Module 1 - Signal estimation

Sérgio M. Jesus
(sjesus@ualg.pt)

Universidade do Algarve,
PT-8005-139 Faro, Portugal
www.siplab.fct.ualg.pt

February 2009

Outline of Module 1

- **Parameter estimation**
  (1) the problem
  (2) example of estimators
- **The underwater channel as a linear system**
  (1) data model and assumptions
  (2) the space-time filter: real data example
- **Data model and optimality**
  (1) the discrete data model
  (2) deterministic signal in noise: the optimal noise enhancer
  (3) the generalized matched filter (GMF)
  (4) the correlated noise case
  (5) the multichannel optimal receiver

A parameter estimation problem

Assume de discrete time observations

$$\mathbf{x}^t = [x(0), x(1), \ldots, x(N-1)] \quad (1)$$

that depends on some parameter θ that we want to estimate.

Estimator $\hat{\theta}$ of θ can be written as a deterministic function g of the observed data set

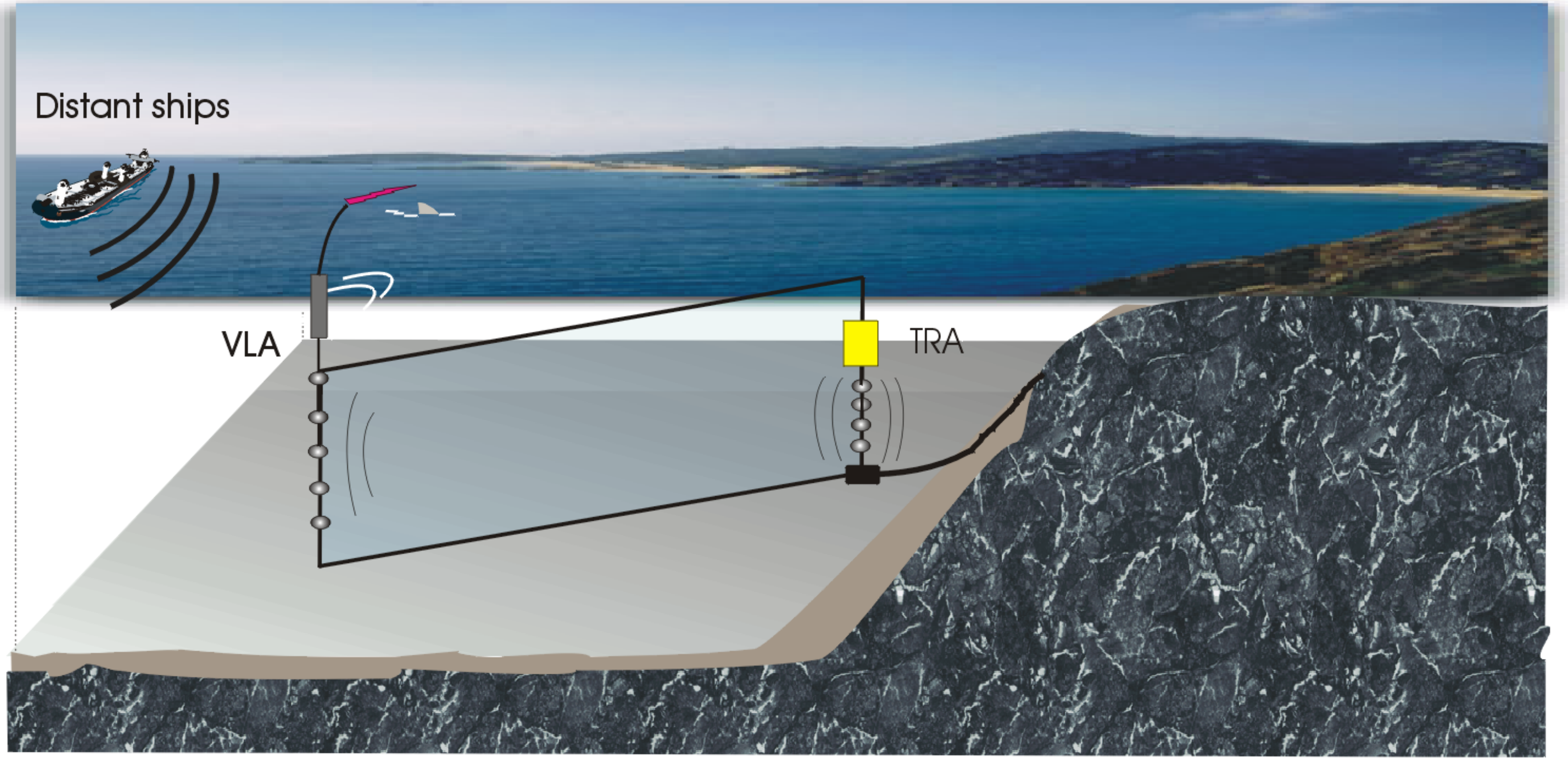
$$\hat{\theta} = g(\mathbf{x}) \quad (2)$$

Our goal is to determine g that provides "the best estimator $\hat{\theta}$ of θ".

$\Rightarrow$ best in which sense ?

$\Rightarrow$ how to determine g ?

Example: the mean of a time series



Our data model in this case will appopriately be $x(n) = A + w(n)$, where A is a constant to be determined and $w(n)$ is a zero-mean random process.

