Special Cases of Maximum-Likelihood Estimation

ECE 6279: Spatial Array Processing
Spring 2007
Lecture 29
4/11/07

Prof. Aaron D. Lanterman

School of Electrical & Computer Engineering
Georgia Institute of Technology
AL: 404-385-2548
<lanterma@ece.gatech.edu>

Our Deterministic Gaussian Model

- ***N*** **sources in additive noise**

$$\mathbf{y}(l) = \sum_{n=1}^{N} \mathbf{e}(\boldsymbol{\theta}_n) s_n(l) + \mathbf{w}(l)$$

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

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

Modeled as deterministic parameters we need to estimate

The Single-Source Case

- **Recall ML estimate of signal for a fixed Θ**

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

- **Suppose we only have one source**

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

A Blast from the Past

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

- **Recall $\mathbf{a}^H \propto \mathbf{e}^H(\boldsymbol{\theta})\mathbf{K}_{\mathbf{w}}^{-1}$ are the weights that maximize SNR**
  - Showed this in Lecture 13

Another Blast from the Past

- If $\mathbf{K}_{\mathbf{w}} = \sigma_w^2 \mathbf{I}$ then

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

$$= \frac{\mathbf{e}^H(\theta)}{M}\mathbf{y}$$

- Old delay-and-sum conventional beamformer

Maximum-Likelihood Procedure

- ML procedure said to maximize

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

- For one source

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

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

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 <u>not</u> the optimal ML procedure
  - But usually still pretty useful!

Take-Home Message (2)

- If you have a single source in <u>white</u> 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 <u>not</u> the optimal ML procedure
  - But usually still pretty useful!

Capon's Argument (1)

- ML signal estimate for one source is
$$\hat{\mathbf{s}} = \frac{\mathbf{e}^H(\theta)\mathbf{K}_{\mathrm{w}}^{-1}}{\mathbf{e}^H(\theta)\mathbf{K}_{\mathrm{w}}^{-1}\mathbf{e}(\theta)}\mathbf{y}$$
- In Capon's application (see p. 358-359 of J&D), he didn't have a good estimate of $\mathbf{K}_{\mathrm{w}}$
- Capon handwaved and said to try replacing $\mathbf{K}_{\mathrm{w}}$ with $\hat{\mathbf{R}}_{\mathrm{y}}$

Capon's Argument (2)

- Totally ad-hoc from an ML standpoint; $\hat{\mathbf{R}}_{\mathrm{y}}$ is a terrible estimate of $\mathbf{K}_{\mathrm{w}}$
- But if we go ahead and do it, we get
$$\hat{\mathbf{s}} = \frac{\mathbf{e}^H(\theta)\hat{\mathbf{R}}_{\mathrm{y}}^{-1}}{\mathbf{e}^H(\theta)\hat{\mathbf{R}}_{\mathrm{y}}^{-1}\mathbf{e}(\theta)}\mathbf{y}$$

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
$$\mathbf{a}^H\mathbf{R}_{\mathrm{y}}\mathbf{a} \text{ s.t. } \mathbf{a}^H\mathbf{e}(\theta) = 1$$
- Yes, this is an optimization problem; but it's not a maximum-likelihood optimization problem
  - Doesn't require a probability model for the data

Our Stochastic Gaussian Model

- *N* sources in additive noise

$$\mathbf{y}(l) = \sum_{n=1}^{N} \mathbf{e}(\boldsymbol{\theta}_n) s_n(l) + \mathbf{w}(l)$$

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

$$\mathbf{s} = \begin{bmatrix} s_1 \\ \vdots \\ s_N \end{bmatrix} \sim CN(0, \mathbf{K}_{\mathbf{s}})$$

Maximum-Likelihood Estimation

- Goal: maximize

$$-L \ln\det \mathbf{K}_{\mathbf{y}}(\boldsymbol{\xi}) - L\,\mathrm{tr}\left[\hat{\mathbf{K}}_{\mathbf{y}} \mathbf{K}_{\mathbf{y}}^{-1}(\boldsymbol{\xi})\right]$$

where

$$\mathbf{K}_{\mathbf{y}}(\boldsymbol{\xi}) = \mathbf{D}(\boldsymbol{\Theta})\mathbf{K}_{\mathbf{s}}\mathbf{D}^H(\boldsymbol{\Theta}) + \mathbf{K}_{\mathbf{w}}$$

over $\boldsymbol{\xi}$, which represents all the parameters (angles, signal powers, correlation coefficients)

Asymptotic ML Est. for Stochastic Model

- If you have
  - –"Lots of snapshots," and
  - –Uncorrelated sources
  - –I.I.D. receiver noise
- Then VT-IV, pp. 985-991 shows that you can get rid of the power parameters $\sigma_1^2, \ldots \sigma_N^2$ and find ML estimates of angles by maximizing…

Asymptotic ML Procedure

- New goal: maximize

$$-\ln\det\left[\mathbf{P}_{\mathbf{D}}\hat{\mathbf{K}}_{\mathbf{y}}\mathbf{P}_{\mathbf{D}} + \sigma_w^2 \mathbf{P}_{\mathbf{D}}^{\perp}\right] - \mathrm{tr}\left[\mathbf{P}_{\mathbf{D}}^{\perp}\hat{\mathbf{K}}_{\mathbf{y}}\right] / \sigma_w^2$$

over $\boldsymbol{\Theta}$

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

$$\mathbf{P}_{\mathbf{D}}^{\perp} = \mathbf{I} - \mathbf{P}_{\mathbf{D}}$$