

Subspace Methods for Multi-Microphone Speech Dereverberation

Sharon Gannot

Technion - Israeli Institute of Technology
Faculty of Electrical Engineering
Signal and Image Processing Laboratory

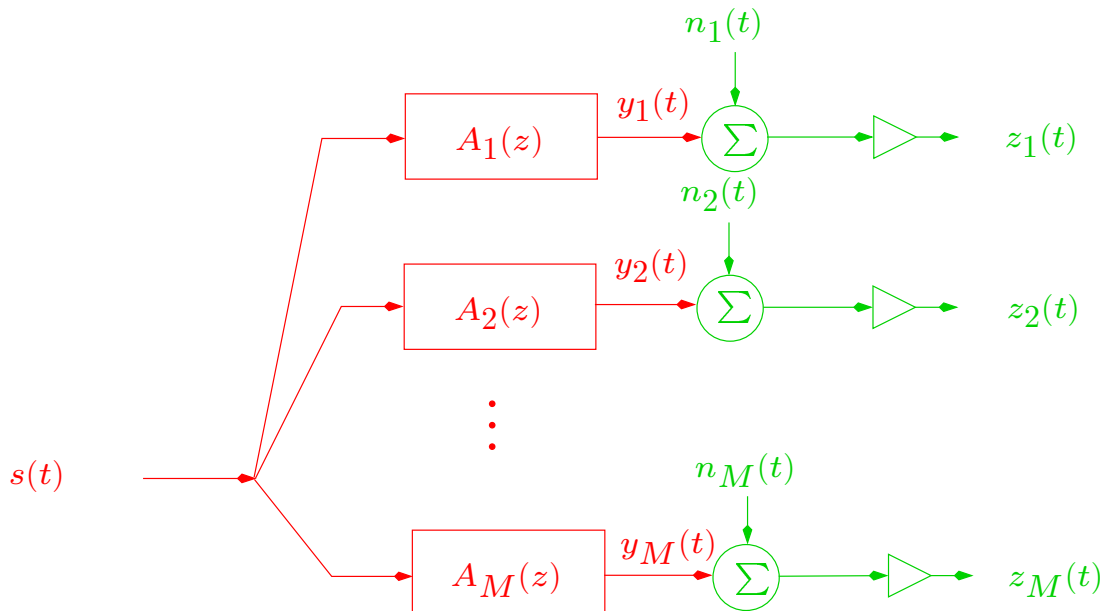
Marc Moonen

Katholieke Universiteit Leuven
Dept. of Electrical Engineering, ESAT-SISTA

May, 2002



The Dereverberation Problem



M Microphones:

$$z_m(t) = a_m(t) * s(t) + n_m(t)$$

$$\sum_{k=0}^{n_a} a_m(k) s(t-k) + n_m(t) ; t = 0, 1, \dots, T$$

Goal

Find desired signal $s(t)$:

- Directly.
- Estimate M ATFs, $A_m(z)$.



Outline

- The data null space - definitions.
- ATFs estimation
 - Two microphone, noiseless case.
 - Two microphone, noisy case.
 - Multi microphone case ($M > 2$).
- Signal reconstruction.
- Preliminary Experimental Study.

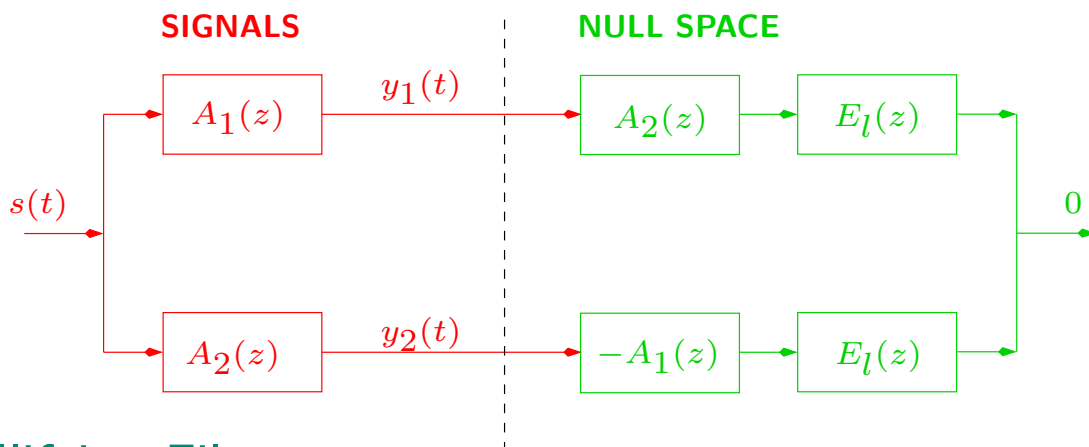


The Data Null Space - Intuition (Two Microphone, Noiseless Case)

