SiML: Sieved Maximum Likelihood for Array Signal Processing

ICASSP, 6-11 June 2021, Toronto

Authors

Matthieu Simeoni*, **Paul Hurley**†

*École Polytechnique Fédérale de Lausanne (EPFL)

†Western Sydney University (WSU)

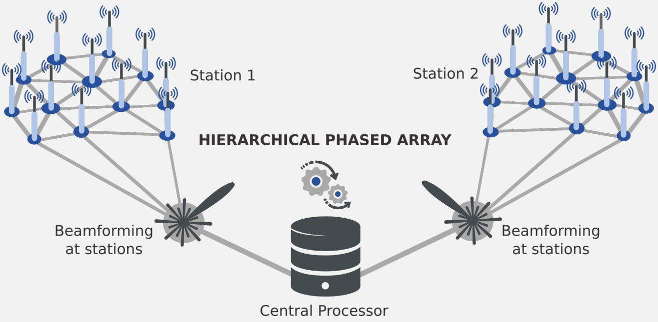
April 19, 2021

Contents

Introduction

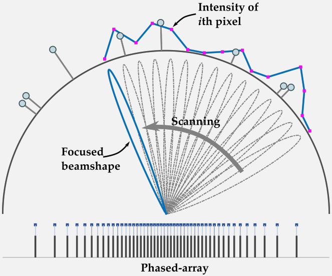
Array Signal Processing

- Array signal processing is concerned with the *sensing*, *processing* and *estimation* of **random wavefields**.
- Myriad of applications: *acoustics*, *radio-interferometry*, *radar* and *sonar* systems, *wireless* networks, and *medical imagery*.
- **Typical task**: estimate the **intensity field** (variance).



Spectral-based & Parametric Methods

- Two competing approaches:
  - ↪ **Spectral-based methods:** sequential scanning by *beamforming* (MB, MVDR, AAR...).
  - ↪ **Parametric methods:** estimate parameters of a *statistical* model.
- Spectral-based methods are *simple*, *computationally attractive* and *generic*. **Limited accuracy** for *low SNR* or *coherent signals*.
- Parametric methods have *excellent performance* but very **intensive computationally**.



Point Source Data Model and DOA

- Parametric methods are concerned with the pb of *Direction of Arrival (DOA)* estimation.
- **Point source model: (Q sources, L sensors)**

$$Y(\boldsymbol{p}_i) = \sum_{q=1}^{Q} S_q \, \exp\left[-\frac{2\pi j}{\lambda}\langle \boldsymbol{r}_q, \boldsymbol{p}_i\rangle\right] + n_i \quad \forall i = 1, \ldots, L,$$

$$\Leftrightarrow \boldsymbol{Y} = A(\rho)\boldsymbol{S} + \boldsymbol{n} \in \mathbb{C}^L.$$

$\hookrightarrow$ $\boldsymbol{S} = [S_1, \cdots, S_Q] \sim \mathbb{C}\mathcal{N}_Q(0, R)$, $R \in \mathbb{C}^{Q\times Q}$ positive semi-definite,
$\hookrightarrow$ $\boldsymbol{n} = [n_1, \cdots, n_L] \sim \mathbb{C}\mathcal{N}_L(0, \sigma I)$, $\sigma > 0$, $\boldsymbol{n}$ independent of $\boldsymbol{S}$,
$\hookrightarrow$ $A(\rho) \in \mathbb{C}^{L\times Q}$ is the *steering matrix*,
$\hookrightarrow$ $\rho = \{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_Q\} \subset \mathbb{S}^2$ are the *unknown source directions*,
$\hookrightarrow$ $\{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_L\} \subset \mathbb{R}^3$ are the sensor locations.

Stochastic Maximum Likelihood (SML)

- **Stochastic Maximum Likelihood (SML)** is the most famous parametric method. Provides closed-form expressions for the ML estimates of $R,\ \sigma,\ \{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_Q\}$.
- For $\boldsymbol{Q} < \boldsymbol{L}$ and $A(\rho)$ **full rank**:

$$\hat{\sigma}(\hat{\rho}) = \frac{\mathrm{Tr}\left[P^{\perp}_{A(\hat{\rho})}\hat{\Sigma}\right]}{L-Q}, \qquad \hat{R}(\hat{\rho}) = A(\hat{\rho})^{\dagger}\left[\hat{\Sigma} - \hat{\sigma}(\hat{\rho})I\right]A(\hat{\rho})^{H,\dagger},$$

$$\hat{\rho} = \arg\min\left\{\log\left|A(\hat{\rho})\hat{R}(\hat{\rho})A(\hat{\rho})^{H} + \hat{\sigma}(\hat{\rho})I_L\right|,\ \rho = \{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_Q\} \subset \mathbb{S}^2\right\}.$$

