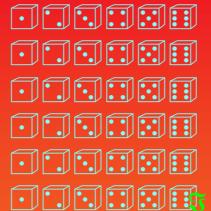
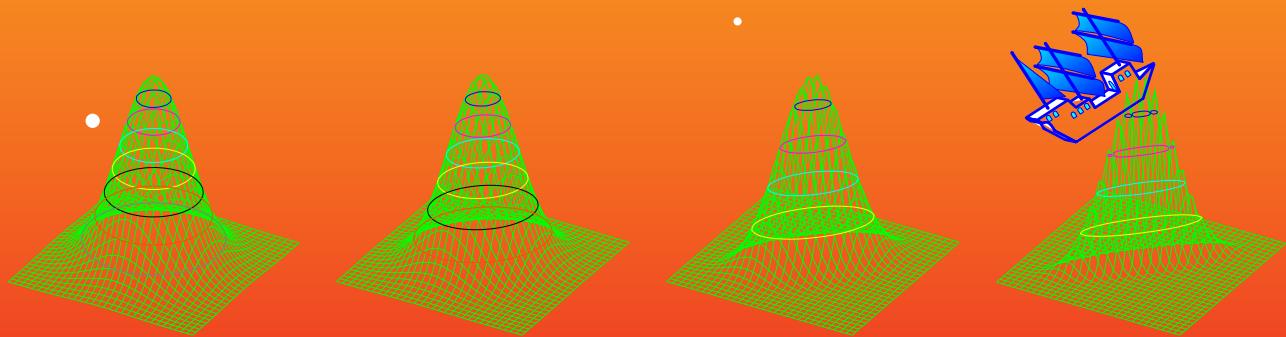
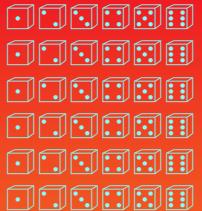


# A Book Concerning Statistical Signal Processing

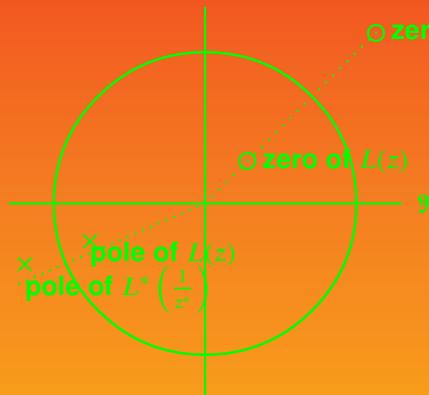
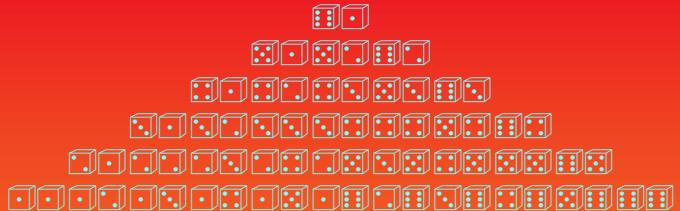
VERSION 0.19



+



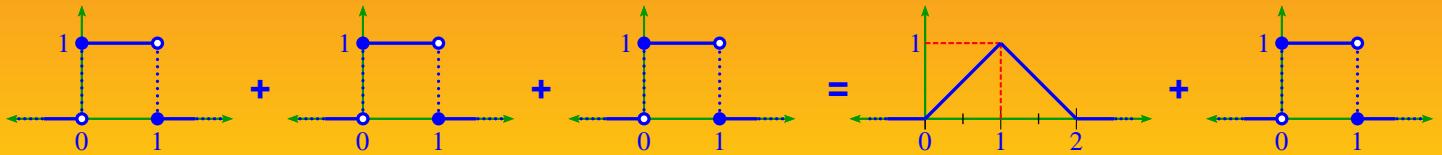
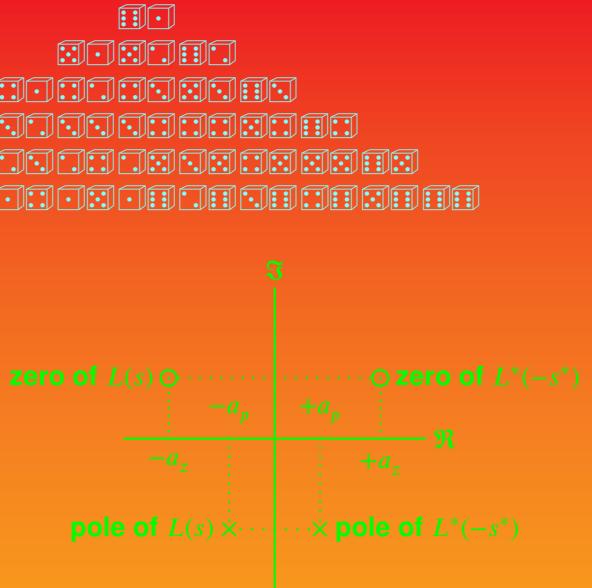
=



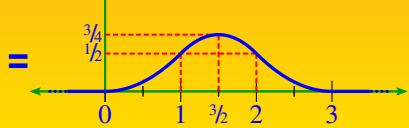
○ zero of  $L^*(\frac{1}{z^*})$

$\Re$

Daniel J. Greenhoe



Signal Processing ABCs series  
volume 2









title: *A Book Concerning Statistical Signal Processing*  
document type: book  
series: *Signal Processing ABCs*  
volume: 2  
author: Daniel J. Greenhoe  
version: VERSION 0.19  
time stamp: 2019 October 30 (Wednesday) 02:27pm UTC  
copyright: Copyright © 2019 Daniel J. Greenhoe  
license: Creative Commons license CC BY-NC-ND 4.0  
typesetting engine: X<sub>E</sub>La<sub>T</sub>E<sub>X</sub>  
document url: <https://github.com/dgreenhoe/pdfs/blob/master/abcstat.pdf>  
<https://www.researchgate.net/project/Signal-Processing-ABCs>

This text was typeset using X<sub>E</sub><sup>A</sup>T<sub>E</sub>X, which is part of the T<sub>E</sub>Xfamily of typesetting engines, which is arguably the greatest development since the Gutenberg Press. Graphics were rendered using the *pstricks* and related packages, and L<sub>A</sub>T<sub>E</sub>X 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 T<sub>E</sub>X-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.<sup>1</sup>



---

<sup>1</sup>  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) <sup>2</sup>



“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 <sup>3</sup>



“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. <sup>4</sup>



<sup>2</sup> quote:  Housman (1936), page 64 (“Smooth Between Sea and Land”),  Hardy (1940) (section 7)

image: <http://en.wikipedia.org/wiki/Image:Housman.jpg>

<sup>3</sup> quote:  Ewen (1961), page 408,  Ewen (1950)

image: [http://en.wikipedia.org/wiki/Image:Igor\\_Stravinsky.jpg](http://en.wikipedia.org/wiki/Image:Igor_Stravinsky.jpg)

<sup>4</sup> quote:  Heijenoort (1967), page 127

image: <http://www-history.mcs.st-andrews.ac.uk/PictDisplay/Russell.html>

## 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 <sup>5</sup>



**“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, <sup>6</sup>

## Symbol list

symbol	description
numbers:	
$\mathbb{Z}$	integers
$\mathbb{W}$	whole numbers
$\mathbb{N}$	natural numbers
$\mathbb{Z}^+$	non-positive integers

*...continued on next page...*

<sup>5</sup>quote: [Descartes \(1684a\)](#) (rugula XVI), translation: [Descartes \(1684b\)](#) (rule XVI), image: Frans Hals (circa 1650), <http://en.wikipedia.org/wiki/Descartes>, public domain

<sup>6</sup>quote: [Cajori \(1993\)](#) (paragraph 540), image: [http://en.wikipedia.org/wiki/File:Gottfried\\_Wilhelm\\_von\\_Leibniz.jpg](http://en.wikipedia.org/wiki/File:Gottfried_Wilhelm_von_Leibniz.jpg), public domain

symbol	description
$\mathbb{Z}^-$	negative integers $\dots, -3, -2, -1$
$\mathbb{Z}_o$	odd integers $\dots, -3, -1, 1, 3, \dots$
$\mathbb{Z}_e$	even integers $\dots, -4, -2, 0, 2, 4, \dots$
$\mathbb{Q}$	rational numbers $\frac{m}{n}$ with $m \in \mathbb{Z}$ and $n \in \mathbb{Z} \setminus 0$
$\mathbb{R}$	real numbers completion of $\mathbb{Q}$
$\mathbb{R}^+$	non-negative real numbers $[0, \infty)$
$\mathbb{R}^-$	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 $\mathbb{C}$ )
$\infty$	positive infinity
$-\infty$	negative infinity
$\pi$	pi 3.14159265 ...
relations:	
$\circledR$	relation
$\circledcirc$	relational and
$X \times Y$	Cartesian product of $X$ and $Y$
$(\Delta, \nabla)$	ordered pair
$ z $	absolute value of a complex number $z$
$=$	equality relation
$\triangleq$	equality by definition
$\rightarrow$	maps to
$\in$	is an element of
$\notin$	is not an element of
$D(\circledR)$	domain of a relation $\circledR$
$I(\circledR)$	image of a relation $\circledR$
$R(\circledR)$	range of a relation $\circledR$
$N(\circledR)$	null space of a relation $\circledR$
set relations:	
$\subseteq$	subset
$\subsetneq$	proper subset
$\supseteq$	super set
$\supsetneq$	proper superset
$\not\subseteq$	is not a subset of
$\not\subsetneq$	is not a proper subset of
operations on sets:	
$A \cup B$	set union
$A \cap B$	set intersection
$A \Delta B$	set symmetric difference
$A \setminus B$	set difference
$A^c$	set complement
$ \cdot $	set order
$\mathbb{1}_A(x)$	set indicator function or characteristic function
logic:	
1	“true” condition
0	“false” condition
$\neg$	logical NOT operation

*...continued on next page...*

symbol	description
$\wedge$	logical AND operation
$\vee$	logical inclusive OR operation
$\oplus$	logical exclusive OR operation
$\Rightarrow$	“implies”;
$\Leftarrow$	“implied by”;
$\Leftrightarrow$	“if and only if”;
$\forall$	universal quantifier:
$\exists$	existential quantifier:
order on sets:	
$\vee$	join or least upper bound
$\wedge$	meet or greatest lower bound
$\leq$	reflexive ordering relation
$\geq$	reflexive ordering relation
$<$	irreflexive ordering relation
$>$	irreflexive ordering relation
measures on sets:	
$ X $	order or counting measure of a set $X$
distance spaces:	
$d$	metric or distance function
linear spaces:	
$\ \cdot\ $	vector norm
$\ \cdot\ $	operator norm
$\langle \Delta   \nabla \rangle$	inner-product
$\text{span}(V)$	span of a linear space $V$
algebras:	
$\Re$	real part of an element in a $*$ -algebra
$\Im$	imaginary part of an element in a $*$ -algebra
set structures:	
$T$	a topology of sets
$R$	a ring of sets
$A$	an algebra of sets
$\emptyset$	empty set
$2^X$	power set on a set $X$
sets of set structures:	
$\mathcal{T}(X)$	set of topologies on a set $X$
$\mathcal{R}(X)$	set of rings of sets on a set $X$
$\mathcal{A}(X)$	set of algebras of sets on a set $X$
classes of relations/functions/operators:	
$2^{XY}$	set of <i>relations</i> from $X$ to $Y$
$Y^X$	set of <i>functions</i> from $X$ to $Y$
$S_j(X, Y)$	set of <i>surjective</i> functions from $X$ to $Y$
$I_j(X, Y)$	set of <i>injective</i> functions from $X$ to $Y$
$B_j(X, Y)$	set of <i>bijective</i> functions from $X$ to $Y$
$B(X, Y)$	set of <i>bounded</i> functions/operators from $X$ to $Y$
$L(X, Y)$	set of <i>linear bounded</i> functions/operators from $X$ to $Y$
$C(X, Y)$	set of <i>continuous</i> functions/operators from $X$ to $Y$
specific transforms/operators:	
$\tilde{F}$	<i>Fourier Transform</i> operator (Definition P.2 page 337)
$\hat{F}$	<i>Fourier Series</i> operator

*...continued on next page...*

symbol	description
$\tilde{\mathbf{F}}$	<i>Discrete Time Fourier Series operator</i> (Definition Q.1 page 347)
$\mathbf{Z}$	<i>Z-Transform operator</i> (Definition R.4 page 358)
$\tilde{f}(\omega)$	<i>Fourier Transform of a function <math>f(x) \in L^2_{\mathbb{R}}</math></i>
$\tilde{x}(\omega)$	<i>Discrete Time Fourier Transform of a sequence <math>(x_n \in \mathbb{C})_{n \in \mathbb{Z}}</math></i>
$\tilde{x}(z)$	<i>Z-Transform of a sequence <math>(x_n \in \mathbb{C})_{n \in \mathbb{Z}}</math></i>

---

---

## SYMBOL INDEX

$\bar{\delta}_n$ , 271	$x^+$ , 204	$\text{span}$ , 209	$[\cdot : \cdot]$ , 285, 286
$(A, \ \cdot\ , *)$ , 257	$d$ , 207	$*$ , 254	$Y^X$ , 308
$\ \cdot\ $ , 275	$I_m$ , 255	$\ \cdot\ $ , 312	$\rho$ , 180, 253
$\perp$ , 206	$I$ , 308	$\odot$ , 282, 284	$\sigma_c$ , 180
$\star$ , 357	$R_e$ , 255	$\star$ , 340	$\sigma_p$ , 180
$\sigma$ , 371	$Z$ , 358	$B(X, Y)$ , 315	$\sigma_r$ , 180
$\tilde{F}$ , 338	$N_n(x)$ , 371	$(\cdot : \cdot)$ , 285, 286	$\sigma$ , 180, 253
$ x $ , 204	$\text{epi}(f)$ , 287	$(\cdot : \cdot]$ , 285, 286	$r$ , 253
$x^-$ , 204	$\text{hyp}(f)$ , 287	$[\cdot : \cdot)$ , 285, 286	



# CONTENTS

Title page . . . . .	v
Typesetting . . . . .	vi
Quotes . . . . .	vii
Symbol list . . . . .	ix
Symbol index . . . . .	xiii
Contents . . . . .	xv
<b>I Statistical Analysis</b>	<b>1</b>
<b>1 Expectation operator</b>	<b>3</b>
1.1 Definitions . . . . .	3
1.2 Expectation as a linear operator . . . . .	4
1.3 Expectation inequalities . . . . .	6
1.4 Conditional expectation . . . . .	7
1.5 Expectation inner product space . . . . .	8
<b>2 Random Sequences</b>	<b>11</b>
2.1 Definitions . . . . .	11
2.2 Properties . . . . .	12
2.3 Wide Sense Stationary processes . . . . .	12
2.4 Spectral density . . . . .	15
2.5 Spectral Power . . . . .	17
<b>3 Continuous Random Processes</b>	<b>19</b>
3.1 Definitions . . . . .	19
3.2 Properties . . . . .	20
<b>4 Random Process Eigen-Analysis</b>	<b>21</b>
4.1 Definitions . . . . .	21
4.2 Properties . . . . .	21
4.3 Quasi-basis . . . . .	25
<b>II Statistical Processing</b>	<b>27</b>
<b>5 Operations on Random Variables</b>	<b>29</b>
5.1 Functions of one random variable . . . . .	29
5.2 Functions of two random variables . . . . .	35
<b>6 Operators on Discrete Random Sequences</b>	<b>39</b>
6.1 LTI operators on random sequences . . . . .	39
6.2 LTI operators on WSS random sequences . . . . .	40
6.3 Parallel operators on WSS random sequences . . . . .	41
6.4 Whitening discrete random sequences . . . . .	42

<b>7 Operators on Continuous Random Sequences</b>	<b>45</b>
7.1 LTI Operations on non-stationary random processes . . . . .	45
7.2 LTI Operations on WSS random processes . . . . .	47
7.3 Whitening continuous random sequences . . . . .	52
<b>8 Additive noise on random sequences</b>	<b>55</b>
8.1 Additive noise and correlation . . . . .	55
8.2 Additive noise and operators . . . . .	57
8.3 Additive noise and LTI operators . . . . .	60
8.4 Additive noise and dual operators . . . . .	62
<b>III Statistical Estimation</b>	<b>65</b>
<b>9 Estimation Overview</b>	<b>67</b>
9.1 Model-based estimation . . . . .	67
9.1.1 Estimation models . . . . .	67
9.1.2 Estimation algorithms . . . . .	68
9.1.3 Estimation criterion function . . . . .	69
9.2 Estimation applications . . . . .	70
9.3 Measures of estimator quality . . . . .	70
<b>10 Least Squares Algorithms</b>	<b>73</b>
10.1 Minimum mean square estimation . . . . .	74
10.2 Least squares . . . . .	76
10.2.1 General results . . . . .	76
10.2.2 Examples using polynomial modeling . . . . .	77
10.3 Recursive forms . . . . .	80
10.4 Direct search . . . . .	82
<b>11 Projection Statistic Algorithms</b>	<b>83</b>
11.1 Projection Statistics . . . . .	83
11.2 Sufficient Statistics . . . . .	85
11.3 Additive noise . . . . .	88
11.4 ML estimates . . . . .	90
11.5 Example data . . . . .	98
11.6 Matched Filter Algorithm . . . . .	99
11.7 Colored noise . . . . .	100
<b>12 Bayesian Network Algorithms</b>	<b>103</b>
12.1 Sequential decoding . . . . .	103
12.2 References . . . . .	103
<b>IV System Estimation / Modeling</b>	<b>105</b>
<b>13 Transfer Function Estimation</b>	<b>107</b>
13.1 Estimation techniques . . . . .	107
13.2 Additive noise system models . . . . .	108
13.3 Transfer function estimate definitions and interpretation . . . . .	109
13.4 Estimator relationships . . . . .	114
13.5 Alternate forms . . . . .	119
13.6 Least squares estimates of non-linear systems . . . . .	121
13.7 Least squares estimates of linear systems . . . . .	125
13.8 Coherence . . . . .	129
13.8.1 Application . . . . .	129
13.8.2 Definitions . . . . .	130
13.8.3 A warning . . . . .	130
<b>14 Noise Estimation</b>	<b>133</b>
<b>15 Moment Estimation</b>	<b>135</b>
15.1 Mean Estimation . . . . .	135

15.2 Variance Estimation . . . . .	138
15.3 Estimates in terms of moment estimates . . . . .	141
15.4 Recursive forms . . . . .	142
<b>16 Density Estimation</b>	<b>143</b>
16.1 Introduction . . . . .	143
16.2 Parametric density estimation . . . . .	143
16.3 Nonparametric density estimation . . . . .	144
16.4 Smoothness . . . . .	144
<b>17 Correlation Estimation</b>	<b>145</b>
17.1 Time series method . . . . .	145
17.2 Spectral methods . . . . .	145
<b>18 Spectral Estimation</b>	<b>147</b>
18.1 Nonparametric spectral estimation . . . . .	147
18.2 Bandwidth-Time Product . . . . .	148
<b>19 Dimensionality Reduction</b>	<b>149</b>

## V Appendices 151

<b>A Probability Space</b>	<b>153</b>
A.1 Probability functions . . . . .	154
A.2 Probability Space . . . . .	155
A.3 Properties . . . . .	156
A.4 Examples . . . . .	158
A.5 Probability subspaces . . . . .	162
<b>B Probability Density Functions</b>	<b>163</b>
B.1 Random variables . . . . .	163
B.2 Probability distributions . . . . .	164
B.3 Properties . . . . .	164
<b>C Some Probability Density Functions</b>	<b>169</b>
C.1 Discrete distributions . . . . .	169
C.2 Continuous distributions . . . . .	169
C.2.1 Uniform distribution . . . . .	169
C.2.2 Gaussian distribution . . . . .	170
C.2.3 Gamma distribution . . . . .	171
C.2.4 Chi-squared distributions . . . . .	172
C.2.5 Radial distributions . . . . .	175
C.3 Joint Gaussian distributions . . . . .	176
<b>D Spectral Theory</b>	<b>179</b>
D.1 Operator Spectrum . . . . .	179
D.2 Fredholm kernels . . . . .	181
<b>E Matrix Calculus</b>	<b>183</b>
E.1 First derivative of a vector with respect to a vector . . . . .	183
E.2 First derivative of a matrix with respect to a scalar . . . . .	190
E.3 Second derivative of a scalar with respect to a vector . . . . .	192
E.4 Multiple derivatives of a vector with respect to a scalar . . . . .	193
<b>F Algebraic structures</b>	<b>195</b>
<b>G Linear spaces</b>	<b>197</b>
G.1 Definition and basic results . . . . .	197
G.2 Order on Linear Spaces . . . . .	202
<b>H Linear Combinations</b>	<b>209</b>
H.1 Linear combinations in linear spaces . . . . .	209

H.2	Bases in topological linear spaces . . . . .	213
H.3	Schauder bases in Banach spaces . . . . .	214
H.4	Linear combinations in inner product spaces . . . . .	216
H.5	Orthonormal bases in Hilbert spaces . . . . .	219
H.6	Riesz bases in Hilbert spaces . . . . .	227
H.7	Frames in Hilbert spaces . . . . .	232
<b>I</b>	<b>Linear Subspaces</b>	<b>239</b>
I.1	Subspaces of a linear space . . . . .	239
I.2	Subspaces of an inner product space . . . . .	245
I.3	Subspaces of a Hilbert Space . . . . .	249
I.4	Subspace Metrics . . . . .	251
I.5	Literature . . . . .	252
<b>J</b>	<b>Normed Algebras</b>	<b>253</b>
J.1	Algebras . . . . .	253
J.2	Star-Algebras . . . . .	254
J.3	Normed Algebras . . . . .	257
J.4	C* Algebras . . . . .	257
<b>K</b>	<b>Inner Product Spaces</b>	<b>259</b>
K.1	Definition and basic results . . . . .	259
K.2	Relationship between norms and inner products . . . . .	263
K.2.1	Norms induced by inner products . . . . .	263
K.2.2	Inner products induced by norms . . . . .	265
K.3	Orthogonality . . . . .	271
<b>L</b>	<b>Normed Linear Spaces</b>	<b>275</b>
L.1	Definition and basic results . . . . .	275
L.2	Relationship between metrics and norms . . . . .	277
L.2.1	Metrics generated by norms . . . . .	277
L.2.2	Norms generated by metrics . . . . .	280
L.3	Orthogonality on normed linear spaces . . . . .	282
<b>M</b>	<b>Intervals and Convexity</b>	<b>285</b>
M.1	Intervals . . . . .	285
M.2	Convex sets . . . . .	286
M.3	Convex functions . . . . .	287
M.4	Literature . . . . .	290
<b>N</b>	<b>Finite Sums</b>	<b>293</b>
N.1	Summation . . . . .	293
N.2	Means . . . . .	294
N.2.1	Weighted $\phi$ -means . . . . .	294
N.2.2	Power means . . . . .	296
N.3	Inequalities on power means . . . . .	299
N.4	Power Sums . . . . .	304
<b>O</b>	<b>Operators on Linear Spaces</b>	<b>307</b>
O.1	Operators on linear spaces . . . . .	307
O.1.1	Operator Algebra . . . . .	307
O.1.2	Linear operators . . . . .	308
O.2	Operators on Normed linear spaces . . . . .	312
O.2.1	Operator norm . . . . .	312
O.2.2	Bounded linear operators . . . . .	315
O.2.3	Adjoints on normed linear spaces . . . . .	317
O.2.4	More properties . . . . .	318
O.3	Operators on Inner product spaces . . . . .	319
O.3.1	General Results . . . . .	319
O.3.2	Operator adjoint . . . . .	320
O.4	Special Classes of Operators . . . . .	322
O.4.1	Projection operators . . . . .	322

O.4.2 Self Adjoint Operators . . . . .	324
O.4.3 Normal Operators . . . . .	325
O.4.4 Isometric operators . . . . .	327
O.4.5 Unitary operators . . . . .	330
O.5 Operator order . . . . .	335
<b>P Fourier Transform</b>	<b>337</b>
P.1 Definitions . . . . .	337
P.2 Operator properties . . . . .	338
P.3 Convolution . . . . .	340
P.4 Real valued functions . . . . .	341
P.5 Moment properties . . . . .	342
P.6 Examples . . . . .	344
<b>Q Discrete Time Fourier Transform</b>	<b>347</b>
Q.1 Definition . . . . .	347
Q.2 Properties . . . . .	347
Q.3 Derivatives . . . . .	355
<b>R Operations on Sequences</b>	<b>357</b>
R.1 Convolution operator . . . . .	357
R.2 Z-transform . . . . .	358
R.3 From z-domain back to time-domain . . . . .	360
R.4 Zero locations . . . . .	361
R.5 Pole locations . . . . .	362
R.6 Mirroring for real coefficients . . . . .	363
R.7 Rational polynomial operators . . . . .	364
R.8 Filter Banks . . . . .	365
R.9 Inverting non-minimum phase filters . . . . .	369
<b>S B-Splines</b>	<b>371</b>
S.1 Definitions . . . . .	371
S.2 Algebraic properties . . . . .	375
S.3 Projection properties . . . . .	384
S.4 Fourier analysis . . . . .	385
S.5 Basis properties . . . . .	387
S.5.1 Uniqueness properties . . . . .	387
S.5.2 Partition of unity properties . . . . .	388
S.5.3 Riesz basis properties . . . . .	389
S.6 Mutiresolution properties . . . . .	395
S.6.1 Introduction . . . . .	395
S.6.2 B-spline dyadic decomposition . . . . .	395
S.6.3 B-spline MRA scaling functions . . . . .	397
S.6.4 B-spline MRA coefficient sequences . . . . .	398
<b>T Source Code</b>	<b>403</b>
<b>Back Matter</b>	<b>407</b>
References . . . . .	408
Reference Index . . . . .	441
Subject Index . . . . .	445
License . . . . .	463
End of document . . . . .	465



# **Part I**

## **Statistical Analysis**



# CHAPTER 1

## EXPECTATION OPERATOR

### 1.1 Definitions

In a *probability space*  $(\Omega, \mathbb{E}, P)$  (Definition A.2 page 155), 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*  $E$ .

**Definition 1.1.** Let  $(\Omega, \mathbb{E}, P)$  be a PROBABILITY SPACE (Definition A.2 page 155) and  $X$  a RANDOM VARIABLE (Definition B.1 page 164) on  $(\Omega, \mathbb{E}, P)$  with PROBABILITY DENSITY FUNCTION  $p_x$ .

**D E F** The *expectation operator*  $E_x$  on  $X$  is defined as

$$E_x X \triangleq \int_{x \in \mathbb{E}} x p_x(x) dx.$$

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 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, even though  $X$  appears to be random, the expected value  $E_x X$  of  $X$  is **not random**. Rather it is a fixed value (like 0 or 7.9 or -2.6).

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  $(\Omega, \mathbb{E}, P)$ .

- D E F**
- (1). The **mean**  $\mu_X$  of  $X$  is  $\mu_X \triangleq E_x X$
  - (2). The **variance**  $\text{Var}(X)$  or  $\sigma_X^2$  of  $X$  is  $\text{Var}(X) \triangleq E_x [(X - E_x X)^2]$

## 1.2 Expectation as a linear operator

The next theorem demonstrates that the operator  $E$  is a *linear operator* (Definition O.3 page 308)—which in turn makes  $E$  part of a distinguished club of operators along with fellow member operators differentiation  $\frac{d}{dx}$ , integration  $\int dx$ , Laplace  $L$ , Fourier  $\tilde{F}$ , z-transform  $Z$ , etc. Because  $E$  is a linear operator, it immediately inherits all the properties that its linear operator birthright grants it (Corollary 1.1 page 4).

**Theorem 1.1** (Linearity of  $E$ ). <sup>1</sup> Let  $X$  be a RANDOM VARIABLE on a PROBABILITY SPACE  $(\Omega, \mathbb{E}, P)$ .

T H M	$E_x(aX + bY + c) = (aE_x X) + (bE_x Y) + c \quad \forall a, b, c \in \mathbb{R} \quad (\text{LINEAR})$
-------------	---

PROOF:

$$\begin{aligned}
 E_{xy}(aX + bY + c) &\triangleq \int_{x \in \mathbb{R}} \int_{y \in \mathbb{R}} [ax + by + c] p_{xy}(x, y) dy dx \quad \text{by definition of } E \text{ (Definition 1.1 page 3)} \\
 &= \int_{x \in \mathbb{R}} \int_{y \in \mathbb{R}} ax p_{xy}(x, y) dy dx + \int_{x \in \mathbb{R}} \int_{y \in \mathbb{R}} by p_{xy}(x, y) dy dx + \int_{x \in \mathbb{R}} \int_{y \in \mathbb{R}} c p_{xy}(x, y) dy dx \\
 &= \int_{x \in \mathbb{R}} ax \underbrace{\int_{y \in \mathbb{R}} p_{xy}(x, y) dy}_{p_x(x)} dx + \int_{y \in \mathbb{R}} by \underbrace{\int_{x \in \mathbb{R}} p_{xy}(x, y) dx}_{p_y(y)} dy + c \underbrace{\int_{y \in \mathbb{R}} \int_{x \in \mathbb{R}} p_{xy}(x, y) dx dy}_{1} \\
 &= a \underbrace{\int_{x \in \mathbb{R}} x p_x(x) dx}_{EX} + b \underbrace{\int_{y \in \mathbb{R}} y p_y(y) dy}_{EY} + c \\
 &= \boxed{(aE_x X) + (bE_y Y) + c}
 \end{aligned}$$



**Corollary 1.1.** Let  $E$  be the EXPECTATION OPERATOR over a PROBABILITY SPACE  $(\Omega, \mathbb{E}, P)$ . Let  $sPLL$  be a VECTOR SPACE OF RANDOM VARIABLES over  $(\Omega, \mathbb{E}, P)$ .

C O R	(1). $E\emptyset = \emptyset$ and (2). $E(-X) = -(EX) \quad \forall X \in L^2_F$ and (3). $E(X - Y) = EX - EY \quad \forall X, Y \in L^2_F$ and	(4). $E\left(\sum_{n=1}^N \alpha_n X_n\right) = \sum_{n=1}^N \alpha_n (EX_n) \quad \forall \alpha_n \in \mathbb{F}, \quad \forall X \in L^2_F$
-------------	---	--

PROOF: These all follow immediately from the fact that  $E$  is a *linear operator* and from Theorem O.1 (page 308). ⇒

*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  $\tilde{F}$ , Laplace  $L$ , and z-transform  $Z$  ...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  $\tilde{F}x(n)$  of a random sequence  $x(n)$  is another random sequence.

One way to overcome this difficulty is to simply invoke a *sampling* operator  $Sx(n)$ , 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.

<sup>1</sup> Haykin (2014) page 107 (“PROPERTY 1 Linearity”), Wilks (1963b), page 73 §§3.2 “Mean value of a random variable”



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

The good news here is that because  $E$  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  $M$  and  $N$  be LINEAR OPERATORS (Definition O.3 page 308).

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

PROOF: These all follow immediately from the fact that  $E$  is a *linear operator* (Theorem 1.1 page 4) and from properties of all linear operators (Theorem O.4 page 311).  $\Rightarrow$

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

C O R	$\text{Var}(aX + b) = a^2 \text{Var}(X) \quad \forall a, b \in \mathbb{R}$
	$\text{Var}(X) = E_x(X^2) - (E_x X)^2$

PROOF:

$$\begin{aligned}
 \text{Var}(X) &\triangleq E_x[(X - E_x X)^2] && \text{by definition of } \text{Var} \quad (\text{Definition 1.2 page 3}) \\
 &= E_x[X^2 - 2XE_x X + (E_x X)^2] && \text{by Binomial Theorem} \\
 &= E_x X^2 - E_x[2XE_x X] + E_x(E_x X)^2 && \text{by linearity of } E \quad (\text{Theorem 1.1 page 4}) \\
 &= E_x X^2 - 2(E_x X)[E_x X] + (E_x X)^2 \\
 &= E_x(X^2) - (E_x X)^2
 \end{aligned}$$

$$\begin{aligned}
 \text{Var}(aX + b) &= E_x(aX + b)^2 - [E_x(aX + b)]^2 \\
 &= E_x(a^2 X^2 + 2abX + b^2) - [a(E_x X) + b]^2 \\
 &= a^2 E_x X^2 + 2ab E_x X + b^2 - [a^2 [E_x X]^2 + 2ab E_x X + b^2] && \text{by linearity of } E \quad (\text{Theorem 1.1 page 4}) \\
 &= a^2 [E_x X^2 - (E_x X)^2] \\
 &\triangleq a^2 \text{Var}(X) && \text{by previous result}
 \end{aligned}$$

$\Rightarrow$

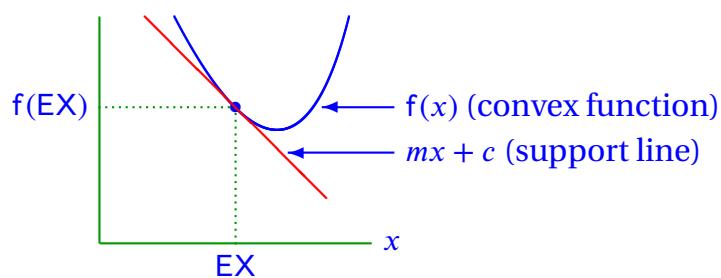


Figure 1.1: Jensen's inequality

*Jensen's inequality* is an extremely useful application of *convexity* (Definition M.9 page 288) to the *expectation* operator. Jensen's inequality is stated in Corollary 1.4 (next) and illustrated in Figure 1.1 (page 5).

**Corollary 1.4** (Jensen's inequality). <sup>2</sup> Let  $f$  be a function in  $\mathbb{R}^{\mathbb{R}}$  and  $X$  be a RANDOM VARIABLE on  $(\Omega, \mathbb{E}, P)$ .

C O R	$\{f \text{ is CONVEX}\} \implies \{f(\mathbb{E}X) \leq \mathbb{E}f(X)\}$
-------------	---

PROOF:

1. Proof 1: Let  $mx + c$  be a “support line” under  $f(x)$  (Figure 1.1 page 5) such that

$$\begin{aligned} mx + c &< f(x) \quad \text{for } x \neq \mathbb{E}X \\ mx + c &= f(x) \quad \text{for } x = \mathbb{E}X. \end{aligned}$$

Then

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

2. Proof 2 (alternate proof):

$$\begin{aligned} f(\mathbb{E}X) &\triangleq f\left(\sum_{x \in \mathbb{E}} x P(x)\right) \\ &\leq \sum_{x \in \mathbb{E}} f(x)P(x) \quad \text{by Jensen's inequality for convex sets} \quad (\text{Theorem M.1 page 288}) \end{aligned}$$

Example 1.1. <sup>3</sup> Some examples of *Jensen's Inequality* (Corollary 1.4 page 6) applied to the *expectation operator* are the following:

E X	$(\mathbb{E}X)^{-1} < \mathbb{E}(X^{-1})$   $\mathbb{E}(\log X) < \log(\mathbb{E}X)$   $e^{-\mathbb{E}X} \leq \mathbb{E}[e^{-X}]$
--------	---

**Theorem 1.2** (Law of the Unconscious Statistician). <sup>4</sup>

T H M	$\mathbb{E}[g(X)] = \int_{x \in \mathbb{R}} g(x)p_x(x) dx$
-------------	--

## 1.3 Expectation inequalities

**Theorem 1.3** (Markov's inequality). <sup>5</sup> Let  $X : \Omega \rightarrow [0, \infty)$  be a non-negative valued RANDOM VARIABLE and  $a \in (0, \infty)$ . Then

T H M	$P\{X \geq a\} \leq \frac{1}{a} \mathbb{E}X$
-------------	--

<sup>2</sup> Shao (2003) page 31 (“1.3 Distributions and Their Characteristics”), Cover and Thomas (1991), page 25, Jensen (1906), pages 179–180

<sup>3</sup> Shao (2003) pages 31–32 (“Example 1.18”), Dekking et al. (2006) page 110 (“8.5 Solutions to the quick exercises”)

<sup>4</sup> 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)

<sup>5</sup> Ross (1998), page 395



PROOF:

$$\begin{aligned} I &\triangleq \begin{cases} 1 & \text{for } X \geq a \\ 0 & \text{for } X < a \end{cases} \\ aI &\leq X \\ I &\leq \frac{1}{a}X \\ EI &\leq E\left(\frac{1}{a}X\right) \end{aligned}$$

$$\begin{aligned} P\{X \geq a\} &= 1 \cdot P\{X \geq a\} + 0 \cdot P\{X < a\} \\ &= EI \\ &\leq E\left(\frac{1}{a}X\right) \\ &= \frac{1}{a}EX \end{aligned}$$



**Theorem 1.4** (Chebyshev's inequality). <sup>6</sup> Let  $X$  be a RANDOM VARIABLE with mean  $\mu$  and variance  $\sigma^2$ .

T H M	$P\{ X - \mu  \geq a\} \leq \frac{\sigma^2}{a^2}$
-------------	---

PROOF:

$$\begin{aligned} P\{|X - \mu| \geq a\} &= P\{(X - \mu)^2 \geq a^2\} \\ &\leq \frac{1}{a^2} E(X - \mu)^2 && \text{by Markov's inequality} && (\text{Theorem 1.3 page 6}) \\ &= \frac{\sigma^2}{a^2} \end{aligned}$$



**Theorem 1.5** (Kolmogorov's inequality). <sup>7</sup> Let  $X$  be a RANDOM VARIABLE with mean  $\mu$  and variance  $\sigma^2$ .

T H M	$\left\{ \begin{array}{ll} (A). & (x_n) \text{ are INDEPENDENT} \\ (B). & \text{Each } x_n \text{ has ZERO-MEAN} \\ (C). & \text{Each } x_n \text{ has variance } \sigma^2 \end{array} \text{ and } \right\} \implies P\left[\left \sum_{n=1}^N x_n\right  < \lambda \sum_{n=1}^N x_n^2\right] \geq 1 - \frac{1}{\lambda^2}$
-------------	--

## 1.4 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  $E_x X$  of  $X$  is **not random**. On the other hand, note that  $E(X|Y)$  is **random**. This is because  $E(X|Y)$  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  $E(X|Y = y)$  is also fixed. However, if we don't know the value of  $Y$ , then  $Y$  is still a *random variable* and the expression  $E(X|Y)$  is also random (a function of *random variable*  $Y$ ).

<sup>6</sup> Ross (1998), page 396

<sup>7</sup> Wilks (1963b), page 107 (§4.5 “Kolmogorov's inequality”)

**Theorem 1.6.** <sup>8</sup> Let  $X$  and  $Y$  be RANDOM VARIABLES. Then

T H M	$E_x X = E_y E_{x y}(X Y)$
-------------	----------------------------

PROOF:

$$\begin{aligned}
 E_y E_{x|y}(X|Y) &\triangleq E_y \left[ \int_{x \in \mathbb{R}} x p(X=x|Y) dx \right] && \text{by definition of } E && (\text{Definition 1.1 page 3}) \\
 &\triangleq \int_{y \in \mathbb{R}} \left[ \int_{x \in \mathbb{R}} x p(x|Y=y) dx \right] p(y) dy && \text{by definition of } E && (\text{Definition 1.1 page 3}) \\
 &= \int_{y \in \mathbb{R}} \int_{x \in \mathbb{R}} x p(x|y) p(y) dx dy \\
 &= \int_{x \in \mathbb{R}} x \int_{y \in \mathbb{R}} p(x,y) dy dx && \text{by Theorem B.3 page 165} \\
 &= \int_{x \in \mathbb{R}} x p(x) dx && \text{by Theorem B.3 page 165} \\
 &\triangleq E_x X && \text{by definition of } E && (\text{Definition 1.1 page 3})
 \end{aligned}$$



## 1.5 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.  $E(XY)$ ) 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*, *normed spaces*, *vector spaces*, *metric spaces*, and *topological spaces*.

**Theorem 1.7.** <sup>9</sup> Let  $R$  be a ring,  $(\Omega, \mathbb{E}, P)$  be a PROBABILITY SPACE,  $E$  the expectation operator, and  $V = \{X|X : \Omega \rightarrow R\}$  be the set of all random vectors in PROBABILITY SPACE  $(\Omega, \mathbb{E}, P)$ .

- |             |  |
|-------------|--|
| T<br>H<br>M | (1). $V \triangleq \{X X : \Omega \rightarrow R\}$ is a VECTOR SPACE.<br>(2). $\langle X   Y \rangle \triangleq E(XY^*)$ is an INNER PRODUCT.<br>(3). $\ X\  \triangleq \sqrt{E(XX^*)}$ is a NORM.<br>(4). $(V, \langle \Delta   \nabla \rangle)$ is an INNER PRODUCT SPACE. |
|-------------|--|

PROOF:

<sup>8</sup> Jazwinski (1970), page 40 (“Theorem 2.9 (Conditional Expectations)”), Jazwinski (2007) page 40

<sup>9</sup> Lindquist and Picci (2015) pages 25–26 (2.1 Hilbert Space of Second-Order Random Variables.  $\langle \xi | \eta \rangle = E\{\xi \bar{\eta}\}$ ), Caines (1988) page 21 ( $E(XY) = \int_{\Omega} x(\omega)y(\omega)dP(\omega)$ ), Caines (2018) page 21 ( $E(XY) = \int_{\Omega} x(\omega)y(\omega)dP(\omega)$ ), Moon and Stirling (2000), pages 105–106

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

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

2. Proof that  $\langle X | Y \rangle \triangleq E(XY^*)$  is an *inner product*.

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

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

4. Proof that  $(\mathbf{V}, \langle \cdot | \cdot \rangle)$  is an *inner product space*: Because  $\mathbf{V}$  is a vector space and  $\langle \cdot | \cdot \rangle$  is an *inner product*,  $(\mathbf{V}, \langle \cdot | \cdot \rangle)$  is an *inner product space*.



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

**Theorem 1.8.** Let  $(\Omega, \mathcal{E}, P)$  be a PROBABILITY SPACE with EXPECTATION functional  $E$ .

<b>T H M</b>	$(1). \quad \sqrt{E\left(\sum_{n=1}^N X_n\right)} \leq \sum_{n=1}^N E(X_n X_n^*) \quad (\text{GENERALIZED TRIANGLE INEQUALITY})$ $(2). \quad  E(XY^*) ^2 \leq E(XX^*) E(YY^*) \quad (\text{CAUCHY-SCHWARTZ INEQUALITY})$ $(3). \quad 2E(XX^*) + 2E(YY^*) = E[(X + Y)(X + Y)^*] + E[(X - Y)(X - Y)^*] \quad (\text{PARALLELOGRAM LAW})$
----------------------	--

PROOF:

1.  $(\mathbb{R}^\Omega, E(x, y))$  is an *inner product space*. Proof: Theorem 1.7 (page 8).

2. Because  $(\mathbb{R}^\Omega, E(x, y))$  is an *inner product space*, the other properties follow:

1. Generalized triangle inequality: Theorem L.1 page 275
2. Cauchy-Schwartz inequality: Theorem K.2 page 260
3. Parallelogram Law: Theorem K.7 page 267





# CHAPTER 2

## RANDOM SEQUENCES



“A likely impossibility is always preferable to an unconvincing possibility.”<sup>1</sup>  
Aristotle (384 BC – 322 BC)



“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<sup>2</sup>

## 2.1 Definitions

### Definition 2.1.

**D E F** A **random sequence** is a **SEQUENCE**  
over a **PROBABILITY SPACE** (Definition A.2 page 155).

**Definition 2.2.**<sup>3</sup> Let  $x(n)$  and  $y(n)$  be RANDOM SEQUENCES.

<sup>1</sup> quote: <http://en.wikiquote.org/wiki/Aristotle>  
image: <http://en.wikipedia.org/wiki/Aristotle>

<sup>2</sup> quote: [Yates and Mather \(1963\)](#) page 107. image: <http://www.genetics.org/content/154/4/1419>

<sup>3</sup> [Papoulis \(1984\)](#) page 263  $\langle R_{xy}(m) = E\{x(m)y^*(0)\} \rangle$ , [Wilks \(1963b\)](#), page 77 §3.4 “Moments of two-dimensional random variables”, [Cadzow \(1987\)](#) page 341  $\langle r_{xy}(m) = E[x(m)y^*(0)] \rangle$ , [MatLab \(2018b\)](#)  $\langle R_{xy}(m) = E\{x_{n+m}y_n^*\} \rangle$ , [MatLab \(2018a\)](#)  $\langle R_{xy}(m) = E\{x_{n+m}y_n^*\} \rangle$

<b>D E F</b>	<b>The mean</b>	$\mu_X(n)$	of $x(n)$ is	$\mu_X(n) \triangleq E[x(n)]$
	<b>The variance</b>	$\sigma_X^2(n)$	of $x(n)$ is	$\sigma_X^2(n) \triangleq E([x(n) - \mu_X(n)]^2)$
	<b>The cross-correlation</b>	$R_{xy}(n, m)$	of $x(n)$ and $y(n)$ is	$R_{xy}(n, m) \triangleq E[x(n + m)y^*(n)]$
	<b>The auto-correlation</b>	$R_{xx}(n, m)$	of $x(n)$ is	$R_{xx}(n, m) \triangleq R_{xy}(n, m) _{y=x}$

## 2.2 Properties

### Theorem 2.1.

<b>T H M</b>	$R_{xx}(n, m) = R_{xx}^*(n + m, -m)$
	$R_{xy}(n, m) = R_{yx}^*(n + m, -m)$

PROOF:

$$\begin{aligned}
 R_{xy}(n, m) &\triangleq E[x(n + m)y^*(n)] && \text{by definition of } R_{xy}(n, m) \quad (\text{Definition 2.2 page 11}) \\
 &= E[y^*(n)x(n + m)] && \text{by commutative property of } (\mathbb{C}, +, \cdot, 0, 1) \quad (\text{Definition F.5 page 196}) \\
 &= (E[y(n)x^*(n + m)])^* && \text{by distributive property of } *-\text{algebras} \quad (\text{Definition J.3 page 254}) \\
 &= (E[y(n + m - m)x^*(n + m)])^* && \text{by additive identity property of } (\mathbb{R}, +, \cdot, 0, 1) \quad (\text{Definition F.5 page 196}) \\
 &\triangleq R_{yx}^*(n + m, -m) && \text{by definition of } R_{xy}(n, m) \quad (\text{Definition 2.2 page 11})
 \end{aligned}$$
  

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



## 2.3 Wide Sense Stationary processes

**Definition 2.3.** Let  $x(n)$  be a RANDOM SEQUENCE with MEAN  $\mu_X(n)$  and VARIANCE  $\sigma_X^2(n)$  (Definition 2.2 page 11).

<b>D E F</b>	$x(n)$ is wide sense stationary (WSS) if
	1. $\mu_X(n)$ is CONSTANT with respect to $n$ (STATIONARY IN THE 1ST MOMENT) and
	2. $\sigma_X^2(n)$ is CONSTANT with respect to $n$ (STATIONARY IN THE 2ND MOMENT)

**Definition 2.4.**<sup>4</sup> Let  $x(n)$  be a RANDOM SEQUENCE with statistics  $\mu_X(n)$ ,  $\sigma_X^2(n)$ ,  $R_{xx}(n, m)$ , and  $R_{xy}(n, m)$  (Definition 2.2 page 11).

<b>D E F</b>	$\{x \text{ and } y \text{ are WIDE SENSE STATIONARY}\} \implies$
	{ 1. The mean $\mu_X$ of $x(n)$ is $\mu_X \triangleq E[x(0)]$ }
	{ 2. The variance $\sigma_X^2$ of $x(n)$ is $\sigma_X^2 \triangleq E([x(0) - \mu_X]^2)$ }
	{ 4. The cross-correlation $R_{xy}(m)$ of $x(n)$ and $y(n)$ is $R_{xy}(m) \triangleq E[x(m)y^*(0)]$ }
	{ 3. The auto-correlation $R_{xx}(m)$ of $x(n)$ is $R_{xx}(m) \triangleq R_{xy}(m) _{y=x}$ }

<sup>4</sup> Papoulis (1984) page 263 “ $R_{xy}(\tau) = E\{x(t + \tau)y^*(t)\}$ ”, Cadzow (1987) page 341  $\langle r_{xy}(n) = E[x(k + n)y^*(k)] \rangle$  (10.41)»

*Remark 2.1.* The  $R_{xy}(n, m)$  of Definition 2.2 (page 11) and the  $R_{xy}(m)$  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  $(\sqrt{3} + i\sqrt{2}) + (e + i\pi)$ . 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.* <sup>5</sup> The definition for  $R_{xy}(m)$  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  $2 \times 2 \times 2 = 8$  different ways to define  $R_{xy}(m)$ . <sup>6</sup> and  $R_{xx}(m)$  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*  $x_l(t)$  of a modulated *narrowband* communications signal  $x(t)$ .
3. Communications channel processing involving phase discrimination (e.g. PSK and QAM).

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

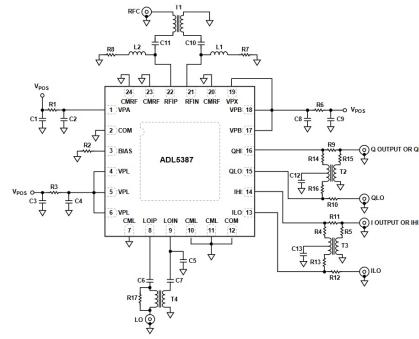
$$\begin{aligned}
 x(t) &\triangleq \underbrace{a(t)\cos[2\pi f_c t + \phi(t)]}_{\text{amplitude-phase form}} && \text{amplitude and phase form} \\
 &= \underbrace{a(t)\cos[\phi(t)]\cos[2\pi f_c t]}_{p(t)} - \underbrace{a(t)\sin[\phi(t)]\sin[2\pi f_c t]}_{q(t)} && \text{by double angle formulas} \\
 &= \underbrace{p(t)\cos[2\pi f_c t] - q(t)\sin[2\pi f_c t]}_{\text{quadrature form}} && \text{quadrature form} \\
 &= \mathbf{R}_e([p(t) + iq(t)][\cos(2\pi f_c t) + isin(2\pi f_c t)]) && \text{by definitions of } \mathbf{R}_e \\
 &= \underbrace{\mathbf{R}_e[x_l(t)e^{i2\pi f_c t}]}_{\text{complex envelope form}} && \text{by Euler's identity}
 \end{aligned}$$

Note that in these equivalent forms, the *complex envelope*  $x_l(t)$  is conveniently represented as a *complex-valued* function in terms of the *quadrature component*  $p(t)$  and the *inphase component*  $q(t)$  such that  $x_l(t) = p(t) + iq(t)$ .

<sup>5</sup>  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)

<sup>6</sup>  Greenhoe (2019a)

In practice (with real hardware), you will likely first have access to the quadrature components  $p(t)$  and  $q(t)$ . Take for example the *Analog Devices ADL5387 Quadrature Demodulator* and evaluation board, as illustrated to the right.<sup>7</sup> Note that *quadrature component p(t)* is available at connector “Q OUTPUT” and *inphase component q(t)* is available at connector “I OUTPUT”.



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

<b>P</b>	<b>R</b>	<b>P</b>	$\left\{ \begin{array}{l} x \text{ and } y \text{ are} \\ \text{WIDE SENSE STATIONARY} \\ (\text{WSS}) \end{array} \right. \text{ (Definition 7.1 page 47)} \right\} \Rightarrow \left\{ \begin{array}{ll} R_{xx}(n, m) & = R_{xx}(m) \\ R_{xy}(n, m) & = R_{xy}(m) \end{array} \right. \forall n \in \mathbb{Z} \quad \left. \begin{array}{l} (\text{Definition 2.2 page 11}) \\ (\text{Definition 2.4 page 12}) \end{array} \right\}$
----------	----------	----------	---

PROOF:

$$\begin{aligned} R_{xy}(n, m) &\triangleq E[x[n + m]y^*[n]] && \text{by definition of } R_{xy}(n, m) && (\text{Definition 2.2 page 11}) \\ &= E[x[n - n + m]y^*[n - n]] && \text{by wide sense stationary hypothesis} \\ &= E[x[m]y^*[0]] \\ &\triangleq R_{xy}(m) && \text{by definition of } R_{xy}(m) && (\text{Definition 2.4 page 12}) \end{aligned}$$

$$\begin{aligned} R_{xx}(n, m) &= R_{xy}(n, m) \Big|_{y=x} \\ &= R_{xy}(m) \Big|_{y=x} && \text{by previous result} \\ &= R_{xx}(m) \end{aligned}$$

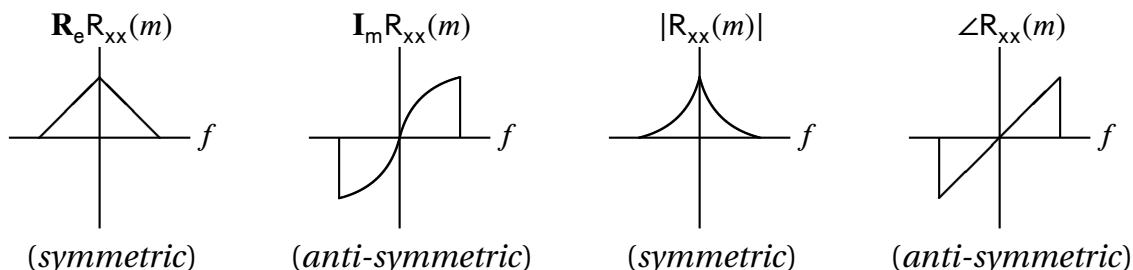


Figure 2.1: auto-correlation  $R_{xx}(m)$

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

<b>C</b>	<b>O</b>	<b>R</b>	$\left\{ \begin{array}{l} (A). \quad x \text{ is WSS} \quad \text{and} \\ (B). \quad y \text{ is WSS} \quad \text{and} \\ (C). \quad S \text{ is LTI} \end{array} \right\} \Rightarrow \left\{ \begin{array}{ll} (1). \quad R_{xy}(m) & = R_{yx}^*(-m) \\ (2). \quad R_{xx}(m) & = R_{xx}^*(-m) \quad (\text{CONJUGATE SYMMETRIC}) \\ (3). \quad R_e R_{xx}(m) & = R_e R_{xx}(-m) \quad (\text{SYMMETRIC}) \\ (4). \quad I_m R_{xx}(m) & = -I_m R_{xx}(-m) \quad (\text{ANTI-SYMMETRIC}) \\ (5). \quad  R_{xx}(m)  & =  R_{xx}(-m)  \quad (\text{SYMMETRIC}) \\ (6). \quad \angle R_{xx}(m) & = -\angle R_{xx}(-m) \quad (\text{ANTI-SYMMETRIC}) \end{array} \right. \text{ and} \end{math}$
----------	----------	----------	--

<sup>7</sup> diagram copied from [Devices \(2016\)](#)



PROOF:

$R_{xy}(m) = R_{xy}(n, m)$	by Proposition 2.1 page 14	and hypotheses (A),(B)
$= R_{yx}^*(n + m, -m)$	by Theorem 2.1 page 12	and hypothesis (B)
$= R_{yx}^*(-m)$	by Proposition 2.1 page 14	and hypothesis (A)
$R_{xx}(m) = R_{xx}(n, m)$	by Proposition 2.1 page 14	and hypothesis (A)
$= R_{xx}^*(n + m, -m)$	by Theorem 2.1 page 12	and hypothesis (B)
$= R_{xx}^*(-m)$	by Proposition 2.1 page 14	and hypothesis (A)



## 2.4 Spectral density

**Definition 2.5.** Let  $x(n)$  and  $y(n)$  be WIDE SENSE STATIONARY RANDOM SEQUENCES with auto-correlation  $R_{xx}(m)$  and cross-correlation  $R_{xy}(m)$ . Let  $Z$  be the Z-TRANSFORM OPERATOR (Definition R.4 page 358).

DEF

The z-domain cross spectral density (CSD)  $\check{S}_{xy}(z)$  of  $x$  and  $y$  is

$$\check{S}_{xy}(z) \triangleq ZR_{xy}(m) \triangleq \sum_{m \in \mathbb{Z}} R_{xy}(m)z^{-m}$$

The z-domain power spectral density (PSD)  $\check{S}_{xx}(z)$  of  $x$  is

$$\check{S}_{xx}(z) \triangleq \check{S}_{xy}(z)|_{y(n)=x(n)}$$

**Definition 2.6.** Let  $x(n)$  and  $y(n)$  be WIDE SENSE STATIONARY RANDOM SEQUENCES with auto-correlation  $R_{xx}(m)$  and cross-correlation  $R_{xy}(m)$ . Let  $\check{F}$  be the DISCRETE TIME FOURIER TRANSFORM (DTFT) operator (Definition Q.1 page 347).

DEF

The auto-spectral density

$$\check{S}_{xx}(z) \text{ of } x \text{ is} \quad \check{S}_{xx}(z) \triangleq \check{F}R_{xx}(m) \triangleq \sum_{m \in \mathbb{Z}} R_{xx}(m)e^{-i\omega m}$$

The cross spectral density

$$(CSD) \check{S}_{xy}(z) \text{ of } x \text{ and } y \text{ is} \quad \check{S}_{xy}(z) \triangleq \check{F}R_{xy}(m) \triangleq \sum_{m \in \mathbb{Z}} R_{xy}(m)e^{-i\omega m}$$

The auto-spectral density is also called power spectral density (PSD).

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

THM

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

PROOF:

$\check{S}_{yx}(z) \triangleq ZR_{yx}(m)$	by definition of $\check{S}_{xy}(z)$	(Definition 2.6 page 15)
$\triangleq \sum_{m \in \mathbb{Z}} R_{yx}(m)z^{-m}$	by definition of $Z$	(Definition R.4 page 358)
$\triangleq \sum_{m \in \mathbb{Z}} R_{xy}^*(-m)z^{-m}$	by Corollary 2.1 page 14	
$= \left[ \sum_{m \in \mathbb{Z}} R_{xy}(-m)(z^*)^{-m} \right]^*$	by antiautomorphic property of *-algebras	(Definition J.3 page 254)

$$\begin{aligned}
 &= \left[ \sum_{-p \in \mathbb{Z}} R_{xy}(p)(z^*)^p \right]^* \quad \text{where } p \triangleq -m \quad \Rightarrow m = -p \\
 &= \left[ \sum_{p \in \mathbb{Z}} R_{xy}(p)(z^*)^p \right]^* \quad \text{by } \textit{absolutely summable} \text{ property} \quad (\text{Definition R.2 page 357}) \\
 &= \left[ \sum_{p \in \mathbb{Z}} R_{xy}(p) \left( \frac{1}{z^*} \right)^{-p} \right]^* \\
 &= \check{S}_{xy}^* \left( \frac{1}{z^*} \right) \quad \text{by definition of } \mathbf{Z} \quad (\text{Definition R.4 page 358}) \\
 \check{S}_{xx}(z) &= \check{S}_{xy}(z)|_{y=x} \\
 &= \check{S}_{yx}^*(z)|_{y=x} \\
 &= \check{S}_{xy}^* \left( \frac{1}{z^*} \right)|_{y=x} \quad \text{by (2)—previous result} \\
 &= \check{S}_{xx}^* \left( \frac{1}{z^*} \right)
 \end{aligned}$$

⇒

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

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

PROOF:

$$\begin{aligned}
 \tilde{S}_{xy}^*(\omega) &= \check{S}_{xy}^*(z)|_{z=e^{i\omega}} \quad \text{by definition of DTFT} \quad (\text{Definition Q.1 page 347}) \\
 &= \check{S}_{yx}^{**} \left( \frac{1}{z^*} \right)|_{z=e^{i\omega}} \quad \text{by Theorem 2.2 page 15} \\
 &= \check{S}_{yx} \left( \frac{1}{z^*} \right)|_{z=e^{i\omega}} \quad \text{by } \textit{involutory} \text{ property of } *-\text{algebras} \quad (\text{Definition J.3 page 254}) \\
 &= \check{S}_{yx} \left( \frac{1}{e^{i\omega*}} \right) \\
 &= \check{S}_{yx} (e^{i\omega}) \\
 &= \tilde{S}_{yx}(\omega) \quad \text{by definition of DTFT} \quad (\text{Definition Q.1 page 347}) \\
 \tilde{S}_{xx}^*(\omega) &= \check{S}_{xx}^*(z)|_{z=e^{i\omega}} \quad \text{by definition of DTFT} \quad (\text{Definition Q.1 page 347}) \\
 &= \check{S}_{xx}^{**} \left( \frac{1}{z^*} \right)|_{z=e^{i\omega}} \quad \text{by Theorem 2.2 page 15} \\
 &= \check{S}_{xx} \left( \frac{1}{z^*} \right)|_{z=e^{i\omega}} \quad \text{by } \textit{involutory} \text{ property of } *-\text{algebras} \quad (\text{Definition J.3 page 254}) \\
 &= \check{S}_{xx} \left( \frac{1}{e^{i\omega*}} \right) \\
 &= \check{S}_{xx} (e^{i\omega}) \\
 &= \tilde{S}_{xx}(\omega) \quad \text{by definition of DTFT} \quad (\text{Definition Q.1 page 347}) \\
 \implies \tilde{S}_{xx}(\omega) &\text{ is real-valued} \\
 \tilde{S}_{xx}^*(\omega) &= \tilde{S}_{xy}^*(\omega)|_{y=x} \\
 &= \tilde{S}_{yx}(\omega)|_{y=x} \quad \text{by previous result} \\
 &= \tilde{S}_{xx}(\omega)
 \end{aligned}$$

⇒



## 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, *Parseval's Theorem* (Proposition H.2 page 220) demonstrates that power in time can also be calculated in frequency. 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).



# CHAPTER 3

## CONTINUOUS RANDOM PROCESSES



“*A likely impossibility is always preferable to an unconvincing possibility.*”<sup>1</sup>  
Aristotle (384 BC – 322 BC)

### 3.1 Definitions

**Definition 3.1.** <sup>2</sup> Let  $(\Omega, \mathbb{E}, \mathbb{P})$  be a PROBABILITY SPACE.

**D E F** The function  $x : \Omega \rightarrow \mathbb{R}$  is a **random variable**.  
The function  $y : \mathbb{R} \times \Omega \rightarrow \mathbb{R}$  is a **random process**.

The random process  $x(t, \omega)$ , 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  $\omega$  are fixed or allowed to vary. These forms are illustrated in Figure 3.1 page 19<sup>3</sup> and Figure 3.2 page 20.

$x(t, \omega)$	fixed $t$	variable $t$
fixed $\omega$	number	time function
variable $\omega$	random variable	random process

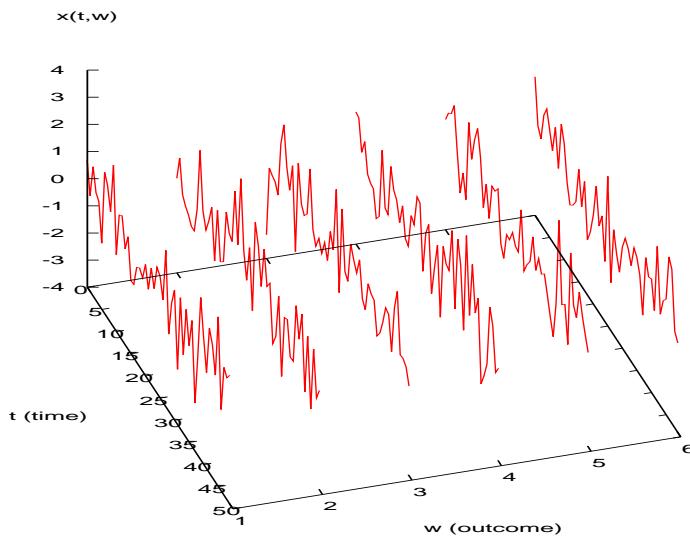
Figure 3.1: Specialized forms of a random process  $x(t, \omega)$

<sup>1</sup> quote: <http://en.wikiquote.org/wiki/Aristotle>

image: <http://en.wikipedia.org/wiki/Aristotle>

<sup>2</sup> [Papoulis \(1991\) page 63](#), [Papoulis \(1991\) page 285](#)

<sup>3</sup> [Papoulis \(1991\) pages 285–286](#)

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

**Definition 3.2.** <sup>4</sup> Let  $x(t)$  and  $y(t)$  be random processes.

DEF	The <b>mean</b> $\mu_X(t)$ of $x(t)$ is	$\mu_X(t) \triangleq E[x(t)]$
	The <b>cross-correlation</b> $R_{xy}(t, u)$ of $x(t)$ and $y(t)$ is	$R_{xy}(t, u) \triangleq E[x(t)y^*(u)]$
	The <b>auto-correlation function</b> $R_{xx}(t, u)$ of $x(t)$ is	$R_{xx}(t, u) \triangleq E[x(t)x^*(u)]$

**Remark 3.1.** <sup>5</sup> The equation  $\int_{u \in \mathbb{R}} R_{xx}(t, u)f(u) du$  is a *Fredholm integral equation of the first kind* and  $R_{xx}(t, u)$  is the *kernel* of the equation.

## 3.2 Properties

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

THM	$R_{xx}(t, u) = R_{xx}^*(u, t)$ (CONJUGATE SYMMETRIC)
	$R_{xy}(t, u) = R_{yx}^*(u, t)$

PROOF:

$$\begin{aligned} R_{xx}(t, u) &\triangleq E[x(t)x^*(u)] &= E[x^*(u)x(t)] = (E[x(u)x^*(t)])^* &\triangleq R_{xx}^*(u, t) \\ R_{xy}(t, u) &\triangleq E[x(t)y^*(u)] &= E[y^*(u)x(t)] = (E[y(u)x^*(t)])^* &\triangleq R_{yx}^*(u, t) \end{aligned}$$

⇒

<sup>4</sup> Papoulis (1984) page 216  $\langle R_{xy}(t_1, t_2) = E\{x(t_1)y^*(t_2)\} \rangle$  (9-35),

<sup>5</sup> Fredholm (1900), Fredholm (1903) page 365, Michel and Herget (1993) page 97, Keener (1988) page 101

# CHAPTER 4

## RANDOM PROCESS EIGEN-ANALYSIS

### 4.1 Definitions

**Definition 4.1.** Let  $x(t)$  be random processes with continuous AUTO-CORRELATION  $R_{xx}(t, u)$  (Definition 3.2 page 20) or discrete AUTO-CORRELATION  $R_{xx}(n, m)$  (Definition 2.2 page 11).

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

$$\begin{aligned} \mathbf{R}\mathbf{f} &\triangleq \underbrace{\int_{u \in \mathbb{R}} R_{xx}(t, u)\mathbf{f}(u) du}_{(\text{continuous case})} \quad \text{or} \quad \mathbf{R}\mathbf{f} \triangleq \underbrace{\sum_{n \in \mathbb{Z}} R_{xx}(n, m)\mathbf{f}(m)}_{(\text{discrete case})} \end{aligned}$$

**Definition 4.2.** Let  $x(t)$  be a RANDOM PROCESS with AUTO-CORRELATION  $R_{xx}(\tau)$  (Definition 3.2 page 20).

A RANDOM PROCESS  $x(t)$  is **white** if  $R_{xx}(\tau) = \delta(\tau)$

If a random process  $x(t)$  is **white** (Definition 4.2 page 21) and the set  $\Psi = \{\psi_1(t), \psi_2(t), \dots, \psi_N(t)\}$  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  $x(t)$  is **colored** (not white), then the innerproducts are not in general uncorrelated. But if the elements of  $\Psi$  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 20), the set  $\{\psi_n(t)\}$  are *orthogonal* and the innerproducts are *uncorrelated* even though  $x(t)$  is not white. This criterion is called the Karhunen-Loève criterion for  $x(t)$ .

### 4.2 Properties

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

$$\begin{aligned} \mathbf{T} \quad \left\{ \begin{array}{l} (A). \quad \langle \mathbf{x} | \mathbf{y} \rangle \triangleq \int_{t \in \mathbb{R}} \mathbf{x}(t)\mathbf{y}^*(t) dt \quad \text{OR} \\ (B). \quad \langle \mathbf{x} | \mathbf{y} \rangle \triangleq \sum_{n \in \mathbb{Z}} \mathbf{x}(n)\mathbf{y}^*(n) \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} (1). \quad \langle \mathbf{Rx} | \mathbf{x} \rangle \geq 0 \quad (\text{NON-NEGATIVE}) \quad \text{and} \\ (2). \quad \langle \mathbf{Rx} | \mathbf{y} \rangle = \langle \mathbf{x} | \mathbf{Ry} \rangle \quad (\text{SELF-ADJOINT}) \end{array} \right\} \end{aligned}$$

PROOF:

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

$$\begin{aligned}
 \langle \mathbf{R}y | y \rangle &= \left\langle \int_{u \in \mathbb{R}} R_{xx}(t, u)y(u) du | y(t) \right\rangle && \text{by definition of } \mathbf{R} \\
 &= \left\langle \int_{u \in \mathbb{R}} E[x(t)x^*(u)]y(u) du | y(t) \right\rangle && \text{by definition of } R_{xx}(t, u) \\
 &= E \left[ \left\langle \int_{u \in \mathbb{R}} x(t)x^*(u)y(u) du | y(t) \right\rangle \right] && \text{by linearity of } \langle \Delta | \nabla \rangle \text{ and } \int \\
 &= E \left[ \int_{u \in \mathbb{R}} x^*(u)y(u) du \langle x(t) | y(t) \rangle \right] && \text{by additivity property of } \langle \Delta | \nabla \rangle \\
 &= E[\langle y(u) | x(u) \rangle \langle x(t) | y(t) \rangle] && \text{by local definition of } \langle \Delta | \nabla \rangle \\
 &= E[\langle x(u) | y(u) \rangle^* \langle x(t) | y(t) \rangle] && \text{by conjugate symmetry prop.} \\
 &= E|\langle x(t) | y(t) \rangle|^2 && \text{by definition of } |\cdot| \\
 &\geq 0 && \text{by strictly positive property of norms} \quad (\text{Definition L.1 page 275})
 \end{aligned}$$

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

$$\begin{aligned}
 \langle [\mathbf{R}x](t) | y \rangle &= \left\langle \int_{u \in \mathbb{R}} R_{xx}(t, u)x(u) du | y(t) \right\rangle && \text{by definition of } \mathbf{R} \\
 &= \int_{u \in \mathbb{R}} x(u) \langle R_{xx}(t, u) | y(t) \rangle du && \text{by additive property of } \langle \Delta | \nabla \rangle \\
 &= \int_{u \in \mathbb{R}} x(u) \langle y(t) | R_{xx}(t, u) \rangle^* du && \text{by conjugate symmetry prop.} \\
 &= \langle x(u) | \langle y(t) | R_{xx}(t, u) \rangle \rangle && \text{by local definition of } \langle \Delta | \nabla \rangle \\
 &= \left\langle x(u) | \int_{t \in \mathbb{R}} y(t) R_{xx}^*(t, u) dt \right\rangle && (\text{Definition K.1 page 259}) \\
 &= \left\langle x(u) | \int_{t \in \mathbb{R}} y(t) R_{xx}(u, t) dt \right\rangle && \text{by property of } R_{xx} \\
 &= \left\langle x(u) | \underbrace{\mathbf{R}y}_{\mathbf{R}^*} \right\rangle && \text{by definition of } \mathbf{R} \\
 \implies \mathbf{R} &= \mathbf{R}^* && \text{by definition of adjoint } \mathbf{R}^* \\
 \implies \mathbf{R} &\text{ is self-adjoint} && \text{by definition of self-adjoint} \quad (\text{Definition O.8 page 324})
 \end{aligned}$$

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



**Theorem 4.2.** <sup>1</sup> 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  $\mathbf{R}\psi_n = \lambda_n\psi_n$ .

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

PROOF:

<sup>1</sup> Keener (1988), pages 114–119



1. Proof that eigenvalues are *real-valued*: Because  $\mathbf{R}$  is *self-adjoint*, its eigenvalues are real (Theorem O.18 page 324).
2. Proof that eigenfunctions associated with distinct eigenvalues are orthogonal: Because  $\mathbf{R}$  is *self-adjoint*, this property follows (Theorem O.18 page 324).
3. Proof that eigenvalues are *non-negative*:

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

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

$$\begin{aligned}
 0 &< \langle \mathbf{R}\psi_n | \psi_n \rangle && \text{by definition of } \textit{positive definite} \\
 &= \langle \lambda_n \psi_n | \psi_n \rangle && \text{by hypothesis} \\
 &= \lambda_n \langle \psi_n | \psi_n \rangle && \text{by } \textit{homogeneous property of } \langle \cdot | \cdot \rangle \quad (\text{Definition K.1 page 259}) \\
 &= \lambda_n \|\psi_n\|^2 && \text{by } \textit{induced norm theorem} \quad (\text{Theorem K.4 page 264})
 \end{aligned}$$



**Theorem 4.3** (Karhunen-Loëve Expansion). <sup>2</sup> Let  $\mathbf{R}$  be the AUTO-CORRELATION OPERATOR (Definition 4.1 page 21) of a RANDOM PROCESS  $x(t)$ . 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	$\underbrace{\ \psi_n(t)\  = 1}_{\{\psi_n(t)\} \text{ are NORMALIZED}} \implies E \left( \underbrace{\left  x(t) - \sum_{n \in \mathbb{Z}} \langle x(t)   \psi_n(t) \rangle \psi_n(t) \right ^2}_{\text{CONVERGENCE IN PROBABILITY}} \right) = 0 \quad (\{\psi_n(t)\} \text{ is a BASIS for } x(t))$
-----	--

PROOF:

1. Define  $\dot{x}_n \triangleq \langle x(t) | \psi_n(t) \rangle$
2. Define  $\mathbf{Rx}(t) \triangleq \int_{u \in \mathbb{R}} R_{xx}(t, u)x(u) du$
3. lemma:  $E[x(t)x(t)] = \sum_{n \in \mathbb{Z}} \lambda_n |\psi_n(t)|^2$ . Proof:

$$E[x(t)x(t)] = \sum_{n \in \mathbb{Z}} \lambda_n |\psi_n(t)|^2 \quad \begin{array}{l} \text{by } \textit{non-negative property} \quad (\text{Theorem 4.1 page 21}) \\ \text{and } \textit{Mercer's Theorem} \quad (\text{Theorem D.4 page 182}) \end{array}$$

<sup>2</sup> Keener (1988), pages 114–119

4. lemma:

$$\begin{aligned}
 & \mathbb{E} \left[ x(t) \left( \sum_{n \in \mathbb{Z}} \dot{x}_n \psi_n(t) \right)^* \right] \\
 & \triangleq \mathbb{E} \left[ x(t) \left( \sum_{n \in \mathbb{Z}} \int_{u \in \mathbb{R}} x(u) \psi_n^*(u) du \psi_n(t) \right)^* \right] && \text{by definition of } \dot{x} && (\text{definition 1 page 23}) \\
 & = \sum_{n \in \mathbb{Z}} \left( \int_{u \in \mathbb{R}} \mathbb{E}[x(t)x^*(u)] \psi_n(u) du \right) \psi_n^*(t) && \text{by linearity} && (\text{Theorem 1.1 page 4}) \\
 & \triangleq \sum_{n \in \mathbb{Z}} \left( \int_{u \in \mathbb{R}} R_{xx}(t, u) \psi_n(u) du \right) \psi_n^*(t) && \text{by definition of } R_{xx}(t, u) && (\text{Definition 3.2 page 20}) \\
 & \triangleq \sum_{n \in \mathbb{Z}} (\mathbf{R} \psi_n(t) \psi_n^*(t)) && \text{by definition of } \mathbf{R} && (\text{definition 2 page 23}) \\
 & = \sum_{n \in \mathbb{Z}} \lambda_n \psi_n(t) \psi_n^*(t) && \text{by property of eigen-system} \\
 & = \sum_{n \in \mathbb{Z}} \lambda_n |\psi_n(t)|^2
 \end{aligned}$$

5. lemma:

$$\begin{aligned}
 & \mathbb{E} \left[ \sum_{n \in \mathbb{Z}} \dot{x}_n \psi_n(t) \left( \sum_{m \in \mathbb{Z}} \dot{x}_m \psi_m(t) \right)^* \right] \\
 & \triangleq \mathbb{E} \left[ \sum_{n \in \mathbb{Z}} \int_{u \in \mathbb{R}} x(u) \psi_n^*(u) du \psi_n(t) \left( \sum_{m \in \mathbb{Z}} \int_v x(v) \psi_m^*(v) dv \psi_m(t) \right)^* \right] && \text{by definition of } \dot{x} && (\text{definition 1 page 23}) \\
 & = \sum_{n \in \mathbb{Z}} \sum_{m \in \mathbb{Z}} \int_u \left( \int_v \mathbb{E}[x(u)x^*(v)] \psi_m(v) dv \right) \psi_n^*(u) du \psi_n(t) \psi_m^*(t) && \text{by linearity} && (\text{Theorem 1.1 page 4}) \\
 & = \sum_{n \in \mathbb{Z}} \sum_{m \in \mathbb{Z}} \int_u \left( \int_v R_{xx}(u, v) \psi_m(v) dv \right) \psi_n^*(u) du \psi_n(t) \psi_m^*(t) && \text{by definition of } R_{xx}(t, u) && (\text{Definition 3.2 page 20}) \\
 & = \sum_{n \in \mathbb{Z}} \sum_{m \in \mathbb{Z}} \int_u (\mathbf{R} \psi_m(u)) \psi_n^*(u) du \psi_n(t) \psi_m^*(t) && \text{by definition of } \mathbf{R} && (\text{definition 2 page 23}) \\
 & = \sum_{n \in \mathbb{Z}} \sum_{m \in \mathbb{Z}} \int_u (\lambda_m \psi_m(u)) \psi_n^*(u) du \psi_n(t) \psi_m^*(t) && \text{by property of eigen-system} \\
 & = \sum_{n \in \mathbb{Z}} \sum_{m \in \mathbb{Z}} \lambda_m \left( \int_{u \in \mathbb{R}} \psi_m(u) \psi_n^*(u) du \right) \psi_n(t) \psi_m^*(t) && \text{by linearity} \\
 & = \sum_{n \in \mathbb{Z}} \sum_{m \in \mathbb{Z}} \lambda_m \|\psi(t)\|^2 \bar{\delta}_{mn} \psi_n(t) \psi_m^*(t) && \text{by orthogonal property} && (\text{Theorem 4.2 page 22}) \\
 & = \sum_{n \in \mathbb{Z}} \sum_{m \in \mathbb{Z}} \lambda_m \bar{\delta}_{mn} \psi_n(t) \psi_m^*(t) && \text{by normalized hypothesis} \\
 & = \sum_{n \in \mathbb{Z}} \lambda_n \psi_n(t) \psi_n^*(t) && \text{by definition of Kronecker delta } \bar{\delta} && (\text{Definition K.3 page 271}) \\
 & = \sum_{n \in \mathbb{Z}} \lambda_n |\psi_n(t)|^2
 \end{aligned}$$

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

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



$$\begin{aligned}
&= E \left( \left[ x(t) - \sum_{n \in \mathbb{Z}} \dot{x}_n \psi_n(t) \right] \left[ x(t) - \sum_{m \in \mathbb{Z}} \dot{x}_m \psi_m(t) \right]^* \right) \\
&= E \left( x(t)x^*(t) - x(t) \left[ \sum_{n \in \mathbb{Z}} \dot{x}_n \psi_n(t) \right]^* - x^*(t) \sum_{n \in \mathbb{Z}} \dot{x}_n \psi_n(t) + \left[ \sum_{n \in \mathbb{Z}} \dot{x}_n \psi_n(t) \right] \left[ \sum_{m \in \mathbb{Z}} \dot{x}_m \psi_m(t) \right]^* \right) \\
&= E(x(t)x^*(t)) - E \left[ x(t) \left[ \sum_{n \in \mathbb{Z}} \dot{x}_n \psi_n(t) \right]^* \right] - E \left[ x^*(t) \sum_{n \in \mathbb{Z}} \dot{x}_n \psi_n(t) \right] + E \left[ \sum_{n \in \mathbb{Z}} \dot{x}_n \psi_n(t) \left[ \sum_{m \in \mathbb{Z}} \dot{x}_m \psi_m(t) \right]^* \right] \\
&\quad \text{by linearity of } E \text{ (Theorem 1.1 page 4)} \\
&= \underbrace{\sum_{n \in \mathbb{Z}} \lambda_n |\psi_n(t)|^2}_{\text{by (3) lemma}} - \underbrace{\sum_{n \in \mathbb{Z}} \lambda_n |\psi_n(t)|^2}_{\text{by (4) lemma}} - \underbrace{\left[ \sum_{n \in \mathbb{Z}} \lambda_n |\psi_n(t)|^2 \right]^*}_{\text{by (4) lemma}} + \underbrace{\sum_{n \in \mathbb{Z}} \lambda_n |\psi_n(t)|^2}_{\text{by (5) lemma}} \\
&= 0
\end{aligned}$$



## 4.3 Quasi-basis

The *auto-correlation operator*  $\mathbf{R}$  (Definition 4.1 page 21) in the discrete case can be approximated using a *correlation matrix*. In the *zero-mean* case, this becomes

$$\mathbf{R} \triangleq \begin{bmatrix} E[y_1 y_1] & E[y_1 y_2] & \cdots & E[y_1 y_n] \\ E[y_2 y_1] & E[y_2 y_2] & & E[y_2 y_n] \\ \vdots & \vdots & \ddots & \vdots \\ E[y_n y_1] & E[y_n y_2] & \cdots & E[y_n y_n] \end{bmatrix}$$

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

**Proposition 4.1.**<sup>3</sup>

P  
R  
P

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

*Remark 4.1.* For more information about the properties of **Toeplitz matrices**, see [Grenander and Szegö \(1958\)](#), [Widom \(1965\)](#), [Gray \(1971\)](#), [Smylie et al. \(1973\) page 408](#) (§“B. PROPERTIES OF THE TOEPLITZ MATRIX”), [GRENANDER AND SZEGÖ \(1984\)](#), [HAYKIN AND KESLER \(1979\)](#), [HAYKIN AND KESLER \(1983\)](#), [S. LAWRENCE MARPLE \(1987\) PAGES 80–92](#) (§“3.8 THE TOEPLITZ MATRIX”), [BÖTTCHER AND SILBERMANN \(1999\) \(ISBN:9780387985701\)](#), [GRAY \(2006\)](#), [S. LAWRENCE MARPLE \(2019\) PAGES 80–93](#) (§“3.8 THE TOEPLITZ MATRIX”).

<sup>3</sup>See [Clarkson \(1993\) page 131](#) (§“Appendix 3A — Positive Semi-Definite Form of the Autocorrelation Matrix”)



## **Part II**

# **Statistical Processing**



# CHAPTER 5

## OPERATIONS ON RANDOM VARIABLES

### 5.1 Functions of one random variable

**Proposition 5.1.** Let  $(\Omega, \mathbb{E}, \mathbb{P})$  be a probability space, and  $X$  a RANDOM VARIABLE with CUMULATIVE DISTRIBUTION FUNCTION  $c_X(x)$ .

<b>P R P</b>	$\left\{ \begin{array}{l} X \text{ is UNIFORMLY DISTRIBUTED} \\ (\text{Definition C.1 page 169}) \end{array} \right\} \iff c_X(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } 0 < x \leq 1 \\ 1 & \text{otherwise} \end{cases}$
----------------------	--

**Theorem 5.1** (Probability integral transform). <sup>1</sup> Let  $(\Omega, \mathbb{E}, \mathbb{P})$  be a probability space. Let  $X$  be a RANDOM VARIABLE with PROBABILITY DENSITY FUNCTION  $p_X(x)$  and CUMULATIVE DISTRIBUTION FUNCTION  $c_X(x)$ . Let  $Y$  be a RANDOM VARIABLE CUMULATIVE DISTRIBUTION FUNCTION  $c_Y(y)$ .

<b>T H M</b>	$\left\{ \begin{array}{l} (1). Y = c_X(X) \\ (2). p_X(x) \text{ is CONTINUOUS} \end{array} \right. \text{ and } \right\} \implies \left\{ \begin{array}{l} Y \text{ is UNIFORMLY DISTRIBUTED} \\ (\text{Definition C.1 page 169}) \end{array} \right\}$
----------------------	---

PROOF:

$$\begin{aligned}
 c_Y(y) &\triangleq \mathbb{P}\{Y \leq y\} && \text{by definition of cdf} && (\text{Definition B.2 page 164}) \\
 &= \mathbb{P}\{c_X(X) \leq y\} && \text{by hypothesis (1)} \\
 &= \mathbb{P}\{X \leq c_X^{-1}(y)\} && \text{by hypothesis (2) and} && \text{Proposition A.2 page 158} \\
 &\triangleq c_X[c_X^{-1}(y)] && \text{by definition of cdf} && (\text{Definition B.2 page 164}) \\
 &= y \\
 \implies Y &\text{ is uniformly distributed} && \text{by} && \text{Proposition 5.1 page 29}
 \end{aligned}$$

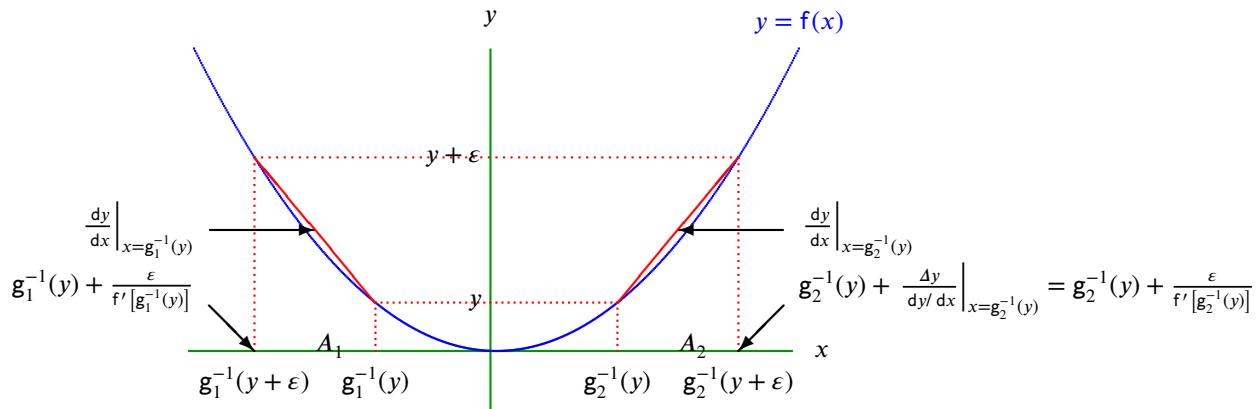
⇒

**Theorem 5.2** (Inverse probability integral transform). <sup>2</sup> Let  $(\Omega, \mathbb{E}, \mathbb{P})$  be a probability space. Let  $X$  be a RANDOM VARIABLE with PROBABILITY DENSITY FUNCTION  $p_X(x)$  and CUMULATIVE DISTRIBUTION FUNCTION  $c_X(x)$ . Let  $Y$  be a RANDOM VARIABLE CUMULATIVE DISTRIBUTION FUNCTION  $c_Y(y)$ .

<b>T H M</b>	$\left\{ \begin{array}{l} (1). Y = c_z^{-1}(X) \\ (2). Y \text{ is UNIFORMLY DISTRIBUTED} \\ (3). p_z(z) \text{ is CONTINUOUS} \end{array} \right. \text{ and } \right\} \implies \left\{ \begin{array}{l} p_Y(y) = p_z(y) \\ (Y \text{ has distribution } p_z(y)) \end{array} \right\}$
----------------------	--

<sup>1</sup> Angus (1994), Roussas (2014) page 232 (Theorem 10), Devroye (1986) page 28 (Theorem 2.1)

<sup>2</sup> Devroye (1986) page 28 (Theorem 2.1), Balakrishnan and Lai (2009) page 624 (14.2.1 Introduction)

Figure 5.1:  $Y = f(X)$ 

PROOF:

$$\begin{aligned}
 c_Y(y) &\triangleq P\{Y \leq y\} && \text{by definition of } c_Y && (\text{Definition B.2 page 164}) \\
 &= P\{c_Z^{-1}(X) \leq y\} && \text{by hypothesis (1)} && \\
 &= P\{X \leq c_Z(y)\} && \text{by hypothesis (3) and} && \text{Proposition A.2 page 158} \\
 &\triangleq c_X[c_Z(y)] && \text{by definition of } c_X && (\text{Definition B.2 page 164}) \\
 &= c_Z(y) && \text{because } 0 \leq c_Z(y) \leq 1 \text{ and by} && \text{Proposition 5.1 page 29} \\
 \implies p_Y(y) &= p_Z(y) && (\text{Y has the distribution of Z}) &&
 \end{aligned}$$

⇒

**Definition 5.1.**<sup>3</sup> Let  $f(x)$  be a DIFFERENTIABLE FUNCTION in  $\mathbb{R}^{\mathbb{R}}$ .

**D E F** A point  $p \in \mathbb{R}$  is a **critical point** off( $x$ ) if  $f'(p) = 0$ .

**Theorem 5.3.**<sup>4</sup> Let  $X$  and  $Y$  be RANDOM VARIABLES in  $\mathbb{R}^{\mathbb{R}}$ . Let  $f$  be a DIFFERENTIABLE FUNCTION in  $\mathbb{R}^{\mathbb{R}}$  with  $N$  CRITICAL POINTS (Definition 5.1 page 30). Let the range of  $X$  be partitioned into  $N + 1$  partitions  $\{A_n | n = 1, 2, \dots, N + 1\}$  with partition boundaries set at the  $N$  CRITICAL POINTS off( $x$ )—as illustrated in Figure 5.1 (page 30). Let  $g_n(x) \triangleq f(x)$  but with domain restricted to  $x \in A_n$ .

**T H M**  $\left\{ \begin{array}{l} (1). \quad Y = f(X) \\ (2). \quad f \text{ is DIFFERENTIABLE} \end{array} \right. \text{ and } \Rightarrow \left\{ p_Y(y) = \sum_{n=1}^{N+1} \frac{p_X(g_n^{-1}(y))}{|f'(g_n^{-1}(y))|} \right\}$

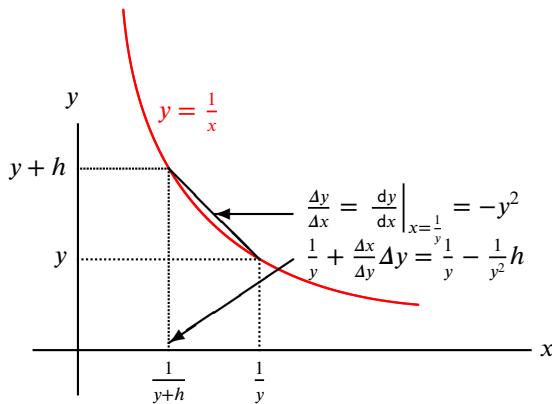
PROOF:

1. The problem with a function  $f(x)$  with at least  $N = 1$  critical point is that  $f^{-1}(y)$  is *not invertible*. That is,  $f^{-1}(y)$  has more than one solution (and thus the *relation*  $f^{-1}(y)$  is not a *function*). However, note that in each partition  $A_n$ ,  $f(x)$  is *invertible* and thus  $f^{-1}(y)$  in that partition has a *unique* solution. Thus, each  $g_n(x)$  is *invertible* in its domain (and each  $g_n^{-1}(y)$  exists as a function).

<sup>3</sup> Callahan (2010) page 189 (Definition 6.1)<sup>4</sup> Papoulis (1984) pages 95–96 (“Fundamental Theorem”), Papoulis (1990) page 157 (“Fundamental Theorem”), Papoulis (1991), page 93, Haykin (1994) page 235 (0471571768)§“4.5 TRANSFORMATIONS OF RANDOM VARIABLES”, Proakis (2001), page 30,

2. Using item (1), the remainder of the proof follows ...

$$\begin{aligned}
 p_Y(y) &\triangleq \frac{d}{dy} P\{Y \leq y\} && \text{by definition of } p_Y \text{ (Definition B.2 page 164)} \\
 &= \frac{d}{dy} P\{f(X) \leq y\} && \text{by hypothesis (1)} \\
 &= \frac{d}{dy} \sum_{n=1}^{N+1} P\{f(X) \leq y | X \in A_n\} && \text{by sum of products (Theorem A.3 page 157)} \\
 &= \frac{d}{dy} \sum_{n=1}^{N+1} P\{f(X) \leq y | X \in A_n\} P\{X \in A_n\} && \text{by definition of } P\{X|Y\} \text{ (Definition A.4 page 156)} \\
 &= \frac{d}{dy} \sum_{n=1}^{N+1} P\{g_n(X) \leq y | X \in A_n\} P\{X \in A_n\} && \text{by definition of } g_n(x) \\
 &= \frac{d}{dy} \left\{ \begin{array}{ll} \sum_{n=1}^{N+1} P\{X \leq g_n^{-1}(y) | X \in A_n\} P\{X \in A_n\} & \text{for } f'(x) \geq 0 \\ \sum_{n=1}^{N+1} P\{X \geq g_n^{-1}(y) | X \in A_n\} P\{X \in A_n\} & \text{otherwise} \end{array} \right\} && \text{by item (1)} \\
 &= \frac{d}{dy} \left\{ \begin{array}{ll} \sum_{n=1}^{N+1} P\{X \leq g_n^{-1}(y) | X \in A_n\} P\{X \in A_n\} & \text{for } f'(x) \geq 0 \\ \sum_{n=1}^{N+1} [1 - P\{X \leq g_n^{-1}(y) | X \in A_n\} P\{X \in A_n\}] & \text{otherwise} \end{array} \right\} \\
 &= \frac{d}{dy} \left\{ \begin{array}{ll} \sum_{n=1}^{N+1} P\{X \leq g_n^{-1}(y) | X \in A_n\} & \text{for } f'(x) \geq 0 \\ \sum_{n=1}^{N+1} [1 - P\{X \leq g_n^{-1}(y) | X \in A_n\}] & \text{otherwise} \end{array} \right\} && \text{by definition of } P\{X|Y\} \\
 &= \frac{d}{dy} \left\{ \begin{array}{ll} \sum_{n=1}^{N+1} (P\{X \leq g_n^{-1}(y)\} - P\{X < \min A_{n-1}\}) & \text{for } f'(x) \geq 0 \\ \sum_{n=1}^{N+1} [1 - (P\{X \leq g_n^{-1}(y)\} - P\{X < \min A_{n-1}\})] & \text{otherwise} \end{array} \right\} \\
 &= \frac{d}{dy} \left\{ \begin{array}{ll} \sum_{n=1}^{N+1} P\{X \leq g_n^{-1}(y)\} & \text{for } f'(x) \geq 0 \\ \sum_{n=1}^{N+1} [1 - P\{X \leq g_n^{-1}(y)\}] & \text{otherwise} \end{array} \right\} && \text{because } \frac{d}{dy} P\{X < \text{a constant}\} = 0 \\
 &= \left\{ \begin{array}{ll} \sum_{n=1}^{N+1} \frac{d}{dy} c_x[g_n^{-1}(y)] & \text{for } f'(x) \geq 0 \\ \sum_{n=1}^{N+1} \frac{d}{dy} [1 - c_x(g_n^{-1}(y))] & \text{otherwise} \end{array} \right\} && \text{by linearity of } \frac{d}{dy} \text{ operator} \\
 &= \left\{ \begin{array}{ll} \sum_{n=1}^{N+1} p_x[g_n^{-1}(y)] \frac{d}{dy}[g_n^{-1}(y)] & \text{for } f'(x) \geq 0 \\ \sum_{n=1}^{N+1} \left[ -p_x[g_n^{-1}(y)] \frac{d}{dy}[g_n^{-1}(y)] \right] & \text{otherwise} \end{array} \right\} && \text{by definition of } p_x \text{ (Definition B.2 page 164) and the chain rule}
 \end{aligned}$$

Figure 5.2:  $Y = \frac{1}{X}$ 

$$\begin{aligned}
 &= \sum_{n=1}^{N+1} p_x(g_n^{-1}(y)) \left| \frac{d}{dy}[g_n^{-1}(y)] \right| \\
 &= \sum_{n=1}^{N+1} \frac{p_x(g_n^{-1}(y))}{|f'(g_n^{-1}(y))|} \quad \text{by Lemma ?? page ??}
 \end{aligned}$$

⇒

**Corollary 5.1.** <sup>5</sup> Let  $X$  and  $Y$  be RANDOM VARIABLES in  $\mathbb{R}^{\mathbb{R}}$ . Let  $a, b \in \mathbb{R}$ .

<b>C O R</b>	$\left\{ \begin{array}{l} (1). \quad Y = aX + b \quad \text{and} \\ (2). \quad a \neq 0 \end{array} \right\} \implies \left\{ p_Y(y) = \frac{1}{ a } p_X\left(\frac{y-b}{a}\right) \right\}$
----------------------	--

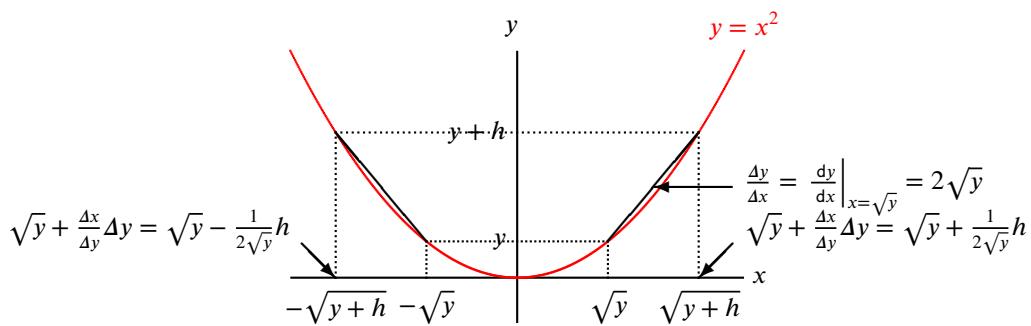
PROOF:

1. Note that  $f(x) = ax + b$  is a *differentiable function* with  $N = 0$  *critical points* and  $f'(x) = a$ .
2. The inverse of  $f(x)$  is  $g_1(y) = f^{-1}(y) = \frac{y-b}{a}$ .
3. It follows that

$$\begin{aligned}
 p_Y(y) &= \sum_{n=1}^{N+1} \frac{p_x(g_n^{-1}(y))}{|f'(g_n^{-1}(y))|} \quad \text{by Theorem 5.3 (page 30)} \\
 &= \frac{p_x(f^{-1}(y))}{|f'(f^{-1}(y))|} \quad \text{because } N = 0 \\
 &= \frac{p_x(f^{-1}(y))}{|a|} \quad \text{by item (1)} \\
 &= \frac{1}{|a|} p_X\left(\frac{y-b}{a}\right) \quad \text{by item (2)}
 \end{aligned}$$

⇒

<sup>5</sup> Papoulis (1984) page 96 (“Illustrations” 1), Papoulis (1991), page 95, Proakis (2001), page 29

Figure 5.3:  $Y = rVX^2$ **Corollary 5.2.**<sup>6</sup>

**COR**  $\left\{ Y = \frac{1}{X} \right\} \implies \left\{ p_Y(y) = \frac{1}{y^2} p_X\left(\frac{1}{y}\right) \text{ for } y > 0 \right\}$

PROOF:

1. Note that  $f(x) = 1/x$  is a *differentiable function* in  $x > 0$  with  $N = 0$  *critical points* and  $f'(x) = -1/x^2$ .
2. The inverse of  $f(x)$  is  $g_1(y) = f^{-1}(y) = \frac{1}{y}$ .
3. It follows that

$$\begin{aligned}
 p_Y(y) &= \sum_{n=1}^{N+1} \frac{p_X(g_n^{-1}(y))}{|f'(g_n^{-1}(y))|} && \text{by Theorem 5.3 (page 30)} \\
 &= \frac{p_X(f^{-1}(y))}{|f'(f^{-1}(y))|} && \text{because } N = 0 \\
 &= \frac{1}{|-1/(1/y)^2|} p_X\left(\frac{1}{y}\right) \\
 &= \frac{1}{y^2} p_X\left(\frac{1}{y}\right)
 \end{aligned}$$

**Corollary 5.3.**<sup>7</sup> Let  $X$  and  $Y$  be RANDOM VARIABLES.

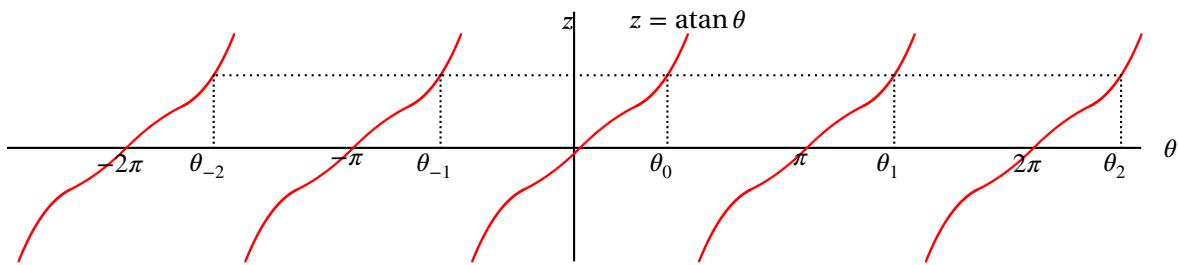
**COR**  $\left\{ Y = X^2 \right\} \implies \left\{ p_Y(y) = \frac{1}{2\sqrt{y}} [p_X(-\sqrt{y}) + p_X(\sqrt{y})] \right\}$

PROOF:

1. The roots of  $y = x^2$  are  $x_1 = -\sqrt{y}$  and  $x_2 = +\sqrt{y}$ .
2. The derivative of  $f(x) \triangleq y = x^2$  is  $f'(x) = 2x$ .

<sup>6</sup> Papoulis (1984) page 97 (Example 5-10), Papoulis (1991), page 94

<sup>7</sup> Papoulis (1984) page 95 (Example 5-9), Devroye (1986) page 27 (Example 4.4), Papoulis (1991), page 95, Proakis (2001), page 29

Figure 5.4:  $Z = \tan \Theta$ 

3. And so it follows that ...

$$\begin{aligned}
 p_Y(y) &= \sum_{n=1}^N \frac{p_X(x_n)}{|f'(x_n)|} && \text{by Theorem 5.3 page 30} \\
 &= \frac{p_X(x_1)}{|f'(x_1)|} + \frac{p_X(x_2)}{|f'(x_2)|} && \text{by definition of } \sum \\
 &= \frac{p_X(-\sqrt{y})}{|f'(-\sqrt{y})|} + \frac{p_X(\sqrt{y})}{|f'(\sqrt{y})|} && \text{by item (1)} \\
 &= \frac{p_X(-\sqrt{y})}{2\sqrt{y}} + \frac{p_X(\sqrt{y})}{2\sqrt{y}} && \text{by item (2)} \\
 &= \frac{1}{2\sqrt{y}} \left[ p_X(-\sqrt{y}) + p_X(\sqrt{y}) \right] && \text{by linearity of } + \text{ operation}
 \end{aligned}$$

⇒

**Corollary 5.4.**<sup>8</sup> Let  $Z = \tan \Theta$ . Then

C O R	$\{Z = \tan \Theta\} \implies \left\{ p_Z(z) = \frac{1}{1+z^2} \sum_{n \in Z} p_\theta(\tan(z) + n\pi) \right\}$
-------------	--

PROOF:

1. The roots of  $z = \tan \theta$  are  $\{\theta_n = \arctan z + n\pi | n \in \mathbb{Z}\}$ .
2. The derivative of  $z = \tan \theta$  is  $f'(\theta) = \sec^2 \theta$ .
3. It follows that

$$\begin{aligned}
 p_Z(z) &= \sum_{n=1}^N \frac{p_\theta(\theta_n)}{|f'(\theta_n)|} \\
 &= \sum_n \frac{p_\theta(\arctan z + n\pi)}{|f'(\arctan z + n\pi)|} \\
 &= \sum_n \frac{p_\theta(\arctan z + n\pi)}{|\sec^2(\arctan z + n\pi)|} \\
 &= \sum_n \cos^2(\arctan z + n\pi) p_\theta(\arctan z + n\pi)
 \end{aligned}$$

<sup>8</sup> Papoulis (1991), pages 99–100

$$\begin{aligned}
 &= \cos^2(\tan z) \sum_n p_\theta(\tan z + n\pi) \\
 &= \frac{1}{1+z^2} \sum_n p_\theta(\tan z + n\pi)
 \end{aligned}$$



## 5.2 Functions of two random variables

**Theorem 5.4.** <sup>9</sup> Let  $X$ ,  $Y$ , and  $Z$  be RANDOM VARIABLES. Let  $\star$  be the CONVOLUTION operator (Definition P.3 page 340).

<b>T</b> <b>H</b> <b>M</b>	$\left\{ \begin{array}{l} (1). \quad Z \triangleq X + Y \\ (2). \quad X \text{ and } Y \text{ are INDEPENDENT} \end{array} \right. \quad \text{and} \quad \begin{array}{l} \text{(Definition A.3 page 155)} \\ \hline \end{array} \right\} \Rightarrow \{p_Z(z) = p_X(z) \star p_Y(z)\}$
----------------------------------	--

PROOF:

$$\begin{aligned}
 p_Z(z) &\triangleq \frac{d}{dz} c_Z(z) && \text{by definition of } p_Z && \text{(Definition B.2 page 164)} \\
 &\triangleq \frac{d}{dz} P\{Z \leq z\} && \text{by definition of } c_Z && \text{(Definition B.2 page 164)} \\
 &= \frac{d}{dz} P\{X + Y \leq z\} && \text{by hypothesis (1)} && \\
 &= \frac{d}{dz} \lim_{\epsilon \rightarrow 0} \sum_{n \in \mathbb{Z}} P\{X + Y \leq z | y + n\epsilon < Y \leq y + (n+1)\epsilon\} && \text{by sum of products} && \text{(Theorem A.3 page 157)} \\
 &= \frac{d}{dz} \int_{y \in \mathbb{R}} P\{X + Y \leq z | Y = y\} p_Y(y) dy && \text{by definiton of } P\{X | Y\} && \text{(Definition A.4 page 156)} \\
 &= \frac{d}{dz} \int_{y \in \mathbb{R}} P\{X \leq z - y | Y = y\} p_Y(y) dy \\
 &= \frac{d}{dz} \int_{y \in \mathbb{R}} P\{X \leq z - y\} p_Y(y) dy && \text{by hypothesis (2)} && \\
 &\triangleq \frac{d}{dz} \int_{y \in \mathbb{R}} c_X(z - y) p_Y(y) dy && \text{by definition of } c_X && \text{(Definition B.2 page 164)} \\
 &= \int_{y \in \mathbb{R}} \frac{d}{dy} [c_X(z - y) p_Y(y)] dy && \text{by linearity of } \frac{d}{dz} && \\
 &= \int_{y \in \mathbb{R}} \left[ \frac{d}{dy} c_X(z - y) \right] p_Y(y) dy && \text{because } y \text{ is fixed inside the integral} && \\
 &\triangleq \int_y p_X(z - y) p_Y(y) dy && \text{by definition of } p_X && \text{(Definition B.2 page 164)} \\
 &= p_X(z) \star p_Y(z) && \text{by definition of } \star && \text{(Definition P.3 page 340)}
 \end{aligned}$$



**Theorem 5.5.** Let

- $X_1$  and  $X_2$  be random variables with joint distribution  $p_{X_1, X_2}(x_1, x_2)$
- $Y_1 = f_1(x_1, x_2)$  and  $Y_2 = f_2(x_1, x_2)$

<sup>9</sup> Papoulis (1990) page 160 (Example 5.16)

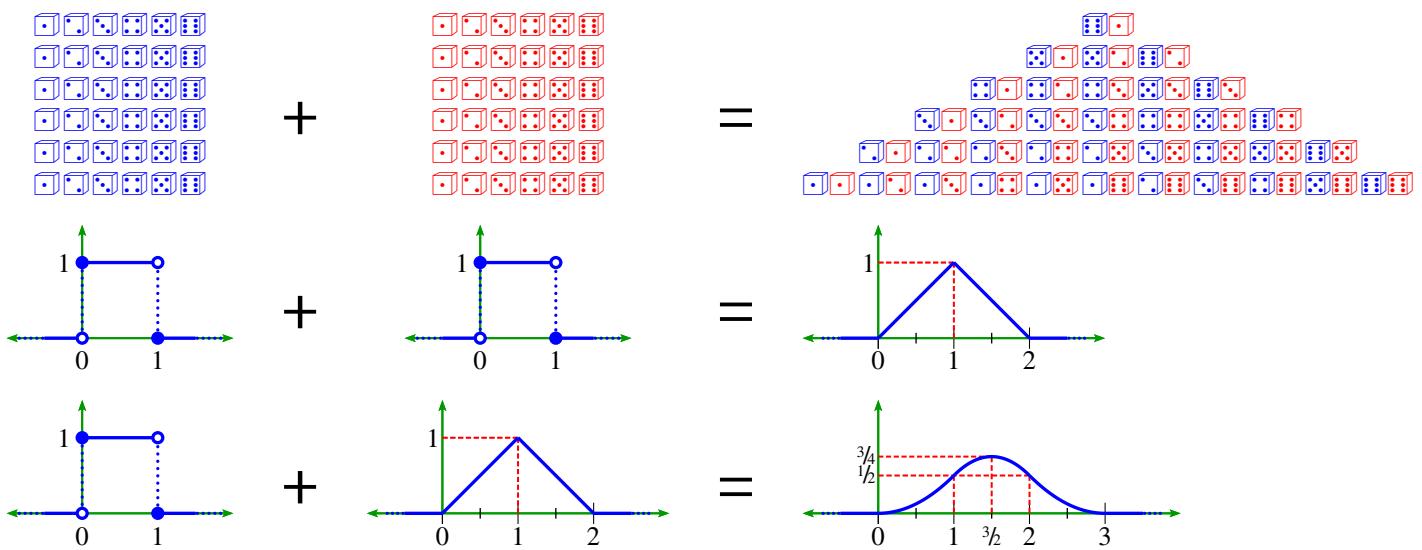


Figure 5.5: Sum of random variables yields convolution of pdfs (Theorem 5.4 page 35)

Then the joint distribution of  $Y_1$  and  $Y_2$  is

$$\text{T H M} \quad p_{Y_1, Y_2}(y_1, y_2) = \frac{p_{x_1, x_2}(x_1, x_2)}{|J(x_1, x_2)|} = \frac{p_{x_1, x_2}(x_1, x_2)}{\begin{vmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{vmatrix}} = \frac{p_{x_1, x_2}(x_1, x_2)}{\frac{\partial f_1}{\partial x_1} \frac{\partial f_2}{\partial x_2} - \frac{\partial f_1}{\partial x_2} \frac{\partial f_2}{\partial x_1}}$$

**Proposition 5.2.** Let  $X$  and  $Y$  be random variables with joint distribution  $p_{XY}(x, y)$  and

$$R^2 \triangleq X^2 + Y^2 \quad \theta \triangleq \tan^{-1} \frac{Y}{X}.$$

Then

$$\text{P R P} \quad p_{R, \theta}(r, \theta) = r p_{XY}(r \cos \theta, r \sin \theta)$$

PROOF:

$$\begin{aligned} p_{R, \theta}(r, \theta) &= \frac{p_{XY}(x, y)}{|J(x, y)|} = \frac{p_{XY}(x, y)}{\begin{vmatrix} \frac{\partial R}{\partial x} & \frac{\partial R}{\partial y} \\ \frac{\partial \theta}{\partial x} & \frac{\partial \theta}{\partial y} \end{vmatrix}} = \frac{p_{XY}(x, y)}{\begin{vmatrix} \frac{x}{\sqrt{x^2+y^2}} & \frac{y}{\sqrt{x^2+y^2}} \\ \frac{-y}{x^2+y^2} & \frac{x}{x^2+y^2} \end{vmatrix}} \\ &= \frac{p_{XY}(x, y)}{\frac{x}{\sqrt{x^2+y^2}} \frac{x}{x^2+y^2} - \frac{y}{\sqrt{x^2+y^2}} \frac{-y}{x^2+y^2}} \\ &= \frac{p_{XY}(x, y)}{\frac{x^2+y^2}{(x^2+y^2)^{3/2}}} \\ &= p_{XY}(x, y) \frac{(x^2+y^2)^{3/2}}{x^2+y^2} \\ &= p_{XY}(r \cos \theta, r \sin \theta) \frac{r^3}{r^2} \\ &= r p_{XY}(r \cos \theta, r \sin \theta) \end{aligned}$$

**Proposition 5.3.** Let  $X \sim N(0, \sigma^2)$  and  $Y \sim N(0, \sigma^2)$  be independent random variables and

$$R^2 \triangleq X^2 + Y^2 \quad \theta \triangleq \tan^{-1} \frac{Y}{X}.$$



Then

- |                                  |   |
|----------------------------------|---|
| <b>P</b><br><b>R</b><br><b>P</b> | 1. <i>R and <math>\Theta</math> are independent with joint distribution</i> $p_{R,\Theta}(r, \theta) = p_R(r)p_\theta(\theta)$<br>2. <i>R has Rayleigh distribution</i> $p_R(r) = \frac{r}{\sigma^2} \exp \frac{-r^2}{2\sigma^2}$<br>3. <i><math>\Theta</math> has uniform distribution</i> $p_\theta(\theta) = \frac{1}{2\pi}$ |
|----------------------------------|---|

PROOF:

$$\begin{aligned}
 p_{R,\Theta}(r, \theta) &= r p_{XY}(r\cos\theta, r\sin\theta) && \text{by Proposition 5.2 (page 36)} \\
 &= r p_X(r\cos\theta) p_Y(r\sin\theta) && \text{by independence hypothesis} \\
 &= r \frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{(r\cos\theta - 0)^2}{-2\sigma^2} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{(r\sin\theta - 0)^2}{-2\sigma^2} \\
 &= \frac{1}{2\pi\sigma^2} r \exp \frac{r^2(\cos^2\theta + \sin^2\theta)}{-2\sigma^2} \\
 &= \frac{1}{2\pi\sigma^2} r \exp \frac{r^2}{-2\sigma^2} \\
 &= \left[ \frac{1}{2\pi} \right] \left[ \frac{r}{\sigma^2} \exp \frac{r^2}{-2\sigma^2} \right]
 \end{aligned}$$

by Proposition 5.2 (page 36)  
by independence hypothesis

**Proposition 5.4.** Let  $X$  and  $Y$  be RANDOM VARIABLES with covariance  $\sigma_{xy}$  on a PROBABILITY SPACE  $(\Omega, \mathbb{E}, P)$ .

<b>P</b> <b>R</b> <b>P</b>	$\left\{ \begin{array}{l} (A). X \text{ is GAUSSIAN with } N(\mu_X, \sigma_X^2) \text{ and} \\ (B). Y \text{ is GAUSSIAN with } N(\mu_Y, \sigma_Y^2) \text{ and} \\ (C). \sigma_{xy} = \text{cov}[X, Y] \end{array} \right\} \Rightarrow \left\{ P\{X > Y\} = Q\left( \frac{-\mu_X + \mu_Y}{\sigma_X^2 + \sigma_Y^2 - 2\sigma_{xy}} \right) \right\}$
----------------------------------	---

PROOF: Because  $X$  and  $Y$  are jointly Gaussian, their linear combination  $Z = rvX - Y$  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  $Z$ .

$$\begin{aligned}
 EZ &= EX - EY \\
 &= \mu_X - \mu_Y
 \end{aligned}$$

$$\begin{aligned}
 \text{var } Z &= EZ^2 - (EZ)^2 \\
 &= E(X - Y)^2 - (EX - EY)^2 \\
 &= E(X^2 - 2XY + Y^2) - [(EX)^2 - 2EXEY + (EY)^2] \\
 &= [EX^2 - (EX)^2] + [Y^2 - (EY)^2] - 2[EXY - EXEY] \\
 &= \text{var } X + \text{var } Y - 2\text{cov}[X, Y] \\
 &\triangleq \sigma_X^2 + \sigma_Y^2 - 2\sigma_{xy}
 \end{aligned}$$

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

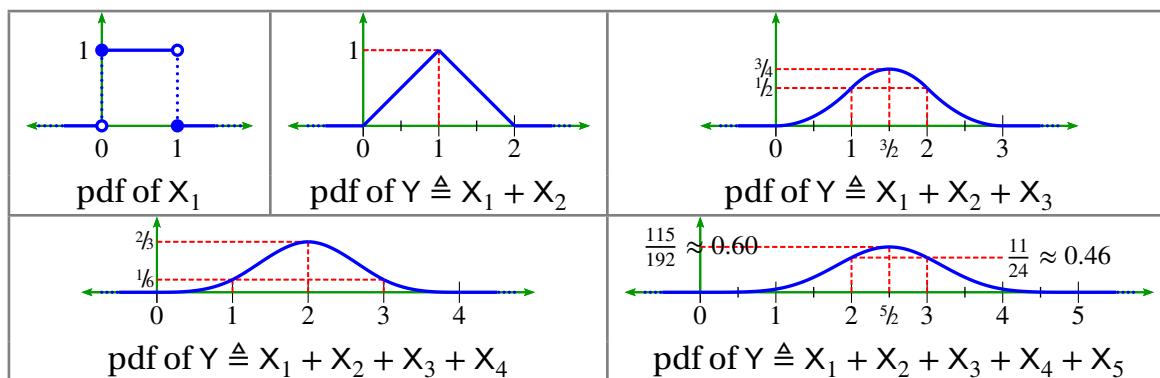


Figure 5.6: The distributions of sums of independent uniformly distributed random variables (Example 5.1 page 38)

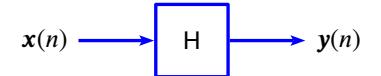
*Example 5.1.* Let  $(X_1, X_2, X_3, \dots)$  be a sequence of *independent* (Definition A.3 page 155) *uniformly distributed* random variables. Let  $p_N(x)$  be the *probability density function* of  $Y \triangleq \sum_{n=1}^N X_n$ . Some of these distributions are illustrated in Figure 5.6 (page 38). Note that the distributions of the sequence  $(p_1, p_2, p_3, \dots)$  are all *B-splines* (Definition S.2 page 371) and all form a *partition of unity*.

# CHAPTER 6

## OPERATORS ON DISCRETE RANDOM SEQUENCES

### 6.1 LTI operators on random sequences

**Theorem 6.1.** <sup>1</sup> Let  $x(n)$  be a RANDOM SEQUENCE with MEAN  $\mu_X$  and  $y(n)$  a RANDOM SEQUENCE with MEAN  $\mu_Y$ . Let  $S$  be a system with IMPULSE RESPONSE  $h(n)$ , INPUT  $x(n)$ , and OUTPUT  $y(n)$ .



**T H M**

$$\{ \text{S is (LTI)} \} \implies \left\{ \begin{array}{l} (1). \quad \mu_Y(n) = \sum_{k \in \mathbb{Z}} h(k) \mu_X(n-k) \triangleq h(n) \star \mu_X(n) \text{ and} \\ (2). \quad R_{xy}(n, m) = \sum_{k \in \mathbb{Z}} h^*(k) R_{xx}(n-k, m+k) \\ (3). \quad R_{yy}(n, m) = \sum_{k \in \mathbb{Z}} h^*(k) R_{yx}(n-k, m+k) \end{array} \right\}$$

PROOF:

$$\begin{aligned}
 \mu_Y(n) &\triangleq E[y(n)] && \text{by definition of } \mu_Y && (\text{Definition 2.2 page 11}) \\
 &= E\left[\sum_{k \in \mathbb{Z}} h(k)x(n-k)\right] && \text{by LTI hypothesis} \\
 &= \sum_{k \in \mathbb{Z}} h(k)E[x(n-k)] && \text{by linear property} \\
 &= \sum_{k \in \mathbb{Z}} h(k)\mu_X(n-k) && \text{by definition of } \mu_X && (\text{Definition 2.2 page 11}) \\
 &\triangleq h(n) \star \mu_X(n) && \text{by definition of convolution} && (\text{Definition R.3 page 357})
 \end{aligned}$$

$$\begin{aligned}
 R_{xy}(n, m) &\triangleq E[x(n+m)y^*(n)] && \text{by definition of } R_{xy}(n, m) && (\text{Definition 2.2 page 11}) \\
 &= E[x(n+m)(h(n) \star x(n))^*] && \text{by LTI hypothesis} \\
 &\triangleq E\left[x(n+m)\left(\sum_{k \in \mathbb{Z}} h(k)x(n-k)\right)^*\right] && \text{by definition of convolution } \star && (\text{Definition R.3 page 357}) \\
 &= E\left[x(n+m) \sum_{k \in \mathbb{Z}} h^*(k)x^*(n-k)\right] && \text{by distributive property of } *-\text{algebras} && (\text{Definition J.3 page 254})
 \end{aligned}$$

<sup>1</sup> Papoulis (1991), page 310

$$\begin{aligned}
 &= E \left[ \sum_{k \in \mathbb{Z}} h^*(k) x(n+m) x^*(n-k) \right] && \text{by } \textit{distributive} \text{ property of } (\mathbb{C}, +, \cdot, 0, 1) \quad (\text{Definition F.5 page 196}) \\
 &= \sum_{k \in \mathbb{Z}} h^*(k) E[x(n-k+k+m)x^*(n-k)] && \text{by } \textit{linear} \text{ property of } E \quad (\text{Theorem 1.1 page 4}) \\
 &\triangleq \sum_{k \in \mathbb{Z}} h^*(k) R_{xx}(n-k, m+k) && \text{by definition of } R_{xx}(n, m) \quad (\text{Definition 2.2 page 11})
 \end{aligned}$$

$$\begin{aligned}
 R_{yy}(n, m) &\triangleq E[y(n+m)y^*(n)] && \text{by definition of } R_{xy}(n, m) \quad (\text{Definition 2.2 page 11}) \\
 &= E[y(n+m)(h(n) \star x(n))^*] && \text{by LTI hypothesis} \\
 &\triangleq E \left[ y(n+m) \left( \sum_{k \in \mathbb{Z}} h(k) x(n-k) \right)^* \right] && \text{by definition of convolution} \quad (\text{Definition R.3 page 357}) \\
 &= E \left[ y(n+m) \sum_{k \in \mathbb{Z}} h^*(k) x^*(n-k) \right] && \text{by } \textit{distributive} \text{ property of } *-\text{algebras} \quad (\text{Definition J.3 page 254}) \\
 &= E \left[ \sum_{k \in \mathbb{Z}} h^*(k) y(n+m) x^*(n-k) \right] && \text{by } \textit{distributive} \text{ property of } (\mathbb{C}, +, \cdot, 0, 1) \quad (\text{Definition F.5 page 196}) \\
 &= \sum_{k \in \mathbb{Z}} h^*(k) E[y(n-k+k+m)x^*(n-k)] && \text{by } \textit{linear} \text{ property of } E \quad (\text{Theorem 1.1 page 4}) \\
 &\triangleq \sum_{k \in \mathbb{Z}} h^*(k) R_{yx}(n-k, m+k) && \text{by definition of } R_{xy}(n, m) \quad (\text{Definition 2.2 page 11})
 \end{aligned}$$

⇒

## 6.2 LTI operators on WSS random sequences

**Corollary 6.1.** Let  $S$  be the system defined in Theorem 6.1 (page 39).

COR

$$\left. \begin{array}{l} (A). \quad S \text{ is LTI} \\ (B). \quad x(n) \text{ is WSS} \end{array} \right\} \implies \left\{ \begin{array}{ll} (1). & \mu_Y = \mu_X \sum_{n \in \mathbb{Z}} h(n) \quad \text{and} \\ (2). & R_{xy}(m) = R_{xx}(m) \star h^*(-m) \triangleq \sum_{k \in \mathbb{Z}} h^*(k) R_{xx}(m+k) \quad \text{and} \\ (3). & R_{yy}(m) = R_{yx}(m) \star h^*(-m) \triangleq \sum_{k \in \mathbb{Z}} h^*(k) R_{xy}(m+k) \quad \text{and} \\ (4). & R_{yy}(m) = R_{xx}^*(m) \star h(-m) \star h^*(-m) \end{array} \right.$$

PROOF:

$$\begin{aligned}
 \mu_Y &= \mu_Y(n) && \text{by Proposition 2.1 page 14} && \text{and hypothesis (A)} \\
 &= \sum_{n \in \mathbb{Z}} h(n) \mu_X(n-k) && \text{by Theorem 2.1 page 12} && \text{and hypothesis (B)} \\
 &= \sum_{n \in \mathbb{Z}} h(n) \mu_X(0) && \text{by Definition 7.1 page 47} && \text{and hypothesis (B)} \\
 &= \mu_X(0) \sum_{n \in \mathbb{Z}} h(n) && \text{by linear property of } \sum && \\
 &= \mu_X \sum_{n \in \mathbb{Z}} h(n) && \text{by Proposition 2.1 page 14} &&
 \end{aligned}$$

<sup>1</sup>  Papoulis (1991), page 323



$$\begin{aligned}
R_{xy}(m) &\triangleq R_{xy}(0, m) && \text{by Proposition 2.1 page 14} && \text{and hypothesis (A)} \\
&= \sum_{k \in \mathbb{Z}} h^*(k) R_{xx}(0 - k, m + k) && \text{by Theorem 6.1 page 39} && \text{and hypothesis (B)} \\
&= \sum_{k \in \mathbb{Z}} h^*(k) R_{xx}(m + k) && \text{by Proposition 2.1 page 14} && \text{and hypothesis (A)} \\
&= h^*(-m) \star R_{xx}(m) && \text{by Proposition R.2 page 358} && \\
R_{yy}(m) &\triangleq R_{yy}(0, m) && \text{by Proposition 2.1 page 14} && \text{and hypothesis (A)} \\
&= \sum_{k \in \mathbb{Z}} h^*(k) R_{yx}(n - k, m + k) && \text{by Theorem 6.1 page 39} && \text{and hypothesis (B)} \\
&= \sum_{k \in \mathbb{Z}} h^*(k) R_{yx}(m + k) && \text{by Proposition 2.1 page 14} && \text{and hypothesis (A)} \\
&= h^*(-m) \star R_{yx}(m) && \text{by Proposition R.2 page 358} && \\
R_{yy}(m) &= h^*(-m) \star R_{yx}(m) && \text{by result (2)} && \\
&= h^*(-m) \star R_{xy}^*(m) && \text{by Corollary 2.1 page 14} && \\
&= h^*(-m) \star [h^*(-m) \star R_{xx}(m)]^* && \text{by result (1)} && \\
&= h^*(-m) \star h(-m) \star R_{xx}^*(m) && \text{by } \textit{distributive property of } *-\textbf{algebras} && \text{(Definition J.3 page 254)}
\end{aligned}$$

⇒

**Corollary 6.2.** <sup>2</sup> Let  $\mathbf{S}$  be a system with IMPULSE RESPONSE  $h(n)$ , INPUT  $x(n)$ , and OUTPUT  $y(n)$ .

COR	$\left\{ \begin{array}{l} (A). \quad h \text{ is LINEAR TIME INVARIANT and} \\ (B). \quad x \text{ and } y \text{ are WIDE SENSE STATIONARY} \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} (1). \quad \check{S}_{xy}(z) = \check{S}_{xx}(z)\check{H}^*\left(\frac{1}{z^*}\right) \text{ and} \\ (2). \quad \check{S}_{yy}(z) = \check{S}_{yx}(z)\check{H}^*\left(\frac{1}{z^*}\right) \text{ and} \\ (3). \quad \check{S}_{yy}(z) = \check{S}_{xx}(z)\check{H}(z)\check{H}^*\left(\frac{1}{z^*}\right) \end{array} \right\}$
-----	--

⇒

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

⇒

**Corollary 6.3.** Let  $\mathbf{S}$  be a system with IMPULSE RESPONSE  $h(n)$ , INPUT  $x(n)$ , and OUTPUT  $y(n)$ .

COR	$\left\{ \begin{array}{l} (A). \quad h \text{ is LTI and} \\ (B). \quad x \text{ and } y \text{ are WSS} \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} (1). \quad \tilde{S}_{xy}(\omega) = \tilde{S}_{xx}(\omega)\tilde{H}^*(\omega) \text{ and} \\ (2). \quad \tilde{S}_{yy}(\omega) = \tilde{S}_{xy}(\omega)\tilde{H}(\omega) \text{ and} \\ (3). \quad \tilde{S}_{yy}(\omega) = \tilde{S}_{xx}(\omega) \tilde{H}(\omega) ^2 \end{array} \right\}$
-----	--

⇒

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

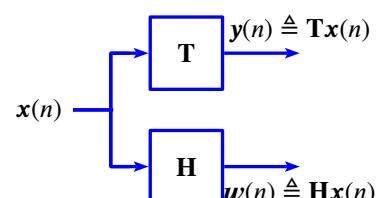
⇒

## 6.3 Parallel operators on WSS random sequences

**Theorem 6.2.** Let  $\mathbf{S}$  be the SYSTEM illustrated to the right, where  $\mathbf{T}$  is NOT NECESSARILY LINEAR. Let

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

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



<sup>2</sup> Papoulis (1991), page 323

THEOREM

$$\left\{ \begin{array}{l} \text{(A). } x(n) \text{ is WSS and} \\ \text{(B). } H \text{ is LTI} \end{array} \right\} \Rightarrow$$

$$\left\{ \begin{array}{ll} \text{(1). } R_{wy}(m) &= \sum_{n \in \mathbb{Z}} h(n)R_{xy}(m-n) \quad (\text{convolution}) \\ &\triangleq h(m) \star R_{xy}(m) \quad \text{and} \\ \text{(2). } \check{S}_{wy}(z) &= \check{H}(z)\check{S}_{xy}(z) \quad \text{and} \\ \text{(3). } \tilde{S}_{wy}(\omega) &= \tilde{H}(\omega)\tilde{S}_{xy}(\omega) \end{array} \right\}$$

PROOF:

$$\begin{aligned} R_{wy}(m) &\triangleq E[w(m)y^*(0)] && \text{by (A) and definition of } R_{wy} && (\text{Definition 2.4 page 12}) \\ &\triangleq E([Hx](m)y^*(0)) && \text{by definition of } S \\ &= HE(x(m)y^*(0)) && \text{by LTI hypothesis} && (\text{B}) \\ &\triangleq HR_{xy}(m) && \text{by definition of } R_{xy} && (\text{Definition 2.4 page 12}) \\ &= \sum_{n \in \mathbb{Z}} h(n)R_{xy}(m-n) && \text{by definition of } H \text{ impulse response } (h(n)) \\ &= [h(m) \star R_{xy}(m)] && \text{by definition of convolution} && (\text{Definition R.3 page 357}) \\ \check{S}_{wy}(z) &\triangleq ZR_{wy}(m) && \text{by definition of } \check{S}_{wy} && (\text{Definition 2.5 page 15}) \\ &= [h(m) \star R_{xy}(m)] && \text{by previous result} \\ &= \check{H}(z)\check{S}_{xy}(z) && \text{by Convolution Theorem} && (\text{Theorem R.2 page 360}) \\ \tilde{S}_{wy}(\omega) &\triangleq \check{F}R_{wy}(m) && \text{by definition of } \tilde{S}_{wy} && (\text{Definition 7.3 page 48}) \\ &= [h(m) \star R_{xy}(m)] && \text{by previous result} \\ &= \tilde{H}(\omega)\tilde{S}_{xy}(\omega) && \text{by Convolution Theorem} && (\text{Theorem R.2 page 360}) \end{aligned}$$



## 6.4 Whitening discrete random sequences

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

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

That is, if  $L(z)$  has a root inside the unit circle, then  $L^*(1/z^*)$  has a root directly opposite across the unit circle boundary (see Figure 6.1 page 43). A causal stable filter  $\check{H}(z)$  must have all of its poles inside the unit circle. A filter has *minimum phase* (Definition R.5 page 361) 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*  $x(n)$  with a filter constructed from a decomposition of  $R_{xx}(m)$ . The technique is stated precisely in Theorem 6.3 page 42 and illustrated in Figure 6.2 page 43. 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”  $x(n)$  from a white noise process  $w(n)$ ) and  $\gamma(n)$  is referred to as the **whitening filter** because it produces a white noise sequence when the input sequence is  $x(n)$ .<sup>3</sup>

**Theorem 6.3.** Let  $x(n)$  be a WSS RANDOM SEQUENCE with auto-correlation  $R_{xx}(m)$  and spectral density  $\check{S}_{xx}(z)$ . If  $\check{S}_{xx}(z)$  has a rational expression, then the following are true:

<sup>3</sup> Papoulis (1991), pages 401–402



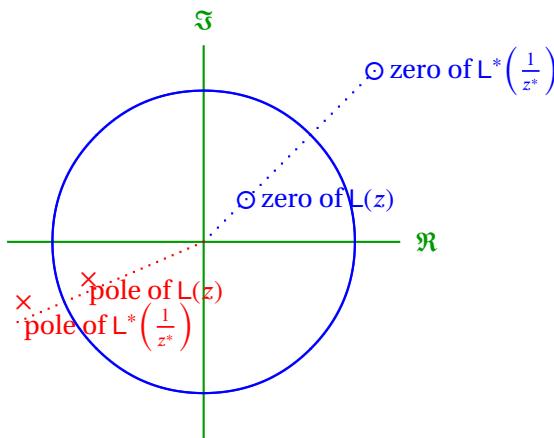


Figure 6.1: Mirrored roots in complex-z plane

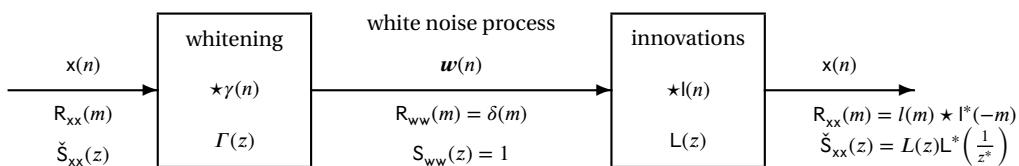


Figure 6.2: Innovations and whitening filters

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

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

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

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

is both causal and stable.

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

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

PROOF:

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





# CHAPTER 7

## OPERATORS ON CONTINUOUS RANDOM SEQUENCES

### 7.1 LTI Operations on non-stationary random processes

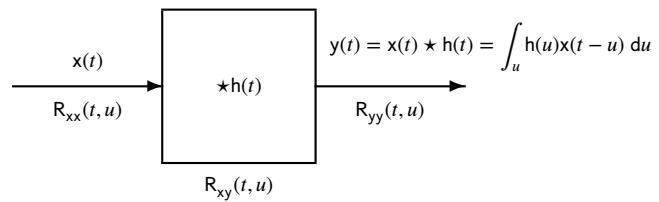


Figure 7.1: Linear system with random process input and output

**Theorem 7.1.** <sup>1</sup> Let  $h : \mathbb{R} \rightarrow \mathbb{C}$  be the impulse response of a linear time-invariant system and let  $y(t) = h(t) \star x(t) \triangleq \int_{u \in \mathbb{R}} h(u)x(t-u) du$  as illustrated in Figure 7.1 page 45. Then

#### Correlation functions

$$\begin{aligned} R_{xy}(t,u) &= R_{xx}(t,u) \star h^*(u) &\triangleq \int_v h^*(v)R_{xx}(t,u-v) dv \\ R_{yy}(t,u) &= R_{xy}(t,u) \star h(t) &\triangleq \int_v h(v)R_{xy}(t-v,u) dv \\ R_{yy}(t,u) &= R_{xx}(t,u) \star h(t) \star h^*(u) &\triangleq \int_w h^*(w) \int_v h(v)R_{xx}(t-v,u-w) dv dw \end{aligned}$$

#### Laplace power spectral density functions

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

#### Power spectral density functions

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

PROOF:

$$\begin{aligned}
 R_{xy}(t, u) &\triangleq E[x(t)y^*(u)] \\
 &= E\left[x(t)\left(\int_v h(v)x(u-v) dv\right)^*\right] \\
 &= E\left[x(t)\int_v h^*(v)x^*(u-v) dv\right] \\
 &= \int_v h^*(v)E[x(t)x^*(u-v)] dv \\
 &= \int_v h^*(v)R_{xx}(t, u-v) dv \\
 &\triangleq R_{xx}(t, u) \star h^*(u)
 \end{aligned}$$

$$\begin{aligned}
 R_{yy}(t, u) &\triangleq E[y(t)y^*(u)] \\
 &= E\left[\left(\int_v h(v)x(t-v) dv\right)y^*(u)\right] \\
 &= \int_v h(v)E[x(t-v)y^*(u)] dv \\
 &= \int_v h(v)R_{xy}(t-v, u) dv \\
 &\triangleq R_{xy}(t, u) \star h(t)
 \end{aligned}$$

$$\begin{aligned}
 R_{yy}(t, u) &\triangleq E[y(t)y^*(u)] \\
 &= E\left[\left(\int_v h(v)x(t-v) dv\right)\left(\int_w h(w)x(u-w) dw\right)^*\right] \\
 &= \int_w h^*(w) \int_v h(v)E[x(t-v)x^*(u-w)] dv dw \\
 &= \int_w h^*(w) \int_v h(v)R_{xx}(t-v, u-w) dv dw \\
 &= \int_w h^*(w)[R_{xx}(t, u-w) \star h(t)] dw \\
 &\triangleq R_{xx}(t, u) \star h(t) \star h^*(u)
 \end{aligned}$$

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

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

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

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

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



## 7.2 LTI Operations on WSS random processes

### Definition 7.1.

**D E F** A random process  $x(t)$  is **wide sense stationary (WSS)** if  
 (1).  $\mu_X(t)$  is CONSTANT with respect to  $t$  (STATIONARY IN THE MEAN) and  
 (2).  $R_{xx}(t + \tau, t)$  is CONSTANT with respect to  $t$  (STATIONARY IN CORRELATION)

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

**Definition 7.2.** Let  $x(t)$  and  $y(t)$  be WSS random processes.

**D E F**

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

**Definition 7.3.** Let  $x(t)$  and  $y(t)$  be WSS random processes.

DEF

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

**Definition 7.4.** <sup>2</sup> Let  $x(t)$  be a random variable that is STATIONARY IN THE MEAN such that  $E[x(t)]$  is constant with respect to  $t$ .

DEF

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

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

**Proposition 7.1.**

PRP

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

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

**Definition 7.5.** <sup>3</sup> Let  $x(t)$  be a WIDE SENSE STATIONARY random process.

DEF

- (1). The average power  $P_{\text{avg}}[x(t)]$  of  $x(t)$  is  $P_{\text{avg}}[x(t)] \triangleq \lim_{\tau \rightarrow \infty} \frac{1}{2\tau} \int_{t \in \mathbb{R}} |x(t)|^2 dt$
- (2). The energy spectral density  $|\tilde{x}(\omega)|^2$  of  $x(t)$  is  $|\tilde{x}(\omega)|^2 \triangleq |\tilde{\mathbf{F}}x(t)|^2$

**Remark 7.1 (spectral power).** Why does  $\tilde{S}_{xx}(\omega)$  deserve the name *power spectral density*? This is answered by Theorem 7.2 (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.

**Theorem 7.2.** Let  $x(t)$  be a RANDOM PROCESS.

THM

$$\left\{ \begin{array}{l} (A). \quad x(t) \text{ is ERGODIC IN THE MEAN} \quad \text{and} \\ (B). \quad \tilde{x}(\omega) \text{ EXISTS} \end{array} \right\} \implies \left\{ \begin{array}{l} (1). \quad \tilde{S}_{xx}(\omega) = \lim_{\tau \rightarrow \infty} \frac{1}{2\tau} \underbrace{\int_{-\tau}^{+\tau} x(t) dt}_{(\text{ESD})} \quad \text{and} \\ (2). \quad P_{\text{avg}}[x(t)] = \int_{\omega \in \mathbb{R}} \tilde{S}_{xx}(\omega) d\omega \end{array} \right\}$$

<sup>2</sup> 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

<sup>3</sup> Bendat and Piersol (2010) page 177

PROOF:

$$\begin{aligned}
 \tilde{S}_{xx}(\omega) &\triangleq \int_{\tau \in \mathbb{R}} R_{xx}(\tau) e^{-i\omega\tau} d\tau && \text{by definition of } \tilde{S}_{xx}(\omega) && (\text{Definition 7.3 page 48}) \\
 &= \int_{\tau \in \mathbb{R}} E[x(t + \tau)x^*(t)] e^{-i\omega\tau} d\tau && \text{by definition of } R_{xx}(t) && (\text{Definition 3.2 page 20}) \\
 &= E\left[x^*(t) \int_{\tau \in \mathbb{R}} x(t + \tau)e^{-i\omega\tau} d\tau\right] && \text{by linearity of } E \text{ operator} \\
 &= E\left[x^*(t) \int_{u \in \mathbb{R}} x(u)e^{-i\omega(u-t)} du\right] && \text{where } u \triangleq t + \tau \implies \tau = u - t \\
 &= E\left[x^*(t)e^{i\omega t} \int_{u \in \mathbb{R}} x(u)e^{-i\omega u} du\right] && \\
 &= E[x^*(t)e^{i\omega t}\tilde{x}(\omega)] && \text{by definition of Fourier Transform} && (\text{Definition P.2 page 337}) \\
 &= E[x^*(t)e^{i\omega t}\tilde{x}(\omega)] && \text{by hypothesis (B)} \\
 &= \left[ \lim_{\tau \rightarrow \infty} \frac{1}{2\tau} \int_{-\tau}^{+\tau} x^*(t)e^{i\omega t} dt \right] \tilde{x}(\omega) && \text{by ergodic in the mean hypothesis} && (\text{Definition 7.4 page 48}) \\
 &= \lim_{\tau \rightarrow \infty} \frac{1}{2\tau} \left[ \int_{t \in \mathbb{R}} x(t)e^{-i\omega t} dt \right]^* \tilde{x}(\omega) \\
 &= \lim_{\tau \rightarrow \infty} \frac{1}{2\tau} \tilde{x}^*(\omega)\tilde{x}(\omega) && \text{by hypothesis (B)} \\
 &= \lim_{\tau \rightarrow \infty} \frac{1}{2\tau} |\tilde{x}(\omega)|^2
 \end{aligned}$$

$$\begin{aligned}
 \int_{\omega \in \mathbb{R}} \tilde{S}_{xx}(\omega) d\omega &= \int_{\omega \in \mathbb{R}} \lim_{\tau \rightarrow \infty} \frac{1}{2\tau} |\tilde{x}(\omega)|^2 d\omega \\
 &= \lim_{\tau \rightarrow \infty} \frac{1}{2\tau} \int_{\omega \in \mathbb{R}} |\tilde{x}(\omega)|^2 d\omega \\
 &= \lim_{\tau \rightarrow \infty} \frac{1}{2\tau} \int_{t \in \mathbb{R}} |x(t)|^2 dt && \text{by Plancheral's formula} && (\text{Theorem P.3 page 339, Theorem H.9 page 222}) \\
 &= P_{avg} && \text{by definition of } P_{avg} && (\text{Definition 7.5 page 48})
 \end{aligned}$$

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

⇒

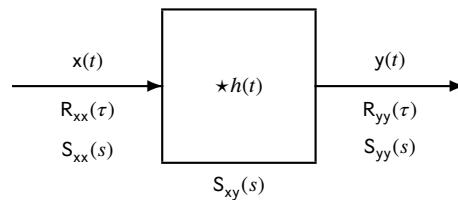


Figure 7.2: Linear system with WSS random process input and output

**Theorem 7.3.** <sup>4</sup> Let  $h : \mathbb{R} \rightarrow \mathbb{C}$  be the impulse response of a linear time-invariant system and let  $y(t) = h(t) \star x(t) \triangleq \int_{u \in \mathbb{R}} h(u)x(t-u) du$  as illustrated in Figure 7.1 page 45. Then

<sup>4</sup> Papoulis (1991) pages 323–324

<b>T H M</b>	$\begin{aligned} R_{xy}(\tau) &= R_{xx}(\tau) \star h^*(-\tau) &\triangleq \int_{u \in \mathbb{R}} h^*(-u) R_{xx}(\tau - u) du \\ R_{yy}(\tau) &= R_{xy}(\tau) \star h(\tau) &\triangleq \int_{u \in \mathbb{R}} h(u) R_{xy}(\tau - u) du \\ R_{yy}(\tau) &= R_{xx}(\tau) \star h(\tau) \star h^*(-\tau) &\triangleq \int_v \int_{u \in \mathbb{R}} h(u - v) h^*(-v) R_{xx}(\tau - u - v) du dv \end{aligned}$ $\begin{aligned} S_{xy}(s) &= S_{xx}(s) \hat{h}^*(-s^*) \\ S_{yy}(s) &= S_{xy}(s) \hat{h}(s) \\ S_{yy}(s) &= S_{xx}(s) \hat{h}(s) \hat{h}^*(-s^*) \end{aligned}$ $\begin{aligned} \tilde{S}_{xy}(\omega) &= \tilde{S}_{xx}(\omega) \tilde{h}^*(\omega) \\ \tilde{S}_{yy}(\omega) &= \tilde{S}_{xy}(\omega) \tilde{h}(\omega) \\ \tilde{S}_{yy}(\omega) &= \tilde{S}_{xx}(\omega)  \tilde{h}(\omega) ^2 \end{aligned}$
----------------------	--

PROOF:

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

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

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

$$\begin{aligned} S_{xy}(s) &\triangleq LR_{xy}(\tau) \\ &\triangleq \int_{\tau \in \mathbb{R}} R_{xy}(\tau) e^{-s\tau} d\tau \\ &= \int_{\tau \in \mathbb{R}} [R_{xx}(\tau) \star h^*(-\tau)] e^{-s\tau} d\tau \\ &= \int_{\tau \in \mathbb{R}} \left[ \int_{u \in \mathbb{R}} h^*(-u) R_{xx}(\tau - u) du \right] e^{-s\tau} d\tau \\ &= \int_{u \in \mathbb{R}} h^*(-u) \int_{\tau \in \mathbb{R}} R_{xx}(\tau - u) e^{-s\tau} d\tau du \end{aligned}$$



$$\begin{aligned}
&= \int_{u \in \mathbb{R}} h^*(-u) \int_v R_{xx}(v) e^{-s(v+u)} dv du \\
&= \int_{u \in \mathbb{R}} h^*(-u) e^{-su} du \int_v R_{xx}(v) e^{-sv} dv \\
&= \int_{u \in \mathbb{R}} h^*(u) e^{-s(-u)} du \int_v R_{xx}(v) e^{-sv} dv \\
&= \left( \int_{u \in \mathbb{R}} h(u) e^{-(s^*)u} du \right)^* \int_v R_{xx}(v) e^{-sv} dv \\
&\triangleq \hat{h}^*(-s^*) S_{xx}(s)
\end{aligned}$$

where  $v = \tau - u \iff \tau = v + u$

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

where  $v = \tau - u \iff \tau = v + u$

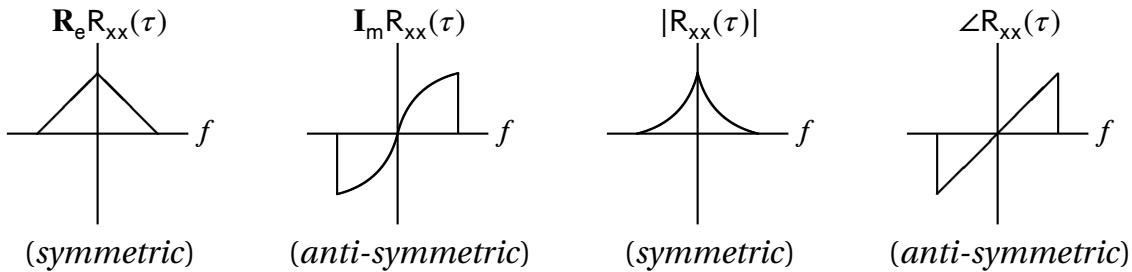
$$\begin{aligned}
S_{yy}(s) &= \hat{h}(s) S_{xy}(s) \\
&= \hat{h}(s) \hat{h}^*(-s^*) S_{xx}(s)
\end{aligned}$$

$$\begin{aligned}
\tilde{S}_{xy}(\omega) &= S_{xy}(s) \Big|_{s=j\omega} \\
&= \hat{h}^*(-s^*) S_{xx}(s) \Big|_{s=j\omega} \\
&= \left( \int_{u \in \mathbb{R}} h(u) e^{-(s^*)u} du \right)^* \int_v R_{xx}(v) e^{-sv} dv \Big|_{s=j\omega} \\
&= \left( \int_{u \in \mathbb{R}} h(u) e^{(-j\omega)^*u} du \right)^* \int_v R_{xx}(v) e^{-j\omega v} dv \\
&= \left( \int_{u \in \mathbb{R}} h(u) e^{-j\omega u} du \right)^* \int_v R_{xx}(v) e^{-j\omega v} dv \\
&\triangleq \tilde{h}^*(\omega) \tilde{S}_{xx}(\omega)
\end{aligned}$$

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

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

⇒

Figure 7.3: auto-correlation  $R_{xx}(\tau)$ 

**Theorem 7.4.** Let  $x : \mathbb{R} \rightarrow \mathbb{C}$  be a WSS random process with auto-correlation  $R_{xx}(\tau)$ . Then  $R_{xx}(\tau)$  is conjugate symmetric such that (see Figure 7.3 page 52)

T H M	$R_{xx}(\tau) = R_{xx}^*(-\tau)$ (CONJUGATE SYMMETRIC) $R_e [R_{xx}(\tau)] = R_e [R_{xx}^*(-\tau)]$ (SYMMETRIC) $I_m [R_{xx}(\tau)] = -I_m [R_{xx}^*(-\tau)]$ (ANTI-SYMMETRIC) $ R_{xx}(\tau)  =  R_{xx}^*(-\tau) $ (SYMMETRIC) $\angle R_{xx}(\tau) = \angle R_{xx}^*(-\tau)$ (ANTI-SYMMETRIC).
-------------	--

PROOF:

$$\begin{aligned}R_{xx}^*(\tau) &\triangleq (E[x(t-\tau)x^*(t)])^* \\ &= E[x^*(t-\tau)x(t)] \\ &= E[x(t)x^*(t-\tau)] \\ &= E[x(u+\tau)x^*(u)] \\ &\triangleq R_{xx}(\tau) \quad \text{where } u \triangleq t-\tau \iff t=u+\tau \\ R_e [R_{xx}(\tau)] &= R_e [R_{xx}^*(-\tau)] \\ I_m [R_{xx}(\tau)] &= I_m [R_{xx}^*(-\tau)] \\ abs R_{xx}(\tau) &= |R_{xx}^*(-\tau)| \\ \angle R_{xx}(\tau) &= \angle R_{xx}^*(-\tau) \\ &= R_e [R_{xx}(-\tau)] \\ &= -I_m [R_{xx}(-\tau)] \\ &= |R_{xx}(-\tau)| \\ &= -\angle R_{xx}(-\tau)\end{aligned}$$

⇒

## 7.3 Whitening continuous random sequences

Simple algebraic operations on white noise processes (processes with autocorrelation  $R_{xx}(\tau) = \delta(\tau)$ ) often produce *colored* noise (processes with autocorrelation  $R_{xx}(\tau) \neq \delta(\tau)$ ). 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–Loëve expansion can be used to decompose colored noise.<sup>5</sup>

**Definition 7.6.** A **rational expression**  $p(s)$  is a polynomial divided by a polynomial such that

**DEF**

$$p(s) = \frac{\sum_{n=0}^N b_n s^n}{\sum_{n=0}^M a_n s^n}$$

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 7.7.** Let  $\check{h}(s)$  be the Laplace transform of the impulse response of a filter. If  $\check{h}(s)$  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 7.4 page 53). 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 reciprocal (zeros become poles and poles become zeros) is also causal and stable.

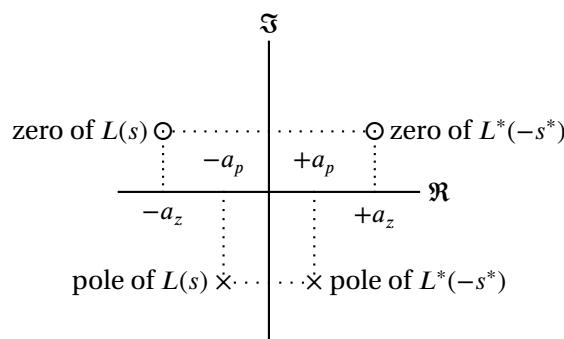


Figure 7.4: Mirrored roots in complex-s plane

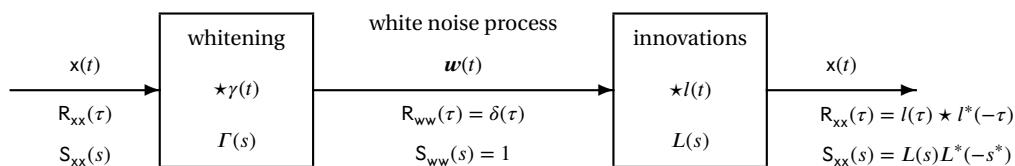


Figure 7.5: Innovations and whitening filters

The next theorem demonstrates a method for “whitening” a random process  $x(t)$  with a filter constructed from a decomposition of  $R_{xx}(\tau)$ . The technique is stated precisely in Theorem 7.5 page 54

<sup>5</sup> Karhunen–Loëve expansion: Section 4.1 page 21

and illustrated in Figure 7.5 page 53. 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”  $x(t)$  from a white noise process  $w(t)$ ) and  $\gamma(t)$  is referred to as the **whitening filter** because it produces a white noise sequence when the input sequence is  $x(t)$ .<sup>6</sup>

**Theorem 7.5.** Let  $x(t)$  be a WSS random process with autocorrelation  $R_{xx}(\tau)$  and spectral density  $S_{xx}(s)$ . If  $S_{xx}(s)$  has a **rational expression**, then the following are true:

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

$$S_{xx}(s) = L(s)L^*(-s^*).$$

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  $x(t)$  is the input to the filter  $\gamma(t)$ , the output  $y(t)$  is a **white noise sequence** such that

$$S_{yy}(s) = 1 \quad R_{yy}(\tau) = \delta(\tau).$$

PROOF:

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

⇒

<sup>6</sup>  Papoulis (1991) pages 401–402

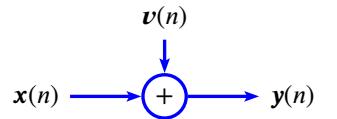


# CHAPTER 8

## ADDITIVE NOISE ON RANDOM SEQUENCES

### 8.1 Additive noise and correlation

**Theorem 8.1.** Let  $S$  be the system illustrated to the right, where  $T$  is NOT NECESSARILY LINEAR.



T H M	(A). $x(n)$ is WSS and (B). $x(n)$ and $v(n)$ are uncorrelated and (C). $v(n)$ is zero-mean and	$\left\{ \begin{array}{l} (1). R_{yy}(m) = R_{vv}(m) \\ (2). R_{xy}(m) = R_{xx}(m) \\ (3). R_{yy}(m) = R_{xx}(m) + R_{vv}(m) \\ (4). R_{xx}(m) = R_{yy}(m) + R_{vv}(m) - 2R_e R_{yy}(m) \end{array} \right.$
-------------	---	--

PROOF:

$R_{yy}(m) \triangleq E[y(m)y^*(0)]$ $\triangleq E[(x(m) + v(m))v^*(0)]$ $= E[x(m)v^*(0)] + E[v(m)v^*(0)]$ $= Ex(m)Ev^*(0) + E[v(m)v^*(0)]$ $= Ex(m)Ev^*(0) + E[v(m)v^*(0)]$ $= R_{vv}(m)$	by (A) and definition of $R_{yy}$ by definition of $y$ by linearity of $E$ by <i>uncorrelated</i> hypothesis by <i>zero-mean</i> hypothesis by definition of $R_{vv}$	(Definition 2.4 page 12) (Theorem 1.1 page 4) (B) (C) (Definition 2.4 page 12) (Definition 2.4 page 12)
$R_{xy}(m) \triangleq E[x(m)y^*(0)]$ $\triangleq E(x(m)[x(0) + v(0)]^*)$ $= E[x(m)x^*(0)] + E[x(m)v^*(0)]$ $= E[x(m)x^*(0)] + E[x(m)]E[v^*(0)]$ $= E[x(m)x^*(0)] + E[x(m)]E[v^*(0)]$ $= R_{xx}(m)$	by (A) and definition of $R_{xy}$ by definition of $y$ by linearity of $E$ by <i>uncorrelated</i> hypothesis by <i>zero-mean</i> hypothesis by definition of $R_{xx}$	(Definition 2.4 page 12) (Theorem 1.1 page 4) (B) (C) (Definition 2.4 page 12)
$R_{yy}(m) \triangleq E[y(m)y^*(0)]$ $\triangleq E[(x(m) + v(m))(x(0) + v(0))^*]$ $= E[x(m)x^*(0)] + E[x(m)v^*(0)] + E[v(m)x^*(0)] + E[v(m)v^*(0)]$ $= E[x(m)x^*(0)] + Ex(m)Ev^*(0) + Ev(m)Ex^*(0) + E[v(m)v^*(0)]$ $= E[x(m)x^*(0)] + Ex(m)Ev^*(0) + Ev(m)Ex^*(0) + E[v(m)v^*(0)]$	by (A) and definition of $R_{yy}$ by definition of $y$ by <i>uncorrelated</i> hypothesis (B) by <i>zero-mean</i> hypothesis (C)	

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

⇒

*Remark 8.1.* Because in Theorem 8.1  $y = x + v$  and  $R_{yy} = R_{xx} + R_{vv}$ , one might assume that  $R$  is a kind of *linear operator* (Definition O.3 page 308) and further assume that because  $x = y - v$  and  $R_{(-v)(-v)} = R_{vv}$ , that  $R_{xx} = R_{yy} + R_{vv}$ . As Theorem 8.1 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 8.1.** Let  $S$  be the system illustrated in Theorem 8.1 (page 55).

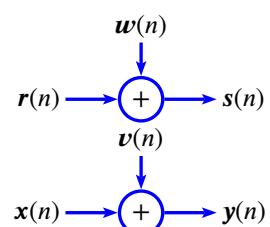
COR

$$\left. \begin{array}{l} \text{hypotheses of} \\ \text{Theorem 8.1 (page 55)} \end{array} \right\} \Rightarrow \left\{ \begin{array}{ll} (1). \quad \check{S}_{yy}(z) &= \check{S}_{xx}(z) + \check{S}_{vv}(z) & \text{and} \\ (2). \quad \check{S}_{yv}(z) &= \check{S}_{vv}(z) & \text{and} \\ (3). \quad \check{S}_{yv}(z) &= \check{S}_{yy}(z) + \check{S}_{vv}(z) + \check{S}_{yv}(z) + \check{S}_{yv}^*(z^*) & \text{and} \\ (4). \quad \tilde{S}_{yy}(\omega) &= \tilde{S}_{xx}(\omega) + \tilde{S}_{vv}(\omega) & \text{and} \\ (5). \quad \tilde{S}_{yv}(\omega) &= \tilde{S}_{vv}(\omega) & \text{and} \\ (6). \quad \tilde{S}_{yv}(\omega) &= \tilde{S}_{yy}(\omega) + \tilde{S}_{vv}(\omega) + \tilde{S}_{yv}(\omega) + \tilde{S}_{yv}^*(-\omega) & \end{array} \right.$$

PROOF:

$$\begin{aligned}
 \check{S}_{yy}(z) &\triangleq ZR_{yy}(m) && \text{by definition of } \check{S}_{xx} && (\text{Definition 2.5 page 15}) \\
 &= ZR_{qq}(m) + ZR_{vv}(m) && \text{by previous result} && (1) \\
 &= \check{S}_{qq}(z) + \check{S}_{vv}(z) && \text{by definition of } \check{S}_{yy} && (\text{Definition 7.3 page 48}) \\
 \tilde{S}_{yy}(\omega) &\triangleq \check{F}R_{yy}(m) && \text{by definition of } \check{S}_{yy} && (\text{Definition 7.3 page 48}) \\
 &= \check{F}R_{qq}(m) + \check{F}R_{vv}(m) && \text{by previous result} && (1) \\
 &= \tilde{S}_{qq}(\omega) + \tilde{S}_{vv}(\omega) && \text{by definition of } \check{S}_{yy} && (\text{Definition 7.3 page 48})
 \end{aligned}$$

⇒



**Theorem 8.2.** Let  $S$  be the system illustrated to the right:

THM

$$\left\{ \begin{array}{ll} \text{(A). } x(n) \text{ and } r(n) \text{ are wide sense stationary} & \text{and} \\ \text{(B). } E[x(n)w(n)] = Ex(n)Ew(n) \text{ (uncorrelated)} & \text{and} \\ \text{(C). } E[r(n)v(n)] = Er(n)Ev(n) \text{ (uncorrelated)} & \text{and} \\ \text{(D). } E[w(n)v(n)] = Ew(n)Ev(n) \text{ (uncorrelated)} & \text{and} \\ \text{(E). } Ev(n) = Ew(n) = 0 & \text{(zero-mean)} \end{array} \right\} \Rightarrow \left\{ \begin{array}{ll} R_{sy}(m) &= R_{sx}(m) \\ &= R_{ry}(m) \\ &= R_{rx}(m) \end{array} \right\}$$

PROOF:

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



**Corollary 8.2.** Let  $S$  be the system illustrated in Theorem 8.2 (page 56).

COR	$\left\{ \begin{array}{l} \text{hypotheses of} \\ \text{Theorem 8.2 (page 56)} \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} (1). \check{S}_{sy}(z) = \check{S}_{sx}(z) = \check{S}_{ry}(z) = \check{S}_{rx}(z) \text{ and} \\ (2). \tilde{S}_{sy}(\omega) = \tilde{S}_{sx}(\omega) = \tilde{S}_{ry}(\omega) = \tilde{S}_{rx}(\omega) \end{array} \right\}$
-----	---

PROOF:

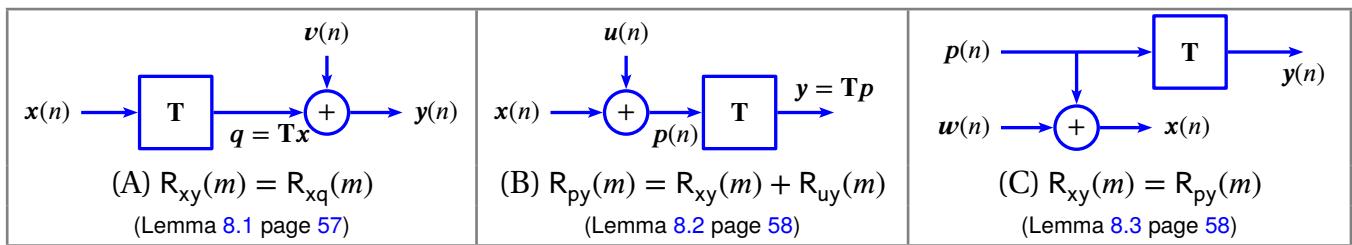
$$\begin{aligned}
 \check{S}_{sy}(\omega) &\triangleq ZR_{sy}(m) && \text{by definition of } \check{S}_{xy} && (\text{Definition 2.5 page 15}) \\
 &= ZR_{rx}(m) && \text{by previous result} && (1) \\
 &= \check{S}_{rx}(z) && \text{by definition of } \check{S}_{xy} && (\text{Definition 2.5 page 15}) \\
 \tilde{S}_{sy}(\omega) &\triangleq \check{F}R_{sy}(m) && \text{by definition of } \check{S}_{xy} && (\text{Definition 7.3 page 48}) \\
 &= \check{F}R_{rx}(m) && \text{by previous result} && (1) \\
 &= \tilde{S}_{rx}(\omega) && \text{by definition of } \check{S}_{xy} && (\text{Definition 7.3 page 48})
 \end{aligned}$$



## 8.2 Additive noise and operators

**Lemma 8.1.** Let  $S$  be the system illustrated in Figure 8.2 (page 59) (A).

LEM	$\left\{ \begin{array}{l} (A). R_{xx}(n_1, m) = R_{xx}(n_2, m) \text{ (WSS)} \\ (B). E[x(n)v(n)] = Ex(n)Ev(n) \text{ (UNCORRELATED)} \\ (E). Ev(n) = 0 \text{ (ZERO-MEAN)} \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} (1). R_{xy}(m) = R_{xq}(m) \text{ and} \\ (2). \check{S}_{xy}(z) = \check{S}_{xq}(z) \text{ and} \\ (3). \tilde{S}_{xy}(\omega) = \tilde{S}_{xq}(\omega) \end{array} \right\}$
-----	---

Figure 8.1: Additive noise with *linear/non-linear* operator **T**

PROOF:

$$\begin{aligned}
 R_{xy}(m) &\triangleq E[x(m)y^*(0)] && \text{by definition of } R_{xy} && (\text{Definition 2.4 page 12}) \\
 &\triangleq E[x(m)(q(0) + v(0))^*] && \text{by definition of } S && (\text{Figure 8.2 page 59}) (A) \\
 &= E[x(m)q^*(0) + p(m)v^*(0)] \\
 &= E[x(m)q^*(0)] + E[x(m)v^*(0)] \\
 &= E[x(m)q^*(0)] + [E x(m)][E v^*(0)] \\
 &= E[x(m)q^*(0)] + [E p(m)][E v^*(0)]^{\cancel{0}} && \text{by uncorrelated hypothesis} && (B) \\
 &= E[x(m)q^*(0)] && \text{by zero-mean hypothesis} && (E) \\
 &= R_{xq}(m) && \text{by definition of } R_{xq} && (\text{Definition 2.4 page 12}) \\
 \check{S}_{xy}(z) &\triangleq ZR_{xy}(m) && \text{by definition of } \check{S}_{xy} && (\text{Definition 2.5 page 15}) \\
 &= ZR_{xq}(m) && \text{by previous result} && (1) \\
 &= \check{S}_{xq}(z) && \text{by definition of } \check{S}_{xq} && (\text{Definition 2.5 page 15}) \\
 \check{S}_{xy}(\omega) &\triangleq \check{F}R_{xy}(m) && \text{by definition of } \check{S}_{xy} && (\text{Definition 7.3 page 48}) \\
 &= \check{F}R_{xq}(m) && \text{by previous result} && (1) \\
 &= \check{S}_{xq}(\omega) && \text{by definition of } \check{S}_{xq} && (\text{Definition 7.3 page 48})
 \end{aligned}$$

⇒

**Lemma 8.2.** Let **S** be the system illustrated in Figure 8.2 (page 59) (B).

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

PROOF:

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

⇒

**Lemma 8.3.** Let **S** be the system illustrated in Figure 8.2 (page 59) (C).

<b>L E M</b>	$\left\{ \begin{array}{ll} (A). & x(n) \text{ is (WSS)} \\ (B). & u(n) \text{ is (ZERO-MEAN)} \\ (C). & x(n) \text{ and } u(n) \text{ are (UNCORRELATED)} \end{array} \right. \text{ and } \right\} \implies \left\{ \begin{array}{ll} (1). & R_{xy}(m) = R_{py}(m) \text{ and} \\ (2). & \check{S}_{xy}(z) = \check{S}_{py}(z) \text{ and} \\ (3). & \check{S}_{xy}(\omega) = \check{S}_{py}(\omega) \end{array} \right\}$
----------------------	---



PROOF:

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

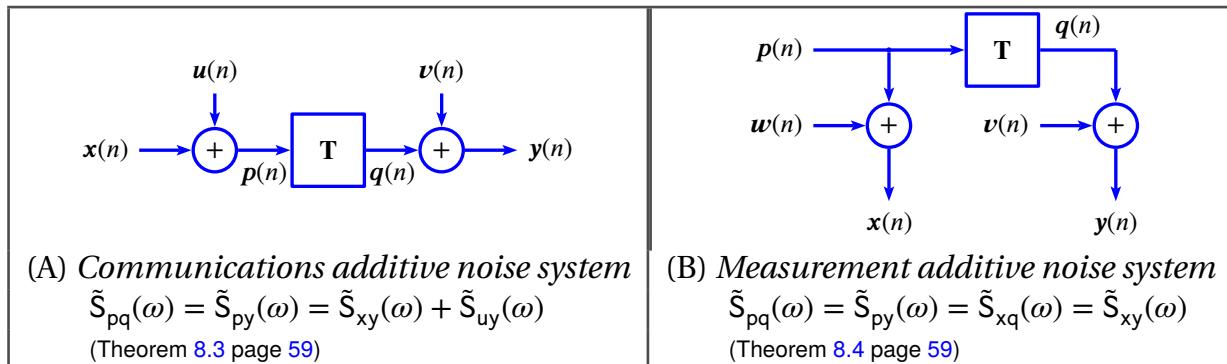


Figure 8.2: linear / non-linear additive noise systems

**Theorem 8.3 (communications additive noise cross-correlation).**

Let  $S$  be the system illustrated in Figure 8.2 page 59 (A).

<b>T H M</b>	(A). $x(n)$ is WSS (B). $u(n)$ is ZERO-MEAN (C). $v(n)$ is ZERO-MEAN (D). $x(n), u(n), v(n)$ are UNCORRELATED	and and and and	$\left\{ \begin{array}{l} (1). R_{pq}(m) = R_{py}(m) = R_{xy}(m) + R_{uy}(m) \text{ and} \\ (2). \tilde{S}_{pq}(z) = \tilde{S}_{py}(z) = \tilde{S}_{xy}(z) + \tilde{S}_{uy}(z) \text{ and} \\ (3). \tilde{S}_{pq}(\omega) = \tilde{S}_{py}(\omega) = \tilde{S}_{xy}(\omega) + \tilde{S}_{uy}(\omega) \end{array} \right.$
--------------	--	--------------------------	---

PROOF:

$$\begin{aligned}
 R_{pq}(m) &= R_{py}(m) && \text{by Lemma 8.1 page 57} \\
 R_{pq}(m) &= R_{xq}(m) + R_{uq}(m) && \text{by Lemma 8.2 page 58} \\
 R_{py}(m) &\triangleq E[p(m)y^*(0)] && \text{by definition } R_{py} && (\text{Definition 2.4 page 12}) \\
 &\triangleq E[(x(m) + u(m))y^*(0)] && \text{by definition } S && (\text{Figure 8.2 page 59}) (A) \\
 &= E[x(m)y^*(0) + u(m)y^*(0)] && && \\
 &= E[x(m)y^*(0)] + E[u(m)y^*(0)] && \text{by linearity of } E && (\text{Theorem 1.1 page 4}) \\
 &= R_{xy}(m) + R_{uy}(m) && \text{by definitions } R_{xy} \text{ and } R_{uy} && (\text{Definition 2.4 page 12})
 \end{aligned}$$



**Theorem 8.4 (measurement additive noise cross-correlation).**

Let  $S$  be the system illustrated in Figure 8.2 page 59 (B).

<b>T H M</b>	(A). $x(n)$ is WSS (B). $u(n)$ is ZERO-MEAN (C). $v(n)$ is ZERO-MEAN (D). $x(n), u(n), v(n)$ are UNCORRELATED	and and and and	$\left\{ \begin{array}{l} (1). R_{pq}(m) = R_{py}(m) = R_{xq}(m) = R_{xy}(m) \text{ and} \\ (2). \tilde{S}_{pq}(z) = \tilde{S}_{py}(z) = \tilde{S}_{xq}(z) = \tilde{S}_{xy}(z) \text{ and} \\ (3). \tilde{S}_{pq}(\omega) = \tilde{S}_{py}(\omega) = \tilde{S}_{xq}(\omega) = \tilde{S}_{xy}(\omega) \end{array} \right.$
--------------	--	--------------------------	---

PROOF:

$$\begin{aligned}
 R_{pq}(m) &= R_{py}(m) && \text{by Lemma 8.1 page 57} \\
 R_{pq}(m) &= R_{xq}(m) && \text{by Lemma 8.3 page 58} \\
 R_{xy}(m) &\triangleq E[x(m)y^*(0)] && \text{by definition } R_{xy} \\
 &\triangleq E([p(m) + u(m)]y^*(0)) && \text{(Definition 2.4 page 12)} \\
 &= E[p(m)y^*(0) + u(m)y^*(0)] && \text{by definition S} \\
 &= E[p(m)y^*(0)] + E[u(m)y^*(0)] && \text{(Figure 8.2 page 59) (B)} \\
 &= E[p(m)y^*(0)] + E[\cancel{u(m)y^*(0)}] && \text{by linearity of E} \\
 &= E[p(m)y^*(0)] && \text{(Theorem 1.1 page 4)} \\
 &= R_{py}(m) && \text{by uncorrelated hypothesis (D)} \\
 & && \text{by definition of } R_{py} \\
 & && \text{(Definition 2.4 page 12)}
 \end{aligned}$$

⇒

## 8.3 Additive noise and LTI operators

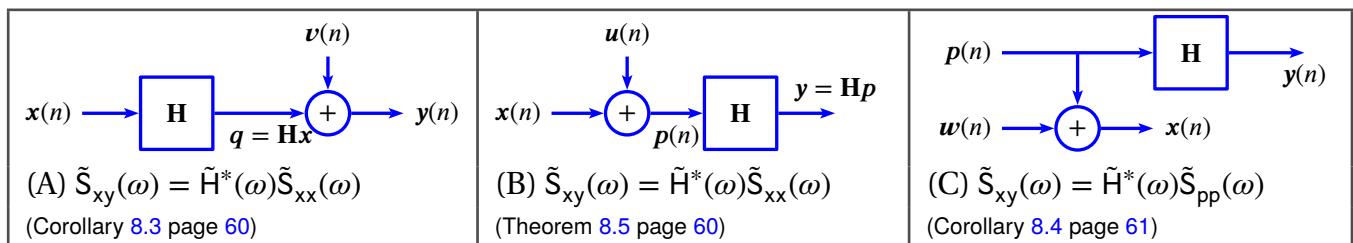


Figure 8.3: Additive noise with LTI operator  $\mathbf{H}$

**Corollary 8.3.** Let  $\mathbf{S}$  be the system illustrated in Figure 8.3 (page 60) (A).

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

PROOF:

$$\begin{aligned}
 \tilde{S}_{xy}(\omega) &= \tilde{S}_{xq}(\omega) && \text{by Lemma 8.1 page 57} \\
 &= \tilde{H}^*(\omega)\tilde{S}_{xx}(\omega) && \text{by Corollary 6.3 page 41}
 \end{aligned}$$

⇒

**Theorem 8.5.** Let  $\mathbf{S}$  be the system illustrated in Figure 8.3 (page 60) (B).

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

PROOF:

- definition: Let  $(h(n))$  be the *impulse response* of operator  $\mathbf{H}$  such that  

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



2. lemma:  $\mathbf{Hx}(n) = \sum_{m \in \mathbb{Z}} h(n)x(m-n) = h(n) \star R_{xx}(n)$ .

Proof: by the *linear time-invariant* hypotheses (D) and definition of *convolution operator*  $\star$  (Definition P.3 page 340)

3. Proof that  $R_{yx}(m) = h(m) \star R_{xx}(m)$ :

$$\begin{aligned}
 R_{yx}(m) &\triangleq E[y(m)x^*(0)] && \text{by definition of } R_{py} && (\text{Definition 2.4 page 12}) \\
 &= E([\mathbf{Hx}(m) + \mathbf{Hu}(m)]x^*(0)) && \text{by linear hypothesis} && (\text{D}) \\
 &= E([\mathbf{Hx}^*(m)]x^*(0) + [\mathbf{Hu}(0)]x^*(0)) && \text{by linearity of } E && (\text{Theorem ?? page ??}) \\
 &= \mathbf{HE}[x(m)x^*(0)] + \mathbf{HE}[u(0)x^*(0)] && \text{by LTI hypotheses} && (\text{D}) \\
 &= \mathbf{HE}[x(m)x^*(0)] + \mathbf{HE}u(m)\mathbf{Ex}^*(0) && \text{by uncorrelated hypothesis} && (\text{C}) \\
 &= \mathbf{HE}[x(m)x^*(0)] + \mathbf{HE}\cancel{u(m)}\mathbf{Ex}^*(0) && \text{by zero-mean hypothesis} && (\text{B}) \\
 &= \mathbf{HR}_{xx}(m) && \text{by definition of } R_{xx} && (\text{Definition 2.4 page 12}) \\
 &= h(m) \star R_{xx}(m) && \text{by (2) lemma} &&
 \end{aligned}$$



When  $\mathbf{H}$  is *LTI*, what effect does the additive uncorrelated noise sources have on the cross-statistical properties of  $x$  and  $y$ ? Corollary 8.5 (next) demonstrates that, amazingly, under very general conditions, the noise sources have **no effect**.

