Microphone-Array Signal Processing

José A. Apolinário Jr. and Marcello L. R. de Campos

{apolin},{mcampos}@ieee.org

COPPE/Poli/UFRJ

IME – Lab. Processamento Digital de Sinais

Outline

1. Introduction and Fundamentals
2. Sensor Arrays and Spatial Filtering
3. Optimal Beamforming
4. Adaptive Beamforming
5. DoA Estimation with Microphone Arrays

5. DOA Estimation with Microphone Arrays

5.0 Signal Preparation

- It is usual to find a delayed signal represented by a multiplication of the signal with exponential $e^{j\omega_o\tau}$

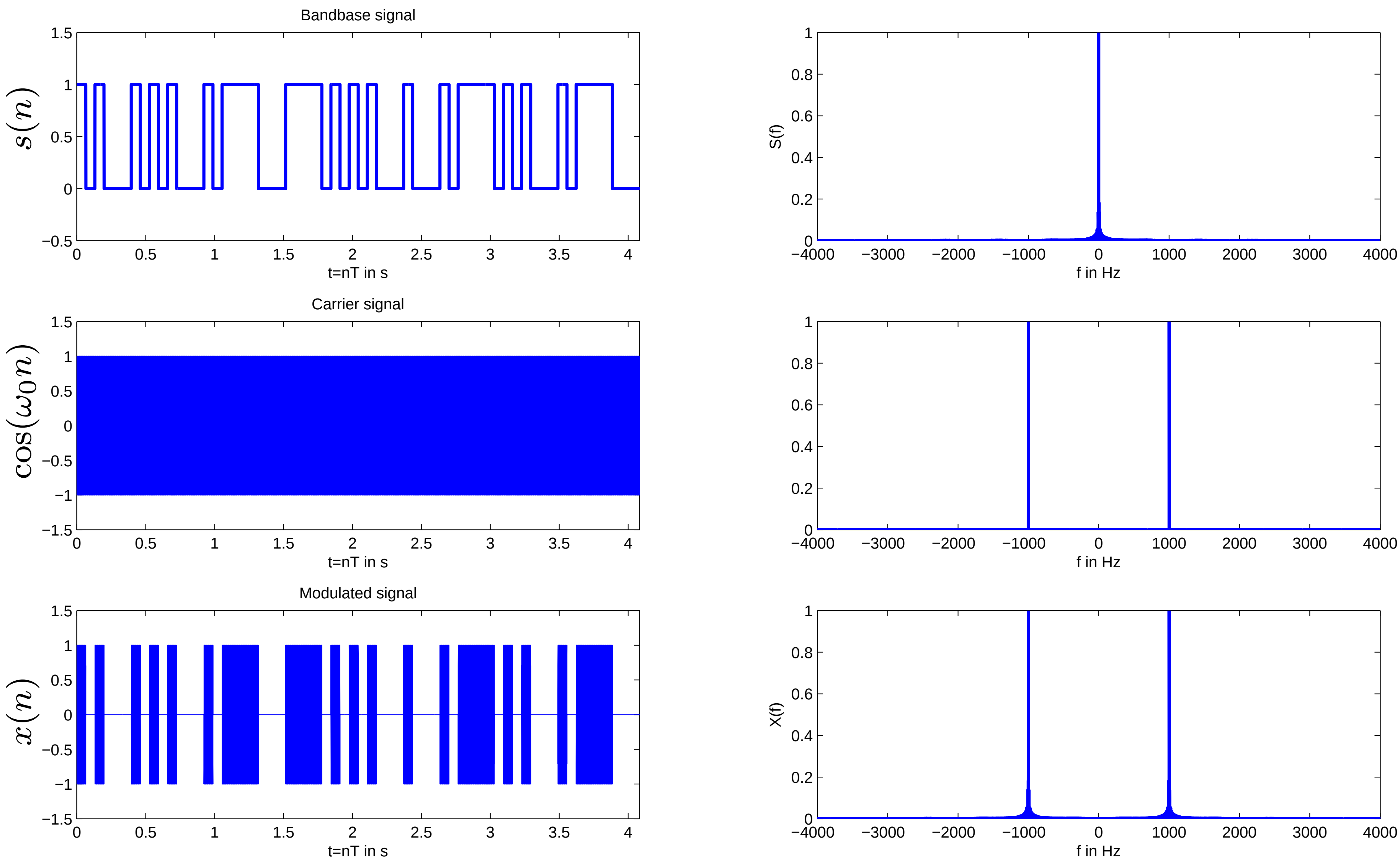
- It is usual to find a delayed signal represented by a multiplication of the signal with exponential $e^{j\omega_o\tau}$
- First thing to note: when this is the case, the signal is narrow band with a center frequency in ω_0 (in the continuous-time domain, it corresponds to a carrier frequency $\Omega_0 = f_s\omega_0$)

- It is usual to find a delayed signal represented by a multiplication of the signal with exponential $e^{j\omega_o\tau}$
- First thing to note: when this is the case, the signal is narrow band with a center frequency in ω_0 (in the continuous-time domain, it corresponds to a carrier frequency $\Omega_0 = f_s\omega_0$)
- But, most importantly, the delay is well represented only if the signal is also analytic, i. e., having only non-negative frequency components.

- It is usual to find a delayed signal represented by a multiplication of the signal with exponential $e^{j\omega_o\tau}$
- First thing to note: when this is the case, the signal is narrow band with a center frequency in ω_0 (in the continuous-time domain, it corresponds to a carrier frequency $\Omega_0 = f_s\omega_0$)
- But, most importantly, the delay is well represented only if the signal is also analytic, i. e., having only non-negative frequency components.
- An analytic signal, mathematically, can be obtained by multiplying its Fourier transform by the continuous Heaviside step function:

$$X_a(e^{j\omega}) = 2X(e^{j\omega})u(\omega), u(\omega) = \begin{cases} 0, \omega < 0 \\ 1, \omega = 0 \\ 1, \omega > 0 \end{cases}$$

- Let $x(n) = s(n)\cos(\omega_0 n)$, $s(n)$ having a maximum frequency component (ω_m) much lower than ω_0:



- If $x(n) = s(n)e^{j\omega_0 n}$, then $x(n)e^{-j\omega_0 \tau} = s(n)e^{j\omega_0 (n-\tau)} \approx x(n-\tau)$ if $\tau \ll 1/\omega_m$

- If $x(n) = s(n)e^{j\omega_0 n}$, then
  $x(n)e^{-j\omega_0 \tau} = s(n)e^{j\omega_0 (n-\tau)} \approx x(n-\tau)$ if $\tau \ll 1/\omega_m$
- But if $x(n) = s(n)cos(\omega_0 n)$, then $x(n)e^{-j\omega_0 \tau} \neq x(n-\tau)$

- If $x(n) = s(n)e^{j\omega_0 n}$, then
  $x(n)e^{-j\omega_0\tau} = s(n)e^{j\omega_0(n-\tau)} \approx x(n-\tau)$ if $\tau \ll 1/\omega_m$

- But if $x(n) = s(n)cos(\omega_0 n)$, then $x(n)e^{-j\omega_0\tau} \neq x(n-\tau)$

- We can make
  $$x(n) = s(n)cos(\omega_0 n) = \underbrace{\frac{s(n)}{2}e^{j\omega_0 n}}_{x_+(n)} + \underbrace{\frac{s(n)}{2}e^{-j\omega_0 n}}_{x_-(n)}$$ such that
  $$x(n-\tau) \approx x_+e^{-j\omega_0\tau} + x_-(n)e^{+j\omega_0\tau} = s(n)cos(\omega_0(n-\tau))$$

- If $x(n) = s(n)e^{j\omega_0 n}$, then
  $x(n)e^{-j\omega_0\tau} = s(n)e^{j\omega_0(n-\tau)} \approx x(n-\tau)$ if $\tau \ll 1/\omega_m$

- But if $x(n) = s(n)cos(\omega_0 n)$, then $x(n)e^{-j\omega_0\tau} \neq x(n-\tau)$

- We can make

$$x(n) = s(n)cos(\omega_0 n) = \underbrace{\frac{s(n)}{2}e^{j\omega_0 n}}_{x_+(n)} + \underbrace{\frac{s(n)}{2}e^{-j\omega_0 n}}_{x_-(n)} \text{ such that}$$

$$x(n-\tau) \approx x_+ e^{-j\omega_0\tau} + x_-(n)e^{+j\omega_0\tau} = s(n)cos(\omega_0(n-\tau))$$

- $\cdots$ but, how to obtain $x_+(n)$ or a scaled copy? Using the Hilbert Transform $x_H(n) = \mathcal{HT}\{x(n)\}$ where

$$X_H(e^{j\omega}) = \begin{cases} jX(e^{j\omega}), -\pi < \omega < 0 \\ X(e^{j\omega}), \omega = 0 \\ -jX(e^{j\omega}), 0 < \omega < \pi \end{cases}$$

- Knowing that $x(n) = x_-(n) + x_+(n) = \mathcal{F}^{-1}\left\{X_-(e^{j\omega}) + X_+(e^{j\omega})\right\}$, we compute $y(n) = x(n) + jx_H(n)$

- Knowing that $x(n) = x_-(n) + x_+(n) = \mathcal{F}^{-1}\left\{X_-(e^{j\omega}) + X_+(e^{j\omega})\right\}$, we compute $y(n) = x(n) + jx_H(n)$

- $y(n) = \mathcal{F}^{-1}\{X_-(e^{j\omega}) + X_+(e^{j\omega}) + j\underbrace{[jX_((e^{j\omega}) - jX_+(e^{j\omega})]}_{X_H(e^{j\omega})}\}$

- Knowing that $x(n) = x_-(n) + x_+(n) = \mathcal{F}^{-1}\left\{X_-(e^{j\omega}) + X_+(e^{j\omega})\right\}$, we compute $y(n) = x(n) + jx_H(n)$

- $$y(n) = \mathcal{F}^{-1}\{X_-(e^{j\omega}) + X_+(e^{j\omega}) + j\underbrace{[jX_(e^{j\omega}) - jX_+(e^{j\omega})]}_{X_H(e^{j\omega})}\}$$

$$= \mathcal{F}^{-1}\{X_-(e^{j\omega}) + X_+(e^{j\omega}) - X_(e^{j\omega}) + X_+(e^{j\omega})\}$$

- Knowing that $x(n) = x_-(n) + x_+(n) = \mathcal{F}^{-1}\left\{X_-(e^{j\omega}) + X_+(e^{j\omega})\right\}$, we compute $y(n) = x(n) + jx_H(n)$

- $$y(n) = \mathcal{F}^{-1}\{X_-(e^{j\omega}) + X_+(e^{j\omega}) + j\underbrace{[jX_(e^{j\omega}) - jX_+(e^{j\omega})]}_{X_H(e^{j\omega})}\}$$

$$= \mathcal{F}^{-1}\{X_-(e^{j\omega}) + X_+(e^{j\omega}) - X_(e^{j\omega}) + X_+(e^{j\omega})\}$$

- Therefore $y(n) = 2\mathcal{F}^{-1}\{X_+(e^{j\omega})\} = s(n)e^{j\omega_0 n}$ which is analytic!

