Introduction to Array Processing

Olivier Besson

Institut Supérieur de l'Aéronautique et de l'Espace
SUPAERO

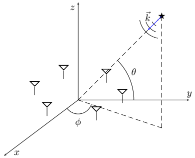
Contents

Context of multichannel processing

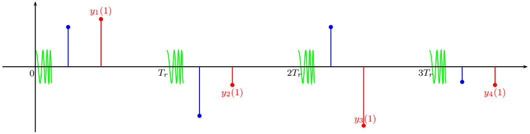
Multichannel processing involves **measurement vectors $\mathbf{y}(k)$**:

$$\mathbf{y}(k) = \begin{bmatrix} y_1(k) \\ y_2(k) \\ \vdots \\ \vdots \\ y_N(k) \end{bmatrix}$$

- N samples collected (possibly at the same time) on N different sensors, e.g., an EM wave received on N antennas:



- N samples taken on a single sensor over a time frame of NT_r. In radar N is the number of pulses sent by a radar each T_r seconds during a coherent processing interval:



Analysis/processing of multichannel measurements

- Analysis of such vectors should be understood as figuring out the relation between $y_n(.)$ and $y_m(.)$:
  - how correlated are the signals $y_n(.)$ and $y_m(.)$?
  - can $\mathbf{y}(.)$ be explained by a small number of variables (principal component analysis)?
- Processing these vectors includes for instance linear filtering, i.e.,

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

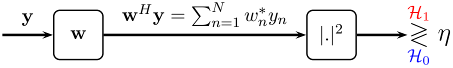
which can be used to improve reception of a signal of interest.

Example: linear filtering for detection purposes

- A classical multichannel problem is to detect/estimate a known signal $\mathbf{a}$ that would be present in $\mathbf{y}$ in addition to some noise $\mathbf{n}$, i.e., decide between the two hypotheses:

$$\begin{cases} \mathcal{H}_0 : & \mathbf{y} = \mathbf{n} \\ \mathcal{H}_1 : & \mathbf{y} = \alpha\mathbf{a} + \mathbf{n} \end{cases}$$

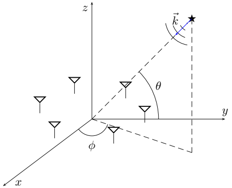
- A simple way is to design a suitable[1] linear filter $\mathbf{w}$ aimed at retrieving $\mathbf{a}$, and to test the energy $|\mathbf{w}^H\mathbf{y}|^2$ at the output:



[1] For instance, if $a_n = e^{i2\pi n f_s}$, one could choose $w_n = e^{i2\pi n f_s}$ so that $\mathbf{w}^H\mathbf{y}$ is the Fourier transform of $\mathbf{y}$ at frequency f_s.

Context of this array processing course

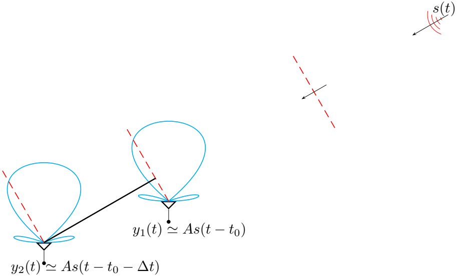
- In this course, we focus on signals received on an array of N antennas placed at different locations with a view to enhance reception of signals coming from preferred directions.



- $y_n(k)$ represents the signal received at antenna number n at time index number k.
- $n \in [1, N]$ is a spatial index as it is related to a particular position of an antenna in space and one is interested in spatial processing of $y_n(.)$, e.g. $\sum_{n=1}^{N} w_n^* y_n(k)$.

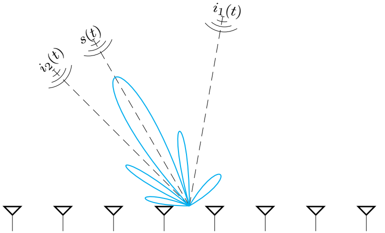
Principle of array processing: illustration with two antennas

- Consider two antennas receiving a signal $s(t)$ emitted by a source in the far-field

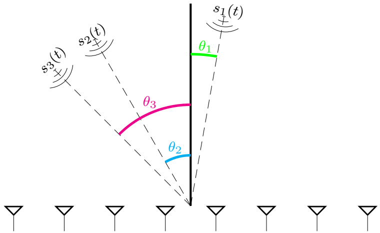


- The time delay Δt depends on the direction of arrival θ of $s(t)$ and on the relative (known) positions of the antennas:
  - if θ is known, one can obtain $s(t)$: spatial filtering (beamforming)
  - if one can estimate Δt from $y_1(t)$ and $y_2(t)$, then θ follows: source localization.

Beamforming and DoA estimation



Goal: retrieve signal from direction of interest and possibly eliminate interference.



Goal: estimate the directions of arrival of incoming signals.

Array of sensors

Potentialities

Array of sensors offer an additional dimension (space) which enables one, possibly in conjunction with temporal or frequency filtering, to perform spatial filtering of signals:

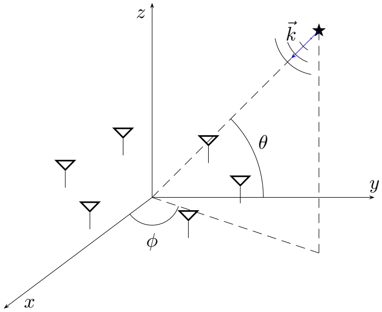
1. source separation
2. direction finding

Fields of application

1. radar, sonar (detection, target localization, anti-jamming)
2. communications (system capacity improvement, enhanced signals reception, spatial focusing of transmissions, interference mitigation)

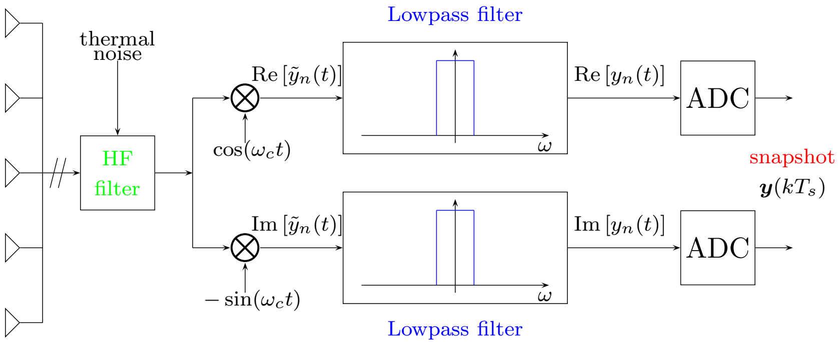
Contents

Arrays and waveforms

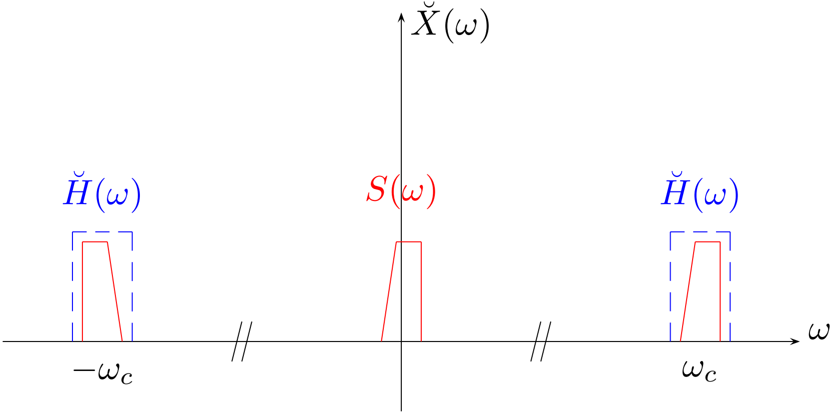


- The array performs spatial sampling of a wavefront impinging from direction(θ, ϕ).
- Assumptions: homogeneous propagation medium, source in the far-field of the array $\rightarrow$ plane wavefront.

Multichannel receiver



Source signal (frequency domain)



Signals and receiver

Source signal (narrowband)

$$\begin{aligned}\breve{x}(t) &= 2\mathrm{Re}\left\{s(t)e^{i\omega_c t}\right\} \\ &\triangleq \mathrm{Re}\left\{\alpha(t)e^{i\phi(t)}e^{i\omega_c t}\right\} \\ &= \alpha(t)\cos\left[\omega_c t + \phi(t)\right]\end{aligned}$$

$\alpha(t)$ and $\phi(t)$ stand for amplitude and phase of $s(t)$, and have slow time-variations relative to f_c.

Channel response

Receive channel number n has impulse response $\breve{h}_n(t)$.

Model of received signals

- Signal received on n-th antenna

$$\breve{y}_n(t) = \alpha \breve{h}_n(t) * \breve{x}(t - \tau_n) + \breve{n}_n(t)$$

where τ_n is the propagation delay to n-th sensor.

- In frequency domain :

$$\breve{Y}_n(\omega) = \alpha \breve{H}_n(\omega) \breve{X}(\omega) e^{-i\omega\tau_n} + \breve{N}_n(\omega)$$

- After demodulation ($\omega \to \omega + \omega_c$) and lowpass filtering:

$$\begin{aligned} Y_n(\omega) &= \alpha \breve{H}_n(\omega + \omega_c) S(\omega) e^{-i(\omega+\omega_c)\tau_n} + \breve{N}_n(\omega + \omega_c) \\ &\simeq \alpha \breve{H}_n(\omega_c) S(\omega) e^{-i\omega_c\tau_n} + \breve{N}_n(\omega + \omega_c) \end{aligned}$$

Model of received signals

- Taking the inverse Fourier transform $\mathcal{F}^{-1}(Y_n(\omega))$ yields

$$y_n(t) \simeq \alpha \breve{H}_n(\omega_c) s(t) e^{-i\omega_c \tau_n} + n_n(t)$$

- The **snapshot** writes

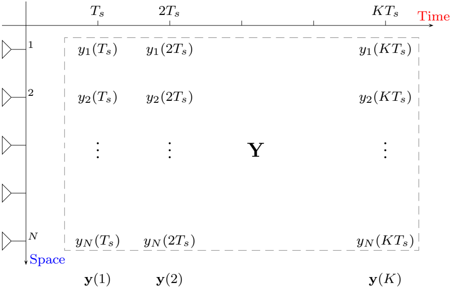
$$\mathbf{y}(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_N(t) \end{bmatrix} = \alpha \begin{bmatrix} \breve{H}_1(\omega_c) e^{-i\omega_c \tau_1} \\ \breve{H}_2(\omega_c) e^{-i\omega_c \tau_2} \\ \vdots \\ \breve{H}_N(\omega_c) e^{-i\omega_c \tau_N} \end{bmatrix} s(t) + \begin{bmatrix} n_1(t \\ n_2(t) \\ \vdots \\ n_N(t) \end{bmatrix}$$

- Assuming all $\breve{H}_n(\omega_c)$ are identical and absorbing α and $\breve{H}_n(\omega_c)$ in $s(t)$, we simply write

$$\boxed{\mathbf{y}(t) = \mathbf{a}(\theta) s(t) + \mathbf{n}(t)}$$

Model of received signals

- The snapshot is then sampled (temporally) at rate T_s to obtain the $N|K$ data matrix $\mathbf{Y} = \begin{bmatrix}\mathbf{y}(1) & \mathbf{y}(2) & \ldots & \mathbf{y}(K)\end{bmatrix}$:



- The k-th snapshot is given by

$$\mathbf{y}(k) = \mathbf{a}(\theta)s(k) + \mathbf{n}(k)$$

where $\mathbf{a}(\theta)$ is the vector of phase shifts, referred to as the **steering vector** since τ_n depends only on the directions(s) of arrival of the source.

Model of received signals

Snapshot at time index k

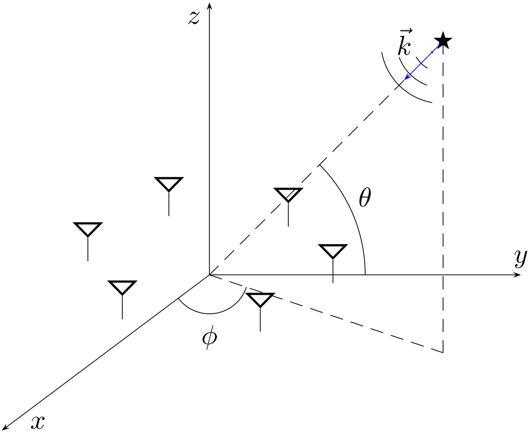
The snapshot received in the presence of P sources is given by

$$\mathbf{y}(k) = \sum_{p=1}^{P} \mathbf{a}(\theta_p) s_p(k) + \mathbf{n}(k)$$

$$= \begin{bmatrix} \mathbf{a}(\theta_1) & \dots & \mathbf{a}(\theta_P) \end{bmatrix} \begin{bmatrix} s_1(k) \\ \vdots \\ s_P(k) \end{bmatrix} + \mathbf{n}(k)$$

$$= \overset{N|P}{\mathbf{A}(\boldsymbol{\theta})} \underset{P|1}{\mathbf{s}(k)} + \mathbf{n}(k)$$

Steering vector

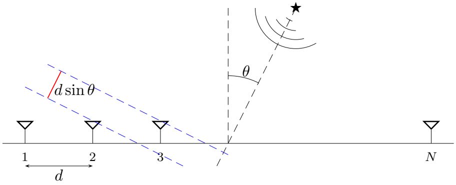


$$\tau_n = \frac{1}{c}\left[x_n \cos\theta \cos\phi + y_n \cos\theta \sin\phi + z_n \sin\theta\right]$$

$$a_n(\theta, \phi) = e^{i\frac{2\pi}{\lambda}\left[x_n \cos\theta \cos\phi + y_n \cos\theta \sin\phi + z_n \sin\theta\right]}$$

Uniform linear array (ULA)

Steering vector



$$\mathbf{a}(\theta) = \begin{bmatrix} 1 & e^{i2\pi f_s} & \cdots & e^{i2\pi (N-1) f_s} \end{bmatrix}^T; \quad f_s = f_c \frac{d \sin\theta}{c} = \frac{d}{\lambda} \sin\theta$$

Spatial sampling requirement

For the phase shift $\Delta\phi = 2\pi \frac{d}{\lambda} \sin\theta$ to be within $[-\pi, \pi]$ for every $\theta \in [-\pi/2, \pi/2]$ one needs to have

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

Covariance matrix

Covariance matrix

- The **covariance matrix** is defined as

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

- The (n, ℓ) entry $\mathbf{R}(n, \ell) = \mathbb{E}\left\{y_n(k)y_\ell^*(k)\right\}$ measures the correlation between signals received at sensors n and ℓ, *at the same time index* k.

