Subspace Averaging in Multi-Sensor Array Processing

Ignacio Santamaría[1], Louis Scharf[2], Vaibhav Garg[1], David Ramírez[3]

[1]Department of Communications Engineering, University of Cantabria, Spain
[2]Department of Mathematics, Colorado State University, Spain
[3]Department of Signal Theory and Communications, University Carlos III de Madrid, Spain

SIAM AG 2019, Bern, July 2019

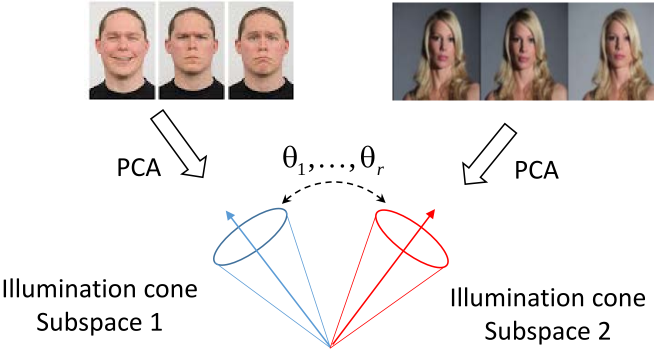
SIAM Conference on Applied Algebraic Geometry
Tuesday 9-Saturday 13 July 2019
Bern, Switzerland

UNIVERSIDAD DE CANTABRIA

Contents

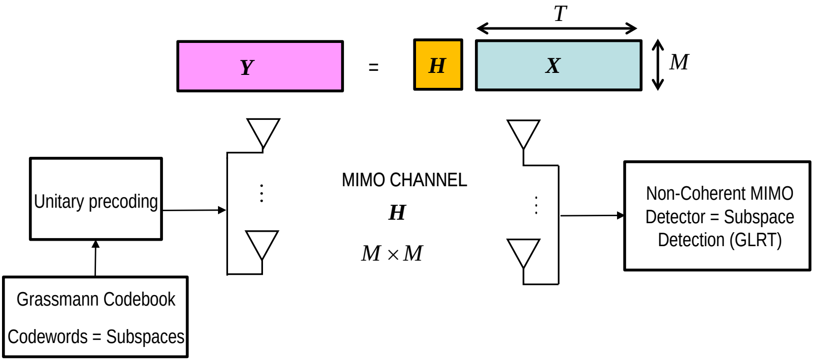
Subspaces as data objects

- In many signal processing problems data sets are high dimensional, but their intrinsic dimension is much smaller than the dimension of the ambient space
- Data objects admit a **subspace** representation
- Example: Image, video processing & computer vision



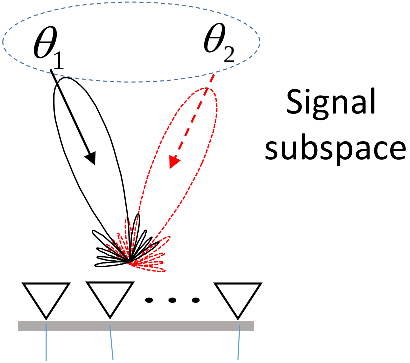
Subspaces in wireless communication problems

- Non-coherent MIMO communications



Subspaces in wireless communication problems

- Multi-sensor array processing: source enumeration



Problem

Given a sequence of experimentally derived subspaces

$$\langle \mathbf{V}_r \rangle \in \mathbb{G}(q_{V_r}, n), \quad r = 1, \dots, R$$

1. to obtain a central subspace and estimate its dimension
2. to apply the resulting algorithm as a method for source enumeration in array processing under the challenging conditions of
   - high-dimensional data (massive MIMO)
   - few snapshots (small sample regime)

Subspace averaging

- The **Karcher mean** or Riemannian center of mass is

$$\langle \mathbf{U} \rangle = \underset{\langle \mathbf{U} \rangle \in \mathbb{G}(s,n)}{\operatorname{argmin}} \quad \frac{1}{R} \sum_{r=1}^{R} \sum_{i=1}^{\min(s, \dim(\mathbf{V}_r))} \theta_{r,i}^2$$

- The **extrinsic mean** is

$$\langle \mathbf{U} \rangle = \underset{\langle \mathbf{U} \rangle \in \mathbb{G}(s,n)}{\operatorname{argmin}} \quad \frac{1}{2R} \sum_{r=1}^{R} \| \mathbf{P}_{\mathbf{V}_r} - \mathbf{P}_{\mathbf{U}} \|_F^2$$

