SI231 Matrix Analysis and Computations
Topic 6: Positive Semidefinite Matrices

Ziping Zhao

Spring Term 2020–2021

School of Information Science and Technology
ShanghaiTech University, Shanghai, China

Topic 6: Positive Semidefinite Matrices

- positive semidefinite matrices
- matrix inequalities
- application: subspace method for super-resolution spectral analysis
- application: Euclidean distance matrices

Hightlights

- a matrix $\mathbf{A} \in \mathbb{S}^n$ is said to be positive semidefinite (PSD) if

$$\mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0, \quad \text{for all } \mathbf{x} \in \mathbb{R}^n;$$

  and positive definite (PD) if

$$\mathbf{x}^T \mathbf{A} \mathbf{x} > 0, \quad \text{for all } \mathbf{x} \in \mathbb{R}^n \text{ with } \mathbf{x} \neq \mathbf{0}$$

- a matrix $\mathbf{A} \in \mathbb{S}^n$ is PSD (resp. PD)
  - if and only if its eigenvalues are all non-negative (resp. positive);
  - if and only if it can be factored as $\mathbf{A} = \mathbf{B}^T \mathbf{B}$ for some $\mathbf{B} \in \mathbb{R}^{m \times n}$
- in this lecture, we will deal with the real-symmetric matrices–the Hermitian case follows along the same lines

Quadratic Form

Let $\mathbf{A} \in \mathbb{S}^n$. For $\mathbf{x} \in \mathbb{R}^n$, the matrix product

$$\mathbf{x}^T\mathbf{A}\mathbf{x}$$

is called a quadratic form.

- some basic facts (try to verify):
  - $\mathbf{x}^T\mathbf{A}\mathbf{x} = \sum_{i=1}^n \sum_{j=1}^n x_i x_j a_{ij} = \sum_{i=1}^n a_{ii}x_i^2 + \sum_{i=1}^{n-1}\sum_{j=i+1}^n 2a_{ij}x_i x_j$
  - $\mathbf{x}^T\mathbf{A}\mathbf{x} = \sum_{i=1}^n a_{ii}x_i^2 + \sum_{i=1}^{n-1}\sum_{j=i+1}^n (a_{ij} + a_{ji})x_i x_j$ for general $\mathbf{A} \in \mathbb{R}^{n\times n}$, there may exist $\mathbf{A}_1$ and $\mathbf{A}_2$ s.t. $\mathbf{x}^T\mathbf{A}_1\mathbf{x} = \mathbf{x}^T\mathbf{A}_2\mathbf{x}$
    * it suffices to consider unique symmetric $\mathbf{A}$ for general $\mathbf{A} \in \mathbb{R}^{n\times n}$ since

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{x}^T\left[\tfrac{1}{2}(\mathbf{A} + \mathbf{A}^T)\right]\mathbf{x}$$

  - complex case:
    * the (complex) quadratic form is defined as $\mathbf{x}^H\mathbf{A}\mathbf{x}$, where $\mathbf{x} \in \mathbb{C}^n$
    * for $\mathbf{A} \in \mathbb{H}^n$, $\mathbf{x}^H\mathbf{A}\mathbf{x}$ is real for any $\mathbf{x} \in \mathbb{C}^n$

Positive Semidefinite Matrices

A matrix $\mathbf{A} \in \mathbb{S}^n$ is said to be

- positive semidefinite (PSD) if $\mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$
- positive definite (PD) if $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x} \in \mathbb{R}^n$ with $\mathbf{x} \neq \mathbf{0}$
- indefinite if both $\mathbf{A}$ and $-\mathbf{A}$ are not PSD

Notation:

- $\mathbf{A} \succeq \mathbf{0}$ means that $\mathbf{A}$ is PSD
- $\mathbf{A} \succ \mathbf{0}$ means that $\mathbf{A}$ is PD
- $\mathbf{A} \not\succeq \mathbf{0}$ means that $\mathbf{A}$ is indefinite

- if $\mathbf{A}$ is PD, then it is also PSD
- a quadratic form is called PSD (resp. PD) if $\mathbf{A}$ is PSD (resp. PD)
- The concepts negative semidefinite and negative definite may be defined by reversing the inequalities or, equivalently, by saying $-\mathbf{A}$ is PSD or PD, respectively.

Example: Covariance Matrices

- let $\mathbf{y}_0, \mathbf{y}_2, \ldots \mathbf{y}_{T-1} \in \mathbb{R}^n$ be a sequence of multi-dimensional data samples
  - examples: patches in image processing, multi-channel signals in signal processing, history of returns of assets in finance **[Brodie-Daubechies-et al.'09]**, ...

- sample mean: $\hat{\boldsymbol{\mu}}_y = \frac{1}{T} \sum_{t=0}^{T-1} \mathbf{y}_t$

- sample covariance: $\hat{\mathbf{C}}_y = \frac{1}{T} \sum_{t=0}^{T-1} (\mathbf{y}_t - \hat{\boldsymbol{\mu}}_y)(\mathbf{y}_t - \hat{\boldsymbol{\mu}}_y)^T$

- a sample covariance is PSD: $\mathbf{x}^T \hat{\mathbf{C}}_y \mathbf{x} = \frac{1}{T} \sum_{t=0}^{T-1} |(\mathbf{y}_t - \hat{\boldsymbol{\mu}}_y)^T \mathbf{x}|^2 \geq 0$

- the (statistical) covariance of $\mathbf{y}_t$ is also PSD
  - to put into context, assume that $\mathbf{y}_t$ is a wide-sense stationary random process
  - the covariance, defined as $\mathbf{C}_y = \mathrm{E}[(\mathbf{y}_t - \boldsymbol{\mu}_y)(\mathbf{y}_t - \boldsymbol{\mu}_y)^T]$ where $\boldsymbol{\mu}_y = \mathrm{E}[\mathbf{y}_t]$, can be shown to be PSD

Example: Hessian

- let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a twice differentiable function

- the Hessian of f, denoted by $\nabla^2 f(\mathbf{x}) \in \mathbb{S}^n$, is a matrix whose (i, j)th entry is given by

$$\left[\nabla^2 f(\mathbf{x})\right]_{i,j} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