Structure of the covariance matrix

Signals covariance matrix

The covariance matrix of the signal component is

$$\mathbb{E}\left\{\mathbf{A}(\boldsymbol{\theta})\mathbf{s}(k)\mathbf{s}^H(k)\mathbf{A}^H(\boldsymbol{\theta})\right\} = \mathbf{A}(\boldsymbol{\theta})\mathbf{R}_s\mathbf{A}^H(\boldsymbol{\theta}) \qquad (\mathbf{R}_s = \mathbb{E}\left\{\mathbf{s}(k)\mathbf{s}^H(k)\right\})$$

$$= \sum_{p=1}^{P} P_p \mathbf{a}(\theta_p)\mathbf{a}^H(\theta_p) \quad (\textit{uncorrelated signals})$$

Provided that $\mathbf{R}_s$ is full-rank (non coherent signals), the signal covariance matrix has **rank P** and its **range space is spanned by the steering vectors $\mathbf{a}(\theta_p)$, $p = 1, \cdots, P$**.

Noise covariance matrix

Assuming spatially white noise (i.e., uncorrelated between channels) with same power on each channel, $\mathbb{E}\left\{\mathbf{n}(k)\mathbf{n}^H(k)\right\} = \sigma^2\mathbf{I}$.

Model limitations

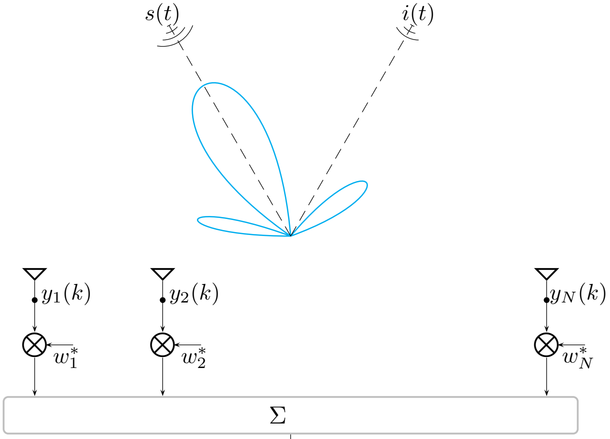
$\mathbf{y}(k) = \mathbf{a}(\theta)s(k) + \mathbf{n}(k)$ is an *idealized model* of the signals received on the array. It does not account for:

- a possibly non homogeneous propagation medium which results in coherence loss and wavefront distortions. This leads to amplitude and phase variations along the array, i.e. $y_n(k) = g_n(k)e^{i\phi_n(k)}a_n(\theta)s(k) + n_n(k)$.
- uncalibrated arrays, i.e., different amplitude and phase responses for each channel.
- wideband signals for which a time delay does not amount to a simple phase shift. In the frequency domain, one has $\mathbf{y}(f) = \mathbf{a}_f(\theta)s(f) + \mathbf{n}(f)$ with $\mathbf{a}_f(\theta) = \begin{bmatrix} 1 & e^{-i2\pi f\tau(\theta)} & \cdots & e^{-i2\pi f(N-1)\tau(\theta)} \end{bmatrix}^T$.
- possibly colored reception noise, i.e. $\mathbb{E}\left\{\mathbf{n}(k)\mathbf{n}^H(k)\right\} \neq \sigma^2\mathbf{I}$.

Contents

Spatial filtering

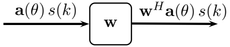
Principle: use a **linear combination of the sensors outputs** in order to point towards a looked direction.



$$y_F(k) = \sum_{n=1}^{N} w_n^* y_n(k) \simeq a s(k)$$

Array beampattern

- For any weight vector $\mathbf{w}$, the corresponding array beampattern (gain at direction θ) is defined as



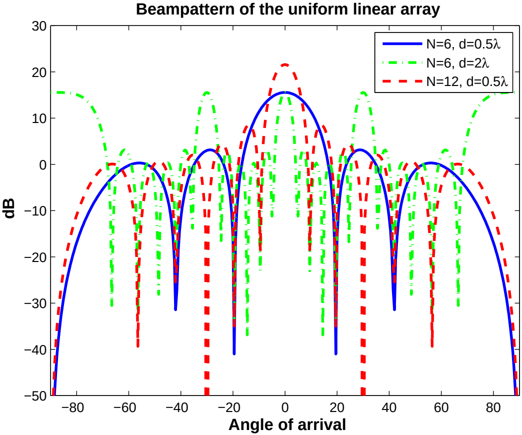
$$\Longrightarrow \quad G_{\mathbf{w}}(\theta) = |g_{\mathbf{w}}(\theta)|^2 = \left|\mathbf{w}^H \mathbf{a}(\theta)\right|^2$$

- For a uniform linear array, the natural beampattern, obtained as a **simple sum ($w_n = 1$) of the sensors outputs**, is given by

$$g(\theta) = \sum_{n=0}^{N-1} e^{i2\pi n \frac{d}{\lambda} \sin\theta} = e^{i\pi(N-1)\frac{d}{\lambda}\sin\theta} \, \frac{\sin\left[\pi N \frac{d}{\lambda} \sin\theta\right]}{\sin\left[\pi \frac{d}{\lambda} \sin\theta\right]}$$

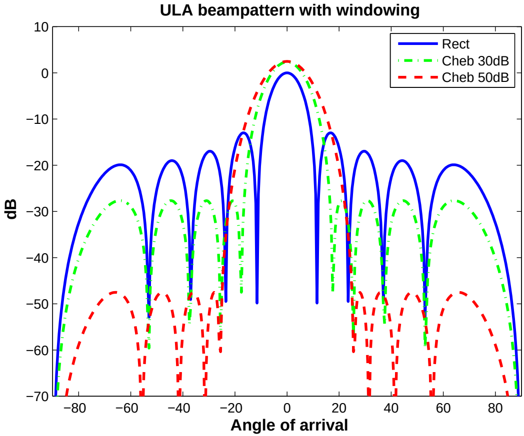
$$\boxed{G(\theta) = |g(\theta)|^2 = \left| \frac{\sin\left[\pi N \frac{d}{\lambda} \sin\theta\right]}{\sin\left[\pi \frac{d}{\lambda} \sin\theta\right]} \right|^2}$$

ULA beampattern



$$\theta_{3\text{dB}} \simeq \frac{0.9\lambda}{N\,d}$$

Windowing



ULA beampattern with windowing
Rect
Cheb 30dB
Cheb 50dB
10
0
−10
−20
−30
−40
−50
−60
−70
dB
−80
−60
−40
−20
0
20
40
60
80
Angle of arrival

Beamforming

Objective

Focus on a given direction in order to enhance reception of the signals impinging from this direction, and possibly mitigate interference located at other directions.

Principle

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

$$y_F(k) = \sum_{n=1}^{N} w_n^* y_n(k) = \begin{bmatrix} w_1^* & w_2^* & \cdots & w_N^* \end{bmatrix} \mathbf{y}(k) = \mathbf{w}^H \mathbf{y}(k).$$

Question

How to choose $\mathbf{w}$ such that, if $\mathbf{y}(k) = \mathbf{a}(\theta_s)s(k) + \cdots$ then at the output $y_F(k) \simeq \alpha s(k)$?

Conventional beamforming

Conventional beamforming: $\mathbf{w} \propto \mathbf{a}(\theta_s)$

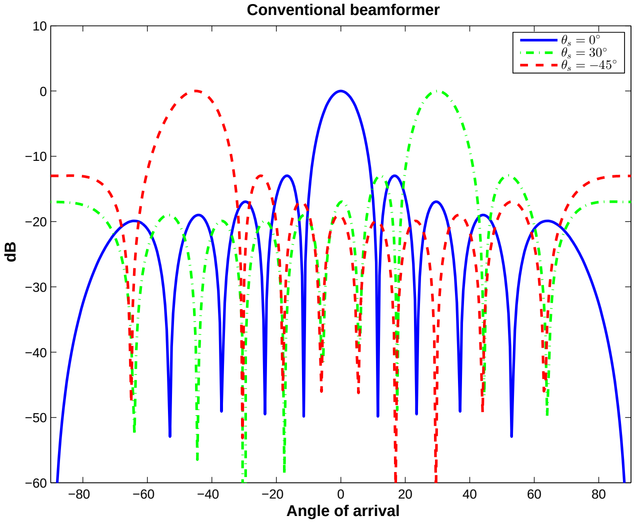
$$
\begin{aligned}
y_F(k) &= \mathbf{a}^H(\theta_s)\mathbf{a}(\theta_s)s(k) \quad [\mathbf{w} = \mathbf{a}(\theta_s), 1 \text{ source at } \theta_s] \\
&= \sum_{n=0}^{N-1} e^{-i2\pi\frac{d}{\lambda}n\sin\theta_s} \times e^{+i2\pi\frac{d}{\lambda}n\sin\theta_s}\ s(k) \\
&= \sum_{n=0}^{N-1} s(k) = N s(k)
\end{aligned}
$$

so that the gain towards θ_s is **maximal** and equal to N. The beamfomer $\mathbf{w}_{\text{CBF}} = \frac{\mathbf{a}(\theta_s)}{\mathbf{a}^H(\theta_s)\mathbf{a}(\theta_s)}$ is referred to as the conventional beamformer.

Principle

One compensates for the phase shift induced by propagation from direction θ_s and then sum **coherently**.

Array beampattern with conventional beamforming



SNR improvement

Before beamforming

$$\mathbf{y}(k) = \mathbf{a}_s s(k) + \mathbf{n}(k); \quad \mathrm{SNR}_{\mathsf{in}} \triangleq \frac{\mathbb{E}\left\{|s(k)|^2\right\}}{\mathbb{E}\left\{|n_n(k)|^2\right\}} = \frac{P}{\sigma^2}$$

After beamforming

$$y_F(k) = \mathbf{w}^H \mathbf{y}(k) = \mathbf{w}^H \mathbf{a}_s s(k) + \mathbf{w}^H \mathbf{n}(k)$$

$$\mathrm{SNR}_{\mathsf{out}} = \frac{|\mathbf{w}^H \mathbf{a}_s|^2}{\|\mathbf{w}\|^2} \mathrm{SNR}_{\mathsf{in}} \leq \|\mathbf{a}_s\|^2 \mathrm{SNR}_{\mathsf{in}} = N \times \mathrm{SNR}_{\mathsf{in}}$$

with equality if $\mathbf{w} \propto \mathbf{a}_s$.

White noise array gain

For any $\mathbf{w}$ such that $\mathbf{w}^H \mathbf{a}_s = 1$, the **white noise array gain** is $A_{\mathsf{WN}} = \mathrm{SNR}_{\mathsf{out}} / \mathrm{SNR}_{\mathsf{in}} = \|\mathbf{w}\|^{-2} \leq N$.

Conventional beamforming versus adaptive beamforming

Conventional beamforming

The conventional beamformer is optimal in white noise: it amounts to minimize $\mathbf{w}^H\mathbf{w}$ (the output power in white noise) under the constraint $\mathbf{w}^H\mathbf{a}(\theta_s) = 1$. Any other direction is deemed to be equivalent $\Rightarrow$ *it does not take into account other signals (interference) present in some directions*.

Adaptive beamforming

Adaptive beamforming takes into account these other signals. It consists in **minimizing the output power** $\mathbb{E}\left\{\left|\mathbf{w}^H\mathbf{y}(k)\right|^2\right\}$ while **maintaining a unit gain towards looked direction** $\Rightarrow$ tends to place nulls towards interfering signals.

Adaptive beamforming

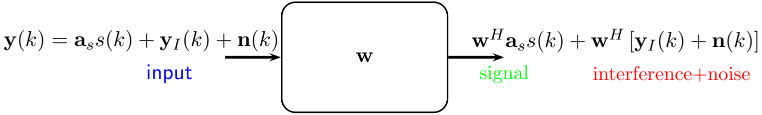
Beamforming-filtering in the presence of interference

- The received (input) signal in the presence of interference and noise is given by

$$\mathbf{y}(k) = \mathbf{a}_s s(k) + \mathbf{y}_I(k) + \mathbf{n}(k)$$

where $\mathbf{a}_s$ is the <u>actual</u> SoI steering vector.

- The output of the beamformer contains the same (albeit filtered) components:



Signal to interference plus noise ratio (SINR)

Definition of SINR

For a given beamformer $\mathbf{w}$, the usual figure of merit is the **signal to interference plus noise ratio (SINR)**, defined as

$$\mathrm{SINR}(\mathbf{w}) = \frac{\mathbb{E}\left\{\left|\mathbf{w}^H \mathbf{a}_s s(k)\right|^2\right\}}{\mathbb{E}\left\{\left|\mathbf{w}^H \left[\mathbf{y}_I(k) + \mathbf{n}(k)\right]\right|^2\right\}} = \frac{P_s \left|\mathbf{w}^H \mathbf{a}_s\right|^2}{\mathbf{w}^H \mathbf{C} \mathbf{w}}$$

where $\mathbf{C} = \mathbb{E}\left\{\left[\mathbf{y}_I(k) + \mathbf{n}(k)\right]\left[\mathbf{y}_I(k) + \mathbf{n}(k)\right]^H\right\}$ stands for the **interference plus noise covariance matrix**.

