

---

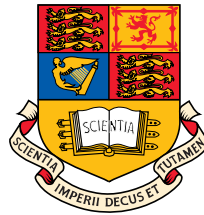
# Adaptive Sig. Proc. & Machine Intel.

## Lecture 2 - Augmented Complex Statistics and Widely Linear Modelling

---

**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](http://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
- Key point 1:  $\mathbb{C}\mathbb{R}$ -Calculus and its application

### Part 2: Complex Statistics

- Data model: Gaussian
- Moving from real to complex
- Key point 2: Circularity and widely linear estimation
- Covariance and pseudocovariance
- Widely linear autoregressive model  $\leadsto$  caters for both second order circular (proper) and non-circular (improper) data

# Motivation for modelling in $\mathbb{C}$

## Much more convenient in a number of applications

---

- Magnetic source imaging (fMRI, MRI, MEG) are recorded in the Fourier domain, that is, the data are inherently complex-valued
- 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 filtering – we like zero mean signals, but in  $\mathbb{R}$  the  $\log$  does not exist for  $x \leq 0$ , yet  $\log z = \log |z| + j\arg(z)$  does
- Optics and seismics - reflection, refraction  $\leftrightarrow$  phase information
- Fractals, associative memory (recognising objects from their parts)
- Much work still to be done – **great opportunity for future research!**

# Fundamental theorem of algebra (FTA)

- Initial work by Albert Girard in 1629

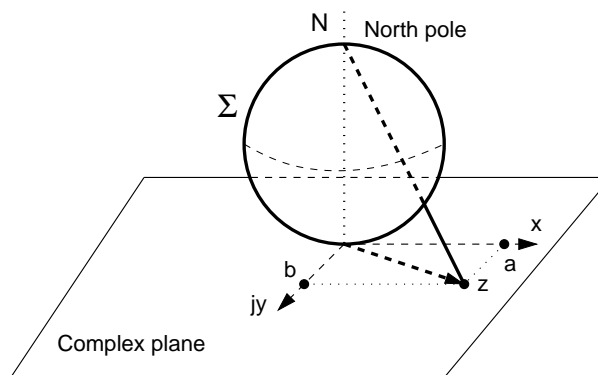
*'there are  $n$  roots to an  $n$ -th order polynomial'*

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

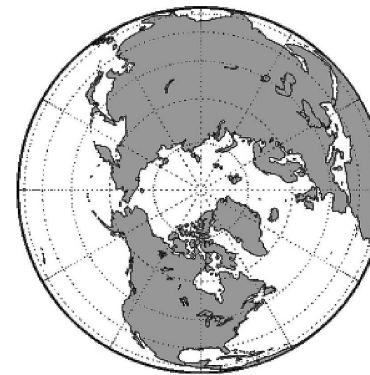
- Descartes in the 1630s **'For every equation of degree  $n$  we can imagine roots which do not correspond to any real quantity'**

- In 1749 Euler proved the FTA

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



(a) Riemann sphere



(b) Earth projection from South pole

## Stereographic projection and Riemann sphere

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

# History of mathematical notation

## Did you know?

---

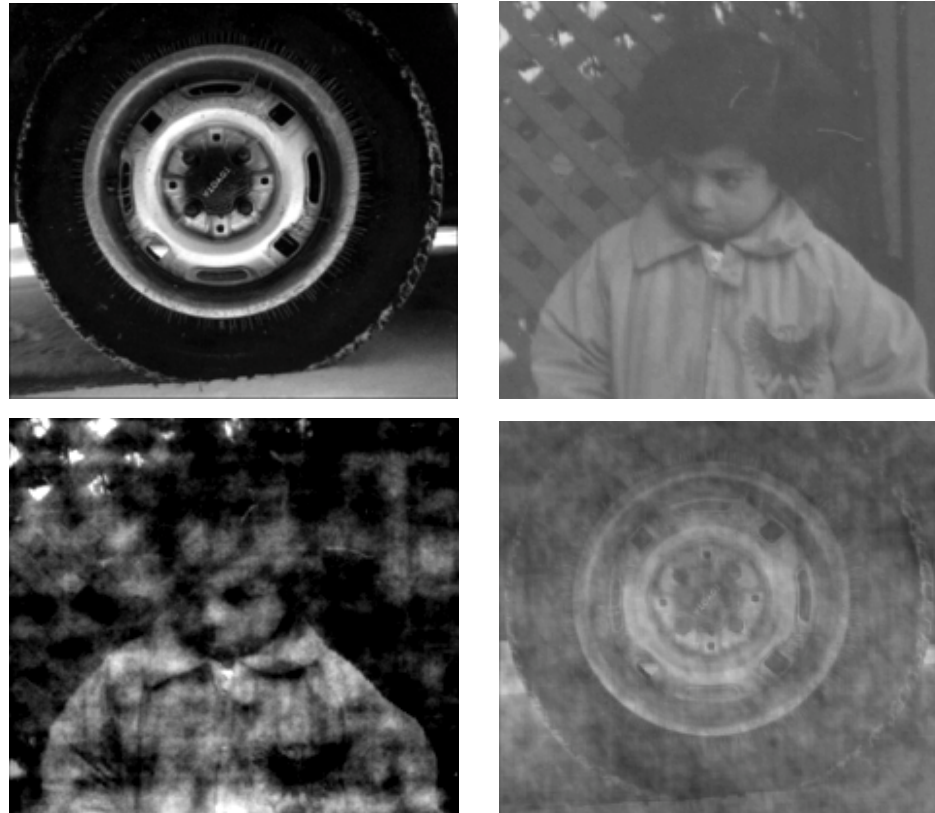
- ⊗ 9<sup>th</sup> century Al Kwarizimi's *Algebra* - solutions descriptive rather than in form of equations
- ⊗ 16<sup>th</sup> century - G. Cardano *Ars Magna* - unknowns denoted by single roman letters
- ⊗ Descartes (1630-s) established general rules
  - lowercase italic letters at the beginning of the alphabet for unknown constants  $a, b, c, d$
  - lowercase italic letters at the end of the alphabet for unknown variables  $x, y, z$
- ⊗  $\sqrt{-1} = i$  – Gauss 1830s, boldface letters for vectors  $\mathbf{x}, \mathbf{v}$  - Oliver Heaviside
- ⊗ Hence  $ax^2 + by + cz = 0$

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

# Example 1: Human visual system

## Importance of phase information

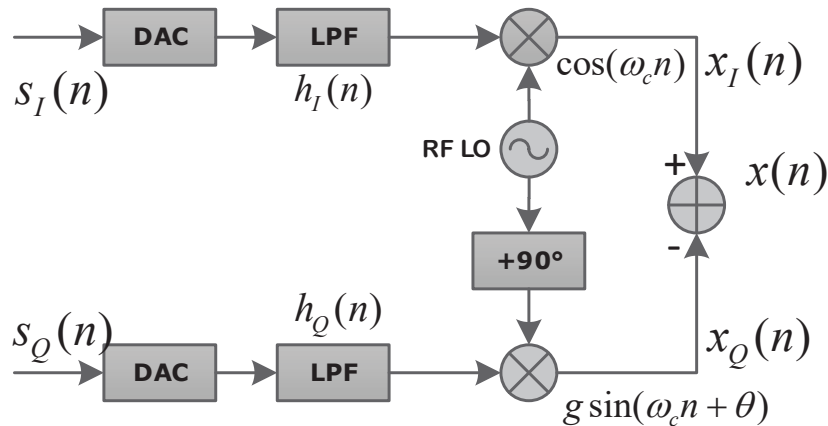
---



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.

# Example 2: Noncircularity arising from I/Q imbalance

One of the key issues in the future 5G networks



Consider the baseband discrete-time input signal,  $s(n)$ , which is complex circular, e.g., 64-QAM. After passing through an I/Q imbalanced modulator, the output  $x(n)$  becomes noncircular, that is

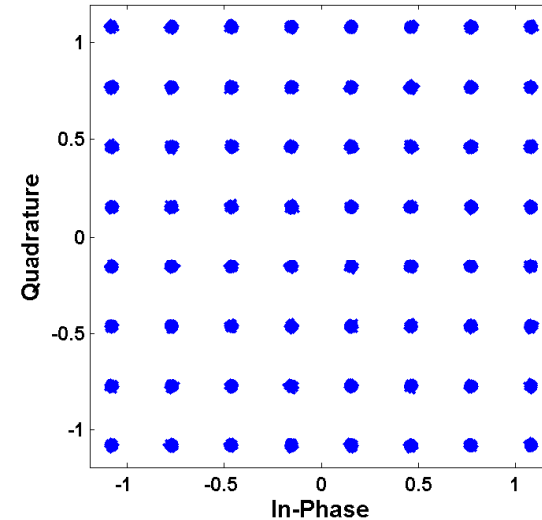
$$x(n) = \mu(n) * s(n) + \nu(n) * s^*(n)$$

where

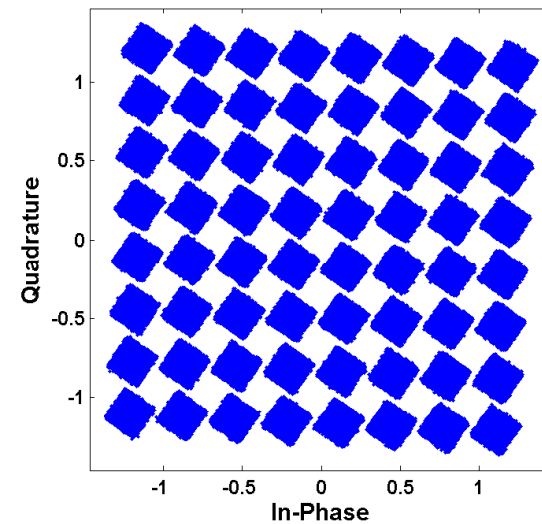
$$\mu(n) = 1/2[h_I(n) + gh_Q(n)e^{-j\theta}]$$

$$\nu(n) = 1/2[h_I(n) - gh_Q(n)e^{-j\theta}]$$

802.11ac 64QAM Original Signal (Noise Added)



802.11ac 64QAM I/Q Imbalanced Signal (Noise Added)



## Modern complex estimation: Numerous opportunities

---

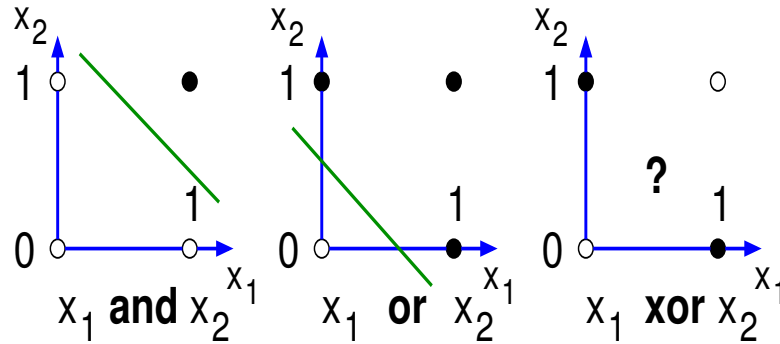
- Complex signals by design (communications, analytic signals, equivalent baseband representation to eliminate spectral redundancy)
- By convenience of representation (radar, sonar, wind field), direction of arrival related problems
- **Problem:** More powerful algebra than  $\mathbb{R}^2$  but no ordering (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  $n_r \perp n_i$  &  $\sigma_{n_r}^2 = \sigma_{n_i}^2$ , doubly white noncircular  $n_r \perp n_i$  &  $\sigma_{n_r}^2 > \sigma_{n_i}^2$ , noncircular noise)



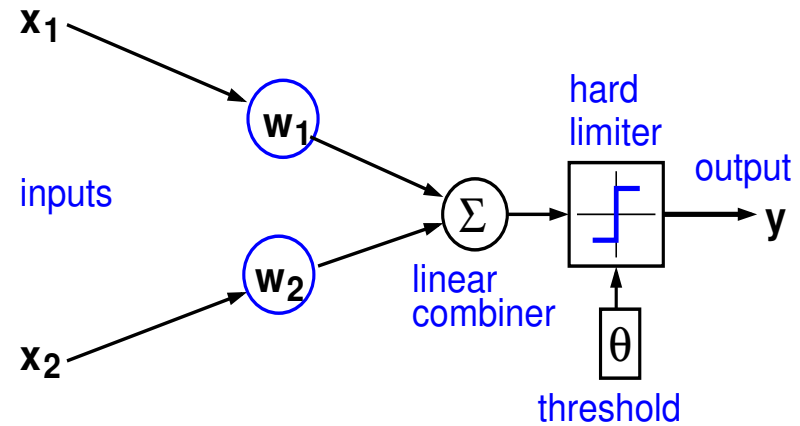
# Jumping ahead a little

(Recall the human-monkey scatter plot)

