Schur Complement

- the Schur complement: let

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

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

$$\mathbf{S}_D = \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{B}^T,$$

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

- We have

$$\mathbf{X} \succeq \mathbf{0} \text{ (resp. } \mathbf{X} \succ \mathbf{0}) \iff \mathbf{S}_D \succeq \mathbf{0} \text{ (resp. } \mathbf{S}_D \succ \mathbf{0})$$

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

$$1 - \mathbf{b}^T\mathbf{D}^{-1}\mathbf{b} \geq 0 \iff \mathbf{D} - \mathbf{b}\mathbf{b}^T \succeq \mathbf{0}$$

  (prove by yourself)

Schur Complement

- let a PD matrix

$$\mathbf{X} = \begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{b}^T & d \end{bmatrix},$$

  where $\mathbf{A} \in \mathbb{S}^{m-1}$, $\mathbf{b} \in \mathbb{R}^m$, $d \in \mathbb{R}$.

- if $\mathbf{b} \neq \mathbf{0}$, the first $m-1$ diagonal elements of $\mathbf{X}^{-1}$ are larger than or equal to the corresponding diagonal elements of $\mathbf{A}^{-1}$ and the mth diagonal element of $\mathbf{X}^{-1}$ is larger than d^{-1}.
- If $\mathbf{b} = \mathbf{0}$, the diagonal elements of $\mathbf{X}^{-1}$ are equal to the corresponding diagonal elements of $\mathbf{A}^{-1}$ and d^{-1}.
- proof (via $s_A = d - \mathbf{b}^T\mathbf{A}^{-1}\mathbf{b}$):

$$\mathbf{X}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{b}(d - \mathbf{b}^T\mathbf{A}^{-1}\mathbf{b})^{-1}\mathbf{b}^T\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{b}(d - \mathbf{b}^T\mathbf{A}^{-1}\mathbf{b})^{-1} \\ -(d - \mathbf{b}^T\mathbf{A}^{-1}\mathbf{b})^{-1}\mathbf{b}^T\mathbf{A}^{-1} & (d - \mathbf{b}^T\mathbf{A}^{-1}\mathbf{b})^{-1} \end{bmatrix}$$

- proof (via $\mathbf{S}_d = \mathbf{A} - d^{-1}\mathbf{b}\mathbf{b}^T$):

$$\mathbf{X}^{-1} = \begin{bmatrix} (\mathbf{A} - d^{-1}\mathbf{b}\mathbf{b}^T)^{-1} & -(\mathbf{A} - d^{-1}\mathbf{b}\mathbf{b}^T)^{-1}\mathbf{b}d^{-1} \\ -d^{-1}\mathbf{b}^T(\mathbf{A} - d^{-1}\mathbf{b}\mathbf{b}^T)^{-1} & d^{-1} + d^{-1}\mathbf{b}^T(\mathbf{A} - d^{-1}\mathbf{b}\mathbf{b}^T)^{-1}\mathbf{b}d^{-1} \end{bmatrix}$$

Application: Linear Transformations of Random Vectors

- scalar multiplication: covariance matrix of $a\mathbf{y}$ is $a^2\boldsymbol{\Sigma}$
- sum: given two uncorrelated random vectors $\mathbf{y}_1$, $\mathbf{y}_2$ with covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$, resp., covariance matrix of $\mathbf{y}_1 + \mathbf{y}_2$ is $\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2$
- affine transformation: define the m-dim. random vector $\mathbf{z} = \mathbf{A}\mathbf{y} + \mathbf{b}$ where $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$ are deterministic, covariance matrix of $\mathbf{z}$ is $\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T$
  - suppose $\mathbf{x}$ is a random (column) vector with non-singular covariance matrix $\boldsymbol{\Sigma}$ and mean $\mathbf{0}$. then the transformation

$$\mathbf{y} = \mathbf{W}\mathbf{x}$$

    with a whitening matrix $\mathbf{W}$ satisfying the condition $\mathbf{W}^T\mathbf{W} = \boldsymbol{\Sigma}^{-1}$ yields the whitened random vector $\mathbf{y}$ with unit diagonal covariance
  - infinitely many possible $\mathbf{W}$: $\mathbf{W} = \boldsymbol{\Sigma}^{-1/2}$ (Mahalanobis or ZCA whitening), $\mathbf{W} = \mathbf{L}^T$ with $\mathbf{L}$ the Cholesky factor of $\boldsymbol{\Sigma}^{-1}$ (Cholesky whitening), the eigensystem of $\boldsymbol{\Sigma}$ (PCA whitening), or CCA whitening considering two random vectors

