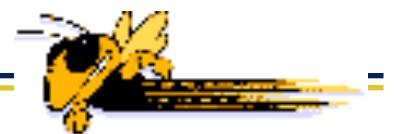

“Deterministic Signal” Gaussian Model

**ECE 6279: Spatial Array Processing
Fall 2013
Lecture 21**

Prof. Aaron D. Lanterman

**School of Electrical & Computer Engineering
Georgia Institute of Technology**

AL: 404-385-2548
[<lanterma@ece.gatech.edu>](mailto:lanterma@ece.gatech.edu)



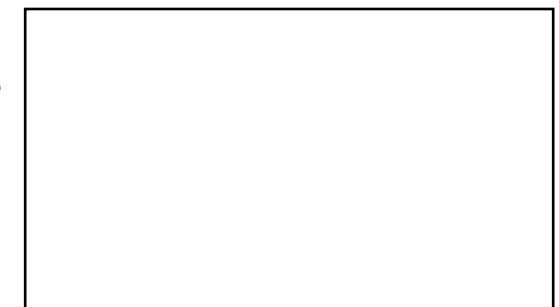
A Deterministic Gaussian Model

- N_s sources in additive noise

$$\underline{\mathbf{y}}(l) = \sum_{n=1}^{N_s} \mathbf{e}(\theta_n) s_n(l) + \underline{\mathbf{n}}(l)$$
$$\underline{\mathbf{n}} \sim \mathcal{CN}(0, \mathbf{K}_n)$$

$$\mathbf{s}(l) = \begin{bmatrix} s_1(l) \\ \vdots \\ s_{N_s}(l) \end{bmatrix}$$

Modeled as
deterministic
parameters we
need to estimate



Joint Estimation Problem

$$\mathbf{y}(l) = \mathbf{D}(\Theta)\mathbf{s}(l) + \mathbf{n}(l)$$

$$\mathbf{D}(\Theta) = [e(\theta_1) \cdots e(\theta_{N_s})]$$

$$\Theta = [\theta_1 \cdots \theta_{N_s}]$$

- Must jointly estimate
 $\mathbf{s}(l)$ and Θ



Probability Density of the Data

$$\underline{\mathbf{y}}(l) \sim \mathcal{CN}(\mathbf{D}(\Theta)\mathbf{s}(l), \mathbf{K}_n)$$

- **Circular (Goodman) Gaussian density for L independent (but not identically distributed!) snapshots:**

$$p(\mathbf{y}) = (\pi^M \det \mathbf{K}_n)^{-L} \times \\ \exp\left\{-\sum_{l=0}^{L-1} [\mathbf{y}(l) - \mathbf{D}(\Theta)\mathbf{s}_n(l)]^H \times \right. \\ \left. \mathbf{K}_n^{-1} [\mathbf{y}(l) - \mathbf{D}(\Theta)\mathbf{s}(l)]\right\}$$



Loglikelihood of the Data

- Loglikelihood is:

$$-\sum_{l=0}^{L-1} [\mathbf{y}(l) - \mathbf{D}(\Theta)\mathbf{s}(l)]^H \times \mathbf{K}_n^{-1} [\mathbf{y}(l) - \mathbf{D}(\Theta)\mathbf{s}(l)]$$

- Luckily, it so happens that if we fix a particular Θ , we can find a closed form solution for the maximizing $\mathbf{s}(l)$

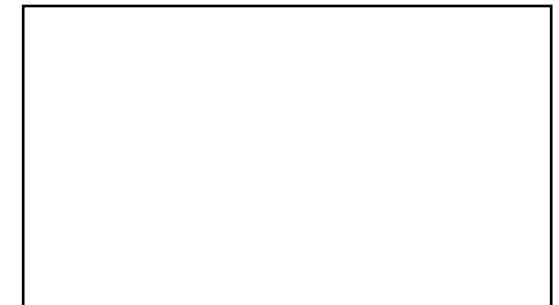


Separation in Time

$$-\sum_{l=0}^{L-1} [\mathbf{y}(l) - \mathbf{D}(\Theta)\mathbf{s}(l)]^H \times \\ \mathbf{K}_n^{-1} [\mathbf{y}(l) - \mathbf{D}(\Theta)\mathbf{s}(l)]$$

- To maximize the total loglikelihood, here it suffices to maximize each term independently

$$-[\mathbf{y}(l) - \mathbf{D}(\Theta)\mathbf{s}(l)]^H \times \\ \mathbf{K}_n^{-1} [\mathbf{y}(l) - \mathbf{D}(\Theta)\mathbf{s}(l)]$$