Signal Model

- Consider $x_m(t)$ the signal from the m-th microphone (prior to the A/D converter) corresponding to audio from D sources (directions θ_1 to θ_D) plus noise:
  $x_m(t) = s_1(t - \bar{\tau}_m(\theta_1)) + \cdots + s_D(t - \bar{\tau}_m(\theta_D)) + n_m(t)$

Signal Model

- Consider $x_m(t)$ the signal from the m-th microphone (prior to the A/D converter) corresponding to audio from D sources (directions θ_1 to θ_D) plus noise:
  $x_m(t) = s_1(t - \bar{\tau}_m(\theta_1)) + \cdots + s_D(t - \bar{\tau}_m(\theta_D)) + n_m(t)$

- Assuming $\bar{\tau}_m(\theta_d) = T\tau_m(\theta_d)$ in s ($\tau_m(\theta_d)$ in number of samples), after the A/D converter and $\{\cdot\} + j\mathcal{HT}\{\cdot\}$ to make it an analytic signal, we could write
  $x_m(n) = s_1(n)e^{-j\omega_0\tau_m(\theta_1)} + \cdots + s_D(n)e^{-j\omega_0\tau_m(\theta_D)} + n_m(n)$

Signal Model

- Consider $x_m(t)$ the signal from the m-th microphone (prior to the A/D converter) corresponding to audio from D sources (directions θ_1 to θ_D) plus noise:
$x_m(t) = s_1(t - \bar{\tau}_m(\theta_1)) + \cdots + s_D(t - \bar{\tau}_m(\theta_D)) + n_m(t)$

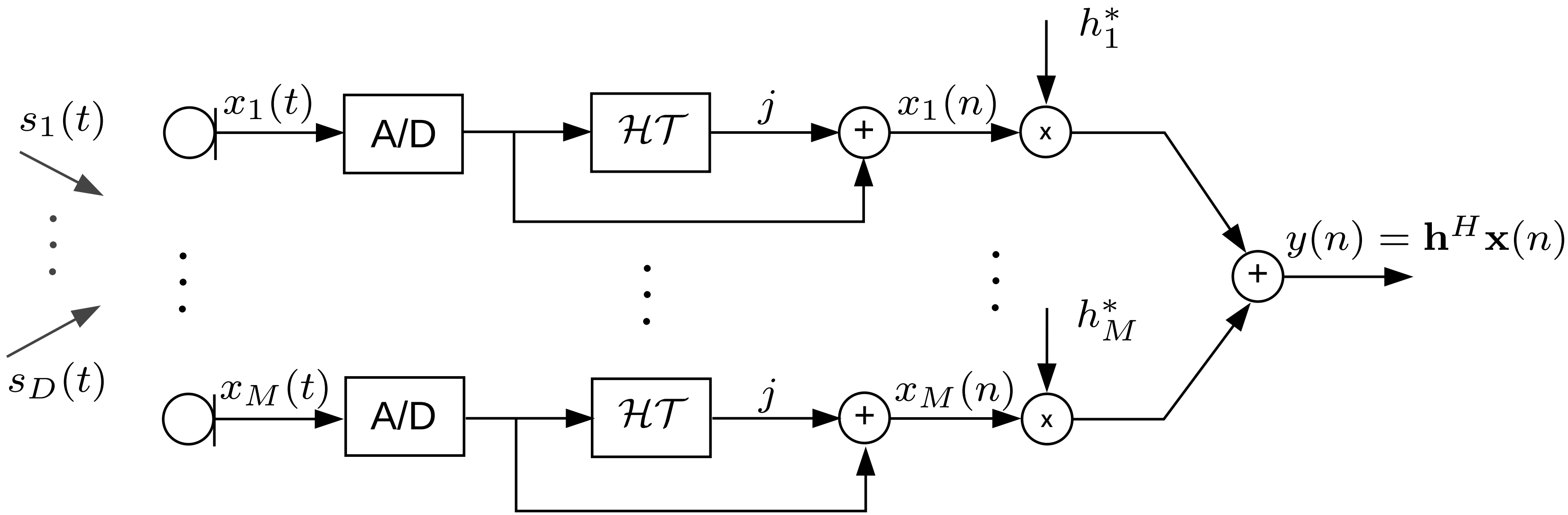
- Assuming $\bar{\tau}_m(\theta_d) = T\tau_m(\theta_d)$ in s ($\tau_m(\theta_d)$ in number of samples), after the A/D converter and $\{.\} + j\mathcal{HT}\{.\}$ to make it an analytic signal, we could write
$x_m(n) = s_1(n)e^{-j\omega_0\tau_m(\theta_1)} + \cdots + s_D(n)e^{-j\omega_0\tau_m(\theta_D)} + n_m(n)$

- For an array with M microphones, we would have:

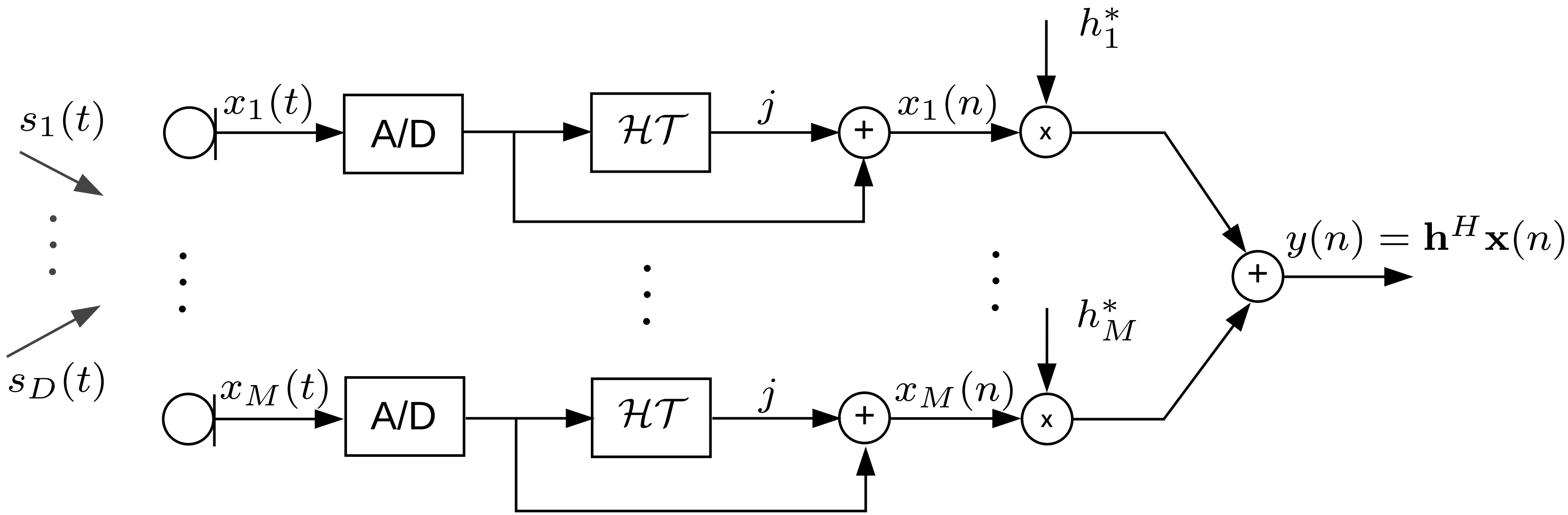
$$\underbrace{\mathbf{x}(n)}_{M\times 1} = \underbrace{\mathbf{A}}_{M\times D}\underbrace{\mathbf{s}(n)}_{D\times 1} + \underbrace{\mathbf{n}(n)}_{M\times 1}$$

5.1 Signal model

- Assume, initially, we have D narrowband signals coming from unknown directions:

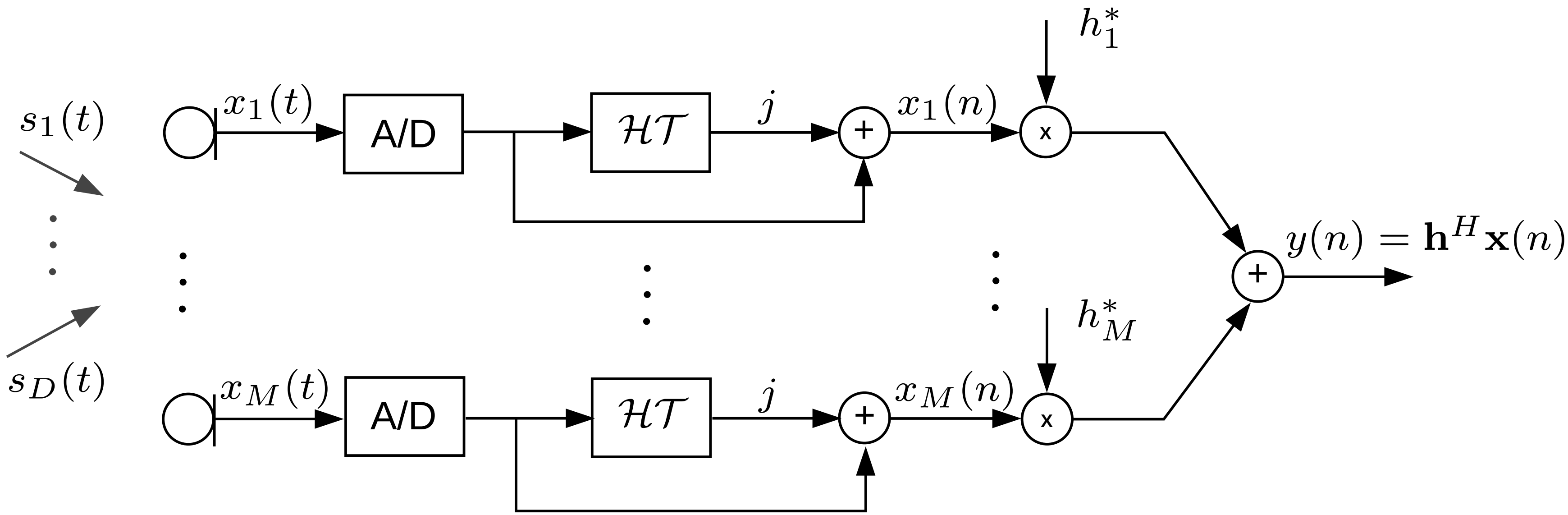


- Assume, initially, we have D narrowband signals coming from unknown directions:



- $$\mathbf{x}(n) = \begin{bmatrix} e^{-j\omega_0 \tau_1(\theta_1)} s_1(n) + \cdots + e^{-j\omega_0 \tau_1(\theta_D)} s_D(n) + n_1(n) \\ \vdots \\ e^{-j\omega_0 \tau_M(\theta_1)} s_1(n) + \cdots + e^{-j\omega_0 \tau_M(\theta_D)} s_D(n) + n_M(n) \end{bmatrix}$$

- Assume, initially, we have D narrowband signals coming from unknown directions:



- $$\mathbf{x}(n) = \begin{bmatrix} e^{-j\omega_0\tau_1(\theta_1)}s_1(n) + \cdots + e^{-j\omega_0\tau_1(\theta_D)}s_D(n) + n_1(n) \\ \vdots \\ e^{-j\omega_0\tau_M(\theta_1)}s_1(n) + \cdots + e^{-j\omega_0\tau_M(\theta_D)}s_D(n) + n_M(n) \end{bmatrix}$$