Application: Factor Model

- suppose a n-dim. random vector $\mathbf{x}$ has covariance matrix

$$\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T + \sigma^2\mathbf{I}$$

- $\mathbf{x}$ can be interpreted as being generated by a model $\mathbf{x} = \boldsymbol{\mu} + \mathbf{A}\mathbf{y} + \mathbf{w}$, where
  - $\boldsymbol{\mu}$ is the mean of $\mathbf{x}$
  - $\mathbf{y}$ is a random variable with mean zero and covariance matrix $\mathbf{I}$
  - $\mathbf{w}$ is random error or noise, uncorrelated with $\mathbf{y}$, with $\mathrm{E}[\mathbf{w}] = \mathbf{0}$, $\mathrm{E}[\mathbf{w}\mathbf{w}^T] = \sigma^2\mathbf{I}$
- in statistics and machine learning, this is known as a factor model
  - components of $\mathbf{y}$ are common factors in $\mathbf{x}$
  - $\mathbf{x} - \boldsymbol{\mu}$ is a vector $\mathbf{A}\mathbf{y}$ in a subspace $\mathcal{R}(\mathbf{A})$ plus noise $\mathbf{w}$

Application: Graph Matrices

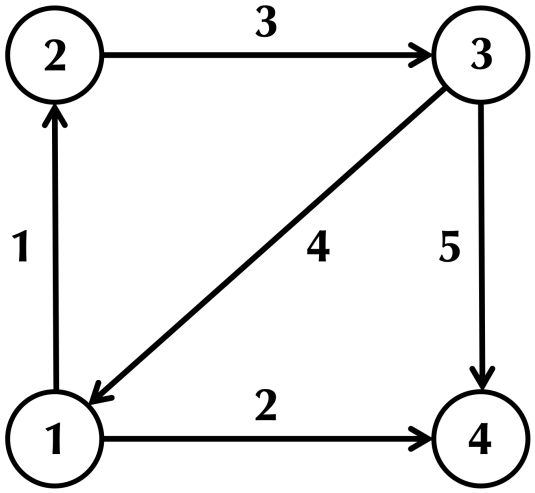
- a graph is made up of vertices (a.k.a. nodes or points) which are connected by edges (a.k.a. links or lines)
- a graph is completely determined by specifying either its adjacency structure or its incidence structure
  - In a graph, two vertices are adjacent if they are connected by an edge.
  - note: two edges are adjacent if they share a common vertex; the edge adjacency relations are not often used since they cannot uniquely determine a graph
  - In a graph, a vertex is incident to an edge (or an edge is incident to a vertex) if the vertex is one of the two vertices that the edge connects.
- these specifications provide far more efficient ways of representing a large or complicated graph than a pictorial representation
- as computers are more adept at manipulating numbers than at recognising pictures, it is standard to use matrix representations for the graph in computers
- the rank and eigenvalue (eigenvector) properties of graph matrices may dictate the structures of the graph

Application: Graph Matrices

- we consider a simple graph, i.e., a graph without loops (a.k.a. self-loops, self-edges, or buckles) and multiple edges (a.k.a. parallel edges or a multi-edge)
- for a directed graph (i.e., digraph) with m vertices and n edges, the adjacency matrix is a matrix $\mathbf{A} \in \mathbb{R}^{m \times m}$ defined by

$$a_{ij} = \begin{cases} 1 & \text{if vertex } i \text{ and vertex } j \text{ are adjacent with an edge from vertex } i \text{ to } j \\ 0 & \text{otherwise} \end{cases}$$

- encodes the relation of "vertex-vertex" pairs; square (typically sparse) with zeros on its diagonal; if graph is undirected (directed), it is symmetric (asymmetric)



$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Application: Graph Matrices

- for a directed graph, in-degree of vertex k is $d_k^{\text{in}} = \sum_{i=1}^{m} a_{ik}$ $(\mathbf{d}^{\text{in}} = \mathbf{A}^T\mathbf{1})$; out-degree of vertex k is $d_k^{\text{out}} = \sum_{i=1}^{m} a_{ki}$ $(\mathbf{d}^{\text{out}} = \mathbf{A}\mathbf{1})$

- degree of vertex k, i.e., number of edges incident with it, is $d_k = d_k^{\text{in}} + d_k^{\text{out}}$ $(\mathbf{d} = \mathbf{d}^{\text{in}} + \mathbf{d}^{\text{out}})$

- symmetrized adjacency matrix $\tilde{\mathbf{A}}$ (i.e., taking the graph as undirected) and (vertex-)degree matrix $\mathbf{D}$ contains infor. about the degree of each vertex; degree of vertex k is $d_k = \sum_{i=1}^{m} \tilde{a}_{ik} = \sum_{i=1}^{m} \tilde{a}_{ki}$, i.e., $\mathbf{d} = \tilde{\mathbf{A}}\mathbf{1} = \tilde{\mathbf{A}}^T\mathbf{1}$

$$\tilde{\mathbf{A}} = \mathbf{A} \vee \mathbf{A}^T = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \qquad \mathbf{D} = \text{Diag}(\mathbf{d}) = \begin{bmatrix} 3 & & & \\ & 2 & & \\ & & 3 & \\ & & & 2 \end{bmatrix}$$

Application: Graph Matrices

- we can associate a nonnegative weight w_k with edge k and $\mathbf{w} = [w_1, \ldots, w_n]^T$; the weighted adjaciency matrix and symmetrized weighted adjacency matrix are

$$\mathbf{A}_w = \begin{bmatrix} 0 & w_1 & 0 & w_2 \\ 0 & 0 & w_3 & 0 \\ w_4 & 0 & 0 & w_5 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad \tilde{\mathbf{A}}_w = \begin{bmatrix} 0 & w_1 & w_4 & w_2 \\ w_1 & 0 & w_3 & 0 \\ w_4 & w_3 & 0 & w_5 \\ w_2 & 0 & w_5 & 0 \end{bmatrix} \neq \mathbf{A}_w + \mathbf{A}_w^T$$