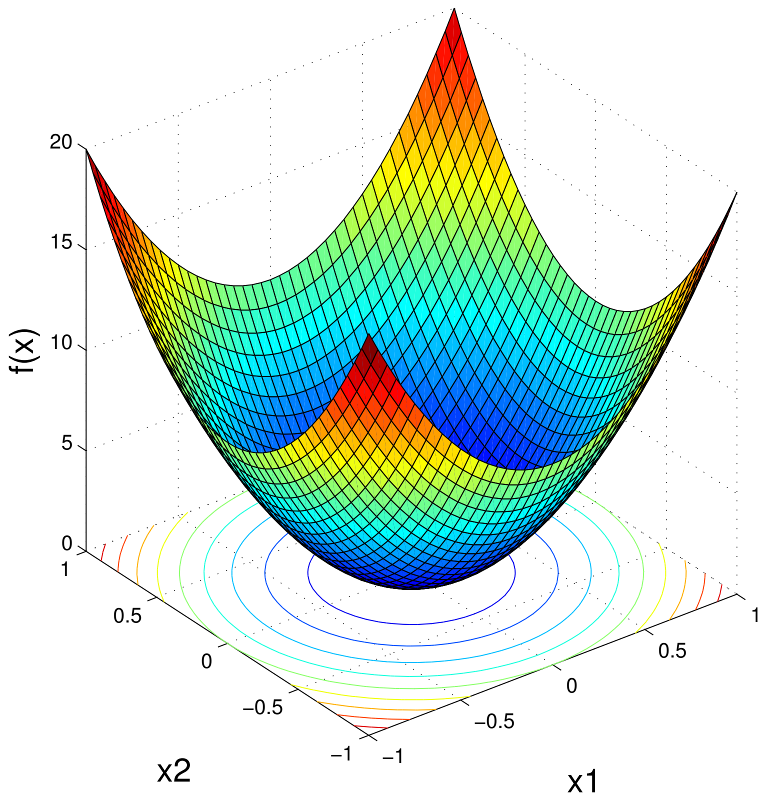
- Fact: f is convex if and only if $\nabla^2 f(\mathbf{x}) \succeq \mathbf{0}$ for all $\mathbf{x}$ in the problem domain

- example: consider the quadratic function

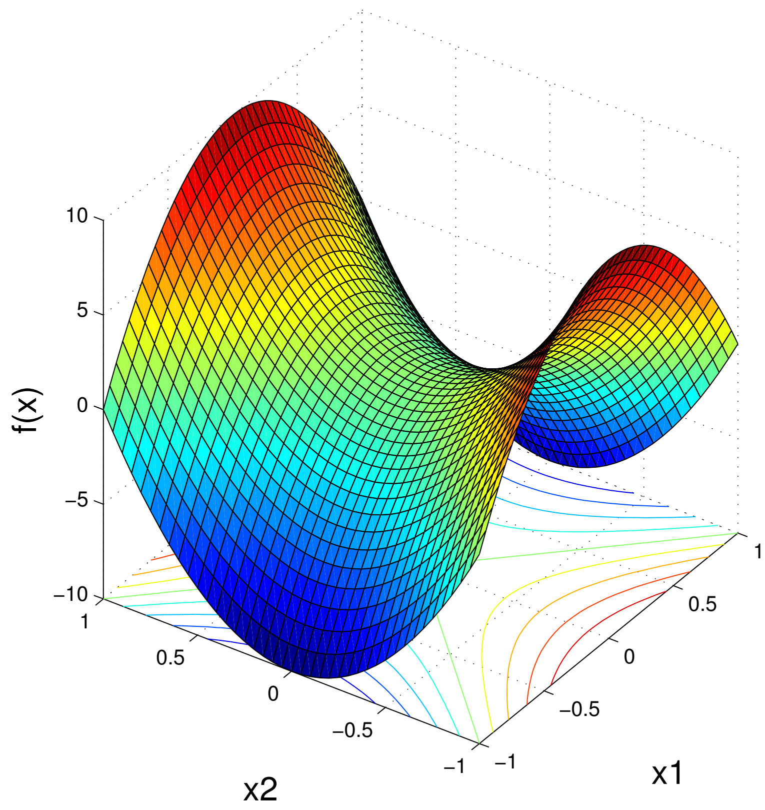
$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{R}\mathbf{x} + \mathbf{q}^T\mathbf{x} + c$$

It can be verified that $\nabla^2 f(\mathbf{x}) = \mathbf{R}$. Thus, f is convex if and only if $\mathbf{R} \succeq \mathbf{0}$

Illustration of Quadratic Functions



(a) PSD $\mathbf{A}$.



(b) indefinite $\mathbf{A}$.

Properties of PSD Matrices: Eigenvalues

Theorem 5.1. Let $\mathbf{A} \in \mathbb{S}^n$, and let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of $\mathbf{A}$. We have

1. $\mathbf{A} \succeq \mathbf{0} \Longleftrightarrow \lambda_i \geq 0$ for $i = 1, \ldots, n$
2. $\mathbf{A} \succ \mathbf{0} \Longleftrightarrow \lambda_i > 0$ for $i = 1, \ldots, n$

- proof: let $\mathbf{A} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T$ be the eigendecomposition of $\mathbf{A}$.

$$
\begin{aligned}
\mathbf{A} \succeq \mathbf{0} \quad & \Longleftrightarrow \quad \mathbf{x}^T\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T\mathbf{x} \geq 0, \quad \text{for all } \mathbf{x} \in \mathbb{R}^n \\
& \Longleftrightarrow \quad \mathbf{z}^T\boldsymbol{\Lambda}\mathbf{z} \geq 0, \quad \text{for all } \mathbf{z} \in \mathcal{R}(\mathbf{V}^T) = \mathbb{R}^n \\
& \Longleftrightarrow \quad \textstyle\sum_{i=1}^n \lambda_i |z_i|^2 \geq 0, \quad \text{for all } \mathbf{z} \in \mathbb{R}^n \\
& \Longleftrightarrow \quad \lambda_i \geq 0 \text{ for all } i
\end{aligned}
$$

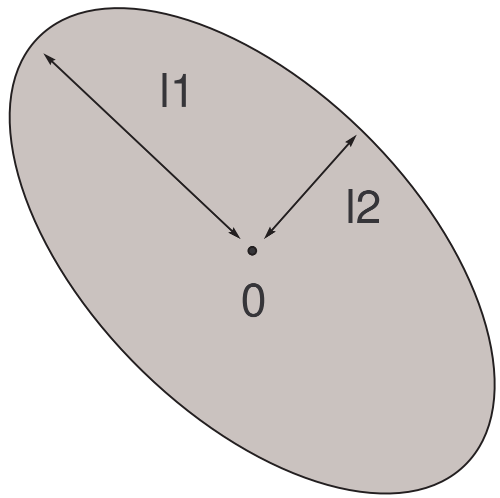
The PD case is proven by the same manner.

Example: Ellipsoid

- an ellipsoid of $\mathbb{R}^n$ centered at $\mathbf{0}$ is defined as

$$\mathcal{E} = \{ \mathbf{x} \in \mathbb{R}^n \mid \mathbf{x}^T \mathbf{P}^{-1} \mathbf{x} \leq 1 \},$$

  for some PD $\mathbf{P} \in \mathbb{S}^n$



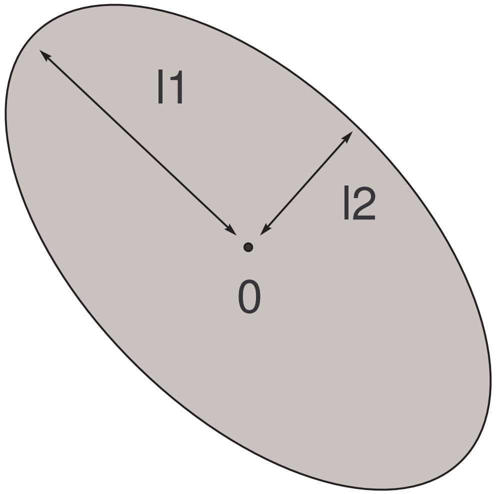
- let $\mathbf{P} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T$ be the eigendecomposition
  - $\mathbf{V}$ determines the directions of the semi-axes
  - $\lambda_1, \ldots, \lambda_n$ determine the lengths of the semi-axes
  - $\ell_i = \lambda_i^{\frac{1}{2}} \mathbf{v}_i$

