure V.1: Direct form 1 order 2 IIR filters

THM

$$\mathbf{Z}\,\underbrace{\left(\langle\!\langle x_n\rangle\!\rangle \star \langle\!\langle y_n\rangle\!\rangle\right)}_{\text{sequence convolution}} = \underbrace{\left(\mathbf{Z}\,\langle\!\langle x_n\rangle\!\rangle\right)\left(\mathbf{Z}\,\langle\!\langle y_n\rangle\!\rangle\right)}_{\text{series multiplication}} \qquad \forall \langle\!\langle x_n\rangle\!\rangle_{n\in\mathbb{Z}}, \langle\!\langle y_n\rangle\!\rangle_{n\in\mathbb{Z}} \in \ell^2_{\mathbb{R}}$$

✎PROOF:

$$\begin{aligned}
[\mathbf{Z}(x\star y)](z) &\triangleq \mathbf{Z}\left(\sum_{m\in\mathbb{Z}} x_m y_{n-m}\right) && \text{by definition of } \star && \text{(Definition D.4 page 201)}\\
&\triangleq \sum_{n\in\mathbb{Z}}\sum_{m\in\mathbb{Z}} x_m y_{n-m} z^{-n} && \text{by definition of } \mathbf{Z} && \text{(Definition V.1 page 429)}\\
&= \sum_{n\in\mathbb{Z}}\sum_{m\in\mathbb{Z}} x_m y_{n-m} z^{-n} && = \sum_{m\in\mathbb{Z}}\sum_{n\in\mathbb{Z}} x_m y_{n-m} z^{-n}\\
&= \sum_{m\in\mathbb{Z}}\sum_{k\in\mathbb{Z}} x_m y_k z^{-(m+k)} && \text{where } k \triangleq n-m && \iff n = m+k\\
&= \left[\sum_{m\in\mathbb{Z}} x_m z^{-m}\right]\left[\sum_{k\in\mathbb{Z}} y_k z^{-k}\right]\\
&\triangleq \left[\mathbf{Z}\,\langle\!\langle x_n\rangle\!\rangle\right]\left[\mathbf{Z}\,\langle\!\langle y_n\rangle\!\rangle\right] && \text{by definition of } \mathbf{Z} && \text{(Definition V.1 page 429)}
\end{aligned}$$

✏

V.2 From z-domain back to time-domain

$$\check{y}(z) = b_0X(z) + b_1z^{-1}X(z) + b_2z^{-2}X(z) - a_1z^{-1}Y(z) + a_2z^{-2}Y(z)$$

$$\mathrm{y}[n] = b_0\mathrm{x}[n] + b_1\mathrm{x}[n-1] + b_2\mathrm{x}[n-2] - a_1\mathrm{y}[n-1] - a_2\mathrm{y}[n-2]$$

Example V.1. See Figure V.1 (page 431)

$$\frac{3z^2+5z+7}{2z^2+10z+12} = \frac{3z^2+5z+7}{2\left(z^2+5z+6\right)} = \frac{\left(\tfrac{3}{2}z^2+\tfrac{5}{2}z+\tfrac{7}{2}\right)}{z^2+5z+6} = \frac{\left(\tfrac{3}{2}+\tfrac{5}{2}z^{-1}+\tfrac{7}{2}z^{-2}\right)}{1+5z^{-1}+6z^{-2}}$$

V.3 Zero locations

The system property of *minimum phase* is defined in Definition V.2 (next) and illustrated in Figure V.2 (page 432).

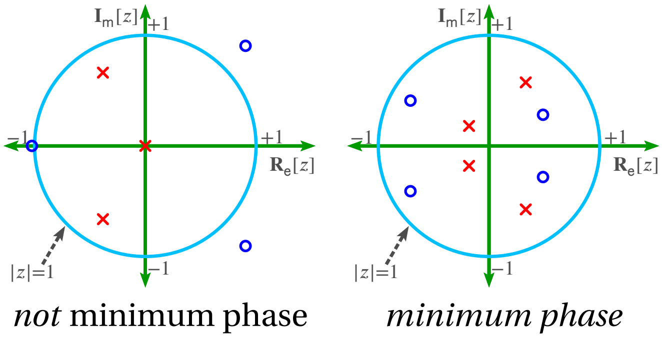


Figure V.2: Minimum Phase filter

Definition V.2. [2] *Let* $\check{x}(z) \triangleq \mathbf{Z}\left(\!\left(x_n\right)\!\right)$ *be the* Z TRANSFORM (Definition V.1 page 429) *of a sequence* $\left(\!\left(x_n\right)\!\right)_{n\in\mathbb{Z}}$ *in* $\ell^2_{\mathbb{R}}$. *Let* $\left(\!\left(z_n\right)\!\right)_{n\in\mathbb{Z}}$ *be the* ZEROS *of* $\check{x}(z)$.

DEF *The sequence* $\left(\!\left(x_n\right)\!\right)$ *is* ***minimum phase*** *if*

$$\underbrace{|z_n| < 1 \qquad \forall n \in \mathbb{Z}}_{\check{x}(z) \text{ has all its ZEROS inside the unit circle}}$$

The impulse response of a minimum phase filter has most of its energy concentrated near the beginning of its support, as demonstrated next.

Theorem V.3 (Robinson's Energy Delay Theorem). [3] *Let* $\mathsf{p}(z) \triangleq \sum_{n=0}^{N} a_n z^{-n}$ *and* $\mathsf{q}(z) \triangleq \sum_{n=0}^{N} b_n z^{-n}$ *be polynomials.*

THM

$$\left\{\begin{array}{lll} \mathsf{p} & \text{is MINIMUM PHASE} & \text{and} \\ \mathsf{q} & \text{is NOT } \textit{minimum phase} & \end{array}\right\} \implies \underbrace{\sum_{n=0}^{m-1} |a_n|^2}_{\text{"energy" of the first } m \text{ coefficients of } \mathsf{p}(z)} \geq \underbrace{\sum_{n=0}^{m-1} |b_n|^2}_{\text{"energy" of the first } m \text{ coefficients of } \mathsf{q}(z)} \qquad \forall 0 \leq m \leq N$$

But for more *symmetry*, put some zeros inside and some outside the unit circle (Figure V.4 page 434).

Example V.2. An example of a minimum phase polynomial is the Daubechies-4 scaling function. The minimum phase polynomial causes most of the energy to be concentrated near the origin, making it very *asymmetric*. In contrast, the Symlet-4 has a design very similar to that of Daubechies-4, but the selected zeros are not all within the unit circle in the complex z plane. This results in a scaling function that is more symmetric and less contrated near the origin. Both scaling functions are illustrated in Figure V.4 (page 434).

[2] Farina and Rinaldi (2000) page 91, Dumitrescu (2007) page 36
[3] Dumitrescu (2007) page 36, Robinson (1962), Robinson (1966) ⟨???⟩, Claerbout (1976) pages 52–53

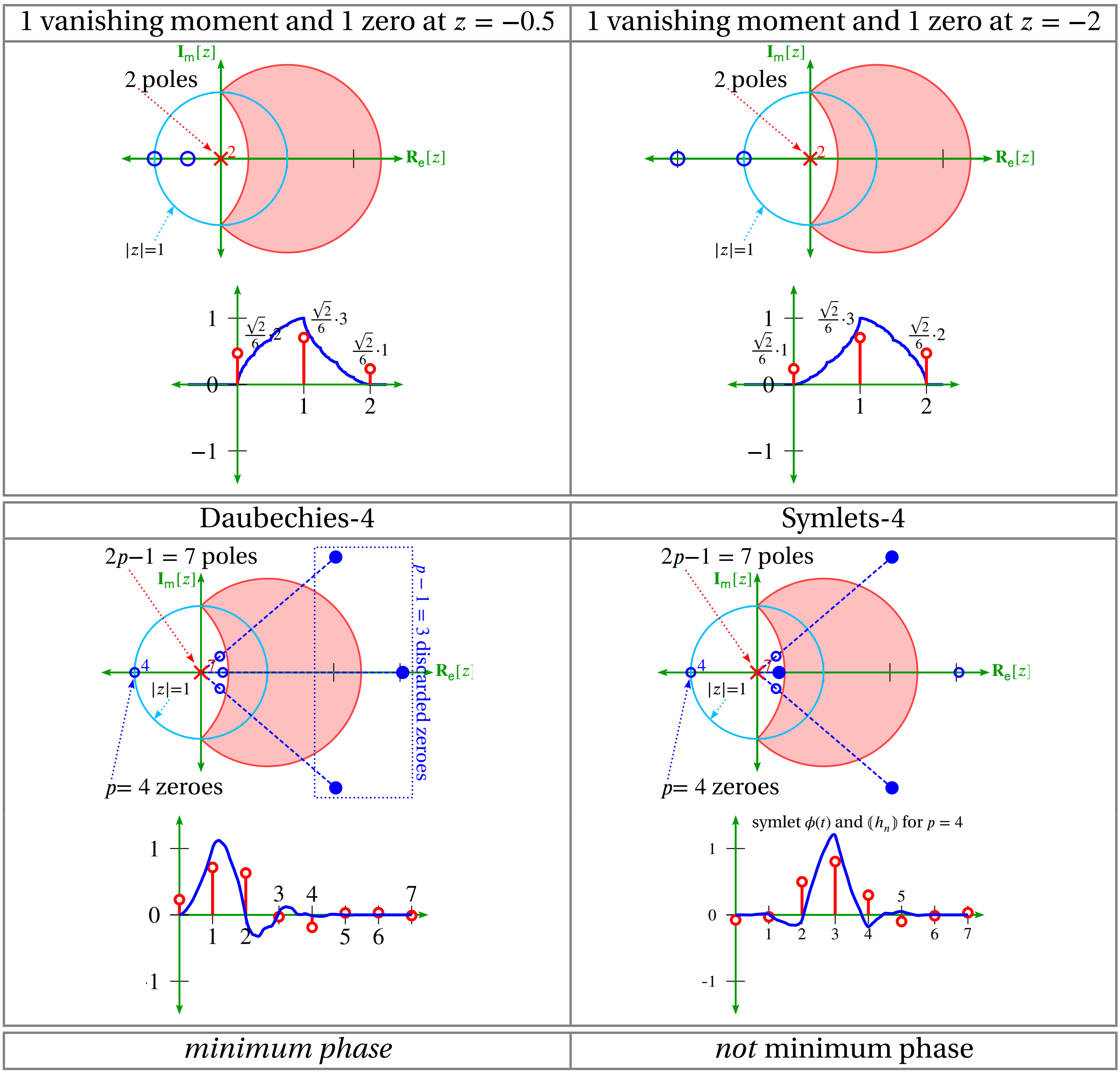


Figure V.3: Minimum/non-minimum phase comparison

V.4 Pole locations

Definition V.3.

DEF *A filter (or system or operator)* **H** *is* ***causal***
if its current output does not depend on future inputs.

Definition V.4.

DEF *A filter (or system or operator)* **H** *is* ***time-invariant***
if the mapping it performs does not change with time.

Definition V.5.

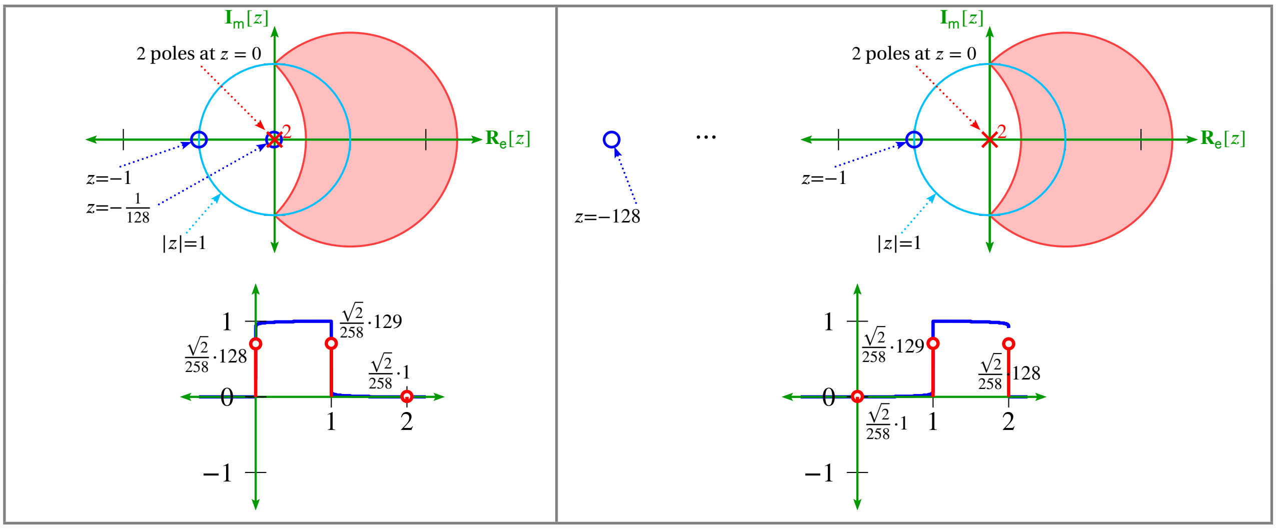


Figure V.4: Another minimum/non-minimum phase comparison

DEF *An operation* $\mathbf{H}$ *is **linear** if any output* y_n *can be described as a linear combination of inputs* x_n *as in*

$$\mathsf{y}_n = \sum_{m\in\mathbb{Z}} \mathsf{h}(m)\mathsf{x}(n-m)\,.$$

For a filter to be *stable*, place all the poles *inside* the unit circle.

Theorem V.4. *A causal LTI filter is **stable** if all of its poles are **inside** the unit circle.*

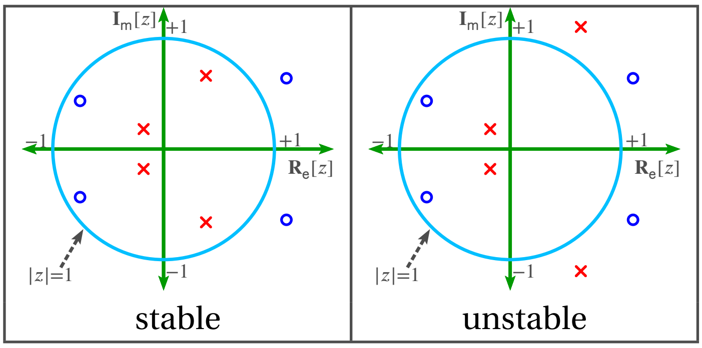


Figure V.5: Pole-zero plot stable/unstable causal LTI filters (Example V.3 page 434)

Example V.3. Stable/unstable filters are illustrated in Figure V.5 (page 434).

True or False? This filter has no poles:

$H(z) = b_0 + b_1 z^{-1} + b_2 z^{-2}$

$$\mathsf{H}(z) = b_0 + b_1 z^{-1} + b_2 z^{-2} = \frac{z^2}{z^2} \times \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{1} = \frac{b_0 z^2 + b_1 z^1 + b_2}{z^2}$$

V.5 Mirroring for real coefficients

If you want real coefficients, choose poles and zeros in conjugate pairs (next).

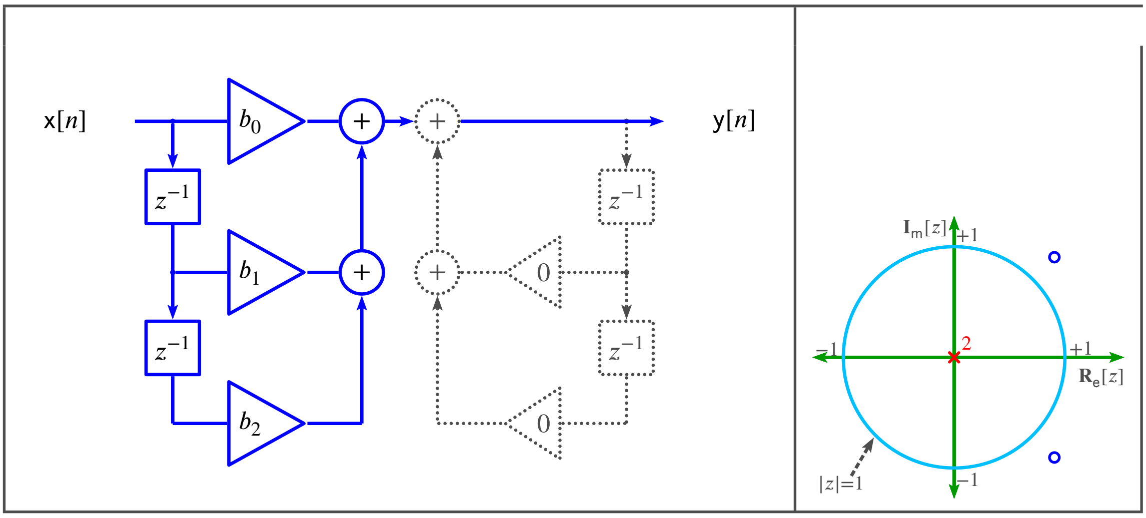


Figure V.6: FIR filters

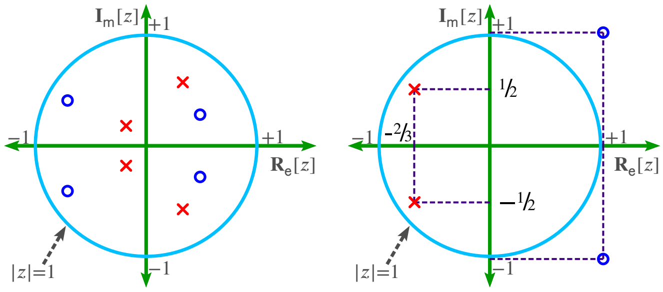


Figure V.7: Conjugate pair structure yielding real coefficients

Proposition V.1.

PRP $\left\{\begin{array}{l}\text{ZEROS } \textit{and} \text{ POLES}\\ \textit{occur in} \text{ CONJUGATE PAIRS}\end{array}\right\} \implies \left\{\begin{array}{l}\text{COEFFICIENTS}\\ \textit{are} \text{ REAL.}\end{array}\right\}$

✎PROOF:

$$\begin{aligned}
\left(z-p_1\right)\left(z-p_1^*\right) &= [z-(a+ib)][z-(a-ib)]\\
&= z^2+[-a+ib-ib-a]z-[ib]^2\\
&= z^2-2az+b^2
\end{aligned}$$

✏

Example V.4. See Figure V.7 (page 435).

$$\begin{aligned}
H(z) &= G\frac{\left[z-z_1\right]\left[z-z_2\right]}{\left[z-p_1\right]\left[z-p_2\right]} = G\frac{[z-(1+i)][z-(1-i)]}{[z-(-\tfrac{2}{3}+i\tfrac{1}{2})][z-(-\tfrac{2}{3}-i\tfrac{1}{2})]}\\
&= G\frac{z^2-z[(1-i)+(1+i)]+(1-i)(1+i)}{z^2-z[(-\tfrac{2}{3}+i\tfrac{1}{2})+(-\tfrac{2}{3}+i\tfrac{1}{2})]+(-\tfrac{2}{3}+i\tfrac{1}{2})(-\tfrac{2}{3}+i\tfrac{1}{2})}\\
&= G\frac{z^2-2z+2}{z^2-\tfrac{4}{3}z+(\tfrac{4}{3}+\tfrac{1}{4})} = G\frac{z^2-2z+2}{z^2-\tfrac{4}{3}z+\tfrac{19}{12}}
\end{aligned}$$

V.6 Rational polynomial operators

A digital filter is simply an operator on $\ell^2_{\mathbb{R}}$. If the digital filter is a causal LTI system, then it can be expressed as a rational polynomial in z as shown next.

Lemma V.1. *A causal LTI operator* **H** *can be expressed as a rational expression* $\check{\mathsf{h}}(z)$.

$$\begin{aligned}\check{\mathsf{h}}(z) &= \frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + \cdots + b_N z^{-N}}{1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_N z^{-N}} \\ &= \frac{\sum\limits_{n=0}^{N} b_n z^{-n}}{1 + \sum\limits_{n=1}^{N} a_n z^{-n}}\end{aligned}$$

A filter operation $\check{\mathsf{h}}(z)$ can be expressed as a product of its roots (poles and zeros).

$$\begin{aligned}\check{\mathsf{h}}(z) &= \frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + \cdots + b_N z^{-N}}{1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_N z^{-N}} \\ &= \alpha \frac{(z - z_1)(z - z_2)\cdots(z - z_N)}{(z - p_1)(z - p_2)\cdots(z - p_N)}\end{aligned}$$

where α is a constant, z_i are the zeros, and p_i are the poles. The poles and zeros of such a rational expression are often plotted in the z-plane with a unit circle about the origin (representing $z = e^{i\omega}$). Poles are marked with $\times$ and zeros with $\bigcirc$. An example is shown in Figure V.8 page 436. Notice that in this figure the zeros and poles are either real or occur in complex conjugate pairs.

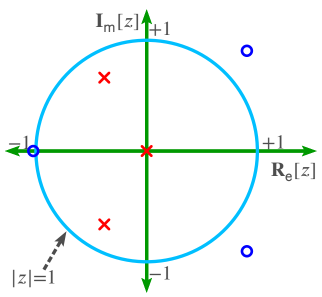


Figure V.8: Pole-zero plot for rational expression with real coefficients

V.7 Filter Banks

Conjugate quadrature filters (next definition) are used in *filter banks*. If $\check{\mathsf{x}}(z)$ is a *low-pass filter*, then the conjugate quadrature filter of $\check{\mathsf{y}}(z)$ is a *high-pass filter*.

Definition V.6. [4] *Let* $(\!(x_n)\!)_{n\in\mathbb{Z}}$ *and* $(\!(y_n)\!)_{n\in\mathbb{Z}}$ *be* SEQUENCES *(Definition D.2 page 200) in* $\ell^2_{\mathbb{R}}$ *(Definition D.3 page 201).*

[4] Strang and Nguyen (1996) page 109, Haddad and Akansu (1992) pages 256–259 ⟨section 4.5⟩, Vaidyanathan (1993) page 342 ⟨(7.2.7), (7.2.8)⟩, Smith and Barnwell (1984a), Smith and Barnwell (1984b), Mintzer (1985)

DEF *The sequence* $(\!(y_n)\!)$ *is a* ***conjugate quadrature filter*** *with shift* N *with respect to* $(\!(x_n)\!)$ *if*
$$y_n = \pm(-1)^n x^*_{N-n}$$
A CONJUGATE QUADRATURE FILTER *is also called a* ***CQF*** *or a* ***Smith-Barnwell filter***.
Any triple $\left((\!(x_n)\!), (\!(y_n)\!), N\right)$ *in this form is said to satisfy the*
conjugate quadrature filter condition *or the* ***CQF condition***.

Theorem V.5 (CQF theorem). [5] *Let* $\check{y}(\omega)$ *and* $\check{x}(\omega)$ *be the* DTFTs (Definition U.1 page 419) *of the sequences* $(\!(y_n)\!)_{n\in\mathbb{Z}}$ *and* $(\!(x_n)\!)_{n\in\mathbb{Z}}$*, respectively, in* $\ell^2_{\mathbb{R}}$ (Definition D.3 page 201).

THM
$$\begin{array}{rcrcll}
\underbrace{y_n = \pm(-1)^n x^*_{N-n}}_{\text{(1) CQF in "time"}} & \iff & \check{y}(z) & = & \pm(-1)^N z^{-N}\check{x}^*\left(\frac{-1}{z^*}\right) & \text{(2) CQF in "z-domain"}\\
& \iff & \check{y}(\omega) & = & \pm(-1)^N e^{-i\omega N}\check{x}^*(\omega+\pi) & \text{(3) CQF in "frequency"}\\
& \iff & x_n & = & \pm(-1)^N(-1)^n y^*_{N-n} & \text{(4) "reversed" CQF in "time"}\\
& \iff & \check{x}(z) & = & \pm z^{-N}\check{y}^*\left(\frac{-1}{z^*}\right) & \text{(5) "reversed" CQF in "z-domain"}\\
& \iff & \check{x}(\omega) & = & \pm e^{-i\omega N}\check{y}^*(\omega+\pi) & \text{(6) "reversed" CQF in "frequency"}
\end{array}$$
$\forall N\in\mathbb{Z}$

✎PROOF:

1. Proof that (1) $\implies$ (2):

$$\begin{aligned}
\check{y}(z) &= \sum_{n\in\mathbb{Z}} y_n z^{-n} && \text{by definition of } z\text{-transform} \quad \text{(Definition V.1 page 429)}\\
&= \sum_{n\in\mathbb{Z}} \underbrace{(\pm)(-1)^n x^*_{N-n}}_{CQF} z^{-n} && \text{by (1)}\\
&= \pm\sum_{m\in\mathbb{Z}} (-1)^{N-m} x^*_m z^{-(N-m)} && \text{where } m \triangleq N-n \implies \quad n = N-m\\
&= \pm(-1)^N z^{-N}\sum_{m\in\mathbb{Z}} (-1)^{-m} x^*_m \left(z^{-1}\right)^{-m}\\
&= \pm(-1)^N z^{-N}\sum_{m\in\mathbb{Z}} x^*_m \left(-\frac{1}{z}\right)^{-m}\\
&= \pm(-1)^N z^{-N}\left[\sum_{m\in\mathbb{Z}} x_m \left(-\frac{1}{z^*}\right)^{-m}\right]^*\\
&= \pm(-1)^N z^{-N}\check{x}^*\left(\frac{-1}{z^*}\right) && \text{by definition of } z\text{-transform} \quad \text{(Definition V.1 page 429)}
\end{aligned}$$

2. Proof that (1) $\impliedby$ (2):

$$\begin{aligned}
\check{y}(z) &= \pm(-1)^N z^{-N}\check{x}^*\left(\frac{-1}{z^*}\right) && \text{by (2)}\\
&= \pm(-1)^N z^{-N}\left[\sum_{m\in\mathbb{Z}} x_m \left(\frac{-1}{z^*}\right)^{-m}\right]^* && \text{by definition of } z\text{-transform} \quad \text{(Definition V.1 page 429)}\\
&= \pm(-1)^N z^{-N}\left[\sum_{m\in\mathbb{Z}} x^*_m \left(-z^{-1}\right)^{-m}\right] && \text{by definition of } z\text{-transform} \quad \text{(Definition V.1 page 429)}\\
&= \sum_{m\in\mathbb{Z}} (\pm)(-1)^{N-m} x^*_m z^{-(N-m)}
\end{aligned}$$

[5] Strang and Nguyen (1996) page 109, Mallat (1999) pages 236–238 ⟨(7.58),(7.73)⟩, Haddad and Akansu (1992) pages 256–259 ⟨section 4.5⟩, Vaidyanathan (1993) page 342 ⟨(7.2.7), (7.2.8)⟩

$$= \sum_{m\in\mathbb{Z}} (\pm)(-1)^n x^*_{N-n} z^{-n} \qquad \text{where } n = N-m \implies \qquad m \triangleq N-n$$
$$\implies x_n = \pm(-1)^n x^*_{N-n}$$

3. Proof that (1) $\implies$ (3):

$$\begin{aligned}
\breve{y}(\omega) &\triangleq \breve{x}(z)\Big|_{z=e^{i\omega}} && \text{by definition of } \textit{DTFT} \text{ (Definition U.1 page 419)}\\
&= \left[\pm(-1)^N z^{-N}\breve{x}\left(\frac{-1}{z^*}\right)\right]_{z=e^{i\omega}} && \text{by (2)}\\
&= \pm(-1)^N e^{-i\omega N}\breve{x}\left(e^{i\pi}e^{i\omega}\right)\\
&= \pm(-1)^N e^{-i\omega N}\breve{x}\left(e^{i(\omega+\pi)}\right)\\
&= \pm(-1)^N e^{-i\omega N}\breve{x}(\omega+\pi) && \text{by definition of } \textit{DTFT} \text{ (Definition U.1 page 419)}
\end{aligned}$$

4. Proof that (1) $\implies$ (6):

$$\begin{aligned}
\breve{x}(\omega) &= \sum_{n\in\mathbb{Z}} y_n e^{-i\omega n} && \text{by definition of } \textit{DTFT} \text{ (Definition U.1 page 419)}\\
&= \sum_{n\in\mathbb{Z}} \underbrace{\pm(-1)^n x^*_{N-n}}_{CQF} e^{-i\omega n} && \text{by (1)}\\
&= \sum_{m\in\mathbb{Z}} \pm(-1)^{N-m} x^*_m e^{-i\omega(N-m)} && \text{where } m \triangleq N-n \implies \quad n = N-m\\
&= \pm(-1)^N e^{-i\omega N} \sum_{m\in\mathbb{Z}} (-1)^m x^*_m e^{i\omega m}\\
&= \pm(-1)^N e^{-i\omega N} \sum_{m\in\mathbb{Z}} e^{i\pi m} x^*_m e^{i\omega m}\\
&= \pm(-1)^N e^{-i\omega N} \sum_{m\in\mathbb{Z}} x^*_m e^{i(\omega+\pi)m}\\
&= \pm(-1)^N e^{-i\omega N} \left[\sum_{m\in\mathbb{Z}} x_m e^{-i(\omega+\pi)m}\right]^*\\
&= \pm(-1)^N e^{-i\omega N} \breve{x}^*(\omega+\pi) && \text{by definition of } \textit{DTFT} \text{ (Definition U.1 page 419)}
\end{aligned}$$

5. Proof that (1) $\impliedby$ (3):

$$\begin{aligned}
y_n &= \frac{1}{2\pi}\int_{-\pi}^{+\pi} \breve{y}(\omega) e^{i\omega n}\,\mathrm{d}\omega && \text{by } \textit{inverse DTFT} \text{ (Theorem U.3 page 425)}\\
&= \frac{1}{2\pi}\int_{-\pi}^{+\pi} \underbrace{\pm(-1)^N e^{-iN\omega}\breve{x}^*(\omega+\pi)}_{\text{right hypothesis}} e^{i\omega n}\,\mathrm{d}\omega && \text{by right hypothesis}\\
&= \pm(-1)^N \frac{1}{2\pi}\int_{-\pi}^{+\pi} \breve{x}^*(\omega+\pi) e^{i\omega(n-N)}\,\mathrm{d}\omega && \text{by right hypothesis}\\
&= \pm(-1)^N \frac{1}{2\pi}\int_{0}^{2\pi} \breve{x}^*(v) e^{i(v-\pi)(n-N)}\,\mathrm{d}v && \text{where } v \triangleq \omega+\pi \implies \quad \omega = v-\pi\\
&= \pm(-1)^N e^{-i\pi(n-N)} \frac{1}{2\pi}\int_{0}^{2\pi} \breve{x}^*(v) e^{iv(n-N)}\,\mathrm{d}v\\
&= \pm(-1)^N \underbrace{(-1)^N}_{e^{i\pi N}} \underbrace{(-1)^n}_{e^{-i\pi n}} \left[\frac{1}{2\pi}\int_{0}^{2\pi} \breve{x}(v) e^{iv(N-n)}\,\mathrm{d}v\right]^*\\
&= \pm(-1)^n x^*_{N-n} && \text{by } \textit{inverse DTFT} \text{ (Theorem U.3 page 425)}
\end{aligned}$$

6. Proof that (1) $\iff$ (4):

$$
\begin{aligned}
y_n = \pm(-1)^n x^*_{N-n} &\iff (\pm)(-1)^n y_n = (\pm)(\pm)(-1)^n(-1)^n x^*_{N-n} \\
&\iff \pm(-1)^n y_n = x^*_{N-n} \\
&\iff \left(\pm(-1)^n y_n\right)^* = \left(x^*_{N-n}\right)^* \\
&\iff \pm(-1)^n y^*_n = x_{N-n} \\
&\iff x_{N-n} = \pm(-1)^n y^*_n \\
&\iff x_m = \pm(-1)^{N-m} y^*_{N-m} && \text{where } m \triangleq N-n \implies n = N-m \\
&\iff x_m = \pm(-1)^{N-m} y^*_{N-m} \\
&\iff x_m = \pm(-1)^{N}(-1)^m y^*_{N-m} \\
&\iff x_n = \pm(-1)^{N}(-1)^n y^*_{N-n} && \text{by change of free variables}
\end{aligned}
$$

7. Proofs for (5) and (6): not included. See proofs for (2) and (3).

✏

Theorem V.6. [6] *Let* $\check{y}(\omega)$ *and* $\check{x}(\omega)$ *be the* DTFTs (Definition U.1 page 419) *of the sequences* $(\!(y_n)\!)_{n\in\mathbb{Z}}$ *and* $(\!(x_n)\!)_{n\in\mathbb{Z}}$, *respectively, in* $\ell^2_{\mathbb{R}}$ (Definition D.3 page 201).

THM *Let* $y_n = \pm(-1)^n x^*_{N-n}$ (CQF CONDITION, *Definition V.6 page 436*). *Then*

$$
\left\{
\begin{array}{llcllll}
\text{(A)} & \left[\frac{\mathsf{d}}{\mathsf{d}\omega}\right]^n \check{y}(\omega)\Big|_{\omega=0} = 0 & \iff & \left[\frac{\mathsf{d}}{\mathsf{d}\omega}\right]^n \check{x}(\omega)\Big|_{\omega=\pi} & = & 0 & \text{(B)} \\
 & & \iff & \sum_{k\in\mathbb{Z}} (-1)^k k^n x_k & = & 0 & \text{(C)} \\
 & & \iff & \sum_{k\in\mathbb{Z}} k^n y_k & = & 0 & \text{(D)}
\end{array}
\right\} \quad \forall n\in\mathbb{W}
$$

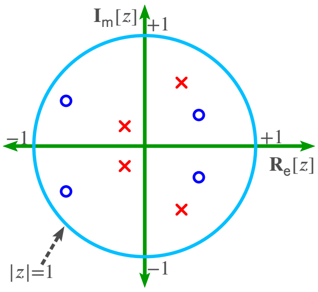
✎PROOF:

1. Proof that (A) $\implies$ (B):

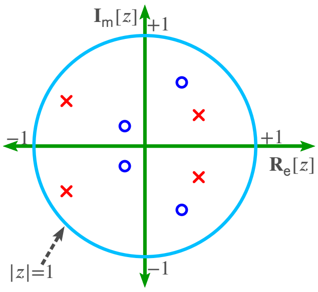
$$
\begin{aligned}
0 &= \left[\frac{\mathsf{d}}{\mathsf{d}\omega}\right]^n \check{y}(\omega)\Big|_{\omega=0} && \text{by (A)} \\
&= \left[\frac{\mathsf{d}}{\mathsf{d}\omega}\right]^n (\pm)(-1)^N e^{-i\omega N}\check{x}^*(\omega+\pi)\Big|_{\omega=0} && \text{by } \textit{CQF theorem} \quad \text{(Theorem V.5 page 437)} \\
&= (\pm)(-1)^N \sum_{\ell=0}^{n} \binom{n}{\ell} \left[\frac{\mathsf{d}}{\mathsf{d}\omega}\right]^\ell \left[e^{-i\omega N}\right] \cdot \left[\frac{\mathsf{d}}{\mathsf{d}\omega}\right]^{n-\ell} \left[\check{x}^*(\omega+\pi)\right]\Big|_{\omega=0} && \text{by } \textit{Leibnitz GPR} \quad \text{(Lemma H.2 page 225)} \\
&= (\pm)(-1)^N \sum_{\ell=0}^{n} \binom{n}{\ell} -iN^\ell e^{-i\omega N} \left[\frac{\mathsf{d}}{\mathsf{d}\omega}\right]^{n-\ell} \left[\check{x}^*(\omega+\pi)\right]\Big|_{\omega=0} \\
&= (\pm)(-1)^N \cancel{e^{-i0N}} \sum_{\ell=0}^{n} \binom{n}{\ell} -iN^\ell \left[\frac{\mathsf{d}}{\mathsf{d}\omega}\right]^{n-\ell} \left[\check{x}^*(\omega+\pi)\right]\Big|_{\omega=0}
\end{aligned}
$$

$$
\begin{aligned}
&\implies \check{x}^{(0)}(\pi) = 0 \\
&\implies \check{x}^{(1)}(\pi) = 0 \\
&\implies \check{x}^{(2)}(\pi) = 0 \\
&\implies \check{x}^{(3)}(\pi) = 0 \\
&\implies \check{x}^{(4)}(\pi) = 0 \\
&\quad\vdots \qquad \vdots \\
&\implies \check{x}^{(n)}(\pi) = 0 \quad \text{for } n = 0, 1, 2, \ldots
\end{aligned}
$$

[6] Vidakovic (1999) pages 82–83, Mallat (1999) pages 241–242



$\times$



$= \quad 1$

$$\frac{(z-z_1)(z-z_2)(z-z_3)(z-z_4)}{(z-p_1)(z-p_2)(z-p_3)(z-p_4)} \times \frac{(z-p_1)(z-p_2)(z-p_3)(z-p_4)}{(z-z_1)(z-z_2)(z-z_3)(z-z_4)} = 1$$

2. Proof that (A) $\Longleftarrow$ (B):

$$\begin{aligned}
0 &= \left[\frac{\mathsf{d}}{\mathsf{d}\omega}\right]^n \breve{\mathsf{x}}(\omega)\bigg|_{\omega=\pi} && \text{by (B)}\\
&= \left[\frac{\mathsf{d}}{\mathsf{d}\omega}\right]^n (\pm)e^{-i\omega N}\breve{\mathsf{y}}^*(\omega+\pi)\bigg|_{\omega=\pi} && \text{by } \textit{CQF theorem} \quad \text{(Theorem V.5 page 437)}\\
&= (\pm)\sum_{\ell=0}^{n}\binom{n}{\ell}\left[\frac{\mathsf{d}}{\mathsf{d}\omega}\right]^{\ell}\left[e^{-i\omega N}\right]\cdot\left[\frac{\mathsf{d}}{\mathsf{d}\omega}\right]^{n-\ell}\left[\breve{\mathsf{y}}^*(\omega+\pi)\right]\bigg|_{\omega=\pi} && \text{by } \textit{Leibnitz GPR} \quad \text{(Lemma H.2 page 225)}\\
&= (\pm)\sum_{\ell=0}^{n}\binom{n}{\ell}(-iN)^{\ell}e^{-i\omega N}\left[\frac{\mathsf{d}}{\mathsf{d}\omega}\right]^{n-\ell}\left[\breve{\mathsf{y}}^*(\omega+\pi)\right]\bigg|_{\omega=\pi}\\
&= (\pm)\cancel{e^{-i\pi N}}\sum_{\ell=0}^{n}\binom{n}{\ell}-iN^{\ell}\left[\frac{\mathsf{d}}{\mathsf{d}\omega}\right]^{n-\ell}\left[\breve{\mathsf{y}}^*(\omega+\pi)\right]\bigg|_{\omega=\pi}\\
&= (\pm)(-1)^N\sum_{\ell=0}^{n}\binom{n}{\ell}-iN^{\ell}\left[\frac{\mathsf{d}}{\mathsf{d}\omega}\right]^{n-\ell}\left[\breve{\mathsf{y}}^*(\omega+\pi)\right]\bigg|_{\omega=\pi}
\end{aligned}$$

$$\begin{aligned}
&\Longrightarrow & \breve{\mathsf{y}}^{(0)}(0) &= 0\\
&\Longrightarrow & \breve{\mathsf{y}}^{(1)}(0) &= 0\\
&\Longrightarrow & \breve{\mathsf{y}}^{(2)}(0) &= 0\\
&\Longrightarrow & \breve{\mathsf{y}}^{(3)}(0) &= 0\\
&\Longrightarrow & \breve{\mathsf{y}}^{(4)}(0) &= 0\\
&\vdots & \vdots &\\
&\Longrightarrow & \breve{\mathsf{y}}^{(n)}(0) &= 0\\
&\Longrightarrow & \breve{\mathsf{y}}^{(n)}(0) &= 0 \quad \text{for } n = 0,1,2,\ldots
\end{aligned}$$

3. Proof that (B) $\Longleftrightarrow$ (C): by Theorem U.5 page 427

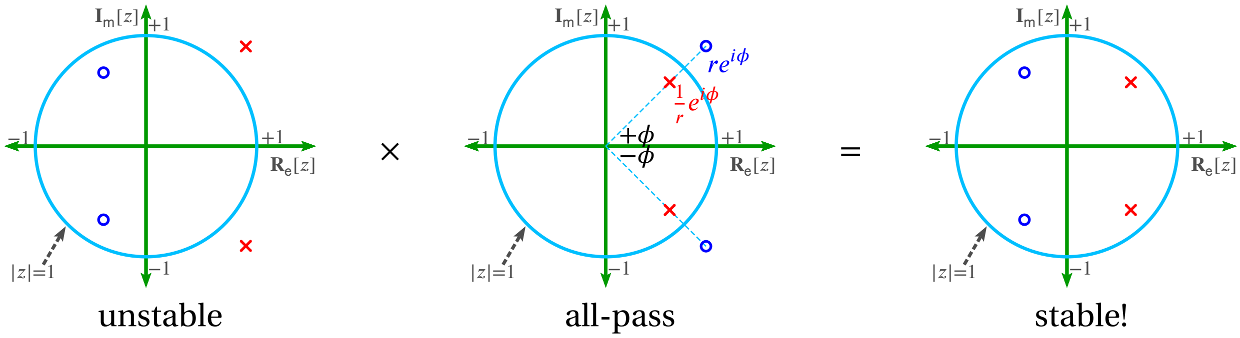
4. Proof that (A) $\Longleftrightarrow$ (D): by Theorem U.5 page 427

5. Proof that (CQF) $\not\Longleftarrow$ (A): Here is a counterexample: $\breve{\mathsf{y}}(\omega) = 0$.

✏

V.8 Inverting non-minimum phase filters

Minimum phase filters are easy to invert: each *zero* becomes a *pole* and each *pole* becomes a *zero*.



$$
\begin{aligned}
|A(z)|_{z=e^{i\omega}} &= \frac{1}{r}\left|\frac{z-re^{i\phi}}{z-\frac{1}{r}e^{i\phi}}\right|_{z=e^{i\omega}}
&&= \left|\frac{z-re^{i\phi}}{rz-e^{i\phi}}\right|_{z=e^{i\omega}} \\
&= \left|e^{i\phi}\left(\frac{e^{-i\phi}z-r}{rz-e^{i\phi}}\right)\right|_{z=e^{i\omega}}
&&= \left|z\left(\frac{e^{-i\phi}-rz^{-1}}{rz-e^{i\phi}}\right)\right|_{z=e^{i\omega}} \\
&= \left|-z\left(\frac{rz^{-1}-e^{-i\phi}}{rz-e^{i\phi}}\right)\right|_{z=e^{i\omega}}
&&= \left|\underbrace{e^{i\pi}}_{-1}e^{i\omega}\left(\frac{re^{-i\omega}-e^{-i\phi}}{re^{i\omega}-e^{i\phi}}\right)\right| \\
&= \left|\frac{1}{e^{-iv}}\left(\frac{re^{-i\omega}-e^{-i\phi}}{\left(re^{i\omega}-e^{i\phi}\right)^{*}}\right)\right|
&&= \left|\frac{re^{-i\omega}-e^{-i\phi}}{re^{-i\omega}-e^{-i\phi}}\right| \\
&= 1
\end{aligned}
$$

APPENDIX W

B-SPLINES

W.1 Definitions

Definition W.1. [1] *Let X be a set.*

DEF

*The **indicator function** $\mathbb{1} \in \{0,1\}^{2^X}$ is defined as*

$$\mathbb{1}_A(x) = \begin{cases} 1 & \text{for } x \in A \\ 0 & \text{otherwise} \end{cases} \qquad \forall x \in X,\ A \in 2^X$$

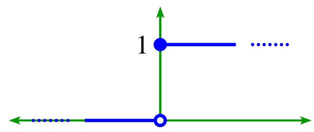
*The indicator function $\mathbb{1}$ is also called the **characteristic function**.*

Definition W.2. *Let X be a set.*

DEF

*The **step function** $\sigma \in \mathbb{R}^{\mathbb{R}}$ is defined as*

$$\sigma(x) \triangleq \mathbb{1}_{[0:\infty)}(x) \qquad \forall x \in \mathbb{R}.$$



Lemma W.1. *Let $\sigma(x)$ be the* STEP FUNCTION (Definition W.2 page 443).

