Fast Covariance Learning algorithms for Sparse Bayesian Learning

Esa Ollila

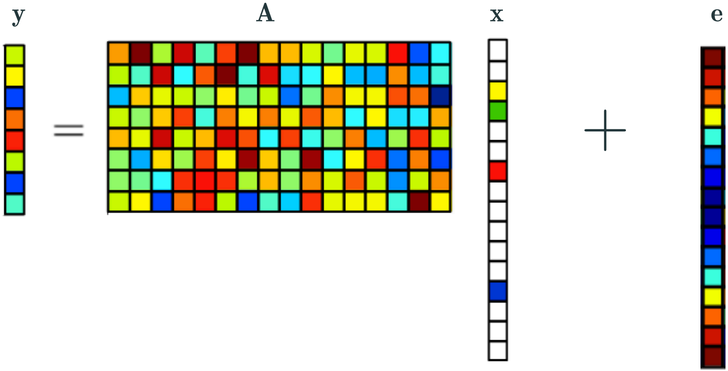
Seminar at Tésa, May 18th, 2026

Table-of-contents

Single Measurement Vector Model (SMV)

Underdetermined linear model:

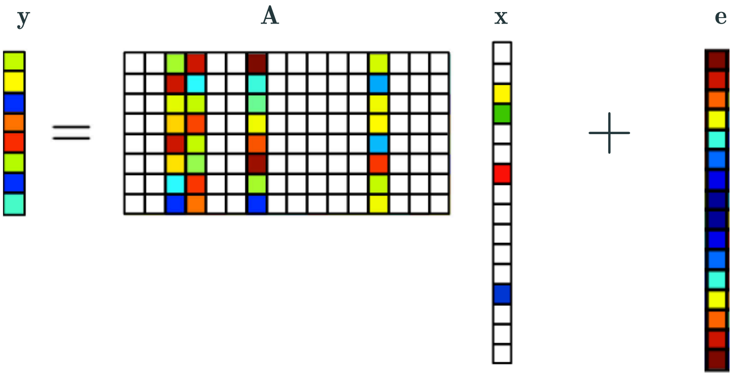
$$\underset{M \times 1}{\mathbf{y}} = \underset{M \times N}{\mathbf{A}} \underset{N \times 1}{\mathbf{x}} + \underset{N \times 1}{\mathbf{e}}$$



Single Measurement Vector Model (SMV)

Underdetermined linear model:

$$\underset{M\times 1}{\mathbf{y}} = \underset{M\times N}{\mathbf{A}}\ \underset{N\times 1}{\mathbf{x}} + \underset{N\times 1}{\mathbf{e}}$$



Greedy approach:

1. Identify the active columns of measurement matrix $\mathbf{A}$, i.e., the support of signal vector $\mathbf{x}$.
2. Estimate non-zero elements of $\mathbf{x}$ via least squares

Optimization approach:

$$\arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1$$

Multiple Measurement Vectors Model (MMV) [DE11]

We acquire L measurement vectors:

$$\mathbf{y}_i = \mathbf{A}\mathbf{x}_i + \mathbf{e}_i, \quad i = 1, \dots, L$$

$$\Leftrightarrow \underset{M \times L}{\mathbf{Y}} = \underset{M \times N}{\mathbf{A}} \; \underset{N \times L}{\mathbf{X}} + \underset{M \times L}{\mathbf{E}}$$

- $\mathbf{Y} = (\mathbf{y}_1 \ \cdots \ \mathbf{y}_L) :=$ matrix of L measurement vectors
- $\mathbf{X} = (\mathbf{x}_1 \ \cdots \ \mathbf{x}_L) :=$ unobserved **signal matrix**
- $\mathbf{E} = (\mathbf{e}_1 \ \cdots \ \mathbf{e}_L) :=$ unobserved **noise matrix**

Multiple Measurement Vectors Model (MMV) [DE11]

We acquire L measurement vectors:

$$\mathbf{y}_i = \mathbf{A}\mathbf{x}_i + \mathbf{e}_i, \quad i = 1, \dots, L$$

$$\Leftrightarrow \underset{M \times L}{\mathbf{Y}} = \underset{M \times N}{\mathbf{A}} \underset{N \times L}{\mathbf{X}} + \underset{M \times L}{\mathbf{E}}$$

***Key assumption*:**
signals $\mathbf{x}_i$ are jointly K-sparse:

$$\|\mathbf{X}\|_0 = |\mathrm{rowsupp}(\mathbf{X})| = K$$

- $\mathbf{Y} = (\mathbf{y}_1 \ \cdots \ \mathbf{y}_L) :=$ matrix of L measurement vectors
- $\mathbf{X} = (\mathbf{x}_1 \ \cdots \ \mathbf{x}_L) :=$ unobserved **signal matrix**
- $\mathbf{E} = (\mathbf{e}_1 \ \cdots \ \mathbf{e}_L) :=$ unobserved **noise matrix**

Multiple Measurement Vectors Model (MMV) [DE11]

We acquire L measurement vectors:

$$\mathbf{y}_i = \mathbf{A}\mathbf{x}_i + \mathbf{e}_i, \quad i = 1, \dots, L$$

$$\Leftrightarrow \underset{M \times L}{\mathbf{Y}} = \underset{M \times N}{\mathbf{A}} \underset{N \times L}{\mathbf{X}} + \underset{M \times L}{\mathbf{E}}$$

- $\mathbf{Y} = (\mathbf{y}_1 \;\cdots\; \mathbf{y}_L) :=$ matrix of L measurement vectors
- $\mathbf{X} = (\mathbf{x}_1 \;\cdots\; \mathbf{x}_L) :=$ unobserved **signal matrix**
- $\mathbf{E} = (\mathbf{e}_1 \;\cdots\; \mathbf{e}_L) :=$ unobserved **noise matrix**

Key assumption:
signals $\mathbf{x}_i$ are jointly K-sparse:

$$\|\mathbf{X}\|_0 = |\mathrm{rowsupp}(\mathbf{X})| = K$$

Goal: estimate

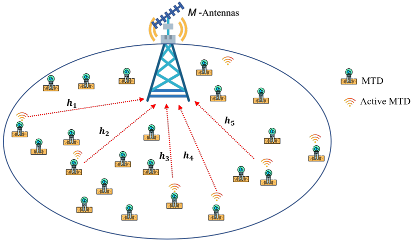
$$\mathcal{M} = \mathrm{rowsupp}(\mathbf{X}) = (i_1, \dots, i_K)$$

and signal matrix $\mathbf{X}$

given $\mathbf{Y}$, measurement matrix $\mathbf{A}$ (, and possibly sparsity level K).

Application 1: active user detection for massive random access

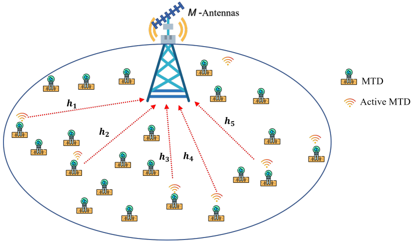
- BS with M antennas, N machine type devices (MTDs)
- In a coherence interval, only $K \ll N$ MTD-s are active.
- $\mathbf{a}_n$ is ***non-orthogonal*** length L pilot sequence unique to n^{th} MTD and known at the BS.
- uplink channel: quasi-static Rayleigh fading $\mathbf{h}_n \sim \mathcal{CN}_M(\mathbf{0}, \mathbf{I})$



Application 1: active user detection for massive random access

- BS with M antennas, N machine type devices (MTDs)
- In a coherence interval, only $K \ll N$ MTD-s are active.
- $\mathbf{a}_n$ is ***non-orthogonal*** length L pilot sequence unique to n^{th} MTD and known at the BS.
- uplink channel: quasi-static Rayleigh fading $\mathbf{h}_n \sim \mathcal{CN}_M(\mathbf{0}, \mathbf{I})$
- $\mathbf{x}_n = \sqrt{\rho_n \beta_n} \alpha_n \mathbf{h}_n :=$ effective channel
  - $\beta_n > 0 :=$ **large-scale fading component (LSFC)**
  - $\rho_n :=$ device's uplink transmission power
  - activity indicator

$$\alpha_n = \begin{cases} 1, & \text{if device } n \text{ is active} \\ 0, & \text{otherwise} \end{cases}.$$



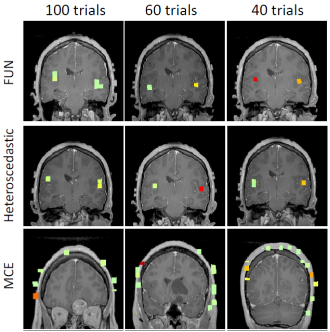
Leads to MMV model [SL18]:

$$\mathbf{Y} = \sum_{n=1}^{N} \sqrt{\rho_n \beta_n} \alpha_n \mathbf{a}_n \mathbf{h}_n^\top + \mathbf{E}$$
$$= \mathbf{A}\mathbf{X} + \mathbf{E} \in \mathbb{C}^{L \times M}$$

$\mathbf{X} = (\mathbf{x}_1 \cdots \mathbf{x}_M)^\top$ is K-rowsparse

Application 2: reconstruction of brain activity using M/EEG based brain source imaging (BSI)

- $\mathbf{Y} \in \mathbb{R}^{M \times L}$:= M/EEG data captures the activity of M sensors at L time instants
- $\mathbf{X} \in \mathbb{R}^{N \times L}$:= unknown activity of N brain source signals at the same time instants
- $\mathbf{A} \in \mathbb{R}^{M \times N}$:= ***lead field matrix*** can be computed using discretization methods (e.g., FEM) for a given head geometry and known electrical conductivities.
- number of sensors M is typically much smaller than the potential brain sources N.
- goal: reconstruct brain activity $\mathbf{X}$ from$\mathbf{Y}$ given $\mathbf{A}$.

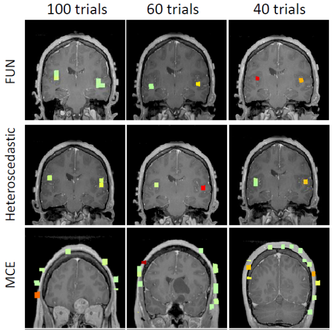


Auditory evoked field (AEF) localization results from one subject [HCG+22]

Application 2: reconstruction of brain activity using M/EEG based brain source imaging (BSI)

- $\mathbf{Y} \in \mathbb{R}^{M \times L}$:= M/EEG data captures the activity of M sensors at L time instants
- $\mathbf{X} \in \mathbb{R}^{N \times L}$:= unknown activity of N brain source signals at the same time instants
- $\mathbf{A} \in \mathbb{R}^{M \times N}$:= ***lead field matrix*** can be computed using discretization methods (e.g., FEM) for a given head geometry and known electrical conductivities.
- number of sensors M is typically much smaller than the potential brain sources N.
- goal: reconstruct brain activity $\mathbf{X}$ from $\mathbf{Y}$ given $\mathbf{A}$.