Example: Ellipsoid

- an ellipsoid of $\mathbb{R}^n$ centered at $\mathbf{0}$ is defined as

$$\mathcal{E} = \{ \ \mathbf{x} \in \mathbb{R}^n \mid \mathbf{x}^T \mathbf{P}^{-1} \mathbf{x} \leq 1 \ \},$$

  for some PD $\mathbf{P} \in \mathbb{S}^n$



- note:
  - in direction $\mathbf{v}_1$, $\mathbf{x}^T \mathbf{P}^{-1} \mathbf{x}$ is large, hence ellipsoid is fat in direction $\mathbf{v}_1$
  - in direction $\mathbf{v}_n$, $\mathbf{x}^T \mathbf{P}^{-1} \mathbf{x}$ is small, hence ellipsoid is thin in direction $\mathbf{v}_n$
  - $\sqrt{\lambda_1 / \lambda_n}$ gives maximum eccentricity
- $\tilde{\mathcal{E}} = \{ \ \mathbf{x} \in \mathbb{R}^n \mid \mathbf{x}^T \mathbf{Q}^{-1} \mathbf{x} \leq 1 \ \}$, for some PD $\mathbf{Q} \in \mathbb{S}^n$, the $\mathcal{E} \supseteq \tilde{\mathcal{E}} \Longleftrightarrow \mathbf{P} \succeq \mathbf{Q}$

Example: Multivariate Gaussian Distribution

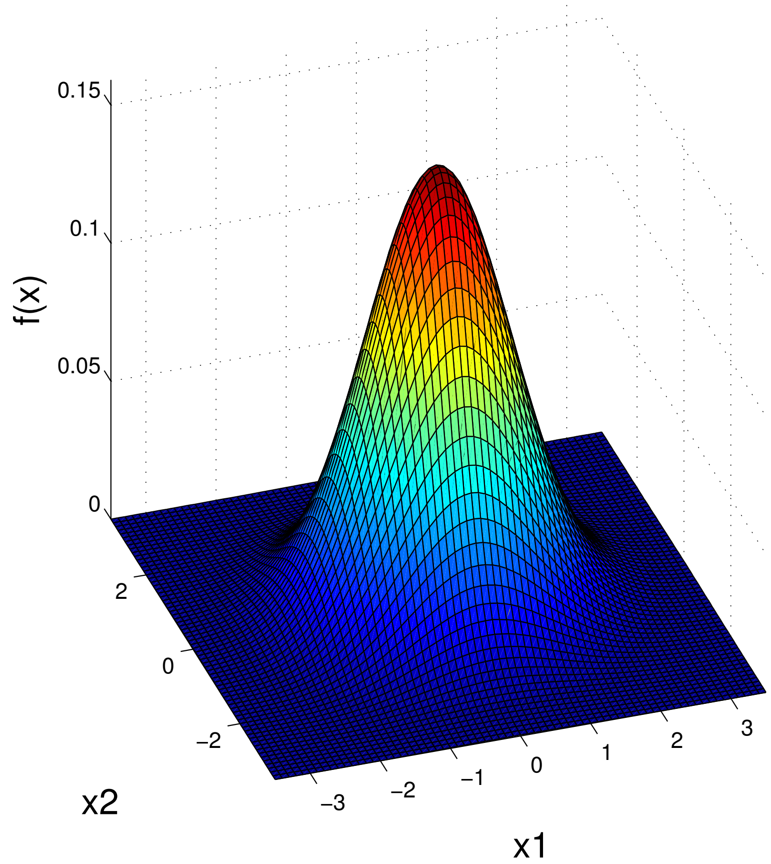
- probability density function for a Gaussian-distributed vector $\mathbf{x} \in \mathbb{R}^n$:

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}(\det(\mathbf{\Sigma}))^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

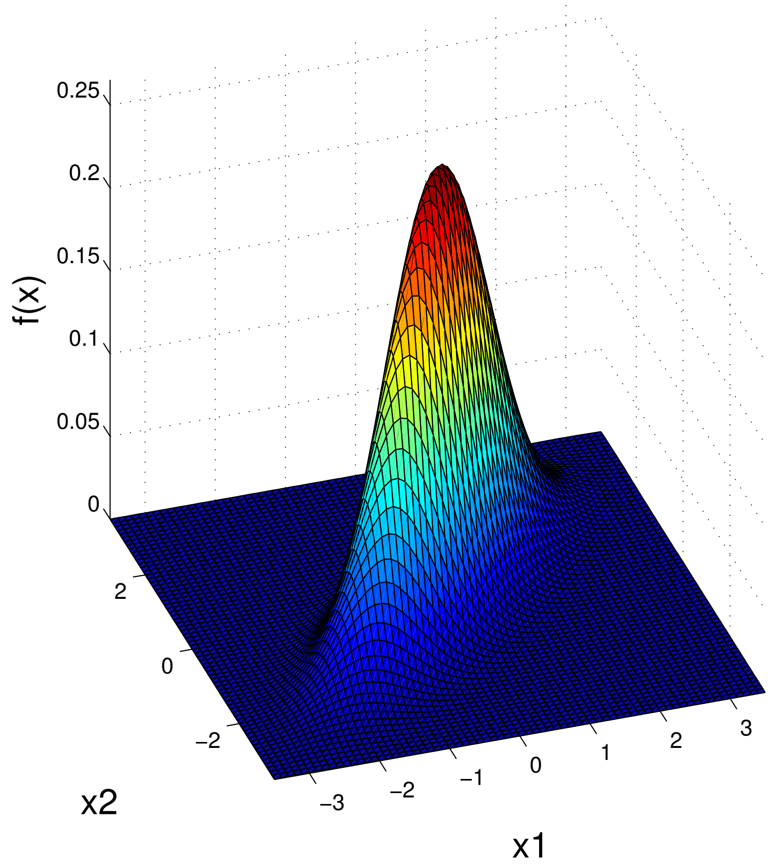
  where $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ are the mean and covariance of $\mathbf{x}$, resp.

  - $\mathbf{\Sigma}$ is PD
  - $\mathbf{\Sigma}$ determines how $\mathbf{x}$ is spread, by the same way as in ellipsoid

Example: Multivariate Gaussian Distribution



(a) $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.



(b) $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}$.

