Four Decades of Array Signal Processing Research: An Optimization Relaxation Technique Perspective

Marius Pesavento, Minh Trinh-Hoang and Mats Viberg

Technical University of Darmstadt
Darmstadt, Germany

Blekinge Institute of Technology
Karlskrona, Sweden

Acknowledgement

Special thanks to

- Christian Steffens
- Yang Yang

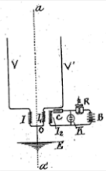
Financial support from

- **EXPRESS II** project (DFG-German Research Foundation Priority Program SSP-1798 CoSIP) under project number PE2080/1-2.
- **PRIDE** project (DFG - German Research Foundation) under project number PE2080/2-1.

Prof. Alex B. Gershman (1962-2011)

Great scientist, teacher and friend.

History



- RF-based direction finding invented by Stone Stone in 1902 who patented a two element array with less than half wavelength [Stone'1902], [Stone'1906-2].
- Later improved upon by De Forest [de Forest'1904], Marconi [Marconi'1906], Bellini and Tosi [Bellini'1909],[Bellini'1910], and Adcock [Adcock'1919].
- See [Schantz'11] for an overview on the origin of RF-based direction finding.
- The shift from analog to digital array processing widely facilitated the processing of data and the diversity of applications, e.g., in seismic applications [Capon'66], [Capon'67].
- In the late seventies the research area greatly advanced with the introduction of the first "super resolution" algorithms [Schmidt'79], [Schmidt'81],[Bienvenu'79],[Barabell'83] [Böhme'84],[Ziskind'99] [Stoica'89], [Böhme'86],[Viberg'91],[Stoica'90].

Motivation

What to expect for the tutorial

- Zhi-Quan (Tom) Luo's tutorial in the morning on advances in robust optimization methods for beamforming under DoA estimation errors.
- Beamforming: Extract signal of interest in presence of interference and noise.
- DoA estimation: Determine directions of multiple superimposed signals in the presence of noise.
- The progress in sensor array processing is closely linked to advances in modern optimization (and sometimes also vise-versa).
- In morning tutorial advanced optimization concepts like convex relaxation, successive (upper bound) approximation have been discussed ...
- ... these concepts will also become important in our tutorial.
- How to perform DoA estimation?
- How to treat interference in DoA estimation?

Motivation

What to expect for the tutorial

- The tutorial addresses both,
  - experienced researchers in sensor array processing, as well as,
  - newcomers to the field.
- In this tutorial, we revisit aspects of four decades of "super-resolution" DoA estimation.
- We approach classical and novel DoA estimation methods from a modern optimization (problem approximation/ problem relaxation) perspective.
- We highlight, how problem approximation and relaxation have always played an important role in developing efficient algorithms:
  - sometimes explicitly in the design ...
  - ... often implicitly, as the consequence of proposed (ad-hoc) algorithms.

Motivation
What to expect for the tutorial

- We show novel derivations for existing algorithm that explicitly highlight the use of relaxation of prior knowledge ...
- ... and introduce a framework for designing novel algorithm under partial relaxation.

Table of Contents

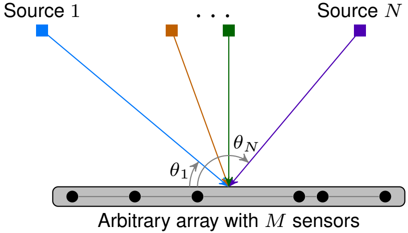
Table of Contents

Applications

- Direction-of-Arrival (DoA) estimation is linked to fundamental problems: harmonic retrieval, frequency estimation, and time-delay estimation.
- One of most widely applied and studied estimation problems.
- Numerous classical applications
  - Radar (military, automotive).
  - Sonar (source localization).
  - Communications (directed transmission, satellite communication).
  - Radio Astronomy (high resolution imaging).
  - Medical Imaging (ultrasound, tomography).
  - Geophysical Exploration (seismic, oil exploration).
  - Biomedical (hearing aids, heart rate monitoring).
- More recent applications
  - Drone localization at airports and public buildings.
  - Parametric channel estimation and user localization in Massive MIMO.

Conventional Signal Model
Assumptions and Signal Model

- Sensor array composed of M sensors.
- N sources in the far-field of the array. $\left(\text{distance} \gg \frac{2\times(\text{diameter of array})^2}{\text{wavelength}}\right)$
- N plane wave narrow-band signals impinge on array.
- We assume that the number of sensors M exceeds the number of source signals N, hence $M > N$.



Conventional Signal Model

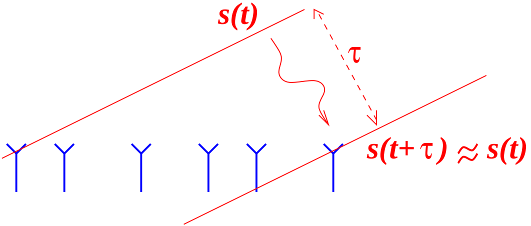
Assumptions and Signal Model

Narrowband condition:

- The relative bandwidth of the signals is small.

$$\text{relative bandwidth} = \frac{\text{signal bandwidth}}{\text{carrier frequency}} \ll \frac{1}{\pi M}$$

- The maximal traveling time $\tau_{\max}$ across the array is substantially smaller than the effective correlation time of signal waveforms.



Conventional Signal Model
Assumptions and Signal Model

Single measurement version for time instant t

$$\boldsymbol{x}(t) = \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{s}(t) + \boldsymbol{n}(t)$$

- $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_N]^{\mathsf{T}}$: DOAs of N source signals.
- W.l.o.g. we consider only azimuth angle estimation $\theta \in \Theta = [0, 180^\circ)$.
- $\boldsymbol{A}(\boldsymbol{\theta}) = [\boldsymbol{a}(\theta_1), ..., \boldsymbol{a}(\theta_N)] \in \mathbb{C}^{M \times N}$: Steering matrix.
- $\boldsymbol{a}(\theta)$: Steering vector from the direction θ.
  - Dependent on the geometry of the sensor array and the direction θ.
  - Example: Uniform Linear Array (ULA) with baseline d:

$$\boldsymbol{a}(\theta) = [1, e^{-\mathrm{j}\frac{2\pi}{\lambda} d\cos(\theta)}, \ldots, e^{-\mathrm{j}\frac{2\pi}{\lambda}(M-1)d\cos(\theta)}]^{\mathsf{T}}.$$

Array manifold

$$\mathcal{A}_N = \left\{\boldsymbol{A} \in \mathbb{C}^{M \times N} \middle|\, \boldsymbol{A} = [\boldsymbol{a}(\vartheta_1), \ldots, \boldsymbol{a}(\vartheta_N)] \text{ with } 0 \leq \vartheta_1 < \ldots < \vartheta_N < 180^\circ\right\}$$

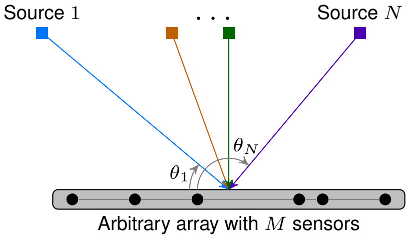
We assume for simplicity w.l.o.g. that the first sensor in the array is the reference sensor with $\boldsymbol{e}_1^{\mathsf{T}}\boldsymbol{A} = \boldsymbol{1}_N^{\mathsf{T}}$.

Conventional Signal Model
Assumptions and Signal Model

Array measurement (snapshot) at time instant t.

$$\boldsymbol{x}(t) = \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{s}(t) + \boldsymbol{n}(t)$$

- $\boldsymbol{x}(t) = [x_1(t), \ldots, x_M(t)]^\mathsf{T} \in \mathbb{C}^{M\times 1}$: Receive signal vector of the M sensors.
- $\boldsymbol{s}(t) = [s_1(t), \ldots, s_N(t)]^\mathsf{T} \in \mathbb{C}^{N\times 1}$: Source signal vector of the N sources.
- $\boldsymbol{n}(t) = [n_1(t), \ldots, n_M(t)]^\mathsf{T} \in \mathbb{C}^{M\times 1}$: Sensor noise vector of the M sensors.



Conventional Signal Model
Assumptions and Signal Model

Sensor noise $\boldsymbol{n}(t)$ modeled as complex circular Gaussian random variable $\mathbf{n}(t)$, with:

- Identical noise variance (power) ν in all sensors (uniform).
- Independent noise in different antennas (spatially white).
- Independent noise in different time instants (temporally white).

Uniform spatially and temporally white noise

- Zero mean: $\mathbb{E}\left\{\mathbf{n}(t)\right\} = \mathbf{0}_M$.
- Covariance matrix: $\mathbb{E}\left\{\mathbf{n}(t)\mathbf{n}^{\mathsf{H}}(t)\right\} = \nu\,\boldsymbol{I}_M \in \mathbb{C}^{M\times M}$.

Conventional Signal Model

Assumptions and Signal Model

Multiple measurement version: T snapshots

$$\boldsymbol{X} = \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{S} + \boldsymbol{N}$$

- $\boldsymbol{X} = [\boldsymbol{x}(1), \boldsymbol{x}(2), ..., \boldsymbol{x}(T)] \in \mathbb{C}^{M \times T}$: Received signal matrix.
- $\boldsymbol{S} = [\boldsymbol{s}(1), \boldsymbol{s}(2), ..., \boldsymbol{s}(T)] \in \mathbb{C}^{N \times T}$: Source signal matrix.
- $\boldsymbol{N} = [\boldsymbol{n}(1), \boldsymbol{n}(2), ..., \boldsymbol{n}(T)] \in \mathbb{C}^{M \times T}$: Sensor noise matrix.
- T : Number of available snapshots.

Objective:

Given the received signal $\boldsymbol{X}$ and the mapping $\boldsymbol{\theta} \mapsto \boldsymbol{A}(\boldsymbol{\theta})$, estimate the DOAs $\boldsymbol{\theta}$

Conventional Signal Model
Stochastic and Deterministic Covariance Model

Signal waveform $\boldsymbol{s}(t)$ modeled as complex circular Gaussian random variable $\mathbf{s}(t)$.

Stochastic (unconditional) signal model

- Zero mean: $\mathbb{E}\{\mathbf{s}(t)\} = \mathbf{0}_N$.
- Signal covariance matrix: $\boldsymbol{P} = \mathbb{E}\left\{\mathbf{s}(t)\mathbf{s}^{\mathsf{H}}(t)\right\} \in \mathbb{C}^{N\times N}$.
- Non-singularity: $\boldsymbol{P} \succ 0$ (not fully coherent signals).
- Gaussian measurements: $\mathbf{x}(t) \sim \mathcal{N}_{\mathsf{C}}\big(\mathbf{0}_M, \boldsymbol{R}\big)$.
- Receive correlation matrix: $\boldsymbol{R} = \mathbb{E}\left\{\mathbf{x}(t)\mathbf{x}^{\mathsf{H}}(t)\right\}$.
  $= \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{P}\boldsymbol{A}^{\mathsf{H}}(\boldsymbol{\theta}) + \nu\boldsymbol{I}_M \in \mathbb{C}^{M\times M}$.
- Parameter characterization: $\boldsymbol{\theta} \in \Theta^N, \boldsymbol{P} \in \mathbb{C}^{N\times N}, \nu \in \mathbb{R}_+$.

Number of parameters independent of number of observations T.

Conventional Signal Model
Stochastic and Deterministic Covariance Model

Signal waveform $\boldsymbol{s}(t)$ modeled as deterministic quantity.
Received signal $\boldsymbol{x}(t)$ modeled as random variable $\mathbf{x}(t) = \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{s}(t) + \mathbf{n}(t)$.

Deterministic (conditional) signal model

- Gaussian measurements: $\mathbf{x}(t) \sim \mathcal{N}_{\mathrm{C}}\big(\boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{s}(t), \nu \boldsymbol{I}\big)$.
- Parameter characterization: $\boldsymbol{\theta} \in \Theta^N$,
  $\boldsymbol{S} = [\boldsymbol{s}(1), \boldsymbol{s}(2), ..., \boldsymbol{s}(T)] \in \mathbb{C}^{N\times T}, \nu \in \mathbb{R}_+$.

Number of parameters grows with number of observations T.

Conventional Signal Model

Stochastic and Deterministic Covariance Model

- In practice the true received signal covariance matrix $\boldsymbol{R}$ is not available and must be estimated from finite samples.
- A commonly use sample covariance/correlation matrix estimator is given as:

Sample correlation/correlation matrix

$$\hat{\boldsymbol{R}} = \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{x}(t)\boldsymbol{x}^{\mathsf{H}}(t) = \frac{1}{T}\boldsymbol{X}\boldsymbol{X}^{\mathsf{H}}$$

Table of Contents

Table of Contents

Performance Bound
Review of Crámer-Rao Bound

Parametric Model

- Random stationary process $\mathbf{x}$.
- Observations over time $\boldsymbol{x}(t) \in \mathcal{X}$ for $t = 1, \ldots, T$ of the random process $\mathbf{x}$.
- Non-redundant deterministic parameter vector $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_I]^\mathsf{T} \in \mathbb{R}^{I \times 1}$.
- Probability density function for a given parameter $f_{\mathbf{x}}(\boldsymbol{x}|\boldsymbol{\alpha})$.

Objective of Parametric Estimation

- Assumption: Independent observations over time drawn from the same probability density function with the true parameter $\boldsymbol{\alpha}_{\text{true}}$.
- Given the observations $\{\boldsymbol{x}(1), \ldots, \boldsymbol{x}(T)\}$ and the family of the probability density functions $f_{\mathbf{x}}(\boldsymbol{x}|\boldsymbol{\alpha})$.
- Estimate $\boldsymbol{\alpha}_{\text{true}}$ by an estimator $\hat{\boldsymbol{\alpha}}$.

Performance Bound
Review of Crámer-Rao Bound

For a given estimator $\hat{\alpha} = T\big(\mathbf{x}(1), \ldots, \mathbf{x}(T)\big)$

- Bias $\boldsymbol{\mu} = \mathbb{E}\{\hat{\boldsymbol{\alpha}}\}$.
- Covariance $\boldsymbol{\Sigma} = \mathbb{E}\left\{(\hat{\boldsymbol{\alpha}} - \boldsymbol{\mu})(\hat{\boldsymbol{\alpha}} - \boldsymbol{\mu})^{\mathsf{H}}\right\}$.

Fisher Information Matrix

Under some regularity conditions, the Fisher Information Matrix (FIM) is defined as

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\alpha}) = -\mathbb{E}\left\{\nabla^2_{\boldsymbol{\alpha}}\big(\log f_{\mathbf{x}}(\boldsymbol{x}|\boldsymbol{\alpha})\big)\right\}.$$

Crámer-Rao Inequality

For any unbiased estimator $\hat{\alpha}$ with the covariance matrix $\boldsymbol{\Sigma}$, we have

$$\boldsymbol{\Sigma} \succeq \boldsymbol{C}(\boldsymbol{\alpha}_{\text{true}}) = \left[\boldsymbol{\mathcal{I}}(\boldsymbol{\alpha}_{\text{true}})\right]^{-1}.$$

Performance Bound
Review of Crámer-Rao Bound

Special Case: Gaussian case

- Parameter vector: $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_I]^{\mathsf{T}}$.
- Circularly-symmetric complex Gaussian observation: $\mathbf{x} \sim \mathcal{N}_{\mathsf{C}}\big(\boldsymbol{m}(\boldsymbol{\alpha}), \boldsymbol{K}(\boldsymbol{\alpha})\big)$.

Slepian-Bangs Formula

The *ij*-th element of the FIM matrix is given by

$$\Big[\boldsymbol{\mathcal{I}}(\boldsymbol{\alpha})\Big]_{ij} = \mathrm{Tr}\left(\boldsymbol{K}(\boldsymbol{\alpha})^{-1}\frac{\partial \boldsymbol{K}(\boldsymbol{\alpha})}{\partial \alpha_i}\boldsymbol{K}(\boldsymbol{\alpha})^{-1}\frac{\partial \boldsymbol{K}(\boldsymbol{\alpha})}{\partial \alpha_j}\right) + 2\mathrm{Re}\left\{\frac{\partial \boldsymbol{m}(\boldsymbol{\alpha})^{\mathsf{H}}}{\partial \alpha_i}\boldsymbol{K}(\boldsymbol{\alpha})^{-1}\frac{\partial \boldsymbol{m}^{\mathsf{H}}(\boldsymbol{\alpha})}{\partial \alpha_j}\right\}.$$

Necessary condition for the invertibility of the FIM matrix

- The parameter vector must be locally identifiable.
- Consequence: the parameters must be non-redundant.

Performance Bound
Review of Crámer-Rao Bound

Partition the FIM matrix

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\alpha}) = \begin{bmatrix} \boldsymbol{\mathcal{I}}_{\boldsymbol{\theta\theta}} & \boldsymbol{\mathcal{I}}_{\boldsymbol{\theta\beta}} \\ \boldsymbol{\mathcal{I}}_{\boldsymbol{\beta\theta}} & \boldsymbol{\mathcal{I}}_{\boldsymbol{\beta\beta}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{C}_{\boldsymbol{\theta\theta}} & \boldsymbol{C}_{\boldsymbol{\theta\beta}} \\ \boldsymbol{C}_{\boldsymbol{\beta\theta}} & \boldsymbol{C}_{\boldsymbol{\beta\beta}} \end{bmatrix}^{-1} \text{ with } \boldsymbol{\alpha} = \left[\boldsymbol{\theta}^{\mathsf{T}}, \boldsymbol{\beta}^{\mathsf{T}}\right]^{\mathsf{T}}$$

- $\boldsymbol{\theta}$ contains desired parameters.
- $\boldsymbol{\beta}$ contains nuisance parameters.

Crámer-Rao bound of the desired parameters $\boldsymbol{\theta}$

$$\boldsymbol{C}_{\boldsymbol{\theta\theta}} = \left(\boldsymbol{\mathcal{I}}_{\boldsymbol{\theta\theta}} - \boldsymbol{\mathcal{I}}_{\boldsymbol{\theta\beta}} \boldsymbol{\mathcal{I}}_{\boldsymbol{\beta\beta}}^{-1} \boldsymbol{\mathcal{I}}_{\boldsymbol{\beta\theta}}\right)^{-1}$$

Performance Bound
Review of Crámer-Rao Bound

Recall the Deterministic Signal Model

$$\mathbf{x}(t) \sim \mathcal{N}_{\mathrm{C}}\big(\boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{s}(t), \nu\boldsymbol{I}\big) \text{ for all } t = 1, \ldots, T.$$

Deterministic Crámer-Rao Bound

$$\boldsymbol{C}_{\mathrm{det}}(\boldsymbol{\theta}) = \boldsymbol{C}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \frac{\nu}{2T}\mathrm{Re}\left\{\hat{\boldsymbol{P}}^{\mathsf{T}} \odot \left(\boldsymbol{D}^{\mathsf{H}}\boldsymbol{\Pi}_{\boldsymbol{A}}^{\perp}\boldsymbol{D}\right)\right\}^{-1}$$

- $\hat{\boldsymbol{P}} = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{s}(t)\boldsymbol{s}^{\mathsf{H}}(t) = \frac{1}{T}\boldsymbol{S}\boldsymbol{S}^{\mathsf{H}}$
- $\boldsymbol{D} = \left[\frac{\mathrm{d}\boldsymbol{a}(\theta_1)}{\mathrm{d}\theta}, \ldots, \frac{\mathrm{d}\boldsymbol{a}(\theta_N)}{\mathrm{d}\theta}\right]$

Performance Bound

Review of Crámer-Rao Bound

Recall the Stochastic Signal Model

$$\mathbf{x}(t) \sim \mathcal{N}_{\mathrm{C}}\left(\mathbf{0}, \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{P}\boldsymbol{A}^{\mathsf{H}}(\boldsymbol{\theta}) + \nu\boldsymbol{I}\right) \text{ for all } t = 1, \ldots, T$$

Stochastic Crámer-Rao Bound

$$\boldsymbol{C}_{\mathrm{sto}}(\boldsymbol{\theta}) = \boldsymbol{C}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \frac{\nu}{2T}\mathrm{Re}\left\{\boldsymbol{M}^{\mathsf{T}} \odot \left(\boldsymbol{D}^{\mathsf{H}}\boldsymbol{\Pi}_{\boldsymbol{A}}^{\perp}\boldsymbol{D}\right)\right\}^{-1}$$

- $\boldsymbol{M} = \boldsymbol{P}\boldsymbol{A}^{\mathsf{H}}\boldsymbol{R}^{-1}\boldsymbol{A}\boldsymbol{P}$
- $\boldsymbol{D} = \left[\frac{\mathrm{d}\boldsymbol{a}(\theta_1)}{\mathrm{d}\theta}, \ldots, \frac{\mathrm{d}\boldsymbol{a}(\theta_N)}{\mathrm{d}\theta}\right]$

Table of Contents

General procedure [Lehmann'98]

- **Step 1:** Determine analytically a multivariate pdf $f\big(\mathbf{x}(1),\ldots,\mathbf{x}(T)|\boldsymbol{\alpha}\big)$ as a function of random observation model vectors and nonrandom parameters $\boldsymbol{\alpha}$.
- **Step 2:** Insert actual observations $\boldsymbol{x}(1),\ldots,\boldsymbol{x}(T)$ instead of "hypothetical" observation model vectors (random variables) $\mathbf{x}(1),\ldots,\mathbf{x}(T)$ to obtain the so-called likelihood function $f\big(\boldsymbol{x}(1),\ldots,\boldsymbol{x}(T)|\boldsymbol{\alpha}\big)$ from the pdf.
- **Step 3:** Maximize the likelihood function w.r.t. all unknown parameters and to ML parameter estimates, i.e.

$$\hat{\boldsymbol{\alpha}}_{\mathrm{ML}} = \arg\max_{\boldsymbol{\alpha}} f\big(\boldsymbol{x}(1),\ldots,\boldsymbol{x}(T)|\boldsymbol{\alpha}\big)$$

Why is Maximum Likelihood important?

- Maximum Likelihood achieves the Cramér-Rao lower-bound (under mild regularity conditions).

