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

- 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), and then any principal minor of $\mathbf{A}$ is nonnegative (resp. positive)

Properties of PSD Matrix

- (Sylvester's criterion). Let $\mathbf{A} \in \mathbb{S}^n$.
  - $\mathbf{A}$ is PD $\Longleftrightarrow$ all its leading (and trailing) principal minors are positive (for $\mathbf{A} \in \mathbb{S}^n$, the positivity of the leading principal minors implies the positivity of all its principal minors)
  - $\mathbf{A}$ is PSD $\Longleftrightarrow$ all its principal minors are nonnegative
  - If the first $n-1$ leading principal minors (and the last $n-1$ trailing principal minors) of $\mathbf{A}$ are positive and det $\det(\mathbf{A}) \geq 0$, then $\mathbf{A}$ is PSD.
- $\mathbf{A}$ is ND $\Longleftrightarrow$ its odd leading principal minors are negative and even are positive
- $\mathbf{A}$ is NSD $\Longleftrightarrow$ its odd principal minors are nonpositive and even are nonnegative
- $\mathbf{A}$ is indefinite $\Longleftrightarrow$ there are two of its odd leading principal minors that have different signs or there is one of its even leading principal minors that is negative

Properties of PSD Matrix

- To obtain conditions for a matrix to be PD or ND, we need to examine the leading principal minors.
- To obtain conditions for a matrix to be PSD or NSD, we need to examine all the principal minors.

- Procedures for checking the definiteness of a matrix
  - find the leading principal minors and check if the conditions for positive or negative definiteness are satisfied; if they are, the the matrix is PD or ND
  - if the conditions are not satisfied, check if they are strictly violated; if they are, then the matrix is indefinite
  - if the conditions are not strictly violated, find all its principal minors and check if the conditions for positive or negative semidefiniteness are satisfied

Application: Classification of Stationary Points

Stationary point (critical point): For a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, a stationary point is a point on the surface of the graph where $\nabla f(\mathbf{x}) = \mathbf{0}$.

- local extrema: given a stationary point $\mathbf{x}$, $\mathbf{x}$ is a local maximum (resp. a local minimum) if there exists a neighborhood $\mathcal{R}$ of $\mathbf{x}$ such that for all $\mathbf{x}' \in \mathcal{R}$, $f(\mathbf{x}') \leq f(\mathbf{x})$ (resp. $\geq f(\mathbf{x})$)
- saddle points: given a stationary point $\mathbf{x}$, $\mathbf{x}$ is a saddle point if for any neighborhood $\mathcal{R}$ of $\mathbf{x}$ there exists a $\mathbf{x}' \in \mathcal{R}$ such that $f(\mathbf{x}') > f(\mathbf{x})$ and a $\mathbf{x}'' \in \mathcal{R}$ such that $f(\mathbf{x}'') < f(\mathbf{x})$
- Suppose f is twice-differentiable, at a stationary point $\mathbf{x}$ we have

$$f(\mathbf{x} + \mathbf{r}) - f(\mathbf{x}) = \frac{1}{2}\mathbf{r}^T \nabla^2 f(\mathbf{x})\mathbf{r} + \mathcal{O}(|\mathbf{r}|^3)$$

  - A twice-differentiable function f has a local minimum (resp. a local maximum, a saddle point) $\mathbf{x}$ if $\nabla f(\mathbf{x}) = \mathbf{0}$ and $\nabla^2 f(\mathbf{x}) \succ \mathbf{0}$ (resp. $-\nabla^2 f(\mathbf{x}) \succ \mathbf{0}$, $\nabla^2 f(\mathbf{x}) \not\succeq \mathbf{0}$) at that point.
  - Unfortunately, it is inconclusive if $\nabla f(\mathbf{x}) = \mathbf{0}$ and $\nabla^2 f(\mathbf{x})$ is only semidefinite and not definite. (consider the example: $f(x_1, x_2) = x_1^3 + x_2^2$)

Properties of PSD Matrix

- $\mathbf{A}$ is PSD, $\mathbf{x}^T\mathbf{A}\mathbf{x} = 0 \Longleftrightarrow \mathbf{A}\mathbf{x} = \mathbf{0}$ for an $\mathbf{x}$. (how to prove it?)
  - proved by eigenvalue properties of PSD matrices
  - alternative proof: the "if" part is easy; the "only if" part: constructing

$$p(\lambda) = (\mathbf{x} + \lambda\mathbf{y})^T\mathbf{A}(\mathbf{x} + \lambda\mathbf{y}) = \lambda^2\mathbf{y}^T\mathbf{A}\mathbf{y} + 2\lambda\mathbf{y}^T\mathbf{A}\mathbf{x} + \mathbf{x}^T\mathbf{A}\mathbf{x}.$$

