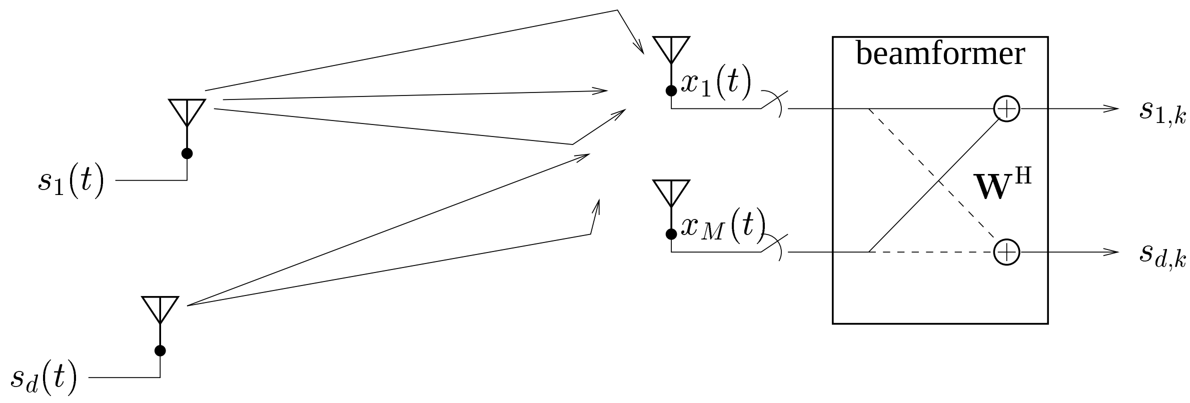
3. SPATIAL PROCESSING TECHNIQUES

Outline

1. Matched and Wiener filters – deterministic approach
2. Matched and Wiener filters – stochastic approach
3. Direction estimation
4. Spatio-temporal generalizations

Data model



- Assume we receive d signals on an antenna array, narrow-band case:

$$\mathbf{x}_k := \mathbf{x}(k) = \sum_{i=1}^{d} \mathbf{a}_i s_i(k) + \mathbf{n}(k) := \sum_{i=1}^{d} \mathbf{a}_i s_{i,k} + \mathbf{n}_k \; = \; \mathbf{A}\mathbf{s}_k + \mathbf{n}_k$$

- Objective:
  - Construct a receiver weight vector $\mathbf{w}_i$ such that

$$\mathbf{w}_i^{\mathrm{H}} \mathbf{x}_k = \hat{s}_{i,k}$$

  - Construct a receiver weight matrix $\mathbf{W}$ such that

$$\mathbf{W}^{\mathrm{H}} \mathbf{x}_k = \hat{\mathbf{s}}_k$$

Deterministic approach

Noiseless case $\mathbf{x}_k = \mathbf{A}\mathbf{s}_k \quad \Leftrightarrow \quad \mathbf{X} = \mathbf{A}\mathbf{S}$

- Objective: find **W** such that $\mathbf{W}^{\mathrm{H}}\mathbf{X} = \mathbf{S}$
- With **A** known (e.g. after channel estimation):

$$\mathbf{X} = \mathbf{A}\mathbf{S} \;\Rightarrow\; \mathbf{S} = \mathbf{A}^{\dagger}\mathbf{X} = (\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{H}}\mathbf{X}$$

  Hence we set

$$\mathbf{W}^{\mathrm{H}} = \mathbf{A}^{\dagger}$$

- With **S** known (e.g. after synchronization and training):

$$\mathbf{W}^{\mathrm{H}}\mathbf{X} = \mathbf{S} \;\Rightarrow\; \mathbf{W}^{\mathrm{H}} = \mathbf{S}\mathbf{X}^{\dagger} = \mathbf{S}\mathbf{X}^{\mathrm{H}}(\mathbf{X}\mathbf{X}^{\mathrm{H}})^{-1}$$

  Further, we have that

$$\mathbf{A} = (\mathbf{W}^{\mathrm{H}})^{\dagger}$$

- In both cases: $\mathbf{W}^{\mathrm{H}}\mathbf{A} = \mathbf{I}$: all interference is cancelled.

Deterministic approach

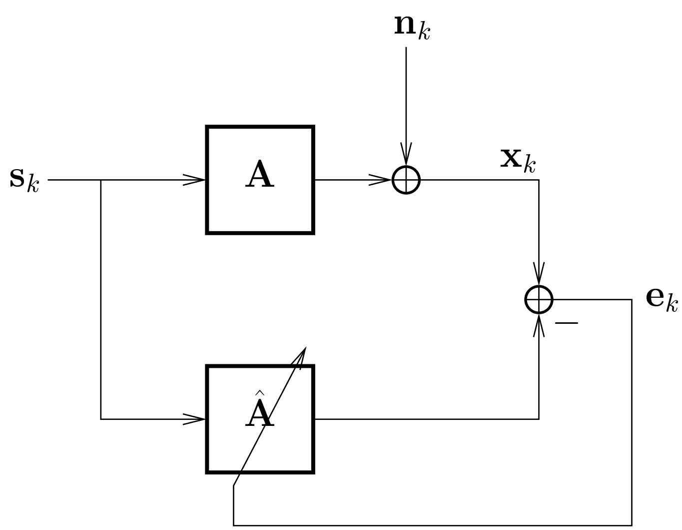
Noisy case: $\mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{N}$

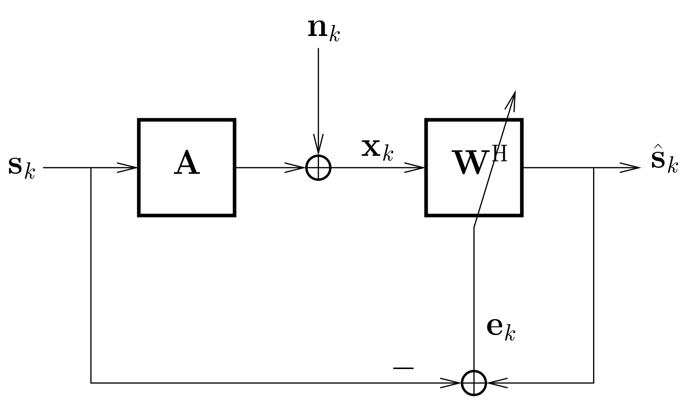
- *Model matching:* minimize residual

$$\min_{\mathbf{S}} \|\mathbf{X} - \mathbf{A}\mathbf{S}\|_F^2, \qquad \text{or} \qquad \min_{\mathbf{A}} \|\mathbf{X} - \mathbf{A}\mathbf{S}\|_F^2$$

- *Output error minimization:*

$$\min_{\mathbf{W}} \|\mathbf{W}^H\mathbf{X} - \mathbf{S}\|_F^2,$$





Deterministic approach

Model matching

- *With* **A** *known:*

$$\hat{\mathbf{S}} = \arg\min_{\mathbf{S}} \|\mathbf{X} - \mathbf{A}\mathbf{S}\|_F^2 \quad \Rightarrow \quad \hat{\mathbf{S}} = \mathbf{A}^\dagger \mathbf{X} \quad \Rightarrow \quad \mathbf{W}^H = \mathbf{A}^\dagger$$

This is the *Zero-Forcing (ZF) receiver*.