- Such that the output signal can be written as
$y(n) = \mathbf{h}^H\mathbf{x}(n) = \mathbf{h}^H\left[\mathbf{A}\mathbf{s}(n) + \mathbf{n}(n)\right]$

- If we now assume one single signal, $s(n)$, coming from direction θ, then
  $\mathbf{x}(n) = s(n)\mathbf{a}(\theta) + \mathbf{n}(n)$

- If we now assume one single signal, $s(n)$, coming from direction θ, then
  $\mathbf{x}(n) = s(n)\mathbf{a}(\theta) + \mathbf{n}(n)$

- And the output signal becomes
  $y(n) = \mathbf{h}^H\mathbf{a}(\theta)s(n) + \mathbf{h}^H\mathbf{n}(n)$

- If we now assume one single signal, $s(n)$, coming from direction θ, then
  $\mathbf{x}(n) = s(n)\mathbf{a}(\theta) + \mathbf{n}(n)$
- And the output signal becomes
  $y(n) = \mathbf{h}^H \mathbf{a}(\theta) s(n) + \mathbf{h}^H \mathbf{n}(n)$
- If we make $\mathbf{h}^H \mathbf{a}(\theta) = 1$, the output signal would correspond to $y(n) = s(n) + \underbrace{\mathbf{h}^H \mathbf{n}(n)}_{\text{noise}}$

- If we now assume one single signal, $s(n)$, coming from direction θ, then
$\mathbf{x}(n) = s(n)\mathbf{a}(\theta) + \mathbf{n}(n)$
- And the output signal becomes
$y(n) = \mathbf{h}^H\mathbf{a}(\theta)s(n) + \mathbf{h}^H\mathbf{n}(n)$
- If we make $\mathbf{h}^H\mathbf{a}(\theta) = 1$, the output signal would correspond to $y(n) = s(n) + \underbrace{\mathbf{h}^H\mathbf{n}(n)}_{\text{noise}}$
- Also note that $E[|y(n)|^2] = \mathbf{h}^H\mathbf{R}_x\mathbf{h}$, $\mathbf{R}_x = E[\mathbf{x}(n)\mathbf{x}^H(n)]$

5.2 Non-parametric methods: BF (beamforming a.k.a. Delay & Sum) and Capon

- If $\mathbf{x}(n)$ were spatially white, i.e. $\mathbf{R}_x = \mathbf{I}$, we would obtain $E[|y(n)|^2] = \mathbf{h}^H\mathbf{h}$

- If $\mathbf{x}(n)$ were spatially white, i.e. $\mathbf{R}_x = \mathbf{I}$, we would obtain $E[|y(n)|^2] = \mathbf{h}^H\mathbf{h}$
- Minimizing $E[|y(n)|^2] = \mathbf{h}^H\mathbf{h}$ s.t. $\mathbf{h}^H\mathbf{a}(\theta) = 1$, the result,

- If $\mathbf{x}(n)$ were spatially white, i.e. $\mathbf{R}_x = \mathbf{I}$, we would obtain $E[|y(n)|^2] = \mathbf{h}^H\mathbf{h}$
- Minimizing $E[|y(n)|^2] = \mathbf{h}^H\mathbf{h}$ s.t. $\mathbf{h}^H\mathbf{a}(\theta) = 1$, the result, after using Lagrange multiplier, taking the gradient,

  and equating to zero, is $\mathbf{h} = \mathbf{a}(\theta)/M$ which leads to $E[|y(n)|^2] = \frac{\mathbf{a}^H(\theta)\mathbf{R}_x\mathbf{a}(\theta)}{M^2}$

DS DoA

- If $\mathbf{x}(n)$ were spatially white, i.e. $\mathbf{R}_x = \mathbf{I}$, we would obtain $E[|y(n)|^2] = \mathbf{h}^H\mathbf{h}$
- Minimizing $E[|y(n)|^2] = \mathbf{h}^H\mathbf{h}$ s.t. $\mathbf{h}^H\mathbf{a}(\theta) = 1$, the result, after using Lagrange multiplier, taking the gradient, and equating to zero, is $\mathbf{h} = \mathbf{a}(\theta)/M$ which leads to $E[|y(n)|^2] = \frac{\mathbf{a}^H(\theta)\mathbf{R}_x\mathbf{a}(\theta)}{M^2}$
- Omitting factor $\frac{1}{M^2}$, we estimate the autocorrelation matrix as $\hat{\mathbf{R}}_x = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}(n)\mathbf{x}^H(n)$ and find the direction of interest by varying θ and obtaining the peak in $\boxed{P_{DS}(\theta) = \mathbf{a}^H(\theta)\hat{\mathbf{R}}_x\mathbf{a}(\theta)}$

Capon

- In the method known as Capon, we minimize $E[|y(n)|^2] = \mathbf{h}^H \mathbf{R}_x \mathbf{h}$ subject to $\mathbf{h}^H \mathbf{a}(\theta) = 1$

- In the method known as Capon, we minimize $E[|y(n)|^2] = \mathbf{h}^H \mathbf{R}_x \mathbf{h}$ subject to $\mathbf{h}^H \mathbf{a}(\theta) = 1$
- Using Lagrange multiplier, we write $\xi = \mathbf{h}^H \mathbf{R}_x \mathbf{h} + \lambda(\mathbf{h}^H \mathbf{a}(\theta) - 1)$, and make $\nabla_{\mathbf{h}} \xi = \mathbf{0}$ such that $\mathbf{h} = \frac{\mathbf{R}_x^{-1} \mathbf{a}(\theta)}{\mathbf{a}^H(\theta) \mathbf{R}_x^{-1} \mathbf{a}(\theta)}$

- In the method known as Capon, we minimize $E[|y(n)|^2] = \mathbf{h}^H \mathbf{R}_x \mathbf{h}$ subject to $\mathbf{h}^H \mathbf{a}(\theta) = 1$
- Using Lagrange multiplier, we write $\xi = \mathbf{h}^H \mathbf{R}_x \mathbf{h} + \lambda(\mathbf{h}^H \mathbf{a}(\theta) - 1)$, and make $\nabla_{\mathbf{h}} \xi = \mathbf{0}$ such that $\mathbf{h} = \frac{\mathbf{R}_x^{-1}\mathbf{a}(\theta)}{\mathbf{a}^H(\theta)\mathbf{R}_x^{-1}\mathbf{a}(\theta)}$
- Replacing the above coefficient vector in $E[|y(n)|^2]$, we obtain $E[|y(n)|^2] = \frac{1}{\mathbf{a}^H(\theta)\mathbf{R}_x^{-1}\mathbf{a}(\theta)}$

- In the method known as Capon, we minimize $E[|y(n)|^2] = \mathbf{h}^H \mathbf{R}_x \mathbf{h}$ subject to $\mathbf{h}^H \mathbf{a}(\theta) = 1$
- Using Lagrange multiplier, we write $\xi = \mathbf{h}^H \mathbf{R}_x \mathbf{h} + \lambda(\mathbf{h}^H \mathbf{a}(\theta) - 1)$, and make $\nabla_{\mathbf{h}} \xi = \mathbf{0}$ such that $\mathbf{h} = \frac{\mathbf{R}_x^{-1}\mathbf{a}(\theta)}{\mathbf{a}^H(\theta)\mathbf{R}_x^{-1}\mathbf{a}(\theta)}$
- Replacing the above coefficient vector in $E[|y(n)|^2]$, we obtain $E[|y(n)|^2] = \frac{1}{\mathbf{a}^H(\theta)\mathbf{R}_x^{-1}\mathbf{a}(\theta)}$
- Therefore, in the Capon DoA, we estimate $\hat{\mathbf{R}}_x = \frac{1}{N}\sum_{n=1}^{N} \mathbf{x}(n)\mathbf{x}^H(n)$ and find the direction of interest by varying θ and obtaining the peak in

$$\boxed{P_{CAPON}(\theta) = \frac{1}{\mathbf{a}^H(\theta)\hat{\mathbf{R}}_x^{-1}\mathbf{a}(\theta)}}$$

5.3 Eigenvalue-Based DoA

- Coming back to the previous model of D sources, we write $\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n) + \mathbf{n}(n)$

- Coming back to the previous model of D sources, we write $\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n) + \mathbf{n}(n)$
- We assume $D < M$ (number of signals lower than the number of sensors); this method is known as *parametric* for we make this assumption

- Coming back to the previous model of D sources, we write $\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n) + \mathbf{n}(n)$
- We assume $D < M$ (number of signals lower than the number of sensors); this method is known as *parametric* for we make this assumption
- Also note that $\mathbf{A}$ is $M \times D$, $\mathbf{s}$ is $D \times 1$, and $\mathbf{n}(n)$ is $M \times 1$

- Coming back to the previous model of D sources, we write $\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n) + \mathbf{n}(n)$
- We assume $D < M$ (number of signals lower than the number of sensors); this method is known as *parametric* for we make this assumption
- Also note that $\mathbf{A}$ is $M \times D$, $\mathbf{s}$ is $D \times 1$, and $\mathbf{n}(n)$ is $M \times 1$
- We then write $\mathbf{R}_x = E\left[\mathbf{x}(n)\mathbf{x}^H(n)\right] = \mathbf{A}\mathbf{R}_s\mathbf{A}^H + \mathbf{R}_n$, this last matrix becoming $\mathbf{R}_n = \sigma_n^2\mathbf{I}$ when assuming spatially white noise; $\mathbf{R}_s$ is the $D \times D$ autocorrelation matrix of the signal vector, i.e., $E\left[\mathbf{s}(n)\mathbf{s}^H(n)\right]$

- $\mathbf{R}_x = \mathbf{A}\mathbf{R}_s\mathbf{A}^H + \mathbf{R}_n$ with $D < M$ implies that $\mathbf{A}\mathbf{R}_s\mathbf{A}^H$ is singular (rank D), its determinant is equal to zero and, therefore, $\det\left[\mathbf{R}_x - \sigma_n^2\mathbf{I}\right] = 0$ and σ_n^2 is a (minimum) eigenvalue with multiplicity $M - D$

- $\mathbf{R}_x = \mathbf{A}\mathbf{R}_s\mathbf{A}^H + \mathbf{R}_n$ with $D < M$ implies that $\mathbf{A}\mathbf{R}_s\mathbf{A}^H$ is singular (rank D), its determinant is equal to zero and, therefore, $\det\left[\mathbf{R}_x - \sigma_n^2\mathbf{I}\right] = 0$ and σ_n^2 is a (minimum) eigenvalue with multiplicity $M - D$

- Spectral decomposition of matrix $\mathbf{R}_x$: vector $\mathbf{e}_m$ being an eigenvector of $\mathbf{R}_x$ means that $\mathbf{R}_x\mathbf{e}_m = \lambda_m\mathbf{e}_m$. Collecting all eigenvectors in matrix $\mathbf{E}$, we may write $\mathbf{R}_x\mathbf{E} = \mathbf{E}\boldsymbol{\Lambda} = \left[\mathbf{e}_1 \cdots \mathbf{e}_M\right]$ diag $\left\{\left[\lambda_1 \cdots \lambda_M\right]\right\}$
  $\Rightarrow \mathbf{R}_x = \mathbf{E}\boldsymbol{\Lambda}\mathbf{E}^H$