LEM

$$\{g(x) > 0\} \implies \left\{\sigma\left[g(x)f(x)\right] = \sigma[f(x)]\right\} \qquad \forall f, g \in \mathbb{R}^{\mathbb{R}}$$

✎PROOF:

$$\begin{aligned}
\sigma\left[g(x)f(x)\right] &\triangleq \mathbb{1}_{[0:\infty)}\left[g(x)f(x)\right] && \text{by definition of } \sigma(x) && \text{(Definition W.2 page 443)} \\
&\triangleq \begin{cases} 1 & \text{for } g(x)f(x) \geq 0 \\ 0 & \text{otherwise} \end{cases} && \text{by definition of } \mathbb{1}_A(x) && \text{(Definition W.1 page 443)} \\
&= \begin{cases} 1 & \text{for } f(x) \geq 0 \\ 0 & \text{otherwise} \end{cases} && \text{by } g(x) > 0 \text{ hypothesis} && \\
&\triangleq \mathbb{1}_{[0:\infty)}[f(x)] && \text{by definition of } \mathbb{1}_A(x) && \text{(Definition W.1 page 443)} \\
&\triangleq \sigma[f(x)] && \text{by definition of } \sigma(x) && \text{(Definition W.2 page 443)}
\end{aligned}$$

✏

[1] de la Vallée-Poussin (1915) page 440, Hausdorff (1937) page 22, Feller (1971) page 104, Aliprantis and Burkinshaw (1998) page 126

Definition W.3. [2] *Let* $\mathbb{1}$ *be the* SET INDICATOR *function* (Definition W.1 page 443). *Let* $\mathsf{f}(x) \star \mathsf{g}(x)$ *represent the* CONVOLUTION *operation* (Definition D.1 page 199).

DEF *The* n***th order cardinal B-spline*** $\mathsf{N}_n(x)$ *for* $n \in \mathbb{W}$ *is defined as*

$$\mathsf{N}_n(x) \triangleq \left\{ \begin{array}{ll} \mathbb{1}_{[0:1)}(x) & \text{for } n = 0 \\ \mathsf{N}_{n-1}(x) \star \mathsf{N}_0(x) & \text{for } n \in \mathbb{W}\backslash 0 \end{array} \right\} \qquad \forall x \in \mathbb{R}$$

Lemma W.2. [3]

LEM
$$\mathsf{N}_n(x) = \int_{\tau=0}^{\tau=1} \mathsf{N}_{n-1}(x-\tau)\,\mathsf{d}\tau \qquad \forall n \in \{1,2,3,\ldots\}$$

✎PROOF:

$$\begin{aligned}
\mathsf{N}_n(x) &\triangleq \mathsf{N}_{n-1}(x) \star \mathsf{N}_0(x) && \text{by definition of } \mathsf{N}_n(x) && \text{(Definition W.3 page 444)} \\
&\triangleq \int_{\mathbb{R}} \mathsf{N}_{n-1}(x-\tau)\mathsf{N}_0(\tau)\,\mathsf{d}\tau && \text{by definition of convolution operation } \star && \text{(Definition D.1 page 199)} \\
&\triangleq \int_{\mathbb{R}} \mathsf{N}_{n-1}(x-\tau)\mathbb{1}_{[0:1)}(\tau)\,\mathsf{d}\tau && \text{by definition of } \mathsf{N}_0(x) && \text{(Definition W.3 page 444)} \\
&= \int_{[0:1)} \mathsf{N}_{n-1}(x-\tau)\,\mathsf{d}\tau && \text{by definition of } \mathbb{1}_A(x) && \text{(Definition W.1 page 443)} \\
&= \int_{[0:1]} \mathsf{N}_{n-1}(x-\tau)\,\mathsf{d}\tau \\
&\triangleq \int_0^1 \mathsf{N}_{n-1}(x-\tau)\,\mathsf{d}\tau
\end{aligned}$$

✏

Lemma W.3. *Let* $\mathsf{f}(x)$ *be a* FUNCTION *in* $\mathbb{R}^{\mathbb{R}}$. *Let* $\mathsf{F}(x)$ *be the* ANTI-DERIVATIVE *of* $\mathsf{f}(x)$. *Let* $\sigma(x)$ *be the* STEP FUNCTION (Definition W.2 page 443).

LEM
$$\begin{aligned}
&\int_{y=a}^{y=b} \mathsf{f}(x-y)\sigma(x-y)\,\mathsf{d}y \\
&= \left\{ \begin{array}{ll} -\int_{y=x-a}^{y=x-b} \mathsf{f}(y)\,\mathsf{d}y & \text{for } x \geq b \\ -\int_{y=x-a}^{y=0} \mathsf{f}(y)\,\mathsf{d}y & \text{for } a \leq x \leq b \\ 0 & \text{for } x \leq a \end{array} \right\} = \left\{ \begin{array}{ll} \mathsf{F}(x-a)-\mathsf{F}(x-b) & \text{for } x \geq b \\ \mathsf{F}(x-a)-\mathsf{F}(0) & \text{for } a \leq x \leq b \\ 0 & \text{for } x \leq a \end{array} \right\} \\
&= [\mathsf{F}(x-a)-\mathsf{F}(0)]\sigma(x-a) + [\mathsf{F}(0)-\mathsf{F}(x-b)]\sigma(x-b)
\end{aligned}$$

✎PROOF:

$$\begin{aligned}
\int_{y=a}^{y=b} \mathsf{f}(x-y)\sigma(x-y)\,\mathsf{d}y &= \left\{ \begin{array}{ll} \int_{y=a}^{y=b} \mathsf{f}(x-y)\,\mathsf{d}y & \text{for } x \geq b \\ \int_{y=a}^{y=x} \mathsf{f}(x-y)\,\mathsf{d}y & \text{for } a \leq x \leq b \\ 0 & \text{for } x \leq a \end{array} \right\} && \text{by definition of } \sigma \text{ (Definition W.2 page 443)} \\
&= \left\{ \begin{array}{ll} -\int_{u=x-a}^{u=x-b} \mathsf{f}(u)\,\mathsf{d}u & \text{for } x \geq b \\ -\int_{u=x-a}^{u=0} \mathsf{f}(u)\,\mathsf{d}u & \text{for } a \leq x \leq b \\ 0 & \text{for } x \leq a \end{array} \right\} && \text{where } u \triangleq x-y \implies y = x-u \\
&= \left\{ \begin{array}{ll} -\int_{y=x-a}^{y=x-b} \mathsf{f}(y)\,\mathsf{d}y & \text{for } x \geq b \\ -\int_{y=x-a}^{y=0} \mathsf{f}(y)\,\mathsf{d}y & \text{for } a \leq x \leq b \\ 0 & \text{for } x \leq a \end{array} \right\} && \text{by change of dummy variable } (u \to y)
\end{aligned}$$

[2] Chui (1992) page 85 ⟨(4.2.1)⟩, Christensen (2008) page 140, Chui (1988) page 1

[3] Christensen (2008) page 140, Chui (1992) page 85 ⟨(4.2.1)⟩, Chui (1988) page 1, Prasad and Iyengar (1997) page 145

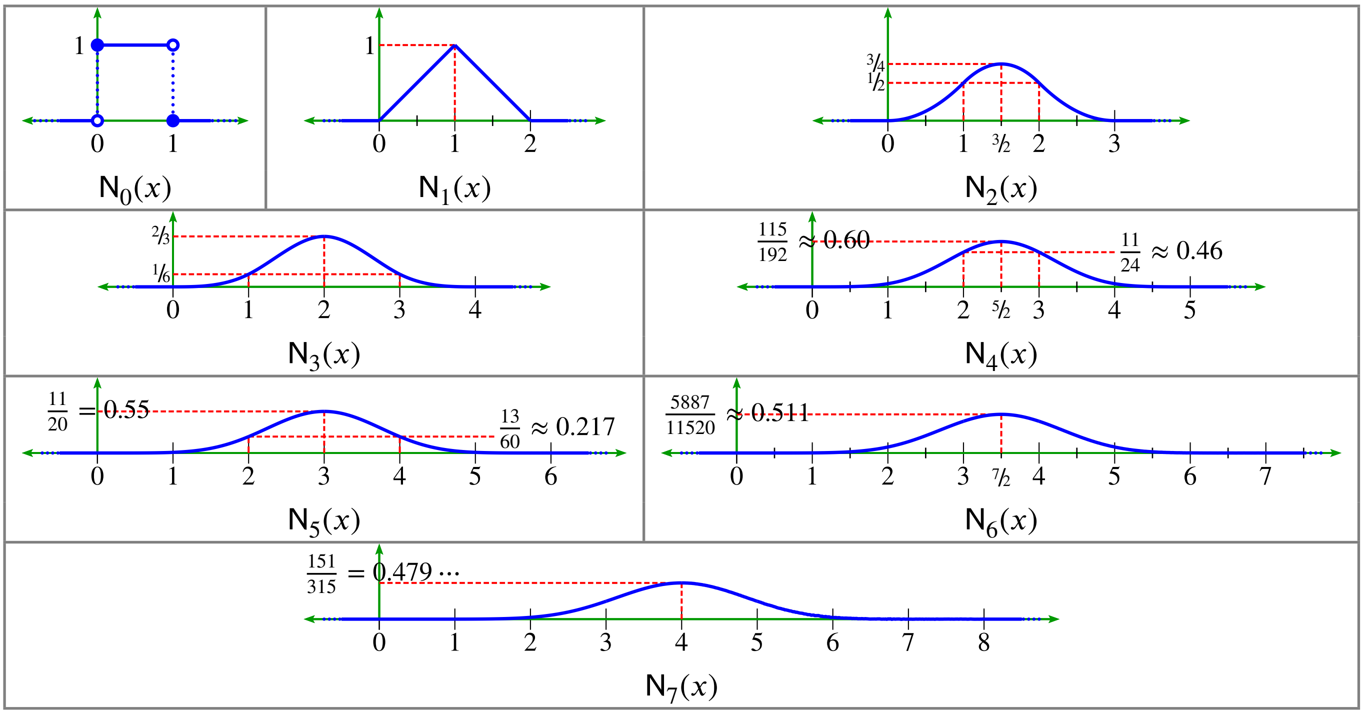


Figure W.1: some low order B-splines (Example W.1 page 445)

$$
\begin{aligned}
&= \left\{\begin{array}{ll} \mathsf{F}(x-a)-\mathsf{F}(x-b) & \text{for } x \ge b \\ \mathsf{F}(x-a)-\mathsf{F}(0) & \text{for } a \le x \le b \\ 0 & \text{for } x \le a \end{array}\right\} && \text{by } \textit{Fundamental Theorem of Calculus} \\
&= [\mathsf{F}(x-a)-\mathsf{F}(x-b)]\sigma(x-b) + [\mathsf{F}(x-a)-\mathsf{F}(0)][\sigma(x-a)-\sigma(x-b)] \\
&= [\mathsf{F}(x-a)-\mathsf{F}(0)]\sigma(x-a) + [\mathsf{F}(x-a)-\mathsf{F}(x-b)-\mathsf{F}(x-a)+\mathsf{F}(0)]\sigma(x-b) \\
&= [\mathsf{F}(x-a)-\mathsf{F}(0)]\sigma(x-a) + [\mathsf{F}(0)-\mathsf{F}(x-b)]\sigma(x-b)
\end{aligned}
$$

✏

Lemma W.4. *Let* $\sigma(x)$ *be the* STEP FUNCTION *(Definition W.2 page 443).*

LEM

$$\int_{\tau=0}^{\tau=1} (x-\tau-k)^n \sigma(x-\tau-k)\,\mathrm{d}\tau = \frac{1}{n+1}\left[(x-k)^{n+1}\sigma(x-k) - (x-k-1)^{n+1}\sigma(x-k-1)\right]$$

✎PROOF:

$$
\begin{aligned}
&\int_{\tau=0}^{\tau=1} (x-\tau-k)^n \sigma(x-\tau-k)\,\mathrm{d}\tau \\
&= \int_{y=k}^{y=k+1} (x-y)^n \sigma(x-y)\,\mathrm{d}y && \text{where } y \triangleq \tau + k \implies \tau = y-k \\
&= [\mathsf{F}(x-k)-\mathsf{F}(0)]\sigma(x-k) + [\mathsf{F}(0)-\mathsf{F}(x-k-1)]\sigma(x-k-1) && \text{by Lemma W.3 (page 444), where } \mathsf{f}(x) \triangleq x^n \\
&= \frac{\left[(x-k)^{n+1}-0\right]\sigma(x-k) + \left[0-(x-k-1)^{n+1}\right]\sigma(x-k-1)}{n+1} && \text{because } \mathsf{F}(x) \triangleq \int \mathsf{f}(x)\,\mathrm{d}x = \frac{x^{n+1}}{n+1}+c \\
&= \frac{1}{n+1}\left[(x-k)^{n+1}\sigma(x-k) - (x-k-1)^{n+1}\sigma(x-k-1)\right]
\end{aligned}
$$

✏

Example W.1. [4] Let $\sigma(x)$ be the *step function* (Definition W.2 page 443). Let $\binom{n}{k}$ be the *binomial coefficient* (Definition ?? page ??). The 0*th order B-spline* (Definition W.3 page 444) $\mathsf{N}_0(x)$ can be expressed as follows:

[4] Schumaker (2007) page 136 ⟨Table 1⟩

EX $$\mathsf{N}_0(x) = \left\{\begin{array}{ll} 1 & \text{for } x \in [0:1) \\ 0 & \text{otherwise}\end{array}\right\} = \left\{\sum_{k=0}^{1}(-1)^k\binom{1}{k}(x-k)^0\sigma(x-k) \quad \forall x\in\mathbb{R}\right\}$$

The B-spline $\mathsf{N}_0(x)$ is illustrated in Figure W.1 (page 445).

✎PROOF:

$$\begin{aligned}
\mathsf{N}_0(x) &= \mathbb{1}_{[0:1)}(x) && \text{by definition of } \mathsf{N}_0(x) && \text{(Definition W.3 page 444)}\\
&= \sigma(x) - \sigma(x-1) && \text{by definition of } \sigma(x) && \text{(Definition W.2 page 443)}\\
&= \left[\binom{1}{0}\sigma(x) - \binom{1}{1}\sigma(x-1)\right] && \text{by definition of \textit{binomial coefficient} } \binom{n}{k} && \text{(Definition ?? page ??)}\\
&= \sum_{k=0}^{1}(-1)^k\binom{1}{k}(x-k)^0\sigma(x-k) && \text{by definition of } \textstyle\sum \text{ operator}
\end{aligned}$$

✏

Example W.2 (1st order B-spline). [5] Let $\sigma(x)$ be the *step function*. Let $\binom{n}{k}$ be the *binomial coefficient*. The 1*st order B-spline* $\mathsf{N}_1(x)$ can be expressed as follows:

EX $$\mathsf{N}_1(x) = \left\{\begin{array}{rl} x & \text{for } x \in [0:1] \\ -x+2 & \text{for } x \in [1:2] \\ 0 & \text{otherwise}\end{array}\right\} = \left\{\sum_{k=0}^{2}(-1)^k\binom{2}{k}(x-k)\sigma(x-k) \quad \forall x\in\mathbb{R}\right\}$$

The B-spline $\mathsf{N}_1(x)$ is illustrated in Figure W.1 (page 445).

✎PROOF:

$$\begin{aligned}
\mathsf{N}_1(x) &= \int_{\tau=0}^{\tau=1} \mathsf{N}_0(x-\tau)\,\mathrm{d}\tau && \text{by Lemma W.2 page 444}\\
&= \int_{\tau=0}^{\tau=1} \sum_{k=0}^{1}(-1)^k\binom{1}{k}(x-\tau-k)^0\sigma(x-\tau-k)\,\mathrm{d}\tau && \text{by Example W.1 page 445}\\
&= \sum_{k=0}^{1}(-1)^k\binom{1}{k}\int_{\tau=0}^{\tau=1}(x-\tau-k)^0\sigma(x-\tau-k)\,\mathrm{d}\tau && \text{by \textit{linearity} of } \int \mathrm{d}\tau \text{ operator}\\
&= \sum_{k=0}^{1}(-1)^k\binom{1}{k}\frac{1}{0+1}\left[(x-k)^{0+1}\sigma(x-k) - (x-k-1)^{0+1}\sigma(x-k-1)\right] && \text{by Lemma W.4 page 445}\\
&= \begin{pmatrix} 1\{(x-0)\sigma(x-0) - (x-1)\sigma(x-1)\} \\ -1\{(x-1)\sigma(x-1) - (x-2)\sigma(x-2)\}\end{pmatrix}\\
&= x\sigma(x) - 2(x-1)\sigma(x-1) + (x-2)\sigma(x-2)\\
&= \sum_{k=0}^{2}(-1)^k\binom{2}{k}(x-k)\sigma(x-k) && \text{by def. of } \binom{n}{k} \text{ (Definition ?? page ??)}\\
&= \left\{\begin{array}{ll} x & \text{for } x \in [0:1) \\ -x+2 & \text{for } x \in [1:2] \\ 0 & \text{otherwise}\end{array}\right\} && \text{by def. of } \sigma(x) \text{ (Definition W.2 page 443)}
\end{aligned}$$

✏

Example W.3 (2nd order B-spline). [6] Let $\sigma(x)$ be the *step function*. Let $\binom{n}{k}$ be the *binomial coefficient*.

[5] Christensen (2008) page 148 ⟨Exercise 6.2⟩, Christensen (2010) page 212 ⟨Exercise 10.2⟩, Heil (2011) pages 142–143 ⟨Definition 4.22 (The Schauder System)⟩, Schumaker (2007) page 136 ⟨Table 1⟩, Stoer and Bulirsch (2002) page 124

[6] Christensen (2008) page 148 ⟨Exercise 6.2⟩, Christensen (2010) page 212 ⟨Exercise 10.2⟩, Schumaker (2007) page 136 ⟨Table 1⟩, Stoer and Bulirsch (2002) page 124

The $2nd$ *order B-spline* $\mathsf{N}_2(x)$ can be expressed as follows:

EX
$$\mathsf{N}_2(x) = \frac{1}{2}\begin{Bmatrix} x^2 & \text{for } x \in [0:1) \\ -2x^2+6x-3 & \text{for } x \in [1:2] \\ x^2-6x+9 & \text{for } x \in [2:3] \\ 0 & \text{otherwise} \end{Bmatrix} = \left\{ \frac{1}{2}\sum_{k=0}^{3}(-1)^k\binom{3}{k}(x-k)^2\sigma(x-k) \quad \forall x\in\mathbb{R} \right\}$$

The B-spline $\mathsf{N}_2(x)$ is illustrated in Figure W.1 (page 445).

✎PROOF:

$$\begin{aligned}
\mathsf{N}_2(x) &= \int_{\tau=0}^{\tau=1} \mathsf{N}_1(x-\tau)\,\mathrm{d}\tau && \text{by Lemma W.2 page 444} \\
&= \int_{\tau=0}^{\tau=1} \sum_{k=0}^{2}(-1)^k\binom{2}{k}(x-\tau-k)\sigma(x-\tau-k)\,\mathrm{d}\tau && \text{by Example W.2 page 446} \\
&= \sum_{k=0}^{2}(-1)^k\binom{2}{k}\int_{\tau=0}^{\tau=1}(x-\tau-k)\sigma(x-\tau-k)\,\mathrm{d}\tau && \text{by } \textit{linearity} \text{ of } \int \mathrm{d}\tau \text{ operator} \\
&= \sum_{k=0}^{2}(-1)^k\binom{2}{k}\frac{1}{1+1}\left[(x-k)^{1+1}\sigma(x-k)-(x-k-1)^{1+1}\sigma(x-k-1)\right] && \text{by Lemma W.4 page 445} \\
&= \frac{1}{2}\begin{pmatrix} 1 & \{(x-0)^2\sigma(x-0)-(x-1)^2\sigma(x-1)\} \\ -2 & \{(x-1)^2\sigma(x-1)-(x-2)^2\sigma(x-2)\} \\ +1 & \{(x-2)^2\sigma(x-2)-(x-3)^2\sigma(x-3)\} \end{pmatrix} \\
&= \frac{1}{2}\left[x^2\sigma(x)-3(x-1)^2\sigma(x-1)+3(x-2)^2\sigma(x-2)-(x-3)^2\sigma(x-3)\right] \\
&= \frac{1}{2}\sum_{k=0}^{3}(-1)^k\binom{3}{k}(x-k)^2\sigma(x-k) && \text{by def. of } \binom{n}{k} \text{ (Definition ?? page ??)} \\
&= \frac{1}{2}\begin{Bmatrix} x^2 & \text{for } x \in [0:1) \\ -2x^2+6x-3 & \text{for } x \in [1:2] \\ x^2-6x+9 & \text{for } x \in [2:3] \\ 0 & \text{otherwise} \end{Bmatrix} && \text{by def. of } \sigma(x) \text{ (Definition W.2 page 443)}
\end{aligned}$$

The final steps of this proof can be calculated "by hand" or by using the free and open source software package *Maxima* along with the script file listed in Section **??** (page **??**). ✏

W.2 Algebraic properties

Theorem W.1 (next) presents a closed form expression for an *nth order B-spline* $\mathsf{N}_n(x)$ based on the definition of $\mathsf{N}_n(x)$ given in Definition W.3 (page 444). Alternatively, Theorem W.1 could serve as the definition and Definition W.3 as a property.

Theorem W.1. [7] *Let* $\mathsf{N}_n(x)$ *be the* nTH ORDER B-SPLINE *(Definition W.3 page 444). Let* $\sigma(x)$ *be the* STEP FUNCTION *(Definition W.2 page 443).*

THM
$$\mathsf{N}_n(x) = \frac{1}{n!}\sum_{k=0}^{n+1}(-1)^k\binom{n+1}{k}(x-k)^n\sigma(x-k) \qquad \forall n \in \{0,1,2,\dots\} = \mathbb{W}$$

✎PROOF: Proof follows by induction:

[7] Christensen (2008) page 142 ⟨Theorem 6.1.3⟩, Chui (1992) page 84 ⟨(4.1.12)⟩

1. base case (choose one):
   - Proof for $n=0$ case: by Example W.1 (page 445).
   - Proof for $n=1$ case: by Example W.2 (page 446).
   - Proof for $n=2$ case: by Example W.3 (page 446).

2. inductive step—proof that n case $\implies$ $n+1$ case:

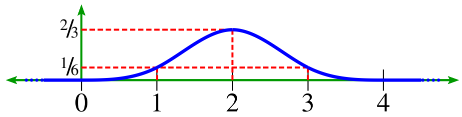
$$\begin{aligned}
&\mathsf{N}_{n+1}(x)\\
&= \int_0^1 \mathsf{N}_n(x-\tau)\,\mathrm{d}\tau && \text{by Lemma W.2 page 444}\\
&= \int_0^1 \frac{1}{n!}\sum_{k=0}^{k=n+1}(-1)^k\binom{n+1}{k}(x-\tau-k)^n\sigma(x-\tau-k)\,\mathrm{d}\tau && \text{by induction hypothesis}\\
&= \frac{1}{n!}\sum_{k=0}^{k=n+1}(-1)^k\binom{n+1}{k}\int_0^1 (x-\tau-k)^n\sigma(x-\tau-k)\,\mathrm{d}\tau && \text{by \textit{linearity} of } \int \mathrm{d}\tau \text{ operator}\\
&= \frac{1}{n!}\sum_{k=0}^{k=n+1}(-1)^k\binom{n+1}{k}\frac{1}{n+1}\left[(x-k)^{n+1}\sigma(x-k)-(x-k-1)^{n+1}\sigma(x-k-1)\right] && \text{by Lemma W.4 page 445}\\
&= \frac{1}{(n+1)!}\sum_{k=0}^{k=n+1}(-1)^k\binom{n+1}{k}\left[(x-k)^{n+1}\sigma(x-k)-(x-k-1)^{n+1}\sigma(x-k-1)\right]\\
&= \frac{1}{(n+1)!}\left[\sum_{k=0}^{k=n+1}(-1)^k\binom{n+1}{k}(x-k)^{n+1}\sigma(x-k)-\sum_{k=0}^{k=n+1}(-1)^k\binom{n+1}{k}(x-k-1)^{n+1}\sigma(x-k-1)\right]\\
&= \frac{1}{(n+1)!}\left[\sum_{k=0}^{k=n+1}(-1)^k\binom{n+1}{k}(x-k)^{n+1}\sigma(x-k)-\sum_{m=1}^{m=n+2}(-1)^{m-1}\binom{n+1}{m-1}(x-m)^{n+1}\sigma(x-m)\right]\\
&\qquad \text{where } m\triangleq k+1 \implies k=m-1\\
&= \frac{1}{(n+1)!}\begin{pmatrix}\displaystyle\sum_{k=0}^{k=n+1}(-1)^k\binom{n+1}{k}(x-k)^{n+1}\sigma(x-k)\\ \displaystyle-\sum_{m=1}^{m=n+2}(-1)^{m-1}\left[\binom{n+2}{m}-\binom{n+1}{m}\right](x-m)^{n+1}\sigma(x-m)\end{pmatrix} && \text{by \textit{Pascal's identity / Stifel formula} (Theorem ?? page ??)}\\
&= \frac{1}{(n+1)!}\begin{pmatrix}\displaystyle\sum_{m=1}^{m=n+2}(-1)^m\binom{n+2}{m}(x-m)^{n+1}\sigma(x-m)\\ \displaystyle-\sum_{m=1}^{m=n+2}(-1)^m\binom{n+1}{m}(x-m)^{n+1}\sigma(x-m)\\ \displaystyle+\sum_{k=0}^{k=n+1}(-1)^k\binom{n+1}{k}(x-k)^{n+1}\sigma(x-k)\end{pmatrix} && \text{note } (-1)^{m-1}=-(-1)^m\\
&= \frac{1}{(n+1)!}\left(\begin{array}{l|ll}
\displaystyle\sum_{m=0}^{m=n+2}(-1)^m\tbinom{n+2}{m}(x-m)^{n+1}\sigma(x-m) & \text{(A)} & \text{desired } n+1 \text{ case}\\
-\quad(-1)^0\tbinom{n+2}{0}(x-0)^{n+1}\sigma(x-0) & \text{(B)} & \text{cancelled by (F)}\\
\displaystyle-\sum_{m=1}^{m=n+1}(-1)^m\tbinom{n+1}{m}(x-m)^{n+1}\sigma(x-m) & \text{(C)} & \text{cancelled by (E)}\\
-\quad(-1)^{n+2}\tbinom{n+1}{n+2}(x-n-2)^{n+1}\sigma(x-n-2) & \text{(D)} & \tbinom{n+1}{n+2}=0 \text{ by Proposition ?? page ??}\\
\displaystyle+\sum_{k=1}^{k=n+1}(-1)^k\tbinom{n+1}{k}(x-k)^{n+1}\sigma(x-k) & \text{(E)} & \text{cancelled by (C)}\\
+\quad(-1)^0\tbinom{n+1}{0}(x-0)^{n+1}\sigma(x-0) & \text{(F)} & \tbinom{n+2}{0}=\tbinom{n+1}{0}=1, \text{ so (F) is cancelled by (B)}
\end{array}\right)
\end{aligned}$$

$$= \frac{1}{(n+1)!} \sum_{m=0}^{m=n+2} (-1)^m \binom{n+2}{m} (x-m)^{n+1} \sigma(x-m) \qquad (n+1 \text{ case})$$

⇨

Example W.4 (3rd order B-spline). [8] Let $\mathsf{N}_3(x)$ be the 3*rd order B-spline* (Definition W.3 page 444).[9]

$$\mathsf{N}_3(x) = \frac{1}{6}\begin{cases} x^3 & \text{for } 0 \le x \le 1 \\ -3x^3 + 12x^2 - 12x + 4 & \text{for } 1 \le x \le 2 \\ 3x^3 - 24x^2 + 60x - 44 & \text{for } 2 \le x \le 3 \\ -x^3 + 12x^2 - 48x + 64 & \text{for } 3 \le x \le 4 \\ 0 & \text{otherwise} \end{cases}$$



✎PROOF: This expression can be calculated "by hand" using Theorem W.1 (page 447) or by using the free and open source software package *Maxima* along with the script file listed in Section **??** (page **??**). ⇨

Example W.5. Let $\mathsf{N}_4(x)$ be the 4*th order B-spline* (Definition W.3 page 444).

$$\mathsf{N}_4(x) = \frac{1}{24}\begin{cases} x^4 & \text{for } 0 \le x \le= 1 \\ -4x^4 + 20x^3 - 30x^2 + 20x - 5 & \text{for } 1 \le x \le= 2 \\ 6x^4 - 60x^3 + 210x^2 - 300x + 155 & \text{for } 2 \le x \le= 3 \\ -4x^4 + 60x^3 - 330x^2 + 780x - 655 & \text{for } 3 \le x \le= 4 \\ x^4 - 20x^3 + 150x^2 - 500x + 625 & \text{for } 4 \le x \le= 5 \\ 0 & \text{otherwise} \end{cases}$$

✎PROOF: This expression can be calculated "by hand" using Theorem W.1 (page 447) or by using the free and open source software package *Maxima* along with the script file listed in Section **??** (page **??**). ⇨

Example W.6. Let $\mathsf{N}_5(x)$ be the 5*th order B-spline* (Definition W.3 page 444).

$$\mathsf{N}_5(x) = \frac{1}{120}\begin{cases} x^5 & \text{for } 0 \le x \le= 1 \\ -5x^5 + 30x^4 - 60x^3 + 60x^2 - 30x + 6 & \text{for } 1 \le x \le= 2 \\ 10x^5 - 120x^4 + 540x^3 - 1140x^2 + 1170x - 474 & \text{for } 2 \le x \le= 3 \\ -10x^5 + 180x^4 - 1260x^3 + 4260x^2 - 6930x + 4386 & \text{for } 3 \le x \le= 4 \\ 5x^5 - 120x^4 + 1140x^3 - 5340x^2 + 12270x - 10974 & \text{for } 4 \le x \le= 5 \\ -x^5 + 30x^4 - 360x^3 + 2160x^2 - 6480x + 7776 & \text{for } 5 \le x \le= 6 \\ 0 & \text{otherwise} \end{cases}$$

The 5*th order B-spline* $\mathsf{N}_5(x)$ is illustrated in Figure W.1 (page 445).

✎PROOF: This expression can be calculated "by hand" using Theorem W.1 (page 447) or by using the free and open source software package *Maxima* along with the script file listed in Section **??** (page **??**). ⇨

Example W.7. Let $\mathsf{N}_6(x)$ be the 6*th order B-spline* (Definition W.3 page 444).

$$\mathsf{N}_6(x) = \frac{1}{720}\begin{cases} x^6 & \text{for } 0 \le x \le 1 \\ -6x^6 + 42x^5 - 105x^4 + 140x^3 - 105x^2 + 42x - 7 & \text{for } 1 \le x \le 2 \\ 15x^6 - 210x^5 + 1155x^4 - 3220x^3 + 4935x^2 - 3990x + 1337 & \text{for } 2 \le x \le 3 \\ -20x^6 + 420x^5 - 3570x^4 + 15680x^3 - 37590x^2 + 47040x - 24178 & \text{for } 3 \le x \le 4 \\ 15x^6 - 420x^5 + 4830x^4 - 29120x^3 + 96810x^2 - 168000x + 119182 & \text{for } 4 \le x \le 5 \\ -6x^6 + 210x^5 - 3045x^4 + 23380x^3 - 100065x^2 + 225750x - 208943 & \text{for } 5 \le x \le 6 \\ x^6 - 42x^5 + 735x^4 - 6860x^3 + 36015x^2 - 100842x + 117649 & \text{for } 6 \le x \le 7 \\ 0 & \text{otherwise} \end{cases}$$

The 6*th order B-spline* $\mathsf{N}_6(x)$ is illustrated in Figure W.1 (page 445).

[8] Schumaker (2007) page 136 ⟨Table 1⟩, Shizgal (2015) page 92 ⟨(2.199)⟩, Szabó and Horváth (2004) page 146 ⟨(4)⟩, Wei and Billings (2006) page 578 ⟨Table 1⟩, Maleknejad et al. (2013) ⟨(9)⟩

[9] For help with plotting *B-splines*, see APPENDIX X (page 475).

✎PROOF: This expression can be calculated "by hand" using Theorem W.1 (page 447) or by using the free and open source software package *Maxima* along with the script file listed in Section ?? (page ??). ✏

Example W.8. Let $\mathsf{N}_7(x)$ be the 7*th order B-spline* (Definition W.3 page 444).

$$7!\mathsf{N}_7(x) = 5040\mathsf{N}_7(x) =$$

$$\left\{\begin{array}{rrrrrrrrl}
x^7 & & & & & & & & \text{for } 0 \le x \le 1\\
-\ 7x^7+ & 56x^6- & 168x^5+ & 280x^4- & 280x^3+ & 168x^2- & 56x+ & 8 & \text{for } 1 \le x \le 2\\
21x^7- & 336x^6+ & 2184x^5- & 7560x^4+ & 15400x^3- & 18648x^2+ & 12488x- & 3576 & \text{for } 2 \le x \le 3\\
-35x^7+ & 840x^6- & 8400x^5+ & 45360x^4- & 143360x^3+ & 267120x^2- & 273280x+ & 118896 & \text{for } 3 \le x \le 4\\
35x^7- & 1120x^6+ & 15120x^5- & 111440x^4+ & 483840x^3- & 1238160x^2+ & 1733760x- & 1027984 & \text{for } 4 \le x \le 5\\
-21x^7+ & 840x^6- & 14280x^5+ & 133560x^4- & 741160x^3+ & 2436840x^2- & 4391240x+ & 3347016 & \text{for } 5 \le x \le 6\\
7x^7- & 336x^6+ & 6888x^5- & 78120x^4+ & 528920x^3- & 2135448x^2+ & 4753336x- & 4491192 & \text{for } 6 \le x \le 7\\
-\ x^7+ & 56x^6- & 1344x^5+ & 17920x^4- & 143360x^3+ & 688128x^2- & 1835008x+ & 2097152 & \text{for } 7 \le x \le 8\\
 & & & & & & & 0 & \text{otherwise}
\end{array}\right\}$$

The 7*th order B-spline* $\mathsf{N}_7(x)$ is illustrated in Figure W.1 (page 445).

