Linear Algebra Review

Dr. Guangliang Chen

August 29, 2018

Outline

Notation: vectors

Vectors are denoted by boldface lowercase letters (such as $\mathbf{a}, \mathbf{b}$).

They are always assumed to be in column form.

To indicate their dimensions, we use notation like $\mathbf{a} \in \mathbb{R}^n$.

The ith element of $\mathbf{a}$ is written as a_i or $\mathbf{a}(i)$.

We denote the constant vector of one as $\mathbf{1}$ (with its dimension implied by the context).

Notation: matrices

Matrices are denoted by boldface uppercase letters (such as $\mathbf{A}, \mathbf{B}$).

Similarly, we write $\mathbf{A} \in \mathbb{R}^{m \times n}$ to indicate its size.

The (i, j) entry of $\mathbf{A}$ is denoted by a_{ij} or $\mathbf{A}(i, j)$.

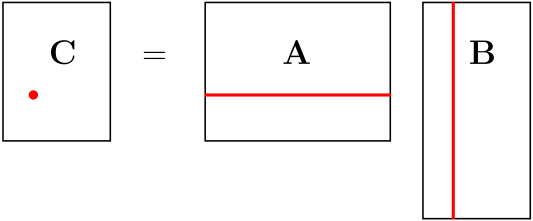
The ith row of $\mathbf{A}$ is denoted by $\mathbf{A}(i, :)$ while its columns are written as $\mathbf{A}(:, j)$, as in MATLAB.

We use $\mathbf{I}$ to denote the identity matrix (with its dimension implied by the context).

Matrix multiplication

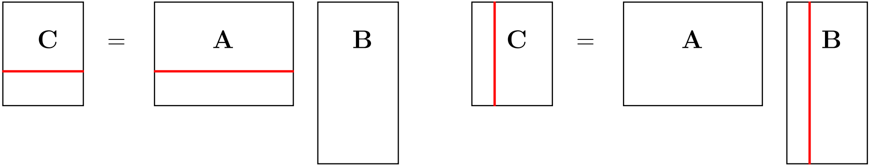
Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{B} \in \mathbb{R}^{n\times k}$. Their product is an $m \times k$ matrix

$$\mathbf{C} = (c_{ij}), \quad c_{ij} = \sum_{\ell=1}^{n} a_{i\ell} b_{\ell j} = \mathbf{A}(i,:) \cdot \mathbf{B}(:,j).$$



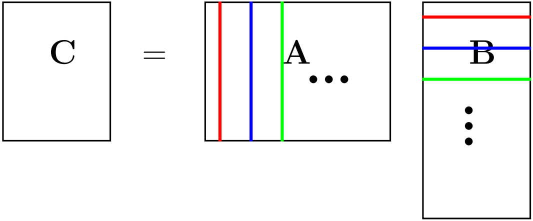
It is possible to obtain one full row (or column) of $\mathbf{C}$ at a time via matrix-vector multiplication:

$$\mathbf{C}(i,:) = \mathbf{A}(i,:) \cdot \mathbf{B}, \qquad \mathbf{C}(:,j) = \mathbf{A} \cdot \mathbf{B}(:,j)$$



The full matrix $\mathbf{C}$ can be written as a sum of rank-1 matrices:

$$\mathbf{C} = \sum_{\ell=1}^{n} \mathbf{A}(:,\ell) \cdot \mathbf{B}(\ell,:).$$



When one of the matrices is a diagonal matrix, we have the following rules:

$$\underbrace{\mathbf{A}}_{\text{diagonal}} \mathbf{B} = \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{pmatrix} \begin{pmatrix} \mathbf{B}(1,:) \\ \vdots \\ \mathbf{B}(n,:) \end{pmatrix} = \begin{pmatrix} a_1\mathbf{B}(1,:) \\ \vdots \\ a_n\mathbf{B}(n,:) \end{pmatrix}$$

$$\begin{aligned} \mathbf{A} \underbrace{\mathbf{B}}_{\text{diagonal}} &= [\mathbf{A}(:,1) \ldots \mathbf{A}(:,n)] \begin{pmatrix} b_1 & & \\ & \ddots & \\ & & b_n \end{pmatrix} \\ &= [b_1\mathbf{A}(:,1) \ldots b_n\mathbf{A}(:,n)] \end{aligned}$$