Signals:

$$y_1(t) = a_1(t) * s(t)$$

$$y_2(t) = a_2(t) * s(t).$$



Nullifying Filters:

$$[y_2(t) * a_1(t) - y_1(t) * a_2(t)] * e_l(t) = 0$$

- ATFs are embedded in nullifying filters.
- Arbitrary unknown multiplying filters, $E_l(z)$.



The Data Null Space - Formulation

(Two Microphone, Noiseless Case)

One Channel Data Matrix:

$$\mathcal{Y}_m^T = \begin{bmatrix} y_m(0) & 0 & \cdots & 0 \\ y_m(1) & y_m(0) & & \\ \vdots & y_m(1) & \ddots & \\ & \vdots & \ddots & \ddots & 0 \\ y_m(\hat{n}_a) & & & & y_m(0) \\ y_m(\hat{n}_a + 1) & & & & y_m(1) \\ \vdots & & & \ddots & \vdots \\ y_m(T) & & & & y_m(\hat{n}_a) \\ 0 & y_m(T) & & & \vdots \\ \vdots & 0 & & & \\ & & & \ddots & \\ 0 & 0 & \cdots & 0 & y_m(T) \end{bmatrix}$$

Two Channel Data Matrix:

$$\mathcal{Y}^T = \begin{bmatrix} \mathcal{Y}_2^T & -\mathcal{Y}_1^T \end{bmatrix}$$

Correlation Matrix:

$$\hat{R}_y = \mathcal{Y}\mathcal{Y}^T / (T + 1)$$



Null Space Vectors

Eigenvalue decomposition (EVD) to \hat{R}_y

OR

Singularvalue decomposition (SVD) to \mathcal{Y} :

$$\begin{aligned} \lambda_l &= 0; & \mathbf{g}_l & \quad l = 0, 1, \dots, \hat{n}_a - n_a \\ \lambda_l &> 0; & & \quad \text{otherwise} \end{aligned}$$

Split,

$$\mathcal{G} = \begin{bmatrix} \mathbf{g}_0 & \mathbf{g}_1 & \cdots & \mathbf{g}_{\hat{n}_a - n_a} \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{a}}_{1,0} & \tilde{\mathbf{a}}_{1,1} & \cdots & \tilde{\mathbf{a}}_{1,\hat{n}_a - n_a} \\ \tilde{\mathbf{a}}_{2,0} & \tilde{\mathbf{a}}_{2,1} & \cdots & \tilde{\mathbf{a}}_{2,\hat{n}_a - n_a} \end{bmatrix}.$$



Null Space: Transfer Functions

$$\tilde{A}_{ml}(z) = \sum_{k=0}^{\hat{n}_a} \tilde{a}_{ml}(k) z^{-k} = A_m(z) E_l(z)$$

$$l = 0, 1, \dots, \hat{n}_a - n_a ; m = 1, 2.$$

- For $m = 1, 2, \dots, M$:
 $\tilde{A}_{ml}(z)$ have $\hat{n}_a - n_a$ common roots $\Rightarrow E_l(z)$.
- For $l = 0, 1, \dots, \hat{n}_a - n_a$:
 $\tilde{A}_{ml}(z)$ have n_a common roots $\Rightarrow A_m(z)$.



