SI231 Matrix Analysis and Computations
Positive Semidefinite Matrices

Ziping Zhao

Spring Term 2022–2023

School of Information Science and Technology
ShanghaiTech University, Shanghai, China

http://si231.sist.shanghaitech.edu.cn

Positive Semidefinite Matrices

- positive semidefinite (PSD) matrices
- properties of PSD matrices
- PSD matrix inequalities
- Schur complement
- application: factor models
- application: graph matrices
- application: log-determinant function
- application: subspace method for super-resolution spectral analysis
- application: Euclidean distance matrices

Hightlights

- a matrix $\mathbf{A} \in \mathbb{S}^n$ is said to be positive semidefinite (PSD) if

$$\mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0, \quad \text{for all } \mathbf{x} \in \mathbb{R}^n;$$

  and positive definite (PD) if

$$\mathbf{x}^T \mathbf{A} \mathbf{x} > 0, \quad \text{for all } \mathbf{x} \in \mathbb{R}^n \text{ with } \mathbf{x} \neq \mathbf{0}$$

- in this case, we are interested in symmetric positive semidefinite matrices
- a matrix $\mathbf{A} \in \mathbb{S}^n$ is PSD (resp. PD)
  - if and only if its eigenvalues are all non-negative (resp. positive);
  - if and only if it can be factored as $\mathbf{A} = \mathbf{B}^T \mathbf{B}$ for some $\mathbf{B} \in \mathbb{R}^{m \times n}$
- there are more general definitions of definiteness, including real non-symmetric matrices, or non-Hermitian ones
- in this lecture, we will deal with the real symmetric matrices—the Hermitian case follows along the same lines

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.

- some basic facts (try to verify):
  - $\mathbf{x}^T \mathbf{A} \mathbf{x} = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j = \sum_{i=1}^n a_{ii} x_i^2 + \sum_{i=1}^{n-1} \sum_{j=i+1}^n 2 a_{ij} x_i x_j$
  - $\mathbf{x}^T \mathbf{A} \mathbf{x} = \sum_{i=1}^n a_{ii} x_i^2 + \sum_{i=1}^{n-1} \sum_{j=i+1}^n (a_{ij} + a_{ji}) x_i x_j$ for general $\mathbf{A} \in \mathbb{R}^{n \times n}$, there may exist $\mathbf{A}_1$ and $\mathbf{A}_2$ s.t. $\mathbf{x}^T \mathbf{A}_1 \mathbf{x} = \mathbf{x}^T \mathbf{A}_2 \mathbf{x}$
    * it suffices to consider unique symmetric $\mathbf{A}$ for general $\mathbf{A} \in \mathbb{R}^{n \times n}$ since

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

  - complex case:
    * the (complex) 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 $\mathbf{x} \in \mathbb{R}^n$ (not very often called nonnegative definite)
- 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 both $\mathbf{A}$ and $-\mathbf{A}$ are 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

- if $\mathbf{A}$ is PD, then it is also PSD
- a quadratic form is called PSD (resp. PD) if $\mathbf{A}$ is PSD (resp. PD)

- Concepts negative semidefinite (NSD) and negative definite (ND) may be defined by reversing the inequalities or, equivalently, by saying $-\mathbf{A}$ is PSD or PD, resp.
- Positive (semi)definite and negative (semi)definite matrices together are called definite matrices.