An intituitive estimator of A would be

$$\hat{A} = \frac{1}{N}\sum_{n=0}^{N-1} x(n) \tag{3}$$

in fact if the model is true, in a mean

$$E[\hat{A}] = \frac{1}{N}\sum_{n=0}^{N-1} E[x(n)] = A \tag{4}$$

Estimator $\hat{A}$ is said to be **unbiased**. But there are many other estimators of A as for example, $\tilde{A} = x(0)$. Which one of this estimators is the best estimator of A ?

Example: the mean of a time series (cont.)

Let us assume that $w(n) : \mathcal{N}(0, \sigma^2)$, then

$$E[\tilde{A}] = E[x(0)] = A + E[w(0)] = A \tag{5}$$

which basically says that $\tilde{A}$ is also unbiased. So, we have to resort to second order statistics...

$$\begin{aligned} V[\hat{A}] &= V\left[\frac{1}{N}\sum_{n=0}^{N-1} x(n)\right] \\ &= \frac{1}{N^2}\sum_{n=0}^{N-1} V[w(n)] \\ &= \frac{\sigma^2}{N} \end{aligned} \tag{6}$$

while for $\tilde{A}$

$$\begin{aligned} V[\tilde{A}] &= V[x(0)] \\ &= V[w(0)] \\ &= \sigma^2 \end{aligned} \tag{7}$$

and therefore the variance of $\hat{A}$ is smaller than that of $\tilde{A}$ by a factor N.

$\Rightarrow$ search for the **Minimum Variance Unbiased Estimator (MVUE)**

Minimum Variance Unbiased Estimator (MVUE)

Minimum variance and no bias are estimator characteristics but do not tell us how close we are from the true value $\Rightarrow$ mean square error (MSE).

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] \tag{8}$$

but

$$\begin{aligned} MSE(\hat{\theta}) &= E\left[[(\hat{\theta} - E[\hat{\theta}]) + (E[\hat{\theta}] - \theta)]^2\right] \\ &= V[\hat{\theta}] + \left[E(\hat{\theta}) - \theta\right]^2 \\ &= V[\hat{\theta}] + b^2(\hat{\theta}) \end{aligned}$$

so, for unbiased estimators $b(\hat{\theta}) = 0$ and therefore minimum variance means minimum MSE.

An MVUE does not always exists and if it exists it is not ensured to be able to find it.

How to find the MVUE

1. determine the Cramer-Rao Lower Bound (CRLB) and check if some estimator satisfies it - that is the MVUE.

2. if no estimator satisfies the bound, use a sufficient statistic and apply the Rao-Blackwell-Lehmann-Scheffe theorem.

3. restrict the class of estimators not only unbiased and of minimum variance but also linear in the data so that g is a linear function. This is the Best Linear Unbiased Estimator (BLUE), which may coincide with the MVUE if it happens to be linear.

CRLB

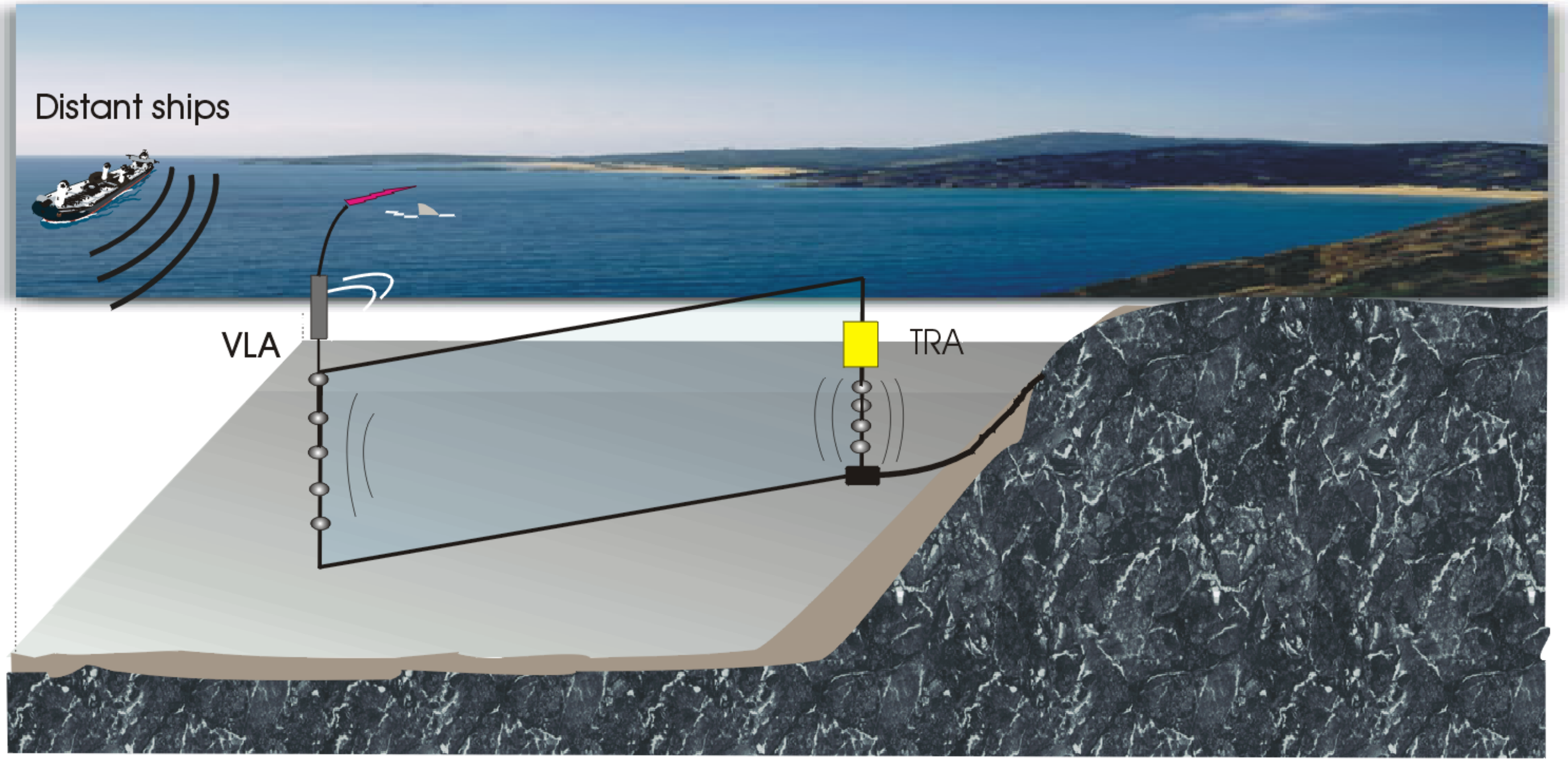
The CRLB states how well a given parameter can be estimated. That ability depends on the sharpness of the Probability Density Function (PDF) against the parameter given the data.

