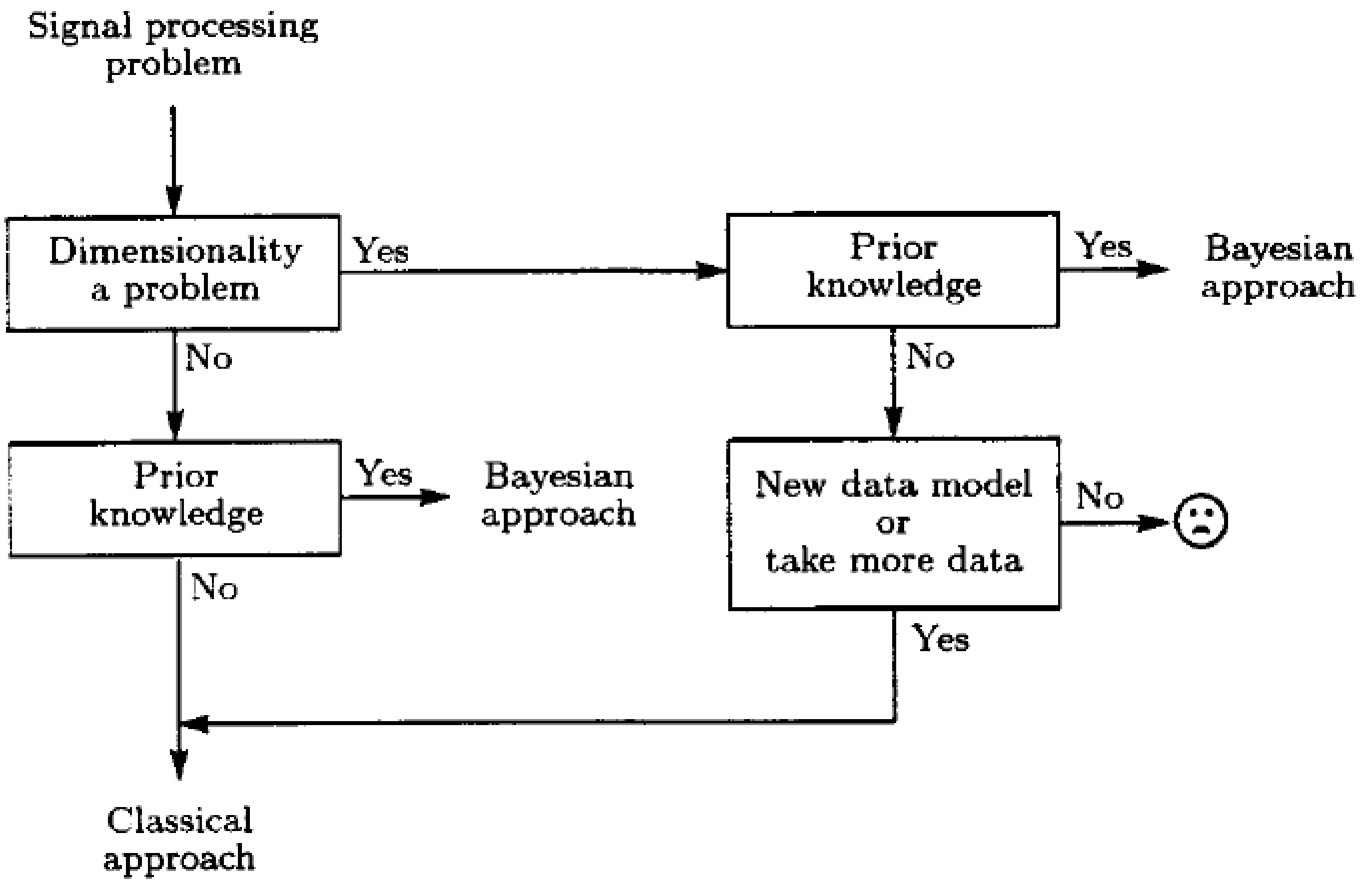
Estimation Theory

Fredrik Rusek

Extra lecture

14.4: Choosing an estimator

$$x[n] = A[n] + w[n] \qquad n = 0, 1, \ldots, N-1$$

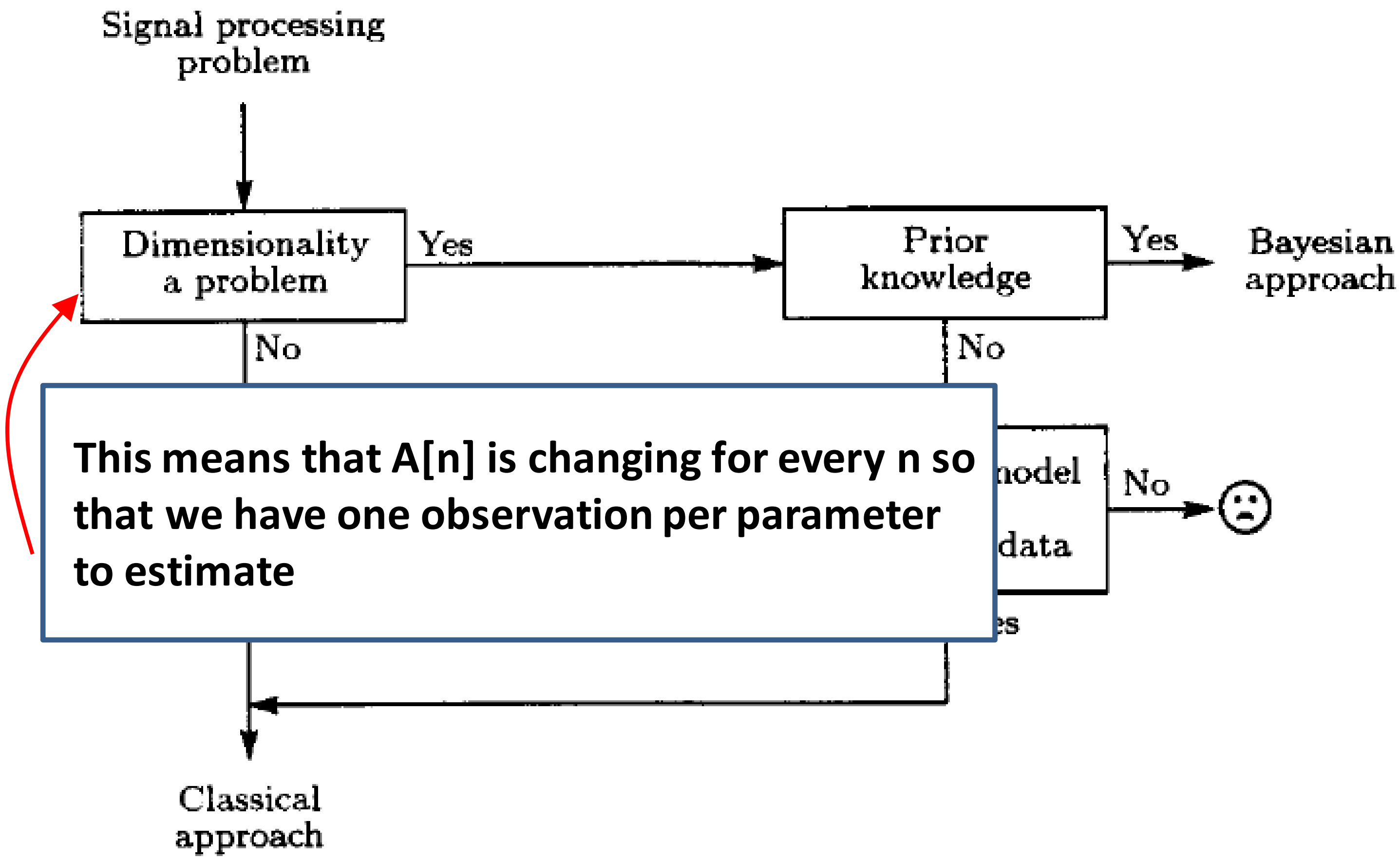


(a) Classical versus Bayesian

Figure 14.1 Decision-making process in estimator selection

14.4: Choosing an estimator

$$x[n] = A[n] + w[n] \qquad n = 0, 1, \ldots, N-1$$



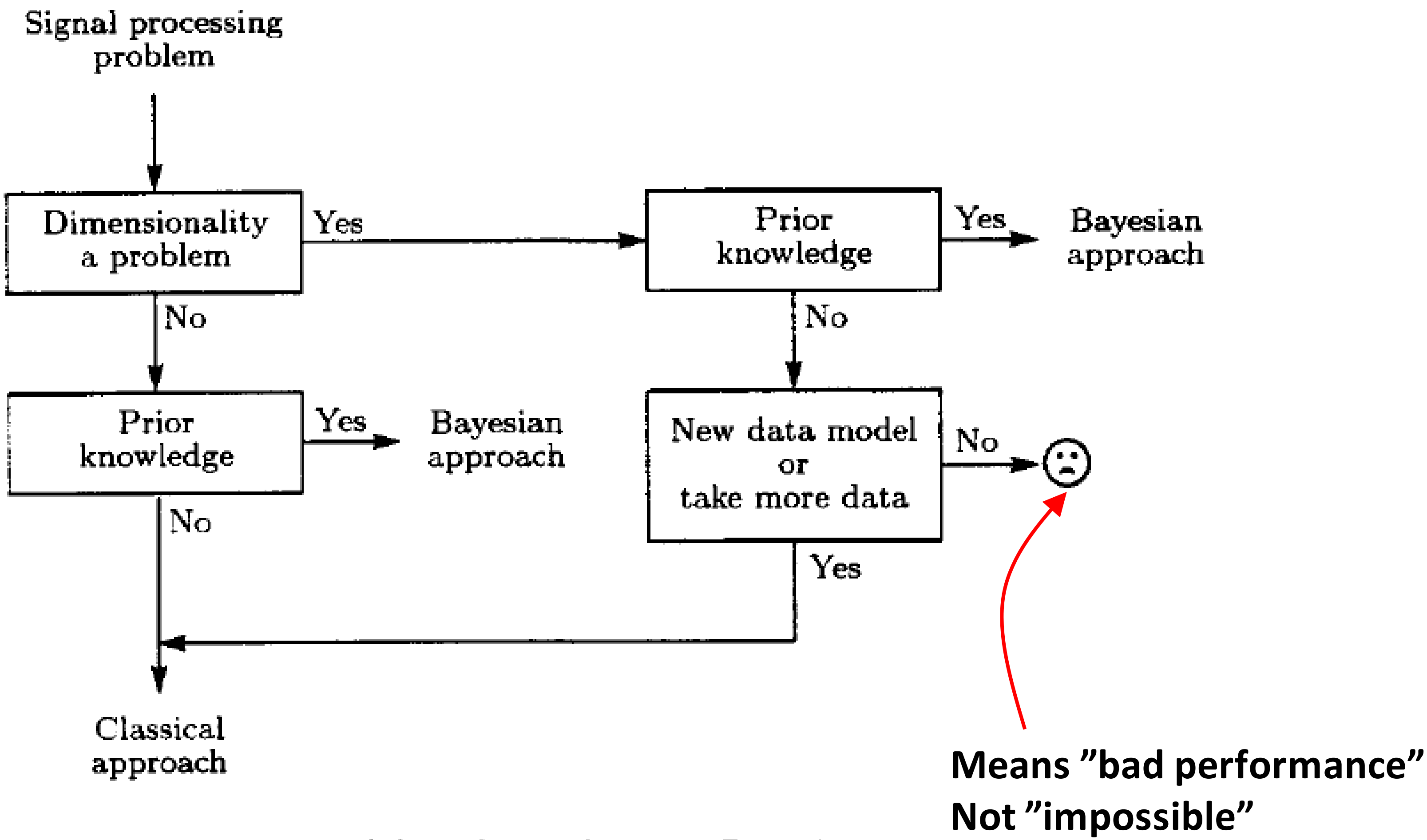
This means that A[n] is changing for every n so that we have one observation per parameter to estimate

