SI231b: Matrix Computations

Lecture 16: Symmetric Positive Definite Matrices

Yue Qiu

qiuyue@shanghaitech.edu.cn

School of Information Science and Technology

ShanghaiTech University

Nov. 5, 2020

Quadratic Form

We use $\mathbb{S}^n$ to denote the set of $n \times n$ symmetric matrices. 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.

Some basic facts (try to verify):

- $\mathbf{x}^T \mathbf{A} \mathbf{x} = \sum_{i=1}^n \sum_{j=1}^n x_i x_j a_{ij}$
- it suffices to consider symmetric $\mathbf{A}$ since 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 $\mathbf{x} \in \mathbb{C}^n$
  - when $\mathbf{A}$ is Hermitian, $\mathbf{x}^H \mathbf{A} \mathbf{x}$ is real for any $\mathbf{x} \in \mathbb{C}^n$

Positive Semidefinite Matrices

When we talk about definite or semidefinite matrices, they are assumed to be symmetric by default.

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 $\mathbf{x} \neq \mathbf{0}$
- indefinite if $\mathbf{A}$ is not PSD

Notation:

- $\mathbf{A} \succeq \mathbf{0}$ means that $\mathbf{A}$ is PSD
- $\mathbf{A} \succ \mathbf{0}$ means that $\mathbf{A}$ is PD
- $\mathbf{A} \not\succeq \mathbf{0}$ means that $\mathbf{A}$ is indefinite

Example: Covariance Matrices

- let $\mathbf{y}_0, \mathbf{y}_2, \ldots \mathbf{y}_{T-1} \in \mathbb{R}^n$ be a sequence of multi-dimensional data samples
  - examples: patches in image processing, multi-channel signals in signal processing, history of returns of assets in finance **[Brodie-Daubechies-et al.'09]**, ...
- 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

Reference

- 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 $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a twice differentiable function
- the Hessian of f, denoted by $\nabla^2 f(\mathbf{x}) \in \mathbb{S}^n$, is a matrix whose (i,j)th entry is given by

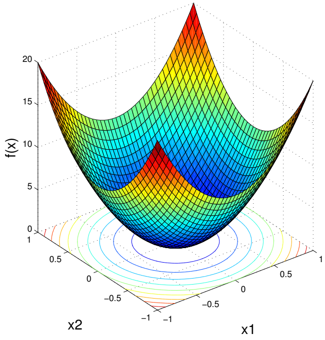
$$\left[\nabla^2 f(\mathbf{x})\right]_{i,j} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

- 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: consider the quadratic function

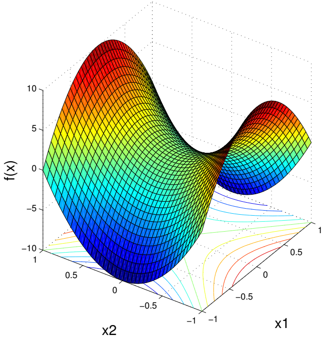
$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{R}\mathbf{x} + \mathbf{q}^T\mathbf{x} + c$$

It can be verified that $\nabla^2 f(\mathbf{x}) = \mathbf{R}$. Thus, f is convex if and only if $\mathbf{R} \succeq \mathbf{0}$

Illustration of Quadratic Functions



(a) PSD **A**.



(b) indefinite **A**.

Theorem. Let $\mathbf{A} \in \mathbb{S}^n$, and let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of $\mathbf{A}$. We have

1. $\mathbf{A} \succeq \mathbf{0} \Longleftrightarrow \lambda_i \geq 0$ for $i = 1, \ldots, n$
2. $\mathbf{A} \succ \mathbf{0} \Longleftrightarrow \lambda_i > 0$ for $i = 1, \ldots, n$

- Proof: let $\mathbf{A} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T$ be the eigendecomposition of $\mathbf{A}$.

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

The PD case is proven by the same manner.

Diagonally Dominant Matrices

The matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is diagonally dominant if

$$|a_{ii}| \geq \sum_{j\neq i} |a_{ij}|, \text{ for all } i$$

When $\mathbf{A}$ is strictly diagonally dominant, the '$\geq$' in the above inequality is replaced by '$>$'.

Property. Let $\mathbf{A} \in \mathbb{S}^n$ be diagonally dominant, then $\mathbf{A}$ is PSD (PD when $\mathbf{A}$ is strictly diagonally dominant).

Proof: you can apply the Gershgorin circle theorem to show this.

