



**HAL**  
open science

## Array Processing

Julien Lesouple

► **To cite this version:**

| Julien Lesouple. Array Processing. Master. France. 2024. hal-04794791

**HAL Id: hal-04794791**

**<https://enac.hal.science/hal-04794791v1>**

Submitted on 21 Nov 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Array Processing

## IENAC22 SAT - ASNAT23

Julien Lesouple (TELECOM/SIGNAV)

[julien.lesouple@enac.fr](mailto:julien.lesouple@enac.fr)



## Class organization

- 10 classes: lectures mixed with Python coding
- Matlab codes will be available on e-campus to replicate figures (and play with it)
- one hour exam, A4 Sheets both sides handwritten, not photocopied

I would very appreciate any feedback on the class

# Plan

- ① Introduction
- ② Model
- ③ Beamforming
- ④ Direction of arrival estimation
- ⑤ Conclusion

# Plan

## ① Introduction

Multichannel processing  
Array processing

## ② Model

## ③ Beamforming

## ④ Direction of arrival estimation

## ⑤ Conclusion

# Plan

## ① Introduction

Multichannel processing

Array processing

## ② Model

## ③ Beamforming

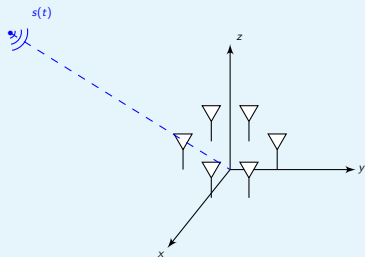
## ④ Direction of arrival estimation

## ⑤ Conclusion

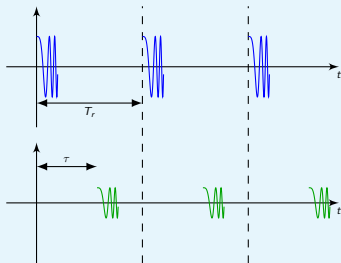
# Multichannel processing

Observations are now stored in a measurement vector:  $\mathbf{y}(\cdot) = \begin{bmatrix} y_1(\cdot) \\ y_2(\cdot) \\ \vdots \\ y_N(\cdot) \end{bmatrix}$

## 1 sample on $N$ sensors



## $N$ samples on 1 sensor



# Analysis/processing of multichannel measurements

- Analysis: what is the relation between  $y_n(\cdot)$  and  $y_m(\cdot)$ ?
  - how correlated are  $y_n(\cdot)$  and  $y_m(\cdot)$ ?
  - can  $\mathbf{y}(\cdot)$  be explained by a small number of variables?
- Processing: combination of the signals, e.g., linear filtering

$$y_F(\cdot) = \sum_{n=1}^N w_n^* y_n(\cdot)$$

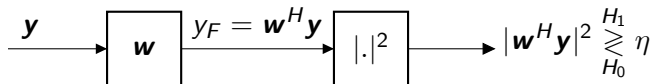


## Example: linear filtering for detection

We want to detect if a known signal  $\mathbf{a}$  is in a received signal  $\mathbf{y}$  with addition of noise  $\mathbf{n}$ , i.e., we have the hypothesis testing

$$\begin{cases} H_0 : \mathbf{y} = \mathbf{n} \\ H_1 : \mathbf{y} = \alpha \mathbf{a} + \mathbf{n} \end{cases}$$

Linear filter:  $\mathbf{w}$  such as  $y_F = \mathbf{w}^H \mathbf{y} \rightarrow \text{energy } |\mathbf{w}^H \mathbf{y}|^2$



For instance, for pure tone  $a_n = e^{j2\pi f_0 n}$ , we can take  $w_n = e^{j2\pi f_0 n}$  leading to the Fourier transform of  $\mathbf{y}$  at  $f_0$

# Plan

## ① Introduction

Multichannel processing  
Array processing

## ② Model

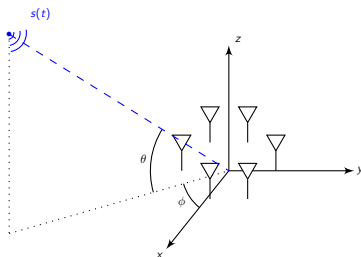
## ③ Beamforming

## ④ Direction of arrival estimation

## ⑤ Conclusion

## Context of the course

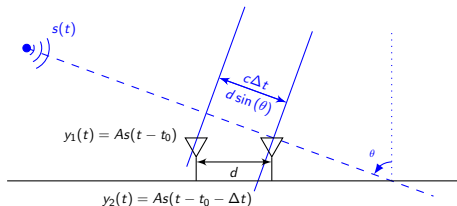
- Signals are received on an array of  $N$  antennas



- $y_n(k)$  represents the signal received on **antenna  $n$**  at **time  $k$**
- $n = 1, \dots, N$  is a spatial index (related to a **position** in space)  $\rightarrow$  spatial processing (e.g.,  $\sum_{n=1}^N w_n^* y_n(\cdot)$ )

# Principle of array processing

Considering two antennas and a source in the far field



## Principle

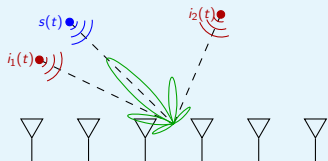
Additional delay  $\Delta t$  only depends on the geometry and the angle of arrival

- if  $\theta$  is known: spatial filtering (beamforming)
- if  $\theta$  is unknown: estimation of  $\Delta t$  (source localization)

# Beamforming vs. DoA estimation

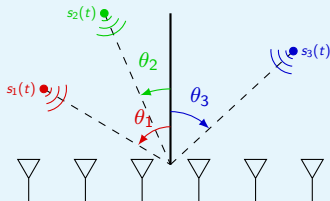
## Beamforming

Retrieve signal from direction of interest, possibly eliminate interference



## DoA estimation

Estimate directions of arrival of incoming signals



# Array of sensors: conclusion

## Possibilities

Sensor array offer an additional dimension (**space**) to perform filtering of signals (possibly coupled with temporal or frequency filtering)

- 1 source separation
- 2 direction finding

## Applications

- 1 Radar, sonar (detection, localization, anti-jamming)
- 2 Communications (capacity improvements, better reception, spatial focusing, interference mitigation)

# Plan

## ① Introduction

## ② Model

- Principle

- Multichannel receiver

- Received signals

- The steering vector

- Covariance matrix

- Limitations

## ③ Beamforming

## ④ Direction of arrival estimation

## ⑤ Conclusion

# Plan

## ① Introduction

## ② Model

### Principle

Multichannel receiver

Received signals

The steering vector

Covariance matrix

Limitations

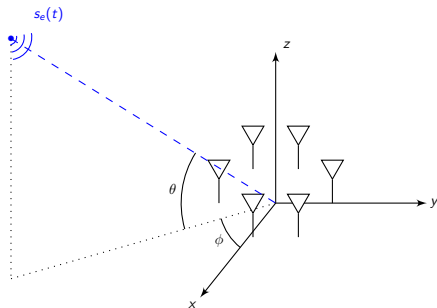
## ③ Beamforming

## ④ Direction of arrival estimation

## ⑤ Conclusion



# Arrays and waveforms



- Wavefront impinging from direction  $(\theta, \phi)$
- Assumptions: homogeneous propagation medium, source in the far field of the array  $\rightarrow$  plane wavefront

# Plan

## ① Introduction

## ② Model

Principle

**Multichannel receiver**

Received signals

The steering vector

Covariance matrix

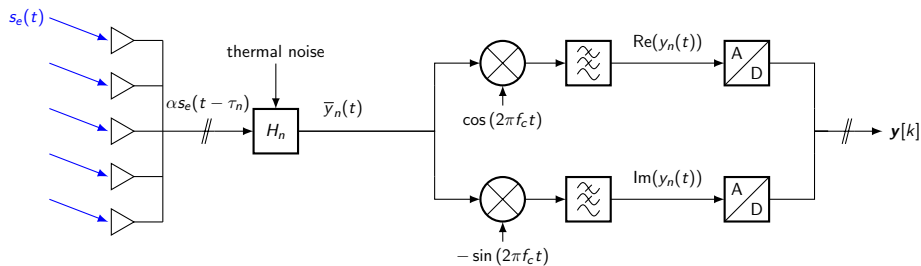
Limitations

## ③ Beamforming

## ④ Direction of arrival estimation

## ⑤ Conclusion

# Multichannel receiver



# Plan

## ① Introduction

## ② Model

Principle

Multichannel receiver

**Received signals**

The steering vector

Covariance matrix

Limitations

## ③ Beamforming

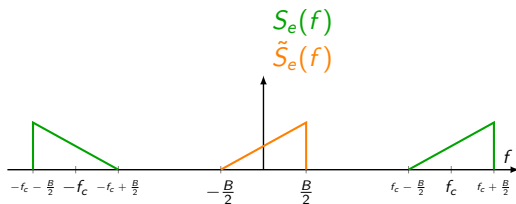
## ④ Direction of arrival estimation

## ⑤ Conclusion

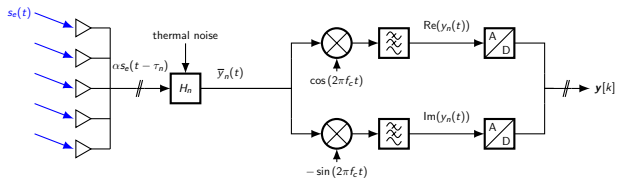
# Emitted signal

Narrowband signal :  $B \ll f_c$

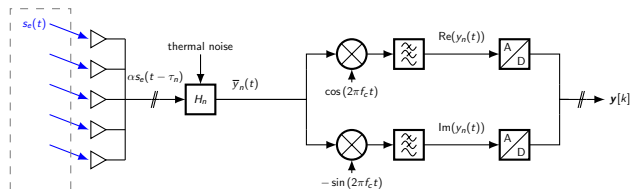
$$s_e(t) = 2\text{Re} \left\{ \tilde{s}_e(t) e^{j2\pi f_c t} \right\}$$



# Received signals

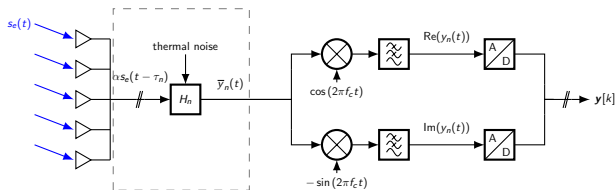


# Received signals



- Narrow band:  $B \ll f_c \rightarrow s_e(t) = 2\text{Re} \{ \tilde{s}_e(t) e^{j2\pi f_c t} \}$

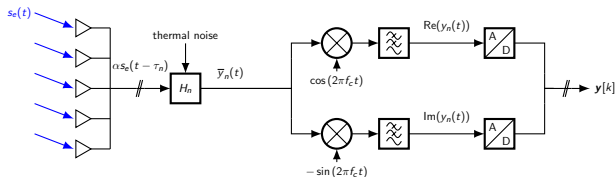
# Received signals



- Narrow band:  $B \ll f_c \rightarrow s_e(t) = 2\text{Re} \{ \tilde{s}_e(t) e^{j2\pi f_c t} \}$
- Impulse response  $h_n$ :  $\bar{y}_n(t) = h_n * (\alpha s_e(t - \tau_n)) + \check{e}_n(t)$

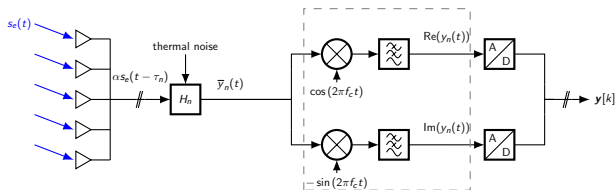


# Received signals



- Narrow band:  $B \ll f_c \rightarrow s_e(t) = 2\text{Re} \{ \tilde{s}_e(t) e^{j2\pi f_c t} \}$
- Impulse response  $h_n$ :  $\bar{y}_n(t) = h_n * (\alpha s_e(t - \tau_n)) + \check{\epsilon}_n(t)$
- Fourier:  $\bar{Y}_n(f) = \alpha H_n(f) S_e(f) e^{-j2\pi f \tau_n} + \check{E}_n(f)$

# Received signals



- Narrow band:  $B \ll f_c \rightarrow s_e(t) = 2\text{Re} \{ \tilde{s}_e(t) e^{j2\pi f_c t} \}$
- Impulse response  $h_n$ :  $\bar{y}_n(t) = h_n * (\alpha s_e(t - \tau_n)) + \check{\epsilon}_n(t)$
- Fourier:  $\bar{Y}_n(f) = \alpha H_n(f) S_e(f) e^{-j2\pi f \tau_n} + \check{E}_n(f)$
- Baseband:  $Y_n(f) = \alpha H_n(f + f_c) \tilde{S}_e(f) e^{-j2\pi(f+f_c)\tau_n} + \check{E}_n(f + f_c)$

## Received signals

Relatively to a given reference ( $\tau_n = \tau_0 + \Delta\tau_n$ )

$$\begin{aligned} Y_n(f) &= \alpha H_n(f + f_c) \tilde{S}_e(f) e^{-j2\pi(f+f_c)\tau_n} + \check{E}_n(f + f_c) \\ &= \alpha H_n(f + f_c) \tilde{S}_e(f) e^{-j2\pi(f+f_c)(\tau_0 + \Delta\tau_n)} + \check{E}_n(f + f_c) \\ &= \alpha H_n(f + f_c) \tilde{S}_e(f) e^{-j2\pi f \tau_0} e^{-j2\pi f_c \tau_0} e^{-j2\pi(f+f_c)\Delta\tau_n} + \check{E}_n(f + f_c) \end{aligned}$$

Short delays and narrowband ( $e^{-j2\pi(f+f_c)\Delta\tau_n} \approx e^{-j2\pi f_c \Delta\tau_n}$ )

$$Y_n(f) = \alpha H_n(f + f_c) \tilde{S}_e(f) e^{-j2\pi f \tau_0} e^{-j2\pi f_c \tau_0} e^{-j2\pi f_c \Delta\tau_n} + \check{E}_n(f + f_c)$$

Narrowband ( $H_n(f + f_c) \tilde{S}_e(f) \approx H_n(f_c) \tilde{S}_e(f)$ )

$$Y_n(f) \approx \alpha e^{-j2\pi f_c \tau_0} H_n(f_c) \tilde{S}_e(f) e^{-j2\pi f \tau_0} e^{-j2\pi f_c \Delta\tau_n} + \check{E}_n(f + f_c)$$

Back to temporal (inverse Fourier)

$$y_n(t) \approx \alpha e^{-j2\pi f_c \tau_0} H_n(f_c) \tilde{s}_e(t - \tau_0) e^{-j2\pi f_c \Delta\tau_n} + e_n(t)$$

## Received signals

Calibrated sensors  $H_n(f_c) = \dots = H_N(f_c)$  and gathering constants in  $\alpha$

$$y_n(t) = \alpha \tilde{s}_e(t - \tau_0) e^{-j2\pi f_c \Delta \tau_n} + e_n(t)$$

### Narrowband signal

A short time shift is equivalent to a phase shift

$$s(t - \Delta \tau_n) \Leftrightarrow s(t) e^{-j2\pi f_c \Delta \tau_n}$$

Main signal on reference point ( $s(t) = \alpha \tilde{s}_e(t - \tau_0)$ )

$$y_n(t) = s(t) e^{-j2\pi f_c \Delta \tau_n} + e_n(t)$$

In the following:  $\Delta \tau_n \leftrightarrow \tau_n$

# Snapshots

Individual signals

$$y_n(t) = s(t)e^{-j2\pi f_c \tau_n} + e_n(t)$$

Output of multichannel receiver

$$\mathbf{y}(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_N(t) \end{bmatrix} = \begin{bmatrix} e^{-j2\pi f_c \tau_1} \\ e^{-j2\pi f_c \tau_2} \\ \vdots \\ e^{-j2\pi f_c \tau_N} \end{bmatrix} s(t) + \begin{bmatrix} e_1(t) \\ e_2(t) \\ \vdots \\ e_N(t) \end{bmatrix}$$

Far field (planar waves):  $\tau_n$  depends only on source location  $(\theta, \phi)$

Snapshot before ADC

$$\mathbf{y}(t) = \mathbf{a}(\theta, \phi)s(t) + \mathbf{e}(t)$$

# Snapshots

Temporal sampling at rate  $T_s$  leads a matrix

$$\mathbf{Y} = [\mathbf{y}[1] \quad \mathbf{y}[2] \quad \cdots \quad \mathbf{y}[K]] \in \mathbb{C}^{N \times K}$$

where

$$\mathbf{y}[k] = \mathbf{y}(kT_s) = \mathbf{a}(\theta, \phi)s[k] + \mathbf{e}[k]$$

## Vector of phase shifts

$\mathbf{a}(\theta, \phi)$  is referred to as the **steering vector** and depends only on the DOA of the source

## In presence of several sources

Snapshot received in the presence of  $P$  sources

$$\begin{aligned}
 \mathbf{y}[k] &= \sum_{p=1}^P \mathbf{y}_p[k] + \mathbf{e}[k] \\
 &= \sum_{p=1}^P \mathbf{a}(\theta_p, \phi_p) s_p[k] + \mathbf{e}[k] \\
 &= [\mathbf{a}(\theta_1, \phi_1) \quad \cdots \quad \mathbf{a}(\theta_P, \phi_P)] \begin{bmatrix} s_1[k] \\ \vdots \\ s_P[k] \end{bmatrix} + \mathbf{e}[k] \\
 &= \mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\phi}) \mathbf{s}[k] + \mathbf{e}[k] \text{ where } \mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\phi}) \in \mathbb{C}^{N \times P} \text{ and } \mathbf{s}[k] \in \mathbb{C}^P
 \end{aligned}$$

In the following:  $\mathbf{a}(\theta_p)$ ,  $\mathbf{A}(\boldsymbol{\theta})$  for clarity

# Plan

## ① Introduction

## ② Model

Principle

Multichannel receiver

Received signals

**The steering vector**

Covariance matrix

Limitations

## ③ Beamforming

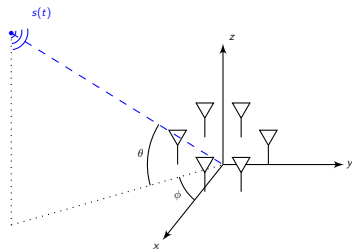
## ④ Direction of arrival estimation

## ⑤ Conclusion



# Steering vector in 3D

Reference at frame origin



Delay

$$\tau_n = \frac{1}{c} (x_n \cos \theta \cos \phi + y_n \cos \theta \sin \phi + z_n \sin \theta)$$

Steering vector

$$\mathbf{a}_n(\theta, \phi) = e^{-j\frac{2\pi}{\lambda} (x_n \cos \theta \cos \phi + y_n \cos \theta \sin \phi + z_n \sin \theta)}$$

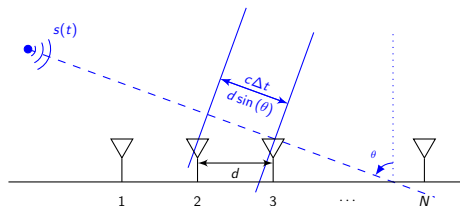
Received signal

$$s_n(t) = s(t) e^{-j\mathbf{r}_n^T \mathbf{k}} = s(t) e^{-j2\pi f_c \tau_n}$$

where

$$\mathbf{r}_n = \begin{bmatrix} x_n \\ y_n \\ z_n \end{bmatrix} \quad \mathbf{k} = \frac{2\pi}{\lambda} \begin{bmatrix} \cos \theta \cos \phi \\ \cos \theta \sin \phi \\ \sin \theta \end{bmatrix}$$

## Simpler 2D case: Uniform Linear Array (ULA)



$$\theta \in \left[ -\frac{\pi}{2}, \frac{\pi}{2} \right]$$

$$\tau_n = (n-1) \frac{d \sin \theta}{c} \Rightarrow \mathbf{a}(\theta) = \left[ 1 \quad e^{-j2\pi f_c \frac{d \sin \theta}{c}} \quad \dots \quad e^{-j2\pi f_c (N-1) \frac{d \sin \theta}{c}} \right]$$

### Spatial sampling frequency

$$f_s = f_c \frac{d \sin \theta}{c} = \frac{d}{\lambda} \sin \theta \Rightarrow \mathbf{a}(\theta) = \left[ 1 \quad e^{-j2\pi f_s} \quad \dots \quad e^{-j2\pi (N-1) f_s} \right]$$

## Simpler 2D case: Uniform Linear Array (ULA)

The phase shift is

$$\Delta\phi = 2\pi f_c \frac{d \sin \theta}{c}$$

Unambiguous if  $\Delta\phi \in [-\pi, \pi]$  for all  $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$

$$-\pi \leq -2\pi f_c \frac{d}{c} \leq \Delta\phi \leq 2\pi f_c \frac{d}{c} \leq \pi$$

Spatial sampling requirement

$$d \leq \frac{\lambda}{2}$$

# Plan

## ① Introduction

## ② Model

Principle

Multichannel receiver

Received signals

The steering vector

**Covariance matrix**

Limitations

## ③ Beamforming

## ④ Direction of arrival estimation

## ⑤ Conclusion

## Autocorrelation matrix of WSS signals

The autocorrelation matrix of a snapshot is defined as

$$\begin{aligned} \mathbf{R} &= \mathbb{E} \left[ \mathbf{y}[k] \mathbf{y}[k]^H \right] \\ &= \mathbb{E} \left[ \begin{bmatrix} y_1[k] \\ y_2[k] \\ \vdots \\ y_N[k] \end{bmatrix} \begin{bmatrix} y_1^*[k] & y_2^*[k] & \dots & y_N^*[k] \end{bmatrix} \right] \end{aligned}$$

The  $(n, l)$  entry of matrix  $\mathbf{R}$  measures the correlation between signals received at sensors  $n$  and  $l$  at the same time index

## Covariance matrix of WSS signals

The covariance matrix of a snapshot is defined as

$$\begin{aligned}\Sigma &= \mathbb{E} \left[ (\mathbf{y}[k] - \mathbb{E}[\mathbf{y}[k]])(\mathbf{y}[k] - \mathbb{E}[\mathbf{y}[k]])^H \right] \\ &= \mathbb{E} \left[ \mathbf{y}[k]\mathbf{y}[k]^H \right] - \mathbb{E}[\mathbf{y}[k]] \mathbb{E}[\mathbf{y}[k]]^H\end{aligned}$$

### Covariance matrix

When signals are zero-mean, the autocorrelation matrix is the same as the covariance matrix

In the following, if not specified, signals are assumed zero-mean

## Structure of the signal covariance matrix

For a mixture of signals

$$\begin{aligned}\mathbb{E} \left[ \mathbf{A}(\boldsymbol{\theta}) \mathbf{s}[k] \mathbf{s}[k]^H \mathbf{A}(\boldsymbol{\theta})^H \right] &= \mathbf{A}(\boldsymbol{\theta}) \mathbb{E} \left[ \mathbf{s}[k] \mathbf{s}[k]^H \right] \mathbf{A}(\boldsymbol{\theta})^H \\ &= \mathbf{A}(\boldsymbol{\theta}) \mathbf{R}_s \mathbf{A}(\boldsymbol{\theta})^H\end{aligned}$$

with  $\mathbf{R}_s \in \mathbb{C}^{P \times P}$  the covariance matrix of incoming signals

