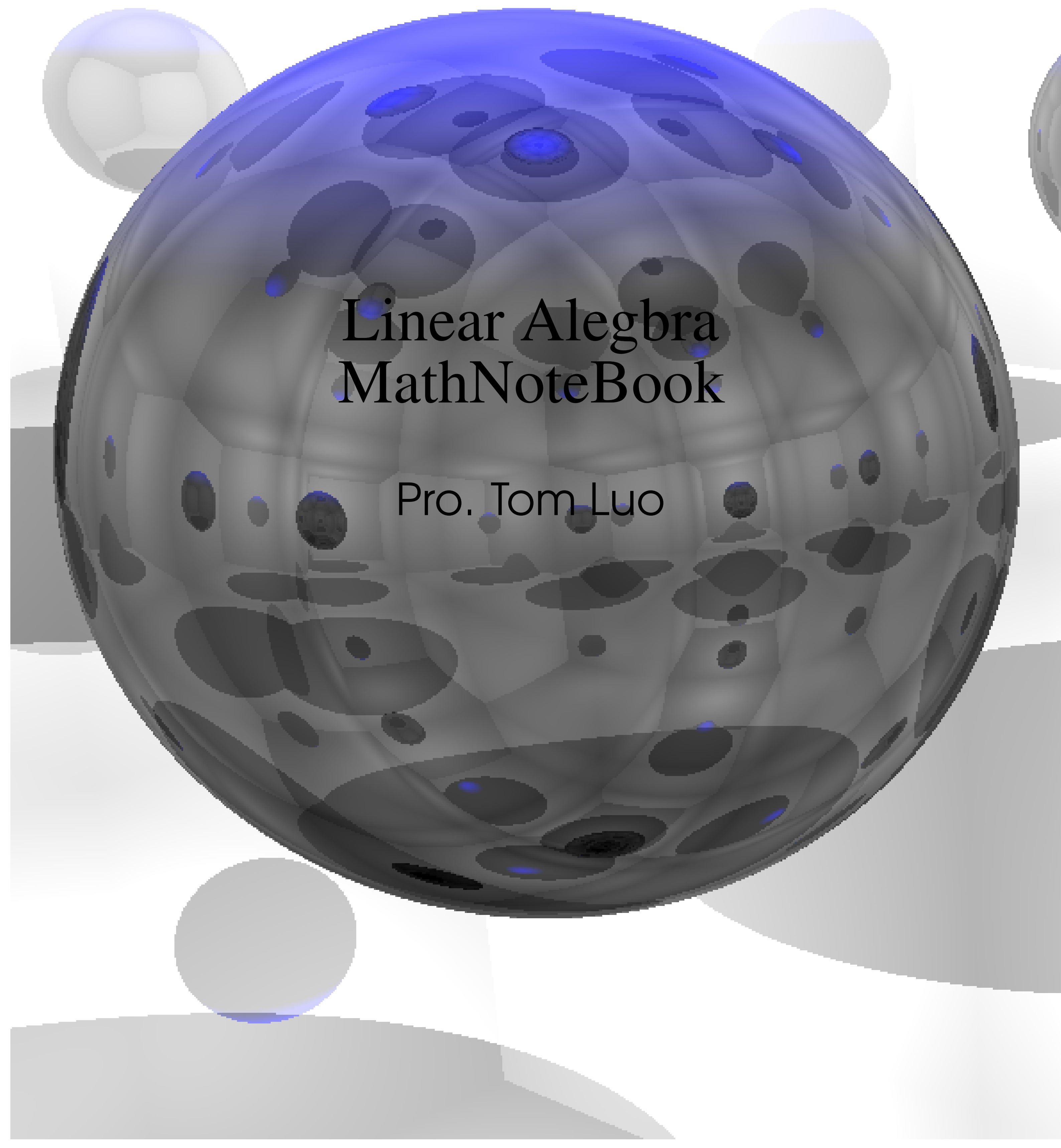

Linear Alegbra MathNoteBook

Pro. Tom Luo

17 — Week7

17.1 Tuesday

17.1.1 *Quadratic form*

The graphs of the following equations are easy to plot:

$$x^2 + y^2 = 1 \implies \text{Circle.} \tag{17.1}$$

$$\frac{x^2}{2} + \frac{y^2}{5} = 1 \implies \text{Elipse.} \tag{17.2}$$

$$\frac{x^2}{2} - \frac{y^2}{5} = 1 \implies \text{Hyperbola.} \tag{17.3}$$

$$\left.\begin{aligned} x^2 &= \alpha y \\ y^2 &= \alpha x \end{aligned}\right\} \implies \text{Parabola.}$$