- It maximizes the output Signal-to-Interference Ratio (SIR).
- It might boost the noise:

  Since $\hat{\mathbf{S}} = \mathbf{W}^H \mathbf{X} = \mathbf{S} + \mathbf{A}^\dagger \mathbf{N}$, the output noise depends on $\mathbf{A}^\dagger$

$$\mathbf{A} = \mathbf{U}_A \Sigma_A \mathbf{V}_A^H \quad \rightarrow \quad \mathbf{A}^\dagger = \mathbf{V}_A \Sigma_A^{-1} \mathbf{U}_A^H ,$$

  If Σ_A^{-1} is large (i.e., **A** is ill conditioned), the output noise is large.

- *With* **S** *known:*

$$\hat{\mathbf{A}} = \arg\min_{\mathbf{A}} \|\mathbf{X} - \mathbf{A}\mathbf{S}\|_F^2 \quad \Rightarrow \quad \hat{\mathbf{A}} = \mathbf{X}\mathbf{S}^\dagger = \mathbf{X}\mathbf{S}^H (\mathbf{S}\mathbf{S}^H)^{-1}$$

This does not specify the beamformer, but it is natural to set $\mathbf{W}^H = \hat{\mathbf{A}}^\dagger$.

Deterministic approach

Output error minimization

- *With* **S** *known*:

$$\mathbf{W}^{\mathrm{H}} = \arg\min_{\mathbf{W}} \| \mathbf{W}^{\mathrm{H}}\mathbf{X} - \mathbf{S} \|_{\mathrm{F}}^2 = \mathbf{S}\mathbf{X}^{\dagger} = \mathbf{S}\mathbf{X}^{\mathrm{H}}(\mathbf{X}\mathbf{X}^{\mathrm{H}})^{-1} = \hat{\mathbf{R}}_{xs}^{\mathrm{H}}\hat{\mathbf{R}}_{x}^{-1}, \quad \mathbf{W} = \hat{\mathbf{R}}_{x}^{-1}\hat{\mathbf{R}}_{xs}$$

$\hat{\mathbf{R}}_x = \frac{1}{N}\mathbf{X}\mathbf{X}^{\mathrm{H}}$: sample data covariance matrix

$\hat{\mathbf{R}}_{xs} = \frac{1}{N}(\mathbf{X}\mathbf{S}^{\mathrm{H}})$: sample correlation between the sources and the received data

- *With* **A** *known*, and assuming $\frac{1}{N}\mathbf{S}\mathbf{S}^{\mathrm{H}} \to \mathbf{I}$, $\frac{1}{N}\mathbf{N}\mathbf{N}^{\mathrm{H}} \to \sigma^2\mathbf{I}$, and $\frac{1}{N}\mathbf{S}\mathbf{N}^{\mathrm{H}} \to \mathbf{0}$:

$$\hat{\mathbf{R}}_x = \frac{1}{N}\mathbf{X}\mathbf{X}^{\mathrm{H}} = \frac{1}{N}\mathbf{A}\mathbf{S}\mathbf{S}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}} + \frac{1}{N}\mathbf{N}\mathbf{N}^{\mathrm{H}} + \frac{1}{N}\mathbf{A}\mathbf{S}\mathbf{N}^{\mathrm{H}} + \frac{1}{N}\mathbf{N}\mathbf{S}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}} \quad \to \quad \mathbf{A}\mathbf{A}^{\mathrm{H}} + \sigma^2\mathbf{I}$$

$$\hat{\mathbf{R}}_{xs} = \frac{1}{N}\mathbf{X}\mathbf{S}^{\mathrm{H}} = \frac{1}{N}\mathbf{A}\mathbf{S}\mathbf{S}^{\mathrm{H}} + \frac{1}{N}\mathbf{N}\mathbf{S}^{\mathrm{H}} \quad \to \quad \mathbf{A}$$

$$\mathbf{W} = (\mathbf{A}\mathbf{A}^{\mathrm{H}} + \sigma^2\mathbf{I})^{-1}\mathbf{A}$$

This is the *Linear Minimum Mean Square Error (LMMSE)* or *Wiener receiver*.

- It makes a compromise between interference and noise cancellation.
- It maximizes the output Signal-to-Interference-plus-Noise Ratio (SINR).

Stochastic approach

Stochastic model matching

$$\mathbf{x}_k = \mathbf{A}\mathbf{s}_k + \mathbf{n}_k \qquad \Leftrightarrow \qquad \mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{N}$$

- **A** and **S** assumed to be deterministic (known or unknown).
- Noise spatially white and *jointly complex Gaussian distributed*:

$$\mathbf{n}_k \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}) \qquad \Leftrightarrow \qquad p(\mathbf{n}_k) = \frac{1}{\sqrt{\pi}\sigma} e^{-\frac{\|\mathbf{n}_k\|^2}{\sigma^2}}$$

- $\mathbf{n}_k = \mathbf{x}_k - \mathbf{A}\mathbf{s}_k$, so the probability to receive a certain vector $\mathbf{x}_k$ given **A** and $\mathbf{s}_k$ is

$$p(\mathbf{x}_k | \mathbf{A}, \mathbf{s}_k) = \frac{1}{\sqrt{\pi}\sigma} e^{-\frac{\|\mathbf{x}_k - \mathbf{A}\mathbf{s}_k\|^2}{\sigma^2}}$$

- If noise is temporally white, we obtain

$$p(\mathbf{X} | \mathbf{A}, \mathbf{S}) = \prod_{k=1}^{N} \frac{1}{\sqrt{\pi}\sigma} e^{-\frac{\|\mathbf{x}_k - \mathbf{A}\mathbf{s}_k\|^2}{\sigma^2}} = \left(\frac{1}{\sqrt{\pi}\sigma}\right)^N e^{-\frac{\|\mathbf{X} - \mathbf{A}\mathbf{S}\|_F^2}{\sigma^2}}$$

Stochastic approach

- Making abstraction of the constant term, we obtain

$$p(\mathbf{X}|\mathbf{A},\mathbf{S}) = \text{const} \cdot e^{-\frac{\|\mathbf{X}-\mathbf{A}\mathbf{S}\|_F^2}{\sigma^2}}$$

- $p(\mathbf{X}|\mathbf{A},\mathbf{S})$ is the *likelihood* of receiving a data matrix **X**, for a given **A** and **S**.

