Matrix Analysis and Applications

Chapter 3: Applications of Eigendecomposition

Instructor: Kai Lu

(http://seit.sysu.edu.cn/teacher/1801)

School of Electronics and Information Technology
Sun Yat-sen University

November 18, 2020

Table of Contents

Table of Contents

Preliminaries

Random processes: consider a scalar random process $\{x_n\}$, where $n \in \mathbb{Z}$ ($\mathbb{Z}$ is the set of integers). Let

$$\mu_x(n) \triangleq \mathbb{E}\{x_n\}, \tag{1}$$

$$r_x(n,l) \triangleq \mathbb{E}\{x_n x_l^*\}, \tag{2}$$

$$c_x(n,l) \triangleq \mathbb{E}\left\{(x_n - \mathbb{E}\{x_n\})(x_l - \mathbb{E}\{x_l\})^*\right\}, \tag{3}$$

denote the mean, auto-correlation function and auto-covariance function of x_n, respectively, where $\mathbb{E}\{\cdot\}$ denotes the expectation.

Preliminaries

Random processes: consider a scalar random process $\{x_n\}$, where $n \in \mathbb{Z}$ ($\mathbb{Z}$ is the set of integers). Let

$$\mu_x(n) \triangleq \mathbb{E}\{x_n\}, \tag{1}$$

$$r_x(n,l) \triangleq \mathbb{E}\{x_n x_l^*\}, \tag{2}$$

$$c_x(n,l) \triangleq \mathbb{E}\left\{(x_n - \mathbb{E}\{x_n\})(x_l - \mathbb{E}\{x_l\})^*\right\}, \tag{3}$$

denote the mean, auto-correlation function and auto-covariance function of x_n, respectively, where $\mathbb{E}\{\cdot\}$ denotes the expectation.

Examples:

- **Discrete-time signal, time series**: x_n is a signal sample or observation at time indicated by n. The functions $r_x(n,l)$ and $c_x(n,l)$ measure how two signal samples at time n, l are related.
- **Image**: x_n stores the value of a pixel, with the pixel's spatial location indicated by n. The functions $r_x(n,l)$ and $c_x(n,l)$ measure the spatial correlations.
- **Data analysis**: x_n is a data point (e.g., score of student n, preference of user n).

Preliminaries (cont'd)

Wide-sense stationary processes: a random process $\{x_n\}$ is called wide-sense stationary (WSS) if, for any n and l, we have

$$\mu_x(n) = \mu_x(l), \tag{4}$$

$$c_x(n, l) = c_x(n + i, l + i). \tag{5}$$

Preliminaries (cont'd)

Wide-sense stationary processes: a random process $\{x_n\}$ is called wide-sense stationary (WSS) if, for any n and l, we have

$$\mu_x(n) = \mu_x(l), \tag{4}$$

$$c_x(n,l) = c_x(n+i, l+i). \tag{5}$$

- A WSS process is a random process whose first-order and second-order statistics do not change with n (e.g., time in discrete-time signals).
- A zero-mean random process $\{x_n\}$, i.e., $\mu_x(n) = 0$ for all n, is WSS if

$$r_x(n,l) = r_x(n+i, l+i) \quad \text{for any } i.$$

- The same concept applies to a vector random process $\{\boldsymbol{x}_n\}$.

Preliminaries (cont'd)

Let $\{\boldsymbol{x}_j\}$, $\boldsymbol{x}_j \in \mathbb{C}^m$, be a zero-mean complex-valued WSS process.

Correlation matrix:

$$\boldsymbol{R}_{\boldsymbol{x}} \triangleq \mathbb{E}\{\boldsymbol{x}_j \boldsymbol{x}_j^H\} = \begin{bmatrix} \mathbb{E}\{|x_{1j}|^2\} & \mathbb{E}\{x_{1j}x_{2j}^*\} & \cdots & \mathbb{E}\{x_{1j}x_{mj}^*\} \\ \mathbb{E}\{x_{2i}x_{1j}^*\} & \mathbb{E}\{|x_{2j}|^2\} & \cdots & \mathbb{E}\{x_{2j}x_{mj}^*\} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}\{x_{mj}x_{1j}^*\} & \cdots & \cdots & \mathbb{E}\{|x_{mj}|^2\} \end{bmatrix}. \quad (6)$$

- Properties:

- $\boldsymbol{R}_{\boldsymbol{x}}$ is Hermitian (and symmetric for the real-valued case).
- $\boldsymbol{R}_{\boldsymbol{x}}$ is positive semidefinite (will be discussed in later lectures).

Representation by Orthonormal Expansion

For a WSS process $\{\boldsymbol{x}_j\}$, $\boldsymbol{x}_j \in \mathbb{C}^m$.

Aim: given a basis $\{\boldsymbol{q}_1, \cdots, \boldsymbol{q}_m\} \subseteq \mathbb{C}^m$, represent each $\boldsymbol{x}_j$ by

$$\boldsymbol{x}_n = \sum_{i=1}^{m} a_{ij}\boldsymbol{q}_i = \boldsymbol{Q}\boldsymbol{a}_j, \quad \text{for some coefficients } a_{ij},\ i = 1, \cdots, m, \tag{7}$$

where $\boldsymbol{a}_j = [a_{1j}, \cdots, a_{mj}]^T$ and $\boldsymbol{Q} = [\boldsymbol{q}_1, \cdots, \boldsymbol{q}_m]$.