Properties of PSD Matrices

- it can be directly seen from the definition that
  - $\mathbf{A} \succeq \mathbf{0} \Longrightarrow a_{ii} \geq 0$ for all i
  - $\mathbf{A} \succ \mathbf{0} \Longrightarrow a_{ii} > 0$ for all i
- $\mathbf{A}$ is PSD, $\mathbf{x}^T \mathbf{A} \mathbf{x} = 0 \Longleftrightarrow \mathbf{A}\mathbf{x} = 0$ for a $\mathbf{x}$. ($\mathbf{A}$ is PD $\Longleftrightarrow$ $\mathbf{A}$ is nonsingular.)
- extension (also direct): partition $\mathbf{A}$ as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}.$$

  Then, $\mathbf{A} \succeq \mathbf{0} \Longrightarrow \mathbf{A}_{11} \succeq \mathbf{0}, \mathbf{A}_{22} \succeq \mathbf{0}$. Also, $\mathbf{A} \succ \mathbf{0} \Longrightarrow \mathbf{A}_{11} \succ \mathbf{0}, \mathbf{A}_{22} \succ \mathbf{0}$

- further extension:
  - a principal submatrix of $\mathbf{A}$, denoted by $\mathbf{A}_{\mathcal{I}}$, where $\mathcal{I} = \{i_1, \ldots, i_m\} \subseteq \{1, \ldots, n\}$, $m < n$, is a submatrix obtained by keeping only the rows and columns indicated by $\mathcal{I}$; i.e., $[\mathbf{A}_{\mathcal{I}}]_{jk} = a_{i_j, i_k}$ for all $j, k \in \{1, \ldots, m\}$
  - if $\mathbf{A}$ is PSD (resp. PD), then any principal submatrix of $\mathbf{A}$ is PSD (resp. PD)

Properties of PSD Matrices

Property 5.1. Let $\mathbf{A} \in \mathbb{S}^n$, $\mathbf{B} \in \mathbb{R}^{n \times m}$, and

$$\mathbf{C} = \mathbf{B}^T \mathbf{A} \mathbf{B}.$$

We have the following properties:

1. $\mathbf{A} \succeq \mathbf{0} \Longrightarrow \mathbf{C} \succeq \mathbf{0}$ (specially, $\mathbf{A} \succ \mathbf{0} \Longrightarrow \mathbf{C} \succeq \mathbf{0}$)
2. suppose $\mathbf{A} \succ \mathbf{0}$. It holds that $\mathbf{C} \succ \mathbf{0} \Longleftrightarrow$ $\mathbf{B}$ has full column rank
3. suppose $\mathbf{B}$ is nonsingular. It holds that $\mathbf{A} \succ \mathbf{0} \Longleftrightarrow \mathbf{C} \succ \mathbf{0}$, and that $\mathbf{A} \succeq \mathbf{0} \Longleftrightarrow \mathbf{C} \succeq \mathbf{0}$.

- proof sketch: the 1st property is trivial. For the 2nd property, observe

$$\mathbf{C} \succ \mathbf{0} \Longleftrightarrow \mathbf{z}^T \mathbf{A} \mathbf{z} > 0, \ \forall \ \mathbf{z} \in \mathcal{R}(\mathbf{B}) \setminus \{\mathbf{0}\}. \qquad (*)$$

  If $\mathbf{A} \succ \mathbf{0}$, $(*)$ reduces to $\mathbf{C} \succ \mathbf{0} \Longleftrightarrow \mathbf{B}\mathbf{x} \neq \mathbf{0}, \ \forall \ \mathbf{x} \neq \mathbf{0}$ (or $\mathbf{B}$ has full column rank). The 3rd property is proven by the similar manner.

Properties of PSD Matrices: Symmetric Factorization

Theorem 5.2. A matrix $\mathbf{A} \in \mathbb{S}^n$ is PSD if and only if it can be factored as

$$\mathbf{A} = \mathbf{B}^T\mathbf{B}$$

for some $\mathbf{B} \in \mathbb{R}^{m\times n}$.

- proof:
  - sufficiency: $\mathbf{A} = \mathbf{B}^T\mathbf{B} \Longrightarrow \mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{x}^T\mathbf{B}^T\mathbf{B}\mathbf{x} = \|\mathbf{B}\mathbf{x}\|_2^2 \geq 0$ for all $\mathbf{x}$
  - necessity: let $\mathbf{\Lambda}^{1/2} = \mathrm{Diag}(\lambda_1^{1/2}, \ldots, \lambda_n^{1/2})$ with $\lambda_i \geq 0$.

    $$\mathbf{A} \succeq \mathbf{0} \Longrightarrow \mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T = (\mathbf{V}\mathbf{\Lambda}^{1/2})(\mathbf{\Lambda}^{1/2}\mathbf{V}^T), \text{ with } \mathbf{\Lambda}^{1/2}\mathbf{V}^T \text{ being real}$$

- corollary: $\mathbf{A}\mathbf{x} = \mathbf{0} \Longleftrightarrow \mathbf{B}\mathbf{x} = \mathbf{0}$, so $\mathcal{N}(\mathbf{A}) = \mathcal{N}(\mathbf{B})$ and $\mathrm{rank}(\mathbf{A}) = \mathrm{rank}(\mathbf{B})$
- corollary: $\mathbf{A} \in \mathbb{R}^{n\times n}$ is PSD with $\mathrm{rank}(\mathbf{A}) = r$ if and only if there exists a $\mathbf{B}$ with $\mathrm{rank}(\mathbf{B}) = r$ such that $\mathbf{A} = \mathbf{B}^T\mathbf{B}$.
- corollary: $\mathbf{A} \in \mathbb{R}^{n\times n}$ is PD if and only if there exists a nonsingular $\mathbf{B} \in \mathbb{R}^{n\times n}$ such that $\mathbf{A} = \mathbf{B}^T\mathbf{B}$.
  - While $\mathbf{B}$ is not unique, there exists one and only one upper-triangular matrix $\mathbf{R}$ with $r_{ii} > 0$ s.t. $\mathbf{A} = \mathbf{R}^T\mathbf{R}$, which is the Cholesky factorization of $\mathbf{A}$.
  - $\mathbf{A}$ has an LU (or LDL) factorization with all pivots being positive.

Properties of PSD Matrices: Symmetric Factorization

- the factorization $\mathbf{A} = \mathbf{B}^T\mathbf{B}$ has *non-unique* factor $\mathbf{B}$
  - for any orthogonal $\mathbf{U} \in \mathbb{R}^{n\times n}$, $\mathbf{B} = \mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{V}^T$ is a factor for $\mathbf{A} = \mathbf{B}^T\mathbf{B}$
  - for any orthogonal $\mathbf{U} \in \mathbb{R}^{n\times n}$, $\mathbf{B} = \mathbf{U}\mathbf{R}$ is a factor for $\mathbf{A} = \mathbf{B}^T\mathbf{B}$

- denote
$$\mathbf{A}^{1/2} = \mathbf{V}\mathbf{\Lambda}^{1/2}\mathbf{V}^T.$$
  - $\mathbf{B} = \mathbf{A}^{1/2}$ is a factor for $\mathbf{A} = \mathbf{B}^T\mathbf{B}$
  - $\mathbf{A}^{1/2}$ is also a symmetric factor
  - $\mathbf{A}^{1/2}$ is the *unique PSD* factor for $\mathbf{A} = \mathbf{B}^T\mathbf{B}$

- $\mathbf{A}^{1/2}$ is called the PSD square root of $\mathbf{A}$
  - note: in general, a matrix $\mathbf{B} \in \mathbb{R}^{n\times n}$ is said to be a square root of another matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ if $\mathbf{A} = \mathbf{B}^2$