Uncorrelated signals:  $\mathbf{R}_s = \text{diag} (P_1, \dots, P_P)$

$$\mathbb{E} \left[ \mathbf{A}(\boldsymbol{\theta}) \mathbf{s}[k] \mathbf{s}[k]^H \mathbf{A}(\boldsymbol{\theta})^H \right] = \sum_{p=1}^P P_p \mathbf{a}(\theta_p) \mathbf{a}^H(\theta_p)$$

## Signal covariance matrix range

The range of a matrix  $\mathbf{A} = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_n] \in \mathbb{C}^{m \times n}$  is the space spanned by its columns

$$\mathcal{R}(\mathbf{A}) = \left\{ \mathbf{y} = \mathbf{A}\mathbf{x} = \sum_{i=1}^n x_i \mathbf{a}_i, \forall \mathbf{x} \in \mathbb{R}^n \right\}$$

### Rank of a matrix

The rank of a matrix is the dimension of its range. A matrix  $\mathbf{A} \in \mathbb{C}^{m \times n}$  is said to be *full rank* if its rank is  $\min(m, n)$

If  $\mathbf{A}(\theta)$  and  $\mathbf{R}_s$  are full rank, the signal covariance matrix has rank  $P$ , and its range is spanned by the steering vectors  $\mathbf{a}(\theta_p), p = 1, \dots, P$



## Structure of the signal covariance matrix

WSS noise

$$\mathbb{E}[\mathbf{e}[k]] = \mathbf{0}, \quad \mathbb{E}[\mathbf{e}[k]\mathbf{e}[k]^H] \text{ indep. of } k$$

Spatially white noise with same power on each channel

$$\mathbb{E}[\mathbf{e}[k]\mathbf{e}[k]^H] = \sigma^2 \mathbf{I}_N$$

Leading for uncorrelated sources (noise and sources independant)

$$\mathbf{R} = \sum_{p=1}^P P_p \mathbf{a}(\theta_p) \mathbf{a}^H(\theta_p) + \sigma^2 \mathbf{I}_N$$

# Plan

## ① Introduction

## ② Model

Principle

Multichannel receiver

Received signals

The steering vector

Covariance matrix

**Limitations**

## ③ Beamforming

## ④ Direction of arrival estimation

## ⑤ Conclusion

## Model limitations

The model  $\mathbf{y}[k] = \mathbf{a}(\theta)s[k] + \mathbf{e}[k]$  is an idealized model which doesn't take into account

- non homogeneous propagation medium

$$y_n[k] = \rho_n[k]e^{j\varphi_n[k]}a_n(\theta)s[k] + e_n[k]$$

- uncalibrated arrays

$$y_n[k] = \alpha H_n(f_c)a_n(\theta)s[k] + e_n[k]$$

- wideband signals (time delay  $\neq$  phase shift)

$$Y_n(f) = \underbrace{\tilde{S}(f)e^{-j2\pi(f+f_c)\tau_n}}_{A_n(f,\theta)} + \check{E}(f + f_c)$$

- colored noise

$$\mathbb{E} \left[ \mathbf{n}[k]\mathbf{n}[k]^H \right] \neq \sigma^2 \mathbf{I}_N$$

# Plan

## ① Introduction

## ② Model

## ③ Beamforming

- Principle

- Array beampattern

- Spatial filtering

- Adaptive beamforming

- Robust adaptive beamforming

- Partially adaptive beamforming

- Summary

## ④ Direction of arrival estimation

## ⑤ Conclusion

# Plan

## ① Introduction

## ② Model

## ③ Beamforming

### Principle

Array beampattern

Spatial filtering

Adaptive beamforming

Robust adaptive beamforming

Partially adaptive beamforming

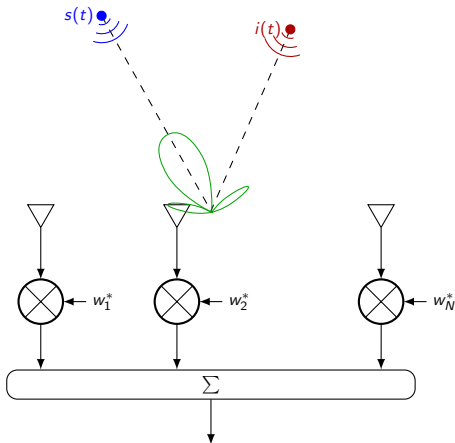
Summary

## ④ Direction of arrival estimation

## ⑤ Conclusion

# Spatial filtering

Principle: linear combination of each sensor to point in a direction



$$y_F[k] = \sum_{n=1}^N w_n^* y_n[k] \simeq \alpha s[k]$$

# Plan

## ① Introduction

## ② Model

## ③ Beamforming

Principle

**Array beampattern**

Spatial filtering

Adaptive beamforming

Robust adaptive beamforming

Partially adaptive beamforming

Summary

## ④ Direction of arrival estimation

## ⑤ Conclusion

# Array beampattern

For a given  $\mathbf{w}$

$$y_F[k] = \sum_{n=1}^N w_n^* y_n[k] = \mathbf{w}^H \mathbf{y}[k] = \mathbf{w}^H \mathbf{a}(\theta) s[k] + \mathbf{w}^H \mathbf{e}[k]$$

Corresponding array beampattern (gain at direction  $\theta$ )

$$G_{\mathbf{w}}(\theta) = |g_{\mathbf{w}}(\theta)|^2 = |\mathbf{w}^H \mathbf{a}(\theta)|^2$$

For ULA, the natural beampattern ( $w_n = 1$ ) is

$$g(\theta) = \sum_{n=0}^{N-1} e^{-j2\pi n \frac{d \sin \theta}{\lambda}} = \frac{1 - e^{-j2\pi N \frac{d \sin \theta}{\lambda}}}{1 - e^{-j2\pi \frac{d \sin \theta}{\lambda}}}$$



## ULA natural beampattern

$$\begin{aligned}
 g(\theta) &= \frac{e^{-j\pi N \frac{d \sin \theta}{\lambda}}}{e^{-j\pi \frac{d \sin \theta}{\lambda}}} \frac{e^{j\pi N \frac{d \sin \theta}{\lambda}} - e^{-j\pi N \frac{d \sin \theta}{\lambda}}}{e^{j\pi \frac{d \sin \theta}{\lambda}} - e^{-j\pi \frac{d \sin \theta}{\lambda}}} \\
 &= e^{-j\pi(N-1) \frac{d \sin \theta}{\lambda}} \frac{\sin\left(\pi N \frac{d \sin \theta}{\lambda}\right)}{\sin\left(\pi \frac{d \sin \theta}{\lambda}\right)}
 \end{aligned}$$

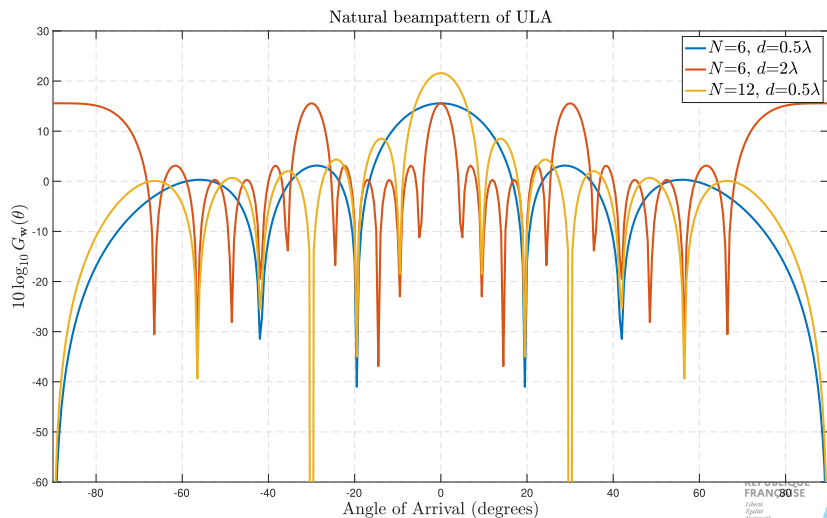
leading

$$G(\theta) = |g(\theta)|^2 = \left| \frac{\sin\left(\pi N \frac{d \sin \theta}{\lambda}\right)}{\sin\left(\pi \frac{d \sin \theta}{\lambda}\right)} \right|^2$$

Property (main lobe width)

$$\theta_{3\text{dB}} \approx \frac{0.9\lambda}{Nd}$$

# ULA natural beampattern



# Plan

## ① Introduction

## ② Model

## ③ Beamforming

Principle

Array beampattern

**Spatial filtering**

Adaptive beamforming

Robust adaptive beamforming

Partially adaptive beamforming

Summary

## ④ Direction of arrival estimation

## ⑤ Conclusion

# Beamforming

Objective: focus on a given direction to enhance reception of some signals, and possibly mitigate signals from fixed directions

## Principle

Each sensor output is weighted by  $w_n^*$  before summation

$$y_F[k] = \sum_{n=1}^N w_n^* y_n[k] = \mathbf{w}^H \mathbf{y}[k]$$

Problem: how to choose  $\mathbf{w}$  when  $\mathbf{y}[k] = \mathbf{a}(\theta_s)s[k] + \dots$  so that  $y_F[k] \approx \alpha s[k]$ ?

# Conventional beamformer

Notation:  $\mathbf{a}_s = \mathbf{a}(\theta_s)$  with  $\theta_s$  the DOA of signal of interest

## Principle

Compensate for the phase shift induced by propagation from  $\theta_s$  and sum coherently

For ULA, considering only the signal of interest

$$\begin{aligned}y_F[k] &= \mathbf{a}_s^H \mathbf{a}_s s[k] \\ &= \sum_{n=0}^{N-1} e^{+j2\pi n \frac{d \sin \theta_s}{\lambda}} e^{-j2\pi n \frac{d \sin \theta_s}{\lambda}} s[k] \\ &= Ns[k]\end{aligned}$$

# Conventional beamformer

## Beampattern

$$g(\theta) = \mathbf{a}_s^H \mathbf{a}(\theta) = \sum_{n=0}^{N-1} e^{j2\pi n \frac{d(\sin \theta_s - \sin \theta)}{\lambda}}$$

## leading

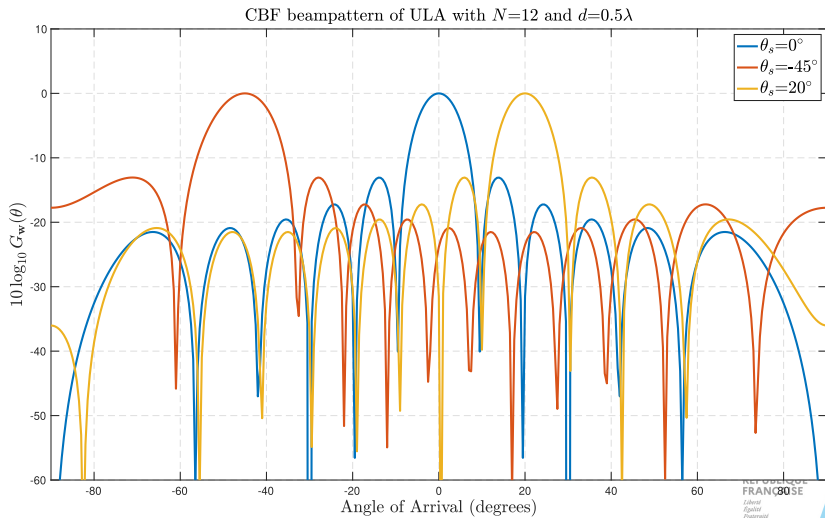
$$G(\theta) = |g(\theta)|^2 = \left| \frac{\sin \left( \pi N \frac{d(\sin \theta_s - \sin \theta)}{\lambda} \right)}{\sin \left( \pi \frac{d(\sin \theta_s - \sin \theta)}{\lambda} \right)} \right|^2$$

Property: max for  $\theta = \theta_s$

## Conventional beamformer

$$\mathbf{w}_{\text{CBF}} = \frac{\mathbf{a}_s}{\mathbf{a}_s^H \mathbf{a}_s}$$

# ULA CBF beampattern



## Impact on SNR

Considering a single source in white uncorrelated noise at input

$$\mathbf{y}[k] = \mathbf{a}_s s[k] + \mathbf{e}[k] \quad \text{SNR}_{\text{in}} = \frac{\mathbb{E}[|s[k]|^2]}{\mathbb{E}[|\mathbf{e}[k]|^2]} = \frac{P}{\sigma^2}$$

After beamforming

$$y_F[k] = \mathbf{w}^H \mathbf{y}[k] = \mathbf{w}^H \mathbf{a}_s s[k] + \mathbf{w}^H \mathbf{e}[k]$$

leading

$$\text{SNR}_{\text{out}} = \frac{\mathbb{E}[|\mathbf{w}^H \mathbf{a}_s s[k]|^2]}{\mathbb{E}[|\mathbf{w}^H \mathbf{e}[k]|^2]} = \frac{|\mathbf{w}^H \mathbf{a}_s|^2}{\|\mathbf{w}\|^2} \text{SNR}_{\text{in}} \leq \|\mathbf{a}_s\|^2 \text{SNR}_{\text{in}}$$



## Impact on SNR

Leading  $\text{SNR}_{\text{out}} \leq N \times \text{SNR}_{\text{in}}$  with equality if and only if  $\mathbf{w} \propto \mathbf{a}_s$

### Property

The CBF is the beamformer that maximizes the SNR at the output with unitary maximum gain

$$\mathbf{w}_{\text{CBF}} = \arg \max_{\mathbf{w}} \text{SNR}_{\text{out}} \quad \text{s.t.} \quad \mathbf{w}^H \mathbf{a}_s = 1$$

### White noise array gain

For any  $\mathbf{w}$  such that  $\mathbf{w}^H \mathbf{a}_s = 1$ , the white noise array gain is

$$A_{\text{WN}} = \frac{\text{SNR}_{\text{out}}}{\text{SNR}_{\text{in}}} = \|\mathbf{w}\|^{-2} \leq N$$

# Plan

## ① Introduction

## ② Model

## ③ Beamforming

Principle

Array beampattern

Spatial filtering

**Adaptive beamforming**

Robust adaptive beamforming

Partially adaptive beamforming

Summary

## ④ Direction of arrival estimation

## ⑤ Conclusion

# Adaptive beamforming

## Conventional beamformer

$$\mathbf{w}_{\text{CBF}} = \arg \max_{\mathbf{w}} \text{SNR}_{\text{out}} \quad \text{s.t.} \quad \mathbf{w}^H \mathbf{a}_s = 1$$

$$\Leftrightarrow \mathbf{w}_{\text{CBF}} = \arg \min_{\mathbf{w}} \|\mathbf{w}\|^2 \quad \text{s.t.} \quad \mathbf{w}^H \mathbf{a}_s = 1$$

The CBF is optimal in white noise: minimize  $\mathbf{w}^H \mathbf{w}$  (i.e., output power when input is  $\mathbf{e}$ ) under  $\mathbf{w}^H \mathbf{a}_s = 1$  (i.e., maintain unit gain in desired direction). Other directions are treated the same

## Adaptive beamforming

Takes into account potentiel interfering signals in some directions

# Adaptive beamforming

In presence of an interfering signal

$$\mathbf{y}[k] = \mathbf{a}_s s[k] + \mathbf{y}_i[k] + \mathbf{e}[k]$$

Filtered signal

$$y_F[k] = \underbrace{\mathbf{w}^H \mathbf{a}_s s[k]}_{\text{signal}} + \underbrace{\mathbf{w}^H (\mathbf{y}_i[k] + \mathbf{e}[k])}_{\text{interference + noise}}$$

## Signal to Interference plus Noise Ratio (SINR)

For a beamformer  $\mathbf{w}$ , the figure of merit is the Signal to Interference plus Noise Ratio (SINR)

$$\begin{aligned} \text{SINR}(\mathbf{w}) &= \frac{\mathbb{E}[|\mathbf{w}^H \mathbf{a}_s s[k]|^2]}{\mathbb{E}[|\mathbf{w}^H (\mathbf{y}_i[k] + \mathbf{e}[k])|^2]} \\ &= \frac{P_s |\mathbf{w}^H \mathbf{a}_s|^2}{\mathbf{w}^H \mathbf{C} \mathbf{w}} \end{aligned}$$

where  $\mathbf{C} = \mathbb{E}[(\mathbf{y}_i[k] + \mathbf{e}[k])(\mathbf{y}_i[k] + \mathbf{e}[k])^H]$  is the interference plus noise covariance matrix

## Interference plus Noise covariance matrix

Under zero mean, independence and narrow band assumptions

$$\begin{aligned}
 \mathbf{R} &= \sum_{p=1}^P P_p \mathbf{a}_p(\theta_p) \mathbf{a}_p^H(\theta_p) + \sigma^2 \mathbf{I}_N \\
 &= P_s \mathbf{a}_s(\theta_s) \mathbf{a}_s^H(\theta_s) + \underbrace{\sum_{j=1}^J P_j \mathbf{a}_j(\theta_j) \mathbf{a}_j^H(\theta_j)}_C + \sigma^2 \mathbf{I}_N
 \end{aligned}$$

where  $J$  is the number of interfering signals, and terms indexed by  $j$  are the corresponding parameters

# Optimal beamformer

CBF: maximises SNR while keeping unit gain in desired direction

## Optimal beamformer

Maximizes the SINR while ensuring unit gain towards  $\theta_s$

$$\begin{aligned} \max_{\mathbf{w}} \frac{P_s |\mathbf{w}^H \mathbf{a}_s|^2}{\mathbf{w}^H \mathbf{C} \mathbf{w}} \quad & \text{s.t.} \quad \mathbf{w}^H \mathbf{a}_s = 1 \\ \Leftrightarrow \min_{\mathbf{w}} \mathbf{w}^H \mathbf{C} \mathbf{w} \quad & \text{s.t.} \quad \mathbf{w}^H \mathbf{a}_s = 1 \end{aligned}$$

For CBF: consequence of its definition. How to solve this for Optimal beamformer?

## Reminder: constrained optimization (1/3)

One wants to solve

$$\min_{\mathbf{w} \in \mathbb{R}^N} f(\mathbf{w}) \quad \text{s.t. } g(\mathbf{w}) = 0 \quad \text{where } f : \mathbb{R}^N \rightarrow \mathbb{R} \text{ and } g : \mathbb{R}^N \rightarrow \mathbb{R}$$

### Lagrangian

We introduce the vector of Lagrange multiplier  $\lambda \in \mathbb{R}$  and define the Lagrangian as

$$\mathcal{L}(\mathbf{w}, \lambda) = f(\mathbf{w}) + \lambda g(\mathbf{w})$$

KKT conditions

$$\frac{\partial \mathcal{L}(\mathbf{w}, \lambda)}{\partial \mathbf{w}} = \mathbf{0} \quad \text{and} \quad \frac{\partial \mathcal{L}(\mathbf{w}, \lambda)}{\partial \lambda} = 0$$

Here, variables are complex



## Reminder: constrained optimization (2/3)

One wants to solve

$$\min_{\mathbf{w} \in \mathbb{C}^N} f(\mathbf{w}) \quad \text{s.t. } g(\mathbf{w}) = 0 \quad \text{where } f : \mathbb{C}^N \rightarrow \mathbb{R} \text{ and } g : \mathbb{C}^N \rightarrow \mathbb{C}$$

### Lagrangian

We introduce the vector of Lagrange multiplier  $\lambda \in \mathbb{C}$  and define the Lagrangian as

$$\mathcal{L}(\mathbf{w}, \lambda) = f(\mathbf{w}) + \text{Re}(\lambda^* g(\mathbf{w})) = f(\mathbf{w}) + \frac{1}{2} (\lambda^* g(\mathbf{w}) + \lambda g^*(\mathbf{w}))$$

KKT conditions

$$\frac{\partial \mathcal{L}(\mathbf{w}, \lambda)}{\partial \mathbf{w}^*} = \mathbf{0} \quad \text{and} \quad \frac{\partial \mathcal{L}(\mathbf{w}, \lambda)}{\partial \lambda^*} = 0$$

In complex derivatives,  $z$  and  $z^*$  are considered independent variables

## Reminder: constrained optimization (3/3)

For convex problems, KKT lead optimal solution

### Optimal beamformer

The problem

$$\min_{\mathbf{w}} \mathbf{w}^H \mathbf{C} \mathbf{w} \quad \text{s.t.} \quad \mathbf{w}^H \mathbf{a}_s = 1$$

is convex

Lagrangian

$$\mathcal{L}(\mathbf{w}, \lambda) = \mathbf{w}^H \mathbf{C} \mathbf{w} + \frac{1}{2} \left( \lambda^* (\mathbf{w}^H \mathbf{a}_s - 1) + \lambda (\mathbf{w}^H \mathbf{a}_s - 1)^* \right)$$

N.B.

$$\frac{\partial \mathbf{a}^T \mathbf{x}}{\partial \mathbf{x}} = \frac{\partial \mathbf{x}^T \mathbf{a}}{\partial \mathbf{x}} = \mathbf{a}$$

# Optimal beamformer

## Optimal beamformer

Maximizes the SINR while ensuring unit gain towards  $\theta_s$

$$\min_{\mathbf{w}} \mathbf{w}^H \mathbf{C} \mathbf{w} \quad \text{s.t.} \quad \mathbf{w}^H \mathbf{a}_s = 1$$

leading

$$\mathbf{w}_{\text{opt}} = \frac{\mathbf{C}^{-1} \mathbf{a}_s}{\mathbf{a}_s^H \mathbf{C}^{-1} \mathbf{a}_s}$$

For such  $\mathbf{w}$

$$\text{SINR}(\mathbf{w}_{\text{opt}}) = P_s \mathbf{a}_s^H \mathbf{C}^{-1} \mathbf{a}_s$$

# Principles

CBF: minimize the output power when input is  $\mathbf{e}$  under the constraint that the actual steering vector  $\mathbf{a}_s$  goes non distorted

Optimal: minimize the output power when input is  $\mathbf{y}_i + \mathbf{e}$  under the constraint that the actual steering vector  $\mathbf{a}_s$  goes non distorted

## Issues

In practice: neither  $\mathbf{a}_s$  nor  $\mathbf{C}$  is known. Actual steering vector might differ from its expected value and  $\mathbf{C}$  needs to be estimated from the data

# Minimum Variance Distortionless Response (MVDR)

## MVDR beamformer

Minimize output power (when input is  $\mathbf{y}_i + \mathbf{e}$ ) under the constraint that the steering vector in the **assumed** direction  $\theta_0$  goes non distorted

$$\min_{\mathbf{w}} \mathbf{w}^H \mathbf{C} \mathbf{w} \text{ subject to } \mathbf{w}^H \mathbf{a}_0 = 1$$

with  $\mathbf{a}_0 = \mathbf{a}(\theta_0)$  the assumed steering vector of the signal of interest (Sol)

$$\mathbf{w}_{\text{MVDR}} = \frac{\mathbf{C}^{-1} \mathbf{a}_0}{\mathbf{a}_0^H \mathbf{C}^{-1} \mathbf{a}_0}$$

$$\text{SINR}(\mathbf{w}_{\text{MVDR}}) = P_s \frac{|\mathbf{a}_0^H \mathbf{C}^{-1} \mathbf{a}_s|^2}{\mathbf{a}_0^H \mathbf{C}^{-1} \mathbf{a}_0}$$

