# A Primer in Econometric Theory

## Lecture 2: Linear Algebra and Matrices

John Stachurski
Lectures by Akshay Shanker

March 26, 2017

# Matrices

Typical $N \times K$ **matrix**:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1K} \\ a_{21} & a_{22} & \cdots & a_{2K} \\ \vdots & \vdots & & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NK} \end{pmatrix}$$

Symbol $a_{nk}$ stands for element in the $n$-th row of the $k$-th column

$N \times K$ matrix also called a

- **row vector** if $N = 1$
- **column vector** if $K = 1$

If $N = K$, then $\mathbf{A}$ is called **square**

If $\mathbf{A}$ is square and $a_{nk} = a_{kn}$ for every $k$ and $n$, then $\mathbf{A}$ is called **symmetric**

Often elements of a matrix $\mathbf{A}$ represent coefficients in a system of linear equations

$$\begin{gathered} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1K}x_K = b_1 \\ \vdots \\ a_{N1}x_1 + a_{N2}x_2 + \cdots + a_{NK}x_K = b_N \end{gathered}$$

For a matrix $\mathbf{A}$, the notation:

- $\mathrm{row}_n\,\mathbf{A}$ refers to the $n$th row of $\mathbf{A}$
- $\mathrm{col}_k\,\mathbf{A}$ to refers to the $k$th column of $\mathbf{A}$

The symbols $\mathbf{0}$ and $\mathbf{1}$ represent matrices with all elements equal to zero and one respectively

For square $\mathbf{A}$, elements $a_{nn}$ called the **principal diagonal**:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NN} \end{pmatrix}$$

**Identity matrix**:

$$\mathbf{I} := \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

Note $\mathrm{col}_n \mathbf{I} = \mathbf{e}_n$, the $n$th canonical basis vector in $\mathbb{R}^N$

# Algebraic Operations for Matrices

Operations for matrices:

- Scalar multiplication
- Addition
- Matrix multiplication

Scalar multiplication is element by element, as with vectors:

$$\gamma \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1K} \\ a_{21} & a_{22} & \cdots & a_{2K} \\ \vdots & \vdots & & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NK} \end{pmatrix} := \begin{pmatrix} \gamma a_{11} & \gamma a_{12} & \cdots & \gamma a_{1K} \\ \gamma a_{21} & \gamma a_{22} & \cdots & \gamma a_{2K} \\ \vdots & \vdots & & \vdots \\ \gamma a_{N1} & \gamma a_{N2} & \cdots & \gamma a_{NK} \end{pmatrix}$$

Addition also element by element:

$$\begin{pmatrix} a_{11} & \cdots & a_{1K} \\ a_{21} & \cdots & a_{2K} \\ \vdots & \vdots & \vdots \\ a_{N1} & \cdots & a_{NK} \end{pmatrix} + \begin{pmatrix} b_{11} & \cdots & b_{1K} \\ b_{21} & \cdots & b_{2K} \\ \vdots & \vdots & \vdots \\ b_{N1} & \cdots & b_{NK} \end{pmatrix}$$

$$:= \begin{pmatrix} a_{11}+b_{11} & \cdots & a_{1K}+b_{1K} \\ a_{21}+b_{21} & \cdots & a_{2K}+b_{2K} \\ \vdots & \vdots & \vdots \\ a_{N1}+b_{N1} & \cdots & a_{NK}+b_{NK} \end{pmatrix}$$

Note matrices must be same dimension

Multiplication of matrices:

Product $\mathbf{AB}$: $i, j$-th element is inner product of $i$-th row of $\mathbf{A}$ and $j$-th column of $\mathbf{B}$

$$c_{ij} = \langle \mathrm{row}_i \, \mathbf{A}, \mathrm{col}_j \, \mathbf{B} \rangle = \sum_{k=1}^{K} a_{ik} b_{kj}$$

Picture for $i = j = 1$:

$$\begin{pmatrix} a_{11} & \cdots & a_{1K} \\ a_{21} & \cdots & a_{2K} \\ \vdots & \vdots & \vdots \\ a_{N1} & \cdots & a_{NK} \end{pmatrix} \begin{pmatrix} b_{11} & \cdots & b_{1J} \\ b_{21} & \cdots & b_{2J} \\ \vdots & \vdots & \vdots \\ b_{K1} & \cdots & b_{KJ} \end{pmatrix} = \begin{pmatrix} c_{11} & \cdots & c_{1J} \\ c_{21} & \cdots & c_{2J} \\ \vdots & \vdots & \vdots \\ c_{N1} & \cdots & c_{NJ} \end{pmatrix}$$

In this picture,

$$c_{11} = \langle \mathrm{row}_1(\mathbf{A}), \mathrm{col}_1(\mathbf{B}) \rangle = \sum_{k=1}^{K} a_{1k} b_{k1}$$

Suppose $\mathbf{A}$ is $N \times K$ and $\mathbf{B}$ is $J \times M$

- $\mathbf{AB}$ defined only if $K = J$
- Resulting matrix $\mathbf{AB}$ is $N \times M$

The rule to remember:

product of $N \times K$ and $K \times M$ is $N \times M$

Multiplication is not commutative: $\mathbf{AB} \neq \mathbf{BA}$

Note $\mathbf{BA}$ is not well-defined unless $N = M$ also holds

**Fact.** (3.1.1) For conformable matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and scalar $\alpha$, we have

1. $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$,
2. $\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB}+\mathbf{AC}$,
3. $(\mathbf{A}+\mathbf{B})\mathbf{C} = \mathbf{AC}+\mathbf{BC}$,
4. $\mathbf{A}\alpha\mathbf{B} = \alpha\mathbf{AB}$, and
5. $\mathbf{AI} = \mathbf{A}$ and $\mathbf{IA} = \mathbf{A}$, where $\mathbf{I}$ is the identity matrix.

Here, "conformable" means operation is defined given matrix dimensions

The $k$**th power** of a square matrix $\mathbf{A}$ is defined as

$$\mathbf{A}^k := \underbrace{\mathbf{A} \cdots \mathbf{A}}_{k \text{ terms}}$$

If $\mathbf{B}$ is such that $\mathbf{B}^2 = \mathbf{A}$, then $\mathbf{B}$ is called the **square root** of $\mathbf{A}$ and written $\sqrt{\mathbf{A}}$

Given $N \times K$ matrix $\mathbf{A}$ and $K \times 1$ column vector $\mathbf{x}$, the product $\mathbf{Ax}$ is:

$$\mathbf{Ax} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1K} \\ a_{21} & a_{22} & \cdots & a_{2K} \\ \vdots & \vdots & & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NK} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_K \end{pmatrix}$$

$$= x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{N1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{N2} \end{pmatrix} + \cdots + x_K \begin{pmatrix} a_{1K} \\ a_{2K} \\ \vdots \\ a_{NK} \end{pmatrix}$$

$$= \sum_{k=1}^{K} x_k \operatorname{col}_k \mathbf{A}$$

# Matrices as Maps

We can think of an $N \times K$ matrix $\mathbf{A}$ as a map from $\mathbb{R}^K$ to $\mathbb{R}^N$:

$$\mathbf{x} \mapsto \mathbf{A}\mathbf{x}$$

Such a map is linear

How about examples of linear functions that don't involve matrices?

...actually, there are none!

The set of linear functions from $\mathbb{R}^K$ to $\mathbb{R}^N$ and the set of $N \times K$ matrices are in one-to-one correspondence:

**Theorem.** (3.1.1) Let $T$ be a function from $\mathbb{R}^K$ to $\mathbb{R}^N$. The following are equivalent:

1. $T$ is linear.
2. There exists an $N \times K$ matrix $\mathbf{A}$ such that $T\mathbf{x} = \mathbf{A}\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^K$.