✎PROOF: This expression can be calculated "by hand" using Theorem W.1 (page 447) or by using the free and open source software package *Maxima* along with the script file listed in Section ?? (page ??). ✏

Example W.9. [10] The $(n+1)^2$ coefficients of the order $n, n-1, \ldots, 0$ monomials of each *B-spline* $\mathsf{N}_n(x)$ multiplied by $n!$ induce an *integer sequence*

$\mathbb{x} \triangleq (\!(1, 1, 0, -1, 2, 1, 0, 0, -2, 6, -3, 1, -6, 9, 1, 0, 0, 0, -3, 12, -12, 4, 3, -24, 60, -44, -1, 12, -48, 64, \ldots)\!)$

as more fully listed in Table W.1 (page 474). In this sequence $\mathbb{x} \triangleq (\!(x_0, x_1, x_2, \ldots)\!)$, the coefficients for the *order* n *B-spline* $\mathsf{N}_n(x)$ begin at the sequence index value

$$p \triangleq \sum_{k=0}^{n} k^2 = \frac{1}{6}n(n+1)(2n+1) \qquad \text{and end at index value } p + (n+1)^2 - 1.$$

For example, the coefficients for $\mathsf{N}_3(x)$ begin at index value $p \triangleq 0 + 1 + 4 + 9 = 14$ and end at index value $p + 4^2 - 1 = 29$. Using these coefficients gives the following expression for $\mathsf{N}_3(x)$:

$$\mathsf{N}_3(x) = \left[\begin{array}{rrrr|l}
1 & 0 & 0 & 0 & \text{for } 0 \le x < 1\\
-3 & 12 & -12 & 4 & \text{for } 1 \le x < 2\\
3 & -24 & 60 & -44 & \text{for } 2 \le x < 3\\
-1 & 12 & -48 & 64 & \text{for } 3 \le x < 4
\end{array}\right]\left[\begin{array}{l} x^3\\ x^2\\ x\\ 1\end{array}\right] = \left\{\begin{array}{rrrrl}
x^3 & & & & \text{for } 0 \le x < 1\\
-3x^3 & +12x^2 & -12x & +4 & \text{for } 1 \le x < 2\\
3x^3 & -24x^2 & +60x & -44 & \text{for } 2 \le x < 3\\
-x^3 & +12x^2 & -48x & +64 & \text{for } 3 \le x < 4\\
 & & & 0 & \text{otherwise}
\end{array}\right.$$

…which agrees with the result presented in Example W.4 (page 449).

✎PROOF:

1. The coefficients for the sequence $\mathbb{x}$ may be computed with assistance from *Maxima* together with the script file listed in Section ?? (page ??).
2. Proof that $\sum_{k=0}^{n} k^2 = \frac{1}{6}n(n+1)(2n+1)$: The summation is a *power sum*. The relation may be proved using *induction*.[11]

   (a) Base case: n=0 case …

$$\sum_{k=0}^{k=0} k^2 = 0$$
$$= \frac{0(0+1)(2\cdot 0+1)}{6}$$

[10] Greenhoe (2017b)

[11] Greenhoe (2017a) pages 186–187 ⟨Proposition 11.2 (Power Sums)⟩

$$= \left.\frac{n(n+1)(2n+1)}{6}\right|_{n=0}$$

(b) Base case: n=1 case …

$$\begin{aligned}\sum_{k=0}^{k=1} k^2 &= 0+1\\ &= \frac{1(1+1)(2\cdot 1+1)}{6}\\ &= \left.\frac{n(n+1)(2n+1)}{6}\right|_{n=1}\end{aligned}$$

(c) inductive step—proof that n case $\implies$ $n+1$ case:

$$\begin{aligned}\sum_{k=0}^{n+1} k^2 &= \left(\sum_{k=0}^{n} k^2\right) + \left(n+1\right)^2\\ &= \left(\frac{n(n+1)(2n+1)}{6}\right) + \left(n+1\right)^2 && \text{by } n \text{ case hypothesis}\\ &= \left(n+1\right)\left(\frac{n(2n+1)+6(n+1)}{6}\right)\\ &= \left(n+1\right)\left(\frac{2n^2+7n+6}{6}\right)\\ &= \left(n+1\right)\left(\frac{(n+2)(2n+3)}{6}\right)\\ &= \frac{(n+1)[(n+1)+1][2(n+1)+1]}{6}\end{aligned}$$

✏

Theorem W.2. [12]

T H M	$\frac{\mathsf{d}}{\mathsf{d}x}\mathsf{N}_n(x) = \mathsf{N}_{n-1}(x) - \mathsf{N}_{n-1}(x-1) \qquad \forall n \in \mathbb{N},\ \forall x \in \mathbb{R}$

✎PROOF:

1. Proof using Lemma W.2 (page 444) and the *Fundamental Theorem of Calculus*:

$$\begin{aligned}\frac{\mathsf{d}}{\mathsf{d}x}\mathsf{N}_n(x) &= \frac{\mathsf{d}}{\mathsf{d}x}\int_0^1 \mathsf{N}_{n-1}(x-\tau)\,\mathsf{d}\tau && \text{by Lemma W.2 page 444}\\ &= \frac{\mathsf{d}}{\mathsf{d}x}\int_{x-u=0}^{x-u=1} \mathsf{N}_{n-1}(u)(-1)\,\mathsf{d}u && \text{where } u \triangleq x-\tau \implies \tau = x-u\\ &= \frac{\mathsf{d}}{\mathsf{d}x}\int_{u=x-1}^{u=x} \mathsf{N}_{n-1}(u)\,\mathsf{d}u\\ &= \left\{\frac{\mathsf{d}}{\mathsf{d}x}\int \mathsf{N}_{n-1}(u)\,\mathsf{d}u\bigg|_{u=x}\right\} - \left\{\frac{\mathsf{d}}{\mathsf{d}x}\int \mathsf{N}_{n-1}(u)\,\mathsf{d}u\bigg|_{u=x-1}\right\} && \text{by \textit{Fundamental Theorem of Calculus}[13]}\\ &= \left\{\mathsf{N}_{n-1}(x)\frac{\mathsf{d}}{\mathsf{d}x}(x)\right\} - \left\{\mathsf{N}_{n-1}(x-1)\frac{\mathsf{d}}{\mathsf{d}x}(x-1)\right\} && \text{by \textit{Chain Rule}[14]}\\ &= \mathsf{N}_{n-1}(x) - \mathsf{N}_{n-1}(x-1)\end{aligned}$$

[12] Höllig (2003) page 25 ⟨3.2⟩, Schumaker (2007) page 121 ⟨Theorem 4.16⟩
[13] Hijab (2011) page 163 ⟨Theorem 4.4.3⟩
[14] Hijab (2011) pages 73–74 ⟨Theorem 3.1.2⟩

2. Proof using Lemma W.2 (page 444) and *induction*:

(a) Base case …proof for $n = 1$ case:

$$\begin{aligned}
\mathsf{N}_0(x) - \mathsf{N}_0(x-1) &= \underbrace{\sigma(x) - \sigma(x-1)}_{\mathsf{N}_0(x)} - \underbrace{[\sigma(x-1) - \sigma(x-2)]}_{\mathsf{N}_0(x-1)} && \text{by Example W.1 page 445}\\
&= \sigma(x) - 2\sigma(x-1) + \sigma(x-2)\\
&= \sum_{k=0}^{2}(-1)^k\binom{2}{k}\sigma(x-k)\\
&= \frac{\mathrm{d}}{\mathrm{d}x}\sum_{k=0}^{2}(-1)^k\binom{2}{k}(x-k)\sigma(x-k)\\
&= \frac{\mathrm{d}}{\mathrm{d}x}\mathsf{N}_1(x) && \text{by Example W.2 page 446}
\end{aligned}$$

(b) Base case …proof for $n = 2$ case:

$$\begin{aligned}
&\mathsf{N}_1(x) - \mathsf{N}_1(x-1)\\
&= \underbrace{x\sigma(x) - 2(x-1)\sigma(x-1) + (x-2)\sigma(x-2)}_{\mathsf{N}_1(x)}\\
&\quad - \underbrace{[(x-1)\sigma(x-1) - 2(x-2)\sigma(x-2) + (x-3)\sigma(x-3)]}_{\mathsf{N}_1(x-1)} && \text{by Example W.2 page 446}\\
&= x\sigma(x) + [-2x+2-x+1]\sigma(x-1) + [x-2+2x-4]\sigma(x-2) + [-x+3]\sigma(x-3)\\
&= x\sigma(x) + [-3x+3]\sigma(x-1) + [3x-6]\sigma(x-2) + [-x+3]\sigma(x-3)\\
&= \frac{\mathrm{d}}{\mathrm{d}x}\left\{\begin{array}{l} \frac{1}{2}x^2\sigma(x) + \left[-\frac{3}{2}x^2 + 3x - \frac{1}{2}\right]\sigma(x-1) + \left[\frac{3}{2}x^2 - 6x + 3\right]\sigma(x-2)\\ \qquad + \left[-\frac{1}{2}x^2 + 3x - \frac{5}{2}\right]\sigma(x-3)\end{array}\right\}\\
&= \frac{\mathrm{d}}{\mathrm{d}x}\mathsf{N}_2(x) && \text{by Example W.3 page 446}
\end{aligned}$$

(c) Proof that n case $\implies$ $n+1$ case:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}\mathsf{N}_{n+1}(x) &= \frac{\mathrm{d}}{\mathrm{d}x}\int_0^1 \mathsf{N}_n(x-\tau)\,\mathrm{d}\tau && \text{by Lemma W.2 page 444}\\
&= \int_0^1 \frac{\partial}{\partial x}\mathsf{N}_n(x-\tau)\,\mathrm{d}\tau && \text{by \textit{Leibniz Integration Rule} (Theorem H.2 page 225)}\\
&= \int_0^1 \left[\mathsf{N}_{n-1}(x-\tau) - \mathsf{N}_{n-1}(x-1-\tau)\right]\mathrm{d}\tau && \text{by left hypothesis}\\
&= \int_0^1 \mathsf{N}_{n-1}(x-\tau)\,\mathrm{d}\tau - \int_0^1 \mathsf{N}_{n-1}(x-1-\tau)\,\mathrm{d}\tau\\
&= \mathsf{N}_n(x) - \mathsf{N}_n(x-1) && \text{by Lemma W.2 page 444}
\end{aligned}$$

✎

Theorem W.3 (*B-spline recursion*). [15] *Let* $\mathsf{N}_n(x)$ *be the* nTH ORDER B-SPLINE *(Definition W.3 page 444).*

THM
$$\mathsf{N}_n(x) = \frac{x}{n}\mathsf{N}_{n-1}(x) + \frac{n+1-x}{n}\mathsf{N}_{n-1}(x-1) \qquad \forall n \in \{1,2,3,\ldots\},\ \forall x \in \mathbb{R}$$

[15] Christensen (2008) page 140 ⟨Theorem 6.1.1⟩, Höllig (2003) page 27 ⟨3.4⟩, Schumaker (2007) page 120 ⟨Theorem 4.15⟩, de Boor (2001) page 90 ⟨B-Spline Property (i)⟩, Chui (1988) page 2 ⟨Theorem 1.1⟩, Wojtaszczyk (1997) page 53 ⟨Theorem 3.7⟩, Cox (1972), de Boor (1972)

✎PROOF:

1. Base case ...proof for $n = 1$ case:

$$\frac{x}{1}\mathsf{N}_0(x) + \frac{1+1-x}{1}\mathsf{N}_0(x-1) = \frac{x}{1}\underbrace{[\sigma(x)-\sigma(x-1)]}_{\mathsf{N}_0(x)} + \frac{1+1-x}{1}\underbrace{[\sigma(x-1)-\sigma(x-2)]}_{\mathsf{N}_0(x-1)}$$
$$= x\sigma(x) + [-x-x+2]\sigma(x-1) + [x-2]\sigma(x-2)$$
$$= \mathsf{N}_1(x) \qquad \text{by Example W.2 page 446}$$

2. Induction step ...proof that n case $\implies$ $n+1$ case:

$$\frac{x}{n+1}\mathsf{N}_n(x) + \frac{n+2-x}{n+1}\mathsf{N}_n(x-1) + c_1$$
$$= \int \frac{\mathsf{d}}{\mathsf{dx}}\left\{\frac{x}{n+1}\mathsf{N}_n(x) + \frac{n+2-x}{n+1}\mathsf{N}_n(x-1)\right\}\,\mathsf{dx}$$
$$= \int \underbrace{\frac{1}{n+1}\mathsf{N}_n(x) + \frac{x}{n+1}\frac{\mathsf{d}}{\mathsf{dx}}\mathsf{N}_n(x)}_{\frac{\mathsf{d}}{\mathsf{dx}}\frac{x}{n+1}\mathsf{N}_n(x)} + \underbrace{\frac{-1}{n+1}\mathsf{N}_n(x-1) + \frac{n+2-x}{n}\frac{\mathsf{d}}{\mathsf{dx}}\mathsf{N}_n(x-1)}_{\frac{\mathsf{d}}{\mathsf{dx}}\frac{n+2-x}{n+1}\mathsf{N}_n(x-1)}\,\mathsf{dx}$$
by *product rule*
$$= \int \frac{1}{n+1}\underbrace{\left[\frac{x}{n}\mathsf{N}_{n-1}(x) + \frac{n+1-x}{n}\mathsf{N}_{n-1}(x-1)\right]}_{\text{by } n \text{ hypothesis}} + \frac{x}{n+1}\underbrace{\left[\mathsf{N}_{n-1}(x) - \mathsf{N}_{n-1}(x-1)\right]}_{\text{by Theorem W.2 page 451}}$$
$$- \underbrace{\left[\frac{x-1}{n^2+n}\mathsf{N}_{n-1}(x-1) + \frac{n-x+2}{n(n+1)}\mathsf{N}_{n-1}(x-2)\right]}_{\text{by induction hypothesis}}$$
$$+ \frac{n+2-x}{n+1}\underbrace{\left[\mathsf{N}_{n-1}(x-1) - \mathsf{N}_{n-1}(x-2)\right]}_{\text{by Theorem W.2 page 451}}\,\mathsf{dx}$$
$$= \int \left[\frac{x}{n(n+1)} + \frac{x}{n+1}\right]\mathsf{N}_{n-1}(x) + \left[\frac{n-x+1}{n(n+1)} - \frac{x-1}{n(n+1)} + \frac{n+2-2x}{n+1}\right]\mathsf{N}_{n-1}(x-1)$$
$$+ \left[\frac{-n-2+x}{n(n+1)} + \frac{-n-2+x}{n+1}\right]\mathsf{N}_{n-1}(x-2)\,\mathsf{dx}$$
$$= \int \left[\frac{x+nx}{n(n+1)}\right]\mathsf{N}_{n-1}(x) + \left[\frac{n+2-2x+n(n+2-2x)}{n(n+1)}\right]\mathsf{N}_{n-1}(x-1)$$
$$+ \left[\frac{-n-2+x+n(-n-2+x)}{n(n+1)}\right]\mathsf{N}_{n-1}(x-2)\,\mathsf{dx}$$
$$= \int \left[\frac{x}{n}\right]\mathsf{N}_{n-1}(x) + \left[\frac{n+2-2x}{n}\right]\mathsf{N}_{n-1}(x-1) + \left[\frac{-n-2+x}{n}\right]\mathsf{N}_{n-1}(x-2)\,\mathsf{dx}$$
$$= \int \underbrace{\left[\frac{x}{n}\right]\mathsf{N}_{n-1}(x) + \left[\frac{n+1-x}{n}\right]\mathsf{N}_{n-1}(x-1)}_{\mathsf{N}_n(x)}$$
$$- \underbrace{\left[\frac{x-1}{n}\right]\mathsf{N}_{n-1}(x-1) - \left[\frac{n+2-x}{n}\right]\mathsf{N}_{n-1}(x-2)}_{\mathsf{N}_{n-1}(x-1)}\,\mathsf{dx}$$
$$= \int \mathsf{N}_n(x) - \mathsf{N}_n(x-1)\,\mathsf{dx} \qquad \text{by } n \text{ hypothesis}$$
$$= \int \frac{\mathsf{d}}{\mathsf{dx}}\mathsf{N}_{n+1}(x)\,\mathsf{dx} \qquad \text{by Theorem W.2 page 451}$$
$$= \mathsf{N}_{n+1}(x) + c_2$$

Proof that $c_1 = c_2$: By item (2) (page 454), $\mathsf{N}_n(x) = 0$ for $x < 0$. Therefore, $c_1 = c_2$.

⇨

Theorem W.4 (*B-spline general form*). [16] *Let* $\mathsf{N}_n(x)$ *be the* nTH ORDER B-SPLINE *(Definition W.3 page 444). Let* $\operatorname{supp}\mathsf{f}$ *be the* SUPPORT *of a function* $\mathsf{f} \in \mathbb{R}^{\mathbb{R}}$.

THM

1.	$\mathsf{N}_n(x) \geq 0$	$\forall n \in \mathbb{W},\ \forall x \in \mathbb{R}$	(NON-NEGATIVE)
2.	$\operatorname{supp}\mathsf{N}_n(x) = [0 : n+1]$	$\forall n \in \mathbb{W}$	(CLOSED SUPPORT)
3.	$\int_{\mathbb{R}} \mathsf{N}_n(x)\,\mathrm{d}x = 1$	$\forall n \in \mathbb{W}$	(UNIT AREA)
4.	$\mathsf{N}_n\left(\frac{n+1}{2} - x\right) = \mathsf{N}_n\left(\frac{n+1}{2} + x\right)$	$\forall n \in \mathbb{N}\ \ \forall x \in \mathbb{R}$	(SYMMETRIC *about* $x = \frac{n+1}{2}$)

✎PROOF:

1. Proof that $\mathsf{N}_n(x) \geq 0$ (proof by induction):

   (a) base case...proof that $\mathsf{N}_0(x) \geq 0$:

   $$\begin{aligned}
   \mathsf{N}_0(x) &\triangleq \mathbb{1}_{[0:1)}(x) && \text{by definition of } \mathsf{N}_0(x) && \text{(Definition W.3 page 444)}\\
   &\geq 0 && \text{by definition of } \mathbb{1}_A(x) && \text{(Definition W.1 page 443)}
   \end{aligned}$$

   (b) inductive step—proof that $\{\mathsf{N}_n(x) \geq 0\} \implies \{\mathsf{N}_{n+1}(x) \geq 0\}$:

   $$\begin{aligned}
   \mathsf{N}_{n+1}(x) &= \int_{\tau=0}^{\tau=1} \mathsf{N}_n(x-\tau)\,\mathrm{d}\tau && \text{by Lemma W.2 page 444}\\
   &\geq 0 && \text{by induction hypothesis } (\mathsf{N}_n(x) \geq 0)
   \end{aligned}$$

2. Proof that $\operatorname{supp}\mathsf{N}_n(x) = [0 : n+1]$ (proof by induction):

   (a) Base case ...proof that $\operatorname{supp}\mathsf{N}_0 = [0 : 1]$:

   $$\begin{aligned}
   \operatorname{supp}\mathsf{N}_0 &\triangleq \operatorname{supp}\mathbb{1}_{[0:1)} && \text{by definition of } \mathsf{N}_0(x) && \text{(Definition W.3 page 444)}\\
   &= \{[0 : 1)\}^- && \text{by definition of } \textit{support} \text{ operator}\\
   &= [0 : 1] && \text{by definition of } \textit{closure} \text{ operator}
   \end{aligned}$$

   (b) Induction step ...proof that $\{\operatorname{supp}\mathsf{N}_n = [0 : n+1]\} \implies \{\operatorname{supp}\mathsf{N}_{n+1} = [0 : n+2]\}$:

   $$\begin{aligned}
   \operatorname{supp}\mathsf{N}_{n+1}(x) &= \operatorname{supp}\int_{\tau=0}^{\tau=1} \mathsf{N}_n(x-\tau)\,\mathrm{d}\tau && \text{by Lemma W.2 page 444}\\
   &= \operatorname{supp}\int_{[0:1]} \mathsf{N}_n(x-\tau)\,\mathrm{d}\tau && \text{by def. of } \textit{Lebesgue integration}\\
   &= \left\{x \in \mathbb{R} | (x-\tau) \in [0 : n+1] \text{ for some } \tau \in [0 : 1]\right\}^- && \text{by induction hypothesis}\\
   &= [0 : n+1] \cup [0+1 : n+1+1]^-\\
   &= [0 : n+2]^-\\
   &= [0 : n+2] && \text{by property of } \textit{closure} \text{ operator}
   \end{aligned}$$

3. Proof that $\int_{\mathbb{R}} \mathsf{N}_n(x)\,\mathrm{d}x = 1$ (proof by induction):

[16] Christensen (2008) page 140 ⟨Theorem 6.1.1⟩, Höllig (2003) page 27 ⟨3.4⟩, Schumaker (2007) page 120 ⟨Theorem 4.15⟩, de Boor (2001) page 90 ⟨B-Spline Property (i)⟩, Chui (1988) page 2 ⟨Theorem 1.1⟩, Wojtaszczyk (1997) page 53 ⟨Theorem 3.7⟩, Cox (1972), de Boor (1972), Prasad and Iyengar (1997) pages 145–146 ⟨Theorem 6.2⟩

(a) Base case ...proof that $\int_{\mathbb{R}} \mathsf{N}_0(x) = 1$:

$$\begin{aligned}
\int_{\mathbb{R}} \mathsf{N}_0(x)\,\mathrm{d}x &= \int_{\mathbb{R}} \mathbb{1}_{[0:1)}\,\mathrm{d}x && \text{by definition of } \mathsf{N}_0(x) && \text{(Definition W.3 page 444)}\\
&= \int_{[0:1)} 1\,\mathrm{d}x && \text{by definition of } \mathbb{1}_A(x) && \text{(Definition W.1 page 443)}\\
&= \int_{[0:1]} 1\,\mathrm{d}x && \text{by property of } \textit{Lebesgue integration}\\
&= 1
\end{aligned}$$

(b) Induction step ...proof that $\left\{\int_{\mathbb{R}} \mathsf{N}_n(x) = 1\right\} \implies \left\{\int_{\mathbb{R}} \mathsf{N}_{n+1} = 1\right\}$:

$$\begin{aligned}
\int_{\mathbb{R}} \mathsf{N}_{n+1}(x)\,\mathrm{d}x &= \int_{\mathbb{R}}\int_0^1 \mathsf{N}_n(x-\tau)\,\mathrm{d}\tau\,\mathrm{d}x && \text{by Lemma W.2 page 444}\\
&= \int_0^1\int_{\mathbb{R}} \mathsf{N}_n(x-\tau)\,\mathrm{d}x\,\mathrm{d}\tau\\
&= \int_0^1\int_{\mathbb{R}} \mathsf{N}_n(u)\,\mathrm{d}u\,\mathrm{d}\tau && \text{where } u \triangleq x-\tau \implies \tau = x-u\\
&= \int_0^1 1\,\mathrm{d}\tau && \text{by induction hypothesis}\\
&= 1
\end{aligned}$$

4. Proof that $\mathsf{N}_n(x)$ is *symmetric* for $n \in \{1, 2, 3, \ldots\}$:

(a) Note that $\mathsf{N}_0(x)$ $(n = 0)$ is *not symmetric* (in particular it fails at $x = 1/2$) because $\mathsf{N}_0\left(\frac{0+1}{2} - \frac{1}{2}\right) = \mathsf{N}_0(0) = 1 \neq 0 = \mathsf{N}_1(1) = \mathsf{N}_0\left(\frac{0+1}{2} + \frac{1}{2}\right)$

(b) Base case ...proof for $n = 1$ case:

$$\begin{aligned}
\mathsf{N}_1\left(\frac{1+1}{2} - x\right) &= \mathsf{N}_1(1-x)\\
&= \left\{\begin{array}{rl} (1-x) & \text{for } 1-x \in [0:1]\\ -(1-x)+2 & \text{for } 1-x \in [1:2]\\ 0 & \text{otherwise}\end{array}\right\} && \text{by Example W.2 page 446}\\
&= \left\{\begin{array}{rl} -x+1 & \text{for } -x \in [-1:0]\\ x+1 & \text{for } -x \in [0:1]\\ 0 & \text{otherwise}\end{array}\right\}\\
&= \left\{\begin{array}{rl} x+1 & \text{for } x \in [-1:0]\\ -x+1 & \text{for } x \in [0:1]\\ 0 & \text{otherwise}\end{array}\right\}\\
&= \left\{\begin{array}{rl} (1+x) & \text{for } 1+x \in [0:1]\\ -(1+x)+2 & \text{for } 1+x \in [1:2]\\ 0 & \text{otherwise}\end{array}\right\}\\
&= \mathsf{N}_1(1+x) && \text{by Example W.2 page 446}\\
&= \mathsf{N}_1\left(\frac{1+1}{2} + x\right)
\end{aligned}$$

(c) Induction step …proof that $n-1$ case $\implies$ n case:

$$\begin{aligned}
&\mathsf{N}_n\left(\frac{n+1}{2}+x\right)\\
&= \frac{\frac{n+1}{2}+x}{n}\mathsf{N}_{n-1}\left(\frac{n+1}{2}+x\right) + \frac{n+1-\left(\frac{n+1}{2}+x\right)}{n}\mathsf{N}_{n-1}\left(\frac{n+1}{2}+x-1\right) && \text{by Theorem W.3 page 452}\\
&= \frac{n+1-\left(\frac{n+1}{2}-x\right)}{n}\mathsf{N}_{n-1}\left(\frac{n}{2}+\left[x+\frac{1}{2}\right]\right) + \frac{\frac{n+1}{2}-x}{n}\mathsf{N}_{n-1}\left(\frac{n}{2}+\left[x-\frac{1}{2}\right]\right)\\
&= \frac{n+1-\left(\frac{n+1}{2}-x\right)}{n}\mathsf{N}_{n-1}\left(\frac{n}{2}-\left[x+\frac{1}{2}\right]\right) + \frac{\frac{n+1}{2}-x}{n}\mathsf{N}_{n-1}\left(\frac{n}{2}-\left[x-\frac{1}{2}\right]\right) && \text{by induction hypothesis}\\
&= \frac{n+1-\left(\frac{n+1}{2}-x\right)}{n}\mathsf{N}_{n-1}\left(\left[\frac{n+1}{2}-x\right]-1\right) + \frac{\frac{n+1}{2}-x}{n}\mathsf{N}_{n-1}\left(\frac{n+1}{2}-x\right)\\
&= \mathsf{N}_n\left(\frac{n+1}{2}-x\right) && \text{by Theorem W.3 page 452}
\end{aligned}$$

✏

W.3 Projection properties

In the case where $\left(\!\left(\mathsf{N}_n(x-k)\right)\!\right)_{k\in\mathbb{Z}}$ is to be used as a basis in some subspace of $\boldsymbol{L}^2_{\mathbb{R}}$, one may want to *project* a function $\mathsf{f}(x)$ onto a basis function $\mathsf{N}_n(x-k)$. This is especially true when $\left(\!\left(\mathsf{N}_n(x-k)\right)\!\right)$ is *orthogonal*; but in the case of *B-splines* this is only true when $n=0$ (Theorem W.8 page 466). Nevertheless, projection of a function onto $\mathsf{N}_n(x-k)$, or the projection of $\mathsf{N}_n(x)$ onto another basis fucntion (such as the complex exponential in the case of *Fourier analysis* as in Lemma W.5 page 458), is still useful. Projection in an *inner product space* is typically performed using the *inner product* $\langle \mathsf{f}(x) \,|\, \mathsf{N}_n(x-k)\rangle$; and in the space $\boldsymbol{L}^2_{\mathbb{R}}$, this inner product is typically defined as an *integral* such that

$$\langle \mathsf{f}(x) \,|\, \mathsf{N}_n(x-k)\rangle \triangleq \int_{\mathbb{R}} \mathsf{f}(x)\mathsf{N}_n(x-k)\,\mathrm{d}x.$$

As it turns out, there is a way to compute this inner product that only involves the function $\mathsf{f}(x)$ and the order parameter n (next theorem).

Theorem W.5. [17] *Let* f *be a* CONTINUOUS *function in* $\boldsymbol{L}^2_{\mathbb{R}}$ *and* $\mathsf{f}^{(n)}$ *the* nTH DERIVATIVE *of* $\mathsf{f}(x)$.

THM

(1). $$\int_{\mathbb{R}} \mathsf{f}(x)\mathsf{N}_n(x)\,\mathrm{d}x = \int_{[0:1)^{n+1}} \mathsf{f}(x_1+x_2+\cdots+x_{n+1})\,\mathrm{d}x_1\,\mathrm{d}x_2\cdots\mathrm{d}x_{n+1}$$

(2). $$\int_{\mathbb{R}} \mathsf{f}^{(n)}(x)\mathsf{N}_n(x)\,\mathrm{d}x = \sum_{k=0}^{n+1}(-1)^{n-k+1}\binom{n+1}{k}\mathsf{f}(k)$$

✎PROOF:

1. Proof for (1) (proof by induction):

   (a) Base case …proof for $n=0$ case:

$$\int_{\mathbb{R}} \mathsf{f}(x)\mathsf{N}_0(x)\,\mathrm{d}x = \int_{[0:1)} \mathsf{f}(x)\,\mathrm{d}x \qquad \text{by definition of } \mathsf{N}_0(x) \qquad \text{(Definition W.3 page 444)}$$

[17] Chui (1992) page 85 ⟨(4.2.2), (4.2.3)⟩, Christensen (2008) page 140 ⟨Theorem 6.1.1⟩

(b) Inductive step—proof that n case $\implies n+1$ case:

$$
\begin{aligned}
&\int_{\mathbb{R}} \mathsf{f}(x)\mathsf{N}_{n+1}(x)\,\mathrm{d}x \\
&= \int_{\mathbb{R}} \left[\int_0^1 \mathsf{N}_n(x-\tau)\,\mathrm{d}\tau\right]\mathsf{f}(x)\,\mathrm{d}x && \text{by Lemma W.2 page 444} \\
&= \int_{[0:1)} \int_{\mathbb{R}} \mathsf{N}_n(x-\tau)\mathsf{f}(x)\,\mathrm{d}x\,\mathrm{d}\tau \\
&= \int_{[0:1)} \int_{\mathbb{R}} \mathsf{N}_n(u)\mathsf{f}(u+\tau)\,\mathrm{d}u\,\mathrm{d}\tau && \text{where } u \triangleq x-\tau \implies x = u+\tau \\
&= \int_{[0:1)} \int_{[0:1)^{n+1}} \mathsf{f}(u_1+u_2+\cdots+u_{n+1}+\tau)\,\mathrm{d}u_1\,\mathrm{d}u_2\cdots\mathrm{d}u_{n+1}\,\mathrm{d}\tau && \text{by induction hypothesis} \\
&= \int_{[0:1)^{n+2}} \mathsf{f}(u_1+u_2+\cdots+u_{n+1}+u_{n+2})\,\mathrm{d}u_1\,\mathrm{d}u_2\cdots\mathrm{d}u_{n+2}\,\mathrm{d}\tau \\
&= \int_{[0:1)^{n+2}} \mathsf{f}(x_1+x_2+\cdots+x_{n+1}+x_{n+2})\,\mathrm{d}x_1\,\mathrm{d}x_2\cdots\mathrm{d}x_{n+2} && \text{by change of variables } u_k \to x_k
\end{aligned}
$$

2. Proof for (2):

$$
\begin{aligned}
\int_{\mathbb{R}} \mathsf{f}^{(n)}(x)\mathsf{N}_n(x)\,\mathrm{d}x &= \int_{[0:1)^{n+1}} \mathsf{f}^{(n)}\left(\sum_{k=1}^{n+1} x_k\right)\mathrm{d}x_1\,\mathrm{d}x_2\cdots\mathrm{d}x_{n+1} && \text{by (1)} \\
&= \sum_{k=0}^{n+1}(-1)^{n-k+1}\binom{n+1}{k}\mathsf{f}(k) && \text{by Theorem H.1 page 224}
\end{aligned}
$$

✎

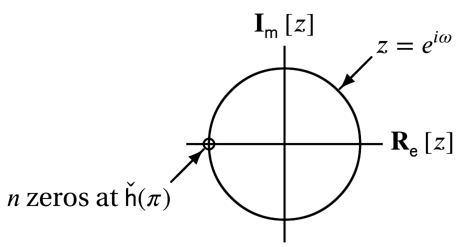


Figure W.2: Zero locations for B-cardinal spline $N_n(x)$ scaling coefficients

W.4 Fourier analysis

Simply put, no matter what new and fancy basis sequences are discovered, the *Fourier transform* never goes out of style. This is largely because the *kernel* of the Fourier transform—the *complex exponential* function—has two properties that makes it extremely special:

- The complex exponential is an eigenvalue of any *linear time invariant* (*LTI*) operator (Theorem I.13 page 244).
- The complex exponential generates a *continuous point spectrum* for the *differential operator*.

Thus, we might expect the projection of the *B-spline* function $\mathsf{N}_n(x)$ onto the complex exponential (essentially the *Fourier transform* of $\mathsf{N}_n(x)$,…next lemma) to be useful. Such a hunch would be confirmed because it is useful for proving that

- the sequence $\langle\!\langle \mathsf{N}_n(x-k)\rangle\!\rangle_{k\in\mathbb{Z}}$ is a *Riesz basis* (Lemma W.6 page 461, Theorem W.8 page 466) and
- the sequence $\langle\!\langle \mathsf{N}_n(x-k)\rangle\!\rangle_{k\in\mathbb{Z}}$ is a *multiresolution analysis* (Theorem W.10 page 469).

Lemma W.5. [18] *Let* $\tilde{\mathbf{F}}$ *be the* FOURIER TRANSFORM *operator* (*Definition T.2 page 408*).

$$\tilde{\mathbf{F}}\mathsf{N}_n(\omega) = \frac{1}{\sqrt{2\pi}}\left(\frac{1-e^{-i\omega}}{i\omega}\right)^{n+1} = \frac{1}{\sqrt{2\pi}}e^{-i\frac{(n+1)\omega}{2}}\left(\frac{\sin(\omega/2)}{\omega/2}\right)^{n+1} \triangleq \frac{1}{\sqrt{2\pi}}e^{-i\frac{(n+1)\omega}{2}}\left(\operatorname{sinc}\frac{\omega}{2}\right)^{n+1}$$

✎PROOF:

1. Proof using Theorem W.5 page 456:

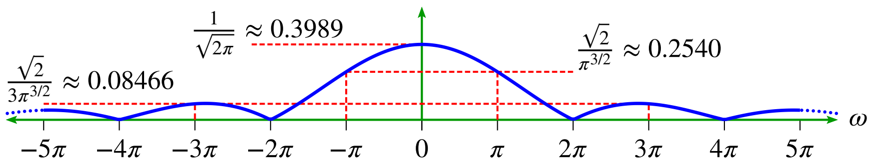
$$\begin{aligned}
\tilde{\mathbf{F}}\mathsf{N}_n(\omega) &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}\mathsf{N}_n(x)e^{-i\omega x}\,\mathrm{d}x && \text{by definition of } \tilde{\mathbf{F}} \quad \text{(Definition T.2 page 408)}\\
&= \frac{1}{\sqrt{2\pi}}\int_{[0:1)^{n+1}} e^{-i\omega(x_1+x_2+\cdots+x_{n+1})}\,\mathrm{d}x_1\ \mathrm{d}x_2\ \cdots,\ \mathrm{d}x_{n+1} && \text{by Theorem W.5}\\
&= \frac{1}{\sqrt{2\pi}}\prod_{k=1}^{n+1}\left(\int_{[0:1)} e^{-i\omega x_k}\,\mathrm{d}x_k\right) && \text{because } e^{x+y}=e^xe^y\\
&= \frac{1}{\sqrt{2\pi}}\left(\int_0^1 e^{-i\omega x}\,\mathrm{d}x\right)^{n+1} = \frac{1}{\sqrt{2\pi}}\left(\left.\frac{e^{-i\omega x}}{-i\omega}\right|_0^1\right)^{n+1}\\
&= \frac{1}{\sqrt{2\pi}}\left(\frac{1-e^{-i\omega}}{i\omega}\right)^{n+1} = \frac{1}{\sqrt{2\pi}}\left[e^{-i\frac{\omega}{2}}\left(\frac{e^{i\frac{\omega}{2}}-e^{-i\frac{\omega}{2}}}{i\omega}\right)\right]^{n+1}\\
&= \frac{1}{\sqrt{2\pi}}\left[e^{-i\frac{\omega}{2}}\left(\frac{2i\sin\left(\frac{\omega}{2}\right)}{\frac{2i\omega}{2}}\right)\right]^{n+1} && \text{by \textit{Euler formulas}} \quad \text{(Corollary I.2 page 235)}\\
&= \frac{1}{\sqrt{2\pi}}e^{-i\frac{(n+1)\omega}{2}}\left(\frac{\sin(\omega/2)}{\omega/2}\right)^{n+1}
\end{aligned}$$

2. Proof using *rectangular pulse* example (Example T.1 page 415) and *Convolution Theorem* (Theorem V.2 page 430):

$$\begin{aligned}
\tilde{\mathbf{F}}\mathsf{N}_n(\omega) &= \left[\sqrt{2\pi}\right]^n\left[\tilde{\mathbf{F}}\mathsf{N}_0\right]^{n+1} && \text{by \textit{Convolution Theorem}} \quad \text{(Theorem V.2 page 430)}\\
&= \frac{1}{\sqrt{2\pi}}\left[\sqrt{2\pi}\frac{c(b-a)}{\sqrt{2\pi}}e^{-i\left(\frac{a+b}{2}\omega\right)}\left(\frac{\sin\left(\frac{b-a}{2}\omega\right)}{\left(\frac{b-a}{2}\omega\right)}\right)\right]^{n+1} && \text{by \textit{rectangular pulse} example with } a=0,\ b=c=1 \quad \text{(Example T.1 page 415)}\\
&= \frac{1}{\sqrt{2\pi}}\left[\sqrt{2\pi}\frac{1}{\sqrt{2\pi}}e^{-i\left(\frac{\omega}{2}\right)}\left(\frac{\sin\left(\frac{\omega}{2}\right)}{(\omega/2)}\right)\right]^{n+1}\\
&= \frac{1}{\sqrt{2\pi}}e^{-i\left(\frac{(n+1)\omega}{2}\right)}\left(\frac{\sin\left(\frac{\omega}{2}\right)}{(\omega/2)}\right)^{n+1}
\end{aligned}$$

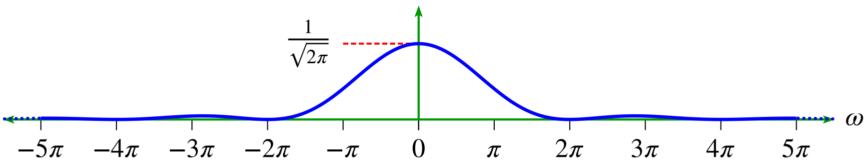
⇰

Example W.10. The *Fourier transform magnitude* $\left|\left[\tilde{\mathbf{F}}\mathsf{N}_0\right](\omega)\right|$ of the *0 order B-spline* $\mathsf{N}_0(x)$ is illustrated to the right.

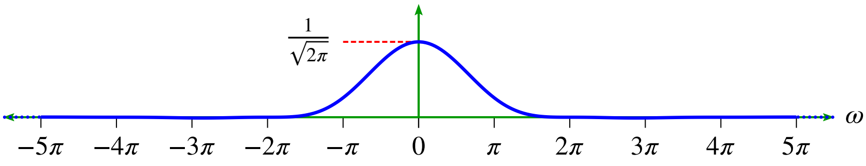


[18] Christensen (2008) page 142 ⟨Corollary 6.1.2⟩

Example W.11. The *Fourier transform magnitude* $\left|\left[\tilde{\mathbf{F}}\mathsf{N}_1\right](\omega)\right|$ of the *1st order B-spline* $\mathsf{N}_1(x)$ (Example W.2 page 446) is illustrated to the right.



Example W.12. The *Fourier transform magnitude* $\left|\left[\tilde{\mathbf{F}}\mathsf{N}_2\right](\omega)\right|$ of the *2nd order B-spline* $\mathsf{N}_2(x)$ is illustrated to the right.