Setting Gradient Equal to Zero

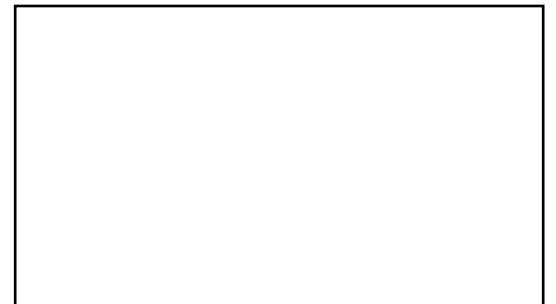
- Dropping explicit notation of l and Θ to keep notation compact

$$\begin{aligned}\ell &= -[\mathbf{y} - \mathbf{Ds}]^H \mathbf{K}_n^{-1} [\mathbf{y} - \mathbf{Ds}] \\ &= -[\mathbf{y}^H - \mathbf{s}^H \mathbf{D}^H] \mathbf{K}_n^{-1} [\mathbf{y} - \mathbf{Ds}]\end{aligned}$$

$$\nabla_{\mathbf{s}^*} \ell = -\mathbf{D}^H \mathbf{K}_n^{-1} [\mathbf{y} - \mathbf{Ds}] = 0$$

$$\mathbf{D}^H \mathbf{K}_n^{-1} \mathbf{y} = \mathbf{D}^H \mathbf{K}_n^{-1} \mathbf{Ds}$$

$$(\mathbf{D}^H \mathbf{K}_n^{-1} \mathbf{D})^{-1} \mathbf{D}^H \mathbf{K}_n^{-1} \mathbf{y} = \hat{\mathbf{s}}$$



Plugging In

$$-\sum_{l=0}^{L-1} [\mathbf{y}(l) - \mathbf{D}(\Theta)\mathbf{s}(l)]^H \times \\ \boxed{\mathbf{K}_n^{-1} [\mathbf{y}(l) - \mathbf{D}(\Theta)\mathbf{s}(l)]}$$

- Can substitute estimate of signal into likelihood

$$\hat{\mathbf{s}} = (\mathbf{D}^H \mathbf{K}_n^{-1} \mathbf{D})^{-1} \mathbf{D}^H \mathbf{K}_n^{-1} \mathbf{y}$$

$$\mathbf{y} - \mathbf{D}\hat{\mathbf{s}} = \mathbf{y} - \mathbf{D}(\mathbf{D}^H \mathbf{K}_n^{-1} \mathbf{D})^{-1} \mathbf{D}^H \mathbf{K}_n^{-1} \mathbf{y}$$



A Simplifying Supposition

Suppose $\mathbf{K}_n = \sigma_n^2 \mathbf{I}$

$$-\sum_{l=0}^{L-1} [\mathbf{y}(l) - \mathbf{D}(\Theta)\mathbf{s}(l)]^H \times [\mathbf{y}(l) - \mathbf{D}(\Theta)\mathbf{s}(l)]$$

- Can substitute estimate of signal into likelihood

$$\hat{\mathbf{s}} = (\mathbf{D}^H \mathbf{D})^{-1} \mathbf{D}^H \mathbf{y}$$

$$\mathbf{y} - \mathbf{D}\hat{\mathbf{s}} = \mathbf{y} - \underbrace{\mathbf{D}(\mathbf{D}^H \mathbf{D})^{-1} \mathbf{D}^H}_{\equiv \mathbf{P}_D} \mathbf{y}$$



Our Projection Matrix

$$\mathbf{P}_D = \mathbf{D}(\mathbf{D}^H \mathbf{D})^{-1} \mathbf{D}^H$$

$$\mathbf{P}_D \mathbf{P}_D = \mathbf{D} \cancel{(\mathbf{D}^H \mathbf{D})^{-1}} \cancel{\mathbf{D}^H} \mathbf{D} \times \\ (\mathbf{D}^H \mathbf{D})^{-1} \mathbf{D}^H$$

$$\mathbf{P}_D \mathbf{P}_D = \mathbf{D}(\mathbf{D}^H \mathbf{D})^{-1} \mathbf{D}^H = \mathbf{P}_D$$

$\mathbf{P}_D \mathbf{P}_D = \mathbf{P}_D$ means that \mathbf{P}_D is a
Projection Matrix (all
eigenvalues are either zero or
one)



“Signal Subspace” Interpretation

$$\mathbf{P}_{\mathbf{D}} \mathbf{y} = \mathbf{D}(\mathbf{D}^H \mathbf{D})^{-1} \mathbf{D}^H \mathbf{y}$$

**This projection matrix projects
the data onto the subspace
spanned by $\mathbf{D}(\Theta)$**



The “Not Signal” Subspace

$$\mathbf{y} - \hat{\mathbf{D}}\hat{\mathbf{s}} = \mathbf{y} - \mathbf{P}_D\mathbf{y} = (\mathbf{I} - \mathbf{P}_D)\mathbf{y}$$