Optimal beamformer: SINR maximization

Optimal beamformer

Maximize SINR while ensuring a unit gain towards $\mathbf{a}_s$:

$$\boxed{\min_{\mathbf{w}} \mathbf{w}^H \mathbf{C} \mathbf{w} \text{ subject to } \mathbf{w}^H \mathbf{a}_s = 1} \qquad \text{(optimal)}$$

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

Remarks

- Principle is to minimize output power (when input $= \mathbf{y}_I + \mathbf{n}$) under the constraint that the **actual** steering vector $\mathbf{a}_s$ goes non distorted.
- Neither $\mathbf{a}_s$ nor $\mathbf{C}$ will be known in practice: the actual steering vector may be different from its expected value and $\mathbf{C}$ needs to be estimated from data (which contain $\mathbf{y}_I + \mathbf{n}$).

Minimum Variance Distortionless Response (MVDR)

Principle of MVDR beamformer

Minimize output power (when input $= \mathbf{y}_I + \mathbf{n}$) under the constraint that the **assumed** steering vector goes non distorted.

Minimization problem and solution

$$\boxed{\min_{\mathbf{w}} \mathbf{w}^H \mathbf{C} \mathbf{w} \text{ subject to } \mathbf{w}^H \mathbf{a}_0 = 1} \qquad \text{(MVDR)}$$

where $\mathbf{a}_0$ is the **assumed** steering vector of the signal of interest (SoI). The solution is given by

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

Minimum Power Distortionless Response (MPDR)

Principle of MPDR beamformer

Minimize output power (when input $= \mathbf{a}_s s + \mathbf{y}_I + \mathbf{n}$) under the constraint that the assumed steering vector goes non distorted:

$$\boxed{\min_{\mathbf{w}} \mathbf{w}^H \mathbf{R} \mathbf{w} \text{ subject to } \mathbf{w}^H \mathbf{a}_0 = 1} \qquad \text{(MPDR)}$$

where $\mathbf{R} (= \mathbf{C} + P_s \mathbf{a}_s \mathbf{a}_s^H)$ stands for the **signal plus interference plus noise** covariance matrix.

Solution

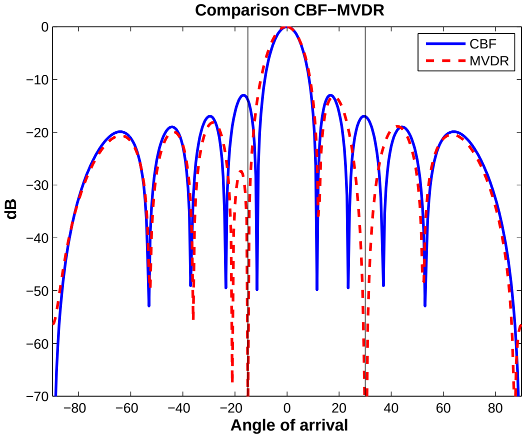
$$\mathbf{w}_{\text{MPDR}} = \frac{\mathbf{R}^{-1}\mathbf{a}_0}{\mathbf{a}_0^H \mathbf{R}^{-1} \mathbf{a}_0}$$

Summary of adaptive beamformers (known covariance matrices)

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

- $\mathbf{a}_s =$ actual steering vector and $\mathbf{a}_0 =$ assumed steering vector
- $\mathbf{C} = \mathrm{cov}(\mathbf{y}_I + \mathbf{n})$ and $\mathbf{R} = \mathrm{cov}(\mathbf{a}_s s + \mathbf{y}_I + \mathbf{n})$

CBF and optimal (MVDR) beampatterns



CBF vs MVDR: the case of a single interference

Derivation of SINR

In the case $\mathbf{C} = P_j \mathbf{a}_j \mathbf{a}_j^H + \sigma^2 \mathbf{I}$ with $\mathrm{INR} = \frac{P_j}{\sigma^2} \gg 1$, it can be shown that

$$\mathrm{SINR}_{\mathsf{CBF}} \simeq \frac{P_s}{\sigma^2} \times \frac{1}{g \times INR}; \quad \mathrm{SINR}_{\mathsf{opt}} \simeq \frac{P_s}{\sigma^2} \times N(1-g)$$

with $g = \cos^2(\mathbf{a}_s, \mathbf{a}_j) = |\mathbf{a}_s^H \mathbf{a}_j|^2 / (\mathbf{a}_s^H \mathbf{a}_s)(\mathbf{a}_j^H \mathbf{a}_j)$.

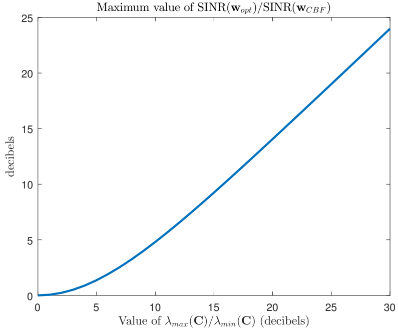
Remarks

- With CBF, the SINR decreases when P_j increases while it is independent of P_j with adaptive beamforming.
- The SINR decreases when $\mathbf{a}_j \to \mathbf{a}_s$ $(g \to 1)$.

CBF vs MVDR: when is the latter useful?

- From Kantorovich's inequality

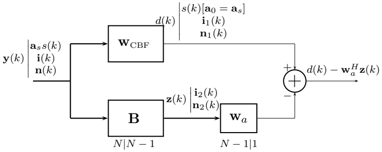
$$1 \le \frac{\mathrm{SINR}(\mathbf{w}_{\mathsf{opt}})}{\mathrm{SINR}(\mathbf{w}_{\mathsf{CBF}})} = \frac{(\mathbf{a}_s^H \mathbf{C}^{-1} \mathbf{a}_s)(\mathbf{a}_s^H \mathbf{C} \mathbf{a}_s)}{(\mathbf{a}_s^H \mathbf{a}_s)^2} \le \frac{(\lambda_{\mathsf{min}}(\mathbf{C}) + \lambda_{\mathsf{max}}(\mathbf{C}))^2}{4\lambda_{\mathsf{min}}(\mathbf{C})\lambda_{\mathsf{max}}(\mathbf{C})}$$

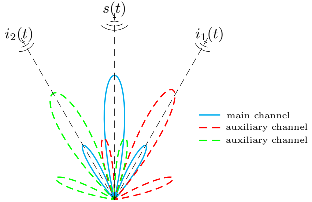


- Adaptive beamforming is adequate if $\lambda_{\mathsf{max}}(\mathbf{C})/\lambda_{\mathsf{min}}(\mathbf{C}) \gg 1$.

Generalized Sidelobe Canceler

Structure

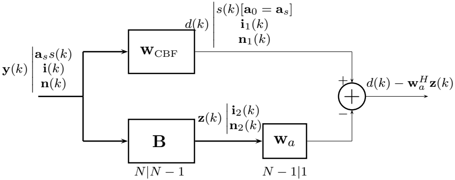


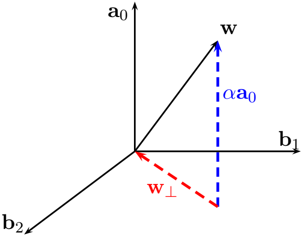


- The $(N-1)$ columns of $\mathbf{B}$ form a basis of the subspace orthogonal to $\mathbf{a}_0$, i.e., $\mathbf{B}^H \mathbf{a}_0 = \mathbf{0}$.
- The $(N-1)$ auxiliary channels $\mathbf{z}(k)$ are free of signal and enable one to infer the part of interference that went through the CBF.
- $\mathbf{w}_a$ enables one to estimate, from $\mathbf{z}(k)$, the part of interference $\mathbf{i}_1(k)$ contained in $d(k)$ since $\mathbf{i}_1(k)$ is **correlated** with $\mathbf{z}(k)$ through $\mathbf{i}_2(k)$.

Generalized Sidelobe Canceler

- The GSC structure decomposes $\mathbf{w}$ into a component along $\mathbf{a}_0$ and a component orthogonal to $\mathbf{a}_0$, i.e., $\mathbf{w} = \alpha\mathbf{a}_0 - \mathbf{w}_\perp$:





- The component along $\mathbf{a}_0$ *ensures that the constraint is fulfilled* since

$$\mathbf{w}^H\mathbf{a}_0 = \alpha^*\mathbf{a}_0^H\mathbf{a}_0 - \mathbf{w}_\perp^H\mathbf{a}_0 = \alpha^*\mathbf{a}_0^H\mathbf{a}_0 + 0 \Rightarrow \alpha = \left(\mathbf{a}_0^H\mathbf{a}_0\right)^{-1}$$

- The orthogonal component $\mathbf{w}_\perp = \mathbf{B}\mathbf{w}_a$ is chosen to minimize output power, in an **unconstrained** way.

Generalized Sidelobe Canceler

- Minimization of the output power can be achieved by solving one of the **two following equivalent problems**:

$$\min_{\mathbf{w}^H \mathbf{a}_0 = 1} \mathbf{w}^H \mathbf{C} \mathbf{w} \quad \Bigg| \quad \min_{\mathbf{w}_a} \left(\mathbf{w}_{\text{CBF}} - \mathbf{B}\mathbf{w}_a\right)^H \mathbf{C} \left(\mathbf{w}_{\text{CBF}} - \mathbf{B}\mathbf{w}_a\right)$$

direct form, constrained | GSC form, unconstrained

- The MVDR beamformer in its GSC form is given by

$$\mathbf{w}_{\text{GSC}} = \mathbf{w}_{\text{CBF}} - \mathbf{B}\mathbf{w}_a^*$$

where $\mathbf{w}_a^*$ solves the above minimization problem.

Generalized Sidelobe Canceler

- The power at the output of the beamformer is given by

$$\begin{aligned}\mathbb{E}\left\{\left|d(k)-\mathbf{w}_a^H\mathbf{z}(k)\right|^2\right\} &= \mathbb{E}\left\{|d(k)|^2\right\}-\mathbf{w}_a^H\mathbf{r}_{d\mathbf{z}}-\mathbf{r}_{d\mathbf{z}}^H\mathbf{w}_a+\mathbf{w}_a^H\mathbf{R}_z\mathbf{w}_a \\ &= \left[\mathbf{w}_a-\mathbf{R}_z^{-1}\mathbf{r}_{d\mathbf{z}}\right]^H\mathbf{R}_z\left[\mathbf{w}_a-\mathbf{R}_z^{-1}\mathbf{r}_{d\mathbf{z}}\right] \\ &\quad+\mathbb{E}\left\{|d(k)|^2\right\}-\mathbf{r}_{d\mathbf{z}}^H\mathbf{R}_z^{-1}\mathbf{r}_{d\mathbf{z}}\end{aligned}$$

  with $\mathbf{r}_{d\mathbf{z}}=\mathbb{E}\left\{\mathbf{z}(k)d^*(k)\right\}$ and $\mathbf{R}_z=\mathbb{E}\left\{\mathbf{z}(k)\mathbf{z}(k)^H\right\}$.
- The weight vector which minimizes output power is thus

$$\mathbf{w}_a^*=\mathbf{R}_z^{-1}\mathbf{r}_{d\mathbf{z}}$$

Generalized Sidelobe Canceler

- The GSC form of the weight vector is given by

$$\begin{aligned}\mathbf{w}_{\mathsf{GSC}} &= \mathbf{w}_{\mathsf{CBF}} - \mathbf{B}\mathbf{R}_z^{-1}\mathbf{r}_{d\mathbf{z}} \\ &= \mathbf{w}_{\mathsf{CBF}} - \mathbf{B}\left(\mathbf{B}^H\mathbf{R}_y\mathbf{B}\right)^{-1}\mathbf{B}^H\mathbf{R}_y\mathbf{w}_{\mathsf{CBF}} \qquad \text{(GSC)}\end{aligned}$$

  where $\mathbf{R}_y = \mathbf{R}$ in a MPDR scenario and $\mathbf{R}_y = \mathbf{C}$ in a MVDR scenario.
- Since they solve the same problem $\mathbf{w}_{\mathsf{GSC}} = \left(\mathbf{a}_0^H\mathbf{R}_y^{-1}\mathbf{a}_0\right)^{-1}\mathbf{R}_y^{-1}\mathbf{a}_0$.
- The SINR is inversely proportional to the output power when $\mathbf{R}_y = \mathbf{C}$, i.e.,

$$\mathrm{SINR}_{\mathsf{GSC}} = P_s\left[\mathbf{w}_{\mathsf{CBF}}^H\mathbf{C}\mathbf{w}_{\mathsf{CBF}} - \mathbf{r}_{d\mathbf{z}}^H\mathbf{R}_z^{-1}\mathbf{r}_{d\mathbf{z}}\right]^{-1}$$

MVDR versus MPDR

The optimal, MVDR and MPDR beamformers are equivalent if and only if

$$\min_{\mathbf{w}} \mathbf{w}^H \left(\mathbf{C} + P_s \mathbf{a}_s \mathbf{a}_s^H\right) \mathbf{w} \text{ subject to } \mathbf{w}^H \mathbf{a}_0 = 1 \qquad \text{(MPDR)}$$

$$\equiv \min_{\mathbf{w}} \mathbf{w}^H \mathbf{C} \mathbf{w} \text{ subject to } \mathbf{w}^H \mathbf{a}_0 = 1 \qquad \text{(MVDR)}$$

$$\equiv \min_{\mathbf{w}} \mathbf{w}^H \mathbf{C} \mathbf{w} \text{ subject to } \mathbf{w}^H \mathbf{a}_s = 1 \qquad \text{(opt)}$$

which is true only when the 2 following conditions are satisfied:

1. the assumed steering vector $\mathbf{a}_0$ coincides with the actual steering vector $\mathbf{a}_s$: in practice, uncalibrated arrays or a pointing error lead to $\mathbf{a}_0 \neq \mathbf{a}_s$;
2. the covariance matrix $\mathbf{R}$ is **known**: in practice, one needs to estimate it which results in estimation errors $\hat{\mathbf{R}} - \mathbf{R}$.