**Proof.**(1. $\implies$ 2.)

Let $T \colon \mathbb{R}^K \to \mathbb{R}^N$ be linear

We aim to construct an $N \times K$ matrix $\mathbf{A}$ such that

$$T\mathbf{x} = \mathbf{A}\mathbf{x}, \qquad \forall\, \mathbf{x} \in \mathbb{R}^K$$

As usual, let $\mathbf{e}_k$ be the $k$-th canonical basis vector in $\mathbb{R}^K$

Define a matrix $\mathbf{A}$ by $\operatorname{col}_k(\mathbf{A}) = T\mathbf{e}_k$. Pick any $\mathbf{x} = (x_1, \ldots, x_K) \in \mathbb{R}^K$. By linearity we have,

$$T\mathbf{x} = T\left[\sum_{k=1}^K x_k \mathbf{e}_k\right] = \sum_{k=1}^K x_k T\mathbf{e}_k = \sum_{k=1}^K x_k \operatorname{col}_k(\mathbf{A}) = \mathbf{A}\mathbf{x}$$

**Proof.**(2 $\implies$ 1) Fix $N \times K$ matrix $\mathbf{A}$ and let $T$ be defined by

$$T \colon \mathbb{R}^K \to \mathbb{R}^N, \qquad T\mathbf{x} = \mathbf{A}\mathbf{x}$$

Pick any $\mathbf{x}$, $\mathbf{y}$ in $\mathbb{R}^K$, and any scalars $\alpha$ and $\beta$

The rules of matrix arithmetic tell us that

$$T(\alpha\mathbf{x} + \beta\mathbf{y}) := \mathbf{A}(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha\mathbf{A}\mathbf{x} + \beta\mathbf{A}\mathbf{y} =: \alpha T\mathbf{x} + \beta T\mathbf{y}$$

Consider solving a system of linear equations such as $\mathbf{A}\mathbf{x} = \mathbf{b}$

Existence: can we find an $\mathbf{x}$ that satisfies this equation, for any given $\mathbf{b}$?

- is the corresponding linear map $T\mathbf{x} = \mathbf{A}\mathbf{x}$ an onto function?
- equivalently, is $\operatorname{rng} T$ equal to all of $\mathbb{R}^N$?

# Column Space

The range of $T$ is all vectors of the form $T\mathbf{x} = \mathbf{A}\mathbf{x}$ where $\mathbf{x}$ varies over $\mathbb{R}^K$

For $\mathbf{x} \in \mathbb{R}^K$, we have

$$\mathbf{A}\mathbf{x} = \sum_{k=1}^{K} x_k \operatorname{col}_k \mathbf{A}$$

Thus, $\operatorname{rng} T$ equals the **column space** of $\mathbf{A}$ – the span of the columns of $\mathbf{A}$

$$\operatorname{colspace} \mathbf{A} := \operatorname{span}\{\operatorname{col}_1 \mathbf{A}, \ldots, \operatorname{col}_K \mathbf{A}\}$$

In summary,

$$\operatorname{colspace} \mathbf{A} = \operatorname{rng} T = \{\mathbf{A}\mathbf{x} : \mathbf{x} \in \mathbb{R}^K\}$$

# Rank

Equivalent questions

- How large is the range of the linear map $T\mathbf{x} = \mathbf{A}\mathbf{x}$?
- How large is the column space of $\mathbf{A}$?

The obvious measure of size for a linear subspace is its dimension

The dimension of $\operatorname{colspace} \mathbf{A}$ is known as the **rank** of $\mathbf{A}$

$$\operatorname{rank} \mathbf{A} := \dim \operatorname{colspace} \mathbf{A}$$

Because $\operatorname{colspace} \mathbf{A}$ is the span of $K$ vectors, we have

$$\operatorname{rank} \mathbf{A} = \dim \operatorname{colspace} \mathbf{A} \leq K$$

$\mathbf{A}$ has **full column rank** if

$$\operatorname{rank} \mathbf{A} = \text{ number of columns of } \mathbf{A}$$

**Fact.** (3.1.2) For any matrix $\mathbf{A}$, the following statements are equivalent:

1. $\mathbf{A}$ is of full column rank
2. The columns of $\mathbf{A}$ are linearly independent
3. If $\mathbf{A}\mathbf{x} = \mathbf{0}$, then $\mathbf{x} = \mathbf{0}$

# Square Matrices and Invertability

Consider the case where $\mathbf{A}$ is $N \times N$

We seek conditions on $\mathbf{A}$ under which, for every $\mathbf{b} \in \mathbb{R}^N$, there exists exactly one $\mathbf{x} \in \mathbb{R}^N$ such that $\mathbf{A}\mathbf{x} = \mathbf{b}$

Let $T$ be the linear map $T\mathbf{x} = \mathbf{A}\mathbf{x}$

- When does each point $\mathbf{b} \in \mathbb{R}^N$ have one and only one preimage under $T$?
- Equivalenty, when is $T$ a bijection?

Recall linear bijections are called nonsingular functions

**Fact.** (3.1.3) For $N \times N$ matrix $\mathbf{A}$, the following are equivalent:

1. The columns of $\mathbf{A}$ are linearly independent.
2. The columns of $\mathbf{A}$ form a basis of $\mathbb{R}^N$.
3. $\operatorname{rank} \mathbf{A} = N$.
4. $\operatorname{colspace} \mathbf{A} = \mathbb{R}^N$.
5. $\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{y} \implies \mathbf{x} = \mathbf{y}$.
6. $\mathbf{A}\mathbf{x} = \mathbf{0} \implies \mathbf{x} = \mathbf{0}$.
7. For each $\mathbf{b} \in \mathbb{R}^N$, the equation $\mathbf{A}\mathbf{x} = \mathbf{b}$ has a solution.
8. For each $\mathbf{b} \in \mathbb{R}^N$, the equation $\mathbf{A}\mathbf{x} = \mathbf{b}$ has a unique solution.

If any of the equivalent conditions in fact 3.1.3 are true we will call not just the map $T$ but also the underlying matrix $\mathbf{A}$ **nonsingular**

If any one—and hence all—of these conditions fail, then $\mathbf{A}$ is called **singular**

Any bijection has an inverse (see §15.2 in ET)

Any nonsingular map $T$ has a nonsingular inverse $T^{-1}$ ( fact 2.1.9 on page 26)

- where $T$ is generated by a matrix $\mathbf{A}$, the inverse $T^{-1}$ is also associated with a matrix, called the inverse of $\mathbf{A}$

**Theorem.** (3.1.2) For nonsingular $\mathbf{A}$ the following statements are true:

1. There exists a square matrix $\mathbf{B}$ such that $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. The matrix $\mathbf{B}$ is called the **inverse** of $\mathbf{A}$, and written as $\mathbf{A}^{-1}$.
2. For each $\mathbf{b} \in \mathbb{R}^N$, the unique solution to $\mathbf{Ax} = \mathbf{b}$ is given by

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$$

Example. Consider the $N$ good linear demand system

$$q_n = \sum_{k=1}^{N} a_{nk} p_k + b_n, \quad n = 1, \ldots, N$$

where $q_n$ and $p_n$ are quantity and price of the $n$th good

We want to compute the inverse demand function which gives prices in terms of quantities

Write the system in matrix form as $\mathbf{q} = \mathbf{A}\mathbf{p} + \mathbf{b}$

If the columns of $\mathbf{A}$ are linearly independent, then we can invert the system– a unique solution exists for each fixed $\mathbf{q}$ and $\mathbf{b}$:

$$\mathbf{p} = \mathbf{A}^{-1}(\mathbf{q} - \mathbf{b})$$

For $N \times N$ matrices $\mathbf{A}$ and $\mathbf{B}$, if

- $\mathbf{B}$ is a **left inverse**, that is, $\mathbf{BA} = \mathbf{I}$
- or $\mathbf{B}$ is a**right inverse**, that is, $\mathbf{AB} = \mathbf{I}$

Then $\mathbf{A}$ is invertible and $\mathbf{B}$ is the inverse of $\mathbf{A}$

**Fact.** (3.1.4) Let $\mathbf{A}$ and $\mathbf{B}$ be $N \times N$ square matrices. If $\mathbf{B}$ is either a left or a right inverse of $\mathbf{A}$, then $\mathbf{A}$ is nonsingular and $\mathbf{B}$ is its inverse.