- the weighted in- and out-degree are $\mathbf{d}_w^{\text{in}} = \mathbf{A}_w^T \mathbf{1}$ and $\mathbf{d}_w^{\text{out}} = \mathbf{A}_w \mathbf{1}$; weighted degree $\mathbf{d}_w = \mathbf{d}_w^{\text{in}} + \mathbf{d}_w^{\text{out}} = \tilde{\mathbf{A}}_w \mathbf{1}$; weighted degree matrix $\mathbf{D}_w = \text{Diag}(\mathbf{d}_w)$

- the (symmetric) normalized adjacency $\bar{\mathbf{A}} = \mathbf{D}^{-1/2} \tilde{\mathbf{A}} \mathbf{D}^{-1/2}$ and the normalized weighted adjacency $\bar{\mathbf{A}}_w = \mathbf{D}_w^{-1/2} \tilde{\mathbf{A}}_w \mathbf{D}_w^{-1/2}$
- the row normalized (a.k.a. left normalized) and column normalized (a.k.a. right normalized) adjacency are $\bar{\mathbf{A}}_{\text{row}} = \mathbf{D}^{-1} \tilde{\mathbf{A}}$ and $\bar{\mathbf{A}}_{\text{col}} = \tilde{\mathbf{A}} \mathbf{D}^{-1}$, resp.; definitions for $\bar{\mathbf{A}}_{w,\text{row}}$ and $\bar{\mathbf{A}}_{w,\text{col}}$ follow
- note the scaling of w_k has no effect on $\bar{\mathbf{A}}_w$, $\bar{\mathbf{A}}_{w,\text{row}}$, and $\bar{\mathbf{A}}_{w,\text{col}}$

Application: Graph Matrices

- If $\mathbf{A}$ is the adjacency matrix of a directed or undirected graph, the (i, j)-th element in $\mathbf{A}^k$ means the number of (directed or undirected) walks of length k from vertex i to vertex j.

- not to be confused with the distance matrix (or vertex-distance matrix or the minimum path matrix) of a graph, which cantains the length of the shortest path, i.e., the minimum number of edges, between any two vertices i and j

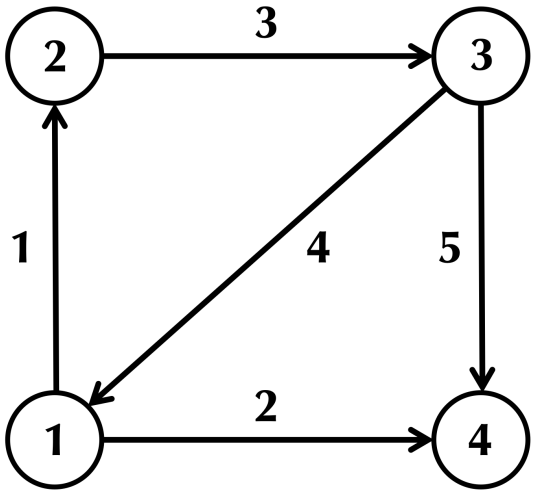
- The adjacency matrix can be used to determine whether a graph is connected.

Application: Graph Matrices

- for a directed graph with m vertices and n edges, the (oriented) incidence matrix is a matrix $\mathbf{B} \in \mathbb{R}^{m \times n}$ defined by

$$b_{ij} = \begin{cases} 1 & \text{if edge } j \text{ ends at vertex } i \\ -1 & \text{if edge } j \text{ starts at vertex } i \\ 0 & \text{otherwise} \end{cases}$$

- it indicates whether "vertex-edge" pairs are incident or not



$$\mathbf{B} = \begin{bmatrix} -1 & -1 & 0 & 1 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

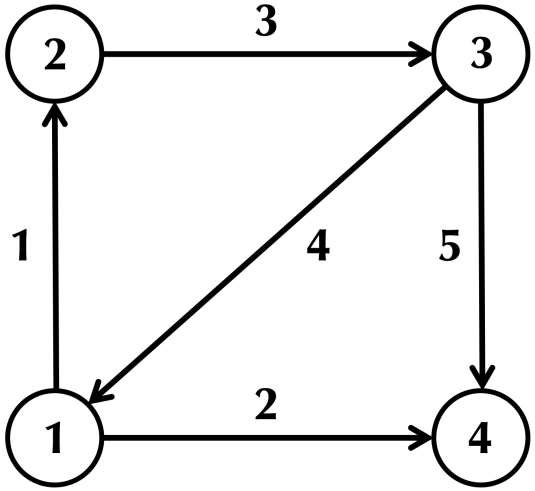
- the "edge-vertex" incidence matrix is $\mathbf{B}^T$
- we can also define the weighted incidence matrix $\mathbf{B}_w = \mathbf{B}\mathbf{D}_w^{1/2}$