$\Rightarrow$ MMV model $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{E} \in \mathbb{R}^{M \times L}$, $\mathbf{X}$ is rowsparse



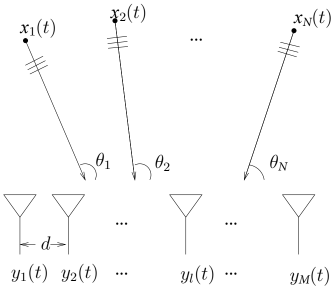
Auditory evoked field (AEF) localization results from one subject [HCG+22]

Application 3: (on-grid) sparse Direction-of-Arrival (DOA) estimation using sensor arrays

- **Goal**: estimate the DOAs $\theta_1, \ldots, \theta_K$ of K narrowband sources using sensor array
- $\{\theta_n\}_{n=1}^{N}$:= **grid** of length N covering the location space $\Theta = [-\pi, \pi)$.
- We acquire L snapshots:

$$\mathbf{y}(t) = \sum_{n=1}^{N} \mathbf{a}_n x_n(t) + \mathbf{e}(t) \in \mathbb{C}^M, \quad t = 1, \ldots, L$$

  - M := # of sensors
  - $\mathbf{a}_n \in \mathbb{C}^M$ is **steering vector** for θ_n ($\|\mathbf{a}_n\|^2 = M$)
  - $\mathbf{e}(t)$ is the noise term.



Uniform linear array (ULA) with d element spacing

Application 3: (on-grid) sparse Direction-of-Arrival (DOA) estimation using sensor arrays

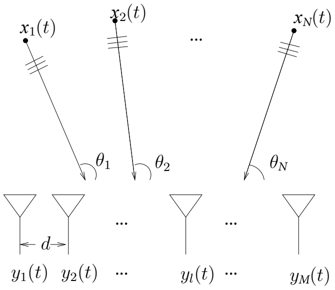
- **Goal**: estimate the DOAs $\theta_1, \ldots, \theta_K$ of K narrowband sources using sensor array
- $\{\theta_n\}_{n=1}^{N} :=$ **grid** of length N covering the location space $\Theta = [-\pi, \pi)$.
- We acquire L snapshots:

$$\mathbf{y}(t) = \sum_{n=1}^{N} \mathbf{a}_n x_n(t) + \mathbf{e}(t) \in \mathbb{C}^M, \quad t = 1, \ldots, L$$

  - $M :=$ # of sensors
  - $\mathbf{a}_n \in \mathbb{C}^M$ is **steering vector** for θ_n ($\|\mathbf{a}_n\|^2 = M$)
  - $\mathbf{e}(t)$ is the noise term.

$\Rightarrow$ MMV model $\mathbf{Y} = \mathbf{AX} + \mathbf{E} \in \mathbb{C}^{M \times L}$

where $\mathbf{X}$ is K-rowsparse for on-grid DOA-s.



Uniform linear array (ULA) with d element spacing

Sparse Bayesian Learning (SBL): key references

Michel Tipping (2001)

Sparse Bayesian learning and the relevance vector machine ME Tipping Journal of machine learning research 1 (Jun), 211-244	9397	2001

David Wipf and Bhaskar Rao (2004) and (2007): analysis and EM algorithm

Sparse Bayesian learning for basis selection DP Wipf, BD Rao IEEE Transactions on Signal processing 52 (8), 2153-2164	1889	2004
An empirical Bayesian strategy for solving the simultaneous sparse approximation problem DP Wipf, BD Rao IEEE Transactions on Signal Processing 55 (7), 3704-3716	1101	2007

Sparse Bayesian Learning (SBL) model

$$\mathbf{y}_i = \mathbf{A}\mathbf{x}_i + \mathbf{e}_i, \quad i = 1, \ldots, L$$

$$\Leftrightarrow \underset{M \times L}{\mathbf{Y}} = \underset{M \times N}{\mathbf{A}} \; \underset{N \times L}{\mathbf{X}} + \underset{M \times L}{\mathbf{E}}$$

Assumptions: σ^2 known, and

$$\mathbf{e}_i \sim \mathcal{CN}_M(\mathbf{0}, \sigma^2\mathbf{I}) \text{ i.i.d.}$$

$$\mathbf{x}_i \sim \mathcal{CN}_N(\mathbf{0}, \mathrm{diag}(\boldsymbol{\gamma})) \text{ i.i.d.}, \; \mathbf{x}_i \perp\!\!\!\perp \mathbf{e}_i$$

uncorrelated sources!

$\Rightarrow$ leads to ***joint Gaussian distribution***:

$$\begin{pmatrix} \mathbf{y}_i \\ \mathbf{x}_i \end{pmatrix} \sim \mathcal{CN}_{M+N}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma} & \mathbf{A}\boldsymbol{\Gamma} \\ \boldsymbol{\Gamma}\mathbf{A}^{\mathsf{H}} & \boldsymbol{\Gamma} \end{pmatrix} \right)$$

$$\boldsymbol{\Gamma} = \mathrm{diag}(\boldsymbol{\gamma}), \qquad \boldsymbol{\Sigma} = \mathbf{A}\,\mathrm{diag}(\boldsymbol{\gamma})\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I}$$

Sparse Bayesian Learning (SBL) model

$$\mathbf{y}_i = \mathbf{A}\mathbf{x}_i + \mathbf{e}_i, \quad i = 1, \dots, L$$

$$\Leftrightarrow \underset{M \times L}{\mathbf{Y}} = \underset{M \times N}{\mathbf{A}} \underset{N \times L}{\mathbf{X}} + \underset{M \times L}{\mathbf{E}}$$

Assumptions: σ^2 known, and

$$\mathbf{e}_i \sim \mathcal{CN}_M(\mathbf{0}, \sigma^2\mathbf{I}) \text{ i.i.d.}$$

$$\mathbf{x}_i \sim \mathcal{CN}_N(\mathbf{0}, \mathrm{diag}(\boldsymbol{\gamma})) \text{ i.i.d.}, \ \mathbf{x}_i \perp\!\!\!\perp \mathbf{e}_i$$

uncorrelated sources!

$\Rightarrow$ leads to ***joint Gaussian distribution***:

$$\begin{pmatrix} \mathbf{y}_i \\ \mathbf{x}_i \end{pmatrix} \sim \mathcal{CN}_{M+N}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma} & \mathbf{A}\boldsymbol{\Gamma} \\ \boldsymbol{\Gamma}\mathbf{A}^{\mathsf{H}} & \boldsymbol{\Gamma} \end{pmatrix} \right)$$

$$\boldsymbol{\Gamma} = \mathrm{diag}(\boldsymbol{\gamma}), \qquad \boldsymbol{\Sigma} = \mathbf{A}\,\mathrm{diag}(\boldsymbol{\gamma})\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I}$$

- The posterior distribution:

$$p(\mathbf{X} \mid \mathbf{Y}, \boldsymbol{\gamma}) = \mathcal{CN}_{NL}(\mathrm{vec}(\mathbf{X}); \mathrm{vec}(\mathbf{M}_x), \mathbf{I}_L \otimes \boldsymbol{\Sigma}_x)$$

$$\mathbf{M}_x = (\boldsymbol{\mu}_1 \ \cdots \ \boldsymbol{\mu}_L) = \boldsymbol{\Gamma}\mathbf{A}^{\mathsf{H}}\boldsymbol{\Sigma}^{-1}\mathbf{Y},$$

$$\boldsymbol{\Sigma}_x = \mathrm{cov}(\mathbf{x}_m \mid \mathbf{y}_m, \boldsymbol{\gamma}) = \boldsymbol{\Gamma} - \boldsymbol{\Gamma}\mathbf{A}^{\mathsf{H}}\boldsymbol{\Sigma}^{-1}\mathbf{A}\boldsymbol{\Gamma}$$

Sparse Bayesian Learning (SBL) model

$$\mathbf{y}_i = \mathbf{A}\mathbf{x}_i + \mathbf{e}_i, \quad i = 1, \ldots, L$$

$$\Leftrightarrow \underset{M \times L}{\mathbf{Y}} = \underset{M \times N}{\mathbf{A}} \underset{N \times L}{\mathbf{X}} + \underset{M \times L}{\mathbf{E}}$$

Assumptions: σ^2 known, and

$$\mathbf{e}_i \sim \mathcal{CN}_M(\mathbf{0}, \sigma^2\mathbf{I}) \text{ i.i.d.}$$

$$\mathbf{x}_i \sim \mathcal{CN}_N(\mathbf{0}, \mathrm{diag}(\boldsymbol{\gamma})) \text{ i.i.d.}, \ \mathbf{x}_i \perp\!\!\!\perp \mathbf{e}_i$$

uncorrelated sources!

$\Rightarrow$ leads to ***joint Gaussian distribution***:

$$\begin{pmatrix} \mathbf{y}_i \\ \mathbf{x}_i \end{pmatrix} \sim \mathcal{CN}_{M+N}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma} & \mathbf{A}\boldsymbol{\Gamma} \\ \boldsymbol{\Gamma}\mathbf{A}^{\mathsf{H}} & \boldsymbol{\Gamma} \end{pmatrix} \right)$$

$$\boldsymbol{\Gamma} = \mathrm{diag}(\boldsymbol{\gamma}), \qquad \boldsymbol{\Sigma} = \mathbf{A}\,\mathrm{diag}(\boldsymbol{\gamma})\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I}$$

- The posterior distribution:

$$p(\mathbf{X} \mid \mathbf{Y}, \boldsymbol{\gamma}) = \mathcal{CN}_{NL}(\mathrm{vec}(\mathbf{X}); \mathrm{vec}(\mathbf{M}_x), \mathbf{I}_L \otimes \boldsymbol{\Sigma}_x)$$

$$\mathbf{M}_x = (\boldsymbol{\mu}_1 \ \cdots \ \boldsymbol{\mu}_L) = \boldsymbol{\Gamma}\mathbf{A}^{\mathsf{H}}\boldsymbol{\Sigma}^{-1}\mathbf{Y},$$

$$\boldsymbol{\Sigma}_x = \mathrm{cov}(\mathbf{x}_m \mid \mathbf{y}_m, \boldsymbol{\gamma}) = \boldsymbol{\Gamma} - \boldsymbol{\Gamma}\mathbf{A}^{\mathsf{H}}\boldsymbol{\Sigma}^{-1}\mathbf{A}\boldsymbol{\Gamma}$$

- Equivalently:

$$p(\mathbf{X} \mid \mathbf{Y}, \boldsymbol{\gamma}) = \prod_{i=1}^{L} \mathcal{CN}_N(\mathbf{x}_i; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_x)$$

$$= |\boldsymbol{\Sigma}_x|^{-L} \prod_{i=1}^{L} \exp(-[\mathbf{x}_i - \boldsymbol{\mu}_i]^{\mathsf{H}}\boldsymbol{\Sigma}_x^{-1}[\mathbf{x}_i - \boldsymbol{\mu}_i])$$

Empirical Bayes approach

Type-II (empirical Bayes): rather than placing a prior on γ, estimate it directly from the data by minimizing the marginal negative log-likelihood (NLLF)

- $\mathbf{y}_i \sim \mathcal{CN}_M(\mathbf{0}, \mathbf{\Sigma})$, $\mathbf{\Sigma} = \mathbf{A}\,\mathrm{diag}(\boldsymbol{\gamma})\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I}$
- The marginal NLLF is

$$\ell(\boldsymbol{\gamma}) = \frac{1}{L}\prod_{i=1}^{L}\mathcal{CN}_M(\mathbf{y}_i; \mathbf{0}, \mathbf{\Sigma})$$

$$= \mathrm{tr}((\mathbf{A\Gamma A}^{\mathsf{H}} + \sigma^2\mathbf{I})^{-1}\hat{\mathbf{\Sigma}}) + \log|\mathbf{A\Gamma A}^{\mathsf{H}} + \sigma^2\mathbf{I}|$$

$$\hat{\mathbf{\Sigma}} = \frac{1}{L}\sum_{i=1}^{L}\mathbf{y}_i\mathbf{y}_i^{\mathsf{H}} := \text{sample cov. matrix}$$