Finally, below are some identities involving the vector $\mathbf{1} \in \mathbb{R}^n$:

$$\begin{aligned}
\mathbf{1}\mathbf{1}^T &= \begin{pmatrix} 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix}, \\
\mathbf{1}^T\mathbf{1} &= n, \\
\mathbf{A}\mathbf{1} &= \sum_j \mathbf{A}(:,j), && \text{(vector of row sums)} \\
\mathbf{1}^T\mathbf{A} &= \sum_i \mathbf{A}(i,:), && \text{(horizontal vector of column sums)} \\
\mathbf{1}^T\mathbf{A}\mathbf{1} &= \sum_i \sum_j \mathbf{A}(i,j) && \text{(total sum of all entries)}
\end{aligned}$$

Example 0.1. Let

$$\mathbf{A} = \begin{pmatrix} 3 & 0 & 0 \\ 5 & 1 & -1 \\ -2 & 2 & 4 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \\ 2 & 3 \end{pmatrix}, \mathbf{\Lambda}_1 = \begin{pmatrix} 1 & & \\ & 0 & \\ & & -1 \end{pmatrix}, \mathbf{\Lambda}_2 = \begin{pmatrix} 2 & \\ & -3 \end{pmatrix}.$$

Find the products $\mathbf{AB}, \mathbf{\Lambda}_1\mathbf{B}, \mathbf{B\Lambda}_2, \mathbf{1}^T\mathbf{B}, \mathbf{B1}$ and verify the above rules.

The entrywise product

Another way to multiply two matrices of the same size, say $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times n}$, is through the Hadamard product, also called the entrywise product:

$$\mathbf{C} = \mathbf{A} \circ \mathbf{B} \in \mathbb{R}^{m\times n}, \qquad \text{with} \quad c_{ij} = a_{ij} b_{ij}.$$

For example,

$$\begin{pmatrix} 0 & 2 & -3 \\ -1 & 0 & -4 \end{pmatrix} \circ \begin{pmatrix} 1 & 0 & -3 \\ 2 & 1 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 9 \\ -2 & 0 & 4 \end{pmatrix}.$$

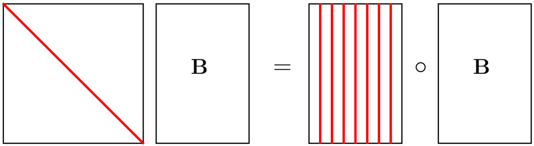
An important application of the entrywise product is in computing the product of a diagonal matrix and a rectangular matrix in software.

Let $\mathbf{A} = \operatorname{diag}(a_1, \ldots, a_n) \in \mathbb{R}^{n \times n}$ and $\mathbf{B} \in \mathbb{R}^{n \times k}$. Define also a vector $\mathbf{a} = (a_1, \ldots, a_n)^T \in \mathbb{R}^n$, which represents the diagonal of $\mathbf{A}$.

Then

$$\mathbf{AB} = \underbrace{[\mathbf{a} \ldots \mathbf{a}]}_{k \text{ copies}} \circ \mathbf{B}.$$

The former takes $\mathcal{O}(n^2 k)$ operations, while the latter takes only $\mathcal{O}(nk)$ operations, which is one magnitude faster.



Matrix rank

Let $\mathbf{A} \in \mathbb{R}^{m\times n}$. The maximal number of linearly independent rows (or columns) is called the rank of $\mathbf{A}$, and often denoted as $\mathrm{rank}(\mathbf{A})$.

It is known that $\mathrm{rank}(\mathbf{A}) \le \min(m, n)$.

A square matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is said to have full rank if $\mathrm{rank}(\mathbf{A}) = n$; otherwise, it is said to be rank deficient.

Matrix trace

The trace of a square matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is defined as the sum of the entries in its diagonal:

$$\mathrm{trace}(\mathbf{A}) = \sum_i a_{ii}.$$

If $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{B}$ is an $n \times m$ matrix, then

$$\mathrm{trace}(\mathbf{AB}) = \mathrm{trace}(\mathbf{BA}).$$

Matrix determinant

The matrix determinant is a rule[1] to evaluate square matrices to numbers:

$$\det : \mathbf{A} \in \mathbb{R}^{n\times n} \mapsto \det(\mathbf{A}) \in \mathbb{R}.$$

The matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is said to be invertible or nonsingular if $\det(A) \neq 0$, which can be shown to be equivalent to being of full rank (i.e., $\operatorname{rank}(\mathbf{A}) = n$).

An important property of matrix determinant is for two square matrices of the same size $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}$,

$$\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B}).$$

[1] `https://en.wikipedia.org/wiki/Determinant`

Example 0.2. For the matrix

$$\mathbf{A} = \begin{pmatrix} 3 & 0 & 0 \\ 5 & 1 & -1 \\ -2 & 2 & 4 \end{pmatrix},$$

find its rank, trace and determinant.

