Sparse Bayesian Learning: A Beamforming and Toeplitz Approximation Perspective

Bhaskar D Rao
University of California, San Diego
Email: brao@ucsd.edu

Some References

Some References

1. M. Tipping, Sparse Bayesian learning and the relevance vector machine, J. Machine Learning Research, vol. 1, pp. 211-244, 2001.

Some References

1. M. Tipping, Sparse Bayesian learning and the relevance vector machine, J. Machine Learning Research, vol. 1, pp. 211-244, 2001.
2. David Wipf and Bhaskar D. Rao, Sparse Bayesian learning for basis selection. IEEE Transactions on Signal processing, 2004.
3. David Wipf and Bhaskar D. Rao,(2007), An empirical Bayesian strategy for solving the simultaneous sparse approximation problem. IEEE Transactions on Signal Processing, 55(7), 3704-3716.

Some References

1. M. Tipping, Sparse Bayesian learning and the relevance vector machine, J. Machine Learning Research, vol. 1, pp. 211-244, 2001.
2. David Wipf and Bhaskar D. Rao, Sparse Bayesian learning for basis selection. IEEE Transactions on Signal processing, 2004.
3. David Wipf and Bhaskar D. Rao,(2007), An empirical Bayesian strategy for solving the simultaneous sparse approximation problem. IEEE Transactions on Signal Processing, 55(7), 3704-3716.
4. D. Wipf, B. Rao, S. Nagarajan, Latent Variable Bayesian Models for Promoting Sparsity, IEEE Trans. Info Theory, 2011.
5. D. Wipf and S. Nagarajan, Iterative Reweighted ℓ_1 and ℓ_2 Methods for Finding Sparse Solutions. IEEE Journal of Selected Topics in Signal Processing, 2010.

Some References

1. M. Tipping, Sparse Bayesian learning and the relevance vector machine, J. Machine Learning Research, vol. 1, pp. 211-244, 2001.
2. David Wipf and Bhaskar D. Rao, Sparse Bayesian learning for basis selection. IEEE Transactions on Signal processing, 2004.
3. David Wipf and Bhaskar D. Rao,(2007), An empirical Bayesian strategy for solving the simultaneous sparse approximation problem. IEEE Transactions on Signal Processing, 55(7), 3704-3716.
4. D. Wipf, B. Rao, S. Nagarajan, Latent Variable Bayesian Models for Promoting Sparsity, IEEE Trans. Info Theory, 2011.
5. D. Wipf and S. Nagarajan, Iterative Reweighted ℓ_1 and ℓ_2 Methods for Finding Sparse Solutions. IEEE Journal of Selected Topics in Signal Processing, 2010.
6. D. Wipf and S. Nagarajan,(2007, June), Beamforming using the relevance vector machine. In Proceedings of the 24th international conference on Machine learning. ACM.

Some References

1. M. Tipping, Sparse Bayesian learning and the relevance vector machine, J. Machine Learning Research, vol. 1, pp. 211-244, 2001.
2. David Wipf and Bhaskar D. Rao, Sparse Bayesian learning for basis selection. IEEE Transactions on Signal processing, 2004.
3. David Wipf and Bhaskar D. Rao,(2007), An empirical Bayesian strategy for solving the simultaneous sparse approximation problem. IEEE Transactions on Signal Processing, 55(7), 3704-3716.
4. D. Wipf, B. Rao, S. Nagarajan, Latent Variable Bayesian Models for Promoting Sparsity, IEEE Trans. Info Theory, 2011.
5. D. Wipf and S. Nagarajan, Iterative Reweighted ℓ_1 and ℓ_2 Methods for Finding Sparse Solutions. IEEE Journal of Selected Topics in Signal Processing, 2010.
6. D. Wipf and S. Nagarajan,(2007, June), Beamforming using the relevance vector machine. In Proceedings of the 24th international conference on Machine learning. ACM.
7. M. Al-Shoukairi, M. and B. Rao, Sparse Signal Recovery Using MPDR Estimation, 2019 ICASSP.

Outline

Outline

- Background

Outline

- Background
  - Beamforming and Minimum Power Distortionless Response (MPDR) beamforming

Outline

- Background
  - Beamforming and Minimum Power Distortionless Response (MPDR) beamforming
  - Sparse Bayesian Learning (SBL)

Outline

- Background
  - Beamforming and Minimum Power Distortionless Response (MPDR) beamforming
  - Sparse Bayesian Learning (SBL)
- Connection between SBL and MPDR beamforming

Outline

- Background
  - Beamforming and Minimum Power Distortionless Response (MPDR) beamforming
  - Sparse Bayesian Learning (SBL)
- Connection between SBL and MPDR beamforming
- Uniform Linear Arrays and Toeplitz Matrix Approximation

Outline

- Background
  - Beamforming and Minimum Power Distortionless Response (MPDR) beamforming
  - Sparse Bayesian Learning (SBL)
- Connection between SBL and MPDR beamforming
- Uniform Linear Arrays and Toeplitz Matrix Approximation
- Nested Arrays and more sources than sensors

Beamforming and MPDR in Array Processing

Beamforming and MPDR in Array Processing

Narrow-band signal model for an array with N sensors.

$$\mathbf{y}(\omega_c, n) = \mathbf{V}(\omega_c, \mathbf{k}_0)x_0[n] + \sum_{l=1}^{D-1} \mathbf{V}(\omega_c, \mathbf{k}_l)x_l[n] + \mathbf{Z}[n], n = 1, .., L$$

ω_c is the carrier frequency, $\mathbf{k}_l$ is the source direction, and $\mathbf{V}(\omega_c, \mathbf{k})$ is the array manifold and is the response of the array to a source at direction $\mathbf{k}$.

Beamforming and MPDR in Array Processing

Narrow-band signal model for an array with N sensors.

$$\mathbf{y}(\omega_c, n) = \mathbf{V}(\omega_c, \mathbf{k}_0)x_0[n] + \sum_{l=1}^{D-1} \mathbf{V}(\omega_c, \mathbf{k}_l)x_l[n] + \mathbf{Z}[n], n = 1, .., L$$

ω_c is the carrier frequency, $\mathbf{k}_l$ is the source direction, and $\mathbf{V}(\omega_c, \mathbf{k})$ is the array manifold and is the response of the array to a source at direction $\mathbf{k}$.

Far-Field model: $\mathbf{V}(\omega_c, \mathbf{k}) = [1, e^{-j\omega_c\tau_1(\mathbf{k})}, \dots, e^{-j\omega_c\tau_{N-1}(\mathbf{k})}]$ and a function of array geometry.

Beamforming and MPDR in Array Processing

Narrow-band signal model for an array with N sensors.

$$\mathbf{y}(\omega_c, n) = \mathbf{V}(\omega_c, \mathbf{k}_0)x_0[n] + \sum_{l=1}^{D-1} \mathbf{V}(\omega_c, \mathbf{k}_l)x_l[n] + \mathbf{Z}[n], n = 1, .., L$$

ω_c is the carrier frequency, $\mathbf{k}_l$ is the source direction, and $\mathbf{V}(\omega_c, \mathbf{k})$ is the array manifold and is the response of the array to a source at direction $\mathbf{k}$.

Far-Field model: $\mathbf{V}(\omega_c, \mathbf{k}) = [1, e^{-j\omega_c \tau_1(\mathbf{k})}, \ldots, e^{-j\omega_c \tau_{N-1}(\mathbf{k})}]$ and a function of array geometry.

Uniform Linear Array (ULA):

Beamforming and MPDR in Array Processing

Narrow-band signal model for an array with N sensors.

$$\mathbf{y}(\omega_c, n) = \mathbf{V}(\omega_c, \mathbf{k}_0)x_0[n] + \sum_{l=1}^{D-1} \mathbf{V}(\omega_c, \mathbf{k}_l)x_l[n] + \mathbf{Z}[n], n = 1, .., L$$

ω_c is the carrier frequency, $\mathbf{k}_l$ is the source direction, and $\mathbf{V}(\omega_c, \mathbf{k})$ is the array manifold and is the response of the array to a source at direction $\mathbf{k}$.

Far-Field model: $\mathbf{V}(\omega_c, \mathbf{k}) = [1, e^{-j\omega_c\tau_1(\mathbf{k})}, \ldots, e^{-j\omega_c\tau_{N-1}(\mathbf{k})}]$ and a function of array geometry.

Uniform Linear Array (ULA): Sensors placed on a line with a unform spacing of d.

Beamforming and MPDR in Array Processing

Narrow-band signal model for an array with N sensors.

$$\mathbf{y}(\omega_c, n) = \mathbf{V}(\omega_c, \mathbf{k}_0)x_0[n] + \sum_{l=1}^{D-1} \mathbf{V}(\omega_c, \mathbf{k}_l)x_l[n] + \mathbf{Z}[n], n = 1, .., L$$

ω_c is the carrier frequency, $\mathbf{k}_l$ is the source direction, and $\mathbf{V}(\omega_c, \mathbf{k})$ is the array manifold and is the response of the array to a source at direction $\mathbf{k}$.

Far-Field model: $\mathbf{V}(\omega_c, \mathbf{k}) = [1, e^{-j\omega_c\tau_1(\mathbf{k})}, \ldots, e^{-j\omega_c\tau_{N-1}(\mathbf{k})}]$ and a function of array geometry.

Uniform Linear Array (ULA): Sensors placed on a line with a unform spacing of d.

$e^{-j\omega_c\tau_m(\mathbf{k_l})} = e^{-j\omega_l m}$, with $\omega_m = 2\pi\frac{d}{\lambda}\cos(\theta_l)$ where θ_l is the elevation angle. Common to use $\frac{d}{\lambda} = \frac{1}{2}$

Beamforming and MPDR in Array Processing

Narrow-band signal model for an array with N sensors.

$$\mathbf{y}(\omega_c, n) = \mathbf{V}(\omega_c, \mathbf{k}_0)x_0[n] + \sum_{l=1}^{D-1} \mathbf{V}(\omega_c, \mathbf{k}_l)x_l[n] + \mathbf{Z}[n], n = 1, .., L$$

ω_c is the carrier frequency, $\mathbf{k}_l$ is the source direction, and $\mathbf{V}(\omega_c, \mathbf{k})$ is the array manifold and is the response of the array to a source at direction $\mathbf{k}$.

Far-Field model: $\mathbf{V}(\omega_c, \mathbf{k}) = [1, e^{-j\omega_c\tau_1(\mathbf{k})}, \ldots, e^{-j\omega_c\tau_{N-1}(\mathbf{k})}]$ and a function of array geometry.

Uniform Linear Array (ULA): Sensors placed on a line with a unform spacing of d.

$e^{-j\omega_c\tau_m(\mathbf{k_l})} = e^{-j\omega_l m}$, with $\omega_m = 2\pi\frac{d}{\lambda}\cos(\theta_l)$ where θ_l is the elevation angle. Common to use $\frac{d}{\lambda} = \frac{1}{2}$
So array manifold can be denoted by $\mathbf{V}(\omega_l)$.

Assumptions

Assumptions

$$
\begin{aligned}
\mathbf{y}(n) &= \sum_{l=0}^{D-1} \mathbf{V}(\omega_l) x_l[n] + \mathbf{Z}[n] \\
&= [\mathbf{V}_0, \mathbf{V}_1, \ldots, \mathbf{V}_{D-1}] \begin{bmatrix} x_0[n] \\ x_1[n] \\ \vdots \\ x_{D-1}[n] \end{bmatrix} + \mathbf{Z}[n] \\
&= \mathbf{V}\mathbf{x}[n] + \mathbf{Z}[n]
\end{aligned}
$$

where $\mathbf{V} \in C^{N \times D}$, and $\mathbf{x}[n] \in C^{D \times 1}$.

