Spectral Estimation and ASP

Lecture 2 - Complex-Valued Signal Processing

Danilo Mandic

room 813, ext: 46271

Department of Electrical and Electronic Engineering

Imperial College London, UK

d.mandic@imperial.ac.uk, URL: www.commsp.ee.ic.ac.uk/~mandic

Outline

Background on:

Complex-Valued Signal Processing

- ◦ Why a complex-valued solution in a real-valued world?
- ◦ History of complex numbers.

Part 1: Complex Calculus

- ◦ Cauchy-Riemann equations
- ◦ $\mathbb{CR}$-Calculus and its application

Part 2: Complex Statistics

- ◦ Data model: Gaussian
- ◦ Moving from real to complex
- ◦ Circularity and widely linear estimation
- ◦ Covariance and pseudocovariance
- ◦ Widely linear autoregressive model

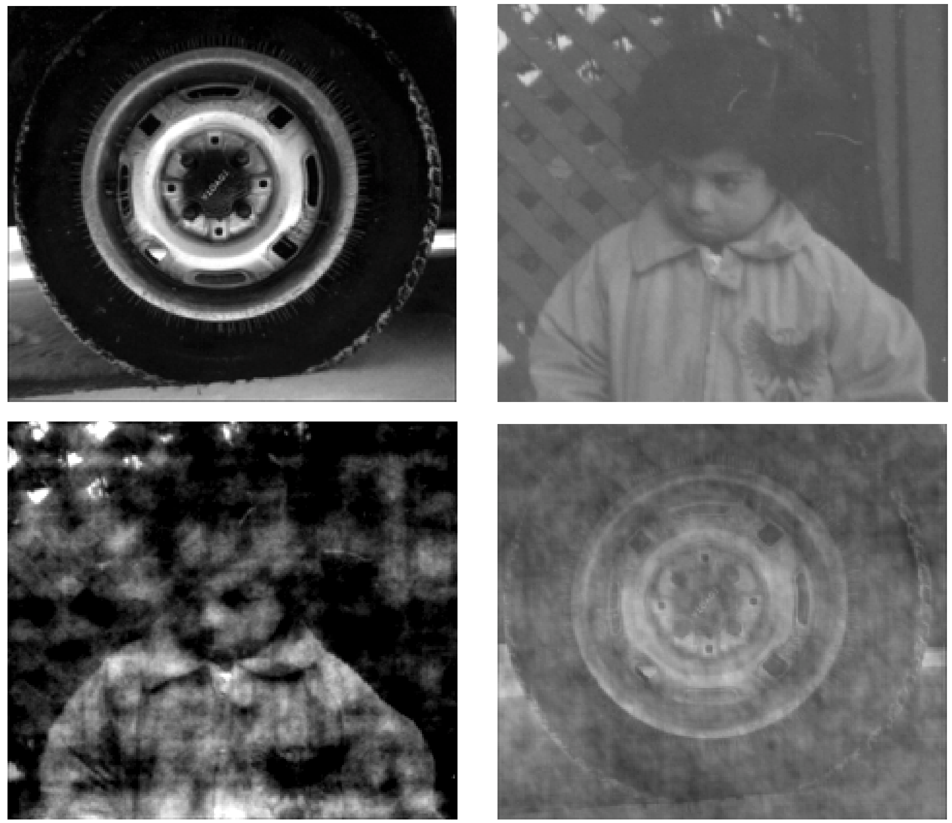
Motivation for Modelling in $\mathbb{C}$

Complex Numbers are Everywhere

- Magnetic source imaging (fMRI, MRI, MEG) are recorded in the Fourier domain
- Interferometric radar - high coherence in order to obtain both the altitude and amplitude introduces speckles
- Array signal processing, antennas, direction of arrival (DoA)
- Transform domain signal processing (DCT, DFT, wavelet)
- Mobile communications (equalisation, I/Q mismatch, nonlinearities)
- Homomorphic fitering – we like zero mean signals, but in $\mathbb{R}$ log does not exist for $x \leq 0$ but $\log z = \log |z| + \jmath arg(z)$ does
- Optics and seismics - reflection, refraction
- Fractals, chaos, and iterated maps (associative memory)
- Much work still to be done – **great opportunity!**

Example: Human Visual System

Importance of Phase Information



Surrogate images. *Top:* Original images I_1 and I_2; *Bottom:* Images $\hat{I}_1$ and $\hat{I}_2$ generated by exchanging the amplitude and phase spectra of the original images.

Usefulness of Complex Numbers

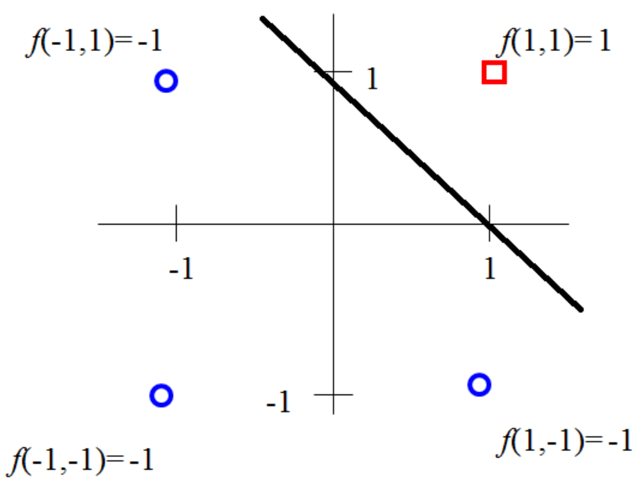
Example: Nonlinear separability of the logical problem XOR

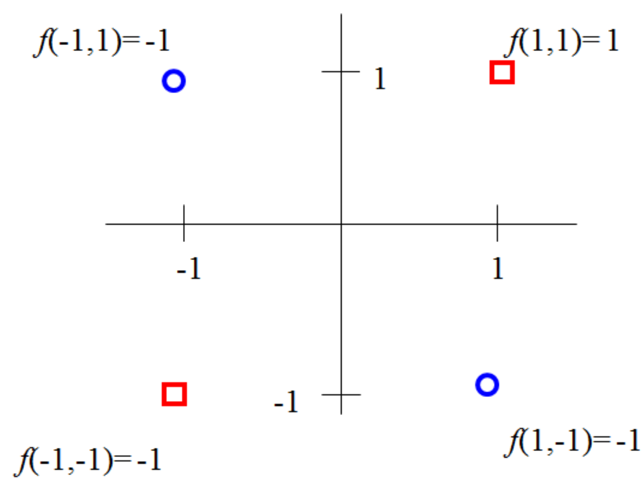
x_1	x_2	z	$P(z)$ = XOR
1	1	$1+\jmath$	1
1	-1	$1-\jmath$	-1
-1	1	$-1+\jmath$	-1
-1	-1	$-1-\jmath$	1

$$P(z) = \begin{cases} 1, & \arg(z) \text{ 1st or 3rd quadrants} \\ -1, & \arg(z) \text{ 2nd or 4th quadrants.} \end{cases}$$

For example, the AND function is linearly separable with a single neuron in $\mathbb{R}$

The XOR function needs a multilayer network in $\mathbb{R}$ but a single neuron in $\mathbb{C}$

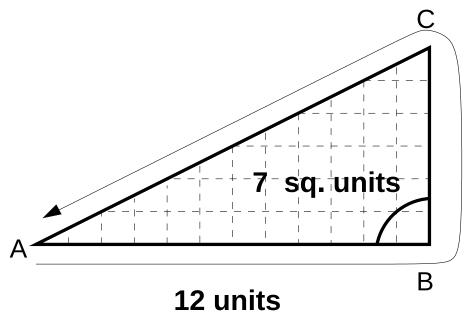




History of Complex Numbers

Find a triangle of Area = 7 and Perimeter = 12

- Heron of Alexandria (60 AD)



To solve this, let the side $|AB| = x$, and the height $|BC| = h$, to yield

$$
\begin{aligned}
area &= \frac{1}{2} x \ h \\
perimeter &= x + h + \sqrt{x^2 + h^2}
\end{aligned}
$$

In order to solve for x we need to find the roots of

$$6x^2 - 43x + 84 = 0$$

However, this equation does not have real roots.

The Depressed Cubic (so called 'cubic formula')

○ In the 16th century Niccolo Tartaglia and G. Cardano considered closed formulas for the roots of third- and fourth-order polynomials.
○ Cardano first introduced complex numbers in his book *Ars Magna* in 1545, as a tool for finding roots of the 'depressed cubic' $x^3 + ax + b = 0$.

$$ay^3 + by^2 + cy + d = 0 \quad \text{substitute} \quad y = x - \frac{1}{3}b \quad \Rightarrow \quad x^3 + \beta x + \gamma = 0$$

○ Scipione del Ferro of Bologna and Tartaglia showed that the depressed cubic can be solved as

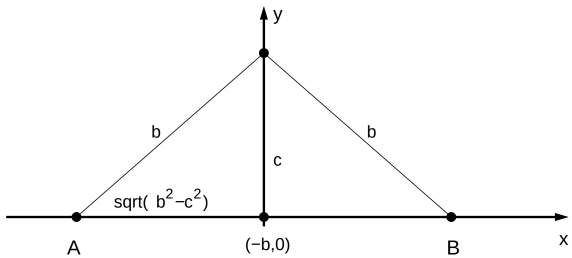
$$x = \sqrt[3]{-\frac{\gamma}{2} + \sqrt{\frac{\gamma^2}{4} + \frac{\beta^3}{27}}} + \sqrt[3]{-\frac{\gamma}{2} - \sqrt{\frac{\gamma^2}{4} + \frac{\beta^3}{27}}}$$

Tartaglia's formula for the roots of $x^3 - x = 0$ is $\frac{1}{\sqrt{3}}\left((\sqrt{-1})^{\frac{1}{3}} + \frac{1}{(\sqrt{-1})^{\frac{1}{3}}}\right)$.
○ In 1572, in his *Algebra*, while solving for $x^3 - 15x - 4 = 0$, R. Bombelli arrived at $\left(2 + \sqrt{-1}\right) + \left(2 - \sqrt{-1}\right) = 4$ and introduced the symbol $\sqrt{-1}$.
○ In 1673 John Wallis realised that the general solution for the form $x^2 + 2bx + c^2 = 0$ is

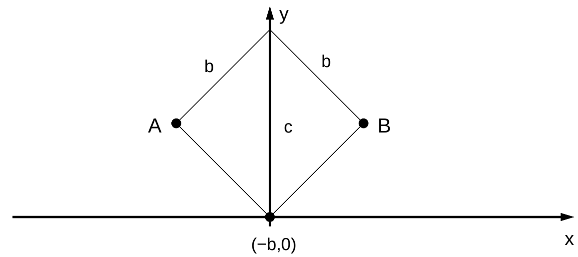
$$x = -b \pm \sqrt{b^2 - c^2}$$

The Role of Geometry

◦ Complex numbers were only accepted after they had a geometric interpretation, but it was only possible for $b^2 - c^2 \geq 0$.
◦ Wallis - complex number a point on the plane (solutions A & B)



Real solution



Complex solution

◦ In 1732 Leonhard Euler, $x^n - 1 = 0 \;\rightarrow\; \cos\theta + \sqrt{-1}\sin\theta$
◦ Abraham de Moivre (1730) and again Euler (1748), introduced the famous formulas

$$\begin{aligned}(\cos\theta + \jmath\sin\theta)^n &= \cos n\theta + \jmath\sin n\theta \\ \cos\theta + \jmath\sin\theta &= e^{\jmath\theta}\end{aligned}$$

◦ In 1806 Argand interpreted $\jmath = \sqrt{-1}$ as a rotation by 90^o and introduced Argand diagram, $z = x + \jmath y$, and the modulus $\sqrt{x^2 + y^2}$.
◦ In 1831 Karl Friedrich Gauss introduced $\imath = \sqrt{-1}$ and complex algebra.

