Enumeration for a Large Number of Sources Based on a Two-Step Difference Operation of Linear Shrinkage Coefficients

Bijo John	Sai Shashank GP	Aditya Khandavally	Haricharan B
EE20B021	EE20B040	EE20B062	EP21B015

Jan-May '25

Problem Formulation

- m - # of Receiver Antennas, n - # of Samples
- Signals from d sources. $\mathbf{s}_t \sim \mathcal{CN}(0, \Sigma_s) \in \mathbb{C}^d$
- Matrix of Steering vectors $\mathbf{A} \ \in \mathbb{C}^{m \times d}$
- Observed Signal: $\mathbf{x}_t = \mathbf{A}\mathbf{s}_t + \mathbf{w}_t \in \mathbb{C}^m$, $\mathbf{w}_t \sim \mathcal{CN}(0, \sigma_w^2 I_m)$
- $m \to \infty$, $n \to \infty$, $c = \frac{m}{n}$ is constant.

Problem Statement

Given the signals $\mathbf{x}_t$ for n time instants, can we determine the number of sources d?

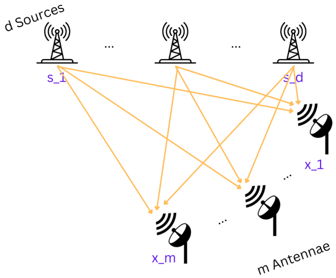


Figure: Schematic Diagram

Linear Shrinkage-based methods

- Solve the optimization problem for k presumed number of sources:

$$\min_{\alpha^{(k)}} \quad g(\alpha^{(k)}) = \mathbb{E}\left[\left\|\mathbf{R}^{(k)} - \mathbf{\Sigma}_{\mathcal{N}}^{(k)}\right\|_F^2\right]$$

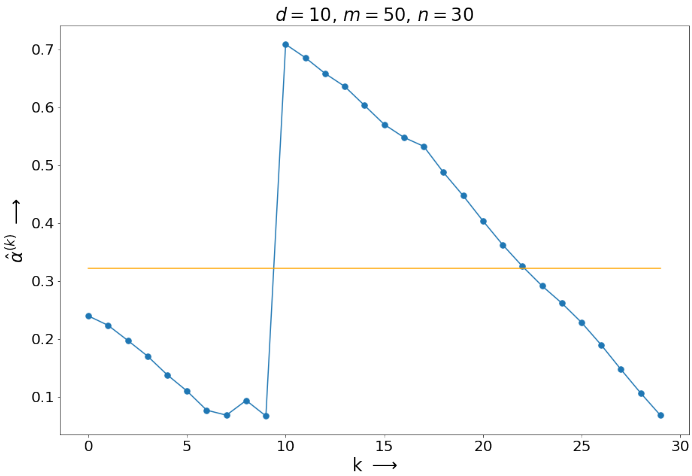
$$\text{subject to} \quad \mathbf{R}^{(k)} = \alpha^{(k)}\mu_k \mathbf{I}_{m-k} + \left(1 - \alpha^{(k)}\right)\mathbf{S}_{\mathcal{N}}^{(k)}$$

$\mathbf{\Sigma}_{\mathcal{N}}^{(k)} = \text{diag}(\lambda_{k+1}, \ldots, \lambda_m)$ and $\mathbf{S}_{\mathcal{N}}^{(k)} = \text{diag}(l_{k+1}, \ldots, l_m)$
$\hat{\alpha}^{(k)} \in [0, 1]$ are the linear shrinkage coefficients
$\mu_k = \text{tr}(\mathbf{\Sigma}_{\mathcal{N}}^{(k)})/(m-k)$ is the regularization term

- Let $\gamma_k = \frac{\frac{1}{m-k}\sum_{i=k+1}^{m} l_i^2}{\left(\frac{1}{m-k}\sum_{i=k+1}^{m} l_i\right)^2}$ and $\bar{m} = \min(m, n)$
- Solution is: [Chen et al.]