Empirical Bayes approach

Type-II (empirical Bayes): rather than placing a prior on γ, estimate it directly from the data by minimizing the marginal negative log-likelihood (NLLF)

- $\mathbf{y}_i \sim \mathcal{CN}_M(\mathbf{0}, \mathbf{\Sigma})$, $\mathbf{\Sigma} = \mathbf{A}\,\mathrm{diag}(\gamma)\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I}$
- The marginal NLLF is

$$\ell(\gamma) = \frac{1}{L}\prod_{i=1}^{L}\mathcal{CN}_M(\mathbf{y}_i; \mathbf{0}, \mathbf{\Sigma})$$

$$= \mathrm{tr}((\mathbf{A}\mathbf{\Gamma}\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I})^{-1}\hat{\mathbf{\Sigma}}) + \log|\mathbf{A}\mathbf{\Gamma}\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I}|$$

$$\hat{\mathbf{\Sigma}} = \frac{1}{L}\sum_{i=1}^{L}\mathbf{y}_i\mathbf{y}_i^{\mathsf{H}} := \text{sample cov. matrix}$$

1. Compute $\hat{\gamma} = \arg\min_{\gamma \geq 0} \ell(\gamma)$, and $\hat{\mathbf{\Gamma}} = \mathrm{diag}(\hat{\gamma})$, $\hat{\mathbf{\Sigma}} = \mathbf{A}\hat{\mathbf{\Gamma}}\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I}$

Empirical Bayes approach

Type-II (empirical Bayes): rather than placing a prior on γ, estimate it directly from the data by minimizing the marginal negative log-likelihood (NLLF)

- $\mathbf{y}_i \sim \mathcal{CN}_M(\mathbf{0}, \mathbf{\Sigma})$, $\mathbf{\Sigma} = \mathbf{A}\,\mathrm{diag}(\boldsymbol{\gamma})\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I}$
- The marginal NLLF is

$$\ell(\boldsymbol{\gamma}) = \frac{1}{L}\prod_{i=1}^{L}\mathcal{CN}_M(\mathbf{y}_i; \mathbf{0}, \mathbf{\Sigma})$$

$$= \mathrm{tr}((\mathbf{A}\mathbf{\Gamma}\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I})^{-1}\hat{\mathbf{\Sigma}}) + \log|\mathbf{A}\mathbf{\Gamma}\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I}|$$

$$\hat{\mathbf{\Sigma}} = \frac{1}{L}\sum_{i=1}^{L}\mathbf{y}_i\mathbf{y}_i^{\mathsf{H}} := \text{sample cov. matrix}$$

1. Compute $\hat{\boldsymbol{\gamma}} = \arg\min_{\boldsymbol{\gamma}\geq 0}\ell(\boldsymbol{\gamma})$, and $\hat{\mathbf{\Gamma}} = \mathrm{diag}(\hat{\boldsymbol{\gamma}})$, $\hat{\mathbf{\Sigma}} = \mathbf{A}\hat{\mathbf{\Gamma}}\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I}$
2. Estimate of posterior: $p(\mathbf{X} \mid \mathbf{Y}, \hat{\boldsymbol{\gamma}})$ using

$$\hat{\mathbf{X}} = \hat{\mathbf{\Gamma}}\mathbf{A}^{\mathsf{H}}\hat{\mathbf{\Sigma}}^{-1}\mathbf{Y}$$

$$\hat{\mathbf{\Sigma}}_x = \hat{\mathbf{\Gamma}} - \hat{\mathbf{\Gamma}}\,\mathbf{A}^{\mathsf{H}}\hat{\mathbf{\Sigma}}^{-1}\mathbf{A}\hat{\mathbf{\Gamma}},$$

as empirical Bayes-estimates of the posterior mean and cov. matrix.

Empirical Bayes approach

Type-II (empirical Bayes): rather than placing a prior on γ, estimate it directly from the data by minimizing the marginal negative log-likelihood (NLLF)

- $\mathbf{y}_i \sim \mathcal{CN}_M(\mathbf{0}, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma} = \mathbf{A}\,\mathrm{diag}(\boldsymbol{\gamma})\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I}$
- The marginal NLLF is

$$\ell(\boldsymbol{\gamma}) = \frac{1}{L}\prod_{i=1}^{L}\mathcal{CN}_M(\mathbf{y}_i; \mathbf{0}, \boldsymbol{\Sigma})$$
$$= \mathrm{tr}((\mathbf{A}\boldsymbol{\Gamma}\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I})^{-1}\hat{\boldsymbol{\Sigma}}) + \log|\mathbf{A}\boldsymbol{\Gamma}\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I}|$$

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{L}\sum_{i=1}^{L}\mathbf{y}_i\mathbf{y}_i^{\mathsf{H}} := \text{sample cov. matrix}$$

1. Compute $\hat{\boldsymbol{\gamma}} = \arg\min_{\boldsymbol{\gamma}\geq 0}\ell(\boldsymbol{\gamma})$, and $\hat{\boldsymbol{\Gamma}} = \mathrm{diag}(\hat{\boldsymbol{\gamma}})$, $\hat{\boldsymbol{\Sigma}} = \mathbf{A}\hat{\boldsymbol{\Gamma}}\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I}$
2. Estimate of posterior: $p(\mathbf{X} \mid \mathbf{Y}, \hat{\boldsymbol{\gamma}})$ using

$$\hat{\mathbf{X}} = \hat{\boldsymbol{\Gamma}}\mathbf{A}^{\mathsf{H}}\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{Y} \leftarrow \text{is row-sparse!}$$
$$\hat{\boldsymbol{\Sigma}}_x = \hat{\boldsymbol{\Gamma}} - \hat{\boldsymbol{\Gamma}}\mathbf{A}^{\mathsf{H}}\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{A}\hat{\boldsymbol{\Gamma}},$$

as empirical Bayes-estimates of the posterior mean and cov. matrix.

Last step: enhancing the row-sparsity of the solution

- Key assumption of MMV model
  $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{E}$ is rowsparsity of $\mathbf{X} \in \mathbb{C}^{M \times L}$:

$$\mathcal{M} = \mathrm{rowsupp}(\mathbf{X}) \subset \{1, \ldots, N\},$$
$$K = |\mathcal{M}| \ll N$$

- Since $\mathbf{\Gamma} = \mathrm{diag}(\boldsymbol{\gamma}) = \mathsf{cov}(\mathbf{x}_i)$, one has that

$$\mathcal{M} = \mathrm{rowsupp}(\mathbf{X}) = \mathrm{supp}(\boldsymbol{\gamma})$$

- After solving

$$\hat{\boldsymbol{\gamma}} = \arg\min_{\boldsymbol{\gamma} \geq 0} \ell(\boldsymbol{\gamma})$$

compute $\hat{\mathbf{\Sigma}} = \mathbf{A}\,\mathrm{diag}(\hat{\boldsymbol{\gamma}})\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I}$ and $\hat{\mathbf{X}}$.

Last step: enhancing the row-sparsity of the solution

- Key assumption of MMV model
  $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{E}$ is rowsparsity of $\mathbf{X} \in \mathbb{C}^{M \times L}$:

$$\mathcal{M} = \mathrm{rowsupp}(\mathbf{X}) \subset \{1, \ldots, N\},$$
$$K = |\mathcal{M}| \ll N$$

- Since $\mathbf{\Gamma} = \mathrm{diag}(\boldsymbol{\gamma}) = \mathsf{cov}(\mathbf{x}_i)$, one has that

$$\mathcal{M} = \mathrm{rowsupp}(\mathbf{X}) = \mathrm{supp}(\boldsymbol{\gamma})$$

- After solving

$$\hat{\boldsymbol{\gamma}} = \arg\min_{\boldsymbol{\gamma} \geq 0} \ell(\boldsymbol{\gamma})$$

compute $\hat{\mathbf{\Sigma}} = \mathbf{A}\,\mathrm{diag}(\hat{\boldsymbol{\gamma}})\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I}$ and $\hat{\mathbf{X}}$.

The activity index $(\alpha_n,\ n = 1, \ldots, n)$ computed using either of the rules to enhance sparsity:

- ***Threshold-rule***:

$$\hat{\alpha}_n = 1\{\hat{\gamma}_n \geq \gamma^{\mathsf{th}}\} = \begin{cases} 1, & \text{if } \hat{\gamma}_n \geq \gamma^{\mathsf{th}} \\ 0 & \text{otherwise} \end{cases}$$

  where γ^{th} is a fixed threshold.

- ***top K-rule***:

$$\hat{\alpha}_n = 1\{\gamma_n \in \text{ the } K \text{ largest entries of } \hat{\boldsymbol{\gamma}}\}$$

  where $1\{\cdot\}$ is the indicator function (and K assumed given/estimated)

M-SBL [WR07]

- **M-SBL** by Wipf and Rao (2007) solves $\ell(\boldsymbol{\gamma})$ using the ***EM algorithm*** treating $\{(\mathbf{y}_i, \mathbf{x}_i)\}_{i=1}^{L}$ as complete data:
    - **slow convergence**, impractical for N large
    - **not sparse**: $\hat{\gamma}_i \neq 0$ for any i

M-SBL [WR07]

- **M-SBL** by Wipf and Rao (2007) solves $\ell(\boldsymbol{\gamma})$ using the ***EM algorithm*** treating $\{(\mathbf{y}_i, \mathbf{x}_i)\}_{i=1}^{L}$ as complete data:
  - **slow convergence**, impractical for N large
  - **not sparse**: $\hat{\gamma}_i \neq 0$ for any i

$\Rightarrow$ *non-sparse Bayesian learning??*

M-SBL [WR07]

- **M-SBL** by Wipf and Rao (2007) solves $\ell(\gamma)$ using the ***EM algorithm*** treating $\{(\mathbf{y}_i, \mathbf{x}_i)\}_{i=1}^L$ as complete data:
  - **slow convergence**, impractical for N large
  - **not sparse**: $\hat{\gamma}_i \neq 0$ for any i

  $\Rightarrow$ *non-sparse Bayesian learning??*
- We need to solve the non-convex objective:

$$\min_{\gamma \geq 0} \ell(\gamma) = g(\gamma) + f(\gamma),$$

$$g(\gamma) = \operatorname{tr}((\mathbf{A} \operatorname{diag}(\gamma)\mathbf{A}^{\mathsf{H}} + \sigma^2 \mathbf{I})^{-1} \hat{\mathbf{\Sigma}})$$

$$f(\gamma) = \log |\mathbf{A} \operatorname{diag}(\gamma)\mathbf{A}^{\mathsf{H}} + \sigma^2 \mathbf{I}|.$$

Some alternatives:

- **Coordinatewise-optimization (CWO)** [FHJC21]: sequential coordinate updates
- **Sequential lightweight SBL (SLW-SBL)** [PR25], a greedy approach.