# Minimum Power Distortionless Response (MPDR)

## MPDR beamformer

Minimize output power (when input is  $\mathbf{a}_s s + \mathbf{y}_i + \mathbf{e}$ ) under the constraint that the steering vector in the **assumed** direction  $\theta_0$  goes non distorted

$$\min_{\mathbf{w}} \mathbf{w}^H \mathbf{R} \mathbf{w} \text{ subject to } \mathbf{w}^H \mathbf{a}_0 = 1$$

with  $\mathbf{R} = P_s \mathbf{a}_s \mathbf{a}_s^H + \mathbf{C}$  the signal plus noise plus interference covariance matrix

$$\mathbf{w}_{\text{MPDR}} = \frac{\mathbf{R}^{-1} \mathbf{a}_0}{\mathbf{a}_0^H \mathbf{R}^{-1} \mathbf{a}_0}$$

$$\text{SINR}(\mathbf{w}_{\text{MPDR}}) = P_s \frac{|\mathbf{a}_0^H \mathbf{R}^{-1} \mathbf{a}_s|^2}{\mathbf{a}_0^H \mathbf{R}^{-1} \mathbf{C} \mathbf{R}^{-1} \mathbf{a}_0}$$

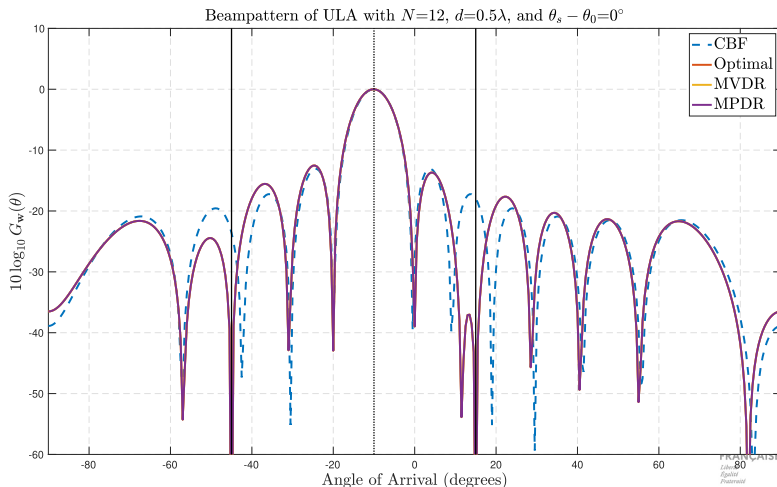
## Summary of adaptive beamformers (known covariances)

Beamformer	Principle	Weight vector
Optimal	$\min_{\mathbf{w}} \underbrace{\mathbf{w}^H \mathbf{C} \mathbf{w}}_{\text{output power}} \text{ s.t. } \underbrace{\mathbf{w}^H \mathbf{a}_s}_{\text{gain constraint}} = 1$	$\mathbf{w}_{\text{opt}} = \frac{\mathbf{C}^{-1} \mathbf{a}_s}{\mathbf{a}_s^H \mathbf{C}^{-1} \mathbf{a}_s}$
MVDR	$\min_{\mathbf{w}} \mathbf{w}^H \mathbf{C} \mathbf{w} \text{ s.t. } \mathbf{w}^H \mathbf{a}_0 = 1$	$\mathbf{w}_{\text{MVDR}} = \frac{\mathbf{C}^{-1} \mathbf{a}_0}{\mathbf{a}_0^H \mathbf{C}^{-1} \mathbf{a}_0}$
MPDR	$\min_{\mathbf{w}} \mathbf{w}^H \mathbf{R} \mathbf{w} \text{ s.t. } \mathbf{w}^H \mathbf{a}_0 = 1$	$\mathbf{w}_{\text{MPDR}} = \frac{\mathbf{R}^{-1} \mathbf{a}_0}{\mathbf{a}_0^H \mathbf{R}^{-1} \mathbf{a}_0}$

- $\mathbf{a}_s$ : true steering vector
- $\mathbf{a}_0$ : assumed steering vector
- $\mathbf{C}$ :  $\text{cov}(\mathbf{y}_i + \mathbf{e})$
- $\mathbf{R}$ :  $\text{cov}(\mathbf{a}_s + s\mathbf{y}_i + \mathbf{e})$

# ULA adaptive beamformer beampatterns

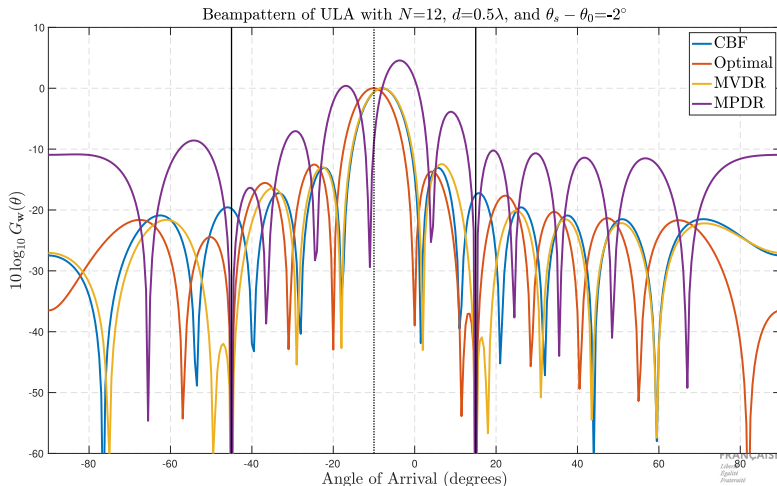
Misalignment:  $0^\circ$





# ULA adaptive beamformer beampatterns

Misalignment:  $2^\circ$



## CBF vs MVDR: single interference

If  $\mathbf{y}[k] = \mathbf{a}_s s[k] + \mathbf{a}_I s_I[k] + \mathbf{e}[k]$  then  $\mathbf{C} = P_I \mathbf{a}_I \mathbf{a}_I^H + \sigma^2 \mathbf{I}_N$

For  $\text{INR} = \frac{P_I}{\sigma^2} \gg 1$  it can be shown

$$\text{SINR}_{\text{CBF}} \approx \frac{P_s}{\sigma^2} \times \frac{1}{g \times \text{INR}} \quad \text{SINR}_{\text{opt}} \approx \frac{P_s}{\sigma^2} \times N(1 - g)$$

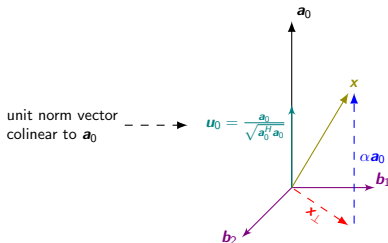
where  $g = \cos^2(\mathbf{a}_s, \mathbf{a}_j) = \frac{|\mathbf{a}_s^H \mathbf{a}_j|^2}{(\mathbf{a}_s^H \mathbf{a}_s)(\mathbf{a}_j^H \mathbf{a}_j)} \in [0, 1]$

### Remarks

- $P_I \nearrow$  leads  $\text{SINR}_{\text{CBF}} \searrow$  while  $\text{SINR}_{\text{opt}}$  is constant
- $\text{SINR} \searrow$  when  $\mathbf{a}_j \rightarrow \mathbf{a}_s$  ( $g \rightarrow 1$ )

## Reminder: linear algebra in Hilbert spaces (1/4)

Given a vector  $\mathbf{a}_0 \in \mathbb{C}^N$  and  $\mathbf{u}_0 = \frac{\mathbf{a}_0}{\sqrt{\mathbf{a}_0^H \mathbf{a}_0}}$ .  $\mathbf{u}_0$  can be completed with vectors  $\{\mathbf{b}_1, \dots, \mathbf{b}_{N-1}\}$  to form an orthonormal basis



Any vector  $\mathbf{x} \in \mathbb{C}^N$  can then be decomposed as

$$\mathbf{x} = \underbrace{\mathbf{u}_0^H \mathbf{x}}_{\alpha \sqrt{\mathbf{a}_0^H \mathbf{a}_0}} \mathbf{u}_0 + \underbrace{\mathbf{x} - \mathbf{u}_0^H \mathbf{x} \mathbf{u}_0}_{\mathbf{x}_\perp}$$

## Reminder: linear algebra in Hilbert spaces (2/4)

The vector space spanned by  $\mathbf{a}_0$  (or its range) is defined as

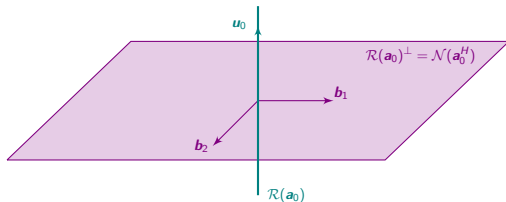
$$\mathcal{R}(\mathbf{a}_0) = \{\alpha \mathbf{a}_0, \forall \alpha \in \mathbb{C}\} \subset \mathbb{C}^N \text{ and } \dim(\mathcal{R}(\mathbf{a}_0)) = 1$$

Its orthogonal subspace is the null-space (or kernel) of  $\mathbf{a}_0^H$

$$\mathcal{R}(\mathbf{a}_0)^\perp = \{\mathbf{x} \in \mathbb{C}^N, \mathbf{a}_0^H \mathbf{x} = 0\} = \mathcal{N}(\mathbf{a}_0^H) \text{ and } \dim(\mathcal{N}(\mathbf{a}_0)) = N - 1$$

and

$$\mathcal{R}(\mathbf{a}_0) \oplus \mathcal{R}(\mathbf{a}_0)^\perp = \mathbb{C}^N$$



## Reminder: linear algebra in Hilbert spaces (3/4)

Expanding  $\mathbf{u}_0$  leads

$$\begin{aligned}\mathbf{x} &= \frac{\mathbf{a}_0^H \mathbf{x} \mathbf{a}_0}{\mathbf{a}_0^H \mathbf{a}_0} + \mathbf{x} - \frac{\mathbf{a}_0^H \mathbf{x} \mathbf{a}_0}{\mathbf{a}_0^H \mathbf{a}_0} \\ &= \Pi_{\mathbf{a}_0} \mathbf{x} + \Pi_{\mathbf{a}_0}^\perp \mathbf{x}\end{aligned}$$

where

$$\Pi_{\mathbf{a}_0} = \frac{\mathbf{a}_0 \mathbf{a}_0^H}{\mathbf{a}_0^H \mathbf{a}_0}$$

is the orthogonal projector<sup>1</sup> on  $\mathcal{R}(\mathbf{a}_0)$  and

$$\Pi_{\mathbf{a}_0}^\perp = I_N - \Pi_{\mathbf{a}_0}$$

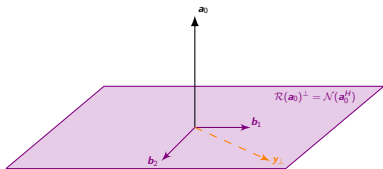
is the orthogonal projector on  $\mathcal{R}(\mathbf{a}_0)^\perp$

---

<sup>1</sup> $\Pi_{\mathbf{a}_0} \times \Pi_{\mathbf{a}_0} = \Pi_{\mathbf{a}_0}$  and  $\Pi_{\mathbf{a}_0}^H = \Pi_{\mathbf{a}_0} \Rightarrow$  orthogonal projector and  $\forall \mathbf{x} \in \mathbb{C}^N, \Pi_{\mathbf{a}_0} \mathbf{x} \in \mathcal{R}(\mathbf{a}_0)$

## Reminder: linear algebra in Hilbert spaces (4/4)

We have  $\mathbf{b}_i^H \mathbf{a}_0 = 0, i = 1, \dots, N - 1$  and the family  $\{\mathbf{b}_1, \dots, \mathbf{b}_{N-1}\}$  is an orthonormal basis of  $\mathcal{R}(\mathbf{a}_0)^\perp$



$$\mathbf{y}_\perp = \sum_{i=1}^{N-1} y_{a,i} \mathbf{b}_i \quad \text{where } y_{a,i} = \mathbf{b}_i^H \mathbf{y}_\perp$$

We consider

$$\mathbf{B} = [\mathbf{b}_1 \quad \dots \quad \mathbf{b}_{N-1}] \in \mathbb{C}^{N \times N-1} \quad \text{and} \quad \mathbf{y}_a = [y_{a,1} \quad \dots \quad y_{a,N-1}]^T \in \mathbb{C}^{N-1}$$

### Properties

$\mathbf{B}$  is rank  $N - 1$ ,  $\mathbf{B}^H \mathbf{a}_0 = \mathbf{0} \in \mathbb{C}^{N-1}$  and  $\mathbf{y}_\perp = \mathbf{B} \mathbf{y}_a$

## Generalized sidelobe canceler

Any beamformer can be decomposed as

$$\mathbf{w} = \alpha \mathbf{a}_0 + \mathbf{w}_\perp \quad \text{where } \mathbf{w}_\perp^H \mathbf{a}_0 = 0$$

- the component along  $\mathbf{a}_0$  ensures the unit gain constraint

$$\mathbf{w}^H \mathbf{a}_0 = \alpha \mathbf{a}_0^H \mathbf{a}_0 + \mathbf{w}_\perp^H \mathbf{a}_0 = \alpha \mathbf{a}_0^H \mathbf{a}_0 \Rightarrow \alpha = (\mathbf{a}_0^H \mathbf{a}_0)^{-1}$$

- orthogonal component is arbitrary, in a space of dimension  $N - 1$

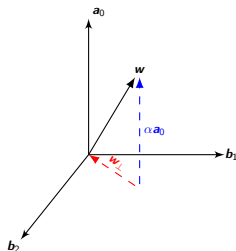
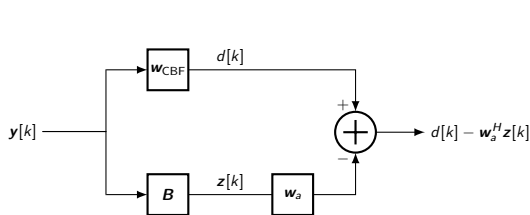
$$\mathbf{w}_\perp = \mathbf{B} \mathbf{w}_a \quad \text{with} \quad \mathbf{B} \in \mathbb{C}^{N \times N-1}, \mathbf{B}^H \mathbf{a}_0 = 0, \mathbf{w}_a \in \mathbb{C}^{N-1}$$

## Generalized sidelobe canceler

We have

$$\mathbf{w} = (\mathbf{a}_0^H \mathbf{a}_0)^{-1} \mathbf{a}_0 - \mathbf{w}_\perp = \mathbf{w}_{\text{CBF}} - \mathbf{B} \mathbf{w}_a$$

$$\Rightarrow \mathbf{w}^H \mathbf{y} = \mathbf{w}_{\text{CBF}}^H \mathbf{y} - \mathbf{w}_a^H \mathbf{B}^H \mathbf{y} = d - \mathbf{w}_a^H \mathbf{z}$$



The  $N - 1$  auxiliary channels in  $\mathbf{z}[k]$  are free of signal and can be used to remove the contribution of interferences



## Generalized sidelobe canceler

MVDR is obtained equivalently by solving<sup>2</sup>

$$\underbrace{\min_{\mathbf{w}, \mathbf{w}^H \mathbf{a}_0 = 1} \mathbf{w}^H \mathbf{C} \mathbf{w}}_{\text{direct form, constrained}} \qquad \underbrace{\min_{\mathbf{w}_a} (\mathbf{w}_{\text{CBF}} - \mathbf{B} \mathbf{w}_a)^H \mathbf{C} (\mathbf{w}_{\text{CBF}} - \mathbf{B} \mathbf{w}_a)}_{\text{GSC form, unconstrained}}$$

The GSC form of the MVDR beamformer is

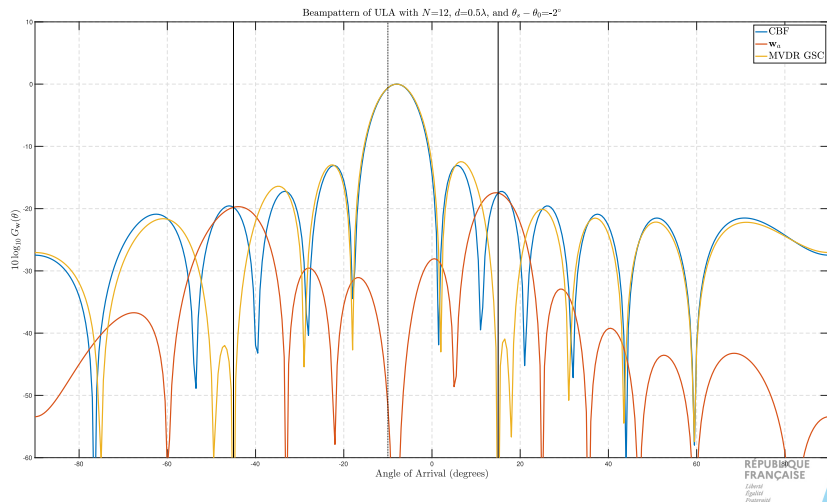
$$\mathbf{w}_{\text{GSC}} = \mathbf{w}_{\text{CBF}} - \mathbf{B} \mathbf{w}_{a, \text{MVDR}}$$

where  $\mathbf{w}_{a, \text{MVDR}}$  solves the above estimation problem

$$\mathbf{w}_{a, \text{MVDR}} = (\mathbf{B}^H \mathbf{C} \mathbf{B})^{-1} \mathbf{B}^H \mathbf{C} \mathbf{w}_{\text{CBF}}$$

<sup>2</sup>MPDR:  $\mathbf{C} \leftrightarrow \mathbf{R}$

# ULA adaptive beamformer beampatterns



## Adaptive beamformer comparison

Three beamformers

- ① Opt:  $\min_{\mathbf{w}} \mathbf{w}^H \mathbf{C} \mathbf{w}$  subject to  $\mathbf{w}^H \mathbf{a}_s = 1$
- ② MVDR:  $\min_{\mathbf{w}} \mathbf{w}^H \mathbf{C} \mathbf{w}$  subject to  $\mathbf{w}^H \mathbf{a}_0 = 1$
- ③ MPDR:  $\min_{\mathbf{w}} \mathbf{w}^H (\mathbf{C} + P_s \mathbf{a}_s \mathbf{a}_s^H) \mathbf{w}$  subject to  $\mathbf{w}^H \mathbf{a}_0 = 1$

The optimal, MVDR, and MPDR beamformers are equivalent when

- the assumed steering vector  $\mathbf{a}_0$  coincides the true steering vector  $\mathbf{a}_s$
- the covariance matrix of the incoming signal  $\mathbf{R}$  is known

In practice

- uncalibrated arrays and pointing errors lead  $\mathbf{a}_0 \neq \mathbf{a}_s$
- the covariance matrix is estimated leading estimation errors  $\hat{\mathbf{R}} - \mathbf{R}$

⇒ Degradation of optimal SINR ( $P_s \mathbf{a}_s^H \mathbf{C}^{-1} \mathbf{a}_s$ )

## Influence of the steering vector (MVDR)

We have

$$\begin{aligned}
 \text{SINR}(\mathbf{w}_{\text{MVDR}}) &= P_s \frac{|\mathbf{a}_0^H \mathbf{C}^{-1} \mathbf{a}_s|^2}{\mathbf{a}_0^H \mathbf{C}^{-1} \mathbf{a}_0} \\
 &= P_s \mathbf{a}_s^H \mathbf{C}^{-1} \mathbf{a}_s \frac{|\mathbf{a}_0^H \mathbf{C}^{-1} \mathbf{a}_s|^2}{(\mathbf{a}_s^H \mathbf{C}^{-1} \mathbf{a}_s)(\mathbf{a}_0^H \mathbf{C}^{-1} \mathbf{a}_0)} \\
 &= \text{SINR}_{\text{opt}} \frac{|\mathbf{a}_0^H \mathbf{C}^{-1} \mathbf{a}_s|^2}{(\mathbf{a}_s^H \mathbf{C}^{-1} \mathbf{a}_s)(\mathbf{a}_0^H \mathbf{C}^{-1} \mathbf{a}_0)} \\
 &= \text{SINR}_{\text{opt}} \cos^2(\mathbf{a}_s, \mathbf{a}_0; \mathbf{C}^{-1}) \\
 &\leq \text{SINR}_{\text{opt}}
 \end{aligned}$$

## Influence of the steering vector (MPDR)

We have

$$\text{SINR}(\mathbf{w}_{\text{MPDR}}) = P_s \frac{|\mathbf{a}_0^H \mathbf{R}^{-1} \mathbf{a}_s|^2}{\mathbf{a}_0^H \mathbf{R}^{-1} \mathbf{C} \mathbf{R}^{-1} \mathbf{a}_0}, \quad \mathbf{R} = P_s \mathbf{a}_s \mathbf{a}_s^H + \mathbf{C}$$

Using matrix inversion lemma<sup>3</sup>

$$\mathbf{R}^{-1} = \mathbf{C}^{-1} - \frac{P_s \mathbf{C}^{-1} \mathbf{a}_s \mathbf{a}_s^H \mathbf{C}^{-1}}{1 + P_s \mathbf{a}_s^H \mathbf{C}^{-1} \mathbf{a}_s} = \mathbf{C}^{-1} - \frac{P_s \mathbf{C}^{-1} \mathbf{a}_s \mathbf{a}_s^H \mathbf{C}^{-1}}{1 + \text{SINR}_{\text{opt}}}$$

hence

$$\mathbf{a}_0^H \mathbf{R}^{-1} \mathbf{a}_s = \mathbf{a}_0^H \mathbf{C}^{-1} \mathbf{a}_s \left( 1 - \frac{\text{SINR}_{\text{opt}}}{1 + \text{SINR}_{\text{opt}}} \right)$$

leading

$$P_s |\mathbf{a}_0^H \mathbf{R}^{-1} \mathbf{a}_s|^2 = \frac{\mathbf{a}_0^H \mathbf{C}^{-1} \mathbf{a}_0 \text{SINR}(\mathbf{w}_{\text{MVDR}})}{(1 + \text{SINR}_{\text{opt}})^2}$$

<sup>3</sup> $(\mathbf{A} + \mathbf{C} \mathbf{B} \mathbf{C}^H)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{C} (\mathbf{B}^{-1} + \mathbf{C}^H \mathbf{A}^{-1} \mathbf{C})^{-1} \mathbf{C}^H \mathbf{A}^{-1}$