Concentration of ML function

Use arbitrary partition $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1^\mathsf{T}, \boldsymbol{\alpha}_2^\mathsf{T}]^\mathsf{T}$ of the parameter vector.
Maximize Likelihood function w.r.t. part of the variables, e.g., partition $\boldsymbol{\alpha}_2$ while considering other variables as constant. Hence,

$$\max_{\boldsymbol{\alpha}} f\big(\boldsymbol{x}(1), \ldots, \boldsymbol{x}(T)|\boldsymbol{\alpha}\big) = \max_{\boldsymbol{\alpha}_1} \underbrace{\max_{\boldsymbol{\alpha}_2} f\big(\boldsymbol{x}(1), \ldots, \boldsymbol{x}(T)|\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2\big)}_{g\left(\boldsymbol{x}(1),\ldots,\boldsymbol{x}(T)|\boldsymbol{\alpha}_1\right)}$$

If possible, find analytic (closed-form) solution $\hat{\boldsymbol{\alpha}}_{2,\mathrm{ML}}(\boldsymbol{\alpha}_1)$ (as a function of $\boldsymbol{\alpha}_1$) for inner optimization problem

$$\begin{aligned} g\big(\boldsymbol{x}(1), \ldots, \boldsymbol{x}(T)|\boldsymbol{\alpha}_1\big) &= \max_{\boldsymbol{\alpha}_1} f\big(\boldsymbol{x}(1), \ldots, \boldsymbol{x}(T)|\boldsymbol{\alpha}_1, \hat{\boldsymbol{\alpha}}_{2,\mathrm{ML}}(\boldsymbol{\alpha}_1)\big), \\ \hat{\boldsymbol{\alpha}}_{1,\mathrm{ML}}^\mathsf{T} &= \arg\max_{\boldsymbol{\alpha}_1} \; g\big(\boldsymbol{x}(1), \ldots, \boldsymbol{x}(T)|\boldsymbol{\alpha}_1\big). \end{aligned}$$

Under the deterministic (unconditional) model [Böhme'84],[Ziskind'99]

$$\mathbf{x}(t) \sim \mathcal{N}_{\mathrm{C}}\big(\boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{s}(t), \nu\boldsymbol{I}\big)$$

with parameter vector $\boldsymbol{\alpha} = [\boldsymbol{\theta}^{\mathsf{T}}, \boldsymbol{s}^{\mathsf{T}}(1), \dots, \boldsymbol{s}^{\mathsf{T}}(T), \nu]^{\mathsf{T}}$.
Hence the corresponding likelihood is

$$f\big(\boldsymbol{x}(1), \dots, \boldsymbol{x}(T)|\boldsymbol{\alpha}\big) = \prod_{t=1}^{T} \frac{1}{(\pi\nu)^M} \exp\left(-\frac{\|\boldsymbol{x}(t) - \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{s}(t)\|^2}{\nu}\right).$$

The negative log-likelihood is

$$\mathcal{L}\big(\boldsymbol{x}(1), \dots, \boldsymbol{x}(T)|\boldsymbol{\alpha}\big) = \sum_{t=1}^{T} M \ln(\pi\nu) + \sum_{t=1}^{T} \frac{1}{\nu} \|\boldsymbol{x}(t) - \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{s}(t)\|^2.$$

Closed form expressions for ML estimates for fixed $\boldsymbol{\theta}$

$$
\begin{aligned}
\hat{\boldsymbol{s}}_{\mathrm{DML}}(t) &= \left(\boldsymbol{A}^{\mathsf{H}}(\boldsymbol{\theta})\boldsymbol{A}(\boldsymbol{\theta})\right)^{-1}\boldsymbol{A}^{\mathsf{H}}(\boldsymbol{\theta})\boldsymbol{x}(t) = \boldsymbol{A}^{\dagger}(\boldsymbol{\theta})\boldsymbol{x}(t) \\
\hat{\nu}_{\mathrm{DML}} &= \frac{1}{M}\mathrm{Tr}\left(\boldsymbol{\Pi}^{\perp}_{\boldsymbol{A}(\boldsymbol{\theta})}\hat{\boldsymbol{R}}\right)
\end{aligned}
$$

and where

$$
\begin{aligned}
\boldsymbol{A}^{\dagger}(\boldsymbol{\theta}) &= \left(\boldsymbol{A}^{\mathsf{H}}(\boldsymbol{\theta})\boldsymbol{A}(\boldsymbol{\theta})\right)^{-1}\boldsymbol{A}^{\mathsf{H}}(\boldsymbol{\theta}) \\
\boldsymbol{\Pi}_{\boldsymbol{A}(\boldsymbol{\theta})} &= \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{A}^{\dagger}(\boldsymbol{\theta}) \\
\text{and} \qquad \boldsymbol{\Pi}^{\perp}_{\boldsymbol{A}(\boldsymbol{\theta})} &= \boldsymbol{I} - \boldsymbol{\Pi}_{\boldsymbol{A}(\boldsymbol{\theta})}
\end{aligned}
$$

denote the pseudo-inverse of $\boldsymbol{A}(\boldsymbol{\theta})$, and projectors onto the range space and nullspace of $\boldsymbol{A}(\boldsymbol{\theta})$, respectively.

Parametric Methods
Maximum Likelihood

Inserting $\hat{\boldsymbol{s}}_{\text{DML}}(t)$ and $\hat{\nu}_{\text{DML}}$ back into the log-likelihood

$$\mathcal{L}\Big(\boldsymbol{x}(1), \ldots, \boldsymbol{x}(T) | \boldsymbol{\theta}\Big) = TM \left(\ln \left(\text{Tr}\Big(\boldsymbol{\Pi}^{\perp}_{\boldsymbol{A}(\boldsymbol{\theta})} \hat{\boldsymbol{R}}\Big) \right) + \ln(\pi) - \ln(M) + 1 \right).$$

Minimization w.r.t. θ: [Böhme'84]

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{\text{DML}} &= \arg\min_{\theta} \mathcal{L}\Big(\boldsymbol{x}(1), \ldots, \boldsymbol{x}(T) | \boldsymbol{\theta}\Big) \\ &= \arg\min_{\boldsymbol{\theta}} \text{Tr}\Big(\boldsymbol{\Pi}^{\perp}_{\boldsymbol{A}(\boldsymbol{\theta})} \hat{\boldsymbol{R}}\Big) \end{aligned}$$

Interpretation: Find DoAs such that the total received energy in noise subspace is minimized.

Parametric Methods
Maximum Likelihood

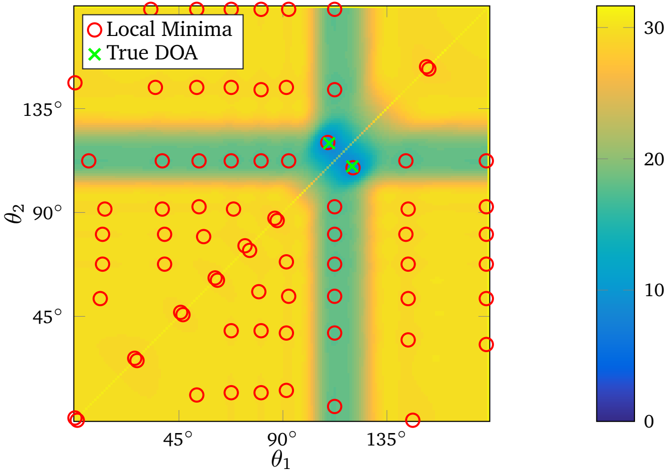
Minimization of concentrated log-likelihood function

$$f_{\mathrm{DML}}(\boldsymbol{\theta}) = \mathrm{Tr}\big(\boldsymbol{\Pi}^{\perp}_{\boldsymbol{A}(\boldsymbol{\theta})}\hat{\boldsymbol{R}}\big)$$

- $f_{\mathrm{DML}}(\boldsymbol{\theta})$ is highly multi-modal, many local optima with cost close to global optimum.
- Minimum can not be computed in closed form.
- Costly N dimensional search over field of view is required.
- Complexity grows exponentially with number of sources N.
- Generally, complexity becomes prohibitive if $N > 3$ sources.

Parametric Methods
Maximum Likelihood

$M = 10, \boldsymbol{\theta} = [110°, 120°]^{\mathrm{T}}, \mathrm{SNR} = 0\ \mathrm{dB}, T = 100$



Under the stochastic (unconditional) model [Böhme'86],[Bresler'88],[Jaffer'88],[Stoica'90-2]

$$\mathbf{x}(t) \sim \mathcal{N}_{\mathrm{C}}\big(\mathbf{0}_M, \boldsymbol{R}\big)$$

with $\boldsymbol{R} = \mathrm{E}\, \boldsymbol{x}(t)\boldsymbol{x}^{\mathsf{H}}(t) = \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{P}\boldsymbol{A}^{\mathsf{H}}(\boldsymbol{\theta}) + \nu \boldsymbol{I}_M$ and parameter vector $\boldsymbol{\alpha} = [\boldsymbol{\theta}^{\mathsf{T}}, \boldsymbol{p}^{\mathsf{T}}, \nu]^{\mathsf{T}}$.

Vector $\boldsymbol{p} \in \mathbb{R}^{N^2}$ contains the N elements on diagonal of matrix $\boldsymbol{P}$ and the $(N^2 - N)$ elements characterizing real and imaginary part of upper triangular of $\boldsymbol{P}$.
Hence the corresponding likelihood is

$$f\big(\boldsymbol{x}(1), \ldots, \boldsymbol{x}(T)|\boldsymbol{\alpha}\big) = \prod_{t=1}^{T} \frac{1}{\pi^M \det(\boldsymbol{R})} \exp\left(-\boldsymbol{x}^{\mathsf{H}}(t)\boldsymbol{R}^{-1}(\boldsymbol{\theta})\boldsymbol{x}(t)\right).$$

Parametric Methods
Stochastic Maximum Likelihood

The negative log-likelihood is

$$\mathcal{L}\big(\boldsymbol{x}(1), \ldots, \boldsymbol{x}(T) | \boldsymbol{\alpha}\big) = T \left(M \ln(\pi) + \ln\det\left(\boldsymbol{R}\right) + \operatorname{Tr} \boldsymbol{R}^{-1} \hat{\boldsymbol{R}} \right)$$

Close form expressions for ML estimates for fixed $\boldsymbol{\theta}$

$$\hat{\nu}_{\text{SML}} = \frac{1}{M-N} \operatorname{Tr} \boldsymbol{\Pi}^{\perp}_{\boldsymbol{A}(\boldsymbol{\theta})} \hat{\boldsymbol{R}}$$

$$\hat{\boldsymbol{P}}_{\text{SML}} = \boldsymbol{A}^{\dagger}(\boldsymbol{\theta}) \left(\hat{\boldsymbol{R}} - \hat{\nu}_{\text{SML}} \boldsymbol{I}_M \right) \boldsymbol{A}^{\dagger \mathsf{H}}(\boldsymbol{\theta})$$

Inserting $\hat{\nu}_{\text{SML}}$ and $\hat{\boldsymbol{P}}_{\text{SML}}$ back and minimizing w.r.t. $\boldsymbol{\theta}$ yields

$$\hat{\boldsymbol{\theta}}_{\text{SML}} = \arg\min_{\boldsymbol{\theta}} \det \left(\boldsymbol{\Pi}_{\boldsymbol{A}(\boldsymbol{\theta})} \hat{\boldsymbol{R}} \boldsymbol{\Pi}_{\boldsymbol{A}(\boldsymbol{\theta})} + \underbrace{\frac{1}{M-N} \operatorname{Tr} \left(\boldsymbol{\Pi}^{\perp}_{\boldsymbol{A}(\boldsymbol{\theta})} \hat{\boldsymbol{R}} \right)}_{\hat{\nu}_{\text{SML}}} \boldsymbol{\Pi}^{\perp}_{\boldsymbol{A}(\boldsymbol{\theta})} \right).$$

Parametric Methods
Weighted Subspace Fitting

Eigendecomposition of array covariance matrix

$$\boldsymbol{R} = \mathrm{E}\, \mathbf{x}(t)\mathbf{x}^{\mathsf{H}}(t) = \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{P}\boldsymbol{A}^{\mathsf{H}}(\boldsymbol{\theta}) + \nu \boldsymbol{I}_M$$
$$= \sum_{m=1}^{M} \lambda_m \boldsymbol{u}_m \boldsymbol{u}_m^{\mathsf{H}}$$

where $\lambda_1 \geq \lambda_2 \ldots \geq \lambda_M \in \mathbb{R}_+$ are sorted eigenvalues of $\boldsymbol{R}$.

From the eigenanalysis of $\boldsymbol{R}$ we obtain that:

$\lambda_m > \nu, \quad m = 1, \ldots, N$ signal subspace eigenvalues

$\lambda_m = \nu, \quad m = N+1, \ldots, M$ noise subspace eigenvalues

with corresponding eigenvectors:

$\boldsymbol{u}_1, \ldots, \quad \boldsymbol{u}_N,$ signal eigenvectors

$\boldsymbol{u}_{N+1}, \ldots, \quad \boldsymbol{u}_M$ noise eigenvectors.

Parametric Methods
Weighted Subspace Fitting

Eigendecomposition in compact matrix notation:

$$\boldsymbol{R} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathsf{H}} = \boldsymbol{U}_{\mathrm{s}}\boldsymbol{\Lambda}_{\mathrm{s}}\boldsymbol{U}_{\mathrm{s}}^{\mathsf{H}} + \boldsymbol{U}_{\mathrm{n}}\boldsymbol{\Lambda}_{\mathrm{n}}\boldsymbol{U}_{\mathrm{n}}^{\mathsf{H}}$$

where we define

$\boldsymbol{U}_{\mathrm{s}} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N] \in \mathbb{C}^{M \times N}$	signal eigenvector matrix
$\boldsymbol{U}_{\mathrm{n}} = [\boldsymbol{u}_{N+1}, \ldots, \boldsymbol{u}_M] \in \mathbb{C}^{M \times (M-N)}$	noise eigenvector matrix
$\boldsymbol{\Lambda}_{\mathrm{s}} = \mathrm{diag}(\lambda_1, \ldots, \lambda_N) \in \mathbb{S}_+^{N \times N}$	diagonal matrix of signal eigenvalues
$\boldsymbol{\Lambda}_{\mathrm{n}} = \nu\, \boldsymbol{I}_{M-N} \in \mathbb{S}_+^{(M-N) \times (M-N)}$	diagonal matrix of noise eigenvalues

and

$\boldsymbol{U} = [\boldsymbol{U}_{\mathrm{s}}, \boldsymbol{U}_{\mathrm{s}}] \in \mathbb{C}^{M \times M}$	unitary matrix of eigenvectors
$\boldsymbol{\Lambda} = \mathrm{blkdiag}\,(\boldsymbol{\Lambda}_{\mathrm{s}}, \boldsymbol{\Lambda}_{\mathrm{n}}) \in \mathbb{S}_+^{M \times M}$	diagonal matrix of eigenvalues.

Parametric Methods
Weighted Subspace Fitting

- $\boldsymbol{U}$ is unitary, i.e. $\boldsymbol{U}^{\mathsf{H}}\boldsymbol{U} = \boldsymbol{I}_M$.
- The columns of the signal subspace eigenvectors $\boldsymbol{U}_{\mathrm{s}}$ span the signal subspace, i.e., the range space spanned by the columns of the steering matrix $\boldsymbol{A}(\boldsymbol{\theta})$ at the true DOAs $\boldsymbol{\theta}$, hence

$$\mathcal{R}(\boldsymbol{U}_{\mathrm{s}}) = \mathcal{R}(\boldsymbol{A}(\boldsymbol{\theta})).$$

- There exists a non-singular matrix $\boldsymbol{K} \in \mathbb{C}^{N \times N}$ such that $\boldsymbol{U}_{\mathrm{s}} = \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{K}$.
- The columns of the noise subspace eigenvectors $\boldsymbol{U}_{\mathrm{n}}$ span the noise-space, i.e., the null-space of the Hermitian of the true steering matrix $\boldsymbol{A}(\boldsymbol{\theta})$

$$\mathcal{R}(\boldsymbol{U}_{\mathrm{n}}) = \mathcal{N}(\boldsymbol{A}^{\mathsf{H}}(\boldsymbol{\theta})).$$

- Hence, the columns of the noise subspace eigenvectors $\boldsymbol{U}_{\mathrm{n}}$ are orthogonal to the column-space of the true steering matrix $\boldsymbol{A}(\boldsymbol{\theta})$, i.e.,

$$\boldsymbol{U}_{\mathrm{n}}^{\mathsf{H}}\boldsymbol{A}(\boldsymbol{\theta}) = \boldsymbol{0}_{(M-N)\times N}.$$

Parametric Methods
Weighted Subspace Fitting

The eigendecomposition of the finite sample covariance matrix $\hat{\boldsymbol{R}}$ is given by:

$$\hat{\boldsymbol{R}} = \hat{\boldsymbol{U}}\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{U}}^{\mathsf{H}} = \hat{\boldsymbol{U}}_{\mathrm{s}}\hat{\boldsymbol{\Lambda}}_{\mathrm{s}}\hat{\boldsymbol{U}}_{\mathrm{s}}^{\mathsf{H}} + \hat{\boldsymbol{U}}_{\mathrm{n}}\boldsymbol{\Lambda}_{\mathrm{n}}\hat{\boldsymbol{U}}_{\mathrm{n}}^{\mathsf{H}}$$

where we define for $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_M$

$\hat{\boldsymbol{U}}_{\mathrm{s}} = [\hat{\boldsymbol{u}}_1, \ldots, \hat{\boldsymbol{u}}_N] \in \mathbb{C}^{M \times N}$ sample signal eigenvector matrix

$\hat{\boldsymbol{U}}_{\mathrm{n}} = [\hat{\boldsymbol{u}}_{N+1}, \ldots, \hat{\boldsymbol{u}}_M] \in \mathbb{C}^{M \times (M-N)}$ sample noise eigenvector matrix

$\hat{\boldsymbol{\Lambda}}_{\mathrm{s}} = \mathrm{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_N) \in \mathbb{S}_{+}^{N \times N}$ sample signal eigenvalues

$\hat{\boldsymbol{\Lambda}}_{\mathrm{n}} = \mathrm{diag}(\hat{\lambda}_{N+1}, \ldots, \hat{\lambda}_M) \in \mathbb{S}_{+}^{(M-N) \times (M-N)}$ sample noise eigenvalues

and

$\hat{\boldsymbol{U}} = \left[\hat{\boldsymbol{U}}_{\mathrm{s}}, \hat{\boldsymbol{U}}_{\mathrm{s}}\right] \in \mathbb{C}^{M \times M}$ unitary matrix of eigenvectors

$\hat{\boldsymbol{\Lambda}} = \mathrm{blkdiag}\left(\hat{\boldsymbol{\Lambda}}_{\mathrm{s}}, \hat{\boldsymbol{\Lambda}}_{\mathrm{n}}\right) \in \mathbb{S}_{+}^{M \times M}$ diagonal matrix of eigenvalues.

Parametric Methods
Weighted Subspace Fitting

The DML cost function

$$f_{\mathrm{DML}}(\boldsymbol{\theta}) = \mathrm{Tr}\left(\boldsymbol{\Pi}^{\perp}_{\boldsymbol{A}(\boldsymbol{\theta})}\hat{\boldsymbol{R}}\right)$$

is equivalently obtained from minimizing the Least-Squares fitting problem w.r.t. to the fitting matrix $\boldsymbol{F}$:

$$f_{\mathrm{LS}}(\boldsymbol{\theta}, \boldsymbol{F}) = \|\boldsymbol{X} - \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{F}\|_{\mathsf{F}}^2.$$

The minimization yields the LS estimate

$$\hat{\boldsymbol{F}}_{\mathrm{LS}} = \left(\boldsymbol{A}^{\mathsf{H}}(\boldsymbol{\theta})\boldsymbol{A}(\boldsymbol{\theta})\right)^{-1}\boldsymbol{A}^{\mathsf{H}}(\boldsymbol{\theta})\boldsymbol{X} = \boldsymbol{A}^{\dagger}(\boldsymbol{\theta})\boldsymbol{X}$$

which, if substituted back in the LS function yields the DML function above.

Parametric Methods
Weighted Subspace Fitting

The LS fitting problem can be generalized. A general data matrix $\boldsymbol{M}$ (as some transformation of the data $\boldsymbol{X}$) can be used instead of $\boldsymbol{X}$.

Examples are $\boldsymbol{M} = \hat{\boldsymbol{U}}_\mathrm{s}$ and $\boldsymbol{M} = \hat{\boldsymbol{U}}_\mathrm{s}\hat{\boldsymbol{\Lambda}}_\mathrm{s}^{\frac{1}{2}}$ or most generally

$$\boldsymbol{M} = \hat{\boldsymbol{U}}_\mathrm{s}\boldsymbol{W}^{\frac{1}{2}}$$

for arbitrary weighting matrix $\boldsymbol{W}$.
The corresponding weighted subspace fitting (WSF) problem becomes [Viberg'91],[Ottersten'90],[Stoica'90]

$$f_\mathrm{WSF}(\boldsymbol{\theta}, \boldsymbol{F}) = \|\boldsymbol{M} - \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{F}\|_\mathsf{F}^2$$

or after concentration w.r.t. $\boldsymbol{F}$ with $\hat{\boldsymbol{F}}_\mathrm{WSF} = \boldsymbol{A}^\dagger(\boldsymbol{\theta})\boldsymbol{M}$

$$f_\mathrm{WSF}(\boldsymbol{\theta}) = \mathrm{Tr}\left(\boldsymbol{\Pi}^\perp_{\boldsymbol{A}(\boldsymbol{\theta})}\hat{\boldsymbol{U}}_\mathrm{s}\boldsymbol{W}\hat{\boldsymbol{U}}_\mathrm{s}^\mathsf{H}\right).$$

Parametric Methods
Weighted Subspace Fitting

The WSF estimates for the DOAs $\boldsymbol{\theta}$ are obtained as

$$\hat{\boldsymbol{\theta}}_{\mathrm{WSF}} = \arg\min_{\boldsymbol{\theta}} \operatorname{Tr}\left(\boldsymbol{\Pi}^{\perp}_{\boldsymbol{A}(\boldsymbol{\theta})} \hat{\boldsymbol{U}}_{\mathrm{s}} \boldsymbol{W} \hat{\boldsymbol{U}}_{\mathrm{s}}^{\mathsf{H}}\right).$$

- The minimization of the WSF cost function cannot be carried out in closed-form and generally requires multi-dimensional search.
- Similarly to the multi-dimensional ML methods, the complexity associated with the minimization becomes prohibitive if the number of source $N > 3$.
- The choice of the weighting matrix as

$$\boldsymbol{W}_{\mathrm{ao}} = \left(\hat{\boldsymbol{\Lambda}}_{\mathrm{s}} - \hat{\nu}_{\mathrm{w}} \boldsymbol{I}_N\right)^2 \hat{\boldsymbol{\Lambda}}_{\mathrm{s}}^{-1} \text{ for } \hat{\nu}_{\mathrm{w}} = \frac{1}{M-N} \operatorname{Tr} \hat{\boldsymbol{\Lambda}}_{\mathrm{n}}$$

  is asymptotically (for large T) optimal in terms of the Mean-Squared-Error (MSE) of DOA estimates which achieves the CRB under the stochastic model.