- *Deterministic Maximum Likelihood (DML)*:

  Estimate **A** and/or **S** as maximizing the likelihood of the actual received **X**

$$\begin{aligned}(\hat{\mathbf{A}},\hat{\mathbf{S}}) &= \arg\max_{\mathbf{A},\mathbf{S}} e^{-\frac{\|\mathbf{X}-\mathbf{A}\mathbf{S}\|_F^2}{\sigma^2}} \\ &= \arg\min_{\mathbf{A},\mathbf{S}} \|\mathbf{X}-\mathbf{A}\mathbf{S}\|_\mathrm{F}^2\end{aligned}$$

- For white Gaussian noise, DML is equivalent to deterministic model matching

Stochastic approach

Stochastic output error minimization

- Minimize the Linear Minimum Mean Square Error cost:

$$\min_{\mathbf{w}_i} J(\mathbf{w}_i) \;=\; \min_{\mathbf{w}_i} \mathrm{E}\left[|\mathbf{w}_i^{\mathrm{H}}\mathbf{x}_k - s_{i,k}|^2\right].$$

- The solution for **W** is then obtained by stacking the solutions for $\mathbf{w}_i$

- It can be worked out as follows:

$$\begin{aligned} J(\mathbf{w}_i) &= \mathrm{E}\left[|\mathbf{w}_i^{\mathrm{H}}\mathbf{x}_k - s_{i,k}|^2\right] \\ &= \mathbf{w}_i^{\mathrm{H}}\mathrm{E}[\mathbf{x}_k\mathbf{x}_k^{\mathrm{H}}]\mathbf{w}_i \;-\; \mathbf{w}_i^{\mathrm{H}}\mathrm{E}[\mathbf{x}_k\bar{s}_{i,k}] \;-\; \mathrm{E}[s_{i,k}\mathbf{x}_k^{\mathrm{H}}]\mathbf{w} \;+\; \mathrm{E}[|s_{i,k}|^2] \\ &= \mathbf{w}_i^{\mathrm{H}}\mathbf{R}_x\mathbf{w}_i \;-\; \mathbf{w}_i^{\mathrm{H}}\mathbf{r}_{xs,i} \;-\; \mathbf{r}_{xs,i}^{\mathrm{H}}\mathbf{w}_i + 1 \end{aligned}$$

Note that $\mathbf{r}_{xs,i} = \mathrm{E}[\mathbf{x}_k\bar{s}_{i,k}]$ is the i-th column of $\mathbf{R}_{xs} = \mathrm{E}[\mathbf{x}_k\mathbf{s}^{\mathrm{H}}]$.

Stochastic approach

$$J(\mathbf{w}_i) \;=\; \mathbf{w}_i^{\mathrm{H}}\mathbf{R}_x\mathbf{w}_i \;-\; \mathbf{w}_i^{\mathrm{H}}\mathbf{r}_{xs,i} \;-\; \mathbf{r}_{xs,i}^{\mathrm{H}}\mathbf{w}_i + r_{s,i}$$

- Differentiate with respect to $\mathbf{w}_i$:

  Let $\mathbf{w}_i = \mathbf{u} + j\mathbf{v}$ with $\mathbf{u}$ and $\mathbf{v}$ real-valued, then the gradient is *defined* as

$$\nabla_{\mathbf{w}_i} J = \frac{1}{2}(\nabla_{\mathbf{u}} J - j\nabla_{\mathbf{v}} J), \quad \nabla_{\bar{\mathbf{w}}_i} J = \frac{1}{2}(\nabla_{\mathbf{u}} J + j\nabla_{\mathbf{v}} J) \quad \text{with} \quad \nabla_{\mathbf{x}\in\mathbb{R}^N} J = \begin{bmatrix} \frac{\partial}{\partial x_1} J \\ \vdots \\ \frac{\partial}{\partial x_M} J \end{bmatrix},$$

  with properties

$$\nabla_{\bar{\mathbf{w}}_i} \mathbf{w}_i^{\mathrm{H}}\mathbf{r}_{xs,i} = \mathbf{r}_{xs,i}, \quad \nabla_{\bar{\mathbf{w}}_i} \mathbf{r}_{xs,i}^{\mathrm{H}}\mathbf{w}_i = \mathbf{0}, \quad \nabla_{\bar{\mathbf{w}}_i} \mathbf{w}_i^{\mathrm{H}}\mathbf{R}_x\mathbf{w}_i = \mathbf{R}_x\mathbf{w}_i$$

- The minimum of $J(\mathbf{w}_i)$ is attained for

$$\nabla_{\bar{\mathbf{w}}_i} J = \mathbf{R}_x\mathbf{w}_i - \mathbf{r}_{xs,i} = 0 \qquad \Rightarrow \qquad \mathbf{w}_i = \mathbf{R}_x^{-1}\mathbf{r}_{xs,i}$$

- For the total beamforming matrix $\mathbf{W}$, we get

$$\mathbf{W} = [\mathbf{w}_1 \ \cdots \ \mathbf{w}_d] = \mathbf{R}_x^{-1}[\mathbf{r}_{xs,1} \ \cdots \ \mathbf{r}_{xs,d}] = \mathbf{R}_x^{-1}\mathbf{R}_{xs}$$

- We thus obtain the Wiener receiver.

Stochastic approach

Colored noise

Assume noise has a *known* variance $\mathrm{E}[\mathbf{n}_k\mathbf{n}_k^{\mathrm{H}}] = \mathbf{R}_n$.

- *Prewhiten* the data with a square-root factor $\mathbf{R}_n^{-1/2}$:

$$\mathbf{x}_k = \mathbf{A}\mathbf{s}_k + \mathbf{n}_k \quad \Rightarrow \quad \underbrace{\mathbf{R}_n^{-1/2}\mathbf{x}_k}_{\underline{\mathbf{x}}_k} = \underbrace{\mathbf{R}_n^{-1/2}\mathbf{A}}_{\underline{\mathbf{A}}}\mathbf{s}_k + \underbrace{\mathbf{R}_n^{-1/2}\mathbf{n}_k}_{\underline{\mathbf{n}}_k}$$

$$\underline{\mathbf{R}}_n = \mathrm{E}[\underline{\mathbf{n}}_k\underline{\mathbf{n}}_k{}^{\mathrm{H}}] = \mathbf{R}_n^{-1/2}\mathbf{R}_n\mathbf{R}_n^{-1/2} = \mathbf{I} \quad \Rightarrow \quad \underline{\mathbf{n}}_k \text{ is white}$$

- The ZF receiver becomes

$$\mathbf{s}_k = \underline{\mathbf{A}}^{\dagger}\underline{\mathbf{x}}_k = (\underline{\mathbf{A}}^{\mathrm{H}}\underline{\mathbf{A}})^{-1}\underline{\mathbf{A}}^{\mathrm{H}}\underline{\mathbf{x}}_k = (\mathbf{A}^{\mathrm{H}}\mathbf{R}_n^{-1}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{H}}\mathbf{R}_n^{-1}\mathbf{x}_k$$

$$\Rightarrow \quad \mathbf{W} = \mathbf{R}_n^{-1}\mathbf{A}(\mathbf{A}^{\mathrm{H}}\mathbf{R}_n^{-1}\mathbf{A})^{-1}$$

- The Wiener receiver will be the same, since $\mathbf{R}_n$ is not used in the derivation:

$$\underline{\mathbf{W}} = \underline{\mathbf{R}}_x^{-1}\underline{\mathbf{R}}_{xs} = (\mathbf{R}_n^{-1/2}\mathbf{R}_x\mathbf{R}_n^{-1/2})^{-1}\mathbf{R}_n^{-1/2}\mathbf{R}_{xs} = \mathbf{R}_n^{1/2}\mathbf{R}_x^{-1}\mathbf{R}_{xs}$$

$$\Rightarrow \quad \mathbf{W} = \mathbf{R}_n^{-1/2}\underline{\mathbf{W}} = \mathbf{R}_x^{-1}\mathbf{R}_{xs}$$