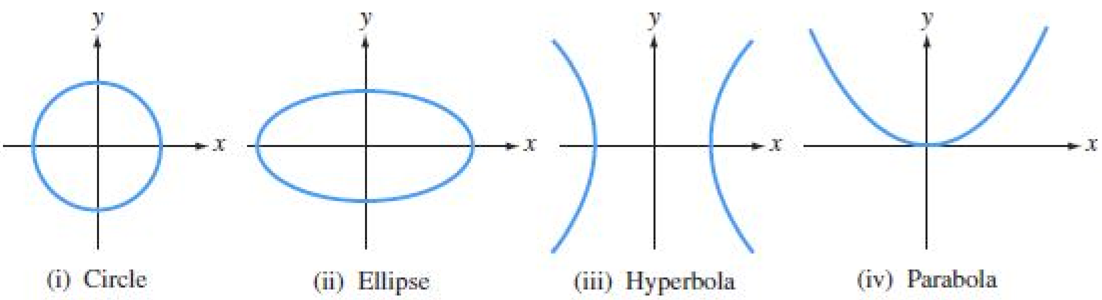

Figure 17.1: graph for quadratic form equations of two variables

The equations (17.1) – (17.3) is the *quadratic form equations of two variables*. Now we give the general form for quadratic equations:

Definition 17.1 — Quadratic form. The formula of **quadratic form** is given by

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$$

where $\boldsymbol{A}$ is $n \times n$ matrix, $\boldsymbol{x} \in \mathbb{R}^n$.
Moreover, sometimes we write $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$ as:

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \sum_{i,j=1}^{n} x_i x_j a_{ij}$$

where x_i is the ith entry of $\boldsymbol{x}$ and a_{ij} are (i,j)th entry of $\boldsymbol{A}$. ■

Moverover, we say an equation is the **conic section of quadratic form** if it can be written as

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = 1.$$

- Note that $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}$. Why?
  If we take the transpose of $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$, since it is a number, so we obtain
  $$(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x})^{\mathrm{T}} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}.$$
  Since $(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x})^{\mathrm{T}} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}$, finally we derive
  $$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}.$$
- Hence given any matrix $\boldsymbol{A}$, we always have
  $$\begin{aligned}\boldsymbol{x}^{\mathrm{T}}\left(\frac{\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}}}{2}\right)\boldsymbol{x} &= \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} + \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x} \\ &= \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} + \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} \\ &= \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}.\end{aligned}$$
  Note that $\left(\frac{\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}}}{2}\right)$ is *symmetric*! Hence given any $\boldsymbol{A}$, if we want to study its quadratic form, we can always convert this matrix into symmetric matrix.

Hence without loss of generality, we assume $\boldsymbol{A} = \boldsymbol{A}^{\mathrm{T}}$ during the section of quadratic form.

■ **Example 17.1**
Given the equation $3x^2+2xy+3y^2=1$, how we transform it into the conic section of quadratic form? And how can we determine its shape in view of matirx?
Actually, It can be written as

$$\begin{pmatrix} x & y \end{pmatrix}\begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = 1. \qquad \textbf{conic section of quadratic form.} \tag{17.4}$$

And we define $\boldsymbol{A} = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$. If we do the eigendecomposition for $\boldsymbol{A}$, we obtain

$$\boldsymbol{A} = \boldsymbol{Q}\Lambda\boldsymbol{Q}^{\mathrm{T}}$$

where $\Lambda = \begin{pmatrix} \lambda_1 & \\ & \lambda_2 \end{pmatrix}$, $\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 \end{bmatrix}$. $\boldsymbol{x}_1,\boldsymbol{x}_2$ is the eigenvectors of $\boldsymbol{A}$ corresponding to eigenvalues λ_1,λ_2 respectively.