(a) Classical versus Bayesian

Figure 14.1 Decision-making process in estimator selection

14.4: Choosing an estimator

$$x[n] = A[n] + w[n] \qquad n = 0, 1, \ldots, N-1$$

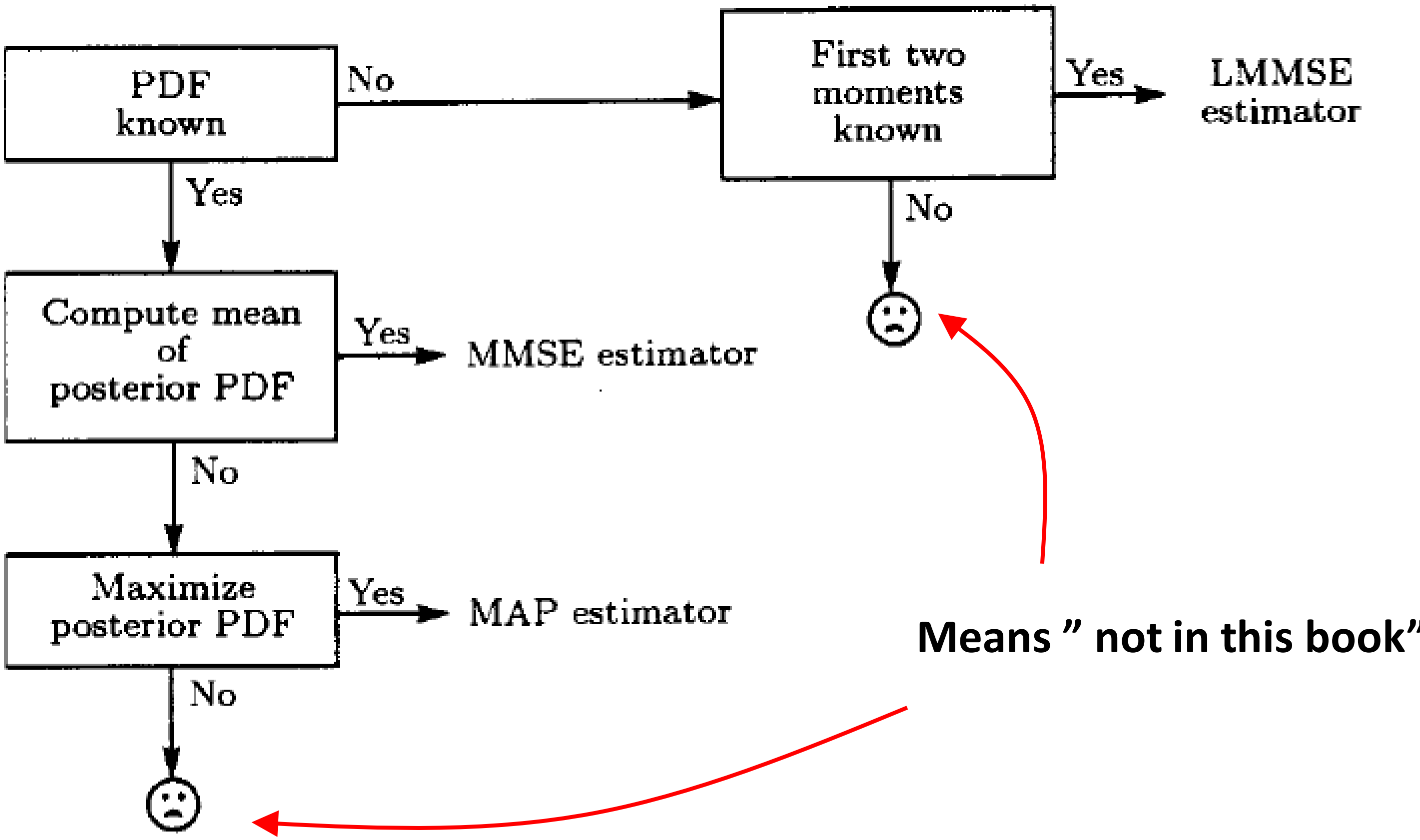


(a) Classical versus Bayesian

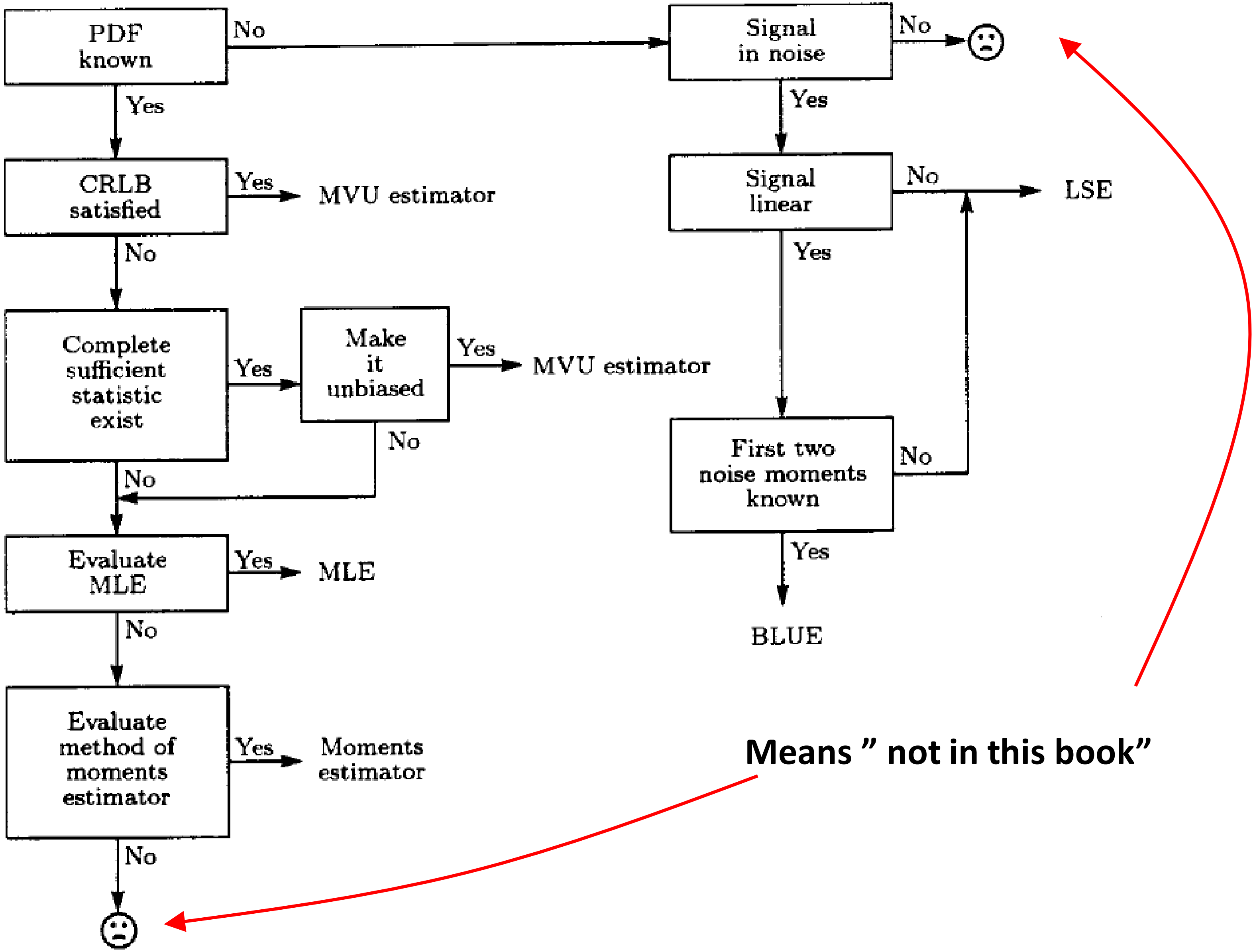
Figure 14.1 Decision-making process in estimator selection

14.4: Choosing an estimator

$$x[n] = A[n] + w[n] \qquad n = 0, 1, \ldots, N-1$$



(b) Bayesian approach



(c) Classical approach

Complex-valued models

Consider the following complex-valued estimation problem

$$J(\tilde{A}) = \sum_{n=0}^{N-1} |\tilde{x}[n] - \tilde{A}\tilde{s}[n]|^2$$

Where all quantities are complex-valued (~ means complex in Kay)

In the real-valued case we can take the differential with respect to A, set to 0, and solve

Complex-valued models

Consider the following complex-valued estimation problem

$$J(\tilde{A}) = \sum_{n=0}^{N-1} |\tilde{x}[n] - \tilde{A}\tilde{s}[n]|^2$$

Where all quantities are complex-valued (~ means complex in Kay)

In the real-valued case we can take the differential with respect to A, set to 0, and solve

To take a differential, we must go back to the definition of a differential

$$\boxed{f'(z_0) = \lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}}$$

Note: z is assumed to be complex valued

Complex-valued models

$$\boxed{f'(z_0) = \lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}}$$

For the definition to be meaningful, the limit must exist, and not depend on the direction that one approahces z_0

Complex-valued models

$$\boxed{f'(z_0) = \lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}}$$

For the definition to be meaningful, the limit must exist, and not depend on the direction that one approahces z_0

When this is true, the function f(z) is said to be **analytic/holomorphic**

Complex-valued models

$$\boxed{f'(z_0) = \lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}}$$

For the definition to be meaningful, the limit must exist, and not depend on the direction that one approahces z_0

When this is true, the function f(z) is said to be **analytic/holomorphic**

Or, in other words: An analytical function allows one to take differentials

Complex-valued models

$$\boxed{f'(z_0) = \lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}}$$

For the definition to be meaningful, the limit must exist, and not depend on the direction that one approahces z_0

When this is true, the function f(z) is said to be **analytic/holomorphic**

A function is analytical when the Cauchy-Riemann equations are satisfied

f(z) = f(x+iy) = u(x,y) + i v(x,y)

$$\boxed{\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \qquad \text{and} \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}}$$

Complex-valued models

$$J(\tilde{A}) = \sum_{n=0}^{N-1} |\tilde{x}[n] - \tilde{A}\tilde{s}[n]|^2$$

Let us now return to the complex-valued estimation problem

Is it analytical so that we can take differentials??

f(z) = f(x+iy) = u(x,y) + i v(x,y)

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

Complex-valued models

$$\boxed{J(\tilde{A}) = \sum_{n=0}^{N-1} |\tilde{x}[n] - \tilde{A}\tilde{s}[n]|^2}$$

Let us now return to the complex-valued estimation problem

Is it analytical so that we can take differentials??

No. Since v(x,y) = 0, we have that dv/dx=dv/dy=0

The only real-valued functions that are analytical are constant functions

= 0

f(z) = f(x+iy) = u(x,y) + i v(x,y)

$$\boxed{\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \quad \text{and} \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}}$$

Complex-valued models

$$\boxed{J(\tilde{A}) = \sum_{n=0}^{N-1} |\tilde{x}[n] - \tilde{A}\tilde{s}[n]|^2}$$

All estimation (in general: optimization) problems have real-valued cost functions

We must abandon the normal definition of differential

$$\boxed{\cancel{f'(z_0) = \lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}}}$$

More fundamentally, we cannot regard z as being one number. We must consider it to be two numbers: its real-part and its complex-part.

Complex-valued models

We will do two things next:

1. **Deal with the complex case as a multivariate optimization problem.**

2. **Show that with clever book-keeping, this multivariate treatment can be written in a form that resembles normal real-valued algebra.**

Recall: **In no cases will the differential be found through the formula**

$$f'(z_0) = \lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}$$

Complex-valued models

Expressing $\tilde{A} = A_R + iA_I$ etc, we have

$$J(\tilde{A}) = \sum_{n=0}^{N-1} |\tilde{x}[n] - \tilde{A}\tilde{s}[n]|^2$$

$$\longrightarrow$$

$$J'(A_R, A_I) = \sum_{n=0}^{N-1} |x_R[n] + jx_I[n] - (A_R + jA_I)(s_R[n] + js_I[n])|^2$$

Complex-valued models

Expressing $\tilde{A} = A_R + iA_I$ etc, we have

$$J(\tilde{A}) = \sum_{n=0}^{N-1} |\tilde{x}[n] - \tilde{A}\tilde{s}[n]|^2$$

$$\begin{aligned} J'(A_R, A_I) &= \sum_{n=0}^{N-1} |x_R[n] + jx_I[n] - (A_R + jA_I)(s_R[n] + js_I[n])|^2 \\ &= \sum_{n=0}^{N-1} (x_R[n] - A_R s_R[n] + A_I s_I[n])^2 + (x_I[n] - A_R s_I[n] - A_I s_R[n])^2 \end{aligned}$$

$|z_R + iz_I|^2 = |z_R|^2 + |z_I|^2$

Complex-valued models

Expressing $\tilde{A} = A_R + iA_I$ etc, we have

$$J(\tilde{A}) = \sum_{n=0}^{N-1} |\tilde{x}[n] - \tilde{A}\tilde{s}[n]|^2$$

$$\longrightarrow$$

$$\begin{aligned} J'(A_R, A_I) &= \sum_{n=0}^{N-1} |x_R[n] + jx_I[n] - (A_R + jA_I)(s_R[n] + js_I[n])|^2 \\ &= \sum_{n=0}^{N-1} (x_R[n] - A_R s_R[n] + A_I s_I[n])^2 + (x_I[n] - A_R s_I[n] - A_I s_R[n])^2 \end{aligned}$$

This is a quadratic form in A_R and A_I Define $\mathbf{A} = [A_R\ A_I]^T$

Some algebra will yield something of the form

$$J'(A_R, A_I) = \mathbf{b}^T\mathbf{A} + \mathbf{A}^T\mathbf{C}\mathbf{A}$$

Where **b** is a 2x1 vector and **C** is 2x2

Complex-valued models

We can now take normal differentials with respect to the real vector **A**

$$J'(A_R, A_I) = \mathbf{b}^T\mathbf{A} + \mathbf{A}^T\mathbf{C}\mathbf{A}$$

Complex-valued models

We can now take normal differentials with respect to the real vector **A**

We have seen this repeatedly in the course, and the final result is

$$\hat{\mathbf{A}} = \begin{bmatrix} \dfrac{\mathbf{s}_R^T\mathbf{x}_R + \mathbf{s}_I^T\mathbf{x}_I}{\mathbf{s}_R^T\mathbf{s}_R + \mathbf{s}_I^T\mathbf{s}_I} \\ \dfrac{\mathbf{s}_R^T\mathbf{x}_I - \mathbf{s}_I^T\mathbf{x}_R}{\mathbf{s}_R^T\mathbf{s}_R + \mathbf{s}_I^T\mathbf{s}_I} \end{bmatrix}$$

x is here a vector containing all x[n] (etc)

$$J'(A_R, A_I) = \mathbf{b}^T\mathbf{A} + \mathbf{A}^T\mathbf{C}\mathbf{A}$$

Complex-valued models

We can now take normal differentials with respect to the real vector **A**

We have seen this repeatedly in the course, and the final result is

$$\hat{\mathbf{A}} = \begin{bmatrix} \dfrac{\mathbf{s}_R^T\mathbf{x}_R + \mathbf{s}_I^T\mathbf{x}_I}{\mathbf{s}_R^T\mathbf{s}_R + \mathbf{s}_I^T\mathbf{s}_I} \\ \dfrac{\mathbf{s}_R^T\mathbf{x}_I - \mathbf{s}_I^T\mathbf{x}_R}{\mathbf{s}_R^T\mathbf{s}_R + \mathbf{s}_I^T\mathbf{s}_I} \end{bmatrix}$$

Going back to complex-valued notation, we get

$$\begin{aligned} \hat{\tilde{A}} &= \frac{\mathbf{s}_R^T\mathbf{x}_R + \mathbf{s}_I^T\mathbf{x}_I + j\mathbf{s}_R^T\mathbf{x}_I - j\mathbf{s}_I^T\mathbf{x}_R}{\mathbf{s}_R^T\mathbf{s}_R + \mathbf{s}_I^T\mathbf{s}_I} \\ &= \frac{(\mathbf{x}_R + j\mathbf{x}_I)^T(\mathbf{s}_R - j\mathbf{s}_I)}{\mathbf{s}_R^T\mathbf{s}_R + \mathbf{s}_I^T\mathbf{s}_I} \\ &= \frac{\sum_{n=0}^{N-1} \tilde{x}[n]\tilde{s}^*[n]}{\sum_{n=0}^{N-1} |\tilde{s}[n]|^2} \end{aligned}$$

Complex-valued models

We can now take normal differentials with respect to the real vector **A**

We have seen this repeatedly in the course, and the final result is

$$\hat{\mathbf{A}} = \begin{bmatrix} \dfrac{\mathbf{s}_R^T\mathbf{x}_R + \mathbf{s}_I^T\mathbf{x}_I}{\mathbf{s}_R^T\mathbf{s}_R + \mathbf{s}_I^T\mathbf{s}_I} \\ \dfrac{\mathbf{s}_R^T\mathbf{x}_I - \mathbf{s}_I^T\mathbf{x}_R}{\mathbf{s}_R^T\mathbf{s}_R + \mathbf{s}_I^T\mathbf{s}_I} \end{bmatrix}$$