M-SBL [WR07]

- **M-SBL** by Wipf and Rao (2007) solves $\ell(\boldsymbol{\gamma})$ using the ***EM algorithm*** treating $\{(\mathbf{y}_i, \mathbf{x}_i)\}_{i=1}^L$ as complete data:
  - **slow convergence**, impractical for N large
  - **not sparse**: $\hat{\gamma}_i \neq 0$ for any i
  - $\Rightarrow$ *non-sparse Bayesian learning??*
- We need to solve the non-convex objective:

$$\min_{\boldsymbol{\gamma} \geq 0} \ell(\boldsymbol{\gamma}) = g(\boldsymbol{\gamma}) + f(\boldsymbol{\gamma}),$$

$$g(\boldsymbol{\gamma}) = \operatorname{tr}((\mathbf{A} \operatorname{diag}(\boldsymbol{\gamma}) \mathbf{A}^{\mathsf{H}} + \sigma^2 \mathbf{I})^{-1} \hat{\boldsymbol{\Sigma}})$$

$$f(\boldsymbol{\gamma}) = \log |\mathbf{A} \operatorname{diag}(\boldsymbol{\gamma}) \mathbf{A}^{\mathsf{H}} + \sigma^2 \mathbf{I}|.$$

Some alternatives:

- **Coordinatewise-optimization (CWO)** [FHJC21]: sequential coordinate updates
- **Sequential lightweight SBL (SLW-SBL)** [PR25], a greedy approach.

Our proposed approaches:

1. **Greedy methods:** Covariance learning matching pursuit (CL-MP and CL-OMP)
2. **Optimization-based:** Covariance learning via MM SCA approaches
3. **Robust** CWO and CL-MP methods, resistant to outliers.

OMP formulation in [Ela10, Table 3.1]:

$$\mathbf{b} = \mathbf{A}\mathbf{x} + \mathbf{e} = \sum_j \mathbf{a}_j x_j + \mathbf{e}, \quad \mathbf{x} \text{ is sparse}$$

Initialization: Initialize $k = 0$, and set

- The initial solution $\mathbf{x}^0 = \mathbf{0}$.
- The initial residual $\mathbf{r}^0 = \mathbf{b} - \mathbf{A}\mathbf{x}^0 = \mathbf{b}$.
- The initial solution support $\mathcal{S}^0 = Support\{\mathbf{x}^0\} = \emptyset$.

Main Iteration: Increment k by 1 and perform the following steps:

- **Sweep:** Compute the errors $\epsilon(j) = \min_{z_j} \|\mathbf{a}_j z_j - \mathbf{r}^{k-1}\|_2^2$ for all j using the optimal choice $z_j^* = \mathbf{a}_j^T \mathbf{r}^{k-1} / \|\mathbf{a}_j\|_2^2$.
- **Update Support:** Find a minimizer, j_0 of $\epsilon(j)$: $\forall\, j \notin \mathcal{S}^{k-1}$, $\epsilon(j_0) \leq \epsilon(j)$, and update $\mathcal{S}^k = \mathcal{S}^{k-1} \cup \{j_0\}$.
- **Update Provisional Solution:** Compute $\mathbf{x}^k$, the minimizer of $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2$ subject to $Support\{\mathbf{x}\} = \mathcal{S}^k$.
- **Update Residual:** Compute $\mathbf{r}^k = \mathbf{b} - \mathbf{A}\mathbf{x}^k$.
- **Stopping Rule:** If $\|\mathbf{r}^k\|_2 < \epsilon_0$, stop. Otherwise, apply another iteration.

Output: The proposed solution is $\mathbf{x}^k$ obtained after k iterations.

OMP formulation in [Ela10, Table 3.1]:

$$\mathbf{b} = \mathbf{A}\mathbf{x} + \mathbf{e} = \sum_j \mathbf{a}_j x_j + \mathbf{e}, \quad \mathbf{x} \text{ is sparse}$$

Initialization: Initialize $k = 0$, and set

- The initial solution $\mathbf{x}^0 = \mathbf{0}$.
- The initial residual $\mathbf{r}^0 = \mathbf{b} - \mathbf{A}\mathbf{x}^0 = \mathbf{b}$.
- The initial solution support $\mathcal{S}^0 = Support\{\mathbf{x}^0\} = \emptyset$.

Main Iteration: Increment k by 1 and perform the following steps:

- **Sweep:** Compute the errors $\epsilon(j) = \min_{z_j} \|\mathbf{a}_j z_j - \mathbf{r}^{k-1}\|_2^2$ for all j using the optimal choice $z_j^* = \mathbf{a}_j^T \mathbf{r}^{k-1} / \|\mathbf{a}_j\|_2^2$.
- **Update Support:** Find a minimizer, j_0 of $\epsilon(j)$: $\forall\, j \notin \mathcal{S}^{k-1}$, $\epsilon(j_0) \leq \epsilon(j)$, and update $\mathcal{S}^k = \mathcal{S}^{k-1} \cup \{j_0\}$.
- **Update Provisional Solution:** Compute $\mathbf{x}^k$, the minimizer of $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2$ subject to $Support\{\mathbf{x}\} = \mathcal{S}^k$.
- **Update Residual:** Compute $\mathbf{r}^k = \mathbf{b} - \mathbf{A}\mathbf{x}^k$.
- **Stopping Rule:** If $\|\mathbf{r}^k\|_2 < \epsilon_0$, stop. Otherwise, apply another iteration.

Output: The proposed solution is $\mathbf{x}^k$ obtained after k iterations.

Sweep: at iteration k, solve univariate Gaussian MLE for i^{th} variable

$$\min_{x_i} \ell(x_i, \mathbf{x}_{\setminus i}^{k-1})$$

keeping other parameters fixed at previous values.

OMP formulation in [Ela10, Table 3.1]:

$$\mathbf{b} = \mathbf{A}\mathbf{x} + \mathbf{e} = \sum_j \mathbf{a}_j x_j + \mathbf{e}, \quad \mathbf{x} \text{ is sparse}$$

Initialization: Initialize $k = 0$, and set

- The initial solution $\mathbf{x}^0 = \mathbf{0}$.
- The initial residual $\mathbf{r}^0 = \mathbf{b} - \mathbf{A}\mathbf{x}^0 = \mathbf{b}$.
- The initial solution support $\mathcal{S}^0 = Support\{\mathbf{x}^0\} = \emptyset$.

Main Iteration: Increment k by 1 and perform the following steps:

- **Sweep:** Compute the errors $\epsilon(j) = \min_{z_j} \|\mathbf{a}_j z_j - \mathbf{r}^{k-1}\|_2^2$ for all j using the optimal choice $z_j^* = \mathbf{a}_j^T \mathbf{r}^{k-1}/\|\mathbf{a}_j\|_2^2$.
- **Update Support:** Find a minimizer, j_0 of $\epsilon(j)$: $\forall\, j \notin \mathcal{S}^{k-1}$, $\epsilon(j_0) \le \epsilon(j)$, and update $\mathcal{S}^k = \mathcal{S}^{k-1} \cup \{j_0\}$.
- **Update Provisional Solution:** Compute $\mathbf{x}^k$, the minimizer of $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2$ subject to $Support\{\mathbf{x}\} = \mathcal{S}^k$.
- **Update Residual:** Compute $\mathbf{r}^k = \mathbf{b} - \mathbf{A}\mathbf{x}^k$.
- **Stopping Rule:** If $\|\mathbf{r}^k\|_2 < \epsilon_0$, stop. Otherwise, apply another iteration.

Output: The proposed solution is $\mathbf{x}^k$ obtained after k iterations.

Sweep: at iteration k, solve univariate Gaussian MLE for i^{th} variable

$$\min_{x_i} \ell(x_i, \mathbf{x}_{\setminus i}^{k-1})$$

keeping other parameters fixed at previous values.

Update provisional solution: solve the full MLE for new model with newly added variable:

$$\min_{\mathbf{x}} \ell(\mathbf{x}) \quad \text{subject to } \operatorname{supp}(\mathbf{x}) = \mathcal{S}^k$$

OMP formulation in [Ela10, Table 3.1]:

$$\mathbf{b} = \mathbf{A}\mathbf{x} + \mathbf{e} = \sum_j \mathbf{a}_j x_j + \mathbf{e}, \quad \mathbf{x} \text{ is sparse}$$

Initialization: Initialize $k = 0$, and set

- The initial solution $\mathbf{x}^0 = \mathbf{0}$.
- The initial residual $\mathbf{r}^0 = \mathbf{b} - \mathbf{A}\mathbf{x}^0 = \mathbf{b}$.
- The initial solution support $\mathcal{S}^0 = Support\{\mathbf{x}^0\} = \emptyset$.

Main Iteration: Increment k by 1 and perform the following steps:

- **Sweep:** Compute the errors $\epsilon(j) = \min_{z_j} \|\mathbf{a}_j z_j - \mathbf{r}^{k-1}\|_2^2$ for all j using the optimal choice $z_j^* = \mathbf{a}_j^T \mathbf{r}^{k-1} / \|\mathbf{a}_j\|_2^2$.
- **Update Support:** Find a minimizer, j_0 of $\epsilon(j)$: $\forall\, j \notin \mathcal{S}^{k-1}$, $\epsilon(j_0) \le \epsilon(j)$, and update $\mathcal{S}^k = \mathcal{S}^{k-1} \cup \{j_0\}$.
- **Update Provisional Solution:** Compute $\mathbf{x}^k$, the minimizer of $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2$ subject to $Support\{\mathbf{x}\} = \mathcal{S}^k$.
- **Update Residual:** Compute $\mathbf{r}^k = \mathbf{b} - \mathbf{A}\mathbf{x}^k$.
- **Stopping Rule:** If $\|\mathbf{r}^k\|_2 < \epsilon_0$, stop. Otherwise, apply another iteration.

Output: The proposed solution is $\mathbf{x}^k$ obtained after k iterations.

Sweep: at iteration k, solve univariate Gaussian MLE for i^{th} variable

$$\min_{x_i} \ell(x_i, \mathbf{x}_{\setminus i}^{k-1})$$

keeping other parameters fixed at previous values.

Update provisional solution: solve the full MLE for new model with newly added variable:

$$\min_{\mathbf{x}} \ell(\mathbf{x}) \quad \text{subject to } \operatorname{supp}(\mathbf{x}) = \mathcal{S}^k$$

$\Rightarrow$ these steps are available for SBL!

Maximum likelihood estimation results [1/2]: sweep

- $\mathbf{y}_i \sim \mathcal{CN}_M(\mathbf{0}, \mathbf{\Sigma})$ with $\mathbf{\Sigma} = \mathbf{A}\,\mathrm{diag}(\boldsymbol{\gamma})\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I}$, $\mathbf{A} \in \mathbb{C}^{M\times N}$, $M \ll N$.
- Both $\boldsymbol{\gamma} \geq 0$ and $\sigma^2 > 0$ are unknown.
- The neg. log-likelihood:

$$\ell(\boldsymbol{\gamma}, \sigma^2) = \mathrm{tr}((\mathbf{A}\,\mathrm{diag}(\boldsymbol{\gamma})\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I})^{-1}\hat{\mathbf{\Sigma}}) + \log|\mathbf{A}\,\mathrm{diag}(\boldsymbol{\gamma})\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I}|$$

where $\hat{\mathbf{\Sigma}} = \frac{1}{L}\sum_{l=1}^{L}\mathbf{y}_l\mathbf{y}_l^{\mathsf{H}}$ is the sample covariance matrix (SCM).