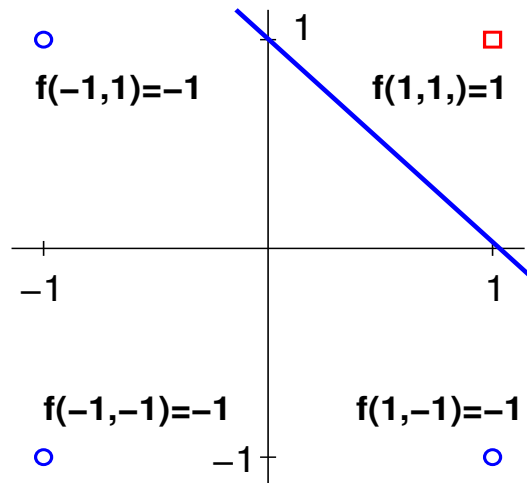
## (Non-)Linear separability and learning machines



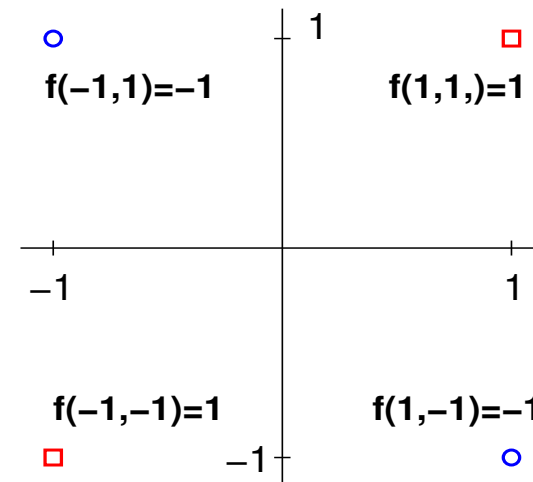
Logic functions and linear separability



Perceptron-type neural network



AND function is linearly separable



How about the XOR

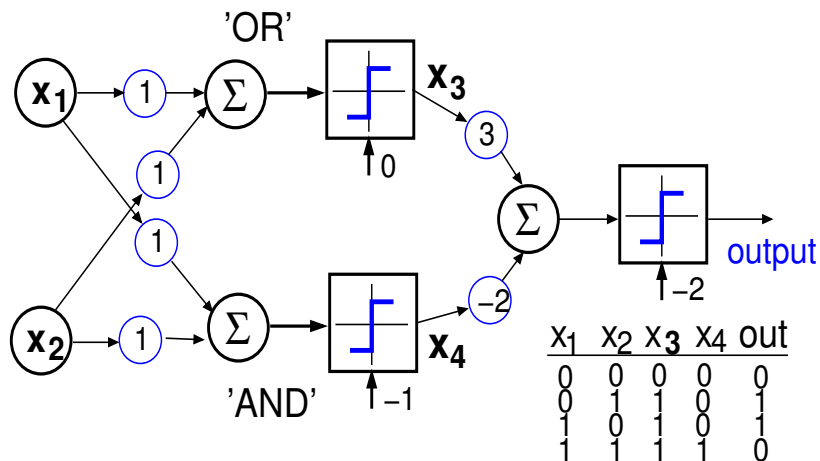
# Usefulness of complex numbers in machine intelligence

Example: A single complex neuron is worth as much as a real-valued MLP

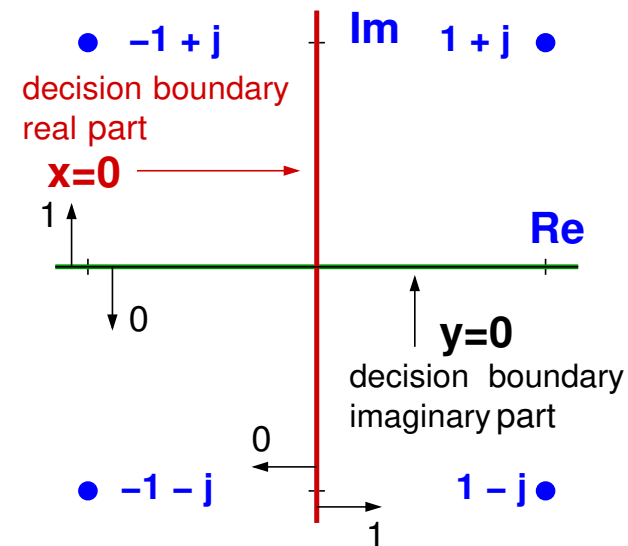
$x_1$	$x_2$	$z$	$P(z) = \text{XOR}$
1	1	$1 + j$	1
1	-1	$1 - j$	-1
-1	1	$-1 + j$	-1
-1	-1	$-1 - j$	1

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

In  $\mathbb{R}$  we need multiple layers, the so called **multilayer perceptron** shown below



But a single neuron in  $\mathbb{C}$  performs the XOR and is amplitude independent



---

# Part 1: Complex Calculus

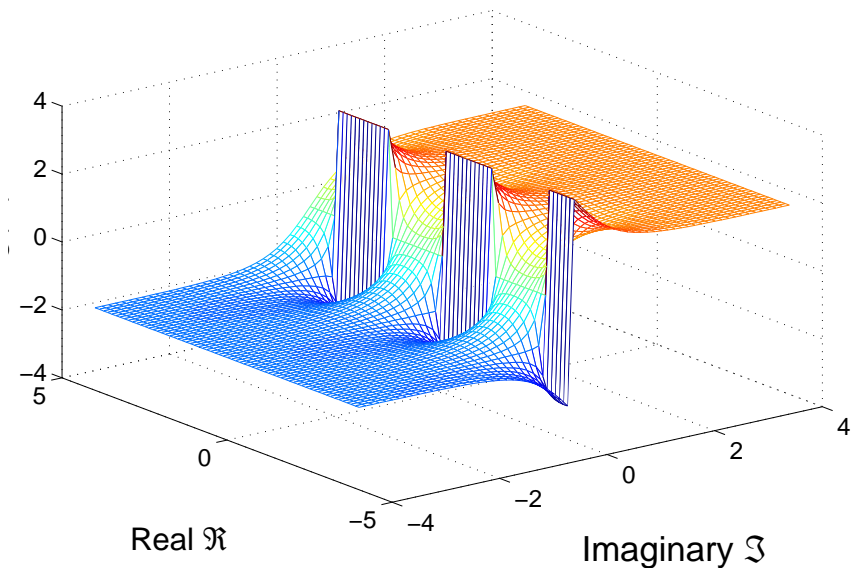
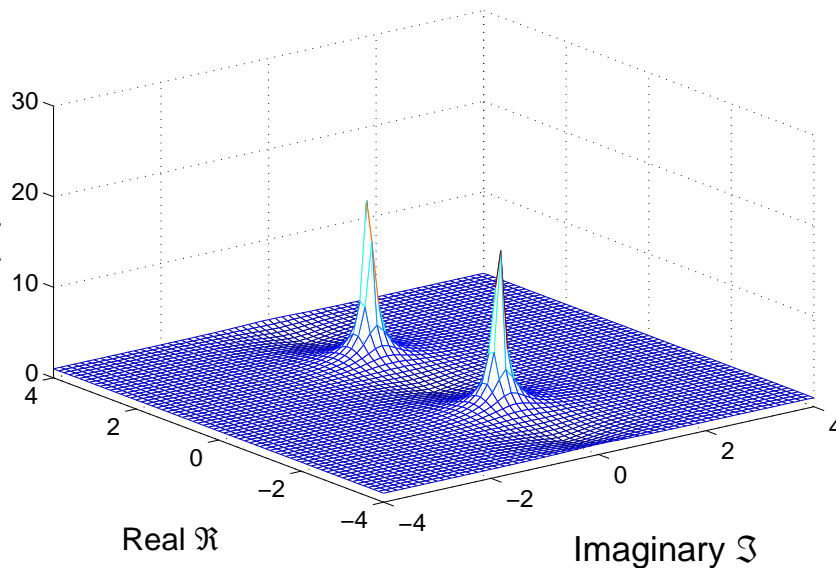
We will now introduce a modern perspective on complex calculus, the so-called **CR calculus** which offers much more flexibility in the differentiation of complex functions, and is indispensable in learning systems where the objective (cost) functions are typically real-valued functions of complex variables.

Such functions are not differentiable using the standard complex differentiation (Cauchy-Riemann), yet gradient based learning schemes require such derivatives.

We show that the CR-calculus applies both to the holomorphic (complex analytic) and non-holomorphic functions of complex variable, and will elucidate the use of the so-called ‘pseudo-gradient’.

## Difference with complex-valued functions

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



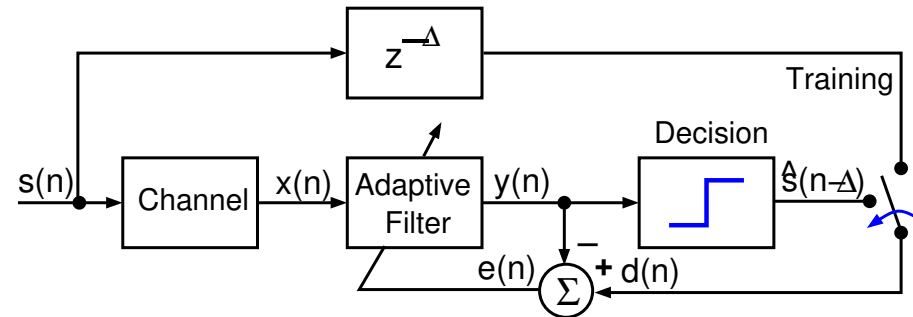
**Singularities:** Isolated singularities (removable singularities, poles, essential singularities), branch points, singularities at  $\infty$ .

In gradient based learning, we seek a coefficient vector  $\mathbf{w}$  using the so called **pseudo-gradient** of the cost function  $J = E\{|e|^2\} = E\{ee^*\}$ ,

$$\nabla_{\mathbf{w}} J(e, e^*) = \frac{\partial J}{\partial \mathbf{w}_r} + j \frac{\partial J}{\partial \mathbf{w}_i}$$

# Motivation: Example of complex-valued optimisation

Channel equalisation in wireless comms (reversing distortion of transmitted signals)



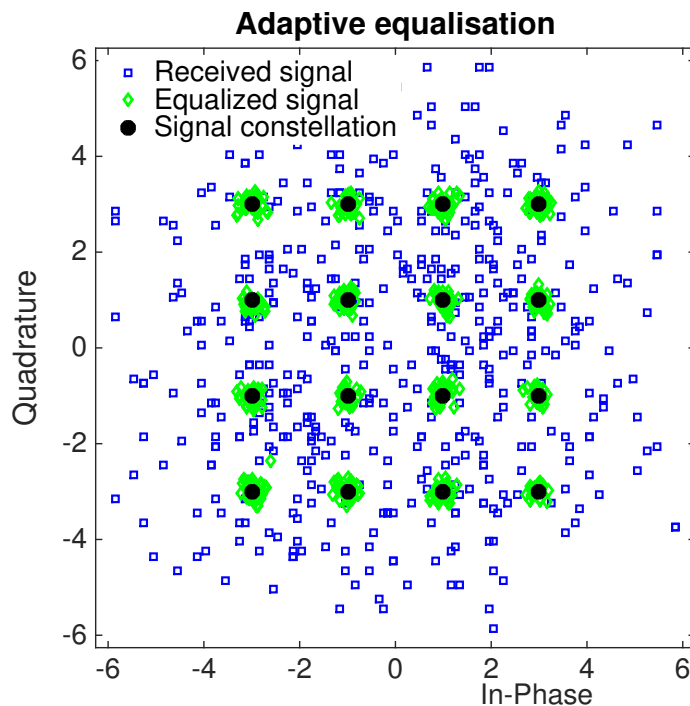
**Objective:** To minimize the cost function

$$J(\mathbf{w}) = |e(k)|^2 = e(k)e^*(k) \\ = e_r^2 + e_i^2 = u(k) + jv(k)$$

Here,  $u(k) = e_r^2 + e_i^2$  and  $v(k) = 0$

But,  $\frac{\partial u}{\partial e_r} = 2e_r \neq \frac{\partial v}{\partial e_i} = 0$ ,  $\frac{\partial u}{\partial e_i} = 2e_i \neq -\frac{\partial v}{\partial e_r} = 0$