PDF of data dependence on a given parameter, assuming Gaussian

$$p(x[0];A) = \frac{1}{\sqrt{2\imath\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x[0]-A)^2\right] \tag{9}$$

$$V[\hat{\theta}] \geq \frac{1}{-E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial\theta}\right]} \tag{10}$$

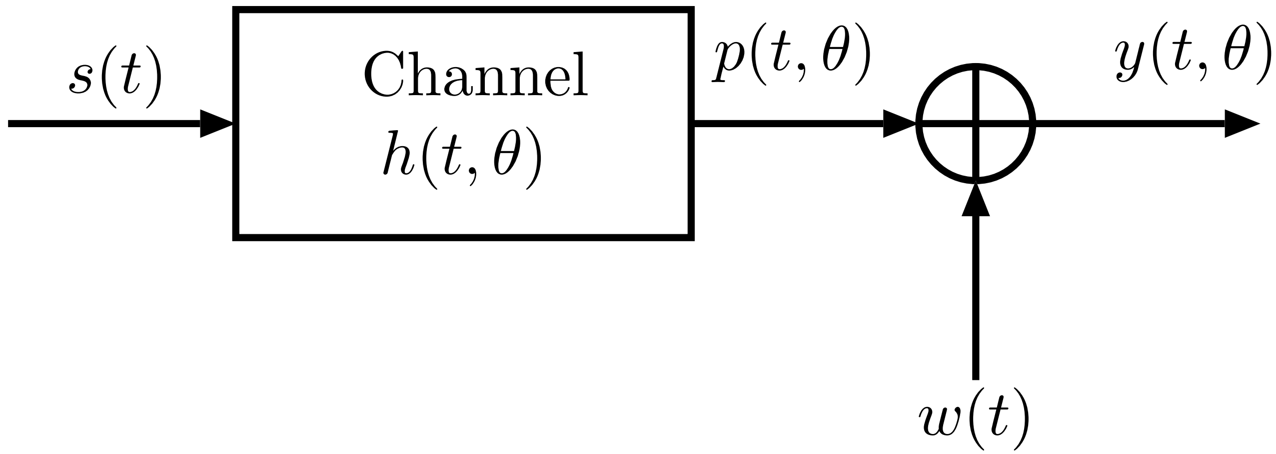
Assume data model, hypothesis on signal(s), noise and interferences.


Distant ships
VLA
TRA

Problems and applications

- **Direct problem:** determine acoustic pressure $\rightarrow$ propagation model(s)
- **Geometric inverse problem:** determine source characteristics
  - *is there a source present ?* detection problem
  - *where is the source emitting from ?* estimate source position,
  - *what is the source emitting ?* estimate emitted signal
- **Environmental inverse problem:** monitoring / exploring the environment
  - tomography: estimate water column temperature in 3D and time
  - bottom properties: estimation bottom properties in 3D

The underwater channel as a linear system



$$\begin{cases} p(t,\theta) & = h(t,\theta) * s(t) \\ y(t,\theta) & = p(t,\theta) + w(t) \end{cases} \quad (11)$$