$\Longrightarrow$ It ensues that degradation compared to $\mathrm{SINR}_{\text{opt}}$ is unavoidable in practice, and it can be quite different between MPDR and MVDR.

Influence of a steering vector error (MVDR)

- We assume that the SoI steering vector is $\mathbf{a}_0$ while it is actually $\mathbf{a}_s$.
- The SINR obtained with $\mathbf{w}_{\text{MVDR}} = \left(\mathbf{a}_0^H \mathbf{C}^{-1} \mathbf{a}_0\right)^{-1} \mathbf{C}^{-1} \mathbf{a}_0$ becomes

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

Influence of a steering vector error (MPDR)

- The MPDR beamformer can be written as

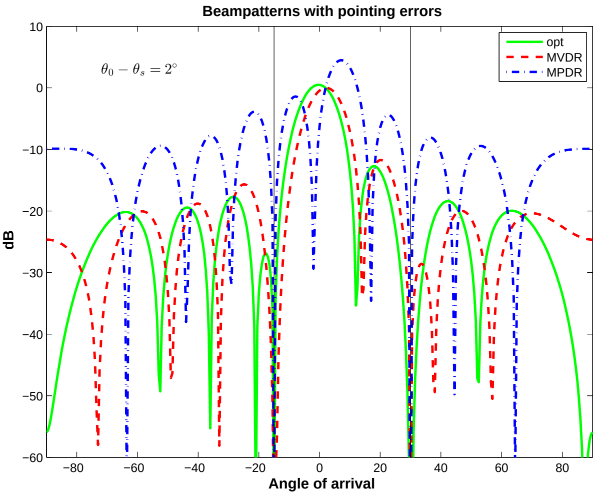
$$\mathbf{w}_{\mathsf{MPDR}} = \frac{\mathbf{R}^{-1}\mathbf{a}_0}{\mathbf{a}_0^H \mathbf{R}^{-1}\mathbf{a}_0};\ \mathbf{R} = P_s \mathbf{a}_s \mathbf{a}_s^H + \mathbf{C}$$

- **Its SINR is decreased compared to that of the MVDR**, viz

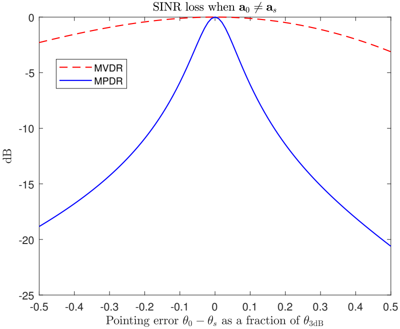
$$\begin{aligned}\mathrm{SINR}_{\mathsf{MPDR}} &= \frac{\mathrm{SINR}_{\mathsf{MVDR}}}{1 + \left(2\mathrm{SINR}_{\mathsf{opt}} + \mathrm{SINR}_{\mathsf{opt}}^2\right)\sin^2\left(\mathbf{a}_s, \mathbf{a}_0; \mathbf{C}^{-1}\right)} \\ &\leq \mathrm{SINR}_{\mathsf{MVDR}}.\end{aligned}$$

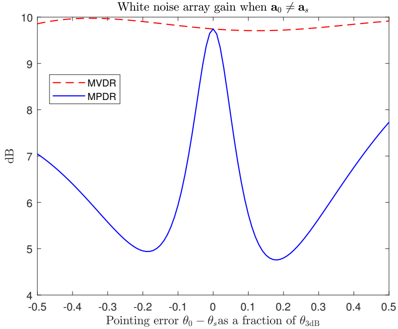
- The degradation is more important as $\mathrm{SINR}_{\mathsf{opt}}$ (hence P_s) increases.

Influence of a steering vector error on beampatterns



Influence of a steering vector error on SINR and WNAG





Case of an uncalibrated array

- Let us consider an uncalibrated array with actual steering vector

$$\tilde{\mathbf{a}}_n(\theta) = (1 + g_n) e^{i\phi_n} \mathbf{a}_n(\theta)$$

where $\{g_n\}$ and $\{\phi_n\}$ are independent random gains and phases.

- For any beamformer $\mathbf{w}$, the average value of the resulting beampattern $\tilde{G}_{\mathbf{w}}(\theta) = |\mathbf{w}^H \tilde{\mathbf{a}}(\theta)|^2$ is related to the nominal beampattern $G_{\mathbf{w}}(\theta) = |\mathbf{w}^H \mathbf{a}(\theta)|^2$ through

$$\mathbb{E}\left\{\tilde{G}_{\mathbf{w}}(\theta)\right\} = |\gamma|^2 G_{\mathbf{w}}(\theta) + \left[1 + \sigma_g^2 - |\gamma|^2\right] \|\mathbf{w}\|^2$$

where $\sigma_g^2 = \mathbb{E}\left\{|g_n|^2\right\}$ and $\gamma = \mathbb{E}\left\{e^{i\phi_n}\right\}$.

- The term proportional to $\|\mathbf{w}\|^2$ leads to sidelobe level increase $\Rightarrow$ better to have high white noise array gain (small $\|\mathbf{w}\|^2$).

Influence of a finite number of snapshots

- In practice, K snapshots are available:

$$\mathbf{y}(k) = \mathbf{a}_s s(k) + \overbrace{\mathbf{y}_I(k) + \mathbf{n}(k)}^{\mathbf{y}_{i+n}(k)}; \qquad k = 1, \ldots, K$$

- The covariance matrices are thus estimated and subsequently one can compute the corresponding beamformers as

$$\hat{\mathbf{R}} = \frac{1}{K} \sum_{k=1}^{K} \mathbf{y}(k)\mathbf{y}^H(k) \qquad \longrightarrow \mathbf{w}_{\mathsf{MPDR}}^{\mathsf{smi}} = \frac{\hat{\mathbf{R}}^{-1}\mathbf{a}_0}{\mathbf{a}_0^H \hat{\mathbf{R}}^{-1}\mathbf{a}_0}$$

$$\hat{\mathbf{C}} = \frac{1}{K} \sum_{k=1}^{K} \mathbf{y}_{i+n}(k)\mathbf{y}_{i+n}^H(k) \qquad \longrightarrow \mathbf{w}_{\mathsf{MVDR}}^{\mathsf{smi}} = \frac{\hat{\mathbf{C}}^{-1}\mathbf{a}_0}{\mathbf{a}_0^H \hat{\mathbf{C}}^{-1}\mathbf{a}_0}$$

where $^{\mathsf{smi}}$ stands for “sample matrix inversion”.

Influence of a finite number of snapshots

- The sample beamformers $\mathbf{w}_{\text{M-DR}}^{\text{smi}}$ will differ from their ensemble counterparts $\mathbf{w}_{\text{M-DR}}$ since $\hat{\mathbf{R}} = \mathbf{R} + \boldsymbol{\Delta}\mathbf{R}$ and $\hat{\mathbf{C}} = \mathbf{C} + \boldsymbol{\Delta}\mathbf{C}$.
- The weight vectors $\mathbf{w}_{\text{M-DR}}^{\text{smi}}$ are random and so are their corresponding signal to noise ratios

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

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

- Important issue is **speed of convergence**, i.e., how large should K be for $\mathbb{E}\left\{\mathrm{SINR}\left(\mathbf{w}_{\text{MPDR}}^{\text{smi}}\right)\right\}$ or $\mathbb{E}\left\{\mathrm{SINR}\left(\mathbf{w}_{\text{MVDR}}^{\text{smi}}\right)\right\}$ to be “close” to $\mathrm{SINR}_{\text{opt}}$?

SINR loss with finite number of snapshots (MVDR)

- When $\mathbf{a}_0 = \mathbf{a}_s$, the **SINR loss** of the MVDR beamformer can be represented as

$$\rho_{\text{MVDR}} = \frac{\text{SINR}\left(\mathbf{w}_{\text{MVDR}}^{\text{smi}}\right)}{\text{SINR}\left(\mathbf{w}_{\text{opt}}\right)} \overset{d}{=} \left[1 + \frac{\chi^2_{2(N-1)}(0)}{\chi^2_{2(K-N+2)}(0)}\right]^{-1}$$

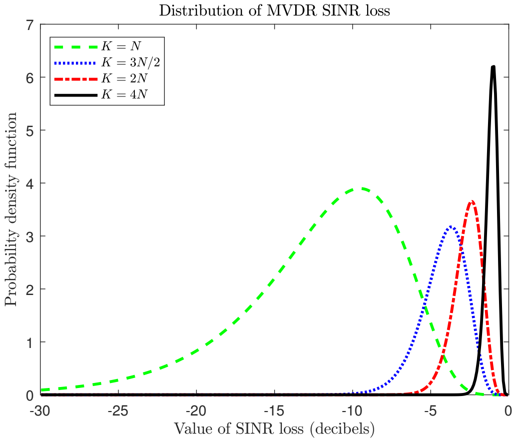
and thus follows a **beta distribution**, i.e.,

$$p_{\text{MVDR}}(\rho) = \frac{\Gamma(K+1)}{\Gamma(K-N+2)\Gamma(N-1)}\, \rho^{K-N+1}(1-\rho)^{N-2}$$

which is independent of $\mathbf{C}$.

- The expected value is $\mathbb{E}\left\{\rho_{\text{MVDR}}\right\} = (K-N+2)/(K+1)$, so that $\text{SINR}\left(\mathbf{w}_{\text{MVDR}}^{\text{smi}}\right)$ is (on average) within 3dB of the optimal SINR for $K_{\text{MVDR}} = 2N-3$.

SINR loss with finite number of snapshots (MVDR)



SINR loss with finite number of snapshots (MPDR)

- As for the MPDR scenario one has

$$\rho_{\text{MPDR}} = \frac{\text{SINR}\left(\mathbf{w}_{\text{MPDR}}^{\text{smi}}\right)}{\text{SINR}\left(\mathbf{w}_{\text{opt}}\right)} \stackrel{d}{=} \left[1 + (1 + \text{SINR}_{\text{opt}}) \frac{\chi^2_{2(N-1)}(0)}{\chi^2_{2(K-N+2)}(0)}\right]^{-1}$$

- The distribution of ρ_{MPDR} is related to that of ρ_{MVDR} as follows:

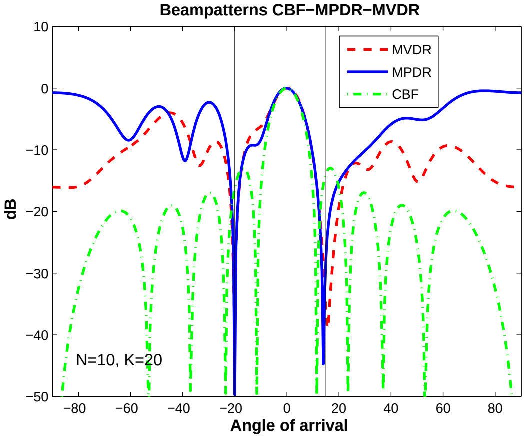
$$p_{\text{MPDR}}(\rho) = p_{\text{MVDR}}(\rho) \times \frac{\left(1 + \text{SINR}_{\text{opt}}\right)^{K-N+2}}{\left(1 + \rho\text{SINR}_{\text{opt}}\right)^{K+1}}$$

- The average number of snapshots to achieve the optimal SINR within 3dB is about

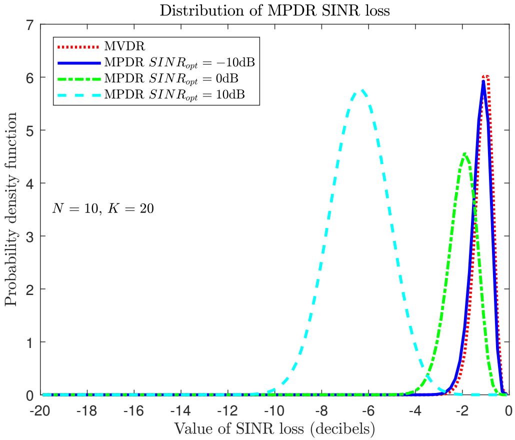
$$K_{\text{MPDR}} \simeq (N-1)\left[1 + \text{SINR}_{\text{opt}}\right]$$

where $\text{SINR}_{\text{opt}} \simeq N\left(\frac{P_s}{\sigma^2}\right)$. In general, $K_{\text{MPDR}} \gg K_{\text{MVDR}}$.

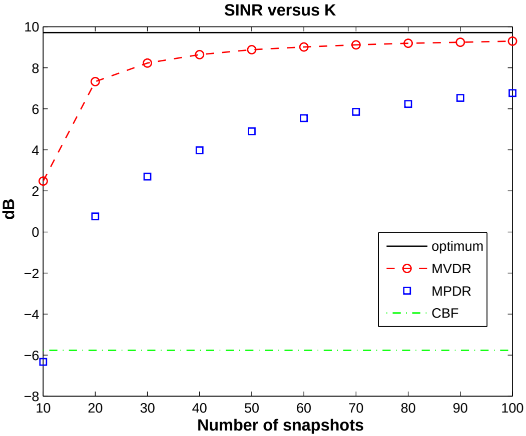
Beampatterns with finite number of snapshots



Distribution of SINR loss



SINR versus number of snapshots



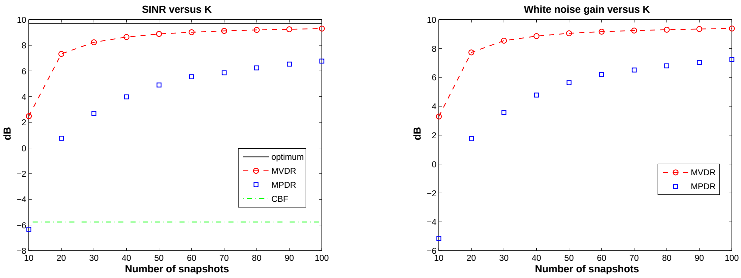
How to make MPDR more robust?