Recall the Covariance Matrix $\boldsymbol{R}$

$$\boldsymbol{R} = \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{P}\boldsymbol{A}^{\mathsf{H}}(\boldsymbol{\theta}) + \nu\boldsymbol{I}$$

Formulation of Covariance Matching Estimation Techniques (COMET) [Ottersten'98]

$$\hat{\boldsymbol{A}}_{\text{COMET}} = \underset{\boldsymbol{A}(\boldsymbol{\theta})\in\mathcal{A}_N}{\arg\min} \min_{\boldsymbol{P}\succeq 0, \nu\geq 0} \left\|\boldsymbol{W} \operatorname{vec}\left(\hat{\boldsymbol{R}} - \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{P}\boldsymbol{A}^{\mathsf{H}}(\boldsymbol{\theta}) - \nu\boldsymbol{I}\right)\right\|_{\mathrm{F}}^2$$

where $\boldsymbol{W} \in \mathbb{C}^{M^2\times M^2}$ is a proper weighting matrix, e.g., $\boldsymbol{W} = \boldsymbol{I}$.

Asymptotically Optimal Weighting Matrix

The MSE of COMET is asymptotically equal to the Stochastic Crámer-Rao bound if the weighting matrix $\boldsymbol{W}$ is chosen as

$$\boldsymbol{W} = \hat{\boldsymbol{W}}_{\text{asymp}} = \left(\hat{\boldsymbol{R}}^{\mathsf{T}} \otimes \hat{\boldsymbol{R}}\right)^{-1/2}.$$

Parametric Methods
Covariance Matching Estimation Techniques

Observation

$$\begin{aligned}\mathrm{vec}\left(\boldsymbol{R}\right) &= \mathrm{vec}\left(\boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{P}\boldsymbol{A}^{\mathsf{H}}(\boldsymbol{\theta}) + \nu\boldsymbol{I}\right)\\ &= \boldsymbol{\Phi}(\boldsymbol{\theta})\boldsymbol{\gamma}\end{aligned}$$

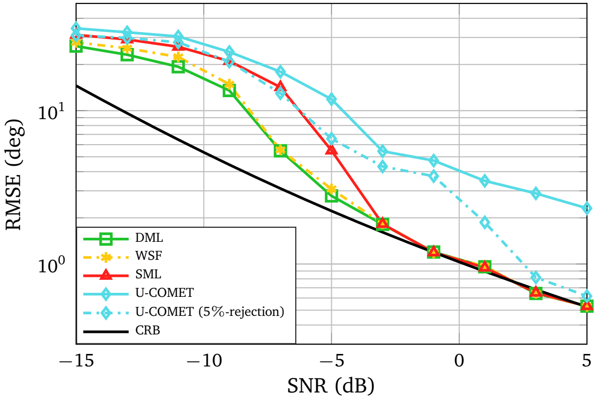
- $\boldsymbol{\Phi} \in \mathbb{C}^{M^2\times(N^2+1)}$ is full-rank matrix depending on the steering matrix $\boldsymbol{A}(\boldsymbol{\theta})$.
- $\boldsymbol{\gamma} \in \mathbb{R}^{(N^2+1)\times 1}$ contains the noise power ν and real-valued entries which characterize the elements on the source covariance matrix $\boldsymbol{P}$.

Relaxed Formulation of COMET

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{\mathrm{COMET}} &= \underset{\boldsymbol{\theta}\in\Theta^N}{\arg\min}\ \min_{\boldsymbol{\gamma}\in\mathbb{C}^{(N^2+1)\times 1}} \left|\left|\boldsymbol{W}\,\mathrm{vec}\left(\hat{\boldsymbol{R}}\right) - \boldsymbol{W}\boldsymbol{\Phi}\left(\boldsymbol{\theta}\right)\boldsymbol{\gamma}\right|\right|_{\mathrm{F}}^2\\ &= \underset{\boldsymbol{\theta}\in\Theta^N}{\arg\min}\ \mathrm{vec}\left(\hat{\boldsymbol{R}}\right)^{\mathsf{H}}\boldsymbol{W}^{\mathsf{H}}\,\boldsymbol{\Pi}^{\perp}_{\boldsymbol{W}\boldsymbol{\Phi}(\boldsymbol{\theta})}\,\boldsymbol{W}\mathrm{vec}\left(\hat{\boldsymbol{R}}\right)\end{aligned}$$

Uncorrelated Source Signals

$$M = 5,\ \boldsymbol{\theta} = [90^\circ, 100^\circ]^\mathsf{T},\ T = 200,\ \rho = 0$$



Correlated Source Signals

$$M = 5,\ \boldsymbol{\theta} = [90^\circ, 100^\circ]^\mathsf{T},\ T = 200,\ \rho = 0.99$$

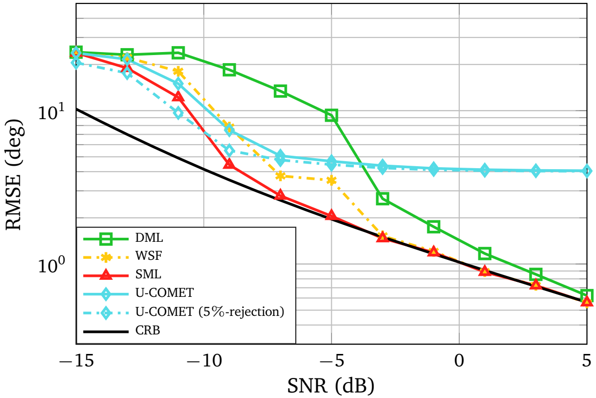


Table of Contents

Table of Contents

Parametric Methods
Summary

General Formulation of Parametric DOA Estimation

$$\boldsymbol{A}\left(\hat{\boldsymbol{\theta}}\right) = \underset{\boldsymbol{A}(\boldsymbol{\theta})\in\mathcal{A}_N}{\arg\min}\, f\left(\boldsymbol{A}\left(\boldsymbol{\theta}\right)\right)$$

Remarks

- Different choices on the cost function $f(\cdot)$ leads to different estimators.
- Generally high computational cost to obtain the global minimum.

Potential Approaches

Original Problem

$$\hat{\boldsymbol{A}} = \min_{\boldsymbol{A} \in \mathcal{A}_N} f(\boldsymbol{A})$$

Relaxation

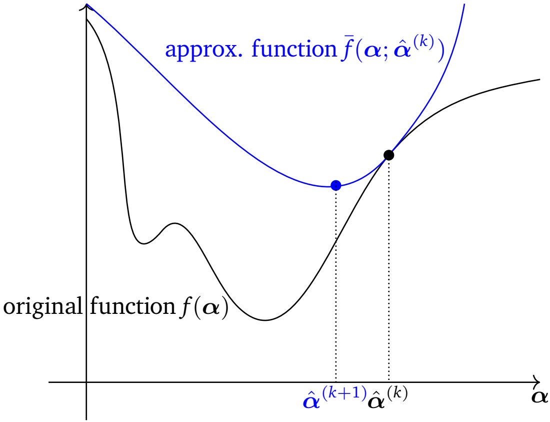
$$\hat{\boldsymbol{A}}_{\text{relaxed}} = \min_{\boldsymbol{A} \in \bar{\mathcal{A}}_N \supset \mathcal{A}_N} f(\boldsymbol{A})$$

Approximation

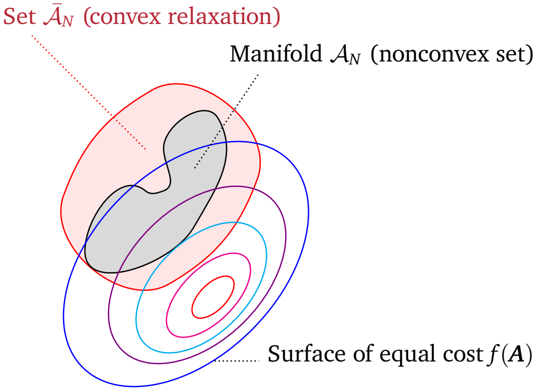
$$\hat{\boldsymbol{A}}_{\text{approx}} = \min_{\boldsymbol{A} \in \mathcal{A}_N} \bar{f}(\boldsymbol{A})$$

- Back-projection is generally required after the relaxation step.
- Possible combination of both relaxation and approximation.

Parametric Methods
Approximation



Parametric Methods
Relaxation



Concept of Relaxation-and-Projection Method

1. Replace the original array manifold $\mathcal{A}_N$ by a relaxed manifold $\bar{\mathcal{A}}_N \supset \mathcal{A}_N$

$$\hat{\boldsymbol{A}} = \underset{\boldsymbol{A}\in\mathcal{A}_N}{\arg\min}\, f(\boldsymbol{A}) \quad \longrightarrow \quad \hat{\boldsymbol{A}}_{\text{relaxed}} = \underset{\boldsymbol{A}\in\bar{\mathcal{A}}_N}{\arg\min}\, f(\boldsymbol{A})\,.$$

2. Project the relaxed estimate $\hat{\boldsymbol{A}}_{\text{relaxed}}$ back to the original array manifold $\mathcal{A}_N$.

Remarks

- The choice on the relaxed array manifold $\bar{\mathcal{A}}_N$ generally depends on the underlying structure of the sensor array.
- Relaxation-and-Projection may, in particular cases, preserve optimality, e.g., in the Extended Invariance Principle (EXIP) [Stoica'89-2].

Table of Contents

Parametric Methods
Root-WSF

For ULA geometries with baseline d the steering matrix is Vandermonde with $z_n = e^{-\mathrm{j}\frac{2\pi}{\lambda} d\cos(\theta_n)}$, i.e.,

$$\boldsymbol{A}(\boldsymbol{\theta}) = \begin{bmatrix} 1 & 1 & \dots & 1 \\ z_1 & z_2 & \cdots & z_N \\ \vdots & \vdots & & \vdots \\ z_1^{M-1} & z_2^{M-1} & \dots & z_N^{M-1} \end{bmatrix} \in \mathbb{C}^{M\times N}.$$

The Root-WSF (RWSF) algorithm uses the Toeplitz reparameterization

$$\boldsymbol{B} = \begin{bmatrix} b_0 & b_1 & \cdots & b_N & 0 & \cdots & 0 \\ 0 & b_0 & b_1 & \cdots & b_N & \ddots & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & b_0 & b_1 & \cdots & b_N \end{bmatrix}^{\mathsf{H}} \in \mathbb{C}^{M\times(M-N)}$$

such that $\boldsymbol{\Pi}^{\perp}_{\boldsymbol{A}(\boldsymbol{\theta})} = \boldsymbol{\Pi}_{\boldsymbol{B}} = \boldsymbol{B}(\boldsymbol{B}^{\mathsf{H}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathsf{H}}$ and $b_n = b^*_{N-n}$, $\mathrm{Re}\{b_0\} = 1$, $\mathrm{Im}\{b_0\} = 0$.

Inserting the reparameterization in the WSF function yields [Stoica'90], [Stoica'90-3],[Kumaresan'82]

$$f_{\text{RWSF}}(\boldsymbol{b}) = \operatorname{Tr}\Big((\boldsymbol{B}^{\mathsf{H}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathsf{H}}\,\hat{\boldsymbol{U}}_{\text{s}}\boldsymbol{W}\hat{\boldsymbol{U}}_{\text{s}}^{\mathsf{H}}\boldsymbol{B}\Big).$$

The RWSF problem is generally solved in three steps.

Step 1: Approximate $(\boldsymbol{B}^{\mathsf{H}}\boldsymbol{B})^{-1} = \boldsymbol{I}_M$

$$\breve{\boldsymbol{b}} = \arg\min_{\boldsymbol{b}} \quad f_{\text{RWSF}}(\boldsymbol{b}) \quad \text{subject to} \quad b_n = b^*_{N-n},\ \operatorname{Re}\{b_0\} = 1,\ \operatorname{Im}\{b_0\} = 0.$$

Step 2: Form matrix $\breve{\boldsymbol{B}}$ from $\breve{\boldsymbol{b}}$ in Step 1 and refine $(\boldsymbol{B}^{\mathsf{H}}\boldsymbol{B})^{-1} = (\breve{\boldsymbol{B}}^{\mathsf{H}}\breve{\boldsymbol{B}})^{-1}$

$$\hat{\boldsymbol{b}} = \arg\min_{\boldsymbol{b}} \quad f_{\text{RWSF}}(\boldsymbol{b}) \quad \text{subject to} \quad b_n = b^*_{N-n},\ \operatorname{Re}\{b_0\} = 1,\ \operatorname{Im}\{b_0\} = 0.$$

Step 3: Compute the roots $\hat{z}_1, \ldots, \hat{z}_N$ of $\hat{b}(z) = \sum_{n=0}^{N} \hat{b}_n z^n = 0$.
Determine DOA estimates as $\hat{\theta}_{n,\text{RWSF}} = \arccos\left(\frac{\lambda}{2\pi d}\arg(\hat{z}_n)\right)$ for $n = 1, \ldots, N$.

Parametric Methods

Root-WSF

Discussion of RWSF in the context of convex relaxation

- The reparameterization allows a (successive) convex approximation and relaxation.
- To see this, note that the conjugate symmetry condition

$$b_n = b^*_{N-n}.$$

  is only a necessary condition for the roots of $b(z)$ to be located on the unit circle (but not a sufficient condition).
- In practice it is not guaranteed that the solutions of the RWSF problem yield roots on the unit circle.
- With the reparameterization $\mathbf{\Pi}_{\boldsymbol{B}}$ the search-space over which the WSF cost function is minimized is increased as compared to the original WSF formulation based on $\mathbf{\Pi}^{\perp}_{\boldsymbol{A}(\boldsymbol{\theta})}$.
- The resulting problems that are solved in each step are convex, hence the reparameterization is a (successive) convex relaxation.

Parametric Methods

Relaxation Based on Geometry Exploitation

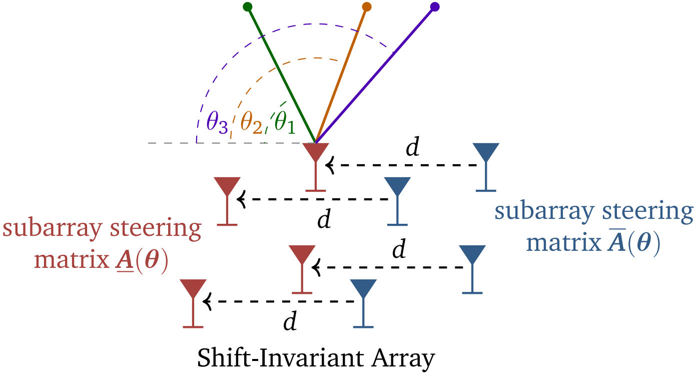


Figure: Antenna array composed of two identical subarrays (subarray 1 in red color) and (subarray 2 in blue color) shifted by baseline d.

Parametric Methods
Relaxation Based on Geometry Exploitation

- ESPRIT (Estimation of Signal Parameters via Rotational Invariance Techniques) is one of the most popular multi-source estimation method. [Roy'86]
- Applicable in shift invariant arrays.
- Search-free technique with simple implementation.

Property

Subarray manifold must not be known (not exploited) in ESPRIT.

Original derivation is based on the algebraic properties of the shift invariant array structure rather than an optimization criteria.

Here: Alternative derivation of ESPRIT in the context of geometry relaxation:
- relate ESPRIT to the aforementioned subspace fitting problem, and
- start the design from relaxation of subarray manifolds.

Parametric Methods
Relaxation Based on Geometry Exploitation

We assume $\frac{M}{2} \geq N$. Given the steering matrix $\underline{\boldsymbol{A}}(\boldsymbol{\theta}) \in \underline{\mathcal{A}}_N$ of the first subarray, the steering matrix $\overline{\boldsymbol{A}}(\boldsymbol{\theta}) \in \overline{\mathcal{A}}_N$ of the second subarray can be expressed as

$$\overline{\boldsymbol{A}}(\boldsymbol{\theta}) = \underline{\boldsymbol{A}}(\boldsymbol{\theta})\boldsymbol{D}(\boldsymbol{\theta}), \quad \boldsymbol{D}(\boldsymbol{\theta}) = \operatorname{diag}\left(e^{-\mathrm{j}\frac{2\pi}{\lambda}d\cos(\theta_1)}, e^{-\mathrm{j}\frac{2\pi}{\lambda}d\cos(\theta_2)}, \cdots, e^{-\mathrm{j}\frac{2\pi}{\lambda}d\cos(\theta_N)}\right).$$

The array steering matrix can be decomposed in subarray responses as

$$\boldsymbol{A}(\boldsymbol{\theta}) = \begin{bmatrix} \underline{\boldsymbol{A}}(\boldsymbol{\theta}) \\ \overline{\boldsymbol{A}}(\boldsymbol{\theta}) \end{bmatrix}.$$

Similarly, let $\boldsymbol{U}_\mathrm{s}$ be partitioned as

$$\boldsymbol{U}_\mathrm{s} = \begin{bmatrix} \underline{\boldsymbol{U}}_\mathrm{s} \\ \overline{\boldsymbol{U}}_\mathrm{s} \end{bmatrix}.$$

Parametric Methods
Relaxation Based on Geometry Exploitation

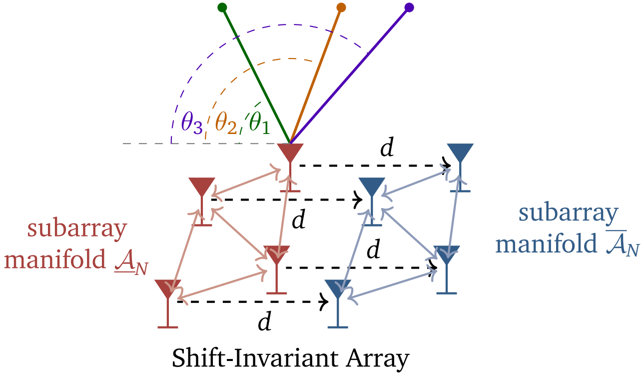


Figure: The subarray displacement (shift) d must be known. $\underline{\mathcal{A}}_N$ and $\overline{\mathcal{A}}_N$ are the manifolds of the identical subarrays.

Parametric Methods
Relaxation Based on Geometry Exploitation

From an optimization perspective ESPRIT can be understood as a subspace matching approach with manifold relaxation.

Recall that $\boldsymbol{A}(\boldsymbol{\theta})$ and $\boldsymbol{U}_\mathrm{s}$ span the same space and consider the variation of the subspace fitting problem

$$f_\mathrm{WSF}(\boldsymbol{\theta}) = \min_{\boldsymbol{A}(\boldsymbol{\theta})\in\mathcal{A}_N} \min_{\boldsymbol{K}\in\mathbb{C}^{N\times N}} \|\hat{\boldsymbol{U}}_\mathrm{s}\boldsymbol{K}^{-1} - \boldsymbol{A}(\boldsymbol{\theta})\|_\mathsf{F}^2$$

which involves a multi-dimensional multi-modal optimization over the manifold $\mathcal{A}_N$:

$$\mathcal{A}_N = \left\{\boldsymbol{A} \in \mathbb{C}^{M\times N} | \, \boldsymbol{A} = [\boldsymbol{a}(\vartheta_1), \ldots, \boldsymbol{a}(\vartheta_N)] \text{ with } 0 \leq \vartheta_1 < \ldots < \vartheta_N < 180^\circ\right\}$$

where we assume for simplicity w.o.l.g. that the first sensor is the reference sensor and $\boldsymbol{e}_1^\mathsf{T}\underline{\boldsymbol{A}} = \mathbf{1}_N^\mathsf{T}$.