- $\mathbf{R}_x = \mathbf{A}\mathbf{R}_s\mathbf{A}^H + \mathbf{R}_n$ with $D < M$ implies that $\mathbf{A}\mathbf{R}_s\mathbf{A}^H$ is singular (rank D), its determinant is equal to zero and, therefore, $\det\left[\mathbf{R}_x - \sigma_n^2\mathbf{I}\right] = 0$ and σ_n^2 is a (minimum) eigenvalue with multiplicity $M - D$

- Spectral decomposition of matrix $\mathbf{R}_x$: vector $\mathbf{e}_m$ being an eigenvector of $\mathbf{R}_x$ means that $\mathbf{R}_x\mathbf{e}_m = \lambda_m\mathbf{e}_m$. Collecting all eigenvectors in matrix $\mathbf{E}$, we may write $\mathbf{R}_x\mathbf{E} = \mathbf{E}\boldsymbol{\Lambda} = \left[\mathbf{e}_1 \cdots \mathbf{e}_M\right] \text{diag}\left\{\left[\lambda_1 \cdots \lambda_M\right]\right\}$ $\Rightarrow \mathbf{R}_x = \mathbf{E}\boldsymbol{\Lambda}\mathbf{E}^H$

- Dividing matrix $\mathbf{E}$ in two parts, the first D columns and the last $N = M - D$ columns, we have:
$$\mathbf{E} = [\underbrace{\mathbf{e}_1 \cdots \mathbf{e}_D}_{\mathbf{E}_S}\, \underbrace{\mathbf{e}_{D+1} \cdots \mathbf{e}_M}_{\mathbf{E}_N}] = [\mathbf{E}_S \;\; \mathbf{E}_N]$$

- Noting that $\mathbf{E}\mathbf{E}^H = \mathbf{I}$, we can write $\mathbf{E}_S\mathbf{E}_S^H + \mathbf{E}_N\mathbf{E}_N^H = \mathbf{I}$

- Noting that $\mathbf{E}\mathbf{E}^H = \mathbf{I}$, we can write $\mathbf{E}_S\mathbf{E}_S^H + \mathbf{E}_N\mathbf{E}_N^H = \mathbf{I}$
- The columns of $\mathbf{E}_S$ span the D-dimensional signal subspace while the columns of $\mathbf{E}_N$ span the N-dimensional noise subspace

- Noting that $\mathbf{E}\mathbf{E}^H = \mathbf{I}$, we can write $\mathbf{E}_S\mathbf{E}_S^H + \mathbf{E}_N\mathbf{E}_N^H = \mathbf{I}$
- The columns of $\mathbf{E}_S$ span the D-dimensional signal subspace while the columns of $\mathbf{E}_N$ span the N-dimensional noise subspace
- A vector in the signal subspace is a linear combination of the columns of $\mathbf{E}_S$. An example: $\sum_{d=1}^{D} x_d \mathbf{e}_d = \mathbf{E}_S\mathbf{x}, \mathbf{x} = [x_1 \cdots x_D]^T$

- Noting that $\mathbf{E}\mathbf{E}^H = \mathbf{I}$, we can write $\mathbf{E}_S\mathbf{E}_S^H + \mathbf{E}_N\mathbf{E}_N^H = \mathbf{I}$
- The columns of $\mathbf{E}_S$ span the D-dimensional signal subspace while the columns of $\mathbf{E}_N$ span the N-dimensional noise subspace
- A vector in the signal subspace is a linear combination of the columns of $\mathbf{E}_S$. An example: $\sum_{d=1}^{D} x_d\mathbf{e}_d = \mathbf{E}_S\mathbf{x}, \mathbf{x} = [x_1 \cdots x_D]^T$
- We can find the distance d from a vector $\mathbf{v}$ to the signal subspace $\mathbf{E}_S$ by obtaining $\mathbf{x}$ that minimizes $d = |\mathbf{v} - \mathbf{E}_S\mathbf{x}|$; the result is $d^2 = \mathbf{v}^H\mathbf{E}_N\mathbf{E}_N^H\mathbf{v}$

- The squared distance from vector $\mathbf{a}(\theta)$ to the signal subspace (spanned by $\mathbf{E}_S$) is $d^2 = \mathbf{a}^H(\theta)\mathbf{E}_N\mathbf{E}_N^H\mathbf{a}(\theta)$

- The squared distance from vector $\mathbf{a}(\theta)$ to the signal subspace (spanned by $\mathbf{E}_S$) is $d^2 = \mathbf{a}^H(\theta)\mathbf{E}_N\mathbf{E}_N^H\mathbf{a}(\theta)$
- When θ belongs to $\{\theta_1 \cdots \theta_D\}$, this distance should be close to zero

- The squared distance from vector $\mathbf{a}(\theta)$ to the signal subspace (spanned by $\mathbf{E}_S$) is $d^2 = \mathbf{a}^H(\theta)\mathbf{E}_N\mathbf{E}_N^H\mathbf{a}(\theta)$
- When θ belongs to $\{\theta_1 \cdots \theta_D\}$, this distance should be close to zero
- Its inverse will present peaks. In algorithm MUSIC, we estimate D from the eigenvalues of $\hat{\mathbf{R}}_x$; from its eigenvectors, we form $\mathbf{E}_S$ and $\mathbf{E}_N$, and by varying θ, we shall find peaks in the directions of θ_1 to θ_D in

$$P_{MUSIC}(\theta) = \frac{1}{d^2_{\mathbf{a}(\theta)}} = \frac{1}{\mathbf{a}^H(\theta)\mathbf{E}_N\mathbf{E}_N^H\mathbf{a}(\theta)}$$

MUSIC

- The squared distance from vector $\mathbf{a}(\theta)$ to the signal subspace (spanned by $\mathbf{E}_S$) is $d^2 = \mathbf{a}^H(\theta)\mathbf{E}_N\mathbf{E}_N^H\mathbf{a}(\theta)$
- When θ belongs to $\{\theta_1 \cdots \theta_D\}$, this distance should be close to zero
- Its inverse will present peaks. In algorithm MUSIC, we estimate D from the eigenvalues of $\hat{\mathbf{R}}_x$; from its eigenvectors, we form $\mathbf{E}_S$ and $\mathbf{E}_N$, and by varying θ, we shall find peaks in the directions of θ_1 to θ_D in

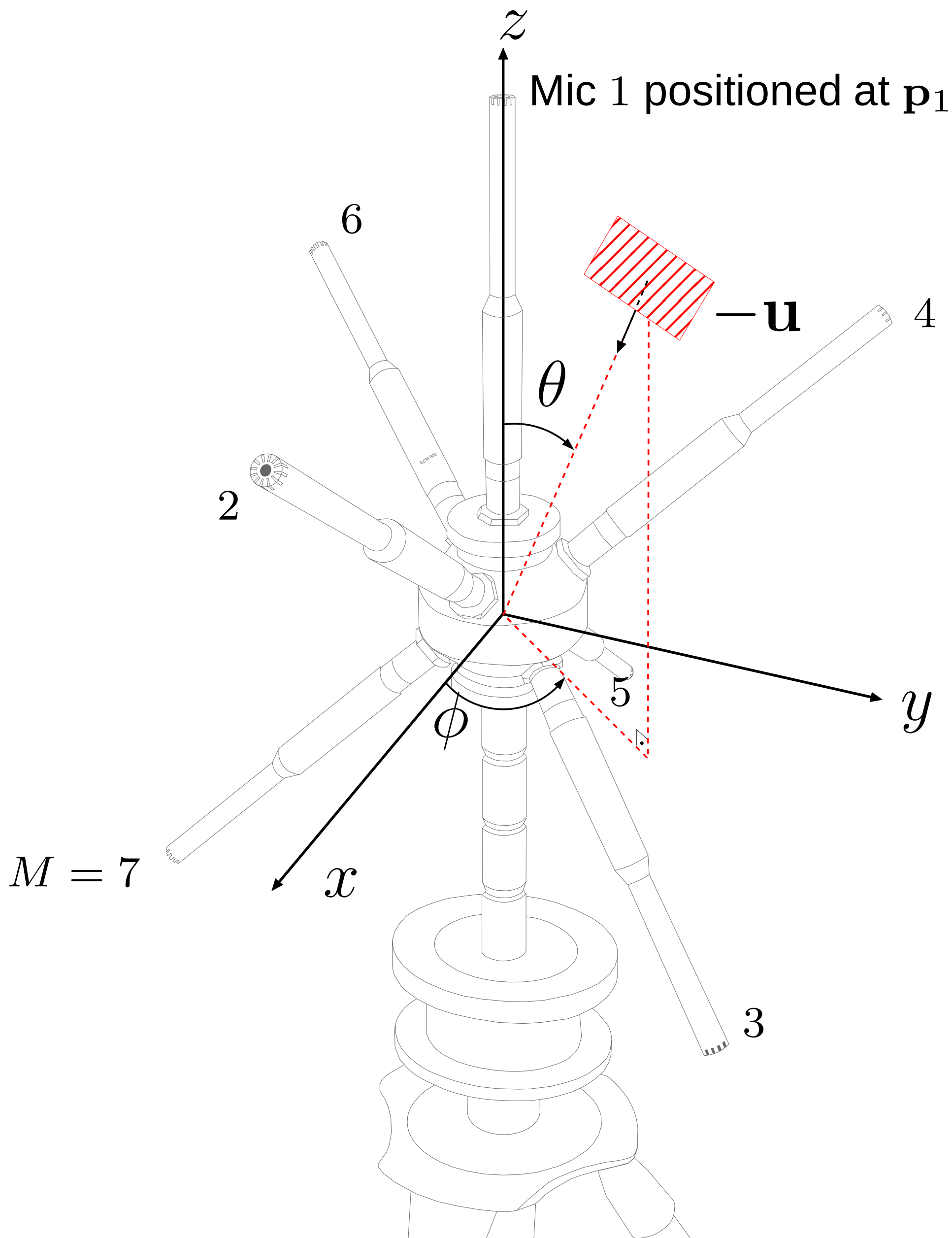
$$\boxed{P_{MUSIC}(\theta) = \frac{1}{d^2_{\mathbf{a}(\theta)}} = \frac{1}{\mathbf{a}^H(\theta)\mathbf{E}_N\mathbf{E}_N^H\mathbf{a}(\theta)}}$$

- If $\mathbf{R}_S$ is required, we compute

$$\mathbf{R}_S = \left(\mathbf{A}^H\mathbf{A}\right)^{-1}\mathbf{A}^H\left(\mathbf{R}_x - \sigma_n^2\mathbf{I}\right)\mathbf{A}\left(\mathbf{A}^H\mathbf{A}\right)^{-1}$$

