More Results from Courant-Fischer (cont'd)

Let $\mathbf{A}, \mathbf{B} \in \mathbb{S}^n$, $\mathbf{z} \in \mathbb{R}^n$

- (Interlacing) $\lambda_{k+1}(\mathbf{A}) \leq \lambda_k(\mathbf{A} \pm \mathbf{z}\mathbf{z}^T) \leq \lambda_{k-1}(\mathbf{A})$ for proper k

$$\lambda_k(A \pm zz^T) = \min_{\substack{S \subseteq \mathbb{R}^n \\ \dim(S)=n-k+1}} \max_{\substack{x \in S \\ \|x\|_2=1}} x^T(A \pm zz^T)x$$

$$\geq \min_{\substack{S \subseteq \mathbb{R}^n \\ \dim(S)=n-k+1}} \max_{\substack{x \in S \cap \operatorname{span}\{z\}^\perp \\ \|x\|_2=1}} \left(x^TAx \pm \underbrace{x^Tzz^Tx}_{=0}\right)$$

Note that $\dim(S \cap \operatorname{span}\{z\}^\perp) = \overset{n-k+1}{\dim(S)} + \overset{n-1}{\dim(\operatorname{span}\{z\}^\perp)} - \underbrace{\dim(S + \operatorname{span}\{z\}^\perp)}_{\leq n} \geq n-k$

It follows that

$$\lambda_k(A \pm zz^T) \geq \min_{\substack{S' \subseteq \mathbb{R}^n :\ \dim(S')=r \\ r \in [n-k,\ n]}} \max_{\substack{x \in S' \\ \|x\|_2=1}} x^TAx \geq \lambda_{k+1}(A)$$

From Courant-Fischer

$r = n-k \rightsquigarrow \lambda_{k+1}(A)$ $\quad r = n \rightsquigarrow \lambda_1(A)$

$r = n-k+1 \rightsquigarrow \lambda_k(A)$

More Results from Courant-Fischer (cont'd)

Let $\mathbf{A}, \mathbf{B} \in \mathbb{S}^n$, $\mathbf{z} \in \mathbb{R}^n$

- If $\mathrm{rank}(\mathbf{B}) \leq r$, then $\lambda_{k+r}(\mathbf{A}) \leq \lambda_k(\mathbf{A}+\mathbf{B}) \leq \lambda_{k-r}(\mathbf{A})$ for proper k

- (Weyl) $\lambda_{j+k-1}(\mathbf{A}+\mathbf{B}) \leq \lambda_j(\mathbf{A}) + \lambda_k(\mathbf{B})$ for proper j, k

- For any semi-orthogonal $\mathbf{U} \in \mathbb{R}^{n\times r}$,
  $\lambda_{k+n-r}(\mathbf{A}) \leq \lambda_k(\mathbf{U}^T\mathbf{A}\mathbf{U}) \leq \lambda_k(\mathbf{A})$ for proper k

Extend Variational Characterization to Sum of Eigenvalues

Theorem

For any $\mathbf{A} \in \mathbb{S}^n$,

U semi-orthogonal

$$\sum_{i=1}^{r} \lambda_i = \max_{\substack{\mathbf{U} \in \mathbb{R}^{n \times r} \\ \|\mathbf{u}_i\|_2 = 1 \ \forall i, \ \mathbf{u}_i^T \mathbf{u}_j = 0 \ \forall i \neq j}} \sum_{i=1}^{r} \mathbf{u}_i^T \mathbf{A} \mathbf{u}_i = \max_{\substack{\mathbf{U} \in \mathbb{R}^{n \times r} \\ \mathbf{U}^T \mathbf{U} = \mathbf{I}}} \operatorname{tr}(\mathbf{U}^T \mathbf{A} \mathbf{U})$$

$r = 1, \cdots, n$

- This can be proved using $\lambda_k(\mathbf{U}^T \mathbf{A} \mathbf{U}) \leq \lambda_k(\mathbf{A})$, but we may try another way of proof to get better understanding of trace, which uses the fact that

$$\max_{\substack{\mathbf{U} \in \mathbb{R}^{n \times r} \\ \mathbf{U}^T \mathbf{U} = \mathbf{I}}} \operatorname{tr}(\mathbf{U}^T \mathbf{A} \mathbf{U}) = \max_{\substack{\mathbf{U} \in \mathbb{R}^{n \times r} \\ \mathbf{U}^T \mathbf{U} = \mathbf{I}}} \operatorname{tr}(\mathbf{U}^T \Lambda \mathbf{U})$$

$$\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

Other Extensions

(Von Neumann) For any $\mathbf{A}, \mathbf{B} \in \mathbb{S}^n$,

$$\mathrm{tr}(\mathbf{A}\mathbf{B}) \leq \sum_{i=1}^{n} \lambda_i(\mathbf{A})\lambda_i(\mathbf{B})$$

