Matrix Computations
Chapter 5: Positive Semidefinite Matrices
Section 5.2 Examples of Applications

Jie Lu
ShanghaiTech University

Application: Spectral Analysis via Subspace

- Consider the complex harmonic time-series

$$y_t = \sum_{i=1}^{k} \alpha_i e^{\mathbf{j}2\pi f_i t} + w_t, \quad t = 0, 1, \ldots, T-1$$

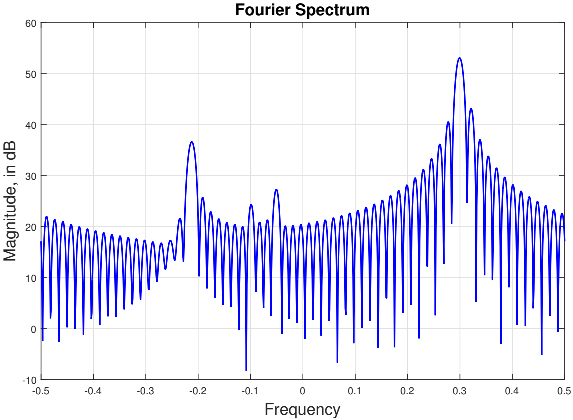
  where $\alpha_i \in \mathbb{C}$ is the amplitude-phase coefficient of the ith sinusoid; $f_i \in \left[-\frac{1}{2}, \frac{1}{2}\right)$ is the frequency of the ith sinusoid; w_t is noise; T is the observation time length

- **Aim**: Estimate the frequencies $f_1, \ldots, f_k$ from $\{y_t\}_{t=0}^{T-1}$
  - Can be done by applying the Fourier transform
  - The spectral resolution of Fourier-based methods is often limited by T

- Our interest: study a subspace approach which can enable "super-resolution"[1]

[1] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*, Prentice Hall, 1997.

Illustration



An illustration of the Fourier spectrum. $T = 64$, $k = 5$, $\{f_1, \ldots, f_k\} = \{-0.213, -0.1, -0.05, 0.3, 0.315\}$

Spectral Analysis: Formulation

Let $z_i = e^{\boldsymbol{j}2\pi f_i}$. Given a positive integer d, let

$$\mathbf{y}_t = \begin{bmatrix} y_t \\ y_{t+1} \\ \vdots \\ y_{t+d-1} \end{bmatrix} = \sum_{i=1}^{k} \alpha_i \begin{bmatrix} z_i^t \\ z_i^{t+1} \\ \vdots \\ z_i^{t+d-1} \end{bmatrix} + \begin{bmatrix} w_t \\ w_{t+1} \\ \vdots \\ w_{t+d-1} \end{bmatrix} = \sum_{i=1}^{k} \alpha_i \underbrace{\begin{bmatrix} 1 \\ z_i \\ \vdots \\ z_i^{d-1} \end{bmatrix}}_{=\mathbf{a}_i} z_i^t + \underbrace{\begin{bmatrix} w_t \\ w_{t+1} \\ \vdots \\ w_{t+d-1} \end{bmatrix}}_{\mathbf{w}_t}$$

Let $\mathbf{Y} = [\ \mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_{T_d-1}\ ]$ where $T_d = T - d + 1$. We can write

$$\mathbf{Y} = \mathbf{ADS} + \mathbf{W},$$

where $\mathbf{A} = [\ \mathbf{a}_1, \ldots, \mathbf{a}_k\ ]$, $\mathbf{D} = \mathrm{Diag}(\alpha_1, \ldots, \alpha_k)$, $\mathbf{W} = [\ \mathbf{w}_0, \ldots, \mathbf{w}_{T_d-1}\ ]$,

$$\mathbf{S} = \begin{bmatrix} 1 & z_1 & z_1^2 & \ldots & z_1^{T_d-1} \\ 1 & z_2 & z_2^2 & \ldots & z_2^{T_d-1} \\ \vdots & & & & \vdots \\ 1 & z_k & z_k^2 & \ldots & z_k^{T_d-1} \end{bmatrix}$$