Thus we convert equation (17.4) into

$$\begin{pmatrix} x & y \end{pmatrix} \boldsymbol{Q}\Lambda\boldsymbol{Q}^{\mathrm{T}} \begin{pmatrix} x \\ y \end{pmatrix} = 1 \implies \tilde{\boldsymbol{x}}^{\mathrm{T}}\Lambda\tilde{\boldsymbol{x}} = 1.$$

where $\tilde{\boldsymbol{x}} = \boldsymbol{Q}^{\mathrm{T}} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{pmatrix}$.

Then how to determine the shape of this equation? We just do matrix multiplication to obtain:

$$\lambda_1 \tilde{x}_1^2 + \lambda_2 \tilde{x}_2^2 = 1.$$

After computation, we find $\lambda_1 = 4, \lambda_2 = 2$. Hence this equation is a **elipse.** ■

R

Matrix Calculus

Now we recall how to compute derivative for matrix:

- $\frac{\partial(f^{\mathrm{T}}g)}{\partial x} = \frac{\partial f(x)}{\partial x} g(x) + \frac{\partial g(x)}{\partial x} f(x)$

Example:

- $\frac{\partial(a^{\mathrm{T}}\boldsymbol{x})}{\partial \boldsymbol{x}} = a$
- $\frac{\partial(a^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{x}} = \frac{\partial((\boldsymbol{A}^{\mathrm{T}}a)^{\mathrm{T}}\boldsymbol{x})}{\partial \boldsymbol{x}} = \boldsymbol{A}^{\mathrm{T}}a$
- $\frac{\partial(\boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{x}} = \boldsymbol{A}^{\mathrm{T}}$
- $\frac{\partial(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}$

■ **Example 17.2**

Given $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}$. We want to do the optimization:

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x})$$

How to find the optimal solution? The direct idea is to take the first order derivative:

$$\begin{aligned} \frac{\partial f}{\partial \boldsymbol{x}} &= \frac{1}{2}\frac{\partial(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{x}} + \frac{\partial(\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x})}{\boldsymbol{x}} \\ &= \frac{1}{2}(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}) + \boldsymbol{b}. \end{aligned}$$

Since $\boldsymbol{A}$ is symmetric, we obtain

$$\frac{\partial f}{\partial \boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}.$$

If $\boldsymbol{x}^*$ is an optimal solution, then it must satisfy:

$$\triangledown f(\boldsymbol{x}^*) = \frac{\partial f(\boldsymbol{x}^*)}{\partial \boldsymbol{x}} = \boldsymbol{0} \implies \boldsymbol{A}\boldsymbol{x}^* + \boldsymbol{b} = \boldsymbol{0}.$$

There may follow these cases:

- If equation $\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} = \boldsymbol{0}$ has no solution, then $f(\boldsymbol{x})$ is unbounded.
  This statement is remained to be proved.

- If equation $\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}=\boldsymbol{0}$ has a solution $\boldsymbol{x}^*$, it doesn't mean $\boldsymbol{x}^*$ is an optimal solution. (Note that the reverse is true.)
  Let's raise a counterexample: if we set
  $$\boldsymbol{A}=\begin{bmatrix}1 & 0\\0 & -1\end{bmatrix},\qquad \boldsymbol{b}=\boldsymbol{0},\qquad \boldsymbol{x}=\begin{pmatrix}x_1\\x_2\end{pmatrix},$$
  then $f(\boldsymbol{x})=\frac{1}{2}(x_1^2-x_2^2)$.
  One solution to $\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}=\boldsymbol{0}$ is $\boldsymbol{x}^*=\begin{pmatrix}0\\0\end{pmatrix}$. Obviously, $\boldsymbol{x}^*$ is not a optimal solution.
  Intutively, if we let $x_1=0, x_2\to\infty$, then $f(\boldsymbol{x})\to-\infty$!