Maximum likelihood estimation results [1/2]: sweep

- $\mathbf{y}_i \sim \mathcal{CN}_M(\mathbf{0}, \mathbf{\Sigma})$ with $\mathbf{\Sigma} = \mathbf{A}\,\mathrm{diag}(\boldsymbol{\gamma})\mathbf{A}^\mathsf{H} + \sigma^2\mathbf{I}$, $\mathbf{A} \in \mathbb{C}^{M\times N}$, $M \ll N$.
- Both $\boldsymbol{\gamma} \geq 0$ and $\sigma^2 > 0$ are unknown.
- The neg. log-likelihood:

$$\ell(\boldsymbol{\gamma}, \sigma^2) = \mathrm{tr}((\mathbf{A}\,\mathrm{diag}(\boldsymbol{\gamma})\mathbf{A}^\mathsf{H} + \sigma^2\mathbf{I})^{-1}\hat{\mathbf{\Sigma}}) + \log|\mathbf{A}\,\mathrm{diag}(\boldsymbol{\gamma})\mathbf{A}^\mathsf{H} + \sigma^2\mathbf{I}|$$

where $\hat{\mathbf{\Sigma}} = \frac{1}{L}\sum_{l=1}^{L}\mathbf{y}_l\mathbf{y}_l^\mathsf{H}$ is the sample covariance matrix (SCM).

The **conditional** neg. LLF for $\gamma_i \geq 0$ (with fixed $\{\gamma_j\}_{j\neq i}$, σ^2) is [FT01, YLS+10]:

$$\begin{aligned}\hat{\gamma}_i &= \arg\min_{\gamma\geq 0}\ \mathrm{tr}((\mathbf{\Sigma}_{\setminus i} + \gamma\mathbf{a}_i\mathbf{a}_i^\mathsf{H})^{-1}\hat{\mathbf{\Sigma}}) + \log|\mathbf{\Sigma}_{\setminus i} + \gamma\mathbf{a}_i\mathbf{a}_i^\mathsf{H}| \\ &= \max\left(\frac{\mathbf{b}_i^\mathsf{H}\hat{\mathbf{\Sigma}}\mathbf{b}_i}{(\mathbf{a}_i^\mathsf{H}\mathbf{b}_i)^2} - \frac{1}{\mathbf{a}_i^\mathsf{H}\mathbf{b}_i}, 0\right),\end{aligned}$$

where $\mathbf{b}_i = \mathbf{\Sigma}_{\setminus i}^{-1}\mathbf{a}_i$ and $\mathbf{\Sigma}_{\setminus i} = \mathbf{\Sigma} - \gamma_i\mathbf{a}_i\mathbf{a}_i^\mathsf{H}$ is the array CM without the i^{th} source.

Maximum likelihood estimation results [2/2]: update provisional solution

Suppose current model $\mathcal{M}^k = (i_1, \ldots, i_k)$ is true: $\mathbf{y}_i \sim \mathcal{CN}_M(\mathbf{0}, \mathbf{A}_\mathcal{M} \operatorname{diag}(\boldsymbol{\gamma}) \mathbf{A}_\mathcal{M}^\mathsf{H} + \sigma^2 \mathbf{I})$ with $\boldsymbol{\gamma} \in \mathbb{R}^k$ and $\mathbf{A}_\mathcal{M}$ is the $M \times k$ matrix of active columns extracted from $M \times N$ matrix $\mathbf{A}$.

Then the unrestricted minimizers of the neg. LLF $\ell(\boldsymbol{\gamma}, \sigma^2)$ are [SN95]:

$$\hat{\sigma}^2 = \frac{1}{N-k} \operatorname{tr}\big((\mathbf{I} - \mathbf{A}_\mathcal{M} \mathbf{A}_\mathcal{M}^+) \hat{\boldsymbol{\Sigma}}\big)$$

$$\hat{\boldsymbol{\gamma}} = \operatorname{diag}(\mathbf{A}_\mathcal{M}^+ (\hat{\boldsymbol{\Sigma}} - \hat{\sigma}^2 \mathbf{I}) \mathbf{A}_\mathcal{M}^{+\mathsf{H}})$$

provided that $\hat{\gamma}_i > 0$ for all $i = 1, \ldots, k$.

Remark. If $\hat{\boldsymbol{\gamma}} \in \mathbb{R}^k$ contains negative elements, then calculate the constrained solution using Bressler's algorithm [Bre88, Algorithm I].

Algorithm 1: CL-OMP: **C**ovariance **L**earning **O**rthogonal **M**atching **P**ursuit algorithm

Input : $\hat{\boldsymbol{\Sigma}} \in \mathbb{C}^{L \times L}$, $\mathbf{A} \in \mathbb{C}^{M \times N}$, K

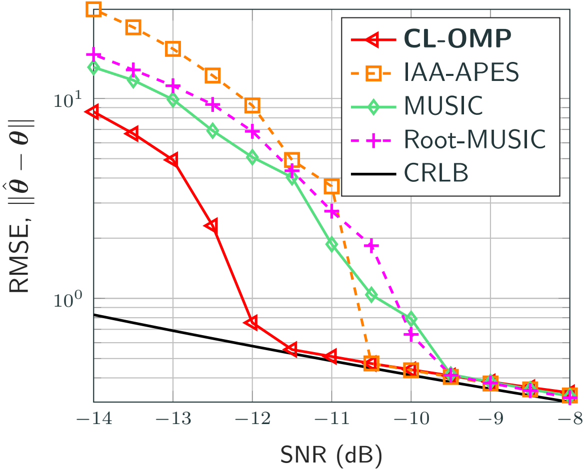
Initialize: $\boldsymbol{\Sigma} = [\mathrm{tr}(\hat{\boldsymbol{\Sigma}})/p]\mathbf{I}$, $\mathcal{M} = \emptyset$

for $k = 1, \ldots, K$ **do**

- $\mathbf{B} \leftarrow \boldsymbol{\Sigma}^{-1}\mathbf{A} = (\mathbf{b}_1 \cdots \mathbf{b}_N)$
- $\gamma_i \leftarrow \max\left(\frac{\mathbf{b}_i^{\mathsf{H}}\hat{\boldsymbol{\Sigma}}\mathbf{b}_i}{(\mathbf{a}_i^{\mathsf{H}}\mathbf{b}_i)^2} - \frac{1}{\mathbf{a}_i^{\mathsf{H}}\mathbf{b}_i}, 0\right)$, $\forall i \in \mathcal{M}^{\complement}$ `// Sweep`
- $\boldsymbol{\epsilon} = (\epsilon_i) \leftarrow \left(\log(1 + \gamma_i \mathbf{a}_i^{\mathsf{H}}\mathbf{b}_i) - \gamma_i \mathbf{a}_i^{\mathsf{H}}\mathbf{b}_i\right)_{N \times 1}$ `// value of neg. LLF at solution`
- $\mathcal{M} \leftarrow \mathcal{M} \cup \{i_k\}$ with $i_k \leftarrow \arg\min_{i \notin \mathcal{M}} \epsilon_i$ `// choose source with smallest error`
- $\sigma^2 \leftarrow \frac{1}{M-k}\mathrm{tr}\big((\mathbf{I} - \mathbf{A}_{\mathcal{M}}\mathbf{A}_{\mathcal{M}}^{+})\hat{\boldsymbol{\Sigma}}\big)$ `// Update provisional solution of noise power`
- $\boldsymbol{\gamma}_{\mathcal{M}} \leftarrow \max\left(\mathrm{diag}\left(\mathbf{A}_{\mathcal{M}}^{+}(\hat{\boldsymbol{\Sigma}} - \sigma^2\mathbf{I})\mathbf{A}_{\mathcal{M}}^{+\mathsf{H}}\right), 0\right)$ `// Update provisional solution`
- If $\boldsymbol{\gamma}_{\mathcal{M}}$ has negative elements, compute the constrained MLE-s using Bressler's algorithm
- $\boldsymbol{\gamma}_{\mathcal{M}^{\complement}} \leftarrow \mathbf{0}$
- $\boldsymbol{\Sigma} \leftarrow \mathbf{A}\,\mathrm{diag}(\boldsymbol{\gamma})\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I}$

Output : $\mathcal{M}$, $\boldsymbol{\gamma}$, σ^2

Application to sparse DOA estimation



- $K = 4$ Gaussian sources;
- $L = 125$ snapshots
- $M = 20$ sensors
- $N = 1801$ grid size ($\Delta\theta = 0.1^o$ resolution)
- MC trials $= 2000$
- Three of the sources are off-grid:

$$\boldsymbol{\theta} = \begin{pmatrix} -30.1^o \\ -20.02^o \\ -10.02^o \\ 3.02^o \end{pmatrix}$$

Case: σ^2 is known

Algorithm 2: Covariance Learning Matching Pursuit (CL-MP) algorithm

Input : SCM $\hat{\mathbf{\Sigma}} \in \mathbb{C}^{L\times L}$, $\mathbf{A} \in \mathbb{C}^{M\times N}$, $\sigma^2 > 0$

Initialize: $\mathbf{B} = (1/\sigma^2)\mathbf{A}$, $\mathcal{M} = \emptyset$

for $k = 1, \dots, N$ **do**

$\quad \gamma_i \leftarrow \max\left(\frac{\mathbf{b}_i^{\mathsf{H}}\hat{\mathbf{\Sigma}}\mathbf{b}_i - \mathbf{a}_i^{\mathsf{H}}\mathbf{b}_i}{(\mathbf{a}_i^{\mathsf{H}}\mathbf{b}_i)^2}, 0\right), \ \forall i \in \mathcal{M}^{\complement}$

$\quad \epsilon_i \leftarrow \log(1 + \gamma_i \mathbf{a}_i^{\mathsf{H}}\mathbf{b}_i) - \gamma_i \mathbf{a}_i^{\mathsf{H}}\mathbf{b}_i, \ \forall i \in \mathcal{M}^{\complement}$

$\quad \mathcal{M} \leftarrow \mathcal{M} \cup \{i_k\}$ with $i_k \leftarrow \arg\min_{i\in\mathcal{M}^{\complement}} \epsilon_i$

$\quad$ **if** *stopping rule is met (see text)* **then**

$\quad\quad$ break

$\quad \mathbf{B} \leftarrow \mathbf{B} - \frac{\gamma_{i_k}}{1 + \gamma_{i_k}\mathbf{a}_{i_k}^{\mathsf{H}}\mathbf{b}_{i_k}} \mathbf{b}_{i_k}\mathbf{a}_{i_k}^{\mathsf{H}}\mathbf{B}$

Output : support $\mathcal{M}$

When σ^2 is known, ***Matching Pursuit*** provides better alternative to OMP:

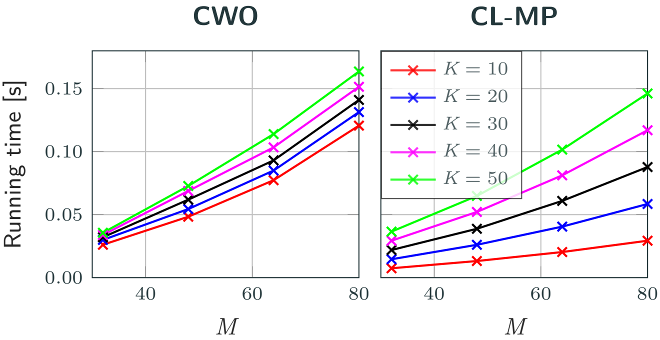
- $\Rightarrow$ provisional solution update can be ignored
- $\Rightarrow$ Further speed-up via Sherman-Morrison formula

$$(\mathbf{\Sigma} + \gamma \mathbf{a}_i \mathbf{a}_i^{\mathsf{H}})^{-1} = \mathbf{\Sigma}^{-1} - \frac{\gamma \mathbf{\Sigma}^{-1}\mathbf{a}_i\mathbf{a}_i^{\mathsf{H}}\mathbf{\Sigma}^{-1}}{1 + \gamma \mathbf{a}_i^{\mathsf{H}}\mathbf{\Sigma}^{-1}\mathbf{a}_i}$$

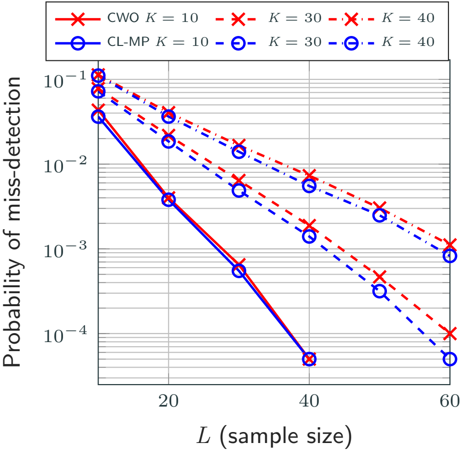
yielding update for $\mathbf{B} = [\mathbf{\Sigma}]^{-1}\mathbf{A}$ as

$$\mathbf{B} \leftarrow \mathbf{B} - \frac{\gamma_i}{1 + \gamma_i \mathbf{a}_i^{\mathsf{H}}\mathbf{b}_i}\mathbf{b}_i\mathbf{a}_i^{\mathsf{H}}\mathbf{B},$$

Coordinatewise optimization (CWO) algorithm versus CL-MP



- Bernoulli measurement matrix $\mathbf{A} \in \mathbb{C}^{M \times N}$
- $\mathbf{Y} \in \mathbb{C}^{M \times L}$, with sample size $L = 40$
- $N = 1000$, $\gamma_n \sim Unif(-15, 0)$ on dB scale.