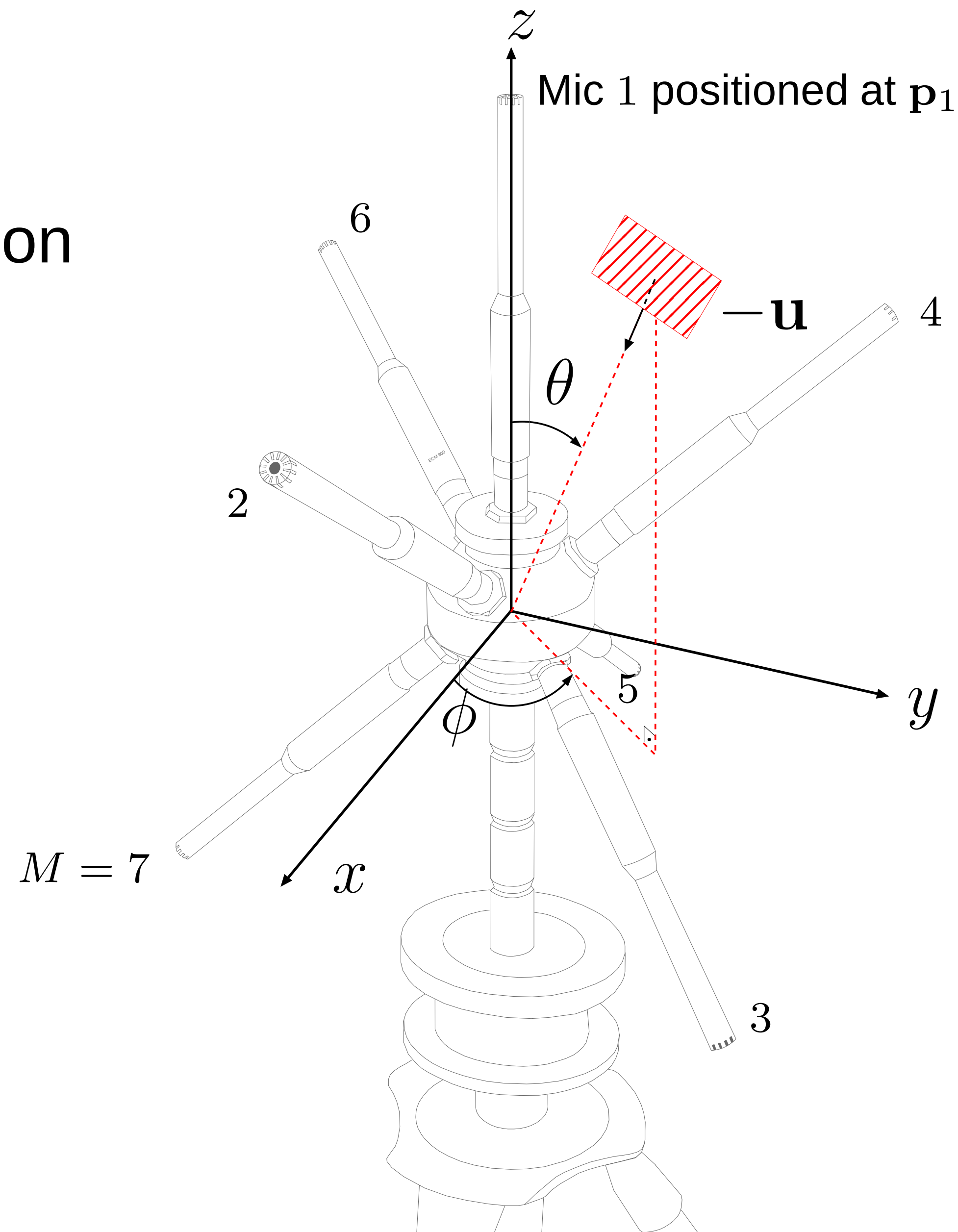
5.4 GCC-Based DoA

- M microphones of an array are in positions $\mathbf{p}_1$ to $\mathbf{p}_M$:

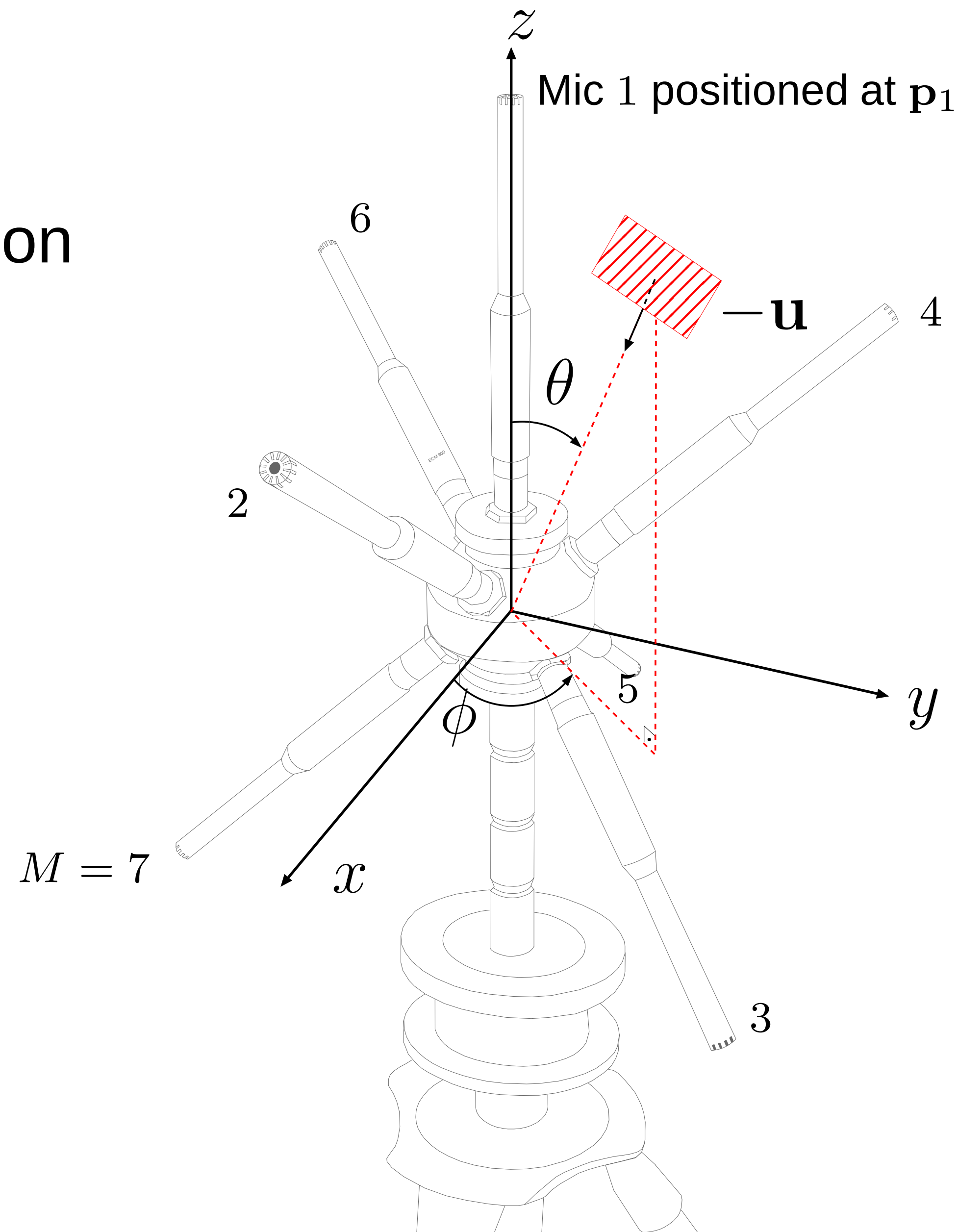


- M microphones of an array are in positions $\mathbf{p}_1$ to $\mathbf{p}_M$:

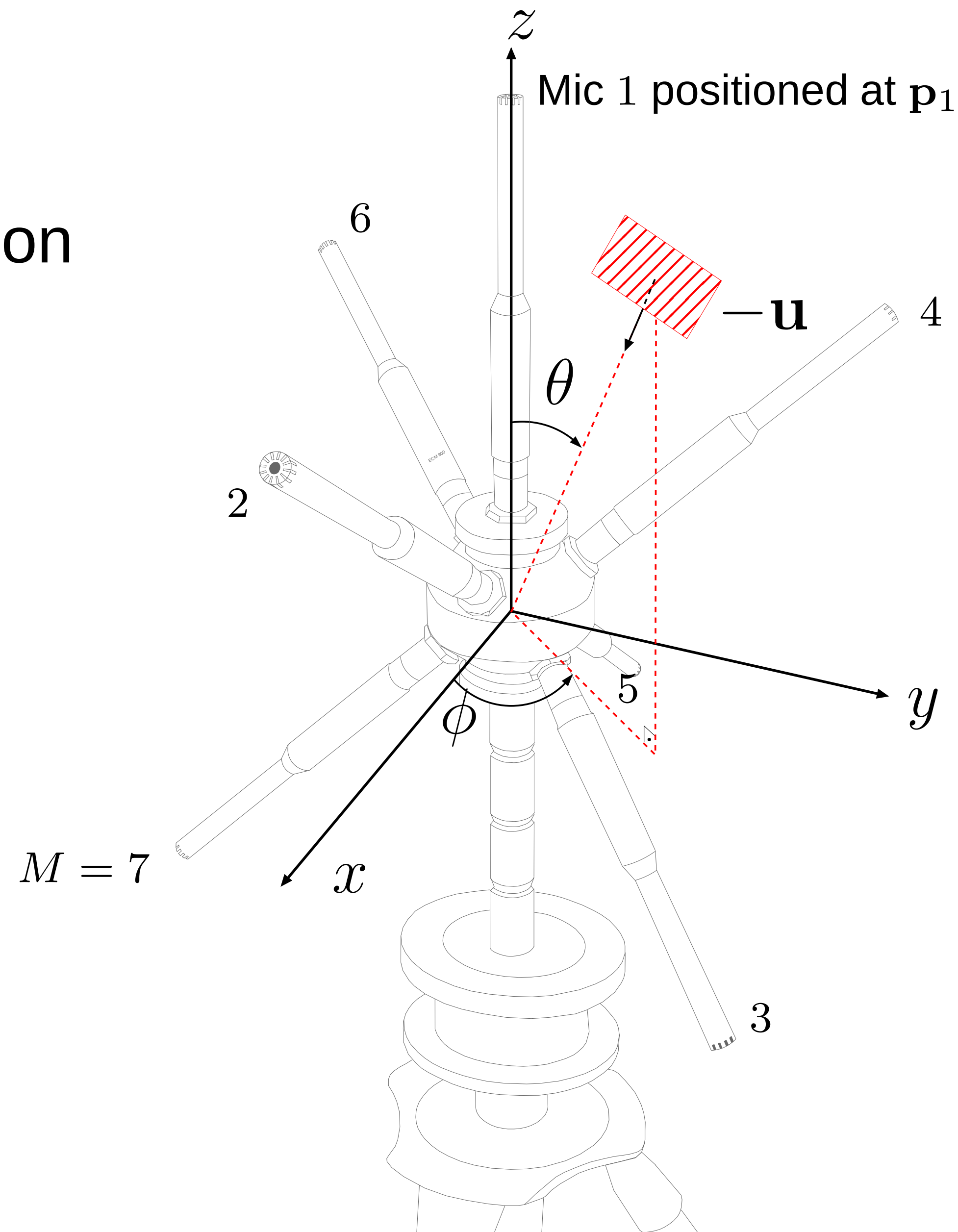
- $-\mathbf{u}$: unit vector in the direction of propagation