Closed-form solution $\mathbf{U}_s^* = (\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_s) = \mathbf{F}_s$ where $\mathbf{F}_s$ is a matrix containing the s largest left eigenvectors of the average projection matrix

$$\overline{\mathbf{P}} = \frac{1}{R} \sum_{r=1}^{R} \mathbf{P}_r = \frac{1}{R} \sum_{r=1}^{R} \mathbf{V}_r \mathbf{V}_r^H$$

Order estimation problem

- The proposed **order estimation criterion** is

$$(s^*, \mathbf{P}_s^*) = \underset{\substack{s \in \{0,1,\ldots,n\} \\ \mathbf{P} \in \mathbb{P}(s,n)}}{\arg\min} \quad \frac{1}{2R} \sum_{r=1}^{R} \|\mathbf{P} - \mathbf{P}_r\|_F^2 \,,$$

  where $\mathbb{P}(s, n)$ denotes the set of rank-s projection matrices
- Writing $\mathbf{P} = \mathbf{U}\mathbf{U}^H$ and expanding the cost function we obtain

$$\min_{\mathbf{U} \in \mathbb{S}(s,n)} \quad \text{tr}\left(\mathbf{U}^H(\mathbf{I} - 2\overline{\mathbf{P}})\mathbf{U}\right),$$

  where $\overline{\mathbf{P}} = \frac{1}{R}\sum_{r=1}^{R} \mathbf{P}_r$ with eigenvalues $0 \leq k_i \leq 1$
- The **optimal order s^*** is the number of negative eigenvalues of the matrix

$$\mathbf{S} = \mathbf{I} - 2\overline{\mathbf{P}},$$

A few comments

- Does not rely on any statistical model for the generated data and is free of penalty terms or tuning parameters, unlike most order determination criteria like MDL, AIC, BIC
- The eigenvectors of the average projection matrix whose eigenvalues are above $1/2$ determine the **signal subspace**
- If all eigenvalues are smaller than $1/2 \rightarrow$ No central subspace, **noise only hypothesis**
- The order fitting rule arises naturally when we force $\overline{\mathbf{P}}$ to be a projection matrix (quantizing its eigenvalues to $0/1$)

A probabilistic interpretation

Given a collection of subspaces $\mathbf{P}_r$ with average $\overline{\mathbf{P}} = \mathbf{FKF}^H$

- The eigenvalues $0 \leq k_i \leq 1$ can be interpreted as **probabilities**
- This allows us to define a **discrete distribution** $\mathcal{D}$ on the set of projection matrices (or subspaces) with orientation matrix $\mathbf{U}$ and concentration parameters $\boldsymbol{\alpha}$

  $$\mathbf{P} \sim \mathcal{D}(\mathbf{U}, \boldsymbol{\alpha}),$$

  useful as a random subspace generation mechanism
- A measure of the spread of the collection of subspaces is given by the **sample entropy**

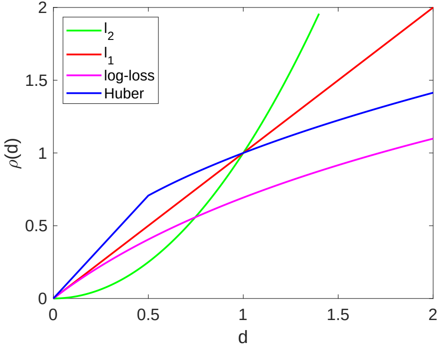
  $$\hat{H} = \frac{1}{n} \sum_{i=1}^{n} \left(-k_i \log(k_i) - (1 - k_i) \log(1 - k_i) \right)$$

  useful for subspace clustering

Robust formulation (outliers)

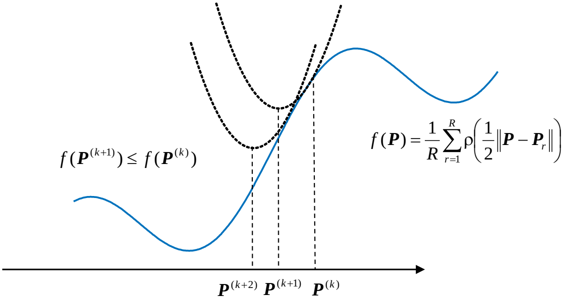
$$\min_{\substack{s \in \{0,1,\dots,n\} \\ \mathbf{P} \in \mathbb{P}(s,n)}} \quad \frac{1}{R} \sum_{r=1}^{R} \rho \left(\frac{1}{2} \|\mathbf{P} - \mathbf{P}_r\|_F^2 \right)$$