$M = 64, \ N = 1000$

References

CL-OMP:

Matching pursuit covariance learning
Esa Ollila
Proc. *EUSIPCO-2024*, pp. 2447-2451, August 26-30 2024, Lyon, France.

CL-MP:

Activity detection for massive random access using covariance-based matching pursuit
Leatile Marata, Esa Ollila, and Hirley Alves
IEEE Transactions on Vehicular Technology, vol. 74, no 11, pp. 17292 - 17303, 2025

SBL using SCA and MM

In SBL, we need to solve the non-convex type-II likelihood:

$$\min_{\gamma \geq 0} \ell(\gamma) = g(\gamma) + f(\gamma),$$

$$g(\gamma) = \mathrm{tr}((\mathbf{A}\,\mathrm{diag}(\gamma)\mathbf{A}^{\mathsf{H}} + \sigma^2\,\mathbf{I})^{-1}\,\hat{\boldsymbol{\Sigma}})$$

$$f(\gamma) = \log|\mathbf{A}\,\mathrm{diag}(\gamma)\mathbf{A}^{\mathsf{H}} + \sigma^2\,\mathbf{I}|.$$

where g is convex, while f is non-convex

We propose two fast **covariance learning (CL-)**algorithms, both quaranteed to converge to a stationary point.

CL-MM:

- uses **majorization-minimization**, majorize both convex and concave terms, then minimize the suggogate analytically
- closed-form update per iteration, avoids the slow convergence of EM-based SBL

CL-SCA:

- linearize non-convex part $\rightarrow$ minimize convex approximation
- All N variables updated simultaneously (parallel).

MM algorithm

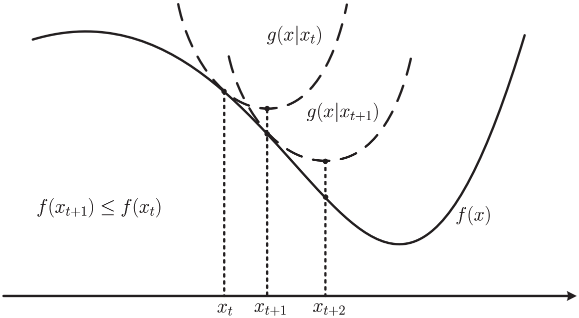


Figure from [SBP16]

(1) $g(x \mid x_t) \geq f(x_t), \ \forall x$

(2) $g(x_t \mid x_t) = f(x_t)$

- Each MM-iterate guarantees $f(x_{t+1}) \leq f(x_t)$
- Convergence to the stationary solution
- With right majorizer, each surrogate has a closed-form solution, fast per-iteration updates

MM algorithm

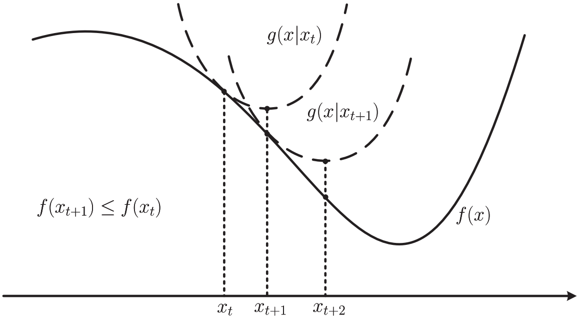


Figure from [SBP16]

(1) $g(x \mid x_t) \geq f(x_t), \ \forall x$

(2) $g(x_t \mid x_t) = f(x_t)$

- Each MM-iterate guarantees $f(x_{t+1}) \leq f(x_t)$
- Convergence to the stationary solution
- With right majorizer, each surrogate has a closed-form solution, fast per-iteration updates

EM is a special case of MM: $\ell(\gamma) \leq Q(\gamma \mid \gamma') + c'$, with c' constant not dependent on γ.

$$Q(\gamma \mid \gamma') = \mathbb{E}_{\{\mathbf{x}_i\}}\big[-\ln p(\{\mathbf{y}_i, \mathbf{x}_i\} \mid \gamma) \ \big| \ \{\mathbf{y}_i\}, \gamma' \big]$$

MM algorithm

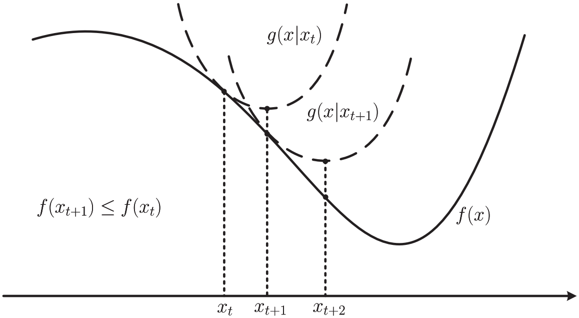


Figure from [SBP16]

(1) $g(x \mid x_t) \geq f(x_t), \ \forall x$

(2) $g(x_t \mid x_t) = f(x_t)$

- Each MM-iterate guarantees $f(x_{t+1}) \leq f(x_t)$
- Convergence to the stationary solution
- With right majorizer, each surrogate has a closed-form solution, fast per-iteration updates

EM is a special case of MM: $\ell(\gamma) \leq Q(\gamma \mid \gamma') + c'$, with c' constant not dependent on γ.

$$Q(\gamma \mid \gamma') = \mathbb{E}_{\{\mathbf{x}_i\}}\left[- \ln p(\{\mathbf{y}_i, \mathbf{x}_i\} \mid \gamma) \mid \{\mathbf{y}_i\}, \gamma' \right]$$

For SBL problem above, $Q(\gamma \mid \gamma')$ does not provide a tight surrogate function

Majorize the convex Trace part

Lemma 1: for $\mathbf{D} \succ \mathbf{0}$ and $\mathbf{A}\mathbf{D}\mathbf{A}^{\mathsf{H}} \succ \mathbf{0}$,

$$\left(\mathbf{A}\mathbf{D}\mathbf{A}^{\mathsf{H}}\right)^{-1} \preceq \left(\mathbf{F}^{k}\right)^{-1}\mathbf{A}\mathbf{D}^{k}\mathbf{D}^{-1}\mathbf{D}^{k}\mathbf{A}^{\mathsf{H}}\left(\mathbf{F}^{k}\right)^{-1}, \quad \mathbf{F}^{(k)} = \mathbf{A}\mathbf{D}^{k}\mathbf{A}^{\mathsf{H}}$$

with equality at $\mathbf{D} = \mathbf{D}^{k}$.

Can we apply Lemma to the matrix $\mathbf{\Sigma}^{-1} = \left(\mathbf{A}\,\mathrm{diag}(\boldsymbol{\gamma})\mathbf{A}^{\mathsf{H}} + \sigma^2\,\mathbf{I}\right)^{-1}$?

Challenge: $\boldsymbol{\gamma} \geq \mathbf{0}$ (not $\boldsymbol{\gamma} > \mathbf{0}$). Lemma 1 requires $\mathbf{D} \succ \mathbf{0}$. Loose bound for $\gamma_i = 0$.

Majorize the convex Trace part

Lemma 1: for $\mathbf{D} \succ \mathbf{0}$ and $\mathbf{A}\mathbf{D}\mathbf{A}^{\mathsf{H}} \succ \mathbf{0}$,

$$\left(\mathbf{A}\mathbf{D}\mathbf{A}^{\mathsf{H}}\right)^{-1} \preceq \left(\mathbf{F}^{k}\right)^{-1}\mathbf{A}\mathbf{D}^{k}\mathbf{D}^{-1}\mathbf{D}^{k}\mathbf{A}^{\mathsf{H}}\left(\mathbf{F}^{k}\right)^{-1}, \quad \mathbf{F}^{(k)} = \mathbf{A}\mathbf{D}^{k}\mathbf{A}^{\mathsf{H}}$$

with equality at $\mathbf{D} = \mathbf{D}^{k}$.

Can we apply Lemma to the matrix $\mathbf{\Sigma}^{-1} = \left(\mathbf{A}\,\mathrm{diag}(\boldsymbol{\gamma})\mathbf{A}^{\mathsf{H}} + \sigma^2\,\mathbf{I}\right)^{-1}$?

Challenge: $\boldsymbol{\gamma} \geq \mathbf{0}$ (not $\boldsymbol{\gamma} > \mathbf{0}$). Lemma 1 requires $\mathbf{D} \succ \mathbf{0}$. Loose bound for $\gamma_i = 0$.

Fix: Since $\sigma^2 > 0$, absorb σ^2/η into $\boldsymbol{\gamma}$ via $\mathbf{B}\mathbf{B}^{\mathsf{H}} = \eta\mathbf{I} - \mathbf{A}\mathbf{A}^{\mathsf{H}} \succeq \mathbf{0}$:

$$\mathbf{\Sigma} = \mathbf{A}\left(\mathrm{diag}(\boldsymbol{\gamma}) + \frac{\sigma^2}{\eta}\mathbf{I}\right)\mathbf{A}^{\mathsf{H}} + \frac{\sigma^2}{\eta}\mathbf{B}\mathbf{B}^{\mathsf{H}}, \quad \eta = \mathrm{tr}\left(\mathbf{A}\mathbf{A}^{\mathsf{H}}\right)$$

$\Rightarrow$ $\bar{\mathbf{D}} = \mathrm{diag}(\boldsymbol{\gamma}) + \frac{\sigma^2}{\eta}\mathbf{I} \succ \mathbf{0}$ while $\eta = \mathrm{tr}\left(\mathbf{A}\mathbf{A}^{\mathsf{H}}\right)$ ensures $\mathbf{B}\mathbf{B}^{\mathsf{H}} \succeq \mathbf{0}$

Majorizing the concave term and the surrogate objective

Linearization of the log-det:

$$\log|\mathbf{\Sigma}| \leq \operatorname{tr}\left((\mathbf{\Sigma}^k)^{-1}\mathbf{\Sigma}\right) + \text{const},$$

where $\mathbf{\Sigma} = \mathbf{A}\,\mathrm{diag}(\boldsymbol{\gamma})\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I}$ and $\mathbf{\Sigma}^k = \mathbf{A}\,\mathrm{diag}(\boldsymbol{\gamma}^k)\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I}$.