Null Space: Impulse Response

In matrix form,

$$\begin{bmatrix} \tilde{a}_{ml}(0) \\ \tilde{a}_{ml}(1) \\ \vdots \\ \tilde{a}_{ml}(\hat{n}_a) \end{bmatrix} = \underbrace{\begin{bmatrix} a_m(0) & 0 & 0 & \cdots & 0 \\ a_m(1) & a_m(0) & 0 & \cdots & 0 \\ \vdots & a_m(1) & \ddots & & \vdots \\ a_m(n_a) & \vdots & \ddots & \ddots & 0 \\ 0 & a_m(n_a) & & \ddots & a_m(0) \\ \vdots & 0 & & & a_m(1) \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & a_m(n_a) \end{bmatrix}}_{\hat{n}_a - n_a + 1} \begin{bmatrix} e_l(0) \\ e_l(1) \\ \vdots \\ e_l(\hat{n}_a - n_a) \end{bmatrix}$$

Compactly,

$$\tilde{\mathbf{a}}_{ml} = \mathcal{A}_m \mathbf{e}_l$$

$$l = 0, 1, \dots, \bar{L} \leq (\hat{n}_a - n_a)$$

$$\mathcal{E} = \begin{bmatrix} \mathbf{e}_0 & \mathbf{e}_1 & \cdots & \mathbf{e}_{\bar{L}} \end{bmatrix},$$

$$\mathcal{G} = \begin{bmatrix} \mathcal{A}_1 \\ \mathcal{A}_2 \end{bmatrix} \mathcal{E} \triangleq \mathcal{A} \mathcal{E}$$



ATFs Extraction

- Full null space ($\bar{L} = \hat{n}_a - n_a$)

$$\mathcal{E}^i = \begin{bmatrix} \mathbf{e}_0^i & \mathbf{e}_1^i & \cdots & \mathbf{e}_{\hat{n}_a - n_a}^i \end{bmatrix} = \text{inv}(\mathcal{E})$$

$$\implies \mathcal{G}\mathcal{E}^i = \mathcal{A}$$

- Partial null space ($\bar{L} < \hat{n}_a - n_a$)

$$\bar{\mathcal{G}} = \begin{bmatrix} \mathcal{G} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathcal{G} & \mathbf{0} & \mathbf{0} \\ \vdots & & \ddots & \mathbf{0} \\ & & & \ddots \\ \mathbf{0} & & & \mathcal{G} \end{bmatrix} = \bar{\mathcal{A}} \underbrace{\begin{bmatrix} \mathcal{E} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathcal{E} & \mathbf{0} & \mathbf{0} \\ \vdots & & \ddots & \mathbf{0} \\ & & & \ddots \\ \mathbf{0} & & & \mathcal{E} \end{bmatrix}}_{L > \hat{n}_a - n_a + \hat{l}} = \bar{\mathcal{A}}\bar{\mathcal{E}}$$

$$\mathcal{E}^{Pi} = \text{Pinv}(\mathcal{E}) = \mathcal{E}^T (\mathcal{E}\mathcal{E}^T)^{-1}$$

$$\implies \mathcal{G}\mathcal{E}^{Pi} = \bar{\mathcal{A}}$$



ATFs Extraction: Efficient Procedure

\mathcal{A} has shifted columns !

$$\underbrace{\begin{bmatrix} \mathcal{G} & \mathcal{O} & \dots & \dots & \dots & \mathcal{O} & -\mathcal{I}^{(0)} \\ \mathcal{O} & \mathcal{G} & \mathcal{O} & \dots & \dots & \mathcal{O} & -\mathcal{I}^{(1)} \\ \vdots & \mathcal{O} & \ddots & & & \vdots & \vdots \\ \vdots & \vdots & & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & & \ddots & \ddots & \mathcal{O} & \vdots \\ \mathcal{O} & \mathcal{O} & \dots & \dots & \mathcal{O} & \mathcal{G} & -\mathcal{I}^{(L)} \end{bmatrix}}_{\tilde{\mathcal{G}}} \begin{bmatrix} \mathbf{e}_0^i \\ \mathbf{e}_1^i \\ \vdots \\ \mathbf{e}_L^i \\ \mathbf{a}_1 \\ \mathbf{a}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

- \mathcal{O} all-zeros matrix. $\mathbf{0}$ all-zero vector.
- $\mathcal{I}^{(l)}$; $l = 0, 1, \dots, L$ a shift by l matrix.

$$\implies \tilde{\mathcal{G}}\mathbf{x} = \mathbf{0}$$

$$\tilde{\mathcal{G}} \Rightarrow \text{Eigenvalue of value } 0 \Rightarrow \mathbf{a}_m$$



Two Microphone Noisy Case

Two Microphone Noisy Case

$$z_m(t) = y_m(t) + n_m(t)$$

Then, the data matrix,

$$\mathcal{Z} = \mathcal{Y} + \mathcal{N}$$

The correlation matrix (for long observation time),

$$\hat{R}_z \approx \hat{R}_y + \hat{R}_n$$

Solve,

$$\tilde{\mathcal{G}}\mathbf{x} = \boldsymbol{\mu}$$

Where, $\boldsymbol{\mu}$ is an error term.

