“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>

A Deterministic Gaussian Model

- N_s **sources in additive noise**

$$\underline{\mathbf{y}}(l) = \sum_{n=1}^{N_s} \mathbf{e}(\theta_n) s_n(l) + \underline{\mathbf{n}}(l)$$

$$\underline{\mathbf{n}} \sim CN(0, \mathbf{K}_{\mathbf{n}})$$

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

Modeled as deterministic parameters we need to estimate

Joint Estimation Problem

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

$$\mathbf{D}(\boldsymbol{\Theta}) = \begin{bmatrix} \mathbf{e}(\boldsymbol{\theta}_1) & \cdots & \mathbf{e}(\boldsymbol{\theta}_{N_s}) \end{bmatrix}$$

$$\boldsymbol{\Theta} = \begin{bmatrix} \boldsymbol{\theta}_1 & \cdots & \boldsymbol{\theta}_{N_s} \end{bmatrix}$$

- **Must jointly estimate $\mathbf{s}(l)$ and $\boldsymbol{\Theta}$**

Probability Density of the Data

$$\underline{\mathbf{y}}(l) \sim CN(\mathbf{D}(\boldsymbol{\Theta})\mathbf{s}(l), \mathbf{K}_{\mathbf{n}})$$

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

$$p(\mathbf{y}) = (\pi^M \det \mathbf{K}_{\mathbf{n}})^{-L} \times$$

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

$$\mathbf{K}_{\mathbf{n}}^{-1} [\mathbf{y}(l) - \mathbf{D}(\boldsymbol{\Theta})\mathbf{s}(l)]\}$$

Loglikelihood of the Data

- **Loglikelihood is:**

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

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

Separation in Time

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

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

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

Setting Gradient Equal to Zero

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

$$\ell = -[\mathbf{y} - \mathbf{D}\mathbf{s}]^H \mathbf{K}_\mathbf{n}^{-1}[\mathbf{y} - \mathbf{D}\mathbf{s}]$$

$$= -[\mathbf{y}^H - \mathbf{s}^H \mathbf{D}^H]\mathbf{K}_\mathbf{n}^{-1}[\mathbf{y} - \mathbf{D}\mathbf{s}]$$

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

$$\mathbf{D}^H \mathbf{K}_\mathbf{n}^{-1}\mathbf{y} = \mathbf{D}^H \mathbf{K}_\mathbf{n}^{-1}\mathbf{D}\mathbf{s}$$

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

Plugging In

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

- Can substitute estimate of signal into likelihood

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

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

A Simplifying Supposition

Suppose $\mathbf{K_n} = \sigma_n^2 \mathbf{I}$

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

- **Can substitute estimate of signal into likelihood**

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

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

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

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

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

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

Minimize Proj. onto "Not Signal" Space

- **Maximize loglikelihood over Θ**

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

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

- **Equivalently minimize**

$\sum_{l=0}^{L-1}\left\|\mathbf{P}_{\mathbf{D}}^{\perp}\mathbf{y}(l)\right\|^2$ **over Θ**

Rewriting

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

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

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

Maximize Proj. on “Signal” space

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

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

$$\ell' = \sum_{l=0}^{L-1} \mathbf{y}^H(l)\mathbf{P}_{\mathbf{D}}\mathbf{y}(l)$$

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

In Terms of Empirical Correlation Matrix

$$\ell' = \sum_{l=0}^{L-1} \mathbf{y}^H(l)\mathbf{P}_\mathbf{D}\mathbf{y}(l)$$

$$= \mathrm{tr}\left\{\sum_{l=0}^{L-1} \mathbf{y}^H(l)\mathbf{P}_\mathbf{D}\mathbf{y}(l)\right\}$$

$$= L\,\mathrm{tr}\left\{\mathbf{P}_\mathbf{D}\underbrace{\frac{1}{L}\sum_{l=0}^{L-1} \mathbf{y}(l)\mathbf{y}^H(l)}_{\equiv \hat{\mathbf{R}}_\mathbf{y}}\right\}$$

$$= \mathrm{tr}\left\{\mathbf{P}_\mathbf{D}\hat{\mathbf{R}}_\mathbf{y}\right\}$$

Practical ML Procedure

Maximize

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

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

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

- Suppose we only have one source

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

A Blast from the Past

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

- **Recall** $\mathbf{a}^H \propto \mathbf{e}^H(\theta)\mathbf{K}_\mathbf{n}^{-1}$ **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

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

- For one source

$$\mathbf{tr}\left\{\mathbf{e}(\mathbf{e}^H\mathbf{e})^{-1}\mathbf{e}^H\hat{\mathbf{R}}_\mathbf{y}\right\} = \frac{1}{M}\mathbf{tr}\left\{\mathbf{e}\mathbf{e}^H\hat{\mathbf{R}}_\mathbf{y}\right\}$$

$$= \frac{1}{M}\mathbf{tr}\left\{\mathbf{e}^H\hat{\mathbf{R}}_\mathbf{y}\mathbf{e}\right\} = \underbrace{\mathbf{e}^H(\theta)\hat{\mathbf{R}}_\mathbf{y}\mathbf{e}(\theta)}_{\text{Power out of our old conventional delay-and-sum beamformer!}}/M$$

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 <u>white Gaussian noise</u>, 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 <u>not</u> the optimal ML procedure**
  - But usually still pretty useful!

Capon's Argument (1)

- ML signal estimate for one source is

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

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

Capon's Argument (2)

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

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

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**

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

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