$$\hat{\alpha}^{(k)} = \min\left(\frac{\gamma_k + (m-k)}{(n+1)(\gamma_k - 1)}, 1\right)$$

Structure of $\hat{\alpha}^{(k)}$



Recap: Algorithms

Algorithm	**Key Principle**	**Limitations**
SCD_{heur}	Largest indexed $\hat{\alpha}^{(k)}$ below the threshold.	**Proposition 1:** Ineffective when $c \to 1$ or $c \geq 1$ and for large d.
First-Step	Maximum difference in LS coefficients $\hat{d}_1 = 1 + \underset{k=1,..,\min(n,m)}{\operatorname{argmax}} \hat{\Delta}^{(k)}$	Effective only in high SNR. Requires $m, m-k \to \infty$.
Second-Step	Modified difference with previous differences $\hat{d}_2 = 1 + \underset{k=i,..,\bar{d}}{\operatorname{argmax}} \dot{\Delta}^{(k)} \equiv \hat{\Delta}^{(k)} + \tau \bar{\Delta}_g^k$	Requires proper i selection.

$$\hat{\Delta}^{(k)} \equiv \hat{\alpha}^{(k)} - \hat{\alpha}^{(k-1)},\ \hat{\Delta}_g^v = \max\left(0, \hat{\Delta}^{(v)}\right),\ \bar{\Delta}_g^k = \frac{\sum_{v=0}^{k-1} \hat{\Delta}_g^v}{k},\ \bar{d} \geq d-1.$$

Propositions Schematic

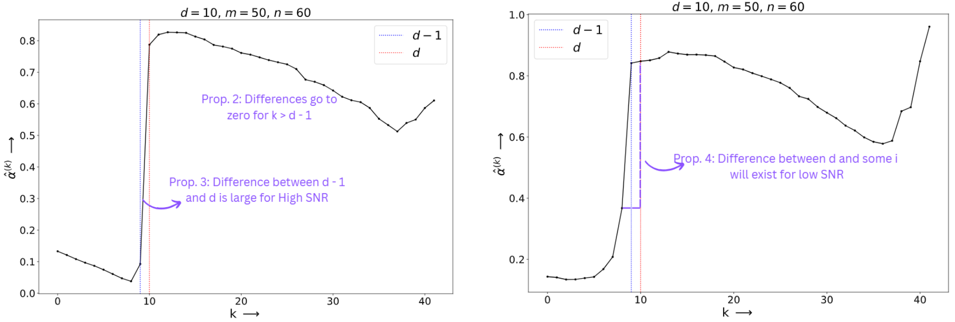


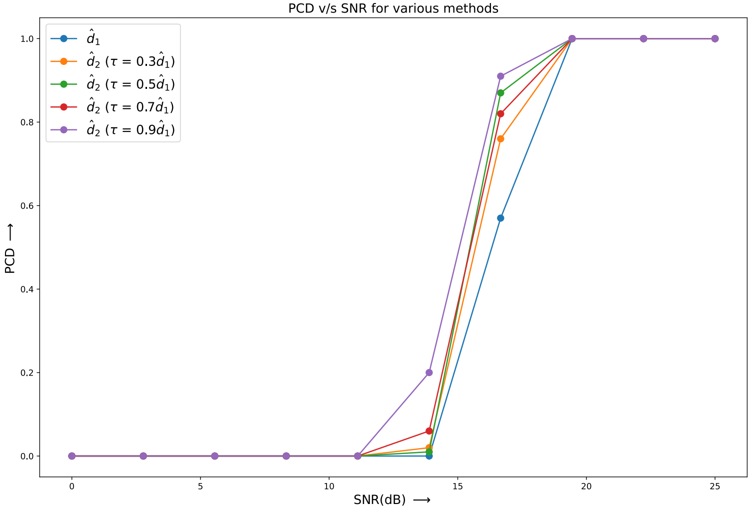
Figure: Schematic of the Propositions for (a) High SNR $= 30dB$ (b) Low SNR $= 0dB$