Representation by Orthonormal Expansion

For a WSS process $\{\boldsymbol{x}_j\}$, $\boldsymbol{x}_j \in \mathbb{C}^m$.

Aim: given a basis $\{\boldsymbol{q}_1, \cdots, \boldsymbol{q}_m\} \subseteq \mathbb{C}^m$, represent each $\boldsymbol{x}_j$ by

$$\boldsymbol{x}_n = \sum_{i=1}^{m} a_{ij}\boldsymbol{q}_i = \boldsymbol{Q}\boldsymbol{a}_j, \quad \text{for some coefficients } a_{ij},\ i = 1, \cdots, m, \tag{7}$$

where $\boldsymbol{a}_j = [a_{1j}, \cdots, a_{mj}]^T$ and $\boldsymbol{Q} = [\boldsymbol{q}_1, \cdots, \boldsymbol{q}_m]$.

Orthonormal expansion: perform (7) via an orthonormal basis, or unitary $\boldsymbol{Q}$.

- Examples: discrete Fourier transform (DFT), discrete cosine transform (DCT), Haar transform.
- For orthonormal expansion, each $\boldsymbol{a}_j$ can be conveniently obtained via

$$\boldsymbol{a}_j = \boldsymbol{Q}^H \boldsymbol{x}_j. \tag{8}$$

Example: Discrete Fourier Transform (DFT)

The DFT basis matrix is $\boldsymbol{Q} = [\boldsymbol{q}_0, \boldsymbol{q}_1, \cdots, \boldsymbol{q}_{m-1}]$, where

$$\boldsymbol{q}_i = \frac{1}{\sqrt{m}} \begin{bmatrix} 1 \\ e^{\frac{j2\pi i}{m}} \\ e^{\frac{j4\pi i}{m}} \\ \vdots \\ e^{\frac{j2\pi(m-1)i}{m}} \end{bmatrix}, \quad i = 0, 1, \cdots, m-1 \tag{9}$$

and $j = \sqrt{-1}$.

Example: Discrete Fourier Transform (DFT)

The DFT basis matrix is $\boldsymbol{Q} = [\boldsymbol{q}_0, \boldsymbol{q}_1, \cdots, \boldsymbol{q}_{m-1}]$, where

$$\boldsymbol{q}_i = \frac{1}{\sqrt{m}} \begin{bmatrix} 1 \\ e^{\frac{j2\pi i}{m}} \\ e^{\frac{j4\pi i}{m}} \\ \vdots \\ e^{\frac{j2\pi(m-1)i}{m}} \end{bmatrix}, \quad i = 0, 1, \cdots, m-1 \tag{9}$$

and $j = \sqrt{-1}$.

Exercise: verify that $\|\boldsymbol{q}_i\|_2^2 = 1$, and $\boldsymbol{q}_i^H \boldsymbol{q}_k = 0$ for all $i \neq k$. Also, if we write $\boldsymbol{Q}^H = [\boldsymbol{u}_0, \cdots, \boldsymbol{u}_{m-1}]$, verify that

$$\boldsymbol{u}_i = \frac{1}{\sqrt{m}} \begin{bmatrix} 1 \\ e^{-\frac{j2\pi i}{m}} \\ e^{-\frac{j4\pi i}{m}} \\ \vdots \\ e^{-\frac{j2\pi(m-1)i}{m}} \end{bmatrix}. \tag{10}$$

Karhunen-Loéve Expansion

Assume a zero-mean WSS $\{\boldsymbol{x}_n\}$. From $\boldsymbol{a}_n = \boldsymbol{Q}^H \boldsymbol{x}_n$, one gets

$$
\begin{aligned}
\boldsymbol{R}_{\boldsymbol{a}} &\triangleq \mathbb{E}\{\boldsymbol{a}_n \boldsymbol{a}_n^H\} && (11)\\
&= \mathbb{E}\{\boldsymbol{Q}^H \boldsymbol{x}_n \boldsymbol{x}_n^H \boldsymbol{Q}\} && (12)\\
&= \boldsymbol{Q}^H \mathbb{E}\{\boldsymbol{x}_n \boldsymbol{x}_n^H\} \boldsymbol{Q} && (13)\\
&= \boldsymbol{Q}^H \boldsymbol{R}_{\boldsymbol{x}} \boldsymbol{Q}. && (14)
\end{aligned}
$$

Karhunen-Loéve Expansion

Assume a zero-mean WSS $\{\boldsymbol{x}_n\}$. From $\boldsymbol{a}_n = \boldsymbol{Q}^H \boldsymbol{x}_n$, one gets

$$\begin{aligned}
\boldsymbol{R}_{\boldsymbol{a}} &\triangleq \mathbb{E}\{\boldsymbol{a}_n \boldsymbol{a}_n^H\} && (11)\\
&= \mathbb{E}\{\boldsymbol{Q}^H \boldsymbol{x}_n \boldsymbol{x}_n^H \boldsymbol{Q}\} && (12)\\
&= \boldsymbol{Q}^H \mathbb{E}\{\boldsymbol{x}_n \boldsymbol{x}_n^H\} \boldsymbol{Q} && (13)\\
&= \boldsymbol{Q}^H \boldsymbol{R}_{\boldsymbol{x}} \boldsymbol{Q}. && (14)
\end{aligned}$$

Problem: find a unitary basis matrix $\boldsymbol{Q}$ such that $\boldsymbol{R}_{\boldsymbol{a}}$ is diagonal.

