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 $\mathbf{A} \in \mathbb{R}^{n \times n}$,

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

- 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 $\mathbf{x} \in \mathbb{C}^n$

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

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

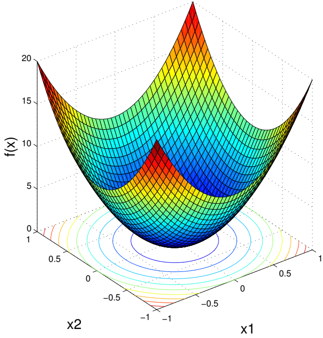
- **Fact**: f is convex if and only if $\nabla^2 f(\mathbf{x}) \succeq \mathbf{0}$ for all $\mathbf{x}$ in the problem domain
- **Example**: The Hessian of the quadratic function

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

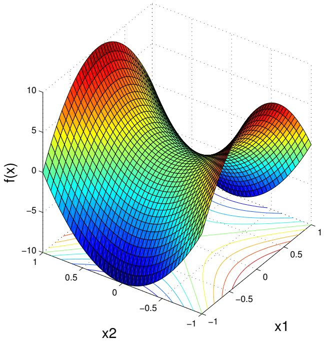
  is given by $\nabla^2 f(\mathbf{x}) = \mathbf{R}$

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

Illustration of Quadratic Functions



(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,*

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

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

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

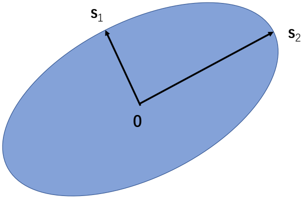
The PD case can be proved in the same way

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

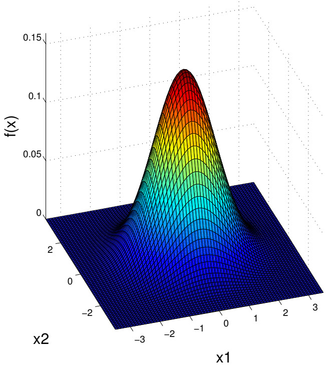
- Probability density function for a Gaussian-distributed vector $\mathbf{x} \in \mathbb{R}^n$:

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

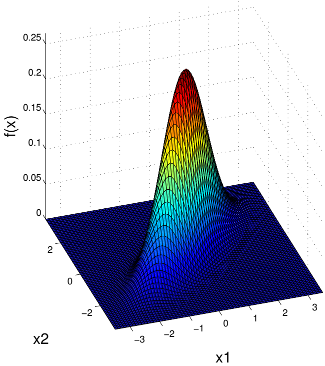
  where $\boldsymbol{\mu}$ and Σ are the mean and covariance of $\mathbf{x}$, respectively

  - 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 $\mathbf{A} = \mathbf{V}\Lambda\mathbf{V}^T$

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

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

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

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

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

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

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 $\mathbf{C} = \mathbf{B}^T \mathbf{A} \mathbf{B}$. The following properties hold:

1. $\mathbf{A} \succeq \mathbf{0} \Longrightarrow \mathbf{C} \succeq \mathbf{0}$
2. With $\mathbf{A} \succ \mathbf{0}$,
$$\mathbf{C} \succ \mathbf{0} \Longleftrightarrow \ \mathbf{B} \text{ has full column rank}$$
3. With nonsingular $\mathbf{B}$,
$$\mathbf{A} \succ \mathbf{0} \Longleftrightarrow \mathbf{C} \succ \mathbf{0}, \quad \mathbf{A} \succeq \mathbf{0} \Longleftrightarrow \mathbf{C} \succeq \mathbf{0}$$

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 $\mathcal{R}(\mathbf{B}) = \mathbb{R}^k$
- $\mathcal{R}(\mathbf{AB}) = \{\mathbf{y} = \mathbf{Az} \mid \mathbf{z} \in \mathcal{R}(\mathbf{B})\} = \{\mathbf{y} = \mathbf{Az} \mid \mathbf{z} \in \mathbb{R}^k\} = \mathcal{R}(\mathbf{A})$

Corollary: 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)