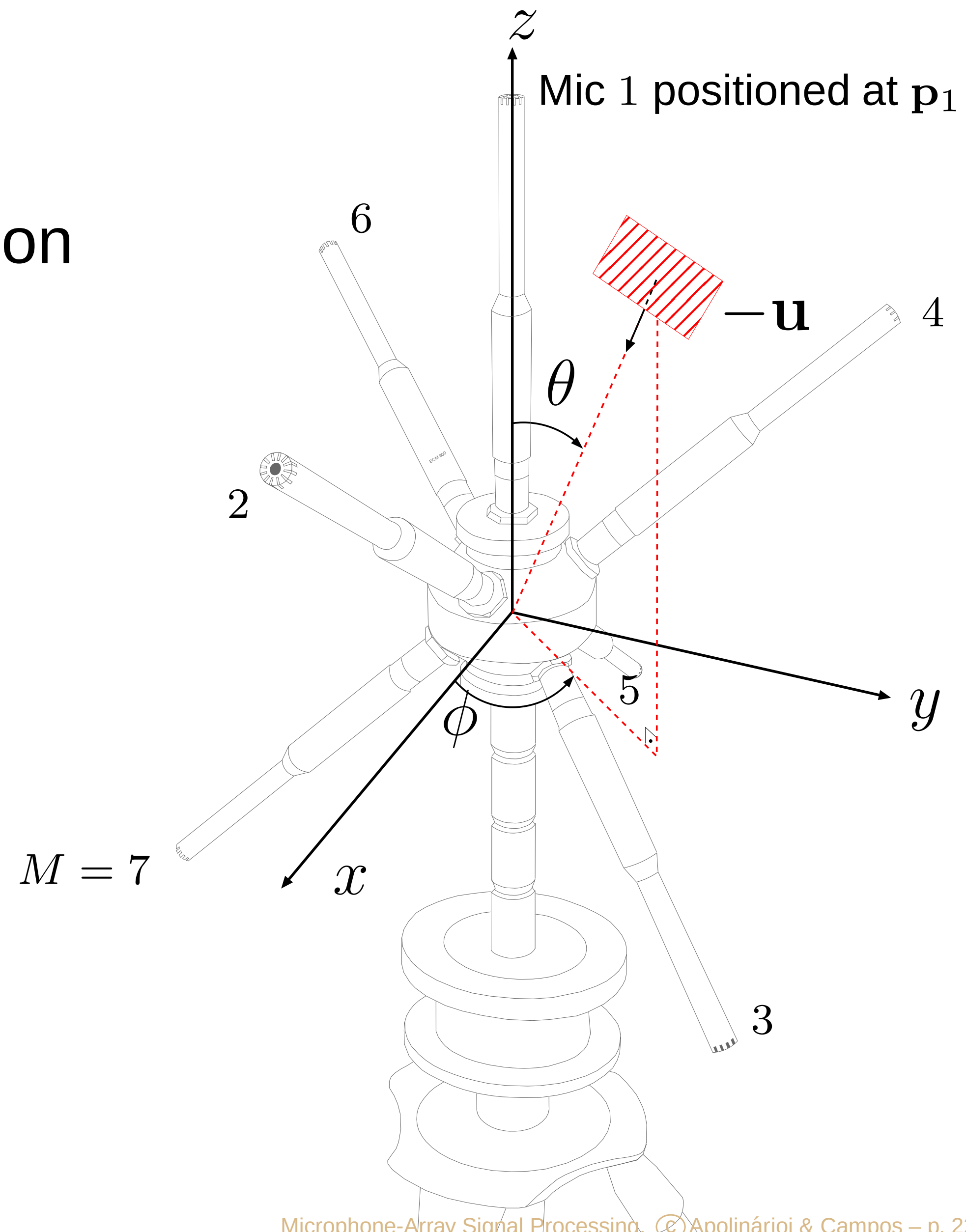
- M microphones of an array are in positions $\mathbf{p}_1$ to $\mathbf{p}_M$:
- $-\mathbf{u}$: unit vector in the direction of propagation
- θ: *grazing angle* ($\frac{\pi}{2}$ - elevation angle)



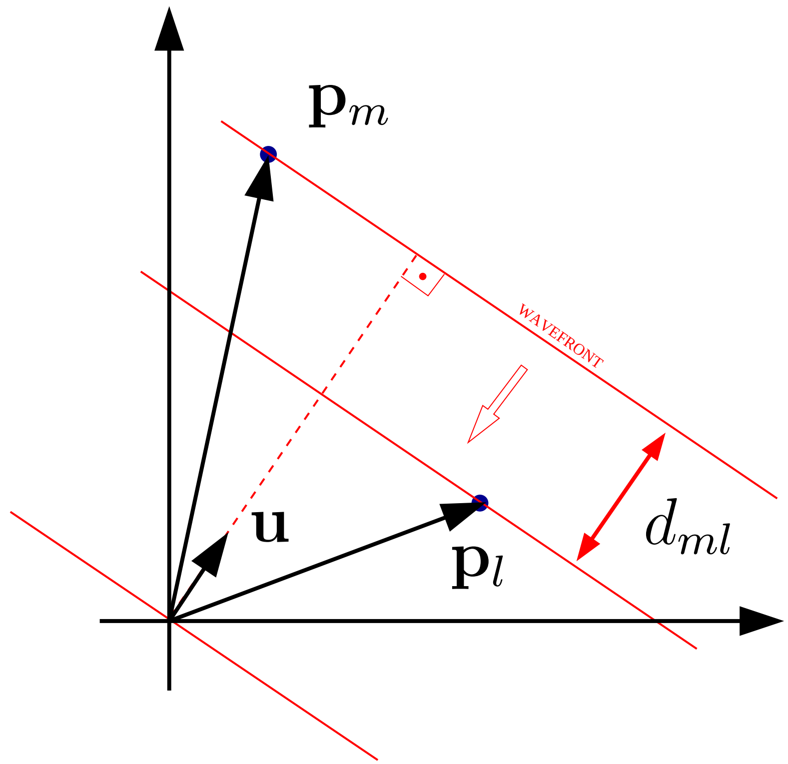
- M microphones of an array are in positions $\mathbf{p}_1$ to $\mathbf{p}_M$:
- $-\mathbf{u}$: unit vector in the direction of propagation
- θ: *grazing angle* ($\frac{\pi}{2}$ - elevation angle)
- ϕ: horizontal angle (*azimuth*)



- M microphones of an array are in positions $\mathbf{p}_1$ to $\mathbf{p}_M$:
- $-\mathbf{u}$: unit vector in the direction of propagation
- θ: *grazing angle* ($\frac{\pi}{2}$ - elevation angle)
- ϕ: horizontal angle (*azimuth*)
- $$\mathbf{u} = \begin{bmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{bmatrix}$$