- Having a diagonal $\boldsymbol{R}_{\boldsymbol{a}}$ means that the elements of $\boldsymbol{a}_n$ are mutually uncorrelated, which is good for applications such as compression.

Karhunen-Loéve Expansion

Assume a zero-mean WSS $\{\boldsymbol{x}_n\}$. From $\boldsymbol{a}_n = \boldsymbol{Q}^H \boldsymbol{x}_n$, one gets

$$\begin{aligned} \boldsymbol{R}_{\boldsymbol{a}} &\triangleq \mathbb{E}\{\boldsymbol{a}_n \boldsymbol{a}_n^H\} && (11) \\ &= \mathbb{E}\{\boldsymbol{Q}^H \boldsymbol{x}_n \boldsymbol{x}_n^H \boldsymbol{Q}\} && (12) \\ &= \boldsymbol{Q}^H \mathbb{E}\{\boldsymbol{x}_n \boldsymbol{x}_n^H\} \boldsymbol{Q} && (13) \\ &= \boldsymbol{Q}^H \boldsymbol{R}_{\boldsymbol{x}} \boldsymbol{Q}. && (14) \end{aligned}$$

Problem: find a unitary basis matrix $\boldsymbol{Q}$ such that $\boldsymbol{R}_{\boldsymbol{a}}$ is diagonal.

- Having a diagonal $\boldsymbol{R}_{\boldsymbol{a}}$ means that the elements of $\boldsymbol{a}_n$ are mutually uncorrelated, which is good for applications such as compression.

Karhunen-Loéve (KL) expansion: set $\boldsymbol{Q} = \boldsymbol{V}$, where $\boldsymbol{V}$ is obtained from the eigendecomposition $\boldsymbol{R}_{\boldsymbol{x}} = \boldsymbol{V} \boldsymbol{\Lambda} \boldsymbol{V}^H$.

- Under KL expansion, $\boldsymbol{R}_{\boldsymbol{a}} = \boldsymbol{\Lambda}$ and $\boldsymbol{R}_{\boldsymbol{a}}$ is diagonal.

More About KL Expansion

Since $\boldsymbol{R_a} = \boldsymbol{\Lambda}$, we have

$$\boldsymbol{R_a} = \begin{bmatrix} \mathbb{E}\{|a_{1n}|^2\} & & & \\ & \mathbb{E}\{|a_{2n}|^2\} & & \\ & & \ddots & \\ & & & \mathbb{E}\{|a_{mn}|^2\} \end{bmatrix} = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_m \end{bmatrix}, \tag{15}$$

or simply

$$\mathbb{E}\{|a_{in}|^2\} = \lambda_i, \ i = 1, \cdots, m. \tag{16}$$

More About KL Expansion

Since $\boldsymbol{R_a} = \boldsymbol{\Lambda}$, we have

$$\boldsymbol{R_a} = \begin{bmatrix} \mathbb{E}\{|a_{1n}|^2\} & & & \\ & \mathbb{E}\{|a_{2n}|^2\} & & \\ & & \ddots & \\ & & & \mathbb{E}\{|a_{mn}|^2\} \end{bmatrix} = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_m \end{bmatrix}, \quad (15)$$

or simply

$$\mathbb{E}\{|a_{in}|^2\} = \lambda_i, \ i = 1, \cdots, m. \quad (16)$$

- The eigenvalues are also the average energies of the KL coefficients.
- There are many situations where the energy in the first few KL coefficients a_{in} dominates that in the remaining ones – which suggests the possibility of approximating $\{\boldsymbol{x}_n\}$ using fewer number of KL coefficients.

Application Example: Coding and Compression

Problem: encode $\{\boldsymbol{a}_n\}$ with a (much) smaller data size than the original $\{\boldsymbol{x}_n\}$; lossy coding is allowed.

Application Example: Coding and Compression

Problem: encode $\{\boldsymbol{a}_n\}$ with a (much) smaller data size than the original $\{\boldsymbol{x}_n\}$; lossy coding is allowed.

Transform coding (for generic bases): given a unitary basis matrix $\boldsymbol{Q}$,

1) Transform each $\boldsymbol{x}_n$ to $\boldsymbol{a}_n = \boldsymbol{Q}^H \boldsymbol{x}_n$;
2) Select a subset of coefficients in $\boldsymbol{a}_n$ to encode, e.g., by selecting a number of r coefficients whose magnitudes are the largest;
   - The selected coefficients may also be quantized.
3) (Decoding) Reconstruct $\boldsymbol{x}_n$ approximately by $\hat{\boldsymbol{x}}_n = \sum_{i \in \mathcal{I}_n} \hat{a}_{in} \boldsymbol{q}_i$, where $\{\hat{a}_{in}\}_{n \in \mathcal{I}_n}$ and $\mathcal{I}_n \subseteq \{1, \cdots, m\}$ are the selected encoded coefficients and its index set, respectively.

KL transform (KLT): choose $\boldsymbol{Q} = \boldsymbol{V}$.