Propositions

- The variation of $\alpha^{(k)}$ with respect to k also relies on c, provided that $k \in [d, \bar{m} - 1]$, and SCD_{heur} is effective only if $\forall k, \gamma_k$ is less than a threshold function g_c. When $c \to 1$ or $c \geq 1$, this condition might be violated, potentially leading to failed estimation.
- The difference between noise LS coefficients $\hat{\alpha}^{(k)}$ and $\hat{\alpha}^{(k+1)} \to 0$ for $k > d - 1$ provided that $m, n, m - k \to \infty$.
- There is a clear gap between $\hat{\alpha}^{(d-1)}$ and $\hat{\alpha}^{(d)}$ for relatively high SNR (β) cases, provided that $m, n, m - d \to \infty$ and $m/n = c \in (0, \infty)$.
- For low SNRs, the clear gap between $\hat{\alpha}^{(d-1)}$ and $\hat{\alpha}^{(d)}$ does not exist. However, for relatively low or medium SNRs where $\rho_d \beta m < \mathcal{O}(m - d)$ but $\sqrt{\sum_{j=i}^{d} \rho_j^2} \beta m > \mathcal{O}(m - d)$, $i < d$, there must be a gap between $\hat{\alpha}^{(i)}$ and $\hat{\alpha}^{(d)}$.

Note: $l_i = \sigma_w^2 \rho_i \beta m$

Recap: Results



Directions from Mid-Review

- Understanding the significance of Proposition 4
- Understanding the intuition behind the second-step estimate
- Determining how to choose i, which is critical in the second-step estimate
- Applying the estimator for coloured noise
- Attempt using this method in DoA estimation

Our mails to the author regarding the selection principle for i did not elicit a response.

Second-Step difference

Significance of Proposition 4

Proposition 4

For sufficiently low SNRs, the clear gap between $\hat{\alpha}^{(d-1)}$ and $\hat{\alpha}^{(d)}$ (mentioned in Proposition 3) does not exist. However, for relatively low or medium SNRs where $\rho_d \beta m < \mathcal{O}(m-d)$ but $\sqrt{\sum_{v=i}^{d} \rho_v^2} \beta m \geq \mathcal{O}(m-d)$, $i < d$, there must exist a gap between $\hat{\alpha}^{(i)}$ and $\hat{\alpha}^{(d)}$.

- **Essential claim:** There is some $i < d$ such that the difference between $\hat{\alpha}^{(i)}$ and $\hat{\alpha}^{(d)}$ is significant.

$$\hat{\alpha}^{(d)} - \hat{\alpha}^{(i)} = \hat{\Delta}^{(d-1)} + \hat{\Delta}^{(d-2)} + \cdots + \hat{\Delta}^{(i)}$$

- Proposition 4 implies that at least one of the $\hat{\Delta}^{(k)}$ terms is significant.
- Second-Step difference tries to make $\hat{\Delta}^{(d-1)}$ the largest term.

Second-Step difference: Why it works

We will demonstrate why the second-step difference works in the following slides.

- Step 1: $\dot{\Delta}^{(d-1)} > \dot{\Delta}^{(k)} \ \forall \ k \in [i, d-1]$
- Step 2: $\dot{\Delta}^{(d)} > \dot{\Delta}^{(k)} \ \forall \ k > d$
- Step 3: $\dot{\Delta}^{(d-1)} > \dot{\Delta}^{(d)}$