History of Mathematical Notation

Did you know?

- $9th$ century Al Kwarizimi's *Algebra* - solutions descriptive rather than in form of equations

- $16th$ century - G. Cardano *Ars Magna* - unknowns denoted by single roman letters

- Descartes (1630-s) established general rules
  - lowercase italic letters at the beginning of the alphabet for unknown constants a, b, c, d
  - lowercase italic letters at the end of the alphabet for unknown variables x, y, z

- $\sqrt{-1} = \imath$ – Gauss 1830s, boldface letters for vectors $\mathbf{x}, \mathbf{v}$ - Oliver Heaviside

- Hence $\quad ax^2 + by + cz = 0$

More detail: F. Cajori, *History of Mathematical Notations*, 1929

Fundamental Theorem of Algebra (FTA)

∘ Initial work by Albert Girard in 1629

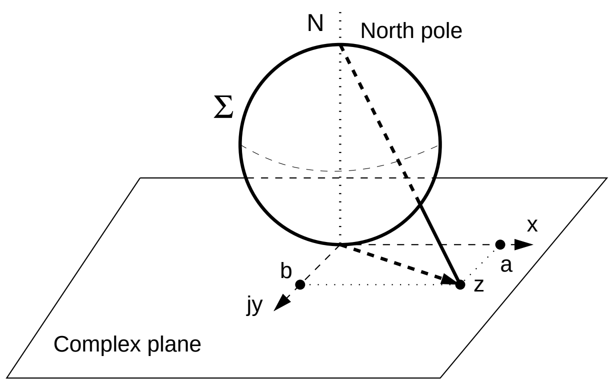
'there are n roots to an n-th order polynomial'

He also introduced the abbreviations $\sin, \cos, \tan$ in 1626.

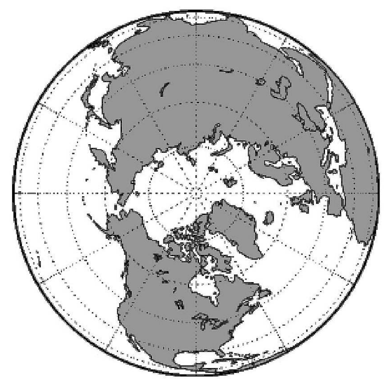
∘ Descartes in the 1630s *'For every equation of degree n we can imagine roots which do not correspond to any real quantity'*

∘ In 1749 Euler proved the FTA

Every n-th order polynomial in $\mathbb{R}$ has exactly n roots in $\mathbb{C}$



(e) Riemann sphere



(f) Earth projection from South pole

Stereographic projection and Riemann sphere

∘ Cauchy $\rightarrow$ *'conjugate'*, Hankel $\rightarrow$ *'direction'*, Weierstrass $\rightarrow$ *'absolute value'*

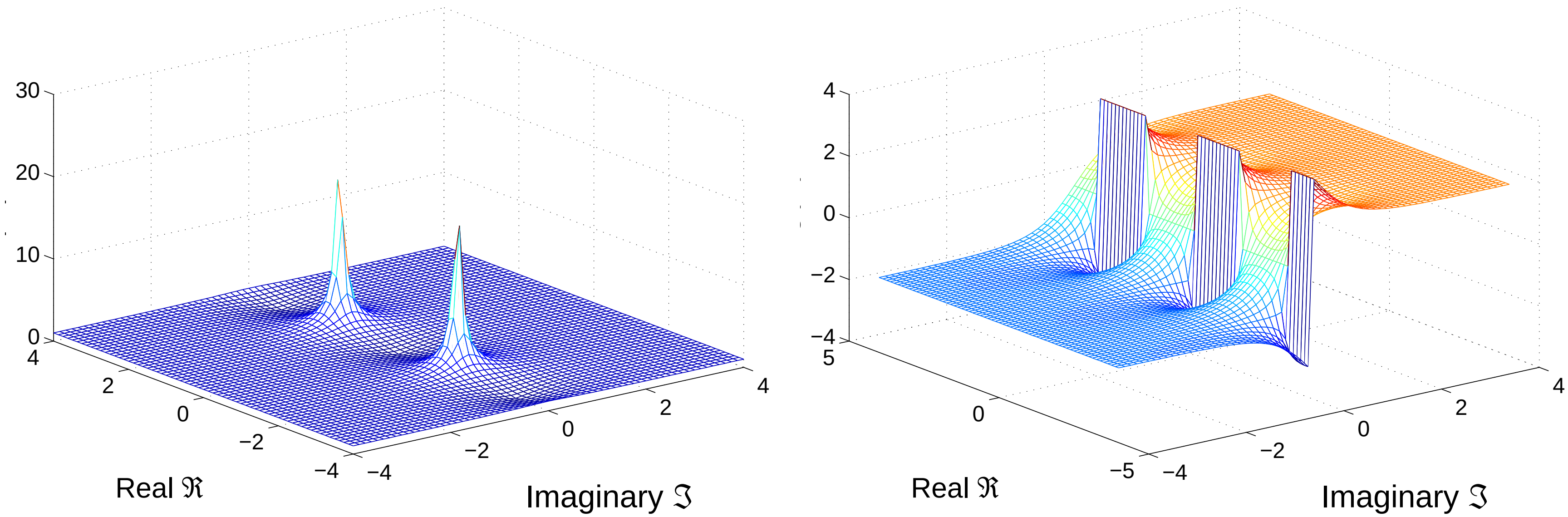
Modern complex estimation: Numerous opportunities

- Complex signals by design (communications, analytic signals, equivalent baseband represenation to eliminate spectral redundancy)
- By convenience of representation (radar, sonar, wind field), direction of arrival related problems
- Problem: Different and more powerful algebra but no ordering (operator "$\leq$" makes no sense!) and the notion of pdf has to be induced from $\mathbb{R}^2$
- Problem: Special form of nonlinearity (the only continuously differentiable function in $\mathbb{C}$ is a constant (Liouville theorem)
- Solution: Special 'augmented' statistics – (started in maths in 1992) – more degrees of freedom and physically meaningful matrix structures
- We can differentiate between several kinds of noises (doubly white circular with various distributions $n_r \perp n_i$ & $\sigma^2_{n_r} = \sigma^2_{n_i}$, doubly white noncircular $n_r \perp n_i$ & $\sigma^2_{n_r} > \sigma^2_{n_i}$, noncircular noise)

Part 1: Complex Calculus

Singularities Exist in Complex-Valued Functions

Observe the magnitude and phase for the function $f(z) = \tanh(\cdot)$



What is a Derivative?

The definition of derivative for $f(x) \in \mathbb{R}$:

$$f'(x) = \lim_{\Delta_x \to 0} \frac{f(x+\Delta_x) - f(x)}{\Delta_x}$$

For a complex function

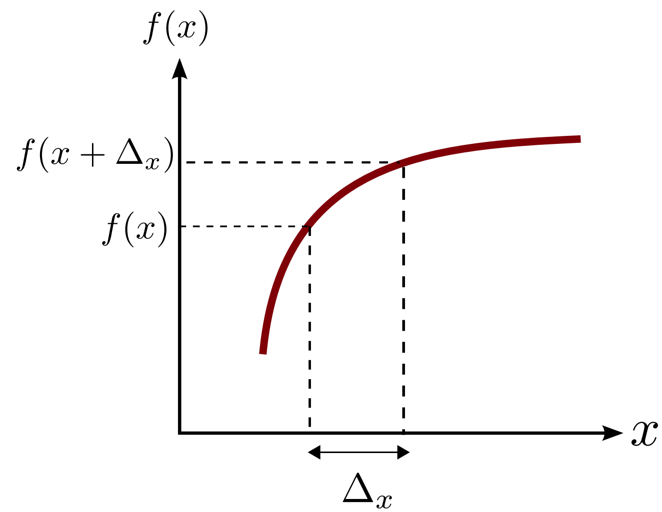
$$f(z) = u(x, y) + \jmath v(x, y)$$

to be differentiable at $z = x + \jmath y$, the limit must converge to a unique complex number no matter how $\Delta z = \Delta_x + \jmath \Delta_y \to 0$.

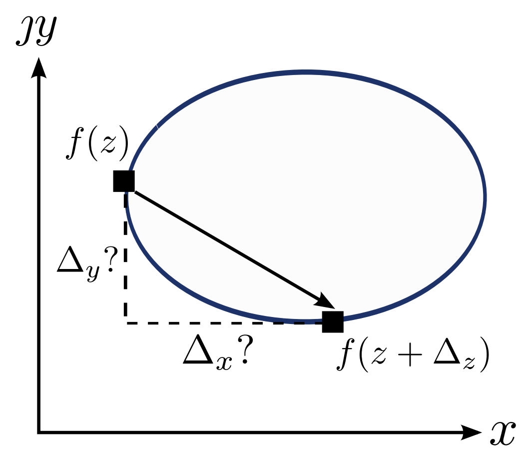
$$f'(z) = \lim_{\Delta_z \to 0} \frac{f(z+\Delta_z) - f(z)}{\Delta_z}$$

So, the complex derivative is only defined for analytic functions.

Real-Domain:



Complex-Domain:



Deriving the Cauchy-Riemann Conditions

Conditions for the Derivative to exist in $\mathbb{C}$

For $f(z)$ to be analytic, a unique limit must exist regardless of how Δz approaches zero

$$f'(z) = \lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \frac{\big[u\left(x+\Delta_x, y+\Delta_y\right) + \jmath v\left(x+\Delta_x, y+\Delta_y\right)\big] - \big[u(x,y) + \jmath v(x,y)\big]}{\Delta_x + \jmath \Delta_y}$$

must exist regardless of how Δz approaches zero. It is convenient to consider the two following cases

Case 1: $\Delta_y = 0$ and $\Delta_x \to 0$, which yields

$$\begin{aligned} f'(z) &= \lim_{\Delta_x \to 0} \frac{\big[u(x+\Delta_x,y)+\jmath v(x+\Delta_x,y)\big]-\big[u(x,y)+\jmath v(x,y)\big]}{\Delta_x} \\ &= \lim_{\Delta_x \to 0} \frac{u(x+\Delta_x,y)-u(x,y)}{\Delta_x} + \jmath \frac{v(x+\Delta_x,y)-v(x,y)}{\Delta_x} \\ &= \frac{\partial u(x,y)}{\partial x} + \jmath \frac{\partial v(x,y)}{\partial x} \end{aligned}$$

Deriving the Cauchy-Riemann Conditions

Conditions for the Derivative to exist in $\mathbb{C}$

Case 2: $\Delta_x = 0$ and $\Delta_y \to 0$, which yields

$$
\begin{aligned}
f'(z) &= \lim_{\Delta_y \to 0} \frac{\left[u\left(x,y+\Delta_y\right)+\jmath v\left(x,y+\Delta_y\right)\right]-\left[u(x,y)+\jmath v(x,y)\right]}{\jmath \Delta_y} \\
&= \lim_{\Delta_y \to 0} \frac{u\left(x,y+\Delta_y\right)-u(x,y)}{\jmath \Delta_y} + \frac{v\left(x,y+\Delta_y\right)-v(x,y)}{\Delta_y} \\
&= \frac{\partial v(x,y)}{\partial y} - \jmath \frac{\partial u(x,y)}{\partial y}
\end{aligned}
$$

For continuity, the limits from Case 1 and Case 2 must be identical, which yields

$$
\frac{\partial u(x,y)}{\partial x} = \frac{\partial v(x,y)}{\partial y}, \qquad \frac{\partial v(x,y)}{\partial x} = -\frac{\partial u(x,y)}{\partial y}
$$

that is, the expressions for the Cauchy-Riemann equations.