Eigenvalues and eigenvectors

Let $\mathbf{A} \in \mathbb{R}^{n\times n}$. The **characteristic polynomial** of $\mathbf{A}$ is

$$p(\lambda) = \det(\mathbf{A} - \lambda \mathbf{I}).$$

The roots of the characteristic equation $p(\lambda) = 0$ are called **eigenvalues** of $\mathbf{A}$.

For a specific eigenvalue λ_i, any nonzero vector $\mathbf{v}_i$ satisfying

$$(\mathbf{A} - \lambda_i \mathbf{I})\mathbf{v}_i = \mathbf{0}$$

or equivalently,

$$\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i$$

is called an **eigenvector** of $\mathbf{A}$ (associated to the eigenvalue λ_i).

All eigenvectors associated to λ_i span a linear subspace, called the **eigenspace**:

$$\mathrm{E}(\lambda_i) = \{\mathbf{v} \in \mathbb{R}^n : (\mathbf{A} - \lambda_i \mathbf{I})\mathbf{v} = \mathbf{0}\}.$$

The dimension g_i of $\mathrm{E}(\lambda_i)$ is called the **geometric multiplicity** of λ_i, while the degree a_i of the factor $(\lambda - \lambda_i)^{a_i}$ in $p(\lambda)$ is called the **algebraic multiplicity** of λ_i.

Note that we must have $\sum a_i = n$ and for all i, $1 \leq g_i \leq a_i$.

Example 0.3. For the matrix $\mathbf{A} = \begin{pmatrix} 3 & 0 & 0 \\ 5 & 1 & -1 \\ -2 & 2 & 4 \end{pmatrix}$, find its eigenvalues and their multiplicities, as well as associated eigenvectors.

Answer. The eigenvalues are $\lambda_1 = 3, \lambda_2 = 2$ with $a_1 = 2, a_2 = 1$ and $g_1 = g_2 = 1$. The corresponding eigenvectors are $\mathbf{v}_1 = (0, 1, -2)^T, \mathbf{v}_2 = (0, 1, -1)^T$.

The following theorem indicates that the trace and determinant of a square matrix can both be computed from the eigenvalues of the matrix.

Theorem 0.1. *Let* $\mathbf{A}$ *be a real square matrix whose eigenvalues are* $\lambda_1, \ldots, \lambda_n$ *(with repetitions). Then*

$$\det(\mathbf{A}) = \prod_{i=1}^{n} \lambda_i \qquad \textit{and} \quad \operatorname{trace}(\mathbf{A}) = \sum_{i=1}^{n} \lambda_i.$$

Example 0.4. For the matrix $\mathbf{A}$ defined previously, verify the identities in the above theorem.

Diagonalizability of square matrices

Definition 0.1. A square matrix $\mathbf{A}$ is **diagonalizable** if it is *similar* to a diagonal matrix, i.e., there exist an invertible matrix $\mathbf{P}$ and a diagonal matrix $\mathbf{\Lambda}$ such that

$$\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}, \qquad \text{or equivalently,} \quad \mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{\Lambda}.$$

Remark. If we write $\mathbf{P} = (\mathbf{p}_1, \ldots, \mathbf{p}_n)$ and $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, then the above equation can be rewritten as

$$\mathbf{A}\mathbf{P} = \mathbf{P}\mathbf{\Lambda}, \qquad \text{or in columns,} \quad \mathbf{A}\mathbf{p}_i = \lambda_i \mathbf{p}_i,\ 1 \le i \le n.$$

This shows that the λ_i are the eigenvalues of $\mathbf{A}$ and $\mathbf{p}_i$ the associated eigenvectors. Thus, the above factorization is called the **eigenvalue decomposition** of $\mathbf{A}$, or sometimes the **spectral decomposition** of $\mathbf{A}$.

Example 0.5. The matrix

$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 3 & 2 \end{pmatrix}$$

is diagonalizable because

$$\begin{pmatrix} 0 & 1 \\ 3 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 3 & -1 \end{pmatrix} \begin{pmatrix} 3 & \\ & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 3 & -1 \end{pmatrix}^{-1}$$

but $\mathbf{B} = \begin{pmatrix} 0 & 1 \\ -1 & 2 \end{pmatrix}$ is not (how can we know this?).