Going back to complex-valued notation, we get

$$\begin{aligned} \hat{\tilde{A}} &= \frac{\mathbf{s}_R^T\mathbf{x}_R + \mathbf{s}_I^T\mathbf{x}_I + j\mathbf{s}_R^T\mathbf{x}_I - j\mathbf{s}_I^T\mathbf{x}_R}{\mathbf{s}_R^T\mathbf{s}_R + \mathbf{s}_I^T\mathbf{s}_I} \\ &= \frac{(\mathbf{x}_R + j\mathbf{x}_I)^T(\mathbf{s}_R - j\mathbf{s}_I)}{\mathbf{s}_R^T\mathbf{s}_R + \mathbf{s}_I^T\mathbf{s}_I} \\ &= \frac{\sum_{n=0}^{N-1}\tilde{x}[n]\tilde{s}^*[n]}{\sum_{n=0}^{N-1}|\tilde{s}[n]|^2} \end{aligned}$$

This approach is valid, but in-efficient (a notational nightmare)

Complex-valued models

Let us now gow back a step $$J(\tilde{A}) = \sum_{n=0}^{N-1} |\tilde{x}[n] - \tilde{A}\tilde{s}[n]|^2$$

We need to compute $\frac{\partial J}{\partial A_R}$ and $\frac{\partial J}{\partial A_I}$

Then, we should set them both to 0

Complex-valued models

Let us now gow back a step $$J(\tilde{A}) = \sum_{n=0}^{N-1} |\tilde{x}[n] - \tilde{A}\tilde{s}[n]|^2$$

We need to compute $\dfrac{\partial J}{\partial A_R}$ and $\dfrac{\partial J}{\partial A_I}$

Then, we should set them both to 0

To keep track of both of them, we can form the quantity

$$\frac{\partial J}{\partial \tilde{A}} = \frac{\partial J}{\partial A_R} + j\frac{\partial J}{\partial A_I}$$

Complex-valued models

Let us now gow back a step $$J(\tilde{A}) = \sum_{n=0}^{N-1} |\tilde{x}[n] - \tilde{A}\tilde{s}[n]|^2$$

We need to compute $\frac{\partial J}{\partial A_R}$ and $\frac{\partial J}{\partial A_I}$

Then, we should set them both to 0

To keep track of both of them, we can form the quantity

$$\cancel{\frac{\partial J}{\partial \tilde{A}} = \frac{\partial J}{\partial A_R} + j\frac{\partial J}{\partial A_I}}$$

However, to obtain some nice formulas later, Kay (but not all other authors) choose to keep track of the differentials as

$$\boxed{\frac{\partial J}{\partial \tilde{A}} = \frac{1}{2}\left(\frac{\partial J}{\partial A_R} - j\frac{\partial J}{\partial A_I}\right)}$$

Complex-valued models

Note that we have

$$\boxed{\frac{\partial J}{\partial \tilde{A}} = 0 \quad \text{if and only if} \quad \frac{\partial J}{\partial A_R} = \mathbf{0} \quad \text{and} \quad \frac{\partial J}{\partial A_I} = \mathbf{0}}$$

With the new book-keeping, our definition of differential ”behaves as normal”

$$\boxed{\frac{\partial J}{\partial \tilde{A}} = \frac{1}{2}\left(\frac{\partial J}{\partial A_R} - j\frac{\partial J}{\partial A_I}\right)}$$

Complex-valued models

Note that we have

$$\boxed{\frac{\partial J}{\partial \tilde{A}} = 0 \quad \text{if and only if} \quad \frac{\partial J}{\partial A_R} = \mathbf{0} \quad \text{and} \quad \frac{\partial J}{\partial A_I} = \mathbf{0}}$$

With the new book-keeping, our definition of differential ”behaves as normal”

RECALL AGAIN $\frac{\partial J}{\partial \tilde{A}} = \frac{1}{2}\left(\frac{\partial J}{\partial A_R} - j\frac{\partial J}{\partial A_I}\right)$ **IS NOT** $f'(z_0) = \lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}$

$$\boxed{\frac{\partial J}{\partial \tilde{A}} = \frac{1}{2}\left(\frac{\partial J}{\partial A_R} - j\frac{\partial J}{\partial A_I}\right)}$$

Complex-valued models

To efficiently use $\frac{\partial J}{\partial \tilde{A}} = \frac{1}{2}\left(\frac{\partial J}{\partial A_R} - j\frac{\partial J}{\partial A_I}\right)$ **we need some properties of it**

Let us consider first $\frac{\partial \theta}{\partial \theta} =$

Complex-valued models

To efficiently use $$\frac{\partial J}{\partial \tilde{A}} = \frac{1}{2}\left(\frac{\partial J}{\partial A_R} - j\frac{\partial J}{\partial A_I}\right)$$ **we need some properties of it**

Let us consider first $$\frac{\partial \theta}{\partial \theta} = \frac{1}{2}\left(\frac{\partial}{\partial \alpha} - j\frac{\partial}{\partial \beta}\right)(\alpha + j\beta)$$

Complex-valued models

To efficiently use $\frac{\partial J}{\partial \tilde{A}} = \frac{1}{2}\left(\frac{\partial J}{\partial A_R} - j\frac{\partial J}{\partial A_I}\right)$ **we need some properties of it**

Let us consider first

$$\begin{aligned} \frac{\partial \theta}{\partial \theta} &= \frac{1}{2}\left(\frac{\partial}{\partial \alpha} - j\frac{\partial}{\partial \beta}\right)(\alpha + j\beta) \\ &= \frac{1}{2}\left(\frac{\partial \alpha}{\partial \alpha} + j\frac{\partial \beta}{\partial \alpha} - j\frac{\partial \alpha}{\partial \beta} + \frac{\partial \beta}{\partial \beta}\right) \end{aligned}$$

Complex-valued models

To efficiently use $\frac{\partial J}{\partial \tilde{A}} = \frac{1}{2}\left(\frac{\partial J}{\partial A_R} - j\frac{\partial J}{\partial A_I}\right)$ **we need some properties of it**

Let us consider first

$$\begin{aligned} \frac{\partial \theta}{\partial \theta} &= \frac{1}{2}\left(\frac{\partial}{\partial \alpha} - j\frac{\partial}{\partial \beta}\right)(\alpha + j\beta) \\ &= \frac{1}{2}\left(\frac{\partial \alpha}{\partial \alpha} + j\frac{\partial \beta}{\partial \alpha} - j\frac{\partial \alpha}{\partial \beta} + \frac{\partial \beta}{\partial \beta}\right) \\ &= \frac{1}{2}(1 + j0 - j0 + 1) \end{aligned}$$

Complex-valued models

To efficiently use $\frac{\partial J}{\partial \tilde{A}} = \frac{1}{2}\left(\frac{\partial J}{\partial A_R} - j\frac{\partial J}{\partial A_I}\right)$ **we need some properties of it**

Let us consider first

$$\begin{aligned}
\frac{\partial \theta}{\partial \theta} &= \frac{1}{2}\left(\frac{\partial}{\partial \alpha} - j\frac{\partial}{\partial \beta}\right)(\alpha + j\beta) \\
&= \frac{1}{2}\left(\frac{\partial \alpha}{\partial \alpha} + j\frac{\partial \beta}{\partial \alpha} - j\frac{\partial \alpha}{\partial \beta} + \frac{\partial \beta}{\partial \beta}\right) \\
&= \frac{1}{2}(1 + j0 - j0 + 1) \\
&= 1
\end{aligned}$$

Complex-valued models

To efficiently use $\frac{\partial J}{\partial \tilde{A}} = \frac{1}{2}\left(\frac{\partial J}{\partial A_R} - j\frac{\partial J}{\partial A_I}\right)$ **we need some properties of it**

Let us now consider

$$\begin{aligned} \frac{\partial \theta^*}{\partial \theta} &= \frac{1}{2}\left(\frac{\partial}{\partial \alpha} - j\frac{\partial}{\partial \beta}\right)(\alpha + j\beta) \\ &= \frac{1}{2}\left(\frac{\partial \alpha}{\partial \alpha} + j\frac{\partial \beta}{\partial \alpha} - j\frac{\partial \alpha}{\partial \beta} + \frac{\partial \beta}{\partial \beta}\right) \\ &= \frac{1}{2}(1 + j0 - j0 + 1) \\ &= 1 \end{aligned}$$

$$\boxed{\frac{\partial \theta}{\partial \theta} = 1}$$

Complex-valued models

To efficiently use $\frac{\partial J}{\partial \tilde{A}} = \frac{1}{2}\left(\frac{\partial J}{\partial A_R} - j\frac{\partial J}{\partial A_I}\right)$ **we need some properties of it**

Let us now consider
$$\begin{aligned}\frac{\partial \theta^*}{\partial \theta} &= \frac{1}{2}\left(\frac{\partial}{\partial \alpha} - j\frac{\partial}{\partial \beta}\right)(\alpha + j\beta) \\ &= \frac{1}{2}\left(\frac{\partial \alpha}{\partial \alpha} + j\frac{\partial \beta}{\partial \alpha} - j\frac{\partial \alpha}{\partial \beta} + \frac{\partial \beta}{\partial \beta}\right) \\ &= \frac{1}{2}(1 + j0 - j0 + 1) \\ &= 1\end{aligned}$$

Change signs here

$$\boxed{\frac{\partial \theta}{\partial \theta} = 1}$$

Complex-valued models

To efficiently use $\frac{\partial J}{\partial \tilde{A}} = \frac{1}{2}\left(\frac{\partial J}{\partial A_R} - j\frac{\partial J}{\partial A_I}\right)$ **we need some properties of it**

Let us now consider
$$\begin{aligned}\frac{\partial \theta^*}{\partial \theta} &= \frac{1}{2}\left(\frac{\partial}{\partial \alpha} - j\frac{\partial}{\partial \beta}\right)(\alpha - j\beta)\\ &= \frac{1}{2}\left(\frac{\partial \alpha}{\partial \alpha} - j\frac{\partial \beta}{\partial \alpha} - j\frac{\partial \alpha}{\partial \beta} - \frac{\partial \beta}{\partial \beta}\right)\\ &= \frac{1}{2}(1 - j0 - j0 - 1)\\ &= 0\end{aligned}$$

$$\boxed{\frac{\partial \theta}{\partial \theta} = 1} \qquad \boxed{\frac{\partial \theta^*}{\partial \theta} = 0}$$

Complex-valued models

To efficiently use $\frac{\partial J}{\partial \tilde{A}} = \frac{1}{2}\left(\frac{\partial J}{\partial A_R} - j\frac{\partial J}{\partial A_I}\right)$ **we need some properties of it**

From the definition, the chain rule continues to hold

$$\frac{\partial}{\partial \theta} X(\theta)Y(\theta) = Y(\theta)\frac{\partial}{\partial \theta} X(\theta) + X(\theta)\frac{\partial}{\partial \theta} Y(\theta)$$

$$\boxed{\frac{\partial \theta}{\partial \theta} = 1} \qquad \boxed{\frac{\partial \theta^*}{\partial \theta} = 0}$$

Complex-valued models

To efficiently use $\frac{\partial J}{\partial \tilde{A}} = \frac{1}{2}\left(\frac{\partial J}{\partial A_R} - j\frac{\partial J}{\partial A_I}\right)$ **we need some properties of it**

From the definition, the chain rule continues to hold

$$\frac{\partial}{\partial \theta} X(\theta)Y(\theta) = Y(\theta)\frac{\partial}{\partial \theta} X(\theta) + X(\theta)\frac{\partial}{\partial \theta} Y(\theta)$$

In particular, this implies that

$$\frac{\partial}{\partial \theta}\theta\theta^* = \frac{\partial \theta}{\partial \theta}\theta^* + \theta\frac{\partial \theta^*}{\partial \theta} = 1 \cdot \theta^* + \theta \cdot 0 = \theta^*$$

$$\boxed{\frac{\partial \theta}{\partial \theta} = 1} \qquad \boxed{\frac{\partial \theta^*}{\partial \theta} = 0} \qquad \boxed{\frac{\partial}{\partial \theta} X(\theta)Y(\theta) = Y(\theta)\frac{\partial}{\partial \theta} X(\theta) + X(\theta)\frac{\partial}{\partial \theta} Y(\theta)}$$

Complex-valued models

To efficiently use $\frac{\partial J}{\partial \tilde{A}} = \frac{1}{2}\left(\frac{\partial J}{\partial A_R} - j\frac{\partial J}{\partial A_I}\right)$ **we need some properties of it**

$$\boxed{\frac{\partial}{\partial \theta}\theta\theta^* = \theta^*}$$

$$\boxed{\frac{\partial \theta}{\partial \theta} = 1} \quad \boxed{\frac{\partial \theta^*}{\partial \theta} = 0} \quad \boxed{\frac{\partial}{\partial \theta} X(\theta)Y(\theta) = Y(\theta)\frac{\partial}{\partial \theta}X(\theta) + X(\theta)\frac{\partial}{\partial \theta}Y(\theta)}$$

Complex-valued models

Let us now return to the "notational nightmare" and solve it with $\frac{\partial J}{\partial \tilde{A}} = \frac{1}{2}\left(\frac{\partial J}{\partial A_R} - j\frac{\partial J}{\partial A_I}\right)$

$$J(\tilde{A}) = \sum_{n=0}^{N-1} |\tilde{x}[n] - \tilde{A}\tilde{s}[n]|^2$$

$$\boxed{\frac{\partial \theta}{\partial \theta} = 1} \qquad \boxed{\frac{\partial \theta^*}{\partial \theta} = 0} \qquad \boxed{\frac{\partial}{\partial \theta}\theta\theta^* = \theta^*}$$

Complex-valued models

Let us now return to the "notational nightmare" and solve it with $\frac{\partial J}{\partial \tilde{A}} = \frac{1}{2}\left(\frac{\partial J}{\partial A_R} - j\frac{\partial J}{\partial A_I}\right)$

$$J(\tilde{A}) = \sum_{n=0}^{N-1} |\tilde{x}[n] - \tilde{A}\tilde{s}[n]|^2$$

$$\frac{\partial J}{\partial \tilde{A}} = \frac{\partial}{\partial \tilde{A}} \sum_{n=0}^{N-1} \left|\tilde{x}[n] - \tilde{A}\tilde{s}[n]\right|^2$$

$$\boxed{\frac{\partial \theta}{\partial \theta} = 1} \qquad \boxed{\frac{\partial \theta^*}{\partial \theta} = 0} \qquad \boxed{\frac{\partial}{\partial \theta} \theta\theta^* = \theta^*}$$