**Corollary 8.4.** Let  $\mathbf{S}$  be the system illustrated in Figure 8.3 (page 60) (C).

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

PROOF:

$$\begin{aligned}
 \tilde{S}_{xy}(\omega) &= \tilde{S}_{py}(\omega) && \text{by Lemma 8.3 page 58} \\
 &= \tilde{H}^*(\omega)\tilde{S}_{pp}(\omega) && \text{by Corollary 6.3 page 41}
 \end{aligned}$$

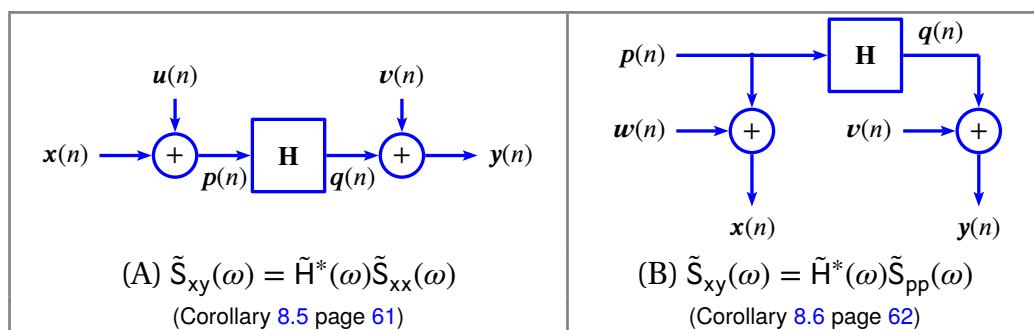


Figure 8.4:

**Corollary 8.5.** Let  $\mathbf{S}$  be the system illustrated in Figure 8.4 page 61 (A).

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

PROOF:

$$\begin{aligned}\tilde{S}_{yx}(\omega) &= \tilde{S}_{qx}(\omega) && \text{by Lemma 8.1 page 57} \\ &= \tilde{H}(\omega)\tilde{S}_{xx}(\omega) && \text{by Corollary 6.3 page 41}\end{aligned}$$

⇒

**Corollary 8.6.** Let  $S$  be the system illustrated in Figure 8.4 page 61 (B).

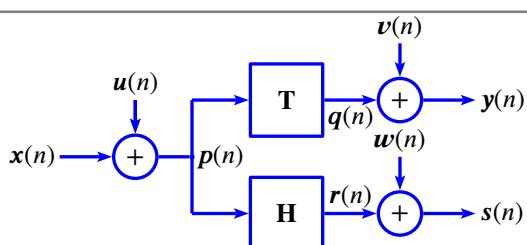
COR	$\left\{ \begin{array}{ll} (A). & x(n) \text{ is WSS} \\ (B). & w(n) \text{ is ZERO-MEAN} \\ (C). & v(n) \text{ is ZERO-MEAN} \\ (D). & x(n), w(n), v(n) \text{ are UNCORRELATED} \\ (E). & H \text{ is LTI} \end{array} \right. \text{ and } \left\{ \begin{array}{l} \tilde{S}_{xy}(\omega) = \tilde{H}^*(\omega)\tilde{S}_{pp}(\omega)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

PROOF:

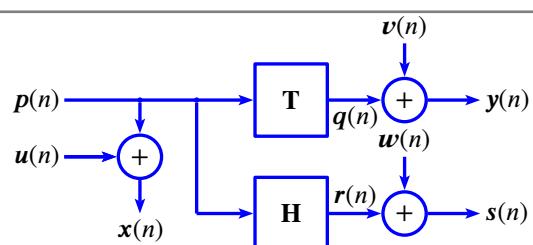
$$\begin{aligned}\tilde{S}_{yx}(\omega) &= \tilde{S}_{qx}(\omega) && \text{by Lemma 8.1 page 57} \\ &= \tilde{S}_{qp}(\omega) && \text{by Lemma 8.1 page 57} \\ &= \tilde{H}(\omega)\tilde{S}_{pp}(\omega) && \text{by Corollary 6.3 page 41}\end{aligned}$$

⇒

## 8.4 Additive noise and dual operators



(A) dual communications additive noise system  
(Corollary 8.7 page 62)



(B) dual measurement additive noise system  
(Corollary 8.8 page 63)

Figure 8.5: Dual Additive Noise Systems

**Corollary 8.7.** Let  $S$  be the system illustrated in Figure 8.5 (page 62) (A).

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

PROOF:

$$\begin{aligned}\check{S}_{sy}(z) &= \check{S}_{rq}(z) && \text{by Corollary 8.2 page 57} && \text{and (B), (C) and (D)} \\ &= \check{H}(z)\check{S}_{pq}(z) && \text{by Theorem 6.2 page 41} && \text{and (A)} \\ &= \check{H}(z)[\check{S}_{xq}(z) + \check{S}_{uq}(z)] && \text{by Lemma 8.2 page 58} \\ &= \check{H}(z)[\check{S}_{xy}(z) + \check{S}_{uy}(z)] && \text{by Lemma 8.1 page 57} \\ \tilde{S}_{sy}(\omega) &= \check{S}_{sy}(z)|_{z=e^{i\omega}} && && \\ &= \check{H}(z)[\check{S}_{xy}(z) + \check{S}_{uy}(z)]|_{z=e^{i\omega}} && \text{by previous result} && (1) \\ &= \tilde{H}(\omega)[\tilde{S}_{xy}(\omega) + \tilde{S}_{uy}(\omega)] && &&\end{aligned}$$





**Corollary 8.8.** Let  $\mathbf{S}$  be the system illustrated in Figure 8.5 (page 62) (B).

<b>C O R</b>	$\left\{ \begin{array}{ll} (A). & \mathbf{H} \text{ is LTI} \\ (B). & \mathbf{x}(n) \text{ is WSS} \\ (C). & \mathbf{u} \text{ and } \mathbf{v} \text{ are ZERO-MEAN} \\ (D). & \mathbf{p}, \mathbf{u}, \mathbf{v} \text{ are UNCORRELATED} \end{array} \right. \text{ and } \left\{ \begin{array}{ll} (1). & \check{\mathbf{S}}_{sy}(z) = \check{\mathbf{H}}(z)\check{\mathbf{S}}_{xy}(z) \text{ and} \\ (2). & \tilde{\mathbf{S}}_{sy}(\omega) = \tilde{\mathbf{H}}(\omega)\tilde{\mathbf{S}}_{xy}(\omega) \end{array} \right\}$
----------------------	--

PROOF:

$$\begin{aligned}
 \check{\mathbf{S}}_{sy}(z) &= \check{\mathbf{S}}_{rq}(z) && \text{by Corollary 8.2 page 57} && \text{and (B), (C) and (D)} \\
 &= \check{\mathbf{H}}(z)\check{\mathbf{S}}_{pq}(z) && \text{by Theorem 6.2 page 41} && \text{and (A)} \\
 &= \check{\mathbf{H}}(z)\check{\mathbf{S}}_{xq}(z) && \text{by Lemma 8.3 page 58} \\
 &= \check{\mathbf{H}}(z)\check{\mathbf{S}}_{xy}(z) && \text{by Lemma 8.1 page 57} \\
 \tilde{\mathbf{S}}_{sy}(\omega) &= \check{\mathbf{S}}_{sy}(z) \Big|_{z=e^{j\omega}} && \text{by definition of } \mathbf{Z} && (\text{Definition R.4 page 358}) \\
 &= \check{\mathbf{H}}(z)\check{\mathbf{S}}_{xy}(z) \Big|_{z=e^{j\omega}} && \text{by previous result} && (1) \\
 &= \tilde{\mathbf{H}}(\omega)\tilde{\mathbf{S}}_{xy}(\omega)
 \end{aligned}$$





## **Part III**

### **Statistical Estimation**



# CHAPTER 9

## ESTIMATION OVERVIEW

### 9.1 Model-based estimation

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

1. The **model**:<sup>2</sup> The presumed system architecture, “the form of which is usually suggested by prior knowledge, physical understanding or guesswork.”<sup>3</sup> 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**.

#### 9.1.1 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 16 page 143).
3. The model architecture (Figure 9.1 page 68)—additive noise model is a common choice, with justification from *Wold's Theorem*.

<sup>1</sup>  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

<sup>2</sup>  Box and Jenkins (1976), pages 173–207 (Chapter 6 “Model Identification”)

<sup>3</sup>  Scargle (1979), page 5

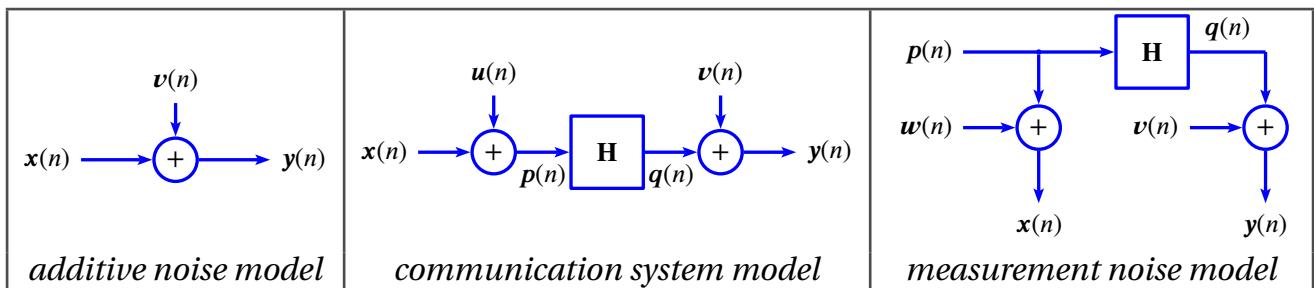


Figure 9.1: Some estimation models

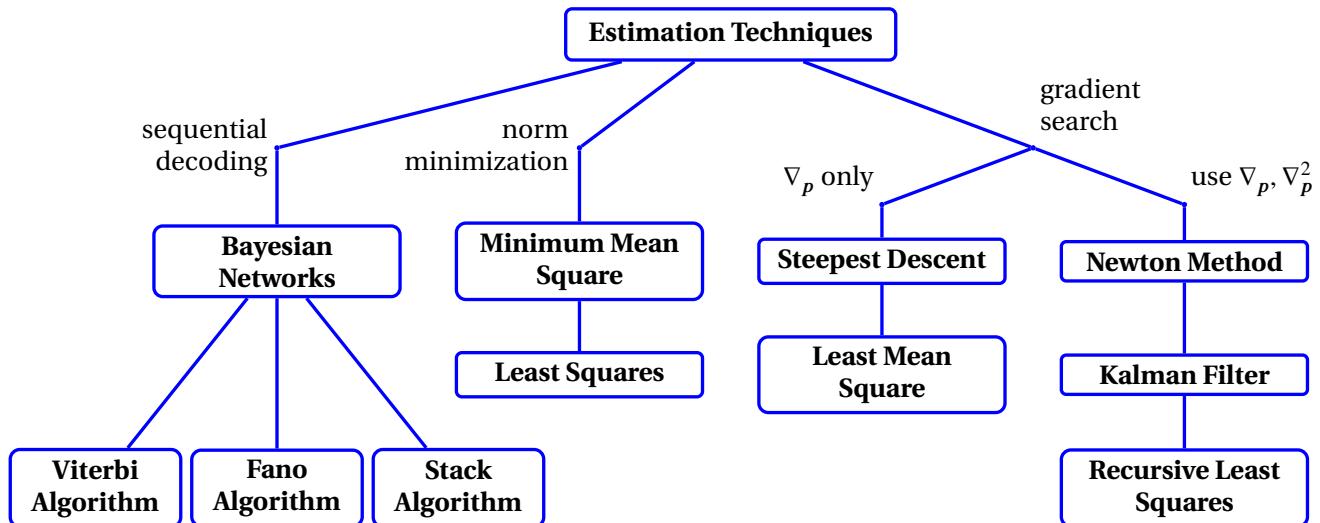


Figure 9.2: Estimation Algorithms

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

### 1. detection:

- ➊ The waveform  $x(t; \theta_n)$  is known except for the value of parameter  $\theta_n$ .
- ➋ The parameter  $\theta_n$  is one of a finite set of values.
- ➌ Estimate  $\theta_n$  and thereby also estimate  $x(t; \theta)$ .

### 2. parametric estimation:

- ➊ The waveform  $x(t; \theta)$  is known except for the value of parameter  $\theta$ .
- ➋ The parameter  $\theta$  is one of an infinite set of values.
- ➌ Estimate  $\theta$  and thereby also estimate  $x(t; \theta)$ .

### 3. nonparametric estimation:

- ➊ The waveform  $x(t)$  is unknown and assumed without any parameter  $\theta$ .
- ➋ Estimate  $x(t)$ .

## 9.1.2 Estimation algorithms

Estimation algorithms include the following (Figure 9.2 page 68):<sup>4</sup>

<sup>4</sup> 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

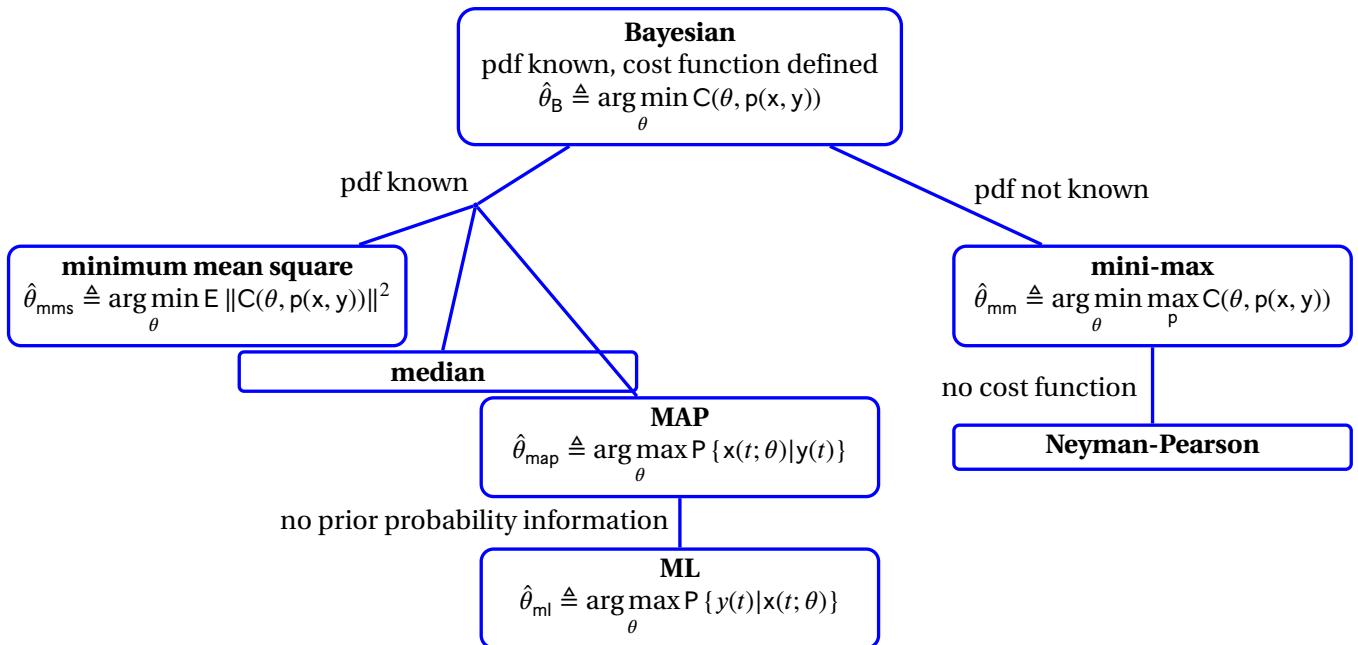


Figure 9.3: Estimation criterion

1. Bayesian Networks (CHAPTER 12 page 103)
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 16 page 143), 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*.<sup>5</sup>

### 9.1.3 Estimation criterion function

Optimization requires a criterion against which the quality of an estimate is measured.<sup>6</sup> 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 9.2 page 68).

**Definition 9.1.** Let

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

Then the following estimates are defined as follows:

<sup>5</sup> (1993) page 276 ('Figure 6.1.1 Options for the adaptive algorithm design.'), <sup>6</sup> Vaseghi (2000) pages 4–5 ('1.2 Signal Processing Methods')

<sup>5</sup> Candy (2009) pages 4–7 ('1.3 SIMULATION-BASED APPROACH TO BAYESIAN PROCESSING'), <sup>6</sup> Liu (2013) page 1 ('1.1 The Need of Monte Carlo Techniques')

<sup>6</sup> Srinath et al. (1996) (013125295X).

DEF

(1). <b>Bayesian estimate</b>	$\hat{\theta}_B \triangleq \arg \min_{\theta} C(\theta, p(x, y))$
(2). <b>Mean square estimate</b> (“MS estimate”)	$\hat{\theta}_{mms} \triangleq \arg \min_{\theta} E \ C(\theta, p(x, y))\ ^2$
(3). <b>mini-max estimate</b> (“MM estimate”)	$\hat{\theta}_{mm} \triangleq \arg \min_{\theta} \max_p C(\theta, p(x, y))$
(4). <b>maximum a-posteriori probability estimate</b> (“MAP estimate”)	$\hat{\theta}_{map} \triangleq \arg \max_{\theta} P\{x(t; \theta)   y(t)\}$
(5). <b>maximum likelihood estimate</b> (“ML estimate”)	$\hat{\theta}_{ml} \triangleq \arg \max_{\theta} P\{y(t)   x(t; \theta)\}$

**Theorem 9.1.** Let  $x(t; \theta)$  be a random process with unknown parameter  $\theta$ .

THM

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

PROOF:

$$\begin{aligned}
 \hat{\theta}_{map} &\triangleq \arg \max_{\theta} P\{x(t; \theta) | y(t)\} && \text{by definition of } \hat{\theta}_{map} && (\text{Definition 9.1 page 69}) \\
 &\triangleq \arg \max_{\theta} \frac{P\{x(t; \theta) \wedge y(t)\}}{P\{r(t)\}} && \text{by definition of } \text{conditional probability} && (\text{Definition A.4 page 156}) \\
 &\triangleq \arg \max_{\theta} \frac{P\{r(t) | x(t; \theta)\} P\{x(t; \theta)\}}{P\{y(t)\}} && \text{by definition of } \text{conditional probability} && (\text{Definition A.4 page 156}) \\
 &= \arg \max_{\theta} P\{y(t) | x(t; \theta)\} P\{x(t; \theta)\} && \text{because } y(t) \text{ is independent of } \theta \\
 &= \arg \max_{\theta} P\{y(t) | x(t; \theta)\} \\
 &\triangleq \hat{\theta}_{ml} && \text{by definition of } \hat{\theta}_{ml} && (\text{Definition 9.1 page 69})
 \end{aligned}$$

⇒

## 9.2 Estimation applications

Applications of estimation include the following:

1. **prediction**—for speech processing, financial markets, etc.
2. **deconvolution**—estimate a system transfer function  $\hat{H}(z)$  and calculate its inverse  $\hat{H}^{-1}(z)$ .
3. **de-noising**—estimate a signal  $x(t)$  that is buried in noise  $v(t)$ .

## 9.3 Measures of estimator quality

**Definition 9.2.** <sup>7</sup>

DEF	The <b>mean square error</b> $\text{mse}(\hat{\theta})$ of an estimate $\hat{\theta}$ of a parameter $\theta$ is defined as	$\text{mse}(\hat{\theta}) \triangleq E[(\hat{\theta} - \theta)^2]$
-----	---	--

<sup>7</sup> 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 9.3.**<sup>8</sup>

**D E F** The **normalized rms error**  $\epsilon(\hat{\theta})$  of an estimate  $\hat{\theta}$  of a parameter  $\theta$  is defined as

$$\epsilon(\hat{\theta}) \triangleq \frac{\sqrt{\text{mse}(\hat{\theta})}}{\theta} \triangleq \frac{\sqrt{\mathbb{E}[(\hat{\theta} - \theta)^2]}}{\theta}$$
**Definition 9.4.**<sup>9</sup>

**D E F** The **mean integrated square error**  $\text{mse}(\hat{\theta})$  of an estimate  $\hat{\theta}$  of a parameter  $\theta$  is defined as

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

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 9.2 in action, see the proof for the  $\text{mse}(\hat{\mu})$  of the *arithmetic mean estimate* as provided in Theorem 15.1 (page 135).

**Theorem 9.2.**<sup>10</sup> Let  $\text{mse}(\hat{\theta})$  be the **MEAN SQUARE ERROR** (Definition 9.2 page 70) and  $\epsilon(\hat{\theta})$  the **NORMALIZED RMS ERROR** (Definition 9.3 page 71) of an estimator  $\hat{\theta}$ .

T H M	$\text{mse}(\hat{\theta}) = \underbrace{\mathbb{E}[(\hat{\theta} - \mathbb{E}\hat{\theta})^2]}_{\text{variance of } \hat{\theta}} + \underbrace{[\mathbb{E}\hat{\theta} - \theta]^2}_{\text{bias of } \hat{\theta} \text{ squared}}$	$\epsilon(\hat{\theta}) = \frac{\sqrt{\mathbb{E}[(\hat{\theta} - \mathbb{E}\hat{\theta})^2] + [\mathbb{E}\hat{\theta} - \theta]^2}}{\theta}$
-------	--	--

☞ PROOF:

$$\begin{aligned}
 \text{mse}(\hat{\theta}) &\triangleq \mathbb{E}[(\hat{\theta} - \theta)^2] && \text{by definition of mse} \quad (\text{Definition 9.2 page 70}) \\
 &= \mathbb{E}\left[\left(\hat{\theta} - \underbrace{\mathbb{E}\hat{\theta} + \mathbb{E}\hat{\theta}}_0 - \theta\right)^2\right] && \text{by additive identity property of } (\mathbb{C}, +, \cdot, 0, 1) \\
 &= \mathbb{E}\left[(\hat{\theta} - \mathbb{E}\hat{\theta})^2 + \underbrace{(\mathbb{E}\hat{\theta} - \theta)^2}_{\text{constant}} - 2(\hat{\theta} - \mathbb{E}\hat{\theta})(\mathbb{E}\hat{\theta} - \theta)\right] && \text{by Binomial Theorem} \\
 &= \mathbb{E}(\hat{\theta} - \mathbb{E}\hat{\theta})^2 + (\mathbb{E}\hat{\theta} - \theta)^2 - 2\mathbb{E}[\hat{\theta}\mathbb{E}\hat{\theta} - \hat{\theta}\theta - \mathbb{E}\hat{\theta}\hat{\theta} + \mathbb{E}\hat{\theta}\theta] && \text{by linearity of E} \quad (\text{Theorem 1.1 page 4}) \\
 &= \mathbb{E}(\hat{\theta} - \mathbb{E}\hat{\theta})^2 + (\mathbb{E}\hat{\theta} - \theta)^2 - 2\underbrace{[\mathbb{E}\hat{\theta}\mathbb{E}\hat{\theta} - \mathbb{E}\hat{\theta}\mathbb{E}\theta - \mathbb{E}\hat{\theta}\mathbb{E}\hat{\theta} + \mathbb{E}\hat{\theta}\mathbb{E}\theta]}_0 && \text{by linearity of E} \quad (\text{Theorem 1.1 page 4}) \\
 &= \mathbb{E}(\hat{\theta} - \mathbb{E}\hat{\theta})^2 + (\mathbb{E}\hat{\theta} - \theta)^2
 \end{aligned}$$

☞

**Definition 9.5.**<sup>11</sup>

**D E F** An estimate  $\hat{\theta}$  of a parameter  $\theta$  is a **minimum variance unbiased estimator (MVUE)** if

- (1).  $\mathbb{E}\hat{\theta} = \theta$  (UNBIASED) and
- (2). no other unbiased estimator  $\hat{\phi}$  has smaller variance  $\text{var}(\hat{\phi})$

<sup>8</sup> ☰ Bendat and Piersol (2010) §“1.4.3 Error Analysis Criteria”

<sup>9</sup> ☰ Silverman (1986) page 35 §“1.3.2 Measures of discrepancy...”, ☱ Rosenblatt (1956) page 835 (“integrated mean square error”)

<sup>10</sup> ☰ Choi (1978) page 76, ☰ Kay (1988) page 45 §“3.3 ESTIMATION THEORY”, ☰ STUART AND ORD (1991) PAGE 629 (“MINIMUM 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”

<sup>11</sup> ☰ Choi (1978) page 76, ☰ Shao (2003) page 161 §“The UMVUE”, ☰ Bolstad (2007) page 164 §“Minimum Variance Unbiased Estimator”



# CHAPTER 10

## LEAST SQUARES ALGORITHMS WHEN MODEL IS UNKNOWN

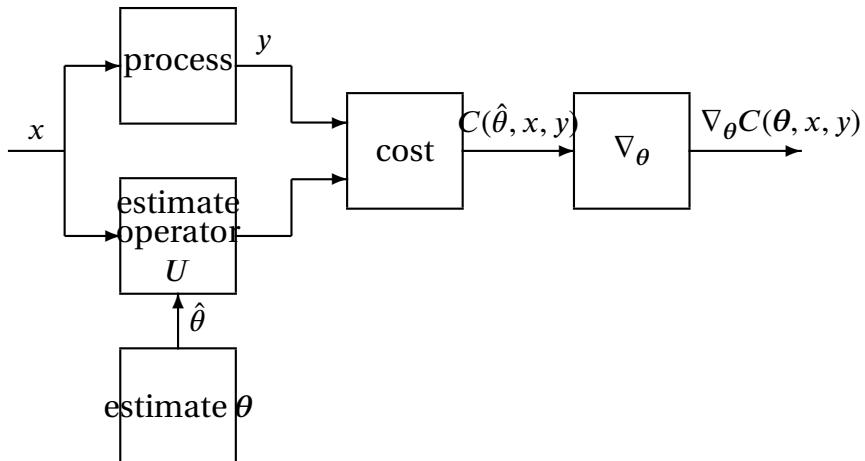


Figure 10.1: Estimation using gradient of cost function

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

1. minimum mean square estimation (MMSE):

The MMS estimate is a *stochastic* estimate. To compute the MMS estimate, 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).<sup>3</sup>

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.

<sup>1</sup> [Clarkson \(1993\) page 1](#) (chapter one Introduction)

<sup>2</sup>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 \(1995\)](#)

<sup>3</sup> [Scargle \(1979\)](#), page 5

Solutions to both are given in terms of two matrices:

$Y$ : Autocorrelation matrix

$W$ : Crosscorrelation matrix.

## 10.1 Minimum mean square estimation

**Definition 10.1.** Let the following vectors, matrices, and functions be defined as follows:

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

**Theorem 10.1** (Minimum mean square estimation). Let

$$\begin{aligned}\hat{y}(\theta) &\triangleq U^H \theta \\ e(\theta) &\triangleq \hat{y} - y \\ C(\theta) &\triangleq E \|e\|^2 \triangleq E[e^H e] \\ \hat{\theta}_{\text{mms}} &\triangleq \arg \min_{\theta} C(\theta) \\ R &\triangleq E[UU^H] \\ W &\triangleq E[Uy]\end{aligned}$$

Then

$$\begin{aligned}\hat{\theta}_{\text{mms}} &= (R_e Y)^{-1} (R_e W) \\ C(\theta) &= \theta^H R \theta - (W^H \theta)^* - W^H \theta + E y^H y \\ \nabla_{\theta} C(\theta) &= 2R_e[Y]\theta - 2R_e W \\ C(\hat{\theta}_{\text{mms}}) &= \begin{cases} E y^H y + (R_e W^H)(R_e Y)^{-1} R (R_e Y)^{-1} (R_e W) - 2(R_e W^H)(R_e Y)^{-1} (R_e W) \\ E y^H y - (R_e W^H) R^{-1} (R_e W) \quad \text{if } R \text{ is real-valued} \end{cases}\end{aligned}$$

PROOF: See APPENDIX E (page 183) for a Matrix Calculus reference.

1. Proof that cost  $C(\theta) = \theta^H R \theta - (W^H \theta)^* - W^H \theta + E y^H y$ :

$$\begin{aligned}C(\theta) &\triangleq E \|e\|^2 && \text{by definition of cost function } C \\ &\triangleq E[e^H e] && \text{by definition of norm } \|\cdot\| \\ &\triangleq E[(\hat{y} - y)^H (\hat{y} - y)] && \text{by definition of error vector } e \\ &\triangleq E[(U^H \theta - y)^H (U^H \theta - y)] && \text{by definition of estimate } \hat{y} \\ &= E[(\theta^H U - y^H)(U^H \theta - y)] && \text{by distributive prop. of } *-\text{algebras (Definition J.3 page 254)} \\ &= E[\theta^H U U^H \theta - \theta^H U y - y^H U^H \theta + y^H y] && \text{by matrix algebra ring property} \\ &= \theta^H E[UU^H] \theta - \theta^H E[Uy] - E[y^H U^H] \theta + E y^H y && \text{by linearity } E \text{ (Theorem 1.1 page 4)} \\ &= \theta^H E[UU^H] \theta - (E[Uy])^H \theta - E[Uy]^H \theta + E y^H y \\ &\triangleq \theta^H R \theta - (W^H \theta)^H - W^H \theta + E y^H y && \text{by definitions of } R \text{ and } W \\ &= \theta^H R \theta - (W^H \theta)^* - W^H \theta + E y^H y && \text{because } W^H \theta \text{ is a scalar} \\ &= \boxed{\theta^H R \theta - (W^H \theta)^* - W^H \theta + E y^H y} \\ &= \theta^H R \theta - 2R_e[W^H] \theta + E y^H y\end{aligned}$$



2. Proof that optimal  $\theta_{\text{opt}} = (\mathbf{R}_e Y)^{-1}(\mathbf{R}_e W)$ :

$$\begin{aligned}
 \nabla_{\theta} C(\theta) &= \nabla_{\theta} [\theta^H \mathbf{R} \theta - (\mathbf{W}^H)^* \theta - \mathbf{W}^H \theta + \mathbf{E} y^H y] \\
 &= \mathbf{R} \theta + \mathbf{R}^T \theta - \nabla_{\theta} [(\mathbf{W}^H)^* \theta + \mathbf{W}^H \theta] + 0 \\
 &= \mathbf{R} \theta + \mathbf{R}^T \theta - [(\mathbf{W}^H)^*]^T - [\mathbf{W}^H]^T \\
 &= \mathbf{R} \theta + (\mathbf{R}^H)^* \theta - \mathbf{W} - \mathbf{W}^* \\
 &= \mathbf{R} \theta + \mathbf{R}^* \theta - \mathbf{W} - \mathbf{W}^* \\
 &= (\mathbf{R} + \mathbf{R}^*) \theta - (\mathbf{W} + \mathbf{W}^*) \\
 &= 2(\mathbf{R}_e Y) \theta - 2\mathbf{R}_e \mathbf{W} \\
 \implies \theta_{\text{opt}} &= (\mathbf{R}_e Y)^{-1}(\mathbf{R}_e \mathbf{W})
 \end{aligned}$$

by item (1)  
by quadratic form result (Theorem E.6 page 188)  
by affine equations result (Theorem E.3 page 187)  
by definition of Hermitian Transpose  ${}^H$   
because  $R$  is Hermitian symmetric  
by ring property  
by definition of  $\mathbf{R}_e$  (Definition J.5 page 255)  
by setting  $\nabla_{\theta} C(\theta) = 0$

3. Cost of optimal  $\theta_{\text{opt}}$ :

$$\begin{aligned}
 C(\theta_{\text{opt}}) &= \theta_{\text{opt}}^H \mathbf{R} \theta_{\text{opt}} - 2\mathbf{R}_e [\mathbf{W}^H] \theta_{\text{opt}} + \mathbf{E} y^H y \\
 &= [(\mathbf{R}_e Y)^{-1}(\mathbf{R}_e \mathbf{W})]^H \mathbf{R} [(\mathbf{R}_e Y)^{-1}(\mathbf{R}_e \mathbf{W})] - 2\mathbf{R}_e [\mathbf{W}^H] [(\mathbf{R}_e Y)^{-1}(\mathbf{R}_e \mathbf{W})] + \mathbf{E} y^H y \quad \text{by item (2)} \\
 &= (\mathbf{R}_e \mathbf{W}^H)(\mathbf{R}_e Y)^{-H} \mathbf{R} (\mathbf{R}_e Y)^{-1}(\mathbf{R}_e \mathbf{W}) - 2\mathbf{R}_e [\mathbf{W}^H](\mathbf{R}_e Y)^{-1}(\mathbf{R}_e \mathbf{W}) + \mathbf{E} y^H y \\
 &= (\mathbf{R}_e \mathbf{W}^H)(\mathbf{R}_e \mathbf{R}^H)^{-1} \mathbf{R} (\mathbf{R}_e Y)^{-1}(\mathbf{R}_e \mathbf{W}) - 2\mathbf{R}_e [\mathbf{W}^H](\mathbf{R}_e Y)^{-1}(\mathbf{R}_e \mathbf{W}) + \mathbf{E} y^H y \\
 &= (\mathbf{R}_e \mathbf{W}^H)(\mathbf{R}_e Y)^{-1} \mathbf{R} (\mathbf{R}_e Y)^{-1}(\mathbf{R}_e \mathbf{W}) - 2(\mathbf{R}_e \mathbf{W}^H)(\mathbf{R}_e Y)^{-1}(\mathbf{R}_e \mathbf{W}) + \mathbf{E} y^H y
 \end{aligned}$$

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

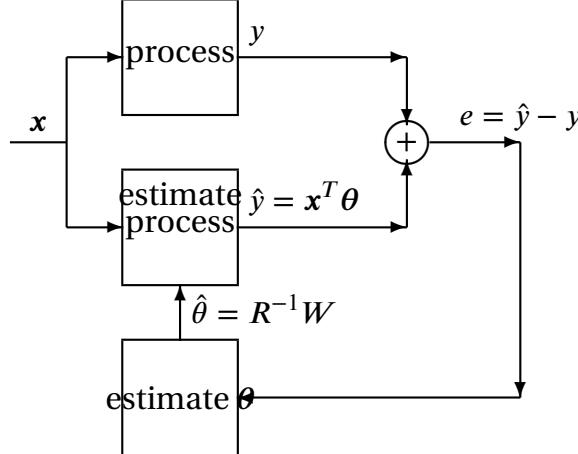


Figure 10.2: Adaptive filter example

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

**Corollary 10.1** (Wiener-Hopf equations). <sup>4</sup>

<sup>4</sup> ↗ 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”)

**COR**

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

PROOF: This is a special case of the more general case discussed in Theorem 10.1 (page 74). Here, the dimension of  $U$  is  $m \times 1$  ( $n=1$ ). As a result,  $y$ ,  $\hat{y}$ , and  $e$  are simply scalar quantities (not vectors). In this special case, we have the following results (Figure 10.2 page 75):

$$\begin{aligned} \hat{y}(\theta) &\triangleq x^T \theta & R &\triangleq E[xx^T] \\ e(\theta) &\triangleq \hat{y} - y & W &\triangleq E[xy] \\ C(\theta) &\triangleq E \|e\|^2 \triangleq E[e^2] & C(\theta) &= \theta^T R \theta - 2W^T \theta + E[y^T y] \\ \hat{\theta}_{\text{mms}} &\triangleq \arg \min_{\theta} C(\theta) & \nabla_{\theta} C(\theta) &= 2R\theta - 2W \\ && C(\hat{\theta}_{\text{mms}})|_R \text{ real} &= E y^T y - W^T R^{-1} W \end{aligned} \Rightarrow$$

## 10.2 Least squares

### 10.2.1 General results

**Theorem 10.2** (Least squares). Let  $\hat{y}(\theta) \triangleq U^H \theta$        $\hat{\theta}_{\text{ls}} \triangleq \arg \min_{\theta} C(\theta)$       Then  
 $e(\theta) \triangleq \hat{y} - y$        $R \triangleq UU^H$   
 $C(\theta) \triangleq \|e\|^2 \triangleq e^H e$        $W \triangleq Uy.$

**THM**

$$\begin{aligned} \hat{\theta}_{\text{ls}} &= (\mathbf{R}_e Y)^{-1} (\mathbf{R}_e W) \\ C(\theta) &= \theta^H R \theta - (W^H \theta)^* - W^H \theta + E y^H y \\ \nabla_{\theta} C(\theta) &= 2\mathbf{R}_e [Y]\theta - 2\mathbf{R}_e W \\ C(\hat{\theta}_{\text{ls}}) &= (\mathbf{R}_e W^H)(\mathbf{R}_e Y)^{-1} R (\mathbf{R}_e Y)^{-1} (\mathbf{R}_e W) - 2(\mathbf{R}_e W^H)(\mathbf{R}_e Y)^{-1} (\mathbf{R}_e W) + E y^H y \\ C(\hat{\theta}_{\text{ls}})|_R \text{ real} &= E y^H y - (\mathbf{R}_e W^H) R^{-1} (\mathbf{R}_e W). \end{aligned}$$

PROOF: See APPENDIX E (page 183) for a Matrix Calculus reference.

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

$$\begin{aligned} \nabla_{\theta} C(\theta) &= \nabla_{\theta} [\theta^H R \theta - (W^H)^* \theta - W^H \theta + y^H y] \\ &= R\theta + R^T \theta - [(W^H)^*]^T - [W^H]^T + 0 \end{aligned}$$



$$\begin{aligned}
&= R\theta + (R^H)^*\theta - W - W^* \\
&= R\theta + R^*\theta - W - W^* \\
&= (R + R^*)\theta - (W + W^*) \\
&= 2(\mathbf{R}_e Y)\theta - 2\mathbf{R}_e W
\end{aligned}$$

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

$$\begin{aligned}
C(\theta_{\text{opt}}) &= \theta_{\text{opt}}^H R \theta_{\text{opt}} - 2\mathbf{R}_e [W^H] \theta_{\text{opt}} + y^H y \\
&= [(\mathbf{R}_e Y)^{-1}(\mathbf{R}_e W)]^H R [(\mathbf{R}_e Y)^{-1}(\mathbf{R}_e W)] - 2\mathbf{R}_e [W^H] [(\mathbf{R}_e Y)^{-1}(\mathbf{R}_e W)] + y^H y \\
&= (\mathbf{R}_e W^H)(\mathbf{R}_e Y)^{-H} R (\mathbf{R}_e Y)^{-1}(\mathbf{R}_e W) - 2\mathbf{R}_e [W^H](\mathbf{R}_e Y)^{-1}(\mathbf{R}_e W) + y^H y \\
&= (\mathbf{R}_e W^H)(\mathbf{R}_e R^H)^{-1} R (\mathbf{R}_e Y)^{-1}(\mathbf{R}_e W) - 2\mathbf{R}_e [W^H](\mathbf{R}_e Y)^{-1}(\mathbf{R}_e W) + y^H y \\
&= (\mathbf{R}_e W^H)(\mathbf{R}_e Y)^{-1} R (\mathbf{R}_e Y)^{-1}(\mathbf{R}_e W) - 2(\mathbf{R}_e W^H)(\mathbf{R}_e Y)^{-1}(\mathbf{R}_e W) + y^H y
\end{aligned}$$

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

⇒

## 10.2.2 Examples using polynomial modeling

*Example 10.1* (Polynomial approximation).

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

$$\begin{aligned}
x &\triangleq \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} & y &\triangleq \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} \\
U^H &\triangleq \underbrace{\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ 1 & x_4 & x_4^2 \\ 1 & x_5 & x_5^2 \end{bmatrix}}_{\text{Vandermonde matrix}} & \hat{\theta} &\triangleq \begin{bmatrix} a \\ b \\ c \end{bmatrix}
\end{aligned}$$

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:<sup>5</sup>

Then, using Theorem 10.2 (page 76), the best coefficients  $\hat{\theta}$  for the polynomial are

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

*Example 10.2.* <sup>6</sup> Find the best fit 3rd order polynomial  $p(x) = dx^3 + cx^2 + bx + a$  for the climate

<sup>5</sup> Horn and Johnson (1990)29

<sup>6</sup> NASA (2019), UEA (2019)

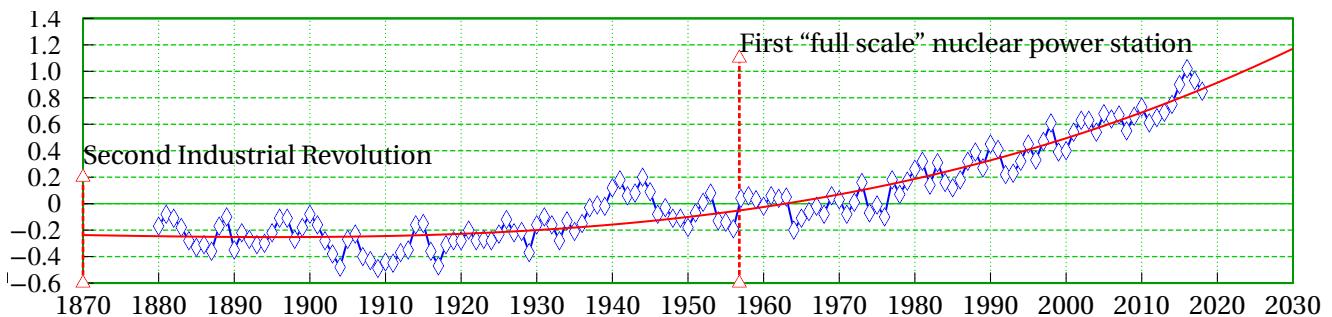


Figure 10.3: Global temperature deviations from average over time (Example 10.2 page 77)

measurements illustrated in Figure 10.3 (page 78).

$$\begin{aligned}\hat{\theta} &= \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = (\mathbf{U}\mathbf{U}^H)^{-1}(\mathbf{U}\mathbf{y}) = \left( \begin{bmatrix} 1 & 0 & 0^2 & 0^3 \\ 1 & 1 & 1^2 & 1^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 138 & 138^2 & 138^3 \end{bmatrix}^H \begin{bmatrix} 1 & 0 & 0^2 & 0^3 \\ 1 & 1 & 1^2 & 1^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 138 & 138^2 & 138^3 \end{bmatrix} \right)^{-1} \left( \begin{bmatrix} 1 & 0 & 0^2 & 0^3 \\ 1 & 1 & 1^2 & 1^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 138 & 138^2 & 138^3 \end{bmatrix}^H \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{138} \end{bmatrix} \right) \\ &= \frac{1}{139} \begin{bmatrix} 1 & 69 & 6371 & 661779 \\ 69 & 6371 & 661779 & 73323839 \\ 6371 & 661779 & 73323839 & 8462609259 \\ 661779 & 73323839 & 8462609259 & 1004606426831 \end{bmatrix}^{-1} \left( \begin{bmatrix} 1 & 0 & 0^2 & 0^3 \\ 1 & 1 & 1^2 & 1^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 138 & 138^2 & 138^3 \end{bmatrix}^H \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{138} \end{bmatrix} \right) \\ &= \begin{bmatrix} -0.24606376973 \\ -0.00076398238 \\ 0.00001657135 \\ 0.00000034336 \end{bmatrix}\end{aligned}$$

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

```

1 Y = [
2 -0.17
3 -0.08
4 -0.11
5 ...
6 0.93
7 0.85
8 ];
9 N = length(Y);
10 X = [0:2018-1880]'; % years - 1880
11 U = [ones(1,N) ; X' , ; X'.^2, ; X'.^3];
12 coefs = inv(U*U')*(U*Y)
13 a=coefs(1)
14 b=coefs(2)
15 c=coefs(3)
16 d=coefs(4)
17 bestFit = d*X.^3 + c*X.^2 + b*X + a;
18 plot(bestFit)
19 x = 2030;
20 predict2030 = d*(x-1880)^3 + c*(x-1880)^2 + b*(x-1880) + a
21 x = 1870;
22 predict1870 = d*(x-1880)^3 + c*(x-1880)^2 + b*(x-1880) + a

```

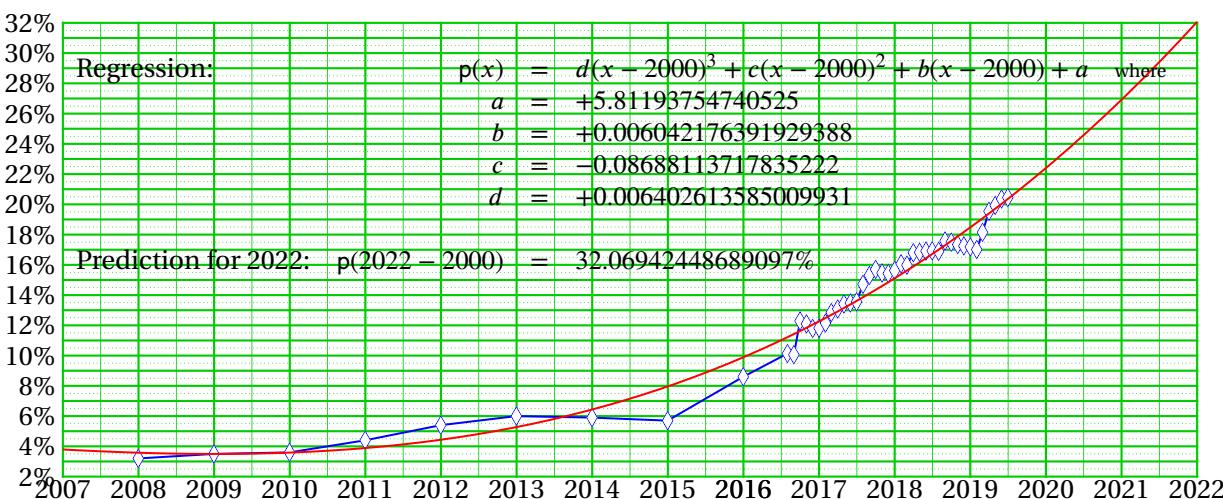


Figure 10.4: Ratio of cagefree chickens in U.S. (Example 10.3 page 79)

*Example 10.3.* <sup>7</sup> Using a 3rd order polynomial fitting (Figure 10.4 page 79), 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:

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

```

1 % years - 2000
2 X = [
3 07.997260 % 2007-12-31 decimal form
4 08.997268 % 2008-12-31 decimal form
5 09.997260 % 2009-12-31 decimal form
6 10.997260 % 2010-12-31 decimal form
7 11.997260 % 2011-12-31 decimal form
8 12.997268 % 2012-12-31 decimal form
9 13.997260 % 2013-12-31 decimal form
10 14.997260 % 2014-12-31 decimal form
11 15.997260 % 2015-12-31 decimal form
12 16.581967 % 2016-08-01 decimal form
13 16.666667 % 2016-09-01 decimal form
14 16.748634 % 2016-10-01 decimal form
15 16.833333 % 2016-11-01 decimal form
16 16.915301 % 2016-12-01 decimal form
17 ...
18 ];
19 Y = [
20 03.2000 % data for 2007-12-31
21 03.5000 % data for 2008-12-31
22 03.6000 % data for 2009-12-31
23 04.4000 % data for 2010-12-31
24 05.4000 % data for 2011-12-31
25 06.0000 % data for 2012-12-31
26 05.9000 % data for 2013-12-31
27 05.7000 % data for 2014-12-31
28 08.6000 % data for 2015-12-31
29 10.1357 % data for 2016-08-01
30 10.0570 % data for 2016-09-01
31 12.2935 % data for 2016-10-01
32 12.1006 % data for 2016-11-01
33 11.7935 % data for 2016-12-01
34 ...

```

<sup>7</sup> Mendez (2019)

```

35 ];
36 format long g
37 pkg load signal;
38 N = length(Y);
39 U = [ones(1,N) ; X', ; X'.^2, ; X'.^3];
40 coefs = inv(U*U')*(U*Y)
41 a=coefs(1)
42 b=coefs(2)
43 c=coefs(3)
44 d=coefs(4)
45 bestFit = d*X.^3 + c*X.^2 + b*X + a;
46 plot(bestFit)
47 x = 2022;
48 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:

```

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

```

## 10.3 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<sup>8</sup>

**T H M**  $\theta_n = \theta_{n-1} - \eta_{n-1} R [\nabla_{\theta} C(\theta_n)]$  where at time  $n$

- |                               |   |          |
|-------------------------------|---|----------|
| $\theta_n$                    | is the <i>state</i>                                       | (vector) |
| $\eta_n$                      | is the <i>step size</i>                                   | (scalar) |
| $Y$                           | is the <i>direction</i>                                   | (matrix) |
| $\nabla_{\theta} C(\theta_n)$ | is the <i>gradient</i> of the cost function $C(\theta_n)$ | (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  $\nabla C$ . Second derivative information, if present, is contained in  $Y$ . Thus, steepest descent

<sup>8</sup>  Nelles (2001), page 90



algorithms do not use second derivative information.

T  
H  
M

$$\theta_n = \theta_{n-1} - \eta_{n-1} [\nabla_{\theta} C(\theta_n)]$$

**Least Mean Squares (LMS).** <sup>9</sup> This is a special case of *steepest descent*. In minimum mean square estimation (Section 10.1 page 74), the cost function  $C(\theta)$  is defined as a *statistical average* of the error vector such that  $C(\theta) = E [e^H e]$ . In this case the gradient  $\nabla C$  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

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

Computing the gradient of this cost function is then just a special case of *least squares estimation* (Section 10.2 page 76). Using LS, we let  $U = \mathbf{x}^T$  and hence

$$\begin{aligned} \nabla_{\theta} C(\theta) &= 2U^T U \theta - 2U^T \mathbf{y} && \text{by Theorem 10.2 page 76} \\ &= 2\mathbf{x}\mathbf{x}^T \theta - 2\mathbf{x}\mathbf{y} && \text{by above definitions} \\ &= 2\mathbf{x}\hat{\mathbf{y}} - 2\mathbf{x}\mathbf{y} \\ &= 2\mathbf{x}(\hat{\mathbf{y}} - \mathbf{y}) \\ &= 2\mathbf{x}e(n) \end{aligned}$$

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

T  
H  
M

$$\theta_n = \theta_{n-1} - 2\eta \mathbf{x}_n e(n)$$

## Second derivative techniques

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

$$\begin{aligned} H_n &\triangleq && \nabla_{\theta} \nabla_{\theta} C(\theta_n) \\ \theta_n &= \theta_{n-1} - \eta_{n-1} H_n^{-1} [\nabla_{\theta} C(\theta_n)] \end{aligned}$$

<sup>9</sup>  Manolakis et al. (2000) page 526

## Kalman filtering <sup>10</sup>

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

**Recursive Least Squares (RLS)** <sup>11</sup> This algorithm is a special case of either the RLS with forgetting or the Kalman filter.

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

## 10.4 Direct search

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

<sup>10</sup>  Nelles (2001), page 66

<sup>11</sup>  Nelles (2001), page 66

# CHAPTER 11

## PROJECTION STATISTIC ALGORITHMS FOR ADDITIVE NOISE MODELS

### 11.1 Projection Statistics

Theorem 11.1 (page 85) (next) shows that the finite set  $Y \triangleq \{\dot{y}_n | n = 1, 2, \dots, N\}$  (a finite number of values) provides just as good an estimate as having the entire  $y(t; \theta)$  waveform (an uncountably infinite number of values) with respect to the following cases:

1. the conditional probability of  $x(t; \theta)$  given  $y(t; \theta)$
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*<sup>1</sup> in “machine learning”. Thus, estimation theory is a form of machine learning.

**Definition 11.1.** Let  $\Psi \triangleq \{\psi_n | n = 1, 2, \dots, N\}$  be an ORTHONORMAL BASIS for a parameterized function  $x(t; \theta)$  with parameter  $\theta$ . Let  $y(t; \theta)$  be  $x(t; \theta)$  plus a RANDOM PROCESS  $v(t)$  such that

$$y(t; \theta) \triangleq x(t; \theta) + v(t)$$

Let  $\dot{y}_n$ ,  $\dot{x}_n$ , and  $\dot{v}_n$  be PROJECTIONS (Definition O.7 page 322) onto the BASIS VECTOR  $\psi_n(t)$  such that

$$\begin{aligned}\dot{y}_n(\theta) &\triangleq \mathbf{P}_n y(t; \theta) \triangleq \langle y(t; \theta) | \psi_n(t) \rangle \triangleq \int_{t \in \mathbb{R}} y(t; \theta) \psi_n(t) dt \\ \dot{x}_n(\theta) &\triangleq \mathbf{P}_n x(t) \triangleq \langle x(t; \theta) | \psi_n(t) \rangle \triangleq \int_{t \in \mathbb{R}} x(t; \theta) \psi_n(t) dt \\ \dot{v}_n &\triangleq \mathbf{P}_n v(t) \triangleq \langle v(t) | \psi_n(t) \rangle \triangleq \int_{t \in \mathbb{R}} v(t) \psi_n(t) dt\end{aligned}$$

Let the set  $Y$  be defined as  $Y \triangleq \{\dot{y}_n(\theta) | 1, 2, \dots, N\}$ . Let  $\hat{\theta}_{\text{map}}$  be the MAP ESTIMATE and  $\hat{\theta}_{\text{ml}}$  be the ML ESTIMATE (Definition 9.1 page 69) of  $\theta$ .

<sup>1</sup>  Bishop (2006), page 2 (9780387310732) “1. INTRODUCTION”

**Lemma 11.1.** Let  $\Psi$ ,  $v(t)$ ,  $\dot{v}_n$ , and  $Y$  be defined as in Definition 11.1 (page 83).

$$\begin{array}{l} \text{L} \\ \text{E} \\ \text{M} \end{array} \quad \left\{ \begin{array}{l} E v(t) = 0 \text{ (ZERO-MEAN)} \end{array} \right\} \implies \left\{ \begin{array}{l} E \dot{v}_n = 0 \text{ (ZERO-MEAN)} \end{array} \right\}$$

PROOF:

$$\begin{aligned} E \dot{v}_n &= E \langle v(t) | \psi_n(t) \rangle && \text{by definition of } \dot{v}_n \\ &= \langle E v(t) | \psi_n(t) \rangle && \text{by linearity of } \langle \Delta | \nabla \rangle \\ &= \langle 0 | \psi_n(t) \rangle && \text{by zero-mean hypothesis} \\ &= 0 \end{aligned} \quad \begin{array}{c} \text{(Definition 11.1 page 83)} \\ \text{by linearity of } \langle \Delta | \nabla \rangle \\ \text{by zero-mean hypothesis} \\ \text{=} \end{array}$$

**Lemma 11.2.** Let  $\Psi$ ,  $v(t)$ ,  $\dot{v}_n$ , and  $Y$  be defined as in Definition 11.1 (page 83).

$$\begin{array}{l} \text{L} \\ \text{E} \\ \text{M} \end{array} \quad \left\{ \begin{array}{l} v(t) \sim N(0, \sigma^2) \text{ (GAUSSIAN)} \end{array} \right\} \implies \left\{ \begin{array}{l} \dot{v}_n \sim N(0, \sigma^2) \text{ (GAUSSIAN)} \end{array} \right\}$$

PROOF: The distribution follows because it is a linear operation on a Gaussian process.  $\Rightarrow$

**Lemma 11.3.** Let  $\Psi$ ,  $y(t; \theta)$ ,  $x(t)$ ,  $v(t)$ ,  $\dot{y}_n$ ,  $\dot{x}_n$ ,  $\dot{v}_n$ , and  $Y$  be defined as in Definition 11.1 (page 83).

$$\begin{array}{l} \text{L} \\ \text{E} \\ \text{M} \end{array} \quad \left\{ \begin{array}{l} (A). \quad E[v(t)] = 0 \\ (B). \quad \text{cov}[v(t), v(u)] = \sigma^2 \delta(t-u) \end{array} \right. \text{ and } \left. \right\} \implies \left\{ \begin{array}{l} (1). \quad E \dot{v}_n = 0 \text{ (ZERO-MEAN)} \\ (2). \quad \text{cov}[\dot{v}_n, \dot{v}_m] = \sigma^2 \bar{\delta}_{n-m} \text{ (UNCORRELATED)} \end{array} \right\}$$

PROOF:

1.

$$E \dot{v}_n = 0 \quad \text{by additive property and Theorem 11.2 page 88}$$

2.

$$\begin{aligned} \text{cov}[\dot{v}_m, \dot{v}_n] &= \text{cov}[\langle v(t) | \psi_m(t) \rangle, \langle v(t) | \psi_n(t) \rangle] && \text{by def. of } \dot{v}_n \quad (\text{Definition 11.1 page 83}) \\ &= \text{cov}\left[\left(\int_{t \in \mathbb{R}} v(t) \psi_m(t) dt\right), \left(\int_{u \in \mathbb{R}} v(u) \psi_n(u) du\right)\right] && \text{by def. of } \langle \Delta | \nabla \rangle \quad (\text{Definition 11.1 page 83}) \\ &= E\left[\left(\int_{t \in \mathbb{R}} v(t) \psi_m(t) dt\right)\left(\int_{u \in \mathbb{R}} v(u) \psi_n(u) du\right)\right] && \text{by def. of Cov} \\ &= E\left[\int_{t \in \mathbb{R}} \int_{u \in \mathbb{R}} v(t) v(u) \psi_m(t) \psi_n(u) dt du\right] \\ &= \int_{t \in \mathbb{R}} \int_{u \in \mathbb{R}} E[v(t)v(u)] \psi_m(t) \psi_n(u) dt du \\ &= \int_{t \in \mathbb{R}} \int_{u \in \mathbb{R}} \sigma^2 \delta(t-u) \psi_m(t) \psi_n(u) dt du && \text{by white hyp.} \quad (B) \\ &= \sigma^2 \int_{t \in \mathbb{R}} \psi_m(t) \psi_n(u) dt \\ &= \sigma^2 \langle \psi_m(t) | \psi_n(u) \rangle && \text{by def. of } \langle \Delta | \nabla \rangle \quad (\text{Definition 11.1 page 83}) \\ &= \begin{cases} \sigma^2 & \text{for } n = m \\ 0 & \text{for } n \neq m. \end{cases} && \text{by orthonormal prop.} \quad (\text{Definition 11.1 page 83}) \end{aligned}$$



**Lemma 11.4.** Let  $\Psi$ ,  $y(t; \theta)$ ,  $x(t)$ ,  $v(t)$ ,  $\dot{y}_n$ ,  $\dot{x}_n$ ,  $\dot{v}_n$ , and  $Y$  be defined as in Definition 11.1 (page 83).

<b>L E M</b>	$\left. \begin{array}{l} (A). \text{ COV}[v(t), v(u)] = \sigma^2 \delta(t-u) \text{ and} \\ (B). \quad v(t) \sim N(0, \sigma^2) \text{ and} \\ (C). \quad \langle \psi_n   \psi_m \rangle = \bar{\delta}_{mn} \end{array} \right\} \Rightarrow \left\{ \begin{array}{ll} (1). & \dot{v}_n \sim N(0, \sigma^2) & (\text{GAUSSIAN}) \\ (2). & \text{COV}[\dot{v}_n, \dot{v}_m] = \sigma^2 \bar{\delta}_{nm} & (\text{UNCORRELATED}) \\ (3). & P\{\dot{v}_n \wedge \dot{v}_m\} = P\{\dot{v}_n\}P\{\dot{v}_m\} & (\text{INDEPENDENT}) \end{array} \right.$
----------------------	--

PROOF:

1. Because the operations are *linear* on processes are *Gaussian* (hypothesis C).

2.

$$\begin{aligned} E\dot{v}_n &= 0 && \text{by AWN properties and Theorem 11.4 page 89} \\ \text{cov}[\dot{v}_m, \dot{v}_n] &= \sigma^2 \bar{\delta}_{mn} && \text{by AWN properties and Lemma 11.3 page 84} \end{aligned}$$

3. Because the processes are *Gaussian, uncorrelated* implies *independent*.



## 11.2 Sufficient Statistics

**Theorem 11.1** (Sufficient Statistic Theorem). <sup>2</sup> Let  $\Psi$ ,  $y(t; \theta)$ ,  $x(t)$ ,  $v(t)$ ,  $\dot{y}_n$ ,  $\dot{x}_n$ ,  $\dot{v}_n$ , and  $Y$  be defined as in Definition 11.1 (page 83). Let  $\hat{\theta}_{\text{map}}$  be the MAP ESTIMATE and  $\hat{\theta}_{\text{ml}}$  be the ML ESTIMATE (Definition 9.1 page 69) of  $\theta$ .

<b>T H M</b>	$\left. \begin{array}{l} (A). \quad v(t) \text{ is ZERO-MEAN} \quad \text{and} \\ (B). \quad v(t) \text{ is WHITE} \quad \text{and} \\ (C). \quad v(t) \text{ is GAUSSIAN} \end{array} \right\} \Rightarrow \underbrace{\left\{ \begin{array}{ll} (1). & P\{x(t; \theta)   y(t; \theta)\} = P\{x(t; \theta)   Y\} \quad \text{and} \\ (2). & \hat{\theta}_{\text{map}} = \arg \max_{\hat{\theta}} P\{x(t; \theta)   Y\} \quad \text{and} \\ (3). & \hat{\theta}_{\text{ml}} = \arg \max_{\hat{\theta}} P\{Y   x(t; \theta)\} \end{array} \right\}}_{\text{the } N \text{ element set } Y \text{ is a SUFFICIENT STATISTIC for estimating } x(t; \theta)}$
----------------------	---

PROOF:

1. definition: Let  $v'(t) \triangleq v(t) - \sum_{n=1}^N \dot{v}_n \psi_n(t)$ .

<sup>2</sup> Fisher (1922) page 316 (“Criterion of Sufficiency”)

2. lemma: The relationship between  $Y$  and  $v'(t)$  is given by

$$\begin{aligned}
 & y(t; \theta) \\
 &= \sum_{n=1}^N \langle y(t; \theta) | \psi_n(t) \rangle \psi_n(t) + \left[ y(t; \theta) - \sum_{n=1}^N \langle y(t; \theta) | \psi_n(t) \rangle \psi_n(t) \right] && \text{by } \textit{additive identity} \text{ property} \\
 &\triangleq \sum_{n=1}^N \langle y(t; \theta) | \psi_n(t) \rangle \psi_n(t) + \left[ y(t; \theta) - \sum_{n=1}^N \langle x(t) + v(t) | \psi_n(t) \rangle \psi_n(t) \right] && \text{by definition of } y(t; \theta) \\
 &= \sum_{n=1}^N \dot{y}_n \psi_n(t) + \underbrace{x(t) + v(t)}_{y(t; \theta)} - \underbrace{\sum_{n=1}^N \langle x(t) | \psi_n(t) \rangle \psi_n(t)}_{x(t)} - \underbrace{\sum_{n=1}^N \langle v(t) | \psi_n(t) \rangle \psi_n(t)}_{v(t) - v'(t)} && \text{by definition of } \dot{y}_n \text{ and} \\
 &&& \text{additive property of } \langle \Delta | \triangleright \rangle \\
 &&& (\text{Definition K.1 page 259}) \\
 &= \sum_{n=1}^N \dot{y}_n \psi_n(t) + x(t) + v(t) - x(t) - [v(t) - v'(t)] \\
 &= \boxed{\sum_{n=1}^N \dot{y}_n \psi_n(t) + v'(t)}
 \end{aligned}$$

3. lemma:  $E[\dot{v}_n v(t)] = N_o \psi_n(t)$ . Proof:

$$\begin{aligned}
 & E[\dot{v}_n v(t)] \\
 &\triangleq E\left[\left(\int_{t \in \mathbb{R}} v(u) \psi_n(u) du\right) v(t)\right] && \text{by definition of } \dot{v}_n(t) \\
 &= E\left[\int_{t \in \mathbb{R}} v(u) v(t) \psi_n(u) du\right] && \text{by } \textit{linearity} \text{ of } \int du \text{ operator} \\
 &= \int_{t \in \mathbb{R}} E[v(u)v(t)] \psi_n(u) du && \text{by } \textit{linearity} \text{ of } E \\
 &= \int_{t \in \mathbb{R}} N_o \delta(u - t) \psi_n(u) du && \text{by } \textit{white hypothesis} \\
 &= N_o \psi_n(t) && \text{by property of } \textit{Dirac delta} \delta(t)
 \end{aligned}$$

4. lemma:  $Y$  and  $v'(t)$  are *uncorrelated*: Proof:

$$\begin{aligned}
 & E[\dot{y}_n v'(t)] \\
 &\triangleq E\left[\langle y(t; \theta) | \psi_n(t) \rangle \left( v(t) - \sum_{n=1}^N \langle v(t) | \psi_n(t) \rangle \psi_n(t) \right)\right] && \text{by definitions of } \dot{y}_n \text{ and } v'(t) \\
 &\triangleq E\left[\langle x(t) + v(t) | \psi_n(t) \rangle \left( v(t) - \sum_{n=1}^N \langle v(t) | \psi_n(t) \rangle \psi_n(t) \right)\right] && \text{by definition of } y(t; \theta) \\
 &= E\left[\left( \langle x(t) | \psi_n(t) \rangle + \langle v(t) | \psi_n(t) \rangle \right) \left( v(t) - \sum_{n=1}^N \langle v(t) | \psi_n(t) \rangle \psi_n(t) \right)\right] && \text{by } \textit{additive property of} \\
 &&& \langle \Delta | \triangleright \rangle (\text{Definition K.1 page 259}) \\
 &= E\left[\left( \dot{x}_n + \dot{v}_n \right) \left( v(t) - \sum_{n=1}^N \dot{v}_n \psi_n(t) \right)\right] && \text{by definitions of } \dot{x}_n \text{ and } \dot{v}_n \\
 &= E\left[\dot{x}_n v(t) - \dot{x}_n \sum_{n=1}^N \dot{v}_n \psi_n(t) + \dot{v}_n v(t) - \dot{v}_n \sum_{n=1}^N \dot{v}_n \psi_n(t)\right] && (\text{Definition 11.1 page 83}) \\
 &= E[\dot{x}_n v(t)] - E\left[\dot{x}_n \sum_{n=1}^N \dot{v}_n \psi_n(t)\right] + E[\dot{v}_n v(t)] - E\left[\sum_{m=1}^N \dot{v}_n \dot{v}_m \psi_m(t)\right] && \text{by } \textit{linearity} \text{ of } E \\
 &&& (\text{Theorem 1.1 page 4})
 \end{aligned}$$



$$\begin{aligned}
 &= \dot{x}_n E[\nu(t)] - \dot{x}_n \sum_{n=1}^N E[\dot{\nu}_n] \psi_n(t) + E[\dot{\nu}_n \nu(t)] - \sum_{m=1}^N E[\dot{\nu}_n \dot{\nu}_m] \psi_m(t) \\
 &= 0 - 0 + E[\dot{\nu}_n \nu(t)] - \sum_{m=1}^N N_o \delta_{mn} \psi_m(t) \\
 &= N_o \psi_n(t) - N_o \psi_n(t) \\
 &= 0
 \end{aligned}$$

by *linearity of E*  
(Theorem 1.1 page 4)

by *white hypothesis*

by (3) lemma

$\implies$  **uncorrelated**

5. lemma:  $Y$  and  $\nu'(t)$  are *independent*. Proof: By (4) lemma,  $\dot{y}_n$  and  $\nu'(t)$  are *uncorrelated*. By hypothesis, they are *Gaussian*, and thus are also **independent**.

6. Proof that  $P\{x(t; \theta) | y(t; \theta)\} = P\{x(t; \theta) | \dot{y}_1, \dot{y}_2, \dots, \dot{y}_N\}$ :

$$\begin{aligned}
 P\{x(t; \theta) | y(t; \theta)\} &= P\left\{x(t; \theta) | \sum_{n=1}^N \dot{y}_n \psi_n(t) + \nu'(t)\right\} \\
 &= P\{x(t; \theta) | Y, \nu'(t)\} \\
 &= \frac{P\{Y, \nu'(t) | x(t; \theta)\} P\{x(t; \theta)\}}{P\{Y, \nu'(t)\}} \\
 &= \frac{P\{Y | x(t; \theta)\} P\{\nu'(t) | x(t; \theta)\} P\{x(t; \theta)\}}{P\{Y\} P\{\nu'(t)\}} \\
 &= \frac{P\{Y | x(t; \theta)\} P\{\nu'(t)\} P\{x(t; \theta)\}}{P\{Y\} P\{\nu'(t)\}} \\
 &= \frac{P\{Y | x(t; \theta)\} P\{x(t; \theta)\}}{P\{Y\}} \\
 &= \frac{P\{Y, x(t; \theta)\}}{P\{Y\}} \\
 &= P\{x(t; \theta) | Y\}
 \end{aligned}$$

because  $Y$  and  $\nu'(t)$  can be extracted by  $\langle \dots | \psi_n(t) \rangle$

by *independence of  $Y$  and  $\nu'(t)$*  ((5) lemma page 87)

by *independence of  $x$  and  $\nu$*

by definition of *conditional probability*  
(Definition A.4 page 156)

7. Proof that  $Y$  is a *sufficient statistic* for the *MAP estimate*:

$$\begin{aligned}
 \hat{\theta}_{\text{map}} &\triangleq \arg \max_{\hat{\theta}} P\{x(t; \theta) | y(t; \theta)\} \\
 &= \arg \max_{\hat{\theta}} P\{x(t; \theta) | Y\}
 \end{aligned}$$

by definition of *MAP estimate* (Definition 9.1 page 69)

by item (6)

8. Proof that  $Y$  is a *sufficient statistic* for the *ML estimate*:

$$\begin{aligned}
 \hat{\theta}_{\text{ml}} &\triangleq \arg \max_{\hat{\theta}} P\{y(t; \theta) | x(t; \theta)\} \\
 &= \arg \max_{\hat{\theta}} P\left\{ \sum_{n=1}^N \dot{y}_n \psi_n(t) + \nu'(t) | x(t; \theta) \right\} \\
 &= \arg \max_{\hat{\theta}} P\{Y, \nu'(t) | x(t; \theta)\} \\
 &= \arg \max_{\hat{\theta}} P\{Y | x(t; \theta)\} P\{\nu'(t)\} x(t; \theta) \\
 &= \arg \max_{\hat{\theta}} P\{Y | x(t; \theta)\} P\{\nu'(t)\} \\
 &= \arg \max_{\hat{\theta}} P\{Y | x(t; \theta)\}
 \end{aligned}$$

by definition of *ML estimate* (Definition 9.1 page 69)

because  $Y$  and  $\nu'(t)$  can be extracted by  $\langle \dots | \psi_n(t) \rangle$

by *independence of  $Y$  and  $\nu'(t)$*  ((5) lemma page 87)

by *independence of  $x(t)$  and  $\nu'(t)$*

by *independence of  $\nu'(t)$  and  $\theta$*



## 11.3 Additive noise

**Theorem 11.2** (Additive noise projection statistics). Let  $\Psi$ ,  $y(t; \theta)$ ,  $x(t)$ ,  $v(t)$ ,  $\dot{y}_n$ ,  $\dot{x}_n$ ,  $\dot{v}_n$ , and  $Y$  be defined as in Definition 11.1 (page 83).

T	H	M	$\left\{ \begin{array}{ll} (A). & y(t; \theta) \triangleq x(t; \theta) + v(t) \quad (\text{additive}) \\ (B). & E[v(t)] = 0 \quad (\text{ZERO-MEAN}) \\ (C). & x(t) \subseteq \text{span } \Psi \quad (\Psi \text{ SPANS } x(t)) \\ (D). & \langle \psi_n   \psi_m \rangle = \bar{\delta}_{mn} \quad (\text{ORTHONORMAL}) \end{array} \right. \text{ and } \left\{ \begin{array}{l} \text{and} \\ \text{and} \\ \text{and} \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} E[\dot{y}_n(\theta)] = \dot{x}_n(\theta) \end{array} \right\}$
---	---	---	---

PROOF:

$$\begin{aligned}
 E[\dot{y}_n(\theta)] &\triangleq E[\langle y(t; \theta) | \psi_n(t) \rangle] && \text{by definition of } \dot{y}_n \\
 &= E\langle x(t; \theta) + v(t) | \psi_n(t) \rangle && \text{by additive hypothesis} \\
 &= E[\langle x(t; \theta) \psi_n(t) | + \rangle \langle v(t) | \psi_n(t) \rangle] && \text{by additive property of } \langle \Delta | \nabla \rangle \\
 &= E\left[ \left\langle \sum_{k=1}^N \dot{x}_k(\theta) \psi_k(t) | \psi_n(t) \right\rangle + \dot{v}_n \right] && \text{by basis hypothesis} \\
 &= E\left[ \sum_{k=1}^N \dot{x}_k(\theta) \langle \psi_k(t) | \psi_n(t) \rangle + \dot{v}_n \right] && \text{by additive property of } \langle \Delta | \nabla \rangle \\
 &= E\left[ \sum_{k=1}^N \dot{x}_k(\theta) \bar{\delta}_{k-n}(t) + \dot{v}_n \right] && \text{by orthonormal hypothesis} \\
 &= E[\dot{x}_n(\theta) + \dot{v}_n] && \text{by definition of } \bar{\delta} \\
 &= E\dot{x}_n(\theta) + E\dot{v}_n && \text{by linearity of } E \\
 &= \dot{x}_n(\theta) && \text{by (B) and Lemma 11.1 page 84}
 \end{aligned}$$

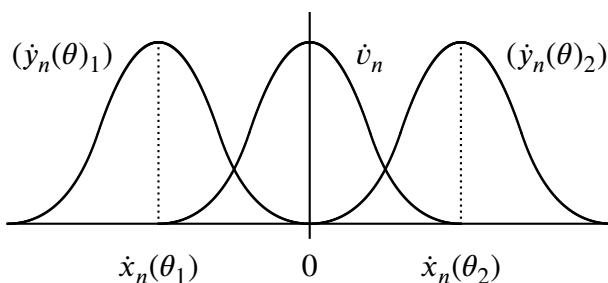


Figure 11.1: Additive Gaussian noise channel Statistics

**Theorem 11.3** (Additive Gaussian noise projection statistics). Let  $\Psi$ ,  $y(t; \theta)$ ,  $x(t)$ ,  $v(t)$ ,  $\dot{y}_n$ ,  $\dot{x}_n$ ,  $\dot{v}_n$ , and  $Y$  be defined as in Definition 11.1 (page 83).

T	H	M	$\left\{ \begin{array}{ll} (A). & y(t; \theta) \triangleq x(t) + v(t) \quad (\text{additive}) \\ (B). & v(t) \sim N(0, \sigma^2) \quad (\text{Gaussian}) \\ (C). & x(t) \subseteq \text{span } \Psi \quad (\Psi \text{ SPANS } x(t)) \\ (D). & \langle \psi_n   \psi_m \rangle = \bar{\delta}_{mn} \quad (\text{ORTHONORMAL}) \end{array} \right. \text{ and } \left\{ \begin{array}{l} \text{and} \\ \text{and} \\ \text{and} \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} \dot{y}_n(\theta) \sim N(\dot{x}_n(\theta), \sigma^2) \quad (\text{GAUSSIAN}) \end{array} \right\}$
---	---	---	--

ADDITIVE GAUSSIAN system

PROOF:

1. Proof for (1): By hypothesis (B) and Lemma 11.1 page 84.

2. Proof for (2):

$$\begin{aligned}
 E[\dot{y}_n(\theta)] &\triangleq E[\langle y(t; \theta) | \psi_n(t) \rangle | \theta] && \text{by definition of } \dot{y}_n && (\text{Definition 11.1 page 83}) \\
 &= E[\langle x(t; \theta) + v(t) | \psi_n(t) \rangle] && \text{by additive hypothesis} && \text{hypothesis (A)} \\
 &= E[\langle x(t; \theta) | \psi_n(t) \rangle] + E[\langle v(t) | \psi_n(t) \rangle] && \text{by additive property of } \langle \Delta | \nabla \rangle && (\text{Definition K.1 page 259}) \\
 &= E\left\langle \sum_{k=1}^N \dot{x}_k(\theta) \psi_k(t) | \psi_n(t) \right\rangle + E\dot{v}_n && \text{by basis hypothesis} && (\text{C}) \\
 &= \sum_{k=1}^N E[\dot{x}_k(\theta)] \langle \psi_k(t) | \psi_n(t) \rangle + E\dot{v}_n && \text{by additive property of } \langle \Delta | \nabla \rangle && (\text{Definition K.1 page 259}) \\
 &= \sum_{k=1}^N E[\dot{x}_k(\theta)] \bar{\delta}_{k-n}(t) + E\dot{v}_n && \text{by orthonormal hypothesis} && (\text{D}) \\
 &= E\dot{x}_n(\theta) + E\dot{v}_n && \text{by definition of } \bar{\delta} && (\text{Definition K.3 page 271}) \\
 &= \dot{x}_n(\theta) + 0 && \text{by Lemma 11.1 page 84}
 \end{aligned}$$

3. Proof for (3): The distribution follows because the process is a linear operations on a Gaussian process.



**Theorem 11.4** (Additive white noise projection statistics). Let  $\Psi$ ,  $y(t; \theta)$ ,  $x(t)$ ,  $v(t)$ ,  $\dot{y}_n$ ,  $\dot{x}_n$ ,  $\dot{v}_n$ , and  $Y$  be defined as in Definition 11.1 (page 83).

<b>T H M</b>	$  \begin{array}{lll}  (A). & y(t; \theta) & \triangleq x(t) + v(t) \quad \text{and} \\  (B). & \text{cov}[v(t), v(u)] & = \sigma^2 \delta(t - u) \quad \text{and} \\  (C). & E[v(t)] & = 0 \quad \text{and} \\  (E). & x(t) & \subseteq \text{span}\Psi \quad \text{and} \\  (E). & \langle \psi_n   \psi_m \rangle & = \bar{\delta}_{mn}  \end{array}  $		$  \underbrace{\qquad\qquad\qquad}_{\text{ADDITIVE WHITE system}} \Rightarrow \left\{  \begin{array}{ll}  (1). & E\dot{v}_n & = 0 & (\text{ZERO-MEAN}) \\  (2). & E(\dot{y}_n(\theta)) & = \dot{x}_n(\theta) \\  (3). & \text{cov}[\dot{v}_n, \dot{v}_m] & = \sigma^2 \bar{\delta}_{nm} & (\text{UNCORRELATED}) \\  (4). & \text{cov}[\dot{y}_n(\theta), \dot{y}_m   \theta] & = \sigma^2 \bar{\delta}_{nm} & (\text{UNCORRELATED})  \end{array}  \right.  $
----------------------	--	--	--

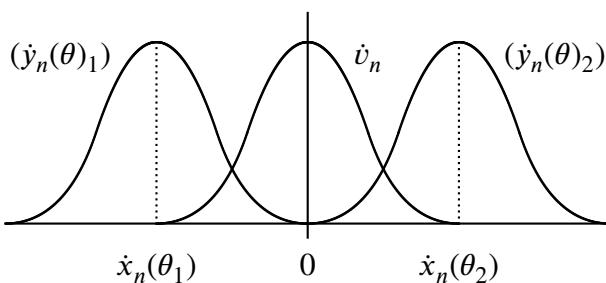
PROOF:

1. Because the noise is *additive* (hypothesis A)...

$$\begin{aligned}
 E\dot{v}_n &= 0 && \text{by additive property and Theorem 11.2 page 88} \\
 (\dot{y}_n(\theta)) &= \dot{x}_n(\theta) + \dot{v}_n && \text{by additive property and Theorem 11.2 page 88} \\
 E(\dot{y}_n | \theta) &= \dot{x}_n(\theta) && \text{by additive property and Theorem 11.2 page 88}
 \end{aligned}$$

2. Proof for (4):

$$\begin{aligned}
 \text{cov}[\dot{y}_n(\theta), \dot{y}_m | \theta] &= E[\dot{y}_n \dot{y}_m | \theta] - [E\dot{y}_n(\theta)][E\dot{y}_m | \theta] \\
 &= E[(\dot{x}_n(\theta) + \dot{v}_n)(\dot{x}_m(\theta) + \dot{v}_m)] - \dot{x}_n(\theta)\dot{x}_m(\theta) \\
 &= E[\dot{x}_n(\theta)\dot{x}_m(\theta) + \dot{x}_n(\theta)\dot{v}_m + \dot{v}_n\dot{x}_m(\theta) + \dot{v}_n\dot{v}_m] - \dot{x}_n(\theta)\dot{x}_m(\theta) \\
 &= \dot{x}_n(\theta)\dot{x}_m(\theta) + \dot{x}_n(\theta)E[\dot{v}_m] + E[\dot{v}_n]\dot{x}_m(\theta) + E[\dot{v}_n\dot{v}_m] - \dot{x}_n(\theta)\dot{x}_m(\theta) \\
 &= 0 + \dot{x}_n(\theta) \cdot 0 + 0 \cdot \dot{x}_m(\theta) + \text{cov}[\dot{v}_n, \dot{v}_m] + [E\dot{v}_n][E\dot{v}_m] \\
 &= \sigma^2 \bar{\delta}_{nm} + 0 \cdot 0 && \text{by Lemma 11.3} \\
 &= \begin{cases} \sigma^2 & \text{for } n = m \\ 0 & \text{for } n \neq m \end{cases}
 \end{aligned}$$

Figure 11.2: Additive white *Gaussian* noise channel statistics

**Theorem 11.5** (AWGN projection statistics). *Let  $\Psi$ ,  $y(t; \theta)$ ,  $x(t)$ ,  $v(t)$ ,  $\dot{y}_n$ ,  $\dot{x}_n$ ,  $\dot{v}_n$ , and  $Y$  be defined as in Definition 11.1 (page 83).*

<b>T H M</b>	$\left. \begin{array}{l} (A). \quad y(t; \theta) \triangleq x(t) + v(t) \text{ and} \\ (B). \quad \text{cov}[v(t), v(u)] = \sigma^2 \delta(t - u) \text{ and} \\ (C). \quad v(t) \sim N(0, \sigma^2) \text{ and} \\ (D). \quad x(t) \subseteq \text{span } \Psi \text{ and} \\ (E). \quad \langle \psi_n   \psi_m \rangle = \bar{\delta}_{mn} \end{array} \right\} \Rightarrow \left\{ \begin{array}{ll} (1). & \dot{y}_n(\theta) \sim N(\dot{x}_n(\theta), \sigma^2) & (\text{GAUSSIAN}) \\ (2). & \text{cov}[\dot{y}_n, \dot{y}_m] = \sigma^2 \bar{\delta}_{nm} & (\text{UNCORRELATED}) \\ (3). & P\{\dot{y}_n \wedge \dot{y}_m\} = P\{\dot{y}_n\} P\{\dot{y}_m\} & (\text{INDEPENDENT}) \end{array} \right.$
ADDITIVE WHITE GAUSSIAN system	

PROOF:

1. Proof for (1) follow because the operations are *linear* on processes are *Gaussian* (hypothesis C).

2.

$$\begin{aligned}
 E\dot{v}_n &= 0 && \text{by AWN properties and Theorem 11.4 page 89} \\
 \dot{y}_n &= \dot{x}_n + \dot{v}_n && \text{by AWN properties and Theorem 11.4 page 89} \\
 E\dot{y}_n &= \dot{x}_n && \text{by AWN properties and Theorem 11.4 page 89} \\
 \text{cov}[\dot{y}_n, \dot{y}_m] &= \sigma^2 \bar{\delta}_{mn} && \text{by AWN properties and Theorem 11.4 page 89}
 \end{aligned}$$

3. Because the processes are *Gaussian*, *uncorrelated* implies *independent*.



## 11.4 ML estimates

The AWGN projection statistics provided by Theorem 11.5 (page 90) help generate the optimal ML-estimates for a number of communication systems. These ML-estimates can be expressed in either of two standard forms:

- ➊ **Spectral decomposition:** The optimal estimate is expressed in terms of *projections* of signals onto orthonormal basis functions.
- ➋ **Matched signal:** The optimal estimate is expressed in terms of the (noisy) received signal correlated with (“matched” with) the (noiseless) transmitted signal.



Theorem 11.6 (page 91) (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 a finite set. In parameter estimation, the estimate is a member of an infinite set (Section 11.4 page 90).

**Theorem 11.6** (General ML estimation). *Let  $\Psi$ ,  $y(t; \theta)$ ,  $x(t)$ ,  $v(t)$ ,  $\dot{y}_n$ ,  $\dot{x}_n$ ,  $\dot{v}_n$ , and  $Y$  be defined as in Definition 11.1 (page 83). Let  $\hat{\theta}_{\text{ml}}$  be the ML ESTIMATE (Definition 9.1 page 69) of  $\theta$ .*

THM	$\begin{aligned}\hat{\theta}_{\text{ml}} &= \arg \min_{\hat{\theta}} \left[ \sum_{n=1}^N [\dot{y}_n - \dot{x}_n(\hat{\theta})]^2 \right] && (\text{spectral decomposition}) \\ &= \arg \max_{\hat{\theta}} \left[ 2 \langle y(t; \theta)   x(t; \theta) \rangle - \ x(t; \theta)\ ^2 \right] && (\text{matched signal})\end{aligned}$
-----	---

PROOF:

$$\begin{aligned}\hat{\theta}_{\text{ml}} &= \arg \max_{\hat{\theta}} P \{ y(t; \theta) | x(t; \theta) \} \\ &= \arg \max_{\hat{\theta}} P \{ \dot{y}_1, \dot{y}_2, \dots, \dot{y}_N | x(t; \theta) \} && \text{by Theorem 11.1 (page 85)}\end{aligned}$$

$$\begin{aligned}&= \arg \max_{\hat{\theta}} \prod_{n=1}^N P \{ \dot{y}_n | x(t; \theta) \} \\ &= \arg \max_{\hat{\theta}} \prod_{n=1}^N p[\dot{y}_n | x(t; \theta)] \\ &= \arg \max_{\hat{\theta}} \prod_{n=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{[\dot{y}_n - \dot{x}_n(\hat{\theta})]^2}{-2\sigma^2} && \text{by Theorem 11.5 (page 90)} \\ &= \arg \max_{\hat{\theta}} \left( \frac{1}{\sqrt{2\pi\sigma^2}} \right)^N \exp \frac{-1}{2\sigma^2} \sum_{n=1}^N [\dot{y}_n - \dot{x}_n(\hat{\theta})]^2\end{aligned}$$

$$= \arg \max_{\hat{\theta}} \left[ - \sum_{n=1}^N [\dot{y}_n - \dot{x}_n(\hat{\theta})]^2 \right]$$

$$= \arg \max_{\hat{\theta}} \left[ - \lim_{N \rightarrow \infty} \sum_{n=1}^N [\dot{y}_n - \dot{x}_n(\hat{\theta})]^2 \right] && \text{by Theorem 11.1 (page 85)}$$

$$= \arg \max_{\hat{\theta}} [- \|y(t; \theta) - x(t; \theta)\|^2] && \text{by Plancheral's formula (Theorem H.9 page 222)}$$

$$= \arg \max_{\hat{\theta}} [- \|y(t; \theta)\|^2 + 2 \mathbf{R}_e \langle y(t; \theta) | x(t; \theta) \rangle - \|x(t; \theta)\|^2]$$

$$= \arg \max_{\hat{\theta}} [2 \langle y(t; \theta) | x(t; \theta) \rangle - \|x(t; \theta)\|^2] && \text{because } y(t; \theta) \text{ independent of } \hat{\theta}$$

**Theorem 11.7** (ML amplitude estimation). <sup>3</sup> Let  $S$  be an additive white gaussian noise system.

<sup>3</sup> Srinath et al. (1996) pages 158–159

THEOREM

$$\left\{ \begin{array}{l} (A). \quad v(t) \text{ is AWGN} \\ (B). \quad y(t; a) = x(t; a) + v(t) \quad \text{and} \\ (C). \quad x(t; a) \triangleq a\lambda(t). \end{array} \right\} \implies \left\{ \begin{array}{ll} (1). & \hat{a}_{\text{ml}} = \frac{1}{\|\lambda(t)\|^2} \sum_{n=1}^N \dot{y}_n \dot{\lambda}_n \\ (2). & E\hat{a}_{\text{ml}} = a \quad (\text{UNBIASED}) \\ (3). & \text{var } \hat{a}_{\text{ml}} = \frac{\sigma^2}{\|\lambda(t)\|^2} \\ (4). & \text{var } \hat{a}_{\text{ml}} = CR \text{ lower bound} \quad (\text{EFFICIENT}) \end{array} \right\}$$

PROOF:

1. *ML estimate* in “matched signal” form:

$$\begin{aligned} \hat{a}_{\text{ml}} &= \arg \max_a [2 \langle y(t; \theta) | x(t; \theta) \rangle - \|x(t; \phi)\|^2] && \text{by Theorem 11.6 (page 91)} \\ &= \arg \max_a [2 \langle y(t; \theta) | a\lambda(t) \rangle - \|a\lambda(t)\|^2] && \text{by hypothesis} \\ &= \arg_a \left[ \frac{\partial}{\partial a} 2a \langle y(t; \theta) | \lambda(t) \rangle - \frac{\partial}{\partial a} a^2 \|\lambda(t)\|^2 = 0 \right] \\ &= \arg_a [2 \langle y(t; \theta) | \lambda(t) \rangle - 2a \|\lambda(t)\|^2 = 0] \\ &= \arg_a [\langle y(t; \theta) | \lambda(t) \rangle = a \|\lambda(t)\|^2] \\ &= \frac{1}{\|\lambda(t)\|^2} \langle y(t; \theta) | \lambda(t) \rangle \end{aligned}$$

2. *ML estimate* in “spectral decomposition” form:

$$\begin{aligned} \hat{a}_{\text{ml}} &= \arg \min_a \left( \sum_{n=1}^N [\dot{y}_n - \dot{x}_n(a)]^2 \right) && \text{by Theorem 11.6 (page 91)} \\ &= \arg_a \left( \frac{\partial}{\partial a} \sum_{n=1}^N [\dot{y}_n - \dot{x}_n(a)]^2 = 0 \right) \\ &= \arg_a \left( 2 \sum_{n=1}^N [\dot{y}_n - \dot{x}_n(a)] \frac{\partial}{\partial a} \dot{x}_n(a) = 0 \right) \\ &= \arg_a \left( \sum_{n=1}^N [\dot{y}_n - \langle a\lambda(t) | \psi_n(t) \rangle] \frac{\partial}{\partial a} \langle a\lambda(t) | \psi_n(t) \rangle = 0 \right) \\ &= \arg_a \left( \sum_{n=1}^N [\dot{y}_n - a \langle \lambda(t) | \psi_n(t) \rangle] \frac{\partial}{\partial a} (a \langle \lambda(t) | \psi_n(t) \rangle) = 0 \right) \\ &= \arg_a \left( \sum_{n=1}^N [\dot{y}_n - a\dot{\lambda}_n] \langle \lambda(t) | \psi_n(t) \rangle = 0 \right) \\ &= \arg_a \left( \sum_{n=1}^N [\dot{y}_n - a\dot{\lambda}_n] \dot{\lambda}_n = 0 \right) \\ &= \arg_a \left( \sum_{n=1}^N \dot{y}_n \dot{\lambda}_n = \sum_{n=1}^N a\dot{\lambda}_n^2 \right) \\ &= \left( \frac{1}{\sum_{n=1}^N \dot{\lambda}_n^2} \right) \sum_{n=1}^N \dot{y}_n \dot{\lambda}_n \\ &= \frac{1}{\|\lambda(t)\|^2} \sum_{n=1}^N \dot{y}_n \dot{\lambda}_n \end{aligned}$$



3. Prove that the estimate  $\hat{a}_{\text{ml}}$  is **unbiased**:

$$\begin{aligned}
 E\hat{a}_{\text{ml}} &= E \frac{1}{\|\lambda(t)\|^2} \int_{t \in \mathbb{R}} y(t; \theta) \lambda(t) dt && \text{by previous result} \\
 &= E \frac{1}{\|\lambda(t)\|^2} \int_{t \in \mathbb{R}} [a\lambda(t) + v(t)] \lambda(t) dt && \text{by hypothesis} \\
 &= \frac{1}{\|\lambda(t)\|^2} \int_{t \in \mathbb{R}} E[a\lambda(t) + v(t)] \lambda(t) dt && \text{by linearity of } \int \cdot dt \text{ and } E \\
 &= \frac{1}{\|\lambda(t)\|^2} a \int_{t \in \mathbb{R}} \lambda^2(t) dt && \text{by } E \text{ operation} \\
 &= \frac{1}{\|\lambda(t)\|^2} a \|\lambda(t)\|^2 && \text{by definition of } \|\cdot\|^2 \\
 &= a
 \end{aligned}$$

4. Compute the variance of  $\hat{a}_{\text{ml}}$ :

$$\begin{aligned}
 E\hat{a}_{\text{ml}}^2 &= E \left[ \frac{1}{\|\lambda(t)\|^2} \int_{t \in \mathbb{R}} y(t; \theta) \lambda(t) dt \right]^2 \\
 &= E \left[ \frac{1}{\|\lambda(t)\|^4} \int_{t \in \mathbb{R}} y(t; \theta) \lambda(t) dt \int_v y(v) \lambda(v) dv \right] \\
 &= E \left[ \frac{1}{\|\lambda(t)\|^4} \int_{t \in \mathbb{R}} \int_v [a\lambda(t) + v(t)][a\lambda(v) + v(v)] \lambda(t) \lambda(v) dv dt \right] \\
 &= E \left[ \frac{1}{\|\lambda(t)\|^4} \int_{t \in \mathbb{R}} \int_v [a^2 \lambda(t) \lambda(v) + a\lambda(t)v(v) + a\lambda(v)v(t) + v(t)v(v)] \lambda(t) \lambda(v) dv dt \right] \\
 &= \left[ \frac{1}{\|\lambda(t)\|^4} \int_{t \in \mathbb{R}} \int_v [a^2 \lambda(t) \lambda(v) + 0 + 0 + \sigma^2 \delta(t - v)] \lambda(t) \lambda(v) dv dt \right] \\
 &= \frac{1}{\|\lambda(t)\|^4} \int_{t \in \mathbb{R}} \int_v a^2 \lambda^2(t) \lambda^2(v) dv dt + \frac{1}{\|\lambda(t)\|^4} \int_{t \in \mathbb{R}} \int_v \sigma^2 \delta(t - v) \lambda(t) \lambda(v) dv dt \\
 &= \frac{1}{\|\lambda(t)\|^4} a^2 \int_{t \in \mathbb{R}} \lambda^2(t) dt \int_v \lambda^2(v) dv + \frac{1}{\|\lambda(t)\|^4} \sigma^2 \int_{t \in \mathbb{R}} \lambda^2(t) dt \\
 &= a^2 \frac{1}{\|\lambda(t)\|^4} \|\lambda(t)\|^2 \|\lambda(v)\|^2 + \frac{1}{\|\lambda(t)\|^4} \sigma^2 \|\lambda(t)\|^2 \\
 &= a^2 + \frac{\sigma^2}{\|\lambda(t)\|^2}
 \end{aligned}$$

$$\begin{aligned}
 \text{var } \hat{a}_{\text{ml}} &= E\hat{a}_{\text{ml}}^2 - (E\hat{a}_{\text{ml}})^2 \\
 &= \left( a^2 + \frac{\sigma^2}{\|\lambda(t)\|^2} \right) - \left( a^2 \right) \\
 &= \frac{\sigma^2}{\|\lambda(t)\|^2}
 \end{aligned}$$

5. Compute the Cramér-Rao Bound:

$$\begin{aligned}
 p[y(t; \theta) | x(t; a)] &= p[\dot{y}_1, \dot{y}_2, \dots, \dot{y}_N | x(t; a)] \\
 &= \prod_{n=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{(\dot{y}_n - a\dot{\lambda}_n)^2}{-2\sigma^2} \\
 &= \left( \frac{1}{\sqrt{2\pi\sigma^2}} \right)^N \exp \frac{1}{-2\sigma^2} \sum_{n=1}^N (\dot{y}_n - a\dot{\lambda}_n)^2
 \end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial a} \ln p[y(t; \theta) | x(t; a)] &= \frac{\partial}{\partial a} \ln \left( \frac{1}{\sqrt{2\pi\sigma^2}} \right)^N \exp \frac{1}{-2\sigma^2} \sum_{n=1}^N (\dot{y}_n - a\dot{\lambda}_n)^2 \\
&= \frac{\partial}{\partial a} \ln \left( \frac{1}{\sqrt{2\pi\sigma^2}} \right)^N + \frac{\partial}{\partial a} \ln \exp \frac{1}{-2\sigma^2} \sum_{n=1}^N (\dot{y}_n - a\dot{\lambda}_n)^2 \\
&= \frac{\partial}{\partial a} \left[ \frac{1}{-2\sigma^2} \sum_{n=1}^N (\dot{y}_n - a\dot{\lambda}_n)^2 \right] \\
&= \frac{1}{-2\sigma^2} \sum_{n=1}^N 2(\dot{y}_n - a\dot{\lambda}_n)(-\dot{\lambda}_n) \\
&= \frac{1}{\sigma^2} \sum_{n=1}^N \dot{\lambda}_n (\dot{y}_n - a\dot{\lambda}_n)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2}{\partial a^2} \ln p[y(t; \theta) | x(t; a)] &= \frac{\partial}{\partial a} \frac{\partial}{\partial a} \ln p[y(t; \theta) | x(t; a)] \\
&= \frac{\partial}{\partial a} \frac{1}{\sigma^2} \sum_{n=1}^N \dot{\lambda}_n (\dot{y}_n - a\dot{\lambda}_n) \\
&= \frac{1}{\sigma^2} \sum_{n=1}^N \dot{\lambda}_n (-\dot{\lambda}_n) \\
&= \frac{-1}{\sigma^2} \sum_{n=1}^N \dot{\lambda}_n^2 \\
&= \frac{-\|\lambda(t)\|^2}{\sigma^2}
\end{aligned}$$

$$\begin{aligned}
\text{var } \hat{a}_{\text{ml}} &\triangleq E[\hat{a}_{\text{ml}} - E\hat{a}_{\text{ml}}]^2 \\
&= E[\hat{a}_{\text{ml}} - a]^2 \\
&\geq \frac{-1}{E\left(\frac{\partial^2}{\partial a^2} \ln p[y(t; \theta) | x(t; a)]\right)} \\
&= \frac{-1}{E\left(\frac{-\|\lambda(t)\|^2}{\sigma^2}\right)} \\
&= \frac{\sigma^2}{\|\lambda(t)\|^2} \quad (\text{Cramér-Rao lower bound of the variance})
\end{aligned}$$

6. Proof that  $\hat{a}_{\text{ml}}$  is an *efficient* estimate:

An estimate is *efficient* if  $\text{var } \hat{a}_{\text{ml}} = \text{CR lower bound}$ . We have already proven this, so  $\hat{a}_{\text{ml}}$  is an *efficient* estimate.

Also, even without explicitly computing the variance of  $\hat{a}_{\text{ml}}$ , the variance equals the *Cramér-Rao lower bound* (and hence  $\hat{a}_{\text{ml}}$  is an *efficient* estimate) if and only if



$$\begin{aligned}
 \hat{a}_{\text{ml}} - a &= \left( \frac{-1}{E\left[ \frac{\partial^2}{\partial a^2} \ln p[y(t; \theta) | x(t; a)] \right]} \right) \left( \frac{\partial}{\partial a} \ln p[y(t; \theta) | x(t; a)] \right) \\
 &\left( \frac{-1}{E\left( \frac{\partial^2}{\partial a^2} \ln p[y(t; \theta) | x(t; a)] \right)} \right) \left( \frac{\partial}{\partial a} \ln p[y(t; \theta) | x(t; a)] \right) = \left( \frac{\sigma^2}{\|\lambda(t)\|^2} \right) \left( \frac{1}{\sigma^2} \sum_{n=1}^N \dot{\lambda}(\dot{y} - a\dot{\lambda}) \right) \\
 &= \frac{1}{\|\lambda(t)\|^2} \sum_{n=1}^N \dot{\lambda}\dot{y} - \frac{1}{\|\lambda(t)\|^2} \sum_{n=1}^N \dot{\lambda}^2 \\
 &= \hat{a}_{\text{ml}} - a
 \end{aligned}$$

⇒

**Theorem 11.8** (ML phase estimation). <sup>4</sup>

T H M	$\left\{ \begin{array}{ll} (A). & v(t) \text{ is AWGN} \\ (B). & y(t; \phi) = x(t; \phi) + v(t) \\ (C). & x(t; \phi) \triangleq A \cos(2\pi f_c t + \phi) \end{array} \right. \text{ and } \right\} \Rightarrow \left\{ \hat{\phi}_{\text{ml}} = -\text{atan} \left( \frac{\langle y(t; \theta)   \sin(2\pi f_c t) \rangle}{\langle y(t; \theta)   \cos(2\pi f_c t) \rangle} \right) \right\}$
-------------	--

PROOF:

$$\begin{aligned}
 \hat{\phi}_{\text{ml}} &= \arg \max_{\phi} [2 \langle y(t; \phi) | x(t; \phi) \rangle - \|x(t; \phi)\|^2] && \text{by Theorem 11.6 (page 91)} \\
 &= \arg \max_{\phi} [2 \langle y(t; \phi) | x(t; \phi) \rangle] && \text{because } \|x(t; \phi)\| \text{ does not depend on } \phi \\
 &= \arg_{\phi} \left[ \frac{\partial}{\partial \phi} \langle y(t; \phi) | x(t; \phi) \rangle = 0 \right] \\
 &= \arg_{\phi} \left[ \left\langle y(t; \phi) | \frac{\partial}{\partial \phi} x(t; \phi) \right\rangle = 0 \right] && \text{because } \langle \Delta | \nabla \rangle \text{ is linear} \\
 &= \arg_{\phi} \left[ \left\langle y(t; \phi) | \frac{\partial}{\partial \phi} A \cos(2\pi f_c t + \phi) \right\rangle = 0 \right] && \text{by definition of } x(t; \phi) \\
 &= \arg_{\phi} [\langle y(t; \phi) | -A \sin(2\pi f_c t + \phi) \rangle = 0] && \text{because } \frac{\partial}{\partial \phi} \cos(x) = -\sin(x) \\
 &= \arg_{\phi} [-A \langle y(t; \phi) | \cos(2\pi f_c t) \sin\phi + \sin(2\pi f_c t) \cos\phi \rangle = 0] && \text{by double angle formulas} \\
 &= \arg_{\phi} [\sin\phi \langle y(t; \phi) | \cos(2\pi f_c t) \rangle = -\cos\phi \langle y(t; \phi) | \sin(2\pi f_c t) \rangle] \\
 &= \arg_{\phi} \left[ \frac{\sin\phi}{\cos\phi} = -\frac{\langle y(t; \phi) | \sin(2\pi f_c t) \rangle}{\langle y(t; \phi) | \cos(2\pi f_c t) \rangle} \right] \\
 &= \arg_{\phi} \left[ \tan\phi = -\frac{\langle y(t; \phi) | \sin(2\pi f_c t) \rangle}{\langle y(t; \phi) | \cos(2\pi f_c t) \rangle} \right] \\
 &= -\text{atan} \left( \frac{\langle y(t; \phi) | \sin(2\pi f_c t) \rangle}{\langle y(t; \phi) | \cos(2\pi f_c t) \rangle} \right)
 \end{aligned}$$

⇒

**Theorem 11.9** (ML estimation of a function of a parameter). <sup>5</sup> Let  $S$  be an additive white gaussian noise system such that  $y(t; \theta) = x(t; \theta) + v(t)$   
 $x(t; \theta) = g(\theta)$ and  $g$  is ONE-TO-ONE AND ONTO (INVERTIBLE).<sup>4</sup> Srinath et al. (1996) pages 159–160<sup>5</sup> Srinath et al. (1996) pages 142–143

THM

Then the optimal ML-estimate of parameter  $\theta$  is

$$\hat{\theta}_{\text{ml}} = g^{-1} \left( \frac{1}{N} \sum_{n=1}^N \dot{y}_n \right).$$

If an ML ESTIMATE  $\hat{\theta}_{\text{ml}}$  is unbiased ( $E\hat{\theta}_{\text{ml}} = \theta$ ) then

$$\text{var } \hat{\theta}_{\text{ml}} \geq \frac{\sigma^2}{N} \frac{1}{\left[ \frac{\partial g(\theta)}{\partial \theta} \right]^2}.$$

If  $g(\theta) = \theta$  then  $\hat{\theta}_{\text{ml}}$  is an **efficient** estimate such that  $\text{var } \hat{\theta}_{\text{ml}} = \frac{\sigma^2}{N}$ .

PROOF:

$$\begin{aligned} \hat{\theta}_{\text{ml}} &= \arg \min_{\theta} \left[ \sum_{n=1}^N [\dot{y}_n - g(\theta)]^2 \right] && \text{by Theorem 11.6 page 91} \\ &= \arg_{\theta} \left[ \frac{\partial}{\partial \theta} \sum_{n=1}^N [\dot{y}_n - g(\theta)]^2 = 0 \right] && \text{because form is quadratic} \\ &= \arg_{\theta} \left[ 2 \sum_{n=1}^N [\dot{y}_n - g(\theta)] \frac{\partial}{\partial \theta} g(\theta) = 0 \right] \\ &= \arg_{\theta} \left[ 2 \sum_{n=1}^N [\dot{y}_n - g(\theta)] = 0 \right] \\ &= \arg_{\theta} \left[ \sum_{n=1}^N \dot{y}_n = N g(\theta) \right] \\ &= \arg_{\theta} \left[ g(\theta) = \frac{1}{N} \sum_{n=1}^N \dot{y}_n \right] \\ &= \arg_{\theta} \left[ \theta = g^{-1} \left( \frac{1}{N} \sum_{n=1}^N \dot{y}_n \right) \right] \\ &= g^{-1} \left( \frac{1}{N} \sum_{n=1}^N \dot{y}_n \right) \end{aligned}$$

If  $\hat{\theta}_{\text{ml}}$  is unbiased ( $E\hat{\theta}_{\text{ml}} = \theta$ ), we can use the *Cramér-Rao bound* to find a lower bound on the variance:

$$\begin{aligned} \text{var } \hat{\theta}_{\text{ml}} &\triangleq E[\hat{\theta}_{\text{ml}} - E\hat{\theta}_{\text{ml}}]^2 \\ &= E[\hat{\theta}_{\text{ml}} - \theta]^2 \\ &\geq \frac{-1}{E \left( \frac{\partial^2}{\partial \theta^2} \ln p[y(t; \theta) | x(t; \theta)] \right)} && \text{by Cramér-Rao Inequality} \\ &= \frac{-1}{E \left( \frac{\partial^2}{\partial \theta^2} \ln p[\dot{y}_1, \dot{y}_2, \dots, \dot{y}_N | x(t; \theta)] \right)} && \text{by Sufficient Statistic Theorem} \\ &&& (\text{Theorem 11.1 page 85}) \\ &= \frac{-1}{E \left( \frac{\partial^2}{\partial \theta^2} \ln \left[ \left( \frac{1}{\sqrt{2\pi\sigma^2}} \right)^N \exp \left( \frac{-1}{2\sigma^2} \sum_{n=1}^N [\dot{y}_n - g(\theta)]^2 \right) \right] \right)} && \text{by AWGN hypothesis} \\ &&& \text{and Theorem 11.5 page 90} \end{aligned}$$



$$\begin{aligned}
&= \frac{-1}{\mathbb{E}\left(\frac{\partial^2}{\partial\theta^2} \ln \left[ \left( \frac{1}{\sqrt{2\pi\sigma^2}} \right)^N \right] + \frac{\partial^2}{\partial\theta^2} \ln \left[ \exp \frac{-1}{2\sigma^2} \sum_{n=1}^N [\dot{y}_n - g(\theta)]^2 \right] \right)} \\
&= \frac{-1}{\mathbb{E}\left(\frac{\partial^2}{\partial\theta^2} \left( \frac{-1}{2\sigma^2} \sum_{n=1}^N [\dot{y}_n - g(\theta)]^2 \right) \right)} \\
&= \frac{2\sigma^2}{\mathbb{E}\left(\frac{\partial}{\partial\theta} \frac{\partial}{\partial\theta} \sum_{n=1}^N [\dot{y}_n - g(\theta)]^2\right)} \\
&= \frac{2\sigma^2}{\mathbb{E}\left(-2 \frac{\partial}{\partial\theta} \frac{\partial g(\theta)}{\partial\theta} \sum_{n=1}^N [\dot{y}_n - g(\theta)]\right)} && \text{by } \textit{Chain Rule} \\
&= \frac{-\sigma^2}{\mathbb{E}\left(\frac{\partial g^2(\theta)}{\partial\theta^2} \sum_{n=1}^N [\dot{y}_n - g(\theta)] + \frac{\partial g(\theta)}{\partial\theta} \frac{\partial}{\partial\theta} \sum_{n=1}^N [\dot{y}_n - g(\theta)]\right)} && \text{by } \textit{Product Rule} \\
&= \frac{-\sigma^2}{\mathbb{E}\left(\frac{\partial g^2(\theta)}{\partial\theta^2} \sum_{n=1}^N [\dot{y}_n - g(\theta)] - N \frac{\partial g(\theta)}{\partial\theta} \frac{\partial g(\theta)}{\partial\theta}\right)} \\
&= \frac{-\sigma^2}{\frac{\partial g^2(\theta)}{\partial\theta^2} \sum_{n=1}^N \mathbb{E}[\dot{y}_n - g(\theta)] - N \frac{\partial g(\theta)}{\partial\theta} \frac{\partial g(\theta)}{\partial\theta}} \\
&= \frac{-\sigma^2}{-N \frac{\partial g(\theta)}{\partial\theta} \frac{\partial g(\theta)}{\partial\theta}} && \text{because derivative of constant} = 0 \\
&= \frac{\sigma^2}{N \left[ \frac{\partial g(\theta)}{\partial\theta} \right]^2}
\end{aligned}$$

The inequality becomes equality (an *efficient* estimate) if and only if

$$\hat{\theta}_{\text{ml}} - \theta = \left( \frac{-1}{\mathbb{E}\left(\frac{\partial^2}{\partial\theta^2} \ln p[y(t; \theta) | x(t; \theta)]\right)} \right) \left( \frac{\partial}{\partial\theta} \ln p[y(t; \theta) | x(t; \theta)] \right).$$

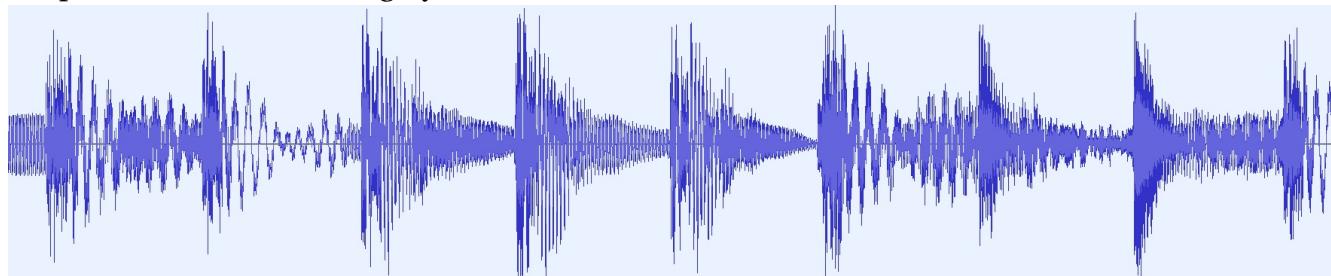
$$\begin{aligned}
&\left( \frac{-1}{\mathbb{E}\left(\frac{\partial^2}{\partial\theta^2} \ln p[y(t; \theta) | x(t; \theta)]\right)} \right) \left( \frac{\partial}{\partial\theta} \ln p[y(t; \theta) | x(t; \theta)] \right) = \left( \frac{\sigma^2}{N} \frac{1}{\left[ \frac{\partial g(\theta)}{\partial\theta} \right]^2} \right) \left( \frac{-1}{2\sigma^2} \left( 2 \frac{\partial g(\theta)}{\partial\theta} \sum_{n=1}^N [\dot{y}_n - g(\theta)] \right) \right) \\
&= -\frac{1}{N} \frac{1}{\frac{\partial g(\theta)}{\partial\theta}} \left( \sum_{n=1}^N [\dot{y}_n - g(\theta)] \right) \\
&= -\frac{1}{\frac{\partial g(\theta)}{\partial\theta}} \left( \frac{1}{N} \sum_{n=1}^N \dot{y}_n - g(\theta) \right) \\
&= -\frac{1}{\frac{\partial g(\theta)}{\partial\theta}} (\hat{\theta}_{\text{ml}} - g(\theta))
\end{aligned}$$

$$= -(\hat{\theta}_{\text{ml}} - \theta)$$

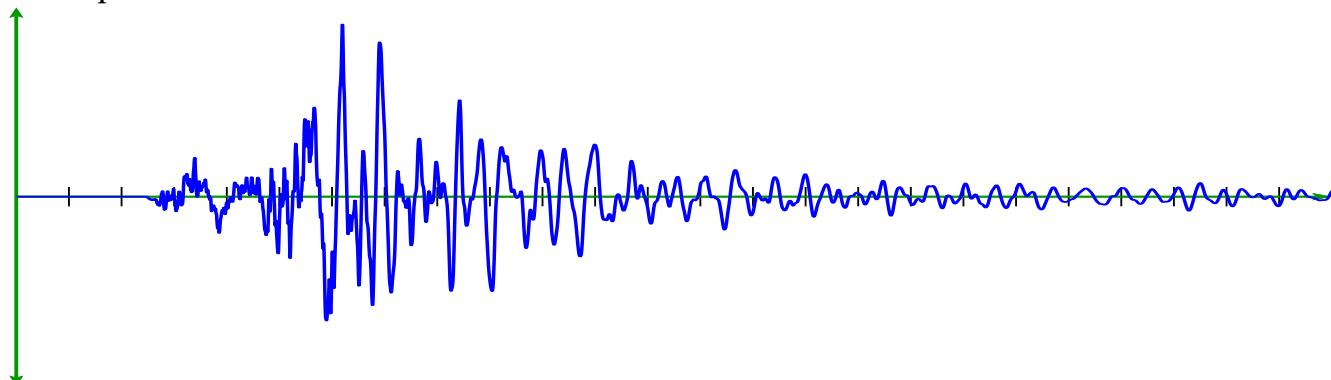
⇒

## 11.5 Example data

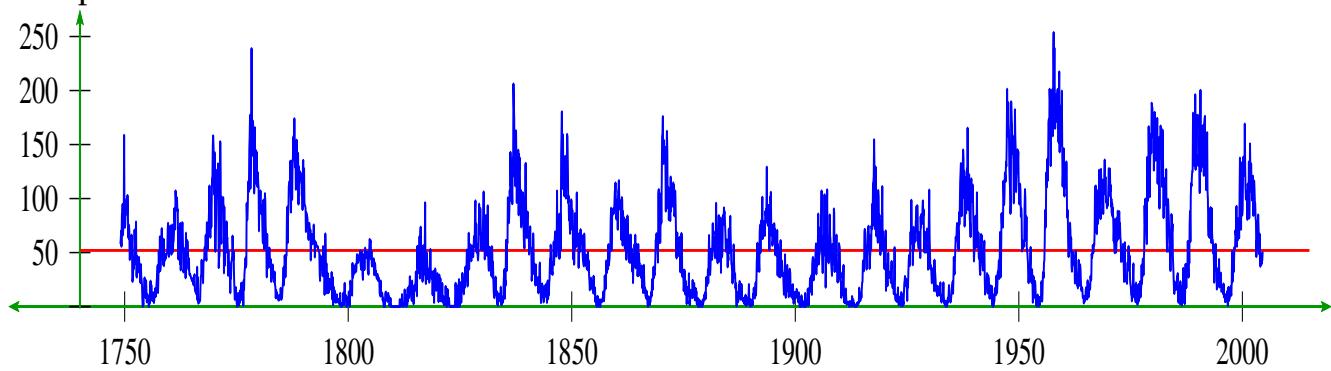
“Pop Goes the World” song by Men Without Hats



Earthquake data



Sunspot data



Dow Jones Industrial Average

<sup>6</sup>[https://www.iris.edu/wilber3/find\\_stations/10953070](https://www.iris.edu/wilber3/find_stations/10953070)

<sup>7</sup><https://d32ogoqmya1dw8.cloudfront.net/files/introgeo/teachingwdata/examples/GreenwichSSNvstime.txt>



$$\begin{aligned}
&= \frac{|\langle \mathbf{x}(t) | f(t) \rangle|^2}{\int_{t \in \mathbb{R}} \int_{\hat{\theta}} \mathbb{E}[\mathbf{v}(t)n^*(\hat{\theta})] \mathbf{x}^*(t)\mathbf{x}(\hat{\theta}) dt du} \\
&= \frac{|\langle \mathbf{x}(t) | \mathbf{x}(t) \rangle|^2}{\int_{t \in \mathbb{R}} \int_{\hat{\theta}} \frac{1}{2} N_o \delta(t - \hat{\theta}) \mathbf{x}^*(t)\mathbf{x}(\hat{\theta}) dt du} \\
&= \frac{|\langle \mathbf{x}(t) | \mathbf{x}(t) \rangle|^2}{\frac{1}{2} N_o \int_{t \in \mathbb{R}} \mathbf{x}^*(t)\mathbf{x}(t) dt} \\
&= \frac{|\langle \mathbf{x}(t) | \mathbf{x}(t) \rangle|^2}{\frac{1}{2} N_o \|\mathbf{x}(t)\|^2} \\
&\leq \frac{\|\mathbf{x}(t)\| \|\mathbf{x}(t)\|^2}{\frac{1}{2} N_o \|\mathbf{x}(t)\|^2} \quad \text{by Cauchy-Schwarz Inequality (Theorem K.2 page 260)} \\
&= \frac{2 \|\mathbf{x}(t)\|^2}{N_o}
\end{aligned}$$

The Cauchy-Schwarz Inequality becomes an equality (SNR is maximized) when  $\mathbf{x}(t) = a\mathbf{x}(t)$ .  $\Rightarrow$

**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  $\langle \mathbf{y}(t; \theta) | f(t) \rangle = \int_0^T \mathbf{y}(t; \theta) f(t) dt$ .

A matched filter introduces a function  $h(t)$  such that  $h(t) = \mathbf{x}(T - t)$  (which implies  $\mathbf{x}(t) = h(T - t)$ ) giving

$$\langle \mathbf{y}(t; \theta) | \mathbf{x}(t) \rangle = \underbrace{\int_0^T \mathbf{y}(t; \theta) \mathbf{x}(t) dt}_{\text{correlator}} = \underbrace{\int_0^\infty \mathbf{x}(\tau) h(t - \tau) d\tau \Big|_{t=T}}_{\text{matched filter}} = \mathbf{x}(t) \star h(t)|_{t=T}.$$

This shows that  $h(t)$  is the impulse response of a filter operation sampled at time  $\tau$ . By Theorem 11.10 (page 99), the optimal impulse response is  $h(\tau - t) = f(t) = \mathbf{x}(t)$ . That is, the optimal  $h(t)$  is just a “flipped” and shifted version of  $\mathbf{x}(t)$ .

## 11.7 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-Loëve basis functions (Section 4.1 page 21)
2. whitening filter<sup>8</sup>

<sup>8</sup> Continuous data whitening: Section 7.3 page 52  
Discrete data whitening: Section 6.4 page 42



**Karhunen-Loève.** If the noise is *white*, the set  $\{\langle y(t; \theta) | \psi_n(t) \rangle | n = 1, 2, \dots, N\}$  is a *sufficient statistic* regardless of which set  $\{\psi_n(t)\}$  of orthonormal basis functions are used. If the noise is *colored*, and if  $\{\psi_n(t)\}$  satisfy the Karhunen-Loève criterion

$$\int_{t_2} R_{xx}(t, u) \psi_n(u) du = \lambda_n \psi_n(t)$$

then the set  $\{\langle y(t; \theta) | \psi_n(t) \rangle\}$  is still a *sufficient statistic*.

**Whitening filter.** The whitening filter makes the received signal  $y(t; \theta)$  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).



# 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 103).

### 12.2 References

1. [Ben-Gal \(2008\) \(ISBN:9780470018613\)](#): Introduction to Bayesian Networks

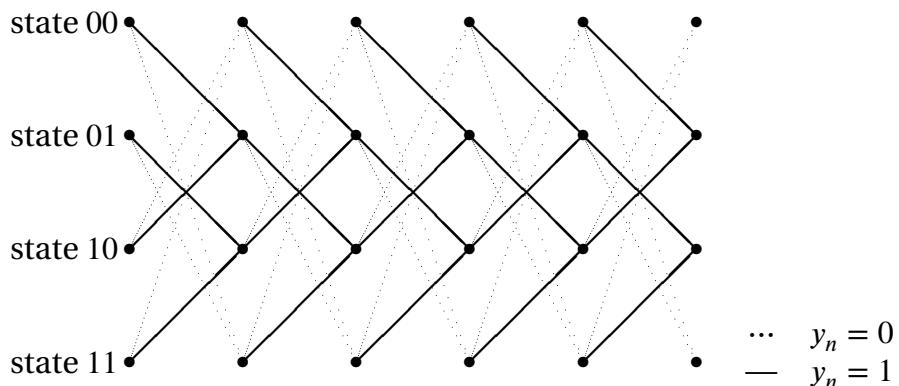


Figure 12.1: Viterbi Algorithm trellis

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 \(1963\)](#)

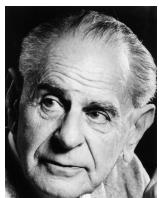
## **Part IV**

# **System Estimation / Modeling**



# CHAPTER 13

## 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)<sup>1</sup>

### 13.1 Estimation techniques

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

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

Determining  $h(n)$  and/or  $\tilde{H}(\omega)$  is part of an operation called “*system identification*”. Determining  $\tilde{H}(\omega)$  in particular is referred to as “*Frequency Response Identification*”<sup>2</sup> or as “*Frequency Response Function*” (“*FRF*”) estimation.<sup>3</sup> *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  $\tilde{H}(\omega)$ —the humble *sine sweep*. That is, they drove the input with a sine wave with slowly increasing (or decreasing) frequency while measuring the resulting output. This technique, although effective, was “very slow”.<sup>4</sup>

<sup>1</sup> quote: [Popper \(1962\)](#), page 231, [Popper \(1963\)](#) page 313

image: [https://en.wikipedia.org/wiki/File:Karl\\_Popper.jpg](https://en.wikipedia.org/wiki/File:Karl_Popper.jpg), “no known copyright restrictions”

<sup>2</sup> [Shin and Hammond \(2008\)](#) page 292

<sup>3</sup> [Cobb \(1988\)](#) page 1 (FRF “measurement”)

<sup>4</sup> [Leuridan et al. \(1986\)](#) 911 “Stepped Sine Testing”, [Cobb \(1988\)](#) page 1 (Chapter 1—Introduction), [Ewins](#)

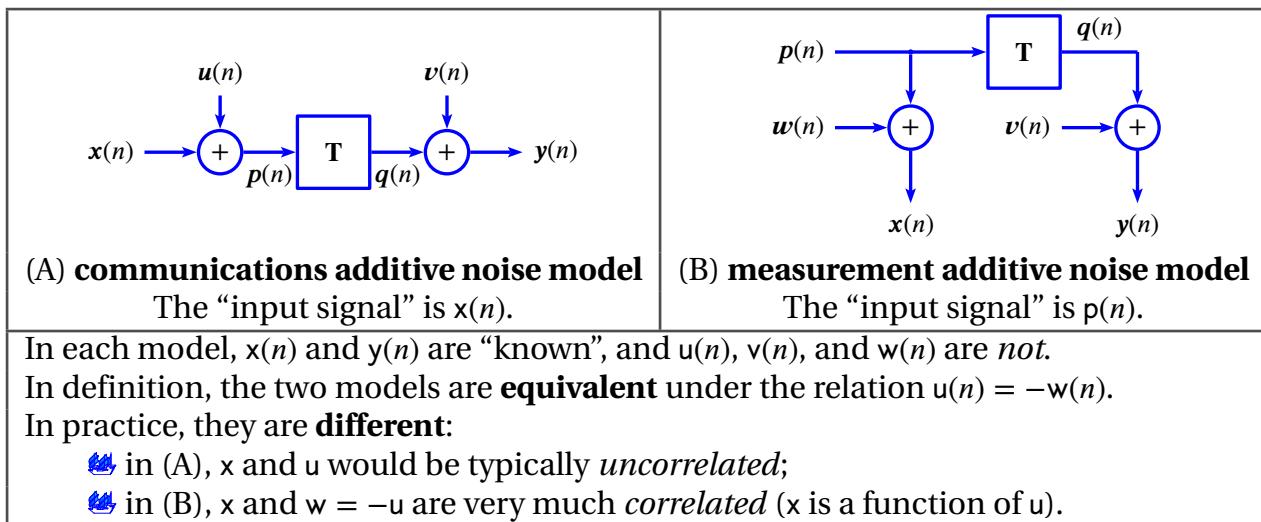


Figure 13.1: Additive noise systems with *linear/non-linear* operator  $\mathbf{T}$

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  $\hat{H}(\omega)$  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  $\hat{H}(\omega)$ , 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”  $EX$ .

The operator  $E$  is also used by the spectral density functions  $\tilde{S}_{xx}(\omega)$  and  $\tilde{S}_{xy}(\omega)$  (Definition 7.3 page 48). And  $\tilde{S}_{xx}(\omega)$  and  $\tilde{S}_{xy}(\omega)$  are what are typically used to calculate an estimate  $\hat{H}(\omega)$ .

## 13.2 Additive noise system models

Consider the additive noise systems illustrated in Figure 14.1 (page 133).

- 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).

Note that the two models are an equivalent system  $S$  under the relation  $u = -w$ . But although one might expect such a sign difference to wreak mathematical havoc in resulting equations, this is

(1986) pages 125–140 {3.7 USE OF DIFFERENT EXCITATION TYPES}



simply not the case here because

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

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

## 13.3 Transfer function estimate definitions and interpretation

As a first attempt at estimating the transfer function  $\mathbf{H}$  of  $\mathbf{S}$ , or at least the magnitude squared of  $\mathbf{H}$ , we might assume  $\mathbf{H}$  to be *LTI*, take a cue from the relation  $\tilde{S}_{yy} = \tilde{S}_{xx}|\tilde{\mathbf{H}}|^2$  of Corollary 6.3 (page 41), and arrive at a function called “*transmissibility*” (next definition).

**Definition 13.1.** <sup>5</sup> Let  $\mathbf{S}$  be a system with input  $x(n)$  and output  $y(n)$ .

**D E F**

The **transmissibility** function  $\tilde{T}_{xy}(\omega)$  is defined as

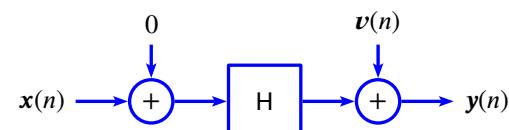
$$\tilde{T}_{xy}(\omega) \triangleq \sqrt{\frac{\tilde{S}_{yy}(\omega)}{\tilde{S}_{xx}(\omega)}}$$

Transmissibility is in essence the ratio of “*spectral power*” (Remark 7.1 page 48) output to *spectral power* input. Note that it is a real-valued function (because  $\tilde{S}_{xx}$  and  $\tilde{S}_{yy}$  are real-valued). We might suspect that we could attain better estimates of  $\mathbf{H}$  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  $\mathbf{S}$  having *zero measurement noise (zero measurement error)* ( $v = u = w = 0$ ),  $\mathbf{h}(n)$  being *linear time invariant (LTI)*, and input  $x(n)$  being *wide sense stationary*...then we can determine (a.k.a “identify”)  $\mathbf{h}(n)$  or  $\tilde{\mathbf{H}}(\omega)$  exactly by  $\tilde{\mathbf{H}}(\omega) = \tilde{S}_{yx}(\omega)/\tilde{S}_{xx}(\omega)$  (Corollary 6.3 page 41).

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  $\hat{\mathbf{H}}_1$  (next definition) is the ideal estimator in the sense that  $\hat{\mathbf{H}}_1 = \tilde{\mathbf{H}}$  (Corollary 13.4 page 127).



**Definition 13.2.** <sup>6</sup> Let  $\mathbf{S}$  be a system with input  $x(n)$  and output  $y(n)$ .

<sup>5</sup> Bendat and Piersol (2010) page 469  $\langle |H(f)| = [G_{yy}(f)/G_{xx}(f)]^{1/2} \rangle$ , Yan and Ren (2012) page 204  $\langle (1) [G_{YY}(s)] = [H(s)][G_{FF}(s)][H^*(s)]^T \rangle$ , Goldman (1999) page 179  $\langle$  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](https://link.springer.com/chapter/10.1007/978-3-319-54109-9_4)

<sup>6</sup> Bendat and Piersol (1993) pages 106–109  $\langle$  5.1.1 Optimality of Calculations  $\rangle$ , Bendat and Piersol (2010) page 185  $\langle H_1(f) = G_{xy}(f)/G_{xx}(f) \rangle$  (6.37), Shin and Hammond (2008) page 293  $\langle H_1(f) = \tilde{S}_{xy}(f)/\tilde{S}_{xx}(f) \rangle$  (9.63); which dif-

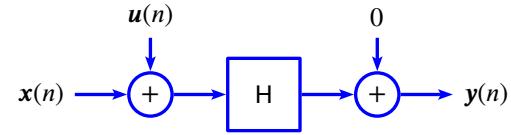
**D  
E  
F**

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

The estimator  $\hat{H}_1$  is a good start. However in the early 1980s, L. D. Mitchell pointed out that in the presence of input noise,  $\hat{H}_1$  is far from ideal in that it is *biased* with respect to  $\tilde{H}$ ; in fact,  $\hat{H}_1$  *under estimates*  $\tilde{H}$  (Corollary 13.4 page 127). Mitchell proposed a new estimator  $\hat{H}_2$  (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  $\hat{H}_2(\omega) = \tilde{H}(\omega)$  (Corollary 13.4 page 127).

Note also that in the case of both no input and no output noise, then  $\hat{H}_1 = \hat{H}_2$  (Corollary 6.3 page 41).



**Definition 13.3.** <sup>7</sup> Let  $S$  be a SYSTEM with input  $x(n)$  and output  $y(n)$ .

**D  
E  
F**

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

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

- $\hat{H}_1$  which is ideal when there is no input noise but *under estimates*  $\tilde{H}$  when there is (Corollary 13.4 page 127)
- $\hat{H}_2$  which is ideal when there is no output noise but *over estimates*  $\tilde{H}$  when there is (Corollary 13.4 page 127).

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  $\hat{H}_1$  and  $\hat{H}_2$  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 13.9 – Definition 13.7 (page 113).

**Definition 13.4.** Let  $S$  be a SYSTEM with input  $x(n)$  and output  $y(n)$ .

**D  
E  
F**

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

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

fers from Definition 13.2, but see [Greenhoe \(2019a\)](#), [Bendat \(1978\)](#)cited by Cobb(1988)—variance estimate for  $\hat{H}_1$ , [Allemand 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  $\hat{H}_1$ , [Allemand et al. \(1987\)](#) pages 54–55 *(5.3.1  $H_1$  Technique;  $[H] = [G_{xF}][G_{FF}]^{-1}$ )* (11)), [Cobb \(1988\)](#) page 2  $\langle^1 \hat{H}(f) = \hat{G}_{yx}(f)/\hat{G}_{xx}(f) \rangle$  (1)), [Goyder \(1984\)](#) page 438  $\langle H(i\omega) = S_{qp}/S_{pp} \rangle$  (3)), [Pintelon and Schoukens \(2012\)](#) page 233  $\langle \hat{G}(\Omega_k) = S_{yu}(j\omega_k)S_{uu}^{-1}(j\omega_k) \rangle$  (7-30)), [White et al. \(2006\)](#) page 678  $\langle H_1(f) = \hat{S}_{x_my_m}(f)/\hat{S}_{x_mx_m}(f) \rangle$  (1) which differs by conjugate, references Bendat and Piersol),

<sup>7</sup> [Shin and Hammond \(2008\)](#) page 293  $\langle H_2(f) = \tilde{S}_{yy}(f)/\tilde{S}_{yx}(f) \rangle$  (9.65); which differs from Definition 13.3, but see [Greenhoe \(2019a\)](#), [Bendat and Piersol \(2010\)](#) page 186  $\langle H_2(f) = G_{yy}(f)/G_{yx}(f) \rangle$  (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 13.3 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  $\langle^2 \hat{H}(f) = \hat{G}_{yy}(f)/\hat{G}_{xy}(f) \rangle$  (1)), [White et al. \(2006\)](#) page 678  $\langle H_2(f) = \hat{S}_{y_my_m}(f)/\hat{S}_{y_mx_m}(f) \rangle$  (2) which differs by conjugate, references Bendat and Piersol)

<sup>8</sup> [Cobb \(1988\)](#) page 3



**Proposition 13.1.** <sup>9</sup> Let  $\mathbf{S}$  be a system with input  $x(n)$  and output  $y(n)$ .

**P R P** 
$$\hat{H}_{\text{am}}(\omega) = \frac{\hat{H}_1(\omega) + \hat{H}_2(\omega)}{2} \quad (\text{arithmetic mean of } \hat{H}_1 \text{ and } \hat{H}_2)$$

PROOF:

$$\begin{aligned} \hat{H}_{\text{am}}(\omega) &\triangleq \frac{|\tilde{S}_{xy}(\omega)|^2 + \tilde{S}_{xx}(\omega)\tilde{S}_{yy}(\omega)}{2\tilde{S}_{xx}(\omega)\tilde{S}_{xy}(\omega)} \quad \text{by definition of } \hat{H}_{\text{am}} \quad (\text{Definition 13.4 page 110}) \\ &= \frac{\tilde{S}_{xy}(\omega)\tilde{S}_{xy}^*(\omega)}{2\tilde{S}_{xx}(\omega)\tilde{S}_{xy}(\omega)} + \frac{\tilde{S}_{xx}(\omega)\tilde{S}_{yy}(\omega)}{2\tilde{S}_{xx}(\omega)\tilde{S}_{xy}(\omega)} = \frac{\tilde{S}_{xy}^*(\omega)}{\tilde{S}_{xx}(\omega)} + \frac{\tilde{S}_{yy}(\omega)}{\tilde{S}_{xy}(\omega)} \\ &= \frac{\hat{H}_1(\omega) + \hat{H}_2(\omega)}{2} \quad \text{by definitions of } \hat{H}_1 \text{ and } \hat{H}_2 \quad (\text{Definition 13.2 page 109, Definition 13.3 page 110}) \end{aligned}$$



**Definition 13.5.** Let  $\mathbf{S}$  be a system with input  $x(n)$  and output  $y(n)$ .

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

**D E F** 
$$\hat{H}_{\text{gm}}(\omega) \triangleq \sqrt{\frac{\tilde{S}_{xy}^*(\omega)}{|\tilde{S}_{xy}(\omega)|}} \sqrt{\frac{\tilde{S}_{yy}(\omega)}{\tilde{S}_{xx}(\omega)}}$$

**Proposition 13.2.** <sup>10</sup> Let  $\mathbf{S}$  be a system with input  $x(n)$  and output  $y(n)$ .

**P R P** 
$$\pm \hat{H}_{\text{gm}}(\omega) = \sqrt{\hat{H}_1(\omega)\hat{H}_2(\omega)} \quad (\text{geometric mean of } \hat{H}_1 \text{ and } \hat{H}_2)$$

PROOF:

$$\begin{aligned} \pm \hat{H}_{\text{gm}}(\omega) &\triangleq \pm \sqrt{\frac{\tilde{S}_{xy}^*(\omega)}{|\tilde{S}_{xy}(\omega)|}} \sqrt{\frac{\tilde{S}_{yy}(\omega)}{\tilde{S}_{xx}(\omega)}} \quad \text{by definition of } \hat{H}_{\text{gm}} \quad (\text{Definition 13.5 page 111}) \\ &= \sqrt{\frac{[\tilde{S}_{xy}^*(\omega)]^2 \tilde{S}_{yy}(\omega)}{|\tilde{S}_{xy}(\omega)|^2 \tilde{S}_{xx}(\omega)}} = \sqrt{\frac{\tilde{S}_{xy}^*(\omega) \tilde{S}_{yy}(\omega)}{\tilde{S}_{xy}(\omega) \tilde{S}_{xx}(\omega)}} = \sqrt{\frac{\tilde{S}_{xy}^*(\omega) \tilde{S}_{yy}(\omega)}{\tilde{S}_{xx}(\omega) \tilde{S}_{xy}(\omega)}} \\ &= \sqrt{\hat{H}_1(\omega)\hat{H}_2(\omega)} \quad \text{by definitions of } \hat{H}_1 \text{ and } \hat{H}_2 \quad (\text{Definition 13.2 page 109, Definition 13.3 page 110}) \\ &= \text{Geometric mean of } \hat{H}_1(\omega) \text{ and } \hat{H}_2(\omega) \end{aligned}$$

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

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

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



Note that the *geometric mean estimator* (Definition 13.5 page 111) and *transmissibility* (Definition 13.1 page 109) are closely related (next).

<sup>9</sup> Mitchell (1982) page 279 (“Frequency Response Calculation: The Average Method”), Zheng et al. (2002) page 918 (“1.3 Arithmetic Mean Estimator  $H_3$ ”)

<sup>10</sup> Zheng et al. (2002) page 918 (“1.4 Geometric Mean Estimator  $H_4$ ”)

<sup>11</sup> Many many thanks to Ben Cleveland for his help with this!!!

**Proposition 13.3.** Let  $\phi(\omega)$  be the PHASE of  $\tilde{S}_{xy}(\omega)$  such that  $\tilde{S}_{xy}(\omega) \triangleq |\tilde{S}_{xy}(\omega)|e^{-i\phi(\omega)}$

P R P	$\hat{H}_{gm}(\omega) = \tilde{T}_{xy}(\omega) e^{-i\phi(\omega)} \quad \left( \begin{array}{l}  \hat{H}_{gm}(\omega)  = \tilde{T}_{xy}(\omega) \text{ is the MAGNITUDE of } \hat{H}_{gm}(\omega) \text{ and} \\ \angle \hat{H}_{gm}(\omega) = -\angle \tilde{S}_{xy}(\omega) \text{ is the PHASE of } \hat{H}_{gm}(\omega) \end{array} \right)$
-------------	--

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

$$\begin{aligned}
 \hat{H}_{gm}(\omega) &\triangleq \sqrt{\hat{H}_1(\omega)\hat{H}_2(\omega)} && \text{by definition of } \hat{H}_{gm} && (\text{Definition 13.5 page 111}) \\
 &\triangleq \sqrt{\frac{\tilde{S}_{xy}^*(\omega)\tilde{S}_{yy}(\omega)}{\tilde{S}_{xx}(\omega)\tilde{S}_{xy}(\omega)}} && \text{by definitions of } \hat{H}_1 \text{ and } \hat{H}_2 && (\text{Definition 13.2 page 109, Definition 13.3 page 110}) \\
 &= \sqrt{\frac{\tilde{S}_{yy}(\omega)\tilde{S}_{xy}^*(\omega)}{\tilde{S}_{xx}(\omega)\tilde{S}_{xy}(\omega)}} \\
 &= \tilde{T}_{xy}(\omega) \sqrt{\frac{\tilde{S}_{xy}^*(\omega)}{\tilde{S}_{xy}(\omega)}} && \text{by definition of } \tilde{T}_{xy} && (\text{Definition 13.1 page 109}) \\
 &= \tilde{T}_{xy}(\omega) \sqrt{\frac{|\tilde{S}_{xy}(\omega)|e^{-i\phi(\omega)}}{|\tilde{S}_{xy}(\omega)|e^{i\phi(\omega)}}} && \text{where } \tilde{S}_{xy}(\omega) \triangleq |\tilde{S}_{xy}(\omega)|e^{-i\phi(\omega)} \\
 &= \tilde{T}_{xy}(\omega) \sqrt{e^{-i2\phi(\omega)}} \\
 &= \tilde{T}_{xy}(\omega) e^{-i\phi(\omega)}
 \end{aligned}$$



*Remark 13.1.* Transmissibility  $\tilde{T}_{xy}$  is a kind of “black sheep” of the system identification function family. All the other members of this family ( $\hat{H}_1$ ,  $\hat{H}_2$ ,  $\hat{H}_v$ ,  $\hat{H}_s$ ) are *complex-valued*, but  $\tilde{T}_{xy}$  is only *real-valued*—a seemingly ordinary Joe born into a super-hero family. But Proposition 13.3 suggests that  $\tilde{T}_{xy}$  is not simply a “black sheep”, but rather a “dark horse” with abilities that can easily be unleashed by slight redefinition. In particular, Proposition 13.3 demonstrates that  $\tilde{T}_{xy}$  is the *magnitude* of the geometric mean of  $\hat{H}_1$  and  $\hat{H}_2$ . We can thus justifiably define a **complex transmissibility** function as  $\hat{H}_{gm}$ ...and the magnitude of this *complex transmissibility* function is the *ordinary transmissibility* function of Definition 13.1 (page 109).

R  
E  
M

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

**Definition 13.6.** Let  $S$  be a SYSTEM with input  $x(n)$  and output  $y(n)$ .

**D  
E  
F** The **Harmonic mean transfer function estimate**  $\hat{H}_{hm}(\omega)$  of  $S$  is defined as

$$\hat{H}_{hm}(\omega) \triangleq \frac{2\tilde{S}_{yy}(\omega)\tilde{S}_{xy}^*(\omega)}{\tilde{S}_{xx}(\omega)\tilde{S}_{yy}(\omega) + |\tilde{S}_{xy}(\omega)|^2}$$

**Proposition 13.4.**<sup>12</sup> Let  $S$  be a SYSTEM with input  $x(n)$  and output  $y(n)$ .

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

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



PROOF:

$$\begin{aligned}
 \hat{H}_{\text{hm}}(\omega) &\triangleq \frac{2\tilde{S}_{yy}(\omega)\tilde{S}_{xy}^*(\omega)}{\tilde{S}_{xx}(\omega)\tilde{S}_{yy}(\omega) + |\tilde{S}_{xy}(\omega)|^2} \quad \text{by definition of } \hat{H}_{\text{hm}} \quad (\text{Definition 13.6 page 112}) \\
 &= \frac{2}{\frac{\tilde{S}_{xx}(\omega)\tilde{S}_{yy}(\omega) + |\tilde{S}_{xy}(\omega)|^2}{\tilde{S}_{yy}(\omega)\tilde{S}_{xy}^*(\omega)}} = \frac{2}{\frac{\tilde{S}_{xx}(\omega)}{\tilde{S}_{xy}^*(\omega)} + \frac{\tilde{S}_{xy}(\omega)}{\tilde{S}_{yy}(\omega)}} \\
 &= \frac{2}{\frac{1}{\hat{H}_1(\omega)} + \frac{1}{\hat{H}_2(\omega)}} \quad \text{by definitions of } \hat{H}_1 \text{ and } \hat{H}_2 \quad (\text{Definition 13.2 page 109, Definition 13.3 page 110}) \\
 &= \text{Harmonic mean of } \hat{H}_1(\omega) \text{ and } \hat{H}_2(\omega)
 \end{aligned}$$



A bit of review reveals  $\hat{H}_1$  at the low end of the estimation problem,  $\hat{H}_2$  at the high end, and  $\hat{H}_{\text{hm}}$ ,  $\hat{H}_{\text{gm}}$ , and  $\hat{H}_{\text{am}}$  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  $\hat{H}_1$  and  $\hat{H}_2$  as limiting cases
- include the between cases
- are optimal in some sense

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

- has  $\hat{H}_1$  and  $\hat{H}_2$  as limiting cases (Theorem 13.1 page 115),
- is optimal in the least squares sense (Theorem 13.6 page 128), and
- allows for a system designer to specify an output-input spectral noise ratio  $\kappa(\omega)$  that can vary with frequency  $\omega$ .

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

- In the case of constant  $\kappa$ ,  $\hat{H}_\kappa$  simplifies to the *Scaling transfer function estimate*  $\hat{H}_s$  (Definition 13.8 page 113).
- In the case of  $\kappa = 1$ ,  $\hat{H}_\kappa$  and  $\hat{H}_s$  simplify to the *Total least squares transfer function estimate*  $\hat{H}_v$  (Definition 13.9 page 114).

**Definition 13.7.** <sup>13</sup> Let  $S$  be a system with input  $x(n)$  and output  $y(n)$ .

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

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

**Definition 13.8.** <sup>14</sup> Let  $S$  be a system with input  $x(n)$  and output  $y(n)$ .

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

<sup>13</sup> White et al. (2006) page 679 ((6)), Shin and Hammond (2008) page 293 ((9.67))

<sup>14</sup> Shin and Hammond (2008) page 293 ((9.67) with  $\kappa(\omega) = s^2$ ), White et al. (2006) page 679 ((6) with  $\kappa(\omega) = s^2$ ), Leclerc 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 13.9.** <sup>15</sup> Let  $\mathbf{S}$  be a system with input  $x(n)$  and output  $y(n)$ .

**D E F** The Total Least Squares transfer function estimate  $\hat{H}_v(\omega)$  of  $\mathbf{S}$  is defined as  
 $\hat{H}_v(\omega) \triangleq \hat{H}_k(\omega; \kappa)$  with  $\kappa(\omega) = 1$

The previous estimators all assumed two signals: an input  $x(n)$  and an output  $y(n)$ . 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 13.10** (Three channel estimate). <sup>16</sup> Let  $\mathbf{S}$  be a system with input  $x(n)$ , output  $y(n)$ , and a driver  $p(n)$ .

**D E F** The transfer function estimate  $\hat{H}_c(\omega)$  is defined as  
 $\hat{H}_c(\omega) \triangleq \frac{\tilde{S}_{py}(\omega)}{\tilde{S}_{px}(\omega)}$

## 13.4 Estimator relationships

**Lemma 13.1.**

**L E M**

$$\begin{aligned} \frac{d}{dp} \left[ \tilde{S}_{yy} - p\tilde{S}_{xx} + \sqrt{(\tilde{S}_{yy} - p\tilde{S}_{xx})^2 + 4p|\tilde{S}_{xy}|^2} \right] &= \frac{\tilde{S}_{xx}(p\tilde{S}_{xx} - \tilde{S}_{yy}) + 2|\tilde{S}_{xy}|^2 - \tilde{S}_{xx}\sqrt{(p\tilde{S}_{xx} - \tilde{S}_{yy})^2 + 4p|\tilde{S}_{xy}|^2}}{\sqrt{(p\tilde{S}_{xx} - \tilde{S}_{yy})^2 + 4p|\tilde{S}_{xy}|^2}} \\ \frac{d}{dp} \left[ p\tilde{S}_{yy} - \tilde{S}_{xx} + \sqrt{(p\tilde{S}_{yy} - \tilde{S}_{xx})^2 + 4p|\tilde{S}_{xy}|^2} \right] &= \frac{\tilde{S}_{yy}(p\tilde{S}_{yy} - \tilde{S}_{xx}) + 2|\tilde{S}_{xy}|^2 + \tilde{S}_{yy}\sqrt{(p\tilde{S}_{yy} - \tilde{S}_{xx})^2 + 4p|\tilde{S}_{xy}|^2}}{\sqrt{(p\tilde{S}_{yy} - \tilde{S}_{xx})^2 + 4p|\tilde{S}_{xy}|^2}} \end{aligned}$$

PROOF:

$$\begin{aligned} \frac{d}{dp} \left[ \tilde{S}_{yy} - p\tilde{S}_{xx} + \sqrt{(\tilde{S}_{yy} - p\tilde{S}_{xx})^2 + 4p|\tilde{S}_{xy}|^2} \right] &= -\tilde{S}_{xx} + \frac{-2\tilde{S}_{xx}(\tilde{S}_{yy} - p\tilde{S}_{xx}) + 4|\tilde{S}_{xy}|^2}{2\sqrt{(\tilde{S}_{yy} - p\tilde{S}_{xx})^2 + 4p|\tilde{S}_{xy}|^2}} \\ &= \frac{4|\tilde{S}_{xy}|^2 - 2\tilde{S}_{xx}(\tilde{S}_{yy} - p\tilde{S}_{xx}) - 2\tilde{S}_{xx}\sqrt{(\tilde{S}_{yy} - p\tilde{S}_{xx})^2 + 4p|\tilde{S}_{xy}|^2}}{2\sqrt{(\tilde{S}_{yy} - p\tilde{S}_{xx})^2 + 4p|\tilde{S}_{xy}|^2}} \\ &= \frac{\tilde{S}_{xx}(p\tilde{S}_{xx} - \tilde{S}_{yy}) + 2|\tilde{S}_{xy}|^2 - \tilde{S}_{xx}\sqrt{(p\tilde{S}_{xx} - \tilde{S}_{yy})^2 + 4p|\tilde{S}_{xy}|^2}}{\sqrt{(p\tilde{S}_{xx} - \tilde{S}_{yy})^2 + 4p|\tilde{S}_{xy}|^2}} \end{aligned}$$

$$\begin{aligned} \frac{d}{dp} \left[ p\tilde{S}_{yy} - \tilde{S}_{xx} + \sqrt{(p\tilde{S}_{yy} - \tilde{S}_{xx})^2 + 4p|\tilde{S}_{xy}|^2} \right] &= \tilde{S}_{yy} + \frac{2\tilde{S}_{yy}(p\tilde{S}_{yy} - \tilde{S}_{xx}) + 4|\tilde{S}_{xy}|^2}{2\sqrt{(p\tilde{S}_{yy} - \tilde{S}_{xx})^2 + 4p|\tilde{S}_{xy}|^2}} \\ &= \frac{4|\tilde{S}_{xy}|^2 + 2\tilde{S}_{yy}(p\tilde{S}_{yy} - \tilde{S}_{xx}) + 2\tilde{S}_{yy}\sqrt{(p\tilde{S}_{yy} - \tilde{S}_{xx})^2 + 4p|\tilde{S}_{xy}|^2}}{2\sqrt{(p\tilde{S}_{yy} - \tilde{S}_{xx})^2 + 4p|\tilde{S}_{xy}|^2}} \end{aligned}$$

<sup>15</sup> White et al. (2006) page 679 ⟨(6)⟩, Shin and Hammond (2008) page 294 ⟨(9.69)⟩

<sup>16</sup> 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)⟩

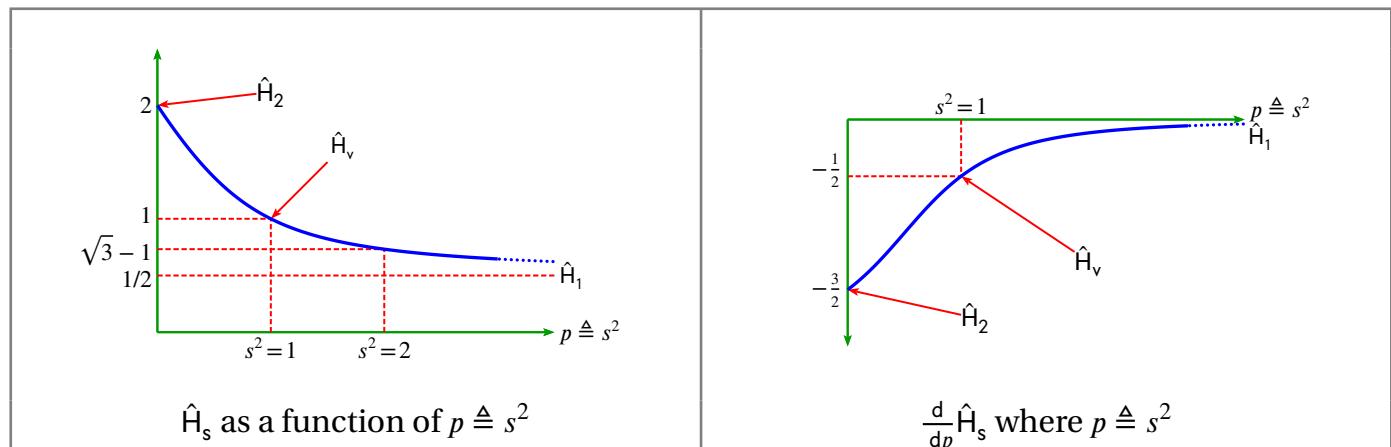
$$= \frac{\tilde{S}_{yy}(p\tilde{S}_{yy} - \tilde{S}_{xx}) + 2|\tilde{S}_{xy}|^2 + \tilde{S}_{yy}\sqrt{(p\tilde{S}_{yy} - \tilde{S}_{xx})^2 + 4p|\tilde{S}_{xy}|^2}}{\sqrt{(p\tilde{S}_{yy} - \tilde{S}_{xx})^2 + 4p|\tilde{S}_{xy}|^2}}$$

**Lemma 13.2.**

<b>L E M</b>	$\tilde{S}_{yy} - p\tilde{S}_{xx} + \sqrt{(\tilde{S}_{yy} - p\tilde{S}_{xx})^2 + 4p \tilde{S}_{xy} ^2} \geq 0$
	$p\tilde{S}_{yy} - \tilde{S}_{xx} + \sqrt{(p\tilde{S}_{yy} - \tilde{S}_{xx})^2 + 4p \tilde{S}_{xy} ^2} \geq 0$

PROOF:

$$\begin{aligned} & \tilde{S}_{yy} - p\tilde{S}_{xx} + \sqrt{(\tilde{S}_{yy} - p\tilde{S}_{xx})^2 + 4p|\tilde{S}_{xy}|^2} \geq 0 \\ \Leftrightarrow & \sqrt{(p\tilde{S}_{xx} - \tilde{S}_{yy})^2 + 4p|\tilde{S}_{xy}|^2} \geq p\tilde{S}_{xx} - \tilde{S}_{yy} \\ \Leftrightarrow & (p\tilde{S}_{xx} - \tilde{S}_{yy})^2 + 4p|\tilde{S}_{xy}|^2 \geq (p\tilde{S}_{xx} - \tilde{S}_{yy})^2 \\ \Leftrightarrow & 4p|\tilde{S}_{xy}|^2 \geq 0 \\ \Leftrightarrow & |\tilde{S}_{xy}| \geq 0 \\ \\ & p\tilde{S}_{yy} - \tilde{S}_{xx} + \sqrt{(p\tilde{S}_{yy} - \tilde{S}_{xx})^2 + 4p|\tilde{S}_{xy}|^2} \geq 0 \\ \Leftrightarrow & \sqrt{(\tilde{S}_{xx} - p\tilde{S}_{yy})^2 + 4p|\tilde{S}_{xy}|^2} \geq \tilde{S}_{xx} - p\tilde{S}_{yy} \\ \Leftrightarrow & (\tilde{S}_{xx} - p\tilde{S}_{yy})^2 + 4p|\tilde{S}_{xy}|^2 \geq (\tilde{S}_{xx} - p\tilde{S}_{yy})^2 \\ \Leftrightarrow & 4p|\tilde{S}_{xy}|^2 \geq 0 \\ \Leftrightarrow & |\tilde{S}_{xy}| \geq 0 \end{aligned}$$

Figure 13.2:  $\hat{H}_s$  with  $\tilde{S}_{xx} = \tilde{S}_{yy} = 1$  and  $\tilde{S}_{xy} = \frac{1}{2}$ **Theorem 13.1.** Let  $\hat{H}_s$  be defined as in Definition 13.8 (page 113).

<b>T H M</b>	$\{s_1 < s_2\} \implies  \hat{H}_s(\omega; s_2)  \leq \hat{H}_s(\omega; s_1) \quad (\hat{H}_s \text{ is monotonically decreasing in } s)$
	$ \hat{H}_1(\omega)  \leq  \hat{H}_s(\omega; s)  \leq  \hat{H}_2(\omega) $
	$\hat{H}_s(\omega; s) _{s=0} = \hat{H}_2(\omega)$
	$\hat{H}_s(\omega; s) _{s=1} = \hat{H}_v(\omega)$
	$\lim_{s \rightarrow \infty} \hat{H}_s(\omega; s) = \hat{H}_1(\omega)$

PROOF: I. Proofs for equalities:

$$\begin{aligned}
 \hat{H}_s(\omega; s) \Big|_{s=0} &\triangleq \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}} \Bigg|_{s=0} && \text{by def. of } \hat{H}_s && (\text{Definition 13.8 page 113}) \\
 &= \frac{\tilde{S}_{yy} - 0 + \sqrt{[\tilde{S}_{yy} - 0]^2 + 0}}{2\tilde{S}_{xy}} = \frac{\tilde{S}_{yy}}{\tilde{S}_{xy}} && && \\
 &\triangleq \hat{H}_2 && \text{by def. of } \hat{H}_2 && (\text{Definition 13.3 page 110}) \\
 \hat{H}_s(\omega; s) \Big|_{s=1} &\triangleq \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}} \Bigg|_{s=1} && \text{by def. of } \hat{H}_s && (\text{Definition 13.8 page 113}) \\
 &= \frac{\tilde{S}_{yy} - \tilde{S}_{xx} + \sqrt{[\tilde{S}_{yy} - \tilde{S}_{xx}]^2 + 4|\tilde{S}_{xy}|^2}}{2\tilde{S}_{xy}} && && \\
 &\triangleq \hat{H}_v && \text{by def. of } \hat{H}_v && (\text{Definition 13.9 page 114}) \\
 \lim_{s \rightarrow \infty} \hat{H}_s(\omega; s) &\triangleq \lim_{s \rightarrow \infty} \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}} && \text{by def. of } \hat{H}_s && (\text{Definition 13.8 page 113}) \\
 &\triangleq \lim_{p \rightarrow 0} \frac{\tilde{S}_{yy} - \frac{1}{p}\tilde{S}_{xx} + \sqrt{[\tilde{S}_{yy} - \frac{1}{p}\tilde{S}_{xx}]^2 + 4\frac{1}{p}|\tilde{S}_{xy}|^2}}{2\tilde{S}_{xy}} && && \\
 &= \lim_{p \rightarrow 0} \frac{p\tilde{S}_{yy} - \tilde{S}_{xx} + \sqrt{[p\tilde{S}_{yy} - \tilde{S}_{xx}]^2 + 4p|\tilde{S}_{xy}|^2}}{2p\tilde{S}_{xy}} && \text{where } p \triangleq \frac{1}{s^2} && \\
 &= \lim_{p \rightarrow 0} \frac{\frac{d}{dp} \left[ p\tilde{S}_{yy} - \tilde{S}_{xx} + \sqrt{[p\tilde{S}_{yy} - \tilde{S}_{xx}]^2 + 4p|\tilde{S}_{xy}|^2} \right]}{\frac{d}{dp} [2p\tilde{S}_{xy}]} && \text{by mult. by } 1 = \frac{p}{p} && \\
 &= \lim_{p \rightarrow 0} \frac{\tilde{S}_{yy}(p\tilde{S}_{yy} - \tilde{S}_{xx}) + 2|\tilde{S}_{xy}|^2 + \tilde{S}_{yy}\sqrt{(p\tilde{S}_{yy} - \tilde{S}_{xx})^2 + 4p|\tilde{S}_{xy}|^2}}{2\tilde{S}_{xy}} && \text{by l'Hôpital's rule} && \\
 &= \lim_{p \rightarrow 0} \frac{\tilde{S}_{yy}(p\tilde{S}_{yy} - \tilde{S}_{xx}) + 2|\tilde{S}_{xy}|^2 + \tilde{S}_{yy}\sqrt{(p\tilde{S}_{yy} - \tilde{S}_{xx})^2 + 4p|\tilde{S}_{xy}|^2}}{2\tilde{S}_{xy}} && \text{by Lemma 13.1 page 114} && \\
 &= \frac{\tilde{S}_{yy}(-\tilde{S}_{xx}) + 2|\tilde{S}_{xy}|^2 + \tilde{S}_{yy}\sqrt{(-\tilde{S}_{xx})^2}}{2\tilde{S}_{xy}\sqrt{(-\tilde{S}_{xx})^2}} && && \\
 &= \frac{2|\tilde{S}_{xy}|^2}{2\tilde{S}_{xx}\tilde{S}_{xy}} = \frac{\tilde{S}_{xy}^*}{\tilde{S}_{xx}} \triangleq \hat{H}_1 && \text{by def. of } \hat{H}_1 && (\text{Definition 13.2 page 109})
 \end{aligned}$$

## II. Proof for monotonicity:

1. Let  $p \triangleq s^2$ 

2. lemma:

$$\begin{aligned}
& \left[ 2|\tilde{S}_{xy}|^2 + \tilde{S}_{xx}(p\tilde{S}_{xx} - \tilde{S}_{yy}) \right]^2 \\
&= 4|\tilde{S}_{xy}|^4 + 4\tilde{S}_{xx}|\tilde{S}_{xy}|^2(p\tilde{S}_{xx} - \tilde{S}_{yy}) + \tilde{S}_{xx}^2(p\tilde{S}_{xx} - \tilde{S}_{yy})^2 \\
&\leq 4|\tilde{S}_{xy}|^2\tilde{S}_{xx}\tilde{S}_{yy} + 4p\tilde{S}_{xx}|\tilde{S}_{xy}|^2\tilde{S}_{xx} - 4\tilde{S}_{xx}|\tilde{S}_{xy}|^2\tilde{S}_{yy} + \tilde{S}_{xx}^2(p\tilde{S}_{xx} - \tilde{S}_{yy})^2 \quad \left( \begin{array}{l} \text{by Cauchy Schwartz Inequality} \\ (\text{Theorem K.2 page 260}) \end{array} \right) \\
&= 4\cancel{\tilde{S}_{xx}}\cancel{\tilde{S}_{yy}}|\tilde{S}_{xy}|^2 + 4p\tilde{S}_{xx}|\tilde{S}_{xy}|^2\tilde{S}_{xx} - 4\cancel{\tilde{S}_{xx}}\cancel{\tilde{S}_{yy}}|\tilde{S}_{xy}|^2 + \tilde{S}_{xx}^2(p\tilde{S}_{xx} - \tilde{S}_{yy})^2 \\
&= \tilde{S}_{xx}^2[(p\tilde{S}_{xx} - \tilde{S}_{yy})^2 + 4p|\tilde{S}_{xy}|^2] \\
&= \left[ \tilde{S}_{xx}\sqrt{(p\tilde{S}_{xx} - \tilde{S}_{yy})^2 + 4p|\tilde{S}_{xy}|^2} \right]^2
\end{aligned}$$

3. lemma:  $2|\tilde{S}_{xy}|^2 + \tilde{S}_{xx}(p\tilde{S}_{xx} - \tilde{S}_{yy}) \leq \tilde{S}_{xx}\sqrt{(p\tilde{S}_{xx} - \tilde{S}_{yy})^2 + 4p|\tilde{S}_{xy}|^2}$ . Proof:

$$\begin{aligned}
& 2|\tilde{S}_{xy}|^2 + \tilde{S}_{xx}(p\tilde{S}_{xx} - \tilde{S}_{yy}) \leq \tilde{S}_{xx}\sqrt{(p\tilde{S}_{xx} - \tilde{S}_{yy})^2 + 4p|\tilde{S}_{xy}|^2} \\
&\Leftrightarrow [2|\tilde{S}_{xy}|^2 + \tilde{S}_{xx}(p\tilde{S}_{xx} - \tilde{S}_{yy})]^2 \leq \left[ \tilde{S}_{xx}\sqrt{(p\tilde{S}_{xx} - \tilde{S}_{yy})^2 + 4p|\tilde{S}_{xy}|^2} \right]^2 \quad \left( \begin{array}{l} \text{because } f(x) \triangleq x^2 \text{ is} \\ \text{strictly monotonic increasing} \end{array} \right)
\end{aligned}$$

The previous inequality is true by (2) lemma, so (3) lemma also true.

4. Proof that  $\frac{d}{dp}|\hat{H}_s| \leq 0$ :

$$\begin{aligned}
\frac{d}{dp}|\hat{H}_s| &\triangleq \frac{d}{dp} \left| \frac{\tilde{S}_{yy} - s^2\tilde{S}_{xx} + \sqrt{(\tilde{S}_{yy} - s^2\tilde{S}_{xx})^2 + 4s^2|\tilde{S}_{xy}|^2}}{2\tilde{S}_{xy}} \right| \quad \text{by def. of } \hat{H}_s \text{ (Definition 13.8 page 113)} \\
&\triangleq \frac{d}{dp} \left| \frac{\tilde{S}_{yy} - p\tilde{S}_{xx} + \sqrt{(\tilde{S}_{yy} - p\tilde{S}_{xx})^2 + 4p|\tilde{S}_{xy}|^2}}{2\tilde{S}_{xy}} \right| \quad \text{by definition of } p \text{ (item (1) page 117)} \\
&= \frac{d}{dp} \left| \frac{\tilde{S}_{yy} - p\tilde{S}_{xx} + \sqrt{(\tilde{S}_{yy} - p\tilde{S}_{xx})^2 + 4p|\tilde{S}_{xy}|^2}}{|2\tilde{S}_{xy}|} \right| \\
&= \frac{d}{dp} \left| \frac{\tilde{S}_{yy} - p\tilde{S}_{xx} + \sqrt{(\tilde{S}_{yy} - p\tilde{S}_{xx})^2 + 4p|\tilde{S}_{xy}|^2}}{2|\tilde{S}_{xy}|} \right| \quad \text{by Lemma 13.2 page 115} \\
&= \frac{\tilde{S}_{xx}(p\tilde{S}_{xx} - \tilde{S}_{yy}) + 2|\tilde{S}_{xy}|^2 - \tilde{S}_{xx}\sqrt{(p\tilde{S}_{xx} - \tilde{S}_{yy})^2 + 4p|\tilde{S}_{xy}|^2}}{\sqrt{(\tilde{S}_{yy} - p\tilde{S}_{xx})^2 + 4p|\tilde{S}_{xy}|^2}} \quad \text{by Lemma 13.1 page 114} \\
&= \frac{\tilde{S}_{xx}(p\tilde{S}_{xx} - \tilde{S}_{yy}) + 2|\tilde{S}_{xy}|^2 - \tilde{S}_{xx}\sqrt{(p\tilde{S}_{xx} - \tilde{S}_{yy})^2 + 4p|\tilde{S}_{xy}|^2}}{2|\tilde{S}_{xy}|} \\
&\leq 0 \quad \text{by (3) lemma}
\end{aligned}$$



**Theorem 13.2.** Let  $S$  be a system with input  $x(n)$  and output  $y(n)$ .

T	H	M	$ \hat{H}_1(\omega)  \leq  \hat{H}_{\text{hm}}(\omega)  \leq  \hat{H}_{\text{gm}}(\omega)  \leq  \hat{H}_{\text{am}}(\omega)  \leq  \hat{H}_2(\omega) $
---	---	---	---



PROOF:

1. lemma:  $\hat{H}_1(\omega) \leq \hat{H}_2(\omega)$ . Proof:

$$\begin{aligned}
 |\hat{H}_1| &\triangleq \left| \frac{\tilde{S}_{yx}(\omega)}{\tilde{S}_{xx}(\omega)} \right| && \text{by definition of } \hat{H}_1 && (\text{Definition 13.2 page 109}) \\
 &= \left| \frac{\langle y | x \rangle}{\|x\|^2} \right| = \frac{|\langle y | x \rangle|}{\|x\|^2} \\
 &\leq \frac{|\langle y | x \rangle|}{\|x\|^2} \left| \frac{\|x\| \|y\|}{\langle y | x \rangle} \right|^2 && \text{by Cauchy Schwartz Inequality} && \text{Theorem K.2 page 260} \\
 &= \frac{\|y\|^2}{|\langle y | x \rangle|} = \left| \frac{\|y\|^2}{\langle x | y \rangle} \right| = \left| \frac{\tilde{S}_{yy}(\omega)}{\tilde{S}_{xy}(\omega)} \right| \\
 &= |\hat{H}_2| && \text{by definition of } \hat{H}_2 && (\text{Definition 13.3 page 110})
 \end{aligned}$$

2. remainder of the proof:

$$\begin{aligned}
 |\hat{H}_1(\omega)| &= \min \{ \hat{H}_1(\omega), \hat{H}_2(\omega) \} && \text{by (1) lemma} \\
 &\leq |\hat{H}_{\text{hm}}(\omega)| && \text{by Corollary N.1 page 299} && \text{with } \lambda_1 = \lambda_2 = \frac{1}{2} \\
 &\leq |\hat{H}_{\text{gm}}(\omega)| && \text{by Corollary N.1 page 299} && \text{with } \lambda_1 = \lambda_2 = \frac{1}{2} \\
 &\leq |\hat{H}_{\text{am}}(\omega)| && \text{by Corollary N.1 page 299} && \text{with } \lambda_1 = \lambda_2 = \frac{1}{2} \\
 &\leq \max \{ \hat{H}_1(\omega), \hat{H}_2(\omega) \} && \text{by Corollary N.1 page 299} && \text{with } \lambda_1 = \lambda_2 = \frac{1}{2} \\
 &= |\hat{H}_2(\omega)| && \text{by (1) lemma}
 \end{aligned}$$



Theorem 13.2 (page 118) 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 13.5** (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  $z$ .

P	R	P	$\angle \hat{H}_1(\omega) = \angle \hat{H}_{\text{hm}}(\omega) = \angle \hat{H}_{\text{gm}}(\omega) = \angle \hat{H}_{\text{am}}(\omega) = \angle \hat{H}_2(\omega) = \angle \hat{H}_s(\omega) = \angle \hat{H}_v(\omega) = \angle \hat{H}_k(\omega)$
---	---	---	---



PROOF:



$$\begin{aligned}
 \angle \hat{H}_1 &\triangleq \angle \frac{\tilde{S}_{yx}}{\tilde{S}_{xx}} & = & = \angle \tilde{S}_{xy}^* = -\angle \tilde{S}_{xy} \\
 &\uparrow \text{(Definition 13.2 page 109)} & & \\
 \angle \hat{H}_{hm} &\triangleq \angle \frac{2\tilde{S}_{yy}\tilde{S}_{xy}^*}{\tilde{S}_{xx}\tilde{S}_{yy} + |\tilde{S}_{xy}|^2} & = & = \angle \tilde{S}_{xy}^* = -\angle \tilde{S}_{xy} \\
 &\uparrow \text{(Definition 13.6 page 112)} & & \\
 \angle \hat{H}_{gm} &\triangleq \angle \frac{\tilde{S}_{xy}^*}{|\tilde{S}_{xy}|} \sqrt{\frac{\tilde{S}_{yy}}{\tilde{S}_{xx}}} & = & = \angle \tilde{S}_{xy}^* = -\angle \tilde{S}_{xy} \\
 &\uparrow \text{(Definition 13.5 page 111)} & & \\
 \angle \hat{H}_{am} &\triangleq \angle \frac{|\tilde{S}_{xy}|^2 + \tilde{S}_{xx}\tilde{S}_{yy}}{2\tilde{S}_{xx}\tilde{S}_{xy}} & = \angle \frac{1}{\tilde{S}_{xy}} & = \angle \tilde{S}_{xy}^* = -\angle \tilde{S}_{xy} \\
 &\uparrow \text{(Definition 13.4 page 110)} & & \\
 \angle \hat{H}_2 &\triangleq \angle \frac{\tilde{S}_{yy}}{\tilde{S}_{xy}} & = \angle \frac{1}{\tilde{S}_{xy}} & = \angle \tilde{S}_{xy}^* = -\angle \tilde{S}_{xy} \\
 &\uparrow \text{(Definition 13.3 page 110)} & & \\
 \angle \hat{H}_s &\triangleq \angle \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}} & = \angle \frac{1}{\tilde{S}_{xy}} & = \angle \tilde{S}_{xy}^* = -\angle \tilde{S}_{xy} \\
 &\uparrow \text{(Definition 13.8 page 113)} & & \\
 \angle \hat{H}_v &\triangleq \angle \frac{\tilde{S}_{yy} - \tilde{S}_{xx} + \sqrt{[\tilde{S}_{yy} - \tilde{S}_{xx}]^2 + 4|\tilde{S}_{xy}|^2}}{2\tilde{S}_{xy}} & = \angle \frac{1}{\tilde{S}_{xy}} & = \angle \tilde{S}_{xy}^* = -\angle \tilde{S}_{xy} \\
 &\uparrow \text{(Definition 13.9 page 114)} & & \\
 \angle \hat{H}_\kappa &\triangleq \angle \frac{\tilde{S}_{yy} - \kappa\tilde{S}_{xx} + \sqrt{[\tilde{S}_{yy} - \kappa\tilde{S}_{xx}]^2 + 4\kappa|\tilde{S}_{xy}|^2}}{2\tilde{S}_{xy}} & = \angle \frac{1}{\tilde{S}_{xy}} & = \angle \tilde{S}_{xy}^* = -\angle \tilde{S}_{xy} \\
 &\uparrow \text{(Definition 13.7 page 113)} & & \\
 \angle C_{xy} &\triangleq \angle \frac{\tilde{S}_{xy}^*}{\sqrt{\tilde{S}_{xx}\tilde{S}_{yy}}} & = & = \angle \tilde{S}_{xy}^* = -\angle \tilde{S}_{xy} \\
 &\uparrow \text{(Definition 13.12 page 130)} & & \\
 \Rightarrow & & &
 \end{aligned}$$

## 13.5 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 travelling from the definition of  $\hat{H}_s$  on page 113, you won't be disappointed when arriving at the next proposition (Proposition 13.6 page 119). 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 13.1—that  $\lim_{s \rightarrow \infty} \hat{H}_s = \hat{H}_1$ .

**Proposition 13.6.** <sup>17</sup> Let  $\hat{H}_\kappa(\omega; \kappa)$  be defined as in Definition 13.7 (page 113).

<sup>17</sup>  Shin and Hammond (2008) page 293 ((9.67)),  Leclere et al. (2014) ((10)  $\kappa(f) = 1/s^2$  and  $x$  and  $y$  swapped)

P  
R  
P

$$\hat{H}_\kappa(\omega; s) = \frac{2\kappa(\omega)\tilde{S}_{yx}(\omega)}{\kappa(\omega)\tilde{S}_{xx}(\omega) - \tilde{S}_{yy}(\omega) + \sqrt{[\kappa(\omega)\tilde{S}_{xx}(\omega) - \tilde{S}_{yy}(\omega)]^2 + 4\kappa(\omega)|\tilde{S}_{xy}(\omega)|^2}}$$

$$= \frac{2\tilde{S}_{xy}^*}{\tilde{S}_{xx} - \frac{1}{\kappa(\omega)}\tilde{S}_{yy} + \sqrt{[\tilde{S}_{xx} - \frac{1}{\kappa(\omega)}\tilde{S}_{yy}]^2 + \frac{4}{\kappa(\omega)}|\tilde{S}_{xy}|^2}}$$

PROOF: We can transform  $\hat{H}_\kappa$  from that found in Definition 13.8 (page 113) to the forms in this proposition by the technique of “rationalizing the denominator”<sup>18</sup>

$$\begin{aligned} \hat{H}_\kappa &\triangleq \frac{\tilde{S}_{yy} - \kappa\tilde{S}_{xx} + \sqrt{[\tilde{S}_{yy} - \kappa\tilde{S}_{xx}]^2 + 4\kappa|\tilde{S}_{xy}|^2}}{2\tilde{S}_{xy}} && \text{by definition of } \hat{H}_\kappa \text{ (Definition 13.8 page 113)} \\ &= \frac{\left[ \tilde{S}_{yy} - \kappa\tilde{S}_{xx} + \sqrt{[\tilde{S}_{yy} - \kappa\tilde{S}_{xx}]^2 + 4\kappa|\tilde{S}_{xy}|^2} \right] \overbrace{\left[ \tilde{S}_{yy} - \kappa\tilde{S}_{xx} - \sqrt{[\tilde{S}_{yy} - \kappa\tilde{S}_{xx}]^2 + 4\kappa|\tilde{S}_{xy}|^2} \right]}^{\text{"rationalizing factor"}}}{2\tilde{S}_{xy} \underbrace{\left[ \tilde{S}_{yy} - \kappa\tilde{S}_{xx} - \sqrt{[\tilde{S}_{yy} - \kappa\tilde{S}_{xx}]^2 + 4\kappa|\tilde{S}_{xy}|^2} \right]}_{\text{"rationalizing factor}}} \\ &= \frac{[\tilde{S}_{yy} - \kappa\tilde{S}_{xx}]^2 - [\tilde{S}_{yy} - \kappa\tilde{S}_{xx}]^2 - 4\kappa|\tilde{S}_{xy}|^2}{2\tilde{S}_{xy} \left[ \tilde{S}_{yy} - \kappa\tilde{S}_{xx} - \sqrt{[\tilde{S}_{yy} - \kappa\tilde{S}_{xx}]^2 + 4\kappa|\tilde{S}_{xy}|^2} \right]} = \frac{-4\kappa|\tilde{S}_{xy}|^2}{2\tilde{S}_{xy} \left[ \tilde{S}_{yy} - \kappa\tilde{S}_{xx} - \sqrt{[\tilde{S}_{yy} - \kappa\tilde{S}_{xx}]^2 + 4\kappa|\tilde{S}_{xy}|^2} \right]} \\ &= \frac{2\kappa\tilde{S}_{xy}^*}{\kappa\tilde{S}_{xx} - \tilde{S}_{yy} + \sqrt{[\kappa\tilde{S}_{xx} - \tilde{S}_{yy}]^2 + 4\kappa|\tilde{S}_{xy}|^2}} = \frac{2\frac{\kappa}{\kappa}\tilde{S}_{xy}^*}{\frac{\kappa}{\kappa}\tilde{S}_{xx} - \frac{1}{\kappa}\tilde{S}_{yy} + \sqrt{\frac{1}{s^4}[\kappa\tilde{S}_{xx} - \tilde{S}_{yy}]^2 + \frac{4\kappa}{s^4}|\tilde{S}_{xy}|^2}} \\ &= \frac{2\tilde{S}_{xy}^*}{\tilde{S}_{xx} - \frac{1}{\kappa}\tilde{S}_{yy} + \sqrt{[\tilde{S}_{xx} - \frac{1}{\kappa}\tilde{S}_{yy}]^2 + \frac{4}{\kappa}|\tilde{S}_{xy}|^2}} \end{aligned}$$

Integrity check for  $s = 0$  and  $s \rightarrow \infty$  cases: Let  $p \triangleq \kappa$ .

$$\begin{aligned} \lim_{p \rightarrow \infty} \hat{H}_\kappa &= \lim_{p \rightarrow \infty} \frac{2\tilde{S}_{yx}}{\tilde{S}_{xx} - \frac{1}{p}\tilde{S}_{yy} + \sqrt{[\tilde{S}_{xx} - \frac{1}{p}\tilde{S}_{yy}]^2 + \frac{4}{p}|\tilde{S}_{xy}|^2}} = \frac{2\tilde{S}_{yx}}{\tilde{S}_{xx} + \sqrt{[\tilde{S}_{xx}]^2}} \\ &= \frac{\tilde{S}_{yx}}{\tilde{S}_{xx}} \triangleq \hat{H}_1 && \text{by def. of } \hat{H}_1 \text{ (Definition 13.2 page 109)} \end{aligned}$$

$$\begin{aligned} \lim_{p \rightarrow 0} \hat{H}_\kappa &= \lim_{p \rightarrow 0} \frac{2p\tilde{S}_{yx}}{p\tilde{S}_{xx} - \tilde{S}_{yy} + \sqrt{[p\tilde{S}_{xx} - \tilde{S}_{yy}]^2 + 4p|\tilde{S}_{xy}|^2}} \\ &= \lim_{p \rightarrow 0} \frac{\frac{d}{dp}(2p\tilde{S}_{yx})}{\frac{d}{dp}\left(p\tilde{S}_{xx} - \tilde{S}_{yy} + \sqrt{[p\tilde{S}_{xx} - \tilde{S}_{yy}]^2 + 4p|\tilde{S}_{xy}|^2}\right)} && \text{by l'Hôpital's rule} \end{aligned}$$

<sup>18</sup> 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 13.6 is being performed essentially in reverse...or rather “rationalizing the numerator”.



$$\begin{aligned}
 &= \lim_{p \rightarrow 0} -\frac{2\tilde{S}_{yx}}{\tilde{S}_{xx}(p\tilde{S}_{xx} - \tilde{S}_{yy}) + 2|\tilde{S}_{xy}|^2 + \tilde{S}_{xx}\sqrt{(p\tilde{S}_{xx} - \tilde{S}_{yy})^2 + 4p|\tilde{S}_{xy}|^2}} \\
 &\quad \text{by Lemma 13.1 page 114} \\
 &= \frac{2\tilde{S}_{yy}\tilde{S}_{yx}}{-\tilde{S}_{xx}\tilde{S}_{yy} + 2|\tilde{S}_{xy}|^2 + \tilde{S}_{xx}\tilde{S}_{yy}} = \frac{2\tilde{S}_{yy}\tilde{S}_{yx}}{2|\tilde{S}_{xy}|^2} = \frac{\tilde{S}_{yy}}{\tilde{S}_{xy}} \\
 &\triangleq \hat{H}_2 \quad \text{by def. of } \hat{H}_2 \text{ (Definition 13.3 page 110)}
 \end{aligned}$$

## 13.6 Least squares estimates of non-linear systems

“The legendary Hungarian mathematician John von Neumann once referred to 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...” <sup>19</sup>

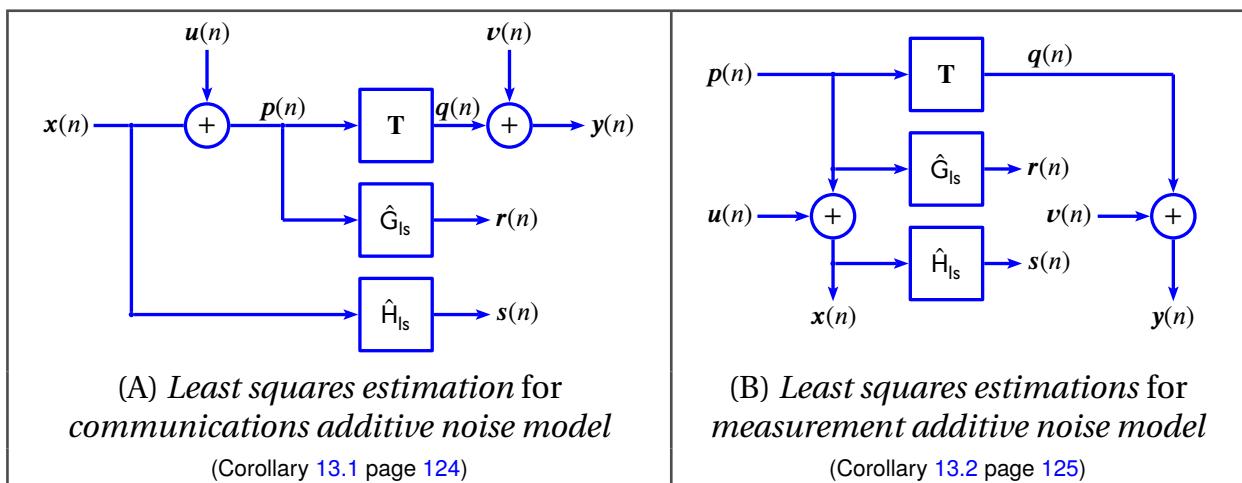
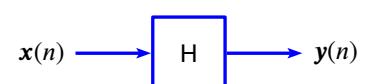


Figure 13.3: Least Square estimation (Theorem 13.3 page 122)

Let  $\mathbf{S}$  be the system illustrated to the right. If there is no measurement noise on the input and output and if  $\mathbf{H}$  is linear time invariant, then  $\tilde{\mathbf{H}} = \tilde{S}_{xy}/\tilde{S}_{xx}$  (Corollary 6.1 page 40). But what if there is output measurement noise? And what if  $\mathbf{H}$  is not LTI? What is the best least-squares estimate of  $\tilde{\mathbf{H}}$ ? The answer depends on how you define “the best”.



The definition of “best” or “optimal” is given by a cost function  $C(\hat{\mathbf{H}})$ . There are several possible cost functions. Definition 13.11 provides some. Theorem 13.3 then demonstrate optimal solutions with respect to these definitions.

**Definition 13.11.** Let  $\mathbf{S}$  be a system defined as in Figure 13.3 (page 121) (A) or (B). Define the following COST FUNCTIONS for spectral LEAST-SQUARES estimates:

DEF	$C_{rq}(\hat{\mathbf{G}}) \triangleq \tilde{\mathbf{F}} \ r(n) - q(n)\ ^2 \triangleq \tilde{\mathbf{F}} \langle r(n) - q(n)   r(0) - q(0) \rangle \triangleq \tilde{\mathbf{F}} \mathbf{E} ([r(n) - q(n)] [r(0) - q(0)]^*)$
DEF	$C_{sy}(\hat{\mathbf{H}}) \triangleq \tilde{\mathbf{F}} \ s(n) - y(n)\ ^2 \triangleq \tilde{\mathbf{F}} \langle s(n) - y(n)   s(0) - y(0) \rangle \triangleq \tilde{\mathbf{F}} \mathbf{E} ([s(n) - y(n)] [s(0) - y(0)]^*)$

<sup>19</sup> Bak (2013) page 29 *⟨§ Systems in Balance Are Not Complex⟩*

**Lemma 13.3.** Let  $C_{rq}(\hat{G})$  and  $C_{sy}(\hat{H})$  be defined as in Definition 13.11 (page 121).

<b>L E M</b>	$C_{rq}(\hat{G}) = \tilde{S}_{pp}(\omega)  \hat{G}(\omega) ^2 - \tilde{S}_{py}(\omega) \hat{G}(\omega) - \tilde{S}_{py}^*(\omega) \hat{G}^*(\omega) + \tilde{S}_{qq}(\omega)$ $C_{sy}(\hat{H}) = \tilde{S}_{xx}(\omega)  \hat{H}(\omega) ^2 - \tilde{S}_{xy}(\omega) \hat{H}(\omega) - \tilde{S}_{xy}^*(\omega) \hat{H}^*(\omega) + \tilde{S}_{yy}(\omega)$
----------------------	---

PROOF:

$$C_{rq}(\hat{G})$$

$$\begin{aligned} &\triangleq \check{\mathbf{F}}\mathbf{E}\left(\left[r(n) - q(n)\right]\left[r(0) - q(0)\right]^*\right) && \text{by definition of } C_{rq} \quad (\text{Definition 13.11 page 121}) \\ &= \check{\mathbf{F}}\left[\mathbf{E}[r(n)r^*(0)] - \mathbf{E}[r(n)q^*(0)] - \mathbf{E}[q(n)r^*(0)] + \mathbf{E}[q(n)q^*(0)]\right] && \text{by linearity of } \mathbf{E} \quad (\text{Theorem 1.1 page 4}) \\ &\triangleq \check{\mathbf{F}}[R_{rr}(m) - R_{rq}(m) - R_{qr}(m) + R_{qq}(m)] && \text{by definition of } R_{xy} \quad (\text{Definition 2.4 page 12}) \\ &\triangleq [\tilde{S}_{rr}(\omega) - \tilde{S}_{rq}(\omega) - \tilde{S}_{qr}(\omega) + \tilde{S}_{qq}(\omega)] && \text{by definition of } \tilde{S}_{xy} \quad (\text{Definition 7.3 page 48}) \\ &= \boxed{\tilde{S}_{pp}(\omega) |\hat{G}(\omega)|^2 - \tilde{S}_{py}(\omega) \hat{G}(\omega) - \tilde{S}_{py}^*(\omega) \hat{G}^*(\omega) + \tilde{S}_{qq}(\omega)} && \text{by (A)–(D) and Corollary 8.8 page 63} \end{aligned}$$

$$C_{sy}(\hat{H})$$

$$\begin{aligned} &\triangleq \check{\mathbf{F}}\mathbf{E}\left(\left[s(n) - y(n)\right]\left[s(0) - y(0)\right]^*\right) && \text{by definition of } C_{sy} \quad (\text{Definition 13.11 page 121}) \\ &= \check{\mathbf{F}}\left[\mathbf{E}[s(n)s^*(0)] - \mathbf{E}[s(n)y^*(0)] - \mathbf{E}[y(n)s^*(0)] + \mathbf{E}[y(n)y^*(0)]\right] && \text{by linearity of } \mathbf{E} \quad (\text{Theorem 1.1 page 4}) \\ &\triangleq \check{\mathbf{F}}[R_{ss}(m) - R_{sy}(m) - R_{ys}(m) + R_{yy}(m)] && \text{by definition of } R_{xy} \quad (\text{Definition 2.4 page 12}) \\ &\triangleq [\tilde{S}_{ss}(\omega) - \tilde{S}_{sy}(\omega) - \tilde{S}_{ys}(\omega) + \tilde{S}_{yy}(\omega)] && \text{by definition of } \tilde{S}_{xy} \quad (\text{Definition 7.3 page 48}) \\ &= \boxed{\tilde{S}_{xx}(\omega) |\hat{H}(\omega)|^2 - \tilde{S}_{xy}(\omega) \hat{H}(\omega) - \tilde{S}_{xy}^*(\omega) \hat{H}^*(\omega) + \tilde{S}_{yy}(\omega)} && \text{by (A)–(D) and Corollary 8.8 (page 63)} \end{aligned}$$



**Theorem 13.3.** Let  $\mathbf{S}$  be the system illustrated in Figure 13.3 page 121 (A) or (B).

<b>T H M</b>	$\left. \begin{array}{l} (A). \ x, u, \text{ and } v \text{ are WSS} \\ (B). \ x, u, \text{ and } v \text{ are UNCORRELATED} \\ (C). \ Eu = Ev = 0 \text{ (ZERO-MEAN)} \\ (D). \ \hat{G}_{ls} \text{ and } \hat{H}_{ls} \text{ are LTI} \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} (1). \ \arg \min_{\hat{G}} C_{rq}(\hat{G}) = \frac{\tilde{S}_{yp}(\omega)}{\tilde{S}_{pp}(\omega)} \\ (2). \ \arg \min_{\hat{H}} C_{sy}(\hat{H}) = \frac{\tilde{S}_{yx}(\omega)}{\tilde{S}_{xx}(\omega)} \end{array} \right.$
----------------------	---

PROOF:

1. Define shorthand function names  $\hat{G} \triangleq \hat{G}_{ls}$  and  $\hat{H} \triangleq \hat{H}_{ls}$ .

2. lemma:

$$0 = \frac{\partial}{\partial \hat{G}_R} C_{rq}(\hat{G}) = \frac{\partial}{\partial \hat{G}_R} \left( \tilde{S}_{pp} |\hat{G}|^2 - \hat{G} \tilde{S}_{py} - \hat{G}^* \tilde{S}_{py}^* + \tilde{S}_{qq} \right)$$

$$= \frac{\partial}{\partial \hat{G}_R} \left( \tilde{S}_{pp} [\hat{G}_R^2 + \hat{G}_I^2] - (\hat{G}_R + i\hat{G}_I) \tilde{S}_{py} - (\hat{G}_R + i\hat{G}_I)^* \tilde{S}_{py}^* + \tilde{S}_{qq} \right)$$

$$= 2\hat{G}_R \tilde{S}_{pp} - \tilde{S}_{py} - \tilde{S}_{py}^* + \frac{\partial}{\partial \hat{G}_R} \tilde{S}_{qq} \xrightarrow{0} \quad \text{because } q \text{ does not vary with } \hat{G}$$

$$= 2\hat{G}_R \tilde{S}_{pp} - 2\mathbf{R}_e \tilde{S}_{py}$$

$$= 2\hat{G}_R \tilde{S}_{pp} - 2\mathbf{R}_e \tilde{S}_{yp}$$

by Corollary 2.2 page 16

$$\Rightarrow \hat{G}_R(\omega) = \frac{\mathbf{R}_e \tilde{S}_{yp}(\omega)}{\tilde{S}_{pp}(\omega)}$$



3. lemma:

$$\begin{aligned}
 0 &= \frac{\partial}{\partial \hat{G}_I} C_{rq}(\hat{G}) \\
 &= \frac{\partial}{\partial \hat{G}_I} \left( \tilde{S}_{pp} |\hat{G}|^2 - \hat{G} \tilde{S}_{py} - \hat{G}^* \tilde{S}_{py}^* + \tilde{S}_{qq} \right) && \text{by Lemma 13.3 page 122} \\
 &= \frac{\partial}{\partial \hat{G}_I} [\tilde{S}_{pp} [\hat{G}_R^2 + \hat{G}_I^2] - (\hat{G}_R + i\hat{G}_I) \tilde{S}_{py} - (\hat{G}_R - i\hat{G}_I) \tilde{S}_{py}^* + \tilde{S}_{qq}] \\
 &= 2\hat{G}_I \tilde{S}_{pp} - i\tilde{S}_{py} + i\tilde{S}_{py}^* + \frac{\partial}{\partial \hat{G}_I} \tilde{S}_{qq} \xrightarrow{0} && \text{because } q \text{ does not vary with } \hat{G} \\
 &= 2\hat{G}_I \tilde{S}_{pp} - 2i(i\mathbf{I}_m \tilde{S}_{py}) \\
 &= 2\hat{G}_I \tilde{S}_{pp} + 2i(i\mathbf{I}_m \tilde{S}_{yp}) && \text{by Corollary 2.2 page 16} \\
 \implies \hat{G}_I(\omega) &= \boxed{\frac{\mathbf{I}_m \tilde{S}_{yp}(\omega)}{\tilde{S}_{pp}(\omega)}}
 \end{aligned}$$

4. Proof for  $\hat{G} \triangleq \hat{G}_{ls}$  expression:

$$\begin{aligned}
 \hat{G}(\omega) &= \hat{G}_R(\omega) + i\hat{G}_I(\omega) \\
 &= \frac{\mathbf{R}_e \tilde{S}_{yp}(\omega)}{\tilde{S}_{pp}(\omega)} + \frac{i\mathbf{I}_m \tilde{S}_{yp}(\omega)}{\tilde{S}_{pp}(\omega)} && \text{by (2) lemma and (3) lemma} \\
 &= \boxed{\frac{\tilde{S}_{yp}(\omega)}{\tilde{S}_{pp}(\omega)}}
 \end{aligned}$$

5. lemma:

$$\begin{aligned}
 0 &= \frac{\partial}{\partial \hat{H}_R} C_{sy}(\hat{H}) \\
 &= \frac{\partial}{\partial \hat{H}_R} \left( \tilde{S}_{xx} |\hat{H}|^2 - \hat{H} \tilde{S}_{xy} - \hat{H}^* \tilde{S}_{xy}^* + \tilde{S}_{yy} \right) && \text{by Lemma 13.3 page 122} \\
 &= \frac{\partial}{\partial \hat{H}_R} (\tilde{S}_{xx} [\hat{H}_R^2 + \hat{H}_I^2] - (\hat{H}_R + i\hat{H}_I) \tilde{S}_{xy} - (\hat{H}_R + i\hat{H}_I)^* \tilde{S}_{xy}^* + \tilde{S}_{yy}) \\
 &= 2\hat{H}_R \tilde{S}_{xx} - \tilde{S}_{xy} - \tilde{S}_{xy}^* + \frac{\partial}{\partial \hat{H}_R} \tilde{S}_{yy} \xrightarrow{0} && \text{because } y \text{ does not vary with } \hat{H} \\
 &= 2\hat{H}_R \tilde{S}_{xx} - 2\mathbf{R}_e \tilde{S}_{xy} \\
 &= 2\hat{H}_R \tilde{S}_{xx} - 2\mathbf{R}_e \tilde{S}_{yx} && \text{by Corollary 2.2 page 16} \\
 \implies \hat{H}_R(\omega) &= \boxed{\frac{\mathbf{R}_e \tilde{S}_{yx}(\omega)}{\tilde{S}_{xx}(\omega)}}
 \end{aligned}$$

6. lemma:

$$\begin{aligned}
 0 &= \frac{\partial}{\partial \hat{H}_I} C_{sy}(\hat{H}) \\
 &= \frac{\partial}{\partial \hat{H}_I} \left( \tilde{S}_{xx} |\hat{H}|^2 - \tilde{S}_{xy} \hat{H} - \tilde{S}_{xy}^* \hat{H}^* + \tilde{S}_{yy} \right) && \text{by Lemma 13.3 page 122} \\
 &= \frac{\partial}{\partial \hat{H}_I} [\tilde{S}_{xx} [\hat{H}_R^2 + \hat{H}_I^2] - \tilde{S}_{xy} (\hat{H}_R + i\hat{H}_I) - \tilde{S}_{xy}^* (\hat{H}_R - i\hat{H}_I) + \tilde{S}_{yy}] \\
 &= 2\hat{H}_I \tilde{S}_{xx} - i\tilde{S}_{xy} + i\tilde{S}_{xy}^* + \frac{\partial}{\partial \hat{H}_R} \tilde{S}_{yy} \xrightarrow{0} && \text{because } q \text{ does not vary with } \hat{H}
 \end{aligned}$$

$$\begin{aligned}
&= 2\hat{H}_I \tilde{S}_{xx} - 2i(i\mathbf{I}_m \tilde{S}_{xy}) \\
&= 2\hat{H}_I \tilde{S}_{xx} + 2i(i\mathbf{I}_m \tilde{S}_{yx}) \\
&= 2\hat{H}_I \tilde{S}_{xx} - 2\mathbf{I}_m \tilde{S}_{yx} \\
\implies \hat{H}_I(\omega) &= \frac{\mathbf{I}_m \tilde{S}_{yx}(\omega)}{\tilde{S}_{xx}(\omega)}
\end{aligned}$$

by Corollary 2.2 page 16

7. Proof for  $\hat{H} \triangleq \hat{H}_{ls}$  expression:

$$\begin{aligned}
\boxed{\hat{H}(\omega)} &= \hat{H}_R(\omega) + i\hat{H}_I(\omega) \\
&= \frac{\mathbf{R}_e \tilde{S}_{yp}(\omega)}{\tilde{S}_{xx}(\omega)} + \frac{i\mathbf{I}_m \tilde{S}_{yp}(\omega)}{\tilde{S}_{xx}(\omega)} \\
&= \frac{\mathbf{R}_e \tilde{S}_{yx}(\omega)}{\tilde{S}_{xx}(\omega)} + \frac{i\mathbf{I}_m \tilde{S}_{yx}(\omega)}{\tilde{S}_{xx}(\omega)} \\
\implies \hat{H}_{ls}(\omega) &= \frac{\tilde{S}_{yx}(\omega)}{\tilde{S}_{xx}(\omega)}
\end{aligned}$$

by (5) lemma and (6) lemma

by Theorem 8.4 page 59



Using Theorem 13.3 (previous) we can see that the optimal **least-squares** operators  $\hat{G}_{ls}$  and  $\hat{H}_{ls}$  for the **non-linear** operator  $\mathbf{T}$  in Figure 13.3 (page 121) (A) and (B) are (next two corollaries)

(1). $\hat{G}_{ls}(\omega) = \frac{\tilde{S}_{yx}(\omega)}{\tilde{S}_{xx}(\omega) + \tilde{S}_{uu}(\omega)}$ for (A)—communication system (2). $\hat{G}_{ls}(\omega) = \frac{\tilde{S}_{yx}(\omega)}{\tilde{S}_{xx}(\omega) - \tilde{S}_{uu}(\omega)}$ for (B)—measurement system (3). $\hat{H}_{ls}(\omega) = \frac{\tilde{S}_{yx}(\omega)}{\tilde{S}_{xx}(\omega)}$ for either (A) or (B)
---

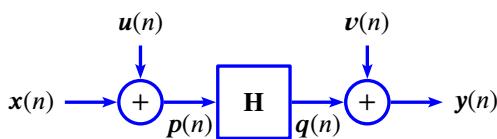
**Corollary 13.1.** Let  $\mathbf{S}$  be the system illustrated in Figure 13.3 page 121 (A).

<b>T H M</b>	$\left\{ \begin{array}{l} \text{hypotheses of Theorem 13.3} \\ \text{page 122} \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} (1). \arg \min_{\hat{G}_{ls}} C_{rq}(\hat{G}_{ls}) = \frac{\tilde{S}_{yx}(\omega)}{\tilde{S}_{xx}(\omega) + \tilde{S}_{uu}(\omega)} \\ (2). \arg \min_{\hat{H}_{ls}} C_{sy}(\hat{H}_{ls}) = \frac{\tilde{S}_{yx}(\omega)}{\tilde{S}_{xx}(\omega)} \end{array} \right\}$
--------------	--

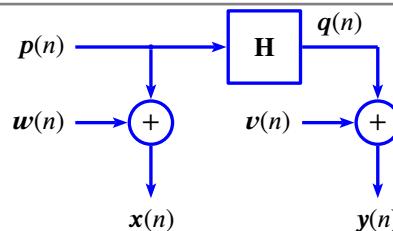
PROOF:

$$\begin{aligned}
\hat{G}_{ls} &= \frac{\tilde{S}_{yx}(\omega)}{\tilde{S}_{pp}(\omega)} && \text{by Theorem 13.3 page 122} \\
&= \frac{\tilde{S}_{yx}(\omega)}{\tilde{S}_{xx}(\omega) + \tilde{S}_{uu}(\omega)} && \text{by Theorem 8.1 page 55} \\
\hat{H}_{ls} &= \frac{\tilde{S}_{yx}(\omega)}{\tilde{S}_{xx}(\omega)} && \text{by Theorem 13.3 page 122}
\end{aligned}$$





(A) communications LTI additive noise model



(B) measurement LTI additive noise model

Figure 13.4: Additive noise systems with LTI operator  $\mathbf{H}$ 

**Corollary 13.2.** Let  $\mathbf{S}$  be the system illustrated in Figure 13.3 page 121 (B).

THM	$\left\{ \begin{array}{l} \text{hypotheses of Theorem 13.3} \\ \text{page 122} \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} (1). \quad \arg \min_{\hat{\mathbf{G}}_{ls}} C_{rq}(\hat{\mathbf{G}}_{ls}) = \frac{\tilde{S}_{yx}(\omega)}{\tilde{S}_{xx}(\omega) - \tilde{S}_{uu}(\omega)} \\ (2). \quad \arg \min_{\hat{\mathbf{H}}_{ls}} C_{sy}(\hat{\mathbf{H}}_{ls}) = \frac{\tilde{S}_{yx}(\omega)}{\tilde{S}_{xx}(\omega)} \end{array} \right\}$
-----	--

PROOF:

$$\begin{aligned} \hat{\mathbf{G}} &= \frac{\tilde{S}_{yx}(\omega)}{\tilde{S}_{pp}(\omega)} && \text{by Theorem 13.3 page 122} \\ &= \frac{\tilde{S}_{yx}(\omega)}{\tilde{S}_{xx}(\omega) - \tilde{S}_{uu}(\omega)} && \text{by Theorem 8.1 page 55} \\ \hat{\mathbf{H}} &= \frac{\tilde{S}_{yx}(\omega)}{\tilde{S}_{xx}(\omega)} && \text{by Theorem 13.3 page 122} \end{aligned}$$

It follows immediately from Corollary 13.1 (page 124) and Corollary 13.2 (page 125) that, in the special case of no input noise ( $u(n) = 0$ ),  $\hat{\mathbf{H}}_1$  is the optimal least-squares estimate of  $\tilde{\mathbf{H}}$  (next corollary).

**Corollary 13.3.**<sup>20</sup> Let  $\mathbf{S}$  be the system illustrated in Figure 13.3 page 121 (A) or (B).

COR	$\left\{ \begin{array}{l} (1). \quad \text{hypotheses of Theorem 13.3 and} \\ (2). \quad u(n) = 0 \end{array} \right\} \Rightarrow \left\{ \hat{\mathbf{G}}_{ls}(\omega) = \hat{\mathbf{H}}_{ls}(\omega) = \hat{\mathbf{H}}_1(\omega) \right\}$
-----	---

## 13.7 Least squares estimates of linear systems

The previous section did assume the estimates  $\hat{\mathbf{H}}_1$  and  $\hat{\mathbf{H}}_2$  to be *linear time invariant (LTI)*, but it did *not* assume that the system transfer function  $\mathbf{T}$  itself to be *LTI*. But making the LTI assumption on  $\mathbf{H}$  yields some interesting and insightful results, such as those in this section.

**Theorem 13.4** (Estimating  $\mathbf{H}$  in communication additive noise system). Let  $\mathbf{S}$  be the system illustrated in Figure 13.4 page 125 (A).

<sup>20</sup> Bendat and Piersol (1980) pages 98–100 (5.1.1 Optimal Character of Calculations; note: proof minimizing  $\tilde{S}_{vv}$  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)

THM

$$\left\{ \begin{array}{l} (A). \quad \mathbf{H} \text{ is} \\ (B). \quad \mathbf{x}(n) \text{ is} \\ (C). \quad \mathbf{x}(n), \mathbf{u}(n), \text{ and } \mathbf{v}(n) \text{ are} \end{array} \right. \begin{array}{l} \text{LINEAR TIME INVARIANT (LTI) and} \\ \text{WIDE-SENSE STATIONARY (WSS) and} \\ \text{UNCORRELATED} \end{array} \right\}$$

$$\Rightarrow \left\{ \begin{array}{l} (1). \quad \hat{\mathbf{H}}_1(\omega) = \tilde{\mathbf{H}}(\omega) \\ (2). \quad \hat{\mathbf{H}}_2(\omega) = \frac{\tilde{S}_{yy}(\omega)}{\tilde{H}^*(\omega)\tilde{S}_{xx}(\omega)} + \tilde{\mathbf{H}}(\omega) \left[ 1 + \frac{\tilde{S}_{uu}(\omega)}{\tilde{S}_{xx}(\omega)} \right] \end{array} \right. \text{and} \right\}$$

PROOF:

$$\begin{aligned} \hat{\mathbf{H}}_1(\omega) &\triangleq \frac{\tilde{S}_{yx}(\omega)}{\tilde{S}_{xx}(\omega)} \\ &= \frac{\tilde{\mathbf{H}}(\omega)\tilde{S}_{xx}(\omega)}{\tilde{S}_{xx}(\omega)} \\ &= \tilde{\mathbf{H}}(\omega) \end{aligned}$$

by definition of  $\hat{\mathbf{H}}_1$  (Definition 13.2 page 109)

$$\begin{aligned} \hat{\mathbf{H}}_2(\omega) &\triangleq \frac{\tilde{S}_{yy}(\omega)}{\tilde{S}_{xy}(\omega)} \\ &= \frac{\tilde{S}_{yy}(\omega)}{\tilde{\mathbf{H}}^*(\omega)\tilde{S}_{xx}(\omega)} \\ &= \frac{\tilde{S}_{vv}(\omega) + \tilde{S}_{qq}(\omega)}{\tilde{\mathbf{H}}^*(\omega)\tilde{S}_{xx}(\omega)} \\ &= \frac{\tilde{S}_{vv}(\omega) + \tilde{\mathbf{H}}^*(\omega)\tilde{\mathbf{H}}(\omega)\tilde{S}_{pp}(\omega)}{\tilde{\mathbf{H}}^*(\omega)\tilde{S}_{xx}(\omega)} \\ &= \frac{\tilde{S}_{vv}(\omega)}{\tilde{\mathbf{H}}^*(\omega)\tilde{S}_{xx}(\omega)} + \frac{\tilde{\mathbf{H}}^*(\omega)\tilde{\mathbf{H}}(\omega)[\tilde{S}_{xx}(\omega) + \tilde{S}_{uu}(\omega)]}{\tilde{\mathbf{H}}^*(\omega)\tilde{S}_{xx}(\omega)} \\ &= \frac{\tilde{S}_{vv}(\omega)}{\tilde{\mathbf{H}}^*(\omega)\tilde{S}_{xx}(\omega)} + \tilde{\mathbf{H}}(\omega) \left[ 1 + \frac{\tilde{S}_{uu}(\omega)}{\tilde{S}_{xx}(\omega)} \right] \end{aligned}$$

by Corollary 8.5 page 61

by definition of  $\hat{\mathbf{H}}_2$  (Definition 13.3 page 110)

by Corollary 8.5 page 61

by Theorem 8.1 page 55

by Corollary 6.3 page 41

⇒

**Theorem 13.5** (Estimating  $\mathbf{H}$  in measurement additive noise system). <sup>21</sup> Let  $\mathbf{S}$  be the system illustrated in Figure 13.4 page 125 (B).

$$\left\{ \begin{array}{l} (A). \quad \mathbf{H} \text{ is} \\ (B). \quad \mathbf{x}(n) \text{ is} \\ (C). \quad \mathbf{x}(n), \mathbf{u}(n), \text{ and } \mathbf{v}(n) \text{ are} \end{array} \right. \begin{array}{l} \text{LINEAR TIME INVARIANT and} \\ \text{WIDE-SENSE STATIONARY and} \\ \text{UNCORRELATED} \end{array} \right\}$$

$$\Rightarrow \left\{ \begin{array}{l} (1). \quad \hat{\mathbf{H}}_1(\omega) = \tilde{\mathbf{H}}(\omega) \left[ \frac{1}{1 + \frac{\tilde{S}_{ww}(\omega)}{\tilde{S}_{pp}(\omega)}} \right] \text{ (UNDER-ESTIMATED) and} \\ (2). \quad \hat{\mathbf{H}}_2(\omega) = \tilde{\mathbf{H}}(\omega) \left[ 1 + \frac{\tilde{S}_{vv}(\omega)}{\tilde{S}_{qq}(\omega)} \right] \text{ (OVER-ESTIMATED)} \end{array} \right\}$$

<sup>21</sup> Shin and Hammond (2008) page 294 ( $H_1(f) = H(f)$  (9.70);  $H_2(f) = H(f)(1 + S_{n_y n_y}(f)/S_{yy}(f))$  (9.71)), Shin and Hammond (2008) page 294 ( $H_1(f) = H(f)/(1 + S_{n_x n_x}/S_{xx}(f))$  (9.72);  $H_2(f) = H(f)$  (9.73)), Mitchell (1982) page 277 ( $H_1(f) = H_0(f)/(1 + G_{nn}/G_{uu})$ ) Mitchell (1982) page 278 ( $H_2(f) = H_0(f)(1 + G_{mm}/G_{vv})$ )



PROOF:

$$\begin{aligned}
 \hat{H}_1(\omega) &\triangleq \frac{\tilde{S}_{yx}(\omega)}{\tilde{S}_{xx}(\omega)} && \text{by definition of } \hat{H}_1 && (\text{Definition 13.2 page 109}) \\
 &= \frac{\tilde{S}_{pp}(\omega)\tilde{H}(\omega)}{\tilde{S}_{xx}(\omega)} && \text{by Corollary 8.4 page 61} \\
 &= \frac{\tilde{S}_{pp}(\omega)\tilde{H}(\omega)}{\tilde{S}_{pp}(\omega) + \tilde{S}_{ww}(\omega)} && \text{by hypothesis (A)} && \text{and Corollary 6.3 page 41} \\
 &= \tilde{H}(\omega) \left[ \frac{1}{1 + \frac{\tilde{S}_{ww}(\omega)}{\tilde{S}_{pp}(\omega)}} \right] \\
 \hat{H}_2(\omega) &\triangleq \frac{\tilde{S}_{yy}(\omega)}{\tilde{S}_{xy}(\omega)} && \text{by definition of } \hat{H}_2 && (\text{Definition 13.3 page 110}) \\
 &= \frac{\tilde{S}_{qq}(\omega) + \tilde{S}_{vv}(\omega)}{\tilde{S}_{xy}(\omega)} && \text{by hypothesis (C)} && \text{and Corollary 8.1 page 56} \\
 &= \frac{\tilde{S}_{qq}(\omega) + \tilde{S}_{vv}(\omega)}{\tilde{S}_{xq}(\omega)} && \text{by hypothesis (C)} && \text{and Theorem 8.4 page 59} \\
 &= \frac{\tilde{S}_{qq}(\omega) + \tilde{S}_{vv}(\omega)}{\tilde{S}_{pq}(\omega)} && \text{by hypothesis (C)} && \text{and Lemma 8.3 page 58} \\
 &= \frac{\tilde{S}_{qq}(\omega) + \tilde{S}_{vv}(\omega)}{\tilde{S}_{qq}(\omega)/\tilde{H}(\omega)} && \text{by LTI hypothesis (A)} && \text{and Corollary 6.3 page 41} \\
 &= \tilde{H}(\omega) \left[ 1 + \frac{\tilde{S}_{vv}(\omega)}{\tilde{S}_{qq}(\omega)} \right] && \text{by hypotheses (A) and (B)} && \text{and Corollary 6.3 page 41}
 \end{aligned}$$



**Corollary 13.4.** Let  $S$  be the system illustrated in Figure 13.4 (page 125).

COR	$\left\{ \begin{array}{l} (A). \text{ hypotheses of Theorem 13.5 and} \\ (B). u(n) = v(n) = 0 \quad (\text{NO INPUT NOISE}) \end{array} \right\} \implies \left\{ \begin{array}{l} \hat{H}_1(\omega) = \tilde{H}(\omega) \quad (\text{UNBIASED}) \end{array} \right\}$
	$\left\{ \begin{array}{l} (A). \text{ hypotheses of Theorem 13.5 and} \\ (B). v(n) = 0 \quad (\text{NO OUTPUT NOISE}) \end{array} \right\} \implies \left\{ \begin{array}{l} \hat{H}_2(\omega) = \tilde{H}(\omega) \quad (\text{UNBIASED}) \end{array} \right\}$

**Lemma 13.4.** Let  $S$  be the system illustrated in Figure 13.4 (page 125).

LEM	$\left\{ \begin{array}{l} \text{There exists } \kappa(\omega) \text{ such that } \tilde{S}_{vv}(\omega) = \kappa(\omega)\tilde{S}_{uu}(\omega) \end{array} \right\}$ $\implies \left\{ \begin{array}{l} \tilde{S}_{uu}(\omega) = \frac{ \hat{H}(\omega) ^2 \tilde{S}_{xx}(\omega) - \hat{H}(\omega)\tilde{S}_{xy}(\omega) - \hat{H}^*(\omega)\tilde{S}_{xy}^*(\omega) + \tilde{S}_{yy}(\omega)}{\kappa(\omega) +  \hat{H}(\omega) ^2} \end{array} \right\}$
-----	---

PROOF:

1. Development based on results of previous chapters:

$$\begin{aligned}
 \tilde{S}_{vv} &= \tilde{S}_{yy} - \tilde{S}_{qq} && \text{by Corollary 8.1 page 56} \\
 &= \tilde{S}_{yy} - \tilde{S}_{pq}\hat{H} && \text{by Corollary 6.3 page 41} \\
 &= \tilde{S}_{yy} - \tilde{S}_{xy}\hat{H} && \text{by Theorem 8.4 page 59} \\
 \tilde{S}_{uu} &= \tilde{S}_{xx} - \tilde{S}_{pp} && \text{by Corollary 8.1 page 56}
 \end{aligned}$$

$$\begin{aligned}
&= \tilde{S}_{xx} - \frac{\tilde{S}_{qp}}{\hat{H}} && \text{by Corollary 6.3 page 41} \\
&= \tilde{S}_{xx} - \frac{\tilde{S}_{yx}}{\hat{H}} && \text{by Theorem 8.4 page 59} \\
\tilde{S}_{uu} \left[ |\hat{H}|^2 + \kappa \right] &= |\hat{H}|^2 \tilde{S}_{uu} + \kappa \tilde{S}_{uu} \\
&\triangleq \tilde{S}_{uu} |\hat{H}|^2 + \tilde{S}_{vv} && \text{by definition of } \kappa(\omega) \\
&= |\hat{H}|^2 \left[ \tilde{S}_{xx} - \frac{\tilde{S}_{yx}}{\hat{H}} \right] + [\tilde{S}_{yy} - \tilde{S}_{xy} \hat{H}] \\
&= |\hat{H}|^2 \tilde{S}_{xx} - \hat{H}^* \tilde{S}_{yx} - \tilde{S}_{xy} \hat{H} + \tilde{S}_{yy} \\
\implies \tilde{S}_{uu}(\omega) &= \frac{|\hat{H}(\omega)|^2 \tilde{S}_{xx}(\omega) - \hat{H}(\omega) \tilde{S}_{xy}(\omega) - \hat{H}^*(\omega) \tilde{S}_{xy}^*(\omega) + \tilde{S}_{yy}(\omega)}{\kappa(\omega) + |\hat{H}(\omega)|^2}
\end{aligned}$$

2. Development of Wicks and Vold ([Wicks and Vold \(1986\)](#)):

$$\begin{aligned}
\tilde{Y} - \tilde{V} &= \tilde{Q} = \hat{H} \tilde{P} = \hat{H}(\tilde{X} - \tilde{U}) && \text{by definition of } \hat{H} \\
\hat{H}\tilde{U} - \tilde{V} &= \hat{H}\tilde{X} - \tilde{Y} && \text{by left distributive property (Theorem O.4 page 311)} \\
E([ \hat{H}\tilde{U} - \tilde{V} ] [ \hat{H}\tilde{U} - \tilde{V} ]^*) &= E([ \hat{H}\tilde{X} - \tilde{Y} ] [ \hat{H}\tilde{X} - \tilde{Y} ]^*) \\
|\hat{H}|^2 \tilde{S}_{uu} - \hat{H} \cancel{\tilde{S}_{uv}}^0 - \hat{H}^* \cancel{\tilde{S}_{vu}}^0 + \tilde{S}_{vv} &= |\hat{H}|^2 \tilde{S}_{xx} - \hat{H} \tilde{S}_{xy} - \hat{H}^* \tilde{S}_{xy}^* + \tilde{S}_{yy} && \text{because } u \text{ and } v \text{ are uncorrelated} \\
|\hat{H}|^2 \tilde{S}_{uu} + \kappa \tilde{S}_{uu} &= |\hat{H}|^2 \tilde{S}_{xx} - \hat{H} \tilde{S}_{xy} - \hat{H}^* \tilde{S}_{xy}^* + \tilde{S}_{yy} && \text{by hypothesis}
\end{aligned}$$

⇒

**Theorem 13.6.** [22](#) Let  $S$  be the system illustrated in Figure 13.4 (page 125). Let  $\hat{H}_k(\omega)$  be the transfer function estimate defined in Definition 13.7 (page 113).

T H M	$\left\{ \begin{array}{l} (1). \quad \text{There exists } \kappa(\omega) \text{ such that} \\ (2). \quad \tilde{S}_{vv}(\omega) = \kappa(\omega) \tilde{S}_{uu}(\omega) \end{array} \right. \text{ and } \right\} \implies \left\{ \begin{array}{l} \arg \min_{\hat{H}} C(\hat{H}) = \hat{H}_k(\omega) \\ (\hat{H}_k \text{ is the "optimal" estimator for minimizing system noise}) \end{array} \right\}$
-------------	--

PROOF:

1. Let  $F \triangleq |\hat{H}(\omega)|^2 \tilde{S}_{xx}(\omega) - \hat{H}(\omega) \tilde{S}_{xy}(\omega) - \hat{H}^*(\omega) \tilde{S}_{xy}^*(\omega) + \tilde{S}_{yy}(\omega)$  (numerator in Lemma 13.4) and  
 $G \triangleq \kappa(\omega) + |\hat{H}(\omega)|^2$  (denominator in Lemma 13.4)

2. lemma  $\left( \frac{\partial}{\partial \hat{H}_R} \tilde{S}_{uu} \right)$ :

$$\begin{aligned}
\boxed{0} &= \frac{1}{2} G^2 0 = \frac{1}{2} G^2 \frac{\partial}{\partial \hat{H}_R} \tilde{S}_{uu} && \text{set } \frac{\partial}{\partial \hat{H}_R} \tilde{S}_{uu} = 0 \text{ to find optimum } \hat{H}_R \\
&= \frac{1}{2} G^2 \frac{\partial}{\partial \hat{H}_R} \frac{F}{G} && \text{by Lemma 13.4 page 127} \\
&= \frac{1}{2} G^2 \frac{(F'G - G'F)}{G^2} && \text{by Quotient Rule} \\
&= \frac{1}{2}(F'G - G'F)
\end{aligned}$$

<sup>22</sup> [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)

$$\begin{aligned}
 &= \frac{1}{2} [2\hat{H}_R \tilde{S}_{xx} - \tilde{S}_{xy} - \tilde{S}_{xy}^*] G - \frac{1}{2} 2\hat{H}_R F \quad \text{by definition of } F, G \\
 &= \boxed{\hat{H}_R \tilde{S}_{xx} G - G \mathbf{R}_e \tilde{S}_{xy} - \hat{H}_R F}
 \end{aligned}$$

(item (1) page 128)

3. lemma  $\left(\frac{\partial}{\partial \hat{H}_I} \tilde{S}_{uu}\right)$ :

$$\begin{aligned}
 [0] &= \frac{1}{2} G^2 0 = \frac{1}{2} G^2 \frac{\partial}{\partial \hat{H}_I} \tilde{S}_{uu} && \text{set } \frac{\partial}{\partial \hat{H}_I} \tilde{S}_{uu} = 0 \text{ to find optimum } \hat{H}_I \\
 &= \frac{1}{2} G^2 0 = \frac{1}{2} G^2 \frac{\partial}{\partial \hat{H}_I} \frac{F}{G} && \text{by Lemma 13.4 page 127} \\
 &= \frac{1}{2} G^2 \frac{(F'G - G'F)}{G^2} && \text{by Quotient Rule} \\
 &= \frac{1}{2} (F'G - G'F) \\
 &= \frac{1}{2} [2\hat{H}_I \tilde{S}_{xx} - i\tilde{S}_{xy} + i\tilde{S}_{xy}^*] G - \frac{1}{2} 2\hat{H}_I F \quad \text{by definition of } F, G \\
 &= \boxed{\hat{H}_I \tilde{S}_{xx} G + G \mathbf{I}_m \tilde{S}_{xy} - \hat{H}_I F}
 \end{aligned}$$

(item (1) page 128)

4. Solve for  $\hat{H}$  ...

$$\begin{aligned}
 0 = 0 + i0 &= \frac{1}{2} G^2 0 + \frac{1}{2} G^2 0 = \frac{1}{2} G^2 \frac{\partial}{\partial \hat{H}_R} \tilde{S}_{uu} + i \frac{1}{2} G^2 \frac{\partial}{\partial \hat{H}_I} \tilde{S}_{uu} \\
 &= [\hat{H}_R \tilde{S}_{xx} G - G \mathbf{R}_e \tilde{S}_{xy} - \hat{H}_R F] + i[\hat{H}_I \tilde{S}_{xx} G + G \mathbf{I}_m \tilde{S}_{xy} - \hat{H}_I F] \quad \text{by (2) lemma and (3) lemma} \\
 &= \hat{H} \tilde{S}_{xx} G - \tilde{S}_{xy}^* G - \hat{H} F \quad \text{because } \mathbf{R}_e(z) + i\mathbf{I}_m(z) = z \text{ and } \mathbf{R}_e(z) - i\mathbf{I}_m(z) = z^* \\
 &= \hat{H} \tilde{S}_{xx} G - \tilde{S}_{yx} G - \hat{H} F \quad \text{by Corollary 2.2 page 16} \\
 &= \hat{H} \tilde{S}_{xx} (\kappa + |\hat{H}|^2) - \tilde{S}_{yx} (\kappa + |\hat{H}|^2) - \hat{H} (|\hat{H}|^2 \tilde{S}_{xx} - \hat{H} \tilde{S}_{xy} - \hat{H}^* \tilde{S}_{xy}^* + \tilde{S}_{yy}) \quad \text{by } F, G \text{ defs.} \\
 &= \hat{H} \tilde{S}_{xx} \left( \kappa + |\hat{H}|^2 \right) - \tilde{S}_{yx} \left( \kappa + |\hat{H}|^2 \right) - \hat{H} \left( |\hat{H}|^2 \tilde{S}_{xx} - \hat{H} \tilde{S}_{xy} - \hat{H}^* \tilde{S}_{xy}^* + \tilde{S}_{yy} \right) \\
 &= \kappa \hat{H} \tilde{S}_{xx} - \tilde{S}_{yx} \left( \kappa + |\hat{H}|^2 \right) + \left( \hat{H}^2 \tilde{S}_{xy} + |\hat{H}|^2 \tilde{S}_{xy}^* - \hat{H} \tilde{S}_{yy} \right) \\
 &= \kappa \hat{H} \tilde{S}_{xx} - \kappa \tilde{S}_{yx} - \tilde{S}_{yx} |\hat{H}|^2 + \left( \hat{H}^2 \tilde{S}_{xy} + |\hat{H}|^2 \tilde{S}_{xy}^* - \hat{H} \tilde{S}_{yy} \right) \\
 &= \hat{H}^2 \tilde{S}_{xy} + \hat{H} [\kappa \tilde{S}_{xx} - \tilde{S}_{yy}] - \kappa \tilde{S}_{xy}^* \\
 \Rightarrow \hat{H} &= \frac{(\tilde{S}_{yy} - \kappa \tilde{S}_{xx}) \pm \sqrt{(\tilde{S}_{yy} - \kappa \tilde{S}_{xx})^2 + 4\kappa |\tilde{S}_{xy}|^2}}{2\tilde{S}_{xy}}
 \end{aligned}$$

by Quadratic Equation



## 13.8 Coherence

### 13.8.1 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  $\hat{H}_1$  and  $\hat{H}_2$  (Definition 13.2 page 109, Definition 13.3 page 110). If the coherence is 0.70 or above, then we can have high confidence that the estimates  $\hat{H}_1$  and  $\hat{H}_2$  are “good” estimates.<sup>23</sup>

### 13.8.2 Definitions

**Definition 13.12.** <sup>24</sup> Let  $S$  be a system with input  $x(n)$  and output  $y(n)$ .

DEF

The **complex coherence function** is defined as  $C_{xy}(\omega) \triangleq \frac{\tilde{S}_{xy}^*(\omega)}{\sqrt{\tilde{S}_{xx}(\omega)\tilde{S}_{yy}(\omega)}}$

The **ordinary coherence function** is defined as  $\gamma_{xy}^2(\omega) \triangleq \frac{|\tilde{S}_{xy}(\omega)|^2}{\tilde{S}_{xx}(\omega)\tilde{S}_{yy}(\omega)}$

**Proposition 13.7.**

**P R P**  $\gamma_{xy}^2(\omega) = \frac{\hat{H}_1(\omega)}{\hat{H}_2(\omega)}$

PROOF:

$$\gamma_{xy}^2(\omega) \triangleq \frac{|\tilde{S}_{xy}|^2}{\tilde{S}_{xx}\tilde{S}_{yy}} \quad \text{by definition of } \gamma_{xy}^2 \quad (\text{Definition 13.12 page 130})$$

$$= \frac{\tilde{S}_{xy}^*/\tilde{S}_{xx}}{\tilde{S}_{yy}/\tilde{S}_{xy}} \triangleq \frac{\hat{H}_1}{\hat{H}_2} \quad \text{by definitions of } \hat{H}_1 \text{ and } \hat{H}_2 \quad (\text{Definition 13.2 page 109, Definition 13.3 page 110})$$

**Remark 13.2.** Note that the *complex transmissibility*  $\tilde{T}'_{xy}$  of Remark 13.1 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:

**R E M**  $C_{xy} \triangleq \sqrt{\frac{\hat{H}_1^*}{\hat{H}_2}}$  whereas  $\tilde{T}'_{xy} \triangleq \sqrt{\hat{H}_1 \hat{H}_2}$

PROOF:

$$\sqrt{\frac{\hat{H}_1^*(\omega)}{\hat{H}_2(\omega)}} \quad \text{by definition of } \hat{H}_{gm} \quad (\text{Definition 13.5 page 111})$$

### 13.8.3 A warning

Estimators yield, as the name implies, estimates. These estimates in general contain some error.

<sup>23</sup> Liang and Lee (2015) pages 363–365 (7.4.2 COHERENCE FUNCTION)

<sup>24</sup> 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))

*Example 13.1* (The K=1 Welch estimate of coherence). Suppose we have two *uncorrelated* stationary sequences  $x(n)$  and  $y(n)$ . Then, there CSD  $S_{xy}(\omega)$  should be 0 because

$$\begin{aligned} S_{xy}(\omega) &\triangleq \check{\mathbf{F}}\mathbf{E}_{xy}(m) \\ &= \check{\mathbf{F}}\mathbf{E}[x(n)y[n+m]] \\ &= \check{\mathbf{F}}[\mathbf{E}_x(n)][\mathbf{E}_y[n+m]] \\ &= \check{\mathbf{F}}[0][0] \\ &= 0 \end{aligned}$$

This will give a coherence of 0 also:

$$C(\omega) = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = 0$$

However, the Welch estimate with  $K = 1$  will yield

$$\begin{aligned} |C(\omega)| &= \left| \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} \right| \\ &= \left| \frac{(\tilde{\mathbf{F}}x)(\tilde{\mathbf{F}}y)^*}{\sqrt{|\tilde{\mathbf{F}}x|^2|\tilde{\mathbf{F}}y|^2}} \right| \\ &= 1 \end{aligned}$$

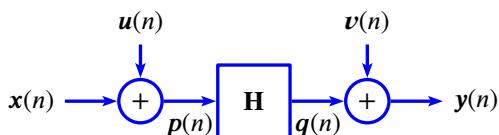


# CHAPTER 14

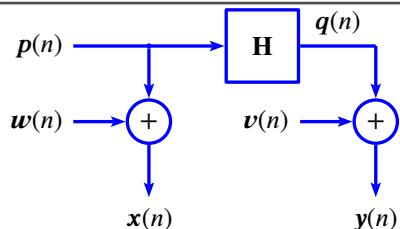
## NOISE ESTIMATION

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 14.1: Additive noise systems with LTI operator  $\mathbf{H}$



# CHAPTER 15

## MOMENT ESTIMATION

### 15.1 Mean Estimation

**Theorem 15.1.** Let  $\hat{\mu} \triangleq \sum_{n=1}^N \lambda_n x_n$  with  $\sum_{n=1}^N \lambda_n = 1$  be the ARITHMETIC MEAN (Definition N.4 page 299).

T  
H  
M

$$\left\{ \begin{array}{l} (A). (\{x_n\}) \text{ is WIDE SENSE STATIONARY} \\ (B). \mu \triangleq E x_n \\ (C). (\{x_n\}) \text{ is UNCORRELATED} \\ (D). \hat{\mu} \triangleq \sum_{n=1}^N \lambda_n x_n \quad (\text{ARITHMETIC MEAN}) \end{array} \right. \text{ and } \left\{ \begin{array}{l} (1). E \hat{\mu} = \mu \quad (\text{UNBIASED}) \text{ and} \\ (2). \text{var}(\hat{\mu}) = \sigma^2 \sum_{n=1}^N \lambda_n^2 \quad \text{and} \\ (3). \text{mse}(\hat{\mu}) = \sigma^2 \sum_{n=1}^N \lambda_n^2 \end{array} \right. \Rightarrow$$

PROOF:

$$\begin{aligned}
 E\hat{\mu} &\triangleq E \sum_{n \in \mathbb{Z}} \lambda_n x_n && \text{by definition of } \textit{arithmetic mean} \quad (\text{Definition N.4 page 299}) \\
 &= \sum_{n \in \mathbb{Z}} \lambda_n E x_n && \text{by } \textit{linearity of } E \quad (\text{Theorem 1.1 page 4}) \\
 &= \mu \sum_{n \in \mathbb{Z}} \lambda_n && \text{by } \textit{WSS hypothesis} \quad (\text{A}) \\
 &= \mu && \text{by } \sum \lambda_n = 1 \text{ hypothesis} \quad (\text{Definition N.4 page 299}) \\
 \text{var}(\hat{\mu}) &\triangleq E(\hat{\mu} - E\hat{\mu})^2 && \text{by definition of } \textit{variance} \\
 &= E(\hat{\mu} - \mu)^2 && \text{by previous result} \\
 &= E \left( \sum_{n=1}^N \lambda_n x_n - \mu \right)^2 && \text{by definition of } \hat{\mu} \\
 &= E \left[ \sum_{n=1}^N \lambda_n x_n - \mu \underbrace{\sum_{n=1}^N \lambda_n}_1 \right]^2 && \text{by } \sum \lambda_n = 1 \text{ hypothesis} \quad (\text{Definition N.4 page 299})
 \end{aligned}$$

$$\begin{aligned}
&= E \left[ \sum_{n=1}^N \lambda_n (x_n - \mu) \right]^2 \\
&= E \left[ \sum_{n=1}^N \lambda_n (x_n - \mu) \sum_{m=1}^N \lambda_m (x_m - \mu) \right] \\
&= \sum_{n=1}^N \sum_{m=1}^N \lambda_n \lambda_m (E[(x_n - \mu)(x_m - \mu)]) \\
&= \sum_{n=1}^N \sum_{m=1}^N \lambda_n \lambda_m (E[x_n x_m] - \mu E[x_n] - \mu E[x_m] + \mu^2) \\
&= \sum_{n=1}^N \sum_{m=1}^N \lambda_n \lambda_m (E[x_n x_m] - \mu^2 - \mu^2 + \mu^2) \\
&= \sum_{n=1}^N \sum_{m=1}^N \lambda_n \lambda_m (E[x_n x_m] - \mu^2) \\
&= \sum_{n=1}^N \lambda_n^2 (E[x_n^2] - \mu^2) + \sum_{n=1}^N \sum_{m \neq n} \lambda_n \lambda_m (E[x_n x_m] - \mu^2) \\
&= \sum_{n=1}^N \lambda_n^2 (E[x_n^2] - \mu^2) + \sum_{n=1}^N \sum_{m \neq n} \lambda_n \lambda_m (E x_n E x_m - \mu^2) \\
&= \sum_{n=1}^N \lambda_n^2 \sigma^2 + \sum_{n=1}^N \sum_{m \neq n} \lambda_n \lambda_m (\mu \mu - \mu^2) \quad \text{by WSS hypothesis} \\
&= \sigma^2 \sum_{n=1}^N \lambda_n^2
\end{aligned} \tag{A}$$

$$\text{mse}(\hat{\mu}) = E(\hat{\mu} - E\hat{\mu})^2 + (E\hat{\mu} - \mu)^2 \quad \text{by Theorem 9.2 page 71}$$

$$\begin{aligned}
&= \sigma^2 \sum_{n=1}^N \lambda_n^2 + (\mu - \mu)^2 \quad \text{by previous results} \\
&= \sigma^2 \sum_{n=1}^N \lambda_n^2
\end{aligned}$$

⇒

### Definition 15.1.

**D E F** The **average**  $\hat{\mu}$  of a length  $N$  sequence  $(x_n)_1^N$  is defined as  $\hat{\mu} \triangleq \frac{1}{N} \sum_{n=1}^N x_n$

### Corollary 15.1.<sup>1</sup>

<b>C O R</b>	$ \left\{ \begin{array}{l} (A). \quad (x_n) \text{ is WIDE SENSE STATIONARY} \quad \text{and} \\ (B). \quad \mu \triangleq E x_n \quad \text{and} \\ (C). \quad (x_n) \text{ is UNCORRELATED} \quad \text{and} \\ (D). \quad \hat{\mu} \triangleq \frac{1}{N} \sum_{n=1}^N x_n \quad (\text{AVERAGE}) \end{array} \right. $	$ \Rightarrow \left\{ \begin{array}{l} (1). \quad E \hat{\mu} = \mu \quad (\text{UNBIASED}) \quad \text{and} \\ (2). \quad \text{var}(\hat{\mu}) = \frac{\sigma^2}{N} \quad \text{and} \\ (3). \quad \text{mse}(\hat{\mu}) = \frac{\sigma^2}{N} \quad (\text{CONSISTENT}) \end{array} \right. $
--------------	---	---

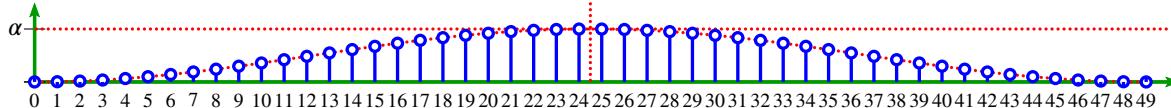
<sup>1</sup>  Kay (1988) page 45 (§“3.3 ESTIMATION THEORY”),  CLARKSON (1993) PAGES 52–54 (§“2.6.2 ESTIMATION OF MOMENTS — TIME AVERAGES”),

PROOF: These results follow from Theorem 15.1 (page 135) with  $\lambda_n = \frac{1}{N}$ .

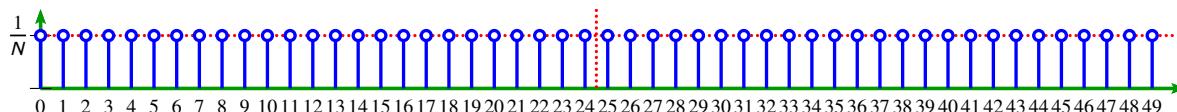
The *arithmetic mean* estimator  $\hat{\mu} \triangleq \sum \lambda_n x_n$  is *unbiased* and *consistent* for any  $\sum \lambda_n = 1$  and yields *mean square error*  $\text{mse}(\hat{\mu}) = \sigma^2 \sum \lambda_n^2$  (Theorem 15.1 page 135). But...

1. Said qualitatively: "What is the 'best' sequence  $(\lambda_n)$  to use?"
2. Said quantitatively: "What sequence  $(\lambda_n)$  yields the smallest  $\text{mse}(\hat{\mu})$ ?"

For example, would fashioning  $(\lambda_n)$  to be a scaled version of a standard window function, like the *Hanning window*<sup>2</sup> illustrated below, yield the best  $\text{mse}(\hat{\mu})$ ?



Theorem 15.2 (page 138) answers question (2) stating that the best sequence in terms of minimal  $\text{mse}$  is the sequence  $(\lambda_n) \triangleq \frac{1}{N} (\dots, 1, 1, 1, \dots)$ , which is the *average* estimator, which yields  $\text{mse}(\hat{\mu}) = \frac{\sigma^2}{N}$  (Corollary 15.1 page 136).



That is, it turns out that  $\frac{1}{N} \leq \sum \lambda_n^2$  for all possible sequences  $(\lambda_n)$ . This fact is demonstrated by Lemma 15.1 (next), which in turn follows more or less directly from the ubiquitous *Cauchy-Schwarz Inequality* (Theorem N.6 page 302, Theorem K.2 page 260).

Even further strengthening the average as choice estimator is Corollary 16.1 (page 143) which demonstrates that in the case where  $(x_n)$  is *uncorrelated* and *Gaussian*, then the optimal maximum likelihood estimator is the average.

### Lemma 15.1.

LEM	$\left\{ \sum_{n=1}^N \lambda_n = 1 \right\}$	$\Rightarrow$	$\left\{ \frac{1}{N} \leq \sum_{n=1}^N \lambda_n^2 \right\}$
-----	---	---------------	--

PROOF:

1. Let the sequence  $(a_n)$  be defined as  $(a_n) \triangleq (\dots, 1, 1, 1, \dots)$
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  $\|a_n\| \triangleq \sqrt{\sum_{n=1}^N a_n^2}$
4. Proof of lemma:

$$\begin{aligned}
 \left[ \frac{1}{N} \right] &= \frac{1}{N} \left( \sum_{n=1}^N \lambda_n \right)^2 && \text{by } \sum_{n=1}^N \lambda_n = 1 \text{ hypothesis} \\
 &= \frac{1}{N} \left( \sum_{n=1}^N a_n \lambda_n \right)^2 && \text{by } (a_n) \triangleq (\dots, 1, 1, 1, \dots) \text{ definition} && \text{(definition 1 page 137)} \\
 &\leq \frac{1}{N} \left( \sum_{n=1}^N a_n^2 \right) \left( \sum_{n=1}^N \lambda_n^2 \right) && \text{by Cauchy-Schwartz inequality} && \text{(Theorem N.6 page 302)}
 \end{aligned}$$

<sup>2</sup> Abdaheer (2009), page 130

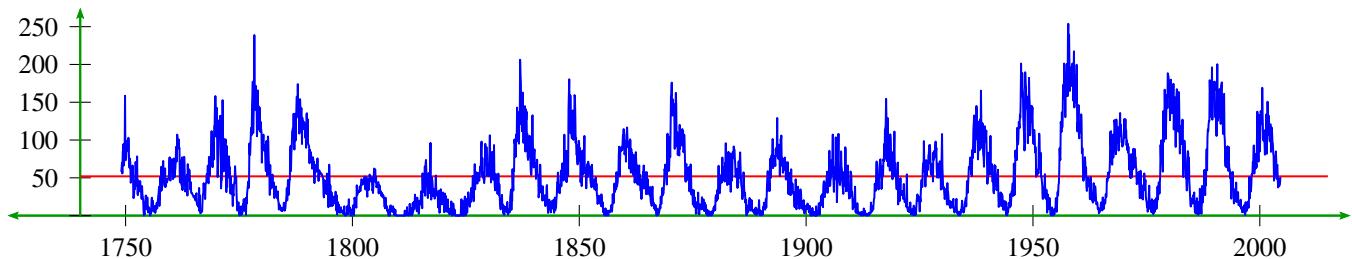


Figure 15.1: Sunspots data from 1749 into 2004

$$\begin{aligned} &\triangleq \frac{1}{N} \left( \sum_{n=1}^N 1^2 \right) \left( \sum_{n=1}^N \lambda_n^2 \right) \quad \text{by definition of } (\lambda_n) \\ &= \boxed{\sum_{n=1}^N \lambda_n^2} \end{aligned} \quad \text{(definition 1 page 137)}$$

⇒

**Theorem 15.2.** Let  $\text{mse}(\text{average mean})$  be the mean square error of the AVERAGE estimator (Corollary 15.1 page 136) and  $\text{mse}(\text{arithmetic mean})$  be the mean square error of the ARITHMETIC estimator (Theorem 15.1 page 135).

**P R P**  $\text{mse}(\text{average mean}) \leq \text{mse}(\text{arithmetic mean})$

PROOF:

$$\begin{aligned} \text{mse}(\text{average mean}) &= \sigma^2 \frac{1}{N} && \text{by Corollary 15.1 page 136} \\ &\leq \sigma^2 \sum_{n=1}^N \lambda_n^2 && \text{by Lemma 15.1 page 137} \\ &= \text{mse}(\text{arithmetic mean}) && \text{by Theorem 15.1 page 135} \end{aligned}$$

⇒

**Example 15.1.** The R code to the right calculates the *average* (Definition N.4 page 299) number of sunspots from 1749 into 2013 (Figure 15.1 page 138).