Since for all $\mathbf{x}$, λ, and $\mathbf{y}$, $p(\lambda) \geq 0$, we have the discriminant for $p(\lambda)$ should be nonpositive, i.e.,

$$4(\mathbf{y}^T\mathbf{A}\mathbf{x})^2 - 4(\mathbf{y}^T\mathbf{A}\mathbf{y})(\mathbf{x}^T\mathbf{A}\mathbf{x}) \leq 0$$

If $\mathbf{x}^T\mathbf{A}\mathbf{x} = 0$, the discriminant is nonpositive only if $\mathbf{y}^T\mathbf{A}\mathbf{x} = 0$ for all $\mathbf{y}$ or, equivalently, if $\mathbf{A}\mathbf{x} = \mathbf{0}$.

- $\mathbf{A}$ is PSD and nonsingular $\Longleftrightarrow$ $\mathbf{A}$ is PD
- for a PSD $\mathbf{A}$, it is PD $\Longleftrightarrow$ $\mathbf{A}$ is nonsingular

Properties of PSD Matrices

Property 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

$$\mathbf{C} \succ \mathbf{0} \Longleftrightarrow \mathbf{x}^T \mathbf{C} \mathbf{x} > 0, \ \forall \ \mathbf{x} \neq \mathbf{0} \Longleftrightarrow \mathbf{z}^T \mathbf{A} \mathbf{z} > 0, \ \forall \ \mathbf{z} = \mathbf{B}\mathbf{x}, \ \mathbf{x} \neq \mathbf{0}. \qquad (*)$$

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.

Example: Correlation Matrix

Given the random variable $\mathbf{y} = [y_1, \ldots, y_n]^T \in \mathbb{R}^n$

- μ_i is the mean or expected value of y_i
- σ_i is the standard deviation and σ_i^2 is the variance of y_i
- σ_{ij}, for i, j, is the covariance of y_i and y_j
- $\rho_{ij} = \sigma_{ij}/(\sigma_i \sigma_j)$, for $i \neq j$, is the correlation between y_i and y_j (variables y_i and y_j are uncorrelated if $\rho_{ij} = 0$, or equivalently, $\sigma_{ij} = 0$)
- the correlation matrix has i, j element $\rho_{ij} = \sigma_{ij}/(\sigma_i \sigma_j)$ for $i \neq j$ and 1 for $i = j$:

$$\mathbf{R} = \begin{bmatrix} 1 & \rho_{12} & \ldots & \rho_{1n} \\ \rho_{21} & 1 & \ldots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \ldots & 1 \end{bmatrix}$$

  which is symmetric and is the covariance matrix of the standardized variables $\tilde{y}_i = (y_i - \mu_i)/\sigma_i$

Example: Correlation Matrix

- the correlation matrix can also be defined as

$$\mathbf{R} = \mathbf{D}\boldsymbol{\Sigma}\mathbf{D}$$

where $\boldsymbol{\Sigma}$ is the covariance matrix and $\mathbf{D}$ is the diagonal matrix

$$\mathbf{D} = \begin{bmatrix} \sigma_1^{-1} & 0 & \dots & 0 \\ 0 & \sigma_2^{-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n^{-1} \end{bmatrix}$$

- since covariance is PSD, the correlation matrix is also PSD

- both covariance and correlation matrices are called second moment matrices

Properties of PSD Matrices

Theorem 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 $\mathbf{B} \in \mathbb{R}^{m\times n}$. (we can also write it as $\mathbf{A} = \mathbf{B}\mathbf{B}^T$ for some $\mathbf{B} \in \mathbb{R}^{n\times m}$)

- 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 $\boldsymbol{\Lambda}^{1/2} = \mathrm{Diag}(\lambda_1^{1/2}, \ldots, \lambda_n^{1/2})$ with $\lambda_i \geq 0$.

    $\mathbf{A} \succeq \mathbf{0} \Longrightarrow \mathbf{A} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T = (\mathbf{V}\boldsymbol{\Lambda}^{1/2})(\boldsymbol{\Lambda}^{1/2}\mathbf{V}^T)$, with $\boldsymbol{\Lambda}^{1/2}\mathbf{V}^T$ being real
- corollary: Given $\mathbf{A} = \mathbf{B}^T\mathbf{B}$, $\mathbf{A}\mathbf{x} = \mathbf{0} \Longleftrightarrow \mathbf{B}\mathbf{x} = \mathbf{0}$ for an $\mathbf{x}$
  - $\mathcal{N}(\mathbf{A}) = \mathcal{N}(\mathbf{B})$, $\mathrm{nullity}(\mathbf{A}) = \mathrm{nullity}(\mathbf{B})$, and $\mathrm{rank}(\mathbf{A}) = \mathrm{rank}(\mathbf{B})$
- corollary: $\mathbf{A} \in \mathbb{S}^n$ 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: $\mathbf{A} \in \mathbb{S}^n$ is PD if and only if there exists a full column-rank $\mathbf{B} \in \mathbb{R}^{m\times n}$ or 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 a unique upper-triangular matrix $\mathbf{B} \in \mathbb{R}^{n\times n}$ with $b_{ii} > 0$ s.t. $\mathbf{A} = \mathbf{B}^T\mathbf{B}$, which is the Cholesky factorization of $\mathbf{A}$.
  - $\mathbf{A}$ is PD iff has an LU (or LDL) factorization with all pivots being positive.