Observations

- Estimation of covariance matrices leads to a significant SINR loss (especially for the MPDR beamformer) due to
  - the interference being less eliminated
  - a sidelobe level increase which results in a lower white noise gain.
- In case of uncalibrated arrays, steering vector errors are all the more emphasized that the white noise gain is low (or $\|\mathbf{w}\|^2$ large).

How to make MPDR more robust?

White noise array gain and SINR



Observation: similarity between the two curves.

A possible remedy

Enforce a minimal white noise array gain or equivalently restrain $\|\mathbf{w}\|^2$ in order to make the MPDR beamformer more robust.

Diagonal loading

Principle

One tries to solve

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

Finding the beamformer

The Lagrangian is given by (with $\lambda \in \mathbb{C}$ and $\mu \in \mathbb{R}^+$)

$$L(\mathbf{w}, \lambda, \mu) = \mathbf{w}^H \hat{\mathbf{R}} \mathbf{w} + \lambda \left(\mathbf{w}^H \mathbf{a}_0 - 1 \right) + \lambda^* \left(\mathbf{a}_0^H \mathbf{w} - 1 \right) + \mu \left(\|\mathbf{w}\|^2 - A_{\mathsf{WN}}^{-1} \right)$$

$$= \left[\mathbf{w} + \lambda \left(\hat{\mathbf{R}} + \mu \mathbf{I} \right)^{-1} \mathbf{a}_0 \right]^H \left(\hat{\mathbf{R}} + \mu \mathbf{I} \right) \left[\mathbf{w} + \lambda \left(\hat{\mathbf{R}} + \mu \mathbf{I} \right)^{-1} \mathbf{a}_0 \right]$$

$$- \lambda - \lambda^* - \mu A_{\mathsf{WN}}^{-1} - |\lambda|^2 \mathbf{a}_0^H \left(\hat{\mathbf{R}} + \mu \mathbf{I} \right)^{-1} \mathbf{a}_0.$$

Diagonal loading

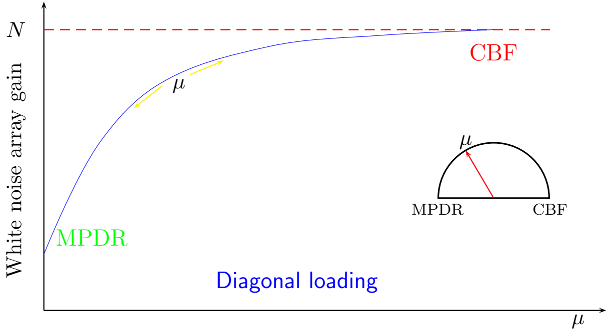
Solution

The solution thus takes the form $\mathbf{w}(\lambda, \mu) = -\lambda \left(\hat{\mathbf{R}} + \mu \mathbf{I}\right)^{-1} \mathbf{a}_0$. Since we must have $\mathbf{w}(\lambda, \mu)^H \mathbf{a}_0 = 1$, it follows that

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

and μ is selected such that $\|\mathbf{w}_{\text{MPDR-DL}}(\mu)\|^{-2} = A_{\text{WN}}$.

Diagonal loading : adaptivity versus robustness



$$\lim_{\mu \to 0} \mathbf{w}_{\text{MPDR-DL}}(\mu) = \mathbf{w}_{\text{MPDR}}^{\text{smi}} \mid \lim_{\mu \to \infty} \mathbf{w}_{\text{MPDR-DL}}(\mu) = \mathbf{w}_{\text{CBF}}$$

Choice of loading level

Many different possibilities have been proposed to set the loading level:

- set A_{WN} (slightly below N) and compute μ from $\|\mathbf{w}_{\text{MPDR-DL}}\|^{-2} = A_{\text{WN}}$.
- set μ directly, generally a few decibels above white noise level (see discussion next slide about beampatterns and eigenvalues).
- set μ using the theory of ridge regression, which enables one to compute μ from data.
- use that diagonal loading is the solution to the following problem

$$\max_{P,\mathbf{a}} \hat{\mathbf{R}} - P\mathbf{a}\mathbf{a}^H \text{ for } \|\mathbf{a} - \mathbf{a}_0\|^2 \leq \varepsilon^2$$

  and compute μ from ε.
- set A_{WN} and compute directly the diagonally loaded beamformer in GSC form without necessarily computing μ.
- ...

An interpretation of diagonal loading and the choice of μ

- The array beampattern with the **true** covariance matrix is given by

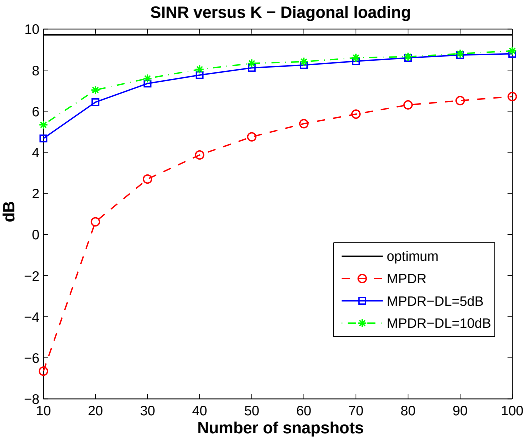
$$g(\theta) = \frac{\alpha}{\sigma^2} \left\{ \mathbf{a}_0^H \mathbf{a}(\theta) - \sum_{n=1}^{J} \frac{\lambda_n}{\lambda_n + \sigma^2} \left[\mathbf{a}_0^H \mathbf{u}_n \right] \mathbf{u}_n^H \mathbf{a}(\theta) \right\}$$

- The array beampattern with an **estimated** covariance matrix becomes

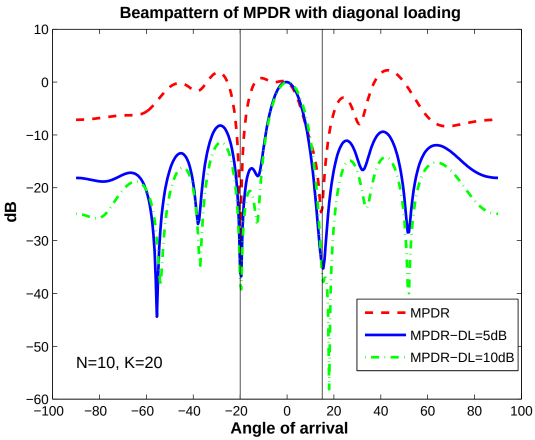
$$g^{\text{smi}}(\theta) = \frac{\alpha}{\hat{\lambda}_{\text{min}}} \left\{ \mathbf{a}_0^H \mathbf{a}(\theta) - \sum_{n=1}^{N} \frac{\hat{\lambda}_n}{\hat{\lambda}_n + \hat{\lambda}_{\text{min}}} \left[\mathbf{a}_0^H \hat{\mathbf{u}}_n \right] \hat{\mathbf{u}}_n^H \mathbf{a}(\theta) \right\}$$

- Degradation is due to $\hat{\lambda}_{J+1} \neq \hat{\lambda}_{J+2} \neq \cdots \hat{\lambda}_N = \hat{\lambda}_{\text{min}}$.
- Replacing $\hat{\mathbf{R}}$ by $\hat{\mathbf{R}} + \mu \mathbf{I}$ enables one to equalize the eigenvalues, provided that $\mu \gg \sigma^2$ and $\mu < \lambda_J$.

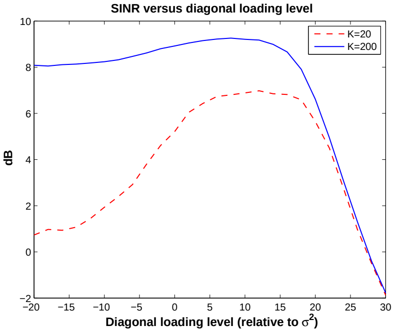
Diagonal loading: SINR versus number of snapshots

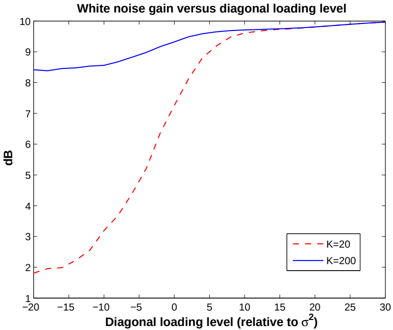


Diagonal loading: beampatterns



Influence of the loading level on SINR and WNAG





Linearly constrained beamforming

- To mitigate pointing errors, one can resort to multiple constraints, i.e. solve the problem

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

whose solution is $\mathbf{w} = \mathbf{C}^{-1}\mathbf{Z}\left(\mathbf{Z}^H\mathbf{C}^{-1}\mathbf{Z}\right)^{-1}\mathbf{d}$.

- One can use a unit gain constraint around the presumed DOA or a smoothness constraint:

$$\mathbf{Z} = \begin{bmatrix} \mathbf{a}(\theta_0) & \mathbf{a}(\theta_0+\delta_1) & \cdots & \mathbf{a}(\theta_0+\delta_L) \end{bmatrix} \quad \mathbf{d} = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^T$$

$$\mathbf{Z} = \begin{bmatrix} \mathbf{a}(\theta_0) & \left.\frac{\partial \mathbf{a}(\theta)}{\partial \theta}\right|_{\theta_0} & \cdots & \left.\frac{\partial^L \mathbf{a}(\theta)}{\partial \theta^L}\right|_{\theta_0} \end{bmatrix} \quad \mathbf{d} = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T$$

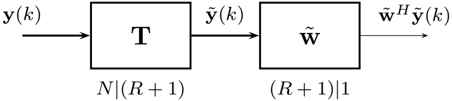
Partially adaptive beamforming

Principle

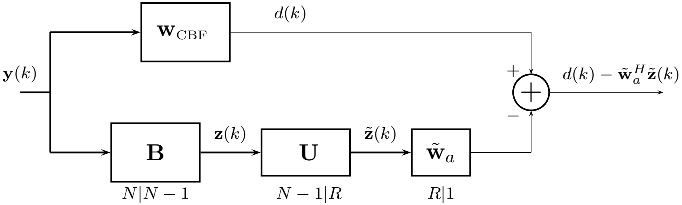
Perform beamforming in a lower dimensional subspace

Structures

- Direct form:



- GSC form ($\mathbf{T} = \begin{bmatrix}\mathbf{w}_{\mathsf{CBF}} & \mathbf{BU}\end{bmatrix}$):



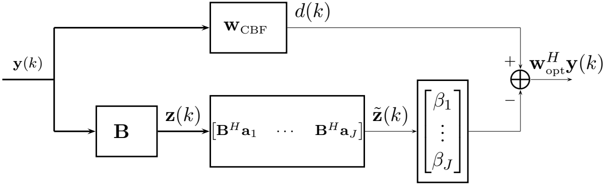
Motivation for partially adaptive beamforming

The optimal beamformer when $\mathbf{C} = \sum_{j=1}^{J} P_j \mathbf{a}_j \mathbf{a}_j^H + \sigma^2 \mathbf{I}$

- With this low-rank + scaled identity matrix form, one has

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

- The optimal beamformer amounts to subtract from the CBF a linear combination of **J** beams steered towards interference:



- **The optimal beamformer is a partially adaptive beamformer.**

Optimality of the partially adaptive beamformer ($\mathbf{a}_0 = \mathbf{a}_s$)

Question: can we possibly have $\mathbf{w}_{\text{PA}} = \mathbf{w}_{\text{opt}}$?

Direct form

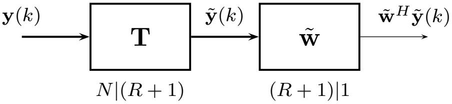
- Answer is **yes**: $\mathbf{w}_{\text{PA-DF}} = \mathbf{w}_{\text{opt}} \Leftrightarrow \mathbf{C}^{-1}\mathbf{a}_s \in \mathcal{R}\{\mathbf{T}\}$
- At first glance, meaningless condition : if $\mathbf{C}^{-1}\mathbf{a}_s$ were known, we would get $\mathbf{w}_{\text{opt}}$ and hence no need for $\mathbf{T}$.
- But if $\mathbf{C} = \sum_{j=1}^{J} P_j \mathbf{a}_j \mathbf{a}_j^H + \sigma^2 \mathbf{I}$ then $\mathbf{C}^{-1}\mathbf{a}_s = \eta_s \mathbf{a}_s + \sum_{j=1}^{J} \eta_j \mathbf{a}_j$.
- $\Rightarrow$ if $\begin{bmatrix} \mathbf{a}_s & \mathbf{a}_1 & \ldots & \mathbf{a}_J \end{bmatrix} \in \mathcal{R}\{\mathbf{T}\}$ 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 $\begin{bmatrix} \mathbf{B}^H\mathbf{a}_1 & \ldots & \mathbf{B}^H\mathbf{a}_J \end{bmatrix} \in \mathcal{R}\{\mathbf{U}\}$ then $\mathbf{w}_{\text{PA-GSC}} = \mathbf{w}_{\text{opt}}$.

Expression of the partially adaptive beamformer

Direct form



- Minimization of the output power

$$\min_{\tilde{\mathbf{w}}} \tilde{\mathbf{w}}^H \mathbf{R}_{\tilde{y}} \tilde{\mathbf{w}} \text{ subject to } \tilde{\mathbf{w}}^H \tilde{\mathbf{a}}_0 = 1 \qquad \text{(PA-DF)}$$

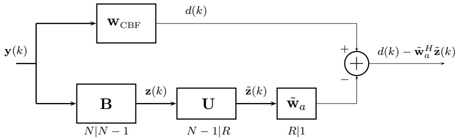
where $\mathbf{R}_{\tilde{y}} = \mathbf{T}^H \mathbf{R}_y \mathbf{T}$ and $\tilde{\mathbf{a}}_0 = \mathbf{T}^H \mathbf{a}_0$.