```

1 data(sunspots);
2 x    = sunspot.month;
3 N    = length(x);
4 avg = sum(x) / N;
5 print(avg);
6 plot(x, col="blue");
7 lines(c(start(x)[1], end(x)[1]), c(avg, avg), col="red");

```

## 15.2 Variance Estimation

If we know the true *mean*  $\mu$  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*:

$$\mathbb{E}\left[\frac{1}{N} \sum_{n=1}^N (x_n - \mu)^2\right] = \frac{1}{N} \sum_{n=1}^N \mathbb{E}[(x_n - \mu)^2] = \frac{1}{N} \sum_{n=1}^N \sigma^2 = \sigma^2$$

Very good. However, in many cases we don't know the **true mean**  $\mu$ , but rather only have an **estimated mean**  $\hat{\mu} \triangleq \frac{1}{N} = \sum_{n=1}^N x_n$ . In this case, substituting in the estimated mean for the true mean as in  $\hat{\text{var}}_B(\langle x_n \rangle)$  (next definition) yields a *biased* variance estimate (next theorem).

**Definition 15.2.** <sup>3</sup> Let  $\hat{\mu}$  be an estimate of the mean of a random sequence  $\langle x_n \rangle$ .

DEF

The **sample variance**  $\hat{\text{var}}(\langle x_n \rangle)$  is defined as

$$\hat{\text{var}}(\langle x_n \rangle) \triangleq \frac{1}{N-1} \sum_{n=1}^N (x_n - \hat{\mu})^2$$

The **biased sample variance**  $\hat{\text{var}}_B(\langle x_n \rangle)$  is here defined as

$$\hat{\text{var}}_B(\langle x_n \rangle) \triangleq \frac{1}{N} \sum_{n=1}^N (x_n - \hat{\mu})^2$$

The factor  $\frac{N}{N-1}$  such that  $\hat{\text{var}}(\langle x_n \rangle) = \frac{N}{N-1} \hat{\text{var}}_B(\langle x_n \rangle)$  is known as "*Bessel's correction*". Why such "correction" would be useful at all is demonstrated by Theorem 15.3 (next). Theorem 15.3 demonstrates that the *biased sample variance*  $\hat{\text{var}}_B(\langle x_n \rangle)$  is *biased*, and multiplication by  $\frac{N}{N-1}$  makes it *unbiased*.

**Theorem 15.3.** <sup>4</sup> Let  $\hat{\mu}$  be the AVERAGE (Definition N.4 page 299) of a sequence  $\langle x_n \rangle$ . Let  $\mu^4 \triangleq E[(x_n - \mu)^4]$  be the 4TH CENTRAL MOMENT of  $x_n$ .

THM

$$\left. \begin{array}{l} (A). \quad \langle x_n \rangle \text{ is WSS and} \\ (B). \quad \mu \triangleq E x_n \quad \text{and} \\ (C). \quad \langle x_n \rangle \text{ is UNCORRELATED} \end{array} \right\} \Rightarrow \left\{ \begin{array}{ll} (1). & E \hat{\text{var}}_B(x_n) = \frac{N-1}{N} \sigma^2 & \text{(BIASED)} \quad \text{and} \\ (2). & E \hat{\text{var}}(x_n) = \sigma^2 & \text{(UNBIASED)} \quad \text{and} \\ (3). & \text{var}[\hat{\text{var}}(x_n)] = \frac{1}{N} \left[ \mu^4 - \left( \frac{N-3}{N-1} \right) \sigma^4 \right] & \text{(CONSISTENT)} \end{array} \right.$$

PROOF:

1. lemma:  $E(x_n \hat{\mu}) = \frac{1}{N} \sigma^2 + \mu^2$ . Proof:

$$\begin{aligned} E(x_n \hat{\mu}) &\triangleq E\left(x_n \frac{1}{N} \sum_{m=1}^N x_m\right) && \text{by definition of average} && \text{(Definition N.4 page 299)} \\ &= E\left(\frac{1}{N} \sum_{m=1}^N x_n x_m\right) \\ &= \frac{1}{N} \sum_{m=1}^N E(x_n x_m) && \text{by linearity of } E && \text{(Theorem 1.1 page 4)} \\ &= \frac{1}{N} \left[ E x_n^2 + \sum_{m \neq n} E(x_n x_m) \right] \\ &= \frac{1}{N} \left[ E x_n^2 + \sum_{m \neq n} (E x_n)(E x_m) \right] && \text{by uncorrelated hypothesis} && \text{(C)} \\ &= \frac{1}{N} [(\sigma^2 + \mu^2) + (N-1)\mu^2] && \text{by Corollary 1.3 (page 5)} \\ &= \frac{1}{N} \sigma^2 + \mu^2 \end{aligned}$$

<sup>3</sup> 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 (???)

<sup>4</sup> 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):

$$\begin{aligned}
 \text{Evâr}_B((x_n)) &\triangleq E\left[\frac{1}{N} \sum_{n=1}^N (x_n - \hat{\mu})^2\right] && \text{by definition of } \text{vâr}_B \quad (\text{Definition 15.2 page 139}) \\
 &= \frac{1}{N} E\left[\sum_{n=1}^N \left(x_n - \underbrace{\mu + \mu - \hat{\mu}}_0\right)^2\right] \\
 &= \frac{1}{N} \sum_{n=1}^N \left[ E(x_n - \mu)^2 + 2E[(x_n - \mu)(\mu - \hat{\mu})] + E\left(\mu - \underbrace{\hat{\mu}}_{E\mu}\right)^2 \right] \\
 &= \frac{1}{N} \sum_{n=1}^N \left[ \sigma^2 + 2E[x_n \mu - x_n \hat{\mu} - \mu^2 + \mu \hat{\mu}] + \frac{1}{N} \sigma^2 \right] \\
 &= \frac{1}{N} \sum_{n=1}^N \left[ \sigma^2 + 2[\mu^2 - E(x_n \hat{\mu}) - \mu^2 + \mu^2] + \frac{1}{N} \sigma^2 \right] \quad \text{by Corollary 15.1 page 136} \\
 &= \frac{1}{N} \sum_{n=1}^N \left[ \sigma^2 + 2\left[\mu^2 - \left(\mu^2 + \frac{1}{N} \sigma^2\right)\right] + \frac{1}{N} \sigma^2 \right] \quad \text{by } \textit{unbiased prop. of } \hat{\mu} \quad ((1) \text{ lemma page 139}) \\
 &= \frac{1}{N} \sum_{n=1}^N \left[ \sigma^2 - \frac{1}{N} \sigma^2 \right] \\
 &= \frac{N-1}{N} \sigma^2
 \end{aligned}$$

3. Proof for (2):

$$\begin{aligned}
 \text{Evâr}((x_n)) &\triangleq E\left[\frac{1}{N-1} \sum_{n=1}^N (x_n - \hat{\mu})^2\right] && \text{by definition of } \text{vâr} \quad (\text{Definition 15.2 page 139}) \\
 &= \frac{N}{N-1} E\left[\frac{1}{N} \sum_{n=1}^{N-1} (x_n - \hat{\mu})^2\right] && \text{by } \textit{linearity of } E \quad (\text{Theorem 1.1 page 4}) \\
 &= \frac{N}{N-1} \left[ \frac{N-1}{N} \sigma^2 \right] && \text{by } \text{vâr}_B \text{ result} \quad (\text{item (2) page 140}) \\
 &= \sigma^2
 \end{aligned}$$

4. lemma:  $E[(\text{vâr})^2] = \frac{\mu^4}{N} + \frac{(N-1)^2+2}{N(N-1)}\sigma^4$ . 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.

$$\begin{aligned}
 E[(\text{vâr})^2] &\triangleq E\left[\left(\frac{1}{N-1} \sum_{n=1}^N (x_n - \hat{\mu})^2\right)^2\right] && \text{by definition of } \text{vâr} \quad (\text{Definition 15.2 page 139}) \\
 &= E\left[\left(\frac{1}{N-1} \sum_{n=1}^N (x_n - \hat{\mu})^2\right) \left(\frac{1}{N-1} \sum_{m=1}^N (x_m - \hat{\mu})^2\right)\right] \\
 &= \left(\frac{1}{N-1}\right)^2 E\left[\sum_{n=1}^N \sum_{m=1}^N (x_n - \hat{\mu})^2 (x_m - \hat{\mu})^2\right] \\
 &= \left(\frac{1}{N-1}\right)^2 E\left[\sum_{n=1}^N \sum_{m=1}^N (x_n^2 - 2x_n \hat{\mu} + \hat{\mu}^2) (x_m^2 - 2x_m \hat{\mu} + \hat{\mu}^2)\right]
 \end{aligned}$$



$$\begin{aligned}
&= \left( \frac{1}{N-1} \right)^2 E \left[ \sum_{n=1}^N \sum_{m=1}^N \left[ + (x_n^2 x_m^2 - 2\hat{\mu} x_n^2 x_m + \hat{\mu}^2 x_n^2) + (-2\hat{\mu} x_n x_m^2 + 4\hat{\mu}^2 x_n x_m - 2x_n \hat{\mu}^3) \right] \right] \\
&= \left( \frac{1}{N-1} \right)^2 \sum_{n=1}^N \sum_{m=1}^N \left[ + E[x_n^2 x_m^2] - 2E[\hat{\mu} x_n^2 x_m] - 2E[\hat{\mu} x_n x_m^2] + E[\hat{\mu}^2 x_n^2] + E[\hat{\mu}^2 x_m^2] \right. \\
&\quad \left. + 4E[\hat{\mu}^2 x_n x_m] - 2E[\hat{\mu}^3 x_n] - 2E[\hat{\mu}^3 x_m] + E[\hat{\mu}^4] \right] \\
&= \left( \frac{1}{N-1} \right)^2 \sum_{n=1}^N \sum_{m=1}^N \left[ + E[x_n^2 x_m^2] - 4E[\hat{\mu} x_n^2 x_m] + 2E[\hat{\mu}^2 x_n^2] \right. \\
&\quad \left. + 4E[\hat{\mu}^2 x_n x_m] - 4E[\hat{\mu}^3 x_n] + E[\hat{\mu}^4] \right] \\
&\stackrel{?}{=} \frac{\mu^4}{N} + \frac{(N-1)^2 + 2}{N(N-1)} \sigma^4
\end{aligned}$$

5. Proof for (3):

$$\begin{aligned}
\text{var}[\hat{\text{var}}(x_n)] &\triangleq E[(\hat{\text{var}} - E\hat{\text{var}})^2] && \text{by definition of } \hat{\text{var}} && (\text{Definition 15.2 page 139}) \\
&= E[(\hat{\text{var}} - \sigma^2)^2] && \text{by (2)} && (\text{item (3) page 140}) \\
&= E[(\hat{\text{var}})^2 - 2\sigma^2 \hat{\text{var}} + (\sigma^2)^2] && \text{by Binomial Theorem} \\
&= E[(\hat{\text{var}})^2] - 2\sigma^2 E[\hat{\text{var}}] + E[(\sigma^2)^2] && \text{by linearity of } E && (\text{Theorem 1.1 page 4}) \\
&= E[(\hat{\text{var}})^2] - 2(\sigma^2)^2 + (\sigma^2)^2 && \text{by (2)} && (\text{item (3) page 140}) \\
&= \left[ \frac{\mu^4}{N} + \left( \frac{(N-1)^2 + 2}{N(N-1)} \right) \sigma^4 \right] - \sigma^4 && \text{by (4) lemma} \\
&= \frac{1}{N} \left[ \mu^4 + \left( \frac{(N^2 - N + 1) + 2 - N(N-1)}{N-1} \right) \sigma^4 \right] \\
&= \frac{1}{N} \left[ \mu^4 - \left( \frac{N-3}{N-1} \right) \sigma^4 \right]
\end{aligned}$$



## 15.3 Estimates in terms of moment estimates

### Definition 15.3.

The **order- $k$  moment estimate** is here defined as

**DEF**

$$\hat{M}_k(x_n) \triangleq \frac{1}{N} \sum_{n=1}^N x_n^k$$

### Proposition 15.1.

P	$\hat{\mu}(x_n) = \hat{M}_1$
P	$\hat{\text{var}}(x_n) = \frac{N}{N-1} \hat{M}_2 - \frac{N}{N-1} \hat{M}_1^2$

PROOF:

$$\begin{aligned}\hat{\mu}(x_n) &\triangleq \frac{1}{N} \sum_{n=1}^N x_n && \text{by definition of } \hat{\mu} && (\text{Definition N.4 page 299}) \\ &\triangleq \hat{M}_1 && \text{by definition of } \hat{M}_1 && (\text{Definition 15.3 page 141})\end{aligned}$$

$$\begin{aligned}\hat{v}\hat{a}\hat{r}(x_n) &\triangleq \frac{1}{N-1} \sum_{n=1}^N (x_n - \hat{\mu})^2 && \text{by definition of } \hat{v}\hat{a}\hat{r} && (\text{Definition 15.2 page 139}) \\ &= \frac{1}{N-1} \sum_{n=1}^N (x_n^2 - 2x_n\hat{\mu} + (\hat{\mu})^2) \\ &= \frac{1}{N-1} \sum_{n=1}^N x_n^2 - 2\frac{1}{N-1} \sum_{n=1}^N x_n\hat{\mu} + \frac{1}{N-1} \sum_{n=1}^N (\hat{\mu})^2 \\ &= \frac{N}{N-1} \underbrace{\frac{1}{N} \sum_{n=1}^N x_n^2}_{\hat{M}_2} - 2\hat{\mu} \underbrace{\frac{N}{N-1} \frac{1}{N} \sum_{n=1}^N x_n}_{\hat{M}_1} + \frac{N}{N-1} \underbrace{(\hat{\mu})^2}_{\hat{M}_1^2} \\ &= \frac{N}{N-1} \hat{M}_2 - \frac{N}{N-1} \hat{M}_1^2\end{aligned}$$

⇒

## 15.4 Recursive forms

In software/firmware implementations, recursive forms are very useful and efficient.

**Proposition 15.2.** <sup>5</sup>

P	R	P	$\hat{\mu}_N = \underbrace{\hat{\mu}_{N-1}}_{\text{new}} + \underbrace{\frac{1}{N} \underbrace{[x_N - \hat{\mu}_{N-1}]}_{\text{error}}}_{\text{previous weight}}$
---	---	---	---

PROOF:

$$\begin{aligned}\hat{\mu}_N &\triangleq \frac{1}{N} \sum_{n=1}^N x_n && \text{by definition of } average && (\text{Definition N.4 page 299}) \\ &= \frac{1}{N} x_N + \frac{1}{N} \sum_{n=1}^{N-1} x_n \\ &= \frac{1}{N} x_N + \frac{N-1}{N} \left( \frac{1}{N-1} \right) \sum_{n=1}^{N-1} x_n \\ &\triangleq \frac{1}{N} x_N + \frac{N-1}{N} \hat{\mu}_{N-1} && \text{by definition of } average && (\text{Definition N.4 page 299}) \\ &= \frac{1}{N} x_N + \hat{\mu}_{N-1} - \frac{1}{N} \hat{\mu}_{N-1} \\ &= \hat{\mu}_{N-1} + \frac{1}{N} [x_N - \hat{\mu}_{N-1}]\end{aligned}$$

⇒

<sup>5</sup> Candy (2009) pages 11–12 (Example 1.3), Candy (2016) pages 12–13 (Example 1.3)



# CHAPTER 16

## DENSITY ESTIMATION

### 16.1 Introduction

Moment estimation is extremely useful for model-building when a *type* of parameterized *probability density function (pdf)*  $p_x(x)$  (Definition B.2 page 164) 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  $\mu$  (Definition C.1 page 169).
- The *Gamma Distribution* with parameter  $b$  (Definition C.3 page 171).
- The *Chi-square Distribution* with parameter variance  $\sigma^2$  (Definition C.5 page 172).
- The *Gaussian Distribution* with parameters mean  $\mu$  and variance  $\sigma^2$  (Definition C.2 page 170).
- The *Rayleigh Distribution* with parameters radius  $r$  and variance  $\sigma^2$  (Definition C.10 page 175).

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**.<sup>1</sup> If a distribution type is not known and not assumed, then the distribution itself must be estimated using **nonparametric density estimation** (CHAPTER 16 page 143)

### 16.2 Parametric density estimation

#### Corollary 16.1.

COR

$$\left\{ \begin{array}{l} (A). \quad (\mathbf{x}_n) \text{ is UNCORRELATED} \quad \text{and} \\ (B). \quad \mathbf{x}_n \text{ is GAUSSIAN} \end{array} \right\} \implies \left\{ \begin{array}{l} (1). \quad \hat{\mu}_{\text{ml}} = \frac{1}{N} \sum_{n=1}^N \mathbf{x}_n \quad \text{and} \\ (2). \quad \hat{\mu}_{\text{ml}} \text{ is CONSISTENT} \quad \text{and} \\ (3). \quad \hat{\mu}_{\text{ml}} \text{ is EFFICIENT} \end{array} \right\}$$

PROOF: This result follows directly from Theorem 11.9 (page 95) with

<sup>1</sup> Gramacki (2017) page 2 (§“1.1 Background”)

$$\begin{aligned}
 y(t) &\triangleq x(t; \theta) + v(t) && \text{where } v(t) \text{ is a zero-mean white Gaussian noise process} \\
 x(t; \theta) &\triangleq g(\theta) \\
 &\triangleq \theta \\
 &\triangleq \mu \\
 x_n &\triangleq \dot{y}_n \\
 &\triangleq \langle y(t) | \psi_n(t) \rangle \\
 &\triangleq \langle y(t) | \delta(t - n\tau) \rangle \\
 &\triangleq \int_{t \in \mathbb{R}} y(t) \delta(t - n\tau) dt \\
 &= y(n\tau)
 \end{aligned}$$

Alternatively, the results follow from Theorem 11.7 (page 91). 

## 16.3 Nonparametric density estimation

Some techniques of *nonparametric density estimation* include<sup>2</sup>

- ➊ Histogram
- ➋ Average Shifted Histogram
- ➌ Kernel Density Estimation (KDE)
- ➍ Fourier-based
- ➎ Wavelet-based
- ➏ Fast Gauss Transform

KDE's are conceptually quite simple, but do have some difficulties:

- ➊ KDEs are good for 6 dimensions or less—the “curse of dimensionality” (CHAPTER 19 page 149).
- ➋ KDEs have a *bandwidth* parameter and the computation of this parameter can be quite demanding.<sup>3</sup>

For cases where the dimension is 6 or greater, one option is to perform *data dimensionality reduction* (CHAPTER 19 page 149).

## 16.4 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 constraint is that the density estimate is not “too wiggly”.<sup>4</sup> Mathematically, this may be expressed as requiring the estimate to be a member of a **Sobolev Space**  $S$

$$S \triangleq \left\{ f \mid \int_{x \in \mathbb{R}} [f''(x)]^2 dx < r^2 \right\}$$

where  $r$  is some constant in  $\mathbb{R}$  or where  $r^2$  is  $\infty$ .

<sup>2</sup>  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”)

<sup>3</sup>  Gramacki (2017) page 4, 63–83 (Chapter 4 “Bandwidth Selectors for Kernel Density Estimation”)

<sup>4</sup>  Wasserman (2013) page 88 (“6.5 Example (Nonparametric density estimation).”)

# CHAPTER 17

## CORRELATION ESTIMATION

### 17.1 Time series method

**Definition 17.1.** <sup>1</sup>

The **windowed auto-correlation estimate**  $\hat{R}_{xx}(m)$  is defined as

$$\hat{R}_{xx}(m) \triangleq \frac{1}{N} \sum_{n=0}^{N-|m|} x(n)x(n+m)$$

**Theorem 17.1.** <sup>2</sup>

$$E[\hat{R}_{xx}(m)] = \left(1 - \frac{|m|}{N}\right) R_{xx}(m) \quad (\text{ASYMPTOTICALLY UNBIASED})$$

$$\text{var}[\hat{R}_{xx}(m)] = \frac{1}{N} \sum_{n \in \mathbb{Z}} [R_{xx}^2(n) + R_{xx}(n-m)R_{xx}(n+m)] \quad (\text{CONSISTENT})$$

### 17.2 Spectral methods

Here are two methods for estimating correlation using spectral methods:

1. Calculate the estimate  $\hat{H}(\omega)$  (CHAPTER 13 page 107) of the true system model  $H(\omega)$  and then calculate the *Inverse Fourier Transform* of  $\hat{H}(\omega)$ .
2. Calculate an *AR*, *MA*, or *ARMA* estimate of the true system model  $H(z)$  and then compute the *Inverse Z-Transform* of the estimate.<sup>3</sup>

<sup>1</sup> 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”)

<sup>2</sup> 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”)

<sup>3</sup> Clarkson (1993) pages 56–59 (§“Appendix 2A — Calculating the Correlation via Contour Integration”)



# CHAPTER 18

## 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*<sup>1</sup>

### 18.1 Nonparametric spectral estimation

Quality of spectral estimators<sup>2</sup>

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 18.1. <sup>3</sup> Using the PSD, we can estimate the periodicity of sunspots (Figure 18.1 page 148) as

<sup>1</sup>  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.

<sup>2</sup>  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

<sup>3</sup>  SILSO World Data Center (2019) ([http://www.sidc.be/silso/DATA/SN\\_m\\_tot\\_V2.0.txt](http://www.sidc.be/silso/DATA/SN_m_tot_V2.0.txt)),  rdocumentation.org (2013)

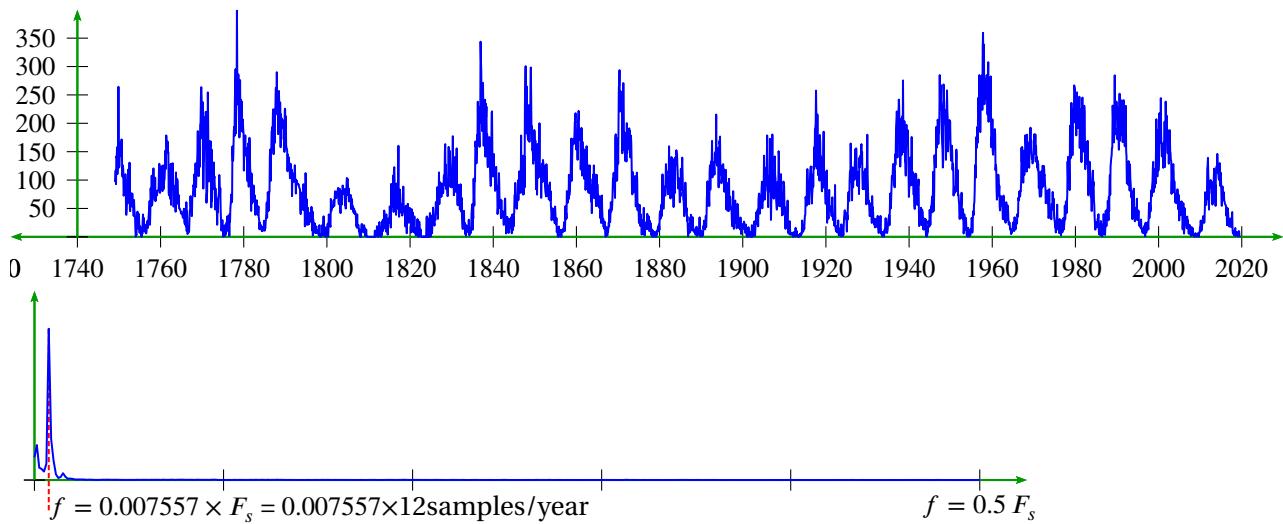


Figure 18.1: Estimation of sunspot periodicity using Welch PSD estimation

E X	number of segments = 2	estimated periodicty = 11.027777777777786 years
	number of segments = 3	estimated periodicty = 11.020833333333321 years
	number of segments = 4	estimated periodicty = 11.027777777777786 years
	number of segments = 8	estimated periodicty = 11.000000000000000 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:

```

1 require(stats);
2 require(graphics);
3 require(datasets);
4 require(ramify);
5 require(bspec);
6 data(sunspots, package="datasets");
7 mypsd = function(x, numSegments=4)
8 {
9   xts      = as.ts(as.vector(x)); # year indices seems to confuse welchPSD
10  N        = length(xts);          # length of time series
11  Fs       = 12;                  # sample rate = 12 samples per year
12  estMean  = mean(xts);          # estimated mean
13  segLength = N / numSegments;   # segment length
14  xpsd     = bspec::welchPSD(xts - estMean, seglength = segLength);
15  psdMax   = max(xpsd$power);    # maximum PSD value
16  binMax   = ramify::argmax(as.matrix(xpsd$power), rows = FALSE);
17  freqMax  = xpsd$frequency[binMax] * Fs; # dominate non-DC frequency
18  periodT  = 1 / freqMax;        # estimated period
19  plot(xpsd$power, type="b", xlim=c(1,50), ylim=c(0,psdMax), col="blue");
20  periodT;                      # return estimated period
21 }
22 estT = mypsd(x, numSegments=4);
23 print(sprintf("estimated period = %.16f years", estT));

```

## 18.2 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”)

# CHAPTER 19

## DIMENSIONALITY REDUCTION

A number of statistical estimation algorithms yield subspace structures with high dimensionality—one key example is *Kernel Density Estimation (KDE)*( CHAPTER 16 page 143). Richard Bellman described this problem as the “curse of dimensionality”.<sup>1</sup>

Actually, a *time series* is a point in infinite dimensional space.<sup>2</sup> Statistical analysis is much to do with approximating this point in low dimensional space.

*Example 19.1.* <sup>3</sup> An example of a one-dimensional representation is the *estimated mean* or average of the time series.

*Example 19.2.* <sup>4</sup> 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.

Here are some references about *dimensionality reduction*:

- Cunningham and Ghahramani (2015): “Linear Dimensionality Reduction:...”
- Sorzano et al. (2014): “A survey of dimensionality reduction techniques”
- Lee and Verleysen (2007): “Nonlinear Dimensionality Reduction”

A very common algorithm for dimensionality reduction is *Principal Component Analysis (PCA)*.<sup>5</sup>

*Example 19.3.* <sup>6</sup> 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 eigenvalues 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*.

<sup>1</sup> ■ 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”)

<sup>2</sup> ■ Wasserman (2006) page 1 (§“1.1 What Is Nonparametric Inference?”)

<sup>3</sup> ■ Wasserman (2013) pages 88–89 (“6.5 Example (Nonparametric estimation of functionals).”)

<sup>4</sup> ■ Wasserman (2013) page 88 (“6.2 Example (Two-dimensional Parametric Estimation)”)

<sup>5</sup> ■ Pearson (1901), ■ Hotelling (1933), ■ Eckart and Young (1936), ■ ?, pages 10–36 (“2. Component Analysis”), ■ Jeffers (1967), ■ Jolliffe (2013) (9781475719048)

<sup>6</sup> ■ Carmona (2013) pages 171–180 (§“3.5 PRINCIPAL COMPONENT ANALYSIS”)

*Remark 19.1.* Despite the description of *PCA* in Example 19.3, it is *not* to say that small eigenvalues are never important. For several counter-examples demonstrating the importance of small eigenvalues, see [Jolliffe \(1982\)](#).



# **Part V**

# **Appendices**



# APPENDIX A

## PROBABILITY SPACE

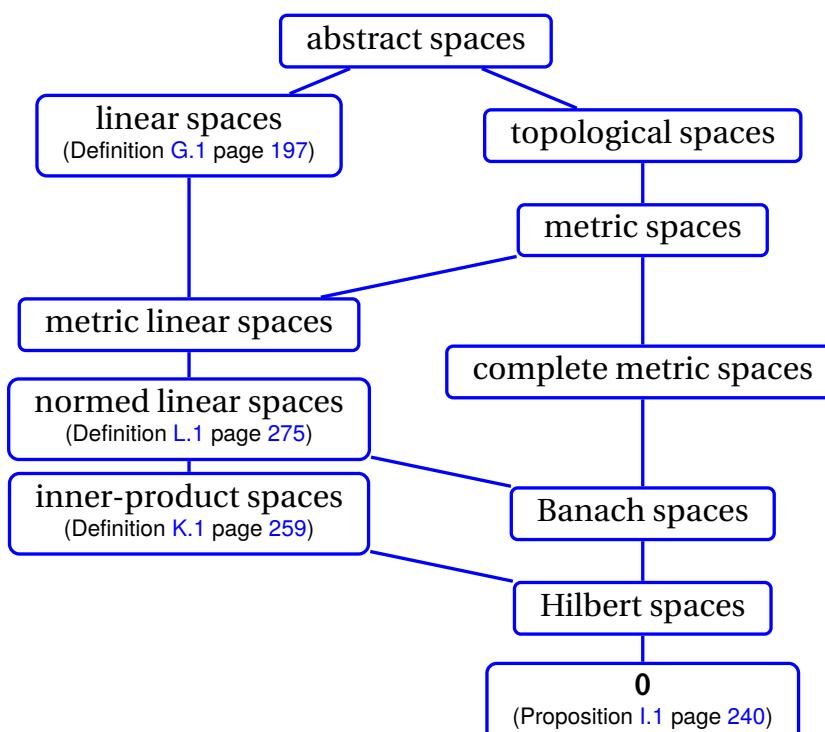


Figure A.1: Lattice of mathematical spaces



“It is not certain that everything is certain.”  
Blaise Pascal (1623–1662), mathematician <sup>1</sup>

<sup>1</sup> quote: [http://en.wikiquote.org/wiki/Blaise\\_Pascal](http://en.wikiquote.org/wiki/Blaise_Pascal)  
image: [http://en.wikipedia.org/wiki/Image:Blaise\\_pascal.jpg](http://en.wikipedia.org/wiki/Image:Blaise_pascal.jpg)

## A.1 Probability functions

**Definition A.1.** <sup>2</sup> Let  $(X, \vee, \wedge, \neg, 0, 1; \leq)$  be a LATTICE WITH NEGATION.

The function  $P$  is a **probability function** if

- |                      |  |
|----------------------|--|
| <b>D<br/>E<br/>F</b> | <ol style="list-style-type: none"> <li>(1). <math>P(1) = 1</math> (NORMALIZED) and</li> <li>(2). <math>P(x) \geq 0 \quad \forall x \in X</math> (NONNEGATIVE) and</li> <li>(3). <math>\bigwedge_{n=1}^{\infty} x_n = 0 \implies P\left(\bigvee_{n=1}^{\infty} x_n\right) = \sum_{n=1}^{\infty} P(x_n) \quad \forall x_n \in X</math> (<math>\sigma</math>-ADDITIVE) .</li> </ol> |
|----------------------|--|

*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  $\sigma$ -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.<sup>3</sup> In fact, Kolmogorov himself provided some argument against  $\sigma$ -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."<sup>4</sup>

There are several other definitions of probability that only require *additivity* rather than  $\sigma$ -additivity. On a *Boolean lattice*, the **traditional probability** function is defined as<sup>5</sup>

- |                      |   |
|----------------------|---|
| <b>D<br/>E<br/>F</b> | <ol style="list-style-type: none"> <li>(1). <math>P(1) = 1</math> (normalized) and</li> <li>(2). <math>P(x) \geq 0 \quad \forall x \in X</math> (nonnegative) and</li> <li>(3). <math>x \wedge y = 0 \implies P(x \vee y) = P(x) + P(y) \quad \forall x, y \in X</math> (additive) .</li> </ol> |
|----------------------|---|

This definition implies (on a *Boolean lattice*) that

- |                      |   |
|----------------------|---|
| <b>D<br/>E<br/>F</b> | <ol style="list-style-type: none"> <li>(a). <math>P(0) = 0</math> (nondegenerate) and</li> <li>(b). <math>P(x) \leq 1 \quad \forall x \in X</math> (upper bounded) and</li> <li>(c). <math>P(x) = 1 - P(x^\perp) \quad \forall x \in X</math> and</li> <li>(d). <math>P(x \vee y) \leq P(x) + P(y) \quad \forall x, y \in X</math> (subadditive) and</li> <li>(e). <math>P(x \vee y) = P(x) + P(y) - P(x \wedge y) \quad \forall x, y \in X</math> and</li> <li>(f). <math>x \leq y \implies P(x) \leq P(y) \quad \forall x, y \in X</math> (monotone) .</li> </ol> |
|----------------------|---|

On a *distributive pseudocomplemented lattice*, the **generalized probability** function has been defined as<sup>6</sup>

- |                      |   |
|----------------------|---|
| <b>D<br/>E<br/>F</b> | <ol style="list-style-type: none"> <li>(1). <math>P(0) = 0</math> (nondegenerate) and</li> <li>(2). <math>P(1) = 1</math> (normalized) and</li> <li>(3). <math>0 \leq P(1) \leq 1</math> and</li> <li>(4). <math>P(x \vee y) = P(x) + P(y) - P(x \wedge y) \quad \forall x, y \in X</math> .</li> </ol> |
|----------------------|---|

On an *orthomodular lattice*, or a *finite modular lattice*, the **quantum probability** function is defined as<sup>7</sup>

- |                      |   |
|----------------------|---|
| <b>D<br/>E<br/>F</b> | <ol style="list-style-type: none"> <li>(1). <math>P(0) = 0</math> (nondegenerate) and</li> <li>(2). <math>P(1) = 1</math> (normalized) and</li> <li>(3). <math>x \perp y \implies P(x \vee y) = P(x) + P(y) \quad \forall x, y \in X</math> (additive) .</li> </ol> |
|----------------------|---|

However, for lattices that are not *distributive*, *modular*, or *orthomodular*, none of these definitions

<sup>2</sup> 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

<sup>3</sup> de Finetti (1930a), Fréchet (1930a), de Finetti (1930b), Fréchet (1930b), de Finetti (1930c), Cifarelli and Regazzini (1996) pages 258–260

<sup>4</sup> Kolmogorov (1933b), page 15

<sup>5</sup> Papoulis (1991) pages 21–22, Kolmogorov (1933b), page 2 (§1. Axioms I–V)

<sup>6</sup> Narens (2014) page 118, Narens (2007) (ISBN:9812708014)

<sup>7</sup> 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  $(O_6, P)$  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

$$a \wedge b = 0 \quad \text{but} \quad P(a \vee b) = P(1) = 1 \neq \frac{1}{3} + \frac{1}{2} = P(a) + P(b).$$

It fails to be a *generalized probability* function because

$$P(a \vee b) = P(1) = 1 \neq \frac{1}{3} + \frac{1}{2} - 0 = P(a) + P(b) - P(0) = P(a) + P(b) - P(a \wedge b).$$

It fails to be an *quantum probability* function because

$$a \perp b = 0 \quad \text{but} \quad P(a \vee b) = P(1) = 1 \neq \frac{1}{3} + \frac{1}{2} = P(a) + P(b).$$

In each of these cases, the function  $P$  fails to be *additive*. The solution of Definition A.1 (page 154) 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  $\sigma$ -*algebra* (*sigma-algebra*) (Definition ?? page ??). A  $\sigma$ -*algebra* is a subset structure that is closed under set complement. A set together with a  $\sigma$ -*algebra* is called a *measurable space*. And a set together with a  $\sigma$ -*algebra* and a *measure* on that  $\sigma$ -*algebra* is called a *measure space* (Definition ?? page ??).

The next definition presents a very important measure space—the *probability space*.

### Definition A.2.

**D E F** The triple  $(\Omega, \mathbb{E}, P)$  is a **probability space** if

- (1).  $\Omega$  is a SET
- (2).  $\mathbb{E}$  is a  $\sigma$ -ALGEBRA on  $\Omega$  (Definition ?? page ??) and
- (3).  $P : \mathbb{E} \rightarrow [0, 1]$  is a MEASURE on  $\mathbb{E}$  (Definition ?? page ??) .

If  $S \triangleq (\Omega, \mathbb{E}, P)$  is a PROBABILITY SPACE then  $x$  is an **outcome** in  $S$  if  $x \in \Omega$ ,  $A$  is an **event** in  $S$  if  $A \in \mathbb{E}$ , and  $PA$  is the **probability** of  $A$  in  $S$  if  $A$  is an EVENT in  $S$ .

### Definition A.3.<sup>8</sup> Let $(\Omega, \mathbb{E}, P)$ be a PROBABILITY SPACE (Definition A.2 page 155).

<sup>8</sup> Papoulis (1990) page 52 (Independent Events)

**D E F** Two EVENTS  $A$  and  $B$  in  $\mathbb{E}$  are **independent** if  
 $P(A \cap B) = P(A)P(B)$

**Definition A.4.** <sup>9</sup> Let  $(\Omega, \mathbb{E}, P)$  be a PROBABILITY SPACE (Definition A.2 page 155). Let  $x$  and  $y$  be EVENTS in  $\mathbb{L}$ .

**D E F** The **conditional probability** of  $x$  given  $y$  is defined as  
 $P(A|B) \triangleq \frac{P(A \cap B)}{P(B)}$

## A.3 Properties

**Proposition A.1.**

**P R P**  $(\Omega, \mathbb{E}, P)$  is a PROBABILITY SPACE  $\implies (\Omega, \mathbb{E}, P)$  is a MEASURE SPACE  
(every probability space is a measure space)

**Theorem A.1.** <sup>10</sup> Let  $(\Omega, \mathbb{E}, P)$  be a PROBABILITY SPACE (Definition A.2 page 155).

**T H M**

(1).	$0 \leq P(x) \leq 1$	$\forall x \in X$	(BOUNDED)	and
(2).	$P(x) = 1 - P(x^\perp)$	$\forall x \in X$	(PARTITION OF UNITY)	and
(3).	$x \leq y \implies P(y^\perp) \leq P(x^\perp)$	$\forall x, y \in X$	(ANTITONE)	

PROOF:

1. Proof for  $0 \leq P(x) \leq 1$ :

$$\begin{aligned} 0 &= P(0) && \text{by by nondegenerate property of } P \text{ (Definition A.2 page 155)} \\ &\leq P(x) && \text{because } 0 \leq x \text{ and monotone property of } P \\ &\leq P(1) && \text{because } x \leq 1 \text{ and monotone property of } P \\ &= 1 && \text{by normalized property of } P \end{aligned}$$

2. Proof for  $P(x) = 1 - P(x^\perp)$ :

(a) Proof that  $P$  is *additive* (Definition A.2 page 155) 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 ...

$$\begin{aligned} 1 - P(x^\perp) &= P(1) - P(x^\perp) && \text{by normalized property of } P && \text{(Definition A.2 page 155)} \\ &= P(x \vee x^\perp) - P(x^\perp) && \text{by excluded middle property of ortho. lat.} \\ &= P(x) + P(x^\perp) - P(x^\perp) && \text{by additive property of } (\Omega, \mathbb{E}, P) && \text{(item (2a) page 156)} \\ &= P(x) && \text{by field property of } (\mathbb{R}, +, \cdot, 0, 1) \end{aligned}$$

3. Proof for  $x \leq y \implies P(y^\perp) \leq P(x^\perp)$ :

$$\begin{aligned} x \leq y &\implies y^\perp \leq x^\perp && \text{by antitone property of orthocomplemented lattices} \\ &\implies P(y^\perp) \leq P(x^\perp) && \text{by monotone property of } P \end{aligned}$$

(Definition A.2 page 155)

<sup>9</sup>  Papoulis (1990) page 45 (2-3 Conditional Probability and Independence)

<sup>10</sup> property (1):  Papoulis (1991) page 21 ((2-11))



**Theorem A.2.** <sup>11</sup> Let  $(\Omega, \mathbb{E}, P)$  be a PROBABILITY SPACE (Definition A.2 page 155).

<b>T H M</b>	$L$ is BOOLEAN (Definition ?? page ??)	$\left\{ \begin{array}{l} 1. \quad P(x \vee y) = P(x) + P(y) - P(x \wedge y) \quad \forall x, y \in X \quad \text{and} \\ 2. \quad P(x \vee y) \leq P(x) + P(y) \quad \forall x, y \in X \quad (\text{BOOLE'S INEQUALITY}) \end{array} \right.$
----------------------	---	---

PROOF:

1. lemma: Proof that  $P((\neg x) \wedge y) = P(y) - P(x \wedge y)$ :

$$\begin{aligned}
 P(y) - P(xy) &= P(1 \wedge y) - P(xy) && \text{by definition of } \wedge \\
 &= P[(x \vee x^\perp)y] - P(xy) && \text{by excluded middle property of Boolean lattices} \\
 &= P(xy \vee x^\perp y) - P(xy) && \text{by distributive property of Boolean lattices} \\
 &= P(xy) + P(x^\perp y) - P(xy) && \text{because } (xy)(x^\perp y) = 0 \text{ and by additive property} \\
 &= P(x^\perp y)
 \end{aligned}$$

2. Proof that  $P(x \vee y) = P(x) + P(y) - P(x \wedge y)$ :

$$\begin{aligned}
 P(x \vee y) &= P(x \vee x^\perp y) && \text{by property of Boolean lattices} \\
 &= P(x) + P(x^\perp y) && \text{because } (x)(x^\perp y) = 0 \text{ and by additive property} \\
 &= P(x) + P(y) - P(x \wedge y) && \text{by item (1) (page 157)}
 \end{aligned}$$



**Theorem A.3** (sum of products). Let  $(X, \vee, \wedge, 0, 1 ; \leq)$  be a BOUNDED LATTICE,  $(\Omega, \mathbb{E}, P)$  a PROBABILITY SPACE (Definition A.2 page 155), and  $\{y, x_1, x_2, x_3, \dots\}$  a subset of  $X$ .

<b>T H M</b>	$\left\{ \begin{array}{l} 1. \quad L \text{ is DISTRIBUTIVE} \\ 2. \quad \{x_1, x_2, \dots\} \text{ is a PARTITION of } y \end{array} \right. \text{ and} \right\} \Rightarrow \left\{ \begin{array}{l} 1. \quad P(y) = \sum_n P(x_n) \quad \text{and} \\ 2. \quad P(y) = \sum_n P(y \wedge x_n) \quad \text{and} \\ 3. \quad P(z \wedge y) = \sum_n P(z \wedge x_n) \end{array} \right\}$
----------------------	--

PROOF:

1. Proof that  $P$  is *additive* (Definition A.2 page 155) on  $(\Omega, \mathbb{E}, P)$ :

(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  $L$  is *distributive*: by *distributive hypothesis*

2. Proof that  $P(y) = \sum_n P(x_n)$

$$\begin{aligned}
 P(y) &= P(yx_1 \vee yx_2 \vee \dots \vee yx_n) && \text{by item (1) and additive property} \\
 &= \sum_n P(yx_n) && \text{by item (1) and additive property} \\
 &= \sum_n P(y|x_n)P(x_n) && \text{by conditional probability} && \text{(Definition A.4 page 156)}
 \end{aligned}$$

<sup>11</sup> Papoulis (1991) page 21 ((2-13)), Feller (1970) pages 22–23 ((7.4),(7.6))



As described in Definition A.2 (page 155), every *probability space*  $(\Omega, \mathbb{E}, P)$  contains a probability *measure*  $P : \mathbb{E} \rightarrow [0, 1]$ . This probability *measure* has some basic properties as described in Theorem A.4 (next).

**Theorem A.4.** Let  $(\Omega, \mathbb{E}, P)$  be a PROBABILITY SPACE. Let  $B$  be a set and  $\{B_n | n = 1, 2, \dots, N\}$  a set of sets.

T  
H  
M

$\left\{ \begin{array}{l} \{B_n | n = 1, 2, \dots, N\} \text{ is a} \\ \text{PARTITION of } B. \end{array} \right\}$

$$\Rightarrow \left\{ \begin{array}{l} (1). \quad P(B) = \sum_{n=1}^N P(B_n) \quad \forall B \in \mathbb{E} \quad \text{and} \\ (2). \quad P(AB) = \sum_{n=1}^N P(AB_n) \quad \forall A, B \in \mathbb{E} \end{array} \right\}$$



PROOF:  $P$  is a *measure* and by Definition ?? (page ??).



**Proposition A.2.** Let  $(\Omega, \mathbb{E}, P)$  be a probability space, and  $X$  a RANDOM VARIABLE (Definition B.1 page 164) with PROBABILITY DENSITY FUNCTION  $p_x(x)$  and CUMULATIVE DISTRIBUTION FUNCTION  $c_x(x)$ .

P  
R  
P

- (1).  $c_x(x)$  is MONOTONE and
- (2).  $p_x(x)$  is CONTINUOUS  $\Rightarrow c_x(x)$  is STRICTLY MONOTONE and
- (3).  $p_x(x)$  is CONTINUOUS  $\Rightarrow c_x(x)$  is INVERTIBLE

**Theorem A.5** (Bayes' Rule). <sup>12</sup>

T  
H  
M

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$$

$$\begin{aligned} P(X|Y) &\triangleq \frac{P(X \cap Y)}{P(Y)} && \text{by definition of conditional probability} \\ &= \frac{P(Y \cap X)}{P(Y)} && \text{by commutative property of } \cap \\ &= \frac{P(Y|X)P(X)}{P(Y)} && \text{by definition of conditional probability} \end{aligned}$$

(Definition A.4 page 156)

by commutative property of  $\cap$

(Definition A.4 page 156)

## A.4 Examples

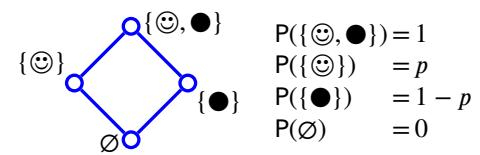
*Example A.1* (single coin toss). Let  $\odot$  represent “heads” and  $\bullet$  represent “tails” in a coin toss. Let  $0 < p < 1$  be the probability of a head. A *probability space*  $(\Omega, \mathbb{E}, P)$  for a single coin toss is as follows:

E  
X

$$\Omega = \{\odot, \bullet\}$$

$$\mathbb{E} = \left\{ \begin{array}{l} \emptyset, \{\odot\}, \{\bullet\}, \{\odot, \bullet\} \\ \text{neither heads tails heads or tails} \end{array} \right\}$$

$$P(X) = \left\{ \begin{array}{lll} 0 & \text{for } X = \emptyset & (\text{neither heads nor tails}) \\ p & \text{for } X = \{\odot\} & (\text{heads}) \\ 1-p & \text{for } X = \{\bullet\} & (\text{tails}) \\ 1 & \text{for } X = \{\odot, \bullet\} & (\text{either heads or tails}) \end{array} \right\}$$



$$\begin{aligned} P(\{\odot, \bullet\}) &= 1 \\ P(\{\odot\}) &= p \\ P(\{\bullet\}) &= 1 - p \\ P(\emptyset) &= 0 \end{aligned}$$

<sup>12</sup> Haykin (2014) page 95 (“Bayes’ Rule”)



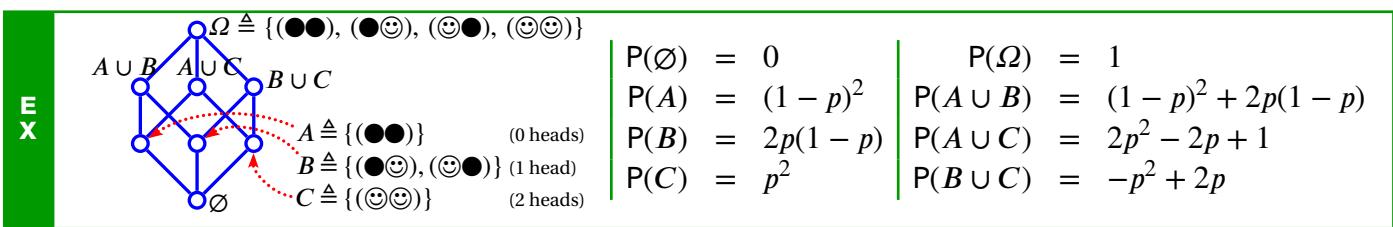


Figure A.2: Double coin toss (Example A.2 page 159)

*Example A.2* (Double coin toss). Let  $\circ$  represent “heads” and  $\bullet$  represent “tails” in a double coin toss in which each toss is *independent* (Definition A.3 page 155) of the other. Let  $0 < p < 1$  be the probability of a head. The *probability space*  $(\Omega, \mathbb{E}, P)$  is illustrated in Figure A.2 (page 159).

PROOF:

$$\begin{aligned}
 P(\Omega) &= 1 && \text{by normalized property of } P && \text{(Definition A.1 page 154)} \\
 P(C) &= P\{\circ\circ\} && \text{by definition of } C \\
 &= P(\circ)\P(\circ) && \text{by definition of } \textit{independence} && \text{(Definition A.3 page 155)} \\
 &= p^2 && \text{by definition of } p \\
 P(A) &= P\{\bullet\bullet\} && \text{by definition of } A \\
 &= P(\bullet)\P(\bullet) && \text{by definition of } \textit{independence} && \text{(Definition A.3 page 155)} \\
 &= \{1 - P(\circ)\}\{1 - P(\circ)\} && \text{by } \textit{antitone} \text{ property of } P && \text{(Theorem A.1 page 156)} \\
 &= (1-p)^2 && \text{by definition of } p \\
 P(B) &= P\{(\bullet\circ), (\circ\bullet)\} && \text{by definition of } B \\
 &= P\{\bullet\circ\} + P\{\circ\bullet\} && \text{by } \textit{additive} \text{ property of } P && \text{(Definition A.1 page 154)} \\
 &= P\{\bullet\}\P(\circ) + P\{\circ\}\P(\bullet) && \text{by definition of } \textit{independence} && \text{(Definition A.3 page 155)} \\
 &= (1-p)p + p(1-p) && \text{by } \textit{antitone} \text{ property of } P && \text{(Theorem A.1 page 156) and definition of } p \\
 &= -2p^2 + p + 1 && && \\
 P(A \cup B) &= P(A) + P(B) - P(A \cap B) && \text{by Theorem A.2} \\
 &= P(A) + P(B) - P(\emptyset) \\
 &= (1-p)^2 + (-2p^2 + p + 1) + 0 && \text{by previous results} \\
 &= -p^2 - p + 1 && && \\
 P(\emptyset) &= 0 && \text{by } \textit{nondegenerate} \text{ property of } P && \text{(Definition A.1 page 154)}
 \end{aligned}$$

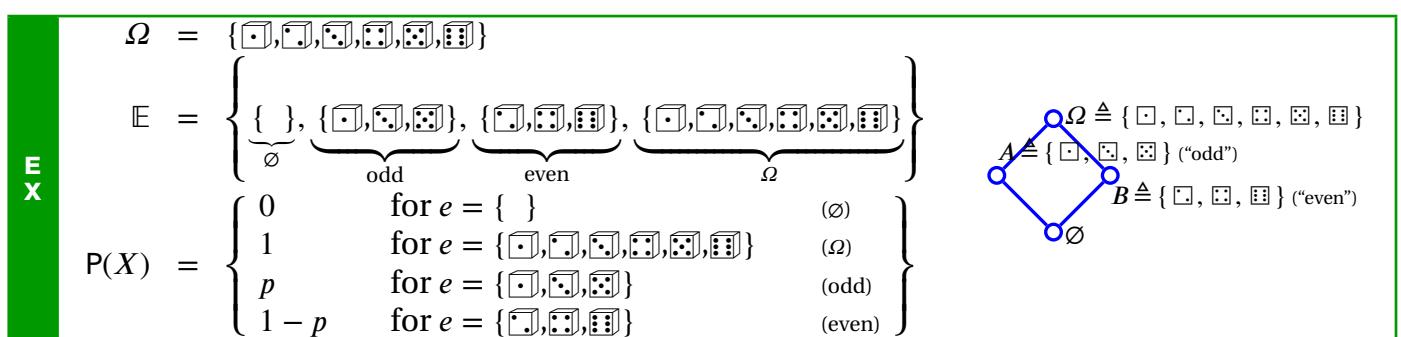


Figure A.3: even/odd die probability space (Example A.3 page 159)

*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 159).

PROOF:

$P(\Omega) = 1$	by <i>normalized</i> property of $P$	(Definition A.1 page 154)
$P(C) = P\{\text{⊕⊕}\}$	by definition of $C$	
$= P(\text{⊕})P(\text{⊕})$	by definition of <i>independence</i>	(Definition A.3 page 155)
$= p^2$	by definition of $p$	
$P(A) = P\{\text{◻}, \text{◻}, \text{◻}\}$	by definition of $A$	
$= p$	by definition of $p$	
$P(B) = P\{\text{◻}, \text{◻}, \text{◻}\}$	by definition of $B$	
$= P\{\text{◻}, \text{◻}, \text{◻}\}^c$	by definition of set complement $c$	
$= PA^c$	by definition of $A$	
$= P(\neg A)$	by definition of $\neg$	
$= 1 - P(A)$	by Theorem A.1 page 156	
$= 1 - p$	by definition of $p$	
$P(\emptyset) = 0$	by <i>nondegenerate</i> property of $P$	(Definition A.1 page 154)



The two previous *even/odd die* example (Example A.5 page 160) is in essence the same as the *single coin toss* (Example A.1 page 158). The next offers a little more complexity.

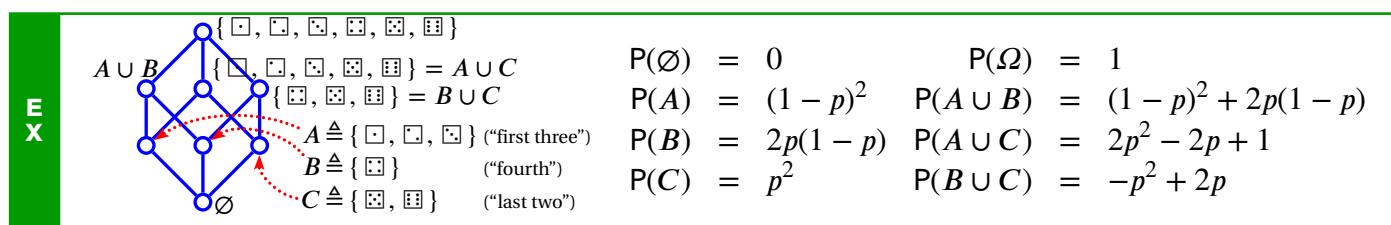


Figure A.4: 3-4-2 die example (Example A.4 page 160)

*Example A.4.* Suppose we have a “fair” die and we are primarily interested in the events of the first three  $\{\text{◻}, \text{◻}, \text{◻}\}$ , the next two,  $\{\text{◻}, \text{◻}\}$  and the final one  $\{\text{◻}\}$ . The resulting *probability space* is illustrated in Figure A.4 (page 160).

The two previous examples (Example A.5 page 160, Example A.4 page 160) illustrate a *probability 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 ( $\{\text{◻}, \text{◻}, \text{◻}, \text{◻}\}$ ) (that is, whether one roll of the die will produce a value in the set  $\{\text{◻}, \text{◻}, \text{◻}, \text{◻}\}$ ) and the last three ( $\{\text{◻}, \text{◻}, \text{◻}\}$ ). However, these events do not by themselves form a  $\sigma$ -algebra. Rather under the  $\cap$  and  $\cup$  operations, these two events generate a total of eight possible events that together form a  $\sigma$ -algebra. The resulting *probability space* is illustrated in Figure A.5 (page 161).

But why go through all the trouble of requiring a  $\sigma$ -algebra? Having a  $\sigma$ -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  $\sigma$ -*additive* property). Example A.6 (next) illustrates how a rather complex  $\sigma$ -algebra (64 elements) can be compactly represented in a measure space.



EX	$\Omega = \{\square, \square, \square, \square, \square, \square\}$ $\mathbb{E} = \left\{ \underbrace{\{\}}_{\emptyset}, \underbrace{\{\square, \square, \square, \square, \square, \square\}}_{\Omega}, \underbrace{\{\square, \square, \square, \square, \square\}}_{\text{first four}}, \underbrace{\{\square, \square, \square\}}_{\text{last three}}, \right.$
	$\left. \begin{array}{l} \underbrace{\{\square\}}_{\{1234\} \cap \{456\}}, \underbrace{\{\square, \square, \square, \square, \square\}}_{\{4\}^c}, \underbrace{\{\square, \square\}}_{\{4\}^c \cap \{456\}}, \underbrace{\{\square, \square, \square\}}_{\{1234\} \cap \{4\}^c} \end{array} \right\}$ $P(e) = \left\{ \begin{array}{lll} 0 & \text{for } e = \{\} & (\emptyset) \\ 1 & \text{for } e = \{\square, \square, \square, \square, \square, \square\} & (\Omega) \\ \frac{2}{3} & \text{for } e = \{\square, \square, \square, \square, \square\} & (\text{first four}) \\ \frac{1}{2} & \text{for } e = \{\square, \square, \square\} & (\text{last three}) \\ \frac{1}{6} & \text{for } e = \{\square\} & (\{1234\} \cap \{456\}) \\ \frac{5}{6} & \text{for } e = \{\square, \square, \square, \square, \square\} & (\{4\}^c) \\ \frac{1}{3} & \text{for } e = \{\square, \square\} & (\{4\}^c \cap \{456\}) \\ \frac{1}{2} & \text{for } e = \{\square, \square, \square\} & (\{1234\} \cap \{4\}^c) \end{array} \right.$

Figure A.5: First 4 / last 3 die example (Example A.5 page 160)

*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  $(\Omega, \mathbb{E}, P)$  where

EX	$\Omega = \{\square, \square, \square, \square, \square, \square\}$ $\mathbb{E} = \mathcal{P}(\Omega)$ $P(e) = \frac{1}{6} e $	(the power-set of $\Omega$ ) ( $\frac{1}{6}$ times the number of possible outcomes in event $e$ )
----	--	--

*Example A.7 (Gaussian distribution on  $\mathbb{R}$ ).* Let  $\mathbf{B}$  be the *Borel algebra on  $\mathbb{R}$* . Let  $\mathbf{L} \triangleq (\mathbf{B}, \subseteq)$  be the lattice formed by the elements of  $\mathbf{B}$ —this lattice is a *Boolean algebra*. Let

$$P(A) \triangleq \frac{1}{\sqrt{2\pi\sigma^2}} \int_A e^{\frac{x^2}{2\sigma^2}} dx \text{ for } A \subseteq \mathbf{B}$$

and where  $\int$  is the *Lebesgue integral* (Definition ?? page ??). Then  $(\mathbf{L}, P)$  is a **probability space**.

*Example A.8 (Gaussian noise).* Let  $X \sim N(0, \sigma^2)$  be a random variable with Gaussian distribution. We can construct the following probability space  $(\Omega, \mathbb{E}, P)$ :

EX	$\Omega = \mathbb{R}$ $\mathbb{E} = \{\emptyset, \Omega\} \cup \{(a, b)   a, b \in \mathbb{R}, a < b\}$ $P_x = \begin{cases} 0 & \text{for } x = \emptyset \\ 1 & \text{for } x = \Omega \\ \frac{1}{\sqrt{2\pi\sigma^2}} \int_a^b e^{\frac{x^2}{2\sigma^2}} dx & \text{otherwise} \end{cases}$
----	---

*Example A.9.* The set of outcomes  $\Omega$  can also be a set of waveforms:

EX	$\Omega = \left\{ \begin{array}{c} \text{[waveform diagram]} \end{array} \right\}$ $\mathbb{E} = \mathcal{P}(\Omega)$ $P(e) = \frac{1}{7} e $
----	---

## A.5 Probability subspaces

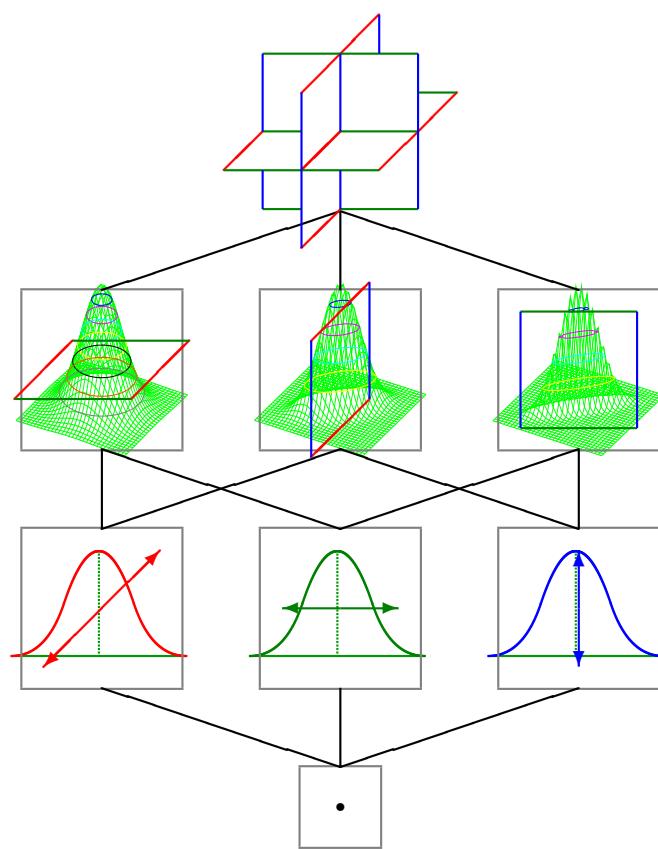
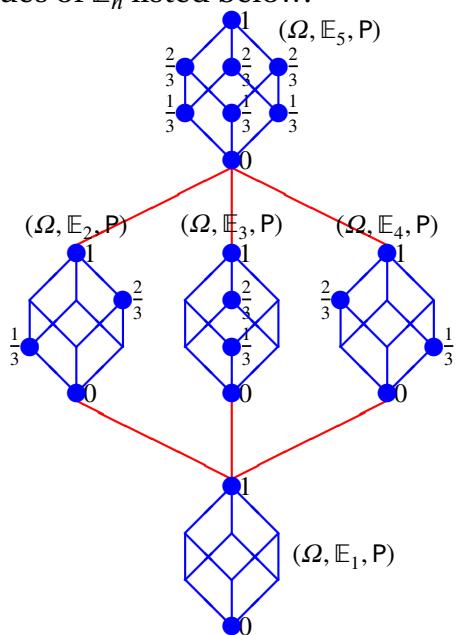


Figure A.6: Euclidean 3-dimensional space partitioned as a power lattice

*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  $\Omega$  and therefore five probability spaces  $(\Omega, \mathbb{E}_n, P)$  on  $\Omega$  with the five values of  $\mathbb{E}_n$  listed below: <sup>13</sup>



$$\begin{aligned}
 \mathbb{E}_1 &= \{ \emptyset, X \} \\
 \mathbb{E}_2 &= \{ \emptyset, \{x\}, \{y, z\}, X \} \\
 \mathbb{E}_3 &= \{ \emptyset, \{y\}, \{x, z\}, X \} \\
 \mathbb{E}_4 &= \{ \emptyset, \{z\}, \{x, y\}, X \} \\
 \mathbb{E}_5 &= \{ \emptyset, \{x\}, \{y\}, \{z\}, \{x, y\}, \{x, z\}, \{y, z\}, X \}
 \end{aligned}$$

Suppose the samples of  $\Omega$  are generated by a physical process such that they are all equally likely to occur. Then by varying the  $\sigma$ -algebra  $\mathbb{E}_n$  effectively varies the probability distribution of the probability space  $(\Omega, \mathbb{E}_n, P)$ . This is illustrated by the figure to the left.

<sup>13</sup> algebra of sets: Definition ?? page ??

## 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.”<sup>1</sup>

Joseph Leonard Doob (1910–2004), pioneer of and key contributor to mathematical probability<sup>1</sup>

## 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.* <sup>2</sup> As pointed out by others, the term “random variable” is a “misnomer”:

RE  
M

- 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*

<sup>1</sup> quote: [Snell \(1997\)](#), page 307, [Snell \(2005\)](#), page 251

image: <http://www.dartmouth.edu/~chance/Doob/conversation.html>

<sup>2</sup> [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  $\omega$  of the experiment appears to be random to the observer. So the *random variable*  $X(\omega)$  is simply a function of an underlying mechanism that appears to be random.

**Definition B.1.** <sup>3</sup> Let  $(\Omega, \mathbb{E}, P)$  be a PROBABILITY SPACE (Definition A.2 page 155).

**D E F** A **random variable**  $X$  is any function in  $\mathbb{R}^{\Omega}$ .

## B.2 Probability distributions

The probability information about  $\sigma$ -algebra  $\mathbb{E}$  in a *probability space* (Definition A.2 page 155) 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.** <sup>4</sup> Let  $X$  be a RANDOM VARIABLE (Definition B.1 page 164) on a PROBABILITY SPACE  $(\Omega, \mathbb{E}, P)$ .

**D E F**

$X$ has <b>cummulative distribution function</b> (cdf) $c_X(x)$ if $c_X(x) \triangleq P\{x \in \mathbb{E}   X < x\}$	$X$ <b>probability density function</b> (pdf) $p_X(x)$ if $p_X(x) \triangleq \frac{d}{dx}c_X(x) \triangleq \frac{d}{dx}P\{x \in \mathbb{E}   X < x\}$
---	--

*Remark B.2.* Suppose  $X$  be a *random variable* on a *probability space*  $(\Omega, \mathbb{E}, P)$ . Note that

- Both  $X$  and  $\mathbb{E}$  are *functions*.
- But  $X$  is a function that maps from  $\Omega$  to  $\mathbb{R}$ ,
- whereas  $P$  is a function that maps from  $\mathbb{E}$  to  $\mathbb{R}$ .

**Definition B.3.** Let  $(\Omega, \mathbb{E}, P)$  be a PROBABILITY SPACE (Definition A.2 page 155) and  $X$  and  $Y : \Omega \rightarrow \mathbb{R}$  random variables. Then a **joint probability density function**  $p_{XY} : \mathbb{E} \times \Omega \rightarrow [0 : 1]$  and a **joint cumulative distribution function**  $c_{XY} : \mathbb{E} \times \Omega \rightarrow [0 : 1]$  are defined as

**D E F**

$c_{XY}(x, y) \triangleq P\{X \leq x   Y \leq y\}$ (JOINT CUMULATIVE DISTRIBUTION FUNCTION)	$p_{XY}(x, y) \triangleq \frac{d}{dy} \frac{d}{dx} c_{XY}(x, y)$ (JOINT PROBABILITY DENSITY FUNCTION)
---	---

**Definition B.4.** Let  $(\Omega, \mathbb{E}, P)$  be a PROBABILITY SPACE (Definition A.2 page 155) and  $X$  a random variable. Then a **conditional probability density function**  $p_X : \mathbb{E} \times \Omega \rightarrow [0 : 1]$  and a **conditional cumulative distribution function**  $c_X : \mathbb{E} \times \Omega \rightarrow [0 : 1]$  are defined as

**D E F**

$c_X(x y) \triangleq P\{X \leq x   Y = y\}$ (CONDITIONAL CUMULATIVE DISTRIBUTION FUNCTION—CDF)	$p_X(x y) \triangleq \frac{d}{dx} c_X(x y)$ (CONDITIONAL PROBABILITY DENSITY FUNCTION—PDF)
--	--

## B.3 Properties

Definition B.2 (page 164) defines the pdf and cdf of a *probability space*  $(\Omega, \mathbb{E}, P)$  in terms of *measure*  $P$ . Conversely, the probability *measure*  $P\{a \leq X < b\}$  of an event  $\{a \leq X < b\}$  can be expressed in terms of either the pdf or cdf.

<sup>3</sup>  Papoulis (1991), page 63

<sup>4</sup>  von der Linden et al. (2014) page 93 (Definitions 7.1, 7.2)



**Proposition B.1.** Let  $X$  a RANDOM VARIABLE with PDF  $p_x$  and CDF  $c_x$  (Definition B.2 page 164) on the PROBABILITY SPACE  $(\Omega, \mathbb{E}, P)$  (Definition A.2 page 155).

P R P	$\begin{aligned} & \left\{ \begin{array}{l} (1). \quad c_x(x) \text{ and } c_y(y) \text{ are CONTINUOUS} \\ (2). \quad p_x(x) \text{ and } p_y(y) \text{ are CONTINUOUS} \end{array} \text{ OR } \right\} \\ \implies & \left\{ \begin{aligned} p_x(x) &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} P\{x \leq X < x + \varepsilon\} \\ p_{XY}(x, y) &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} P\{x \leq X < x + \varepsilon \wedge y \leq Y < y + \varepsilon\} \end{aligned} \right\} \end{aligned}$
-------------	--

PROOF:

$$\begin{aligned} p_x(x) &\triangleq \frac{d}{dx} c_x(x) && \text{by definition of } p_x && (\text{Definition B.2 page 164}) \\ &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} P\{x \in \mathbb{R} | x \leq X < x + \varepsilon\} && \text{by definition of } \frac{d}{dx} && (\text{Definition ?? page ??}) \end{aligned}$$



**Theorem B.1.** Let  $(\Omega, \mathbb{E}, P)$  be a probability space,  $X$  be a random variable, and  $(a, b)$  a real interval.

T H M	$\left\{ \begin{array}{l} (1). \quad c_x(x) \text{ is CONTINUOUS} \\ (2). \quad p_x(x) \text{ is CONTINUOUS} \end{array} \text{ OR } \right\} \implies \left\{ P\{a < X \leq b\} = c_x(b) - c_x(a) = \int_a^b p_x(x) dx \right\}$
-------------	---

PROOF:

$$\begin{aligned} P\{a < X \leq b\} &= P\{X \leq b\} - P\{X < a\} && \text{by sum of products} && (\text{Theorem A.3 page 157}) \\ &= P\{X \leq b\} - P\{X \leq a\} && \text{by continuity hypothesis} \\ &\triangleq c_x(b) - c_x(a) && \text{by definition of } c_x && (\text{Definition B.2 page 164}) \end{aligned}$$

$$\begin{aligned} \int_a^b p_x(x) dx &\triangleq \int_a^b \left[ \frac{d}{dx} c_x(x) \right] dx && \text{by definition of } p_x && (\text{Definition B.2 page 164}) \\ &= c_x(x)|_{x=b} - c_x(x)|_{x=a} && \text{by Fundamental theorem of calculus} \\ &= c_x(b) - c_x(a) \end{aligned}$$



**Theorem B.2.** Let  $(\Omega, \mathbb{E}, P)$  be a PROBABILITY SPACE,  $X$  be a RANDOM VARIABLE, and  $(a : b)$  a REAL INTERVAL.

T H M	$P\{a \leq X < b\} = \int_a^b p_x(x) dx = \int_{-\infty}^b c_x(x) dx - \int_{-\infty}^a c_x(x) dx$
-------------	--

The properties of the pdf follow closely the properties of measure  $P$ .

**Theorem B.3.**<sup>5</sup>

T H M	$\left\{ \begin{array}{l} (A). \quad c_x(x) \text{ is CONTINUOUS} \\ (B). \quad p_x(x) \text{ is CONTINUOUS} \end{array} \text{ OR } \right\} \implies \left\{ \begin{array}{l} (1). \quad p_{XY}(x y) = \frac{p_{XY}(x, y)}{p_Y(y)} \text{ and} \\ (2). \quad p_x(x) = \int_{y \in \mathbb{R}} p_{XY}(x, y) dy \end{array} \right\}$
-------------	---

<sup>5</sup> Papoulis (1990) page 158 (Auxiliary Variable), Jazwinski (1970), page 39 (“(2.102)”), Jazwinski (2007) page 39 (“(2.102)”).

PROOF:

$$\begin{aligned}
 p_{X|Y}(x|y) &\triangleq \frac{d}{dx} c_{X|Y}(x|y) && \text{by definition of } c_X \quad (\text{Definition A.4 page 156}) \\
 &\triangleq \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} P\{x \leq X < x + \varepsilon | Y = y\} && \text{by definition of } \frac{d}{dx} \quad (\text{Definition ?? page ??}) \\
 &\triangleq \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \frac{P\{(x \leq X < x + \varepsilon) \wedge (Y = y)\}}{P\{Y = y\}} && \text{by definition of } P\{A|B\} \quad (\text{Definition A.4 page 156}) \\
 &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \frac{P\{(x \leq X < x + \varepsilon) \wedge (y \leq Y < y + \varepsilon)\}}{P\{y \leq Y < y + \varepsilon\}} && \text{by continuity hypothesis} \\
 &= \frac{\lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} P\{(x \leq X < x + \varepsilon) \wedge (y \leq Y < y + \varepsilon)\}}{\lim_{\varepsilon \rightarrow 0} P\{y \leq Y < y + \varepsilon\}} && \text{by property of } \lim_{\varepsilon \rightarrow 0} \\
 &= \frac{p_{XY}(x, y)}{p_Y(y)} && \text{by Proposition B.1 page 165}
 \end{aligned}$$

$$\begin{aligned}
 \int_{y \in \mathbb{R}} p_{XY}(x, y) dy &\triangleq \int_{y \in \mathbb{R}} \left[ \frac{d}{dy} \frac{d}{dx} c_{XY}(x, y) \right] dy && \text{by definition of } p_X \quad (\text{Definition B.2 page 164}) \\
 &= \frac{d}{dx} c_{XY}(x, y) && \\
 &\triangleq \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \int_{y \in \mathbb{R}} P\{x \leq X < x + \varepsilon, y \leq Y < y + \varepsilon\} dy && \text{by definition of } \frac{d}{dx} \quad (\text{Definition ?? page ??}) \\
 &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} P\{x \leq X < x + \varepsilon\} && \\
 &= p_X(x) && \text{by Proposition B.1 page 165}
 \end{aligned}$$

⇒

### Theorem B.4.

T	$c_X(\sup \mathbb{R}) = 1$
H	$c_X(\inf \mathbb{R}) = 0$

PROOF:

$$\begin{aligned}
 c_X(\sup \mathbb{R}) &\triangleq P\{X \leq \sup \mathbb{R}\} && \text{by definition of } c_X \quad (\text{Definition B.2 page 164}) \\
 &= 1 \\
 c_X(\inf \mathbb{R}) &\triangleq P\{X \leq \inf \mathbb{R}\} && \text{by definition of } c_X \quad (\text{Definition B.2 page 164}) \\
 &= 0
 \end{aligned}$$

⇒

The properties of the pdf follow closely the properties of measure P.

### Theorem B.5.

T	$c_{X Y}(x y) = \frac{\frac{d}{dy} c_{XY}(x, y)}{p_Y(y)}$	$p_{X Y}(x y) = \frac{p_{XY}(x, y)}{p_Y(y)}$
---	---	--



PROOF:

$$\begin{aligned}
 c_{X|Y}(x|y) &\triangleq P\{X \leq x | Y = y\} && \text{by definition of } c_{X|Y} && (\text{Definition B.4 page 164}) \\
 &\triangleq \frac{P\{X \leq x | Y = y\}}{P\{Y = y\}} && \text{by definition of } P\{X|Y\} && (\text{Definition A.4 page 156}) \\
 &= \lim_{\epsilon \rightarrow 0} \frac{P\{X \leq x | y < Y \leq y + \epsilon\}}{P\{y < Y \leq y + \epsilon\}} \\
 &= \lim_{\epsilon \rightarrow 0} \frac{[P\{X \leq x | Y \leq y + \epsilon\} - P\{X \leq x | Y \leq y\}]/\epsilon}{[P\{Y \leq y + \epsilon\} - P\{Y \leq y\}]/\epsilon} \\
 &\triangleq \lim_{\epsilon \rightarrow 0} \frac{[c_{XY}(x, y + \epsilon) - c_{XY}(x, y)]/\epsilon}{[c_Y(y + \epsilon) - c_Y(y)]/\epsilon} && \text{by definition of } c_{XY} && (\text{Definition B.3 page 164}) \\
 &\triangleq \frac{\frac{d}{dy}c_{XY}(x, y)}{\frac{d}{dy}c_Y(y)} && \text{by definition of } \frac{d}{dy}f(y) \\
 &\triangleq \frac{\frac{d}{dy}c_{XY}(x, y)}{p_Y(y)} && \text{by definition of } p_Y && (\text{Definition B.2 page 164}) \\
 &= \frac{\frac{d}{dy}c_{XY}(x, y)}{p_Y(y)} && \text{because } y \text{ is fixed}
 \end{aligned}$$

$$\begin{aligned}
 p_{X|Y}(x|y) &\triangleq \frac{d}{dx}c_{X|Y}(x|y) && \text{by definition of } p_{X|Y} && (\text{Definition B.4 page 164}) \\
 &= \frac{d}{dx} \frac{\frac{d}{dy}c_{XY}(x, y)}{p_Y(y)} && \text{by previous result} \\
 &= \frac{\frac{d}{dx}\frac{d}{dy}c_{XY}(x, y)}{p_Y(y)} && \text{because } p_Y(y) \text{ is not a function of } x \\
 &\triangleq \frac{p_{XY}(x, y)}{p_Y(y)} && \text{by definition of } p_{XY}(x, y) && (\text{Definition B.3 page 164})
 \end{aligned}$$



**Theorem B.6.** Let  $(\Omega, \mathbb{E}, P)$  be a probability space.

<b>T H M</b>	$\int_{x \in \mathbb{R}} p_X(x) dx = 1$ $\int_{y \in \mathbb{R}} p_{XY}(x, y) dy = p_X(x) \quad \forall x \in \Omega$	$\int_{x \in \mathbb{R}} p_{X Y}(x y) dx = 1$ $\int_{x \in \mathbb{R}} \int_{y \in \mathbb{R}} p_{XY}(x, y) dy dx = 1$
----------------------	--	---

PROOF:

$$\begin{aligned}
 \int_{\mathbb{R}} p_X(x) dx &= c_X(\sup \mathbb{R}) - c_X(\inf \mathbb{R}) && \text{by Theorem B.1 page 165} \\
 &= 1 - 0 \\
 &= 1 && \text{because 0 is the additive identity element in } (\mathbb{R}, +, \cdot, 0, 1) \\
 \int_{x \in \mathbb{R}} p_{X|Y}(x|y) dx &\triangleq \int_{x \in \mathbb{R}} \frac{d}{dx}c_{X|Y}(x|y) dx && \text{by definition of } p_{X|Y}(x|y) (\text{Definition B.4 page 164}) \\
 &= c_{X|Y}(\sup \mathbb{R}|y) - c_{X|Y}(\inf \mathbb{R}|y) && \text{by Fundamental theorem of calculus}
 \end{aligned}$$

$$= 1 - 0$$

$$= 1$$

because 0 is the additive identity element in  $(\mathbb{R}, +, \cdot, 0, 1)$

$$\int_{y \in \mathbb{R}} p_{XY}(x, y) dy = \int_{y \in \mathbb{R}} p_{YX}(y, x) dy$$

$$= \int_{y \in \mathbb{R}} p_{Y|X}(y|x) p_X(x) dy$$

by Theorem B.5 page 166

$$= p_X(x) \int_{y \in \mathbb{R}} p_{Y|X}(y|x) dy$$

because  $p_X(x)$  is not a function of  $y$

$$= p_X(x) \cdot 1$$

by previous result

$$= p_X(x)$$

because 1 is the multiplicative identity element in  $(\mathbb{R}, +, \cdot, 0, 1)$

$$\int_{x \in \mathbb{R}} \int_{y \in \mathbb{R}} p_{XY}(x, y) dy dx = \int_{x \in \mathbb{R}} p_X(x) dx$$

by previous result

$$= 1$$

by previous result



## APPENDIX C

### SOME PROBABILITY DENSITY FUNCTIONS

#### C.1 Discrete distributions

*Example C.1.* <sup>1</sup> 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 170) and has probability space as follows:

EX	$\Omega = \{\square\square, \square\bullet\square, \square\square\bullet, \dots, \bullet\bullet\bullet\}$
	$\mathbb{E} = \{2^{n=2,3,\dots,10,11, \text{ or } 12}\}$
	$P(e) = \frac{1}{36}  e $

#### C.2 Continuous distributions

##### C.2.1 Uniform distribution

**Definition C.1.** The **uniform distribution**  $p_x(x)$  is defined as

DEF	$p_x(x) \triangleq \begin{cases} 1 & \text{for } 0 < x \leq 1 \\ 0 & \text{otherwise} \end{cases}$
-----	--

Note that although “simple” in form, in light of Wold's Theorem,<sup>2</sup> the value of the *uniform distribution* should *not* be taken lightly.

<sup>1</sup>  Osgood (2002)

<sup>2</sup>  Wold (1938),  Wold (1954),  Scargle (1979)

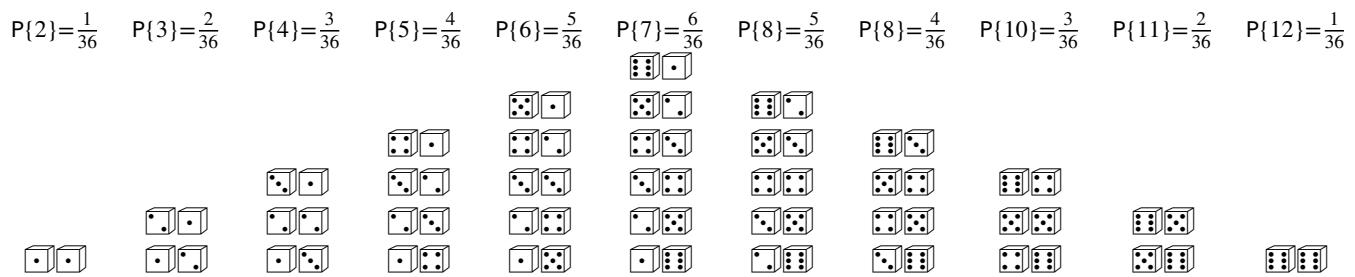


Figure C.1: Probability distribution for two dice (see Example C.1 page 169)

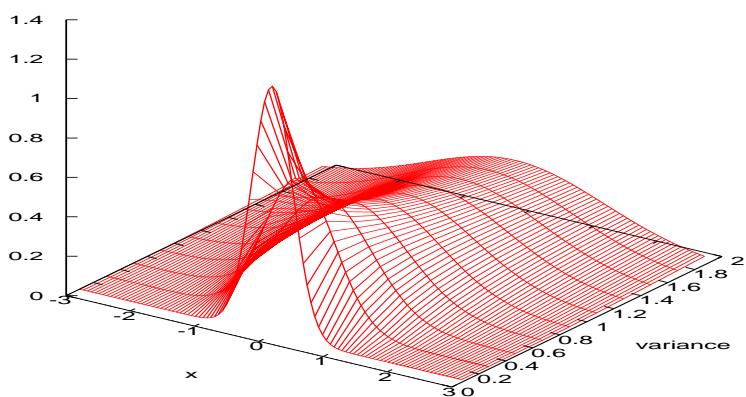
## C.2.2 Gaussian distribution

“Tout le monde y croit cependant, me disait un jour M. Lippmann, car les expérimentateurs s’imagine 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 <sup>3</sup>

Figure C.2: Gaussian pdf with  $\mu = 0$  and  $\sigma \in [0.1, 2]$ .

### Definition C.2.

<sup>3</sup> quote: Poincaré (1912), page 171  
translation: assisted by Google Translate  
image:



The **Gaussian distribution** (or **normal distribution**) has pdf

$$p_x(x) \triangleq \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

A random variable  $X$  with this distribution is denoted

$$X \sim N(\mu, \sigma^2)$$

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

$$Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{u^2}{2}} du$$

### C.2.3 Gamma distribution

**Definition C.3.** <sup>4</sup> Let  $b \in \mathbb{R}$ . The **gamma function**  $\Gamma(b)$  is

$$\text{DEF } \Gamma(b) \triangleq \int_0^\infty x^{b-1} e^{-x} dx$$

**Proposition C.1.** <sup>5</sup> Let  $b \in \mathbb{R}$  and  $n \in \mathbb{N}$ .

$$\begin{aligned} \text{P} & \Gamma(b) = (b-1)\Gamma(b-1) \\ \text{R} & \Gamma(n) = (n-1)! \end{aligned}$$

PROOF: Let

$$\begin{aligned} u &= x^{b-1} & du &= (b-1)x^{b-2} dx \\ dv &= e^{-x} dx & v &= -e^{-x} \end{aligned}$$

$$\begin{aligned} \Gamma(b) &\triangleq \int_0^\infty x^{b-1} e^{-x} dx \\ &= \int_{x=0}^\infty u dv \\ &= uv|_{x=0}^\infty - \int_{x=0}^\infty v du \\ &= -x^{b-1} e^{-x}|_{x=0}^\infty + (b-1) \int_{x=0}^\infty e^{-x} x^{b-1} dx \\ &= (-0+0) + (b-1)\Gamma(b-1) \end{aligned}$$

Note that

$$\Gamma(1) = \int_0^\infty x^{1-1} e^{-x} dx = \int_0^\infty e^{-x} dx = -e^{-x}|_0^\infty = -0+1=1$$

$$\begin{aligned} \Gamma(n) &= (n-1)\Gamma(n-1) \\ &= (n-1)(n-2)\Gamma(n-2) \\ &= (n-1)(n-2)(n-3)\Gamma(n-3) \\ &\vdots \\ &= (n-1)(n-2)(n-3)\cdots(1)\Gamma(1) \\ &= (n-1)(n-2)(n-3)\cdots(1) \\ &\triangleq (n-1)! \end{aligned}$$

<sup>4</sup> Papoulis (1991), page 79, Ross (1998), page 222

<sup>5</sup> Ross (1998), page 223



**Definition C.4.** A **Gamma distribution**  $(b, \lambda)$  has pdf

**D E F**  $p_x(x) \triangleq \frac{\lambda}{\Gamma(b)} e^{-\lambda x} (\lambda x)^{b-1}$

**Theorem C.1.**<sup>6</sup> Let  $X$  and  $Y$  be RANDOM VARIABLES on a PROBABILITY SPACE  $(\Omega, \mathbb{E}, P)$ .

**T H M**  $\left\{ \begin{array}{ll} (A). & X \text{ and } Y \text{ are INDEPENDENT} \\ (B). & X \text{ has GAMMA DISTRIBUTION } (a, \lambda) \quad \text{and} \\ (C). & Y \text{ has GAMMA DISTRIBUTION } (b, \lambda) \quad \text{and} \\ (D). & Z \triangleq X + Y \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} Z \text{ has Gamma distribution} \\ (a+b, \lambda). \end{array} \right\}$

PROOF:

$$p_z(z) = p_x(z) \star p_y(z)$$

$$= \int_{u \in \mathbb{R}} p_x(u) p_y(z-u) du$$

by definition of *convolution* (Definition P.3 page 340)

$$= \int_0^z \frac{1}{\Gamma(a)} \lambda e^{-\lambda u} (\lambda u)^{a-1} \frac{1}{\Gamma(b)} \lambda e^{-\lambda(z-u)} (\lambda(z-u))^{b-1} du$$

$$= \frac{1}{\Gamma(a)} \frac{1}{\Gamma(b)} e^{-\lambda z} \lambda^{1+1+a-1+b-1} \int_0^z u^{a-1} (z-u)^{b-1} du$$

$$= \frac{1}{\Gamma(a)} \frac{1}{\Gamma(b)} \lambda e^{-\lambda z} \lambda^{a+b-1} \int_0^1 (vz)^{a-1} (z-vz)^{b-1} z dv$$

$$= \frac{1}{\Gamma(a)} \frac{1}{\Gamma(b)} \lambda e^{-\lambda z} \lambda^{a+b-1} z^{a-1+b-1+1} \int_0^1 v^{a-1} (1-v)^{b-1} dv$$

$$= \left[ \frac{1}{\Gamma(a)} \frac{1}{\Gamma(b)} \int_0^1 v^{a-1} (1-v)^{b-1} dv \right] \lambda e^{-\lambda z} (\lambda z)^{a+b-1}$$

$$= C \lambda e^{-\lambda z} (\lambda z)^{a+b-1}$$

where  $C$  is some constant

$$= \frac{\lambda}{\Gamma(a+b)} e^{-\lambda z} (\lambda z)^{a+b-1}$$

$C$  must be the value that makes  $\int_z p_z(z) = 1$

$\Rightarrow p_z(z)$  is a  $(a+b, \lambda)$  Gamma distribution



## C.2.4 Chi-squared distributions

**Definition C.5.**<sup>7</sup> Let  $p(x)$  be a PROBABILITY DENSITY FUNCTION on a PROBABILITY SPACE  $(\Omega, \mathbb{E}, P)$ .

**D E F**  $p(x)$  is a **chi-square distribution** if

$$p(x) \triangleq \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{\sqrt{2\pi\sigma^2 x}} \exp\left[-\frac{x}{2\sigma^2}\right] & \text{if } x \geq 0 \end{cases} \quad \text{for } \sigma > 0$$

**Theorem C.2.**<sup>8</sup>

<sup>6</sup> Ross (1998), page 266

<sup>7</sup> Proakis (2001), page 41, Papoulis (1990) page 219 (7-4 Special Distributions of Statistics, (7-78))

<sup>8</sup> Ross (1998), page 267

THM

The following distributions are equivalent:

- (1). chi-squared distribution and
- (2). distribution of  $X^2$  where  $X \sim N(0, \sigma^2)$  and
- (3). Gamma distribution  $\left(\frac{1}{2}, \frac{1}{2\sigma^2}\right)$

PROOF:

1. Proof that  $X^2$  has chi-squared distribution:

$$\begin{aligned} p_Y(y) &= \frac{1}{2\sqrt{y}} \left[ p_X(-\sqrt{y}) + p_X(\sqrt{y}) \right] && \text{by Corollary 5.3 page 33} \\ &= \frac{1}{2\sqrt{y}} \left[ \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{(-\sqrt{y}-0)^2}{2\sigma^2} + \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{(+\sqrt{y}-0)^2}{2\sigma^2} \right] \\ &= \frac{1}{2\sqrt{y}} \left[ 2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{y}{2\sigma^2} \right] \\ &= \frac{1}{\sqrt{2\pi\sigma^2}y} \exp -\frac{y}{2\sigma^2} \end{aligned}$$

2. Proof that chi-distribution is a Gamma distribution  $(b, \lambda)$ :

$$\begin{aligned} b &\triangleq \frac{1}{2} \\ \lambda &\triangleq \frac{1}{2\sigma^2} \end{aligned}$$

$$\begin{aligned} \frac{1}{\sqrt{2\pi\sigma^2}y} \exp -\frac{y}{2\sigma^2} &= \frac{1}{\sqrt{\pi}} \lambda^{1/2} \lambda^{1/2} (\lambda y)^{-1/2} e^{-\lambda y} \\ &= \frac{\lambda}{\sqrt{\pi}} (\lambda y)^{b-1} e^{-\lambda y} \end{aligned}$$



**Definition C.6.** <sup>9</sup> The Chi-squared distribution with  $n$  degrees of freedom has pdf

$$\begin{array}{|c|} \hline \text{DEF} \\ \hline \end{array} p_Y(y) \triangleq \begin{cases} 0 & : y < 0 \\ \frac{1}{2\sigma^2 \Gamma(n/2)} \left(\frac{y}{2\sigma^2}\right)^{\frac{n}{2}-1} \exp -\frac{y}{2\sigma^2} & : y \geq 0 \end{cases}$$

**Theorem C.3.** <sup>10</sup> 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  $\{X_k | X_k \sim N(0, \sigma^2), k = 1, 2, \dots, n\}$  are independent random variables.

3. Gamma distribution  $\left(\frac{n}{2}, \frac{1}{2\sigma^2}\right)$ .

<sup>9</sup> Proakis (2001), page 41

<sup>10</sup> Ross (1998), page 267

PROOF:

- Prove chi-squared distribution with  $n$  degrees of freedom is the Gamma distribution  $\left(\frac{n}{2}, \frac{1}{2\sigma^2}\right)$ :

$$\begin{aligned} \lambda &\triangleq \frac{1}{2\sigma^2} \\ b &\triangleq \frac{1}{2} \\ \frac{1}{2\sigma^2 \Gamma(n/2)} \left(\frac{y}{2\sigma^2}\right)^{\frac{n}{2}-1} \exp -\frac{y}{2\sigma^2} &= \frac{\lambda}{\Gamma(nb)} (\lambda y)^{nb-1} \exp -\lambda y \end{aligned}$$

- 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.** <sup>11</sup> A **noncentral chi-square distribution**  $(\mu, \sigma^2)$  has pdf

**D E F**  $p_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2 y}} \exp \frac{y + \mu^2}{-2\sigma^2} \cosh \frac{\mu\sqrt{y}}{\sigma^2}$

**Theorem C.4.**

**T H M**

The following distributions are equivalent:

- (1). NON-CENTRAL CHI-SQUARED DISTRIBUTION  $(\mu, \sigma^2)$
- (2). distribution of  $X^2$  where  $X \sim N(\mu, \sigma^2)$

PROOF:

- Proof that  $Y = X^2$  has a non-central chi-squared distribution:

$$\begin{aligned} p_Y(y) &= \frac{1}{2\sqrt{y}} \left[ p_X(-\sqrt{y}) + p_X(\sqrt{y}) \right] \quad \text{by Corollary 5.3 page 33} \\ &= \frac{1}{2\sqrt{y}} \left[ \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{(-\sqrt{y} - \mu)^2}{2\sigma^2} + \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{(+\sqrt{y} - \mu)^2}{2\sigma^2} \right] \\ &= \frac{1}{2\sqrt{y}} \left[ \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{y + \mu^2}{2\sigma^2} \exp \frac{-2\mu\sqrt{y}}{2\sigma^2} + \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{y + \mu^2}{2\sigma^2} \exp \frac{2\mu\sqrt{y}}{2\sigma^2} \right] \\ &= \frac{1}{\sqrt{2\pi\sigma^2 y}} \exp -\frac{y + \mu^2}{2\sigma^2} \frac{1}{2} \left[ \exp \frac{2\mu\sqrt{y}}{2\sigma^2} + \exp \frac{-2\mu\sqrt{y}}{2\sigma^2} \right] \\ &= \frac{1}{\sqrt{2\pi\sigma^2 y}} \exp \frac{y + \mu^2}{-2\sigma^2} \cosh \frac{\mu\sqrt{y}}{\sigma^2} \end{aligned}$$



<sup>11</sup> Proakis (2001), page 42

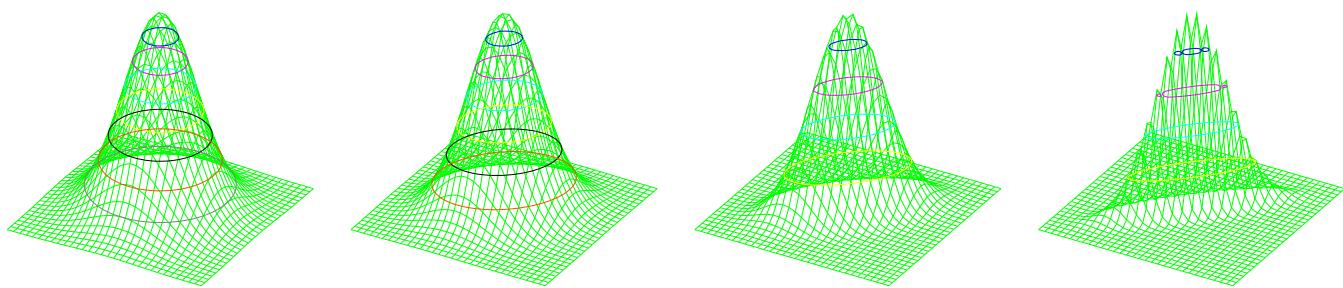


Figure C.3: *Joint Gaussian distributions  $p_{xy}(x, y)$  with varying correlations*

**Definition C.8.** <sup>12</sup> *The  $\alpha$ th-order modified Bessel function of the first kind  $I_\alpha(x)$  is*

$$\text{DEF } I_\alpha(x) = \sum_{k=0}^{\infty} \frac{1}{k! \Gamma(\alpha + k + 1)} \left(\frac{x}{2}\right)^{\alpha+2k}$$

**Definition C.9.** <sup>13</sup> *The noncentral chi-square with  $n$ -degrees of freedom distribution has pdf*

$$\text{DEF } p_Y(y) = \frac{1}{2\sigma^2} \left(\frac{y}{s^2}\right)^{\frac{n-2}{4}} \exp \frac{y+s^2}{-2\sigma^2} I_{n/2-1} \left(\sqrt{y} \frac{s}{\sigma^2}\right) \quad \text{where } s^2 \triangleq \sum_{k=1}^n \mu_k^2$$

## C.2.5 Radial distributions

**Definition C.10.** <sup>14</sup> *The Rayleigh distribution is the pdf*

$$\text{DEF } p_R(r) = \begin{cases} 0 & \text{for } r < 0 \\ \frac{r}{\sigma^2} \exp -\frac{r^2}{2\sigma^2} & \text{for } r \geq 0 \end{cases}$$

Note that by Proposition 5.3, 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  $N(0, \sigma^2)$ .

**Definition C.11.** <sup>15</sup> *The Rice distribution is the pdf*

$$\text{DEF } p_R(r) = \begin{cases} 0 & \text{for } r < 0 \\ \frac{r}{\sigma^2} \exp \frac{r^2+s^2}{-2\sigma^2} I_o \left(\frac{rs}{\sigma^2}\right) & \text{for } r \geq 0 \end{cases}$$

## C.3 Joint Gaussian distributions

**Definition C.12 (Joint Gaussian pdf).** <sup>16</sup>

$$p(x_1, x_2, \dots, x_n) \triangleq \frac{1}{\sqrt{(2\pi)^n |\mathbf{M}|}} \exp -\frac{1}{2}(\mathbf{x} - \mathbf{E}\mathbf{x})^T \mathbf{M}^{-1}(\mathbf{x} - \mathbf{E}\mathbf{x}) \quad (\text{Gaussian joint pdf})$$

DEF

$$\begin{aligned} \mathbf{x} &\triangleq \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\ Z_k &\triangleq X_k - \mathbf{E}X_k \end{aligned} \quad (\text{zero mean random variables})$$

$$\mathbf{M} \triangleq \begin{bmatrix} E[Z_1Z_1] & E[Z_1Z_2] & \cdots & E[Z_1Z_n] \\ E[Z_2Z_1] & E[Z_2Z_2] & \ddots & E[Z_2Z_n] \\ \vdots & \vdots & \ddots & \vdots \\ E[Z_nZ_1] & E[Z_nZ_2] & \cdots & E[Z_nZ_n] \end{bmatrix} \quad (\text{correlation matrix})$$

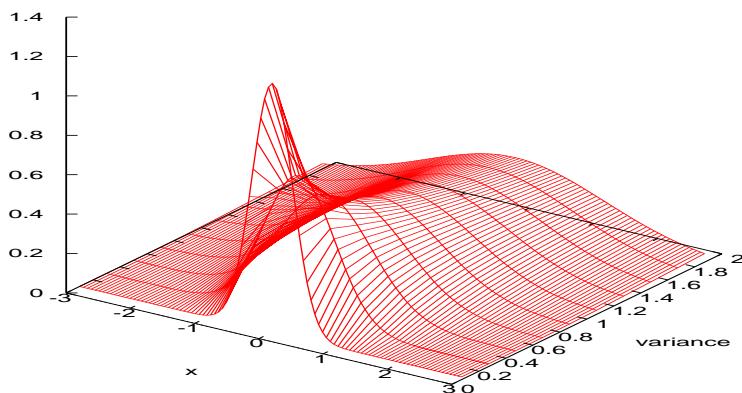


Figure C.4: Gaussian pdf with  $\mu = 0$  and  $\sigma \in [0.1, 2]$ .

*Example C.2* (1 variable joint Gaussian pdf). The **Gaussian distribution** (or **normal distribution**) has pdf

$$\mathbf{E}_{\mathbf{X}} p_x(x) \triangleq \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

<sup>12</sup> Proakis (2001), page 43

<sup>13</sup> Proakis (2001), page 43

<sup>14</sup> Proakis (2001), page 44

<sup>15</sup> Proakis (2001), page 46

<sup>16</sup> 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

$$\begin{aligned}
t &= \arg_t \min_t \left[ \frac{1}{2} \int_t^\infty \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} + \frac{1}{2} \int_{-\infty}^t \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\eta)^2}{2\sigma^2}} \right] \\
&= \arg_t \left\{ \frac{\partial}{\partial t} \left[ \frac{1}{2} \int_t^\infty \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} + \frac{1}{2} \int_{-\infty}^t \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\eta)^2}{2\sigma^2}} \right] = 0 \right\} \\
&= \arg_t \left\{ \frac{1}{2\sqrt{2\pi\sigma^2}} \left[ \frac{\partial}{\partial t} \int_t^\infty e^{\frac{-(x-\mu)^2}{2\sigma^2}} + \frac{\partial}{\partial t} \int_{-\infty}^t e^{\frac{-(x-\eta)^2}{2\sigma^2}} \right] = 0 \right\} \\
&= \arg_t \left\{ \left[ \left( e^{\frac{-(\infty-\mu)^2}{2\sigma^2}} 0 - e^{\frac{-(t-\mu)^2}{2\sigma^2}} 1 \right) + \left( e^{\frac{-(t-\eta)^2}{2\sigma^2}} 1 - e^{\frac{-(\infty-\eta)^2}{2\sigma^2}} 0 \right) \right] = 0 \right\} \\
&= \arg_t \left\{ \left[ e^{\frac{-(t-\eta)^2}{2\sigma^2}} - e^{\frac{-(t-\mu)^2}{2\sigma^2}} \right] = 0 \right\} \\
&= \arg_t \{(t - \eta)^2 = (t - \mu)^2\} \\
&= \frac{\mu + \eta}{2}
\end{aligned}$$

*Example C.3* (2 variable joint Gaussian pdf).

**E  
X**

$$\begin{aligned}
z_1 &\triangleq x_1 - \mathbb{E}x_1 \\
z_2 &\triangleq x_2 - \mathbb{E}x_2 \\
|M| &\triangleq |\mathbb{E}[z_1 z_1] \mathbb{E}[z_2 z_2] - \mathbb{E}[z_1 z_2] \mathbb{E}[z_1 z_2]| \\
p(x_1, x_2) &\triangleq \frac{1}{2\pi\sqrt{|M|}} \exp\left(\frac{z_1^2 \mathbb{E}[z_2 z_2] - 2z_1 z_2 \mathbb{E}[z_1 z_2] + z_2^2 \mathbb{E}[z_1 z_1]}{-2|M|}\right)
\end{aligned}$$



## APPENDIX D

### SPECTRAL THEORY

## D.1 Operator Spectrum

**Definition D.1.** <sup>1</sup> Let  $\mathbf{A} \in \mathcal{B}(X, Y)$  be an operator over the linear spaces  $X = (X, F, \oplus, \otimes)$  and  $Y \triangleq (Y, F, \oplus, \otimes)$ . Let  $\mathcal{N}(\mathbf{A})$  be the NULL SPACE of  $\mathbf{A}$ .

**D E F** An **eigenvalue** of  $\mathbf{A}$  is any value  $\lambda$  such that there exists  $\mathbf{x}$  such that  $\mathbf{Ax} = \lambda\mathbf{x}$ .

The **eigenspace**  $H_\lambda$  of  $\mathbf{A}$  at eigenvalue  $\lambda$  is  $\mathcal{N}(\mathbf{A} - \lambda\mathbf{I})$ .

An **eigenvector** of  $\mathbf{A}$  associated with eigenvalue  $\lambda$  is any element of  $\mathcal{N}(\mathbf{A} - \lambda\mathbf{I})$ .

*Example D.1.* <sup>2</sup> Let  $\mathbf{D}$  be the differential operator.

**E X** The set  $\{e^{\lambda x} | \lambda \in \mathbb{C}\}$  are the eigenvectors of  $\mathbf{D}$ .

$\rho(\mathbf{D}) = \emptyset$  ( $\mathbf{D}$  has no non-spectral points whatsoever)

$\sigma_p(\mathbf{D}) = \sigma(\mathbf{D})$  (the spectrum of  $\mathbf{D}$  is all eigenvalues)

$\sigma_c(\mathbf{D}) = \emptyset$  ( $\mathbf{D}$  has no continuous spectrum)

$\sigma_r(\mathbf{D}) = \emptyset$  ( $\mathbf{D}$  has no resolvent spectrum)

PROOF:

$$\begin{aligned} (\mathbf{D} - \lambda\mathbf{I})e^{\lambda x} &= \mathbf{D}e^{\lambda x} - \lambda\mathbf{I}e^{\lambda x} \\ &= \lambda e^{\lambda x} - \lambda e^{\lambda x} \\ &= 0 \end{aligned} \quad \forall \lambda \in \mathbb{C}$$

This theorem and proof needs more work and investigation to prove/disprove its claims. ⇒

**Definition D.2.** <sup>3</sup> Let  $\mathbf{A} \in \mathcal{B}(X, Y)$  be an operator over the linear spaces  $X = (X, F, \oplus, \otimes)$  and  $Y \triangleq (Y, F, \oplus, \otimes)$ .

<sup>1</sup> Bollobás (1999), page 168, Descartes (1637), Descartes (1954), Cayley (1858), Hilbert (1904), page 67, Hilbert (1912),

<sup>2</sup> Pedersen (2000), page 79

<sup>3</sup> Michel and Herget (1993), page 439

quantity	$\mathcal{N}(\mathbf{A} - \lambda\mathbf{I}) = \{\mathbf{0}\}$ ( $\mathbf{x} = \mathbf{0}$ is the only solution)	$\overline{\mathcal{R}(\mathbf{A} - \lambda\mathbf{I})} = \mathbf{X}$ (dense)	$(\mathbf{A} - \lambda\mathbf{I})^{-1} \in \mathcal{B}(\mathbf{X}, \mathbf{Y})$ (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 D.1: Spectrum of an operator  $\mathbf{A}$ 

The **resolvent set**  $\rho(\mathbf{A})$  of operator  $\mathbf{A}$  is defined as

$$\text{DEF } \rho(\mathbf{A}) \triangleq \left\{ \lambda \in F \mid \begin{array}{l} 1. \quad \mathcal{N}(\mathbf{A} - \lambda\mathbf{I}) = \{\mathbf{0}\} \\ 2. \quad \overline{\mathcal{R}(\mathbf{A} - \lambda\mathbf{I})} = \mathbf{X} \\ 3. \quad (\mathbf{A} - \lambda\mathbf{I})^{-1} \in \mathcal{B}(\mathbf{X}, \mathbf{Y}) \end{array} \right. \begin{array}{l} \text{(no non-zero eigenvectors)} \\ \text{(the range is dense in } \mathbf{X} \text{).} \\ \text{(inverse is continuous/bounded).} \end{array} \right\} \text{and and}$$

The **spectrum**  $\sigma(\mathbf{A})$  of operator  $\mathbf{A}$  is defined as

$$\sigma(\mathbf{A}) \triangleq F \setminus \rho(\mathbf{A}).$$

**Definition D.3.** <sup>4</sup> Let  $\mathbf{A} \in \mathcal{B}(\mathbf{X}, \mathbf{Y})$  be an operator over the linear spaces  $\mathbf{X} = (X, F, \oplus, \otimes)$  and  $\mathbf{Y} \triangleq (Y, F, \oplus, \otimes)$ .

The **point spectrum**  $\sigma_p(\mathbf{A})$  of operator  $\mathbf{A}$  is defined as

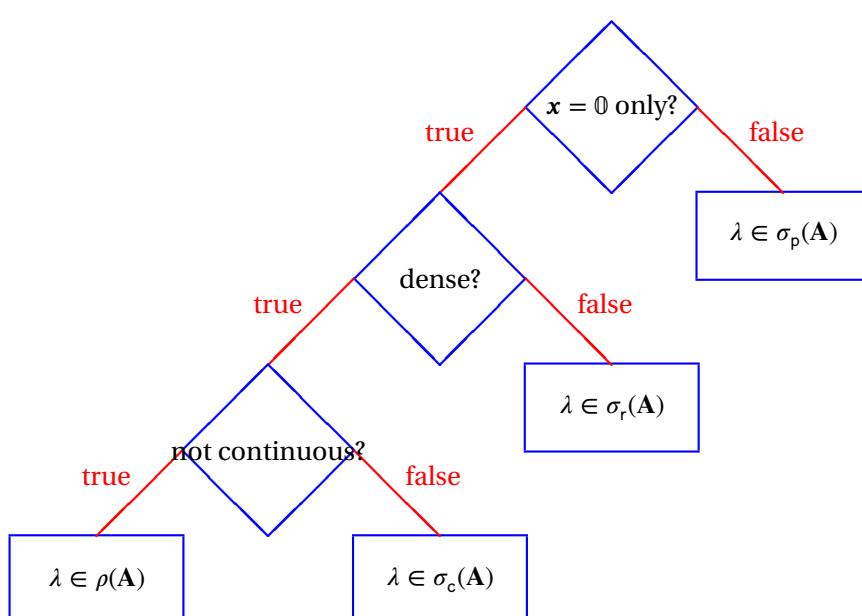
$$\sigma_p(\mathbf{A}) \triangleq \left\{ \lambda \in F \mid \begin{array}{l} 1. \quad \mathcal{N}(\mathbf{A} - \lambda\mathbf{I}) \supsetneq \{\mathbf{0}\} \\ \text{(has non-zero eigenvector)} \end{array} \right\}$$

The **residual spectrum**  $\sigma_r(\mathbf{A})$  of operator  $\mathbf{A}$  is defined as

$$\sigma_r(\mathbf{A}) \triangleq \left\{ \lambda \in F \mid \begin{array}{l} 1. \quad \mathcal{N}(\mathbf{A} - \lambda\mathbf{I}) = \{\mathbf{0}\} \\ 2. \quad \overline{\mathcal{R}(\mathbf{A} - \lambda\mathbf{I})} \neq \mathbf{X} \\ \text{(not dense in } \mathbf{X} \text{—has gaps).} \end{array} \right\} \text{and}$$

The **continuous spectrum**  $\sigma_c(\mathbf{A})$  of operator  $\mathbf{A}$  is defined as

$$\sigma_c(\mathbf{A}) \triangleq \left\{ \lambda \in F \mid \begin{array}{l} 1. \quad \mathcal{N}(\mathbf{A} - \lambda\mathbf{I}) = \{\mathbf{0}\} \\ 2. \quad \overline{\mathcal{R}(\mathbf{A} - \lambda\mathbf{I})} = \mathbf{X} \\ 3. \quad (\mathbf{A} - \lambda\mathbf{I})^{-1} \notin \mathcal{B}(\mathbf{X}, \mathbf{Y}) \end{array} \right. \begin{array}{l} \text{(no non-zero eigenvectors)} \\ \text{(dense in } \mathbf{X}.) \\ \text{(not continuous / not bounded)} \end{array} \right\} \text{and and}$$



The spectral components' definitions are illustrated in the figure to the left and summarized in Table D.1 (page 180). 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.  $\mathbf{B}(\lambda)\mathbf{x} = \mathbf{0}$  if and only if  $\mathbf{x} = \mathbf{0}$
2.  $\mathbf{B}(\lambda)\mathbf{x}$  spans virtually all of  $\mathbf{X}$  as we vary  $\mathbf{x}$
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

<sup>4</sup> [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  $\mathcal{R}(\lambda) = \mathbb{R}$ , and its inverse  $\lambda^{-1}x$  is continuous.

If for some  $\lambda$  the operator  $\mathbf{B}(\lambda)$  does have all these “regular” properties, then that  $\lambda$  part of the *resolvent set* of  $\mathbf{A}$  and  $\lambda$  is called *regular*. However if for some  $\lambda$  the operator  $\mathbf{B}(\lambda)$  fails any of these conditions, then that  $\lambda$  part of the *spectrum* of  $\mathbf{A}$ . And which conditions it fails determines which component of the spectrum it is in.

**Theorem D.1.** <sup>5</sup> Let  $\mathbf{A} \in \mathcal{B}(X, Y)$  be an operator.

T  
H  
M

$$\sigma(\mathbf{A}) = \sigma_p(\mathbf{A}) \cup \sigma_c(\mathbf{A}) \cup \sigma_r(\mathbf{A})$$

**Theorem D.2** (Spectral Theorem). <sup>6</sup> Let  $\mathbf{N} \in Y^X$  be an operator.

T  
H  
M

$$\left. \begin{array}{l} (A). \underbrace{\mathbf{N}^* \mathbf{N} = \mathbf{N} \mathbf{N}^*}_{\mathbf{N} \text{ is NORMAL}} \\ (B). \mathbf{N} \text{ is COMPACT} \end{array} \right\} \Rightarrow$$

$$\left\{ \begin{array}{l} (1). \mathbf{N} = \sum_n \lambda_n \mathbf{P}_n \\ (2). \sum_n \mathbf{P}_n = \mathbf{I} \\ (3). \mathbf{P}_n \mathbf{P}_m = \bar{\delta}_{n-m} \mathbf{P}_n \\ (4). \dim(\mathcal{H}_n) < \infty \\ (5). |\{\lambda_n | \lambda_n \neq 0\}| \text{ is COUNTABLY INFINITE} \end{array} \right.$$

where

$$\begin{aligned} (\lambda_n)_{n \in \mathbb{Z}} &\triangleq \sigma_p(\mathbf{N}) && \text{(eigenvalues of } \mathbf{N}) \\ \mathcal{H}_n &\triangleq \mathcal{N}(\mathbf{N} - \lambda_n \mathbf{I}) && \text{(\lambda}_n \text{ is the eigenspace of } \mathbf{N} \text{ at } \lambda_n \text{ in } Y) \\ \mathbf{H}_n &= \mathbf{P}_n Y && \text{(\mathbf{P}_n \text{ is the projection operator that generates } \mathcal{H}_n)} \end{aligned}$$

## D.2 Fredholm kernels

**Definition D.4.** <sup>7</sup>

D  
E  
F

A **Fredholm operator**  $\mathbf{K}$  is defined as

$$[\mathbf{K}\mathbf{f}](t) \triangleq \underbrace{\int_a^b \kappa(t, s)\mathbf{f}(s) ds}_{\text{kernel}} \quad \forall \mathbf{f} \in L_2([a, b])$$

*Fredholm integral equation of the first kind*<sup>8</sup>

*Example D.2.* Examples of Fredholm operators include

- |                              |   |                                |
|------------------------------|---|--------------------------------|
| 1. Fourier Transform         | $[\tilde{\mathbf{F}}\mathbf{x}](f) = \int_t x(t)e^{-i2\pi ft} dt$                     | $\kappa(t, f) = e^{-i2\pi ft}$ |
| 2. Inverse Fourier Transform | $[\tilde{\mathbf{F}}^{-1}\tilde{\mathbf{x}}](t) = \int_f \tilde{x}(f)e^{i2\pi ft} df$ | $\kappa(f, t) = e^{i2\pi ft}$  |
| 3. Laplace operator          | $[\mathbf{L}\mathbf{x}](s) = \int_t x(t)e^{-st} dt$                                   | $\kappa(t, s) = e^{-st}$       |
| 4. autocorrelation operator  | $[\mathbf{R}\mathbf{x}](t) = \int_s R(t, s)x(s) ds$                                   | $\kappa(t, s) = R(t, s)$       |

**Theorem D.3.** Let  $\mathbf{K}$  be a Fredholm operator with kernel  $\kappa(t, s)$  and adjoint  $\mathbf{K}^*$ .

T  
H  
M

$$[\mathbf{K}\mathbf{f}](t) = \int_A \kappa(t, s)\mathbf{f}(s) ds \iff [\mathbf{K}^*\mathbf{f}](t) = \int_A \kappa^*(s, t)\mathbf{f}(s) ds$$

<sup>5</sup> Michel and Herget (1993), page 440

<sup>6</sup> Michel and Herget (1993), page 457, Bollobás (1999), page 200, Hilbert (1906), Hilbert (1912), von Neumann (1929), de Witt (1659)

<sup>7</sup> Michel and Herget (1993), page 425

<sup>8</sup>The equation  $\int_u \kappa(t, s)\mathbf{f}(s) ds$  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:

$$\begin{aligned}
 [\mathbf{K}\mathbf{f}](t) &= \int_A \kappa(t, s)\mathbf{f}(s) \, ds \\
 \iff \langle [\mathbf{K}\mathbf{f}](t) | g(t) \rangle &= \left\langle \int_s \kappa(t, s)\mathbf{f}(s) \, ds | g(t) \right\rangle \quad \text{by left hypothesis} \\
 &= \int_s \mathbf{f}(s) \langle \kappa(t, s) | g(t) \rangle \, ds \quad \text{by additivity property of } \langle \Delta | \nabla \rangle \text{ (Definition K.1 page 259)} \\
 &= \int_s \mathbf{f}(s) \langle g(t) | \kappa(t, s) \rangle^* \, ds \quad \text{by conjugate symmetry property of } \langle \Delta | \nabla \rangle \text{ (Definition K.1 page 259)} \\
 &= \langle \mathbf{f}(s) | \langle g(t) | \kappa(t, s) \rangle \rangle \quad \text{by local definition of } \langle \Delta | \nabla \rangle \\
 &= \left\langle \mathbf{f}(s) | \underbrace{\int_t \kappa^*(t, s)g(t) \, dt}_{[\mathbf{K}^*\mathbf{g}](s)} \right\rangle \quad \text{by local definition of } \langle \Delta | \nabla \rangle \\
 \iff [\mathbf{K}^*\mathbf{g}](s) &= \int_A \kappa^*(t, s)g(t) \, dt \quad \text{by right hypothesis} \\
 \iff [\mathbf{K}^*\mathbf{g}](\sigma) &= \int_A \kappa^*(\tau, \sigma)g(\tau) \, d\tau \quad \text{by change of variable: } \tau = t, \sigma = s \\
 \iff [\mathbf{K}^*\mathbf{f}](t) &= \int_A \kappa^*(s, t)\mathbf{f}(s) \, ds \quad \text{by change of variable: } t = \sigma, s = \tau, \mathbf{f} = \mathbf{g}
 \end{aligned}$$

⇒

**Corollary D.1.** <sup>9</sup> Let  $\mathbf{K}$  be an Fredholm operator with kernel  $\kappa(t, s)$  and adjoint  $\mathbf{K}^*$ .

C  
O  
R

$$\mathbf{K} = \mathbf{K}^* \iff \underbrace{\kappa(t, s)}_{\text{kernel is conjugate symmetric}} = \underbrace{\kappa^*(s, t)}_{\text{kernel is conjugate symmetric}}$$

PROOF:

$$\begin{aligned}
 \mathbf{K} = \mathbf{K}^* &\iff \int_A \kappa(t, s)\mathbf{f}(s) \, ds = \int_A \kappa^*(s, t)\mathbf{f}(s) \, ds \quad \text{by Theorem D.3 page 181} \\
 &\iff \kappa(t, s) = \kappa^*(s, t)
 \end{aligned}$$

⇒

**Theorem D.4** (Mercer's Theorem). <sup>10</sup> Let  $\mathbf{K}$  be an Fredholm operator with kernel  $\kappa(t, s)$  and eigen-system  $(\lambda_n, \phi_n(t))_{n \in \mathbb{Z}}$ .

T  
H  
M

$$\left\{ \begin{array}{l} (A). \underbrace{\int_a^b \int_a^b \kappa(t, s)\mathbf{f}(t)\mathbf{f}^*(s) \, dt \geq 0}_{\text{positive}} \quad \text{and} \\ (B). \kappa(t, s) \text{ is CONTINUOUS on} \\ [a : b] \times [a : b] \end{array} \right\} \implies \left\{ \begin{array}{l} (1). \kappa(t, s) = \sum_n \lambda_n \phi_n(t) \phi_n^*(s) \quad \text{and} \\ (2). \kappa(t, s) \text{ CONVERGES ABSOLUTELY} \\ \text{and UNIFORMLY on} \\ [a : b] \times [a : b] \end{array} \right\}$$

<sup>9</sup> Michel and Herget (1993), page 430

<sup>10</sup> Gohberg et al. (2003), page 198, Courant and Hilbert (1930), pages 138–140, Mercer (1909), page 439

## APPENDIX E

### 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 natural 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.<sup>1</sup>

## E.1 First derivative of a vector with respect to a vector

### Definition E.1.

$\mathbf{x}$  is a vector with the following properties:

$$1. \quad \mathbf{x} \triangleq \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad (\text{n element column vector})$$

$$2. \quad \frac{\partial}{\partial x_k} x_j = \delta_{kj} \quad ((x_1, x_2, \dots, x_n) \text{ are mutually independent})$$

**Definition E.2 (Jacobian matrix).** <sup>2</sup> The gradient of  $\mathbf{y}$  with respect to  $\mathbf{x}$ , as well as the gradient of  $\mathbf{y}^T$  with respect to  $\mathbf{x}$ , is defined as

<sup>1</sup> [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 \(1999\)](#)

<sup>2</sup> [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”)

**D E F**  $\frac{\partial \mathbf{y}}{\partial \mathbf{x}} \triangleq \frac{\partial \mathbf{y}^T}{\partial \mathbf{x}} \triangleq \underbrace{\begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} & \dots & \frac{\partial y_m}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} & \dots & \frac{\partial y_m}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial x_n} & \frac{\partial y_2}{\partial x_n} & \dots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}}_{n \times m \text{ matrix}}$   $\forall \mathbf{y} \in \mathbb{C}^m$

*Remark E.1.* Depending on whether  $\mathbf{x}$  and  $\mathbf{y}$  are scalars or vectors,  $\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$  takes on the following forms:<sup>3</sup>

	$y$ scalar	$y$ vector
$x$ scalar	$\frac{\partial y}{\partial x}$	$\frac{\partial \mathbf{y}}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x} & \frac{\partial y_2}{\partial x} & \dots & \frac{\partial y_m}{\partial x} \end{bmatrix}$
$x$ vector	$\frac{\partial y}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{bmatrix}$	$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} \triangleq \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} & \dots & \frac{\partial y_m}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} & \dots & \frac{\partial y_m}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial x_n} & \frac{\partial y_2}{\partial x_n} & \dots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}$

**Lemma E.1.** Let  $\mathbf{x} \in \mathbb{R}^n$  be a vector. Then

**L E M**  $\frac{\partial}{\partial x_k} x_i x_j = \bar{\delta}_{ik} x_j + \bar{\delta}_{jk} x_i = \begin{cases} 2x_k & \text{for } i = j = k \\ x_j & \text{for } i = k \text{ and } j \neq k \\ x_i & \text{for } i \neq k \text{ and } j = k \\ 0 & \text{otherwise} \end{cases}$

**Lemma E.2.**

**L E M**  $(\mathbf{x}^H \mathbf{A} \mathbf{x}) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i^* x_j \quad \forall \begin{array}{l} \mathbf{A} \in (\mathbb{C}^n \times \mathbb{C}^n) \quad (n \times n \text{ array}) \\ \mathbf{x} \in \mathbb{C}^n \quad (n \text{ element column vector}) \end{array}$  and

PROOF:

$$\begin{aligned}
 \mathbf{x}^H \mathbf{A} \mathbf{x} &\triangleq [x_1 \ x_2 \ \dots \ x_n]^* \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{by definitions of } \mathbf{A} \text{ and } \mathbf{x} \\
 &= [x_1 \ x_2 \ \dots \ x_n]^* \sum_{i=1}^n x_i \begin{bmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{ni} \end{bmatrix} \\
 &= \sum_{i=1}^n x_i [x_1 \ x_2 \ \dots \ x_n]^* \begin{bmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{ni} \end{bmatrix} \\
 &= \sum_{i=1}^n x_i \sum_{j=1}^n a_{ji} x_j^*
 \end{aligned}$$

<sup>3</sup>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 E.3.****L  
E  
M**

$$\frac{\partial}{\partial \mathbf{x}} [a(\mathbf{x}) b(\mathbf{x})] = a(\mathbf{x}) \left[ \frac{\partial}{\partial \mathbf{x}} b(\mathbf{x}) \right] + \left[ \frac{\partial}{\partial \mathbf{x}} a(\mathbf{x}) \right] b(\mathbf{x})$$

$\underbrace{\forall a, b : \mathbb{R}^n \rightarrow \mathbb{R}}$

$a(\mathbf{x}), b(\mathbf{x})$  are functions from a vector  $\mathbf{x}$  to a scalar in  $\mathbb{R}$

PROOF:

$$\begin{aligned}
 \frac{\partial}{\partial \mathbf{x}} [a(\mathbf{x}) b(\mathbf{x})] &= \begin{bmatrix} \frac{\partial}{\partial x_1} [a(\mathbf{x}) b(\mathbf{x})] \\ \frac{\partial}{\partial x_2} [a(\mathbf{x}) b(\mathbf{x})] \\ \vdots \\ \frac{\partial}{\partial x_n} [a(\mathbf{x}) b(\mathbf{x})] \end{bmatrix} && \text{by definition of } \frac{\partial}{\partial \mathbf{x}} && (\text{Definition E.2 page 183}) \\
 &= \begin{bmatrix} a(\mathbf{x}) \frac{\partial b(\mathbf{x})}{\partial x_1} + b(\mathbf{x}) \frac{\partial a(\mathbf{x})}{\partial x_1} \\ a(\mathbf{x}) \frac{\partial b(\mathbf{x})}{\partial x_2} + b(\mathbf{x}) \frac{\partial a(\mathbf{x})}{\partial x_2} \\ \vdots \\ a(\mathbf{x}) \frac{\partial b(\mathbf{x})}{\partial x_n} + b(\mathbf{x}) \frac{\partial a(\mathbf{x})}{\partial x_n} \end{bmatrix} && \text{by linearity of } \frac{\partial}{\partial \mathbf{x}} \\
 &= \begin{bmatrix} a(\mathbf{x}) \frac{\partial b(\mathbf{x})}{\partial x_1} \\ a(\mathbf{x}) \frac{\partial b(\mathbf{x})}{\partial x_2} \\ \vdots \\ a(\mathbf{x}) \frac{\partial b(\mathbf{x})}{\partial x_n} \end{bmatrix} + \begin{bmatrix} \frac{\partial a(\mathbf{x})}{\partial x_1} b(\mathbf{x}) \\ \frac{\partial a(\mathbf{x})}{\partial x_2} b(\mathbf{x}) \\ \vdots \\ \frac{\partial a(\mathbf{x})}{\partial x_n} b(\mathbf{x}) \end{bmatrix} && \text{by linearity of vector addition} \\
 &= a(\mathbf{x}) \left[ \frac{\partial b(\mathbf{x})}{\partial \mathbf{x}} \right] + \left[ \frac{\partial a(\mathbf{x})}{\partial \mathbf{x}} \right] b(\mathbf{x})
 \end{aligned}$$

**Theorem E.1.**<sup>4</sup>**L  
E  
M**

$$\frac{\partial}{\partial \mathbf{x}} \mathbf{x} = \mathbf{I} \quad \forall \mathbf{x} \in \mathbb{R}^n$$

PROOF:

$$\begin{aligned}
 \frac{\partial}{\partial \mathbf{x}} \mathbf{x} &= \begin{bmatrix} \frac{\partial x_1}{\partial x_1} & \frac{\partial x_2}{\partial x_1} & \dots & \frac{\partial x_n}{\partial x_1} \\ \frac{\partial x_1}{\partial x_2} & \frac{\partial x_1}{\partial x_2} & \dots & \frac{\partial x_n}{\partial x_2} \\ \frac{\partial x_1}{\partial x_2} & \frac{\partial x_2}{\partial x_2} & \dots & \frac{\partial x_n}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_1}{\partial x_n} & \frac{\partial x_2}{\partial x_n} & \dots & \frac{\partial x_n}{\partial x_n} \end{bmatrix} && \text{by Definition E.2 page 183} \\
 &= \begin{bmatrix} \bar{\delta}_{11} & \bar{\delta}_{21} & \dots & \bar{\delta}_{n1} \\ \bar{\delta}_{12} & \bar{\delta}_{22} & \dots & \bar{\delta}_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{\delta}_{1n} & \bar{\delta}_{2n} & \dots & \bar{\delta}_{nn} \end{bmatrix} && \text{by Definition E.1 page 183 (mutual independence property)}
 \end{aligned}$$

<sup>4</sup> Scharf (1991), page 274, Trees (2002), page 1398

$$\begin{aligned}
 &= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} && \text{by definition of kronecker delta function } \delta \\
 &= \mathbf{I} && \text{by definition of identity operator } \mathbf{I}
 \end{aligned}$$

⇒

**Theorem E.2.**

THM	$\frac{\partial}{\partial \mathbf{x}}(\mathbf{A}\mathbf{x}) = \mathbf{A}^T + \sum_{i=1}^n \left( \frac{\partial}{\partial \mathbf{x}} \begin{bmatrix} a_{1i} & a_{2i} & \cdots & a_{mi} \end{bmatrix} \right) \mathbf{x}_i \quad \forall \mathbf{x} \in \mathbb{R}^n, \mathbf{A} \in \mathbb{C}^m \times \mathbb{C}^n$
-----	--

PROOF: Let

$$\mathbf{x} \triangleq \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \mathbf{A} \triangleq \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

$$\frac{\partial}{\partial \mathbf{x}} (\mathbf{A}\mathbf{x}) = \frac{\partial}{\partial \mathbf{x}} \left( \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right)$$

by definition of  $A$  and  $x$

$$= \frac{\partial}{\partial \mathbf{x}} \begin{bmatrix} \sum_{i=1}^n a_{1i}x_i \\ \sum_{i=1}^n a_{2i}x_i \\ \vdots \\ \sum_{i=1}^n a_{mi}x_i \end{bmatrix}$$

by matrix multiplication

$$= \frac{\partial}{\partial \mathbf{x}} \sum_{i=1}^n \begin{bmatrix} a_{1i}x_i \\ a_{2i}x_i \\ \vdots \\ a_{mi}x_i \end{bmatrix}$$

$$= \sum_{i=1}^n \frac{\partial}{\partial \mathbf{x}} \begin{bmatrix} a_{1i}x_i \\ a_{2i}x_i \\ \vdots \\ a_{mi}x_i \end{bmatrix}$$

$$= \sum_{i=1}^n \begin{bmatrix} \frac{\partial a_{1i}x_i}{\partial x_1} & \frac{\partial a_{2i}x_i}{\partial x_1} & \cdots & \frac{\partial a_{mi}x_i}{\partial x_1} \\ \frac{\partial a_{1i}x_i}{\partial x_2} & \frac{\partial a_{2i}x_i}{\partial x_2} & \cdots & \frac{\partial a_{mi}x_i}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial a_{1i}x_i}{\partial x_n} & \frac{\partial a_{2i}x_i}{\partial x_n} & \cdots & \frac{\partial a_{mi}x_i}{\partial x_n} \end{bmatrix}$$

$$= \sum_{i=1}^n \begin{bmatrix} a_{1i} \frac{\partial x_i}{\partial x_1} + \frac{\partial a_{1i}}{\partial x_1} x_i & a_{2i} \frac{\partial x_i}{\partial x_1} + \frac{\partial a_{2i}}{\partial x_1} x_i & \cdots & a_{mi} \frac{\partial x_i}{\partial x_1} + \frac{\partial a_{mi}}{\partial x_1} x_i \\ a_{1i} \frac{\partial x_i}{\partial x_2} + \frac{\partial a_{1i}}{\partial x_2} x_i & a_{2i} \frac{\partial x_i}{\partial x_2} + \frac{\partial a_{2i}}{\partial x_2} x_i & \cdots & a_{mi} \frac{\partial x_i}{\partial x_2} + \frac{\partial a_{mi}}{\partial x_2} x_i \\ \vdots & \vdots & \ddots & \vdots \\ a_{1i} \frac{\partial x_i}{\partial x_n} + \frac{\partial a_{1i}}{\partial x_n} x_i & a_{2i} \frac{\partial x_i}{\partial x_n} + \frac{\partial a_{2i}}{\partial x_n} x_i & \cdots & a_{mi} \frac{\partial x_i}{\partial x_n} + \frac{\partial a_{mi}}{\partial x_n} x_i \end{bmatrix}$$

by Definition E.2 page 183

by Lemma E.3 page 185



$$= \sum_{i=1}^n \begin{bmatrix} a_{1i} \frac{\partial x_i}{\partial x_1} & a_{2i} \frac{\partial x_i}{\partial x_1} & \cdots & a_{mi} \frac{\partial x_i}{\partial x_1} \\ a_{1i} \frac{\partial x_i}{\partial x_2} & a_{2i} \frac{\partial x_i}{\partial x_2} & \cdots & a_{mi} \frac{\partial x_i}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1i} \frac{\partial x_i}{\partial x_n} & a_{2i} \frac{\partial x_i}{\partial x_n} & \cdots & a_{mi} \frac{\partial x_i}{\partial x_n} \end{bmatrix} + \sum_{i=1}^n \begin{bmatrix} \frac{\partial a_{1i}}{\partial x_1} x_i & \frac{\partial a_{2i}}{\partial x_1} x_i & \cdots & \frac{\partial a_{mi}}{\partial x_1} x_i \\ \frac{\partial a_{1i}}{\partial x_2} x_i & \frac{\partial a_{2i}}{\partial x_2} x_i & \cdots & \frac{\partial a_{mi}}{\partial x_2} x_i \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial a_{1i}}{\partial x_n} x_i & \frac{\partial a_{2i}}{\partial x_n} x_i & \cdots & \frac{\partial a_{mi}}{\partial x_n} x_i \end{bmatrix}$$

$$= \sum_{i=1}^n \begin{bmatrix} a_{1i} \bar{\delta}_{i1} & a_{2i} \bar{\delta}_{i1} & \cdots & a_{mi} \bar{\delta}_{i1} \\ a_{1i} \bar{\delta}_{i2} & a_{2i} \bar{\delta}_{i2} & \cdots & a_{mi} \bar{\delta}_{i2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1i} \bar{\delta}_{in} & a_{2i} \bar{\delta}_{in} & \cdots & a_{mi} \bar{\delta}_{in} \end{bmatrix} + \sum_{i=1}^n \left( \frac{\partial}{\partial \mathbf{x}} [ a_{1i} \ a_{2i} \ \cdots \ a_{mi} ] \right) x_i \quad \text{by Lemma E.1}$$

$$= \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix} + \sum_{i=1}^n \left( \frac{\partial}{\partial \mathbf{x}} [ a_{1i} \ a_{2i} \ \cdots \ a_{mi} ] \right) x_i \quad \text{by definition of } \bar{\delta}$$

$$= \mathbf{A}^T + \sum_{i=1}^n \left( \frac{\partial}{\partial \mathbf{x}} [ a_{1i} \ a_{2i} \ \cdots \ a_{mi} ] \right) x_i$$

⇒

**Theorem E.3** (Affine equations). <sup>5</sup>

THM	<b>A and B are independent of x</b> ⇒ $\begin{cases} \frac{\partial}{\partial \mathbf{x}}(\mathbf{Ax}) = \mathbf{A}^T & \forall \mathbf{x} \in \mathbb{R}^n, \mathbf{A} \in \mathbb{C}^m \times \mathbb{C}^n \\ \frac{\partial}{\partial \mathbf{x}}(\mathbf{x}^T \mathbf{B}) = \mathbf{B} & \forall \mathbf{x} \in \mathbb{R}^n, \mathbf{B} \in \mathbb{C}^n \times \mathbb{C}^m \end{cases}$
-----	--

PROOF: Let  $\mathbf{B} \triangleq \mathbf{A}^T$ .

1. Proof that  $\frac{\partial}{\partial \mathbf{x}}(\mathbf{Ax}) = \mathbf{A}^T$ :

$$\begin{aligned} \frac{\partial}{\partial \mathbf{x}}(\mathbf{Ax}) &= \mathbf{A}^T + \sum_{i=1}^n \left( \frac{\partial}{\partial \mathbf{x}} [ a_{1i} \ a_{2i} \ \cdots \ a_{mi} ] \right) x_i && \text{by Theorem E.2 page 186} \\ &= \mathbf{A}^T + \sum_{i=1}^n \left[ \frac{\partial}{\partial \mathbf{x}} a_{1i} \ \frac{\partial}{\partial \mathbf{x}} a_{2i} \ \cdots \ \frac{\partial}{\partial \mathbf{x}} a_{mi} \right] x_i \\ &= \mathbf{A}^T + \sum_{i=1}^n \left[ \begin{array}{cccc} 0 & 0 & \cdots & 0 \end{array} \right] x_i && \text{by left hypothesis} \\ &= \mathbf{A}^T \end{aligned}$$

2. Proof that  $\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}^T \mathbf{B}) = \mathbf{B}$ :

$$\begin{aligned} \frac{\partial}{\partial \mathbf{x}}(\mathbf{x}^T \mathbf{B}) &= \frac{\partial}{\partial \mathbf{x}}(\mathbf{x}^T \mathbf{A}^T) && \text{by definition of } \mathbf{B} \\ &= \frac{\partial}{\partial \mathbf{x}}[(\mathbf{Ax})^T] && \\ &= \frac{\partial}{\partial \mathbf{x}}(\mathbf{Ax}) && \text{by Definition E.2 page 183} \\ &= \mathbf{A}^T && \text{by Theorem E.3 page 187} \\ &= \mathbf{B} && \text{by definition of } \mathbf{B} \end{aligned}$$

⇒

<sup>5</sup> [Graham \(1981\)](#), page 54, [Graham \(2018\)](#), page 549780486824178§“4.2 The Derivatives of Vectors”

**Theorem E.4** (Product rule). <sup>6</sup> Let  $\mathbf{y}$  and  $\mathbf{z}$  be functions of  $\mathbf{x}$  and

T H M	$\frac{\partial}{\partial \mathbf{x}} \mathbf{z}^T \mathbf{y} = \frac{\partial \mathbf{z}}{\partial \mathbf{x}} \mathbf{y} + \frac{\partial \mathbf{y}}{\partial \mathbf{x}} \mathbf{z}$	$\forall \mathbf{x} \in \mathbb{R}^n, \mathbf{y} \in \mathbb{R}^m, \mathbf{z} \in \mathbb{R}^m$
-------------	--	---

PROOF:

$$\begin{aligned}
 \frac{\partial}{\partial \mathbf{x}} \mathbf{z}^T \mathbf{y} &= \frac{\partial}{\partial \mathbf{x}} \sum_{k=1}^m z_k y_k \\
 &= \sum_{k=1}^m \frac{\partial}{\partial \mathbf{x}} z_k y_k \\
 &= \sum_{k=1}^m \frac{\partial z_k}{\partial \mathbf{x}} y_k + \sum_{k=1}^m \frac{\partial y_k}{\partial \mathbf{x}} z_k \quad \text{by Lemma E.3 page 185} \\
 &= \left[ \begin{array}{cccc} \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{array} \right] + \left[ \begin{array}{cccc} \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{array} \right] \\
 &= \left[ \begin{array}{ccc} \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 & \vdots \\ \frac{\partial z_1}{\partial x_1} & \frac{\partial z_2}{\partial x_1} & \cdots & \frac{\partial z_n}{\partial x_1} \end{array} \right] \left[ \begin{array}{c} y_1 \\ y_2 \\ \vdots \\ y_n \end{array} \right] + \left[ \begin{array}{ccc} \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 & \vdots \\ \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_1} \end{array} \right] \left[ \begin{array}{c} z_1 \\ z_2 \\ \vdots \\ z_n \end{array} \right] \\
 &= \frac{\partial \mathbf{z}}{\partial \mathbf{x}} \mathbf{y} + \frac{\partial \mathbf{y}}{\partial \mathbf{x}} \mathbf{z}
 \end{aligned}$$

⇒

**Theorem E.5.**

T H M	$\frac{\partial}{\partial \mathbf{x}} (\mathbf{x}^T \mathbf{A} \mathbf{x}) = \mathbf{A} \mathbf{x} + \mathbf{A}^T \mathbf{x} + \left[ \sum_{i=1}^n \left( \frac{\partial}{\partial \mathbf{x}} [ a_{1i} \ a_{2i} \ \cdots \ a_{ni} ] \right) x_i \right] \mathbf{x}$	$\forall \mathbf{x} \in \mathbb{R}^n, \mathbf{A} \in \mathbb{R}^n \times \mathbb{R}^n$
-------------	--	--

PROOF:

$$\begin{aligned}
 \frac{\partial}{\partial \mathbf{x}} (\mathbf{x}^T \mathbf{A} \mathbf{x}) &= \left[ \frac{\partial}{\partial \mathbf{x}} \mathbf{x} \right] \mathbf{A} \mathbf{x} + \left[ \frac{\partial}{\partial \mathbf{x}} \mathbf{A} \mathbf{x} \right] \mathbf{x} \quad \text{by Theorem E.4 page 188} \\
 &= \mathbf{I} \mathbf{A} \mathbf{x} + \left[ \mathbf{A}^T + \sum_{i=1}^n \left( \frac{\partial}{\partial \mathbf{x}} [ a_{1i} \ a_{2i} \ \cdots \ a_{ni} ] \right) x_i \right] \mathbf{x} \quad \text{by Theorem E.1 and Theorem E.2} \\
 &= \mathbf{A} \mathbf{x} + \mathbf{A}^T \mathbf{x} + \left[ \sum_{i=1}^n \left( \frac{\partial}{\partial \mathbf{x}} [ a_{1i} \ a_{2i} \ \cdots \ a_{ni} ] \right) x_i \right] \mathbf{x} \quad \text{by definition of identity operator } \mathbf{I}
 \end{aligned}$$

⇒

**Theorem E.6** (Quadratic form). <sup>7</sup>

T H M	$\mathbf{A}$ is independent of $\mathbf{x}$ $\implies \frac{\partial}{\partial \mathbf{x}} (\mathbf{x}^T \mathbf{A} \mathbf{x}) = \mathbf{A} \mathbf{x} + \mathbf{A}^T \mathbf{x}$	$\forall \mathbf{x} \in \mathbb{R}^n, \mathbf{A} \in \mathbb{R}^n \times \mathbb{R}^n$
-------------	--	--

<sup>6</sup> Scharf (1991), page 274, Trees (2002), page 1398

<sup>7</sup> Graham (1981), page 54

PROOF:

$$\begin{aligned}\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}^T \mathbf{A} \mathbf{x}) &= \left[ \frac{\partial}{\partial \mathbf{x}} \mathbf{x} \right] \mathbf{A} \mathbf{x} + \left[ \frac{\partial}{\partial \mathbf{x}} \mathbf{A} \mathbf{x} \right] \mathbf{x} && \text{by Theorem E.4 page 188} \\ &= \mathbf{I} \mathbf{A} \mathbf{x} + \mathbf{A}^T \mathbf{x} && \text{by Theorem E.1 page 185 and Theorem E.3 page 187}\end{aligned}$$

**Corollary E.1.**<sup>8</sup>

COR	$\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}^T \mathbf{x}) = 2\mathbf{x} \quad \forall \mathbf{x} \in \mathbb{R}^n$
-----	---



PROOF:

$$\begin{aligned}\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}^T \mathbf{x}) &= \frac{\partial}{\partial \mathbf{x}}(\mathbf{x}^T \mathbf{I} \mathbf{x}) && \text{by property of identity operator } I \\ &= \mathbf{I} \mathbf{x} + \mathbf{I}^T \mathbf{x} && \text{by previous result 3.} \\ &= \mathbf{x} + \mathbf{x} && \text{by property of identity operator } I \\ &= 2\mathbf{x}\end{aligned}$$

**Theorem E.7** (Chain rule).<sup>9</sup> Let  $\mathbf{z}$  be a function of  $\mathbf{y}$  and  $\mathbf{y}$  a function of  $\mathbf{x}$  and

$$\mathbf{x} \triangleq \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \mathbf{y} \triangleq \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \quad \mathbf{z} \triangleq \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_k \end{bmatrix}$$

THM	$\frac{\partial}{\partial \mathbf{x}} \mathbf{z} = \frac{\partial \mathbf{y}}{\partial \mathbf{x}} \frac{\partial \mathbf{z}}{\partial \mathbf{y}}$
-----	---



PROOF:

$$\begin{aligned}\frac{\partial \mathbf{z}}{\partial \mathbf{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 \mathbf{y}}{\partial \mathbf{x}} \frac{\partial \mathbf{z}}{\partial \mathbf{y}}\end{aligned}$$

<sup>8</sup> Graham (1981), page 54<sup>9</sup> Graham (1981), pages 54–55



## E.2 First derivative of a matrix with respect to a scalar

**Definition E.3.** Let  $x \in \mathbb{R}$ ,  $\{y_{jk} \in \mathbb{C} | j = 1, 2, \dots, m; k = 1, 2, \dots, n\}$  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

**D E F**

$$\frac{dY}{dx} \triangleq \underbrace{\begin{bmatrix} \frac{dy_{11}}{dx} & \frac{dy_{12}}{dx} & \cdots & \frac{dy_{1n}}{dx} \\ \frac{dy_{21}}{dx} & \frac{dy_{22}}{dx} & \cdots & \frac{dy_{2n}}{dx} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{dy_{m1}}{dx} & \frac{dy_{m2}}{dx} & \cdots & \frac{dy_{mn}}{dx} \end{bmatrix}}_{m \times n \text{ matrix}}$$

**Theorem E.8.**<sup>10</sup> Let  $x \in \mathbb{R}$ ,  $\{y_{jp} \in \mathbb{C} | j = 1, 2, \dots, m; p = 1, 2, \dots, n\}$ ,  $\{w_{jp} \in \mathbb{C} | j = 1, 2, \dots, n; p = 1, 2, \dots, k\}$ , and

$$Y = \underbrace{\begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mn} \end{bmatrix}}_{m \times n \text{ matrix}} \quad W = \underbrace{\begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1k} \\ w_{21} & w_{22} & \cdots & w_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ w_{p1} & w_{p2} & \cdots & w_{pk} \end{bmatrix}}_{p \times k \text{ matrix}}$$

**T H M**

$\frac{d}{dx}(Y + W) = \frac{d}{dx}Y + \frac{d}{dx}W$	(for $p = m, k = n$ )
$\frac{d}{dx}(YW) = \left(\frac{d}{dx}Y\right)W + Y\left(\frac{d}{dx}W\right)$	(for $p = n$ )
$\frac{d}{dx}(Y^T) = \left(\frac{d}{dx}Y\right)^T$	
$\frac{d}{dx}(Y^{-1}) = -Y^{-1}\left(\frac{d}{dx}Y\right)Y^{-1}$	(for $m = n$ and $Y$ invertible)

PROOF:

$$\begin{aligned} \frac{d}{dx}(Y + W) &= \frac{d}{dx} \left( \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mn} \end{bmatrix} + \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{p1} & w_{p2} & \cdots & w_{pn} \end{bmatrix} \right) \\ &= \frac{d}{dx} \begin{bmatrix} y_{11} + w_{11} & y_{12} + w_{12} & \cdots & y_{1n} + w_{1n} \\ y_{21} + w_{21} & y_{22} + w_{22} & \cdots & y_{2n} + w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} + w_{m1} & y_{m2} + w_{m2} & \cdots & y_{mn} + w_{mn} \end{bmatrix} \end{aligned}$$

<sup>10</sup> Gradshteyn and Ryzhik (1980), pages 1106–1107

$$\begin{aligned}
&= \begin{bmatrix} (y_{11} + w_{11})' & (y_{12} + w_{12})' & \cdots & (y_{1n} + w_{1n})' \\ (y_{21} + w_{21})' & (y_{22} + w_{22})' & \cdots & (y_{2n} + w_{2n})' \\ \vdots & \vdots & \ddots & \vdots \\ (y_{m1} + w_{m1})' & (y_{m2} + w_{m2})' & \cdots & (y_{mn} + w_{mn})' \end{bmatrix} \\
&= \begin{bmatrix} y'_{11} + w'_{11} & y'_{12} + w'_{12} & \cdots & y'_{1n} + w'_{1n} \\ y'_{21} + w'_{21} & y'_{22} + w'_{22} & \cdots & y'_{2n} + w'_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y'_{m1} + w'_{m1} & y'_{m2} + w'_{m2} & \cdots & y'_{mn} + w'_{mn} \end{bmatrix} \\
&= \begin{bmatrix} y'_{11} & y'_{12} & \cdots & y'_{1n} \\ y'_{21} & y'_{22} & \cdots & y'_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y'_{m1} & y'_{m2} & \cdots & y'_{mn} \end{bmatrix} + \begin{bmatrix} w'_{11} & w'_{12} & \cdots & w'_{1n} \\ w'_{21} & w'_{22} & \cdots & w'_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w'_{m1} & w'_{m2} & \cdots & w'_{mn} \end{bmatrix} \\
&= \frac{d}{dx} \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mn} \end{bmatrix} + \frac{d}{dx} \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m1} & w_{m2} & \cdots & w_{mn} \end{bmatrix} \\
&= \frac{d}{dx} Y + \frac{d}{dx} W
\end{aligned}$$

$$\begin{aligned}
\frac{d}{dx}(YW) &= \frac{d}{dx} \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{d}{dx} \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{d}{dx} \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{d}{dx} \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{d}{dx}(y_{1j} w_{j1}) & \frac{d}{dx}(y_{1j} w_{j2}) & \cdots & \frac{d}{dx}(y_{1j} w_{jk}) \\ \frac{d}{dx}(y_{2j} w_{j1}) & \frac{d}{dx}(y_{2j} w_{j2}) & \cdots & \frac{d}{dx}(y_{2j} w_{jk}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{d}{dx}(y_{mj} w_{j1}) & \frac{d}{dx}(y_{mj} w_{j2}) & \cdots & \frac{d}{dx}(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}$$

$$\begin{aligned}
&= \sum_{j=1}^n \left( \begin{bmatrix} y'_{1j} w_{j1} & y'_{1j} w_{j2} & \cdots & y'_{1j} w_{jk} \\ y'_{2j} w_{j1} & y'_{2j} w_{j2} & \cdots & y'_{2j} w_{jk} \\ \vdots & \vdots & \ddots & \vdots \\ y'_{mj} w_{j1} & y'_{mj} w_{j2} & \cdots & y'_{mj} w_{jk} \end{bmatrix} + \begin{bmatrix} y_{1j} w'_{j1} & y_{1j} w'_{j2} & \cdots & y_{1j} w'_{jk} \\ y_{2j} w'_{j1} & y_{2j} w'_{j2} & \cdots & y_{2j} w'_{jk} \\ \vdots & \vdots & \ddots & \vdots \\ y_{mj} w'_{j1} & y_{mj} w'_{j2} & \cdots & y_{mj} w'_{jk} \end{bmatrix} \right) \\
&= \left( \frac{d}{dx} Y \right) W + Y \left( \frac{d}{dx} W \right)
\end{aligned}$$

$$\begin{aligned}
\frac{d}{dx} (Y^T) &= \frac{d}{dx} \left( \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mn} \end{bmatrix}^T \right) \\
&= \frac{d}{dx} \begin{bmatrix} y_{11} & y_{21} & \cdots & y_{n1} \\ y_{12} & y_{22} & \cdots & y_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ y_{1n} & y_{2n} & \cdots & y_{nn} \end{bmatrix} \\
&= \begin{bmatrix} y'_{11} & y'_{21} & \cdots & y'_{n1} \\ y'_{12} & y'_{22} & \cdots & y'_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ y'_{1n} & y'_{2n} & \cdots & y'_{nn} \end{bmatrix} \\
&= \begin{bmatrix} y'_{11} & y'_{12} & \cdots & y'_{1n} \\ y'_{21} & y'_{22} & \cdots & y'_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y'_{m1} & y'_{m2} & \cdots & y'_{mn} \end{bmatrix}^T \\
&= \left( \frac{d}{dx} \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mn} \end{bmatrix} \right)^T
\end{aligned}$$

$$\begin{aligned}
\frac{d}{dx} (Y^{-1}) &= \frac{d}{dx} \frac{\text{adj} Y}{|Y|} \\
&\vdots \\
&\text{no proof at this time} \\
&\vdots \\
&= -Y^{-1} \left( \frac{d}{dx} Y \right) Y^{-1}
\end{aligned}$$

⇒

## E.3 Second derivative of a scalar with respect to a vector

**Definition E.4.**<sup>11</sup> Let

$$\mathbf{x} \triangleq \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

<sup>11</sup> Lieb and Loss (2001), page 240, Horn and Johnson (1990), page 167

The **Hessian matrix** of a scalar  $y$  with respect to the vector  $\mathbf{x}$  is

**D E F**

$$\frac{\partial^2 y}{\partial \mathbf{x}^2} = \frac{\partial}{\partial \mathbf{x}} \left( \frac{\partial y}{\partial \mathbf{x}} \right) = \frac{\partial}{\partial \mathbf{x}} \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{\partial}{\partial x_1} \frac{\partial y}{\partial x_1} & \frac{\partial}{\partial x_1} \frac{\partial y}{\partial x_2} & \cdots & \frac{\partial}{\partial x_1} \frac{\partial y}{\partial x_n} \\ \frac{\partial}{\partial x_2} \frac{\partial y}{\partial x_1} & \frac{\partial}{\partial x_2} \frac{\partial y}{\partial x_2} & \cdots & \frac{\partial}{\partial x_2} \frac{\partial y}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_n} \frac{\partial y}{\partial x_1} & \frac{\partial}{\partial x_n} \frac{\partial y}{\partial x_2} & \cdots & \frac{\partial}{\partial x_n} \frac{\partial y}{\partial x_n} \end{bmatrix}}_{n \times n \text{ matrix}}$$

## E.4 Multiple derivatives of a vector with respect to a scalar

**Definition E.5.** Let

$$\mathbf{y} \triangleq \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

The derivative of a vector  $\mathbf{y}$  with respect to the scalar  $x$  is

**D E F**

$$\begin{bmatrix} \mathbf{y} \\ \frac{d}{dx} \mathbf{y} \\ \frac{d^2}{dx^2} \mathbf{y} \\ \vdots \\ \frac{d^n}{dx^n} \mathbf{y} \end{bmatrix} = \underbrace{\begin{bmatrix} y_1 & y_2 & \cdots & y_m \\ \frac{d}{dx} y_1 & \frac{d}{dx} y_2 & \cdots & \frac{d}{dx} y_m \\ \frac{d^2}{dx^2} y_1 & \frac{d^2}{dx^2} y_2 & \cdots & \frac{d^2}{dx^2} y_m \\ \vdots & \vdots & \ddots & \vdots \\ \frac{d^n}{dx^n} y_1 & \frac{d^n}{dx^n} y_2 & \cdots & \frac{d^n}{dx^n} y_m \end{bmatrix}}_{(n+1) \times m \text{ matrix}}$$



## APPENDIX F

### 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<sup>1</sup>

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 E.1.** <sup>2</sup> Let  $X$  be a set and  $\diamond : X \times X \rightarrow X$  be an operation on  $X$ .

The pair  $(X, \diamond)$  is a **group** if

- |     |   |
|-----|---|
| DEF | 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 E.2.** <sup>3</sup> Let  $+ : X \times X \rightarrow X$  and  $* : X \times X \rightarrow X$  be operations on a set  $X$ . Furthermore, let the operation  $*$  also be represented by juxtaposition as in  $a * b \equiv ab$ .

The triple  $(X, +, *)$  is a **ring** if

- |     |   |
|-----|---|
| DEF | 1. $(X, +)$ is a group. (additive group) and  |
|     | 2. $x(yz) = (xy)z \quad \forall x, y, z \in X$ (ASSOCIATIVE with respect to $*$ ) and           |
|     | 3. $x(y + z) = xy + xz \quad \forall x, y, z \in X$ ( $*$ is LEFT DISTRIBUTIVE over $+$ ) and   |
|     | 4. $(x + y)z = xz + yz \quad \forall x, y, z \in X$ ( $*$ is RIGHT DISTRIBUTIVE over $+$ ). and |

**Definition E.3.** <sup>4</sup>

<sup>1</sup> quote: Cardano (1545), page 1

image: <http://en.wikipedia.org/wiki/Image:Cardano.jpg>

<sup>2</sup> Durbin (2000), page 29

<sup>3</sup> Durbin (2000), pages 114–115

<sup>4</sup> Durbin (2000), page 118

**D E F** A triple  $(X, +, *)$  is a **commutative ring** if

1.  $(X, +, *)$  is a RING and
2.  $xy = yx \quad \forall x, y \in X$  (COMMUTATIVE).

**Definition F.4.** <sup>5</sup> Let  $R$  be a COMMUTATIVE RING (Definition F.3 page 195).

A function  $|\cdot|$  in  $\mathbb{R}$  is an **absolute value** (or **modulus**) if

1.  $|x| \geq 0 \quad x \in \mathbb{R}$  (NON-NEGATIVE) and
2.  $|x| = 0 \iff x = 0 \quad x \in \mathbb{R}$  (NONDEGENERATE) and
3.  $|xy| = |x| \cdot |y| \quad x, y \in \mathbb{R}$  (HOMOGENEOUS / SUBMULTIPLICATIVE) and
4.  $|x + y| \leq |x| + |y| \quad x, y \in \mathbb{R}$  (SUBADDITIVE / TRIANGLE INEQUALITY)

**Definition F.5.** <sup>6</sup>

The structure  $F \triangleq (X, +, \cdot, 0, 1)$  is a **field** if

1.  $(X, +, *)$  is a ring (ring) and
2.  $xy = yx \quad \forall x, y \in X$  (commutative with respect to  $*$ ) and
3.  $(X \setminus \{0\}, *)$  is a group (group with respect to  $*$ ).

**Definition F.6.** <sup>7</sup> Let  $V = (F, +, \cdot)$  be a vector space and  $\otimes : V \times V \rightarrow V$  be a vector-vector multiplication operator.

An **algebra** is any pair  $(V, \otimes)$  that satisfies ( $\otimes$  is represented by juxtaposition)

1.  $(ux)y = u(xy) \quad \forall u, x, y \in V$  (ASSOCIATIVE) and
2.  $u(x + y) = (ux) + (uy) \quad \forall u, x, y \in V$  (LEFT DISTRIBUTIVE) and
3.  $(u + x)y = (uy) + (xy) \quad \forall u, x, y \in V$  (RIGHT DISTRIBUTIVE) and
4.  $\alpha(xy) = (\alpha x)y = x(\alpha y) \quad \forall x, y \in V \text{ and } \alpha \in F$  (SCALAR COMMUTATIVE) .

<sup>5</sup> Cohn (2002) page 312

<sup>6</sup> Durbin (2000), page 123, Weber (1893)

<sup>7</sup> Abramovich and Aliprantis (2002), page 3, Michel and Herget (1993), page 56

## APPENDIX G

### 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.”<sup>1</sup>

Giuseppe Peano (1858–1932), Italian mathematician, credited with being one of the first to introduce the concept of the *linear space* (*vector space*).<sup>1</sup>

## G.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 F page 195) 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 K.1 page 259), or a *norm* giving a *normed linear space* (Definition L.1 page 275).

**Definition G.1.** <sup>2</sup> Let  $(\mathbb{F}, +, \cdot, 0, 1)$  be a FIELD (Definition F.5 page 196). Let  $X$  be a set, let  $+$  be an OPERATOR (Definition O.1 page 307) in  $X^{X^2}$ , and let  $\otimes$  be an operator in  $X^{\mathbb{F} \times X}$ .

<sup>1</sup> quote: Peano (1888b), page ix

image [http://en.wikipedia.org/wiki/File:Giuseppe\\_Peano.jpg](http://en.wikipedia.org/wiki/File:Giuseppe_Peano.jpg), public domain

<sup>2</sup> 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

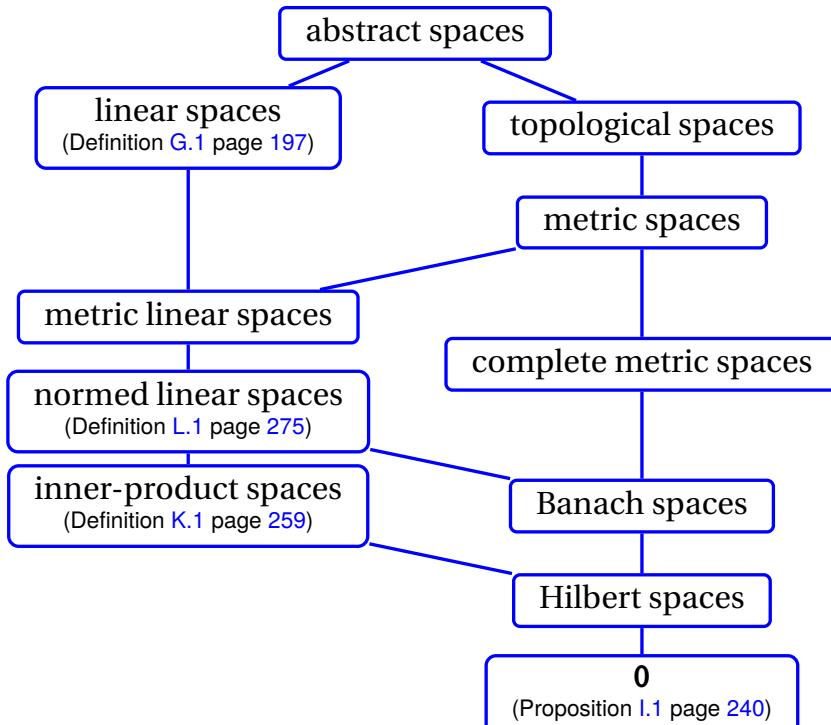


Figure G.1: Lattice of mathematical spaces

The structure  $\Omega \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$  is a **linear space** over  $(\mathbb{F}, +, \cdot, 0, 1)$  if

- |     |   |                               |
|-----|---|-------------------------------|
| DEF | 1. $\exists \mathbb{0} \in X$ such that $x + \mathbb{0} = x \quad \forall x \in X$  | (+ IDENTITY)                  |
|     | 2. $\exists y \in X$ such that $x + y = \mathbb{0} \quad \forall x \in X$   | (+ INVERSE)                   |
|     | 3. $(x + y) + z = x + (y + z) \quad \forall x, y, z \in X$  | (+ is ASSOCIATIVE)            |
|     | 4. $x + y = y + x \quad \forall x, y \in X$   | (+ is COMMUTATIVE)            |
|     | 5. $1 \cdot x = x \quad \forall x \in X$  | (· IDENTITY)                  |
|     | 6. $\alpha \cdot (\beta \cdot x) = (\alpha \cdot \beta) \cdot x \quad \forall \alpha, \beta \in S \text{ and } x \in X$   | (· ASSOCIATES with ·)         |
|     | 7. $\alpha \cdot (x + y) = (\alpha \cdot x) + (\alpha \cdot y) \quad \forall \alpha \in S \text{ and } x, y \in X$        | (· DISTRIBUTES over +)        |
|     | 8. $(\alpha + \beta) \cdot x = (\alpha \cdot x) + (\beta \cdot x) \quad \forall \alpha, \beta \in S \text{ and } x \in X$ | (· 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  $\Omega$  is a **real linear space**. If  $\mathbb{F} \triangleq \mathbb{C}$ , then  $\Omega$  is a **complex linear space**.

**Definition G.2.** Let  $L_1 \triangleq (X_1, +, \cdot, (\mathbb{F}_1, \dot{+}, \dot{\times}))$  and  $L_2 \triangleq (X_2, +, \cdot, (\mathbb{F}_2, \dot{+}, \dot{\times}))$ .

$\Omega_2$  is a **linear subspace** of  $\Omega_1$  if

- |     |  |
|-----|--|
| DEF | 1. $L_1$ is a LINEAR SPACE (Definition G.1 page 197) and |
|     | 2. $L_2$ is a LINEAR SPACE (Definition G.1 page 197) and |
|     | 3. $\mathbb{F}_2 \subseteq \mathbb{F}_1$ and             |
|     | 4. $X_2 \subseteq X_1$ and                               |

**Remark G.1.**<sup>3</sup> By the first four conditions (\*) listed in Definition G.1,  $(X, +)$  is a **commutative group** (or **abelian group**).

<sup>3</sup> 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.  $\alpha x$  is equivalent to  $\alpha \cdot x$ ). 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.<sup>4</sup>

*Example G.1* (tuples in  $\mathbb{F}^N$ ).<sup>5</sup> Let  $(x_n)_1^N$  be an  $N$ -tuple (Definition R.1 page 357) over a field (Definition F.5 page 196)  $(\mathbb{F}, +, \cdot, 0, 1)$ .

EX	$\begin{aligned} X &\triangleq \{(x_n)_1^N \mid x_n \in \mathbb{F}\} && \text{and} \\ (x_n)_1^N + (y_n)_1^N &\triangleq (x_n + y_n)_1^N && \forall [x_n] \in X \quad \text{and} \\ \alpha \cdot (x_n)_1^N &\triangleq (\alpha \dot{x}_n)_1^N && \forall [x_n] \in X, \alpha \in \mathbb{F}. \end{aligned}$
----	--

Then the structure  $(X, +, \cdot, (\mathbb{F}, +, \cdot))$  is a *linear space*.

*Example G.2* (real numbers).<sup>6</sup> Let  $(\mathbb{R}, +, \cdot, 0, 1)$  be the field of real numbers.

EX	<p>The structure <math>(\mathbb{R}, +, \cdot, (\mathbb{R}, +, \cdot))</math> is a <i>linear space</i>. That is, the field of real numbers forms a linear space over itself.</p>
----	---

*Example G.3* (functions).<sup>7</sup> 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	$\begin{aligned} [f + g](x) &\triangleq f(x) + g(x) && \forall f, g \in Y^X \quad (\text{pointwise addition}) \quad \text{and} \\ [\alpha \cdot f](x) &\triangleq \alpha \cdot [f(x)] && \forall f \in Y^X, \alpha \in \mathbb{F}. \end{aligned}$
----	---

Then the structure  $(Y^X, +, \cdot, (\mathbb{F}, +, \cdot))$  is a *linear space*.

*Example G.4* (functions onto  $\mathbb{F}$ ).<sup>8</sup> Let  $(\mathbb{F}, +, \cdot, 0, 1)$  be a field. Let  $\mathbb{F}^X$  be the set of all functions with domain  $X$  and range  $\mathbb{F}$ .

EX	$\begin{aligned} [f + g](x) &\triangleq f(x) + g(x) && \forall f, g \in \mathbb{F}^X \quad (\text{pointwise addition}) \quad \text{and} \\ [\alpha \cdot f](x) &\triangleq \alpha \cdot [f(x)] && \forall f \in \mathbb{F}^X, \alpha \in \mathbb{F}. \end{aligned}$
----	---

Then the structure  $(\mathbb{F}^X, +, \cdot, (\mathbb{F}, +, \cdot))$  is a *linear space*.

**Theorem G.1** (Additive identity properties).<sup>9</sup> Let  $(X, +, \cdot, (\mathbb{F}, +, \cdot))$  be a linear space,  $0$  the ADDITIVE IDENTITY ELEMENT (Definition F.1 page 195) with respect to  $+$ , and  $\mathbb{0}$  the ADDITIVE IDENTITY ELEMENT with respect to  $\cdot$ .

THM	$\begin{aligned} 1. \quad 0x &= \mathbb{0} && \forall x \in X \\ 2. \quad \alpha \mathbb{0} &= \mathbb{0} && \forall \alpha \in \mathbb{F} \\ 3. \quad \alpha x &= \mathbb{0} &\implies \alpha = 0 \text{ or } x = \mathbb{0} \\ 4. \quad x + x &= x &\implies x = \mathbb{0} \\ 5. \quad \alpha \neq 0 \text{ and } x \neq \mathbb{0} &\implies \alpha x \neq \mathbb{0} \end{aligned}$
-----	--

PROOF:

<sup>4</sup> *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.

<sup>5</sup> Kubrusly (2001) page 41 (Example 2D)

<sup>6</sup> Kubrusly (2001) page 41 (Example 2D), Hamel (1905)

<sup>7</sup> Kubrusly (2001) page 42 (Example 2F)

<sup>8</sup> Kubrusly (2001) page 41 (Example 2E)

<sup>9</sup> Berberian (1961) page 6 (Theorem 1), Michel and Herget (1993) page 77

1. Proof that  $0x = \emptyset$ :

$$\begin{aligned}
 0x &= 0x + 0\emptyset && \text{by definition of } + \text{ additive identity element} \\
 &= 0x + 0x + (-0x) && \text{by definition of } + \text{ additive inverse} \\
 &= (0 + 0)x + (-0 \cdot x) && \text{by definition of } + \text{ additive identity element} \\
 &= 0x + (-0x) && \text{by Definition G.1 property 4} \\
 &= \emptyset && \text{by definition of } + \text{ additive identity element}
 \end{aligned}$$

2. Proof that  $\alpha\emptyset = \emptyset$ :

$$\begin{aligned}
 \alpha\emptyset &= \alpha(0x) && \text{by item 1} \\
 &= (\alpha 0)x && \text{by Definition G.1 property 6} \\
 &= 0x && \\
 &= \emptyset && \text{by item 1}
 \end{aligned}$$

3. Proof that  $\alpha \neq 0$  and  $x \neq \emptyset \implies \alpha x \neq \emptyset$ : Suppose  $\alpha x = \emptyset$ . Then

$$\begin{aligned}
 x &= \left(\frac{1}{\alpha}\right)x \\
 &= \frac{1}{\alpha}(\alpha x) && \text{by item 2} \\
 &= \frac{1}{\alpha}\emptyset \\
 &= \emptyset \\
 &\implies x = \emptyset
 \end{aligned}$$

This is a *contradiction* and so  $\alpha x \neq \emptyset$ .

4. Proof that  $\alpha x = \emptyset \implies \alpha = 0$  or  $x = \emptyset$ : contrapositive argument of item 3

5. Proof that  $x + x = x \implies x = \emptyset$ :

$$\begin{aligned}
 x &= x + \emptyset && \text{by } \textit{additive identity property} \text{ (Definition G.1 page 197)} \\
 &= x + [x + (-x)] && \text{by } \textit{additive inverse property} \text{ (Definition G.1 page 197)} \\
 &= [x + x] + (-x) && \text{by } \textit{associative property} \text{ (Definition G.1 page 197)} \\
 &= x + (-x) && \text{by left hypothesis} \\
 &= \emptyset && \text{by } \textit{additive inverse property} \text{ (Definition G.1 page 197)}
 \end{aligned}$$

⇒  
**Definition G.3.** <sup>10</sup> Let  $\Omega \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$  be a linear space with vectors  $x, y \in X$ . Let  $-y$  be the additive inverse of  $y$  such that  $y + (-y) = \emptyset$ .

**D E F** The **difference** of  $x$  and  $y$  is  $x + (-y)$  and is denoted  $x - y$ .

**Theorem G.2** (Additive inverse properties). <sup>11</sup> Let  $\Omega \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$  be a linear space,  $\emptyset$  the ADDITIVE IDENTITY ELEMENT with respect to  $+$ , and  $-x$  the ADDITIVE INVERSE (Definition F.1 page 195) of  $x$  with respect to  $+$ .

T H M	1. $x + y = \emptyset \implies x = -y \quad \forall x, y \in X \quad (\text{additive inverse is UNIQUE})$ 2. $(-\alpha)x = -(\alpha x) = \alpha(-x) \quad \forall x \in X, \alpha \in \mathbb{F}$ 3. $\alpha(x - y) = \alpha x - \alpha y \quad \forall x, y \in X, \alpha \in \mathbb{F} \quad (\text{DISTRIBUTIVE})$ 4. $(\alpha - \beta)x = \alpha x - \beta x \quad \forall x \in X, \alpha, \beta \in \mathbb{F} \quad (\text{DISTRIBUTIVE})$
-------	---

<sup>10</sup> Berberian (1961) page 7 (Definition 1)

<sup>11</sup> 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  $x + y = 0 \implies x = -y$ :

$$\begin{aligned} x &= x - 0 \\ &= x - (x + y) && \text{by left hypothesis} \\ &= (x - x) - y \\ &= 0 - y \\ &= -y \end{aligned}$$

2. Proof that  $(-\alpha)x = -(\alpha x)$ :

$$\begin{aligned} 0 &= 0x && \text{by Theorem G.1 page 199} \\ &= (\alpha - \alpha)x && \text{by field property of } \mathbb{F} \\ &= [\alpha + (-\alpha)]x && \text{by field property of } \mathbb{F} \\ &= \alpha x + (-\alpha)x && \text{by Definition G.1 page 197} \\ \implies -(\alpha x) &= (-\alpha)x && \text{by item (1) page 201} \end{aligned}$$

3. Proof that  $\alpha(-x) = -(\alpha x)$ :

$$\begin{aligned} 0 &= \alpha 0 && \text{by Theorem G.1 page 199} \\ &= \alpha[x + (-x)] && \text{by definition of additive identity element } -x \\ &= \alpha x + \alpha(-x) && \text{by Definition G.1 page 197} \\ &= \alpha x + \alpha(-x) \\ \implies -(\alpha x) &= \alpha(-x) && \text{by item (1) page 201} \end{aligned}$$

4. Proof that  $\alpha(x - y) = \alpha x - \alpha y$ :

$$\begin{aligned} \alpha(x - y) &= \alpha[x + (-y)] && \text{by Definition G.3 page 200} \\ &= \alpha x + \alpha(-y) && \text{by Definition G.1 page 197} \\ &= \alpha x + (-\alpha y) && \text{by item (3) page 201} \\ &= \alpha x - \alpha y && \text{by Definition G.3 page 200} \end{aligned}$$

5. Proof that  $(\alpha - \beta)x = \alpha x - \beta x$ :

$$\begin{aligned} (\alpha - \beta)x &= [\alpha + (-\beta)]x && \text{by field properties of } \mathbb{F} \\ &= \alpha x + (-\beta)x && \text{by Definition G.1} \\ &= \alpha x + [-(\beta x)] && \text{by item (2) page 201} \\ &= \alpha x - (\beta x) && \text{by Definition G.3 page 200} \end{aligned}$$

⇒

**Theorem G.3.** <sup>12</sup> Let  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$  be a linear space,  $0$  the additive identity element with respect to  $+$ , and  $-x$  additive inverse of  $x$  with respect to  $+$ .

- |                      |  |
|----------------------|--|
| <b>T<br/>H<br/>M</b> | 1. $\alpha x = \alpha y$ and $\alpha \neq 0 \implies x = y \quad \forall x, y \in X$<br>2. $\alpha x = \beta x$ and $x \neq 0 \implies \alpha = \beta \quad \forall x, y \in X, \alpha, \beta \in \mathbb{F}$<br>3. $z + x = z + y \implies x = y \quad \forall x, y, z \in X$ |
|----------------------|--|

<sup>12</sup> Michel and Herget (1993) page 77, Peano (1888a) (Chapter IX), Peano (1888b) pages 119–120, Banach (1922) pages 134–135

PROOF:

1. Proof that  $\alpha\mathbf{x} = \alpha\mathbf{y}$  and  $\alpha \neq 0 \implies \mathbf{x} = \mathbf{y}$ :

$$\begin{aligned} 0 &= \frac{1}{\alpha}(0) && \text{by left hypothesis } (\alpha \neq 0) \\ &= \frac{1}{\alpha}(\alpha\mathbf{x} - \alpha\mathbf{y}) && \text{by left hypothesis } (\alpha\mathbf{x} = \alpha\mathbf{y}) \\ &= \frac{1}{\alpha}\alpha(\mathbf{x} - \mathbf{y}) && \text{by Definition G.1 page 197} \\ &= \mathbf{x} - \mathbf{y} \end{aligned}$$

2. Proof that  $\alpha\mathbf{x} = \beta\mathbf{x}$  and  $\mathbf{x} \neq 0 \implies \alpha = \beta$ :

$$\begin{aligned} 0 &= \alpha\mathbf{x} + (-\alpha\mathbf{x}) && \text{by definition of additive inverse} \\ &= \beta\mathbf{x} + (-\alpha\mathbf{x}) && \text{by left hypothesis} \\ &= \beta\mathbf{x} + (-\alpha)\mathbf{x} && \text{by Theorem G.2 page 200} \\ &= [\beta + (-\alpha)]\mathbf{x} && \text{by Definition G.1 page 197} \\ \implies \beta - \alpha &= 0 && \text{by Theorem G.1 page 199} \\ \implies \alpha &= \beta && \text{by field properties of } \mathbb{F} \end{aligned}$$

3. Proof that  $\mathbf{z} + \mathbf{x} = \mathbf{z} + \mathbf{y} \implies \mathbf{x} = \mathbf{y}$ :

$$\begin{aligned} 0 &= (\mathbf{z} + \mathbf{x}) - (\mathbf{z} + \mathbf{y}) && \text{by Definition G.1 property 1} \\ &= (\mathbf{x} + \mathbf{z}) - (\mathbf{z} + \mathbf{y}) && \text{by Definition G.1 property 3} \\ &= (\mathbf{x} + \mathbf{z}) + [(-1)\mathbf{z} + (-1)\mathbf{y}] && \text{by previous result 2.} \\ &= (\mathbf{x} + \mathbf{z}) + (-\mathbf{z} - \mathbf{y}) \\ &= \mathbf{x} + (\mathbf{z} - \mathbf{z}) - \mathbf{y} \\ &= \mathbf{x} - \mathbf{y} \end{aligned}$$

⇒

## G.2 Order on Linear Spaces

**Definition G.4.** <sup>13</sup> Let  $\Omega \triangleq (X, +, \cdot, (\mathbb{R}, \dot{+}, \dot{\times}))$  be a real linear space.

The pair  $(\Omega, \leq)$  is an ordered linear space if

- |              |  |     |
|--------------|--|-----|
| <b>D E F</b> | 1. $\mathbf{x} \leq \mathbf{y} \implies \mathbf{x} + \mathbf{z} \leq \mathbf{y} + \mathbf{z} \quad \forall \mathbf{z} \in X$ | and |
|              | 2. $\mathbf{x} \leq \mathbf{y} \implies \alpha\mathbf{x} \leq \alpha\mathbf{y} \quad \forall \alpha \in \mathbb{F}$          |     |

A vector  $\mathbf{x}$  is positive if  $0 \leq \mathbf{x}$ .

The positive cone  $X^+$  of  $(X, \leq)$  is the set  $X^+ \triangleq \{\mathbf{x} \in X \mid 0 \leq \mathbf{x}\}$ .

**Definition G.5.** <sup>14</sup> Let  $(X, \leq)$  be an ordered linear space.

**D E F** The tuple  $L \triangleq (X, \vee, \wedge; \leq)$  is a Riesz space if  $L$  is a lattice.

A RIESZ SPACE is also called a vector lattice.

**Theorem G.4.** <sup>15</sup> Let  $(X, \vee, \wedge; \leq)$  be a Riesz space (Definition G.5 page 202).

<b>T H M</b>	$\mathbf{x} \vee \mathbf{y} = -[(-\mathbf{x}) \wedge (-\mathbf{y})]$ $\mathbf{x} + (\mathbf{y} \vee \mathbf{z}) = (\mathbf{x} + \mathbf{y}) \vee (\mathbf{x} + \mathbf{z})$ $\alpha(\mathbf{x} \vee \mathbf{y}) = (\alpha\mathbf{x}) \vee (\alpha\mathbf{y})$ $\mathbf{x} + \mathbf{y} = (\mathbf{x} \wedge \mathbf{y}) + (\mathbf{x} \vee \mathbf{y})$	$\mathbf{x} \wedge \mathbf{y} = -[(-\mathbf{x}) \vee (-\mathbf{y})]$ $\mathbf{x} + (\mathbf{y} \wedge \mathbf{z}) = (\mathbf{x} + \mathbf{y}) \wedge (\mathbf{x} + \mathbf{z})$ $\alpha(\mathbf{x} \wedge \mathbf{y}) = (\alpha\mathbf{x}) \wedge (\alpha\mathbf{y})$ $\forall \mathbf{x}, \mathbf{y} \in X$ $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in X$ $\forall \mathbf{x}, \mathbf{y} \in X, \alpha \geq 0$ $\forall \mathbf{x}, \mathbf{y} \in X, \alpha \in \mathbb{F}$
--------------	--	---

<sup>13</sup> Aliprantis and Burkinshaw (2006) pages 1–2

<sup>14</sup> Aliprantis and Burkinshaw (2006) page 2

<sup>15</sup> Aliprantis and Burkinshaw (2006) page 3 (Theorem 1.2)



PROOF:

1. Proof that  $x \vee y = -[(-x) \wedge (-y)]$ :

$(-x) \wedge (-y) \leq -x$	$(-x) \wedge (-y) \leq -y$
$x \leq -[(-x) \wedge (-y)]$	$y \leq -[(-x) \wedge (-y)]$
$x \vee y \leq -[(-x) \wedge (-y)]$	
$x \leq x \vee y$	$y \leq x \vee y$
$-(x \vee y) \leq -x$	$-(x \vee y) \leq -y$
$-(x \vee y) \leq (-x) \wedge (-y)$	
$-[(-x) \wedge (-y)] \leq x \vee y$	

2. Proof that  $x \wedge y = -[(-x) \vee (-y)]$ :

$x \vee y = -[(-x) \wedge (-y)]$	by item (1)
$(-x) \vee (-y) = -[(-(-x)) \wedge (-(-y))]$	replace $x$ with $-x$ and $y$ with $y$
$(-x) \vee (-y) = -[x \wedge y]$	$-(-x) = x$
$-[x \wedge y] = (-x) \vee (-y)$	by symmetry of $=$ relation
$x \wedge y = -[(-x) \vee (-y)]$	multiply both sides by $-1$

3. Proof that  $x + (y \vee z) = (x + y) \vee (x + z)$ :

$x + y \leq x + (y \vee z)$	$x + z \leq x + (y \vee z)$
$(x + y) \vee (x + z) \leq x + (y \vee z)$	
$y = -x + (x + y)$	$z = -x + (x + z)$
$\leq -x + [(x + y) \vee (x + z)]$	$\leq -x + [(x + y) \vee (x + z)]$
$y \vee z \leq -x + [(x + y) \vee (x + z)]$	
$x + (y \vee z) \leq (x + y) \vee (x + z)$	

4. Proof that  $x + (y \wedge z) = (x + y) \wedge (x + z)$ :

$x + y \geq x + (y \wedge z)$	$x + z \geq x + (y \wedge z)$
$(x + y) \wedge (x + z) \geq x + (y \wedge z)$	
$y = -x + (x + y)$	$z = -x + (x + z)$
$\geq -x + [(x + y) \wedge (x + z)]$	$\geq -x + [(x + y) \wedge (x + z)]$
$y \wedge z \geq -x + [(x + y) \wedge (x + z)]$	
$x + (y \wedge z) \geq (x + y) \wedge (x + z)$	

5. Proof that  $\alpha(x \vee y) = (\alpha x) \vee (\alpha y)$  for  $\alpha \geq 0$ :

$x \leq x \vee y$	$y \leq x \vee y$	
$\alpha x \leq \alpha(x \vee y)$	$\alpha y \leq \alpha(x \vee y)$	by Definition G.4 page 202
$(\alpha x) \vee (\alpha y) \leq \alpha(x \vee y)$		
$\alpha x \leq (\alpha x) \vee (\alpha y)$	$\alpha y \leq (\alpha x) \vee (\alpha y)$	
$x \leq \alpha^{-1}(\alpha x) \vee (\alpha y)$	$y \leq \alpha^{-1}(\alpha x) \vee (\alpha y)$	
$x \vee y \leq \alpha^{-1}(\alpha x) \vee (\alpha y)$		
$\alpha(x \vee y) \leq (\alpha x) \vee (\alpha y)$		

6. Proof that  $\alpha(x \wedge y) = (\alpha x) \wedge (\alpha y)$  for  $\alpha \geq 0$ :

$x \geq x \wedge y$	$y \geq x \wedge y$	
$\alpha x \geq \alpha(x \wedge y)$	$\alpha y \geq \alpha(x \wedge y)$	
$(\alpha x) \wedge (\alpha y) \geq \alpha(x \wedge y)$		by Definition G.4 page 202

$\alpha x \geq (\alpha x) \wedge (\alpha y)$	$\alpha y \geq (\alpha x) \wedge (\alpha y)$	
$x \geq \alpha^{-1}(\alpha x) \wedge (\alpha y)$	$y \geq \alpha^{-1}(\alpha x) \wedge (\alpha y)$	
$x \wedge y \geq \alpha^{-1}(\alpha x) \wedge (\alpha y)$		
$\alpha(x \wedge y) \geq (\alpha x) \wedge (\alpha y)$		

7. Proof that  $x + y = (x \wedge y) + (x \vee y)$ :

$x \leq x \vee y$	$y \leq x \vee y$
$x + y \leq (x \vee y) + y$	$x + vy \leq x + (x \vee y)$
$x + y - (x \vee y) \leq y$	$x + vy - (x \vee y) \leq x$
$x + y - (x \vee y) \leq x \wedge y$	
$x + y \leq (x \vee y) + (x \wedge y)$	
$x \wedge y \leq x$	$x \wedge y \leq y$
$0 \leq x - (x \wedge y)$	$0 \leq y - (x \wedge y)$
$y \leq y + x - (x \wedge y)$	$x \leq x + y - (x \wedge y)$
$y \leq x + y - (x \wedge y)$	$x \leq x + y - (x \wedge y)$
$x \vee y \leq x + y - (x \wedge y)$	
$(x \wedge y) + (x \vee y) \leq x + y$	

⇒

**Definition G.6.** <sup>16</sup> Let  $(X, \vee, \wedge; \leq)$  be a RIESZ SPACE (Definition G.5 page 202).

**D E F**  $x^+$  is defined as  $x^+ \triangleq x \vee \emptyset$  and is called the **positive part** of  $x$ .  
 $x^-$  is defined as  $x^- \triangleq (-x) \vee \emptyset$  and is called the **negative part** of  $x$ .  
 $|x|$  is defined as  $|x| \triangleq x \vee (-x)$  and is called the **absolute value** of  $x$ .

**Theorem G.5.** <sup>17</sup> Let  $(X, \vee, \wedge; \leq)$  be a RIESZ SPACE (Definition G.5 page 202).

T H M	$y - z = x$ and $y \wedge z = \emptyset$	$\Leftrightarrow$	$\left\{ \begin{array}{l} y = x^+ \text{ and} \\ z = x^- \end{array} \right.$
-------	---	-------------------	---

PROOF:

1. Proof that left hypothesis  $\Rightarrow$  right hypothesis:

$$\begin{aligned}
 x^+ &= x \vee \emptyset && \text{by definition of } x^+ \text{ Definition G.6 page 204} \\
 &= (y - z) \vee \emptyset && \text{by left hypothesis} \\
 &= (y - z) \vee (z - z) \\
 &= (y \vee z) - z && \text{by Theorem G.4 page 202} \\
 &= [y + z - (y \wedge z)] - z && \text{by Theorem G.4 page 202} \\
 &= y - (y \wedge z) \\
 &= y - \emptyset && \text{by left hypothesis} \\
 &= y \\
 x^- &= (-x) \vee \emptyset && \text{by definition of } x^- \text{ Definition G.6 page 204} \\
 &= (z - y) \vee \emptyset && \text{by left hypothesis} \\
 &= (z - y) \vee (y - y) \\
 &= (z \vee y) - y && \text{by Theorem G.4 page 202}
 \end{aligned}$$

<sup>16</sup> Aliprantis and Burkinshaw (2006) page 4, Istrătescu (1987) page 129

<sup>17</sup> Aliprantis and Burkinshaw (2006) page 4 (Theorem 1.3)

$$\begin{aligned}
 &= [z + y - (z \wedge y)] - z && \text{by Theorem G.4 page 202} \\
 &= z - (z \wedge y) \\
 &= z - \emptyset && \text{by left hypothesis} \\
 &= z
 \end{aligned}$$

2. Proof that left hypothesis  $\iff$  right hypothesis:

$$\begin{aligned}
 y - z &= x^+ - x^- && \text{by right hypothesis} \\
 &= [x \vee \emptyset] - [(-x) \vee \emptyset] && \text{by Definition G.6 page 204} \\
 &= (x \vee \emptyset) + (x \wedge \emptyset) && \text{by Theorem G.4 page 202} \\
 &= x && \text{by Theorem G.4 page 202} \\
 y \wedge z &= x^+ \wedge x^- && \text{by right hypothesis} \\
 &= [x^- + (x^+ - x^-)] \wedge [x^- + \emptyset] && \\
 &= x^- + [(x^+ - x^-) \wedge \emptyset] && \text{by Theorem G.4 page 202} \\
 &= x^- + [(y - z) \wedge \emptyset] && \text{by right hypothesis} \\
 &= x^- + (x \wedge \emptyset) && \text{by previous result} \\
 &= x^- - [-x \vee \emptyset] && \text{by Theorem G.4 page 202} \\
 &= x^- - x && \text{by definition of } x^- \text{ (Definition G.6 page 204)} \\
 &= \emptyset
 \end{aligned}$$



**Theorem G.6.**<sup>18</sup> Let  $(X, \vee, \wedge; \leq)$  be a RIESZ SPACE (Definition G.5 page 202). Let  $x^+$  the POSITIVE PART of  $x \in X$ ,  $x^-$  the NEGATIVE PART of  $x \in X$ , and  $|x|$  the ABSOLUTE VALUE (Definition G.6 page 204) of  $x \in X$ .

T H M	$ x  = x^+ + x^-$ $x = (x - y)^+ + (x \wedge y)$ $\forall x \in X$
-------------	---

PROOF:

$$\begin{aligned}
 |x| &= x \vee (-x) && \text{by definition of } |x| \text{ (Definition G.6 page 204)} \\
 &= (2x - x) \vee (\emptyset - x) \\
 &= (-x + 2x) \vee (-x + \emptyset) && \text{by commutative property (Definition G.1 page 197)} \\
 &= (-x) + (2x \vee \emptyset) && \text{by Theorem G.4 page 202} \\
 &= (2x \vee \emptyset) - x && \text{by the commutative property (Definition G.1 page 197)} \\
 &= 2(x \vee \emptyset) - x && \text{by Theorem G.4 page 202} \\
 &= 2x^+ - x && \text{by definition of } x^+ \text{ (Definition G.6 page 204)} \\
 &= 2x^+ - (x^+ - x^-) && \text{by 1} \\
 &= x^+ + x^- \\
 x &= x + \emptyset \\
 &= (x \vee y) + (x \wedge y) - y && \text{by Theorem G.4 page 202} \\
 &= [(x - y) \vee (y - y)] + (x \wedge y) && \text{by Theorem G.4 page 202} \\
 &= [(x - y) \vee \emptyset] + (x \wedge y) \\
 &= (x - y)^+ + (x \wedge y) && \text{by definition of } x^+ \text{ (Definition G.6 page 204)}
 \end{aligned}$$



<sup>18</sup> Aliprantis and Burkinshaw (2006) page 4

**Theorem G.7.** <sup>19</sup> Let  $(X, \vee, \wedge; \leq)$  be a RIESZ SPACE (Definition G.5 page 202). Let  $x^+$  the POSITIVE PART of  $x \in X$ ,  $x^-$  the NEGATIVE PART of  $x \in X$ , and  $|x|$  the ABSOLUTE VALUE (Definition G.6 page 204) of  $x \in X$ .

T  
H  
M

1.  $x \vee y = \frac{1}{2}(x + y + |x - y|) \quad \forall x, y \in X$
2.  $x \wedge y = \frac{1}{2}(x + y - |x - y|) \quad \forall x, y \in X$
3.  $|x - y| = (x \vee y) - (x \wedge y) \quad \forall x, y \in X$
4.  $|x| \vee |y| = \frac{1}{2}(|x + y| + |x - y|) \quad \forall x, y \in X$
5.  $|x| \wedge |y| = \frac{1}{2}(|x + y| - |x - y|) \quad \forall x, y \in X$

PROOF:

$$\begin{aligned}
 (x + y + |x - y|) &= (x + y) + [(x - y) \vee (y - x)] && \text{by Definition G.6 page 204} \\
 &= [(x + y) + (x - y)] \vee [(x + y) + (y - x)] && \text{by Theorem G.4 page 202} \\
 &= (2x) \vee (2y) \\
 &= 2(x \vee y) && \text{by Theorem G.4 page 202} \\
 (x + y - |x - y|) &= (x + y) - [(x - y) \vee (y - x)] && \text{by Definition G.6 page 204} \\
 &= (x + y) - [(-(y - x)) \vee (-(x - y))] \\
 &= (x + y) + [(y - x) \wedge (x - y)] \\
 &= [(x + y) + (y - x)] \wedge [(x + y) + (x - y)] \\
 &= (2y) \wedge (2x) \\
 &= 2(y \wedge x) \\
 &= 2(x \wedge y) && \text{by Theorem G.4 page 202} \\
 |x - y| &= \frac{1}{2}(x + y + |x - y|) - \frac{1}{2}(x + y - |x - y|) \\
 &= (x \vee y) - (x \wedge y) && \text{by 1 and 2} \\
 |x + y| + |x - y| &= \frac{1}{2}(\emptyset + |2x + 2y|) + |x - y| \\
 &= \frac{1}{2}[(x + y) + (-x - y) + |(x + y) - (-x - y)|] + |x - y| \\
 &= [(x + y) \vee (-x - y)] + |x - y| && \text{by 1} \\
 &= [(x + y) + |x - y|] \vee [(-x - y) + |x - y|] \\
 &= 2(x \vee y) \vee 2[(-y) + (-x) + |(-y) - (-x)|] \\
 &= 2(x \vee y) \vee 2[(-y) \vee (-x)] \\
 &= 2([x \vee (-x)] \vee (y \vee (-y))) \\
 &= 2(|x| \vee |y|) && \text{by Definition G.6 page 204} \\
 ||x + y| - |x - y|| &= 2(|x + y| \vee |x - y|) - (|x + y| + |x - y|) \\
 &= (|x + y + x - y| + |x + y - x + y|) - 2(|x| \vee |y|) && \text{by 1} \\
 &= 2(|x| + |y|) - 2(|x| \vee |y|) \\
 &= 2(|x| \vee |y|) && \text{by Theorem G.4 page 202}
 \end{aligned}$$

⇒

**Definition G.7.** <sup>20</sup> Let  $(X, \vee, \wedge; \leq)$  be a RIESZ SPACE (Definition G.5 page 202). Let  $x^+$  the POSITIVE PART of  $x \in X$ ,  $x^-$  the NEGATIVE PART of  $x \in X$ , and  $|x|$  the ABSOLUTE VALUE (Definition G.6 page 204) of  $x \in X$ .

D  
E  
F

**x and y are disjoint**, denoted by  $x \perp y$ , if

$$|x| \wedge |y| = \emptyset.$$

**Two subsets U and V of X are disjoint**, denoted by  $U \perp V$  if

$$x \perp y \quad \forall x \in U \text{ and } y \in V$$

<sup>19</sup> Aliprantis and Burkinshaw (2006) page 5 (Theorem 1.4)

<sup>20</sup> Aliprantis and Burkinshaw (2006) page 5



**Definition G.8.** <sup>21</sup> Let  $(X, \vee, \wedge; \leq)$  be a RIESZ SPACE (Definition G.5 page 202). Let  $x^+$  the POSITIVE PART of  $x \in X$ ,  $x^-$  the NEGATIVE PART of  $x \in X$ , and  $|x|$  the ABSOLUTE VALUE (Definition G.6 page 204) of  $x \in X$ . Let  $Y$  be a subset of  $X$ .

**DEF**  $Y^d$  is the **disjoint complement** of  $Y$  if  $Y^d \triangleq \{x \in X | x \perp y \quad \forall y \in Y\}$ .  
The quantity  $Y^{dd}$  is defined as  $(Y^d)^d$ .

**Definition G.9.** <sup>22</sup> Let  $(X, \vee, \wedge; \leq)$  be a RIESZ SPACE (Definition G.5 page 202). Let  $x^+$  the POSITIVE PART of  $x \in X$ ,  $x^-$  the NEGATIVE PART of  $x \in X$ , and  $|x|$  the ABSOLUTE VALUE (Definition G.6 page 204) of  $x \in X$ .

D <small>E</small> F	$ A  \triangleq \{ a    a \in A\}$ $A^+ \triangleq \{a^+   a \in A\}$ $A^- \triangleq \{a^-   a \in A\}$ $A \vee B \triangleq \{a \vee b   a \in A \text{ and } b \in B\}$ $A \wedge B \triangleq \{a \wedge b   a \in A \text{ and } b \in B\}$ $x \vee A \triangleq \{x \vee a   a \in A\}$ $x \wedge A \triangleq \{x \wedge a   a \in A\}$
----------------------	--

<sup>21</sup>  Aliprantis and Burkinshaw (2006) page 5

<sup>22</sup>  Aliprantis and Burkinshaw (2006) page 7



## APPENDIX H

### LINEAR COMBINATIONS

## H.1 Linear combinations in linear spaces

A *linear space* (Definition G.1 page 197) 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 N.1 page 293) only (finite “ $N$ ”).

**Definition H.1.** <sup>1</sup> Let  $\{x_n \in X \mid n=1,2,\dots,N\}$  be a set of vectors in a LINEAR SPACE  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$ .

**D E F** A vector  $x \in X$  is a **linear combination** of the vectors in  $\{x_n\}$  if

there exists  $\{\alpha_n \in \mathbb{F} \mid n=1,2,\dots,N\}$  such that 
$$x = \sum_{n=1}^N \alpha_n x_n.$$

**Definition H.2.** <sup>2</sup> Let  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$  be a linear space and  $Y$  be a subset of  $X$ .

**D E F** The **linear span** of  $Y$  is defined as  $\text{span}Y \triangleq \left\{ \sum_{\gamma \in \Gamma} \alpha_\gamma y_\gamma \mid \alpha_\gamma \in \mathbb{F}, y_\gamma \in Y \right\}.$

The set  $Y$  **spans** a set  $A$  if  $A \subseteq \text{span}Y.$

**Proposition H.1.** <sup>3</sup> Let  $\{x_n \in X \mid n=1,2,\dots,N\}$  be a set of vectors in a LINEAR SPACE  $L \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times})).$

- P R P**
1.  $\text{span}\{x_n\}$  is a LINEAR SPACE (Definition G.1 page 197) and
  2.  $\text{span}\{x_n\}$  is a LINEAR SUBSPACE of  $L$  (Definition G.2 page 198).

**Definition H.3.** <sup>4</sup> Let  $L \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$  be a LINEAR SPACE.

**D E F** The set  $Y \triangleq \{x_n \in X \mid n=1,2,\dots,N\}$  is **linearly independent** in  $L$  if 
$$\left\{ \sum_{n=1}^N \alpha_n x_n = 0 \right\} \implies \{\alpha_1 = \alpha_2 = \dots = \alpha_N = 0\}.$$

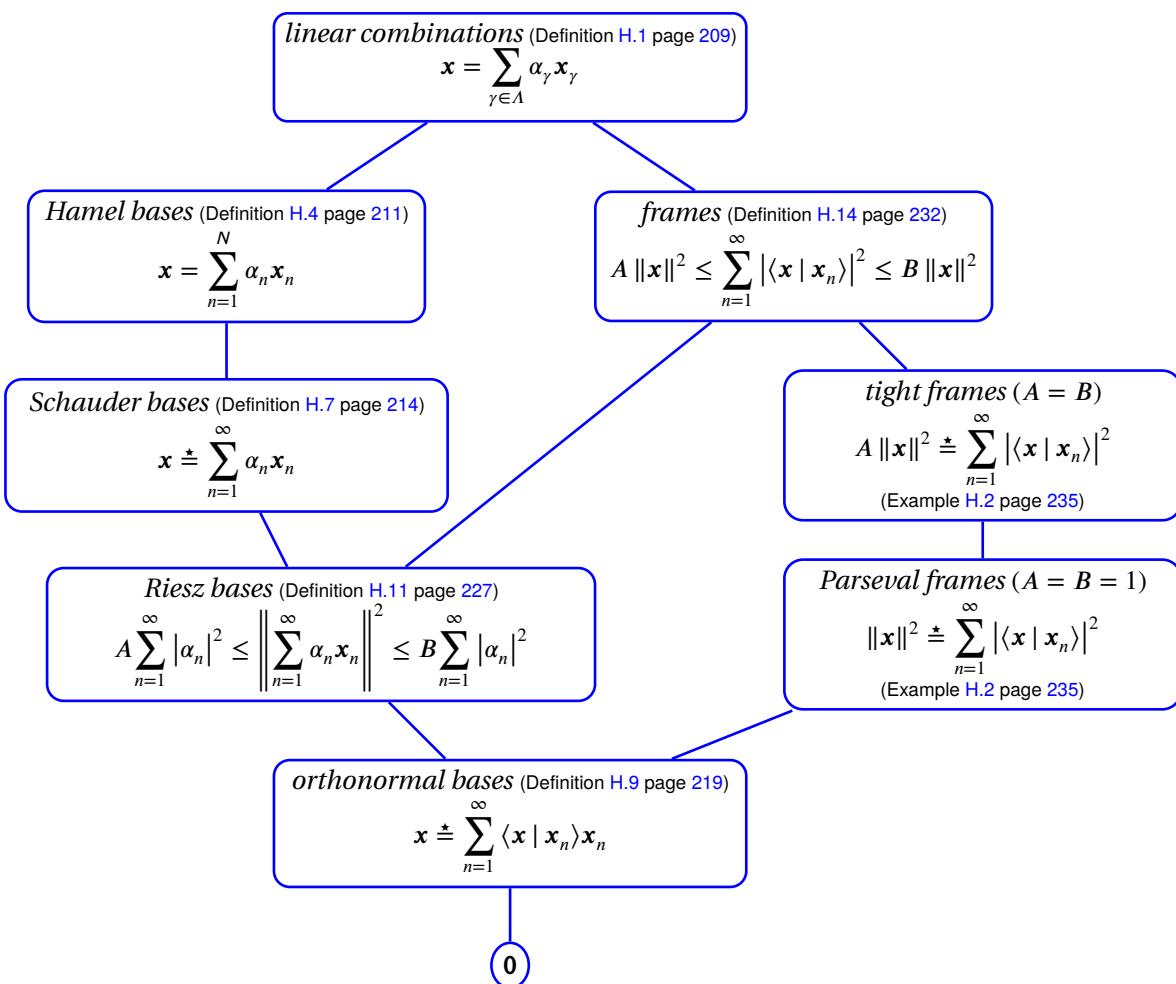
The set  $Y$  is **linearly dependent** in  $L$  if  $Y$  is not linearly independent in  $L$ .

<sup>1</sup> Berberian (1961) page 11 (Definition I.4.1), Kubrusly (2001) page 46

<sup>2</sup> Michel and Herget (1993) page 86 (3.3.7 Definition), Kurdila and Zabarankin (2005) page 44, Searcoid (2002) page 71 (Definition 3.2.5—more general definition)

<sup>3</sup> Kubrusly (2001) page 46

<sup>4</sup> Bachman and Narici (1966) pages 3–4, Christensen (2003) page 2, Heil (2011) page 156 (Definition 5.7)

Figure H.1: Lattice of *linear combinations*

**Definition H.4.** <sup>5</sup> Let  $\{x_n \in X | n=1,2,\dots,N\}$  be a set of vectors in a LINEAR SPACE  $L \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$ .

**D E F** The set  $\{x_n\}$  is a **Hamel basis** for  $L$  if

1.  $\{x_n\}$  SPANS  $L$  (Definition H.2 page 209) and
2.  $\{x_n\}$  is LINEARLY INDEPENDENT in  $L$  (Definition H.1 page 209) .

A HAMEL BASIS is also called a **linear basis**.

**Definition H.5.** <sup>6</sup> Let  $L \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$  be a LINEAR SPACE. Let  $x$  be a VECTOR in  $L$  and  $Y \triangleq \{x_n \in X | n=1,2,\dots,N\}$  be a set of vectors in  $L$ .

**D E F** The expression  $\sum_{n=1}^N \alpha_n x_n$  is the **expansion** of  $x$  on  $Y$  in  $L$  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  $L$ . If  $\alpha_N \neq 0$ , then  $N$  is the **dimension**  $\dim L$  of  $L$ .

**Theorem H.1.** <sup>7</sup> Let  $\{x_n | n=1,2,\dots,N\}$  be a HAMEL BASIS (Definition H.4 page 211) for a LINEAR SPACE  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$ .

**T H M** 
$$\left\{ x = \sum_{n=1}^N \alpha_n x_n = \sum_{n=1}^N \beta_n x_n \right\} \implies \underbrace{\alpha_n = \beta_n}_{\text{coordinates of } x \text{ are UNIQUE}} \quad \forall x \in X$$

PROOF:

$$0 = x - x$$

$$= \sum_{n=1}^N \alpha_n x_n - \sum_{n=1}^N \beta_n x_n$$

$$= \sum_{n=1}^N (\alpha_n - \beta_n) x_n$$

$\implies \{x_n\}$  is linearly dependent if  $(\alpha_n - \beta_n) \neq 0 \quad \forall n = 1, 2, \dots, N$

$\implies (\alpha_n - \beta_n) = 0 \quad \forall n = 1, 2, \dots, N$  (because  $\{x_n\}$  is a basis and therefore must be linearly independent)

$\implies \alpha_n = \beta_n$  for  $n = 1, 2, \dots, N$

**Theorem H.2.** <sup>8</sup> Let  $L \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$  be a LINEAR SPACE.

**T H M** 
$$\begin{aligned} & \left\{ \begin{array}{l} 1. \quad \{x_n \in X | n=1,2,\dots,N\} \text{ is a HAMEL BASIS for } L \\ 2. \quad \{y_n \in X | n=1,2,\dots,M\} \text{ is a set of LINEARLY INDEPENDENT vectors in } L \end{array} \right. \quad \text{and} \\ & \implies \left\{ \begin{array}{l} 1. \quad M \leq N \\ 2. \quad M = N \implies \{y_n | n=1,2,\dots,M\} \text{ is a BASIS for } L \\ 3. \quad M \neq N \implies \{y_n | n=1,2,\dots,M\} \text{ is NOT a basis for } L \end{array} \right. \quad \text{and} \end{aligned}$$

PROOF:

<sup>5</sup> 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)

<sup>6</sup> 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)

<sup>7</sup> Michel and Herget (1993) pages 89–90 (Theorem 3.3.25)

<sup>8</sup> Michel and Herget (1993) pages 90–91 (Theorem 3.3.26)

1. Proof that  $\{y_1, x_1, \dots, x_{N-1}\}$  is a *basis* for  $L$ :

(a) Proof that  $\{y_1, x_1, \dots, x_{N-1}\}$  spans  $L$ :

i. Because  $\{x_n|_{n=1,2,\dots,N}\}$  is a *basis* for  $L$ , there exists  $\beta \in \mathbb{F}$  and  $\{\alpha_n \in \mathbb{F}|_{n=1,2,\dots,N}\}$  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  $\{y_1, x_1, \dots, x_{N-1}\}$  spans  $L$ :

(b) Proof that  $\{y_1, x_1, \dots, x_{N-1}\}$  is *linearly independent*:

i. If  $\{y_1, x_1, \dots, x_{N-1}\}$  is *linearly dependent*, then there exists  $\{\epsilon, \epsilon_1, \dots, \epsilon_{N-1}\}$  such that

$$\epsilon y_1 + \left( \sum_{n=1}^{N-1} \epsilon_{n \in \mathbb{Z}} x_n \right) + 0 x_n = 0.$$

ii. item (1(b)i) implies that the coordinate of  $y_1$  associated with  $x_n$  is 0.

$$y_1 = -\left( \sum_{n=1}^{N-1} \frac{\epsilon_n}{\epsilon} x_n \right) + 0 x_n = 0.$$

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  $\{y_1, x_1, \dots, x_{N-1}\}$  is *linearly independent*.

2. Proof that  $\{y_1, y_2, x_1, \dots, x_{N-2}\}$  is a *basis*: Repeat item (1).

3. Suppose  $m = n$ . Proof that  $\{y_1, y_2, \dots, y_M\}$  is a *basis*: Repeat item (1)  $M - 1$  times.

4. Proof that  $M \not> N$ :

(a) Suppose that  $M = N + 1$ .

(b) Then because  $\{y_n|_{n=1,2,\dots,N}\}$  is a *basis*, there exists  $\{\zeta_n|_{n=1,2,\dots,N+1}\}$  such that

$$\sum_{n=1}^{N+1} \zeta_{n \in \mathbb{Z}} y_{n \in \mathbb{Z}} = 0.$$

(c) This implies that  $\{y_n|_{n=1,2,\dots,N+1}\}$  is *linearly dependent*.



(d) This implies that  $\{y_n|_{n=1,2,\dots,N+1}\}$  is *not* a basis.

(e) This implies that  $M > N$ .

5. Proof that  $M \neq N \implies \{y_n|_{n=1,2,\dots,M}\}$  is *not* a basis for  $L$ :

(a) Proof that  $M > N \implies \{y_n|_{n=1,2,\dots,M}\}$  is *not* a basis for  $L$ : same as in item (4).

(b) Proof that  $M < N \implies \{y_n|_{n=1,2,\dots,M}\}$  is *not* a basis for  $L$ :

i. Suppose  $M = N - 1$ .

ii. Then  $\{y_n|_{n=1,2,\dots,N-1}\}$  is a *basis* and there exists  $\lambda$  such that

$$\sum_{n=1}^N \lambda_{n \in \mathbb{Z}} y_{n \in \mathbb{Z}} = 0.$$

iii. This implies that  $\{y_n|_{n=1,2,\dots,N}\}$  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 H.1.** <sup>9</sup> Let  $L \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$  be a linear space.

**COR**  $\left\{ \begin{array}{l} 1. \quad \{x_n \in X | n=1,2,\dots,N\} \text{ is a HAMEL BASIS for } L \text{ and} \\ 2. \quad \{y_n \in X | n=1,2,\dots,M\} \text{ is a HAMEL BASIS for } L \end{array} \right\} \implies \{N = M\}$

(all Hamel bases for  $L$  have the same number of vectors)



PROOF: This follows from Theorem H.2 (page 211).

## H.2 Bases in topological linear spaces

A linear space supports the concept of the *span* of a set of vectors (Definition H.2 page 209). In a topological linear space  $\Omega \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), T)$ , a set  $A$  is said to be *total* in  $\Omega$  if the span of  $A$  is *dense* in  $\Omega$ . 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”. <sup>10</sup> In this text, except for these comments and Definition H.6, “complete” refers to the metric space definition only.

If a set is both *total* and *linearly independent* (Definition H.3 page 209) in  $\Omega$ , then that set is a *Hamel basis* (Definition H.4 page 211) for  $\Omega$ .

**Definition H.6.** <sup>11</sup> Let  $A^-$  be the CLOSURE of  $A$  in a TOPOLOGICAL LINEAR SPACE  $\Omega \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), T)$ . Let  $\text{span } A$  be the SPAN (Definition H.2 page 209) of a set  $A$ .

**DEF** A set of vectors  $A$  is **total** (or **complete** or **fundamental**) in  $\Omega$  if  
 $(\text{span } A)^- = \Omega$       (SPAN of  $A$  is DENSE in  $\Omega$ ).

<sup>9</sup> Kubrusly (2001) page 52 (Theorem 2.7), Michel and Herget (1993) page 91 (Theorem 3.3.31)

<sup>10</sup> 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)

<sup>11</sup> 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)

## H.3 Schauder bases in Banach spaces

**Definition H.7.** <sup>12</sup> Let  $\mathcal{B} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$  be a BANACH SPACE. Let  $\doteq$  represent STRONG CONVERGENCE in  $\mathcal{B}$ .

The countable set  $\{x_n \in X \mid n \in \mathbb{N}\}$  is a **Schauder basis** for  $\mathcal{B}$  if for each  $x \in X$

$$1. \quad \exists (\alpha_n \in \mathbb{F})_{n \in \mathbb{N}} \quad \text{such that} \quad x \doteq \sum_{n=1}^{\infty} \alpha_n x_n \quad (\text{STRONG CONVERGENCE in } \mathcal{B} \text{ and})$$

$$2. \quad \left\{ \sum_{n=1}^{\infty} \alpha_n x_n \doteq \sum_{n=1}^{\infty} \beta_n x_n \right\} \implies \{(\alpha_n) = (\beta_n)\} \quad (\text{COEFFICIENT FUNCTIONALS are UNIQUE})$$

DEF

In this case,  $\sum_{n=1}^{\infty} \alpha_n x_n$  is the **expansion** of  $x$  on  $\{x_n \mid n \in \mathbb{N}\}$  and

the elements of  $(\alpha_n)$  are the **coefficient functionals** associated with the basis  $\{x_n\}$ . Coefficient functionals are also called **coordinate functionals**.

In a Banach space, the existence of a Schauder basis implies that the space is *separable* (Theorem H.3 page 214). The question of whether the converse is also true was posed by Banach himself in 1932,<sup>13</sup> and became known 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.<sup>14</sup> 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 H.11 page 226).

**Theorem H.3.** <sup>15</sup> Let  $\mathcal{B} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$  be a BANACH SPACE. Let  $\mathbb{Q}$  be the field of rational numbers.

$$\begin{array}{c} \text{T} \\ \text{H} \\ \text{M} \end{array} \left\{ \begin{array}{l} 1. \quad \mathcal{B} \text{ has a SCHAUDER BASIS and} \\ 2. \quad \mathbb{Q} \text{ is DENSE in } \mathbb{F}. \end{array} \right\} \implies \{ \mathcal{B} \text{ is SEPARABLE} \}$$

PROOF:

1. lemma:

$$\begin{aligned} \left| \left\{ x \mid \exists (\alpha_n \in \mathbb{Q})_{n \in \mathbb{N}} \text{ such that } \lim_{N \rightarrow \infty} \left\| x - \sum_{n=1}^N \alpha_n x_n \right\| = 0 \right\} \right| &= |\mathbb{Q} \times \mathbb{N}| \\ &= |\mathbb{Z} \times \mathbb{Z}| \\ &= |\mathbb{Z}| \\ &= \text{countably infinite} \end{aligned}$$

<sup>12</sup> 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)

<sup>13</sup> Banach (1932a), page 111

<sup>14</sup> Enflo (1973), Lindenstrauss and Tzafriri (1977) pages 84–95 (Section 2.d)

<sup>15</sup> Bachman et al. (2000) page 112 (3.4.8), Giles (2000) page 17, Heil (2011) page 21 (Theorem 1.27)



2. remainder of proof:

$\mathcal{B}$  has a Schauder basis  $(\mathbf{x}_n)_{n \in \mathbb{N}}$

$\implies$  for every  $\mathbf{x} \in \mathcal{B}$ , there exists  $(\alpha_n \in \mathbb{F})_{n \in \mathbb{N}}$  such that  $\mathbf{x} \doteq \sum_{n=1}^{\infty} \alpha_n \mathbf{x}_n$  by Definition H.7 page 214

$\implies$  for every  $\mathbf{x} \in \mathcal{B}$ , there exists  $(\alpha_n \in \mathbb{F})_{n \in \mathbb{N}}$  such that  $\lim_{N \rightarrow \infty} \left\| \mathbf{x} - \sum_{n=1}^N \alpha_n \mathbf{x}_n \right\| = 0$

$\implies$  for every  $\mathbf{x} \in \mathcal{B}$ , there exists  $(\alpha_n \in \mathbb{Q})_{n \in \mathbb{N}}$  such that  $\lim_{N \rightarrow \infty} \left\| \mathbf{x} - \sum_{n=1}^N \alpha_n \mathbf{x}_n \right\| = 0$  because  $\mathbb{Q}^- = \mathbb{F}$

$\implies \mathcal{B} = \left\{ \mathbf{x} \mid \exists (\alpha_n \in \mathbb{Q})_{n \in \mathbb{N}} \text{ such that } \lim_{N \rightarrow \infty} \left\| \mathbf{x} - \sum_{n=1}^N \alpha_n \mathbf{x}_n \right\| = 0 \right\}$

$\implies \mathcal{B} = \left\{ \mathbf{x} \mid \exists (\alpha_n \in \mathbb{Q})_{n \in \mathbb{N}} \text{ such that } \mathbf{x} = \lim_{N \rightarrow \infty} \sum_{n=1}^N \alpha_n \mathbf{x}_n \right\}$

$\implies \mathcal{B}$  is separable by (1) lemma page 214



**Definition H.8.** <sup>16</sup> Let  $\{\mathbf{x}_n | n \in \mathbb{N}\}$  and  $\{\mathbf{y}_n | n \in \mathbb{N}\}$  be SCHAUDER BASES of a BANACH SPACE  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$ .

**D E F**  $\{\mathbf{x}_n\}$  is equivalent to  $\{\mathbf{y}_n\}$   
if there exists a BOUNDED INVERTIBLE operator  $\mathbf{R}$  in  $X^X$  such that  $\mathbf{R}\mathbf{x}_n = \mathbf{y}_n \quad \forall n \in \mathbb{Z}$

**Theorem H.4.** <sup>17</sup> Let  $\{\mathbf{x}_n | n \in \mathbb{N}\}$  and  $\{\mathbf{y}_n | n \in \mathbb{N}\}$  be SCHAUDER BASES of a BANACH SPACE  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$ .

**T H M**  $\{\{\mathbf{x}_n\} \text{ is EQUIVALENT to } \{\mathbf{y}_n\}\}$   
 $\iff \left\{ \sum_{n=1}^{\infty} \alpha_n \mathbf{x}_n \text{ is CONVERGENT} \iff \sum_{n=1}^{\infty} \alpha_n \mathbf{y}_n \text{ is CONVERGENT} \right\}$

**Lemma H.1.** <sup>18</sup> Let  $\Omega \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), T)$  be a topological linear space. Let  $\text{span} A$  be the SPAN of a set  $A$  (Definition H.2 page 209). Let  $\tilde{f}(\omega)$  and  $\tilde{g}(\omega)$  be the FOURIER TRANSFORMS (Definition P.2 page 337) of the functions  $f(x)$  and  $g(x)$ , respectively, in  $L^2_{\mathbb{R}}$  (Definition ?? page ??). Let  $\check{a}(\omega)$  be the DTFT (Definition Q.1 page 347) of a sequence  $(a_n)_{n \in \mathbb{Z}}$  in  $\ell^2_{\mathbb{R}}$  (Definition R.2 page 357).

**L E M**  $\left\{ \begin{array}{l} (1). \quad \left\{ \mathbf{T}^n f | n \in \mathbb{Z} \right\} \text{ is a SCHAUDER BASIS for } \Omega \quad \text{and} \\ (2). \quad \left\{ \mathbf{T}^n g | n \in \mathbb{Z} \right\} \text{ is a SCHAUDER BASIS for } \Omega \end{array} \right\} \implies \left\{ \begin{array}{l} \exists (a_n)_{n \in \mathbb{Z}} \text{ such that} \\ \tilde{f}(\omega) = \check{a}(\omega) \tilde{g}(\omega) \end{array} \right\}$

PROOF: Let  $V'_0$  be the space spanned by  $\{\mathbf{T}^n \phi | n \in \mathbb{Z}\}$ .

$$\begin{aligned} \tilde{f}(\omega) &\triangleq \tilde{F}f && \text{by definition of } \tilde{F} \\ &= \tilde{F} \sum_{n \in \mathbb{Z}} a_n Tg && \text{by (2)} \\ &= \sum_{n \in \mathbb{Z}} a_n \tilde{F}Tg \end{aligned}$$

<sup>16</sup> Young (2001) page 25 (Definition 1.8.1, Theorem 1.8.7)

<sup>17</sup> Young (2001) page 25 (Definition 1.8.1, Theorem 1.8.7)

<sup>18</sup> Daubechies (1992), page 140

$$\begin{aligned}
 &= \underbrace{\sum_{n \in \mathbb{Z}} a_n e^{-i\omega n} \tilde{\mathbf{F}} \mathbf{g}}_{\check{a}(\omega)} \quad \text{by Corollary ?? page ??} \\
 &= \check{a}(\omega) \tilde{\mathbf{g}}(\omega) \quad \text{by definition of } \check{\mathbf{F}} \text{ and } \tilde{\mathbf{F}} \quad \text{by (Definition Q.1 page 347, Definition P.2 page 337)}
 \end{aligned}$$

$$\begin{aligned}
 V_0 &\triangleq \left\{ f(x) | f(x) = \sum_{n \in \mathbb{Z}} b_n T^n g(x) \right\} \\
 &= \left\{ f(x) | \tilde{\mathbf{F}} f(x) = \tilde{\mathbf{F}} \sum_{n \in \mathbb{Z}} b_n T^n g(x) \right\} \\
 &= \left\{ f(x) | \tilde{f}(\omega) = \tilde{b}(\omega) \tilde{\mathbf{g}}(\omega) \right\} \\
 &= \left\{ f(x) | \tilde{f}(\omega) = \tilde{b}(\omega) \check{a}(\omega) \tilde{f}(\omega) \right\} \\
 &= \left\{ f(x) | \tilde{f}(\omega) = \tilde{c}(\omega) \tilde{f}(\omega) \right\} \quad \text{where } \tilde{c}(\omega) \triangleq \tilde{b}(\omega) \check{a}(\omega) \\
 &= \left\{ f(x) | f(x) = \sum_{n \in \mathbb{Z}} c_n f(x - n) \right\} \\
 &\triangleq V'_0
 \end{aligned}$$



## H.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 H.5.** <sup>19</sup> Let  $\{x_n \in X | n=1,2,\dots,N\}$  be a set of vectors in an INNER PRODUCT SPACE (Definition K.1 page 259)  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), (\Delta | \nabla))$ .

<b>T H M</b>	$\left\{ \begin{array}{l} \{x_n\} \text{ is ORTHOGONAL} \\ (\text{Definition K.4 page 271}) \end{array} \right\} \implies \left\{ \begin{array}{l} \{x_n\} \text{ is LINEARLY INDEPENDENT} \\ (\text{Definition H.1 page 209}) \end{array} \right\}$
----------------------	--

PROOF:

1. Proof using *Pythagorean theorem* (Theorem K.10 page 273):

Let  $(\alpha_n)_{n \in \mathbb{N}}$  be a sequence with at least one nonzero element.

$$\begin{aligned}
 \left\| \sum_{n=1}^N \alpha_n x_n \right\|^2 &= \sum_{n=1}^N \|\alpha_n x_n\|^2 \quad \text{by left hypoth. and Pythagorean Theorem (Theorem K.10 page 273)} \\
 &= \sum_{n=1}^N |\alpha_n|^2 \|x_n\|^2 \quad \text{by definition of } \|\cdot\| \quad (\text{Definition L.1 page 275}) \\
 &> 0 \\
 \implies \sum_{n=1}^N \alpha_n x_n &\neq 0 \\
 \implies (\alpha_n)_{n \in \mathbb{N}} \text{ is linearly independent} &\quad \text{by definition of linear independence} \quad (\text{Definition H.3 page 209})
 \end{aligned}$$

<sup>19</sup> Aliprantis and Burkinshaw (1998) page 283 (Corollary 32.8), Kubrusly (2001) page 352 (Proposition 5.34)

2. Alternative proof:

$$\begin{aligned}
 \sum_{n=1}^N \alpha_n \mathbf{x}_n = \mathbb{0} &\implies \left\langle \sum_{n=1}^N \alpha_n \mathbf{x}_n \mid \mathbf{x}_m \right\rangle = \langle \mathbb{0} \mid \mathbf{x}_m \rangle \\
 &\implies \sum_{n=1}^N \alpha_n \langle \mathbf{x}_n \mid \mathbf{x}_m \rangle = 0 \\
 &\implies \sum_{n=1}^N \alpha_n \bar{\delta}(k-m) = 0 \\
 &\implies \alpha_m = 0 \quad \text{for } m = 1, 2, \dots, N
 \end{aligned}$$

⇒

**Theorem H.6** (Bessel's Equality). <sup>20</sup> Let  $\{\mathbf{x}_n \in X \mid n=1, 2, \dots, N\}$  be a set of vectors in an INNER PRODUCT SPACE (Definition K.1 page 259)  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), (\triangle \mid \nabla))$  and with  $\|\mathbf{x}\| \triangleq \sqrt{\langle \mathbf{x} \mid \mathbf{x} \rangle}$  (Definition K.2 page 265).

T  
H  
M

$$\left\{ \begin{array}{l} \{\mathbf{x}_n\} \text{ is ORTHONORMAL} \\ (\text{Definition K.4 page 271}) \end{array} \right\} \implies \left\{ \underbrace{\left\| \mathbf{x} - \sum_{n=1}^N \langle \mathbf{x} \mid \mathbf{x}_n \rangle \mathbf{x}_n \right\|^2}_{\text{approximation error}} = \|\mathbf{x}\|^2 - \sum_{n=1}^N |\langle \mathbf{x} \mid \mathbf{x}_n \rangle|^2 \quad \forall \mathbf{x} \in X \right\}$$

PROOF:

$$\begin{aligned}
 &\left\| \mathbf{x} - \sum_{n=1}^N \langle \mathbf{x} \mid \mathbf{x}_n \rangle \mathbf{x}_n \right\|^2 \\
 &= \|\mathbf{x}\|^2 + \left\| \sum_{n=1}^N \langle \mathbf{x} \mid \mathbf{x}_n \rangle \mathbf{x}_n \right\|^2 - 2\Re \left\langle \mathbf{x} \left| \sum_{n=1}^N \langle \mathbf{x} \mid \mathbf{x}_n \rangle \mathbf{x}_n \right. \right\rangle \quad \text{by polar identity} \quad (\text{Lemma K.1 page 263}) \\
 &= \|\mathbf{x}\|^2 + \left\| \sum_{n=1}^N \langle \mathbf{x} \mid \mathbf{x}_n \rangle \mathbf{x}_n \right\|^2 - 2\Re \left[ \left( \sum_{n=1}^N \langle \mathbf{x} \mid \mathbf{x}_n \rangle \right)^* \langle \mathbf{x} \mid \mathbf{x}_n \rangle \right] \quad \text{by property of } (\triangle \mid \nabla) \quad (\text{Definition K.1 page 259}) \\
 &= \|\mathbf{x}\|^2 + \sum_{n=1}^N \|\langle \mathbf{x} \mid \mathbf{x}_n \rangle \mathbf{x}_n\|^2 - 2\Re \left[ \left( \sum_{n=1}^N \langle \mathbf{x} \mid \mathbf{x}_n \rangle \right)^* \langle \mathbf{x} \mid \mathbf{x}_n \rangle \right] \quad \text{by Pythagorean Theorem} \quad (\text{Theorem K.10 page 273}) \\
 &= \|\mathbf{x}\|^2 + \sum_{n=1}^N \|\langle \mathbf{x} \mid \mathbf{x}_n \rangle \mathbf{x}_n\|^2 - 2\Re \left( \sum_{n=1}^N \langle \mathbf{x} \mid \mathbf{x}_n \rangle^* \langle \mathbf{x} \mid \mathbf{x}_n \rangle \right) \\
 &= \|\mathbf{x}\|^2 + \sum_{n=1}^N |\langle \mathbf{x} \mid \mathbf{x}_n \rangle|^2 \underbrace{\|\mathbf{x}_n\|^2}_1 - 2\Re \left( \sum_{n=1}^N \langle \mathbf{x} \mid \mathbf{x}_n \rangle^* \langle \mathbf{x} \mid \mathbf{x}_n \rangle \right) \quad \text{by property of } \|\cdot\| \quad (\text{Definition L.1 page 275}) \\
 &= \|\mathbf{x}\|^2 + \sum_{n=1}^N |\langle \mathbf{x} \mid \mathbf{x}_n \rangle|^2 \cdot 1 - 2\Re \left( \sum_{n=1}^N \langle \mathbf{x} \mid \mathbf{x}_n \rangle^* \langle \mathbf{x} \mid \mathbf{x}_n \rangle \right) \quad \text{by def. of orthonormality} \quad (\text{Definition K.4 page 271}) \\
 &= \|\mathbf{x}\|^2 + \sum_{n=1}^N |\langle \mathbf{x} \mid \mathbf{x}_n \rangle|^2 - 2\Re \sum_{n=1}^N |\langle \mathbf{x} \mid \mathbf{x}_n \rangle|^2 \\
 &= \|\mathbf{x}\|^2 + \sum_{n=1}^N |\langle \mathbf{x} \mid \mathbf{x}_n \rangle|^2 - 2 \sum_{n=1}^N |\langle \mathbf{x} \mid \mathbf{x}_n \rangle|^2 \quad \text{because } |\cdot| \text{ is real}
 \end{aligned}$$

<sup>20</sup> Bachman et al. (2000) page 103, Pedersen (2000) pages 38–39

$$= \|\mathbf{x}\|^2 - \sum_{n=1}^N |\langle \mathbf{x} | \mathbf{x}_n \rangle|^2$$

⇒

**Theorem H.7** (Bessel's inequality). <sup>21</sup> Let  $\{\mathbf{x}_n \in X \mid n=1,2,\dots,N\}$  be a set of vectors in an INNER PRODUCT SPACE (Definition K.1 page 259) ( $X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \Delta | \nabla \rangle$ ) and with  $\|\mathbf{x}\| \triangleq \sqrt{\langle \mathbf{x} | \mathbf{x} \rangle}$  (Definition K.2 page 265).

T H M	$\left\{ \begin{array}{l} \{\mathbf{x}_n\} \text{ is ORTHONORMAL} \\ (\text{Definition K.4 page 271}) \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} \sum_{n=1}^N  \langle \mathbf{x}   \mathbf{x}_n \rangle ^2 \leq \ \mathbf{x}\ ^2 \quad \forall \mathbf{x} \in X \end{array} \right\}$
-------------	---

PROOF:

$$\begin{aligned} 0 &\leq \left\| \mathbf{x} - \sum_{n=1}^N \langle \mathbf{x} | \mathbf{x}_n \rangle \mathbf{x}_n \right\|^2 && \text{by definition of } \|\cdot\| && (\text{Definition L.1 page 275}) \\ &= \|\mathbf{x}\|^2 - \sum_{n=1}^N |\langle \mathbf{x} | \mathbf{x}_n \rangle|^2 && \text{by Bessel's Equality} && (\text{Theorem H.6 page 217}) \end{aligned}$$

⇒

The *Best Approximation Theorem* (next) shows that

- ➊ the best sequence for representing a vector is the sequence of projections of the vector onto the sequence of basis functions
- ➋ the error of the projection is orthogonal to the projection.

**Theorem H.8** (Best Approximation Theorem). <sup>22</sup> Let  $\{\mathbf{x}_n \in X \mid n=1,2,\dots,N\}$  be a set of vectors in an INNER PRODUCT SPACE (Definition K.1 page 259) ( $X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \Delta | \nabla \rangle$ ) and with  $\|\mathbf{x}\| \triangleq \sqrt{\langle \mathbf{x} | \mathbf{x} \rangle}$  (Definition K.2 page 265).

T H M	$\left\{ \begin{array}{l} \{\mathbf{x}_n\} \text{ is} \\ \text{ORTHONORMAL} \\ (\text{Definition K.4 page 271}) \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} 1. \arg \min_{(\alpha_n)_{n=1}^N} \left\  \mathbf{x} - \sum_{n=1}^N \alpha_n \mathbf{x}_n \right\  = \underbrace{(\langle \mathbf{x}   \mathbf{x}_n \rangle)_{n=1}^N}_{\text{best } \alpha_n = \langle \mathbf{x}   \mathbf{x}_n \rangle} \quad \forall \mathbf{x} \in X \quad \text{and} \\ 2. \underbrace{\left( \sum_{n=1}^N \langle \mathbf{x}   \mathbf{x}_n \rangle \mathbf{x}_n \right)}_{\text{approximation}} \perp \underbrace{\left( \mathbf{x} - \sum_{n=1}^N \langle \mathbf{x}   \mathbf{x}_n \rangle \mathbf{x}_n \right)}_{\text{approximation error}} \quad \forall \mathbf{x} \in X \end{array} \right\}$
-------------	---

PROOF:

<sup>21</sup>  Giles (2000) pages 54–55 (3.13 Bessel's inequality),  Bollobás (1999) page 147,  Aliprantis and Burkinshaw (1998) page 284

<sup>22</sup>  Walter and Shen (2001), pages 3–4,  Pedersen (2000), page 39,  Edwards (1995), pages 94–100,  Weyl (1940)

1. Proof that  $(\langle \mathbf{x} | \mathbf{x}_n \rangle)$  is the best sequence:

$$\begin{aligned}
 & \left\| \mathbf{x} - \sum_{n=1}^N \alpha_n \mathbf{x}_n \right\|^2 \\
 &= \| \mathbf{x} \|^2 - 2 \Re \left\langle \mathbf{x} \mid \sum_{n=1}^N \alpha_n \mathbf{x}_n \right\rangle + \left\| \sum_{n=1}^N \alpha_n \mathbf{x}_n \right\|^2 \\
 &= \| \mathbf{x} \|^2 - 2 \Re \left( \sum_{n=1}^N \alpha_n^* \langle \mathbf{x} | \mathbf{x}_n \rangle \right) + \sum_{n=1}^N \| \alpha_n \mathbf{x}_n \|^2 \quad \text{by Pythagorean Theorem} \quad (\text{Theorem K.10 page 273}) \\
 &= \| \mathbf{x} \|^2 - 2 \Re \left( \sum_{n=1}^N \alpha_n^* \langle \mathbf{x} | \mathbf{x}_n \rangle \right) + \sum_{n=1}^N | \alpha_n |^2 + \underbrace{\left[ \sum_{n=1}^N | \langle \mathbf{x} | \mathbf{x}_n \rangle |^2 - \sum_{n=1}^N | \langle \mathbf{x} | \mathbf{x}_n \rangle |^2 \right]}_0 \\
 &= \left[ \| \mathbf{x} \|^2 - \sum_{n=1}^N | \langle \mathbf{x} | \mathbf{x}_n \rangle |^2 \right] + \sum_{n=1}^N \left[ | \langle \mathbf{x} | \mathbf{x}_n \rangle |^2 - 2 \Re_e [\alpha_n^* \langle \mathbf{x} | \mathbf{x}_n \rangle] + | \alpha_n |^2 \right] \\
 &= \left[ \| \mathbf{x} \|^2 - \sum_{n=1}^N | \langle \mathbf{x} | \mathbf{x}_n \rangle |^2 \right] + \sum_{n=1}^N [| \langle \mathbf{x} | \mathbf{x}_n \rangle |^2 - \alpha_n^* \langle \mathbf{x} | \mathbf{x}_n \rangle - \alpha_n \langle \mathbf{x} | \mathbf{x}_n \rangle^* + | \alpha_n |^2] \\
 &= \left\| \mathbf{x} - \sum_{n=1}^N \langle \mathbf{x} | \mathbf{x}_n \rangle \mathbf{x}_n \right\|^2 + \sum_{n=1}^N | \langle \mathbf{x} | \mathbf{x}_n \rangle - \alpha_n |^2 \quad \text{by Bessel's Equality} \quad (\text{Theorem H.6 page 217}) \\
 &\geq \left\| \mathbf{x} - \sum_{n=1}^N \langle \mathbf{x} | \mathbf{x}_n \rangle \mathbf{x}_n \right\|^2
 \end{aligned}$$

2. Proof that the approximation and approximation error are orthogonal:

$$\begin{aligned}
 \left\langle \sum_{n=1}^N \langle \mathbf{x} | \mathbf{x}_n \rangle \mathbf{x}_n \mid \mathbf{x} - \sum_{n=1}^N \langle \mathbf{x} | \mathbf{x}_n \rangle \mathbf{x}_n \right\rangle &= \left\langle \sum_{n=1}^N \langle \mathbf{x} | \mathbf{x}_n \rangle \mathbf{x}_n \mid \mathbf{x} \right\rangle - \left\langle \sum_{n=1}^N \langle \mathbf{x} | \mathbf{x}_n \rangle \mathbf{x}_n \mid \sum_{n=1}^N \langle \mathbf{x} | \mathbf{x}_n \rangle \mathbf{x}_n \right\rangle \\
 &= \sum_{n=1}^N \langle \mathbf{x} | \mathbf{x}_n \rangle^* \langle \mathbf{x} | \mathbf{x}_n \rangle - \sum_{n=1}^N \sum_{m=1}^N \langle \mathbf{x} | \mathbf{x}_n \rangle \langle \mathbf{x} | \mathbf{x}_m \rangle^* \langle \mathbf{x}_n | \mathbf{x}_m \rangle \\
 &= \sum_{n=1}^N | \langle \mathbf{x} | \mathbf{x}_n \rangle |^2 - \sum_{n=1}^N \sum_{m=1}^N \langle \mathbf{x} | \mathbf{x}_n \rangle \langle \mathbf{x} | \mathbf{x}_m \rangle^* \bar{\delta}_{nm} \\
 &= \sum_{n=1}^N | \langle \mathbf{x} | \mathbf{x}_n \rangle |^2 - \sum_{n=1}^N | \langle \mathbf{x} | \mathbf{x}_n \rangle |^2 \\
 &= 0
 \end{aligned}$$

⇒

## H.5 Orthonormal bases in Hilbert spaces

**Definition H.9.** Let  $\{ \mathbf{x}_n \in X \mid n=1,2,\dots,N \}$  be a set of vectors in an INNER PRODUCT SPACE (Definition K.1 page 259)  $\Omega \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle | \triangledown \rangle)$ .

**D E F** The set  $\{x_n\}$  is an **orthogonal basis** for  $\Omega$  if  $\{x_n\}$  is ORTHOGONAL and is a SCHAUDER BASIS for  $\Omega$ .  
The set  $\{x_n\}$  is an **orthonormal basis** for  $\Omega$  if  $\{x_n\}$  is ORTHONORMAL and is a SCHAUDER BASIS for  $\Omega$ .