where $\rho(\cdot)$ is a smooth concave function



Majorization-minimization (MM) algorithms

- At each iteration, use a majorizer of the objective function



- As a majorizer, we linearize the concave function $\rho(\cdot)$

$$\min_{\mathbf{P}\in\mathbb{P}(s,n)} \quad \frac{1}{R}\sum_{r=1}^{R}\rho\left(d_r^2\left(\mathbf{P}^{(k)}\right)\right)+\rho'\left(d_r^2\left(\mathbf{P}^{(k)}\right)\right)\left(d_r^2\left(\mathbf{P}\right)-d_r^2\left(\mathbf{P}^{(k)}\right)\right)$$

where $d_r^2\left(\mathbf{P}\right)=\frac{1}{2}\left\|\mathbf{P}-\mathbf{P}_r\right\|_F^2$

- At each iteration we solve a weighted SA problem

$$\min_{\substack{s \in \{0,1,\ldots,n\} \\ \mathbf{P} \in \mathbb{P}(s,n)}} \quad \frac{1}{2} \sum_{r=1}^{R} \bar{w}_r^{(k)} \left\| \mathbf{P} - \mathbf{P}_r \right\|_F^2 .$$

where

$$\bar{w}_r^{(k)} = \frac{\rho'\left(d_r^2\left(\mathbf{P}^{(k)}\right)\right)}{\sum_{r=1}^{R} \rho'\left(d_r^2\left(\mathbf{P}^{(k)}\right)\right)}, \quad \bar{w}_r^{(k)} \geq 0, \quad \sum_r \bar{w}_r^{(k)} = 1,$$

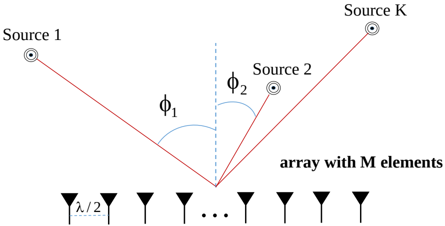
- The optimal order at iteration $k+1$, $s^{(k+1)}$, is the number of negative eigenvalues of the matrix

$$\mathbf{S}^{(k)} = \mathbf{I} - 2\overline{\mathbf{P}}_w^{(k)}.$$

where $\overline{\mathbf{P}}_w^{(k)}$ is now a weighted average projection matrix

$$\overline{\mathbf{P}}_w^{(k)} = \sum_{r=1}^{R} \bar{w}_r^{(k)} \mathbf{P}_r .$$

Application to multi-sensor array processing



- Uniform linear array (ULA) with M antennas
- K sources
- Electrical angles: $\theta_k = \frac{2\pi d}{\lambda} \sin(\phi_k)$
- $M >> K$ antennas (e.g., massive MIMO, large-scale arrays)
- Small-sample regime: few snapshots

- Received signal

$$\mathbf{x}[t] = \begin{bmatrix} 1 & \dots & 1 \\ e^{j\theta_1} & \dots & e^{j\theta_K} \\ \vdots & \vdots & \vdots \\ e^{j(M-1)\theta_1} & \dots & e^{j(M-1)\theta_K} \end{bmatrix} \begin{bmatrix} s_1[t] \\ \vdots \\ s_K[t] \end{bmatrix} + \begin{bmatrix} e_1[t] \\ e_2[t] \\ \vdots \\ e_M[t] \end{bmatrix} = \mathbf{A}\mathbf{s}[t]+\mathbf{e}[t],$$

- $\mathbf{e}[t] \sim \mathcal{CN}_M(\mathbf{0}, \sigma^2\mathbf{I})$
- $\mathbf{s}[t] \sim \mathcal{CN}_K(\mathbf{0}, \mathbf{S})$
- $\mathbf{R} = E\left[\mathbf{x}[t]\mathbf{x}^H[t]\right] = \mathbf{ASA}^H + \sigma^2\mathbf{I}$

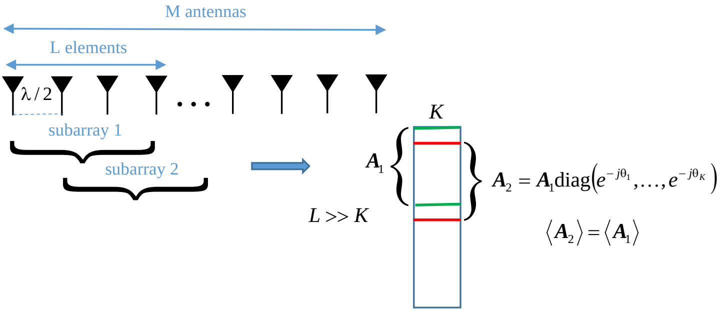
Source enumeration (order estimation) problem