Learning: Cauchy–Riemann Equations

$f(z) = u(x,y) + \jmath v(x,y) \rightarrow f'(z) = \partial u(x,y)/\partial x + \jmath \partial v(x,y)/\partial x$

$$\frac{\partial u(x,y)}{\partial x} = \frac{\partial v(x,y)}{\partial y}, \qquad \frac{\partial v(x,y)}{\partial x} = -\frac{\partial u(x,y)}{\partial y}$$

Intuition: The Jacobian matrix of $f(z) = u + \jmath v$, is given by

$$\mathbf{J} = \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix} \qquad \Leftrightarrow \qquad \begin{bmatrix} '1' & '1' \\ '-1' & '1' \end{bmatrix}$$

Thus, $f(z) = z^*$ is not analytic as its Jacobian $\mathbf{J} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$.

Functions which depend on both $z = x + \jmath y$ and $z^* = x - \jmath y$ are not analytic, for example

$$J(z, z^*) = zz^* = x^2 + y^2 \quad \Rightarrow \quad \mathbf{J} = \begin{bmatrix} 2x & 2y \\ 0 & 0 \end{bmatrix} \quad \Leftrightarrow \quad \frac{\partial u}{\partial x} \neq \frac{\partial v}{\partial y} \quad \frac{\partial v}{\partial x} \neq -\frac{\partial u}{\partial y}$$

Another typical example is the cost function $J = \frac{1}{2} e(k) e^*(k) = \frac{1}{2} |e(k)|^2$

The Key: $\mathbb{CR}$-derivatives

Can we exploit results from Multivariate Calculus in $\mathbb{R}^2$?

GOAL: Find the derivative of a complex function $f(z)$ w.r.t. $z = x + \jmath y$.
In standard Multivariate Calculus in $\mathbb{R}^{N\times 1}$ the derivative of a function $g(\mathbf{x})$, $\mathbf{x} = [x_1, x_2, \ldots, x_N]$ is defined as $\frac{\partial g}{\partial \mathbf{x}} = \left[\frac{\partial g}{\partial x_1}, \ldots, \frac{\partial g}{\partial x_N}\right]^T$

- **Step 1:** Define the vector $\mathbf{x} = [x, \jmath y]^T$, hence $z = \mathbf{1}^T\mathbf{x}$.

- **Step 2:** Express the derivative of f with respect to "real" vector $\mathbf{x}$ i.e $\frac{\partial f}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\jmath \partial y} \end{bmatrix}^T$

- **Step 3:** Transform derivative vector in Step 2 back into $\mathbb{C}$

$$\frac{\partial f}{\partial z} = \mathbf{1}^T \frac{\partial f}{\partial \mathbf{x}} = \frac{\partial f}{\partial x} + \frac{\partial f}{\jmath \partial y} = \frac{\partial f}{\partial x} - \jmath \frac{\partial f}{\partial y}$$

- **Step 4:** Normalise the derivative since f is "differentiated twice"

$\mathbb{R} - \text{der} : \frac{\partial f}{\partial z} = \frac{1}{2}\left[\frac{\partial f}{\partial x} - \jmath \frac{\partial f}{\partial y}\right]$. Similarly, $\mathbb{R}^* - \text{der} : \frac{\partial f}{\partial z^*} = \frac{1}{2}\left[\frac{\partial f}{\partial x} + \jmath \frac{\partial f}{\partial y}\right]$

The Key: $\mathbb{CR}$-derivatives

Relationship between $\mathbb{CR}$-derivatives and standard $\mathbb{C}$-derivatives

- If a function $f = f(z, z^*) = u(x, y) + \jmath v(x, y)$ is holomorphic, then the Cauchy–Riemann conditions are satisfied, that is

$$\frac{\partial u(x,y)}{\partial x} = \frac{\partial v(x,y)}{\partial y} \quad \text{and} \quad \frac{\partial v(x,y)}{\partial x} = -\frac{\partial u(x,y)}{\partial y}$$

Therefore the $\mathbb{R}-$ and $\mathbb{R}^*-$derivatives are

$$\mathbb{R} - \text{der. } : \frac{\partial f}{\partial z}\Big|_{z^*=\text{const.}} = \frac{1}{2}\left[\frac{\partial f}{\partial x} - \jmath\frac{\partial f}{\partial y}\right] = \frac{1}{2}\left[2\frac{\partial u}{\partial x} + 2\jmath\frac{\partial v}{\partial x}\right] = f'(z)$$

$$\mathbb{R}^* - \text{der. } : \frac{\partial f}{\partial z^*}\Big|_{z=\text{const.}} = \frac{1}{2}\left[\frac{\partial f}{\partial x} + \jmath\frac{\partial f}{\partial y}\right] = 0$$

$\Rightarrow$ **For holomorphic functions the $\mathbb{R}^*$-derivative vanishes and the $\mathbb{R}$-derivative is equivalent to the standard complex derivative $f'(z)$**

Examples: $\mathbb{CR}$-derivatives

Prove these from the definitions of the $\mathbb{R}$ and $\mathbb{R}^*$ derivatives

For the $\mathbb{R}$ − derivative, the function is partially differentiated w.r.t z while keeping z^* constant, and vice versa for the $\mathbb{R}^*$ − derivative.

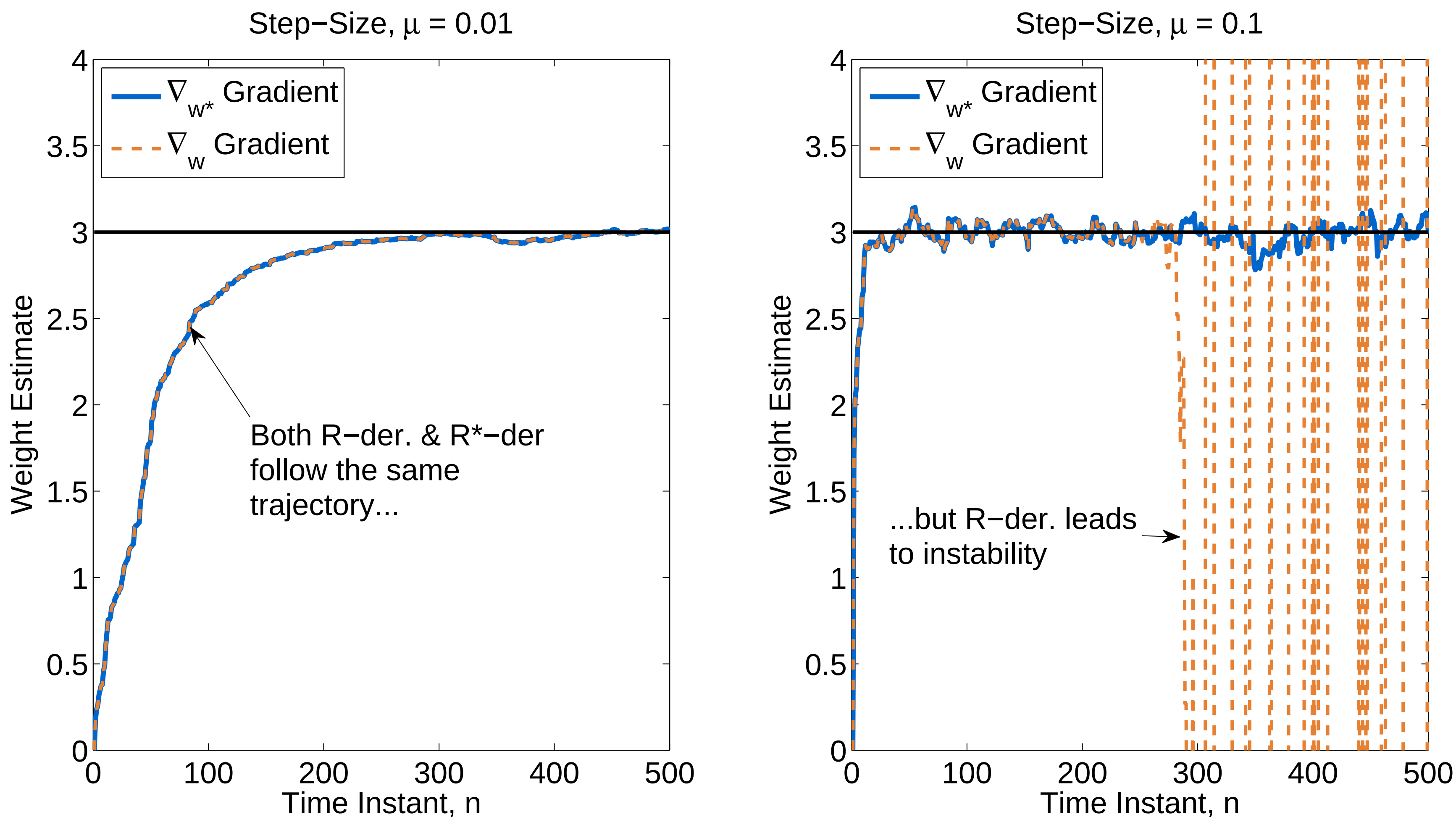
$f(z, z^*)$	$\mathbb{R}$−der	$\mathbb{R}^*$−der	$\mathbb{C}$−der
z	1	0	1
z^*	0	1	Undefined
$\|z\|^2 = zz^*$	z^*	z	Undefined
$z^2 z^*$	$2\|z\|^2$	z^2	Undefined
e^z	e^z	0	e^z

If $f(z, z^*)$ is independent of z^*, then the $\mathbb{R}$-derivative of $f(z)$ is equivalent to the standard $\mathbb{C}$-derivative;

Which derivative to we choose to compute the gradient?

$\mathbb{R}$-der vs. $\mathbb{R}^*$-der?

