# Subspace Methods for Multi-Microphone Speech Dereverberation

**Sharon Gannot**

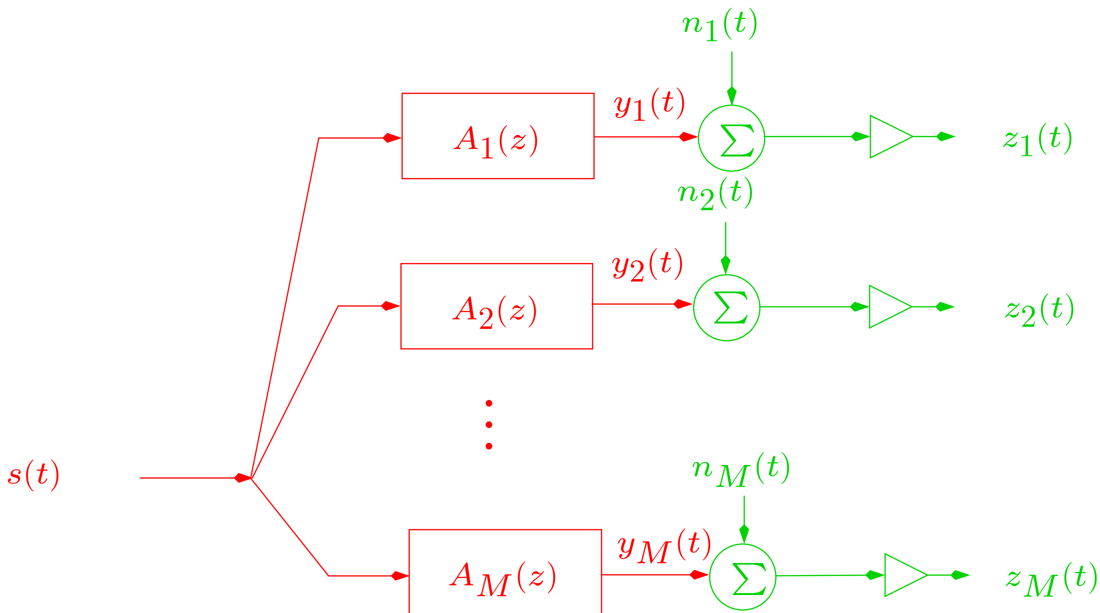
Technion - Israeli Institute of Technology
Faculty of Electrical Engineering
Signal and Image Processing Laboratory

**Marc Moonen**

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

May, 2002

# The Dereverberation Problem




## $M$ Microphones:

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

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

### Goal

Find desired signal $s(t)$:

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

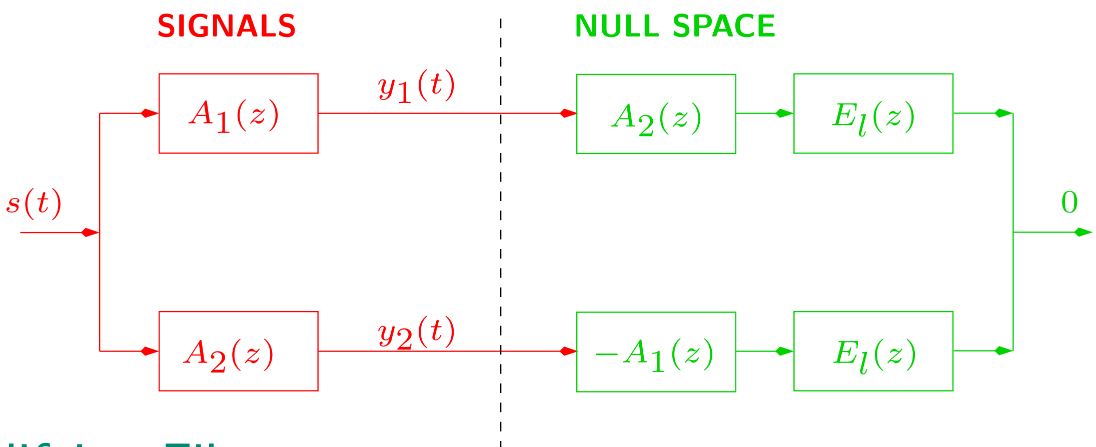
# Outline

- 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:

$$
\begin{aligned}
y_1(t) &= a_1(t) * s(t) \\
y_2(t) &= a_2(t) * s(t).
\end{aligned}
$$