  - $\hookrightarrow$ $\hat{\Sigma}$ is the ML estimate of $\mathbb{E}[\boldsymbol{Y}\boldsymbol{Y}^H]$.
  - $\hookrightarrow$ Consistent estimates, asymptotically efficient (attain CRB).
  - $\hookrightarrow$ Estimation involves minimisation in $(\mathbb{S}^2)^Q$ of a highly non-linear function (Newton method).

SML is too Restrictive...

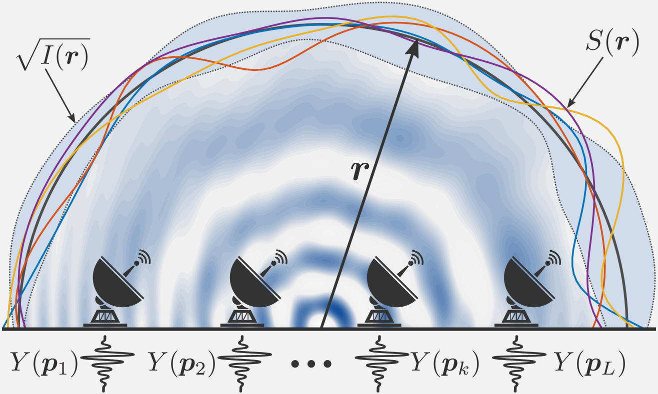
- SML is too **computationally intensive**.
- SML requires $Q < L$: in some applications $Q \sim 10^6$!
  - $\hookrightarrow$ Building arrays sufficiently large is **unrealistic**.
- Modern arrays can make the point source approximation break.
- Need a data model allowing to handle **extended sources**.

A Functional Data Model

Random Amplitude Function

- **Random amplitude function**: $\mathcal{S} = \{S(\boldsymbol{r}) : \Omega \to \mathbb{C}, \boldsymbol{r} \in \mathbb{S}^2\}$.
- From **Huygens-Fresnel** principle and **Fraunhofer** equation:

$$Y(\boldsymbol{p}_i) = \int_{\mathbb{S}^2} S(\boldsymbol{r}) \exp\left(-j\frac{2\pi}{\lambda}\langle \boldsymbol{r}, \boldsymbol{p}_i\rangle\right) d\boldsymbol{r} \; + \; n_i.$$



Sampling Operator Φ^*

- Defining $\phi_i(\boldsymbol{r}) := \exp\left(j2\pi\langle \boldsymbol{r}, \boldsymbol{p}_i\rangle/\lambda\right)$, we can write

$$\boldsymbol{Y} = \begin{bmatrix} Y(\boldsymbol{p}_1) \\ \vdots \\ Y(\boldsymbol{p}_L) \end{bmatrix} = \begin{bmatrix} \langle S, \phi_1\rangle \\ \vdots \\ \langle S, \phi_L\rangle \end{bmatrix} + \begin{bmatrix} n_1 \\ \vdots \\ n_L \end{bmatrix} = \Phi^* S + \boldsymbol{n}.$$

  - ↪ Sample paths $s_\omega : \mathbb{S}^2 \to \mathbb{C}$ of S are in $\mathcal{H} = \mathcal{L}^2(\mathbb{S}^2)$,
  - ↪ $\Phi^* : \mathcal{H} \to \mathbb{C}^L$ is the **sampling operator**.

- Assuming S to be Gaussian, we have $\boldsymbol{Y} \sim \mathbb{C}\mathcal{N}_L(\mathbf{0}, \Sigma)$, where

$$(\Sigma)_{ij} = \iint\limits_{\mathbb{S}^2\times\mathbb{S}^2} \kappa(\boldsymbol{r}, \boldsymbol{\rho})\, \phi_i^*(\boldsymbol{r})\phi_j(\boldsymbol{\rho})\, d\boldsymbol{r} d\boldsymbol{\rho} \;+\; \sigma\,\delta_{ij}, \quad i,j = 1,\ldots,L.$$

  - ↪ $\kappa(\boldsymbol{r}, \boldsymbol{\rho}) = \mathbb{E}[S(\boldsymbol{r})S^*(\boldsymbol{\rho})]$ is the **covariance kernel**,
  - ↪ Characterises S. When uncorrelated, the **intensity function** $I(\boldsymbol{r}) = \kappa(\boldsymbol{r}, \boldsymbol{r})$ is enough.

