Exercise session 5 (tips on HW 2)
Bayesian estimation, array signal processing, and Kalman filtering

ELEC-E5440 Statistical Signal Processing

Robin Rajamäki

November 25, 2021

1 Bayesian estimation

Three ingredients of Bayesian estimation (x is unknown parameter, y is data):

- Bayes rule:

$$p(x; y) = p(x|y)p(y) = p(y|x)p(x)$$

- Minimum mean squared error (MMSE) estimator:

$$\hat{x}_{\text{MS}} \triangleq \arg\min_{\hat{x}} \mathbb{E}(|x - \hat{x}(y)|^2) = \mathbb{E}(x|y) = \int xp(x|y)dx$$

- Maximum a posteriori (MAP) estimator:

$$\hat{x}_{\text{MAP}} \triangleq \arg\max_{x} p(x|y) = \arg\max_{x} \frac{p(y|x)p(x)}{p(y)} = \arg\max_{x} p(y|x)p(x).$$

For details and examples complementary to the lecture handouts, see

- S. M. Kay, *Fundamentals of statistical signal processing: Estimation theory.* Prentice Hall PTR, 1993 (check library or try googling...)

2 Sensor array processing

Consider an M sensor linear array observing a scalar wavefield produced by K uncorrelated far field point sources, as illustrated in Fig. 1. A snapshot in time of the received signal $\boldsymbol{x} \in \mathbb{C}^M$ is

$$\boldsymbol{x}(n) = \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{s}(n) + \boldsymbol{v}(n), \tag{1}$$

where $\boldsymbol{A} \in \mathbb{C}^{M \times K}$ is the array steering matrix whose columns are parametrized by the (distinct) source angles in vector $\boldsymbol{\theta} \in \mathbb{R}^K$. In the case of the Uniform

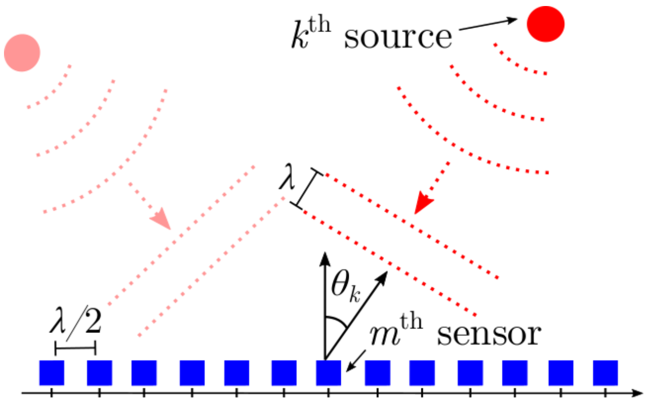


Figure 1: Canonical array processing model.

Linear Array (ULA) with omnidirectional sensors and a half wavelength inter-sensor spacing (see Fig. 1), the (m,k)th entry of $\boldsymbol{A}$ assumes the form

$$A_{m,k} = \exp(j\pi(m-1)\sin\theta_k).$$

Furthermore, $\boldsymbol{s} \in \mathbb{C}^K$ is the source signal vector and $\boldsymbol{v} \in \mathbb{C}^M$ a noise vector, both typically assumed circularly symmetric complex Gaussian random vectors. Under standard assumptions (see lecture handouts p. 401), the covariance matrix of the measurements becomes

$$\boldsymbol{\Sigma} \triangleq \mathbb{E}(\boldsymbol{x}\boldsymbol{x}^{\mathrm{H}}) = \boldsymbol{A}\boldsymbol{P}\boldsymbol{A}^{\mathrm{H}} + \sigma^2\boldsymbol{I}, \tag{2}$$

where $\boldsymbol{P} \triangleq \mathbb{E}(\boldsymbol{s}\boldsymbol{s}^{\mathrm{H}}) = \mathrm{diag}([p_1, p_2, \ldots, p_K])$ is the diagonal source power matrix with p_k denoting the power of the kth source, and σ^2 is the noise variance. In practice, we compute a finite sample estimate of the covariance matrix:

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}(n)\boldsymbol{x}^{\mathrm{H}}(n),$$

where N denotes the number of snapshots (time samples). Note that $N \geq M$ is necessary for $\hat{\boldsymbol{\Sigma}}$ to be full rank (this is crucial when $\hat{\boldsymbol{\Sigma}}$ needs to be inverted).

A typical task of array processing is to estimate quantities of interest from (1), such as the source angles $\boldsymbol{\theta}$, signals waveforms $\boldsymbol{s}$, or signal covariance $\boldsymbol{P}$. We briefly consider these three cases next.