W.5 Basis properties

W.5.1 Uniqueness properties

Coefficients of a *basis sequence* are not always *unique*. Take for example a very trivial sequence (α_1, α_2) in which the coefficients are summed. If $\mathsf{f}(x) \triangleq \alpha_1 + \alpha_2$ and $\mathsf{g}(x) \triangleq \beta_1 + \beta_2$,

$$\begin{array}{lll} \text{then} & \left\{(\alpha_1,\alpha_2) = (\beta_1,\beta_2)\right\} \implies & \mathsf{f}(x) = \mathsf{g}(x) \\ \text{but} & \mathsf{f}(x) = \mathsf{g}(x) \quad \not\Longrightarrow & \left\{(\alpha_1,\alpha_2) = (\beta_1,\beta_2)\right\}, \end{array}$$

because for example if $(\alpha_1, \alpha_2) = (1, 2)$ and $(\beta_1, \beta_2) = (-6, 9)$, then $\mathsf{f}(x) = \mathsf{g}(x)$, but $(\alpha_1, \alpha_2) \neq (\beta_1, \beta_2)$. This example demonstrates that the "if and only if" condition $\iff$ does not hold and coefficients are not unique in all *basis sequence*s. But arguably a minimal requirement for any practical basis sequence is that the coefficients are *unique* (the "if and only if" condition $\iff$ holds). And indeed, in a *B-spline* basis sequence $\left(\!\left(\mathsf{N}_n(x-k)\right)\!\right)_{k\in\mathbb{Z}}$, the coefficients $\left(\!\left(\alpha_k\right)\!\right)_{k\in\mathbb{Z}}$ are *unique*, as demonstrated by Theorem W.6 (next).

Theorem W.6. [19] *Let* $\mathsf{N}_n(x)$ *be the* nTH-ORDER B-SPLINE *(Definition W.3 page 444). Let*

$$\mathsf{f}(x) \triangleq \sum_{k\in\mathbb{Z}} \alpha_k \mathsf{N}_n(x-k) \quad \text{and} \quad \mathsf{g}(x) \triangleq \sum_{k\in\mathbb{Z}} \beta_k \mathsf{N}_n(x-k).$$

THM

$$\underbrace{\left\{\ \mathsf{f}(x) = \mathsf{g}(x) \quad \forall x \in \mathbb{R}\ \right\} \iff \left\{\left(\!\left(\alpha_k\right)\!\right)_{k\in\mathbb{Z}} = \left(\!\left(\beta_k\right)\!\right)_{k\in\mathbb{Z}}\right\}}_{\text{coefficients are UNIQUE}}$$

✎PROOF:

1. Proof that $\Longleftarrow$ condition holds:

$$\begin{aligned} \mathsf{f}(x) &\triangleq \sum_{k\in\mathbb{Z}} \alpha_k \mathsf{N}_n(x-k) && \text{by definition of } \mathsf{f}(x) \\ &= \sum_{k\in\mathbb{Z}} \beta_k \mathsf{N}_n(x-k) && \text{by right hypothesis} \\ &\triangleq \mathsf{g}(x) && \text{by definition of } \mathsf{g}(x) \end{aligned}$$

2. Proof that $\Longrightarrow$ condition holds (proof by contradiction):

   (a) Suppose it does *not* hold.

[19] Wojtaszczyk (1997) page 55 ⟨Theorem 3.11⟩

(b) Then there exists sequences $(\!(\alpha_k)\!)_{k\in\mathbb{Z}}$ and $(\!(\beta_k)\!)_{k\in\mathbb{Z}}$ such that
$(\!(\alpha_k)\!) - (\!(\beta_k)\!) \triangleq (\!(alpha_k - \beta_k)\!) \neq (\!(0, 0, 0, \cdots)\!)$
but also such that $\mathsf{f}(x) - \mathsf{g}(x) = 0 \; \forall x \in \mathbb{R}$.

(c) If this were possible, then

$$\begin{aligned}
0 &= \mathsf{f}(x) - \mathsf{g}(x) \\
&= \sum_{m\in\mathbb{Z}} \alpha_m \mathsf{N}_n(x-m) - \sum_{m\in\mathbb{Z}} \beta_m \mathsf{N}_n(x-m) \\
&= \sum_{m\in\mathbb{Z}} (\alpha_m - \beta_m)\mathsf{N}_n(x-m) \\
&= \sum_{m=0}^{m=n} (\alpha_m - \beta_m) \frac{1}{n!}\left[\sum_{k=0}^{n+1} (-1)^k \binom{n+1}{k} (x-k)^n \sigma(x-k)\right] && \text{by Theorem W.1 page 447}
\end{aligned}$$

(d) But this is *impossible* because $\mathsf{N}(x)$ is *non-negative* (Theorem W.4 page 454).

(e) Therefore, there is a contradiction, and the $\Longrightarrow$ condition *does* hold.

✏

W.5.2 Partition of unity properties

In the case in which a sequence of *B-splines* $(\!(\mathsf{N}_n(x-k))\!)_{k\in\mathbb{Z}}$ is to be used as a *basis* for some subspace of $\boldsymbol{L}^2_{\mathbb{R}}$, arguably one of the most important properties for the sequence to have is the *partition of unity* property such that $\sum_{k\in\mathbb{Z}} \mathsf{N}_n(x-k) = 1$. This allows for convenient representation of the most basic functions, such as constants.[20] As it turns out, B-splines *do* have this property (next theorem).

Theorem W.7 (*B-spline partition of unity*). [21] *Let $\mathsf{N}_n(x)$ be the nTH ORDER B-SPLINE (Definition W.3 page 444).*

THM	
	$\displaystyle\sum_{k\in\mathbb{Z}} \mathsf{N}_n(x-k) \;=\; 1 \quad \forall n \in \mathbb{W}$ (PARTITION OF UNITY)

✎PROOF:

1. lemma: $\sum_{k\in\mathbb{Z}} \mathsf{N}_0(x-k) = 1$. Proof:

$$\begin{aligned}
\sum_{k\in\mathbb{Z}} \mathsf{N}_0(x-k) &= \sum_{k\in\mathbb{Z}} \mathbb{1}_{[0:1)}(x-k) && \text{by definition of } \mathsf{N}_0(x) && \text{(Definition W.3 page 444)} \\
&= 1 && \text{by definition of } \mathbb{1}_A(x) && \text{(Definition W.1 page 443)}
\end{aligned}$$

2. Proof for this theorem follows from the $n = 0$ case ((1) lemma page 460), the definition of $\mathsf{N}_n(x)$ (Definition W.3 page 444), and Corollary ?? (page ??).

3. Alternatively, this theorem can be proved by *induction*:

   (a) Base case ($n = 0$ case): by (1) lemma.

[20] Jawerth and Sweldens (1994) page 8

[21] Christensen (2008) page 140 ⟨Theorem 6.1.1⟩, Höllig (2003) page 27 ⟨3.4⟩, Schumaker (2007) page 120 ⟨Theorem 4.15⟩, de Boor (2001) page 90 ⟨B-Spline Property (i)⟩, Chui (1988) page 2 ⟨Theorem 1.1⟩, Wojtaszczyk (1997) page 53 ⟨Theorem 3.7⟩, Cox (1972), de Boor (1972)

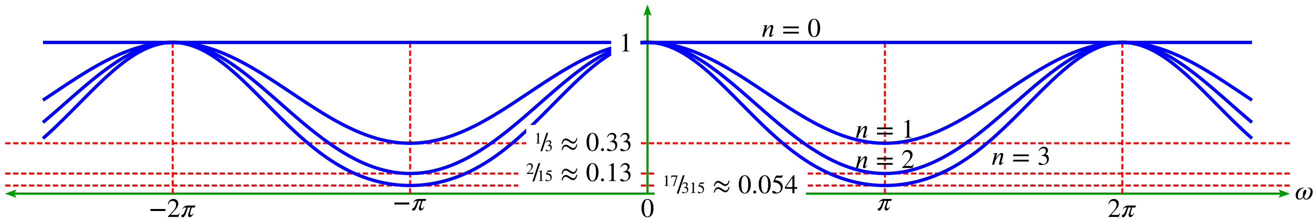


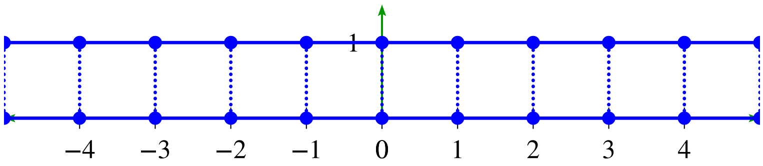
Figure W.3: *auto-power spectrum* $\tilde{S}_n(\omega)$ plots of *B-splines* $\mathsf{N}_n(x)$ (Lemma W.6 page 461) For C and LaTeX source code to generate such a plot, see Section **??** (page **??**).

(b) Inductive step—proof that $\sum_{k\in\mathbb{Z}} \mathsf{N}_n(x-k) = 1 \implies \sum_{k\in\mathbb{Z}} \mathsf{N}_{n+1}(x-k) = 1$:

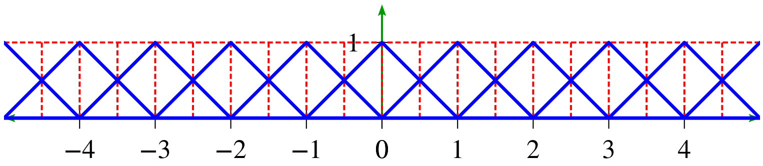
$$\begin{aligned}
\sum_{k\in\mathbb{Z}} \mathsf{N}_{n+1}(x-k) &= \sum_{k\in\mathbb{Z}} \int_{\tau=0}^{\tau=1} \mathsf{N}_n(x-k-\tau)\,\mathrm{d}\tau && \text{by Lemma W.2 page 444} \\
&= \sum_{k\in\mathbb{Z}} \int_{x-u=0}^{x-u=1} \mathsf{N}_n(u-k)(-1)\,\mathrm{d}u && \text{where } u \triangleq x-\tau \implies \tau = x-u \\
&= \sum_{k\in\mathbb{Z}} \int_{u=x-1}^{u=x} \mathsf{N}_n(u-k)\,\mathrm{d}u \\
&= \int_{u=x-1}^{u=x} \left(\sum_{k\in\mathbb{Z}} \mathsf{N}_n(u-k)\right) \mathrm{d}u \\
&= \int_{u=x-1}^{u=x} 1\,\mathrm{d}u && \text{by induction hypothesis} \\
&= 1
\end{aligned}$$

✏

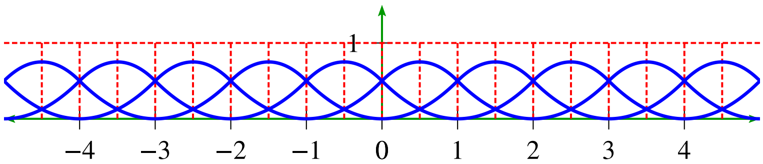
Example W.13. The *partition of unity* property for the 0 *order B-spline* $\mathsf{N}_0(x)$ (Example W.1 page 445) is illustrated to the right.



Example W.14. The *partition of unity* property for the 1*st order B-spline* $\mathsf{N}_1(x)$ (Example W.2 page 446) is illustrated to the right.



Example W.15. The *partition of unity* property for the 2*nd order B-spline* $\mathsf{N}_2(x)$ (Example W.3 page 446) is illustrated to the right.



W.5.3 Riesz basis properties

Lemma W.6. *Let* $\mathsf{N}_n(x)$ *be the* nTH ORDER B-SPLINE *(Definition W.3 page 444).*
Let $\tilde{S}_n(\omega) \triangleq 2\pi \sum_{k\in\mathbb{Z}} \left|\tilde{\mathbf{F}}\mathsf{N}_n(\omega - 2\pi k)\right|^2$ *be the* AUTO-POWER SPECTRUM *(Definition **??** page **??**) of* $\mathsf{N}_n(x)$.

LEM

(1). $0 < \tilde{S}_n(\omega) \leq 1 \quad \forall \omega \in \mathbb{R} \quad , \quad \forall n \in \mathbb{W}$	(3). $\tilde{S}_n(0) = 1 \quad \forall n \in \mathbb{W}$
(2). $\tilde{S}_n(\omega) = 1 \quad \forall \omega \in \mathbb{R} \quad , \quad \textit{for } n = 0$	(4). $\tilde{S}_n(\pi) \leq \frac{1}{3} \quad \forall n \in \mathbb{W} \backslash \{0\}$

(Note: see illustration in Figure W.3 page 461.)

✎PROOF:

1. lemma: $\tilde{S}_n(\omega) = \sum_{k\in\mathbb{Z}} \left[\frac{\sin\left(\frac{\omega}{2}\right)}{k\pi - \frac{\omega}{2}}\right]^{2(n+1)}$. Proof:

$$\begin{aligned}
\tilde{S}_n(\omega) &\triangleq 2\pi \sum_{k\in\mathbb{Z}} \left|\tilde{\mathbf{F}}\mathsf{N}_n(\omega - 2\pi k)\right|^2 && \text{by Definition ?? page ??} \\
&= 2\pi \sum_{k\in\mathbb{Z}} \left| \frac{1}{\sqrt{2\pi}} e^{-i\frac{(n+1)(\omega-2\pi k)}{2}} \left(\frac{\sin\left(\frac{\omega-2\pi k}{2}\right)}{\frac{\omega-2\pi k}{2}}\right)^{n+1} \right|^2 && \text{by Lemma W.5 page 458} \\
&= \sum_{k\in\mathbb{Z}} \left[\frac{\sin\left(\frac{\omega-2\pi k}{2}\right)}{\frac{\omega-2\pi k}{2}}\right]^{2(n+1)} \\
&= \sum_{k\in\mathbb{Z}} \left[\frac{\sin\left(\frac{\omega}{2} - k\pi\right)}{k\pi - \frac{\omega}{2}}\right]^{2(n+1)} \\
&= \sum_{k\in\mathbb{Z}} \left[\frac{(-1)^k \sin\left(\frac{\omega}{2}\right)}{k\pi - \frac{\omega}{2}}\right]^{2(n+1)} \\
&= \sum_{k\in\mathbb{Z}} \left[\frac{\sin\left(\frac{\omega}{2}\right)}{k\pi - \frac{\omega}{2}}\right]^{2(n+1)}
\end{aligned}$$

2. lemma (one sided series form):

$$\begin{aligned}
\tilde{S}_n(\omega) &= \sum_{k\in\mathbb{Z}} \left[\frac{\sin\left(\frac{\omega}{2}\right)}{k\pi - \frac{\omega}{2}}\right]^{2(n+1)} && \text{by (1) lemma} \\
&= \left[\frac{\sin(\omega/2)}{(\omega/2)}\right]^{2(n+1)} + \left[\frac{\sin(\omega/2)}{(\pi/2)}\right]^{2(n+1)} \left(\sum_{k=1}^{\infty}\left[\frac{1}{2k - \frac{\omega}{\pi}}\right]^{2(n+1)} + \sum_{k=1}^{\infty}\left[\frac{1}{2k + \frac{\omega}{\pi}}\right]^{2(n+1)}\right)
\end{aligned}$$

3. lemma: $\tilde{S}_n(\omega)$ is *continuous* for all $\omega \in \mathbb{R}$.
Proof: $\sin(\omega/2)$ and $\omega/2$ are *continuous*, so $\tilde{S}_n(\omega)$ is *continuous* as well.

4. lemma: $\tilde{S}_n(\omega)$ is *periodic* with period 2π (and so we only need to examine $\tilde{S}_n(\omega)$ for $\omega \in [0 : 2\pi)$). Proof of *periodicity*: This follows directly from Proposition ?? (page ??).

5. lemma: $\tilde{S}_n(-\omega) = \tilde{S}_n(\omega)$ (*symmetric* about 0) and $\tilde{S}_n(\pi - \omega) = \tilde{S}_n(\pi + \omega)$ (*symmetric* about π). Proof: This follows directly from Proposition ?? (page ??).

6. Proof that $\tilde{S}_n(0) = 1$:

$$\begin{aligned}
\tilde{S}_n(0) &= \lim_{\omega\to 0}\tilde{S}_n(\omega) && \text{by (3) lemma}\\
&= \lim_{\omega\to 0}\left[\left[\frac{\sin(\omega/2)}{(\omega/2)}\right]^{2(n+1)} + \left[\frac{\sin(\omega/2)}{(\pi/2)}\right]^{2(n+1)}\left(\sum_{k=1}^{\infty}\left[\frac{1}{2k-\frac{\omega}{\pi}}\right]^{2(n+1)} + \sum_{k=1}^{\infty}\left[\frac{1}{2k+\frac{\omega}{\pi}}\right]^{2(n+1)}\right)\right] && \text{by (2) lemma}\\
&= \lim_{\omega\to 0}\left[\frac{\cos\left(\frac{\omega}{2}\right)}{-\frac{1}{2}}\right]^{2(n+1)} + 0 \qquad \text{by \textit{l'Hôpital's rule}}\\
&= (-1)^{2(n+1)} = 1
\end{aligned}$$

7. Proof that $\tilde{S}_n(\pi)$ converges to some value > 0:

(a) Proof that $\tilde{S}_n(\pi) > 0$:

$$\begin{aligned}
\tilde{S}_n(\pi) &= \left[\frac{\sin(\pi/2)}{(\pi/2)}\right]^{2(n+1)} + \left[\frac{\sin(\pi/2)}{(\pi/2)}\right]^{2(n+1)}\left(\sum_{k=1}^{\infty}\left[\frac{1}{2k-\frac{\pi}{\pi}}\right]^{2(n+1)} + \sum_{k=1}^{\infty}\left[\frac{1}{2k+\frac{\pi}{\pi}}\right]^{2(n+1)}\right) && \text{by (2) lemma}\\
&= \left(\frac{2}{\pi}\right)^{2(n+1)}\left[1 + \left(\frac{1}{1}\right)^{2(n+1)} + \left(\frac{1}{3}\right)^{2(n+1)} + \left(\frac{1}{3}\right)^{2(n+1)} + \left(\frac{1}{5}\right)^{2(n+1)} + \left(\frac{1}{5}\right)^{2(n+1)} + \cdots\right]\\
&= 2\left(\frac{2}{\pi}\right)^{2(n+1)} \underbrace{\sum_{k=1}^{\infty}\left[\frac{1}{2k-1}\right]^{2(n+1)}}_{\textit{Dirichlet Lambda function } \lambda(2n+2)}\\
&> 0 \qquad \text{because } x^2 > 0 \text{ for all } x \in \mathbb{R}\backslash\{0\}
\end{aligned}$$

(b) Proof that $\tilde{S}_n(\pi)$ *converges*:

$$\begin{aligned}
\tilde{S}_n(\pi) &= 2\left(\frac{2}{\pi}\right)^{2(n+1)}\sum_{k=1}^{\infty}\left(\frac{1}{2k-1}\right)^{2(n+1)} && \text{by item (7a)}\\
&\le 2\left(\frac{2}{\pi}\right)^{2(n+1)}\sum_{k=1}^{\infty}\left(\frac{1}{k}\right)^{2(n+1)}\\
&\le 2\left(\frac{2}{\pi}\right)^{2(n+1)}\sum_{k=1}^{\infty}\left(\frac{1}{k}\right)^{2}\\
&\implies \textit{convergence} && \text{by \textit{comparison test}}
\end{aligned}$$

(c) Tighter bounds for $\tilde{S}_n(\pi)$ for certain values of $n \in \{0, 1, 2, 3, 4\}$:

$$\begin{aligned}
\tilde{S}_n(\pi) &= 2\left(\frac{2}{\pi}\right)^{2(n+1)}\sum_{k=1}^{\infty}\left(\frac{1}{2k-1}\right)^{2(n+1)} && \text{by item (7a)}\\
&= 2\left(\frac{2}{\pi}\right)^{2(n+1)} U_{2(n+1)} && \text{by Jolley (1961) pages 56–57 ⟨(307)⟩}\\
&= 2\left(\frac{2}{\pi}\right)^{2(n+1)}\left[\frac{\pi^{2(n+1)}\alpha_{n+1}}{(4)[(2n+2)!]}\right] && \text{by Jolley (1961) pages 56–57 ⟨(307)⟩}\\
&= \frac{2^{2n+1}\alpha_{n+1}}{(2n+2)!}
\end{aligned}$$

$$= \begin{Bmatrix} \frac{2^1(1)}{2!} & \text{for } n=0 & (\alpha_1 = 1) \\ \frac{2^3(1)}{4!} & \text{for } n=1 & (\alpha_2 = 1) \\ \frac{2^5(3)}{6!} & \text{for } n=2 & (\alpha_3 = 3) \\ \frac{2^7(17)}{8!} & \text{for } n=3 & (\alpha_4 = 17) \\ \frac{2^9(155)}{10!} & \text{for } n=4 & (\alpha_5 = 155) \end{Bmatrix} \qquad \text{by Jolley (1961) page 234 } \langle(1130)\rangle$$

$$= \begin{Bmatrix} 1 & \text{for } n=0 \\ \frac{1}{3} & \text{for } n=1 \\ \frac{2}{15} & \text{for } n=2 \\ \frac{17}{315} & \text{for } n=3 \\ \frac{62}{2835} & \text{for } n=4 \end{Bmatrix} = \begin{Bmatrix} 1 & \text{for } n=0 \\ 0.3333333333333\cdots & \text{for } n=1 \\ 0.1333333333333\cdots & \text{for } n=2 \\ 0.0539682539682\cdots & \text{for } n=3 \\ 0.0218694885361\cdots & \text{for } n=4 \end{Bmatrix}$$

(d) Being important for the $n = 0$ case, note that[22]

$$\underbrace{\sum_{k=1}^{\infty}\left(\frac{1}{2k-1}\right)^2}_{\textit{Dirichlet Lambda function } \lambda(2)} = 1 + \frac{1}{3^2} + \frac{1}{5^2} + \frac{1}{7^2} + \cdots = \frac{\pi^2}{8}$$

(e) Proof that $\tilde{S}_n(\pi) \le \frac{1}{3}$: because $\tilde{S}_n(\pi) = \frac{1}{3}$ for $n = 1$ (item (7c) page 463) and because $\tilde{S}_n(\pi)$ is decreasing for increasing n.

8. lemma: $\tilde{S}_n(\omega)$ *converges* to some value $> 0\ \forall\omega \in \mathbb{R}$. Proof:

(a) For $\omega = 0$, $\tilde{S}_n(\omega) = 1$ by item (6).

(b) Proof that $\tilde{S}_n(\omega) > 0$ for $\omega \in (0 : 2\pi)$:

$$\begin{aligned} \tilde{S}_n(\omega) &= \left[\frac{\sin(\omega/2)}{(\omega/2)}\right]^{2(n+1)} + \left[\frac{\sin(\omega/2)}{(\pi/2)}\right]^{2(n+1)}\left(\sum_{k=1}^{\infty}\left[\frac{1}{2k-\frac{\omega}{\pi}}\right]^{2(n+1)} + \sum_{k=1}^{\infty}\left[\frac{1}{2k+\frac{\omega}{\pi}}\right]^{2(n+1)}\right) && \text{by (2) lemma} \\ &> 0 \end{aligned}$$

(c) Proof that $\tilde{S}_n(\omega)$ *converges*:

i. lemma: $\displaystyle\sum_{k=1}^{\infty}\left[\frac{1}{2k\pm\frac{\omega}{\pi}}\right]^{2(n+1)}$ *converges*. Proof:

$$\begin{aligned} \lim_{b\to\infty}\int_1^b \left[\frac{1}{2y\pm\frac{\omega}{\pi}}\right]^{2(n+1)} \mathrm{d}y &= \lim_{b\to\infty}\int_1^b \left[2y\pm\frac{\omega}{\pi}\right]^{-2n-2} \mathrm{d}y \\ &= \lim_{b\to\infty} \left.\frac{\left[2y\pm\frac{\omega}{\pi}\right]^{-2n-1}}{2(-2n-1)}\right|_1^b \\ &= \lim_{b\to\infty}\left(\frac{-1}{2(2n+1)}\right)\left[\frac{1}{\left[2b\pm\frac{\omega}{\pi}\right]^{2n+1}} - \frac{1}{\left[2\pm\frac{\omega}{\pi}\right]^{2n+1}}\right] \\ &= 0 + \frac{1}{2(2n+1)\left[2\pm\frac{\omega}{\pi}\right]^{2n+1}} \\ &< \infty && \forall\omega\in[0:2\pi) \end{aligned}$$

[22] Nahin (2011) page 153, Bailey et al. (2013a) page 334 ⟨Catalan's Constant⟩, Bailey et al. (2013b) page 849 ⟨Catalan's Constant⟩, Bailey et al. (2011) page 15 ⟨4.1 Catalan's constan⟩, Wells (1987) page 36 ⟨Dictionary entry for π: pages 31–37⟩, Heinbockel (2010) page 94 ⟨(2.27) *Dirichlet Lambda function*⟩

$$\implies \sum_{k=1}^{\infty}\left[\frac{1}{2k \pm \frac{\omega}{\pi}}\right]^{2(n+1)} \textit{ converges} \qquad \text{by } \textit{integral test}$$

ii. completion of proof using (8(c)i) lemma ...

$$\begin{aligned}
&\tilde{S}_n(\omega)\\
&= \left[\frac{\sin(\omega/2)}{(\omega/2)}\right]^{2(n+1)} + \left[\frac{\sin(\omega/2)}{(\pi/2)}\right]^{2(n+1)}\left(\sum_{k=1}^{\infty}\left[\frac{1}{2k-\frac{\omega}{\pi}}\right]^{2(n+1)} + \sum_{k=1}^{\infty}\left[\frac{1}{2k+\frac{\omega}{\pi}}\right]^{2(n+1)}\right) && \text{by (2) lemma}\\
&\implies \tilde{S}_n(\omega) \textit{ converges } \forall\omega\in(0:2\pi) && \text{by (8(c)i) lemma}
\end{aligned}$$

9. lemma (an expression for $\tilde{S}'_n(\omega)$):

$$\begin{aligned}
\tilde{S}'_n(\omega) &\triangleq \frac{\mathrm{d}}{\mathrm{d}\omega}\tilde{S}_n(\omega)\\
&= \frac{\mathrm{d}}{\mathrm{d}\omega}\sum_{k\in\mathbb{Z}}\left[\frac{\sin\left(\frac{\omega}{2}\right)}{k\pi-\frac{\omega}{2}}\right]^{2(n+1)} && \text{by (1) lemma page 462}\\
&= \sum_{k\in\mathbb{Z}}\frac{\mathrm{d}}{\mathrm{d}\omega}\left[\frac{\sin\left(\frac{\omega}{2}\right)}{k\pi-\frac{\omega}{2}}\right]^{2(n+1)} && \text{by } \textit{linearity} \text{ of } \frac{\mathrm{d}}{\mathrm{d}\omega} \text{ operator}\\
&= \sum_{k\in\mathbb{Z}} 2(n+1)\left[\frac{\sin\left(\frac{\omega}{2}\right)}{k\pi-\frac{\omega}{2}}\right]^{2n+1}\frac{\mathrm{d}}{\mathrm{d}\omega}\left[\frac{\sin\left(\frac{\omega}{2}\right)}{k\pi-\frac{\omega}{2}}\right] && \text{by } \textit{power rule}\\
&= 2(n+1)\sum_{k\in\mathbb{Z}}\left[\frac{\sin\left(\frac{\omega}{2}\right)}{k\pi-\frac{\omega}{2}}\right]^{2n+1}\left[\frac{\frac{1}{2}\cos\left(\frac{\omega}{2}\right)\left(k\pi-\frac{\omega}{2}\right)-\sin\left(\frac{\omega}{2}\right)\left(-\frac{1}{2}\right)}{\left(k\pi-\frac{\omega}{2}\right)^2}\right] && \text{by } \textit{quotient rule}\\
&= (n+1)\sum_{k\in\mathbb{Z}}\left[\frac{\sin\left(\frac{\omega}{2}\right)}{k\pi-\frac{\omega}{2}}\right]^{2n+1}\left[\frac{\cos\left(\frac{\omega}{2}\right)\left(k\pi-\frac{\omega}{2}\right)+\sin\left(\frac{\omega}{2}\right)}{\left(k\pi-\frac{\omega}{2}\right)^2}\right]
\end{aligned}$$

10. lemma: $\tilde{S}'_n(0) = \tilde{S}'_n(\pi) = 0$. Proof: This follows from Proposition **??** (page **??**). Here is alternate proof:

$$\begin{aligned}
\tilde{S}'_n(0) &= \lim_{\omega\to 0}\tilde{S}'_n(\omega)\\
&= \lim_{\omega\to 0}(n+1)\sum_{k\in\mathbb{Z}}\left[\frac{\sin\left(\frac{\omega}{2}\right)}{k\pi-\frac{\omega}{2}}\right]^{2n+1}\left[\frac{\cos\left(\frac{\omega}{2}\right)\left(k\pi-\frac{\omega}{2}\right)+\sin\left(\frac{\omega}{2}\right)}{\left(k\pi-\frac{\omega}{2}\right)^2}\right] && \text{by (9) lemma}\\
&= \lim_{\omega\to 0}(n+1)\left[\frac{\sin\left(\frac{\omega}{2}\right)}{-\frac{\omega}{2}}\right]^{2n+1}\left[\frac{\cos\left(\frac{\omega}{2}\right)\left(-\frac{\omega}{2}\right)+\sin\left(\frac{\omega}{2}\right)}{\left(-\frac{\omega}{2}\right)^2}\right]\\
&= (n+1)\lim_{\omega\to 0}\left[\frac{\sin\left(\frac{\omega}{2}\right)}{-\frac{\omega}{2}}\right]^{2n+1}\lim_{\omega\to 0}\left[\frac{\cos\left(\frac{\omega}{2}\right)\left(-\frac{\omega}{2}\right)+\sin\left(\frac{\omega}{2}\right)}{\left(-\frac{\omega}{2}\right)^2}\right]\\
&= (n+1)[-1]^{2n+1}\lim_{\omega\to 0}\left[\frac{-\frac{1}{2}\sin\left(\frac{\omega}{2}\right)\left(-\frac{\omega}{2}\right)+\cos\left(\frac{\omega}{2}\right)\left(-\frac{1}{2}\right)+\cos\left(\frac{\omega}{2}\right)\left(\frac{1}{2}\right)}{-\frac{2}{2}\left(-\frac{\omega}{2}\right)}\right] && \text{by } \textit{l'Hôpital's rule}\\
&= (1)(0)
\end{aligned}$$

$$= 0$$

$$\begin{aligned}
\tilde{S}_n'(\pi) &= (n+1)\sum_{k\in\mathbb{Z}}\left[\frac{\sin\left(\frac{\pi}{2}\right)}{k\pi-\frac{\pi}{2}}\right]^{2n+1}\left[\frac{\cos\left(\frac{\pi}{2}\right)\left(k\pi-\frac{\pi}{2}\right)+\sin\left(\frac{\pi}{2}\right)}{\left(k\pi-\frac{\pi}{2}\right)^2}\right] && \text{by (9) lemma}\\
&= (n+1)\sum_{k\in\mathbb{Z}}\left[\frac{1}{k\pi-\frac{\pi}{2}}\right]^{2n+1}\left[\frac{0\left(k\pi-\frac{\pi}{2}\right)+1}{\left(k\pi-\frac{\pi}{2}\right)^2}\right]\\
&= (n+1)\left(\frac{2}{\pi}\right)^{2n+3}\sum_{k\in\mathbb{Z}}\left[\frac{1}{2k-1}\right]^{2n+3}\\
&= (n+1)\left(\frac{2}{\pi}\right)^{2n+3}\left[\left(\frac{1}{1}\right)^{2n+3}+\left(\frac{1}{-1}\right)^{2n+3}+\left(\frac{1}{3}\right)^{2n+3}+\left(\frac{1}{-3}\right)^{2n+3}+\ldots\right]\\
&= (n+1)\left(\frac{2}{\pi}\right)^{2n+3}\sum_{k=1}^{\infty}(-1)^{k+1}\alpha_k \qquad \text{where } \alpha_k \triangleq \left\{\begin{array}{ll}\left(\frac{1}{k}\right)^{2n+3} & \text{for } k \text{ odd}\\ \left(\frac{1}{k-1}\right)^{2n+3} & \text{for } k \text{ even}\end{array}\right\}\\
&= 0 \qquad \text{because } \lim_{k\to\infty}\alpha_k = 0 \text{ and by } \textit{Alternating Series Test}
\end{aligned}$$

11. lemma: $\tilde{S}_n(\omega)$ is *decreasing* with respect to $\omega\in[0:\pi]$. Proof:

$$\begin{aligned}
\tilde{S}_n'(\omega) &= (n+1)\sum_{k\in\mathbb{Z}}\left[\frac{\sin\left(\frac{\omega}{2}\right)}{k\pi-\frac{\omega}{2}}\right]^{2n+1}\left[\frac{\cos\left(\frac{\omega}{2}\right)\left(k\pi-\frac{\omega}{2}\right)+\sin\left(\frac{\omega}{2}\right)}{\left(k\pi-\frac{\omega}{2}\right)^2}\right] && \text{by (9) lemma page 465}\\
&= \underbrace{(n+1)\left(\sin\frac{\omega}{2}\right)^{2n+1}}_{\geq 0 \text{ for } \omega\in[0:2\pi)}\underbrace{\sum_{k\in\mathbb{Z}}\left[\frac{1}{k\pi-\frac{\omega}{2}}\right]^{2n+2}}_{>0}\left[\underbrace{\left(\cos\frac{\omega}{2}\right)}_{\text{sign change at } \omega=\pi}+\frac{\overbrace{\sin\frac{\omega}{2}}^{>0 \text{ for } \omega\in(0:2\pi)}}{\underbrace{k\pi-\frac{\omega}{2}}_{\textit{decreasing} \text{ w.r.t. } \omega\in\mathbb{R}}}\right]
\end{aligned}$$

12. lemma: $\tilde{S}_n(\omega)$ is *increasing* with respect to $\omega\in[\pi:2\pi)$. Proof: This is true because $\tilde{S}_n(\omega)$ is *decreasing* in $[0:\pi]$ ((11) lemma) and because $\tilde{S}_n(\omega)$ is *symmetric* about $\omega=\pi$ ((5) lemma).

13. Proof that $0<\tilde{S}_n(\omega)\leq 1$:
    (a) $\tilde{S}_n(\omega)>0$ by (8) lemma and
    (b) $\tilde{S}_n(0)=1$ by item (6) and
    (c) $\tilde{S}_n(\omega)$ is *decreasing* from $\omega=0$ to $\omega=\pi$ by (11) lemma and
    (d) $\tilde{S}_n(\omega)$ is *increasing* from $\omega=\pi$ to $\omega=2\pi$ by (12) lemma and
    (e) $\tilde{S}_n(2\pi)=1$ because $\tilde{S}_n(2\pi)=\tilde{S}_n(0)$ by (4) lemma.

⇒

Theorem W.8. [23]

THM

1. $\left(\mathsf{N}_n(x-k)\right)_{k\in\mathbb{Z}}$ *is a* RIESZ BASIS *for* span $\left(\mathsf{N}_n(x-k)\right)_{k\in\mathbb{Z}}$ $\quad\forall n\in\mathbb{W}$
2. $\left(\mathsf{N}_n(x-k)\right)_{k\in\mathbb{Z}}$ *is an* ORTHONORMAL BASIS *for* span $\left(\mathsf{N}_n(x-k)\right)_{k\in\mathbb{Z}}$ $\iff n=0$

[23] Wojtaszczyk (1997) page 56 ⟨Proposition 3.12⟩, Prasad and Iyengar (1997) page 148 ⟨Theorem 6.3⟩, Forster and Massopust (2009) page 66 ⟨Theorem 2.25⟩

✎PROOF:

1. Proof that $(\!(\mathsf{N}_n(x-k))\!)_{k\in\mathbb{Z}}$ is a *Riesz basis* for $\operatorname{span}(\!(\mathsf{N}_n(x-k))\!)_{k\in\mathbb{Z}}$:

$$\begin{aligned} & 0 < \tilde{S}_n(\omega) \le 1 && \text{by Lemma W.6 page 461 (1)} \\ \implies & (\!(\mathsf{N}_n(x-k))\!)_{k\in\mathbb{Z}} \text{ is a } \textit{Riesz basis} \text{ for } \operatorname{span}(\!(\mathsf{N}_n(x-k))\!)_{k\in\mathbb{Z}} && \text{by Theorem ?? page ??} \end{aligned}$$

2. Proof that $\{n=0\} \iff (\!(\mathsf{N}_n(x-k))\!)_{k\in\mathbb{Z}}$ is an *orthonormal basis* for $\operatorname{span}(\!(\mathsf{N}_n(x-k))\!)_{k\in\mathbb{Z}}$:

$$\begin{aligned} n=0 \iff & \tilde{S}_n(\omega) = 1 && \text{by Lemma W.6 page 461 (2), (4)} \\ \iff & (\!(\mathsf{N}_n(x-k))\!)_{k\in\mathbb{Z}} \text{ is an } \textit{orthonormal basis} \text{ for } \operatorname{span}(\!(\mathsf{N}_n(x-k))\!) && \text{by Theorem ?? page ??} \end{aligned}$$

✏

W.6 Mutiresolution properties

W.6.1 Introduction

In 1989, Stéphane G. Mallat introduced the *Multiresolution Analysis* (MRA) structure (Definition ?? page ??) An MRA is very powerful because it can be used to approximate functions at incrementally increasing "scales" of resolution, and furthermore induces a *wavelet*. In fact, the MRA has become the dominate wavelet construction method. Moreover, P.G. Lemarié has proved that all wavelets with *compact support* are generated by an MRA. [24]

W.6.2 B-spline dyadic decomposition

One key feature of an MRA is *dyadic decomposition* such that $\mathsf{N}_n(x) = \sum_k \alpha_n \mathsf{N}_n(2x-n)$ for some sequence $(\!(\alpha_n)\!)$. As it turns out, *B-spline*s also have this property (next theorem).

Theorem W.9 (*B-spline dyadic decomposition*). [25] *Let* $\mathsf{N}_n(x)$ *be the* nTH ORDER B-SPLINE.