# Influence of the steering vector (MPDR)

Denominator

$$\begin{aligned}
 & \mathbf{a}_0^H \mathbf{R}^{-1} \mathbf{C} \mathbf{R}^{-1} \mathbf{a}_0 \\
 &= \mathbf{a}_0^H \mathbf{C}^{-1} \left( \mathbf{I}_N - \frac{P_s \mathbf{a}_s \mathbf{a}_s^H \mathbf{C}^{-1}}{1 + \text{SINR}_{\text{opt}}} \right) \left( \mathbf{I}_N - \frac{P_s \mathbf{a}_s \mathbf{a}_s^H \mathbf{C}^{-1}}{1 + \text{SINR}_{\text{opt}}} \right) \mathbf{a}_0 \\
 &= \mathbf{a}_0^H \mathbf{C}^{-1} \left( \mathbf{a}_0 - 2 \frac{P_s \mathbf{a}_s \mathbf{a}_s^H \mathbf{C}^{-1} \mathbf{a}_0}{1 + \text{SINR}_{\text{opt}}} + \frac{P_s^2 \mathbf{a}_s \mathbf{a}_s^H \mathbf{C}^{-1} \mathbf{a}_s}{(1 + \text{SINR}_{\text{opt}})^2} \mathbf{a}_s^H \mathbf{C}^{-1} \mathbf{a}_0 \right) \\
 &= \left( \mathbf{a}_0^H \mathbf{C}^{-1} \mathbf{a}_0 - 2 \frac{P_s \mathbf{a}_0^H \mathbf{C}^{-1} \mathbf{a}_s \mathbf{a}_s^H \mathbf{C}^{-1} \mathbf{a}_0}{1 + \text{SINR}_{\text{opt}}} + \frac{P_s^2 \mathbf{a}_0^H \mathbf{C}^{-1} \mathbf{a}_s \mathbf{a}_s^H \mathbf{C}^{-1} \mathbf{a}_s}{(1 + \text{SINR}_{\text{opt}})^2} \mathbf{a}_s^H \mathbf{C}^{-1} \mathbf{a}_0 \right) \\
 &= \mathbf{a}_0^H \mathbf{C}^{-1} \mathbf{a}_0 \left( 1 - 2 \frac{P_s |\mathbf{a}_0^H \mathbf{C}^{-1} \mathbf{a}_s|^2}{\mathbf{a}_0^H \mathbf{C}^{-1} \mathbf{a}_0 (1 + \text{SINR}_{\text{opt}})} + \frac{P_s^2 |\mathbf{a}_0^H \mathbf{C}^{-1} \mathbf{a}_s|^2 \mathbf{a}_s^H \mathbf{C}^{-1} \mathbf{a}_s}{\mathbf{a}_0^H \mathbf{C}^{-1} \mathbf{a}_0 (1 + \text{SINR}_{\text{opt}})^2} \right)
 \end{aligned}$$

## Influence of the steering vector (MPDR)

Denominator

$$\begin{aligned} & \mathbf{a}_0^H \mathbf{C}^{-1} \mathbf{a}_0 \left( 1 - 2 \frac{P_s \cos(\mathbf{a}_s, \mathbf{a}_0; \mathbf{C}^{-1})^2 \mathbf{a}_s^H \mathbf{C}^{-1} \mathbf{a}_s}{(1 + \text{SINR}_{\text{opt}})} + \frac{P_s^2 \cos(\mathbf{a}_s, \mathbf{a}_0; \mathbf{C}^{-1})^2 (\mathbf{a}_s^H \mathbf{C}^{-1} \mathbf{a}_s)^2}{(1 + \text{SINR}_{\text{opt}})^2} \right) \\ &= \mathbf{a}_0^H \mathbf{C}^{-1} \mathbf{a}_0 \left( 1 - 2 \frac{\cos(\mathbf{a}_s, \mathbf{a}_0; \mathbf{C}^{-1})^2 \text{SINR}_{\text{opt}}}{(1 + \text{SINR}_{\text{opt}})} + \frac{\cos(\mathbf{a}_s, \mathbf{a}_0; \mathbf{C}^{-1})^2 \text{SINR}_{\text{opt}}^2}{(1 + \text{SINR}_{\text{opt}})^2} \right) \end{aligned}$$

Hence

$$\text{SINR}(\mathbf{w}_{\text{MPDR}}) = \frac{\text{SINR}(\mathbf{w}_{\text{MVDR}})}{\left(1 + \text{SINR}_{\text{opt}}\right)^2 \left(1 - 2 \frac{\cos(\mathbf{a}_s, \mathbf{a}_0; \mathbf{C}^{-1})^2 \text{SINR}_{\text{opt}}}{(1 + \text{SINR}_{\text{opt}})} + \frac{\cos(\mathbf{a}_s, \mathbf{a}_0; \mathbf{C}^{-1})^2 \text{SINR}_{\text{opt}}^2}{(1 + \text{SINR}_{\text{opt}})^2}\right)}$$

# Influence of the steering vector (MPDR)

## Leading

$$\begin{aligned}
 & \text{SINR}(\mathbf{w}_{\text{MPDR}}) \\
 &= \frac{\text{SINR}(\mathbf{w}_{\text{MVDR}})}{\left(1 + \text{SINR}_{\text{opt}}\right)^2 - 2 \cos^2(\mathbf{a}_s, \mathbf{a}_0; \mathbf{C}^{-1})^2 \text{SINR}_{\text{opt}}(1 + \text{SINR}_{\text{opt}}) + \cos^2(\mathbf{a}_s, \mathbf{a}_0; \mathbf{C}^{-1})^2 \text{SINR}_{\text{opt}}^2} \\
 &= \frac{\text{SINR}(\mathbf{w}_{\text{MVDR}})}{1 + 2\text{SINR}_{\text{opt}} + \text{SINR}_{\text{opt}}^2 - 2 \cos^2(\mathbf{a}_s, \mathbf{a}_0; \mathbf{C}^{-1})\text{SINR}_{\text{opt}} - \cos^2(\mathbf{a}_s, \mathbf{a}_0; \mathbf{C}^{-1})\text{SINR}_{\text{opt}}^2} \\
 &= \frac{\text{SINR}(\mathbf{w}_{\text{MVDR}})}{1 + \left(2\text{SINR}_{\text{opt}} + \text{SINR}_{\text{opt}}^2\right) \left(1 - \cos^2(\mathbf{a}_s, \mathbf{a}_0; \mathbf{C}^{-1})\right)} \\
 &= \frac{\text{SINR}(\mathbf{w}_{\text{MVDR}})}{1 + \left(2\text{SINR}_{\text{opt}} + \text{SINR}_{\text{opt}}^2\right) \sin^2(\mathbf{a}_s, \mathbf{a}_0; \mathbf{C}^{-1})}
 \end{aligned}$$



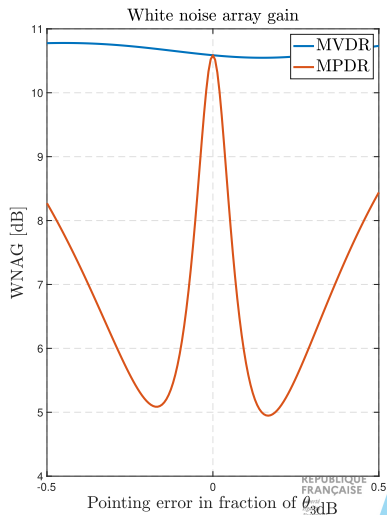
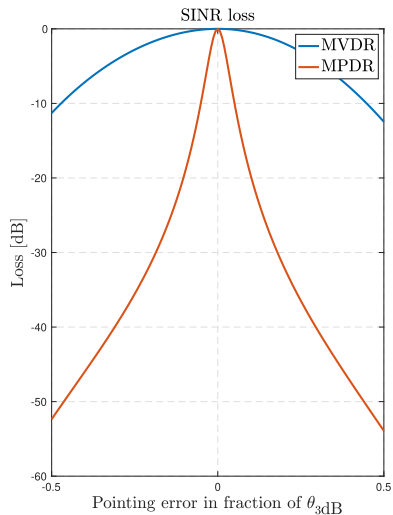
## Influence of the steering vector (MPDR)

Finally

$$\begin{aligned} \text{SINR}(\mathbf{w}_{\text{MPDR}}) &= \frac{\text{SINR}(\mathbf{w}_{\text{MVDR}})}{1 + (2\text{SINR}_{\text{opt}} + \text{SINR}_{\text{opt}}^2) \sin^2(\mathbf{a}_s, \mathbf{a}_0; \mathbf{C}^{-1})} \\ &\leq \text{SINR}(\mathbf{w}_{\text{MVDR}}) \end{aligned}$$

The degradation is more important as  $\text{SINR}_{\text{opt}}$  (hence  $P_s$ ) increases

# Influence of a steering vector error on SINR and WNAG



## Case of an uncalibrated array

Uncalibrated arrays lead different gains and phases

$$y_n[k] = \underbrace{(1 + g_n)e^{j\phi_n} a_n(\theta)}_{\tilde{a}_n(\theta)} s[k] + e_n[k]$$

with  $g_n$  and  $\phi_n$  independent random gains and phases such as

$$\mathbb{E}[g_n] = 0, \quad \mathbb{E}[g_n^2] = \sigma_g^2, \quad \mathbb{E}[e^{j\phi_n}] = \gamma$$

For any beamformer  $\mathbf{w}$

$$\mathbb{E}[\tilde{G}_{\mathbf{w}}(\theta)] = |\gamma|^2 G_{\mathbf{w}}(\theta) + \underbrace{(1 + \sigma_g^2 - |\gamma|^2) \|\mathbf{w}\|^2}_{\text{sidelobe level increase}}$$

## Influence of the number of snapshots

In practice, limited number of snapshots  $K$

$$\mathbf{y}[k] = \mathbf{a}_s s[k] + \underbrace{\mathbf{y}_i[k] + \mathbf{e}[k]}_{\mathbf{y}_{i+e}[k]}, \quad k = 1, \dots, K$$

Covariance matrices are estimated from these snapshots

$$\hat{\mathbf{R}} = \frac{1}{K} \sum_{k=1}^K \mathbf{y}[k] \mathbf{y}^H[k] \quad \Rightarrow \quad \mathbf{w}_{\text{MPDR}}^{\text{smi}} = \frac{\hat{\mathbf{R}}^{-1} \mathbf{a}_0}{\mathbf{a}_0^H \hat{\mathbf{R}}^{-1} \mathbf{a}_0}$$

$$\hat{\mathbf{C}} = \frac{1}{K} \sum_{k=1}^K \mathbf{y}_{i+e}[k] \mathbf{y}_{i+e}^H[k] \quad \Rightarrow \quad \mathbf{w}_{\text{MVDR}}^{\text{smi}} = \frac{\hat{\mathbf{C}}^{-1} \mathbf{a}_0}{\mathbf{a}_0^H \hat{\mathbf{C}}^{-1} \mathbf{a}_0}$$

smi: "sample matrix inversion"

## Influence of the number of snapshots

Beamformers  $\mathbf{w}_{M \times DR}^{\text{smi}}$  will differ from  $\mathbf{w}_{M \times DR}$  since

$$\hat{\mathbf{R}} = \mathbf{R} + \Delta \mathbf{R} \quad \hat{\mathbf{C}} = \mathbf{C} + \Delta \mathbf{C}$$

$\mathbf{w}_{M \times DR}^{\text{smi}}$  are random vectors, and so are their SINR

$$\text{SINR}(\mathbf{w}_{\text{MPDR}}^{\text{smi}}) = P_s \frac{|\mathbf{a}_0^H \hat{\mathbf{R}}^{-1} \mathbf{a}_s|^2}{\mathbf{a}_0^H \hat{\mathbf{R}}^{-1} \mathbf{C} \hat{\mathbf{R}}^{-1} \mathbf{a}_0}$$

$$\text{SINR}(\mathbf{w}_{\text{MVDR}}^{\text{smi}}) = P_s \frac{|\mathbf{a}_0^H \hat{\mathbf{C}}^{-1} \mathbf{a}_s|^2}{\mathbf{a}_0^H \hat{\mathbf{C}}^{-1} \mathbf{C} \hat{\mathbf{C}}^{-1} \mathbf{a}_0}$$

Importance of the convergence speed of the estimators

# SINR loss with finite number of snapshots (MVDR)

Defining

$$\rho_{\text{MVDR}} = \frac{\text{SINR}(\mathbf{w}_{\text{MVDR}}^{\text{smi}})}{\text{SINR}_{\text{opt}}}$$

one can show

$$\rho_{\text{MVDR}} \sim \text{Beta}(K - N + 2, N - 1)$$

which is independent of  $\mathbf{C}$ , and

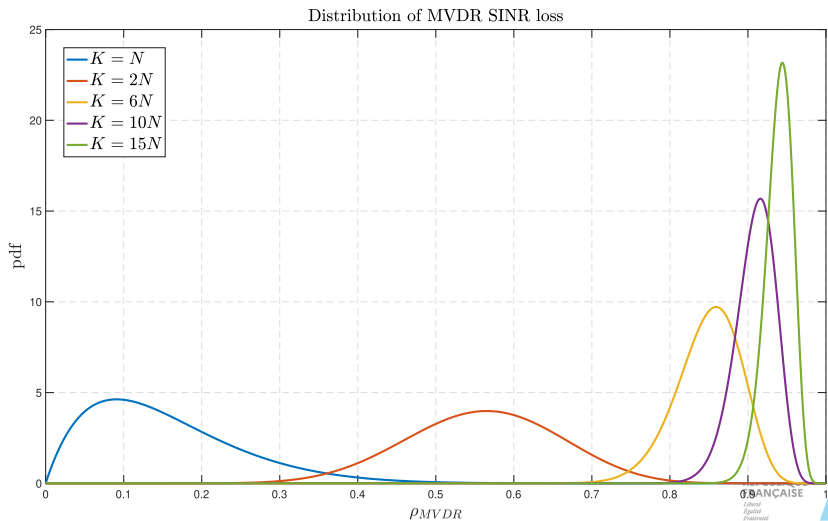
$$\mathbb{E}[\rho_{\text{MVDR}}] = \frac{K - N + 2}{K + 1}$$

For  $K_{\text{MVDR}} = 2N - 3$

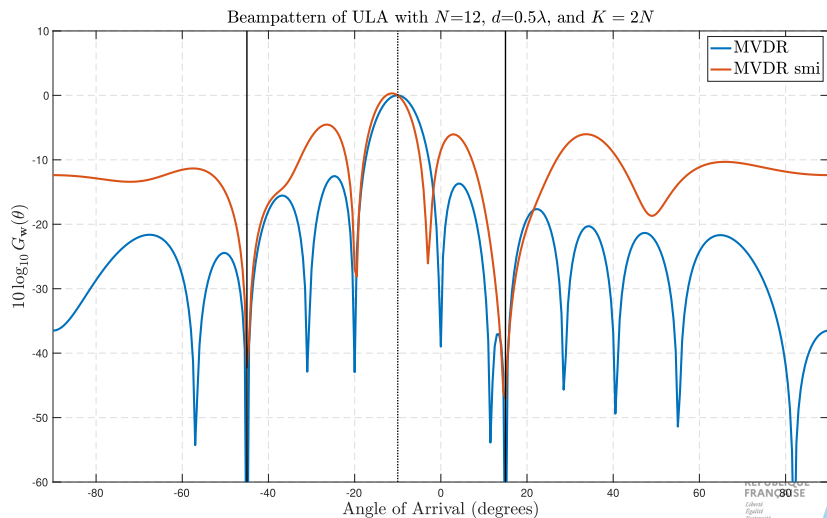
$$\mathbb{E}[\text{SINR}(\mathbf{w}_{\text{MVDR}}^{\text{smi}})] = \frac{\text{SINR}_{\text{opt}}}{2} \quad (3 \text{ dB loss on average})$$

Rule of thumb:  $K = 2N$

# SINR loss with finite number of snapshots (MVDR)



# MVDR versus smi version





## SINR loss with finite number of snapshots (MPDR)

Defining

$$\rho_{\text{MPDR}} = \frac{\text{SINR}(\mathbf{w}_{\text{MPDR}}^{\text{smi}})}{\text{SINR}_{\text{opt}}}$$

one can show that

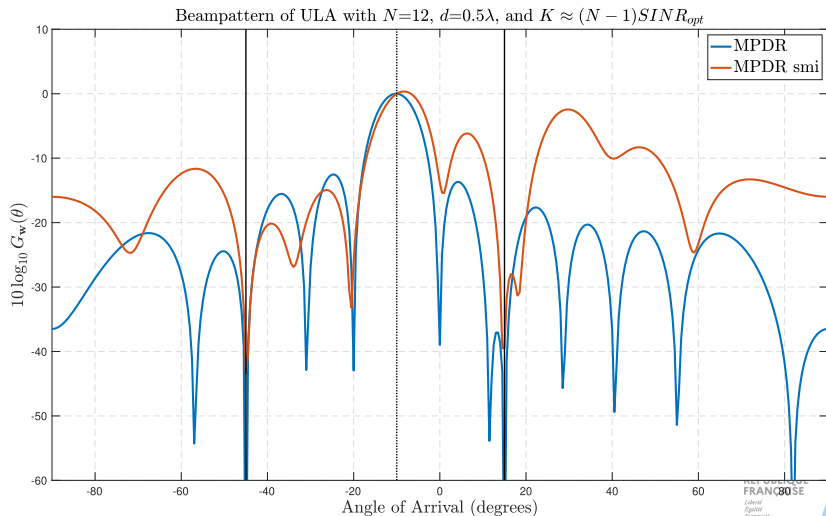
$$\mathbb{E}[\rho_{\text{MPDR}}] \approx \frac{K}{K + \text{SINR}_{\text{opt}}(N - 1)}$$

and to be within 3 dB of the optimal value

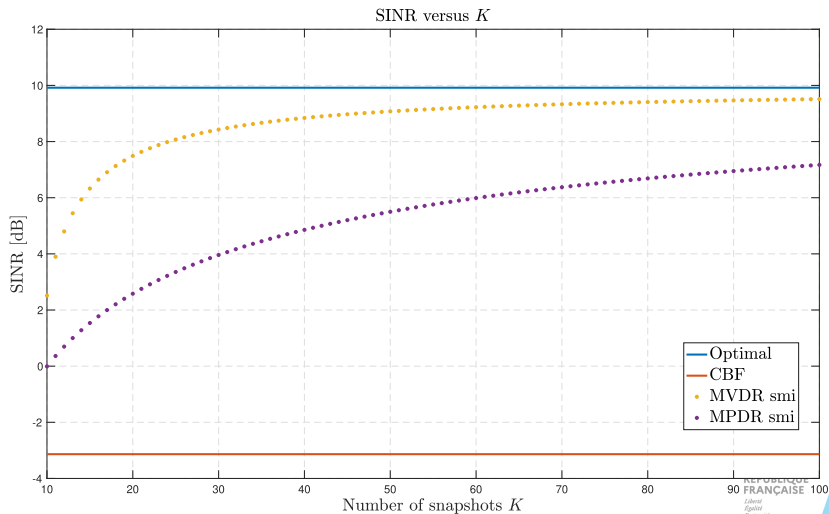
$$K_{\text{MPDR}} \approx (N - 1)\text{SINR}_{\text{opt}}$$

and in general,  $K_{\text{MPDR}} \gg K_{\text{MVDR}}$

# MPDR versus smi version



# Influence of $K$



# Conclusion

## In practice

- Estimation of covariance matrices  $\Rightarrow$  SINR loss (especially MPDR)
  - interference is less eliminated
  - sidelobe level increase ( $\Rightarrow$  lower white noise gain)
- Uncalibrated arrays: white noise gain  $\searrow$  (or  $\|\mathbf{w}\|^2 \nearrow$ )  $\Rightarrow$  steering vector errors  $\nearrow$

How to make these beamformers more robust?

# Plan

## ① Introduction

## ② Model

## ③ Beamforming

Principle

Array beampattern

Spatial filtering

Adaptive beamforming

**Robust adaptive beamforming**

Partially adaptive beamforming

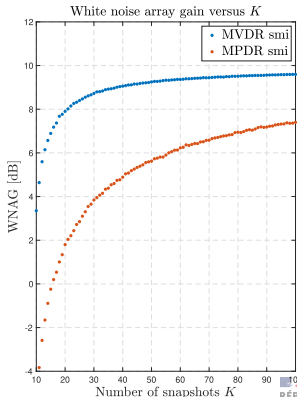
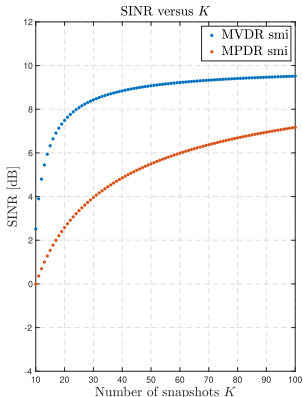
Summary

## ④ Direction of arrival estimation

## ⑤ Conclusion

# Motivation

- For calibration error, error grows with  $\|\mathbf{w}\|^2$
- For limited snapshot numbers



Idea: minimize  $\|\mathbf{w}\|^2$  (i.e., maximize WNAG)

# Diagonal loading

## Principle

Solve

$$\min_{\mathbf{w}} \mathbf{w}^H \hat{\mathbf{R}} \mathbf{w} \text{ subject to } \mathbf{w}^H \mathbf{a}_0 = 1 \text{ and } \|\mathbf{w}\|^2 \leq A_{\text{WN}}^{-1}$$

where  $\|\mathbf{w}\|^{-2} \leq N$  imposes  $A_{\text{WN}} \leq N$

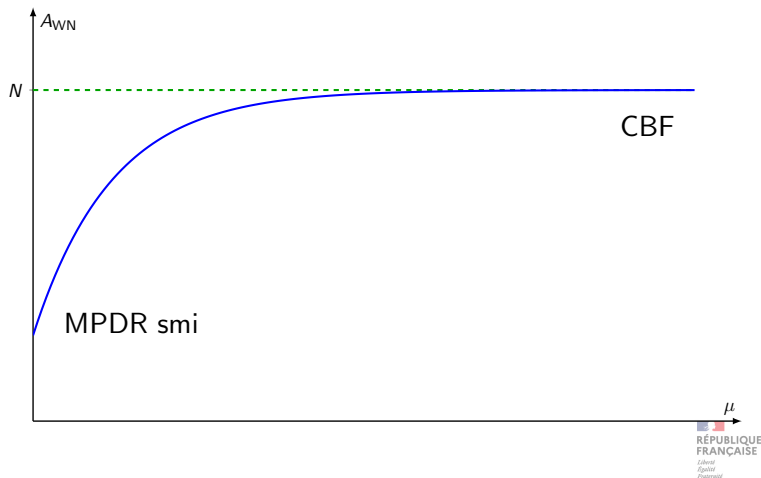
## Solution

We have

$$\mathbf{w}_{\text{MPDR-DL}}(\mu) = \frac{\left(\hat{\mathbf{R}} + \mu \mathbf{I}_N\right)^{-1} \mathbf{a}_0}{\mathbf{a}_0^H \left(\hat{\mathbf{R}} + \mu \mathbf{I}_N\right)^{-1} \mathbf{a}_0}$$

with  $\mu$  such as  $\|\mathbf{w}_{\text{MPDR-DL}}(\mu)\|^{-2} = A_{\text{WN}}$

# Diagonal loading: adaptivity versus robustness





## Setting the value of $\mu$

Several methods in the literature to set  $\mu$

- set  $A_{WN}$  and compute  $\mu$  from  $\|\mathbf{w}_{MPDR-DL}(\mu)\|^{-2} = A_{WN}$
- directly, a few dB above noise level (see next part)
- from the theory of ridge regression (and compute  $\mu$  from the data)
- it can be shown DL is equivalent to

$$\max_{P, \mathbf{a}} \left\| \hat{\mathbf{R}} - P\mathbf{a}\mathbf{a}^H \right\|^2 \quad \text{subject to } \|\mathbf{a} - \mathbf{a}_0\|^2 \leq \varepsilon^2$$

deduce  $\mu$  from  $\varepsilon$

- set  $A_{WN}$  and compute DL from GSC form
- ...