Parametric Methods
Relaxation Based on Geometry Exploitation

To make the problem tractable the original array manifold $\mathcal{A}_N$ is replaced by the relaxed manifold $\mathcal{A}_N^{\text{ESPRIT}}$

$$\mathcal{A}_N^{\text{ESPRIT}} = \left\{ \boldsymbol{A} \in \mathbb{C}^{M \times N} \middle| \, \boldsymbol{A} = \begin{bmatrix} \underline{\boldsymbol{A}} \\ \underline{\boldsymbol{A}} \boldsymbol{D}(\boldsymbol{\vartheta}) \end{bmatrix}, \, \underline{\boldsymbol{A}} \in \mathbb{C}^{\frac{M}{2} \times N}, \boldsymbol{e}_1^{\mathsf{T}} \underline{\boldsymbol{A}} = \mathbf{1}_N^{\mathsf{T}}, \boldsymbol{\vartheta} \in \Theta^N \right\}$$

where $\underline{\boldsymbol{A}} \in \mathbb{C}^{\frac{M}{2} \times N}$ is an arbitrary complex matrix and

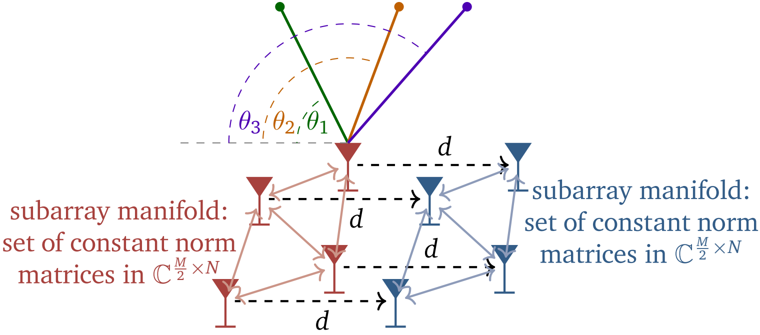
$$\boldsymbol{D}(\boldsymbol{\vartheta}) = \operatorname{diag}\left(e^{-\mathrm{j} \frac{2\pi}{\lambda} d \cos(\vartheta_1)}, e^{-\mathrm{j} \frac{2\pi}{\lambda} d \cos(\vartheta_2)}, \cdots, e^{-\mathrm{j} \frac{2\pi}{\lambda} d \cos(\vartheta_N)} \right).$$

The condition $\boldsymbol{e}_1^{\mathsf{T}} \underline{\boldsymbol{A}} = \mathbf{1}_N^{\mathsf{T}}$ selects w.l.o.g. the first sensor in the array as the reference sensor. Let

$$\mathcal{D}_N = \left\{ \boldsymbol{D} \in \mathbb{S}^{N \times N} \middle| \, \boldsymbol{D}(\boldsymbol{\vartheta}) = \operatorname{diag}\left(e^{-\mathrm{j} \frac{2\pi}{\lambda} d \cos(\vartheta_1)}, \cdots, e^{-\mathrm{j} \frac{2\pi}{\lambda} d \cos(\vartheta_N)} \right), \boldsymbol{\vartheta} \in \Theta^N \right\}$$

denote the corresponding manifold.

Application Example: Multidimensional Frequency Estimation



Shift-Invariant Array

Figure: The subarray displacement (shift) d must be known. The original manifold $\mathcal{A}_N$ of the shift-invariant array is relaxed to manifold $\mathcal{A}_N^{\text{ESPRIT}}$.

Parametric Methods
Relaxation Based on Geometry Exploitation

The subspace fitting problem over manifold $\mathcal{A}_N^{\text{ESPRIT}}$ becomes the ESPRIT problem

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{\text{ESPRIT}} &= \underset{\boldsymbol{A}(\boldsymbol{\theta})\in\mathcal{A}_N^{\text{ESPRIT}}}{\arg\min}\ \min_{\boldsymbol{K}\in\mathbb{C}^{N\times N}} \left\|\hat{\boldsymbol{U}}_{\text{s}}\boldsymbol{K}^{-1} - \boldsymbol{A}(\boldsymbol{\theta})\right\|_{\mathsf{F}}^2 \\ &= \underset{\boldsymbol{\theta}\in\Theta^N}{\arg\min}\ \min_{\boldsymbol{K}\in\mathbb{C}^{N\times N}}\ \min_{\substack{\boldsymbol{e}_1^{\mathsf{T}}\underline{\boldsymbol{A}}=\boldsymbol{1}_N^{\mathsf{T}}\\ \underline{\boldsymbol{A}}\in\mathbb{C}^{(M/2)\times N}}} \left(\left\|\underline{\hat{\boldsymbol{U}}}_{\text{s}}\boldsymbol{K}^{-1} - \underline{\boldsymbol{A}}\right\|_{\mathsf{F}}^2 + \left\|\overline{\hat{\boldsymbol{U}}}_{\text{s}}\boldsymbol{K}^{-1} - \underline{\boldsymbol{A}}\boldsymbol{D}(\boldsymbol{\theta})\right\|_{\mathsf{F}}^2\right) \\ &= \underset{\boldsymbol{\theta}\in\Theta^N}{\arg\min}\ \min_{\boldsymbol{K}\in\mathbb{C}^{N\times N}}\ \min_{\substack{\boldsymbol{e}_1^{\mathsf{T}}\underline{\boldsymbol{A}}=\boldsymbol{1}_N^{\mathsf{T}}\\ \underline{\boldsymbol{A}}\in\mathbb{C}^{(M/2)\times N}}} \left\|[\underline{\hat{\boldsymbol{U}}}_{\text{s}}\boldsymbol{K}^{-1}, \overline{\hat{\boldsymbol{U}}}_{\text{s}}\boldsymbol{K}^{-1}] - \underline{\boldsymbol{A}}\ [\boldsymbol{I}, \boldsymbol{D}(\boldsymbol{\theta})]\right\|_{\mathsf{F}}^2.\end{aligned}$$

The minimizer for the inner optimization problem is the Least-Square (LS) estimator

$$\hat{\underline{\boldsymbol{A}}}_{\text{LS}} = \frac{1}{2}\left(\underline{\hat{\boldsymbol{U}}}_{\text{s}}\boldsymbol{K}^{-1} + \overline{\hat{\boldsymbol{U}}}_{\text{s}}\boldsymbol{K}^{-1}\boldsymbol{D}^*(\boldsymbol{\theta})\right)$$

where a scaling constraint applies to the design of $\boldsymbol{K}$ to ensure $\boldsymbol{e}_1^{\mathsf{T}}\hat{\underline{\boldsymbol{A}}}_{\text{LS}} = \boldsymbol{1}_N^{\mathsf{T}}$ that we drop for simplicity (for reasons that become apparent later).

Inserting $\underline{\hat{\boldsymbol{A}}}_{\mathrm{LS}} = \frac{1}{2}\left(\underline{\hat{\boldsymbol{U}}}_{\mathrm{s}}\boldsymbol{K}^{-1} + \bar{\hat{\boldsymbol{U}}}_{\mathrm{s}}\boldsymbol{K}^{-1}\boldsymbol{D}^{*}(\boldsymbol{\theta})\right)$ back into the relaxed subspace fitting problem yields

$$f_{\mathrm{ESPRIT}}(\boldsymbol{\theta}) = \min_{\boldsymbol{\theta}\in\Theta^N} \min_{\boldsymbol{K}\in\mathbb{C}^{N\times N}} \left\| \underline{\hat{\boldsymbol{U}}}_{\mathrm{s}}\boldsymbol{K}^{-1}\boldsymbol{D}(\boldsymbol{\theta}) - \bar{\hat{\boldsymbol{U}}}_{\mathrm{s}}\boldsymbol{K}^{-1} \right\|_{\mathsf{F}}^2$$
$$= \min_{\boldsymbol{D}\in\mathcal{D}_N} \min_{\boldsymbol{K}\in\mathbb{C}^{N\times N}} \left\| \underline{\hat{\boldsymbol{U}}}_{\mathrm{s}}\boldsymbol{K}^{-1}\boldsymbol{D} - \bar{\hat{\boldsymbol{U}}}_{\mathrm{s}}\boldsymbol{K}^{-1} \right\|_{\mathsf{F}}^2.$$

If we further relax the last problem by replacing the set $\mathcal{D}_N$ over which the variable $\boldsymbol{D}$ is minimized by the set of arbitrary complex diagonal $N \times N$ matrices denoted by $\mathcal{S}^N$ then we obtain the Eigenvalue problem:

$$\hat{\boldsymbol{D}}_{\mathrm{ESPRIT}} = \arg\min_{\boldsymbol{D}\in\mathcal{S}^N} \min_{\boldsymbol{K}\in\mathbb{C}^{N\times N}} \left\| \underline{\hat{\boldsymbol{U}}}_{\mathrm{s}}\boldsymbol{K}^{-1}\boldsymbol{D} - \bar{\hat{\boldsymbol{U}}}_{\mathrm{s}}\boldsymbol{K}^{-1} \right\|_{\mathsf{F}}^2 = \boldsymbol{K}\boldsymbol{\Psi}\boldsymbol{K}^{-1}.$$

where $\boldsymbol{\Psi} = (\underline{\hat{\boldsymbol{U}}}_{\mathrm{s}}^{\mathsf{H}}\underline{\hat{\boldsymbol{U}}}_{\mathrm{s}})^{-1}\underline{\hat{\boldsymbol{U}}}_{\mathrm{s}}^{\mathsf{H}}\bar{\hat{\boldsymbol{U}}}_{\mathrm{s}}$ and $\boldsymbol{K}$ is the matrix that diagonalizes $\boldsymbol{\Psi}$.

Hence the eigenvalues of $\boldsymbol{\Psi}$ form the diagonal element of $\hat{\boldsymbol{D}}_{\mathrm{ESPRIT}}$.

Parametric Methods

Relaxation Based on Geometry Exploitation

To obtain estimates in set $\mathcal{D}^N$ the solution $\hat{\boldsymbol{D}}_{\text{ESPRIT}}$ is projected back to the unit-circle.

To summarize, the LS-ESPRIT algorithm is carried out in the following steps:

Step 1: Compute the eigendecomposition of the sample covariance matrix $\hat{\boldsymbol{R}}$ and obtain the sample signal-subspace $\hat{\boldsymbol{U}}_{\text{s}}$.

Step 2: Form the matrices $\overline{\hat{\boldsymbol{U}}}_{\text{s}}$ and $\underline{\hat{\boldsymbol{U}}}_{\text{s}}$.

Step 3: Compute

$$\hat{\boldsymbol{\Psi}} = (\underline{\hat{\boldsymbol{U}}}_{\text{s}}^{\mathsf{H}} \underline{\hat{\boldsymbol{U}}}_{\text{s}})^{-1} \underline{\hat{\boldsymbol{U}}}_{\text{s}}^{\mathsf{H}} \overline{\hat{\boldsymbol{U}}}_{\text{s}}$$

Step 4: Find the eigenvalues $\lambda_n(\hat{\boldsymbol{\Psi}})$, $n = 1, 2, \ldots, N$ of $\hat{\boldsymbol{\Psi}}$ and determine DOA estimates as $\hat{\theta}_{n,\text{ESPRIT}} = \arccos\left(\frac{\lambda}{2\pi d} \arg\left(\lambda_n(\hat{\boldsymbol{\Psi}})\right)\right)$, for $n = 1, \ldots, N$.

Parametric Methods
Relaxation Based on Geometry Exploitation

Recall the Formulation of LS-ESPRIT

$$\hat{\boldsymbol{\theta}}_{\text{ESPRIT}} = \underset{\boldsymbol{A}(\boldsymbol{\theta}) \in \mathcal{A}_N^{\text{ESPRIT}}}{\arg\min} \min_{\boldsymbol{K} \in \mathbb{C}^{N \times N}} \left\| \hat{\boldsymbol{U}}_{\text{s}} \boldsymbol{K}^{-1} - \boldsymbol{A}(\boldsymbol{\theta}) \right\|_{\mathsf{F}}^2$$

Formulation of Total Least Squares ESPRIT

$$\hat{\boldsymbol{\theta}}_{\text{TLS}-\text{ESPRIT}} = \underset{\boldsymbol{A}(\boldsymbol{\theta}) \in \mathcal{A}_N^{\text{ESPRIT}}}{\arg\min} \min_{\boldsymbol{K} \in \mathbb{C}^{N \times N}} \left\| \hat{\boldsymbol{U}}_{\text{s}} - \boldsymbol{A}(\boldsymbol{\theta}) \boldsymbol{K} \right\|_{\mathsf{F}}^2$$

- Both LS-ESPRIT and TLS-ESPRIT technique are search-free approaches.
- The subarray manifold must not be known.

Parametric Methods

Relaxation Based on Geometry Exploitation

In the ESPRIT algorithm the subarrays can also overlap, such as in the case of ULA:

$$\boldsymbol{A}(\boldsymbol{\theta}) = \begin{bmatrix} 1 & 1 & \dots & 1 \\ e^{-\mathrm{j}\frac{2\pi}{\lambda}d\cos(\theta_1)} & e^{-\mathrm{j}\frac{2\pi}{\lambda}d\cos(\theta_2)} & \dots & e^{-\mathrm{j}\frac{2\pi}{\lambda}d\cos(\theta_N)} \\ \vdots & \vdots & & \vdots \\ e^{-\mathrm{j}\frac{2\pi}{\lambda}(M-1)d\cos(\theta_1)} & e^{-\mathrm{j}\frac{2\pi}{\lambda}(M-1)d\cos(\theta_2)} & \dots & e^{-\mathrm{j}\frac{2\pi}{\lambda}(M-1)d\cos(\theta_N)} \end{bmatrix}$$

with partition $\overline{\boldsymbol{A}}(\boldsymbol{\theta})$ and $\underline{\boldsymbol{A}}(\boldsymbol{\theta})$ denoting the matrices with eliminated first and last row, respectively.

Parametric Methods
Relaxation Based on Geometry Exploitation

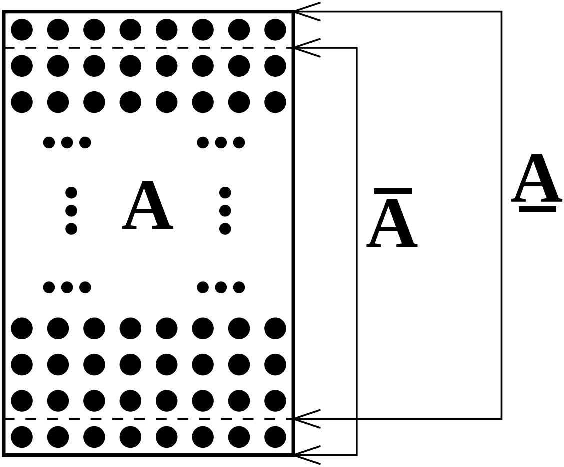


Table of Contents

Sparse Relaxation Techniques
ℓ_1-relaxation Techniques

To avoid the difficulty of the multi-dimensional multimodal optimization over a nonconvex manifold $\mathcal{A}_N$ the compressed sensing (CS) approach is to sample the field of view Ω on a fine grid of DOAs

$$\tilde{\boldsymbol{\theta}} = [\tilde{\theta}_1, \tilde{\theta}_2, \ldots, \tilde{\theta}_K]^\mathsf{T} \in \Theta^K$$

with $K \gg N$ constructing an fixed overcomplete (fat) dictionary (sensing) matrix

$$\tilde{\boldsymbol{A}} = \boldsymbol{A}(\tilde{\boldsymbol{\theta}}) \in \mathcal{A}_K.$$

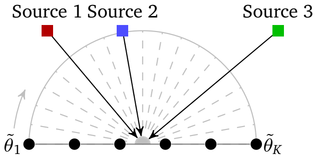
In the following we assume for simplicity that the true source DoAs in vector $\boldsymbol{\theta}$ lie on the grid, hence

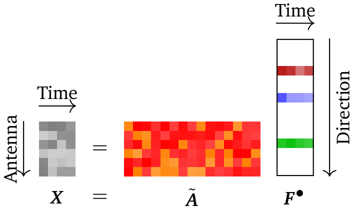
$$\theta_n \in \tilde{\Theta} = \{\tilde{\theta}_1, \ldots, \tilde{\theta}_K\} \text{ for } n = 1, \ldots, N.$$

- Observe T snapshots of N source signals impinging on array of M sensors
- Sparse representation of $M \times T$ measurement matrix

$$\boldsymbol{X} = \tilde{\boldsymbol{A}}\boldsymbol{F}^{\bullet} + \boldsymbol{N}$$

with
- $M \times K$ sensing matrix $\tilde{\boldsymbol{A}} = \left[\boldsymbol{a}(\tilde{\theta}_1), \ldots, \boldsymbol{a}(\tilde{\theta}_K)\right]$
- $K \times T$ joint sparse signal matrix $\boldsymbol{F}^{\bullet} = [\boldsymbol{f}^{\bullet}(1), \ldots, \boldsymbol{f}^{\bullet}(T)]$
- $M \times T$ sensor noise matrix $\boldsymbol{N} = [\boldsymbol{n}(1), \ldots, \boldsymbol{n}(T)]$.

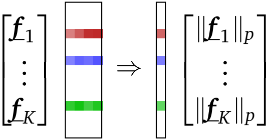




Sparse Relaxation Techniques
ℓ_1-relaxation Techniques

- $\ell_{p,q}$ mixed-norm of matrix $\boldsymbol{F} = \left[\boldsymbol{f}_1, \ldots, \boldsymbol{f}_K\right]^{\mathsf{T}}$:

$$\|\boldsymbol{F}\|_{p,q} = \left(\sum_{k=1}^{K} \|\boldsymbol{f}_k\|_p^q\right)^{\frac{1}{q}}.$$



- Nonlinear coupling of elements in row vectors $\boldsymbol{f}_k$ by ℓ_p-norm.
- Ideal for sparse reconstruction: $\ell_{p,0}$-norm with $p \geq 2$.

Sparse Relaxation Techniques

ℓ_1-relaxation Techniques

With dictionary $\tilde{\boldsymbol{A}}$ the LS fitting problem can be equivalently reformulated as

$$\begin{aligned} \min_{\boldsymbol{F}^\bullet \in \mathbb{C}^{K \times T}} \quad & \|\boldsymbol{X} - \tilde{\boldsymbol{A}} \boldsymbol{F}^\bullet\|_\mathsf{F}^2 \\ \text{subject to} \quad & \|\boldsymbol{F}^\bullet\|_{p,0} = N. \end{aligned}$$

- Note, that the sensing matrix $\tilde{\boldsymbol{A}}$ is fat, hence the equation $\boldsymbol{X} = \tilde{\boldsymbol{A}} \boldsymbol{F}^\bullet$ has infinitely many exact solutions.
- Hence, in the $\ell_{p,0}$-constrained problem we search for an N-row sparse solution that minimizes the fitting error.
- Dictionary $\tilde{\boldsymbol{A}}$ is constant, hence the optimization over manifold $\mathcal{A}_N$ has been avoided in the problem reformulation.
- However, the $\ell_{p,0}$-constraint is still nonconvex and combinatorial.

Sparse Relaxation Techniques
ℓ_1-relaxation Techniques

To solve the problem Lagrangian relaxation can be applied. The corresponding dual function is

$$d(\lambda) = \min_{\boldsymbol{F}^\bullet \in \mathbb{C}^{K \times T}} \frac{1}{2} \|\boldsymbol{X} - \tilde{\boldsymbol{A}} \boldsymbol{F}^\bullet\|_{\mathsf{F}}^2 + \lambda \|\boldsymbol{F}^\bullet\|_{p,0} - \lambda N$$

for $\lambda \geq 0$.

- The Lagrange multiplier λ marks the cost associated with the violation of the $\ell_{p,0}$ constraint.
- The Lagrangian minimization problem provides a lower bound for the objective function value of the $\ell_{p,0}$ constrained LS matching problem above that is tight for an appropriate choice of λ.
- We will later discuss a practical procedure for finding a suitable λ.
- The relaxed problem is still nonconvex due to the nonconvexity of the $\ell_{p,0}$ mixed-norm, hence convex approximation techniques can be applied.

Sparse Relaxation Techniques
ℓ_1-relaxation Techniques

- A common convex approximation of the $\ell_{p,0}$-pseudo-norm that is known to promote sparse solutions is the $\ell_{p,1}$-norm. This approximation is commonly termed ℓ_1-norm relaxation,...
- ... even though depending on the choice of λ it may not necessarily represent a relaxation of the the ℓ_0 constrained LS matching problem above in the optimization relaxation sense (the lower bound property is not necessarily satisfied).
- Further, for fixed λ dropping constant terms we obtain the ℓ_1 regularized LS problem also known as LASSO [Yang'18].

$$\hat{\boldsymbol{F}}^{\bullet}_{\lambda} = \min_{\boldsymbol{F}^{\bullet} \in \mathbb{C}^{K \times T}} \frac{1}{2} \|\boldsymbol{X} - \tilde{\boldsymbol{A}} \boldsymbol{F}^{\bullet}\|_{\mathsf{F}}^2 + \lambda \|\boldsymbol{F}^{\bullet}\|_{p,1}$$

where $\lambda \geq 0$.

Sparse Relaxation Techniques
ℓ_1-relaxation Techniques

Multiple Snapshot Problem – Mixed-Norm Regularization

- $\ell_{2,1}$ Mixed-norm minimization [Malioutov'05], [Yuan'05]

$$\min_{\boldsymbol{F}^{\bullet}} \frac{1}{2} \left\| \boldsymbol{X} - \tilde{\boldsymbol{A}}\boldsymbol{F}^{\bullet} \right\|_{\mathsf{F}}^2 + \lambda \left\| \boldsymbol{F}^{\bullet} \right\|_{2,1}.$$

- Problem: For large number of snapshots N or large number of candidate frequencies K the problem becomes computationally intractable.
- Heuristic approach: Reduction of the dimension of measurement matrix $\boldsymbol{X}$ by ℓ_1-SVD and adaptive grid refinement,

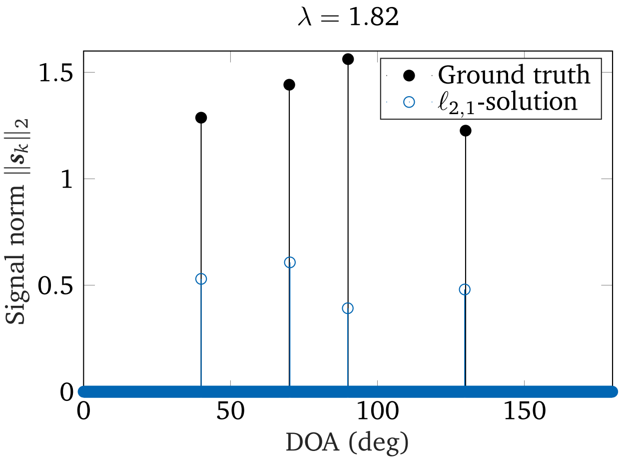
Choice of regularization parameter λ

- It can be proven that with the choice

$$\lambda \geq \lambda_{\max} = \max_{k=1,\dots,K} \|\tilde{\boldsymbol{a}}_k^{\mathsf{H}} \boldsymbol{X}\|_2$$

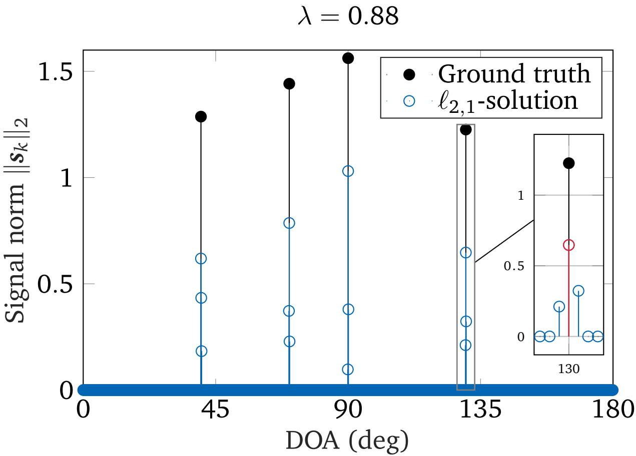
  the all zero matrix $\hat{\boldsymbol{F}}_{\lambda}^{\bullet} = \hat{\boldsymbol{F}}_{\lambda_{\max}}^{\bullet} = \boldsymbol{0}_{K \times T}$ is always the optimal solution of the $\ell_{2,1}$ mixed-norm problem.
- Hence $\lambda_{\max}$ provides an upper bound for the choice of λ.
- The bisection algorithm can be used to find the smallest value of $\lambda_{N,\min}$ for which an N-sparse solution vector $\hat{\boldsymbol{F}}_{\lambda_{N,\min}}^{\bullet}$ is obtained, i.e., $\|\hat{\boldsymbol{F}}_{\lambda_{N,\min}}^{\bullet}\|_{2,0} = N$.