THM
$$\mathsf{N}_n(x) = \frac{1}{2^n}\sum_{k=0}^{n+1}\binom{n+1}{k}\mathsf{N}_n(2x-k) \qquad \forall n\in\mathbb{W},\ \forall x\in\mathbb{R}$$

✎PROOF:

1. Base case ...proof for $n=0$ case:

$$\begin{aligned} \mathsf{N}_0(x) &= \mathbb{1}_{[0:1)}(x) && \text{by definition of } \mathbb{1}_A(x) \quad \text{(Definition W.1 page 443)} \\ &= \mathbb{1}_{[0:\frac{1}{2})}(x) + \mathbb{1}_{[\frac{1}{2}:1)}(x) \\ &= \mathbb{1}_{[2\times 0:2\times\frac{1}{2})}(2x) + \mathbb{1}_{[2\times\frac{1}{2}-1:2\times 1-1)}(2x-1) \\ &= \mathbb{1}_{[0:1)}(2x) + \mathbb{1}_{[0:1)}(2x-1) \\ &= \frac{1}{2^0}\sum_{k=0}^{0+1}\binom{0+1}{k}\mathsf{N}_n(2x-k) \end{aligned}$$

[24] Mallat (1999) page 240, Definition ?? (page ??)
[25] Prasad and Iyengar (1997) pages 151–152 ⟨proof using Fourier transform⟩

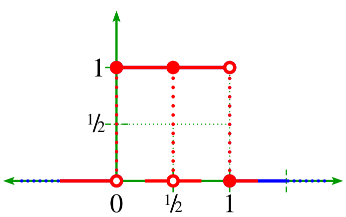
2. Induction step…proof that n case $\implies$ $n+1$ case:

$$
\begin{aligned}
\mathsf{N}_{n+1}(x) &= \int_0^1 \mathsf{N}_n(x-\tau)\,\mathrm{d}\tau && \text{by Lemma W.2 page 444}\\
&= \int_0^1 \frac{1}{2^n}\sum_{k=0}^{n+1}\binom{n+1}{k}\mathsf{N}_n(2x-2\tau-k)\,\mathrm{d}\tau && \text{by induction hypothesis}\\
&= \frac{1}{2^n}\sum_{k=0}^{n+1}\binom{n+1}{k}\int_{\tau=0}^{\tau=1}\mathsf{N}_n(2x-2\tau-k)\,\mathrm{d}\tau && \text{by \textit{linearity} of }\textstyle\sum\text{ operator}\\
&= \frac{1}{2^n}\sum_{k=0}^{n+1}\binom{n+1}{k}\int_{u=0}^{u=2}\mathsf{N}_n(2x-u-k)\frac{1}{2}\,\mathrm{d}u && \text{where } u\triangleq 2\tau \implies \tau=\frac{1}{2}u\\
&= \frac{1}{2^{n+1}}\sum_{k=0}^{n+1}\binom{n+1}{k}\left[\int_{u=0}^{u=1}\mathsf{N}_n(2x-k-u)\,\mathrm{d}u+\int_{u=1}^{u=2}\mathsf{N}_n(2x-k-u)\,\mathrm{d}u\right]\\
&= \frac{1}{2^{n+1}}\sum_{k=0}^{n+1}\binom{n+1}{k}\left[\int_{u=0}^{u=1}\mathsf{N}_n(2x-k-u)\,\mathrm{d}u+\int_{v=0}^{v=1}\mathsf{N}_n(2x-k-v-1)\,\mathrm{d}v\right] && \begin{array}{l}\text{where } v\triangleq u-1\\ \implies u=v+1\end{array}\\
&= \frac{1}{2^{n+1}}\sum_{k=0}^{n+1}\binom{n+1}{k}\left[\mathsf{N}_n(2x-k)+\mathsf{N}_n(2x-k-1)\right] && \text{by Lemma W.2 page 444}\\
&= \frac{1}{2^{n+1}}\left[\sum_{k=0}^{n+1}\binom{n+1}{k}\mathsf{N}_n(2x-k)+\sum_{m=1}^{n+2}\binom{n+1}{m-1}\mathsf{N}_n(2x-m)\right] && \text{where } m\triangleq k+1 \implies k=m-1\\
&= \frac{1}{2^{n+1}}\left[\underbrace{\sum_{k=1}^{n+1}\left[\binom{n+1}{k}+\binom{n+1}{k-1}\right]\mathsf{N}_n(2x-k)}_{\text{common indices of above two summations}}+\underbrace{\binom{n+1}{0}\mathsf{N}_n(2x-0)}_{k=0\text{ term}}+\underbrace{\binom{n+2}{n+2}\mathsf{N}_n(2x-n-2)}_{m=n+2\text{ term}}\right]\\
&= \frac{1}{2^{n+1}}\left[\underbrace{\sum_{k=1}^{n+1}\binom{n+2}{k}\mathsf{N}_n(2x-k)}_{\text{by \textit{Stifel formual} (Theorem ?? page ??)}}+\underbrace{\binom{n+2}{0}\mathsf{N}_n(2x-0)}_{\text{because }\binom{n+1}{0}=1=\binom{n+2}{0}}+\binom{n+2}{n+2}\mathsf{N}_n(2x-n-2)\right]\\
&= \frac{1}{2^{n+1}}\sum_{k=0}^{n+2}\binom{n+2}{k}\mathsf{N}_n(2x-k)
\end{aligned}
$$

Example W.16. [26]The 0 *order B-spline dyadic decomposition*

$$\mathsf{N}_0(x)=\frac{1}{1}\sum_{k=0}^{k=1}\binom{1}{k}\mathsf{N}_0(2x-k)$$

is illustrated to the right.

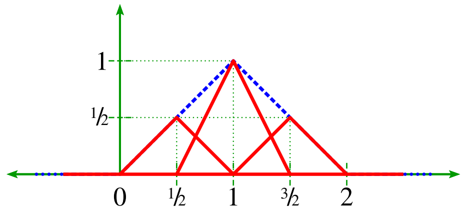


[26] Strang (1989) page 615 ⟨Box function⟩, Strang and Nguyen (1996) page 441 ⟨Box function⟩

Example W.17. [27]The 1*st order B-spline dyadic decomposition*

$$\mathsf{N}_1(x) = \frac{1}{2}\sum_{k=0}^{k=2}\binom{2}{k}\mathsf{N}_1(2x-k)$$

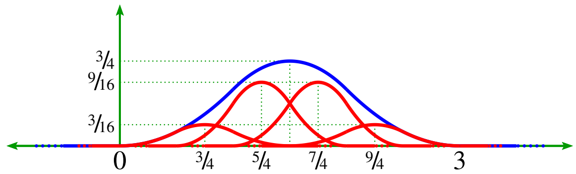
is illustrated to the right.



Example W.18. The 2*nd order B-spline dyadic decomposition*

$$\mathsf{N}_2(x) = \frac{1}{4}\sum_{k=0}^{k=3}\binom{3}{k}\mathsf{N}_2(2x-k)$$

is illustrated to the right.



W.6.3 B-spline MRA scaling functions

Theorem W.10. *Let f $N_n(x)$ be the* nTH ORDER B-SPLINE *(Definition W.3 page 444)*.
Let $\boldsymbol{V}_j \triangleq \operatorname{span}\left(\!\left(\mathsf{N}_n\left(2^j x - k\right)\right)\!\right)_{k\in\mathbb{Z}}$.

THM $\left(\!\left(\boldsymbol{V}_j\right)\!\right)_{j\in\mathbb{Z}}$ *is a* MULTIRESOLUTION ANALYSIS *on* $\boldsymbol{L}^2_{\mathbb{R}}$ *with* SCALING FUNCTION $\phi(x) \triangleq \mathsf{N}_n(x)$

✎PROOF:

1. lemma: $(\!(\mathsf{N}_n(x-k))\!)_{k\in\mathbb{Z}}$ is a *Riesz sequence*. Proof: by Theorem W.8 (page 466).
2. lemma: $\exists\,(\!(h_k)\!)$ such that $\mathsf{N}_n(x) = \sum_{k\in\mathbb{Z}} h_k \mathsf{N}_n(2x-k)$. Proof: by Theorem W.9 (page 467). In fact, note that $h_k = \frac{1}{2^n\sqrt{2}}\binom{n+1}{k}$
3. lemma: $\tilde{\mathsf{F}}\mathsf{N}_n(\omega)$ is *continuous* at 0. Proof:

$$\begin{aligned}\tilde{\mathsf{F}}\mathsf{N}_n(\omega) &= \frac{1}{\sqrt{2\pi}}e^{-i\frac{(n+1)\omega}{2}}\left(\operatorname{sinc}\frac{\omega}{2}\right)^{n+1} && \text{by Lemma W.5 page 458}\\ &\implies \textit{continuous}\text{ at }0 && \text{by known property of sinc function}\end{aligned}$$

4. lemma: $\tilde{\phi}(0) \neq 0$. Proof:

$$\begin{aligned}\tilde{\mathsf{F}}\mathsf{N}_n(0) &= \left.\frac{1}{\sqrt{2\pi}}e^{-i\frac{(n+1)\omega}{2}}\left(\operatorname{sinc}\frac{\omega}{2}\right)^{n+1}\right|_{\omega=0} && \text{by Lemma W.5 page 458}\\ &= 1\cdot\frac{1}{1/2} = 2 && \text{by \textit{l'Hôpital's rule}}\\ &\neq 0\end{aligned}$$

5. The completion of this proof follows directly from (1) lemma, (2) lemma, (3) lemma, (4) lemma, and Theorem **??** (page **??**).

✏

[27] Strang (1989) page 615 ⟨Hat function⟩, Strang and Nguyen (1996) page 442 ⟨Hat function⟩, Heil (2011) page 380 ⟨Fig. 12.10⟩

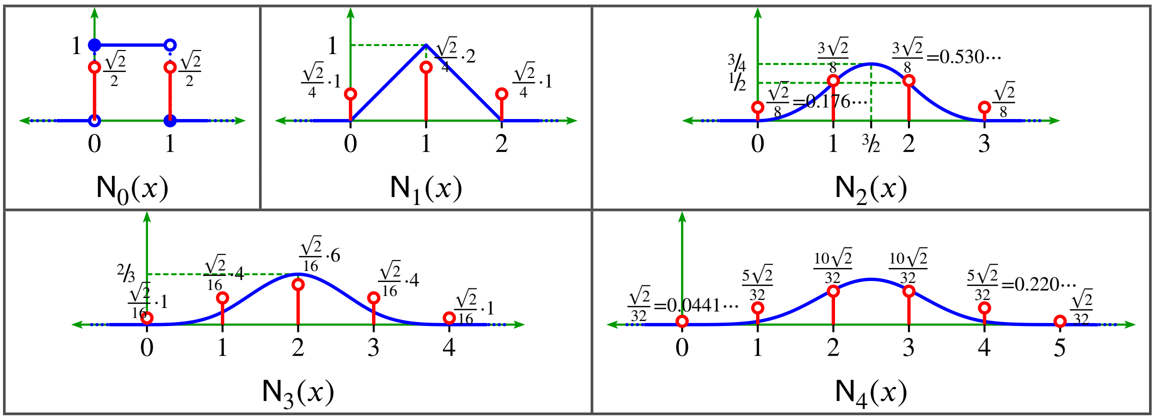


Figure W.4: *dilation equation* demonstrations for selected B-splines (Example W.19 page 470)

W.6.4 B-spline MRA coefficient sequences

Because each *B-spline* $\mathsf{N}_n(x)$ is the *scaling function* for an *MRA* (Theorem W.10 page 469), each *B-spline* also satisfies the *dilation equation* (Theorem ?? page ??) such that

$$\mathsf{N}_n(x) = \sqrt{2}\sum_{k\in\mathbb{Z}} h_k \mathsf{N}(2x-k) \quad \text{where} \quad h_k = \left\{\begin{array}{ll} \frac{\sqrt{2}}{2^{n+1}}\binom{n+1}{k} & \text{for } n=0,1,\ldots,n \\ 0 & \text{otherwise}\end{array}\right\}$$

The resulting *sequence* $(\!(h_k)\!)_{k\in\mathbb{Z}}$ is the *order n B-spline MRA coefficient sequence* induced by the *order n B-spline MRA scaling sequence* $\phi(x) \triangleq \mathsf{N}_n(x)$.[28]

Example W.19. See Figure W.4 (page 470) for some *dilation equation* demonstrations of selected B-splines.

Theorem W.11 (B-spline scaling coefficients). *Let* $\left(\boldsymbol{L}^2_{\mathbb{R}}, (\!(\boldsymbol{V}_j)\!), \phi, (\!(h_n)\!)\right)$ *be an* MRA SYSTEM *(Definition ?? page ??). Let* $\mathsf{N}_n(x)$ *be a* nTH ORDER B-SPLINE *(Definition W.3 page 444).*

THM

$$\underbrace{\phi(x) \triangleq \mathsf{N}_n(x)}_{\text{(1) B-spline scaling function}} \implies (\!(h_k)\!) = \left\{\begin{array}{ll} \frac{\sqrt{2}}{2^{n+1}}\binom{n}{k} & \text{for } k=0,1,\ldots,n \\ 0 & \text{otherwise}\end{array}\right. \quad \text{(2) scaling sequence in "time"}$$

$$\iff \check{\mathsf{h}}(z)\Big|_{z\triangleq e^{i\omega}} = \frac{\sqrt{2}}{2^n}\left(1+z^{-1}\right)^{n+1}\Big|_{z\triangleq e^{i\omega}} \quad \text{(3) scaling sequence in "z domain"}$$

$$\iff \tilde{\mathsf{h}}(\omega) = 2\sqrt{2}e^{-i\frac{n+1}{2}\omega}\left[\cos\left(\frac{\omega}{2}\right)\right]^{n+1} \quad \text{(4) scaling sequence in "frequency"}$$

✎PROOF:

1. Proof that (1) $\implies$ (3): By Theorem W.10 page 469 we know that $\mathsf{N}_n(x)$ is a *scaling function* (Definition ?? page ??). So then we know that we can use Lemma ?? page ??.

$$\begin{aligned}
\tilde{\mathsf{h}}(\omega) &= \sqrt{2}\frac{\tilde{\phi}(2\omega)}{\tilde{\phi}(\omega)} && \text{by Lemma ?? page ??}\\
&= \sqrt{2}\frac{\tilde{\mathsf{N}}_n(2\omega)}{\tilde{\mathsf{N}}_n(\omega)} && \text{by (1)}\\
&= \sqrt{2}\frac{\frac{1}{\sqrt{2\pi}}\left(\frac{1-e^{-i2\omega}}{2i\omega}\right)^{n+1}}{\frac{1}{\sqrt{2\pi}}\left(\frac{1-e^{-i\omega}}{i\omega}\right)^{n+1}} && \text{by Lemma W.5 page 458}
\end{aligned}$$

[28] For *Octave/MatLab* code useful for plotting a function given a sequence of coefficients $(\!(h_k)\!)$, see Section ?? (page ??).

$$
\begin{aligned}
&= \frac{\sqrt{2}}{2^{n+1}} \left(\frac{1 - z^{-2}}{1 - z^{-1}} \right)^{n+1} \Bigg|_{z=e^{i\omega}} \\
&= \frac{\sqrt{2}}{2^{n+1}} \left[\left(\frac{1 - z^{-2}}{1 - z^{-1}} \right) \left(\frac{1 + z^{-1}}{1 + z^{-1}} \right) \right]^{n+1} \Bigg|_{z=e^{i\omega}} \\
&= \frac{\sqrt{2}}{2^{n+1}} \left(\frac{(1 - z^{-2})(1 + z^{-1})}{1 - z^{-2}} \right)^{n+1} \Bigg|_{z=e^{i\omega}} \\
&= \frac{\sqrt{2}}{2^{n}} \left(1 + z^{-1}\right)^{n+1} \Big|_{z \triangleq e^{i\omega}}
\end{aligned}
$$

2. Proof that (3) $\iff$ (2):

$$
\begin{aligned}
\check{\mathsf{h}}(z)\Big|_{z \triangleq e^{i\omega}} &= \frac{\sqrt{2}}{2^{n}} \left(1 + z^{-1}\right)^{n+1} \Big|_{z \triangleq e^{i\omega}} && \text{by (3)} \\
&= \frac{\sqrt{2}}{2^{n}} \left(\sum_{k=0}^{n+1} \binom{n}{k} z^{-k} \right) \Bigg|_{z \triangleq e^{i\omega}} && \text{by \textit{binomial theorem}} \\
\iff \qquad h_k &= \frac{\sqrt{2}}{2^{n+1}} \binom{n}{k} && \text{by definition of \textit{Z transform} (Definition V.1 page 429)}
\end{aligned}
$$

3. Proof that (3) $\implies$ (4):

$$
\begin{aligned}
\tilde{\mathsf{h}}(\omega) &= \check{\mathsf{h}}(z)\Big|_{z \triangleq e^{i\omega}} && \text{by definition of \textit{DTFT} (Definition U.1 page 419)} \\
&= \frac{\sqrt{2}}{2^{n}} \left(1 + z^{-1}\right)^{n+1} \Big|_{z \triangleq e^{i\omega}} && \text{by (3)} \\
&= \frac{\sqrt{2}}{2^{n}} \left(1 + e^{-i\omega}\right)^{n+1} && \text{by definition of } z \\
&= \frac{\sqrt{2}}{2^{n}} \left[e^{-i\frac{1}{2}\omega} \left(e^{i\frac{\omega}{2}} + e^{-i\frac{\omega}{2}} \right) \right]^{n+1} \\
&= \frac{\sqrt{2}}{2^{n}} e^{-i\frac{n+1}{2}\omega} \left[2\cos\left(\frac{\omega}{2}\right) \right]^{n+1} \\
&= 2\sqrt{2} e^{-i\frac{n+1}{2}\omega} \left[\cos\left(\frac{\omega}{2}\right) \right]^{n+1}
\end{aligned}
$$

4. Proof that (3) $\impliedby$ (4):

$$
\begin{aligned}
\check{\mathsf{h}}(z)\Big|_{z \triangleq e^{i\omega}} &= \check{\mathsf{h}}\left(e^{i\omega}\right) \\
&= \tilde{\mathsf{h}}(\omega) \\
&= 2\sqrt{2} e^{-i\frac{n+1}{2}\omega} \left[\cos\left(\frac{\omega}{2}\right) \right]^{n+1} && \text{by (4)} \\
&= \frac{\sqrt{2}}{2^{n}} e^{-i\frac{n+1}{2}\omega} \left[2\cos\left(\frac{\omega}{2}\right) \right]^{n+1} \\
&= \frac{\sqrt{2}}{2^{n}} \left[e^{-i\frac{1}{2}\omega} \left(e^{i\frac{\omega}{2}} + e^{-i\frac{\omega}{2}} \right) \right]^{n+1} \\
&= \frac{\sqrt{2}}{2^{n}} \left(1 + e^{-i\omega}\right)^{n+1} \\
&= \frac{\sqrt{2}}{2^{n}} \left(1 + z^{-1}\right)^{n+1} \Big|_{z \triangleq e^{i\omega}}
\end{aligned}
$$

Example W.20 (2 coefficient case). [29] Let $\left(L^2_{\mathbb{R}}, \left(\!\left(V_j\right)\!\right), \phi, \left(\!\left(h_n\right)\!\right)\right)$ be an *MRA system* (Definition ?? page ??).

$$\underbrace{\left\{\begin{array}{ll} 1. & \operatorname{supp}\phi(x) = [0:1] \quad \text{and} \\ 2. & \left(\!\left(\phi(x-k)\right)\!\right) \text{ forms a} \\ & \textit{partition of unity} \end{array}\right\}}_{(A)} \iff \underbrace{h_n = \left\{\begin{array}{ll} \frac{\sqrt{2}}{2} & \text{for } n=0 \\ \frac{\sqrt{2}}{2} & \text{for } n=1 \\ 0 & \text{otherwise} \end{array}\right\}}_{(B)} \iff \underbrace{\left\{\phi(x) = \mathsf{N}_0(x)\right\}}_{(C)}$$

✎PROOF:

1. Proof that (A) $\implies$ (B):

   (a) lemma: Only h_0 and h_1 are *non-zero*; All other coefficients h_k are 0. Proof: This follows from $\operatorname{supp}\phi(x) = [0:1]$ (Definition ?? page ??) and by Theorem ?? page ??.

   (b) lemma (equations for $\left(\!\left(h_k\right)\!\right)$): Because $\left(\!\left(h_k\right)\!\right)$ is a *scaling coefficient sequence* (Definition ?? page ??), it must satisfy the *admissibility equation* (Theorem ?? page ??). And because $\left(\!\left(\phi(x-k)\right)\!\right)$ forms a *partition of unity*, it must satisfy the equations given by Theorem ?? (page ??). (1a) lemma and these two constraints yield two simultaneous equations and two unknowns:

   $$\begin{array}{rcll} h_0 + h_1 &=& \sqrt{2} & (\textit{admissibility condition}) \\ h_0 - h_1 &=& 0 & (\textit{partition of unity / zero at } -1 \textit{ / vanishing 0th moment}) \end{array}$$

   (c) lemma: The equations provided by (1b) lemma can be expressed in matrix algebra form as follows...

   $$\underbrace{\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}}_{\mathbf{A}} \begin{bmatrix} h_0 \\ h_1 \end{bmatrix} = \begin{bmatrix} \sqrt{2} \\ 0 \end{bmatrix}$$

   (d) lemma: The *inverse* $\mathbf{A}^{-1}$ of $\mathbf{A}$ can be expressed as demonstrated below...

   $$\left[\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 1 & -1 & 0 & 1 \end{array}\right] \to \left[\begin{array}{cc|cc} 2 & 0 & 1 & 1 \\ 1 & -1 & 0 & 1 \end{array}\right] \to \left[\begin{array}{cc|cc} 2 & 0 & 1 & 1 \\ 0 & -1 & -\tfrac{1}{2} & \tfrac{1}{2} \end{array}\right] \to \left[\begin{array}{cc|cc} 1 & 0 & \tfrac{1}{2} & \tfrac{1}{2} \\ 0 & 1 & \tfrac{1}{2} & -\tfrac{1}{2} \end{array}\right]$$

   $$\implies \mathbf{A}^{-1} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

   (e) Proof for the values of $\left(\!\left(h_k\right)\!\right)$ (B):

   $$\begin{aligned} \begin{bmatrix} h_0 \\ h_1 \end{bmatrix} = \mathbf{A}^{-1}\mathbf{A}\begin{bmatrix} h_0 \\ h_1 \end{bmatrix} &= \mathbf{A}^{-1}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} h_0 \\ h_1 \end{bmatrix} && \text{by (1c) lemma} \\ &= \mathbf{A}^{-1}\begin{bmatrix} \sqrt{2} \\ 0 \end{bmatrix} && \text{by (1c) lemma} \\ &= \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} \sqrt{2} \\ 0 \end{bmatrix} && \text{by (1d) lemma} \\ &= \frac{\sqrt{2}}{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix} \end{aligned}$$

[29] Haar (1910), Wojtaszczyk (1997) pages 14–15 ⟨"Sources and comments"⟩

2. Proof that (B) $\implies$ (C):

$$\begin{aligned}
\text{(B)} \implies \phi(x) &= \sum_{k\in\mathbb{Z}} h_k \sqrt{2}\phi(2x-k) && \textit{dilation equation} && \text{(Theorem ?? page ??)}\\
&= \sum_{k=0}^{k=1} \left(\frac{\sqrt{2}}{2}\right)\sqrt{2}\phi(2x-k) && \text{by item (1e) page 472}\\
&= \sum_{k=0}^{k=1} \phi(2x-k)\\
&= \sum_{k=0}^{k=1} \binom{1}{k}\phi(2x-k) && \text{by definition of } \binom{n}{k} && \text{(Definition ?? page ??)}\\
&\implies \text{(D)} && \text{by } \textit{B-spline dyadic decomposition} && \text{(Theorem W.9 page 467)}
\end{aligned}$$

3. Proof that (B) $\impliedby$ (C):

$$\begin{aligned}
\text{(C)} \implies \mathsf{N}_0(x) &= \sum_{k=0}^{k=1} \binom{1}{k}\mathsf{N}_0(2x-k) && \text{by } \textit{B-spline dyadic decomposition} && \text{(Theorem W.9 page 467)}\\
&= \sum_{k=0}^{k=1} \left(\frac{\sqrt{2}}{2}\right)\sqrt{2}\mathsf{N}_0(2x-k) && \text{by definition of } \binom{n}{k} && \text{(Definition ?? page ??)}\\
&= \sum_{k\in\mathbb{Z}} h_k\sqrt{2}\mathsf{N}_0(2x-k) && \text{by definition of } \binom{n}{k} && \text{(Definition ?? page ??)}\\
&\implies \text{(B)}
\end{aligned}$$

4. Proof that (A) $\impliedby$ (C):
   1. Proof that (C) $\implies$ $\operatorname{supp}\phi(x) = [0:1]$: by Theorem W.4 (page 454)
   2. Proof that (C) $\implies$ $\langle\!\langle \phi(x-k)\rangle\!\rangle$ forms a *partition of unity*: by Theorem W.7 (page 460)

✏

EX

n=0,	(÷0!)	1;								
n=1,	(÷1!)	1,	0;							
		−1,	2;							
n=2,	(÷2!)	1,	0,	0;						
		−2,	6,	−3;						
		1,	−6,	9;						
n=3,	(÷3!)	1,	0,	0,	0;					
		−3,	12,	−12,	4;					
		3,	−24,	60,	−44;					
		−1,	12,	−48,	64;					
n=4,	(÷4!)	1,	0,	0,	0,	0;				
		−4,	20,	−30,	20,	−5;				
		6,	−60,	210,	−300,	155;				
		−4,	60,	−330,	780,	−655;				
		1,	−20,	150,	−500,	625;				
n=5,	(÷5!)	1,	0,	0,	0,	0,	0;			
		−5,	30,	−60,	60,	−30,	6;			
		10,	−120,	540,	−1140,	1170,	−474;			
		−10,	180,	−1260,	4260,	−6930,	4386;			
		5,	−120,	1140,	−5340,	12270,	−10974;			
		−1,	30,	−360,	2160,	−6480,	7776;			
n=6,	(÷6!)	1,	0,	0,	0,	0,	0,	0;		
		−6,	42,	−105,	140,	−105,	42,	−7;		
		15,	−210,	1155,	−3220,	4935,	−3990,	1337;		
		−20,	420,	−3570,	15680,	−37590,	47040,	−24178;		
		15,	−420,	4830,	−29120,	96810,	−168000,	119182;		
		−6,	210,	−3045,	23380,	−100065,	225750,	−208943;		
		1,	−42,	735,	−6860,	36015,	−100842,	117649;		
n=7,	(÷7!)	1,	0,	0,	0,	0,	0,	0,	0;	
		−7,	56,	−168,	280,	−280,	168,	−56,	8;	
		21,	−336,	2184,	−7560,	15400,	−18648,	12488,	−3576;	
		−35,	840,	−8400,	45360,	−143360,	267120,	−273280,	118896;	
		35,	−1120,	15120,	−111440,	483840,	−1238160,	1733760,	−1027984;	
		−21,	840,	−14280,	133560,	−741160,	2436840,	−4391240,	3347016;	
		7,	−336,	6888,	−78120,	528920,	−2135448,	4753336,	−4491192;	
		−1,	56,	−1344,	17920,	−143360,	688128,	−1835008,	2097152;	
n=8,	(÷8!)	1,	0,	0,	0,	0,	0,	0,	0,	0;
		−8,	72,	−252,	504,	−630,	504,	−252,	72,	−9
		28,	−504,	3780,	−15624,	39690,	−64008,	64260,	−36792,	9207
		−56,	1512,	−17388,	111384,	−436590,	1079064,	−1650348,	1432872,	−541917
		70,	−2520,	39060,	−340200,	1821330,	−6146280,	12800340,	−15082200,	7715619
		−56,	2520,	−49140,	541800,	−3691170,	15903720,	−42324660,	63667800,	−41503131
		28,	−1512,	35532,	−474264,	3929310,	−20674584,	67410252,	−124449192,	99584613
		−8,	504,	−13860,	217224,	−2121210,	13208328,	−51179940,	112731192,	−107948223
		1,	−72,	2268,	−40824,	459270,	−3306744,	14880348,	−38263752,	43046721

Table W.1: Coefficients of the *B-spline*s $\mathsf{N}_n(x)$ multiplied by $n!$ (Example W.9 page 450)

APPENDIX X

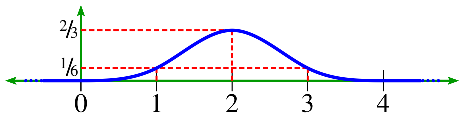
SOURCE CODE

The free and open source software package Maxima has been used to compute some of the algebraic expressions for *B-spline*s used in APPENDIX W (page 443):

```
/*=========================================================================*
 * Daniel J. Greenhoe                                                      *
 * Maxima script file                                                      *
 * To execute this script, start Maxima in a command window               *
 * in the subdirectory containing this file (e.g. c:\math\maxima\)         *
 * and then after the (%i...) prompt enter                                 *
 *    batchload("bspline.max")$                                            *
 * Data produced will be written to the file "bsplineout.txt".            *
 * reference: http://maxima.sourceforge.net/documentation.html             *
 *=========================================================================*/
/*---------------------------------------
 * initialize script
 *---------------------------------------*/
reset();
kill(all);
load(orthopoly);
display2d:false; /* 2-dimensional display */
writefile("bsplineout.txt");
/*---------------------------------------
 * n = B-spline order parameter
 * may be set to any value in {1,2,3,...}
 *---------------------------------------*/
n:2;
print("=====================================================================");
print("Daniel J. Greenhoe");
print("Output file for nth order B-spline Nn(x) calculation, n=",n," .");
print("Output produced using Maxima running the script file bspline.max");
print("=====================================================================");
Nnx:(1/n!)*sum((-1)^k*binomial(n+1,k)*(x-k)^n*unit_step(x-k),k,0,n+1);
print("------------------------------------------------------------");
print("             n+1      k  (n+1)       n                       ");
print(" n! Nn(x) = SUM (-1)   (    ) (x-k)   step(x-k)  ,n=",n);
print("             k=0         ( k )                               ");
print("             ",n+1,"       k (",n+1,")       ",n);
print(n,"! Nn(x) = SUM (-1)   (    ) (x-k)   step(x-k)");
print("             k=0         ( k )");
print("           = ",expand(n!*Nnx));
print("------------------------------------------------------------");
assume(x<=0);    print(n!,"N(x)= ",expand(n!*Nnx)," for x<=0");  forget(x<=0);
for i:0 thru n step 1 do(
  assume(x>i,x<(i+1)),
  print(n!,"N(x)= ",expand(n!*Nnx)," for ",i,"<x<",i+1),
  tex(expand(n!*Nnx),"djg.tex"),/*write output in TeX format to file "djg.tex"*/
  forget(x>i,x<(i+1))
  );
assume(x>(n+1)); print(n!,"N(x)= ",expand(n!*Nnx)," for x>",n+1); forget(x>(n+1));
```

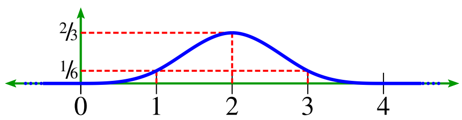
```
print("-------------------------------------------------------");
print(" values at some specific points x:                     ");
print("-------------------------------------------------------");
y:Nnx,x=(n+1)/2;print("N(",(n+1)/2,")= ",y," (center value)");
y:Nnx,x=(n+2)/2;print("N(",(n+2)/2,")= ",y);
y:Nnx,x=(n+3)/2;print("N(",(n+3)/2,")= ",y);
/*-------------------------------------
 * close output file
 *-------------------------------------*/
closefile();
```

Once the polynomial expressions for a *B-spline* have been calculated, they can be plotted within a LaTeX environment using the pst-plot package along with a LaTeX source file such as the following:[1]




```
%==========================================================================
% Daniel J. Greenhoe
% LaTeX file
% N_3(x) B-spline
% nominal unit = 10mm
%==========================================================================
\begin{pspicture}(-1,-0.5)(5,1)
  %--------------------------------------
  % parameters
  %--------------------------------------
  \psset{plotpoints=64,labelsep=1pt}%
  %--------------------------------------
  % axes
  %--------------------------------------
  \psaxes[linewidth=0.75pt,linecolor=axis,yAxis=false,ticks=x,labels=x]{<->}(0,0)(-1,0)(5,1)% x axis
  \psaxes[linewidth=0.75pt,linecolor=axis,xAxis=false,ticks=x,labels=x]{->}(0,0)(-1,0)(5,1)%  y axis
  %--------------------------------------
  % annotation
  %--------------------------------------
  \psline[linestyle=dashed,linewidth=0.75pt,linecolor=red](2,0)(2,0.667)%
  \psline[linestyle=dashed,linewidth=0.75pt,linecolor=red](0,0.667)(2,0.667)%
  \psline[linestyle=dashed,linewidth=0.75pt,linecolor=red](1,0)(1,0.1667)%
  \psline[linestyle=dashed,linewidth=0.75pt,linecolor=red](3,0)(3,0.1667)%
  \psline[linestyle=dashed,linewidth=0.75pt,linecolor=red](0,0.1667)(3,0.1667)%
  \uput[180](0,0.667){$\sfrac{2}{3}$}%
  \uput[180](0,0.1667){$\sfrac{1}{6}$}%
  %--------------------------------------
  % function plot
  %--------------------------------------
  \psplot{0}{1}{+1 x 3 exp mul                                          6 div}% for 0<=x<=1
  \psplot{1}{2}{-3 x 3 exp mul +12 x 2 exp mul add -12 x mul add  +4 add 6 div}% for 1<=x<=2
  \psplot{2}{3}{+3 x 3 exp mul -24 x 2 exp mul add +60 x mul add -44 add 6 div}% for 2<=x<=3
  \psplot{3}{4}{-1 x 3 exp mul +12 x 2 exp mul add -48 x mul add +64 add 6 div}% for 3<=x<=4
  \psline(0,0)(-0.5,0)\psline[linestyle=dotted](-0.5,0)(-0.75,0)%             % for x<=0
  \psline(4,0)(4.5,0)\psline[linestyle=dotted](4.5,0)(4.75,0)%                % for x>=4
\end{pspicture}%
```

Alternatively, one can plot $\mathsf{N}_3(x)$ more or less directly from the equation given in Theorem W.1 (page 447) without first calculating the polynomial expressions:



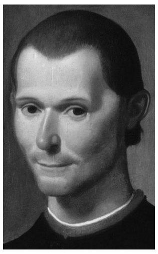

```
%==========================================================================
% Daniel J. Greenhoe
% LaTeX file
% N_3(x) B-spline
% nominal unit = 10mm
%==========================================================================
\begin{pspicture}(-1,-0.5)(5,1)
  %--------------------------------------
  % parameters
  %--------------------------------------
  \psset{plotpoints=64,labelsep=1pt}%
```

[1]For help with PostScript®math operators, see Adobe (1999) pages 508–509 ⟨Arithmetic and Math Operators⟩.

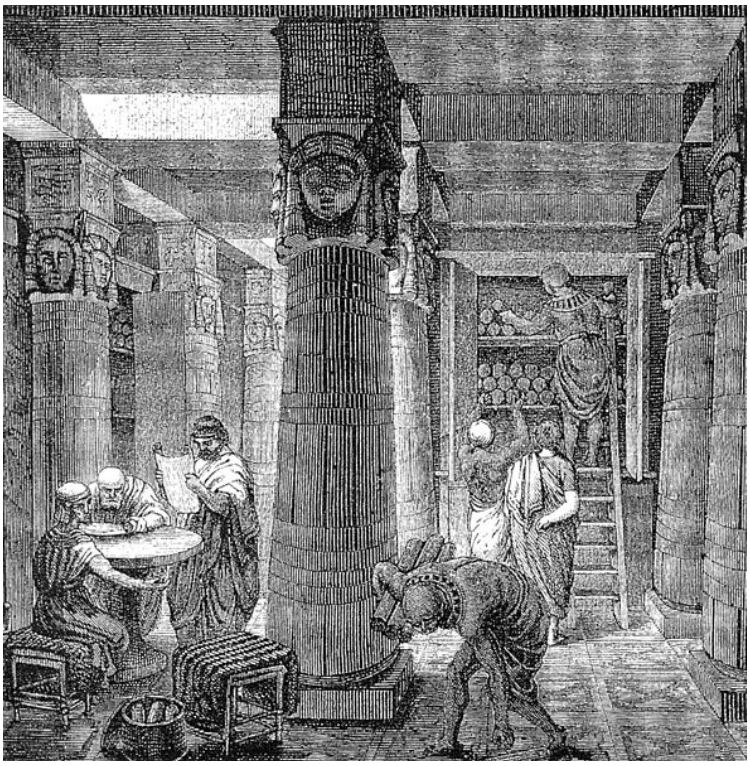
```
  %---------------------------------------
  % axes
  %---------------------------------------
  \psaxes[linewidth=0.75pt,linecolor=axis,yAxis=false,ticks=x,labels=x]{<->}(0,0)(-1,0)(5,1)% x axis
  \psaxes[linewidth=0.75pt,linecolor=axis,xAxis=false,ticks=x,labels=x]{->}(0,0)(-1,0)(5,1)%  y axis
  %---------------------------------------
  % annotation
  %---------------------------------------
  \psline[linestyle=dashed,linewidth=0.75pt,linecolor=red](2,0)(2,0.667)%
  \psline[linestyle=dashed,linewidth=0.75pt,linecolor=red](0,0.667)(2,0.667)%
  \psline[linestyle=dashed,linewidth=0.75pt,linecolor=red](1,0)(1,0.1667)%
  \psline[linestyle=dashed,linewidth=0.75pt,linecolor=red](3,0)(3,0.1667)%
  \psline[linestyle=dashed,linewidth=0.75pt,linecolor=red](0,0.1667)(3,0.1667)%
  \uput[180](0,0.667){$\sfrac{2}{3}$}%
  \uput[180](0,0.1667){$\sfrac{1}{6}$}%
  %-------------------------------------------------------------------------
  % for n=3
  %         1   n+1         (n+1)                   1    4            ( 4 )
  % N_n(x)--- SUM (-1)^k  (     ) (x-k)^n s(x-k) = --- SUM  (-1)^k (    ) (x-k)^3 s(x-k)
  %         n!  k=1         ( k )                   3!  k=1           ( k )
  %
  % where s(x) = 0 for x<0 and 1 for x>=0  (step function)
  %-------------------------------------------------------------------------
  \psplot{0}{1}  {1 x 0 sub 3 exp mul   6 div}%
  \psplot{1}{2}  {1 x 0 sub 3 exp mul   4 x 1 sub 3 exp mul sub   6 div}%
  \psplot{2}{3}  {1 x 0 sub 3 exp mul   4 x 1 sub 3 exp mul sub   6 x 2 sub 3 exp mul add   6 div}%
  \psplot{3}{4}  {1 x 0 sub 3 exp mul   4 x 1 sub 3 exp mul sub   6 x 2 sub 3 exp mul add   4 x 3 sub
          3 exp mul sub   6 div}%
  \psplot{4}{4.5}{1 x 0 sub 3 exp mul   4 x 1 sub 3 exp mul sub   6 x 2 sub 3 exp mul add   4 x 3 sub
          3 exp mul sub   1 x 4 sub 3 exp mul add 6 div}%
  %    N_3(x) = [(4choose0)(x-0)^3    - (4choose1)(x-1)^3       + (4choose2)(x-2)^3       -
          (4choose3)(x-3)^3       + (4choose4)(x-4)^3       ]/3!
  %           = [         1 (x-0)^3    -          4 (x-1)^3       +          6 (x-2)^3       -          4
          (x-3)^3       +          1 (x-4)^3       ]/6
  \psline(0,0)(-0.5,0)%
  \psline[linestyle=dotted](-0.5,0)(-0.75,0)%
  \psline[linestyle=dotted](4.5,0)(4.75,0)%
\end{pspicture}%
```

Back Matter

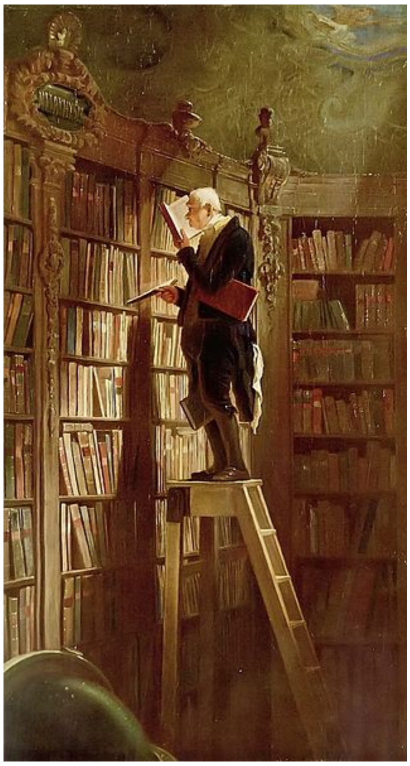
❝*It appears to me that if one wants to make progress in mathematics, one should study the masters and not the pupils.*❞
Niels Henrik Abel (1802–1829), Norwegian mathematician [2]



❝*When evening comes, I return home and go to my study. On the threshold I strip naked, taking off my muddy, sweaty workaday clothes, and put on the robes of court and palace, and in this graver dress I enter the courts of the ancients and am welcomed by them, and there I taste the food that alone is mine, and for which I was born. And there I make bold to speak to them and ask the motives of their actions, and they, in their humanity reply to me. And for the space of four hours I forget the world, remember no vexation, fear poverty no more, tremble no more at death; I pass indeed into their world.*❞
Niccolò Machiavelli (1469–1527), Italian political philosopher, in a 1513 letter to friend Francesco Vettori. [3]



ancient library of Alexandria



The Book Worm by *Carl Spitzweg*, circa 1850 [4]

[2] quote: Simmons (2007) page 187.
image: http://en.wikipedia.org/wiki/Image:Niels_Henrik_Abel.jpg, public domain

[3] quote: Machiavelli (1961) page 139?.
image: http://commons.wikimedia.org/wiki/File:Santi_di_Tito_-_Niccolo_Machiavelli%27s_portrait_headcrop.jpg, public domain

[4] http://en.wikipedia.org/wiki/File:Ancientlibraryalex.jpg, public domain http://en.wikipedia.org/wiki/File:Carl_Spitzweg_021.jpg,

❝*To sit alone in the lamplight with a book spread out before you, and hold intimate converse with men of unseen generations—such is a pleasure beyond compare.*❞
Yoshida Kenko (Urabe Kaneyoshi) (1283? – 1350?), Japanese author and Buddhist monk [5]

[5] quote: Kenko (circa 1330)
image: http://en.wikipedia.org/wiki/Yoshida_Kenko

BIBLIOGRAPHY

M. S. Abdaheer. *Digital Signal Processing*. University Science Press, New Delhi, 2009.

M. Abom. A note on random errors in frequency response estimators. *Journal of Sound and Vibration*, 107(2):355–358, 1986. doi: 10.1016/0022-460X(86)90244-0. URL http://adsabs.harvard.edu/abs/1986JSV...107..355A.