∴ **Our  $J(\mathbf{w})$  is not differentiable in the Cauchy–Riemann sense** (here, error power is a real function of complex variables  $e(k)$  and  $e^*(k)$ ) **but the equaliser has to work!**



# Cauchy–Riemann derivatives are very restrictive!

**Recall:**  $f(z) = u(x, y) + jv(x, y) \rightarrow f'(z) = \partial u(x, y)/\partial x + j\partial v(x, y)/\partial x$

---

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

**Intuition:** The Jacobian matrix of  $f(z) = u + jv$ , is given by

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

**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  $z = x + jy$  and  $z^* = x - jy$  are not analytic**, for example

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

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

## The key: $\mathbb{C}$ -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.  $z = x + jy$ .

In standard Multivariate Calculus in  $\mathbb{R}^{N \times 1}$  the derivative of a function  $g(\mathbf{x})$ ,  $\mathbf{x} = [x_1, x_2, \dots, x_N]$  is defined as  $\frac{\partial g}{\partial \mathbf{x}} = \left[ \frac{\partial g}{\partial x_1}, \dots, \frac{\partial g}{\partial x_N} \right]^T$

- **Step 1:** Define the vector  $\mathbf{x} = [x, jy]^T$ , hence  $z = \mathbf{1}^T \mathbf{x}$ .
- **Step 2:** Express the derivative of  $f$  with respect to “real” vector  $\mathbf{x}$  i.e.  
$$\frac{\partial f}{\partial \mathbf{x}} = \left[ \frac{\partial f}{\partial x} \quad \frac{\partial f}{\partial jy} \right]^T \quad \text{(see the Appendix 3 for vector-valued derivatives)}$$
- **Step 3:** Transform the derivative vector in Step 2 back into  $\mathbb{C}$

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

- **Step 4:** Normalise the derivative since  $f$  is “differentiated twice”, to give the  $\mathbb{R}$ -derivatives (cf. differentiate wrt  $z^*$  for  $\mathbb{R}^*$ -derivatives)

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

# CR-derivatives of holomorphic functions

## CR-derivatives vs. standard $\mathbb{C}$ -derivatives for $f(z) = z$ & $f(z) = z^*$

If a function  $f = f(z, z^*) = u(x, y) + jv(x, y)$  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

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

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

↪ **For holomorphic functions the  $\mathbb{R}^*$ -derivative vanishes and the  $\mathbb{R}$ -derivative is equivalent to the standard complex derivative**

**Example:** (i)  $f(z) = z = x + jy \Rightarrow \mathbb{R} - \text{der} = 1 \quad \& \quad \mathbb{R}^* - \text{der} = 0$ ,  
(ii)  $f(z) = z^* = x - jy \Rightarrow \mathbb{R} - \text{der} = 0 \quad \& \quad \mathbb{R}^* - \text{der} = 1$



## Example 3: The CR-calculus for non-holomorphic functions

---

Consider a real function of complex variable  $f(z) = |z|^2 = zz^* = x^2 + y^2$ , where  $z = x + jy$  (clearly non-holomorphic). Assuming  $z \perp z^*$ , the  $\mathbb{R}$ -derivative and the conjugate  $\mathbb{R}^*$ -derivative are

$$\frac{\partial f}{\partial z} = \frac{\partial(zz^*)}{\partial z} = z^* \quad \text{and} \quad \frac{\partial f}{\partial z^*} = \frac{\partial(zz^*)}{\partial z^*} = z$$

To verify, start from

$$f(z) = f(u(x, y) + jv(x, y)) = f(u, v) = x^2 + y^2$$

where  $u(x, y) = x^2 + y^2$  and  $v(x, y) = 0$ . Therefore,

$$\text{R-der :} \quad \frac{\partial f}{\partial z} = \frac{1}{2} \left[ \frac{\partial(x^2 + y^2)}{\partial x} - j \frac{\partial(x^2 + y^2)}{\partial y} \right] = \frac{1}{2} [2x - j2y] = x - jy = z^*$$

$$\text{R}^* \text{-der :} \quad \frac{\partial f}{\partial z^*} = \frac{1}{2} \left[ \frac{\partial(x^2 + y^2)}{\partial x} + j \frac{\partial(x^2 + y^2)}{\partial y} \right] = \frac{1}{2} [2x + j2y] = x + jy = z$$

## Example 4: Some typical $\mathbb{C}\mathbb{R}$ -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.

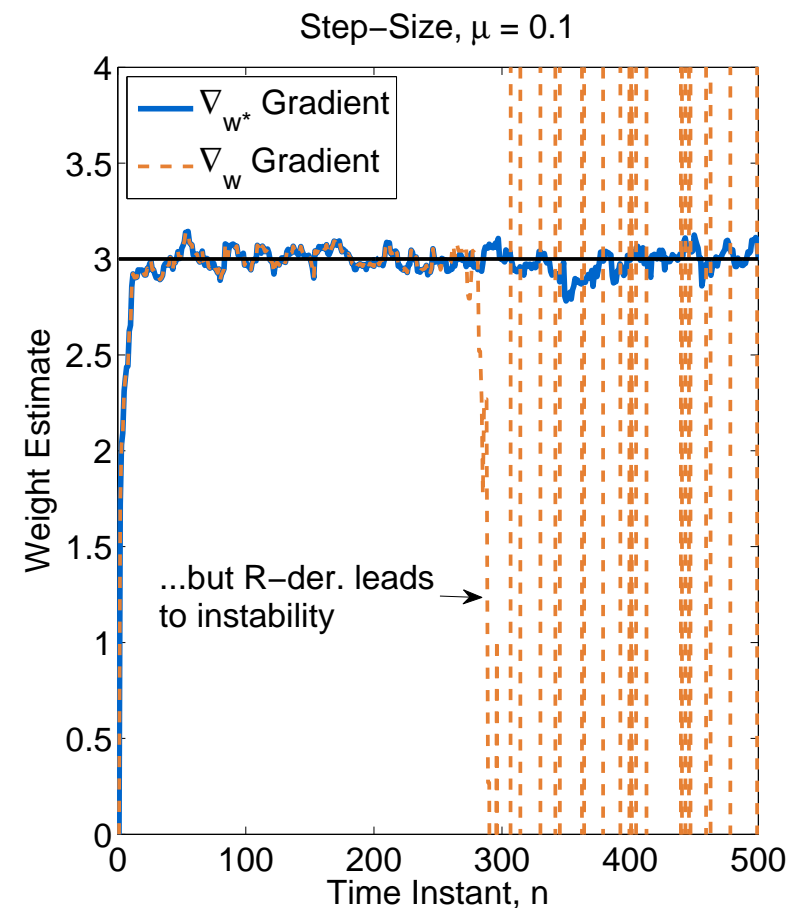
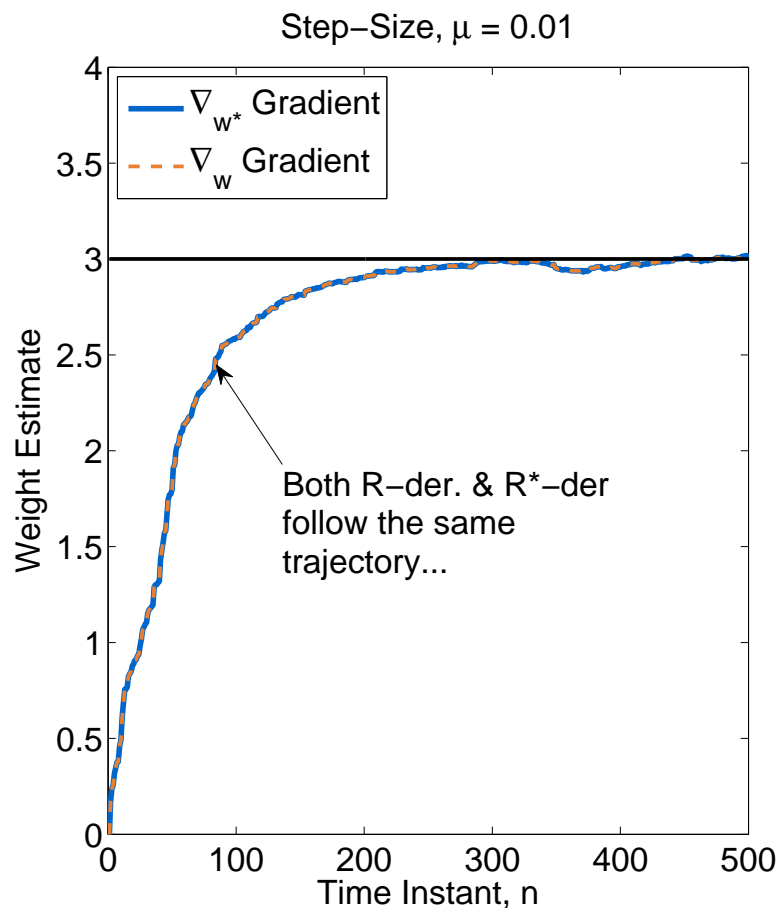
$f(z, z^*)$	$\mathbb{R}$ –der	$\mathbb{R}^*$ –der	$\mathbb{C}$ –der
$z$	1	0	1
$z^*$	0	1	Undefined
$ z ^2 = zz^*$	$z^*$	$z$	Undefined
$z^2 z^*$	$2 z ^2$	$z^2$	Undefined
$e^z$	$e^z$	0	$e^z$

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?

An example from learning systems:  $\mathbb{R}$ -der vs.  $\mathbb{R}^*$ -der? (more later)

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



---

## Part 2: Complex Statistics

Now that we have familiarised ourselves with the concept of (non-)circularity, we will examine how to use the concept in the domain of second-order statistics and how to design so-called **widely linear** estimators which are second-order optimal for both second-order circular (proper) and second-order noncircular (improper) data.

# Isomorphism between $\mathbb{C}$ and $\mathbb{R}^2$

## Moving from real-valued to complex-valued data

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

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

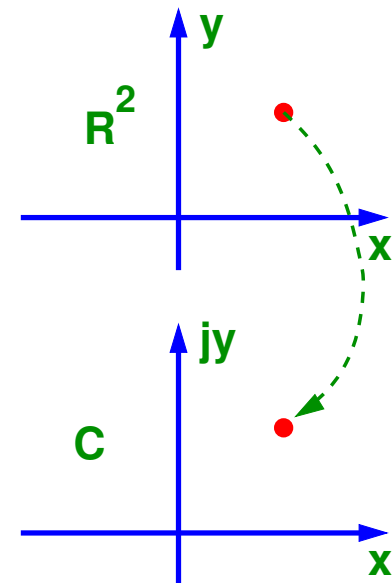
$$\mathbf{z} \rightarrow \mathbf{z}^a \quad \leftrightarrow \quad \begin{bmatrix} \mathbf{z} \\ \mathbf{z}^* \end{bmatrix} = \begin{bmatrix} \mathbf{I} & j\mathbf{I} \\ \mathbf{I} & -j\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}$$

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

$$\mathbf{v} = [z_1, z_1^*, \dots, z_N, z_N^*]^T$$

$$\mathbf{v} = \mathbf{A}\mathbf{w}, \quad \mathbf{w} = [x_1, y_1, \dots, x_N, y_N]^T$$

where matrix  $\mathbf{A} = \text{diag}(\mathbf{J}, \dots, \mathbf{J}) \in \mathbb{C}^{2N \times 2N}$  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  $\mapsto$  pdf introduced via duality with  $\mathbb{R}$

---

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

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

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.

