Spectral Analysis: Subspace Properties (cont'd)

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

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

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

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

$\mathcal{R}(A) =$

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

$\supseteq$ $\qquad$ $\|$

$\mathcal{R}(V_1 \Lambda V_1^H)$

Spectral Analysis: Subspace Properties (cont'd)

How to find V_1?

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

$$\underbrace{\mathbf{R}_y}_{\text{known from } Y} = \begin{bmatrix} \mathbf{V}_1 & \mathbf{V}_2 \end{bmatrix} \begin{bmatrix} \Lambda_1 + \sigma^2 \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \sigma^2 \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^H \\ \mathbf{V}_2^H \end{bmatrix}$$

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 $\mathbf{a}(e^{j2\pi f}) \in \mathcal{R}(\mathbf{A})$

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,

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

Proof: $(\Rightarrow)$ $z \in \{z_1, \cdots, z_k\} \Rightarrow a(z)$ is a column of A

$\Rightarrow a(z) \in \mathcal{R}(A)$.

$(\Leftarrow)$ Assume to the contrary that $\exists\, z' \notin \{z_1, \cdots, z_k\}$

s.t. $a(z') \in \mathcal{R}(A)$.

Consider $A' = [\, a(z') \quad A \,] \in \mathbb{C}^{d \times (k+1)}$

$a(z') \in \mathcal{R}(A) \Rightarrow A'$ has linearly dependent columns $\Rightarrow$ A' doesn't have full column rank

However, A' is Vandemonde with distinct roots $z', z_1, \cdots, z_k$

Since $d \geq k+1$. A' has full column rank. Contradiction!

Spectral Analysis: Algorithm

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

$= \mathcal{R}(V_2)^\perp$

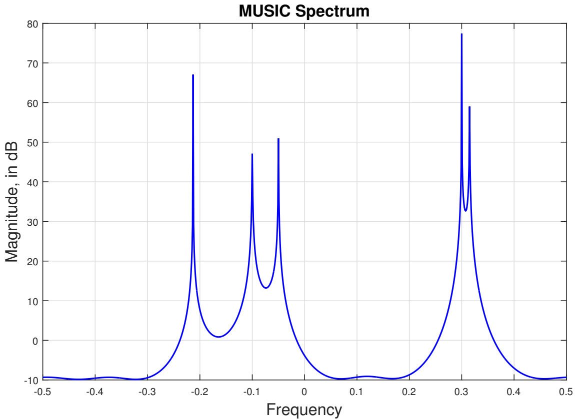
MUSIC uses the fact that $\mathbf{a}(e^{\boldsymbol{j}2\pi f}) \in \mathcal{R}(\mathbf{A}) \Longleftrightarrow \mathbf{V}_2^H \mathbf{a}(e^{\boldsymbol{j}2\pi f}) = \mathbf{0}$

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

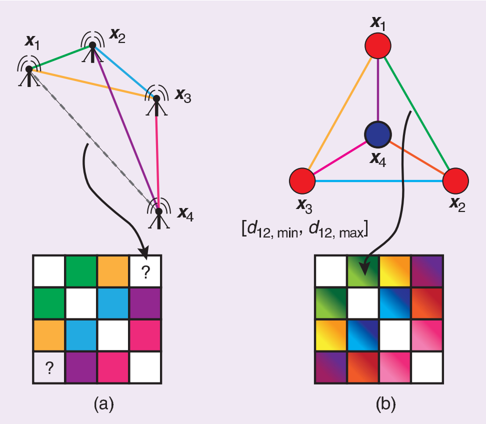


An illustration of the MUSIC spectrum. $T = 64$, $k = 5$,
$\{f_1, \ldots, f_k\} = \{-0.213, -0.1, -0.05, 0.3, 0.315\}$

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 known $\mathbf{R} \in \mathbb{R}^{n \times n}$ be s.t. $r_{ij} = d_{ij}^2$ $= \|x_i - x_j\|_2^2$ for all $i, j = 1, \ldots, n$
- Note from

$= (x_i - x_j)^T (x_i - x_j)$

$$r_{ij} = d_{ij}^2 = \mathbf{x}_i^T \mathbf{x}_i - 2\mathbf{x}_i^T \mathbf{x}_j + \mathbf{x}_j^T \mathbf{x}_j$$

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 $\mathrm{diag}(\mathbf{Y}) := [y_{11}, \ldots, y_{nn}]^T$ for any square matrix $\mathbf{Y}$

- **Observation:** $(*)$ Also holds if we replace $\mathbf{X}$ by — ambiguity in X
  - $\tilde{\mathbf{X}} = [\mathbf{x}_1 + \mathbf{b}, \ldots, \mathbf{x}_n + \mathbf{b}]$ for any $\mathbf{b} \in \mathbb{R}^d$ $(d_{ij} = \|\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j\|_2)$
  - $\tilde{\mathbf{X}} = \mathbf{Q}\mathbf{X}$ 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