Maximum Ratio Combining

- Single signal in white noise: $\mathbf{x}_k = \mathbf{a}s_k + \mathbf{n}_k\,, \quad \mathrm{E}[\mathbf{n}_k\mathbf{n}_k^{\mathrm{H}}] = \sigma^2\mathbf{I}$
- The ZF beamformer is given by

$$\mathbf{w} \;=\; \mathbf{a}(\mathbf{a}^{\mathrm{H}}\mathbf{a})^{-1} \;=\; \gamma_1\mathbf{a}$$

- Single signal in colored noise: $\mathbf{x}_k = \mathbf{a}s_k + \mathbf{n}_k\,, \quad \mathrm{E}[\mathbf{n}_k\mathbf{n}_k^{\mathrm{H}}] = \mathbf{R}_n$
- The ZF beamformer is given by

$$\mathbf{w} \;=\; \mathbf{R}_n^{-1}\mathbf{a}(\mathbf{a}^{\mathrm{H}}\mathbf{R}_n^{-1}\mathbf{a})^{-1} \;=\; \gamma_2\mathbf{R}_n^{-1}\mathbf{a}$$

- Note: a scalar multiplication does not change the output SNR.
- $\mathbf{w} = \mathbf{a}$ (white noise) and $\mathbf{w} = \mathbf{R}_n^{-1}\mathbf{a}$ (non-white noise) are known as:

  matched filter, *classical beamformer*, or *Maximum Ratio Combining* (MRC)

Maximum Ratio Combining

- Also the Wiener filter will lead to MRC
  - Wiener receiver in white noise

$$\mathbf{w} = \mathbf{R}_x^{-1}\mathbf{r}_{xs} = (\mathbf{a}\mathbf{a}^{\mathrm{H}} + \sigma^2\mathbf{I})^{-1}\mathbf{a} = \mathbf{a}(\mathbf{a}^{\mathrm{H}}\mathbf{a} + \sigma^2)^{-1} \quad \sim \quad \mathbf{a}$$

  - Wiener receiver in colored noise

$$\mathbf{w} = \mathbf{R}_x^{-1}\mathbf{r}_{xs} = (\mathbf{a}\mathbf{a}^{\mathrm{H}} + \mathbf{R}_n)^{-1}\mathbf{a} = \mathbf{R}_n^{-1}\mathbf{a}(\mathbf{a}^{\mathrm{H}}\mathbf{R}_n^{-1}\mathbf{a} + 1)^{-1} \quad \sim \quad \mathbf{R}_n^{-1}\mathbf{a}$$

- The colored noise case is relevant also for the following reason:

  with more than one signal, we can write the model as

$$\mathbf{x}_k = \mathbf{A}\mathbf{s}_k + \mathbf{n}_k = \mathbf{a}_1 s_{1,k} + (\mathbf{A}'\mathbf{s}'_k + \mathbf{n}_k)$$

  This is of the form

$$\mathbf{x}_k = \mathbf{a}s_k + \mathbf{n}_k\,, \qquad \mathbf{R}_n = \mathbf{A}'\mathbf{A}'^H + \sigma^2\mathbf{I}$$

  where the “noise” is colored due to the contribution of the interfering sources.

Matched filtering

Maximizing the output SNR

- The matched filter $\mathbf{w} = \mathbf{R}_n^{-1}\mathbf{a}$ maximizes the output SNR.

- Proof:

  We can write $\mathbf{R}_x = \mathbf{R}_a + \mathbf{R}_n$, with $\mathbf{R}_a = \mathbf{a}\mathbf{a}^{\mathrm{H}}$ and $\mathbf{R}_n = \mathrm{E}[\mathbf{n}\mathbf{n}^{\mathrm{H}}]$.

$$\mathrm{SNR}_{out}(\mathbf{w}) = \frac{\mathbf{w}^{\mathrm{H}}\mathbf{R}_a\mathbf{w}}{\mathbf{w}^{\mathrm{H}}\mathbf{R}_n\mathbf{w}}$$

$$\mathbf{w} = \arg\max_{\mathbf{w}} \frac{\mathbf{w}^{\mathrm{H}}\mathbf{R}_a\mathbf{w}}{\mathbf{w}^{\mathrm{H}}\mathbf{R}_n\mathbf{w}}$$

  This is a *Rayleigh quotient*.

  The solution is known to follow from the eigenvalue equation

$$\mathbf{R}_n^{-1}\mathbf{R}_a\mathbf{w} = \lambda_{\max}\mathbf{w}$$

  Easy to see if $\mathbf{R}_n = \mathbf{I}$, otherwise prewhiten.

Matched filtering

Maximizing the output SNR

$$\mathbf{R}_n^{-1}\mathbf{R}_a\mathbf{w} = \lambda_{\max}\mathbf{w}$$

Closed form solution: insert $\mathbf{R}_a = \mathbf{a}\mathbf{a}^{\mathrm{H}}$:

$$\begin{aligned}
\Leftrightarrow \quad & \mathbf{R}_n^{-1}\mathbf{a}\mathbf{a}^{\mathrm{H}}\mathbf{w} &=& \ \lambda_{\max}\mathbf{w} \\
\Leftrightarrow \quad & (\mathbf{R}_n^{-1/2}\mathbf{a})(\mathbf{a}^{\mathrm{H}}\mathbf{R}_n^{-1/2})(\mathbf{R}_n^{1/2}\mathbf{w}) &=& \ \lambda_{\max}(\mathbf{R}_n^{1/2}\mathbf{w}) \\
\Leftrightarrow \quad & \underline{\mathbf{a}}\underline{\mathbf{a}}^{\mathrm{H}}\underline{\mathbf{w}} &=& \ \lambda_{\max}\underline{\mathbf{w}} \\
\Leftrightarrow \quad & \underline{\mathbf{w}} &=& \ \underline{\mathbf{a}}\,, \quad \lambda_{\max} = \underline{\mathbf{a}}^{\mathrm{H}}\underline{\mathbf{a}} \\
\Leftrightarrow \quad & \mathbf{w} &=& \ \mathbf{R}_n^{-1}\mathbf{a}
\end{aligned}$$

Linearly constrained Minimum Variance (LCMV) (MVDR)

Linearly constrained Minimum Variance / Minimum Variance Distortionless Response

- If **a** is known, then we can constrain the beamformer **w** to

$$\mathbf{w}^{\mathrm{H}}\mathbf{a} = 1$$

in order to have a fixed response towards the source.

- The remaining freedom is used to minimize the total output power ("response" or "variance") after beamforming:

$$\min_{\mathbf{w}} \mathbf{w}^{\mathrm{H}}\mathbf{R}_x\mathbf{w} \qquad \text{such that} \quad \mathbf{w}^{\mathrm{H}}\mathbf{a} = 1$$

- Via Lagrange multipliers:

$$\mathbf{w} = \mathbf{R}_x^{-1}\mathbf{a}(\mathbf{a}^{\mathrm{H}}\mathbf{R}_x^{-1}\mathbf{a})^{-1}$$

- Thus, **w** is a scalar multiple of the Wiener receiver.