Remarks

1) To perform KL expansion, information of $\boldsymbol{R}_{\boldsymbol{x}}$ is required. One way is to estimate $\boldsymbol{R}_{\boldsymbol{x}}$ from data: given a number of N samples $\{\boldsymbol{x}_n\}_{n=1}^{N}$, compute the sampled correlation matrix

$$\hat{\boldsymbol{R}}_{\boldsymbol{x}} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}_n \boldsymbol{x}_n^H. \tag{17}$$

2) Consider the context of coding
   - While the decoder may not require information of $\boldsymbol{R}_{\boldsymbol{x}}$, it still requires information of the eigen-basis vectors $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n$.
   - We can encode $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n$ too, but that may incur overhead in data size.
   - On the other hand, one may apply fixed models for $\boldsymbol{R}_{\boldsymbol{x}}$. It has been found that for a certain class of correlation models, DCT is, or well approximates, KLT.

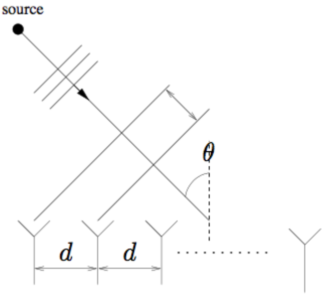
3) KLT is almost the same as **principal component analysis (PCA)**, which is extensively used for data analysis and dimensionality reduction.

Table of Contents

Sensor Array Processing

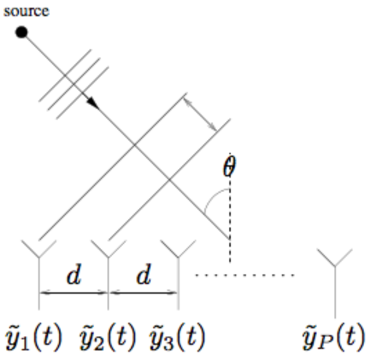
Sensor array processing deals with signal processing problems arising from the use of an array of sensors.

Applications: radar, sonar, wireless communications, seismology, radio astronomy, audio and speech,...



A key problem in this context is **direction-of-arrival (DoA) estimation** of sources.

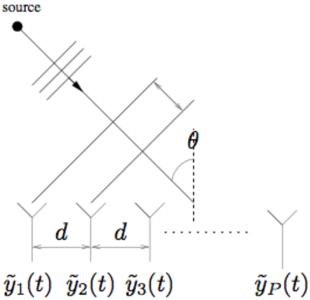
DoA Estimation: Preliminaries



Settings:

- Uniform linear array (ULA) with inter-sensor distance d and with P sensors,
- Far-field wave propagation so that the arriving source waves are planar.

DoA Estimation: Preliminaries (cont'd)



For one radiating source, the output of sensor p can be modeled as

$$\tilde{y}_p(t) = x\left(t - (p-1)\frac{d\sin\theta}{c}\right), \tag{18}$$

where $x(t)$ is the source signal impinging on sensor 1, $\theta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ denotes the source's DoA, and $c = 3 \times 10^8$ m/s refers to the wave propagation velocity.

DoA Estimation: Preliminaries (cont'd)

Assumption: narrowband carrier-modulated source waveform

$$x(t) = e^{j\omega_c t} s(t), \tag{19}$$

where ω_c is the center frequency, and $s(t)$ is narrowband in the sense that it satisfies

$$s(t-\tau) \simeq s(t), \quad \text{for any } \tau \in [-T_0, T_0] \tag{20}$$

and for some time duration T_0.

DoA Estimation: Preliminaries (cont'd)

Assumption: narrowband carrier-modulated source waveform

$$x(t) = e^{j\omega_c t} s(t), \tag{19}$$

where ω_c is the center frequency, and $s(t)$ is narrowband in the sense that it satisfies

$$s(t-\tau) \simeq s(t), \quad \text{for any } \tau \in [-T_0, T_0] \tag{20}$$

and for some time duration T_0.

Let $y_p(t) = e^{-j\omega_c t}\tilde{y}_p(t)$ be a demodulated signal for sensor p. Assuming $T_0 > Pd/c$, we get

$$\begin{aligned} y_p(t) &= e^{-j\omega_c t} x\left(t - (p-1)\frac{d\sin\theta}{c}\right) && (21)\\ &= e^{-j(p-1)\omega_c \frac{d\sin\theta}{c}} s\left(t - (p-1)\frac{d\sin\theta}{c}\right) && (22)\\ &\simeq e^{-j(p-1)\omega_c \frac{d\sin\theta}{c}} s(t). && (23) \end{aligned}$$

DoA Estimation: Preliminaries (cont'd)

Let $\boldsymbol{y}(t) = [y_1(t), \cdots, y_P(t)]^T$, we know

$$\boldsymbol{y}(t) = \boldsymbol{a}(\theta)s(t), \tag{24}$$

where

$$\boldsymbol{a}(\theta) \triangleq \left[1, e^{-j\phi(\theta)}, e^{-j2\phi(\theta)}, \cdots, e^{-j(P-1)\phi(\theta)}\right]^T, \tag{25}$$

is called the **steering vector**, in which

$$\phi(\theta) \triangleq \omega_c \frac{d \sin\theta}{c} = 2\pi \frac{d \sin\theta}{\lambda}, \tag{26}$$

and λ is the wavelength corresponding to the center frequency ω_c.