Complex-valued models

Let us now return to the "notational nightmare" and solve it with $\frac{\partial J}{\partial \tilde{A}} = \frac{1}{2}\left(\frac{\partial J}{\partial A_R} - j\frac{\partial J}{\partial A_I}\right)$

$$J(\tilde{A}) = \sum_{n=0}^{N-1} |\tilde{x}[n] - \tilde{A}\tilde{s}[n]|^2$$

$$\begin{aligned}
\frac{\partial J}{\partial \tilde{A}} &= \frac{\partial}{\partial \tilde{A}} \sum_{n=0}^{N-1} \left|\tilde{x}[n] - \tilde{A}\tilde{s}[n]\right|^2 \\
&= \sum_{n=0}^{N-1} \frac{\partial}{\partial \tilde{A}} \left(|\tilde{x}[n]|^2 - \tilde{x}[n]\tilde{A}^*\tilde{s}^*[n] - \tilde{A}\tilde{s}[n]\tilde{x}^*[n] + \tilde{A}\tilde{A}^*|\tilde{s}[n]|^2\right)
\end{aligned}$$

$$\boxed{\frac{\partial \theta}{\partial \theta} = 1} \quad \boxed{\frac{\partial \theta^*}{\partial \theta} = 0} \quad \boxed{\frac{\partial}{\partial \theta}\theta\theta^* = \theta^*}$$

Complex-valued models

Let us now return to the ”notational nightmare” and solve it with $\frac{\partial J}{\partial \tilde{A}} = \frac{1}{2}\left(\frac{\partial J}{\partial A_R} - j\frac{\partial J}{\partial A_I}\right)$

$$J(\tilde{A}) = \sum_{n=0}^{N-1} |\tilde{x}[n] - \tilde{A}\tilde{s}[n]|^2$$

$$\begin{aligned}\frac{\partial J}{\partial \tilde{A}} &= \frac{\partial}{\partial \tilde{A}} \sum_{n=0}^{N-1} \left|\tilde{x}[n] - \tilde{A}\tilde{s}[n]\right|^2 \\ &= \sum_{n=0}^{N-1} \frac{\partial}{\partial \tilde{A}} \left(|\tilde{x}[n]|^2 - \tilde{x}[n]\tilde{A}^*\tilde{s}^*[n] - \tilde{A}\tilde{s}[n]\tilde{x}^*[n] + \tilde{A}\tilde{A}^*|\tilde{s}[n]|^2\right)\end{aligned}$$

$$\boxed{\frac{\partial\theta}{\partial\theta} = 1} \qquad \boxed{\frac{\partial\theta^*}{\partial\theta} = 0} \qquad \boxed{\frac{\partial}{\partial\theta}\,\theta\theta^* = \theta^*}$$

Complex-valued models

Let us now return to the "notational nightmare" and solve it with $\frac{\partial J}{\partial \tilde{A}} = \frac{1}{2}\left(\frac{\partial J}{\partial A_R} - j\frac{\partial J}{\partial A_I}\right)$

$$J(\tilde{A}) = \sum_{n=0}^{N-1} |\tilde{x}[n] - \tilde{A}\tilde{s}[n]|^2$$

$$\begin{aligned}
\frac{\partial J}{\partial \tilde{A}} &= \frac{\partial}{\partial \tilde{A}} \sum_{n=0}^{N-1} \left|\tilde{x}[n] - \tilde{A}\tilde{s}[n]\right|^2 \\
&= \sum_{n=0}^{N-1} \frac{\partial}{\partial \tilde{A}} \left(|\tilde{x}[n]|^2 - \tilde{x}[n]\tilde{A}^*\tilde{s}^*[n] - \tilde{A}\tilde{s}[n]\tilde{x}^*[n] + \tilde{A}\tilde{A}^*|\tilde{s}[n]|^2\right) \\
&= \sum_{n=0}^{N-1} \left(0 - 0 - \tilde{s}[n]\tilde{x}^*[n] + \tilde{A}^*|\tilde{s}[n]|^2\right).
\end{aligned}$$

$$\boxed{\frac{\partial \theta}{\partial \theta} = 1} \qquad \boxed{\frac{\partial \theta^*}{\partial \theta} = 0} \qquad \boxed{\frac{\partial}{\partial \theta}\theta\theta^* = \theta^*}$$

Complex-valued models

Let us now return to the "notational nightmare" and solve it with $\frac{\partial J}{\partial \tilde{A}} = \frac{1}{2}\left(\frac{\partial J}{\partial A_R} - j\frac{\partial J}{\partial A_I}\right)$

$$J(\tilde{A}) = \sum_{n=0}^{N-1} |\tilde{x}[n] - \tilde{A}\tilde{s}[n]|^2$$

$$\begin{aligned}
\frac{\partial J}{\partial \tilde{A}} &= \frac{\partial}{\partial \tilde{A}} \sum_{n=0}^{N-1} \left|\tilde{x}[n] - \tilde{A}\tilde{s}[n]\right|^2 \\
&= \sum_{n=0}^{N-1} \frac{\partial}{\partial \tilde{A}} \left(|\tilde{x}[n]|^2 - \tilde{x}[n]\tilde{A}^*\tilde{s}^*[n] - \tilde{A}\tilde{s}[n]\tilde{x}^*[n] + \tilde{A}\tilde{A}^*|\tilde{s}[n]|^2\right) \\
&= \sum_{n=0}^{N-1} \left(0 - 0 - \tilde{s}[n]\tilde{x}^*[n] + \tilde{A}^*|\tilde{s}[n]|^2\right).
\end{aligned}$$

$$\boxed{\frac{\partial \theta}{\partial \theta} = 1}$$

$$\boxed{\frac{\partial \theta^*}{\partial \theta} = 0}$$

$$\boxed{\frac{\partial}{\partial \theta}\theta\theta^* = \theta^*}$$

$$\boxed{\hat{\tilde{A}} = \frac{\sum_{n=0}^{N-1} \tilde{x}[n]\tilde{s}^*[n]}{\sum_{n=0}^{N-1} |\tilde{s}[n]|^2}}$$

Complex-valued models

For vector valued parameters we can, in the same fashion, reach the formulas

$$\frac{\partial \mathbf{b}^H \boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = \mathbf{b}^* \qquad \frac{\partial \boldsymbol{\theta}^H \mathbf{b}}{\partial \boldsymbol{\theta}} = 0$$

$$\frac{\partial \boldsymbol{\theta}^H \mathbf{A} \boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = (\mathbf{A}\boldsymbol{\theta})^*, \quad \text{where } \mathbf{A}^H = \mathbf{A}$$

Complex-valued models

Complex Gaussian PDFs

Let $\tilde{x} = u + jv$ where u and v are independent Gaussians

$\mathcal{N}(\mu_u, \sigma^2/2)$ and $\mathcal{N}(\mu_v, \sigma^2/2)$

Complex-valued models

Complex Gaussian PDFs

Let $\tilde{x} = u + jv$ where u and v are independent Gaussians $\mathcal{N}(\mu_u, \sigma^2/2)$ and $\mathcal{N}(\mu_v, \sigma^2/2)$

Variance is defined as

$$\mathrm{var}(\tilde{x}) = E\left(|\tilde{x} - E(\tilde{x})|^2\right) = E(|\tilde{x}|^2) - |E(\tilde{x})|^2$$

Complex-valued models

Complex Gaussian PDFs

Let $\tilde{x} = u + jv$ where u and v are independent Gaussians

$$\mathcal{N}(\mu_u, \sigma^2/2) \text{ and } \mathcal{N}(\mu_v, \sigma^2/2)$$

Variance is defined as

$$\mathrm{var}(\tilde{x}) = E\left(|\tilde{x} - E(\tilde{x})|^2\right) = E(|\tilde{x}|^2) - |E(\tilde{x})|^2$$

$$= E(|x_R|^2) + E(|x_I|^2) - E(x_R)^2 - E(x_I)^2$$

$$= var(x_R) + var(x_I)$$

Complex-valued models

Complex Gaussian PDFs

Let $\tilde{x} = u + jv$ where u and v are independent Gaussians

$$\mathcal{N}(\mu_u, \sigma^2/2) \text{ and } \mathcal{N}(\mu_v, \sigma^2/2)$$

The pdf of u,v is

$$\begin{aligned} p(u,v) &= \frac{1}{\sqrt{2\pi\frac{\sigma^2}{2}}} \exp\left[-\frac{1}{2(\frac{\sigma^2}{2})}(u-\mu_u)^2\right] \\ &\quad \cdot \frac{1}{\sqrt{2\pi\frac{\sigma^2}{2}}} \exp\left[-\frac{1}{2(\frac{\sigma^2}{2})}(v-\mu_v)^2\right] \\ &= \frac{1}{\pi\sigma^2} \exp\left[-\frac{1}{\sigma^2}\left((u-\mu_u)^2 + (v-\mu_v)^2\right)\right] \end{aligned}$$

Twice the variance

Complex-valued models

Complex Gaussian PDFs

Let $\tilde{x} = u + jv$ where u and v are independent Gaussians

$\mathcal{N}(\mu_u, \sigma^2/2)$ and $\mathcal{N}(\mu_v, \sigma^2/2)$

The pdf of u,v is

$$
\begin{aligned}
p(u,v) &= \frac{1}{\sqrt{2\pi\frac{\sigma^2}{2}}} \exp\left[-\frac{1}{2(\frac{\sigma^2}{2})}(u-\mu_u)^2\right] \\
&\quad \cdot \frac{1}{\sqrt{2\pi\frac{\sigma^2}{2}}} \exp\left[-\frac{1}{2(\frac{\sigma^2}{2})}(v-\mu_v)^2\right] \\
&= \frac{1}{\pi\sigma^2} \exp\left[-\frac{1}{\sigma^2}\left((u-\mu_u)^2 + (v-\mu_v)^2\right)\right]
\end{aligned}
$$

Square root

Complex-valued models

Complex Gaussian PDFs

Let $\tilde{x} = u + jv$ where u and v are independent Gaussians $\mathcal{N}(\mu_u, \sigma^2/2)$ and $\mathcal{N}(\mu_v, \sigma^2/2)$

The pdf of u,v is

$$\begin{aligned} p(u,v) &= \frac{1}{\sqrt{2\pi\frac{\sigma^2}{2}}} \exp\left[-\frac{1}{2(\frac{\sigma^2}{2})}(u-\mu_u)^2\right] \\ &\quad \cdot \frac{1}{\sqrt{2\pi\frac{\sigma^2}{2}}} \exp\left[-\frac{1}{2(\frac{\sigma^2}{2})}(v-\mu_v)^2\right] \\ &= \frac{1}{\pi\sigma^2} \exp\left[-\frac{1}{\sigma^2}\left((u-\mu_u)^2 + (v-\mu_v)^2\right)\right] \end{aligned}$$

Define $\tilde{\mu} = E(\tilde{x}) = \mu_u + j\mu_v$

Then $p(\tilde{x}) = \dfrac{1}{\pi\sigma^2} \exp\left[-\dfrac{1}{\sigma^2}|\tilde{x} - \tilde{\mu}|^2\right]$

Complex-valued models

Complex Gaussian PDFs

Let $\tilde{x} = u + jv$ where u and v are independent Gaussians $\mathcal{N}(\mu_u, \sigma^2/2)$ and $\mathcal{N}(\mu_v, \sigma^2/2)$

The pdf of u,v is

$$\begin{aligned} p(u, v) &= \frac{1}{\sqrt{2\pi\frac{\sigma^2}{2}}} \exp\left[-\frac{1}{2(\frac{\sigma^2}{2})}(u-\mu_u)^2\right] \\ &\quad \cdot \frac{1}{\sqrt{2\pi\frac{\sigma^2}{2}}} \exp\left[-\frac{1}{2(\frac{\sigma^2}{2})}(v-\mu_v)^2\right] \\ &= \frac{1}{\pi\sigma^2} \exp\left[-\frac{1}{\sigma^2}\left((u-\mu_u)^2 + (v-\mu_v)^2\right)\right] \end{aligned}$$

Define $\tilde{\mu} = E(\tilde{x}) = \mu_u + j\mu_v$

Then $p(\tilde{x}) = \frac{1}{\pi\sigma^2} \exp\left[-\frac{1}{\sigma^2}|\tilde{x} - \tilde{\mu}|^2\right]$

Variance (since *$var(\tilde{x}) = var(x_R) + var(x_I)$*)

Complex-valued models

Complex Gaussian PDFs

Let $\tilde{x} = u + jv$ where u and v are independent Gaussians

$\mathcal{N}(\mu_u, \sigma^2/2)$ and $\mathcal{N}(\mu_v, \sigma^2/2)$

The pdf of u,v is

$$\begin{aligned} p(u,v) &= \frac{1}{\sqrt{2\pi\frac{\sigma^2}{2}}} \exp\left[-\frac{1}{2(\frac{\sigma^2}{2})}(u-\mu_u)^2\right] \\ &\quad \cdot \frac{1}{\sqrt{2\pi\frac{\sigma^2}{2}}} \exp\left[-\frac{1}{2(\frac{\sigma^2}{2})}(v-\mu_v)^2\right] \\ &= \frac{1}{\pi\sigma^2} \exp\left[-\frac{1}{\sigma^2}\left((u-\mu_u)^2 + (v-\mu_v)^2\right)\right] \end{aligned}$$

Define $\tilde{\mu} = E(\tilde{x}) = \mu_u + j\mu_v$

Then $p(\tilde{x}) = \frac{1}{\pi\sigma^2} \exp\left[-\frac{1}{\sigma^2}|\tilde{x} - \tilde{\mu}|^2\right]$