# Circular complex random variables

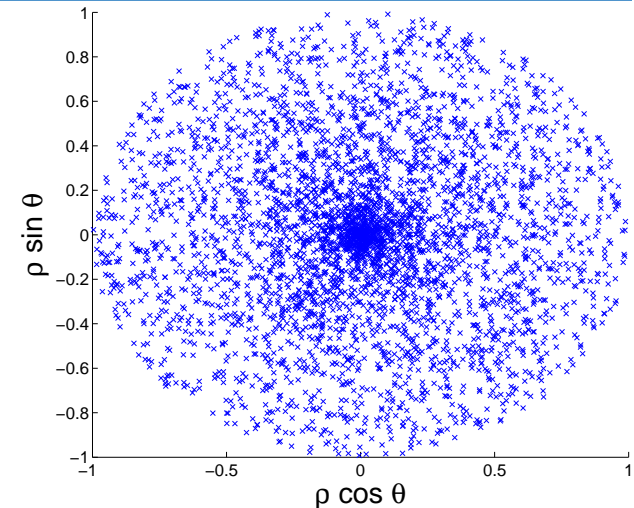
Try to generate complex ran. var. from various distrib. in MATLAB

**Circularity**  $\leadsto$  **Rotation invariant distrib.**

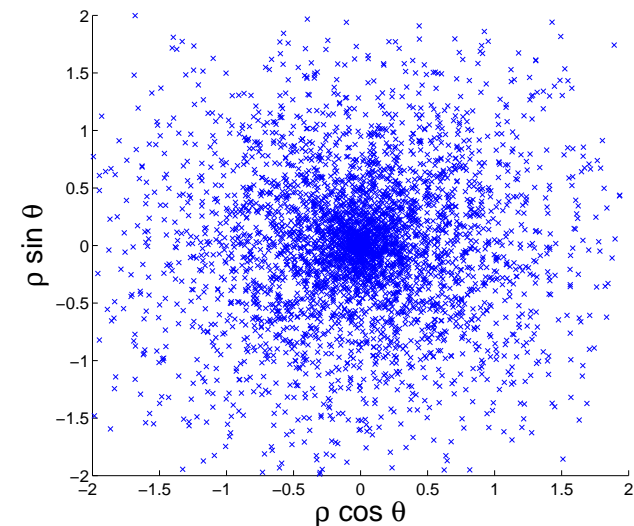
$$p(\rho, \theta) = p(\rho, \theta - \phi)$$

1. The name of the distribution takes after the distribution of the real-valued random variable  $\rho$  with a pdf  $p(\rho)$ ;
2. It can be Gaussian, uniform, etc.
3. Take another real-valued random variable  $\theta$ , which must be uniformly distributed on  $[0, 2\pi]$  and independent of  $\rho$ ;
4. Construct the complex random variable  $Z = X + jY$  as

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



(g) Uniform circular



(h) Gaussian circular

## Complex circularity

---

**Definition:** A complex-valued random is called **circular** if its probability distribution is not dependent on the angle, that is, the distribution is “**rotation invariant**”.

For simplicity, we consider univariate complex-valued random variables; the concepts are readily extended to the multivariate case.

Recall that for an iid complex-valued random variable  $Z = X + jY$ , the pdf

$$\mathcal{P}_Z(z) = \mathcal{P}_X(x)\mathcal{P}_Y(y)$$

On the other hand, in the case of a **rotation invariant**  $\mathcal{P}_Z(z)$ , its pdf is only be dependent of the **Euclidean distance** from the origin in the complex domain. Therefore, if the random variable  $Z$  is circular, we have

$$g(r) = \mathcal{P}_Z(z) = \mathcal{P}_X(x)\mathcal{P}_Y(y)$$

where  $r = \sqrt{x^2 + y^2}$  and  $g(\cdot)$  is a general function.

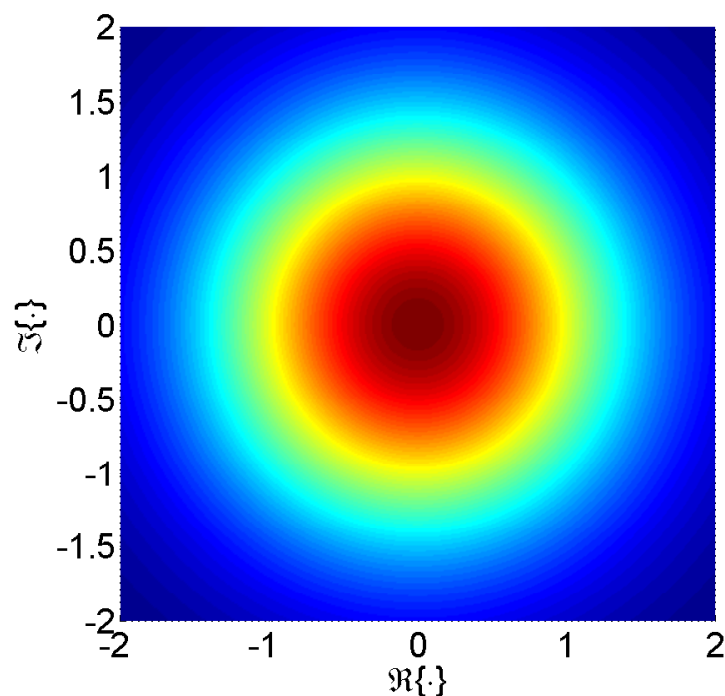


# Circularity (inherently assumed in standard processing in $\mathbb{C}$ )

## Some circular distributions

### Circular complex-valued random variables

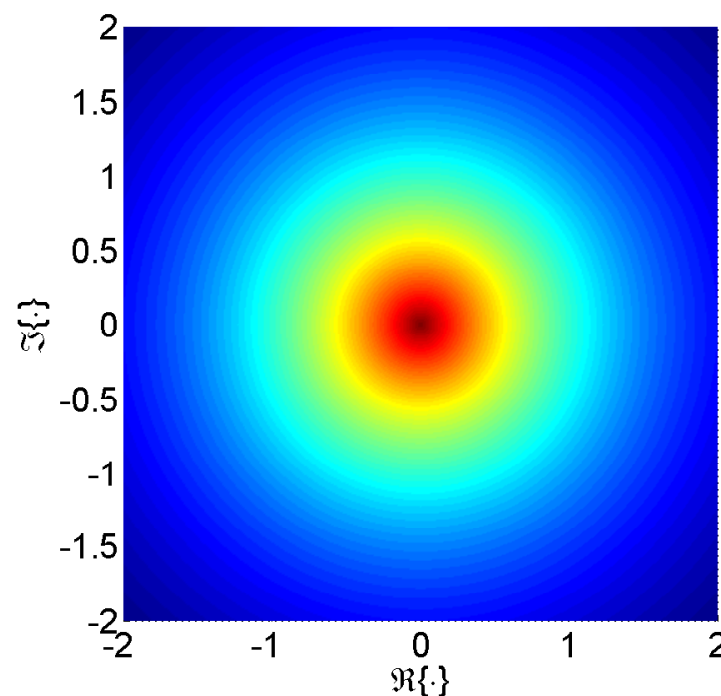
The distribution of  $R$  is Rayleigh.  
Thus, the distributions of the real  
and imaginary parts are Gaussian.



circular Rayleigh distribution

The distribution of  $R$  is exponential

$$\mathcal{P}_R(r) = \lambda e^{-\lambda r}, \quad \lambda = 1$$



circular exponential distribution

# More on circularity

## A noncircular distribution

### Independent real & imaginary distributions but not circular!

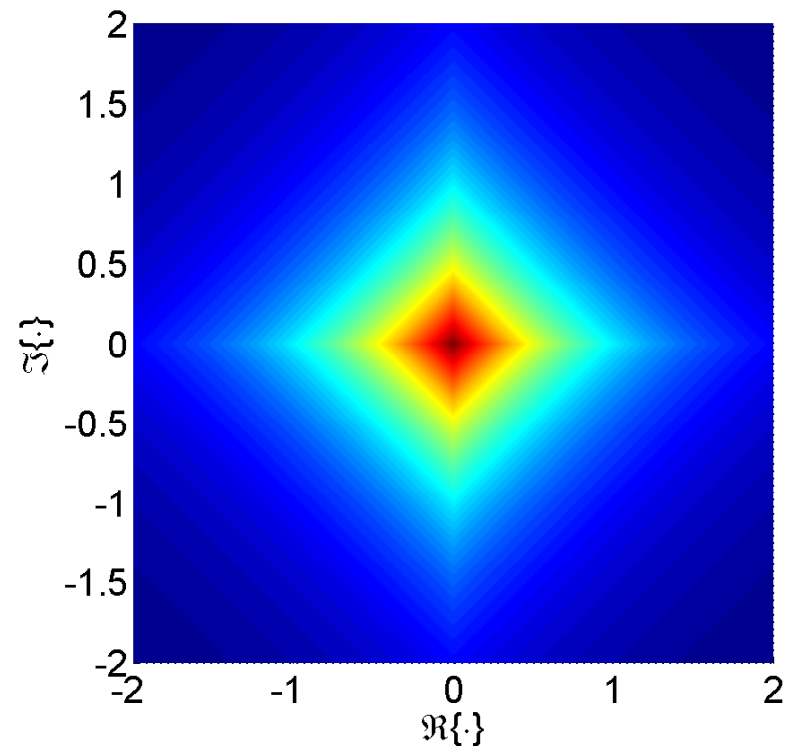
Distributions of the real and imaginary part are **independent Laplace distributions**

$$\mathcal{P}_X(x) = \frac{1}{2}e^{-|x|} \text{ and } \frac{1}{2}\mathcal{P}_Y(y) = \frac{1}{2}e^{-|y|}$$

Thus,

$$\mathcal{P}_Z(z = x + jy) = \frac{1}{4}e^{-(|x|+|y|)}$$

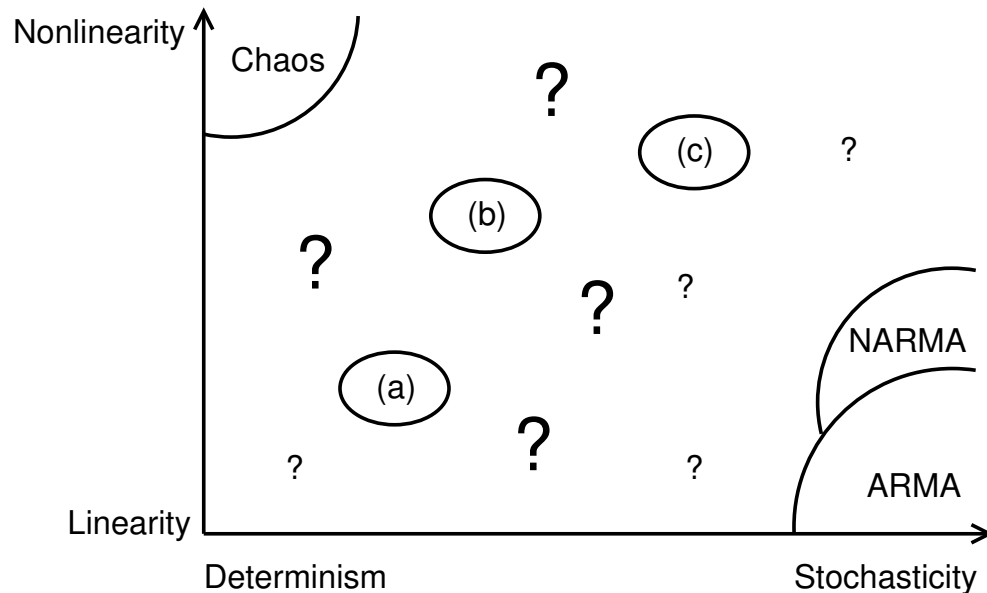
Although the distributions on the real and imaginary axes are independent and hence uncorrelated, the resulting distribution is not rotation invariant, that is, it is non-circular.