**Definition H.10.**<sup>23</sup> Let  $H \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \Delta | \nabla \rangle)$  be a Hilbert space.

**D E F** Suppose there exists a set  $\{x_n \in X \mid n \in \mathbb{N}\}$  such that  $x \triangleq \sum_{n=1}^{\infty} \langle x | x_n \rangle x_n$ . Then the quantities  $\langle x | x_n \rangle$  are called the **Fourier coefficients** of  $x$  and the sum  $\sum_{n=1}^{\infty} \langle x | x_n \rangle x_n$  is called the **Fourier expansion** of  $x$  or the **Fourier series** for  $x$ .

**Lemma H.2** (Perfect reconstruction). Let  $\{x_n \in X \mid n \in \mathbb{N}\}$  be a set of vectors in a HILBERT SPACE  $H \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \Delta | \nabla \rangle)$ .

**L E M**  $\left\{ \begin{array}{l} (1). \quad \{x_n\} \text{ is a BASIS for } H \\ (2). \quad \{x_n\} \text{ is ORTHONORMAL} \end{array} \right. \text{ and } \Rightarrow x \triangleq \underbrace{\sum_{n=1}^{\infty} \underbrace{\langle x | x_n \rangle}_{\text{Fourier coefficient}} x_n}_{\text{Fourier expansion}} \quad \forall x \in X$

PROOF:

$$\begin{aligned} \langle x | x_n \rangle &= \left\langle \sum_{m \in \mathbb{Z}} \alpha_m x_m | x_n \right\rangle && \text{by left hypothesis (1)} \\ &= \sum_{m \in \mathbb{Z}} \alpha_m \langle x_m | x_n \rangle && \text{by homogeneous property of } \langle \Delta | \nabla \rangle \quad (\text{Definition K.1 page 259}) \\ &= \sum_{m \in \mathbb{Z}} \alpha_m \delta_{n-m} && \text{by left hypothesis (2)} \quad (\text{Definition K.4 page 271}) \\ &= \alpha_n \end{aligned}$$



**Proposition H.2.**<sup>24</sup> Let  $\{x_n \in X \mid n \in \mathbb{N}\}$  be a set of vectors in a HILBERT SPACE  $H \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \Delta | \nabla \rangle)$ .

**P R P**  $\|x\|^2 \triangleq \underbrace{\sum_{n=1}^{\infty} |\langle x | x_n \rangle|^2}_{\text{PARSEVAL FRAME}} \iff x \triangleq \underbrace{\sum_{n=1}^{\infty} \langle x | x_n \rangle x_n}_{\text{FOURIER EXPANSION (Definition H.10 page 220)}} \quad \forall x \in X$

PROOF:

<sup>23</sup> Fabian et al. (2010) page 27 (Theorem 1.55), Young (2001) page 6, Young (1980) page 6

<sup>24</sup> Han et al. (2007) pages 93–94 (Proposition 3.11)

1. Proof that *Parseval frame*  $\iff$  *Fourier expansion*

$$\begin{aligned}
 \|x\|^2 &\triangleq \langle x | x \rangle && \text{by definition of } \|\cdot\| && (\text{Definition L.1 page 275}) \\
 &= \left\langle \sum_{n=1}^{\infty} \langle x | x_n \rangle x | x_n \right\rangle && \text{by right hypothesis} \\
 &\stackrel{*}{=} \sum_{n=1}^{\infty} \langle x | x_n \rangle \langle x | x_n \rangle && \text{by property of } \langle \Delta | \nabla \rangle && (\text{Definition K.1 page 259}) \\
 &\stackrel{*}{=} \sum_{n=1}^{\infty} \langle x | x_n \rangle \langle x | x_n \rangle^* && \text{by property of } \langle \Delta | \nabla \rangle && (\text{Definition K.1 page 259}) \\
 &\stackrel{*}{=} \sum_{n=1}^{\infty} |\langle x | x_n \rangle|^2 && \text{by property of } \mathbb{C} && (\text{Definition J.7 page 257})
 \end{aligned}$$

2. Proof that *Parseval frame*  $\implies$  *Fourier expansion*

(a) Let  $(e_n)_{n \in \mathbb{N}}$  be the *standard orthonormal basis* such that the  $n$ th element of  $e_n$  is 1 and all other elements are 0.

(b) Let  $M$  be an operator in  $H$  such that  $Mx \triangleq \sum_{n=1}^{\infty} \langle x | x_n \rangle e_n$ .

(c) lemma:  $M$  is *isometric*. Proof:

$$\begin{aligned}
 \|Mx\|^2 &= \left\| \sum_{n=1}^{\infty} \langle x | x_n \rangle e_n \right\|^2 && \text{by definition of } M && (\text{item (2b) page 221}) \\
 &= \sum_{n=1}^{\infty} \|\langle x | x_n \rangle e_n\|^2 && \text{by Pythagorean Theorem} && (\text{Theorem K.10 page 273}) \\
 &= \sum_{n=1}^{\infty} |\langle x | x_n \rangle|^2 \|e_n\|^2 && \text{by homogeneous property of } \|\cdot\| && (\text{Definition L.1 page 275}) \\
 &= \sum_{n=1}^{\infty} |\langle x | x_n \rangle|^2 && \text{by definition of orthonormal} && (\text{Definition K.4 page 271}) \\
 &= \|x\|^2 && \text{by Parseval frame hypothesis} \\
 \implies M &\text{ is isometric} && \text{by definition of isometric} && (\text{Definition O.10 page 327})
 \end{aligned}$$

(d) Let  $(u_n)_{n \in \mathbb{N}}$  be an *orthonormal basis* for  $H$ .

(e) Proof for *Fourier expansion*:

$$\begin{aligned}
 x &= \sum_{n=1}^{\infty} \langle x | u_n \rangle u_n && \text{by Fourier expansion (Proposition H.3 page 224)} \\
 &= \sum_{n=1}^{\infty} \langle Mx | Mu_n \rangle u_n && \text{by (2c) lemma page 221 and Theorem O.23 page 328} \\
 &= \sum_{n=1}^{\infty} \left\langle \sum_{m=1}^{\infty} \langle x | x_m \rangle e_m | \sum_{k=1}^{\infty} \langle u_n | x_k \rangle e_k \right\rangle u_n && \text{by item (2b) page 221} \\
 &= \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \langle x | x_m \rangle \sum_{k=1}^{\infty} \langle u_n | x_k \rangle^* \langle e_m | e_k \rangle u_n && \text{by Definition K.1 page 259} \\
 &= \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \langle x | x_m \rangle \langle u_n | x_m \rangle^* u_n && \text{by item (2a) page 221 and Definition K.4 page 271}
 \end{aligned}$$

$$\begin{aligned}
 &= \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \langle \mathbf{x} | \mathbf{x}_m \rangle \langle \mathbf{x}_m | \mathbf{u}_n \rangle \mathbf{u}_n && \text{by Definition K.1 page 259} \\
 &= \sum_{m=1}^{\infty} \langle \mathbf{x} | \mathbf{x}_m \rangle \sum_{n=1}^{\infty} \langle \mathbf{x}_m | \mathbf{u}_n \rangle \mathbf{u}_n \\
 &= \sum_{m=1}^{\infty} \langle \mathbf{x} | \mathbf{x}_m \rangle \mathbf{x}_m && \text{by item (2d) page 221}
 \end{aligned}$$

⇒

When is a set of orthonormal vectors in a Hilbert space  $\mathbf{H}$  *total*? Theorem H.9 (next) offers some help.

**Theorem H.9** (The Fourier Series Theorem). <sup>25</sup> Let  $\{\mathbf{x}_n \in X \mid n \in \mathbb{N}\}$  be a set of vectors in a HILBERT SPACE  $\mathbf{H} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), (\triangle | \nabla))$  and let  $\|\mathbf{x}\| \triangleq \sqrt{\langle \mathbf{x} | \mathbf{x} \rangle}$  (Definition K.2 page 265).

T  
H  
M

$$\left. \begin{array}{l}
 \text{(A) } \{\mathbf{x}_n\} \text{ is ORTHONORMAL in } \mathbf{H} \implies \\
 \text{(1). } (\text{span}\{\mathbf{x}_n\})^\perp = \mathbf{H} \quad (\{\mathbf{x}_n\} \text{ is TOTAL in } \mathbf{H}) \\
 \iff \text{(2). } \langle \mathbf{x} | \mathbf{y} \rangle \triangleq \sum_{n=1}^{\infty} \langle \mathbf{x} | \mathbf{x}_n \rangle \langle \mathbf{y} | \mathbf{x}_n \rangle^* \quad \forall \mathbf{x}, \mathbf{y} \in X \quad (\text{GENERALIZED PARSEVAL'S IDENTITY}) \\
 \iff \text{(3). } \|\mathbf{x}\|^2 \triangleq \sum_{n=1}^{\infty} |\langle \mathbf{x} | \mathbf{x}_n \rangle|^2 \quad \forall \mathbf{x} \in X \quad (\text{PARSEVAL'S IDENTITY}) \\
 \iff \text{(4). } \mathbf{x} \triangleq \sum_{n=1}^{\infty} \langle \mathbf{x} | \mathbf{x}_n \rangle \mathbf{x}_n \quad \forall \mathbf{x} \in X \quad (\text{FOURIER SERIES EXPANSION})
 \end{array} \right\}$$

⇒ PROOF:

1. Proof that (1)  $\implies$  (2):

$$\begin{aligned}
 \langle \mathbf{x} | \mathbf{y} \rangle &\triangleq \left\langle \sum_{n=1}^{\infty} \langle \mathbf{x} | \mathbf{x}_n \rangle \mathbf{x}_n \mid \sum_{m=1}^{\infty} \langle \mathbf{y} | \mathbf{x}_m \rangle \mathbf{x}_m \right\rangle && \text{by (A) and (1)} \\
 &\triangleq \sum_{n=1}^{\infty} \langle \mathbf{x} | \mathbf{x}_n \rangle \left\langle \mathbf{x}_n \mid \sum_{m=1}^{\infty} \langle \mathbf{y} | \mathbf{x}_m \rangle \mathbf{x}_m \right\rangle && \text{by property of } \langle \cdot | \cdot \rangle \quad (\text{Definition K.1 page 259}) \\
 &\triangleq \sum_{n=1}^{\infty} \langle \mathbf{x} | \mathbf{x}_n \rangle \sum_{m=1}^{\infty} \langle \mathbf{y} | \mathbf{x}_m \rangle^* \langle \mathbf{x}_n | \mathbf{x}_m \rangle && \text{by property of } \langle \cdot | \cdot \rangle \quad (\text{Definition K.1 page 259}) \\
 &\triangleq \sum_{n=1}^{\infty} \langle \mathbf{x} | \mathbf{x}_n \rangle \sum_{m=1}^{\infty} \langle \mathbf{y} | \mathbf{x}_m \rangle^* \bar{\delta}_{mn} && \text{by (A)} \\
 &\triangleq \sum_{n=1}^{\infty} \langle \mathbf{x} | \mathbf{x}_n \rangle \langle \mathbf{y} | \mathbf{x}_n \rangle^* && \text{by definition of } \bar{\delta}_n \quad (\text{Definition K.3 page 271})
 \end{aligned}$$

<sup>25</sup> 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):

$$\begin{aligned} \|\mathbf{x}\|^2 &\triangleq \langle \mathbf{x} | \mathbf{x} \rangle && \text{by definition of } \textit{induced norm} && (\text{Theorem K.4 page 264}) \\ &= \sum_{n=1}^{\infty} \langle \mathbf{x} | \mathbf{x}_n \rangle \langle \mathbf{x} | \mathbf{x}_n \rangle^* && \text{by (2)} \\ &= \sum_{n=1}^{\infty} |\langle \mathbf{x} | \mathbf{x}_n \rangle|^2 \end{aligned}$$

3. Proof that (3)  $\iff$  (4) *not* using (A): by Proposition H.2 page 220

4. Proof that (3)  $\implies$  (1) (proof by contradiction):

- (a) Suppose  $\{\mathbf{x}_n\}$  is *not total*.
- (b) Then there must exist a vector  $\mathbf{y}$  in  $H$  such that the set  $B \triangleq \{\mathbf{x}_n\} \cup \mathbf{y}$  is *orthonormal*.
- (c) Then  $1 = \|\mathbf{y}\|^2 \neq \sum_{n=1}^{\infty} |\langle \mathbf{y} | \mathbf{x}_n \rangle|^2 = 0$ .
- (d) But this contradicts (3), and so  $\{\mathbf{x}_n\}$  must be *total* and (3)  $\implies$  (1).

5. Extraneous proof that (3)  $\implies$  (4) (this proof is not really necessary here):

$$\begin{aligned} \left\| \mathbf{x} - \sum_{n=1}^{\infty} \langle \mathbf{x} | \mathbf{x}_n \rangle \mathbf{x}_n \right\|^2 &= \|\mathbf{x}\|^2 - \sum_{n=1}^{\infty} |\langle \mathbf{x} | \mathbf{x}_n \rangle|^2 && \text{by Bessel's Equality} && (\text{Theorem H.6 page 217}) \\ &= 0 && \text{by (3)} \\ \implies \mathbf{x} &\stackrel{*}{=} \sum_{n=1}^{\infty} \langle \mathbf{x} | \mathbf{x}_n \rangle \mathbf{x}_n && \text{by definition of } \stackrel{*}{=} \end{aligned}$$

6. Extraneous proof that (A)  $\implies$  (4) (this proof is not really necessary here)

- (a) The sequence  $\sum_{n=1}^N |\langle \mathbf{x} | \mathbf{x}_n \rangle|^2$  is *monotonically increasing* in  $n$ .
- (b) By Bessel's inequality (page 218), the sequence is upper bounded by  $\|\mathbf{x}\|^2$ :

$$\sum_{n=1}^N |\langle \mathbf{x} | \mathbf{x}_n \rangle|^2 \leq \|\mathbf{x}\|^2$$

- (c) Because this sequence is both monotonically increasing and bounded in  $n$ , it must equal its bound in the limit as  $n$  approaches infinity:

$$\lim_{N \rightarrow \infty} \sum_{n=1}^N |\langle \mathbf{x} | \mathbf{x}_n \rangle|^2 = \|\mathbf{x}\|^2 \tag{H.1}$$

- (d) If we combine this result with *Bessel's Equality* (Theorem H.6 page 217) we have

$$\begin{aligned} \lim_{N \rightarrow \infty} \left\| \mathbf{x} - \sum_{n=1}^N \langle \mathbf{x} | \mathbf{x}_n \rangle \mathbf{x}_n \right\|^2 &= \|\mathbf{x}\|^2 - \lim_{N \rightarrow \infty} \sum_{n=1}^N |\langle \mathbf{x} | \mathbf{x}_n \rangle|^2 && \text{by Bessel's Equality (Theorem H.6 page 217)} \\ &= \|\mathbf{x}\|^2 - \|\mathbf{x}\|^2 && \text{by equation (H.1) page 223} \\ &= 0 \end{aligned}$$



**Proposition H.3** (Fourier expansion). Let  $\{x_n \in X | n \in \mathbb{N}\}$  be a set of vectors in a HILBERT SPACE  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle | \nabla \rangle)$ .

P R P	$\underbrace{\{x_n\} \text{ is an ORTHONORMAL BASIS for } H}_{(A)} \quad \Rightarrow \quad \underbrace{\left\{ x \doteq \sum_{n=1}^{\infty} \alpha_n x_n \quad \Leftrightarrow \quad \underbrace{\alpha_n = \langle x   x_n \rangle}_{(2)} \right\}}_{(I)}$
-------------	---

PROOF:

1. Proof that (1)  $\Rightarrow$  (2): by Lemma H.2 page 220

2. Proof that (1)  $\Leftarrow$  (2):

$$\begin{aligned}
 \left\| x - \sum_{n=1}^{\infty} \alpha_n x_n \right\|^2 &= \left\| x - \sum_{n=1}^{\infty} \langle x | x_n \rangle x_n \right\|^2 && \text{by right hypothesis} \\
 &= \|x\|^2 - \sum_{n=1}^{\infty} |\langle x | x_n \rangle|^2 && \text{by Bessel's equality} \quad (\text{Theorem H.6 page 217}) \\
 &= 0 && \text{by Parseval's Identity} \quad (\text{Theorem H.9 page 222}) \\
 \stackrel{\text{def}}{\Leftrightarrow} \quad x &\doteq \sum_{n=1}^{\infty} \langle x | x_n \rangle x_n && \text{by definition of strong convergence}
 \end{aligned}$$

$\Rightarrow$

**Proposition H.4** (Riesz-Fischer Theorem). <sup>26</sup> Let  $\{x_n \in X | n \in \mathbb{N}\}$  be a set of vectors in a HILBERT SPACE  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle | \nabla \rangle)$ .

P R P	$\underbrace{\{x_n\} \text{ is an ORTHONORMAL BASIS for } H}_{(A)} \quad \Rightarrow \quad \left\{ \underbrace{\sum_{n=1}^{\infty}  \alpha_n ^2 < \infty}_{(1)} \quad \Leftrightarrow \quad \underbrace{\exists x \in H \text{ such that } \alpha_n = \langle x   x_n \rangle}_{(2)} \right\}$
-------------	--

PROOF:

1. Proof that (1)  $\Rightarrow$  (2):

(a) If (1) is true, then let  $x \doteq \sum_{n \in \mathbb{N}} \alpha_n x_n$ .

(b) Then

$$\begin{aligned}
 \langle x | x_n \rangle &= \left\langle \sum_{m \in \mathbb{N}} \alpha_m x_m | x_n \right\rangle && \text{by definition of } x \\
 &= \sum_{m \in \mathbb{N}} \alpha_m \langle x_m | x_n \rangle && \text{by homogeneous property of } \langle \triangle | \nabla \rangle \quad (\text{Definition K.1 page 259}) \\
 &= \sum_{m \in \mathbb{N}} \alpha_m \bar{\delta}_{mn} && \text{by (A)} \\
 &= \sum_{m \in \mathbb{N}} \alpha_m && \text{by definition of } \bar{\delta} \quad (\text{Definition K.3 page 271})
 \end{aligned}$$

<sup>26</sup> Young (2001) page 6

2. Proof that (1)  $\iff$  (2):

$$\begin{aligned} \sum_{n \in \mathbb{N}} |\alpha_n|^2 &= \sum_{n \in \mathbb{N}} |\langle x | x_n \rangle|^2 && \text{by (2)} \\ &\leq \|x\|^2 && \text{by Bessel's Inequality} && \text{(Theorem H.7 page 218)} \\ &\leq \infty \end{aligned}$$



### Theorem H.10. <sup>27</sup>

**T H M** All SEPARABLE HILBERT SPACES are ISOMORPHIC. That is,

$$\left\{ \begin{array}{l} \mathbf{X} \text{ is a separable} \\ \text{Hilbert space} \\ \mathbf{Y} \text{ is a separable} \\ \text{Hilbert space} \end{array} \right. \text{and} \quad \Rightarrow \quad \left\{ \begin{array}{l} \text{there is a BIJECTIVE operator } \mathbf{M} \in \mathbf{Y}^{\mathbf{X}} \text{ such that} \\ (1). \quad \mathbf{y} = \mathbf{M}\mathbf{x} \quad \forall \mathbf{x} \in \mathbf{X}, \mathbf{y} \in \mathbf{Y} \quad \text{and} \\ (2). \quad \|\mathbf{M}\mathbf{x}\| = \|\mathbf{x}\| \quad \forall \mathbf{x} \in \mathbf{X} \quad \text{and} \\ (3). \quad \langle \mathbf{M}\mathbf{x} | \mathbf{M}\mathbf{y} \rangle = \langle \mathbf{x} | \mathbf{y} \rangle \quad \forall \mathbf{x} \in \mathbf{X}, \mathbf{y} \in \mathbf{Y} \end{array} \right\}$$

PROOF:

1. Let  $\mathbf{X} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle | \triangleright \rangle)$  be a *separable Hilbert space* with *orthonormal basis*  $\{x_n | n \in \mathbb{N}\}$ . Let  $\mathbf{Y} \triangleq (Y, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle | \triangleright \rangle)$  be a *separable Hilbert space* with *orthonormal basis*  $\{y_n | n \in \mathbb{N}\}$ .

2. Proof that there exists *bijection* operator  $\mathbf{M}$  and its inverse  $\mathbf{M}^{-1}$  between  $\{x_n\}$  and  $\{y_n\}$ :

- (a) Let  $\mathbf{M}$  be defined such that  $y_n \triangleq \mathbf{M}x_n$ .
- (b) Thus  $\mathbf{M}$  is a *bijection* between  $\{x_n\}$  and  $\{y_n\}$ .
- (c) Because  $\mathbf{M}$  is a *bijection* between  $\{x_n\}$  and  $\{y_n\}$ ,  $\mathbf{M}$  has an inverse operator  $\mathbf{M}^{-1}$  between  $\{x_n\}$  and  $\{y_n\}$  such that  $x_n = \mathbf{M}^{-1}y_n$ .

3. Proof that  $\mathbf{M}$  and  $\mathbf{M}^{-1}$  are *bijection* operators between  $\mathbf{X}$  and  $\mathbf{Y}$ :

- (a) Proof that  $\mathbf{M}$  maps  $\mathbf{X}$  into  $\mathbf{Y}$ :

$$\begin{aligned} \mathbf{x} \in \mathbf{X} &\iff \mathbf{x} \triangleq \sum_{n \in \mathbb{N}} \langle \mathbf{x} | x_n \rangle x_n && \text{by Fourier expansion} && \text{(Theorem H.9 page 222)} \\ &\implies \exists \mathbf{y} \in \mathbf{Y} \quad \text{such that} \quad \langle \mathbf{y} | y_n \rangle = \langle \mathbf{x} | x_n \rangle && \text{by Riesz-Fischer Thm.} && \text{(Proposition H.4 page 224)} \\ &\implies \\ \mathbf{y} &= \sum_{n \in \mathbb{N}} \langle \mathbf{y} | y_n \rangle y_n && \text{by Fourier expansion} && \text{(Theorem H.9 page 222)} \\ &= \sum_{n \in \mathbb{N}} \langle \mathbf{x} | x_n \rangle y_n && \text{by Riesz-Fischer Thm.} && \text{(Proposition H.4 page 224)} \\ &= \sum_{n \in \mathbb{N}} \langle \mathbf{x} | x_n \rangle \mathbf{M}x_n && \text{by definition of } \mathbf{M} && \text{(item (2a) page 225)} \\ &= \mathbf{M} \sum_{n \in \mathbb{N}} \langle \mathbf{x} | x_n \rangle x_n && \text{by prop. of linear ops.} && \text{(Theorem O.1 page 308)} \\ &= \mathbf{M}\mathbf{x} && \text{by definition of } \mathbf{x} && \end{aligned}$$

<sup>27</sup> Young (2001) page 6

(b) Proof that  $\mathbf{M}^{-1}$  maps  $\mathbf{Y}$  into  $\mathbf{X}$ :

$$\begin{aligned}
 y \in \mathbf{Y} &\iff y = \sum_{n \in \mathbb{N}} \langle y | y_n \rangle y_n && \text{by Fourier expansion (Theorem H.9 page 222)} \\
 &\implies \exists x \in \mathbf{X} \text{ such that } \langle x | x_n \rangle = \langle y | y_n \rangle \text{ by Riesz-Fischer Thm.} && \text{(Proposition H.4 page 224)} \\
 &\implies \\
 x &= \sum_{n \in \mathbb{N}} \langle x | x_n \rangle x_n && \text{by Fourier expansion (Theorem H.9 page 222)} \\
 &= \sum_{n \in \mathbb{N}} \langle y | y_n \rangle x_n && \text{by Riesz-Fischer Thm. (Proposition H.4 page 224)} \\
 &= \sum_{n \in \mathbb{N}} \langle y | y_n \rangle \mathbf{M}^{-1} y_n && \text{by definition of } \mathbf{M}^{-1} \text{ (item (2c) page 225)} \\
 &= \mathbf{M}^{-1} \sum_{n \in \mathbb{N}} \langle y | y_n \rangle y_n && \text{by prop. of linear ops. (Theorem O.1 page 308)} \\
 &= \mathbf{M}^{-1} y && \text{by definition of } y
 \end{aligned}$$

4. Proof for (2):

$$\begin{aligned}
 \|\mathbf{M}x\|^2 &= \left\| \mathbf{M} \sum_{n \in \mathbb{N}} \langle x | x_n \rangle x_n \right\|^2 && \text{by Fourier expansion (Theorem H.9 page 222)} \\
 &= \left\| \sum_{n \in \mathbb{N}} \langle x | x_n \rangle \mathbf{M}x_n \right\|^2 && \text{by property of linear operators (Theorem O.1 page 308)} \\
 &= \left\| \sum_{n \in \mathbb{N}} \langle x | x_n \rangle y_n \right\|^2 && \text{by definition of } \mathbf{M} \text{ (item (2a) page 225)} \\
 &= \sum_{n \in \mathbb{N}} |\langle x | x_n \rangle|^2 && \text{by Parseval's Identity (Proposition H.4 page 224)} \\
 &= \left\| \sum_{n \in \mathbb{N}} \langle x | x_n \rangle x_n \right\|^2 && \text{by Parseval's Identity (Proposition H.4 page 224)} \\
 &= \|x\|^2 && \text{by Fourier expansion (Theorem H.9 page 222)}
 \end{aligned}$$

5. Proof for (3): by (2) and Theorem O.23 page 328



**Theorem H.11.**<sup>28</sup> Let  $\mathbf{H}$  be a HILBERT SPACE.

T	H	M	$\mathbf{H}$ has a SCHAUDER BASIS	$\iff$	$\mathbf{H}$ is SEPARABLE
---	---	---	-----------------------------------	--------	---------------------------

**Theorem H.12.**<sup>29</sup> Let  $\mathbf{H}$  be a HILBERT SPACE.

T	H	M	$\mathbf{H}$ has an ORTHONORMAL BASIS	$\iff$	$\mathbf{H}$ is SEPARABLE
---	---	---	---------------------------------------	--------	---------------------------

<sup>28</sup> Bachman et al. (2000) page 112 (3.4.8), Berberian (1961) page 53 (Theorem II§8.3)

<sup>29</sup> Kubrusly (2001) page 357 (Proposition 5.43)

## H.6 Riesz bases in Hilbert spaces

**Definition H.11.** <sup>30</sup> 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 H.8 page 215) to some ORTHONORMAL BASIS (Definition H.9 page 219) in  $H$ .

**Definition H.12.** <sup>31</sup> 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

$$\exists A, B \in \mathbb{R}^+ \text{ such that } A \sum_{n=1}^{\infty} |\alpha_n|^2 \leq \left\| \sum_{n=1}^{\infty} \alpha_n x_n \right\|^2 \leq B \sum_{n=1}^{\infty} |\alpha_n|^2 \quad \forall (\alpha_n) \in \ell_{\mathbb{F}}^2.$$

**Definition H.13.** Let  $X \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle | \nabla \rangle)$  be an INNER PRODUCT SPACE (Definition K.1 page 259).

**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  $\langle x_n | y_m \rangle = \delta_{nm}$

**Lemma H.3.** <sup>32</sup> 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**  $\left\{ \begin{array}{l} (i). \quad \{x_n\} \text{ is TOTAL in } X \\ (ii). \quad \text{There exists } A > 0 \text{ such that } A \sum_{n \in C} |\alpha_n|^2 \leq \left\| \sum_{n \in C} \alpha_n x_n \right\|^2 \text{ for finite } C \\ (iii). \quad \text{There exists } B > 0 \text{ such that } \left\| \sum_{n=1}^{\infty} b_n y_n \right\|^2 \leq B \sum_{n=1}^{\infty} |b_n|^2 \quad \forall (b_n)_{n \in \mathbb{N}} \in \ell_{\mathbb{F}}^2 \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} (1). \quad \mathbf{R}^\circ \text{ is a linear bounded operator that maps from } \text{span}\{x_n\} \text{ to } \text{span}\{y_n\} \\ \text{where } \mathbf{R}^\circ \sum_{n \in C} c_n x_n \triangleq \sum_{n \in C} c_n y_n, \text{ for some sequence } (c_n) \text{ and finite set } C \\ (2). \quad \mathbf{R} \text{ has a unique extension to a bounded operator } \mathbf{R} \text{ that maps from } X \text{ to } Y \\ (3). \quad \|\mathbf{R}^\circ\| \leq \frac{B}{A} \\ (4). \quad \|\mathbf{R}\| \leq \frac{B}{A} \end{array} \right\}$

**Theorem H.13.** <sup>33</sup> 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**  $\left\{ \begin{array}{l} \{x_n\} \text{ is a RIESZ BASIS} \\ \text{for } H \end{array} \right\} \iff \left\{ \begin{array}{l} (1). \quad \{x_n\} \text{ is TOTAL in } H \\ (2). \quad \exists A, B \in \mathbb{R}^+ \text{ such that } A \sum_{n=1}^{\infty} |\alpha_n|^2 \leq \left\| \sum_{n=1}^{\infty} \alpha_n x_n \right\|^2 \leq B \sum_{n=1}^{\infty} |\alpha_n|^2 \end{array} \right\}$

PROOF:

<sup>30</sup> Young (2001) page 27 (Definition 1.8.2), Christensen (2003) page 63 (Definition 3.6.1), Heil (2011) page 196 (Definition 7.9)

<sup>31</sup> Christensen (2003) pages 66–68 (page 68 and (3.24) on page 66), Wojtaszczyk (1997) page 20 (Definition 2.6)

<sup>32</sup> Christensen (2003) pages 65–66 (Lemma 3.6.5)

<sup>33</sup> 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 ( $\implies$ ) case:(a) Proof that *Riesz basis* hypothesis  $\implies$  (1): all bases for  $H$  are *total* in  $H$ .(b) Proof that *Riesz basis* hypothesis  $\implies$  (2):i. Let  $(\mathbf{u}_n)_{n \in \mathbb{N}}$  be an *orthonormal basis* for  $H$ .ii. Let  $\mathbf{R}$  be a *bounded bijective* operator such that  $\mathbf{x}_n = \mathbf{R}\mathbf{u}_n$ .iii. Proof for upper bound  $B$ :

$$\begin{aligned}
 \left\| \sum_{n=1}^{\infty} \alpha_n \mathbf{x}_n \right\|^2 &= \left\| \sum_{n=1}^{\infty} \alpha_n \mathbf{R}\mathbf{u}_n \right\|^2 && \text{by definition of } \mathbf{R} && \text{(item (1(b)ii))} \\
 &= \left\| \mathbf{R} \sum_{n=1}^{\infty} \alpha_n \mathbf{u}_n \right\|^2 && \text{by Theorem O.1 page 308} \\
 &\leq \|\mathbf{R}\|^2 \left\| \sum_{n=1}^{\infty} \alpha_n \mathbf{u}_n \right\|^2 && \text{by Theorem O.6 page 314} \\
 &= \|\mathbf{R}\|^2 \sum_{n=1}^{\infty} \|\alpha_n \mathbf{u}_n\|^2 && \text{by Pythagorean Theorem} && \text{(Theorem K.10 page 273)} \\
 &= \|\mathbf{R}\|^2 \sum_{n=1}^{\infty} |\alpha_n|^2 \|\mathbf{u}_n\|^2 && \text{by homogeneous property of norms} && \text{(Definition L.1 page 275)} \\
 &= \underbrace{\|\mathbf{R}\|^2}_{B} \sum_{n=1}^{\infty} |\alpha_n|^2 && \text{by definition of orthonormality} && \text{(Definition K.4 page 271)}
 \end{aligned}$$

iv. Proof for lower bound  $A$ :

$$\begin{aligned}
 \left\| \sum_{n=1}^{\infty} \alpha_n \mathbf{x}_n \right\|^2 &= \frac{\|\mathbf{R}^{-1}\|^2}{\|\mathbf{R}^{-1}\|^2} \left\| \sum_{n=1}^{\infty} \alpha_n \mathbf{x}_n \right\|^2 && \text{because } \|\mathbf{R}^{-1}\| > 0 && \text{(Proposition O.1 page 312)} \\
 &\geq \frac{1}{\|\mathbf{R}^{-1}\|^2} \left\| \mathbf{R}^{-1} \sum_{n=1}^{\infty} \alpha_n \mathbf{x}_n \right\|^2 && \text{by Theorem O.6 page 314} \\
 &= \frac{1}{\|\mathbf{R}^{-1}\|^2} \left\| \mathbf{R}^{-1} \sum_{n=1}^{\infty} \alpha_n \mathbf{R}\mathbf{u}_n \right\|^2 && \text{by definition of } \mathbf{R} && \text{(item (1(b)ii) page 228)} \\
 &= \frac{1}{\|\mathbf{R}^{-1}\|^2} \left\| \mathbf{R}^{-1} \mathbf{R} \sum_{n=1}^{\infty} \alpha_n \mathbf{u}_n \right\|^2 && \text{by property of linear operators} && \text{(Theorem O.1 page 308)} \\
 &= \frac{1}{\|\mathbf{R}^{-1}\|^2} \left\| \sum_{n=1}^{\infty} \alpha_n \mathbf{u}_n \right\|^2 && \text{by definition of inverse op.} && \text{(Definition O.2 page 307)} \\
 &= \frac{1}{\|\mathbf{R}^{-1}\|^2} \sum_{n=1}^{\infty} \|\alpha_n \mathbf{u}_n\|^2 && \text{by Pythagorean Theorem} && \text{(Theorem K.10 page 273)} \\
 &= \frac{1}{\|\mathbf{R}^{-1}\|^2} \sum_{n=1}^{\infty} |\alpha_n|^2 \|\mathbf{u}_n\|^2 && \text{by } \|\cdot\| \text{ homogeneous prop.} && \text{(Definition L.1 page 275)} \\
 &= \underbrace{\frac{1}{\|\mathbf{R}^{-1}\|^2} \sum_{n=1}^{\infty} |\alpha_n|^2}_{A} && \text{by def. of orthonormality} && \text{(Definition K.4 page 271)}
 \end{aligned}$$

2. Proof for ( $\implies$ ) case:

- (a) Let  $\{u_n\}_{n \in \mathbb{N}}$  be an *orthonormal basis* for  $H$ .
- (b) Using (2) and Lemma H.3 (page 227), construct an bounded extension operator  $R$  such that  $Ru_n = x_n$  for all  $n \in \mathbb{N}$ .
- (c) Using (2) and Lemma H.3 (page 227), construct an bounded extension operator  $S$  such that  $Sx_n = u_n$  for all  $n \in \mathbb{N}$ .
- (d) Then,  $RVx = VRx \implies V = R^{-1}$ , and so  $R$  is a bounded invertible operator
- (e) and  $\{x_n\}$  is a *Riesz sequence*.



**Theorem H.14.** <sup>34</sup> Let  $H \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \Delta | \nabla \rangle)$  be a SEPARABLE HILBERT SPACE.

<b>T H M</b>	$\left\{ \begin{array}{l} (x_n \in H)_{n \in \mathbb{Z}} \text{ is a} \\ \text{RIESZ BASIS for } H \end{array} \right\} \implies \left\{ \begin{array}{l} \text{There exists } (y_n \in H)_{n \in \mathbb{Z}} \text{ such that} \\ (1). \quad (x_n) \text{ and } (y_n) \text{ are BIORTHOGONAL and} \\ (2). \quad (y_n) \text{ is also a RIESZ BASIS for } H \text{ and} \\ (3). \quad \exists B > A > 0 \text{ such that} \\ A \sum_{n=1}^{\infty}  a_n ^2 \leq \left\  \sum_{n=1}^{\infty} a_n x_n \right\ ^2 = \left\  \sum_{n=1}^{\infty} a_n y_n \right\ ^2 \leq B \sum_{n=1}^{\infty}  a_n ^2 \\ \forall (a_n)_{n \in \mathbb{N}} \in \ell_{\mathbb{F}}^2 \end{array} \right\}$
----------------------	---

PROOF:

1. Proof for (1):

- (a) Let  $e_n$  be the *unit vector* in  $H$  such that the  $n$ th element of  $e_n$  is 1 and all other elements are 0.
- (b) Let  $M$  be an operator on  $H$  such that  $Me_n = x_n$ .
- (c) Note that  $M$  is *isometric*, and as such  $\|Mx\| = \|x\| \quad \forall x \in H$ .
- (d) Let  $y_n \triangleq (M^{-1})^*$ .
- (e) Then,

$$\begin{aligned}
 \langle y_n | x_m \rangle &= \left\langle (M^{-1})^* e_n | M e_m \right\rangle \\
 &= \langle e_n | M^{-1} M e_m \rangle \\
 &= \langle e_n | e_m \rangle \\
 &= \bar{\delta}_{nm} \\
 \implies \{x_n\} \text{ and } \{y_n\} \text{ are biorthogonal} &\quad \text{by Definition K.4 page 271}
 \end{aligned}$$

<sup>34</sup> Wojtaszczyk (1997) page 20 (Lemma 2.7(a))

2. Proof for (3):

$$\begin{aligned}
 \left\| \sum_{n \in \mathbb{Z}} \alpha_n y_n \right\| &= \left\| \sum_{n \in \mathbb{Z}} \alpha_n (\mathbf{M}^{-1})^* e_n \right\| && \text{by definition of } y_n && \text{(Proposition 1d page 229)} \\
 &= \left\| (\mathbf{M}^{-1})^* \sum_{n \in \mathbb{Z}} \alpha_n e_n \right\| && \text{by property of linear ops.} \\
 &= \left\| \sum_{n \in \mathbb{Z}} \alpha_n e_n \right\| && \text{because } (\mathbf{M}^{-1})^* \text{ is isometric} && \text{(Definition O.10 page 327)} \\
 &= \left\| \mathbf{M} \sum_{n \in \mathbb{Z}} \alpha_n e_n \right\| && \text{because } \mathbf{M} \text{ is isometric} && \text{(Definition O.10 page 327)} \\
 &= \left\| \sum_{n \in \mathbb{Z}} \alpha_n \mathbf{M} e_n \right\| && \text{by property of linear operators} \\
 &= \left\| \sum_{n \in \mathbb{Z}} \alpha_n x_n \right\| && \text{by definition of } \mathbf{M} \\
 \implies \{y_n\} &\text{ is a Riesz basis} && \text{by left hypothesis}
 \end{aligned}$$

3. Proof for (2): by (3) and definition of *Riesz basis* (Definition H.11 page 227)



**Proposition H.5.**<sup>35</sup> Let  $\{x_n | n \in \mathbb{N}\}$  be a set of vectors in a HILBERT SPACE  $\mathbf{H} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle | \triangleright \rangle)$ .

P R P	$  \left\{  \begin{array}{l}  \{x_n\} \text{ is a RIESZ BASIS for } \mathbf{H} \text{ with} \\  A \sum_{n=1}^{\infty}  a_n ^2 \leq \left\  \sum_{n=1}^{\infty} a_n x_n \right\ ^2 \leq B \sum_{n=1}^{\infty}  a_n ^2 \\  \forall \{a_n\} \in \ell_{\mathbb{F}}^2  \end{array}  \right\} \Rightarrow \left\{  \begin{array}{l}  \{x_n\} \text{ is a FRAME for } \mathbf{H} \text{ with} \\  \underbrace{\frac{1}{B} \ x\ ^2 \leq \sum_{n=1}^{\infty}  \langle x   x_n \rangle ^2 \leq \frac{1}{A} \ x\ ^2}_{\text{STABILITY CONDITION}} \\  \forall x \in \mathbf{H}  \end{array}  \right\}  $
-------------	---

PROOF:

1. Let  $\{y_n | n \in \mathbb{N}\}$  be a *Riesz basis* that is *biorthonormal* to  $\{x_n | n \in \mathbb{N}\}$  (Theorem H.14 page 229).

2. Let  $x \triangleq \sum_{n=1}^{\infty} a_n y_n$ .

3. lemma:

$$\begin{aligned}
 \sum_{n=1}^{\infty} |\langle x | x_n \rangle|^2 &= \sum_{n=1}^{\infty} \left| \left\langle \sum_{m=1}^{\infty} a_m y_m | x_n \right\rangle \right|^2 && \text{by definition of } x && \text{(item (2) page 230)} \\
 &= \sum_{n=1}^{\infty} \left| \sum_{m=1}^{\infty} a_m \langle y_m | x_n \rangle \right|^2 && \text{by homogeneous property of } \langle \triangle | \triangleright \rangle && \text{(Definition K.1 page 259)} \\
 &= \sum_{n=1}^{\infty} \left| \sum_{m=1}^{\infty} a_m \bar{\delta}_{mn} \right|^2 && \text{by definition of biorthonormal} && \text{(Definition H.13 page 227)} \\
 &= \sum_{n=1}^{\infty} |a_n|^2 && \text{by definition of } \bar{\delta} && \text{(Definition K.3 page 271)}
 \end{aligned}$$

<sup>35</sup> Igari (1996) page 220 (Lemma 9.8), Wojtaszczyk (1997) pages 20–21 (Lemma 2.7(a))

4. Then

$$\begin{aligned}
 A \sum_{n=1}^{\infty} |a_n|^2 &\leq \left\| \sum_{n=1}^{\infty} a_n \mathbf{x}_n \right\|^2 \leq B \sum_{n=1}^{\infty} |a_n|^2 && \text{by definition of } \{\mathbf{y}_n\} \text{ (item (1) page 230)} \\
 \implies A \sum_{n=1}^{\infty} |a_n|^2 &\leq \left\| \sum_{n=1}^{\infty} a_n \mathbf{y}_n \right\|^2 \leq B \sum_{n=1}^{\infty} |a_n|^2 && \text{by definition of } \{\mathbf{y}_n\} \text{ (item (1) page 230)} \\
 \implies A \sum_{n=1}^{\infty} |a_n|^2 &\leq \|\mathbf{x}\|^2 \leq B \sum_{n=1}^{\infty} |a_n|^2 && \text{by definition of } \mathbf{x} \text{ (item (2) page 230)} \\
 \implies A \sum_{n=1}^{\infty} |\langle \mathbf{x} | \mathbf{x}_n \rangle|^2 &\leq \|\mathbf{x}\|^2 \leq B \sum_{n=1}^{\infty} |\langle \mathbf{x} | \mathbf{x}_n \rangle|^2 && \text{by (3) lemma} \\
 \implies \frac{1}{B} \|\mathbf{x}\|^2 &\leq \sum_{n=1}^{\infty} |\langle \mathbf{x} | \mathbf{x}_n \rangle|^2 \leq \frac{1}{A} \|\mathbf{x}\|^2
 \end{aligned}$$

⇒

**Theorem H.15** (Battle-Lemarié orthogonalization). <sup>36</sup> Let  $\tilde{f}(\omega)$  be the FOURIER TRANSFORM (Definition P.2 page 337) of a function  $f \in L^2_{\mathbb{R}}$ .

T H M	$  \left\{  \begin{array}{l}  1. \quad \left\{ \mathbf{T}^n g \mid n \in \mathbb{Z} \right\} \text{ is a RIESZ BASIS for } L^2_{\mathbb{R}} \quad \text{and} \\  2. \quad \tilde{f}(\omega) \triangleq \frac{\tilde{g}(\omega)}{\sqrt{2\pi \sum_{n \in \mathbb{Z}}  \tilde{g}(\omega + 2\pi n) ^2}}  \end{array}  \right\} \Rightarrow \left\{ \begin{array}{l} \left\{ \mathbf{T}^n f \mid n \in \mathbb{Z} \right\} \\ \text{is an ORTHONORMAL BASIS for } L^2_{\mathbb{R}} \end{array} \right\}  $
-------------	---

PROOF:

1. Proof that  $\{\mathbf{T}^n f \mid n \in \mathbb{Z}\}$  is orthonormal:

$$\begin{aligned}
 \tilde{S}_{\phi\phi}(\omega) &= 2\pi \sum_{n \in \mathbb{Z}} |\tilde{f}(\omega + 2\pi n)|^2 && \text{by Theorem ?? page ??} \\
 &= 2\pi \sum_{n \in \mathbb{Z}} \left| \frac{\tilde{g}(\omega + 2\pi n)}{\sqrt{2\pi \sum_{m \in \mathbb{Z}} |\tilde{g}(\omega + 2\pi(m-n))|^2}} \right|^2 && \text{by left hypothesis} \\
 &= \sum_{n \in \mathbb{Z}} \left| \frac{\tilde{g}(\omega + 2\pi n)}{\sqrt{\sum_{m \in \mathbb{Z}} |\tilde{g}(\omega + 2\pi m)|^2}} \right|^2 \\
 &= \sum_{n \in \mathbb{Z}} \left| \frac{1}{\sqrt{\sum_{m \in \mathbb{Z}} |\tilde{g}(\omega + 2\pi m)|^2}} \right|^2 |\tilde{g}(\omega + 2\pi n)|^2 \\
 &= \frac{1}{\sum_{m \in \mathbb{Z}} |\tilde{g}(\omega + 2\pi m)|^2} \sum_{n \in \mathbb{Z}} |\tilde{g}(\omega + 2\pi n)|^2 \\
 &= 1 \\
 \implies \{f_n \mid n \in \mathbb{Z}\} &\text{ is orthonormal} && \text{by Theorem ?? page ??}
 \end{aligned}$$

<sup>36</sup> 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  $\{\mathbf{T}^n f | n \in \mathbb{Z}\}$  is a basis for  $V_0$ : by Lemma H.1 page 215.



## H.7 Frames in Hilbert spaces

**Definition H.14.** <sup>37</sup> Let  $\{x_n \in X | n \in \mathbb{N}\}$  be a set of vectors in a HILBERT SPACE  $H \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \Delta | \nabla \rangle)$ .

The set  $\{x_n\}$  is a **frame** for  $H$  if (STABILITY CONDITION)

$$\exists A, B \in \mathbb{R}^+ \text{ such that } A \|x\|^2 \leq \sum_{n=1}^{\infty} |\langle x | x_n \rangle|^2 \leq B \|x\|^2 \quad \forall x \in X.$$

The quantities  $A$  and  $B$  are **frame bounds**.

The quantity  $A'$  is the **optimal lower frame bound** if

$$A' = \sup \{A \in \mathbb{R}^+ | A \text{ is a lower frame bound}\}.$$

The quantity  $B'$  is the **optimal upper frame bound** if

$$B' = \inf \{B \in \mathbb{R}^+ | B \text{ is an upper frame bound}\}.$$

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  $\{x_n | n \in \mathbb{N}\}$  is an **exact frame** if for some  $m \in \mathbb{Z}$ ,  $\{x_n | n \in \mathbb{N}\} \setminus \{x_m\}$  is NOT a frame.

A frame is a *Parseval frame* (Definition H.14) if it satisfies *Parseval's Identity* (Theorem H.9 page 222). All orthonormal bases are Parseval frames (Theorem H.9 page 222); but not all Parseval frames are orthonormal bases.

**Definition H.15.** Let  $\{x_n\}$  be a **frame** (Definition H.14 page 232) for the HILBERT SPACE  $H \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \Delta | \nabla \rangle)$ . Let  $S$  be an OPERATOR on  $H$ .

**D E F**  $S$  is a **frame operator** for  $\{x_n\}$  if  $Sf(x) = \sum_{n \in \mathbb{Z}} \langle f | x_n \rangle x_n \quad \forall f \in H$ .

**Theorem H.16.** <sup>38</sup> Let  $S$  be a FRAME OPERATOR (Definition H.15 page 232) of a FRAME  $\{x_n\}$  (Definition H.14 page 232) for the HILBERT SPACE  $H \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \Delta | \nabla \rangle)$ .

**T H M**

- (1).  $S$  is INVERTIBLE.
- (2).  $f(x) = \sum_{n \in \mathbb{Z}} \langle f | S^{-1} x_n \rangle x_n = \sum_{n \in \mathbb{Z}} \langle f | x_n \rangle S^{-1} x_n \quad \forall f \in H$

**Theorem H.17.** <sup>39</sup> Let  $\{x_n \in X | n=1,2,\dots,N\}$  be a set of vectors in a HILBERT SPACE  $H \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \Delta | \nabla \rangle)$ .

**T H M**  $\{x_n\}$  is a FRAME for  $\text{span}\{x_n\}$ .

PROOF:

<sup>37</sup> 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

<sup>38</sup> Christensen (2008) pages 100–102 (Theorem 5.1.7)

<sup>39</sup> Christensen (2003) page 3

1. Upper bound: Proof that there exists  $B$  such that  $\sum_{n=1}^N |\langle \mathbf{x} | \mathbf{x}_n \rangle|^2 \leq B \|\mathbf{x}\|^2 \quad \forall \mathbf{x} \in \mathcal{H}$ :

$$\begin{aligned} \sum_{n=1}^N |\langle \mathbf{x} | \mathbf{x}_n \rangle|^2 &\leq \sum_{n=1}^N \langle \mathbf{x}_n | \mathbf{x}_n \rangle \langle \mathbf{x} | \mathbf{x} \rangle \quad \text{by Cauchy-Schwarz inequality (Theorem K.2 page 260)} \\ &= \underbrace{\left\{ \sum_{n=1}^N \|\mathbf{x}_n\|^2 \right\}}_B \|\mathbf{x}\|^2 \end{aligned}$$

2. Lower bound: Proof that there exists  $A$  such that  $A \|\mathbf{x}\|^2 \leq \sum_{n=1}^N |\langle \mathbf{x} | \mathbf{x}_n \rangle|^2 \quad \forall \mathbf{x} \in \mathcal{H}$ :

$$\begin{aligned} \sum_{n=1}^N |\langle \mathbf{x} | \mathbf{x}_n \rangle|^2 &= \sum_{n=1}^N \left| \left\langle \mathbf{x}_n | \frac{\mathbf{x}}{\|\mathbf{x}\|} \right\rangle \right|^2 \|\mathbf{x}\|^2 \\ &\geq \underbrace{\left( \inf_y \left\{ \sum_{n=1}^N |\langle \mathbf{x}_n | y \rangle|^2 \mid \|y\| = 1 \right\} \right)}_A \|\mathbf{x}\|^2 \end{aligned}$$

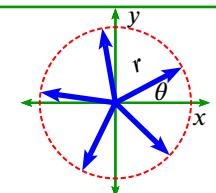
*Example H.1.* Let  $(\mathbb{R}^2, +, \cdot, (\mathbb{R}, \dot{+}, \dot{\times}), \langle \triangle | \nabla \rangle)$  be an inner product space with  $\left\langle \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} | \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} \right\rangle \triangleq x_1 x_2 + y_1 y_2$ . Let  $\mathbf{S}$  be the *frame operator* (Definition H.15 page 232) with *inverse*  $\mathbf{S}^{-1}$ .

EX

Let  $N \in \{3, 4, 5, \dots\}$ ,  $\theta \in \mathbb{R}$ , and  $r \in \mathbb{R}^+$  ( $r > 0$ ).

Let  $\mathbf{x}_n \triangleq r \begin{bmatrix} \cos(\theta + 2n\pi/N) \\ \sin(\theta + 2n\pi/N) \end{bmatrix} \quad \forall n \in \{0, 1, \dots, N-1\}$ .

Then,  $(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{N-1})$  is a **tight frame** for  $\mathbb{R}^2$  with *frame bound*  $A = \frac{Nr^2}{2}$ .



Moreover,  $\mathbf{S} = \frac{Nr^2}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ ,  $\mathbf{S}^{-1} = \frac{2}{Nr^2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ , and  $\mathbf{v} = \frac{2}{Nr^2} \sum_{n=0}^{N-1} \langle \mathbf{v} | \mathbf{x}_n \rangle \mathbf{x}_n \quad \forall \mathbf{v} \in \mathbb{R}^2$ .

PROOF:

1. Proof that  $(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{N-1})$  is a *tight frame* with *frame bound*  $A = \frac{Nr^2}{2}$ : Let  $\mathbf{v} \triangleq (x, y) \in \mathbb{R}^2$ .

$$\begin{aligned} \sum_{n=0}^{N-1} |\langle \mathbf{v} | \mathbf{x}_n \rangle|^2 &\triangleq \sum_{n=0}^{N-1} \left| \mathbf{v}^H \begin{bmatrix} \cos\left(\theta + \frac{2n\pi}{N}\right) \\ \sin\left(\theta + \frac{2n\pi}{N}\right) \end{bmatrix} \right|^2 && \text{by definitions of } \mathbf{v} \text{ of } \langle \mathbf{y} | \mathbf{x} \rangle \\ &\triangleq \sum_{n=0}^{N-1} r^2 \left| x \cos\left(\theta + \frac{2n\pi}{N}\right) + y \sin\left(\theta + \frac{2n\pi}{N}\right) \right|^2 && \text{by definition of } \mathbf{y}^H \mathbf{x} \text{ operation} \\ &= r^2 x^2 \sum_{n=0}^{N-1} \cos^2\left(\theta + \frac{2n\pi}{N}\right) + r^2 y^2 \sum_{n=0}^{N-1} \sin^2\left(\theta + \frac{2n\pi}{N}\right) + r^2 xy \sum_{n=0}^{N-1} \cos\left(\theta + \frac{2n\pi}{N}\right) \sin\left(\theta + \frac{2n\pi}{N}\right) \\ &= r^2 x^2 \frac{N}{2} + r^2 y^2 \frac{N}{2} + r^2 xy 0 && \text{by Corollary ?? page ??} \\ &= (x^2 + y^2) \frac{Nr^2}{2} = \underbrace{\left( \frac{Nr^2}{2} \right)}_A \mathbf{v}^H \mathbf{v} \triangleq \left( \frac{Nr^2}{2} \right) \|\mathbf{v}\|^2 && \text{by definition of } \|\mathbf{v}\| \end{aligned}$$

2. Proof that  $\mathbf{S} = \frac{Nr^2}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ :

(a) Let  $e_1 \triangleq \begin{bmatrix} 1 \\ 0 \end{bmatrix}$  and  $e_2 \triangleq \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ .

(b) lemma:  $\mathbf{S}e_1 = \frac{Nr^2}{2}e_1$ . Proof:

$$\begin{aligned}\mathbf{S}e_1 &= \sum_{n=0}^{N-1} \langle e_1 | \mathbf{x}_n \rangle \mathbf{x}_n \\ &= \sum_{n=0}^{N-1} r \cos\left(\theta + \frac{2n\pi}{N}\right) r \begin{bmatrix} \cos\left(\theta + \frac{2n\pi}{N}\right) \\ \sin\left(\theta + \frac{2n\pi}{N}\right) \end{bmatrix} = r^2 \sum_{n=0}^{N-1} \begin{bmatrix} \cos^2\left(\theta + \frac{2n\pi}{N}\right) \\ \cos\left(\theta + \frac{2n\pi}{N}\right) \sin\left(\theta + \frac{2n\pi}{N}\right) \end{bmatrix} \\ &= r^2 \begin{bmatrix} N/2 \\ 0 \end{bmatrix} = \frac{Nr^2}{2}e_1 \quad \text{by Summation around unit circle (Corollary ?? page ??)}$$

(c) lemma:  $\mathbf{S}e_2 = \frac{Nr^2}{2}e_2$ . Proof:

$$\begin{aligned}\mathbf{S}e_2 &= \sum_{n=0}^{N-1} \langle e_2 | \mathbf{x}_n \rangle \mathbf{x}_n \\ &= \sum_{n=0}^{N-1} r \sin\left(\theta + \frac{2n\pi}{N}\right) r \begin{bmatrix} \cos\left(\theta + \frac{2n\pi}{N}\right) \\ \sin\left(\theta + \frac{2n\pi}{N}\right) \end{bmatrix} = r^2 \sum_{n=0}^{N-1} \begin{bmatrix} \sin\left(\theta + \frac{2n\pi}{N}\right) \cos\left(\theta + \frac{2n\pi}{N}\right) \\ \sin^2\left(\theta + \frac{2n\pi}{N}\right) \end{bmatrix} \\ &= r^2 \begin{bmatrix} 0 \\ N/2 \end{bmatrix} = \frac{Nr^2}{2}e_2 \quad \text{by Summation around unit circle (Corollary ?? page ??)}$$

(d) Complete the proof of item (2) using Eigendecomposition  $\mathbf{S} = \mathbf{Q}\Lambda\mathbf{Q}^{-1}$ :

$$\mathbf{S}e_1 = \frac{Nr^2}{2}e_1 \quad \text{by (2c) lemma}$$

$\Rightarrow e_1$  is an eigenvector of  $\mathbf{S}$  with eigenvalue  $\frac{Nr^2}{2}$

$$\mathbf{S}e_2 = \frac{Nr^2}{2}e_2 \quad \text{by (2c) lemma}$$

$\Rightarrow e_2$  is an eigenvector of  $\mathbf{S}$  with eigenvalue  $\frac{Nr^2}{2}$

$$\underbrace{\mathbf{S} = \underbrace{\begin{bmatrix} 1 & 1 \\ e_1 & e_2 \\ 1 & 1 \end{bmatrix}}_{\mathbf{Q}} \underbrace{\begin{bmatrix} \frac{Nr^2}{2} & 0 \\ 0 & \frac{Nr^2}{2} \end{bmatrix}}_{\Lambda} \underbrace{\begin{bmatrix} 1 & 1 \\ e_1 & e_2 \\ 1 & 1 \end{bmatrix}}_{\mathbf{Q}^{-1}}^{-1}}_{\text{Eigendecomposition of } \mathbf{S}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{Nr^2}{2} & 0 \\ 0 & \frac{Nr^2}{2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^{-1} = \frac{Nr^2}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

3. Proof that  $\mathbf{S}^{-1} = \frac{2}{Nr^2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ :

$$\mathbf{S}\mathbf{S}^{-1} = \frac{Nr^2}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \frac{2}{Nr^2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \triangleq \mathbf{I}_2 \quad \text{by item (2)}$$

$$\mathbf{S}^{-1}\mathbf{S} = \frac{2}{Nr^2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \frac{Nr^2}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \triangleq \mathbf{I}_2 \quad \text{by item (2)}$$

4. Proof that  $\mathbf{v} = \frac{2}{Nr^2} \sum_{n=0}^{N-1} (\mathbf{v}^H \mathbf{x}_n) \mathbf{x}_n$ :

$$\mathbf{v} = \sum_{n=0}^{N-1} \langle \mathbf{v} | \mathbf{S}^{-1} \mathbf{x}_n \rangle \mathbf{x}_n = \sum_{n=0}^{N-1} \left\langle \mathbf{v} \mid \frac{2}{Nr^2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{x}_n \right\rangle \mathbf{x}_n \quad \text{by item (3)}$$

$$= \frac{2}{Nr^2} \sum_{n=0}^{N-1} \langle \mathbf{v} | \mathbf{x}_n \rangle \mathbf{x}_n = \frac{2}{Nr^2} \sum_{n=0}^{N-1} (\mathbf{v}^H \mathbf{x}_n) \mathbf{x}_n \quad \text{by definition of } \langle \mathbf{y} | \mathbf{x} \rangle$$



*Example H.2 (Peace Frame/Mercedes Frame).* <sup>40</sup> Let  $(\mathbb{R}^2, +, \cdot, (\mathbb{R}, \dot{+}, \dot{\times}), \langle \triangle | \nabla \rangle)$  be an inner product space with  $\langle (x_1, y_1) | (x_2, y_2) \rangle \triangleq x_1y_1 + x_2y_2$ . Let  $\mathbf{S}$  be the *frame operator* (Definition H.15 page 232) with inverse  $\mathbf{S}^{-1}$ .

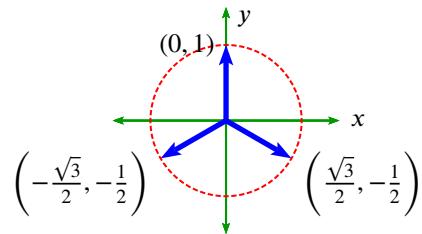
**E  
X**

Let  $\mathbf{x}_1 \triangleq \frac{1}{2} \begin{bmatrix} 0 \\ 2 \end{bmatrix}$ ,  $\mathbf{x}_2 \triangleq \frac{1}{2} \begin{bmatrix} -\sqrt{3} \\ -1 \end{bmatrix}$ , and  $\mathbf{x}_3 \triangleq \frac{1}{2} \begin{bmatrix} \sqrt{3} \\ -1 \end{bmatrix}$ .

Then,  $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$  is a **tight frame** for  $\mathbb{R}^2$  with *frame bound*  $A = \frac{3}{2}$ .

Moreover,  $\mathbf{S} = \frac{3}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ ,  $\mathbf{S}^{-1} = \frac{2}{3} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ ,

and  $\mathbf{v} = \frac{2}{3} \sum_{n=1}^3 \langle \mathbf{v} | \mathbf{x}_n \rangle \mathbf{x}_n \triangleq \frac{2}{3} \sum_{n=1}^3 (\mathbf{v}^H \mathbf{x}_n) \mathbf{x}_n \quad \forall \mathbf{v} \in \mathbb{R}^2$ .



PROOF:

1. This frame is simply a special case of the frame presented in Example H.1 (page 233) with  $r = 1$ ,  $N = 3$ , and  $\theta = \pi/2$ .

2. Let's give it a try! Let  $\mathbf{v} \triangleq \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ .

$$\begin{aligned} \sum_{n=1}^3 \langle \mathbf{v} | \mathbf{S}^{-1} \mathbf{x}_n \rangle \mathbf{x}_n &= \frac{2}{3} \sum_{n=1}^3 (\mathbf{v}^H \mathbf{x}_n) \mathbf{x}_n && \text{by Example H.1 page 233} \\ &= (\mathbf{v}^H \mathbf{x}_1) \mathbf{x}_1 + (\mathbf{v}^H \mathbf{x}_2) \mathbf{x}_2 + (\mathbf{v}^H \mathbf{x}_3) \mathbf{x}_3 \\ &= \frac{2}{3} \left( \left( \mathbf{v}^H \frac{1}{2} \begin{bmatrix} 0 \\ 2 \end{bmatrix} \right) \mathbf{x}_1 + \left( \mathbf{v}^H \frac{1}{2} \begin{bmatrix} -\sqrt{3} \\ -1 \end{bmatrix} \right) \mathbf{x}_2 + \left( \mathbf{v}^H \frac{1}{2} \begin{bmatrix} \sqrt{3} \\ -1 \end{bmatrix} \right) \mathbf{x}_3 \right) \\ &= \frac{2}{3} \cdot \frac{1}{2} \left( \left( \mathbf{v}^H \begin{bmatrix} 0 \\ 2 \end{bmatrix} \right) \mathbf{x}_1 + \left( \mathbf{v}^H \begin{bmatrix} -\sqrt{3} \\ -1 \end{bmatrix} \right) \mathbf{x}_2 + \left( \mathbf{v}^H \begin{bmatrix} \sqrt{3} \\ -1 \end{bmatrix} \right) \mathbf{x}_3 \right) \\ &= \frac{1}{3} \left( (2) \frac{1}{2} \begin{bmatrix} 0 \\ 2 \end{bmatrix} + (-\sqrt{3}-1) \frac{1}{2} \begin{bmatrix} -\sqrt{3} \\ -1 \end{bmatrix} + (\sqrt{3}-1) \frac{1}{2} \begin{bmatrix} \sqrt{3} \\ -1 \end{bmatrix} \right) \\ &= \frac{1}{6} \left[ \begin{array}{lcl} 2(0) & + & (-\sqrt{3}-1)(-\sqrt{3}) & + & (\sqrt{3}-1)(\sqrt{3}) \\ 2(2) & + & (-\sqrt{3}-1)(-1) & + & (\sqrt{3}-1)(-1) \end{array} \right] \\ &= \frac{1}{6} \left[ \begin{array}{lcl} 0 & + & (3+\sqrt{3}) & + & (3-\sqrt{3}) \\ 4 & + & (1+\sqrt{3}) & + & (1-\sqrt{3}) \end{array} \right] = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \triangleq \mathbf{v} \end{aligned}$$



In Example H.1 (page 233) and Example H.2 (page 235), 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 H.3 (next).

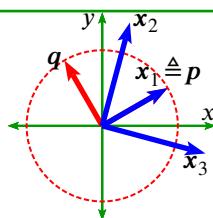
*Example H.3.* <sup>41</sup> Let  $(\mathbb{R}^2, +, \cdot, (\mathbb{R}, \dot{+}, \dot{\times}), \langle \triangle | \nabla \rangle)$  be an inner product space with  $\langle (x_1, y_1) | (x_2, y_2) \rangle \triangleq x_1y_1 + x_2y_2$ . Let  $\mathbf{S}$  be the *frame operator* (Definition H.15 page 232) with inverse  $\mathbf{S}^{-1}$ .

<sup>40</sup> 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)

<sup>41</sup> Christensen (2003) pages 7–8 (?)

**E  
X**

Let  $p$  and  $q$  be orthonormal vectors in  $\mathbf{X} \triangleq \text{span}\{p, q\}$ .  
 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  $\mathbf{X}$  with *frame bounds*  $A = 0$  and  $B = 5$ .



Moreover,  
 $S^{-1}x_1 = \frac{1}{3}p$  and  
 $S^{-1}x_2 = \frac{1}{3}p + \frac{1}{2}q$  and  
 $S^{-1}x_3 = \frac{1}{3}p - \frac{1}{2}q$ .

PROOF:

1. Proof that  $(x_1, x_2, x_3)$  is a *frame* with *frame bounds*  $A = 0$  and  $B = 5$ :

$$\begin{aligned} \sum_{n=1}^3 |\langle v | x_n \rangle|^2 &\triangleq |\langle v | p \rangle|^2 + |\langle v | p + q \rangle|^2 + |\langle v | p - q \rangle|^2 && \text{by definitions of } x_1, x_2, \text{ and } x_3 \\ &= |\langle v | p \rangle|^2 + |\langle v | p \rangle + \langle v | q \rangle|^2 + |\langle v | p \rangle - \langle v | q \rangle|^2 && \text{by additivity of } \langle \Delta | \nabla \rangle \text{ (Definition K.1 page 259)} \\ &= |\langle v | p \rangle|^2 + (|\langle v | p \rangle|^2 + |\langle v | q \rangle|^2 + \langle v | p \rangle \langle v | q \rangle^* + \langle v | q \rangle \langle v | p \rangle^*) \\ &\quad + (|\langle v | p \rangle|^2 + |\langle v | q \rangle|^2 - \langle v | p \rangle \langle v | q \rangle^* - \langle v | q \rangle \langle v | p \rangle^*) \\ &= 3|\langle v | p \rangle|^2 + 2|\langle v | q \rangle|^2 \\ &\leq 3\|v\| \|p\| + 2\|v\| \|q\| && \text{by CS Inequality (Theorem K.2 page 260)} \\ &= \|v\| (3\|p\| + 2\|q\|) \\ &= \boxed{5\|v\|} && \text{by orthonormality of } p \text{ and } q \end{aligned}$$

2. lemma:  $Sp = 3p$ ,  $Sq = 2q$ ,  $S^{-1}p = \frac{1}{3}p$ , and  $S^{-1}q = \frac{1}{2}q$ . Proof:

$$\begin{aligned} Sp &\triangleq \sum_{n=1}^3 \langle p | x_n \rangle x_n \\ &= \langle p | p \rangle p + \langle p | p + q \rangle (p + q) + \langle p | p - q \rangle (p - q) \\ &= (1)p + (1+0)(p+q) + (1-0)(p-q) \\ &= 3p \\ \implies S^{-1}p &= \frac{1}{3}p \\ Sq &\triangleq \sum_{n=1}^3 \langle q | x_n \rangle x_n \\ &= \langle q | p \rangle p + \langle q | p + q \rangle (p + q) + \langle q | p - q \rangle (p - q) \\ &= (0)q + (0+1)(p+q) + (0-1)(p-q) \\ &= 2q \\ \implies S^{-1}q &= \frac{1}{2}q \end{aligned}$$

3. Remark: Without knowing  $p$  and  $q$ , from (2) lemma it follows that it is not possible to compute  $S$  or  $S^{-1}$  explicitly.

4. Proof that  $S^{-1}x_1 = \frac{1}{3}p$ ,  $S^{-1}x_2 = \frac{1}{3}p + \frac{1}{2}q$  and  $S^{-1}x_3 = \frac{1}{3}p - \frac{1}{2}q$ :

$$\begin{aligned} S^{-1}x_1 &\triangleq S^{-1}p && \text{by definition of } x_1 \\ &= \frac{1}{3}p && \text{by (2) lemma} \\ S^{-1}x_2 &\triangleq S^{-1}(p + q) && \text{by definition of } x_2 \\ &= \frac{1}{3}p + \frac{1}{2}q && \text{by (2) lemma} \end{aligned}$$

$$\begin{aligned} \mathbf{S}^{-1}\mathbf{x}_3 &\triangleq \mathbf{S}^{-1}(\mathbf{p} - \mathbf{q}) \\ &= \frac{1}{3}\mathbf{p} - \frac{1}{2}\mathbf{q} \end{aligned}$$

by definition of  $\mathbf{x}_2$   
by (2) lemma

5. Check that  $\mathbf{v} = \sum_n \langle \mathbf{v} | \mathbf{x}_n \rangle \mathbf{x}_n = \langle \mathbf{v} | \mathbf{p} \rangle \mathbf{p} + \langle \mathbf{v} | \mathbf{q} \rangle \mathbf{q}$ :

$$\begin{aligned} \mathbf{v} &= \sum_{n=1}^3 \langle \mathbf{v} | \mathbf{S}^{-1}\mathbf{x}_n \rangle \mathbf{x}_n \\ &= \langle \mathbf{v} | \mathbf{S}^{-1}\mathbf{p} \rangle \mathbf{p} + \langle \mathbf{v} | \mathbf{S}^{-1}(\mathbf{p} + \mathbf{q}) \rangle (\mathbf{p} + \mathbf{q}) + \langle \mathbf{v} | \mathbf{S}^{-1}(\mathbf{p} - \mathbf{q}) \rangle (\mathbf{p} - \mathbf{q}) \\ &= \left\langle \mathbf{v} \mid \frac{1}{3}\mathbf{p} \right\rangle \mathbf{p} + \left\langle \mathbf{v} \mid \frac{1}{3}\mathbf{p} + \frac{1}{2}\mathbf{q} \right\rangle (\mathbf{p} + \mathbf{q}) + \left\langle \mathbf{v} \mid \frac{1}{3}\mathbf{p} - \frac{1}{2}\mathbf{q} \right\rangle (\mathbf{p} - \mathbf{q}) \\ &= \left( \frac{1}{3} + \frac{1}{3} + \frac{1}{3} \right) \langle \mathbf{v} | \mathbf{p} \rangle \mathbf{p} + \left( \frac{1}{3} - \frac{1}{3} \right) \langle \mathbf{v} | \mathbf{p} \rangle \mathbf{q} + \left( \frac{1}{2} - \frac{1}{2} \right) \langle \mathbf{v} | \mathbf{q} \rangle \mathbf{p} + \left( \frac{1}{2} + \frac{1}{2} \right) \langle \mathbf{v} | \mathbf{q} \rangle \mathbf{q} \\ &= \langle \mathbf{v} | \mathbf{p} \rangle \mathbf{p} + \langle \mathbf{v} | \mathbf{q} \rangle \mathbf{q} \end{aligned}$$





## I.1 Subspaces of a linear space

*Linear spaces* (Definition G.1 page 197) can be decomposed into a collection of *linear subspaces* (Definition I.1 page 240). Often such a collection along with an *order relation* forms a *lattice*.

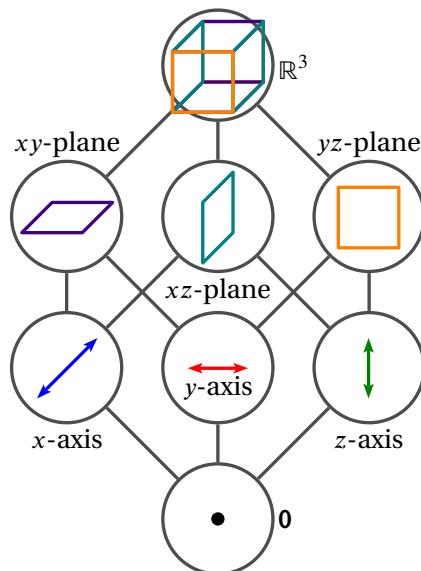
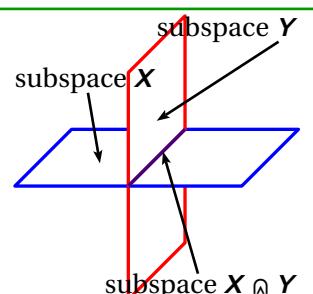


Figure I.1: lattice of subspaces of  $\mathbb{R}^3$  (Example I.1 page 239)

EX

*Example I.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 I.1 (page 239).



**Definition I.1.** <sup>1</sup> Let  $\Omega \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$  be a LINEAR SPACE (Definition G.1 page 197).

A ttuple  $(Y, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$  is a **linear subspace** of  $\Omega$  if

- |  |   |   |
|--|---|---|
| 1. $Y \neq \emptyset$                    | <i>(<math>Y</math> must contain at least one element)</i> | <i>and</i>  |
| 2. $Y \subseteq X$                       | <i>(<math>Y</math> is a subset of <math>X</math>)</i>     | <i>and</i>  |
| 3. $x, y \in Y$                          | $\implies x + y \in Y$                                    | <i>(closed under vector addition)</i>               |
| 4. $x \in Y$ and $\alpha \in \mathbb{F}$ | $\implies \alpha x \in Y$                                 | <i>(closed under scalar-vector multiplication).</i> |

A linear subspace is also called a **linear manifold**.

Every *linear space* (Definition G.1 page 197)  $X$  has at least two *linear subspaces*—itself and  $\mathbf{0}$  (Proposition I.1 page 240), called the *trivial linear space*. The *linear span* (Definition H.2 page 209) of every subset of a linear linear space is a subspace (Proposition I.2 page 241). Every *linear subspace* contains the “zero” vector  $\mathbf{0}$ , and is *convex* (Definition M.6 page 286, Proposition I.3 page 241).

**Proposition I.1.** <sup>2</sup> Let  $X \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$  and  $\mathbf{0} \triangleq (\{\mathbf{0}\}, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$ .

<b>P</b>	<b>R</b>	<b>P</b>	$\left\{ \begin{array}{l} X \text{ is a LINEAR SPACE} \\ (\text{Definition G.1 page 197}) \end{array} \right\} \implies \left\{ \begin{array}{l} 1. \mathbf{0} \text{ is a LINEAR SUBSPACE of } X \text{ and} \\ 2. X \text{ is a LINEAR SUBSPACE of } X \end{array} \right\}$
----------	----------	----------	--

PROOF: For a structure to be a linear subspace of  $X$ , it must satisfy the requirements of Definition I.1 (page 240).

1. Proof that  $\{\mathbf{0}\}$  is a linear subspace:

(a) Note that  $\{\mathbf{0}\} \neq \emptyset$ .

(b) Proof that  $x, y \in \{\mathbf{0}\} \implies x + y \in \{\mathbf{0}\}$ :

$$\begin{aligned} x + y &= \mathbf{0} + \mathbf{0} && \text{by } x, y \in \{\mathbf{0}\} \text{ hypothesis} \\ &= \mathbf{0} \\ &\in \{\mathbf{0}\} \end{aligned}$$

(c) Proof that  $x \in \{\mathbf{0}\}, \alpha \in \mathbb{F} \implies \alpha x \in \{\mathbf{0}\}$ :

$$\begin{aligned} \alpha x &= \alpha \mathbf{0} && \text{by } x \in \{\mathbf{0}\} \text{ hypothesis} \\ &= \mathbf{0} && \text{by definition of } \mathbf{0} \\ &\in \{\mathbf{0}\} \end{aligned}$$

2. Proof that  $\Omega$  is a linear subspace of itself:

(a) Proof that  $X \neq \emptyset$ :

$$X \neq \emptyset$$

(b) Proof that  $x, y \in X \implies x + y \in X$ :

$$x + y \in \{\mathbf{0}\} \quad \text{because } + : X \times X \rightarrow X \text{ (} X \text{ is closed under vector addition)}$$

(c) Proof that  $x \in X, \alpha \in \mathbb{F} \implies \alpha x \in X$ :

$$\alpha x \in X \quad \text{because } \cdot : \mathbb{F} \times X \rightarrow X \text{ (} X \text{ is closed under scalar-vector multiplication)}$$

<sup>1</sup> Michel and Herget (1993) page 81 (Definition 3.2.1), Berberian (1961) page 13 (Definition I.5.1), Halmos (1958), page 16

<sup>2</sup> Michel and Herget (1993) pages 81–83, Haaser and Sullivan (1991) page 43



**Proposition I.2.** <sup>3</sup> Let  $\mathbf{X} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$  be a LINEAR SPACE (Definition G.1 page 197). Let  $\text{span}$  be the LINEAR SPAN of a set  $Y$  in  $\mathbf{X}$ .

P R P	$\left\{ \begin{array}{l} Y \text{ is a SUBSET of the set } X \\ (Y \subseteq X) \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} \text{span}Y \text{ is a LINEAR SUBSPACE of } \mathbf{X}. \end{array} \right\}$
-------------	--

**Proposition I.3.** <sup>4</sup> Let  $\mathbf{X} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$  be a LINEAR SPACE and  $\mathbf{0}$  the zero vector of  $\mathbf{X}$ .

P R P	$\left\{ \begin{array}{l} Y \text{ is a LINEAR SUBSPACE of } \mathbf{X} \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} 1. \quad \mathbf{0} \in Y \\ 2. \quad Y \text{ is CONVEX in } \mathbf{X} \end{array} \right. \text{ and } \right\}$
-------------	---

PROOF:

$$\begin{aligned} Y \text{ is a subspace} &\implies \exists(\alpha y) \in Y \quad \forall \alpha \in \mathbb{F} && \text{by Definition I.1 page 240} \\ &\implies \exists 0 \in Y && \text{because } \alpha = 0 \in \mathbb{F} \end{aligned}$$

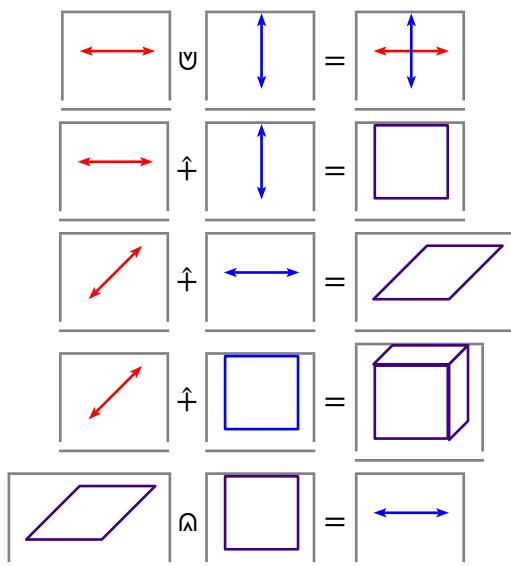
$$\begin{aligned} Y \text{ is a linear subspace} &\implies x + y \in Y \quad \forall x, y \in Y \\ &\implies \lambda x + (1 - \lambda)y \in Y \quad \forall x, y \in Y \\ &\implies Y \text{ is convex} \end{aligned}$$



**Definition I.2.** <sup>5</sup> Let  $\mathbf{X} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$  and  $\mathbf{Y} \triangleq (Y, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$  be LINEAR SUBSPACES (Definition I.1 page 240) of a LINEAR SPACE (Definition G.1 page 197)  $\Omega \triangleq (\Omega, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$ .

DEF	$X \hat{+} Y \triangleq (\{x + y   x \in X \text{ and } y \in Y\}, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$ (Minkowski addition)
	$X \uplus Y \triangleq (X \cup Y, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$ (subspace union)
	$X \Cap Y \triangleq (X \cap Y, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$ (subspace intersection)

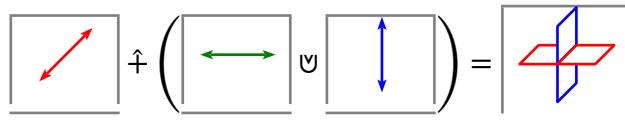
*Example I.2.* Some examples of operations on subspaces in  $\mathbb{R}^3$  are illustrated next:



<sup>3</sup> Michel and Herget (1993) page 86

<sup>4</sup> Michel and Herget (1993) page 81

<sup>5</sup> Wedderburn (1907) page 79

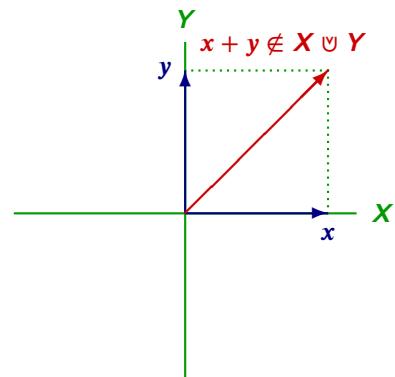


*Remark I.1.*

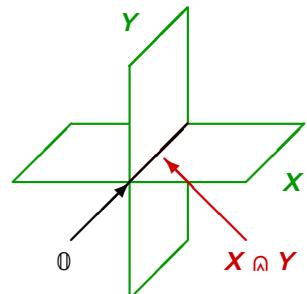
Notice the similarities between the properties of linear subspaces in a linear space (Proposition I.4 page 242) and the properties of closed sets in a topological space:

linear subspaces	closed sets
$\emptyset$	$\emptyset$
$\Omega$	$\Omega$
$X \dagger Y$	$X \cup Y$
$\bigcap_{n=1}^N X_n$	$\bigcap_{\gamma \in \Gamma} X_\gamma$

One key difference is that the union of two linear subspaces is not in general a linear subspace. For example, if  $x$  is the vector  $[1 0]$  in the  $x$  direction linear subspace of  $\mathbb{R}^2$  and  $y$  is the vector  $[0 1]$  in the  $y$  direction linear subspace, then  $x + y$  is not in the union of the two linear subspaces (it is not on the  $x$  axis or  $y$  axis but rather at  $(1, 1)$ ).<sup>6</sup>



In general, the set of all linear subspaces of a linear space  $\Omega$  is *not* closed under the subspace union ( $\cup$ ) operation; that is, the union of two linear subspaces is *not* necessarily a linear subspace. However the set *is* closed under Minkowski sum ( $\dagger$ ) and subspace intersection ( $\cap$ ). Proposition I.4 (next) shows four useful objects are always subspaces. Some of these in Euclidean space  $\mathbb{R}^3$  are illustrated to the right.



**Proposition I.4.** <sup>7</sup> Let  $X$  be a LINEAR SPACE (Definition G.1 page 197).

P R P	$\left\{ X_n \mid n=1,2,\dots,N \right\}$ are LINEAR SUBSPACES of $X$	$\Rightarrow \left\{ \begin{array}{l} 1. X_1 \dagger X_2 \dagger \dots \dagger X_N \text{ is a LINEAR SUBSPACE of } X \\ \text{and} \\ 2. X_1 \cap X_2 \cap \dots \cap X_N \text{ is a LINEAR SUBSPACE of } X \end{array} \right.$
-------------	---	--

PROOF: For a structure to be a linear subspace of  $X$ , it must satisfy the requirements of Definition I.1 (page 240).

1. Proof that  $X_1 \dagger X_2 \dagger \dots \dagger X_N$  is a *linear subspace* (proof by induction):

- (a) proof for  $N = 1$  case: by left hypothesis.
- (b) proof for  $N = 2$  case:

<sup>6</sup> Michel and Herget (1993) page 82

<sup>7</sup> Michel and Herget (1993) pages 81–83

i. proof that  $\mathbf{X}_1 \hat{+} \mathbf{X}_2 \neq \emptyset$ :

$$\begin{aligned}\mathbf{X}_1 \hat{+} \mathbf{X}_2 &= \{\mathbf{v} + \mathbf{w} \mid \mathbf{v} \in \mathbf{X}_1 \text{ and } \mathbf{w} \in \mathbf{Y}\} && \text{by Definition I.2 page 241} \\ &\supseteq \{\mathbf{v} + \mathbf{w} \mid \mathbf{v} \in \{\mathbf{0}\} \subseteq \mathbf{X}_1 \text{ and } \mathbf{w} \in \{\mathbf{0}\} \subseteq \mathbf{Y}\} \\ &= \{\mathbf{0} + \mathbf{0}\} \\ &= \{\mathbf{0}\} \\ &\neq \emptyset\end{aligned}$$

ii. proof that  $\mathbf{x}, \mathbf{y} \in \mathbf{X}_1 \hat{+} \mathbf{X}_2 \implies \mathbf{x} + \mathbf{y} \in \mathbf{X}_1 \hat{+} \mathbf{X}_2$ :

$$\begin{aligned}\mathbf{x} + \mathbf{y} &= (\mathbf{v}_1 + \mathbf{w}_1) + (\mathbf{v}_2 + \mathbf{w}_2) && \text{by } \mathbf{x}, \mathbf{y} \in \mathbf{X}_1 \hat{+} \mathbf{X}_2 \text{ hypothesis} \\ &= \underbrace{(\mathbf{v}_1 + \mathbf{v}_2)}_{\text{in } \mathbf{X}_1} + \underbrace{(\mathbf{w}_1 + \mathbf{w}_2)}_{\text{in } \mathbf{X}_2 \text{ because } \mathbf{X}_2 \text{ is a linear subspace}} && \text{by Definition G.1 page 197} \\ &\in \{\mathbf{v} + \mathbf{w} \mid \mathbf{v} \in \mathbf{X}_1 \text{ and } \mathbf{w} \in \mathbf{Y}\} \\ &= \mathbf{X}_1 \hat{+} \mathbf{X}_2 && \text{by Definition I.2 page 241}\end{aligned}$$

iii. proof that  $\mathbf{v} \in \mathbf{X}_1 \hat{+} \mathbf{X}_2, \alpha \in F \implies \alpha\mathbf{v} \in \mathbf{X}_1 \hat{+} \mathbf{X}_2$ :

$$\begin{aligned}\alpha\mathbf{x} &= \alpha(\mathbf{v}_1 + \mathbf{w}_1) && \text{by } \mathbf{x} \in \mathbf{X}_1 \hat{+} \mathbf{X}_2 \text{ hypothesis} \\ &= \underbrace{\alpha\mathbf{v}_1}_{\text{in } \mathbf{X}_1} + \underbrace{\alpha\mathbf{w}_1}_{\text{in } \mathbf{X}_2 \text{ because } \mathbf{X}_2 \text{ is a linear subspace}} && \text{by Definition G.1 page 197} \\ &\in \{\mathbf{v} + \mathbf{w} \mid \mathbf{v} \in \mathbf{X}_1 \text{ and } \mathbf{w} \in \mathbf{Y}\} \\ &= \mathbf{X}_1 \hat{+} \mathbf{X}_2 && \text{by Definition I.2 page 241}\end{aligned}$$

(c) Proof that [N case]  $\implies$  [N + 1 case]:

$$\begin{aligned}\mathbf{X}_1 \hat{+} \mathbf{X}_2 \hat{+} \cdots \hat{+} \mathbf{X}_{N+1} &= \underbrace{(\mathbf{X}_1 \hat{+} \mathbf{X}_2 \hat{+} \cdots \hat{+} \mathbf{X}_N)}_{\text{linear subspace by } N \text{ case hypothesis}} \hat{+} \mathbf{X}_{N+1} \\ &\implies \text{linear subspace by } N = 2 \text{ case (item (1b) page 242)}\end{aligned}$$

2. Proof that  $\mathbf{X}_1 \wedge \mathbf{X}_2 \wedge \cdots \wedge \mathbf{X}_N$  is a *linear subspace* (proof by induction):

(a) proof for  $N = 1$  case:  $\mathbf{X}_1$  is a linear subspace by left hypothesis.

(b) Proof for  $N = 2$  case:

i. proof that  $\mathbf{X} \wedge \mathbf{Y} \neq \emptyset$ :

$$\begin{aligned}\mathbf{X} \wedge \mathbf{Y} &= \{\mathbf{x} \in X \mid \mathbf{x} \in \mathbf{X} \text{ and } \mathbf{w} \in \mathbf{Y}\} \\ &\supseteq \{\mathbf{x} \in X \mid \mathbf{x} \in \{\mathbf{0}\} \subseteq \mathbf{X} \text{ and } \mathbf{w} \in \{\mathbf{0}\} \subseteq \mathbf{Y}\} \\ &= \{\mathbf{0} + \mathbf{0}\} \\ &= \{\mathbf{0}\} \\ &\neq \emptyset\end{aligned}$$

ii. proof that  $\mathbf{x}, \mathbf{y} \in \mathbf{X} \wedge \mathbf{Y} \implies \mathbf{x} + \mathbf{y} \in \mathbf{X} \wedge \mathbf{Y}$ :

$$\begin{aligned}\mathbf{x}, \mathbf{y} \in \mathbf{X} \wedge \mathbf{Y} &\implies \mathbf{x}, \mathbf{y} \in \mathbf{X} \text{ and } \mathbf{x}, \mathbf{y} \in \mathbf{Y} \\ &\implies \mathbf{x} + \mathbf{y} \in \mathbf{X} \text{ and } \mathbf{x} + \mathbf{y} \in \mathbf{Y} \quad \text{because } \mathbf{X} \text{ and } \mathbf{Y} \text{ are linear subspaces} \\ &\implies \mathbf{x} + \mathbf{y} \in \mathbf{X} \wedge \mathbf{Y}\end{aligned}$$

iii. proof that  $\mathbf{v} \in \mathbf{X} \wedge \mathbf{Y}, \alpha \in F \implies \alpha\mathbf{v} \in \mathbf{X} \wedge \mathbf{Y}$ :

$$\begin{aligned}\mathbf{x} \in \mathbf{X} \wedge \mathbf{Y} &\implies \mathbf{x} \in \mathbf{X} \text{ and } \mathbf{x} \in \mathbf{Y} \\ &\implies \alpha\mathbf{x} \in \mathbf{X} \text{ and } \alpha\mathbf{x} \in \mathbf{Y} \quad \text{because } \mathbf{X} \text{ and } \mathbf{Y} \text{ are linear subspaces} \\ &\implies \alpha\mathbf{x} \in \mathbf{X} \wedge \mathbf{Y}\end{aligned}$$

(c) Proof that [ $N$  case]  $\implies$  [ $N + 1$  case]:

$$\begin{aligned} \mathbf{X}_1 \wedge \mathbf{X}_2 \wedge \cdots \wedge \mathbf{X}_{N+1} &= \underbrace{(\mathbf{X}_1 \wedge \mathbf{X}_2 \wedge \cdots \wedge \mathbf{X}_N)}_{\text{linear subspace by } N \text{ case hypothesis}} \wedge \mathbf{X}_{N+1} \\ &\implies \text{linear subspace by } N = 2 \text{ case (item (2b) page 243)} \end{aligned}$$

⇒

Every linear subspace contains the zero vector  $\mathbb{0}$  (Proposition I.3 page 241). But if a pair of linear subspaces of a linear space  $\mathbf{X}$  *only* have  $\mathbb{0}$  in common, then any vector in  $\mathbf{X}$  can be *uniquely* represented by a single vector from each of the two subspaces (next).

**Theorem I.1.**<sup>8</sup> Let  $\mathbf{X} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$  and  $\mathbf{Y} \triangleq (Y, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$  be LINEAR SUBSPACES (Definition I.1 page 240) of a LINEAR SPACE (Definition G.1 page 197)  $\Omega \triangleq (\Omega, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$ .

T	H	M	$X \cap Y = \{\mathbb{0}\} \iff \left\{ \begin{array}{l} \text{for every } u \in X \hat{+} Y \text{ there exist } x \in X \text{ and } y \in Y \text{ such that} \\ \quad \text{1. } u = x + y \quad \text{and} \\ \quad \text{2. } x \text{ and } y \text{ are UNIQUE.} \end{array} \right\}$
---	---	---	--

PROOF:

1. Proof that  $X \cap Y = \{\mathbb{0}\} \implies \text{unique } x, y$ :

Suppose that  $x$  and  $y$  are not unique, but rather  $u = x_1 + y_1 = x_2 + y_2$  where  $x_1, x_2 \in X$  and  $y_1, y_2 \in Y$ .

$$\begin{aligned} u = x_1 + y_1 = x_2 + y_2 &\implies \underbrace{x_1 - x_2}_{\in X} = \underbrace{y_2 - y_1}_{\in Y} \\ &\implies x_1 - x_2, y_2 - y_1 \in X \cap Y \\ &\implies x_1 - x_2 = y_2 - y_1 = \mathbb{0} \quad \text{by left hypothesis} \\ &\implies x_1 = x_2 \quad \text{and} \quad y_2 = y_1 \\ &\implies x \text{ and } y \text{ are unique} \end{aligned}$$

2. Proof that  $X \cap Y = \{\mathbb{0}\} \iff \text{unique } x, y$ :

$$\begin{aligned} u &= x + y \\ &= x + y + y - y \quad \text{for some vector } y \in X \cap Y \\ &= \underbrace{(x + y)}_{\in X} + \underbrace{(y - y)}_{\in Y} \quad \text{because } x \in X \text{ and } y \in X \cap Y \dots \\ &\implies x \text{ and } y \text{ are not unique if } y \neq \mathbb{0} \\ &\implies y = \mathbb{0} \quad \text{by right hypothesis} \\ &\implies X \cap Y = \{\mathbb{0}\} \end{aligned}$$

⇒

**Theorem I.2.**<sup>9</sup> Let  $\Omega$  be a linear subspace and  $\mathcal{Z}^\Omega$  the set of closed linear subspaces of  $\Omega$ .

$(\mathcal{Z}^\Omega, \hat{+}, \wedge, \mathbb{0}, \Omega; \subseteq)$  is a LATTICE. In particular

T	H	M	$\begin{aligned} \mathbf{X} \hat{+} \mathbf{X} &= \mathbf{X} & \mathbf{X} \wedge \mathbf{X} &= \mathbf{X} & \forall \mathbf{X} \in \mathcal{Z}^\Omega \\ \mathbf{X} \hat{+} \mathbf{Y} &= \mathbf{Y} \hat{+} \mathbf{X} & \mathbf{X} \wedge \mathbf{Y} &= \mathbf{Y} \wedge \mathbf{X} & \forall \mathbf{X}, \mathbf{Y} \in \mathcal{Z}^\Omega \\ (\mathbf{X} \hat{+} \mathbf{Y}) \hat{+} \mathbf{Z} &= \mathbf{X} \hat{+} (\mathbf{Y} \hat{+} \mathbf{Z}) & (\mathbf{X} \wedge \mathbf{Y}) \wedge \mathbf{Z} &= \mathbf{X} \wedge (\mathbf{Y} \wedge \mathbf{Z}) & \forall \mathbf{X}, \mathbf{Y}, \mathbf{Z} \in \mathcal{Z}^\Omega \\ \mathbf{X} \hat{+} (\mathbf{X} \wedge \mathbf{Y}) &= \mathbf{X} & \mathbf{X} \wedge (\mathbf{X} \hat{+} \mathbf{Y}) &= \mathbf{X} & \forall \mathbf{X}, \mathbf{Y} \in \mathcal{Z}^\Omega \end{aligned}$
---	---	---	---

PROOF: These results follow directly from the properties of lattices.

⇒

<sup>8</sup> Michel and Herget (1993) page 83 (Theorem 3.2.12), Kubrusly (2001) page 67 (Theorem 2.14)

<sup>9</sup> Iturrioz (1985) pages 56–57



## I.2 Subspaces of an inner product space

**Definition I.3.** <sup>10</sup> Let  $\Omega \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle | \nabla \rangle)$  be an INNER PRODUCT SPACE (Definition K.1 page 259).

**D E F** The **orthogonal complement**  $A^\perp$  in  $\Omega$  of a set  $A \subseteq X$  is

$$A^\perp \triangleq \{x \in X \mid \langle x | y \rangle = 0 \quad \forall y \in A\}.$$

The expression  $A^{\perp\perp}$  is defined as  $(A^\perp)^\perp$ .

**Proposition I.5.** <sup>11</sup> Let  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle | \nabla \rangle)$  be an INNER PRODUCT SPACE (Definition K.1 page 259).

**P R P**  $A \subseteq B \implies B^\perp \subseteq A^\perp \quad \forall A, B \in 2^X$  (ANTITONE)

PROOF:

$$\begin{aligned} B^\perp &\triangleq \{x \in X \mid \langle x | y \rangle = 0 \quad \forall y \in B\} && \text{by definition of } B^\perp \text{ (Definition I.3 page 245)} \\ &\subseteq \{x \in X \mid \langle x | y \rangle = 0 \quad \forall y \in A\} && \text{by } A \subseteq B \text{ hypothesis} \\ &= A^\perp && \text{by definition of } A^\perp \text{ (Definition I.3 page 245)} \end{aligned}$$



Every *linear space*  $X$  contains  $\mathbf{0}$  and  $X$  as *linear subspaces* (Proposition I.1 page 240). If  $X$  is also an *inner product space*, then  $\mathbf{0}$  and  $X$  are *orthogonal complements* of each other (next proposition).

**Proposition I.6.** <sup>12</sup> Let  $\Omega \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle | \nabla \rangle)$  be an INNER PRODUCT SPACE (Definition K.1 page 259) and  $\mathbf{0}$  the VECTOR ADDITIVE IDENTITY ELEMENT (Definition G.1 page 197) in  $\Omega$ .

**P R P**

1.	$\{\mathbf{0}\}^\perp = X$
2.	$X^\perp = \{\mathbf{0}\}$

PROOF:

$$\begin{aligned} \{\mathbf{0}\}^\perp &= \{x \in X \mid \langle x | y \rangle = 0 \quad \forall y \in \{\mathbf{0}\}\} && \text{by definition of } \perp \text{ (Definition I.3 page 245)} \\ &= \{x \in X \mid \langle x | \mathbf{0} \rangle = 0\} \\ &= X \\ X^\perp &= \{x \in X \mid \langle x | y \rangle = 0 \quad \forall y \in X\} && \text{by definition of } \perp \text{ (Definition I.3 page 245)} \\ &= \{x \in X \mid \langle x | x \rangle = 0\} \\ &= \{\mathbf{0}\} \end{aligned}$$



For any set  $A$  contained in a linear space  $X$ ,  $A^{\perp\perp}$  is a *linear subspace*, and it is the smallest linear subspace containing the set  $A$  ( $A^{\perp\perp} = \text{span}A$ , next theorem). In the case that  $A$  is a *linear subspace* rather than just a subset, results simplify significantly (next corollary).

<sup>10</sup> Berberian (1961) page 59 (Definition III.2.1), Michel and Herget (1993) page 382, Kubrusly (2001) page 328

<sup>11</sup> Berberian (1961) page 60 (Theorem III.2.2), Kubrusly (2011) page 326

<sup>12</sup> Kubrusly (2011) page 326, Michel and Herget (1993) page 383

**Theorem I.3.** <sup>13</sup> Let  $\Omega \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle | \nabla \rangle)$  be an INNER PRODUCT SPACE (Definition K.1 page 259). Let  $\text{span}A$  be the span of a set  $A$  (Definition H.2 page 209).

THEM

$$\left\{ \begin{array}{l} A \text{ is a subset of } X \\ (A \subseteq X) \end{array} \right\} \implies \left\{ \begin{array}{l} 1. \quad A \cap A^\perp = \begin{cases} \{\emptyset\} & \text{if } \emptyset \in A \\ \emptyset & \text{if } \emptyset \notin A \end{cases} \text{ and} \\ 2. \quad A \subseteq A^{\perp\perp} = \text{span}A \\ 3. \quad A^\perp = A^{\perp\perp\perp} = A^{\perp^-} = A^{-\perp} = (\text{span}A)^\perp \text{ and} \\ 4. \quad A^\perp \text{ is a subspace of } \Omega \end{array} \right\}$$

PROOF:

1. Proof that  $A \cap A^\perp = \dots$ :

$$\begin{aligned} A \cap A^\perp &= \{x \in X | x \in A\} \cap \{x \in X | \langle x | y \rangle \quad \forall y \in A\} && \text{by definition of } A^\perp \\ &= \{x \in X | x \in A \text{ and } \langle x | y \rangle \quad \forall y \in A\} \\ &= \begin{cases} \{\emptyset\} & \text{if } \emptyset \in A \\ \emptyset & \text{if } \emptyset \notin A \end{cases} \end{aligned}$$

2. Proof that  $A \subseteq A^{\perp\perp} = \text{span}A$ :

$$\begin{aligned} x \in A &\implies \{x\}^{\perp\perp} \subseteq A^{\perp\perp} \\ &\implies x \in \{x\}^{\perp\perp} \subseteq A^{\perp\perp} \\ &\implies x \in A^{\perp\perp} \end{aligned}$$

but

$$x \in A^{\perp\perp} \not\implies x \in A$$

Here is an example for the  $\not\implies$  part using the linear space  $\mathbb{R}^3$ :

- (a) Let  $A \triangleq \{i\}$ , where  $i$  is the unit vector on the x-axis.
- (b) Then  $A^\perp = \{x \in X | x \in \text{yz plane}\}$ .
- (c) Then  $A^{\perp\perp} = \{x \in X | x \in \text{x axis}\}$ .
- (d) Therefore,  $A \subsetneq A^{\perp\perp}$

3. Proof for  $A^\perp$  equivalent expressions:

- (a) Proof that  $A^\perp = A^{\perp\perp\perp}$ :

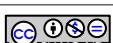
$$\begin{aligned} A^\perp &\subseteq (A^\perp)^{\perp\perp} && \text{by item (2)} \\ &= (A^{\perp\perp})^\perp \\ &= A^{\perp\perp\perp} && \text{by Definition I.3 page 245} \\ A^{\perp\perp\perp} &= (A^{\perp\perp})^\perp && \text{by Definition I.3 page 245} \\ &\subseteq A^\perp && \text{by item (2) and Proposition I.5 (page 245)} \end{aligned}$$

- (b) Proof that  $A^{\perp\perp\perp} = (\text{span}A)^\perp$ : follows directly from item (2) ( $A^\perp = \text{span}A$ ).

- (c) Proof that  $A^\perp = A^{\perp^-}$ :

- i. Let  $(x_n)$  be an  $A^\perp$ -valued sequence that converges to the limit  $x$  in  $X$ .

<sup>13</sup> Michel and Herget (1993) page 383, Kubrusly (2011) page 326



ii. The limit point  $x$  must be in  $A^\perp$  because for all  $y \in A$

$$\begin{aligned}\langle x | y \rangle &= \langle \lim x_n | y \rangle && \text{by definition of the sequence } (\mathbf{x}_n) \\ &= \lim \langle x_n | y \rangle \\ &= 0 && \text{because } (\mathbf{x}_n) \text{ is } A^\perp\text{-valued}\end{aligned}$$

iii. Because  $\langle x | y \rangle = 0 \quad \forall y \in A$ ,  $x$  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  $(y_n)$  be an  $A^\perp$ -valued sequence that converges in  $X$  to  $y$ .

iii. Thus  $A^\perp \perp A^-$  because

$$\begin{aligned}\langle y | x \rangle &= \langle \lim y_n | x \rangle && \text{by definition of } (y_n) \\ &= \lim \langle y_n | x \rangle \\ &= 0 && \text{because } (y_n) \text{ is } A^\perp\text{-valued}\end{aligned}$$

iv. Because  $A^\perp \perp A^-$ , so  $A^\perp \subseteq A^{\perp^-}$ .

v. But  $A^{\perp^-} \subseteq A^\perp$  because

$$A \subseteq A^- \implies A^{\perp^-} \subseteq A^\perp \quad \text{by } \textit{antitone} \text{ property (Proposition I.5 page 245)}$$

vi. And so  $A^\perp = A^{\perp^-}$ .

4. Proof that  $A^\perp$  is a **subspace** of  $\Omega$  (must satisfy the conditions of Definition I.1 page 240):

(a) Proof that  $A^\perp \neq \emptyset$ :  $A^\perp$  has at least one element, the element  $0$ ...

$$\begin{aligned}\langle 0 | y \rangle &= 0 \quad \forall y \in A && \text{by definition of } 0 \\ \implies 0 &\in A^\perp && \text{by definition of } A^\perp \text{ (Definition I.3 page 245)}\end{aligned}$$

(b) Proof that  $A^\perp \subseteq X$ :

$$\begin{aligned}u \in A^\perp &\implies u \in \{x \in X \mid \langle x | y \rangle = 0 \quad \forall y \in A\} && \text{by definition of } A^\perp \text{ (Definition I.3 page 245)} \\ &\implies u \in X && \text{by definition of sets}\end{aligned}$$

(c) Proof that  $u, v \in A^\perp \implies (u + v) \in A^\perp$ :

$$\begin{aligned}u, v \in A^\perp &\implies \langle u | y \rangle = \langle v | y \rangle = 0 \quad \forall y \in A && \text{by definition of } A^\perp \text{ (Definition I.3 page 245)} \\ &\implies \langle u | y \rangle + \langle v | y \rangle = 0 \quad \forall y \in A \\ &\implies \langle u + v | y \rangle = 0 \quad \forall y \in A && \text{by } \textit{additive} \text{ property of } \langle \triangle | \nabla \rangle \text{ (Definition K.1 page 259)} \\ &\implies u + v \in A^\perp && \text{by definition of } A^\perp \text{ (Definition I.3 page 245)}\end{aligned}$$

(d) Proof that  $v \in \Omega \implies \alpha v \in A^\perp$ :

$$\begin{aligned}v \in A^\perp &\implies \langle v | y \rangle = 0 \quad \forall y \in A && \text{by definition of } A^\perp \text{ (Definition I.3 page 245)} \\ &\implies \alpha \langle v | y \rangle = \alpha \cdot 0 \quad \forall y \in A \\ &\implies \langle \alpha v | y \rangle = 0 \quad \forall y \in A && \text{by } \textit{homogeneous} \text{ property of } \langle \triangle | \nabla \rangle \text{ (Definition K.1 page 259)} \\ &\implies \alpha v \in A^\perp && \text{by definition of } A^\perp \text{ (Definition I.3 page 245)}\end{aligned}$$



**Corollary I.1.** Let  $\mathbf{X} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle | \nabla \rangle)$  and  $\mathbf{Y} \triangleq (Y, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle | \nabla \rangle)$  be INNER PRODUCT SPACES. Let  $\text{span } Y$  be the span of the set  $Y$  (Definition H.2 page 209).

C  
O  
R

$$\left\{ \begin{array}{l} \mathbf{Y} \text{ is a linear subspace of } \mathbf{X} \end{array} \right\} \implies \left\{ \begin{array}{ll} 1. & Y \cap Y^\perp = \{\mathbb{0}\} \\ 2. & Y = Y^{\perp\perp} = \text{span } Y \\ 3. & Y^\perp = Y^{\perp\perp\perp} \\ 4. & Y^\perp \text{ is a subspace of } \mathbf{X} \end{array} \right. \text{ and}$$

PROOF:

1. Proof that  $Y \cap Y^\perp = \{\mathbb{0}\}$ : This follows from Theorem I.3 (page 246) and the fact that all subspaces contain the zero vector  $\mathbb{0}$  (Proposition I.3 page 241).
2. Proof that  $Y = Y^{\perp\perp} = \text{span } Y$ : This follows directly from Theorem I.3 (page 246).
3. Proof that  $Y^\perp = Y^{\perp\perp\perp}$ : This follows directly from Theorem I.3 (page 246).
4. Proof that  $Y^\perp$  is a **subspace** of  $\mathbf{X}$ : This follows directly from Theorem I.3 (page 246).

⇒

**Theorem I.4.** <sup>14</sup> Let  $\mathbf{Y} \triangleq (Y, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle | \nabla \rangle)$  and  $\mathbf{Z} \triangleq (Z, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle | \nabla \rangle)$  be LINEAR SUBSPACES of an INNER PRODUCT SPACE  $\mathbf{Q} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle | \nabla \rangle)$ .

T  
H  
M

$$Y \perp Z \implies Y \cap Z = \{\mathbb{0}\}$$

⇒

**Theorem I.5.** <sup>15</sup> Let  $\mathbf{Y} \triangleq (Y, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle | \nabla \rangle)$  and  $\mathbf{Z} \triangleq (Z, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle | \nabla \rangle)$  be linear subspaces of an INNER PRODUCT SPACE  $\mathbf{Q} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle | \nabla \rangle)$ .

T  
H  
M

$$\left\{ \begin{array}{l} 1. \quad Y \perp Z \text{ and} \\ 2. \quad x \in Y \hat{+} Z \end{array} \right\} \implies \left\{ \begin{array}{l} 1. \quad \text{There exists } y \in Y \text{ and } z \in Z \text{ such that } x = y + z \text{ and} \\ 2. \quad y \text{ and } z \text{ are UNIQUE.} \end{array} \right\}$$

PROOF:

1. Proof that  $y$  and  $z$  exist: by definition of Minkowski addition operator  $\hat{+}$  (Definition I.2 page 241).
2. Proof that  $y$  and  $z$  are *unique*:

(a) Suppose  $x = y_1 + z_1 = y_2 + z_2$  for  $y_1, y_2 \in Y$  and  $z_1, z_2 \in Z$ .

<sup>14</sup> Kubrusly (2001) page 324

<sup>15</sup> Berberian (1961) page 61 (Theorem III.2.3)

(b) This implies

$$\begin{aligned} \mathbb{0} &= \mathbf{x} - \mathbf{x} \\ &= (\mathbf{y}_1 + \mathbf{z}_1) - (\mathbf{y}_1 + \mathbf{z}_2) \\ &= \underbrace{(\mathbf{y}_1 - \mathbf{y}_2)}_{\text{in } Y} + \underbrace{(\mathbf{z}_1 - \mathbf{z}_2)}_{\text{in } Z} \end{aligned}$$

- (c) Because  $\mathbf{y}_1 - \mathbf{y}_2 \in Y$ ,  $\mathbf{z}_1 - \mathbf{z}_2 \in Z$ ,  $(\mathbf{y}_1 - \mathbf{y}_2) + (\mathbf{z}_1 - \mathbf{z}_2) = \mathbb{0}$ , and  $\langle \mathbf{y}_1 - \mathbf{y}_2 | \mathbf{z}_1 - \mathbf{z}_2 \rangle = 0$ , then by Theorem K.9 (page 272),  $\mathbf{y}_1 - \mathbf{y}_2 = \mathbb{0}$  and  $\mathbf{z}_1 - \mathbf{z}_2 = \mathbb{0}$ .
- (d) This implies  $\mathbf{y}_1 = \mathbf{y}_2$  and  $\mathbf{z}_1 = \mathbf{z}_2$ .
- (e) This implies  $\mathbf{y}$  and  $\mathbf{z}$  are *unique*.



## I.3 Subspaces of a Hilbert Space

**Theorem I.6.** <sup>16</sup> Let  $H \triangleq (X, +, \cdot, (\mathbb{F}, +, \times), \langle \Delta | \nabla \rangle)$  be a HILBERT SPACE (Definition ?? page ??). Let  $Y$  be a SUBSET of  $X$ , and let  $d(x, Y) \triangleq \inf_{y \in Y} \|x - y\|$ .

T H M	$\left\{ \begin{array}{l} 1. \quad Y \neq \emptyset \\ 2. \quad Y \text{ is CLOSED} \\ 3. \quad Y \text{ is CONVEX} \end{array} \right. \quad \begin{array}{l} \text{and} \\ \text{and} \end{array} \quad \begin{array}{l} \text{(Definition M.6 page 286)} \\ \text{by parallelogram law (page 267)} \end{array}$	$\Rightarrow$	$\left\{ \begin{array}{l} \text{There exists } p \in Y \text{ such that} \\ \begin{array}{l} 1. \quad d(x, Y) = \ x - p\  \quad \text{and} \\ 2. \quad p \text{ is UNIQUE.} \end{array} \end{array} \right.$
-------------	--	---------------	--

PROOF:

- Let  $\delta \triangleq \inf \{x - y | y \in Y\}$ .
- Let  $(\mathbf{y}_n)_{n \in \mathbb{Z}}$  be a sequence such that  $\|\mathbf{x} - \mathbf{y}_n\| \rightarrow \delta$ .
- Proof that  $(\mathbf{y}_n)$  is *Cauchy*:

$$\begin{aligned} &\lim_{m,n \rightarrow \infty} \|\mathbf{y}_n - \mathbf{y}_m\|^2 \\ &= \lim_{m,n \rightarrow \infty} \|(\mathbf{y}_n - \mathbf{x}) + (\mathbf{x} - \mathbf{y}_m)\|^2 \\ &= \lim_{m,n \rightarrow \infty} \left\{ -\|(\mathbf{y}_n - \mathbf{x}) - (\mathbf{x} - \mathbf{y}_m)\|^2 + 2\|\mathbf{y}_n - \mathbf{x}\|^2 + 2\|\mathbf{x} - \mathbf{y}_m\|^2 \right\} \quad \text{by parallelogram law (page 267)} \\ &= \lim_{m,n \rightarrow \infty} \left\{ -4 \left\| \underbrace{\left( \frac{1}{2}\mathbf{y}_n + \frac{1}{2}\mathbf{y}_m \right)}_{\text{in } Y \text{ by convexity}} - \mathbf{x} \right\|^2 + 2\|\mathbf{y}_n - \mathbf{x}\|^2 + 2\|\mathbf{x} - \mathbf{y}_m\|^2 \right\} \\ &\leq \lim_{m,n \rightarrow \infty} \left\{ -4\delta^2 + 2\|\mathbf{y}_n - \mathbf{x}\|^2 + 2\|\mathbf{x} - \mathbf{y}_m\|^2 \right\} \quad \text{by definition of } \delta \text{ (item (1))} \\ &= -4\delta^2 + \lim_{m,n \rightarrow \infty} \left\{ 2\|\mathbf{y}_n - \mathbf{x}\|^2 \right\} + \lim_{m,n \rightarrow \infty} \left\{ 2\|\mathbf{x} - \mathbf{y}_m\|^2 \right\} \\ &= -4\delta^2 + 2\delta^2 + 2\delta^2 \quad \text{by definition of } \delta \text{ (item (1))} \\ &= 0 \end{aligned}$$

<sup>16</sup> Kubrusly (2001) page 330 (Theorem 5.13), Aliprantis and Burkinshaw (1998) page 290 (Theorem 33.6), Berberian (1961) page 68 (Theorem III.5.1)

4. Proof that  $d(x, Y) = \|x - y\|$ : because  $(y_n)$  is *Cauchy* (item (1)) and by the *closed* hypothesis.
5. Proof that  $y$  is *unique*: Because in a metric space, the limit of a convergent sequence is *unique*.

⇒

**Theorem I.7.** <sup>17</sup> Let  $H \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle | \nabla \rangle)$  be a HILBERT SPACE (Definition ?? page ??). Let  $d(x, Y) \triangleq \inf_{y \in Y} \|x - y\|$ . Let  $Y \triangleq (Y, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle | \nabla \rangle)$  and  $Y^\perp$  the ORTHOGONAL COMPLEMENT of  $Y$ .

T  
H  
M

$$\{ Y \text{ is a SUBSPACE of } H \} \implies \left\{ \begin{array}{l} \text{There exists } p \in Y \text{ such that} \\ \begin{array}{ll} 1. & d(x, Y) = \|x - p\| \text{ and} \\ 2. & p \text{ is UNIQUE} \text{ and} \\ 3. & x - p \in Y^\perp. \end{array} \end{array} \right\}$$

**Theorem I.8** (Projection Theorem). <sup>18</sup> Let  $H \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle | \nabla \rangle)$  be a Hilbert space.

T  
H  
M

$$\{ Y \text{ is a SUBSPACE of } H \} \implies \{ Y \dagger Y^\perp = H \}$$

PROOF:

$$\begin{aligned} Y \dagger Y^\perp &= [Y \dagger Y^\perp]^\perp && \text{by Corollary I.1 page 248} \\ &= [Y^\perp \wedge Y^{\perp\perp}]^\perp && \text{by Proposition I.5 (page 245)} \\ &= \{\emptyset\}^\perp && \text{by Corollary I.1 page 248} \\ &= H && \text{by Proposition I.6 page 245} \end{aligned}$$

⇒

The inclusion relation  $\subseteq$  is an order relation on the set of subspaces of a linear space  $\Omega$ .

**Proposition I.7.** Let  $S$  be the set of subspaces of a linear space  $\Omega$ . Let  $\subseteq$  be the inclusion relation.

P  
R  
P

$(S, \subseteq)$  is an ordered set

PROOF:  $(S, \subseteq)$  is an *ordered set* and because

- |    |  |                         |                  |  |     |                          |
|----|--|-------------------------|------------------|--|-----|--------------------------|
| 1. | $X \subseteq X$  | $\forall X \in S$       | (reflexive)      |  |     | $\boxed{\quad}$ preorder |
| 2. | $X \subseteq Y$ and $Y \subseteq Z \implies X \subseteq Z$ | $\forall X, Y, Z \in S$ | (transitive)     |  | and | ⇒                        |
| 3. | $X \subseteq Y$ and $Y \subseteq X \implies X = Y$         | $\forall X, Y \in S$    | (anti-symmetric) |  |     |                          |

**Theorem I.9.** <sup>19</sup> Let  $H$  be a Hilbert space and  $\mathcal{Z}^H$  the set of closed linear subspaces of  $H$ .

T  
H  
M

$(\mathcal{Z}^H, \dagger, \wedge, \mathbf{0}, H; \subseteq)$  is an ORTHOMODULAR LATTICE. In particular

- |    |  |                      |                         |  |
|----|--|----------------------|-------------------------|--|
| 1. | $X \dagger X^\perp = H$                              | $\forall X \in H$    | (COMPLEMENTED)          |  |
| 2. | $X \wedge X^\perp = \mathbf{0}$                      | $\forall X \in H$    | (COMPLEMENTED)          |  |
| 3. | $(X^\perp)^\perp = X$                                | $\forall X \in H$    | (INVOLUTORY)            |  |
| 4. | $X \leq Y \implies Y^\perp \leq X^\perp$             | $\forall X, Y \in H$ | (ANTITONE)              |  |
| 5. | $X \leq Y \implies X \dagger (X^\perp \wedge Y) = Y$ | $\forall X, Y \in H$ | (ORTHOMODULAR IDENTITY) |  |

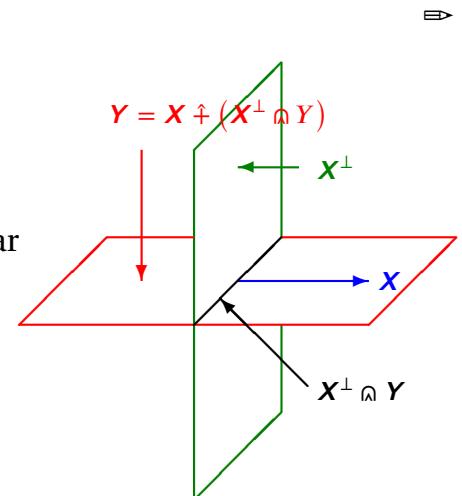
<sup>17</sup> Kubrusly (2001) page 330 (Theorem 5.13)

<sup>18</sup> Bachman and Narici (1966) page 172 (Theorem 10.8), Kubrusly (2001) page 339 (Theorem 5.20)

<sup>19</sup> Iturrioz (1985) pages 56–57

PROOF:

1. Proof for *complemented* (1) property: by *Projection Theorem* (Theorem I.8 page 250).
2. Proof for *complemented* (2) property: by Corollary I.1 (page 248).
3. Proof for *involutory* property: by Corollary I.1 (page 248).
4. Proof for *antitone* property: by Proposition I.5 (page 245).
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

$$X \subseteq Y \implies Y = X + (X^\perp \cap Y).$$

**Corollary I.2.** Let  $H$  be a Hilbert space with orthogonality operation  $\perp$ . Let  $(2^H, \hat{+}, \wedge, \mathbf{0}, \mathbf{1}, H; \subseteq)$  be the lattice of subspaces of  $H$ .

C O R	$(X \hat{+} Y)^\perp = X^\perp \wedge Y^\perp \quad \forall X, Y \in 2^H \quad (\text{DE MORGAN}) \quad \text{and}$
	$(X \wedge Y)^\perp = X^\perp \hat{+} Y^\perp \quad \forall X, Y \in 2^H \quad (\text{DE MORGAN})$

PROOF: By properties of *orthocomplemented lattices*.

## I.4 Subspace Metrics

**Definition I.4** (Hilbert space gap metric). <sup>20</sup> Let  $X$  be a **Hilbert space** and  $S$  the set of subspaces of  $X$ . Then we define the following metric between subspaces of  $X$ .

D E F	$d(V, W) \triangleq \ P - Q\  \quad \forall V, W \in S$ (the distance between subspaces $V$ and $W$ is the size of the difference of their projection operators) $P$ is the projection operator that generates the subspace $V$ ) $Q$ is the projection operator that generates the subspace $W$ ).
	$V \triangleq PX$ $W \triangleq QX$

**Definition I.5** (Banach space gap metric). <sup>21</sup> Let  $X$  be a **Banach space** and  $S$  the set of subspaces of  $X$ . Then we define the following metric between subspaces of  $X$ .

D E F	$d(V, W) \triangleq \max \left\{ \sup_{v \in V, \ v\ =1} p(v, W), \sup_{w \in W, \ w\ =1} p(w, V) \right\} \quad \forall V, W \in S$ $p(v, W) \triangleq \inf_{w \in W} \ v - w\  \quad (\text{metric from the point } v \text{ to the subspace } W)$
-------------	--

<sup>20</sup> Deza and Deza (2006), page 235, Akhiezer and Glazman (1993), page 69, Berkson (1963), page 8, Krein and Krasnoselski (1947)

<sup>21</sup> Akhiezer and Glazman (1993), page 70, Berkson (1963), page 8, Krein et al. (1948)

**Definition I.6** (Schäffer's metric). <sup>22</sup>

DEF	$d(V, W) = \log(1 + \max\{r(V, W), r(W, V)\}) \quad \text{where}$ $r(V, W) \triangleq \begin{cases} \inf\{\ A - I\  \mid AV = W\} & \text{if } A \text{ and } A^{-1} \text{ both exist} \\ 1 & \text{otherwise} \end{cases}$
-----	--

## I.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)



<sup>22</sup>📘 Massera and Schäffer (1958), pages 562–563, 📝 Berkson (1963), pages 7–8



## APPENDIX J

### NORMED ALGEBRAS

#### J.1 Algebras

All *linear spaces* (Definition G.1 page 197) 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**.<sup>1</sup>

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.”<sup>2</sup>

**Definition J.1.** <sup>3</sup> Let  $\mathbf{A}$  be an ALGEBRA.

**D E F** An algebra  $\mathbf{A}$  is **unital** if  $\exists u \in \mathbf{A}$  such that  $ux = xu = x \quad \forall x \in \mathbf{A}$

**Definition J.2.** <sup>4</sup> Let  $\mathbf{A}$  be an UNITAL ALGEBRA (Definition J.1 page 253) with unit  $e$ .

**D E F** The **spectrum** of  $x \in \mathbf{A}$  is  $\sigma(x) \triangleq \{\lambda \in \mathbb{C} \mid \lambda e - x \text{ is not invertible}\}$ .  
The **resolvent** of  $x \in \mathbf{A}$  is  $\rho_x(\lambda) \triangleq (\lambda e - x)^{-1} \quad \forall \lambda \notin \sigma(x)$ .  
The **spectral radius** of  $x \in \mathbf{A}$  is  $r(x) \triangleq \sup \{|\lambda| \mid \lambda \in \sigma(x)\}$ .

<sup>1</sup> Fuchs (1995) page 2

<sup>2</sup> Hazewinkel (2000) page v

<sup>3</sup> Folland (1995) page 1

<sup>4</sup> Folland (1995) pages 3–4

## J.2 Star-Algebras

**Definition J.3.** <sup>5</sup> Let  $A$  be an ALGEBRA.

The pair  $(A, *)$  is a  **$*$ -algebra**, or **star-algebra**, if

- D E F
1.  $(x + y)^* = x^* + y^* \quad \forall x, y \in A$  (DISTRIBUTIVE) and
  2.  $(\alpha x)^* = \bar{\alpha} x^* \quad \forall x \in A, \alpha \in \mathbb{C}$  (CONJUGATE LINEAR) and
  3.  $(xy)^* = y^* x^* \quad \forall x, y \in A$  (ANTIAUTOMORPHIC) and
  4.  $x^{**} = x \quad \forall x \in A$  (INVOLUTORY)

The operator  $*$  is called an **involution** on the algebra  $A$ .

**Proposition J.1.** <sup>6</sup> Let  $(A, *)$  be an UNITAL  $*$ -ALGEBRA.

P R P

$$x \text{ is invertible} \implies \begin{cases} 1. \quad x^* \text{ is INVERTIBLE} \quad \forall x \in A \text{ and} \\ 2. \quad (x^*)^{-1} = (x^{-1})^* \quad \forall x \in A \end{cases}$$

PROOF: Let  $e$  be the unit element of  $(A, *)$ .

1. Proof that  $e^* = e$ :

$$\begin{aligned} x e^* &= (x e^*)^{**} && \text{by involutory property of } * && (\text{Definition J.3 page 254}) \\ &= (x^* e^{**})^* && \text{by antiautomorphic property of } * && (\text{Definition J.3 page 254}) \\ &= (x^* e)^* && \text{by involutory property of } * && (\text{Definition J.3 page 254}) \\ &= (x^*)^* && \text{by definition of } e && \\ &= x && \text{by involutory property of } * && (\text{Definition J.3 page 254}) \\ e^* x &= (e^* x)^{**} && \text{by involutory property of } * && (\text{Definition J.3 page 254}) \\ &= (e^{**} x^*)^* && \text{by antiautomorphic property of } * && (\text{Definition J.3 page 254}) \\ &= (e x^*)^* && \text{by involutory property of } * && (\text{Definition J.3 page 254}) \\ &= (x^*)^* && \text{by definition of } e && \\ &= x && \text{by involutory property of } * && (\text{Definition J.3 page 254}) \end{aligned}$$

2. Proof that  $(x^*)^{-1} = (x^{-1})^*$ :

$$\begin{aligned} (x^{-1})^* (x^*) &= [x (x^{-1})]^* && \text{by antiautomorphic and involution properties of } * && (\text{Definition J.3 page 254}) \\ &= e^* \\ &= e && \text{by item (1) page 254} \\ (x^*) (x^{-1})^* &= [x^{-1} x]^* && \text{by antiautomorphic and involution properties of } * && (\text{Definition J.3 page 254}) \\ &= e^* \\ &= e && \text{by item (1) page 254} \end{aligned}$$

**Definition J.4.** <sup>7</sup> Let  $(A, \|\cdot\|)$  be a  $*$ -ALGEBRA (Definition J.3 page 254).

D E F

- An element  $x \in A$  is **hermitian** or **self-adjoint** if  $x^* = x$ .

- An element  $x \in A$  is **normal** if  $xx^* = x^*x$ .

- An element  $x \in A$  is a **projection** if  $xx = x$  (INVOLUTORY) and  $x^* = x$  (HERMITIAN).

<sup>5</sup> Rickart (1960), page 178, Gelfand and Naimark (1964), page 241

<sup>6</sup> Folland (1995) page 5

<sup>7</sup> Rickart (1960), page 178, Gelfand and Naimark (1964), page 242

**Theorem J.1.** <sup>8</sup> Let  $(A, \|\cdot\|)$  be a  $*$ -ALGEBRA (Definition J.3 page 254).

<b>T H M</b>	$\underbrace{x = x^* \text{ and } y = y^*}_{x \text{ and } y \text{ are HERMITIAN}} \implies \begin{cases} x + y = (x + y)^* & (x + y \text{ is selfadjoint}) \\ x^* = (x^*)^* & (x^* \text{ is selfadjoint}) \\ \underbrace{xy = (xy)^*}_{(xy) \text{ is HERMITIAN}} \iff \underbrace{xy = yx}_{\text{commutative}} & \end{cases}$
----------------------	---

PROOF:

$$(x + y)^* = x^* + y^* \quad \begin{matrix} \text{by distributive property of } * \\ \text{by left hypothesis} \end{matrix} \quad (\text{Definition J.3 page 254})$$

$$(x^*)^* = x \quad \begin{matrix} \text{by involutory property of } * \\ \text{by Definition J.3 page 254} \end{matrix}$$

Proof that  $xy = (xy)^* \implies xy = yx$

$$\begin{aligned} xy &= (xy)^* && \text{by left hypothesis} \\ &= y^* x^* && \text{by antiautomorphic property of } * \\ &= yx && \text{by left hypothesis} \end{aligned} \quad (\text{Definition J.3 page 254})$$

Proof that  $xy = (xy)^* \iff xy = yx$

$$\begin{aligned} (xy)^* &= (yx)^* && \text{by left hypothesis} \\ &= x^* y^* && \text{by antiautomorphic property of } * \\ &= xy && \text{by left hypothesis} \end{aligned} \quad (\text{Definition J.3 page 254})$$

**Definition J.5** (Hermitian components). <sup>9</sup> Let  $(A, \|\cdot\|)$  be a  $*$ -ALGEBRA (Definition J.3 page 254).

<b>D E F</b>	$\mathbf{R}_e x \triangleq \frac{1}{2}(x + x^*)$ $\mathbf{I}_m x \triangleq \frac{1}{2i}(x - x^*)$
----------------------	---

**Theorem J.2.** <sup>10</sup> Let  $(A, *)$  be a  $*$ -ALGEBRA (Definition J.3 page 254).

<b>T H M</b>	$\mathbf{R}_e x = (\mathbf{R}_e x)^* \quad \forall x \in A \quad (\mathbf{R}_e x \text{ is HERMITIAN})$ $\mathbf{I}_m x = (\mathbf{I}_m x)^* \quad \forall x \in A \quad (\mathbf{I}_m x \text{ is HERMITIAN})$
----------------------	--

PROOF:

$$\begin{aligned} (\mathbf{R}_e x)^* &= \left(\frac{1}{2}(x + x^*)\right)^* && \text{by definition of } \mathfrak{R} \\ &= \frac{1}{2}(x^* + x^{**}) && \text{by distributive property of } * \\ &= \frac{1}{2}(x^* + x) && \text{by involutory property of } * \\ &= \mathbf{R}_e x && \text{by definition of } \mathfrak{R} \\ (\mathbf{I}_m x)^* &= \left(\frac{1}{2i}(x - x^*)\right)^* && \text{by definition of } \mathfrak{I} \end{aligned} \quad (\text{Definition J.5 page 255})$$

<sup>8</sup> Michel and Herget (1993) page 429

<sup>9</sup> Michel and Herget (1993) page 430, Rickart (1960), page 179, Gelfand and Naimark (1964), page 242

<sup>10</sup> Michel and Herget (1993) page 430, Halmos (1998a) page 42

$$\begin{aligned}
 &= \frac{1}{2i}(x^* - x^{**}) && \text{by } \textit{distributive} \text{ property of } * && (\text{Definition J.3 page 254}) \\
 &= \frac{1}{2i}(x^* - x) && \text{by } \textit{involutory} \text{ property of } * && (\text{Definition J.3 page 254}) \\
 &= \mathbf{I}_m x && \text{by definition of } \mathfrak{I} && (\text{Definition J.5 page 255})
 \end{aligned}$$

⇒

**Theorem J.3** (Hermitian representation). <sup>11</sup> Let  $(A, *)$  be a  $*$ -ALGEBRA (Definition J.3 page 254).

T	H	M	$a = x + iy \iff x = \mathbf{R}_e a \text{ and } y = \mathbf{I}_m a$
---	---	---	--

PROOF:

Proof that  $a = x + iy \implies x = \mathbf{R}_e a$  and  $y = \mathbf{I}_m a$ :

$$\begin{aligned}
 &a = x + iy && \text{by left hypothesis} \\
 \implies &a^* = (x + iy)^* && \text{by definition of } \textit{adjoint} && (\text{Definition J.4 page 254}) \\
 &= x^* - iy && \text{by } \textit{distributive} \text{ property of } * && (\text{Definition J.3 page 254}) \\
 &= x - iy && \text{by Theorem J.2 page 255} \\
 \implies &x = a - iy && \text{by solving for } x \text{ in } a = x + iy \text{ equation} \\
 &x = a^* + iy && \text{by solving for } x \text{ in } a^* = x - iy \text{ equation} \\
 \implies &x + x = a + a^* && \text{by adding previous 2 equations} \\
 \implies &2x = a + a^* && \text{by solving for } x \text{ in previous equation} \\
 \implies &x = \frac{1}{2}(a + a^*) && \\
 &= \mathbf{R}_e a && \text{by definition of } \mathfrak{R} && (\text{Definition J.5 page 255}) \\
 \\
 &iy = a - x && \text{by solving for } iy \text{ in } a = x + iy \text{ equation} \\
 &iy = -a^* + x && \text{by solving for } iy \text{ in } a = x + iy \text{ equation} \\
 \implies &iy + iy = a - a^* && \text{by adding previous 2 equations} \\
 \implies &y = \frac{1}{2i}(a - a^*) && \text{by solving for } iy \text{ in previous equations} \\
 &= \mathbf{I}_m a && \text{by definition of } \mathfrak{I} && (\text{Definition J.5 page 255})
 \end{aligned}$$

Proof that  $a = x + iy \iff x = \mathbf{R}_e a$  and  $y = \mathbf{I}_m a$ :

$$\begin{aligned}
 x + iy &= \mathbf{R}_e a + i \mathbf{I}_m a && \text{by right hypothesis} \\
 &= \underbrace{\frac{1}{2}(a + a^*)}_{\mathbf{R}_e a} + i \underbrace{\frac{1}{2i}(a - a^*)}_{\mathbf{I}_m a} && \text{by definition of } \mathfrak{R} \text{ and } \mathfrak{I} && (\text{Definition J.5 page 255}) \\
 &= \left(\frac{1}{2}a + \frac{1}{2}a\right) + \left(\frac{1}{2}a^* - \frac{1}{2}a^*\right) \xrightarrow{0} 0 \\
 &= a
 \end{aligned}$$

⇒

<sup>11</sup> Michel and Herget (1993) page 430, Rickart (1960), page 179, Gelfand and Neumark (1943b), page 7

## J.3 Normed Algebras

**Definition J.6.** <sup>12</sup> Let  $\mathbf{A}$  be an algebra.

D  
E  
F

The pair  $(\mathbf{A}, \|\cdot\|)$  is a **normed algebra** if

$$\|xy\| \leq \|x\| \|y\| \quad \forall x, y \in \mathbf{A} \quad (\text{multiplicative condition})$$

A normed algebra  $(\mathbf{A}, \|\cdot\|)$  is a **Banach algebra** if  $(\mathbf{A}, \|\cdot\|)$  is also a Banach space.

**Proposition J.2.**

P  
R  
P

$(\mathbf{A}, \|\cdot\|)$  is a normed algebra  $\implies$  multiplication is **continuous** in  $(\mathbf{A}, \|\cdot\|)$

PROOF:

1. Define  $f(x) \triangleq zx$ . That is, the function  $f$  represents multiplication of  $x$  times some arbitrary value  $z$ .
2. Let  $\delta \triangleq \|x - y\|$  and  $\epsilon \triangleq \|f(x) - f(y)\|$ .
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  $\epsilon$  arbitrarily small for some  $\delta > 0$ .
4. And here is the proof that multiplication is indeed continuous in  $(\mathbf{A}, \|\cdot\|)$ :

$$\begin{aligned}
 \|f(x) - f(y)\| &\triangleq \|zx - zy\| && \text{by definition of } f && \text{(item (1) page 257)} \\
 &= \|z(x - y)\| \\
 &\leq \|z\| \|x - y\| && \text{by definition of normed algebra} && \text{(Definition J.6 page 257)} \\
 &\triangleq \|z\| \delta && \text{by definition of } \delta && \text{(item (2) page 257)} \\
 &\leq \epsilon && \text{for some value of } \delta > 0
 \end{aligned}$$



**Theorem J.4** (Gelfand-Mazur Theorem). <sup>13</sup> Let  $\mathbb{C}$  be the field of complex numbers.

T  
H  
M

$(\mathbf{A}, \|\cdot\|)$  is a Banach algebra  
every nonzero  $x \in \mathbf{A}$  is invertible }  $\implies \mathbf{A} \cong \mathbb{C}$  ( $\mathbf{A}$  is isomorphic to  $\mathbb{C}$ )

## J.4 C\* Algebras

**Definition J.7.** <sup>14</sup>

D  
E  
F

The triple  $(\mathbf{A}, \|\cdot\|, *)$  is a **C\* algebra** if

1.  $(\mathbf{A}, \|\cdot\|)$  is a Banach algebra and
2.  $(\mathbf{A}, *)$  is a  $*$ -algebra and
3.  $\|x^* x\| = \|x\|^2 \quad \forall x \in \mathbf{A}$ .

A **C\* algebra**  $(\mathbf{A}, \|\cdot\|, *)$  is also called a **C star algebra**.

<sup>12</sup> Rickart (1960), page 2, Berberian (1961) page 103 (Theorem IV.9.2)

<sup>13</sup> Folland (1995) page 4, Mazur (1938) ((statement)), Gelfand (1941) ((proof))

<sup>14</sup> Folland (1995) page 1, Gelfand and Naimark (1964), page 241, Gelfand and Neumark (1943a), Gelfand and Neumark (1943b)

**Theorem J.5.** <sup>15</sup> Let  $A$  be an algebra.

T  
H  
M

$$(A, \|\cdot\|, *) \text{ is a } C^* \text{ algebra} \quad \Rightarrow \quad \|x^*\| = \|x\|$$

PROOF:

$$\begin{aligned}
 \|x\| &= \frac{1}{\|x\|} \|x\|^2 \\
 &= \frac{1}{\|x\|} \|x^* x\| && \text{by definition of } C^* \text{-algebra} && (\text{Definition J.7 page 257}) \\
 &\leq \frac{1}{\|x\|} \|x^*\| \|x\| && \text{by definition of normed algebra} && (\text{Definition J.6 page 257}) \\
 &= \|x^*\| \\
 \|x^*\| &\leq \|x^{**}\| && \text{by previous result} \\
 &= \|x\| && \text{by involution property of } * && (\text{Definition J.3 page 254})
 \end{aligned}$$

⇒

<sup>15</sup>  Folland (1995) page 1,  Gelfand and Neumark (1943b), page 4,  Gelfand and Neumark (1943a)

# APPENDIX K

## INNER PRODUCT SPACES

### K.1 Definition and basic results

**Definition K.1.** <sup>1</sup> Let  $\Omega \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$  be a LINEAR SPACE (Definition G.1 page 197).

A FUNCTIONAL  $\langle \Delta | \nabla \rangle \in \mathbb{F}^{X \times X}$  is an **inner product** on  $\Omega$  if

- |     |  |                       |     |
|-----|--|-----------------------|-----|
| DEF | 1. $\langle \alpha x   y \rangle = \alpha \langle x   y \rangle$ <span style="float: right;"><math>\forall x, y \in X, \forall \alpha \in \mathbb{C}</math></span> | (HOMOGENEOUS)         | and |
|     | 2. $\langle x + y   u \rangle = \langle x   u \rangle + \langle y   u \rangle$ <span style="float: right;"><math>\forall x, y, u \in X</math></span>               | (ADDITIONAL)          | and |
|     | 3. $\langle x   y \rangle = \langle y   x \rangle^*$ <span style="float: right;"><math>\forall x, y \in X</math></span>  | (CONJUGATE SYMMETRIC) | and |
|     | 4. $\langle x   x \rangle \geq 0$ <span style="float: right;"><math>\forall x \in X</math></span>  | (NON-NEGATIVE)        | and |
|     | 5. $\langle x   x \rangle = 0 \iff x = \emptyset$ <span style="float: right;"><math>\forall x \in X</math></span>  | (NON-ISOTROPIC)       |     |

An inner product is also called a **scalar product**.

The tuple  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \Delta | \nabla \rangle)$  is called an **inner product space**.

**Theorem K.1.** <sup>2</sup> Let  $\Omega \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \Delta | \nabla \rangle)$  be a LINEAR SPACE (Definition G.1 page 197).

- |     |  |
|-----|--|
| THM | 1. $\langle x   y + z \rangle = \langle x   y \rangle + \langle x   z \rangle$ <span style="float: right;"><math>\forall x, y, z \in X</math></span>         |
|     | 2. $\langle x   \alpha y \rangle = \alpha^* \langle x   y \rangle$ <span style="float: right;"><math>\forall x, y \in X, \alpha \in \mathbb{F}</math></span> |
|     | 3. $\langle x   \emptyset \rangle = \langle \emptyset   x \rangle = 0$ <span style="float: right;"><math>\forall x \in X</math></span>                       |
|     | 4. $\langle x - y   z \rangle = \langle x   z \rangle - \langle y   z \rangle$ <span style="float: right;"><math>\forall x, y, z \in X</math></span>         |
|     | 5. $\langle x   y - z \rangle = \langle x   y \rangle - \langle x   z \rangle$ <span style="float: right;"><math>\forall x, y, z \in X</math></span>         |
|     | 6. $\langle x   z \rangle = \langle y   z \rangle$ <span style="float: right;"><math>\forall z \in X \neq \{0\}</math></span> $\iff x = y$                   |
|     | 7. $\langle x   y \rangle = 0$ <span style="float: right;"><math>\forall x \in X</math></span> $\iff y = \emptyset$  |

PROOF:

$$\begin{aligned}
 \langle x | y + z \rangle &= \langle y + z | x \rangle^* && \text{by conjugate symmetric property of } \langle \Delta | \nabla \rangle && \text{(Definition K.1 page 259)} \\
 &= (\langle y | x \rangle + \langle z | x \rangle)^* && \text{by additive property of } \langle \Delta | \nabla \rangle && \text{(Definition K.1 page 259)} \\
 &= \langle y | x \rangle^* + \langle z | x \rangle^* && \text{by distributive property of } * && \text{(Definition J.3 page 254)} \\
 &= \langle x | y \rangle + \langle x | z \rangle && \text{by conjugate symmetric property of } \langle \Delta | \nabla \rangle && \text{(Definition K.1 page 259)} \\
 \langle x | \alpha y \rangle &= \langle \alpha y | x \rangle^* && \text{by conjugate symmetric property of } \langle \Delta | \nabla \rangle && \text{(Definition K.1 page 259)}
 \end{aligned}$$

<sup>1</sup> Istrătescu (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

<sup>2</sup> Berberian (1961) page 27, Haaser and Sullivan (1991) page 277

$= (\alpha \langle y   x \rangle)^*$	by <i>homogeneous</i> property of $\langle \triangle   \nabla \rangle$	(Definition K.1 page 259)
$= \alpha^* \langle y   x \rangle^*$	by <i>antiautomorphic</i> property of $*$	(Definition J.3 page 254)
$= \alpha^* \langle x   y \rangle$	by <i>conjugate symmetric</i> property of $\langle \triangle   \nabla \rangle$	(Definition K.1 page 259)
$\langle x   0 \rangle = \langle 0   x \rangle^*$	by <i>conjugate symmetric</i> property of $\langle \triangle   \nabla \rangle$	(Definition K.1 page 259)
$= \langle 0 \cdot y   x \rangle^*$		
$= (0 \cdot \langle y   x \rangle)^*$	by <i>homogeneous</i> property of $\langle \triangle   \nabla \rangle$	(Definition K.1 page 259)
$= 0$		
$\langle 0   x \rangle = \langle 0 \cdot y   x \rangle$		
$= (0 \cdot \langle y   x \rangle)$	by <i>homogeneous</i> property of $\langle \triangle   \nabla \rangle$	(Definition K.1 page 259)
$= 0$		
$\langle x - y   z \rangle = \langle x + (-y)   z \rangle$	by definition of $+$	
$= \langle x   z \rangle + \langle -y   z \rangle$	by <i>additive</i> property of $\langle \triangle   \nabla \rangle$	(Definition K.1 page 259)
$= \langle x   z \rangle - \langle y   z \rangle$	by <i>homogeneous</i> property of $\langle \triangle   \nabla \rangle$	(Definition K.1 page 259)
$\langle x   y - z \rangle = \langle y - z   x \rangle^*$	by <i>conjugate symmetric</i> property of $\langle \triangle   \nabla \rangle$	(Definition K.1 page 259)
$= (\langle y   x \rangle - \langle z   x \rangle)^*$	by 4.	
$= \langle y   x \rangle^* - \langle z   x \rangle^*$	by <i>distributive</i> property of $*$	(Definition J.3 page 254)
$= \langle x   y \rangle - \langle x   z \rangle$	by <i>conjugate symmetric</i> property of $\langle \triangle   \nabla \rangle$	(Definition K.1 page 259)

$$\begin{aligned} & \langle x | z \rangle = \langle y | z \rangle && \forall z \\ \iff & \langle x | z \rangle - \langle y | z \rangle = 0 && \forall z \quad \text{by property of complex numbers} \\ \iff & \langle x - y | z \rangle = 0 && \forall z \quad \text{by 4.} \\ \iff & x - y = 0 && \forall z \quad \text{by } \textit{non-isotropic} \text{ property of } \langle \triangle | \nabla \rangle \text{ (Definition K.1 page 259)} \end{aligned}$$

Proof that  $\langle x | y \rangle = 0 \implies y = 0$ :

1. Suppose  $y \neq 0$ ;
2. Then  $\langle y | y \rangle \neq 0$  by the *non-isotropic* property of  $\langle \triangle | \nabla \rangle$  (Definition K.1 page 259)
3. But because  $y \in X$ , the left hypothesis implies that  $\langle y | y \rangle = 0$ .
4. This is a *contradiction*.
5. Therefore  $y \neq 0$  must be incorrect and  $y = 0$  must be correct.

Proof that  $\langle x | y \rangle = 0 \iff y = 0$ :

$$\begin{aligned} \langle x | y \rangle &= \langle x | 0 \rangle && \text{by right hypothesis} \\ &= 0 && \text{by Theorem K.1 page 259} \end{aligned}$$

⇒

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 K.2** (Cauchy-Schwarz Inequality). <sup>3</sup> Let  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle | \nabla \rangle)$  be an INNER PRODUCT SPACE and  $|\cdot| \in \mathbb{R}^{\mathbb{C}}$  an ABSOLUTE VALUE function (Definition F.4 page 196). Let  $\|\cdot\|$  be a function in  $\mathbb{R}^{\mathbb{F}}$  such

<sup>3</sup> Haaser and Sullivan (1991) page 278, Aliprantis and Burkinshaw (1998), page 278, Cauchy (1821) page 455, Bunyakovsky (1859) page 6, Schwarz (1885)



that  $\|x\| \triangleq \sqrt{\langle x | x \rangle}$ .<sup>4</sup>

<b>T H M</b>	$ \langle x   y \rangle ^2 \leq \langle x   x \rangle \langle y   y \rangle \quad \forall x, y \in X$ $ \langle x   y \rangle ^2 = \langle x   x \rangle \langle y   y \rangle \iff \exists \alpha \in \mathbb{F} \text{ such that } y = \alpha x \quad \forall x, y \in X$ $ \langle x   y \rangle  \leq \ x\  \ y\  \quad \forall x, y \in X$ $ \langle x   y \rangle  = \ x\  \ y\  \iff \exists \alpha \in \mathbb{F} \text{ such that } y = \alpha x \quad \forall x, y \in X$
----------------------	--

PROOF:

1. Proof that  $|\langle x | y \rangle| \leq \|x\| \|y\|$ :<sup>5</sup>

(a)  $y = \emptyset$  case:

$$\begin{aligned}
 |\langle x | y \rangle|^2 &= |\langle x | \emptyset \rangle|^2 && \text{by } y = \emptyset \text{ hypothesis} \\
 &= |\langle \emptyset | x \rangle|^2 && \text{by Definition K.1 page 259} \\
 &= |\langle \emptyset \emptyset | x \rangle|^2 && \text{by Definition G.1 page 197} \\
 &= |0 \langle \emptyset | x \rangle|^2 && \text{by Definition K.1 page 259} \\
 &= 0 \\
 &= \langle x | x \rangle \langle \emptyset | \emptyset \rangle \\
 &= \langle x | x \rangle \langle y | y \rangle && \text{by } y = \emptyset \text{ hypothesis}
 \end{aligned}$$

(b)  $y \neq \emptyset$  case: Let  $\lambda \triangleq \frac{\langle x | y \rangle}{\langle y | y \rangle}$ .

$$\begin{aligned}
 0 &\leq \langle x - \lambda y | x - \lambda y \rangle && \text{by Definition K.1} \\
 &= \langle x | x - \lambda y \rangle + \langle -\lambda y | x - \lambda y \rangle && \text{by Definition K.1} \\
 &= \langle x - \lambda y | x \rangle^* - \lambda \langle y | x - \lambda y \rangle && \text{by Definition K.1} \\
 &= \langle x | x \rangle^* + \langle -\lambda y | x \rangle^* - \lambda \langle x - \lambda y | y \rangle^* && \text{by Definition K.1} \\
 &= \langle x | x \rangle^* - \lambda^* \langle y | x \rangle^* - \lambda \langle x | y \rangle^* - \lambda \langle -\lambda y | y \rangle^* && \text{by Definition K.1} \\
 &= \langle x | x \rangle - \lambda^* \langle x | y \rangle - \lambda \langle x | y \rangle^* + \lambda \lambda^* \langle y | y \rangle^* && \text{by Definition K.1} \\
 &= \langle x | x \rangle + \left[ \frac{\langle x | y \rangle}{\langle y | y \rangle} \lambda^* \langle y | y \rangle - \lambda^* \langle x | y \rangle \right] - \frac{\langle x | y \rangle}{\langle y | y \rangle} \langle x | y \rangle^* && \text{by definition of } \lambda \\
 &= \langle x | x \rangle - \frac{1}{\langle y | y \rangle} |\langle x | y \rangle|^2 \\
 \implies |\langle x | y \rangle|^2 &\leq \langle x | x \rangle \langle y | y \rangle
 \end{aligned}$$

2. Proof that  $|\langle x | y \rangle|^2 = \langle x | x \rangle \langle y | y \rangle \iff y = ax$ :

Let  $\frac{1}{a} \triangleq \lambda \triangleq \frac{\langle x | y \rangle}{\langle y | y \rangle}$ . Then...

$$\begin{aligned}
 y &= ax \\
 \iff x &= \lambda y \\
 \iff x - \lambda y &= \emptyset \\
 \iff 0 &= \langle x - \lambda y | x - \lambda y \rangle && \text{by Definition K.1 page 259} \\
 &= \langle x | x - \lambda y \rangle + \langle -\lambda y | x - \lambda y \rangle && \text{by Definition K.1 page 259} \\
 &= \langle x - \lambda y | x \rangle^* - \lambda \langle y | x - \lambda y \rangle && \text{by Definition K.1 page 259} \\
 &\vdots && \text{(same steps as in 1(b))}
 \end{aligned}$$

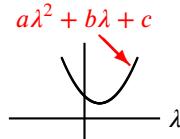
<sup>4</sup>The function  $\|\cdot\|$  is a *norm* (Theorem K.4 page 264) and is called the *norm induced by the inner product*  $\langle \Delta | \nabla \rangle$  (Definition K.2 page 265).

<sup>5</sup> Haaser and Sullivan (1991), page 278

$$\iff |\langle \mathbf{x} | \mathbf{y} \rangle|^2 = \langle \mathbf{x} | \mathbf{x} \rangle \langle \mathbf{y} | \mathbf{y} \rangle - \frac{1}{\langle \mathbf{y} | \mathbf{y} \rangle} |\langle \mathbf{x} | \mathbf{y} \rangle|^2$$

3. Alternate proof for  $|\langle \mathbf{x} | \mathbf{y} \rangle| \leq \|\mathbf{x}\| \|\mathbf{y}\|$ : (Note: This is essentially the same proof as used by Schwarz).<sup>6</sup>

- (a) Proof that  $\{a\lambda^2 + b\lambda + c \geq 0 \quad \forall \lambda \in \mathbb{R}\} \implies \{b^2 \leq 4ac\}$  (quadratic discriminant 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

$$\begin{aligned} 0 &\leq a\lambda^2 + b\lambda + c && \text{by left hypothesis} \\ &= k(\lambda - r_1)(\lambda - r_2) && \text{by definition of } r_1 \text{ and } r_2 \\ &= k(\lambda^2 - r_1\lambda - r_2\lambda + r_1r_2) \\ \implies \lambda^2 - r_1\lambda - r_2\lambda + r_1r_2 &\geq 0 \\ \implies r_1 &= r_2^* && \text{because } r_1r_2 \geq 0 \text{ for } \lambda = 0 \end{aligned}$$

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  $b^2 \leq 4ac$  (the **discriminate** is non-positive).

- (b) Proof that  $\langle \mathbf{y} | \mathbf{y} \rangle \lambda^2 + 2|\langle \mathbf{x} | \mathbf{y} \rangle| \lambda + \langle \mathbf{x} | \mathbf{x} \rangle \geq 0 \quad \forall \lambda \in \mathbb{R}$ :

$$\begin{aligned} 0 &\leq \langle \mathbf{x} + \alpha \mathbf{y} | \mathbf{x} + \alpha \mathbf{y} \rangle && \text{by Definition K.1 page 259} \\ &= \langle \mathbf{x} | \mathbf{x} + \alpha \mathbf{y} \rangle + \langle \alpha \mathbf{y} | \mathbf{x} + \alpha \mathbf{y} \rangle && \text{by Definition K.1 page 259} \\ &= \langle \mathbf{x} | \mathbf{x} + \alpha \mathbf{y} \rangle + \alpha \langle \mathbf{y} | \mathbf{x} + \alpha \mathbf{y} \rangle && \text{by Definition K.1 page 259} \\ &= \langle \mathbf{x} + \alpha \mathbf{y} | \mathbf{x} \rangle^* + \alpha \langle \mathbf{x} + \alpha \mathbf{y} | \mathbf{y} \rangle^* && \text{by Definition K.1 page 259} \\ &= \langle \mathbf{x} | \mathbf{x} \rangle^* + \langle \alpha \mathbf{y} | \mathbf{x} \rangle^* + \alpha \langle \mathbf{x} | \mathbf{y} \rangle^* + \alpha \langle \alpha \mathbf{y} | \mathbf{y} \rangle^* && \text{by Definition K.1 page 259} \\ &= \langle \mathbf{x} | \mathbf{x} \rangle^* + \alpha^* \langle \mathbf{y} | \mathbf{x} \rangle^* + \alpha \langle \mathbf{x} | \mathbf{y} \rangle^* + \alpha \alpha^* \langle \mathbf{y} | \mathbf{y} \rangle^* && \text{by Definition K.1 page 259} \\ &= \langle \mathbf{x} | \mathbf{x} \rangle + \alpha^* \langle \mathbf{x} | \mathbf{y} \rangle + (\alpha^* \langle \mathbf{x} | \mathbf{y} \rangle)^* + |\alpha|^2 \langle \mathbf{y} | \mathbf{y} \rangle && \text{by Definition K.1 page 259} \\ &= \langle \mathbf{x} | \mathbf{x} \rangle + 2\Re(\alpha^* \langle \mathbf{x} | \mathbf{y} \rangle) + |\alpha|^2 \langle \mathbf{y} | \mathbf{y} \rangle && \text{by Definition K.1 page 259} \\ &\leq \langle \mathbf{x} | \mathbf{x} \rangle + 2|\alpha^* \langle \mathbf{x} | \mathbf{y} \rangle| + |\alpha|^2 \langle \mathbf{y} | \mathbf{y} \rangle && \text{by Definition K.1 page 259} \\ &= \langle \mathbf{x} | \mathbf{x} \rangle + 2|\langle \mathbf{x} | \mathbf{y} \rangle||\alpha| + \langle \mathbf{y} | \mathbf{y} \rangle |\alpha|^2 && \text{by Definition K.1 page 259} \\ &= \langle \mathbf{y} | \mathbf{y} \rangle |\alpha|^2 + 2|\langle \mathbf{x} | \mathbf{y} \rangle| |\alpha| + \langle \mathbf{x} | \mathbf{x} \rangle && \text{by Definition K.1 page 259} \\ &= \underbrace{\langle \mathbf{y} | \mathbf{y} \rangle}_{a} \lambda^2 + \underbrace{2|\langle \mathbf{x} | \mathbf{y} \rangle|}_{b} \lambda + \underbrace{\langle \mathbf{x} | \mathbf{x} \rangle}_{c} && \text{because } \lambda \triangleq |\alpha| \in \mathbb{R} \end{aligned}$$

<sup>6</sup> 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).

$$\begin{aligned} \underbrace{\left(2|\langle x | y \rangle|\right)^2}_{b} &\leq 4 \underbrace{\langle y | y \rangle}_{a} \underbrace{\langle x | x \rangle}_{c} \quad \text{by the results of parts (a) and (b)} \\ \implies |\langle x | y \rangle|^2 &\leq \langle x | x \rangle \langle y | y \rangle \end{aligned}$$

4. Proof that  $|\langle x | y \rangle| \leq \|x\| \|y\|$ :

This follows directly from the definition  $\|x\| \triangleq \sqrt{\langle x | x \rangle}$ .

5. Proof that  $|\langle x | y \rangle| = \|x\| \|y\| \iff \exists \alpha \in \mathbb{C} \text{ such that } y = \alpha x$ :

This follows directly from the definition  $\|x\| \triangleq \sqrt{\langle x | x \rangle}$ .



**Corollary K.1.** <sup>7</sup> Let  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle | \nabla \rangle)$  be an INNER PRODUCT SPACE.

**C  
O  
R**

$\langle x | y \rangle$  is CONTINUOUS in both  $x$  and  $y$ .

PROOF: Let  $\|x\| \triangleq \sqrt{\langle x | x \rangle}$ .

$$\begin{aligned} |\langle x + \epsilon | y \rangle - \langle x | y \rangle|^2 &= |\langle x + \epsilon - x | y \rangle|^2 \quad \text{by additivity of } \langle \triangle | \nabla \rangle \quad (\text{Definition K.1 page 259}) \\ &= |\langle \epsilon | y \rangle|^2 \\ &\leq \|\epsilon\|^2 \|y\|^2 \quad \text{by Cauchy-Schwarz Inequality} \quad (\text{Theorem K.2 page 260}) \end{aligned}$$



## K.2 Relationship between norms and inner products

### K.2.1 Norms induced by inner products

**Lemma K.1** (Polar Identity). <sup>8</sup> Let  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle | \nabla \rangle)$  be an INNER PRODUCT SPACE (Definition K.1 page 259). Let  $\Re z$  represent the real part of  $z \in \mathbb{C}$ . Let  $\|\cdot\|$  be a function in  $\mathbb{R}^{\mathbb{F}}$  such that  $\|x\| \triangleq \sqrt{\langle x | x \rangle}$ .<sup>9</sup>

**L  
E  
M**

$$\|x + y\|^2 = \|x\|^2 + 2\Re_e [\langle x | y \rangle] + \|y\|^2 \quad \forall x, y \in X$$

<sup>7</sup> Bollobás (1999) page 132, Aliprantis and Burkinshaw (1998) page 279 (Lemma 32.4)

<sup>8</sup> Conway (1990) page 4, Heil (2011) page 27 (Lemma 1.36(a))

<sup>9</sup>The function  $\|\cdot\|$  is a norm (Theorem K.4 page 264) and is called the norm induced by the inner product  $\langle \triangle | \nabla \rangle$  (Definition K.2 page 265).

PROOF:

$$\begin{aligned}
 \|x + y\|^2 &= \langle x + y | x + y \rangle && \text{by definition of } \textit{induced norm} && (\text{Theorem K.4 page 264}) \\
 &= \langle x | x + y \rangle + \langle y | x + y \rangle && \text{by Definition K.1 page 259} \\
 &= \langle x + y | x \rangle^* + \langle x + y | y \rangle^* && \text{by Definition K.1 page 259} \\
 &= \langle x | x \rangle^* + \langle y | x \rangle^* + \langle x | y \rangle^* + \langle y | y \rangle^* && \text{by Definition K.1 page 259} \\
 &= \langle x | x \rangle + \langle x | y \rangle + \langle x | y \rangle^* + \langle y | y \rangle && \text{by definition of } \textit{inner product} && (\text{Definition K.1 page 259}) \\
 &= \|x\|^2 + 2\Re \langle x | y \rangle + \|y\|^2 && \text{by definition of } \textit{induced norm} && (\text{Theorem K.4 page 264})
 \end{aligned}$$

⇒

**Theorem K.3** (Minkowski's Inequality). <sup>10</sup> Let  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle | \nabla \rangle)$  be an INNER PRODUCT SPACE. Let  $\|\cdot\|$  be a function in  $\mathbb{R}^{\mathbb{F}}$  such that  $\|x\| \triangleq \sqrt{\langle x | x \rangle}$ .<sup>11</sup>

**T H M**  $\|x + y\| \leq \|x\| + \|y\| \quad \forall x, y \in X$

PROOF:

$$\begin{aligned}
 \|x + y\|^2 &= \|x\|^2 + 2\Re \langle x | y \rangle + \|y\|^2 && \text{by } \textit{Polar Identity} && (\text{Lemma K.1 page 263}) \\
 &\leq \|x\|^2 + 2|\langle x | y \rangle| + \|y\|^2 \\
 &\leq \|x\|^2 + 2\sqrt{\langle x | x \rangle}\sqrt{\langle y | y \rangle} + \|y\|^2 && \text{by } \textit{Cauchy-Schwarz Inequality} && (\text{Theorem K.2 page 260}) \\
 &= (\|x\| + \|y\|)^2
 \end{aligned}$$

⇒

**Theorem K.4** (induced norm). <sup>12</sup> Let  $\Omega \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle | \nabla \rangle)$  be an INNER PRODUCT SPACE (Definition K.1 page 259).

**T H M**  $\|x\| \triangleq \sqrt{\langle x | x \rangle} \implies \|\cdot\| \text{ is a NORM}$

PROOF: For a function to be a norm, it must satisfy the four properties listed in Definition L.1 (page 275).

1. Proof that  $\|\cdot\|$  is a norm:

- (a) Proof that  $\|x\| > 0$  for  $x \neq 0$  (non-negative):  
By Definition K.1 page 259, all inner products have this property.
- (b) Proof that  $\|x\| = 0 \iff x = 0$  (non-isometric):  
By Definition K.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 page ??

<sup>10</sup>  Aliprantis and Burkinshaw (1998) pages 278–279 (Theorem 32.3),  Maligranda (1995),  Minkowski (1910) page 115

<sup>11</sup> The function  $\|\cdot\|$  is a *norm* (Theorem K.4 page 264) and is called the *normed induced by the inner product*  $\langle \triangle | \nabla \rangle$  (Definition K.2 page 265).

<sup>12</sup>  Aliprantis and Burkinshaw (1998), pages 278–279,  Haaser and Sullivan (1991) page 278

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 K.4 (previous theorem) demonstrates that in any inner product space  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle | \triangleright \rangle)$ , 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 K.2.** Let  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle | \triangleright \rangle)$  be an INNER PRODUCT SPACE (Definition K.1 page 259).

**D E F** The norm induced by the inner product  $\langle \triangle | \triangleright \rangle$  is defined as

$$\|x\| \triangleq \sqrt{\langle x | x \rangle}$$

## K.2.2 Inner products induced by norms

Theorem K.4 (page 264) demonstrates that if a *linear space* (Definition G.1 page 197) has an *inner product* (Definition K.1 page 259), then that inner product always induces a *norm* (Definition L.1 page 275), and the relationship between the two is simply  $\|x\| = \sqrt{\langle x | x \rangle}$  (Definition K.2 page 265). 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 K.5 page 265).
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 K.6 page 266).
3. Which norms can induce an inner product and which ones cannot? The answer is that norms that satisfy the *parallelogram law* (Theorem K.7 page 267) **can** induce an inner product; and the ones that don't, cannot (Theorem K.7 page 267).

**Theorem K.5.**<sup>13</sup> Let  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$  be a NORMED LINEAR SPACE (Definition L.1 page 275).

**T H M** 
$$\underbrace{\exists \langle \triangle | \triangleright \rangle \text{ and } (\cdot | \cdot) \text{ such that } \|x\|^2 = \langle x | x \rangle = (x | x) \quad \forall x \in X}_{\text{If a norm is induced by two inner products...}} \Rightarrow \underbrace{\langle x | y \rangle = (x | y)}_{\dots \text{then those two inner products are equivalent.}} \quad \forall x, y \in X$$

<sup>13</sup> Aliprantis and Burkinshaw (1998), page 280, Bollobás (1999), page 132, Jordan and von Neumann (1935) page 721

PROOF:

$$\begin{aligned}
 2 \langle x | y \rangle &= [\langle x | y \rangle + \langle y | x \rangle] + [\langle x | y \rangle - \langle y | x \rangle] \\
 &= [\langle x | y \rangle + \langle y | x \rangle] + i [-i \langle x | y \rangle + i \langle y | x \rangle] \\
 &= [\langle x | y \rangle + \langle y | x \rangle] + i [-\langle ix | y \rangle - \langle y | ix \rangle] \\
 &= \left( \underbrace{[\langle x | y \rangle + \langle y | x \rangle + \langle x | x \rangle + \langle y | y \rangle]}_{\langle x+y | x+y \rangle} - \underbrace{[\langle x | x \rangle + \langle y | y \rangle]}_{\text{residue}} \right) \\
 &\quad - i \left( \underbrace{[\langle ix | y \rangle + \langle y | ix \rangle + \langle ix | ix \rangle + \langle y | y \rangle]}_{\langle ix+y | ix+y \rangle} - \underbrace{[\langle ix | ix \rangle + \langle y | y \rangle]}_{\text{residue}} \right) \\
 &= ([\langle x+y | x+y \rangle] - [\langle x | x \rangle + \langle y | y \rangle]) - i ([\langle ix+y | ix+y \rangle] - [\langle ix | ix \rangle + \langle y | y \rangle]) \\
 &= ([\langle x+y | x+y \rangle] - [\langle x | x \rangle + \langle y | y \rangle]) - i ([\langle ix+y | ix+y \rangle] - [\langle ix | ix \rangle + \langle y | y \rangle]) \\
 &= \left( \underbrace{[\langle x | y \rangle + \langle y | x \rangle + \langle x | x \rangle + \langle y | y \rangle]}_{\langle x+y | x+y \rangle} - \underbrace{[\langle x | x \rangle + \langle y | y \rangle]}_{\text{residue}} \right) \\
 &\quad - i \left( \underbrace{[\langle ix | y \rangle + \langle y | ix \rangle + \langle ix | ix \rangle + \langle y | y \rangle]}_{\langle ix+y | ix+y \rangle} - \underbrace{[\langle ix | ix \rangle + \langle y | y \rangle]}_{\text{residue}} \right) \\
 &= [\langle x | y \rangle + \langle y | x \rangle] + i [-\langle ix | y \rangle - \langle y | ix \rangle] \\
 &= [\langle x | y \rangle + \langle y | x \rangle] + i [-i \langle x | y \rangle + i \langle y | x \rangle] \\
 &= [\langle x | y \rangle + \langle y | x \rangle] + [\langle x | y \rangle - \langle y | x \rangle] \\
 &= 2 \langle x | y \rangle
 \end{aligned}$$



**Theorem K.6** (Polarization Identities). <sup>14</sup> Let  $\Omega \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$  be a linear space,  $\langle \Delta | \nabla \rangle \in \mathbb{F}^{X \times X}$  a function, and  $\|x\| \triangleq \sqrt{\langle x | x \rangle}$ .

T  
H  
M

$(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \Delta | \nabla \rangle)$  is an inner product space  $\implies$

$$4 \langle x | y \rangle = \underbrace{\begin{cases} \|x+y\|^2 - \|x-y\|^2 + i \|x+iy\|^2 - i \|x-iy\|^2 & \text{for } \mathbb{F} = \mathbb{C} \quad \forall x, y \in X \\ \|x+y\|^2 - \|x-y\|^2 & \text{for } \mathbb{F} = \mathbb{R} \quad \forall x, y \in X \end{cases}}_{\text{inner product induced by norm}}$$

PROOF:

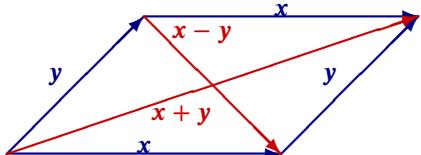
- These follow directly from properties of *bilinear functionals* (Theorem ?? page ??).

<sup>14</sup> Berberian (1961) pages 29–30 (Theorem II.3.3), Istrățescu (1987) page 110 (Proposition 4.1.5), Bollobás (1999), page 132, Jordan and von Neumann (1935) page 721

2. Alternative proof for  $\mathbb{F} = \mathbb{C}$  case:

$$\begin{aligned}
& \|x + y\|^2 - \|x - y\|^2 + i \|x + iy\|^2 - i \|x - iy\|^2 \\
&= \underbrace{\|x\|^2 + \|y\|^2 + 2\Re \langle x | y \rangle}_{\langle x+y | x+y \rangle} - \underbrace{(\|x\|^2 + \|y\|^2 + 2\Re \langle x | -y \rangle)}_{\langle x-y | x-y \rangle} \\
&\quad + i \underbrace{(\|x\|^2 + \|y\|^2 + 2\Re \langle x | iy \rangle)}_{i \langle x+iy | x+iy \rangle} - i \underbrace{(\|x\|^2 + \|y\|^2 + 2\Re \langle x | -iy \rangle)}_{i \langle x-iy | x-iy \rangle} \quad \text{by Lemma K.1 page 263} \\
&= \underbrace{\|x\|^2 + \|y\|^2 + 2\Re \langle x | y \rangle}_{\langle x+y | x+y \rangle} - \underbrace{(\|x\|^2 + \|y\|^2 - 2\Re \langle x | y \rangle)}_{\langle x-y | x-y \rangle} \\
&\quad + i \underbrace{(\|x\|^2 + \|y\|^2 + 2\Re \langle x | iy \rangle)}_{i \langle x+iy | x+iy \rangle} - i \underbrace{(\|x\|^2 + \|y\|^2 - 2\Re \langle x | iy \rangle)}_{i \langle x-iy | x-iy \rangle} \quad \text{by Definition K.1 page 259} \\
&= 4\Re \langle x | y \rangle + 4i\Re \langle x | iy \rangle \\
&= 2 \underbrace{(\langle x | y \rangle + \langle x | y \rangle^*)}_{4\Re \langle x | y \rangle} + 2i \underbrace{(\langle x | iy \rangle + \langle x | iy \rangle^*)}_{4i\Re \langle x | iy \rangle} \\
&= 2 (\langle x | y \rangle + \langle x | y \rangle^*) + 2i (i^* \langle x | y \rangle + (i^{**}) \langle x | y \rangle^*) \\
&= 2 (\langle x | y \rangle + \langle x | y \rangle^*) + 2i (-i \langle x | y \rangle + i \langle x | y \rangle^*) \quad \text{by Definition K.1 page 259} \\
&= 2 \langle x | y \rangle + 2 \langle x | y \rangle^* + 2 \langle x | y \rangle - 2 \langle x | y \rangle^* \\
&= 4 \langle x | y \rangle
\end{aligned}$$

⇒



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.

Actually, 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 K.7** (Parallelogram law). <sup>15</sup> Let  $\Omega \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle | \nabla \rangle)$  and  $\|x\| \triangleq \sqrt{\langle x | x \rangle}$ .

T H M	$\Omega$ is an inner product space	$\iff$	$\underbrace{2\ x\ ^2 + 2\ y\ ^2}_{\text{PARALLELOGRAM LAW / VON NEUMANN-JORDAN CONDITION}} = \ x + y\ ^2 + \ x - y\ ^2$	$\forall x, y \in \Omega$
-------------	------------------------------------	--------	--	---------------------------

PROOF:

1. Proof that  $[\exists \langle x | y \rangle \text{ such that } \|x\|^2 = \langle x | x \rangle] \implies [\text{parallelogram law is true}]$ :

$$\begin{aligned}
\|x + y\|^2 + \|x - y\|^2 &= [\|x\|^2 + \|y\|^2 + 2\Re_e [2\langle x | y \rangle]] + [\|x\|^2 + \|y\|^2 + 2\Re_e [2\langle x | -y \rangle]] \\
&\quad \text{by Lemma K.1 page 263} \\
&= [\|x\|^2 + \|y\|^2 + 2\Re_e [2\langle x | y \rangle]] + [\|x\|^2 + \|y\|^2 - 2\Re_e [2\langle x | y \rangle]] \\
&= 2\|x\|^2 + 2\|y\|^2
\end{aligned}$$

<sup>15</sup> Amir (1986), page 8, Istrățescu (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

2. Proof that  $[\exists \langle x | y \rangle \text{ such that } \|x\|^2 = \langle x | x \rangle] \iff [\text{parallelogram law is true}]$ :

Note that if an inner product exists in the norm linear space  $(\Omega, \|\cdot\|)$ , then that norm linear space is actually an inner product space  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \Delta | \nabla \rangle)$ . And if it is an inner product space, then by Theorem K.6 page 266 that inner product must be given by the **Polarization Identity**

$$\langle x | y \rangle = \|ax + y\|^2 - \|ax - y\|^2 + i \|ax + iy\|^2 - i \|ax - iy\|^2.$$

Therefore, here we must use the parallelogram law to show that the bilinear function  $f(x, y) \triangleq \langle x | y \rangle$  given on the left hand side of the “=” relation is indeed an inner product—that is, that it satisfies the requirements of Definition K.1 page 259.

(a) Proof that  $\langle x | x \rangle \geq 0$  (non-negative):

$$\begin{aligned} 4 \langle x | x \rangle &\triangleq \|x + x\|^2 - \cancel{\|x - x\|^2}^0 + i \|x + ix\|^2 - i \|x - ix\|^2 && \text{by Polarization Identity} \\ &= \|2x\|^2 - 0 + i (\|x + ix\|^2 - \|x - ix\|^2) && \text{by Definition L.1 page 275} \\ &= |2|^2 \|x\|^2 + i (\|x + ix\|^2 - |i| \|x - ix\|^2) \\ &= 4 \|x\|^2 + i (\|x + ix\|^2 - \|ix + x\|^2) && \text{by Definition L.1 page 275} \\ &= 4 \|x\|^2 && \cancel{\|ix + x\|^2}^0 \\ &\geq 0 && \text{by Definition L.1 page 275} \end{aligned}$$

(b) Proof that  $\langle x | x \rangle = 0 \iff x = 0$  (non-isotropic):

$$\begin{aligned} 4 \langle x | x \rangle &= 4 \|x\|^2 && \text{by result of part (a)} \\ &= 0 \iff x = 0 && \text{by Definition L.1 page 275} \end{aligned}$$

(c) Proof that  $\langle x + u | y \rangle = \langle x | y \rangle + \langle u | y \rangle$  (additive):<sup>16</sup>

$$\begin{aligned} 4 \langle x + y | z \rangle &= 8 \left\langle \frac{x+y}{2} | z \right\rangle && \text{by Definition K.1 page 259} \\ &= 2 \left\| \frac{x+y}{2} + z \right\|^2 - 2 \left\| \frac{x+y}{2} - z \right\|^2 \\ &\quad + 2i \left\| \frac{x+y}{2} + z \right\|^2 - 2i \left\| \frac{x+y}{2} - iz \right\|^2 && \text{by Polarization Identity} \\ &= \left( 2 \left\| \frac{x+y}{2} + z \right\|^2 + 2 \left\| \frac{x-y}{2} \right\|^2 \right) \\ &\quad - \left( 2 \left\| \frac{x+y}{2} - z \right\|^2 + 2 \left\| \frac{x-y}{2} \right\|^2 \right) \\ &\quad + \left( 2i \left\| \frac{x+y}{2} + z \right\|^2 + 2i \left\| \frac{x-y}{2} \right\|^2 \right) \\ &\quad - \left( 2i \left\| \frac{x+y}{2} - iz \right\|^2 + 2i \left\| \frac{x-y}{2} \right\|^2 \right) \\ &= (\|x+z\|^2 + \|y+z\|^2) - (\|x-z\|^2 + \|y-z\|^2) \\ &\quad + (i \|x+z\|^2 + i \|y+z\|^2) - (i \|x-iz\|^2 + i \|y-iz\|^2) && \text{by parallelogram law} \\ &= (\|x+z\|^2 - \|x-z\|^2 + i \|x+z\|^2 - i \|x-iz\|^2) \\ &\quad + (\|y+z\|^2 - \|y-z\|^2 + i \|y+z\|^2 - i \|y-iz\|^2) \\ &= 4 \langle x | z \rangle + 4 \langle y | z \rangle && \text{by Polarization Identity} \end{aligned}$$

<sup>16</sup> Aliprantis and Burkinshaw (1998), page 281



(d) Proof that  $\langle \mathbf{x} | \mathbf{y} \rangle = \langle \mathbf{y} | \mathbf{y} \rangle^*$  (*conjugate symmetric*):

$$\begin{aligned}
 4 \langle \mathbf{x} | \mathbf{y} \rangle &\triangleq \|\mathbf{x} + \mathbf{y}\|^2 - \|\mathbf{x} - \mathbf{y}\|^2 + i \|\mathbf{x} + i\mathbf{y}\|^2 - i \|\mathbf{x} - i\mathbf{y}\|^2 && \text{by Polarization Identity} \\
 &= \|\mathbf{y} + \mathbf{x}\|^2 - \|\mathbf{y} - \mathbf{x}\|^2 + i \|i(\mathbf{y} - i\mathbf{x})\|^2 - i \| -i(\mathbf{y} + i\mathbf{x})\|^2 && \text{by Definition G.1 page 197} \\
 &= \|\mathbf{y} + \mathbf{x}\|^2 - \|\mathbf{y} - \mathbf{x}\|^2 + i \|\mathbf{y} - i\mathbf{x}\|^2 - i \|\mathbf{y} + i\mathbf{x}\|^2 && \text{by Definition L.1 page 275} \\
 &= (\|\mathbf{y} + \mathbf{x}\|^2 - \|\mathbf{y} - \mathbf{x}\|^2 - i \|\mathbf{y} - i\mathbf{x}\|^2 + i \|\mathbf{y} + i\mathbf{x}\|^2)^* \\
 &= (\|\mathbf{y} + \mathbf{x}\|^2 - \|\mathbf{y} - \mathbf{x}\|^2 + i \|\mathbf{y} + i\mathbf{x}\|^2 - i \|\mathbf{y} - i\mathbf{x}\|^2)^* \\
 &\triangleq 4 \langle \mathbf{y} | \mathbf{x} \rangle^* && \text{by } \textit{Polarization Identity}
 \end{aligned}$$

(e) Proof that  $\langle \alpha \mathbf{x} | \mathbf{y} \rangle = \alpha \langle \mathbf{x} | \mathbf{y} \rangle$  (*homogeneous*):<sup>17</sup>

i. Proof that  $\langle \alpha \mathbf{x} | \mathbf{y} \rangle$  is linear in  $\alpha$ :

$$\begin{aligned}
 0 &\leq \|\alpha \mathbf{x} + \mathbf{y}\| - \|\beta \mathbf{x} + \mathbf{y}\| && \text{by Definition F.4 page 196} \\
 &\leq \|(\alpha \mathbf{x} + \mathbf{y}) - (\beta \mathbf{x} + \mathbf{y})\| && \text{by Theorem L.2 page 276} \\
 &\leq \|(\alpha - \beta)\mathbf{x}\|
 \end{aligned}$$

This implies that as  $\alpha \rightarrow \beta$ ,  $\|\alpha \mathbf{x} + \mathbf{y}\| \rightarrow \|\beta \mathbf{x} + \mathbf{y}\|$ , which by definition implies that  $\|\alpha \mathbf{x} + \mathbf{y}\|$  linear in  $\alpha$ . And by the parallelogram law,  $\langle \alpha \mathbf{x} | \mathbf{y} \rangle$  is also linear in  $\alpha$ .

ii. Proof that  $\langle n \mathbf{x} | \mathbf{y} \rangle = n \langle \mathbf{x} | \mathbf{y} \rangle$  for  $n \in \mathbb{Z}$  (integer case):

A. Proof for  $n = \pm 1$ :

$$\begin{aligned}
 \langle n \mathbf{x} | \mathbf{y} \rangle &= \langle \pm 1 \mathbf{x} | \mathbf{y} \rangle && \text{by } n = \pm 1 \text{ hypothesis} \\
 &= \pm 1 \langle \mathbf{x} | \mathbf{y} \rangle && \text{by definition of inner product} && \text{(Definition K.1 page 259)} \\
 &= n \langle \mathbf{x} | \mathbf{y} \rangle && \text{by } n = \pm 1 \text{ hypothesis}
 \end{aligned}$$

B. Proof for  $n = 0$ :

$$\begin{aligned}
 \langle n \mathbf{x} | \mathbf{y} \rangle &= \langle 0 \mathbf{x} | \mathbf{y} \rangle && \text{by } n = 0 \text{ hypothesis} \\
 &= \langle \mathbf{x} - \mathbf{x} | \mathbf{y} \rangle \\
 &= \langle \mathbf{x} | \mathbf{y} \rangle + \langle -\mathbf{x} | \mathbf{y} \rangle \\
 &= \langle \mathbf{x} | \mathbf{y} \rangle - 1 \langle \mathbf{x} | \mathbf{y} \rangle \\
 &= \langle \mathbf{x} | \mathbf{y} \rangle - \langle \mathbf{x} | \mathbf{y} \rangle \\
 &= 0 \langle \mathbf{x} | \mathbf{y} \rangle \\
 &= n \langle \mathbf{x} | \mathbf{y} \rangle && \text{by } n = 0 \text{ hypothesis}
 \end{aligned}$$

C. Proof for  $n = \pm 2$ :

$$\begin{aligned}
 \langle n \mathbf{x} | \mathbf{y} \rangle &= \langle \pm 2 \mathbf{x} | \mathbf{y} \rangle && \text{by } n = \pm 1 \text{ hypothesis} \\
 &= \langle \pm(\mathbf{x} + \mathbf{x}) | \mathbf{y} \rangle \\
 &= \pm \langle \mathbf{x} + \mathbf{x} | \mathbf{y} \rangle && \text{by definition of inner product} && \text{(Definition K.1 page 259)} \\
 &= \pm (\langle \mathbf{x} | \mathbf{y} \rangle + \langle \mathbf{x} | \mathbf{y} \rangle) && \text{by additive property} \\
 &= \pm 2 \langle \mathbf{x} | \mathbf{y} \rangle \\
 &= n \langle \mathbf{x} | \mathbf{y} \rangle && \text{by } n = \pm 2 \text{ hypothesis}
 \end{aligned}$$

D. Proof that  $[n \text{ case}] \implies [n \pm 1 \text{ case}]$ :

$$\begin{aligned}
 \langle (n \pm 1) \mathbf{x} | \mathbf{y} \rangle &= \langle n \mathbf{x} \pm 1 \mathbf{x} | \mathbf{y} \rangle \\
 &= \langle n \mathbf{x} | \mathbf{y} \rangle + \langle \pm 1 \mathbf{x} | \mathbf{y} \rangle && \text{by additive property} \\
 &= n \langle \mathbf{x} | \mathbf{y} \rangle \pm 1 \langle \mathbf{x} | \mathbf{y} \rangle && \text{by left hypothesis} \\
 &= (n \pm 1) \langle \mathbf{x} | \mathbf{y} \rangle
 \end{aligned}$$

<sup>17</sup>  Aliprantis and Burkinshaw (1998), page 138

iii. Proof that  $\langle qx | y \rangle = q \langle x | y \rangle$  for  $q \in \mathbb{Q}$  (rational number case):

$$\begin{aligned} \frac{n}{m} \langle x | y \rangle &= \frac{n}{m} \left\langle \frac{m}{m} x | y \right\rangle && \text{where } n, m \in \mathbb{Z} \text{ and } m \neq 0 \\ &= \frac{nm}{m} \left\langle \frac{1}{m} x | y \right\rangle && \text{by previous result} \\ &= \frac{m}{m} \left\langle \frac{n}{m} x | y \right\rangle && \text{by previous result} \\ &= \left\langle \frac{n}{m} x | y \right\rangle \end{aligned}$$

iv. Proof that  $\langle rx | y \rangle = r \langle x | y \rangle$  for all  $r \in \mathbb{R}$  (real number case):

Because  $\mathbb{Q}$  is dense in  $\mathbb{R}$  and because  $\|ax + y\|$  is continuous in  $a$ , so  $\langle ax | y \rangle = a \langle x | y \rangle$  for all  $a \in \mathbb{R}$ .

v. Proof that  $\langle cx | y \rangle = c \langle x | y \rangle$  for all  $c \in \mathbb{C}$  (complex number case):

No proof at this time.



*Remark K.1.* <sup>18</sup> The inner product has already been defined in Definition K.1 (page 259) 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 K.7 page 267) together with the *Polarization Identity* (Theorem K.6 page 266). 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  $\ell_F^p$  norms, only the norm for  $p = 2$  induces an inner product (Proposition K.1, next).

**Proposition K.1.** <sup>19</sup> Let  $\|(x_n)_{n \in \mathbb{Z}}\|_p$  be the  $\ell_F^p$  norm of the sequence  $(x_n)$  in the space  $\ell_F^p$ .

P R P	$\ (x_n)\ _p$ induces an inner product $\iff p = 2$
-------------	---

PROOF:

1. Proof that  $\|\cdot\|_p$  induces an inner product  $\iff p = 2$  (using the *Parallelogram law* page 267):

$$\begin{aligned} &\|x + y\|_p^2 + \|x - y\|_p^2 \\ &= \|x + y\|_2^2 + \|x - y\|_2^2 && \text{by right hypothesis} \\ &= \left( \sum_{n \in \mathbb{Z}} |x_n + y_n|^2 \right)^{\frac{2}{2}} + \left( \sum_{n \in \mathbb{Z}} |x_n - y_n|^2 \right)^{\frac{2}{2}} && \text{by definition of } \|\cdot\|_p \\ &= \sum_{n \in \mathbb{Z}} (x_n + y_n)(x_n + y_n)^* + \sum_{n \in \mathbb{Z}} (x_n - y_n)(x_n - y_n)^* \\ &= \sum_{n \in \mathbb{Z}} \left( |x_n|^2 + |y_n|^2 + 2\Re(x_n y_n) \right) + \sum_{n \in \mathbb{Z}} \left( |x_n|^2 + |y_n|^2 - 2\Re(x_n y_n) \right) \\ &= 2 \sum_{n \in \mathbb{Z}} |x_n|^2 + 2 \sum_{n \in \mathbb{Z}} |y_n|^2 \\ &= 2 \|x\|_2^2 + 2 \|y\|_2^2 && \text{by definition of } \|\cdot\|_p \\ &= 2 \|x\|_p^2 + 2 \|y\|_p^2 && \text{by right hypothesis} \\ &\implies \|\cdot\|_2 \text{ induces an inner product} && \text{by Theorem K.7 page 267} \end{aligned}$$

<sup>18</sup> Loomis (1953), pages 23–24, Kubrusly (2001) page 317

<sup>19</sup> Kubrusly (2001) pages 318–319 (Example 5B)

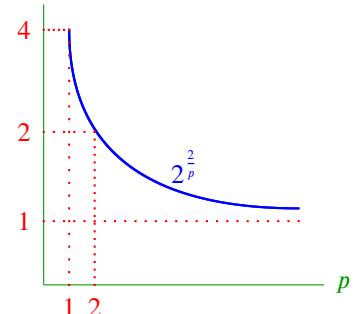
2. Proof that  $\|\cdot\|_p$  induces an inner product  $\implies p = 2$ :

(a) Let  $x \triangleq (1, 0)$  and  $y \triangleq (0, 1)$ . Then <sup>20</sup>

$$\begin{aligned}
 \|x + y\|_p^2 + \|x - y\|_p^2 &= \left( \sum_{n \in \mathbb{Z}} |x_n + y_n|^p \right)^{\frac{2}{p}} + \left( \sum_{n \in \mathbb{Z}} |x_n - y_n|^p \right)^{\frac{2}{p}} && \text{by definition of } \|\cdot\|_p \\
 &= (|1+0|^p + |0+1|^p)^{\frac{2}{p}} + (|1-0|^p + |0-1|^p)^{\frac{2}{p}} && \text{by definitions of } x \text{ and } y \\
 &= 2^{\frac{2}{p}} + 2^{\frac{2}{p}} \\
 &= 2 \cdot 2^{\frac{2}{p}} \\
 2 \|x\|_p^2 + 2 \|y\|_p^2 &= 2 \left( \sum_{n \in \mathbb{Z}} |x_n|^p \right)^{\frac{2}{p}} + 2 \left( \sum_{n \in \mathbb{Z}} |y_n|^p \right)^{\frac{2}{p}} && \text{by definition of } \|\cdot\|_p \\
 &= 2(|1|^p + |0|^p)^{\frac{2}{p}} + 2(|1|^p + |0|^p)^{\frac{2}{p}} && \text{by definitions of } x \text{ and } y \\
 &= 2 + 2 \\
 &= 4 \\
 2 \cdot 2^{\frac{2}{p}} = 4 &\iff 2^{\frac{2}{p}} = 2 \\
 &\implies p = 2
 \end{aligned}$$

(b) Proof that  $2^{2/p}$  is monotonic decreasing in  $p$  (and so  $p = 2$  is the only solution):

$$\begin{aligned}
 \frac{d}{dp} 2^{\frac{2}{p}} &= \frac{d}{dp} e^{\ln 2^{\frac{2}{p}}} \\
 &= \left( e^{\ln 2^{\frac{2}{p}}} \right) \frac{d}{dp} \ln 2^{\frac{2}{p}} \\
 &= \left( 2^{\frac{2}{p}} \right) \frac{d}{dp} (2 \ln 2) \frac{1}{p} \\
 &= \left( 2^{\frac{2}{p}} \right) 2 \ln 2 \left( -\frac{1}{p^2} \right) \\
 &< 0 \quad \forall p \in (0, \infty)
 \end{aligned}$$



## K.3 Orthogonality

### Definition K.3.

**D E F** The **Kronecker delta function**  $\bar{\delta}_n$  is defined as  $\bar{\delta}_n \triangleq \begin{cases} 1 & \text{for } n = 0 \\ 0 & \text{for } n \neq 0 \end{cases}$   $\forall n \in \mathbb{Z}$

**Definition K.4.** Let  $(X, +, \cdot, (\mathbb{F}, +, \times), \langle \triangle | \triangleright \rangle)$  be an INNER PRODUCT SPACE (Definition K.1 page 259).

<sup>20</sup> <http://groups.google.com/group/sci.math/msg/531b1173f08871e9>

Two vectors  $x$  and  $y$  in  $X$  are **orthogonal** if

$$\langle x | y \rangle = \begin{cases} 0 & \text{for } x \neq y \\ c \in \mathbb{F} \setminus 0 & \text{for } x = y \end{cases}$$

The notation  $x \perp y$  implies  $x$  and  $y$  are **ORTHOGONAL**.

A set  $Y \in 2^X$  is **orthogonal** if  $x \perp y \quad \forall x, y \in Y$ .

A set  $Y$  is **orthonormal** if it is ORTHOGONAL and  $\langle y | y \rangle = 1 \quad \forall y \in Y$ .

A sequence  $(x_n \in X)_{n \in \mathbb{Z}}$  is **orthogonal** if  $\langle x_n | x_m \rangle = c \delta_{nm}$  for some  $c \in \mathbb{R} \setminus 0$ .

A sequence  $(x_n \in X)_{n \in \mathbb{Z}}$  is **orthonormal** if  $\langle x_n | x_m \rangle = \delta_{nm}$ .

The definition of the orthogonality relation  $\perp$  has several immediate consequences (next theorem):

**Theorem K.8.** <sup>21</sup> Let  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle | \triangledown \rangle)$  be an INNER PRODUCT SPACE.

- |                      |  |
|----------------------|--|
| <b>T<br/>H<br/>M</b> | <ol style="list-style-type: none"> <li>1. <math>x \perp x \iff x = 0 \quad \forall x \in X</math></li> <li>2. <math>x \perp y \implies \alpha x \perp y \quad \forall \alpha \in \mathbb{R}, x, y \in X \quad (\text{HOMOGENEOUS})</math></li> <li>3. <math>x \perp y \iff y \perp x \quad \forall x, y \in X \quad (\text{SYMMETRY})</math></li> <li>4. <math>x \perp y \text{ and } y \perp z \implies x \perp (y + z) \quad \forall x, y, z \in X \quad (\text{ADDITIVE})</math></li> <li>5. <math>\exists \beta \in \mathbb{R} \text{ such that } x \perp (\beta x + y) \quad \forall x \in X \setminus 0, y \in X</math></li> </ol> |
|----------------------|--|

**Theorem K.9.** Let  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle | \triangledown \rangle)$  be an INNER PRODUCT SPACE.

- |                      |  |
|----------------------|--|
| <b>T<br/>H<br/>M</b> | <ol style="list-style-type: none"> <li>1. <math>\langle x   y \rangle = 0 \text{ and }</math></li> <li>2. <math>x + y = 0</math> </li> </ol> $\iff \left\{ \begin{array}{l} 1. \quad x = 0 \text{ and} \\ 2. \quad y = 0 \end{array} \right. \quad \forall x, y \in X$ |
|----------------------|--|

PROOF:

1. Proof that  $x = y = 0$ :

$$\begin{aligned}
 0 &= \langle 0 | 0 \rangle && \text{by non-isotropic property of } \langle \triangle | \triangledown \rangle \text{ (Definition K.1 page 259)} \\
 &= \langle x + y | x + y \rangle && \text{by left hypothesis 2} \\
 &= \langle x | x + y \rangle + \langle y | x + y \rangle && \text{by additive property of } \langle \triangle | \triangledown \rangle \text{ (Definition K.1 page 259)} \\
 &= \langle x | x \rangle + \langle x | y \rangle + \langle x | y \rangle^* + \langle y | y \rangle && \text{by conjugate symmetric and additive properties of } \langle \triangle | \triangledown \rangle \\
 &= \underbrace{\langle x | x \rangle}_{\geq 0} + 0 + 0 + \underbrace{\langle y | y \rangle}_{\geq 0} && \text{by left hypothesis 1} \\
 &\implies x = 0 \text{ and } y = 0 && \text{by non-negative and non-isotropic properties of } \langle \triangle | \triangledown \rangle
 \end{aligned}$$

2. Proof that  $\langle x | y \rangle = 0$ :

$$\begin{aligned}
 \langle x | y \rangle &= \langle 0 | 0 \rangle && \text{by right hypotheses} \\
 &= 0 && \text{by non-isotropic property of } \langle \triangle | \triangledown \rangle \text{ (Definition K.1 page 259)}
 \end{aligned}$$

3. Proof that  $x + y = 0$ :

$$\begin{aligned}
 x + y &= 0 + 0 && \text{by right hypotheses} \\
 &= 0
 \end{aligned}$$

⇒

The *triangle inequality* for vectors in a normed linear space (Theorem L.1 page 275) demonstrates that

$\left\| \sum_{n=1}^N x_n \right\| \leq \sum_{n=1}^N \|x_n\|$ . The *Pythagorean Theorem* (next) demonstrates that this *inequality* becomes *equality* when the set  $\{x_n\}$  is *orthogonal*.

<sup>21</sup> James (1945), page 292, Drljević (1989) page 232



**Theorem K.10** (Pythagorean Theorem). <sup>22</sup> Let  $\{x_n \in X \mid n=1,2,\dots,N\}$  be a set of vectors in an INNER PRODUCT SPACE (Definition K.1 page 259)  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle | \nabla \rangle)$  and let  $\|x\| \triangleq \sqrt{\langle x | x \rangle}$  (Definition K.2 page 265).

T  
H  
M

$$\{x_n\} \text{ is ORTHOGONAL} \iff \left\| \sum_{n=1}^N x_n \right\|^2 = \sum_{n=1}^N \|x_n\|^2 \quad \forall N \in \mathbb{N}$$

PROOF: 1. Proof for ( $\implies$ ) case:

$$\begin{aligned} \left\| \sum_{n=1}^N x_n \right\|^2 &= \left\langle \sum_{n=1}^N x_n \mid \sum_{m=1}^N x_m \right\rangle && \text{by def. of } \|\cdot\| && \text{(Definition L.1 page 275)} \\ &= \sum_{n=1}^N \sum_{m=1}^N \langle x_n \mid x_m \rangle && \text{by def. of } \langle \triangle | \nabla \rangle && \text{(Definition K.1 page 259)} \\ &= \sum_{n=1}^N \sum_{m=1}^N \langle x_n \mid x_m \rangle \bar{\delta}_{n-m} && \text{by left hypothesis} \\ &= \sum_{n=1}^N \langle x_n \mid x_n \rangle && \text{by def. of } \bar{\delta} && \text{(Definition K.3 page 271)} \\ &= \sum_{n=1}^N \|x_n\|^2 && \text{by def. of } \|\cdot\| && \text{(Definition L.1 page 275)} \end{aligned}$$

2. Proof for ( $\impliedby$ ) case:

$$\begin{aligned} 4 \langle x \mid y \rangle &= \|x + y\|^2 - \|x - y\|^2 + i \|x + iy\|^2 - i \|x - iy\|^2 && \text{by polarization identity (Theorem K.6 page 266)} \\ &= (\|x\|^2 + \|y\|^2) - (\|x\|^2 + \| -y \|^2) + i (\|x\|^2 + \|iy\|^2) - i (\|x\|^2 + \| -iy \|^2) && \text{by right hypothesis} \\ &= (\|x\|^2 + \|y\|^2) - (\|x\|^2 + |-1|^2 \|y\|^2) + i (\|x\|^2 + |i|^2 \|y\|^2) - i (\|x\|^2 + |-i|^2 \|y\|^2) && \text{by definition of } \|\cdot\| \\ &= (\|x\|^2 + \|y\|^2) - (\|x\|^2 + \|y\|^2) + i (\|x\|^2 + \|y\|^2) - i (\|x\|^2 + \|y\|^2) && \text{by def. of } |\cdot| \text{ (Definition F.4 page 196)} \\ &= 0 \end{aligned}$$



<sup>22</sup> Aliprantis and Burkinshaw (1998) pages 282–283 (Theorem 32.7), Kubrusly (2001) page 324 (Proposition 5.8), Bollobás (1999) pages 132–133 (Theorem 3)