No square root

Theorem 15.1 (Complex Multivariate Gaussian PDF) *If a real random vector* $\mathbf{x}$ *of dimension* $2n \times 1$ *can be partitioned as*

$$\mathbf{x} = \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = \begin{bmatrix} n \times 1 \\ n \times 1 \end{bmatrix}$$

where $\mathbf{u}, \mathbf{v}$ *are real random vectors and* $\mathbf{x}$ *has the PDF*

$$\mathbf{x} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_u \\ \boldsymbol{\mu}_v \end{bmatrix}, \begin{bmatrix} \mathbf{C}_{uu} & \mathbf{C}_{uv} \\ \mathbf{C}_{vu} & \mathbf{C}_{vv} \end{bmatrix} \right)$$

and $\mathbf{C}_{uu} = \mathbf{C}_{vv}$ *and* $\mathbf{C}_{uv} = -\mathbf{C}_{vu}$
then defining the $n \times 1$ *complex random vector* $\tilde{\mathbf{x}} = \mathbf{u} + j\mathbf{v}$, $\tilde{\mathbf{x}}$ *has the complex multivariate Gaussian PDF*

$$\tilde{\mathbf{x}} \sim \mathcal{CN}(\tilde{\boldsymbol{\mu}}, \mathbf{C}_{\tilde{x}})$$

where

$$\begin{aligned} \tilde{\boldsymbol{\mu}} &= \boldsymbol{\mu}_u + j\boldsymbol{\mu}_v \\ \mathbf{C}_{\tilde{x}} &= 2(\mathbf{C}_{uu} + j\mathbf{C}_{vu}) \end{aligned}$$

or more explicitly

$$p(\tilde{\mathbf{x}}) = \frac{1}{\pi^n \det(\mathbf{C}_{\tilde{x}})} \exp\left[-(\tilde{\mathbf{x}} - \tilde{\boldsymbol{\mu}})^H \mathbf{C}_{\tilde{x}}^{-1} (\tilde{\mathbf{x}} - \tilde{\boldsymbol{\mu}})\right]. \tag{15.22}$$

Theorem 15.1 (Complex Multivariate Gaussian PDF) *If a real random vector* $\mathbf{x}$ *of dimension* $2n \times 1$ *can be partitioned as*

$$\mathbf{x} = \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = \begin{bmatrix} n \times 1 \\ n \times 1 \end{bmatrix}$$

where $\mathbf{u}, \mathbf{v}$ *are real random vectors and* $\mathbf{x}$ *has the PDF*

$$\mathbf{x} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_u \\ \boldsymbol{\mu}_v \end{bmatrix}, \begin{bmatrix} \mathbf{C}_{uu} & \mathbf{C}_{uv} \\ \mathbf{C}_{vu} & \mathbf{C}_{vv} \end{bmatrix} \right)$$

and $\mathbf{C}_{uu} = \mathbf{C}_{vv}$ *and* $\mathbf{C}_{uv} = -\mathbf{C}_{vu}$

Important remark: Not all vectors with complex Gaussian elements are multivariate complex Gaussian. Not even if the underlying real Gaussians are jointly Gaussian

It MUST hold that $\mathbf{C}_{uu} = \mathbf{C}_{vv}$ *and* $\mathbf{C}_{uv} = -\mathbf{C}_{vu}$

Meaning of this is, e.g., that the variance is equal for the real and imag parts (circularily symmetric)

Complex-valued models

Example

Let $\tilde{\mathbf{x}} = \mathbf{H}\boldsymbol{\theta} + \tilde{\mathbf{w}}$ $\tilde{\mathbf{w}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$

Thus, $\tilde{\mathbf{x}} \sim \mathcal{CN}(\mathbf{H}\boldsymbol{\theta}, \mathbf{C})$

Find MLE

Complex-valued models

Example

Let $\tilde{\mathbf{x}} = \mathbf{H}\boldsymbol{\theta} + \tilde{\mathbf{w}}$ $\qquad \tilde{\mathbf{w}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$

Thus, $\tilde{\mathbf{x}} \sim \mathcal{CN}(\mathbf{H}\boldsymbol{\theta}, \mathbf{C})$

The pdf is $$p(\tilde{\mathbf{x}}; \boldsymbol{\theta}) = \frac{1}{\pi^N \det(\mathbf{C})} \exp\left[-(\tilde{\mathbf{x}} - \mathbf{H}\boldsymbol{\theta})^H \mathbf{C}^{-1} (\tilde{\mathbf{x}} - \mathbf{H}\boldsymbol{\theta})\right]$$

Complex-valued models

Example

Let $\tilde{\mathbf{x}} = \mathbf{H}\boldsymbol{\theta} + \tilde{\mathbf{w}}$ $\qquad \tilde{\mathbf{w}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$

Thus, $\tilde{\mathbf{x}} \sim \mathcal{CN}(\mathbf{H}\boldsymbol{\theta}, \mathbf{C})$

The pdf is $p(\tilde{\mathbf{x}};\boldsymbol{\theta}) = \frac{1}{\pi^N \det(\mathbf{C})} \exp\left[-(\tilde{\mathbf{x}} - \mathbf{H}\boldsymbol{\theta})^H \mathbf{C}^{-1} (\tilde{\mathbf{x}} - \mathbf{H}\boldsymbol{\theta})\right]$

Take log, and differential

$$\frac{\partial \ln p(\tilde{\mathbf{x}};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\frac{\partial (\tilde{\mathbf{x}} - \mathbf{H}\boldsymbol{\theta})^H \mathbf{C}^{-1} (\tilde{\mathbf{x}} - \mathbf{H}\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

Complex-valued models

Example

Let $\tilde{\mathbf{x}} = \mathbf{H}\boldsymbol{\theta} + \tilde{\mathbf{w}}$ $\tilde{\mathbf{w}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$

Thus, $\tilde{\mathbf{x}} \sim \mathcal{CN}(\mathbf{H}\boldsymbol{\theta}, \mathbf{C})$

The pdf is $p(\tilde{\mathbf{x}}; \boldsymbol{\theta}) = \frac{1}{\pi^N \det(\mathbf{C})} \exp\left[-(\tilde{\mathbf{x}} - \mathbf{H}\boldsymbol{\theta})^H \mathbf{C}^{-1} (\tilde{\mathbf{x}} - \mathbf{H}\boldsymbol{\theta})\right]$

Take log, and differential

$$\frac{\partial \ln p(\tilde{\mathbf{x}}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\frac{\partial (\tilde{\mathbf{x}} - \mathbf{H}\boldsymbol{\theta})^H \mathbf{C}^{-1} (\tilde{\mathbf{x}} - \mathbf{H}\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

Expand $J = \tilde{\mathbf{x}}^H \mathbf{C}^{-1} \tilde{\mathbf{x}} - \tilde{\mathbf{x}}^H \mathbf{C}^{-1} \mathbf{H}\boldsymbol{\theta} - \boldsymbol{\theta}^H \mathbf{H}^H \mathbf{C}^{-1} \tilde{\mathbf{x}} + \boldsymbol{\theta}^H \mathbf{H}^H \mathbf{C}^{-1} \mathbf{H}\boldsymbol{\theta}$

$$\frac{\partial J}{\partial \boldsymbol{\theta}} =$$

Complex-valued models

Example

$$\frac{\partial \mathbf{b}^H \boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = \mathbf{b}^* \qquad \frac{\partial \boldsymbol{\theta}^H \mathbf{b}}{\partial \boldsymbol{\theta}} = 0$$

$$\frac{\partial \boldsymbol{\theta}^H \mathbf{A} \boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = (\mathbf{A}\boldsymbol{\theta})^*, \quad \text{where } \mathbf{A}^H = \mathbf{A}$$

Expand

$$J = \tilde{\mathbf{x}}^H \mathbf{C}^{-1} \tilde{\mathbf{x}} - \tilde{\mathbf{x}}^H \mathbf{C}^{-1} \mathbf{H}\boldsymbol{\theta} - \boldsymbol{\theta}^H \mathbf{H}^H \mathbf{C}^{-1} \tilde{\mathbf{x}} + \boldsymbol{\theta}^H \mathbf{H}^H \mathbf{C}^{-1} \mathbf{H}\boldsymbol{\theta}$$

$$\frac{\partial J}{\partial \boldsymbol{\theta}} =$$

Complex-valued models

Example

$$\frac{\partial \mathbf{b}^H \boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = \mathbf{b}^* \qquad \frac{\partial \boldsymbol{\theta}^H \mathbf{b}}{\partial \boldsymbol{\theta}} = \mathbf{0}$$

$$\frac{\partial \boldsymbol{\theta}^H \mathbf{A} \boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = (\mathbf{A}\boldsymbol{\theta})^*, \quad \text{where } \mathbf{A}^H = \mathbf{A}$$

Expand

$$J = \tilde{\mathbf{x}}^H \mathbf{C}^{-1} \tilde{\mathbf{x}} - \tilde{\mathbf{x}}^H \mathbf{C}^{-1} \mathbf{H}\boldsymbol{\theta} - \boldsymbol{\theta}^H \mathbf{H}^H \mathbf{C}^{-1} \tilde{\mathbf{x}} + \boldsymbol{\theta}^H \mathbf{H}^H \mathbf{C}^{-1} \mathbf{H}\boldsymbol{\theta}$$

$$\begin{aligned} \frac{\partial J}{\partial \boldsymbol{\theta}} &= \mathbf{0} - \left(\mathbf{H}^H \mathbf{C}^{-1} \tilde{\mathbf{x}}\right)^* - \mathbf{0} + \left(\mathbf{H}^H \mathbf{C}^{-1} \mathbf{H}\boldsymbol{\theta}\right)^* \\ & \quad \left[\mathbf{H}^H \mathbf{C}^{-1} (\tilde{\mathbf{x}} - \mathbf{H}\boldsymbol{\theta})\right]^* \end{aligned}$$

Complex-valued models

Example

$$\frac{\partial \mathbf{b}^H \boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = \mathbf{b}^* \qquad \frac{\partial \boldsymbol{\theta}^H \mathbf{b}}{\partial \boldsymbol{\theta}} = 0$$

$$\frac{\partial \boldsymbol{\theta}^H \mathbf{A} \boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = (\mathbf{A}\boldsymbol{\theta})^*, \quad \text{where } \mathbf{A}^H = \mathbf{A}$$

Expand

$$J = \tilde{\mathbf{x}}^H \mathbf{C}^{-1} \tilde{\mathbf{x}} - \tilde{\mathbf{x}}^H \mathbf{C}^{-1} \mathbf{H}\boldsymbol{\theta} - \boldsymbol{\theta}^H \mathbf{H}^H \mathbf{C}^{-1} \tilde{\mathbf{x}} + \boldsymbol{\theta}^H \mathbf{H}^H \mathbf{C}^{-1} \mathbf{H}\boldsymbol{\theta}$$

$$\frac{\partial J}{\partial \boldsymbol{\theta}} = 0 - (\mathbf{H}^H \mathbf{C}^{-1} \tilde{\mathbf{x}})^* - 0 + (\mathbf{H}^H \mathbf{C}^{-1} \mathbf{H}\boldsymbol{\theta})^*$$

$$\left[\mathbf{H}^H \mathbf{C}^{-1} (\tilde{\mathbf{x}} - \mathbf{H}\boldsymbol{\theta})\right]^*$$

$$\boxed{\hat{\boldsymbol{\theta}} = (\mathbf{H}^H \mathbf{C}^{-1} \mathbf{H})^{-1} \mathbf{H}^H \mathbf{C}^{-1} \tilde{\mathbf{x}}}$$

Continuous time – discrete time

I will briefly cover two parts (if time permits)

1. Karhunen-Loeve decomposition
2. Practical (mathematical) conversion

On Bandwidth

DAVID SLEPIAN, FELLOW, IEEE

***Abstract*—It is easy to argue that real signals must be bandlimited. It is also easy to argue that they cannot be so.**

THE DILEMMA

ARE SIGNALS really bandlimited? They seem to be, and yet they seem not to be.

On the one hand, a pair of solid copper wires will not propagate electromagnetic waves at optical frequencies, and so the signals I receive over such a pair must be bandlimited. In fact, it makes little physical sense to talk of energy received over wires at frequencies higher than some finite cutoff W, say 10^{20} Hz. It would seem, then, that signals must be bandlimited.

On the other hand, however, signals of limited bandwidth W are finite Fourier transforms,

On Bandwidth

DAVID SLEPIAN, FELLOW, IEEE

$$s(t) = \int_{-W}^{W} e^{2\pi i f t} S(f)\, df$$

and irrefutable mathematical arguments show them to be extremely smooth. They possess derivatives of all orders. Indeed, such integrals are entire functions of t, completely predictable from any little piece, and they cannot vanish on any t interval unless they vanish everywhere. Such signals cannot start or stop, but must go on forever. Surely *real signals* start and stop, and so they cannot be bandlimited!

Continuous time – discrete time

Karhunen-Loeve decomposition

Recall the operations of the MMSE estimator in discrete time:

It starts by whitening the observations

Correlated signals carries little information, so a clear goal of the conversion should be to produce uncorrelated samples

Continuous time – discrete time

Karhunen-Loeve decomposition

We seek to expand the signal in the form

$$x(t) = \lim_{N \to \infty} \sum_{i=1}^{N} x_i \phi_i(t), \qquad 0 \leq t \leq T,$$

$$x_i \triangleq \int_0^T x(t)\, \phi_i(t)\, dt.$$

We work with convergence in the MSE sense, i.e.

$$\lim_{N \to \infty} E\left[\left(x_t - \sum_{i=1}^{N} x_i\, \phi_i(t)\right)^2\right] = 0, \qquad 0 \leq t \leq T.$$

Continuous time – discrete time

Karhunen-Loeve decomposition

We seek to expand the signal in the form

$$x(t) = \lim_{N \to \infty} \sum_{i=1}^{N} x_i \phi_i(t), \qquad 0 \le t \le T,$$

$$x_i \triangleq \int_0^T x(t)\, \phi_i(t)\, dt.$$

The constraint that the x_i's should be uncorrelated reads

$$E(x_i x_j) = \lambda_i \delta_{ij}.$$

Continuous time – discrete time

Karhunen-Loeve decomposition

We seek to expand the signal in the form

$$x(t) = \lim_{N \to \infty} \sum_{i=1}^{N} x_i \phi_i(t), \qquad 0 \le t \le T,$$

$$x_i \triangleq \int_0^T x(t)\, \phi_i(t)\, dt.$$

1. The value x_i^2 has a simple physical interpretation. It corresponds to the *energy* along the coordinate function $\phi_i(t)$ in a particular sample function.

2. Similarly, $E(x_i^2) = \lambda_i$ corresponds to the *expected* value of the energy along $\phi_i(t)$, assuming that $m_i = 0$. Clearly, $\lambda_i \ge 0$ for all i.