Properties for Symmetric Factorization

Property 5.2. Let $\mathbf{A} \in \mathbb{R}^{m\times k}$ and $\mathbf{B} \in \mathbb{R}^{k\times n}$, and suppose that $\mathbf{B}$ has full row rank. Then

$$\mathcal{R}(\mathbf{AB}) = \mathcal{R}(\mathbf{A})$$

- proof:
  - observe that $\dim \mathcal{R}(\mathbf{B}) = \mathrm{rank}(\mathbf{B}) = k$, which implies $\mathcal{R}(\mathbf{B}) = \mathbb{R}^k$.
  - we have $\mathcal{R}(\mathbf{AB}) = \{\mathbf{y} = \mathbf{Az} \mid \mathbf{z} \in \mathcal{R}(\mathbf{B})\} = \{\mathbf{y} = \mathbf{Az} \mid \mathbf{z} \in \mathbb{R}^k\} = \mathcal{R}(\mathbf{A})$.
- corollary: if $\mathbf{R}$ is a PSD matrix with factorization $\mathbf{R} = \mathbf{BB}^T$ for some full-column rank $\mathbf{B}$, then $\mathcal{R}(\mathbf{R}) = \mathcal{R}(\mathbf{B})$.

Properties for Symmetric Factorization

Property 5.3. Let $\mathbf{B} \in \mathbb{R}^{n\times k}, \mathbf{C} \in \mathbb{R}^{n\times k}$ be full-column rank matrices. It holds that

$$\mathbf{B}\mathbf{B}^T = \mathbf{C}\mathbf{C}^T \quad \Longleftrightarrow \quad \mathbf{C} = \mathbf{B}\mathbf{Q} \text{ for some orthogonal } \mathbf{Q} \in \mathbb{R}^{k\times k}$$

- proof: we consider "$\Longrightarrow$" only, as "$\Longleftarrow$" is trivial
  - suppose $\mathbf{B}\mathbf{B}^T = \mathbf{C}\mathbf{C}^T$.
  - from

    $$\mathbf{I} = (\mathbf{B}^\dagger\mathbf{B})(\mathbf{B}^\dagger\mathbf{B})^T = \mathbf{B}^\dagger(\mathbf{B}\mathbf{B}^T)(\mathbf{B}^\dagger)^T = \mathbf{B}^\dagger(\mathbf{C}\mathbf{C}^T)(\mathbf{B}^\dagger)^T = (\mathbf{B}^\dagger\mathbf{C})(\mathbf{B}^\dagger\mathbf{C})^T,$$

    we see that $\mathbf{B}^\dagger\mathbf{C}$ is orthogonal (note that $\mathbf{B}^\dagger\mathbf{C}$ is square).
  - let $\mathbf{Q} = \mathbf{B}^\dagger\mathbf{C}$. We have $\mathbf{B}\mathbf{Q} = \mathbf{B}\mathbf{B}^\dagger\mathbf{C} = \mathbf{P}_\mathbf{B}\mathbf{C}$, or equivalently,

    $$\mathbf{B}\mathbf{q}_i = \Pi_{\mathcal{R}(\mathbf{B})}(\mathbf{c}_i), \quad i = 1, \ldots, k.$$
  - from Property 5.2 we see that $\mathcal{R}(\mathbf{B}) = \mathcal{R}(\mathbf{B}\mathbf{B}^T) = \mathcal{R}(\mathbf{C}\mathbf{C}^T) = \mathcal{R}(\mathbf{C})$. It follows that $\Pi_{\mathcal{R}(\mathbf{B})}(\mathbf{c}_i) = \mathbf{c}_i$ for all i.

PSD Matrix Inequalities

- the notion of PSD matrices can be used to define inequalities for matrices
- PSD matrix inequalities are frequently used in topics like semidefinite programming (SDP)
- definition:
  - $\mathbf{A} \succeq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is PSD
  - $\mathbf{A} \succ \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is PD
  - $\mathbf{A} \not\succeq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is indefinite
- results that immediately follow from the definition: let $\mathbf{A}, \mathbf{B}, \mathbf{C} \in \mathbb{S}^n$.
  - $\mathbf{A} \succeq \mathbf{0}, \alpha \geq 0$ (resp. $\mathbf{A} \succ \mathbf{0}, \alpha > 0$) $\Longrightarrow \alpha\mathbf{A} \succeq \mathbf{0}$ (resp. $\alpha\mathbf{A} \succ \mathbf{0}$)
  - $\mathbf{A}, \mathbf{B} \succeq \mathbf{0}$ (resp. $\mathbf{A} \succeq \mathbf{0}, \mathbf{B} \succ \mathbf{0}$) $\Longrightarrow \mathbf{A} + \mathbf{B} \succeq \mathbf{0}$ (resp. $\mathbf{A} + \mathbf{B} \succ \mathbf{0}$)
  - $\mathbf{A} \succeq \mathbf{B}, \mathbf{B} \succeq \mathbf{C}$ (resp. $\mathbf{A} \succeq \mathbf{B}, \mathbf{B} \succ \mathbf{C}$) $\Longrightarrow \mathbf{A} \succeq \mathbf{C}$ (resp. $\mathbf{A} \succ \mathbf{C}$)
  - $\mathbf{A} \not\succeq \mathbf{B}$ does not imply $\mathbf{B} \succeq \mathbf{A}$

PSD Matrix Inequalities

- more results: let $\mathbf{A}, \mathbf{B} \in \mathbb{S}^n$.
  - $\mathbf{A} \succeq \mathbf{B} \Longrightarrow \lambda_k(\mathbf{A}) \geq \lambda_k(\mathbf{B})$ for all k; the converse is not always true
  - $\mathbf{A} \succeq \mathbf{I}$ (resp. $\mathbf{A} \succ \mathbf{I}$) $\Longleftrightarrow \lambda_k(\mathbf{A}) \geq 1$ for all k (resp. $\lambda_k(\mathbf{A}) > 1$ for all k)
  - $\mathbf{I} \succeq \mathbf{A}$ (resp. $\mathbf{I} \succ \mathbf{A}$) $\Longleftrightarrow \lambda_k(\mathbf{A}) \leq 1$ for all k (resp. $\lambda_k(\mathbf{A}) < 1$ for all k)
  - if $\mathbf{A}, \mathbf{B} \succ \mathbf{0}$ then $\mathbf{A} \succeq \mathbf{B} \Longleftrightarrow \mathbf{B}^{-1} \succeq \mathbf{A}^{-1}$

- some results as consequences of the above results:
  - for $\mathbf{A} \succeq \mathbf{B} \succeq \mathbf{0}$, $\det(\mathbf{A}) \geq \det(\mathbf{B})$
  - for $\mathbf{A} \succeq \mathbf{B}$, $\operatorname{tr}(\mathbf{A}) \geq \operatorname{tr}(\mathbf{B})$
  - for $\mathbf{A} \succeq \mathbf{B} \succ \mathbf{0}$, $\operatorname{tr}(\mathbf{A}^{-1}) \leq \operatorname{tr}(\mathbf{B}^{-1})$

PSD Matrix Inequalities

- the Schur complement: let

$$\mathbf{X} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{C} \end{bmatrix},$$

  where $\mathbf{A} \in \mathbb{S}^m$, $\mathbf{B} \in \mathbb{R}^{m \times n}$, $\mathbf{C} \in \mathbb{S}^n$ with $\mathbf{C} \succ \mathbf{0}$. Let

$$\mathbf{S}_C = \mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^T,$$

  which is called the Schur complement of $\mathbf{C}$.