(Lidskii) For any $\mathbf{A}, \mathbf{B} \in \mathbb{S}^n$ and any $1 \leq i_1 \leq i_2 \leq \cdots \leq i_k \leq n$,

$$\sum_{j=1}^{k} \lambda_{i_j}(\mathbf{A} + \mathbf{B}) \leq \sum_{j=1}^{k} \lambda_{i_j}(\mathbf{A}) + \sum_{j=1}^{k} \lambda_{i_j}(\mathbf{B})$$

Matrix Computations
Chapter 5: Positive Semidefinite Matrices
Section 5.1 Properties of Positive Semidefinite Matrices

Jie Lu
ShanghaiTech University

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

Facts:

- $\mathbf{x}^T \mathbf{A} \mathbf{x} = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j$
- It suffices to consider symmetric $\mathbf{A}$ because for general $\mathbf{A} \in \mathbb{R}^{n \times n}$,

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

- Complex case: The 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 $\mathbf{x} \in \mathbb{C}^n$

$$(x^H A x)^* \overset{\text{scalar}}{=} (x^H A x)^H = x^H A^H x \overset{A \in \mathbb{H}^n}{=} x^H A x$$

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 $\mathbf{x} \neq \mathbf{0}$
- negative semidefinite (NSD) if $-\mathbf{A}$ is PSD
- negative definite (ND) if $-\mathbf{A}$ is PD
- indefinite if $\mathbf{A}$ is neither PSD nor NSD

Notation:

- $\mathbf{A} \succeq \mathbf{0}$ means that $\mathbf{A}$ is PSD
- $\mathbf{A} \succ \mathbf{0}$ means that $\mathbf{A}$ is PD
- $\mathbf{A} \preceq \mathbf{0}$ means that $\mathbf{A}$ is NSD
- $\mathbf{A} \prec \mathbf{0}$ means that $\mathbf{A}$ is ND
- $\mathbf{A} \not\succeq \mathbf{0}$ or $\mathbf{A} \not\preceq \mathbf{0}$ means that $\mathbf{A}$ is indefinite

Example: Covariance Matrices

- Let $\mathbf{y}_0, \mathbf{y}_1, \ldots \mathbf{y}_{T-1} \in \mathbb{R}^n$ be multi-dimensional data samples
  - Examples: patches in image processing, multi-channel signals in signal processing, history of returns of assets in finance, etc.[1]
- 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

[1] 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.

Example: Hessian

- Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be a twice continuously differentiable function
- The Hessian (matrix) 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**: The Hessian of the quadratic function

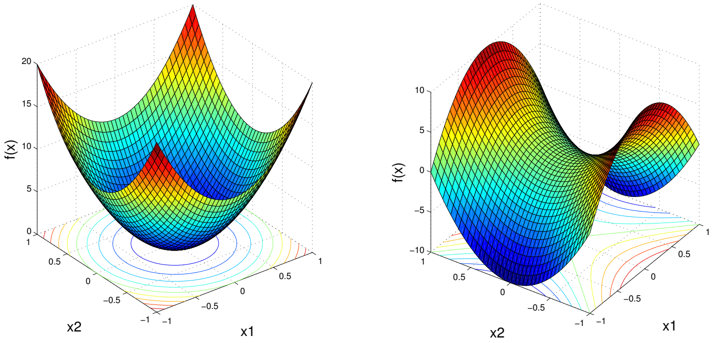
$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{R}\mathbf{x} + \mathbf{q}^T\mathbf{x} + c$$

is given by $\nabla^2 f(\mathbf{x}) = \mathbf{R}$

f is convex if and only if $\mathbf{R} \succeq \mathbf{0}$

Illustration of Quadratic Functions

$f(x) = x^T A x \qquad x \in \mathbb{R}^2$



(a) PSD **A**

(b) indefinite **A**

PSD Matrices and Eigenvalues

Theorem

Let $\mathbf{A} \in \mathbb{S}^n$ *with eigenvalues* $\lambda_1, \ldots, \lambda_n$. *Then,*