2.1 Source signal estimation

The source signals $\boldsymbol{s}$ can be estimated using the MMSE estimator:

$$\hat{\boldsymbol{s}}_{\mathrm{MS}} = \boldsymbol{P}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{\Sigma}^{-1}\boldsymbol{x}.$$

Note that $\boldsymbol{P}$ and $\boldsymbol{A}(\boldsymbol{\theta})$ are typically unknown and need to be estimated. The estimation of the source angles $\boldsymbol{\theta}$ is known as *direction-of-arrival* (DoA) estimation.

2.2 DoA estimation

2.2.1 Beamforming

Beamforming is a simple and versatile nonparametric approach for spatial filtering, which can be employed at both the transmitter and the receiver. For example, consider the received power of the array steered in direction ϕ after applying beamforming weight vector $\boldsymbol{w} \in \mathbb{C}^M$:

$$b(\phi) \triangleq \mathbb{E}(|\boldsymbol{w}^{\mathrm{H}}(\phi)\boldsymbol{x}|^2) = \boldsymbol{w}^{\mathrm{H}}(\phi)\boldsymbol{\Sigma}\boldsymbol{w}(\phi) \geq 0.$$

The choice of $\boldsymbol{w}$ affects the shape of the beampattern by trading off main lobe width (resolution) and side lobe levels (interference suppression capability), as demonstrated in Fig. 2. A common choice is the *spatial matched filter*, for which

$$\boldsymbol{w}(\phi) = \mathbf{a}(\phi)/\|\mathbf{a}(\phi)\|_2^2.$$

The DoA estimates are given by the peaks of $b(\phi)$. For simplicity, assume that $\|\boldsymbol{a}(\phi)\|_2^2 = 1 \ \forall\phi$ (e.g., omnidirectional sensors). For a single source with power p in direction θ, the matched beamformer output power is thus

$$b(\phi) = \mathbf{a}^{\mathrm{H}}(\phi)\boldsymbol{\Sigma}\mathbf{a}(\phi) = \mathbf{a}^{\mathrm{H}}(\phi)(p\mathbf{a}(\theta)\mathbf{a}^{\mathrm{H}}(\theta) + \sigma^2\boldsymbol{I})\mathbf{a}(\phi) = p|\mathbf{a}^{\mathrm{H}}(\phi)\mathbf{a}(\theta)|^2 + \sigma^2.$$

The Cauchy-Schwartz inequality therefore implies that

$$|\mathbf{a}^{\mathrm{H}}(\phi)\mathbf{a}(\theta)|^2 \leq 1,$$

where equality holds if and only if $\mathbf{a}(\phi)$ and $\mathbf{a}(\theta)$ are linearly dependent, i.e., $\mathbf{a}(\phi) = e^{j\varphi}\mathbf{a}(\theta)$ for some $\varphi \in \mathbb{R}$. Suppose that $\mathbf{a}(\phi)$ and $\mathbf{a}(\theta)$ are linearly independent if $\phi \neq \theta$. For example, the ULA with omnidirectional sensors and an inter-sensor spacing of half a wavelength or less satisfies this property. In this case, the beamformer power is maximized by the true DoA, i.e.,

$$\theta = \arg\max_{\phi} b(\phi).$$

For a finite sample realization of b, it is unknown whether the angle corresponding to the peak of the beamformer power spectrum yields an unbiased estimate of θ. In the case of $K > 1$ sources, the peaks do not generally yield unbiased estimates of the source DoAs, even when the number of snapshots approaches infinity. The modest statistical performance of beamforming, as well as its limited resolution, motivates developing alternative DoA estimators.

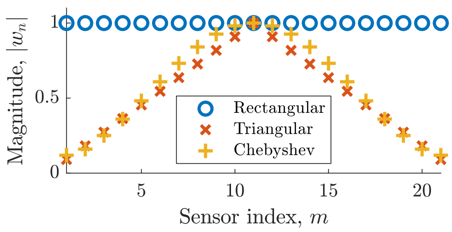
2.2.2 MVDR

In contrast to the data-independent classical beamformer, the minimum variance distortionless response (MVDR) adapts its weight vector based on the measurements. In particular, the MVDR minimizes the output power while maintaining a distortionless response in the direction of interest ϕ:

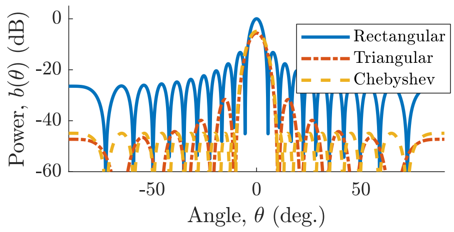
$$\underset{\boldsymbol{w}\in\mathbb{C}^M}{\text{minimize}}\ \boldsymbol{w}^{\mathrm{H}}\boldsymbol{\Sigma}\boldsymbol{w} \text{ subject to } \boldsymbol{w}^{\mathrm{H}}\boldsymbol{a}(\phi) = 1$$

The solution of this optimization problem has a closed-form expression that depends on the (inverse of the) data covariance matrix $\boldsymbol{\Sigma}$ as follows:

$$\boldsymbol{w}_{\mathrm{MVDR}}(\phi) = \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{a}(\phi)}{\boldsymbol{a}^{\mathrm{H}}(\phi)\boldsymbol{\Sigma}^{-1}\boldsymbol{a}(\phi)}.$$



(a) Beamforming weights



(b) Beamformer output

Figure 2: Different beamforming weight choices trade off between a narrow main lobe and low side-lobe levels.