## APPENDIX L

### NORMED LINEAR SPACES

#### L.1 Definition and basic results

**Definition L.1.** <sup>1</sup> Let  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$  be a LINEAR SPACE (Definition G.1 page 197) and  $|\cdot| \in \mathbb{R}^{\mathbb{F}}$  the ABSOLUTE VALUE function (Definition F.4 page 196).

A functional  $\|\cdot\|$  in  $\mathbb{R}^X$  is a **norm** if

- |     |                                    |  |                                    |     |
|-----|------------------------------------|--|------------------------------------|-----|
| DEF | 1. $\ x\  \geq 0$                  | $\forall x \in X$                        | (STRICTLY POSITIVE)                | and |
|     | 2. $\ x\  = 0 \iff x = 0$          | $\forall x \in X$                        | (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 INEQUALITY). |     |

A normed linear space is the tuple  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$ .

The definition of the *norm* (Definition L.1 page 275) requires that any two vectors in a norm space be *subadditive* (they satisfy the *triangle inequality* property) such that  $\|x_1 + x_2\| \leq \|x_1\| + \|x_2\|$ . Actually, in **any** normed linear space, this property holds true for **any** finite number of vectors—not just two—such that  $\|x_1 + x_2 + \dots + x_N\| \leq \|x_1\| + \|x_2\| + \dots + \|x_N\|$  (next theorem).

**Theorem L.1** (triangle inequality). <sup>2</sup> Let  $(x_n \in X)_1^N$  be an N-TUPLE (Definition R.1 page 357) of vectors in a NORMED LINEAR SPACE  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$ .

THM	$\left\  \sum_{n=1}^N x_n \right\  \leq \sum_{n=1}^N \ x_n\  \quad \forall N \in \mathbb{N}, x_n \in V$
-----	---

PROOF: Proof is by induction:

<sup>1</sup> Aliprantis and Burkinshaw (1998), pages 217–218, Banach (1932a), page 53, Banach (1932b), page 33, Banach (1922) page 135

<sup>2</sup> Michel and Herget (1993), page 344, Euclid (circa 300BC) (Book I Proposition 20)

1. Proof for the  $N = 1$  case:

$$\begin{aligned} \left\| \sum_{n=1}^1 \mathbf{x}_n \right\| &= \|\mathbf{x}_1\| \\ &= \sum_{n=1}^1 \|\mathbf{x}_1\| \end{aligned}$$

2. Proof for the  $N = 2$  case:

$$\begin{aligned} \left\| \sum_{n=1}^2 \mathbf{x}_n \right\| &= \left\| \sum_{n=1}^2 \mathbf{x}_n \right\| \\ &= \|\mathbf{x}_1 + \mathbf{x}_2\| \\ &\leq \|\mathbf{x}_1\| + \|\mathbf{x}_2\| && \text{by Definition L.1 page 275 (triangle inequality)} \\ &= \sum_{n=1}^2 \|\mathbf{x}_n\| \end{aligned}$$

3. Proof that [ $N$  case]  $\implies$  [ $N + 1$  case]:

$$\begin{aligned} \left\| \sum_{n=1}^{N+1} \mathbf{x}_n \right\| &= \left\| \sum_{n=1}^N \mathbf{x}_n + \mathbf{x}_{N+1} \right\| \\ &\leq \left\| \sum_{n=1}^N \mathbf{x}_n \right\| + \|\mathbf{x}_{N+1}\| && \text{by Definition L.1 page 275 (triangle inequality)} \\ &\leq \sum_{n=1}^N \|\mathbf{x}_n\| + \|\mathbf{x}_{N+1}\| && \text{by left hypothesis} \\ &= \sum_{n=1}^{N+1} \|\mathbf{x}_n\| \end{aligned}$$



**Theorem L.2** (Reverse Triangle Inequality). <sup>3</sup> Let  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$  be a NORMED LINEAR SPACE (Definition L.1 page 275).

T	$\underbrace{\ \mathbf{x}\  - \ \mathbf{y}\  \leq \ \mathbf{x} - \mathbf{y}\ }_{\text{REVERSE TRIANGLE INEQUALITY}} \leq \ \mathbf{x}\  + \ \mathbf{y}\  \quad \forall \mathbf{x}, \mathbf{y} \in X$	■
---	--	---

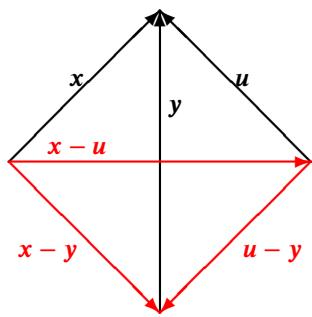
PROOF:

$$\begin{aligned} \|\mathbf{x}\| - \|\mathbf{y}\| &= \|(\mathbf{x} - \mathbf{y}) + \mathbf{y}\| - \|\mathbf{y}\| \\ &\leq \|\mathbf{x} - \mathbf{y}\| + \|\mathbf{y}\| - \|\mathbf{y}\| && \text{by Definition L.1 page 275} \\ &= \|\mathbf{x} - \mathbf{y}\| \\ &= \|\mathbf{x} - \mathbf{y}\| && \text{by Definition L.1 page 275} \end{aligned}$$

$$\begin{aligned} \|\mathbf{x} - \mathbf{y}\| &\leq \|\mathbf{x} - \mathbf{0}\| + \|\mathbf{0} - \mathbf{y}\| \\ &= \|\mathbf{x}\| + |-1| \|\mathbf{y}\| && \text{by previous result with } u = 0 \\ &= \|\mathbf{x}\| + \|\mathbf{y}\| && \text{by Definition L.1 page 275} \end{aligned}$$



<sup>3</sup> 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 L.1.** <sup>4</sup> Let  $(X, +, \cdot, (\mathbb{F}, +, \dot{\times}), \|\cdot\|)$  be a NORMED LINEAR SPACE (Definition L.1 page 275).

**P** **R** **P** 
$$\|x - y\| \leq \|x - u\| + \|u - y\| \quad \forall x, y, u \in X$$

PROOF:

$$\begin{aligned} \|x - y\| &= \|(x - u) + (u - y)\| \\ &\leq \|x - u\| + \|u - y\| \end{aligned} \quad \text{by Definition L.1 page 275}$$

*Example L.1 (The usual norm).* <sup>5</sup> Let  $\mathbb{R}^\mathbb{R}$  be the set of all functions with domain and range the set of *real numbers*  $\mathbb{R}$ .

**E** **X** The absolute value (Definition F.4 page 196)  $|\cdot| \in \mathbb{R}^\mathbb{R}$  is a norm.

*Example L.2 ( $l_p$  norms).* Let  $(x_n)_{n \in \mathbb{Z}}$  be a sequence (Definition R.1 page 357) of real numbers. An uncountably infinite number of norms is provided by the  $\ell_p^{\mathbb{F}}$  norms  $\|(x_n)\|_p$ :

**E** **X** 
$$\|(x_n)\|_p \triangleq \left( \sum_{n \in \mathbb{Z}} |x_n|^p \right)^{\frac{1}{p}}$$
 is a norm for  $p \in [1 : \infty]$

## L.2 Relationship between metrics and norms

### L.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,

$$\text{normed linear space} \implies \text{metric space} \implies \text{topological space}.$$

<sup>4</sup> Aliprantis and Burkinshaw (1998), page 218

<sup>5</sup> Giles (1987), page 3

**Theorem L.3.** <sup>6</sup> Let  $d \in \mathbb{R}^{X \times X}$  be a function on a REAL normed linear space  $(X, +, \cdot, (\mathbb{R}, \dot{+}, \dot{\times}), \|\cdot\|)$ . Let  $B(x, r) \triangleq \{y \in X \mid \|y - x\| < r\}$  be the OPEN BALL of radius  $r$  centered at a point  $x$ .

**T H M**  $d(x, y) \triangleq \|x - y\|$  is a metric on  $X$

PROOF: The proof follows directly from the definition of a metric (not included in this text) the definition of *norm* (Definition L.1 page 275).  $\Rightarrow$

The previous theorem defined a metric  $d(x, y)$  induced by the norm  $\|x\|$ . The next definition defines this metric formally.

**Definition L.2.** <sup>7</sup> Let  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$  be a NORMED LINEAR SPACE (Definition L.1 page 275).

**D E F** The metric induced by the norm  $\|\cdot\|$  is the function  $d \in \mathbb{R}^X$  such that  

$$d(x, y) \triangleq \|x - y\| \quad \forall x, y \in X$$

Due to its algebraic structure, every norm is *continuous* (Corollary L.1 page 278). This makes norm spaces very useful in analysis. For a function  $f$  to be *continuous*, for every  $\epsilon > 0$  there must exist a  $\delta > 0$  such that  $|f(x + \delta) - f(x)| < \epsilon$ . The *Reverse Triangle Inequality* (Theorem L.2 page 276) shows this to be true when  $f(\cdot) \triangleq \|\cdot\|$ .

**Corollary L.1.** <sup>8</sup> Let  $\Omega \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$  be a NORMED LINEAR SPACE (Definition L.1 page 275).

**C O R** The norm  $\|\cdot\|$  is CONTINUOUS in  $\Omega$ .

PROOF: This follows from these concepts:

1. The fact that  $d(x, y) \triangleq \|x - y\|$  is a *metric* (Theorem L.3 page 278).
2. *Continuity* in a metric space.
3. The *Reverse Triangle Inequality* (Theorem L.2 page 276).

Theorem L.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 L.4.** <sup>9</sup> Let  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), d)$  be a METRIC LINEAR SPACE (Definition ?? page ??). Let  $B$  be the OPEN BALL  $B(p, r) \triangleq \{x \in X \mid d(p, x) < r\}$  (open ball with respect to metric  $d$  centered at point  $p$  and with radius  $r$ ).

**T H M** 
$$\left. \begin{array}{l} \exists \|\cdot\| \in \mathbb{R}^X \text{ such that} \\ d(x, y) = \|y - x\| \\ d \text{ is generated by a norm} \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} 1. \quad B(x, r) = x + B(0, r) \\ 2. \quad B(0, r) = r B(0, 1) \\ 3. \quad B(x, r) \text{ is CONVEX} \\ 4. \quad x \in B(0, r) \iff -x \in B(0, r) \quad (\text{SYMMETRIC}) \end{array} \right.$$

<sup>6</sup> Michel and Herget (1993), page 344, Banach (1932a) page 53

<sup>7</sup> Giles (2000) page 1 (1.1 Definition)

<sup>8</sup> Giles (2000) page 2

<sup>9</sup> Giles (2000) page 2 (1.2 Remarks), Giles (1987) pages 22–26 (2.4 Theorem, 2.11 Theorem)

PROOF:

1. Proof that  $d(x + z, y + vz) = d(x, y)$  (invariant):

$$\begin{aligned} d(x + z, y + vz) &= \|(y + vz) - (x + z)\| && \text{by left hypothesis} \\ &= \|y - x\| \\ &= d(x, y) && \text{by left hypothesis} \end{aligned}$$

2. Proof that  $B(x, r) = x + B(0, r)$ :

$$\begin{aligned} B(x, r) &= \{y \in X | d(x, y) < r\} && \text{by definition of open ball } B \\ &= \{y \in X | d(y - x, y - x) < r\} && \text{by right result 1.} \\ &= \{y \in X | d(0, y - x) < r\} \\ &= \{u + x \in X | d(0, u) < r\} && \text{let } u \triangleq y - x \\ &= x + \{u \in X | d(0, u) < r\} \\ &= x + B(0, r) && \text{by definition of open ball } B \end{aligned}$$

3. Proof that  $B(0, r) = r B(0, 1)$ :

$$\begin{aligned} B(0, r) &= \{y \in X | d(0, y) < r\} && \text{by definition of open ball } B \\ &= \left\{ y \in X | \frac{1}{r} d(0, y) < 1 \right\} \\ &= \left\{ y \in X | \frac{1}{r} \|y - 0\| < 1 \right\} && \text{by left hypothesis} \\ &= \left\{ y \in X | \left\| \frac{1}{r} y - \frac{1}{r} 0 \right\| < 1 \right\} && \text{by homogeneous property of } \|\cdot\| \text{ page 275} \\ &= \left\{ y \in X | d\left(\frac{1}{r} 0, \frac{1}{r} y\right) < 1 \right\} && \text{by left hypothesis} \\ &= \{ru \in X | d(0, u) < 1\} && \text{let } u \triangleq \frac{1}{r} y \\ &= r \{u \in X | d(0, u) < 1\} \\ &= r B(0, 1) && \text{by definition of open ball } B \end{aligned}$$

4. Proof that  $B(p, r)$  is convex:

We must prove that for any pair of points  $x$  and  $y$  in the open ball  $B(p, r)$ , any point  $\lambda x + (1 - \lambda)y$  is also in the open ball. That is, the distance from any point  $\lambda x + (1 - \lambda)y$  to the ball's center  $p$  must be less than  $r$ .

$$\begin{aligned} d(p, \lambda x + (1 - \lambda)y) &= \|p - \lambda x - (1 - \lambda)y\| && \text{by left hypothesis} \\ &= \left\| \underbrace{\lambda p + (1 - \lambda)p - \lambda x - (1 - \lambda)y}_{p} \right\| \\ &= \|\lambda p - \lambda x + (1 - \lambda)p - (1 - \lambda)y\| \\ &\leq \|\lambda p - \lambda x\| + \|(1 - \lambda)p - (1 - \lambda)y\| && \text{by subadditivity property of } \|\cdot\| \text{ page 275} \\ &= |\lambda| \|p - x\| + |1 - \lambda| \|p - y\| && \text{by homogeneous property of } \|\cdot\| \text{ page 275} \\ &= \lambda \|p - x\| + (1 - \lambda) \|p - y\| && \text{because } 0 \leq \lambda \leq 1 \\ &\leq \lambda r + (1 - \lambda)r && \text{because } x, y \text{ are in the ball } B(p, r) \\ &= r \end{aligned}$$

5. Proof that  $x \in B(\mathbf{0}, r) \iff -x \in B(\mathbf{0}, r)$  (symmetric):

$$\begin{aligned}
 x \in B(\mathbf{0}, r) &\iff x \in \{y \in X \mid d(\mathbf{0}, y) < r\} && \text{by definition of open ball } B \\
 &\iff x \in \{y \in X \mid \|y - \mathbf{0}\| < r\} && \text{by left hypothesis} \\
 &\iff x \in \{y \in X \mid \|y\| < r\} \\
 &\iff x \in \{y \in X \mid \|(-1)(-y)\| < r\} \\
 &\iff x \in \{y \in X \mid \| -1 \| \| -y \| < r\} && \text{by homogeneous property of } \|\cdot\| \text{ page 275} \\
 &\iff x \in \{y \in X \mid \| -y - \mathbf{0} \| < r\} \\
 &\iff x \in \{y \in X \mid d(\mathbf{0}, -y) < r\} && \text{by left hypothesis} \\
 &\iff x \in \{-u \in X \mid d(\mathbf{0}, u) < r\} && \text{let } u \triangleq -y \\
 &\iff x \in (-\{u \in X \mid d(\mathbf{0}, u) < r\}) \\
 &\iff x \in (-B(\mathbf{0}, r)) \\
 &\iff -x \in B(\mathbf{0}, r)
 \end{aligned}$$

⇒

Theorem L.4 (page 278) demonstrates that if a metric  $d$  in a metric space  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), d)$  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  $(Y, p)$  to be *convex*, but yet the metric  $p$  not be generated by a norm.

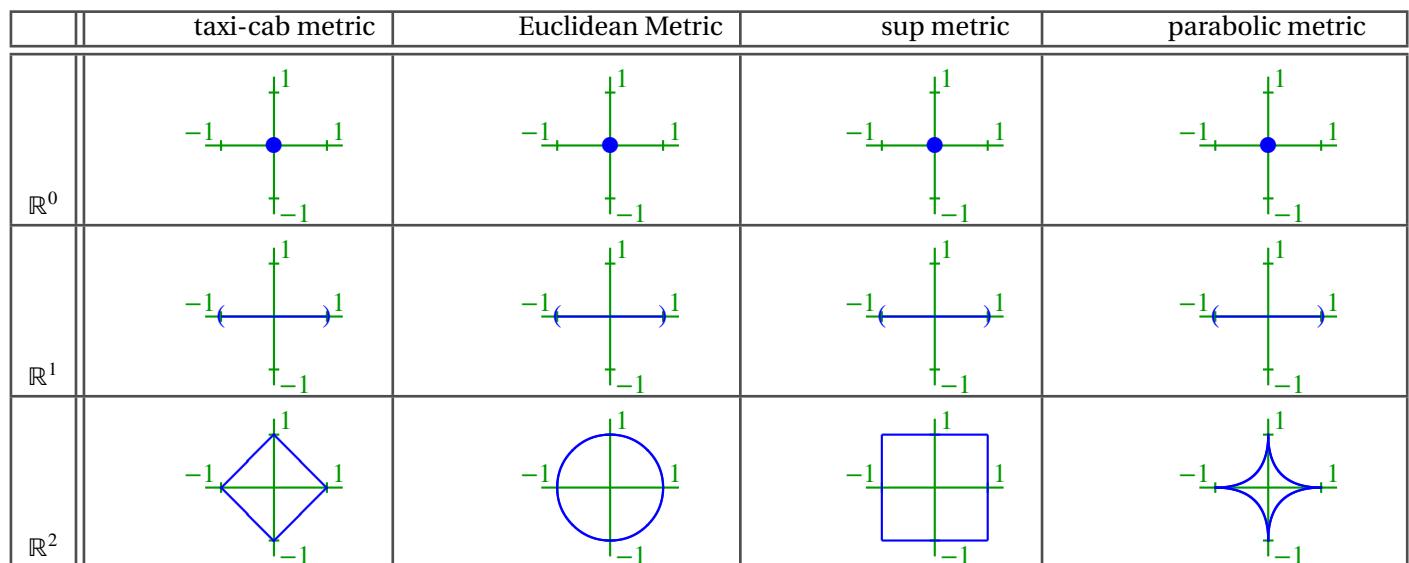


Figure L.1: Open balls in  $(\mathbb{R}^0, d_n)$ ,  $(\mathbb{R}, d_n)$ ,  $(\mathbb{R}^2, d_n)$ , and  $(\mathbb{R}^3, d_n)$ .

## L.2.2 Norms generated by metrics

Every normed linear space is also a metric linear space (Theorem L.3 page 278). 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 L.5 page 281.

**Lemma L.1.**<sup>10</sup> Let  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), d)$  be a METRIC LINEAR SPACE. Let  $\|x\| \triangleq d(x, \mathbf{0}) \forall x \in X$ .

<sup>10</sup> Oikhberg and Rosenthal (2007), page 599



**L E M**

$$\underbrace{d(x+z, y+z) = d(x, y) \quad \forall x, y, z \in X}_{\text{TRANSLATION INVARIANT}} \implies \begin{cases} 1. \quad \|x\| = \|-x\| & \forall x \in X \quad \text{and} \\ 2. \quad \|x\| = 0 \iff x = 0 & \forall x \in X \quad \text{and} \\ 3. \quad \|x+y\| \leq \|x\| + \|y\| & \forall x, y \in X \end{cases}$$

PROOF:

1. Proof that  $\|x\| = \|-x\|$ :

$$\begin{aligned} \|x\| &= d(x, 0) && \text{by definition of } \|\cdot\| \\ &= d(x - x, 0 - x) && \text{by translation invariance hypothesis} \\ &= d(0, -x) \\ &= \|-x\| && \text{by definition of } \|\cdot\| \end{aligned}$$

2a. Proof that  $\|x\| = 0 \implies x = 0$ :

$$\begin{aligned} 0 &= \|x\| && \text{by left hypothesis} \\ &= d(x, 0) && \text{by definition of } \|\cdot\| \\ &= d(x, 0) && \text{by definition of } \|\cdot\| \\ &\implies x = 0 && \text{by property of metrics} \end{aligned}$$

2b. Proof that  $\|x\| = 0 \iff x = 0$ :

$$\begin{aligned} \|x\| &= d(x, 0) && \text{by definition of } \|\cdot\| \\ &= d(0, 0) && \text{by right hypothesis} \\ &= 0 && \text{by property of metrics} \end{aligned}$$

3. Proof that  $\|x+y\| \leq \|x\| + \|y\|$ :

$$\begin{aligned} \|x+y\| &= d(x+y, 0) && \text{by definition of } \|\cdot\| \\ &= d(x+y - y, 0 - y) && \text{by translation invariance hypothesis} \\ &= d(x, -y) \\ &\leq d(x, 0) + d(0, y) && \text{by property of metrics} \\ &= d(x, 0) + d(y, 0) && \text{by property of metrics} \\ &= \|x\| + \|y\| && \text{by definition of } \|\cdot\| \end{aligned}$$



**Theorem L.5.** <sup>11</sup> Let  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$  be a LINEAR SPACE. Let  $d(x, y) \triangleq \|x - y\| \forall x, y \in X$ .

**T H M**

$$\left. \begin{array}{l} 1. \quad d(x+z, y+z) = d(x, y) \quad \forall x, y, z \in X \quad (\text{TRANSLATION INVARIANT}) \\ 2. \quad d(\alpha x, \alpha y) = |\alpha| d(x, y) \quad \forall x, y \in X, \alpha \in \mathbb{F} \quad (\text{HOMOGENEOUS}) \end{array} \right\} \iff \|\cdot\| \text{ is a NORM}$$

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 L.1 (page 280).
- (c) Proof that  $\|\cdot\|$  is *homogeneous*: This follows from the second left hypothesis.

<sup>11</sup> Bollobás (1999), page 21

(d) Proof that  $\|\cdot\|$  satisfies the *triangle-inequality*: This follows directly from Lemma L.1 (page 280).

2. Proof of  $\Leftarrow$  assertion:

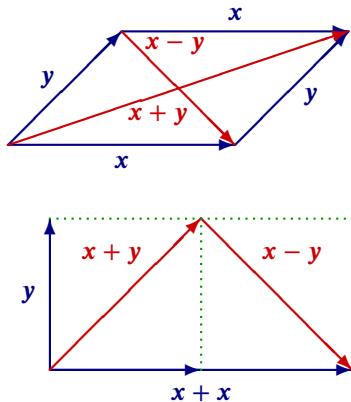
$$\begin{aligned}
 d(x+z, y+z) &= \|(x+z) - (y+z)\| && \text{by definition of } d \\
 &= \|x - y\| \\
 &= d(x, y) && \text{by definition of } d \\
 d(\alpha x, \alpha y) &= \|(\alpha x) - (\alpha y)\| && \text{by definition of } d \\
 &= \|\alpha(x - y)\| \\
 &= |\alpha| \|x - y\| && \text{by definition of } \|\cdot\| \text{ page 275} \\
 &= |\alpha| d(x, y) && \text{by definition of } d
 \end{aligned}$$

⇒

## L.3 Orthogonality on normed linear spaces

Traditionally, *orthogonality* (Definition K.4 page 271) is a property defined in *inner product spaces* (Definition K.1 page 259). However, the concept of orthogonality can be extended to *normed linear spaces* (Definition L.1 page 275). Here are some examples:

- ① *Isosceles orthogonality*: Definition L.3 page 282
- ② *Pythagorean orthogonality*: Definition L.4 page 284
- ③ *Birkhoff orthogonality*: Definition L.5 page 284



*Isosceles orthogonality* (Definition L.3 page 282) 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 ( $\|x + y\| = \|x - y\|$ ). Isosceles orthogonality can also be illustrated with a triangle where the sides are of lengths  $\|x + y\|$  and  $\|x - y\|$  and base of length  $\|x + x\|$ . In this case if  $x$  and  $y$  are orthogonal, then the triangle is *isosceles*. This is illustrated in figure to the lower left. Isosceles orthogonality is formally defined next.

**Definition L.3.** <sup>12</sup> Let  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$  be a NORMED LINEAR SPACE (Definition L.1 page 275).

**D E F** Two vectors  $x$  and  $y$  are **orthogonal in the sense of James** if

$$\|x + y\| = \|x - y\|.$$

This property is also called **isosceles orthogonality** or **James orthogonality**.

**Theorem L.6.** Let  $(X, +, \cdot, (\mathbb{R}, \dot{+}, \dot{\times}), \langle \Delta | \nabla \rangle)$  be an INNER-PRODUCT SPACE (Definition K.1 page 259) with induced norm  $\|x\| \triangleq \sqrt{\langle x | x \rangle}$ , ISOSCELES ORTHOGONALITY (Definition L.3 page 282) relation ①, and inner-product relation ORTHOGONALITY (Definition K.4 page 271) relation ⊥.

T H M	$\underbrace{x \oplus y}_{\text{orthogonal in the sense of James}}$	$\iff$	$\underbrace{x \perp y}_{\text{orthogonal in the sense of inner-product space}}$
-------------	---	--------	--

<sup>12</sup> James (1945) page 292 (DEFINITION 2.1), Amir (1986) page 24, Dunford and Schwartz (1957), page 93

PROOF:

1. Proof that  $x \odot y \implies x \perp y$ :

$$\begin{aligned}
 & 4 \langle x | y \rangle \\
 &= \underbrace{\|x + y\|^2 - \|x - y\|^2}_{0 \text{ by } x \odot y \text{ hypothesis}} + i \|x + iy\|^2 - i \|x - iy\|^2 \quad \text{by polarization identity (Theorem K.6 page 266)} \\
 &= 0 + i \|x + iy\|^2 - i \|x - iy\|^2 \quad \text{by } x \odot y \text{ hypothesis} \\
 &= i [\|x\|^2 + \|y\|^2 + 2\Re \langle x | iy \rangle] \\
 &\quad - i [\|x\|^2 + \|y\|^2 + 2\Re \langle x | -iy \rangle] \quad \text{by Polar Identity (Lemma K.1 page 263)} \\
 &= i [\|x\|^2 + \|y\|^2 + 2\Re \langle x | iy \rangle] \\
 &\quad - i [\|x\|^2 + \|y\|^2 + 2\Re \langle x | -iy \rangle] \quad \text{by Definition L.1 page 275 and Definition K.1 page 259} \\
 &= 4i\Re \langle x | iy \rangle \\
 &= 4i\Re [i^* \langle x | y \rangle] \\
 &= 0 \quad \text{because inner-product space is real } (\mathbb{F} = \mathbb{R})
 \end{aligned}$$

2. Proof that  $x \odot y \iff x \perp y$ :

$$\begin{aligned}
 \|x + y\|^2 &= \|x\|^2 + \|y\|^2 + 2\Re \langle x | y \rangle \quad \text{by Polar Identity (Lemma K.1 page 263)} \\
 &= \|x\|^2 + \|y\|^2 + 0 \quad \text{by } x \perp y \text{ hypothesis} \\
 &= \|x\|^2 + \|y\|^2 - 2\Re \langle x | y \rangle \quad \text{0 when } x \perp y \quad \text{by } x \perp y \text{ hypothesis} \\
 &= \|x\|^2 + \|y\|^2 + 2\Re \langle x | -y \rangle \\
 &= \|x - y\|^2 \quad \text{by Polar Identity (Lemma K.1 page 263)}
 \end{aligned}$$



**Theorem L.7.** <sup>13</sup> Let  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$  be a normed linear space and with ISOSCELES ORTHOGONALITY (Definition L.3 page 282) relation  $\odot$ .

T H M	$x \odot y \iff y \odot x \iff \alpha x \odot \alpha y \quad \forall \alpha \in \mathbb{F}$
-------------	---

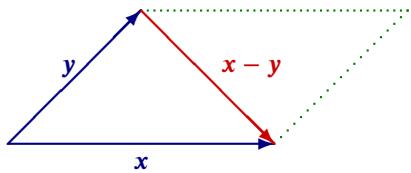
PROOF:

$$\begin{aligned}
 x \odot y \implies & \|x + y\| = \|x - y\| \quad \text{by Definition L.3 page 282} \\
 \implies & \|x + y\| = |-1| \|x - y\| \\
 \implies & \|x + y\| = \|(x - y)\| \quad \text{by Definition L.1 page 275} \\
 \implies & \|y + x\| = \|y - x\| \quad \text{by Definition G.1 page 197} \\
 \implies & y \odot x \quad \text{by Definition L.3 page 282} \\
 \\ 
 y \odot x \implies & \|y + x\| = \|y - x\| \quad \text{by Definition L.3 page 282} \\
 \implies & |\alpha| \|y + x\| = |\alpha| \|y - x\| \\
 \implies & \|\alpha(y + x)\| = \|\alpha(y - x)\| \quad \text{by Definition L.1 page 275} \\
 \implies & \|\alpha y + \alpha x\| = \|\alpha y - \alpha x\| \\
 \implies & \|\alpha x + \alpha y\| = \|-(\alpha x - \alpha y)\| \quad \text{by Definition G.1 page 197}
 \end{aligned}$$

<sup>13</sup> Amir (1986) page 24

$$\begin{aligned} \Rightarrow \| \alpha x + \alpha y \| &= | -1 | \| \alpha x - \alpha y \| \\ \Rightarrow \| \alpha x + \alpha y \| &= \| \alpha x - \alpha y \| \\ \Rightarrow \alpha x \oplus \alpha y & \end{aligned} \quad \begin{array}{l} \text{by Definition L.1 page 275} \\ \text{by Definition F.4 page 196} \\ \text{by Definition L.3 page 282} \end{array}$$

$$\begin{aligned} \alpha x \oplus \alpha y \Rightarrow \| \alpha x + \alpha y \| &= \| \alpha x - \alpha y \| \\ \Rightarrow \| \alpha(x + y) \| &= \| \alpha(x - y) \| \\ \Rightarrow | \alpha | \| x + y \| &= | \alpha | \| x - y \| \\ \Rightarrow \| x + y \| &= \| x - y \| \\ \Rightarrow x \oplus y & \end{aligned} \quad \begin{array}{l} \text{by Definition L.3 page 282} \\ \text{by Definition G.1 page 197} \\ \text{by Definition L.1 page 275} \\ \text{by Definition L.1 page 275} \\ \text{by Definition L.3 page 282} \end{array}$$



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 K.10 page 273),  $a^2 + b^2 = c^2$ . 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 L.4.** <sup>14</sup> Let  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$  be a NORMED LINEAR SPACE (Definition L.1 page 275).

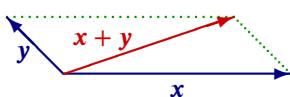
**D E F** Two vectors  $x$  and  $y$  are **orthogonal in the Pythagorean sense** if

$$\|x - y\|^2 = \|x\|^2 + \|y\|^2.$$

This relationship is also called **Pythagorean orthogonality**.

**Theorem L.8.** <sup>15</sup> Let  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle | \triangleright \rangle)$  be an INNER-PRODUCT SPACE (Definition K.1 page 259) with induced norm  $\|x\| \triangleq \sqrt{\langle x | x \rangle}$ , PYTHAGOREAN ORTHOGONALITY (Definition L.4 page 284) relation  $\oplus$ , and inner-product relation ORTHOGONALITY (Definition K.4 page 271) relation  $\perp$ .

T H M	$\underbrace{x \oplus y}_{\text{orthogonal in the Pythagorean sense}}$	$\iff$	$\underbrace{x \perp y}_{\text{orthogonal in the sense of inner-product space}}$
-------	--	--------	--



Besides *isosceles orthogonality* (Definition L.3 page 282), orthogonality in normed linear spaces can be defined using *Birkhoff orthogonality*, as defined in Definition L.5 (next) and illustrated to the left.

**Definition L.5.** <sup>16</sup> Let  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$  be a NORMED LINEAR SPACE (Definition L.1 page 275).

**D E F** Two vectors  $x$  and  $y$  are **orthogonal in the sense of Birkhoff** if

$$\|x\| \leq \|x + \alpha y\| \quad \forall \alpha \in \mathbb{F}.$$

This relationship is also called **Birkhoff orthogonality**.

**Theorem L.9.** Let  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \langle \triangle | \triangleright \rangle)$  be an INNER-PRODUCT SPACE (Definition K.1 page 259) with induced norm  $\|x\| \triangleq \sqrt{\langle x | x \rangle}$ , BIRKHOFF ORTHOGONALITY relation  $\oplus$  (Definition L.5 page 284), and inner-product relation ORTHOGONALITY relation  $\perp$  (Definition K.4 page 271).

T H M	$\underbrace{x \oplus y}_{\text{orthogonal in the sense of Birkhoff}}$	$\iff$	$\underbrace{x \perp y}_{\text{orthogonal in the sense of inner-product space}}$
-------	--	--------	--

<sup>14</sup> James (1945) page 292 (DEFINITION 2.2), Amir (1986) page 57, Drljević (1989) page 232

<sup>15</sup> Amir (1986) page 57

<sup>16</sup> Amir (1986) page 33, Dunford and Schwartz (1957), page 93, James (1947) page 265

# APPENDIX M

## INTERVALS AND CONVEXITY

### M.1 Intervals

In the real number system, for  $a \leq b$ , the *interval*  $[a : b]$  is the set  $a$  and  $b$  and all the numbers inbetween, as in  $[a : b] \triangleq \{x \in \mathbb{R} | a \leq x \leq b\}$ . This concept can be easily generalized:

- In an **ordered set**, if two elements  $x$  and  $y$  are *comparable* and  $x \leq y$ , then we say that  $x$  and  $y$  and all the elements inbetween, as determined by the ordering relation  $\leq$ , 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 (M.1) is restricted to the case in which  $x$  and  $y$  are *comparable*. This restriction can be lifted (Definition M.2 page 285) with the additional structure of upper and lower bounds provided by lattices.
- A **metric space** in general has no *order relation*  $\leq$ . But intervals can still be defined (Definition M.4 page 286) in a metric space in terms of the *triangle inequality*.
- A **linear space** (Definition G.1 page 197) over a real or complex field in general has no *order relation* that compares *vectors* in the space, but the standard order relation  $\leq$  for real numbers  $\mathbb{R}$  can still be used (Definition M.5 page 286) to define an interval in a linear space.

**Definition M.1 (intervals on ordered sets).** <sup>1</sup> Let  $(X, \leq)$  be an ORDERED SET.

DEF	The set $[x : y] \triangleq \{z \in X   x \leq z \leq y\}$ is called a <b>closed interval</b> and
	The set $(x : y] \triangleq \{z \in X   x < z \leq y\}$ is called a <b>half-open interval</b> and
	The set $[x : y) \triangleq \{z \in X   x \leq z < y\}$ is called a <b>half-open interval</b> and
	The set $(x : y) \triangleq \{z \in X   x < z < y\}$ is called an <b>open interval</b> .

**Definition M.2 (intervals on lattices).** <sup>2</sup> Let  $(X, \vee, \wedge; \leq)$  be a LATTICE.

DEF	The set $[x : y] \triangleq \{z \in X   x \wedge y \leq z \leq x \vee y\}$ is called a <b>closed interval</b> .
	The set $(x : y] \triangleq \{z \in X   x \wedge y < z \leq x \vee y\}$ is called a <b>half-open interval</b> .
	The set $[x : y) \triangleq \{z \in X   x \wedge y \leq z < x \vee y\}$ is called a <b>half-open interval</b> .
	The set $(x : y) \triangleq \{z \in X   x \wedge y < z < x \vee y\}$ is called an <b>open interval</b> .

When  $x$  and  $y$  are comparable and  $x \leq y$ , then Definition M.2 (previous) simplifies to item (M.1)

<sup>1</sup> Apostol (1975) page 4, Ore (1935) page 409

<sup>2</sup> Duthie (1942) page 2, Ore (1935) page 425 (*quotient structures*)

(page 285).

**Definition M.3.**<sup>3</sup> Let  $L \triangleq (X, \vee, \wedge; \leq)$  be a LATTICE with dual  $L^*$ . Let  $[x : y]$  be a CLOSED INTERVAL (Definition M.2 page 285) on set  $X$ . The sublattices  $L[x : y]$  and  $L^*[x : y]$  are defined as follows:

DEF	$L[x : y] \triangleq \{z \in L   z \in [x : y]\} \quad \forall x, y \in X$
DEF	$L^*[x : y] \triangleq \{z \in L^*   z \in [x : y]\} \quad \forall x, y \in X$

**Definition M.4.**<sup>4</sup>

DEF	In a METRIC SPACE $(X, d)$ , the set $[a : b]$ is the <b>closed interval</b> from $x$ to $y$ and is defined as $[x : y] \triangleq \{z \in X   d(x, z) + d(z, y) = d(x, y)\}$ . An element $z \in X$ is <b>geodesically between</b> $x$ and $y$ if $z \in [x : y]$ .
-----	---

**Definition M.5.**<sup>5</sup>

DEF	In a LINEAR SPACE $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$ (Definition G.1 page 197), $[x : y] \triangleq \{\lambda x + (1 - \lambda)y = z   0 \leq \lambda \leq 1\}$ is called a <b>closed interval</b> and $(x : y] \triangleq \{\lambda x + (1 - \lambda)y = z   0 < \lambda \leq 1\}$ is called a <b>half-open interval</b> and $[x : y) \triangleq \{\lambda x + (1 - \lambda)y = z   0 \leq \lambda < 1\}$ is called a <b>half-open interval</b> and $(x : y) \triangleq \{\lambda x + (1 - \lambda)y = z   0 < \lambda < 1\}$ is called an <b>open interval</b> .
-----	--

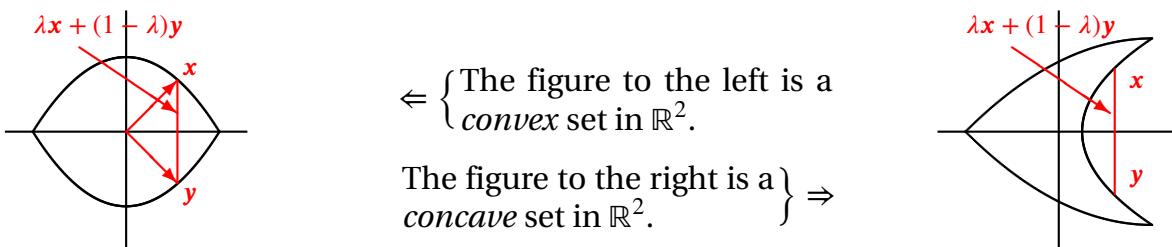
## M.2 Convex sets

Using the concept of the *interval* (previous section), we can define the *convex set* (next definition).

**Definition M.6.**<sup>6</sup> 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  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$ .

DEF	A subset $D \subseteq X$ is a <b>convex set</b> in $X$ if $x, y \in D \implies [x : y] \subseteq D$ . A set that is <b>not</b> convex is <b>concave</b> .
-----	---

*Example M.1.* Consider the Euclidean space  $\mathbb{R}^2$  (a special case of a linear space).



*Example M.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

<sup>3</sup> Maeda and Maeda (1970), page 1

<sup>4</sup> van de Vel (1993) page 8

<sup>5</sup> Barvinok (2002) page 2

<sup>6</sup> Barvinok (2002) page 5

Examples of metrics generating balls which are *not* convex include the following:

- Parabolic metric
- Exponential metric

## M.3 Convex functions

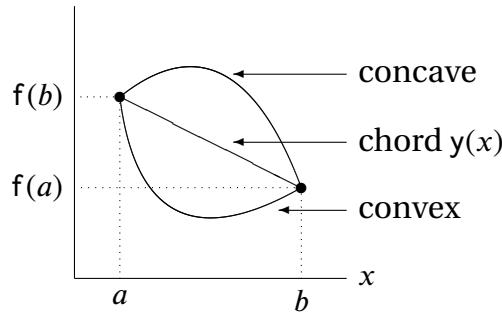


Figure M.1: Convex and concave functions

**Definition M.7.** <sup>7</sup> Let  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$  be a LINEAR SPACE (Definition G.1 page 197) and  $D$  a CONVEX SET (Definition M.6 page 286) in  $X$ .

A function  $f \in F^D$  is **convex** if

$$f(\lambda x + [1 - \lambda]y) \leq \lambda f(x) + (1 - \lambda) f(y) \quad \forall x, y \in D \text{ and } \forall \lambda \in (0, 1)$$

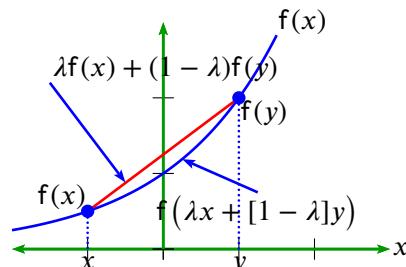
A function  $g \in F^D$  is **strictly convex** if

$$g(\lambda x + [1 - \lambda]y) = \lambda g(x) + (1 - \lambda) g(y) \quad \forall x, y \in D, x \neq y, \text{ and } \forall \lambda \in (0, 1)$$

A function  $f \in F^D$  is **concave** if  $-f$  is CONVEX.

A function  $f \in F^D$  is **affine** iff is CONVEX and CONCAVE.

**Example M.3.** The function  $f(x) = 2^x$  is a **convex function** (Definition M.7 page 287), as illustrated to the right.



**Definition M.8.** <sup>8</sup> Let  $(X, +, \cdot, (\mathbb{R}, \dot{+}, \dot{\times}))$  be a LINEAR SPACE (Definition G.1 page 197).

The **epigraph**  $\text{epi}(f)$  and **hypograph**  $\text{hyp}(f)$  of a functional  $f \in \mathbb{R}^X$  are defined as

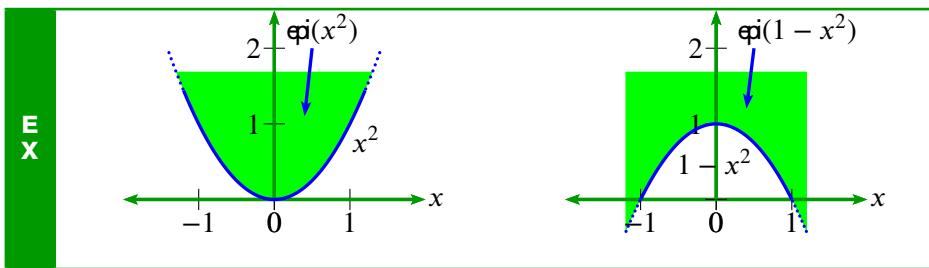
$$\text{epi}(f) \triangleq \{(x, y) \in X \times \mathbb{R} | y \geq f(x)\}$$

$$\text{hyp}(f) \triangleq \{(x, y) \in X \times \mathbb{R} | y \leq f(x)\}$$

**Example M.4.**

<sup>7</sup> Simon (2011) page 2, Barvinok (2002) page 2, Bollobás (1999), page 3, Jensen (1906), page 176, Clarkson (1936) (strictly convex)

<sup>8</sup> Beer (1993) page 13 (§1.3), Aubin and Frankowska (2009) page 222, Aubin (2011) page 223



**Proposition M.1.**<sup>9</sup> Let  $(X, +, \cdot, (\mathbb{R}, \dot{+}, \dot{\times}))$  be a LINEAR SPACE (Definition G.1 page 197). Let  $f$  be a FUNCTIONAL in  $\mathbb{R}^X$ .

P R P	$\left\{ \begin{array}{l} f \text{ is a} \\ \text{CONVEX FUNCTION} \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} \text{epi}(f) \text{ is a} \\ \text{CONVEX SET} \end{array} \right\}$
-------------	--

Often a function can be proven to be *convex* or *concave*. *Convex* and *concave* functions are defined in Definition M.9 (page 288) (next) and illustrated in Figure M.1 (page 287).

**Definition M.9.** Let

$$y(x) = \frac{f(b) - f(a)}{b - a}(x - a) + f(a).$$

A function  $f : \mathbb{R} \rightarrow \mathbb{R}$  is

D E F	(1). <b>convex</b> in $(a : b)$ if $f(x) \leq y(x)$ for $x \in (a : b)$ (2). <b>concave</b> in $(a : b)$ if $f(x) \geq y(x)$ for $x \in (a : b)$ (3). <b>strictly convex</b> in $(a : b)$ if $f(x) < y(x)$ for $x \in (a : b)$ (4). <b>strictly concave</b> in $(a : b)$ if $f(x) > y(x)$ for $x \in (a : b)$
-------------	--

**Theorem M.1** (Jensen's Inequality).<sup>10</sup> Let  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$  be a LINEAR SPACE (Definition G.1 page 197),  $D$  a subset of  $X$ , and  $f$  a functional in  $\mathbb{F}^D$ . Let  $\sum$  be the SUMMATION OPERATOR (Definition N.1 page 293).

T H M	$\left\{ \begin{array}{ll} 1. & D \text{ is CONVEX} & \text{and} \\ 2. & f \text{ is CONVEX} & \text{and} \\ 3. & \sum_{n=1}^N \lambda_n = 1 & (\text{WEIGHTS}) \end{array} \right\} \implies f\left(\sum_{n=1}^N \lambda_n x_n\right) \leq \sum_{n=1}^N \lambda_n f(x_n) \quad \forall x_n \in D, N \in \mathbb{N}$
-------------	--

PROOF: Proof is by induction:

1. Proof that statement is true for  $N = 1$ :

$$\begin{aligned} f\left(\sum_{n=1}^{N=1} \lambda_n x_n\right) &= f(\lambda_1 x_1) \\ &\leq f(\lambda_1 x_1) \\ &= \sum_{n=1}^{N=1} \lambda_n f(x_n) \end{aligned}$$

<sup>9</sup> 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)

<sup>10</sup> Mitrovic 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  $N = 2$ :

$$\begin{aligned}
 f\left(\sum_{n=1}^{N=2} \lambda_n x_n\right) &= f(\lambda_1 x_1 + \lambda_2 x_2) \\
 &\leq \lambda_1 f(x_1) + \lambda_2 f(x_2) && \text{by convexity hypothesis} \\
 &= \sum_{n=1}^{N=2} \lambda_n f(x_n)
 \end{aligned}$$

3. Proof that if the statement is true for  $N$ , then it is also true for  $N + 1$ :

$$\begin{aligned}
 f\left(\sum_{n=1}^{N+1} \lambda_n x_n\right) &= f\left(\sum_{n=1}^N \lambda_n x_n + \lambda_{N+1} x_{N+1}\right) \\
 &= f\left([1 - \lambda_{N+1}] \sum_{n=1}^N \frac{\lambda_n}{1 - \lambda_{N+1}} x_n + \lambda_{N+1} x_{N+1}\right) \\
 &\leq [1 - \lambda_{N+1}] f\left(\sum_{n=1}^N \frac{\lambda_n}{1 - \lambda_{N+1}} x_n\right) + \lambda_{N+1} f(x_{N+1}) && \text{by convexity hypothesis} \\
 &\leq [1 - \lambda_{N+1}] \sum_{n=1}^N \frac{\lambda_n}{1 - \lambda_{N+1}} f(x_n) + \lambda_{N+1} f(x_{N+1}) && \text{by "true for } N\text" hypothesis} \\
 &= \sum_{n=1}^N \lambda_n f(x_n) + \lambda_{N+1} f(x_{N+1}) \\
 &= \sum_{n=1}^{N+1} \lambda_n f(x_n)
 \end{aligned}$$

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  $N = 1, 2, 3, 4, \dots$



The next theorem gives another form of convex functions that is a little less intuitive but provides powerful analytic results.

**Theorem M.2.** Let  $f : \mathbb{R} \rightarrow \mathbb{R}$ . For every  $x_1, x_2 \in (a, b)$  and  $\lambda \in [0, 1]$

T  
H  
M

$f$  is convex in  $(a, b) \iff f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2)$

PROOF:

1. prove  $f$  is convex  $\implies f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2)$ :

$$\begin{aligned}
 f(\lambda x_1 + (1 - \lambda)x_2) &\leq \frac{f(b) - f(a)}{b - a} [\lambda x_1 + (1 - \lambda)x_2 - a] + f(a) \\
 &= \frac{f(x_2) - f(x_1)}{x_2 - x_1} [\lambda x_1 + (1 - \lambda)x_2 - x_1] + f(x_1) \\
 &= \frac{f(x_2) - f(x_1)}{x_2 - x_1} [(x_2 - x_1)(1 - \lambda)] + f(x_1) \\
 &= (1 - \lambda)f(x_2) - (1 - \lambda)f(x_1) + f(x_1) \\
 &= \lambda f(x_1) + (1 - \lambda)f(x_2)
 \end{aligned}$$

2. prove  $f$  is convex  $\iff f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2)$ :

Let  $x = \lambda(b - a) + a$  Notice that as  $\lambda$  varies from 0 to 1,  $x$  varies from  $b$  to  $a$ . So free variable  $\lambda$  works as a change of variable for free variable  $x$ .

$$\begin{aligned}\lambda &= \frac{x - a}{b - a} \\ f(x) &= f(\lambda(b - a) + a) \\ &\leq \lambda f(b) + (1 - \lambda)f(a) \\ &= \lambda[f(b) - f(a)] + f(a) \\ &= \frac{f(b) - f(a)}{b - a}(x - a) + f(a)\end{aligned}$$

⇒

Taking the second derivative of a function provides a convenient test for whether that function is convex.

**Theorem M.3.** <sup>11</sup>

T  
H  
M

$f''(x) > 0 \implies f$  is convex

PROOF:

$$\begin{aligned}f(x) &= f(x_0) + f'(x_0)(x - x_0) + f''(c)(x - x_0)^2 \\ &\geq f(x_0) + f'(x_0)(x - x_0) \\ &= f(x_0) + f'(x_0)(x - \lambda x_1 - (1 - \lambda)x_2)\end{aligned}$$

$$\begin{aligned}f(x_1) &\geq f(x_0) + f'(x_0)(x_1 - \lambda x_1 - (1 - \lambda)x_2) \\ &= f(x_0) + f'(x_0)(1 - \lambda)(x_1 - x_2) \\ &= f(x_0) - f'(x_0)(1 - \lambda)(x_2 - x_1)\end{aligned}$$

$$\begin{aligned}f(x_2) &\geq f(x_0) + f'(x_0)(x_2 - \lambda x_1 - (1 - \lambda)x_2) \\ &= f(x_0) + f'(x_0)\lambda(x_2 - x_1)\end{aligned}$$

$$\begin{aligned}\lambda f(x_1) + (1 - \lambda)f(x_2) &\geq \lambda [f(x_0) - f'(x_0)(1 - \lambda)(x_2 - x_1)] + (1 - \lambda) [f(x_0) + f'(x_0)\lambda(x_2 - x_1)] \\ &= \lambda [f(x_0) - f'(x_0)(1 - \lambda)(x_2 - x_1)] + [f(x_0) + f'(x_0)\lambda(x_2 - x_1)] - \lambda [f(x_0) + f'(x_0)\lambda(x_2 - x_1)] \\ &= f(x_0) \\ &= f(\lambda x_1 + (1 - \lambda)x_2)\end{aligned}$$

By Theorem M.2 (page 289),  $f(x)$  is convex.

⇒

## M.4 Literature

 **Literature survey:**

<sup>11</sup>  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 N

## 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.””<sup>1</sup>

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.<sup>1</sup>

## N.1 Summation

**Definition N.1.** <sup>2</sup> Let  $+$  be an addition operator on a tuple  $(x_n)_m^N$ .

The **summation** of  $(x_n)$  from index  $m$  to index  $N$  with respect to  $+$  is

$$\sum_{n=m}^N x_n \triangleq \begin{cases} 0 & \text{for } N < m \\ \left( \sum_{n=m}^{N-1} x_n \right) + x_N & \text{for } N \geq m \end{cases}$$

**Theorem N.1** (Generalized associative property). <sup>3</sup> Let  $+$  be an addition operator on a tuple  $(x_n)_m^N$ .

$$\begin{aligned} + \text{ is ASSOCIATIVE} \implies & \sum_{n=m}^L x_n + \left( \sum_{n=L+1}^M x_n + \sum_{n=M+1}^N x_n \right) = \left( \sum_{n=m}^L x_n + \sum_{n=L+1}^M x_n \right) + \sum_{n=M+1}^N x_n \quad \text{for } m < L < M \leq N \\ & \underbrace{\sum_{n=m}^N x_n}_{\text{is ASSOCIATIVE}} \end{aligned}$$

PROOF:

<sup>1</sup> quote: [Hardy \(1929\)](#), page 64

image: [http://www-history.mcs.st-andrews.ac.uk/PictDisplay/Bohr\\_Harald.html](http://www-history.mcs.st-andrews.ac.uk/PictDisplay/Bohr_Harald.html)

<sup>2</sup> reference: [Berberian \(1961\) page 8](#) (Definition I.3.1)

“ $\Sigma$ ” notation: [Fourier \(1820\)](#) page 280

<sup>3</sup> [Berberian \(1961\)](#) pages 9–10 (Theorem I.3.1)

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:

$$\begin{aligned} \sum_{n=m}^{m+2} x_n &= \left( \sum_{n=m}^{m+1} x_n \right) + x_{m+2} && \text{by Definition N.1 page 293} \\ &= (x_m + x_{m+1}) + x_{m+2} && \text{by item (3)} \\ &= x_m + (x_{m+1} + x_{m+2}) && \text{by left hypothesis} \end{aligned}$$

5. Proof that  $N$  case  $\implies N + 1$  case:

$$\begin{aligned} \sum_{n=m}^{N+1} x_n &= \underbrace{\left( \sum_{n=m}^N x_n \right)}_{\text{associative}} + x_{N+1} && \text{by Definition N.1 page 293} \\ &= \left( \sum_{n=m}^L x_n + \left( \sum_{n=L+1}^M x_n + \sum_{n=M+1}^N x_n \right) \right) + x_{N+1} && = \left( \left( \sum_{n=m}^L x_n + \sum_{n=L+1}^M x_n \right) + \sum_{n=M+1}^N x_n \right) + x_{N+1} \\ &= \left( \sum_{n=m}^L x_n + \sum_{n=L+1}^M x_n \right) + \left( \sum_{n=M+1}^N x_n + x_{N+1} \right) && = \left( \sum_{n=m}^L x_n + \sum_{n=L+1}^M x_n \right) + \left( \sum_{n=M+1}^{N+1} x_n \right) \end{aligned}$$

$\iff$

## N.2 Means

### N.2.1 Weighted $\phi$ -means

**Definition N.2.**<sup>4</sup>

The  $(\lambda_n)_1^N$  weighted  $\phi$ -mean of a tuple  $(x_n)_1^N$  is defined as

$$M_\phi((x_n)) \triangleq \phi^{-1} \left( \sum_{n=1}^N \lambda_n \phi(x_n) \right)$$

where  $\phi$  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  $\sum_{n=1}^N \lambda_n = 1$ .

D  
E  
F

<sup>4</sup>  Bollobás (1999) page 5



**Lemma N.1.** <sup>5</sup> Let  $M_\phi(\langle x_n \rangle)$  be the  $(\lambda_n)_1^N$  weighted  $\phi$ -mean of a tuple  $\langle x_n \rangle_1^N$ . Let the property CONVEX be defined as in Definition M.7 (page 287).

LEM

$\phi\psi^{-1}$ is CONVEX and $\phi$ is INCREASING	$\implies M_\phi(\langle x_n \rangle) \geq M_\psi(\langle x_n \rangle)$
$\phi\psi^{-1}$ is CONVEX and $\phi$ is DECREASING	$\implies M_\phi(\langle x_n \rangle) \leq M_\psi(\langle x_n \rangle)$
$\phi\psi^{-1}$ is CONCAVE and $\phi$ is INCREASING	$\implies M_\phi(\langle x_n \rangle) \leq M_\psi(\langle x_n \rangle)$
$\phi\psi^{-1}$ is CONCAVE and $\phi$ is DECREASING	$\implies M_\phi(\langle x_n \rangle) \geq M_\psi(\langle x_n \rangle)$

PROOF:

1. Case where  $\phi\psi^{-1}$  is *convex* and  $\phi$  is *increasing*:

$$\begin{aligned}
 M_\phi(\langle x_n \rangle) &\triangleq \phi^{-1}\left(\sum_{n=1}^N \lambda_n \phi(x_n)\right) && \text{by definition of } M_\phi && (\text{Definition N.2 page 294}) \\
 &= \phi^{-1}\left(\sum_{n=1}^N \lambda_n \phi\psi^{-1}\psi(x_n)\right) && \text{by definition of } \psi^{-1} \\
 &\geq \phi^{-1}\left(\phi\psi^{-1} \sum_{n=1}^N \lambda_n \psi(x_n)\right) && \text{by Jensen's Inequality} && (\text{Theorem M.1 page 288}) \\
 &= \left(\psi^{-1} \sum_{n=1}^N \lambda_n \psi(x_n)\right) && \text{by definition of } \psi^{-1} \\
 &\triangleq M_\psi(\langle x_n \rangle) && \text{by definition of } M_\psi && (\text{Definition N.2 page 294})
 \end{aligned}$$

2. Case where  $\phi\psi^{-1}$  is *convex* and  $\phi$  is *decreasing*:

$$\begin{aligned}
 M_\phi(\langle x_n \rangle) &\triangleq \phi^{-1}\left(\sum_{n=1}^N \lambda_n \phi(x_n)\right) && \text{by definition of } M_\phi && (\text{Definition N.2 page 294}) \\
 &= \phi^{-1}\left(\sum_{n=1}^N \lambda_n \phi\psi^{-1}\psi(x_n)\right) && \text{by definition of } \psi^{-1} \\
 &\leq \phi^{-1}\left(\phi\psi^{-1} \sum_{n=1}^N \lambda_n \psi(x_n)\right) && \text{by Jensen's Inequality} && (\text{Theorem M.1 page 288}) \\
 &&& \text{and because } \phi^{-1} \text{ is decreasing} && (\text{by hypothesis}) \\
 &= \left(\psi^{-1} \sum_{n=1}^N \lambda_n \psi(x_n)\right) && \text{by definition of } \psi^{-1} \\
 &\triangleq M_\psi(\langle x_n \rangle) && \text{by definition of } M_\psi && (\text{Definition N.2 page 294})
 \end{aligned}$$



One of the most well known inequalities in mathematics is *Minkowski's Inequality* (1910, Theorem N.5 page 301). In 1946, H.P. Mulholland submitted a result<sup>6</sup> that generalizes Minkowski's Inequality to an equal weighted  $\phi$ -mean. And Milovanović and Milovanović (1979) generalized this even further to a *weighted*  $\phi$ -mean (Theorem N.2, next).

## Theorem N.2. <sup>7</sup>

<sup>5</sup> Pečarić et al. (1992) page 107, Bollobás (1999) page 5, Hardy et al. (1952) page 75

<sup>6</sup> Mulholland (1950)

<sup>7</sup> Milovanović and Milovanović (1979), Bullen (2003) page 306 (Theorem 9)

**T H M**

$$\left\{ \begin{array}{l} (1). \phi \text{ is CONVEX} \\ (2). \phi \text{ is STRICTLY MONOTONIC} \end{array} \right. \text{ and } \left\{ \begin{array}{l} (3). \phi(0) = 0 \\ (4). \log \circ \phi \circ \exp \text{ is CONVEX} \end{array} \right. \text{ and } \left\{ \begin{array}{l} \left\{ \phi^{-1} \left( \sum_{n=1}^N \lambda_n \phi(x_n + y_n) \right) \leq \phi^{-1} \left( \sum_{n=1}^N \lambda_n \phi(x_n) \right) + \phi^{-1} \left( \sum_{n=1}^N \lambda_n \phi(y_n) \right) \right\} \end{array} \right\} \Rightarrow$$

## N.2.2 Power means

**Definition N.3.**<sup>8</sup> Let  $M_{\phi(x;r)}(\{x_n\})$  be the  $(\lambda_n)_1^N$  weighted  $\phi$ -mean of a NON-NEGATIVE tuple  $(x_n)_1^N$  (Definition N.2 page 294).

**D E F** A mean  $M_{\phi(x;r)}(\{x_n\})$  is a **power mean** with parameter  $r$  if  $\phi(x) \triangleq x^r$ . That is,

$$M_{\phi(x;r)}(\{x_n\}) = \left( \sum_{n=1}^N \lambda_n (x_n)^r \right)^{\frac{1}{r}}$$

**Theorem N.3.**<sup>9</sup> Let  $M_{\phi(x;r)}(\{x_n\})$  be POWER MEAN with parameter  $r$  of an  $N$ -tuple  $(x_n)_1^N$ . Let  $\mathbb{R}^*$  be the set of extended real numbers ( $\mathbb{R}^* \triangleq \mathbb{R} \cup \{-\infty, \infty\}$ ).<sup>10</sup>

**T H M**

$$M_{\phi(x;r)}(\{x_n\}) \triangleq \left( \sum_{n=1}^N \lambda_n (x_n)^r \right)^{\frac{1}{r}} \text{ is CONTINUOUS and STRICTLY INCREASING in } \mathbb{R}^*.$$

$$M_{\phi(x;r)}(\{x_n\}) = \begin{cases} \min_{n=1,2,\dots,N} \{x_n\} & \text{for } r = -\infty \\ \prod_{n=1}^N x_n^{\lambda_n} & \text{for } r = 0 \\ \max_{n=1,2,\dots,N} \{x_n\} & \text{for } r = +\infty \end{cases}$$

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  $s$ :
 

$r < 0$ ( $\phi_r$ decreasing)	$r > 0$ ( $\phi_r$ increasing)
--------------------------------	--------------------------------

$s < 0$	convex	(not possible)
$s > 0$	convex	concave

(d) Therefore by Lemma N.1 (page 295),

$$-\infty < r < s < \infty \implies M_{\phi(x;r)}(\{x_n\}) < M_{\phi(x;s)}(\{x_n\}).$$

2. Proof that  $M_{\phi(x;r)}$  is continuous in  $r$  for  $r \in \mathbb{R} \setminus 0$ : The sum of continuous functions is continuous. For the cases of  $r \in \{-\infty, 0, \infty\}$ , see the items that follow.

<sup>8</sup> Bullen (2003) page 175, Bollobás (1999) page 6

<sup>9</sup> Bullen (2003) pages 175–177 (see also page 203), Bollobás (1999) pages 6–8, Besso (1879), Bienaymé (1840) page 68

<sup>10</sup> Rana (2002) pages 385–388 (Appendix A)

3. Lemma:  $M_{\phi(x;-r)}(\langle x_n \rangle) = \{M_{\phi(x;r)}(\langle x_n^{-1} \rangle)\}^{-1}$ . Proof:

$$\begin{aligned} \{M_{\phi(x;r)}(\langle x_n^{-1} \rangle)\}^{-1} &= \left\{ \left( \sum_{n=1}^N \lambda_n (x_n^{-1})^r \right)^{\frac{1}{r}} \right\}^{-1} && \text{by definition of } M_\phi \\ &= \left( \sum_{n=1}^N \lambda_n (x_n)^{-r} \right)^{\frac{1}{-r}} \\ &= M_{\phi(x;-r)}(\langle x_n \rangle) && \text{by definition of } M_\phi \end{aligned}$$

4. Proof that  $\lim_{r \rightarrow \infty} M_\phi(\langle x_n \rangle) = \max_{n \in \mathbb{Z}} \langle x_n \rangle$ :

(a) Let  $x_m \triangleq \max_{n \in \mathbb{Z}} \langle x_n \rangle$

(b) Note that  $\lim_{r \rightarrow \infty} M_\phi \leq \max_{n \in \mathbb{Z}} \langle x_n \rangle$  because

$$\begin{aligned} \lim_{r \rightarrow \infty} M_\phi(\langle x_n \rangle) &= \lim_{r \rightarrow \infty} \left( \sum_{n=1}^N \lambda_n x_n^r \right)^{\frac{1}{r}} && \text{by definition of } M_\phi \\ &\leq \lim_{r \rightarrow \infty} \left( \sum_{n=1}^N \lambda_n x_m^r \right)^{\frac{1}{r}} && \text{by definition of } x_m \text{ in item (4a) and because } \\ &&& \phi(x) \triangleq x^r \text{ and } \phi^{-1} \text{ are both increasing or both decreasing} \\ &= \lim_{r \rightarrow \infty} \left( x_m^r \underbrace{\sum_{n=1}^N \lambda_n}_1 \right)^{\frac{1}{r}} && \text{because } x_m \text{ is a constant} \\ &= \lim_{r \rightarrow \infty} (x_m^r \cdot 1)^{\frac{1}{r}} \\ &= x_m \\ &= \max_{n \in \mathbb{Z}} \langle x_n \rangle && \text{by definition of } x_m \text{ in item (4a)} \end{aligned}$$

(c) But also note that  $\lim_{r \rightarrow \infty} M_\phi \geq \max_{n \in \mathbb{Z}} \langle x_n \rangle$  because

$$\begin{aligned} \lim_{r \rightarrow \infty} M_\phi(\langle x_n \rangle) &= \lim_{r \rightarrow \infty} \left( \sum_{n=1}^N \lambda_n x_n^r \right)^{\frac{1}{r}} && \text{by definition of } M_\phi \\ &\geq \lim_{r \rightarrow \infty} (w_m x_m^r)^{\frac{1}{r}} && \text{by definition of } x_m \text{ in item (4a) and because } \\ &&& \phi(x) \triangleq x^r \text{ and } \phi^{-1} \text{ are both increasing or both decreasing} \\ &= \lim_{r \rightarrow \infty} w_m^{\frac{1}{r}} x_m^r \\ &= x_m \\ &= \max_{n \in \mathbb{Z}} \langle x_n \rangle && \text{by definition of } x_m \text{ in item (4a)} \end{aligned}$$

(d) Combining items (b) and (c) we have  $\lim_{r \rightarrow \infty} M_\phi = \max_{n \in \mathbb{Z}} \langle x_n \rangle$ .

5. Proof that  $\lim_{r \rightarrow -\infty} M_\phi(\langle x_n \rangle) = \min_{n \in \mathbb{Z}} \langle x_n \rangle$ :

$$\begin{aligned}
 \lim_{r \rightarrow -\infty} M_{\phi(x;r)}(\langle x_n \rangle) &= \lim_{r \rightarrow \infty} M_{\phi(x;-r)}(\langle x_n \rangle) && \text{by change of variable } r \\
 &= \lim_{r \rightarrow \infty} \{M_{\phi(x;r)}(\langle x_n^{-1} \rangle)\}^{-1} && \text{by Lemma in item (3) page 297} \\
 &= \lim_{r \rightarrow \infty} \frac{1}{M_{\phi(x;r)}(\langle x_n^{-1} \rangle)} && \\
 &= \frac{\lim_{r \rightarrow \infty} 1}{\lim_{r \rightarrow \infty} M_{\phi(x;r)}(\langle x_n^{-1} \rangle)} && \text{by property of lim } ^{11} \\
 &= \frac{1}{\max_{n \in \mathbb{Z}} \langle x_n^{-1} \rangle} && \text{by item (4)} \\
 &= \frac{1}{\left( \min_{n \in \mathbb{Z}} \langle x_n \rangle \right)^{-1}} \\
 &= \min_{n \in \mathbb{Z}} \langle x_n \rangle
 \end{aligned}$$

6. Proof that  $\lim_{r \rightarrow 0} M_\phi(\langle x_n \rangle) = \prod_{n=1}^N x_n^{\lambda_n}$ :

$$\begin{aligned}
 \lim_{r \rightarrow 0} M_\phi(\langle x_n \rangle) &= \lim_{r \rightarrow 0} \exp \{ \ln \{ M_\phi(\langle x_n \rangle) \} \} \\
 &= \lim_{r \rightarrow 0} \exp \left\{ \ln \left\{ \left( \sum_{n=1}^N \lambda_n (x_n^r) \right)^{\frac{1}{r}} \right\} \right\} && \text{by definition of } M_\phi \\
 &= \exp \left\{ \frac{\frac{\partial}{\partial r} \ln \left( \sum_{n=1}^N \lambda_n (x_n^r) \right)}{\frac{\partial}{\partial r} r} \right\}_{r=0} && \text{by l'Hôpital's rule } ^{12} \\
 &= \exp \left\{ \frac{\sum_{n=1}^N \lambda_n \frac{\partial}{\partial r} (x_n^r)}{\sum_{n=1}^N \lambda_n (x_n^r)} \right\}_{r=0} \\
 &= \exp \left\{ \frac{\sum_{n=1}^N \lambda_n \frac{\partial}{\partial r} \exp(r \ln(x_n))}{1} \right\}_{r=0} \\
 &= \exp \left\{ \sum_{n=1}^N \lambda_n \exp \{ r \ln(x_n) \} \ln(x_n) \right\}_{r=0} \\
 &= \exp \left\{ \sum_{n=1}^N \lambda_n \ln(x_n) \right\} \\
 &= \exp \left\{ \ln \prod_{n=1}^N x_n^{\lambda_n} \right\} = \prod_{n=1}^N x_n^{\lambda_n}
 \end{aligned}$$

<sup>11</sup>  Rudin (1976) page 85 (4.4 Theorem)

<sup>12</sup>  Rudin (1976) page 109 (5.13 Theorem)



**Definition N.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  $\mu_g \triangleq \prod_{n=1}^N x_n^{\lambda_n}$  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  $\mu_a \triangleq \frac{1}{N} \sum_{n=1}^N x_n$

## N.3 Inequalities on power means

**Corollary N.1.** <sup>13</sup> Let  $(x_n)_1^N$  be a tuple. Let  $(\lambda_n)_1^N$  be a tuple of weighting values.

COR

$$\min(x_n) \leq \underbrace{\left( \sum_{n=1}^N \lambda_n \frac{1}{x_n} \right)^{-1}}_{\text{harmonic mean}} \leq \underbrace{\prod_{n=1}^N x_n^{\lambda_n}}_{\text{geometric mean}} \leq \underbrace{\sum_{n=1}^N \lambda_n x_n}_{\text{arithmetic mean}} \leq \max(x_n) \quad \text{where } \sum_{n=1}^N \lambda_n = 1$$

PROOF:

- These five means are all special cases of the *power mean*  $M_{\phi(x:r)}$  (Definition N.3 page 296):

$$\begin{aligned} r = \infty: & \max(x_n) \\ r = 1: & \text{arithmetic mean} \\ r = 0: & \text{geometric mean} \\ r = -1: & \text{harmonic mean} \\ r = -\infty: & \min(x_n) \end{aligned}$$

- The inequalities follow directly from Theorem N.3 (page 296).
- 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*:

$$\begin{aligned} \sum_{n=1}^N \lambda_n x_n &= b^{\log_b(\sum_{n=1}^N \lambda_n x_n)} \geq b^{\left( \sum_{n=1}^N \lambda_n \log_b x_n \right)} \quad \text{by Jensen's Inequality (Theorem M.1 page 288)} \\ &= \prod_{n=1}^N b^{(\lambda_n \log_b x_n)} = \prod_{n=1}^N b^{(\log_b x_n) \lambda_n} = \prod_{n=1}^N x_n^{\lambda_n} \end{aligned}$$



<sup>13</sup> Bullen (2003) page 71, Bollobás (1999) page 5, Cauchy (1821), pages 457–459 (Note II, theorem 17), Jensen (1906) page 183

**Lemma N.2** (Young's Inequality). <sup>14</sup>

LEM

$$\begin{aligned} xy &< \frac{x^p}{p} + \frac{y^q}{q} \quad \text{with } \frac{1}{p} + \frac{1}{q} = 1 \quad \forall 1 < p < \infty, x, y \geq 0, \quad \text{but } y \neq x^{p-1} \\ xy &= \frac{x^p}{p} + \frac{y^q}{q} \quad \text{with } \frac{1}{p} + \frac{1}{q} = 1 \quad \forall 1 < p < \infty, x, y \geq 0, \quad \text{and } y = x^{p-1} \end{aligned}$$

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  $v = u^{p-1} \iff u = v^{q-1}$ :

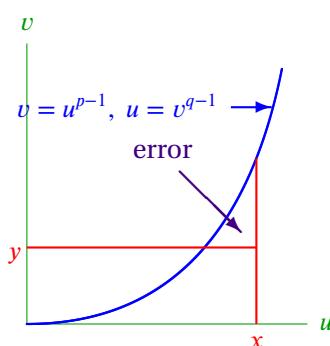
$$\begin{aligned} u &= v^{\frac{1}{p-1}} && \text{by left hypothesis} \\ &= v^{q-1} && \text{by item (1)} \end{aligned}$$

3. Proof that  $v = u^{p-1}$  is propemonotonically increasing in  $u$  and  $u = v^{q-1}$  is propemonotonically increasing in  $v$ :

$$\begin{aligned} \frac{dv}{du} &= \frac{d}{du} u^{p-1} &= (p-1)u^{p-2} &> 0 \\ \frac{du}{dv} &= \frac{d}{dv} v^{q-1} &= (q-1)v^{q-2} &> 0 \end{aligned}$$

4. Proof that  $xy \leq \frac{x^p}{p} + \frac{y^q}{q}$ :

$$\begin{aligned} xy &\leq \int_0^x u^{p-1} du + \int_0^y v^{q-1} dv \\ &= \frac{u^p}{p} \Big|_0^x + \frac{v^q}{q} \Big|_0^y \\ &= \frac{x^p}{p} + \frac{y^q}{q} \end{aligned}$$



<sup>14</sup> Carothers (2000), page 43, Tolsted (1964), page 5, Maligranda (1995), page 257, Hardy et al. (1952) (Theorem 24), Young (1912) page 226



**Theorem N.4** (Hölder's Inequality). <sup>15</sup> Let  $\langle x_n \rangle \in \mathbb{C} \rangle_1^N$  and  $\langle y_n \rangle \in \mathbb{C} \rangle_1^N$  be complex  $N$ -tuples.

T H M	$\underbrace{\sum_{n=1}^N  x_n y_n }_{\ \mathbf{x} \cdot \mathbf{y}\ _1} \leq \underbrace{\left( \sum_{n=1}^N  x_n ^p \right)^{\frac{1}{p}}}_{\ \mathbf{x}\ _p} \underbrace{\left( \sum_{n=1}^N  y_n ^q \right)^{\frac{1}{q}}}_{\ \mathbf{y}\ _q}$	<i>with</i> $\frac{1}{p} + \frac{1}{q} = 1 \quad \forall 1 < p < \infty$
-------------	--	--

PROOF: Let  $\|x_n\|_p \triangleq \left( \sum_{n=1}^N |x_n|^p \right)^{\frac{1}{p}}$ .

$$\begin{aligned}
 \sum_{n=1}^N |x_n y_n| &= \|(\langle x_n \rangle\|_p \|(\langle y_n \rangle\|_q \sum_{n=1}^N \frac{|x_n y_n|}{\|(\langle x_n \rangle\|_p \|(\langle y_n \rangle\|_q} \\
 &= \|(\langle x_n \rangle\|_p \|(\langle y_n \rangle\|_q \sum_{n=1}^N \frac{|x_n|}{\|(\langle x_n \rangle\|_p} \frac{|y_n|}{\|(\langle y_n \rangle\|_q} \\
 &\leq \|(\langle x_n \rangle\|_p \|(\langle y_n \rangle\|_q \sum_{n=1}^N \left( \frac{1}{p} \frac{|x_n|^p}{\|(\langle x_n \rangle\|_p^p} + \frac{1}{q} \frac{|y_n|^q}{\|(\langle y_n \rangle\|_q^q} \right) \quad \text{by Young's Inequality} \quad (\text{Lemma N.2 page 300}) \\
 &= \|(\langle x_n \rangle\|_p \|(\langle y_n \rangle\|_q \left( \frac{1}{p} \cdot \frac{\sum |x_n|^p}{\|(\langle x_n \rangle\|_p^p} + \frac{1}{q} \cdot \frac{\sum |y_n|^q}{\|(\langle y_n \rangle\|_q^q} \right) \\
 &= \|(\langle x_n \rangle\|_p \|(\langle y_n \rangle\|_q \left( \frac{1}{p} \frac{\|(\langle x_n \rangle\|_p^p}{\|(\langle x_n \rangle\|_p^p} + \frac{1}{q} \frac{\|(\langle y_n \rangle\|_q^q}{\|(\langle y_n \rangle\|_q^q} \right) \quad \text{by definition of } \|\cdot\| \\
 &= \|(\langle x_n \rangle\|_p \|(\langle y_n \rangle\|_q \underbrace{\left( \frac{1}{p} + \frac{1}{q} \right)}_1 \\
 &= \|(\langle x_n \rangle\|_p \|(\langle y_n \rangle\|_q \quad \text{by } \frac{1}{p} + \frac{1}{q} = 1 \text{ constraint}
 \end{aligned}$$



**Theorem N.5** (Minkowski's Inequality for sequences). <sup>16</sup> Let  $\langle x_n \rangle \in \mathbb{C} \rangle_1^N$  and  $\langle y_n \rangle \in \mathbb{C} \rangle_1^N$  be complex  $N$ -tuples.

T H M	$\left( \sum_{n=1}^N  x_n + y_n ^p \right)^{\frac{1}{p}} \leq \left( \sum_{n=1}^N  x_n ^p \right)^{\frac{1}{p}} + \left( \sum_{n=1}^N  y_n ^p \right)^{\frac{1}{p}} \quad \forall 1 < p < \infty$
-------------	---

PROOF:

1. Define  $q \triangleq \frac{p}{p-1}$

<sup>15</sup> 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)

<sup>16</sup> 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  $\|x\|_p \triangleq \left( \sum_{n=1}^N |x_n|^p \right)^{\frac{1}{p}}$ .

3. Proof that  $\|x_n + y_n\|_p \leq \|x_n\|_p + \|y_n\|_p$ :

$$\|x_n + y_n\|_p^p$$

$$= \sum_{n=1}^N |x_n + y_n|^p$$

by definition of  $\|\cdot\|_p$  (definition 2 page 302)

$$= \sum_{n=1}^N |x_n + y_n| |x_n + y_n|^{p-1}$$

by *homogeneous* property of  $|\cdot|$

$$\leq \sum_{n=1}^N |x_n| |x_n + y_n|^{p-1} + \sum_{n=1}^N |y_n| |x_n + y_n|^{p-1}$$

by *subadditive* property of  $|\cdot|$

$$= \sum_{n=1}^N |x_n(x_n + y_n)^{p-1}| + \sum_{n=1}^N |y_n(x_n + y_n)^{p-1}|$$

by *homogeneous* property of  $|\cdot|$

$$\leq \|x_n\|_p \|(x_n + y_n)^{p-1}\|_q + \|y_n\|_p \|(x_n + y_n)^{p-1}\|_q$$

by *Hölder's Inequality* (Theorem N.4 page 301)

$$= (\|x_n\|_p + \|y_n\|_p) \|(x_n + y_n)^{p-1}\|_q$$

$$= (\|x_n\|_p + \|y_n\|_p) \left( \sum_{n=1}^N |(x_n + y_n)^{p-1}|^q \right)^{\frac{1}{q}}$$

by definition of  $\|\cdot\|_p$  (definition 2 page 302)

$$= (\|x_n\|_p + \|y_n\|_p) \left( \sum_{n=1}^N |(x_n + y_n)^{p-1}|^{\frac{p}{p-1}} \right)^{\frac{p-1}{p}}$$

by definition 1

$$= (\|x_n\|_p + \|y_n\|_p) \left( \sum_{n=1}^N |(x_n + y_n)|^p \right)^{\frac{p-1}{p}}$$

by definition of  $\|\cdot\|_p$  (definition 2 page 302)

$$= (\|x_n\|_p + \|y_n\|_p) \|x_n + y_n\|^{p-1}$$

$$\implies \|x_n + y_n\|_p \leq \|x_n\|_p + \|y_n\|_p$$



“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 <sup>17</sup>

⇒

**Theorem N.6** (Cauchy-Schwarz Inequality for sequences). <sup>18</sup> Let  $\{x_n \in \mathbb{C}\}_1^N$  and  $\{y_n \in \mathbb{C}\}_1^N$  be complex  $N$ -tuples.

<sup>17</sup> 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](http://en.wikipedia.org/wiki/File:Augustin-Louis_Cauchy_1901.jpg), public domain

<sup>18</sup> [Aliprantis and Burkinshaw \(1998\)](#), page 278, [Scharz \(1885\)](#), [Bouniakowsky \(1859\)](#), [Hardy et al. \(1952\) page 25](#) (Theorem 11), [Cauchy \(1821\)](#), page 455 (???)

THM

$$\begin{aligned} \left| \sum_{n=1}^N x_n y_n^* \right|^2 &\leq \left( \sum_{n=1}^N |x_n|^2 \right) \left( \sum_{n=1}^N |y_n|^2 \right) & \forall x, y \in X \\ \left| \sum_{n=1}^N x_n y_n^* \right|^2 &= \left( \sum_{n=1}^N |x_n|^2 \right) \left( \sum_{n=1}^N |y_n|^2 \right) & \Leftrightarrow \exists a \in \mathbb{C} \text{ such that } y = ax & \forall x, y \in X \end{aligned}$$

PROOF:

1. The *Cauchy-Schwarz Inequality for sequences* is a special case of the *Hölder inequality* (Theorem N.4 page 301) 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)  $\langle x_n | y_n \rangle \triangleq \sum_{n=1}^N x_n y_n$  is an inner-product and  $(\|x_n\|, \langle \cdot | \cdot \rangle)$  is an inner-product space.
- (b) By the more general *Cauchy-Schwarz Inequality for inner-product spaces*,

$$\begin{aligned} \left( \sum_{n=1}^N a_n \lambda_n \right)^2 &\triangleq \langle a_n | \lambda_n \rangle^2 && \text{by definition of } \langle x_n | y_n \rangle \\ &\leq \|x_n\|^2 \|y_n\|^2 && \text{by Cauchy-Schwarz Inequality for inner-product spaces} \\ &\triangleq \left( \sum_{n=1}^N x_n^2 \right) \left( \sum_{n=1}^N y_n^2 \right) && \text{by definition of } \|\cdot\| \end{aligned}$$

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.<sup>19</sup>



### Proposition N.1. <sup>20</sup>

PRP

$$(x + y)^p \leq 2^p(x^p + y^p) \quad \forall x, y \geq 0, 1 < p < \infty$$

PROOF:

$$\begin{aligned} (x + y)^p &\leq (2 \max \{x, y\})^p \\ &= 2^p(\max \{x, y\})^p \\ &= 2^p(\max \{x^p, y^p\}) \\ &\leq 2^p(x^p + y^p) \end{aligned}$$



<sup>19</sup> Bullen (2003) pages 183–185 (Theorem 5)

<sup>20</sup> Carothers (2000), page 43

## N.4 Power Sums

**Theorem N.7** (Geometric Series). <sup>21</sup>

**T H M** 
$$\sum_{k=0}^{n-1} r^k = \frac{1 - r^n}{1 - r} \quad \forall r \in \mathbb{C} \setminus \{0\}$$

PROOF:

$$\begin{aligned} \left[ \sum_{k=0}^{n-1} r^k \right] &= \left( \frac{1}{1-r} \right) \left[ (1-r) \sum_{k=0}^{n-1} r^k \right] = \left( \frac{1}{1-r} \right) \left[ \sum_{k=0}^{n-1} r^k - r \sum_{k=0}^{n-1} r^k \right] = \left( \frac{1}{1-r} \right) \left[ \sum_{k=0}^{n-1} r^k - \left( \sum_{k=0}^{n-1} r^k - 1 + r^n \right) \right] \\ &= \left( \frac{1}{1-r} \right) [1 - r^n] = \boxed{\frac{1 - r^n}{1 - r}} \end{aligned}$$

⇒

**Lemma N.3.** Let  $f(x)$  be a function.

**L E M**  $S(x) \triangleq \sum_{n \in \mathbb{Z}} f(x + n\tau) = S(x + \tau) \quad (S(x) \text{ is PERIODIC with period } \tau)$

PROOF:

$$\begin{aligned} S(x + \tau) &\triangleq \sum_{n \in \mathbb{Z}} f(x + \tau + n\tau) = \sum_{n \in \mathbb{Z}} f(x + (n+1)\tau) = \sum_{m \in \mathbb{Z}} f(x + m\tau) \quad (\text{where } m \triangleq n+1) \\ &\triangleq S(x) \end{aligned}$$

⇒

**Proposition N.2** (Power Sums). <sup>22</sup>

**P R P** 
$$\begin{aligned} \sum_{k=1}^n k &= \frac{n(n+1)}{2} & \forall n \in \mathbb{N} \\ \sum_{k=1}^n k^2 &= \frac{n(n+1)(2n+1)}{6} & \forall n \in \mathbb{N} \end{aligned} \quad \begin{aligned} \sum_{k=1}^n k^3 &= \frac{n^2(n+1)^2}{4} & \forall n \in \mathbb{N} \\ \sum_{k=1}^n k^4 &= \frac{n(n+1)(2n+1)(3n^2+3n-1)}{30} & \forall n \in \mathbb{N} \end{aligned}$$

PROOF:

1. Proof that  $\sum_{k=1}^n k = \frac{n(n+1)}{2}$ : (proof by induction)

$$\begin{aligned} \sum_{k=1}^{n=1} k &= 1 = \frac{1(1+1)}{2} = \frac{n(n+1)}{2} \Big|_{n=1} \\ \sum_{k=1}^{n+1} k &= \left( \sum_{k=1}^n k \right) + (n+1) = \underbrace{\left( \frac{n(n+1)}{2} \right)}_{\text{by left hypothesis}} + (n+1) = (n+1) \left( \frac{n}{2} + 1 \right) \\ &= (n+1) \left( \frac{n+2}{2} \right) = \frac{(n+1)(n+2)}{2} \end{aligned}$$

<sup>21</sup>  Hall and Knight (1894), page 39 (article 55)

<sup>22</sup>  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)

$$\begin{aligned}\sum_{k=1}^{n=1} k^2 &= 1 = \frac{1(1+1)(2+1)}{6} = \frac{n(n+1)(2n+1)}{6} \Big|_{n=1} \\ \sum_{k=1}^{n+1} k^2 &= \left( \sum_{k=1}^n k^2 \right) + (n+1)^2 = \underbrace{\left( \frac{n(n+1)(2n+1)}{6} \right)}_{\text{by left hypothesis}} + (n+1)^2 = (n+1) \left( \frac{n(2n+1) + 6(n+1)}{6} \right) \\ &= (n+1) \left( \frac{2n^2 + 7n + 6}{6} \right) = (n+1) \left( \frac{(n+2)(2n+3)}{6} \right) = \frac{(n+1)[(n+1)+1][2(n+1)+1]}{6}\end{aligned}$$





# APPENDIX O

## 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.<sup>1</sup>

## O.1 Operators on linear spaces

### O.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 O.1.** <sup>2</sup>

**D E F** A function  $A$  in  $Y^X$  is an **operator** in  $Y^X$  if  
 $X$  and  $Y$  are both LINEAR SPACES (Definition G.1 page 197).

Two operators  $A$  and  $B$  in  $Y^X$  are **equal** if  $Ax = Bx$  for all  $x \in X$ . The inverse relation of an operator  $A$  in  $Y^X$  always exists as a *relation* in  $2^{XY}$ , but may not always be a *function* (may not always be an operator) in  $Y^X$ .

The operator  $I \in X^X$  is the *identity* operator if  $Ix = I$  for all  $x \in X$ .

**Definition O.2.** <sup>3</sup> Let  $X^X$  be the set of all operators with from a LINEAR SPACE  $X$  to  $X$ . Let  $I$  be an operator in  $X^X$ . Let  $\mathbb{I}(X)$  be the IDENTITY ELEMENT in  $X^X$ .

<sup>1</sup> quote: Leibniz (1679) pages 248–249

image: [http://en.wikipedia.org/wiki/File:Gottfried\\_Wilhelm\\_von\\_Leibniz.jpg](http://en.wikipedia.org/wiki/File:Gottfried_Wilhelm_von_Leibniz.jpg), public domain

<sup>2</sup> Heil (2011) page 42

<sup>3</sup> Michel and Herget (1993) page 411

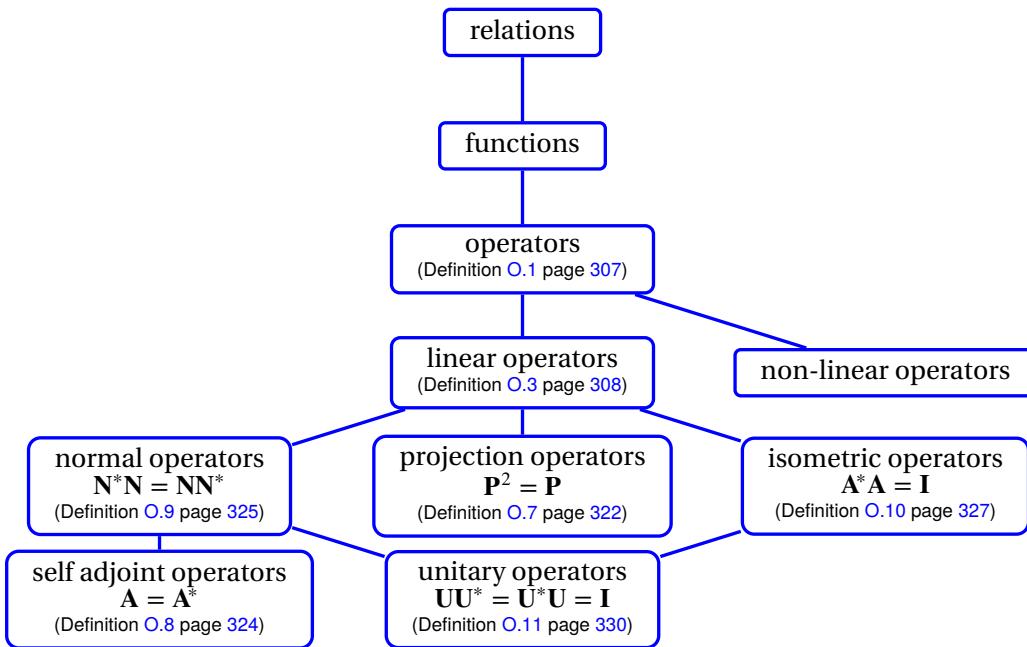


Figure O.1: Some operator types

DEF

**I** is the **identity operator** in  $\mathbf{X}^{\mathbf{X}}$  if  $\mathbf{I} = \mathbb{I}(\mathbf{X})$ .

## O.1.2 Linear operators

**Definition O.3.**<sup>4</sup> Let  $\mathbf{X} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$  and  $\mathbf{Y} \triangleq (Y, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}))$  be linear spaces.

DEF

An operator  $\mathbf{L} \in \mathbf{Y}^{\mathbf{X}}$  is **linear** if

1.  $\mathbf{L}(x + y) = \mathbf{L}x + \mathbf{L}y \quad \forall x, y \in \mathbf{X}$  (ADDITIVE) and
2.  $\mathbf{L}(\alpha x) = \alpha \mathbf{L}x \quad \forall x \in \mathbf{X}, \alpha \in \mathbb{F}$  (HOMOGENEOUS).

The set of all linear operators from  $\mathbf{X}$  to  $\mathbf{Y}$  is denoted  $\mathcal{L}(\mathbf{X}, \mathbf{Y})$  such that  $\mathcal{L}(\mathbf{X}, \mathbf{Y}) \triangleq \{\mathbf{L} \in \mathbf{Y}^{\mathbf{X}} | \mathbf{L} \text{ is linear}\}$ .

**Theorem O.1.**<sup>5</sup> Let  $\mathbf{L}$  be an operator from a linear space  $\mathbf{X}$  to a linear space  $\mathbf{Y}$ , both over a field  $\mathbb{F}$ .

THM

$$\{\mathbf{L} \text{ is LINEAR}\} \implies \left\{ \begin{array}{lcl} 1. \mathbf{L}\emptyset & = & \emptyset \\ 2. \mathbf{L}(-x) & = & -(\mathbf{L}x) \quad \forall x \in \mathbf{X} \quad \text{and} \\ 3. \mathbf{L}(x - y) & = & \mathbf{L}x - \mathbf{L}y \quad \forall x, y \in \mathbf{X} \quad \text{and} \\ 4. \mathbf{L}\left(\sum_{n=1}^N \alpha_n x_n\right) & = & \sum_{n=1}^N \alpha_n (\mathbf{L}x_n) \quad x_n \in \mathbf{X}, \alpha_n \in \mathbb{F} \end{array} \right\}$$

PROOF:

<sup>4</sup> Kubrusly (2001) page 55, Aliprantis and Burkinshaw (1998) page 224, Hilbert et al. (1927) page 6, Stone (1932) page 33

<sup>5</sup> Berberian (1961) page 79 (Theorem IV.1.1)

1. Proof that  $\mathbf{L}\mathbf{0} = \mathbf{0}$ :

$$\begin{aligned}\mathbf{L}\mathbf{0} &= \mathbf{L}(0 \cdot \mathbf{0}) && \text{by additive identity property} && (\text{Theorem G.1 page 199}) \\ &= 0 \cdot (\mathbf{L}\mathbf{0}) && \text{by } \textit{homogeneous} \text{ property of } \mathbf{L} && (\text{Definition O.3 page 308}) \\ &= \mathbf{0} && \text{by } \textit{additive identity} \text{ property} && (\text{Theorem G.1 page 199})\end{aligned}$$

2. Proof that  $\mathbf{L}(-\mathbf{x}) = -(\mathbf{Lx})$ :

$$\begin{aligned}\mathbf{L}(-\mathbf{x}) &= \mathbf{L}(-1 \cdot \mathbf{x}) && \text{by } \textit{additive inverse} \text{ property} && (\text{Theorem G.2 page 200}) \\ &= -1 \cdot (\mathbf{Lx}) && \text{by } \textit{homogeneous} \text{ property of } \mathbf{L} && (\text{Definition O.3 page 308}) \\ &= -(\mathbf{Lx}) && \text{by } \textit{additive inverse} \text{ property} && (\text{Theorem G.2 page 200})\end{aligned}$$

3. Proof that  $\mathbf{L}(\mathbf{x} - \mathbf{y}) = \mathbf{Lx} - \mathbf{Ly}$ :

$$\begin{aligned}\mathbf{L}(\mathbf{x} - \mathbf{y}) &= \mathbf{L}(\mathbf{x} + (-\mathbf{y})) && \text{by } \textit{additive inverse} \text{ property} && (\text{Theorem G.2 page 200}) \\ &= \mathbf{L}(\mathbf{x}) + \mathbf{L}(-\mathbf{y}) && \text{by } \textit{linearity} \text{ property of } \mathbf{L} && (\text{Definition O.3 page 308}) \\ &= \mathbf{Lx} - \mathbf{Ly} && \text{by item (2)} &&\end{aligned}$$

4. Proof that  $\mathbf{L}\left(\sum_{n=1}^N \alpha_n \mathbf{x}_n\right) = \sum_{n=1}^N \alpha_n (\mathbf{Lx}_n)$ :

(a) Proof for  $N = 1$ :

$$\begin{aligned}\mathbf{L}\left(\sum_{n=1}^N \alpha_n \mathbf{x}_n\right) &= \mathbf{L}(\alpha_1 \mathbf{x}_1) && \text{by } N = 1 \text{ hypothesis} \\ &= \alpha_1 (\mathbf{Lx}_1) && \text{by } \textit{homogeneous} \text{ property of } \mathbf{L} && (\text{Definition O.3 page 308})\end{aligned}$$

(b) Proof that  $N$  case  $\implies N + 1$  case:

$$\begin{aligned}\mathbf{L}\left(\sum_{n=1}^{N+1} \alpha_n \mathbf{x}_n\right) &= \mathbf{L}\left(\alpha_{N+1} \mathbf{x}_{N+1} + \sum_{n=1}^N \alpha_n \mathbf{x}_n\right) \\ &= \mathbf{L}(\alpha_{N+1} \mathbf{x}_{N+1}) + \mathbf{L}\left(\sum_{n=1}^N \alpha_n \mathbf{x}_n\right) && \text{by } \textit{linearity} \text{ property of } \mathbf{L} && (\text{Definition O.3 page 308}) \\ &= \alpha_{N+1} \mathbf{L}(\mathbf{x}_{N+1}) + \sum_{n=1}^N \mathbf{L}(\alpha_n \mathbf{x}_n) && \text{by left } N + 1 \text{ hypothesis} \\ &= \sum_{n=1}^{N+1} \mathbf{L}(\alpha_n \mathbf{x}_n)\end{aligned}$$



**Theorem O.2.**<sup>6</sup> Let  $\mathcal{L}(\mathbf{X}, \mathbf{Y})$  be the set of all linear operators from a linear space  $\mathbf{X}$  to a linear space  $\mathbf{Y}$ . Let  $\mathcal{N}(\mathbf{L})$  be the NULL SPACE of an operator  $\mathbf{L}$  in  $\mathbf{Y}^\mathbf{X}$  and  $\mathcal{I}(\mathbf{L})$  the IMAGE SET of  $\mathbf{L}$  in  $\mathbf{Y}^\mathbf{X}$ .

T H M	$\mathcal{L}(\mathbf{X}, \mathbf{Y})$ is a linear space $\mathcal{N}(\mathbf{L})$ is a linear subspace of $\mathbf{X}$ $\mathcal{I}(\mathbf{L})$ is a linear subspace of $\mathbf{Y}$	(space of linear transforms) $\forall \mathbf{L} \in \mathbf{Y}^\mathbf{X}$ $\forall \mathbf{L} \in \mathbf{Y}^\mathbf{X}$
-------------	---	--

PROOF:

<sup>6</sup> Michel and Herget (1993) pages 98–104, Berberian (1961) pages 80–85 (Theorem IV.1.4 and Theorem IV.3.1)

1. Proof that  $\mathcal{N}(\mathbf{L})$  is a linear subspace of  $\mathbf{X}$ :

- (a)  $0 \in \mathcal{N}(\mathbf{L}) \implies \mathcal{N}(\mathbf{L}) \neq \emptyset$
- (b)  $\mathcal{N}(\mathbf{L}) \triangleq \{x \in \mathbf{X} | \mathbf{L}x = 0\} \subseteq \mathbf{X}$
- (c)  $x + y \in \mathcal{N}(\mathbf{L}) \implies 0 = \mathbf{L}(x + y) = \mathbf{L}(y + x) \implies y + x \in \mathcal{N}(\mathbf{L})$
- (d)  $\alpha \in \mathbb{F}, x \in \mathbf{X} \implies 0 = \mathbf{L}x \implies 0 = \alpha \mathbf{L}x \implies 0 = \mathbf{L}(\alpha x) \implies \alpha x \in \mathcal{N}(\mathbf{L})$

2. Proof that  $\mathcal{I}(\mathbf{L})$  is a linear subspace of  $\mathbf{Y}$ :

- (a)  $0 \in \mathcal{I}(\mathbf{L}) \implies \mathcal{I}(\mathbf{L}) \neq \emptyset$
- (b)  $\mathcal{I}(\mathbf{L}) \triangleq \{y \in \mathbf{Y} | \exists x \in \mathbf{X} \text{ such that } y = \mathbf{L}x\} \subseteq \mathbf{Y}$
- (c)  $x + y \in \mathcal{I}(\mathbf{L}) \implies \exists v \in \mathbf{X} \text{ such that } \mathbf{L}v = x + y = y + x \implies y + x \in \mathcal{I}(\mathbf{L})$
- (d)  $\alpha \in \mathbb{F}, x \in \mathcal{I}(\mathbf{L}) \implies \exists x \in \mathbf{X} \text{ such that } y = \mathbf{L}x \implies \alpha y = \alpha \mathbf{L}x = \mathbf{L}(\alpha x) \implies \alpha x \in \mathcal{I}(\mathbf{L})$

⇒

*Example O.1.* <sup>7</sup> Let  $C([a : b], \mathbb{R})$  be the set of all *continuous* functions from the closed real interval  $[a : b]$  to  $\mathbb{R}$ .

**E** **X**  $C([a : b], \mathbb{R})$  is a linear space.

**Theorem O.3.** <sup>8</sup> Let  $\mathcal{L}(\mathbf{X}, \mathbf{Y})$  be the set of linear operators from a linear space  $\mathbf{X}$  to a linear space  $\mathbf{Y}$ . Let  $\mathcal{N}(\mathbf{L})$  be the NULL SPACE of a linear operator  $\mathbf{L} \in \mathcal{L}(\mathbf{X}, \mathbf{Y})$ .

<b>T</b>	$\mathbf{L}x = \mathbf{L}y \iff x - y \in \mathcal{N}(\mathbf{L})$
<b>H</b>	$\mathbf{L}$ is INJECTIVE $\iff \mathcal{N}(\mathbf{L}) = \{0\}$

PROOF:

1. Proof that  $\mathbf{L}x = \mathbf{L}y \implies x - y \in \mathcal{N}(\mathbf{L})$ :

$$\begin{aligned} \mathbf{L}(x - y) &= \mathbf{L}x - \mathbf{L}y && \text{by Theorem O.1 page 308} \\ &= 0 && \text{by left hypothesis} \\ &\implies x - y \in \mathcal{N}(\mathbf{L}) && \text{by definition of null space} \end{aligned}$$

2. Proof that  $\mathbf{L}x = \mathbf{L}y \iff x - y \in \mathcal{N}(\mathbf{L})$ :

$$\begin{aligned} \mathbf{L}y &= \mathbf{L}y + 0 && \text{by definition of linear space (Definition G.1 page 197)} \\ &= \mathbf{L}y + \mathbf{L}(x - y) && \text{by right hypothesis} \\ &= \mathbf{L}y + (\mathbf{L}x - \mathbf{L}y) && \text{by Theorem O.1 page 308} \\ &= (\mathbf{L}y - \mathbf{L}y) + \mathbf{L}x && \text{by associative and commutative properties (Definition G.1 page 197)} \\ &= \mathbf{L}x \end{aligned}$$

3. Proof that  $\mathbf{L}$  is *injective*  $\iff \mathcal{N}(\mathbf{L}) = \{0\}$ :

$$\begin{aligned} \mathbf{L} \text{ is injective} &\iff \{(\mathbf{L}x = \mathbf{L}y \iff x = y) \quad \forall x, y \in \mathbf{X}\} \\ &\iff \{[\mathbf{L}x - \mathbf{L}y = 0 \iff (x - y) = 0] \quad \forall x, y \in \mathbf{X}\} \\ &\iff \{[\mathbf{L}(x - y) = 0 \iff (x - y) = 0] \quad \forall x, y \in \mathbf{X}\} \\ &\iff \mathcal{N}(\mathbf{L}) = \{0\} \end{aligned}$$

<sup>7</sup> Eidelman et al. (2004) page 3

<sup>8</sup> Berberian (1961) page 88 (Theorem IV.1.4)



**Theorem O.4.**<sup>9</sup> Let  $\mathcal{W}$ ,  $\mathcal{X}$ ,  $\mathcal{Y}$ , and  $\mathcal{Z}$  be linear spaces over a field  $\mathbb{F}$ .

T H M	1. $L(MN) = (LM)N$ 2. $L(M \dotplus N) = (LM) \dotplus (LN)$ 3. $(L \dotplus M)N = (LN) \dotplus (MN)$ 4. $\alpha(LM) = (\alpha L)M = L(\alpha M)$	$\forall L \in \mathcal{L}(\mathcal{Z}, \mathcal{W}), M \in \mathcal{L}(\mathcal{Y}, \mathcal{Z}), N \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ $\forall L \in \mathcal{L}(\mathcal{Y}, \mathcal{Z}), M \in \mathcal{L}(\mathcal{X}, \mathcal{Y}), N \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ $\forall L \in \mathcal{L}(\mathcal{Y}, \mathcal{Z}), M \in \mathcal{L}(\mathcal{Y}, \mathcal{Z}), N \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ $\forall L \in \mathcal{L}(\mathcal{Y}, \mathcal{Z}), M \in \mathcal{L}(\mathcal{X}, \mathcal{Y}), \alpha \in \mathbb{F}$	(ASSOCIATIVE) (LEFT DISTRIBUTIVE) (RIGHT DISTRIBUTIVE) (HOMOGENEOUS)
-------------	---	--	---

PROOF:

1. Proof that  $L(MN) = (LM)N$ : Follows directly from property of *associative* operators.

2. Proof that  $L(M \dotplus N) = (LM) \dotplus (LN)$ :

$$\begin{aligned} [L(M \dotplus N)]x &= L[(M \dotplus N)x] \\ &= L[(Mx) \dotplus (Nx)] \\ &= [L(Mx)] \dotplus [L(Nx)] \quad \text{by } \textit{additive} \text{ property Definition O.3 page 308} \\ &= [(LM)x] \dotplus [(LN)x] \end{aligned}$$

3. Proof that  $(L \dotplus M)N = (LN) \dotplus (MN)$ : Follows directly from property of *associative* operators.

4. Proof that  $\alpha(LM) = (\alpha L)M$ : Follows directly from *associative* property of linear operators.

5. Proof that  $\alpha(LM) = L(\alpha M)$ :

$$\begin{aligned} [\alpha(LM)]x &= \alpha[(LM)x] \\ &= L[\alpha(Mx)] \quad \text{by } \textit{homogeneous} \text{ property Definition O.3 page 308} \\ &= L[(\alpha M)x] \\ &= [L(\alpha M)]x \end{aligned}$$



**Theorem O.5** (Fundamental theorem of linear equations). Michel and Herget (1993) page 99 Let  $\mathcal{Y}^{\mathcal{X}}$  be the set of all operators from a linear space  $\mathcal{X}$  to a linear space  $\mathcal{Y}$ . Let  $\mathcal{N}(L)$  be the NULL SPACE of an operator  $L$  in  $\mathcal{Y}^{\mathcal{X}}$  and  $\mathcal{I}(L)$  the IMAGE SET of  $L$  in  $\mathcal{Y}^{\mathcal{X}}$  (Definition ?? page ??).

T H M	$\dim \mathcal{I}(L) + \dim \mathcal{N}(L) = \dim \mathcal{X}$ $\forall L \in \mathcal{Y}^{\mathcal{X}}$
-------------	--

PROOF: Let  $\{\psi_k | k = 1, 2, \dots, p\}$  be a basis for  $\mathcal{X}$  constructed such that  $\{\psi_{p-n+1}, \psi_{p-n+2}, \dots, \psi_p\}$  is a basis for

<sup>9</sup> Berberian (1961) page 88 (Theorem IV.5.1)

$\mathcal{N}(\mathbf{L})$ .

Let  $p \triangleq \dim \mathbf{X}$ .

Let  $n \triangleq \dim \mathcal{N}(\mathbf{L})$ .

$$\begin{aligned}
 \dim \mathcal{I}(\mathbf{L}) &= \dim \{y \in \mathbf{Y} \mid \exists x \in \mathbf{X} \text{ such that } y = \mathbf{L}x\} \\
 &= \dim \left\{ y \in \mathbf{Y} \mid \exists (\alpha_1, \alpha_2, \dots, \alpha_p) \text{ such that } y = \mathbf{L} \sum_{k=1}^p \alpha_k \psi_k \right\} \\
 &= \dim \left\{ y \in \mathbf{Y} \mid \exists (\alpha_1, \alpha_2, \dots, \alpha_p) \text{ such that } y = \sum_{k=1}^p \alpha_k \mathbf{L}\psi_k \right\} \\
 &= \dim \left\{ y \in \mathbf{Y} \mid \exists (\alpha_1, \alpha_2, \dots, \alpha_p) \text{ such that } y = \sum_{k=1}^{p-n} \alpha_k \mathbf{L}\psi_k + \sum_{k=1}^n \alpha_k \mathbf{L}\psi_k \right\} \\
 &= \dim \left\{ y \in \mathbf{Y} \mid \exists (\alpha_1, \alpha_2, \dots, \alpha_p) \text{ such that } y = \sum_{k=1}^{p-n} \alpha_k \mathbf{L}\psi_k + \mathbb{0} \right\} \\
 &= p - n \\
 &= \dim \mathbf{X} - \dim \mathcal{N}(\mathbf{L})
 \end{aligned}$$

Note: This “proof” may be missing some necessary detail.



## O.2 Operators on Normed linear spaces

### O.2.1 Operator norm

**Definition O.4.** <sup>10</sup> Let  $\mathcal{L}(\mathbf{X}, \mathbf{Y})$  be the space of linear operators over normed linear spaces  $\mathbf{X}$  and  $\mathbf{Y}$ .  
<sup>11</sup>

**D E F** The **operator norm**  $\|\cdot\|$  is defined as  

$$\|\mathbf{A}\| \triangleq \sup_{x \in \mathbf{X}} \{\|\mathbf{Ax}\| \mid \|x\| \leq 1\} \quad \forall \mathbf{A} \in \mathcal{L}(\mathbf{X}, \mathbf{Y})$$
  
The pair  $(\mathcal{L}(\mathbf{X}, \mathbf{Y}), \|\cdot\|)$  is the **normed space of linear operators** on  $(\mathbf{X}, \mathbf{Y})$ .

Proposition O.1 (next) shows that the functional defined in Definition O.4 (previous) is a **norm** (Definition L.1 page 275).

**Proposition O.1.** <sup>12</sup> Let  $(\mathcal{L}(\mathbf{X}, \mathbf{Y}), \|\cdot\|)$  be the normed space of linear operators over the normed linear spaces  $\mathbf{X} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$  and  $\mathbf{Y} \triangleq (Y, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$ .

**P R P** The functional  $\|\cdot\|$  is a **norm** on  $\mathcal{L}(\mathbf{X}, \mathbf{Y})$ . In particular,

1.  $\|\mathbf{A}\| \geq 0 \quad \forall \mathbf{A} \in \mathcal{L}(\mathbf{X}, \mathbf{Y}) \quad (\text{NON-NEGATIVE}) \quad \text{and}$
2.  $\|\mathbf{A}\| = 0 \iff \mathbf{A} \stackrel{\circ}{=} \mathbb{0} \quad \forall \mathbf{A} \in \mathcal{L}(\mathbf{X}, \mathbf{Y}) \quad (\text{NONDEGENERATE}) \quad \text{and}$
3.  $\|\alpha \mathbf{A}\| = |\alpha| \|\mathbf{A}\| \quad \forall \mathbf{A} \in \mathcal{L}(\mathbf{X}, \mathbf{Y}), \alpha \in \mathbb{F} \quad (\text{HOMOGENEOUS}) \quad \text{and}$
4.  $\|\mathbf{A} \dot{+} \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\| \quad \forall \mathbf{A} \in \mathcal{L}(\mathbf{X}, \mathbf{Y}) \quad (\text{SUBADDITIVE}).$

Moreover,  $(\mathcal{L}(\mathbf{X}, \mathbf{Y}), \|\cdot\|)$  is a **normed linear space**.

<sup>10</sup> Rudin (1991) page 92, Aliprantis and Burkinshaw (1998) page 225

<sup>11</sup> The operator norm notation  $\|\cdot\|$  is introduced (as a Matrix norm) in

Horn and Johnson (1990), page 290

<sup>12</sup> Rudin (1991) page 93

PROOF:

1. Proof that  $\|A\| > 0$  for  $A \neq 0$ :

$$\begin{aligned} \|A\| &\triangleq \sup_{x \in X} \{\|Ax\| \mid \|x\| \leq 1\} \\ &> 0 \end{aligned} \quad \text{by definition of } \|\cdot\| \text{ (Definition O.4 page 312)}$$

2. Proof that  $\|A\| = 0$  for  $A = 0$ :

$$\begin{aligned} \|A\| &\triangleq \sup_{x \in X} \{\|Ax\| \mid \|x\| \leq 1\} \\ &= \sup_{x \in X} \{\|0x\| \mid \|x\| \leq 1\} \\ &= 0 \end{aligned} \quad \text{by definition of } \|\cdot\| \text{ (Definition O.4 page 312)}$$

3. Proof that  $\|\alpha A\| = |\alpha| \|A\|$ :

$$\begin{aligned} \|\alpha A\| &\triangleq \sup_{x \in X} \{\|\alpha Ax\| \mid \|x\| \leq 1\} \\ &= \sup_{x \in X} \{|\alpha| \|Ax\| \mid \|x\| \leq 1\} \\ &= |\alpha| \sup_{x \in X} \{\|Ax\| \mid \|x\| \leq 1\} \\ &= |\alpha| \|A\| \end{aligned} \quad \begin{array}{l} \text{by definition of } \|\cdot\| \text{ (Definition O.4 page 312)} \\ \text{by definition of } \|\cdot\| \text{ (Definition O.4 page 312)} \\ \text{by definition of sup} \\ \text{by definition of } \|\cdot\| \text{ (Definition O.4 page 312)} \end{array}$$

4. Proof that  $\|A + B\| \leq \|A\| + \|B\|$ :

$$\begin{aligned} \|A + B\| &\triangleq \sup_{x \in X} \{\|(A + B)x\| \mid \|x\| \leq 1\} \\ &= \sup_{x \in X} \{\|Ax + Bx\| \mid \|x\| \leq 1\} \\ &\leq \sup_{x \in X} \{\|Ax\| + \|Bx\| \mid \|x\| \leq 1\} \\ &\leq \sup_{x \in X} \{\|Ax\| \mid \|x\| \leq 1\} + \sup_{x \in X} \{\|Bx\| \mid \|x\| \leq 1\} \\ &\triangleq \|A\| + \|B\| \end{aligned} \quad \begin{array}{l} \text{by definition of } \|\cdot\| \text{ (Definition O.4 page 312)} \\ \text{by definition of } \|\cdot\| \text{ (Definition O.4 page 312)} \\ \text{by definition of } \|\cdot\| \text{ (Definition O.4 page 312)} \end{array}$$



**Lemma O.1.** Let  $(\mathcal{L}(X, Y), \|\cdot\|)$  be the normed space of linear operators over normed linear spaces  $X \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$  and  $Y \triangleq (Y, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$ .

L  
E  
M

$$\|L\| = \sup_x \{\|Lx\| \mid \|x\| = 1\} \quad \forall x \in \mathcal{L}(X, Y)$$

PROOF: <sup>13</sup>

1. Proof that  $\sup_x \{\|Lx\| \mid \|x\| \leq 1\} \geq \sup_x \{\|Lx\| \mid \|x\| = 1\}$ :

$$\sup_x \{\|Lx\| \mid \|x\| \leq 1\} \geq \sup_x \{\|Lx\| \mid \|x\| = 1\} \quad \text{because } A \subseteq B \implies \sup A \leq \sup B$$

13



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

$$Y \triangleq \left\{ y \in X \mid \begin{array}{l} 1. \quad \|Ly\| = \sup_{x \in X} \{\|Lx\| \mid \|x\| \leq 1\} \text{ and} \\ 2. \quad 0 < \|y\| \leq 1 \end{array} \right\}$$

3. Proof that  $\sup_x \{\|Lx\| \mid \|x\| \leq 1\} \leq \sup_x \{\|Lx\| \mid \|x\| = 1\}$ :

$$\begin{aligned} \sup_{x \in X} \{\|Lx\| \mid \|x\| \leq 1\} &= \|Ly\| && \text{by definition of set } Y \\ &= \frac{\|y\|}{\|y\|} \|Ly\| \\ &= \|y\| \left\| \frac{1}{\|y\|} Ly \right\| && \text{by homogeneous property (page 275)} \\ &= \|y\| \left\| L \frac{y}{\|y\|} \right\| && \text{by homogeneous property (page 308)} \\ &\leq \|y\| \sup_{y \in Y} \left\{ \left\| L \frac{y}{\|y\|} \right\| \right\} && \text{by definition of supremum} \\ &= \|y\| \sup_{y \in Y} \left\{ \left\| L \frac{y}{\|y\|} \right\| \mid \left\| \frac{y}{\|y\|} \right\| = 1 \right\} && \text{because } \left\| \frac{y}{\|y\|} \right\| = 1 \text{ for all } y \in Y \\ &\leq \sup_{y \in Y} \left\{ \left\| L \frac{y}{\|y\|} \right\| \mid \left\| \frac{y}{\|y\|} \right\| = 1 \right\} && \text{because } 0 < \|y\| \leq 1 \\ &\leq \sup_{x \in X} \{\|Lx\| \mid \|x\| = 1\} && \text{because } \frac{y}{\|y\|} \in X \quad \forall y \in Y \end{aligned}$$

4. By (1) and (3),

$$\sup_{x \in X} \{\|Lx\| \mid \|x\| \leq 1\} = \sup_{x \in X} \{\|Lx\| \mid \|x\| = 1\}$$



**Proposition O.2.** <sup>14</sup> Let  $\mathbf{I}$  be the identity operator in the normed space of linear operators  $(\mathcal{L}(X, X), \|\cdot\|)$ .

P	R	P
		$\ \mathbf{I}\  = 1$

PROOF:

$$\begin{aligned} \|\mathbf{I}\| &\triangleq \sup \{\|\mathbf{Ix}\| \mid \|x\| \leq 1\} && \text{by definition of } \|\cdot\| \text{ (Definition O.4 page 312)} \\ &= \sup \{\|x\| \mid \|x\| \leq 1\} && \text{by definition of } \mathbf{I} \text{ (Definition O.2 page 307)} \\ &= 1 \end{aligned}$$



**Theorem O.6.** <sup>15</sup> Let  $(\mathcal{L}(X, Y), \|\cdot\|)$  be the normed space of linear operators over normed linear spaces  $X$  and  $Y$ .

T	$\ Lx\  \leq \ \mathbf{L}\  \ x\  \quad \forall L \in \mathcal{L}(X, Y), x \in X$
H	$\ \mathbf{KL}\  \leq \ \mathbf{K}\  \ \mathbf{L}\  \quad \forall K, L \in \mathcal{L}(X, Y)$

<sup>14</sup> Michel and Herget (1993) page 410

<sup>15</sup> Rudin (1991) page 103, Aliprantis and Burkinshaw (1998) page 225

PROOF:

1. Proof that  $\|\mathbf{L}\mathbf{x}\| \leq \|\mathbf{L}\| \|\mathbf{x}\|$ :

$$\begin{aligned}
 \|\mathbf{L}\mathbf{x}\| &= \frac{\|\mathbf{x}\|}{\|\mathbf{x}\|} \|\mathbf{L}\mathbf{x}\| \\
 &= \|\mathbf{x}\| \left\| \frac{1}{\|\mathbf{x}\|} \mathbf{L}\mathbf{x} \right\| \\
 &= \|\mathbf{x}\| \left\| \mathbf{L} \frac{\mathbf{x}}{\|\mathbf{x}\|} \right\| \\
 &\triangleq \|\mathbf{x}\| \|\mathbf{Ly}\| \\
 &\leq \|\mathbf{x}\| \sup_y \|\mathbf{Ly}\| \\
 &= \|\mathbf{x}\| \sup_y \{ \|\mathbf{Ly}\| \mid \|\mathbf{y}\| = 1 \} \\
 &\triangleq \|\mathbf{x}\| \|\mathbf{L}\|
 \end{aligned}$$

by property of norms  
by property of linear operators  
where  $y \triangleq \frac{\mathbf{x}}{\|\mathbf{x}\|}$   
by definition of supremum  
because  $\|\mathbf{y}\| = \left\| \frac{\mathbf{x}}{\|\mathbf{x}\|} \right\| = \frac{\|\mathbf{x}\|}{\|\mathbf{x}\|} = 1$   
by definition of operator norm

2. Proof that  $\|\mathbf{KL}\| \leq \|\mathbf{K}\| \|\mathbf{L}\|$ :

$$\begin{aligned}
 \|\mathbf{KL}\| &\triangleq \sup_{x \in X} \{ \|(\mathbf{KL})\mathbf{x}\| \mid \|\mathbf{x}\| \leq 1 \} \\
 &= \sup_{x \in X} \{ \|\mathbf{K}(\mathbf{L}\mathbf{x})\| \mid \|\mathbf{x}\| \leq 1 \} \\
 &\leq \sup_{x \in X} \{ \|\mathbf{K}\| \|\mathbf{L}\mathbf{x}\| \mid \|\mathbf{x}\| \leq 1 \} \\
 &\leq \sup_{x \in X} \{ \|\mathbf{K}\| \|\mathbf{L}\| \|\mathbf{x}\| \mid \|\mathbf{x}\| \leq 1 \} \\
 &= \sup_{x \in X} \{ \|\mathbf{K}\| \|\mathbf{L}\| 1 \mid \|\mathbf{x}\| \leq 1 \} \\
 &= \|\mathbf{K}\| \|\mathbf{L}\|
 \end{aligned}$$

by Definition O.4 page 312 ( $\|\cdot\|$ )  
by 1.  
by 1.  
by definition of sup  
by definition of sup

## O.2.2 Bounded linear operators

**Definition O.5.** <sup>16</sup> Let  $(\mathcal{L}(X, Y), \|\cdot\|)$  be a normed space of linear operators.

**D E F** An operator  $\mathbf{B}$  is **bounded** if  $\|\mathbf{B}\| < \infty$ .

The quantity  $\mathcal{B}(X, Y)$  is the set of all **bounded linear operators** on  $(X, Y)$  such that  $\mathcal{B}(X, Y) \triangleq \{ \mathbf{L} \in \mathcal{L}(X, Y) \mid \|\mathbf{L}\| < \infty \}$ .

**Theorem O.7.** <sup>17</sup> Let  $(\mathcal{L}(X, Y), \|\cdot\|)$  be the set of linear operators over normed linear spaces  $X \triangleq (X, +, \cdot, (\mathbb{F}, +, \times), \|\cdot\|)$  and  $Y \triangleq (Y, +, \cdot, (\mathbb{F}, +, \times), \|\cdot\|)$ .

The following conditions are all EQUIVALENT:

- |              |  |
|--------------|--|
| <b>T H M</b> | 1. $\mathbf{L}$ is continuous at A SINGLE POINT $\mathbf{x}_0 \in X$ $\forall \mathbf{L} \in \mathcal{L}(X, Y)$ $\iff$<br>2. $\mathbf{L}$ is CONTINUOUS (at every point $\mathbf{x} \in X$ ) $\forall \mathbf{L} \in \mathcal{L}(X, Y)$ $\iff$<br>3. $\ \mathbf{L}\  < \infty$ ( $\mathbf{L}$ is BOUNDED) $\forall \mathbf{L} \in \mathcal{L}(X, Y)$ $\iff$<br>4. $\exists M \in \mathbb{R}$ such that $\ \mathbf{L}\mathbf{x}\  \leq M \ \mathbf{x}\ $ $\forall \mathbf{L} \in \mathcal{L}(X, Y), \mathbf{x} \in X$ |
|--------------|--|

<sup>16</sup> Rudin (1991) pages 92–93

<sup>17</sup> Aliprantis and Burkinshaw (1998) page 227

PROOF:

1. Proof that 1  $\implies$  2:

$$\begin{aligned}
 \epsilon &> \|Lx - Lx_0\| && \text{by hypothesis 1} \\
 &= \|L(x - x_0)\| && \text{by linearity (Definition O.3 page 308)} \\
 &= \|L(x + y - x_0 - y)\| \\
 &= \|L(x + y) - L(x_0 + y)\| && \text{by linearity (Definition O.3 page 308)} \\
 \implies L &\text{ is continuous at point } x + y \\
 \implies L &\text{ is continuous at every point in } X && \text{(hypothesis 2)}
 \end{aligned}$$

2. Proof that 2  $\implies$  1: obvious:

3. Proof that 4  $\implies$  2:<sup>18</sup>

$$\begin{aligned}
 \|Lx\| \leq M \|x\| &\implies \|L(x - y)\| \leq M \|x - y\| && \text{by hypothesis 4} \\
 &\implies \|Lx - Ly\| \leq M \|x - y\| && \text{by linearity of } L \text{ (Definition O.3 page 308)} \\
 &\implies \|Lx - Ly\| \leq \epsilon \text{ whenever } M \|x - y\| < \epsilon \\
 &\implies \|Lx - Ly\| \leq \epsilon \text{ whenever } \|x - y\| < \frac{\epsilon}{M} && \text{(hypothesis 2)}
 \end{aligned}$$

4. Proof that 3  $\implies$  4:

$$\begin{aligned}
 \|Lx\| &\leq \underbrace{\|L\|}_M \|x\| && \text{by Theorem O.6 page 314} \\
 &= M \|x\| && \text{where } M \triangleq \|L\| < \infty \text{ (by hypothesis 1)}
 \end{aligned}$$

5. Proof that 1  $\implies$  3:<sup>19</sup>

$$\begin{aligned}
 \|L\| = \infty &\implies \{\|Lx\| \mid \|x\| \leq 1\} = \infty \\
 &\implies \exists (x_n) \text{ such that } \|x_n\| = 1 \text{ and } \|L\| = \{\|Lx_n\| \mid \|x_n\| \leq 1\} = \infty \\
 &\implies \|x_n\| = 1 \text{ and } \infty = \|L\| = \|Lx_n\| \\
 &\implies \|x_n\| = 1 \text{ and } \|Lx_n\| \geq n \\
 &\implies \frac{1}{n} \|x_n\| = \frac{1}{n} \text{ and } \frac{1}{n} \|Lx_n\| \geq 1 \\
 &\implies \left\| \frac{x_n}{n} \right\| = \frac{1}{n} \text{ and } \left\| L \frac{x_n}{n} \right\| \geq 1 \\
 &\implies \lim_{n \rightarrow \infty} \left\| \frac{x_n}{n} \right\| = 0 \text{ and } \lim_{n \rightarrow \infty} \left\| L \frac{x_n}{n} \right\| \geq 1 \\
 &\implies L \text{ is not continuous at } 0
 \end{aligned}$$

But by hypothesis,  $L$  is continuous. So the statement  $\|L\| = \infty$  must be *false* and thus  $\|L\| < \infty$  ( $L$  is *bounded*).

<sup>18</sup> Bollobás (1999), page 29

<sup>19</sup> Aliprantis and Burkinshaw (1998), page 227

### O.2.3 Adjoint on normed linear spaces

**Definition O.6.** Let  $\mathcal{B}(X, Y)$  be the space of BOUNDED LINEAR OPERATORS on normed linear spaces  $X$  and  $Y$ . Let  $X^*$  be the TOPOLOGICAL DUAL SPACE of  $X$ .

**DEF**  $B^*$  is the **adjoint** of an operator  $B \in \mathcal{B}(X, Y)$  if  
 $f(Bx) = [B^*f](x) \quad \forall f \in X^*, x \in X$

**Theorem O.8.** <sup>20</sup> Let  $\mathcal{B}(X, Y)$  be the space of BOUNDED LINEAR OPERATORS on NORMED LINEAR SPACES  $X$  and  $Y$ .

**THM**  $(A + B)^* = A^* + B^* \quad \forall A, B \in \mathcal{B}(X, Y)$   
 $(\lambda A)^* = \lambda A^* \quad \forall A, B \in \mathcal{B}(X, Y)$   
 $(AB)^* = B^*A^* \quad \forall A, B \in \mathcal{B}(X, Y)$

PROOF:

$$[A + B]^*f(x) = f([A + B]x) \quad \text{by definition of adjoint} \quad (\text{Definition O.6 page 317})$$

$$= f(Ax + Bx) \quad \text{by definition of linear operators} \quad (\text{Definition O.3 page 308})$$

$$= f(Ax) + f(Bx) \quad \text{by definition of linear functional}$$

$$= A^*f(x) + B^*f(x) \quad \text{by definition of adjoint} \quad (\text{Definition O.6 page 317})$$

$$= [A^* + B^*]f(x) \quad \text{by definition of linear functional}$$

$$[\lambda A]^*f(x) = f([\lambda A]x) \quad \text{by definition of adjoint} \quad (\text{Definition O.6 page 317})$$

$$= \lambda f(Ax) \quad \text{by definition of linear functional}$$

$$= [\lambda A^*]f(x) \quad \text{by definition of adjoint}$$

$$[AB]^*f(x) = f([AB]x) \quad \text{by definition of adjoint} \quad (\text{Definition O.6 page 317})$$

$$= f(A[Bx]) \quad \text{by definition of linear operators} \quad (\text{Definition O.3 page 308})$$

$$= [A^*f](Bx) \quad \text{by definition of adjoint} \quad (\text{Definition O.6 page 317})$$

$$= B^*[A^*f](x) \quad \text{by definition of adjoint} \quad (\text{Definition O.6 page 317})$$

$$= [B^*A^*]f(x) \quad \text{by definition of adjoint} \quad (\text{Definition O.6 page 317})$$

**Theorem O.9.** <sup>21</sup> Let  $\mathcal{B}(X, Y)$  be the space of BOUNDED LINEAR OPERATORS on normed linear spaces  $X$  and  $Y$ . Let  $B^*$  be the adjoint of an operator  $B$ .

**THM**  $\|B\| = \|B^*\| \quad \forall B \in \mathcal{B}(X, Y)$

PROOF:

$$\|B\| \triangleq \sup \{\|Bx\| \mid \|x\| \leq 1\} \quad \text{by Definition O.4 page 312}$$

$$\triangleq \sup \{|g(Bx; y^*)| \mid \|x\| \leq 1, \|y^*\| \leq 1\}$$

$$\triangleq \sup \{|f(x; B^*y^*)| \mid \|x\| \leq 1, \|y^*\| \leq 1\}$$

$$\triangleq \sup \{\|B^*y^*\| \mid \|x\| \leq 1, \|y^*\| \leq 1\}$$

$$\triangleq \sup \{\|B^*y^*\| \mid \|y^*\| \leq 1\}$$

$$\triangleq \|B^*\| \quad \text{by Definition O.4 page 312}$$

<sup>20</sup> Bollobás (1999), page 156

<sup>21</sup> Rudin (1991) page 98

## O.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.”<sup>22</sup>

Stanislaus M. Ulam (1909–1984), Polish mathematician <sup>22</sup>

**Theorem O.10** (Mazur-Ulam theorem). <sup>23</sup> Let  $\phi \in \mathcal{L}(\mathbf{X}, \mathbf{Y})$  be a function on normed linear spaces  $(\mathbf{X}, \|\cdot\|_{\mathbf{X}})$  and  $(\mathbf{Y}, \|\cdot\|_{\mathbf{Y}})$ . Let  $\mathbf{I} \in \mathcal{L}(\mathbf{X}, \mathbf{X})$  be the identity operator on  $(\mathbf{X}, \|\cdot\|_{\mathbf{X}})$ .

T H M	$\left. \begin{array}{l} 1. \underbrace{\phi^{-1}\phi = \phi\phi^{-1} = \mathbf{I}}_{\text{bijective}} \\ 2. \underbrace{\ \phi\mathbf{x} - \phi\mathbf{y}\ _{\mathbf{Y}} = \ \mathbf{x} - \mathbf{y}\ _{\mathbf{X}}}_{\text{isometric}} \end{array} \right\} \text{and} \quad \Rightarrow \underbrace{\phi([1 - \lambda]\mathbf{x} + \lambda\mathbf{y}) = [1 - \lambda]\phi\mathbf{x} + \lambda\phi\mathbf{y} \forall \lambda \in \mathbb{R}}_{\text{affine}}$
-------------	---

PROOF: Proof not yet complete.

1. Let  $\psi$  be the reflection of  $\mathbf{z}$  in  $\mathbf{X}$  such that  $\psi\mathbf{x} = 2\mathbf{z} - \mathbf{x}$

$$(a) \|\psi\mathbf{x} - \mathbf{z}\| = \|\mathbf{x} - \mathbf{z}\|$$

2. Let  $\lambda \triangleq \sup_g \{\|g\mathbf{z} - \mathbf{z}\|\}$

3. Proof that  $g \in W \implies g^{-1} \in W$ :

Let  $\hat{\mathbf{x}} \triangleq g^{-1}\mathbf{x}$  and  $\hat{\mathbf{y}} \triangleq g^{-1}\mathbf{y}$ .

$$\begin{aligned} \|g^{-1}\mathbf{x} - g^{-1}\mathbf{y}\| &= \|\hat{\mathbf{x}} - \hat{\mathbf{y}}\| && \text{by definition of } \hat{\mathbf{x}} \text{ and } \hat{\mathbf{y}} \\ &= \|g\hat{\mathbf{x}} - g\hat{\mathbf{y}}\| && \text{by left hypothesis} \\ &= \|gg^{-1}\mathbf{x} - gg^{-1}\mathbf{y}\| && \text{by definition of } \hat{\mathbf{x}} \text{ and } \hat{\mathbf{y}} \\ &= \|\mathbf{x} - \mathbf{y}\| && \text{by definition of } g^{-1} \end{aligned}$$

<sup>22</sup> quote: [Ulam \(1991\)](#), page 33

image: <http://www-history.mcs.st-andrews.ac.uk/Biographies/Ulam.html>

<sup>23</sup> [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  $gz = z$ :

$$\begin{aligned}
 2\lambda &= 2 \sup \{ \|gz - z\| \} && \text{by definition of } \lambda \text{ item (2)} \\
 &\leq 2 \|gz - z\| && \text{by definition of sup} \\
 &= \|2z - 2gz\| && \\
 &= \|\psi gz - gz\| && \text{by definition of } \psi \text{ item (1)} \\
 &= \|g^{-1}\psi gz - g^{-1}gz\| && \text{by item (3)} \\
 &= \|g^{-1}\psi gz - z\| && \text{by definition of } g^{-1} \\
 &= \|\psi g^{-1}\psi gz - z\| && \\
 &= \|g^*z - z\| && \\
 &\leq \lambda && \text{by definition of } \lambda \text{ item (2)} \\
 &\implies 2\lambda \leq \lambda \\
 &\implies \lambda = 0 \\
 &\implies gz = z
 \end{aligned}$$

5. Proof that  $\phi\left(\frac{1}{2}\mathbf{x} + \frac{1}{2}\mathbf{y}\right) = \frac{1}{2}\phi\mathbf{x} + \frac{1}{2}\phi\mathbf{y}$ :

$$\begin{aligned}
 \phi\left(\frac{1}{2}\mathbf{x} + \frac{1}{2}\mathbf{y}\right) &= \\
 &= \frac{1}{2}\phi\mathbf{x} + \frac{1}{2}\phi\mathbf{y}
 \end{aligned}$$

6. Proof that  $\phi([1 - \lambda]\mathbf{x} + \lambda\mathbf{y}) = [1 - \lambda]\phi\mathbf{x} + \lambda\phi\mathbf{y}$ :

$$\begin{aligned}
 \phi([1 - \lambda]\mathbf{x} + \lambda\mathbf{y}) &= \\
 &= [1 - \lambda]\phi\mathbf{x} + \lambda\phi\mathbf{y}
 \end{aligned}$$



**Theorem O.11** (Neumann Expansion Theorem). <sup>24</sup> Let  $\mathbf{A} \in \mathbf{X}^\mathbf{X}$  be an operator on a linear space  $\mathbf{X}$ . Let  $\mathbf{A}^0 \triangleq \mathbf{I}$ .

T H M	$  \left. \begin{array}{l} 1. \quad \mathbf{A} \in \mathcal{B}(\mathbf{X}, \mathbf{X}) \quad (\mathbf{A} \text{ is bounded}) \\ 2. \quad \ \mathbf{A}\  < 1 \end{array} \right\} \implies \left\{ \begin{array}{ll} 1. & (\mathbf{I} - \mathbf{A})^{-1} \quad \text{exists} \\ 2. & \ (\mathbf{I} - \mathbf{A})^{-1}\  \leq \frac{1}{1 - \ \mathbf{A}\ } \\ 3. & (\mathbf{I} - \mathbf{A})^{-1} = \sum_{n=0}^{\infty} \mathbf{A}^n \\ & \text{with uniform convergence} \end{array} \right.  $
-------------	--

## O.3 Operators on Inner product spaces

### O.3.1 General Results

**Theorem O.12.** <sup>25</sup> Let  $\mathbf{A}, \mathbf{B} \in \mathcal{B}(\mathbf{X}, \mathbf{X})$  be BOUNDED LINEAR OPERATORS on an inner product space  $\mathbf{X} \triangleq (X, +, \cdot, (\mathbb{F}, +, \hat{\times}), \langle \triangle | \nabla \rangle)$ .

T H M	$  \begin{array}{llll}  \langle \mathbf{Bx}   x \rangle & = & 0 & \forall x \in X \iff \mathbf{Bx} = \mathbf{0} \quad \forall x \in X \\  \langle \mathbf{Ax}   x \rangle & = & \langle \mathbf{Bx}   x \rangle & \forall x \in X \iff \mathbf{A} = \mathbf{B}  \end{array}  $
-------------	--

<sup>24</sup> Michel and Herget (1993) page 415

<sup>25</sup> Rudin (1991) page 310 (Theorem 12.7, Corollary)

PROOF:

1. Proof that  $\langle \mathbf{Bx} | x \rangle = 0 \implies \mathbf{Bx} = \mathbb{0}$ :

$$\begin{aligned}
 0 &= \langle \mathbf{B}(x + \mathbf{Bx}) | (x + \mathbf{Bx}) \rangle + i \langle \mathbf{B}(x + i\mathbf{Bx}) | (x + i\mathbf{Bx}) \rangle && \text{by left hypothesis} \\
 &= \{\langle \mathbf{Bx} + \mathbf{B}^2 x | x + \mathbf{Bx} \rangle\} + i\{\langle \mathbf{Bx} + i\mathbf{B}^2 x | x + i\mathbf{Bx} \rangle\} && \text{by Definition O.3 page 308} \\
 &= \{\langle \mathbf{Bx} | x \rangle + \langle \mathbf{Bx} | \mathbf{Bx} \rangle + \langle \mathbf{B}^2 x | x \rangle + \langle \mathbf{B}^2 x | \mathbf{Bx} \rangle\} && \text{by Definition K.1 page 259} \\
 &\quad + i\{\langle \mathbf{Bx} | x \rangle - i\langle \mathbf{Bx} | \mathbf{Bx} \rangle + i\langle \mathbf{B}^2 x | x \rangle - i^2\langle \mathbf{B}^2 x | \mathbf{Bx} \rangle\} \\
 &= \{0 + \langle \mathbf{Bx} | \mathbf{Bx} \rangle + \langle \mathbf{B}^2 x | x \rangle + 0\} + i\{0 - i\langle \mathbf{Bx} | \mathbf{Bx} \rangle + i\langle \mathbf{B}^2 x | x \rangle - i^2 0\} && \text{by left hypothesis} \\
 &= \{\langle \mathbf{Bx} | \mathbf{Bx} \rangle + \langle \mathbf{B}^2 x | x \rangle\} + \{\langle \mathbf{Bx} | \mathbf{Bx} \rangle - \langle \mathbf{B}^2 x | x \rangle\} \\
 &= 2\langle \mathbf{Bx} | \mathbf{Bx} \rangle \\
 &= 2\|\mathbf{Bx}\|^2 \\
 \implies \mathbf{Bx} &= \mathbb{0} && \text{by Definition L.1 page 275}
 \end{aligned}$$

2. Proof that  $\langle \mathbf{Bx} | x \rangle = 0 \iff \mathbf{Bx} = \mathbb{0}$ : by property of inner products (Theorem K.1 page 259).

3. Proof that  $\langle \mathbf{Ax} | x \rangle = \langle \mathbf{Bx} | x \rangle \implies \mathbf{A} \doteq \mathbf{B}$ :

$$\begin{aligned}
 0 &= \langle \mathbf{Ax} | x \rangle - \langle \mathbf{Bx} | x \rangle && \text{by left hypothesis} \\
 &= \langle \mathbf{Ax} - \mathbf{Bx} | x \rangle && \text{by } \textit{additivity} \text{ property of } \langle \triangle | \nabla \rangle \text{ (Definition K.1 page 259)} \\
 &= \langle (\mathbf{A} - \mathbf{B})x | x \rangle && \text{by definition of operator addition} \\
 \implies (\mathbf{A} - \mathbf{B})x &= \mathbb{0} && \text{by item 1} \\
 \implies \mathbf{A} &= \mathbf{B} && \text{by definition of operator subtraction}
 \end{aligned}$$

4. Proof that  $\langle \mathbf{Ax} | x \rangle = \langle \mathbf{Bx} | x \rangle \iff \mathbf{A} \doteq \mathbf{B}$ :

$$\langle \mathbf{Ax} | x \rangle = \langle \mathbf{Bx} | x \rangle \quad \text{by } \mathbf{A} \doteq \mathbf{B} \text{ hypothesis}$$

⇒

### O.3.2 Operator adjoint

A fundamental concept of operators on inner product spaces is the *operator adjoint* (Proposition O.3 page 320). The adjoint of an operator is a kind of generalization of the conjugate of a complex number in that

- Both are *star-algebras* (Theorem O.13 page 321).
- Both support decomposition into “real” and “imaginary” parts (Theorem J.3 page 256).

Structurally, the operator adjoint provides a convenient symmetric relationship between the *range space* and *null space* of an operator (Theorem O.14 page 322).

**Proposition O.3.** <sup>26</sup> Let  $\mathcal{B}(\mathbf{H}, \mathbf{H})$  be the space of BOUNDED LINEAR OPERATORS (Definition O.5 page 315) on a HILBERT SPACE  $\mathbf{H}$ .

**P** An operator  $\mathbf{B}^*$  is the **adjoint** of  $\mathbf{B} \in \mathcal{B}(\mathbf{H}, \mathbf{H})$  if  
**R**  $\langle \mathbf{Bx} | y \rangle = \langle x | \mathbf{B}^* y \rangle \quad \forall x, y \in \mathbf{H}$ .

<sup>26</sup> 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  $y$ ,  $f(x) \triangleq \langle x | y \rangle$  is a *functional* in  $\mathbb{F}^X$ .

2.  $B^*$  is the *adjoint* of  $B$  because

$$\begin{aligned}\langle Bx | y \rangle &\triangleq f(Bx) \\ &\triangleq B^*f(x) && \text{by definition of operator adjoint} \\ &= \langle x | B^*y \rangle\end{aligned}\quad (\text{Definition O.6 page 317})$$



### Example O.2.

**E  
X**

In matrix algebra (“linear algebra”)

- The inner product operation  $\langle x | y \rangle$  is represented by  $y^H x$ .
- The linear operator is represented as a matrix  $A$ .
- The operation of  $A$  on a vector  $x$  is represented as  $Ax$ .
- The adjoint of matrix  $A$  is the Hermitian matrix  $A^H$ .



PROOF:

$$\langle Ax | y \rangle \triangleq y^H Ax = [(Ax)^H y]^H = [x^H A^H y]^H = (A^H y)^H x \triangleq \langle x | A^H y \rangle$$

Structures that satisfy the four conditions of the next theorem are known as *\*-algebras* (“star-algebras” (Definition J.3 page 254)). Other structures which are \*-algebras include the *field of complex numbers*  $\mathbb{C}$  and any *ring of complex square  $n \times n$  matrices*.<sup>27</sup>

**Theorem O.13** (operator star-algebra). <sup>28</sup> Let  $H$  be a HILBERT SPACE with operators  $A, B \in \mathcal{B}(H, H)$  and with adjoints  $A^*, B^* \in \mathcal{B}(H, H)$ . Let  $\bar{\alpha}$  be the complex conjugate of some  $\alpha \in \mathbb{C}$ .

**T  
H  
M**

The pair  $(H, *)$  is a \*-ALGEBRA (STAR-ALGEBRA). In particular,

1.  $(A + B)^* = A^* + B^*$   $\forall A, B \in H$  (DISTRIBUTIVE) and
2.  $(\alpha A)^* = \bar{\alpha} A^*$   $\forall A, B \in H$  (CONJUGATE LINEAR) and
3.  $(AB)^* = B^* A^*$   $\forall A, B \in H$  (ANTIAUTOMORPHIC) and
4.  $A^{**} = A$   $\forall A, B \in H$  (INVOLUTARY)

PROOF:

$$\begin{aligned}\langle x | (A + B)^* y \rangle &= \langle (A + B)x | y \rangle && \text{by definition of adjoint} \\ &= \langle Ax | y \rangle + \langle Bx | y \rangle && \text{by definition of inner product} \\ &= \langle x | A^* y \rangle + \langle x | B^* y \rangle && \text{by definition of operator addition} \\ &= \langle x | A^* y + B^* y \rangle && \text{by definition of inner product} \\ &= \langle x | (A^* + B^*) y \rangle && \text{by definition of operator addition}\end{aligned}\quad (\text{Proposition O.3 page 320})$$

$$\begin{aligned}\langle x | (\alpha A)^* y \rangle &= \langle (\alpha A)x | y \rangle && \text{by definition of adjoint} \\ &= \langle \alpha(Ax) | y \rangle && \text{by definition of scalar multiplication} \\ &= \alpha \langle Ax | y \rangle && \text{by definition of inner product}\end{aligned}\quad (\text{Proposition O.3 page 320})$$

<sup>27</sup> Sakai (1998) page 1

<sup>28</sup> Halmos (1998a), pages 39–40, Rudin (1991) page 311

$$\begin{aligned} &= \alpha \langle \mathbf{x} | \mathbf{A}^* \mathbf{y} \rangle && \text{by definition of adjoint} && (\text{Proposition O.3 page 320}) \\ &= \langle \mathbf{x} | \alpha^* \mathbf{A}^* \mathbf{y} \rangle && \text{by definition of inner product} && (\text{Definition K.1 page 259}) \end{aligned}$$

$$\begin{aligned} \langle \mathbf{x} | (\mathbf{AB})^* \mathbf{y} \rangle &= \langle (\mathbf{AB})\mathbf{x} | \mathbf{y} \rangle && \text{by definition of adjoint} && (\text{Proposition O.3 page 320}) \\ &= \langle \mathbf{A}(\mathbf{B}\mathbf{x}) | \mathbf{y} \rangle && \text{by definition of operator multiplication} && \\ &= \langle (\mathbf{B}\mathbf{x}) | \mathbf{A}^* \mathbf{y} \rangle && \text{by definition of adjoint} && (\text{Proposition O.3 page 320}) \\ &= \langle \mathbf{x} | \mathbf{B}^* \mathbf{A}^* \mathbf{y} \rangle && \text{by definition of adjoint} && (\text{Proposition O.3 page 320}) \end{aligned}$$

$$\begin{aligned} \langle \mathbf{x} | \mathbf{A}^{**} \mathbf{y} \rangle &= \langle \mathbf{A}^* \mathbf{x} | \mathbf{y} \rangle && \text{by definition of adjoint} && (\text{Proposition O.3 page 320}) \\ &= \langle \mathbf{y} | \mathbf{A}^* \mathbf{x} \rangle^* && \text{by definition of inner product} && (\text{Definition K.1 page 259}) \\ &= \langle \mathbf{A}\mathbf{y} | \mathbf{x} \rangle^* && \text{by definition of adjoint} && (\text{Proposition O.3 page 320}) \\ &= \langle \mathbf{x} | \mathbf{A}\mathbf{y} \rangle && \text{by definition of inner product} && (\text{Definition K.1 page 259}) \end{aligned}$$

⇒

**Theorem O.14.** <sup>29</sup> Let  $\mathcal{Y}^X$  be the set of all operators from a linear space  $\mathbf{X}$  to a linear space  $\mathbf{Y}$ . Let  $\mathcal{N}(\mathbf{L})$  be the NULL SPACE of an operator  $\mathbf{L}$  in  $\mathcal{Y}^X$  and  $\mathcal{I}(\mathbf{L})$  the IMAGE SET of  $\mathbf{L}$  in  $\mathcal{Y}^X$ .

T	$\mathcal{N}(\mathbf{A}) = \mathcal{I}(\mathbf{A}^*)^\perp$
H	
M	$\mathcal{N}(\mathbf{A}^*) = \mathcal{I}(\mathbf{A})^\perp$

PROOF:

$$\begin{aligned} \mathcal{I}(\mathbf{A}^*)^\perp &= \{y \in \mathbf{H} \mid \langle y | u \rangle = 0 \quad \forall u \in \mathcal{I}(\mathbf{A}^*)\} \\ &= \{y \in \mathbf{H} \mid \langle y | \mathbf{A}^* \mathbf{x} \rangle = 0 \quad \forall \mathbf{x} \in \mathbf{H}\} \\ &= \{y \in \mathbf{H} \mid \langle \mathbf{A}\mathbf{y} | \mathbf{x} \rangle = 0 \quad \forall \mathbf{x} \in \mathbf{H}\} && \text{by definition of } \mathbf{A}^* && (\text{Proposition O.3 page 320}) \\ &= \{y \in \mathbf{H} \mid \mathbf{A}\mathbf{y} = 0\} \\ &= \mathcal{N}(\mathbf{A}) && \text{by definition of } \mathcal{N}(\mathbf{A}) \end{aligned}$$

$$\begin{aligned} \mathcal{I}(\mathbf{A})^\perp &= \{y \in \mathbf{H} \mid \langle y | u \rangle = 0 \quad \forall u \in \mathcal{I}(\mathbf{A})\} \\ &= \{y \in \mathbf{H} \mid \langle y | \mathbf{Ax} \rangle = 0 \quad \forall \mathbf{x} \in \mathbf{H}\} && \text{by definition of } \mathcal{I} \\ &= \{y \in \mathbf{H} \mid \langle \mathbf{A}^* \mathbf{y} | \mathbf{x} \rangle = 0 \quad \forall \mathbf{x} \in \mathbf{H}\} && \text{by definition of } \mathbf{A}^* && (\text{Proposition O.3 page 320}) \\ &= \{y \in \mathbf{H} \mid \mathbf{A}^* \mathbf{y} = 0\} \\ &= \mathcal{N}(\mathbf{A}^*) && \text{by definition of } \mathcal{N}(\mathbf{A}) \end{aligned}$$

⇒

## O.4 Special Classes of Operators

### O.4.1 Projection operators

**Definition O.7.** <sup>30</sup> Let  $\mathcal{B}(\mathbf{X}, \mathbf{Y})$  be the space of BOUNDED LINEAR OPERATORS on normed linear spaces  $\mathbf{X}$  and  $\mathbf{Y}$ . Let  $\mathbf{P}$  be a bounded linear operator in  $\mathcal{B}(\mathbf{X}, \mathbf{Y})$ .

<sup>29</sup> Rudin (1991) page 312

<sup>30</sup> Rudin (1991) page 133 (5.15 Projections), Kubrusly (2001) page 70, Bachman and Narici (1966) page 6, Halmos (1958) page 73 (§41. Projections)



**D  
E  
F**

**P** is a **projection operator** if  $\mathbf{P}^2 = \mathbf{P}$ .

**Theorem O.15.** <sup>31</sup> Let  $\mathcal{B}(X, Y)$  be the space of BOUNDED LINEAR OPERATORS on normed linear spaces  $X$  and  $Y$ . Let  $\mathbf{P}$  be a bounded linear operator in  $\mathcal{B}(X, Y)$  with NULL SPACE  $\mathcal{N}(\mathbf{P})$  and IMAGE SET  $\mathcal{I}(\mathbf{P})$ .

**T  
H  
M**

$$\left. \begin{array}{ll} 1. \quad \mathbf{P}^2 = \mathbf{P} & (\mathbf{P} \text{ is a projection operator}) \\ 2. \quad \Omega = X \dagger Y & (Y \text{ complements } X \text{ in } \Omega) \\ 3. \quad \mathbf{P}\Omega = X & (\mathbf{P} \text{ projects onto } X) \end{array} \right\} \Rightarrow \left\{ \begin{array}{ll} 1. \quad \mathcal{I}(\mathbf{P}) = X & \text{and} \\ 2. \quad \mathcal{N}(\mathbf{P}) = Y & \text{and} \\ 3. \quad \Omega = \mathcal{I}(\mathbf{P}) \dagger \mathcal{N}(\mathbf{P}) & \end{array} \right.$$

PROOF:

$$\begin{aligned} \mathcal{I}(\mathbf{P}) &= \mathbf{P}\Omega \\ &= \mathbf{P}(\Omega_1 + \Omega_2) \\ &= \mathbf{P}\Omega_1 + \mathbf{P}\Omega_2 \\ &= \Omega_1 + \{\mathbf{0}\} \\ &= \Omega_1 \end{aligned}$$

$$\begin{aligned} \mathcal{N}(\mathbf{P}) &= \{x \in \Omega \mid \mathbf{P}x = \mathbf{0}\} \\ &= \{x \in (\Omega_1 + \Omega_2) \mid \mathbf{P}x = \mathbf{0}\} \\ &= \{x \in \Omega_1 \mid \mathbf{P}x = \mathbf{0}\} + \{x \in \Omega_2 \mid \mathbf{P}x = \mathbf{0}\} \\ &= \{\mathbf{0}\} + \Omega_2 \\ &= \Omega_2 \end{aligned}$$

⇒

**Theorem O.16.** <sup>32</sup> Let  $\mathcal{B}(X, Y)$  be the space of BOUNDED LINEAR OPERATORS on normed linear spaces  $X$  and  $Y$ . Let  $\mathbf{P}$  be a bounded linear operator in  $\mathcal{B}(X, Y)$ .

**T  
H  
M**

$$\underbrace{\mathbf{P}^2 = \mathbf{P}}_{\mathbf{P} \text{ is a projection operator}} \iff \underbrace{(\mathbf{I} - \mathbf{P})^2 = (\mathbf{I} - \mathbf{P})}_{(\mathbf{I} - \mathbf{P}) \text{ is a projection operator}}$$

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  $\mathbf{P}^2 = \mathbf{P} \iff (\mathbf{I} - \mathbf{P})^2 = (\mathbf{I} - \mathbf{P})$ :

$$\begin{aligned} \mathbf{P}^2 &= \underbrace{\mathbf{I} - \mathbf{P} - \mathbf{P} + \mathbf{P}^2}_{(\mathbf{I} - \mathbf{P})^2} - (\mathbf{I} - \mathbf{P} - \mathbf{P}) \\ &= (\mathbf{I} - \mathbf{P})^2 - (\mathbf{I} - \mathbf{P} - \mathbf{P}) \\ &= (\mathbf{I} - \mathbf{P}) - (\mathbf{I} - \mathbf{P} - \mathbf{P}) && \text{by right hypothesis} \\ &= \mathbf{P} \end{aligned}$$

⇒

<sup>31</sup> Michel and Herget (1993) pages 120–121

<sup>32</sup> Michel and Herget (1993) page 121

**Theorem O.17.** <sup>33</sup> Let  $\mathbf{H}$  be a HILBERT SPACE and  $\mathbf{P}$  an operator in  $\mathcal{B}(\mathbf{H})$  with adjoint  $\mathbf{P}^*$ , NULL SPACE  $\mathcal{N}(\mathbf{P})$ , and IMAGE SET  $\mathcal{I}(\mathbf{P})$ .

If  $\mathbf{P}$  is a PROJECTION OPERATOR, then the following are equivalent:

- |             |   |  |        |
|-------------|---|--|--------|
| T<br>H<br>M | 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. $\langle \mathbf{P}\mathbf{x}   \mathbf{x} \rangle = \ \mathbf{P}\mathbf{x}\ ^2 \quad \forall \mathbf{x} \in \mathbf{X}$ |  |        |

PROOF: This proof is incomplete at this time.

Proof that (1)  $\implies$  (2):

$$\begin{aligned} \mathbf{P}^*\mathbf{P} &= \mathbf{P}^*\mathbf{P}^*\mathbf{P}^* && \text{by (1)} \\ &= \mathbf{P}\mathbf{P}^* && \text{by Theorem O.13 page 321} \end{aligned}$$

Proof that (1)  $\implies$  (3):

$$\begin{aligned} \mathcal{I}(\mathbf{P}) &= \mathcal{N}(\mathbf{P}^*)^\perp && \text{by Theorem O.14 page 322} \\ &= \mathcal{N}(\mathbf{P})^\perp && \text{by (1)} \end{aligned}$$

Proof that (3)  $\implies$  (4):

Proof that (4)  $\implies$  (1):



## O.4.2 Self Adjoint Operators

**Definition O.8.** <sup>34</sup> Let  $\mathbf{B} \in \mathcal{B}(\mathbf{H}, \mathbf{H})$  be a BOUNDED operator with adjoint  $\mathbf{B}^*$  on a HILBERT SPACE  $\mathbf{H}$ .

**D E F** The operator  $\mathbf{B}$  is said to be **self-adjoint** or **hermitian** if  $\mathbf{B} = \mathbf{B}^*$ .

*Example O.3* (Autocorrelation operator). Let  $\mathbf{x}(t)$  be a random process with autocorrelation

$$R_{xx}(t, u) \triangleq \underbrace{\mathbb{E}[x(t)x^*(u)]}_{\text{expectation}}$$

Let an autocorrelation operator  $\mathbf{R}$  be defined as  $[\mathbf{R}\mathbf{f}](t) \triangleq \int_{\mathbb{R}} R_{xx}(t, u)\mathbf{f}(u) du$ .

**E X**  $\mathbf{R} = \mathbf{R}^*$  (The auto-correlation operator  $\mathbf{R}$  is *self-adjoint*)

**Theorem O.18.** <sup>35</sup> Let  $\mathbf{S} : \mathbf{H} \rightarrow \mathbf{H}$  be an operator over a HILBERT SPACE  $\mathbf{H}$  with eigenvalues  $\{\lambda_n\}$  and eigenfunctions  $\{\psi_n\}$  such that  $\mathbf{S}\psi_n = \lambda_n\psi_n$  and let  $\|\mathbf{x}\| \triangleq \sqrt{\langle \mathbf{x} | \mathbf{x} \rangle}$ .

**T H M**  $\left\{ \begin{array}{l} \mathbf{S} = \mathbf{S}^* \\ \mathbf{S} \text{ is selfadjoint} \end{array} \right\} \implies \left\{ \begin{array}{ll} 1. \langle \mathbf{S}\mathbf{x} | \mathbf{x} \rangle \in \mathbb{R} & (\text{the hermitian quadratic form of } \mathbf{S} \text{ is REAL-VALUED}) \\ 2. \lambda_n \in \mathbb{R} & (\text{eigenvalues of } \mathbf{S} \text{ are REAL-VALUED}) \\ 3. \lambda_n \neq \lambda_m \implies \langle \psi_n | \psi_m \rangle = 0 & (\text{eigenvectors are ORTHOGONAL}) \end{array} \right\}$

<sup>33</sup> Rudin (1991) page 314

<sup>34</sup> Historical works regarding self-adjoint operators: von Neumann (1929), page 49, "linearer Operator R selbstdjugiert oder Hermitesch", Stone (1932), page 50 ('self-adjoint transformations')

<sup>35</sup> 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...”)

PROOF:

1. Proof that  $S = S^* \implies \langle Sx | x \rangle \in \mathbb{R}$ :

$$\begin{aligned} \langle x | Sx \rangle &= \langle Sx | x \rangle && \text{by left hypothesis} \\ &= \langle x | Sx \rangle^* && \text{by definition of } \langle \cdot | \cdot \rangle \text{ Definition K.1 page 259} \end{aligned}$$

2. Proof that  $S = S^* \implies \lambda_n \in \mathbb{R}$ :

$$\begin{aligned} \lambda_n \|\psi_n\|^2 &= \lambda_n \langle \psi_n | \psi_n \rangle && \text{by definition} \\ &= \langle \lambda_n \psi_n | \psi_n \rangle && \text{by definition of } \langle \cdot | \cdot \rangle \text{ Definition K.1 page 259} \\ &= \langle S\psi_n | \psi_n \rangle && \text{by definition of eigenpairs} \\ &= \langle \psi_n | S\psi_n \rangle && \text{by left hypothesis} \\ &= \langle \psi_n | \lambda_n \psi_n \rangle && \text{by definition of eigenpairs} \\ &= \lambda_n^* \langle \psi_n | \psi_n \rangle && \text{by definition of } \langle \cdot | \cdot \rangle \text{ Definition K.1 page 259} \\ &= \lambda_n^* \|\psi_n\|^2 && \text{by definition} \end{aligned}$$

3. Proof that  $S = S^* \implies [\lambda_n \neq \lambda_m \implies \langle \psi_n | \psi_m \rangle = 0]$ :

$$\begin{aligned} \lambda_n \langle \psi_n | \psi_m \rangle &= \langle \lambda_n \psi_n | \psi_m \rangle && \text{by definition of } \langle \cdot | \cdot \rangle \text{ Definition K.1 page 259} \\ &= \langle S\psi_n | \psi_m \rangle && \text{by definition of eigenpairs} \\ &= \langle \psi_n | S\psi_m \rangle && \text{by left hypothesis} \\ &= \langle \psi_n | \lambda_m \psi_m \rangle && \text{by definition of eigenpairs} \\ &= \lambda_m^* \langle \psi_n | \psi_m \rangle && \text{by definition of } \langle \cdot | \cdot \rangle \text{ Definition K.1 page 259} \\ &= \lambda_m \langle \psi_n | \psi_m \rangle && \text{because } \lambda_m \text{ is real} \end{aligned}$$

This implies for  $\lambda_n \neq \lambda_m \neq 0$ ,  $\langle \psi_n | \psi_m \rangle = 0$ .



### O.4.3 Normal Operators

**Definition O.9.** <sup>36</sup> Let  $\mathcal{B}(X, Y)$  be the space of BOUNDED LINEAR OPERATORS on normed linear spaces  $X$  and  $Y$ . Let  $N^*$  be the adjoint of an operator  $N \in \mathcal{B}(X, Y)$ .

**D E F**  $N$  is **normal** if  $N^*N = NN^*$ .

**Theorem O.19.** <sup>37</sup> Let  $\mathcal{B}(H, H)$  be the space of BOUNDED LINEAR OPERATORS on a HILBERT SPACE  $H$ . Let  $\mathcal{N}(N)$  be the NULL SPACE of an operator  $N$  in  $\mathcal{B}(H, H)$  and  $\mathcal{I}(N)$  the IMAGE SET of  $N$  in  $\mathcal{B}(H, H)$ .

T H M	$\underbrace{N^*N = NN^*}_{N \text{ is normal}}$	$\iff$	$\ N^*x\  = \ Nx\  \quad \forall x \in H$
-------------	--	--------	---

<sup>36</sup> Rudin (1991) page 312, Michel and Herget (1993) page 431, Dieudonné (1969), page 167, Frobenius (1878), Frobenius (1968), page 391

<sup>37</sup> Rudin (1991) pages 312–313

PROOF:

1. Proof that  $\mathbf{N}^*\mathbf{N} = \mathbf{N}\mathbf{N}^* \implies \|\mathbf{N}^*\mathbf{x}\| = \|\mathbf{N}\mathbf{x}\|$ :

$$\begin{aligned}
 \|\mathbf{N}\mathbf{x}\|^2 &= \langle \mathbf{N}\mathbf{x} | \mathbf{N}\mathbf{x} \rangle && \text{by definition} \\
 &= \langle \mathbf{x} | \mathbf{N}^*\mathbf{N}\mathbf{x} \rangle && \text{by Proposition O.3 page 320 (definition of } \mathbf{N}^*) \\
 &= \langle \mathbf{x} | \mathbf{N}\mathbf{N}^*\mathbf{x} \rangle && \text{by left hypothesis (N is normal)} \\
 &= \langle \mathbf{N}\mathbf{x} | \mathbf{N}^*\mathbf{x} \rangle && \text{by Proposition O.3 page 320 (definition of } \mathbf{N}^*) \\
 &= \|\mathbf{N}^*\mathbf{x}\|^2 && \text{by definition}
 \end{aligned}$$

2. Proof that  $\mathbf{N}^*\mathbf{N} = \mathbf{N}\mathbf{N}^* \iff \|\mathbf{N}^*\mathbf{x}\| = \|\mathbf{N}\mathbf{x}\|$ :

$$\begin{aligned}
 \langle \mathbf{N}^*\mathbf{N}\mathbf{x} | \mathbf{x} \rangle &= \langle \mathbf{N}\mathbf{x} | \mathbf{N}^*\mathbf{x} \rangle && \text{by Proposition O.3 page 320 (definition of } \mathbf{N}^*) \\
 &= \langle \mathbf{N}\mathbf{x} | \mathbf{N}\mathbf{x} \rangle && \text{by Theorem O.13 page 321 (property of adjoint)} \\
 &= \|\mathbf{N}\mathbf{x}\|^2 && \text{by definition} \\
 &= \|\mathbf{N}^*\mathbf{x}\|^2 && \text{by right hypothesis (\|N}^*\mathbf{x}\| = \|\mathbf{N}\mathbf{x}\|)} \\
 &= \langle \mathbf{N}^*\mathbf{x} | \mathbf{N}^*\mathbf{x} \rangle && \text{by definition} \\
 &= \langle \mathbf{N}\mathbf{N}^*\mathbf{x} | \mathbf{x} \rangle && \text{by Proposition O.3 page 320 (definition of } \mathbf{N}^*)
 \end{aligned}$$

⇒

**Theorem O.20.**<sup>38</sup> Let  $\mathcal{B}(\mathbf{H}, \mathbf{H})$  be the space of BOUNDED LINEAR OPERATORS on a HILBERT SPACE  $\mathbf{H}$ . Let  $\mathcal{N}(\mathbf{N})$  be the NULL SPACE of an operator  $\mathbf{N}$  in  $\mathcal{B}(\mathbf{H}, \mathbf{H})$  and  $\mathcal{I}(\mathbf{N})$  the IMAGE SET of  $\mathbf{N}$  in  $\mathcal{B}(\mathbf{H}, \mathbf{H})$ .

<b>T H M</b>	$\underbrace{\mathbf{N}^*\mathbf{N} = \mathbf{N}\mathbf{N}^*}_{\mathbf{N} \text{ is normal}}$	$\implies$	$\underbrace{\mathcal{N}(\mathbf{N}^*) = \mathcal{N}(\mathbf{N})}_{\mathbf{N} \text{ and } \mathbf{N}^* \text{ have the same null space}}$
----------------------	---	------------	--

PROOF:

$$\begin{aligned}
 \mathcal{N}(\mathbf{N}^*) &= \{ \mathbf{x} | \mathbf{N}^*\mathbf{x} = 0 \quad \forall \mathbf{x} \in \mathbf{X} \} && \text{(definition of } \mathcal{N}^*) \\
 &= \{ \mathbf{x} | \| \mathbf{N}^*\mathbf{x} \| = 0 \quad \forall \mathbf{x} \in \mathbf{X} \} && \text{by definition of } \| \cdot \| \text{ (Definition L.1 page 275)} \\
 &= \{ \mathbf{x} | \| \mathbf{N}\mathbf{x} \| = 0 \quad \forall \mathbf{x} \in \mathbf{X} \} && \text{by definition of } \| \cdot \| \text{ (Definition L.1 page 275)} \\
 &= \{ \mathbf{x} | \mathbf{N}\mathbf{x} = 0 \quad \forall \mathbf{x} \in \mathbf{X} \} && \text{(definition of } \mathcal{N})
 \end{aligned}$$

⇒

**Theorem O.21.**<sup>39</sup> Let  $\mathcal{B}(\mathbf{H}, \mathbf{H})$  be the space of BOUNDED LINEAR OPERATORS on a HILBERT SPACE  $\mathbf{H}$ . Let  $\mathcal{N}(\mathbf{N})$  be the NULL SPACE of an operator  $\mathbf{N}$  in  $\mathcal{B}(\mathbf{H}, \mathbf{H})$  and  $\mathcal{I}(\mathbf{N})$  the IMAGE SET of  $\mathbf{N}$  in  $\mathcal{B}(\mathbf{H}, \mathbf{H})$ .

<b>T H M</b>	$\left\{ \underbrace{\mathbf{N}^*\mathbf{N} = \mathbf{N}\mathbf{N}^*}_{\mathbf{N} \text{ is normal}} \right\}$	$\implies$	$\left\{ \underbrace{\lambda_n \neq \lambda_m \implies \langle \psi_n   \psi_m \rangle = 0}_{\text{eigenfunctions associated with distinct eigenvalues are orthogonal}} \right\}$
----------------------	--	------------	---

PROOF: The proof in (1) is flawed. This implies that (2) is also flawed. [Rudin] claims both to be true. (Rudin, 1991)313

<sup>38</sup> Rudin (1991) pages 312–313

<sup>39</sup> Rudin (1991) pages 312–313

1. Proof that  $\mathbf{N}^*\mathbf{N} = \mathbf{N}\mathbf{N}^* \implies \mathbf{N}^*\psi = \lambda^*\psi$ :

$$\begin{aligned}
 \mathbf{N}\psi &= \lambda\psi \\
 \iff 0 &= \mathcal{N}(\mathbf{N} - \lambda\mathbf{I}) \\
 &= \mathcal{N}([\mathbf{N} - \lambda\mathbf{I}]^*) && \text{by } \mathcal{N}(\mathbf{N}) = \mathcal{N}(\mathbf{N}^*) \\
 &= \mathcal{N}(\mathbf{N}^* - [\lambda\mathbf{I}]^*) && \text{by Theorem O.13 page 321} \\
 &= \mathcal{N}(\mathbf{N}^* - \lambda^*\mathbf{I}^*) && \text{by Theorem O.13 page 321} \\
 &= \mathcal{N}(\mathbf{N}^* - \lambda^*\mathbf{I}) \\
 \implies (\mathbf{N}^* - \lambda^*\mathbf{I})\psi &= 0 \\
 \iff \mathbf{N}^*\psi &= \lambda^*\psi
 \end{aligned}$$

2. Proof that  $\mathbf{N}^*\mathbf{N} = \mathbf{N}\mathbf{N}^* \implies [\lambda_n \neq \lambda_m \implies \langle \psi_n | \psi_m \rangle = 0]$ :

$$\begin{aligned}
 \lambda_n \langle \psi_n | \psi_m \rangle &= \langle \lambda_n \psi_n | \psi_m \rangle && \text{by definition of } \langle \triangle | \triangleright \rangle \text{ Definition K.1 page 259} \\
 &= \langle \mathbf{N}\psi_n | \psi_m \rangle && \text{by definition of eigenpairs} \\
 &= \langle \psi_n | \mathbf{N}^*\psi_m \rangle && \text{by Proposition O.3 page 320 (definition of adjoint)} \\
 &= \langle \psi_n | \lambda_m^* \psi_m \rangle && \text{by (4.)} \\
 &= \lambda_m \langle \psi_n | \psi_m \rangle && \text{by definition of } \langle \triangle | \triangleright \rangle \text{ Definition K.1 page 259}
 \end{aligned}$$

This implies for  $\lambda_n \neq \lambda_m \neq 0$ ,  $\langle \psi_n | \psi_m \rangle = 0$ .



#### O.4.4 Isometric operators

An operator on a pair of normed linear spaces is *isometric* (next definition) if it is an *isometry*.

**Definition O.10.** Let  $(X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$  and  $(Y, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$  be NORMED LINEAR SPACES (Definition L.1 page 275).

**D E F** An operator  $\mathbf{M} \in \mathcal{L}(X, Y)$  is **isometric** if  $\|\mathbf{M}\mathbf{x}\| = \|\mathbf{x}\| \quad \forall \mathbf{x} \in X$ .

**Theorem O.22.**<sup>40</sup> 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  $\mathcal{L}(X, Y)$ .

T H M	$\underbrace{\ \mathbf{M}\mathbf{x}\  = \ \mathbf{x}\ }_{\text{isometric in length}} \quad \iff \quad \underbrace{\ \mathbf{M}\mathbf{x} - \mathbf{M}\mathbf{y}\  = \ \mathbf{x} - \mathbf{y}\ }_{\text{isometric in distance}} \quad \forall \mathbf{x}, \mathbf{y} \in X$
-------------	---

PROOF:

1. Proof that  $\|\mathbf{M}\mathbf{x}\| = \|\mathbf{x}\| \implies \|\mathbf{M}\mathbf{x} - \mathbf{M}\mathbf{y}\| = \|\mathbf{x} - \mathbf{y}\|$ :

$$\begin{aligned}
 \|\mathbf{M}\mathbf{x} - \mathbf{M}\mathbf{y}\| &= \|\mathbf{M}(\mathbf{x} - \mathbf{y})\| && \text{by definition of linear operators (Definition O.3 page 308)} \\
 &= \|\mathbf{M}\mathbf{u}\| && \text{let } \mathbf{u} \triangleq \mathbf{x} - \mathbf{y} \\
 &= \|\mathbf{x} - \mathbf{y}\| && \text{by left hypothesis}
 \end{aligned}$$

<sup>40</sup> Kubrusly (2001) page 239 (Proposition 4.37), Berberian (1961) page 27 (Theorem IV.7.5)

2. Proof that  $\|\mathbf{M}\mathbf{x}\| = \|\mathbf{x}\| \iff \|\mathbf{M}\mathbf{x} - \mathbf{M}\mathbf{y}\| = \|\mathbf{x} - \mathbf{y}\|$ :

$$\begin{aligned}\|\mathbf{M}\mathbf{x}\| &= \|\mathbf{M}(\mathbf{x} - \mathbf{0})\| \\ &= \|\mathbf{M}\mathbf{x} - \mathbf{M}\mathbf{0}\| && \text{by definition of linear operators (Definition O.3 page 308)} \\ &= \|\mathbf{x} - \mathbf{0}\| && \text{by right hypothesis} \\ &= \|\mathbf{x}\|\end{aligned}$$



Isometric operators have already been defined (Definition O.10 page 327) in the more general normed linear spaces, while Theorem O.22 (page 327) demonstrated that in a normed linear space  $\mathbf{X}$ ,  $\|\mathbf{M}\mathbf{x}\| = \|\mathbf{x}\| \iff \|\mathbf{M}\mathbf{x} - \mathbf{M}\mathbf{y}\| = \|\mathbf{x} - \mathbf{y}\|$  for all  $\mathbf{x}, \mathbf{y} \in \mathbf{X}$ . Here in the more specialized inner product spaces, Theorem O.23 (next) demonstrates two additional equivalent properties.

**Theorem O.23.**<sup>41</sup> Let  $\mathcal{B}(\mathbf{X}, \mathbf{X})$  be the space of BOUNDED LINEAR OPERATORS on a normed linear space  $\mathbf{X} \triangleq (X, +, \cdot, (\mathbb{F}, \dot{+}, \dot{\times}), \|\cdot\|)$ . Let  $\mathbf{N}$  be a bounded linear operator in  $\mathcal{L}(\mathbf{X}, \mathbf{X})$ , and  $\mathbf{I}$  the identity operator in  $\mathcal{L}(\mathbf{X}, \mathbf{X})$ . Let  $\|\mathbf{x}\| \triangleq \sqrt{\langle \mathbf{x} | \mathbf{x} \rangle}$ .

T H M	<i>The following conditions are all equivalent:</i>
	1. $\mathbf{M}^* \mathbf{M} = \mathbf{I} \iff$
	2. $\langle \mathbf{M}\mathbf{x}   \mathbf{M}\mathbf{y} \rangle = \langle \mathbf{x}   \mathbf{y} \rangle \quad \forall \mathbf{x}, \mathbf{y} \in X \quad (\mathbf{M} \text{ is surjective}) \iff$
	3. $\ \mathbf{M}\mathbf{x} - \mathbf{M}\mathbf{y}\  = \ \mathbf{x} - \mathbf{y}\  \quad \forall \mathbf{x}, \mathbf{y} \in X \quad (\text{isometric in distance}) \iff$
	4. $\ \mathbf{M}\mathbf{x}\  = \ \mathbf{x}\  \quad \forall \mathbf{x} \in X \quad (\text{isometric in length})$

PROOF:

1. Proof that (1)  $\implies$  (2):

$$\begin{aligned}\langle \mathbf{M}\mathbf{x} | \mathbf{M}\mathbf{y} \rangle &= \langle \mathbf{x} | \mathbf{M}^* \mathbf{M}\mathbf{y} \rangle && \text{by Proposition O.3 page 320 (definition of adjoint)} \\ &= \langle \mathbf{x} | \mathbf{I}\mathbf{y} \rangle && \text{by (1)} \\ &= \langle \mathbf{x} | \mathbf{y} \rangle && \text{by Definition O.2 page 307 (definition of } \mathbf{I}\text{)}\end{aligned}$$

2. Proof that (2)  $\implies$  (4):

$$\begin{aligned}\|\mathbf{M}\mathbf{x}\| &= \sqrt{\langle \mathbf{M}\mathbf{x} | \mathbf{M}\mathbf{x} \rangle} && \text{by definition of } \|\cdot\| \\ &= \sqrt{\langle \mathbf{x} | \mathbf{x} \rangle} && \text{by right hypothesis} \\ &= \|\mathbf{x}\| && \text{by definition of } \|\cdot\|\end{aligned}$$

3. Proof that (2)  $\iff$  (4):

$$\begin{aligned}4 \langle \mathbf{M}\mathbf{x} | \mathbf{M}\mathbf{y} \rangle &= \|\mathbf{M}\mathbf{x} + \mathbf{M}\mathbf{y}\|^2 - \|\mathbf{M}\mathbf{x} - \mathbf{M}\mathbf{y}\|^2 + i \|\mathbf{M}\mathbf{x} + i\mathbf{M}\mathbf{y}\|^2 - i \|\mathbf{M}\mathbf{x} - i\mathbf{M}\mathbf{y}\|^2 && \text{by polarization id.} \\ &= \|\mathbf{M}(\mathbf{x} + \mathbf{y})\|^2 - \|\mathbf{M}(\mathbf{x} - \mathbf{y})\|^2 + i \|\mathbf{M}(\mathbf{x} + i\mathbf{y})\|^2 - i \|\mathbf{M}(\mathbf{x} - i\mathbf{y})\|^2 && \text{by Definition O.3} \\ &= \|\mathbf{x} + \mathbf{y}\|^2 - \|\mathbf{x} - \mathbf{y}\|^2 + i \|\mathbf{x} + i\mathbf{y}\|^2 - i \|\mathbf{x} - i\mathbf{y}\|^2 && \text{by left hypothesis}\end{aligned}$$

4. Proof that (3)  $\iff$  (4): by Theorem O.22 page 327

<sup>41</sup> Michel and Herget (1993) page 432 (Theorem 7.5.8), Kubrusly (2001) page 391 (Proposition 5.72)



5. Proof that (4)  $\implies$  (1):

$$\begin{aligned}
 \langle \mathbf{M}^* \mathbf{M} \mathbf{x} | \mathbf{x} \rangle &= \langle \mathbf{M} \mathbf{x} | \mathbf{M}^* \mathbf{x} \rangle && \text{by Proposition O.3 page 320 (definition of adjoint)} \\
 &= \langle \mathbf{M} \mathbf{x} | \mathbf{M} \mathbf{x} \rangle && \text{by Theorem O.13 page 321 (property of adjoint)} \\
 &= \|\mathbf{M} \mathbf{x}\|^2 && \text{by definition} \\
 &= \|\mathbf{x}\|^2 && \text{by left hypothesis with } y = 0 \\
 &= \langle \mathbf{x} | \mathbf{x} \rangle && \text{by definition} \\
 &= \langle \mathbf{I} \mathbf{x} | \mathbf{x} \rangle && \text{by Definition O.2 page 307 (definition of I)} \\
 \implies \mathbf{M}^* \mathbf{M} &= \mathbf{I} && \forall \mathbf{x} \in X
 \end{aligned}$$



**Theorem O.24.**<sup>42</sup> Let  $\mathcal{B}(\mathbf{X}, \mathbf{Y})$  be the space of BOUNDED LINEAR OPERATORS on normed linear spaces  $\mathbf{X}$  and  $\mathbf{Y}$ . Let  $\mathbf{M}$  be a bounded linear operator in  $\mathcal{B}(\mathbf{X}, \mathbf{Y})$ , and  $\mathbf{I}$  the identity operator in  $\mathcal{L}(\mathbf{X}, \mathbf{X})$ . Let  $\Lambda$  be the set of eigenvalues of  $\mathbf{M}$ . Let  $\|\mathbf{x}\| \triangleq \sqrt{\langle \mathbf{x} | \mathbf{x} \rangle}$ .

THM	$\underbrace{\mathbf{M}^* \mathbf{M} = \mathbf{I}}$	$\implies \left\{ \begin{array}{l} \ \mathbf{M}\  = 1 \quad (\text{UNIT LENGTH}) \quad \text{and} \\  \lambda  = 1 \quad \forall \lambda \in \Lambda \end{array} \right.$
M is isometric		



PROOF:

1. Proof that  $\mathbf{M}^* \mathbf{M} = \mathbf{I} \implies \|\mathbf{M}\| = 1$ :

$$\begin{aligned}
 \|\mathbf{M}\| &= \sup_{\mathbf{x} \in \mathbf{X}} \{ \|\mathbf{M} \mathbf{x}\| \mid \|\mathbf{x}\| = 1 \} && \text{by Definition O.4 page 312} \\
 &= \sup_{\mathbf{x} \in \mathbf{X}} \{ \|\mathbf{x}\| \mid \|\mathbf{x}\| = 1 \} && \text{by Theorem O.23 page 328} \\
 &= \sup_{\mathbf{x} \in \mathbf{X}} \{ 1 \} \\
 &= 1
 \end{aligned}$$

2. Proof that  $|\lambda| = 1$ : Let  $(\mathbf{x}, \lambda)$  be an eigenvector-eigenvalue pair.

$$\begin{aligned}
 1 &= \frac{1}{\|\mathbf{x}\|} \|\mathbf{x}\| \\
 &= \frac{1}{\|\mathbf{x}\|} \|\mathbf{M} \mathbf{x}\| && \text{by Theorem O.23 page 328} \\
 &= \frac{1}{\|\mathbf{x}\|} \|\lambda \mathbf{x}\| && \text{by definition of } \lambda \\
 &= \frac{1}{\|\mathbf{x}\|} |\lambda| \|\mathbf{x}\| && \text{by homogeneous property of } \|\cdot\| \\
 &= |\lambda|
 \end{aligned}$$



*Example O.4 (One sided shift operator).*<sup>43</sup> Let  $\mathbf{X}$  be the set of all sequences with range  $\mathbb{W}(0, 1, 2, \dots)$  and shift operators defined as

1.  $\mathbf{S}_r((x_0, x_1, x_2, \dots)) \triangleq (0, x_0, x_1, x_2, \dots)$  (right shift operator)
2.  $\mathbf{S}_l((x_0, x_1, x_2, \dots)) \triangleq (x_1, x_2, x_3, \dots)$  (left shift operator)

EX	1. $\mathbf{S}_r$ is an isometric operator. 2. $\mathbf{S}_r^* = \mathbf{S}_l$
----	---



<sup>42</sup> Michel and Herget (1993) page 432

<sup>43</sup> Michel and Herget (1993) page 441

PROOF:

1. Proof that  $\mathbf{S}_r^* = \mathbf{S}_l$ :

$$\begin{aligned}
 \langle \mathbf{S}_r(x_0, x_1, x_2, \dots) | (y_0, y_1, y_2, \dots) \rangle &= \langle (0, x_0, x_1, x_2, \dots) | (y_0, y_1, y_2, \dots) \rangle \\
 &= \sum_{n=1}^{\infty} x_{n-1} y_n^* \\
 &= \sum_{n=0}^{\infty} x_n y_{n+1}^* \\
 &= \sum_{n=0}^{\infty} x_n y_{n+1}^* \\
 &= \langle (x_0, x_1, x_2, \dots) | (y_1, y_2, y_3, \dots) \rangle \\
 &= \left\langle (x_0, x_1, x_2, \dots) | \underbrace{\mathbf{S}_l(y_0, y_1, y_2, \dots)}_{\mathbf{S}_r^*} \right\rangle
 \end{aligned}$$

2. Proof that  $\mathbf{S}_r$  is isometric ( $\mathbf{S}_r^* \mathbf{S}_r = \mathbf{I}$ ):

$$\begin{aligned}
 \mathbf{S}_r^* \mathbf{S}_r &= \mathbf{S}_l \mathbf{S}_r \\
 &= \mathbf{I}
 \end{aligned}
 \quad \text{by 1.}$$

⇒

## O.4.5 Unitary operators

**Definition O.11.** <sup>44</sup> Let  $\mathcal{B}(X, Y)$  be the space of BOUNDED LINEAR OPERATORS on normed linear spaces  $X$  and  $Y$ . Let  $\mathbf{U}$  be a bounded linear operator in  $\mathcal{B}(X, Y)$ , and  $\mathbf{I}$  the identity operator in  $\mathcal{B}(X, X)$ .

D  
E  
F

The operator  $\mathbf{U}$  is **unitary** if  $\mathbf{U}^* \mathbf{U} = \mathbf{U} \mathbf{U}^* = \mathbf{I}$ .

**Proposition O.4.** Let  $\mathcal{B}(X, Y)$  be the space of BOUNDED LINEAR OPERATORS on normed linear spaces  $X$  and  $Y$ . Let  $\mathbf{U}$  and  $\mathbf{V}$  be BOUNDED LINEAR OPERATORS in  $\mathcal{B}(X, Y)$ .

P       $\mathbf{U}$  is UNITARY    and     $\left. \begin{array}{l} \mathbf{V} \text{ is UNITARY} \\ \mathbf{V} \text{ is UNITARY} \end{array} \right\} \Rightarrow (\mathbf{UV}) \text{ is UNITARY.}$

PROOF:

$$\begin{aligned}
 (\mathbf{UV})(\mathbf{UV})^* &= (\mathbf{UV})(\mathbf{V}^* \mathbf{U}^*) && \text{by Theorem O.8 page 317} \\
 &= \mathbf{U}(\mathbf{V} \mathbf{V}^*) \mathbf{U}^* && \text{by associative property} \\
 &= \mathbf{U} \mathbf{I} \mathbf{U}^* && \text{by definition of unitary operators} && \text{(Definition O.11 page 330)} \\
 &= \mathbf{I} && \text{by definition of unitary operators} && \text{(Definition O.11 page 330)}
 \end{aligned}$$

$$\begin{aligned}
 (\mathbf{UV})^*(\mathbf{UV}) &= (\mathbf{V}^* \mathbf{U}^*)(\mathbf{UV}) && \text{by Theorem O.8 page 317} \\
 &= \mathbf{V}^* (\mathbf{U}^* \mathbf{U}) \mathbf{V} && \text{by associative property} \\
 &= \mathbf{V}^* \mathbf{I} \mathbf{V} && \text{by definition of unitary operators} && \text{(Definition O.11 page 330)} \\
 &= \mathbf{I} && \text{by definition of unitary operators} && \text{(Definition O.11 page 330)}
 \end{aligned}$$

<sup>44</sup> Rudin (1991) page 312, Michel and Herget (1993) page 431, Autonne (1901) page 209, Autonne (1902), Schur (1909), Steen (1973)





**Theorem O.25.** <sup>45</sup> Let  $\mathcal{B}(H, H)$  be the space of BOUNDED LINEAR OPERATORS on a HILBERT SPACE  $H$ . Let  $\mathcal{I}(U)$  be the IMAGE SET of  $U$ .

If  $U$  is a bounded linear operator ( $U \in \mathcal{B}(H, H)$ ), then the following conditions are equivalent:

- |                      |   |                                      |
|----------------------|---|--------------------------------------|
| <b>T<br/>H<br/>M</b> | 1. $UU^* = U^*U = I$ <span style="float: right;">(UNITARY)</span><br>2. $\langle Ux   Uy \rangle = \langle U^*x   U^*y \rangle = \langle x   y \rangle$ and $\mathcal{I}(U) = X$ <span style="float: right;">(SURJECTIVE)</span><br>3. $\ Ux - Uy\  = \ U^*x - U^*y\  = \ x - y\ $ and $\mathcal{I}(U) = X$ <span style="float: right;">(ISOMETRIC IN DISTANCE)</span><br>4. $\ Ux\  = \ x\ $ and $\mathcal{I}(U) = X$ <span style="float: right;">(ISOMETRIC IN LENGTH)</span> | $\iff$<br>$\iff$<br>$\iff$<br>$\iff$ |
|----------------------|---|--------------------------------------|

PROOF:

1. Proof that (1)  $\implies$  (2):

$$(a) \langle Ux | Uy \rangle = \langle U^*x | U^*y \rangle = \langle x | y \rangle \text{ by Theorem O.23 (page 328).}$$

(b) Proof that  $\mathcal{I}(U) = X$ :

$$\begin{aligned} X &\supseteq \mathcal{I}(U) && \text{because } U \in X^X \\ &\supseteq \mathcal{I}(UU^*) \\ &= \mathcal{I}(I) && \text{by left hypothesis } (U^*U = UU^* = I) \\ &= X && \text{by Definition O.2 page 307 (definition of } I) \end{aligned}$$

2. Proof that (2)  $\iff$  (3)  $\iff$  (4): by Theorem O.23 page 328.

3. Proof that (3)  $\implies$  (1):

(a) Proof that  $\|Ux - Uy\| = \|x - y\| \implies U^*U = I$ : by Theorem O.23 page 328

(b) Proof that  $\|U^*x - U^*y\| = \|x - y\| \implies UU^* = I$ :

$$\begin{aligned} \|U^*x - U^*y\| = \|x - y\| \implies U^{**}U^* = I && \text{by Theorem O.23 page 328} \\ UU^* = I && \text{by Theorem O.13 page 321} \end{aligned}$$



**Theorem O.26.** Let  $\mathcal{B}(H, H)$  be the space of BOUNDED LINEAR OPERATORS on a HILBERT SPACE  $H$ . Let  $U$  be a bounded linear operator in  $\mathcal{B}(H, H)$ ,  $\mathcal{N}(U)$  the NULL SPACE of  $U$ , and  $\mathcal{I}(U)$  the IMAGE SET of  $U$ .

<b>T H M</b>	$\underbrace{UU^* = U^*U = I}_{U \text{ is unitary}} \implies \left\{ \begin{array}{lcl} U^{-1} & = & U^* & \text{and} \\ \mathcal{I}(U) & = & \mathcal{I}(U^*) & = & X & \text{and} \\ \mathcal{N}(U) & = & \mathcal{N}(U^*) & = & \{\emptyset\} & \text{and} \\ \ U\  & = & \ U^*\  & = & 1 & (\text{UNIT LENGTH}) \end{array} \right\}$
----------------------	--

PROOF:

1. Note that  $U$ ,  $U^*$ , and  $U^{-1}$  are all both *isometric* and *normal*:

$$\begin{aligned} U^*U &= I \implies U \text{ is isometric} \\ UU^* &= U^*U = I \implies U^* \text{ is isometric} \\ U^{-1} &= U^* \implies U^{-1} \text{ is isometric} \end{aligned}$$

$$\begin{aligned} U^*U &= UU^* = I \implies U \text{ is normal} \\ UU^* &= U^*U = I \implies U^* \text{ is normal} \\ U^{-1} &= U^* \implies U^{-1} \text{ is normal} \end{aligned}$$

<sup>45</sup> Rudin (1991) pages 313–314 (Theorem 12.13), Knapp (2005a) page 45 (Proposition 2.6)

2. Proof that  $\mathbf{U}^*\mathbf{U} = \mathbf{U}\mathbf{U}^* = \mathbf{I} \implies \mathcal{I}(\mathbf{U}) = \mathcal{I}(\mathbf{U}^*) = \mathbf{H}$ : by Theorem O.25 page 331.

3. Proof that  $\mathbf{U}^*\mathbf{U} = \mathbf{U}\mathbf{U}^* = \mathbf{I} \implies \mathcal{N}(\mathbf{U}) = \mathcal{N}(\mathbf{U}^*) = \mathcal{N}(\mathbf{U}^{-1})$ :

$$\begin{aligned}\mathcal{N}(\mathbf{U}^*) &= \mathcal{N}(\mathbf{U}) && \text{because } \mathbf{U} \text{ and } \mathbf{U}^* \text{ are both } \textit{normal} \text{ and by Theorem O.21 page 326} \\ &= \mathcal{I}(\mathbf{U})^\perp && \text{by Theorem O.14 page 322} \\ &= X^\perp && \text{by above result} \\ &= \{\emptyset\} && \text{by Proposition I.6 page 245}\end{aligned}$$

4. Proof that  $\mathbf{U}^*\mathbf{U} = \mathbf{U}\mathbf{U}^* = \mathbf{I} \implies \|\mathbf{U}^{-1}\| = \|\mathbf{U}^*\| = \|\mathbf{U}\| = 1$ :

Because  $\mathbf{U}$ ,  $\mathbf{U}^*$ , and  $\mathbf{U}^{-1}$  are all isometric and by Theorem O.24 page 329.



*Example O.5 (Rotation matrix).* <sup>46</sup>

<b>E</b> <b>X</b>	$\left\{ \mathbf{R}_\theta \triangleq \underbrace{\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}}_{\text{rotation matrix } \mathbf{R}_\theta : \mathbb{R}^2 \rightarrow \mathbb{R}^2} \right\}$	$\implies$	$\left\{ \begin{array}{l} (1). \quad \mathbf{R}_{-\theta}^{-1} = \mathbf{R}_{-\theta} \quad \text{and} \\ (2). \quad \mathbf{R}_\theta^* = \mathbf{R}_{-\theta}^{-1} \quad (\mathbf{R} \text{ is unitary}) \end{array} \right\}$
----------------------	--	------------	--

PROOF:

$$\begin{aligned}\mathbf{R}^* &= \mathbf{R}^H \\ &= \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}^H && \text{by definition of } \mathbf{R} \\ &= \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} && \text{by definition of } \textit{Hermitian transpose operator } H \\ &= \begin{bmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{bmatrix} && \text{by Theorem ?? page ??} \\ &= \mathbf{R}_{-\theta} && \text{by definition of } \mathbf{R} \\ &= \mathbf{R}^{-1} && \text{by 1.}\end{aligned}$$



*Example O.6.* <sup>47</sup> Let  $\mathbf{A}$  and  $\mathbf{B}$  be matrix operators.

<b>E</b> <b>X</b>	$\mathbf{A} \triangleq \begin{bmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\mathbf{B} \triangleq \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$
$\mathbf{A}$ is a <i>rotation operator</i> . $\mathbf{B}$ is a <i>reflection operator</i> .		
$\mathbf{A}$ and $\mathbf{B}$ are <b>unitary</b> .		

*Example O.7.* Examples of *Fredholm integral operators* include

<b>E</b> <b>X</b>	1. <b>Fourier Transform</b> $[\tilde{\mathbf{F}}\mathbf{x}](f) = \int_{t \in \mathbb{R}} \mathbf{x}(t) e^{-i2\pi f t} dt$ $\kappa(t, f) = e^{-i2\pi f t}$ 2. <b>Inverse Fourier Transform</b> $[\tilde{\mathbf{F}}^{-1}\tilde{\mathbf{x}}](t) = \int_{f \in \mathbb{R}} \tilde{\mathbf{x}}(f) e^{i2\pi f t} df$ $\kappa(f, t) = e^{i2\pi f t}$ 3. <b>Laplace operator</b> $[\mathbf{L}\mathbf{x}](s) = \int_{t \in \mathbb{R}} \mathbf{x}(t) e^{-st} dt$ $\kappa(t, s) = e^{-st}$
----------------------	---

*Example O.8 (Translation operator).* Let  $\mathbf{T} = L^2_{\mathbb{R}}$  and  $\mathbf{T} \in \mathbf{X}^{\mathbf{X}}$  be defined as

$$\mathbf{T}\mathbf{f}(x) \triangleq \mathbf{f}(x - 1) \quad \forall \mathbf{f} \in L^2_{\mathbb{R}} \quad (\text{translation operator})$$

<sup>46</sup> Noble and Daniel (1988) page 311

<sup>47</sup> Gel'fand (1963), page 4, Gelfand et al. (2018) page 4

<b>E</b>	1. $\mathbf{T}^{-1}\mathbf{f}(x) = \mathbf{f}(x+1)$ $\forall \mathbf{f} \in L^2_{\mathbb{R}}$ (inverse translation operator)
<b>X</b>	2. $\mathbf{T}^* = \mathbf{T}^{-1}$ ( $\mathbf{T}$ is invertible)
<b>X</b>	3. $\mathbf{T}^*\mathbf{T} = \mathbf{T}\mathbf{T}^* = \mathbf{I}$ ( $\mathbf{T}$ is unitary)

PROOF:

1. Proof that  $\mathbf{T}^{-1}\mathbf{f}(x) = \mathbf{f}(x+1)$ :

$$\begin{aligned}\mathbf{T}^{-1}\mathbf{T} &= \mathbf{I} \\ \mathbf{T}\mathbf{T}^{-1} &= \mathbf{I}\end{aligned}$$

2. Proof that  $\mathbf{T}$  is unitary:

$$\begin{aligned}\langle \mathbf{T}\mathbf{f}(x) | \mathbf{g}(x) \rangle &= \langle \mathbf{f}(x-1) | \mathbf{g}(x) \rangle && \text{by definition of } \mathbf{T} \\ &= \int_x \mathbf{f}(x-1)\mathbf{g}^*(x) dx \\ &= \int_x \mathbf{f}(x)\mathbf{g}^*(x+1) dx \\ &= \langle \mathbf{f}(x) | \mathbf{g}(x+1) \rangle \\ &= \left\langle \mathbf{f}(x) | \underbrace{\mathbf{T}^{-1}\mathbf{g}(x)}_{\mathbf{T}^*} \right\rangle && \text{by 1.}\end{aligned}$$



Example O.9 (Dilation operator). Let  $\mathbf{X} = L^2_{\mathbb{R}}$  and  $\mathbf{T} \in \mathbf{X}^{\mathbf{X}}$  be defined as

$$\mathbf{D}\mathbf{f}(x) \triangleq \sqrt{2}\mathbf{f}(2x) \quad \forall \mathbf{f} \in L^2_{\mathbb{R}} \quad (\text{dilation operator})$$

<b>E</b>	1. $\mathbf{D}^{-1}\mathbf{f}(x) = \frac{1}{\sqrt{2}}\mathbf{f}\left(\frac{1}{2}x\right)$ $\forall \mathbf{f} \in L^2_{\mathbb{R}}$ (inverse dilation operator)
<b>X</b>	2. $\mathbf{D}^* = \mathbf{D}^{-1}$ ( $\mathbf{D}$ is invertible)
<b>X</b>	3. $\mathbf{D}^*\mathbf{D} = \mathbf{D}\mathbf{D}^* = \mathbf{I}$ ( $\mathbf{D}$ is unitary)

PROOF:

1. Proof that  $\mathbf{D}^{-1}\mathbf{f}(x) = \frac{1}{\sqrt{2}}\mathbf{f}\left(\frac{1}{2}x\right)$ :

$$\begin{aligned}\mathbf{D}^{-1}\mathbf{D} &= \mathbf{I} \\ \mathbf{D}\mathbf{D}^{-1} &= \mathbf{I}\end{aligned}$$

2. Proof that  $\mathbf{D}$  is unitary:

$$\begin{aligned}\langle \mathbf{D}\mathbf{f}(x) | \mathbf{g}(x) \rangle &= \left\langle \sqrt{2}\mathbf{f}(2x) | \mathbf{g}(x) \right\rangle && \text{by definition of } \mathbf{D} \\ &= \int_x \sqrt{2}\mathbf{f}(2x)\mathbf{g}^*(x) dx \\ &= \int_{u \in \mathbb{R}} \sqrt{2}\mathbf{f}(u)\mathbf{g}^*\left(\frac{1}{2}u\right) \frac{1}{2} du && \text{let } u \triangleq 2x \implies dx = \frac{1}{2} du \\ &= \int_{u \in \mathbb{R}} \mathbf{f}(u) \left[ \frac{1}{\sqrt{2}}\mathbf{g}\left(\frac{1}{2}u\right) \right]^* du \\ &= \left\langle \mathbf{f}(x) | \underbrace{\frac{1}{\sqrt{2}}\mathbf{g}\left(\frac{1}{2}x\right)}_{\mathbf{D}^*} \right\rangle \\ &= \left\langle \mathbf{f}(x) | \underbrace{\mathbf{D}^{-1}\mathbf{g}(x)}_{\mathbf{D}^*} \right\rangle && \text{by 1.}\end{aligned}$$



*Example O.10* (Delay operator). Let  $\mathbf{X}$  be the set of all sequences and  $\mathbf{D} \in \mathbf{X}^{\mathbf{X}}$  be a delay operator.

**E X** 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

$$\mathbf{D}^{-1}((x_n)_{n \in \mathbb{Z}}) \triangleq ((x_{n+1})_{n \in \mathbb{Z}}).$$

$$\begin{aligned} \langle \mathbf{D}((x_n)) | (y_n) \rangle &= \langle ((x_{n-1})) | (y_n) \rangle && \text{by definition of } \mathbf{D} \\ &= \sum_n x_{n-1} y_n^* \\ &= \sum_n x_n y_{n+1}^* \\ &= \langle ((x_n)) | ((y_{n+1})) \rangle \\ &= \left\langle ((x_n)) | \underbrace{\mathbf{D}^{-1}((y_n))}_{\mathbf{D}^*} \right\rangle \end{aligned}$$

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.  $\Rightarrow$

*Example O.11* (Fourier transform). Let  $\tilde{\mathbf{F}}$  be the *Fourier Transform* and  $\tilde{\mathbf{F}}^{-1}$  the *inverse Fourier Transform* operator

$$[\tilde{\mathbf{F}}\mathbf{x}](f) \triangleq \int_t \mathbf{x}(t) e^{-i2\pi f t} dt \quad [\tilde{\mathbf{F}}^{-1}\tilde{\mathbf{x}}](t) \triangleq \int_f \tilde{\mathbf{x}}(f) e^{i2\pi f t} df.$$

**E X**  $\tilde{\mathbf{F}}^* = \tilde{\mathbf{F}}^{-1}$  (the Fourier Transform operator  $\tilde{\mathbf{F}}$  is unitary)

PROOF:

$$\begin{aligned} \langle \tilde{\mathbf{F}}\mathbf{x} | \tilde{\mathbf{y}} \rangle &= \left\langle \int_t \mathbf{x}(t) e^{-i2\pi f t} dt | \tilde{\mathbf{y}}(f) \right\rangle \\ &= \int_t \mathbf{x}(t) \left\langle e^{-i2\pi f t} | \tilde{\mathbf{y}}(f) \right\rangle dt \\ &= \int_t \mathbf{x}(t) \int_f e^{-i2\pi f t} \tilde{\mathbf{y}}^*(f) df dt \\ &= \int_t \mathbf{x}(t) \left[ \int_f e^{i2\pi f t} \tilde{\mathbf{y}}(f) df \right]^* dt \\ &= \left\langle \mathbf{x}(t) | \int_f \tilde{\mathbf{y}}(f) e^{i2\pi f t} df \right\rangle \\ &= \left\langle \mathbf{x} | \underbrace{\tilde{\mathbf{F}}^{-1}\tilde{\mathbf{y}}}_{\tilde{\mathbf{F}}^*} \right\rangle \end{aligned}$$

This implies that  $\tilde{\mathbf{F}}$  is unitary ( $\tilde{\mathbf{F}}^* = \tilde{\mathbf{F}}^{-1}$ ).  $\Rightarrow$



## O.5 Operator order

**Definition O.12.** <sup>48</sup> Let  $P \in Y^X$  be an operator.

**D E F** **P is positive if**  $\langle Px | x \rangle \geq 0 \forall x \in X$ .  
**This condition is denoted**  $P \geq 0$ .

**Theorem O.27.** <sup>49</sup>

<b>T H M</b>	$\underbrace{P \geq 0 \text{ and } Q \geq 0}_{P \text{ and } Q \text{ are both positive}}$	$\implies$	$\begin{cases} (P + Q) \geq 0 & (\text{$(P + Q)$ is positive}) \\ A^*PA \geq 0 & (\text{$A^*PA$ is positive}) \\ A^*A \geq 0 & (\text{$A^*A$ is positive}) \end{cases}$
--------------	--	------------	---

PROOF:

$$\begin{aligned}
 \langle (P + Q)x | x \rangle &= \langle Px | x \rangle + \langle Qx | x \rangle && \text{by additive property of } \langle \cdot | \cdot \rangle \text{ (Definition K.1 page 259)} \\
 &\geq \langle Px | x \rangle && \text{by left hypothesis} \\
 &\geq 0 && \text{by left hypothesis} \\
 \langle A^*PAx | x \rangle &= \langle PAx | Ax \rangle && \text{by definition of adjoint (Proposition O.3 page 320)} \\
 &= \langle Py | y \rangle && \text{where } y \triangleq Ax \\
 &\geq 0 && \text{by left hypothesis} \\
 \langle Ix | x \rangle &= \langle x | x \rangle && \text{by definition of } I \text{ (Definition O.2 page 307)} \\
 &\geq 0 && \text{by non-negative property of } \langle \cdot | \cdot \rangle \text{ (Definition K.1 page 259)} \\
 &\implies I \text{ is positive} && \\
 \langle A^*Ax | x \rangle &= \langle A^*IAx | x \rangle && \text{by definition of } I \text{ (Definition O.2 page 307)} \\
 &\geq 0 && \text{by two previous results}
 \end{aligned}$$



**Definition O.13.** <sup>50</sup> Let  $A, B \in \mathcal{B}(X, Y)$  be BOUNDED operators.

**D E F**  **$A \geq B$  ("A is greater than or equal to B") if**  
 **$A - B \geq 0$  ("( $A - B$ ) is positive")**

<sup>48</sup> Michel and Herget (1993) page 429 (Definition 7.4.12)

<sup>49</sup> Michel and Herget (1993) page 429

<sup>50</sup> Michel and Herget (1993) page 429



# APPENDIX P

## FOURIER TRANSFORM



“The analytical equations ... extend to all general phenomena. There cannot be a language more universal and more simple, more free from errors and from obscurities, ... mathematical analysis is as extensive as nature itself; ... it preserves every principle which it has once acquired; it grows and strengthens itself incessantly in the midst of the many variations and errors of the human mind. Its chief attribute is clearness; it has no marks to express confused notions. It brings together phenomena the most diverse, and discovers the hidden analogies which unite them.”

Joseph Fourier (1768–1830)<sup>1</sup>

### P.1 Definitions

This chapter deals with the *Fourier Transform* in the space of *Lebesgue square-integrable functions*  $L^2_{(\mathbb{R}, \mathcal{B}, \mu)}$ , where  $\mathbb{R}$  is the set of real numbers,  $\mathcal{B}$  is the set of *Borel sets* on  $\mathbb{R}$ ,  $\mu$  is the standard *Borel measure* on  $\mathbb{R}$ , and

$$L^2_{(\mathbb{R}, \mathcal{B}, \mu)} \triangleq \left\{ f \in \mathbb{R}^\mathbb{R} \mid \int_{\mathbb{R}} |f|^2 d\mu < \infty \right\}.$$

Furthermore,  $\langle \Delta | \nabla \rangle$  is the *inner product* induced by the operator  $\int_{\mathbb{R}} d\mu$  such that

$$\langle f | g \rangle \triangleq \int_{\mathbb{R}} f(x)g^*(x) dx,$$

and  $(L^2_{(\mathbb{R}, \mathcal{B}, \mu)}, \langle \Delta | \nabla \rangle)$  is a *Hilbert space*.

**Definition P.1.** Let  $\kappa$  be a FUNCTION in  $\mathbb{C}^{\mathbb{R}^2}$ .

**D E F** The function  $\kappa$  is the **Fourier kernel** if  $\kappa(x, \omega) \triangleq e^{i\omega x} \quad \forall x, \omega \in \mathbb{R}$

**Definition P.2.** <sup>2</sup> Let  $L^2_{(\mathbb{R}, \mathcal{B}, \mu)}$  be the space of all LEBESGUE SQUARE-INTEGRABLE FUNCTIONS.

<sup>1</sup> quote: [Fourier \(1878\)](#), pages 7–8 (Preliminary Discourse)

image: <http://en.wikipedia.org/wiki/File:Fourier2.jpg>, public domain

<sup>2</sup> [Bachman et al. \(2000\)](#) 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?

**D E F** The Fourier Transform operator  $\tilde{F}$  is defined as

$$[\tilde{F}f](\omega) \triangleq \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) e^{-i\omega x} dx \quad \forall f \in L^2_{(\mathbb{R}, \mathcal{B}, \mu)}$$

This definition of the Fourier Transform is also called the **unitary Fourier Transform**.

**Remark P.1 (Fourier transform scaling factor).**<sup>3</sup> If the Fourier transform operator  $\tilde{F}$  and inverse Fourier transform operator  $\tilde{F}^{-1}$  are defined as

$$\tilde{F}f(x) \triangleq F(\omega) \triangleq A \int_{\mathbb{R}} f(x) e^{-i\omega x} dx \quad \text{and} \quad \tilde{F}^{-1}\tilde{f}(\omega) \triangleq B \int_{\mathbb{R}} F(\omega) e^{i\omega x} d\omega$$

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  $[\tilde{F}f(x)](\omega) \triangleq \int_{\mathbb{R}} f(x) e^{-i\omega x} dx$ . In this case, the inverse Fourier transform operator  $\tilde{F}^{-1}$  is either defined as

- $[\tilde{F}^{-1}f(x)](f) \triangleq \int_{\mathbb{R}} f(x) e^{i2\pi f x} dx$  (using oscillatory frequency free variable  $f$ ) or
- $[\tilde{F}^{-1}f(x)](\omega) \triangleq \frac{1}{2\pi} \int_{\mathbb{R}} f(x) e^{i\omega x} dx$  (using angular frequency free variable  $\omega$ ).

In short, the  $2\pi$  has to show up somewhere, either in the argument of the exponential ( $e^{-i2\pi f t}$ ) or in front of the integral ( $\frac{1}{2\pi} \int \dots$ ). One could argue that it is unnecessary to burden the exponential argument with the  $2\pi$  factor ( $e^{-i2\pi f t}$ ), and thus could further argue in favor of using the angular frequency variable  $\omega$  thus giving the inverse operator definition  $[\tilde{F}^{-1}f(x)](\omega) \triangleq \frac{1}{2\pi} \int_{\mathbb{R}} f(x) e^{-i\omega x} dx$ . But this causes a new problem. In this case, the Fourier operator  $\tilde{F}$  is not *unitary* (see Theorem P.2 page 338)—in particular,  $\tilde{F}\tilde{F}^* \neq I$ , where  $\tilde{F}^*$  is the *adjoint* of  $\tilde{F}$ ; but rather,  $\tilde{F} \left( \frac{1}{2\pi} \tilde{F}^* \right) = \left( \frac{1}{2\pi} \tilde{F}^* \right) \tilde{F} = I$ . But if we define the operators  $\tilde{F}$  and  $\tilde{F}^{-1}$  to both have the scaling factor  $\frac{1}{\sqrt{2\pi}}$ , then  $\tilde{F}$  and  $\tilde{F}^{-1}$  are inverses and  $\tilde{F}$  is *unitary*—that is,  $\tilde{F}\tilde{F}^* = \tilde{F}^*\tilde{F} = I$ .

## P.2 Operator properties

**Theorem P.1** (Inverse Fourier transform).<sup>4</sup> Let  $\tilde{F}$  be the Fourier Transform operator (Definition P.2 page 337). The inverse  $\tilde{F}^{-1}$  of  $\tilde{F}$  is

**T H M**

$$[\tilde{F}^{-1}\tilde{f}](x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \tilde{f}(\omega) e^{i\omega x} d\omega \quad \forall \tilde{f} \in L^2_{(\mathbb{R}, \mathcal{B}, \mu)}$$

**Theorem P.2.** Let  $\tilde{F}$  be the Fourier Transform operator with inverse  $\tilde{F}^{-1}$  and adjoint  $\tilde{F}^*$ .

**T H M**

$$\tilde{F}^* = \tilde{F}^{-1}$$

PROOF:

$$\begin{aligned} \langle \tilde{F}f | g \rangle &= \left\langle \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) e^{-i\omega x} dx | g(\omega) \right\rangle && \text{by definition of } \tilde{F} \text{ page 337} \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) \langle e^{-i\omega x} | g(\omega) \rangle dx && \text{by additive property of } \langle \cdot | \cdot \rangle \text{ page 259} \\ &= \int_{\mathbb{R}} f(x) \frac{1}{\sqrt{2\pi}} \langle g(\omega) | e^{-i\omega x} \rangle^* dx && \text{by conjugate symmetric property of } \langle \cdot | \cdot \rangle \text{ page 259} \end{aligned}$$

<sup>3</sup> Chorin and Hald (2009) page 13, Jeffrey and Dai (2008) pages xxxi–xxxii, Knapp (2005b) pages 374–375

<sup>4</sup> Chorin and Hald (2009) page 13

$$\begin{aligned}
 &= \left\langle f(x) \mid \frac{1}{\sqrt{2\pi}} \langle g(\omega) \mid e^{-i\omega x} \rangle \right\rangle && \text{by definition of } \langle \triangle \mid \nabla \rangle \\
 &= \left\langle f \mid \underbrace{\tilde{F}^{-1}g}_{\tilde{F}^*} \right\rangle && \text{by Theorem P.1 page 338}
 \end{aligned}$$



The Fourier Transform operator has several nice properties:

- $\tilde{F}$  is *unitary*<sup>5</sup> (Corollary P.1—next corollary).
- Because  $\tilde{F}$  is unitary, it automatically has several other nice properties (Theorem P.3 page 339).

**Corollary P.1.** Let  $\mathbf{I}$  be the identity operator and let  $\tilde{F}$  be the Fourier Transform operator with adjoint  $\tilde{F}^*$  and inverse  $\tilde{F}^{-1}$ .

C O R	$\tilde{F}\tilde{F}^* = \tilde{F}^*\tilde{F} = \mathbf{I}$ ( $\tilde{F}$ is unitary)
$\tilde{F}^* = \tilde{F}^{-1}$	



PROOF: This follows directly from the fact that  $\tilde{F}^* = \tilde{F}^{-1}$  (Theorem P.2 page 338).  $\Rightarrow$

**Theorem P.3.** Let  $\tilde{F}$  be the Fourier transform operator with adjoint  $\tilde{F}^*$  and inverse  $\tilde{F}$ . Let  $\|\cdot\|$  be the operator norm with respect to the vector norm  $\|\cdot\|$  with respect to the Hilbert space  $(\mathbb{C}^{\mathbb{R}}, \langle \triangle \mid \nabla \rangle)$ . Let  $\mathcal{R}(A)$  be the range of an operator  $A$ .

T H M	$  \begin{aligned}  \mathcal{R}(F\tau) &= \mathcal{R}(\tilde{F}^{-1}) &= L^2_{\mathbb{R}} \\  \ \tilde{F}\  &= \ \tilde{F}^{-1}\  &= 1 & \text{(UNITARY)} \\  \langle \tilde{F}f \mid \tilde{F}g \rangle &= \langle \tilde{F}^{-1}f \mid \tilde{F}^{-1}g \rangle &= \langle f \mid g \rangle & \text{(PARSEVAL'S EQUATION)} \\  \ \tilde{F}f\  &= \ \tilde{F}^{-1}f\  &= \ f\  & \text{(PLANCHEREL'S FORMULA)} \\  \ \tilde{F}f - \tilde{F}g\  &= \ \tilde{F}^{-1}f - \tilde{F}^{-1}g\  &= \ f - g\  & \text{(ISOMETRIC)}  \end{aligned}  $
-------------	---



PROOF: These results follow directly from the fact that  $\tilde{F}$  is unitary (Corollary P.1 page 339) and from the properties of unitary operators (Theorem O.26 page 331).  $\Rightarrow$

**Theorem P.4** (Shift relations). Let  $\tilde{F}$  be the Fourier transform operator.

T H M	$  \begin{aligned}  \tilde{F}[f(x-u)](\omega) &= e^{-i\omega u} [\tilde{F}f](\omega) \\  [\tilde{F}(e^{ivx}g(x))](\omega) &= [\tilde{F}g](\omega - v)  \end{aligned}  $
-------------	---

PROOF:

$$\begin{aligned}
 \tilde{F}[f(x-u)](\omega) &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x-u)e^{-i\omega x} dx && \text{by definition of } \tilde{F} && (\text{Definition P.2 page 337}) \\
 &= \frac{1}{\sqrt{2\pi}} \int_{v \in \mathbb{R}} f(v)e^{-i\omega(u+v)} dv && \text{where } v \triangleq x - u \implies t = u + v \\
 &= e^{-i\omega u} \frac{1}{\sqrt{2\pi}} \int_{v \in \mathbb{R}} f(v)e^{-i\omega v} dv \\
 &= e^{-i\omega u} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x)e^{-i\omega x} dx && \text{by change of variable } t = v \\
 &= e^{-i\omega u} [\tilde{F}f](\omega) && \text{by definition of } \tilde{F} && (\text{Definition P.2 page 337}) \\
 [\tilde{F}(e^{ivx}g(x))](\omega) &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{ivx}g(x)e^{-i\omega x} dx && \text{by definition of } \tilde{F} && (\text{Definition P.2 page 337})
 \end{aligned}$$

<sup>5</sup>unitary operators: Definition O.11 page 330

$$\begin{aligned}
 &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} g(x) e^{-i(\omega-v)x} dx \\
 &= [\tilde{\mathbf{F}}g(x)](\omega - v) \quad \text{by definition of } \tilde{\mathbf{F}} \quad (\text{Definition P.2 page 337})
 \end{aligned}$$

⇒

**Theorem P.5** (Complex conjugate). *Let  $\tilde{\mathbf{F}}$  be the Fourier Transform operator and  $*$  represent the complex conjugate operation on the set of complex numbers.*

T H M	$\tilde{\mathbf{F}}f^*(-x) = -[\tilde{\mathbf{F}}f(x)]^* \quad \forall f \in L^2_{(\mathbb{R}, \mathcal{B}, \mu)}$ $f \text{ is real} \implies \tilde{f}(-\omega) = [\tilde{f}(\omega)]^* \quad \forall \omega \in \mathbb{R}$ <span style="float: right;">REALITY CONDITION</span>
-------------	--

PROOF:

$$\begin{aligned}
 [\tilde{\mathbf{F}}f^*(-x)](\omega) &\triangleq \frac{1}{\sqrt{2\pi}} \int f^*(-x) e^{-i\omega x} dx \quad \text{by definition of } \tilde{\mathbf{F}} \quad (\text{Definition P.2 page 337}) \\
 &= \frac{1}{\sqrt{2\pi}} \int f^*(u) e^{i\omega u} (-1) du \quad \text{where } u \triangleq -x \implies dx = -du \\
 &= - \left[ \frac{1}{\sqrt{2\pi}} \int f(u) e^{-i\omega u} du \right]^* \\
 &\triangleq -[\tilde{\mathbf{F}}f(x)]^* \quad \text{by definition of } \tilde{\mathbf{F}} \quad (\text{Definition P.2 page 337}) \\
 \tilde{f}(-\omega) &\triangleq \frac{1}{\sqrt{2\pi}} \int f(x) e^{-i(-\omega)x} dx \quad \text{by definition of } \tilde{\mathbf{F}} \quad (\text{Definition P.2 page 337}) \\
 &= \left[ \frac{1}{\sqrt{2\pi}} \int f^*(x) e^{-i\omega x} dx \right]^* \\
 &= \left[ \frac{1}{\sqrt{2\pi}} \int f(x) e^{-i\omega x} dx \right]^* \quad \text{by } f \text{ is real hypothesis} \\
 &\triangleq \tilde{f}^*(\omega) \quad \text{by definition of } \tilde{\mathbf{F}} \quad (\text{Definition P.2 page 337})
 \end{aligned}$$

⇒

## P.3 Convolution

**Definition P.3.** <sup>6</sup>

**D E F** *The convolution operation is defined as*

D E F	$[f \star g](x) \triangleq f(x) \star g(x) \triangleq \int_{u \in \mathbb{R}} f(u)g(x-u) du \quad \forall f, g \in L^2_{(\mathbb{R}, \mathcal{B}, \mu)}$
-------------	--

Theorem R.2 (next) demonstrates that multiplication in the “time domain” is equivalent to convolution in the “frequency domain” and vice-versa.

**Theorem P.6** (convolution theorem). <sup>7</sup> *Let  $\tilde{\mathbf{F}}$  be the Fourier Transform operator and  $\star$  the convolution operator.*

<sup>6</sup>  Bachman (1964), page 6,  Bracewell (1978) page 108 (Convolution theorem)

<sup>7</sup>  Bracewell (1978) page 110

<b>T H M</b>	$\underbrace{\tilde{F}[f(x) \star g(x)](\omega)}_{\text{convolution in "time domain"} } = \underbrace{\sqrt{2\pi} [\tilde{F}f](\omega) [\tilde{F}g](\omega)}_{\text{multiplication in "frequency domain"} } \quad \forall f, g \in L^2_{(\mathbb{R}, \mathcal{B}, \mu)}$ $\underbrace{\tilde{F}[f(x)g(x)](\omega)}_{\text{multiplication in "time domain"} } = \underbrace{\frac{1}{\sqrt{2\pi}} [\tilde{F}f](\omega) \star [\tilde{F}g](\omega)}_{\text{convolution in "frequency domain"} } \quad \forall f, g \in L^2_{(\mathbb{R}, \mathcal{B}, \mu)}.$
----------------------	---

PROOF:

$$\begin{aligned}
 \tilde{F}[f(x) \star g(x)](\omega) &= \tilde{F}\left[\int_{u \in \mathbb{R}} f(u)g(x-u) du\right](\omega) && \text{by definition of } \star \text{ (Definition P.3 page 340)} \\
 &= \int_{u \in \mathbb{R}} f(u)[\tilde{F}g(x-u)](\omega) du \\
 &= \int_{u \in \mathbb{R}} f(u)e^{-i\omega u} [\tilde{F}g(x)](\omega) du && \text{by Theorem P.4 page 339} \\
 &= \sqrt{2\pi} \left( \underbrace{\frac{1}{\sqrt{2\pi}} \int_{u \in \mathbb{R}} f(u)e^{-i\omega u} du}_{[\tilde{F}f](\omega)} \right) [\tilde{F}g](\omega) \\
 &= \sqrt{2\pi} [\tilde{F}f](\omega) [\tilde{F}g](\omega) && \text{by definition of } \tilde{F} \text{ (Definition P.2 page 337)} \\
 \tilde{F}[f(x)g(x)](\omega) &= \tilde{F}[(\tilde{F}^{-1}\tilde{F}f(x))g(x)](\omega) && \text{by definition of operator inverse (page 307)} \\
 &= \tilde{F}\left[\left(\frac{1}{\sqrt{2\pi}} \int_{v \in \mathbb{R}} [\tilde{F}f(x)](v)e^{ivx} dv\right) g(x)\right](\omega) && \text{by Theorem P.1 page 338} \\
 &= \frac{1}{\sqrt{2\pi}} \int_{v \in \mathbb{R}} [\tilde{F}f(x)](v) [\tilde{F}(e^{ivx} g(x))](\omega, v) dv \\
 &= \frac{1}{\sqrt{2\pi}} \int_{v \in \mathbb{R}} [\tilde{F}f(x)](v) [\tilde{F}g(x)](\omega - v) dv && \text{by Theorem P.4 page 339} \\
 &= \frac{1}{\sqrt{2\pi}} [\tilde{F}f](\omega) \star [\tilde{F}g](\omega) && \text{by definition of } \star \text{ (Definition P.3 page 340)}
 \end{aligned}$$



## P.4 Real valued functions

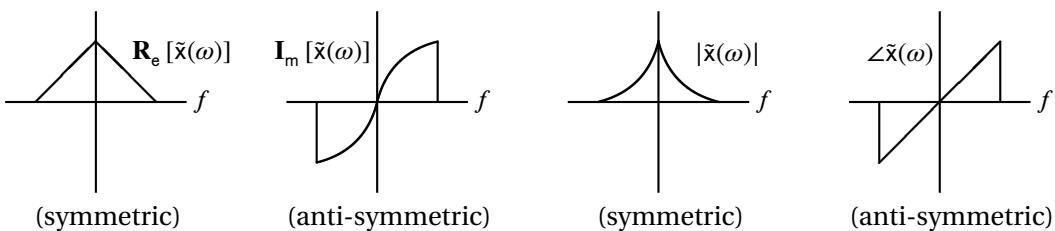


Figure P.1: Fourier transform components of real-valued signal

**Theorem P.7.** Let  $f(x)$  be a function in  $L^2_{\mathbb{R}}$  and  $\tilde{f}(\omega)$  the FOURIER TRANSFORM of  $f(x)$ .

<b>T H M</b>	$  \left\{ \begin{array}{l} f(x) \text{ is REAL-VALUED} \\ (f \in \mathbb{R}^{\mathbb{R}}) \end{array} \right\} \implies \left\{ \begin{array}{lcl} \tilde{f}(\omega) & = & \tilde{f}^*(-\omega) & (\text{HERMITIAN SYMMETRIC}) \\ \mathbf{R}_e [\tilde{f}(\omega)] & = & \mathbf{R}_e [\tilde{f}(-\omega)] & (\text{SYMMETRIC}) \\ \mathbf{I}_m [\tilde{f}(\omega)] & = & -\mathbf{I}_m [\tilde{f}(-\omega)] & (\text{ANTI-SYMMETRIC}) \\  \tilde{f}(\omega)  & = &  \tilde{f}(-\omega)  & (\text{SYMMETRIC}) \\ \angle \tilde{f}(\omega) & = & \angle \tilde{f}(-\omega) & (\text{ANTI-SYMMETRIC}). \end{array} \right\}  $
----------------------	---

PROOF:

$$\begin{aligned}
 \tilde{f}(\omega) &\triangleq [\tilde{\mathbf{F}}f(x)](\omega) \triangleq \langle f(x) | e^{i\omega x} \rangle = \langle f(x) | e^{i(-\omega)x} \rangle^* \triangleq \tilde{f}^*(-\omega) \\
 \mathbf{R}_e[\tilde{f}(\omega)] &= \mathbf{R}_e[\tilde{f}^*(-\omega)] = \mathbf{R}_e[\tilde{f}(-\omega)] \\
 \mathbf{I}_m[\tilde{f}(\omega)] &= \mathbf{I}_m[\tilde{f}^*(-\omega)] = -\mathbf{I}_m[\tilde{f}(-\omega)] \\
 |\tilde{f}(\omega)| &= |\tilde{f}^*(-\omega)| = |\tilde{f}(-\omega)| \\
 \angle\tilde{f}(\omega) &= \angle\tilde{f}^*(-\omega) = -\angle\tilde{f}(-\omega)
 \end{aligned}$$

⇒

## P.5 Moment properties

Definition P.4.<sup>8</sup>

DEF

The quantity  $M_n$  is the ***n*th moment** of a function  $f(x) \in L^2_{\mathbb{R}}$  if

$$M_n \triangleq \int_{\mathbb{R}} x^n f(x) dx \quad \text{for } n \in \mathbb{W}.$$

**Lemma P.1.**<sup>9</sup> Let  $M_n$  be the ***n*th moment** (Definition P.4 page 342) and  $\tilde{f}(\omega) \triangleq [\tilde{\mathbf{F}}f](\omega)$  the Fourier transform (Definition P.2 page 337) of a function  $f(x)$  in  $L^2_{\mathbb{R}}$  (Definition ?? page ??).

LEM

$$M_n = \sqrt{2\pi}(i)^n \left[ \frac{d}{d\omega} \right]^n \tilde{f}(\omega) \Big|_{\omega=0} \quad \forall n \in \mathbb{W}, f \in L^2_{\mathbb{R}}$$

$$\left[ \frac{d}{d\omega} \right]^n \tilde{f}(\omega) \Big|_{\omega=0} = \frac{1}{\sqrt{2\pi}} (-i)^n M_n \quad \forall n \in \mathbb{W}, f \in L^2_{\mathbb{R}}$$

PROOF:

$$\begin{aligned}
 \sqrt{2\pi}(i)^n \left[ \left[ \frac{d}{d\omega} \right]^n \tilde{f}(\omega) \right]_{\omega=0} &= \sqrt{2\pi}(i)^n \left[ \left[ \frac{d}{d\omega} \right]^n \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) e^{-i\omega x} dx \right]_{\omega=0} \quad \text{by definition of } \tilde{\mathbf{F}} \text{ (Definition P.2 page 337)} \\
 &= (i)^n \int_{\mathbb{R}} f(x) \left[ \left[ \frac{d}{d\omega} \right]^n e^{-i\omega x} \right] dx \Big|_{\omega=0} \\
 &= (i)^n \int_{\mathbb{R}} f(x) [(-i)^n x^n e^{-i\omega x}] dx \Big|_{\omega=0} \\
 &= (-i^2)^n \int_{\mathbb{R}} f(x) x^n dx \\
 &= \int_{\mathbb{R}} f(x) x^n dx \\
 &\triangleq M_n \quad \text{by definition of } M_n \text{ (Definition P.4 page 342)}
 \end{aligned}$$

⇒

**Lemma P.2.**<sup>10</sup> Let  $M_n$  be the ***n*th moment** (Definition P.4 page 342) and  $\tilde{f}(\omega) \triangleq [\tilde{\mathbf{F}}f](\omega)$  the Fourier transform (Definition P.2 page 337) of a function  $f(x)$  in  $L^2_{\mathbb{R}}$  (Definition ?? page ??).

LEM

$$M_n = 0 \iff \left[ \frac{d}{d\omega} \right]^n \tilde{f}(\omega) \Big|_{\omega=0} = 0 \quad \forall n \in \mathbb{W}$$

PROOF:

<sup>8</sup> Jawerth and Sweldens (1994), pages 16–17, Sweldens and Piessens (1993), page 2, Vidakovic (1999), page 83

<sup>9</sup> Goswami and Chan (1999), pages 38–39

<sup>10</sup> Vidakovic (1999), pages 82–83, Mallat (1999), pages 241–242

1. Proof for ( $\implies$ ) case:

$$\begin{aligned} 0 &= \langle f(x) | x^n \rangle && \text{by left hypothesis} \\ &= \sqrt{2\pi}(-i)^{-n} \left[ \frac{d}{d\omega} \right]^n \tilde{f}(\omega) \Big|_{\omega=0} && \text{by Lemma P.1 page 342} \\ &\implies \left[ \frac{d}{d\omega} \right]^n \tilde{f}(\omega) \Big|_{\omega=0} = 0 \end{aligned}$$

2. Proof for ( $\Leftarrow$ ) case:

$$\begin{aligned} 0 &= \left[ \frac{d}{d\omega} \right]^n \tilde{f}(\omega) \Big|_{\omega=0} && \text{by right hypothesis} \\ &= \left[ \frac{d}{d\omega} \right]^n \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) e^{-i\omega x} dx \Big|_{\omega=0} && \text{by definition of } \tilde{f}(\omega) \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) \left[ \frac{d}{d\omega} \right]^n e^{-i\omega x} dx \Big|_{\omega=0} \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) [(-i)^n x^n e^{-i\omega x}] dx \Big|_{\omega=0} \\ &= (-i)^n \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) x^n dx \\ &= (-i)^n \frac{1}{\sqrt{2\pi}} \langle f(x) | x^n \rangle && \text{by definition of } \langle \cdot | \cdot \rangle \text{ in } L^2_{\mathbb{R}} \text{ (Definition ?? page ??)} \end{aligned}$$



**Lemma P.3** (Strang-Fix condition). <sup>11</sup> Let  $f(x)$  be a function in  $L^2_{\mathbb{R}}$  and  $M_n$  the  $n$ TH MOMENT (Definition P.4 page 342) off  $f(x)$ . Let  $T$  be the TRANSLATION OPERATOR (Definition ?? page ??).

<b>L</b> <b>E</b> <b>M</b>	$\sum_{k \in \mathbb{Z}} \underbrace{T^k x^n f(x)}_{\text{STRANG-FIX CONDITION in "time"}} = M_n \iff \underbrace{\left[ \frac{d}{d\omega} \right]^n \tilde{f}(\omega) \Big _{\omega=2\pi k} = \frac{1}{\sqrt{2\pi}} (-i)^n \bar{\delta}_k M_n}_{\text{STRANG-FIX CONDITION in "frequency"}}$
----------------------------------	---

PROOF:

1. Proof for ( $\implies$ ) case:

$$\begin{aligned} \left[ \left[ \frac{d}{d\omega} \right]^n \tilde{f}(\omega) \right]_{\omega=2\pi k} &= \sum_{k \in \mathbb{Z}} \left[ \left[ \frac{d}{d\omega} \right]^n \tilde{f}(\omega) \right]_{\omega=2\pi k} e^{i2\pi kx} \bar{\delta}_k \\ &= \frac{1}{\sqrt{2\pi}} \sum_{k \in \mathbb{Z}} \left[ \left[ \frac{d}{d\omega} \right]^n \int_{\mathbb{R}} f(x) e^{-i\omega x} dx \right]_{\omega=2\pi k} e^{i2\pi kx} \bar{\delta}_k && \text{by Definition P.2 page 337} \\ &= \frac{1}{\sqrt{2\pi}} \sum_{k \in \mathbb{Z}} \left[ \int_{\mathbb{R}} f(x) (-ix)^n e^{-i\omega x} dx \right]_{\omega=2\pi k} e^{i2\pi kx} \bar{\delta}_k \\ &= (-i)^n \frac{1}{\sqrt{2\pi}} \sum_{k \in \mathbb{Z}} \left[ \int_{\mathbb{R}} x^n f(x) e^{-i\omega x} dx \right]_{\omega=2\pi k} e^{i2\pi kx} \bar{\delta}_k \\ &= (-i)^n \frac{1}{\sqrt{2\pi}} \sum_{k \in \mathbb{Z}} (x - k)^n f(x - k) \bar{\delta}_k && \text{by PSF (Theorem ?? page ??)} \\ &= \frac{1}{\sqrt{2\pi}} (-i)^n \bar{\delta}_k M_n && \text{by left hypothesis} \end{aligned}$$

<sup>11</sup> 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)

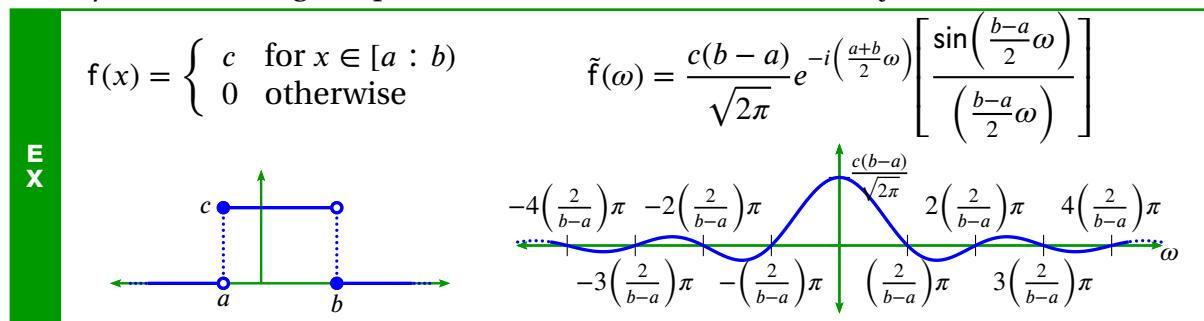
2. Proof for ( $\Leftarrow$ ) case:

$$\begin{aligned}
 \frac{1}{\sqrt{2\pi}}(-i)^n \mathbf{M}_n &= \frac{1}{\sqrt{2\pi}} \sum_{k \in \mathbb{Z}} [(-i)^n \bar{\delta}_k \mathbf{M}_n] e^{-i2\pi kx} && \text{by definition of } \bar{\delta} \quad (\text{Definition K.3 page 271}) \\
 &= \sum_{k \in \mathbb{Z}} \left[ \left[ \frac{d}{d\omega} \right]^n \tilde{f}(\omega) \right] \Big|_{\omega=2\pi k} e^{-i2\pi kx} && \text{by right hypothesis} \\
 &= \sum_{k \in \mathbb{Z}} \left[ \left[ \frac{d}{d\omega} \right]^n \int_{\mathbb{R}} f(x) e^{-i\omega x} dx \right] \Big|_{\omega=2\pi k} e^{-i2\pi kx} \\
 &= \sum_{k \in \mathbb{Z}} \left[ \int_{\mathbb{R}} f(x) (-ix)^n e^{-i\omega x} dx \right] \Big|_{\omega=2\pi k} e^{-i2\pi kx} \\
 &= (-i)^n \sum_{k \in \mathbb{Z}} \left[ \int_{\mathbb{R}} x^n f(x) e^{-i\omega x} dx \right] \Big|_{\omega=2\pi k} e^{-i2\pi kx} \\
 &= (-i)^n \sum_{k \in \mathbb{Z}} (x - k)^n f(x - k) && \text{by PSF} \quad (\text{Theorem ?? page ??})
 \end{aligned}$$

⇒

## P.6 Examples

*Example P.1* (rectangular pulse). Let  $\tilde{f}(\omega)$  be the Fourier transform of a function  $f(x) \in L^2_{\mathbb{R}}$ .



PROOF:

$$\begin{aligned}
 \tilde{f}(\omega) &= \tilde{\mathbf{F}}[f(x)](\omega) && \text{by definition of } \tilde{f}(\omega) \\
 &= e^{-i\left(\frac{a+b}{2}\right)\omega} \tilde{\mathbf{F}}\left[f\left(x - \frac{a+b}{2}\right)\right](\omega) && \text{by shift relation} \quad (\text{Theorem P.4 page 339}) \\
 &= e^{-i\left(\frac{a+b}{2}\right)\omega} \tilde{\mathbf{F}}\left[c \mathbb{1}_{[a:b]}\left(x - \frac{a+b}{2}\right)\right](\omega) && \text{by definition of } f(x) \\
 &= e^{-i\left(\frac{a+b}{2}\right)\omega} \tilde{\mathbf{F}}\left[c \mathbb{1}_{\left[-\frac{b-a}{2} : \frac{b-a}{2}\right]}(x)\right](\omega) && \text{by definition of } \mathbb{1} \quad (\text{Definition ?? page ??}) \\
 &= \frac{1}{\sqrt{2\pi}} e^{-i\left(\frac{a+b}{2}\right)\omega} \int_{\mathbb{R}} c \mathbb{1}_{\left[-\frac{b-a}{2} : \frac{b-a}{2}\right]}(x) e^{-i\omega x} dx && \text{by definition of } \tilde{\mathbf{F}} \quad (\text{Definition P.2 page 337}) \\
 &= \frac{1}{\sqrt{2\pi}} e^{-i\left(\frac{a+b}{2}\right)\omega} \int_{-\frac{b-a}{2}}^{\frac{b-a}{2}} c e^{-i\omega x} dx && \text{by definition of } \mathbb{1} \quad (\text{Definition ?? page ??}) \\
 &= \frac{c}{\sqrt{2\pi}} e^{-i\left(\frac{a+b}{2}\right)\omega} \frac{1}{-i\omega} e^{-i\omega x} \Big|_{-\frac{b-a}{2}}^{\frac{b-a}{2}} \\
 &= \frac{2c}{\sqrt{2\pi}\omega} e^{-i\left(\frac{a+b}{2}\right)\omega} \left[ \frac{e^{i\left(\frac{b-a}{2}\omega\right)} - e^{-i\left(\frac{b-a}{2}\omega\right)}}{2i} \right]
 \end{aligned}$$



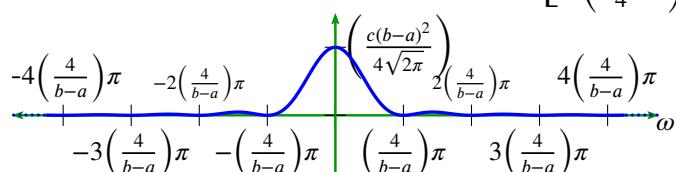
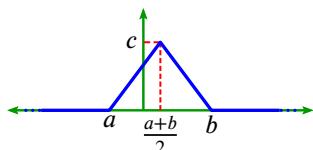
$$= \frac{c(b-a)}{\sqrt{2\pi}} e^{-i(\frac{a+b}{2}\omega)} \left[ \frac{\sin\left(\frac{b-a}{2}\omega\right)}{\left(\frac{b-a}{2}\omega\right)} \right]$$

by Euler formulas

(Corollary ?? page ??)

*Example P.2 (triangle).* Let  $\tilde{f}(\omega)$  be the Fourier transform of a function  $f(x) \in L^2_{\mathbb{R}}$ .

$f(x) = \begin{cases} c \left[ 1 - \frac{ 2x-b-a }{b-a} \right] & \text{for } x \in [a : b) \\ 0 & \text{otherwise} \end{cases}$	$\tilde{f}(\omega) = \frac{c(b-a)^2}{4\sqrt{2\pi}} e^{-i(\frac{a+b}{2}\omega)} \left[ \frac{\sin\left(\frac{b-a}{4}\omega\right)}{\left(\frac{b-a}{4}\omega\right)} \right]^2$
--	--

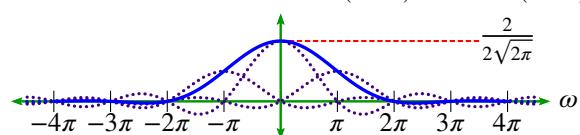
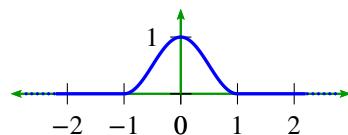
**E  
X**

PROOF:

$$\begin{aligned}
 \tilde{f}(\omega) &= \tilde{F}[f(x)](\omega) && \text{by definition of } \tilde{f}(\omega) \\
 &= e^{-i(\frac{a+b}{2})\omega} \tilde{F}\left[f\left(x - \frac{a+b}{2}\right)\right](\omega) && \text{by shift relation} \quad (\text{Theorem P.4 page 339}) \\
 &= \tilde{F}\left[c\left(1 - \frac{|2x-b-a|}{b-a}\right) \mathbb{1}_{[a:b)}(x)\right](\omega) && \text{by definition of } f(x) \\
 &= c \tilde{F}\left[\mathbb{1}_{[\frac{a}{2}:\frac{b}{2}]}(x) \star \mathbb{1}_{[\frac{a}{2}:\frac{b}{2}]}(x)\right](\omega) \\
 &= c \sqrt{2\pi} \tilde{F}\left[\mathbb{1}_{[\frac{a}{2}:\frac{b}{2}]}\right] \tilde{F}\left[\mathbb{1}_{[\frac{a}{2}:\frac{b}{2}]}\right] && \text{by convolution theorem} \quad (\text{Theorem R.2 page 360}) \\
 &= c \sqrt{2\pi} \left( \tilde{F}\left[\mathbb{1}_{[\frac{a}{2}:\frac{b}{2}]}\right] \right)^2 \\
 &= c \sqrt{2\pi} \left( \frac{\left(\frac{b}{2} - \frac{a}{2}\right)}{\sqrt{2\pi}} e^{-i(\frac{a+b}{4}\omega)} \left[ \frac{\sin\left(\frac{b-a}{4}\omega\right)}{\left(\frac{b-a}{4}\omega\right)} \right] \right)^2 && \text{by Rectangular pulse ex.} \quad \text{Example P.1 page 344} \\
 &= \frac{c(b-a)^2}{4\sqrt{2\pi}} e^{-i(\frac{a+b}{2}\omega)} \left[ \frac{\sin\left(\frac{b-a}{4}\omega\right)}{\left(\frac{b-a}{4}\omega\right)} \right]^2
 \end{aligned}$$

*Example P.3.* Let a function  $f$  be defined in terms of the cosine function (Definition ?? page ??) as follows:

$f(x) \triangleq \begin{cases} \cos^2\left(\frac{\pi}{2}x\right) & \text{for }  x  \leq 1 \\ 0 & \text{otherwise} \end{cases}$	$\tilde{f}(\omega) = \frac{1}{2\sqrt{2\pi}} \left[ \underbrace{\frac{2\sin\omega}{\omega}}_{2\operatorname{sinc}(\omega)} + \underbrace{\frac{\sin(\omega+\pi)}{(\omega+\pi)}}_{\operatorname{sinc}(\omega+\pi)} + \underbrace{\frac{\sin(\omega-\pi)}{(\omega-\pi)}}_{\operatorname{sinc}(\omega-\pi)} \right]$
--	--

**E  
X**

PROOF: Let  $\mathbb{1}_A(x)$  be the *set indicator function* (Definition ?? page ??) on a set  $A$ .

$$\begin{aligned}
 \tilde{f}(\omega) &\triangleq \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x)e^{-i\omega x} dx && \text{by definition of } \tilde{f}(\omega) \text{ (Definition P.2)} \\
 &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \cos^2\left(\frac{\pi}{2}x\right) \mathbb{1}_{[-1:1]}(x) e^{-i\omega x} dx && \text{by definition of } f(x) \\
 &= \frac{1}{\sqrt{2\pi}} \int_{-1}^1 \cos^2\left(\frac{\pi}{2}x\right) e^{-i\omega x} dx && \text{by definition of } \mathbb{1} \text{ (Definition ??)} \\
 &= \frac{1}{\sqrt{2\pi}} \int_{-1}^1 \left[ \frac{e^{i\frac{\pi}{2}x} + e^{-i\frac{\pi}{2}x}}{2} \right]^2 e^{-i\omega x} dx && \text{by Corollary ?? page ??} \\
 &= \frac{1}{4\sqrt{2\pi}} \int_{-1}^1 [2 + e^{i\pi x} + e^{-i\pi x}] e^{-i\omega x} dx \\
 &= \frac{1}{4\sqrt{2\pi}} \int_{-1}^1 2e^{-i\omega x} + e^{-i(\omega+\pi)x} + e^{-i(\omega-\pi)x} dx \\
 &= \frac{1}{4\sqrt{2\pi}} \left[ 2 \frac{e^{-i\omega x}}{-i\omega} + \frac{e^{-i(\omega+\pi)x}}{-i(\omega+\pi)} + \frac{e^{-i(\omega-\pi)x}}{-i(\omega-\pi)} \right]_{-1}^1 \\
 &= \frac{1}{2\sqrt{2\pi}} \left[ 2 \frac{e^{-i\omega} - e^{+i\omega}}{-2i\omega} + \frac{e^{-i(\omega+\pi)} - e^{+i(\omega+\pi)}}{-2i(\omega+\pi)} + \frac{e^{-i(\omega-\pi)} - e^{+i(\omega-\pi)}}{-2i(\omega-\pi)} \right]_{-1}^1 \\
 &= \frac{1}{2\sqrt{2\pi}} \left[ \underbrace{\frac{2\sin\omega}{\omega}}_{2\text{sinc}(\omega)} + \underbrace{\frac{\sin(\omega+\pi)}{(\omega+\pi)}}_{\text{sinc}(\omega+\pi)} + \underbrace{\frac{\sin(\omega-\pi)}{(\omega-\pi)}}_{\text{sinc}(\omega-\pi)} \right]
 \end{aligned}$$

⇒

# APPENDIX Q

## DISCRETE TIME FOURIER TRANSFORM

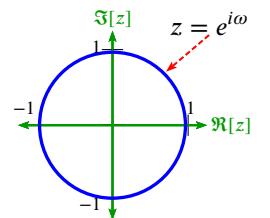
### Q.1 Definition

#### Definition Q.1.

**D E F** The **discrete-time Fourier transform**  $\check{F}$  of  $(x_n)_{n \in \mathbb{Z}}$  is defined as

$$[\check{F}(x_n)](\omega) \triangleq \sum_{n \in \mathbb{Z}} x_n e^{-i\omega n} \quad \forall (x_n)_{n \in \mathbb{Z}} \in \ell^2_{\mathbb{R}}$$

If we compare the definition of the *Discrete Time Fourier Transform* (Definition Q.1 page 347) to the definition of the Z-transform (Definition R.4 page 358), 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”  $\omega$  in the DTFT is the unit circle in the much larger z-plane, as illustrated to the right.



### Q.2 Properties

**Proposition Q.1** (DTFT periodicity). Let  $\check{x}(\omega) \triangleq \check{F}[(x_n)](\omega)$  be the DISCRETE-TIME FOURIER TRANSFORM (Definition Q.1 page 347) of a sequence  $(x_n)_{n \in \mathbb{Z}}$  in  $\ell^2_{\mathbb{R}}$ .

**P R P**  $\underbrace{\check{x}(\omega) = \check{x}(\omega + 2\pi n)}_{\text{PERIODIC with period } 2\pi} \quad \forall n \in \mathbb{Z}$

PROOF:

$$\begin{aligned} \check{x}(\omega + 2\pi n) &= \sum_{m \in \mathbb{Z}} x_m e^{-i(\omega+2\pi n)m} \\ &= \sum_{m \in \mathbb{Z}} x_m e^{-i\omega m} e^{-i2\pi nm} \\ &= \check{x}(\omega) \end{aligned}$$

**Theorem Q.1.** Let  $\tilde{x}(\omega) \triangleq \check{F}[(x[n])](\omega)$  be the DISCRETE-TIME FOURIER TRANSFORM (Definition Q.1 page 347) of a sequence  $(x_n)_{n \in \mathbb{Z}}$  in  $\ell^2_{\mathbb{R}}$ .

<b>T H M</b>	$\left\{ \begin{array}{l} \tilde{x}(\omega) \triangleq \check{F}(x[n]) \\ \end{array} \right\} \implies \left\{ \begin{array}{l} (1). \quad \check{F}(x[-n]) = \tilde{x}(-\omega) \text{ and} \\ (2). \quad \check{F}(x^*[n]) = \tilde{x}^*(-\omega) \text{ and} \\ (3). \quad \check{F}(x^*[-n]) = \tilde{x}^*(\omega) \end{array} \right\}$
----------------------	---

PROOF:

$$\begin{aligned} \check{F}(x[-n]) &\triangleq \sum_{n \in \mathbb{Z}} x[-n]e^{-i\omega n} && \text{by definition of DTFT} && (\text{Definition Q.1 page 347}) \\ &= \sum_{m \in \mathbb{Z}} x[m]e^{i\omega m} && \text{where } m \triangleq -n \implies n = -m \\ &= \sum_{m \in \mathbb{Z}} x[m]e^{-i(-\omega)m} \\ &\triangleq \tilde{x}(-\omega) && \text{by left hypothesis} \end{aligned}$$

$$\begin{aligned} \check{F}(x^*[n]) &\triangleq \sum_{n \in \mathbb{Z}} x^*[n]e^{-i\omega n} && \text{by definition of DTFT} && (\text{Definition Q.1 page 347}) \\ &= \left( \sum_{n \in \mathbb{Z}} x[n]e^{i\omega n} \right)^* && \text{by distributive property of } *-\text{algebras} && (\text{Definition J.3 page 254}) \\ &= \left( \sum_{n \in \mathbb{Z}} x[n]e^{-i(-\omega)n} \right)^* \\ &\triangleq \tilde{x}^*(-\omega) && \text{by left hypothesis} \end{aligned}$$

$$\begin{aligned} \check{F}(x^*[-n]) &\triangleq \sum_{n \in \mathbb{Z}} x^*[-n]e^{-i\omega n} && \text{by definition of DTFT} && (\text{Definition Q.1 page 347}) \\ &= \left( \sum_{n \in \mathbb{Z}} x[-n]e^{i\omega n} \right)^* && \text{by distributive property of } *-\text{algebras} && (\text{Definition J.3 page 254}) \\ &= \left( \sum_{m \in \mathbb{Z}} x[m]e^{-i\omega m} \right)^* \\ &\triangleq \tilde{x}^*(\omega) && \text{by left hypothesis} \end{aligned}$$

⇒

**Theorem Q.2.** Let  $\tilde{x}(\omega) \triangleq \check{F}[(x[n])](\omega)$  be the DISCRETE-TIME FOURIER TRANSFORM (Definition Q.1 page 347) of a sequence  $(x[n])_{n \in \mathbb{Z}}$  in  $\ell^2_{\mathbb{R}}$ .

<b>T H M</b>	$\left\{ \begin{array}{l} (1). \quad \tilde{x}(\omega) \triangleq \check{F}(x[n]) \\ (2). \quad (x[n]) \text{ is REAL-VALUED} \end{array} \right\} \implies \left\{ \begin{array}{l} (1). \quad \check{F}(x[-n]) = \tilde{x}(-\omega) \text{ and} \\ (2). \quad \check{F}(x^*[n]) = \tilde{x}^*(-\omega) = \tilde{x}(\omega) \text{ and} \\ (3). \quad \check{F}(x^*[-n]) = \tilde{x}^*(\omega) = \tilde{x}(-\omega) \end{array} \right\}$
----------------------	--

PROOF:

$$\begin{aligned} \check{F}(x[-n]) &\triangleq \sum_{n \in \mathbb{Z}} x[-n]e^{-i\omega n} && \text{by definition of DTFT} && (\text{Definition Q.1 page 347}) \\ &= \sum_{m \in \mathbb{Z}} x[m]e^{i\omega m} && \text{where } m \triangleq -n \implies n = -m \\ &= \sum_{m \in \mathbb{Z}} x[m]e^{-i(-\omega)m} \end{aligned}$$

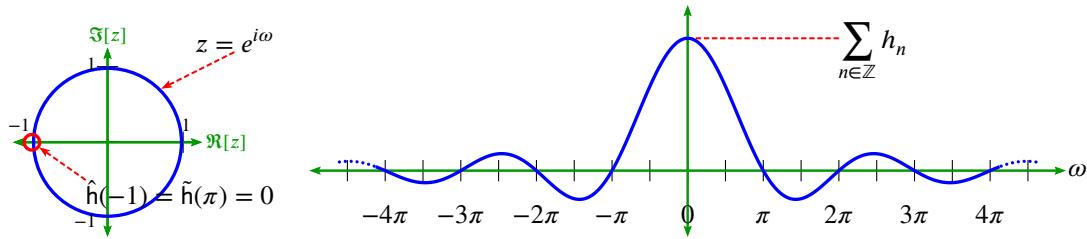


$$\triangleq \tilde{x}(-\omega) \quad \text{by left hypothesis}$$

$$\begin{aligned} \tilde{x}^*(-\omega) &= \check{\mathbf{F}}(\mathbf{x}^*[n]) && \text{by Theorem Q.1 page 348} \\ &= \check{\mathbf{F}}(\mathbf{x}[n]) && \text{by real-valued hypothesis} \\ &= [\tilde{x}(\omega)] && \text{by definition of } \tilde{x}(\omega) \quad (\text{Definition Q.1 page 347}) \end{aligned}$$

$$\begin{aligned} \tilde{x}^*(\omega) &= \check{\mathbf{F}}(\mathbf{x}^*[-n]) && \text{by Theorem Q.1 page 348} \\ &= \check{\mathbf{F}}(\mathbf{x}[-n]) && \text{by real-valued hypothesis} \\ &= [\tilde{x}(-\omega)] && \text{by result (1)} \end{aligned}$$

⇒



**Proposition Q.2.** Let  $\check{x}(z)$  be the Z-TRANSFORM (Definition R.4 page 358) and  $\check{x}(\omega)$  the DISCRETE-TIME FOURIER TRANSFORM (Definition Q.1 page 347) of  $(x_n)$ .

<b>P</b> <b>R</b> <b>P</b>	$\underbrace{\left\{ \sum_{n \in \mathbb{Z}} x_n = c \right\}}_{(1) \text{ time domain}} \iff \underbrace{\left\{ \check{x}(z) \Big _{z=1} = c \right\}}_{(2) z \text{ domain}} \iff \underbrace{\left\{ \check{x}(\omega) \Big _{\omega=0} = c \right\}}_{(3) \text{ frequency domain}} \quad \forall (x_n)_{n \in \mathbb{Z}} \in \ell^2_{\mathbb{R}}, c \in \mathbb{R}$
----------------------------------	--

PROOF:

1. Proof that (1)  $\implies$  (2):

$$\begin{aligned} \check{x}(z) \Big|_{z=1} &= \sum_{n \in \mathbb{Z}} x_n z^{-n} \Big|_{z=1} && \text{by definition of } \check{x}(z) \text{ (Definition R.4 page 358)} \\ &= \sum_{n \in \mathbb{Z}} x_n && \text{because } z^n = 1 \text{ for all } n \in \mathbb{Z} \\ &= c && \text{by hypothesis (1)} \end{aligned}$$

2. Proof that (2)  $\implies$  (3):

$$\begin{aligned} \check{x}(\omega) \Big|_{\omega=0} &= \sum_{n \in \mathbb{Z}} x_n e^{-i\omega n} \Big|_{\omega=0} && \text{by definition of } \check{x}(\omega) \quad (\text{Definition Q.1 page 347}) \\ &= \sum_{n \in \mathbb{Z}} x_n z^{-n} \Big|_{z=1} \\ &= \check{x}(z) \Big|_{z=1} && \text{by definition of } \check{x}(z) \quad (\text{Definition R.4 page 358}) \\ &= c && \text{by hypothesis (2)} \end{aligned}$$

3. Proof that (3)  $\implies$  (1):

$$\begin{aligned} \sum_{n \in \mathbb{Z}} x_n &= \sum_{n \in \mathbb{Z}} x_n e^{-i\omega n} \Big|_{\omega=0} \\ &= \check{x}(\omega) && \text{by definition of } \check{x}(\omega) && \text{(Definition Q.1 page 347)} \\ &= c && \text{by hypothesis (3)} \end{aligned}$$



**Proposition Q.3.** *If the coefficients are real, then the magnitude response (MR) is symmetric.*

PROOF:

$$\begin{aligned} |\tilde{h}(-\omega)| &\triangleq |\check{h}(z)|_{z=e^{-i\omega}} \\ &= \left| \sum_{m \in \mathbb{Z}} x[m] e^{i\omega m} \right| \\ &= \left| \underbrace{\left( \sum_{m \in \mathbb{Z}} x[m] e^{-i\omega m} \right)^*}_{\text{if } x[m] \text{ is real}} \right| \\ &\triangleq |\check{h}(z)|_{z=e^{-i\omega}} && \triangleq |\tilde{h}(\omega)| \end{aligned}$$

$$\begin{aligned} &\triangleq \left| \sum_{m \in \mathbb{Z}} x[m] z^{-m} \right|_{z=e^{-i\omega}} \\ &= \left| \left( \sum_{m \in \mathbb{Z}} x^*[m] e^{-i\omega m} \right)^* \right| \\ &= \left| \sum_{m \in \mathbb{Z}} x[m] e^{-i\omega m} \right| \end{aligned}$$



**Proposition Q.4.**<sup>1</sup>

P  
R  
P

$$\begin{aligned} \underbrace{\sum_{n \in \mathbb{Z}} (-1)^n x_n = c}_{(1) \text{ in "time"}} &\iff \underbrace{\check{x}(z)|_{z=-1} = c}_{(2) \text{ in "z domain"}} \iff \underbrace{\check{x}(\omega)|_{\omega=\pi} = c}_{(3) \text{ in "frequency"}} \\ &\iff \underbrace{\left( \sum_{n \in \mathbb{Z}} h_{2n}, \sum_{n \in \mathbb{Z}} h_{2n+1} \right)}_{(4) \text{ sum of even, sum of odd}} = \left( \frac{1}{2} \left( \sum_{n \in \mathbb{Z}} h_n + c \right), \frac{1}{2} \left( \sum_{n \in \mathbb{Z}} h_n - c \right) \right) \\ &\forall c \in \mathbb{R}, (x_n)_{n \in \mathbb{Z}}, (y_n)_{n \in \mathbb{Z}} \in \ell^2_{\mathbb{R}} \end{aligned}$$

PROOF:

1. Proof that (1)  $\implies$  (2):

$$\begin{aligned} \check{x}(z)|_{z=-1} &= \sum_{n \in \mathbb{Z}} x_n z^{-n} \Big|_{z=-1} \\ &= \sum_{n \in \mathbb{Z}} (-1)^n x_n \\ &= c && \text{by (1)} \end{aligned}$$

<sup>1</sup> Chui (1992) page 123

2. Proof that (2)  $\implies$  (3):

$$\begin{aligned} \sum_{n \in \mathbb{Z}} x_n e^{-i\omega n} \Big|_{\omega=\pi} &= \sum_{n \in \mathbb{Z}} (-1)^n x_n \\ &= \sum_{n \in \mathbb{Z}} (-1)^{-n} x_n && = \sum_{n \in \mathbb{Z}} z^{-n} x_n \Big|_{z=-1} \\ &= c && \text{by (2)} \end{aligned}$$

3. Proof that (3)  $\implies$  (1):

$$\begin{aligned} \sum_{n \in \mathbb{Z}} (-1)^n x_n &= \sum_{n \in \mathbb{Z}} (-1)^{-n} x_n \\ &= \sum_{n \in \mathbb{Z}} e^{-i\omega n} x_n \Big|_{\omega=\pi} \\ &= c && \text{by (3)} \end{aligned}$$

4. Proof that (2)  $\implies$  (4):

$$(a) \text{ Define } A \triangleq \sum_{n \in \mathbb{Z}} h_{2n} \quad B \triangleq \sum_{n \in \mathbb{Z}} h_{2n+1}.$$

(b) Proof that  $A - B = c$ :

$$\begin{aligned} c &= \sum_{n \in \mathbb{Z}} (-1)^n x_n && \text{by (2)} \\ &= \underbrace{\sum_{n \in \mathbb{Z}_e} (-1)^n x_n}_{\text{even terms}} + \underbrace{\sum_{n \in \mathbb{Z}_o} (-1)^n x_n}_{\text{odd terms}} \\ &= \sum_{n \in \mathbb{Z}} (-1)^{2n} x_{2n} + \sum_{n \in \mathbb{Z}} (-1)^{2n+1} x_{2n+1} \\ &= \underbrace{\sum_{n \in \mathbb{Z}} x_{2n}}_A - \underbrace{\sum_{n \in \mathbb{Z}} x_{2n+1}}_B \\ &\triangleq A - B && \text{by definitions of } A \text{ and } B \end{aligned}$$

$$(c) \text{ Proof that } A + B = \sum_{n \in \mathbb{Z}} x_n:$$

$$\begin{aligned} \sum_{n \in \mathbb{Z}} x_n &= \sum_{n \text{ even}} x_n + \sum_{n \text{ odd}} x_n \\ &= \underbrace{\sum_{n \in \mathbb{Z}} x_{2n}}_A + \underbrace{\sum_{n \in \mathbb{Z}} x_{2n+1}}_B \\ &= A + B && \text{by definitions of } A \text{ and } B \end{aligned}$$

(d) This gives two simultaneous equations:

$$A - B = c$$

$$A + B = \sum_{n \in \mathbb{Z}} x_n$$

(e) Solutions to these equations give

$$\begin{aligned}\sum_{n \in \mathbb{Z}} x_{2n} &\triangleq A \\ \sum_{n \in \mathbb{Z}} x_{2n+1} &\triangleq B\end{aligned}\begin{aligned}&= \frac{1}{2} \left( \sum_{n \in \mathbb{Z}} x_n + c \right) \\ &= \frac{1}{2} \left( \sum_{n \in \mathbb{Z}} x_n - c \right)\end{aligned}$$

5. Proof that (2)  $\Leftarrow$  (4):

$$\begin{aligned}\sum_{n \in \mathbb{Z}} (-1)^n x_n &= \underbrace{\sum_{n \in \mathbb{Z}_e} (-1)^n x_n}_{\text{even terms}} + \underbrace{\sum_{n \in \mathbb{Z}_o} (-1)^n x_n}_{\text{odd terms}} \\ &= \sum_{n \in \mathbb{Z}} (-1)^{2n} x_{2n} + \sum_{n \in \mathbb{Z}} (-1)^{2n+1} x_{2n+1} \\ &= \sum_{n \in \mathbb{Z}} x_{2n} - \sum_{n \in \mathbb{Z}} x_{2n+1} \\ &= \frac{1}{2} \left( \sum_{n \in \mathbb{Z}} x_n + c \right) - \frac{1}{2} \left( \sum_{n \in \mathbb{Z}} x_n - c \right) \quad \text{by (3)} \\ &= c\end{aligned}$$

$\Rightarrow$

**Lemma Q.1.** Let  $\tilde{f}(\omega)$  be the DTFT (Definition Q.1 page 347) of a sequence  $(x_n)_{n \in \mathbb{Z}}$ .

<b>L E M</b>	$\underbrace{(x_n \in \mathbb{R})_{n \in \mathbb{Z}}}_{\text{REAL-VALUED sequence}}$	$\Rightarrow$	$\underbrace{ \check{x}(\omega) ^2 =  \check{x}(-\omega) ^2}_{\text{EVEN}}$	$\forall (x_n)_{n \in \mathbb{Z}} \in \ell^2_{\mathbb{R}}$
----------------------	--	---------------	---	--

PROOF:

$$\begin{aligned}|\check{x}(\omega)|^2 &= |\check{x}(z)|^2 \Big|_{z=e^{i\omega}} \\ &= \check{x}(z)\check{x}^*(z) \Big|_{z=e^{i\omega}} \\ &= \left[ \sum_{n \in \mathbb{Z}} x_n z^{-n} \right] \left[ \sum_{m \in \mathbb{Z}} x_m z^{-m} \right]^* \Big|_{z=e^{i\omega}} \\ &= \left[ \sum_{n \in \mathbb{Z}} x_n z^{-n} \right] \left[ \sum_{m \in \mathbb{Z}} x_m^* (z^*)^{-m} \right] \Big|_{z=e^{i\omega}} \\ &= \sum_{n \in \mathbb{Z}} \sum_{m \in \mathbb{Z}} x_n x_m^* z^{-n} (z^*)^{-m} \Big|_{z=e^{i\omega}} \\ &= \sum_{n \in \mathbb{Z}} \left[ |x_n|^2 + \sum_{m>n} x_n x_m^* z^{-n} (z^*)^{-m} + \sum_{m<n} x_n x_m^* z^{-n} (z^*)^{-m} \right] \Big|_{z=e^{i\omega}} \\ &= \sum_{n \in \mathbb{Z}} \left[ |x_n|^2 + \sum_{m>n} x_n x_m e^{i\omega(m-n)} + \sum_{m<n} x_n x_m^* e^{i\omega(m-n)} \right] \\ &= \sum_{n \in \mathbb{Z}} \left[ |x_n|^2 + \sum_{m>n} x_n x_m e^{i\omega(m-n)} + \sum_{m>n} x_n x_m e^{-i\omega(m-n)} \right] \\ &= \sum_{n \in \mathbb{Z}} \left[ |x_n|^2 + \sum_{m>n} x_n x_m (e^{i\omega(m-n)} + e^{-i\omega(m-n)}) \right]\end{aligned}$$

$$\begin{aligned}
 &= \sum_{n \in \mathbb{Z}} \left[ |x_n|^2 + \sum_{m>n} x_n x_m 2\cos[\omega(m-n)] \right] \\
 &= \sum_{n \in \mathbb{Z}} |x_n|^2 + 2 \sum_{n \in \mathbb{Z}} \sum_{m>n} x_n x_m \cos[\omega(m-n)]
 \end{aligned}$$

Since  $\cos$  is real and even, then  $|\check{x}(\omega)|^2$  must also be real and even.  $\Rightarrow$

**Theorem Q.3** (inverse DTFT). <sup>2</sup> Let  $\check{x}(\omega)$  be the DISCRETE-TIME FOURIER TRANSFORM (Definition Q.1 page 347) of a sequence  $(x_n)_{n \in \mathbb{Z}} \in \ell^2_{\mathbb{R}}$ . Let  $\check{x}^{-1}$  be the inverse of  $\check{x}$ .

T H M	$\underbrace{\left\{ \check{x}(\omega) \triangleq \sum_{n \in \mathbb{Z}} x_n e^{-i\omega n} \right\}}_{\check{x}(\omega) \triangleq \check{F}(x_n)} \quad \Rightarrow \quad \underbrace{\left\{ x_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} \check{x}(\omega) e^{i\omega n} d\omega \quad \forall \omega \in \mathbb{R} \right\}}_{(x_n) = \check{F}^{-1}(\check{x}(\omega))} \quad \forall (x_n)_{n \in \mathbb{Z}} \in \ell^2_{\mathbb{R}}$
-------------	---

$\Leftarrow$  PROOF:

$$\begin{aligned}
 \frac{1}{2\pi} \int_{-\pi}^{\pi} \check{x}(\omega) e^{i\omega n} d\omega &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \underbrace{\left[ \sum_{m \in \mathbb{Z}} x_m e^{-i\omega m} \right]}_{\check{x}(\omega)} e^{i\omega n} d\omega && \text{by definition of } \check{x}(\omega) \\
 &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{m \in \mathbb{Z}} x_m e^{-i\omega(m-n)} d\omega \\
 &= \frac{1}{2\pi} \sum_{m \in \mathbb{Z}} x_m \int_{-\pi}^{\pi} e^{-i\omega(m-n)} d\omega \\
 &= \frac{1}{2\pi} \sum_{m \in \mathbb{Z}} x_m [2\pi \delta_{m-n}] \\
 &= x_n
 \end{aligned}$$

**Theorem Q.4** (orthonormal quadrature conditions). <sup>3</sup> Let  $\check{x}(\omega)$  be the DISCRETE-TIME FOURIER TRANSFORM (Definition Q.1 page 347) 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 K.3 page 271).

T H M	$  \begin{aligned}  \sum_{m \in \mathbb{Z}} x_m y_{m-2n}^* &= 0 \iff \check{x}(\omega) \check{y}^*(\omega) + \check{x}(\omega + \pi) \check{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  \check{x}(\omega) ^2 +  \check{x}(\omega + \pi) ^2 = 2 && \forall n \in \mathbb{Z}, \forall (x_n), (y_n) \in \ell^2_{\mathbb{R}}  \end{aligned}  $
-------------	--

$\Leftarrow$  PROOF: Let  $z \triangleq e^{j\omega}$ .

<sup>2</sup>  J.S.Chitode (2009a) page 3-95 ((3.6.2))

<sup>3</sup>  Daubechies (1992) pages 132–137 ((5.1.20),(5.1.39))

1. Proof that  $2 \sum_{n \in \mathbb{Z}} \left[ \sum_{k \in \mathbb{Z}} x_k y_{k-2n}^* \right] e^{-i2\omega n} = \check{x}(\omega) \check{y}^*(\omega) + \check{x}(\omega + \pi) \check{y}^*(\omega + \pi)$ :

$$\begin{aligned}
 & 2 \sum_{n \in \mathbb{Z}} \left[ \sum_{k \in \mathbb{Z}} x_k y_{k-2n}^* \right] e^{-i2\omega n} \\
 &= 2 \sum_{k \in \mathbb{Z}} x_k \sum_{n \in \mathbb{Z}} y_{k-2n}^* z^{-2n} \\
 &= 2 \sum_{k \in \mathbb{Z}} x_k \sum_{n \text{ even}} y_{k-n}^* z^{-n} \\
 &= \sum_{k \in \mathbb{Z}} x_k \sum_{n \in \mathbb{Z}} y_{k-n}^* z^{-n} (1 + e^{i\pi n}) \\
 &= \sum_{k \in \mathbb{Z}} x_k \sum_{n \in \mathbb{Z}} y_{k-n}^* z^{-n} + \sum_{k \in \mathbb{Z}} x_k \sum_{n \in \mathbb{Z}} y_{k-n}^* z^{-n} e^{i\pi n} \\
 &= \sum_{k \in \mathbb{Z}} x_k \sum_{m \in \mathbb{Z}} y_m^* z^{-(k-m)} + \sum_{k \in \mathbb{Z}} x_k \sum_{m \in \mathbb{Z}} y_m^* e^{-i(\omega+\pi)(k-m)} \quad \text{where } m \triangleq k - n \\
 &= \sum_{k \in \mathbb{Z}} x_k z^{-k} \sum_{m \in \mathbb{Z}} y_m^* z^m + \sum_{k \in \mathbb{Z}} x_k e^{-i(\omega+\pi)k} \sum_{m \in \mathbb{Z}} y_m^* e^{+i(\omega+\pi)m} \\
 &= \sum_{k \in \mathbb{Z}} x_k e^{-i\omega k} \left[ \sum_{m \in \mathbb{Z}} y_m^* e^{-i\omega m} \right]^* + \sum_{k \in \mathbb{Z}} x_k e^{-i(\omega+\pi)k} \left[ \sum_{m \in \mathbb{Z}} y_m^* e^{-i(\omega+\pi)m} \right]^* \\
 &\triangleq \check{x}(\omega) \check{y}^*(\omega) + \check{x}(\omega + \pi) \check{y}^*(\omega + \pi)
 \end{aligned}$$

2. Proof that  $\sum_{m \in \mathbb{Z}} x_m y_{m-2n}^* = 0 \implies \check{x}(\omega) \check{y}^*(\omega) + \check{x}(\omega + \pi) \check{y}^*(\omega + \pi) = 0$ :

$$\begin{aligned}
 0 &= 2 \sum_{n \in \mathbb{Z}} \left[ \sum_{k \in \mathbb{Z}} x_k y_{k-2n}^* \right] e^{-i2\omega n} && \text{by left hypothesis} \\
 &= \check{x}(\omega) \check{y}^*(\omega) + \check{x}(\omega + \pi) \check{y}^*(\omega + \pi) && \text{by item (1)}
 \end{aligned}$$

3. Proof that  $\sum_{m \in \mathbb{Z}} x_m y_{m-2n}^* = 0 \iff \check{x}(\omega) \check{y}^*(\omega) + \check{x}(\omega + \pi) \check{y}^*(\omega + \pi) = 0$ :

$$\begin{aligned}
 2 \sum_{n \in \mathbb{Z}} \left[ \sum_{k \in \mathbb{Z}} x_k y_{k-2n}^* \right] e^{-i2\omega n} &= \check{x}(\omega) \check{y}^*(\omega) + \check{x}(\omega + \pi) \check{y}^*(\omega + \pi) && \text{by item (1)} \\
 &= 0 && \text{by right hypothesis}
 \end{aligned}$$

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  $\sum_{m \in \mathbb{Z}} x_m x_{m-2n}^* = \bar{\delta}_n \implies |\check{x}(\omega)|^2 + |\check{x}(\omega + \pi)|^2 = 2$ :  
Let  $g_n \triangleq x_n$ .

$$\begin{aligned}
 2 &= 2 \sum_{n \in \mathbb{Z}} \bar{\delta}_n e^{-i2\omega n} \\
 &= 2 \sum_{n \in \mathbb{Z}} \left[ \sum_{k \in \mathbb{Z}} x_k y_{k-2n}^* \right] e^{-i2\omega n} && \text{by left hypothesis} \\
 &= \check{x}(\omega) \check{y}^*(\omega) + \check{x}(\omega + \pi) \check{y}^*(\omega + \pi) && \text{by item (1)}
 \end{aligned}$$

5. Proof that  $\sum_{m \in \mathbb{Z}} x_m x_{m-2n}^* = \bar{\delta}_n \iff |\check{x}(\omega)|^2 + |\check{x}(\omega + \pi)|^2 = 2$ :  
Let  $g_n \triangleq x_n$ .

$$\begin{aligned}
 2 \sum_{n \in \mathbb{Z}} \left[ \sum_{k \in \mathbb{Z}} x_k y_{k-2n}^* \right] e^{-i2\omega n} &= \check{x}(\omega) \check{y}^*(\omega) + \check{x}(\omega + \pi) \check{y}^*(\omega + \pi) && \text{by item (1)} \\
 &= 2 && \text{by right hypothesis}
 \end{aligned}$$



Thus by the above equation,  $\sum_{n \in \mathbb{Z}} [\sum_{k \in \mathbb{Z}} x_k y_{k-2n}^*] e^{-i2\omega n} = 1$ . The only way for this to be true is if  $\sum_{k \in \mathbb{Z}} x_k y_{k-2n}^* = \delta_n$ .



## Q.3 Derivatives

**Theorem Q.5.**<sup>4</sup> Let  $\check{x}(\omega)$  be the DTFT (Definition Q.1 page 347) of a sequence  $(x_n)_{n \in \mathbb{Z}}$ .

T H M	$(A) \quad \left[ \frac{d}{d\omega} \right]^n \check{x}(\omega) \Big _{\omega=0} = 0 \iff \sum_{k \in \mathbb{Z}} k^n x_k = 0 \quad (B) \quad \forall n \in \mathbb{W}$ $(C) \quad \left[ \frac{d}{d\omega} \right]^n \check{x}(\omega) \Big _{\omega=\pi} = 0 \iff \sum_{k \in \mathbb{Z}} (-1)^k k^n x_k = 0 \quad (D) \quad \forall n \in \mathbb{W}$
-------------	--

PROOF:

1. Proof that (A)  $\implies$  (B):

$$\begin{aligned}
 0 &= \left[ \frac{d}{d\omega} \right]^n \check{x}(\omega) \Big|_{\omega=0} && \text{by hypothesis (A)} \\
 &= \left[ \frac{d}{d\omega} \right]^n \sum_{k \in \mathbb{Z}} x_k e^{-i\omega k} \Big|_{\omega=0} && \text{by definition of } \check{x}(\omega) \text{ (Definition Q.1 page 347)} \\
 &= \sum_{k \in \mathbb{Z}} x_k \left[ \frac{d}{d\omega} \right]^n e^{-i\omega k} \Big|_{\omega=0} \\
 &= \sum_{k \in \mathbb{Z}} x_k \left[ (-i)^n k^n e^{-i\omega k} \right] \Big|_{\omega=0} \\
 &= (-i)^n \sum_{k \in \mathbb{Z}} k^n x_k
 \end{aligned}$$

2. Proof that (A)  $\iff$  (B):

$$\begin{aligned}
 \left[ \frac{d}{d\omega} \right]^n \check{x}(\omega) \Big|_{\omega=0} &= \left[ \frac{d}{d\omega} \right]^n \sum_{k \in \mathbb{Z}} x_k e^{-i\omega k} \Big|_{\omega=0} && \text{by definition of } \check{x}(\omega) \\
 &= \sum_{k \in \mathbb{Z}} x_k \left[ \left[ \frac{d}{d\omega} \right]^n e^{-i\omega k} \right] \Big|_{\omega=0} \\
 &= \sum_{k \in \mathbb{Z}} x_k \left[ (-i)^n k^n e^{-i\omega k} \right] \Big|_{\omega=0} \\
 &= (-i)^n \sum_{k \in \mathbb{Z}} k^n x_k \\
 &= 0 && \text{by hypothesis (B)}
 \end{aligned}$$

<sup>4</sup> Vidakovic (1999), pages 82–83, Mallat (1999), pages 241–242

3. Proof that  $(C) \implies (D)$ :

$$\begin{aligned}
 0 &= \left[ \frac{d}{d\omega} \right]^n \check{x}(\omega) \Big|_{\omega=\pi} && \text{by hypothesis (C)} \\
 &= \left[ \frac{d}{d\omega} \right]^n \sum_{k \in \mathbb{Z}} x_k e^{-i\omega k} \Big|_{\omega=\pi} && \text{by definition of } \check{x} \text{ (Definition Q.1 page 347)} \\
 &= \sum_{k \in \mathbb{Z}} x_k \left[ \frac{d}{d\omega} \right]^n e^{-i\omega k} \Big|_{\omega=\pi} \\
 &= \sum_{k \in \mathbb{Z}} x_k \left[ (-i)^n k^n e^{-i\omega k} \right] \Big|_{\omega=\pi} \\
 &= \sum_{k \in \mathbb{Z}} x_k \left[ (-i)^n k^n (-1)^k \right] \\
 &= (-i)^n \sum_{k \in \mathbb{Z}} (-1)^k k^n x_k
 \end{aligned}$$

4. Proof that  $(C) \iff (D)$ :

$$\begin{aligned}
 \left[ \frac{d}{d\omega} \right]^n \check{x}(\omega) \Big|_{\omega=\pi} &= \left[ \frac{d}{d\omega} \right]^n \sum_{k \in \mathbb{Z}} x_k e^{-i\omega k} \Big|_{\omega=\pi} && \text{by definition of } \check{x} \text{ (Definition Q.1 page 347)} \\
 &= \sum_{k \in \mathbb{Z}} x_k \left[ \frac{d}{d\omega} \right]^n e^{-i\omega k} \Big|_{\omega=\pi} \\
 &= \sum_{k \in \mathbb{Z}} x_k \left[ (-i)^n k^n e^{-i\omega k} \right] \Big|_{\omega=\pi} \\
 &= \sum_{k \in \mathbb{Z}} x_k \left[ (-i)^n k^n (-1)^k \right] \\
 &= (-i)^n \sum_{k \in \mathbb{Z}} (-1)^k k^n x_k \\
 &= 0 && \text{by hypothesis (D)}
 \end{aligned}$$



# APPENDIX R

## OPERATIONS ON SEQUENCES

### R.1 Convolution operator

**Definition R.1.** <sup>1</sup> 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.

**D E F** 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  $(x_n)$ .

**Definition R.2.** <sup>2</sup> Let  $(\mathbb{F}, +, \cdot, 0, 1)$  be a FIELD (Definition F.5 page 196).

The space of all absolutely square summable sequences  $\ell_{\mathbb{F}}^2$  over  $\mathbb{F}$  is defined as

$$\ell_{\mathbb{F}}^2 \triangleq \left\{ (x_n)_{n \in \mathbb{Z}} \mid \sum_{n \in \mathbb{Z}} |x_n|^2 < \infty \right\}$$

The space  $\ell_{\mathbb{R}}^2$  is an example of a *separable Hilbert space*. In fact,  $\ell_{\mathbb{R}}^2$  is the *only* separable Hilbert space in the sense that all separable Hilbert spaces are isomorphically equivalent. For example,  $\ell_{\mathbb{R}}^2$  is isomorphic to  $L_{\mathbb{R}}^2$ , the space of all absolutely square Lebesgue integrable functions.

**Definition R.3.**

The **convolution operation  $\star$**  is defined as

$$(x_n) \star (y_n) \triangleq \left( \left( \sum_{m \in \mathbb{Z}} x_m y_{n-m} \right) \right)_{n \in \mathbb{Z}} \quad \forall (x_n)_{n \in \mathbb{Z}}, (y_n)_{n \in \mathbb{Z}} \in \ell_{\mathbb{R}}^2$$

**Proposition R.1.** Let  $\star$  be the CONVOLUTION OPERATOR (Definition R.3 page 357).

**P R P**  $(x_n) \star (y_n) = (y_n) \star (x_n) \quad \forall (x_n)_{n \in \mathbb{Z}}, (y_n)_{n \in \mathbb{Z}} \in \ell_{\mathbb{R}}^2 \quad (\star \text{ is COMMUTATIVE})$

<sup>1</sup> Bromwich (1908), page 1, Thomson et al. (2008) page 23 (Definition 2.1), Joshi (1997) page 31

<sup>2</sup> Kubrusly (2011) page 347 (Example 5.K)

PROOF:

$$\begin{aligned}
 [x \star y](n) &\triangleq \sum_{m \in \mathbb{Z}} x_m y_{n-m} && \text{by Definition R.3 page 357} \\
 &= \sum_{k \in \mathbb{Z}} x_{n-k} y(k) && \text{where } k \triangleq n - m \implies m = n - k \\
 &= \sum_{k \in \mathbb{Z}} x_{n-k} y(k) && \text{by commutativity of addition} \\
 &= \sum_{m \in \mathbb{Z}} x_{n-m} y_m && \text{by change of variables} \\
 &= \sum_{m \in \mathbb{Z}} y_m x_{n-m} && \text{by commutative property of the field over } \mathbb{C} \\
 &\triangleq (y \star x)_n && \text{by Definition R.3 page 357}
 \end{aligned}$$

⇒

**Proposition R.2.** Let  $\star$  be the CONVOLUTION OPERATOR (Definition R.3 page 357). Let  $\ell^2_{\mathbb{R}}$  be the set of ABSOLUTELY SUMMABLE sequences (Definition R.2 page 357).

$$\boxed{\begin{array}{l} \textbf{P} \\ \textbf{R} \\ \textbf{P} \end{array} \left\{ \begin{array}{l} (A). \quad x(n) \in \ell^2_{\mathbb{R}} \text{ and} \\ (B). \quad y(n) \in \ell^2_{\mathbb{R}} \end{array} \right\} \Rightarrow \left\{ \sum_{k \in \mathbb{Z}} x[k]y[n+k] = x[-n] \star y(n) \right\}}$$

PROOF:

$$\begin{aligned}
 \sum_{k \in \mathbb{Z}} x[k]y[n+k] &= \sum_{-p \in \mathbb{Z}} x[-p]y[n-p] && \text{where } p \triangleq -k && \implies k = -p \\
 &= \sum_{p \in \mathbb{Z}} x[-p]y[n-p] && \text{by absolutely summable hypothesis} && (\text{Definition R.2 page 357}) \\
 &= \sum_{p \in \mathbb{Z}} x'[p]y[n-p] && \text{where } x'[n] \triangleq x[-n] && \implies x[-n] = x'[n] \\
 &\triangleq x'[n] \star y[n] && \text{by definition of convolution } \star && (\text{Definition R.3 page 357}) \\
 &\triangleq x[-n] \star y[n] && \text{by definition of } x'[n]
 \end{aligned}$$

⇒

## R.2 Z-transform

**Definition R.4.**<sup>3</sup>

$$\boxed{\begin{array}{l} \textbf{D} \\ \textbf{E} \\ \textbf{F} \end{array} \text{The } z\text{-transform } \mathbf{Z} \text{ of } (x_n)_{n \in \mathbb{Z}} \text{ is defined as} \\
 [\mathbf{Z}(x_n)](z) \triangleq \underbrace{\sum_{n \in \mathbb{Z}} x_n z^{-n}}_{\text{Laurent series}} \quad \forall (x_n) \in \ell^2_{\mathbb{R}}}$$

**Theorem R.1.** Let  $X(z) \triangleq \mathbf{Z}x[n]$  be the Z-TRANSFORM of  $x[n]$ .

$$\boxed{\begin{array}{l} \textbf{T} \\ \textbf{H} \\ \textbf{M} \end{array} \left\{ \check{x}(z) \triangleq \mathbf{Z}(x[n]) \right\} \implies \left\{ \begin{array}{l} (1). \quad \mathbf{Z}(\alpha x[n]) = \alpha \check{x}(z) \quad \forall (x_n) \in \ell^2_{\mathbb{R}} \text{ and} \\ (2). \quad \mathbf{Z}(x[n-k]) = z^{-k} \check{x}(z) \quad \forall (x_n) \in \ell^2_{\mathbb{R}} \text{ and} \\ (3). \quad \mathbf{Z}(x[-n]) = \check{x}\left(\frac{1}{z}\right) \quad \forall (x_n) \in \ell^2_{\mathbb{R}} \text{ and} \\ (4). \quad \mathbf{Z}(x^*[n]) = \check{x}^*\left(z^*\right) \quad \forall (x_n) \in \ell^2_{\mathbb{R}} \text{ and} \\ (5). \quad \mathbf{Z}(x^*[-n]) = \check{x}^*\left(\frac{1}{z^*}\right) \quad \forall (x_n) \in \ell^2_{\mathbb{R}} \end{array} \right\}}$$

<sup>3</sup>Laurent series:  Abramovich and Aliprantis (2002) page 49



PROOF:

$$\begin{aligned}
 \alpha \mathbb{Z} \check{x}(z) &\triangleq \alpha \mathbb{Z}(\check{x}[n]) && \text{by definition of } \check{x}(z) \\
 &\triangleq \alpha \sum_{n \in \mathbb{Z}} x[n] z^{-n} && \text{by definition of } \mathbb{Z} \text{ operator} \\
 &\triangleq \sum_{n \in \mathbb{Z}} (\alpha x[n]) z^{-n} && \text{by distributive property} \\
 &\triangleq \mathbb{Z}(\alpha x[n]) && \text{by definition of } \mathbb{Z} \text{ operator} \\
 z^{-k} \check{x}(z) &= z^{-k} \mathbb{Z}(x[n]) && \text{by definition of } \check{x}(z) \quad (\text{left hypothesis}) \\
 &\triangleq z^{-k} \sum_{n=-\infty}^{n=+\infty} x[n] z^{-n} && \text{by definition of } \mathbb{Z} \\
 &= \sum_{n=-\infty}^{n=+\infty} x[n] z^{-n-k} && (\text{Definition R.4 page 358}) \\
 &= \sum_{m=k-\infty}^{m=+\infty} x[m-k] z^{-m} && \text{where } m \triangleq n+k \quad \implies n = m - k \\
 &= \sum_{m=-\infty}^{m=+\infty} x[m-k] z^{-m} \\
 &= \sum_{n=-\infty}^{n=+\infty} x[n-k] z^{-n} && \text{where } n \triangleq m \\
 &\triangleq \mathbb{Z}(x[n-k]) && \text{by definition of } \mathbb{Z} \quad (\text{Definition R.4 page 358}) \\
 \mathbb{Z}(x^*[n]) &\triangleq \sum_{n \in \mathbb{Z}} x^*[n] z^{-n} && \text{by definition of } \mathbb{Z} \quad (\text{Definition R.4 page 358}) \\
 &\triangleq \left( \sum_{n \in \mathbb{Z}} x[n] (z^*)^{-n} \right)^* && \text{by definition of } \mathbb{Z} \quad (\text{Definition R.4 page 358}) \\
 &\triangleq \check{x}^*(z^*) && \text{by definition of } \mathbb{Z} \quad (\text{Definition R.4 page 358}) \\
 \mathbb{Z}(x[-n]) &\triangleq \sum_{n \in \mathbb{Z}} x[-n] z^{-n} && \text{by definition of } \mathbb{Z} \quad (\text{Definition R.4 page 358}) \\
 &= \sum_{-m \in \mathbb{Z}} x[m] z^m && \text{where } m \triangleq -n \quad \implies n = -m \\
 &= \sum_{m \in \mathbb{Z}} x[m] z^m && \text{by absolutely summable property} \quad (\text{Definition R.2 page 357}) \\
 &= \sum_{m \in \mathbb{Z}} x[m] \left( \frac{1}{z} \right)^{-m} && \text{by absolutely summable property} \quad (\text{Definition R.2 page 357}) \\
 &\triangleq \check{x}\left(\frac{1}{z}\right) && \text{by definition of } \mathbb{Z} \quad (\text{Definition R.4 page 358}) \\
 \mathbb{Z}(x^*[-n]) &\triangleq \sum_{n \in \mathbb{Z}} x^*[-n] z^{-n} && \text{by definition of } \mathbb{Z} \quad (\text{Definition R.4 page 358}) \\
 &= \sum_{-m \in \mathbb{Z}} x^*[m] z^m && \text{where } m \triangleq -n \quad \implies n = -m \\
 &= \sum_{m \in \mathbb{Z}} x^*[m] z^m && \text{by absolutely summable property} \quad (\text{Definition R.2 page 357}) \\
 &= \sum_{m \in \mathbb{Z}} x^*[m] \left( \frac{1}{z} \right)^{-m} && \text{by absolutely summable property} \quad (\text{Definition R.2 page 357}) \\
 &= \left( \sum_{m \in \mathbb{Z}} x[m] \left( \frac{1}{z^*} \right)^{-m} \right)^* && \text{by absolutely summable property} \quad (\text{Definition R.2 page 357})
 \end{aligned}$$

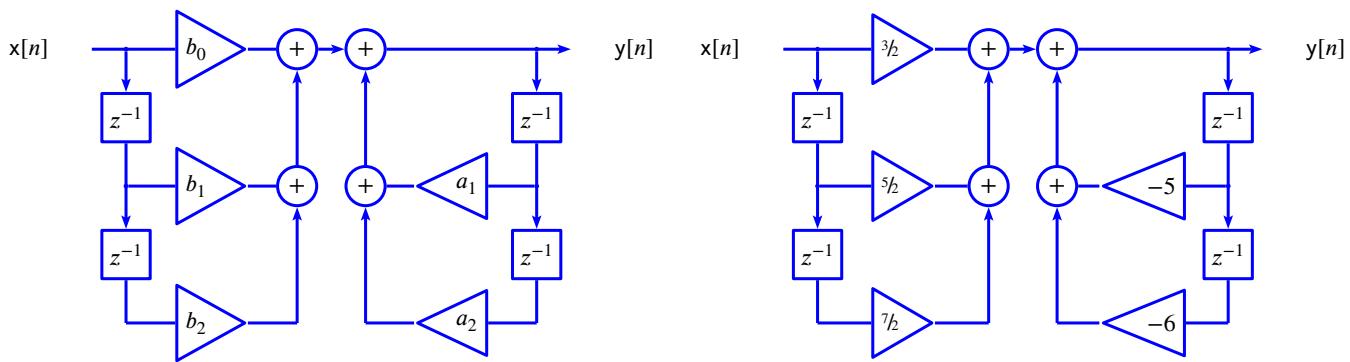


Figure R.1: Direct form 1 order 2 IIR filters

$$\triangleq \check{x}^* \left( \frac{1}{z^*} \right) \quad \text{by definition of } \mathbf{Z} \quad (\text{Definition R.4 page 358})$$

⇒

**Theorem R.2** (convolution theorem). Let  $\star$  be the convolution operator (Definition R.3 page 357).

<b>T</b> <b>H</b> <b>M</b>	$\mathbf{Z} \underbrace{\left( \langle x_n \rangle \star \langle y_n \rangle \right)}_{\text{sequence convolution}} = \underbrace{\left( \mathbf{Z} \langle x_n \rangle \right) \left( \mathbf{Z} \langle y_n \rangle \right)}_{\text{series multiplication}}$	$\forall \langle x_n \rangle_{n \in \mathbb{Z}}, \langle y_n \rangle_{n \in \mathbb{Z}} \in \ell^2_{\mathbb{R}}$
----------------------------------	--	--

⇒

PROOF:

$$\begin{aligned}
 [\mathbf{Z}(x \star y)](z) &\triangleq \mathbf{Z} \left( \sum_{m \in \mathbb{Z}} x_m y_{n-m} \right) && \text{by Definition R.3 page 357} \\
 &\triangleq \sum_{n \in \mathbb{Z}} \sum_{m \in \mathbb{Z}} x_m y_{n-m} z^{-n} && \text{by Definition R.4 page 358} \\
 &= \sum_{n \in \mathbb{Z}} \sum_{m \in \mathbb{Z}} x_m y_{n-m} z^{-n} \\
 &= \sum_{m \in \mathbb{Z}} \sum_{n \in \mathbb{Z}} x_m y_{n-m} z^{-n} \\
 &= \sum_{m \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} x_m y_k z^{-(m+k)} && \text{where } k = n - m \iff n = m + k \\
 &= \left[ \sum_{m \in \mathbb{Z}} x_m z^{-m} \right] \left[ \sum_{k \in \mathbb{Z}} y_k z^{-k} \right] \\
 &\triangleq (\mathbf{Z} \langle x_n \rangle) (\mathbf{Z} \langle y_n \rangle) && \text{by Definition R.4 page 358}
 \end{aligned}$$

⇒

### R.3 From z-domain back to time-domain

$$\check{y}(z) = b_0 X(z) + b_1 z^{-1} X(z) + b_2 z^{-2} X(z) - a_1 z^{-1} Y(z) + a_2 z^{-2} Y(z)$$

$$y[n] = b_0 x[n] + b_1 x[n-1] + b_2 x[n-2] - a_1 y[n-1] - a_2 y[n-2]$$

*Example R.1.* See Figure R.1 (page 360)

$$\frac{3z^2 + 5z + 7}{2z^2 + 10z + 12} = \frac{3z^2 + 5z + 7}{2(z^2 + 5z + 6)} = \frac{(3z^2 + 5z + 7)z^{-2}}{z^2 + 5z + 6} = \frac{(3z^2 + 5z + 7)z^{-2}}{1 + 5z^{-1} + 6z^{-2}}$$

## R.4 Zero locations

The system property of *minimum phase* is defined in Definition R.5 (next) and illustrated in Figure R.2 (page 361).

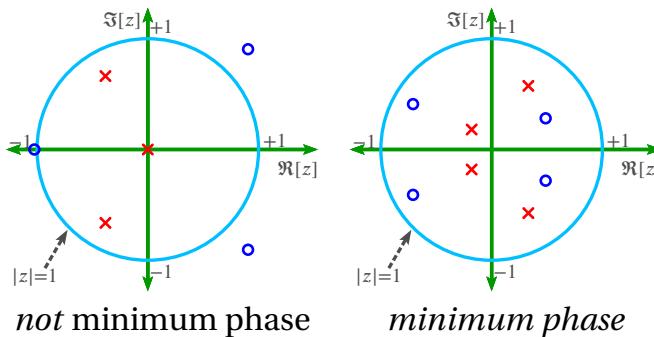


Figure R.2: Minimum Phase filter

**Definition R.5.**<sup>4</sup> Let  $\check{x}(z) \triangleq \mathbf{Z}(x_n)$  be the Z TRANSFORM (Definition R.4 page 358) of a sequence  $(x_n)_{n \in \mathbb{Z}}$  in  $\ell_R^2$ . Let  $(z_n)_{n \in \mathbb{Z}}$  be the ZEROS of  $\check{x}(z)$ .

**DEF** The sequence  $(x_n)$  is **minimum phase** if  

$$\underbrace{|z_n| < 1}_{\check{x}(z) \text{ has all its ZEROS inside the unit circle}} \quad \forall n \in \mathbb{Z}$$

The impulse response of a minimum phase filter has most of its energy concentrated near the beginning of its support, as demonstrated next.

**Theorem R.3** (Robinson's Energy Delay Theorem).<sup>5</sup> Let  $p(z) \triangleq \sum_{n=0}^N a_n z^{-n}$  and  $q(z) \triangleq \sum_{n=0}^N b_n z^{-n}$  be polynomials.

**THM**  $\left\{ \begin{array}{l} p \text{ is MINIMUM PHASE} \\ q \text{ is NOT minimum phase} \end{array} \right. \text{ and } \right\} \Rightarrow \underbrace{\sum_{n=0}^{m-1} |a_n|^2}_{\substack{\text{"energy"} \\ \text{of} \\ \text{the first } m \\ \text{coefficients} \\ \text{of} \\ p(z)}} \geq \underbrace{\sum_{n=0}^{m-1} |b_n|^2}_{\substack{\text{"energy"} \\ \text{of} \\ \text{the first } m \\ \text{coefficients} \\ \text{of} \\ q(z)}} \quad \forall 0 \leq m \leq N$

But for more *symmetry*, put some zeros inside and some outside the unit circle (Figure R.3 page 362).

*Example R.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

<sup>4</sup> Farina and Rinaldi (2000) page 91, Dumitrescu (2007) page 36

<sup>5</sup> Dumitrescu (2007) page 36, Robinson (1962), Robinson (1966) *(???)*, Claerbout (1976), pages 52–53

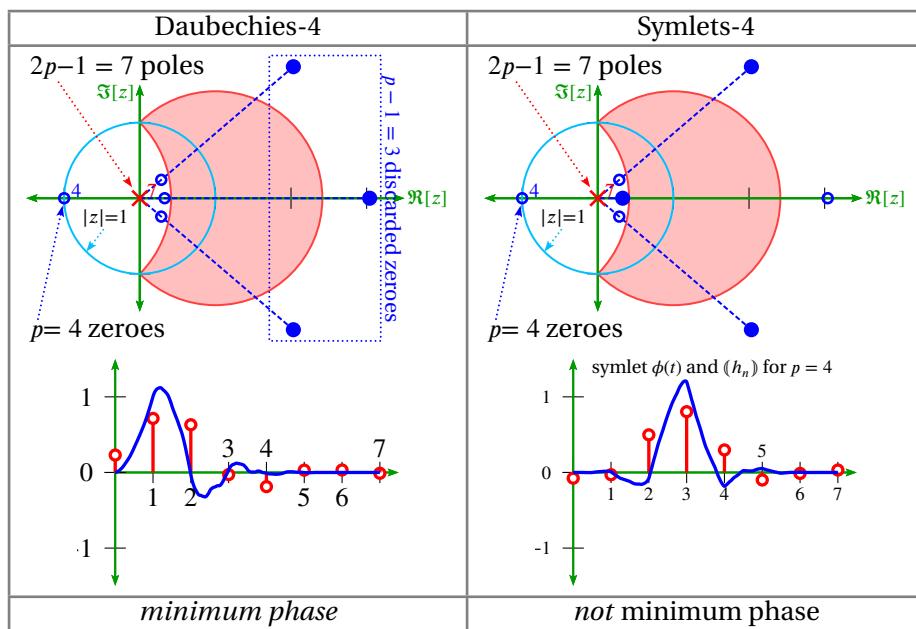


Figure R.3: Daubechies-4 and Symlet-4 scaling functions pole-zero plots

scaling function that is more symmetric and less contracted near the origin. Both scaling functions are illustrated in Figure R.3 (page 362).

## R.5 Pole locations

### Definition R.6.

**D E F** A filter (or system or operator)  $\mathbf{H}$  is **causal** if its current output does not depend on future inputs.

### Definition R.7.

**D E F** A filter (or system or operator)  $\mathbf{H}$  is **time-invariant** if the mapping it performs does not change with time.

### Definition R.8.

**D E F** An operation  $\mathbf{H}$  is **linear** if any output  $y_n$  can be described as a linear combination of inputs  $x_n$  as in

$$y_n = \sum_{m \in \mathbb{Z}} h(m)x(n-m).$$

For a filter to be *stable*, place all the poles *inside* the unit circle.

**Theorem R.4.** A causal LTI filter is **stable** if all of its poles are **inside** the unit circle.

**Example R.3.** Stable/unstable filters are illustrated in Figure R.4 (page 363).

True or False? This filter has no poles:



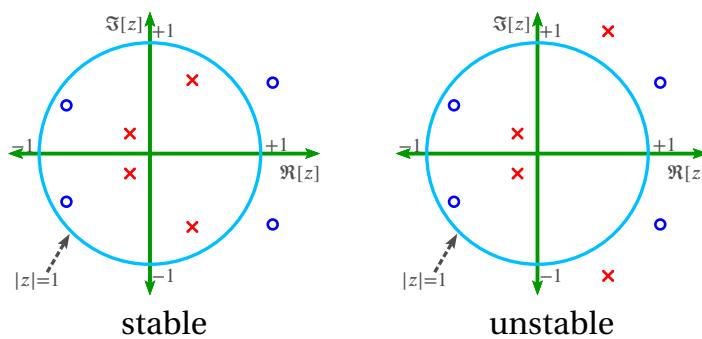
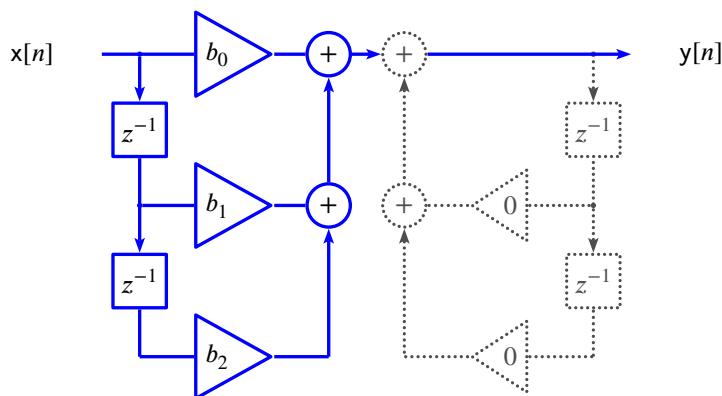
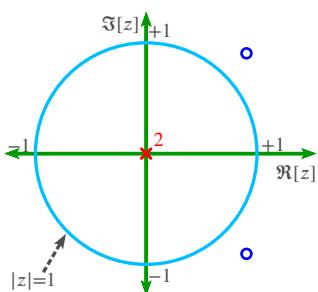


Figure R.4: Pole-zero plot stable/unstable causal LTI filters (Example R.3 page 362)

$$H(z) = b_0 + b_1 z^{-1} + b_2 z^{-2}$$



$$H(z) = b_0 + b_1 z^{-1} + b_2 z^{-2} = \frac{z^2}{z^2} \times \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{1} = \frac{b_0 z^2 + b_1 z^1 + b_2}{z^2}$$



## R.6 Mirroring for real coefficients

If you want real coefficients, choose poles and zeros in conjugate pairs (next).

### Proposition R.3.

P	$\left\{ \begin{array}{l} \text{ZEROS and POLES} \\ \text{occur in CONJUGATE PAIRS} \end{array} \right\}$	$\Rightarrow$	$\left\{ \begin{array}{l} \text{COEFFICIENTS} \\ \text{are REAL.} \end{array} \right\}$
---	---	---------------	---

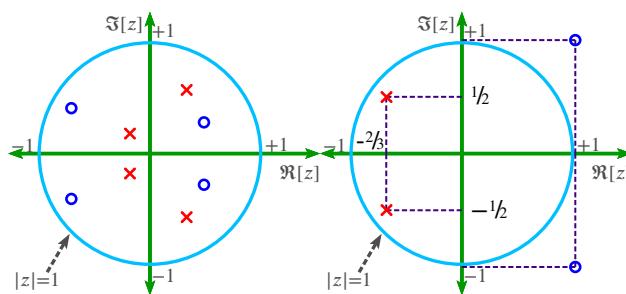


Figure R.5: Conjugate pair structure yielding real coefficients

PROOF:

$$\begin{aligned}(z - p_1)(z - p_1^*) &= [z - (a + ib)][z - (a - ib)] \\ &= z^2 + [-a + ib - ib - a]z - [ib]^2 \\ &= z^2 - 2az + b^2\end{aligned}$$

⇒

Example R.4. See Figure R.5 (page 364).

$$\begin{aligned}H(z) &= G \frac{[z - z_1][z - z_2]}{[z - p_1][z - p_2]} = G \frac{[z - (1+i)][z - (1-i)]}{[z - (-^{2/3} + i^{1/2})][z - (-^{2/3} - i^{1/2})]} \\ &= G \frac{z^2 - z[(1-i) + (1+i)] + (1-i)(1+i)}{z^2 - z[(-^{2/3} + i^{1/2}) + (-^{2/3} - i^{1/2})] + (-^{2/3} + i^{1/2})(-^{2/3} - i^{1/2})} \\ &= G \frac{z^2 - 2z + 2}{z^2 - \frac{4}{3}z + (\frac{4}{3} + \frac{1}{4})} = G \frac{z^2 - 2z + 2}{z^2 - \frac{4}{3}z + \frac{19}{12}}\end{aligned}$$

## R.7 Rational polynomial operators

A digital filter is simply an operator on  $\ell_{\mathbb{R}}^2$ . If the digital filter is a causal LTI system, then it can be expressed as a rational polynomial in  $z$  as shown next.

**Lemma R.1.** A causal LTI operator  $H$  can be expressed as a rational expression  $\check{h}(z)$ .

$$\begin{aligned}\check{h}(z) &= \frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + \dots + b_N z^{-N}}{1 + a_1 z^{-1} + a_2 z^{-2} + \dots + a_N z^{-N}} \\ &= \frac{\sum_{n=0}^N b_n z^{-n}}{1 + \sum_{n=1}^N a_n z^{-n}}\end{aligned}$$

A filter operation  $\check{h}(z)$  can be expressed as a product of its roots (poles and zeros).

$$\begin{aligned}\check{h}(z) &= \frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + \dots + b_N z^{-N}}{1 + a_1 z^{-1} + a_2 z^{-2} + \dots + a_N z^{-N}} \\ &= \alpha \frac{(z - z_1)(z - z_2) \dots (z - z_N)}{(z - p_1)(z - p_2) \dots (z - p_N)}\end{aligned}$$

where  $\alpha$  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  $\circ$ . An example is shown in Figure R.6 page 365. Notice that in this figure the zeros and poles are either real or occur in complex conjugate pairs.

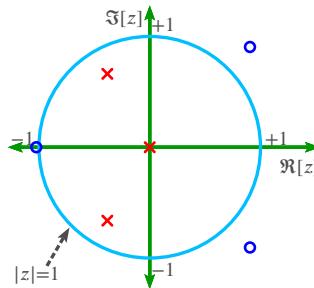


Figure R.6: Pole-zero plot for rational expression with real coefficients

## R.8 Filter Banks

*Conjugate quadrature filters* (next definition) are used in *filter banks*. If  $\check{x}(z)$  is a *low-pass filter*, then the conjugate quadrature filter of  $\check{y}(z)$  is a *high-pass filter*.

**Definition R.9.**<sup>6</sup> Let  $(x_n)_{n \in \mathbb{Z}}$  and  $(y_n)_{n \in \mathbb{Z}}$  be SEQUENCES (Definition R.1 page 357) in  $\ell^2_{\mathbb{R}}$  (Definition R.2 page 357).

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  $((x_n), (y_n), N)$  in this form is said to satisfy the

**conjugate quadrature filter condition** or the **CQF condition**.

**Theorem R.5** (CQF theorem).<sup>7</sup> Let  $\check{y}(\omega)$  and  $\check{x}(\omega)$  be the DTFTs (Definition Q.1 page 347) of the sequences  $(y_n)_{n \in \mathbb{Z}}$  and  $(x_n)_{n \in \mathbb{Z}}$ , respectively, in  $\ell^2_{\mathbb{R}}$  (Definition R.2 page 357).

T H M	$\underbrace{y_n = \pm(-1)^n x_{N-n}^*}_{(1) \text{ CQF in "time"} } \iff \check{y}(z) = \pm(-1)^N z^{-N} \check{x}^*\left(\frac{-1}{z^*}\right) \quad (2) \text{ CQF in "z-domain"}$ $\iff \check{y}(\omega) = \pm(-1)^N e^{-i\omega N} \check{x}^*(\omega + \pi) \quad (3) \text{ CQF in "frequency"}$ $\iff x_n = \pm(-1)^N (-1)^n y_{N-n}^* \quad (4) \text{ "reversed" CQF in "time"}$ $\iff \check{x}(z) = \pm z^{-N} \check{y}^*\left(\frac{-1}{z^*}\right) \quad (5) \text{ "reversed" CQF in "z-domain"}$ $\iff \check{x}(\omega) = \pm e^{-i\omega N} \check{y}^*(\omega + \pi) \quad (6) \text{ "reversed" CQF in "frequency"}$
-------------	--

$\forall N \in \mathbb{Z}$

PROOF:

<sup>6</sup> 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)

<sup>7</sup> 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))

1. Proof that (1)  $\Rightarrow$  (2):

$$\begin{aligned}
 \check{y}(z) &= \sum_{n \in \mathbb{Z}} y_n z^{-n} && \text{by definition of } z\text{-transform} && (\text{Definition R.4 page 358}) \\
 &= \sum_{n \in \mathbb{Z}} (\pm)(-1)^n x_{N-n}^* z^{-n} && \text{by (1)} \\
 &= \pm \sum_{m \in \mathbb{Z}} (-1)^{N-m} x_m^* z^{-(N-m)} && \text{where } m \triangleq N - n \Rightarrow && n = N - m \\
 &= \pm(-1)^N z^{-N} \sum_{m \in \mathbb{Z}} (-1)^{-m} x_m^* (z^{-1})^{-m} \\
 &= \pm(-1)^N z^{-N} \sum_{m \in \mathbb{Z}} x_m^* \left(-\frac{1}{z}\right)^{-m} \\
 &= \pm(-1)^N z^{-N} \left[ \sum_{m \in \mathbb{Z}} x_m \left(-\frac{1}{z^*}\right)^{-m} \right]^* \\
 &= \pm(-1)^N z^{-N} \check{x}^* \left(\frac{-1}{z^*}\right)
 \end{aligned}$$

by definition of  $z$ -transform (Definition R.4 page 358)

2. Proof that (1)  $\Leftarrow$  (2):

$$\begin{aligned}
 \check{y}(z) &= \pm(-1)^N z^{-N} \check{x}^* \left(\frac{-1}{z^*}\right) && \text{by (2)} \\
 &= \pm(-1)^N z^{-N} \left[ \sum_{m \in \mathbb{Z}} x_m \left(\frac{-1}{z^*}\right)^{-m} \right]^* && \text{by definition of } z\text{-transform} && (\text{Definition R.4 page 358}) \\
 &= \pm(-1)^N z^{-N} \left[ \sum_{m \in \mathbb{Z}} x_m^* (-z^{-1})^{-m} \right] && \text{by definition of } z\text{-transform} && (\text{Definition R.4 page 358}) \\
 &= \sum_{m \in \mathbb{Z}} (\pm)(-1)^{N-m} x_m^* z^{-(N-m)} \\
 &= \sum_{m \in \mathbb{Z}} (\pm)(-1)^n x_{N-n}^* z^{-n} && \text{where } n = N - m \Rightarrow && m \triangleq N - n \\
 &\Rightarrow x_n = \pm(-1)^n x_{N-n}^*
 \end{aligned}$$

3. Proof that (1)  $\Rightarrow$  (3):

$$\begin{aligned}
 \check{y}(\omega) &\triangleq \check{y}(z) \Big|_{z=e^{i\omega}} && \text{by definition of DTFT (Definition Q.1 page 347)} \\
 &= \left[ \pm(-1)^N z^{-N} \check{x} \left(\frac{-1}{z^*}\right) \right]_{z=e^{i\omega}} && \text{by (2)} \\
 &= \pm(-1)^N e^{-i\omega N} \check{x} (e^{i\pi} e^{i\omega}) \\
 &= \pm(-1)^N e^{-i\omega N} \check{x} (e^{i(\omega+\pi)}) \\
 &= \pm(-1)^N e^{-i\omega N} \check{x}(\omega + \pi)
 \end{aligned}$$

by definition of DTFT (Definition Q.1 page 347)

4. Proof that (1)  $\Rightarrow$  (6):

$$\begin{aligned}
 \check{x}(\omega) &= \sum_{n \in \mathbb{Z}} y_n e^{-i\omega n} && \text{by definition of DTFT} && (\text{Definition Q.1 page 347}) \\
 &= \sum_{n \in \mathbb{Z}} \underbrace{\pm(-1)^n x_{N-n}^*}_{CQF} e^{-i\omega n} && \text{by (1)} \\
 &= \sum_{m \in \mathbb{Z}} \pm(-1)^{N-m} x_m^* e^{-i\omega(N-m)} && \text{where } m \triangleq N - n \Rightarrow && n = N - m \\
 &= \pm(-1)^N e^{-i\omega N} \sum_{m \in \mathbb{Z}} (-1)^m x_m^* e^{i\omega m}
 \end{aligned}$$



$$\begin{aligned}
&= \pm(-1)^N e^{-i\omega N} \sum_{m \in \mathbb{Z}} e^{i\pi m} x_m^* e^{i\omega m} \\
&= \pm(-1)^N e^{-i\omega N} \sum_{m \in \mathbb{Z}} x_m^* e^{i(\omega+\pi)m} \\
&= \pm(-1)^N e^{-i\omega N} \left[ \sum_{m \in \mathbb{Z}} x_m e^{-i(\omega+\pi)m} \right]^* \\
&= \pm(-1)^N e^{-i\omega N} \check{x}^*(\omega + \pi) \quad \text{by definition of DTFT} \quad (\text{Definition Q.1 page 347})
\end{aligned}$$

5. Proof that (1)  $\iff$  (3):

$$\begin{aligned}
y_n &= \frac{1}{2\pi} \int_{-\pi}^{+\pi} \check{y}(\omega) e^{i\omega n} d\omega && \text{by inverse DTFT} \quad (\text{Theorem Q.3 page 353}) \\
&= \frac{1}{2\pi} \int_{-\pi}^{+\pi} \underbrace{\pm(-1)^N e^{-iN\omega}}_{\text{right hypothesis}} \check{x}^*(\omega + \pi) e^{i\omega n} d\omega && \text{by right hypothesis} \\
&= \pm(-1)^N \frac{1}{2\pi} \int_{-\pi}^{+\pi} \check{x}^*(\omega + \pi) e^{i\omega(n-N)} d\omega && \text{by right hypothesis} \\
&= \pm(-1)^N \frac{1}{2\pi} \int_0^{2\pi} \check{x}^*(v) e^{i(v-\pi)(n-N)} dv && \text{where } v \triangleq \omega + \pi \implies \omega = v - \pi \\
&= \pm(-1)^N e^{-i\pi(n-N)} \frac{1}{2\pi} \int_0^{2\pi} \check{x}^*(v) e^{iv(n-N)} dv \\
&= \pm(-1)^N \underbrace{(-1)^N}_{e^{i\pi N}} \underbrace{(-1)^n}_{e^{-i\pi n}} \left[ \frac{1}{2\pi} \int_0^{2\pi} \check{x}(v) e^{iv(N-n)} dv \right]^* \\
&= \pm(-1)^n x_{N-n}^* && \text{by inverse DTFT} \quad (\text{Theorem Q.3 page 353})
\end{aligned}$$

6. Proof that (1)  $\iff$  (4):

$$\begin{aligned}
y_n = \pm(-1)^n x_{N-n}^* &\iff (\pm)(-1)^n y_n = (\pm)(\pm)(-1)^n (-1)^n x_{N-n}^* \\
&\iff \pm(-1)^n y_n = x_{N-n}^* \\
&\iff (\pm(-1)^n y_n)^* = (x_{N-n}^*)^* \\
&\iff \pm(-1)^n y_n^* = x_{N-n} \\
&\iff x_{N-n} = \pm(-1)^n y_n^* \\
&\iff x_m = \pm(-1)^{N-m} y_{N-m}^* && \text{where } m \triangleq N - n \implies n = N - m \\
&\iff x_m = \pm(-1)^{N-m} y_{N-m}^* \\
&\iff x_m = \pm(-1)^N (-1)^m y_{N-m}^* \\
&\iff x_n = \pm(-1)^N (-1)^n y_{N-n}^* && \text{by change of free variables}
\end{aligned}$$

7. Proofs for (5) and (6): not included. See proofs for (2) and (3).



**Theorem R.6.**<sup>8</sup> Let  $\check{y}(\omega)$  and  $\check{x}(\omega)$  be the DTFTs (Definition Q.1 page 347) of the sequences  $(y_n)_{n \in \mathbb{Z}}$  and  $(x_n)_{n \in \mathbb{Z}}$ , respectively, in  $\ell^2_{\mathbb{R}}$  (Definition R.2 page 357).

T H M	<p>Let <math>y_n = \pm(-1)^n x_{N-n}^*</math> (CQF CONDITION, Definition R.9 page 365). Then</p> $ \left\{ \begin{aligned} (A) \quad \left[ \frac{d}{d\omega} \right]^n \check{y}(\omega) \Big _{\omega=0} &= 0 \iff \left[ \frac{d}{d\omega} \right]^n \check{x}(\omega) \Big _{\omega=\pi} &= 0 & (B) \\ &\iff \sum_{k \in \mathbb{Z}} (-1)^k k^n x_k &= 0 & (C) \\ &\iff \sum_{k \in \mathbb{Z}} k^n y_k &= 0 & (D) \end{aligned} \right\} \forall n \in \mathbb{W} $
-------------	--

<sup>8</sup> Vidakovic (1999), pages 82–83, Mallat (1999), pages 241–242

PROOF:

1. Proof that (A)  $\implies$  (B):

$$\begin{aligned}
 0 &= \left[ \frac{d}{d\omega} \right]^n \check{y}(\omega) \Big|_{\omega=0} && \text{by (A)} \\
 &= \left[ \frac{d}{d\omega} \right]^n (\pm)(-1)^N e^{-i\omega N} \check{x}^*(\omega + \pi) \Big|_{\omega=0} && \text{by CQF theorem (Theorem R.5 page 365)} \\
 &= (\pm)(-1)^N \sum_{\ell=0}^n \binom{n}{\ell} \left[ \frac{d}{d\omega} \right]^\ell [e^{-i\omega N}] \cdot \left[ \frac{d}{d\omega} \right]^{n-\ell} [\check{x}^*(\omega + \pi)] \Big|_{\omega=0} && \text{by Leibnitz GPR (Lemma ?? page ??)} \\
 &= (\pm)(-1)^N \sum_{\ell=0}^n \binom{n}{\ell} -iN^\ell e^{-i\omega N} \left[ \frac{d}{d\omega} \right]^{n-\ell} [\check{x}^*(\omega + \pi)] \Big|_{\omega=0} \\
 &= (\pm)(-1)^N e^{-i0N} \sum_{\ell=0}^n \binom{n}{\ell} -iN^\ell \left[ \frac{d}{d\omega} \right]^{n-\ell} [\check{x}^*(\omega + \pi)] \Big|_{\omega=0} \\
 &\implies \check{x}^{(0)}(\pi) = 0 \\
 &\implies \check{x}^{(1)}(\pi) = 0 \\
 &\implies \check{x}^{(2)}(\pi) = 0 \\
 &\implies \check{x}^{(3)}(\pi) = 0 \\
 &\implies \check{x}^{(4)}(\pi) = 0 \\
 &\vdots \quad \vdots \\
 &\implies \check{x}^{(n)}(\pi) = 0 \quad \text{for } n = 0, 1, 2, \dots
 \end{aligned}$$

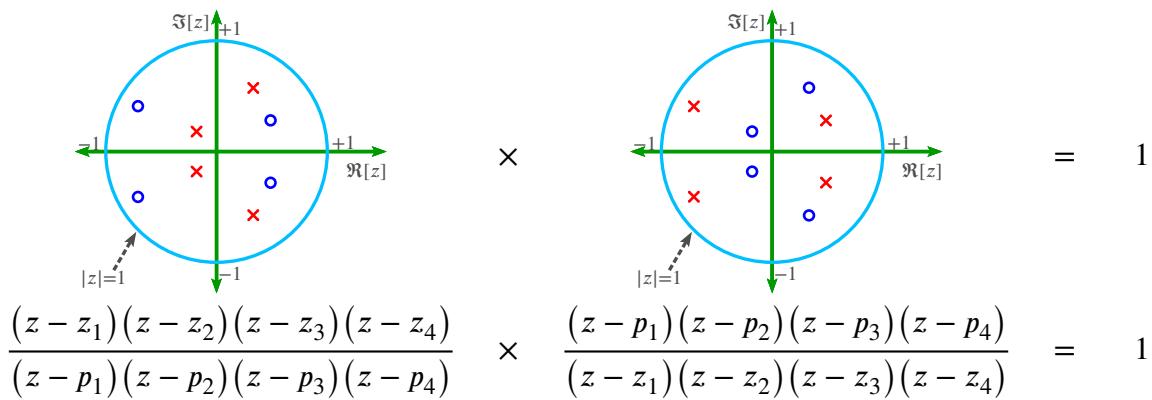
2. Proof that (A)  $\iff$  (B):

$$\begin{aligned}
 0 &= \left[ \frac{d}{d\omega} \right]^n \check{x}(\omega) \Big|_{\omega=\pi} && \text{by (B)} \\
 &= \left[ \frac{d}{d\omega} \right]^n (\pm) e^{-i\omega N} \check{y}^*(\omega + \pi) \Big|_{\omega=\pi} && \text{by CQF theorem (Theorem R.5 page 365)} \\
 &= (\pm) \sum_{\ell=0}^n \binom{n}{\ell} \left[ \frac{d}{d\omega} \right]^\ell [e^{-i\omega N}] \cdot \left[ \frac{d}{d\omega} \right]^{n-\ell} [\check{y}^*(\omega + \pi)] \Big|_{\omega=\pi} && \text{by Leibnitz GPR (Lemma ?? page ??)} \\
 &= (\pm) \sum_{\ell=0}^n \binom{n}{\ell} (-iN)^\ell e^{-i\omega N} \left[ \frac{d}{d\omega} \right]^{n-\ell} [\check{y}^*(\omega + \pi)] \Big|_{\omega=\pi} \\
 &= (\pm) e^{-i\pi N} \sum_{\ell=0}^n \binom{n}{\ell} -iN^\ell \left[ \frac{d}{d\omega} \right]^{n-\ell} [\check{y}^*(\omega + \pi)] \Big|_{\omega=\pi} \\
 &= (\pm)(-1)^N \sum_{\ell=0}^n \binom{n}{\ell} -iN^\ell \left[ \frac{d}{d\omega} \right]^{n-\ell} [\check{y}^*(\omega + \pi)] \Big|_{\omega=\pi} \\
 &\implies \check{y}^{(0)}(0) = 0 \\
 &\implies \check{y}^{(1)}(0) = 0 \\
 &\implies \check{y}^{(2)}(0) = 0 \\
 &\implies \check{y}^{(3)}(0) = 0 \\
 &\implies \check{y}^{(4)}(0) = 0 \\
 &\vdots \quad \vdots \\
 &\implies \check{y}^{(n)}(0) = 0 \quad \text{for } n = 0, 1, 2, \dots
 \end{aligned}$$

3. Proof that (B)  $\iff$  (C): by Theorem Q.5 page 355

4. Proof that (A)  $\iff$  (D): by Theorem Q.5 page 355



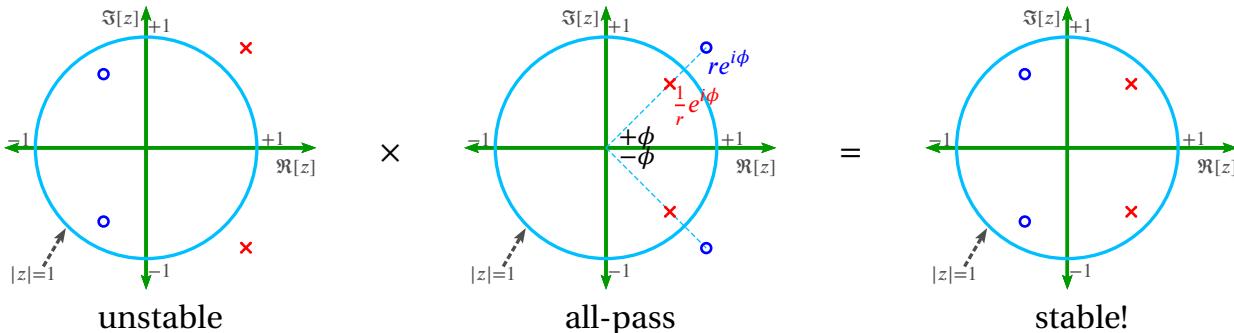


5. Proof that (CQF)  $\Leftrightarrow$  (A): Here is a counterexample:  $\tilde{y}(\omega) = 0$ .



## R.9 Inverting non-minimum phase filters

*Minimum phase* filters are easy to invert: each *zero* becomes a *pole* and each *pole* becomes a *zero*.



$$\begin{aligned}
 |A(z)|_{z=e^{i\omega}} &= \frac{1}{r} \left| \frac{z - re^{i\phi}}{z - \frac{1}{r}e^{i\phi}} \right|_{z=e^{i\omega}} \\
 &= \left| e^{i\phi} \left( \frac{e^{-i\phi}z - r}{rz - e^{i\phi}} \right) \right|_{z=e^{i\omega}} \\
 &= \left| -z \left( \frac{rz^{-1} - e^{-i\phi}}{rz - e^{i\phi}} \right) \right|_{z=e^{i\omega}} \\
 &= \left| \frac{1}{e^{-iv}} \left( \frac{re^{-i\omega} - e^{-i\phi}}{(re^{i\omega} - e^{i\phi})^*} \right) \right| \\
 &= 1
 \end{aligned}
 \quad
 \begin{aligned}
 &= \left| \frac{z - re^{i\phi}}{rz - e^{i\phi}} \right|_{z=e^{i\omega}} \\
 &= \left| z \left( \frac{e^{-i\phi} - rz^{-1}}{rz - e^{i\phi}} \right) \right|_{z=e^{i\omega}} \\
 &= \left| e^{i\pi} e^{i\omega} \left( \frac{re^{-i\omega} - e^{-i\phi}}{re^{i\omega} - e^{i\phi}} \right) \right| \\
 &= \left| \frac{re^{-i\omega} - e^{-i\phi}}{re^{-i\omega} - e^{i\phi}} \right|
 \end{aligned}$$



# APPENDIX S

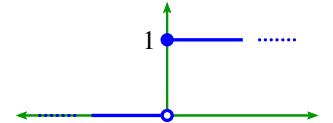
## B-SPLINES

### S.1 Definitions

**Definition S.1.** Let  $X$  be a set.

**D E F** The **step function**  $\sigma \in \mathbb{R}^{\mathbb{R}}$  is defined as  

$$\sigma(x) \triangleq \mathbb{1}_{[0:\infty)}(x) \quad \forall x \in \mathbb{R}.$$



**Lemma S.1.** Let  $\sigma(x)$  be the STEP FUNCTION (Definition S.1 page 371).

**L E M**  $\{g(x) > 0\} \implies \{\sigma[g(x)f(x)] = \sigma[f(x)]\} \quad \forall f, g \in \mathbb{R}^{\mathbb{R}}$

PROOF:

$$\begin{aligned}
 \sigma[g(x)f(x)] &\triangleq \mathbb{1}_{[0:\infty)}[g(x)f(x)] && \text{by definition of } \sigma(x) && (\text{Definition S.1 page 371}) \\
 &\triangleq \begin{cases} 1 & \text{for } g(x)f(x) \geq 0 \\ 0 & \text{otherwise} \end{cases} && \text{by definition of } \mathbb{1}_A(x) && (\text{Definition ?? page ??}) \\
 &= \begin{cases} 1 & \text{for } f(x) \geq 0 \\ 0 & \text{otherwise} \end{cases} && \text{by } g(x) > 0 \text{ hypothesis} && \\
 &\triangleq \mathbb{1}_{[0:\infty)}[f(x)] && \text{by definition of } \mathbb{1}_A(x) && (\text{Definition ?? page ??}) \\
 &\triangleq \sigma[f(x)] && \text{by definition of } \sigma(x) && (\text{Definition S.1 page 371})
 \end{aligned}$$

**Definition S.2.**<sup>1</sup> Let  $\mathbb{1}$  be the SET INDICATOR function (Definition ?? page ??). Let  $f(x) \star g(x)$  represent the CONVOLUTION operation (Definition P.3 page 340).

**D E F** The  **$n$ th order cardinal B-spline**  $N_n(x)$  for  $n \in \mathbb{W}$  is defined as  

$$N_n(x) \triangleq \begin{cases} \mathbb{1}_{[0:1)}(x) & \text{for } n = 0 \\ N_{n-1}(x) \star N_0(x) & \text{for } n \in \mathbb{W} \setminus 0 \end{cases} \quad \forall x \in \mathbb{R}$$

**Lemma S.2.**<sup>2</sup>

<sup>1</sup> Chui (1992) page 85 ((4.2.1)), Christensen (2008) page 140, Chui (1988) page 1

<sup>2</sup> Christensen (2008) page 140, Chui (1992) page 85 ((4.2.1)), Chui (1988) page 1, Prasad and Iyengar (1997) page 145

**L E M**  $N_n(x) = \int_{\tau=0}^{\tau=1} N_{n-1}(x - \tau) d\tau \quad \forall n \in \{1, 2, 3, \dots\}$

PROOF:

$$\begin{aligned}
 N_n(x) &\triangleq N_{n-1}(x) \star N_0(x) && \text{by definition of } N_n(x) && (\text{Definition S.2 page 371}) \\
 &\triangleq \int_{\mathbb{R}} N_{n-1}(x - \tau) N_0(\tau) d\tau && \text{by definition of convolution operation } \star && (\text{Definition P.3 page 340}) \\
 &\triangleq \int_{\mathbb{R}} N_{n-1}(x - \tau) \mathbb{1}_{[0:1]}(\tau) d\tau && \text{by definition of } N_0(x) && (\text{Definition S.2 page 371}) \\
 &= \int_{[0:1]} N_{n-1}(x - \tau) d\tau && \text{by definition of } \mathbb{1}_A(x) && (\text{Definition ?? page ??}) \\
 &= \int_{[0:1]} N_{n-1}(x - \tau) d\tau \\
 &\triangleq \int_0^1 N_{n-1}(x - \tau) d\tau
 \end{aligned}$$

⇒

**Lemma S.3.** Let  $f(x)$  be a FUNCTION in  $\mathbb{R}^{\mathbb{R}}$ . Let  $F(x)$  be the ANTI-DERIVATIVE of  $f(x)$ .

Let  $\sigma(x)$  be the STEP FUNCTION (Definition S.1 page 371).

**L E M**

$$\begin{aligned}
 &\int_{y=a}^{y=b} f(x - y) \sigma(x - y) dy \\
 &= \left\{ \begin{array}{ll} - \int_{y=x-a}^{y=x-b} f(y) dy & \text{for } x \geq b \\ - \int_{y=x-a}^{y=0} f(y) dy & \text{for } a \leq x \leq b \\ 0 & \text{for } x \leq a \end{array} \right\} = \left\{ \begin{array}{ll} F(x - a) - F(x - b) & \text{for } x \geq b \\ F(x - a) - F(0) & \text{for } a \leq x \leq b \\ 0 & \text{for } x \leq a \end{array} \right\} \\
 &= [F(x - a) - F(0)]\sigma(x - a) + [F(0) - F(x - b)]\sigma(x - b)
 \end{aligned}$$

PROOF:

$$\begin{aligned}
 \int_{y=a}^{y=b} f(x - y) \sigma(x - y) dy &= \left\{ \begin{array}{ll} \int_{y=a}^{y=b} f(x - y) dy & \text{for } x \geq b \\ \int_{y=a}^{y=x} f(x - y) dy & \text{for } a \leq x \leq b \\ 0 & \text{for } x \leq a \end{array} \right\} && \text{by definition of } \sigma \text{ (Definition S.1 page 371)} \\
 &= \left\{ \begin{array}{ll} - \int_{u=x-a}^{u=x-b} f(u) du & \text{for } x \geq b \\ - \int_{u=x-a}^{u=0} f(u) du & \text{for } a \leq x \leq b \\ 0 & \text{for } x \leq a \end{array} \right\} && \text{where } u \triangleq x - y \implies y = x - u \\
 &= \left\{ \begin{array}{ll} - \int_{y=x-a}^{y=x-b} f(y) dy & \text{for } x \geq b \\ - \int_{y=x-a}^{y=0} f(y) dy & \text{for } a \leq x \leq b \\ 0 & \text{for } x \leq a \end{array} \right\} && \text{by change of dummy variable } (u \rightarrow y) \\
 &= \left\{ \begin{array}{ll} F(x - a) - F(x - b) & \text{for } x \geq b \\ F(x - a) - F(0) & \text{for } a \leq x \leq b \\ 0 & \text{for } x \leq a \end{array} \right\} && \text{by Fundamental Theorem of Calculus} \\
 &= [F(x - a) - F(x - b)]\sigma(x - b) + [F(x - a) - F(0)][\sigma(x - a) - \sigma(x - b)] \\
 &= [F(x - a) - F(0)]\sigma(x - a) + [F(x - a) - F(x - b) - F(x - a) + F(0)]\sigma(x - b) \\
 &= [F(x - a) - F(0)]\sigma(x - a) + [F(0) - F(x - b)]\sigma(x - b)
 \end{aligned}$$

⇒

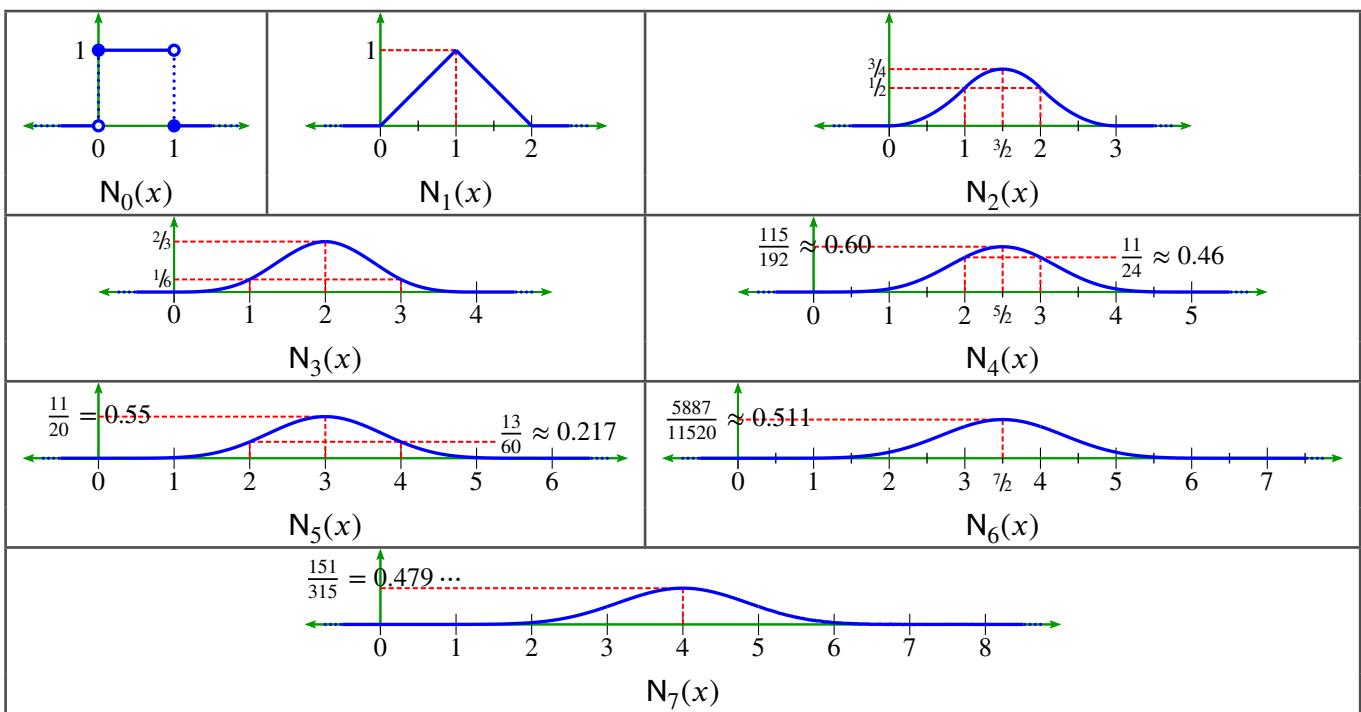


Figure S.1: some low order B-splines (Example S.1 page 373)

**Lemma S.4.** Let  $\sigma(x)$  be the STEP FUNCTION (Definition S.1 page 371).

$$\boxed{\text{LEM}} \quad \int_{\tau=0}^{\tau=1} (x - \tau - k)^n \sigma(x - \tau - k) d\tau = \frac{1}{n+1} [(x - k)^{n+1} \sigma(x - k) - (x - k - 1)^{n+1} \sigma(x - k - 1)]$$

PROOF:

$$\begin{aligned}
 & \int_{\tau=0}^{\tau=1} (x - \tau - k)^n \sigma(x - \tau - k) d\tau \\
 &= \int_{y=k}^{y=k+1} (x - y)^n \sigma(x - y) dy && \text{where } y \triangleq \tau + k \implies \tau = y - k \\
 &= [\mathcal{F}(x - k) - \mathcal{F}(0)] \sigma(x - k) + [\mathcal{F}(0) - \mathcal{F}(x - k - 1)] \sigma(x - k - 1) && \text{by Lemma S.3 (page 372), where } f(x) \triangleq x^n \\
 &= \frac{[(x - k)^{n+1} - 0] \sigma(x - k) + [0 - (x - k - 1)^{n+1}] \sigma(x - k - 1)}{n+1} && \text{because } \mathcal{F}(x) \triangleq \int f(x) dx = \frac{x^{n+1}}{n+1} + c \\
 &= \frac{1}{n+1} [(x - k)^{n+1} \sigma(x - k) - (x - k - 1)^{n+1} \sigma(x - k - 1)]
 \end{aligned}$$

☞

**Example S.1.**<sup>3</sup> Let  $\sigma(x)$  be the step function (Definition S.1 page 371). Let  $\binom{n}{k}$  be the binomial coefficient (Definition ?? page ??). The 0th order B-spline (Definition S.2 page 371)  $N_0(x)$  can be expressed as follows:

$$\boxed{\text{EX}} \quad N_0(x) = \left\{ \begin{array}{ll} 1 & \text{for } x \in [0 : 1) \\ 0 & \text{otherwise} \end{array} \right\} = \left\{ \sum_{k=0}^1 (-1)^k \binom{1}{k} (x - k)^0 \sigma(x - k) \quad \forall x \in \mathbb{R} \right\}$$

The B-spline  $N_0(x)$  is illustrated in Figure S.1 (page 373).

<sup>3</sup> Schumaker (2007) page 136 (Table 1)

PROOF:

$$\begin{aligned}
 N_0(x) &= \mathbb{1}_{[0:1)}(x) && \text{by definition of } N_0(x) \\
 &= \sigma(x) - \sigma(x-1) && \text{by definition of } \sigma(x) \\
 &= \left[ \binom{1}{0} \sigma(x) - \binom{1}{1} \sigma(x-1) \right] && \text{by definition of binomial coefficient } \binom{n}{k} \\
 &= \sum_{k=0}^1 (-1)^k \binom{1}{k} (x-k)^0 \sigma(x-k) && \text{by definition of } \sum \text{ operator}
 \end{aligned}$$

*Example S.2.* <sup>4</sup> Let  $\sigma(x)$  be the step function. Let  $\binom{n}{k}$  be the binomial coefficient.

The 1st order B-spline  $N_1(x)$  can be expressed as follows:

$$\boxed{\mathbf{E} \quad X \quad N_1(x) = \begin{cases} x & \text{for } x \in [0 : 1] \\ -x+2 & \text{for } x \in [1 : 2] \\ 0 & \text{otherwise} \end{cases} = \left\{ \sum_{k=0}^2 (-1)^k \binom{2}{k} (x-k) \sigma(x-k) \quad \forall x \in \mathbb{R} \right\}}$$

The B-spline  $N_1(x)$  is illustrated in Figure S.1 (page 373).

PROOF:

$$\begin{aligned}
 N_1(x) &= \int_{\tau=0}^{\tau=1} N_0(x-\tau) d\tau && \text{by Lemma S.2 page 371} \\
 &= \int_{\tau=0}^{\tau=1} \sum_{k=0}^1 (-1)^k \binom{1}{k} (x-\tau-k)^0 \sigma(x-\tau-k) d\tau && \text{by Example S.1 page 373} \\
 &= \sum_{k=0}^1 (-1)^k \binom{1}{k} \int_{\tau=0}^{\tau=1} (x-\tau-k)^0 \sigma(x-\tau-k) d\tau && \text{by linearity of } \int d\tau \text{ operator} \\
 &= \sum_{k=0}^1 (-1)^k \binom{1}{k} \frac{1}{0+1} [(x-k)^{0+1} \sigma(x-k) - (x-k-1)^{0+1} \sigma(x-k-1)] && \text{by Lemma S.4 page 373} \\
 &= \begin{pmatrix} 1\{(x-0)\sigma(x-0) - (x-1)\sigma(x-1)\} \\ -1\{(x-1)\sigma(x-1) - (x-2)\sigma(x-2)\} \end{pmatrix} \\
 &= x\sigma(x) - 2(x-1)\sigma(x-1) + (x-2)\sigma(x-2) \\
 &= \sum_{k=0}^2 (-1)^k \binom{2}{k} (x-k) \sigma(x-k) && \text{by def. of } \binom{n}{k} \text{ (Definition ?? page ??)} \\
 &= \begin{cases} x & \text{for } x \in [0 : 1] \\ -x+2 & \text{for } x \in [1 : 2] \\ 0 & \text{otherwise} \end{cases} && \text{by def. of } \sigma(x) \text{ (Definition S.1 page 371)}
 \end{aligned}$$

*Example S.3.* <sup>5</sup> Let  $\sigma(x)$  be the step function. Let  $\binom{n}{k}$  be the binomial coefficient.

The 2nd order B-spline  $N_2(x)$  can be expressed as follows:

$$\boxed{\mathbf{E} \quad X \quad N_2(x) = \frac{1}{2} \begin{cases} x^2 & \text{for } x \in [0 : 1) \\ -2x^2+6x-3 & \text{for } x \in [1 : 2] \\ x^2-6x+9 & \text{for } x \in [2 : 3] \\ 0 & \text{otherwise} \end{cases} = \left\{ \frac{1}{2} \sum_{k=0}^3 (-1)^k \binom{3}{k} (x-k)^2 \sigma(x-k) \quad \forall x \in \mathbb{R} \right\}}$$

The B-spline  $N_2(x)$  is illustrated in Figure S.1 (page 373).

<sup>4</sup>  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

<sup>5</sup>  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



PROOF:

$$\begin{aligned}
 N_2(x) &= \int_{\tau=0}^{\tau=1} N_1(x - \tau) d\tau && \text{by Lemma S.2 page 371} \\
 &= \int_{\tau=0}^{\tau=1} \sum_{k=0}^2 (-1)^k \binom{2}{k} (x - \tau - k) \sigma(x - \tau - k) d\tau && \text{by Example S.2 page 374} \\
 &= \sum_{k=0}^2 (-1)^k \binom{2}{k} \int_{\tau=0}^{\tau=1} (x - \tau - k) \sigma(x - \tau - k) d\tau && \text{by linearity of } \int d\tau \text{ operator} \\
 &= \sum_{k=0}^2 (-1)^k \binom{2}{k} \frac{1}{1+1} [(x - k)^{1+1} \sigma(x - k) - (x - k - 1)^{1+1} \sigma(x - k - 1)] && \text{by Lemma S.4 page 373} \\
 &= \frac{1}{2} \left( \begin{array}{l} 1 \quad \{(x-0)^2 \sigma(x-0) - (x-1)^2 \sigma(x-1)\} \\ -2 \quad \{(x-1)^2 \sigma(x-1) - (x-2)^2 \sigma(x-2)\} \\ +1 \quad \{(x-2)^2 \sigma(x-2) - (x-3)^2 \sigma(x-3)\} \end{array} \right) \\
 &= \frac{1}{2} [x^2 \sigma(x) - 3(x-1)^2 \sigma(x-1) + 3(x-2)^2 \sigma(x-2) - (x-3)^2 \sigma(x-3)] \\
 &= \frac{1}{2} \sum_{k=0}^3 (-1)^k \binom{3}{k} (x-k)^2 \sigma(x-k) && \text{by def. of } \binom{n}{k} \text{ (Definition ?? page ??)} \\
 &= \frac{1}{2} \left\{ \begin{array}{ll} x^2 & \text{for } x \in [0 : 1] \\ -2x^2 + 6x - 3 & \text{for } x \in [1 : 2] \\ x^2 - 6x + 9 & \text{for } x \in [2 : 3] \\ 0 & \text{otherwise} \end{array} \right\} && \text{by def. of } \sigma(x) \text{ (Definition S.1 page 371)}
 \end{aligned}$$

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 ??).  $\Rightarrow$

## S.2 Algebraic properties

Theorem S.1 (next) presents a closed form expression for an *n*th order B-spline  $N_n(x)$  based on the definition of  $N_n(x)$  given in Definition S.2 (page 371). Alternatively, Theorem S.1 could serve as the definition and Definition S.2 as a property.

**Theorem S.1.**<sup>6</sup> Let  $N_n(x)$  be the *n*th ORDER B-SPLINE (Definition S.2 page 371). Let  $\sigma(x)$  be the STEP FUNCTION (Definition S.1 page 371).

T H M	$N_n(x) = \frac{1}{n!} \sum_{k=0}^{n+1} (-1)^k \binom{n+1}{k} (x-k)^n \sigma(x-k) \quad \forall n \in \{0, 1, 2, \dots\} = \mathbb{W}$
-------------	--

PROOF: Proof follows by induction:

1. base case (choose one):
  - Proof for  $n = 0$  case: by Example S.1 (page 373).
  - Proof for  $n = 1$  case: by Example S.2 (page 374).
  - Proof for  $n = 2$  case: by Example S.3 (page 374).

<sup>6</sup> Christensen (2008) page 142 (Theorem 6.1.3), Chui (1992) page 84 ((4.1.12))

2. inductive step—proof that  $n$  case  $\implies n + 1$  case:

$$\begin{aligned}
 N_{n+1}(x) &= \int_0^1 N_n(x - \tau) d\tau && \text{by Lemma S.2 page 371} \\
 &= \int_0^1 \frac{1}{n!} \sum_{k=0}^{k=n+1} (-1)^k \binom{n+1}{k} (x - \tau - k)^n \sigma(x - \tau - k) d\tau && \text{by induction hypothesis} \\
 &= \frac{1}{n!} \sum_{k=0}^{k=n+1} (-1)^k \binom{n+1}{k} \int_0^1 (x - \tau - k)^n \sigma(x - \tau - k) d\tau && \text{by linearity of } \int d\tau \text{ operator} \\
 &= \frac{1}{n!} \sum_{k=0}^{k=n+1} (-1)^k \binom{n+1}{k} \frac{1}{n+1} [(x - k)^{n+1} \sigma(x - k) - (x - k - 1)^{n+1} \sigma(x - k - 1)] && \text{by Lemma S.4 page 373} \\
 &= \frac{1}{(n+1)!} \sum_{k=0}^{k=n+1} (-1)^k \binom{n+1}{k} [(x - k)^{n+1} \sigma(x - k) - (x - k - 1)^{n+1} \sigma(x - k - 1)] \\
 &= \frac{1}{(n+1)!} \left[ \sum_{k=0}^{k=n+1} (-1)^k \binom{n+1}{k} (x - k)^{n+1} \sigma(x - k) - \sum_{k=0}^{k=n+1} (-1)^k \binom{n+1}{k} (x - k - 1)^{n+1} \sigma(x - k - 1) \right] \\
 &= \frac{1}{(n+1)!} \left[ \sum_{k=0}^{k=n+1} (-1)^k \binom{n+1}{k} (x - k)^{n+1} \sigma(x - k) - \sum_{m=1}^{m=n+2} (-1)^{m-1} \binom{n+1}{m-1} (x - m)^{n+1} \sigma(x - m) \right]
 \end{aligned}$$

where  $m \triangleq k + 1 \implies k = m - 1$

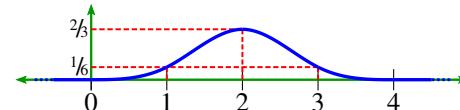
$$\begin{aligned}
 &= \frac{1}{(n+1)!} \left( \sum_{k=0}^{k=n+1} (-1)^k \binom{n+1}{k} (x - k)^{n+1} \sigma(x - k) - \sum_{m=1}^{m=n+2} (-1)^{m-1} \left[ \binom{n+2}{m} - \binom{n+1}{m} \right] (x - m)^{n+1} \sigma(x - m) \right) && \text{by Pascal's identity / Stifel formula (Theorem ?? page ??)} \\
 &= \frac{1}{(n+1)!} \left( \sum_{m=1}^{m=n+2} (-1)^m \binom{n+2}{m} (x - m)^{n+1} \sigma(x - m) - \sum_{m=1}^{m=n+2} (-1)^m \binom{n+1}{m} (x - m)^{n+1} \sigma(x - m) + \sum_{k=0}^{k=n+1} (-1)^k \binom{n+1}{k} (x - k)^{n+1} \sigma(x - k) \right) && \text{note } (-1)^{m-1} = -(-1)^m \\
 &= \frac{1}{(n+1)!} \left( \sum_{m=0}^{m=n+2} (-1)^m \binom{n+2}{m} (x - m)^{n+1} \sigma(x - m) - (-1)^0 \binom{n+2}{0} (x - 0)^{n+1} \sigma(x - 0) - \sum_{m=1}^{m=n+1} (-1)^m \binom{n+1}{m} (x - m)^{n+1} \sigma(x - m) - (-1)^{n+2} \binom{n+1}{n+2} (x - n - 2)^{n+1} \sigma(x - n - 2) + \sum_{k=1}^{k=n+1} (-1)^k \binom{n+1}{k} (x - k)^{n+1} \sigma(x - k) + (-1)^0 \binom{n+1}{0} (x - 0)^{n+1} \sigma(x - 0) \right) && \begin{array}{ll} \text{(A)} & \text{desired } n + 1 \text{ case} \\ \text{(B)} & \text{cancelled by (F)} \\ \text{(C)} & \text{cancelled by (E)} \\ \text{(D)} & \binom{n+1}{n+2} = 0 \text{ by Proposition ?? page ??} \\ \text{(E)} & \text{cancelled by (C)} \\ \text{(F)} & \binom{n+2}{0} = \binom{n+1}{0} = 1, \text{ so (F) is cancelled by (B)} \end{array} \\
 &= \frac{1}{(n+1)!} \sum_{m=0}^{m=n+2} (-1)^m \binom{n+2}{m} (x - m)^{n+1} \sigma(x - m) && (n + 1 \text{ case})
 \end{aligned}$$





*Example S.4.* <sup>7</sup> Let  $N_3(x)$  be the 3rd order B-spline (Definition S.2 page 371).<sup>8</sup>

**E**X 
$$N_3(x) = \frac{1}{6} \begin{cases} x^3 & \text{for } 0 \leq x \leq 1 \\ -3x^3 + 12x^2 - 12x + 4 & \text{for } 1 \leq x \leq 2 \\ 3x^3 - 24x^2 + 60x - 44 & \text{for } 2 \leq x \leq 3 \\ -x^3 + 12x^2 - 48x + 64 & \text{for } 3 \leq x \leq 4 \\ 0 & \text{otherwise} \end{cases}$$



PROOF: This expression can be calculated “by hand” using Theorem S.1 (page 375) or by using the free and open source software package *Maxima* along with the script file listed in Section ?? (page ??).  $\Rightarrow$

*Example S.5.* Let  $N_4(x)$  be the 4th order B-spline (Definition S.2 page 371).

**E**X 
$$N_4(x) = \frac{1}{24} \begin{cases} x^4 & \text{for } 0 \leq x \leq 1 \\ -4x^4 + 20x^3 - 30x^2 + 20x - 5 & \text{for } 1 \leq x \leq 2 \\ 6x^4 - 60x^3 + 210x^2 - 300x + 155 & \text{for } 2 \leq x \leq 3 \\ -4x^4 + 60x^3 - 330x^2 + 780x - 655 & \text{for } 3 \leq x \leq 4 \\ x^4 - 20x^3 + 150x^2 - 500x + 625 & \text{for } 4 \leq x \leq 5 \\ 0 & \text{otherwise} \end{cases}$$

PROOF: This expression can be calculated “by hand” using Theorem S.1 (page 375) or by using the free and open source software package *Maxima* along with the script file listed in Section ?? (page ??).  $\Rightarrow$

*Example S.6.* Let  $N_5(x)$  be the 5th order B-spline (Definition S.2 page 371).

**E**X 
$$N_5(x) = \frac{1}{120} \begin{cases} x^5 & \text{for } 0 \leq x \leq 1 \\ -5x^5 + 30x^4 - 60x^3 + 60x^2 - 30x + 6 & \text{for } 1 \leq x \leq 2 \\ 10x^5 - 120x^4 + 540x^3 - 1140x^2 + 1170x - 474 & \text{for } 2 \leq x \leq 3 \\ -10x^5 + 180x^4 - 1260x^3 + 4260x^2 - 6930x + 4386 & \text{for } 3 \leq x \leq 4 \\ 5x^5 - 120x^4 + 1140x^3 - 5340x^2 + 12270x - 10974 & \text{for } 4 \leq x \leq 5 \\ -x^5 + 30x^4 - 360x^3 + 2160x^2 - 6480x + 7776 & \text{for } 5 \leq x \leq 6 \\ 0 & \text{otherwise} \end{cases}$$

The 5th order B-spline  $N_5(x)$  is illustrated in Figure S.1 (page 373).

PROOF: This expression can be calculated “by hand” using Theorem S.1 (page 375) or by using the free and open source software package *Maxima* along with the script file listed in Section ?? (page ??).  $\Rightarrow$

*Example S.7.* Let  $N_6(x)$  be the 6th order B-spline (Definition S.2 page 371).

**E**X 
$$N_6(x) = \frac{1}{720} \begin{cases} x^6 & \text{for } 0 \leq x \leq 1 \\ -6x^6 + 42x^5 - 105x^4 + 140x^3 - 105x^2 + 42x - 7 & \text{for } 1 \leq x \leq 2 \\ 15x^6 - 210x^5 + 1155x^4 - 3220x^3 + 4935x^2 - 3990x + 1337 & \text{for } 2 \leq x \leq 3 \\ -20x^6 + 420x^5 - 3570x^4 + 15680x^3 - 37590x^2 + 47040x - 24178 & \text{for } 3 \leq x \leq 4 \\ 15x^6 - 420x^5 + 4830x^4 - 29120x^3 + 96810x^2 - 168000x + 119182 & \text{for } 4 \leq x \leq 5 \\ -6x^6 + 210x^5 - 3045x^4 + 23380x^3 - 100065x^2 + 225750x - 208943 & \text{for } 5 \leq x \leq 6 \\ x^6 - 42x^5 + 735x^4 - 6860x^3 + 36015x^2 - 100842x + 117649 & \text{for } 6 \leq x \leq 7 \\ 0 & \text{otherwise} \end{cases}$$

The 6th order B-spline  $N_6(x)$  is illustrated in Figure S.1 (page 373).

PROOF: This expression can be calculated “by hand” using Theorem S.1 (page 375) or by using the free and open source software package *Maxima* along with the script file listed in Section ?? (page ??).  $\Rightarrow$

<sup>7</sup> 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))

<sup>8</sup>For help with plotting B-splines, see APPENDIX T (page 403).

*Example S.8.* Let  $N_7(x)$  be the 7th order B-spline (Definition S.2 page 371).

**E**xample 7!  $N_7(x) = 5040N_7(x) =$

$$\left\{ \begin{array}{ll} x^7 & \text{for } 0 \leq x \leq 1 \\ -7x^7 + 56x^6 - 168x^5 + 280x^4 - 280x^3 + 168x^2 - 56x + 8 & \text{for } 1 \leq x \leq 2 \\ 21x^7 - 336x^6 + 2184x^5 - 7560x^4 + 15400x^3 - 18648x^2 + 12488x - 3576 & \text{for } 2 \leq x \leq 3 \\ -35x^7 + 840x^6 - 8400x^5 + 45360x^4 - 143360x^3 + 267120x^2 - 273280x + 118896 & \text{for } 3 \leq x \leq 4 \\ 35x^7 - 1120x^6 + 15120x^5 - 111440x^4 + 483840x^3 - 1238160x^2 + 1733760x - 1027984 & \text{for } 4 \leq x \leq 5 \\ -21x^7 + 840x^6 - 14280x^5 + 133560x^4 - 741160x^3 + 2436840x^2 - 4391240x + 3347016 & \text{for } 5 \leq x \leq 6 \\ 7x^7 - 336x^6 + 6888x^5 - 78120x^4 + 528920x^3 - 2135448x^2 + 4753336x - 4491192 & \text{for } 6 \leq x \leq 7 \\ -x^7 + 56x^6 - 1344x^5 + 17920x^4 - 143360x^3 + 688128x^2 - 1835008x + 2097152 & \text{for } 7 \leq x \leq 8 \\ 0 & \text{otherwise} \end{array} \right\}$$

The 7th order B-spline  $N_7(x)$  is illustrated in Figure S.1 (page 373).

PROOF: This expression can be calculated “by hand” using Theorem S.1 (page 375) or by using the free and open source software package *Maxima* along with the script file listed in Section ?? (page ??).  $\Rightarrow$

*Example S.9.*<sup>9</sup> The  $(n+1)^2$  coefficients of the order  $n, n-1, \dots, 0$  monomials of each B-spline  $N_n(x)$  multiplied by  $n!$  induce an *integer sequence*

$\mathbf{x} \triangleq (1, 1, 0, -1, 2, 1, 0, 0, -2, 6, -3, 1, -6, 9, 1, 0, 0, 0, -3, 12, -12, 4, 3, -24, 60, -44, -1, 12, -48, 64, \dots)$  as more fully listed in Table S.1 (page 402). In this sequence  $\mathbf{x} \triangleq (x_0, x_1, x_2, \dots)$ , the coefficients for the order  $n$  B-spline  $N_n(x)$  begin at the sequence index value

$$p \triangleq \sum_{k=0}^n k^2 = \frac{1}{6}n(n+1)(2n+1) \quad \text{and end at index value } p + (n+1)^2 - 1.$$

For example, the coefficients for  $N_3(x)$  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  $N_3(x)$ :

$$N_3(x) = \left[ \begin{array}{cccc} 1 & 0 & 0 & 0 \\ -3 & 12 & -12 & 4 \\ 3 & -24 & 60 & -44 \\ -1 & 12 & -48 & 64 \end{array} \middle| \begin{array}{l} \text{for } 0 \leq x < 1 \\ \text{for } 1 \leq x < 2 \\ \text{for } 2 \leq x < 3 \\ \text{for } 3 \leq x < 4 \end{array} \right] \left[ \begin{array}{c} x^3 \\ x^2 \\ x \\ 1 \end{array} \right] = \left\{ \begin{array}{ll} x^3 & \text{for } 0 \leq x < 1 \\ -3x^3 + 12x^2 - 12x + 4 & \text{for } 1 \leq x < 2 \\ 3x^3 - 24x^2 + 60x - 44 & \text{for } 2 \leq x < 3 \\ -x^3 + 12x^2 - 48x + 64 & \text{for } 3 \leq x < 4 \\ 0 & \text{otherwise} \end{array} \right\}$$

...which agrees with the result presented in Example S.4 (page 377).

PROOF:

1. The coefficients for the sequence  $\mathbf{x}$  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*.<sup>10</sup>
  - (a) Base case:  $n=0$  case ...

$$\begin{aligned} \sum_{k=0}^{n=0} k^2 &= 0 \\ &= \frac{0(0+1)(2 \cdot 0 + 1)}{6} \\ &= \frac{n(n+1)(2n+1)}{6} \Big|_{n=0} \end{aligned}$$

<sup>9</sup> Greenhoe (2017b)

<sup>10</sup> Greenhoe (2017a), pages 186–187 (Proposition 11.2 (Power Sums))

(b) Base case:  $n=1$  case ...

$$\begin{aligned}\sum_{k=0}^{k=1} k^2 &= 0 + 1 \\ &= \frac{1(1+1)(2 \cdot 1 + 1)}{6} \\ &= \frac{n(n+1)(2n+1)}{6} \Big|_{n=1}\end{aligned}$$

(c) inductive step—proof that  $n$  case  $\implies n+1$  case:

$$\begin{aligned}\sum_{k=0}^{n+1} k^2 &= \left( \sum_{k=0}^n k^2 \right) + (n+1)^2 \\ &= \left( \frac{n(n+1)(2n+1)}{6} \right) + (n+1)^2 && \text{by } n \text{ case hypothesis} \\ &= (n+1) \left( \frac{n(2n+1) + 6(n+1)}{6} \right) \\ &= (n+1) \left( \frac{2n^2 + 7n + 6}{6} \right) \\ &= (n+1) \left( \frac{(n+2)(2n+3)}{6} \right) \\ &= \frac{(n+1)[(n+1)+1][2(n+1)+1]}{6}\end{aligned}$$



### Theorem S.2. <sup>11</sup>

T H M	$\frac{d}{dx} N_n(x) = N_{n-1}(x) - N_{n-1}(x-1) \quad \forall n \in \mathbb{N}, \forall x \in \mathbb{R}$
-------------	--

PROOF:

1. Proof using Lemma S.2 (page 371) and the *Fundamental Theorem of Calculus*:

$$\begin{aligned}\frac{d}{dx} N_n(x) &= \frac{d}{dx} \int_0^1 N_{n-1}(x-\tau) d\tau && \text{by Lemma S.2 page 371} \\ &= \frac{d}{dx} \int_{x-u=0}^{x-u=1} N_{n-1}(u)(-1) du && \text{where } u \triangleq x - \tau \implies \tau = x - u \\ &= \frac{d}{dx} \int_{u=x-1}^{u=x} N_{n-1}(u) du \\ &= \left\{ \frac{d}{dx} \int N_{n-1}(u) du \Big|_{u=x} \right\} - \left\{ \frac{d}{dx} \int N_{n-1}(u) du \Big|_{u=x-1} \right\} && \text{by Fundamental Theorem of Calculus}^{12} \\ &= \left\{ N_{n-1}(x) \frac{d}{dx}(x) \right\} - \left\{ N_{n-1}(x-1) \frac{d}{dx}(x-1) \right\} && \text{by Chain Rule}^{13} \\ &= N_{n-1}(x) - N_{n-1}(x-1)\end{aligned}$$

<sup>11</sup> Höllig (2003) page 25 (3.2), Schumaker (2007) page 121 (Theorem 4.16)

<sup>12</sup> Hijab (2011) page 163 (Theorem 4.4.3)

<sup>13</sup> Hijab (2011) pages 73–74 (Theorem 3.1.2)

2. Proof using Lemma S.2 (page 371) and *induction*:

(a) Base case ...proof for  $n = 1$  case:

$$\begin{aligned}
 N_0(x) - N_0(x-1) &= \underbrace{\sigma(x) - \sigma(x-1)}_{N_0(x)} - \underbrace{[\sigma(x-1) - \sigma(x-2)]}_{N_0(x-1)} \quad \text{by Example S.1 page 373} \\
 &= \sigma(x) - 2\sigma(x-1) + \sigma(x-2) \\
 &= \sum_{k=0}^2 (-1)^k \binom{2}{k} \sigma(x-k) \\
 &= \frac{d}{dx} \sum_{k=0}^2 (-1)^k \binom{2}{k} (x-k) \sigma(x-k) \\
 &= \frac{d}{dx} N_1(x) \quad \text{by Example S.2 page 374}
 \end{aligned}$$

(b) Base case ...proof for  $n = 2$  case:

$$\begin{aligned}
 N_1(x) - N_1(x-1) &= \underbrace{x\sigma(x) - 2(x-1)\sigma(x-1) + (x-2)\sigma(x-2)}_{N_1(x)} \\
 &\quad - \underbrace{[(x-1)\sigma(x-1) - 2(x-2)\sigma(x-2) + (x-3)\sigma(x-3)]}_{N_1(x-1)} \quad \text{by Example S.2 page 374} \\
 &= x\sigma(x) + [-2x + 2 - x + 1]\sigma(x-1) + [x - 2 + 2x - 4]\sigma(x-2) + [-x + 3]\sigma(x-3) \\
 &= x\sigma(x) + [-3x + 3]\sigma(x-1) + [3x - 6]\sigma(x-2) + [-x + 3]\sigma(x-3) \\
 &= \frac{d}{dx} \left\{ \begin{array}{l} \frac{1}{2}x^2\sigma(x) + \left[ -\frac{3}{2}x^2 + 3x - \frac{1}{2} \right] \sigma(x-1) + \left[ \frac{3}{2}x^2 - 6x + 3 \right] \sigma(x-2) \\ \quad + \left[ -\frac{1}{2}x^2 + 3x - \frac{5}{2} \right] \sigma(x-3) \end{array} \right\} \\
 &= \frac{d}{dx} N_2(x) \quad \text{by Example S.3 page 374}
 \end{aligned}$$

(c) Proof that  $n$  case  $\implies n+1$  case:

$$\begin{aligned}
 \frac{d}{dx} N_{n+1}(x) &= \frac{d}{dx} \int_0^1 N_n(x-\tau) d\tau \quad \text{by Lemma S.2 page 371} \\
 &= \int_0^1 \frac{d}{d\tau} N_n(x-\tau) d\tau \quad \text{by Leibniz Integration Rule (Theorem ?? page ??)} \\
 &= \int_0^1 [N_{n-1}(x-\tau) - N_{n-1}(x-1-\tau)] d\tau \quad \text{by left hypothesis} \\
 &= \int_0^1 N_{n-1}(x-\tau) d\tau - \int_0^1 N_{n-1}(x-1-\tau) d\tau \\
 &= N_n(x) - N_n(x-1) \quad \text{by Lemma S.2 page 371}
 \end{aligned}$$



**Theorem S.3 (B-spline recursion).** <sup>14</sup> Let  $N_n(x)$  be the  $n$ TH ORDER B-SPLINE (Definition S.2 page 371).

T H M	$N_n(x) = \frac{x}{n} N_{n-1}(x) + \frac{n+1-x}{n} N_{n-1}(x-1) \quad \forall n \in \{1, 2, 3, \dots\}, \forall x \in \mathbb{R}$
-------------	---

<sup>14</sup> 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:

$$\begin{aligned} \frac{x}{1} N_0(x) + \frac{1+1-x}{1} N_0(x-1) &= \underbrace{\frac{x}{1} [\sigma(x) - \sigma(x-1)]}_{N_0(x)} + \underbrace{\frac{1+1-x}{1} [\sigma(x-1) - \sigma(x-2)]}_{N_0(x-1)} \\ &= x\sigma(x) + [-x - x + 2]\sigma(x-1) + [x - 2]\sigma(x-2) \\ &= N_1(x) \quad \text{by Example S.2 page 374} \end{aligned}$$

2. Induction step ...proof that  $n$  case  $\implies n+1$  case:

$$\begin{aligned} &\frac{x}{n+1} N_n(x) + \frac{n+2-x}{n+1} N_n(x-1) + c_1 \\ &= \int \frac{d}{dx} \left\{ \frac{x}{n+1} N_n(x) + \frac{n+2-x}{n+1} N_n(x-1) \right\} dx \\ &= \int \underbrace{\frac{1}{n+1} N_n(x) + \frac{x}{n+1} \frac{d}{dx} N_n(x)}_{\frac{d}{dx} \frac{x}{n+1} N_n(x)} + \underbrace{\frac{-1}{n+1} N_n(x-1) + \frac{n+2-x}{n} \frac{d}{dx} N_n(x-1)}_{\frac{d}{dx} \frac{n+2-x}{n+1} N_n(x-1)} dx \\ &\quad \text{by product rule} \\ &= \int \underbrace{\frac{1}{n+1} \left[ \frac{x}{n} N_{n-1}(x) + \frac{n+1-x}{n} N_{n-1}(x-1) \right]}_{\text{by } n \text{ hypothesis}} + \underbrace{\frac{x}{n+1} [N_{n-1}(x) - N_{n-1}(x-1)]}_{\text{by Theorem S.2 page 379}} \\ &\quad - \underbrace{\left[ \frac{x-1}{n^2+n} N_{n-1}(x-1) + \frac{n-x+2}{n(n+1)} N_{n-1}(x-2) \right]}_{\text{by induction hypothesis}} \\ &\quad + \underbrace{\frac{n+2-x}{n+1} [N_{n-1}(x-1) - N_{n-1}(x-2)]}_{\text{by Theorem S.2 page 379}} dx \\ &= \int \left[ \frac{x}{n(n+1)} + \frac{x}{n+1} \right] N_{n-1}(x) + \left[ \frac{n-x+1}{n(n+1)} - \frac{x-1}{n(n+1)} + \frac{n+2-2x}{n+1} \right] N_{n-1}(x-1) \\ &\quad + \left[ \frac{-n-2+x}{n(n+1)} + \frac{-n-2+x}{n+1} \right] N_{n-1}(x-2) dx \\ &= \int \left[ \frac{x+nx}{n(n+1)} \right] N_{n-1}(x) + \left[ \frac{n+2-2x+n(n+2-2x)}{n(n+1)} \right] N_{n-1}(x-1) \\ &\quad + \left[ \frac{-n-2+x+n(-n-2+x)}{n(n+1)} \right] N_{n-1}(x-2) dx \\ &= \int \left[ \frac{x}{n} \right] N_{n-1}(x) + \left[ \frac{n+2-2x}{n} \right] N_{n-1}(x-1) + \left[ \frac{-n-2+x}{n} \right] N_{n-1}(x-2) dx \\ &= \int \underbrace{\left[ \frac{x}{n} \right] N_{n-1}(x)}_{N_n(x)} + \underbrace{\left[ \frac{n+1-x}{n} \right] N_{n-1}(x-1)}_{N_{n-1}(x-1)} \\ &\quad - \underbrace{\left[ \frac{x-1}{n} \right] N_{n-1}(x-1) - \left[ \frac{n+2-x}{n} \right] N_{n-1}(x-2)}_{N_{n-1}(x-1)} dx \\ &= \int N_n(x) - N_n(x-1) dx \quad \text{by } n \text{ hypothesis} \\ &= \int \frac{d}{dx} N_{n+1}(x) dx \quad \text{by Theorem S.2 page 379} \\ &= N_{n+1}(x) + c_2 \end{aligned}$$

Proof that  $c_1 = c_2$ : By item (2) (page 382),  $N_n(x) = 0$  for  $x < 0$ . Therefore,  $c_1 = c_2$ .



**Theorem S.4 (B-spline general form).** <sup>15</sup> Let  $N_n(x)$  be the  $n$ TH ORDER B-SPLINE (Definition S.2 page 371). Let  $\text{supp } f$  be the SUPPORT of a function  $f \in \mathbb{R}^{\mathbb{R}}$ .

T H M	1. $N_n(x) \geq 0 \quad \forall n \in \mathbb{W}, \quad \forall x \in \mathbb{R}$ (NON-NEGATIVE) 2. $\text{supp } N_n(x) = [0 : n + 1] \quad \forall n \in \mathbb{W}$ (CLOSED SUPPORT) 3. $\int_{\mathbb{R}} N_n(x) dx = 1 \quad \forall n \in \mathbb{W}$ (UNIT AREA) 4. $N_n\left(\frac{n+1}{2} - x\right) = N_n\left(\frac{n+1}{2} + x\right) \quad \forall n \in \mathbb{N} \quad \forall x \in \mathbb{R}$ (SYMMETRIC about $x = \frac{n+1}{2}$ )
-------------	--

PROOF:

1. Proof that  $N_n(x) \geq 0$  (proof by induction):

(a) base case...proof that  $N_0(x) \geq 0$ :

$$\begin{aligned} N_0(x) &\triangleq \mathbb{1}_{[0:1]}(x) && \text{by definition of } N_0(x) && (\text{Definition S.2 page 371}) \\ &\geq 0 && \text{by definition of } \mathbb{1}_A(x) && (\text{Definition ?? page ??}) \end{aligned}$$

(b) inductive step—proof that  $\{N_n(x) \geq 0\} \implies \{N_{n+1}(x) \geq 0\}$ :

$$\begin{aligned} N_{n+1}(x) &= \int_{\tau=0}^{\tau=1} N_n(x - \tau) d\tau && \text{by Lemma S.2 page 371} \\ &\geq 0 && \text{by induction hypothesis } (N_n(x) \geq 0) \end{aligned}$$

2. Proof that  $\text{supp } N_n(x) = [0 : n + 1]$  (proof by induction):

(a) Base case ...proof that  $\text{supp } N_0 = [0 : 1]$ :

$$\begin{aligned} \text{supp } N_0 &\triangleq \text{supp } \mathbb{1}_{[0:1]} && \text{by definition of } N_0(x) && (\text{Definition S.2 page 371}) \\ &= \{[0 : 1]\}^- && \text{by definition of } \text{support operator} \\ &= [0 : 1] && \text{by definition of } \text{closure operator} \end{aligned}$$

(b) Induction step ...proof that  $\{\text{supp } N_n = [0 : n + 1]\} \implies \{\text{supp } N_{n+1} = [0 : n + 2]\}$ :

$$\begin{aligned} \text{supp } N_{n+1}(x) &= \text{supp } \int_{\tau=0}^{\tau=1} N_n(x - \tau) d\tau && \text{by Lemma S.2 page 371} \\ &= \text{supp } \int_{[0:1]} N_n(x - \tau) d\tau && \text{by def. of Lebesgue integration} \\ &= \{x \in \mathbb{R} | (x - \tau) \in [0 : n + 1] \text{ for some } \tau \in [0 : 1]\}^- && \text{by induction hypothesis} \\ &= [0 : n + 1] \cup [0 + 1 : n + 1 + 1]^- \\ &= [0 : n + 2]^- \\ &= [0 : n + 2] && \text{by property of } \text{closure operator} \end{aligned}$$

3. Proof that  $\int_{\mathbb{R}} N_n(x) dx = 1$  (proof by induction):

<sup>15</sup> 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  $\int_{\mathbb{R}} N_0(x) dx = 1$ :

$$\begin{aligned} \int_{\mathbb{R}} N_0(x) dx &= \int_{\mathbb{R}} \mathbb{1}_{[0:1]} dx && \text{by definition of } N_0(x) && (\text{Definition S.2 page 371}) \\ &= \int_{[0:1)} 1 dx && \text{by definition of } \mathbb{1}_A(x) && (\text{Definition ?? page ??}) \\ &= \int_{[0:1]} 1 dx && \text{by property of Lebesgue integration} \\ &= 1 \end{aligned}$$

(b) Induction step ...proof that  $\{\int_{\mathbb{R}} N_n(x) dx = 1\} \implies \{\int_{\mathbb{R}} N_{n+1} dx = 1\}$ :

$$\begin{aligned} \int_{\mathbb{R}} N_{n+1}(x) dx &= \int_{\mathbb{R}} \int_0^1 N_n(x - \tau) d\tau dx && \text{by Lemma S.2 page 371} \\ &= \int_0^1 \int_{\mathbb{R}} N_n(x - \tau) dx d\tau \\ &= \int_0^1 \int_{\mathbb{R}} N_n(u) du d\tau && \text{where } u \triangleq x - \tau \implies \tau = x - u \\ &= \int_0^1 1 d\tau && \text{by induction hypothesis} \\ &= 1 \end{aligned}$$

4. Proof that  $N_n(x)$  is *symmetric* for  $n \in \{1, 2, 3, \dots\}$ :

(a) Note that  $N_0(x)$  ( $n = 0$ ) is *not symmetric* (in particular it fails at  $x = 1/2$ ) because

$$N_0\left(\frac{0+1}{2} - \frac{1}{2}\right) = N_0(0) = 1 \neq 0 = N_1(1) = N_0\left(\frac{0+1}{2} + \frac{1}{2}\right)$$

(b) Base case ...proof for  $n = 1$  case:

$$\begin{aligned} N_1\left(\frac{1+1}{2} - x\right) &= N_1(1-x) \\ &= \begin{cases} (1-x) & \text{for } 1-x \in [0 : 1] \\ -(1-x)+2 & \text{for } 1-x \in [1 : 2] \\ 0 & \text{otherwise} \end{cases} && \text{by Example S.2 page 374} \\ &= \begin{cases} -x+1 & \text{for } -x \in [-1 : 0] \\ x+1 & \text{for } -x \in [0 : 1] \\ 0 & \text{otherwise} \end{cases} \\ &= \begin{cases} x+1 & \text{for } x \in [-1 : 0] \\ -x+1 & \text{for } x \in [0 : 1] \\ 0 & \text{otherwise} \end{cases} \\ &= \begin{cases} (1+x) & \text{for } 1+x \in [0 : 1] \\ -(1+x)+2 & \text{for } 1+x \in [1 : 2] \\ 0 & \text{otherwise} \end{cases} \\ &= N_1(1+x) && \text{by Example S.2 page 374} \\ &= N_1\left(\frac{1+1}{2} + x\right) \end{aligned}$$

(c) Induction step ...proof that  $n - 1$  case  $\implies n$  case:

$$\begin{aligned}
 & N_n\left(\frac{n+1}{2} + x\right) \\
 &= \frac{\frac{n+1}{2} + x}{n} N_{n-1}\left(\frac{n+1}{2} + x\right) + \frac{n+1 - \left(\frac{n+1}{2} + x\right)}{n} N_{n-1}\left(\frac{n+1}{2} + x - 1\right) \quad \text{by Theorem S.3 page 380} \\
 &= \frac{n+1 - \left(\frac{n+1}{2} - x\right)}{n} N_{n-1}\left(\frac{n}{2} + \left[x + \frac{1}{2}\right]\right) + \frac{\frac{n+1}{2} - x}{n} N_{n-1}\left(\frac{n}{2} + \left[x - \frac{1}{2}\right]\right) \\
 &= \frac{n+1 - \left(\frac{n+1}{2} - x\right)}{n} N_{n-1}\left(\frac{n}{2} - \left[x + \frac{1}{2}\right]\right) + \frac{\frac{n+1}{2} - x}{n} N_{n-1}\left(\frac{n}{2} - \left[x - \frac{1}{2}\right]\right) \quad \text{by induction hypothesis} \\
 &= \frac{n+1 - \left(\frac{n+1}{2} - x\right)}{n} N_{n-1}\left(\left[\frac{n+1}{2} - x\right] - 1\right) + \frac{\frac{n+1}{2} - x}{n} N_{n-1}\left(\frac{n+1}{2} - x\right) \\
 &= N_n\left(\frac{n+1}{2} - x\right) \quad \text{by Theorem S.3 page 380}
 \end{aligned}$$

⇒

## S.3 Projection properties

In the case where  $(N_n(x - k))_{k \in \mathbb{Z}}$  is to be used as a basis in some subspace of  $L^2_{\mathbb{R}}$ , one may want to *project* a function  $f(x)$  onto a basis function  $N_n(x - k)$ . This is especially true when  $(N_n(x - k))$  is *orthogonal*; but in the case of *B-splines* this is only true when  $n = 0$  (Theorem S.8 page 394). Nevertheless, projection of a function onto  $N_n(x - k)$ , or the projection of  $N_n(x)$  onto another basis function (such as the complex exponential in the case of *Fourier analysis* as in Lemma S.5 page 386), is still useful. Projection in an *inner product space* is typically performed using the *inner product*  $\langle f(x) | N_n(x - k) \rangle$ ; and in the space  $L^2_{\mathbb{R}}$ , this inner product is typically defined as an *integral* such that

$$\langle f(x) | N_n(x - k) \rangle \triangleq \int_{\mathbb{R}} f(x) N_n(x - k) dx.$$

As it turns out, there is a way to compute this inner product that only involves the function  $f(x)$  and the order parameter  $n$  (next theorem).

**Theorem S.5.** <sup>16</sup> Let  $f$  be a continuous function in  $L^2_{\mathbb{R}}$  and  $f^{(n)}$  the  $n$ th derivative of  $f(x)$ .

THEM

$$\begin{aligned}
 (1). \quad \int_{\mathbb{R}} f(x) N_n(x) dx &= \int_{[0:1]^{n+1}} f(x_1 + x_2 + \dots + x_{n+1}) dx_1 dx_2 \dots dx_{n+1} \\
 (2). \quad \int_{\mathbb{R}} f^{(n)}(x) N_n(x) dx &= \sum_{k=0}^{n+1} (-1)^{n-k+1} \binom{n+1}{k} f(k)
 \end{aligned}$$

PROOF:

1. Proof for (1) (proof by induction):

(a) Base case ...proof for  $n = 0$  case:

$$\int_{\mathbb{R}} f(x) N_0(x) dx = \int_{[0:1]} f(x) dx \quad \text{by definition of } N_0(x) \quad (\text{Definition S.2 page 371})$$

<sup>16</sup> 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:

$$\begin{aligned}
 & \int_{\mathbb{R}} f(x) N_{n+1}(x) dx \\
 &= \int_{\mathbb{R}} \left[ \int_0^1 N_n(x - \tau) d\tau \right] f(x) dx && \text{by Lemma S.2 page 371} \\
 &= \int_{[0:1)} \int_{\mathbb{R}} N_n(x - \tau) f(x) dx d\tau \\
 &= \int_{[0:1)} \int_{\mathbb{R}} N_n(u) f(u + \tau) du d\tau && \text{where } u \triangleq x - \tau \implies x = u + \tau \\
 &= \int_{[0:1)} \int_{[0:1)^{n+1}} f(u_1 + u_2 + \dots + u_{n+1} + \tau) du_1 du_2 \dots du_{n+1} d\tau && \text{by induction hypothesis} \\
 &= \int_{[0:1)^{n+2}} f(u_1 + u_2 + \dots + u_{n+1} + u_{n+2}) du_1 du_2 \dots du_{n+2} d\tau \\
 &= \int_{[0:1)^{n+2}} f(x_1 + x_2 + \dots + x_{n+1} + x_{n+2}) dx_1 dx_2 \dots dx_{n+2} && \text{by change of variables } u_k \rightarrow x_k
 \end{aligned}$$

2. Proof for (2):

$$\begin{aligned}
 \int_{\mathbb{R}} f^{(n)}(x) N_n(x) dx &= \int_{[0:1)^{n+1}} f^{(n)} \left( \sum_{k=1}^{n+1} x_k \right) dx_1 dx_2 \dots dx_{n+1} && \text{by (1)} \\
 &= \sum_{k=0}^{n+1} (-1)^{n-k+1} \binom{n+1}{k} f(k) && \text{by Theorem ?? page ??}
 \end{aligned}$$

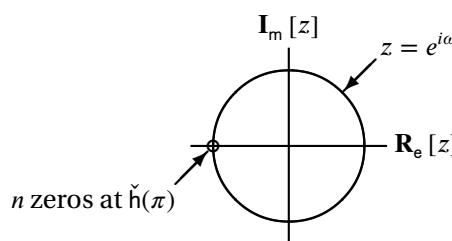


Figure S.2: Zero locations for B-cardinal spline  $N_n(x)$  scaling coefficients

## S.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:

- ➊ The complex exponential is an eigenvalue of any *linear time invariant* (LTI) operator (Theorem ?? page ??).
- ➋ The complex exponential generates a *continuous point spectrum* for the *differential operator*.

Thus, we might expect the projection of the *B-spline* function  $N_n(x)$  onto the complex exponential (essentially the *Fourier transform* of  $N_n(x)$ ,...next lemma) to be useful. Such a hunch would be confirmed because it is useful for proving that

- ☞ the sequence  $(N_n(x - k))_{k \in \mathbb{Z}}$  is a *Riesz basis* (Lemma S.6 page 389, Theorem S.8 page 394) and
- ☞ the sequence  $(N_n(x - k))_{k \in \mathbb{Z}}$  is a *multiresolution analysis* (Theorem S.10 page 397).

**Lemma S.5.** <sup>17</sup> Let  $\tilde{\mathbf{F}}$  be the FOURIER TRANSFORM operator (Definition P.2 page 337).

L E M	$\tilde{\mathbf{F}}N_n(\omega) = \frac{1}{\sqrt{2\pi}} \left( \frac{1 - e^{-i\omega}}{i\omega} \right)^{n+1} = \frac{1}{\sqrt{2\pi}} e^{-i\frac{(n+1)\omega}{2}} \left( \frac{\sin(\omega/2)}{\omega/2} \right)^{n+1} \triangleq \frac{1}{\sqrt{2\pi}} e^{-i\frac{(n+1)\omega}{2}} \left( \text{sinc} \frac{\omega}{2} \right)^{n+1}$
-------------	---

☞ PROOF:

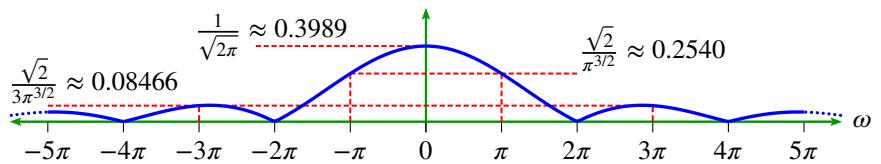
1. Proof using Theorem S.5 page 384:

$$\begin{aligned}
 \tilde{\mathbf{F}}N_n(\omega) &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} N_n(x) e^{-i\omega x} dx && \text{by definition of } \tilde{\mathbf{F}} \quad (\text{Definition P.2 page 337}) \\
 &= \frac{1}{\sqrt{2\pi}} \int_{[0:1]^{n+1}} e^{-i\omega(x_1+x_2+\dots+x_{n+1})} dx_1 dx_2 \dots dx_{n+1} && \text{by Theorem S.5} \\
 &= \frac{1}{\sqrt{2\pi}} \prod_{k=1}^{n+1} \left( \int_{[0:1]} e^{-i\omega x_k} dx_k \right) && \text{because } e^{x+y} = e^x e^y \\
 &= \frac{1}{\sqrt{2\pi}} \left( \int_0^1 e^{-i\omega x} dx \right)^{n+1} = \frac{1}{\sqrt{2\pi}} \left( \left. \frac{e^{-i\omega x}}{-i\omega} \right|_0^1 \right)^{n+1} \\
 &= \frac{1}{\sqrt{2\pi}} \left( \frac{1 - e^{-i\omega}}{i\omega} \right)^{n+1} = \frac{1}{\sqrt{2\pi}} \left[ e^{-i\frac{\omega}{2}} \left( \frac{e^{i\frac{\omega}{2}} - e^{-i\frac{\omega}{2}}}{i\omega} \right) \right]^{n+1} \\
 &= \frac{1}{\sqrt{2\pi}} \left[ e^{-i\frac{\omega}{2}} \left( \frac{2i \sin\left(\frac{\omega}{2}\right)}{\frac{2i\omega}{2}} \right) \right]^{n+1} && \text{by Euler formulas} \quad (\text{Corollary ?? page ??}) \\
 &= \frac{1}{\sqrt{2\pi}} e^{-i\frac{(n+1)\omega}{2}} \left( \frac{\sin(\omega/2)}{\omega/2} \right)^{n+1}
 \end{aligned}$$

2. Proof using *rectangular pulse* example (Example P.1 page 344) and *Convolution Theorem* (Theorem R.2 page 360):

$$\begin{aligned}
 \tilde{\mathbf{F}}N_n(\omega) &= \left[ \sqrt{2\pi} \right]^n [\tilde{\mathbf{F}}N_0]^{n+1} && \text{by Convolution Theorem} \quad (\text{Theorem R.2 page 360}) \\
 &= \frac{1}{\sqrt{2\pi}} \left[ \sqrt{2\pi} \frac{c(b-a)}{\sqrt{2\pi}} e^{-i\left(\frac{a+b}{2}\omega\right)} \left( \frac{\sin\left(\frac{b-a}{2}\omega\right)}{\left(\frac{b-a}{2}\omega\right)} \right) \right]^{n+1} && \text{by rectangular pulse example} \\
 &= \frac{1}{\sqrt{2\pi}} \left[ \sqrt{2\pi} \frac{1}{\sqrt{2\pi}} e^{-i\left(\frac{\omega}{2}\right)} \left( \frac{\sin\left(\frac{\omega}{2}\right)}{(\omega/2)} \right) \right]^{n+1} && \text{with } a = 0, b = c = 1 \quad (\text{Example P.1 page 344}) \\
 &= \frac{1}{\sqrt{2\pi}} e^{-i\left(\frac{(n+1)\omega}{2}\right)} \left( \frac{\sin\left(\frac{\omega}{2}\right)}{(\omega/2)} \right)^{n+1}
 \end{aligned}$$

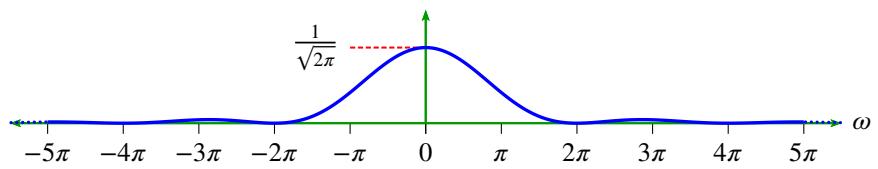
**Example S.10.** The Fourier transform magnitude  $|\tilde{\mathbf{F}}N_0](\omega)|$  of the 0 order B-spline  $N_0(x)$  is illustrated to the right.



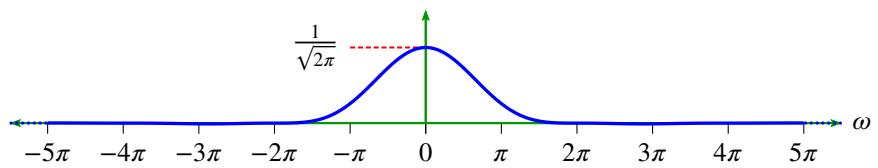
<sup>17</sup>☞ Christensen (2008) page 142 (Corollary 6.1.2)



*Example S.11.* The Fourier transform magnitude  $|[\tilde{F}N_1](\omega)|$  of the 1st order B-spline  $N_1(x)$  is illustrated to the right.



*Example S.12.* The Fourier transform magnitude  $|[\tilde{F}N_2](\omega)|$  of the 2nd order B-spline  $N_2(x)$  is illustrated to the right.



## S.5 Basis properties

### S.5.1 Uniqueness properties

Coefficients of a *basis sequence* are not always *unique*. Take for example a very trivial sequence  $(\alpha_1, \alpha_2)$  in which the coefficients are summed. If  $f(x) \triangleq \alpha_1 + \alpha_2$  and  $g(x) \triangleq \beta_1 + \beta_2$ ,

$$\begin{aligned} \text{then } \{(\alpha_1, \alpha_2) = (\beta_1, \beta_2)\} &\implies f(x) = g(x) \\ \text{but } f(x) = g(x) &\implies \{(\alpha_1, \alpha_2) = (\beta_1, \beta_2)\}, \end{aligned}$$

because for example if  $(\alpha_1, \alpha_2) = (1, 2)$  and  $(\beta_1, \beta_2) = (-6, 9)$ , then  $f(x) = g(x)$ , 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 sequences*. 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  $(N_n(x - k))_{k \in \mathbb{Z}}$ , the coefficients  $(\alpha_k)_{k \in \mathbb{Z}}$  are *unique*, as demonstrated by Theorem S.6 (next).

**Theorem S.6.** <sup>18</sup> Let  $N_n(x)$  be the *n*TH-ORDER B-SPLINE (Definition S.2 page 371). Let

$$f(x) \triangleq \sum_{k \in \mathbb{Z}} \alpha_k N_n(x - k) \quad \text{and} \quad g(x) \triangleq \sum_{k \in \mathbb{Z}} \beta_k N_n(x - k).$$

T H M	$\{ f(x) = g(x) \quad \forall x \in \mathbb{R} \} \iff \{ (\alpha_k)_{k \in \mathbb{Z}} = (\beta_k)_{k \in \mathbb{Z}} \}$
coefficients are UNIQUE	

PROOF:

1. Proof that  $\iff$  condition holds:

$$\begin{aligned} f(x) &\triangleq \sum_{k \in \mathbb{Z}} \alpha_k N_n(x - k) && \text{by definition of } f(x) \\ &= \sum_{k \in \mathbb{Z}} \beta_k N_n(x - k) && \text{by right hypothesis} \\ &\triangleq g(x) && \text{by definition of } g(x) \end{aligned}$$

2. Proof that  $\implies$  condition holds (proof by contradiction):

(a) Suppose it does *not* hold.

<sup>18</sup> 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  
 $(\alpha_k) - (\beta_k) \triangleq (\alpha_k - \beta_k) \neq (0, 0, 0, \dots)$   
but also such that  $f(x) - g(x) = 0 \forall x \in \mathbb{R}$ .

- (c) If this were possible, then

$$\begin{aligned} 0 &= f(x) - g(x) \\ &= \sum_{m \in \mathbb{Z}} \alpha_m N_n(x-m) - \sum_{m \in \mathbb{Z}} \beta_m N_n(x-m) \\ &= \sum_{m \in \mathbb{Z}} (\alpha_m - \beta_m) N_n(x-m) \\ &= \sum_{m=0}^{m=n} (\alpha_m - \beta_m) \frac{1}{n!} \left[ \sum_{k=0}^{n+1} (-1)^k \binom{n+1}{k} (x-k)^n \sigma(x-k) \right] \end{aligned} \quad \text{by Theorem S.1 page 375}$$

- (d) But this is *impossible* because  $N(x)$  is *non-negative* (Theorem S.4 page 382).  
(e) Therefore, there is a contradiction, and the  $\Rightarrow$  condition *does* hold.

⇒

## S.5.2 Partition of unity properties

In the case in which a sequence of *B-splines*  $(N_n(x-k))_{k \in \mathbb{Z}}$  is to be used as a *basis* for some subspace of  $L^2_{\mathbb{R}}$ , arguably one of the most important properties for the sequence to have is the *partition of unity* property such that  $\sum_{k \in \mathbb{Z}} N_n(x-k) = 1$ . This allows for convenient representation of the most basic functions, such as constants.<sup>19</sup> As it turns out, B-splines *do* have this property (next theorem).

**Theorem S.7 (B-spline partition of unity).** <sup>20</sup> Let  $N_n(x)$  be the *n*TH ORDER B-SPLINE (Definition S.2 page 371).

T H M	$\sum_{k \in \mathbb{Z}} N_n(x-k) = 1 \quad \forall n \in \mathbb{W}$	(PARTITION OF UNITY)
-------------	---	----------------------

PROOF:

1. lemma:  $\sum_{k \in \mathbb{Z}} N_0(x-k) = 1$ . Proof:

$$\begin{aligned} \sum_{k \in \mathbb{Z}} N_0(x-k) &= \sum_{k \in \mathbb{Z}} \mathbb{1}_{[0:1]}(x-k) && \text{by definition of } N_0(x) && \text{(Definition S.2 page 371)} \\ &= 1 && \text{by definition of } \mathbb{1}_A(x) && \text{(Definition ?? page ??)} \end{aligned}$$

2. Proof for this theorem follows from the  $n = 0$  case ((1) lemma page 388), the definition of  $N_n(x)$  (Definition S.2 page 371), and Corollary ?? (page ??).

3. Alternatively, this theorem can be proved by *induction*:

- (a) Base case ( $n = 0$  case): by (1) lemma.

<sup>19</sup> Jawerth and Sweldens (1994) page 8

<sup>20</sup> 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)

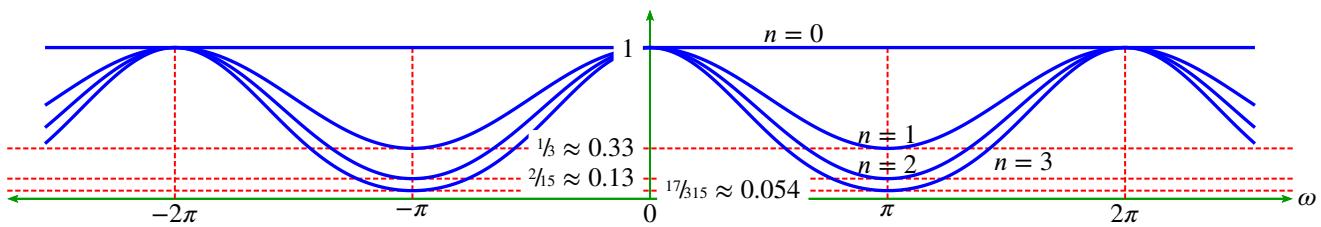


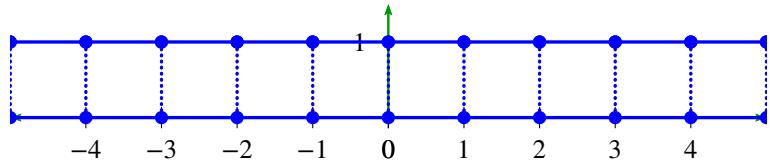
Figure S.3: *auto-power spectrum*  $\tilde{S}_n(\omega)$  plots of *B-splines*  $N_n(x)$  (Lemma S.6 page 389) For C and L<sup>A</sup>T<sub>E</sub>X source code to generate such a plot, see Section ?? (page ??).

(b) Inductive step—proof that  $\sum_{k \in \mathbb{Z}} N_n(x - k) = 1 \implies \sum_{k \in \mathbb{Z}} N_{n+1}(x - k) = 1$ :

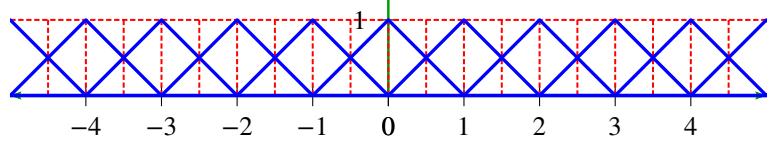
$$\begin{aligned}
 \sum_{k \in \mathbb{Z}} N_{n+1}(x - k) &= \sum_{k \in \mathbb{Z}} \int_{\tau=0}^{\tau=1} N_n(x - k - \tau) d\tau && \text{by Lemma S.2 page 371} \\
 &= \sum_{k \in \mathbb{Z}} \int_{x-u=0}^{x-u=1} N_n(u - k)(-1) du && \text{where } u \triangleq x - \tau \implies \tau = x - u \\
 &= \sum_{k \in \mathbb{Z}} \int_{u=x-1}^{u=x} N_n(u - k) du \\
 &= \int_{u=x-1}^{u=x} \left( \sum_{k \in \mathbb{Z}} N_n(u - k) \right) du \\
 &= \int_{u=x-1}^{u=x} 1 du && \text{by induction hypothesis} \\
 &= 1
 \end{aligned}$$

⇒

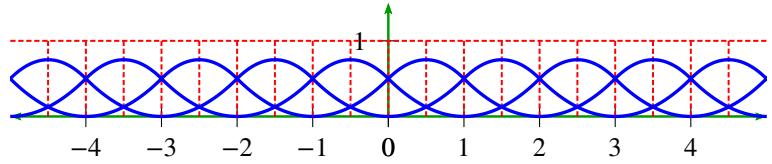
*Example S.13.* The *partition of unity* property for the *0 order B-spline*  $N_0(x)$  (Example S.1 page 373) is illustrated to the right.



*Example S.14.* The *partition of unity* property for the *1st order B-spline*  $N_1(x)$  (Example S.2 page 374) is illustrated to the right.



*Example S.15.* The *partition of unity* property for the *2nd order B-spline*  $N_2(x)$  (Example S.3 page 374) is illustrated to the right.



### S.5.3 Riesz basis properties

**Lemma S.6.** Let  $N_n(x)$  be the *n*th ORDER B-SPLINE (Definition S.2 page 371).

Let  $\tilde{S}_n(\omega) \triangleq 2\pi \sum_{k \in \mathbb{Z}} |\tilde{F}N_n(\omega - 2\pi k)|^2$  be the AUTO-POWER SPECTRUM (Definition ?? page ??) of  $N_n(x)$ .

LEM	(1). $0 < \tilde{S}_n(\omega) \leq 1 \quad \forall \omega \in \mathbb{R} \quad , \quad \forall n \in \mathbb{W}$ (2). $\tilde{S}_n(\omega) = 1 \quad \forall \omega \in \mathbb{R} \quad , \quad \text{for } n = 0$	(3). $\tilde{S}_n(0) = 1 \quad \forall n \in \mathbb{W}$ (4). $\tilde{S}_n(\pi) \leq \frac{1}{3} \quad \forall n \in \mathbb{W} \setminus \{0\}$	(Note: see illustration in Figure S.3 page 389.)
-----	--	---	---

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:

$$\begin{aligned}
 \tilde{S}_n(\omega) &\triangleq 2\pi \sum_{k \in \mathbb{Z}} |\tilde{\mathbf{F}}\mathbf{N}_n(\omega - 2\pi k)|^2 && \text{by Definition ?? page ??} \\
 &= 2\pi \sum_{k \in \mathbb{Z}} \left| \frac{1}{\sqrt{2\pi}} e^{-i\frac{(n+1)(\omega-2\pi k)}{2}} \left( \frac{\sin\left(\frac{\omega-2\pi k}{2}\right)}{\frac{\omega-2\pi k}{2}} \right)^{n+1} \right|^2 && \text{by Lemma S.5 page 386} \\
 &= \sum_{k \in \mathbb{Z}} \left[ \frac{\sin\left(\frac{\omega-2\pi k}{2}\right)}{\frac{\omega-2\pi k}{2}} \right]^{2(n+1)} \\
 &= \sum_{k \in \mathbb{Z}} \left[ \frac{\sin\left(\frac{\omega}{2} - k\pi\right)}{k\pi - \frac{\omega}{2}} \right]^{2(n+1)} \\
 &= \sum_{k \in \mathbb{Z}} \left[ \frac{(-1)^k \sin\left(\frac{\omega}{2}\right)}{k\pi - \frac{\omega}{2}} \right]^{2(n+1)} \\
 &= \sum_{k \in \mathbb{Z}} \left[ \frac{\sin\left(\frac{\omega}{2}\right)}{k\pi - \frac{\omega}{2}} \right]^{2(n+1)}
 \end{aligned}$$

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  $2\pi$  (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  $\pi$ ). Proof: This follows directly from Proposition ?? (page ??).



6. Proof that  $\tilde{S}_n(0) = 1$ :

$$\begin{aligned}
 \tilde{S}_n(0) &= \lim_{\omega \rightarrow 0} \tilde{S}_n(\omega) && \text{by (3) lemma} \\
 &= \lim_{\omega \rightarrow 0} \left[ \left[ \frac{\sin(\omega/2)}{(\omega/2)} \right]^{2(n+1)} + \left[ \frac{\sin(\omega/2)}{(\pi/2)} \right]^{2(n+1)} \left( \sum_{k=1}^{\infty} \left[ \frac{1}{2k - \frac{\omega}{\pi}} \right]^{2(n+1)} + \sum_{k=1}^{\infty} \left[ \frac{1}{2k + \frac{\omega}{\pi}} \right]^{2(n+1)} \right) \right] && \text{by (2) lemma} \\
 &= \lim_{\omega \rightarrow 0} \left[ \frac{\cos\left(\frac{\omega}{2}\right)}{-\frac{1}{2}} \right]^{2(n+1)} + 0 && \text{by l'Hôpital's rule} \\
 &= (-1)^{2(n+1)} = 1
 \end{aligned}$$

7. Proof that  $\tilde{S}_n(\pi)$  converges to some value  $> 0$ :

(a) Proof that  $\tilde{S}_n(\pi) > 0$ :

$$\begin{aligned}
 \tilde{S}_n(\pi) &= \left[ \frac{\sin(\pi/2)}{(\pi/2)} \right]^{2(n+1)} + \left[ \frac{\sin(\pi/2)}{(\pi/2)} \right]^{2(n+1)} \left( \sum_{k=1}^{\infty} \left[ \frac{1}{2k - \frac{\pi}{\pi}} \right]^{2(n+1)} + \sum_{k=1}^{\infty} \left[ \frac{1}{2k + \frac{\pi}{\pi}} \right]^{2(n+1)} \right) && \text{by (2) lemma} \\
 &= \left( \frac{2}{\pi} \right)^{2(n+1)} \left[ 1 + \left( \frac{1}{1} \right)^{2(n+1)} + \left( \frac{1}{3} \right)^{2(n+1)} + \left( \frac{1}{3} \right)^{2(n+1)} + \left( \frac{1}{5} \right)^{2(n+1)} + \left( \frac{1}{5} \right)^{2(n+1)} + \dots \right] \\
 &= 2 \left( \frac{2}{\pi} \right)^{2(n+1)} \underbrace{\sum_{k=1}^{\infty} \left[ \frac{1}{2k-1} \right]^{2(n+1)}}_{\text{Dirichlet Lambda function } \lambda(2n+2)} \\
 &> 0 && \text{because } x^2 > 0 \text{ for all } x \in \mathbb{R} \setminus \{0\}
 \end{aligned}$$

(b) Proof that  $\tilde{S}_n(\pi)$  converges:

$$\begin{aligned}
 \tilde{S}_n(\pi) &= 2 \left( \frac{2}{\pi} \right)^{2(n+1)} \sum_{k=1}^{\infty} \left( \frac{1}{2k-1} \right)^{2(n+1)} && \text{by item (7a)} \\
 &\leq 2 \left( \frac{2}{\pi} \right)^{2(n+1)} \sum_{k=1}^{\infty} \left( \frac{1}{k} \right)^{2(n+1)} \\
 &\leq 2 \left( \frac{2}{\pi} \right)^{2(n+1)} \sum_{k=1}^{\infty} \left( \frac{1}{k} \right)^2 \\
 &\implies \text{convergence} && \text{by comparison test}
 \end{aligned}$$

(c) Tighter bounds for  $\tilde{S}_n(\pi)$  for certain values of  $n \in \{0, 1, 2, 3, 4\}$ :

$$\begin{aligned}
 \tilde{S}_n(\pi) &= 2 \left( \frac{2}{\pi} \right)^{2(n+1)} \sum_{k=1}^{\infty} \left( \frac{1}{2k-1} \right)^{2(n+1)} && \text{by item (7a)} \\
 &= 2 \left( \frac{2}{\pi} \right)^{2(n+1)} U_{2(n+1)} && \text{by } \text{Jolley (1961), pages 56–57 ((307))} \\
 &= 2 \left( \frac{2}{\pi} \right)^{2(n+1)} \left[ \frac{\pi^{2(n+1)} \alpha_{n+1}}{(4)[(2n+2)!]} \right] && \text{by } \text{Jolley (1961), pages 56–57 ((307))} \\
 &= \frac{2^{2n+1} \alpha_{n+1}}{(2n+2)!}
 \end{aligned}$$

$$\begin{aligned}
 &= \left\{ \begin{array}{ll} \frac{2^1(1)}{2!} & \text{for } n = 0 \quad (\alpha_1 = 1) \\ \frac{2^2(1)}{4!} & \text{for } n = 1 \quad (\alpha_2 = 1) \\ \frac{2^5(3)}{6!} & \text{for } n = 2 \quad (\alpha_3 = 3) \\ \frac{2^7(17)}{8!} & \text{for } n = 3 \quad (\alpha_4 = 17) \\ \frac{2^9(155)}{10!} & \text{for } n = 4 \quad (\alpha_5 = 155) \end{array} \right\} \quad \text{by } \text{Jolley (1961), page 234 (1130)} \\
 &= \left\{ \begin{array}{ll} 1 & \text{for } n = 0 \\ \frac{1}{3} & \text{for } n = 1 \\ \frac{2}{15} & \text{for } n = 2 \\ \frac{17}{315} & \text{for } n = 3 \\ \frac{62}{2835} & \text{for } n = 4 \end{array} \right\} = \left\{ \begin{array}{ll} 1 & \text{for } n = 0 \\ 0.3333333333333333 \dots & \text{for } n = 1 \\ 0.1333333333333333 \dots & \text{for } n = 2 \\ 0.0539682539682 \dots & \text{for } n = 3 \\ 0.0218694885361 \dots & \text{for } n = 4 \end{array} \right\}
 \end{aligned}$$

(d) Being important for the  $n = 0$  case, note that<sup>21</sup>

$$\underbrace{\sum_{k=1}^{\infty} \left( \frac{1}{2k-1} \right)^2}_{\text{Dirichlet Lambda function } \lambda(2)} = 1 + \frac{1}{3^2} + \frac{1}{5^2} + \frac{1}{7^2} + \dots = \frac{\pi^2}{8}$$

(e) Proof that  $\tilde{S}_n(\pi) \leq \frac{1}{3}$ : because  $\tilde{S}_n(\pi) = \frac{1}{3}$  for  $n = 1$  (item (7c) page 391) 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:  $\sum_{k=1}^{\infty} \left[ \frac{1}{2k \pm \frac{\omega}{\pi}} \right]^{2(n+1)}$  converges. Proof:

$$\begin{aligned}
 \lim_{b \rightarrow \infty} \int_1^b \left[ \frac{1}{2y \pm \frac{\omega}{\pi}} \right]^{2(n+1)} dy &= \lim_{b \rightarrow \infty} \int_1^b \left[ 2y \pm \frac{\omega}{\pi} \right]^{-2n-2} dy \\
 &= \lim_{b \rightarrow \infty} \frac{\left[ 2y \pm \frac{\omega}{\pi} \right]^{-2n-1}}{2(-2n-1)} \Big|_1^b \\
 &= \lim_{b \rightarrow \infty} \left( \frac{-1}{2(2n+1)} \right) \left[ \frac{1}{\left[ 2b \pm \frac{\omega}{\pi} \right]^{2n+1}} - \frac{1}{\left[ 2 \pm \frac{\omega}{\pi} \right]^{2n+1}} \right]
 \end{aligned}$$

$$\begin{aligned}
 &= 0 + \frac{1}{2(2n+1) \left[ 2 \pm \frac{\omega}{\pi} \right]^{2n+1}} \\
 &< \infty \quad \forall \omega \in [0 : 2\pi)
 \end{aligned}$$

$$\Rightarrow \sum_{k=1}^{\infty} \left[ \frac{1}{2k \pm \frac{\omega}{\pi}} \right]^{2(n+1)} \text{ converges} \quad \text{by integral test}$$

<sup>21</sup> [Nahin \(2011\) page 153](#), [Bailey et al. \(2013\) page 334](#) (Catalan's Constant), [Bailey et al. \(2011\) \(15\)](#), [Wells \(1987\) page 36](#) (Dictionary entry for  $\pi$ : pages 31–37), [Heinbockel \(2010\) page 94](#) ((2.27) Dirichlet Lambda function)

ii. completion of proof using (8(c)i) lemma ...

$$\begin{aligned}\tilde{S}_n(\omega) &= \left[ \frac{\sin(\omega/2)}{(\omega/2)} \right]^{2(n+1)} + \left[ \frac{\sin(\omega/2)}{(\pi/2)} \right]^{2(n+1)} \left( \sum_{k=1}^{\infty} \left[ \frac{1}{2k - \frac{\omega}{\pi}} \right]^{2(n+1)} + \sum_{k=1}^{\infty} \left[ \frac{1}{2k + \frac{\omega}{\pi}} \right]^{2(n+1)} \right) \text{ by (2) lemma} \\ &\implies \tilde{S}_n(\omega) \text{ converges } \forall \omega \in (0 : 2\pi) \quad \text{by (8(c)i) lemma}\end{aligned}$$

9. lemma (an expression for  $\tilde{S}'_n(\omega)$ ):

$$\begin{aligned}\tilde{S}'_n(\omega) &\triangleq \frac{d}{d\omega} \tilde{S}_n(\omega) \\ &= \frac{d}{d\omega} \sum_{k \in \mathbb{Z}} \left[ \frac{\sin\left(\frac{\omega}{2}\right)}{k\pi - \frac{\omega}{2}} \right]^{2(n+1)} \quad \text{by (1) lemma page 390} \\ &= \sum_{k \in \mathbb{Z}} \frac{d}{d\omega} \left[ \frac{\sin\left(\frac{\omega}{2}\right)}{k\pi - \frac{\omega}{2}} \right]^{2(n+1)} \quad \text{by linearity of } \frac{d}{d\omega} \text{ operator} \\ &= \sum_{k \in \mathbb{Z}} 2(n+1) \left[ \frac{\sin\left(\frac{\omega}{2}\right)}{k\pi - \frac{\omega}{2}} \right]^{2n+1} \frac{d}{d\omega} \left[ \frac{\sin\left(\frac{\omega}{2}\right)}{k\pi - \frac{\omega}{2}} \right] \quad \text{by power rule} \\ &= 2(n+1) \sum_{k \in \mathbb{Z}} \left[ \frac{\sin\left(\frac{\omega}{2}\right)}{k\pi - \frac{\omega}{2}} \right]^{2n+1} \left[ \frac{\frac{1}{2} \cos\left(\frac{\omega}{2}\right) \left(k\pi - \frac{\omega}{2}\right) - \sin\left(\frac{\omega}{2}\right) \left(-\frac{1}{2}\right)}{\left(k\pi - \frac{\omega}{2}\right)^2} \right] \quad \text{by quotient rule} \\ &= (n+1) \sum_{k \in \mathbb{Z}} \left[ \frac{\sin\left(\frac{\omega}{2}\right)}{k\pi - \frac{\omega}{2}} \right]^{2n+1} \left[ \frac{\cos\left(\frac{\omega}{2}\right) \left(k\pi - \frac{\omega}{2}\right) + \sin\left(\frac{\omega}{2}\right)}{\left(k\pi - \frac{\omega}{2}\right)^2} \right]\end{aligned}$$

10. lemma:  $\tilde{S}'_n(0) = \tilde{S}'_n(\pi) = 0$ . Proof: This follows from Proposition ?? (page ??). Here is alternate proof:

$$\begin{aligned}\tilde{S}'_n(0) &= \lim_{\omega \rightarrow 0} \tilde{S}'_n(\omega) \\ &= \lim_{\omega \rightarrow 0} (n+1) \sum_{k \in \mathbb{Z}} \left[ \frac{\sin\left(\frac{\omega}{2}\right)}{k\pi - \frac{\omega}{2}} \right]^{2n+1} \left[ \frac{\cos\left(\frac{\omega}{2}\right) \left(k\pi - \frac{\omega}{2}\right) + \sin\left(\frac{\omega}{2}\right)}{\left(k\pi - \frac{\omega}{2}\right)^2} \right] \quad \text{by (9) lemma} \\ &= \lim_{\omega \rightarrow 0} (n+1) \left[ \frac{\sin\left(\frac{\omega}{2}\right)}{-\frac{\omega}{2}} \right]^{2n+1} \left[ \frac{\cos\left(\frac{\omega}{2}\right) \left(-\frac{\omega}{2}\right) + \sin\left(\frac{\omega}{2}\right)}{\left(-\frac{\omega}{2}\right)^2} \right] \\ &= (n+1) \lim_{\omega \rightarrow 0} \left[ \frac{\sin\left(\frac{\omega}{2}\right)}{-\frac{\omega}{2}} \right]^{2n+1} \lim_{\omega \rightarrow 0} \left[ \frac{\cos\left(\frac{\omega}{2}\right) \left(-\frac{\omega}{2}\right) + \sin\left(\frac{\omega}{2}\right)}{\left(-\frac{\omega}{2}\right)^2} \right] \\ &= (n+1) [-1]^{2n+1} \lim_{\omega \rightarrow 0} \left[ \frac{-\frac{1}{2} \sin\left(\frac{\omega}{2}\right) \left(-\frac{\omega}{2}\right) + \cos\left(\frac{\omega}{2}\right) \left(-\frac{1}{2}\right) + \cos\left(\frac{\omega}{2}\right) \left(\frac{1}{2}\right)}{-\frac{2}{2} \left(-\frac{\omega}{2}\right)} \right] \quad \text{by l'Hôpital's rule} \\ &= (1)(0) \\ &= 0\end{aligned}$$

$$\begin{aligned}
\tilde{S}'_n(\pi) &= (n+1) \sum_{k \in \mathbb{Z}} \left[ \frac{\sin\left(\frac{\pi}{2}\right)}{k\pi - \frac{\pi}{2}} \right]^{2n+1} \left[ \frac{\cos\left(\frac{\pi}{2}\right)\left(k\pi - \frac{\pi}{2}\right) + \sin\left(\frac{\pi}{2}\right)}{\left(k\pi - \frac{\pi}{2}\right)^2} \right] \\
&= (n+1) \sum_{k \in \mathbb{Z}} \left[ \frac{1}{k\pi - \frac{\pi}{2}} \right]^{2n+1} \left[ \frac{0\left(k\pi - \frac{\pi}{2}\right) + 1}{\left(k\pi - \frac{\pi}{2}\right)^2} \right] \\
&= (n+1) \left( \frac{2}{\pi} \right)^{2n+3} \sum_{k \in \mathbb{Z}} \left[ \frac{1}{2k-1} \right]^{2n+3} \\
&= (n+1) \left( \frac{2}{\pi} \right)^{2n+3} \left[ \left( \frac{1}{1} \right)^{2n+3} + \left( \frac{1}{-1} \right)^{2n+3} + \left( \frac{1}{3} \right)^{2n+3} + \left( \frac{1}{-3} \right)^{2n+3} + \dots \right] \\
&= (n+1) \left( \frac{2}{\pi} \right)^{2n+3} \sum_{k=1}^{\infty} (-1)^{k+1} \alpha_k \quad \text{where } \alpha_k \triangleq \begin{cases} \left( \frac{1}{k} \right)^{2n+3} & \text{for } k \text{ odd} \\ \left( \frac{1}{k-1} \right)^{2n+3} & \text{for } k \text{ even} \end{cases} \\
&= 0 \quad \text{because } \lim_{k \rightarrow \infty} \alpha_k = 0 \text{ and by Alternating Series Test}
\end{aligned}$$

11. lemma:  $\tilde{S}_n(\omega)$  is *decreasing* with respect to  $\omega \in [0 : \pi]$ . Proof:

$$\begin{aligned}
\tilde{S}'_n(\omega) &= (n+1) \sum_{k \in \mathbb{Z}} \left[ \frac{\sin\left(\frac{\omega}{2}\right)}{k\pi - \frac{\omega}{2}} \right]^{2n+1} \left[ \frac{\cos\left(\frac{\omega}{2}\right)\left(k\pi - \frac{\omega}{2}\right) + \sin\left(\frac{\omega}{2}\right)}{\left(k\pi - \frac{\omega}{2}\right)^2} \right] \quad \text{by (9) lemma page 393} \\
&= (n+1) \underbrace{\left( \sin \frac{\omega}{2} \right)^{2n+1}}_{\geq 0 \text{ for } \omega \in [0 : 2\pi]} \sum_{k \in \mathbb{Z}} \underbrace{\left[ \frac{1}{k\pi - \frac{\omega}{2}} \right]^{2n+2}}_{> 0} \left[ \underbrace{\left( \cos \frac{\omega}{2} \right)}_{\text{sign change at } \omega = \pi} + \underbrace{\frac{\sin \frac{\omega}{2}}{k\pi - \frac{\omega}{2}}}_{\substack{\text{decreasing w.r.t. } \omega \in \mathbb{R} \\ > 0 \text{ for } \omega \in (0 : 2\pi)}} \right]
\end{aligned}$$

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 S.8. <sup>22</sup>

T H M	1. $(N_n(x-k))_{k \in \mathbb{Z}}$ is a RIESZ BASIS	<i>for</i> $\text{span}(N_n(x-k))_{k \in \mathbb{Z}}$	$\forall n \in \mathbb{W}$
	2. $(N_n(x-k))_{k \in \mathbb{Z}}$ is an ORTHONORMAL BASIS	<i>for</i> $\text{span}(N_n(x-k))_{k \in \mathbb{Z}}$	$\iff n = 0$

PROOF:

<sup>22</sup> Wojaszczyk (1997) page 56 (Proposition 3.12), Prasad and Iyengar (1997) page 148 (Theorem 6.3), Forster and Massopust (2009) page 66 (Theorem 2.25)



1. Proof that  $(N_n(x - k))_{k \in \mathbb{Z}}$  is a *Riesz basis* for  $\text{span}(N_n(x - k))_{k \in \mathbb{Z}}$ :

$$\begin{aligned} 0 < \tilde{S}_n(\omega) &\leq 1 && \text{by Lemma S.6 page 389 (1)} \\ \implies (N_n(x - k))_{k \in \mathbb{Z}} &\text{ is a } Riesz \text{ basis for } \text{span}(N_n(x - k))_{k \in \mathbb{Z}} && \text{by Theorem ?? page ??} \end{aligned}$$

2. Proof that  $\{n = 0\} \iff (N_n(x - k))_{k \in \mathbb{Z}}$  is an *orthonormal basis* for  $\text{span}(N_n(x - k))_{k \in \mathbb{Z}}$ :

$$\begin{aligned} n = 0 \iff \tilde{S}_n(\omega) &= 1 && \text{by Lemma S.6 page 389 (2), (4)} \\ \iff (N_n(x - k))_{k \in \mathbb{Z}} &\text{ is an orthonormal basis for } \text{span}(N_n(x - k)) && \text{by Theorem ?? page ??} \end{aligned}$$



## S.6 Mutiresolution properties

### S.6.1 Introduction

In 1989, Stéphane G. Mallat introduced the *Mutiresolution 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.<sup>23</sup>

### S.6.2 B-spline dyadic decomposition

One key feature of an MRA is *dyadic decomposition* such that  $N_n(x) = \sum_k \alpha_n N_n(2x - k)$  for some sequence  $(\alpha_n)$ . As it turns out, *B-splines* also have this property (next theorem).

**Theorem S.9** (*B-spline dyadic decomposition*).<sup>24</sup> Let  $N_n(x)$  be the  $n$ th ORDER B-SPLINE.

T H M	$N_n(x) = \frac{1}{2^n} \sum_{k=0}^{n+1} \binom{n+1}{k} N_n(2x - k) \quad \forall n \in \mathbb{W}, \forall x \in \mathbb{R}$
-------------	---

PROOF:

1. Base case ...proof for  $n = 0$  case:

$$\begin{aligned} N_0(x) &= \mathbb{1}_{[0:1]}(x) && \text{by definition of } \mathbb{1}_A(x) \quad (\text{Definition ?? page ??}) \\ &= \mathbb{1}_{[0:\frac{1}{2}]}(x) + \mathbb{1}_{[\frac{1}{2}:1]}(x) \\ &= \mathbb{1}_{[2x0:2x\frac{1}{2}]}(2x) + \mathbb{1}_{[2x\frac{1}{2}-1:2x1-1]}(2x - 1) \\ &= \mathbb{1}_{[0:1]}(2x) + \mathbb{1}_{[0:1]}(2x - 1) \\ &= \frac{1}{2^0} \sum_{k=0}^{0+1} \binom{0+1}{k} N_0(2x - k) \end{aligned}$$

<sup>23</sup> Mallat (1999) page 240, Definition ?? (page ??)

<sup>24</sup> Prasad and Iyengar (1997) pages 151–152 (proof using Fourier transform)

2. Induction step...proof that  $n$  case  $\implies n + 1$  case:

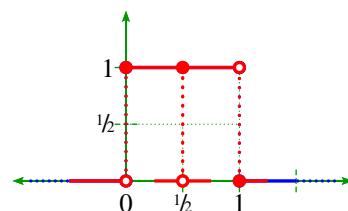
$$\begin{aligned}
 N_{n+1}(x) &= \int_0^1 N_n(x - \tau) d\tau && \text{by Lemma S.2 page 371} \\
 &= \int_0^1 \frac{1}{2^n} \sum_{k=0}^{n+1} \binom{n+1}{k} N_n(2x - 2\tau - k) d\tau && \text{by induction hypothesis} \\
 &= \frac{1}{2^n} \sum_{k=0}^{n+1} \binom{n+1}{k} \int_{\tau=0}^{\tau=1} N_n(2x - 2\tau - k) d\tau && \text{by linearity of } \sum \text{ operator} \\
 &= \frac{1}{2^n} \sum_{k=0}^{n+1} \binom{n+1}{k} \int_{u=0}^{u=2} N_n(2x - u - k) \frac{1}{2} du && \text{where } u \triangleq 2\tau \implies \tau = \frac{1}{2}u \\
 &= \frac{1}{2^{n+1}} \sum_{k=0}^{n+1} \binom{n+1}{k} \left[ \int_{u=0}^{u=1} N_n(2x - k - u) du + \int_{u=1}^{u=2} N_n(2x - k - u) du \right] \\
 &= \frac{1}{2^{n+1}} \sum_{k=0}^{n+1} \binom{n+1}{k} \left[ \int_{u=0}^{u=1} N_n(2x - k - u) du + \int_{v=0}^{v=1} N_n(2x - k - v - 1) dv \right] && \text{where } v \triangleq u - 1 \implies u = v + 1 \\
 &= \frac{1}{2^{n+1}} \sum_{k=0}^{n+1} \binom{n+1}{k} [N_n(2x - k) + N_n(2x - k - 1)] && \text{by Lemma S.2 page 371} \\
 &= \frac{1}{2^{n+1}} \left[ \sum_{k=0}^{n+1} \binom{n+1}{k} N_n(2x - k) + \sum_{m=1}^{n+2} \binom{n+1}{m-1} N_n(2x - m) \right] && \text{where } m \triangleq k + 1 \implies k = m - 1 \\
 &= \frac{1}{2^{n+1}} \left[ \underbrace{\sum_{k=1}^{n+1} \left[ \binom{n+1}{k} + \binom{n+1}{k-1} \right] N_n(2x - k)}_{\text{common indices of above two summations}} + \underbrace{\binom{n+1}{0} N_n(2x - 0)}_{k=0 \text{ term}} + \underbrace{\binom{n+2}{n+2} N_n(2x - n - 2)}_{m=n+2 \text{ term}} \right] \\
 &= \frac{1}{2^{n+1}} \left[ \underbrace{\sum_{k=1}^{n+1} \binom{n+2}{k} N_n(2x - k)}_{\text{by Stifel formula (Theorem ?? page ??)}} + \underbrace{\binom{n+2}{0} N_n(2x - 0)}_{\text{because } \binom{n+1}{0} = 1 = \binom{n+2}{0}} + \underbrace{\binom{n+2}{n+2} N_n(2x - n - 2)}_{m=n+2 \text{ term}} \right] \\
 &= \frac{1}{2^{n+1}} \sum_{k=0}^{n+2} \binom{n+2}{k} N_n(2x - k)
 \end{aligned}$$

⇒

*Example S.16.* <sup>25</sup>The 0 order B-spline dyadic decomposition

$$N_0(x) = \frac{1}{1} \sum_{k=0}^{k=1} \binom{1}{k} N_0(2x - k)$$

is illustrated to the right.

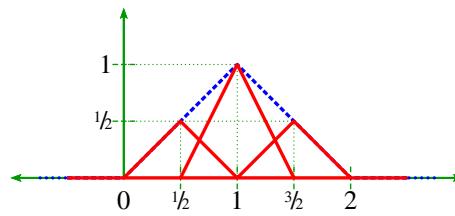


<sup>25</sup> Strang (1989) page 615 (Box function), Strang and Nguyen (1996) page 441 (Box function)

*Example S.17.* <sup>26</sup>The 1st order B-spline dyadic decomposition

$$N_1(x) = \frac{1}{2} \sum_{k=0}^{k=2} \binom{2}{k} N_1(2x - k)$$

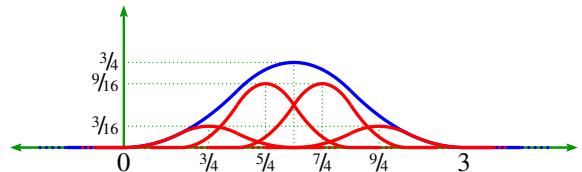
is illustrated to the right.



*Example S.18.* The 2nd order B-spline dyadic decomposition

$$N_2(x) = \frac{1}{4} \sum_{k=0}^{k=3} \binom{3}{k} N_2(2x - k)$$

is illustrated to the right.



### S.6.3 B-spline MRA scaling functions

**Theorem S.10.** Let  $f N_n(x)$  be the  $n$ TH ORDER B-SPLINE (Definition S.2 page 371).

Let  $V_j \triangleq \text{span}(\{N_n(2^j x - k)\}_{k \in \mathbb{Z}})$ .

**T H M**  $(V_j)_{j \in \mathbb{Z}}$  is a MULTIRESOLUTION ANALYSIS on  $L^2_{\mathbb{R}}$  with SCALING FUNCTION  $\phi(x) \triangleq N_n(x)$

PROOF:

1. lemma:  $(N_n(x - k))_{k \in \mathbb{Z}}$  is a *Riesz sequence*. Proof: by Theorem S.8 (page 394).

2. lemma:  $\exists (h_k) \text{ such that } N_n(x) = \sum_{k \in \mathbb{Z}} h_k N_n(2x - k)$ . Proof: by Theorem S.9 (page 395). In fact, note that  $h_k = \frac{1}{2^n \sqrt{2}} \binom{n+1}{k}$

3. lemma:  $\tilde{F}N_n(\omega)$  is *continuous* at 0. Proof:

$$\tilde{F}N_n(\omega) = \frac{1}{\sqrt{2\pi}} e^{-i\frac{(n+1)\omega}{2}} \left( \text{sinc} \frac{\omega}{2} \right)^{n+1} \quad \text{by Lemma S.5 page 386}$$

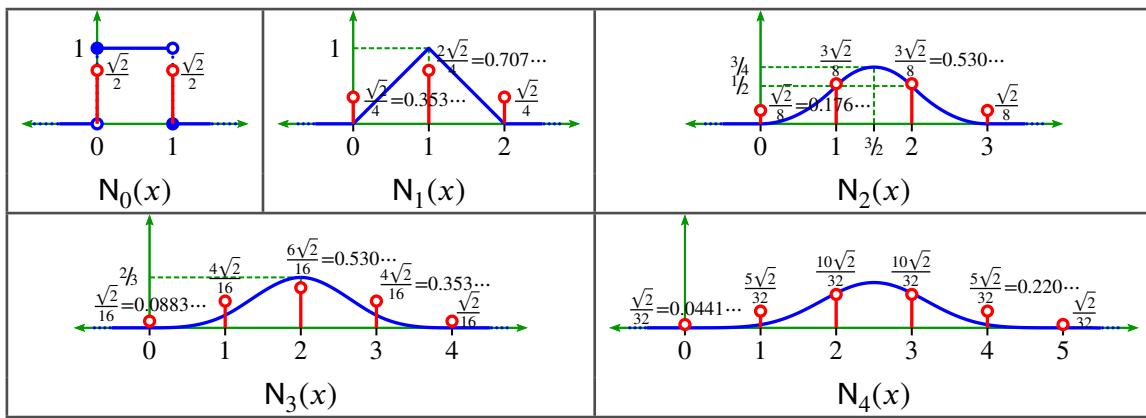
$\implies$  continuous at 0 by known property of sinc function

4. lemma:  $\tilde{\phi}(0) \neq 0$ . Proof:

$$\begin{aligned} \tilde{F}N_n(0) &= \frac{1}{\sqrt{2\pi}} e^{-i\frac{(n+1)\omega}{2}} \left( \text{sinc} \frac{\omega}{2} \right)^{n+1} \Big|_{\omega=0} && \text{by Lemma S.5 page 386} \\ &= 1 \cdot \frac{1}{1/2} = 2 && \text{by } l'Hôpital's \text{ rule} \\ &\neq 0 \end{aligned}$$

5. The completion of this proof follows directly from (1) lemma, (2) lemma, (3) lemma, (4) lemma, and Theorem ?? (page ??).

<sup>26</sup> Strang (1989) page 615 (Hat function), Strang and Nguyen (1996) page 442 (Hat function), Heil (2011) page 380 (Fig. 12.10)

Figure S.4: *dilation equation* demonstrations for selected B-splines (Example S.19 page 398)

## S.6.4 B-spline MRA coefficient sequences

Because each *B-spline*  $N_n(x)$  is the *scaling function* for an *MRA* (Theorem S.10 page 397), each *B-spline* also satisfies the *dilation equation* (Theorem ?? page ??) such that

$$N_n(x) = \sqrt{2} \sum_{k \in \mathbb{Z}} h_k N(2x - k) \quad \text{where} \quad h_k = \begin{cases} \frac{\sqrt{2}}{2^{n+1}} \binom{n+1}{k} & \text{for } n = 0, 1, \dots, n \\ 0 & \text{otherwise} \end{cases}$$

The resulting sequence  $(h_k)_{k \in \mathbb{Z}}$  is the *ordern B-spline MRA coefficient sequence* induced by the *order n B-spline MRA scaling sequence*  $\phi(x) \triangleq N_n(x)$ .<sup>27</sup>

*Example S.19.* See Figure S.4 (page 398) for some *dilation equation* demonstrations of selected B-splines.

**Theorem S.11** (*B-spline scaling coefficients*). *Let  $(L^2_{\mathbb{R}}, (\mathcal{V}_j), \phi, (h_n))$  be an MRA SYSTEM (Definition ?? page ??). Let  $N_n(x)$  be a nTH ORDER B- SPLINE (Definition S.2 page 371).*

T H M	$\underbrace{\phi(x) \triangleq N_n(x)}_{(1) \text{ B-spline scaling function}} \implies (h_k) = \begin{cases} \frac{\sqrt{2}}{2^{n+1}} \binom{n}{k} & \text{for } k = 0, 1, \dots, n \\ 0 & \text{otherwise} \end{cases} \quad (2) \text{ scaling sequence in "time"}$ $\iff \check{h}(z) \Big _{z \triangleq e^{i\omega}} = \frac{\sqrt{2}}{2^n} (1 + z^{-1})^{n+1} \Big _{z \triangleq e^{i\omega}} \quad (3) \text{ scaling sequence in "z domain"}$ $\iff \check{h}(\omega) = 2\sqrt{2} e^{-i\frac{n+1}{2}\omega} \left[ \cos\left(\frac{\omega}{2}\right) \right] \quad (4) \text{ scaling sequence in "frequency"}$
-------------	--

PROOF:

1. Proof that (1)  $\implies$  (3): By Theorem S.10 page 397 we know that  $N_n(x)$  is a *scaling function* (Definition ?? page ??). So then we know that we can use Lemma ?? page ??.

$$\begin{aligned}
 \check{h}(\omega) &= \sqrt{2} \frac{\tilde{\phi}(2\omega)}{\tilde{\phi}(\omega)} && \text{by Lemma ?? page ??} \\
 &= \sqrt{2} \frac{\tilde{N}_n(2\omega)}{\tilde{N}_n(\omega)} && \text{by (1)} \\
 &= \sqrt{2} \frac{\frac{1}{\sqrt{2\pi}} \left( \frac{1-e^{-i2\omega}}{2i\omega} \right)^{n+1}}{\frac{1}{\sqrt{2\pi}} \left( \frac{1-e^{-i\omega}}{i\omega} \right)^{n+1}} && \text{by Lemma S.5 page 386}
 \end{aligned}$$

<sup>27</sup>For Octave/ MatLab code useful for plotting a function given a sequence of coefficients  $(h_k)$ , see Section ?? (page ??).

$$\begin{aligned}
&= \frac{\sqrt{2}}{2^{n+1}} \left( \frac{1 - z^{-2}}{1 - z^{-1}} \right)^{n+1} \Big|_{z=e^{i\omega}} \\
&= \frac{\sqrt{2}}{2^{n+1}} \left[ \left( \frac{1 - z^{-2}}{1 - z^{-1}} \right) \left( \frac{1 + z^{-1}}{1 + z^{-1}} \right) \right]^{n+1} \Big|_{z=e^{i\omega}} \\
&= \frac{\sqrt{2}}{2^{n+1}} \left( \frac{(1 - z^{-2})(1 + z^{-1})}{1 - z^{-2}} \right)^{n+1} \Big|_{z=e^{i\omega}} \\
&= \frac{\sqrt{2}}{2^n} (1 + z^{-1})^{n+1} \Big|_{z \triangleq e^{i\omega}}
\end{aligned}$$

2. Proof that (3)  $\iff$  (2):

$$\begin{aligned}
\check{h}(z) \Big|_{z \triangleq e^{i\omega}} &= \frac{\sqrt{2}}{2^n} (1 + z^{-1})^{n+1} \Big|_{z \triangleq e^{i\omega}} && \text{by (3)} \\
&= \frac{\sqrt{2}}{2^n} \left( \sum_{k=0}^{n+1} \binom{n}{k} z^{-k} \right) \Big|_{z \triangleq e^{i\omega}} && \text{by binomial theorem} \\
\iff h_k &= \frac{\sqrt{2}}{2^{n+1}} \binom{n}{k} && \text{by definition of } Z \text{ transform (Definition R.4 page 358)}
\end{aligned}$$

3. Proof that (3)  $\implies$  (4):

$$\begin{aligned}
\tilde{h}(\omega) &= \check{h}(z) \Big|_{z \triangleq e^{i\omega}} && \text{by definition of DTFT (Definition Q.1 page 347)} \\
&= \frac{\sqrt{2}}{2^n} (1 + z^{-1})^{n+1} \Big|_{z \triangleq e^{i\omega}} && \text{by (3)} \\
&= \frac{\sqrt{2}}{2^n} (1 + e^{-i\omega})^{n+1} && \text{by definition of } z \\
&= \frac{\sqrt{2}}{2^n} \left[ e^{-i\frac{1}{2}\omega} \left( e^{i\frac{\omega}{2}} + e^{-i\frac{\omega}{2}} \right) \right]^{n+1} \\
&= \frac{\sqrt{2}}{2^n} e^{-i\frac{n+1}{2}\omega} \left[ 2\cos\left(\frac{\omega}{2}\right) \right]^{n+1} \\
&= 2\sqrt{2} e^{-i\frac{n+1}{2}\omega} \left[ \cos\left(\frac{\omega}{2}\right) \right]^{n+1}
\end{aligned}$$

4. Proof that (3)  $\iff$  (4):

$$\begin{aligned}
\check{h}(z) \Big|_{z \triangleq e^{i\omega}} &= \check{h}(e^{i\omega}) \\
&= \tilde{h}(\omega) \\
&= 2\sqrt{2} e^{-i\frac{n+1}{2}\omega} \left[ \cos\left(\frac{\omega}{2}\right) \right]^{n+1} && \text{by (4)} \\
&= \frac{\sqrt{2}}{2^n} e^{-i\frac{n+1}{2}\omega} \left[ 2\cos\left(\frac{\omega}{2}\right) \right]^{n+1} \\
&= \frac{\sqrt{2}}{2^n} \left[ e^{-i\frac{1}{2}\omega} \left( e^{i\frac{\omega}{2}} + e^{-i\frac{\omega}{2}} \right) \right]^{n+1} \\
&= \frac{\sqrt{2}}{2^n} (1 + e^{-i\omega})^{n+1} \\
&= \frac{\sqrt{2}}{2^n} (1 + z^{-1})^{n+1} \Big|_{z \triangleq e^{i\omega}}
\end{aligned}$$



*Example S.20 (2 coefficient case).* <sup>28</sup> Let  $(L^2_{\mathbb{R}}, (\mathcal{V}_j), \phi, (h_n))$  be an MRA system (Definition ?? page ??).

<b>E</b>	$\left\{ \begin{array}{l} 1. \text{ supp}\phi(x) = [0 : 1] \quad \text{and} \\ 2. (\phi(x - k)) \text{ forms a} \\ \text{partition of unity} \end{array} \right\}$	$\iff h_n = \left\{ \begin{array}{ll} \frac{\sqrt{2}}{2} & \text{for } n = 0 \\ \frac{\sqrt{2}}{2} & \text{for } n = 1 \\ 0 & \text{otherwise} \end{array} \right\} \iff \underbrace{\{\phi(x) = N_0(x)\}}_{(C)}$
----------	--	---

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  $\text{supp}\phi(x) = [0 : 1]$  (Definition ?? page ??) and by Theorem ?? page ??.
- (b) lemma (equations for  $(h_k)$ ): Because  $(h_k)$  is a *scaling coefficient sequence* (Definition ?? page ??), it must satisfy the *admissibility equation* (Theorem ?? page ??). And because  $(\phi(x - k))$  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:

$$\begin{aligned} h_0 + h_1 &= \sqrt{2} && \text{(admissibility condition)} \\ h_0 - h_1 &= 0 && \text{(partition of unity/zero at -1/vanishing 0th moment)} \end{aligned}$$

- (c) lemma: The equations provided by (1b) lemma can be expressed in matrix algebra form as follows...

$$\underbrace{\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}}_A \begin{bmatrix} h_0 \\ h_1 \end{bmatrix} = \begin{bmatrix} \sqrt{2} \\ 0 \end{bmatrix}$$

- (d) lemma: The *inverse A*<sup>-1</sup> of A can be expressed as demonstrated below...

$$\begin{aligned} \left[ \begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 1 & -1 & 0 & 1 \end{array} \right] &\rightarrow \left[ \begin{array}{cc|cc} 2 & 0 & 1 & 1 \\ 1 & -1 & 0 & 1 \end{array} \right] \rightarrow \left[ \begin{array}{cc|cc} 2 & 0 & 1 & 1 \\ 0 & -1 & -\frac{1}{2} & \frac{1}{2} \end{array} \right] \rightarrow \left[ \begin{array}{cc|cc} 1 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 1 & \frac{1}{2} & -\frac{1}{2} \end{array} \right] \\ \implies A^{-1} &= \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \end{aligned}$$

- (e) Proof for the values of  $(h_k)$  (B):

$$\begin{aligned} \begin{bmatrix} h_0 \\ h_1 \end{bmatrix} &= A^{-1} A \begin{bmatrix} h_0 \\ h_1 \end{bmatrix} = A^{-1} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} h_0 \\ h_1 \end{bmatrix} && \text{by (1c) lemma} \\ &= A^{-1} \begin{bmatrix} \sqrt{2} \\ 0 \end{bmatrix} && \text{by (1c) lemma} \\ &= \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \sqrt{2} \\ 0 \end{bmatrix} && \text{by (1d) lemma} \\ &= \frac{\sqrt{2}}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \end{aligned}$$

<sup>28</sup> Haar (1910), Wojtaszczyk (1997) pages 14–15 (“Sources and comments”)

2. Proof that (B)  $\implies$  (C):

$$\begin{aligned}
 (B) \implies \phi(x) &= \sum_{k \in \mathbb{Z}} h_k \sqrt{2} \phi(2x - k) && \text{dilation equation} \\
 &= \sum_{k=0}^{k=1} \left( \frac{\sqrt{2}}{2} \right) \sqrt{2} \phi(2x - k) && \text{by item (1e) page 400} \\
 &= \sum_{k=0}^{k=1} \phi(2x - k) \\
 &= \sum_{k=0}^{k=1} \binom{1}{k} \phi(2x - k) && \text{by definition of } \binom{n}{k} \\
 \implies (D) && \text{by } B\text{-spline dyadic decomposition} & \text{(Theorem S.9 page 395)}
 \end{aligned}$$

3. Proof that (B)  $\Leftarrow$  (C):

$$\begin{aligned}
 (C) \implies N_0(x) &= \sum_{k=0}^{k=1} \binom{1}{k} N_0(2x - k) && \text{by } B\text{-spline dyadic decomposition} & \text{(Theorem S.9 page 395)} \\
 &= \sum_{k=0}^{k=1} \left( \frac{\sqrt{2}}{2} \right) \sqrt{2} N_0(2x - k) && \text{by definition of } \binom{n}{k} & \text{(Definition ?? page ??)} \\
 &= \sum_{k \in \mathbb{Z}} h_k \sqrt{2} N_0(2x - k) && \text{by definition of } \binom{n}{k} & \text{(Definition ?? page ??)} \\
 \implies (B) &
 \end{aligned}$$

4. Proof that (A)  $\Leftarrow$  (C):

1. Proof that (C)  $\implies \text{supp } \phi(x) = [0 : 1]$ : by Theorem S.4 (page 382)
2. Proof that (C)  $\implies (\phi(x - k))$  forms a *partition of unity*: by Theorem S.7 (page 388)



E X	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 S.1: Coefficients of the *B-splines*  $N_n(x)$  multiplied by  $n!$  (Example S.9 page 378)

## APPENDIX T

### SOURCE CODE

The free and open source software package Maxima has been used to compute some of the algebraic expressions for *B-splines* used in APPENDIX S (page 371):

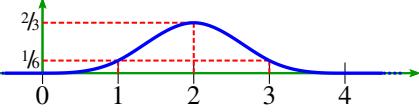
```
1 /*=====
2 * Daniel J. Greenhoe
3 * Maxima script file
4 * To execute this script, start Maxima in a command window
5 * in the subdirectory containing this file (e.g. c:\math\maxima\
6 * and then after the (%i...) prompt enter
7 * batchload("bspline.max")$
8 * Data produced will be written to the file "bsplineout.txt".
9 * reference: http://maxima.sourceforge.net/documentation.html
10 */
11 /*
12 * initialize script
13 */
14 reset();
15 kill(all);
16 load(orthopoly);
17 display2d:false; /* 2-dimensional display */
18 writefile("bsplineout.txt");
19 /*
20 * n = B-spline order parameter
21 * may be set to any value in {1,2,3,...}
22 */
23 n:2;
24 print("=====");
25 print("Daniel J. Greenhoe");
26 print("Output file for nth order B-spline Nn(x) calculation, n=",n," .");
27 print("Output produced using Maxima running the script file bspline.max");
28 print("=====");
29 Nnx:(1/n!)*sum((-1)^k*binomial(n+1,k)*(x-k)^n*unit_step(x-k),k,0,n+1);
30 print("=====");
31 print("      n+1      k (n+1)      n      ");
32 print("      n! Nn(x) = SUM (-1) ( ) (x-k)  step(x-k) ,n=",n," ");
33 print("      k=0      ( k )      ");
34 print("      ,n+1,"      k (",n+1,")      ,n);
35 print(n,"! Nn(x) = SUM (-1) ( ) (x-k)  step(x-k)");
36 print("      k=0      ( k )");
37 print("      = ",expand(n!*Nnx));
38 print("=====");
39 assume(x<=0);   print(n!,"N(x)= ",expand(n!*Nnx)," for x<=0");   forget(x<=0);
40 for i:0 thru n step 1 do(
41   assume(x>i,x<(i+1)),
42   print(n!,"N(x)= ",expand(n!*Nnx)," for ",i,"<x<",i+1),
43   tex(expand(n!*Nnx),"djh.tex"),/*write output in TeX format to file "djh.tex"*/
44   forget(x>i,x<(i+1))
45 );
46 assume(x>(n+1)); print(n!,"N(x)= ",expand(n!*Nnx)," for x>",n+1); forget(x>(n+1));
```