- We have

$$\mathbf{X} \succeq \mathbf{0} \text{ (resp. } \mathbf{X} \succ \mathbf{0}) \quad \Longleftrightarrow \quad \mathbf{S}_C \succeq \mathbf{0} \text{ (resp. } \mathbf{S}_C \succ \mathbf{0})$$

  – example: let $\mathbf{C}$ be PD. By the Schur complement,

$$1 - \mathbf{b}^T\mathbf{C}^{-1}\mathbf{b} \geq 0 \quad \Longleftrightarrow \quad \mathbf{C} - \mathbf{b}\mathbf{b}^T \succeq \mathbf{0}$$

  (prove by yourself)

Application: Spectral Analysis

- consider the complex harmonic time-series

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

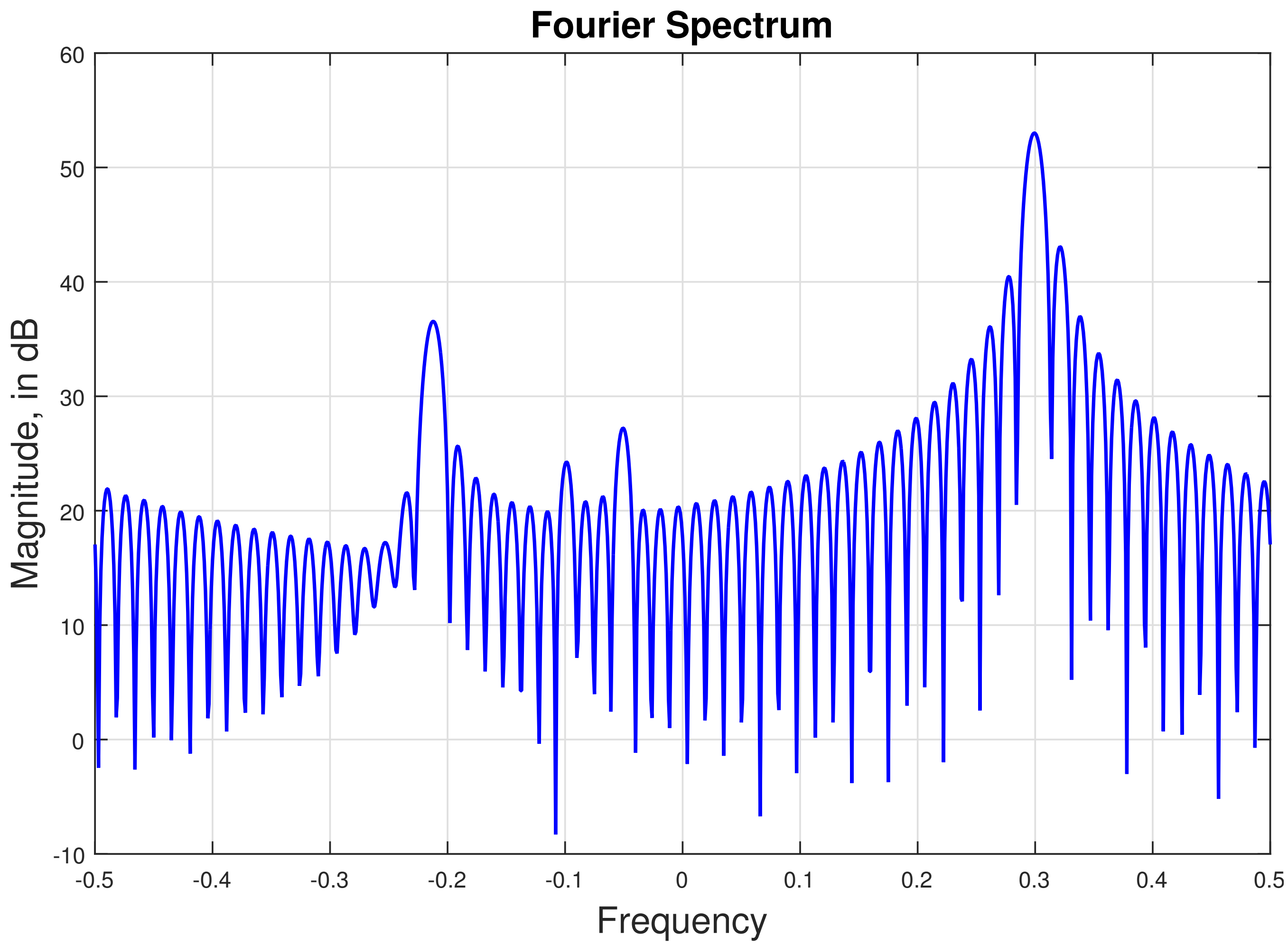
  where $\alpha_i \in \mathbb{C}$ is the amplitude-phase coefficient of the ith sinusoid; $f_i \in \left[-\frac{1}{2}, \frac{1}{2}\right)$ is the frequency of the ith sinusoid; w_t is noise; T is the observation time length

- 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"

- suggested reading: **[Stoica-Moses'97]**

Application: Spectral Analysis



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

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

Spectral Analysis via Subspace: Formulation

- 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 $\mathbf{y}_t$. We have

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

- (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 $T_d \to \infty$,

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

Spectral Analysis via Subspace: Formulation

- let us summarize
- Model: the correlation matrix $\mathbf{R}_y = \frac{1}{T_d}\mathbf{Y}\mathbf{Y}^H$ is modeled as

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

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

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

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

- observation: $\mathbf{A}$ and $\mathbf{S}$ are both Vandemonde

Spectral Analysis via Subspace: Subspace Properties

- Assumptions: i) $\alpha_i \neq 0$ for all i, ii) $f_i \neq f_j$ for all $i \neq j$, iii) $d > k$, iv) $T_d \geq k$
- results:
  - $\mathbf{A}$ 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 $\mathbf{S}^H \mathbf{D}^H \mathbf{x} = \mathbf{0}$ if and only if $\mathbf{S}^H$ does not have full column rank
  - $\mathcal{R}(\mathbf{A}\mathbf{\Phi}\mathbf{A}^H) = \mathcal{R}(\mathbf{A})$, by Property 5.2
  - $\operatorname{rank}(\mathbf{A}\mathbf{\Phi}\mathbf{A}^H) = \operatorname{rank}(\mathbf{A}) = k$, thus $\mathbf{A}\mathbf{\Phi}\mathbf{A}^H$ has k nonzero eigenvalues

Spectral Analysis via Subspace: Subspace Properties

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

- since $\lambda_i > 0$ for $i = 1, \ldots, k$ and $\lambda_i = 0$ for $i = k+1, \ldots, d$,

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

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

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

Spectral Analysis via Subspace: Subspace Properties

- consider the eigendecomposition of $\mathbf{R}_y$. Observe

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

- results:
  - $\mathbf{V}(\mathbf{\Lambda} + \sigma^2 \mathbf{I})\mathbf{V}^H$ 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 via Subspace: Subspace Properties

- let us summarize
- 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 $f \in \{f_1, \ldots f_k\}$ 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 $f \in \{f_1, \ldots f_k\}$?