## 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:

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

## 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 $\mathcal{Y}$:

$$\boxed{\begin{array}{lll} \lambda_l = 0; & \mathbf{g}_l & l = 0, 1, \ldots, \hat{n}_a - n_a \\ \lambda_l > 0; & & \text{otherwise} \end{array}}$$

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

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

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

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

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

# Null Space: Impulse Response

In matrix form,

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

Compactly,

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

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

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

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

# ATFs Extraction

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

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

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

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

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

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

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

# ATFs Extraction: Efficient Procedure

$\boxed{\mathcal{A} \text{ has shifted columns !}}$

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

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

$$\boxed{\Longrightarrow \tilde{\mathcal{G}}\mathbf{x} = \mathbf{0}}$$

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

# 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,

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

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 $\sigma^2$.
- $\mathcal{G}$ 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 $(M > 2)$

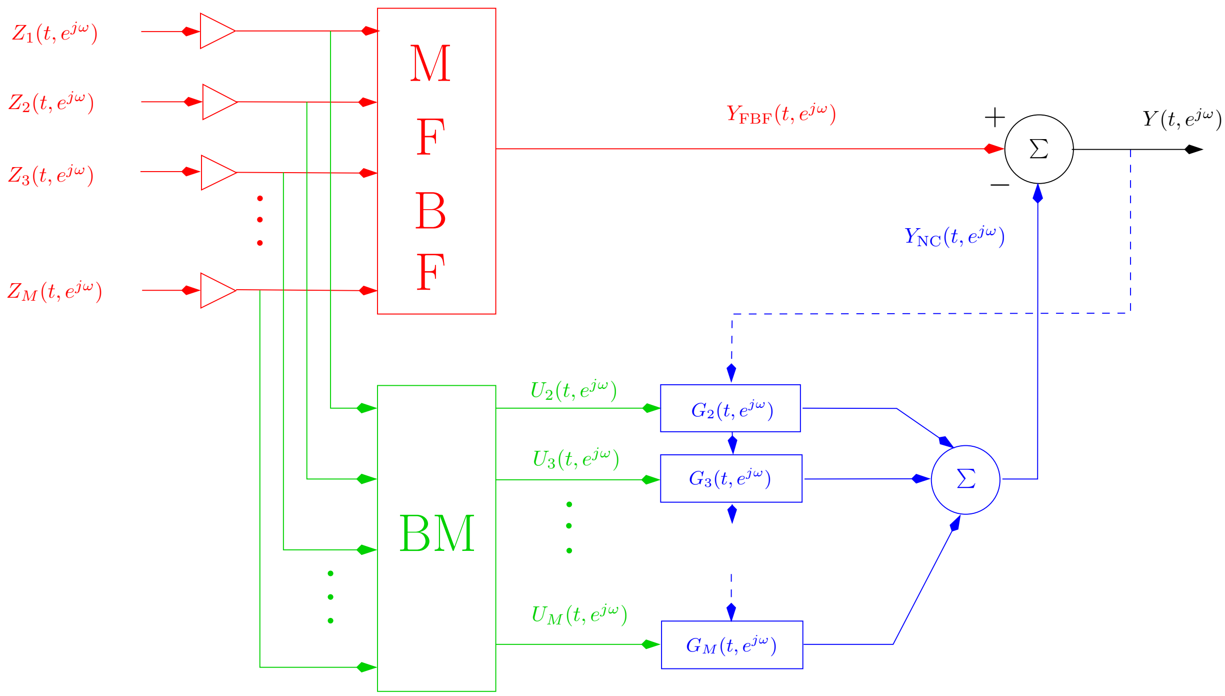
## Channel paring

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

## Data matrix

$$\mathcal{Z} = \begin{bmatrix} \mathcal{Z}_2 & \mathcal{Z}_3 & \cdots & & \mathcal{Z}_M & \mathcal{O} & \cdots & \mathcal{O} & \cdots & \mathcal{O} \\ -\mathcal{Z}_1 & \mathcal{O} & \cdots & & & \mathcal{Z}_3 & \cdots & \mathcal{Z}_M & & \mathcal{O} \\ \mathcal{O} & -\mathcal{Z}_1 & & & & -\mathcal{Z}_2 & & \mathcal{O} & & \vdots \\ \vdots & \mathcal{O} & \ddots & & & & & \vdots & & \mathcal{O} \\ & \vdots & & \ddots & & & & \mathcal{O} & & \mathcal{Z}_M \\ \mathcal{O} & \mathcal{O} & \cdots & & -\mathcal{Z}_1 & \cdots & & -\mathcal{Z}_2 & \cdots & -\mathcal{Z}_{M-1} \end{bmatrix}$$

# Signal Reconstruction



# Signal Reconstruction (Details)

- Noisy signals (time-frequency).

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

$$m = 1, \ldots, M$$

- Fixed beamformer (MFBF).

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

- Blocking Matrix (BM).

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

$$m = 2, \ldots, M$$

- Noise canceller(NC).

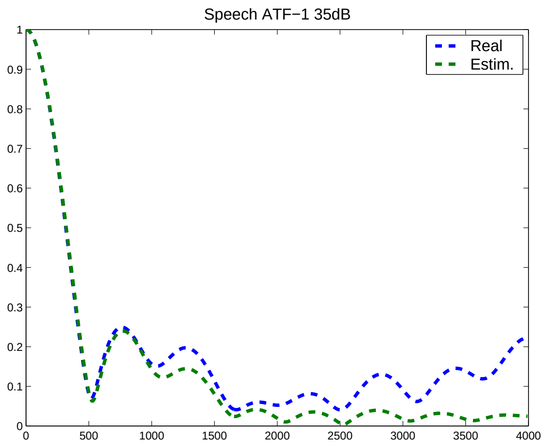
$$\tilde{G}_m(t+1, e^{j\omega}) = G_m(t, e^{j\omega}) + \mu \frac{U_m(t, e^{j\omega}) Y^*(t, e^{j\omega})}{P_{\mathsf{est}}(t, e^{j\omega})}$$

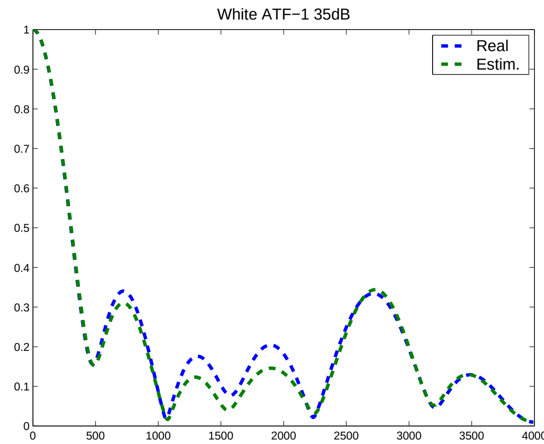
$$G_m(t+1, e^{j\omega}) \xleftarrow{\text{FIR}} \tilde{G}_m(t+1, e^{j\omega})$$

$$m = 2, \ldots, M$$

# Preliminary Results

- Real ATFs order: $n_a = 15$.
- Overestimated ATFs order: $\hat{n}_a = 22$.
- Signal: Speech (left), White Noise (right).
- Signal Length: $T = 4 \quad \text{seconds}$.
- 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.

$\boxed{\Rightarrow \text{ Use Frequency domain or Subband method}}$

**Warning:**

- Perfect reconstruction.
- Gain Ambiguity.