$R = \mathbb{1}(\mathrm{diag}(X^TX))^T - 2X^TX + \mathrm{diag}(X^TX)\,\mathbb{1}^T$

EDM: Formulation (cont'd)

Verification of $(*)$

$$X = \begin{bmatrix} | & & | \\ x_1 & \cdots, & x_n \\ | & & | \end{bmatrix}$$

$$r_{ij} = d_{ij}^2 = x_i^T x_i - 2x_i^T x_j + x_j^T x_j$$

$$X^TX = \begin{bmatrix} — & x_1^T & — \\ & \vdots & \\ — & x_n^T & — \end{bmatrix} \begin{bmatrix} | & & | \\ x_1 & \cdots & x_n \\ | & & | \end{bmatrix} = \begin{bmatrix} x_1^Tx_1 & x_1^Tx_2 & \cdots & x_1^Tx_n \\ x_2^Tx_1 & x_2^Tx_2 & \cdots & x_2^Tx_n \\ \vdots & \vdots & & \vdots \\ x_n^Tx_1 & x_n^Tx_2 & \cdots & x_n^Tx_n \end{bmatrix}$$

$$\mathrm{diag}(X^TX) = [x_1^Tx_1 \ \cdots \ x_n^Tx_n]^T$$

$$\mathbb{1}\,\mathrm{diag}(X^TX)^T = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} [x_1^Tx_1 \ \cdots \ x_n^Tx_n] = \begin{bmatrix} x_1^Tx_1 & \cdots & x_n^Tx_n \\ \vdots & \vdots & \vdots \\ x_1^Tx_1 & \cdots & x_n^Tx_n \end{bmatrix}$$

$$\mathrm{diag}(X^TX)\,\mathbb{1}^T = \begin{bmatrix} x_1^Tx_1 & \cdots & x_1^Tx_1 \\ \vdots & \ddots & \vdots \\ x_n^Tx_n & \cdots & x_n^Tx_n \end{bmatrix}$$

$$\text{R.H.S. of } (*) = \begin{bmatrix} \underbrace{x_1^Tx_1 - 2x_1^Tx_1 + x_1^Tx_1}_{0} & \cdots & \underbrace{x_1^Tx_1 - 2x_1^Tx_n + x_n^Tx_n}_{\|x_1 - x_n\|_2^2 = d_{1n}^2} \\ \vdots & \vdots & \vdots \end{bmatrix}$$

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

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

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 $\lambda_1 \geq \ldots \geq \lambda_n$

$= V_1 \Lambda_1 V_1^T = V_1 \Lambda^{\frac{1}{2}} \Lambda^{\frac{1}{2}} V_1^T$

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

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.

Matrix Computations
Chapter 5: Positive Semidefinite Matrices
Section 5.3 Matrix Inequalities and Schur Complement

Jie Lu
ShanghaiTech University

PSD Matrix Inequalities

- Inequalities for matrices are defined based on the notion of PSD matrices
- PSD matrix inequalities are frequently used in topics like semidefinite programming
- Definitions:
  - $\mathbf{A} \succeq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is PSD $A - B \succeq 0$
  - $\mathbf{A} \succ \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is PD
  - $\mathbf{A} \not\succeq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is indefinite

$A \preceq B$

$\prec$

$\not\preceq$

$x^TAx \geq x^TCx$

$\Updownarrow$

$x^T(A-C)x \geq 0$

- Consequences immediately from the definitions: For any $\mathbf{A}, \mathbf{B}, \mathbf{C} \in \mathbb{S}^n$,   scalar
  - $\mathbf{A} \succeq \mathbf{0}, \alpha \geq 0$ (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. $\mathbf{A} \succ \mathbf{0}, \mathbf{B} \succ \mathbf{0}$) $\Longrightarrow \mathbf{A} + \mathbf{B} \succeq \mathbf{0}$ (resp. $\mathbf{A} + \mathbf{B} \succ \mathbf{0}$)
  - $\mathbf{A} \succeq \mathbf{B}, \mathbf{B} \succeq \mathbf{C}$ (resp. $\mathbf{A} \succ \mathbf{B}, \mathbf{B} \succ \mathbf{C}$) $\Longrightarrow \mathbf{A} \succeq \mathbf{C}$ (resp. $\mathbf{A} \succ \mathbf{C}$)
  - $\mathbf{A} \not\succeq \mathbf{B}$ does not imply $\mathbf{B} \succeq \mathbf{A}$

$\forall x \in \mathbb{R}^n$, $A \succeq B \Rightarrow 0 \leq x^T(A-B)x \Leftrightarrow x^TAx \geq x^TBx$. $B \succeq C \Rightarrow x^TBx \geq x^TCx$

Properties of PSD Matrix Inequalities

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

- $\mathbf{A} \succeq \mathbf{B} \implies \lambda_k(\mathbf{A}) \geq \lambda_k(\mathbf{B})$ for all k; the converse is not always true