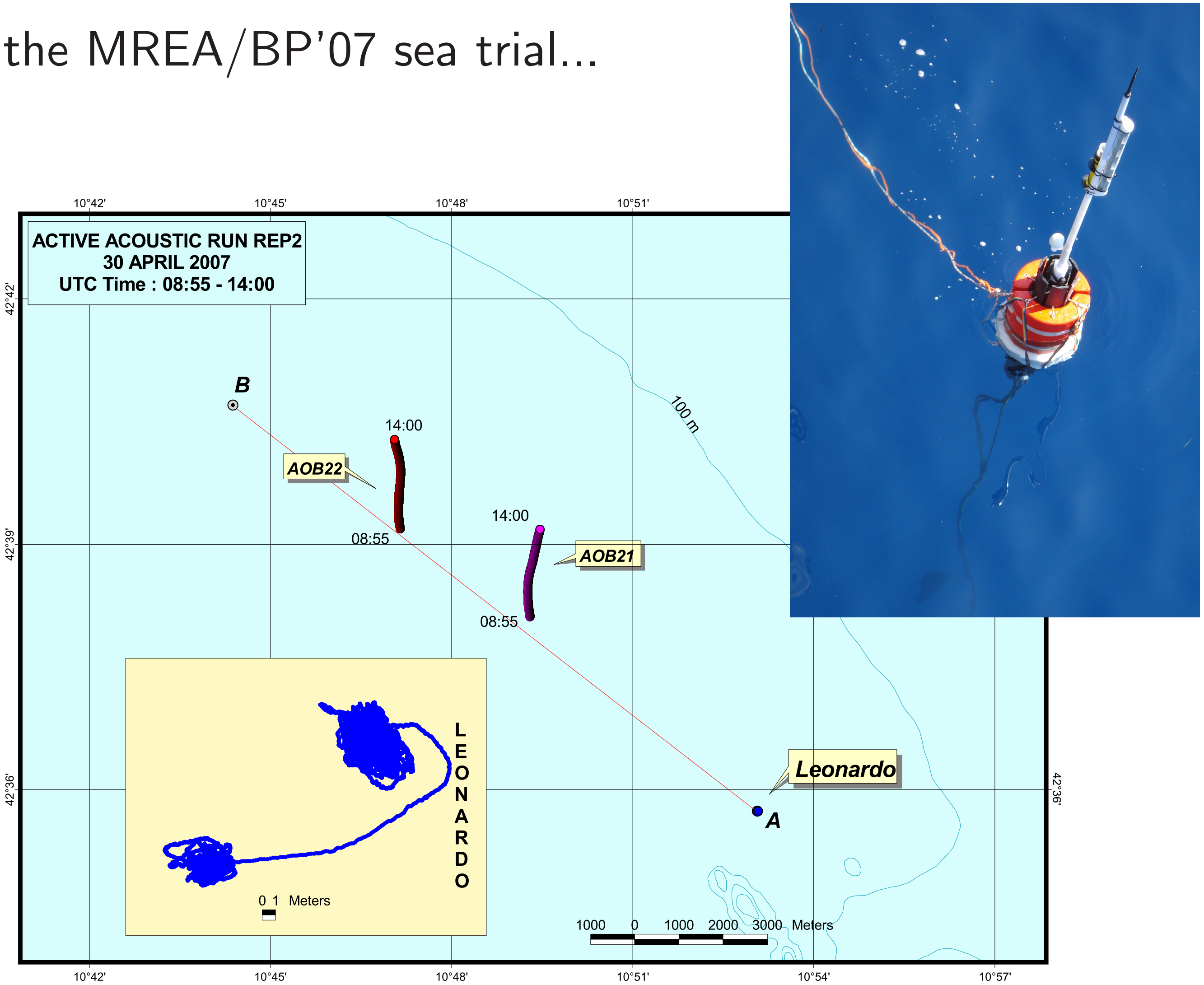
- $h(t,\theta)$: channel impulse response, θ is a vector with geometric and environmental known, unknown or partially known parameters, assumed deterministic or random

- $s(t)$: emitted signal, assumed known, unknown, random or deterministic.

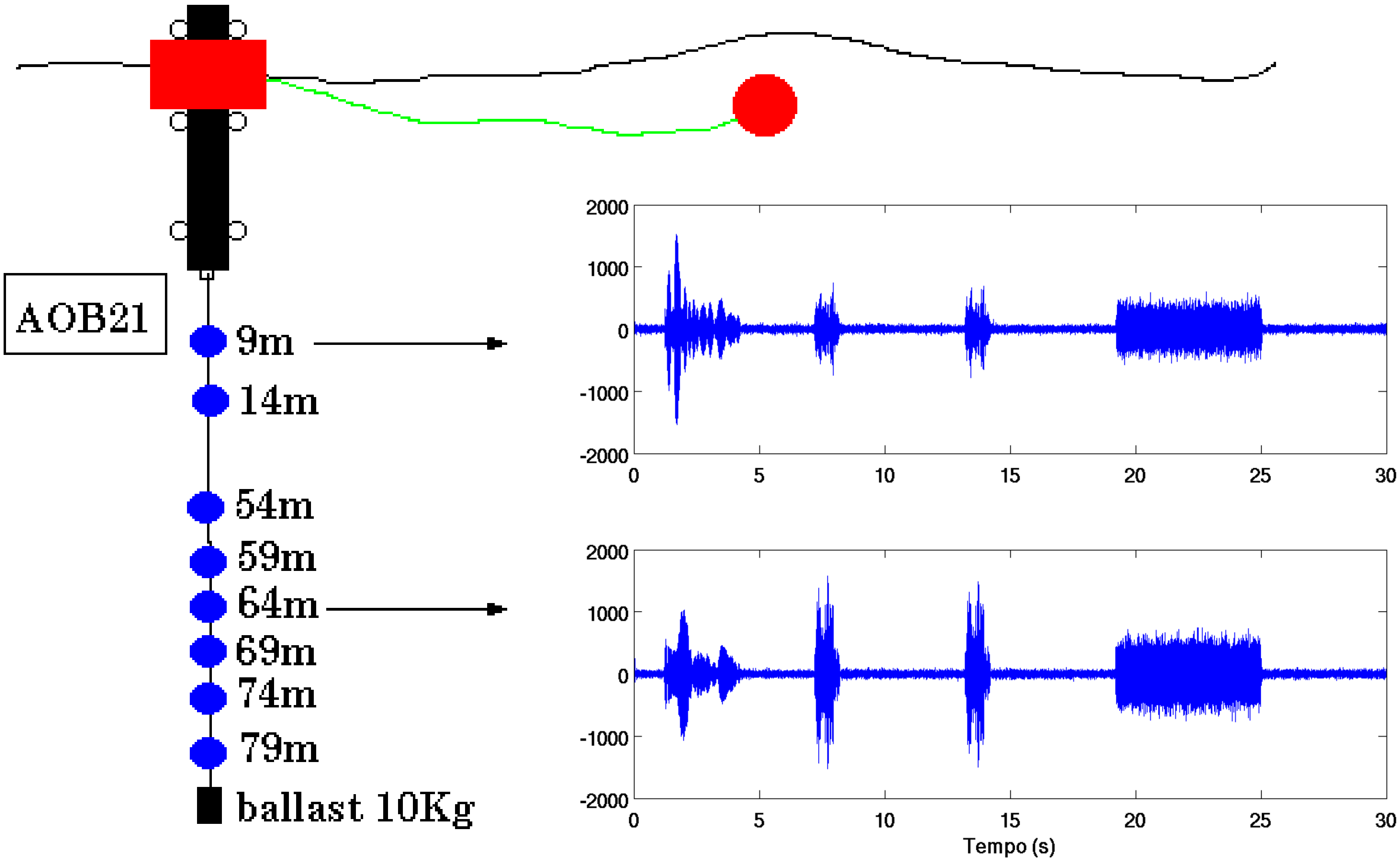
- $w(t)$: observation additive noise, non-correlated with signal, zero mean, space-time correlated or uncorrelated