Spectral Analysis: Formulation (cont'd)

Let $\mathbf{R}_y = \frac{1}{T_d}\sum_{t=0}^{T_d-1} \mathbf{y}_t \mathbf{y}_t^H = \frac{1}{T_d}\mathbf{Y}\mathbf{Y}^H$ be the correlation matrix of $\mathbf{y}_t$

$$\mathbf{R}_y = \mathbf{A}\underbrace{\left(\frac{1}{T_d}\mathbf{D}\mathbf{S}\mathbf{S}^H\mathbf{D}^H\right)}_{=\mathbf{\Phi}}\mathbf{A}^H + \frac{1}{T_d}\mathbf{A}\mathbf{D}\mathbf{S}\mathbf{W}^H + \frac{1}{T_d}\mathbf{W}\mathbf{S}^H\mathbf{D}^H\mathbf{A}^H + \frac{1}{T_d}\mathbf{W}\mathbf{W}^H$$

(This requires knowledge of random processes) Assume that w_t is a temporally white circular Gaussian process with mean zero and variance σ^2 . Then, as $T_d \to \infty$,

$$\frac{1}{T_d}\mathbf{S}\mathbf{W}^H \to \mathbf{0}, \qquad \frac{1}{T_d}\mathbf{W}\mathbf{W}^H \to \sigma^2\mathbf{I}$$

Therefore, we can approximate $\mathbf{R}_y$ by

$$\mathbf{R}_y = \mathbf{A}\mathbf{\Phi}\mathbf{A}^H + \sigma^2\mathbf{I}$$

Spectral Analysis: Formulation (cont'd)

Model: The correlation matrix $\mathbf{R}_y = \frac{1}{T_d}\mathbf{Y}\mathbf{Y}^H$ is modeled as

$$\mathbf{R}_y = \mathbf{A}\boldsymbol{\Phi}\mathbf{A}^H + \sigma^2\mathbf{I}$$

where $\sigma^2 > 0$ is the noise power; $\boldsymbol{\Phi} = \frac{1}{T_d}\mathbf{D}\mathbf{S}\mathbf{S}^H\mathbf{D}^H$; $\mathbf{D} = \mathrm{Diag}(\alpha_1, \ldots, \alpha_k)$

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ z_1 & z_2 & & z_k \\ \vdots & \vdots & & \vdots \\ z_1^{d-1} & z_2^{d-1} & \ldots & z_k^{d-1} \end{bmatrix} \in \mathbb{C}^{d\times k}, \quad \mathbf{S} = \begin{bmatrix} 1 & z_1 & z_1^2 & \ldots & z_1^{T_d-1} \\ 1 & z_2 & z_2^2 & \ldots & z_2^{T_d-1} \\ \vdots & & & & \vdots \\ 1 & z_k & z_k^2 & \ldots & z_k^{T_d-1} \end{bmatrix} \in \mathbb{C}^{k\times T_d},$$

with $z_i = e^{\boldsymbol{j}2\pi f_i}$

Observation: $\mathbf{A}$ and $\mathbf{S}^H$ are both Vandemonde

Spectral Analysis: Subspace Properties

Assumptions:

1. $\alpha_i \neq 0$ for all i
2. $f_i \neq f_j$ for all $i \neq j$
3. $d > k$
4. $T_d \geq k$

Consequences:

- $\mathbf{A}$ has full column rank, $\mathbf{S}$ has full row rank
- $\mathbf{\Phi}$ is positive definite (and thus nonsingular)
  - Proof: $\mathbf{x}^H\mathbf{D}\mathbf{S}\mathbf{S}^H\mathbf{D}^H\mathbf{x} = \|\mathbf{S}^H\mathbf{D}^H\mathbf{x}\|_2^2$, and $\mathbf{S}^H\mathbf{D}^H\mathbf{x} = \mathbf{0} \iff \mathbf{x} = \mathbf{0}$ because $\mathbf{S}^H$ has full column rank and $\mathbf{D}$ has full rank
- $\mathcal{R}(\mathbf{A}\mathbf{\Phi}\mathbf{A}^H) = \mathcal{R}(\mathbf{A})$
  - Proof: $\mathbf{A}^H$ has full row rank $\Longrightarrow$ rank$(\mathbf{\Phi}\mathbf{A}^H) =$ rank$(\mathbf{\Phi})$. Since $\mathbf{\Phi}$ is PD (and thus full rank), $\mathbf{\Phi}\mathbf{A}^H$ has full row rank. Then use the property on the last page of Section 5.1
- $\mathrm{rank}(\mathbf{A}\mathbf{\Phi}\mathbf{A}^H) = \mathrm{rank}(\mathbf{A}) = k$, and $\mathbf{A}\mathbf{\Phi}\mathbf{A}^H$ has k nonzero eigenvalues

Spectral Analysis: Subspace Properties (cont'd)

Consider the eigendecomposition of $\mathbf{A}\boldsymbol{\Phi}\mathbf{A}^H$. Let $\mathbf{A}\boldsymbol{\Phi}\mathbf{A}^H = \mathbf{V}\Lambda\mathbf{V}^H$ and assume $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d$ are the eigenvalues of $\mathbf{A}\boldsymbol{\Phi}\mathbf{A}^H$

Since $\mathbf{A}\boldsymbol{\Phi}\mathbf{A}^H$ is PSD, we have $\lambda_i > 0$ for $i = 1, \ldots, k$ and $\lambda_i = 0$ for $i = k+1, \ldots, d$

$$\mathbf{A}\boldsymbol{\Phi}\mathbf{A}^H = \begin{bmatrix} \mathbf{V}_1 & \mathbf{V}_2 \end{bmatrix} \begin{bmatrix} \Lambda_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^H \\ \mathbf{V}_2^H \end{bmatrix} = \mathbf{V}_1 \Lambda_1 \mathbf{V}_1^H$$

where $\mathbf{V}_1 = [\ \mathbf{v}_1, \ldots, \mathbf{v}_k\ ] \in \mathbb{C}^{d \times k}$, $\mathbf{V}_2 = [\ \mathbf{v}_{k+1}, \ldots, \mathbf{v}_d\ ] \in \mathbb{C}^{d \times (d-k)}$, $\Lambda_1 = \mathrm{Diag}(\lambda_1, \ldots, \lambda_k)$

Consequence: $\mathcal{R}(\mathbf{A}\boldsymbol{\Phi}\mathbf{A}^H) = \mathcal{R}(\mathbf{V}_1)$, $\mathcal{R}(\mathbf{A}\boldsymbol{\Phi}\mathbf{A}^H)^\perp = \mathcal{R}(\mathbf{V}_2)$

Spectral Analysis: Subspace Properties (cont'd)

Now consider the eigendecomposition of $\mathbf{R}_y$

$$\mathbf{R}_y = \begin{bmatrix} \mathbf{V}_1 & \mathbf{V}_2 \end{bmatrix} \begin{bmatrix} \Lambda_1 + \sigma^2 \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \sigma^2 \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^H \\ \mathbf{V}_2^H \end{bmatrix}$$

Consequences:

- $\mathbf{V}(\Lambda + \sigma^2 \mathbf{I})\mathbf{V}^H$ is the eigendecomposition of $\mathbf{R}_y$
- $\mathbf{V}_1$ can be obtained from $\mathbf{R}_y$ by finding the eigenvectors associated with the first k largest eigenvalues of $\mathbf{R}_y$

Spectral Analysis: Subspace Properties (cont'd)