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: multivariate features in machine learning, patches in image processing, multi-channel signals in signal processing, history of returns of assets in finance **[Brodie-Daubechies-et al.'09]**, ...
- sample mean: $\hat{\boldsymbol{\mu}} = \frac{1}{T}\sum_{t=0}^{T-1} \mathbf{y}_t$
- sample covariance: $\hat{\boldsymbol{\Sigma}} = \frac{1}{T}\sum_{t=0}^{T-1}(\mathbf{y}_t - \hat{\boldsymbol{\mu}})(\mathbf{y}_t - \hat{\boldsymbol{\mu}})^T$
- a sample covariance is PSD: $\mathbf{x}^T\hat{\boldsymbol{\Sigma}}\mathbf{x} = \frac{1}{T}\sum_{t=0}^{T-1} |(\mathbf{y}_t - \hat{\boldsymbol{\mu}})^T\mathbf{x}|^2 \geq 0$
- the (statistical) covariance matrix (or variance-covariance mat.) of $\mathbf{y}_t$ is also PSD
  - to put into context, assume that $\mathbf{y}_t$ is a wide-sense (or weakly) stationary random process
  - the covariance, defined as $\boldsymbol{\Sigma} = \mathrm{E}[(\mathbf{y}_t - \boldsymbol{\mu})(\mathbf{y}_t - \boldsymbol{\mu})^T]$ where the mean $\boldsymbol{\mu} = \mathrm{E}[\mathbf{y}_t]$, is PSD: $\mathbf{x}^T\boldsymbol{\Sigma}\mathbf{x} = \mathbf{x}^T\mathrm{E}[(\mathbf{y}_t - \boldsymbol{\mu})(\mathbf{y}_t - \boldsymbol{\mu})^T]\mathbf{x} = \mathrm{E}[(|\mathbf{y}_t - \boldsymbol{\mu})^T\mathbf{x}|^2] \geq 0$

Example: Covariance Matrices

- define $\mathbf{Y} = [\mathbf{y}_0\ \mathbf{y}_1\ \ldots\ \mathbf{y}_{T-1}] \in \mathbb{R}^{n\times T}$
- the sample mean: $\hat{\boldsymbol{\mu}} = \frac{1}{T}\mathbf{Y}\mathbf{1}$
- subtracting $\hat{\boldsymbol{\mu}}$ from each column gives the row-centered data matrix

$$\mathbf{Y}_c = \mathbf{Y} - \hat{\boldsymbol{\mu}}\mathbf{1}^T = \mathbf{Y} - \frac{1}{T}\mathbf{Y}\mathbf{1}\mathbf{1}^T = \mathbf{Y}(\mathbf{I} - \frac{1}{T}\mathbf{1}\mathbf{1}^T) = \mathbf{Y}\mathbf{C}$$

  where $\mathbf{C} = \mathbf{I} - \frac{1}{T}\mathbf{1}\mathbf{1}^T$ is called the centering matrix
  - a symmetric, PSD, and idempotent matrix, and $\mathrm{rank}(\mathbf{C}) = T - 1$
  - singular and has the eigenvalue 1 of multiplicity $n-1$ and eig. 0 of multip. 1
  - has a nullspace of dimension 1, along the vector $\mathbf{1}$
  - an orthogonal projection matrix (projection to a subspace of all n-dim. vectors whose components sum to zero)
  - left-multiplying it with a vector has the same effect as subtracting the mean of the components of the vector from every component of that vector
- the sample covariance matrix based on $\mathbf{Y}_c$ (a Gram matrix) is

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{T}\mathbf{Y}_c\mathbf{Y}_c^T = \frac{1}{T}\mathbf{Y}\mathbf{C}\mathbf{C}^T\mathbf{Y}^T = \frac{1}{T}\mathbf{Y}\mathbf{C}^2\mathbf{Y}^T = \frac{1}{T}\mathbf{Y}\mathbf{C}\mathbf{Y}^T = \frac{1}{T}\mathbf{Y}\mathbf{Y}^T - \hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}^T$$

Properties of PSD Matrices

results that immediately follow from the definition: let $\mathbf{A}, \mathbf{B}, \mathbf{C} \in \mathbb{S}^n$.

- $\mathbf{A} \succeq \mathbf{0}, \alpha \geq 0$ (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. $\mathbf{A} \succeq \mathbf{0}, \mathbf{B} \succ \mathbf{0}$) $\Longrightarrow \mathbf{A} + \mathbf{B} \succeq \mathbf{0}$ (resp. $\mathbf{A} + \mathbf{B} \succ \mathbf{0}$)

- $\mathbf{A} \succ \mathbf{0}, \mathbf{B} \succeq \mathbf{0} \Longrightarrow \mathbf{A} + \mathbf{B} \succeq \mathbf{0}$

- $\mathbf{A} \succ \mathbf{0} \Longleftrightarrow \mathbf{A}^{-1} \succ \mathbf{0}$

- $\mathbf{A} \succeq \mathbf{0}$ (resp. $\mathbf{A} \succ \mathbf{0}$) $\Longrightarrow \mathrm{tr}(\mathbf{A}) \geq 0$ and $\det(\mathbf{A}) \geq 0$ (resp. $\mathrm{tr}(\mathbf{A}) > 0$ and $\det(\mathbf{A}) > 0$)

- The set of PSD matrices is convex. (the basis of semidefinite programming (SDP), a subfield of convex optimization in matrix variables)

- ...