- To estimate K from $\hat{\mathbf{R}} = \frac{1}{N}\sum_{t=1}^{N} \mathbf{x}[t]\mathbf{x}^H[t]$
- Typically solved by information-theoretic criteria such as MDL (penalized functions of the eigenvalues of $\hat{\mathbf{R}}$)
- These methods underperform in the small-sample regime

Subspace averaging for source enumeration

- To apply SA we need a collection of subspaces to start with
- The extracted subspaces should overlap as much as possible with the true signal subspace
- But the noise portions of each subspace should be "as independent as possible"
- How can we generate a good collection of subspaces for this problem?
  1. Exploiting the **shift-invariance** property of ULAs
  2. **Random sampling** (bootstrapping)

Shift invariance property



- Number of subarray $S = M - L + 1$
- For each L-dimensional subarray:
  1. Estimate the sample covariance matrix $\hat{\mathbf{R}}_s$, $s = 1, \ldots, S$
  2. Extract a subspace of dimension k_{max} $(K < k_{max} << L)$

Random sampling

More than one subspace per subarray? $\rightarrow$ Draw subspaces from an appropriate distribution $\mathbf{P} \sim \mathcal{D}(\mathbf{U}, \boldsymbol{\alpha})$

$$\hat{\boldsymbol{R}}_s = \underbrace{[\boldsymbol{u}_1 \quad \boldsymbol{u}_2 \quad \cdots \quad \boldsymbol{u}_L]} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_L \end{bmatrix} \begin{bmatrix} \boldsymbol{u}_1^H \\ \boldsymbol{u}_2^H \\ \vdots \\ \boldsymbol{u}_L^H \end{bmatrix}$$

$\boldsymbol{U} =$ **Orientations**

$$\Delta\lambda_i = \begin{cases} \lambda_i - \lambda_i, & i = 1,\ldots,M\text{-}1 \\ 0, & i = M \end{cases} \Rightarrow \underbrace{\alpha_i = \frac{\Delta\lambda_i}{\sum_i \Delta\lambda_i}}$$

$\boldsymbol{\alpha} =$ **Concentrations**

Each random subspace is iteratively constructed as follows:

1. Initialize $\langle \mathbf{V} \rangle = \emptyset$
2. While rank$(\mathbf{V}) \leq k_{max}$ do
   2.1 Generate a random draw $\langle \mathbf{G} \rangle \sim \mathcal{D}(\mathbf{U}, \boldsymbol{\alpha})$
   2.2 $\langle \mathbf{V} \rangle = \langle \mathbf{V} \rangle \bigcup \langle \mathbf{G} \rangle$

SA algorithm

- **Input**: $\hat{\mathbf{R}}$, L, T and k_{max}
- **Output**: $\hat{k}_{SA}$
- **For** $s = 1, \ldots, S$ **do**
  1. Extract $\hat{\mathbf{R}}_s$ from $\hat{\mathbf{R}}$ and obtain $\hat{\mathbf{R}}_s = \mathbf{U}_s \mathbf{\Sigma}_s \mathbf{U}_s^H$
  2. Generate T random subspaces from $\hat{\mathbf{R}}_s$
  3. Compute the projection matrices $\mathbf{P}_{st} = \mathbf{V}_{st} \mathbf{V}_{st}^H$
- Compute

$$\overline{\mathbf{P}} = \frac{1}{ST} \sum_{s=1}^{S} \sum_{t=1}^{T} \mathbf{P}_{st}$$

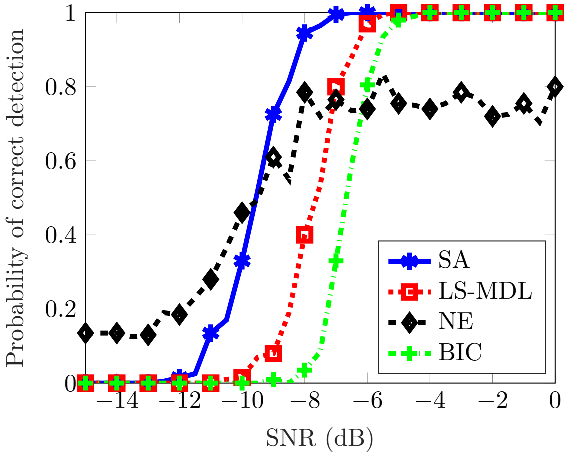
  and its eigenvalues $(k_1, \ldots, k_L)$
- Estimate $\hat{k}_{SA}$ as the number of eigenvalues of $\overline{\mathbf{P}}$ larger than $1/2$