**Fact.** (3.1.5) If $\mathbf{A}$ and $\mathbf{B}$ are nonsingular and $\alpha \neq 0$, then

1. $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$,
2. $(\alpha\mathbf{A})^{-1} = \alpha^{-1}\mathbf{A}^{-1}$, and
3. $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$

# Determinants

Determinant is a unique number for each square matrix $\mathbf{A}$

Let $S(N)$ be the set of all bijections from $\{1, \ldots, N\}$ to itself

For $\pi \in S(N)$, define the **signature** of $\pi$

$$\operatorname{sgn}(\pi) := \prod_{m<n} \frac{\pi(m) - \pi(n)}{m - n}$$

The **determinant** of $N \times N$ matrix $\mathbf{A}$ is

$$\det \mathbf{A} := \sum_{\pi \in S(N)} \operatorname{sgn}(\pi) \prod_{n=1}^{N} a_{\pi(n)n}$$

**Fact.** (3.1.6) If $\mathbf{I}$ is the $N \times N$ identity, $\mathbf{A}$ and $\mathbf{B}$ are $N \times N$ matrices and $\alpha \in \mathbb{R}$, then

1. $\det \mathbf{I} = 1$,
2. $\mathbf{A}$ is nonsingular if and only if $\det \mathbf{A} \neq 0$,
3. $\det(\mathbf{AB}) = \det(\mathbf{A}) \det(\mathbf{B})$,
4. $\det(\alpha \mathbf{A}) = \alpha^N \det(\mathbf{A})$, and
5. $\det(\mathbf{A}^{-1}) = (\det(\mathbf{A}))^{-1}$.

In the $2 \times 2$ case one can show that the determinant satisfies

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$$

**Proof.**(Fact 3.1.4)

Fix square matrix $\mathbf{A}$ and suppose a right inverse $\mathbf{B}$ exists:

$$\mathbf{AB} = \mathbf{I}$$

Then both $\mathbf{A}$ and $\mathbf{B}$ are nonsingular, since using the rules in fact 3.1.6:

$$\det(\mathbf{A})\det(\mathbf{B}) = 1$$

Both $\det \mathbf{A}$ and $\det \mathbf{B}$ are nonzero and hence both matrices are nonsingular. Next, $\mathbf{AB} = \mathbf{I}$, hence

$$\mathbf{BAB} = \mathbf{B}$$

Post-multiplying by $\mathbf{B}^{-1}$ gives $\mathbf{BA} = \mathbf{I}$

# Diagonal and Triangular Matrices

A square matrix is called **lower triangular** if every element strictly above the principle diagonal is zero

Example.

$$\mathbf{L} := \begin{pmatrix} 1 & 0 & 0 \\ 2 & 5 & 0 \\ 3 & 6 & 1 \end{pmatrix}$$

A square matrix is called **upper triangular** if every element strictly below the principle diagonal is zero

Example.

$$\mathbf{U} := \begin{pmatrix} 1 & 2 & 3 \\ 0 & 5 & 6 \\ 0 & 0 & 1 \end{pmatrix}$$

A square matrix is called **triangular** if it is either upper or lower triangular

**Fact.** (3.2.1) If $\mathbf{A} = (a_{mn})$ is triangular, then $\det \mathbf{A} = \prod_{n=1}^{N} a_{nn}$.

Associated linear equations also simple to solve

Example.

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & 5 & 0 \\ 3 & 6 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

becomes

$$\begin{gathered} x_1 = b_1 \\ 2x_1 + 5x_2 = b_2 \\ 3x_1 + 6x_2 + x_3 = b_3 \end{gathered}$$

Top equation involves only $x_1$, so can solve for it directly

Plug that value into second equation, solve out for $x_2$, etc.

Consider a square $N \times N$ matrix $\mathbf{A}$

The $N$ elements of the form $a_{nn}$ are called the **principal diagonal**

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NN} \end{pmatrix}$$

A square matrix $\mathbf{D}$ is called **diagonal** if all entries off the principal diagonal are zero

$$\mathbf{D} = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & d_N \end{pmatrix}$$

Often written as

$$\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_N)$$

Diagonal systems are very easy to solve

Example.

$$\begin{pmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

is equivalent to

$$\begin{aligned} d_1x_1 &= b_1 \\ d_2x_2 &= b_2 \\ d_3x_3 &= b_3 \end{aligned}$$

**Fact.** (3.2.2) If $\mathbf{C} = \operatorname{diag}(c_1, \ldots, c_N)$ and $\mathbf{D} = \operatorname{diag}(d_1, \ldots, d_N)$, then

1. $\mathbf{C} + \mathbf{D} = \operatorname{diag}(c_1 + d_1, \ldots, c_N + d_N)$
2. $\mathbf{C}\mathbf{D} = \operatorname{diag}(c_1 d_1, \ldots, c_N d_N)$
3. $\mathbf{D}^k = \operatorname{diag}(d_1^k, \ldots, d_N^k)$ for any $k \in \mathbb{N}$
4. $d_n \geq 0$ for all $n \implies \mathbf{D}^{1/2}$ exists and equals
$$\operatorname{diag}(\sqrt{d_1}, \ldots, \sqrt{d_N})$$
5. $d_n \neq 0$ for all $n \implies \mathbf{D}$ is nonsingular and
$$\mathbf{D}^{-1} = \operatorname{diag}(d_1^{-1}, \ldots, d_N^{-1})$$

Proofs: Check 1 and 2 directly. Parts 3-5 then follow from 1 and 2.

# Trace, Transpose, Symmetry

The **trace** of a square matrix is defined by

$$\operatorname{trace}\begin{pmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & & \vdots \\ a_{N1} & \cdots & a_{NN} \end{pmatrix} = \sum_{n=1}^{N} a_{nn}$$

**Fact.** (3.2.3) If $\mathbf{A}$ and $\mathbf{B}$ are square matrices and $\alpha, \beta \in \mathbb{R}$, then

$$\operatorname{trace}(\alpha\mathbf{A} + \beta\mathbf{B}) = \alpha \operatorname{trace}(\mathbf{A}) + \beta \operatorname{trace}(\mathbf{B})$$

If $\mathbf{A}$ is $N \times M$ and $\mathbf{B}$ is $M \times N$, then $\operatorname{trace}(\mathbf{AB}) = \operatorname{trace}(\mathbf{BA})$

The **transpose** of $N \times K$ matrix $\mathbf{A}$ is a $K \times N$ matrix $\mathbf{A}^\top$ defined by

$$\mathrm{col}_n(\mathbf{A}') = \mathrm{row}_n(\mathbf{A})$$

Example. If

$$\mathbf{A} := \begin{pmatrix} 10 & 40 \\ 20 & 50 \\ 30 & 60 \end{pmatrix} \quad \text{then} \quad \mathbf{A}' = \begin{pmatrix} 10 & 20 & 30 \\ 40 & 50 & 60 \end{pmatrix}$$