# Signal modality – So why are complex signals different?

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

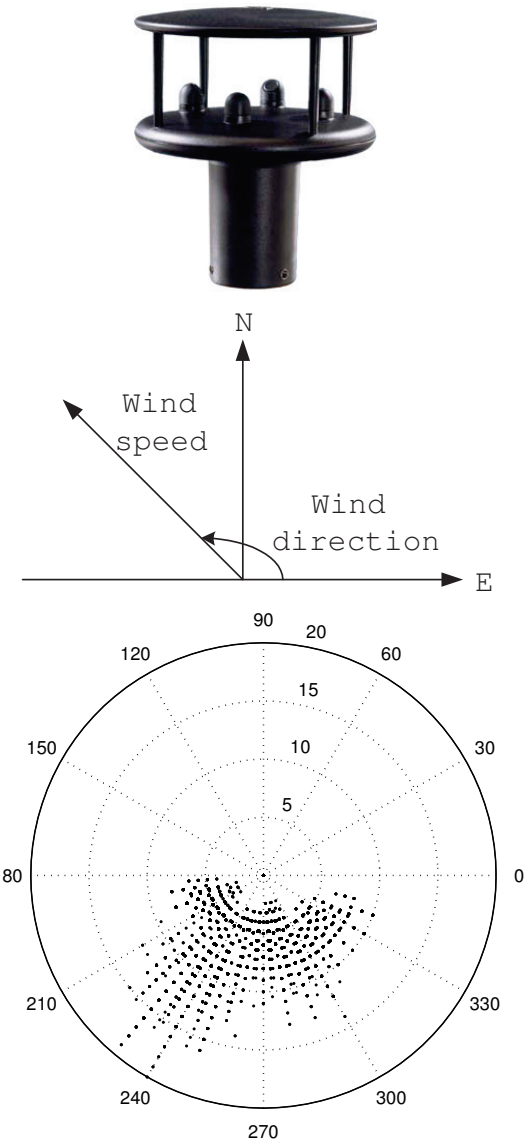
Deterministic vs. Stochastic nature  
Linear vs. Nonlinear nature



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?
- $\rightarrow$  **degree of noncircularity**



## Other definitions of circularity

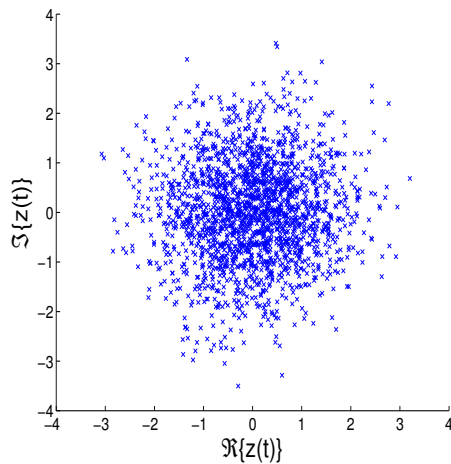
Via Probability density function, Characteristic Function, Cumulants

- *Probability density function.* A complex random variable  $Z$  is circular if its pdf is a function of only the product  $zz^*$ , that is<sup>1</sup>

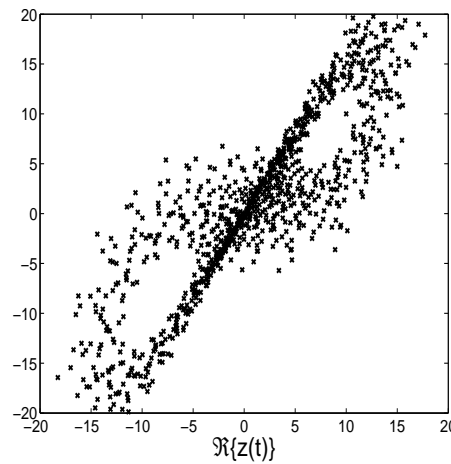
$$p_{Z,Z^*}(z, z^*) = p_{Z_\phi, Z_\phi^*}(z_\phi, z_\phi^*)$$

and for Gaussian CCRVs we have

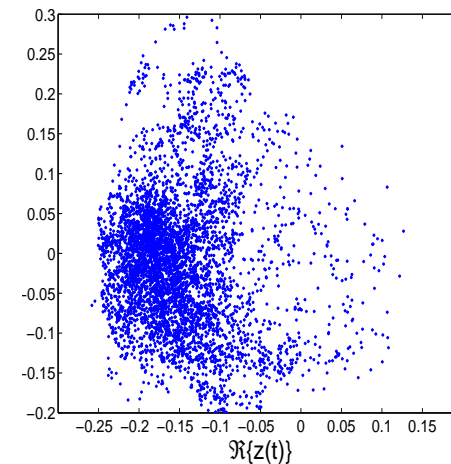
$$p_{Z,Z^*}(z, z^*) = \frac{1}{\pi\sigma^2} e^{-zz^*/\sigma^2}$$



(i) Complex AR(4)



(j) Complex Lorenz

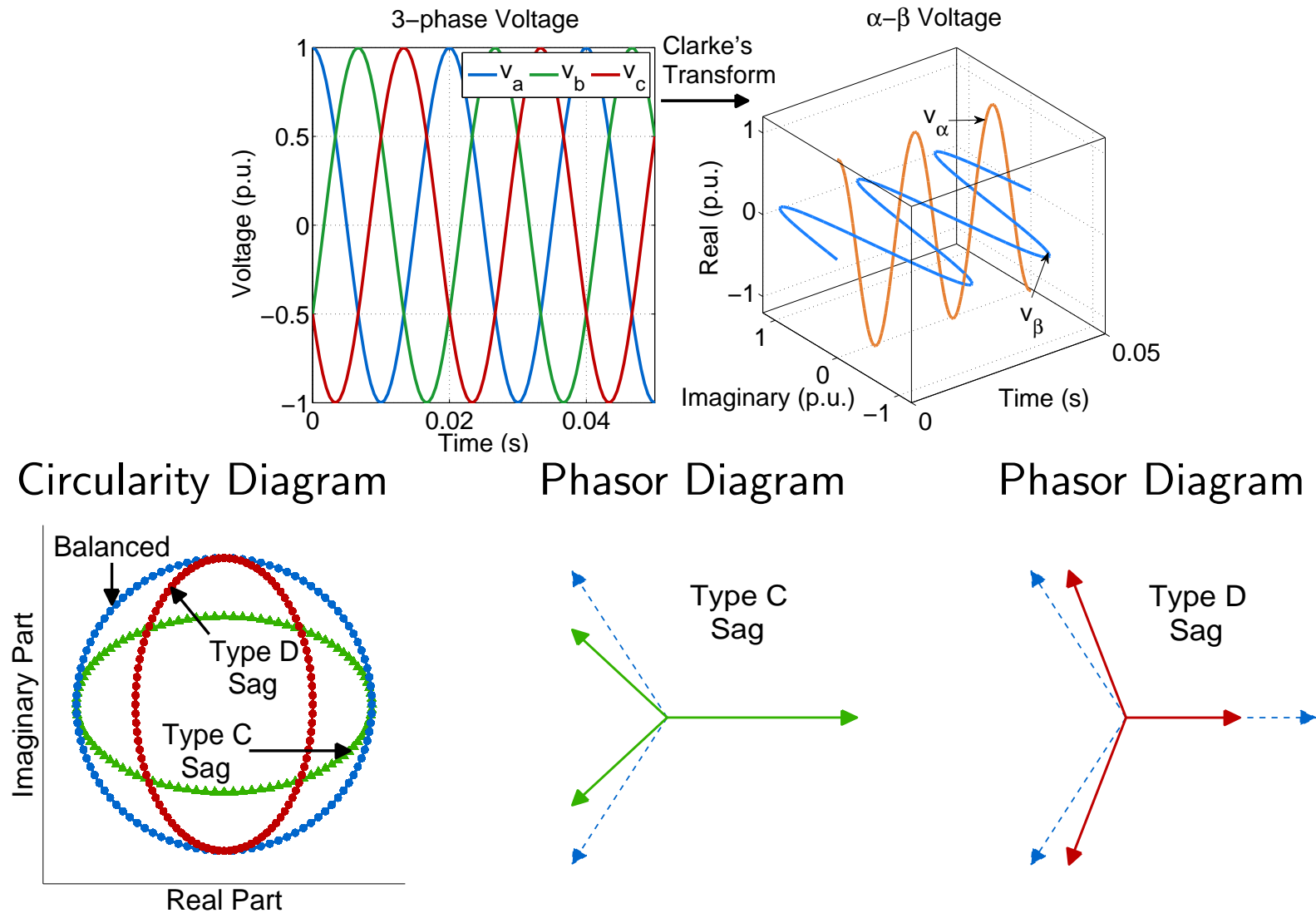


(k) Complex wind

<sup>1</sup>The pdf of a circular complex random variable is function of only the modulus of  $z$ , **and not of  $z^*$** .

# Example: Noncircularity, a fault signature in power grid

## Visualising the Clarke transform and noncircular voltage sags



## **Key:** What are we doing wrong $\leadsto$ 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

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

$$\text{thus} \quad \hat{y} = E[y_r|x_r, x_i] + jE[y_i|x_r, x_i]$$

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

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

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

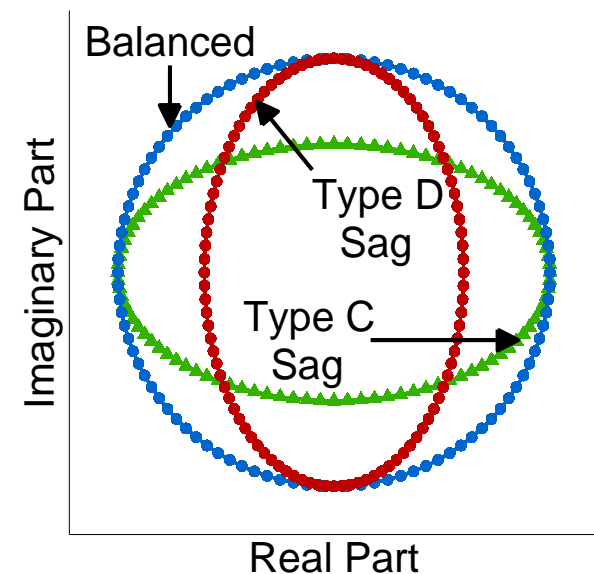
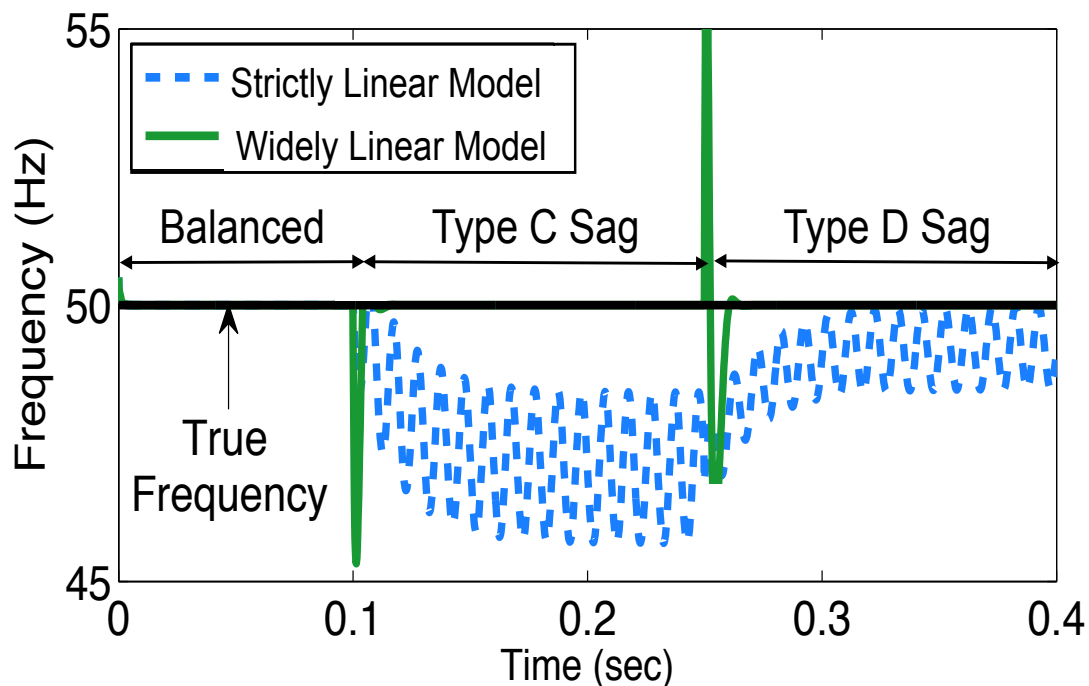
$$y = \mathbf{h}^H \mathbf{x} + \mathbf{g}^H \mathbf{x}^*$$

**We can now process general (noncircular) complex signals!**

## Example: Enhanced degrees of freedom using WLM

### WLM for frequency estimation in power grid

- We are jumping a little ahead, here the idea is just to show that by the additional degree of freedom in WLM, we can greatly improve the performance of any signal processing algorithm in  $\mathbb{C}$
- ↪ Figure below: The widely linear model was 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  $\mathbf{z} = \mathbf{x} + jy$ , ‘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