Simulation parameters

- K narrowband incoherent unit-power signals, with DOAs separated by Δ_θ in electrical angle
- ULA with M antennas and half-wavelength element separation
- $L = M - 5 \implies$ total number of subarrays $S = 6$
- For each subarray we generate $T = 20$ random subspaces of dimension $k_{max} = \lfloor M/5 \rfloor$
- 120 subspaces on $\mathbb{G}(k_{max}, L)$ to average
- $\text{SNR} = 10 \log_{10}(1/\sigma^2)$
- Methods under comparison:
  - LS-MDL criterion (Huang/So TSP 2013)
  - NE criterion (Nadakuditi/Edelman TSP 2008)
  - BIC method for large-scale arrays (Huang *et. al.* TVT 2016)

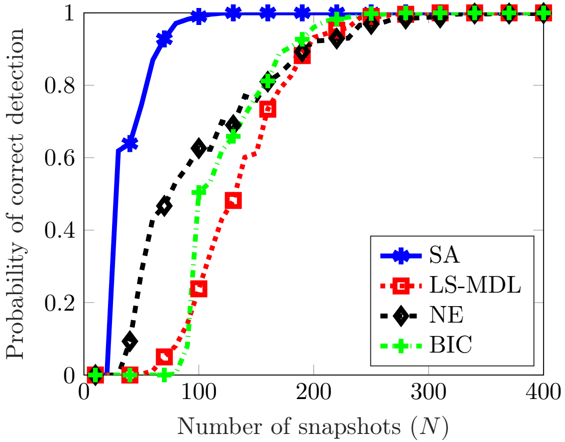
Scenario 1

- $K = 3$ sources separated $\Delta_\theta = 2^\circ$
- $M = 100$ antennas, $N = 60$ snapshots, $L = \lfloor M - 5 \rfloor$



Scenario 2

- $K = 3$ sources separated $\Delta_\theta = 10^\circ$
- $M = 100$ antennas, $SNR = -16$ dB, $L = \lfloor M - 5 \rfloor$,



Conclusions

- An automatic order-fitting rule for extracting the dimension of the average subspace that minimizes the extrinsic distance
  - Quantization of the average projection matrix
  - Free of penalty terms
- Scale-independent subspace modeling vs scale-dependent covariance modeling
- Application to source enumeration in array processing
  - Generation of a collection of subspaces:
    - Exploiting the shift invariance property of ULAs
    - Generating random draws from $\mathcal{D}(\mathbf{U}, \boldsymbol{\alpha})$
  - Competitive results in problems with large number of antennas (high-dimensional ambient spaces) and relatively few snapshots

Thank you for your attention!

References

[1] V. Garg, I. Santamaria, D. Ramirez, and L. L. Scharf, "Subspace Averaging and Order Determination for Source Enumeration", *IEEE Transactions on Signal Processing*, vol. 67, issue 11, pp. 3028-3041, June, 2019.

[2] I. Santamaria, L. L. Scharf, C. Peterson, M. Kirby, and J. Francos, "An order fitting rule for optimal subspace averaging", *IEEE Workshop on Statistical Signal Processing* (SSP), Palma de Mallorca, June , 2016.

[3] I. Santamaria, D. Ramirez, and L. L. Scharf, "Subspace Averaging for Source Enumeration in Large Array", *IEEE Statistical Signal Processing Workshop* (SSP), Freiburg, Germany, June, 2018.

[4] V. Garg, I. Santamaria, "Source Enumeration in Non-White Noise and Small Sample Size via Subspace Averaging", *European Signal Processing Conference*, La Corua, Spain, Sept. 2019.

[5] P. Turaga, A. Veeraraghavan, A. Srivastava, and R. Chellappa, "Statistical computations on Grassmann and Stiefel manifolds for image and video-based recognition" *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 11, pp. 22732286, 2011.

[6] R. Marrinan, J. R. Beveridge, B. Draper, M. Kirby, and C. Peterson, "Finding the subspace mean or median to fit your need" *Proc. of Computer Vision and Pattern Recognition* (CVPR), Columbus, OH, USA, pp. 10821089, Jun. 2014.

[7] A. Edelman, T. Arias, S. T. Smith, "The geometry of algo-rithms with orthogonality constraints", *SIAM J. Matrix Anal. Appl.*, vol. 20, no. 2, pp. 303-353, 1998.

[8] A. Srivastava, E. Klassen,"Monte Carlo extrinsic estimatorsof manifold-valued parameters",*IEEE Trans. Signal Process.*,vol. 50, no. 2, pp. 299-308, Aug. 2002.