Assumptions

$$\begin{aligned}
\mathbf{y}(n) &= \sum_{l=0}^{D-1} \mathbf{V}(\omega_l) x_l[n] + \mathbf{Z}[n] \\
&= [\mathbf{V}_0, \mathbf{V}_1, \ldots, \mathbf{V}_{D-1}] \begin{bmatrix} x_0[n] \\ x_1[n] \\ \vdots \\ x_{D-1}[n] \end{bmatrix} + \mathbf{Z}[n] \\
&= \mathbf{V}\mathbf{x}[n] + \mathbf{Z}[n]
\end{aligned}$$

where $\mathbf{V} \in C^{N\times D}$, and $\mathbf{x}[n] \in C^{D\times 1}$.

Assumptions: $E(\mathbf{x}[n]) = \mathbf{0}_{D\times 1}$, $E(\mathbf{x}[n]\mathbf{Z}^H[n]) = \mathbf{0}_{D\times N}$ and temporally uncorrelated.

Covariance Matrices of interest

Covariance Matrices of interest

Source Covariance matrix:

Covariance Matrices of interest

Source Covariance matrix: $\mathbf{R}_x = E(\mathbf{x}[n]\mathbf{x}^H[n])$ where $\mathbf{R}_x \in C^{D \times D}$, and the diagonal elements are $p_0, p_1, ..., p_{D-1}$, the source powers

Covariance Matrices of interest

Source Covariance matrix: $\mathbf{R}_x = E(\mathbf{x}[n]\mathbf{x}^H[n])$ where $\mathbf{R}_x \in C^{D \times D}$, and the diagonal elements are $p_0, p_1, ..., p_{D-1}$, the source powers

For uncorrelated sources $\mathbf{R}_x$ is a diagonal matrix, i.e. $\mathbf{R}_x = \text{diag}(p_0, p_1, \ldots, p_{D-1})$.

Covariance Matrices of interest

Source Covariance matrix: $\mathbf{R}_x = E(\mathbf{x}[n]\mathbf{x}^H[n])$ where $\mathbf{R}_x \in C^{D\times D}$, and the diagonal elements are $p_0, p_1, ..., p_{D-1}$, the source powers

For uncorrelated sources $\mathbf{R}_x$ is a diagonal matrix, i.e. $\mathbf{R}_x = \mathrm{diag}(p_0, p_1, \ldots, p_{D-1})$.

Data Covariance Matrix

$$\mathbf{R}_y = E(\mathbf{y}[n]\mathbf{y}^H[n]) = \mathbf{R}_s + \sigma_z^2\mathbf{I} = \mathbf{V}\mathbf{R}_x\mathbf{V}^H + \sigma_z^2\mathbf{I}$$

$\mathbf{R}_y \in C^{N\times N}$, $\mathbf{R}_s \in C^{N\times N}$, $\mathbf{R}_x \in C^{D\times D}$, and $\mathbf{V} \in C^{N\times D}$

Covariance Matrices of interest

Source Covariance matrix: $\mathbf{R}_x = E(\mathbf{x}[n]\mathbf{x}^H[n])$ where $\mathbf{R}_x \in C^{D \times D}$, and the diagonal elements are $p_0, p_1, ..., p_{D-1}$, the source powers

For uncorrelated sources $\mathbf{R}_x$ is a diagonal matrix, i.e. $\mathbf{R}_x = \text{diag}(p_0, p_1, \ldots, p_{D-1})$.

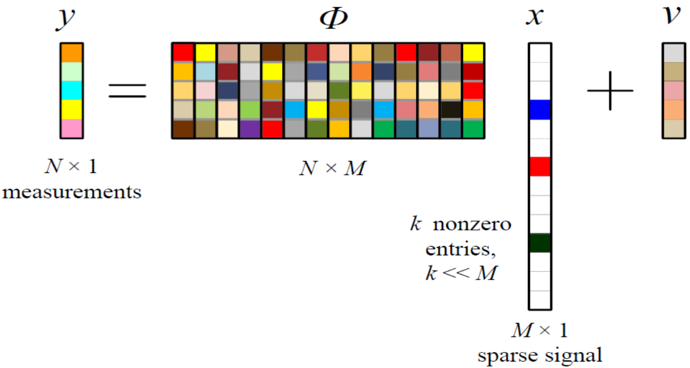
Data Covariance Matrix

$$\mathbf{R}_y = E(\mathbf{y}[n]\mathbf{y}^H[n]) = \mathbf{R}_s + \sigma_z^2\mathbf{I} = \mathbf{V}\mathbf{R}_x\mathbf{V}^H + \sigma_z^2\mathbf{I}$$

$\mathbf{R}_y \in C^{N \times N}$, $\mathbf{R}_s \in C^{N \times N}$, $\mathbf{R}_x \in C^{D \times D}$, and $\mathbf{V} \in C^{N \times D}$

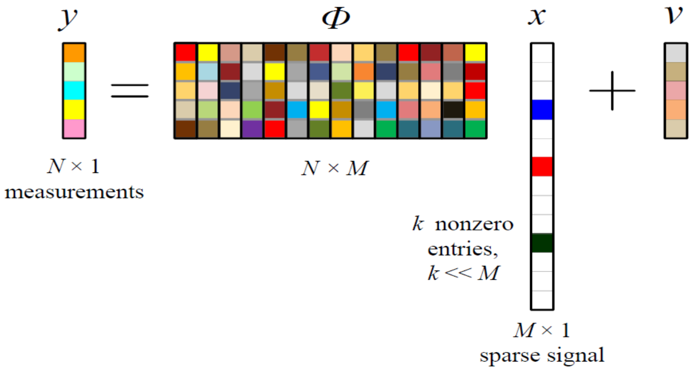
For ULA and uncorrelated sources, $\mathbf{R}_y$ is a Toeplitz matrix.

Sparse Signal Recovery (SSR)



- y is a $N \times 1$ measurement vector and x is $M \times 1$ desired vector which is sparse with k non zero entries.
- Φ is $N \times M$ dictionary matrix where $M >> N$.

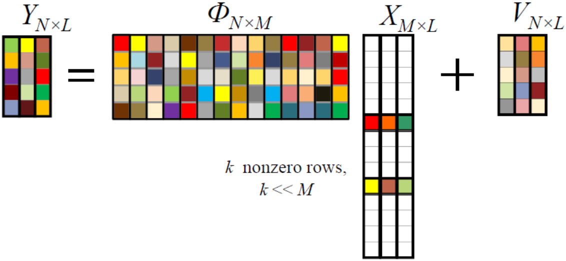
Sparse Signal Recovery (SSR)



- y is a $N \times 1$ measurement vector and x is $M \times 1$ desired vector which is sparse with k non zero entries.
- Φ is $N \times M$ dictionary matrix where $M >> N$. For array processing, Φ is obtained by employing a grid, i.e. lth column is $\mathbf{V}(\omega_l)$.

Multiple Measurement Vectors (MMV)

- Model



- Multiple measurements: L measurements
- Common Sparsity Profile: k nonzero rows

Beamforming(BF)

Beamforming(BF)

Pick a direction of interest ω_s, and select the linear combining weights W such that $r[n] = W^H \mathbf{y}[n]$ contains mostly the signal from direction ω_s

Beamforming(BF)

Pick a direction of interest ω_s, and select the linear combining weights W such that $r[n] = W^H \mathbf{y}[n]$ contains mostly the signal from direction ω_s

Conventional beamforming: $W = \mathbf{V}(\omega_s)$ and the power in direction ω_s is $E(|r[n]|^2) = \mathbf{V}(\omega_s)^H \mathbf{R}_y \mathbf{V}(\omega_s)$.

Beamforming(BF)

Pick a direction of interest ω_s, and select the linear combining weights W such that $r[n] = W^H \mathbf{y}[n]$ contains mostly the signal from direction ω_s

Conventional beamforming: $W = \mathbf{V}(\omega_s)$ and the power in direction ω_s is $E(|r[n]|^2) = \mathbf{V}(\omega_s)^H \mathbf{R}_y \mathbf{V}(\omega_s)$.

If you scan the spatial angles and look for the direction with most power, it has similarity to the search step of OMP.

Beamforming(BF)

Pick a direction of interest ω_s, and select the linear combining weights W such that $r[n] = W^H \mathbf{y}[n]$ contains mostly the signal from direction ω_s

Conventional beamforming: $W = \mathbf{V}(\omega_s)$ and the power in direction ω_s is $E(|r[n]|^2) = \mathbf{V}(\omega_s)^H \mathbf{R}_y \mathbf{V}(\omega_s)$.

If you scan the spatial angles and look for the direction with most power, it has similarity to the search step of OMP.

BF with Null constraints: Incorporate constraints, usually nulls in certain directions, in the BF design.

ULA and Beamforming in Pictures

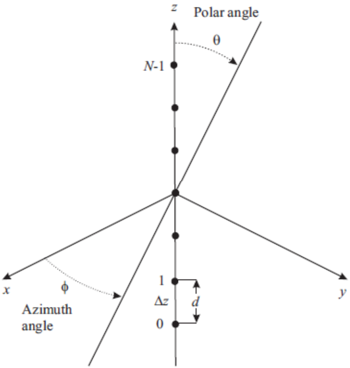


Figure: ULA on the z-axis

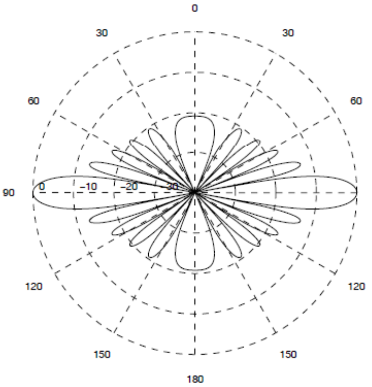


Figure: BF pattern(polar plot, $N = 11$)

Minimum power distortionless response (MPDR) beamformer

Minimum power distortionless response (MPDR) beamformer

Pick direction of interest ω_s

Minimum power distortionless response (MPDR) beamformer

Pick direction of interest ω_s

Distortionless constraint on beamformer W: $W^H \mathbf{V}_s = 1$, where $\mathbf{V}_s = \mathbf{V}(\omega_s)$.

Minimum power distortionless response (MPDR) beamformer

Pick direction of interest ω_s

Distortionless constraint on beamformer W: $W^H \mathbf{V}_s = 1$, where $\mathbf{V}_s = \mathbf{V}(\omega_s)$.

Minimum Power objective: Choose W to minimize $E(|W^H|\mathbf{y}[n]|^2)$

Minimum power distortionless response (MPDR) beamformer

Pick direction of interest ω_s

Distortionless constraint on beamformer W: $W^H \mathbf{V}_s = 1$, where $\mathbf{V}_s = \mathbf{V}(\omega_s)$.

Minimum Power objective: Choose W to minimize $E(|W^H|\mathbf{y}[n]|^2)$

MPDR BF design

$$\min_W W^H \mathbf{R}_y W \text{ subject to } W^H \mathbf{V}_s = 1.$$

Minimum power distortionless response (MPDR) beamformer

Pick direction of interest ω_s

Distortionless constraint on beamformer W: $W^H \mathbf{V}_s = 1$, where $\mathbf{V}_s = \mathbf{V}(\omega_s)$.