Data Model (Population Version)

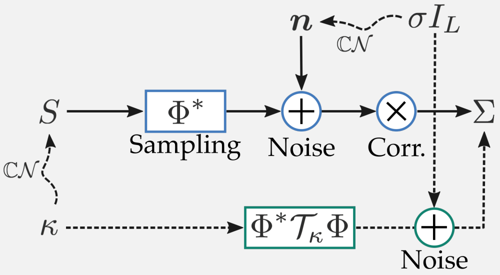
- Using the sampling operator Φ^* and its adjoint Φ we get:

$$\Sigma = \Phi^* \mathcal{T}_\kappa \Phi \; + \; \sigma I_L.$$

- $\mathcal{T}_\kappa : \mathcal{H} \to \mathcal{H}$ is the **covariance operator** associated to κ:

$$(\mathcal{T}_\kappa f)(\boldsymbol{r}) := \int_{\mathbb{S}^2} \kappa(\boldsymbol{r}, \boldsymbol{\rho}) f(\boldsymbol{\rho}) d\boldsymbol{\rho}, \quad f \in \mathcal{H}, \boldsymbol{r} \in \mathbb{S}^2,$$

$\hookrightarrow$ It is customary to write: $\kappa = \text{vec}(T_\kappa)$.



Data Model (Empirical Version)

- In practice Σ is estimated from N i.i.d. observations of $\boldsymbol{Y}$:

$$\hat{\Sigma} = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{y}_i \boldsymbol{y}_i^H.$$

- $\hat{\Sigma}$ follows a L-variate **complex Wishart distribution**:

$$N\hat{\Sigma} \overset{d}{\sim} \mathbb{C}\mathcal{W}_L(N, \Sigma).$$

with density

$$f(W) \propto \begin{cases} \frac{|W|^{N-L}}{|\Sigma|^N} \exp\left(-\mathrm{tr}\left(\Sigma^{-1} W\right)\right), & \text{for } W \text{ definite-positive,} \\ 0, & \text{otherwise.} \end{cases}$$

Sieved Maximum Likelihood

Likelihood Function

- We wish to construct **ML estimates** for κ and σ.
- Given the data $\hat{\Sigma}$ the negative log-likelihood function is given by

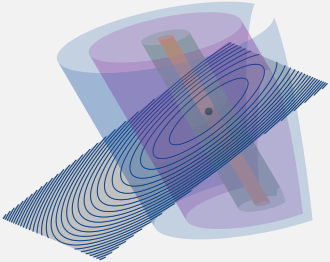
$$\ell\left(\kappa, \sigma|\hat{\Sigma}\right) = \mathrm{Tr}\left[\left(\Phi^*\mathcal{T}_\kappa\Phi + \sigma I_L\right)^{-1}\hat{\Sigma}\right] + \log\left|\Phi^*\mathcal{T}_\kappa\Phi + \sigma I_L\right|.$$

$\hookrightarrow$ Well-defined for any noise power $\sigma > 0$.

- **Infinitely many solutions!** Indeed, take $f \in \mathcal{N}(\Phi^*) = \mathcal{R}(\Phi)^\perp$. Then, adding $\overline{f} \otimes f$ to $\mathcal{T}_\kappa$ does not change ℓ. Indeed,

$$\begin{aligned}\Phi^*\left(\mathcal{T}_\kappa + \overline{f} \otimes f\right)\Phi &= \Phi^*\mathcal{T}_\kappa\Phi + \Phi^*\left(\overline{f} \otimes f\right)\Phi \\ &= \Phi^*\mathcal{T}_\kappa\Phi + \left(\bar{\Phi}^* \otimes \Phi^*\right)\left(\overline{f} \otimes f\right) \\ &= \Phi^*\mathcal{T}_\kappa\Phi + \underbrace{\overline{\Phi^* f} \otimes \Phi^* f}_{=0} = \Phi^*\mathcal{T}_\kappa\Phi.\end{aligned}$$

Sieved Maximum Likelihood



- We need to **constrain** the optimisation problem:

$$\kappa = \left(\bar{\Psi} \otimes \Psi\right) \mathrm{vec}(R) = \sum_{i,j=1}^{M} R_{ij} \, \bar{\psi}_j \otimes \psi_i,$$

$$\Leftrightarrow \mathcal{T}_\kappa = \Psi R \Psi^*,$$

- Maximum likelihood estimates of R, σ are given by

$$\hat{R}, \hat{\sigma} = \arg \min_{\substack{R \in \mathbb{C}^{M^2} \\ \sigma > 0}} \mathrm{Tr}\left[\left(GRG^H + \sigma I_L\right)^{-1} \hat{\Sigma}\right] + \log\left|GRG^H + \sigma I_L\right|.$$

$\hookrightarrow$ $G = \Phi^* \Psi \in \mathbb{C}^{L \times M}$ is the **Gram matrix**.