The underwater channel as a space-time filter

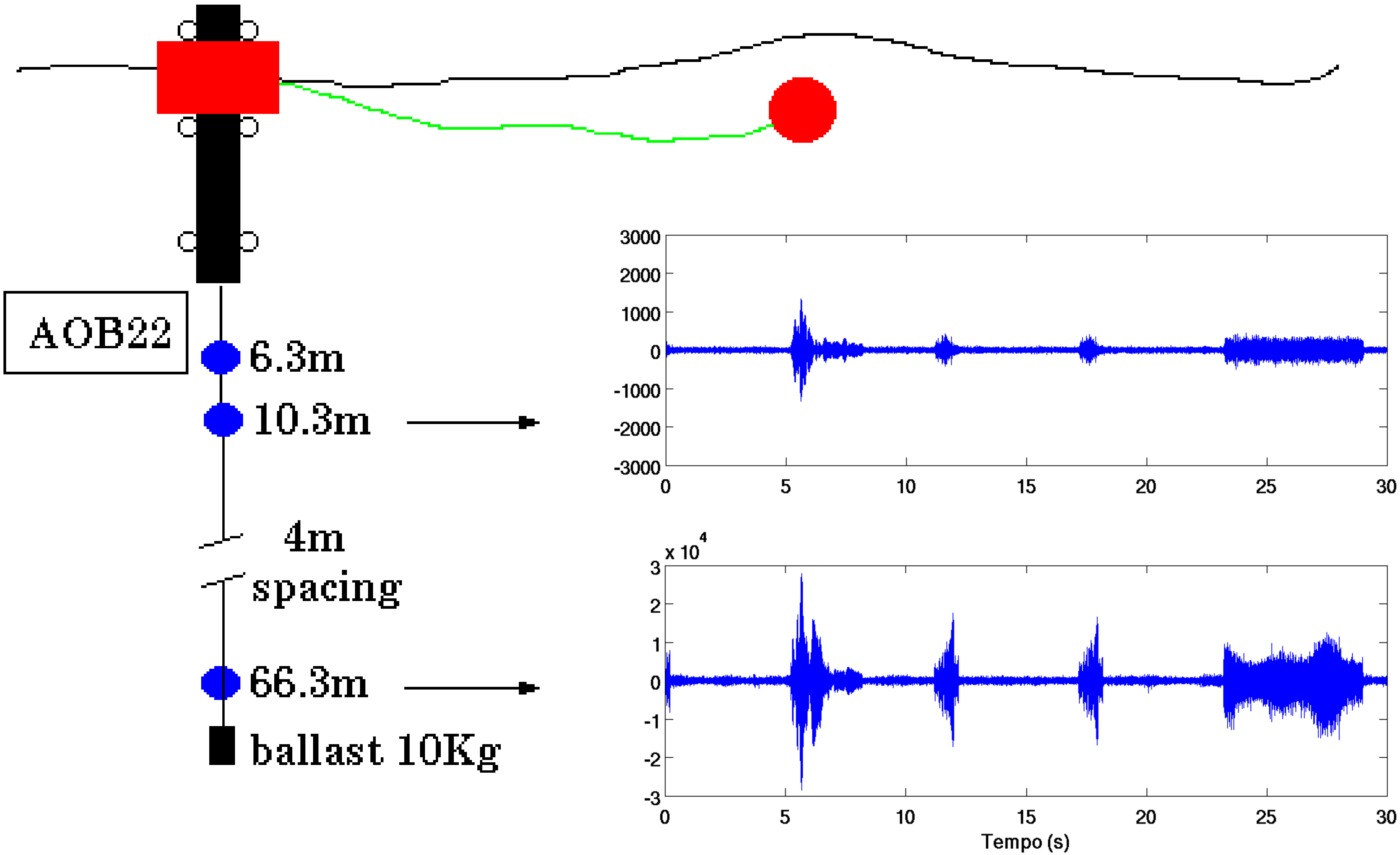
Example from the MREA/BP'07 sea trial...