Minimum Power objective: Choose W to minimize $E(|W^H|\mathbf{y}[n]|^2)$

MPDR BF design

$$\min_W W^H \mathbf{R}_y W \text{ subject to } W^H \mathbf{V}_s = 1.$$

Solution: $W_{mpdr} = \frac{1}{\mathbf{V}_s^H \mathbf{R}_y^{-1} \mathbf{V}_s} \mathbf{R}_y^{-1} \mathbf{V}_s$

MPDR Spatial Power Spectrum

MPDR Spatial Power Spectrum

$$W_{mpdr} = \frac{1}{\mathbf{V}_s^H \mathbf{R}_y^{-1} \mathbf{V}_s} \mathbf{R}_y^{-1} \mathbf{V}_s$$

MPDR Spatial Power Spectrum

$$W_{mpdr} = \frac{1}{\mathbf{V}_s^H \mathbf{R}_y^{-1} \mathbf{V}_s} \mathbf{R}_y^{-1} \mathbf{V}_s$$

Benefit:

MPDR Spatial Power Spectrum

$W_{mpdr} = \frac{1}{\mathbf{V}_s^H \mathbf{R}_y^{-1} \mathbf{V}_s} \mathbf{R}_y^{-1} \mathbf{V}_s$

Benefit:

- $\mathbf{R}_y$ is easier to determine making it computationally attractive

$$\mathbf{R}_y \approx \frac{1}{L} \sum_{n=1}^{L-1} \mathbf{y}[n]\mathbf{y}^H[n]$$

MPDR Spatial Power Spectrum

$$W_{mpdr} = \frac{1}{\mathbf{V}_s^H \mathbf{R}_y^{-1} \mathbf{V}_s} \mathbf{R}_y^{-1} \mathbf{V}_s$$

Benefit:

- $\mathbf{R}_y$ is easier to determine making it computationally attractive

$$\mathbf{R}_y \approx \frac{1}{L} \sum_{n=1}^{L-1} \mathbf{y}[n] \mathbf{y}^H[n]$$

- Same $\mathbf{R}_y$ is needed if you change your mind on direction of interest. Can deal with multiple signals of interest with considerable ease.

MPDR Spatial Power Spectrum

$$W_{mpdr} = \frac{1}{\mathbf{V}_s^H \mathbf{R}_y^{-1} \mathbf{V}_s} \mathbf{R}_y^{-1} \mathbf{V}_s$$

Benefit:

- $\mathbf{R}_y$ is easier to determine making it computationally attractive

$$\mathbf{R}_y \approx \frac{1}{L} \sum_{n=1}^{L-1} \mathbf{y}[n]\mathbf{y}^H[n]$$

- Same $\mathbf{R}_y$ is needed if you change your mind on direction of interest. Can deal with multiple signals of interest with considerable ease.

Spatial Power Spectrum using MPDR:

$$P_{mpdr}(\omega_s) = E(|W_{mpdr}^H \mathbf{y}[n]|^2) = W_{mpdr}^H \mathbf{R}_y W_{mpdr} = \frac{1}{\mathbf{V}_s^H \mathbf{R}_y^{-1} \mathbf{V}_s}$$

MPDR intuition

MPDR intuition

$$\begin{aligned} W^H \mathbf{y}[n] &= W^H \mathbf{V}_s x_s[n] + W^H \mathbf{l}[n] \\ &= \underbrace{x_s[n]}_{\text{distortionless constraint}} + q[n], \quad \text{where} \quad q[n] = W^H \mathbf{l}[n] \end{aligned}$$

MPDR intuition

$$
\begin{aligned}
W^H\mathbf{y}[n] &= W^H\mathbf{V}_s x_s[n] + W^H\mathbf{I}[n] \\
&= \underbrace{x_s[n]}_{\text{distortionless constraint}} + q[n], \quad \text{where} \quad q[n] = W^H\mathbf{I}[n]
\end{aligned}
$$

Minimum Power objective: Choose W to minimize $E(|W^H|\mathbf{y}[n]|^2)$, the power at the output of the beamformer

MPDR intuition

$$\begin{aligned} W^H \mathbf{y}[n] &= W^H \mathbf{V}_s x_s[n] + W^H \mathbf{I}[n] \\ &= \underbrace{x_s[n]}_{\text{distortionless constraint}} + q[n], \quad \text{where} \quad q[n] = W^H \mathbf{I}[n] \end{aligned}$$

Minimum Power objective: Choose W to minimize $E(|W^H|\mathbf{y}[n]|^2)$, the power at the output of the beamformer

If the interference is uncorrelated with the desired signal, then minimization of $E(|q[n]|^2)$ is achieved, i.e. interference is minimized.

MPDR intuition

$$\begin{aligned} W^H\mathbf{y}[n] &= W^H\mathbf{V}_s x_s[n] + W^H\mathbf{I}[n] \\ &= \underbrace{x_s[n]}_{\text{distortionless constraint}} + q[n], \quad \text{where} \quad q[n] = W^H\mathbf{I}[n] \end{aligned}$$

Minimum Power objective: Choose W to minimize $E(|W^H|\mathbf{y}[n]|^2)$, the power at the output of the beamformer

If the interference is uncorrelated with the desired signal, then minimization of $E(|q[n]|^2)$ is achieved, i.e. interference is minimized.

Signal of interest has been isolated and interference minimized (SINR maximized)

MPDR intuition

$$\begin{aligned} W^H\mathbf{y}[n] &= W^H\mathbf{V}_s x_s[n] + W^H\mathbf{I}[n] \\ &= \underbrace{x_s[n]}_{\text{distortionless constraint}} + q[n], \quad \text{where} \quad q[n] = W^H\mathbf{I}[n] \end{aligned}$$

Minimum Power objective: Choose W to minimize $E(|W^H|\mathbf{y}[n]|^2)$, the power at the output of the beamformer

If the interference is uncorrelated with the desired signal, then minimization of $E(|q[n]|^2)$ is achieved, i.e. interference is minimized.

Signal of interest has been isolated and interference minimized (SINR maximized)

If the interference is correlated with the desired source, we have cancellation and the desired source is not preserved.

MPDR: Uncorrelated sources

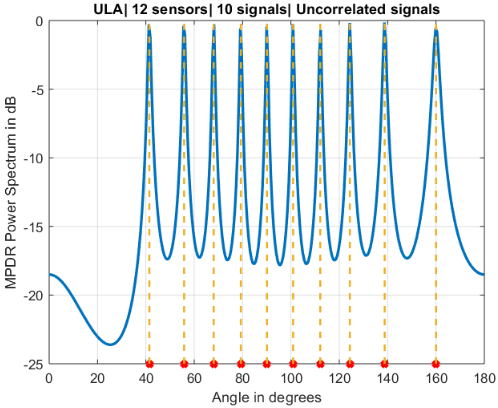


Figure: nsignals = 10 and nsensors = 12

MPDR: Beamformer for uncorrelated sources

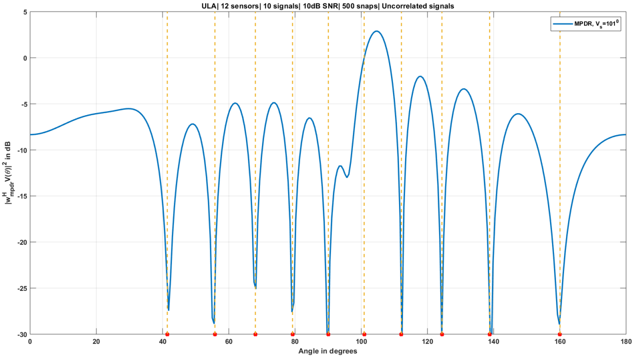


Figure: nsignals = 10 and nsensors = 12

MPDR: correlated sources

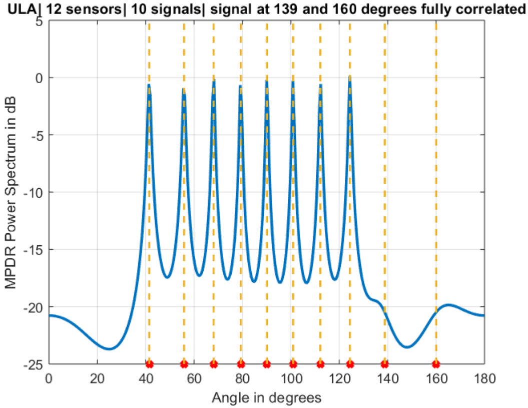


Figure: nsignals = 10, two correlated, and nsensors = 12

Sparse Bayesian Learning

$$\mathbf{y} = \Phi\mathbf{x} + v$$

Sparse Bayesian Learning

$$\mathbf{y} = \Phi\mathbf{x} + v$$

SBL uses a Bayesian framework with a separable prior $p(\mathbf{x}) = \Pi p(x_i)$.

Sparse Bayesian Learning

$$\mathbf{y} = \Phi\mathbf{x} + v$$

SBL uses a Bayesian framework with a separable prior $p(\mathbf{x}) = \Pi p(x_i)$.

Gaussian Scale Mixtures (GSM)

$$p(x_i) = \int p(x_i|\gamma_i)p(\gamma_i)d\gamma_i = \int N(x_i; 0, \gamma_i)p(\gamma_i)d\gamma_i$$

Sparse Bayesian Learning

$$\mathbf{y} = \Phi\mathbf{x} + v$$

SBL uses a Bayesian framework with a separable prior $p(\mathbf{x}) = \Pi p(x_i)$.

Gaussian Scale Mixtures (GSM)

$$p(x_i) = \int p(x_i|\gamma_i)p(\gamma_i)d\gamma_i = \int N(x_i; 0, \gamma_i)p(\gamma_i)d\gamma_i$$

Theorem

A density $p(x)$ which is symmetric with respect to the origin, can be represented by a GSM iff $p(\sqrt{x})$ is completely monotonic on $(0, \infty)$.

Sparse Bayesian Learning

$$\mathbf{y} = \Phi\mathbf{x} + v$$

SBL uses a Bayesian framework with a separable prior $p(\mathbf{x}) = \Pi p(x_i)$.

Gaussian Scale Mixtures (GSM)

$$p(x_i) = \int p(x_i|\gamma_i)p(\gamma_i)d\gamma_i = \int N(x_i; 0, \gamma_i)p(\gamma_i)d\gamma_i$$

Theorem

A density $p(x)$ which is symmetric with respect to the origin, can be represented by a GSM iff $p(\sqrt{x})$ is completely monotonic on $(0, \infty)$.

Most of the interesting priors over x can be represented in this GSM form. [Palmer et al., 2006]

Sparse Bayesian Learning

$$\mathbf{y} = \Phi \mathbf{x} + v$$

SBL uses a Bayesian framework with a separable prior $p(\mathbf{x}) = \Pi p(x_i)$.

Gaussian Scale Mixtures (GSM)

$$p(x_i) = \int p(x_i|\gamma_i)p(\gamma_i)d\gamma_i = \int N(x_i; 0, \gamma_i)p(\gamma_i)d\gamma_i$$

Theorem

A density $p(x)$ which is symmetric with respect to the origin, can be represented by a GSM iff $p(\sqrt{x})$ is completely monotonic on $(0, \infty)$.