- Compute the eigenvector matrix $\mathbf{V} \in \mathbb{C}^{d \times d}$ of $\mathbf{R}_y$. Partition $\mathbf{V} = [\ \mathbf{V}_1, \mathbf{V}_2\ ]$ where $\mathbf{V}_1 \in \mathbb{C}^{n \times k}$ corresponds the first k largest eigenvalues. Then,

$$\mathcal{R}(\mathbf{V}_1) = \mathcal{R}(\mathbf{A}), \qquad \mathcal{R}(\mathbf{V}_2) = \mathcal{R}(\mathbf{A})^{\perp}$$

- **Idea of subspace methods**: Let

$$\mathbf{a}(z) = \begin{bmatrix} 1 \\ z \\ \vdots \\ z^{d-1} \end{bmatrix}.$$

Find any $f \in [-\frac{1}{2}, \frac{1}{2})$ that satisfies $\mathbf{a}(e^{\boldsymbol{j}2\pi f}) \in \mathcal{R}(\mathbf{A})$

Spectral Analysis via Subspace: Subspace Properties

Question: it is true that $f \in \{f_1, \dots f_k\}$ implies $\mathbf{a}(e^{\boldsymbol{j}2\pi f}) \in \mathcal{R}(\mathbf{A})$. But is it also true that $\mathbf{a}(e^{\boldsymbol{j}2\pi f}) \in \mathcal{R}(\mathbf{A})$ implies $f \in \{f_1, \dots f_k\}$?
Answer: Yes if $d > k$

Theorem

Let $\mathbf{A} \in \mathbb{C}^{d \times k}$ any Vandemonde matrix with distinct roots $z_1, \dots, z_k$ and with $d \geq k+1$. Then,

$$z \in \{z_1, \dots, z_k\} \quad \Longleftrightarrow \quad \mathbf{a}(z) \in \mathcal{R}(\mathbf{A}).$$

Spectral Analysis: Algorithm

There are many subspace methods, and Multiple Signal Classification (MUSIC) is most well-known

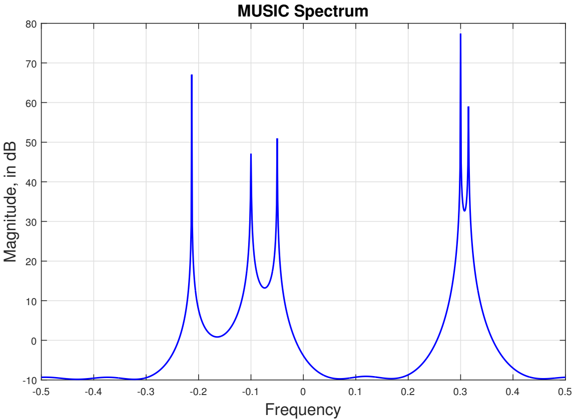
MUSIC uses the fact that $\mathbf{a}(e^{\boldsymbol{j}2\pi f}) \in \mathcal{R}(\mathbf{A}) \iff \mathbf{V}_2^H \mathbf{a}(e^{\boldsymbol{j}2\pi f}) = \mathbf{0}$

Algorithm: MUSIC
input: the correlation matrix $\mathbf{R}_y \in \mathbb{C}^{d\times d}$ and the model order $k < d$
Perform eigendecomposition $\mathbf{R}_y = \mathbf{V}\Lambda\mathbf{V}^H$ with $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d$.
Let $\mathbf{V}_2 = [\ \mathbf{v}_{k+1}, \ldots, \mathbf{v}_d\ ]$, and compute

$$S(f) = \frac{1}{\|\mathbf{V}_2^H \mathbf{a}(e^{\boldsymbol{j}2\pi f})\|_2^2}$$

for $f \in \left[-\frac{1}{2}, \frac{1}{2}\right)$ (done by discretization).
output: $S(f)$

Spectral Analysis: Algorithm (cont'd)