Find the smallest eigenvector of the matrix $\tilde{\mathcal{G}}$ (TLS).



Noise Correlation Matrix Dependency

How \hat{R}_n effects $\tilde{\mathcal{G}}$?

White Noise Case

- $\hat{R}_n \approx \sigma^2 I$.
- Choose eigenvectors corresponding to the eigenvalues of value σ^2 .
- \mathcal{G} remains intact.

Colored Noise Case

- Calculate Generalized EVD of \hat{R}_z and \hat{R}_n (or Generalized SVD of \mathcal{Z} and \mathcal{N}).
- Choose eigenvectors corresponding to the eigenvalues of value 1.
- Subsequent stages remain intact.



Multi Microphone Case ($M > 2$)

Channel paring

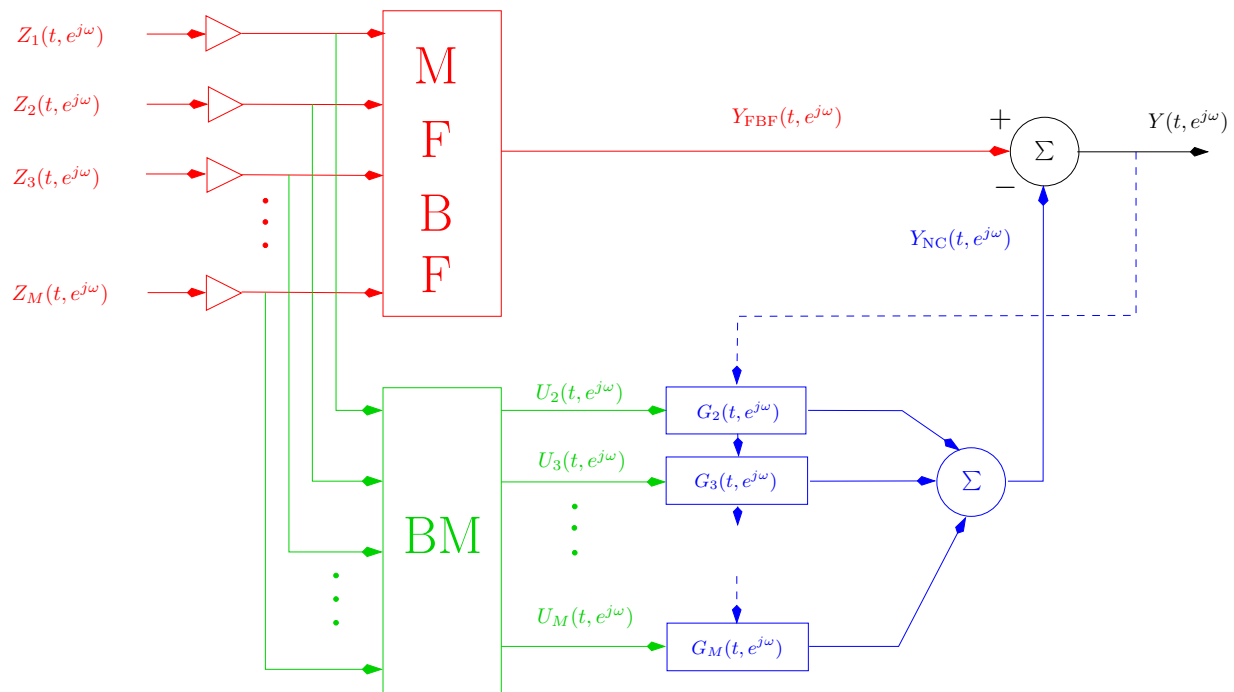
$$[y_i(t) * a_j(t) - y_j(t) * a_i(t)] * e_l(t) = 0.$$