- To avoid **spatial aliasing**, i.e., $\boldsymbol{a}(\theta_1) = \boldsymbol{a}(\theta_2)$ for some $\theta_1 \neq \theta_2$, $\theta_1, \theta_2 \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, we need
$$d < \frac{\lambda}{2}. \tag{27}$$

DoA Estimation: Problem Setup and Model

Scenario: K radiating sources, noisy observation.

Model: consider discrete-time sampling $\boldsymbol{y}[n] = \boldsymbol{y}(nT_s)$, where T_s is the sampling period. The signal $\boldsymbol{y}[n]$ is modeled as

$$\boldsymbol{y}[n] = \sum_{k=1}^{K} \boldsymbol{a}(\theta_k) s_k[n] + \boldsymbol{\nu}[n] \tag{28}$$

$$= \boldsymbol{A}\boldsymbol{s}[n] + \boldsymbol{\nu}[n],\ n = 1, 2, \cdots, \tag{29}$$

where

- $\boldsymbol{s}_k[n]$ is the k^{th} source signal.
- θ_k is the DoA of the k^{th} source.
- $\boldsymbol{\nu}[n]$ is zero-mean spatially white noise, i.e., $\mathbb{E}\{\boldsymbol{\nu}[n]\boldsymbol{\nu}^H[n]\} = \sigma_{\boldsymbol{\nu}}^2 \boldsymbol{I}$.
- $\boldsymbol{A} = [\boldsymbol{a}(\theta_1), \cdots, \boldsymbol{a}(\theta_K)]$.
- $\boldsymbol{s}[n] = [s_1[n], \cdots, s_K[n]]^T$.

DoA Estimation: Problem Setup and Model (cont'd)

Assumptions:

- every $s_k[n]$ is zero-mean WSS.
- $\boldsymbol{s}[n]$ and $\boldsymbol{\nu}[n]$ are uncorrelated.

The correlation matrix of $\boldsymbol{y}[n]$ is given by

$$\begin{aligned}\boldsymbol{R}_{\boldsymbol{y}} &\triangleq \mathbb{E}\left\{\boldsymbol{y}[n]\boldsymbol{y}^H[n]\right\} &(30)\\ &= \mathbb{E}\left\{\boldsymbol{A}\boldsymbol{s}[n]\boldsymbol{s}^H[n]\boldsymbol{A}^H\right\} + \mathbb{E}\left\{\boldsymbol{\nu}[n]\boldsymbol{\nu}^H[n]\right\} &(31)\\ &= \boldsymbol{A}\boldsymbol{R}_{\boldsymbol{s}}\boldsymbol{A}^H + \sigma_\nu^2\boldsymbol{I}, &(32)\end{aligned}$$

where $\boldsymbol{R}_{\boldsymbol{s}} = \mathbb{E}\left\{\boldsymbol{s}[n]\boldsymbol{s}^H[n]\right\}$.

DoA Estimation: Problem Setup and Model (cont'd)

Assumptions:

- every $s_k[n]$ is zero-mean WSS.
- $\boldsymbol{s}[n]$ and $\boldsymbol{\nu}[n]$ are uncorrelated.

The correlation matrix of $\boldsymbol{y}[n]$ is given by

$$\boldsymbol{R}_{\boldsymbol{y}} \triangleq \mathbb{E}\left\{\boldsymbol{y}[n]\boldsymbol{y}^H[n]\right\} \tag{30}$$

$$= \mathbb{E}\left\{\boldsymbol{A}\boldsymbol{s}[n]\boldsymbol{s}^H[n]\boldsymbol{A}^H\right\} + \mathbb{E}\left\{\boldsymbol{\nu}[n]\boldsymbol{\nu}^H[n]\right\} \tag{31}$$

$$= \boldsymbol{A}\boldsymbol{R}_{\boldsymbol{s}}\boldsymbol{A}^H + \sigma_\nu^2\boldsymbol{I}, \tag{32}$$

where $\boldsymbol{R}_{\boldsymbol{s}} = \mathbb{E}\left\{\boldsymbol{s}[n]\boldsymbol{s}^H[n]\right\}$.

Problem: estimate $\theta_1, \cdots, \theta_K$ from $\boldsymbol{R}_{\boldsymbol{y}}$, assuming the model

$$\boldsymbol{R}_{\boldsymbol{y}} = \boldsymbol{A}\boldsymbol{R}_{\boldsymbol{s}}\boldsymbol{A}^H + \sigma_\nu^2\boldsymbol{I}. \tag{33}$$

DoA Estimation: Subspace Approach

Idea: use eigendecomposition of $\boldsymbol{R_y}$ and subspace properties to identify $\theta_1, \cdots, \theta_K$.

Assumptions:

- $P > K$ (more sensors than sources).
- $\theta_1, \cdots, \theta_K$ are distinct.
- $\boldsymbol{R_s}$ has full rank (i.e., the sources may be correlated, but not coherent).

Property 1

The array response matrix $\boldsymbol{A}$ has full rank.

Vandemonde Matrix

A matrix $\boldsymbol{A}$ is said to be **Vandemonde** if it takes the form

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_n \\ z_1^2 & z_2^2 & \cdots & z_n^2 \\ \vdots & \dots & \ddots & \vdots \\ z_1^{m-1} & z_2^{m-1} & \cdots & z_n^{m-1} \end{bmatrix} \in \mathbb{C}^{m \times n}, \tag{34}$$

where $z_1, z_2, \cdots, z_n \in \mathbb{C}$.