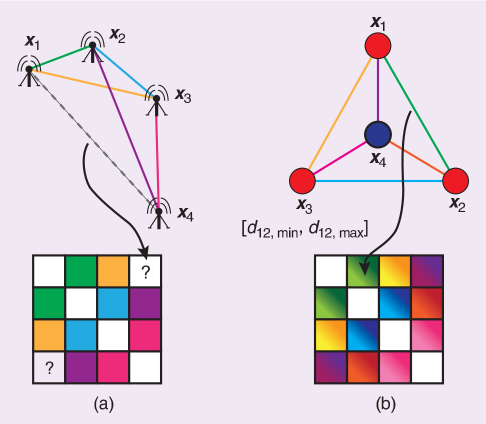


An illustration of the MUSIC spectrum. $T = 64$, $k = 5$,
$\{f_1, \ldots, f_k\} = \{-0.213, -0.1, -0.05, 0.3, 0.315\}$

Application: Euclidean Distance Matrices

- Let $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^d$ be a collection of points, and let $\mathbf{X} = [\ \mathbf{x}_1, \ldots, \mathbf{x}_n\ ]$
- Let $d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|_2$ be the Euclidean distance between points i and j
- **Problem**: Given d_{ij} for all $i, j \in \{1, \ldots, n\}$, recover $\mathbf{X}$
  - This is called the Euclidean distance matrix (EDM) problem
- Applications: sensor network localization (SNL), molecule conformation, etc.

Applications of EDM



(a) Sensor network localization (SNL) (b) Molecular transformation[2]

[2] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*, Prentice Hall, 1997.

EDM: Formulation

- Let $\mathbf{R} \in \mathbb{R}^{n \times n}$ be s.t. $r_{ij} = d_{ij}^2$ for all $i, j = 1, \dots, n$
- Note from
$$r_{ij} = d_{ij}^2 = \mathbf{x}_i^T \mathbf{x}_i - 2\mathbf{x}_i^T \mathbf{x}_j + \mathbf{x}_j^T \mathbf{x}_j$$
that $\mathbf{R}$ can be written as
$$\mathbf{R} = \mathbf{1}(\mathrm{diag}(\mathbf{X}^T\mathbf{X}))^T - 2\mathbf{X}^T\mathbf{X} + (\mathrm{diag}(\mathbf{X}^T\mathbf{X}))\mathbf{1}^T \qquad (*)$$
where $\mathrm{diag}(\mathbf{Y}) := [y_{11}, \dots, y_{nn}]^T$ for any square matrix $\mathbf{Y}$
- **Observation:** $(*)$ Also holds if we replace $\mathbf{X}$ by
  - $\tilde{\mathbf{X}} = [\mathbf{x}_1 + \mathbf{b}, \dots, \mathbf{x}_n + \mathbf{b}]$ for any $\mathbf{b} \in \mathbb{R}^d$ $(d_{ij} = \|\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j\|_2)$
  - $\tilde{\mathbf{X}} = \mathbf{Q}\mathbf{X}$ for any orthogonal $\mathbf{Q}$ $(\tilde{\mathbf{X}}^T\tilde{\mathbf{X}} = \mathbf{X}^T\mathbf{X})$
- **Implication**: recovery of $\mathbf{X}$ from $\mathbf{R}$ is subjected to translations and rotations/reflections
  - In SNL we can use anchors to fix this issue

EDM: Formulation (cont'd)

EDM: Formulation (cont'd)