The MVDR automatically supresses interference, such as other source signals, in directions $\theta \neq \phi$. By adapting to the environment, the MVDR achieves better resolution and interference suppression capability than the classical beamformer. Similarly to the classical beamformer, the DoAs are found from the peaks of the spatial power spectrum, which in case of the MVDR reduces to

$$b(\phi) = \frac{1}{\boldsymbol{a}^{\mathrm{H}}(\phi)\boldsymbol{\Sigma}^{-1}\boldsymbol{a}(\phi)}.$$

The previous derivation assumes that $\boldsymbol{\Sigma}$ is known exactly, which is equivalent to assuming that the number of snapshots $N \to \infty$. For a finite N, spurious peaks may occur in the power spectrum since a finite-sample estimate of $\boldsymbol{\Sigma}$ has to be employed.

2.2.3 MUSIC

MUSIC (MUltiple SIgnal Classification) is a widely used parametric line spectrum estimation algorithm applicable to DoA estimation. The main advantage of MUSIC is its ability to surpass the resolution limit of beamforming. The main steps of the MUSIC algorithm can be summarized as follows:

Step I: Compute (estimate) covariance matrix $\boldsymbol{\Sigma}$

Step II: Evaluate eigenvalue decomposition of $\boldsymbol{\Sigma}$

$$\boldsymbol{\Sigma} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{H}} = \begin{bmatrix} \boldsymbol{U}_{\mathrm{s}} & \boldsymbol{U}_{\mathrm{n}} \end{bmatrix} \begin{bmatrix} \boldsymbol{L} + \sigma^2 \boldsymbol{I}_K & \boldsymbol{0} \\ \boldsymbol{0} & \sigma^2 \boldsymbol{I}_{M-K} \end{bmatrix} \begin{bmatrix} \boldsymbol{U}_{\mathrm{s}}^{\mathrm{H}} \\ \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}} \end{bmatrix}$$

Step III: Find peaks of pseudospectrum

$$\tilde{b}(\phi) \triangleq \frac{1}{\boldsymbol{a}^{\mathrm{H}}(\phi)\boldsymbol{U}_{\mathrm{n}}\boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\boldsymbol{a}(\phi)}.$$

Here $\boldsymbol{L} = \mathrm{diag}([l_1, l_2, \ldots, l_K])$ is a diagonal matrix containing the nonzero eigenvalues of the signal component of $\boldsymbol{\Sigma}$, i.e.,

$$\boldsymbol{A}\boldsymbol{P}\boldsymbol{A}^{\mathrm{H}} = \boldsymbol{U}_{\mathrm{s}}\boldsymbol{L}\boldsymbol{U}_{\mathrm{s}}^{\mathrm{H}}.$$

Since $\boldsymbol{APA}^{\mathrm{H}}$ is positive semi-definite, we have $l_k > 0$. Sorting the eigenvalues of $\boldsymbol{\Sigma}$ is descending order,

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_K \geq \ldots \geq \lambda_M,$$

we have

$$\lambda_m = \begin{cases} l_m + \sigma^2, & \text{if } 1 \leq m \leq K \\ \sigma^2, & \text{otherwise.} \end{cases}$$

Consequently, the K largest eigenvalues are associated with the signal subspace spanned by the range space of $\boldsymbol{A}$ (or columns of $\boldsymbol{U}_{\mathrm{s}}$), with the remaining $M-K$ eigenvalues associated with the noise subspace, i.e., the orthogonal complement of the range space of $\boldsymbol{A}$ (or columns of $\boldsymbol{U}_{\mathrm{n}}$). The denominator of the MUSIC pseudospectrum is therefore zero if the steering angle equals a true source direction, i.e., $\phi \in \{\theta_k\}_{k=1}^K$. Most notably, the zeros are unique for the previously discussed ULA. Uniqueness means that the peaks in the MUSIC pseudospectrum correspond to only the true DoAs.

In practice, the eigenvalues corresponding to the noise subspace may be difficult to distinguish from those of weak sources, since a finite-sample estimate of $\boldsymbol{\Sigma}$ has to be used.