Linearly constrained Minimum Variance (LCMV) (MVDR)

Generalization

- Introduce a constraint matrix **C** : $M \times L$ ($M > L$) and an L-dimensional vector **f**
- The general LCMV or MVDR problem can then be written as

$$\min_{\mathbf{w}} \mathbf{w}^{\mathrm{H}} \mathbf{R}_x \mathbf{w} \qquad \text{such that} \quad \mathbf{C}^{\mathrm{H}} \mathbf{w} = \mathbf{f}$$

- Solution:

$$\mathbf{w} = \mathbf{R}_x^{-1} \mathbf{C} (\mathbf{C}^{\mathrm{H}} \mathbf{R}_x^{-1} \mathbf{C})^{-1} \mathbf{f}$$

Generalized Sidelobe Canceler

- Decompose

$$\mathbf{w} = \mathbf{w}_0 - \mathbf{v}, \qquad \text{with} \quad \mathbf{w}_0 \in \text{ran}(\mathbf{C}) \quad \perp \quad \mathbf{v} \in \ker(\mathbf{C}^{\mathrm{H}})$$

- Since $\mathbf{C}^{\mathrm{H}}\mathbf{v} = \mathbf{0}$, we obtain

$$\mathbf{C}^{\mathrm{H}}\mathbf{w} = \mathbf{f} \qquad \Rightarrow \qquad \mathbf{C}^{\mathrm{H}}\mathbf{w}_0 = \mathbf{f} \qquad \Rightarrow \qquad \mathbf{w}_0 = \mathbf{C}(\mathbf{C}^{\mathrm{H}}\mathbf{C})^{-1}\mathbf{f}$$

- Let $\mathbf{C}_n$ be a basis for $\ker(\mathbf{C}^{\mathrm{H}})$, then $\mathbf{v} = \mathbf{C}_n\mathbf{w}_n$ for some $\mathbf{w}_n$

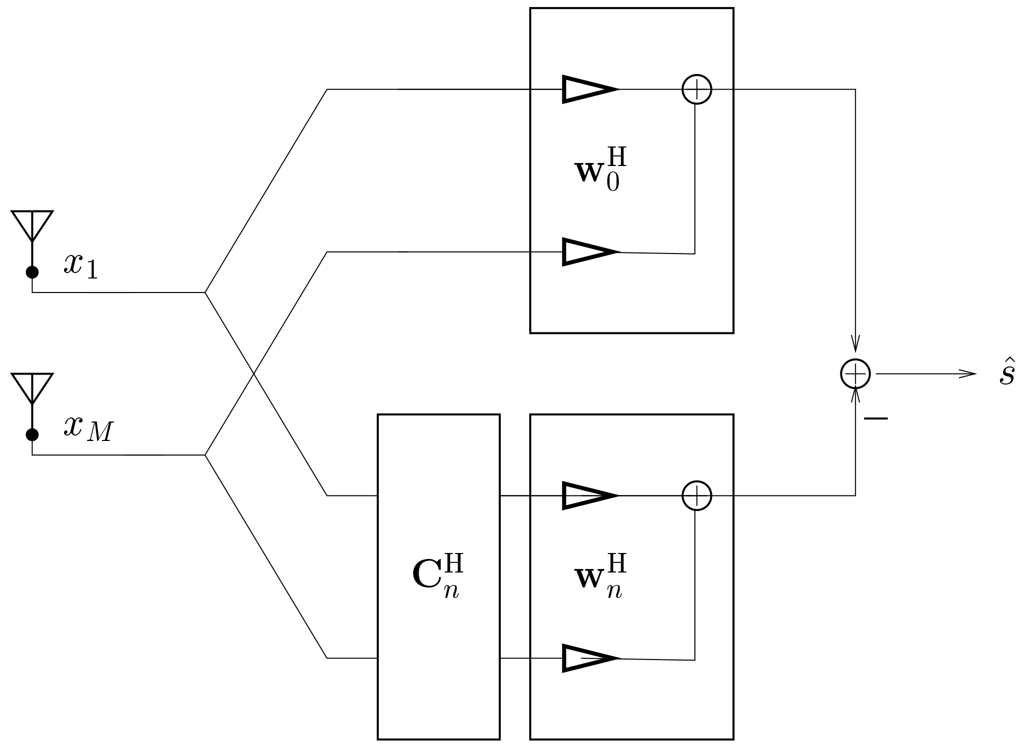
$$\min_{\mathbf{w}} \mathbf{w}^{\mathrm{H}}\mathbf{R}_x\mathbf{w} \qquad \text{such that} \quad \mathbf{C}^{\mathrm{H}}\mathbf{w} = \mathbf{f}$$

$$\Rightarrow \qquad \min_{\mathbf{w}_n} [\mathbf{w}_0 - \mathbf{C}_n\mathbf{w}_n]^{\mathrm{H}}\mathbf{R}_x[\mathbf{w}_0 - \mathbf{C}_n\mathbf{w}_n]\,.$$

- The solution is

$$\mathbf{w}_n \;=\; (\mathbf{C}_n^{\mathrm{H}}\mathbf{R}_x\mathbf{C}_n)^{-1}\mathbf{C}_n^{\mathrm{H}}\mathbf{R}_x\mathbf{w}_0\,.$$

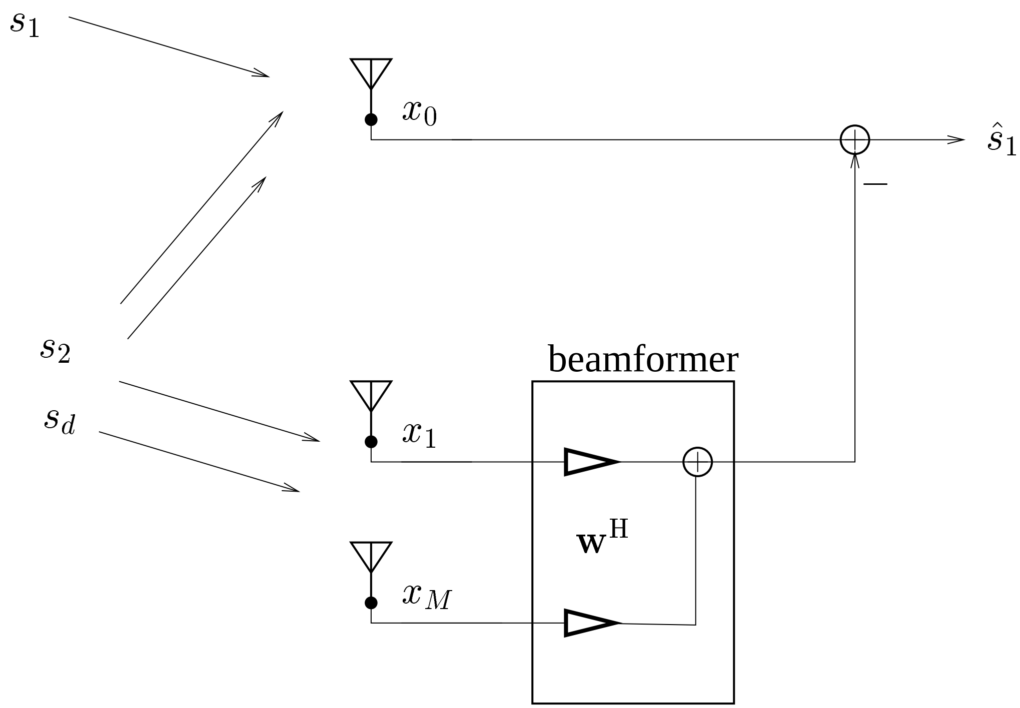
Generalized Sidelobe Canceler



Advantages of this scheme:

- The constraint is always satisfied;
- The size of $\mathbf{w}_n$ is smaller than the size of $\mathbf{w}$;
- It is easy to make $\mathbf{w}_n$ *adaptive*.