Properties of PSD Matrices

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

- denote

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

  - $\mathbf{B} = \mathbf{A}^{1/2}$ 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 symmetric 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 $\mathbf{A} = \mathbf{B}^2$

Example: PD Covariance Matrices

- the sample covariance matrix based on $\mathbf{Y}_c \in \mathbb{R}^{n \times T}$ is

$$\hat{\mathbf{\Sigma}} = \frac{1}{T}\mathbf{Y}_c\mathbf{Y}_c^T = \frac{1}{T}\mathbf{Y}\mathbf{C}\mathbf{C}^T\mathbf{Y}^T$$

- based on Theorem 2, conditions for $\hat{\mathbf{\Sigma}}$ to be PD is $\mathrm{rank}(\mathbf{Y}_c) = \mathrm{rank}(\mathbf{Y}_c^T) = n$, i.e., $\mathbf{Y}_c$ is full row-rank

- we further have $\mathrm{rank}(\mathbf{Y}_c) = \mathrm{rank}(\mathbf{Y}\mathbf{C}) \leq \min\{\mathrm{rank}(\mathbf{Y}), \mathrm{rank}(\mathbf{C})\} = \min\{\mathrm{rank}(\mathbf{Y}), T-1\} \leq \min\{n, T, T-1\} = \min\{n, T-1\}$
  - conditions for $\hat{\mathbf{\Sigma}}$ to be PD is $\mathbf{Y}$ has full rank and $T \geq n+1$

- if the samples $\mathbf{y}_i$'s are random, we at least need $T = n+1$ samples

Properties of PSD Matrices

Property 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 $\mathbf{A}$ is a PSD matrix with factorization $\mathbf{A} = \mathbf{BB}^T$ for some full-column rank $\mathbf{B}$, then $\mathcal{R}(\mathbf{A}) = \mathcal{R}(\mathbf{B})$.

Properties of PSD Matrices

Property 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,

    $$\mathbf{B}\mathbf{q}_i = \Pi_{\mathcal{R}(\mathbf{B})}(\mathbf{c}_i), \quad i = 1, \dots, k.$$

  - from Property 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.

Properties of PSD Matrix

- For PSD matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ of equal size,

$$\det(\mathbf{A}+\mathbf{B}+\mathbf{C})+\det(\mathbf{C}) \geq \det(\mathbf{A}+\mathbf{C})+\det(\mathbf{B}+\mathbf{C})$$

with the corollary $\det(\mathbf{A}+\mathbf{B}) \geq \det(\mathbf{A})+\det(\mathbf{B})$.

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
- 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
- This defines a partial ordering and a strict partial ordering on the set of all square matrices, which is called the Loewner order.
- results that immediately follow from the definition: let $\mathbf{A}, \mathbf{B}, \mathbf{C} \in \mathbb{S}^n$.
  - $\mathbf{A} \succeq \mathbf{B}, \mathbf{B} \succeq \mathbf{C}$ (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 $\mathbf{A} \succeq \mathbf{B}$, $\mathrm{tr}(\mathbf{A}) \geq \mathrm{tr}(\mathbf{B})$
  - for $\mathbf{A} \succeq \mathbf{B} \succ \mathbf{0}$, $\det(\mathbf{A}^{-1}) \leq \det(\mathbf{B}^{-1})$ and $\mathrm{tr}(\mathbf{A}^{-1}) \leq \mathrm{tr}(\mathbf{B}^{-1})$

- Example: $\mathcal{E}(\mathbf{Q}) = \{ \mathbf{x} \in \mathbb{R}^n \mid \mathbf{x}^T\mathbf{Q}^{-1}\mathbf{x} \leq 1 \}$, for some PD $\mathbf{Q} \in \mathbb{S}^n$, $\mathcal{E}(\mathbf{P}) \supseteq \mathcal{E}(\mathbf{Q}) \Longleftrightarrow \mathbf{P} \succeq \mathbf{Q}$