Checking diagonalizability of a square matrix

Theorem 0.2. *A matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is diagonalizable if and only if it has n linearly independent eigenvectors (i.e., $\sum g_i = n$).*

Corollary 0.3. *The following matrices are diagonalizable:*

- *Any matrix whose eigenvalues all have identical geometric and algebraic multiplicities, i.e., $g_i = a_i$ for all i;*
- *Any matrix with n distinct eigenvalues ($g_i = a_i = 1$ for all i);*

Example 0.6. The matrix $\mathbf{B} = \begin{pmatrix} 0 & 1 \\ -1 & 2 \end{pmatrix}$ is not diagonalizable because it has only one distinct eigenvalue $\lambda_1 = 1$ with $a_1 = 2$ and $g_1 = 1$.

Special square matrices

- **Symmetric matrices** $\mathbf{A} \in \mathbb{R}^{n\times n}$: $\mathbf{A}^T = \mathbf{A}$
- **Orthogonal matrices** $\mathbf{Q} \in \mathbb{R}^{n\times n}$: $\mathbf{Q}^{-1} = \mathbf{Q}^T$ (i.e. $\mathbf{Q}\mathbf{Q}^T = \mathbf{Q}^T\mathbf{Q} = \mathbf{I}$).

  Note that the columns of an orthogonal matrix $\mathbf{Q} = [\mathbf{q}_1 \ldots \mathbf{q}_n]$ are an orthonormal basis for $\mathbb{R}^n$:

$$\mathbf{q}_i^T\mathbf{q}_j = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}$$

  Geometrically, an orthogonal matrix multiplying a vector (i.e., $\mathbf{Q}\mathbf{x} \in \mathbb{R}^n$) represents an rotation of the vector in the space.

Spectral decomposition of symmetric matrices

Theorem 0.4. *Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Then there exist an orthogonal matrix $\mathbf{Q} = [\mathbf{q}_1 \ldots \mathbf{q}_n]$ and a diagonal matrix $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, such that*

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T \qquad (\textit{we say that } \mathbf{A} \textit{ is orthogonally diagonalizable in this case})$$

Note that the above equation is equivalent to

$$\mathbf{A}\mathbf{q}_i = \lambda_i \mathbf{q}_i, \quad i = 1, \ldots, n$$

Therefore, the λ_i's represent eigenvalues of $\mathbf{A}$ while the $\mathbf{q}_i$'s are the associated eigenvectors (with unit norm).

For convenience the diagonal elements of $\mathbf{\Lambda}$ are often sorted in decreasing order $\lambda_{\max} \equiv \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \equiv \lambda_{\min}$ (with same ordering of the eigenvectors).

Example 0.7. Find the spectral decomposition of the following matrix

$$\mathbf{A} = \begin{pmatrix} 0 & 2 \\ 2 & 3 \end{pmatrix}$$

Answer.

$$\mathbf{A} = \underbrace{\frac{1}{\sqrt{5}} \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix}}_{\mathbf{Q}} \cdot \underbrace{\begin{pmatrix} 4 & \\ & -1 \end{pmatrix}}_{\mathbf{\Lambda}} \cdot \underbrace{\frac{1}{\sqrt{5}} \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix}^T}_{\mathbf{Q}^T}$$

Rayleigh quotients

Theorem 0.5. *For any given symmetric matrix* $\mathbf{A} \in \mathbb{R}^{n\times n}$,

$$\max_{\mathbf{x}\in\mathbb{R}^n:\,\mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} = \lambda_{\max} \quad (\textit{when } \mathbf{x} = \textit{“largest” eigenvector of } \mathbf{A})$$

$$\min_{\mathbf{x}\in\mathbb{R}^n:\,\mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} = \lambda_{\min} \quad (\textit{when } \mathbf{x} = \textit{“smallest” eigenvector of } \mathbf{A})$$

Remark. The quantity $\frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$ is called a **Rayleigh quotient**.

Example 0.8. For the matrix $\mathbf{A}$ in the preceding example, the maximum of the Rayleigh quotient is 4, achieved when $\mathbf{x} = \frac{1}{\sqrt{5}}\begin{pmatrix}1\\2\end{pmatrix}$.

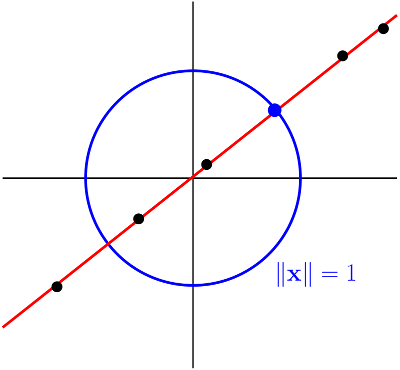
We prove the theorem on the preceding slide in two ways.