Simulation for the CLMS derived using $\mathbb{R}$-der. and $\mathbb{R}^*$-der. $(\mathbf{w}_o = 3)$



Stochastic Gradient Optimisation - Complex Gradient

Cost function $J(e, e^*) = |e|^2 = ee^*$, where $e(k) = d(k) - \mathbf{w}^H(k)\mathbf{x}(k)$

Gradient: $$\nabla_{\mathbf{w}} J = \frac{\partial J(e,e^*)}{\partial \mathbf{w}} = \left[\frac{\partial J(e,e^*)}{\partial w_1}, \ldots, \frac{\partial J(e,e^*)}{\partial w_N}\right]^T$$

For the minima $$\frac{\partial J(e,e^*)}{\partial \mathbf{w}} = \mathbf{0} \qquad \text{and} \qquad \frac{\partial J(e,e^*)}{\partial \mathbf{w}^*} = \mathbf{0}$$

The first term of Taylor series expansion (since $J(e, e^*)$ is real).

$$\Delta J(e,e^*) = \left[\frac{\partial J}{\partial \mathbf{w}}\right]^T \Delta\mathbf{w} + \left[\frac{\partial J}{\partial \mathbf{w}^*}\right]^T \Delta\mathbf{w}^* = 2\Re\left\{\left[\frac{\partial J}{\partial \mathbf{w}}\right]^H \Delta\mathbf{w}^*\right\} = 2\Re\left\{\left[\frac{\partial J}{\partial \mathbf{w}^*}\right]^T \Delta\mathbf{w}^*\right\}$$

The scalar product

$$< \partial J/\partial \mathbf{w}, \Delta\mathbf{w}^* > = \left[\frac{\partial J}{\partial \mathbf{w}}\right]^H \Delta\mathbf{w}^* = \| \partial J/\partial \mathbf{w} \| \| \Delta\mathbf{w}^* \| \cos\angle(\partial J/\partial \mathbf{w}, \Delta\mathbf{w}^*)$$

achieves its maximum value when $\frac{\partial J}{\partial \mathbf{w}} \parallel \Delta\mathbf{w}^*$.

Thus, **the maximum change of the gradient of the cost function is in the direction of the conjugate weight vector**, and

$$\nabla_{\mathbf{w}} J = \nabla_{\mathbf{w}^*} J \qquad\qquad \text{Brandwood 1984}$$

Vectorial scalar function

$f(\mathbf{x} = f(x_1, \ldots, x_N)$

$$\text{Gradient } \nabla_x f(\mathbf{x})) = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_N} \end{bmatrix} = \mathbf{0} \text{ and the Hessian matrix } \mathbf{H}_x > \mathbf{0}.$$

where the elements of the Hessian matrix are $\{H_x\}_{i,j} = \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}$

Theorem: If $f(\mathbf{z}, \mathbf{z}^*)$ is a real-valued function of the complex vectors $\mathbf{z}$ and $\mathbf{z}^*$, the vector pointing in the direction of the maximum rate of change of $f(\mathbf{z}, \mathbf{z}^*)$ is $\nabla_{\mathbf{z}^*} f(\mathbf{z}, \mathbf{z}^*)$, the derivative of $f(\mathbf{z}, \mathbf{z}^*)$ wrt $\mathbf{z}^*$. [Hayes 1996].

Thus, the turning points of $f(\mathbf{z}, \mathbf{z}^*)$ are solutions to $\nabla_{\mathbf{z}^*} f(\mathbf{z}, \mathbf{z}^*) = \mathbf{0}$,

$$\text{where } \nabla_{\mathbf{z}^*} = \frac{1}{2} \begin{bmatrix} \frac{\partial}{\partial x_1} + \jmath \frac{\partial}{\partial y_1} \\ \vdots \\ \frac{\partial}{\partial x_n} + \jmath \frac{\partial}{\partial y_n} \end{bmatrix} \quad , \nabla_{\mathbf{z}} \mathbf{a}^H \mathbf{z} = \mathbf{a}^*, \quad , \nabla_{\mathbf{z}^*} \mathbf{a}^H \mathbf{z} = \mathbf{0}$$

Some useful examples from $\mathbb{CR}$-Calculus

For proofs see lecture supplement

$$
\begin{aligned}
\text{Linear Form:} \quad & \frac{\partial}{\partial \mathbf{x}^*}\left\{\mathbf{x}^T\mathbf{a}\right\} = \mathbf{0} \\
\text{Linear Form:} \quad & \frac{\partial}{\partial \mathbf{x}^*}\left\{\mathbf{x}^H\mathbf{a}\right\} = \mathbf{a} \\
\text{Quadratic Form:} \quad & \frac{\partial}{\partial \mathbf{x}^*}\left\{\mathbf{x}^H\mathbf{C}\mathbf{x}\right\} = \mathbf{C}\mathbf{x} \\
\text{Quadratic Form:} \quad & \frac{\partial}{\partial \mathbf{x}^*}\left\{\mathbf{x}^T\mathbf{C}\mathbf{x}^*\right\} = \mathbf{C}^T\mathbf{x} \\
\text{Vector Form:} \quad & \mathbf{y} = \mathbf{A}\mathbf{x}, \ \frac{\partial \mathbf{y}^H}{\partial \mathbf{x}^*} = \mathbf{A}^H
\end{aligned}
$$

Some useful examples from $\mathbb{CR}$-Calculus

Chain Rule

$$\text{Linear Form:} \quad \frac{\partial}{\partial \mathbf{z}^*}\left\{\mathbf{x}^H \mathbf{a}\right\} = \frac{\partial \mathbf{x}^H}{\partial \mathbf{z}^*}\mathbf{a} + \frac{\partial \mathbf{a}^T}{\partial \mathbf{z}^*}\mathbf{x}^*$$

$$\text{Quadratic Form:} \quad \frac{\partial}{\partial \mathbf{z}^*}\left\{\mathbf{x}^H \mathbf{C}\mathbf{x}\right\} = \frac{\partial \mathbf{x}^H}{\partial \mathbf{z}^*}\mathbf{C}\mathbf{x} + \frac{\partial \mathbf{x}^T}{\partial \mathbf{z}^*}\mathbf{C}^T\mathbf{x}^*$$

$$\text{Vector Form:} \quad \mathbf{y} = \mathbf{A}\mathbf{x}, \;\; \frac{\partial \mathbf{y}^H}{\partial \mathbf{z}^*} = \frac{\partial \mathbf{x}^H}{\partial \mathbf{z}^*}\mathbf{A}^H, \;\; \frac{\partial \mathbf{y}^T}{\partial \mathbf{z}^*} = \frac{\partial \mathbf{x}^T}{\partial \mathbf{z}^*}\mathbf{A}^T$$

Matrix Derivatives

$$\text{Linear Form:} \quad \frac{\partial}{\partial \mathbf{B}^*}\left\{\mathrm{Tr}\,\mathbf{B}^*\mathbf{C}\right\} = \mathbf{C}^T$$

$$\text{Quadratic Form:} \quad \frac{\partial}{\partial \mathbf{A}^*}\left\{\mathrm{Tr}\,\mathbf{A}\mathbf{C}\mathbf{A}^H\right\} = \mathbf{A}\mathbf{C}$$

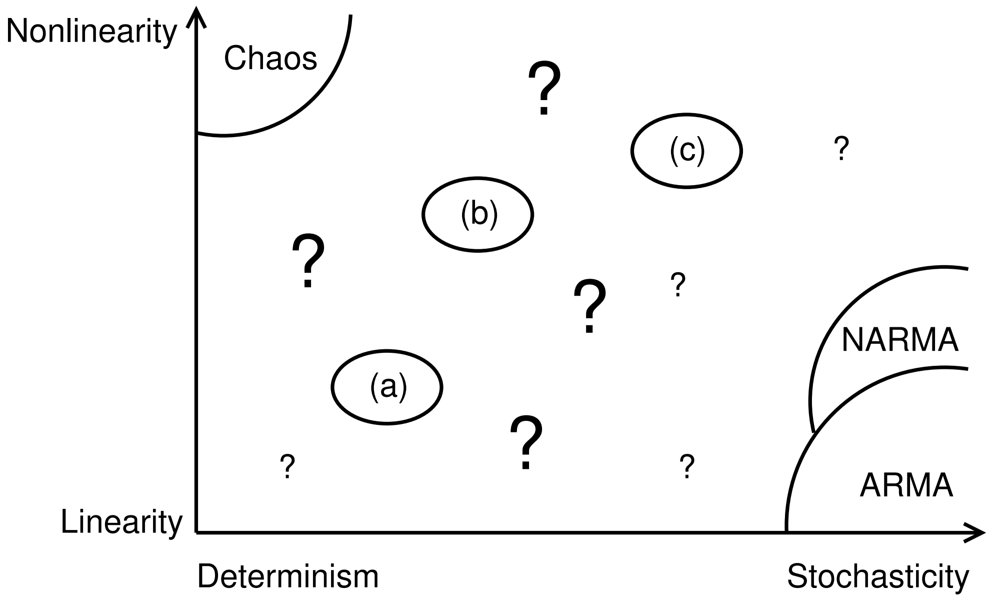
Part 2: Complex Statistics

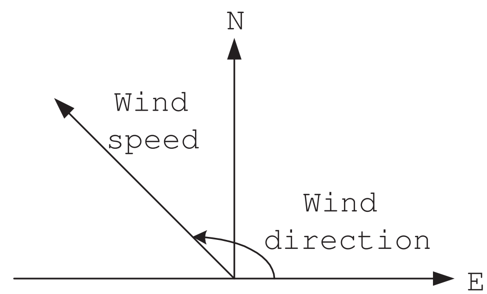
Signal modality – So why are complex signals different?

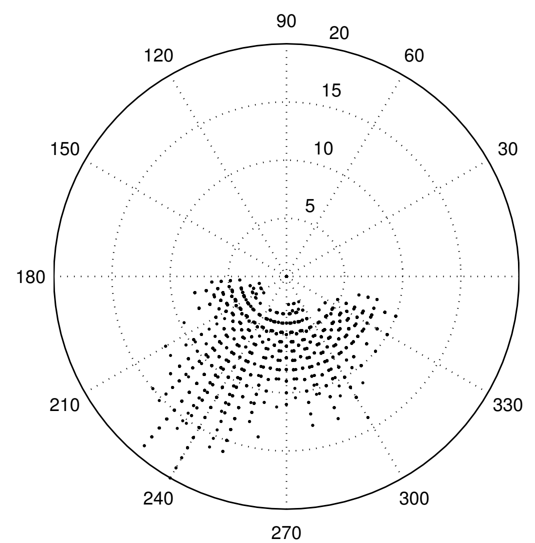
(many expressions are conformal $\looparrowright$ but dangerous to directly apply real tools!)

Deterministic vs. Stochastic nature

Linear vs. Nonlinear nature







Change in signal modality can indicate e.g. health hazard (fMRI, HRV)

Real world signals are denoted by '???'