Combining both majorizations, yields simplified surrogate objective:

$$\boxed{\underset{\boldsymbol{\gamma}\geq\mathbf{0}}{\text{minimize}} \sum_{i=1}^{N} \frac{(\gamma_i^k + \sigma^2/\eta)^2}{\gamma_i + \sigma^2/\eta}[\mathbf{b}_i^k]^{\mathsf{H}}\hat{\mathbf{\Sigma}}\mathbf{b}_i^k + \sum_{i=1}^{N}\gamma_i\mathbf{a}_i^{\mathsf{H}}\mathbf{b}_i^k}, \qquad \mathbf{b}_i^k = (\mathbf{\Sigma}^k)^{-1}\mathbf{a}_i$$

Solution:

$$\gamma_i^{k+1} = \left[\left(\gamma_i^k + \frac{\sigma^2}{\eta}\right)\sqrt{\frac{[\mathbf{b}_i^k]^{\mathsf{H}}\hat{\mathbf{\Sigma}}\mathbf{b}_i^k}{\mathbf{a}_i^{\mathsf{H}}\mathbf{b}_i^k}} - \frac{\sigma^2}{\eta}\right]_+, \qquad i = 1,\ldots,N,$$

Algorithm 3: CL-MM algorithm

Input : SCM $\hat{\mathbf{\Sigma}}$, measurement matrix $\mathbf{A} \in \mathbb{C}^{M \times N}$, noise variance $\sigma^2 > 0$
max. # of iterations $I_{\max}$, initial power $\gamma_{\text{init}} \in \mathbb{R}^N_{\geq 0}$

Initialize: $\eta = \operatorname{tr}\left(\mathbf{A}\mathbf{A}^{\mathsf{H}}\right)$ $(,\gamma_{\text{init}} = \mathbf{0})$

$\gamma^0 = \gamma_{\text{init}}$, $\delta = 10^{-3}$

for $k = 1, \ldots, I_{\max}$ **do**

  $\mathbf{\Sigma}^k = \mathbf{A}\,\operatorname{diag}(\gamma^k)\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I}$,

  $\mathbf{B}^k = \left(\mathbf{\Sigma}^k\right)^{-1}\mathbf{A} = (\mathbf{b}_1^k \cdots \mathbf{b}_N^k)$.

  For all $i \in \{1, \ldots, N\}$ solve in parallel:

$$\gamma_i^{(k+1)} = \left[\left(\gamma_i^k + \frac{\sigma^2}{\eta}\right)\sqrt{\frac{[\mathbf{b}_i^k]^{\mathsf{H}}\hat{\mathbf{\Sigma}}\mathbf{b}_i^k}{\mathbf{a}_i^{\mathsf{H}}\mathbf{b}_i^k}} - \frac{\sigma^2}{\eta}\right]_+$$

  if $\left\|\gamma^{k+1} - \gamma^k\right\|_2 / \left\|\gamma^k\right\|_2 < \delta$ **then**

    break

Output : $\hat{\gamma} = \gamma^{k+1}$.

Successive Convex Approximation (SCA)

In SBL, we need to solve the non-convex type-II likelihood:

$$\min_{\gamma \geq 0} \ell(\gamma) = g(\gamma) + f(\gamma),$$

$$g(\gamma) = \mathrm{tr}((\mathbf{A}\,\mathrm{diag}(\gamma)\mathbf{A}^{\mathsf{H}} + \sigma^2\,\mathbf{I})^{-1}\,\hat{\boldsymbol{\Sigma}})$$

$$f(\gamma) = \log|\mathbf{A}\,\mathrm{diag}(\gamma)\mathbf{A}^{\mathsf{H}} + \sigma^2\,\mathbf{I}|.$$

where g is convex, while f is non-convex

We **linearize** non-convex term $f(\gamma)$ and use SCA framework to minimize the surrogate.

Note: $\nabla_{\gamma_i} f(\gamma) = \mathbf{a}_i^{\mathsf{H}}\boldsymbol{\Sigma}^{-1}\mathbf{a}_i$

Successive Convex Approximation (SCA)

In SBL, we need to solve the non-convex type-II likelihood:

$$\min_{\boldsymbol{\gamma}\geq 0} \ell(\boldsymbol{\gamma}) = g(\boldsymbol{\gamma}) + f(\boldsymbol{\gamma}),$$

$$g(\boldsymbol{\gamma}) = \mathrm{tr}((\mathbf{A}\,\mathrm{diag}(\boldsymbol{\gamma})\mathbf{A}^{\mathsf{H}} + \sigma^2\,\mathbf{I})^{-1}\,\hat{\boldsymbol{\Sigma}})$$

$$f(\boldsymbol{\gamma}) = \log|\mathbf{A}\,\mathrm{diag}(\boldsymbol{\gamma})\mathbf{A}^{\mathsf{H}} + \sigma^2\,\mathbf{I}|.$$

where g is convex, while f is non-convex

We **linearize** non-convex term $f(\boldsymbol{\gamma})$ and use SCA framework to minimize the surrogate.

Note: $\nabla_{\gamma_i} f(\boldsymbol{\gamma}) = \mathbf{a}_i^{\mathsf{H}}\boldsymbol{\Sigma}^{-1}\mathbf{a}_i$

At iteration $k+1$, by linearizing $f(\boldsymbol{\gamma})$, the N functions to be minimized in parallel are

$$\begin{aligned}\tilde{\ell}_i(\gamma_i \mid \boldsymbol{\gamma}^k) &= g(\gamma_i, \boldsymbol{\gamma}^k_{\setminus i}) + \nabla_{\gamma_i} f(\boldsymbol{\gamma}^k)\,(\gamma_i - \gamma_i^k)\\ &= -\frac{\gamma_i[\mathbf{c}_i^k]^{\mathsf{H}}\hat{\boldsymbol{\Sigma}}\mathbf{c}_i^k}{1+\gamma_i\,\mathbf{a}_i^{\mathsf{H}}\mathbf{c}_i^k} + \gamma_i\,\mathbf{a}_i^{\mathsf{H}}\mathbf{b}_i^k + \text{const}\,,\end{aligned}$$

where

$$\mathbf{c}_i^k = (\boldsymbol{\Sigma}^k - \gamma_i^k\mathbf{a}_i\mathbf{a}_i^{\mathsf{H}})^{-1}\mathbf{a}_i,$$

$$\boldsymbol{\Sigma}^k = \mathbf{A}\,\mathrm{diag}(\boldsymbol{\gamma}^k)\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I},$$

$$\mathbf{b}_i^k = \left(\boldsymbol{\Sigma}^k\right)^{-1}\mathbf{a}_i.$$

SCA surrogate objective

The surrogate objective has closed form solution:

$$\hat{\gamma}_i(\boldsymbol{\gamma}^k) = \arg\min_{\gamma_i \geq 0} \tilde{\ell}_i(\gamma_i \mid \boldsymbol{\gamma}^k)$$
$$= \left[\gamma_i^k + \sqrt{\frac{[\mathbf{b}_i^k]^{\mathsf{H}} \hat{\boldsymbol{\Sigma}} \mathbf{b}_i^k}{(\mathbf{a}_i^{\mathsf{H}} \mathbf{b}_i^k)^3} - \frac{1}{\mathbf{a}_i^{\mathsf{H}} \mathbf{b}_i^k}} \right]_+$$

with $\boldsymbol{\Sigma}^k = \mathbf{A}\,\mathrm{diag}(\boldsymbol{\gamma}^k)\mathbf{A}^{\mathsf{H}} + \sigma^2 \mathbf{I}$ and $\mathbf{b}_i^k = \left(\boldsymbol{\Sigma}^k\right)^{-1} \mathbf{a}_i$.

Then apply smoothing:

$$\boldsymbol{\gamma}^{k+1} = (1-\alpha_k)\boldsymbol{\gamma}^k + \alpha_k \hat{\boldsymbol{\gamma}}\left(\boldsymbol{\gamma}^k\right)$$
$$= \boldsymbol{\gamma}^k + \alpha_k \left(\hat{\boldsymbol{\gamma}}(\boldsymbol{\gamma}^k) - \boldsymbol{\gamma}^k\right)$$

$\alpha_k \in (0,1] :=$ ***step-size sequence*** controls the update along $\mathbf{d}^k = \hat{\boldsymbol{\gamma}}(\boldsymbol{\gamma}^k) - \boldsymbol{\gamma}^k$, verifying:

$$\sum_k^{\infty} \alpha_k = \infty, \quad \sum_{k=0}^{\infty} (\alpha_k)^2 < +\infty$$

We use [SS18]:

$$\alpha_k = \alpha_{k-1}\left(1 - \epsilon\, \alpha_{k-1}\right)$$

with $\epsilon \in (0,1)$ is a constant, and $\alpha_0 < 1/\epsilon$.

Default: $\alpha_0 = 0.99$, $\epsilon = 0.06$

Algorithm 4: CL-SCA algorithm

Input : $\hat{\boldsymbol{\Sigma}}$, $\mathbf{A} \in \mathbb{C}^{M \times N}$, $\sigma^2 > 0$, $\boldsymbol{\gamma}_{\text{init}} \in \mathbb{R}^M_{\geq 0}$)

Initialize: $\alpha_0 = 0.99$, $\epsilon = 0.06$ (, $\boldsymbol{\gamma}_{\text{init}} = \mathbf{0}$)

$k = 0$, $\boldsymbol{\gamma}^0 = \boldsymbol{\gamma}_{\text{init}}$, $\delta = 1e^{-2}$.

for $k = 1, \ldots, I_{\max}$ **do**

$\boldsymbol{\Sigma}^k = \mathbf{A}\,\mathrm{diag}(\boldsymbol{\gamma}^k)\mathbf{A}^{\mathsf{H}} + \sigma^2 \mathbf{I}$

$\mathbf{B}^k = (\boldsymbol{\Sigma}^k)^{-1}\mathbf{A} = (\mathbf{b}_1^k \cdots \mathbf{b}_N^k)$

For all $i \in \{1, \ldots, M\}$ solve in parallel:

$$\hat{\gamma}_i(\boldsymbol{\gamma}^k) = \left[\gamma_i^k + \sqrt{\frac{[\mathbf{b}_i^k]^{\mathsf{H}} \hat{\boldsymbol{\Sigma}} \mathbf{b}_i^k}{\left(\mathbf{a}_i^{\mathsf{H}} \mathbf{b}_i^k\right)^3}} - \frac{1}{\mathbf{a}_i^{\mathsf{H}} \mathbf{b}_i^k} \right]_+$$

$\boldsymbol{\gamma}^{k+1} = \boldsymbol{\gamma}^k + \alpha_k \mathbf{d}^k$ with $\mathbf{d}^k = \hat{\boldsymbol{\gamma}}(\boldsymbol{\gamma}^k) - \boldsymbol{\gamma}^k$

if $\|\mathbf{d}^k\|_2 < \delta$ **then**

break

$\alpha_k = \alpha_{k-1}\left(1 - \epsilon\,\alpha_{k-1}\right)$

Output : $\hat{\boldsymbol{\gamma}}$

Complexity: $\mathcal{O}(NM^2)$

Same order as CWO, CL-MM, but fewer iterations.