Properties of PSD Matrices: Eigenvalues

Theorem 1. 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 $i = 1, \ldots, n$
3. $\mathbf{A} \not\succeq \mathbf{0} \Longleftrightarrow \lambda_i > 0$ for some i and $\lambda_i < 0$ for some i

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

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

The PD and indefinite cases are proven by the same manner.

Example: Hessian

- let $f : \mathbb{R}^n \to \mathbb{R}$ 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

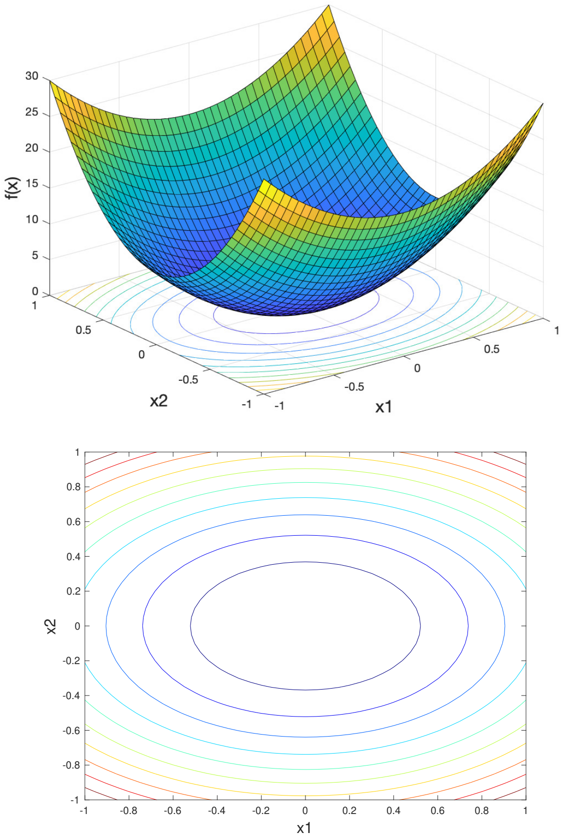
- example: consider the quadratic function with $\mathbf{R} \in \mathbb{S}^n$

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

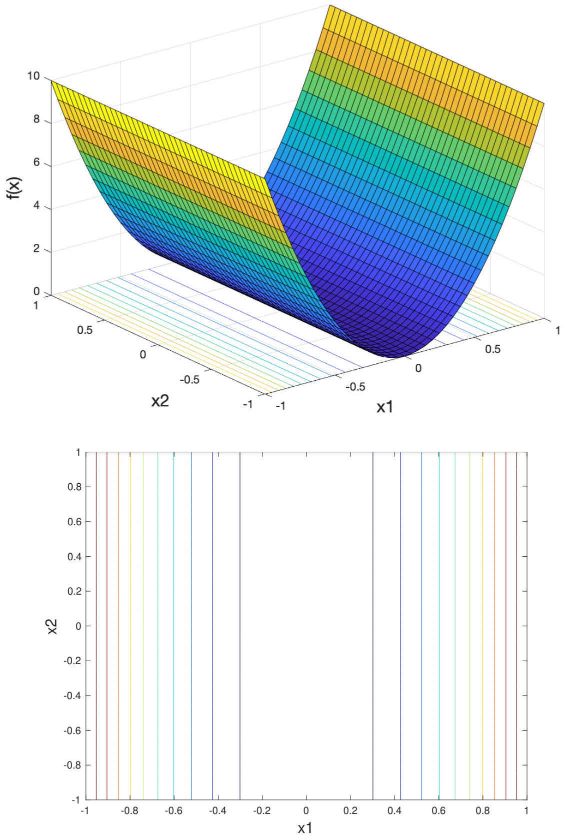
  It can be verified that $\nabla^2 f(\mathbf{x}) = \mathbf{R}$. Thus, f is convex (resp. strictly convex) if and only if $\mathbf{R} \succeq \mathbf{0}$ (resp. $\succ \mathbf{0}$)