Application: Graph Matrices

- matrix $\mathbf{L} = \mathbf{B}\mathbf{B}^T$ is called (symmetric signed) Laplacian matrix or graph Laplacian
- a symmetric $m \times m$ matrix with elements

$$\ell_{ij} = \begin{cases} \text{degree of vertex } i & \text{if } i = j \\ -1 & \text{if } i \neq j \text{ and vertices } i \text{ and } j \text{ are adjacent} \\ 0 & \text{otherwise} \end{cases}$$

- does not depend on the orientation of the edges; square and typically sparse



$$\mathbf{L} = \mathbf{B}\mathbf{B}^T = \begin{bmatrix} 3 & -1 & -1 & -1 \\ -1 & 2 & -1 & 0 \\ -1 & -1 & 3 & -1 \\ -1 & 0 & -1 & 2 \end{bmatrix}$$

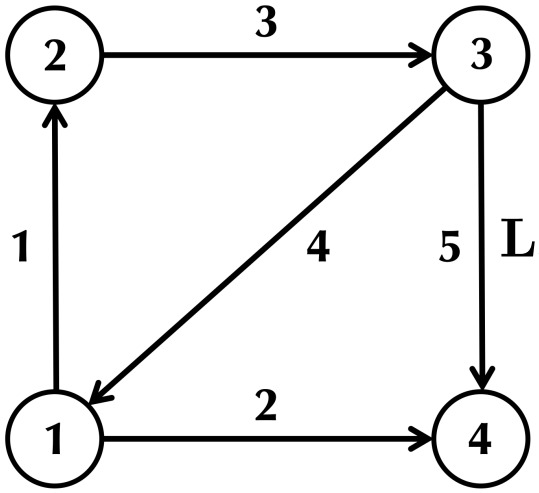
- $\mathbf{L}$ is PSD; $\mathrm{rank}(\mathbf{L}) = \mathrm{rank}(\mathbf{B})$; $\lambda_{\min}(\mathbf{L}) = 0$; $\mathbf{L} = \mathbf{D} - \tilde{\mathbf{A}}$ and $\lambda(\mathbf{L}) = 1 - \lambda(\tilde{\mathbf{A}})$
- an "augmented vertex-adjacency matrix"; edge-based Laplacian is not often used
- symm. normalized Laplacian $\bar{\mathbf{L}} = \mathbf{I} - \bar{\mathbf{A}} = \mathbf{D}^{-1/2}(\mathbf{D} - \tilde{\mathbf{A}})\mathbf{D}^{-1/2} = \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}$

Application: Graph Matrices

- the weighted graph Laplacian is the matrix $\mathbf{L}_w = \mathbf{B}_w\mathbf{B}_w^T = \mathbf{B}\mathbf{D}_w\mathbf{B}^T$

$$\ell_{w,ij} = \begin{cases} \sum_{k\in\mathcal{N}_i} w_k & \text{if } i = j \text{ (where } \mathcal{N}_i \text{ is the set of edges incident to vertex } i) \\ -w_k & \text{if } i \neq j \text{ and edge } k \text{ is between vertices } i \text{ and } j \\ 0 & \text{otherwise} \end{cases}$$

- still a PSD matrix; $\mathrm{rank}(\mathbf{L}_w) = \mathrm{rank}(\mathbf{B}_w)$; $\lambda_{\min}(\mathbf{L}_w) = 0$; $\mathbf{L}_w = \mathbf{D}_w - \tilde{\mathbf{A}}_w$



$$\mathbf{L}_w = \begin{bmatrix} w_1+w_2+w_4 & -w_1 & -w_4 & -w_2 \\ -w_1 & w_1+w_3 & -w_3 & 0 \\ -w_4 & -w_3 & w_3+w_4+w_5 & -w_5 \\ -w_2 & 0 & -w_5 & w_2+w_5 \end{bmatrix}$$

- it is the conductance matrix of a resistive circuit (w_k is conductance in branch k)
- the symmetric normalized weighted Laplacian $\bar{\mathbf{L}}_w = \mathbf{I} - \bar{\mathbf{A}}_w = \mathbf{D}_w^{-1/2}(\mathbf{D}_w - \tilde{\mathbf{A}}_w)\mathbf{D}_w^{-1/2} = \mathbf{D}_w^{-1/2}\mathbf{L}_w\mathbf{D}_w^{-1/2}$

Application: Graph Matrices

- Laplacian matrix is used to characterize the "smoothness" of a graph
- smooth: if two vertices are adjacient, values of the two vertices should be close
- the Laplacian quadratic form (a.k.a. Dirichlet energy)

$$\mathbf{x}^T\mathbf{L}_w\mathbf{x} = \|\mathbf{B}_w^T\mathbf{x}\|_2^2 = \left\| \begin{bmatrix} -\sqrt{w_1} & \sqrt{w_1} & 0 & 0 \\ -\sqrt{w_2} & 0 & 0 & \sqrt{w_2} \\ 0 & -\sqrt{w_3} & \sqrt{w_3} & 0 \\ \sqrt{w_4} & 0 & -\sqrt{w_4} & 0 \\ 0 & 0 & -\sqrt{w_5} & \sqrt{w_5} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \right\|_2^2 = \sum_{k\sim(i,j)} w_k(x_i - x_j)^2$$