- The answer is yes if $d > k$. The following matrix result gives the answer.

  Theorem 5.3. Let $\mathbf{A} \in \mathbb{C}^{d \times k}$ any Vandemonde matrix with distinct roots $z_1, \ldots, z_k$ and with $d \geq k+1$. Then it holds that

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

Spectral Analysis via Subspace: Subspace Properties

- proof of Theorem 5.3: "$\Longrightarrow$" is trivial, and we consider "$\Longleftarrow$"
  - suppose there exists $\bar{z} \notin \{z_1, \ldots, z_k\}$ such that $\mathbf{a}(\bar{z}) \in \mathcal{R}(\mathbf{A})$.
  - let $\tilde{\mathbf{A}} = [\ \mathbf{a}(\bar{z})\ \mathbf{A}\ ] \in \mathbb{C}^{d \times (k+1)}$.
  - $\mathbf{a}(\bar{z}) \in \mathcal{R}(\mathbf{A})$ implies that $\tilde{\mathbf{A}}$ has linearly dependent columns
  - however, $\tilde{\mathbf{A}}$ is Vandemonde with distinct roots $\bar{z}, z_1, \ldots, z_k$, and for $d \geq k+1$ $\tilde{\mathbf{A}}$ must have linearly independent columns—a contradiction

Spectral Analysis via Subspace: Algorithm

- there are many subspace methods, and multiple signal classification (MUSIC) is most well-known

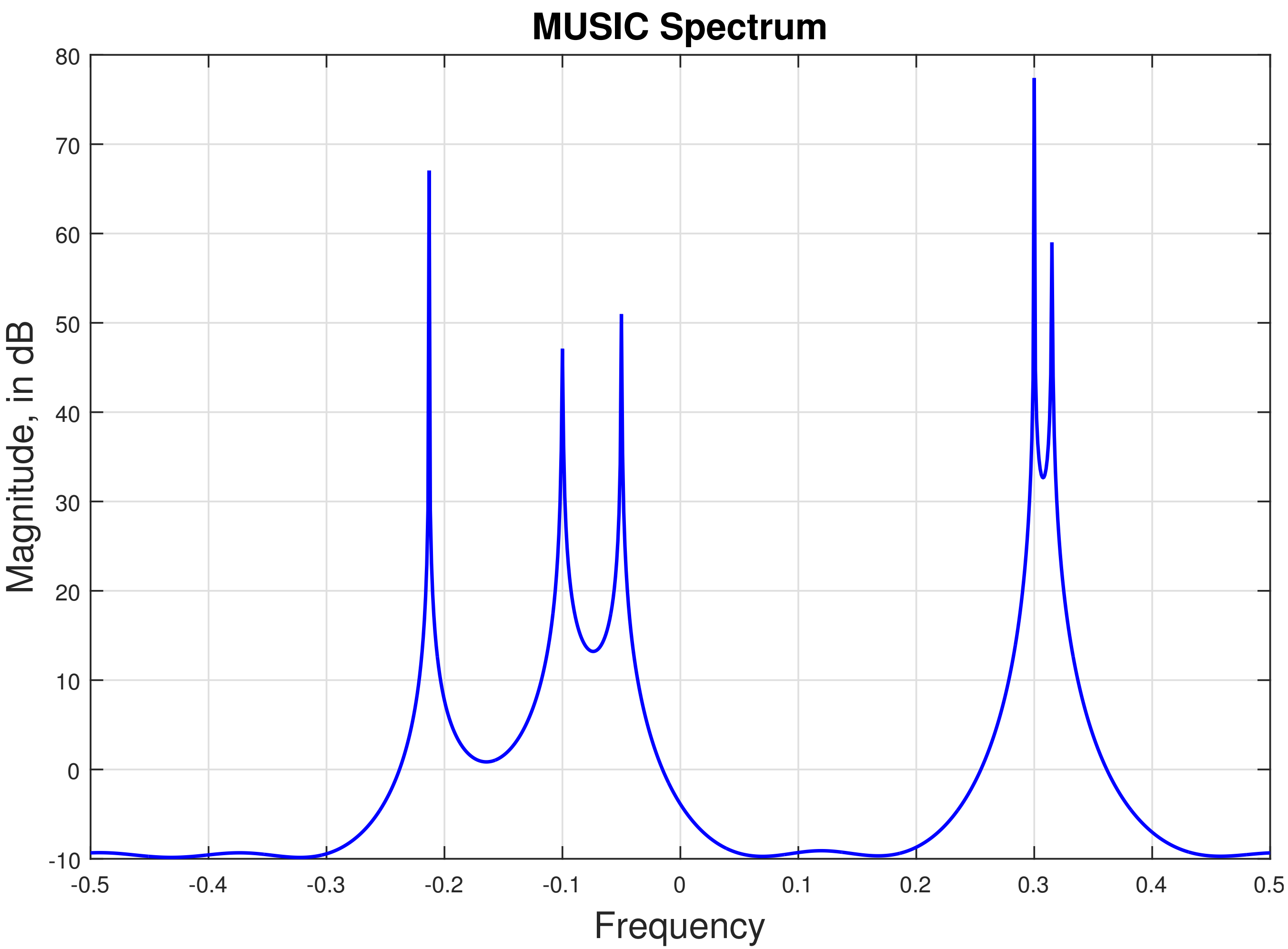
- MUSIC uses the fact that $\mathbf{a}(e^{\boldsymbol{j}2\pi f}) \in \mathcal{R}(\mathbf{A}) \Longleftrightarrow \mathbf{V}_2^H \mathbf{a}(e^{\boldsymbol{j}2\pi f}) = \mathbf{0}$

Algorithm: MUSIC
input: the correlation matrix $\mathbf{R}_y \in \mathbb{C}^{d \times d}$ and the model order $k < d$
Perform eigendecomposition $\mathbf{R}_y = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H$ 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 via Subspace: Algorithm