Theorem [Gershgorin circle theorem]. For an $n \times n$ matrix $\mathbf{A}$, define

$$r_i = \sum_{j\neq i} |a_{ij}|, \text{ for all } i.$$

Then each eigenvalue of $\mathbf{A}$ is in at least one of the disks

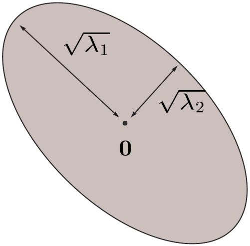
$$\{z : |z - a_{ii}| \leq r_i\}, \text{ for all } i.$$

Example: Ellipsoid

- an ellipsoid of $\mathbb{R}^n$ is defined as

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

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



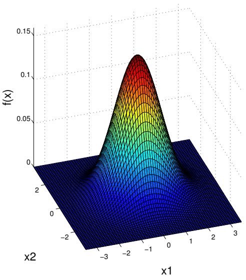
- let $\mathbf{P} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ be the eigendecomposition
  - $\mathbf{V}$ determines the directions of the semi-axes
  - $\lambda_1, \ldots, \lambda_n$ determine the lengths of the semi-axes

Example: Multivariate Gaussian Distribution

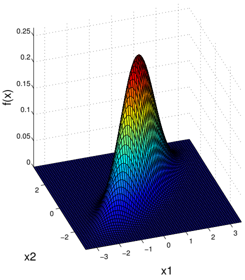
Probability density function for a multivariate Gaussian distribution

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

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean and covariance of $\mathbf{x}$, resp. $\boldsymbol{\Sigma}$ is PD and determines how $\mathbf{x}$ is spread, by the same way as in ellipsoid



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



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

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: $\mathbf{A} = \mathbf{B}^T\mathbf{B} \Longrightarrow \mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{x}^T\mathbf{B}^T\mathbf{B}\mathbf{x} = \|\mathbf{B}\mathbf{x}\|_2^2 \geq 0$ for all $\mathbf{x}$
- necessity: let $\mathbf{\Lambda}^{1/2} = \mathrm{diag}(\lambda_1^{1/2}, \ldots, \lambda_n^{1/2})$.

$$\mathbf{A} \succeq \mathbf{0} \Longrightarrow \mathbf{A} = (\mathbf{V}\mathbf{\Lambda}^{1/2})(\mathbf{\Lambda}^{1/2}\mathbf{V}^T), \text{ with } \mathbf{\Lambda}^{1/2}\mathbf{V}^T \text{ being real}$$

PSD Matrices and Square Root

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

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

  - $\mathbf{B} = \mathbf{A}^{1/2}$ 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 $\mathbf{A} = \mathbf{B}^T\mathbf{B}$ (how to prove it?)
- $\mathbf{A}^{1/2}$ 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 $\mathbf{A} = \mathbf{B}^2$

- 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)

Property. 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. $\mathbf{A} \succeq \mathbf{0} \Longrightarrow \mathbf{C} \succeq \mathbf{0}$
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:

$$\mathbf{C} \succ \mathbf{0} \Longleftrightarrow \mathbf{z}^T \mathbf{A} \mathbf{z} > 0, \ \forall\, \mathbf{z} \in \mathcal{R}(\mathbf{B}) \setminus \{\mathbf{0}\}. \qquad (*)$$

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.

Lemma. Let $\mathbf{A} \in \mathbb{R}^{m\times k}$ and $\mathbf{B} \in \mathbb{R}^{k\times n}$ $(k \leq \{m, n\})$, and suppose that $\mathbf{B}$ has full row rank. Then

$$\mathcal{R}(\mathbf{AB}) = \mathcal{R}(\mathbf{A})$$

Proof:

- observe that $\dim \mathcal{R}(\mathbf{B}) = \text{rank}(\mathbf{B}) = k$, which implies $\mathcal{R}(\mathbf{B}) = \mathbb{R}^k$.
- we have $\mathcal{R}(\mathbf{AB}) = \{\mathbf{y} = \mathbf{Az} \mid \mathbf{z} \in \mathcal{R}(\mathbf{B})\} = \{\mathbf{y} = \mathbf{Az} \mid \mathbf{z} \in \mathbb{R}^k\} = \mathcal{R}(\mathbf{A})$.

Property: let $\mathbf{R}$ be a PSD matrix. Suppose that we factor $\mathbf{R}$ as $\mathbf{R} = \mathbf{BB}^T$ for some full-column rank $\mathbf{B}$. Then, $\mathcal{R}(\mathbf{R}) = \mathcal{R}(\mathbf{B})$.

Property. 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,

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

- according to the lemma, 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.