- the normalized Laplacian quadratic form

$$\mathbf{x}^T\bar{\mathbf{L}}_w\mathbf{x} = \|\mathbf{B}_w^T\mathbf{D}_w^{-1/2}\mathbf{x}\|_2^2 = \sum_{k\sim(i,j)} w_k\left(\frac{x_i}{\sqrt{d_{w,i}}} - \frac{x_j}{\sqrt{d_{w,j}}}\right)^2$$

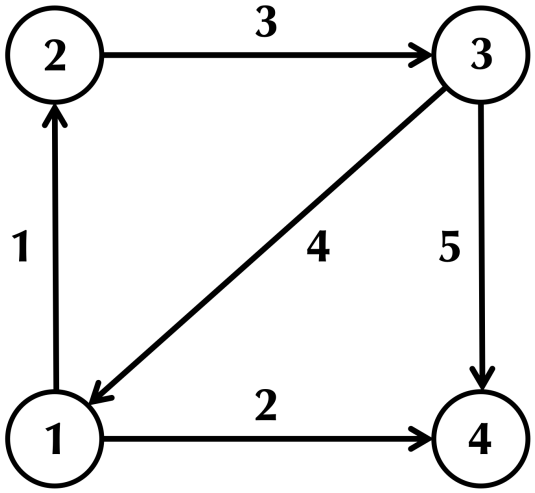
- if $\mathbf{x}$ is a signal on the graph, minimize this enengy is to minimize the variation of the graph signal

Application: Graph Matrices

- for a directed graph with m vertices and n edges, the (unoriented) incidence matrix is a matrix $\tilde{\mathbf{B}} \in \mathbb{R}^{m \times n}$ defined by

$$\tilde{b}_{ij} = \begin{cases} 1 & \text{if vertex } i \text{ is incident to edge } j \\ 0 & \text{otherwise} \end{cases}$$

- it indicates whether "vertex-edge" pairs are incident or not



$$\tilde{\mathbf{B}} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

- the signless Laplacian is $\tilde{\mathbf{L}} = \tilde{\mathbf{B}}\tilde{\mathbf{B}}^T$ which is PSD and $\tilde{\mathbf{L}} = \mathbf{D} + \tilde{\mathbf{A}}$
- we can also define the weighted incidence matrix $\tilde{\mathbf{B}}\mathbf{D}_w^{1/2}$ and the "edge-vertex" incidence matrix $\tilde{\mathbf{B}}^T$

Convexity of Log-Determinant Function

- The logarithmic determinant (log-determinant) function is a function from $\mathbb{S}^n$, with domain the set of PD matrices, and with values

$$f(\mathbf{X}) = \begin{cases} \log\det\mathbf{X} & \text{if } \mathbf{X} \succ \mathbf{0} \\ +\infty & \text{otherwise.} \end{cases}$$

  Through eigendecomposition, the function can be expressed in terms of the eigenvalues of $\mathbf{X}$. It provides a measure of the volume of an ellipsoid.
- Precisely, the volume of the ellipsoid $\mathcal{E}(\mathbf{P}) = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x}^T\mathbf{P}^{-1}\mathbf{x} \leq 1\}$ is given by $\mathrm{vol}(\mathcal{E}(\mathbf{P})) = C_n \prod_{i=1}^n \lambda_i^{1/2}(\mathbf{P})$, where C_n is a constant (given by the volume of the unit sphere in $\mathbb{R}^n$). Thus,

$$\log\mathrm{vol}(\mathcal{E}(\mathbf{P})) = \frac{1}{2}f(\mathbf{P}) + \text{constant.}$$

- This means that the volume of the ellipsoid is a function of the product of the eigenvalues of the matrix $\mathbf{P}$.

Convexity of Log-Determinant Function

- Log-determinant function $f(\mathbf{X}) = \log\det(\mathbf{X})$ is concave for PD $\mathbf{X}$.
- Fact: a function is convex if and only if its restriction to a line is.
- We can consider an arbitrary line, given by $\mathbf{X} = \mathbf{Z} + t\mathbf{V}$, where $\mathbf{Z}, \mathbf{V} \in \mathbb{S}^n$.
- We define $g(t) = f(\mathbf{Z} + t\mathbf{V})$, and restrict g to the interval of values of t for which $\mathbf{Z} + t\mathbf{V} \succ \mathbf{0}$. Without loss of generality, we can assume that $t = 0$ is inside this interval, i.e., $\mathbf{Z} \succ \mathbf{0}$. We have

$$\begin{aligned} g(t) &= \log\det(\mathbf{Z} + t\mathbf{V}) \\ &= \log\det\left(\mathbf{Z}^{1/2}\left(\mathbf{I} + t\mathbf{Z}^{-1/2}\mathbf{V}\mathbf{Z}^{-1/2}\right)\mathbf{Z}^{1/2}\right) \\ &= \sum_{i=1}^{n} \log\left(1 + t\lambda_i\right) + \log\det\mathbf{Z} \end{aligned}$$

  where λ_i is the eigenvalue of $\mathbf{Z}^{-1/2}\mathbf{V}\mathbf{Z}^{-1/2}$.
- It is easy to verify $g(t)$ is concave, and hence $f(\mathbf{X})$ is concave.