- Define $\mathbf{P}_D^\perp = \mathbf{I} - \mathbf{P}_D$
- It's easy to show \mathbf{P}_D^\perp is also a projection matrix

$$\mathbf{y} - \hat{\mathbf{D}}\hat{\mathbf{s}} = \mathbf{P}_D^\perp\mathbf{y}$$



Minimize Proj. onto “Not Signal” Space

- Maximize loglikelihood over Θ

$$-\sum_{l=0}^{L-1} [(\mathbf{I} - \mathbf{P}_D) \mathbf{y}(l)]^H [(\mathbf{I} - \mathbf{P}_D) \mathbf{y}(l)]$$

$$= -\sum_{l=0}^{L-1} \left\| \mathbf{P}_D^\perp \mathbf{y}(l) \right\|^2$$

- Equivalently minimize

$$\sum_{l=0}^{L-1} \left\| \mathbf{P}_D^\perp \mathbf{y}(l) \right\|^2 \text{ over } \Theta$$



Rewriting

$$\ell = - \sum_{l=0}^{L-1} \left\| \mathbf{P}_D^\perp \mathbf{y}(l) \right\|^2$$

$$= - \sum_{l=0}^{L-1} [\mathbf{P}_D^\perp \mathbf{y}(l)]^H [\mathbf{P}_D^\perp \mathbf{y}(l)]$$

$$= - \sum_{l=0}^{L-1} \mathbf{y}^H(l) \mathbf{P}_D^\perp \mathbf{y}(l)$$



Maximize Proj. on “Signal” space

$$\ell = - \sum_{l=0}^{L-1} \mathbf{y}^H(l) \mathbf{P}_D^\perp \mathbf{y}(l)$$

$$= - \sum_{l=0}^{L-1} \mathbf{y}^H(l) (\mathbf{I} - \mathbf{P}_D) \mathbf{y}(l)$$

$$\begin{aligned}\ell' &= \sum_{l=0}^{L-1} \mathbf{y}^H(l) \mathbf{P}_D \mathbf{y}(l) \\ &= \sum_{l=0}^{L-1} \|\mathbf{P}_D \mathbf{y}(l)\|^2\end{aligned}$$



In Terms of Empirical Correlation Matrix

$$\begin{aligned}\ell' &= \sum_{l=0}^{L-1} \mathbf{y}^H(l) \mathbf{P}_D \mathbf{y}(l) \\ &= \text{tr} \left\{ \sum_{l=0}^{L-1} \mathbf{y}^H(l) \mathbf{P}_D \mathbf{y}(l) \right\} \\ &= L \text{tr} \left\{ \mathbf{P}_D \underbrace{\frac{1}{L} \sum_{l=0}^{L-1} \mathbf{y}(l) \mathbf{y}^H(l)}_{\hat{\mathbf{R}}_y} \right\} \\ &= \text{tr} \left\{ \mathbf{P}_D \hat{\mathbf{R}}_y \right\}\end{aligned}$$



Practical ML Procedure

Maximize

$$\text{tr} \left\{ \mathbf{P}_{\mathbf{D}}(\Theta) \hat{\mathbf{R}}_y \right\} = \text{tr} \left\{ \mathbf{D} (\mathbf{D}^H \mathbf{D})^{-1} \mathbf{D}^H \hat{\mathbf{R}}_y \right\}$$

over Θ



How Can We Maximize Over Angles?

- No closed form solution in general
- Still an open research problem!
- One approach: try an **expectation-maximization** algorithm (sometimes ECE7251 covers EM algorithms)

M.I. Miller and D.R. Fuhrmann,
“Maximum-Likelihood Narrow-Band
Direction Finding and the EM
Algorithm,” IEEE Trans. ASSP, 38(9),
Sept. 1990, pp. 1560-1577.



The Single-Source Case

- Recall ML estimate of signal for a fixed Θ

$$\hat{s} = \left[D^H(\Theta) K_n^{-1} D(\Theta) \right]^{-1} D(\Theta)^H K_n^{-1} y$$

- Suppose we only have one source

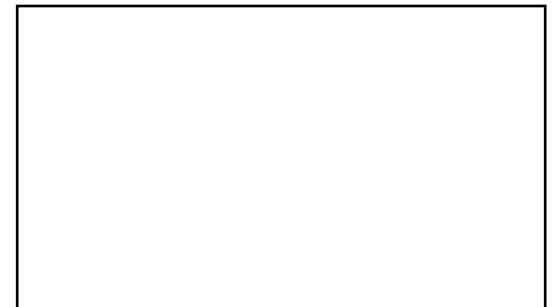
$$\hat{s} = \left[e^H(\theta) K_n^{-1} e(\theta) \right]^{-1} e(\theta)^H K_n^{-1} y$$