Yuri A. Abramovich and Charalambos D. Aliprantis. *An Invitation to Operator Theory*. American Mathematical Society, Providence, Rhode Island, 2002. ISBN 0-8218-2146-6. URL http://books.google.com/books?vid=ISBN0821821466.

Milton Abramowitz and Irene A. Stegun, editors. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. National Bureau of Standards, 1972. URL http://www.cs.bham.ac.uk/~aps/research/projects/as/book.php.

Adobe. *PostScript Language Reference*. Addison-Wesley Publishing Company, Reading, Massachusetts, 1999. ISBN 0-201-37922-8. URL http://wwwimages.adobe.com/content/dam/Adobe/en/devnet/postscript/pdfs/PLRM.pdf.

N. I. Akhiezer and I. M. Glazman. *Theory of Linear Operators in Hilbert Spaces*, volume 1. Dover, New York, 1993. URL http://books.google.com/books?vid=ISBN0486677486. Translated from the original Russian text *Teoriia lineinykh operatorov v Gil'bertovom prostranstve*.

Charalambos D. Aliprantis and Owen Burkinshaw. *Principles of Real Analysis*. Acedemic Press, London, 3 edition, 1998. ISBN 9780120502578. URL http://www.amazon.com/dp/0120502577.

Charalambos D. Aliprantis and Owen Burkinshaw. *Positive Operators*. Springer, Dordrecht, 2006. ISBN 9781402050077. URL http://books.google.com/books?vid=ISBN1402050070. reprint of Academic Press 1985 edition.

Randall J. Allemang, R. Zimmerman, and David L. Brown. Technlqucs of reducing noise in frquency response measurements. *ASME Winter Annual Meeting*, 1979. session on Application of System Identlfication Techniques.

Randall J. Allemang, David L. Brown, and Robert W. Rost. *Measurement Techniques for Experimental Modal Analysis*, 2(AFWAL-TR-87-3069), December 1987. URL https://apps.dtic.mil/dtic/tr/fulltext/u2/a195145.pdf.

Theodore T. Allen. *Introduction to Engineering Statistics and Lean Six Sigma: Statistical Quality Control and Design of Experiments and Systems*. Springer, 3 edition, 2018. ISBN 9781447174202.

Herbert Amann and Joachim Escher. *Analysis II*. Birkhäauser Verlag AG, Basel–Boston–Berlin, 2008. ISBN 978-3-7643-7472-3. URL http://books.google.com/books?vid=ISBN3764374721.

Ichiro Amemiya and Huzihiro Araki. A remark on piron's paper. *Publications of the Research Institute for Mathematical Sciences, Kyoto University*, 2(3):423–427, 1966. ISSN 0034-5318. doi: 10.2977/prims/1195195769. URL http://projecteuclid.org/euclid.prims/1195195769.

Dan Amir. *Characterizations of Inner Product Spaces*, volume 20 of *Operator Theory: Advances and Applications*. Birkhäuser Verlag, Basel, 1986. ISBN 3-7643-1774-4. URL http://www.worldcat.org/isbn/3764317744.

Theodore Wilbur Anderson. *An introduction to multivariate statistical analysis*. Wiley series in probability and mathematical statistics. Wiley, 1 edition, 1958. URL https://books.google.com/books?id=7YpqAAAAMAAJ.

Theodore Wilbur Anderson. *An Introduction to Multivariate Statistical Analysis*, volume 114 of *Wiley Series in Probability and Statistics—Applied Probability and Statistics Section Series*. Wiley, 2 edition, 1984. ISBN 9780471889878. URL http://books.google.com/books?vid=ISBN9780471889878.

John E. Angus. The probability integral transform and related results. *SIAM Review*, 36(4):652–654, February 25 1994. doi: 10.1137/1036146. URL http://dx.doi.org/10.1137/1036146.

Tom M. Apostol. *Mathematical Analysis*. Addison-Wesley series in mathematics. Addison-Wesley, Reading, 2 edition, 1975. ISBN 986-154-103-9. URL http://books.google.com/books?vid=ISBN0201002884.

Jean-Pierre Aubin. *Applied Functional Analysis*, volume 47 of *Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts*. John Wiley & Sons, 2 edition, September 30 2011. ISBN 9781118030974. URL http://books.google.com/books?vid=ISBN1118030974.

Jean-Pierre Aubin and Hélène Frankowska. *Set-Valued Analysis*. Modern Birkhäuser Classics. Springer, March 2 2009. ISBN 9780817648480. URL http://books.google.com/books?vid=ISBN0817648488.

Léon Autonne. Sur l'hermitien (on the hermitian). In *Comptes Rendus Des SéAnces De L'AcadéMie Des Sciences*, volume 133, pages 209–268. De L'Académie des sciences (Academy of Sciences), Paris, 1901. URL http://visualiseur.bnf.fr/Visualiseur?O=NUMM-3089. Comptes Rendus Des SéAnces De L'AcadéMie Des Sciences (Reports Of the Meetings Of the Academy of Science).

Léon Autonne. Sur l'hermitien (on the hermitian). *Rendiconti del Circolo Matematico di Palermo*, 16:104–128, 1902. Rendiconti del Circolo Matematico di Palermo (Statements of the Mathematical Circle of Palermo).

George Bachman. *Elements of Abstract Harmonic Analysis*. Academic paperbacks. Academic Press, New York, 1964. ISBN 9781483267562. URL http://books.google.com/books?vid=ISBN9781483267562.

George Bachman and Lawrence Narici. *Functional Analysis*. Academic Press textbooks in mathematics; Pure and Applied Mathematics Series. Academic Press, 1 edition, 1966. ISBN 9780486402512. URL http://books.google.com/books?vid=ISBN0486402517. "unabridged republication" available from Dover (isbn 0486402517).

George Bachman, Lawrence Narici, and Edward Beckenstein. *Fourier and Wavelet Analysis*. Universitext Series. Springer, 2002. ISBN 9780387988993. URL http://books.google.com/books?vid=ISBN9780387988993. "Corrected second printing, 2002".

David H. Bailey, Jonathan M. Borwein, Andrew Mattingly, and Glenn Wightwick. The computation of previously inaccessible digits of π^2 and catalan's constant. page 16, July 03 2011. URL https://carma.newcastle.edu.au/resources/jon/bbp-bluegene.pdf.

David H. Bailey, Jonathan M. Borwein, and Andrew Mattingly. 20. the computation of previously inaccessible digits of π^2 and catalan's constant (2013). In *Pi: The Next Generation: A Sourcebook on the Recent History of Pi and Its Computation*, pages 324–339. Springer, 2013a. ISBN 9783319323770. URL http://books.google.com/books?vid=ISBN3319323776.

David H. Bailey, Jonathan M. Borwein, Andrew Mattingly, and Glenn Wightwick. The computation of previously inaccessible digits of π^2 and catalan's constant. *Notices of the American Mathematical Society*, 60(7):844–854, June 10 2013b. URL https://www.ams.org/journals/notices/201307/rnoti-p844.pdf.

Om Prakash Bajpai. *Foundations of statistics*. Asia Publishing House, 1967. URL https://books.google.com/books?id=QrPuAAAAMAAJ.

Per Bak. *How Nature Works: the science of self-organized criticality*. Springer Science & Business Media, 2013. ISBN 9781475754261. URL http://books.google.com/books?vid=ISBN1475754264.

N Balakrishnan and Chin-Diew Lai. *Continuous Bivariate Distributions*. Springer Science & Business Media, 2 edition, May 31 2009. ISBN 9780387096148. URL http://books.google.com/books?vid=ISBN0387096140.

Stefan Banach. Sur les opérations dans les ensembles abstraits et leur applications aux équations intégrales (on abstract operations and their applications to the integral equations). *Fundamenta Mathematicae*, 3:133–181, 1922. URL http://matwbn.icm.edu.pl/ksiazki/fm/fm3/fm3120.pdf.

Stefan Banach. *Théorie des opérations linéaires*. Monografje Matematyczne, Warsaw, Poland, 1932a. URL http://matwbn.icm.edu.pl/kstresc.php?tom=1&wyd=10. (Theory of linear operations).

Stefan Banach. *Theory of Linear Operations*, volume 38 of *North-Holland mathematical library*. North-Holland, Amsterdam, 1932b. ISBN 0444701842. URL http://www.amazon.com/dp/0444701842/. English translation of 1932 French edition, published in 1987.

Alexander Barvinok. *A Course in Convexity*, volume 54 of *Graduate studies in mathematics*. American Mathematical Society, 2002. ISBN 9780821872314. URL http://books.google.com/books?vid=ISBN0821872311.

Gerald Beer. *Topologies on Closed and Closed Convex Sets*, volume 268 of *Mathematics and Its Applications*. Springer Science & Business Media, October 31 1993. ISBN 9780792325314. URL http://books.google.com/books?vid=ISBN0792325311.

Eric Temple Bell. *Men of Mathematics*. Simon & Schuster, New York, 1986. ISBN 9780671628185. URL http://books.google.com/books?vid=ISBN0671628186.

Richard Bellman. Dynamic programming of continuous processes. Technical Report R-271, Rand Corporation, 1954. URL https://books.google.com/books?id=sAwu9I2z4HoC.

Richard Bellman. *Adaptive Control Processes: A Guided Tour*. Rand Corporation Research studies. Princeton University Press, 1961. URL https://books.google.com/books?id=cXJEAAAAIAAJ.

Richard Bellman. *Introduction to the Mathematical Theory of Control Processes*, volume II: Nonlinear Processes of *Mathematics in Science and Engineering*. Academic Press, New York and London, 1971. URL https://archive.org/details/in.ernet.dli.2015.133515.

Irad Ben-Gal. Bayesian networks. In Fabrizio Ruggeri, Ron Kenett, Ron S. Kenett, and Frederick W. Faltin, editors, *Encyclopedia of Statistics in Quality and Reliability*, volume 1. John Wiley, 2008. ISBN 9780470018613. URL https://onlinelibrary.wiley.com/doi/full/10.1002/9780470061572.eqr089. http://www.eng.tau.ac.il/~bengal/BN.pdf.

Adi Ben-Israel and Robert P. Gilbert. *Computer-supported calculus*. Texts and monographs in symbolic computation. Springer, 2002. ISBN 3-211-82924-5. URL http://books.google.com/books?vid=ISBN3211829245.

J. S. Bendat. Statistical errors in measurement of coherence functions and input/output quantities. *Journal of Sound and Vibration*, 59(3):405–421, August 08 1978. doi: 10.1016/S0022-460X(78)80007-8. URL https://doi.org/10.1016/S0022-460X(78)80007-8.

Julius S. Bendat and Allan G. Piersol. *Measurement and Analysis of Random Data*. John Wiley & Sons, 1966.

Julius S. Bendat and Allan G. Piersol. *Engineering Applications of Correlation and Spectral Analysis*. John Wiley & Sons, 1980. ISBN 9780471058878. URL http://www.amazon.com/dp/0471058874.

Julius S. Bendat and Allan G. Piersol. *Engineering Applications of Correlation and Spectral Analysis*. Wiley-Interscience, 2 edition, 1993. ISBN 9780471570554. URL http://www.amazon.com/dp/0471570559.

Julius S. Bendat and Allan G. Piersol. *Random Data: Analysis and Measurement Procedures*, volume 729 of *Wiley Series in Probability and Statistics*. John Wiley & Sons, 4 edition, 2010. ISBN 9781118210826. URL http://books.google.com/books?vid=ISBN1118210824.

Elmer Bendiner. *The Fall of Fortresses*. Penguin Adult HC/TR, May 12 1980. ISBN 9780399123726. URL https://books.google.com/books?id=sNBzAAAAIAAJ.

Sterling Khazag Berberian. *Introduction to Hilbert Space*. Oxford University Press, New York, 1961. URL http://books.google.com/books?vid=ISBN0821819127.

Earl Berkson. Some metrics on the subspaces of a banach space. *Pacific Journal of Mathematics*, 13(1):7–22, 1963. URL http://projecteuclid.org/euclid.pjm/1103035953.

M. Bertero and P. Boccacci. *Introduction to Inverse Problems in Imaging*. CRC Press, 1998. ISBN 9781439822067. URL http://books.google.com/books?vid=ISBN9781439822067.

D. Besso. Teoremi elementari sui massimi i minimi. *Annuari Ist. Tech. Roma*, pages 7–24, 1879. see Bullen(2003) pages 453, 203.

Josef Betten. *Creep Mechanics*, chapter A. The Heaviside and Dirac Functions, pages 285–294. Springer Science & Business Media, 3 edition, 2008a. ISBN 9783540850519. URL https://link.springer.com/content/pdf/bbm:978-3-540-85051-9/1.pdf. https://www.springer.com/gp/book/9783540850502.

Josef Betten. *Creep Mechanics*, chapter B. The Laplace Transformation, pages 295–343. Springer Science & Business Media, 3 edition, 2008b. ISBN 9783540850519. URL https://link.springer.com/content/pdf/bbm:978-3-540-85051-9/1.pdf. https://www.springer.com/gp/book/9783540850502.

M. Jales Bienaymé. Société philomatique de paris—extraits des procès-verbaux. *Scéance*, pages 67–68, June 13 1840. URL http://www.archive.org/details/extraitsdesproc46183941soci. see Bullen(2003) pages 453, 203.

Patrick Billingsley. *Probability And Measure*. Wiley series in probability and mathematical statistics. Wiley, 3 edition, 1995. ISBN 978-0471007104. URL http://books.google.com/books?vid=ISBN0471007102.

Garrett Birkhoff and John Von Neumann. The logic of quantum mechanics. *The Annals of Mathematics*, 37(4):823–843, October 1936. URL http://www.jstor.org/stable/1968621.

Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer, New York, 1 edition, 2006. ISBN 9780387310732. URL http://users.isr.ist.utl.pt/~wurmd/Livros/school/Bishop%20-%20Pattern%20Recognition%20And%20Machine%20Learning%20-%20Springer%20%202006.pdf.

Leonard Mascot Blumenthal. *Theory and Applications of Distance Geometry*. Chelsea Publishing Company, Bronx, New York, USA, 2 edition, 1970. ISBN 0-8284-0242-6. URL http://books.google.com/books?vid=ISBN0828402426.

Béla Bollobás. *Linear Analysis; an introductory course*. Cambridge mathematical textbooks. Cambridge University Press, Cambridge, 2 edition, March 1 1999. ISBN 978-0521655774. URL http://books.google.com/books?vid=ISBN0521655773.

William M. Bolstad. *Introduction to Bayesian Statistics*. Wiley, 2 edition, 2007. ISBN 9780470141151. URL http://books.google.com/books?vid=ISBN9780470141151.

Umberto Bottazzini. *The Higher Calculus: A History of Real and Complex Analysis from Euler to Weierstrass*. Springer-Verlag, New York, 1986. ISBN 0-387-96302-2. URL http://books.google.com/books?vid=ISBN0387963022.

A. Böttcher and B. Silbermann. *Introduction to Large Truncated Toeplitz Matrices*. Springer, 1999. ISBN 9780387985701. URL http://books.google.com/books?vid=ISBN9780387985701.

V. Bouniakowsky. Sur quelques inégalités concernant les intégrales ordinaires et les intégrales aux différences finies. *Mémoires de lÁcad. de St.-Pétersbourg*, 1(9), 1859. URL http://www-stat.wharton.upenn.edu/~steele/Publications/Books/CSMC/CSMC_index.html.

George E. P. Box and Gwilym M. Jenkins. *Time Series Analysis: forecasting and control*. Holden-Day, revised edition, 1976. URL https://archive.org/details/in.ernet.dli.2015.134151/page/n2.

Carl Benjamin Boyer and Uta C. Merzbach. *A History of Mathematics*. Wiley, New York, 2 edition, 1991. ISBN 0471543977. URL http://books.google.com/books?vid=ISBN0471543977.

Carl Benjamin Boyer and Uta C. Merzbach. *A History of Mathematics*. Wiley, New York, 3 edition, January 11 2011. ISBN 9780470525487. URL http://books.google.com/books?vid=ISBN0470525487.

Ronald Newbold Bracewell. *The Fourier transform and its applications*. McGraw-Hill electrical and electronic engineering series. McGraw-Hill, 2, illustrated, international student edition edition, 1978. ISBN 9780070070134. URL http://books.google.com/books?vid=ISBN007007013X.

Thomas John I'Anson Bromwich. *An Introduction to the Theory of Infinite Series*. Macmillan and Company, 1 edition, 1908. ISBN 9780821839768. URL http://www.archive.org/details/anintroductiont00bromgoog.

P. S. Bullen. *Handbook of Means and Their Inequalities*, volume 560 of *Mathematics and Its Applications*. Kluwer Academic Publishers, Dordrecht, Boston, 2 edition, 2003. ISBN 9781402015229. URL http://books.google.com/books?vid=ISBN1402015224.

Viktor Yakovlevich Bunyakovsky. Sur quelques inégalités concernant les intégrales ordinaires et les intégrales aux différences finies. *Mémoires de L'Académie Impériale des Sciences de St.-Pétersbourg*, 1(9):1–18, 1859. URL http://www-stat.wharton.upenn.edu/~steele/Publications/Books/CSMC/bunyakovsky.pdf. (Mémoires of the Imperial Academy of Sciences of Saint Petersburg); (On Some Inequalities concerning the Ordinary Integrals and Integrals with the Finite Differences).

Dmitri Burago, Yuri Burago, and Sergei Ivanov. *A Course in Metric Geometry*, volume 33 of *Graduate Studies in Mathematics*. American Mathematical Society, 2001. ISBN 978-0821821299. URL http://books.google.com/books?vid=ISBN0821821296.

Charles L. Byrne. *Signal Processing: A Mathematical Approach*. AK Peters Series. A K Peters, 2005. ISBN 1568812426.

James A. Cadzow. *Foundations of Digital Signal Processing and Data Analysis*. Macmillan Publishing Company, New York, 1987. ISBN 0023180102. URL https://books.google.com/books?vid=ISBN9780023180101.

Peter E. Caines. *Linear Stochastic Systems*, volume 100 of *Wiley Series in Probability and Statistics*. J. Wiley, 1988. ISBN 9780471081012. URL http://books.google.com/books?vid=ISBN0471081019.

Peter E. Caines. *Linear Stochastic Systems*, volume 77 of *Classics in Applied Mathematics*. SIAM, 2018. ISBN 9781611974713. URL http://books.google.com/books?vid=ISBN1611974712.

Florian Cajori. A history of mathematical notations; notations mainly in higher mathematics. In *A History of Mathematical Notations; Two Volumes Bound as One*, volume 2. Dover, Mineola, New York, USA, 1993. ISBN 0-486-67766-4. URL http://books.google.com/books?vid=ISBN0486677664. reprint of 1929 edition by *The Open Court Publishing Company*.

James V. Candy. *Signal processing: the model-based approach*. McGraw-Hill Series in Electrical and Computer Engineering. McGraw-Hill Book Company, 1985. ISBN 9780070097254.

James V. Candy. *Signal processing: the modern approach*. McGraw-Hill Series in Electrical and Computer Engineering. McGraw-Hill Book Company, 1 edition, 1988. ISBN 9780070097513.

James V. Candy. *Model-Based Signal Processing*, volume 36 of *Adaptive and Cognitive Dynamic Systems: Signal Processing, Learning, Communications and Control*. John Wiley & Sons, 2005. ISBN 9780471732662. URL http://books.google.com/books?vid=ISBN9780471732662.

James V. Candy. *Bayesian Signal Processing: Classical, Modern, and Particle Filtering Methods*, volume 54 of *Adaptive and Cognitive Dynamic Systems: Signal Processing, Learning, Communications and Control*. Wiley, 2009. ISBN 9780470430576. URL http://books.google.com/books?vid=ISBN9780470430576.

James V. Candy. *Bayesian Signal Processing: Classical, Modern, and Particle Filtering Methods*, volume 54 of *Adaptive and Cognitive Dynamic Systems: Signal Processing, Learning, Communications and Control*. John Wiley & Sons, 2 edition, 2016. ISBN 9781119125457. URL http://books.google.com/books?vid=ISBN9781119125457. http://www.amazon.com/dp/9781119125457.

James V. Candy. *Model-Based Processing: An Applied Subspace Identification Approach*. John Wiley & Sons, 2019. ISBN 9781119457763. URL http://books.google.com/books?vid=ISBN9781119457763.

James V. Candy and E. J. Sullivan. Ocean acoustic signal processing: A model-based approach, August 1992. URL https://doi.org/10.1121/1.404168.

Gerolamo Cardano. *Ars Magna or the Rules of Algebra*. Dover Publications, Mineola, New York, 1545. ISBN 0486458733. URL http://www.amazon.com/dp/0486458733. English translation of the Latin *Ars Magna* edition, published in 2007.

René Carmona. *Statistical Analysis of Financial Data in R*. Springer Texts in Statistics. Springer Science & Business Media, 2 edition, 2013. ISBN 9781461487883. URL http://books.google.com/books?vid=ISBN9781461487883.

Thomas G. Carne and C. R. Dohrmann. Improving experimental frequency response function matrices for admittance modeling. In *Proceedings of the XXIV International Modal Analysis Conference (IMAC)*, St. Louis, Missouri, USA, 2006. URL https://www.researchgate.net/publication/265790751.

N. L. Carothers. *A Short Course on Banach Space Theory*. Number 64 in London Mathematical Society Student Texts. Cambridge University Press, Cambridge, 2005. ISBN 0521842832. URL http://www.comm.musashi-tech.ac.jp/~arimoto/Drarimoto/preprints/calothers/Assemble.ps.

N.L. Carothers. *Real Analysis*. Cambridge University Press, Cambridge, 2000. ISBN 978-0521497565. URL http://books.google.com/books?vid=ISBN0521497566.

Augustin-Louis Cauchy. *Part 1: Analyse Algebrique*. Cours D'Analyse de L'école Royale Polytechnique. Debure frères, Paris, 1821. ISBN 2-87647-053-5. URL http://www.archive.org/details/coursdanalysede00caucgoog. Systems design course of the Polytechnic Royal School; 1st Part: Algebraic analysis).

Arthur Cayley. A memoir on the theory of matrices. *Philosophical Transactions of the Royal Society of London*, 148:17–37, 1858. ISSN 1364-503X. URL http://www.jstor.org/view/02610523/ap000059/00a00020/0.

Joshua Chapmann. *Markov Models: Introduction to Markov Chains, Hidden Markov Models and Bayesian Networks*. Advanced Data Analytics. CreateSpace Independent Publishing Platform, 2017. ISBN 9781978304871. URL https://www.amazon.com/dp/1978304870.

Si-Wei Chen, Xue-Song Wang, and Motoyuki Sato. Polinsar complex coherence estimation based on covariance matrix similarity test. *IEEE Transactions on Geoscience and Remote Sensing*, 50(11):4699–4710, November 2012. doi: 10.1109/TGRS.2012.2192937. URL https://www.researchgate.net/profile/Si_Wei_Chen/publication/281271737.

Sung C. Choi. *Introductory applied statistics in science*. Prentice-Hall, 1978. ISBN 9780135016190. URL http://books.google.com/books?vid=ISBN9780135016190.

Alexandre J. Chorin and Ole H. Hald. *Stochastic Tools in Mathematics and Science*, volume 1 of *Surveys and Tutorials in the Applied Mathematical Sciences*. Springer, New York, 2 edition, 2009. ISBN 978-1-4419-1001-1. URL http://books.google.com/books?vid=ISBN9781441910011.

Ole Christensen. *An Introduction to Frames and Riesz Bases*. Applied and Numerical Harmonic Analysis. Birkhäuser, Boston/Basel/Berlin, 2003. ISBN 0-8176-4295-1. URL http://books.google.com/books?vid=ISBN0817642951.

Ole Christensen. *Frames and bases: An Introductory Course*. Applied and Numerical Harmonic Analysis. Birkhäuser, Boston/Basel/Berlin, 2008. ISBN 9780817646776. URL http://books.google.com/books?vid=ISBN0817646779.

Ole Christensen. *Functions, Spaces, and Expansions: Mathematical Tools in Physics and Engineering*. Applied and Numerical Harmonic Analysis. Birkhäuser, Boston/Basel/Berlin, 2010. ISBN 978-0-8176-4979-1. doi: 10.1007/978-0-8176-4980-7. URL http://books.google.com/books?vid=ISBN0817649794.

Charles K. Chui. *Multivariate Splines*, volume 54 of *CBMS-NSF regional conference series in applied mathematics*. SIAM, 1988. ISBN 1611970172. URL http://books.google.com/books?vid=ISBN1611970172.

Charles K. Chui. *An Introduction to Wavelets*. Academic Press, San Diego, California, USA, January 3 1992. ISBN 9780121745844. URL http://books.google.com/books?vid=ISBN0121745848.

Donato Michele Cifarelli and Eugenio Regazzini. De finetti's contribution to probability and statistics. *Statistical Science*, 11(4):253–282, 1996. URL http://www.jstor.org/stable/2246020.

Erhan Çinlar and Robert J Vanderbei. *Real and Convex Analysis*. Undergraduate Texts in Mathematics. Springer, January 4 2013. ISBN 1461452570. URL http://books.google.com/books?vid=ISBN1461452570.

Jon F. Claerbout. *Fundamentals of Geophysical Data Processing with Applications to Petroleum Prospecting*. International series in the earth and planetary sciences, Tab Mastering Electronics Series. McGraw-Hill, New York, 1976. ISBN 9780070111172. URL http://sep.stanford.edu/sep/prof/.

James A. Clarkson. Uniformly convex spaces. *Transactions of the American Mathematical Society*, 40(3):396–414, December 1936. URL http://www.jstor.org/stable/1989630.

Peter M. Clarkson. *Optimal and Adaptive Signal Processing*. Electronic Engineering Systems Series. CRC Press, 1993. ISBN 0849386098. URL http://books.google.com/books?vid=isbn0849386098.

R. E. Cobb and L. D. Mitchell. Estimation of uncorrelated content in experimentally measured frequency response functions using three measurement channels. *Mechanical Systems and Signal Processing*, 4(6):449–461, November 1990. URL https://doi.org/10.1016/0888-3270(90)90045-M.

Richard E. Cobb. *Confidence Bands, Measurement Noise, and Multiple Input - Multiple Output Measurements Using the Three-Channel Frequency Response Function Estimator*. Ph.d. dissertation, Virginia Polytechnic Institute and State University, May 1988. URL http://hdl.handle.net/10919/53675. "Dissertation submitted to the Faculty of the Virginia Polytechnic Institute and State University in partial fulullment of the requirements for the degree of Doctor of Philosophy in Mechanical Engineering".

Paul M. Cohn. *Basic Algebra; Groups, Rings and Fields*. Springer, December 6 2002. ISBN 1852335874. URL http://books.google.com/books?vid=isbn1852335874.

John B. Conway. *A Course in Functional Analysis*, volume 96 of *Graduate Texts in Mathematics*. Springer, 2 edition, 1990. ISBN 0-387-97245-5. URL http://books.google.com/books?vid=ISBN0387972455.

R. Courant and D. Hilbert. *Methods of Mathematical Physics*, volume 1. Interscience Publishers, New York, 1930. URL http://www.worldcat.org/isbn/0471504475.

T.M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., New York, 1991. ISBN 0471062596. URL http://www.amazon.com/dp/0471062596.

M. G. Cox. The numerical evaluation of b-splines. *IMA Journal of Applied Mathematics*, 10(2):134–149, 1972. doi: 10.1093/imamat/10.2.134. URL http://imamat.oxfordjournals.org/content/10/2/134.abstract.

John P. Cunningham and Zoubin Ghahramani. Linear dimensionality reduction: Survey, insights, and generalizations. *Journal of Machine Learning Research*, 16:2859–2900, December 2015. URL http://www.jmlr.org/papers/volume16/cunningham15a/cunningham15a.pdf.

Guy L. Curry and Richard M. Feldman. *Manufacturing Systems Modeling and Analysis*. Technology & Engineering. Springer Science & Business Media, 2 edition, 2010. ISBN 9783642166181. URL http://books.google.com/books?vid=ISBN3642166180.

Adnan Darwiche. *Modeling and Reasoning with Bayesian Networks*. Cambridge University Press, reprint edition, 2009. ISBN 9780521884389. URL http://books.google.com/books?vid=ISBN9780521884389.

Ingrid Daubechies. *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, Philadelphia, 1992. ISBN 0-89871-274-2. URL http://www.amazon.com/dp/0898712742.

Ingrid Daubechies, A. Grossman, and Y. Meyer. Painless nonorthogonal expansions. *Journal of Mathematical Physics*, 27(5):1271–1283, May 1986. ISSN 0022-2488. URL link.aip.org/link/?jmp/27/1271.

Mahlon M. Day. Some characterizations of inner-product spaces. *Transactions of the American Mathematical Society*, 62(2):320–337, September 1947. URL http://www.jstor.org/stable/1990458.

Mahlon Marsh Day. *Normed Linear Spaces*, volume 21 of *Ergebnisse der Mathematik und ihrer Grenzgebiete*. Springer-Verlag, Berlin, Heidelberg, New York, 3 edition, 1973. ISBN 0387061487. URL http://books.google.com/books?id=UZ3vSAAACAAJ.

Bruno de Finetti. Sui passaggi al limite nel calcolo delle probabilità. *Rendiconti del Reale Istituto Lombardo di Scienze e Lettere*, 63:155–166, 1930a.

Bruno de Finetti. A proposito dell'estensione del teorema delle probabilita totali alle classi numerabili. *Rendiconti del Reale Istituto Lombardo di Scienze e Lettere*, 63:901–905, 1930b.

Bruno de Finetti. Ancora sull'estensione alle classi numerabili del teorema delle probabilità totali. *Rendiconti del Reale Istituto Lombardo di Scienze e Lettere*, 63:1063–1069, 1930c.

Charles Jean de la Vallée-Poussin. Sur l'intégrale de lebesgue. *Transactions of the American Mathematical Society*, 16(4):435–501, October 1915. URL http://www.jstor.org/stable/1988879.

Johan de Witt. Elementa curvarum linearum (elements of curves). In *La Géométrie*. 1659. Elementa curvarum linearum was an appendix to de Witt's translation of of Descartes' *La Géométrie*.

Carl R. de Boor. On calculating with b-splines. *Journal of Approximation Theory*, pages 50–62, 1972.

Carl R. de Boor. *A Practical Guide to Splines*, volume 27 of *Applied Mathematical Sciences*. Springer, revised edition, 2001. ISBN 0387953663. URL http://books.google.com/books?vid=ISBN0387953663.

F.M. Dekking, C. Kraaikamp, H.P. Lopuhaa, and L.E. Meester. *A Modern Introduction to Probability and Statistics: Understanding Why and How*. Springer Texts in Statistics. Springer Science & Business Media, 2006. ISBN 9781846281686. URL http://books.google.com/books?vid=ISBN1846281687.

René Descartes. *La géométrie*. 1637. URL http://historical.library.cornell.edu/math/math_D.html.

René Descartes. *Regulae ad directionem ingenii*. 1684a. URL http://www.fh-augsburg.de/~harsch/Chronologia/Lspost17/Descartes/des_re00.html.

René Descartes. *Rules for Direction of the Mind*. 1684b. URL http://en.wikisource.org/wiki/Rules_for_the_Direction_of_the_Mind.

René Descartes. *The Geometry of Rene Descartes*. Courier Dover Publications, June 1 1954. ISBN 0486600688. URL http://books.google.com/books?vid=isbn0486600688. orginally published by Open Court Publishing, Chicago, 1925; translation of La géométrie.

Analog Devices. Adl5387. Technical report, 2016. URL https://www.analog.com/media/en/technical-documentation/data-sheets/ADL5387.pdf. "30 MHz to 2 GHzQuadrature Demodulator".

Luc Devroye. *Non-Uniform Random Variate Generation*. Springer-Verlag, New York, 1986. ISBN 0387963057. URL http://www.eirene.de/Devroye.pdf.

Luc Devroye, Laszlo Györfi, and Gabor Lugosi. *A Probabilistic Theory of Pattern Recognition*, volume 31 of *Stochastic Modelling and Applied Probability*. Springer Science & Business Media, 2013. ISBN 9781461207115.

Elena Deza and Michel-Marie Deza. *Dictionary of Distances*. Elsevier Science, Amsterdam, 2006. ISBN 0444520872. URL http://books.google.com/books?vid=ISBN0444520872.

Jean Alexandre Dieudonné. *Foundations of Modern Analysis*. Academic Press, New York, 1969. ISBN 1406727911. URL http://books.google.com/books?vid=ISBN1406727911.

Ivan H. Dimovski. *Convolutional Calculus*, volume 43 of *Mathematics and its Applications*. Springer Science & Business Media, 2012. ISBN 9789400905276.

Gustav Doetsch. *Einführung in Theorie und Anwendung der Laplace-Transformation*, volume 24 of *Mathematische Reihe*. Birkhäuser Verlag Basel Und Stuttgart, 1958. URL https://archive.org/details/einfuhrungintheo0000doet/mode/2up.

J. Alejandro Dominguez-Torres. The origin and history of convolution i: continuous and discrete convolution operations. online, November 02 2010. URL http://www.slideshare.net/Alexdfar/origin-adn-history-of-convolution.

J. Alejandro Dominguez-Torres. A history of the convolution operation. *IEEE Pulse, Retroscope*, January/February 2015. URL https://www.embs.org/pulse/articles/history-convolution-operation/.

Hamid Drljević. On the representation of functionals and the stability of mappings in hilbert and banach spaces. In Themistocles M Rassias, editor, *Topics in mathematical analysis: a volume dedicated to the memory of A.L. Cauchy*, volume II of *Series in Pure Mathematics*, pages 231–245. World Scientific Publishing Company, 1989. ISBN 9971506661. URL http://books.google.com/books?vid=ISBN9971506661.

R. J. Duffin and A. C. Schaeffer. A class of nonharmonic fourier series. *Transactions of the American Mathematical Society*, 72(2):341–366, March 1952. ISSN 1088-6850. URL http://www.jstor.org/stable/1990760.

Bogdan Dumitrescu. *Positive Trigonometric Polynomials and Signal Processing Applications*. Signals and Communication Technology. Springer, 2007. ISBN 978-1-4020-5124-1. URL gen.lib.rus.ec/get?md5=5346e169091b2d928d8333cd053300f9.

Nelson Dunford and Jacob T. Schwartz. *Linear operators. Part 1, General Theory*, volume 7 of *Pure and applied mathematics*. Interscience Publishers, New York, 1957. ISBN 0471226394. URL http://www.amazon.com/dp/0471608483. with the assistance of William G. Bade and Robert G. Bartle.

John R. Durbin. *Modern Algebra; An Introduction*. John Wiley & Sons, Inc., 4 edition, 2000. ISBN 0-471-32147-8. URL http://www.worldcat.org/isbn/0471321478.

W.D. Duthie. Segments of ordered sets. *Transactions of the American Mathematical Society*, 51(1): 1–14, January 1942. doi: 10.2307/1989978. URL http://www.jstor.org/stable/1989978.

Jacques Dutka. On gauss' priority in the discovery of the method of least squares. *Archive for History of Exact Sciences*, 49(4):355–370, December 1995. URL https://www.jstor.org/stable/41134010.

C. Eckart and G. Young. *Psychometrika*, 1(3):211–218, 1936. ISSN 0033-3123. doi: https://doi.org/10.1007/BF02288367. URL http://www.math.zju.edu.cn:8080/wjd/TN/T3.pdf.

Paul H. Edelman. Abstract convexity and meet-distributive lattices. In Ivan Rival, editor, *Combinatorics and ordered sets: Proceedings of the AMS-IMS-SIAM joint summer research conference, held August 11–17, 1985*, volume 57 of *Contemporary Mathematics*, pages 127–150, Providence RI, 1986. American Mathematical Society. ISBN 0821850512. URL http://books.google.com/books?vid=ISBN0821850512. conference held in Arcata California.

Paul H. Edelman and Robert E. Jamison. The theory of convex geometries. *Geometriae Dedicata*, 19(3):247–270, December 1985. ISSN 0046-5755. doi: 10.1007/BF00149365. URL http://www.springerlink.com/content/n4344856887387gw/.

Robert E. Edwards. *Functional Analysis: Theory and Applications*. Dover books on mathematics. Dover, New York, 1995. ISBN 0-486-68143-2. URL http://books.google.com/books?vid=ISBN0486681432.

Yuli Eidelman, Vitali D. Milman, and Antonis Tsolomitis. *Functional Analysis: An Introduction*, volume 66 of *Graduate Studies in Mathematics*. American Mathematical Society, 2004. ISBN 0821836463. URL http://books.google.com/books?vid=ISBN0821836463.

Per Enflo. A counterexample to the approximation problem in banach spaces. *Acta Mathematica*, 130:309–317, 1973. URL http://link.springer.com/content/pdf/10.1007/BF02392270.

Euclid. *Elements*. circa 300BC. URL http://farside.ph.utexas.edu/euclid.html.

Leonhard Euler. *Introductio in analysin infinitorum*, volume 1. Marcum-Michaelem Bousquet & Socios, Lausannæ, 1748. URL http://www.math.dartmouth.edu/~euler/pages/E101.html. Introduction to the Analysis of the Infinite.

Leonhard Euler. *Introduction to the Analysis of the Infinite*. Springer, 1988. ISBN 0387968245. URL http://books.google.com/books?vid=ISBN0387968245. translation of 1748 Introductio in analysin infinitorum.

David Ewen. *The Book of Modern Composers*. Alfred A. Knopf, New York, 1950. URL http://books.google.com/books?id=yHw4AAAAIAAJ.

David Ewen. *The New Book of Modern Composers*. Alfred A. Knopf, New York, 3 edition, 1961. URL http://books.google.com/books?id=bZIaAAAAMAAJ.

D. J. Ewins. *Modal Testing: Theory and Practice*. Research Studies Press, "reprinted with amendments april 1986" edition, April 1986. ISBN 0863800173. URL http://www.amazon.com/dp/086380036X.

Marián Fabian, Petr Habala, Petr Hájek, Vicente Montesinos, Vaclav Zizler, and Václav Zizler. *Banach Space Theory: The Basis for Linear and Nonlinear Analysis*. CMS Books in Mathematics. Springer, 2010. ISBN 1441975144. URL http://books.google.com/books?vid=isbn1441975144.

Robert M. Fano. A heuristic discussion of probabilistic decoding. *IEEE Transactions on Information Theory*, 9(2):64–74, April 1963. doi: 10.1109/TIT.1963.1057827. URL https://doi.org/10.1109/TIT.1963.1057827.

Lorenzo Farina and Sergio Rinaldi. *Positive Linear Systems: Theory and Applications*. Pure and applied mathematics. John Wiley & Sons, 1 edition, July 3 2000. ISBN 9780471384564. URL http://books.google.com/books?vid=ISBN0471384569.

Richard M. Feldman and Ciriaco Valdez-Flores. *Applied Probability and Stochastic Processes*. Technology & Engineering. Springer Science & Business Media, 2 edition, 2010. ISBN 9783642051586. URL http://books.google.com/books?vid=ISBN3642051588.