Application: Spectral Analysis

- consider the complex harmonic time-series

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

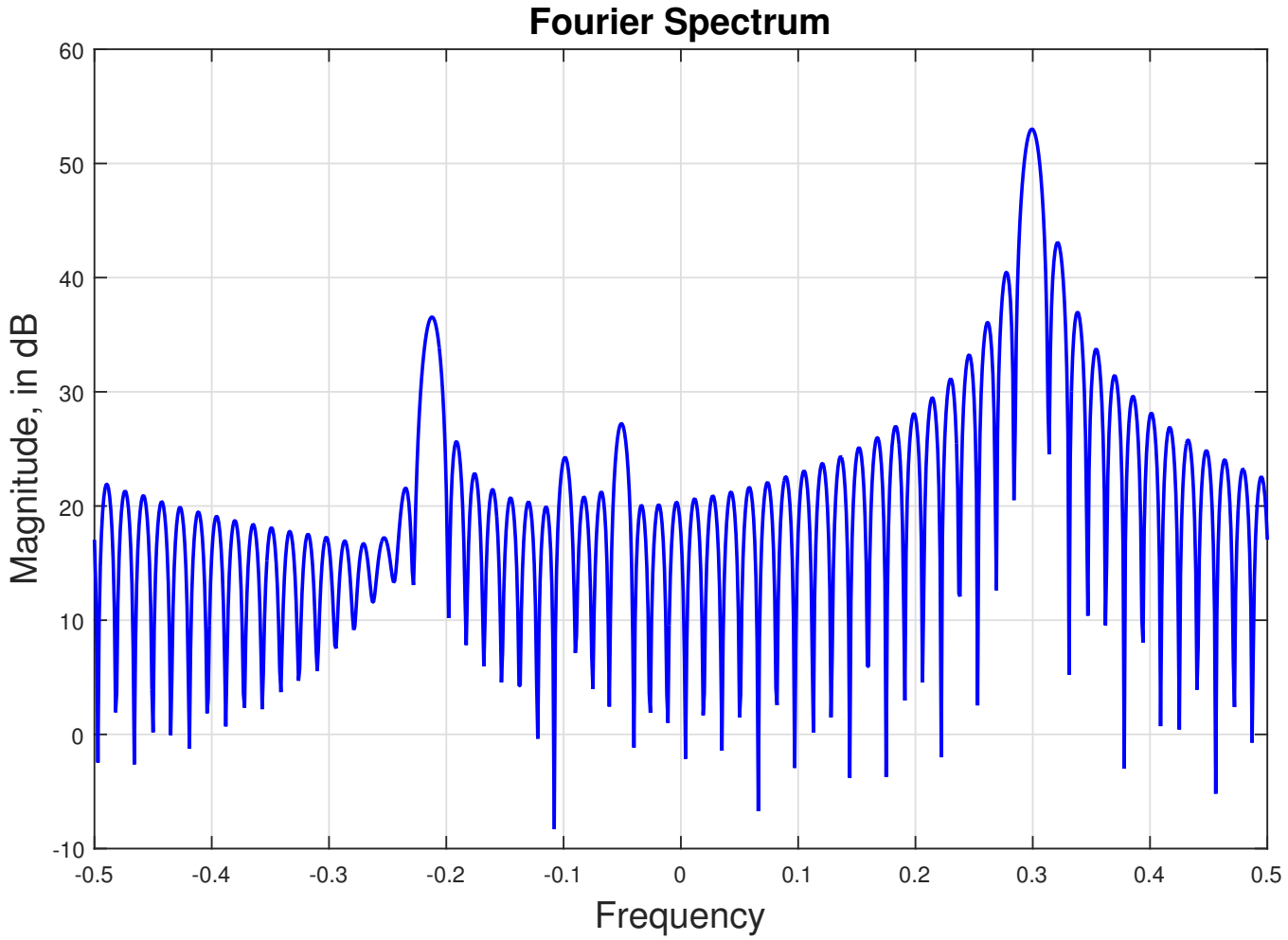
  where $\alpha_i \in \mathbb{C}$ is the amplitude-phase coefficient of the ith sinusoid; $f_i \in \left[-\frac{1}{2}, \frac{1}{2}\right)$ is the frequency of the ith sinusoid; w_t is noise; T is the observation time length

- 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 $z_i = e^{\jmath 2\pi f_i}$. 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

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

- (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 $T_d \to \infty$,

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

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; $\mathbf{\Phi} = \frac{1}{T_d}\mathbf{D}\mathbf{S}\mathbf{S}^H\mathbf{D}^H$; $\mathbf{D} = \mathrm{Diag}(\alpha_1, \ldots, \alpha_k)$;

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

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

- observation: $\mathbf{A}$ and $\mathbf{S}$ 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 2
  - $\mathrm{rank}(\mathbf{A}\mathbf{\Phi}\mathbf{A}^H) = \mathrm{rank}(\mathbf{A}) = k$, thus $\mathbf{A}\mathbf{\Phi}\mathbf{A}^H$ has k nonzero eigenvalues