Most of the interesting priors over x can be represented in this GSM form. [Palmer et al., 2006]

Definition

A function $f(x)$ is completely monotonic on (a, b) if $(-1)^n f^{(n)}(x) \geq 0, \; n = 0, 1, .., $ where $f^{(n)}(x)$ denotes the nth derivative

GSM Hierarchy Viewpoint

GSM Hierarchy Viewpoint

$$\underbrace{\gamma}_{p(\gamma)} \rightarrow X \sim N(x; 0, \gamma)$$

GSM Hierarchy Viewpoint

$$\underbrace{\gamma}_{p(\gamma)} \rightarrow X \sim N(x; 0, \gamma)$$

Alternatively

$$X = \sqrt{\gamma} G$$

where $G \sim N(g; 0, 1)$ and γ and G are independent.

GSM Hierarchy Viewpoint

$$\underbrace{\gamma}_{p(\gamma)} \rightarrow X \sim N(x; 0, \gamma)$$

Alternatively

$$X = \sqrt{\gamma}G$$

where $G \sim N(g; 0, 1)$ and γ and G are independent.

Kurtosis: Note

$$E(X^2) = E(\gamma)E(G^2) = E(\gamma), \text{ and } E(X^4) = E(\gamma^2)E(G^4) = 3E(\gamma^2)$$

GSM Hierarchy Viewpoint

$$\underbrace{\gamma}_{p(\gamma)} \to X \sim N(x; 0, \gamma)$$

Alternatively

$$X = \sqrt{\gamma} G$$

where $G \sim N(g; 0, 1)$ and γ and G are independent.

Kurtosis: Note

$$E(X^2) = E(\gamma)E(G^2) = E(\gamma), \quad \text{and} \quad E(X^4) = E(\gamma^2)E(G^4) = 3E(\gamma^2)$$

$$\begin{aligned}\mathrm{Kurt}(X) &= E(X^4) - 3(E(X^2))^2 = 3E(\gamma^2) - 3(E(\gamma))^2 \\ &= 3(E(\gamma^2) - (E(\gamma))^2) = 3\mathrm{Var}(\gamma) \geq 0\end{aligned}$$

Examples of Gaussian Scale Mixtures

Laplacian density

$$p(x;\beta) = \frac{1}{2\beta} exp(-\frac{|x|}{\beta})$$

Scale mixing density: $p(\gamma) = \frac{1}{2\beta^2}\exp(-\frac{1}{2\beta^2}\gamma), \gamma \geq 0.$

Examples of Gaussian Scale Mixtures

Laplacian density

$$p(x;\beta) = \frac{1}{2\beta} exp(-\frac{|x|}{\beta})$$

Scale mixing density: $p(\gamma) = \frac{1}{2\beta^2}\exp(-\frac{1}{2\beta^2}\gamma), \gamma \geq 0.$

Student-t Distribution

$$p(x;a,b) = \frac{b^a\Gamma(a+1/2)}{(2\pi)^{0.5}\Gamma(a)}\frac{1}{(b+x^2/2)^{a+1/2}}$$

Scale mixing density: Inverse-Gamma Distribution.

Examples of Gaussian Scale Mixtures

Laplacian density

$$p(x;\beta) = \frac{1}{2\beta} exp(-\frac{|x|}{\beta})$$

Scale mixing density: $p(\gamma) = \frac{1}{2\beta^2}\exp(-\frac{1}{2\beta^2}\gamma), \gamma \geq 0.$

Student-t Distribution

$$p(x;a,b) = \frac{b^a\Gamma(a+1/2)}{(2\pi)^{0.5}\Gamma(a)}\frac{1}{(b+x^2/2)^{a+1/2}}$$

Scale mixing density: Inverse-Gamma Distribution.

Generalized Gaussian

$$p(x;p) = \frac{1}{2\Gamma(1+\frac{1}{p})}e^{-|x|^p}$$

Scale mixing density: Positive alpha stable density of order $p/2$.

Two Options for Estimation with GSM priors

Two Options for Estimation with GSM priors

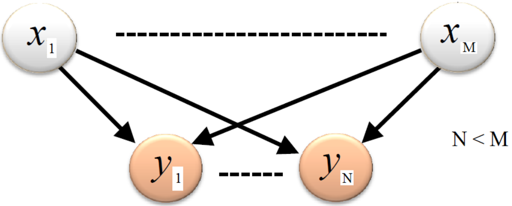
MAP Estimation (Type I)

Two Options for Estimation with GSM priors

MAP Estimation (Type I)

Evidence Maximization (Type II)

MAP Estimation Framework (Type I)

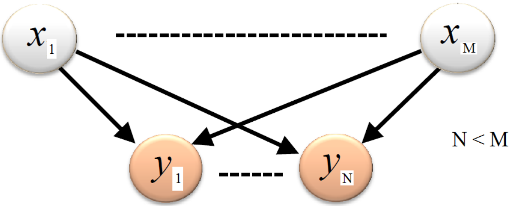


Problem Statement

$$\hat{x} = \arg\max_{x} p(x|y) = \arg\max_{x} p(y|x)p(x) = \arg\max[\log p(y|x) + \log p(x)]$$

[1]Giri, R., and Rao, B. D. (2016). Type I and Type II Bayesian Methods for Sparse Signal Recovery Using Scale Mixtures. IEEE Trans. Signal Processing, 64(13), 3418-3428.

MAP Estimation Framework (Type I)



Problem Statement

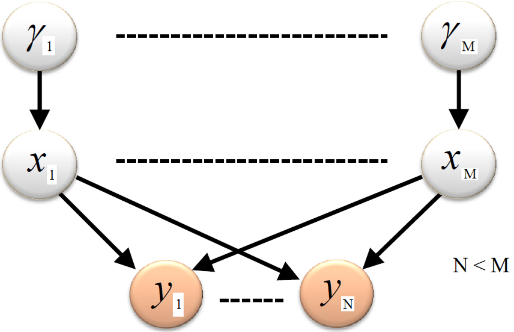
$$\hat{x} = \arg\max_x p(x|y) = \arg\max_x p(y|x)p(x) = \arg\max[\log p(y|x) + \log p(x)]$$

Can derive a general version that includes many past reweighted algorithms using a generalized-t distribution in one setting[1]

[1] Giri, R., and Rao, B. D. (2016). Type I and Type II Bayesian Methods for Sparse Signal Recovery Using Scale Mixtures. IEEE Trans. Signal Processing, 64(13), 3418-3428.

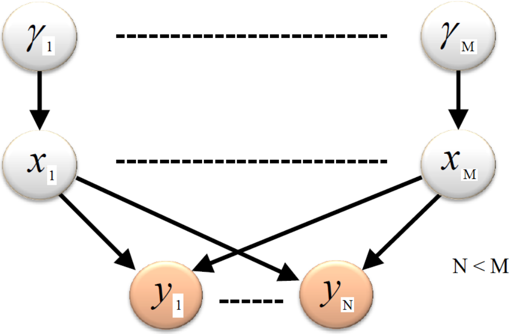
Evidence Maximization (Type II)

Evidence Maximization (Type II)



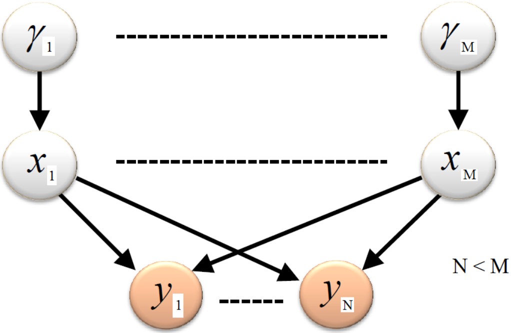
Find a MAP estimate of γ, i.e. $\hat{\gamma} = \arg\max p(\gamma|y)$.

Evidence Maximization (Type II)



Find a MAP estimate of γ, i.e. $\hat{\gamma} = \arg\max p(\gamma|y)$.
Estimate of the posterior distribution for x using estimated $\hat{\gamma}$; i.e. $p(x|y;\hat{\gamma})$.

Evidence Maximization (Type II)



Find a MAP estimate of γ, i.e. $\hat{\gamma} = \arg\max p(\gamma|y)$.
Estimate of the posterior distribution for x using estimated $\hat{\gamma}$; i.e.$p(x|y;\hat{\gamma})$.

This leads to Sparse Bayesian Learning (SBL).

Evidence Maximization Framework

Potential Advantages

Evidence Maximization Framework

Potential Advantages

- Averaging over x leads to fewer minima in $p(\gamma|y) = \int p(\gamma, x|y)dx$.

Evidence Maximization Framework

Potential Advantages

- Averaging over x leads to fewer minima in $p(\gamma|y) = \int p(\gamma, x|y)dx$.
- γ can tie several parameters, leading to fewer parameters.

Evidence Maximization Framework

Potential Advantages

- Averaging over x leads to fewer minima in $p(\gamma|y) = \int p(\gamma, x|y)dx$.
- γ can tie several parameters, leading to fewer parameters. For MMV, a single γ_i for row i.

Evidence Maximization Framework

Potential Advantages

- Averaging over x leads to fewer minima in $p(\gamma|y) = \int p(\gamma, x|y)dx$.
- γ can tie several parameters, leading to fewer parameters. For MMV, a single γ_i for row i.
- Maximizing the true posterior mass over the subspaces spanned by non zero indexes instead of looking for the mode.

Sparse Bayesian Learning

Sparse Bayesian Learning

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$$

Sparse Bayesian Learning

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$$

Solving for MAP estimate of γ

$$\hat{\gamma} = \arg\max_{\gamma} p(\gamma|y) = \arg\max_{\gamma} p(y|\gamma)p(\gamma)$$

Sparse Bayesian Learning

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$$

Solving for MAP estimate of γ

$$\hat{\gamma} = \arg\max_{\gamma} p(\gamma|y) = \arg\max_{\gamma} p(y|\gamma)p(\gamma)$$

What is $p(y|\gamma)$

Sparse Bayesian Learning

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$$

Solving for MAP estimate of γ

$$\hat{\gamma} = \arg\max_{\gamma} p(\gamma|y) = \arg\max_{\gamma} p(y|\gamma)p(\gamma)$$

What is $p(y|\gamma)$

Given γ, $\mathbf{x}$ is Gaussian with mean zero and Covariance matrix Γ with $\Gamma = \text{diag}(\gamma)$, i.e. $p(x|\gamma) = N(x; 0, \Gamma) = \Pi N(x_i; 0, \gamma_i)$.

Sparse Bayesian Learning

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$$

Solving for MAP estimate of γ

$$\hat{\gamma} = \arg\max_{\gamma} p(\gamma|y) = \arg\max_{\gamma} p(y|\gamma)p(\gamma)$$

What is $p(y|\gamma)$

Given γ, $\mathbf{x}$ is Gaussian with mean zero and Covariance matrix Γ with $\Gamma = \text{diag}(\gamma)$, i.e. $p(x|\gamma) = N(x; 0, \Gamma) = \Pi N(x_i; 0, \gamma_i)$.