Illustration of Quadratic Functions

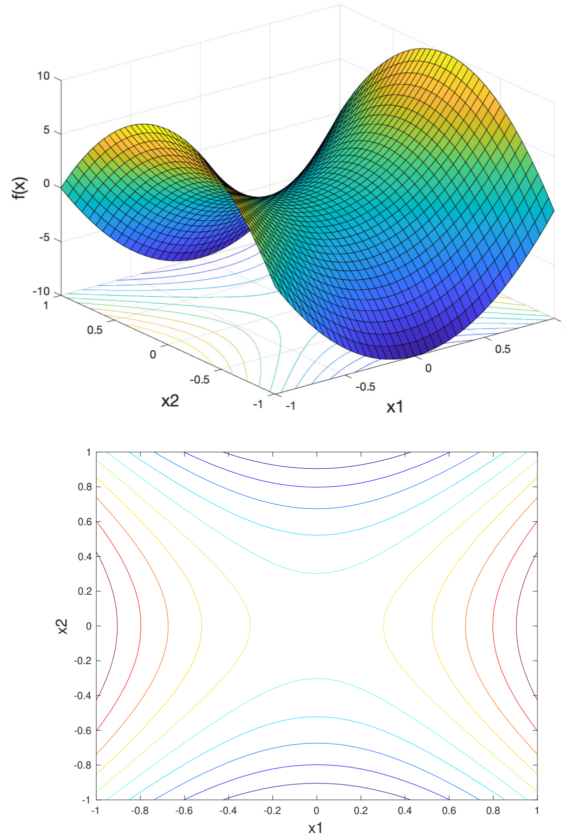
$$f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$$



(a) PSD $\mathbf{A}$ with all $\lambda_i > 0$.



(b) PSD $\mathbf{A}$ with some $\lambda_i = 0$.



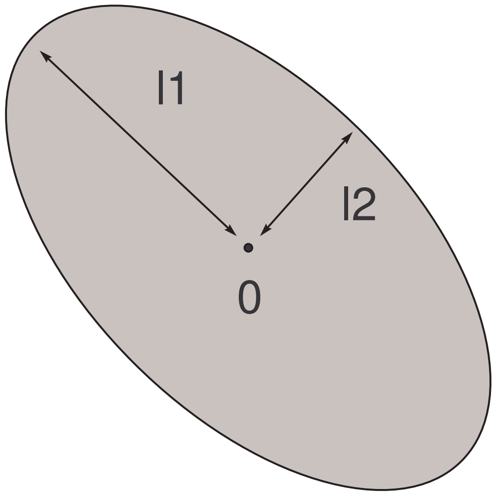
(c) indefinite $\mathbf{A}$.

Example: Ellipsoid

- an ellipsoid of $\mathbb{R}^n$ centered at $\mathbf{0}$ is defined as

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

  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

– $\ell_i = \lambda_i^{\frac{1}{2}} \mathbf{v}_i$

- note:
  – in direction $\mathbf{v}_1$, $\mathbf{x}^T \mathbf{P}^{-1} \mathbf{x}$ is large, hence ellipsoid is fat in direction $\mathbf{v}_1$
  – in direction $\mathbf{v}_n$, $\mathbf{x}^T \mathbf{P}^{-1} \mathbf{x}$ is small, hence ellipsoid is thin in direction $\mathbf{v}_n$
  – $\sqrt{\lambda_1/\lambda_n}$ gives maximum eccentricity