Spectral Analysis via Subspace: Subspace Properties

- consider the eigendecomposition of $\mathbf{A}\mathbf{\Phi}\mathbf{A}^H$. Let $\mathbf{A}\mathbf{\Phi}\mathbf{A}^H = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$ 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 $i = k+1, \ldots, d$,

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

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

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 $f \in \{f_1, \dots f_k\}$ implies $\mathbf{a}(e^{\boldsymbol{\jmath} 2\pi f}) \in \mathcal{R}(\mathbf{A})$. But is it also true that $\mathbf{a}(e^{\boldsymbol{\jmath} 2\pi f}) \in \mathcal{R}(\mathbf{A})$ implies $f \in \{f_1, \dots f_k\}$?

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

  Theorem 3. 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 it holds that

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

Spectral Analysis via Subspace: Subspace Properties

- proof of Theorem 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

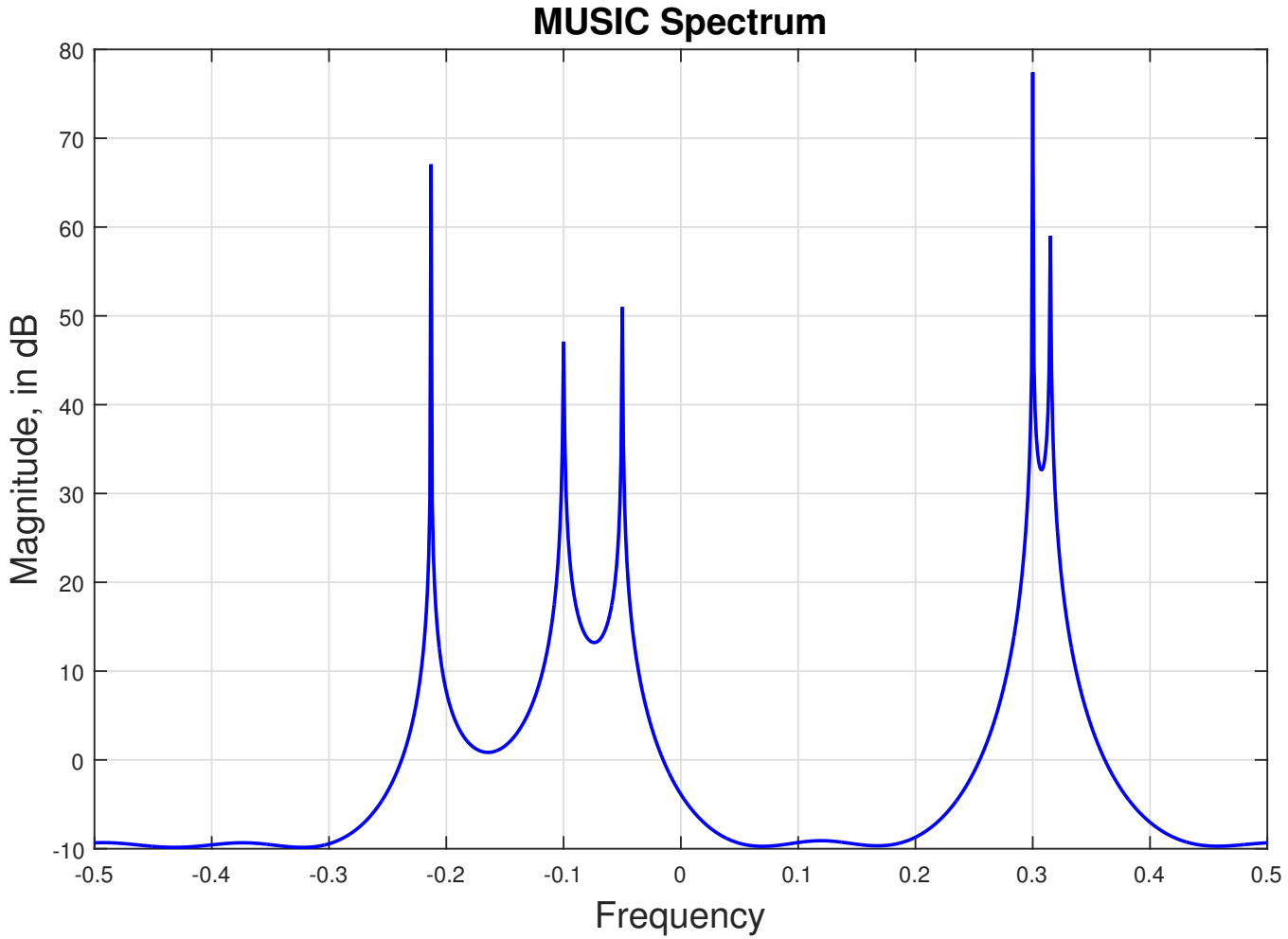
- MUSIC uses the fact that $\mathbf{a}(e^{\boldsymbol{\jmath} 2\pi f}) \in \mathcal{R}(\mathbf{A}) \Longleftrightarrow \mathbf{V}_2^H \mathbf{a}(e^{\boldsymbol{\jmath} 2\pi f}) = \mathbf{0}$