(1) **Linear algebra approach**:

$$\max_{\mathbf{x}\in\mathbb{R}^n:\, \|\mathbf{x}\|=1} \mathbf{x}^T \mathbf{A}\mathbf{x}$$

(2) **Multivariable calculus approach**:

$$\max_{\mathbf{x}\in\mathbb{R}^n} \mathbf{x}^T \mathbf{A}\mathbf{x} \quad \text{subject to } \|\mathbf{x}\|^2 = 1$$



Linear algebra approach

***Proof*.** Let $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$ be the spectral decomposition, where $\mathbf{Q} = [\mathbf{q}_1, \ldots, \mathbf{q}_n]$ is orthogonal and $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ is diagonal with sorted diagonals from large to small. Then for any unit vector $\mathbf{x}$,

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{x}^T(\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T)\mathbf{x} = (\mathbf{x}^T\mathbf{Q})\mathbf{\Lambda}(\mathbf{Q}^T\mathbf{x}) = \mathbf{y}^T\mathbf{\Lambda}\mathbf{y}$$

where $\mathbf{y} = \mathbf{Q}^T\mathbf{x}$ is also a unit vector:

$$\|\mathbf{y}\|^2 = \mathbf{y}^T\mathbf{y} = (\mathbf{Q}^T\mathbf{x})^T(\mathbf{Q}^T\mathbf{x}) = \mathbf{x}^T\mathbf{Q}\mathbf{Q}^T\mathbf{x} = \mathbf{x}^T\mathbf{x} = 1.$$

So the original optimization problem becomes the following one:

$$\max_{\mathbf{y}\in\mathbb{R}^n:\,\|\mathbf{y}\|=1} \mathbf{y}^T \underbrace{\mathbf{\Lambda}}_{\text{diagonal}} \mathbf{y}$$

To solve this new problem, write $\mathbf{y} = (y_1, \ldots, y_n)^T$. It follows that

$$\mathbf{y}^T \mathbf{\Lambda} \mathbf{y} = \sum_{i=1}^{n} \underbrace{\lambda_i}_{\text{fixed}} y_i^2 \qquad (\text{subject to } y_1^2 + y_2^2 + \cdots + y_n^2 = 1)$$

Because $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, when $y_1^2 = 1, y_2^2 = \cdots = y_n^2 = 0$ (i.e., $\mathbf{y} = \pm \mathbf{e}_1$), the objective function attains its maximum value $\mathbf{y}^T \mathbf{\Lambda} \mathbf{y} = \lambda_1$.

In terms of the original variable $\mathbf{x}$, the maximizer is

$$\mathbf{x}^* = \mathbf{Q}\mathbf{y}^* = \mathbf{Q}(\pm \mathbf{e}_1) = \pm \mathbf{q}_1.$$

In conclusion, when $\mathbf{x} = \pm \mathbf{q}_1$ (largest eigenvector), $\mathbf{x}^T \mathbf{A} \mathbf{x}$ attains its maximum value λ_1 (largest eigenvalue).

Multivariable calculus approach

Proof. First, we form the Lagrangian function

$$L(\mathbf{x}, \lambda) = \mathbf{x}^T \mathbf{A}\mathbf{x} - \lambda(\|\mathbf{x}\|^2 - 1).$$

Next, we need to find all of its critical points by solving

$$\frac{\partial L}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x} - \lambda(2\mathbf{x}) = 0 \quad \longrightarrow \quad \mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$
$$\frac{\partial L}{\partial \lambda} = \|\mathbf{x}\|^2 - 1 = 0 \quad \longrightarrow \quad \|\mathbf{x}\|^2 = 1$$

This implies that $\mathbf{x}, \lambda$ must be an eigenpair of $\mathbf{A}$. Among them, $\mathbf{v}_1$ (corresponding to largest eigenvalue λ_1 of $\mathbf{A}$) is the global optimal solution, and it yields the absolute maximum value

$$\mathbf{v}_1^T \mathbf{A}\mathbf{v}_1 = \mathbf{v}_1(\lambda_1 \mathbf{v}_1) = \lambda_1.$$

Positive (semi)definite matrices

Definition 0.2. A symmetric matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is said to be **positive semidefinite** if $\mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$.

If the equality holds true only for $\mathbf{x} = \mathbf{0}$ (i.e., $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$), then $\mathbf{A}$ is said to be **positive definite**.

Example 0.9. For any rectangular matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, show that both of the matrices $\mathbf{A}\mathbf{A}^T \in \mathbb{R}^{m\times m}$ and $\mathbf{A}^T\mathbf{A} \in \mathbb{R}^{n\times n}$ are positive semidefinite.

Theorem. A symmetric matrix $\mathbf{A}$ is positive definite (semidefinite) if and only if all of its eigenvalues are positive (nonnegative).