- $\exists$ a unique signature of complex signals?
- $\looparrowright$ **degree of noncircularity**

Data model: Gaussianity

Starting From Real-valued Data

Why Gaussian? **Justification: Central Limit Theorem**

If we form a sum of independent measurements

$\Rightarrow$ the distribution of the sum tends to a Gaussian distribution

$$p(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}} \qquad x \sim \mathcal{N}(\mu_x, \sigma_x^2)$$

$\Rightarrow$ **distribution defined by its mean and variance!!!**

If $x \sim \mathcal{N}(0, \sigma_x^2)$ then $E\{x^{2n-1}\} = 1, 3, \ldots, (2n-1)\sigma_x^{2n}, \quad \forall n$

In the vector case (N Gaussian random variables)

$$p\big(x[0], x[1], \ldots, x[N-1]\big) = \frac{1}{(2\pi)^{N/2} det(\mathbf{C}_{xx})^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_x)^T \mathbf{C}_{xx}^{-1}(\mathbf{x}-\boldsymbol{\mu}_x)}$$

where $\mathbf{C}_{xx} = E\{(\mathbf{x}-\boldsymbol{\mu}_x)(\mathbf{x}-\boldsymbol{\mu}_x)^T\}$ is the **covariance matrix.**

Isomorphism Between $\mathbb{C}$ and $\mathbb{R}^2$

Moving from Real-valued to Complex-valued Data

$$z \rightarrow z^a \quad \leftrightarrow \quad \begin{bmatrix} z \\ z^* \end{bmatrix} = \begin{bmatrix} 1 & \jmath \\ 1 & -\jmath \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

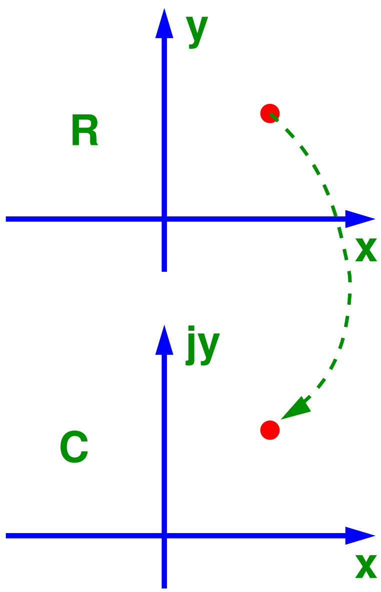
whereas in the case of complex–valued signals, we have

$$\mathbf{z} \rightarrow \mathbf{z}^a \quad \leftrightarrow \quad \begin{bmatrix} \mathbf{z} \\ \mathbf{z}^* \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \jmath\mathbf{I} \\ \mathbf{I} & -\jmath\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}$$

For convenience, the "augmented" complex vector $\mathbf{v} \in \mathbb{C}^{2N \times 1}$ can be introduced as

$$\begin{aligned} \mathbf{v} &= [z_1, z_1^*, \ldots, z_N, z_N^*]^T \\ \mathbf{v} &= \mathbf{A}\mathbf{w}, \qquad \mathbf{w} = [x_1, y_1, \ldots, x_N, y_N]^T \end{aligned}$$

where matrix $\mathbf{A} = diag(\mathbf{J}, \ldots, \mathbf{J}) \in \mathbb{C}^{2N \times 2N}$ is block diagonal and transforms the **composite** real vector $\mathbf{w}$ into the augmented complex vector $\mathbf{v}$.



The Multivariate Complex Normal Distribution

We cannot introduce a CDF $\looparrowright$ pdf's introduced via duality with $\mathbb{R}$

Recall, the relationships like "$<$" or "$\geq$" make no sense in $\mathbb{C}$.

$$\mathbf{V} = cov(\mathbf{v}) = E[\mathbf{v}\mathbf{v}^H] = \mathbf{A}\mathbf{W}\mathbf{A}^H$$

Using the result by Vanden Bos 1995

$$\begin{aligned}
\mathbf{w} &= \mathbf{A}^{-1}\mathbf{v} = \frac{1}{2}\mathbf{A}^H\mathbf{v} \\
det(\mathbf{W}) &= \left(\frac{1}{2}\right)^{2N} det(\mathbf{V}) \\
\mathbf{w}^T\mathbf{W}^{-1}\mathbf{w} &= \mathbf{v}^H\mathbf{V}^{-1}\mathbf{v}
\end{aligned}$$

The multivariate *generalised complex normal distribution* (GCND) can now be expressed as

$$f(\mathbf{v}) = \frac{1}{\pi^N\sqrt{det(\mathbf{V})}}\, e^{-\frac{1}{2}\mathbf{v}^H\mathbf{V}^{-1}\mathbf{v}}$$

and has been derived without any restriction.

Try this in MATLAB

Circularity $\looparrowright$ Rotation invariant distrib.

$$p(\rho, \theta) = p(\rho, \theta - \phi)$$

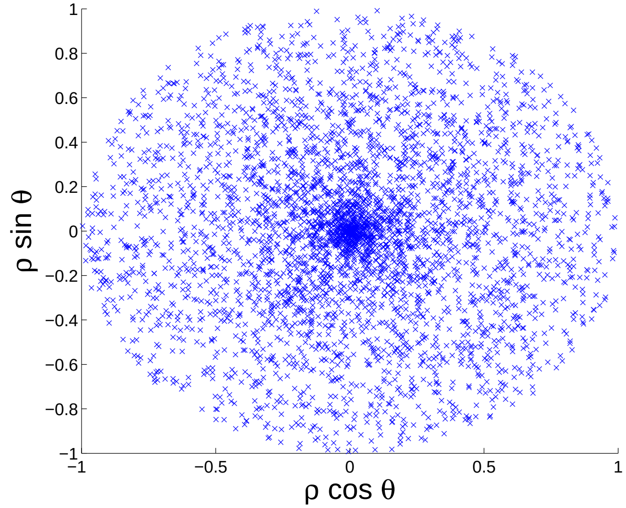
1. The name of the distribution takes after the distribution of the real–valued random variable ρ with a pdf $p(\rho)$;

2. It can be Gaussian, uniform, etc.

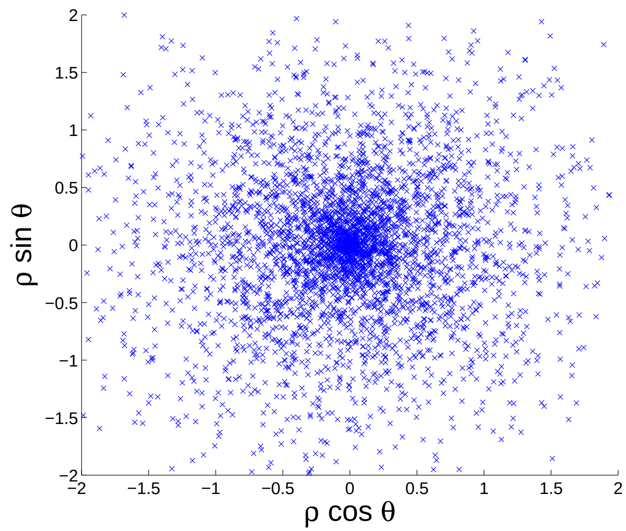
3. Take another real–valued random variable θ, which must be uniformly distributed on $[0, 2\pi]$ and independent of ρ;

4. Construct the complex random variable $Z = X + jY$ as

$$X = \rho\cos(\theta), \qquad Y = \rho\sin(\theta)$$



(i) Uniform circular



(j) Gaussian circular

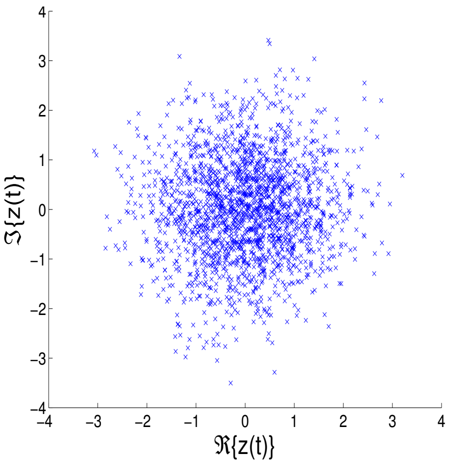
Other Definitions of Circularity

Via Probability density function, Characteristic Function, Cumulants

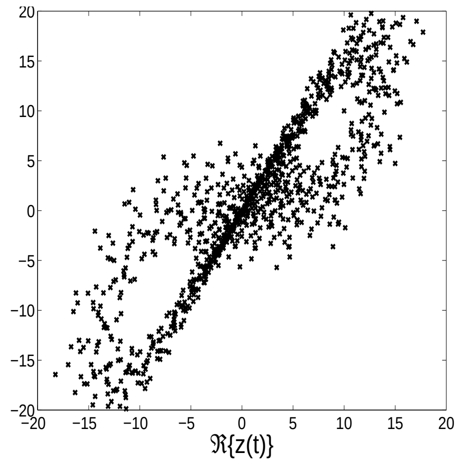
- *Probability density function.* A complex random variable Z is circular if its pdf is a function of only the product zz^*, that is[1]

$$p_{Z,Z^*}(z, z^*) = p_{Z_\phi, Z_\phi^*}(z_\phi, z_\phi^*)$$

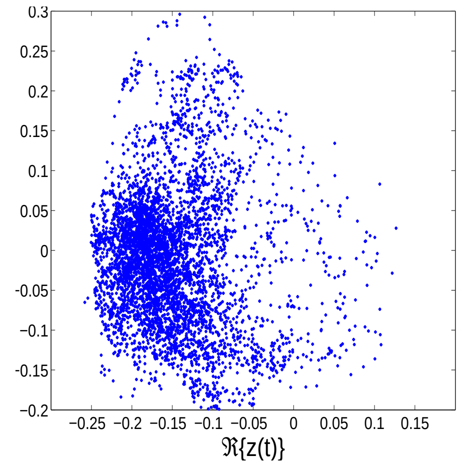
and for for Gaussian CCRVs we have
$$p_{Z,Z^*}(z, z^*) = \frac{1}{\pi\sigma^2} e^{-zz^*/\sigma^2}$$



(k) Complex AR(4)



(l) Complex Lorenz

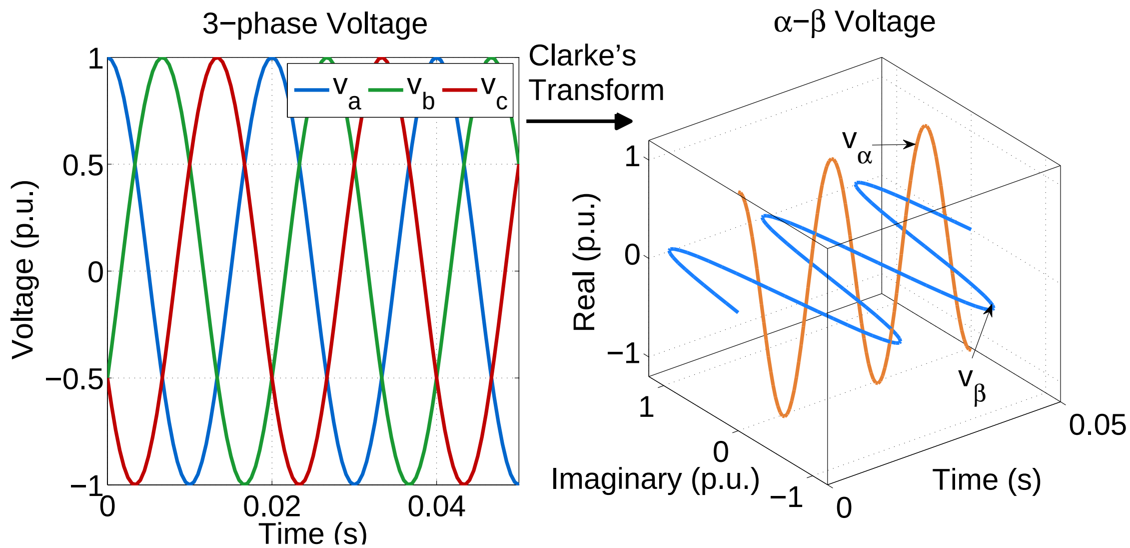


(m) Complex wind

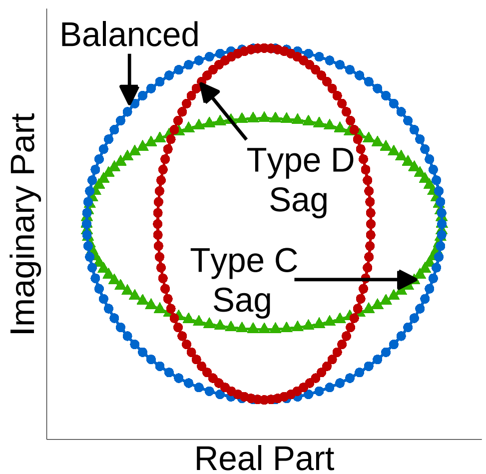
[1]The pdf of a circular complex random variable is function of only the modulus of z, **and not of z^*.**

Does Circularity Influence Estimation in $\mathbb{C}$?