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

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

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

Spectral Analysis via Subspace: Algorithm

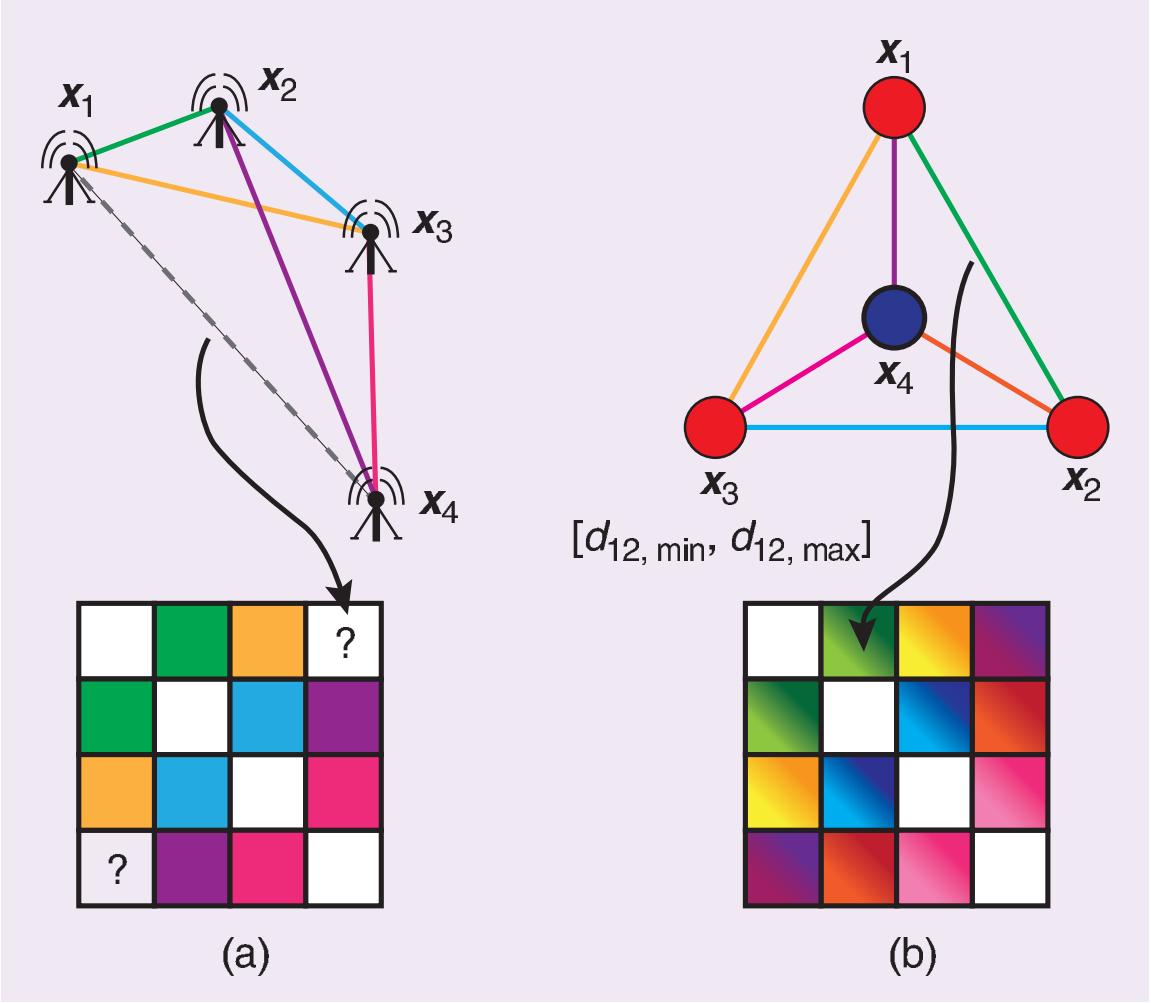


An illustration of the MUSIC spectrum. $T = 64$, $k = 5$, $\{f_1, \ldots, f_k\} = \{-0.213, -0.1, -0.05, 0.3, 0.315\}$.

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
  - it is related to multidimensional scaling (MDS) in machine learning for dimensionality reduction
- 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
$$\mathbf{R} = \mathbf{1}(\mathrm{diag}(\mathbf{X}^T\mathbf{X}))^T - 2\mathbf{X}^T\mathbf{X} + (\mathrm{diag}(\mathbf{X}^T\mathbf{X}))\mathbf{1}^T \qquad (*)$$
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,

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

- 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 $\mathbf{G} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$. Assuming $\lambda_1 \geq \ldots \geq \lambda_n$, it takes the form

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

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

- EDM solution: take $\hat{\mathbf{X}} = \mathbf{\Lambda}^{1/2}\mathbf{V}_1^T$ as an estimate of $\mathbf{X}$
- recovery guarantee: by Property 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 M. Vetterli, "Euclidean distance matrices," *IEEE Signal Processing Magazine*, vol. 32, no. 6, pp. 12–30, Nov. 2015.