Reference channels – Multiple sidelobe canceler



- Interference is first estimated from the reference antenna array.
- It is then subtracted from the primary antenna x_0.

$$\min_{\mathbf{w}} \mathrm{E}\|x_0 - \mathbf{w}^{\mathrm{H}}\mathbf{x}\,\|^2 \quad \Rightarrow \quad \mathbf{w} = \mathbf{R}_x^{-1}\mathbf{r}, \quad \mathbf{R}_x := \mathrm{E}[\mathbf{x}\mathbf{x}^{\mathrm{H}}] \quad \mathbf{r} := \mathrm{E}[\mathbf{x}\bar{x}_0]$$

- This is a special case of LCMV or MVDR.

TUDelft

Direction estimation

Model: $\mathbf{x}_k = \mathbf{a}(\theta_0)s_k + \mathbf{n}_k$

Objective: estimate θ_0: *direction finding*

The classical beamformer

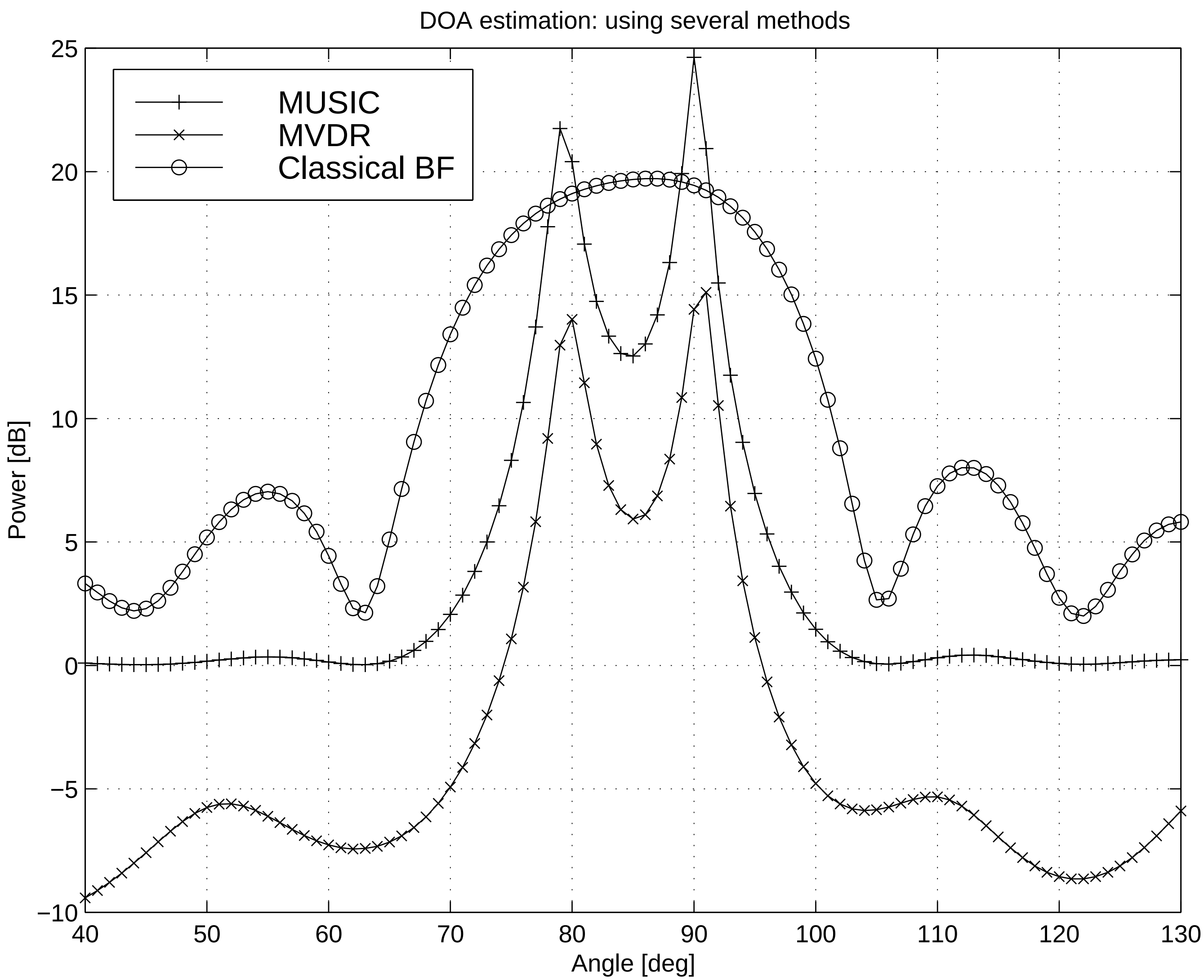
- The *classical beamformer* (Bartlett beamformer) is $\mathbf{w} = \mathbf{a}(\theta)$.
- This corresponds to the matched filter assuming spatially white noise.
- Find $\mathbf{w} = \mathbf{a}(\theta)$ that maximizes the output power

$$\hat{\theta}_0 = \max_{\theta} \frac{\mathbf{a}(\theta)^{\mathrm{H}}\mathbf{R}_x\mathbf{a}(\theta)}{\mathbf{a}(\theta)^{\mathrm{H}}\mathbf{a}(\theta)} .$$

  - For finite data, replace $\mathbf{R}_x$ by the sample covariance matrix $\hat{\mathbf{R}}_x$.
  - With known colored noise, replace denominator by $\mathbf{a}(\theta)^{\mathrm{H}}\mathbf{R}_n\mathbf{a}(\theta)$.
  - For multiple signals, choose the d largest local maxima.

    Interference and thus noise color generally not known $\Rightarrow$ *biased estimates*

Direction estimation



MVDR

- In MVDR we try to minimize the output power, while constraining the power towards the direction θ:

$$\hat{\theta}_0 = \max_{\theta}\{\min_{\mathbf{w}} \mathbf{w}^{\mathrm{H}}\hat{\mathbf{R}}_x\mathbf{w} \qquad \text{subject to} \quad \mathbf{w}^{\mathrm{H}}\mathbf{a}(\theta) = 1\}\,.$$

This yields

$$\mathbf{w} = \frac{\hat{\mathbf{R}}_x^{-1}\mathbf{a}(\theta)}{\mathbf{a}(\theta)^{\mathrm{H}}\hat{\mathbf{R}}_x^{-1}\mathbf{a}(\theta)}$$

$$\hat{\theta}_0 = \max_{\theta} \frac{1}{\mathbf{a}(\theta)^{\mathrm{H}}\hat{\mathbf{R}}_x^{-1}\mathbf{a}(\theta)}$$

- For multiple signals, choose again the d largest local maxima.