Second-Step difference: When $k < d$

- The parameter i is chosen such that $\hat{\Delta}^{(k)} > \bar{\Delta}_g^k$ holds $\forall k \in [i, d-1]$
- Since $\hat{\alpha}^{(k)} \to 0$ for small k, it can be seen that such an interval must exist in most cases.
- This has also been verified through a large number of simulations.
- For $k \in [i, d-1]$,

$$\bar{\Delta}_g^k = \frac{\hat{\Delta}_g^{k-1} + \sum_{v=0}^{k-2} \hat{\Delta}_g^v}{k} > \frac{(\frac{1}{k-1} + 1) \sum_{v=0}^{k-2} \hat{\Delta}_g^v}{k} = \bar{\Delta}_g^{k-1}$$

- Defining $\tilde{\Delta}^k = \bar{\Delta}_g^k - \bar{\Delta}_g^{k-1} > 0$, we have

$$\dot{\Delta}^{(i)} < \dot{\Delta}^{(i+1)} < \cdots < \dot{\Delta}^{(d-1)}$$

as long as $\hat{\Delta}^{(k-1)} < \hat{\Delta}^{(k)} + \tau \tilde{\Delta}^k$ holds for $k \in [i, d-1]$

Second-Step difference: When $k \geq d$

For $k \geq d$:

- By Proposition 2, $\hat{\Delta}^{(k)} \to 0$.
- The definition of $\bar{\Delta}_g^{(k)}$ gives:

$$\bar{\Delta}_g^{(k)} = \frac{\sum_{v=0}^{k-1} \hat{\Delta}_g^{(v)}}{k} = \frac{\sum_{v=0}^{d-1} \hat{\Delta}_g^{(v)} + \sum_{v=d}^{k-1} \hat{\Delta}_g^{(v)}}{k} \approx \frac{\sum_{v=0}^{d-1} \hat{\Delta}_g^{(v)}}{k}$$

- Substituting in the expression for $\dot{\Delta}^{(k)}$:

$$\dot{\Delta}^{(k)} = \hat{\Delta}^{(k)} + \tau \bar{\Delta}_g^{(k)} \approx \frac{\tau}{k} \sum_{v=0}^{d-1} \hat{\Delta}_g^{(v)}$$

- Since $k \geq d$, we can bound this by:

$$\dot{\Delta}^{(k)} \leq \frac{\tau}{d} \sum_{v=0}^{d-1} \hat{\Delta}_g^{(v)} \leq \dot{\Delta}^{(d)}$$

Second-Step difference: Comparison between $d-1$ and d

- Now, we compare between $\dot{\Delta}^{(d-1)}$ and $\dot{\Delta}^{(d)}$
- By Proposition 2, $\hat{\Delta}^{(d)} \to 0$
- Hence, we have

$$\begin{aligned}
\dot{\Delta}^{(d)} &\approx \frac{\tau}{d}\left(\sum_{v=0}^{d-1} \hat{\Delta}_g^v\right) \\
&= \frac{\tau}{d}\Delta_g^{d-1} + \frac{\tau(d-1)}{d}\frac{\sum_{v=0}^{d-2}\hat{\Delta}_g^v}{d-1} \\
&< \Delta_g^{d-1} + \tau\bar{\Delta}_g^k \\
&= \dot{\Delta}^{(d-1)}
\end{aligned}$$

- Thus, the second-step difference method correctly chooses $\hat{d}_2 = 1 + (d-1) = d$, provided the assumptions hold.

Second-Step difference: Choice of i

- $\forall i \leq k < d$, it must hold that $\hat{\Delta}^{(k-1)} < \hat{\Delta}^{(k)} + \tau(\bar{\Delta}_g^k - \bar{\Delta}_g^{k-1})$
- We choose i as the first index k such that $\hat{\Delta}^{(k-1)} < \hat{\Delta}^{(k)} + \tau(\bar{\Delta}_g^k - \bar{\Delta}_g^{k-1})$