Sparse Relaxation Techniques
ℓ_1-relaxation Techniques



Sparse Relaxation Techniques

ℓ_1-relaxation Techniques



- If the solution is not N-row sparse, choose the N-largest local maxima.

Sparse Relaxation Techniques
Equivalent Formulation

SPARROW Formulation [Steffen'16]

The $\ell_{2,1}$ mixed-norm minimization problem

$$\min_{\boldsymbol{F}^\bullet \in \mathbb{C}^{K \times T}} \frac{1}{2} \left\| \boldsymbol{X} - \tilde{\boldsymbol{A}} \boldsymbol{F}^\bullet \right\|_\mathsf{F}^2 + \lambda \sqrt{T} \, \| \boldsymbol{F}^\bullet \|_{2,1}$$

is equivalent to SPARse ROW-norm reconstruction (SPARROW)

$$\min_{\boldsymbol{G} \in \mathbb{D}_+^K} \mathrm{Tr}\Big((\tilde{\boldsymbol{A}} \boldsymbol{G} \tilde{\boldsymbol{A}}^\mathsf{H} + \lambda \boldsymbol{I})^{-1} \hat{\boldsymbol{R}} \Big) + \mathrm{Tr}\,(\boldsymbol{G}),$$

with $\hat{\boldsymbol{R}} = \boldsymbol{X}\boldsymbol{X}^\mathsf{H}/T$ and minimizers $\hat{\boldsymbol{F}}^\bullet = [\hat{\boldsymbol{f}}_1^\bullet \ldots, \hat{\boldsymbol{f}}_K^\bullet]^\mathsf{T}$ and $\hat{\boldsymbol{G}} = \mathrm{diag}(\hat{g}_1, \ldots, \hat{g}_K)$ as

$$\hat{\boldsymbol{F}}^\bullet = \hat{\boldsymbol{G}} \tilde{\boldsymbol{A}}^\mathsf{H} (\tilde{\boldsymbol{A}} \hat{\boldsymbol{G}} \tilde{\boldsymbol{A}}^\mathsf{H} + \lambda \boldsymbol{I})^{-1} \boldsymbol{X} \quad \text{and} \quad \hat{g}_k = \| \hat{\boldsymbol{f}}_k^\bullet \|_2 / \sqrt{T} \quad \text{for } k = 1, \ldots, K.$$

Sparse Relaxation Techniques
Equivalent Formulation

- SPARROW formulation

$$\min_{\boldsymbol{G}\in\mathbb{D}_+^K} \operatorname{Tr}\left((\tilde{\boldsymbol{A}}\boldsymbol{G}\tilde{\boldsymbol{A}}^{\mathsf{H}} + \lambda\boldsymbol{I})^{-1}\hat{\boldsymbol{R}}\right) + \operatorname{Tr}(\boldsymbol{G}).$$

- SDP implementation for oversampled case $T > M$

$$\min_{\boldsymbol{G}\in\mathbb{D}_+^K, \boldsymbol{U}_M} \operatorname{Tr}(\boldsymbol{U}_M\hat{\boldsymbol{R}}) + \operatorname{Tr}(\boldsymbol{G})$$

$$\text{subject to} \quad \begin{bmatrix} \boldsymbol{U}_M & \boldsymbol{I}_M \\ \boldsymbol{I}_M & \tilde{\boldsymbol{A}}\boldsymbol{G}\tilde{\boldsymbol{A}}^{\mathsf{H}} + \lambda\boldsymbol{I}_M \end{bmatrix} \succeq \boldsymbol{0} \quad \Leftrightarrow \quad \boldsymbol{U}_M \succeq \left(\tilde{\boldsymbol{A}}\boldsymbol{G}\tilde{\boldsymbol{A}}^{\mathsf{H}} + \lambda\boldsymbol{I}_M\right)^{-1}.$$

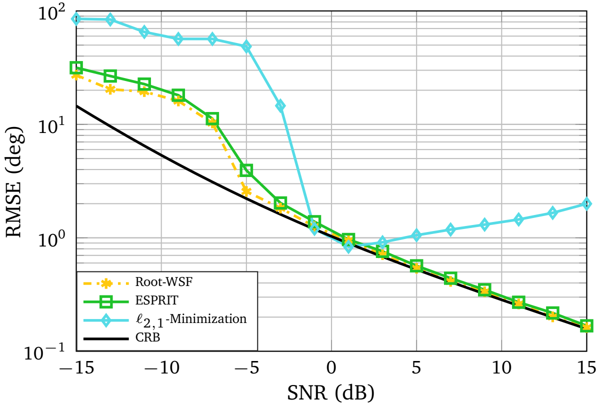
- SDP implementation for undersampled case $N \leq M$

$$\min_{\boldsymbol{G}\in\mathbb{D}_+^K, \boldsymbol{U}_T} \frac{1}{T}\operatorname{Tr}(\boldsymbol{U}_T) + \operatorname{Tr}(\boldsymbol{G})$$

$$\text{subject to} \quad \begin{bmatrix} \boldsymbol{U}_T & \boldsymbol{X}^{\mathsf{H}} \\ \boldsymbol{X} & \tilde{\boldsymbol{A}}\boldsymbol{G}\tilde{\boldsymbol{A}}^{\mathsf{H}} + \lambda\boldsymbol{I}_M \end{bmatrix} \succeq \boldsymbol{0} \quad \Leftrightarrow \quad \boldsymbol{U}_T \succeq \boldsymbol{X}^{\mathsf{H}}\left(\tilde{\boldsymbol{A}}\boldsymbol{G}\tilde{\boldsymbol{A}}^{\mathsf{H}} + \lambda\boldsymbol{I}_M\right)^{-1}\boldsymbol{X}.$$

Uncorrelated Source Signals

$$M = 5,\ \boldsymbol{\theta} = [90^\circ, 100^\circ]^\mathsf{T},\ T = 200,\ \rho = 0.99,\ \lambda = \sqrt{\nu M T \log M}$$



Correlated Source Signals

$$M = 5,\ \boldsymbol{\theta} = [90^\circ, 100^\circ]^\mathsf{T},\ T = 200,\ \rho = 0.99,\ \lambda = \sqrt{\nu MT \log M}$$

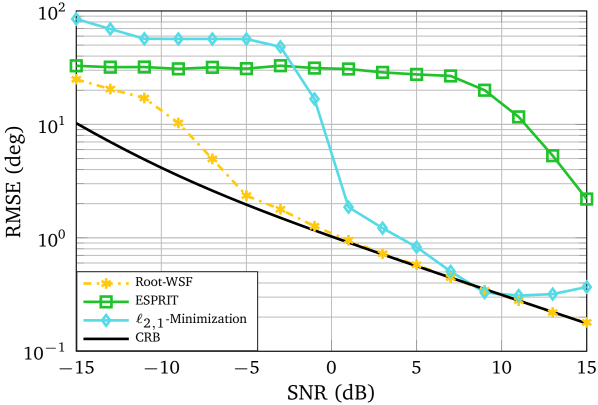
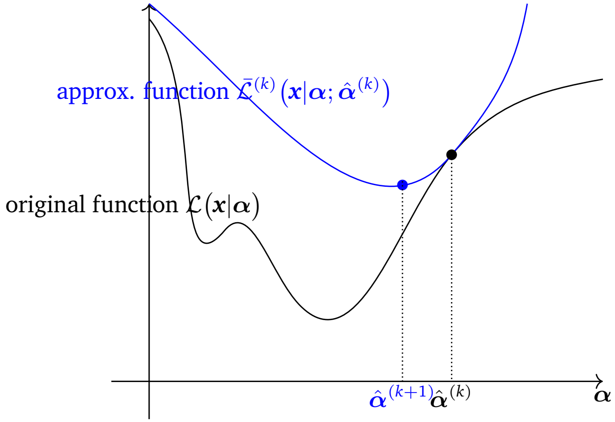


Table of Contents

Approximation Methods
Expectation-Maximization

- Multi-source criteria such as ML achieve excellent threshold and asymptotic estimation performance.
- Full N-dimensional search required.
- Prohibitive complexity for scenarios where $N > 3$.
- Approximation techniques such as Alternating Projection, Block Coordinate Descent, viable options for local convergence.
- Majorization-minimization (MM) approach is an iterative optimization technique.
- Original optimization problem approximated by a sequence of upper bound problems.
- The approximate problems much easier to solve than the original problem (e.g. closed form).



approx. function $\bar{\mathcal{L}}^{(k)}(\boldsymbol{x}|\boldsymbol{\alpha};\hat{\boldsymbol{\alpha}}^{(k)})$
original function $\mathcal{L}(\boldsymbol{x}|\boldsymbol{\alpha})$
$\hat{\boldsymbol{\alpha}}^{(k+1)}$
$\hat{\boldsymbol{\alpha}}^{(k)}$
$\boldsymbol{\alpha}$

ML problem:

$$\hat{\boldsymbol{\alpha}}_{\text{ML}} = \arg\min_{\boldsymbol{\alpha}} \mathcal{L}\big(\boldsymbol{x}|\boldsymbol{\alpha}\big).$$

Approximate problem at point $\hat{\boldsymbol{\alpha}}^{(k)}$ in iteration k:

$$\hat{\boldsymbol{\alpha}}^{(k+1)} = \arg\min_{\boldsymbol{\alpha}} \bar{\mathcal{L}}^{(k)}\big(\boldsymbol{x}|\boldsymbol{\alpha};\hat{\boldsymbol{\alpha}}^{(k)}\big)$$

where the approximate function $\bar{\mathcal{L}}^{(k)}\big(\boldsymbol{x}|\boldsymbol{\alpha};\hat{\boldsymbol{\alpha}}^{(k)}\big)$ is chosen such that it satisfies

- upper bound property:

$$\bar{\mathcal{L}}^{(k)}\big(\boldsymbol{x}|\boldsymbol{\alpha};\hat{\boldsymbol{\alpha}}^{(k)}\big) \geq \mathcal{L}\big(\boldsymbol{x}|\boldsymbol{\alpha}\big), \quad \forall \boldsymbol{\alpha}$$

- tightness at $\hat{\boldsymbol{\alpha}}^{(k)}$:

$$\bar{\mathcal{L}}^{(k)}\big(\boldsymbol{x}|\hat{\boldsymbol{\alpha}}^{(k)};\hat{\boldsymbol{\alpha}}^{(k)}\big) = \mathcal{L}\big(\boldsymbol{x}|\hat{\boldsymbol{\alpha}}^{(k)}\big).$$

Approximation Methods
Expectation-Maximization

- Expectation-maximization (EM) algorithm [Miller'90] [Dempster'77] is a special case of the MM algorithm [Hunter'04], [Luo'16].
- Unobserved data $\boldsymbol{y}$ only available through mapping $\boldsymbol{x} = \mathcal{T}(\boldsymbol{y})$, hence given $\boldsymbol{y}$ the observed data $\boldsymbol{x}$ is fully determined.
- $f(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\alpha})$ is conditional pdf of observations $\boldsymbol{x}$ given unobserved data $\boldsymbol{y}$ with parameterization $\boldsymbol{\alpha}$.
- $f(\boldsymbol{y}|\boldsymbol{\alpha})$ is pdf of unobserved data $\boldsymbol{y}$ with parameterization $\boldsymbol{\alpha}$.
- In the EM algorithm the negative likelihood is approximated by Jensen's inequality

$$\begin{aligned}\mathcal{L}\left(\boldsymbol{x}|\boldsymbol{\alpha}\right) &= -\ln \mathrm{E}_{\boldsymbol{y}|\boldsymbol{\alpha}}\left(f(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\alpha})\right) \\ &\leq -\mathrm{E}_{\boldsymbol{y}|\boldsymbol{x}, \hat{\boldsymbol{\alpha}}^{(k)}}\left(\ln\left(f(\boldsymbol{y}|\boldsymbol{\alpha})\right)\right) + \text{constant} \triangleq \bar{\mathcal{L}}^{(k)}\left(\boldsymbol{x}|\boldsymbol{\alpha}; \hat{\boldsymbol{\alpha}}^{(k)}\right).\end{aligned}$$

Approximation Methods
Expectation-Maximization

- Consider example of DML signal model with known noise variance ν

$$\boldsymbol{x}(t) = \sum_{n=1}^{N} \boldsymbol{a}(\theta_n) s_n(t) + \boldsymbol{n}(t)$$

where $\boldsymbol{A} = [\boldsymbol{a}(\theta_1), \ldots, \boldsymbol{a}(\theta_N)] \in \mathcal{A}_N$ and $\boldsymbol{n}(t) \sim \mathcal{N}_\mathsf{C}\big(\mathbf{0}_M, \nu \boldsymbol{I}_M\big)$.

- Define unobserved data $\boldsymbol{y}^\mathsf{T}(t) = [\boldsymbol{y}_1^\mathsf{T}(t), \ldots, \boldsymbol{y}_N^\mathsf{T}(t)]$ as individual source contributions

$$\boldsymbol{y}_n(t) = \boldsymbol{a}(\theta_n) s_n(t) + \boldsymbol{n}_n(t), \quad n = 1, \ldots, N$$

with i.i.d. $\boldsymbol{n}_n(t) \sim \mathcal{N}_\mathsf{C}\big(\mathbf{0}_{M\times 1}, \nu_n \boldsymbol{I}_M\big)$ and $\sum_{n=1}^{N} \nu_n = \nu$.

- Then

$$\boldsymbol{x}(t) = \sum_{n=1}^{N} \boldsymbol{y}_n(t) = \sum_{n=1}^{N} \boldsymbol{a}(\theta_n) s_n(t) + \boldsymbol{n}(t), \quad \text{where} \quad \boldsymbol{n}(t) = \sum_{n=1}^{N} \boldsymbol{n}_n(t).$$

Approximation Methods
Expectation-Maximization

Expectation Step

At point $\hat{\boldsymbol{\alpha}}^{(k)} = [\hat{\boldsymbol{\theta}}^{(k)\mathsf{T}}, \hat{\boldsymbol{s}}^{(k)\mathsf{T}}]^\mathsf{T}$ in iteration k, the approximate upper bound function can be characterized as

$$\bar{\mathcal{L}}^{(k)}(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{s} | \hat{\boldsymbol{\theta}}^{(k)}, \hat{\boldsymbol{s}}^{(k)}) \propto \sum_{n=1}^{N} \mathrm{E}_{\boldsymbol{y}_n | \boldsymbol{x}, \hat{\boldsymbol{\alpha}}^{(k)}} \Big(\ln \big(f(\boldsymbol{y}_n | \boldsymbol{\alpha}) \big) \Big)$$

$$\propto - \sum_{n=1}^{N} \Big\| \underbrace{\boldsymbol{a}(\hat{\theta}_n^{(k)}) \hat{s}_n^{(k)} - \frac{1}{N} \Big(\boldsymbol{x} - \boldsymbol{A}(\hat{\boldsymbol{\theta}}^{(k)}) \hat{\boldsymbol{s}}^{(k)} \Big)}_{\hat{\boldsymbol{y}}_n^{(k)}(t)} - \boldsymbol{a}(\theta_n) s_n \Big\|^2$$

where we omitted constant terms.

Maximization Step

$$\Big(\hat{\theta}_n^{(k+1)}, \hat{s}_n^{(k+1)} \Big) = \underset{\theta_n, s_n(1), \ldots, s_n(T)}{\arg\min} \sum_{t=1}^{T} \Big\| \boldsymbol{a}(\theta_n) s_n(t) - \hat{\boldsymbol{y}}_n^{(k)}(t) \Big\|^2, \quad \text{for } n = 1, \ldots, N.$$

Solved in parallel or sequentially. Each subproblem is simple to solve.

Table of Contents

Table of Contents

Single-source Approximation Techniques
Concept

Suboptimal solutions of the DOA estimation problem can be obtained by adopting the Single-source Approximation.

Recall the General DOA Estimation Problem

$$\boldsymbol{A}\left(\hat{\boldsymbol{\theta}}\right) = \underset{\boldsymbol{A}\left(\boldsymbol{\theta}\right) \in \mathcal{A}_N}{\arg\min}\ f\left(\boldsymbol{A}\left(\boldsymbol{\theta}\right)\right)$$

Single-source Approximation

Spectral sweep to find the N deepest local minima $\hat{\boldsymbol{\theta}} = \left[\hat{\theta}_1, \ldots, \hat{\theta}_N\right]^{\mathsf{T}}$ of $f\left(\boldsymbol{a}\left(\theta\right)\right)$

$$\boldsymbol{A}\left(\hat{\boldsymbol{\theta}}\right) = {}^{N}\underset{\boldsymbol{a}(\theta) \in \mathcal{A}_1}{\arg\min} f\left(\boldsymbol{a}(\theta)\right).$$

Interpretation: The cost function measures the goodness-of-fit under the assumption of only one source signal located at the candidate DOA $\theta \in \Theta$.

Original Derivation

- Output power of the receive signal $\boldsymbol{x}(t)$ after spatial filtering with the beamforming vector $\boldsymbol{w}(\theta)$

$$P(\theta) = \mathbb{E}\left\{\left|\boldsymbol{w}^{\mathsf{H}}(\theta)\boldsymbol{x}(t)\right|^2\right\}$$
$$= \boldsymbol{w}^{\mathsf{H}}(\theta)\boldsymbol{R}\boldsymbol{w}(\theta).$$

- In practice, the true covariance matrix $\boldsymbol{R}$ of the receive signal $\boldsymbol{x}(t)$ is not available and therefore replaced by the sample covariance matrix $\hat{\boldsymbol{R}}$

$$\hat{P}(\theta) = \frac{1}{T}\sum_{t=1}^{T}\left|\boldsymbol{w}^{\mathsf{H}}(\theta)\boldsymbol{x}(t)\right|^2$$
$$= \boldsymbol{w}^{\mathsf{H}}(\theta)\hat{\boldsymbol{R}}\boldsymbol{w}(\theta).$$

Single-source Approximation Techniques

Conventional Beamformer

Beamformer Vector

$$\boldsymbol{w}_{\mathrm{CBF}}(\theta) = \frac{\boldsymbol{a}(\theta)}{||\boldsymbol{a}(\theta)||}$$

Conventional Beamforming Estimator [Bartlett'48]

Find the N highest local maxima of the beamformer spectrum

$$\hat{P}_{\mathrm{CBF}}(\theta) = \frac{\boldsymbol{a}^{\mathsf{H}}(\theta)\hat{\boldsymbol{R}}\boldsymbol{a}(\theta)}{||\boldsymbol{a}(\theta)||^2}.$$

Interpretation

- $\boldsymbol{w}_{\mathrm{CBF}}(\theta)$ can be considered as a spatially matched filter that maximizes the power impinging on the sensor array from the direction θ.

Single-source Approximation Techniques

Conventional Beamformer

Alternative Derivation: Starting from the Covariance Matrix $\boldsymbol{R}$

$$\boldsymbol{R} = \boldsymbol{A}\boldsymbol{P}\boldsymbol{A}^{\mathsf{H}} + \nu \boldsymbol{I}$$

Single-source approximation of Covariance Fitting Problem

$$\begin{aligned}\hat{\sigma}_{\mathrm{s}}^2 &= \arg\min_{\sigma_{\mathrm{s}}^2} \left|\left|\hat{\boldsymbol{R}} - \sigma_{\mathrm{s}}^2 \boldsymbol{a}\boldsymbol{a}^{\mathsf{H}}\right|\right|_{\mathrm{F}}^2 \\ &= \frac{\boldsymbol{a}^{\mathsf{H}}\hat{\boldsymbol{R}}\boldsymbol{a}}{\left(\boldsymbol{a}^{\mathsf{H}}\boldsymbol{a}\right)^2}\end{aligned}$$

- Conventional beamformer spectrum measures the power impinging at the sensor array from the direction $\boldsymbol{a} = \boldsymbol{a}(\theta)$.
- Disadvantage: limited angular resolution.

Single-source Approximation Techniques
Capon Beamformer

Design of the Capon beamformer

For each direction $\boldsymbol{a} = \boldsymbol{a}(\theta)$, find the beamformer vector $\boldsymbol{w} = \boldsymbol{w}(\theta)$ such that

- the power from the direction $\boldsymbol{a}$ is maintained
- the power from remaining directions is suppressed as much as possible.

Optimization Problem

$$\min_{\boldsymbol{w}} \boldsymbol{w}^{\mathsf{H}} \hat{\boldsymbol{R}} \boldsymbol{w}$$
$$\text{subject to } \boldsymbol{w}^{\mathsf{H}} \boldsymbol{a} = 1$$

- Also known as Minimum Variance Distortionless Response beamformer.
- Optimal beamformer vector $\boldsymbol{w}_{\text{Capon}} = \dfrac{\hat{\boldsymbol{R}}^{-1}\boldsymbol{a}}{\boldsymbol{a}^{\mathsf{H}}\hat{\boldsymbol{R}}^{-1}\boldsymbol{a}}$.