If

$$\mathbf{B} := \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix} \quad \text{then} \quad \mathbf{B}' := \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}$$

A square matrix $\mathbf{A}$ is called **symmetric** if $\mathbf{A}^\mathsf{T} = \mathbf{A}$

- equivalently, $a_{nk} = a_{kn}$ for every $k$ and $n$

The matricies $\mathbf{A}^\mathsf{T}\mathbf{A}$ and $\mathbf{A}\mathbf{A}^\mathsf{T}$ are always well-defined and symmetric

**Fact.** (3.2.4) For conformable matrices $\mathbf{A}$ and $\mathbf{B}$, transposition satisfies

1. $(\mathbf{A}^\intercal)^\intercal = \mathbf{A}$,
2. $(\mathbf{A}\mathbf{B})^\intercal = \mathbf{B}^\intercal\mathbf{A}^\intercal$,
3. $(\mathbf{A}+\mathbf{B})^\intercal = \mathbf{A}^\intercal + \mathbf{B}^\intercal$, and
4. $(c\mathbf{A})^\intercal = c\mathbf{A}^\intercal$ for any constant $c$.

**Fact.** (3.2.5) For each square matrix $\mathbf{A}$, we have

1. $\operatorname{trace}(\mathbf{A}) = \operatorname{trace}(\mathbf{A}^\top)$ and
2. $\det(\mathbf{A}^\top) = \det(\mathbf{A})$.
3. If $\mathbf{A}$ is nonsingular, then so is $\mathbf{A}^\top$, and its inverse is $(\mathbf{A}^\top)^{-1} = (\mathbf{A}^{-1})^\top$.

If $\mathbf{a}$ and $\mathbf{b}$ are $N \times 1$ vectors, then the matrix product $\mathbf{a}^\top \mathbf{b} = \mathbf{b}^\top \mathbf{a}$ is equal to $\sum_{n=1}^N a_n b_n$,

- same as the inner product $\langle \mathbf{a}, \mathbf{b} \rangle$

# Eigenvalues and Eigenvectors

Let $\mathbf{A}$ be $N \times N$

In general $\mathbf{A}$ maps $\mathbf{x}$ to some arbitrary new location $\mathbf{A}\mathbf{x}$

But sometimes $\mathbf{x}$ will only be scaled:

$$\mathbf{A}\mathbf{x} = \lambda \mathbf{x} \quad \text{for some scalar } \lambda \tag{1}$$

If (1) holds and $\mathbf{x}$ is nonzero, then

1. $\mathbf{x}$ is called an **eigenvector** of $\mathbf{A}$ and $\lambda$ is called an **eigenvalue**
2. $(\mathbf{x}, \lambda)$ is called an **eigenpair**

Clearly $(\mathbf{x}, \lambda)$ is an eigenpair of $\mathbf{A}$ $\implies$ $(\alpha \mathbf{x}, \lambda)$ is an eigenpair of $\mathbf{A}$ for any nonzero $\alpha$

Example. Let

$$\mathbf{A} := \begin{pmatrix} 1 & -1 \\ 3 & 5 \end{pmatrix}$$

Then

$$\lambda = 2 \quad \text{and} \quad \mathbf{x} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

form an eigenpair because $\mathbf{x} \neq \mathbf{0}$ and

$$\mathbf{A}\mathbf{x} = \begin{pmatrix} 1 & -1 \\ 3 & 5 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 2 \\ -2 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \lambda \mathbf{x}$$

Example. Consider the matrix

$$\mathbf{R} := \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

Induces counter-clockwise rotation on any point by $90^\circ$

Hence no point $\mathbf{x}$ is scaled

Hence there exists <u>no</u> pair $\lambda \in \mathbb{R}$ and $\mathbf{x} \neq \mathbf{0}$ such that

$$\mathbf{R}\mathbf{x} = \lambda\mathbf{x}$$

- In other words, no <u>real-valued</u> eigenpairs exist

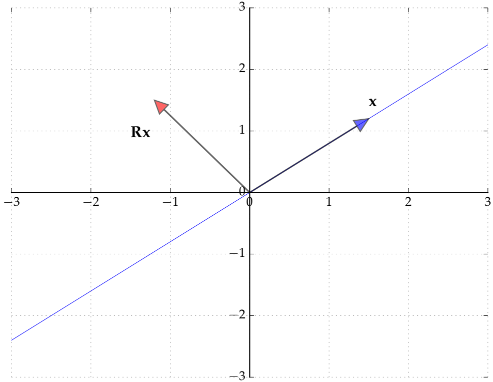


Figure: The matrix $\mathbf{R}$ rotates points by $90^\circ$

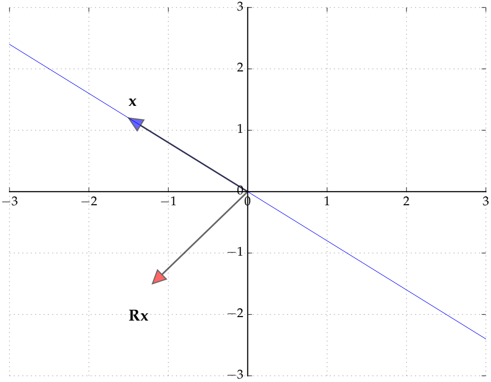


Figure: The matrix $\mathbf{R}$ rotates points by $90^{\circ}$

But $\mathbf{R}\mathbf{x} = \lambda\mathbf{x}$ can hold <u>if</u> we allow complex values

Example.

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ -i \end{pmatrix} = \begin{pmatrix} i \\ 1 \end{pmatrix} = i \begin{pmatrix} 1 \\ -i \end{pmatrix}$$

That is,

$$\mathbf{R}\mathbf{x} = \lambda\mathbf{x} \quad \text{for} \quad \lambda := i \quad \text{and} \quad \mathbf{x} := \begin{pmatrix} 1 \\ -i \end{pmatrix}$$

Hence $(\mathbf{x}, \lambda)$ is an eigenpair provided we admit complex values

**Fact.** (3.2.6) For any square matrix $\mathbf{A}$

$$\lambda \text{ is an eigenvalue of } \mathbf{A} \iff \det(\mathbf{A} - \lambda \mathbf{I}) = 0$$

**Proof.** Let $\mathbf{A}$ by $N \times N$ and let $\mathbf{I}$ be the $N \times N$ identity

We have

$$\begin{aligned} \det(\mathbf{A} - \lambda \mathbf{I}) = 0 &\iff \mathbf{A} - \lambda \mathbf{I} \text{ is singular} \\ &\iff \exists\, \mathbf{x} \neq \mathbf{0} \text{ s.t. } (\mathbf{A} - \lambda \mathbf{I})\mathbf{x} = \mathbf{0} \\ &\iff \exists\, \mathbf{x} \neq \mathbf{0} \text{ s.t. } \mathbf{A}\mathbf{x} = \lambda \mathbf{x} \\ &\iff \lambda \text{ is an eigenvalue of } \mathbf{A} \end{aligned}$$

Example. In the $2 \times 2$ case,

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \Longrightarrow \quad \mathbf{A} - \lambda \mathbf{I} = \begin{pmatrix} a - \lambda & b \\ c & d - \lambda \end{pmatrix}$$

$$\begin{aligned} \therefore \quad \det(\mathbf{A} - \lambda \mathbf{I}) &= (a - \lambda)(d - \lambda) - bc \\ &= \lambda^2 - (a + d)\lambda + (ad - bc) \end{aligned}$$