- The solution is given by

$$\tilde{\mathbf{w}} = \alpha \mathbf{R}_{\tilde{y}}^{-1} \tilde{\mathbf{a}}_0 \Rightarrow \mathbf{w}_{\text{PA-DF}} = \alpha \mathbf{T} \mathbf{R}_{\tilde{y}}^{-1} \tilde{\mathbf{a}}_0$$

Expression of the partially adaptive beamformer

GSC form



- Minimization of the output power $[\tilde{\mathbf{z}}(k) = \mathbf{U}^H \mathbf{z}(k)]$

$$\min_{\tilde{\mathbf{w}}_a} \mathbb{E}\left\{ \left| d(k) - \tilde{\mathbf{w}}_a^H \tilde{\mathbf{z}}(k) \right|^2 \right\} \qquad \text{(PA-GSC)}$$

- The solution is given by

$$\tilde{\mathbf{w}}_a = \mathbf{R}_{\tilde{z}}^{-1} \mathbf{r}_{d\tilde{\mathbf{z}}} = \left(\mathbf{U}^H \mathbf{R}_z \mathbf{U}\right)^{-1} \mathbf{U}^H \mathbf{r}_{d\mathbf{z}}$$

$$\mathbf{w}_{\text{PA-GSC}} = \mathbf{w}_{\text{CBF}} - \mathbf{B}\mathbf{U}\mathbf{R}_{\tilde{z}}^{-1} \mathbf{r}_{d\tilde{\mathbf{z}}}$$

Analysis of the partially adaptive MVDR

SINR loss for fixed $\mathbf{T}$ ($\mathbf{R}_y = \mathbf{C}$, $\mathbf{a}_0 = \mathbf{a}_s$)

The SINR loss of the partially adaptive beamformer $\mathbf{w} = \mathbf{T}\tilde{\mathbf{w}} = \mathbf{T}\tilde{\mathbf{C}}^{-1}\tilde{\mathbf{a}}_0$ with fixed $\mathbf{T}$ is distributed according to

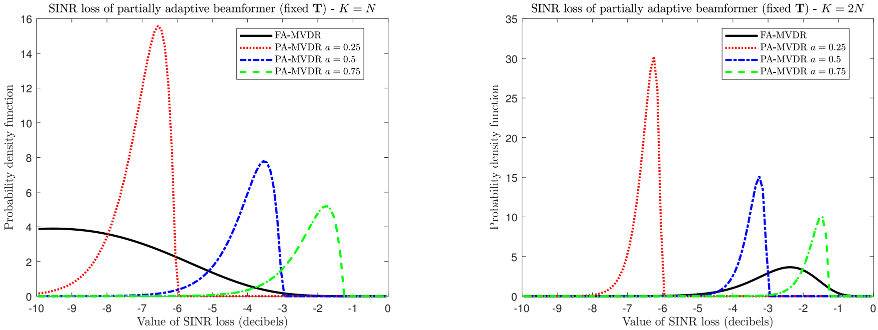
$$\rho_{\text{PA-MVDR}} \overset{d}{=} a \left[1 + \frac{\chi^2_{2R}(0)}{\chi^2_{2(K-R+1)}(0)}\right]^{-1}$$

where

$$a = \frac{\tilde{\mathbf{a}}_0^H \tilde{\mathbf{C}}^{-1} \tilde{\mathbf{a}}_0}{\mathbf{a}_0^H \mathbf{C}^{-1} \mathbf{a}_0} = \frac{\mathbf{a}_0^H \mathbf{T}(\mathbf{T}^H \mathbf{C} \mathbf{T})^{-1} \mathbf{T}^H \mathbf{a}_0}{\mathbf{a}_0^H \mathbf{C}^{-1} \mathbf{a}_0}$$

$$= \frac{\text{energy of } \mathbf{C}^{-1/2}\mathbf{a}_0 \text{ in } \mathcal{R}\left\{\mathbf{C}^{1/2}\mathbf{T}\right\}}{\text{energy of } \mathbf{C}^{-1/2}\mathbf{a}_0} \leq 1$$

Analysis of the partially adaptive MVDR



$\Rightarrow$ partially adaptive beamforming is potentially very effective in low sample support, provided that $\mathbf{T}$ is well chosen.

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

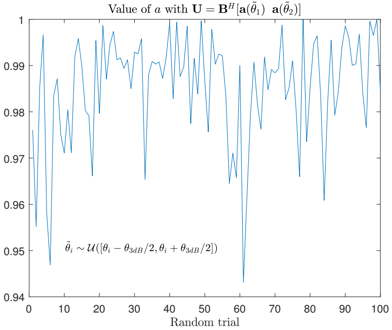
Fixed transformations

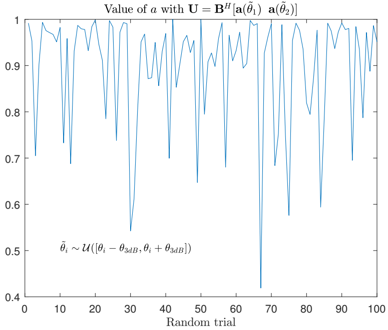
- For instance using pre-steered beams, i.e.

$$\mathbf{T} = \begin{bmatrix} \mathbf{a}(\theta_s) & \mathbf{a}(\tilde{\theta}_1) & \mathbf{a}(\tilde{\theta}_2) & \cdots & \mathbf{a}(\tilde{\theta}_R) \end{bmatrix}$$
$$\mathbf{U} = \mathbf{B}^H \begin{bmatrix} \mathbf{a}(\tilde{\theta}_1) & \mathbf{a}(\tilde{\theta}_2) & \cdots & \mathbf{a}(\tilde{\theta}_R) \end{bmatrix}$$

- In this case, the columns of $\mathbf{U}$ can be viewed as beamformers aimed at intercepting the interference.
- Require some prior knowledge about the interference DOA in order for them to pass through the beams.

Value of a with pre-steered beams





Case of 2 interferences located at θ_1, θ_2. Value of a when $\mathbf{U} = \mathbf{B}^H \left[\mathbf{a}(\tilde{\theta}_1) \quad \mathbf{a}(\tilde{\theta}_2)\right]$ and $\tilde{\theta}_i$ drawn randomly around θ_i.

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

Adaptive transformations ($\mathbf{T}$ or $\mathbf{U}$ depend on the snapshots)

- The optimal beamformer writes $\mathbf{w}_{\text{opt}} = \mathbf{w}_{\text{CBF}} - \mathbf{B}\sum_{j=1}^{J}\beta_j(\mathbf{B}^H\mathbf{a}_j)$ $\Rightarrow$ all we need is a basis for $\mathcal{R}\left\{\begin{bmatrix}\mathbf{B}^H\mathbf{a}_1 & \ldots & \mathbf{B}^H\mathbf{a}_J\end{bmatrix}\right\}$.
- The eigenvalue decomposition of $\mathbf{R}_z = \mathbf{B}^H\mathbf{C}\mathbf{B}$ is given by

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

- Property: $\mathcal{R}\left\{\begin{bmatrix}\mathbf{B}^H\mathbf{a}_1 & \ldots & \mathbf{B}^H\mathbf{a}_J\end{bmatrix}\right\} = \mathcal{R}\left\{\begin{bmatrix}\mathbf{q}_1 & \ldots & \mathbf{q}_J\end{bmatrix}\right\}$.

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

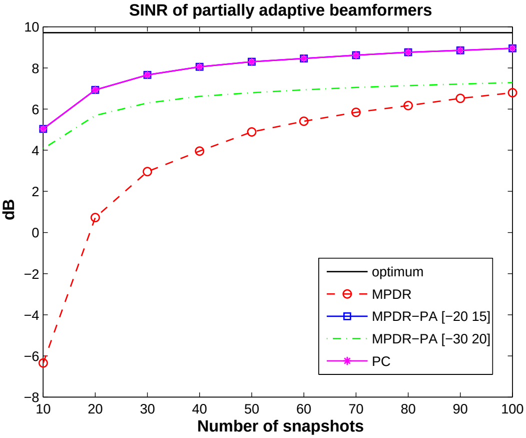
Adaptive transformations ($\mathbf{T}$ or $\mathbf{U}$ depend on the snapshots)

- A logical choice for $\mathbf{U}$ is thus to select the R principal eigenvectors of $\mathbf{R}_z$ (Principal Component), i.e. $\mathbf{U} = \begin{bmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_R \end{bmatrix}$.
- With this choice $\mathbf{R}_{\tilde{z}} = \mathbf{U}^H \mathbf{R}_z \mathbf{U} = \mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_R)$ and

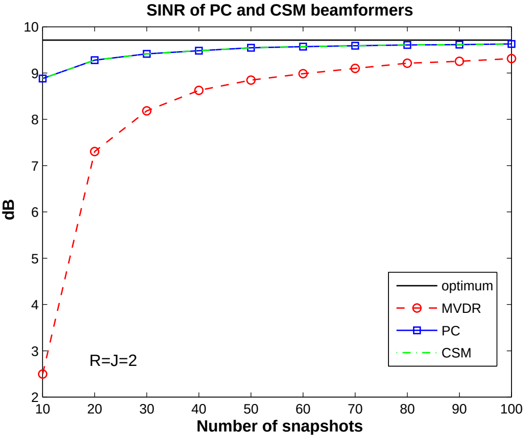
$$\mathbf{w}_{\text{pc-gsc}} = \mathbf{w}_{\text{CBF}} - \mathbf{B}\mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}^H \mathbf{r}_{d\mathbf{z}}$$

- Another interesting choice is to select the R eigenvectors which contribute most to increasing the SINR (Cross Spectral Metric).

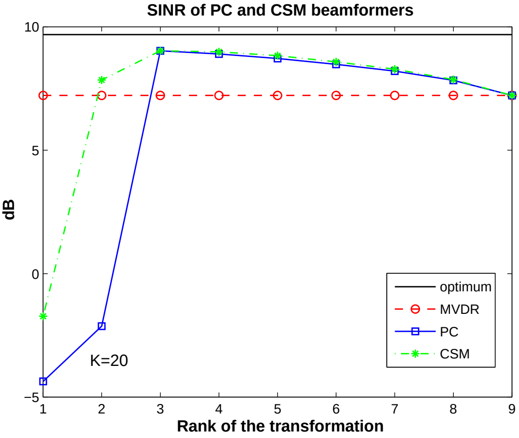
Partially adaptive beamforming: SINR versus K



Partially adaptive beamforming: SINR versus K



Partially adaptive beamforming: SINR versus R



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

Random transformations

- The idea[a] is to use L matrices $\mathbf{U}_\ell$ drawn from a uniform distribution on the manifold of semi-unitary $(N-1) \times R$ matrices, i.e.

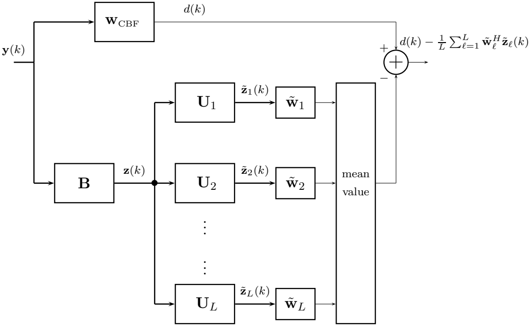
$$\mathbf{U}_\ell = \mathbf{X}_\ell \left(\mathbf{X}_\ell^H \mathbf{X}_\ell\right)^{-H/2}; \quad \mathbf{X}_\ell \overset{d}{=} \mathbb{C}\mathcal{N}\left(\mathbf{0}, \mathbf{I}_{N-1}, \mathbf{I}_R\right)$$

and to average the corresponding weight vectors $\tilde{\mathbf{w}}_\ell$, yielding

$$\begin{aligned} \mathbf{w} &= \mathbf{w}_{\text{CBF}} - \mathbf{B}\left[\frac{1}{L}\sum_{\ell=1}^{L} \mathbf{U}_\ell \left(\mathbf{U}_\ell^H \mathbf{R}_z \mathbf{U}_\ell\right)^{-1} \mathbf{U}_\ell^H \mathbf{r}_{d\mathbf{z}}\right] \\ &= \mathbf{w}_{\text{CBF}} - \mathbf{B}\left[\frac{1}{L}\sum_{\ell=1}^{L} \mathbf{X}_\ell \left(\mathbf{X}_\ell^H \mathbf{R}_z \mathbf{X}_\ell\right)^{-1} \mathbf{X}_\ell^H \mathbf{r}_{d\mathbf{z}}\right] \end{aligned}$$

[a] T. Marzetta, G. Tucci, S. Simon, "A random matrix-theoretic approach to handling singular covariance matrices", *IEEE Transactions Information Theory*, September 2011

Marzetta's method based on random $\mathbf{U}$



The matrices $\mathbf{U}_\ell$ are drawn from a uniform distribution on the manifold of semi-unitary matrices, or from a Gaussian distribution $\mathbb{C}\mathcal{N}\left(\mathbf{0}, \mathbf{I}_{N-1}, \mathbf{I}_R\right)$.

Beamforming: synthesis

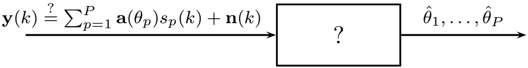
- **Conventional beamforming** $\mathbf{w}_{\text{CBF}} = \left(\mathbf{a}_0^H \mathbf{a}_0\right)^{-1} \mathbf{a}_0$. *Optimal in white noise*, $\theta_{3\text{dB}} = 0.9 \left(N \frac{d}{\lambda}\right)^{-1}$, sidelobes at -13dB.
- **Adaptive beamforming** $\mathbf{w}_{\text{opt}} \propto \mathbf{C}^{-1}\mathbf{a}_s$, $\mathbf{w}_{\text{MVDR}} \propto \mathbf{C}^{-1}\mathbf{a}_0$, $\mathbf{w}_{\text{MPDR}} \propto \mathbf{R}^{-1}\mathbf{a}_0$
    - all equivalent if $\mathbf{R}$, $\mathbf{C}$ known and $\mathbf{a}_s = \mathbf{a}_0$
    - $\text{SINR}_{\text{opt}} \gtrsim \text{SINR}_{\text{MVDR}} \gg \text{SINR}_{\text{MPDR}}$ when $\mathbf{a}_s \neq \mathbf{a}_0$
    - $\text{SINR}_{\text{MVDR-SMI}} \gg \text{SINR}_{\text{MPDR-SMI}}$: convergence for about $2N$ snapshots for MVDR, $N \times \text{SINR}_{\text{opt}}$ for MPDR
- **Diagonal loading:** *helps to mitigate both finite-sample errors and steering vector errors*. Especially useful in MPDR context with low power signal of interest.
- **Partially adaptive beamforming:** enables one to achieve *faster convergence* by operating in low-dimensional subspace. Especially effective with strong, low-rank interference.