Single-source Approximation Techniques
Capon Beamformer

Capon spectrum [Capon'66]

$$\hat{P}_{\text{Capon}}(\theta) = \boldsymbol{w}_{\text{Capon}}^{\mathsf{H}}(\theta)\hat{\boldsymbol{R}}\boldsymbol{w}_{\text{Capon}}(\theta) = \frac{1}{\boldsymbol{a}^{\mathsf{H}}(\theta)\hat{\boldsymbol{R}}^{-1}\boldsymbol{a}(\theta)}$$

- Estimate the DOAs $\hat{\boldsymbol{\theta}}$ from the N highest peaks of $\hat{P}_{\text{Capon}}(\theta)$.
- Higher resolution capability than the conventional beamformer.
- Applicable if the sample covariance matrix $\hat{\boldsymbol{R}}$ is full rank.
- Values of Capon peaks are roughly proportional to the signal power of the sources.

Single-source Approximation Techniques
Capon Beamformer

Recall the Conventional Beamfomer

$$\hat{\sigma}_{\mathrm{s}}^2 = \arg\min_{\sigma_{\mathrm{s}}^2} \left|\left|\hat{\boldsymbol{R}} - \sigma_{\mathrm{s}}^2 \boldsymbol{a}\boldsymbol{a}^{\mathsf{H}}\right|\right|_{\mathrm{F}}^2$$

Alternative Formulation of the Capon Spectrum

$$\hat{\sigma}_{\mathrm{s}}^2 = \arg\min_{\sigma_{\mathrm{s}}^2} \left|\left|\hat{\boldsymbol{R}} - \sigma_{\mathrm{s}}^2 \boldsymbol{a}\boldsymbol{a}^{\mathsf{H}}\right|\right|_{\mathrm{F}}^2$$
$$\text{subject to } \hat{\boldsymbol{R}} - \sigma_{\mathrm{s}}^2 \boldsymbol{a}\boldsymbol{a}^{\mathsf{H}} \succeq \mathbf{0}$$

Remarks

- Both formulations are based on covariance fitting criteria under single-source approximation.
- Constraint in the Capon formulation prevents the residual matrix to be indefinite.

Single-source Approximation Techniques
MUSIC

Recall the Eigendecomposition of the Covariance Matrix $\boldsymbol{R}$

$$\boldsymbol{R} = \boldsymbol{A}\boldsymbol{P}\boldsymbol{A}^{\mathsf{H}} + \nu \boldsymbol{I} = \boldsymbol{U}_{\mathrm{s}}\boldsymbol{\Lambda}_{\mathrm{s}}\boldsymbol{U}_{\mathrm{s}}^{\mathsf{H}} + \nu \boldsymbol{U}_{\mathrm{n}}\boldsymbol{U}_{\mathrm{n}}^{\mathsf{H}}$$

- Assumption: Non-coherent source signals.
- Key observation: $\boldsymbol{U}_{\mathrm{n}}^{\mathsf{H}}\boldsymbol{a}(\theta) = \boldsymbol{0}$ iff θ coincides with one of the true DOAs $\boldsymbol{\theta}$.

MUSIC Pseudo-spectrum [Schmidt'79]

$$\hat{P}_{\mathrm{MUSIC}}(\theta) = \frac{1}{\left\| \hat{\boldsymbol{U}}_{\mathrm{n}}^{\mathsf{H}}\boldsymbol{a}(\theta) \right\|_2^2} = \frac{1}{\boldsymbol{a}^{\mathsf{H}}(\theta)\hat{\boldsymbol{U}}_{\mathrm{n}}\hat{\boldsymbol{U}}_{\mathrm{n}}^{\mathsf{H}}\boldsymbol{a}(\theta)}$$

- MUSIC pseudo-spectrum is inversely proportional to the distance between the steering vector $\boldsymbol{a}(\theta)$ and the sample noise subspace $\mathrm{span}(\boldsymbol{U}_{\mathrm{n}})$.

Single-source Approximation Techniques
MUSIC

Recall the WSF Estimator [Viberg'91]

$$\hat{\boldsymbol{A}} = \underset{\boldsymbol{A} \in \mathcal{A}_N}{\arg\min} \min_{\boldsymbol{F}} \left\| \hat{\boldsymbol{U}}_{\mathrm{s}} - \boldsymbol{A}\boldsymbol{F} \right\|_{\mathrm{F}}^2$$

MUSIC Null-spectrum

$$f_{\mathrm{MUSIC}}(\theta) = \boldsymbol{a}^{\mathsf{H}}(\theta)\hat{\boldsymbol{U}}_{\mathrm{n}}\hat{\boldsymbol{U}}_{\mathrm{n}}^{\mathsf{H}}\boldsymbol{a}(\theta)$$

Alternative Interpretation

$$f_{\mathrm{MUSIC}}(\theta) \propto \min_{\boldsymbol{f}} \left\| \hat{\boldsymbol{U}}_{\mathrm{s}} - \boldsymbol{a}(\theta)\boldsymbol{f}^{\mathsf{T}} \right\|_{\mathrm{F}}^2$$

- MUSIC can be considered as a single-source approximation of WSF with identity weighting.

Single-source Approximation Techniques
Root-MUSIC

- For ULA geometries with baseline d the steering vector

$$\boldsymbol{a}(\theta) = [1, e^{-\mathrm{j}\frac{2\pi}{\lambda} d\cos(\theta)}, \cdots, e^{-\mathrm{j}\frac{2\pi}{\lambda}(M-1)d\cos(\theta)}]^{\mathsf{T}} \in \mathbb{C}^{M\times 1}$$

  exhibits Vandermonde structure with unit modulus entries.
- In this case the MUSIC method has an efficient variation, that is both computationally more efficient and that shows improved resolution capabilities.
- Defining the unit root $z = e^{-\mathrm{j}\frac{2\pi}{\lambda} d\cos(\theta)}$ the steering vector reads

$$\boldsymbol{a}(z) = [1, z \cdots, z^{M-1}]^{\mathsf{T}} \in \mathbb{C}^{M\times 1}.$$

- With the definition above and the property $z^* = z^{-1}$ for $|z| = 1$ the MUSIC null-spectrum can be expressed as the polynomial [Barabell'83]

$$f_{\text{MUSIC}}(z) = \boldsymbol{a}^{\mathsf{H}}(z)\hat{\boldsymbol{U}}_{\text{n}}\hat{\boldsymbol{U}}_{\text{n}}^{\mathsf{H}}\boldsymbol{a}(z) = \boldsymbol{a}^{\mathsf{T}}(1/z)\hat{\boldsymbol{U}}_{\text{n}}\hat{\boldsymbol{U}}_{\text{n}}^{\mathsf{H}}\boldsymbol{a}(z)$$

  of degree $2M-2$.

Single-source Approximation Techniques

Root-MUSIC

The spectral MUSIC algorithm evaluates the MUSIC polynomial on the unit circle and seeks the N deepest minima.

Hence, the signal roots are determined as:

Spectral MUSIC null-spectrum

$$\{\hat{z}_{\text{MUSIC}}\} = {}^{N}\!\!\operatorname*{arg\,min}_{z \in \mathbb{C},\ |z|=1} \boldsymbol{a}^{\mathsf{T}}(1/z)\hat{\boldsymbol{U}}_{\text{n}}\hat{\boldsymbol{U}}_{\text{n}}^{\mathsf{H}}\boldsymbol{a}(z).$$

However, the set $\{z \in \mathbb{C} \big| \ |z| = 1\}$ is nonconvex and the minimization of the polynomial on the unit circle requires full spectral search.

Single-source Approximation Techniques
Root-MUSIC

- The root-MUSIC algorithm can be understood as a relaxation of the search space over which the MUSIC polynomial is minimized.
- Instead of minimizing the null-spectrum on the unit circle, hence over the set $\{z \in \mathbb{C} \big| \, |z| = 1\}$, the unit circle constraint is relaxed to the full complex space $z \in \mathbb{C}$.
- We remark that the resulting MUSIC polynomial function may take complex values outside the unit circle. Hence, the absolute value is considered outside the unit circle, resulting in optimization problem:

$$\min_{z \in \mathbb{C}} \left| \boldsymbol{a}^{\mathsf{T}}(1/z) \hat{\boldsymbol{U}}_{\mathrm{n}} \hat{\boldsymbol{U}}_{\mathrm{n}}^{\mathsf{H}} \boldsymbol{a}(z) \right| .$$

Single-source Approximation Techniques
Root-MUSIC

- The objective of problem

$$\min_{z \in \mathbb{C}} \left| \boldsymbol{a}^{\mathsf{T}}(1/z)\hat{\boldsymbol{U}}_{\mathrm{n}}\hat{\boldsymbol{U}}_{\mathrm{n}}^{\mathsf{H}}\boldsymbol{a}(z) \right|$$

  is non-negative and the minima are obtained by simply computing the roots of $f_{\mathrm{MUSIC}}(z)$, i.e., by solving equation

$$f_{\mathrm{MUSIC}}(z) = \boldsymbol{a}^{\mathsf{T}}(1/z)\hat{\boldsymbol{U}}_{\mathrm{n}}\hat{\boldsymbol{U}}_{\mathrm{n}}^{\mathsf{H}}\boldsymbol{a}(z) = 0.$$

- The MUSIC polynomial is of degree $2M-2$ and exhibits exactly $2M-2$ roots.
- Hence, due to the relaxation of the feasible set, there exist $2M-2$ global minima of the relaxed optimization problem above (instead of N).
- In the following a procedure will be described to partition the set of $2M-2$ roots into a set of signal roots that correspond to the true signals and a set of spurious roots that result from the relaxation.

Single-source Approximation Techniques
Root-MUSIC

The MUSIC polynomial has the order $2M - 2$, and, therefore, it has $2M - 2$ roots.
We select only N closest to the unit circle roots inside it ($|z| \leq 1$).

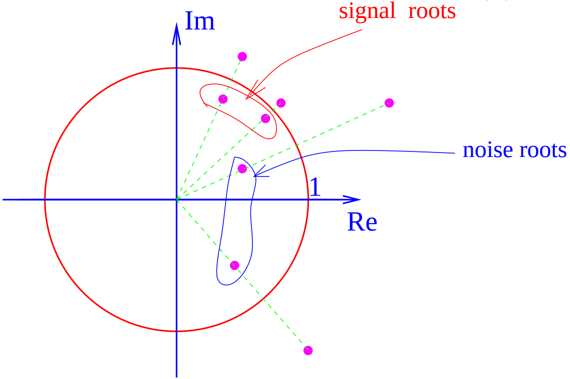


Table of Contents

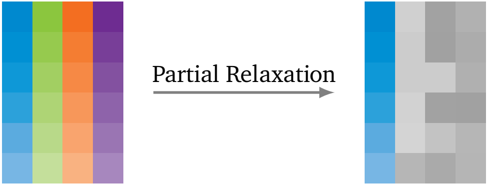
Table of Contents

Formulation of the Multi-dimensional Search

$$\left\{\hat{\boldsymbol{A}}\right\} = \underset{\boldsymbol{A}\in\mathcal{A}_N}{\arg\min} f\left(\boldsymbol{A}\right)$$

Relaxed Array Manifold

$$\bar{\mathcal{A}}_N = \left\{\boldsymbol{A}\in\mathbb{C}^{M\times N} \mid \boldsymbol{A} = \left[\boldsymbol{a}(\theta), \boldsymbol{B}\right], \boldsymbol{B}\in\mathbb{C}^{M\times(N-1)} \text{ and } \operatorname{rank}\left(\boldsymbol{A}\right) = N\right\}$$



$\boldsymbol{A}\in\mathcal{A}_N$ $\boldsymbol{A}\in\bar{\mathcal{A}}_N$

Partial Relaxation Techniques
General Concept

Formulation of the Multi-dimensional Search

$$\left\{\hat{\boldsymbol{A}}\right\} = \underset{\boldsymbol{A}\in\mathcal{A}_N}{\arg\min}\, f\left(\boldsymbol{A}\right)$$

Relaxed Array Manifold

$$\bar{\mathcal{A}}_N = \left\{\boldsymbol{A} \in \mathbb{C}^{M\times N} \mid \boldsymbol{A} = \left[\boldsymbol{a}(\theta), \boldsymbol{B}\right], \boldsymbol{B} \in \mathbb{C}^{M\times (N-1)} \text{ and } \operatorname{rank}\left(\boldsymbol{A}\right) = N\right\}$$

Formulation of Partial Relaxation (PR) Framework [Trinh-Hoang'18]

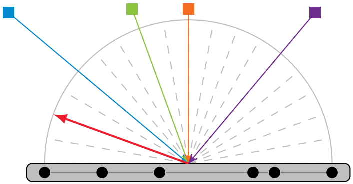
$$\left\{\hat{\boldsymbol{a}}_{\mathrm{PR}}\right\} =^{N} \underset{\boldsymbol{a}\in\mathcal{A}_1}{\arg\min} \min_{\boldsymbol{B}\in\mathbb{C}^{M\times(N-1)}} f\left(\left[\boldsymbol{a}, \boldsymbol{B}\right]\right)$$

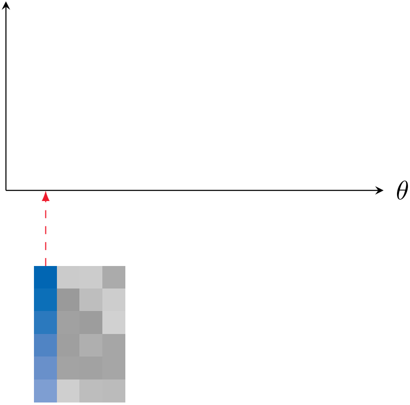
- Compute the null-spectrum $f_{\mathrm{PR}}(\theta) = \min_{\boldsymbol{B}\in\mathbb{C}^{M\times(N-1)}} f\left(\left[\boldsymbol{a}(\theta), \boldsymbol{B}\right]\right)$.
- N-deepest local minimizers of $f_{\mathrm{PR}}(\theta)$ are the DOA estimates.

Partial Relaxation Techniques

General Concept

$$\min_{\boldsymbol{B}\in\mathbb{C}^{M\times(N-1)}} f\left(\left[\boldsymbol{a}(\theta), \boldsymbol{B}\right]\right)$$

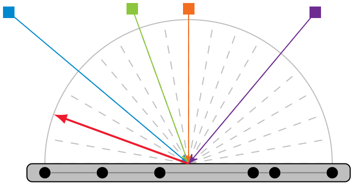


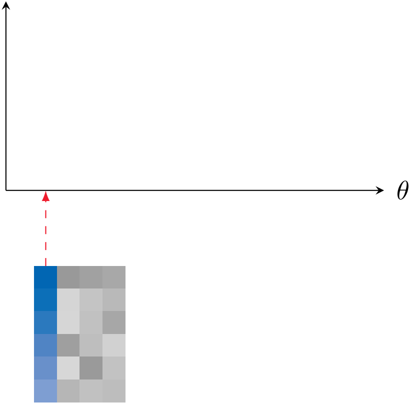


$$\left[\boldsymbol{a}(\theta), \boldsymbol{B}\right]$$

Partial Relaxation Techniques
General Concept

$$\min_{\boldsymbol{B}\in\mathbb{C}^{M\times(N-1)}} f\left(\left[\boldsymbol{a}(\theta),\boldsymbol{B}\right]\right)$$

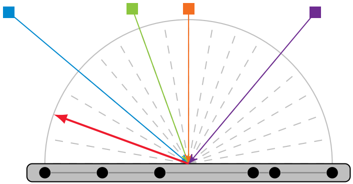


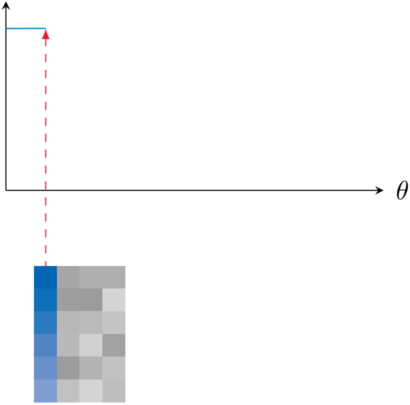


$\left[\boldsymbol{a}(\theta),\boldsymbol{B}\right]$

Partial Relaxation Techniques
General Concept

$$\min_{\boldsymbol{B}\in\mathbb{C}^{M\times(N-1)}} f\left(\left[\boldsymbol{a}(\theta),\boldsymbol{B}\right]\right)$$

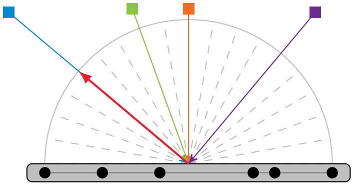


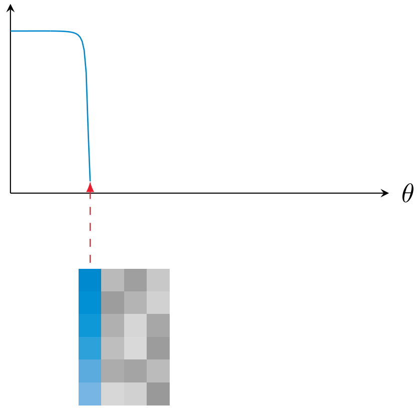


$\left[\boldsymbol{a}(\theta),\boldsymbol{B}\right]$

Partial Relaxation Techniques
General Concept

$$\min_{\boldsymbol{B} \in \mathbb{C}^{M \times (N-1)}} f\left(\left[\boldsymbol{a}(\theta), \boldsymbol{B}\right]\right)$$



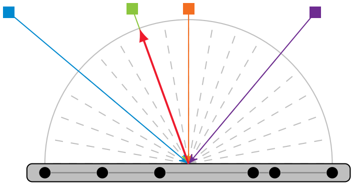


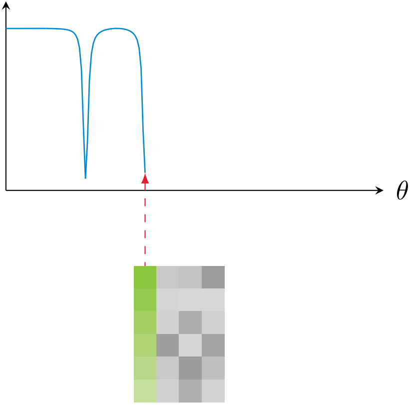
$\left[\boldsymbol{a}(\theta), \boldsymbol{B}\right]$

Partial Relaxation Techniques

General Concept

$$\min_{\boldsymbol{B}\in\mathbb{C}^{M\times(N-1)}} f\left(\left[\boldsymbol{a}(\theta),\boldsymbol{B}\right]\right)$$



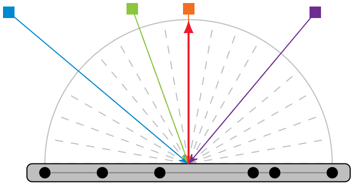


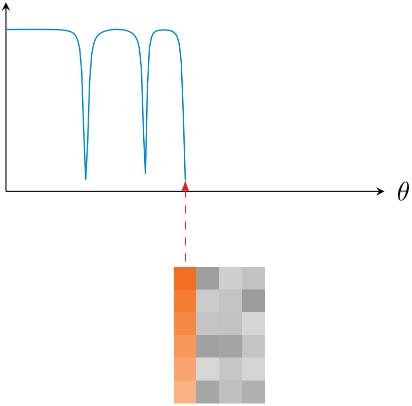
$$\left[\boldsymbol{a}(\theta),\boldsymbol{B}\right]$$

Partial Relaxation Techniques

General Concept

$$\min_{\boldsymbol{B} \in \mathbb{C}^{M \times (N-1)}} f\left(\left[\boldsymbol{a}(\theta), \boldsymbol{B}\right]\right)$$

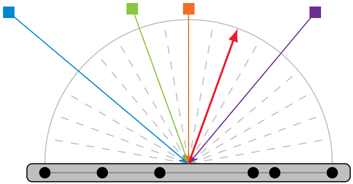


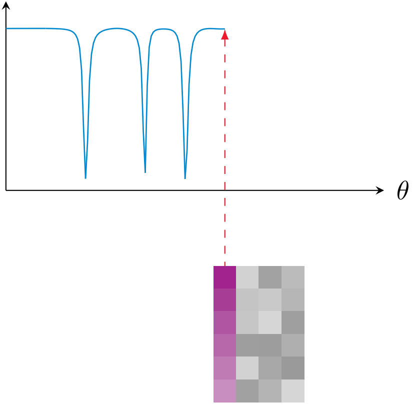


$\left[\boldsymbol{a}(\theta), \boldsymbol{B}\right]$

Partial Relaxation Techniques
General Concept

$$\min_{\boldsymbol{B}\in\mathbb{C}^{M\times(N-1)}} f\left(\left[\boldsymbol{a}(\theta),\boldsymbol{B}\right]\right)$$

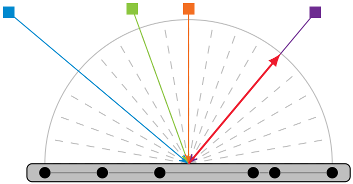


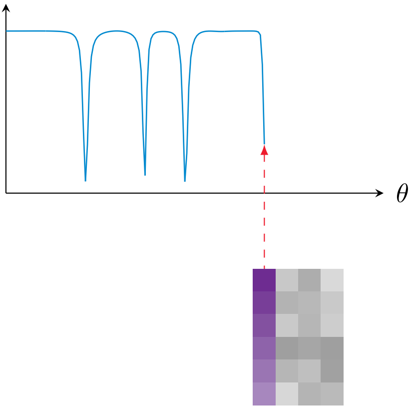


$\left[\boldsymbol{a}(\theta),\boldsymbol{B}\right]$

Partial Relaxation Techniques
General Concept

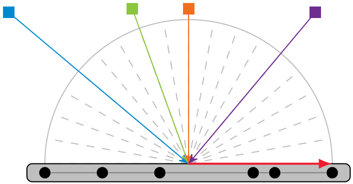
$$\min_{\boldsymbol{B}\in\mathbb{C}^{M\times(N-1)}} f\left(\left[\boldsymbol{a}(\theta), \boldsymbol{B}\right]\right)$$

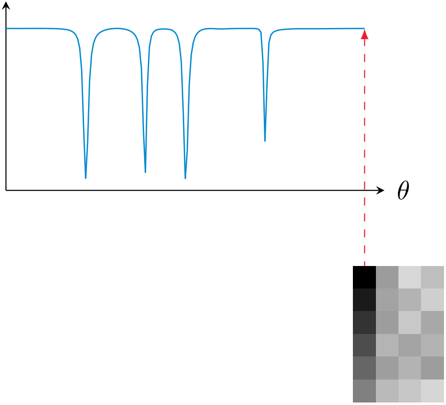




$\left[\boldsymbol{a}(\theta), \boldsymbol{B}\right]$

$$\min_{\boldsymbol{B}\in\mathbb{C}^{M\times(N-1)}} f\left(\left[\boldsymbol{a}(\theta), \boldsymbol{B}\right]\right)$$

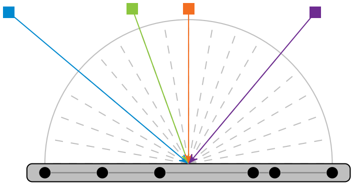


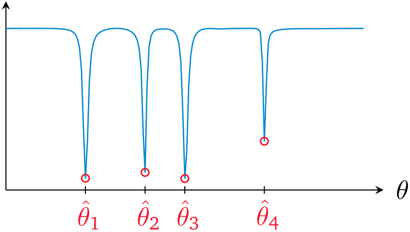


$$\left[\boldsymbol{a}(\theta), \boldsymbol{B}\right]$$

Partial Relaxation Techniques
General Concept

$$\min_{\boldsymbol{B}\in\mathbb{C}^{M\times(N-1)}} f\left(\left[\boldsymbol{a}(\theta),\boldsymbol{B}\right]\right)$$





- Relax the manifold structure of the signals from “interfering” directions.
- Generally lower complexity than multi-dimensional search.