Hence the eigenvalues of $\mathbf{A}$ are given by the two roots of

$$\lambda^2 - (a + d)\lambda + (ad - bc) = 0$$

Equivalently,

$$\lambda^2 - \mathrm{trace}(\mathbf{A})\lambda + \det(\mathbf{A}) = 0$$

# Existence of Eigenvalues

Fix $N \times N$ matrix $\mathbf{A}$

**Fact.** There exist complex numbers $\lambda_1, \ldots, \lambda_N$ such that

$$\det(\mathbf{A} - \lambda \mathbf{I}) = \prod_{n=1}^{N} (\lambda_n - \lambda)$$

Each such $\lambda_i$ is an eigenvalue of $\mathbf{A}$ because

$$\det(\mathbf{A} - \lambda_i \mathbf{I}) = \prod_{n=1}^{N} (\lambda_n - \lambda_i) = 0$$

Important: not all eigenvalues are necessarily distinct — there can be repeats

**Fact.** (3.2.7) Given $N \times N$ matrix $\mathbf{A}$ with eigenvalues $\lambda_1, \ldots, \lambda_N$ we have

1. $\det(\mathbf{A}) = \prod_{n=1}^{N} \lambda_n$
2. $\operatorname{trace}(\mathbf{A}) = \sum_{n=1}^{N} \lambda_n$
3. If $\mathbf{A}$ is symmetric, then $\lambda_n \in \mathbb{R}$ for all $n$
4. If $\mathbf{A}$ is nonsingular, then eigenvalues of $\mathbf{A}^{-1} = 1/\lambda_1, \ldots, 1/\lambda_N$
5. If $\mathbf{A} = \operatorname{diag}(d_1, \ldots, d_N)$, then $\lambda_n = d_n$ for all $n$

Hence $\mathbf{A}$ is nonsingular $\iff$ all eigenvalues are nonzero

# Quadratic Forms

Fix $N \times N$ matrix $\mathbf{A}$

The **quadratic function** or **quadratic form** on $\mathbb{R}^N$ associated with $\mathbf{A}$ is the map $Q$ defined by

$$Q(\mathbf{x}) := \mathbf{x}^\top \mathbf{A} \mathbf{x} = \sum_{j=1}^{N} \sum_{i=1}^{N} a_{ij} x_i x_j$$

Example. Let $N = 2$ and let $\mathbf{A}$ be the identity matrix $\mathbf{I}$. In this case,

$$Q(\mathbf{x}) = \|\mathbf{x}\|^2 = x_1^2 + x_2^2$$

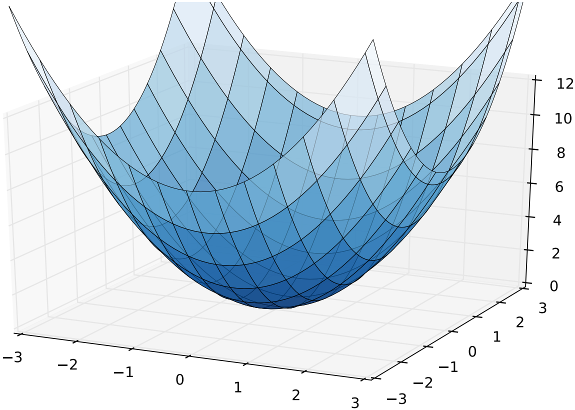


Figure: Quadratic function $Q(\mathbf{x}) = x_1^2 + x_2^2$

Notice:

- The graph for $Q(\mathbf{x}) = \|\mathbf{x}\|^2 = x_1^2 + x_2^2$ lies everywhere above zero

Matrix $\mathbf{A}$ with Quadratic form with the above property $Q(\mathbf{x}) \geq 0$ is called *positive definite*

More generally, an $N \times N$ symmetric matrix $\mathbf{A}$ is called

- **nonnegative definite** if $\mathbf{x}^\mathsf{T} \mathbf{A} \mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^N$,
- **positive definite** if $\mathbf{x}^\mathsf{T} \mathbf{A} \mathbf{x} > 0$ for all $\mathbf{x} \in \mathbb{R}^N$ with $\mathbf{x} \neq \mathbf{0}$,
- **nonpositive definite** if $\mathbf{x}^\mathsf{T} \mathbf{A} \mathbf{x} \leq 0$ for all $\mathbf{x} \in \mathbb{R}^N$, and
- **negative definite** if $\mathbf{x}^\mathsf{T} \mathbf{A} \mathbf{x} < 0$ for all $\mathbf{x} \in \mathbb{R}^N$ with $\mathbf{x} \neq \mathbf{0}$.

If $\mathbf{A}$ fits none of these categories, then $\mathbf{A}$ is called **indefinite**

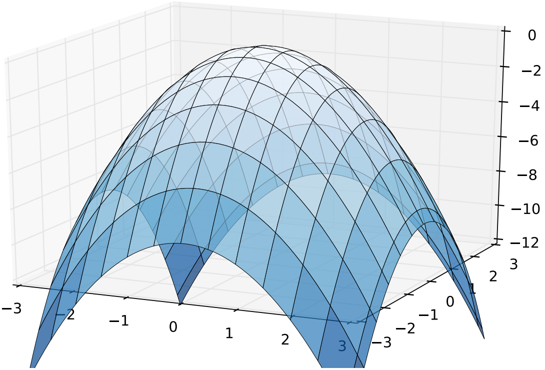


Figure: Quadratic function $Q(\mathbf{x}) = -x_1^2 - x_2^2$

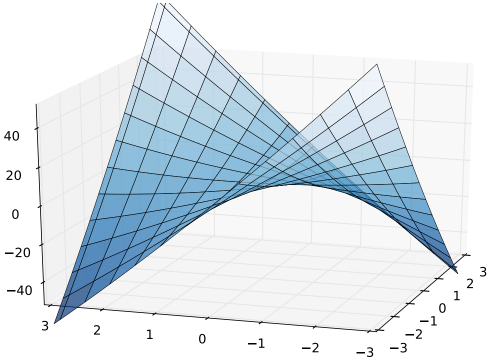


Figure: Quadratic function $Q(\mathbf{x}) = x_1^2/2 + 8x_1x_2 + x_2^2/2$

When the matrix $\mathbf{A}$ is diagonal:

$$\mathbf{A} = \mathrm{diag}(d_1, \ldots, d_N) \quad \text{implies} \quad Q(\mathbf{x}) = d_1 x_1^2 + \cdots + d_N x_N^2$$

A diagonal matrix is positive definite if and only if all diagonal elements are positive

**Fact.** (3.2.8) Let $\mathbf{A}$ be any symmetric matrix. $\mathbf{A}$ is

1. positive definite if and only if its eigenvalues are all positive,
2. negative definite if and only if its eigenvalues are all negative,

...and similarly for nonpositive and nonnegative definite

**Fact.** (3.2.9) If $\mathbf{A}$ is positive definite, then $\mathbf{A}$ is nonsingular, with $\det \mathbf{A} > 0$

A necessary (but not sufficient) condition for each kind of definiteness:

**Fact.** (3.2.10) If $\mathbf{A}$ is positive definite, then each element $a_{nn}$ on the principal diagonal is positive, and the same for nonnegative, nonpositive and negative.