Contents

The direction of arrival estimation problem

Problem formulation

Given a collection of K snapshots which can possibly be modeled as $\mathbf{y}(k) = \sum_{p=1}^{P} \mathbf{a}(\theta_p) s_p(k) + \mathbf{n}(k)$, estimate the directions of arrival (DoA) $\theta_1, \ldots, \theta_P$:

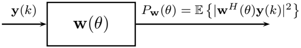


Approaches

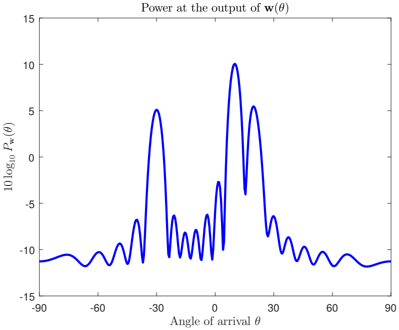
- Non parametric approaches which do not necessarily rely on a model for $\mathbf{y}(k)$: similar to Fourier-based methods in time domain;
- Parametric approaches where a model is assumed and its properties (algebraic structure, distribution) are exploited.

Beamforming for direction finding purposes

- The idea is **to form a beam $\mathbf{w}(\theta)$ for each angle θ and to evaluate the power $\mathbb{E}\left\{|y_F(k)|^2\right\} = \mathbb{E}\left\{\left|\mathbf{w}^H(\theta)\mathbf{y}(k)\right|^2\right\} = \mathbf{w}^H(\theta)\mathbf{R}\mathbf{w}(\theta)$ at the output of the beamformer versus θ**:



- Large peaks should provide the directions of arrival:



Conventional beamforming for direction finding purposes

Conventional beamformer

The conventional beamformer $\mathbf{w}_{\text{CBF}}(\theta) = \mathbf{a}(\theta)/N$ can be used, which yields the output power

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

In practice

With K snapshots available, $\mathbf{R}$ is estimated as

$$\hat{\mathbf{R}} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{y}(k)\mathbf{y}^H(k)$$

and subsequently the output power as

$$P_{\text{CBF}}(\theta) = N^{-2}\mathbf{a}^H(\theta)\hat{\mathbf{R}}\mathbf{a}(\theta)$$

CBF and Fourier analysis

- The estimated power at the output of the CBF writes

$$
\begin{aligned}
P_{\text{CBF}}(\theta) &= \frac{1}{N^2}\mathbf{a}^H(\theta)\hat{\mathbf{R}}\mathbf{a}(\theta) \\
&= \frac{1}{KN^2}\sum_{k=1}^{K}\left|\mathbf{a}^H(\theta)\mathbf{y}(k)\right|^2 \\
&= \frac{1}{KN^2}\sum_{k=1}^{K}\left|\sum_{n=1}^{N}\mathbf{y}_n(k)e^{-i2\pi(n-1)f}\right|^2
\end{aligned}
$$

where $f = \frac{d}{\lambda}\sin\theta$.

- The inner sum is recognized as the (spatial) **Fourier transform** of each snapshot.

MPDR beamforming for direction finding purposes

Capon's method

If the MPDR beamformer

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

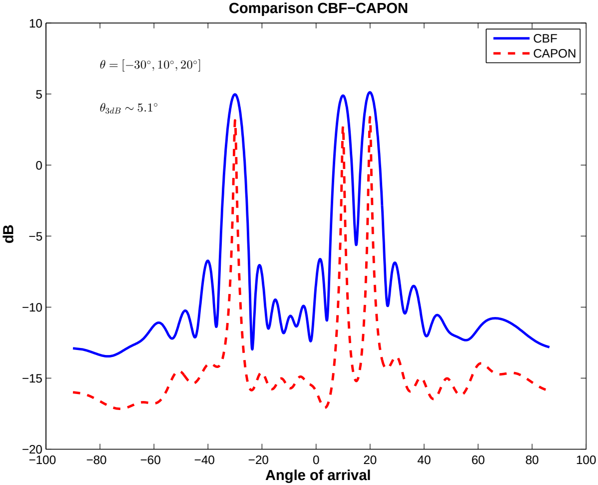
is used, the output power then writes

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

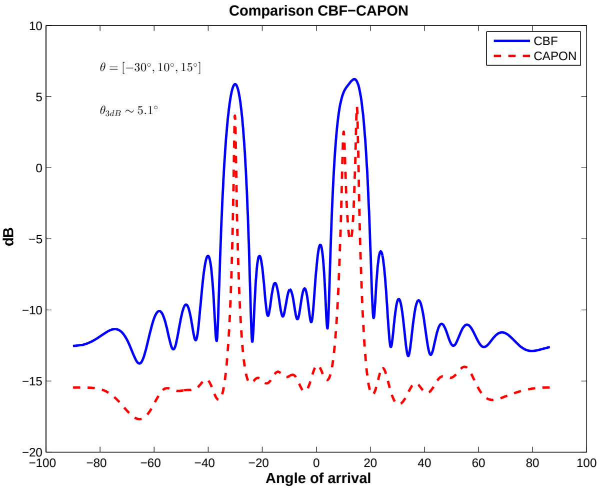
which in practice yields

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

Comparison CBF-Capon (low resolution scenario)



Comparison CBF-Capon (high resolution scenario)



Model-based methods

Principle

Based on the model

$$\mathbf{y}(k) = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}(k) + \mathbf{n}(k)$$

where $\boldsymbol{\theta} = \begin{bmatrix}\theta_1 & \theta_2 & \cdots & \theta_P\end{bmatrix}^T$,

$$\mathbf{A}(\boldsymbol{\theta}) = \begin{bmatrix}\mathbf{a}(\theta_1) & \mathbf{a}(\theta_2) & \cdots & \mathbf{a}(\theta_P)\end{bmatrix}$$
$$\mathbf{s}(k) = \begin{bmatrix}s_1(k) & s_2(k) & \cdots & s_P(k)\end{bmatrix}^T$$

and $\mathbf{a}(\theta)$ stands for the steering vector.

Classes of methods

- **Maximum Likelihood methods** are based on maximizing the likelihood function, which amounts to finding the unknown parameters which make the observed data the more likely.
- **Subspace-based methods** rely on the fact that the signal subspace coincides with the subspace spanned by the principal eigenvectors of $\mathbf{R}$. Moreover, the latter is orthogonal to the subspace spanned by the minor eigenvectors. These two algebraic properties are exploited for direction finding.
- **Covariance matching** relies on a model $\mathbf{R}(\boldsymbol{\eta})$ for the covariance matrix and looks for the model parameters which minimize the distance between $\mathbf{R}(\boldsymbol{\eta})$ and the sample covariance matrix $\hat{\mathbf{R}}$.

Maximum Likelihood Estimation

- The MLE consists in finding the parameter vector $\boldsymbol{\eta}$ which maximizes the likelihood function $p(\mathbf{Y};\boldsymbol{\eta})$ of the snapshots $\mathbf{Y} = \begin{bmatrix}\mathbf{y}(1) & \mathbf{y}(2) & \cdots & \mathbf{y}(k)\end{bmatrix}$, where $\boldsymbol{\eta}$ is the model parameter vector.

☺ Asymptotically efficient.

☹ Multi-dimensional optimization problem (usually) $\Rightarrow$ computational complexity, possible convergence to local maxima.

Stochastic (unconditional) MLE

- Assume that $\mathbf{s}(k)$ is Gaussian distributed with $\mathbb{E}\left\{\mathbf{s}(k)\right\} = \mathbf{0}$, and a covariance matrix $\mathbf{R}_s = \mathbb{E}\left\{\mathbf{s}(k)\mathbf{s}^H(k)\right\}$ which is *full rank*.
- The distribution of the snapshots is thus given by

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

- The likelihood function can be written as

$$p(\mathbf{Y};\boldsymbol{\eta}) = \prod_{k=1}^{K} \pi^{-N}|\mathbf{R}|^{-1}e^{-\mathbf{y}(k)^H\mathbf{R}^{-1}\mathbf{y}(k)}$$

Stochastic (unconditional) MLE

- The ML estimate is obtained as

$$\hat{\boldsymbol{\eta}} = \arg \min_{\boldsymbol{\theta}, \mathbf{R}_s, \sigma^2} -\log p(\mathbf{Y}; \boldsymbol{\eta})$$
$$= \arg \min_{\boldsymbol{\theta}, \mathbf{R}_s, \sigma^2} \log |\mathbf{R}| + \mathrm{Tr}\left\{\mathbf{R}^{-1}\hat{\mathbf{R}}\right\}$$

- Closed-form solutions for σ^2 and $\mathbf{R}_s$ can be obtained so that the likelihood function is concentrated, yielding a minimization over the angles only:

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

Deterministic (conditional) MLE

- The signal waveforms are assumed deterministic so that

$$\mathbf{y}(k) \sim \mathcal{CN}\left(\mathbf{A}(\boldsymbol{\theta})\mathbf{s}(k), \sigma^2 \mathbf{I}\right)$$

- The MLE is now given by

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

- The likelihood function can be concentrated with respect to all $\mathbf{s}(k)$ and σ^2, and finally

$$\hat{\boldsymbol{\theta}}^{\mathsf{det}} = \arg \min_{\boldsymbol{\theta}} \operatorname{Tr}\left\{\mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta})\hat{\mathbf{R}}\right\}$$

- For a single source $\hat{\theta}^{\mathsf{det}} = \arg\max_{\theta} \frac{1}{N}\mathbf{a}^H(\theta)\hat{\mathbf{R}}\mathbf{a}(\theta) \equiv$ CBF.

Subspace-based methods

Eigenvalue decomposition of the covariance matrix

If P signals are present, one has

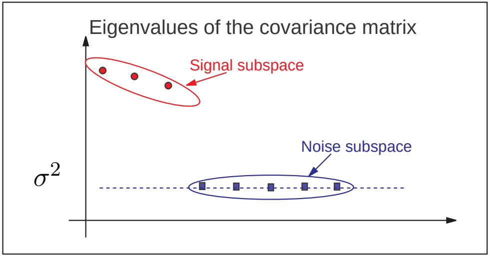
$$\begin{aligned}
\mathbf{R} &= \overbrace{\mathbf{A}(\boldsymbol{\theta})\mathbf{R}_s\mathbf{A}^H(\boldsymbol{\theta})}^{\mathcal{R}\{\mathbf{A}(\boldsymbol{\theta})\},\mathrm{rank}=P} + \sigma^2\mathbf{I} \qquad (\mathbf{R}_s \text{ assumed full-rank}) \\
&= \sum_{p=1}^{P} \lambda_p \mathbf{u}_p\mathbf{u}_p^H + 0 \sum_{p=P+1}^{N} \mathbf{u}_p\mathbf{u}_p^H + \sigma^2\mathbf{I} \\
&= \sum_{p=1}^{P} \lambda_p \mathbf{u}_p\mathbf{u}_p^H + 0 \sum_{p=P+1}^{N} \mathbf{u}_p\mathbf{u}_p^H + \sigma^2 \sum_{p=1}^{N} \mathbf{u}_p\mathbf{u}_p^H \\
&= \mathbf{U}_s\boldsymbol{\Lambda}_s\mathbf{U}_s^H + \sigma^2\mathbf{U}_n\mathbf{U}_n^H
\end{aligned}$$

where $\mathbf{U}_s = \begin{bmatrix}\mathbf{u}_1 & \ldots \mathbf{u}_P\end{bmatrix} \perp \mathbf{U}_n = \begin{bmatrix}\mathbf{u}_{P+1} & \ldots \mathbf{u}_N\end{bmatrix}$.

Subspace-based methods

Signal and noise subspaces

- ◐ $\mathcal{R}\{\mathbf{U}_s\} = \mathcal{R}\{\mathbf{A}(\boldsymbol{\theta})\}$: the signal subspace is spanned by $\mathbf{U}_s$ and hence $\mathbf{U}_s = \mathbf{A}(\boldsymbol{\theta})\mathbf{T}$ for some non-singular matrix $\mathbf{T}$.
- ◑ $\mathcal{R}\{\mathbf{U}_n\}$ is orthogonal to $\mathcal{R}\{\mathbf{U}_s\} = \mathcal{R}\{\mathbf{A}(\boldsymbol{\theta})\} \Rightarrow \mathbf{A}^H(\boldsymbol{\theta})\mathbf{U}_n = \mathbf{0}$.



$\Rightarrow$ Subspace-based methods rely on either ◐ or ◑.

MUSIC

- **The signal steering vectors are orthogonal to U_n**

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

- One looks for the P largest maxima of

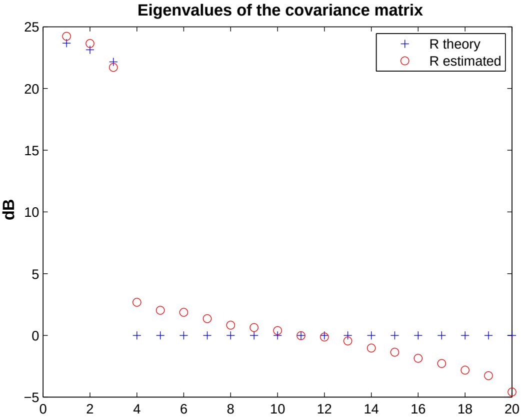
$$P_{\mathsf{MUSIC}}(\theta) = \frac{1}{\mathbf{a}^H(\theta)\hat{\mathbf{U}}_n \hat{\mathbf{U}}_n^H \mathbf{a}(\theta)} = \frac{1}{\sum_{n=P+1}^{N} |\mathbf{a}^H(\theta)\hat{\mathbf{u}}_n|^2}$$

  on the rationale that, as K grows large, $\hat{\mathbf{U}}_n \to \mathbf{U}_n$ and hence $P_{\mathsf{MUSIC}}(\theta_p) \to \infty$.
- Many variants around MUSIC, e.g., SSMUSIC (Mc Cloud & Scharf).