Continuous time – discrete time

Karhunen-Loeve decomposition

Let us now examine the requirements on the basis for having uncorrelated coefficients

$$\begin{aligned}\lambda_i \delta_{ij} &= E(x_i x_j) \\ &= E\left[\int_0^T x(t)\,\phi_i(t)\,dt \int_0^T x(u)\,\phi_j(u)\,du\right] \\ &= \int_0^T \phi_i(t)\,dt \int_0^T K_x(t,u)\,\phi_j(u)\,du, \qquad \text{for all } i \text{ and } j.\end{aligned}$$

$K_x(t,u) = E(x(t)x(u))$

Continuous time – discrete time

Karhunen-Loeve decomposition

Let us now examine the requirements on the basis for having uncorrelated coefficients

$$
\begin{aligned}
\lambda_i \delta_{ij} &= E(x_i x_j) \\
&= E\left[\int_0^T x(t)\,\phi_i(t)\,dt \int_0^T x(u)\,\phi_j(u)\,du\right] \\
&= \int_0^T \phi_i(t)\,dt \underbrace{\int_0^T K_x(t,u)\,\phi_j(u)\,du,}_{=\ \lambda_j\,\phi_j(t)} \qquad \text{for all } i \text{ and } j.
\end{aligned}
$$

Continuous time – discrete time

Karhunen-Loeve decomposition

Let us now examine the requirements on the basis for having uncorrelated coefficients

$$
\begin{aligned}
\lambda_i \delta_{ij} &= E(x_i x_j) \\
&= E\left[\int_0^T x(t)\,\phi_i(t)\,dt \int_0^T x(u)\,\phi_j(u)\,du\right] \\
&= \int_0^T \phi_i(t)\,dt \underbrace{\int_0^T K_x(t, u)\,\phi_j(u)\,du,}_{=\ \lambda_j\,\phi_j(t)} \qquad \text{for all } i \text{ and } j.
\end{aligned}
$$

$$
\boxed{\lambda_j\,\phi_j(t) = \int_0^T K_x(t, u)\,\phi_j(u)\,du, \qquad 0 \le t \le T.}
$$

Continuous time – discrete time

Karhunen-Loeve decomposition

Most important properties:

$$E\left[\int_0^T x^2(t)\,dt\right] = \int_0^T K_x(t,t)\,dt = \sum_{i=1}^{\infty} \lambda_i$$

Mercer's theorem $$K_x(t,u) = \sum_{i=1}^{\infty} \lambda_i\,\phi_i(t)\,\phi_i(u), \qquad 0 \le t, u \le T,$$

$$\boxed{\lambda_j\,\phi_j(t) = \int_0^T K_x(t,u)\,\phi_j(u)\,du, \qquad 0 \le t \le T.}$$

Continuous time – discrete time

Karhunen-Loeve decomposition

Most important properties:

Reconstruction precision

$$\xi_N(t) \triangleq E\left[\left(x(t) - \sum_{i=1}^{N} x_i\,\phi_i(t)\right)^2\right]$$

for any $\epsilon > 0$ there exists an N_1 *independent of t* such that $\xi_N(t) < \epsilon$ for all $N > N_1$

$$\boxed{\lambda_j\,\phi_j(t) = \int_0^T K_x(t, u)\,\phi_j(u)\,du, \qquad 0 \le t \le T.}$$

Continuous time – discrete time

Karhunen-Loeve decomposition

Most important properties:

Reconstruction precision

$$\xi_N(t) \triangleq E\left[\left(x(t) - \sum_{i=1}^{N} x_i\, \phi_i(t)\right)^2\right]$$

for any $\epsilon > 0$ there exists an N_1 *independent of* t such that $\xi_N(t) < \epsilon$ for all $N > N_1$

Thus, with N_1 basis functions, we can represent x(t) with a small loss ε

$$\boxed{\lambda_j\, \phi_j(t) = \int_0^T K_x(t, u)\, \phi_j(u)\, du, \qquad 0 \le t \le T.}$$

Continuous time – discrete time

Karhunen-Loeve decomposition

A very important application area of Karhunen – Loeve: Time variant Gaussian process

For more details, see Van Trees 1968

Continuous time – discrete time

Karhunen-Loeve decomposition

Bandlimited processes

Spectrum of process: $$S_x(\omega) = \begin{cases} \dfrac{P}{2W}, & |f| \leq W, \\ 0, & |f| > W, \end{cases}$$

Observation time of the process: T seconds

Continuous time – discrete time

Karhunen-Loeve decomposition

Bandlimited processes

Spectrum of process: $$S_x(\omega) = \begin{cases} \dfrac{P}{2W}, & |f| \leq W, \\ 0, & |f| > W, \end{cases}$$

Observation time of the process: T seconds

After some math $$K_x(t, u) = P \frac{\sin \alpha(t - u)}{\alpha(t - u)}$$

And we need to solve $$\lambda \, \phi(t) = \int_{-T/2}^{+T/2} P \frac{\sin \alpha(t - u)}{\alpha(t - u)} \, \phi(u) \, du.$$

Continuous time – discrete time

Karhunen-Loeve decomposition

Bandlimited processes

Spectrum of process:

$$S_x(\omega) = \begin{cases} \frac{P}{2W}, & |f| \leq W, \\ 0, & |f| > W, \end{cases}$$

Observation time of the process: T seconds

After some math

$$K_x(t, u) = P \frac{\sin \alpha(t - u)}{\alpha(t - u)}$$

And we need to solve

$$\lambda \, \phi(t) = \int_{-T/2}^{+T/2} P \frac{\sin \alpha(t - u)}{\alpha(t - u)} \phi(u) \, du.$$

Solutions are well known: prolate spheroidal wave functions

Continuous time – discrete time

Karhunen-Loeve decomposition

A process of bandwidth W observed for T seconds is 2WT+1 dimensional

Interesting discovery

Eigenvalues above 2WT+1 are almost zero

$2WT = 2.55$
$\lambda_0 = 0.996 \frac{P}{2W}$
$\lambda_1 = 0.912 \frac{P}{2W}$
$\lambda_2 = 0.519 \frac{P}{2W}$
$\lambda_3 = 0.110 \frac{P}{2W}$
$\lambda_4 = 0.009 \frac{P}{2W}$
$\lambda_5 = 0.0004 \frac{P}{2W}$

$2WT = 5.10$
$\lambda_0 = 1.000 \frac{P}{2W}$
$\lambda_1 = 0.999 \frac{P}{2W}$
$\lambda_2 = 0.997 \frac{P}{2W}$
$\lambda_3 = 0.961 \frac{P}{2W}$
$\lambda_4 = 0.748 \frac{P}{2W}$
$\lambda_5 = 0.321 \frac{P}{2W}$
$\lambda_6 = 0.061 \frac{P}{2W}$
$\lambda_7 = 0.006 \frac{P}{2W}$
$\lambda_8 = 0.0004 \frac{P}{2W}$

Continuous time – discrete time

Sampling theorem (≈ asymptotic version of the prolate spheroidal wave functions)

A signal s(t) whose Fourier transform S(f) is limited to B (positive) Hz can be perfectly reconstructed from its samples $s(nT_0)$, n=…-1,0,1,2,…. Where $T_0 = 1/2B$

Continuous time – discrete time

Sampling theorem (≈ asymptotic version of the prolate spheroidal wave functions)

A signal s(t) whose Fourier transform S(f) is limited to B (positive) Hz can be perfectly reconstructed from its samples $s(nT_0)$, n=…-1,0,1,2,…. Where $T_0 = 1/2B$

In reality, the signal may be approximately band-limited, but the noise is typically not

Following the sampling theorem yields horrible results

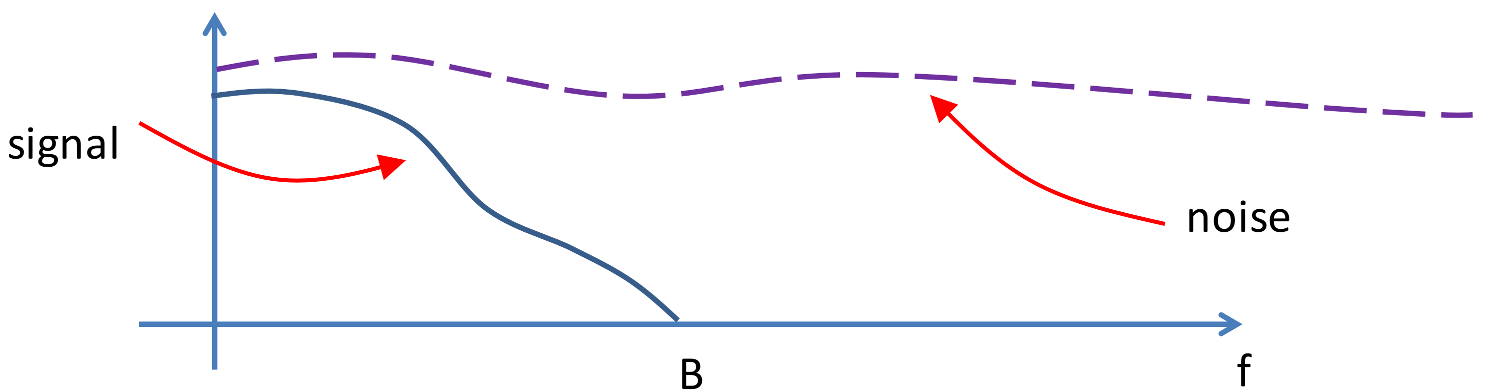
Continuous time – discrete time

Sampling theorem (≈ asymptotic version of the prolate spheroidal wave functions)

A signal s(t) whose Fourier transform S(f) is limited to B (positive) Hz can be perfectly reconstructed from its samples $s(nT_0)$, n=…-1,0,1,2,…. Where $T_0 = 1/2B$

In reality, the signal may be approximately band-limited, but the noise is typically not

Following the sampling theorem (based on signal only) yields horrible results



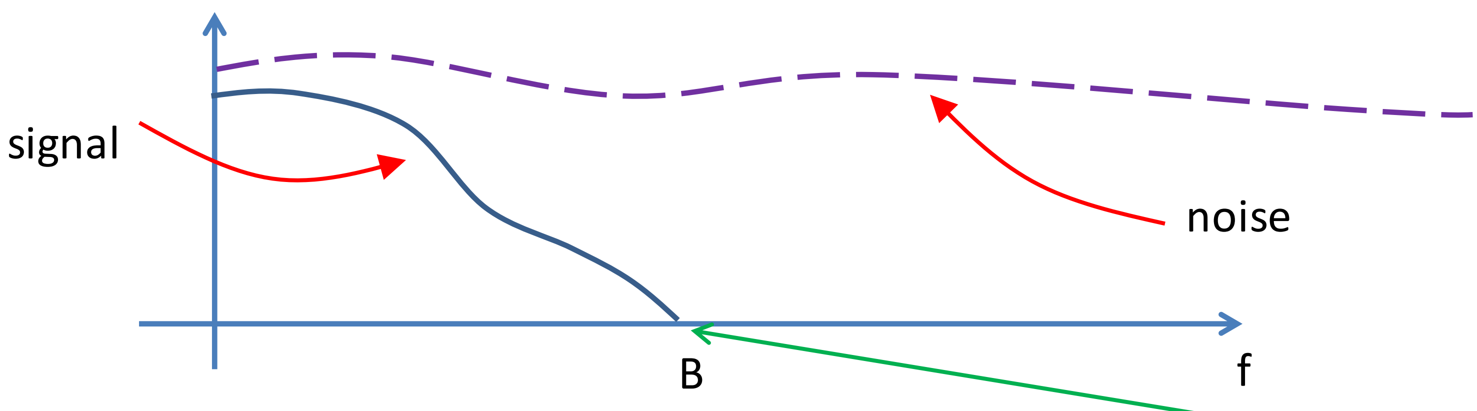
Continuous time – discrete time

Sampling theorem (≈ asymptotic version of the prolate spheroidal wave functions)

A signal s(t) whose Fourier transform S(f) is limited to B (positive) Hz can be perfectly reconstructed from its samples $s(nT_0)$, n=…-1,0,1,2,…. Where $T_0 = 1/2B$

In reality, the signal may be approximately band-limited, but the noise is typically not

Following the sampling theorem (based on signal only) yields horrible results



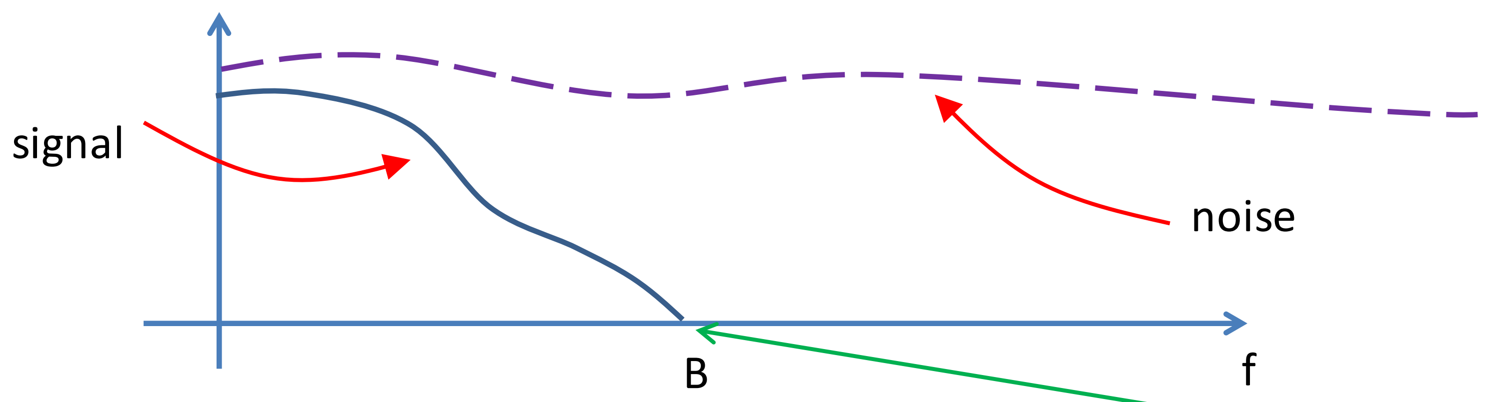
Selecting the sampling rate as 2B yields a discrete time spectrum with folding here

Continuous time – discrete time

Sampling theorem (≈ asymptotic version of the prolate spheroidal wave functions)

A signal s(t) whose Fourier transform S(f) is limited to B (positive) Hz can be perfectly reconstructed from its samples $s(nT_0)$, n=…-1,0,1,2,…. Where $T_0 = 1/2B$

$s[n] = s(n/2B)$ $S(w) = ????$



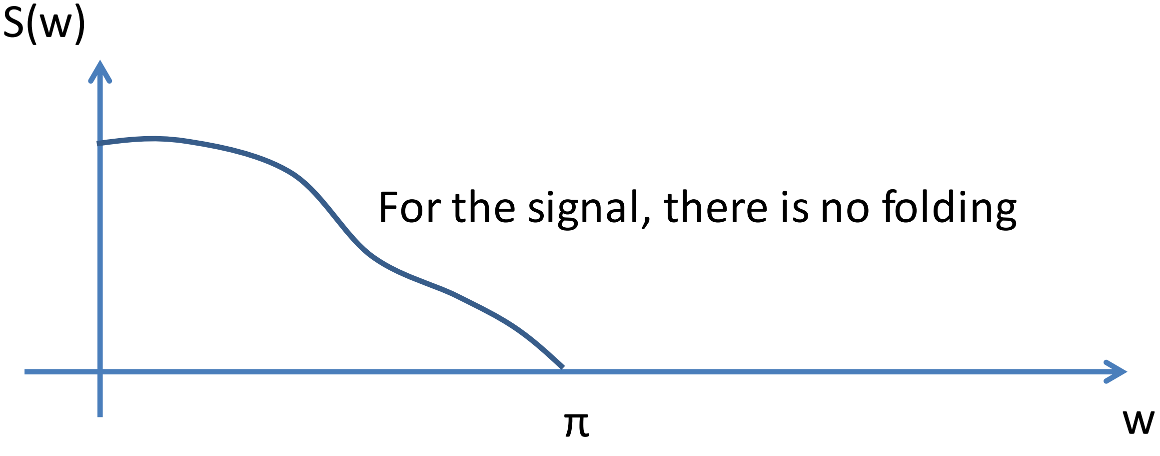
Selecting the sampling rate as 2B yields a discrete time spectrum with folding here

Continuous time – discrete time

Sampling theorem (≈ asymptotic version of the prolate spheroidal wave functions)

A signal s(t) whose Fourier transform S(f) is limited to B (positive) Hz can be perfectly reconstructed from its samples $s(nT_0)$, n=…-1,0,1,2,…. Where $T_0 = 1/2B$

s[n] = s(n/2B)   S(w) = ????