Partial Relaxation Techniques
PR Deterministic Maximum Likelihood

Recall the DML estimator

$$\left\{\hat{\boldsymbol{A}}_{\text{DML}}\right\} = \underset{\boldsymbol{A}\in\mathcal{A}_N}{\arg\min}\ \text{Tr}\left(\boldsymbol{\Pi}_{\boldsymbol{A}}^{\perp}\hat{\boldsymbol{R}}\right)$$

Partially-relaxed (PR) Formulation

$$\begin{aligned}\{\hat{\boldsymbol{a}}_{\text{PR-DML}}\} &=^{N} \underset{\boldsymbol{a}\in\mathcal{A}_1}{\arg\min}\ \min_{\boldsymbol{B}\in\mathbb{C}^{M\times(N-1)}} \text{Tr}\left(\boldsymbol{\Pi}_{[\boldsymbol{a},\boldsymbol{B}]}^{\perp}\hat{\boldsymbol{R}}\right)\\ &=^{N} \underset{\boldsymbol{a}\in\mathcal{A}_1}{\arg\min}\ \min_{\boldsymbol{B}\in\mathbb{C}^{M\times(N-1)}} \text{Tr}\left(\boldsymbol{\Pi}_{\boldsymbol{a}}^{\perp}\hat{\boldsymbol{R}}\right) - \text{Tr}\left(\boldsymbol{\Pi}_{\boldsymbol{\Pi}_{\boldsymbol{a}}^{\perp}\boldsymbol{B}}\hat{\boldsymbol{R}}\right)\end{aligned}$$

Null-spectrum of the PR-DML Estimator with $\boldsymbol{a} = \boldsymbol{a}(\theta)$

$$f_{\text{PR-DML}}(\theta) = \text{Tr}\left(\boldsymbol{\Pi}_{\boldsymbol{a}}^{\perp}\hat{\boldsymbol{R}}\right) - \max_{\boldsymbol{B}\in\mathbb{C}^{M\times(N-1)}} \text{Tr}\left(\boldsymbol{\Pi}_{\boldsymbol{\Pi}_{\boldsymbol{a}}^{\perp}\boldsymbol{B}}\hat{\boldsymbol{R}}\right)$$

Partial Relaxation Techniques
PR Deterministic Maximum Likelihood

New Optimization Problem

$$\max_{\boldsymbol{B}\in\mathbb{C}^{M\times(N-1)}} \operatorname{Tr}\left(\boldsymbol{\Pi}_{\boldsymbol{\Pi}_{\boldsymbol{a}}^{\perp}\boldsymbol{B}}\hat{\boldsymbol{R}}\right)$$

Eigenvalue Decomposition of $\boldsymbol{\Pi}_{\boldsymbol{\Pi}_{\boldsymbol{a}}^{\perp}\boldsymbol{B}}$

$$\boldsymbol{\Pi}_{\boldsymbol{\Pi}_{\boldsymbol{a}}^{\perp}\boldsymbol{B}} = \boldsymbol{Z}\boldsymbol{Z}^{\mathsf{H}} \text{ with } \boldsymbol{Z}\in\mathbb{C}^{M\times K}$$

- $\operatorname{rank}\left(\boldsymbol{\Pi}_{\boldsymbol{\Pi}_{\boldsymbol{a}}^{\perp}\boldsymbol{B}}\right) = K \leq N-1$
- $\boldsymbol{Z}^{\mathsf{H}}\boldsymbol{a} = \boldsymbol{0}$

Equivalent Reformulation

$$\max_{\boldsymbol{Z}\in\mathbb{C}^{M\times K}} \operatorname{Tr}\left(\boldsymbol{Z}^{\mathsf{H}}\boldsymbol{\Pi}_{\boldsymbol{a}}^{\perp}\hat{\boldsymbol{R}}\boldsymbol{\Pi}_{\boldsymbol{a}}^{\perp}\boldsymbol{Z}\right) = \sum_{k=1}^{N-1}\lambda_k(\boldsymbol{\Pi}_{\boldsymbol{a}}^{\perp}\hat{\boldsymbol{R}}\boldsymbol{\Pi}_{\boldsymbol{a}}^{\perp}) = \sum_{k=1}^{N-1}\lambda_k(\boldsymbol{\Pi}_{\boldsymbol{a}}^{\perp}\hat{\boldsymbol{R}})$$

subject to $\boldsymbol{Z}^{\mathsf{H}}\boldsymbol{a} = \boldsymbol{0}$

$\boldsymbol{Z}^{\mathsf{H}}\boldsymbol{Z} = \boldsymbol{I}$

Null-spectrum of the PR-DML Estimator

$$
\begin{aligned}
f_{\text{PR-DML}}(\theta) &= \operatorname{Tr}\left(\boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}(\theta)}\hat{\boldsymbol{R}}\right) - \max_{\boldsymbol{B}\in\mathbb{C}^{M\times(N-1)}} \operatorname{Tr}\left(\boldsymbol{\Pi}_{\boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}(\theta)}\boldsymbol{B}}\hat{\boldsymbol{R}}\right) \\
&= \sum_{k=N}^{M} \lambda_k(\boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}(\theta)}\hat{\boldsymbol{R}}) \\
&= \sum_{k=N}^{M} \lambda_k\left(\hat{\boldsymbol{R}} - \frac{1}{||\boldsymbol{a}(\theta)||^2}\hat{\boldsymbol{R}}^{1/2}\boldsymbol{a}(\theta)\boldsymbol{a}^{\mathsf{H}}(\theta)\hat{\boldsymbol{R}}^{1/2}\right)
\end{aligned}
$$

Remarks

- Multiple minimizers for $\boldsymbol{B}$.
- Closed-form expressions for the null-spectrum.
- $(M-N+1)$- smallest eigenvalues are required.

Alternative Derivation of Null-spectrum of PR-DML

$$f_{\text{PR-DML}}(\theta) = \min_{\boldsymbol{B}\in\mathbb{C}^{M\times(N-1)}} \operatorname{Tr}\left(\boldsymbol{\Pi}^{\perp}_{[\boldsymbol{a}(\theta),\boldsymbol{B}]}\hat{\boldsymbol{R}}\right)$$

$$= \min_{\boldsymbol{B}\in\mathbb{C}^{M\times(N-1)}} \min_{\boldsymbol{s}\in\mathbb{C}^{T\times 1},\boldsymbol{H}\in\mathbb{C}^{(N-1)\times T}} \frac{1}{T}\left|\left|\boldsymbol{X} - \boldsymbol{a}(\theta)\boldsymbol{s}^{\mathsf{T}} - \boldsymbol{B}\boldsymbol{H}\right|\right|_{\mathrm{F}}^{2}$$

Substitute $\boldsymbol{E} = \boldsymbol{B}\boldsymbol{H}$ and Concentrate with Respect to $\boldsymbol{s}$

$$f_{\text{PR-DML}}(\theta) = \min_{\operatorname{rank}(\boldsymbol{E})\leq N-1} \frac{1}{T}\left|\left|\boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}(\theta)}\boldsymbol{X} - \boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}(\theta)}\boldsymbol{E}\right|\right|_{\mathrm{F}}^{2}$$

$$= \frac{1}{T}\sum_{k=N}^{M} \sigma_k^2\left(\boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}(\theta)}\boldsymbol{X}\right)$$

$$= \sum_{k=N}^{M} \lambda_k\left(\boldsymbol{\Pi}^{\perp}_{\boldsymbol{a}(\theta)}\hat{\boldsymbol{R}}\right)$$

Partial Relaxation Techniques
PR Weighted Subspace Fitting

Recall the WSF estimator

$$\left\{\hat{\boldsymbol{A}}_{\mathrm{WSF}}\right\} = \underset{\boldsymbol{A}\in\mathcal{A}_N}{\arg\min}\ \mathrm{Tr}\left(\boldsymbol{\Pi}_{\boldsymbol{A}}^{\perp}\hat{\boldsymbol{U}}_{\mathrm{s}}\boldsymbol{W}\hat{\boldsymbol{U}}_{\mathrm{s}}^{\mathsf{H}}\right)$$

Partially-relaxed (PR) Formulation

$$\{\hat{\boldsymbol{a}}_{\text{PR-WSF}}\} = {}^{N}\underset{\boldsymbol{a}\in\mathcal{A}_1}{\arg\min}\ \underset{\boldsymbol{B}\in\mathbb{C}^{M\times(N-1)}}{\min}\ \mathrm{Tr}\left(\boldsymbol{\Pi}_{[\boldsymbol{a},\boldsymbol{B}]}^{\perp}\hat{\boldsymbol{U}}_{\mathrm{s}}\boldsymbol{W}\hat{\boldsymbol{U}}_{\mathrm{s}}^{\mathsf{H}}\right)$$

Null-spectrum of the PR-WSF Estimator

$$f_{\text{PR-WSF}}(\theta) = \lambda_N\left(\boldsymbol{\Pi}_{\boldsymbol{a}(\theta)}^{\perp}\hat{\boldsymbol{U}}_{\mathrm{s}}\boldsymbol{W}\hat{\boldsymbol{U}}_{\mathrm{s}}^{\mathsf{H}}\right)$$

- Only one eigenvalue required.
- PR-WSF with $\boldsymbol{W} = \boldsymbol{I}$ is equivalent to MUSIC estimator.

Partial Relaxation Techniques
PR Constrained Covariance Fitting

Recall the Covariance Matrix $\boldsymbol{R}$

$$\boldsymbol{R} = \boldsymbol{A}\boldsymbol{P}\boldsymbol{A}^{\mathsf{H}} + \nu\boldsymbol{I}$$
$$= \begin{bmatrix}\boldsymbol{a} & \boldsymbol{B}\end{bmatrix}\begin{bmatrix}\sigma_s^2 & \boldsymbol{\rho}^{\mathsf{H}} \\ \boldsymbol{\rho} & \boldsymbol{Q}\end{bmatrix}\begin{bmatrix}\boldsymbol{a}^{\mathsf{H}} \\ \boldsymbol{B}^{\mathsf{H}}\end{bmatrix} + \nu\boldsymbol{I}$$

Formulation of PR-Constrained Covariance Fitting (PR-CCF)

$$\{\hat{\boldsymbol{a}}_{\text{PR-CCF}}\} = {}^{N}\underset{\boldsymbol{a}\in\mathcal{A}_1}{\arg\min}\ \min_{\boldsymbol{B},\sigma_s^2\geq 0,\boldsymbol{Q}\succeq\boldsymbol{0}} \left\|\hat{\boldsymbol{R}} - \sigma_s^2\boldsymbol{a}\boldsymbol{a}^{\mathsf{H}} - \boldsymbol{B}\boldsymbol{Q}\boldsymbol{B}^{\mathsf{H}}\right\|_{\mathrm{F}}^2$$
$$\text{subject to } \hat{\boldsymbol{R}} - \sigma_s^2\boldsymbol{a}\boldsymbol{a}^{\mathsf{H}} - \boldsymbol{B}\boldsymbol{Q}\boldsymbol{B}^{\mathsf{H}} \succeq \boldsymbol{0}$$

- Neglect the correlation between source signals.
- Replace the noise component with the positive-semidefinite constraint.

Partial Relaxation Techniques

PR Constrained Covariance Fitting

Equivalent formulation of the inner optimization

$$\min_{\sigma_s^2 \geq 0} \sum_{k=N}^{M} \lambda_k^2 \left(\hat{\boldsymbol{R}} - \sigma_s^2 \boldsymbol{a}\boldsymbol{a}^{\mathsf{H}} \right)$$

$$\text{subject to } \hat{\boldsymbol{R}} - \sigma_s^2 \boldsymbol{a}\boldsymbol{a}^{\mathsf{H}} \succeq \boldsymbol{0}$$

Closed-form solution for the minimizer $\hat{\sigma}^2_{s,\,C}$

$$\hat{\sigma}^2_{s,\,C} = \frac{1}{\boldsymbol{a}^{\mathsf{H}} \hat{\boldsymbol{R}}^{-1} \boldsymbol{a}}$$

Null-spectrum of the PR-CCF Estimator

$$f_{\text{PR-CCF}}(\theta) = \sum_{k=N}^{M} \lambda_k^2 \left(\hat{\boldsymbol{R}} - \frac{1}{\boldsymbol{a}^{\mathsf{H}}(\theta) \hat{\boldsymbol{R}}^{-1} \boldsymbol{a}(\theta)} \boldsymbol{a}(\theta) \boldsymbol{a}^{\mathsf{H}}(\theta) \right)$$

Partial Relaxation Techniques
PR Unconstrained Covariance Fitting

Formulation of PR-Unconstrained Covariance Fitting (PR-UCF)

$$\{\hat{\boldsymbol{a}}_{\text{PR-UCF}}\} = {}^{N} \underset{\boldsymbol{a}\in\mathcal{A}_1}{\arg\min} \min_{\boldsymbol{B},\sigma_{\text{s}}^2\geq 0,\boldsymbol{Q}\succeq\boldsymbol{0}} \left|\left|\hat{\boldsymbol{R}} - \sigma_{\text{s}}^2\boldsymbol{a}\boldsymbol{a}^{\mathsf{H}} - \boldsymbol{B}\boldsymbol{Q}\boldsymbol{B}^{\mathsf{H}}\right|\right|_{\text{F}}^2$$

Null-spectrum of the PR-UCF Estimator with $\boldsymbol{a} = \boldsymbol{a}(\theta)$

$$f_{\text{PR-UCF}}(\theta) = \min_{\sigma_{\text{s}}^2\geq 0} \sum_{k=N}^{M} \lambda_k^2\left(\hat{\boldsymbol{R}} - \sigma_{\text{s}}^2\boldsymbol{a}\boldsymbol{a}^{\mathsf{H}}\right)$$

- No closed-form solution for the minimizer $\hat{\sigma}_{s,\text{U}}^2$.
- $\bar{\lambda}_k(\sigma_{\text{s}}^2) = \lambda_k\left(\hat{\boldsymbol{R}} - \sigma_{\text{s}}^2\boldsymbol{a}\boldsymbol{a}^{\mathsf{H}}\right)$ is continuously differentiable with respect to σ_s^2

$$\frac{\mathrm{d}\bar{\lambda}_k\left(\sigma_{\text{s}}^2\right)}{\mathrm{d}\sigma_{\text{s}}^2} = -\frac{1}{\sigma_{\text{s}}^4\boldsymbol{a}^{\mathsf{H}}\left(\hat{\boldsymbol{R}} - \bar{\lambda}_k(\sigma_{\text{s}}^2)\boldsymbol{I}_M\right)^{-2}\boldsymbol{a}}.$$

Partial Relaxation Techniques

PR Unconstrained Covariance Fitting

Define

$$g(\sigma_s^2) = \sum_{k=N}^{M} \lambda_k^2 \left(\hat{\boldsymbol{R}} - \sigma_s^2 \boldsymbol{a}\boldsymbol{a}^{\mathsf{H}}\right)$$

Objective: Find $\hat{\sigma}_{s,\mathrm{U}}^2$ where the derivative $g'(\sigma_s^2)$ vanishes

$$g'(\sigma_s^2) = -\sum_{k=N}^{M} \frac{2\bar{\lambda}_k(\sigma_s^2)}{\sigma_s^4 \boldsymbol{a}^{\mathsf{H}} \left(\hat{\boldsymbol{R}} - \bar{\lambda}_k(\sigma_s^2)\boldsymbol{I}_M\right)^{-2} \boldsymbol{a}}$$

- If $\sigma_s^2 \to 0 \implies g'(\sigma_s^2) < 0$
- If $\sigma_s^2 \to \infty \implies g(\sigma_s^2) \approx \sigma_s^4 \, ||\boldsymbol{a}||_2^4 \implies g'(\sigma_s^2) > 0$

Solution: Find an interval where $g'(\sigma_s^2)$ changes sign and perform bisection search

Formulation of PR-Full Covariance Fitting (PR-FCF)

$$\{\hat{\boldsymbol{a}}_{\text{PR-UCF}}\} =^{N} \underset{\boldsymbol{a}\in\mathcal{A}_1}{\arg\min} \min_{\boldsymbol{B},\sigma_{\text{s}}^2\geq 0,\boldsymbol{Q}\succeq\boldsymbol{0},\nu\geq 0} \left|\left|\hat{\boldsymbol{R}} - \sigma_{\text{s}}^2 \boldsymbol{a}\boldsymbol{a}^{\mathsf{H}} - \boldsymbol{B}\boldsymbol{Q}\boldsymbol{B} - \nu\boldsymbol{I}\right|\right|_{\text{F}}^2$$

Null-spectrum of the PR-FCF Estimator with $\boldsymbol{a} = \boldsymbol{a}(\theta)$

$$f_{\text{PR-FCF}}(\theta) = \min_{\sigma_{\text{s}}^2\geq 0} \sum_{k=N}^{M} \lambda_k^2\left(\hat{\boldsymbol{R}} - \sigma_{\text{s}}^2\boldsymbol{a}\boldsymbol{a}^{\mathsf{H}}\right) - \frac{\left(\sum_{k=N}^{M}\lambda_k\left(\hat{\boldsymbol{R}} - \sigma_{\text{s}}^2\boldsymbol{a}\boldsymbol{a}^{\mathsf{H}}\right)\right)^2}{M-N+1}$$

- No closed-form solution for the minimizer $\hat{\sigma}_{s,\text{ F}}^2$.
- Numerical suboptimal solution obtained from Newton's method.

Partial Relaxation Techniques
Insights and Relation

Methods	Multi-dimensional Search	Partial Relaxation	Single-source Approximation
Signal Fitting	DML	PR-DML	Conv. Beamformer
Subspace Fitting	WSF	PR-WSF	Weighted MUSIC
Covariance Fitting	Unweighted COMET	PR-CCF PR-UCF PR-FCF	Capon Beamformer Conv. Beamformer

- Degraded performance of PR methods in the case of correlated signals.
- Null-spectra of PR methods require the computation of eigenvalues.

Explanation of Performance Degradation of PR Methods

Case study: Two fully coherent source signals without sensor noise

$$\begin{aligned} \boldsymbol{X} &= \boldsymbol{a}(\theta_1)\boldsymbol{s}^\mathsf{T} + \boldsymbol{a}(\theta_2)\boldsymbol{s}^\mathsf{T} \\ &= \Big(\boldsymbol{a}(\theta_1) + \boldsymbol{a}(\theta_2)\Big)\boldsymbol{s}^\mathsf{T}. \end{aligned}$$

Null-spectrum of the PR-DML estimator for $N = 2$ source signals

$$f_{\text{PR-DML}}(\theta) = \min_{\boldsymbol{b}\in\mathbb{C}^{M\times 1}} \min_{\boldsymbol{s}\in\mathbb{C}^{T\times 1}, \boldsymbol{h}\in\mathbb{C}^{T\times 1}} \frac{1}{T}\left|\left|\boldsymbol{X} - \boldsymbol{a}(\theta)\boldsymbol{s}^\mathsf{T} - \boldsymbol{b}\boldsymbol{h}^\mathsf{T}\right|\right|_\text{F}^2$$

- Cost function is non-negative.
- Perfect match is achieved if $\boldsymbol{b} = \boldsymbol{a}(\theta_1) + \boldsymbol{a}(\theta_2)$ regardless of θ.
- Flat null-spectrum for all look-direction $\theta \implies$ no reliable DOA estimation.

Null-spectrum of the PR-DML Estimator

$$f_{\text{PR-DML}}(\theta) = \sum_{k=N}^{M} \lambda_k \left(\hat{\boldsymbol{R}} - \frac{1}{||\boldsymbol{a}||^2} \hat{\boldsymbol{R}}^{1/2} \boldsymbol{a}\boldsymbol{a}^{\mathsf{H}} \hat{\boldsymbol{R}}^{1/2} \right) \text{ with } \boldsymbol{a} = \boldsymbol{a}(\theta)$$

Null-spectrum of the PR-CCF Estimator

$$f_{\text{PR-CCF}}(\theta) = \sum_{k=N}^{M} \lambda_k^2 \left(\hat{\boldsymbol{R}} - \frac{1}{\boldsymbol{a}^{\mathsf{H}} \hat{\boldsymbol{R}}^{-1} \boldsymbol{a}} \boldsymbol{a}\boldsymbol{a}^{\mathsf{H}} \right) \text{ with } \boldsymbol{a} = \boldsymbol{a}(\theta)$$

- Dependent on eigenvalues but not on eigenvectors.
- Similar structure of the matrix argument.