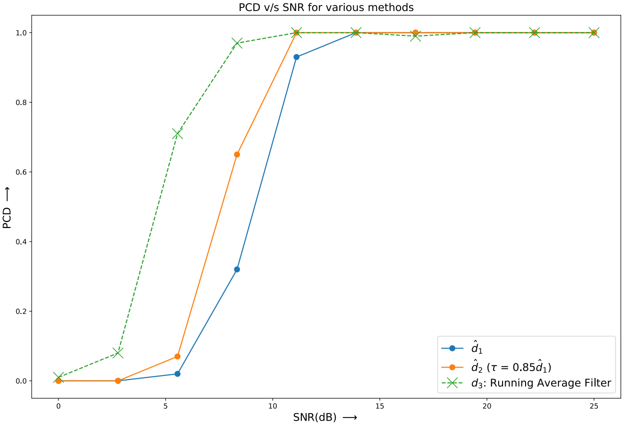
Our attempts: Imagining the Problem as Change Detection

- We have a series of $\hat{\alpha}^{(k)}$ values such that there is a significant difference between $\hat{\alpha}^{(d)}$ and $\hat{\alpha}^{(i)}$ for some i
- This is a change detection problem, and there exist classical techniques to achieve this.
- We tried out two change detection methods: Inflection-point based filter and Running average filter.

Applying Filters over $\hat{\alpha}^k$

Running average filter

$$\hat{d}_3 = \underset{k=i,..,\bar{d}}{\text{argmax}}\ \rho^{(k)} \equiv \hat{\alpha}^{(k)} - \frac{\sum_{i=0}^{k-1} \hat{\alpha}^{(i)}}{k}$$



Explanation of Running Average Filter

We consider the following cases:

- Step 1: $\rho^{(k)} \approx 0$ for $k < i$
- Step 2: $\rho^{(k+1)} > \rho^{(k)}$ for $k \in [i, d-1]$
- Step 3: $\rho^{(k+1)} < \rho^{(k)}$ for $k \geq d$

Running Average Filter: Step 1

- For $k < i$, $\rho^{(k)} \approx 0$ as $\hat{\alpha}^{(0)}, \ldots, \hat{\alpha}^{(i)} \to 0$.

Running Average Filter: Step 2

- For $k \in [i, d-1]$:

It can be deduced that $\hat{\Delta}^{(k)} \gg \bar{\Delta}_g^k$ for a very small range of $k \in [i, d-1]$, which has also been verified via a large number of simulations. Thus,

$$\hat{\Delta}^k > \frac{\hat{\alpha}^{(k)} - \hat{\alpha}^{(0)}}{k}$$
$$\implies \hat{\alpha}^{(k+1)} - \hat{\alpha}^{(k)} > \frac{\hat{\alpha}^{(k)}}{k} \text{ (Since } \hat{\alpha}^{(0)} \text{ is small)}$$

Some rearranging yields $\rho^{(k+1)} > \rho^{(k)}\ \forall\ k \in [i, d-1]$, where we use:

$$\rho^{(k)} = \left(\hat{\alpha}^{(k)} - \frac{\sum_{j=0}^{k-1} \hat{\alpha}^{(j)}}{k}\right) = \left(\frac{\sum_{j=0}^{k-1} (\hat{\alpha}^{(k)} - \hat{\alpha}^{(j)})}{k}\right)$$

Running Average Filter: Step 3

- For $k \geq d$:

According to Proposition 2, we have $\hat{\Delta}^k \to 0$. This implies $\hat{\alpha}^{(k+1)} \approx \hat{\alpha}^{(k)}$.

$$\begin{aligned}
\rho^{(k+1)} &= \frac{\sum_{j=0}^{k} \hat{\alpha}^{k+1} - \hat{\alpha}^j}{k+1} \\
&\approx \frac{\sum_{j=0}^{k-1} \hat{\alpha}^{k} - \hat{\alpha}^j}{k+1} \quad (\text{Since } \hat{\alpha}^{k+1} \approx \hat{\alpha}^{k}) \\
&< \frac{\sum_{j=0}^{k-1} \hat{\alpha}^{k} - \hat{\alpha}^j}{k} \\
&= \rho^{(k)}
\end{aligned}$$

Therefore, $\rho^{(k+1)} < \rho^{(k)} \ \forall \ k \geq d$.
From this, we see that in the low-medium SNR when Prop. 4 holds, the method can be proven to predict d with a high accuracy.