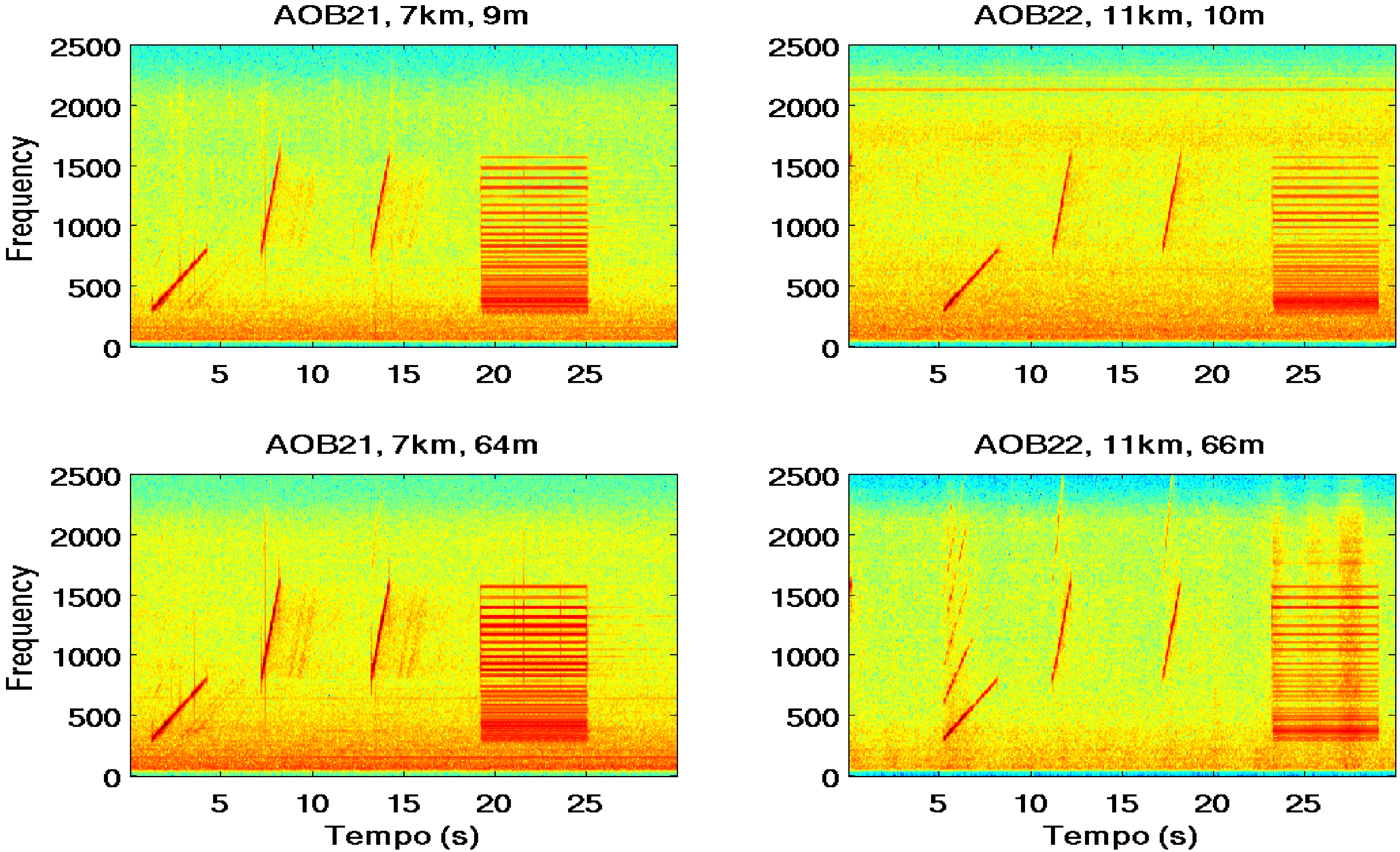
Received signals in the AO Buoy 1 @ 7km



Received signals in the AO Buoy 2 @ 11km



Comparing received signals



Data model and optimality (1)

Continuous to discrete time data model

$$\begin{cases} p(t,\theta) & = h(t,\theta) * s(t) \\ y(t,\theta) & = p(t,\theta) + w(t) \end{cases} \Rightarrow \begin{cases} \mathbf{p}(\theta) & = \mathbf{H}(\theta)\mathbf{s} \\ \mathbf{y}(\theta) & = \mathbf{p}(\theta) + \mathbf{w} \end{cases}$$ **with**

$$\mathbf{p}^t(\theta) = [p(0,\theta), p(1,\theta), \ldots, p(N-1,\theta)]$$

$$\mathbf{H}(\theta) = \begin{bmatrix} h(0,\theta) & 0 & \ldots & 0 \\ h(1,\theta) & h(0,\theta) & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ h(N-1,\theta) & h(N-2,\theta) & \ldots & h(0,\theta) \end{bmatrix} = \begin{bmatrix} \mathbf{h}^t(0,\theta) \\ \vdots \\ \mathbf{h}^t(n,\theta) \\ \vdots \\ \mathbf{h}^t(N-1,\theta) \end{bmatrix}$$