MUSIC (Multiple SIgnal Classification) algorithm

- Eigenvalue-based technique (assume $d < M$):

$$\mathbf{R}_x = \mathbf{A}\mathbf{R}_s\mathbf{A}^{\mathrm{H}} + \sigma^2\mathbf{I}_M = \mathbf{U}_s(\mathbf{\Lambda}_s + \sigma^2\mathbf{I}_d)\mathbf{U}_s^{\mathrm{H}} + \mathbf{U}_n(\sigma^2\mathbf{I}_{M-d})\mathbf{U}_n^{\mathrm{H}}$$

$$\text{span}(\mathbf{U}_s) = \text{span}(\mathbf{A}), \qquad \mathbf{U}_n^{\mathrm{H}}\mathbf{A} = 0, \qquad \text{where } \mathbf{A} = [\mathbf{a}(\theta_1), \cdots, \mathbf{a}(\theta_d)].$$

- Choose $[\theta_1, \cdots, \theta_d]$ to make **A** *fit* span($\mathbf{U}_s$):

$$\mathbf{U}_n^{\mathrm{H}}\mathbf{a}(\theta_i) = 0, \qquad (1 \le i \le d)$$

- Choose the d lowest local minima of the cost function

$$J_{MUSIC}(\theta) = \frac{\|\hat{\mathbf{U}}_n^{\mathrm{H}}\mathbf{a}(\theta)\|^2}{\|\mathbf{a}(\theta)\|^2} = \frac{\mathbf{a}(\theta)^{\mathrm{H}}\hat{\mathbf{U}}_n\hat{\mathbf{U}}_n^{\mathrm{H}}\mathbf{a}(\theta)}{\mathbf{a}(\theta)^{\mathrm{H}}\mathbf{a}(\theta)}$$

  - In a graph, we plot the inverse of $J_{MUSIC}(\theta)$.
  - If number of sources smaller than number of sensors ($d < M$), we get the *exact* DOAs for $N \to \infty$ or SNR $\to \infty \Rightarrow$ *statistically consistent* estimates.

Spatial-temporal generalizations

- Let us consider the general single-user model (see introduction)

$$\mathcal{X} = \begin{bmatrix} \mathbf{x}_0 & \mathbf{x}_1 & \cdots & \mathbf{x}_{N-1} \\ \mathbf{x}_{-1} & \mathbf{x}_0 & \cdots & \mathbf{x}_{N-2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_{-m+1} & \mathbf{x}_{-m+2} & \cdots & \mathbf{x}_{N-m} \end{bmatrix}$$

$$:= \begin{bmatrix} \boxed{\mathbf{H}} & & \mathbf{0} \\ & \boxed{\mathbf{H}} & \\ & \ddots & \\ \mathbf{0} & & \boxed{\mathbf{H}} \end{bmatrix} \begin{bmatrix} s_0 & s_1 & \cdots & s_{N-1} \\ s_{-1} & s_0 & \cdots & s_{N-2} \\ \vdots & \vdots & \ddots & \vdots \\ s_{-L-m+2} & s_{-L-m+3} & \cdots & s_{N-L-m+1} \end{bmatrix} = \mathcal{H}\mathcal{S}$$

- Adding also a white noise matrix $\mathcal{N}$, we get $\mathcal{X} = \mathcal{H}\mathcal{S} + \mathcal{N}$.

- We can also view the ISI as noise

$$\mathcal{X} = \mathbf{hs} + \mathcal{H}'\mathcal{S}' + \mathcal{N} = \mathbf{hs} + \mathcal{N}', \qquad \mathbf{h} = \mathcal{H}\mathbf{e}, \quad \mathcal{H}' = \mathcal{H}\mathbf{e}'$$

h is a specific column of $\mathcal{H}$ (**s** is the corresponding row of $\mathcal{S}$)

$\mathcal{H}'$ contains all other columns of $\mathcal{H}$ ($\mathcal{S}'$ constains the corresponding rows of $\mathcal{S}$)

Spatial-temporal generalizations

Receivers

- Matched filter:
  - ISI viewed as noise: $\mathbf{w} = \mathbf{R}_{\mathcal{N}'}^{-1}\mathbf{h} = (\mathcal{H}'\mathcal{H}'^{\mathrm{H}} + \sigma^2\mathbf{I})^{-1}\mathbf{h}$
  - ISI not viewed as noise: $\mathbf{w} = \mathcal{H}\mathbf{e} = \mathbf{h}$
  - These two matched filters are NOT THE SAME
- ZF receiver:
  - ISI viewed as noise: $\mathbf{w} = \mathbf{R}_{\mathcal{N}'}^{-1}\mathbf{h}(\mathbf{h}^{\mathrm{H}}\mathbf{R}_{\mathcal{N}'}^{-1}\mathbf{h})^{-1} \sim \mathbf{R}_{\mathcal{N}'}^{-1}\mathbf{h}$ (same as MF)
  - ISI not viewed as noise: $\mathbf{w} = \mathcal{H}(\mathcal{H}^{\mathrm{H}}\mathcal{H})^{-1}\mathbf{e} = (\mathcal{H}\mathcal{H}^{\mathrm{H}})^{\dagger}\mathcal{H}\mathbf{e} = (\mathcal{H}\mathcal{H}^{\mathrm{H}})^{\dagger}\mathbf{h}$
  - These two ZF receivers are NOT THE SAME
- Wiener filter:
  - ISI viewed as noise: $\mathbf{w} = \mathbf{R}_{\mathcal{X}}^{-1}\mathbf{r}_{\mathcal{X}s} = (\mathbf{h}\mathbf{h}^{\mathrm{H}} + \mathbf{R}_{\mathcal{N}'})^{-1}\mathbf{h} \sim \mathbf{R}_{\mathcal{N}'}^{-1}\mathbf{h}$ (same as MF)
  - ISI not viewed as noise: $\mathbf{w} = \mathbf{R}_{\mathcal{X}}^{-1}\mathbf{R}_{\mathcal{X}\mathcal{S}}\mathbf{e} = (\mathcal{H}\mathcal{H}^{\mathrm{H}} + \sigma^2\mathbf{I})^{-1}\mathcal{H}\mathbf{e} = (\mathcal{H}\mathcal{H}^{\mathrm{H}} + \sigma^2\mathbf{I})^{-1}\mathbf{h}$
  - These two Wiener receivers are THE SAME

Spatial-temporal generalizations

Joint Angle-Delay estimation

- Suppose only a single ray is present: $\mathbf{h}(t) = \mathbf{a}(\theta)\,\beta\,g(t-\tau)$.
- This means that $\mathcal{H}$ and $\mathbf{h}$ depend on θ and τ.
- Conventional: scan the output of $\mathbf{w} = \mathbf{h}(\theta,\tau)$ and maximize the power
- MUSIC algorithm:

$$\mathbf{R}_{\mathcal{X}} = \mathbf{U}_s \mathbf{\Lambda}_s \mathbf{U}_s^{\mathrm{H}} \qquad \text{span}\{\mathbf{U}_s\} = \text{span}\{\mathcal{H}(\theta,\tau)\}$$