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

**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:*

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

**Remark #4:** General complex random processes are *improper*.

**‘Properness’ is a second order statistical property and ‘circularity’ is a property of the probability density function.**



## Measuring improperness $\leadsto$ intuitive example

---

Consider the estimation of a zero-mean complex r.v.  $z \in \mathbb{C}$  from its conjugate, that is

$$\hat{z} = h z^*$$

**Solution:** Find an estimate of  $h$  that minimises

$$J_{\text{MSE}} = E[|e|^2] = E[|z^* - \hat{z}^*|^2]$$

The Wiener solution is then

$$h_{\text{opt}} = E[zz^*]^{-1} E[zz] = \frac{p}{c} = \rho_z$$

where  $\rho_z$  is referred to as the **circularity quotient**. We now have

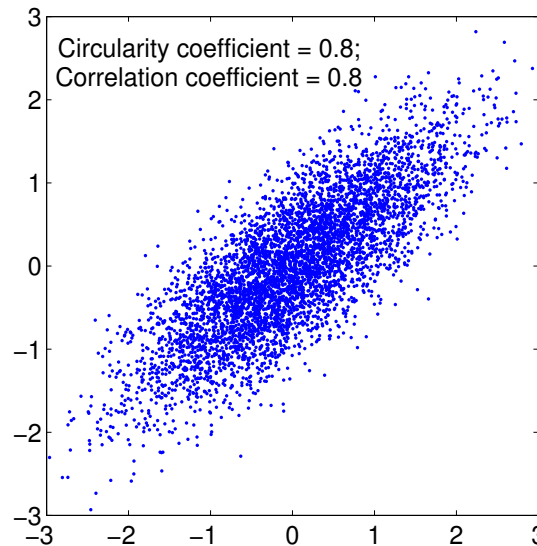
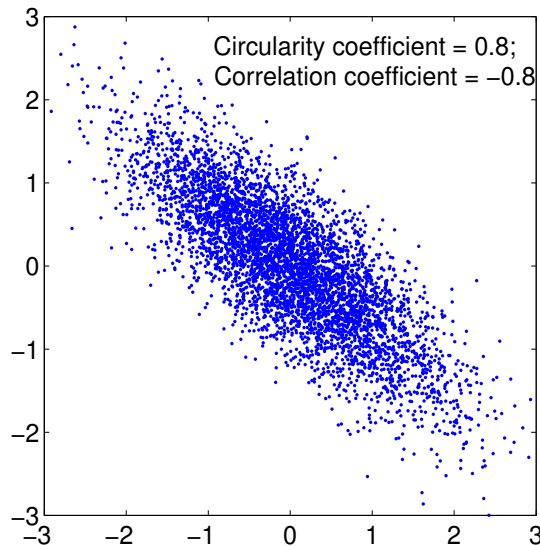
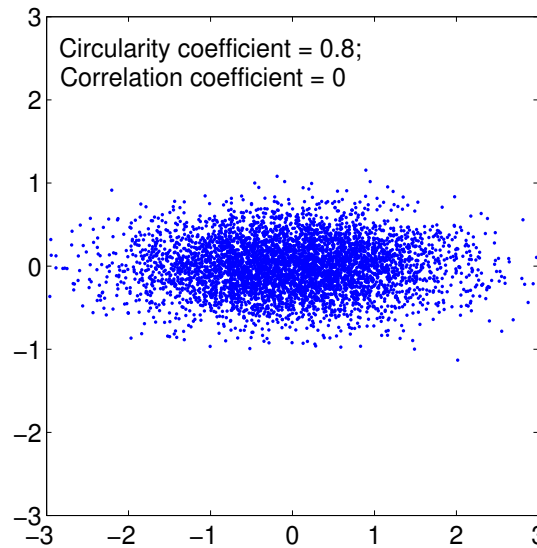
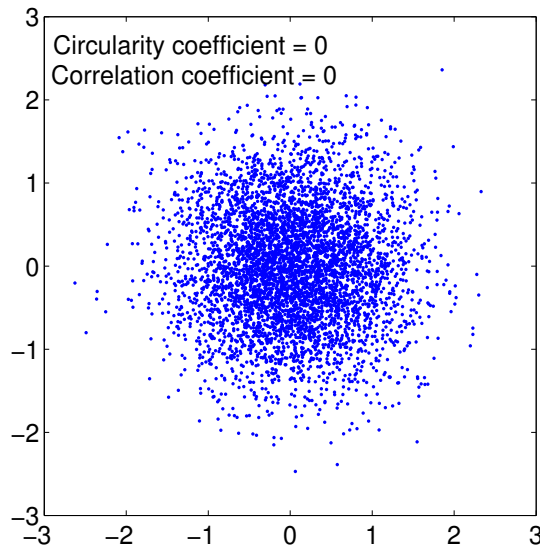
$$\rho_z = \frac{1}{c} \left( \sigma_x^2 - \sigma_y^2 + 2j c_{xy} \right)$$

where the real part of  $\rho_z$  gives the power difference between the real and imaginary parts while the imaginary part of  $\rho_z$  models their correlation (both normalized by total power).

Now, the **circularity coefficient**  $\eta = \frac{|p|}{c} \quad 0 \leq \eta \leq 1$

# 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!

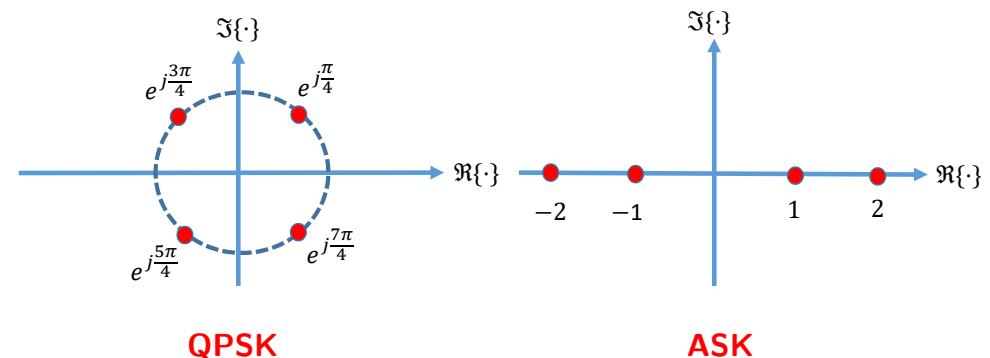
**Recall:** Real valued ICA cannot separate two Gaussian signals.

# Circularity

## Constellations in communications, 4 symbols

Consider a communication system with 4 complex-valued symbols.

The most widely used modulation schemes are quadrature phase shift keying (**QPSK**) and amplitude shift keying (**ASK**).



Although these constellations are arranged so that the distances of each point to its nearest neighbour is equal in both cases, the **QPSK is more compact**.

### **QPSK second-order statistics:**

$$\text{covariance : } c = E[zz^*] = 1$$

$$\text{pseudocov. : } p = E[zz] = 0$$

### **ASK second-order statistics:**

$$\text{covariance : } c = E[zz^*] = 2.5$$

$$\text{pseudocov. : } p = E[zz] = 2.5$$

In the case of the **QPSK** there is no power difference or correlation between the real and imaginary components, resulting in the impropriety measure of  $\rho = 0$ .

In the case of the **ASK** all the information is on the real axis, resulting in the impropriety measure of  $\rho = 1$  (real-valued signals are maximally non-circular).

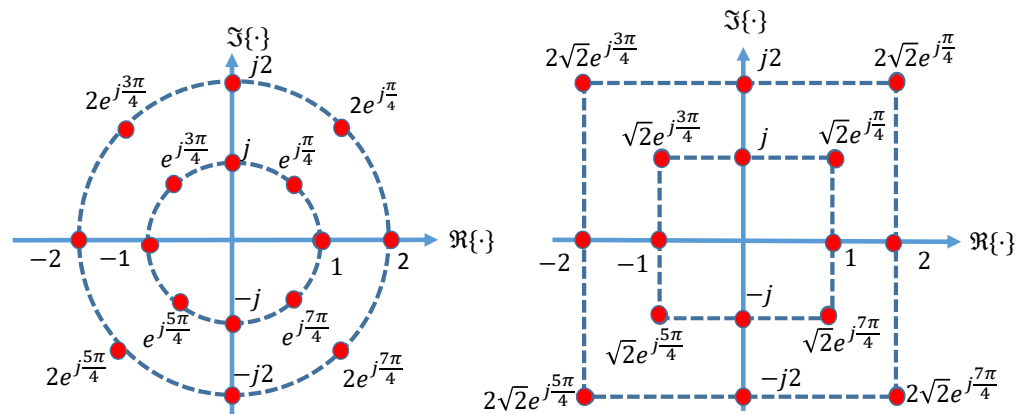
# Circularity in communications

## Constellations in communications, 16 symbols

Now, consider a communication system with 16 complex-valued symbols.

The most widely used modulation schemes are the amplitude and quadrature phase shift keying (**APSK**) and quadrature amplitude modulation (**QAM**).

Note that the constellation for **16-APSK** is more **compact** than that of the **16-QAM**.



**16-APSK**

**16-QAM**

**16-APSK second-order statistics:**

$$c = E[zz^*] = 2.5$$

$$p = E[zz] = 0$$

**16-QAM second-order statistics:**

$$c = E[zz^*] = 3.75$$

$$p = E[zz] = 0$$

Although **both methods are proper**, **only the 16-APSK is circular** (loosely speaking). Note that **circular** constellations offer better **energy efficiency**, whereas **non-circular constellations are more resilient to noise**, especially when using widely-linear processing.

## 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

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

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

Widely linear model

Widely linear normal equations

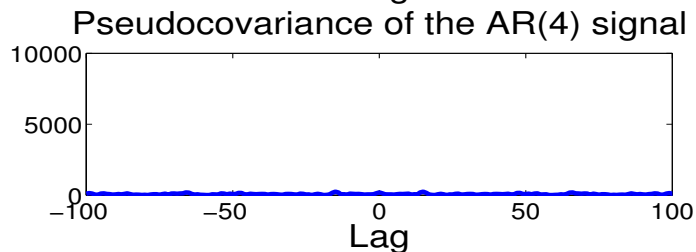
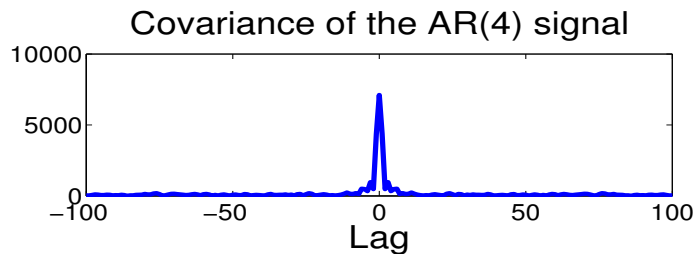
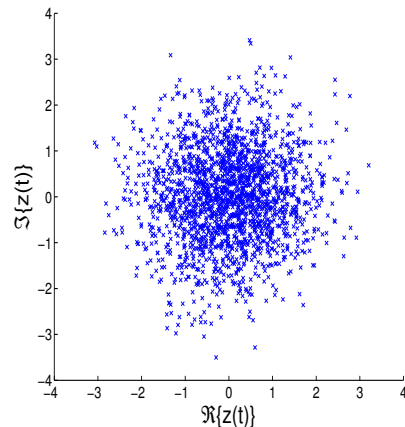
$$y(k) = \mathbf{h}^T(k) \mathbf{x}(k) + \mathbf{g}^T(k) \mathbf{x}^*(k) + q(k)$$
$$\begin{bmatrix} \mathbf{h}^* \\ \mathbf{g}^* \end{bmatrix} = \begin{bmatrix} \mathbf{C} & \mathbf{P} \\ \mathbf{P}^* & \mathbf{C}^* \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{c} \\ \mathbf{p}^* \end{bmatrix}$$

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

## Example 5: Pseudocovariance $\longleftrightarrow$ properness

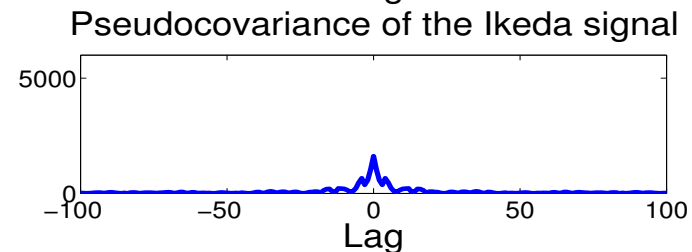
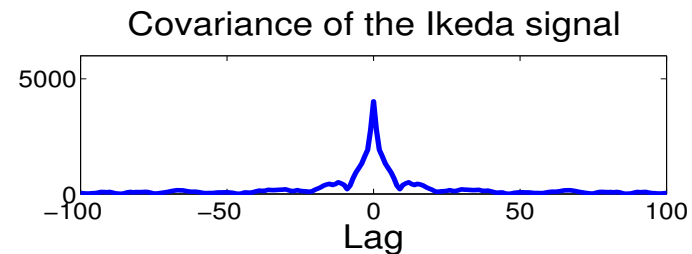
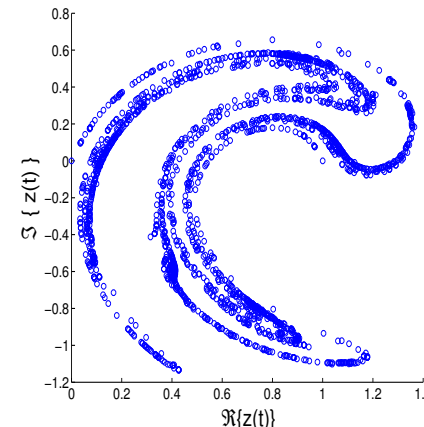
Real-world data are rarely circular (if nothing else  $\rightarrow$  short length, artefacts)?