- This procedure is known as the **method of sieves**.

Case 1: Known Noise Power

- **Suppose σ known**. We need to have $M \leq L$ and G full rank for identifiability.
- Closed-form formula for $\hat{R}$:
$$\hat{R} = G^{\dagger} \left[\hat{\Sigma} - \sigma I_L\right] \left(G^{\dagger}\right)^H .$$
- The maximum likelihood of $\hat{\kappa}$ is then given by
$$\begin{aligned}\hat{\kappa} &= \sum_{i,j=1}^{M} \hat{R}_{ij} \, \bar{\psi}_j \otimes \psi_i \\ &= \left(\bar{\Psi} \otimes \Psi\right) \mathrm{vec}\left(G^{\dagger} \left[\hat{\Sigma} - \sigma I_L\right] \left(G^{\dagger}\right)^H\right) \\ &= \left(\bar{\Psi} \otimes \Psi\right) \left[\bar{G}^{\dagger} \otimes G^{\dagger}\right] \left[\mathrm{vec}(\hat{\Sigma}) - \sigma \mathrm{vec}(I_L)\right],\end{aligned}$$

Case 1: Known Noise Power ($M = L$)

- When $M = L$ there is a nice **geometrical interpretation**.
- Indeed, on expectation we have:
$$\mathbb{E}[\hat{\kappa}] = \left(\bar{\Psi} \otimes \Psi\right) \left[\bar{G}^{-1} \otimes G^{-1}\right] \left(\bar{\Phi} \otimes \Phi\right)^* \kappa.$$
- The **interpolation operator** $\left(\bar{\Psi} \otimes \Psi\right) \left[\bar{G}^{-1} \otimes G^{-1}\right]$ is *consistent* with $\left(\bar{\Phi} \otimes \Phi\right)^*$:
$$\left(\bar{\Phi} \otimes \Phi\right)^* \left(\bar{\Psi} \otimes \Psi\right) \left[\bar{G}^{-1} \otimes G^{-1}\right] = I_{L^2}.$$
- $\hat{\kappa}$ is hence unbiased, consistent and efficient estimate of the **oblique projection** of κ on $\mathcal{R}(\bar{\Psi} \otimes \Psi)$.

Case 2: Unknown Noise Power

- **Suppose** σ **unknown**. We need to have $M < L$ and G full rank for identifiability (one more parameter to estimate!).
- Closed-form formula for $\hat{R}$ and $\hat{\sigma}$:

$$\hat{\sigma} = \frac{\mathrm{Tr}\left(\hat{\Sigma} - GG^{\dagger}\hat{\Sigma}\right)}{L - M}, \qquad \hat{R} = G^{\dagger}\left[\hat{\Sigma} - \hat{\sigma}I\right]\left(G^{\dagger}\right)^{H}.$$

- Again, the maximum likelihood of $\hat{\kappa}$ is given by

$$\hat{\kappa} = \left(\bar{\Psi} \otimes \Psi\right)\left[\bar{G}^{\dagger} \otimes G^{\dagger}\right]\left[\mathrm{vec}(\hat{\Sigma}) - \sigma\mathrm{vec}(I_L)\right].$$

  - $\hookrightarrow$ No geometric interpretation in general.
  - $\hookrightarrow$ When $M \simeq L$ then it is *almost* an oblique projection.