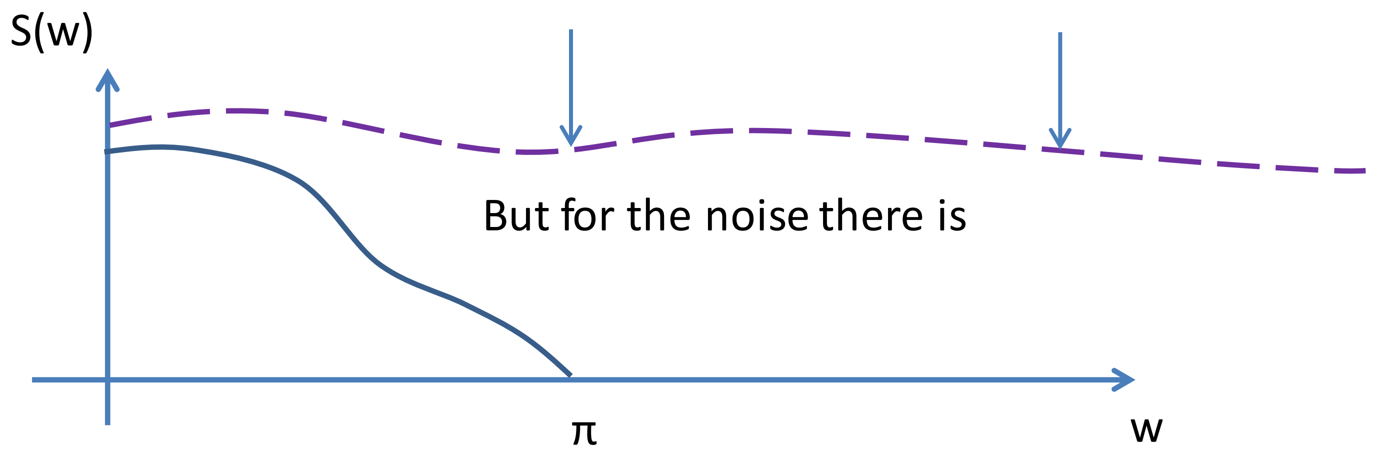


Continuous time – discrete time

Sampling theorem (≈ asymptotic version of the prolate spheroidal wave functions)

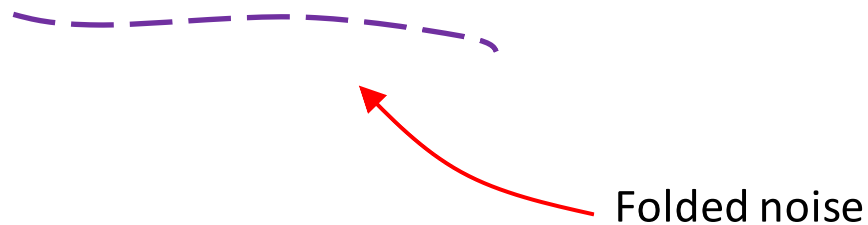
A signal s(t) whose Fourier transform S(f) is limited to B (positive) Hz can be perfectly reconstructed from its samples $s(nT_0)$, n=…-1,0,1,2,…. Where $T_0 = 1/2B$

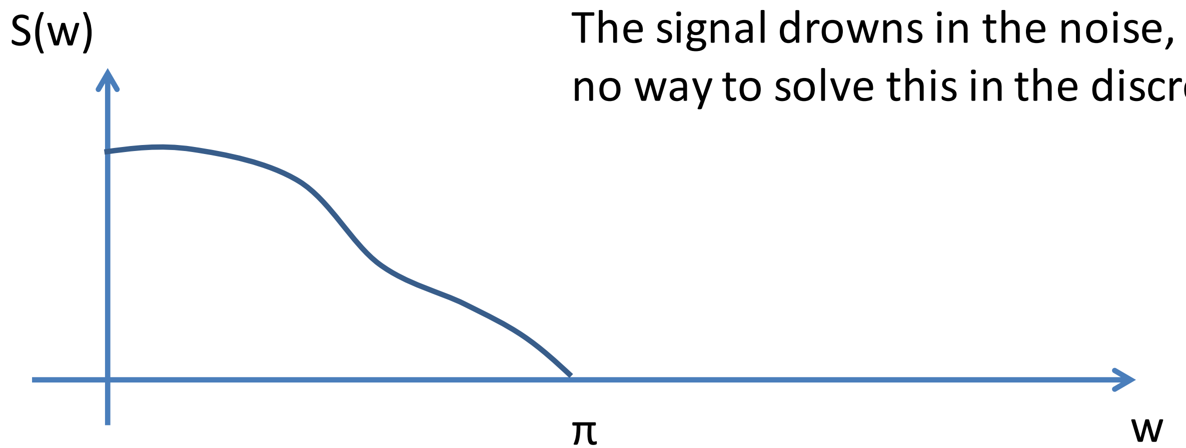
s[n] = s(n/2B) S(w) = ????



Continuous time – discrete time

Sampling theorem (≈ asymptotic version of the prolate spheroidal wave functions)





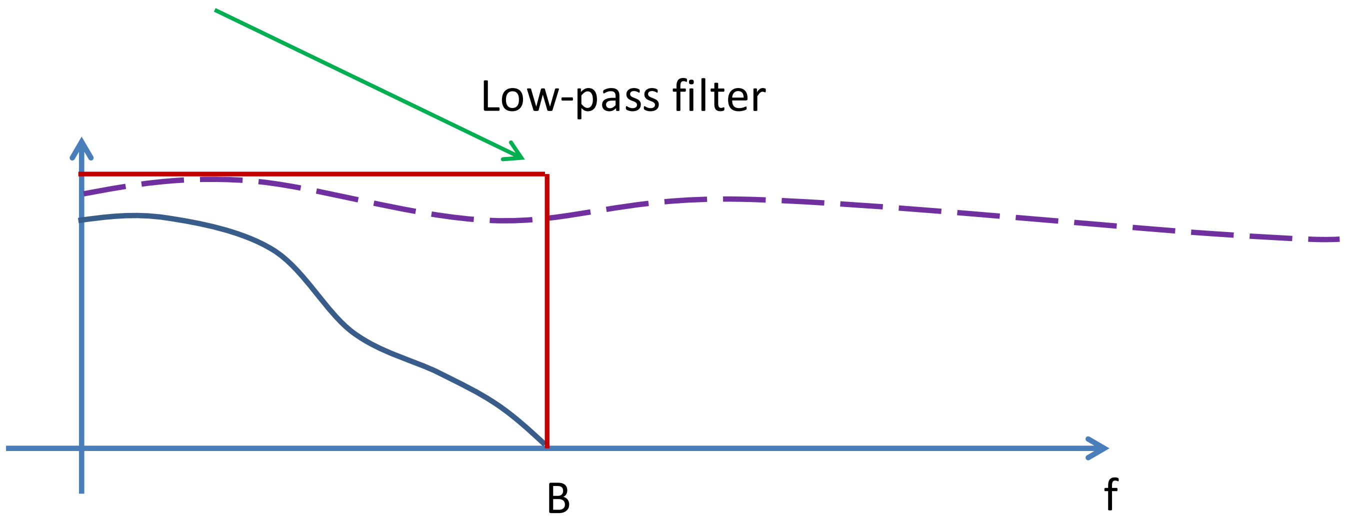
The signal drowns in the noise, and there is no way to solve this in the discrete domain

Continuous time – discrete time

Sampling theorem (≈ asymptotic version of the prolate spheroidal wave functions)

To handle this, we must first filter the signal in order to avoid to fold back any noise

Ideally, we should filter here

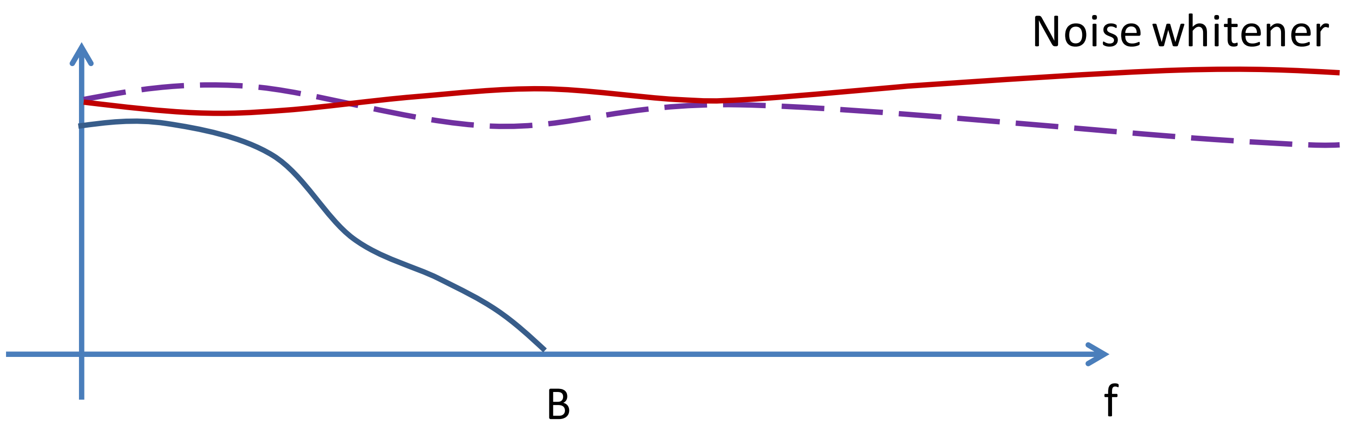


Continuous time – discrete time

Sampling theorem (≈ asymptotic version of the prolate spheroidal wave functions)

To handle this, we must first filter the signal in order to avoid to fold back any noise

However, let us first start with noise whitening. This is invertable, so nothing is lost

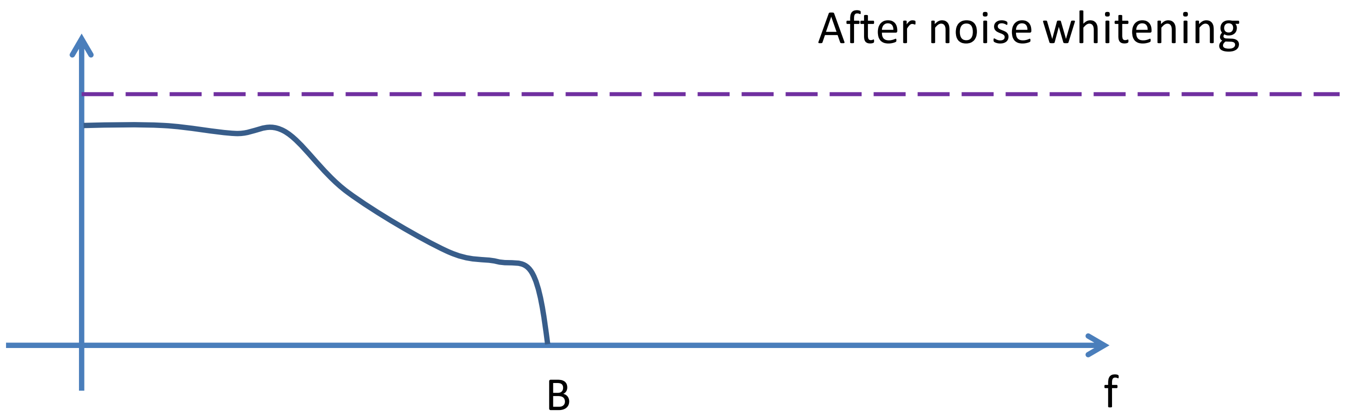


Continuous time – discrete time

Sampling theorem (≈ asymptotic version of the prolate spheroidal wave functions)

To handle this, we must first filter the signal in order to avoid to fold back any noise

However, let us first start with noise whitening. This is invertable, so nothing is lost