## Linear algebra reminder: eigenvalue decomposition

If  $\mathbf{A} \in \mathbb{C}^{N \times N}$  is Hermitian ( $\mathbf{A}^H = \mathbf{A}$ ), there exists a unitary matrix  $\mathbf{U} \in \mathbb{C}^{N \times N}$  ( $\mathbf{U}^H \mathbf{U} = \mathbf{U} \mathbf{U}^H = \mathbf{I}_N$ ) such as

$$\mathbf{A} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^H$$

where  $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_N) = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_N \end{bmatrix}$  is a diagonal matrix

containing the eigenvalues of  $\mathbf{A}$  (Hermitian  $\Rightarrow \lambda_n \in \mathbb{R}$ )

### Convention

If not stated otherwise, we assume  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$

## Application to signal covariance matrix

We have the (Hermitian) covariance matrix

$$\mathbf{R} = \mathbf{A}(\theta)\mathbf{R}_s\mathbf{A}(\theta)^H + \sigma^2\mathbf{I}_N$$

$\mathbf{A}(\theta)\mathbf{R}_s\mathbf{A}(\theta)^H$  is Hermitian, positive and with rank  $P$

eigenvalues are  $\underbrace{\lambda_1, \dots, \lambda_P}_{>0}, \underbrace{0, \dots, 0}_{N-P}$

$\sigma^2\mathbf{I}_N$  directly diagonal

$\exists \mathbf{U} \in \mathbb{C}^{N \times N}, \mathbf{R} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$  with  $\mathbf{\Lambda} = \text{diag}(\lambda_1 + \sigma^2, \dots, \lambda_P + \sigma^2, \sigma^2, \dots, \sigma^2)$

and  $\mathbf{R}^{-1} = \mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}^H$

## Application to signal covariance matrix

With 1 signal of interest and  $J$  interferences:  $P = J + 1$ . Denoting

$$\mathbf{U} = [\mathbf{u}_1 \quad \cdots \quad \mathbf{u}_N]$$

we have

$$\begin{aligned} \mathbf{R}^{-1} &= \sum_{n=1}^{J+1} \frac{1}{\lambda_n + \sigma^2} \mathbf{u}_n \mathbf{u}_n^H + \sum_{n=J+2}^N \frac{1}{\sigma^2} \mathbf{u}_n \mathbf{u}_n^H \\ &= \sum_{n=1}^{J+1} \frac{1}{\lambda_n + \sigma^2} \mathbf{u}_n \mathbf{u}_n^H + \sum_{n=1}^N \frac{1}{\sigma^2} \mathbf{u}_n \mathbf{u}_n^H - \sum_{n=1}^{J+1} \frac{1}{\sigma^2} \mathbf{u}_n \mathbf{u}_n^H \\ &= \frac{1}{\sigma^2} \mathbf{I}_N - \frac{1}{\sigma^2} \sum_{n=1}^{J+1} \frac{\lambda_n}{\lambda_n + \sigma^2} \mathbf{u}_n \mathbf{u}_n^H \end{aligned}$$

## Consequence on MPDR (or MVDR)

The beamformer is expressed as

$$\mathbf{w}_{\text{MPDR}} = \frac{\mathbf{R}^{-1} \mathbf{a}_0}{\mathbf{a}_0^H \mathbf{R}^{-1} \mathbf{a}_0}$$

leading

$$\begin{aligned} g_{\text{MPDR}}(\theta) &= \left( \frac{\mathbf{R}^{-1} \mathbf{a}_0}{\mathbf{a}_0^H \mathbf{R}^{-1} \mathbf{a}_0} \right)^H \mathbf{a}(\theta) \\ &= \alpha \mathbf{a}_0^H (\mathbf{R}^{-1})^H \mathbf{a}(\theta) \\ &= \alpha \mathbf{a}_0^H \left( \frac{1}{\sigma^2} \mathbf{I}_N - \frac{1}{\sigma^2} \sum_{n=1}^{J+1} \frac{\lambda_n}{\lambda_n + \sigma^2} \mathbf{u}_n \mathbf{u}_n^H \right)^H \mathbf{a}(\theta) \end{aligned}$$

## Interpretation of DL and the choice of $\mu$

Beampattern with true covariance matrix

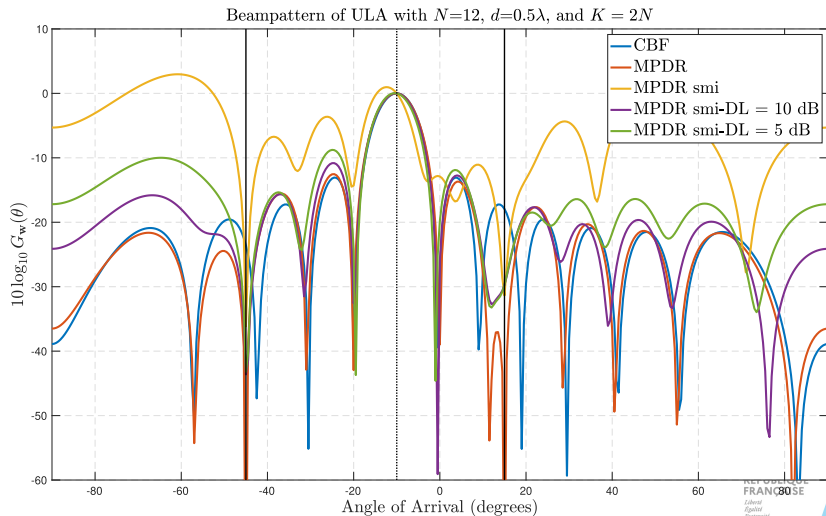
$$g_{\text{MPDR}}(\theta) = \frac{\alpha}{\sigma^2} \left( \mathbf{a}_0^H \mathbf{a}(\theta) - \sum_{n=1}^J \frac{\lambda_n}{\lambda_n + \sigma^2} (\mathbf{a}_0^H \mathbf{u}_n)(\mathbf{u}_n^H \mathbf{a}(\theta)) \right)$$

With estimated covariance matrix

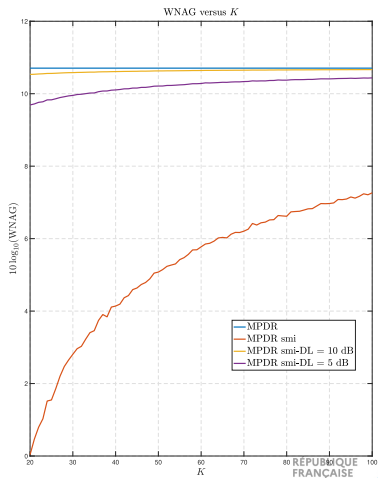
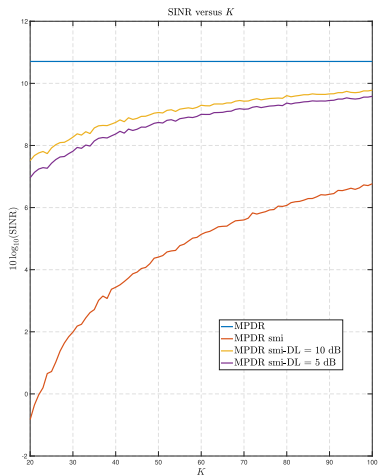
$$g_{\text{MPDR}}^{\text{smi}}(\theta) = \frac{\alpha}{\hat{\lambda}_{\min}} \left( \mathbf{a}_0^H \mathbf{a}(\theta) - \sum_{n=1}^N \frac{\hat{\lambda}_n}{\hat{\lambda}_n + \hat{\lambda}_{\min}} (\mathbf{a}_0^H \hat{\mathbf{u}}_n)(\hat{\mathbf{u}}_n^H \mathbf{a}(\theta)) \right)$$

- degradation is due to  $\hat{\lambda}_{J+2} \neq \hat{\lambda}_{J+3} \neq \dots \neq \hat{\lambda}_N = \hat{\lambda}_{\min}$
- replacing  $\hat{\mathbf{R}}$  with  $\hat{\mathbf{R}} + \mu \mathbf{I}_N$  replaces  $\hat{\lambda}_n$  with  $\hat{\lambda}_n + \mu$
- if  $\mu \gg \sigma^2$  and  $\mu < \lambda_{J+1}$ : equalization of last eigenvalues

# Diagonal loading: beampatterns

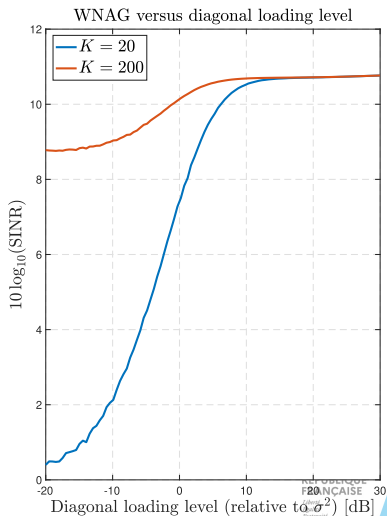
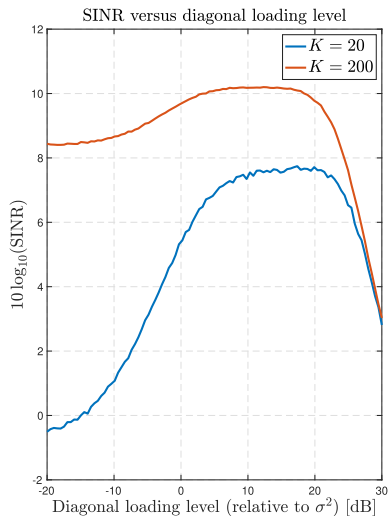


# Diagonal loading: influence of $K$ on SINR and WNAG





# Diagonal loading: influence of $\mu$ on SINR and WNAG



# Linearly constrained beamforming

To mitigate pointing errors: multiple constraints

$$\min_{\mathbf{w}} \mathbf{w}^H \mathbf{C} \mathbf{w} \text{ subject to } \mathbf{Z}^H \mathbf{w} = \mathbf{d}$$

Solution:  $\mathbf{w} = \mathbf{C}^{-1} \mathbf{Z} (\mathbf{Z}^H \mathbf{C}^{-1} \mathbf{Z})^{-1} \mathbf{d}$

Constraint exemples

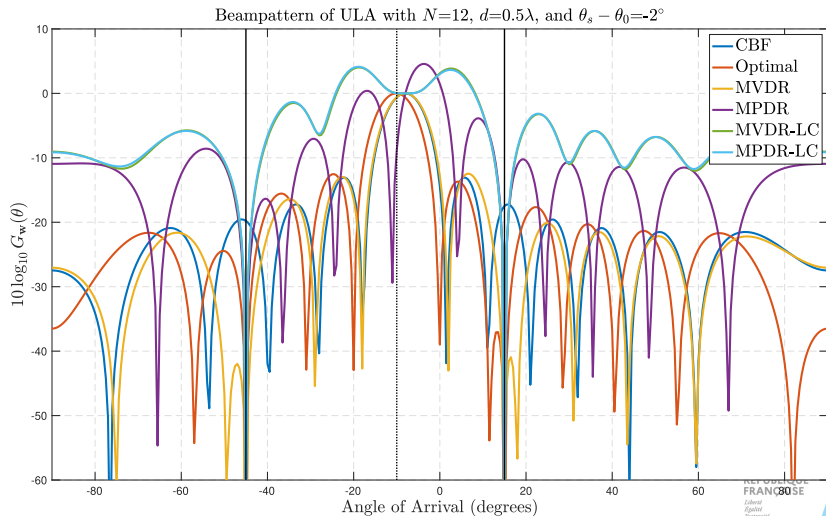
- unit gain around presumed DOA

$$\mathbf{Z} = [\mathbf{a}(\theta_0) \quad \mathbf{a}(\theta_0 + \delta_1) \quad \cdots \quad \mathbf{a}(\theta_0 + \delta_L)], \quad \mathbf{d} = [1 \quad 1 \quad \cdots \quad 1]^T$$

- smoothness around presumed DOA

$$\mathbf{Z} = \left[ \mathbf{a}(\theta_0) \quad \frac{\partial \mathbf{a}(\theta)}{\partial \theta} \quad \cdots \quad \frac{\partial^L \mathbf{a}(\theta)}{\partial \theta^L} \right], \quad \mathbf{d} = [1 \quad 0 \quad \cdots \quad 0]^T$$

# Linearly constrained beamforming: beampatterns



# Plan

## ① Introduction

## ② Model

## ③ Beamforming

- Principle

- Array beampattern

- Spatial filtering

- Adaptive beamforming

- Robust adaptive beamforming

- Partially adaptive beamforming**

- Summary

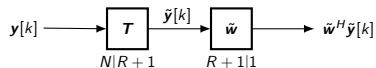
## ④ Direction of arrival estimation

## ⑤ Conclusion

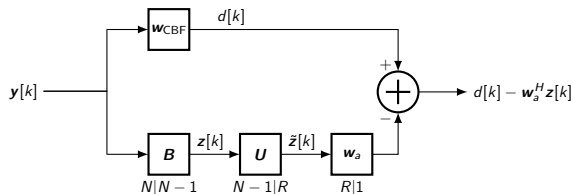
# Partially adaptive beamforming

Principle: work in a lower dimensional subspace

- Direct form



- GSC form ( $\mathbf{T} = [\mathbf{w}_{\text{CBF}} \quad \mathbf{B}\mathbf{U}]$ )

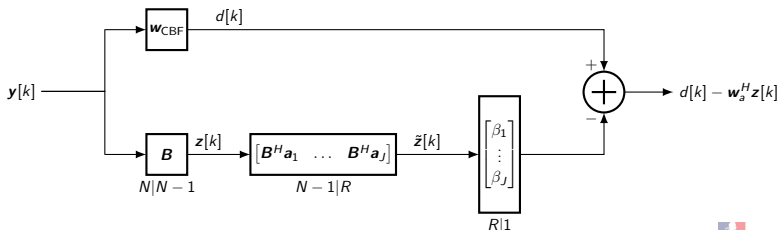


## Motivation for partially adaptive beamforming

When  $\mathbf{C} = \sum_{j=1}^J P_j \mathbf{a}_j \mathbf{a}_j^H + \sigma^2 \mathbf{I}_N$ , one can show

$$\mathbf{w}_{\text{opt}} = \mathbf{w}_{\text{CBF}} - \mathbf{B} \sum_{j=1}^J \beta_j \mathbf{B}^H \mathbf{a}_j$$

The optimal beamformer subtracts a linear combinations  $J$  beams steered towards interference



The optimal beamformer is a partially adaptive beamformer!

# Optimality of partially adaptive beamformer

Is it possible to obtain  $\mathbf{w}_{\text{opt}}$ ?

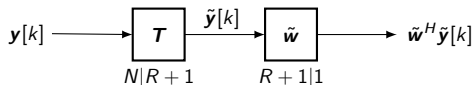
## Direct form

- $\mathbf{w}_{\text{PA-DF}} = \mathbf{w}_{\text{opt}} \Leftrightarrow \mathbf{C}^{-1} \mathbf{a}_s \in \mathcal{R}(\mathbf{T})$
- Not useful: if  $\mathbf{C}^{-1} \mathbf{a}_s$  is known, optimal beamformer can be implemented
- if  $\mathbf{C} = \sum_{j=1}^J P_j \mathbf{a}_j \mathbf{a}_j^H + \sigma^2 \mathbf{I}_N \Rightarrow \mathbf{C}^{-1} \mathbf{a}_s = \eta_s \mathbf{a}_s + \sum_{j=1}^J \eta_j \mathbf{a}_j$
- so if  $[\mathbf{a}_s \quad \mathbf{a}_1 \quad \dots \quad \mathbf{a}_J] \in \mathcal{R}(\mathbf{T})$  then  $\mathbf{w}_{\text{PA-DF}} = \mathbf{w}_{\text{opt}}$

## GSC form

- $\mathbf{w}_{\text{PA-GSC}} = \mathbf{w}_{\text{opt}} \Leftrightarrow \mathbf{B}^H \mathbf{C}^{-1} \mathbf{a}_s \in \mathcal{R}(\mathbf{U})$
- if  $[\mathbf{B}^H \mathbf{a}_1 \quad \dots \quad \mathbf{B}^H \mathbf{a}_J] \in \mathcal{R}(\mathbf{U})$  then  $\mathbf{w}_{\text{PA-GSC}} = \mathbf{w}_{\text{opt}}$

## Expression of partially adaptive beamformer



### Direct form

- Minimize output power

$$\min_{\tilde{w}} \tilde{w}^H R_{\tilde{y}} \tilde{w} \quad \text{subject to} \quad \tilde{w}^H \tilde{a}_0 = 1$$

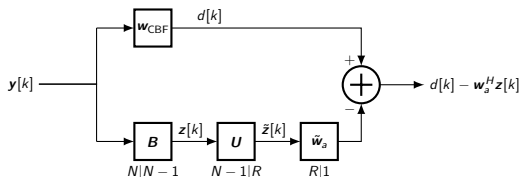
where  $R_{\tilde{y}} = T^H R_y T$  and  $\tilde{a}_0 = T^H a_0$

- Solution

$$\tilde{w} = \alpha R_{\tilde{y}}^{-1} \tilde{a}_0 \Rightarrow w_{\text{PA-DF}} = \alpha T R_y^{-1} a_0$$



## Expression of partially adaptive beamformer



### GSC form

- Minimize output power

$$\min_{\tilde{\mathbf{w}}_a} \mathbb{E} \left\{ |d[k] - \tilde{\mathbf{w}}_a^H \tilde{\mathbf{z}}[k]|^2 \right\}$$

- Solution ( $\mathbf{r}_{dz} = \mathbb{E} \{ \mathbf{z}[k] d^*[k] \}$ )

$$\tilde{\mathbf{w}}_a = \mathbf{R}_{\tilde{\mathbf{z}}}^{-1} \mathbf{r}_{d\tilde{\mathbf{z}}} = \left( \mathbf{U}^H \mathbf{R}_z \mathbf{U} \right)^{-1} \mathbf{U}^H \mathbf{r}_{dz} \Rightarrow \mathbf{w}_{\text{PA-GSC}} = \mathbf{w}_{\text{CBF}} - \mathbf{B} \mathbf{U} \mathbf{R}_{\tilde{\mathbf{z}}}^{-1} \mathbf{r}_{d\tilde{\mathbf{z}}}$$

## Selection of $\mathbf{T}$ and $\mathbf{U}$

Fixed transformations: pre-steered beams

$$\mathbf{T} = \left[ \mathbf{a}(\theta_s) \quad \mathbf{a}(\tilde{\theta}_1) \quad \mathbf{a}(\tilde{\theta}_2) \quad \dots \quad \mathbf{a}(\tilde{\theta}_R) \right]$$
$$\mathbf{U} = \mathbf{B}^H \left[ \mathbf{a}(\tilde{\theta}_1) \quad \mathbf{a}(\tilde{\theta}_2) \quad \dots \quad \mathbf{a}(\tilde{\theta}_R) \right]$$

Requires prior knowledge about interferences DOA

## Selection of $\mathbf{T}$ and $\mathbf{U}$

### Adaptive transformations

- optimal beamformer:  $\mathbf{w}_{\text{opt}} = \mathbf{w}_{\text{CBF}} - \mathbf{B} \sum_{j=1}^J \beta_j \mathbf{B}^H \mathbf{a}_j$
- only need a basis for  $\mathcal{R} \{ [\mathbf{B}^H \mathbf{a}_1 \ \dots \ \mathbf{B}^H \mathbf{a}_J] \}$
- for  $\mathbf{R}_z = \mathbf{B}^H \mathbf{C} \mathbf{B}$

$$\begin{aligned} \mathbf{R}_z &= \sum_{j=1}^J P_j (\mathbf{B}^H \mathbf{a}_j) (\mathbf{B}^H \mathbf{a}_j)^H + \sigma^2 \mathbf{I}_{N-1} \\ &= \sum_{n=1}^{N-1} \lambda_n \mathbf{q}_n \mathbf{q}_n^H \quad \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{N-1} \end{aligned}$$

- Property:  $\mathcal{R} \{ [\mathbf{B}^H \mathbf{a}_1 \ \dots \ \mathbf{B}^H \mathbf{a}_J] \} = \mathcal{R} \{ [\mathbf{q}_1 \ \dots \ \mathbf{q}_J] \}$

## Selection of $\mathbf{T}$ and $\mathbf{U}$

### Principal Component

Choose the  $R$  principal eigenvectors of  $\mathbf{R}_z$ ,  $\mathbf{U} = [\mathbf{q}_1 \ \dots \ \mathbf{q}_R]$  leading

$$\mathbf{R}_z = \mathbf{U}^H \mathbf{R}_z \mathbf{U} = \mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_R)$$

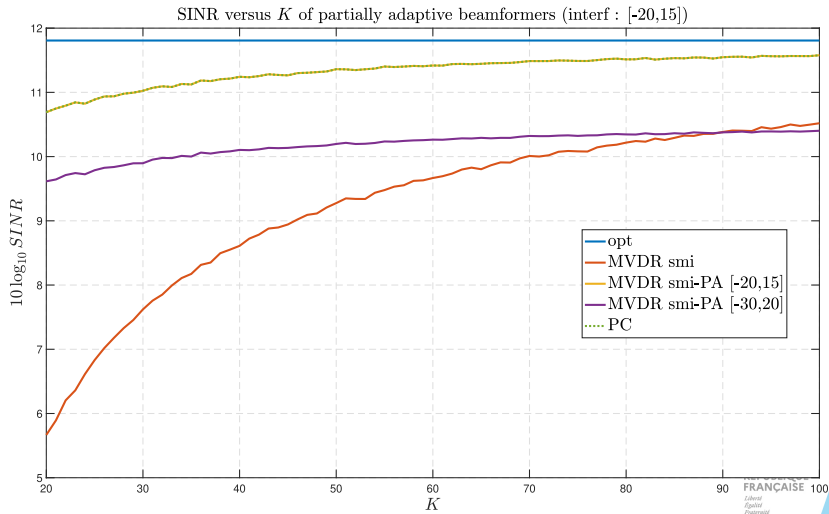
and

$$\mathbf{w}_{\text{PA-GSC}} = \mathbf{w}_{\text{CBF}} - \mathbf{B}\mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{r}_{dz}$$

### Cross Spectral Metric

Choose the  $R$  eigenvectors which contribute most to increase the SINR

# Partially adaptive beamforming: influence of $K$



# Plan

## ① Introduction

## ② Model

## ③ Beamforming

- Principle

- Array beampattern

- Spatial filtering

- Adaptive beamforming

- Robust adaptive beamforming

- Partially adaptive beamforming

- Summary**

## ④ Direction of arrival estimation

## ⑤ Conclusion

## Beamforming: synthesis

- Conventional beamforming:  $\mathbf{w}_{\text{CBF}} = (\mathbf{a}_0^H \mathbf{a}_0)^{-1} \mathbf{a}_0$ , optimal in white noise,  $\theta_{3\text{dB}} \approx 0.9 (N \frac{d}{\lambda})^{-1}$
- Adaptive beamforming:
  - $\mathbf{w}_{\text{opt}} \propto \mathbf{C}^{-1} \mathbf{a}_s$ ,  $\mathbf{w}_{\text{MVDR}} \propto \mathbf{C}^{-1} \mathbf{a}_0$ ,  $\mathbf{w}_{\text{MPDR}} \propto \mathbf{R}^{-1} \mathbf{a}_0$ ,
  - all equivalent if  $\mathbf{R}$ ,  $\mathbf{C}$  known and  $\mathbf{a}_0 = \mathbf{a}_s$
  - $\text{SINR}_{\text{opt}} \geq \text{SINR}_{\text{MVDR}} \gg \text{SINR}_{\text{MPDR}}$  when  $\mathbf{a}_s \neq \mathbf{a}_0$
  - $\text{SINR}_{\text{MVDR-SMI}} \gg \text{SINR}_{\text{MPDR-SMI}}$  and convergence for about  $2N$  snapshots for MVDR and  $N \times \text{SINR}_{\text{opt}}$  for MPDR
- Diagonal loading: helps to mitigate finite sample errors and steering vector errors. Especially useful in MPDR context with low power signal of interest
- Partially adaptive beamforming: enables to achieve faster convergence by operating in low-dimensional subspace. Especially effective with strong, low-rank interference

# Plan

① Introduction

② Model

③ Beamforming

④ Direction of arrival estimation

- Problem formulation

- Non parametric methods

- Parametric methods

- Sparse estimation

⑤ Conclusion



# Plan

① Introduction

② Model

③ Beamforming

④ Direction of arrival estimation

Problem formulation

Non parametric methods

Parametric methods

Sparse estimation

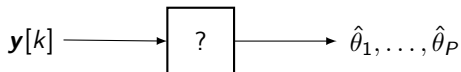
⑤ Conclusion

## DOA estimation problem

Collection of  $K$  snapshots of  $P$  signals impinging the array

$$\mathbf{y}[k] = \sum_{p=1}^P \mathbf{a}(\theta_p) s_p[k] + \mathbf{e}[k] = \mathbf{A}(\boldsymbol{\theta}) \mathbf{s}[k] + \mathbf{e}[k]$$

How to retrieve  $\boldsymbol{\theta} = \{\theta_1, \dots, \theta_P\}$ ?