# Projection Matrices

Recall given any subspace of $\mathbb{R}^N$, $S$, the corresponding projection $\mathbf{P} = \operatorname{proj} S$ is a linear map from $\mathbb{R}^N$ to $\mathbb{R}^N$

Recall Theorem 3.1.1: there exists an $N \times N$ matrix $\hat{\mathbf{P}}$ such that $\mathbf{P}\mathbf{x} = \hat{\mathbf{P}}\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^N$

- from now on $\mathbf{P}$ will also represent the corresponding matrix

What does this matrix look like?

**Theorem.** (3.3.1) Let $S$ be a subspace of $\mathbb{R}^N$. If $\mathbf{P} = \operatorname{proj} S$, then

$$\mathbf{P} = \mathbf{B}(\mathbf{B}^\intercal \mathbf{B})^{-1}\mathbf{B}^\intercal \tag{2}$$

for every matrix $\mathbf{B}$ such that the columns of $\mathbf{B}$ form a basis of $S$

See exercise 3.5.30 for proof

- The matrix $\mathbf{M} = \mathbf{I} - \mathbf{P}$ denotes the residual projection (see page 32)

Example. Recall example 2.2.1 on page 30

We found the projection of $\mathbf{y} \in \mathbb{R}^N$ onto $\text{span}\{\mathbf{1}\}$ is $\bar{y}\mathbf{1}$

Same result using Theorem (3.3.1):

- Since $\mathbf{1}$ is a basis for $\text{span}\{\mathbf{1}\}$:

$$\mathbf{P} = \text{proj span}\{\mathbf{1}\} \implies \mathbf{P} = \mathbf{1}(\mathbf{1}^\intercal\mathbf{1})^{-1}\mathbf{1}^\intercal = \frac{1}{N}\mathbf{1}\mathbf{1}^\intercal$$

- Thus, $\mathbf{P}\mathbf{y} = \bar{y}\mathbf{1}$, as expected
- Corresponding residual projection is

$$\mathbf{M}_c = \mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^\intercal$$

**Fact.** (3.3.1) In the setting of theorem 3.3.1, we have

1. $\mathbf{MB} = \mathbf{0}$
2. $\mathbf{PB} = \mathbf{B}$

Proof is an exercise (ex. 3.5.31 in ET)

Easy to see $\mathbf{M}_c$ in the previous example maps $\mathbf{1}$ to $\mathbf{0}$

A square matrix $\mathbf{A}$ is **idempotent** if $\mathbf{A}\mathbf{A} = \mathbf{A}$

**Fact.** (3.3.2) Both $\mathbf{P}$ and $\mathbf{M}$ are symmetric and idempotent

(Exercise: check by direct calculation)

Intuition: projecting onto a subspace twice is the same as projecting once – recall fact 2.2.8 on page 33

**Fact.** (3.3.4) Let $S$ be a linear subspace of $\mathbb{R}^N$. If $\mathbf{P} = \operatorname{proj} S$ and $\mathbf{M}$ is the residual projection, then

1. $\operatorname{rank} \mathbf{P} = \operatorname{trace} \mathbf{P} = \dim S$ and
2. $\operatorname{rank} \mathbf{M} = \operatorname{trace} \mathbf{M} = N - \dim S$

**Proof.**

- The rank of a linear map is the dimension of its range. When $\mathbf{P} = \operatorname{proj} S$, the range of $\mathbf{P}$ is exactly $S$
- To show that $\operatorname{trace} \mathbf{P} = \dim S$ also holds, use fact 3.3.3– $\operatorname{trace} \mathbf{P} = \dim S$,
- It follows that $\operatorname{trace} \mathbf{M} = N - \dim S$, because

$$\operatorname{trace} \mathbf{M} = \operatorname{trace}(\mathbf{I} - \mathbf{P}) = \operatorname{trace} \mathbf{I} - \operatorname{trace} \mathbf{P} = N - \dim S$$

# Overdetermined Systems of Equations

Consider systems of equations of the form $\mathbf{Ax} = \mathbf{b}$ when:

- The matrix $\mathbf{A}$ is $N \times K$ and has full column rank
- The vector $\mathbf{x}$ is $K \times 1$
- The vector $\mathbf{b}$ is $N \times 1$
- $K \leq N$

Taking $\mathbf{A}$ and $\mathbf{b}$ as given, we seek $\mathbf{x} \in \mathbb{R}^K$ such that $\mathbf{Ax} = \mathbf{b}$

If $K = N$, then system has precisely one solution

When $N > K$, the system of equations said to be **overdetermined**:

- number of equations $>$ number of unknowns
- number of constraints $>$ degrees of freedom

May not be able find a $\mathbf{b}$ that satisfies all $N$ equations

Recall the linear map $T \colon \mathbb{R}^K \to \mathbb{R}^N$ corresponding to $\mathbf{A}$ is $T\mathbf{x} = \mathbf{A}\mathbf{x}$

The following statements are equivalent:

1. there exists an $\mathbf{x} \in \mathbb{R}^K$ with $\mathbf{A}\mathbf{x} = \mathbf{b}$
2. the vector $\mathbf{b} \in \operatorname{colspace} \mathbf{A}$
3. the vector $\mathbf{b} \in \operatorname{rng} T$

Theorem 2.1.8 on page 26: when $K < N$, the function $T$ cannot be onto – possible $\mathbf{b}$ lies outside the range of $T$

When $K < N$, the scenario $\mathbf{b} \in \operatorname{colspace} \mathbf{A}$ is "very rare" because:

- the point $\mathbf{b}$ is an arbitrary point in $\mathbb{R}^N$
- the space $\operatorname{colspace} \mathbf{A}$ has dimension $K$
- $K$-dimensional subspaces of $\mathbb{R}^N$ have "Lebesgue measure zero" – the "chance" of $\mathbf{b}$ happening to lie in this subspace is tiny

Standard approach: admit an exact solution may not exist

Focus on finding $\mathbf{x} \in \mathbb{R}^K$ to make $\mathbf{A}\mathbf{x}$ as close to $\mathbf{b}$ as possible

- close in terms of ordinary Euclidean norm

The minimization problem, called the **least squares problem**

$$\hat{\mathbf{x}} := \operatorname*{argmin}_{\mathbf{x} \in \mathbb{R}^K} \|\mathbf{b} - \mathbf{A}\mathbf{x}\| \tag{3}$$

Assuming $\mathbf{A}$ is $N \times K$ with $K \leq N$ and $\mathbf{b}$ is $N \times 1$, we can use the orthogonal projection theorem to solve (3)

**Theorem.** (3.3.2) If $\mathbf{A}$ has full column rank, then (3) has the unique solution

$$\hat{\mathbf{x}} := (\mathbf{A}^\intercal \mathbf{A})^{-1} \mathbf{A}^\intercal \mathbf{b} \tag{4}$$

**Proof.**

Let:

- $\mathbf{A}$ and $\mathbf{b}$ be as in the statement of the theorem
- $\hat{\mathbf{x}}$ be as in (4) and
- $S := \operatorname{colspace} \mathbf{A}$

By full column rank assumption, the columns of $\mathbf{A}$ form a basis for $S$. Applying theorem 3.3.1, orthogonal projection of $\mathbf{b}$ onto $S$ is

$$\mathbf{P}\mathbf{b} := \mathbf{A}(\mathbf{A}^\intercal \mathbf{A})^{-1}\mathbf{A}^\intercal \mathbf{b} = \mathbf{A}\hat{\mathbf{x}} \tag{5}$$

Since the orthogonal projection theorem gives a unique minimizer in terms of the closest point in $S$ to $\mathbf{b}$,

$$\|\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}\| < \|\mathbf{b} - \mathbf{y}\| \quad \text{for all } \mathbf{y} \in S, \ \mathbf{y} \neq \mathbf{A}\hat{\mathbf{x}} \tag{6}$$