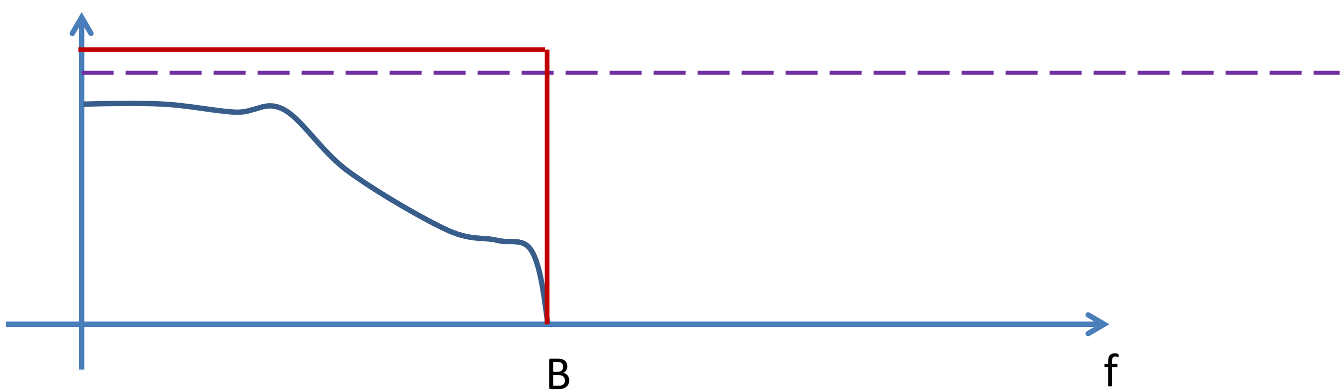


Continuous time – discrete time

Sampling theorem (≈ asymptotic version of the prolate spheroidal wave functions)

To handle this, we must first filter the signal in order to avoid to fold back any noise

We can now apply the low-pass filter

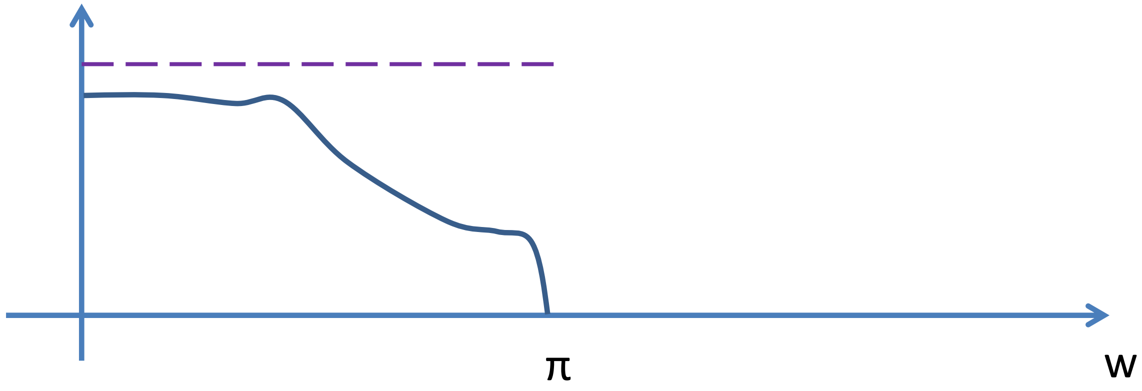


Continuous time – discrete time

Sampling theorem (≈ asymptotic version of the prolate spheroidal wave functions)

To handle this, we must first filter the signal in order to avoid to fold back any noise

Discrete time model will be



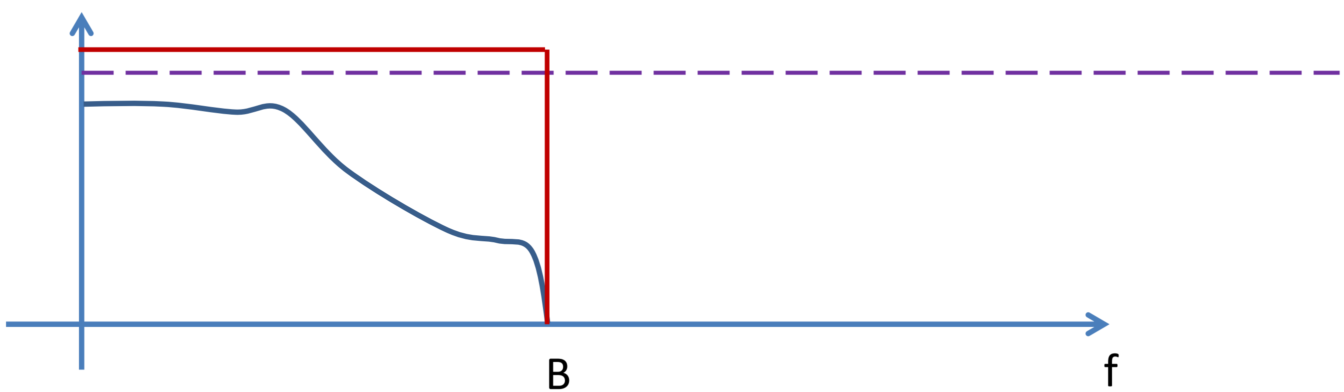
Thus, the spectrum of the discrete time model coincides with useful part of the continiuous signal spectrum. Optimal A2D

Continuous time – discrete time

Sampling theorem (≈ asymptotic version of the prolate spheroidal wave functions)

To handle this, we must first filter the signal in order to avoid to fold back any noise

However, this LPF is hard to implement

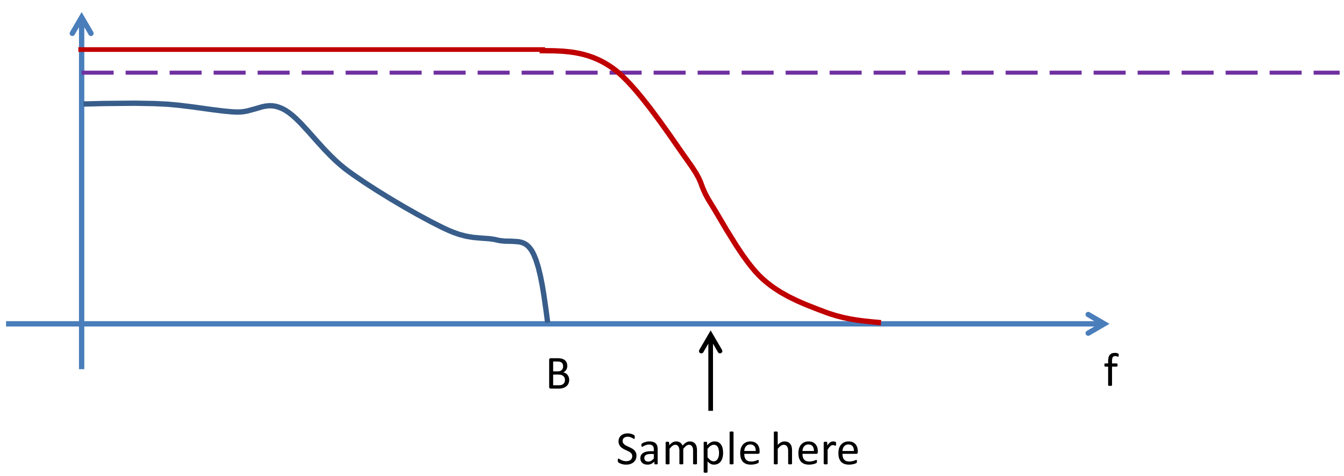


Continuous time – discrete time

Sampling theorem (≈ asymptotic version of the prolate spheroidal wave functions)

To handle this, we must first filter the signal in order to avoid to fold back any noise

Let us filter with a realistic filter, and change the sampling rate

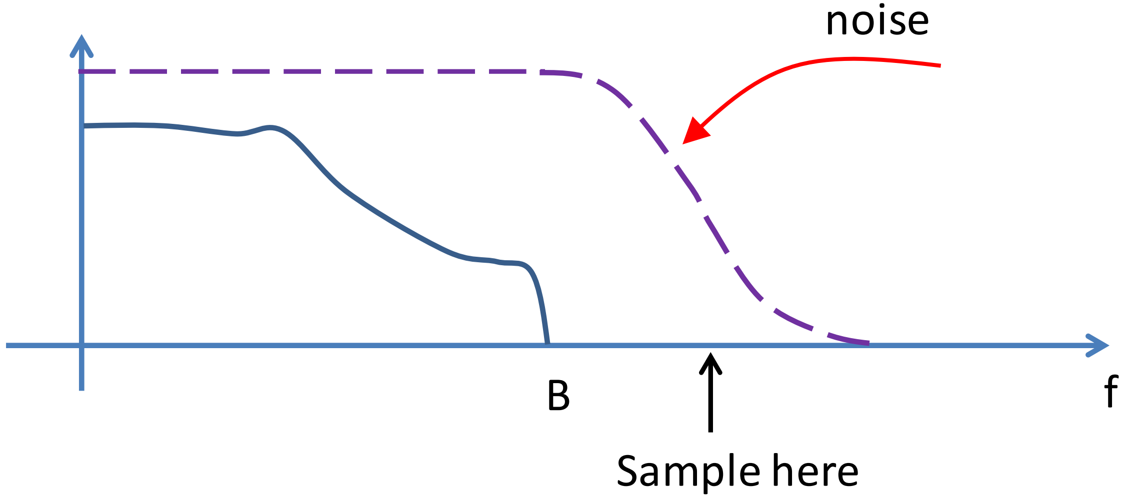


Continuous time – discrete time

Sampling theorem (≈ asymptotic version of the prolate spheroidal wave functions)

To handle this, we must first filter the signal in order to avoid to fold back any noise

After filtering (but before sampling)

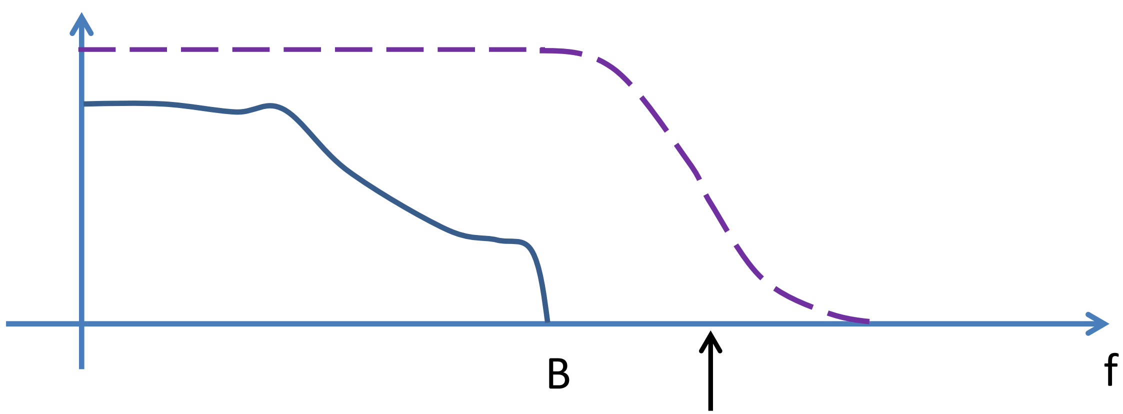


Continuous time – discrete time

Sampling theorem (≈ asymptotic version of the prolate spheroidal wave functions)

To handle this, we must first filter the signal in order to avoid to fold back any noise

Now sample



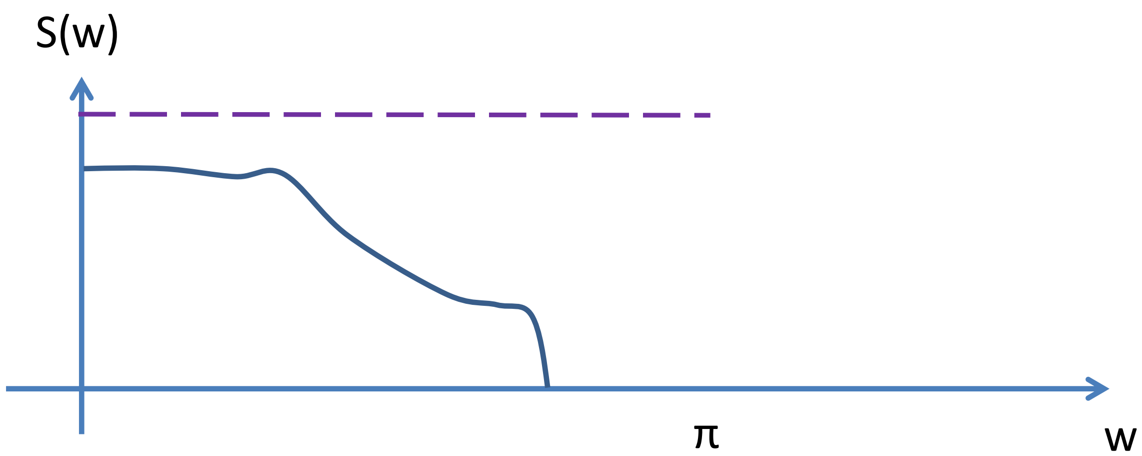
We should take the part to the right and fold back

Continuous time – discrete time

Sampling theorem (≈ asymptotic version of the prolate spheroidal wave functions)

To handle this, we must first filter the signal in order to avoid to fold back any noise

Result



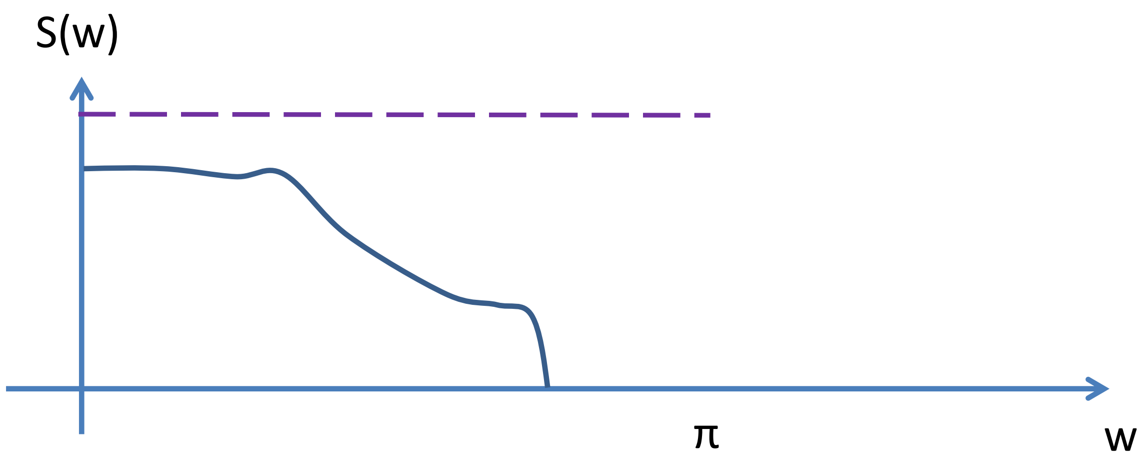
Is it lossy or not?

Continuous time – discrete time

Sampling theorem (≈ asymptotic version of the prolate spheroidal wave functions)

To handle this, we must first filter the signal in order to avoid to fold back any noise

Result



Is it lossy or **not**? The fact that there is ”more noise than necessary” can be handled by the discrete-time processing. System is over-sampled