Second optimality condition

If $\boldsymbol{x}^*$ is a optimal solution to $f(\boldsymbol{x})$, what else condition should $\boldsymbol{x}^*$ satisfy?
Let's take $f(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}$ as an example, we want to find $\boldsymbol{x}^*$ s.t. $\min f(\boldsymbol{x})=f(\boldsymbol{x}^*)$.
Firstly, we convert $f(\boldsymbol{x})$ into its *taylor expansion*:

$$f(\boldsymbol{x})=f(\boldsymbol{x}^*)+\langle\triangledown f(\boldsymbol{x}^*),\boldsymbol{x}-\boldsymbol{x}^*\rangle+\frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}^*)^{\mathrm{T}}\triangledown^2 f(\boldsymbol{x}^*)(\boldsymbol{x}-\boldsymbol{x}^*).$$

Note that $\triangledown^2 f(\boldsymbol{x}^*)$ is the Hessian matrix of $f(\boldsymbol{x}^*)$, which is defined as

$$\triangledown^2 f(\boldsymbol{x}^*):=\left[\frac{\partial^2 f(\boldsymbol{x}^*)}{\partial x_i\partial x_j}\right]=\triangledown(\triangledown f(\boldsymbol{x}^*)).$$

Firstly we compute $\triangledown f(\boldsymbol{x})$ and $\triangledown^2 f(\boldsymbol{x})$:

$$\triangledown f(\boldsymbol{x})=\frac{1}{2}(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x})+\boldsymbol{b}.$$
$$\triangledown^2 f(\boldsymbol{x})=\triangledown\left[\frac{1}{2}(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x})+\boldsymbol{b}\right]=\frac{1}{2}\triangledown(\boldsymbol{A}\boldsymbol{x})+\frac{1}{2}\triangledown(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x})=\frac{1}{2}(\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}}).$$

If we assume $\boldsymbol{A}$ is **symmetric**, then we have $\triangledown f(\boldsymbol{x})=(\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}})\boldsymbol{x}+\boldsymbol{b}$ and $\triangledown^2 f(\boldsymbol{x})=\boldsymbol{A}$.
Since the optimal solution $\boldsymbol{x}^*$ must satisfy $\triangledown f(\boldsymbol{x}^*)=\boldsymbol{0}$, we deive

$$\langle\triangledown f(\boldsymbol{x}^*),\boldsymbol{x}-\boldsymbol{x}^*\rangle=0.\implies f(\boldsymbol{x})=f(\boldsymbol{x}^*)+\frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}^*)^{\mathrm{T}}\triangledown^2 f(\boldsymbol{x}^*)(\boldsymbol{x}-\boldsymbol{x}^*).$$

Hence $f(\boldsymbol{x})-f(\boldsymbol{x}^*)=\frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{x}^*)$.
Since $\boldsymbol{x}^*$ is optimal that minimize $f(\boldsymbol{x})$, $LHS=f(\boldsymbol{x})-f(\boldsymbol{x}^*)\geq 0$.

$$\implies\frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{x}^*)\geq 0$$

Or equivalently,

$$(\boldsymbol{x}-\boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{x}^*)\geq 0 \text{ for } \forall\boldsymbol{x}.\iff\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}\geq 0\text{ for }\forall\boldsymbol{x}.$$

Our conclusion is that if there exists a optimal solution for $f(\boldsymbol{x})$, then the matrix $\boldsymbol{A}$ should satisfy $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}\geq 0$ for $\forall\boldsymbol{x}$. We have a specific name for such $\boldsymbol{A}$.

17.1.2 Positive Definite Matrices

Definition 17.2 — Positive-definite.

- Matrix $\boldsymbol{A}$ is *positive-semidefinite* (PSD) if $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} \geq 0$ for $\forall \boldsymbol{x}$. And we denote it as $\boldsymbol{A} \succeq 0$.
- Matrix $\boldsymbol{A}$ is *positive-definite* (PD) if $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} > 0$ for $\forall \boldsymbol{x} \neq \boldsymbol{0}$. And we denote it as $\boldsymbol{A} \succ 0$.
- Matrix $\boldsymbol{A}$ is *indefinite* if there exist some $\boldsymbol{x}$ and $\boldsymbol{y}$ s.t.

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} < 0 < \boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{y}.$$

■

R If a matrix is PSD or PD, it is usually assumed to be symmetric by default. Even in other textbooks, the definition for PSD and PD contains the *symmetric* condition.

Theorem 17.1 Let $\boldsymbol{A}$ be a $n \times n$ real symmetric matrix, the following are equivalent:

1. $\boldsymbol{A}$ is PD.
2. All eigenvalues of $\boldsymbol{A}$ are positive.
3. All n *upper left square submatrices* $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_n$ all have positive determinants.
4. $\boldsymbol{A}$ could be factorized as $\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}$, where $\boldsymbol{R}$ is nonsingular.