- $\mathbf{A} \succeq \mathbf{0} \iff \lambda_i \geq 0 \ \forall i = 1, \ldots, n$
- $\mathbf{A} \succ \mathbf{0} \iff \lambda_i > 0 \ \forall i = 1, \ldots, n$

V orthogonal

Proof: Let $\mathbf{A} = \mathbf{V}\Lambda\mathbf{V}^T$ be the eigendecomposition of $\mathbf{A}$ (always exists for $\mathbf{A} \in \mathbb{S}^n$)

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

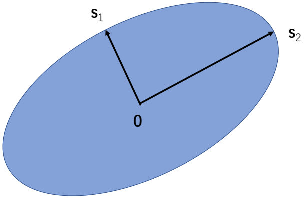
The PD case can be proved in the same way

Example: Ellipsoid

- An ellipsoid of $\mathbb{R}^n$ centered at the origin is defined as

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

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



- Let $\mathbf{P} = \mathbf{V}\Lambda\mathbf{V}^T$ be the eigendecomposition ($\mathbf{V}$ orthogonal). Then, each semi-axis of the ellipsoid is given by

$$\mathbf{s}_i = \lambda_i^{-\frac{1}{2}} \mathbf{v}_i$$

  - The orthonormal eigenvectors determine the directions of the semi-axes
  - The eigenvalues determine the lengths of the semi-axes

Example: Multivariate Gaussian Distribution

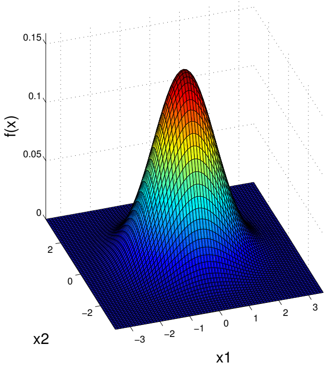
- Probability density function for a Gaussian-distributed vector $\mathbf{x} \in \mathbb{R}^n$:

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

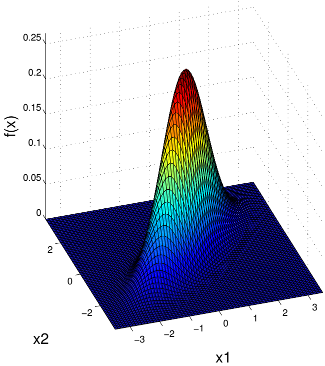
where $\boldsymbol{\mu}$ and Σ are the mean and covariance of $\mathbf{x}$, respectively

  - Σ is PD
  - Σ determines how $\mathbf{x}$ is spread

Example: Multivariate Gaussian Distribution (cont'd)



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



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

PSD Matrices and Square Root

Theorem

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

Proof:

- Sufficiency ($\Longleftarrow$): $\mathbf{A} = \mathbf{B}^T\mathbf{B} \Longrightarrow \mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{x}^T\mathbf{B}^T\mathbf{B}\mathbf{x} = \|\mathbf{B}\mathbf{x}\|_2^2 \geq 0$ for all $\mathbf{x}$
- Necessity ($\Longrightarrow$): Let $\mathbf{A} = \mathbf{V}\Lambda\mathbf{V}^T$

$$\Rightarrow \Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \succeq 0$$

$$\mathbf{A} \succeq \mathbf{0} \Longrightarrow \mathbf{A} = (\mathbf{V}\Lambda^{1/2})(\Lambda^{1/2}\mathbf{V}^T), \text{ where } \Lambda^{1/2} = \mathrm{Diag}(\lambda_1^{1/2}, \ldots, \lambda_n^{1/2})$$

where $\Lambda^{1/2}\mathbf{V}^T$ is real because Λ and $\mathbf{V}$ are real

PSD Matrices and Square Root (cont'd)

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

$$\mathbf{A}^{1/2} = \mathbf{V}\Lambda^{1/2}\mathbf{V}^T$$

$$\Lambda^{\frac{1}{2}} = \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{bmatrix}$$

  - $\mathbf{B} = \mathbf{A}^{1/2}$ 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 $\mathbf{A}$
  - 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 of PSD Matrices

It is straightforward to see 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

A straightforward extension: Partition $\mathbf{A}$ as

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

A_{11}, A_{22} square

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

Further extension:

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

Properties of PSD Matrices (cont'd)

Let $\mathbf{A} \in \mathbb{S}^n$, $\mathbf{B} \in \mathbb{R}^{n\times m}$, and $\mathbf{C} = \mathbf{B}^T\mathbf{AB}$. The following properties hold:

1. $\mathbf{A} \succeq \mathbf{0} \Longrightarrow \mathbf{C} \succeq \mathbf{0}$ For any $x \in \mathbb{R}^m$, $x^TCx = \underbrace{x^TB^T}A\underbrace{Bx}$
   $= (Bx)^TA(Bx) \geq 0$
2. With $\mathbf{A} \succ \mathbf{0}$,

$$\mathbf{C} \succ \mathbf{0} \Longleftrightarrow \mathbf{B} \text{ has full column rank}$$

3. With nonsingular $\mathbf{B}$,

$$\mathbf{A} \succ \mathbf{0} \Longleftrightarrow \mathbf{C} \succ \mathbf{0}, \quad \mathbf{A} \succeq \mathbf{0} \Longleftrightarrow \mathbf{C} \succeq \mathbf{0}$$

$\forall z \in \mathbb{R}^n$, $z \neq 0$,
$z^TAz > 0$

$\forall x \in \mathbb{R}^m$, $x \neq 0$,
$x^TCx = (Bx)^TA(Bx) > 0$

columns of B linearly independent

$\Updownarrow$

$Bx = 0$ if and only if $x = 0$

Properties for Symmetric Factorization

Property: Let $\mathbf{A} \in \mathbb{R}^{m\times k}$ and $\mathbf{B} \in \mathbb{R}^{k\times n}$. Suppose $\mathbf{B}$ 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 $\mathcal{R}(\mathbf{B}) = \mathbb{R}^k$
- $\mathcal{R}(\mathbf{AB}) = \{\mathbf{y} = \underset{=}{\mathbf{A}\mathbf{z}} \mid \mathbf{z} \in \mathcal{R}(\mathbf{B})\} = \{\mathbf{y} = \mathbf{Az} \mid \mathbf{z} \in \mathbb{R}^k\} = \mathcal{R}(\mathbf{A})$

$= \{y = ABx \mid x \in \mathbb{R}^n\}$

Corollary: Let $\mathbf{R}$ be a PSD matrix. Suppose $\mathbf{R} = \mathbf{B}\mathbf{B}^T$ for some full-column rank $\mathbf{B}$. Then, $\mathcal{R}(\mathbf{R}) = \mathcal{R}(\mathbf{B})$

Property: Suppose $\mathbf{B}, \mathbf{C} \in \mathbb{R}^{n\times k}$ have full column rank. Then,

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

The proof needs pseudo inverse (later)

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

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

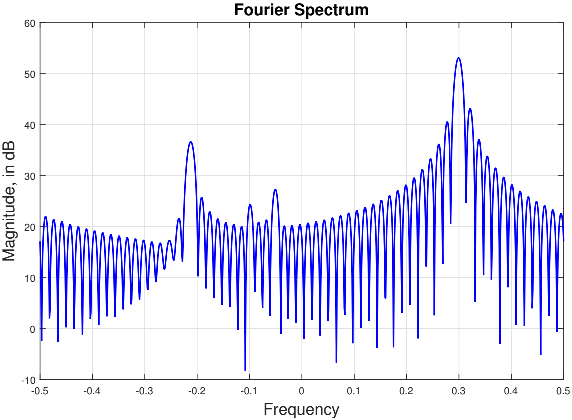
  where $\alpha_i \in \mathbb{C}$ is the amplitude-phase coefficient of the ith sinusoid; $f_i \in \left[-\frac{1}{2}, \frac{1}{2}\right)$ is the frequency of the ith sinusoid; w_t is noise; T is the observation time length

- **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"[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 $z_i = e^{j2\pi f_i}$. 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 $T_d \rightarrow \infty$,

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

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
- $\mathcal{R}(\mathbf{A\Phi A}^H) = \mathcal{R}(\mathbf{A})$
  - Proof: $\mathbf{A}^H$ has full row rank $\Longrightarrow$ $\mathrm{rank}(\mathbf{\Phi A}^H) = \mathrm{rank}(\mathbf{\Phi})$. Since $\mathbf{\Phi}$ is PD (and thus full rank), $\mathbf{\Phi A}^H$ has full row rank. Then use the property on the last page of Section 5.1
- $\mathrm{rank}(\mathbf{A\Phi A}^H) = \mathrm{rank}(\mathbf{A}) = k$, and $\mathbf{A\Phi A}^H$ has k nonzero eigenvalues

not true for any rank k matrix