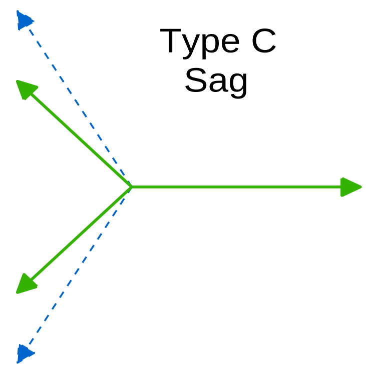
Visualising the Clarke Transform and Noncircular Voltage Signals



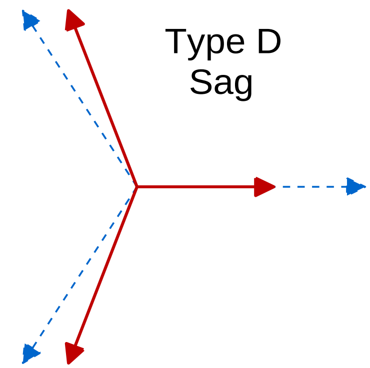
Circularity Diagram



Phasor Diagram



Phasor Diagram

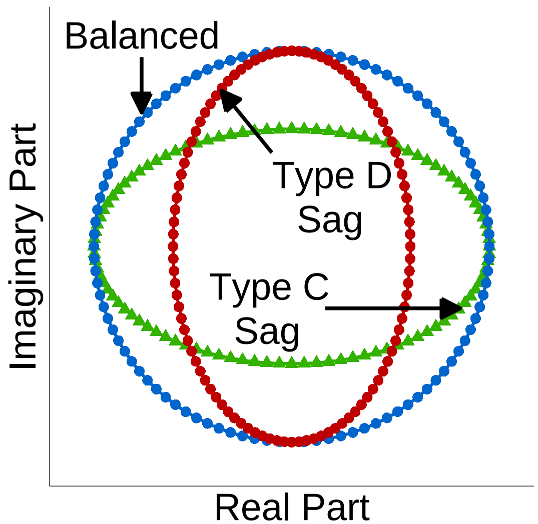


Does Circularity Influence Estimation in $\mathbb{C}$?

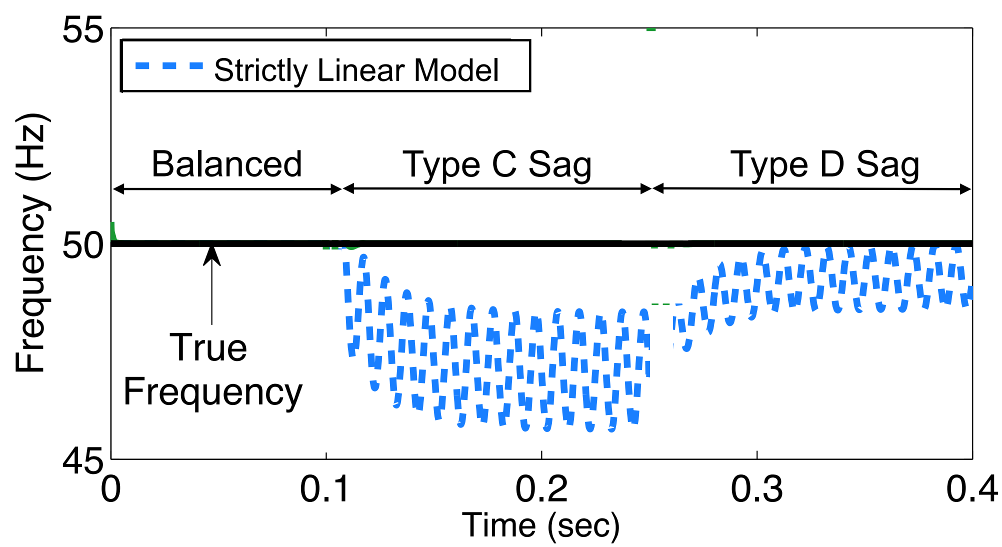
Voltage Sag: A magnitude and/or phase imbalance

- For balanced systems, $v(k) = A(k)e^{\jmath\omega k\Delta T}$ → circular trajectory.
- Unbalanced systems, $v(k) = A(k)e^{\jmath\omega k\Delta T} + B(k)\mathbf{e}^{-\jmath\omega \mathbf{k}\boldsymbol{\Delta}\mathbf{T}}$ are influenced by the "conjugate" component.
- We need the complex conjugate when the modelling the signal.

Circularity Diagram



Strictly linear model yields biased estimates when system is unbalanced



What are we doing wrong $\looparrowright$ Widely Linear Model

Consider the MSE estimator of a signal y in terms of another observation x

$$\hat{y} = E[y|x]$$

For zero mean, jointly normal y and x, the solution is

$$\hat{y} = \mathbf{h}^T\mathbf{x}$$

In standard MSE in the complex domain $\hat{y} = \mathbf{h}^H\mathbf{x}$, however

$$\hat{y}_r = E[y_r|x_r, x_i] \quad \& \quad \hat{y}_i = E[y_i|x_r, x_i]$$

$$thus \qquad \hat{y} = E[y_r|x_r, x_i] + \jmath E[y_i|x_r, x_i]$$

Upon employing the identities $x_r = (x + x^*)/2$ and $x_i = (x - x^*)/2\jmath$

$$\hat{y} = E[y_r|x, x^*] + \jmath E[y_i|x, x^*]$$

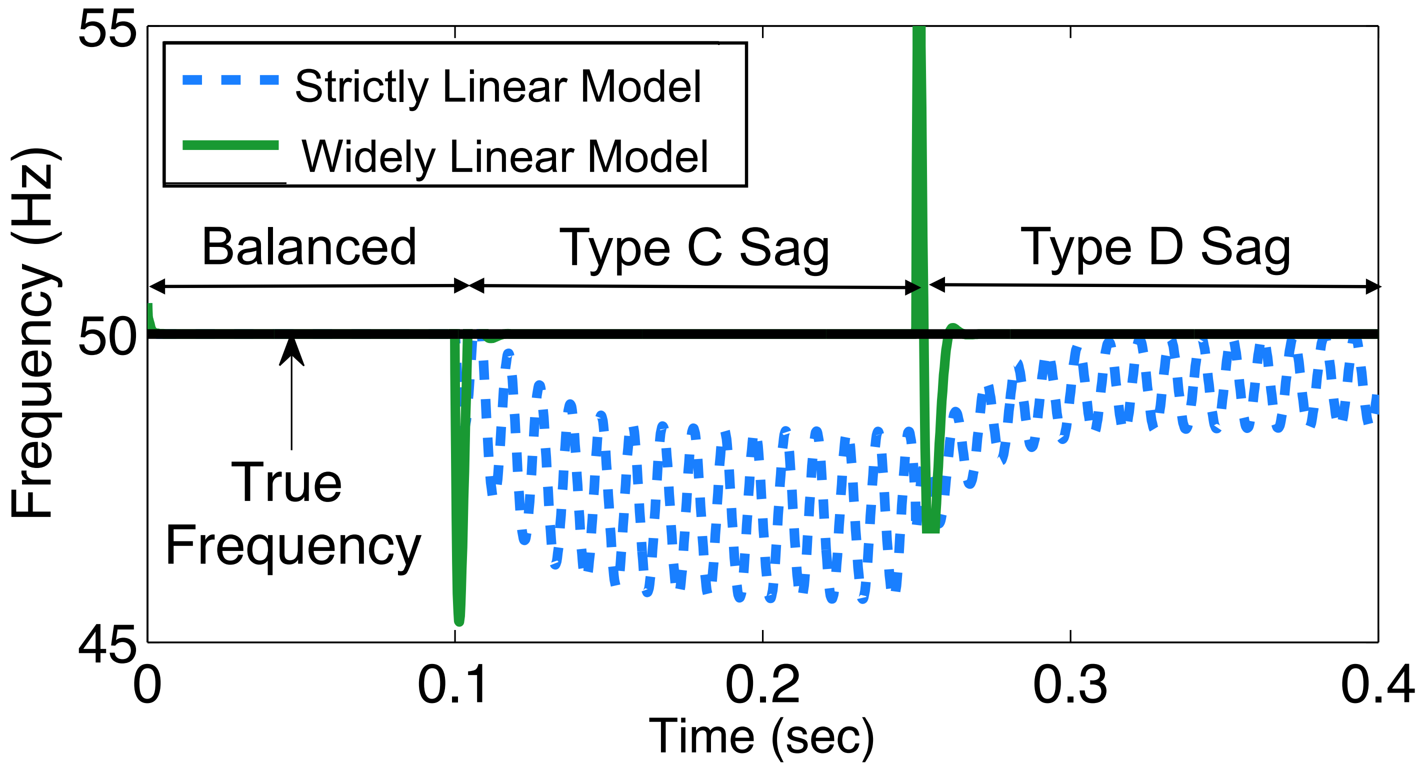
and thus arrive at the **widely linear** estimator for general complex signals

$$y = \mathbf{h}^H\mathbf{x} + \mathbf{g}^H\mathbf{x}^*$$

We can now process general (noncircular) complex signals!

Using the widely linear model for frequency estimation

The widely linear model is able to estimate the frequency for both **circular** (balanced) and **noncircular** (unbalanced) voltages.