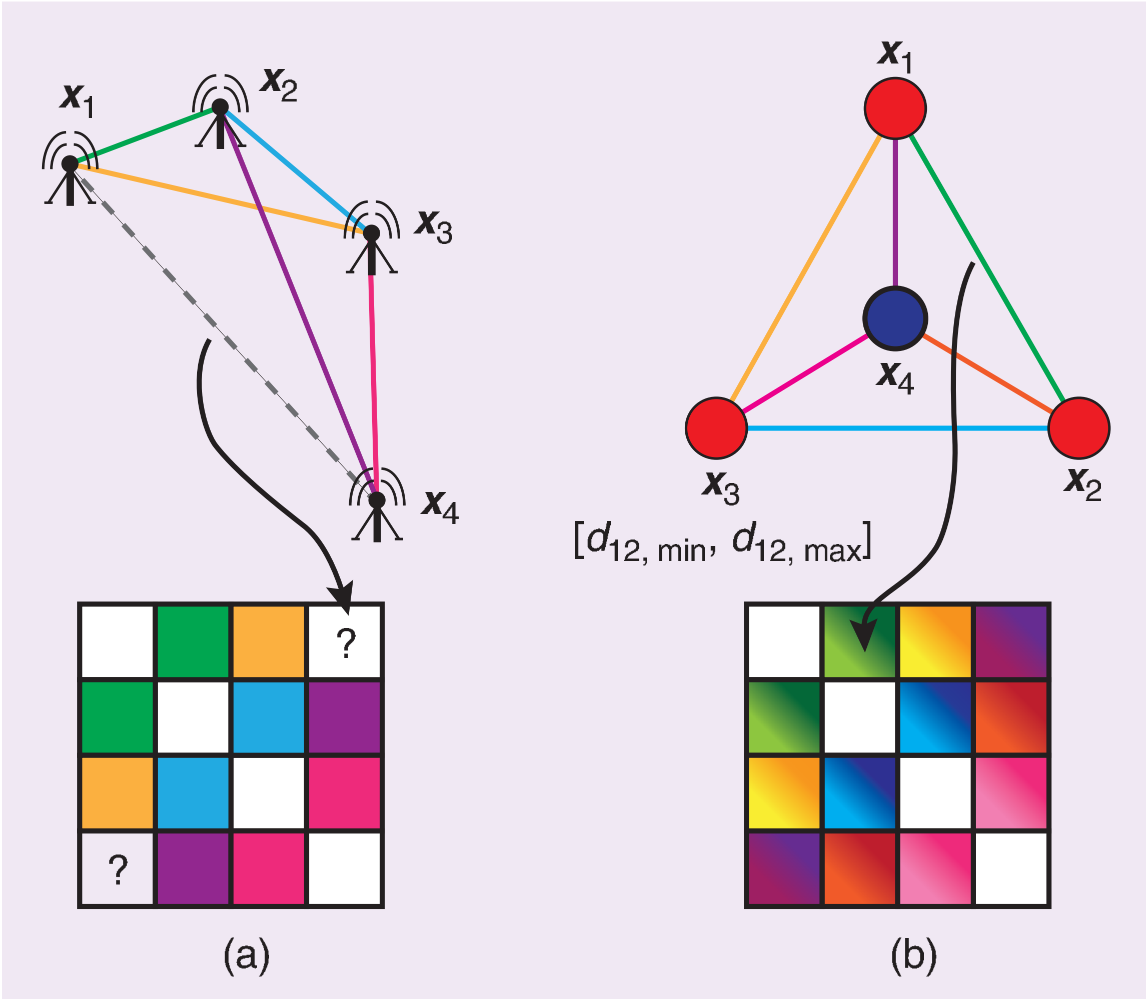


An illustration of the MUSIC spectrum. $T = 64$, $k = 5$, $\{f_1, \ldots, f_k\} = \{-0.213, -0.1, -0.05, 0.3, 0.315\}$.

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}'s for all $i, j \in \{1, \ldots, n\}$, recover $\mathbf{X}$
  - this problem is called the Euclidean distance matrix (EDM) problem
- applications: sensor network localization (SNL), molecular conformation, ...
- suggested reading: **[Dokmanić-Parhizkar-et al.'15]**

EDM Applications



(a) SNL. (b) Molecular transformation. Source: **[Dokmanić-Parhizkar-et al.'15]**

EDM: Formulation

- let $\mathbf{R} \in \mathbb{S}^n$ be matrix whose entries are $r_{ij} = d_{ij}^2$ for all i, j
- from
$$r_{ij} = d_{ij}^2 = \|\mathbf{x}_i\|_2^2 - 2\mathbf{x}_i^T\mathbf{x}_j + \|\mathbf{x}_j\|_2^2,$$
we see that $\mathbf{R}$ can be written as
$$\mathbf{R} = \mathbf{1}(\mathrm{diag}(\mathbf{X}^T\mathbf{X}))^T - 2\mathbf{X}^T\mathbf{X} + (\mathrm{diag}(\mathbf{X}^T\mathbf{X}))\mathbf{1}^T \tag{$*$}$$
where the notation diag means that $\mathrm{diag}(\mathbf{Y}) = [y_{11}, \ldots, y_{nn}]^T$ for any square $\mathbf{Y}$
- observation: $(*)$ also holds if we replace $\mathbf{X}$ by
  - $\tilde{\mathbf{X}} = [\mathbf{x}_1 + \mathbf{b}, \ldots, \mathbf{x}_n + \mathbf{b}]$ for any $\mathbf{b} \in \mathbb{R}^d$ ($d_{ij} = \|\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j\|_2$ is also true)
  - $\tilde{\mathbf{X}} = \mathbf{Q}\mathbf{X}$ 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

- assume $\mathbf{x}_1 = \mathbf{0}$ w.l.o.g. Then,

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

- construct from $\mathbf{R}$ the following matrix

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

  We have

$$\mathbf{G} = \mathbf{X}^T\mathbf{X}$$

- idea: do a symmetric factorization for $\mathbf{G}$ to try to recover $\mathbf{X}$

EDM: Method

- assumption: $\mathbf{X}$ has full row rank
- $\mathbf{G}$ is PSD and has $\mathrm{rank}(\mathbf{G}) = d$
- denote the eigendecomposition of $\mathbf{G}$ as $\mathbf{G} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T$. Assuming $\lambda_1 \geq \ldots \geq \lambda_n$, it takes the form

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

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

- EDM solution: take $\hat{\mathbf{X}} = \boldsymbol{\Lambda}^{1/2}\mathbf{V}_1^T$ as an estimate of $\mathbf{X}$
- recovery guarantee: by Property 5.3, we have $\hat{\mathbf{X}} = \mathbf{Q}\mathbf{X}$ for some orthogonal $\mathbf{Q}$

EDM: Further Discussion

- in applications such as SNL, not all pairwise distances d_{ij}'s are available
- or, there are missing entries with $\mathbf{R}$
- possible solution: apply low-rank matrix completion to try to recover the full $\mathbf{R}$
- to use low-rank matrix completion, we need to know a rank bound on $\mathbf{R}$
- by the result $\mathrm{rank}(\mathbf{A}+\mathbf{B}) \leq \mathrm{rank}(\mathbf{A}) + \mathrm{rank}(\mathbf{B})$, we get

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

- other issues: noisy distance measurements, resolving the orthogonal rotation problem with $\hat{\mathbf{X}}$. See the suggested reference **[Dokmanić-Parhizkar-et al.'15]**.

References

[Brodie-Daubechies-et al.'09] J. Brodie, I. Daubechies, C. De Mol, D. Giannone, and I. Loris, "Sparse and stable Markowitz portfolios," *Proceedings of the National Academy of Sciences*, vol. 106, no. 30, pp. 12267–12272, 2009.

[Stoica-Moses'97] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*, Prentice Hall, 1997.

[Dokmanić-Parhizkar-et al.'15] I. Dokmanić, R. Parhizkar, J. Ranieri, and Vetterli, "Euclidean distance matrices," *IEEE Signal Processing Magazine*, vol. 32, no. 6, pp. 12–30, Nov. 2015.