**Proof.**(cont.) Pick any $\mathbf{x} \in \mathbb{R}^K$ such that $\mathbf{x} \neq \hat{\mathbf{x}}$

We have $\mathbf{A}\mathbf{x} \in S$

In addition, since $\mathbf{x} \neq \hat{\mathbf{x}}$, and since $\mathbf{A}$ has full column rank, it must be that $\mathbf{A}\mathbf{x} \neq \mathbf{A}\hat{\mathbf{x}}$ (ex. 3.5.4)

Hence

$$\|\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}\| < \|\mathbf{b} - \mathbf{A}\mathbf{x}\| \quad \text{for all } \mathbf{x} \in \mathbb{R}^K,\ \mathbf{x} \neq \hat{\mathbf{x}}$$

In other words, $\hat{\mathbf{x}}$ is the unique solution to (3)

In (4), the matrix $(\mathbf{A}^\intercal\mathbf{A})^{-1}\mathbf{A}^\intercal$ is called the **pseudoinverse** of $\mathbf{A}$

If $K = N$, then the least squares solution $\hat{\mathbf{x}}$ in (4) reduces to:

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$$

What happens if the columns of $\mathbf{A}$ are not linearly independent?

- the set $\operatorname{colspace} \mathbf{A}$ is still a linear subspace and the orthogonal projection theorem still gives us a closest point $\mathbf{Pb}$ to $\mathbf{b}$ in $\operatorname{colspace} \mathbf{A}$
- since $\mathbf{Pb} \in \operatorname{colspace} \mathbf{A}$, there still exists a vector $\hat{\mathbf{x}}$ such that $\mathbf{Pb} = \mathbf{A}\hat{\mathbf{x}}$
- but there exists an infinity of such vectors

See Exercise 3.5.34

# QR Decomposition

The **QR decomposition** of a given matrix $\mathbf{A}$ is a product of the form $\mathbf{QR}$

- first matrix has orthonormal columns and
- the second is upper triangular

Applications include least squares problems and the computation of eigenvalues

**Theorem.** (3.3.3) If $\mathbf{A}$ is an $N \times K$ matrix with full column rank, then there exists a factorization $\mathbf{A} = \mathbf{Q}\mathbf{R}$ where

1. $\mathbf{R}$ is $K \times K$, upper triangular and nonsingular, and
2. $\mathbf{Q}$ is $N \times K$, with orthonormal columns

See page 64 in ET for a proof

Given the decomposition $\mathbf{A} = \mathbf{QR}$, the least squares solution $\hat{\mathbf{x}}$ defined in (4) can also be written as:

$$\hat{\mathbf{x}} = \mathbf{R}^{-1}\mathbf{Q}^{\intercal}\mathbf{b}$$

See Ex. 3.5.32

Premultiplying by $\mathbf{R}$:

$$\mathbf{R}\hat{\mathbf{x}} = \mathbf{Q}^{\intercal}\mathbf{b}$$

# Diagonalisation and Spectral Theory

If $f\colon A \to A$ and $g\colon B \to B$, then $g$ is said to be **topologically conjugate** to $f$ whenever there exists a continuous bijection $\tau\colon B \to A$ such that

$$f = \tau \circ g \circ \tau^{-1}$$

Can be beneficial if $g$ is somehow simpler than $f$

A square matrix $\mathbf{A}$ is said to be **similar** to another matrix $\mathbf{B}$ if there exists an invertible matrix $\mathbf{P}$ such that $\mathbf{A} = \mathbf{P}\mathbf{B}\mathbf{P}^{-1}$

$$
\begin{tikzcd}
\bullet \arrow[r, "\mathbf{B}"] & \bullet \arrow[d, "\mathbf{P}"] \\
\underset{\mathbf{x}}{\bullet} \arrow[u, "\mathbf{P}^{-1}"] \arrow[r, "\mathbf{A}"] & \underset{\mathbf{A}\mathbf{x}}{\bullet}
\end{tikzcd}
$$

Figure: $\mathbf{A}$ is similar to $\mathbf{B}$

If $\mathbf{A}$ is similar to a diagonal matrix, then $\mathbf{A}$ is called **diagonalizable**

We are interested in similarity to simple matrices, and diagonal matrices are the simplest kind

**Fact.** (3.3.5) If $\mathbf{A}$ is similar to $\mathbf{B}$, then $\mathbf{A}^t$ is similar to $\mathbf{B}^t$ for all $t \in \mathbb{N}$

Example. We want to calculate $\mathbf{A}^t$ for some given $t \in \mathbb{N}$

If $\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}$ for some $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_N)$, then by fact 3.3.5 and fact 3.2.2, we have

$$\mathbf{A}^t = \mathbf{P} \operatorname{diag}(\lambda_1^t, \ldots, \lambda_N^t) \mathbf{P}^{-1}$$

# Diagonalization and Eigenvalues

**Fact.** (3.3.6) If $\mathbf{A} = \mathbf{P} \mathbf{\Lambda} \mathbf{P}^{-1}$ for some $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$, then $(\mathrm{col}_n \, \mathbf{P}, \lambda_n)$ is an eigenpair of $\mathbf{A}$ for each $n$

**Proof.** Observe $\mathbf{A} = \mathbf{P} \mathbf{\Lambda} \mathbf{P}^{-1}$ implies $\mathbf{A}\mathbf{P} = \mathbf{P}\mathbf{\Lambda}$

Equating the $n$th column on each side gives

$$\mathbf{A}\mathbf{p}_n = \lambda_n \mathbf{p}_n$$

Where $\mathbf{p}_n := \mathrm{col}_n \, \mathbf{P}$

Note $\mathbf{p}_n$ is not the zero vector because $\mathbf{P}$ is invertible

But when is $\mathbf{A}$ diagonalizable?

**Fact.** (3.3.7) An $N \times N$ matrix $\mathbf{A}$ is diagonalizable if and only if it has $N$ linearly independent eigenvectors

In some cases, we can get an even simpler matrix decomposition if the matrix $\mathbf{P}$ has orthogonal columns

These kinds of matrices are called **orthogonal matrices**

**Fact.** (3.3.8) If $\mathbf{Q}$ and $\mathbf{P}$ are $N \times N$ orthogonal matrices, then

1. $\mathbf{Q}^\intercal$ is orthogonal and $\mathbf{Q}^{-1} = \mathbf{Q}^\intercal$,
2. $\mathbf{QP}$ is orthogonal, and
3. $\det \mathbf{Q} \in \{-1, 1\}$.