Complex AR(4) process (circular)



Complex AR(4) process (proper)

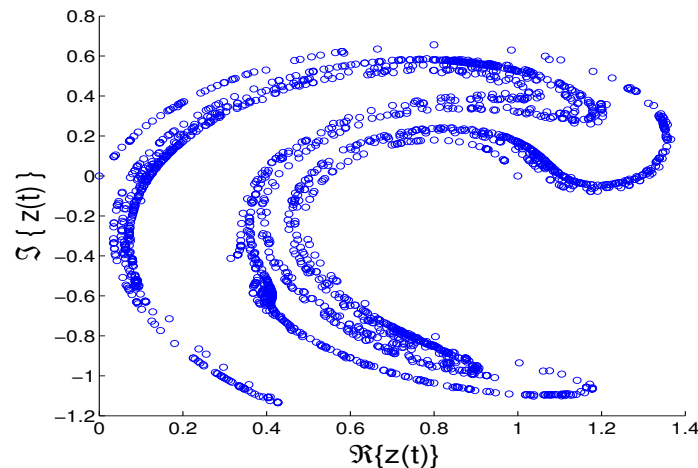
Complex Ikeda map (noncircular)



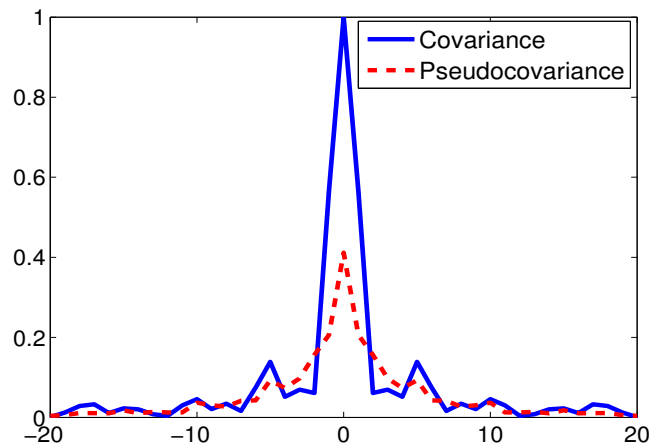
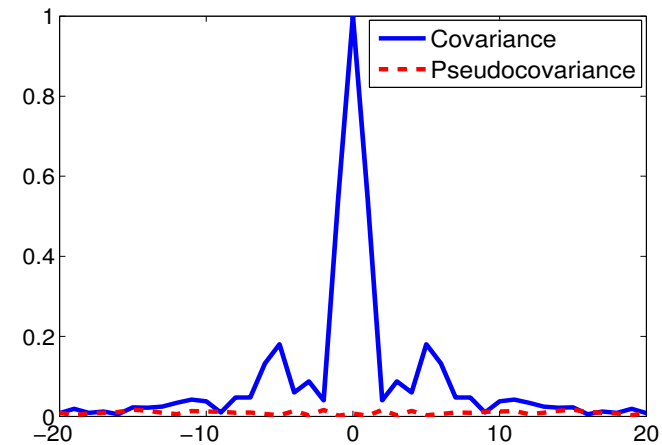
Complex Ikeda map (improper)

**Key:** WLM caters for both proper and improper data  
This is a rigorous way to model general complex signals!

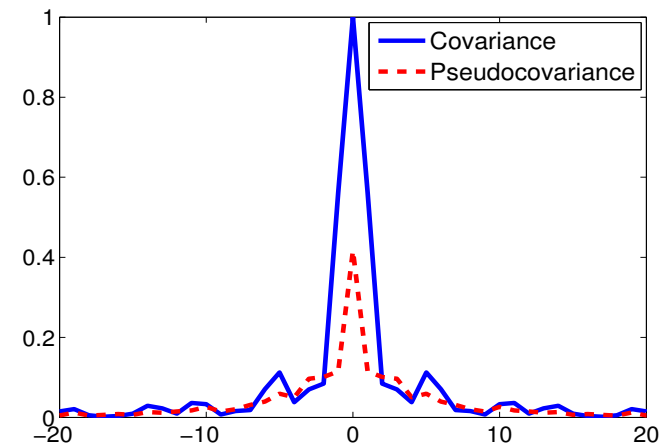
Circularity for Ikeda map



AR model of Ikeda signal



Covariances: Original Ikeda



Widely linear AR of Ikeda

## Lecture summary

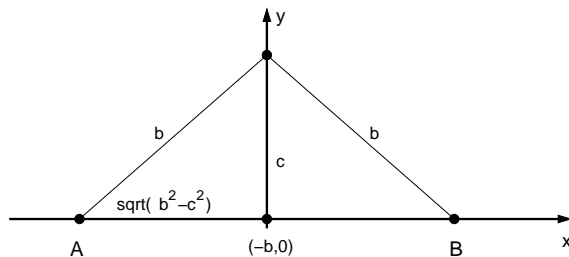
---

- We have demystified several basic concepts in complex calculus
- Problems with the Cauchy-Riemann derivatives
- The CR-calculus deals with both analytic and non-analytic functions
- This has made possible optimisation of real functions of complex variables (typical minimum error power cost functions)
- Complex noncircularity  $\leadsto$  a mathematical microscope into data behaviour, and a signature of many conditions
- Circularity  $\leadsto$  property of a probability distribution, properness is a second order statistical property (pseudocovariance vs covariance)
- Widely linear modelling  $\leadsto$  deals with both proper and improper signals
- Examples in communications and smart grid

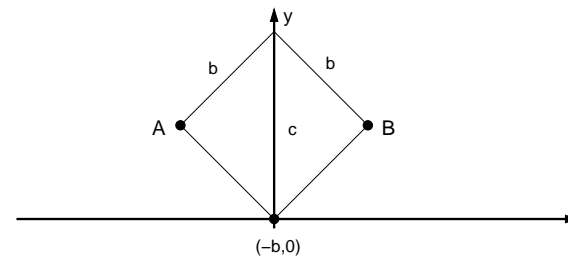


## Some trivia: The role of geometry

- Complex numbers were only accepted after they had a geometric interpretation, but it was initially 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

$$(\cos \theta + j \sin \theta)^n = \cos n\theta + j \sin n\theta$$

$$\cos \theta + j \sin \theta = e^{j\theta}$$

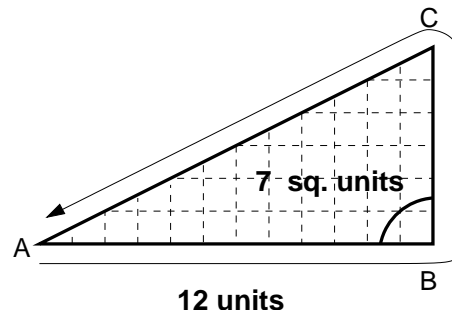
- In 1806 Argand interpreted  $j = \sqrt{-1}$  as a rotation by  $90^\circ$  and introduced Argand diagram,  $z = x + jy$ , and the modulus  $\sqrt{x^2 + y^2}$ .
- In 1831 Karl Friedrich Gauss introduced  $i = \sqrt{-1}$  and complex algebra.

## Some trivia: 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

$$area = \frac{1}{2} x h$$

$$perimeter = x + h + \sqrt{x^2 + h^2}$$

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.

# Recap: What is a derivative?

we need to understand where the pseudo-gradient comes from

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

$$f'(x) = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

For a complex function

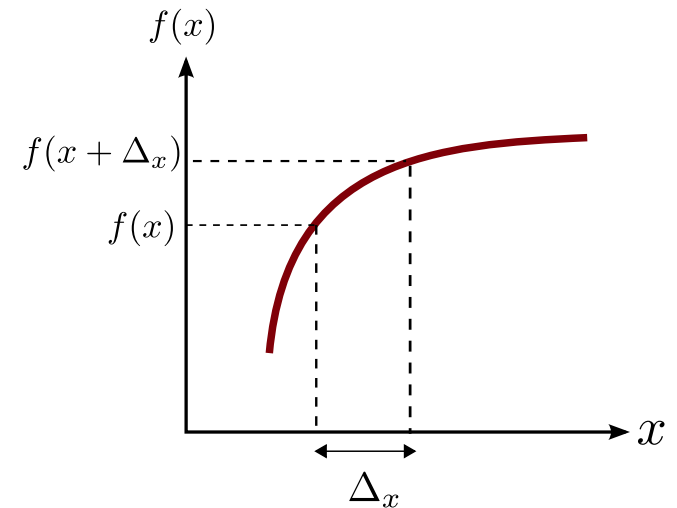
$$f(z) = u(x, y) + jv(x, y)$$

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

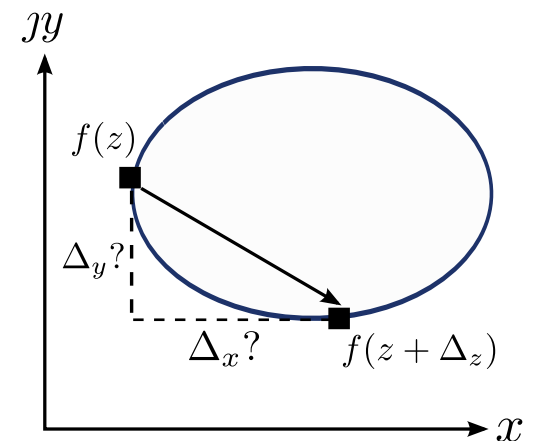
$$f'(z) = \lim_{\Delta z \rightarrow 0} \frac{f(z + \Delta z) - f(z)}{\Delta z}$$

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

Real-Domain:



Complex-Domain:



# Recap: Complex derivatives, 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  $\Delta z$  approaches zero

$$f'(z) = \lim_{\substack{\Delta x \rightarrow 0 \\ \Delta y \rightarrow 0}} \frac{[u(x + \Delta x, y + \Delta y) + jv(x + \Delta x, y + \Delta y)] - [u(x, y) + jv(x, y)]}{\Delta x + j\Delta y}$$

must exist regardless of how  $\Delta z$  approaches zero. It is convenient to consider the two following cases for the  $\mathbb{C}$ – derivatives:

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

$$\begin{aligned} f'(z) &= \lim_{\Delta x \rightarrow 0} \frac{[u(x + \Delta x, y) + jv(x + \Delta x, y)] - [u(x, y) + jv(x, y)]}{\Delta x} \\ &= \lim_{\Delta x \rightarrow 0} \frac{u(x + \Delta x, y) - u(x, y)}{\Delta x} + j \frac{v(x + \Delta x, y) - v(x, y)}{\Delta x} \\ &= \frac{\partial u(x, y)}{\partial x} + j \frac{\partial v(x, y)}{\partial x} \end{aligned}$$

# Recap: Complex derivatives, Cauchy-Riemann conditions

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

---

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

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

For continuity, the limits from Case 1 and Case 2 must be identical, which yields **the Cauchy-Riemann equations**

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

This introduces a tremendous amount of structure (restrictions) in the calculus, as shown in an intuitive example on the next two slides.