You may be confused about the "upper left submatrices", they are the 1 by 1, 2 by 2,...,n by n submatrices of $\boldsymbol{A}$ on the upper left. The n by n submatrix is exactly $\boldsymbol{A}$. Before we geive a detailed proof, let's show how to test some matrices for positive definiteness by using this theorem:

■ **Example 17.3** Test these matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ for positive definiteness:

$$\boldsymbol{A} = \begin{bmatrix} 1 & & & \\ & 2 & & \\ & & 2 & \\ & & & 2 \end{bmatrix} \quad \text{and} \quad \boldsymbol{B} = \begin{bmatrix} 1 & -1 & & \\ -1 & 2 & -1 & \\ & -1 & 2 & -1 \\ & & -1 & 2 \end{bmatrix}$$

- For matrix $\boldsymbol{A}$, its eigenvalues are $\{1,2,2,2\}$. So all eigenvalues of $\boldsymbol{A}$ are positive, $\boldsymbol{A}$ is PD. Moverover, we can test its positive definiteness by definition:

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = x_1^2 + 2x_2^2 + 2x_3^2 + 2x_4^2 > 0.$$

for $\forall \boldsymbol{x} = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \end{pmatrix}^{\mathrm{T}} \neq \boldsymbol{0}$.
- For matrix $\boldsymbol{B}$, all *upper left square submatrices* is given by

$$\boldsymbol{B}_1 = \begin{bmatrix} 1 \end{bmatrix} \quad \boldsymbol{B}_2 = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \quad \boldsymbol{B}_3 = \begin{bmatrix} 1 & -1 & \\ -1 & 2 & -1 \\ & -1 & 2 \end{bmatrix} \quad \boldsymbol{B}_4 = \begin{bmatrix} 1 & -1 & & \\ -1 & 2 & -1 & \\ & -1 & 2 & -1 \\ & & -1 & 2 \end{bmatrix}$$

After messy computation, we obtain

$$\det(\boldsymbol{B}_1) = 1 \quad \det(\boldsymbol{B}_2) = 1 \quad \det(\boldsymbol{B}_3) = 1 \quad \det(\boldsymbol{B}_4) = 1.$$

Hence all *upper left square determinants* are positive, $\boldsymbol{B}$ is PD.

■

Then we begin to give a proof for this theorem:

Proof.

- (1) $\implies$ (2) : Suppose λ is any eigenvalue of $\boldsymbol{A}$. Then $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$ for some $\boldsymbol{x} \neq \boldsymbol{0}$. By postmutliplying $\boldsymbol{x}^{\mathrm{T}}$ both sides we obtain:

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = \lambda\|\boldsymbol{x}\|^2 \implies \lambda = \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\|\boldsymbol{x}\|^2} > 0.$$

- (2) $\implies$ (1) : Assume all eigenvalues $\lambda_i > 0$ for $i = 1, 2, \dots, n$.
  For $\forall \boldsymbol{x} \neq \boldsymbol{0}$, our goal is to show $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} > 0$:
  Since $\boldsymbol{A}$ is real symmetric matrix, we do eigendecomposition of $\boldsymbol{A}$:

$$\boldsymbol{A} = \boldsymbol{Q}\Lambda\boldsymbol{Q}^{\mathrm{T}} \qquad \boldsymbol{Q} \text{ is orthonormal matrix.}$$

  Hence

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\Lambda\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{x} = (\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{x})^{\mathrm{T}}\Lambda(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{x}).$$

  If we set $\tilde{\boldsymbol{x}} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{x} = \begin{bmatrix}\tilde{x_1} & \dots & \tilde{x_n}\end{bmatrix}$, then $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$ can be written as

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \tilde{\boldsymbol{x}}^{\mathrm{T}}\Lambda\tilde{\boldsymbol{x}} = \sum_{i=1}^{n}\lambda_i\tilde{x_i}^2 \geq 0.$$