On the choice of Ψ

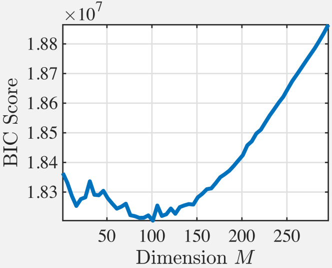
- **Plenty of possibilities!** Canonical choice:

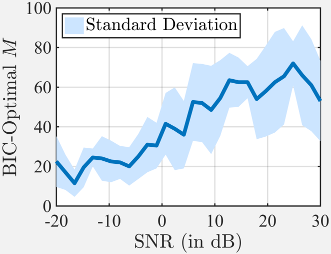
$$\Psi = \Phi W, \qquad W \in \mathbb{C}^{L \times M}.$$

- Convenient as Gram is **computable analytically**.
- Indeed, $G = \Phi^* \Phi W = HW$, where $H \in \mathbb{C}^{L \times L}$ is given by:

$$(H)_{ij} = 4\pi \operatorname{sinc}(2\pi \|\boldsymbol{p}_i - \boldsymbol{p}_j\|_2 / \lambda), \quad i, j = 1, \ldots, L.$$

- Choose M as minimising $BIC(M) = -2\hat{\ell}_M + 2M^2 \log(L)$.





Comparison with SML

	SML	SiML
Type of sources	Point Sources	Arbitrary Sources
Nb. of sources	Less than antennas	Unlimited
Identifiability	Steering matrix full-rank	Gram matrix full-rank
Statistical Efficiency	Consistent, efficient	Consistent, efficient
Computational Efficiency	Very intensive	Fast (Kronecker)

Results

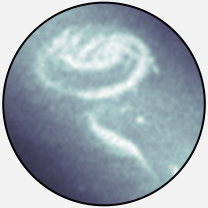
Experimental Setup

- Extended source, no correlation.
- We image the intensity field:

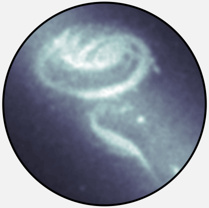
$$\hat{I}(\boldsymbol{r}) = \sum_{i,j=1}^{L} \hat{R}_{ij}\psi_i(\boldsymbol{r})\bar{\psi}_i(\boldsymbol{r}).$$

- $L = 300$ antennas.
- $N = 2000$ samples.
- Algorithms: SiML, MB, MVDR, AAR.
- Metrics: MSE, Contrast RMS.

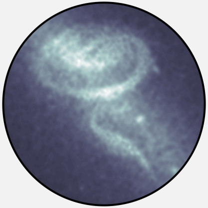
Results



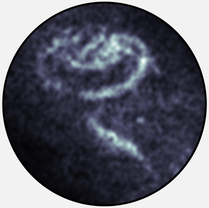
(a) MB estimate.



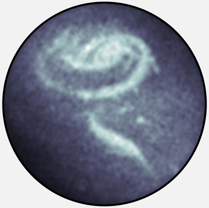
(b) MVDR estimate.



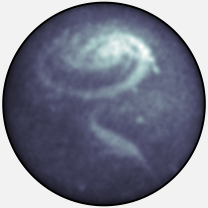
(c) AAR estimate.



(d) SiML estimate ($M = 15$).



(e) SiML estimate (BIC-selected $M = 101$).

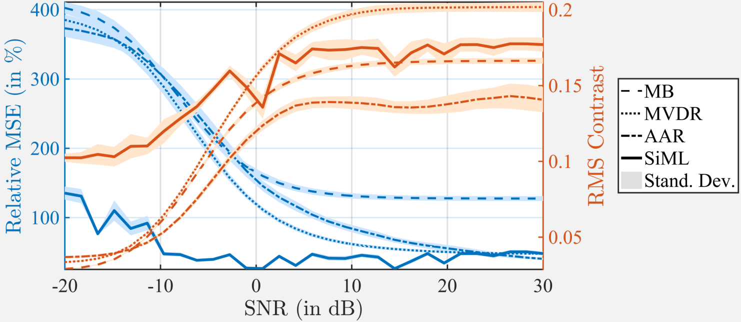


(f) SiML estimate ($M = 296$).

Results

	Relative MSE (in %)			RMS Contrast		
MB	148			0.16		
MVDR	79			**0.2**		
AAR	71			0.16		
SiML	M=15	M=101	M=295	M=15	M=101	M=295
	22	53	46	0.15	0.18	0.15

(g) Relative MSE and RMS contrast scores.



(h) Performance of the different algorithms for various SNR.

Conclusion

Conclusions & Future Work

- SiML generalises the SML to a wider class of signals,
- Allows for arbitrarily shaped, possibly correlated, sources
- Nice geometrical interpretation in the case of known noise power.
- Superior to state-of-the-art subspace-based methods, both in terms of accuracy and contrast.
- The tensor product structure makes SiML very computationally efficient.

- Derive SVD of $\mathcal{T}_{\hat{\kappa}}$,
- Derive distribution and confidence intervals for $\mathcal{T}_{\hat{\kappa}}$.
- Investigate different Ψ.