A Blast from the Past

$$\hat{S} = \frac{\mathbf{e}^H(\theta) \mathbf{K}_n^{-1}}{\underbrace{\mathbf{e}^H(\theta) \mathbf{K}_n^{-1} \mathbf{e}(\theta)}_{\equiv \mathbf{a}^H}} \mathbf{y} = \mathbf{a}^H \mathbf{y}$$

- Recall $\mathbf{a}^H \propto \mathbf{e}^H(\theta) \mathbf{K}_n^{-1}$ are the weights that maximize SNR
 - Showed this in earlier lecture



Another Blast from the Past

- If $\mathbf{K}_n = \sigma_n^2 \mathbf{I}$ then

$$\hat{s} = \frac{\mathbf{e}^H(\theta) \mathbf{K}_n^{-1}}{\mathbf{e}^H(\theta) \mathbf{K}_n^{-1} \mathbf{e}(\theta)} \mathbf{y} = \frac{\mathbf{e}^H(\theta)}{\mathbf{e}^H(\theta) \mathbf{e}(\theta)} \mathbf{y}$$
$$= \frac{\mathbf{e}^H(\theta)}{M} \mathbf{y}$$



- Old delay-and-sum conventional beamformer



Maximum-Likelihood Procedure

- ML procedure said to maximize

$$\text{tr} \left\{ \mathbf{D} (\mathbf{D}^H \mathbf{D})^{-1} \mathbf{D}^H \hat{\mathbf{R}}_y \right\} \text{ over } \Theta$$

- For one source

$$\begin{aligned} \text{tr} \left\{ \mathbf{e} (\mathbf{e}^H \mathbf{e})^{-1} \mathbf{e}^H \hat{\mathbf{R}}_y \right\} &= \frac{1}{M} \text{tr} \left\{ \mathbf{e} \mathbf{e}^H \hat{\mathbf{R}}_y \right\} \\ &= \frac{1}{M} \text{tr} \left\{ \mathbf{e}^H \hat{\mathbf{R}}_y \mathbf{e} \right\} = \underbrace{\mathbf{e}^H(\theta) \hat{\mathbf{R}}_y \mathbf{e}(\theta)}_{M} / M \end{aligned}$$

Power out of our old conventional
delay-and-sum beamformer!



Take-Home Messages (1)

- If you have a **single source** in **Gaussian noise**, maximizing the SNR turns out to be optimal (in the ML estimation sense) under the deterministic Gaussian model
- If you have non-Gaussian noise or multiple sources, maximizing SNR is **not** the optimal ML procedure
 - But usually still pretty useful!



Take-Home Message (2)

- If you have a **single source** in white Gaussian noise, the conventional delay-and-sum beamformer is optimal in the ML sense under the deterministic Gaussian model
- If you have non-Gaussian noise or multiple sources the conventional beamformer is not the optimal ML procedure
 - But usually still pretty useful!



Capon's Argument (1)

- ML signal estimate for one source is

$$\hat{s} = \frac{e^H(\theta)K_n^{-1}}{e^H(\theta)K_n^{-1}e(\theta)} y$$

- In Capon's application (see p. 358-359 of J&D), he didn't have a good estimate of K_n
- Capon handwaved and said to try replacing K_n with \hat{R}_y



Capon's Argument (2)

- Totally ad-hoc from an ML standpoint; $\hat{\mathbf{R}}_y$ is a terrible estimate of \mathbf{K}_n
- But if we go ahead and do it, we get

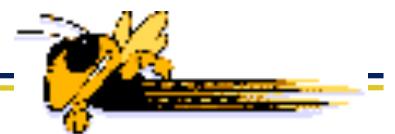
$$\hat{s} = \frac{\mathbf{e}^H(\theta) \hat{\mathbf{R}}_y^{-1}}{\mathbf{e}^H(\theta) \hat{\mathbf{R}}_y^{-1} \mathbf{e}(\theta)} \mathbf{y}$$

which is the MVDR beamformer!



A Great Confusion

- MVDR (a.k.a. Capon's method) is sometimes erroneously referred to as “maximum likelihood”
 - That's nonsense! MVDR does not arise from any known statistical ML procedure
- Our earlier derivation of MVDR from a constrained optimization procedure is due to Lacoss (1971)



Recollections and Observations

- Remember MVDR was a solution to

$$\mathbf{w}^H \mathbf{R}_y \mathbf{w} \text{ s.t. } \mathbf{w}^H \mathbf{e}(\theta) = 1$$

- Yes, this is an optimization problem; but it's not a maximum-likelihood optimization problem
 - Doesn't require a probability model for the data