# Appendix 1: Noncircularity and I/Q imbalance $\leadsto$ A proof

---

Derivation:

The modulated passband signal  $x_p(n)$  is given by

$$\begin{aligned} x_p(n) &= [s_I(n) * h_I(n)] \cos \omega_c n - [s_Q(n) * h_Q(n)] g \sin(\omega_c n + \varphi) \\ &= \underbrace{[s_I(n) * h_I(n) + g \sin \varphi s_Q(n) * h_Q(n)]}_{x_I(n)} \cos \omega_c n - \underbrace{g \cos \varphi}_{x_Q(n)} \sin \omega_c n \end{aligned}$$

Upon extracting the baseband signal from  $x_p(n)$ , and taking the in-phase and quadrature branches as the real and imaginary parts of  $x(n)$ , we have

$$\begin{aligned} x(n) &= x_I(n) + jx_Q(n) \\ &= \underbrace{\frac{1}{2}[h_I(n) + ge^{-j\varphi}h_Q(n)]}_{\mu(n)} * s(n) + \underbrace{\frac{1}{2}[h_I(n) - ge^{-j\varphi}h_Q(n)]}_{\nu(n)} * s^*(n) \end{aligned}$$

where  $s(n) = s_I(n) + js_Q(n)$

In a narrow-band scenario, the I/Q imbalance becomes frequency-independent, that is,  $h_I(n) = h_Q(n) \approx \delta(n)$ , and so

$$x(n) = \underbrace{\frac{1}{2}[1 + ge^{-j\varphi}]}_{\mu} s(n) + \underbrace{\frac{1}{2}[1 - ge^{-j\varphi}]}_{\nu} s^*(n)$$

## Appendix 2: The depressed cubic (so called 'cubic formula') implicitly uses complex numbers

---

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

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

## Appendix 3: Derivatives of a multivariate function

---

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

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

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  $\nabla_{\mathbf{z}^*} f(\mathbf{z}, \mathbf{z}^*) = \mathbf{0}$ ,

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



## Appendix 4: Some useful examples from CR-calculus

---

For proofs see lecture supplement

$$\text{Linear Form: } \frac{\partial}{\partial \mathbf{x}^*} \{ \mathbf{x}^T \mathbf{a} \} = \mathbf{0}$$

$$\text{Linear Form: } \frac{\partial}{\partial \mathbf{x}^*} \{ \mathbf{x}^H \mathbf{a} \} = \mathbf{a}$$

$$\text{Quadratic Form: } \frac{\partial}{\partial \mathbf{x}^*} \{ \mathbf{x}^H \mathbf{C} \mathbf{x} \} = \mathbf{C} \mathbf{x}$$

$$\text{Quadratic Form: } \frac{\partial}{\partial \mathbf{x}^*} \{ \mathbf{x}^T \mathbf{C} \mathbf{x}^* \} = \mathbf{C}^T \mathbf{x}$$

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

## Appendix 4: Some useful examples from $\mathbb{CR}$ -calculus

---

### Chain Rule

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

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

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

### Matrix Derivatives

$$\text{Linear Form: } \frac{\partial}{\partial \mathbf{B}^*} \{ \text{Tr } \mathbf{B}^* \mathbf{C} \} = \mathbf{C}^T$$

$$\text{Quadratic Form: } \frac{\partial}{\partial \mathbf{A}^*} \{ \text{Tr } \mathbf{A} \mathbf{C} \mathbf{A}^H \} = \mathbf{A} \mathbf{C}$$

## Appendix 5: 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:  $v(k) = v_\alpha(k) + jv_\beta(k)$ :

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

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

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

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

$$B(k) = 0$$

## Appendix 6: CR calculus and learning alg. (more later)

### The derivative of the cost function $\frac{1}{2}e(k)e^*(k)$ and CLMS

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

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

and the gradient

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

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

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

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

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

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

**Thus, no tedious computations  $\leadsto$  the CLMS is derived in one line.**

## App 7: Stochastic gradient optimis. $\leadsto$ complex gradient

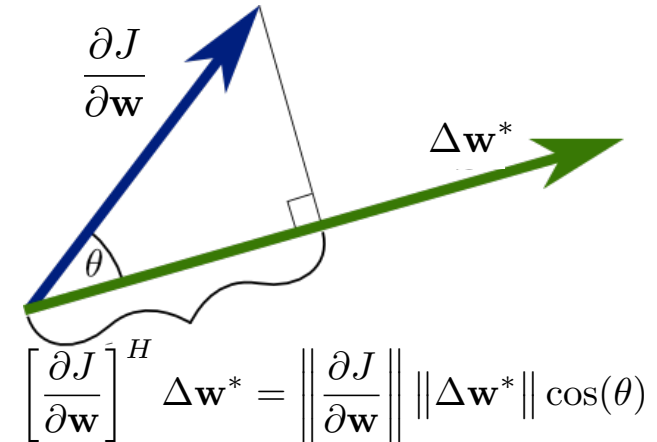
**Cost function**  $J(e, e^*) = |e|^2 = ee^*$ , where  $e(k) = d(k) - \mathbf{w}^H(k)\mathbf{x}(k)$

---

Gradient:  $\nabla_{\mathbf{w}} J = \frac{\partial J}{\partial \mathbf{w}} = \left[ \frac{\partial J}{\partial w_1}, \dots, \frac{\partial J}{\partial w_N} \right]^T$

**For the minima :**  $\frac{\partial J}{\partial \mathbf{w}} = \mathbf{0}$  and  $\frac{\partial J}{\partial \mathbf{w}^*} = \mathbf{0}$

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


$$\left[ \frac{\partial J}{\partial \mathbf{w}} \right]^H \Delta \mathbf{w}^* = \left\| \frac{\partial J}{\partial \mathbf{w}} \right\| \left\| \Delta \mathbf{w}^* \right\| \cos(\theta)$$

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

Therefore, the scalar product

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

achieves its maximum value when  $\frac{\partial J}{\partial \mathbf{w}} \parallel \Delta \mathbf{w}^*$ , that is, for  $\nabla_{\mathbf{w}} J = \nabla_{\mathbf{w}^*} J$ .

**The maximum change of the gradient of the cost function is in the direction of the conjugate weight vector ( $R^*$ -derivative)  $\leadsto$  equivalent to pseudogradient .**

## Appendix 8: Performance advantage of widely linear modelling in $\mathbb{C}$

The MSE of the widely linear and the strictly linear estimator of a variable  $y$  are respectively given by

$$E[|e_{\text{WL}}|^2] = E[|y - y_{\text{WL}}|^2] = E[|y|^2] - (\mathbf{h}^H \mathbf{c} + \mathbf{g}^H \mathbf{p}^*)$$

$$E[|e_{\text{SL}}|^2] = E[|y - y_{\text{SL}}|^2] = E[|y|^2] - \mathbf{c}^H \mathcal{C}^{-1} \mathbf{c}$$

The performance advantage of using the widely linear model in  $\mathbb{C}$  is then

$$\Delta \text{MSE} = E[|e_{\text{SL}}|^2] - E[|e_{\text{WL}}|^2] = \underbrace{[\mathbf{p} - \mathcal{P} \mathcal{C}^{-*} \mathbf{c}^*]^H [\mathcal{C} - \mathcal{P} \mathcal{C}^{-*} \mathcal{P}^*]^{-1} [\mathbf{p} - \mathcal{P} \mathcal{C}^{-*} \mathbf{c}^*]}$$

Is this term always nonnegative?

A joint diagonalisation of  $\mathcal{C}$  and  $\mathcal{P}$  can be achieved by using the strong uncorrelating transform to give  $\mathcal{C} = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^H$  and  $\mathcal{P} = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^T$ , where  $\mathbf{\Lambda} = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_N\}$  and  $1 \geq \lambda_1 \geq \dots \geq \lambda_N \geq 0$ . Upon such joint diagonalisation, we have

$$\Delta \text{MSE} = \underbrace{[\mathbf{Q}^{-1} \mathbf{p} - \mathbf{\Lambda} \mathbf{Q}^{-*} \mathbf{c}^*]^H}_{\mathbf{b}^H} \frac{\mathbf{I}}{\mathbf{I} - \mathbf{\Lambda}^2} \underbrace{[\mathbf{Q}^{-1} \mathbf{p} - \mathbf{\Lambda} \mathbf{Q}^{-*} \mathbf{c}^*]}_{\mathbf{b}} = \sum_{n=1}^N \frac{|b_n|^2}{1 - \lambda_n^2} \geq 0$$

## Appendix 9: Data model $\leadsto$ 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

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

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

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

In the vector case ( $N$  Gaussian random variables)

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

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

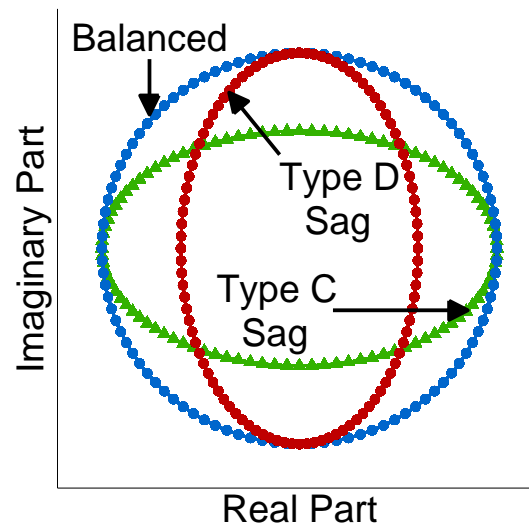
# App. 10: Does degree of circularity influence estimation in $\mathbb{C}$ ?

## Voltage sag: A magnitude and/or phase imbalance

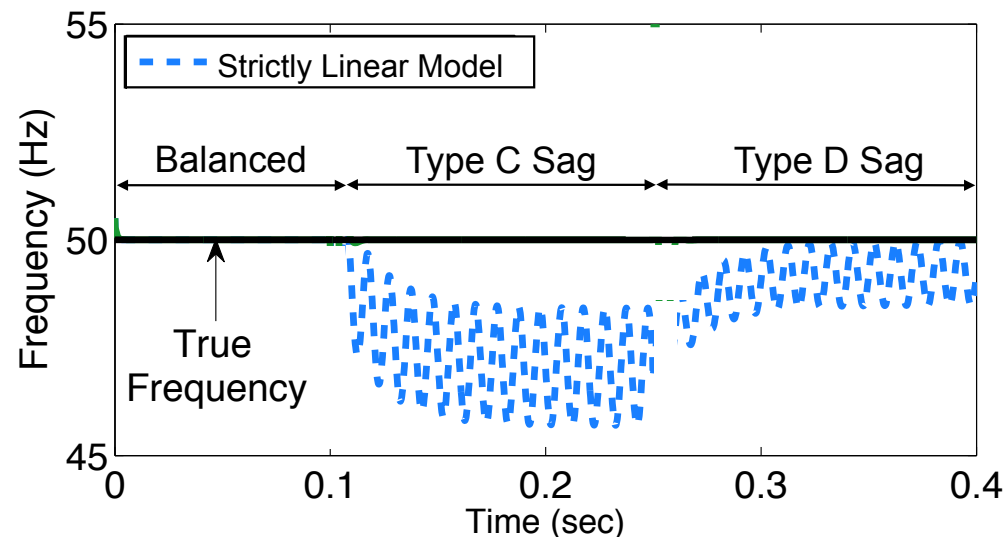
- For balanced systems,  $v(k) = A(k)e^{j\omega k\Delta T} \rightarrow$  circular trajectory.
- Unbalanced systems,  $v(k) = A(k)e^{j\omega k\Delta T} + B(k)e^{-j\omega k\Delta T}$  are influenced by the “conjugate” component.

→ **We need the complex conjugate of the signal too.**

Circularity Diagram



The strictly linear model,  $\hat{v} = f(v)$ , yields biased estimates when the system is unbalanced





# Notes:

---

○

# Notes:

---

○

# Notes:

---

○

# Notes:

---

○