Dealing with Complex Statistics

Provides us with a tremendous amount of structure

For $\mathbf{z} = \mathbf{x} + \jmath \mathbf{y}$, 'augmented' vectors $\mathbf{w}^a = [\mathbf{h}^T, \mathbf{g}^T]^T$ and $\mathbf{z}^a = [\mathbf{z}^T, \mathbf{z}^H]^T$

$$y = \mathbf{w}^{aH} \mathbf{z}^a$$

so the 'augmented' covariance matrix

$$\mathbf{C}^a_{zz} = E \begin{bmatrix} \mathbf{z} \\ \mathbf{z}^* \end{bmatrix} \left[\mathbf{z}^H \mathbf{z}^T \right] = \begin{bmatrix} \mathbf{C}_{zz} & \mathbf{P}_{zz} \\ \mathbf{P}^*_{zz} & \mathbf{C}^*_{zz} \end{bmatrix}$$

Remark #1: In general, the covariance matrix $\mathbf{C}_{zz} = E\{\mathbf{z}\mathbf{z}^H\}$ *does not* completely describe the second order statistics of $\mathbf{z}$

Remark #2: The **pseudocovariance** or **complementary covariance** $\mathbf{P}_{zz} = E\{\mathbf{z}\mathbf{z}^T\}$ needs also to be taken into account;

Remark #3: For second-order circular (proper data) $\mathbf{P}_{zz} = \mathbf{0}$ *vanishes because:*

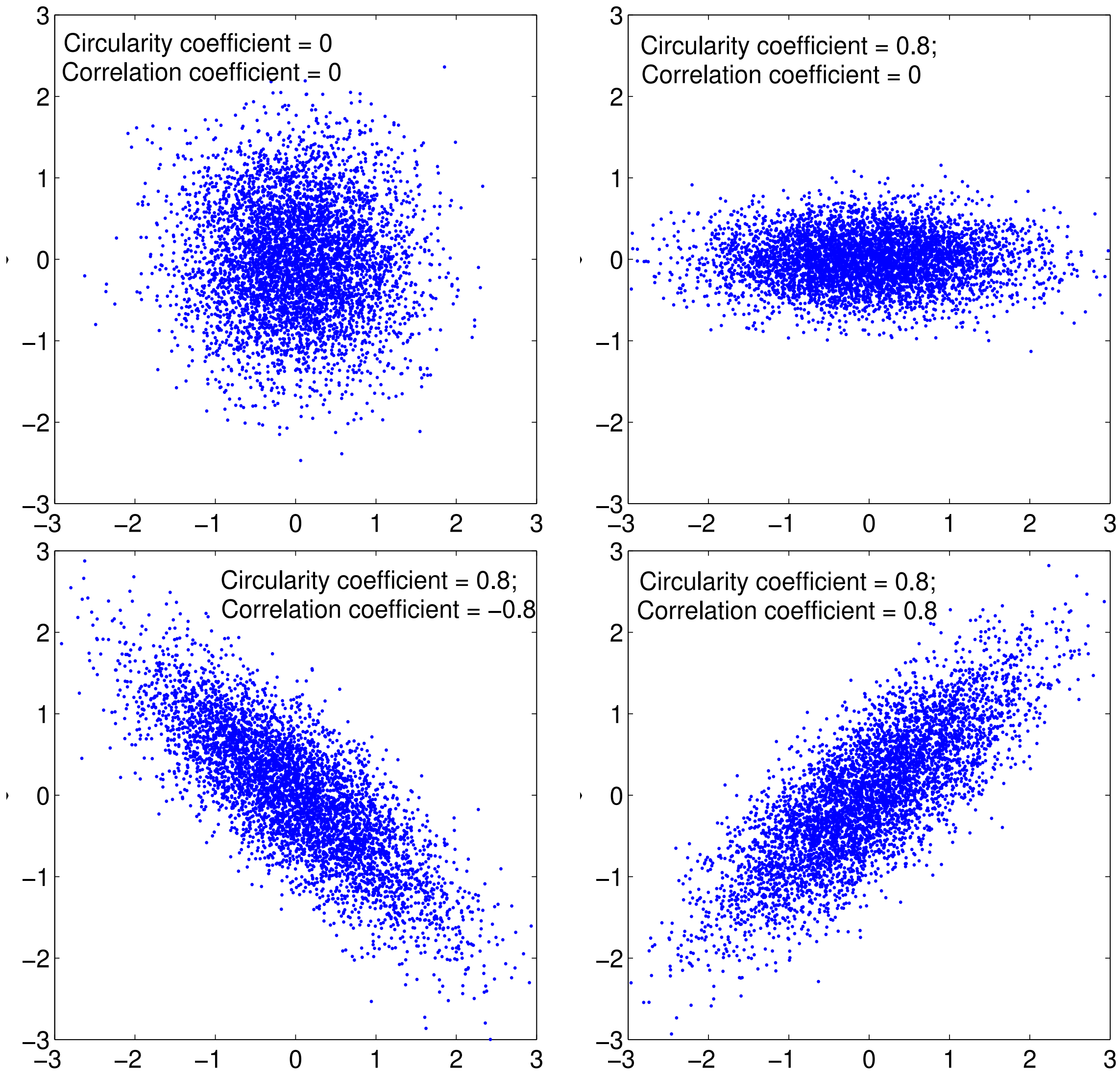
$$E\{z \times z^T\} = E\{x^2\} - E\{y^2\} + 2\jmath E\{xy\} = \sigma_x^2 - \sigma_y^2 + 2\jmath \rho_{xy}$$

Remark #4: However, general complex random processes are *improper*.

'Properness' is a second order statistical property and 'circularity' is a property of the probability density function.

Different kinds of noncircularity

'Noncircular' and 'Improper' used interchangeably, but these are not identical



So, **the degree of circularity** can be used as a **fingerprint** of a signal, allowing us enormous additional freedom in estimation, compared with standard **strictly linear** systems.

For instance, we can now differentiate between different Gaussian signals!

Recall: Real valued ICA cannot separate two Gaussian signals.

Autoregressive Modelling in $\mathbb{C}$

Standard AR model of order n is given by

$$z(k) = a_1 z(k-1) + \cdots + a_n z(k-n) + q(k) = \mathbf{a}^T \mathbf{z}(k) + q(k),$$

Using the Yule-Walker equations the AR coefficients are found from

$$\begin{aligned}
\mathbf{a}^* &= \mathcal{C}^{-1}\mathbf{c} \\
\begin{bmatrix} a_1^* \\ a_2^* \\ \vdots \\ a_n^* \end{bmatrix} &= \begin{bmatrix} c(0) & c^*(1) & \dots & c^*(n-1) \\ c(1) & c(0) & \dots & c^*(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ c(n-1) & c(n-2) & \dots & c(0) \end{bmatrix}^{-1} \begin{bmatrix} c(1) \\ c(2) \\ \vdots \\ c(n) \end{bmatrix}
\end{aligned}$$

where $\mathbf{c} = [c(1), c(2), \dots, c(n)]^T$ is the time shifted correlation vector.

Widely linear model

$$y(k) = \mathbf{h}^T(k)\mathbf{x}(k) + \mathbf{g}^T(k)\mathbf{x}^*(k) + q(k)$$

Widely linear normal equations

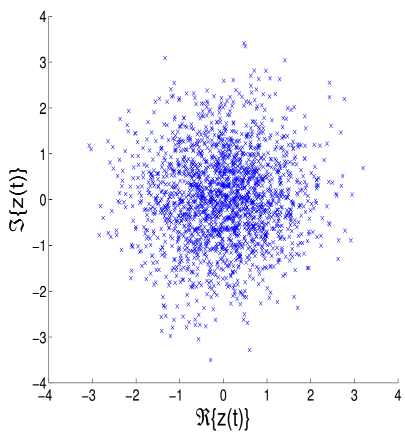
$$\begin{bmatrix} \mathbf{h}^* \\ \mathbf{g}^* \end{bmatrix} = \begin{bmatrix} \mathcal{C} & \mathcal{P} \\ \mathcal{P}^* & \mathcal{C}^* \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{c} \\ \mathbf{p}^* \end{bmatrix}$$

where $\mathbf{h}$ and $\mathbf{g}$ are coefficient vectors and $\mathbf{x}$ the regressor vector.

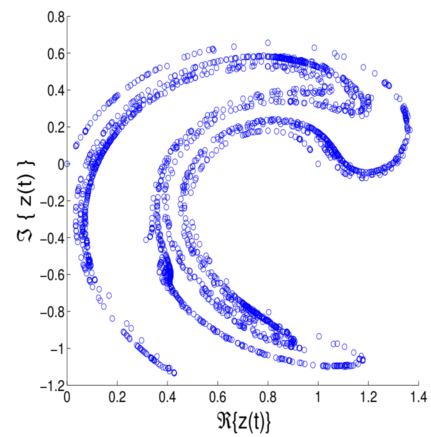
Practical Example

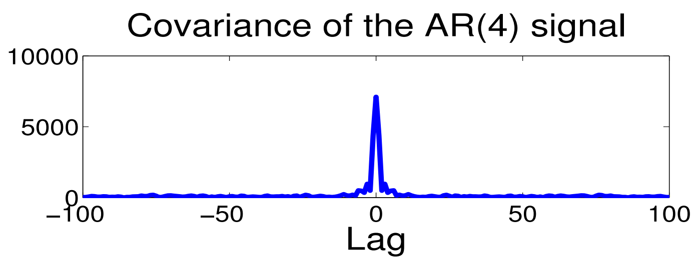
Do we ever know that the data are circular (short length, aftefacts)?

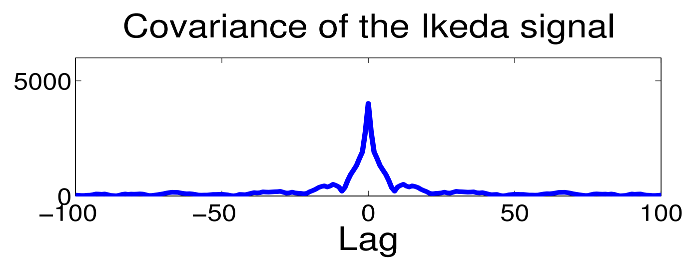
Complex AR(4) process (circular)

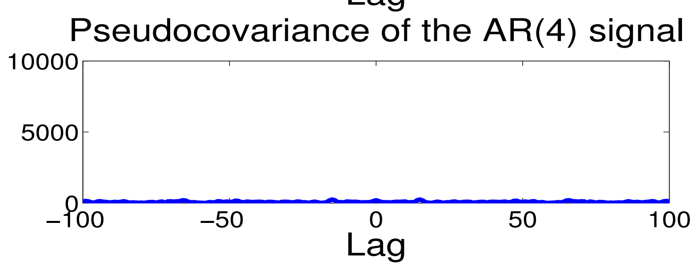


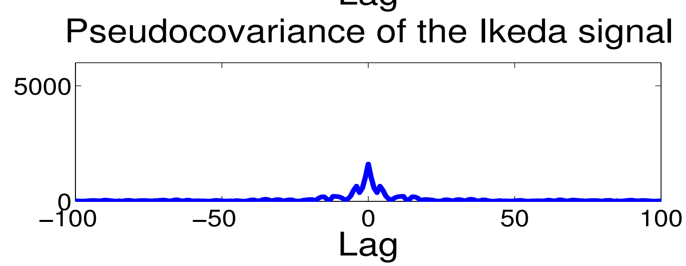
Complex Ikeda map (noncircular)