  Then we aruge that $\sum_{i=1}^{n}\lambda_i\tilde{x_i}^2 \neq 0$. Actually we only need to show $\|\boldsymbol{x}\| \neq 0$:
  Since previously we have shown $\|\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{x}\| = \|\boldsymbol{x}\|$, we obtain:

$$\|\tilde{\boldsymbol{x}}\| = \|\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{x}\| = \|\boldsymbol{x}\| \neq 0.$$

- (1) $\implies$ (3) : We only need to show $\det(\boldsymbol{A}_k) > 0$ for $k = 1, \dots, n$.
  Given any $\tilde{\boldsymbol{x}} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix} \in \mathbb{R}^k$, we construct $\boldsymbol{x} = \begin{pmatrix} \tilde{\boldsymbol{x}} \\ \boldsymbol{0} \end{pmatrix} \in \mathbb{R}^n$.
  Since $\boldsymbol{A} \succ 0$, we find

$$\begin{aligned}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} &= \begin{pmatrix}\tilde{\boldsymbol{x}}^{\mathrm{T}} & \boldsymbol{0}\end{pmatrix}\boldsymbol{A}\begin{pmatrix} \tilde{\boldsymbol{x}} \\ \boldsymbol{0} \end{pmatrix} \\ &= \tilde{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{A}_k\tilde{\boldsymbol{x}} > 0.\end{aligned}$$

  Since $\tilde{\boldsymbol{x}}$ is arbitrary vector in $\mathbb{R}^k$, we derive $\boldsymbol{A}_k \succ 0$.
  By (2) of this theorem, all eigenvalues of $\boldsymbol{A}_k$ are positive.
  Thus $\det(\boldsymbol{A}_k)$ = product of all eigenvalues of $\boldsymbol{A} > 0$.
- (3) $\implies$ (4) :
  - We want to show all pivots of $\boldsymbol{A}$ are positive first:
    We do row transform to convert $\boldsymbol{A}$ into upper triangular matrix $\tilde{\boldsymbol{A}}$:

$$\begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{bmatrix} \Longrightarrow \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \end{bmatrix}$$

    And during the row transformation, the determinant doesn't change. Moreover, the correponding *upper left submatrices* determinants don't change. In other words, we obtain

$$\det(\tilde{\boldsymbol{A}}_i) = \det(\boldsymbol{A}_i) \text{ for } i = 1, \dots, n.$$

And moreover, $\tilde{\boldsymbol{A}}_i$ always contains $\tilde{\boldsymbol{A}}_{i-1}$ on its upper left side:

$$\tilde{\boldsymbol{A}}_i = \begin{bmatrix} \tilde{\boldsymbol{A}}_{i-1} & \boldsymbol{B} \\ \boldsymbol{0} & \tilde{a}_{ii} \end{bmatrix}$$

And we notice $\tilde{\boldsymbol{A}}_i$'s are also upper triangular matrix. The determinant of a upper triangular matrix is the product of its diagonal entries. Hence we obtain

$$\det(\tilde{\boldsymbol{A}}_i) = \tilde{a_{ii}} \det(\tilde{\boldsymbol{A}}_{i-1}) \text{ for } i = 2,\dots,n.$$

Thus $\tilde{a_{ii}} = \frac{\det(\tilde{\boldsymbol{A}}_i)}{\det(\tilde{\boldsymbol{A}}_{i-1})} = \frac{\det(\boldsymbol{A}_i)}{\det(\boldsymbol{A}_{i-1})}$ for $i = 2,\dots,n$.
Due to (3) of this theorem, $\tilde{a_{ii}} > 0$ for $i = 2,\dots,n$. And $\tilde{a_{11}} = \det(\tilde{\boldsymbol{A}}_1) = \det(\boldsymbol{A}_1) > 0$.
In conclusion, all pivots $\tilde{a_{ii}} > 0$ for $i = 1,\dots,n$.