Null-spectrum of the PR-CCF Estimator

$$f_{\text{PR-CCF}}(\theta) = \sum_{k=N}^{M} \lambda_k^2 \left(\hat{\boldsymbol{R}} - \frac{1}{\boldsymbol{a}^{\mathsf{H}} \hat{\boldsymbol{R}}^{-1} \boldsymbol{a}} \boldsymbol{a}\boldsymbol{a}^{\mathsf{H}} \right) \text{ with } \boldsymbol{a} = \boldsymbol{a}(\theta)$$

- Dependent on eigenvalues but not on eigenvectors.
- Similar structure of the matrix argument.

Core Numerical Problem: Efficient Computation of Eigenvalues

$$\bar{d}_k = \lambda_k \left(\boldsymbol{D} - \bar{\rho} \boldsymbol{z}\boldsymbol{z}^{\mathsf{H}} \right) \text{ with } \rho > 0$$

- $\boldsymbol{D} = \text{diag}(d_1, \ldots, d_K) \in \mathbb{R}^{K \times K}$ with $d_1 > \ldots > d_K$.
- $\boldsymbol{z} = [z_1, \ldots, z_K]^{\mathsf{T}} \in \mathbb{C}^{K \times 1}$ has no zero entry.

Partial Relaxation Techniques
Efficient Implementation

Remarks

- Corresponding to the routine `dlaed4()` in LAPACK [Anderson'99].
- Applicable to PR estimators using orthogonal transformation.
- Adaptive initialization using previous eigenvalues.
- Reduction in execution time using alternative expressions.

Example: PR-DML Estimator

$$\{\hat{\boldsymbol{a}}_{\text{PR-DML}}\} =^N \underset{\boldsymbol{a}\in\mathcal{A}_1}{\arg\min} \sum_{k=N}^{M} \lambda_k \left(\hat{\boldsymbol{R}} - \frac{1}{||\boldsymbol{a}||^2} \hat{\boldsymbol{R}}^{1/2} \boldsymbol{a}\boldsymbol{a}^{\mathsf{H}} \hat{\boldsymbol{R}}^{1/2}\right)$$

$$=^N \underset{\boldsymbol{a}\in\mathcal{A}_1}{\arg\min}\ \mathrm{Tr}\left(\hat{\boldsymbol{R}}\right) - \frac{\boldsymbol{a}^{\mathsf{H}}\hat{\boldsymbol{R}}\boldsymbol{a}}{\boldsymbol{a}^{\mathsf{H}}\boldsymbol{a}} - \sum_{k=1}^{N-1} \lambda_k \left(\hat{\boldsymbol{\Lambda}} - \frac{1}{||\boldsymbol{a}||_2^2} \hat{\boldsymbol{\Lambda}}^{1/2} \hat{\boldsymbol{U}}^{\mathsf{H}} \boldsymbol{a}\boldsymbol{a}^{\mathsf{H}} \hat{\boldsymbol{U}} \hat{\boldsymbol{\Lambda}}^{1/2}\right)$$

Relaxed Array Manifold

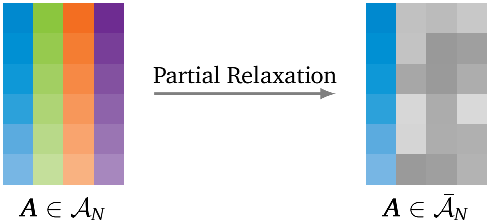
$$\bar{\mathcal{A}}_N = \left\{ \boldsymbol{A} | \boldsymbol{A} = [\boldsymbol{a}(\theta), \boldsymbol{B}], \boldsymbol{a}(\theta) \in \mathcal{A}_1, \boldsymbol{B} \in \mathbb{C}^{M \times (N-1)} \text{ and rank}(\boldsymbol{A}) = N \right\}$$

Partial Relaxation Model for Time Instant t

$$\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t) + \boldsymbol{n}(t) \text{ with } \boldsymbol{A} \in \bar{\mathcal{A}}_N.$$

Relaxed Array Manifold

$$\bar{\mathcal{A}}_N = \left\{ \boldsymbol{A} | \boldsymbol{A} = [\boldsymbol{a}(\theta), \boldsymbol{B}], \boldsymbol{a}(\theta) \in \mathcal{A}_1, \boldsymbol{B} \in \mathbb{C}^{M \times (N-1)} \text{ and } \operatorname{rank}(\boldsymbol{A}) = N \right\}$$



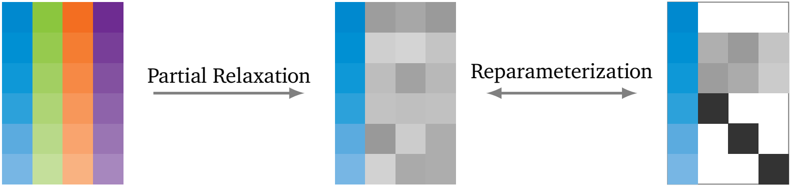
How does the array manifold relaxation affect the DOA estimation?

Partial Relaxation Techniques

Crámer-Rao Bound for Partial Relaxation Model

Reparameterization for Redundancy Elimination [Trinh-Hoang'20-2]

$$\boldsymbol{A}(\boldsymbol{\theta}) \in \mathcal{A}_N \qquad \boldsymbol{A} = \begin{bmatrix} a_1(\vartheta) & \boldsymbol{b}_1^{\mathrm{T}} \\ \boldsymbol{a}_2(\vartheta) & \boldsymbol{B}_2 \\ \boldsymbol{a}_3(\vartheta) & \boldsymbol{B}_3 \end{bmatrix} \in \bar{\mathcal{A}}_N \qquad \bar{\boldsymbol{A}} = \boldsymbol{A}\boldsymbol{T} = \begin{bmatrix} a_1(\vartheta) & \boldsymbol{0}^{\mathrm{T}} \\ \boldsymbol{a}_2(\vartheta) & \bar{\boldsymbol{B}} \\ \boldsymbol{a}_3(\vartheta) & \boldsymbol{I}_{N-1} \end{bmatrix}$$



$$\boldsymbol{R} = \boldsymbol{A}\boldsymbol{P}\boldsymbol{A}^{\mathrm{H}} + \nu \boldsymbol{I}_M \qquad \boldsymbol{R} = \boldsymbol{A}\boldsymbol{P}\boldsymbol{A}^{\mathrm{H}} + \nu \boldsymbol{I}_M \qquad \boldsymbol{R} = \bar{\boldsymbol{A}}\bar{\boldsymbol{P}}\bar{\boldsymbol{A}}^{\mathrm{H}} + \nu \boldsymbol{I}_M$$

- Structure of the desired direction is unaltered.
- Non-redundancy of the parameterization is ensured.

Partial Relaxation Techniques
Expression of the PR-CRB

Recall the conventional Crámer-Rao Bound

$$\boldsymbol{C}_{\text{sto}}(\boldsymbol{\theta}) = \frac{\nu}{2T}\text{Re}\left\{\boldsymbol{M} \odot \left(\boldsymbol{D}^{\mathsf{H}}\boldsymbol{\Pi}_{\boldsymbol{A}}^{\perp}\boldsymbol{D}\right)\right\}^{-1}$$

- $\boldsymbol{M} = \left(\boldsymbol{P}\boldsymbol{A}^{\mathsf{H}}\boldsymbol{R}^{-1}\boldsymbol{A}\boldsymbol{P}\right)^{\mathsf{T}} = \begin{bmatrix} M_{11} & \boldsymbol{M}_{21}^{\mathsf{H}} \\ \boldsymbol{M}_{21} & \boldsymbol{M}_{22} \end{bmatrix}$
- $\boldsymbol{D} = \left[\frac{\mathrm{d}\boldsymbol{a}(\theta_1)}{\mathrm{d}\theta}, \ldots, \frac{\mathrm{d}\boldsymbol{a}(\theta_N)}{\mathrm{d}\theta}\right] = [\boldsymbol{d}, \boldsymbol{D}_2]$

Crámer-Rao Bound for $\vartheta = \theta_1$ under the PR model

$$C_{\text{PR-CRB}}(\vartheta) = \frac{\nu}{2T}\left(\left(M_{11} - \boldsymbol{M}_{21}^{\mathsf{H}}\boldsymbol{M}_{22}^{-1}\boldsymbol{M}_{21}\right)\boldsymbol{d}^{\mathsf{H}}\boldsymbol{\Pi}_{\boldsymbol{A}}^{\perp}\boldsymbol{d}\right)^{-1}.$$

Crámer-Rao Bounds

$$\boldsymbol{C}_{\text{sto}}(\boldsymbol{\theta}) = \frac{\nu}{2T} \text{Re}\left\{\boldsymbol{M} \odot \left(\boldsymbol{D}^{\mathsf{H}} \boldsymbol{\Pi}_{\boldsymbol{A}}^{\perp} \boldsymbol{D}\right)\right\}^{-1}$$

$$C_{\text{PR-CRB}}(\vartheta) = \frac{\nu}{2T}\left(\left(M_{11} - \boldsymbol{M}_{21}^{\mathsf{H}} \boldsymbol{M}_{22}^{-1} \boldsymbol{M}_{21}\right) \boldsymbol{d}^{\mathsf{H}} \boldsymbol{\Pi}_{\boldsymbol{A}}^{\perp} \boldsymbol{d}\right)^{-1}$$

- PR-CRB is always lower-bounded by the conventional CRB.

- In the case of high SNR and uncorrelated source signals, the two bounds are approximately equal.

Recall the null-spectrum of PR-DML and PR-WSF estimator

$$f_{\text{PR-DML}}(\boldsymbol{a}) = \sum_{k=N}^{M} \lambda_k \left(\boldsymbol{\Pi}_{\boldsymbol{a}}^{\perp} \hat{\boldsymbol{R}} \right)$$

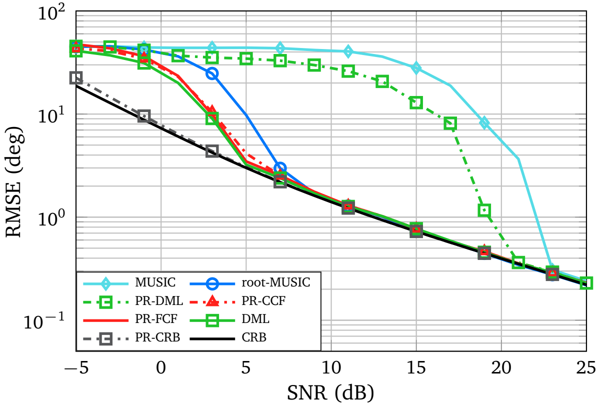
$$f_{\text{PR-WSF}}(\boldsymbol{a}) = \lambda_N \left(\boldsymbol{\Pi}_{\boldsymbol{a}}^{\perp} \hat{\boldsymbol{U}}_{\text{s}} \boldsymbol{W} \hat{\boldsymbol{U}}_{\text{s}}^{\mathsf{H}} \right)$$

Asymptotically as $T \to \infty$,

- The mean-square error of PR-WSF achieves PR-CRB for all positive definite weighting matrix $\boldsymbol{W}$.
- The mean-square error of PR-WSF, PR-DML and MUSIC are identical.

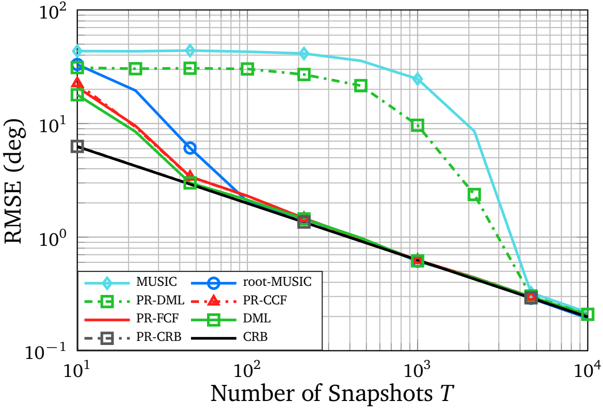
Uncorrelated Source Signals

$M = 5,\ \boldsymbol{\theta} = [135°, 140°]^\mathsf{T},\ T = 150$



Uncorrelated Source Signals

$$M = 5,\ \boldsymbol{\theta} = [135^\circ, 140^\circ]^\mathsf{T},\ \mathrm{SNR} = 10\mathrm{dB}$$



Concluding Remarks

Problem relaxation

Deliberately ignoring part of the prior knowledge is a powerful approach to make complicated estimation problems computationally tractable (without sacrificing much performance).

- Partial array geometry relaxation.
- Relaxation of interference structure.

Extensions?

- Revisit established algorithms for more advanced measurement models and design your own relaxation algorithms!!!

Use PR models in the performance analysis:

- Understand which model information is relaxed in a particular algorithm.

MATLAB Code is available at

https://git.rwth-aachen.de/minh.trinh_hoang/sam-2020-tutorial-code

Thank you for your attention!

References I

- John Stone Stone, “Method of determining the direction of space telegraph signals,” US Patent 716,134, December 16, 1902.
- John Stone Stone, “Apparatus for determining the direction of space telegraph signals,” US Patent 716,135, December 16, 1902.
- Lee de Forest, “Wireless signaling apparatus,” US Patent 771,819, October 11, 1904.
- G. Marconi, John Ambrose Fleming,“On methods whereby the radiation of electric waves may be mainly confined to certain directions, and whereby the receptivity of a receiver may be restricted to electric waves emanating from certain directions,” Proc. R. Soc. Lond. 77, pp.413–421.

References II

- E. Bellini, A. Tosi, “ System of directed wireless telegraphy,” US Patent 943,960, December 21, 1909.
- E. Bellini, A. Tosi, “ Directed wireless telegraphy,” US Patent 948,086, February 1, 1910.
- F. Adcock, “Improvement in Means for Determining the Direction of a Distant Source of Electro-magnetic Radiation,” UK Patent 130,490, August 7, 1919.
- R. Keen, “Wireless Direction Finding,” London, UK: Iliffe & Son Dorset House, 1938.
- H. G. Schantz, “On the origins of RF-based location” 2011 IEEE Topical Conference on Wireless Sensors and Sensor Networks, Phoenix, AZ, pp. 21-24, 2011.

References III

- D. Malioutov, M. Cetin, A. Willsky, “A sparse signal reconstruction perspective for source localization with sensor arrays,” IEEE Transactions on Signal Processing, vol. 53, no. 8, pp. 3010–3022, 2005.
- J. Capon, R.J . Greenfield, R. T. Lacoss, “Off-line signal processing results for the large aperture seismic array,” Mass. Inst. Tech. Lincoln Lab., Lexington, Mass., Tech. Note 1966-37, July 11, 1966
- J. Capon, R.J . Greenfield, R.J. Kolker, “Multidimensional maximum-likelihood processing of a large aperture seismic array,” Proceedings of the IEEE, vol. 55, no. 2, pp. 192-211, 1967.
- R.O. Schmidt, “Multiple emitter location and signal parameter estimation,” Proc. RADC Spectrum Estimation Workshop, Griffiths AFB, Rome, New York, pp. 243-258, 1979.

References IV

- R.O. Schmidt, “Multiple emitter location and signal parameter estimation,” PhD Dissertation, Stanford University, Stanford, California, 1981.
- R.O. Schmidt, “Multiple emitter location and signal parameter estimation,” IEEE Transactions on Antennas and Propagation, vol. 34, no. 3,pp.276–280, 1986.
- G. Bienvenu, L. Kopp, “Principle de la goniometrie passive adaptive,” in Proceedings of the 7’eme Colloque GRESIT, Nice, France, pp. 106/1–106/10, 1979.
- A.J. Barabell, “Improving the resolution performance of eigenstructure-based direction-finding algorithms,” Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, Boston, MA, pp. 336–339, 1983.

References V

- R. Roy, A. Paulraj, T. Kailath, “ESPRIT - a subspace rotation approach to estimation of parameters of cisoids in noise,” IEEE Trans. Acoust. Speech Signal Process., vol 34, pp. 1340–1342, 1986.
- P. Stoica, K. Sharman, “Maximum likelihood methods for direction-of-arrival estimation,” IEEE Trans. Acoust. Speech Signal Process., vol. 38, no. 7, 1132–1143, 1990.
- I. Ziskind, M. Wax,“Maximum likelihood localization of multiple sources by alternating projection,” in IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 36, no. 10, pp. 1553-1560, 1988
- J. Böhme, “Estimation of source parameters by Maximum Likelihood and nonlinear regression,” Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, volume 9, pages 271–274, 1984.

References VI

- M. Wax. “Detection and Localization of Multiple Sources in Noise with Unknown Covariance.” IEEE Trans. Acoust. Speech Signal Process, vol. 40, no. 1, pp. 245-249, 1992.
- P. Stoica, A. Nehorai, “MUSIC, maximum likelihood and Cramer-Rao bound”, IEEE Trans. Acoust. Speech Signal Process., vol. 37, no. 5, pp. 720–741, 1989.
- J.F. Böhme, “Estimation of spectral parameters of correlated signals in wavefields,” Signal Process., vol. 11, pp. 329–337, 1986.
- J.F. Böhme, “Separated estimation of wave parameters and spectral parameters by maximum likelihood,” Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, Tokyo, Japan, 1986, pp. 2818–2822.

References VII

- Y. Bresler, “Maximum likelihood estimation of linearly structured covariance with application to antenna array processing,” in Proceedings of the 4th ASSP Workshop on Spectrum Estimation and Modeling, Minneapolis, MN, pp. 172–175, 1988.
- P. Stoica, T. Söderström, “On reparametrization of loss functions used in estimation and the invariance principle,” Signal Processing, vol. 17, no. 4, pp. 383-387, 1989.
- R. Kumaresan, D.W. Tufts, “Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise,” IEEE Trans. Acoust. Speech Signal Process. ASSP-30, pp. 833–840, 1982.

References VIII

- M. I. Miller and D. R. Fuhrmann,“Maximum-likelihood narrow-band direction finding and the EM algorithm” in IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 38, no. 9, pp. 1560-1577, 1990.
- D. R. Hunter and K. Lange, “A tutorial on MM algorithms,” Am. Stat., vol. 58, no. 1, pp. 30–37, 2004.
- A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” J. R. Stat. Soc. Series B, vol. 39, no. 1, pp. 1–38, 1977.
- M. Hong, M. Razaviyayn, Z. Luo and J. Pang, “A Unified Algorithmic Framework for Block-Structured Optimization Involving Big Data: With applications in machine learning and signal processing” in IEEE Signal Processing Magazine, vol. 33, no. 1, pp. 57-77, Jan. 2016

References IX

- B. Ottersten, P. Stoica, R. Roy, "Covariance matching estimation techniques for array signal processing applications," Digital Signal Process., vol. 8, no. 3, pp. 185–210, 1998.
- Y. Bresler, V.U. Reddy, T. Kailath, "Optimum beamforming for coherent signal and interferences," IEEE Trans. Acoust. Speech Signal Process. 36 (6) (1988) 833–843.
- B. Ottersten, M. Viberg, T. Kailath, "Analysis of subspace fitting and ML techniques for parameter estimation from sensor array data," IEEE Trans. Signal Process., vol. 40, no. 3, pp. 590–599, 1992.
- P. Stoica, A. Nehorai, "Performance study of conditional and unconditional direction of arrival estimation," IEEE Trans. Acoust. Speech Signal Process., vol. 38, no. 19, pp. 1783–1795, 1990.

References X

- P. Stoica and A. Nehorai, “MODE, maximum likelihood and Cramer-Rao bound: conditional and unconditional results,” Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, Albuquerque, NM, USA, pp. 2715-2718 vol.5, 1990
- M. Viberg, B. Ottersten, “Sensor Array Processing Based on Subspace Fitting,” IEEE Transactions on Signal Processing, vol. 39, no. 5, pp. 1110–1121, 1991.
- E.L. Lehmann, G. Casella, “Theory of Point Estimation,” second ed., Springer, New York, 1998.
- A. G. Jaffer, “Maximum likelihood direction finding of stochastic sources: a separable solution,“ Proc. International Conference on Acoustics, Speech, and Signal Processing, New York, NY, USA, pp. 2893-2896, 1988.

References XI

- C. Steffens, M. Pesavento, M. E. Pfetsch, “A compact formulation for the $\ell_{2,1}$ mixed-norm minimization problem,” IEEE Transactions on Signal Processing, vol. 66, no. 6, pp. 1483-1497, 2018.
- Z. Yang, Jian Li, Petre Stoica, Lihua Xie, “Chapter 11 - Sparse methods for direction-of-arrival estimation,” Editor(s): Rama Chellappa, Sergios Theodoridis, Academic Press Library in Signal Processing, Volume 7, Academic Press, pp. 509-581, 2018.
- Yuan, M. and Lin, Y. L. Ming Yuan, “Model selection and estimation in regression with grouped variables,” Journal of the Royal Statistical Society. Series B (Statistical Methodology), vol. 68, no. 1, pp. 49–67, 2006.

References XII

- M. Trinh-Hoang, W. Ma and M. Pesavento, “Cramer-Rao Bound for DOA Estimators under the Partial Relaxation Framework: Derivation and Comparison,” IEEE Transactions on Signal Processing, 2020.
- M. Trinh-Hoang, W. Ma and M. Pesavento, “A Partial Relaxation DOA Estimator Based on Orthogonal Matching Pursuit,” Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, Barcelona, Spain, pp. 4806-4810, 2020.
- M. Trinh-Hoang, W. Ma and M. Pesavento, “Partial Relaxation Approach: An Eigenvalue-Based DOA Estimator Framework,” IEEE Transactions on Signal Processing, vol. 66, no. 23, pp. 6190-6203, 2018.
- M.S.Bartlett, “Smoothing Periodograms from Time-Series with Continuous Spectra,” Nature, 161:686-687, 1948

References XIII

E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen, "LAPACK Users' Guide", 3rd ed. Philadelphia, Society for Industrial and Applied Mathematics, 1999.