```

47 print("-----");
48 print(" values at some specific points x:           ");
49 print("-----");
50 y:Nnx,x=(n+1)/2;print("N(",(n+1)/2,")= ",y," (center value)");
51 y:Nnx,x=(n+2)/2;print("N(",(n+2)/2,")= ",y);
52 y:Nnx,x=(n+3)/2;print("N(",(n+3)/2,")= ",y);
53 /*-----*/
54 * close output file
55 *-----*/
56 closefile();

```

Once the polynomial expressions for a *B-spline* have been calculated, they can be plotted within a  $\text{\LaTeX}$  environment using the [pst-plot package](#) along with a  $\text{\LaTeX}$  source file such as the following:<sup>1</sup>

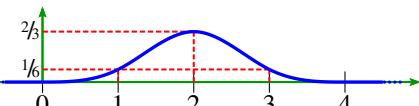


```

1 %=====
2 % Daniel J. Greenhoe
3 % LaTeX file
4 % N_3(x) B-spline
5 % nominal unit = 10mm
6 %=====
7 \begin{pspicture}(-1,-0.5)(5,1)
8 %
9 % parameters
10 %
11 \psset{plotpoints=64,labelsep=1pt}%
12 %
13 % axes
14 %
15 \psaxes[linewidth=0.75pt, linecolor=axis ,yAxis=false ,ticks=x, labels=x]{<->}(0,0)(-1,0)(5,1)% x axis
16 \psaxes[linewidth=0.75pt, linecolor=axis ,xAxis=false ,ticks=x, labels=x]{->}(0,0)(-1,0)(5,1)% y axis
17 %
18 % annotation
19 %
20 \psline [linestyle=dashed, linewidth=0.75pt, linecolor=red](2,0)(2,0.667)%
21 \psline [linestyle=dashed, linewidth=0.75pt, linecolor=red](0,0.667)(2,0.667)%
22 \psline [linestyle=dashed, linewidth=0.75pt, linecolor=red](1,0)(1,0.1667)%
23 \psline [linestyle=dashed, linewidth=0.75pt, linecolor=red](3,0)(3,0.1667)%
24 \psline [linestyle=dashed, linewidth=0.75pt, linecolor=red](0,0.1667)(3,0.1667)%
25 \uput[180](0,0.667){$\frac{2}{3}$}%
26 \uput[180](0,0.1667){$\frac{1}{6}$}%
27 %
28 % function plot
29 %
30 \psplot{0}{1}{+1 x 3 exp mul}                                6 div% for 0<=x<=1
31 \psplot{1}{2}{-3 x 3 exp mul +12 x 2 exp mul add -12 x mul add +4 add 6 div}%
32 % for 1<=x<=2
33 \psplot{2}{3}{+3 x 3 exp mul -24 x 2 exp mul add +60 x mul add -44 add 6 div}%
34 % for 2<=x<=3
35 \psplot{3}{4}{-1 x 3 exp mul +12 x 2 exp mul add -48 x mul add +64 add 6 div}%
36 % for 3<=x<=4
37 \psline(0,0)(-0.5,0)\psline[linestyle=dotted](-0.5,0)(-0.75,0)%          % for x<=0
38 \psline(4,0)(4.5,0)\psline[linestyle=dotted](4.5,0)(4.75,0)%          % for x>=4
39 \end{pspicture}%

```

Alternatively, one can plot  $N_3(x)$  more or less directly from the equation given in Theorem S.1 (page 375) without first calculating the polynomial expressions:



```

1 %=====
2 % Daniel J. Greenhoe
3 % LaTeX file
4 % N_3(x) B-spline
5 % nominal unit = 10mm
6 %=====
7 \begin{pspicture}(-1,-0.5)(5,1)
8 %
9 % parameters
10 %
11 \psset{plotpoints=64,labelsep=1pt}%

```

<sup>1</sup>For help with PostScript®math operators, see [Adobe \(1999\)](#), pages 508–509 (Arithmetic and Math Operators).

```

12 %
13 % axes
14 %
15 \psaxes[linewidth=0.75pt, linecolor=axis, yAxis=false, ticks=x, labels=x]{<->}(0,0)(-1,0)(5,1)% x axis
16 \psaxes[linewidth=0.75pt, linecolor=axis, xAxis=false, ticks=x, labels=x]{->}(0,0)(-1,0)(5,1)% y axis
17 %
18 % annotation
19 %
20 \psline[linestyle=dashed, linewidth=0.75pt, linecolor=red](2,0)(2,0.667)%
21 \psline[linestyle=dashed, linewidth=0.75pt, linecolor=red](0,0.667)(2,0.667)%
22 \psline[linestyle=dashed, linewidth=0.75pt, linecolor=red](1,0)(1,0.1667)%
23 \psline[linestyle=dashed, linewidth=0.75pt, linecolor=red](3,0)(3,0.1667)%
24 \psline[linestyle=dashed, linewidth=0.75pt, linecolor=red](0,0.1667)(3,0.1667)%
25 \put[180](0,0.667){$ \frac{2}{3} $}%
26 \put[180](0,0.1667){$ \frac{1}{6} $}%
27 %
28 % for n=3
29 % 
$$\frac{1}{n!} \sum_{k=1}^{n+1} (-1)^k \binom{n+1}{k} (x-k)^n s(x-k) = \frac{1}{3!} \sum_{k=1}^4 (-1)^k \binom{4}{k} (x-k)^3 s(x-k)$$

30 % where  $s(x) = 0$  for  $x < 0$  and  $1$  for  $x \geq 0$  (step function)
31 %
32 %
33 \psplot{0}{1}{1 x 0 sub 3 exp mul 6 div}%
34 \psplot{1}{2}{1 x 0 sub 3 exp mul 4 x 1 sub 3 exp mul sub 6 div}%
35 \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}%
36 \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}%
37 \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}%
38 %
39 % 
$$N_3(x) = \frac{[(4 \text{choose} 0)(x-0)^3 - (4 \text{choose} 1)(x-1)^3 + (4 \text{choose} 2)(x-2)^3 - (4 \text{choose} 3)(x-3)^3 + (4 \text{choose} 4)(x-4)^3]/3!}{6}$$

40 % 
$$= \frac{1}{6} [(x-0)^3 - 4(x-1)^3 + 6(x-2)^3 - (x-3)^3 + (x-4)^3]$$

41 %
42 \psline(0,0)(-0.5,0)%
43 \psline[linestyle=dotted](-0.5,0)(-0.75,0)%
44 \psline[linestyle=dotted](4.5,0)(4.75,0)%
45 \end{pspicture}%

```



## Back Matter



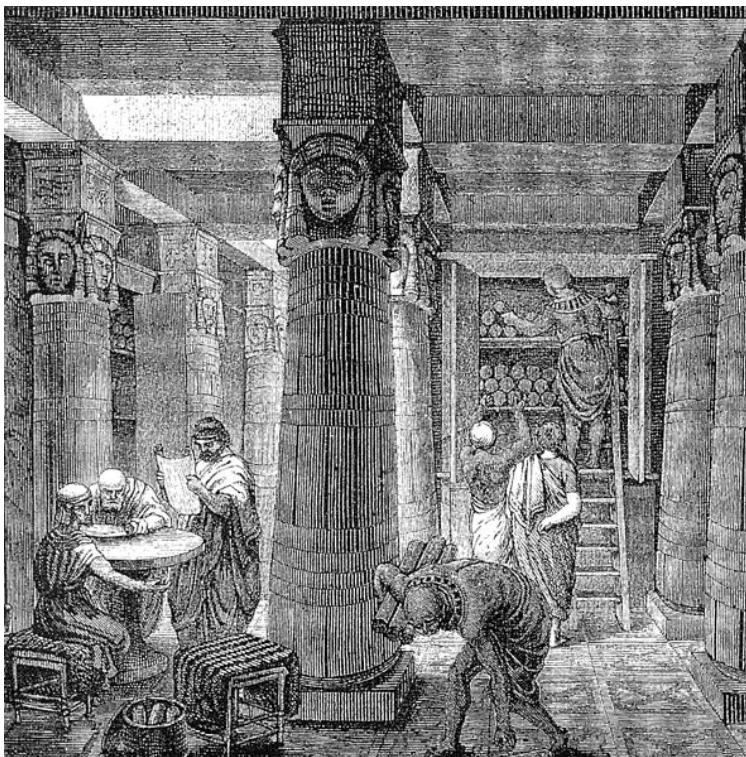
**“It appears to me that if one wants to make progress in mathematics, one should study the masters and not the pupils.”**

Niels Henrik Abel (1802–1829), Norwegian mathematician <sup>2</sup>

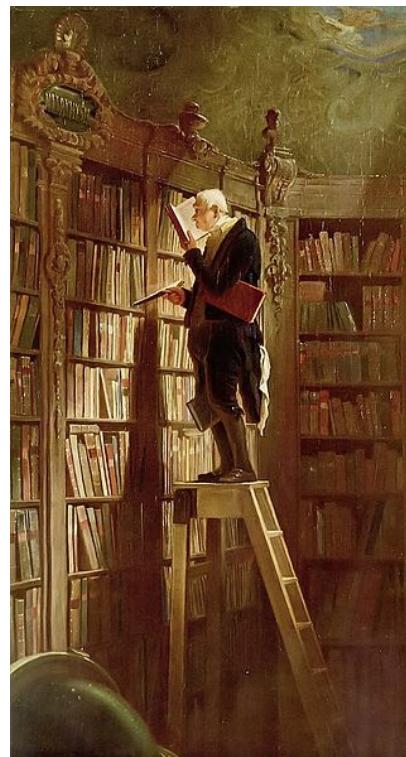


**“When evening comes, I return home and go to my study. On the threshold I strip naked, taking off my muddy, sweaty workaday clothes, and put on the robes of court and palace, and in this graver dress I enter the courts of the ancients and am welcomed by them, and there I taste the food that alone is 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. <sup>3</sup>



ancient library of Alexandria



The Book Worm by Carl Spitzweg, circa 1850 <sup>4</sup>



**“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 <sup>5</sup>

<sup>2</sup> quote: [Simmons \(2007\)](#), page 187.

image: [http://en.wikipedia.org/wiki/Image:Niels\\_Henrik\\_Abel.jpg](http://en.wikipedia.org/wiki/Image:Niels_Henrik_Abel.jpg), public domain

<sup>3</sup> quote: [Machiavelli \(1961\)](#), page 139?.

image: [http://commons.wikimedia.org/wiki/File:Santi\\_di\\_Tito\\_-\\_Niccolo\\_Machiavelli%27s\\_portrait\\_headcrop.jpg](http://commons.wikimedia.org/wiki/File:Santi_di_Tito_-_Niccolo_Machiavelli%27s_portrait_headcrop.jpg), public domain

<sup>4</sup> <http://en.wikipedia.org/wiki/File:Ancientlibraryalex.jpg>, public domain [http://en.wikipedia.org/wiki/File:Carl\\_Spitzweg\\_021.jpg](http://en.wikipedia.org/wiki/File:Carl_Spitzweg_021.jpg)

<sup>5</sup> quote: [肯所集](#)  
image: [https://en.wikipedia.org/wiki/Yoshida\\_Kenko](https://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>.

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*. Academic 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. Allemand, R. Zimmerman, and David L. Brown. Techniques of reducing noise in frequency response measurements. *ASME Winter Annual Meeting*, 1979. session on Application of System Identification Techniques.

Randall J. Allemand, 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*. Birkhäuser 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. URL <http://books.google.com/books?id=ZP8-AAAAIAAJ>.

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 re-publication” available from Dover (isbn 0486402517).

George Bachman, Lawrence Narici, and Edward Beckenstein. *Fourier and Wavelet Analysis*. Universitext Series. Springer, 2000. ISBN 9780387988993. URL <http://books.google.com/books?vid=ISBN0387988998>.

David H. Bailey, Jonathan M. Borwein, and Andrew Mattingly. The computation of previously inaccessible digits of  $\pi^2$  and catalan's constant (2013). page 23, March 24 2011. URL <https://pdfs.semanticscholar.org/0b15/22a1b2f76729e19be2be1bca9f57031a4db9.pdf>.



- David H. Bailey, Jonathan M. Borwein, and Andrew Mattingly. 20. the computation of previously inaccessible digits of  $\pi^2$  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, 2013. ISBN 9783319323770. URL <http://books.google.com/books?vid=ISBN3319323776>.
- 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*. 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=cXJEAAAIAAJ>.
- 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*, 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>.
- 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](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>.

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.

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>.

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 l'Acad. de St.-Pétersbourg*, 1(9), 1859. URL [http://www-stat.wharton.upenn.edu/~steele/Publications/Books/CSMC/CSMC\\_index.html](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/inernet.dli.2015.134151/page/n2>.

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/anintroduction00bromgoog>.

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 [http://frontweb.vuse.vanderbilt.edu/vuse\\_web/directory/facultybio.asp?FacultyID=9](http://frontweb.vuse.vanderbilt.edu/vuse_web/directory/facultybio.asp?FacultyID=9).

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 J. Callahan. *Advanced Calculus: A Geometric View*. Springer Science & Business Media, September 9 2010. ISBN 9781441973320. URL <http://books.google.com/books?vid=ISBN144197332X>.

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/coursdanalyse00caucgoog>. 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 Chapman. *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](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](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 fulfillment 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](http://link.aip.org/link/?jmp/27/1271).
- 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 probabilità 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.

Johan de Witt. *Elementa curvarum linearum (Elements of Curves)*. 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](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](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](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>.

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>.

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>.

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>.

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 Transactions 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.

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=TDQJAAAIAAJ>.

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'événements. *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'événements. *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). *I:Oefversigt af Kongl. Sv. Vetenskaps-Academiens Förfhandlingar*, 57:39–66, 1900.

Erik Ivar Fredholm. Sur une classe d'équations 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. Über 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. Über 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):3–24, 1941.

Israel M. Gelfand and Mark A. Naimark. *Normed Rings with an Involution and their Representations*, 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*, pages 3–19. 1943b. ISBN 0821851756. 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. *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 0-471-19748-3. URL [http://vadkudr.boom.ru/Collection/fundwave\\_contents.html](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](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*. 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  $0,1,2,\dots$  and  $m$  in  $1,2,3,\dots,(n+1)2$ . *Online Encyclopedia of Integer Sequences*, 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*. Academic 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-Verlag, New York, 2 edition, 1958. ISBN 0-387-90093-4. URL <http://books.google.com/books?vid=isbn0387900934>.
- Paul R. Halmos. *Introduction 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](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>.

S. Haykin and S. Kesler. *Prediction-Error Filtering and Maximum-Entropy Spectral Estimation*, 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*, 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>.

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>.

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. Istrătescu. *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. Orthogonality and linear functionals in normed linear spaces. *Transactions of the American Mathematical Society*, 61(2):265–292, March 1947. ISSN 1088-6850. doi: 10.2307/1990220. URL <http://www.jstor.org/stable/1990220>.

Bjorn Jawerth and Wim Sweldens. An overview of wavelet based multiresolutional analysis. *SIAM Review*, 36:377–412, September 1994. URL <http://cm.bell-labs.com/who/wim/papers/papers.html#overview>.

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](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 inégalités 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 Society*, 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>.
- 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, Massachusetts, 1988. ISBN 0201156741. URL <http://www.worldcat.org/isbn/0201156741>.
- Yoshida Kenko. *The Tsurezure 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 York, 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>.

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. Leclerc, 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 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. pages 908–918, 1986. URL <https://www.researchgate.net/publication/283994172>.

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 Society, 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>.

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\\_1XCFRkC](http://books.google.com/books?id=P3Vl_1XCFRkC).

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=4oeBAAAIAAJ>.



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/bbe0/55f40768bbe9e6476bdce701effb9243f7.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](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/z2gxu/>.

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. lupaš. *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/geometriederzah100minkrich>. 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>.
- 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/rosenth1/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: Stanford 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>.
- Anthanasis 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>.
- Anthanasis Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, New York, 3 edition, 1991. ISBN 0070484775. URL <http://books.google.com/books?vid=ISBN0070484775>.
- Anthanasis 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>.
- Athanasis 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>.
- 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 present: *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=PVrvAAAMAAJ>.

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. *Geophysics*, 31(3):482–500, June 1966. doi: 10.1190/1.1439788. URL <http://dx.doi.org/10.1190/1.1439788>.

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://link.springer.com/content/pdf/10.1007/978-1-4419-8339-8_13.pdf). [https://projecteuclid.org/download/pdf\\_1/euclid.aoms/1177728190](https://projecteuclid.org/download/pdf_1/euclid.aoms/1177728190).

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. *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>.
- 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 variation-srechnung. *Acta Soc. Scient. Fen.*, 15:315–362, 1885. URL [http://www-stat.wharton.upenn.edu/~steele/Publications/Books/CSMC/CSMC\\_index.html](http://www-stat.wharton.upenn.edu/~steele/Publications/Books/CSMC/CSMC_index.html).
- 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. Über 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 commemorate the birthdays of Mr. Karl Weierstrass).
- Mícheál Ó Searcoid. *Elements of Abstract Analysis*. Springer Undergraduate Mathematics Series. Springer, 2002. ISBN 9781852334246. URL <http://books.google.com/books?vid=ISBN185233424X>.
- 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>.

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/elementaryalgebr00slaugh/>.

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*, volume 13 of *Geophysics*, 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](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](https://www.researchgate.net/profile/Gabor_Horvath/publication/263107833).
- 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](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>.

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=3bjq0gAACAAJ>.

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/weih11\\_NPG06-13\(5\)pp577-584.pdf](http://eprints.whiterose.ac.uk/1945/1/weih11_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*, 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/6cccd/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>.



A Book Concerning Statistical Signal Processing [VERSION 0.19]

<https://github.com/dgreenhoe/pdfs/blob/master/abcstat.pdf>

2019 OCTOBER 30 (WEDNESDAY) 02:27PM UTC  
COPYRIGHT © 2019 DANIEL J. GREENHOE

---

## REFERENCE INDEX

- Devices (2016), 14  
Abramovich and Aliprantis (2002), 196, 358  
Aliprantis and Burkinshaw (2006), 202, 204–207  
Abdaheer (2009), 137, 147  
Abom (1986), 110  
Aliprantis and Burkinshaw (1998), 216, 218, 222, 249, 259, 260, 262–265, 267–269, 273, 275–277, 302, 308, 312, 314–316  
Adobe (1999), 404  
Akhiezer and Glazman (1993), 198, 251  
Allemand et al. (1979), 110  
Allemand et al. (1987), 110  
Allen (2018), 6  
Amann and Escher (2008), 304  
Amemiya and Araki (1966), 252  
Amir (1986), 267, 282–284  
Anderson (1958), 176, 183  
Anderson (1984), 176, 183  
Angus (1994), 29  
Apostol (1975), 285  
Aubin and Frankowska (2009), 287  
Aubin (2011), 287  
Autonne (1901), 330  
Autonne (1902), 330  
Bachman (1964), 340  
Bachman and Narici (1966), 209, 211, 213, 222, 250, 322, 324  
Bachman et al. (2000), 214, 217, 226, 337  
Bailey et al. (2011), 392  
Bailey et al. (2013), 392  
Bajpai (1967), 139  
Bak (2013), 121  
Balakrishnan and Lai (2009), 29  
Banach (1922), 197, 200, 201, 275, 276  
Banach (1932b), 275  
Banach (1932a), 214, 275, 278  
Barvinok (2002), 286, 287  
Beer (1993), 287  
Bell (1986), 302  
Bellman (1954), 149  
Bellman (1961), 149  
Bellman (1971), 149  
Bendat and Piersol (1966), 70, 71  
Bendat (1978), 110  
Bendat and Piersol (1980), 71, 125  
Bendat and Piersol (1993), 109, 125  
Bendat and Piersol (2010), 48, 70, 71, 109, 110, 125  
Berberian (1961), 199, 200, 209, 222, 226, 240, 245, 248, 249, 257, 259, 266, 293, 308–311, 327  
Berkson (1963), 251, 252  
Bertero and Boccacci (1998), 324  
Besso (1879), 296  
Bienaymé (1840), 296  
Billingsley (1995), 154  
Birkhoff and Neumann (1936), 252  
Bishop (2006), 83, 149  
Blumenthal (1970), 291  
Bollobás (1999), 179–181, 218, 259, 263, 265, 266, 273, 281, 287, 288, 294–296, 299, 316, 317  
Bolstad (2007), 71  
Bouniakowsky (1859), 302  
Box and Jenkins (1976), 67  
Boyer and Merzbach (2011), 302  
Bracewell (1978), 222, 340  
Bromwich (1908), 357  
Bullen (2003), 295, 296, 299, 301, 303  
Bunyakovsky (1859), 260  
Burago et al. (2001), 252  
Byrne (2005), 235  
Cadzow (1987), 11, 12, 48  
Caines (1988), 8  
Caines (2018), 8  
Callahan (2010), 30  
NASA (2019), 77  
Candy and Sullivan (1992), 67  
Candy (2005), 67  
Candy (2009), 69, 142  
Candy (2016), 142  
Candy (2019), 67  
Cardano (1545), 195  
Carmona (2013), 149  
Carne and Dohrmann (2006), 112  
Carothers (2000), 300, 301, 303  
Carothers (2005), 211, 214  
Cauchy (1821), 260, 299, 302  
Cayley (1858), 179  
Chen et al. (2012), 130  
J.S.Chitode (2009b), 147  
J.S.Chitode (2009a), 353  
Choi (1978), 71  
Chorin and Hald (2009), 337, 338  
Christensen (2003), 209, 214, 222, 227, 232, 235  
Christensen (2008), 227, 232, 371, 374, 375, 380, 382, 384, 386, 388  
Christensen (2010), 374  
Chui (1988), 371, 380, 382,

- 388  
 Chui (1992), 350, 371, 375,  
 384  
 Cifarelli and Regazzini  
 (1996), 154  
 Cinlar and Vanderbei (2013),  
 288  
 Claerbout (1976), 361  
 Clarkson (1993), 25, 69–71,  
 73, 136, 145  
 Cobb (1988), 107, 110, 114  
 Cobb and Mitchell (1990),  
 114  
 Cohn (2002), 196  
 Conway (1990), 263  
 Courant and Hilbert (1930),  
 182  
 Cover and Thomas (1991), 6,  
 290  
 Cox (1972), 380, 382, 388  
 Cunningham and Ghahra-  
 mani (2015), 149  
 Curry and Feldman (2010),  
 163  
 Daubechies (1992), 215, 231,  
 353  
 Day (1973), 267  
 de Witt (1659), 181  
 de Boor (1972), 380, 382, 388  
 de Boor (2001), 380, 382, 388  
 de Finetti (1930a), 154  
 de Finetti (1930b), 154  
 de Finetti (1930c), 154  
 Dekking et al. (2006), 6  
 Descartes (1954), 179  
 Descartes (1637), 179  
 Devroye (1986), 29, 33  
 Deza and Deza (2006), 251  
 Daubechies et al. (1986), 232  
 Dieudonné (1969), 325  
 Drljević (1989), 272, 284  
 Duffin and Schaeffer (1952),  
 232  
 Dumitrescu (2007), 361  
 Dunford and Schwartz  
 (1957), 282, 284, 318  
 Durbin (2000), 195, 196  
 Duthie (1942), 285  
 Dutka (1995), 73  
 Eckart and Young (1936), 149  
 Edelman and Jamison  
 (1985), 291  
 Edelman (1986), 291  
 Edwards (1995), 218  
 Eidelman et al. (2004), 310  
 Enflo (1973), 214  
 Ewen (1950), viii  
 Ewen (1961), viii  
 Ewins (1986), 108, 130  
 Fabian et al. (2010), 220  
 Fano (1963), 104  
 Farina and Rinaldi (2000),  
 361  
 Feldman and Valdez-Flores  
 (2010), 163  
 Felippa (1999), 183  
 Feller (1970), 157  
 Fisher (1922), 85  
 Fix and Strang (1969), 343  
 Folland (1995), 253, 254, 257,  
 258  
 Forster and Massopust  
 (2009), 394  
 Fourier (1820), 293  
 Fourier (1878), 337  
 Fourier (1822), 337  
 Fréchet (1930a), 154  
 Fréchet (1930b), 154  
 Fredholm (1900), 20, 181  
 Fredholm (1903), 20, 181  
 Frobenius (1968), 325  
 Frobenius (1878), 325  
 Fuchs (1995), 253  
 Gelfand (1941), 257  
 Gelfand and Neumark  
 (1943b), 256–258  
 Gelfand and Neumark  
 (1943a), 257, 258  
 Gel'fand (1963), 332  
 Gelfand and Naimark (1964),  
 254, 255, 257  
 Gelfand et al. (2018), 332  
 Giles (1987), 277, 278  
 Giles (2000), 214, 218, 276,  
 278, 318, 320  
 Gohberg et al. (2003), 182  
 Goldman (1999), 109  
 Goswami and Chan (1999),  
 342  
 Goyder (1984), 110, 114  
 Gradshteyn and Ryzhik  
 (1980), 190  
 Graham (2018), 183, 187  
 Graham (1981), 183, 187–189  
 Gramacki (2017), 143, 144,  
 149  
 Gray (1971), 25  
 Gray (2006), 25  
 Greechie (1971), 154  
 Greenhoe (2017a), 378  
 Greenhoe (2019a), 13, 110  
 Grenander and Szegő (1958),  
 25  
 Grenander and Szegő (1984),  
 25  
 Gudder (1979), 252  
 Gudder (2005), 252  
 Gupta (1998), 213  
 Haar (1910), 400  
 Haaser and Sullivan (1991),  
 197, 198, 213, 240, 259–261,  
 264  
 Haddad and Akansu (1992),  
 365  
 Hall and Knight (1894), 304  
 Halmos (1948), 197  
 Halmos (1958), 240, 322  
 Halmos (1998b), 252  
 Halmos (1998a), 255, 267,  
 321  
 Hamel (1905), 199, 211  
 Han et al. (2007), 220, 235  
 Hardy (1929), 293  
 Hardy et al. (1952), 295, 300–  
 302  
 Haykin and Kesler (1979), 25  
 Haykin and Kesler (1983), 25  
 Haykin (1994), 30  
 Haykin (2014), 4, 148, 158  
 Hazewinkel (2000), 253  
 Heijenoort (1967), viii  
 Heil (2011), 209, 211, 213,  
 214, 222, 227, 232, 235, 263,  
 307, 374, 397  
 Heinbockel (2010), 392  
 Helstrom (1991), 48  
 Hijab (2011), 379  
 Hilbert (1904), 179  
 Hilbert (1906), 180, 181  
 Hilbert (1912), 179, 181  
 Hilbert et al. (1927), 308  
 Hölder (1889), 301  
 Holland (1970), 252  
 Höllig (2003), 379, 380, 382,  
 388  
 Hörmander (1994), 291  
 Horn and Johnson (1990), 77,  
 192, 312  
 Hotelling (1933), 149  
 Housman (1936), viii  
 Husimi (1937), 252  
 Ifeachor and Jervis (1993), 75  
 Ifeachor and Jervis (2002),  
 75, 147  
 Igari (1996), 230  
 Istrătescu (1987), 204, 259,  
 266, 267  
 Iturrioz (1985), 244, 250, 252  
 James (1945), 272, 282, 284  
 James (1947), 284  
 Jawerth and Sweldens  
 (1994), 342, 343, 388  
 Jazwinski (1970), 8, 165  
 Jazwinski (2007), 8, 165  
 Jeffers (1967), 149  
 Jeffrey and Dai (2008), 338  
 Jenkins and Watts (1968), 145  
 Jensen (1906), 6, 287, 288,  
 299  
 Johnstone and Silverman  
 (1997), 101  
 Jolley (1961), 391, 392  
 Jolliffe (1982), 150

- Jordan and von Neumann (1935), 265–267  
 Jørgensen et al. (2008), 232  
 Joshi (1997), 357  
 Kalmbach (1986), 154  
 Kamel and Sim (2004), 133  
 Kay (1988), 48, 71, 75, 136  
 Keener (1988), 20, 22, 23, 181, 324, ?, 149  
 Kenko (circa 1330), 408  
 Kenney (1947), 139  
 Khamsi and Kirk (2001), 291  
 Kim and Kamel (2003), 133  
 Knapp (2005a), 331  
 Knapp (2005b), 337, 338  
 Kolmogorov (1933b), 154  
 Kolmogorov (1933a), 154  
 Koopmans (1995), 48  
 Krein and Krasnoselski (1947), 251  
 Krein et al. (1948), 251  
 Kubrusly (2001), 197, 199, 209, 211, 213, 216, 222, 226, 244, 245, 248–250, 270, 273, 308, 322, 327, 328  
 Kubrusly (2011), 245, 246, 357  
 Kurdila and Zabarankin (2005), 209, 288  
 Lax (2002), 324  
 Lay (1982), 288  
 Leclerc et al. (2014), 113, 119  
 Lee and Verleysen (2007), 149  
 Leibniz (1679), 307  
 Leuridan et al. (1986), 107, 110  
 Liang and Lee (2015), 130  
 Lieb and Loss (2001), 192  
 von der Linden et al. (2014), 164  
 Lindenstrauss and Tzafriri (1977), 214  
 Lindquist and Picci (2015), 8  
 Liu (2013), 69  
 Loomis (1953), 270  
 Loomis (1955), 252  
 Loomis and Bolker (1965), 337  
 Machiavelli (1961), 407  
 Maeda and Maeda (1970), 286  
 Maleknejad et al. (2013), 377  
 Maligranda (1995), 264, 300, 301  
 Mallat (1989), 231  
 Mallat (1999), 231, 342, 343, 355, 365, 367, 395  
 Manolakis et al. (2000), 81  
 Manolakis et al. (2005), 48  
 S. Lawrence Marple (1987), 13, 25, 148  
 S. Lawrence Marple (2019), 13, 25  
 Massera and Schäffer (1958), 252  
 MatLab (2018a), 11  
 MatLab (2018b), 11  
 Mazur and Ulam (1932), 318  
 Mazur (1938), 257  
 Mendez (2019), 79  
 Menger (1928), 291  
 Menini and Oystaeyen (2004), 304  
 Mercer (1909), 182  
 Michel and Herget (1993), 20, 179, 181, 182, 196, 199–201, 209, 211, 213, 240–242, 244–246, 255, 256, 275, 278, 307, 309, 311, 314, 319, 320, 323, 325, 328–330, 335  
 Miller (2006), 163  
 Milovanović and Milovanović (1979), 295  
 Minkowski (1910), 264, 301  
 Mintzer (1985), 365  
 Mitchell (1980), 110  
 Mitchell (1982), 110, 111, 126  
 Mitrinović et al. (2010), 288  
 Moon and Stirling (2000), 8, 176  
 Mulholland (1950), 295  
 Nahin (2011), 392  
 Narens (2014), 154  
 Nelles (2001), 68, 80, 82  
 Noble and Daniel (1988), 332  
 Greenhoe (2017b), 378  
 Oikhberg and Rosenthal (2007), 280, 318  
 Ore (1935), 285  
 Osgood (2002), 169  
 Paine (2000), vi  
 Pap (1995), 154  
 Papoulis (1984), 11, 12, 20, 30, 32, 33, 48  
 Papoulis (1990), 6, 30, 35, 155, 156, 165, 172  
 Papoulis (2002), 48  
 Papoulis (1991), 19, 30, 32–34, 39–42, 45, 49, 54, 154, 156, 157, 164, 171  
 Peano (1888b), 197, 200, 201, 259  
 Peano (1888a), 200, 201  
 Pearson (1901), 149  
 Pečarić et al. (1992), 295  
 Pedersen (2000), 179, 217, 218  
 Pintelon and Schoukens (2012), 110, 114  
 Plackett (1972), 73  
 Poincaré (1912), 170  
 Popper (1962), 107  
 Popper (1963), 107  
 Prasad and Iyengar (1997), 371, 382, 394, 395  
 Proakis and Manolakis (1996), 147  
 Proakis (2002), 147  
 Proakis (2001), 30, 32, 33, 172–174, 176  
 Rana (2002), 296  
 Rao and Swamy (2018), 147  
 Rice (2012), 147  
 Rickart (1960), 254–257  
 Riesz (1934), 267  
 Robinson (1962), 361  
 Robinson (1966), 361  
 Rockafellar (1970), 288  
 Rosenblatt (1956), 71  
 Rossing (Springer), 67  
 Ross (1998), 6, 7, 171–173  
 Roussas (2014), 29  
 Rudin (1991), 312, 314, 315, 317, 319–322, 324–326, 330, 331  
 Rudin (1976), 298  
 Rynne and Youngson (2008), 213  
 Sakai (1998), 321  
 Salivahanan and Vallavaraj (2001), 147  
 Sasaki (1954), 252  
 Scargle (1979), 67, 73, 169  
 Scharf (1991), 183, 185, 188  
 Schauder (1927), 214  
 Schauder (1928), 214  
 Schumaker (2007), 373, 374, 377, 379, 380, 382, 388  
 Schur (1909), 330  
 Scharz (1885), 302  
 Schwarz (1885), 260  
 Searcoid (2002), 209  
 Shao (2003), 6, 71  
 Shin and Hammond (2008), 107, 110, 113, 114, 119, 126, 128  
 Shizgal (2015), 377  
 Silverman (1986), 70, 71  
 Simmons (2007), 407  
 Simon (2011), 287  
 Singer (1970), 214  
 Slaughter and Lennes (1915), 120  
 Smith and Barnwell (1984a), 365  
 Smith and Barnwell (1984b), 365  
 Smylie et al. (1973), 25  
 Snell (1997), 163  
 Snell (2005), 163  
 Sohrab (2003), 213

- Sorenson (1970), 73  
Sorzano et al. (2014), 149  
Srinath et al. (1996), 69, 91, 95  
Steele (2004), 262  
Steen (1973), 179, 330  
Stigler (1981), 73  
Stoer and Bulirsch (2002), 374  
Stone (1932), 308, 320, 324  
Strang and Nguyen (1996), 365, 396, 397  
Strang (1989), 396, 397  
Stuart and Ord (1991), 71, 139  
Suhov et al. (2005), 6  
Sweldens and Piessens (1993), 342, 343  
Szabó and Horváth (2004), 377  
Thomson et al. (2008), 357  
Thong et al. (2001), 133  
Tolsted (1964), 300, 301  
Trivedi (2016), 163
- Tsybakov (2008), 144  
Tucker (1965), 139  
Udriste (1994), 288  
UEA (2019), 77  
Ulam (1991), 318  
Vaidyanathan (1993), 365  
Väisälä (2003), 318  
Trees (2002), 183, 185, 188  
Vaseghi (2000), 69, 145  
van de Vel (1993), 286, 291  
Vidakovic (1999), 101, 144, 231, 342, 343, 355, 367  
von Neumann (1929), 181, 320, 324  
von Neumann (1960), 252  
Walter and Shen (2001), 101, 218  
Wasserman (2006), 149  
Wasserman (2013), 144, 149  
Weber (1893), 196  
Wedderburn (1907), 241  
Wei and Billings (2006), 377  
Wells (1987), 392  
Weyl (1940), 218
- White et al. (2006), 110, 113, 114, 128  
Wicks and Vold (1986), 113, 128  
Widom (1965), 25  
Wilks (1963a), 139, 140  
Wilks (1963b), 4, 7, 11, 139  
Wojtaszczyk (1997), 227, 229–231, 380, 382, 387, 388, 394, 400  
Wold (1938), 169  
Wold (1954), 169  
Wornell and Oppenheim (1992), 101  
Yan and Ren (2012), 109  
Yates and Mather (1963), 11  
Young (1912), 300  
Young (1980), 220  
Young (2001), 211, 213–215, 220, 224, 225, 227, 232  
Zhang et al. (2016), 109  
Zheng et al. (2002), 111, 113, 133  
Zhou and Wahab (2018), 109

## SUBJECT INDEX

0 order B-spline, 389  
0 order B-spline dyadic decomposition, 396  
0th order B-spline, 373  
1st order B-spline, 374, 389  
1st order B-spline dyadic decomposition, 397  
2nd order B-spline, 374, 389  
2nd order B-spline dyadic decomposition, 397  
3rd order B-spline, 377  
4th order B-spline, 377  
5th order B-spline, 377  
6th order B-spline, 377  
7th order B-spline, 378  
 $C^*$  algebra, 257, 258  
 $C^*$ -algebra, 258  
 $\Gamma(b)$ , 171  
\*-algebra, 12, 16, 39–41, 254, 254–256, 321, 348  
 $\langle \Delta | \nabla \rangle$ , 259  
 $\alpha$ th-order modified Bessel function of the first kind, 175  
 $n$ th derivative, 384  
 $n$ th moment, 342, 342, 343  
 $n$ th order B-spline, 375, 380, 382, 388, 389, 395, 397, 398  
 $n$ th order cardinal B-spline, 371  
 $n$ th-order B-spline, 387  
 $N$ -tuple, 199, 275  
 $\phi$ -mean, 294  
(ESD), 48  
\*-algebra, 74  
\*-algebras, 321  
 $\text{\LaTeX}$ , vi  
 $\text{\TeX-Gyre Project}$ , vi  
 $\text{\Xe\LaTeX}$ , vi  
**attention markers**, 270, 318  
    problem, 179, 312, 317, 324, 326  
 $\sigma$ -additive, 154, 160

$\sigma$ -additivity, 154  
 $\sigma$ -algebra, 155, 160, 162  
0 order B-spline, 386  
1st order B-spline, 387  
2 coefficient case, 400  
2nd order B-spline, 387  
4th central moment, 139  
  
Abel, 302  
Abel, Niels Henrik, 407  
abelian group, 198  
absolute value, x, 196, 204, 205–207, 260, 275, 277  
absolutely summable, 16, 358, 359  
ADC, 109  
additive, 22, 84, 86, 88, 89, 154–157, 159, 247, 259, 260, 270, 272, 308, 311  
additive Gaussian, 88  
Additive Gaussian noise projection statistics, 88  
additive identity, 12, 71, 86, 199–201, 309  
additive identity element, 199, 200  
additive inverse, 200, 201, 309  
additive noise model, 67, 68  
Additive noise projection statistics, 88  
additive property, 338  
additive white, 89  
additive white Gaussian, 90  
Additive white noise projection statistics, 89  
additivity, 22, 154, 155, 236, 263, 320  
adjoint, 22, 256, 317, 317, 320, 321, 338  
admissibility condition, 400  
admissibility equation, 400

Adobe Systems Incorporated, vi  
affine, 180, 287, 318  
Affine equations, 187  
affine equations, 75  
algebra, 195, 196, 197, 253, 253, 254  
algebra of sets, xi, 162  
algebraic, 197  
algebras  
     $C^*$ -algebra, 257  
    \*-algebra, 254  
algebras of sets, 162  
algorithm, 67, 67  
Alternating Series Test, 394  
AM-GM inequality, 299  
Analog Devices ADL5387  
Quadrature Demodulator, 14  
analog-to-digital converter, 109  
AND, xi  
anti-derivative, 372  
anti-symmetric, 14, 52, 250, 341  
anti-symmetric, 52  
antiautomorphic, 15, 254, 255, 260, 321  
antilinear, 255  
antitone, 156, 159, 245, 247, 250, 251  
aperature, 251  
AR, 145  
Aristotle, 11, 19  
arithmetic, 138  
arithmetic mean, 111, 135, 137, 299  
arithmetic mean estimate, 71  
arithmetic mean geometric mean inequality, 299  
Arithmetic Mean transfer

function estimate, 110  
 ARMA, 145  
 associates, 198  
 associative, 5, 195, 196, 198, 200, 293, 294, 310, 311, 330  
 asymmetric, 361  
 asymptotically unbiased, 145  
 auto-correlation, 12, 14, 21, 52  
 auto-correlation function, 20  
 auto-correlation matrix, 25  
 auto-correlation operator, 21, 23, 25  
 auto-power spectrum, 389  
 auto-spectral density, 15, 17  
 autocorrelation, 324  
 autocorrelation matrix, 74  
 Avant-Garde, vi  
 average, 136, 136–139, 142, 299  
 average power, 48  
 AWGN, 92, 95, 96  
 AWGN projection statistics, 90  
 AWN, 85, 90  
 Axiom of Continuity, 154  
 B-spline, 38, 377, 378, 384, 385, 387, 388, 395, 398, 402–404  
 B-spline dyadic decomposition, 395, 401  
 B-spline general form, 382  
 B-spline partition of unity, 388  
 B-spline recursion, 380  
 B-spline scaling coefficients, 398  
 B-splines, 389  
 Bak, Per, 121  
 ball, 286  
 Banach algebra, 257  
 Banach space, 214, 215  
 bandwidth, 144  
 Bartlett Method, 147  
 basis, 23, 24, 88, 89, 211–213, 220, 388  
     frame, 232  
     orthogonal, 219  
     orthonormal, 219  
     Riesz, 227  
     tight frame, 232  
 basis sequence, 387  
 basis vector, 83  
 Battle-Lemarié orthogonalization, 231  
 Bayes' Rule, 158  
 Bayesian, 69  
 Bayesian estimate, 70  
 Bessel's correction, 139  
 Bessel's Equality, 217, 218, 219, 223  
 Bessel's Inequality, 225  
 Bessel's inequality, 218  
 Best Approximation Theorem, 218, 218  
 biased, 110, 139  
 biased sample variance, 139, 139  
 bijection, 225  
 bijective, xi, 225, 318  
 bilinear functionals, 266  
 binomial coefficient, 373, 374  
 Binomial Theorem, 5, 71, 141  
 binomial theorem, 399  
 biorthogonal, 227, 229, 230  
 Birkhoff orthogonality, 282, 284, 284  
 Blackman-Tukey Method, 147  
 Bohr, Harald, 293  
 Boole's inequality, 157  
 Boolean, 155, 157  
 Boolean algebra, 161  
 Boolean lattice, 154, 157  
 Borel algebra on  $\mathbb{R}$ , 161  
 Borel measure, 337  
 Borel sets, 337  
 bounded, xi, 156, 215, 315, 324, 335  
 bounded bijective, 228  
 bounded lattice, 157  
 bounded linear operator, 331  
 bounded linear operators, 315, 317, 319, 320, 322, 323, 325, 326, 328–331  
 bounded operator, 315  
 C star algebra, 257  
 calculus, 197  
 Cardano, Gerolamo, 195  
 Carl Spitzweg, 407  
 Cartesian product, x  
 Cauchy, 249, 250  
 Cauchy Schwartz Inequality, 117, 118  
 Cauchy-Bunyakovsky-Schwarz Inequality, 260  
 Cauchy-Schwartz inequality, 9, 137  
 Cauchy-Schwarz Inequality, 100, 137, 260, 260, 263, 264  
 Cauchy-Schwarz inequality, 233, 303  
 Cauchy-Schwarz Inequality for inner-product spaces, 303  
 Cauchy-Schwarz inequality for inner-product spaces, 303  
 Cauchy-Schwarz Inequality for sequences, 302, 303  
 Cauchy-Schwarz inequality for sequences, 303  
 causal, 362, 362  
 cdf, 29, 164, 164, 165  
 Chain Rule, 97, 379  
 chain rule, 31, 189  
 characteristic function, x  
 Chebyshev's inequality, 7  
 Chi-square Distribution, 143  
 chi-square distribution, 172  
 Chi-squared distribution, 172  
 chi-squared distribution, 172  
 Chi-squared distribution with  $n$  degrees of freedom, 173, 173  
 chi-squared distribution with  $n$  degrees of freedom, 173  
 Cholesky Decomposition, 25  
 closed, 247, 249, 250  
 closed interval, 285, 286, 286  
 Closed Set Theorem, 247  
 closed sets, 242  
 closed support, 382  
 closure, 213, 382  
 coefficient functionals, 214, 214  
 coefficients, 363  
 coherence, 129, 130  
 colored, 21, 101  
 communication system, 124  
 communication system model, 68  
 communications additive noise cross-correlation, 59  
 communications additive noise model, 108, 121  
 Communications additive noise system, 59  
 communications LTI additive noise model, 125, 133  
 commutative, 12, 158, 196, 198, 205, 310, 357, 358  
 commutative group, 198  
 commutative ring, 195, 196, 196  
 commutativity, 358  
 compact, 181  
 compact support, 395  
 comparable, 285  
 comparison test, 391  
 complement, x  
 complemented, 250, 251  
 complete, 213  
 complete metric space, 213  
 complete set, 213  
 complex, 118

complex coherence, 130  
 complex envelope, 13  
 complex exponential, 385  
 complex linear space, 198  
 complex transmissibility, 112, 112, 130  
 complex-valued, 13, 112  
 concave, 286, 286, 287, 287, 288, 288, 295, 296  
 conditional cumulative distribution function, 164, 164  
 conditional probability, 70, 87, 157, 158  
 conditional probability density function, 164, 164  
 conditional probability of  $x$  given  $y$ , 156  
 conjugate symmetric, 20  
 conjugate linear, 254, 255, 321  
 conjugate pairs, 363  
 conjugate quadrature filter, 365, 365  
 conjugate quadrature filter condition, 365  
 Conjugate quadrature filters, 365  
 conjugate symmetric, 14, 16, 20, 52, 182, 259, 260, 269, 270, 272  
 conjugate symmetric property, 338  
 conjugate symmetry, 22  
 conjugate-symmetric, 16  
 consistent, 136, 137, 139, 143, 145  
 constant, 12, 47, 70  
 Continuity, 278  
 continuity, 165, 166  
 continuous, xi, 29, 158, 165, 182, 257, 263, 278, 294, 296, 310, 384, 390, 397  
 Continuous data whitening, 100  
 continuous point spectrum, 385  
 continuous spectrum, 180  
 contradiction, 200  
 convergence, 391  
 convergence in probability, 23  
 convergent, 48, 215  
 converges, 391–393  
 converges absolutely, 182  
 Convex, 288  
 convex, 5, 6, 240, 241, 249, 278, 280, 286, 287, 287, 288, 288, 295, 296  
     functional, 287  
     strictly, 287  
 convex function, 287, 288  
 convex set, 286, 286–288  
 convexity, 249  
 convolution, 35, 36, 39, 40, 42, 49, 61, 172, 340, 357, 357, 358, 371  
 convolution operation, 340  
 convolution operator, 357, 358  
 Convolution Theorem, 42, 386  
 convolution theorem, 340, 345, 360  
 coordinate functionals, 214  
 coordinates, 211  
 correlated, 56, 108, 109  
 correlation matrix, 25  
 cost function, 67, 69, 74, 121  
 countably infinite, 181, 214  
 counting measure, xi  
 CQF, 365, 365, 366  
 CQF condition, 365, 367  
 CQF theorem, 365, 368  
 Cramér-Rao Bound, 93  
 Cramér-Rao bound, 96  
 Cramér-Rao Inequality, 96  
 Cramér-Rao lower bound, 94  
 criterion function, 67  
 critical point, 30, 30, 32, 33  
 cross spectral density, 15  
 cross-correlation, 12, 14, 20, 20  
 cross-correlation vector, 74  
 CS Inequality, 236  
 CSD, 15  
 cummulative distribution function, 164, 164  
 cumulative distribution function, 29, 158  
 data dimensionality reduction, 144  
 data vector, 74  
 de Morgan, 251  
 de-noising, 70  
 deconvolution, 70  
 decreasing, 295, 394  
 definitions  
      $C^*$  algebra, 257, 258  
     \*-algebra, 12, 16, 39–41, 254, 348  
     nth order cardinal B-spline, 371  
         absolute value, 204  
         algebra, 196, 253  
         Banach algebra, 257  
         biorthogonal, 227  
         Birkhoff orthogonality, 284  
         bounded linear operators, 315  
         C star algebra, 257  
         closed interval, 285, 286  
 coefficient functionals, 214  
 commutative ring, 196  
 complete, 213  
 complex linear space, 198  
 conditional probability of  $x$  given  $y$ , 156  
 continuous spectrum, 180  
 convex set, 286  
 coordinate functionals, 214  
 coordinates, 211  
 CQF, 365  
 difference, 200  
 disjoint, 206  
 disjoint complement, 207  
 eigenspace, 179  
 epigraph, 287  
 equivalent, 215  
 even/odd die toss, 159  
 event, 155  
 exact frame, 232  
 expansion, 211, 214  
 field, 196  
 Fourier coefficients, 220  
 Fourier expansion, 220  
 Fourier series, 220  
 frame, 232  
 frame bounds, 232  
 fundamental, 213  
 group, 195  
 half-open interval, 285, 286  
 Hamel basis, 211  
 hermitian, 254  
 Hessian matrix, 193  
 hypograph, 287  
 inner product space, 259  
 intervals on lattices, 285  
 intervals on ordered sets, 285  
 isosceles orthogonality, 282  
 James orthogonality, 282  
 linear basis, 211  
 linear combination, 209  
 linear manifold, 240  
 linear space, 198  
 linear subspace, 198, 240  
 model, 67  
 model-based approach, 67  
 multiplicative condition, 257  
 negative part, 204

norm, 275  
 norm induced by the inner product, 265  
 normal, 254  
 normalized tight frame, 232  
 normed algebra, 257  
 normed linear space, 312  
 normed space of linear operators, 312  
 open interval, 285, 286  
 optimal lower frame bound, 232  
 optimal upper frame bound, 232  
 ordered linear space, 202  
 orthogonal basis, 220  
 orthogonal complement, 245  
 orthogonal in the Pythagorean sense, 284  
 orthogonal in the sense of Birkhoff, 284  
 orthogonal in the sense of James, 282  
 orthonormal basis, 220  
 outcome, 155  
 Parseval frame, 232  
 point spectrum, 180  
 positive, 202  
 positive cone, 202  
 positive part, 204  
 power mean, 296  
 probability space, 155, 161  
 projection, 254  
 Pythagorean orthogonality, 284  
 real linear space, 198  
 residual spectrum, 180  
 resolvent, 253  
 resolvent set, 180  
 Riesz basis, 227  
 Riesz space, 202  
 ring, 195  
 scalars, 198  
 Schauder basis, 214  
 self-adjoint, 254  
 Smith-Barnwell filter, 365  
 Sobelov Space, 144  
 space of all absolutely square summable sequences, 357  
 spans, 209  
 spectral radius, 253  
 spectrum, 180, 253  
 star-algebra, 254  
 subspace intersection, 241  
 subspace union, 241  
 summation, 293  
 tight frame, 232  
 total, 213  
 underlying set, 198  
 unital, 253  
 vector lattice, 202  
 vector space, 198  
 vectors, 198  
 delay, 334  
 dense, 213, 214  
 Descartes, René, ix  
 detection, 68  
 deterministic, 73  
 difference, x, 200  
 differentiable, 30  
 differentiable function, 30, 32, 33  
 differential operator, 385  
 dilation, 333  
 dilation equation, 398, 401  
 dimension, 211  
 dimensionality reduction, 149  
 Dirac delta, 86  
 direction, 80  
 Dirichlet Lambda function, 391, 392  
 Discrete data whitening, 100  
 Discrete Time Fourier Series, xii  
 Discrete Time Fourier Transform, xii, 15, 347  
 discrete-time Fourier transform, 347, 347–349, 353  
 discriminate, 262  
 disjoint, 206  
 disjoint complement, 207  
 distance, 277  
 distributes, 198  
 distributive, 12, 39–41, 74, 154–157, 200, 254–256, 259, 260, 321, 348, 359  
 distributive pseudocomplemented lattice, 154  
 domain, x  
 double angle formulas, 13, 95  
 Double coin toss, 159, 159  
 DTFT, 15, 16, 107, 215, 348, 352, 355, 365–367, 399  
 DTFT periodicity, 347  
 dyadic decomposition, 395  
 efficient, 92, 94, 96, 97, 143  
 eigen-decomposition, 149  
 eigen-system, 24  
 Eigendecomposition, 234  
 eigenspace, 179  
 eigenvalue, 179, 234  
 eigenvector, 179, 234  
 empty set, xi  
 energy, 361  
 energy spectral density, 48  
 ensemble average, 48  
 epigraph, 287  
 equal, 307  
 equalities  
     Bessel's, 217  
 equality  
     triangle, 276  
 equality by definition, x  
 equality relation, x  
 equations  
     parallelogram law, 267  
 equivalent, 215, 215, 227, 331  
 ergodic in the mean, 48, 48, 49  
 error vector, 74  
 estimate, 69, 107  
 estimated mean, 149  
 Euclidean metric, 280  
 Euclidean space, 286  
 Euler formulas, 345, 386  
 Euler's identity, 13  
 even, 352  
 even/odd die, 160  
 even/odd die probability space, 159  
 even/odd die toss, 159  
 event, 155, 155, 156  
 events, 155  
 exact frame, 232  
 examples  
     2 coefficient case, 400  
     Double coin toss, 159  
     even/odd die, 160  
     even/odd die toss, 159  
     Mercedes Frame, 235  
     Peace Frame, 235  
     Rectangular pulse, 345  
     rectangular pulse, 344, 386  
     single coin toss, 158, 160  
     The usual norm, 277  
     triangle, 345  
     tuples in  $\mathbb{F}^N$ , 199  
 excluded middle, 156, 157  
 exclusive OR, xi  
 existential quantifier, xi  
 exists, 48  
 expansion, 211, 214  
 expectation, 5, 9, 108  
 expectation operator, 3, 3, 4, 6  
 false, x  
 feature extraction, 83  
 FFT, 13  
 field, 195, 196, 197, 199, 357  
 field of complex numbers, 321  
 field of probability, 154  
 filter, 67

filter banks, 365  
finite modular lattice, 154  
FontLab Studio, vi  
for each, xi  
Fourier analysis, 384  
fourier analysis, 337  
Fourier coefficients, 220  
Fourier expansion, 220, 220, 221, 224, 225, 226  
Fourier kernel, 337  
Fourier Series, xi  
Fourier series, 220  
Fourier series expansion, 222  
Fourier Transform, xi, xii, 49, 231, 332, 334, 337, 338, 341, 386  
    adjoint, 338  
Fourier transform, 215, 334, 342, 344, 345, 385  
    inverse, 338  
Fourier transform magnitude, 386, 387  
Fourier transform scaling factor, 338  
Fourier, Joseph, 337  
frame, 230, 232, 232, 236  
frame bound, 233, 235  
frame bounds, 232, 236  
frame operator, 232, 232, 233, 235  
frames, 210  
Fredholm integral equation of the first kind, 20, 181  
Fredholm integral operators, 332  
Fredholm operator, 181, 181, 182  
Fredholm operators, 181  
Free Software Foundation, vi  
Frequency Response Function, 107  
Frequency Response Identification, 107  
FRF, 107  
function, 30, 67, 163, 164, 307, 337, 372  
    even, 352  
    step, 371  
function overload, 13  
functional, 259, 288, 321  
functions, xi  
    0 order B-spline, 389  
    0th order B-spline, 373  
    1st order B-spline, 374, 389  
    2nd order B-spline, 374, 389  
    3rd order B-spline, 377  
    4th order B-spline, 377  
    5th order B-spline, 377  
    6th order B-spline, 377  
    7th order B-spline, 378  
     $\alpha$ th-order modified Bessel function of the first kind, 175  
    nth derivative, 384  
    nth moment, 342  
    nth order B-spline, 375, 380, 382, 388, 389, 395, 397, 398  
    nth-order B-spline, 387  
     $\phi$ -mean, 294  
    (ESD), 48  
    0 order B-spline, 386  
    1st order B-spline, 387  
    2nd order B-spline, 387  
    4th central moment, 139  
    absolute value, 196, 205, 260, 275  
    adjoint, 256  
    anti-derivative, 372  
    arithmetic, 138  
    arithmetic mean, 111, 135, 137, 299  
    arithmetic mean estimate, 71  
    Arithmetic Mean transfer function estimate, 110  
    auto-correlation, 12, 14, 21, 52  
    auto-correlation function, 20  
    auto-power spectrum, 389  
    auto-spectral density, 17  
    average, 136, 136–139, 142, 299  
    average power, 48  
    B-spline, 38, 377, 378, 384, 385, 387, 388, 395, 398, 402–404  
    B-splines, 389  
    basis vector, 83  
    Bayesian estimate, 70  
    biased sample variance, 139, 139  
    binomial coefficient, 373, 374  
    Borel measure, 337  
    cdf, 29, 164, 164, 165  
    Chi-square Distribution, 143  
    chi-square distribution, 172  
    Chi-squared distribution with  $n$  degrees of freedom, 173  
    coherence, 129, 130  
    complex coherence, 130  
    complex envelope, 13  
    complex exponential,

385  
complex transmissibility, 112, 112, 130  
conditional cumulative distribution function, 164, 164  
conditional probability, 70, 87, 158  
conditional probability density function, 164, 164  
conjugate quadrature filter, 365  
continuous point spectrum, 385  
cost function, 67, 69, 74, 121  
criterion function, 67  
cross-correlation, 12, 14, 20  
cross-correlation vector, 74  
cummulative distribution function, 164, 164  
cumulative distribution function, 29, 158  
data vector, 74  
dilation equation, 398, 401  
Dirac delta, 86  
Dirichlet Lambda function, 391, 392  
Discrete Time Fourier Transform, 347  
discrete-time Fourier transform, 347–349  
    DTFT, 352, 355, 366, 367  
eigenvector, 234  
energy spectral density, 48  
error vector, 74  
estimate, 69  
estimated mean, 149  
Fourier kernel, 337  
Fourier transform, 342, 344, 345  
Fredholm integral equation of the first kind, 181  
    function, 67, 164  
Gamma Distribution, 143  
Gamma distribution, 172  
gamma function, 171  
Gaussian Distribution, 143  
    Gaussian distribution, 171, 176  
    generalized probability, 154, 155  
    Geometric mean, 111  
geometric mean, 111,

- 299**  
 geometric mean estimator, 111  
 Geometric mean transfer function estimate, 111  
 Hanning window, 137  
 Harmonic mean, 112, 113  
 harmonic mean, 299  
 Harmonic mean transfer function estimate, 112  
 impulse response, 15, 16, 39, 41, 42, 60, 107  
 induced norm, 223, 264  
 inner product, 8, 9, 137, 259, 264, 269, 337, 384  
 inphase component, 13, 14  
 input, 15, 16, 39, 41  
 integer sequence, 378  
 Inverse Method transfer function estimate, 110  
 Jensen's inequality, 5  
 joint cumulative distribution function, 164, 164  
 Joint Gaussian distribution, 175  
 Joint Gaussian pdf, 176  
 joint probability density function, 164, 164  
 kernel, 385  
 Kronecker delta, 24  
 Kronecker delta function, 271  
 kronecker delta function, 353  
 Least Squares Technique, 110  
 Least Squares transfer function estimate, 110  
 linear functional, 317  
 magnitude, 112  
 MAP estimate, 70, 83, 85  
 maximum a-posteriori probability estimate, 70  
 maximum likelihood estimate, 70  
 mean, 3, 12, 12, 20, 39  
 mean integrated square error, 71  
 mean square error, 70, 71, 137  
 Mean square estimate, 70  
 measure, 3, 154, 155, 158, 164, 165  
 measure-theoretic, 155  
 mini-max estimate, 70  
 ML estimate, 70, 83, 85, 87, 91, 92, 96  
 MM estimate, 70
- modulus, 196  
 MS estimate, 70  
 non-central chi-squared distribution, 174  
 noncentral chi-square distribution, 174  
 noncentral chi-square with  $n$ -degrees of freedom, 175  
 norm, 8, 9, 74, 137, 312  
 normal distribution, 171, 176  
 normalized rms error, 71, 71  
 operator norm, 312  
 order  $n$  B-spline, 378  
 order  $n$  B-spline MRA coefficient sequence, 398  
 order  $n$  B-spline MRA scaling sequence, 398  
 order-k moment estimate, 141  
 ordinary coherence, 130  
 ordinary transmissibility, 112  
 output, 15, 16, 39, 41  
 parameter vector, 74  
 pdf, 143, 164, 164, 165  
 phase, 112, 118  
 power mean, 296, 299  
 power spectral density, 17, 48  
 power sum, 378  
 probability, 155  
 probability density function, 3, 29, 38, 143, 158, 164, 164, 172  
 probability function, 154  
 processed data estimate vector, 74  
 processed data vector, 74  
 quadrature component, 13, 14  
 quantization noise, 109  
 quantum probability, 154, 155  
 random process, 19, 21, 23, 48, 83  
 random sequence, 11, 11, 12, 14, 15, 39, 42, 108  
 random variable, 3–8, 19, 29, 30, 32, 33, 35, 37, 158, 163, 164, 164, 165, 172  
 Rayleigh Distribution, 143  
 Rayleigh distribution, 175  
 Rice distribution, 175  
 Riesz sequence, 227
- sample variance, 139  
 scalar product, 259  
 scaling coefficient sequence, 400  
 scaling function, 397, 398  
 Scaling transfer function estimate, 113, 113  
 sequence, 11, 38, 357, 398  
 set indicator function, 346  
 signal-to-noise ratio, 109  
 sine sweep, 107  
 SNR, 109  
 spectral power, 17, 48, 109  
 step function, 371, 371–375  
 time series, 149  
 Total Least Squares transfer function estimate, 114  
 Total least squares transfer function estimate, 113  
 traditional probability, 154, 155  
 transfer function estimate  $\hat{H}_k(\omega; \kappa)$ , 113  
 transfer function estimate  $\hat{H}_c(\omega)$ , 114  
 translation operator, 343  
 Transmissibility, 112  
 transmissibility, 109, 109, 111  
 Uniform Distribution, 143  
 uniform distribution, 169, 169  
 variance, 3, 12, 12, 135  
 wavelet, 395  
 windowed auto-correlation estimate, 145  
 Z transform, 399  
 z transform, 361  
 Z-transform, 349  
 z-transform, 366  
 fundamental, 213  
 Fundamental Theorem of Calculus, 372, 379  
 Fundamental theorem of calculus, 165, 167  
 Fundamental theorem of linear equations, 311  
 Gamma Distribution, 143  
 Gamma distribution, 172, 172, 173  
 gamma function, 171, 171  
 gap metric, 251

Gaussian, 37, 84, 85, 87, 88, 90, 137, 143  
 Gaussian Distribution, 143  
 Gaussian distribution, 170, 171, 176  
 Gelfand-Mazur Theorem, 257  
 General ML estimation, 91  
 Generalized AM-GM inequality, 299  
 generalized arithmetic mean geometric mean inequality, 299  
 Generalized associative property, 293  
 Generalized Parseval's Identity, 222  
 generalized probability, 154, 155  
 Generalized triangle inequality, 9  
 geodesically between, 286  
 Geometric mean, 111  
 geometric mean, 111, 299  
 geometric mean estimator, 111  
 Geometric mean transfer function estimate, 111  
 Geometric Series, 304  
 geometry, 197  
 GNU Octave  
     as, 148  
     c, 138, 148  
     char, 80  
     col, 80, 138, 148  
     comment, 80  
     csv, 80  
     data, 138, 148  
     date, 80  
     end, 138  
     for, 403  
     format, 80  
     frequency, 148  
     function, 148  
     graphics, 148  
     install, 80  
     inv, 78, 80  
     length, 78, 80, 138, 148  
     lines, 138  
     load, 80  
     matrix, 148  
     max, 148  
     mean, 148  
     package, 148  
     packages, 80  
     plot, 78, 80, 138, 148  
     power, 148  
     print, 138, 148  
     read, 80  
     require, 80, 148  
     start, 138

step, 403  
 sum, 138  
 ts, 148  
 vector, 148  
 Golden Hind, vi  
 gradient, 80  
 gradient of  $y$  with respect to  $x$ , 183  
 gradient of  $y^T$  with respect to  $x$ , 183  
 greatest lower bound, xi  
 group, 195, 195  
 Gutenberg Press, vi  
 Hölder inequality, 303  
 Hölder's Inequality, 301, 302  
 half-open interval, 285, 286  
 Hamel bases, 210  
 Hamel basis, 211, 211, 213  
 Handbook of Algebras, 253  
 Hanning window, 137  
 Hardy, G.H., 293  
 harmonic analysis, 337  
 Harmonic mean, 112, 113  
 harmonic mean, 299  
 Harmonic mean transfer function estimate, 112  
 Hermetian transpose, 332  
 hermitian, 254, 254, 255, 324  
 hermitian components, 256  
 Hermitian representation, 256  
 Hermitian symmetric, 75, 341  
 Hermitian Transpose, 75  
 Hessian, 81  
 Hessian matrix, 192, 193  
 Heuristica, vi  
 high-pass filter, 365  
 Hilbert space, 220, 222, 224, 226, 227, 230, 232, 249, 250, 320, 321, 324–326, 331, 337  
 homogeneous, 5, 23, 196, 220, 221, 224, 228, 230, 247, 259, 260, 264, 269, 270, 272, 275, 281, 302, 308, 309, 311, 312  
 Housman, Alfred Edward, vii  
 hypograph, 287  
 hypotenuse, 284  
 identities  
     Fourier expansion, 220, 221  
     Parseval frame, 220, 221  
 identity, 195, 198, 307  
 identity element, 307  
 identity operator, 186, 307, 308  
 if, xi  
 if and only if, xi  
 image, x

image set, 309, 311, 322–326, 331  
 imaginary part, xi, 255  
 implied by, xi  
 implies, xi  
 implies and is implied by, xi  
 impossible, 388  
 impulse response, 15, 16, 39, 41, 42, 60, 107  
 inclusive OR, xi  
 increasing, 295, 394  
 independence, 87, 159, 160  
 independent, 7, 35, 38, 85, 87, 90, 91, 156, 159, 172  
 indicator function, x  
 induced norm, 23, 223, 264, 264  
 induces, 277  
 induction, 378, 380, 388  
 inequalities  
     AM-GM, 299  
     Cauchy-Bunyakovsky-Schwarz, 260, 302  
     Cauchy-Bunyakovsky-Schwarz Inequality, 260  
     Cauchy-Schwartz inequality, 9, 137  
     Cauchy-Schwarz, 260, 302  
     Cauchy-Schwarz Inequality, 100, 137, 260, 260, 263, 264  
     Cauchy-Schwarz inequality, 303  
     Cauchy-Schwarz Inequality for inner-product spaces, 303  
     Cauchy-Schwarz inequality for inner-product spaces, 303  
     Cauchy-Schwarz Inequality for sequences, 302, 303  
     Cauchy-Schwarz inequality for sequences, 303  
     Cramér-Rao Bound, 93  
     Cramér-Rao Inequality, 96  
     Generalized triangle inequality, 9  
     Hölder, 301  
     Hölder inequality, 303  
     Hölder's Inequality, 301, 302  
     Jensen's, 288  
     Jensen's Inequality, 295, 299  
     Minkowski (sequences), 301  
     Minkowski's Inequality, 264, 295

Minkowski's Inequality for sequences, 301  
 quadratic discriminant inequality, 262  
 triangle inequality for vectors, 272  
 Young, 300  
 Young's Inequality, 300  
 inequality  
 Bessel's, 218  
 triangle, 275, 312  
 infinite sum, 209  
 injective, xi, 310  
 inner product, 8, 9, 137, 197, 259, 264, 265, 269, 337, 384  
 Polarization Identity, 266  
 uniqueness, 265  
 inner product space, 8, 9, 197, 216–219, 227, 245, 246, 248, 259, 260, 263–265, 267, 271–273, 282, 384  
 inner-product, xi  
 inner-product space, 282, 284  
 inphase component, 13, 14  
 input, 15, 16, 39, 41  
 inside, 362  
 instantaneous error, 81  
 integer sequence, 378  
 integral, 384  
 integral domain, 195  
 integral test, 392  
 intersection, x, 155  
 interval, 285, 286  
 intervals on lattices, 285  
 intervals on ordered sets, 285  
 into, 225, 226  
 inverse, 195, 198, 233, 235, 400  
 inverse DTFT, 353, 367  
 Inverse Fourier Transform, 145, 332  
 Inverse Fourier transform, 338  
 inverse Fourier Transform, 334  
 Inverse Method transfer function estimate, 110  
 Inverse probability integral transform, 29  
 Inverse Z-Transform, 145  
 invertible, 30, 95, 158, 215, 232, 254  
 involuntary, 321  
 involution, 254, 254, 258  
 involutory, 16, 250, 251, 254–256  
 irreflexive ordering relation, xi  
 isometric, 221, 229, 230, 318,

327, 327, 331, 339  
 isometric in distance, 331  
 isometric in length, 331  
 isometric operator, 328–330  
 isometry, 327  
 isomorphic, 225  
 isosceles, 282  
 Isosceles orthogonality, 282  
 isosceles orthogonality, 282, 282–284  
 Jacobian matrix, 183, 183  
 James orthogonality, 282  
 Jensen's Inequality, 6, 288, 295, 299  
 Jensen's inequality, 5, 6, 6  
 join, xi  
 joint cumulative distribution function, 164, 164  
 Joint Gaussian distribution, 175  
 Joint Gaussian pdf, 176  
 joint probability density function, 164, 164  
 Kaneyoshi, Urabe, 407  
 Karhunen-Loëve Expansion, 23  
 KDE, 144, 149  
 Kenko, Yoshida, 407  
 kernel, 20, 385  
 Kernel Density Estimation, 144, 149  
 Kolmogorov's inequality, 7  
 Kronecker delta, 24  
 Kronecker delta function, 271  
 kronecker delta function, 186, 353  
 kronecker product, 184  
 kronecker products, 184  
 l'Hôpital's rule, 116, 119, 120, 298, 391, 393, 397  
 Laplace operator, 332  
 lattice, 202, 239, 244, 285, 286  
 lattice with negation, 154  
 Laurent series, 358  
 Law of the Unconscious Statistician, 6  
 Least squares, 76  
 least squares, 76, 77  
 Least squares estimation, 121  
 least squares estimation, 81  
 Least squares estimations, 121  
 Least Squares Technique, 110  
 Least Squares transfer function estimate, 110  
 least upper bound, xi  
 least-squares, 121, 124  
 Lebesgue integral, 161  
 Lebesgue integration, 382, 383  
 Lebesgue square-integrable functions, 337  
 left distributive, 5, 128, 195, 196, 311  
 Leibnitz GPR, 368  
 Leibniz Integration Rule, 380  
 Leibniz, Gottfried, ix, 307  
 length, 277  
 linear, 4, 5, 39, 40, 58, 59, 61, 85, 90, 95, 108, 118, 308, 308, 362  
 linear basis, 211  
 linear bounded, xi  
 linear combination, 209  
 linear combinations, 210  
 linear functional, 317  
 linear independence, 216  
 linear manifold, 240  
 linear operator, 4, 5, 56, 58, 59, 228, 230  
 linear operators, 226, 308, 317  
 linear ops., 226, 230  
 Linear space, 239  
 linear space, 197, 198, 198, 199, 209, 211, 240–242, 244, 245, 253, 259, 265, 275, 281, 285–288, 307  
 linear spaces, 307  
 linear span, 209, 240, 241  
 linear subspace, 198, 209, 239, 240, 240–245, 248  
 linear subspaces, 242  
 linear time invariant, 41, 45, 109, 121, 125, 126, 385  
 linear time invariant systems, 49  
 linear time-invariant, 61  
 linearity, 5, 22, 24, 25, 31, 34, 35, 49, 55, 59–61, 71, 74, 84, 86–88, 122, 135, 139–141, 185, 309, 374–376, 393, 396  
 Linearity of E, 4  
 linearly dependent, 209, 211–213  
 linearly independent, 209, 211–213, 216  
 Lippmann, Bernard A., 170  
 Liquid Crystal, vi  
 low-pass filter, 365  
 lower bounded, 157  
 LTI, 14, 16, 39–42, 49, 60–63, 109, 110, 121, 122, 125–127, 133, 385  
 operations on non-stationary random processes, 45

operations on WSS random processes, 47  
MA, 145  
Machiavelli, Niccolò, 407  
magnitude, 112  
MAP, 69, 85  
MAP estimate, 70, 83, 85, 87  
maps to, x  
Markov's inequality, 6, 7  
MatLab, 398  
matrix  
    rotation, 332  
matrix algebra, 74  
matrix calculus, 183  
matrix:quadratic form, 188, 189  
Maxima, 375, 377, 378  
maximum a-posteriori, 69  
maximum a-posteriori probability estimate, 70  
maximum a-posteriori probability estimation, 85  
maximum likelihood, 69  
maximum likelihood estimate, 70  
maximum likelihood estimation, 85, 90  
    general, 91  
    phase, 95  
Mazur-Ulam theorem, 318  
mean, 3, 3, 12, 12, 20, 39, 138  
mean integrated square error, 71  
mean square error, 70, 71, 137  
Mean square estimate, 70  
measurable space, 155  
measure, 3, 154, 155, 158, 164, 165  
measure space, 155, 156  
measure-theoretic, 155  
measurement additive noise cross-correlation, 59  
measurement additive noise model, 108, 121  
Measurement additive noise system, 59  
measurement LTI additive noise model, 125, 133  
measurement noise model, 68  
measurement system, 124  
meet, xi  
Mercedes Frame, 235  
Mercer's Theorem, 23, 182, 182  
method, 67  
metric, xi, 197, 277, 278  
    Euclidean, 280  
    generated by norm, 278  
    induced by norm, 278

parabolic, 280  
sup, 280  
taxi-cab, 280  
metric induced by the norm, 278  
metric linear space, 197, 277, 278, 280  
metric space, 197, 278, 285, 286  
metrics  
    gap, 251  
    Schäffer, 252  
mini-max estimate, 70  
Minimum mean square estimation, 74  
Minimum phase, 369  
minimum phase, 42, 53, 53, 361, 361, 362  
    energy, 361  
minimum variance unbiased estimator, 71  
Minkowski addition, 241  
Minkowski's Inequality, 264, 295  
Minkowski's Inequality for sequences, 301  
ML, 69, 85  
ML amplitude estimation, 91  
ML estimate, 70, 83, 85, 87, 91, 92, 96  
ML estimation of a function of a parameter, 95  
ML phase estimation, 95  
MM estimate, 70  
model, 67  
model-based approach, 67  
modular, 154  
modulus, 196  
monotone, 154, 156, 158  
monotonically decreasing, 115  
Monte Carlo Integration, 69  
MRA, 398  
MRA system, 398, 400  
MS estimate, 70  
multi-variate, 149  
multiplicative condition, 257  
Multiresolution Analysis, 395  
multiresolution analysis, 386, 397  
mutually exclusive, 157, 160  
MVUE, 71  
narrowband, 13  
negative part, 204, 205–207  
Neumann Expansion Theorem, 319  
Newton's method, 80  
no input noise, 127  
no output noise, 127  
noise

colored, 100  
non-central chi-squared distribution, 174  
non-contradiction, 156  
non-isometric, 264  
non-isotropic, 248, 259, 260, 268, 270, 272  
non-linear, 58, 59, 108, 124  
non-negative, 21–23, 196, 259, 264, 268, 270, 272, 296, 312, 382, 388  
non-negative definite, 23  
non-positive, 262  
non-stationary, 48  
non-uniform sampling, 79  
non-zero, 400  
Noncentral chi-square distribution, 174  
noncentral chi-square distribution, 174  
noncentral chi-square with  $n$ -degrees of freedom, 175  
nondegenerate, 154, 156, 159, 160, 196, 275, 281, 312  
nonnegative, 154  
nonparametric, 68  
nonparametric density estimation, 143, 144  
norm, 8, 9, 74, 137, 197, 261, 263–265, 275, 275, 277, 278, 281, 312  
    induced by inner product, 264, 265  
    Polarization Identity, 266  
    usual, 277  
norm induced by the inner product, 265, 265  
normal, 181, 254, 324, 325, 325, 326, 331, 332  
normal distribution, 170, 171, 176  
normal operator, 325, 330  
normalized, 23, 24, 154, 156, 159, 160  
normalized rms error, 71, 71  
normalized tight frame, 232  
normed algebra, 257, 257, 258  
normed linear space, 197, 265, 272, 275–278, 280, 282, 284, 312  
normed linear spaces, 317, 327  
normed space of linear operators, 312  
NOT, x  
not ergodic in the mean, 48  
not included  
    metric, 278  
not invertible, 30

- not necessarily linear, 41, 55  
 not symmetric, 383  
 not total, 223  
 not unique, 244  
 null space, x, 179, 309–311, 320, 322–326, 331  
 $O_6$  lattice, 155  
 Octave, 398  
 one sided shift operator, 329  
 one-to-one and onto, 95  
 only if, xi  
 open ball, 278  
 open interval, 285, 286  
 opening, 251  
 operations  
     0 order B-spline dyadic decomposition, 396  
     1st order B-spline dyadic decomposition, 397  
     2nd order B-spline dyadic decomposition, 397  
     adjoint, 22, 317, 317, 320  
     algorithm, 67, 67  
     AR, 145  
     ARMA, 145  
     auto-correlation, 21  
     auto-correlation matrix, 25  
     auto-correlation operator, 21, 23, 25  
     auto-spectral density, 15  
     autocorrelation matrix, 74  
     Bartlett Method, 147  
     Blackman-Tukey Method, 147  
     Cholesky Decomposition, 25  
     closure, 382  
     communications additive noise model, 108  
     communications LTI additive noise model, 125, 133  
     Continuous data whitening, 100  
     convolution, 35, 36, 39, 40, 42, 61, 172, 357, 358, 371  
     convolution operation, 340  
     correlation matrix, 25  
     cross spectral density, 15  
     CSD, 15  
     data dimensionality reduction, 144  
     de-noising, 70  
     deconvolution, 70  
     detection, 68  
     differential operator, 385  
     dimensionality reduction, 149  
     Discrete data whitening, 100  
     Discrete Time Fourier Series, xii  
     Discrete Time Fourier Transform, xii, 15  
     discrete-time Fourier transform, 347  
     DTFT, 15, 16, 107, 215, 348, 365, 367, 399  
     eigen-decomposition, 149  
     Eigendecomposition, 234  
     ensemble average, 48  
     estimate, 107  
     expectation, 5, 9, 108  
     expectation operator, 3, 3, 4, 6  
     feature extraction, 83  
     FFT, 13  
     filter, 67  
     Fourier analysis, 384  
     Fourier Series, xi  
     Fourier Transform, xi, xii, 49, 231, 332, 334, 338, 341  
     Fourier transform, 215, 385  
     Fourier transform magnitude, 386, 387  
     frame operator, 232, 232, 233, 235  
     Fredholm integral equation of the first kind, 20  
     Fredholm operator, 181  
     Frequency Response Function, 107  
     Frequency Response Identification, 107  
     FRF, 107  
     gradient of  $y$  with respect to  $x$ , 183  
     gradient of  $y^T$  with respect to  $x$ , 183  
     Hermetian transpose, 332  
     Hermitian Transpose, 75  
     identity operator, 308  
     imaginary part, 255  
     integral, 384  
     inverse, 233, 235, 400  
     Inverse Fourier Transform, 145, 332  
     inverse Fourier Transform, 334  
     Inverse Z-Transform, 145  
 involution, 254  
 Jacobian matrix, 183  
 KDE, 144, 149  
 kernel, 20  
 Kernel Density Estimation, 144, 149  
 Laplace operator, 332  
 Least squares estimation, 121  
 Least squares estimations, 121  
 Lebesgue integral, 161  
 Lebesgue integration, 382, 383  
 linear operator, 4, 5, 56, 58, 59  
 linear operators, 317  
 linear span, 209  
 MA, 145  
 measure, 155  
 measurement additive noise model, 108  
 measurement LTI additive noise model, 125, 133  
 method, 67  
 Monte Carlo Integration, 69  
 non-uniform sampling, 79  
 nonparametric density estimation, 143, 144  
 operator, 232, 307  
 operator adjoint, 321  
 parametric density estimation, 143  
 PCA, 149, 150  
 Periodogram, 147  
 power spectral density, 15  
 prediction, 70  
 Principal Component Analysis, 149  
 processor, 67  
 project, 384  
 projection, 323  
 projections, 83  
 PSD, 15  
 rationalizing the denominator, 120  
 real part, 255  
 reflection operator, 332  
 regression matrix, 74  
 rotation matrix, 332  
 rotation operator, 332  
 sampling, 4  
 Sum, 36  
 summation operator, 288  
 support, 382  
 system identification, 107  
 time average, 48



- Toeplitz matrices, 25  
 unitary Fourier Transform, 338  
 variance, 48  
 vector addition, 185  
 Welch Method 0% overlap, 147  
 Welch Method 50% overlap, 147  
 z-domain cross spectral density, 15  
 z-domain power spectral density, 15  
 Z-Transform, xii  
 z-transform, 358, 358  
 Z-transform operator, 15  
 operator, 197, 232, 307  
 adjoint, 255  
 autocorrelation, 324  
 bounded, 315  
 definition, 307  
 delay, 334  
 dilation, 333  
 identity, 307  
 isometric, 328–330  
 linear, 308  
 norm, 312  
 normal, 325, 326, 330  
 null space, 322  
 positive, 335  
 projection, 322  
 range, 322  
 self-adjoint, 324  
 shift, 329  
 translation, 332  
 unbounded, 315  
 unitary, 330, 331, 339  
 operator adjoint, 320, 321  
 operator norm, xi, 312  
 operator star-algebra, 321  
 optimal lower frame bound, 232  
 optimal receiver, 85  
 optimal upper frame bound, 232  
 order, x, xi  
 order  $n$  B-spline, 378  
 order  $n$  B-spline MRA coefficient sequence, 398  
 order  $n$  B-spline MRA scaling sequence, 398  
 order relation, 239, 285  
 order-k moment estimate, 141  
 ordered linear space, 202  
 ordered pair, x  
 ordered set, 250, 285, 286  
 ordinary coherence, 130  
 ordinary transmissibility, 112  
 orthocomplemented lattice, 156, 251  
 orthogonal, 21, 22, 24, 216, 219, 220, 272, 272, 273, 324, 384  
 orthogonal basis, 220  
 orthogonal complement, 245, 245, 250  
 orthogonal in the Pythagorean sense, 284  
 orthogonal in the sense of Birkhoff, 284  
 orthogonal in the sense of James, 282  
 orthogonality, 216, 282, 284  
 Birkhoff, 284  
 inner product space, 271  
 isosceles, 282  
 James, 282  
 Pythagorean, 284  
 orthomodular, 154, 251  
 orthomodular identity, 250, 251  
 orthomodular lattice, 154, 250, 251  
 orthonomal, 272  
 orthonormal, 84, 88, 89, 217–223, 236, 272  
 orthonormal bases, 210  
 orthonormal basis, 83, 220, 224–229, 231, 394, 395  
 Orthonormal decomposition, 99  
 orthonormal expansion, 220, 224  
 orthonormal quadrature conditions, 353  
 orthonormality, 217, 228, 236  
 orthornormal basis, 221  
 outcome, 155  
 outcomes, 155  
 output, 15, 16, 39, 41  
 over estimate, 110  
 over-estimated, 126  
 parabolic metric, 280  
 parallelogram, 267, 282  
 Parallelogram Law, 9  
 Parallelogram law, 267, 270  
 parallelogram law, 249, 265, 267, 270  
 parameter vector, 74  
 parametric, 68  
 parametric density estimation, 143  
 Parseval frame, 220, 221, 232, 232  
 Parseval frames, 210  
 Parseval's equation, 339  
 Parseval's Identity, 222, 224, 226, 232  
 Parseval's Theorem, 17  
 partition, 157, 158  
 partition of unity, 38, 156, 388, 389, 400, 401  
 Pascal's identity, 376  
 Pascal, Blaise, 153  
 PCA, 149, 150  
 pdf, 143, 164, 164, 165  
 Peace Frame, 235  
 Per Enflo, 214  
 Perfect reconstruction, 220  
 periodic, 304, 347, 390  
 periodicity, 390  
 Periodogram, 147  
 phase, 112, 118  
 Plancheral's formula, 49, 91  
 Plancherel's formula, 339  
 Poincaré, Henri, 170  
 point spectrum, 180  
 pointwise addition, 199  
 Polar Identity, 263, 264, 283  
 polar identity, 217  
 Polarization Identities, 266  
 Polarization Identity, 265, 268–270  
 polarization identity, 273, 283  
 pole, 369  
 poles, 53, 363  
 polynomial approximation, 77  
 Popper, Karl, 107  
 positive, 22, 23, 202, 335  
 positive cone, 202  
 positive definite, 22, 23  
 positive part, 204, 205–207  
 power mean, 296, 296, 299  
 power rule, 393  
 power set, xi  
 power spectral density, 15, 17, 48  
 power sum, 378  
 Power Sums, 304  
 prediction, 70  
 Principal Component Analysis, 149  
 probability, 155  
 probability density function, 3, 29, 38, 143, 158, 164, 164, 172  
 probability function, 154  
 Probability integral transform, 29  
 probability measure, 155  
 probability space, 3–5, 8, 9, 11, 19, 37, 155, 155–160, 161, 164, 165, 172  
 probability space, 160  
 processed data estimate vector, 74  
 processed data vector, 74  
 processor, 67

Product Rule, 97  
 product rule, 185, 188, 381  
 project, 384  
 projection, 254, 323  
 projection operator, 322, 324  
 projection statistics  
     Additive *Gaussian* noise  
 channel, 88  
     Additive noise channel,  
 88  
     Additive white *Gaussian*  
 noise channel, 90  
     Additive white noise  
 channel, 89  
 Projection Theorem, 250,  
 251  
 projections, 83, 90  
 proper subset, x  
 proper superset, x  
 properties  
      $\sigma$ -additive, 154, 160  
      $\sigma$ -additivity, 154  
     absolute value, x  
     absolutely summable,  
 16, 358, 359  
     additive, 22, 84, 86, 88,  
 89, 154–157, 159, 247, 259,  
 260, 270, 272, 308, 311  
     additive Gaussian, 88  
     additive identity, 12, 71,  
 86, 200, 309  
     additive inverse, 200,  
 309  
     additive white, 89  
     additive white Gaussian,  
 90  
     additivity, 22, 154, 155,  
 236, 263, 320  
     admissibility condition,  
 400  
     admissibility equation,  
 400  
     affine, 180, 287, 318  
     algebra of sets, xi  
     AND, xi  
     anti-symmetric, 14, 52,  
 250, 341  
     anti-symmetric, 52  
     antiautomorphic, 15,  
 254, 255, 260, 321  
     antitone, 156, 159, 245,  
 247, 250, 251  
     associates, 198  
     associative, 5, 195, 196,  
 198, 200, 293, 294, 310, 311,  
 330  
     asymptotically unbiased,  
 145  
     AWGN, 92, 95, 96  
     AWN, 85, 90  
     basis, 88, 89  
     Bayesian, 69  
     biased, 110, 139  
     bijection, 225  
     bijective, 225, 318  
     biorthogonal, 229, 230  
     Birkhoff orthogonality,  
 284  
     Boole's inequality, 157  
     Boolean, 155, 157  
     Boolean lattice, 157  
     bounded, 156, 215, 315,  
 324, 335  
     bounded bijective, 228  
     Cartesian product, x  
     Cauchy, 249, 250  
     causal, 362  
     characteristic function,  
 x  
     closed, 247, 249, 250  
     closed support, 382  
     closure, 213  
     colored, 21, 101  
     commutative, 12, 158,  
 196, 198, 205, 310, 357, 358  
     commutativity, 358  
     compact, 181  
     compact support, 395  
     comparable, 285  
     complement, x  
     complemented, 250, 251  
     complex, 118  
     complex-valued, 13, 112  
     concave, 286, 286, 287,  
 287, 288, 295, 296  
     conjugate linear, 254,  
 321  
     conjugate quadrature  
     filter condition, 365  
     conjugate symmetric,  
 14, 16, 20, 52, 182, 259, 260,  
 269, 270, 272  
     conjugate symmetry, 22  
     conjugate-symmetric,  
 16  
     consistent, 136, 137,  
 139, 143, 145  
     constant, 12, 47, 70  
     Continuity, 278  
     continuity, 165, 166  
     continuous, 29, 158,  
 165, 182, 263, 278, 294, 296,  
 310, 384, 390, 397  
     convergence, 391  
     convergence in proba-  
 bility, 23  
     convergent, 48, 215  
     converges, 391–393  
     converges absolutely,  
 182  
     convex, 5, 6, 240, 241,  
 249, 278, 280, 286, 287, 287,

288, 288, 295, 296  
     convexity, 249  
     correlated, 56, 108, 109  
     countably infinite, 181,  
 214  
     counting measure, xi  
     CQF condition, 365, 367  
     critical point, 30, 30, 32,  
 33  
     de Morgan, 251  
     decreasing, 295, 394  
     dense, 213, 214  
     deterministic, 73  
     difference, x  
     differentiable, 30  
     differentiable function,  
 32, 33  
     distributes, 198  
     distributive, 12, 39–41,  
 74, 154–157, 200, 254–256,  
 259, 260, 321, 348, 359  
     domain, x  
     dyadic decomposition,  
 395  
     efficient, 92, 94, 96, 97,  
 143  
     empty set, xi  
     equal, 307  
     equality by definition, x  
     equality relation, x  
     equivalent, 215, 227, 331  
     ergodic in the mean, 48,  
 48, 49  
     even, 352  
     excluded middle, 156,  
 157  
     exclusive OR, xi  
     existential quantifier, xi  
     exists, 48  
     false, x  
     for each, xi  
     Fredholm operator, 181,  
 182  
     Fredholm operators, 181  
     Gamma distribution,  
 172  
     Gaussian, 37, 84, 85, 87,  
 88, 90, 137, 143  
     geodesically between,  
 286  
     greatest lower bound, xi  
     hermitian, 254, 255, 324  
     Hermitian symmetric,  
 75, 341  
     homogeneous, 5, 23,  
 196, 220, 221, 224, 228, 230,  
 247, 259, 260, 264, 269, 270,  
 272, 275, 281, 302, 308, 309,  
 311, 312  
     hypotenuse, 284  
     identity, 195, 198

- identity operator, 186  
 if, xi  
 if and only if, xi  
 image, x  
 imaginary part, xi  
 implied by, xi  
 implies, xi  
 implies and is implied by, xi  
 impossible, 388  
 inclusive OR, xi  
 increasing, 295, 394  
 independence, 87, 159, 160  
 independent, 7, 35, 38, 85, 87, 90, 91, 156, 159, 172  
 indicator function, x  
 injective, 310  
 inner-product, xi  
 inside, 362  
 intersection, x  
 into, 225, 226  
 inverse, 195  
 invertible, 30, 95, 158, 215, 232, 254  
 involuntary, 321  
 involution, 254, 258  
 involutory, 16, 250, 251, 254–256  
 irreflexive ordering relation, xi  
 isometric, 221, 229, 230, 318, 327, 327, 331, 339  
 isometric in distance, 331  
 isometric in length, 331  
 isomorphic, 225  
 Isosceles orthogonality, 282  
 isosceles orthogonality, 282–284  
 join, xi  
 kronecker delta function, 186  
 least upper bound, xi  
 least-squares, 121, 124  
 left distributive, 5, 128, 195, 196, 311  
 linear, 4, 5, 39, 40, 58, 59, 61, 85, 90, 95, 108, 118, 308, 308  
 linear independence, 216  
 linear space, 199  
 linear time invariant, 41, 109, 121, 125, 126, 385  
 linear time-invariant, 61  
 linearity, 5, 22, 24, 25, 31, 34, 35, 49, 55, 59–61, 71, 74, 84, 86–88, 122, 135, 139–141, 185, 309, 374–376, 393, 396  
 linearly dependent, 209, 211–213  
 linearly independent, 209, 211–213, 216  
 lower bounded, 157  
 LTI, 14, 16, 39–42, 60–63, 109, 110, 121, 122, 125–127, 133, 385  
 maps to, x  
 meet, xi  
 metric, xi  
 Minimum phase, 369  
 minimum phase, 42, 53, 361, 361, 362  
 minimum variance unbiased estimator, 71  
 modular, 154  
 monotone, 154, 156, 158  
 monotonically decreasing, 115  
 multi-variate, 149  
 mutually exclusive, 157, 160  
 MVUE, 71  
 narrowband, 13  
 no input noise, 127  
 no output noise, 127  
 non-contradiction, 156  
 non-isometric, 264  
 non-isotropic, 248, 259, 260, 268, 270, 272  
 non-linear, 58, 59, 108, 124  
 non-negative, 21–23, 196, 259, 264, 268, 270, 272, 296, 312, 382, 388  
 non-negative definite, 23  
 non-positive, 262  
 non-stationary, 48  
 non-zero, 400  
 nondegenerate, 154, 156, 159, 160, 196, 275, 312  
 nonnegative, 154  
 nonparametric, 68  
 normal, 181, 324, 325, 331, 332  
 normalized, 23, 24, 154, 156, 159, 160  
 NOT, x  
 not ergodic in the mean, 48  
 not invertible, 30  
 not necessarily linear, 41, 55  
 not symmetric, 383  
 not total, 223  
 not unique, 244  
 null space, x  
 one-to-one and onto, 95  
 only if, xi  
 operator norm, xi  
 order, x, xi  
 ordered pair, x  
 orthogonal, 21, 22, 24, 216, 220, 272, 272, 273, 324, 384  
 orthogonality, 216, 282, 284  
 orthomodular, 154, 251  
 orthomodular identity, 250, 251  
 orthonomal, 272  
 orthonormal, 84, 88, 89, 217, 218, 220–223, 236, 272  
 orthonormality, 217, 228, 236  
 over estimate, 110  
 over-estimated, 126  
 parallelogram, 267, 282  
 parallelogram law, 270  
 parametric, 68  
 Parseval's Theorem, 17  
 partition of unity, 38, 156, 388, 389, 400, 401  
 periodic, 304, 347, 390  
 periodicity, 390  
 pointwise addition, 199  
 Polarization Identity, 270  
 positive, 22, 23, 335  
 positive definite, 22, 23  
 power set, xi  
 proper subset, x  
 proper superset, x  
 pseudo-distributes, 198  
 Pythagorean orthogonality, 284  
 Pythagorean Theorem, 284  
 quadratic, 96  
 randomness, 163  
 range, x  
 real, 278, 283, 350, 363  
 real part, xi  
 real-valued, 16, 22, 23, 74, 112, 324, 341, 348, 349, 352  
 reality condition, 340  
 reflexive, 250  
 reflexive ordering relation, xi  
 regular, 181  
 relation, x  
 relational and, x  
 reverse triangle inequality, 276  
 right distributive, 5, 195, 196, 311  
 ring, 75  
 ring of sets, xi  
 scalar, 74

- scalar commutative, 196  
 self adjoint, 324  
 self-adjoint, 21–23, 324,  
 324  
 separable, 214, 215, 226  
 set of algebras of sets, xi  
 set of rings of sets, xi  
 set of topologies, xi  
 space of linear transforms, 309  
 span, xi  
 spans, 88, 211, 212  
 stability condition, 230,  
 232  
 stable, 362  
 stationary in correlation, 47  
     stationary in the 1st moment, 12  
     stationary in the 2nd moment, 12  
     stationary in the mean, 47, 48  
 stochastic, 73  
 Strang-Fix condition, 343  
 strictly concave, 288  
 strictly convex, 287, 288  
 strictly increasing, 296  
 strictly monotone, 158  
 strictly monotonic, 294,  
 296  
 strictly monotonic increasing, 117  
 strictly positive, 22, 275  
 strong convergence, 214, 224  
 subadditive, 154, 196,  
 264, 275, 302, 312  
 submultiplicative, 196  
 subset, x  
 sufficient, 83, 99  
 sufficient statistic, 85,  
 87, 101  
 sufficient statistics, 83  
 super set, x  
 surjective, 331  
 symmetric, 14, 52, 278,  
 341, 350, 382, 383, 390, 394  
 symmetric, 52  
 symmetric difference, x  
 symmetry, 272, 361  
 there exists, xi  
 tight frame, 233  
 time-invariant, 362  
 Toeplitz, 25  
 topology of sets, xi  
 total, 213, 222, 223, 227,  
 228  
 transitive, 250  
 translation invariant,
- 281  
 triangle inequality, 196,  
 285  
 triangle inequality, 275  
 true, x  
 unbiased, 71, 92, 93, 127,  
 135–140  
 uncorrelated, 21, 55–63,  
 84–87, 89, 90, 108–110, 122,  
 126, 128, 131, 135–137, 139,  
 143  
 under estimate, 110  
 under estimates, 110  
 under-estimated, 126  
 uniformly, 182  
 uniformly distributed,  
 29, 38  
 union, x  
 unique, 30, 200, 211,  
 214, 244, 248–250, 387  
 unit area, 382  
 unit length, 329, 331  
 unitary, 330, 330–332,  
 339  
 universal quantifier, xi  
 upper bounded, 154  
 vanishing 0th moment,  
 400  
 vector norm, xi  
 von Neumann-Jordan condition, 267  
 white, 21, 21, 84–87, 101  
 wide sense stationary, 12, 12, 14, 15, 41, 47, 47, 48,  
 56, 109, 135, 136  
 wide-sense stationary, 126  
 WSS, 12, 14, 16, 40–42,  
 47, 47, 55, 57–63, 109, 110,  
 122, 126, 135, 136, 139  
 zero at –1, 400  
 zero measurement error, 109  
 zero measurement noise, 109  
 zero-mean, 7, 25, 55–63,  
 84, 85, 88, 89, 122, 144  
 PSD, 15  
 pseudo-distributes, 198  
 PSF, 343, 344  
 pstricks, vi  
 Pythagorean orthogonality, 282, 284, 284  
 Pythagorean Theorem, 216, 217, 219, 221, 228, 272, 273, 284  
 Pythagorean theorem, 216  
 quadratic, 96  
 quadratic discriminant inequality, 262  
 Quadratic Equation, 129  
 quadratic equation, 262  
 Quadratic form, 188  
 quadratic form, 75, 188, 189  
 quadrature component, 13,  
 14  
 quantization noise, 109  
 quantum probability, 154,  
 155  
 quotes  
 Abel, 302  
 Abel, Niels Henrik, 407  
 Aristotle, 11, 19  
 Bak, Per, 121  
 Bohr, Harald, 293  
 Cardano, Gerolamo, 195  
 Descartes, René, ix  
 Fourier, Joseph, 337  
 Hardy, G.H., 293  
 Housman, Alfred Edward, vii  
 Kaneyoshi, Urabe, 407  
 Kenko, Yoshida, 407  
 Leibniz, Gottfried, ix,  
 307  
 Lippmann, Bernard A.,  
 170  
 Machiavelli, Niccolò,  
 407  
 Pascal, Blaise, 153  
 Poincaré, Henri, 170  
 Popper, Karl, 107  
 Russull, Bertrand, vii  
 Stravinsky, Igor, vii  
 Ulam, Stanislaus M., 318  
 Quotient Rule, 128, 129  
 quotient rule, 393  
 quotient structures, 285  
 random process, 19, 19, 21,  
 23, 48, 83  
 random sequence, 11, 11, 12,  
 14, 15, 39, 42, 108  
 random variable, 3–8, 19, 19,  
 29, 30, 32, 33, 35, 37, 158, 163,  
 164, 164, 165, 172  
 randomness, 163  
 range, x  
 range space, 320  
 rational expression, 53  
 rationalizing factor, 120  
 rationalizing the denominator, 120  
 Rayleigh Distribution, 143  
 Rayleigh distribution, 175,  
 175  
 real, 278, 283, 350, 363  
 real interval, 165  
 real linear space, 198  
 real numbers, 277  
 real part, xi, 255  
 real-valued, 16, 22, 23, 74,  
 112, 324, 341, 348, 349, 352

- reality condition, 340  
 Rectangular pulse, 345  
 rectangular pulse, 344, 386  
 recursively, 80  
 reflection, 318  
 reflection operator, 332  
 reflexive, 250  
 reflexive ordering relation, xi  
 regression matrix, 74  
 regular, 181  
 relation, x, 30, 307  
 relational and, x  
 relations, xi  
     order relation, 285  
 residual spectrum, 180  
 resolvent, 253  
 resolvent set, 180, 181  
 Reverse Triangle Inequality, 276, 278  
 reverse triangle inequality, 276  
 Rice distribution, 175, 175  
 Riesz bases, 210  
 Riesz basis, 227, 227–231, 386, 394, 395  
 Riesz sequence, 227, 229, 397  
 Riesz space, 202, 202, 204–207  
 Riesz-Fischer Theorem, 224  
 Riesz-Fischer Thm., 225, 226  
 right distributive, 5, 195, 196, 311  
 ring, 75, 195, 195, 196  
     absolute value, 196  
     commutative, 195  
     modulus, 196  
 ring of complex square  $n \times n$  matrices, 321  
 ring of sets, xi  
 Robinson's Energy Delay Theorem, 361  
 rotation matrix, 332  
 rotation operator, 332  
 Russull, Bertrand, vii  
 sample variance, 139  
 sampling, 4  
 scalar, 74  
 scalar commutative, 196  
 scalar product, 259  
 scalars, 198  
 scaling coefficient sequence, 400  
 scaling function, 113, 397, 398  
 scaling parameter, 113  
 Scaling transfer function estimate, 113, 113  
 Schauder bases, 210, 215  
 Schauder basis, 214, 214, 215, 220, 226  
 Schäffer's metric, 252  
 self adjoint, 324  
 self-adjoint, 21–23, 254, 324, 324  
 semilinear, 255  
 separable, 214, 215, 226  
 separable Hilbert space, 225, 227, 229, 357  
 separable Hilbert spaces, 225  
 sequence, 11, 38, 277, 357, 398  
 sequences, 365  
 set, 155, 197, 286  
 set indicator, 371  
 set indicator function, 346  
 set of algebras of sets, xi  
 set of rings of sets, xi  
 set of topologies, xi  
 sets  
     open ball, 278  
     real numbers, 277  
 shift operator, 329  
 shift relation, 344, 345  
 sigma-algebra, 155  
 Signal matching, 99  
 signal to noise ratio, 99  
 signal-to-noise ratio, 109  
 sinc, 344, 345  
 sine sweep, 107  
 single coin toss, 158, 160  
 Smith-Barnwell filter, 365  
 SNR, 109  
 Sobelov Space, 144  
 source code, 403  
 space, 155  
     inner product, 259  
     linear, 197  
     linear subspace, 240  
     metric, 278  
     Minkowski addition, 241  
     normed vector, 275  
     orthogonal, 245  
     vector, 197  
 space of all absolutely square Lebesgue integrable functions, 357  
 space of all absolutely square summable sequences, 357  
 space of linear transforms, 309  
 span, xi, 213, 215  
 spans, 88, 209, 211, 212  
 spectral power, 17, 48, 109  
 spectral radius, 253  
 Spectral Theorem, 181  
 spectrum, 180, 181, 253  
 stability, 362  
 stability condition, 230, 232  
 stable, 362  
 standard orthonormal basis, 221  
 star-algebra, 254, 254, 321  
 star-algebras, 320, 321  
 state, 80  
 stationary in correlation, 47  
 stationary in the 1st moment, 12  
 stationary in the 2nd moment, 12  
 stationary in the mean, 47, 48  
 statistical average, 81  
 steepest descent, 80, 81  
 step function, 371, 371–375  
 step size, 80  
 Stifel formula, 396  
 Stifel formula, 376  
 stochastic, 73  
 Strang-Fix condition, 343, 343  
 Stravinsky, Igor, vii  
 strictly concave, 288  
 strictly convex, 287, 287, 288  
 strictly increasing, 296  
 strictly monotone, 158  
 strictly monotonic, 294, 296  
 strictly monotonic increasing, 117  
 strictly positive, 22, 275, 281  
 strong convergence, 214, 224  
 structures  
      $C^*$  algebra, 257  
      $C^*$ -algebra, 258  
     \*-algebra, 12, 16, 39–41, 254, 254–256, 321, 348  
      $N$ -tuple, 199, 275  
     \*-algebra, 74  
     \*-algebras, 321  
      $\sigma$ -algebra, 155, 160, 162  
     absolute value, 206, 207  
     additive identity element, 199, 200  
     additive inverse, 200  
     additive noise model, 67, 68  
     adjoint, 321  
     algebra, 196, 197, 253, 253, 254  
     algebraic, 197  
     ball, 286  
     Banach space, 214, 215  
     basis, 23, 24, 211–213, 220, 388  
     basis sequence, 387  
     bijection, 225  
     bijective, 225  
     bilinear functionals, 266  
     Boolean algebra, 161  
     Boolean lattice, 154, 157  
     Borel algebra on  $\mathbb{R}$ , 161  
     Borel sets, 337  
     bounded lattice, 157  
     bounded linear operator, 331

bounded linear operators, 315, 317, 319, 320, 322, 323, 325, 326, 328–331  
 C star algebra, 257  
 calculus, 197  
 closed interval, 285, 286, 286  
 coefficient functionals, 214  
 coefficients, 363  
 communication system, 124  
 communication system model, 68  
 communications additive noise model, 121  
 Communications additive noise system, 59  
 commutative ring, 195, 196, 196  
 complete metric space, 213  
 complete set, 213  
 complex linear space, 198  
 conditional probability, 157  
 conditional probability of  $x$  given  $y$ , 156  
 conjugate pairs, 363  
 conjugate quadrature filter, 365  
 Conjugate quadrature filters, 365  
 continuous spectrum, 180  
 convex function, 287, 288  
 convex set, 286, 286–288  
 convolution operator, 357, 358  
 coordinates, 211  
 CQE, 365, 365, 366  
 critical point, 30  
 differentiable function, 30  
 discrete-time Fourier transform, 353  
 distributive pseudo-complemented lattice, 154  
 Double coin toss, 159  
 eigen-system, 24  
 eigenspace, 179  
 epigraph, 287  
 Euclidean space, 286  
 even/odd die probability space, 159  
 even/odd die toss, 159  
 event, 155, 155, 156  
 expansion, 211  
 field, 195, 196, 197, 199, 357  
 field of complex numbers, 321  
 field of probability, 154  
 filter banks, 365  
 finite modular lattice, 154  
 Fourier Transform, 337  
 frame, 230, 232, 232, 236  
 frames, 210  
 function, 30, 163, 337, 372  
 functional, 259, 288, 321  
 geometry, 197  
 group, 195, 195  
 half-open interval, 285, 286  
 Hamel bases, 210  
 Hamel basis, 211, 211, 213  
 high-pass filter, 365  
 Hilbert space, 220, 222, 224, 226, 227, 230, 232, 249, 250, 320, 321, 324–326, 331, 337  
 hypograph, 287  
 identity, 198  
 identity element, 307  
 image set, 309, 311, 322–326, 331  
 infinite sum, 209  
 inner product, 197, 265  
 inner product space, 8, 9, 197, 216–219, 227, 245, 246, 248, 259, 260, 263–265, 267, 271–273, 282, 384  
 inner-product space, 282, 284  
 integral domain, 195  
 interval, 285, 286  
 intervals on lattices, 285  
 intervals on ordered sets, 285  
 inverse, 198  
 isometry, 327  
 lattice, 239, 244, 285, 286  
 lattice with negation, 154  
 Laurent series, 358  
 Lebesgue square-integrable functions, 337  
 linear basis, 211  
 linear combination, 209  
 linear combinations, 210  
 linear operator, 228, 230  
 linear operators, 226  
 linear ops., 226, 230  
 Linear space, 239  
 linear space, 197, 198, 198, 209, 211, 240–242, 244, 245, 253, 259, 265, 275, 281, 285–288, 307  
 linear spaces, 307  
 linear span, 240, 241  
 linear subspace, 198, 209, 239, 240, 240–245, 248  
 low-pass filter, 365  
 matrix algebra, 74  
 measurable space, 155  
 measure, 155  
 measure space, 155, 156  
 measurement additive noise model, 121  
 Measurement additive noise system, 59  
 measurement noise model, 68  
 measurement system, 124  
 metric, 197, 278  
 metric linear space, 197, 277, 278, 280  
 metric space, 197, 278, 285, 286  
 model, 67  
 model-based approach, 67  
 MRA, 398  
 MRA system, 398, 400  
 Multiresolution Analysis, 395  
 multiresolution analysis, 386, 397  
 negative part, 205–207  
 norm, 197, 261, 263–265, 275, 277, 278, 281  
 normalized tight frame, 232  
 normed algebra, 257, 258  
 normed linear space, 197, 265, 272, 275–278, 280, 282, 284, 312  
 normed linear spaces, 317, 327  
 normed space of linear operators, 312  
 null space, 179, 310, 324–326, 331  
 $O_6$  lattice, 155  
 open ball, 278  
 open interval, 285, 286  
 operator, 197  
 order relation, 239  
 ordered set, 250, 285, 286  
 orthocomplemented lattice, 156, 251  
 orthogonal basis, 220  
 orthogonal complement, 245, 250

- orthomodular lattice, 154, 250, 251  
 orthonormal bases, 210  
 orthonormal basis, 83, 220, 224–229, 231, 394, 395  
 orthonormal basis, 221  
 outcome, 155  
 Parseval frame, 232  
 Parseval frames, 210  
 Parseval's equation, 339  
 partition, 157, 158  
 Plancherel's formula, 339  
 point spectrum, 180  
 pole, 369  
 poles, 363  
 positive part, 205–207  
 probability space, 3–5, 8, 9, 11, 19, 37, 155, 155–160, 161, 164, 165, 172  
 probability space, 160  
 projection operator, 324  
 quotient structures, 285  
 random variable, 164  
 real interval, 165  
 real linear space, 198  
 relation, 30  
 residual spectrum, 180  
 resolvent, 253  
 resolvent set, 180, 181  
 Riesz bases, 210  
 Riesz basis, 227–231, 386, 394, 395  
 Riesz sequence, 229, 397  
 Riesz space, 202, 204–207  
 ring, 195, 195, 196  
 ring of complex square  $n \times n$  matrices, 321  
 scalars, 198  
 Schauder bases, 210, 215  
 Schauder basis, 214, 214, 215, 220, 226  
 separable Hilbert space, 225, 227, 229, 357  
 separable Hilbert spaces, 225  
 sequence, 277  
 sequences, 365  
 set, 155, 197, 286  
 sigma-algebra, 155  
 Smith-Barnwell filter, 365  
 Sobolev Space, 144  
 space, 155  
 space of all absolutely square Lebesgue integrable functions, 357  
 space of all absolutely square summable se-
- quences, 357  
 span, 213, 215  
 spectral radius, 253  
 spectrum, 180, 181, 253  
 standard orthonormal basis, 221  
 star-algebra, 254, 321  
 star-algebras, 320  
 subset, 241, 246, 249  
 subspace, 241, 246–248, 250  
 support, 382  
 system, 14, 41, 55–63, 107, 109–114, 118, 121, 122, 124–128, 130  
 tight frame, 232, 233, 235  
 tight frames, 210  
 topological dual space, 317  
 topological linear space, 197, 213  
 topology, 155, 197, 209  
 total set, 213  
 trivial linear space, 240  
 underlying set, 198  
 unit vector, 229  
 unital  $*$ -algebra, 254  
 unital algebra, 253  
 vector, 211  
 vector additive identity element, 245  
 vector space, 8, 197, 198  
 vector space of random variables, 4  
 vectors, 198, 285  
 weights, 288  
 zero, 369  
 zeros, 361, 363  
 subadditive, 154, 196, 264, 275, 302, 312  
 submultiplicative, 196  
 subset, x, 241, 246, 249  
 subspace, 241, 246–248, 250  
 subspace intersection, 241  
 subspace union, 241  
 sufficient, 83, 99  
 sufficient statistic, 85, 87, 101  
 Sufficient Statistic Theorem, 85, 96  
 sufficient statistics, 83  
 Sum, 36  
 sum of products, 31, 35, 157, 165  
 summation, 293  
 Summation around unit circle, 234  
 summation operator, 288  
 sup metric, 280  
 super set, x  
 support, 382  
 surjective, xi, 331  
 symmetric, 14, 52, 278, 341, 350, 382, 383, 390, 394  
 conjugate, 20  
 symmetric, 52  
 symmetric difference, x  
 symmetry, 272, 361  
 system, 14, 41, 55–63, 107, 109–114, 118, 121, 122, 124–128, 130  
 system identification, 107  
 taxi-cab metric, 280  
 The basis problem, 214  
 The Book Worm, 407  
 The Fourier Series Theorem, 222  
 The usual norm, 277  
 theorems  
 Additive Gaussian noise projection statistics, 88  
 Additive noise projection statistics, 88  
 Additive white noise projection statistics, 89  
 Affine equations, 187  
 affine equations, 75  
 Alternating Series Test, 394  
 AWGN projection statistics, 90  
 B-spline dyadic decomposition, 395, 401  
 B-spline general form, 382  
 B-spline partition of unity, 388  
 B-spline recursion, 380  
 B-spline scaling coefficients, 398  
 Battle-Lemarié orthogonalization, 231  
 Bayes' Rule, 158  
 Bessel's Equality, 217, 218, 219, 223  
 Bessel's Inequality, 225  
 Bessel's inequality, 218  
 Best Approximation Theorem, 218, 218  
 Binomial Theorem, 5, 71, 141  
 binomial theorem, 399  
 Cauchy Schwartz Inequality, 117, 118  
 Cauchy-Schwarz inequality, 233  
 Chain Rule, 97  
 chain rule, 31  
 Chebyshev's inequality, 7  
 Closed Set Theorem, 247  
 communications addi-

- tive noise cross-correlation, 59  
 comparison test, 391  
 Convolution Theorem, 42, 386  
 convolution theorem, 340, 345, 360  
 CQF theorem, 365, 368  
 CS Inequality, 236  
 dilation equation, 398  
 double angle formulas, 13, 95  
 DTFT periodicity, 347  
 Euler formulas, 345, 386  
 Euler's identity, 13  
 Fourier expansion, 221, 224, 225, 226  
 Fourier series expansion, 222  
 Fundamental Theorem of Calculus, 372, 379  
 Fundamental theorem of calculus, 165, 167  
 Fundamental theorem of linear equations, 311  
 Gelfand-Mazur Theorem, 257  
 General ML estimation, 91  
 Generalized AM-GM inequality, 299  
 Generalized associative property, 293  
 Generalized Parseval's Identity, 222  
 Geometric Series, 304  
 Hermitian representation, 256  
 induced norm, 23, 264  
 induction, 378, 380, 388  
 integral test, 392  
 inverse DTFT, 353, 367  
 Inverse Fourier transform, 338  
 Inverse probability integral transform, 29  
 Jensen's Inequality, 6, 288  
 Jensen's inequality, 6, 6  
 Karhunen-Loëve Expansion, 23  
 Kolmogorov's inequality, 7  
 l'Hôpital's rule, 116, 119, 120, 298, 391, 393, 397  
 Law of the Unconscious Statistician, 6  
 Least squares, 76  
 Leibnitz GPR, 368  
 Leibniz Integration Rule, 380  
 Linearity of E, 4  
 Markov's inequality, 6, 7  
 Mazur-Ulam theorem, 318  
 measurement additive noise cross-correlation, 59  
 Mercer's Theorem, 23, 182, 182  
 Minimum mean square estimation, 74  
 ML amplitude estimation, 91  
 ML estimation of a function of a parameter, 95  
 ML phase estimation, 95  
 Neumann Expansion Theorem, 319  
 operator star-algebra, 321  
 orthonormal quadrature conditions, 353  
 Parallelogram Law, 9  
 Parallelogram law, 267  
 parallelogram law, 249, 267  
 Parseval's Identity, 222, 224, 226, 232  
 Pascal's identity, 376  
 Perfect reconstruction, 220  
 Plancheral's formula, 49, 91  
 Polar Identity, 263, 264, 283  
 polar identity, 217  
 Polarization Identities, 266  
 Polarization Identity, 268–270  
 polarization identity, 273, 283  
 power rule, 393  
 Power Sums, 304  
 Probability integral transform, 29  
 Product Rule, 97  
 Projection Theorem, 250, 251  
 PSF, 343, 344  
 Pythagorean Theorem, 216, 217, 219, 221, 228, 272, 273  
 Pythagorean theorem, 216  
 Quadratic Equation, 129  
 Quadratic form, 188  
 quadratic form, 75  
 Quotient Rule, 128, 129  
 quotient rule, 393  
 Reverse Triangle Inequality, 276, 278  
 Riesz-Fischer Theorem, 224  
 Riesz-Fischer Thm., 225, 226  
 Robinson's Energy Delay Theorem, 361  
 shift relation, 344, 345  
 Spectral Theorem, 181  
 Stifel formual, 396  
 Stifel formula, 376  
 Strang-Fix condition, 343  
 Sufficient Statistic Theorem, 85, 96  
 sum of products, 31, 35, 157, 165  
 Summation around unit circle, 234  
 The Fourier Series Theorem, 222  
 triangle inequality, 275  
 Wiener-Hopf equations, 75, 75  
 Wold's Theorem, 67, 169  
 Young's Inequality, 301  
 there exists, xi  
 tight frame, 232, 233, 235  
 tight frames, 210  
 time average, 48  
 time series, 149  
 time-invariant, 362, 362  
 Toeplitz, 25  
 Toeplitz matrices, 25  
 topological dual space, 317  
 topological linear space, 197, 213  
 topological space, 155, 277  
 topology, 155, 197, 209  
 topology of sets, xi  
 total, 213, 213, 222, 223, 227, 228  
 Total Least Squares transfer function estimate, 114  
 Total least squares transfer function estimate, 113  
 total set, 213  
 traditional probability, 154, 155  
 transfer function estimate  $\hat{H}_k(\omega; \kappa)$ , 113  
 transfer function estimate  $\hat{H}_c(\omega)$ , 114  
 transform  
   inverse Fourier, 338  
 transitive, 250  
 translation, 332  
 translation invariant, 281  
 translation operator, 343  
 Transmissibility, 112  
 transmissibility, 109, 109, 111

- triangle, 345  
 triangle inequality, 196, 275, 275, 285, 312  
 triangle inequality for vectors, 272  
 triangle inequality, 275  
 triangle-inequality, 282  
 trivial linear space, 240  
 true, x  
 tuples in  $\mathbb{F}^N$ , 199  
 Ulam, Stanislaus M., 318  
 unbiased, 71, 92, 93, 127, 135–140  
 uncorrelated, 21, 55–63, 84–87, 89, 90, 108–110, 122, 126, 128, 131, 135–137, 139, 143  
 under estimate, 110  
 under estimates, 110  
 under-estimated, 126  
 underlying set, 198  
 Uniform Distribution, 143  
 uniform distribution, 169, 169  
 uniformly, 182  
 uniformly distributed, 29, 38  
 union, x, 155  
 unique, 30, 200, 211, 214, 244, 248–250, 387  
 unit area, 382  
 unit length, 329, 331  
 unit vector, 229  
 unital, 253  
 unital  $*$ -algebra, 254  
 unital algebra, 253  
 unitary, 330, 330–332, 334, 338, 339  
 unitary Fourier Transform, 338  
 unitary operator, 330  
 universal quantifier, xi  
 upper bounded, 154  
 usual norm, 277  
 Utopia, vi  
 values  
 $n$ th moment, 342  
 bandwidth, 144  
 Bessel's correction, 139  
 Cramér-Rao bound, 96  
 Cramér-Rao lower bound, 94  
 dimension, 211  
 eigenvalue, 179, 234  
 eigenvector, 179  
 frame bound, 233, 235  
 frame bounds, 236  
 MAP estimate, 87  
 mean, 138  
 rationalizing factor, 120  
 scaling function, 113  
 scaling parameter, 113  
 Vandermonde matrix, 77  
 vanishing 0th moment, 400  
 vanishing moments, 342, 355  
 variance, 3, 3, 12, 12, 48, 135  
 vector, 211  
 vector addition, 185  
 vector additive identity element, 245  
 vector lattice, 202  
 vector norm, xi  
 vector space, 8, 197, 198  
 vector space of random variables, 4  
 vectors, 198, 285  
 von Neumann-Jordan condition, 267  
 wavelet, 395  
 weighted, 295  
 weights, 288  
 Welch Method 0% overlap, 147  
 Welch Method 50% overlap, 147  
 white, 21, 21, 84–87, 101  
 whitening filter, 42, 52  
 wide sense stationary, 12, 12, 14, 15, 41, 47, 47, 48, 56, 109, 135, 136  
 wide-sense stationary, 126  
 Wiener-Hopf equations, 75, 75  
 windowed auto-correlation estimate, 145  
 Wold's Theorem, 67, 169  
 Wronskian, 193  
 WSS, 12, 12, 14, 16, 40–42, 47, 47, 55, 57–63, 109, 110, 122, 126, 135, 136, 139  
 Young's Inequality, 300, 301  
 Z transform, 399  
 z transform, 361  
 z-domain cross spectral density, 15  
 z-domain power spectral density, 15  
 Z-Transform, xii  
 Z-transform, 349  
 z-transform, 358, 358, 366  
 Z-transform operator, 15  
 zero, 369  
 zero at  $-1$ , 350, 400  
 zero measurement error, 109  
 zero measurement noise, 109  
 zero-mean, 7, 25, 55–63, 84, 85, 88, 89, 122, 144  
 zeros, 53, 361, 363

## License

This document is provided under the terms of the Creative Commons license CC BY-NC-ND 4.0. For an exact statement of the license, see

<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>

The icon  appearing throughout this document is based on one that was once at <https://creativecommons.org/>

where it was stated, “Except where otherwise noted, content on this site is licensed under a Creative Commons Attribution 4.0 International license.”





...last page ...please stop reading ...