Vandemonde Matrix

A matrix $\boldsymbol{A}$ is said to be **Vandemonde** if it takes the form

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_n \\ z_1^2 & z_2^2 & \cdots & z_n^2 \\ \vdots & \dots & \ddots & \vdots \\ z_1^{m-1} & z_2^{m-1} & \cdots & z_n^{m-1} \end{bmatrix} \in \mathbb{C}^{m \times n}, \tag{34}$$

where $z_1, z_2, \cdots, z_n \in \mathbb{C}$.

Property 2

A Vandemonde matrix with $z_i \neq z_k$ for all $i \neq k$ has full rank.

- The array response matrix $\boldsymbol{A}$ in DoA estimation is a special case of the Vandemonde matrix where $z_i = e^{-j\phi(\theta_i)}$.
- Property 1 directly follows from Property 2.

DoA Estimation: Subspace Approach

Denote the eigendecomposition of the signal correlation matrix $\boldsymbol{A}\boldsymbol{R_s}\boldsymbol{A}^H$ by

$$\boldsymbol{A}\boldsymbol{R}_s\boldsymbol{A}^H = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^H. \tag{35}$$

Assume $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_P$.

Property 3

The eigenvalues $\lambda_1, \cdots, \lambda_P$ *of* $\boldsymbol{A}\boldsymbol{R_s}\boldsymbol{A}^H$ *satisfy*

$$\lambda_i > 0,\ i = 1, \cdots, K \textit{ while } \lambda_{K+1} = \cdots = \lambda_P = 0. \tag{36}$$

DoA Estimation: Subspace Approach

Denote the eigendecomposition of the signal correlation matrix $\boldsymbol{A}\boldsymbol{R}_{\boldsymbol{s}}\boldsymbol{A}^H$ by

$$\boldsymbol{A}\boldsymbol{R}_s\boldsymbol{A}^H = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^H. \tag{35}$$

Assume $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_P$.

Property 3

The eigenvalues $\lambda_1, \cdots, \lambda_P$ of $\boldsymbol{A}\boldsymbol{R}_{\boldsymbol{s}}\boldsymbol{A}^H$ satisfy

$$\lambda_i > 0,\ i = 1, \cdots, K \text{ while } \lambda_{K+1} = \cdots = \lambda_P = 0. \tag{36}$$

Property 4

The eigendecomposition of $\boldsymbol{R}_{\boldsymbol{y}}$ is

$$\boldsymbol{R}_{\boldsymbol{y}} = \boldsymbol{V}(\boldsymbol{\Lambda} + \sigma_\nu^2\boldsymbol{I})\boldsymbol{V}^{\boldsymbol{H}}, \tag{37}$$

where $\boldsymbol{V}$ and $\boldsymbol{\Lambda}$ denote the eigenvector and eigenvalue matrices of $\boldsymbol{A}\boldsymbol{R}_{\boldsymbol{s}}\boldsymbol{A}^H$, resp.

- Property 4 means that the eigenvector matrix $\boldsymbol{V}$ of the signal correlation matrix can be obtained from the eigendecomposition of $\boldsymbol{R}_{\boldsymbol{y}}$.

DoA Estimation: Subspace Approach (cont'd)

Theorem 1

Partition $\boldsymbol{V} = [\boldsymbol{V}_1, \boldsymbol{V}_2]$, *where*

$$\boldsymbol{V}_1 = [\boldsymbol{v}_1, \cdots, \boldsymbol{v}_K], \quad \boldsymbol{V}_2 = [\boldsymbol{v}_{K+1}, \cdots, \boldsymbol{v}_P]. \tag{38}$$

The equation

$$\boldsymbol{V}_2^H \boldsymbol{a}(\theta) = \boldsymbol{0}, \quad \theta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \tag{39}$$

is satisfied if and only if $\theta = \theta_i$ *for any* $i = 1, \cdots, K$.

- The subspace approach amounts to using $\boldsymbol{V}_2^H \boldsymbol{a}(\theta) = \boldsymbol{0}$ (or its approximation) to identify the source DoAs.
- There are many ways to do so, leading to many algorithms.
- A classical and most popular one is multiple signal classification (MUSIC).

DoA Estimation: The MUSIC Algorithm

MUSIC: Multiple Signal Classification[1]

- **Step1.** compute the sample correlation matrix $\hat{\boldsymbol{R}}_{\boldsymbol{y}} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{y}_n[n]\,\boldsymbol{y}_n^H[n]$.
- **Step2.** compute the eigenvector matrix of $\hat{\boldsymbol{R}}_{\boldsymbol{y}}$, denoted by $\hat{\boldsymbol{V}}$.
- **Step3.** evaluate the 'pseudo-spectrum'

$$P_{\mathsf{MUSIC}}(\theta) = \frac{1}{\|\hat{\boldsymbol{V}}_2^H \boldsymbol{a}(\theta)\|_2^2}, \quad \text{for } \theta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right], \tag{40}$$

where $\hat{\boldsymbol{V}} = [\hat{\boldsymbol{v}}_{K+1}, \cdots, \hat{\boldsymbol{v}}_P]$, and determine the DoAs by finding the peaks of $P_{\mathsf{MUSIC}}(\theta)$.

[1] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antenna Propagation*, vol. AP-34, pp. 276–290, March 1986.