2.3 Source covariance matrix estimation

Estimating the source covariance matrix $\boldsymbol{P}$ is straightforward once the DoAs have been estimated. The maximum likelihood estimator (MLE) of $\boldsymbol{P}$ is [2]

$$\hat{\boldsymbol{P}} = \boldsymbol{A}^\dagger(\boldsymbol{\Sigma} - \sigma^2\boldsymbol{I})(\boldsymbol{A}^\dagger)^{\mathrm{H}},$$

where $\boldsymbol{A}^\dagger = (\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{H}}$ is the pseudoinverse of $\boldsymbol{A}$. Recall that $\boldsymbol{A}$ is a function of the source angles $\boldsymbol{\theta}$. The MLE of the noise variance σ^2 can in turn be computed using the noise subspace of $\boldsymbol{\Sigma}$ (see Section 2.2.3) as [2]

$$\hat{\sigma}^2 = \frac{1}{M-K}\mathrm{tr}(\boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\boldsymbol{\Sigma}\boldsymbol{U}_{\mathrm{n}}),$$

where $\mathrm{tr}(\cdot)$ is the trace operator (sum of diagonal elements). In practice, quantities $\boldsymbol{\Sigma}$, $\boldsymbol{U}_{\mathrm{n}}$, σ^2 and $\boldsymbol{\theta}$ above should be replaced by their finite sample estimates.

2.4 Source number estimation

The number of signals K is often unknown in practice. In this case, K also needs to be estimated. For this, one may use model order selection techniques, such as minimum description length (MDL), or the Akaike or Bayesian information criteria [3], which balance the model complexity (number of parameters, i.e., sources) against data fidelity to avoid overfitting or underfitting.

E.g., the MDL estimates K by minimizing (see lecture handouts p. 479)

$$\mathrm{MDL}(k) \triangleq -(M-k)N\log\left(\frac{\left(\prod_{m=k+1}^{M}\lambda_m\right)^{1/(M-k)}}{\frac{1}{M-k}\sum_{m=k+1}^{M}\lambda_i}\right) + \frac{1}{2}k(2M-k)\log N$$

for $k \in \{0, 1, \ldots, M-1\}$, i.e., $\hat{K}_{\text{MDL}} = \arg\min_k \text{MDL}(k)$. The ratio of the geometric and arithmetic mean of the $M-k$ smallest eigenvalues λ_m (of matrix $\mathbf{\Sigma}$, or a finite-sample estimate thereof) tests the equality of the hypothesized noise eigenvalues. This ratio equals one if $\lambda_m = c > 0$ for $m = k+1, \ldots, M$. Otherwise it is less than one (but nonnegative).

Further resources

MATLAB tutorials (you should program your own functions and *not* use the Phased Array System Toolbox):

- DoA estimation using beamforming, MVDR, and MUSIC:
  https://se.mathworks.com/help/phased/ug/direction-of-arrival-estimation-with-beamscan-mvdr-and-music.html
- High resoution DoA estimation using MUSIC, ESPRIT, and root-WSF:
  https://se.mathworks.com/help/phased/ug/high-resolution-direction-of-arrival-estimation.html

Textbooks:

- P. Stoica and R. L. Moses, *Spectral analysis of signals.* Pearson Prentice Hall Upper Saddle River, NJ, 2005 (http://user.it.uu.se/~ps/SAS-new.pdf)
- S. J. Orfanidis, *Electromagnetic waves and antennas.* Rutgers University New Brunswick, NJ, 2014 (http://eceweb1.rutgers.edu/~orfanidi/ewa/) — especially Ch. 20 on antenna arrays (Ch. 22 in 2016 edition)

3 Kalman filtering

The Kalman filter is a sequential MMSE estimator and generalizaton of the Wiener filter. It can be thought of as a dynamical filter, which estimates the internal state of interest using partial state measurements, and a prediction based on the assumed dynamical model. For details and examples beyond the lecture handouts, see Kay's book [1] and

- Wikipedia example: https://en.wikipedia.org/wiki/Kalman_filter#Example_application,_technical

References

[1] S. M. Kay, *Fundamentals of statistical signal processing: Estimation theory.* Prentice Hall PTR, 1993.

[2] B. Ottersten, R. Roy, and T. Kailath, "Signal waveform estimation in sensor array processing," in *23rd Asilomar Conference on Signals, Systems and Computers*, vol. 2, 1989, pp. 787–791.

[3] P. Stoica and Y. Selen, "Model-order selection: a review of information criterion rules," *IEEE Signal Processing Magazine*, vol. 21, no. 4, pp. 36–47, July 2004.

[4] P. Stoica and R. L. Moses, *Spectral analysis of signals.* Pearson Prentice Hall Upper Saddle River, NJ, 2005.

[5] S. J. Orfanidis, *Electromagnetic waves and antennas.* Rutgers University New Brunswick, NJ, 2014.