$\mathbf{h}(\theta,\tau)$ is in the span of $\mathbf{U}_s$, therefore,

$$\mathbf{h}(\theta,\tau) \perp \mathbf{U}_n \equiv (\mathbf{U}_s)^{\perp}$$

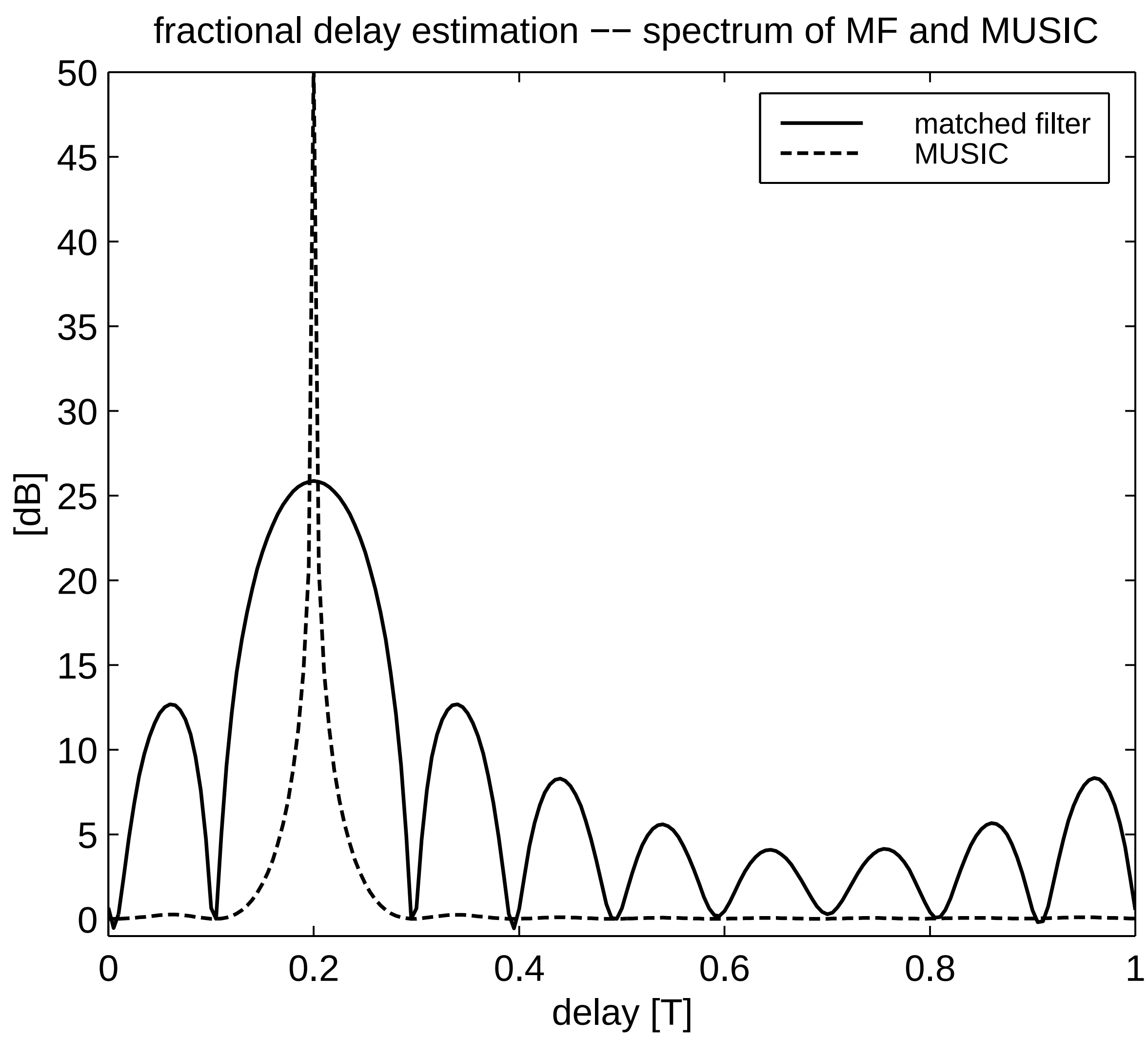
Thus, the MUSIC cost

$$J_{MUSIC}(\theta,\tau) = \frac{\mathbf{h}(\theta,\tau)^{\mathrm{H}} \hat{\mathbf{U}}_n \hat{\mathbf{U}}_n^{\mathrm{H}} \mathbf{h}(\theta,\tau)}{\mathbf{h}(\theta,\tau)^{\mathrm{H}} \mathbf{h}(\theta,\tau)}$$

will be exactly zero when θ and τ match the true values.

Spatial-temporal generalizations

Delay estimation



TU Delft

Spatial-temporal generalizations

Channel estimation

- We slightly change our single-user data model to

$$\begin{bmatrix} \mathbf{x}_{N-1} \\ \vdots \\ \mathbf{x}_0 \end{bmatrix} = \begin{bmatrix} \mathbf{h}_0 & \cdots & \mathbf{h}_{L-1} & & \mathbf{0} \\ & \ddots & & \ddots & \\ \mathbf{0} & & \mathbf{h}_0 & \cdots & \mathbf{h}_{L-1} \end{bmatrix} \begin{bmatrix} s_{N-1} \\ \vdots \\ s_{-L+1} \end{bmatrix}$$

- Using the commutativity of the convolution, this can be rewritten as

$$\begin{bmatrix} \mathbf{x}_{N-1} \\ \vdots \\ \mathbf{x}_0 \end{bmatrix} = \begin{bmatrix} s_{N-1}\mathbf{I} & \cdots & s_{N-L}\mathbf{I} \\ \vdots & & \vdots \\ s_0\mathbf{I} & \cdots & s_{-L+1}\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{h}_0 \\ \vdots \\ \mathbf{h}_{L-1} \end{bmatrix} \quad \Leftrightarrow \quad \mathbf{x} = (\mathbf{S} \otimes \mathbf{I})\mathbf{h}$$

where

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_{N-1} \\ \vdots \\ \mathbf{x}_0 \end{bmatrix} \qquad \mathbf{S} = \begin{bmatrix} s_{N-1} & \cdots & s_{N-L} \\ \vdots & & \vdots \\ s_0 & \cdots & s_{-L+1} \end{bmatrix} \qquad \mathbf{h} = \begin{bmatrix} \mathbf{h}_0 \\ \vdots \\ \mathbf{h}_{L-1} \end{bmatrix}$$

- If $\mathbf{S}$ is known, we can estimate $\mathbf{h}$ as $\hat{\mathbf{h}} = (\mathbf{S}^\dagger \otimes \mathbf{I})\mathbf{x}$

Spatial-temporal generalizations

Channel estimation

- For multiple users, we obtain

$$\mathbf{x} = (\mathbf{S}^{(1)} \otimes \mathbf{I})\mathbf{h}^{(1)} + \cdots + (\mathbf{S}^{(d)} \otimes \mathbf{I})\mathbf{h}^{(d)} \quad (1)$$

$$= \left(\begin{bmatrix} \mathbf{S}^{(1)} & \cdots & \mathbf{S}^{(d)} \end{bmatrix} \otimes \mathbf{I} \right) \begin{bmatrix} \mathbf{h}^{(1)} \\ \vdots \\ \mathbf{h}^{(d)} \end{bmatrix} \quad (2)$$

- So if the pilot symbols from all users, i.e., $\mathbf{S}^{(1)}, \ldots, \mathbf{S}^{(d)}$, are known, we can estimate all the channels jointly as

$$\begin{bmatrix} \hat{\mathbf{h}}^{(1)} \\ \vdots \\ \hat{\mathbf{h}}^{(d)} \end{bmatrix} = \left(\begin{bmatrix} \mathbf{S}^{(1)} & \cdots & \mathbf{S}^{(d)} \end{bmatrix}^{\dagger} \otimes \mathbf{I} \right) \mathbf{x}$$