– Then we do the LDU composition for $\boldsymbol{A}$. Since $\boldsymbol{A}$ is symmetric, we obtain

$$\boldsymbol{A} = \boldsymbol{L}\boldsymbol{D}\boldsymbol{L}^{\mathrm{T}}$$

where $\boldsymbol{D} = \operatorname{diag}(d_1,\dots,d_n)$. The diagonal entries of $\boldsymbol{D}$ are pivots of $\boldsymbol{A}$. And $\boldsymbol{L}$ is lower triangular matrix with 1's on the diagonal entries.
Since all pivots of $\boldsymbol{A}$ are positive, we define $\sqrt{\boldsymbol{D}} := \operatorname{diag}(\sqrt{d_1},\dots,\sqrt{d_n})$.
Hence $\boldsymbol{A}$ could be written as:

$$\boldsymbol{A} = \boldsymbol{L}\begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix}\boldsymbol{L}^{\mathrm{T}} = \boldsymbol{L}\sqrt{\boldsymbol{D}}\sqrt{\boldsymbol{D}}\boldsymbol{L}^{\mathrm{T}} = (\sqrt{\boldsymbol{D}}\boldsymbol{L}^{\mathrm{T}})^{\mathrm{T}}(\sqrt{\boldsymbol{D}}\boldsymbol{L}^{\mathrm{T}}).$$

Fefine $\boldsymbol{R} = \sqrt{\boldsymbol{D}}\boldsymbol{L}^{\mathrm{T}}$. Since $\sqrt{\boldsymbol{D}}$ and $\boldsymbol{L}^{\mathrm{T}}$ are nonsingular, $\boldsymbol{D}$ is nonsingular.
Hence $\boldsymbol{A} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}$, where $\boldsymbol{R}$ is nonsingular matrix.

- $(4) \implies (1)$: Suppose $\boldsymbol{A} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}$, where $\boldsymbol{R}$ is nonsingular. Then for any $\boldsymbol{x} \in \mathbb{R}^n$, we have

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{x} = \|\boldsymbol{R}\boldsymbol{x}\|^2 \geq 0.$$

Then we only need to show that if $\boldsymbol{x} \neq \boldsymbol{0}$, then $\|\boldsymbol{R}\boldsymbol{x}\| \neq 0$.:
Since $\boldsymbol{R}$ is nonsinguar, when $\boldsymbol{x} \neq \boldsymbol{0}$, we obtain $\boldsymbol{R}\boldsymbol{x} \neq \boldsymbol{0}$. Hence $\|\boldsymbol{R}\boldsymbol{x}\| \neq 0$.

■

We may ask is there any quick ways to determine the positive definiteness of a matrix? The answer is yes. Let's introduce some definitions first.

Definition 17.3 — Submatrix. If $\boldsymbol{A}$ is a $n \times n$ matrix, then a submatrix of $\boldsymbol{A}$ is obtained by keeping some collection of rows and columns. ■

■ **Example 17.4** If $\boldsymbol{A} = \begin{bmatrix} 1 & -1 & & \\ -1 & 2 & -1 & \\ & -1 & 2 & -1 \\ & & -1 & 2 \end{bmatrix}$, then if we keep the (1,3,4)th row of $\boldsymbol{A}$ and (1,2)th column of $\boldsymbol{A}$, our submatrix is denoted as

$$\boldsymbol{A}_{(1,3,4),(1,2)} = \begin{bmatrix} 1 & -1 \\ 0 & -1 \\ 0 & 0 \end{bmatrix}$$

■

Definition 17.4 — principal submatrix. If A is a $n \times n$ matrix, then a principal submatrix of A is obtained by keeping the same collection of rows and columns. For example, if we want to keep the (5,7)th row of A, in order to construct a principal submatrix, we must keep the (5,7)th column of A as well.

■ **Example 17.5** If $A = \begin{bmatrix} 1 & -1 & & \\ -1 & 2 & -1 & \\ & -1 & 2 & -1 \\ & & -1 & 2 \end{bmatrix}$, then if we keep the (1,3,4)th row of A, in order to construct a principal submatrix, we have to keep (1,3,4)th column of A as well. Our principal submatrix is denoted as

$$A_{(1,3,4),(1,3,4)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

Definition 17.5 — leading principal submatrix. If A is a $n \times n$ matrix, then a leading principal submatrix of A is obtained by keeping the first k rows and columns of A, where $k \in \{1, 2, \ldots, n\}$.

Note that the leading principal submatrix is just the upper left submatrix we have mentioned before.

Corollary 17.1 If $A \succ 0$, then all principal submatrices of $A \succ 0$.