Then $p(y|\gamma) = N(y; 0, \Sigma_y)$, where $\Sigma_y = \sigma^2 I + \mathbf{A}\Gamma\mathbf{A}^T$,

$$p(y|\gamma) = \frac{1}{\sqrt{(2\pi)^N |\Sigma_y|}} e^{-\frac{1}{2} y^T \Sigma_y^{-1} y}$$

Solving for the optimal γ

Solving for the optimal γ

$$\hat{\gamma} = \arg\max_{\gamma} p(\gamma|y) = \arg\max_{\gamma} p(y|\gamma)p(\gamma)$$
$$= \arg\min_{\gamma} \log|\Sigma_y| + y^T \Sigma_y^{-1} y - 2\sum_i \log p(\gamma_i)$$
$$= \arg\min_{\gamma} \log|\Sigma_y| + \text{Trace}\ \Sigma_y^{-1}\ yy^T - 2\sum_i \log p(\gamma_i)$$

where, $\Sigma_y = \sigma^2 I + \mathbf{A}\Gamma\mathbf{A}^T$ and $\Gamma = \text{diag}(\gamma)$

Solving for the optimal γ

$$\begin{aligned}\hat{\gamma} &= \arg\max_{\gamma} p(\gamma|y) = \arg\max_{\gamma} p(y|\gamma)p(\gamma) \\ &= \arg\min_{\gamma} \log|\Sigma_y| + y^T\Sigma_y^{-1}y - 2\sum_i \log p(\gamma_i) \\ &= \arg\min_{\gamma} \log|\Sigma_y| + \text{Trace}\left(\Sigma_y^{-1}\, yy^T\right) - 2\sum_i \log p(\gamma_i)\end{aligned}$$

where, $\Sigma_y = \sigma^2 I + \mathbf{A}\Gamma\mathbf{A}^T$ and $\Gamma = \text{diag}(\gamma)$

For some of the discussion, we will ignore the prior on γ, i.e. $p(\gamma)$. This can be viewed as a non-informative prior on γ or γ as deterministic but unknown.

Solving for the optimal γ

$$
\begin{aligned}
\hat{\gamma} &= \arg\max_{\gamma} p(\gamma|y) = \arg\max_{\gamma} p(y|\gamma)p(\gamma) \\
&= \arg\min_{\gamma} \log|\Sigma_y| + y^T\Sigma_y^{-1}y - 2\sum_i \log p(\gamma_i) \\
&= \arg\min_{\gamma} \log|\Sigma_y| + \text{Trace}\ \Sigma_y^{-1}\ yy^T - 2\sum_i \log p(\gamma_i)
\end{aligned}
$$

where, $\Sigma_y = \sigma^2 I + \mathbf{A}\Gamma\mathbf{A}^T$ and $\Gamma = \text{diag}(\gamma)$

For some of the discussion, we will ignore the prior on γ, i.e. $p(\gamma)$. This can be viewed as a non-informative prior on γ or γ as deterministic but unknown.

Given the separable nature of the prior, the prior is easy to incorporate.

Solving for the optimal γ

$$\hat{\gamma} = \arg\max_{\gamma} p(\gamma|y) = \arg\max_{\gamma} p(y|\gamma)p(\gamma)$$
$$= \arg\min_{\gamma} \log|\Sigma_y| + y^T\Sigma_y^{-1}y - 2\sum_i \log p(\gamma_i)$$
$$= \arg\min_{\gamma} \log|\Sigma_y| + \text{Trace } \Sigma_y^{-1}\ yy^T\ - 2\sum_i \log p(\gamma_i)$$

where, $\Sigma_y = \sigma^2 I + \mathbf{A}\Gamma\mathbf{A}^T$ and $\Gamma = \text{diag}(\gamma)$

For some of the discussion, we will ignore the prior on γ, i.e. $p(\gamma)$. This can be viewed as a non-informative prior on γ or γ as deterministic but unknown.

Given the separable nature of the prior, the prior is easy to incorporate.

For MMV

$$\hat{\gamma} = \arg\min_{\gamma} \log|\Sigma_y| + \text{Trace } \Sigma_y^{-1}\ \hat{\mathbf{R}}_y - \frac{2}{L}\sum_i \log p(\gamma_i)$$

where $\hat{\mathbf{R}}_y = \frac{1}{L}\sum_{n=1}^{L} \mathbf{y}[n]\mathbf{y}^T[n]$

Algorithmic Variants

- Fixed Point iteration based on setting the derivative of the objective function to zero (Tipping)
- Expectation-Maximization (EM) Algorithm
- Sequential search for the significant γ's (Tipping and Faul)
- Majorization-Minimization based approach (Wipf and Nagarajan)
- Reweighted ℓ_1 and ℓ_2 algorithms (Wipf and Nagarajan)
- Approximate Message Passing (AlShoukairi, Schniter and Rao)

Computing Posterior $p(x|y;\hat{\gamma})$

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$$

Computing Posterior $p(x|y;\hat{\gamma})$

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$$

Now because of our convenient GSM choice, posterior can be easily computed, i.e, $p(x|y;\hat{\gamma}) = N(\mu_x, \Sigma_x)$ where,

$$\mu_x = E[x|y;\hat{\gamma}] = \hat{\Gamma}\mathbf{A}^T\Sigma_y^{-1}\mathbf{y} = \hat{\Gamma}\mathbf{A}^T(\sigma^2 I + \mathbf{A}\hat{\Gamma}\mathbf{A}^T)^{-1}\mathbf{y}$$

$$\Sigma_{\tilde{x}} = Cov[x|y;\hat{\gamma}] = \hat{\Gamma} - \hat{\Gamma}\mathbf{A}^T\Sigma_y^{-1}\mathbf{A}\hat{\Gamma} = \hat{\Gamma} - \hat{\Gamma}\mathbf{A}^T(\sigma^2 I + \mathbf{A}\hat{\Gamma}\mathbf{A}^T)^{-1}\mathbf{A}\hat{\Gamma}$$

Computing Posterior $p(x|y;\hat{\gamma})$

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$$

Now because of our convenient GSM choice, posterior can be easily computed, i.e, $p(x|y;\hat{\gamma}) = N(\mu_x, \Sigma_x)$ where,

$$\mu_x = E[x|y;\hat{\gamma}] = \hat{\Gamma}\mathbf{A}^T\Sigma_y^{-1}\mathbf{y} = \hat{\Gamma}\mathbf{A}^T(\sigma^2 I + \mathbf{A}\hat{\Gamma}\mathbf{A}^T)^{-1}\mathbf{y}$$

$$\Sigma_{\tilde{x}} = Cov[x|y;\hat{\gamma}] = \hat{\Gamma} - \hat{\Gamma}\mathbf{A}^T\Sigma_y^{-1}\mathbf{A}\hat{\Gamma} = \hat{\Gamma} - \hat{\Gamma}\mathbf{A}^T(\sigma^2 I + \mathbf{A}\hat{\Gamma}\mathbf{A}^T)^{-1}\mathbf{A}\hat{\Gamma}$$

μ_x can be used as a point estimate.

Computing Posterior $p(x|y;\hat{\gamma})$

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$$

Now because of our convenient GSM choice, posterior can be easily computed, i.e, $p(x|y;\hat{\gamma}) = N(\mu_x, \Sigma_x)$ where,

$$\mu_x = E[x|y;\hat{\gamma}] = \hat{\Gamma}\mathbf{A}^T\Sigma_y^{-1}\mathbf{y} = \hat{\Gamma}\mathbf{A}^T(\sigma^2 I + \mathbf{A}\hat{\Gamma}\mathbf{A}^T)^{-1}\mathbf{y}$$

$$\Sigma_{\tilde{x}} = Cov[x|y;\hat{\gamma}] = \hat{\Gamma} - \hat{\Gamma}\mathbf{A}^T\Sigma_y^{-1}\mathbf{A}\hat{\Gamma} = \hat{\Gamma} - \hat{\Gamma}\mathbf{A}^T(\sigma^2 I + \mathbf{A}\hat{\Gamma}\mathbf{A}^T)^{-1}\mathbf{A}\hat{\Gamma}$$

μ_x can be used as a point estimate.
Sparsity is achieved by having many of the γ_i's have zero value.

Computing Posterior $p(x|y;\hat{\gamma})$

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$$

Now because of our convenient GSM choice, posterior can be easily computed, i.e, $p(x|y;\hat{\gamma}) = N(\mu_x, \Sigma_x)$ where,

$$\mu_x = E[x|y;\hat{\gamma}] = \hat{\Gamma}\mathbf{A}^T\Sigma_y^{-1}\mathbf{y} = \hat{\Gamma}\mathbf{A}^T(\sigma^2 I + \mathbf{A}\hat{\Gamma}\mathbf{A}^T)^{-1}\mathbf{y}$$

$$\Sigma_{\tilde{x}} = Cov[x|y;\hat{\gamma}] = \hat{\Gamma} - \hat{\Gamma}\mathbf{A}^T\Sigma_y^{-1}\mathbf{A}\hat{\Gamma} = \hat{\Gamma} - \hat{\Gamma}\mathbf{A}^T(\sigma^2 I + \mathbf{A}\hat{\Gamma}\mathbf{A}^T)^{-1}\mathbf{A}\hat{\Gamma}$$

μ_x can be used as a point estimate.
Sparsity is achieved by having many of the γ_i's have zero value.

Computing Posterior $p(x|y;\hat{\gamma})$

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$$

Now because of our convenient GSM choice, posterior can be easily computed, i.e, $p(x|y;\hat{\gamma}) = N(\mu_x, \Sigma_x)$ where,

$$\mu_x = E[x|y;\hat{\gamma}] = \hat{\Gamma}\mathbf{A}^T\Sigma_y^{-1}\mathbf{y} = \hat{\Gamma}\mathbf{A}^T(\sigma^2 I + \mathbf{A}\hat{\Gamma}\mathbf{A}^T)^{-1}\mathbf{y}$$

$$\Sigma_{\tilde{x}} = Cov[x|y;\hat{\gamma}] = \hat{\Gamma} - \hat{\Gamma}\mathbf{A}^T\Sigma_y^{-1}\mathbf{A}\hat{\Gamma} = \hat{\Gamma} - \hat{\Gamma}\mathbf{A}^T(\sigma^2 I + \mathbf{A}\hat{\Gamma}\mathbf{A}^T)^{-1}\mathbf{A}\hat{\Gamma}$$

μ_x can be used as a point estimate.
Sparsity is achieved by having many of the γ_i's have zero value.

The conditional mean and variance computation constitute a LMMSE BF!

EM algorithm: Updating γ

EM algorithm: Updating γ

Treating $(\mathbf{y}, \mathbf{x})$ as complete data and vector $\mathbf{x}$ as hidden variable.

$$\log p(y, x, \gamma) = \log p(y|x) + \log p(x|\gamma) + \log p(\gamma)$$

EM algorithm: Updating γ

Treating $(\mathbf{y}, \mathbf{x})$ as complete data and vector $\mathbf{x}$ as hidden variable.

$$\log p(y, x, \gamma) = \log p(y|x) + \log p(x|\gamma) + \log p(\gamma)$$

E step

$$Q(\gamma|\gamma^k) = \mathbb{E}_{x|y;\gamma^k}[\log p(y|x) + \log p(x|\gamma) + \log p(\gamma)]$$

EM algorithm: Updating γ

Treating $(\mathbf{y}, \mathbf{x})$ as complete data and vector $\mathbf{x}$ as hidden variable.

$$\log p(y, x, \gamma) = \log p(y|x) + \log p(x|\gamma) + \log p(\gamma)$$

E step

$$Q(\gamma|\gamma^k) = \mathbb{E}_{x|y;\gamma^k}[\log p(y|x) + \log p(x|\gamma) + \log p(\gamma)]$$

M step

$$\begin{aligned}
\gamma^{k+1} &= \text{argmax}_\gamma Q(\gamma|\gamma^k) = \text{argmax}_\gamma \mathbb{E}_{x|y;\gamma^k}[\log p(x|\gamma) + \log p(\gamma)] \\
&= \text{argmin}_\gamma \mathbb{E}_{x|y;\gamma^k}[\sum_{i=1}^{M} \left(\frac{x_i^2}{2\gamma_i} + \frac{1}{2}\log \gamma_i \right) - \log p(\gamma)]
\end{aligned}$$