Example: Multivariate Gaussian Distribution

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

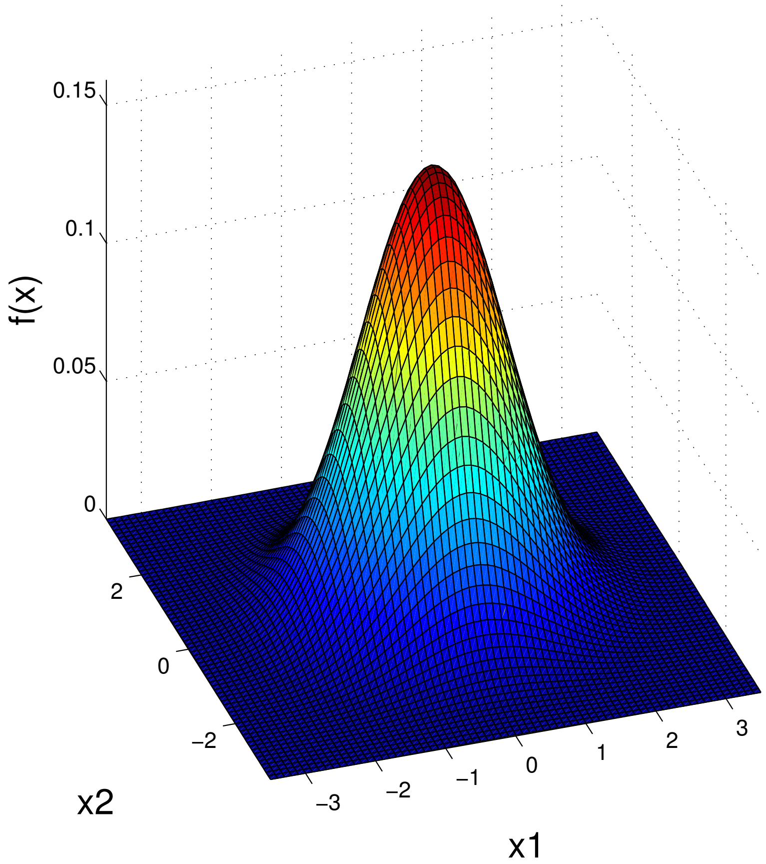
$$p(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{n}{2}}(\det(\boldsymbol{\Sigma}))^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right)$$

  where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean and covariance matrix of $\mathbf{x}$, resp.

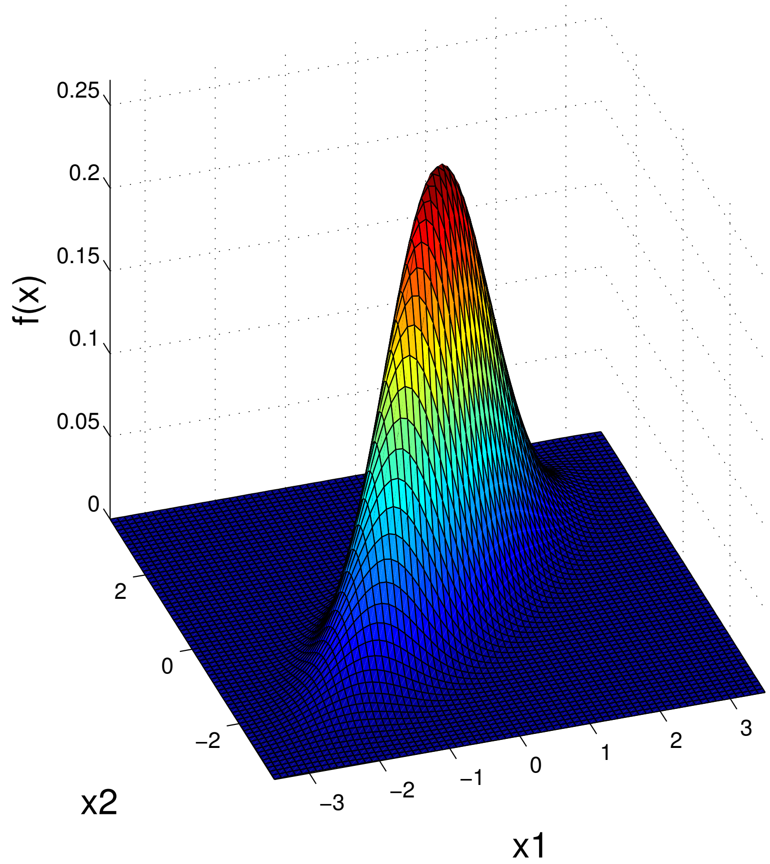
  – $\boldsymbol{\Sigma}$ is PD

  – $\boldsymbol{\Sigma}$ determines how $\mathbf{x}$ is spread, by the same way as in ellipsoid