Proof. Our goal is to show $A_{\alpha,\alpha} \succ 0$, where $\alpha \in \{1, 2, \ldots, n\}$.
For any $x_\alpha \in \mathbb{R}^{|\alpha|}$, we only need to show $x_\alpha^{\mathrm{T}} A_{\alpha,\alpha} x_\alpha > 0$. Note that $|\alpha|$ denotes the number of elements in set α.
We construct $x \in \mathbb{R}^n$ s.t. the ith entry of x is

$$x_i = \begin{cases} (x_\alpha)_i & i \in \alpha \\ 0 & i \notin \alpha \end{cases}$$

It's obvious that

$$\begin{aligned} x^{\mathrm{T}} A x &= \sum_{i,j=1}^{n} x_i x_j A_{ij} \\ &= \sum_{i,j \in \alpha} (x_\alpha)_i (x_\alpha)_j (A_{\alpha,\alpha})_{ij} \\ &= x_\alpha^{\mathrm{T}} A_{\alpha,\alpha} x_\alpha > 0. \end{aligned}$$

■

How to use this corollary to test the positive definiteness?
For example, given $A = \begin{bmatrix} 2 & -1 & 1 \\ -1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}$, immediately we find one principal matrix is $A_{2,2} = 0$.
Hence it is not PD.
Also, there are many equivalent statements related to PSD.

Theorem 17.2 Let $\boldsymbol{A}$ be a $n \times n$ real symmetric matrix, the following are equivalent:

1. $\boldsymbol{A}$ is PSD.
2. All eigenvalues of $\boldsymbol{A}$ are nonnegative.
3. $\boldsymbol{A}$ could be factorized as $\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}$, where $\boldsymbol{R}$ is square.

R Is $\boldsymbol{A} \succeq 0$ equivalent to $\boldsymbol{A}_{ij} \geq 0$? No. Let's raise a counterexample:

$$\boldsymbol{A} = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix} \succeq 0.$$

PSD has many interesting properties. Before we introduce one, let's extend the definiton of inner product into matrix form:

Definition 17.6 — Frobenius inner product. For two real $n \times n$ matrix $\boldsymbol{A}$ and $\boldsymbol{B}$, the **Frobenius inner product** is given by

$$\langle \boldsymbol{A}, \boldsymbol{B} \rangle = \sum_{i,j=1}^{n} \boldsymbol{A}_{ij}\boldsymbol{B}_{ij}$$

■

Proposition 17.1 If two real $n \times n$ symmetric matrix $\boldsymbol{A} \succeq 0, \boldsymbol{B} \succeq 0$, then $\langle \boldsymbol{A}, \boldsymbol{B} \rangle \geq 0$.

Proof. Since $\boldsymbol{A} \succeq 0$, there exists square matrix $\boldsymbol{R} = \begin{bmatrix} \boldsymbol{r}_1 & \dots & \boldsymbol{r}_n \end{bmatrix}$ s.t.

$$\boldsymbol{A} = \boldsymbol{R}\boldsymbol{R}^{\mathrm{T}} = \sum_{k=1}^{n} \boldsymbol{r}_k \boldsymbol{r}_k^{\mathrm{T}}$$

Hence our inner product is given by

$$\begin{aligned}
\langle \boldsymbol{A}, \boldsymbol{B} \rangle &= \langle \sum_{k=1}^{n} \boldsymbol{r}_k \boldsymbol{r}_k^{\mathrm{T}}, \boldsymbol{B} \rangle \\
&= \sum_{k=1}^{n} \langle \boldsymbol{r}_k \boldsymbol{r}_k^{\mathrm{T}}, \boldsymbol{B} \rangle \\
&= \sum_{k=1}^{n} (\sum_{i,j=1}^{n} \boldsymbol{B}_{ij}\boldsymbol{R}_{ki}\boldsymbol{R}_{kj}) \\
&= \sum_{k=1}^{n} \boldsymbol{r}_k^{\mathrm{T}} \boldsymbol{B} \boldsymbol{r}_k
\end{aligned}$$

Since $\boldsymbol{B} \succeq 0$, we obtain $\langle \boldsymbol{A}, \boldsymbol{B} \rangle = \sum_{k=1}^{n} \boldsymbol{r}_k^{\mathrm{T}} \boldsymbol{B} \boldsymbol{r}_k \geq 0$. ■