The optimization involves M scalar optimization problems of the form

$$J(\gamma_l) = \frac{E(x_l^2|\mathbf{y}, \gamma^k)}{2\gamma_l} + \frac{1}{2}\log \gamma_l - \log p(\gamma_l)$$

Scalar Optimization

$$J(\gamma_l) = \frac{E(x_l^2|\mathbf{y}, \gamma^k)}{2\gamma_l} + \frac{1}{2}\log\gamma_l - \log p(\gamma_l)$$

Note that $E(x_l^2|\mathbf{y}, \gamma^k) = \mu_x^{(k)}(l)^2 + \Sigma_{\tilde{x}}^{(k)}(l, l)$

Non-informative prior results in an objective function

$$J(\gamma_l) = \frac{E(x_l^2|\mathbf{y}, \gamma^k)}{2\gamma_l} + \frac{1}{2}\log\gamma_l$$

Update equation

$$\gamma_l^{k+1} = E(x_l^2|\mathbf{y}, \gamma^k) = \mu_x^{(k)}(l)^2 + \Sigma_{\tilde{x}}^{(k)}(l, l)$$

Scalar Optimization

$$J(\gamma_l) = \frac{E(x_l^2|\mathbf{y}, \gamma^k)}{2\gamma_l} + \frac{1}{2}\log\gamma_l - \log p(\gamma_l)$$

Note that $E(x_l^2|\mathbf{y}, \gamma^k) = \mu_x^{(k)}(l)^2 + \Sigma_{\tilde{x}}^{(k)}(l, l)$

Non-informative prior results in an objective function

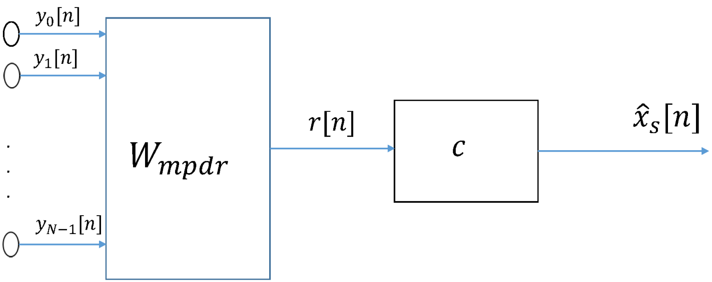
$$J(\gamma_l) = \frac{E(x_l^2|\mathbf{y}, \gamma^k)}{2\gamma_l} + \frac{1}{2}\log\gamma_l$$

Update equation

$$\gamma_l^{k+1} = E(x_l^2|\mathbf{y}, \gamma^k) = \mu_x^{(k)}(l)^2 + \Sigma_{\tilde{x}}^{(k)}(l, l)$$

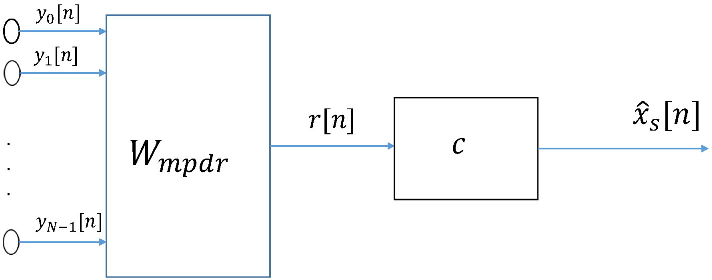
Source powers updated in an iterated manner

Connection between SBL and MPDR



LMMSE Beamformer viewed as a MPDR beamformer followed by scaling for the uncorrelated source case

Connection between SBL and MPDR

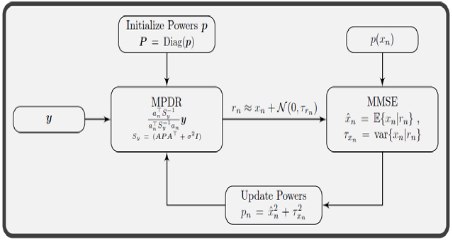


LMMSE Beamformer viewed as a MPDR beamformer followed by scaling for the uncorrelated source case

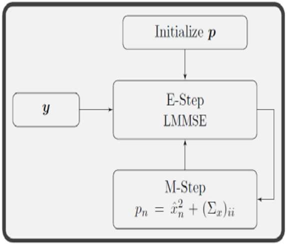
Note: $r[n] = x_s[n] + q[n]$

MPDR view of SBL

MPDR view of SBL

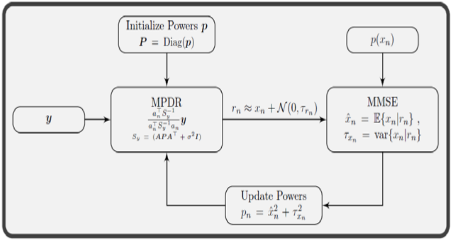


(a) MPDR Based Algorithm

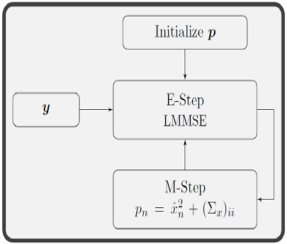


(b) EM-SBL Algorithm

MPDR view of SBL



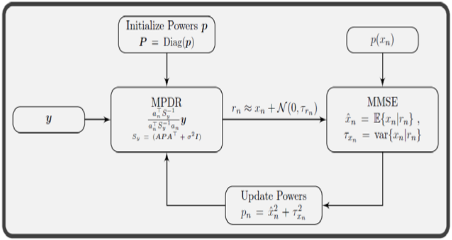
(a) MPDR Based Algorithm



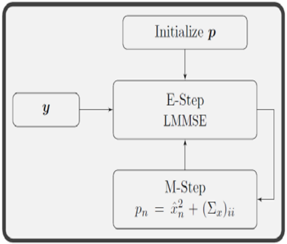
(b) EM-SBL Algorithm

- When $p(x_n) = \mathcal{N}(0, p_n)$ the MPDR + MMSE steps are equivalent to the LMMSE step in the EM SBL and the two algorithm are equivalent.

MPDR view of SBL



(a) MPDR Based Algorithm



(b) EM-SBL Algorithm

- When $p(x_n) = \mathcal{N}(0, p_n)$ the MPDR + MMSE steps are equivalent to the LMMSE step in the EM SBL and the two algorithm are equivalent.
- More general priors can be used within the MPDR framework. The EM-SBL has a closed form solution for the GSM prior only.

MPDR versus SBL

MPDR versus SBL

MPDR:

MPDR versus SBL

MPDR:

1. $\mathbf{y}[n] \to \hat{\mathbf{R}}_y = \frac{1}{L}\sum_{n=1}^{L} \mathbf{y}[n]\mathbf{y}^H[n]$

MPDR versus SBL

MPDR:

1. $\mathbf{y}[n] \rightarrow \hat{\mathbf{R}}_y = \frac{1}{L}\sum_{n=1}^{L}\mathbf{y}[n]\mathbf{y}^H[n]$
2. MPDR BF using $\hat{\mathbf{R}}_y$, i.e. $W_{mpdr} = \frac{1}{\mathbf{V}_s^H\hat{\mathbf{R}}_y^{-1}\mathbf{V}_s}\hat{\mathbf{R}}_y^{-1}\mathbf{V}_s$

MPDR versus SBL

MPDR:

1. $\mathbf{y}[n] \rightarrow \hat{\mathbf{R}}_y = \frac{1}{L}\sum_{n=1}^{L} \mathbf{y}[n]\mathbf{y}^H[n]$
2. MPDR BF using $\hat{\mathbf{R}}_y$, i.e. $W_{mpdr} = \frac{1}{\mathbf{V}_s^H \hat{\mathbf{R}}_y^{-1} \mathbf{V}_s} \hat{\mathbf{R}}_y^{-1} \mathbf{V}_s$
3. MPDR spatial power spectrum $P_{mpdr}(\omega_s) = \frac{1}{\mathbf{V}_s^H \hat{\mathbf{R}}_y^{-1} \mathbf{V}_s}$

MPDR versus SBL

MPDR:

1. $\mathbf{y}[n] \rightarrow \hat{\mathbf{R}}_y = \frac{1}{L}\sum_{n=1}^{L} \mathbf{y}[n]\mathbf{y}^H[n]$
2. MPDR BF using $\hat{\mathbf{R}}_y$, i.e. $W_{mpdr} = \frac{1}{\mathbf{V}_s^H \hat{\mathbf{R}}_y^{-1} \mathbf{V}_s} \hat{\mathbf{R}}_y^{-1} \mathbf{V}_s$
3. MPDR spatial power spectrum $P_{mpdr}(\omega_s) = \frac{1}{\mathbf{V}_s^H \hat{\mathbf{R}}_y^{-1} \mathbf{V}_s}$

SBL

MPDR versus SBL

MPDR:

1. $\mathbf{y}[n] \rightarrow \hat{\mathbf{R}}_y = \frac{1}{L}\sum_{n=1}^{L} \mathbf{y}[n]\mathbf{y}^H[n]$
2. MPDR BF using $\hat{\mathbf{R}}_y$, i.e. $W_{mpdr} = \frac{1}{\mathbf{V}_s^H \hat{\mathbf{R}}_y^{-1} \mathbf{V}_s} \hat{\mathbf{R}}_y^{-1} \mathbf{V}_s$
3. MPDR spatial power spectrum $P_{mpdr}(\omega_s) = \frac{1}{\mathbf{V}_s^H \hat{\mathbf{R}}_y^{-1} \mathbf{V}_s}$

SBL

1. Guess power of sources γ and employ uncorrelated model for correlation matrix $\Sigma_y = \Phi \mathbf{\Gamma} \Phi^T + \lambda \mathbf{I}$, where $\mathbf{\Gamma} = \mathrm{diag}(\gamma)$.

MPDR versus SBL

MPDR:

1. $\mathbf{y}[n] \to \hat{\mathbf{R}}_y = \frac{1}{L}\sum_{n=1}^{L} \mathbf{y}[n]\mathbf{y}^H[n]$
2. MPDR BF using $\hat{\mathbf{R}}_y$, i.e. $W_{mpdr} = \frac{1}{\mathbf{V}_s^H \hat{\mathbf{R}}_y^{-1}\mathbf{V}_s}\hat{\mathbf{R}}_y^{-1}\mathbf{V}_s$
3. MPDR spatial power spectrum $P_{mpdr}(\omega_s) = \frac{1}{\mathbf{V}_s^H \hat{\mathbf{R}}_y^{-1}\mathbf{V}_s}$

SBL

1. Guess power of sources γ and employ uncorrelated model for correlation matrix $\Sigma_y = \Phi\mathbf{\Gamma}\Phi^T + \lambda\mathbf{I}$, where $\mathbf{\Gamma} = \text{diag}(\gamma)$.
2. MPDR BF using Σ_y, i.e. $W_{mpdr} = \frac{1}{\mathbf{V}_s^H \Sigma_y^{-1}\mathbf{V}_s}\Sigma_y^{-1}\mathbf{V}_s$

MPDR versus SBL

MPDR:

1. $\mathbf{y}[n] \rightarrow \hat{\mathbf{R}}_y = \frac{1}{L}\sum_{n=1}^{L} \mathbf{y}[n]\mathbf{y}^H[n]$
2. MPDR BF using $\hat{\mathbf{R}}_y$, i.e. $W_{mpdr} = \frac{1}{\mathbf{V}_s^H \hat{\mathbf{R}}_y^{-1} \mathbf{V}_s} \hat{\mathbf{R}}_y^{-1} \mathbf{V}_s$
3. MPDR spatial power spectrum $P_{mpdr}(\omega_s) = \frac{1}{\mathbf{V}_s^H \hat{\mathbf{R}}_y^{-1} \mathbf{V}_s}$