- WLOG, assume $\mathbf{x}_1 = \mathbf{0}$

$$\mathbf{r}_1 = \begin{bmatrix} \|\mathbf{x}_1 - \mathbf{x}_1\|_2^2 \\ \|\mathbf{x}_2 - \mathbf{x}_1\|_2^2 \\ \vdots \\ \|\mathbf{x}_n - \mathbf{x}_1\|_2^2 \end{bmatrix} = \begin{bmatrix} 0 \\ \|\mathbf{x}_2\|_2^2 \\ \vdots \\ \|\mathbf{x}_n\|_2^2 \end{bmatrix}, \qquad \mathrm{diag}(\mathbf{X}^T\mathbf{X}) = \begin{bmatrix} \|\mathbf{x}_1\|_2^2 \\ \|\mathbf{x}_2\|_2^2 \\ \vdots \\ \|\mathbf{x}_n\|_2^2 \end{bmatrix} = \mathbf{r}_1$$

- Combining the above with $(*)$ gives

$$\mathbf{X}^T\mathbf{X} = -\frac{1}{2}(\mathbf{R} - \mathbf{1}\mathbf{r}_1^T - \mathbf{r}_1\mathbf{1}^T)$$

- **Idea**: do a symmetric factorization $\mathbf{G} = \mathbf{X}^T\mathbf{X}$ for

$$\mathbf{G} = -\frac{1}{2}(\mathbf{R} - \mathbf{1}\mathbf{r}_1^T - \mathbf{r}_1\mathbf{1}^T)$$

to recover $\mathbf{X}$

EDM: Method

- **Assumption**: $\mathbf{X}$ has full row rank
- It can be shown that $\mathbf{G}$ is PSD, $\mathrm{rank}(\mathbf{G}) = d$, and $\mathbf{G}$ has d nonzero eigenvalues
- Let $\mathbf{G} = \mathbf{V}\Lambda\mathbf{V}^T$ be the eigendecomposition of $\mathbf{G}$ with eigenvalues $\lambda_1 \geq \ldots \geq \lambda_n$

$$\mathbf{G} = \begin{bmatrix} \mathbf{V}_1 & \mathbf{V}_2 \end{bmatrix} \begin{bmatrix} \Lambda_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T \end{bmatrix} = (\Lambda^{1/2}\mathbf{V}_1^T)^T(\Lambda^{1/2}\mathbf{V}_1^T)$$

  where $\mathbf{V}_1 \in \mathbb{R}^{d \times d}$, $\Lambda_1 = \mathrm{Diag}(\lambda_1, \ldots, \lambda_d)$

- **EDM solution**: Take $\hat{\mathbf{X}} = \Lambda^{1/2}\mathbf{V}_1^T$ as an estimate of $\mathbf{X}$
- Recovery guarantee: From the last property of Section 5.1, $\hat{\mathbf{X}} = \mathbf{QX}$ for some orthogonal $\mathbf{Q}$

EDM: Further Discussion

- In applications such as SNL, not all pairwise distances d_{ij}'s are available, so that there are missing entries in $\mathbf{R}$
- Possible solution: Apply low-rank matrix completion (cf. Section 3.4) to recover the full $\mathbf{R}$
- To use low-rank matrix completion, we need a bound on rank($\mathbf{R}$)
- Using $\mathrm{rank}(\mathbf{A}+\mathbf{B}) \leq \mathrm{rank}(\mathbf{A}) + \mathrm{rank}(\mathbf{B})$, we have

$$\begin{aligned}\mathrm{rank}(\mathbf{R}) \leq & \mathrm{rank}(\mathbf{1}(\mathrm{diag}(\mathbf{X}^T\mathbf{X}))^T) + \mathrm{rank}(-2\mathbf{X}^T\mathbf{X}) \\ & + \mathrm{rank}((\mathrm{diag}(\mathbf{X}^T\mathbf{X}))\mathbf{1}^T) \\ \leq & 1 + d + 1 = d + 2\end{aligned}$$

- Other issues:[3] Noisy distance measurements, resolving the orthogonal rotation problem with $\hat{\mathbf{X}}$

[3] I. Dokmanić, R. Parhizkar, J. Ranieri, and Vetterli, "Euclidean distance matrices," *IEEE Signal Processing Magazine*, vol. 32, no. 6, pp. 12–30, Nov. 2015.