DoA Estimation: Numerical Demonstration

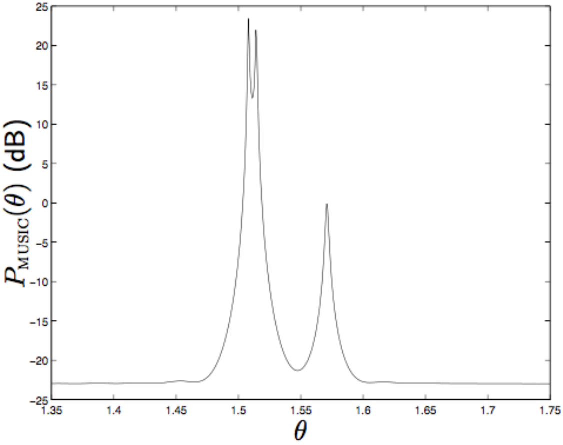


Table of Contents

Preliminary: Circulant Matrix

A matrix $\boldsymbol{H}$ is said to be **circulant** if it takes the following structure

$$\boldsymbol{H} = \begin{bmatrix} h_0 & h_{N-1} & \cdots & h_2 & h_1 \\ h_1 & h_0 & \cdots & h_3 & h_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ h_{N-1} & h_{N-2} & \cdots & h_1 & h_0 \end{bmatrix} \tag{41}$$

Preliminary: Circulant Matrix

A matrix $\boldsymbol{H}$ is said to be **circulant** if it takes the following structure

$$\boldsymbol{H} = \begin{bmatrix} h_0 & h_{N-1} & \cdots & h_2 & h_1 \\ h_1 & h_0 & \cdots & h_3 & h_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ h_{N-1} & h_{N-2} & \cdots & h_1 & h_0 \end{bmatrix} \tag{41}$$

Eigenvectors: let $\boldsymbol{f}_k \triangleq \frac{1}{\sqrt{N}} \left[1, e^{j\frac{2\pi k}{N}}, \cdots, e^{j\frac{2\pi(N-1)k}{N}}\right]^T \in \mathbb{C}^N$. It can be verified that

$$\boldsymbol{H}\boldsymbol{f}_k = d_k \boldsymbol{f}_k, \quad k = 0, \cdots, N-1, \tag{42}$$

where

$$d_k = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} h_n e^{-j\frac{2\pi nk}{N}}. \tag{43}$$

- $\{d_k\}$ is the normalized DFT of $\{h_n\}$.

Preliminary: Circulant Matrix (cont'd)

Let $\boldsymbol{F} \triangleq [\boldsymbol{f}_0, \cdots, \boldsymbol{f}_{N-1}] \in \mathbb{C}^{N\times N}$.

- $\boldsymbol{F}$ is unitary.
- $\boldsymbol{F}$ is seen to be the inverse DFT (IDFT) matrix, while $\boldsymbol{F}^{-1} = \boldsymbol{F}^H$ the DFT matrix.
- DFT and IDFT can be efficiently computed (complexity order of $N\log(N)$).

Preliminary: Circulant Matrix (cont'd)

Let $\boldsymbol{F} \triangleq [\boldsymbol{f}_0, \cdots, \boldsymbol{f}_{N-1}] \in \mathbb{C}^{N \times N}$.

- $\boldsymbol{F}$ is unitary.
- $\boldsymbol{F}$ is seen to be the inverse DFT (IDFT) matrix, while $\boldsymbol{F}^{-1} = \boldsymbol{F}^H$ the DFT matrix.
- DFT and IDFT can be efficiently computed (complexity order of $N \log(N)$).

By the unitarity of $\boldsymbol{F}$, a circulant $\boldsymbol{H}$ admits an eigendecomposition

$$\boldsymbol{H} = \boldsymbol{F}\boldsymbol{D}\boldsymbol{F}^H, \tag{44}$$

$$\boldsymbol{D} = \text{diag}(d_0, \cdots, d_{N-1}). \tag{45}$$

- Application: **orthogonal frequency division multiplexing (OFDM)** in wireless communications.

OFDM: Background

Scenario: digital transmission over a linear time-invariant channel.
System Model:

$$y[n] = \sum_{l=0}^{L} h_l s[n-l] + \nu[n], \quad n \in \mathbb{Z}, \tag{46}$$

where

- $s[n]$: the transmitted data symbol stream,
- $y[n]$: the received signal (discrete-time sampled),
- h_l: the channel impulse response, with length $L+1$,
- $\nu[n]$: AWGN noise.

OFDM: Background

Scenario: digital transmission over a linear time-invariant channel.
System Model:

$$y[n] = \sum_{l=0}^{L} h_l s[n-l] + \nu[n], \quad n \in \mathbb{Z}, \tag{46}$$

where

- $s[n]$: the transmitted data symbol stream,
- $y[n]$: the received signal (discrete-time sampled),
- h_l: the channel impulse response, with length $L+1$,
- $\nu[n]$: AWGN noise.

Problem formulation:
Originally, we wish to detect each $s[n]$ from $y[n]$ – which works if $h_0 \neq 0$, $h_l = 0$ for $l > 1$. However, in general, we have **intersymbol interference (ISI)**

$$y[n] = h_0 s[n] + \underbrace{\sum_{l \neq 0} h_l s[n-l]}_{\text{ISI}} + \nu[n]. \tag{47}$$

OFDM

Aim: precode $\{s[n]\}$ so that it can be received in a zero-ISI manner.