SBL

1. Guess power of sources γ and employ uncorrelated model for correlation matrix $\Sigma_y = \Phi \mathbf{\Gamma} \Phi^T + \lambda \mathbf{I}$, where $\mathbf{\Gamma} = \mathrm{diag}(\gamma)$.
2. MPDR BF using Σ_y, i.e. $W_{mpdr} = \frac{1}{\mathbf{V}_s^H \Sigma_y^{-1} \mathbf{V}_s} \Sigma_y^{-1} \mathbf{V}_s$
3. Apply BF to data $\mathbf{y}[n]$, use BF output to compute new source power estimate, and update Σ_y

MPDR versus SBL

MPDR:

1. $\mathbf{y}[n] \to \hat{\mathbf{R}}_y = \frac{1}{L}\sum_{n=1}^{L}\mathbf{y}[n]\mathbf{y}^H[n]$
2. MPDR BF using $\hat{\mathbf{R}}_y$, i.e. $W_{mpdr} = \frac{1}{\mathbf{V}_s^H\hat{\mathbf{R}}_y^{-1}\mathbf{V}_s}\hat{\mathbf{R}}_y^{-1}\mathbf{V}_s$
3. MPDR spatial power spectrum $P_{mpdr}(\omega_s) = \frac{1}{\mathbf{V}_s^H\hat{\mathbf{R}}_y^{-1}\mathbf{V}_s}$

SBL

1. Guess power of sources γ and employ uncorrelated model for correlation matrix $\Sigma_y = \Phi\mathbf{\Gamma}\Phi^T + \lambda\mathbf{I}$, where $\mathbf{\Gamma} = \text{diag}(\gamma)$.
2. MPDR BF using Σ_y, i.e. $W_{mpdr} = \frac{1}{\mathbf{V}_s^H\Sigma_y^{-1}\mathbf{V}_s}\Sigma_y^{-1}\mathbf{V}_s$
3. Apply BF to data $\mathbf{y}[n]$, use BF output to compute new source power estimate, and update Σ_y
4. Repeat steps 2 and 3 till convergence

Toeplitz Matrix Approximation: ULA

Toeplitz Matrix Approximation: ULA

$\mathbf{R}_y = \mathbf{R}_s + \sigma^2 \mathbf{I}$ and $\mathbf{R}_s$ is low rank and Toeplitz for uncorrelated sources

Toeplitz Matrix Approximation: ULA

$\mathbf{R}_y = \mathbf{R}_s + \sigma^2\mathbf{I}$ and $\mathbf{R}_s$ is low rank and Toeplitz for uncorrelated sources

Theorem: A low rank Toeplitz $\mathbf{R}$, rank D, is uniquely represented as $\mathbf{R} = \sum_{l=1}^{D} \lambda_l \mathbf{V}(\omega_l)\mathbf{V}^H(\omega_l)$

Toeplitz Matrix Approximation: ULA

$\mathbf{R}_y = \mathbf{R}_s + \sigma^2\mathbf{I}$ and $\mathbf{R}_s$ is low rank and Toeplitz for uncorrelated sources

Theorem: A low rank Toeplitz $\mathbf{R}$, rank D, is uniquely represented as $\mathbf{R} = \sum_{l=1}^{D} \lambda_l \mathbf{V}(\omega_l)\mathbf{V}^H(\omega_l)$

$\Sigma_y = \sum_{l=1}^{M} \gamma_l \mathbf{V}(\omega_l)\mathbf{V}^H(\omega_l) + \lambda\mathbf{I}$ in SBL has the appropriate structure.

Toeplitz Matrix Approximation: ULA

$\mathbf{R}_y = \mathbf{R}_s + \sigma^2 \mathbf{I}$ and $\mathbf{R}_s$ is low rank and Toeplitz for uncorrelated sources

Theorem: A low rank Toeplitz $\mathbf{R}$, rank D, is uniquely represented as $\mathbf{R} = \sum_{l=1}^{D} \lambda_l \mathbf{V}(\omega_l)\mathbf{V}^H(\omega_l)$

$\Sigma_y = \sum_{l=1}^{M} \gamma_l \mathbf{V}(\omega_l)\mathbf{V}^H(\omega_l) + \lambda \mathbf{I}$ in SBL has the appropriate structure.

ML estimate: As $L \to \infty$, $KL(p^*||p)$ is minimized, where p is the model assumed by SBL, i.e. $p(y; \Sigma_y)$, and p^* is the actual data density, i.e. $p(y; \mathbf{R}_y)$.

Toeplitz Matrix Approximation: ULA

$\mathbf{R}_y = \mathbf{R}_s + \sigma^2\mathbf{I}$ and $\mathbf{R}_s$ is low rank and Toeplitz for uncorrelated sources

Theorem: A low rank Toeplitz $\mathbf{R}$, rank D, is uniquely represented as $\mathbf{R} = \sum_{l=1}^{D} \lambda_l \mathbf{V}(\omega_l)\mathbf{V}^H(\omega_l)$

$\Sigma_y = \sum_{l=1}^{M} \gamma_l \mathbf{V}(\omega_l)\mathbf{V}^H(\omega_l) + \lambda\mathbf{I}$ in SBL has the appropriate structure.

ML estimate: As $L \to \infty$, $KL(p^*||p)$ is minimized, where p is the model assumed by SBL, i.e. $p(y;\Sigma_y)$, and p^* is the actual data density, i.e. $p(y;\mathbf{R}_y)$.

Uncorrelated sources: then true model and SBL model is the same and SBL is finding a ML estimate of a Toeplitz matrix.

Toeplitz Matrix Approximation: ULA

$\mathbf{R}_y = \mathbf{R}_s + \sigma^2\mathbf{I}$ and $\mathbf{R}_s$ is low rank and Toeplitz for uncorrelated sources

Theorem: A low rank Toeplitz $\mathbf{R}$, rank D, is uniquely represented as $\mathbf{R} = \sum_{l=1}^{D} \lambda_l \mathbf{V}(\omega_l)\mathbf{V}^H(\omega_l)$

$\Sigma_y = \sum_{l=1}^{M} \gamma_l \mathbf{V}(\omega_l)\mathbf{V}^H(\omega_l) + \lambda\mathbf{I}$ in SBL has the appropriate structure.

ML estimate: As $L \to \infty$, $KL(p^*||p)$ is minimized, where p is the model assumed by SBL, i.e. $p(y; \Sigma_y)$, and p^* is the actual data density, i.e. $p(y; \mathbf{R}_y)$.

Uncorrelated sources: then true model and SBL model is the same and SBL is finding a ML estimate of a Toeplitz matrix.

Correlated sources: True model $\mathbf{R}_y$ is no longer Toeplitz and SBL is finding the best Toeplitz matrix in the KL sense.

Toeplitz Matrix Approximation: ULA

$\mathbf{R}_y = \mathbf{R}_s + \sigma^2 \mathbf{I}$ and $\mathbf{R}_s$ is low rank and Toeplitz for uncorrelated sources

Theorem: A low rank Toeplitz $\mathbf{R}$, rank D, is uniquely represented as $\mathbf{R} = \sum_{l=1}^{D} \lambda_l \mathbf{V}(\omega_l)\mathbf{V}^H(\omega_l)$

$\Sigma_y = \sum_{l=1}^{M} \gamma_l \mathbf{V}(\omega_l)\mathbf{V}^H(\omega_l) + \lambda \mathbf{I}$ in SBL has the appropriate structure.

ML estimate: As $L \to \infty$, $KL(p^* || p)$ is minimized, where p is the model assumed by SBL, i.e. $p(y; \Sigma_y)$, and p^* is the actual data density, i.e. $p(y; \mathbf{R}_y)$.

Uncorrelated sources: then true model and SBL model is the same and SBL is finding a ML estimate of a Toeplitz matrix.

Correlated sources: True model $\mathbf{R}_y$ is no longer Toeplitz and SBL is finding the best Toeplitz matrix in the KL sense.

The uncorrelated source model allows SBL to be largely unaffected by correlated sources.

MPDR: Uncorrelated sources

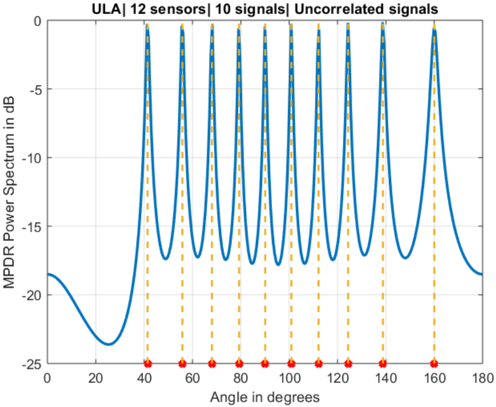


Figure: nsignals = 10 and nsensors = 12

MPDR: correlated sources

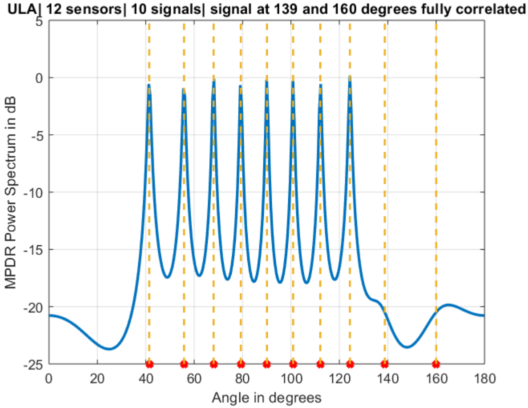


Figure: nsignals = 10, two correlated, and nsensors = 12

Correlated signals with MUSIC

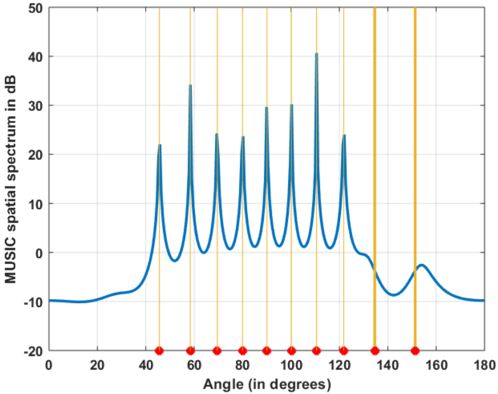


Figure: MUSIC: nsignals = 10 (Last two correlated), nsensors = 12

SBL to the rescue!

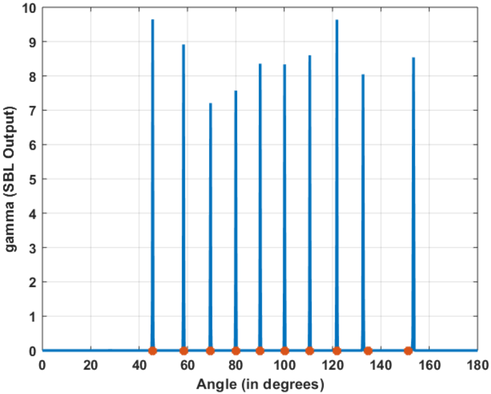


Figure: SBL: nsignals = 10 (Last two correlated), nsensors = 12

Beam Space processing and Nested Arrays

Beam Space processing and Nested Arrays

Consider $\mathbf{y}_{na}[n] = \mathbf{S}\mathbf{y}[n]$, where $\mathbf{S}$ is user chosen and $\mathbf{S}^{P \times N}$, with $P \leq N$.