### Approaches

- Non parametric approaches: don't rely on a model for  $\mathbf{y}[k]$  (similar to Fourier in time domain)
- Parametric approaches: a model is assumed and its properties are exploited (algebraic structure, distribution)

# Plan

① Introduction

② Model

③ Beamforming

④ Direction of arrival estimation

Problem formulation

**Non parametric methods**

Parametric methods

Sparse estimation

⑤ Conclusion

## Beamforming for direction finding purposes

Idea: form a beam  $\mathbf{w}(\theta)$  for each angle  $\theta$  and evaluate the power at the output

$$\mathbf{y}[k] \longrightarrow \boxed{\mathbf{w}(\theta)} \longrightarrow P_{\mathbf{w}}(\theta) = \mathbb{E} \left\{ |\mathbf{w}^H(\theta)\mathbf{y}[k]|^2 \right\} = \mathbf{w}^H(\theta)\mathbf{R}_y\mathbf{w}(\theta)$$

Large peaks should correspond to the DOA

## CBF for direction finding purposes

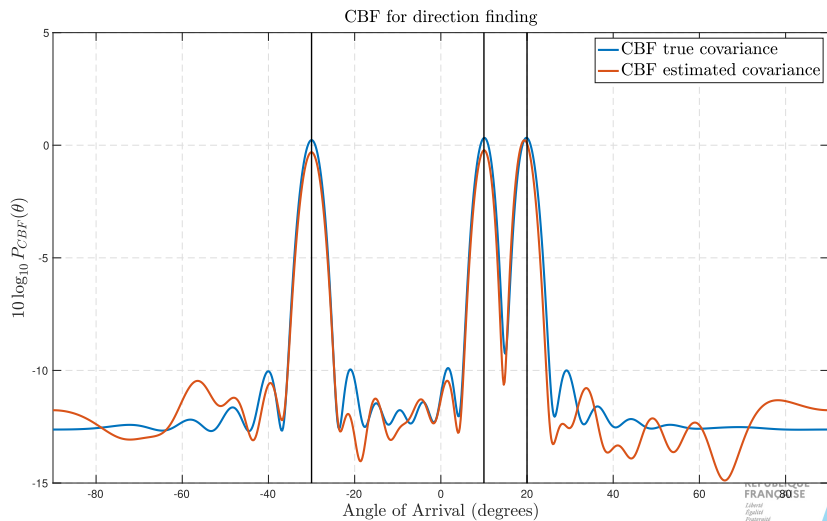
Conventional beamformer

$$\mathbf{w}_{\text{CBF}}(\theta) = \frac{\mathbf{a}(\theta)}{N} \Rightarrow \mathbf{w}_{\text{CBF}}^H(\theta) \mathbf{R} \mathbf{w}_{\text{CBF}}(\theta) = \frac{\mathbf{a}^H(\theta) \mathbf{R} \mathbf{a}(\theta)}{N^2}$$

In practice,  $\mathbf{R}$  is unknown

$$\hat{\mathbf{R}} = \frac{1}{K} \sum_{k=1}^K \mathbf{y}[k] \mathbf{y}^H[k] \Rightarrow P_{\text{CBF}}(\theta) = \frac{\mathbf{a}^H(\theta) \hat{\mathbf{R}} \mathbf{a}(\theta)}{N^2}$$

# CBF for direction finding purposes



## Interpretation of CBF for direction finding

The estimated power at the output of the CBF is

$$\begin{aligned} P_{\text{CBF}}(\theta) &= \frac{1}{N^2} \mathbf{a}^H(\theta) \hat{\mathbf{R}} \mathbf{a}(\theta) \\ &= \frac{1}{KN^2} \sum_{k=1}^K \left| \mathbf{a}^H(\theta) \mathbf{y}[k] \right|^2 \\ &= \frac{1}{KN^2} \sum_{k=1}^K \left| \sum_{n=1}^N y_n[k] e^{-j2\pi(n-1)f_s} \right|^2 \end{aligned}$$

with  $f_s = \frac{d}{\lambda} \sin \theta$

## Interpretation of CBF for direction finding

The estimated power at the output of the CBF is

$$\begin{aligned}
 P_{\text{CBF}}(\theta) &= \frac{1}{N^2} \mathbf{a}^H(\theta) \hat{\mathbf{R}} \mathbf{a}(\theta) \\
 &= \frac{1}{KN^2} \sum_{k=1}^K \left| \mathbf{a}^H(\theta) \mathbf{y}[k] \right|^2 \\
 &= \frac{1}{KN^2} \sum_{k=1}^K \left| \sum_{n=1}^N y_n[k] e^{-j2\pi(n-1)f_s} \right|^2
 \end{aligned}$$

with  $f_s = \frac{d}{\lambda} \sin \theta$

⇒ Spatial Fourier transform of each snapshot



# MPDR for direction finding purposes: Capon's method

MPDR beamformer

$$\mathbf{w}_{\text{MPDR}}(\theta) = \frac{\mathbf{R}^{-1} \mathbf{a}(\theta)}{\mathbf{a}^H(\theta) \mathbf{R}^{-1} \mathbf{a}(\theta)}$$

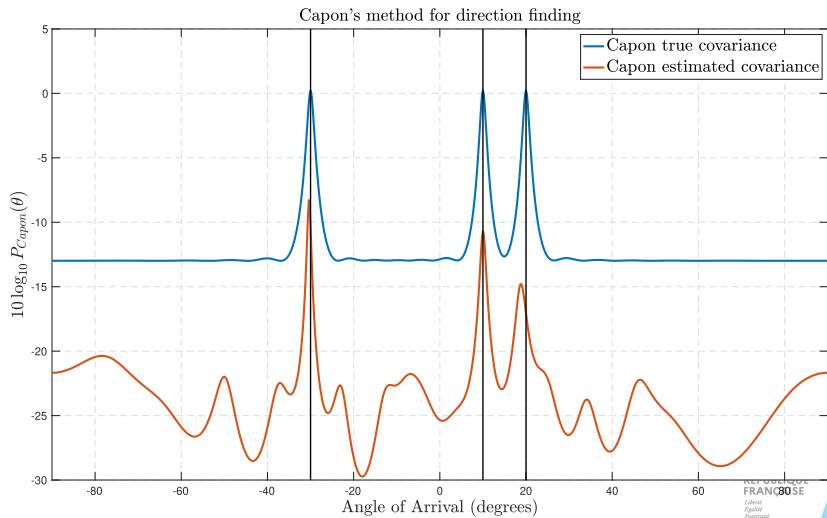
leads

$$\mathbf{w}_{\text{MPDR}}^H(\theta) \mathbf{R} \mathbf{w}_{\text{MPDR}}(\theta) = \frac{\mathbf{a}^H(\theta) \mathbf{R}^{-1} \mathbf{R} \mathbf{R}^{-1} \mathbf{a}(\theta)}{(\mathbf{a}^H(\theta) \mathbf{R}^{-1} \mathbf{a}(\theta))^2} = \frac{1}{\mathbf{a}^H(\theta) \mathbf{R}^{-1} \mathbf{a}(\theta)}$$

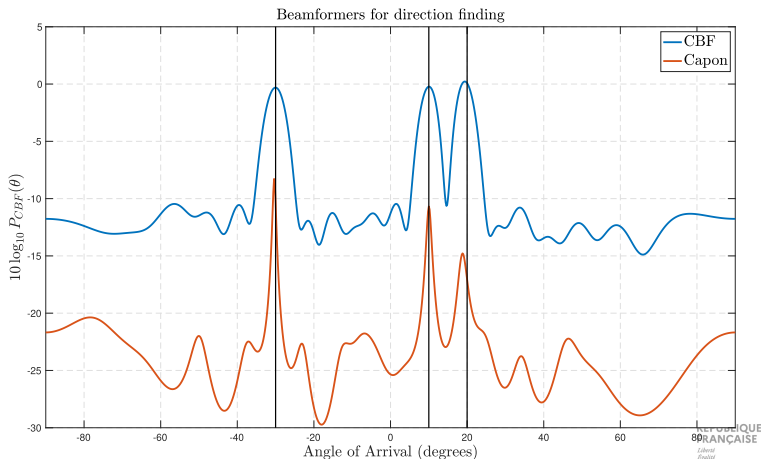
In practice,

$$P_{\text{Capon}}(\theta) = \frac{1}{\mathbf{a}^H(\theta) \hat{\mathbf{R}}^{-1} \mathbf{a}(\theta)}$$

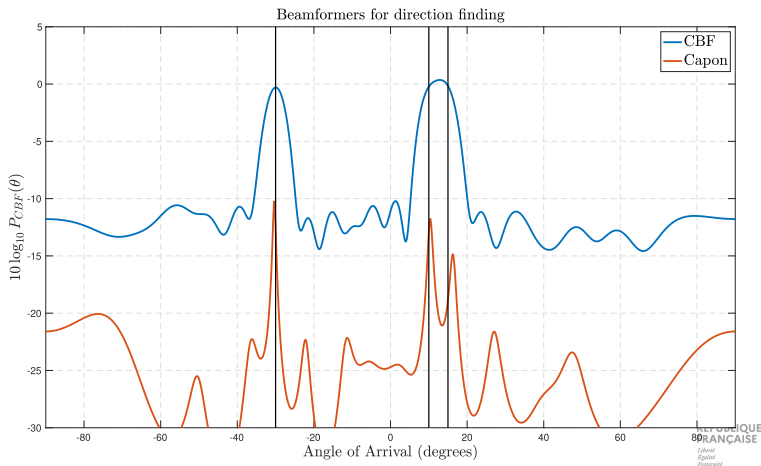
# Capon for direction finding purposes



# Comparison of beamformers for direction finding purposes (low resolution)



# Comparison of beamformers for direction finding purposes (high resolution)



# Plan

① Introduction

② Model

③ Beamforming

④ Direction of arrival estimation

Problem formulation

Non parametric methods

**Parametric methods**

Maximum likelihood estimation

Subspace-based methods

Covariance fitting

Sparse estimation

⑤ Conclusion

# Model-based methods

Principle: based on the model

$$\mathbf{y}[k] = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}[k] + \mathbf{e}[k]$$

with

$$\begin{aligned}\boldsymbol{\theta} &= [\theta_1 \quad \dots \quad \theta_P]^T \\ \mathbf{A}(\boldsymbol{\theta}) &= [\mathbf{a}(\theta_1) \quad \dots \quad \mathbf{a}(\theta_P)] \\ \mathbf{s}[k] &= [s_1[k] \quad \dots \quad s_P[k]]^T\end{aligned}$$

# Classes of methods

- **Maximum likelihood methods:** maximize the likelihood function  $\rightarrow$  find the parameters that best explain the data
- **Subspace-based methods:** signal subspace coincides with the subspace spanned by the principal eigenvectors of  $\mathbf{R}$ , which is orthogonal to the one spanned by the minor eigenvectors
- **Covariance matching:** relies on a model  $\mathbf{R}(\eta)$  of the covariance matrix, finds the parameter that makes it the closest to  $\hat{\mathbf{R}}$

## Reminder: MLE

We have  $K$  independent observations from the same distribution parametrized by  $\boldsymbol{\eta}$

$$\mathbf{y}_k \sim p(\mathbf{y}_k; \boldsymbol{\eta}), \quad k = 1, \dots, K$$

The likelihood is defined as

$$\begin{aligned} \mathcal{L}(\boldsymbol{\eta}; \mathbf{y}_1, \dots, \mathbf{y}_K) &= p(\mathbf{y}_1, \dots, \mathbf{y}_K; \boldsymbol{\eta}) \\ &= \prod_{k=1}^K p(\mathbf{y}_k; \boldsymbol{\eta}) \end{aligned}$$

MLE: find  $\boldsymbol{\eta}$  that best explains the observations

$$\begin{aligned} \hat{\boldsymbol{\eta}}_{\text{ML}} &= \arg \max_{\boldsymbol{\eta}} \mathcal{L}(\boldsymbol{\eta}; \mathbf{y}_1, \dots, \mathbf{y}_K) \\ &= \arg \min_{\boldsymbol{\eta}} (-\log \mathcal{L}(\boldsymbol{\eta}; \mathbf{y}_1, \dots, \mathbf{y}_K)) \end{aligned}$$



# Reminder: MLE

## Advantages

- asymptotically unbiased

$$\lim_{K \rightarrow +\infty} \mathbb{E} \{ \hat{\boldsymbol{\eta}}_{\text{ML}} \} = \boldsymbol{\eta}$$

- asymptotically efficient

$$\lim_{K \rightarrow +\infty} \text{Var} ( \hat{\boldsymbol{\eta}}_{\text{ML}} ) = \text{CRB} ( \boldsymbol{\eta} )$$

## Drawback

Usually multi-dimensional optimization problem: computationally intensive, possible to reach local maxima

## MLE for direction finding

There are two expressions, based on the signal model

- **Deterministic MLE:** the received signal is assumed deterministic (Conditional signal model)
- **Stochastic MLE:** the received signal is assumed random (Unconditional signal model)

For each model, it is assumed

$$\mathbf{e}[k] \sim \mathcal{CN}(\mathbf{e}[k]; \mathbf{0}, \sigma^2 \mathbf{I}_N)$$

where

$$p(\mathbf{e}; \boldsymbol{\mu}, \mathbf{R}) = \frac{1}{\pi^N |\mathbf{R}|} \exp\left(-(\mathbf{e} - \boldsymbol{\mu})^H \mathbf{R}^{-1} (\mathbf{e} - \boldsymbol{\mu})\right)$$

## Deterministic MLE

We have

$$\mathbf{y}[k] = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}[k] + \mathbf{e}[k]$$

hence

$$\mathbf{y}[k] \sim \mathcal{CN}(\mathbf{y}[k]; \mathbf{A}(\boldsymbol{\theta})\mathbf{s}[k], \sigma^2 \mathbf{I}_N)$$

with unknown parameters  $\boldsymbol{\eta} = \{\sigma^2, \mathbf{s}[1], \dots, \mathbf{s}[K], \boldsymbol{\theta}\}$

### Likelihood

$$\mathcal{L}(\boldsymbol{\eta}; \mathbf{y}_1, \dots, \mathbf{y}_K) = \prod_{k=1}^K \frac{1}{(\sigma^2 \pi)^N} \exp\left(-\frac{1}{\sigma^2} \|\mathbf{y}[k] - \mathbf{A}(\boldsymbol{\theta})\mathbf{s}[k]\|^2\right)$$

# Deterministic MLE

Maximum likelihood estimator

$$\hat{\boldsymbol{\eta}}_{\text{ML}} = \arg \min_{\boldsymbol{\eta}} NK \log(\sigma^2) + \frac{1}{\sigma^2} \sum_{k=1}^K \|\mathbf{y}[k] - \mathbf{A}(\boldsymbol{\theta})\mathbf{s}[k]\|^2$$

leading <sup>4</sup>

$$\hat{\boldsymbol{\theta}}_{\text{ML}}^{\text{det}} = \arg \min_{\boldsymbol{\theta}} \text{Tr} \left( \boldsymbol{\Pi}_{\mathbf{A}(\boldsymbol{\theta})}^{\perp} \hat{\mathbf{R}} \right) = \arg \max_{\boldsymbol{\theta}} \text{Tr} \left( \boldsymbol{\Pi}_{\mathbf{A}(\boldsymbol{\theta})} \hat{\mathbf{R}} \right)$$

For a single source

$$\hat{\theta}_{\text{ML}}^{\text{det}} = \arg \max_{\theta} \mathbf{a}^H(\theta) \hat{\mathbf{R}} \mathbf{a}(\theta) \Rightarrow \text{CBF}$$

<sup>4</sup> $\boldsymbol{\Pi}_{\mathbf{A}} = \mathbf{A} (\mathbf{A}\mathbf{A}^H)^{-1} \mathbf{A}^H$  and  $\boldsymbol{\Pi}_{\mathbf{A}}^{\perp} = \mathbf{I} - \boldsymbol{\Pi}_{\mathbf{A}}$

## Gaussian vectors: reminder

### Property: independence

Let  $\mathbf{x}$  and  $\mathbf{y}$  two Gaussian vectors. Then

$$\mathbf{x} \text{ and } \mathbf{y} \text{ are independent} \Leftrightarrow \text{cov}(\mathbf{x}, \mathbf{y}) = \mathbf{0}$$

### Property: affine transformation

Let  $\mathbf{x} \sim \mathcal{N}(\mathbf{m}, \Sigma) \in \mathbb{R}^n$ ,  $\mathbf{A} \in \mathbb{R}^{p \times n}$  ( $p \leq n$ ) with full rank and  $\mathbf{b} \in \mathbb{R}^p$ , and let  $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$ . Then

$$\mathbf{y} \sim \mathcal{N}(\mathbf{A}\mathbf{m} + \mathbf{b}, \mathbf{A}\Sigma\mathbf{A}^T)$$

These results can be extended to complex Gaussian vectors (with Hermitian instead of transpose)

## Stochastic MLE

We have

$$\mathbf{y}[k] = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}[k] + \mathbf{e}[k]$$

and we assume

$$\mathbf{s}[k] \sim \mathcal{CN}(\mathbf{s}[k]; \mathbf{0}, \mathbf{R}_s)$$

and independent with  $\mathbf{e}[k]$ . Then

$$\mathbf{y}[k] \sim \mathcal{CN}(\mathbf{y}[k]; \mathbf{0}, \mathbf{R} = \mathbf{A}(\boldsymbol{\theta})\mathbf{R}_s\mathbf{A}^H(\boldsymbol{\theta}) + \sigma^2\mathbf{I}_N)$$

with unknown parameters  $\boldsymbol{\eta} = \{\sigma^2, \mathbf{R}_s, \boldsymbol{\theta}\}$

### Likelihood

$$\mathcal{L}(\boldsymbol{\eta}; \mathbf{y}_1, \dots, \mathbf{y}_K) = \prod_{k=1}^K \frac{1}{\pi^N |\mathbf{R}|} \exp\left(-\mathbf{y}[k]^H \mathbf{R}^{-1} \mathbf{y}[k]\right)$$

# Stochastic MLE

## Maximum likelihood estimator

$$\begin{aligned}\hat{\boldsymbol{\eta}}_{\text{ML}} &= \arg \min_{\boldsymbol{\eta}} K \log(|\mathbf{R}|) + \sum_{k=1}^K \mathbf{y}^H[k] \mathbf{R}^{-1} \mathbf{y}[k] \\ &= \arg \min_{\boldsymbol{\eta}} \log(|\mathbf{R}|) + \text{Tr} \left( \mathbf{R}^{-1} \hat{\mathbf{R}} \right)\end{aligned}$$

closed-form expressions for  $\hat{\sigma}^2(\boldsymbol{\theta})$  and  $\hat{\mathbf{R}}_s(\boldsymbol{\theta})$  can be found (book<sup>5</sup>, Eqs. (14.64) and (14.65)) yielding

$$\hat{\boldsymbol{\theta}}_{\text{ML}}^{\text{sto}} = \arg \min_{\boldsymbol{\theta}} \log \left| \mathbf{A}(\boldsymbol{\theta}) \hat{\mathbf{R}}_s(\boldsymbol{\theta}) \mathbf{A}^H(\boldsymbol{\theta}) + \hat{\sigma}^2(\boldsymbol{\theta}) \mathbf{I}_N \right|$$

<sup>5</sup>Rama Chellappa Abdelhak M. Zoubir Mats Viberg and Sergios Theodoridis. *Array and Statistical Signal Processing*. Academic Press, 2014.

# Solving MLE

Several methods can be employed to solve these problems, as

- extensive grid search: try all possible values of  $\theta \rightarrow$  complexity increases exponentially as  $P$  increases
- EM algorithm: find unobserved variables that simplifies the problem, iterate between estimation of these variables and the parameters of interest
- SAGE algorithm: same approach, instead the unobserved variables change  $\rightarrow$  speeds up the previous one

All can be used for conditional or stochastic MLE, here only conditional



# EM<sup>6</sup> approach

## Expectation maximization

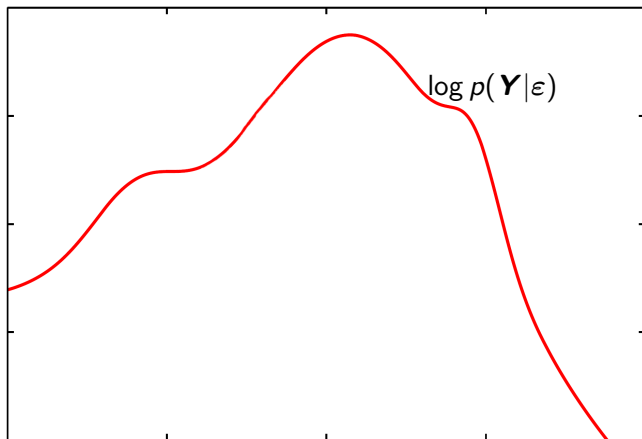
Principle: augment the observations  $\mathbf{Y}$  with imputed values  $\mathbf{X}$  to simplify the maximization, and apply the augmented data to estimate the unknown parameters

Find a many-to-one mapping  $\mathcal{M} : \mathbf{X} \mapsto \mathcal{M}(\mathbf{X}) = \mathbf{Y}$

- EM algorithm for estimating  $\varepsilon$  when observing  $\mathbf{y}$ 
  - E-step: compute  $Q(\varepsilon|\varepsilon^{(t)}) = \mathbb{E}_{\mathbf{X}|\mathbf{Y},\varepsilon^{(t)}}[\log p(\mathbf{X}|\varepsilon)]$
  - M-step: solve  $\varepsilon^{(t+1)} = \arg \max_{\varepsilon} Q(\varepsilon|\varepsilon^{(t)})$

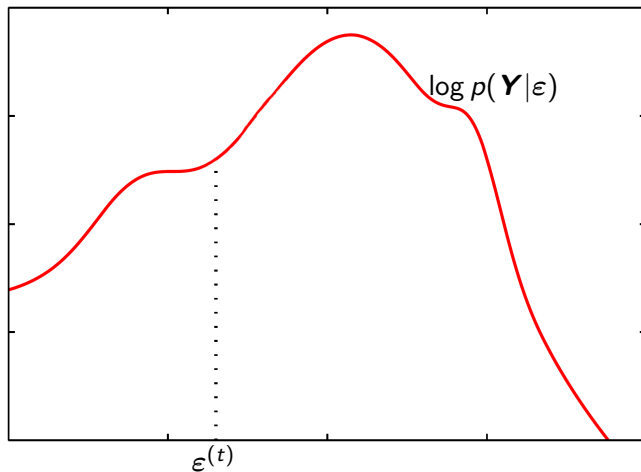
<sup>6</sup>A.P. Dempster, N.M. Laird, and Do. Rubin. "Maximum Likelihood from Incomplete Data via the EM Algorithm". In *Stat. Soc. Series B* 39.1 (1977), pp. 1–38.