BIC-based Source Enumeration

Another paper dealing with $m \to \infty$

On Accurate Source Enumeration: A New Bayesian Information Criterion, by Xiaochuan Ke, Yuan Zhao, Lei Huang in 2021

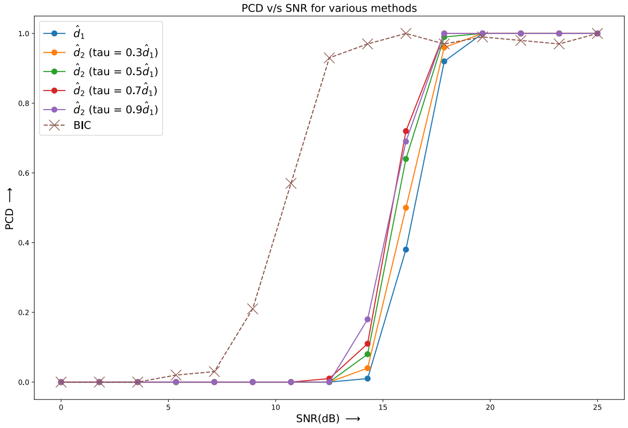
- Assuming that each k (# of sources) are equally likely, we maximise likelihood function.

$$\hat{d} = \underset{k=1,..,\bar{m}}{\operatorname{argmax}} f\left(X|d=k\right)$$

- The likelihood function is approximated to second order around ML estimate of $\hat{\theta}_k$ (unknown parameter), gives

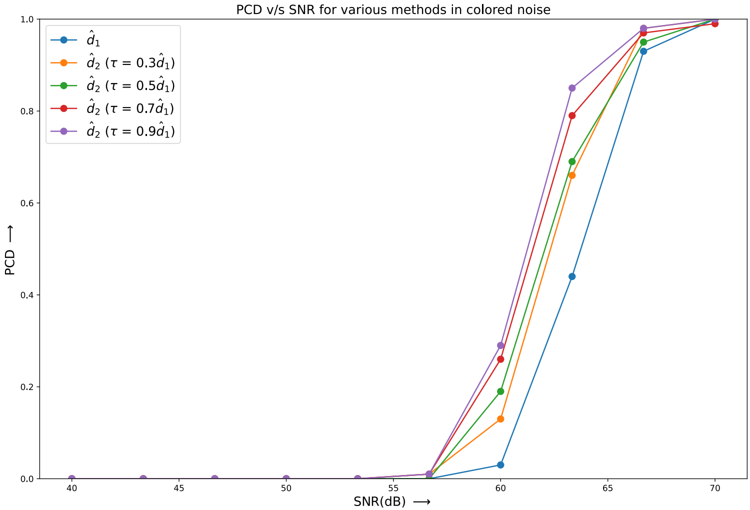
$$\boxed{\mathrm{BIC} = \mathrm{LLF} + \mathrm{PPF} + \mathrm{PF}}$$

BIC based Source Enumeration



Coloured Noise

First-step and second-step estimators in case of colored noise: A case of misspecified model.



Conclusions

- In the low SNR regime, incorporating the running average of the first-step differences scales the first-step difference to maximize the d^{th} term (as shown in Propositions 2 and 4), leading to improved performance of $\hat{d}_2$.
- We proposed an empirical method for selecting the index i in $\hat{d}_2$ calculation, based on the correctness proof of $\hat{d}_2$.
- We experimented with a running average filter on the sequence of shrinkage coefficients and proved correctness.
- We observe that in the coloured noise setting, the algorithms demonstrate good performance only in the high SNR regime.
- We also explored a source enumeration technique that leverages Bayesian information criterion (BIC) to improve PCD.