Consider a block transmission, where transmission is parsed into time blocks, and every block has N symbols, $s[0], s[1], \cdots, s[N-1]$, to transmit.

Model for one time block:

$$y[n] = \sum_{l=0}^{L} h_l \tilde{x}[n-l] + \nu[n], \quad n = 0, 1, \cdots, N+L-1, \tag{48}$$

where

$$\left.\begin{aligned} \tilde{x}[0] &= x[N-L] \\ \tilde{x}[1] &= x[N-L+1] \\ &\vdots \\ \tilde{x}[L-1] &= x[N-1] \end{aligned}\right\} \text{cyclic prefix} \qquad \begin{aligned} \tilde{x}[L] &= x[0] \\ \tilde{x}[L+1] &= x[1] \\ &\vdots \\ \tilde{x}[N+L-1] &= x[N-1] \end{aligned}$$

and $x[n]$ is the precoded signal, which will be specified shortly.

OFDM (cont'd)

It can be verified that

$$\underbrace{\begin{bmatrix} y[L] \\ y[L+1] \\ \vdots \\ \vdots \\ \vdots \\ y[N+L-1] \end{bmatrix}}_{\triangleq \boldsymbol{y} \in \mathbb{C}^N} = \underbrace{\begin{bmatrix} h_0 & \cdots & \cdots & h_L & \cdots & h_1 \\ h_1 & h_0 & \cdots & \cdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \cdots & \cdots & h_L \\ h_L & & & \ddots & & \\ & \ddots & & & \ddots & \\ & & h_L & \cdots & h_1 & h_0 \end{bmatrix}}_{\triangleq \boldsymbol{H} \in \mathbb{C}^{N\times N}} \underbrace{\begin{bmatrix} x[0] \\ x[1] \\ \vdots \\ \vdots \\ \vdots \\ x[N-1] \end{bmatrix}}_{\triangleq \boldsymbol{x} \in \mathbb{C}^N} + \underbrace{\begin{bmatrix} \nu[L] \\ \nu[L+1] \\ \vdots \\ \vdots \\ \vdots \\ \nu[N+L-1] \end{bmatrix}}_{\triangleq \boldsymbol{\nu} \in \mathbb{C}^N}, \quad (49)$$

where a linear system $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{\nu}$ with a circulant system response $\boldsymbol{H}$ is obtained, in other words, we have

$$\boldsymbol{y} = \boldsymbol{F}\boldsymbol{D}\boldsymbol{F}^H\boldsymbol{x} + \boldsymbol{\nu}. \quad (50)$$

OFDM (cont'd)

1) At the transmitter side (IDFT), let $\boldsymbol{s} = [s[0], \cdots, s[N-1]]^T$.
OFDM: choose $\boldsymbol{x} = \boldsymbol{F}\boldsymbol{s}$ (i.e., IDFT the symbol block).
The system model reduces to

$$\boldsymbol{y} = \boldsymbol{F}\boldsymbol{D}\boldsymbol{F}^H\boldsymbol{x} + \boldsymbol{\nu} = \boldsymbol{F}\boldsymbol{D}\boldsymbol{F}^H\boldsymbol{F}\boldsymbol{s} + \boldsymbol{\nu} = \boldsymbol{F}\boldsymbol{D}\boldsymbol{s} + \boldsymbol{\nu}. \quad (51)$$

2) At the receiver side (DFT), consider a transformation $\boldsymbol{z} = \boldsymbol{F}^H\boldsymbol{y}$ (i.e., DFT $\boldsymbol{y}$, done by the receiver). Since

$$\boldsymbol{z} = \boldsymbol{D}\boldsymbol{s} + \boldsymbol{F}^H\boldsymbol{\nu} \Longleftrightarrow z_k = d_k s_k + [\boldsymbol{F}^H\boldsymbol{\nu}]_k, \quad k = 0, \cdots, N-1, \quad (52)$$

zero ISI is achieved.

- The noise term $\boldsymbol{F}^H\boldsymbol{\nu}$ is only rotated (by the unitarity of $\boldsymbol{F}$), but not magnified.
- A related scheme is single-carrier modulation (SCM), which chooses $\boldsymbol{x} = \boldsymbol{s}$. The receiver can achieve zero ISI by $\boldsymbol{z} = \boldsymbol{F}\boldsymbol{D}^{-1}\boldsymbol{F}^H\boldsymbol{y}$ (can be efficiently computed via DFT/IDFT), assuming an invertible $\boldsymbol{D}$. However, noise may be magnified.

OFDM (cont'd)

Remark 1

1) *To capture the central idea, we have introduced OFDM using a simplified model, namely, no interference between time blocks is assumed, but actually there is.*
   - *The inter-block interference can be easily handled by the so-called guard time removal; find textbooks for details, see e.g., [R1].*[2]

2) *OFDM (also SCM) uses a block length of $N + L$ to transmit N symbols. For good efficiency, N should be much greater than L.*

[2] [R1] J. Li, X. Wu, and R. Laroia, *OFDMA Mobile Broadband Communications: A Systems Approach*, Cambridge University Press, 2013.

Thank you for your attention!

Kai Lu

E-mail: lukai86@mail.sysu.edu.cn

Web: http://seit.sysu.edu.cn/teacher/1801