Carlos A. Felippa. *Matrix Calculus*. University of Colorado at Boulder, August 18 1999. URL http://caswww.colorado.edu/courses.d/IFEM.d/.

William Feller. *An Introduction to Probability Theory and its Applications Volume I*. Wiley series in probability and mathematical statistics. John Wiley & Sons, 3, revised edition, 1970. ISBN 9780471257080. URL http://www.amazon.com/dp/0471257087.

William Feller. *An Introduction to Probability Theory and its Applications Volume II*. John Wiley & Sons/ Mei Ya Publications, New York/Taipei, 2 edition, 1971.

Norman Fenton and Martin Neil. *Risk Assessment and Decision Analysis with Bayesian Networks*. CRC Press, 2 edition, 2018. ISBN 9781351978965. URL http://books.google.com/books?vid=ISBN9781351978965.

R. A. Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transacations of the Royal Society*, January 1922. URL https://doi.org/10.1098/rsta.1922.0009.

G.L. Fix and G. Strang. Fourier analysis of the finite element method in ritz-galerkin theory. *Studies in Applied Mathematics*, 48:265–273, 1969.

Harley Flanders. Differentiation under the integral sign. *The American Mathematical Monthly*, 80(6):615–627, June–July 1973. URL http://sgpwe.izt.uam.mx/files/users/uami/jdf/proyectos/Derivar_inetegral.pdf. http://www.jstor.org/pss/2319163.

Francis J. Flanigan. *Complex Variables; Harmonic and Analytic Functions*. Dover, New York, 1983. ISBN 9780486613888. URL http://books.google.com/books?vid=ISBN0486613887.

Gerald B. Folland. *Fourier Analysis and its Applications*. Wadsworth & Brooks / Cole Advanced Books & Software, Pacific Grove, California, USA, 1992. ISBN 0-534-17094-3. URL http://www.worldcat.org/isbn/0534170943.

Gerald B. Folland. *A Course in Abstract Harmonic Analysis*. Studies in Advanced Mathematics. CRC Press, Boca Raton, 1995. ISBN 0-8493-8490-7. URL http://books.google.com/books?vid=ISBN0849384907.

Brigitte Forster and Peter Massopust, editors. *Four Short Courses on Harmonic Analysis: Wavelets, Frames, Time-Frequency Methods, and Applications to Signal and Image Analysis*. Applied and Numerical Harmonic Analysis. Springer, November 19 2009. ISBN 9780817648909. URL http://books.google.com/books?vid=ISBN0817648909.

Jean-Baptiste-Joseph Fourier. Refroidissement séculaire du globe terrestre". In M. Gaston Darboux, editor, *Œuvres De Fourier*, volume 2, pages 271–288. Ministère de L'instruction Publique, Paris, France, April 1820. URL http://gallica.bnf.fr/ark:/12148/bpt6k33707/f276.image. original paper at pages 58–70.

Jean-Baptiste-Joseph Fourier. *Théorie Analytique de la Chaleur (The Analytical Theory of Heat)*. Chez Firmin Didot, pere et fils, Paris, 1822. URL http://books.google.com/books?vid=04X2vlqZx7hydlQUWEq&id=TDQJAAAAIAAJ.

Jean-Baptiste-Joseph Fourier. *The Analytical Theory of Heat (Théorie Analytique de la Chaleur)*. Cambridge University Press, Cambridge, February 20 1878. URL http://www.archive.org/details/analyticaltheory00fourrich. 1878 English translation of the original 1822 French edition. A 2003 Dover edition is also available: isbn 0486495310.

Maurice Fréchet. Sur l'extension du théorème des probabilités totales au cas d'une suite infinie d'evénéments. *Ren- diconti del Reale Istituto Lombardo di Scienze e Lettere*, 63:899–900, 1930a.

Maurice Fréchet. Sur l'extension du théorème des probabilités totales au cas d'une suite infinie d'evénéments. *Ren- diconti del Reale Istituto Lombardo di Scienze e Lettere*, 63:1059–1062, 1930b.

Erik Ivar Fredholm. Sur une nouvelle méthode pour la résolution du problème de dirichlet (on a new method for the resolution of the problem of dirichlet). */:Oefversigt af Kongl. Sv. Vetenskaps-Academiens Förhandlingar*, 57:39–66, 1900.

Erik Ivar Fredholm. Sur une classe d'equations fonctionnelles (on a class of functional equations). *Acta Mathematica*, 27(1):365–390, December 1903. ISSN 0001-5962. doi: 10.1007/BF02421317. URL http://www.springerlink.com/content/c41371137837p252/.

Ferdinand Georg Frobenius. Uber lineare substitutionen und bilineare formen. *Journal für die reine und angewandte Mathematik (Crelle's Journal)*, 84:1–63, 1878. ISSN 0075-4102. URL http://www.digizeitschriften.de/home/services/pdfterms/?ID=509796.

Ferdinand Georg Frobenius. Uber lineare substitutionen und bilineare formen. In Jean Pierre Serre, editor, *Gesammelte Abhandlungen (Collected Papers)*, volume I, pages 343–405. Springer, Berlin, 1968. URL http://www.worldcat.org/oclc/253015. reprint of Frobenius' 1878 paper.

Jürgen Fuchs. *Affine Lie Algebras and Quantum Groups: An Introduction, With Applications in Conformal Field Theory*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, 1995. ISBN 052148412X. URL http://books.google.com/books?vid=ISBN052148412X.

Israel M. Gelfand. Normierte ringe. *Mat. Sbornik*, 9(51)(1):3–24, 1941. URL http://www.mathnet.ru/links/d19894c5545d2f75b3e1037162f86d58/sm6046.pdf.

Israel M. Gelfand and Mark A. Naimark. Normed rings with an involution and their representations. In *Commutative Normed Rings*, number 170 in AMS Chelsea Publishing Series, pages 240–274. Chelsea Publishing Company, Bronx, 1964. ISBN 9780821820223. URL http://books.google.com/books?vid=ISBN0821820222.

Israel M. Gelfand and Mark A. Neumark. On the imbedding of normed rings into the ring of operators in hilbert space. *Mat. Sbornik*, 12(54:2):197–217, 1943a.

Israel M. Gelfand and Mark A. Neumark. On the imbedding of normed rings into the ring of operators in hilbert space. In Robert S. Doran, editor, *C*-algebras: 1943–1993: a Fifty Year Celebration: Ams Special Session Commenorating the First Fifty Years of C*-Algebra Theory January 13–14, 1993*, pages 3–19. American Mathematical Society, 1943b. ISBN 978-0821851753. URL http://books.google.com/books?vid=ISBN0821851756.

Israel M. Gelfand, R. A. Minlos, and Z. .Ya. Shapiro. *Representations of the rotation and Lorentz groups and their applications*. Courier Dover Publications, reprint edition, 2018. ISBN 9780486823850. Izrail' Moiseevich Gel'fand.

Izrail' Moiseevich Gel'fand. *Representations of the rotation and Lorentz groups and their applications*. Pergamon Press book, 1963. 2018 Dover edition available.

John Robilliard Giles. *Introduction to the Analysis of Metric Spaces*. Number 3 in Australian Mathematical Society lecture series. Cambridge University Press, Cambridge, 1987. ISBN 978-0521359283. URL http://books.google.com/books?vid=ISBN0521359287.

John Robilliard Giles. *Introduction to the Analysis of Normed Linear Spaces*. Number 13 in Australian Mathematical Society lecture series. Cambridge University Press, Cambridge, 2000. ISBN 0-521-65375-4. URL http://books.google.com/books?vid=ISBN0521653754.

Israel Gohberg, Seymour Goldberg, and Marinus A. Kaashoek. *Basic Classes of Linear Operators*. Birkhäuser, Basel, 1 edition, 2003. ISBN 3764369302. URL http://books.google.com/books?vid=ISBN3764369302.

Steve Goldman. *Vibration Spectrum Analysis: A Practical Approach*. Industrial Press Inc., 1999. ISBN 9780831130886. URL http://books.google.com/books?vid=ISBN0831130881.

Jaideva C. Goswami and Andrew K. Chan. *Fundamentals of Wavelets; Theory, Algorithms, and Applications*. John Wiley & Sons, Inc., 1999. ISBN 9780471197485. URL http://vadkudr.boom.ru/Collection/fundwave_contents.html.

H. G. D. Goyder. Foolproof methods for frequency response measurements. Technical Report ADP003677, United Kingdom Atomic Energy Authority London (England), 1984. URL https://apps.dtic.mil/dtic/tr/fulltext/u2/a143301.pdf#page=118.

I.S. Gradshteyn and I.M. Ryzhik. *Table of Integrals, Series, and Products*. 1980. ISBN 0-12-294760-6. URL http://www.worldcat.org/isbn/0122947606.

Alexander Graham. *Kronecker Products and Matrix Calculus: With Applications*. Ellis Horwood Series; Mathematics and its Applications. Ellis Horwood Limited, Chichester, 1981. ISBN 0-85312-391-8. URL http://books.google.com/books?vid=ISBN0853123918.

Alexander Graham. *Kronecker Products and Matrix Calculus: With Applications*. Dover Books on Mathematics. Courier Dover Publications, 2018. ISBN 9780486824178. URL http://books.google.com/books?vid=ISBN9780486824178.

Artur Gramacki. *Nonparametric Kernel Density Estimation and Its Computational Aspects*, volume 37 of *Studies in Big Data*. Springer, 2017. ISBN 9783319716886. URL http://books.google.com/books?vid=ISBN9783319716886.

Robert M. Gray. Toeplitz and circulant matrices: A review. Technical Report AD0727139, Stanford University California Stanford Electronics Labs, Norwell, Massachusetts, June 1971. URL https://apps.dtic.mil/docs/citations/AD0727139.

Robert M. Gray. Toeplitz and circulant matrices: A review. *Foundations and Trends® in Communications and Information Theory*, 2(3):155–239, January 31 2006. doi: http://dx.doi.org/10.1561/0100000006. URL https://ee.stanford.edu/~gray/toeplitz.pdf.

R.J Greechie. Orthomodular lattices admitting no states. *Journal of Combinatorial Theory, Series A*, 10(2):119–132, March 1971. URL http://dx.doi.org/10.1016/0097-3165(71)90015-X.

Daniel J. Greenhoe. *Structure and Analysis of Mathematical Spaces*, volume 2 of *Mathematical Structure and Design series*. ResearchGate, version 0.30 edition, January 2017a. URL https://www.researchgate.net/publication/312201065.

Daniel J. Greenhoe. The sequence a(n,m) of the m polynomial coefficients of the n-th order b-spline scaled by n!, read by rows, with n in 0,1,2,... and m in 1,2,3,...,(n+1)2. *Online Encyclopedia of Integer Sequences*, (A289358), July 2017b. URL https://oeis.org/A289358.

Daniel J. Greenhoe. The effects of the assorted cross-correlation definitions. *ResearchGate.com*, June 2019a. doi: 10.13140/RG.2.2.15881.70240. URL https://www.researchgate.net/publication/333882235.

Daniel J. Greenhoe. *A Book Concerning Digital Communications*, volume 4 of *Signal Processing ABCs*. 2019b. URL https://github.com/dgreenhoe/pdfs/blob/master/abcdc.pdf.

Ulf Grenander and Gábor Szegö. *Toeplitz Forms and Their Applications*. California monographs in mathematical sciences. University of California Press, 1958. URL https://www.worldcat.org/title/toeplitz-forms-and-their-applications/oclc/648601989.

Ulf Grenander and Gábor Szegö. *Toeplitz Forms and Their Applications*. Chelsea Publishing Company, 2 edition, 1984. ISBN 9780828403214. URL http://books.google.com/books?vid=ISBN9780828403214.

Stanley Gudder. *Stochastic Methods in Quantum Mechanics*. North Holland, 1979. ISBN 0444002995. URL http://books.google.com/books?vid=ISBN0444002995.

Stanley Gudder. *Stochastic Methods in Quantum Mechanics*. Dover, Mineola NY, 2005. ISBN 0486445321. URL http://books.google.com/books?vid=ISBN0486445321.

Amritava Gupta. *Real & Abstract Analysis*. Academic Publishers Calcutta, 1998. ISBN 9788186358443. URL http://books.google.com/books?vid=ISBN8186358447. first published in 1998, reprinted in 2000 and 2006.

Alfréd Haar. Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*, 69:331–371, September 1910. ISSN 1432-1807. doi: 10.1007/BF01456326.

Norman B. Haaser and Joseph A. Sullivan. *Real Analysis*. Dover Publications, New York, 1991. ISBN 0-486-66509-7. URL http://books.google.com/books?vid=ISBN0486665097.

Paul R. Haddad and Ali N. Akansu. *Multiresolution Signal Decomposition: Transforms, Subbands, and Wavelets*. Acedemic Press, October 1 1992. ISBN 0323138365. URL http://books.google.com/books?vid=ISBN0323138365.

Henry Sinclair Hall and Samuel Ratcliffe Knight. *Higher algebra, a sequel to elementary algebra for schools*. Macmillan, London, 1894. URL http://www.archive.org/details/higheralgebraas00kniggoog.

Paul R. Halmos. *Finite Dimensional Vector Spaces*. Princeton University Press, Princeton, 1 edition, 1948. ISBN 0691090955. URL http://books.google.com/books?vid=isbn0691090955.

Paul R. Halmos. *Finite Dimensional Vector Spaces*. Springer Science & Business Media, New York, 1 edition, 1958. ISBN 9781461263876. URL http://books.google.com/books?vid=isbn9781461263876. 2012 "Softcover reprint of the hardcover 1st edition 1987"; *Undergraduate Texts in Mathematics*.

Paul R. Halmos. *Intoduction to Hilbert Space and the Theory of Spectral Multiplicity*. Chelsea Publishing Company, New York, 2 edition, 1998a. ISBN 0821813781. URL http://books.google.com/books?vid=ISBN0821813781.

Paul R. Halmos. *Introduction to Hilbert Space and the Theory of Spectral Multiplicity*. AMS Chelsea, Providence RI, 2 edition, 1998b. ISBN 0821813781. URL http://books.google.com/books?vid=ISBN0821813781.

Georg Hamel. Eine basis aller zahlen und die unstetigen lösungen der funktionalgleichung $f(x+y) = f(x) + f(y)$. *Mathematische Annalen*, 60(3):459–462, 1905. URL http://gdz.sub.uni-goettingen.de/dms/load/img/?PPN=GDZPPN002260395&IDDOC=28580.

Deguang Han, Keri Kornelson, David Larson, and Eric Weber. *Frames for Undergraduates*, volume 40 of *Student Mathematical Library*. American Mathematical Society, 2007. ISBN 0821842129. URL http://books.google.com/books?vid=ISBN0821842129. Deguang Han = ???

G.H. Hardy. Prolegomena to a chapter on inequalities. *Journal of the London Mathematical Society*, 1–4:61–78, November 8 1929. URL http://jlms.oxfordjournals.org/content/vols1-4/issue13/index.dtl#PRESIDENTIAL_ADDRESS. "Presidential Address" to the London Mathematical Society.

Godfrey H. Hardy. *A Mathematician's Apology*. Cambridge University Press, Cambridge, 1940. URL http://www.math.ualberta.ca/~mss/misc/A%20Mathematician's%20Apology.pdf.

Godfrey Harold Hardy, John Edensor Littlewood, and George Pólya. *Inequalities*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2 edition, 1952. URL http://books.google.com/books?vid=ISBN0521358809.

Felix Hausdorff. *Set Theory*. Chelsea Publishing Company, New York, 3 edition, 1937. ISBN 0828401195. URL http://books.google.com/books?vid=ISBN0828401195. 1957 translation of the 1937 German *Grundzüge der Mengenlehre*.

S. Haykin and S. Kesler. Prediction-error filtering and maximum-entropy spectral estimation. In S. Haykin, editor, *Nonlinear Methods of Spectral Analysis*, volume 34 of *Topics in Applied Physics*, pages 9–72. Springer-Verlag, 1 edition, 1979.

S. Haykin and S. Kesler. Prediction-error filtering and maximum-entropy spectral estimation. In S. Haykin, editor, *Nonlinear Methods of Spectral Analysis*, volume 34 of *Topics in Applied Physics*, pages 9–72. Springer-Verlag, "second corrected and updated edition" edition, 1983.

Simon Haykin. *Communication Systems*. John Wiley & Sons, 3 edition, 1994. ISBN 0471571768. URL http://www.amazon.com/dp/0471571768.

Simon Haykin. *Adaptive Filter Theory*. Prentice Hall, Upper Saddle River, 4 edition, September 24 2001. ISBN 978-0130901262. URL http://books.google.com/books?vid=isbn0130901261.

Simon Haykin. *Digital Communication Systems*. Wiley Global Education, 4 edition, 2014. ISBN 9781118476772. URL http://books.google.com/books?vid=isbn9781118476772.

Michiel Hazewinkel, editor. *Handbook of Algebras*, volume 2. North-Holland, Amsterdam, 1 edition, 2000. ISBN 044450396X. URL http://books.google.com/books?vid=ISBN044450396X.

Jean Van Heijenoort. *From Frege to Gödel : A Source Book in Mathematical Logic, 1879-1931*. Harvard University Press, Cambridge, Massachusetts, 1967. URL http://www.hup.harvard.edu/catalog/VANFGX.html.

Christopher Heil. *A Basis Theory Primer*. Applied and Numerical Harmonic Analysis. Birkhäuser, Boston, expanded edition edition, 2011. ISBN 9780817646868. URL http://books.google.com/books?vid=ISBN9780817646868.

J. H. Heinbockel. *Introduction to Finite and Infinite Series and Related Topics*. Trafford Publishing, 2010. ISBN 9781426949548. URL http://books.google.com/books?vid=ISBN9781426949545.

Carl W. Helstrom. *Probability and Stochastic Processes for Engineers*. Maxwell Macmillan international editions in engineering. Macmillan, 2 edition, 1991. ISBN 9780023535710. URL http://books.google.com/books?vid=ISBN0023535717.

Hugo Hernandez. Modelling the effect of fluctuation in nonlinear systems using variance algebra - application to light scattering of ideal gases. Technical Report FRR 2016-1, ForsChem Research, December 2016. URL https://www.researchgate.net/publication/311424052.

Omar Hijab. *Introduction to Calculus and Classical Analysis*. Undergraduate Texts in Mathematics. Springer, 3 edition, 2011. ISBN 9781441994875. URL http://books.google.com/books?vid=ISBN1441994874.

David Hilbert. Grundzüge einer allgemeinen theorie der linearen integralgleichungen (fundamentals of a general theory of the linear integral equations). *Mathematisch-Physikalische Klasse (Mathematical-Physical Class)*, pages 49–91, March 1904. URL http://dz-srv1.sub.uni-goettingen.de/sub/digbib/loader?did=D57552. Report 1 of 6.

David Hilbert. Grundzüge einer allgemeinen theorie der linearen integralgleichungen (fundamentals of a general theory of the linear integral equations). *Mathematisch-Physikalische Klasse (Mathematical-Physical Class)*, pages 157–228, March 1906. URL http://dz-srv1.sub.uni-goettingen.de/sub/digbib/loader?did=D58133. Report 4 of 6.

David Hilbert. *Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen (Fundamentals of a general theory of the linear integral equations)*, volume 26 of *Fortschritte der mathematischen Wissenschaften in Monographien (Progress of the mathematical sciences in Monographs)*. B.G. Teubner, Leipzig und Berlin, 1912. URL http://www.worldcat.org/oclc/13468199.

David Hilbert, Lothar Nordheim, and John von Neumann. über die grundlagen der quantenmechanik (on the bases of quantum mechanics). *Mathematische Annalen*, 98:1–30, 1927. ISSN 0025-5831 (print) 1432-1807 (online). URL http://dz-srv1.sub.uni-goettingen.de/cache/toc/D27776.html.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, March 1963. doi: 10.1080/01621459.1963.10500830. URL https://repository.lib.ncsu.edu/bitstream/handle/1840.4/2170/ISMS_1962_326.pdf.

Otto Hölder. üeber einen mittelwerthssatz. *Göttingen Nachrichten*, pages 38–47, 1889. URL http://www.digizeitschriften.de/dms/img/?PPN=GDZPPN00252421X.

Samuel S. Holland, Jr. The current interest in orthomodular lattices. In James C. Abbott, editor, *Trends in Lattice Theory*, pages 41–126. Van Nostrand-Reinhold, New York, 1970. URL http://books.google.com/books?id=ZfA-AAAAIAAJ. from Preface: "The present volume contains written versions of four talks on lattice theory delivered to a symposium on Trends in Lattice Theory held at the United States Naval Academy in May of 1966.".

Klaus Höllig. *Finite Element Methods With B-Splines*, volume 26 of *Frontiers in Applied Mathematics*. SIAM, 2003. ISBN 0-89871-533-4. URL http://books.google.com/books?vid=ISBN0898715334.

Lars Hörmander. *Notions of Convexity*. Birkhäuser, Boston, 1994. ISBN 0817637990. URL http://books.google.com/books?vid=ISBN0817637990.

Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, 1990. ISBN 0-521-30586-1. URL http://books.google.com/books?vid=isbn0521305861. Library: QA188H66 1985.

Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6):417–441, 1933. URL http://dx.doi.org/10.1037/h0071325.

Alfred Edward Housman. *More Poems*. Alfred A. Knopf, 1936. URL http://books.google.com/books?id=rTMiAAAAMAAJ.

K Husimi. Studies on the foundations of quantum mechanics i. *Proceedings of the Physico-Mathematical Society of Japan*, 19:766–789, 1937.

Emmanuel C. Ifeachor and Barrie W. Jervis. *Digital Signal Processing: A Practical Approach*. Electronic systems engineering series. Prentice Hall, 1993. ISBN 020154413X. URL http://www.amazon.com/dp/020154413X.

Emmanuel C. Ifeachor and Barrie W. Jervis. *Digital Signal Processing: A Practical Approach*. Electronic systems engineering series. Prentice Hall, 2002. ISBN 9780201596199. URL http://www.amazon.com/dp/0201596199.

Satoru Igari. *Real Analysis—With an Introduction to Wavelet Theory*, volume 177 of *Translations of mathematical monographs*. American Mathematical Society, 1996. ISBN 9780821821046. URL http://books.google.com/books?vid=ISBN0821821040.

Vasile I. Istrăţescu. *Inner Product Structures: Theory and Applications*. Mathematics and Its Applications. D. Reidel Publishing Company, 1987. ISBN 9789027721822. URL http://books.google.com/books?vid=ISBN9027721823.

Luisa Iturrioz. Ordered structures in the description of quantum systems: mathematical progress. In *Methods and applications of mathematical logic: proceedings of the VII Latin American Symposium on Mathematical Logic held July 29-August 2, 1985*, volume 69, pages 55–75, Providence Rhode Island, July 29–August 2 1985. Sociedade Brasileira de Lógica, Sociedade Brasileira de Matemática, and the Association for Symbolic Logic, AMS Bookstore (1988). ISBN 0821850768.

Robert C. James. Orthogonality in normed linear spaces. *Duke Mathematical Journal*, 12(2):291–302, 1945. ISSN 0012-7094. doi: 10.1215/S0012-7094-45-01223-3. URL http://projecteuclid.org/euclid.dmj/1077473105.

Robert C. James. Inner products in normed linear spaces. *Bulletin of the American Mathematical Society*, 53:559–566, 1947a. ISSN 0002-9904. doi: 10.1090/S0002-9904-1947-08833-9. URL http://www.ams.org/bull/1947-53-06/S0002-9904-1947-08831-5/.

Robert C. James. Orthogonality and linear functionals in normed linear spaces. *Transactions of the American Mathematical Society*, 61(2):265–292, March 1947b. ISSN 1088-6850. doi: 10.2307/1990220. URL http://www.jstor.org/stable/1990220.

Bjorn Jawerth and Wim Sweldens. An overview of wavelet based multiresolution analysis. *SIAM Review*, 36:377–412, September 1994. URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.45.5864&rep=rep1&type=pdf.

Andrew H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press. Dover Books on Electrical Engineering, 1970.

Andrew H. Jazwinski. *Stochastic Processes and Filtering Theory*. Dover Books on Electrical Engineering Series. Dover, "this dover edition, first published in 2007, is an unabridged republication of the work originally published by academic press, inc., new york, in 1970." edition, 2007. ISBN 9780486318196.

J. N. R. Jeffers. Two case studies in the application of principal component analysis. *Royal Statistical Society*, 16(3):225–236, 1967. URL https://doi.org/10.2307/2985919.

Alan Jeffrey and Hui Hui Dai. *Handbook of Mathematical Formulas and Integrals*. Handbook of Mathematical Formulas and Integrals Series. Academic Press, 4 edition, January 18 2008. ISBN 9780080556840. URL http://books.google.com/books?vid=ISBN0080556841.

Gwilym M. Jenkins and Donald G. Watts. *Spectral analysis and its applications*. Holden-Day series in time series analysis. Holden-Day, 1968. URL https://archive.org/details/spectralanalysis00jenk_0.

Finn Verner Jensen. *Bayesian Networks and Decision Graphs*. Information Science and Statistics. Springer Science & Business Media, 2013. ISBN 9781475735024. URL http://books.google.com/books?vid=ISBN9781475735024.

J. L. W. V. Jensen. Sur les fonctions convexes et les ine'galite's entre les valeurs moyennes (on the convex functions and the inequalities between the average values). *Acta Mathematica*, 30 (1):175–193, December 1906. ISSN 0001-5962. doi: 10.1007/BF02418571. URL http://www.springerlink.com/content/r55q1411g840j446/.

Iain M. Johnstone and Bernard W. Silverman. Wavelet threshold estimators for data with correlated noise. *Royal Statistical Society*, 59(2):319–351, 1997. doi: https://doi.org/10.1111/1467-9868.00071. URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.50.9891&rep=rep1&type=pdf.

Leonard Benjamin William Jolley. *Summation of Series*. Dover, New York, second revised edition, 1961. URL http://plouffe.fr/simon/math/SummationofSeries.pdf.

I. T. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics. Springer Science & Business Media, 2013. ISBN 9781475719048. URL http://books.google.com/books?vid=ISBN9781475719048.

Ian T. Jolliffe. A note on the use of principal components in regression. *Royal Statistical Society*, 31 (2):300–303, 1982. URL https://doi.org/10.2307/2348005.

P. Jordan and J. von Neumann. On inner products in linear, metric spaces. *Annals of Mathematics*, 36(3):719–723, July 1935. doi: 10.2307/1968653. URL http://www.jstor.org/stable/1968653.

Palle E. T. Jørgensen, Kathy Merrill, Judith Packer, and Lawrence W. Baggett. *Representations, Wavelets and Frames: A Celebration of the Mathematical Work of Lawrence Baggett*. Applied and Numerical Harmonic Analysis. Birkhäuser, 2008. ISBN 9780817646820. URL http://books.google.com/books?vid=ISBN0817646825.

K. D. Joshi. *Applied Discrete Structures*. New Age International, New Delhi, 1997. ISBN 8122408265. URL http://books.google.com/books?vid=ISBN8122408265.

J.S.Chitode. *Signals And Systems*. Technical Publications, 2009a. ISBN 9788184316780. URL http://books.google.com/books?vid=ISBN818431678X.

J.S.Chitode. *Digital Signal Processing*. Technical Publications, 2009b. ISBN 9788184316469. URL http://books.google.com/books?vid=ISBN9788184316469.

Shizuo Kakutani and George W. Mackey. Ring and lattice characterizations of complex hilbert space. *Bulletin of the American Mathematical Socieity*, 52:727–733, 1946. doi: 10.1090/S0002-9904-1946-08644-9. URL http://www.ams.org/bull/1946-52-08/S0002-9904-1946-08644-9/.

Gudrun Kalmbach. *Measures and Hilbert Lattices*. World Scientific, Singapore, 1986. ISBN 9789971500092. URL http://books.google.com/books?vid=ISBN9971500094.

Nidal S. Kamel and K. S. Sim. Image signal-to-noise ratio and noise variance estimation using autoregressive model. *The Journal of Scanning Microscopes*, 26:277–281, September 15 2004. URL https://onlinelibrary.wiley.com/doi/pdf/10.1002/sca.4950260605.

Edward Kasner and James Roy Newman. *Mathematics and the Imagination*. Simon and Schuster, 1940. ISBN 0486417034. URL http://books.google.com/books?vid=ISBN0486417034. “unabridged and unaltered republication” available from Dover.

Steven M. Kay. *Modern Spectral Estimation: Theory and Application*. Prentice-Hall signal processing series. Prentice Hall, 1988. ISBN 9788131733561. URL http://books.google.com/books?vid=ISBN8131733564.

James P. Keener. *Principles of Applied Mathematics; Transformation and Approximation*. Addison-Wesley Publishing Company, Reading, Massachusets, 1988. ISBN 0201156741. URL http://www.worldcat.org/isbn/0201156741.

M. G. Kendall. *A Course in Multivariate Analysis*, volume 2 of *Griffin's Statistical Monographs & Courses*. Hafner Publishing Company, New York, “fourth impression” edition, 1968. URL https://archive.org/details/courseinmultivar0000kend.

Yoshida Kenko. *The Tsuredzure Gusa of Yoshida No Kaneyoshi. Being the meditations of a recluse in the 14th Century (Essays in Idleness)*. circa 1330. URL http://www.humanistictexts.org/kenko.htm. 1911 translation of circa 1330 text.

John Francis Kenney. *Mathematics of Statistics, Volume 2*. D. Van Nostrand Company, New York, second edition, second printing edition, 1947.

Mohamed A. Khamsi and W.A. Kirk. *An Introduction to Metric Spaces and Fixed Point Theory*. John Wiley, New York, 2001. ISBN 978-0471418252. URL http://books.google.com/books?vid=isbn0471418250.

K.S. Kim and Nidal S. Kamel. Image signal-to-noise ratio and noise variance estimation using autoregressive model. *The Journal of Scanning Microscopes*, 26(6):135–139, September 24 2003. URL https://doi.org/10.1002/sca.4950260605.

Anthony W Knapp. *Advanced Real Analysis*. Cornerstones. Birkhäuser, Boston, Massachusetts, USA, 1 edition, July 29 2005a. ISBN 0817643826. URL http://books.google.com/books?vid=ISBN0817643826.

Anthony W Knapp. *Basic Real Analysis*. Cornerstones. Birkhäuser, Boston, Massachusetts, USA, 1 edition, July 29 2005b. ISBN 0817632506. URL http://books.google.com/books?vid=ISBN0817632506.

Andrei N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, Berlin, 1933a.

Andrei Nikolaevich Kolmogorov. *Foundations of the theory of probability*. Chelsea Publishing Company, New Yourk, 2 edition, 1933b. ISBN B0006AUOGO. URL http://statweb.stanford.edu/~cgates/PERSI/Courses/Phil166-266/Kolmogorov-Foundations.pdf. 1956 2nd edition English translation of A. N. Kolmogorov's 1933 "Grundbegriffe der Wahrscheinlichkeitsrechnung".

Lambert Herman Koopmans. volume 22 of *Probability and mathematical statistics*. Academic Press, 2 edition, 1995. ISBN 9780124192515.

M. G. Krein and M. A. Krasnoselski. Fundamental theorems concerning the extension of hermitian operators and some of their applications to the theory of orthogonal polynomials and the moment problem. *Uspekhi Matematicheskikh Nauk*, 2(3), 1947. URL http://www.turpion.org/php/homes/pa.phtml?jrnid=rm. (Russian Mathematical Surveys).

M. G. Krein, M. A. Krasnoselski, and D. P. Milman. Concerning the deficiency numbers of linear operators in banach space and some geometric questions. *Sbornik Trudov Inst. Matem. AN Ukrainian SSR*, 11:97–112, 1948.

Carlos S. Kubrusly. *The Elements of Operator Theory*. Springer, 1 edition, 2001. ISBN 9780817641740. URL http://books.google.com/books?vid=ISBN0817641742.

Carlos S. Kubrusly. *The Elements of Operator Theory*. Springer, 2 edition, 2011. ISBN 9780817649975. URL http://books.google.com/books?vid=ISBN0817649972.

Andrew Kurdila and Michael Zabarankin. *Convex Functional Analysis*. Systems & Control: Foundations & Applications. Birkhäuser, Boston, 2005. ISBN 9783764321987. URL http://books.google.com/books?vid=ISBN3764321989.

Traian Lalescu. *Sur les équations de Volterra*. PhD thesis, University of Paris, 1908. advisor was Émile Picard.

Traian Lalescu. *Introduction à la théorie des équations intégrales (Introduction to the Theory of Integral Equations)*. Librairie Scientifique A. Hermann, Paris, 1911. URL http://www.worldcat.org/oclc/1278521. first book about integral equations ever published.

Pierre-Simon Laplace. *Theorie Analytique des Probabilités*. 1812.

Pierre-Simon Laplace. *Theorie Analytique des Probabilités*. Paris, 2 edition, 1814. URL https://archive.org/details/bub_gb_6MRLAAAAMAAJ/.

Pierre-Simon Laplace. *A Philosophical Essay on Probabilities*. Wiley, 1902. URL https://books.google.com/books?id=DfI7AQAAMAAJ.

Peter D. Lax. *Functional Analysis*. John Wiley & Sons Inc., USA, 2002. ISBN 0-471-55604-1. URL http://www.worldcat.org/isbn/0471556041. QA320.L345 2002.

Steven R. Lay. *Convex Sets and their Applications*. John Wiley & Sons, New York, 1982. ISBN 0-471-09584-2. URL http://books.google.com/books?vid=isbn0471095842.

Q. Leclere, Bert Roozen, and Celine Sandier. On the use of the hs estimator for the experimental assessment of transmissibility matrices. *Mechanical Systems and Signal Processing*, 43(1-2):237–245, 2014. URL https://hal.archives-ouvertes.fr/hal-00911916/document. <hal-00911916>.

John A. Lee and Michel Verleysen. *Nonlinear Dimensionality Reduction*. Information Science and Statistics. 2007. ISBN 9780387393513. URL http://books.google.com/books?vid=ISBN9780387393513.

Gottfried W. Leibniz. Symbolismus memorabilis calculi algebraici et infinitesimalis, in comparatione potentiarum et differentiarum; et de lege homogeneorum transcendentali. *Miscellanea Berolinensia ad incrementum scientiarum, ex scriptis Societati Regiae scientarum*, pages 160–165, 1710. URL http://bibliothek.bbaw.de/bibliothek-digital/digitalequellen/schriften/anzeige/index_html?band=01-misc/1&seite:int=184.

Gottfried Wilhelm Leibniz. Letter to christian huygens, 1679. In Leroy E. Loemker, editor, *Philosophical Papers and Letters*, volume 2 of *The New Synthese Historical Library*, chapter 27, pages 248–249. Kluwer Academic Press, Dordrecht, 2 edition, September 8 1679. ISBN 902770693X. URL http://books.google.com/books?vid=ISBN902770693X.

J. Leuridan, J. De Vis, and Van der Auweraer H. F. Lembregts D. ”a comparison of some frequency response function measurement techniques. In *Proceedings of the 4th International Modal Analysis Conference IMAC*, pages 908–918, Los Angeles, CA, 1986. URL https://www.researchgate.net/publication/283994172.

Ezra C. Levy. *A New Table of Laplace Transformation Pairs*. Space Technology Laboratories, Inc., P.O. Box 95001 Los Angeles 45, California, November 1958. “This Laplace transform table contains 1024 possible transformation pairs in nine sections.…In this table we shall deal with the one-sided Laplace transformation, as contrasted to the two-sided transformation…The reason for this choice is that for the systems that are most commonly considered the solution of interest is that corresponding to real values of time, and which assumes knowledge of the initial conditions, i. e. , the dynamic characteristics of the system at $t = 0$.”.

Zach Liang and George C. Lee. *Random Vibration: Mechanical, Structural, and Earthquake Engineering Applications*. Advances in Earthquake Engineering. Taylor & Francis, illustrated edition, 2015. ISBN 9781498702348. URL http://www.amazon.com/dp/1498702341.

Elliott H. Lieb and Michael Loss. *Analysis*, volume 14 of *Graduate studies in mathematics*. American Mathematical Soceity, Providence, Rhode Island, USA, 2 edition, 2001. ISBN 0821827839. URL http://books.google.com/books?vid=ISBN0821827839.

Joram Lindenstrauss and Lior Tzafriri. *Classical Banach Spaces I: Sequence Spaces*, volume 92 of *Ergebnisse der Mathematik und ihrer Grenzgebiete*. Springer-Verlag, 1977. ISBN 978-3642665592. URL http://www.amazon.com/dp/3642665594.

Anders Lindquist and Giorgio Picci. *Linear Stochastic Systems: A Geometric Approach to Modeling, Estimation and Identification*, volume 1 of *Series in Contemporary Mathematics*. Springer, 2015. ISBN 9783662457504. URL http://books.google.com/books?vid=ISBN3662457504.

J. Liouville. Sur l'integration d'une classe d'équations différentielles du second ordre en quantités finies explicites. *Journal De Mathematiques Pures Et Appliquees*, 4:423–456, 1839. URL http://gallica.bnf.fr/ark:/12148/bpt6k16383z.

Jun S. Liu. *Monte Carlo Strategies in Scientific Computing*. Springer Series in Statistics. Springer Science & Business Media, 2013. ISBN 9780387763712. URL http://books.google.com/books?vid=ISBN9780387763712.

Lynn H. Loomis. *An Introduction to Abstract Harmonic Analysis*. The University Series in Higher Mathematics. D. Van Nostrand Company, Toronto, 1953. URL http://books.google.com/books?id=aNg-AAAAIAAJ.

Lynn H. Loomis. *The Lattice Theoretic Background of the Dimension Theory of Operator Algebras*, volume 18 of *Memoirs of the American Mathematical Society*. American Mathematical Society, Providence RI, 1955. ISBN 0821812181. URL http://books.google.com/books?id=P3Vl_lXCFRkC.

Lynn H. Loomis and Ethan D. Bolker. *Harmonic analysis*. Mathematical Association of America, 1965. URL http://books.google.com/books?id=MEfvAAAAMAAJ.

Niccolò Machiavelli. *The Literary Works of Machiavelli: Mandragola, Clizia, A Dialogue on Language, and Belfagor, with Selections from the Private Correspondence*. Oxford University Press, 1961. ISBN 0313212481. URL http://www.worldcat.org/isbn/0313212481.

Fumitomo Maeda and Shûichirô Maeda. *Theory of Symmetric lattices*, volume 173 of *Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen*. Springer-Verlag, Berlin/New York, 1970. URL http://books.google.com/books?id=4oeBAAAAIAAJ.

Khosrow Maleknejad, Monireh Nosrati Sahlan, and Azadeh Ostad. Numerical solution of fractional integro-differential equation by using cubic b-spline wavelets. In *Proceedings of the World Congress on Engineering*, volume I, London, July 3–5 2013. URL https://pdfs.semanticscholar.org/bbe0/55f40768bbefe9e6476bdce701effb9243f7.pdf.

Lech Maligranda. A simple proof of the hölder and the minkowski inequality. *The American Mathematical Monthly*, 102(3):256–259, March 1995. URL http://www.jstor.org/stable/2975013.

Stéphane G. Mallat. Multiresolution approximations and wavelet orthonormal bases of $l^2(\mathbb{R})$. *Transactions of the American Mathematical Society*, 315(1):69–87, September 1989. URL http://blanche.polytechnique.fr/~mallat/papiers/math_multiresolution.pdf.

Stéphane G. Mallat. *A Wavelet Tour of Signal Processing*. Elsevier, 2 edition, September 15 1999. ISBN 9780124666061. URL http://books.google.com/books?vid=ISBN012466606X.