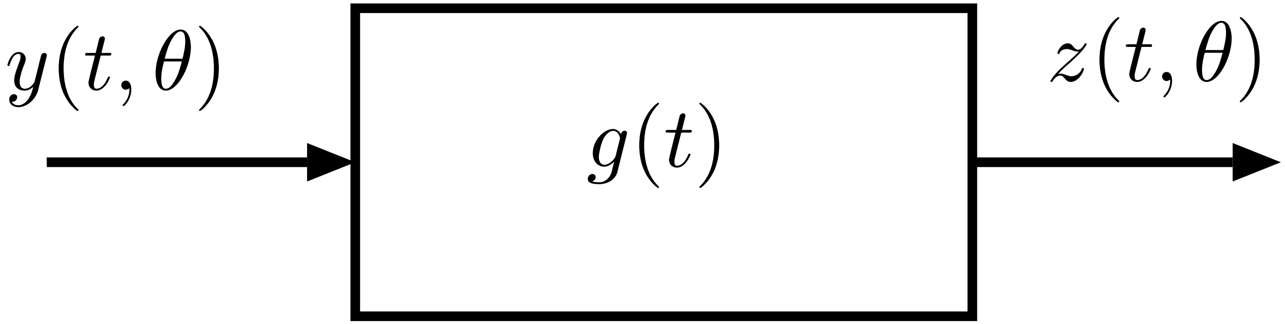
$$\mathbf{s}^t = [s(0), s(1), \ldots, s(N-1)]$$

$\mathbf{w}$ is $[\mathbf{0}, \mathbf{C}_w]$ if time correlated, $\mathbf{C}_w = \sigma^2\mathbf{I}$ if time uncorrelated.

$\mathbf{y}$ is $[\mathbf{p}(\theta), \mathbf{C}_w]$ if deterministic signal

Data model and optimality (2)

Objective: find filter $g(t)$ to optimally reduce noise on the received acoustic pressure.



$$\mathbf{y}(\theta) = \mathbf{p}(\theta) + \mathbf{w} \quad \Rightarrow \quad \mathbf{z}(\theta) = \mathbf{z}_o(\theta) + \mathbf{w}_o$$

$$\begin{aligned} \mathbf{z}_o(\theta) &= \mathbf{G}\mathbf{p}(\theta) \\ \mathbf{w}_o &= \mathbf{G}\mathbf{w} \end{aligned}$$

Generalized Matched Filter (GMF)

Signal-to-noise ratio in: $\rho_{\rm in} = \frac{|\mathbf{p}(\theta)|^2}{E[|\mathbf{w}|^2]}$

Signal-to-noise ratio out: $\rho_{\rm out} = \frac{|\mathbf{z}_o(\theta)|^2}{E[|\mathbf{w}_o|^2]}$

Instantaneous SNR out:

$$\rho(n,\theta) = \frac{|\mathbf{g}^t(n)\mathbf{p}(\theta)|^2}{E[|\mathbf{g}^t(n)\mathbf{w}|^2]}$$

$\Rightarrow$ Find $\mathbf{G}$ so as to **maximize** $\rho(n,\theta)$

GMF: the white noise case

Assuming: temporally white noise, $\mathbf{C}_w = \sigma^2 \mathbf{I}$

$$\begin{aligned}\rho(n,\theta) &= \frac{|\mathbf{g}^t(n)\mathbf{p}(\theta)|^2}{E[|\mathbf{g}^t(n)\mathbf{w}|^2]} = \frac{|\mathbf{g}^t(n)\mathbf{p}(\theta)|^2}{\mathbf{g}^t(n)E[\mathbf{w}\mathbf{w}^t]\mathbf{g}(n)} \\ &= \frac{1}{\sigma^2}\frac{|\mathbf{g}^t(n)\mathbf{p}(\theta)|^2}{\mathbf{g}^t(n)\mathbf{g}(n)}\end{aligned}$$

in virtue of the Schwartz inequality $|\mathbf{x}^t\mathbf{y}|^2 \leq |\mathbf{x}|^2|\mathbf{y}|^2$ with equality iff $\mathbf{x} = \lambda\mathbf{y}$, $\rho(n,\theta)$ will be maximum for

$$\boxed{\mathbf{g}(n) = \lambda\mathbf{p}(\theta) = \lambda\mathbf{H}(\theta)\mathbf{s}}$$

the filter is said to be matched to the incoming signal $\rightarrow$ **matched filter**

GMF: the correlated noise case (1)

Assuming: temporally correlated noise, $E[\mathbf{w}\mathbf{w}^t] = \mathbf{C}_w$, with $\mathbf{C}_w$ definite positive $\mathbf{C}_w^{-1} = \mathbf{D}^t\mathbf{D}$, with $\mathbf{D}$ non-singular, therefore

$$\mathbf{y}(\theta) \rightarrow \tilde{\mathbf{y}}(\theta) = \mathbf{D}\mathbf{y}(\theta)$$