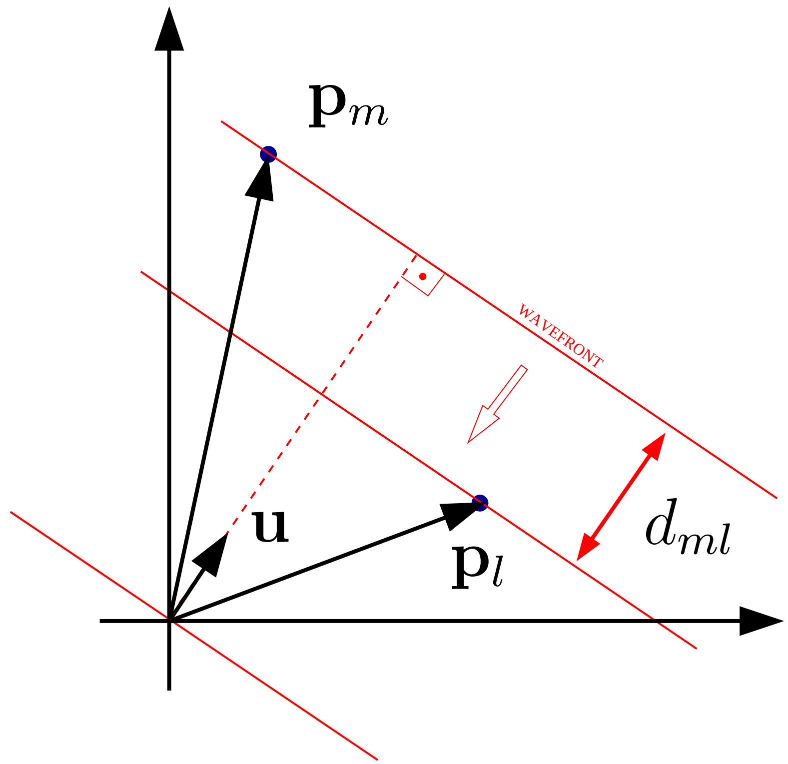


- We are interested in the TDoA between mics m and l

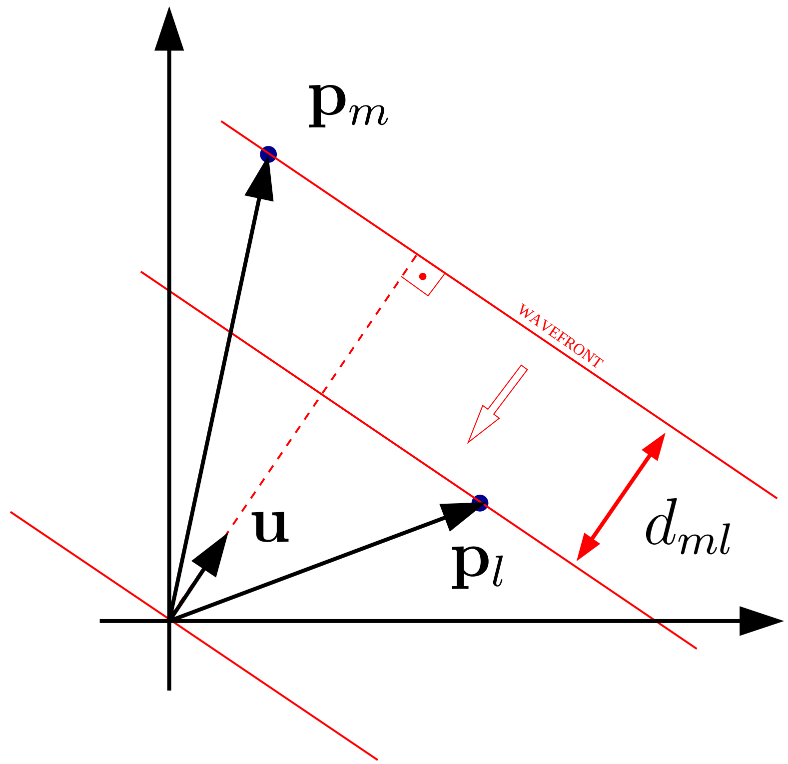


- We are interested in the TDoA between mics m and l

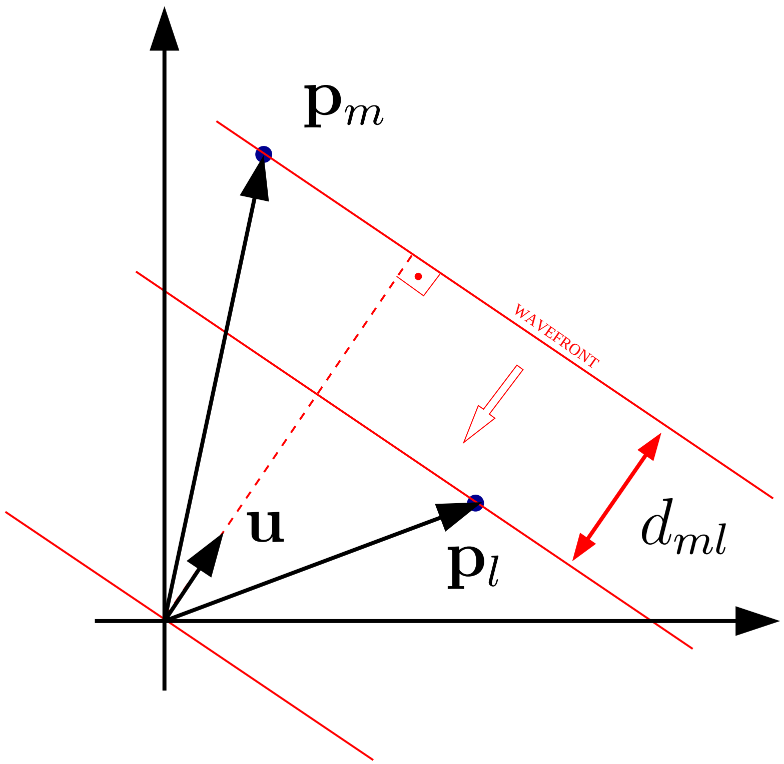
- Note that $d_{ml} = \mathbf{u}^T \underbrace{(\mathbf{p}_m - \mathbf{p}_l)}_{\Delta \mathbf{p}_{ml}}$



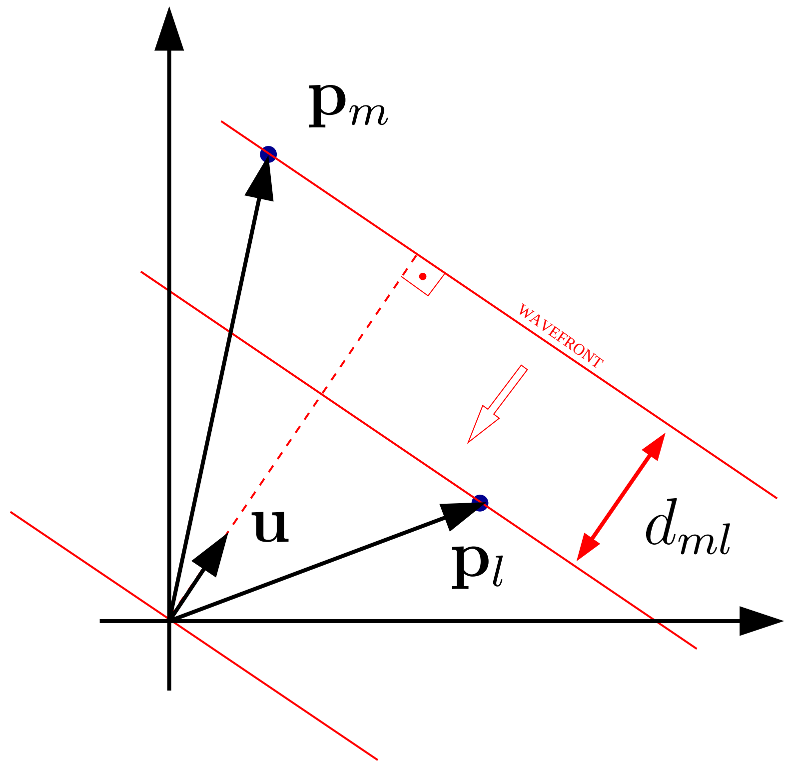
- We are interested in the TDoA between mics m and l
- Note that $d_{ml} = \mathbf{u}^T \underbrace{(\mathbf{p}_m - \mathbf{p}_l)}_{\Delta \mathbf{p}_{ml}}$
- TDoA:
$\bar{\tau}_{ml} = \frac{d_{ml}}{v_{som}} = \tau_{ml} T = \frac{\tau_{ml}}{f_s}$



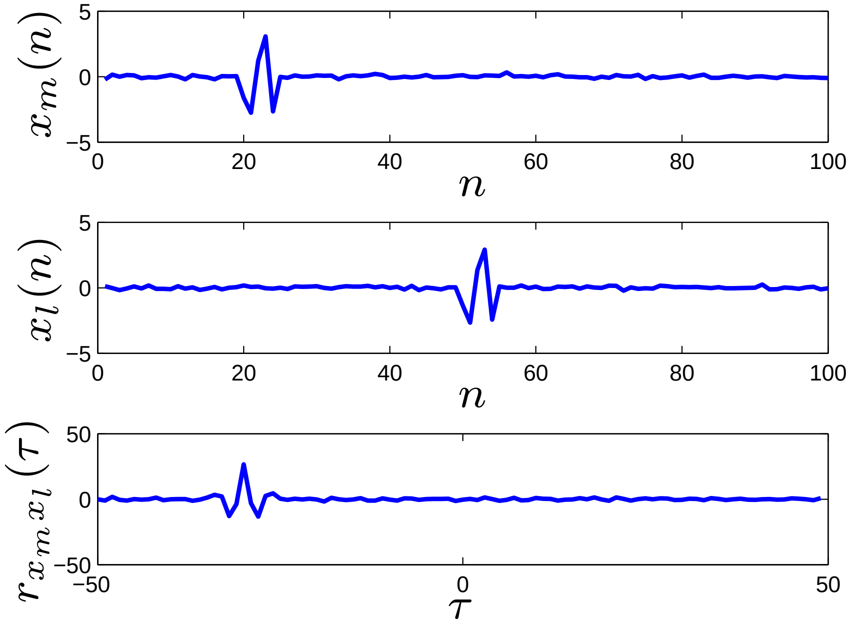
- We are interested in the TDoA between mics m and l
- Note that $d_{ml} = \mathbf{u}^T \underbrace{(\mathbf{p}_m - \mathbf{p}_l)}_{\Delta \mathbf{p}_{ml}}$
- TDoA:
  $\bar{\tau}_{ml} = \frac{d_{ml}}{v_{som}} = \tau_{ml} T = \frac{\tau_{ml}}{f_s}$
- τ_{ml} (in number of samples) is to be obtained from the peak of $\hat{r}_{x_m x_l}(\tau)$



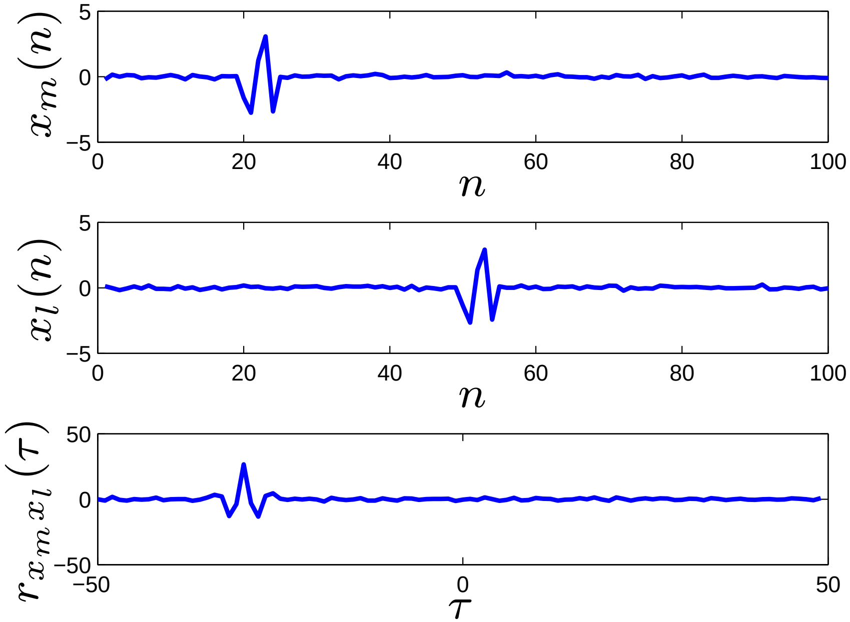
- We are interested in the TDoA between mics m and l
- Note that $d_{ml} = \mathbf{u}^T \underbrace{(\mathbf{p}_m - \mathbf{p}_l)}_{\Delta \mathbf{p}_{ml}}$
- TDoA:
  $\bar{\tau}_{ml} = \frac{d_{ml}}{v_{som}} = \tau_{ml} T = \frac{\tau_{ml}}{f_s}$
- τ_{ml} (in number of samples) is to be obtained from the peak of $\hat{r}_{x_m x_l}(\tau)$
- $r_{x_m x_l}(\tau) = E[x_m(n) x_l(n-\tau)]$



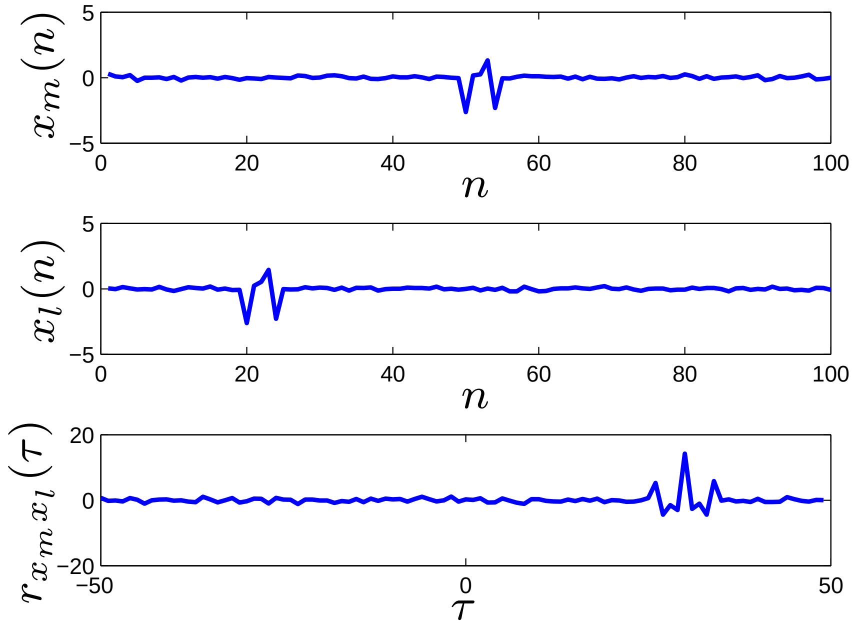
- When the sound frontwave first hits microphone m ($\tau_{ml} < 0$):



- When the sound frontwave first hits microphone m ($\tau_{ml} < 0$):



- When it first hits mic l ($\tau_{ml} > 0$):



GCC

- An estimate for the correlation can be given as:

$$\hat{r}_{x_m x_l}(\tau) = \sum_{-\infty}^{\infty} x_m(n) x_l(n-\tau) = x_m(\tau) * x_l(-\tau)$$