Dimitris G. Manolakis, Vinay K. Ingle, and Stephen M. Kogon. *Statistical and Adaptive Signal Processing*. Mc-Graw Hill, Boston, 2000. ISBN 0070400512. URL http://www.amazon.com/dp/0070400512.

Dimitris G. Manolakis, Vinay K. Ingle, and Stephen M. Kogon. *Statistical and Adaptive Signal Processing: Spectral Estimation, Signal Modeling, Adaptive Filtering, and Array Processing*. Artech House signal processing library. Artech House, illustrated, reprint edition, 2005. ISBN 9781580536103. URL http://www.amazon.com/dp/1580536107.

J. L. Massera and J. J. Schäffer. Linear differential equations and functional analysis, i. *The Annals of Mathematics, 2nd Series*, 67(3):517–573, May 1958. URL http://www.jstor.org/stable/1969871.

MatLab. cpsd: Cross power spectral density. *MatLab*, accessed 2018 November 26 2018a. URL https://www.mathworks.com/help/signal/ref/cpsd.html.

MatLab. xcorr: Cross-correlation. *MatLab*, accessed 2018 November 26 2018b. URL https://www.mathworks.com/help/signal/ref/xcorr.html.

Stefan Mazur. Sur les anneaux linéaires. *Comptes rendus de l'Académie des sciences*, 207:1025–1027, 1938.

Stefan Mazur and Stanislaus M. Ulam. Sur les transformations isométriques d'espaces vectoriels normées. *Comptes rendus de l'Académie des sciences*, 194:946–948, 1932.

Samara Mendez. Us egg production data set. Technical Report E008R01, The Human League Labs, 2019. URL https://osf.io/z2gxn/.

José Mendoza and Tijani Pakhrou. On some characterizations of inner product spaces. *Journal of Mathematical Analysis and Applications*, 282(1):369–382, June 1 2003. URL http://linkinghub.elsevier.com/retrieve/pii/S0022247X03001690.

Karl Menger. Untersuchungen über allgemeine metrik. *Mathematische Annalen*, 100:75–163, 1928. ISSN 0025-5831. URL http://link.springer.com/article/10.1007/BF01455705. (Investigations on general metric).

Claudia Menini and Freddy Van Oystaeyen. *Abstract Algebra; A Comprehensive Treatment*. Marcel Dekker Inc, New York, April 2004. ISBN 0-8247-0985-3. URL http://books.google.com/books?vid=isbn0824709853.

James Mercer. Functions of positive and negative type and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society of London*, 209:415–446, 1909. ISSN 02643952. URL http://www.jstor.org/stable/91043.

Anthony N. Michel and Charles J. Herget. *Applied Algebra and Functional Analysis*. Dover Publications, Inc., 1993. ISBN 048667598X. URL http://books.google.com/books?vid=ISBN048667598X. original version published by Prentice-Hall in 1981.

Gregory K. Miller. *Probability: Modeling and Applications to Random Processes*. Wiley, August 25 2006. URL http://www.amazon.com/dp/0471458929.

Gradimir V. Milovanović and Igorž. Milovanović. On a generalization of certain results of a. ostrowski and a. lupaş. *Publikacije Elektrotehničkog Fakulteta (Publications Electrical Engineering)*, 634(677):62–69, 1979. URL http://www.mi.sanu.ac.rs/~gvm/radovi/643.pdf.

Hermann Minkowski. *Geometrie der Zahlen*. Druck und Verlag von B.G. Teubner, Leipzig, 1910. URL http://www.archive.org/details/geometriederzahl00minkrich. Geometry of Numbers.

Fred Mintzer. Filters for distortion-free two-band multi-rate filter banks. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 32, 1985.

Larry D. Mitchell. An improved method for the fast fourier transform (fft) calculation of the frequency response function. In M. Kunt and F. De Coulon, editors, *Proceedings of EUSIPCO-80, First European Signal Processing Conference*, Signal processing, theories and applications, pages 159–160, Lausanne, Switzerland, September 16–19 1980. Elsevier. URL http://www.worldcat.org/oclc/63416846. citation from Cobb(1988).

Larry D. Mitchell. Improved methods for the fast fourier transform (fft) calculation of the frequency response function. *Journal of Mechanical Design*, 104(2):277–279, April 01 1982. doi: 10.1115/1.3256336. URL http://mechanicaldesign.asmedigitalcollection.asme.org/article.aspx?articleid=1451670.

Dragoslav S. Mitrinović, J. E. Pečarić, and Arlington M. Fink. *Classical and New Inequalities in Analysis*, volume 61 of *Mathematics and its Applications (East European Series)*. Kluwer Academic Publishers, Dordrecht, Boston, London, 2010. ISBN 978-90-481-4225-5. URL http://www.amazon.com/dp/0792320646.

Todd K. Moon and Wynn C. Stirling. *Mathematical Methods and Algorithms for Signal Processing*. Prentice Hall, Upper Saddle River, 2000. ISBN 0-201-36186-8. URL http://books.google.com/books?vid=isbn0201361868.

Robert Édouard Moritz. *Memorabilia Mathematica; Or, The Philomath's Quotation-book*. Macmillan, 1914. URL https://archive.org/details/memorabiliamath02morigoog.

H. P. Mulholland. On generalizations of minkowski's inequality in the form of a triangle inequality. *Proceedings of the London Mathematical Society*, s2-51:294–307, 1950. URL http://plms.oxfordjournals.org/content/s2-51/1/294.extract. received 1946 October 10, read 1947 June 19.

Paul J. Nahin. *Dr. Euler's Fabulous Formula: Cures Many Mathematical Ills*. Princeton Science Library. Princeton University Press, reprint, revised edition, April 25 2011. ISBN 9781400838479. URL http://books.google.com/books?vid=ISBN1400838479.

Louis Narens. *Theories of Probability: An examination of logical and qualitative foundations*, volume 2 of *Advanced series on mathematical psychology*. World Scientific, January 2007. ISBN 9789812708014. URL http://www.amazon.com/dp/9812708014.

Louis Narens. Alternative probability theories for cognitive psychology. *Topics in Cognitive Science*, 6(1):114–120, January 2014. ISSN 1756-8757. doi: 10.1111/tops.12071. URL http://onlinelibrary.wiley.com/doi/10.1111/tops.12071/full.

NASA. Global temperature. *NASA Global Climate Change: Vital Signs of the Planet*, 2019. URL https://climate.nasa.gov/vital-signs/global-temperature/.

Oliver Nelles. *Nonlinear System Identification*. Springer, New York, 2001. ISBN 9783540673699.

Ben Noble and James W. Daniel. *Applied Linear Algebra*. Prentice-Hall, Englewood Cliffs, NJ, USA, 3 edition, 1988. ISBN 0130412600. URL http://www.worldcat.org/isbn/0130412600.

Timur Oikhberg and Haskell Rosenthal. A metric characterization of normed linear spaces. *Rocky Mountain Journal Of Mathematics*, 37(2):597–608, 2007. URL http://www.ma.utexas.edu/users/rosenthl/pdf-papers/95-oikh.pdf.

Oystein Ore. On the foundation of abstract algebra. i. *The Annals of Mathematics*, 36(2):406–437, April 1935. URL http://www.jstor.org/stable/1968580.

Brad Osgood. The fourier transform and its applications; ee261: Standford university course. university course notes, 2002. URL http://www.stanford.edu/class/ee261/.

Lincoln P. Paine. *Warships of the World to 1900*. Ships of the World Series. Houghton Mifflin Harcourt, 2000. ISBN 9780395984147. URL http://books.google.com/books?vid=ISBN9780395984149.

Endre Pap. *Null-Additive Set Functions*, volume 337 of *Mathematics and Its Applications*. Kluwer Academic Publishers, 1995. ISBN 0792336585. URL http://www.amazon.com/dp/0792336585.

Anthanasios Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill Series in Electrical Engineering. McGraw-Hill Book Company, New York, 2 edition, 1984. ISBN 9780070484689. URL http://books.google.com/books?vid=ISBN0070484686.

Anthanasios Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, New York, 3 edition, 1991. ISBN 0070484775. URL http://books.google.com/books?vid=ISBN0070484775.

Anthanasios Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill series in electrical and computer engineering. Tata McGraw-Hill, New York, 4 edition, 2002. ISBN 9780071226615. URL http://books.google.com/books?vid=ISBN0071226613.

Athanasios Papoulis. *Probability & Statistics*. Prentice-Hall international editions. Prentice Hall, illustrated edition, 1990. ISBN 9780137116980. URL http://books.google.com/books?vid=ISBN0137116985.

Giuseppe Peano. *Calcolo geometrico secondo l'Ausdehnungslehre di H. Grassmann preceduto dalle operazioni della logica deduttiva*. Fratelli Bocca Editori, Torino, 1888a. Geometric Calculus: According to the *Ausdehnungslehre* of H. Grassmann.

Giuseppe Peano. *Geometric Calculus: According to the Ausdehnungslehre of H. Grassmann*. Springer (2000), 1888b. ISBN 0817641262. URL http://books.google.com/books?vid=isbn0817641262. originally published in 1888 in Italian.

Karl Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2:559–572, 1901. URL https://doi.org/10.1080/14786440109462720.

Michael Pedersen. *Functional Analysis in Applied Mathematics and Engineering*. Chapman & Hall/CRC, New York, 2000. ISBN 9780849371691. URL http://books.google.com/books?vid=ISBN0849371694. Library QA320.P394 1999.

J. E. Pečarić, Frank Proschan, and Yung Liang Tong. *Convex Functions, Partial Orderings, and Statistical Applications*, volume 187 of *Mathematics in Science and Engineering*. Academic Press, San Diego, California, 1992. ISBN 978-0125492508. URL http://books.google.com/books?vid=ISBN0125492502.

Rik Pintelon and Johan Schoukens. *System Identification: A Frequency Domain Approach*. John Wiley & Sons, 2 edition, 2012. ISBN 9781118287392. URL http://books.google.com/books?vid=ISBN1118287398.

C Piron. Axiomatique quantique. *Helvetica Physica Acta*, 37:439–468, 1964a. ISSN 0018-0238. English translation completed by M. Cole.

C Piron. Qunatum axiomatics. *Helvetica Physica Acta*, 1964b. English translation of *Axiomatique quantique*, RB4 Technical memo 107/106/104, GPO Engineering Department (London).

R. L. Plackett. Studies in the history of probability and statistics. xxix: The discovery of the method of least squares. *Biometrika*, 59(2):239–251, August 1972. doi: 10.2307/2334569. URL https://www.jstor.org/stable/2334569. https://doi.org/10.1093/biomet/59.2.239.

Henri Poincaré. *Calcul Des Probabilités*. De L'ecole Polytechnique, Paris, 1912. URL http://www.archive.org/details/calculdeprobabil00poinrich.

Karl Raimund Popper. *Conjectures and refutations: the growth of scientific knowledge*. Basic Books, 1962. URL https://books.google.com/books?id=OBgNAQAAIAAJ.

Karl Raimund Popper. *Conjectures and refutations: the growth of scientific knowledge*. Routledge, 2 edition, 1963. ISBN 9781135971304.

Olivier Pourret, Patrick Naïm, and Bruce Marcot, editors. *Bayesian Networks: A Practical Guide to Applications*, volume 73 of *Statistics in Practice*. John Wiley & Sons, 2008. ISBN 9780470994542. URL http://books.google.com/books?vid=ISBN9780470994542.

Lakshman Prasad and Sundararaja S. Iyengar. *Wavelet Analysis with Applications to Image Processing*. CRC Press LLC, Boca Raton, 1997. ISBN 978-0849331695. URL http://books.google.com/books?vid=ISBN0849331692. Library TA1637.P7 1997.

John G. Proakis. *Digital Communications*. McGraw Hill, 4 edition, 2001. ISBN 0-07-232111-3. URL http://www.mhhe.com/.

John G. Proakis, editor. *Algorithms for Statistical Signal Processing*. Prentice Hall, 2002. ISBN 9780130622198. URL http://books.google.com/books?vid=ISBN9780130622198.

John G. Proakis and Dimitris G. Manolakis. *Digital Signal Processing: Principles, Algorithms, and Applications*. Prentice Hall international editions. Prentice Hall, 3 edition, 1996. ISBN 9780133737622. URL http://books.google.com/books?vid=ISBN0133737624.

Murray H. Protter and Charles B. Jr. Morrey. *Intermediate Calculus*. Undergraduate Texts in Mathematics. Springer Science & Business Media, 2, illustrated edition, 2012. ISBN 9781461210863. URL http://books.google.com/books?vid=ISBN1461210860. “Softcover reprint of the hardcover 2nd edition 1985” (page iv).

Ptolemy. *Ptolemy's Almagest*. Springer-Verlag (1984), New York, circa 100AD. ISBN 0387912207. URL http://gallica.bnf.fr/ark:/12148/bpt6k3974x.

Inder K. Rana. *An Introduction to Measure and Integration*, volume 45 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, R.I., 2 edition, 2002. ISBN 978-0821829745. URL http://books.google.com/books?vid=ISBN0821829742.

K. Deergha Rao and M. N. S. Swamy. *Digital Signal Processing: Theory and Practice*. Springer, 2018. ISBN 9789811080814. URL http://books.google.com/books?vid=ISBN9789811080814.

rdocumentation.org. Monthly sunspot data, from 1749 to p̈resenẗ. *RDocumentation*, 2013. URL https://www.rdocumentation.org/packages/datasets/versions/3.6.1/topics/sunspot.month.

Helen Steiner Rice. *Helen Steiner Rice: A Collection of Faith, Hope, and Love*. Inspirational Book Bargains. Barbour Publishing, 2012. ISBN 9781620294345. URL http://books.google.com/books?vid=ISBN9781620294345.

Charles Earl Rickart. *General Theory of Banach Algebras*. University series in higher mathematics. D. Van Nostrand Company, Yale University, 1960. URL http://books.google.com/books?id=PVrvAAAAMAAJ.

Frigyes Riesz. Zur theorie des hilbertschen raumes. *Acta Scientiarum Mathematicarum*, 7:34–38, 1934. URL http://www.math.u-szeged.hu/acta/. The Theory of Hilbert Space.

Enders A. Robinson. *Random Wavelets and Cybernetic Systems*, volume 9 of *Griffins Statistical Monographs & Courses*. Lubrecht & Cramer Limited, London, June 1962. ISBN 0852640757. URL http://books.google.com/books?vid=ISBN0852640757.

Enders A. Robinson. Multichannel z-transforms and minimum delay. *Geophyics*, 31(3):482–500, June 1966. doi: 10.1190/1.1439788. URL http://dx.doi.org/10.1190/1.1439788.

Enders A. Robinson. A historical perspective of spectrum estimation. *Proceedings of the IEEE*, 70 (9):885–907, September 1982. doi: 10.1109/PROC.1982.12423. URL http://www.archive.ece.cmu.edu/~ece792/handouts/Robinson82.pdf.

R. Tyrrell Rockafellar. *Convex Analysis*. Princeton landmarks in mathematics. Princeton University Press, 1970. ISBN 9780691015866. URL http://books.google.com/books?vid=ISBN0691015864.

Murray Rosenblatt. Remarks on some non-parametric estimates of a density function. *Annals of Mathematical Statistics*, 27(3):832–837, September 1956. URL https://link.springer.com/content/pdf/10.1007/978-1-4419-8339-8_13.pdf. https://projecteuclid.org/download/pdf_1/euclid.aoms/1177728190.

Maxwell Rosenlicht. *Introduction to Analysis*. Dover Publications, New York, 1968. ISBN 0-486-65038-3. URL http://books.google.com/books?vid=ISBN0486650383.

Sheldon Ross. *A First Course in Probability*. Prentice Hall, Upper Saddle River, New Jersey, USA 07458, 5 edition, 1998. ISBN 0-13-746314-6. URL http://www.amazon.com/dp/0137463146.

Thomas Rossing, editor. *Springer Handbook of Acoustics*. Springer Handbooks. Springer, 2 edition, Springer. ISBN 9781493907557. URL http://books.google.com/books?vid=ISBN9781493907557.

George G. Roussas. *An Introduction to Probability and Statistical Inference*. Academic Press, 2 edition, October 21 2014. ISBN 9780128004371. URL http://books.google.com/books?vid=ISBN0128004371.

Walter Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, New York, 3 edition, 1976. ISBN 007054235X. URL http://books.google.com/books?vid=ISBN007054235X. Library QA300.R8 1976.

Walter Rudin. *Real and Complex Analysis*. McGraw-Hill Book Company, New York, New York, USA, 3 edition, 1987. ISBN 9780070542341. URL http://www.amazon.com/dp/0070542341. Library QA300.R8 1976.

Walter Rudin. *Functional Analysis*. McGraw-Hill, New York, 2 edition, 1991. ISBN 0-07-118845-2. URL http://www.worldcat.org/isbn/0070542252. Library QA320.R83 1991.

Bryan P. Rynne and Martin A. Youngson. *Linear Functional Analysis*. Springer undergraduate mathematics series. Springer, 2 edition, January 1 2008. ISBN 9781848000056. URL http://books.google.com/books?vid=ISBN1848000057.

Jr. S. Lawrence Marple. *Digital Spectral Analysis with Applications*. Prentice Hall, 1 edition, 1987. ISBN 9780132141499. URL http://books.google.com/books?vid=ISBN9780132141499.

Jr. S. Lawrence Marple. *Digital Spectral Analysis*. Dover Books on Electrical Engineering. Courier Dover Publications, 2019. ISBN 9780486780528. URL http://books.google.com/books?vid=ISBN9780486780528.

Shôichirô Sakai. *C*-Algebras and W*-Algebras*. Springer-Verlag, Berlin, 1 edition, 1998. ISBN 9783540636335. URL http://books.google.com/books?vid=ISBN3540636331. reprint of 1971 edition.

S. Salivahanan and A. Vallavaraj. *Digital Signal Processing*. Tata McGraw-Hill Education, 2001. ISBN 9780074639962. URL http://books.google.com/books?vid=ISBN9780074639962.

Usa Sasaki. Orthocomplemented lattices satisfying the exchange axiom. *Journal of Science of the Hiroshima University*, 17:293–302, 1954. ISSN 0386-3034. URL http://journalseek.net/cgi-bin/journalseek/journalsearch.cgi?field=issn&query=0386-3034.

Frank Moulton Saxelby. *A course in practical mathematics*. Longmans, Green, and Co., 39 Paternoster Row, London, 6 edition, 1920. URL https://archive.org/details/courseinpractica00saxerich/.

Jeffrey D. Scargle. Studies in astronomical time series analysis: Modeling random processes in the time domain. Technical Report NASA-TM-81148, A-7959, NASA Ames Research Center, Moffett Field, CA, United States, December 01 1979. URL https://ntrs.nasa.gov/search.jsp?R=19800007594. Document ID 19800007594, Accession Number: 80N15854.

Louis L. Scharf. *Statistical Signal Processing*. Addison-Wesley Publishing Company, Reading, MA, 1991. ISBN 0-201-19038-9.

H. A. Scharz. über ein die flächen kleinsten flächeninhalts betreffendes problem der variationsrechnung. *Acta Soc. Scient. Fen.*, 15:315–362, 1885. URL http://www-stat.wharton.upenn.edu/~steele/Publications/Books/CSMC/CSMC_index.html.

Michelle Schatzman. *Numerical Analysis: A Mathematical Introduction*. Clarendon Press, 2002. ISBN 9780198508526. URL https://books.google.com/books?id=3SuNiR1hzxUC.

Juljusz Schauder. Zur theorie stetiger abbildungen in funktionalräumen. *Mathematische Zeitschrift*, 26:47–65, 1927. URL http://eudml.org/doc/167911;jsessionid=156A34EBAB6C0E2DDAAC8C1232D23E8F.

Juljusz Schauder. Eine eigenschaft des haarschen orthogonalsystems. *Mathematische Zeitschrift*, 28:317–320, 1928.

Larry L. Schumaker. *Spline Functions: Basic Theory*. Cambridge Mathematical Library. Cambridge University Press, 3 edition, August 2007. ISBN 978-0-521-70512-7. URL http://books.google.com/books?vid=ISBN0521705126.

Isaac Schur. Uber die charakterischen wurzeln einer linearen substitution mit enier anwendung auf die theorie der integralgleichungen (over the characteristic roots of one linear substitution with an application to the theory of the integral). *Mathematische Annalen*, 66:488–510, 1909. URL http://dz-srv1.sub.uni-goettingen.de/cache/toc/D38231.html.

Karl Hermann Amandus Schwarz. *Ueber Ein Die Flachen Kleinsten Flächeninhalts; Betreffendes Problem Der Variationsrechnung (Over flattening smallest flat contents;*. October 31 1885. URL http://www-stat.wharton.upenn.edu/~steele/Publications/Books/CSMC/Schwarz.pdf. Festschrift Zum Jubelgeburtstage des Herrn Karl Weierstrass (anniversary publication to comemerate the birthdays of Mr. Karl Weierstrass).

Mícheál Ó Searcóid. *Elements of Abstract Analysis*. Springer Undergraduate Mathematics Series. Springer, 2002. ISBN 9781852334246. URL http://books.google.com/books?vid=ISBN185233424X.

Morteza Shafii-Mousavi. *Laplace Transforms Essentials*. Essentials. Research & Education Association, Piscataway, New Jersey, USA, 2015. ISBN 9780738672458. URL http://books.google.com/books?vid=ISBN9780738672458.

Jun Shao. *Mathematical Statistics*. Springer Texts in Statistics. Springer Science & Business Media, 2003. ISBN 9780387953823. URL http://books.google.com/books?vid=ISBN0387953825.

Kihong Shin and Joseph Hammond. *Fundamentals of Signal Processing for Sound and Vibration Engineers*. John Wiley & Sons, 2008. ISBN 9780470725641. URL https://archive.org/details/FubdamentalsOfSignalShinHammond/.

Bernard Shizgal. *Spectral Methods in Chemistry and Physics: Applications to Kinetic Theory and Quantum Mechanics*. Scientific Computation. Springer, January 7 2015. ISBN 9789401794541. URL http://books.google.com/books?vid=ISBN9401794545.

SILSO World Data Center. The international sunspot number. *International Sunspot Number Monthly Bulletin and online catalogue*, 2019. URL http://www.sidc.be/silso/datafiles.

SILSO World Data Center. The international sunspot number. *International Sunspot Number Monthly Bulletin and online catalogue*, May 24 2021. URL http://www.sidc.be/silso/datafiles. http://www.sidc.be/silso/INFO/snmtotcsv.php.

Bernard. W. Silverman. *Density Estimation for Statistics and Data Analysis*, volume 26 of *Monographs on Statistics & Applied Probability*. Chapman & Hall/CRC, illustrated, reprint edition, 1986. ISBN 9780412246203. URL http://books.google.com/books?vid=ISBN9780412246203.

George Finlay Simmons. *Calculus Gems: Brief Lives and Memorable Mathematicians*. Mathematical Association of America, Washington DC, 2007. ISBN 0883855615. URL http://books.google.com/books?vid=ISBN0883855615.

Barry Simon. *Convexity: An Analytic Viewpoint*, volume 187 of *Cambridge Tracts in Mathematics*. Cambridge University Press, May 19 2011. ISBN 9781107007314. URL http://books.google.com/books?vid=ISBN1107007313.

Iván Singer. *Bases in Banach Spaces I*, volume 154 of *Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen mit besonderer Berücksichtigung der Anwendungsgebiete*. Springer-Verlag New York, New York, 1970. ISBN 9780387048338.

Herbert Ellsworth Slaught and Nels Johann Lennes. *Elementary Algebra*. Allyn and Bacon, 1915. URL https://archive.org/details/elementaryalgebr00slaurich/.

M.J.T. Smith and T.P. Barnwell. A procedure for designing exact reconstruction filter banks for tree-structured subband coders. *IEEE International Conference on Acoustics, Speech and Signal Processing*, 9:421–424, 1984a. T.P. Barnwell is T.P. Barnwell III.

M.J.T. Smith and T.P. Barnwell. The design of digital filters for exact reconstruction in subband coding. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 34(3):434–441, June 1984b. ISSN 0096-3518. doi: 10.1109/TASSP.1986.1164832. T.P. Barnwell is T.P. Barnwell III.

D. E. Smylie, G. K. C. Clarke, and T. J. Ulrych. Analysis of irregularities in the earth's rotation. In Bruce Bolt, editor, *Methods in Computational Physics: Advances in Research and Applications*, volume 13 of *Geophyics*, pages 391–340. Academic Press, 1973. ISBN 9780323148368. URL http://books.google.com/books?vid=ISBN9780323148368.

J. Laurie Snell. A conversation with joe doob. *Statistical Science*, 12(4):301–311, 1997. URL http://www.jstor.org/stable/2246220. similar article on web at http://www.dartmouth.edu/~chance/Doob/conversation.html.

J. Laurie Snell. Obituary: Joseph leonard doob. *Journal of Applied Probability*, 42(1):247–256, 2005. URL http://projecteuclid.org/euclid.jap/1110381384.

Houshang H. Sohrab. *Basic Real Analysis*. Birkhäuser, Boston, 1 edition, 2003. ISBN 978-0817642112. URL http://books.google.com/books?vid=ISBN0817642110.

H. W. Sorenson. Least-squares estimation: from gauss to kalman. *IEEE Spectrum*, pages 63–68, July 1970. URL https://www2.pv.infn.it/~fontana/download/lect/Sorenson.pdf.

C.O.S. Sorzano, J. Vargas, and A. Pascual Montano. A survey of dimensionality reduction techniques. *arXiv.org*, (arXiv:1403.2877 [stat.ML]), March 12 2014. URL https://arxiv.org/abs/1403.2877.

Mandyam D. Srinath, P.K. Rajasekaran, and R. Viswanathan. *Introduction to Statistical Signal Processing with Applications*. Prentice Hall Inc, Upper Saddle River, 1996. ISBN 013125295X. URL http://engr.smu.edu/ee/mds/.

J. Michael Steele. *The Cauchy-Schwarz Master Class: An Introduction to the Art of Mathematical Inequalities*. Cambridge University Press, Cambridge, April 26 2004. ISBN 052154677X. URL http://www-stat.wharton.upenn.edu/~steele/Publications/Books/CSMC/CSMC_index.htm.

Lynn Arthur Steen. Highlights in the history of spectral theory. *The American Mathematical Monthly*, 80(4):359–381, April 1973. ISSN 00029890. URL http://www.jstor.org/stable/2319079.

Stephen M. Stigler. Gauss and the invention of least squares. *The Annals of Statistics*, 9(3):465–474, May 1981. URL https://projecteuclid.org/euclid.aos/1176345451. https://www.jstor.org/stable/2240811.

Josef Stoer and R. Bulirsch. *Introduction to Numerical Analysis*, volume 12 of *Texts in Applied Mathematics*. Springer Science & Business Media, 3 edition, August 21 2002. ISBN 9780387954523. URL http://books.google.com/books?vid=ISBN038795452X.

Marshall Harvey Stone. *Linear transformations in Hilbert space and their applications to analysis*, volume 15 of *American Mathematical Society. Colloquium publications*. American Mathematical Society, New York, 1932. URL http://books.google.com/books?vid=ISBN0821810154. 1990 reprint of the original 1932 edition.

Gilbert Strang. Wavelets and dilation equations: A brief introduction. *SIAM Review*, 31(4): 614–627, December 1989. ISSN 0036-1445. URL https://pdfs.semanticscholar.org/d118/a25a91488b4b37457cdc506472819de90523.pdf.

Gilbert Strang and Truong Nguyen. *Wavelets and Filter Banks*. Wellesley-Cambridge Press, Wellesley, MA, 1996. ISBN 9780961408879. URL http://books.google.com/books?vid=ISBN0961408871.

Alan Stuart and J. Keith Ord. *Kendall's Advanced Theory of Statistics Volume 2 Classical Inference and Relationship*. Hodder & Stoughton, 5 edition, 1991. ISBN 9780340560235. URL http://books.google.com/books?vid=ISBN9780340560235.

Yuri M. Suhov, Michael Kelbert, and Mark Kelbert. *Probability and Statistics by Example: Volume 1, Basic Probability and Statistics*, volume 1 of *Probability and Statistics by Example*. Cambridge University Press, illustrated, reprint edition, October 13 2005. ISBN 9780521847667.

Wim Sweldens and Robert Piessens. Wavelet sampling techniques. In *1993 Proceedings of the Statistical Computing Section*, pages 20–29. American Statistical Association, August 1993. URL http://citeseer.ist.psu.edu/18531.html.

Tamás Szabó and Gábor Horváth. An efficient hardware implementation of feed-forward neural networks. *Applied Intelligence*, 1:143–158, 2004. URL https://www.researchgate.net/profile/Gabor_Horvath/publication/263107833.

Erik Talvila. Necessary and sufficient conditions for differentiating under the integral sign. *The American Mathematical Monthly*, 108(6):544–548, June–July 2001. URL http://arxiv.org/abs/math/0101012.

Brian S. Thomson, Andrew M. Bruckner, and Judith B. Bruckner. *Elementary Real Analysis*. www.classicalrealanalysis.com, 2 edition, 2008. ISBN 9781434843678. URL http://classicalrealanalysis.info/com/Elementary-Real-Analysis.php.

J. T. L. Thong, K. S. Sim, and J. C. H. Phang. Single-image signal-to-noise ratio estimation. *The Journal of Scanning Microscopes*, 23:328–336, December 07 2001. URL https://doi.org/10.1002/sca.4950230506.

Elmer Tolsted. An elementary derivation of cauchy, hölder, and minkowski inequalities from young's inequality. *Mathematics Magazine*, 37:2–12, 1964. URL http://mathdl.maa.org/mathDL/22/?pa=content&sa=viewDocument&nodeId=3036.

Harry L. Van Trees. *Otimum Array Processing; Part IV of Detection, Estimation, and Modulation Theory*. Wiley-Interscience, New York, 2002. ISBN 0-471-09390-4. URL http://ece.gmu.edu/faculty_info/van.html.

Kishor S. Trivedi. *Probability and Statistics with Reliability, Queuing, and Computer Science Applications*. John Wiley & Sons, 2 edition, 2016. ISBN 9781119314202.

Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer Science & Business Media, 2008. ISBN 9780387790527. URL http://books.google.com/books?vid=ISBN9780387790527.

Howard G. Tucker. *An Introduction to Probability and Mathematical Statistics*. Academic press textbooks in mathematics. Academic Press, 1 edition, 1965.

Constantin Udriste. *Convex Functions and Optimization Methods on Riemannian Manifolds*, volume 297 of *Mathematics and Its Applications*. Springer, July 31 1994. ISBN 0792330021. URL http://books.google.com/books?vid=ISBN0792330021.

UEA. Hadcrut4 global temperature dataset. *Climate Research Unit Data*, 2019. URL http://www.cru.uea.ac.uk/data.

Stanislaw Marcin Ulam. *Adventures of a Mathematician*. University of California Press, Berkeley, 1991. ISBN 0520071549. URL http://books.google.com/books?vid=ISBN0520071549.

P.P. Vaidyanathan. *Multirate Systems and Filter Banks*. Prentice Hall Signal Processing Series. Prentice Hall, 1993. ISBN 0136057187. URL http://books.google.com/books?vid=ISBN0136057187.

Jussi Väisälä. A proof of the mazur-ulam theorem. *The American Mathematical Monthly*, 110(7): 633–635, August–September 2003. URL http://www.helsinki.fi/~jvaisala/mazurulam.pdf.

M.L.J. van de Vel. *Theory of Convex Structures*, volume 50 of *North-Holland Mathematical Library*. North-Holland, Amsterdam, 1993. ISBN 978-0444815057. URL http://books.google.com/books?vid=ISBN0444815058.

Balthasar van der Pol and Hendricus Bremmer. *Operational Calculus based on the Two-Sided Laplace Integral*. Cambridge at the University Press, 2nd, reprinted edition, 1959. URL https://archive.org/details/in.ernet.dli.2015.141269/.

Saeed V. Vaseghi. *Advanced Digital Signal Processing and Noise Reduction*. John Wiley & Sons, 2 edition, 2000. ISBN 9780471626923. URL http://www.amazon.com/dp/0471626929.

Brani Vidakovic. *Statistical Modeling by Wavelets*. John Wiley & Sons, Inc, New York, 1999. ISBN 9780471293651. URL http://www.amazon.com/dp/0471293652.

Wolfgang von der Linden, Volker Dose, and Udo Von Toussaint. *Bayesian Probability Theory: Applications in the Physical Sciences*. Cambridge University Press, 2014. ISBN 9781107035904. URL http://books.google.com/books?vid=ISBN1107035902.

John von Neumann. Allgemeine eigenwerttheorie hermitescher funktionaloperatoren. *Mathematische Annalen*, 102(1):49–131, 1929. ISSN 0025-5831 (print) 1432-1807 (online). URL http://resolver.sub.uni-goettingen.de/purl?GDZPPN002273535. General eigenvalue theory of Hermitian functional operators.

John von Neumann. *Continuous Geometry*. Princeton mathematical series. Princeton University Press, Princeton, 1960. URL http://books.google.com/books?id=3bjqOgAACAAJ.

Gilbert G. Walter and XiaoPing Shen. *Wavelets and Other Orthogonal Systems*. Chapman and Hall/CRC, New York, 2 edition, 2001. ISBN 9781584882275. URL http://books.google.com/books?vid=ISBN1584882271.

Larry Wasserman. *All of Nonparametric Statistics*. Springer Texts in Statistics. Springer Science & Business Media, 2006. ISBN 9780387306230. URL http://books.google.com/books?vid=ISBN9780387306230.

Larry Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer Texts in Statistics. Springer Science & Business Media, 2013. ISBN 9780387217369. URL http://books.google.com/books?vid=ISBN9780387217369.

Heinrich Martin Weber. Die allgemeinen grundlagen der galois'schen gleichungstheorie. *Mathematische Annalen*, 43(4):521–549, December 1893. URL http://resolver.sub.uni-goettingen.de/purl?GDZPPN002254670. The general foundation of Galois' equation theory.

Joseph H. Maclagan Wedderburn. On hypercomplex numbers. *Proceedings of the London Mathematical Society*, 6:77–118, 1907. URL http://plms.oxfordjournals.org/cgi/reprint/s2-6/1/77.

H. L. Wei and S. A. Billings. An efficient nonlinear cardinal b-spline model for high tide forecasts at the venice lagoon. *Nonlinear Processes in Geophysics*, 13:577–584, 2006. URL https://www.nonlin-processes-geophys.net/13/577/2006/. http://eprints.whiterose.ac.uk/1945/1/weihl1_NPG06-13(5)pp577-584.pdf.

David Wells. *The Penguin Dictionary of Curious and Interesting Numbers*. Penguin Books, revised edition, April 7 1987. ISBN 9780140080292. URL http://www.amazon.com/dp/0140080295.

Hermann Weyl. The method of orthogonal projection in potential theory. *Duke Mathematical Journal*, 7(1):411–444, 1940. URL http://projecteuclid.org/euclid.dmj/1077492266.

P. R. White, M. H. Tan, and J. K. Hammond. Analysis of the maximum likelihood, total least squares and principal component approaches for frequency response function estimation. *Journal of Sound and Vibration*, 290(3–5):676–689, March 2006. URL https://doi.org/10.1016/j.jsv.2005.04.029.

Alfred L. Wicks and Harvard Vold. The h_s frequency response function estimator. In *Proceedings of the 4th International Modal Analysis Conference*, pages 897–899, Los Angeles Airport Marriott Hotel, California, USA, February 3–6 1986. URL https://www.tib.eu/en/search/id/TIBKAT%3A246489537/Proceedings-of-the-4th-International-Modal-Analysis/.

Harold Widom. Toeplitz matrices. In Jr. I. I. Hirschmann, editor, *Studies in Real and Complex Analysis*, volume 3 of *MAA Studies in Mathematics*, pages 179–209. Mathematical Association of America; distributed by Prentice-Hall, 1965. ISBN 9780883851036. URL https://www.worldcat.org/title/studies-in-real-and-complex-analysis/oclc/506377.

W. John Wilbur. Quantum logic and the locally convex spaces. *Transactions of the American Mathematical Society*, 207:343–360, June 1975. URL http://www.jstor.org/stable/1997181.

Samuel Stanley Wilks. *Mathematical Statistics*. Wiley series in probability and mathematical statistics. Wiley, 1 edition, 1963a.

Samuel Stanley Wilks. *Mathematical Statistics*. Wiley series in probability and mathematical statistics. Wiley, "second printing with corrections" edition, 1963b. URL https://books.google.com/books?id=-sJLAAAAMAAJ.

P. Wojtaszczyk. *A Mathematical Introduction to Wavelets*, volume 37 of *London Mathematical Society student texts*. Cambridge University Press, February 13 1997. ISBN 9780521578943. URL http://books.google.com/books?vid=ISBN0521578949.

Herman Wold. *A Study in the Analysis of Stationary Time Series*. Uppsala, 1 edition, 1938. URL https://archive.org/details/in.ernet.dli.2015.262214. https://archive.org/details/in.ernet.dli.2015.261682.

Herman Wold. *A Study in the Analysis of Stationary Time Series*. Almqvist & Wiksell, 2 edition, 1954. URL https://archive.org/details/in.ernet.dli.2015.262214.

G. W. Wornell and A. V. Oppenheim. Estimation of fractal signals from noisy measurements using wavelets. *IEEE Transactions on Signal Processing*, 40(3):611–623, March 1992. ISSN print: 1053-587X, electronic: 1941-0476. URL https://ieeexplore.ieee.org/abstract/document/120804.

W. J. Yan and W. X. Ren. Operational modal parameter identification from power spectrum density transmissibility. *Computer-Aided Civil and Infrastructure Engineering*, 27(3):202–217, 2012. URL https://doi.org/10.1111/j.1467-8667.2011.00735.x.

Frank Yates and Kenneth Mather. Ronald aylmer fisher, 1890–1962. *Biographical Memoirs of Fellows of the Royal Society of London*, 9:90–120, November 1963. URL https://doi.org/10.1098/rsbm.1963.0006. print ISSN:0080-4606, online ISSN:1748-8494, https://pdfs.semanticscholar.org/6ccd/4303f68f3543fc794c376ce00395947a79b6.pdf.

Robert M. Young. *An introduction to nonharmonic Fourier series*, volume 93 of *Pure and applied mathematics*. Academic Press, 1 edition, 1980. ISBN 0127728503. URL http://books.google.com/books?vid=ISBN0127728503.

Robert M. Young. *An introduction to nonharmonic Fourier series*, volume 93 of *Pure and applied mathematics*. Academic Press, revised first edition, May 16 2001. ISBN 0127729550. URL http://books.google.com/books?vid=ISBN0127729550.

William Henry Young. On classes of summable functions and their fourier series. *Proceedings of the Royal Society of London*, 87(594):225–229, August 1912. URL http://www.archive.org/details/philtrans02496252.

Yu Zhang, Tong Zhu, and Jing Zhou. Understanding power spectrum density transmissibility. *Shock and Vibration*, 2016(8479721):9, 2016. URL http://dx.doi.org/10.1155/2016/8479721.

Mu-Qiao Zheng, She-Ping Hu, and Li Li. Improved methods for the estimate of frequency response functions. pages 917–921, 2002. URL https://pdfs.semanticscholar.org/edfc/0572a727633e729204bc9a8cbce4eb3211e5.pdf.

Yun-Lai Zhou and Magd Abdel Wahab. Republished paper. rapid early damage detection using transmissibility with distance measure analysis under unknown excitation in long-term health monitoring. *Journal of Vibroengineering*, 20(1):823–831, 2018. doi: 10.21595/jve.2016.19718. URL https://doi.org/10.21595/jve.2016.19718.

Gary Zukav. *The Dancing Wu Li Masters : An Overview of the New Physics*. Bantam Books, New York, 1980. ISBN 055326382X. URL http://books.google.com/books?vid=ISBN055326382X.

REFERENCE INDEX

SUBJECT INDEX

License

…last page …please stop reading …