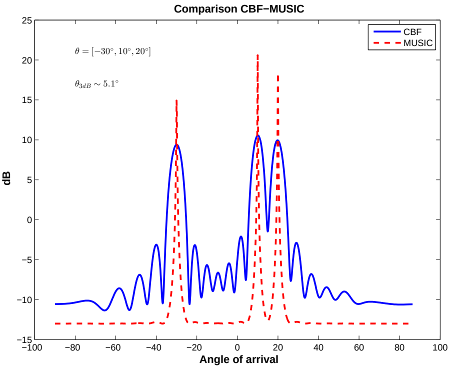
Root-MUSIC

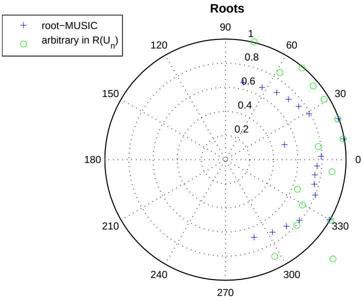
- Let $\mathbf{a}(z) = \begin{bmatrix} 1 & z & \cdots & z^{N-1} \end{bmatrix}^T$. For a ULA, one can compute the P roots of
$$P_{\text{MUSIC}}(z) = \mathbf{a}^T(z^{-1})\hat{\mathbf{U}}_n\hat{\mathbf{U}}_n^H\mathbf{a}(z)$$
closest to the unit circle. The reason is that
$$\mathbf{a}^T(e^{-i2\pi\frac{d}{\lambda}\sin\theta_p})\mathbf{U}_n\mathbf{U}_n^H\mathbf{a}(e^{i2\pi\frac{d}{\lambda}\sin\theta_p}) = \mathbf{a}^H(\theta_p)\mathbf{U}_n\mathbf{U}_n^H\mathbf{a}(\theta_p) = 0$$
- $P_{\text{MUSIC}}(z) = \sum_{n=-(N-1)}^{N-1} p_n z^{-n}$ has $2(N-1)$ roots, $(N-1)$ of which inside the unit circle since
$$\begin{aligned} P_{\text{MUSIC}}(1/z^*) &= \mathbf{a}^T(z^*)\hat{\mathbf{U}}_n\hat{\mathbf{U}}_n^H\mathbf{a}(1/z^*) \\ &= \mathbf{a}^H(z)\hat{\mathbf{U}}_n\hat{\mathbf{U}}_n^H\mathbf{a}^*(z^{-1}) \\ &= \mathbf{a}^T(z^{-1})\hat{\mathbf{U}}_n\hat{\mathbf{U}}_n^H\mathbf{a}(z) \\ &= P_{\text{MUSIC}}(z) \end{aligned}$$

Low-resolution scenario

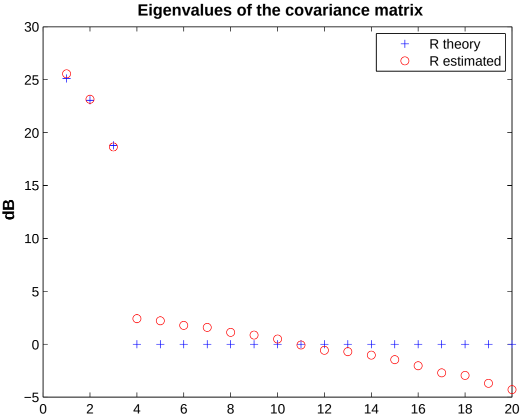


Low-resolution scenario

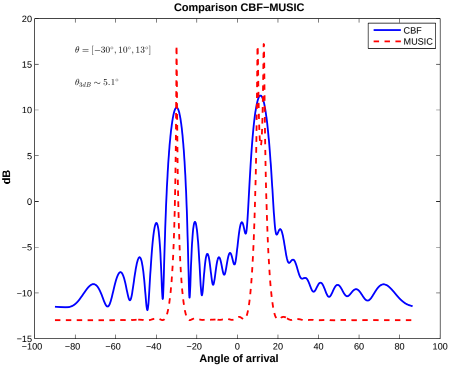


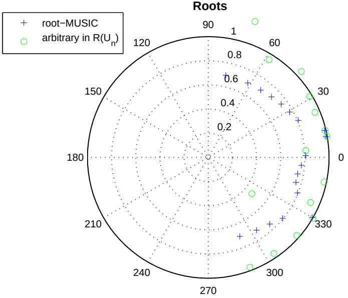


High-resolution scenario



High-resolution scenario





Min-norm

- Let $\mathbf{d} = \mathbf{U}_n \eta = \begin{bmatrix} d_0 & d_1 & \cdots & d_{N-1} \end{bmatrix}^T$ be an arbitrary vector in the noise subspace.
- Since $\mathbf{d} \perp \mathbf{a}(\theta_p)$, $D(z) = \sum_{n=0}^{N-1} d_n z^{-n}$ has P of its roots equal to $e^{i2\pi \frac{d}{\lambda} \sin\theta_p}$ and hence can serve to estimate θ_p.
- The min-norm method searches the **vector in $\mathcal{R}\{\mathbf{U}_n\}$ with minimal norm**. To avoid $\mathbf{d} = \mathbf{0}$, one considers

$$\min_{\mathbf{d} \in \mathcal{R}\{\mathbf{U}_n\}} \|\mathbf{d}\|^2 \text{ s. t. } d_0 = 1 \Leftrightarrow \min_{\eta} \|\eta\|^2 \text{ s. t. } \eta^H \mathbf{U}_n^H \mathbf{e}_1 = 1$$

where $\mathbf{e}_1 = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \end{bmatrix}^T$. The solution is

$$\eta_\star = \frac{\mathbf{U}_n^H \mathbf{e}_1}{\mathbf{e}_1^H \mathbf{U}_n \mathbf{U}_n^H \mathbf{e}_1} \Rightarrow \mathbf{d}_{\text{Min-Norm}} = \frac{\mathbf{U}_n \mathbf{U}_n^H \mathbf{e}_1}{\mathbf{e}_1^H \mathbf{U}_n \mathbf{U}_n^H \mathbf{e}_1}$$

ESPRIT

- Let us partition $\mathbf{A} = \mathbf{A}(\boldsymbol{\theta})$ as

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

  where $\mathbf{A}_1$ [resp. $\mathbf{A}_2$] contains all but the last [resp. first] row of $\mathbf{A}$.
- Then, for a ULA, we have

$$\mathbf{A}_2 = \mathbf{A}_1 \boldsymbol{\Phi}; \quad \boldsymbol{\Phi} = \mathrm{diag}\left(\{ e^{i2\pi \frac{d}{\lambda} \sin\theta_p} \}_{p=1}^{P} \right) \tag{1}$$

- $\boldsymbol{\Phi}$ conveys the useful information and can be deduced from $(\mathbf{A}_1, \mathbf{A}_2)$. The latter are unknown but $\mathbf{U}_s = \mathbf{AT} \Rightarrow$ can we find a similar relation for $\mathbf{U}_s$?

ESPRIT

- Let us partition $\mathbf{U}_s$ as $\mathbf{A}$, i.e. $\mathbf{U}_s = \begin{bmatrix} \mathbf{U}_{s1} \\ - \end{bmatrix} = \begin{bmatrix} - \\ \mathbf{U}_{s2} \end{bmatrix}$. Then

$$\mathbf{U}_s = \mathbf{A}\mathbf{T} \Rightarrow \begin{cases} \mathbf{U}_{s1} = \mathbf{A}_1\mathbf{T} \\ \mathbf{U}_{s2} = \mathbf{A}_2\mathbf{T} = \mathbf{A}_1\mathbf{\Phi}\mathbf{T} \end{cases}$$

$$\Rightarrow \mathbf{U}_{s2} = \mathbf{U}_{s1}\mathbf{T}^{-1}\mathbf{\Phi}\mathbf{T}$$

$$\Rightarrow \mathbf{U}_{s2} = \mathbf{U}_{s1}\mathbf{\Psi}$$

- **$\mathbf{\Psi}$ and $\mathbf{\Phi}$ share the same eigenvalues** since $\mathbf{\Phi}\mathbf{u} = \lambda\mathbf{u}$ implies that $\mathbf{\Psi}\mathbf{T}^{-1}\mathbf{u} = \mathbf{T}^{-1}\mathbf{\Phi}\mathbf{u} = \lambda\mathbf{T}^{-1}\mathbf{u}$.
- It follows that the eigenvalues of $\mathbf{\Psi}$ are $\left\{e^{i2\pi\frac{d}{\lambda}\sin\theta_p}\right\}_{p=1}^{P}$.
- In practice one solves in a least-squares sense $\hat{\mathbf{U}}_{s2} = \hat{\mathbf{U}}_{s1}\mathbf{\Psi}$ and computes the eigenvalues of $\hat{\mathbf{\Psi}}$.

Subspace Fitting

- Since $\mathcal{R}\left\{\mathbf{U}_s\right\} = \mathcal{R}\left\{\mathbf{A}(\boldsymbol{\theta})\right\}$, there exists a full-rank matrix $\mathbf{T}$ $(P \times P)$ such that

$$\boxed{\mathbf{U}_s = \mathbf{A}(\boldsymbol{\theta})\mathbf{T}}$$

- The idea is to look for the DOA which minimize the error between the subspaces spanned by $\hat{\mathbf{U}}_s$ and $\mathbf{A}(\boldsymbol{\theta})$:

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

Subspace Fitting

- There exists a closed-form solution for $\mathbf{T}$ and finally

$$\hat{\boldsymbol{\theta}}^{\mathsf{SSF}} = \arg\min_{\boldsymbol{\theta}} \mathrm{Tr}\left\{\mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta})\hat{\mathbf{U}}_s\mathbf{W}\hat{\mathbf{U}}_s^H\right\}$$

- Alternative: use the fact that

$$\boxed{\mathcal{R}\left\{\mathbf{U}_n\right\} = \mathcal{N}\left\{\mathbf{A}^H(\boldsymbol{\theta})\right\} \Rightarrow \mathbf{U}_n^H\mathbf{A}(\boldsymbol{\theta}) = \mathbf{0}}$$

and estimate the angles as

$$\hat{\boldsymbol{\theta}}^{\mathsf{NSF}} = \arg\min_{\boldsymbol{\theta}} \left\|\hat{\mathbf{U}}_n^H\mathbf{A}(\boldsymbol{\theta})\right\|_{\mathbf{W}}^2$$

Covariance fitting

- The covariance matrix is given by $\mathbf{R}(\boldsymbol{\theta}, \mathbf{P}, \sigma) = \mathbf{R}_s(\boldsymbol{\theta}, \mathbf{P}) + \mathbf{Q}(\sigma)$

$$\mathbf{r} = \text{vec}(\mathbf{R}) = \boldsymbol{\Psi}(\boldsymbol{\theta})\mathbf{P} + \boldsymbol{\Sigma}\sigma = \begin{bmatrix} \boldsymbol{\Psi}(\boldsymbol{\theta}) & \boldsymbol{\Sigma} \end{bmatrix} \begin{bmatrix} \mathbf{P} \\ \sigma \end{bmatrix} \triangleq \boldsymbol{\Phi}(\boldsymbol{\theta})\boldsymbol{\alpha}$$

- The parameters are estimated by minimizing the error between $\mathbf{R}$ and its estimate $\hat{\mathbf{R}}$:

$$\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\alpha}} = \arg\min \left[\hat{\mathbf{r}} - \boldsymbol{\Phi}(\boldsymbol{\theta})\boldsymbol{\alpha}\right] \mathbf{W}^{-1} \left[\hat{\mathbf{r}} - \boldsymbol{\Phi}(\boldsymbol{\theta})\boldsymbol{\alpha}\right]$$

- The criterion can be concentrated with respect to $\boldsymbol{\alpha}$: minimization with respect to $\boldsymbol{\theta}$ only.

Covariance fitting

- In case of independent Gaussian distributed snaphots, $\mathbf{W}_{\text{opt}} = \mathbf{R}^T \otimes \mathbf{R}$ and covariance matching estimates are asymptotically (i.e. when $K \to \infty$) equivalent to ML estimates.
- In contrast to MLE, no need for assumptions on the pdf, only an assumption on $\mathbf{R}$. The criterion is usually simpler to minimize.
- Covariance matching can be used with full-rank covariance matrix $\mathbf{R}_s$ while subspace methods require the latter to be rank deficient.

Synthesis

	Hypotheses	Algorithm	Performance	Problems
ML	distribution	optimization	optimal	Computational cost
COMET	$\mathbf{R}$	optimization	$\simeq$ optimal	Computational cost
MUSIC	$\mathbf{R}$	EVD	$\simeq$ optimal	Coherent signals

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.

References

1. H.L. Van Trees, *Optimum Array Processing*, John Wiley, 2002
2. D.G. Manolakis, V.K. Ingle et S.M. Kogon, *Statistical and Adaptive Signal Processing*, ch. 11, McGraw-Hill, 2000
3. Y. Hua, A.B. Gershman et Q. Cheng (Editeurs), *High-Resolution and Robust Signal Processing*, Marcel Dekker, 2004
4. J.R. Guerci, *Space-Time Adaptive Processing*, Artech House, 2003
5. S. A. Vorobyov, *Adaptive and robust beamforming*, A. M. Zoubir, M. Viberg, R. Chellappa, and S. Theodoridis, editors, Academic Press Library in Signal Processing, vol. 3, pp. 503-552, Elsevier, 2014.