GCC

- An estimate for the correlation can be given as:
  $\hat{r}_{x_m x_l}(\tau) = \sum_{-\infty}^{\infty} x_m(n) x_l(n-\tau) = x_m(\tau) * x_l(-\tau)$
- The cross-power spectrum density (CPSD):
  $\hat{R}_{x_m x_l}(e^{j\omega}) = \mathcal{F}\{x_m(\tau) * x_l(-\tau)\} = X_m(e^{j\omega}) X_l(e^{-j\omega})$

GCC

- An estimate for the correlation can be given as:
  $\hat{r}_{x_m x_l}(\tau) = \sum_{-\infty}^{\infty} x_m(n) x_l(n-\tau) = x_m(\tau) * x_l(-\tau)$
- The cross-power spectrum density (CPSD):
  $\hat{R}_{x_m x_l}(e^{j\omega}) = \mathcal{F}\{x_m(\tau) * x_l(-\tau)\} = X_m(e^{j\omega}) X_l(e^{-j\omega})$
- We may assume the model
  $x_m(n) = s(n) * h_m(n) + n_m(n)$ and similarly for $x_l(n)$

GCC

- An estimate for the correlation can be given as:
  $\hat{r}_{x_m x_l}(\tau) = \sum_{-\infty}^{\infty} x_m(n) x_l(n-\tau) = x_m(\tau) * x_l(-\tau)$
- The cross-power spectrum density (CPSD):
  $\hat{R}_{x_m x_l}(e^{j\omega}) = \mathcal{F}\{x_m(\tau) * x_l(-\tau)\} = X_m(e^{j\omega}) X_l(e^{-j\omega})$
- We may assume the model
  $x_m(n) = s(n) * h_m(n) + n_m(n)$ and similarly for $x_l(n)$
- Hence, considering very small additive error and real sequences, we find
  $\hat{R}_{x_m x_l}(e^{j\omega}) \approx |S(e^{j\omega})|^2 H_m(e^{j\omega}) H_l^*(e^{j\omega})$ and
  $\hat{r}_{x_m x_l}(\tau) \approx \frac{1}{2\pi} \int_{-\pi}^{\pi} H_m(e^{j\omega}) H_l^*(e^{j\omega}) \hat{R}_s(e^{j\omega}) e^{j\omega\tau} d\omega$

GCC

- An estimate for the correlation can be given as:
$\hat{r}_{x_m x_l}(\tau) = \sum_{-\infty}^{\infty} x_m(n) x_l(n-\tau) = x_m(\tau) * x_l(-\tau)$
- The cross-power spectrum density (CPSD):
$\hat{R}_{x_m x_l}(e^{j\omega}) = \mathcal{F}\{x_m(\tau) * x_l(-\tau)\} = X_m(e^{j\omega}) X_l(e^{-j\omega})$
- We may assume the model
$x_m(n) = s(n) * h_m(n) + n_m(n)$ and similarly for $x_l(n)$
- Hence, considering very small additive error and real sequences, we find
$\hat{R}_{x_m x_l}(e^{j\omega}) \approx |S(e^{j\omega})|^2 H_m(e^{j\omega}) H_l^*(e^{j\omega})$ and
$\hat{r}_{x_m x_l}(\tau) \approx \frac{1}{2\pi} \int_{-\pi}^{\pi} H_m(e^{j\omega}) H_l^*(e^{j\omega}) \hat{R}_s(e^{j\omega}) e^{j\omega\tau} d\omega$
- Which motivates the GCC:

$$\boxed{r_{x_m x_l}^{G}(\tau) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \psi(\omega) \hat{R}_{x_m x_l}(e^{j\omega}) e^{j\omega\tau} d\omega}$$

Types of $\psi(\omega)$

- Classical cross-correlation:
  $\psi(\omega) = 1$

Types of $\psi(\omega)$

- Classical cross-correlation:

  $$\psi(\omega) = 1$$

- Maximum Likelihood (ML):

  $$\psi(\omega) = \frac{|X_m(e^{j\omega})||X_l(e^{j\omega})|}{\hat{R}_{n_n}(e^{j\omega})\hat{R}_{x_m}(e^{j\omega}) + \hat{R}_{n_l}(e^{j\omega})\hat{R}_{x_l}(e^{j\omega})}$$

Types of $\psi(\omega)$

- Classical cross-correlation:
  $\psi(\omega) = 1$

- Maximum Likelihood (ML):
  $$\psi(\omega) = \frac{|X_m(e^{j\omega})||X_l(e^{j\omega})|}{\hat{R}_{n_n}(e^{j\omega})\hat{R}_{x_m}(e^{j\omega}) + \hat{R}_{n_l}(e^{j\omega})\hat{R}_{x_l}(e^{j\omega})}$$
  - $\hat{R}_{x_m}(e^{j\omega}) = |X_m(e^{j\omega})|^2$
  - $\hat{R}_{x_l}(e^{j\omega}) = |X_l(e^{j\omega})|^2$
  - $\hat{R}_{n_m}(e^{j\omega}) = |N_m(e^{j\omega})|^2$ (estimated during silence interval)
  - $\hat{R}_{n_l}(e^{j\omega}) = |N_l(e^{j\omega})|^2$ (estimated during silence interval)

- PHAT (Phase Transform):
  $\psi(\omega) = \frac{1}{|\hat{R}_{x_m x_l}(e^{j\omega})}|$

- PHAT (Phase Transform):
  $\psi(\omega) = \frac{1}{|\hat{R}_{x_m x_l}(e^{j\omega})}|$

- Replacing this function in the expression of $r^{G}_{x_m x_l}(\tau)$:
  $r^{PHAT}_{x_m x_l}(\tau) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\hat{R}_{x_m x_l}(e^{j\omega})}{|\hat{R}_{x_m x_l}(e^{j\omega})|} e^{j\omega\tau} d\omega$ in which,
  after making $\hat{R}_{x_m x_l}(e^{j\omega}) = |S(e^{j\omega})|^2 H_m(e^{j\omega}) H_l^*(e^{j\omega})$,
  we have $r^{PHAT}_{x_m x_l}(\tau) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{j(\sphericalangle H_m - \sphericalangle H_l + \omega\pi)} d\omega$

- PHAT (Phase Transform):
  $\psi(\omega) = \frac{1}{|\hat{R}_{x_m x_l}(e^{j\omega})|}$

- Replacing this function in the expression of $r^G_{x_m x_l}(\tau)$:
  $r^{PHAT}_{x_m x_l}(\tau) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\hat{R}_{x_m x_l}(e^{j\omega})}{|\hat{R}_{x_m x_l}(e^{j\omega})|} e^{j\omega\tau} d\omega$ in which,
  after making $\hat{R}_{x_m x_l}(e^{j\omega}) = |S(e^{j\omega})|^2 H_m(e^{j\omega}) H_l^*(e^{j\omega})$,
  we have $r^{PHAT}_{x_m x_l}(\tau) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{j(\sphericalangle H_m - \sphericalangle H_l + \omega\pi)} d\omega$

- For the PHAT, in case of having
  $h_m(n) = \alpha_m \delta(n)$ and $h_l(n) = \alpha_l \delta(n - \Delta\tau)$,
  the cross-correlation would be
  $r^{PHAT}_{x_m x_l}(\tau) = \delta(\tau + \Delta\tau) \Rightarrow$ peak in $\tau_{ml} = -\Delta\tau$
  (a perfect indication of a temporal delay!)

LS solution

- Assuming we have all possible ($M(M-1)/2$) delays τ_{ml}, we want angles ϕ and θ

LS solution

- Assuming we have all possible ($M(M-1)/2$) delays τ_{ml}, we want angles ϕ and θ
- We define a cost function:

$$\xi = \left(\bar{\tau}_{12} - \Delta\mathbf{p}_{12}^T\mathbf{u}\right)^2 + \cdots + \left(\bar{\tau}_{(M-1)M} - \Delta\mathbf{p}_{(M-1)M}^T\mathbf{u}\right)^2$$

with $\bar{\tau}_{ml} = \tau_{ml}/f_s$

LS solution

- Assuming we have all possible ($M(M-1)/2$) delays τ_{ml}, we want angles ϕ and θ
- We define a cost function:
  $$\xi = \left(\bar{\tau}_{12} - \Delta\mathbf{p}_{12}^T\mathbf{u}\right)^2 + \cdots + \left(\bar{\tau}_{(M-1)M} - \Delta\mathbf{p}_{(M-1)M}^T\mathbf{u}\right)^2$$
  with $\bar{\tau}_{ml} = \tau_{ml}/f_s$
- We then find $\mathbf{u}$ that minimizes ξ by making $\nabla_{\mathbf{u}}\xi = \mathbf{0}$:

$$\mathbf{A}\mathbf{u} = \mathbf{b}$$

where $\mathbf{A} = \Delta\mathbf{p}_{12}\Delta\mathbf{p}_{12}^T + \cdots + \Delta\mathbf{p}_{(M-1)M}\Delta\mathbf{p}_{(M-1)M}^T$
and $\mathbf{b} = \bar{\tau}_{12}\Delta\mathbf{p}_{12} + \cdots + \bar{\tau}_{(M-1)M}\Delta\mathbf{p}_{(M-1)M}$

- Assuming we have all possible $(M(M-1)/2)$ delays τ_{ml}, we want angles ϕ and θ
- We define a cost function:
  $$\xi = \left(\bar{\tau}_{12} - \Delta\mathbf{p}_{12}^T\mathbf{u}\right)^2 + \cdots + \left(\bar{\tau}_{(M-1)M} - \Delta\mathbf{p}_{(M-1)M}^T\mathbf{u}\right)^2$$
  with $\bar{\tau}_{ml} = \tau_{ml}/f_s$
- We then find $\mathbf{u}$ that minimizes ξ by making $\nabla_{\mathbf{u}}\xi = \mathbf{0}$:
  $$\mathbf{A}\mathbf{u} = \mathbf{b}$$
  where $\mathbf{A} = \Delta\mathbf{p}_{12}\Delta\mathbf{p}_{12}^T + \cdots + \Delta\mathbf{p}_{(M-1)M}\Delta\mathbf{p}_{(M-1)M}^T$
  and $\mathbf{b} = \bar{\tau}_{12}\Delta\mathbf{p}_{12} + \cdots + \bar{\tau}_{(M-1)M}\Delta\mathbf{p}_{(M-1)M}$
- And this unit vector is given as $\mathbf{u} = \begin{bmatrix} u_x \\ u_y \\ u_z \end{bmatrix} = \mathbf{A}^{-1}\mathbf{b}$

- Knowing $\mathbf{u}$ and also the fact that it corresponds to

$$\begin{bmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{bmatrix}, \cdots$$

- Knowing $\mathbf{u}$ and also the fact that it corresponds to

$$\begin{bmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{bmatrix}, \cdots$$

- $\cdots$ we compute the azimuth:

$$\boxed{\phi = \arctan\frac{u_y}{u_x}}$$

- Knowing $\mathbf{u}$ and also the fact that it corresponds to

$$\begin{bmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{bmatrix}, \cdots$$

- $\cdots$ we compute the azimuth:

$$\boxed{\phi = \arctan\frac{u_y}{u_x}}$$

- And the elevation:

$$\boxed{\text{elevation} = 90^\circ - \theta = 90^\circ - \arccos u_z}$$

Thank you!