Beam Space processing and Nested Arrays

Consider $\mathbf{y}_{na}[n] = \mathbf{Sy}[n]$, where $\mathbf{S}$ is user chosen and $\mathbf{S}^{P \times N}$, with $P \leq N$.

Options

Beam Space processing and Nested Arrays

Consider $\mathbf{y}_{na}[n] = \mathbf{S}\mathbf{y}[n]$, where $\mathbf{S}$ is user chosen and $\mathbf{S}^{P \times N}$, with $P \leq N$.

Options

- Could be a random matrix as in CS

Beam Space processing and Nested Arrays

Consider $\mathbf{y}_{na}[n] = \mathbf{S}\mathbf{y}[n]$, where $\mathbf{S}$ is user chosen and $\mathbf{S}^{P \times N}$, with $P \leq N$.

Options

- Could be a random matrix as in CS
- Could be chosen to cover a region in space

Beam Space processing and Nested Arrays

Consider $\mathbf{y}_{na}[n] = \mathbf{S}\mathbf{y}[n]$, where $\mathbf{S}$ is user chosen and $\mathbf{S}^{P \times N}$, with $P \leq N$.

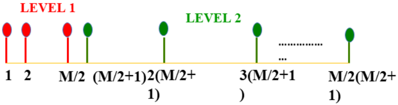
Options

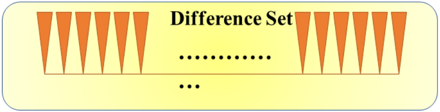
- Could be a random matrix as in CS
- Could be chosen to cover a region in space
- Could be used to thin the array (low complexity sensor array)

Inspiration from Sister Field: Array Signal Processing

Inspiration from Sister Field: Array Signal Processing

Nested Arrays: Structure & Properties

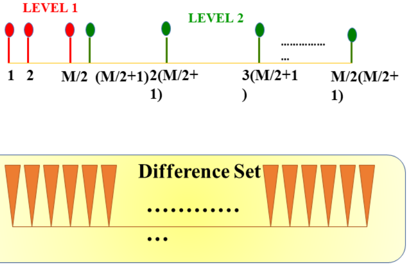




$M^2/2+M-1$ elements

Inspiration from Sister Field: Array Signal Processing

Nested Arrays: Structure & Properties

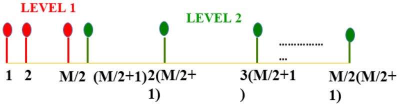


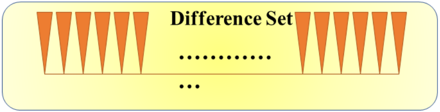
$M^2/2+M-1$ elements

- A nested array with N antennas has a filled difference set with $O(N^2)$ elements

Inspiration from Sister Field: Array Signal Processing

Nested Arrays: Structure & Properties



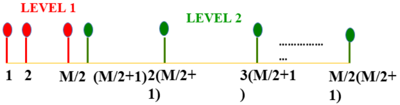


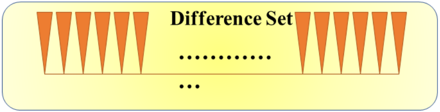
$M^2/2+M-1$ elements

- A nested array with N antennas has a filled difference set with $O(N^2)$ elements
- Drastically increases the spatial degrees of freedom

Inspiration from Sister Field: Array Signal Processing

Nested Arrays: Structure & Properties





$M^2/2+M-1$ elements

- A nested array with N antennas has a filled difference set with $O(N^2)$ elements
- Drastically increases the spatial degrees of freedom
- For source localization, this results in estimating location of more sources than sensors!

DoA Estimation with Nested Arrays[2]

With Nested Arrays, we can transform an order-N covariance matrix into order-$\frac{N^2+2N}{4}$ Toeplitz matrix

Transformation:

[2] Pal, P., and Vaidyanathan, P. P. (2010). Nested arrays: A novel approach to array processing with enhanced degrees of freedom. IEEE Transactions on Signal Processing, 58(8), 4167-4181.

DoA Estimation with Nested Arrays[2]

With Nested Arrays, we can transform an order-N covariance matrix into order-$\frac{N^2+2N}{4}$ Toeplitz matrix

Transformation:

Consider as an example $N = 4$, nested array sensor positions: $\{1, 2, 3, 6\}$

$$\begin{bmatrix} R_0 & R_{-1} & R_{-2} & R_{-5} \\ R_1 & R_0 & R_{-1} & R_{-4} \\ R_2 & R_1 & R_0 & R_{-3} \\ R_5 & R_4 & R_3 & R_0 \end{bmatrix} \rightarrow \begin{bmatrix} R_0 & R_{-1} & R_{-2} & R_{-3} & R_{-4} & R_{-5} \\ R_1 & R_0 & R_{-1} & R_{-2} & R_{-3} & R_{-4} \\ R_2 & R_1 & R_0 & R_{-1} & R_{-2} & R_{-3} \\ R_3 & R_2 & R_1 & R_0 & R_{-1} & R_{-2} \\ R_4 & R_3 & R_2 & R_1 & R_0 & R_{-1} \\ R_5 & R_4 & R_3 & R_2 & R_1 & R_0 \end{bmatrix}$$

[2] Pal, P., and Vaidyanathan, P. P. (2010). Nested arrays: A novel approach to array processing with enhanced degrees of freedom. IEEE Transactions on Signal Processing, 58(8), 4167-4181.

DoA Estimation with Nested Arrays[2]

With Nested Arrays, we can transform an order-N covariance matrix into order-$\frac{N^2+2N}{4}$ Toeplitz matrix

Transformation:

Consider as an example $N = 4$, nested array sensor positions: $\{1, 2, 3, 6\}$

$$\begin{bmatrix} R_0 & R_{-1} & R_{-2} & R_{-5} \\ R_1 & R_0 & R_{-1} & R_{-4} \\ R_2 & R_1 & R_0 & R_{-3} \\ R_5 & R_4 & R_3 & R_0 \end{bmatrix} \rightarrow \begin{bmatrix} R_0 & R_{-1} & R_{-2} & R_{-3} & R_{-4} & R_{-5} \\ R_1 & R_0 & R_{-1} & R_{-2} & R_{-3} & R_{-4} \\ R_2 & R_1 & R_0 & R_{-1} & R_{-2} & R_{-3} \\ R_3 & R_2 & R_1 & R_0 & R_{-1} & R_{-2} \\ R_4 & R_3 & R_2 & R_1 & R_0 & R_{-1} \\ R_5 & R_4 & R_3 & R_2 & R_1 & R_0 \end{bmatrix}$$

Number of DoAs that can be estimated is $5 > 3$ (for ULA)

[2] Pal, P., and Vaidyanathan, P. P. (2010). Nested arrays: A novel approach to array processing with enhanced degrees of freedom. IEEE Transactions on Signal Processing, 58(8), 4167-4181.

SBL and Nested Arrays

[3] Nannuru, S., and Gerstoft, P. (2019, May). 2D Beamforming on Sparse Arrays with Sparse Bayesian Learning. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 4355-4359).

SBL and Nested Arrays

Simply apply SBL[3] to $\mathbf{y}_{na}[n]$, where $\mathbf{y}_{na}[n] = \mathbf{S}\mathbf{y}[n]$

[3]Nannuru, S., and Gerstoft, P. (2019, May). 2D Beamforming on Sparse Arrays with Sparse Bayesian Learning. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 4355-4359).

SBL and Nested Arrays

Simply apply SBL[3] to $\mathbf{y}_{na}[n]$, where $\mathbf{y}_{na}[n] = \mathbf{S}\mathbf{y}[n]$

Actual covariance $\mathbf{S}\mathbf{R}_y\mathbf{S}^H$, and SBL covariance model is $\mathbf{S}\Sigma_y\mathbf{S}^H$, where Σ_y is Toeplitz

---

[3] Nannuru, S., and Gerstoft, P. (2019, May). 2D Beamforming on Sparse Arrays with Sparse Bayesian Learning. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 4355-4359).

SBL and Nested Arrays

Simply apply SBL[3] to $\mathbf{y}_{na}[n]$, where $\mathbf{y}_{na}[n] = \mathbf{S}\mathbf{y}[n]$

Actual covariance $\mathbf{S}\mathbf{R}_y\mathbf{S}^H$, and SBL covariance model is $\mathbf{S}\Sigma_y\mathbf{S}^H$, where Σ_y is Toeplitz

The ML estimation minimizes the KL distance between the actual density and model density assumed by SBL and so tries to best match the two covariance matrices.

[3] Nannuru, S., and Gerstoft, P. (2019, May). 2D Beamforming on Sparse Arrays with Sparse Bayesian Learning. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 4355-4359).

SBL and Nested Arrays

Simply apply SBL[3] to $\mathbf{y}_{na}[n]$, where $\mathbf{y}_{na}[n] = \mathbf{S}\mathbf{y}[n]$

Actual covariance $\mathbf{S}\mathbf{R}_y\mathbf{S}^H$, and SBL covariance model is $\mathbf{S}\Sigma_y\mathbf{S}^H$, where Σ_y is Toeplitz

The ML estimation minimizes the KL distance between the actual density and model density assumed by SBL and so tries to best match the two covariance matrices.

Figure: Nested array with 12 sensors

[3]Nannuru, S., and Gerstoft, P. (2019, May). 2D Beamforming on Sparse Arrays with Sparse Bayesian Learning. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 4355-4359).

Nested Arrays: SBL robustness

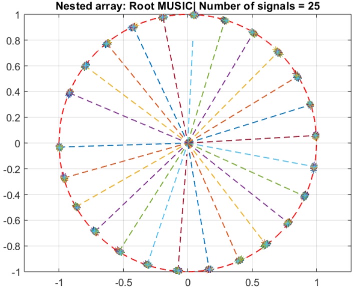


Figure: MUSIC

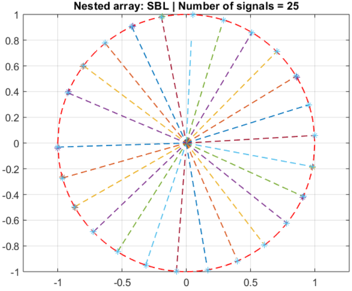


Figure: SBL

Nested Arrays: SBL efficiency

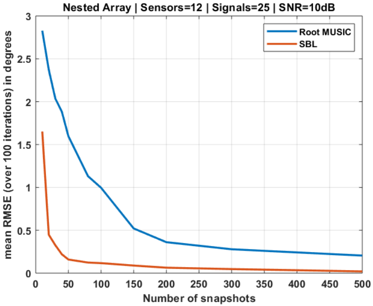


Figure: nsens $= 12$, nsignals $= 25$, SNR $= 10$dB, nsnapshots $= 500$

Summary

Summary

- Established a connection between SBL and MPDR beamforming.
  - Provides better insight into effective BF
  - Enables an approach to deal with more intractable inference problems

Summary

- Established a connection between SBL and MPDR beamforming.
  - Provides better insight into effective BF
  - Enables an approach to deal with more intractable inference problems
- Discussed Uniform Linear Arrays and Toeplitz Matrix Approximation property of SBL

Summary

- Established a connection between SBL and MPDR beamforming.
  - Provides better insight into effective BF
  - Enables an approach to deal with more intractable inference problems
- Discussed Uniform Linear Arrays and Toeplitz Matrix Approximation property of SBL
- Shown effectiveness of SBL for Nested Arrays to identify more sources than sensors