Example: Multivariate Gaussian Distribution



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



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

Example: Multivariate Generalized Gaussian Distribution

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

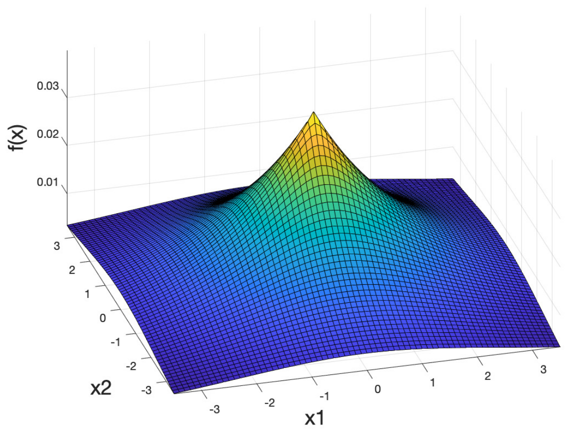
$$p(\mathbf{x}; \mathbf{m}, \mathbf{C}, \nu) = \frac{\frac{\nu}{2}\Gamma(\frac{n}{2})}{(2^{\frac{2}{\nu}}\pi)^{\frac{n}{2}}\Gamma(\frac{n}{\nu})(\det(\mathbf{C}))^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\left((\mathbf{x}-\mathbf{m})^T\mathbf{C}^{-1}(\mathbf{x}-\mathbf{m})\right)^{\frac{\nu}{2}}\right)$$

  where $\mathbf{m}$ and $\mathbf{C}$ are the location and scatter matrix (a.k.a. scale matrix) of $\mathbf{x}$, resp., and ν is the shape parameter (a.k.a. form parameter)
  - $\mathbf{C}$ is PD and determines how $\mathbf{x}$ is spread
  - ν affect the shape of the distribution
    * it becomes the multivariate Gaussian distribution when when $\nu = 2$
    * more peaky with heavy tails if $\nu < 2$ (it becomes a multivariate Laplacian distribution when $\nu = 1$)
    * less peaky with light tails if $\nu > 2$ (it tends to converge to a multivariate uniform distribution when $\nu \to \infty$)

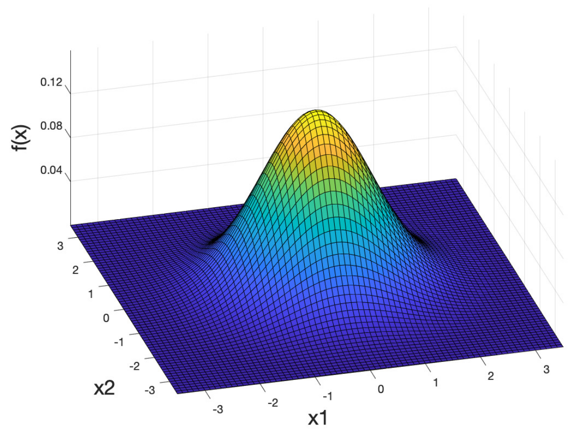
- For Gaussian distribution, the location and scatter parameters correspond to the mean and covariance, resp. But this is not necessarily true for other distributions.

Example: Multivariate Generalized Gaussian Distribution

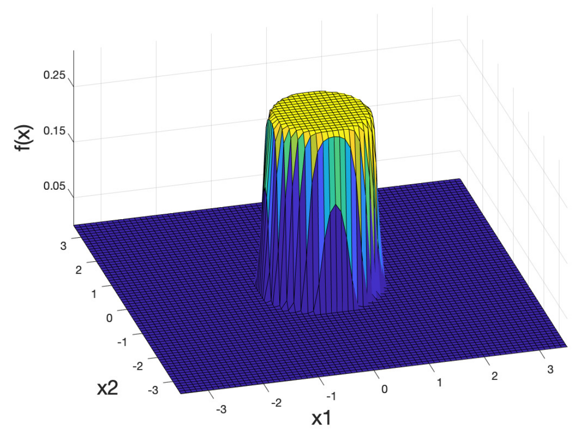
$$\mathbf{m} = \mathbf{0}, \quad \mathbf{C} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$



(a) $\nu = 1$.



(b) $\nu = 2$.



(c) $\nu = 30$.

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 (resp., 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 (resp., 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

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
- $\mathbf{A} \succ \mathbf{0} \Longleftrightarrow \mathbf{A}^{-1} \succ \mathbf{0}$