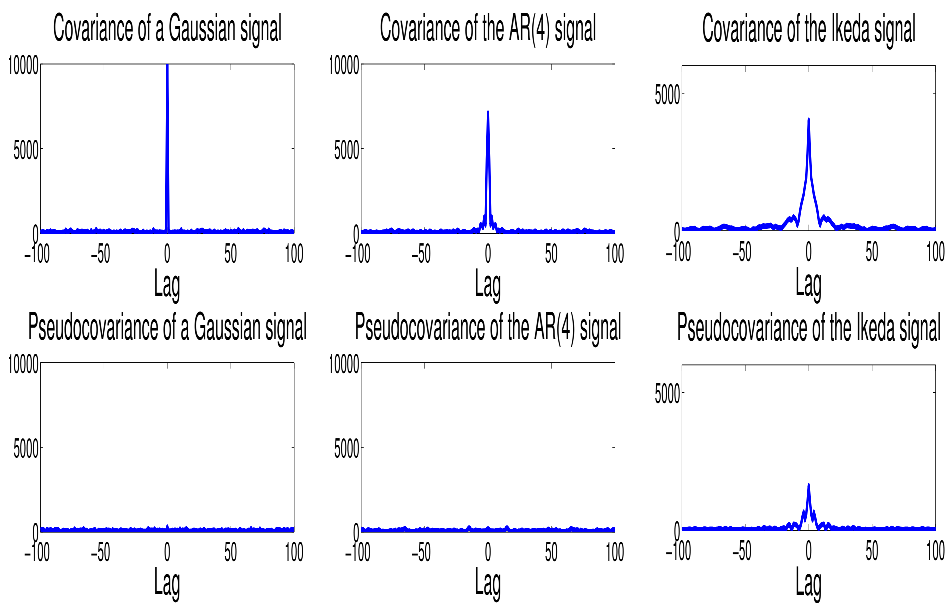


Complex AR(4) process (proper) Complex Ikeda map (improper)

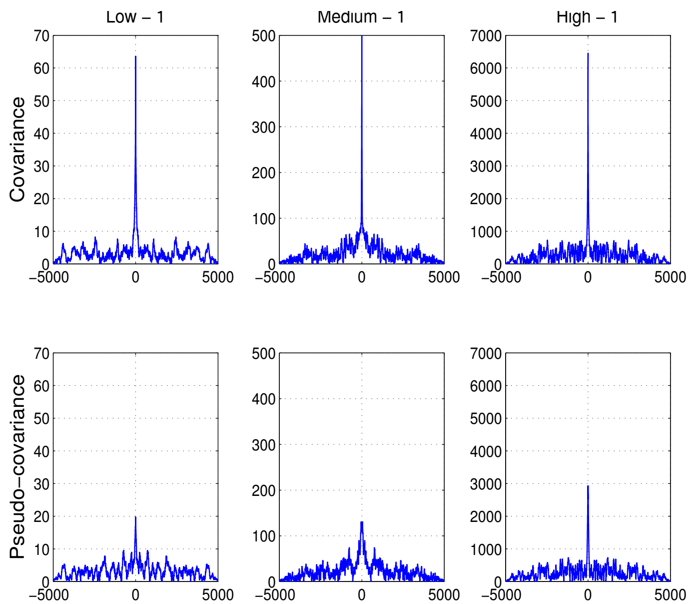
The augmented covariance matrix

$$\mathbf{C}^a_{zz} = E\left\{\mathbf{z}^a\mathbf{z}^{aH}\right\} = \begin{bmatrix} \mathbf{z}\mathbf{z}^H & \mathbf{z}\mathbf{z}^T \\ \mathbf{z}^*\mathbf{z}^H & \mathbf{z}^*\mathbf{z}^T \end{bmatrix} = \begin{bmatrix} \mathbf{C}_{zz} & \mathbf{P}_{zz} \\ \mathbf{P}^*_{zz} & \mathbf{C}^*_{zz} \end{bmatrix}$$

The *augmented covariance matrix* $\mathbf{C}^a_{zz}$ is Hermitian and has real eigenvalues.



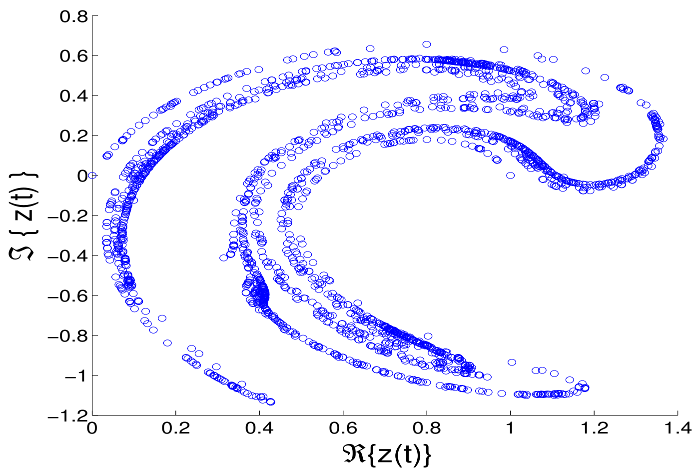
Synthetic data



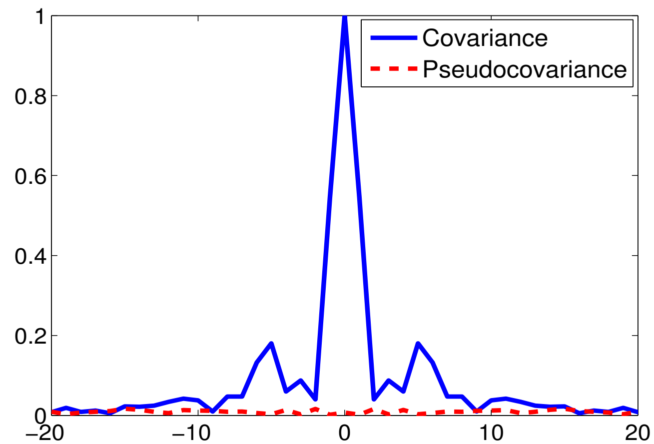
Real world wind signal

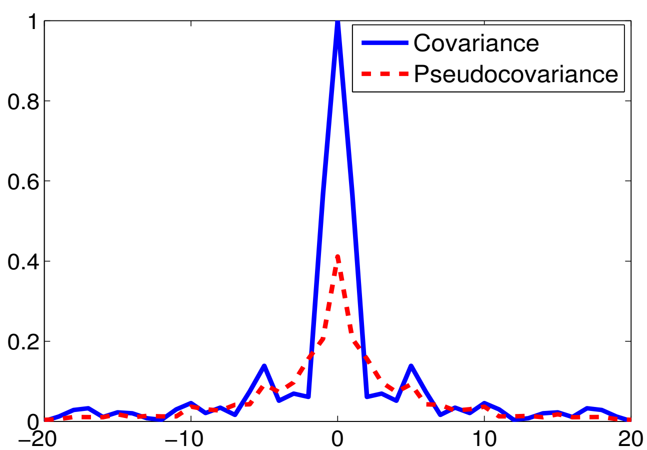
This is a rigorous way to model general complex signals!

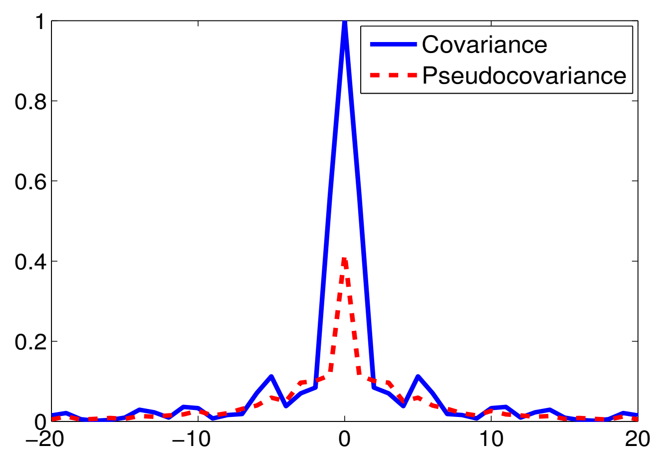
Circularity for Ikeda map



AR model of Ikeda signal







Covariances: Original Ikeda   Widely linear AR of Ikeda

Appendix: CR calculus and learning algorithms (covered later)

The Derivative of a Cost Function $\frac{1}{2}e(k)e^*(k)$ and CLMS

As $\mathbb{C}$-derivatives are not defined for real functions of complex variable

$$\mathbb{R}-\text{der:}\quad \frac{\partial}{\partial \mathbf{z}} = \frac{1}{2}\left[\frac{\partial}{\partial \mathbf{x}} - \jmath\frac{\partial}{\partial \mathbf{y}}\right] \qquad \mathbb{R}^*-\text{der:}\quad \frac{\partial}{\partial \mathbf{z}^*} = \frac{1}{2}\left[\frac{\partial}{\partial \mathbf{x}} + \jmath\frac{\partial}{\partial \mathbf{y}}\right]$$

and the gradient

$$\nabla_{\mathbf{w}} J = \frac{\partial J(e,e^*)}{\partial \mathbf{w}} = \left[\frac{\partial J(e,e^*)}{\partial w_1},\ldots,\frac{\partial J(e,e^*)}{\partial w_N}\right]^T = 2\frac{\partial J}{\partial \mathbf{w}^*} = \underbrace{\frac{\partial J}{\partial \mathbf{w}^r} + \jmath\frac{\partial J}{\partial \mathbf{w}^i}}_{pseudogradient}$$

The standard Complex Least Mean Square (CLMS) (Widrow *et al.* 1975)

$$y(k) = \mathbf{w}^H(k)\mathbf{x}(k)$$

$$e(k) = d(k) - \mathbf{w}^H(k)\mathbf{x}(k) \qquad e^*(k) = d^*(k) - \mathbf{x}^H(k)\mathbf{w}(k)$$

$$\text{and}\quad \nabla_{\mathbf{w}} J = \nabla_{\mathbf{w}^*} J$$

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \mu\frac{\partial \frac{1}{2}e(k)e^*(k)}{\partial \mathbf{w}^*(k)} = \mathbf{w}(k) + \mu e^*(k)\mathbf{x}(k)$$

Thus, no need for tedious computations – The CLMS is derived in one line.

Appendix: Does Circularity Influence Estimation in $\mathbb{C}$?

Real-world example: Estimation in the Smart Grid

Three-phase voltages can be represented as a single-channel complex signal by first using the **Clarke Transform**,

$$\begin{bmatrix} v_0(k) \\ v_\alpha(k) \\ v_\beta(k) \end{bmatrix} = \underbrace{\sqrt{\frac{2}{3}} \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix}}_{\text{Clarke Matrix}} \underbrace{\begin{bmatrix} V_a(k)\cos(\omega nT + \phi_a) \\ V_b(k)\cos(\omega nT + \phi_b - \frac{2\pi}{3}) \\ V_c(k)\cos(\omega nT + \phi_c + \frac{2\pi}{3}) \end{bmatrix}}_{\text{Three-phase voltage}}$$

Then by forming the complex-valued $\alpha\beta$ voltage: $v(k) = v_\alpha(k) + \jmath v_\beta(k)$:

$$v(k) = v_\alpha(k) + \jmath v_\beta(k) = A(k)e^{\jmath\omega kT} + B(k)e^{-\jmath\omega kT}$$

$$A(k) = \frac{\sqrt{6}}{6}\left[V_a(k)e^{\jmath\phi_a} + V_b(k)e^{\jmath\phi_b} + V_c(k)e^{\jmath\phi_c}\right],$$

$$B(k) = \frac{\sqrt{6}}{6}\left[V_a(k)e^{-\jmath\phi_a} + V_b(k)e^{-\jmath\left(\phi_b+\frac{2\pi}{3}\right)} + V_c(k)e^{-\jmath\left(\phi_c-\frac{2\pi}{3}\right)}\right]$$

For balanced systems i.e. $V_a(k) = V_b(k) = V_c(k)$ and $\phi_a = \phi_b = \phi_c$,

$$B(k) = 0$$

Notes:

-

Notes:

○