# EM principle



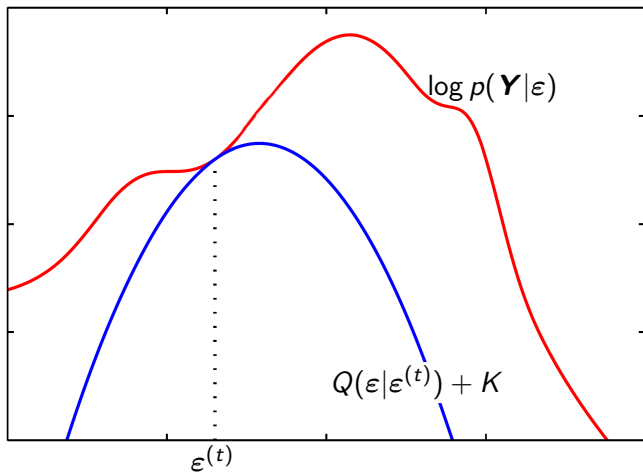
Adapted from Christopher M. Bishop, "Pattern Recognition and Machine Learning," Chap. 9

# EM principle



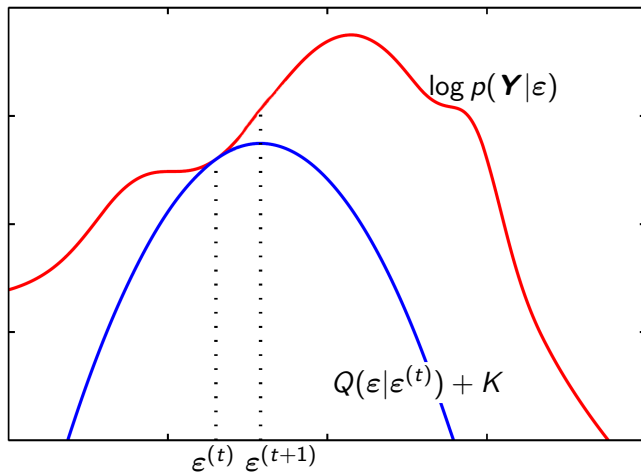
Adapted from Christopher M. Bishop, "Pattern Recognition and Machine Learning," Chap. 9

# EM principle



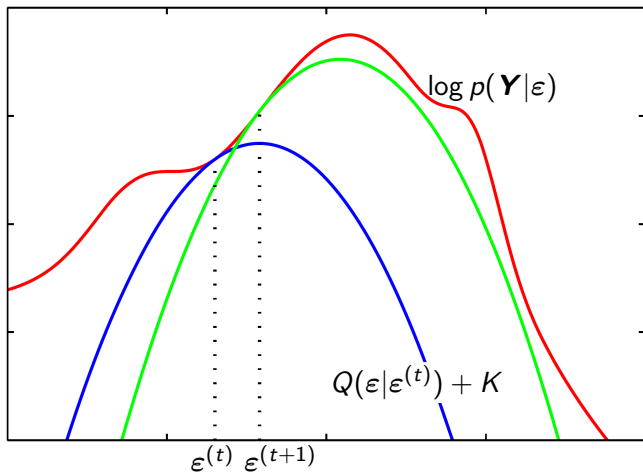
Adapted from Christopher M. Bishop, "Pattern Recognition and Machine Learning," Chap. 9

# EM principle



Adapted from Christopher M. Bishop, "Pattern Recognition and Machine Learning," Chap. 9

# EM principle



Adapted from Christopher M. Bishop, "Pattern Recognition and Machine Learning," Chap. 9

## Application of EM to MLE (conditional)

We define

$$\mathbf{x}[k] = \begin{bmatrix} \mathbf{x}_1[k] \\ \vdots \\ \mathbf{x}_P[k] \end{bmatrix} = \begin{bmatrix} \mathbf{a}(\theta_1)s_1[k] + \mathbf{e}_1[k] \\ \vdots \\ \mathbf{a}(\theta_P)s_P[k] + \mathbf{e}_P[k] \end{bmatrix}$$

such as

$$\mathbf{y}[k] = \sum_{p=1}^P \mathbf{x}_p[k] = [\mathbf{I}_N \quad \dots \quad \mathbf{I}_N] \mathbf{x}[k]$$

and

$$\mathbf{X} = \{\mathbf{x}[1], \dots, \mathbf{x}[K]\}, \quad \mathbf{Y} = \{\mathbf{y}[1], \dots, \mathbf{y}[K]\}$$

## Application of EM to MLE (conditional)

We define the unknown parameters

$$\boldsymbol{\varepsilon} = \{\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_P\}, \quad \boldsymbol{\varepsilon}_p = \{\theta_p, s_p[1], \dots, s_p[K], \sigma_p^2\}$$

and we have

$$\mathbf{x}_p[k] | \boldsymbol{\varepsilon}_p \sim \mathcal{CN}(\mathbf{x}_p[k]; \mathbf{a}(\theta_p) s_p[k], \sigma_p^2 \mathbf{I}_N)$$

where  $\sum_{p=1}^P \sigma_p^2 = \sigma^2$  (e.g.,  $\sigma_p^2 = \frac{\sigma^2}{P}$ ), leading

$$\log p(\mathbf{X} | \boldsymbol{\varepsilon}) = C - \sum_{k=1}^K \sum_{p=1}^P \frac{1}{\sigma_p^2} \|\mathbf{x}_p[k] - \mathbf{a}(\theta_p) s_p[k]\|^2$$



## Application of EM to MLE (conditional)

One can show<sup>7</sup>

$$\mathbb{E}_{\mathbf{X}|\mathbf{Y},\boldsymbol{\varepsilon}^{(t)}}[\mathbf{x}_p[k]] = \hat{\mathbf{x}}_p[k] = \mathbf{a}(\theta_p^{(t)})s_p^{(t)}[k] + \frac{1}{P} \left( \mathbf{y}[k] - \mathbf{A}(\boldsymbol{\theta}^{(t)})\mathbf{s}^{(t)}[k] \right)$$

leading (when not interested in  $\sigma_p^2$ )

$$Q(\boldsymbol{\varepsilon}|\boldsymbol{\varepsilon}^{(t)}) = D - \sum_{k=1}^K \sum_{p=1}^P \frac{1}{\sigma_p^2} \|\hat{\mathbf{x}}_p[k] - \mathbf{a}(\theta_p)s_p[k]\|^2$$

The M step is equivalent to

$$\boldsymbol{\varepsilon}^{(t+1)} = \arg \min_{\boldsymbol{\varepsilon}} \sum_{p=1}^P \sum_{k=1}^K \|\hat{\mathbf{x}}_p[k] - \mathbf{a}(\theta_p)s_p[k]\|^2$$

<sup>7</sup>Rama Chellappa Abdelhak M. Zoubir Mats Viberg and Sergios Theodoridis. *Array and Statistical Signal Processing*. Academic Press, 2014.

# Application of EM to MLE (conditional): wrap up

## EM approach to solve conditional MLE

- 1 initialize  $\epsilon^{(0)}$
- 2 iterate until convergence
  - $\hat{\mathbf{x}}_p[k] \leftarrow \mathbf{a}(\theta_p^{(t)})s_p^{(t)}[k] + \frac{1}{P} (\mathbf{y}[k] - \mathbf{A}(\theta^{(t)})\mathbf{s}^{(t)}[k])$
  - $\hat{\mathbf{R}}_{x_p} \leftarrow \frac{1}{K} \sum_{k=1}^K \hat{\mathbf{x}}_p[k]\hat{\mathbf{x}}_p^H[k]$
  - $\hat{\theta}_p^{(t+1)} = \arg \max_{\theta_p} \mathbf{a}^H(\theta_p)\hat{\mathbf{R}}_{x_p}\mathbf{a}(\theta_p)$
  - $s_p^{(t+1)}[k] = \frac{\mathbf{a}^H(\hat{\theta}_p^{(t+1)})\hat{\mathbf{x}}_p[k]}{N}$

# SAGE<sup>8</sup> approach

## Space Alternating Generalized EM

Principle: break the problem into several smaller problems by conditioning sequentially on a subset of the parameters, then applies EM to each reduced problem.

Each iteration has  $C$  cycles: at cycle  $c$ ,

- E-step: compute  $Q^{(c)}(\epsilon_c | \epsilon^{(t,c-1)}) = \mathbb{E}_{\mathbf{Z} | \mathbf{Y}, \epsilon^{(t,c-1)}} [\log p(\mathbf{Z}_c | \epsilon_c)]$
- M-step: solve  $\epsilon^{(t,c)} = \arg \max_{\epsilon_c} Q^{(c)}(\epsilon_c | \epsilon^{(t,c-1)})$

at the end of the last cycle:  $\epsilon^{(t+1,0)} = \epsilon^{(t,C)}$

<sup>8</sup>J.A. Fessler and A.O. Hero. "Space-Alternating Generalized Expectation-Maximization algorithm". In: *IEEE Transactions on Signal Processing* 42.10 (1994), pp. 2664–2677.

## Application of SAGE to MLE (conditional)

We use  $P$  cycles and define

$$\mathbf{z}_p[k] = \mathbf{a}(\theta_p)s_p[k] + \mathbf{e}[k]$$

Using same reasoning as for EM

$$\hat{\mathbf{z}}_p[k] = \mathbf{a}(\theta_p^{(t,p-1)})s_p^{(t,p-1)}[k] + \mathbf{y}[k] - \mathbf{A}(\theta^{(t,p-1)})\mathbf{s}^{(t,p-1)}[k]$$

Leading the update

$$\theta_p^{(t,c)} = \arg \max_{\theta_p} \mathbf{a}(\theta_p^H) \hat{R}_{z_p} \mathbf{a}(\theta_p)$$

where

$$\hat{R}_{z_p} = \frac{1}{K} \hat{\mathbf{z}}_p[k] \hat{\mathbf{z}}_p^H[k]$$

# Application of SAGE to MLE (conditional): wrap up

## SAGE approach to solve conditional MLE

- 1 initialize  $\boldsymbol{\varepsilon}^{(0,0)}$
- 2 iterate until convergence
  - For  $p = 1, \dots, P$ 
    - $\hat{\mathbf{z}}_p[k] \leftarrow \mathbf{a}(\theta_p^{(t,p-1)})s_p^{(t,p-1)}[k] + \mathbf{y}[k] - \mathbf{A}(\theta^{(t,p-1)})\mathbf{s}^{(t,p-1)}[k]$
    - $\hat{\mathbf{R}}_{z_p} \leftarrow \frac{1}{K} \hat{\mathbf{z}}_p[k] \hat{\mathbf{z}}_p^H[k]$
    - $\hat{\theta}_p^{(t,p)} = \arg \max_{\theta_p} \mathbf{a}(\theta_p^H) \hat{\mathbf{R}}_{z_p} \mathbf{a}(\theta_p)$
    - $s_p^{(t,p)}[k] = \frac{\mathbf{a}^H(\hat{\theta}_p^{(t,p)}) \hat{\mathbf{z}}_p[k]}{N}$
  - $\boldsymbol{\varepsilon}^{(t+1,0)} \leftarrow \boldsymbol{\varepsilon}^{(t,P)}$

# Initialization of EM or SAGE

## Several possibilities

- easy one
  - 1 assume a single source:

$$\hat{\theta}_1^{(0)} = \arg \max_{\theta} \text{Tr} \left( \mathbf{\Pi}_{\mathbf{a}(\theta)} \hat{\mathbf{R}} \right)$$

- 2 augment the steering matrix with a new source:

$$\mathbf{A}(\theta) = \begin{bmatrix} \mathbf{a}(\hat{\theta}_1^{(0)}) & \mathbf{a}(\theta) \end{bmatrix} \text{ and } \hat{\theta}_2^{(0)} = \arg \max_{\theta} \text{Tr} \left( \mathbf{\Pi}_{\mathbf{A}(\theta)} \hat{\mathbf{R}} \right)$$

- 3 continue until the total number of sources is reached
- any other method

## Subspace-based methods: principle

- Theorem (reminder): if  $\mathbf{A} \in \mathbb{C}^N$  is an Hermitian matrix, then its eigenvalues are real and there is an orthonormal basis formed by eigenvectors of  $\mathbf{A}$

Eigenvalue decomposition of the covariance matrix ( $P < N$  signals)

$$\begin{aligned}
 \mathbf{R} &= \mathbf{A}(\boldsymbol{\theta})\mathbf{R}_s\mathbf{A}^H(\boldsymbol{\theta}) + \sigma^2\mathbf{I}_N \quad (\mathbf{R}_s \text{ full rank}) \\
 &= \sum_{p=1}^P \lambda_p \mathbf{u}_p \mathbf{u}_p^H + \sigma^2 \sum_{p=1}^N \mathbf{u}_p \mathbf{u}_p^H \quad (\lambda_1 \geq \dots \geq \lambda_P) \\
 &= \sum_{p=1}^P (\lambda_p + \sigma^2) \mathbf{u}_p \mathbf{u}_p^H + \sigma^2 \sum_{p=P+1}^N \mathbf{u}_p \mathbf{u}_p^H \quad (\lambda_1 \geq \dots \geq \lambda_P) \\
 &= \mathbf{U}_s \boldsymbol{\Lambda}_s \mathbf{U}_s^H + \sigma^2 \mathbf{U}_e \mathbf{U}_e^H
 \end{aligned}$$

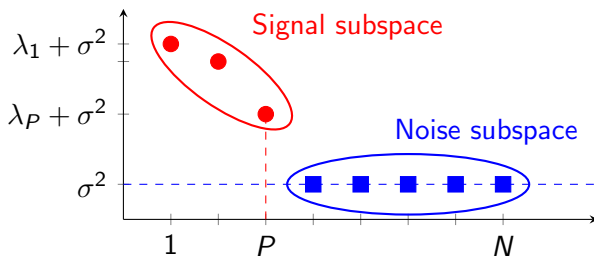
where  $\mathbf{U}_s = [\mathbf{u}_1 \ \dots \ \mathbf{u}_P] \perp \mathbf{U}_e = [\mathbf{u}_{P+1} \ \dots \ \mathbf{u}_N]$

## Signal and noise subspaces

Subspace-based methods rely on either

- $\mathcal{R}\{\mathbf{U}_s\} = \mathcal{R}\{\mathbf{A}(\boldsymbol{\theta})\}$ : the signal subspace is spanned by  $\mathbf{U}_s$  so there is a non-singular matrix  $\mathbf{T}$  such that  $\mathbf{U}_s = \mathbf{A}(\boldsymbol{\theta})\mathbf{T}$
- $\mathcal{R}\{\mathbf{U}_e\}$  is orthogonal to  $\mathcal{R}\{\mathbf{U}_s\} = \mathcal{R}\{\mathbf{A}(\boldsymbol{\theta})\}$  so  $\mathbf{A}^H(\boldsymbol{\theta})\mathbf{U}_e = \mathbf{0}$

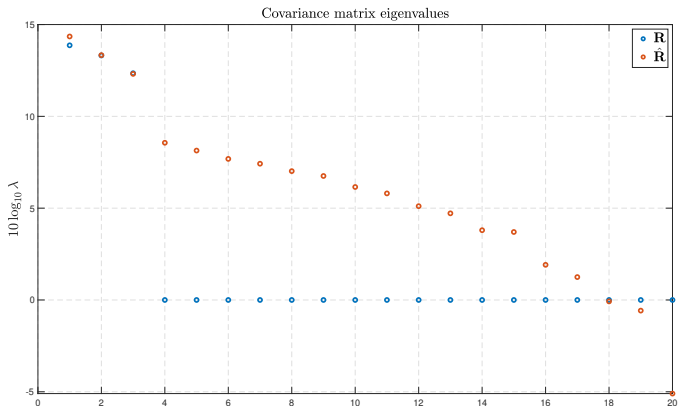
Eigenvalues of the covariance matrix





# Principle

In practice: estimated covariance matrix



# MUSIC (MULTiple Signal Classification)

The steering vectors are orthogonal to  $\mathcal{R}\{\mathbf{U}_e\}$

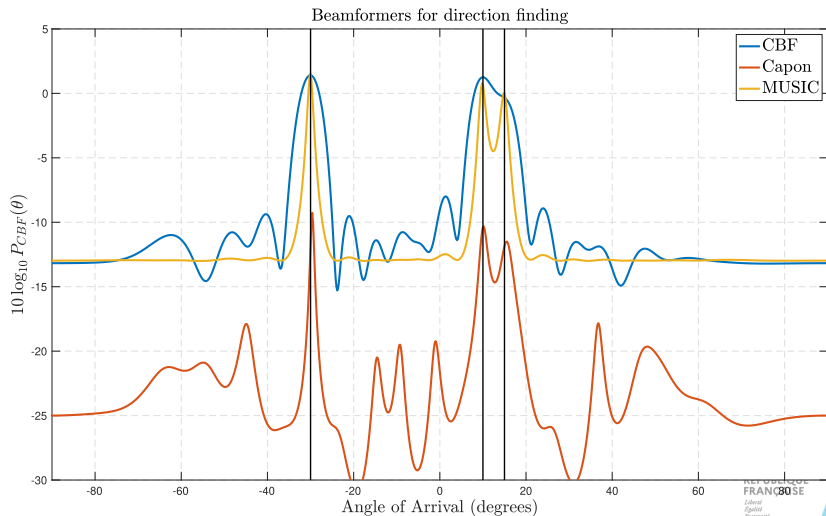
$$\mathbf{U}_e^H \mathbf{a}(\theta_p) = \mathbf{0} \Rightarrow \mathbf{u}_n^H \mathbf{a}(\theta_p) = 0 \text{ for } n = P + 1, \dots, N$$

Principle: look for the  $P$  largest maxima of

$$P_{\text{MUSIC}}(\theta) = \frac{1}{\mathbf{a}^H(\theta) \hat{\mathbf{U}}_e \hat{\mathbf{U}}_e^H \mathbf{a}(\theta)} = \frac{1}{\sum_{n=P+1}^N |\hat{\mathbf{u}}_n^H \mathbf{a}(\theta)|^2}$$

When  $\sigma^2 \rightarrow 0$  or  $K \rightarrow +\infty$  one has  $P_{\text{MUSIC}}(\theta) \rightarrow +\infty$

# MUSIC



# Root-MUSIC

Let  $\mathbf{a}_z(z) = [1 \quad z \quad \dots \quad z^{N-1}]^T$ . For ULA

$$\mathbf{a}(\theta) = \left[ 1 \quad e^{-j2\pi f_c \frac{d \sin \theta}{c}} \quad \dots \quad e^{-j2\pi f_c (N-1) \frac{d \sin \theta}{c}} \right]^T = \mathbf{a}_z \left( e^{-j2\pi f_c \frac{d \sin \theta}{c}} \right)$$

leading for a DOA

$$\mathbf{a}^H(\theta) \hat{\mathbf{U}}_e \hat{\mathbf{U}}_e^H \mathbf{a}(\theta) = \mathbf{a}_z^T \left( e^{j2\pi f_c \frac{d \sin \theta}{c}} \right) \hat{\mathbf{U}}_e \hat{\mathbf{U}}_e^H \mathbf{a}_z \left( e^{-j2\pi f_c \frac{d \sin \theta}{c}} \right) = 0$$

One can look for the roots of

$$P_{\text{MUSIC}}(z) = \mathbf{a}_z^T(z^{-1}) \hat{\mathbf{U}}_e \hat{\mathbf{U}}_e^H \mathbf{a}_z(z)$$

closest to the unit circle

## Root-MUSIC

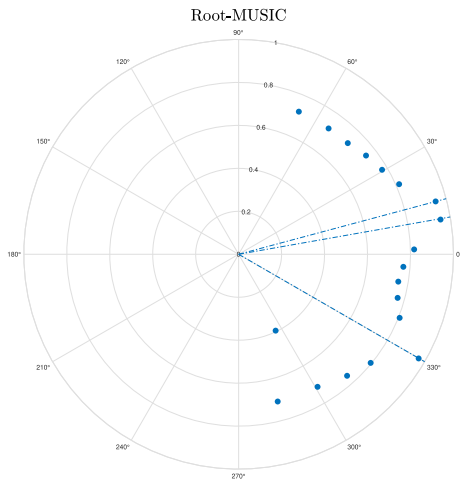
Moreover,  $P_{\text{MUSIC}}(z) = \sum_{n=-(N-1)}^{N-1} p_n z^n$  has  $2(N-1)$  roots, but  $\mathbf{a}_z(z^*) = \mathbf{a}_z^*(z)$  leads

$$\begin{aligned}
 P_{\text{MUSIC}}^* \left( \frac{1}{z^*} \right) &= \left( \mathbf{a}_z^T(z^*) \hat{\mathbf{U}}_e \hat{\mathbf{U}}_e^H \mathbf{a}_z \left( \frac{1}{z^*} \right) \right)^* \\
 &= \mathbf{a}_z^T(z) \hat{\mathbf{U}}_e^* \hat{\mathbf{U}}_e^T \mathbf{a}_z(z^{-1}) \\
 &= \mathbf{a}_z^T(z^{-1}) \hat{\mathbf{U}}_e \hat{\mathbf{U}}_e^H \mathbf{a}_z(z) \\
 &= P_{\text{MUSIC}}(z)
 \end{aligned}$$

so only  $(N-1)$  are inside the unit circle

No need to try several values of  $\theta$ , directly given by the roots

# Root-MUSIC



# ESPRIT (Estimation of Signal Parameters via Rotational Invariance Techniques)

Partition of  $\mathbf{A} = \mathbf{A}(\theta)$

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 \\ - \end{bmatrix} = \begin{bmatrix} - \\ \mathbf{A}_2 \end{bmatrix}$$

where  $\mathbf{A}_1$  (resp.  $\mathbf{A}_2$ ) contains all but last (resp. first) row of  $\mathbf{A}$ .