- CWO: $\sim 2\times$ slower
- MSBL using EM: $> 10\times$ slower, 100s of iterations

CL-SCA

Input : $\hat{\boldsymbol{\Sigma}}$, $\mathbf{A} \in \mathbb{C}^{M \times N}$, $\sigma^2 > 0$, $\boldsymbol{\gamma}_{\text{init}} \in \mathbb{R}^M_{\geq 0}$)

Initialize: $\alpha_0 = 0.99$, $\epsilon = 0.06$ (, $\boldsymbol{\gamma}_{\text{init}} = \mathbf{0}$)

$k = 0$, $\boldsymbol{\gamma}^0 = \boldsymbol{\gamma}_{\text{init}}$, $\delta = 1e^{-2}$.

for $k = 1, \ldots, I_{\max}$ **do**

- $\boldsymbol{\Sigma}^k = \mathbf{A}\,\mathrm{diag}(\boldsymbol{\gamma}^k)\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I}$
- $\mathbf{B}^k = (\boldsymbol{\Sigma}^k)^{-1}\mathbf{A} = (\mathbf{b}_1^k \ \cdots \mathbf{b}_N^k)$
- For all $i \in \{1, \ldots, M\}$ solve in parallel:

$$\hat{\gamma}_i(\boldsymbol{\gamma}^k) = \left[\gamma_i^k + \sqrt{\frac{[\mathbf{b}_i^k]^{\mathsf{H}}\hat{\boldsymbol{\Sigma}}\mathbf{b}_i^k}{\left(\mathbf{a}_i^{\mathsf{H}}\mathbf{b}_i^k\right)^3}} - \frac{1}{\mathbf{a}_i^{\mathsf{H}}\mathbf{b}_i^k}\right]_+$$

- $\boldsymbol{\gamma}^{k+1} = \boldsymbol{\gamma}^k + \alpha_k\,\mathbf{d}^k$ with $\mathbf{d}^k = \hat{\boldsymbol{\gamma}}(\boldsymbol{\gamma}^k) - \boldsymbol{\gamma}^k$
- **if** $\|\mathbf{d}^k\|_2 < \delta$ **then**
  - break
- $\alpha_k = \alpha_{k-1}\left(1 - \epsilon\,\alpha_{k-1}\right)$

Output : $\hat{\boldsymbol{\gamma}}$

CL-MM

Input : $\hat{\boldsymbol{\Sigma}}$, $\mathbf{A} \in \mathbb{C}^{M \times N}$, $\sigma^2 > 0$, $\boldsymbol{\gamma}_{\text{init}} \in \mathbb{R}^N_{\geq 0}$

Initialize: $\eta = \mathrm{tr}\left(\mathbf{A}\mathbf{A}^{\mathsf{H}}\right)$ (,$\boldsymbol{\gamma}_{\text{init}} = \mathbf{0}$)

$\boldsymbol{\gamma}^0 = \boldsymbol{\gamma}_{\text{init}}$, $\delta = 10^{-3}$

for $k = 1, \ldots, I_{\max}$ **do**

- $\boldsymbol{\Sigma}^k = \mathbf{A}\,\mathrm{diag}(\boldsymbol{\gamma}^k)\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I}$,
- $\mathbf{B}^k = \left(\boldsymbol{\Sigma}^k\right)^{-1}\mathbf{A} = (\mathbf{b}_1^k \cdots \mathbf{b}_N^k)$.
- For all $i \in \{1, \ldots, N\}$ solve in parallel:

$$\gamma_i^{k+1} = \left[\left(\gamma_i^k + \frac{\sigma^2}{\eta}\right)\sqrt{\frac{[\mathbf{b}_i^k]^{\mathsf{H}}\hat{\boldsymbol{\Sigma}}\mathbf{b}_i^k}{\mathbf{a}_i^{\mathsf{H}}\mathbf{b}_i^k}} - \frac{\sigma^2}{\eta}\right]_+$$

- **if** $\left\|\boldsymbol{\gamma}^{k+1} - \boldsymbol{\gamma}^k\right\|_2 / \left\|\boldsymbol{\gamma}^k\right\|_2 < \delta$ **then**
  - break

Output : $\hat{\boldsymbol{\gamma}} = \boldsymbol{\gamma}^{k+1}$.

Comparison to M-SBL of Wipf and Rao (2007)

1. **Initialization**: Set $\boldsymbol{\Gamma}^0 = \mathrm{diag}(\boldsymbol{\gamma}^0)$ for some initial $\boldsymbol{\gamma}^0 \in \mathbb{R}^M_{>0}$, and set $k = 0$.
2. **E-step**:

$$\boldsymbol{\Sigma}^k = \mathbf{A}\boldsymbol{\Gamma}^k\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I},$$
$$\mathbf{X}^k = \boldsymbol{\Gamma}^k\mathbf{A}^{\mathsf{H}}(\boldsymbol{\Sigma}^k)^{-1}\mathbf{Y},$$
$$\boldsymbol{\Sigma}_x^k = \boldsymbol{\Gamma}^k - \boldsymbol{\Gamma}^k\mathbf{A}^{\mathsf{H}}(\boldsymbol{\Sigma}^k)^{-1}\mathbf{A}\boldsymbol{\Gamma}^k.$$

3. **M-step**:

$$\boldsymbol{\Gamma}^{k+1} = \mathrm{diag}\Big(L^{-1}\mathbf{X}^k(\mathbf{X}^k)^{\mathsf{H}} + \boldsymbol{\Sigma}_x^k\Big).$$

4. **Convergence check**: If not converged, increment k and repeat steps 2-3

CL-MM

$$\gamma_i^{k+1} = \left[\left(\gamma_i^k + \frac{\sigma^2}{\eta}\right)\sqrt{\frac{[\mathbf{b}_i^k]^{\mathsf{H}}\hat{\boldsymbol{\Sigma}}\mathbf{b}_i^k}{\mathbf{a}_i^{\mathsf{H}}\mathbf{b}_i^k}} - \frac{\sigma^2}{\eta}\right]_+$$

CL-SCA:

$$\hat{\gamma}_i(\boldsymbol{\gamma}^k) = \left[\gamma_i^k + \sqrt{\frac{[\mathbf{b}_i^k]^{\mathsf{H}}\hat{\boldsymbol{\Sigma}}\mathbf{b}_i^k}{(\mathbf{a}_i^{\mathsf{H}}\mathbf{b}_i^k)^3}} - \frac{1}{\mathbf{a}_i^{\mathsf{H}}\mathbf{b}_i^k}\right]_+$$

$$\gamma_i^{k+1} = \gamma_i^k + \alpha^k\left(\hat{\gamma}_i(\boldsymbol{\gamma}^k) - \gamma_i^k\right)$$

References

CL-MM:

Majorization-Minimization Covariance Learning Algorithm for Sparse Signal Recovery
Majdoddin Esfandiari, Esa Ollila and Daniel P. Palomar
Proc. *EUSIPCO-2026*, pp. 1-5, August 31-Sept. 4 2026, Bruges, Belgium.

CL-SCA:

Joint Activity Detection and Channel Estimation for Massive Random Access Using SBL and SCA
Esa Ollila, Majdoddin Esfandiari, and Daniel P. Palomar
Proc. *SPAWC-2026*, pp. 1-5, Sept. 6-9 2026, Athens, Greece.

Conclusions

- (M-)SBL is versatile empirical Bayes approach for sparse approximation in MMV model.
- Applications: device activity detection in IoT, sparse DoA estimation, reconstruction of brain actitivity using EEG, etc.
- EM algorithm for SBL has ***slow convergence*** and ***does not provide sparse solution***
- We proposed greedy approaches (CL-MP and CL-OMP) and optimization approaches (CL-MM and CL-SCA) that are fast to compute, outperform the state-of-the-art in diverse settings and applications.
- **Robustness** was not mentioned, **deep learning** was not mentioned!

Conclusions

- (M-)SBL is versatile empirical Bayes approach for sparse approximation in MMV model.
- Applications: device activity detection in IoT, sparse DoA estimation, reconstruction of brain actitivity using EEG, etc.
- EM algorithm for SBL has ***slow convergence*** and ***does not provide sparse solution***
- We proposed greedy approaches (CL-MP and CL-OMP) and optimization approaches (CL-MM and CL-SCA) that are fast to compute, outperform the state-of-the-art in diverse settings and applications.
- **Robustness** was not mentioned, **deep learning** was not mentioned!
- For robust generalization of CWO and CL-MP methods, see:

> Jobust Activity Detection for Massive Random Access
> Xinjue Wang, Esa Ollila, Sergiy A. Vorobyov
> *IEEE Transactions on Signal Processing*, vol. 73, pp. 3513 - 3527, 2025.

References i

- Yoram Bresler, *Maximum likelihood estimation of a linearly structured covariance with application to antenna array processing*, Fourth Annual ASSP Workshop on Spectrum Estimation and Modeling, IEEE, 1988, pp. 172–175.
- M. F. Duarte and Y. C. Eldar, *Structured compressed sensing: From theory to applications*, IEEE Transactions on Signal Processing **59** (2011), no. 9, 4053–4085.
- M. Elad, *Sparse and redundant representations*, Springer, New York, 2010.
- Alexander Fengler, Saeid Haghighatshoar, Peter Jung, and Giuseppe Caire, *Non-Bayesian activity detection, large-scale fading coefficient estimation, and unsourced random access with a massive MIMO receiver*, IEEE Transactions on Information Theory **67** (2021), no. 5, 2925–2951.

References ii

- Anita Faul and Michael Tipping, *Analysis of sparse bayesian learning*, Advances in neural information processing systems **14** (2001).
- Ali Hashemi, Chang Cai, Yijing Gao, Sanjay Ghosh, Klaus-Robert Müller, Srikantan S Nagarajan, and Stefan Haufe, *Joint learning of full-structure noise in hierarchical bayesian regression models*, IEEE Transactions on Medical Imaging **43** (2022), no. 2, 610–624.
- Rohan R Pote and Bhaskar D Rao, *Theory and practice of light-weight sequential sbl algorithm: An alternative to omp*, IEEE Transactions on Signal Processing (2025).
- Ying Sun, Prabhu Babu, and Daniel P Palomar, *Majorization-minimization algorithms in signal processing, communications, and machine learning*, IEEE Transactions on Signal Processing **65** (2016), no. 3, 794–816.

References iii

- Kamil Senel and Erik G Larsson, *Grant-free massive MTC-enabled massive MIMO: A compressive sensing approach*, IEEE Transactions on Communications **66** (2018), no. 12, 6164–6175.
- Petre Stoica and Arye Nehorai, *On the concentrated stochastic likelihood function in array signal processing*, Circuits, Systems and Signal Processing **14** (1995), no. 5, 669–674.
- Gesualdo Scutari and Ying Sun, *Parallel and distributed successive convex approximation methods for big-data optimization*, Multi-Agent Optimization: Cetraro, Italy 2014, Springer, 2018, pp. 141–308.
- David P Wipf and Bhaskar D Rao, *An empirical Bayesian strategy for solving the simultaneous sparse approximation problem*, IEEE Transactions on Signal Processing **55** (2007), no. 7, 3704–3716.

References iv

Tarik Yardibi, Jian Li, Petre Stoica, Ming Xue, and Arthur B Baggeroer, *Source localization and sensing: A nonparametric iterative adaptive approach based on weighted least squares*, IEEE Transactions on Aerospace and Electronic Systems **46** (2010), no. 1, 425–443.