Data matrix

$$Z = \begin{bmatrix} z_2 & z_3 & \cdots & z_M & 0 & \cdots & 0 & \cdots & 0 \\ -z_1 & 0 & \cdots & & z_3 & \cdots & z_M & & 0 \\ 0 & -z_1 & & & -z_2 & & 0 & & \vdots \\ \vdots & 0 & \ddots & & & & \vdots & & 0 \\ & \vdots & & \ddots & & & 0 & & z_M \\ 0 & 0 & \cdots & -z_1 & \cdots & -z_2 & \cdots & -z_{M-1} \end{bmatrix}$$



Signal Reconstruction



Signal Reconstruction (Details)

- Noisy signals (time-frequency).

$$Z_m(t, e^{j\omega}) = A_m(e^{j\omega})S(t, e^{j\omega}) + N_m(t, e^{j\omega})$$

$$m = 1, \dots, M$$

- Fixed beamformer (MFBF).

$$Y_{\text{FBF}}(t, e^{j\omega}) = \frac{1}{\|\hat{A}(e^{j\omega})\|^2} \sum_{m=1}^M Z_m(t, e^{j\omega}) \hat{A}_m^*(e^{j\omega})$$

- Blocking Matrix (BM).

$$U_m(t, e^{j\omega}) = Z_m(t, e^{j\omega}) A_1(e^{j\omega}) - Z_1(t, e^{j\omega}) A_m(e^{j\omega})$$

$$m = 2, \dots, M$$

- Noise canceller(NC).

$$\tilde{G}_m(t+1, e^{j\omega}) = G_m(t, e^{j\omega}) + \mu \frac{U_m(t, e^{j\omega}) Y^*(t, e^{j\omega})}{P_{\text{est}}(t, e^{j\omega})}$$

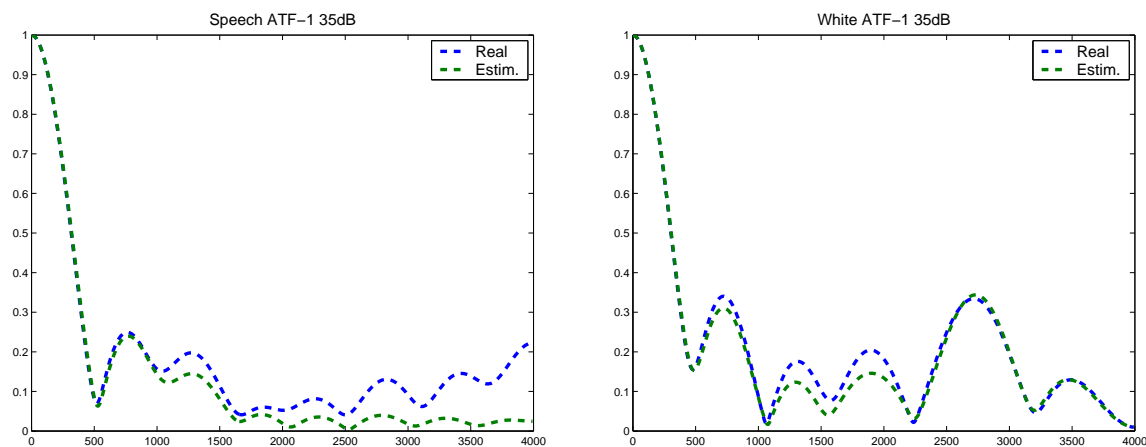
$$G_m(t+1, e^{j\omega}) \stackrel{\text{FIR}}{\leftarrow} \tilde{G}_m(t+1, e^{j\omega})$$

$$m = 2, \dots, M$$



Preliminary Results

- Real ATFs order: $n_a = 15$.
- Overestimated ATFs order: $\hat{n}_a = 22$.
- Signal: Speech (left), White Noise (right).
- Signal Length: $T = 4$ seconds.
- No. of microphones: $M = 3$.
- Noise: Directional, Non-White.



Conclusions

Properties:

- "Good" ATFs estimation (for short ATFs only).
- Limited performance in low energy bands (eigenvalue spread).
- Applicable to equalization.
- High computational burden.

⇒ Use Frequency domain or Subband method

Warning:

- Perfect reconstruction.
- Gain Ambiguity.