For ULA

$$\mathbf{A}_2 = \mathbf{A}_1 \Phi \text{ with } \Phi = \text{diag} \left( \left\{ e^{-j2\pi \frac{d}{\lambda} \sin \theta_p} \right\}_{p=1}^P \right)$$

The DOAs are in  $\Phi$ , which can be deduced from  $\mathbf{A}_1$  and  $\mathbf{A}_2$  which are unknown

# ESPRIT

We know  $\mathbf{U}_s = \mathbf{A}\mathbf{T}$ . We can decompose  $\mathbf{U}_s = \begin{bmatrix} \mathbf{U}_{s,1} \\ - \end{bmatrix} = \begin{bmatrix} - \\ \mathbf{U}_{s,2} \end{bmatrix}$  leading

$$\begin{cases} \mathbf{U}_{s,1} = \mathbf{A}_1 \mathbf{T} \\ \mathbf{U}_{s,2} = \mathbf{A}_2 \mathbf{T} = \mathbf{A}_1 \Phi \mathbf{T} \end{cases}$$

$$\Rightarrow \mathbf{U}_{s,2} = \mathbf{U}_{s,1} \mathbf{T}^{-1} \Phi \mathbf{T}$$

$$\Rightarrow \mathbf{U}_{s,2} = \mathbf{U}_{s,1} \Psi$$

Moreover, if  $\mathbf{u}$  is an eigenvector of  $\Phi$ , then  $\Psi \mathbf{T}^{-1} \mathbf{u} = \mathbf{T}^{-1} \Phi \mathbf{u} = \lambda \mathbf{T}^{-1} \mathbf{u}$

$\rightarrow \Phi$  and  $\Psi$  share the same eigenvalues  $e^{-j2\pi \frac{d}{\lambda} \sin \theta_p}$ ,  $p = 1, \dots, P$



# ESPRIT

In practice, one solves  $\hat{\mathbf{U}}_{s,2} = \hat{\mathbf{U}}_{s,1} \Psi$  with a least squares approach

$$\hat{\Psi} = (\hat{\mathbf{U}}_{s,1}^H \hat{\mathbf{U}}_{s,1})^{-1} \hat{\mathbf{U}}_{s,1}^H \hat{\mathbf{U}}_{s,2}$$

and compute the eigenvalues of  $\hat{\Psi}$

As root-MUSIC, no need for a grid

## Subspace fitting

Principle: minimize error between subspaces spanned by  $\hat{\mathbf{U}}_s$  and  $\mathbf{A}(\theta)$  using a weighted least square criterion

$$\begin{aligned}\hat{\theta}, \hat{\mathbf{T}} &= \arg \min_{\theta, \mathbf{T}} \|\hat{\mathbf{U}}_s - \mathbf{A}(\theta)\mathbf{T}\|_W^2 \\ &= \arg \min_{\theta, \mathbf{T}} \text{Tr} \left( (\hat{\mathbf{U}}_s - \mathbf{A}(\theta)\mathbf{T})\mathbf{W}(\hat{\mathbf{U}}_s - \mathbf{A}(\theta)\mathbf{T})^H \right)\end{aligned}$$

As for ML, we replace with the estimator of  $\mathbf{T}$ , leading

$$\hat{\theta} = \arg \min_{\theta} \text{Tr} \left( \mathbf{\Pi}_{\mathbf{A}(\theta)}^\perp \hat{\mathbf{U}}_s \mathbf{W} \hat{\mathbf{U}}_s^H \right)$$

# Subspace fitting

Possible weights<sup>9</sup>

$$\mathbf{W} = \hat{\hat{\Lambda}}^2 \hat{\Lambda}_S^{-1}, \quad \hat{\hat{\Lambda}} = \hat{\Lambda}_S - \hat{\sigma}^2 \mathbf{I}_P, \quad \hat{\sigma}^2 = \frac{1}{N-P} \sum_{n=P+1}^N \hat{\lambda}_n$$

Alternative: use the fact that  $\mathbf{A}^H(\boldsymbol{\theta})\mathbf{U}_n = \mathbf{0}$  and solve

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \|\hat{\mathbf{U}}_e^H \mathbf{A}(\boldsymbol{\theta})\|_{\mathbf{W}}^2$$

---

<sup>9</sup>Rama Chellappa Abdelhak M. Zoubir Mats Viberg and Sergios Theodoridis. *Array and Statistical Signal Processing*. Academic Press, 2014.

## Some matrix operations

Kronecker product:  $\mathbf{A} \in \mathbb{C}^{m \times n}$ ,  $\mathbf{B} \in \mathbb{C}^{r \times q}$

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} A_{1,1}\mathbf{B} & A_{1,2}\mathbf{B} & \dots & A_{1,n}\mathbf{B} \\ A_{2,1}\mathbf{B} & A_{2,2}\mathbf{B} & \dots & A_{2,n}\mathbf{B} \\ \vdots & \vdots & \vdots & \vdots \\ A_{m,1}\mathbf{B} & A_{m,2}\mathbf{B} & \dots & A_{m,n}\mathbf{B} \end{bmatrix} \in \mathbb{C}^{mr \times nq}$$

Vec (vectorize) operator (linear)

$$\mathbf{A} = \begin{bmatrix} A_{1,1} & A_{1,2} & \dots & A_{1,n} \\ A_{2,1} & A_{2,2} & \dots & A_{1,n} \\ \vdots & \vdots & \vdots & \vdots \\ A_{m,1} & A_{m,2} & \dots & A_{m,n} \end{bmatrix} = [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \dots \quad \mathbf{a}_n] \Rightarrow \text{vec}(\mathbf{A}) = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_n \end{bmatrix}$$

Property

$$\text{vec}(\mathbf{AXB}) = (\mathbf{B}^T \otimes \mathbf{A})\text{vec}(\mathbf{X})$$

# Covariance fitting

Covariance matrix

$$\mathbf{R} = \mathbf{A}(\boldsymbol{\theta})\mathbf{R}_s\mathbf{A}^H(\boldsymbol{\theta}) + \sigma^2\mathbf{I}_N$$

leads

$$\begin{aligned} \mathbf{r} = \text{vec}(\mathbf{R}) &= \text{vec}(\mathbf{A}(\boldsymbol{\theta})\mathbf{R}_s\mathbf{A}^H(\boldsymbol{\theta})) + \text{vec}(\sigma^2\mathbf{I}_N) \\ &= (\mathbf{A}^*(\boldsymbol{\theta}) \otimes \mathbf{A}(\boldsymbol{\theta}))\text{vec}(\mathbf{R}_s) + \sigma^2\text{vec}(\mathbf{I}_N) \\ &= \boldsymbol{\Psi}(\boldsymbol{\theta})\mathbf{P} + \sigma^2\boldsymbol{\Sigma} \\ &= \begin{bmatrix} \boldsymbol{\Psi}(\boldsymbol{\theta}) & \boldsymbol{\Sigma} \end{bmatrix} \begin{bmatrix} \mathbf{P} \\ \sigma^2 \end{bmatrix} \\ &= \boldsymbol{\Phi}(\boldsymbol{\theta})\boldsymbol{\alpha} \end{aligned}$$

## Covariance fitting

Principle: minimize the error between  $\mathbf{R}$  and its estimate  $\hat{\mathbf{R}}$

$$\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\alpha}} = \arg \min_{\boldsymbol{\theta}, \boldsymbol{\alpha}} \|\hat{\mathbf{r}} - \Phi(\boldsymbol{\theta})\boldsymbol{\alpha}\|_{\mathbf{W}}^2 = (\hat{\mathbf{r}} - \Phi(\boldsymbol{\theta})\boldsymbol{\alpha})^H \mathbf{W} (\hat{\mathbf{r}} - \Phi(\boldsymbol{\theta})\boldsymbol{\alpha})$$

Solving for  $\boldsymbol{\alpha}$  then  $\boldsymbol{\theta}$  leads

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \|\mathbf{\Pi}^{\perp} \mathbf{W}^{\frac{1}{2}} \Phi(\boldsymbol{\theta}) \mathbf{W}^{\frac{1}{2}} \hat{\mathbf{r}}\|^2$$

Proposed weights

$$\hat{\mathbf{W}} = \left( \hat{\mathbf{R}}^T \otimes \hat{\mathbf{R}} \right)^{-1}$$

# Covariance fitting

## Avantages

- asymptotically equivalent to ML estimates
- no need for assumptions on the pdf (only  $\mathbf{R}$ )
- criterion usually simpler to minimize
- can be used with full-rank covariance  $\mathbf{R}_s$  (subspace methods require it to be rank deficient)

## Drawback

- requires an important number of samples

# Plan

① Introduction

② Model

③ Beamforming

④ Direction of arrival estimation

Problem formulation

Non parametric methods

Parametric methods

**Sparse estimation**

⑤ Conclusion



## Overcomplete representation

Signal model,  $\boldsymbol{\theta}$  and  $\mathbf{s}[k]$  unknown

$$\boldsymbol{\theta} = \{\theta_1, \theta_2, \dots, \theta_P\}, \quad \mathbf{y}[k] = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}[k] + \mathbf{e}[k]$$

Overcomplete representation, only  $\tilde{\mathbf{s}}[k]$  unknown

$$\tilde{\boldsymbol{\theta}} = \{\tilde{\theta}_1, \tilde{\theta}_2, \dots, \tilde{\theta}_{N_\theta}\}, \quad \mathbf{y}[k] = \tilde{\mathbf{A}}(\tilde{\boldsymbol{\theta}})\tilde{\mathbf{s}}[k] + \mathbf{e}[k], \quad N_\theta \gg P$$

where

$$\tilde{s}_j[k] = 0 \text{ if } \tilde{\theta}_j \notin \boldsymbol{\theta}, \quad j = 1, \dots, N_\theta$$

Only few non-zero elements in  $\tilde{\mathbf{s}}[k]$  (sparse vector)

# Regularization

Overcomplete representation

$$\mathbf{y}[k] = \tilde{\mathbf{A}}(\tilde{\boldsymbol{\theta}})\tilde{\mathbf{s}}[k] + \mathbf{e}[k] \text{ with } \tilde{\mathbf{A}}(\tilde{\boldsymbol{\theta}}) \in \mathbb{C}^{N \times N_{\theta}} \text{ and } N_{\theta} \gg N$$

Problem:  $\tilde{\mathbf{A}}(\tilde{\boldsymbol{\theta}})$  not full-rank (doesn't admit pseudoinverse)

Solution: regularization (add constraints to the vector to estimate)

$$\arg \min_{\tilde{\mathbf{s}}[k]} \underbrace{\|\mathbf{y}[k] - \tilde{\mathbf{A}}(\tilde{\boldsymbol{\theta}})\tilde{\mathbf{s}}[k]\|_2^2}_{\text{data-fidelity term}} + \lambda \underbrace{f(\tilde{\mathbf{s}}[k])}_{\text{regularization term}}$$

Hyperparameter  $\lambda$  controls the balance between fidelity and regularization

# Regularization

For instance: minimize signal energy (ridge regression)

$$\arg \min_{\tilde{\mathbf{s}}[k]} \|\mathbf{y}[k] - \tilde{\mathbf{A}}(\tilde{\boldsymbol{\theta}})\tilde{\mathbf{s}}[k]\|_2^2 + \lambda \|\tilde{\mathbf{s}}[k]\|_2^2$$

Solution

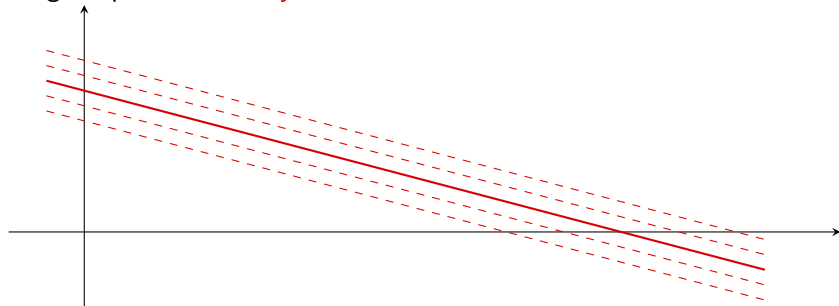
$$\hat{\tilde{\mathbf{s}}}_{\ell_2}[k] = \left( \tilde{\mathbf{A}}(\tilde{\boldsymbol{\theta}})^H \tilde{\mathbf{A}}(\tilde{\boldsymbol{\theta}}) + \lambda \mathbf{I}_{N_\theta} \right)^{-1} \tilde{\mathbf{A}}(\tilde{\boldsymbol{\theta}})^H \mathbf{y}[k]$$

We want sparsity, are there regularizations providing sparsity?

$$\ell_1 \text{ norm: } \|\mathbf{x}\|_1 = \sum_{i=1}^N |x_i| \text{ provides sparsity}$$

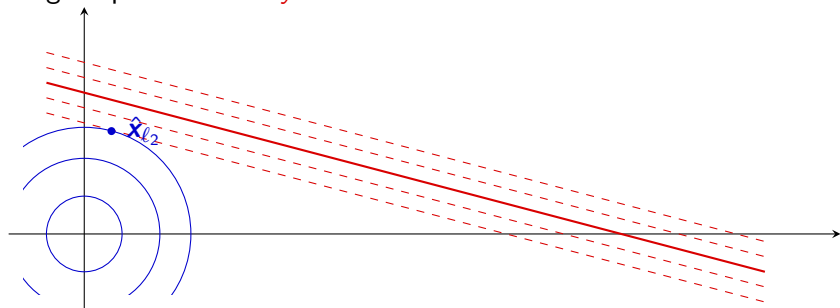
# Sparse regularization: illustration

Single equation in 2D  $y = \mathbf{h}^T \mathbf{x}$



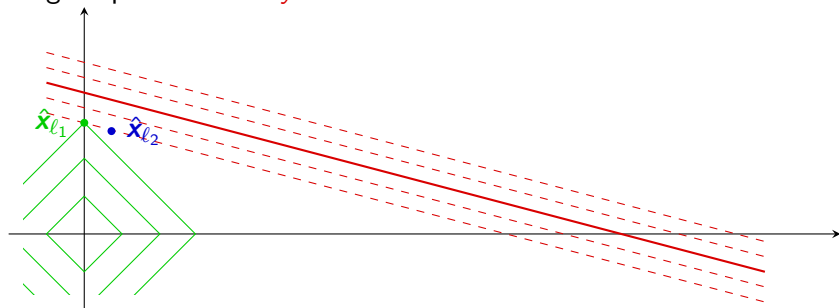
# Sparse regularization: illustration

Single equation in 2D  $y = \mathbf{h}^T \mathbf{x}$



# Sparse regularization: illustration

Single equation in 2D  $y = \mathbf{h}^T \mathbf{x}$



# Sparse DOA with a single snapshot

Solve

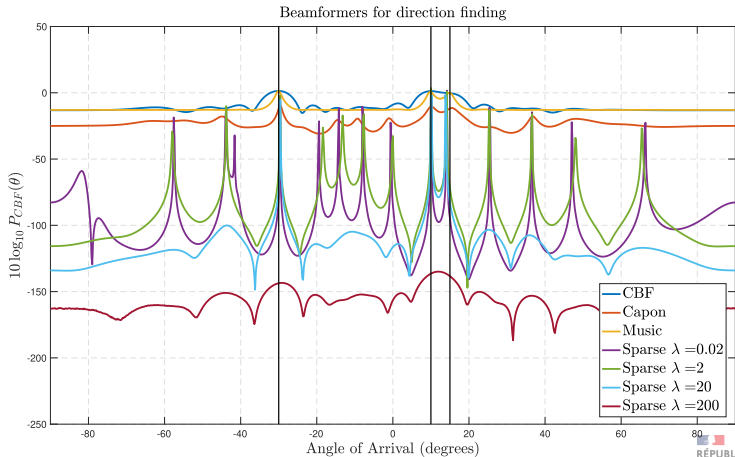
$$\arg \min_{\tilde{\mathbf{s}}[k]} \|\mathbf{y}[k] - \tilde{\mathbf{A}}(\tilde{\boldsymbol{\theta}})\tilde{\mathbf{s}}[k]\|_2^2 + \lambda \|\tilde{\mathbf{s}}[k]\|_1$$

We can plot a pseudospectrum for all values of  $\tilde{\theta}_j$  as  $|\tilde{s}_j[k]|^2$

## Influence of $\lambda$

$\lambda$  controls the level of sparsity of the solution

# Sparse DOA with a single snapshot





## Sparse DOA with all snapshots

Define

$$\mathbf{Y} = [\mathbf{y}[1] \quad \dots \quad \mathbf{y}[K]], \quad \tilde{\mathbf{S}} = [\tilde{\mathbf{s}}[1] \quad \dots \quad \tilde{\mathbf{s}}[K]] \quad \text{and} \quad \mathbf{E} = [\mathbf{e}[1] \quad \dots \quad \mathbf{e}[K]]$$

then

$$\mathbf{Y} = \tilde{\mathbf{A}}(\tilde{\boldsymbol{\theta}})\tilde{\mathbf{S}} + \mathbf{E}$$

We want to induce sparsity on the rows of matrix  $\tilde{\mathbf{S}}$ . Define

$$\tilde{\mathbf{s}}_j^{(\ell_2)} = \left\| [\tilde{\mathbf{s}}_j[1] \quad \dots \quad \tilde{\mathbf{s}}_j[K]]^T \right\|_2, \quad j = 1, \dots, N_\theta$$

and

$$\tilde{\mathbf{s}}^{(\ell_2)} = \begin{bmatrix} \tilde{\mathbf{s}}_1^{(\ell_2)} & \dots & \tilde{\mathbf{s}}_{N_\theta}^{(\ell_2)} \end{bmatrix}$$

## Reminder: Singular Value Decomposition (SVD)

Let  $\mathbf{M} \in \mathbb{C}^{N \times K}$  a rectangular matrix. Then, there exists  $\mathbf{U} \in \mathbb{C}^{N \times N}$ ,  $\mathbf{\Sigma} \in (\mathbb{R}^+)^{N \times K}$  and  $\mathbf{V} \in \mathbb{C}^{K \times K}$  such as

- $\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$
- $\mathbf{U}^H\mathbf{U} = \mathbf{I}_N$
- $\mathbf{V}^H\mathbf{V} = \mathbf{I}_K$
- $\mathbf{\Sigma}$  is a rectangular diagonal matrix

## Sparse DOA with all snapshots

Solve

$$\arg \min_{\tilde{\mathbf{S}}} \|\mathbf{Y} - \tilde{\mathbf{A}}(\tilde{\boldsymbol{\theta}})\tilde{\mathbf{S}}\|_2^2 + \lambda \|\tilde{\mathbf{s}}^{(\ell_2)}\|_1$$

SVD

$$\mathbf{Y} = \mathbf{U}\mathbf{L}\mathbf{V}^H, \quad \mathbf{U} \in \mathbb{C}^{N \times N}, \quad \mathbf{L} \in (\mathbb{R}^+)^{N \times K}, \quad \mathbf{V} \in \mathbb{C}^{K \times K}$$

can help to reduce the dimension of  $\mathbf{Y}$

$$\mathbf{Y}_s = \mathbf{U}\mathbf{L}\mathbf{D}_P = \mathbf{Y}\mathbf{V}\mathbf{D}_P, \quad \mathbf{D}_P = [\mathbf{I}_P \quad \mathbf{0}_{(K-P) \times K}]^T$$

Related to eigenvalue decomposition of  $\hat{\mathbf{R}}$

$$\hat{\mathbf{R}} = \frac{1}{K} \mathbf{Y}\mathbf{Y}^H = \frac{1}{K} \mathbf{U}\mathbf{L}\mathbf{V}^H \mathbf{V}\mathbf{L}\mathbf{U}^H = \frac{1}{K} \mathbf{U}\mathbf{L}^2 \mathbf{U}^H$$

# Sparse DOA with all snapshots

Defining

$$\tilde{\mathbf{S}}_s = \tilde{\mathbf{S}} \mathbf{V} \mathbf{D}_P, \quad \mathbf{E}_s = \mathbf{E} \mathbf{V} \mathbf{D}_P,$$

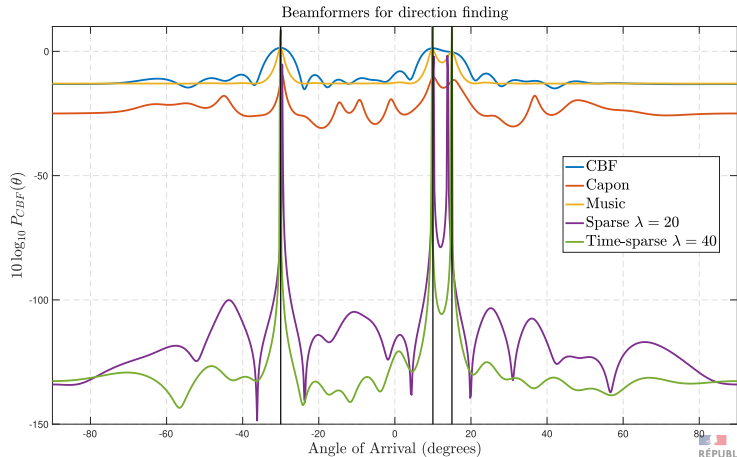
we have

$$\mathbf{Y}_s = \tilde{\mathbf{A}}(\tilde{\boldsymbol{\theta}}) \tilde{\mathbf{S}}_s + \mathbf{E}_s$$

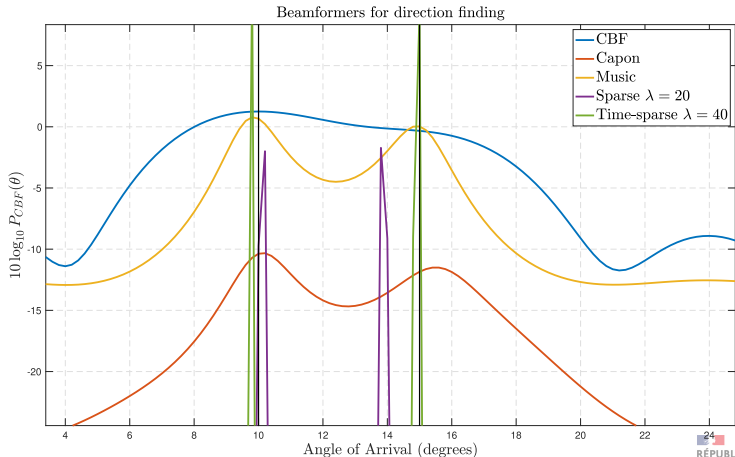
Solve

$$\arg \min_{\tilde{\mathbf{S}}_s} \|\mathbf{Y}_s - \tilde{\mathbf{A}}(\tilde{\boldsymbol{\theta}}) \tilde{\mathbf{S}}_s\|_2^2 + \lambda \|\tilde{\mathbf{S}}_s^{(\ell_2)}\|_1$$

# Sparse DOA with all snapshots



# Sparse DOA with all snapshots



# Conclusion on sparse DOA

## Advantages

- sharp peaks
- lower side lobes

## Drawbacks

- $\lambda$  selection
- depends on a grid
- numerical optimisation

# Plan

- ① Introduction
- ② Model
- ③ Beamforming
- ④ Direction of arrival estimation
- ⑤ Conclusion



# Conclusions

- Array processing, thanks to additional degrees of freedom, enables one to perform spatial filtering of signals
- Adaptive beamforming, possibly with reduced-rank transformations, enables one to achieve high SINR with a fast rate of convergence in adverse conditions (interference, noise)
- Robustness issues are of utmost importance in practical systems, and should be given a careful attention
- Non-parametric direction finding methods are simple and robust but may suffer from a lack of resolution
- Parametric methods offer high resolution, often at the price of degraded robustness

Please provide any useful comment on [SIDE!](#)