$\tilde{\mathbf{y}}(\theta)$ is said to be a pre-whitened version of the observation vector $\mathbf{y}(\theta)$.

$$\begin{aligned}\tilde{\rho}(n,\theta) &= \frac{|\tilde{\mathbf{g}}^t(n)\tilde{\mathbf{p}}(\theta)|^2}{E[|\tilde{\mathbf{g}}^t(n)\tilde{\mathbf{w}}|^2]} = \frac{|\tilde{\mathbf{g}}^t(n)\tilde{\mathbf{p}}(\theta)|^2}{\tilde{\mathbf{g}}^t(n)\mathbf{D}E[\mathbf{w}\mathbf{w}^t]\mathbf{D}^t\tilde{\mathbf{g}}(n)} \\ &= \frac{|\tilde{\mathbf{g}}^t(n)\tilde{\mathbf{p}}(\theta)|^2}{\tilde{\mathbf{g}}^t(n)\mathbf{D}\mathbf{C}_w\mathbf{D}^t\tilde{\mathbf{g}}(n)} = \frac{|\tilde{\mathbf{g}}^t(n)\tilde{\mathbf{p}}(\theta)|^2}{\tilde{\mathbf{g}}^t(n)\tilde{\mathbf{g}}(n)}\end{aligned}$$

GMF: the correlated noise case (2)

As before $\tilde{\rho}(n,\theta)$ will be maximum for $\tilde{\mathbf{g}}(n) = \lambda\tilde{\mathbf{p}}(\theta)$, since $\mathbf{g}(n) = \mathbf{D}^t\tilde{\mathbf{g}}(n)$ we have

$$\begin{aligned} \tilde{\mathbf{g}}(n) &= \lambda\tilde{\mathbf{p}}(\theta) = \lambda\mathbf{D}\mathbf{p}(\theta) \\ \mathbf{g}(n) &= \lambda\mathbf{D}^t\mathbf{D}\mathbf{p}(\theta) = \lambda\mathbf{C}_w^{-1}\mathbf{p}(\theta) \end{aligned}$$

in practice the GMF is a time-reversed replica of the signal

$$\tilde{g}(N-1-n,\theta) = \lambda\tilde{p}(n,\theta), \qquad n = 0,\ldots,N-1$$

$$\tilde{g}(n,\theta) = \lambda\tilde{p}(N-1-n,\theta), \qquad n = 0,\ldots,N-1$$

$\Rightarrow$ <u>filter by a time-reversed signal = correlate with that signal.</u>

GMF: the multichannel (spatial) version (1)

Let us assume K sensors, $\rightarrow KN$ sample augmented vector

$$\mathbf{y}_a^t(\theta) = [\mathbf{y}^t(0,\theta), \mathbf{y}^t(1,\theta), \ldots, \mathbf{y}^t(N-1,\theta)]$$

with $\mathbf{y}(n,\theta)$ a K-dimensional vector with the K sensor entries at time n,

$$\mathbf{y}^t(n,\theta) = [y_1(n,\theta), y_2(n,\theta), \ldots, y_K(n,\theta)]$$

is the **temporal-ordering**. Similarly the **spatial ordering** is

$$\mathbf{y}_a^t(\theta) = [\mathbf{y}_1^t(\theta), \mathbf{y}_2^t(\theta), \ldots, \mathbf{y}_K^t(\theta)]$$

with

$$\mathbf{y}_k^t(\theta) = [y_k(0,\theta), y_k(1,\theta), \ldots, y_k(N-1,\theta)].$$

GMF: the multichannel (spatial) version (2)

The augmented data model (spatially-ordered)

$$\mathbf{y}_a(\theta) = \mathbf{H}_a(\theta)\mathbf{s} + \mathbf{w}_a$$

with

$$\mathbf{H}_a(\theta) = [\mathbf{H}_1(\theta)|\mathbf{H}_2(\theta)|\ldots|\mathbf{H}_K(\theta)]^t$$

and the appropriate notation for $\mathbf{w}_a$, allows for the multichannel GMF,

$$\begin{aligned}\mathbf{g}_a(n) &= \lambda\mathbf{p}_a(\theta)\\ &= \lambda\mathbf{H}_a(\theta)\mathbf{s}\end{aligned}$$

$$z(n,\theta) = \sum_{n'=0}^{n-1}\mathbf{g}^t(n')\mathbf{y}(n',\theta) = \sum_{k=1}^{K}\mathbf{g}_k^t(n)\mathbf{y}_k(\theta)$$

Bibliography

1. W.B. Davenport Jr. and W.L. Root, "An Introduction to the Theory of Random Signals and Noise", IEEE Press, New York, 1987.
2. L. Scharf, "Statistica Signal Processing: Detection, Estimation and Time Series Analysis", Addison-Wesley, New York, 1991.
3. S.M. Kay, "Fundamentals of Statistical Signal Processing: Detection Theory", Vol.2, Prentice-Hall, New Jersey (USA), 1998.
4. S.M. Jesus, C. Martins and F. Zabel, "Maritime Rapid Environmental Assessment / Blue Planet'07: Acoustic Oceanographic Buoys Data Report", Rep. 03/07, SiPLAB Report, University of Algarve, August 2007.
5. F.B. Jensen, W.A. Kuperman, M.B. Porter and H. Schmidt, "Computational Ocean Acoustics", AIP, New York (USA), 1993.