$A \succeq B \iff A-B \succeq 0 \iff \lambda_k(A-B) \geq 0 \quad \forall k$

$\lambda_k(A) = \lambda_k(\underline{B} + \underline{A-B}) \geq \lambda_k(B) + \underbrace{\lambda_n(A-B)}_{\geq 0} \geq \lambda_k(B)$

Weyl's Inequality (Sec 4.6)

- $\mathbf{A} \succeq \mathbf{I}$ (resp. $\mathbf{A} \succ \mathbf{I}$) $\iff \lambda_k(\mathbf{A}) \geq 1$ for all k (resp. $\lambda_k(\mathbf{A}) > 1$ for all k)
- $\mathbf{I} \succeq \mathbf{A}$ (resp. $\mathbf{I} \succ \mathbf{A}$) $\iff \lambda_k(\mathbf{A}) \leq 1$ for all k (resp. $\lambda_k(\mathbf{A}) < 1$ for all k)

Another way of proof: Let $A = V \Lambda V^T$ be an decomposition of A.
Then,
$A - I = V(\Lambda - I)V^T$ is an eigendecomposition of $A - I$.

$\lambda_k(A-I) = \lambda_k(A) - 1 \quad \forall k$

$A \succeq I \iff \lambda_k(A-I) \geq 0 \iff \lambda_k(A) \geq 1.$

- If $\mathbf{A}, \mathbf{B} \succ \mathbf{0}$ then $\mathbf{A} \succeq \mathbf{B} \iff \mathbf{B}^{-1} \succeq \mathbf{A}^{-1}$

Properties of PSD Matrix Inequalities (cont'd)

- For $\mathbf{A} \succeq \mathbf{B} \succeq \mathbf{0}$, $\det(\mathbf{A}) \geq \det(\mathbf{B})$

determinant = Product of all eigenvalues

- For $\mathbf{A} \succeq \mathbf{B}$, $\mathrm{tr}(\mathbf{A}) \geq \mathrm{tr}(\mathbf{B})$

trace = sum of all eigenvalues

- For $\mathbf{A} \succeq \mathbf{B} \succ \mathbf{0}$, $\mathrm{tr}(\mathbf{A}^{-1}) \leq \mathrm{tr}(\mathbf{B}^{-1})$

$A^{-1} \preceq B^{-1}$

Schur Complement

Let $\mathbf{X} \in \mathbb{R}^{n \times n}$ be partitioned as

$$\mathbf{X} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}$$

where $\mathbf{A} \in \mathbb{R}^{m \times m}$ and $\mathbf{D} \in \mathbb{R}^{(n-m)\times(n-m)}$

Assume $\mathbf{D}$ is nonsingular and solve the linear system

$$\begin{cases} Ax + By = c & ① \\ Cx + Dy = d & ② \end{cases} \iff \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{c} \\ \mathbf{d} \end{bmatrix} \qquad \begin{matrix} x \in \mathbb{R}^m & c \in \mathbb{R}^m \\ y \in \mathbb{R}^{n-m} & d \in \mathbb{R}^{n-m} \end{matrix}$$

If $\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}$ is nonsingular, we obtain

$$\mathbf{x} = (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}(\mathbf{c} - \mathbf{B}\mathbf{D}^{-1}\mathbf{d})$$

$$\mathbf{y} = \mathbf{D}^{-1}(\mathbf{d} - \mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}(\mathbf{c} - \mathbf{B}\mathbf{D}^{-1}\mathbf{d}))$$

The matrix $\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}$ is called the Schur Complement of $\mathbf{D}$ in $\mathbf{X}$

Similarly, when $\mathbf{A}$ is nonsingular, the matrix $\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}$ is the Schur Complement of $\mathbf{A}$ in $\mathbf{X}$

$$② \Rightarrow y = D^{-1}(d - Cx) \overset{①}{\Longrightarrow} Ax + BD^{-1}(d - Cx) = c$$

$$(A - BD^{-1}C)x = c - BD^{-1}d$$

Schur Complement (cont'd)

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} \begin{bmatrix} c \\ d \end{bmatrix}$$

Suppose $\mathbf{D}$ and the Schur complement $\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}$ are nonsingular
Rewrite the solution of the linear system as

$$\begin{aligned}
\mathbf{x} &= (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{c} - (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1}\mathbf{d} \\
\mathbf{y} &= -\mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{c} + (\mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1})\mathbf{d}
\end{aligned}$$

Then, we derive the inverse of $\mathbf{X}$ as

$$\begin{aligned}
\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} &= \begin{bmatrix} (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & -(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \end{bmatrix} \\
&= \begin{bmatrix} (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & \mathbf{0} \\ -\mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & \mathbf{D}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\mathbf{B}\mathbf{D}^{-1} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{D}^{-1}\mathbf{C} & \mathbf{I} \end{bmatrix} \begin{bmatrix} (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\mathbf{B}\mathbf{D}^{-1} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}
\end{aligned}$$

It follows that

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{B}\mathbf{D}^{-1} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{D}^{-1}\mathbf{C} & \mathbf{I} \end{bmatrix}$$