If $\mathbf{A} = \mathbf{Q}\Lambda\mathbf{Q}^{-1}$ and $\mathbf{Q}$ has orthonormal columns, then

$$\mathbf{A} = \mathbf{Q}\Lambda\mathbf{Q}^{\intercal}$$

Clearly, $\mathbf{A}$ must be symmetric. Next theorem shows symmetry of $\mathbf{A}$ is also sufficient

**Theorem.** (3.3.4) If $\mathbf{A}$ is symmetric, then $\mathbf{A}$ can be diagonalized as $\mathbf{A} = \mathbf{Q}\Lambda\mathbf{Q}^{\intercal}$, where $\mathbf{Q}$ is an orthogonal matrix and $\boldsymbol{\Lambda}$ is the diagonal matrix formed from the eigenvalues of $\mathbf{A}$

Above theorem was a version of the **spectral decomposition theorem**

$\mathbf{Q}\Lambda\mathbf{Q}^\intercal$ is called the **symmetric eigenvalue decomposition** of $\mathbf{A}$ – action of $\mathbf{A}$ on an $N \times 1$ vector $\mathbf{x}$:

$$\mathbf{A}\mathbf{x} = \sum_{n=1}^{N} \lambda_n (\mathbf{u}_n^\intercal \mathbf{x}) \mathbf{u}_n$$

where $\lambda_n$ is the $n$th eigenvalue of $\mathbf{A}$ and $\mathbf{u}_n = \mathrm{col}_n\, \mathbf{Q}$

Compare with $\mathbf{x} = \sum_{n=1}^{N} (\mathbf{u}_n^\intercal \mathbf{x}) \mathbf{u}_n$

**Fact.** (3.3.9) If $\mathbf{A}$ is nonnegative definite, then $\sqrt{\mathbf{A}}$ exists and equals $\mathbf{Q}\sqrt{\mathbf{\Lambda}}\mathbf{Q}^\intercal$. The matrix $\sqrt{\mathbf{\Lambda}}$ is given by $\mathrm{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_N})$

**Fact.** (3.3.10) If $\mathbf{A}$ is positive definite, then there exists a nonsingular, upper triangular matrix $\mathbf{R}$ such that $\mathbf{A} = \mathbf{R}^\intercal \mathbf{R}$

This decomposition is called the **Cholesky decomposition**

**Proof.**(Cholesky decomposition) We can write:

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\intercal = \mathbf{Q}\sqrt{\mathbf{\Lambda}}\sqrt{\mathbf{\Lambda}}\mathbf{Q}^\intercal = (\sqrt{\mathbf{\Lambda}}\mathbf{Q}^\intercal)^\intercal\sqrt{\mathbf{\Lambda}}\mathbf{Q}^\intercal$$

Then apply the QR decomposition to $\sqrt{\mathbf{\Lambda}}\mathbf{Q}^\intercal$:

$$\sqrt{\mathbf{\Lambda}}\mathbf{Q}^\intercal = \tilde{\mathbf{Q}}\mathbf{R}$$

where $\mathbf{R}$ is nonsingular and upper triangular, and $\tilde{\mathbf{Q}}$ has orthonormal columns

Because the columns of $\tilde{\mathbf{Q}}$ are orthonormal,

$$\mathbf{A} = (\tilde{\mathbf{Q}}\mathbf{R})^\intercal\tilde{\mathbf{Q}}\mathbf{R} = \mathbf{R}^\intercal\tilde{\mathbf{Q}}^\intercal\tilde{\mathbf{Q}}\mathbf{R} = \mathbf{R}^\intercal\mathbf{R}$$

# Norms and Continuity

Given vector sequence $\{\mathbf{x}_n\}$ in $\mathbb{R}^K$ and any point $\mathbf{x} \in \mathbb{R}^K$, we say that $\{\mathbf{x}_n\}$ **converges** to $\mathbf{x}$ and write $\mathbf{x}_n \to \mathbf{x}$ if, for any $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that $\|\mathbf{x}_n - \mathbf{x}\| < \epsilon$ whenever $n \geq N$

Equivalently, the real-valued sequence $z_n := \|\mathbf{x}_n - \mathbf{x}\|$ converges to zero in $\mathbb{R}$ as $n \to \infty$

**Fact.** (3.3.11) The following results hold:

1. If $\mathbf{x}_n \to \mathbf{x}$ and $\mathbf{y}_n \to \mathbf{y}$, then $\mathbf{x}_n + \mathbf{y}_n \to \mathbf{x} + \mathbf{y}$.
2. If $\mathbf{x}_n \to \mathbf{x}$ and $\alpha \in \mathbb{R}$, then $\alpha \mathbf{x}_n \to \alpha \mathbf{x}$.
3. $\mathbf{x}_n \to \mathbf{x}$ if and only if $\mathbf{a}^\top \mathbf{x}_n \to \mathbf{a}^\top \mathbf{x}$ for all $\mathbf{a} \in \mathbb{R}^K$.

We want to extend notion of convergence to matrices

The **matrix norm** of $N \times K$ matrix $\mathbf{A}$:

$$\|\mathbf{A}\| := \max \left\{ \|\mathbf{A}\mathbf{x}\| \, : \, \mathbf{x} \in \mathbb{R}^K, \; \|\mathbf{x}\| = 1 \right\} \tag{7}$$

The value of the matrix norm is not easy to solve for in general

However, the matrix norm behaves like the vector norm

**Fact.** (3.3.12) For any conformable matrices $\mathbf{A}$ and $\mathbf{B}$, the matrix norm satisfies

1. $\|\mathbf{A}\| \geq 0$ and $\|\mathbf{A}\| = 0$ if and only if all entries of $\mathbf{A}$ are zero,
2. $\|\alpha\mathbf{A}\| = |\alpha|\|\mathbf{A}\|$ for any scalar $\alpha$,
3. $\|\mathbf{A}+\mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$, and
4. $\|\mathbf{A}\mathbf{B}\| \leq \|\mathbf{A}\|\|\mathbf{B}\|$.

**Fact.** (3.3.13) For any $J \times K$ matrix $\mathbf{A}$ with elements $a_{jk}$, we have

$$\|\mathbf{A}\| \leq \sqrt{JK} \max_{jk} |a_{jk}|$$

If every element of $\mathbf{A}$ is close to zero then $\|\mathbf{A}\|$ is also close to zero

## Neumann Series

Later on, we study dynamic systems of the form

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{b}$$

Does there exist a "stationary" vector $\mathbf{x} \in \mathbb{R}^N$, in the sense that $\mathbf{x}_t = \mathbf{x}$ implies $\mathbf{x}_{t+1} = \mathbf{x}$?

We seek an $\mathbf{x} \in \mathbb{R}^N$ that solves the system of equations

$$\mathbf{x} = \mathbf{A}\mathbf{x} + \mathbf{b} \qquad (\mathbf{A} \text{ is } N \times N \text{ and } \mathbf{b} \text{ is } N \times 1) \tag{8}$$

Consider the scalar case $x = ax + b$

If $|a| < 1$, then there is a unique solution

$$\bar{x} = \frac{b}{1-a} = b \sum_{k=0}^{\infty} a^k$$

The Neumann series Lemma helps generalise to $\mathbb{R}^N$

**Theorem.** (3.3.5) If $\mathbf{A}$ is square and $\|\mathbf{A}^j\| < 1$ for some $j \in \mathbb{N}$, then $\mathbf{I} - \mathbf{A}$ is invertible, and

$$(\mathbf{I} - \mathbf{A})^{-1} = \sum_{i=0}^{\infty} \mathbf{A}^i$$

When the condition of the Neumann series lemma holds, (8) has the unique solution

$$\bar{\mathbf{x}} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{b} = \sum_{i=0}^{\infty} \mathbf{A}^i \mathbf{b}$$

To test the condition, we use the **spectral radius** of $\mathbf{A}$:

$$\varrho(\mathbf{A}) := \max\{|\lambda| : \lambda \text{ is an eigenvalue of } \mathbf{A}\}$$

$|\lambda|$ is the **modulus** of the possibly complex number $\lambda$

**Fact.** If $\varrho(\mathbf{A}) < 1$, then $\|\mathbf{A}^j\| < 1$ for some $j \in \mathbb{N}$

Why is $\varrho(\mathbf{A}) < 1$ is sufficient?

We need $\sum_{i=0}^{t} \mathbf{A}^i(\mathbf{I} - \mathbf{A})$ to close be $\mathbf{I}$ for large $t$

We have:

$$\sum_{i=0}^{t} \mathbf{A}^i(\mathbf{I} - \mathbf{A}) = \sum_{i=0}^{t} \mathbf{A}^i - \sum_{i=0}^{t} \mathbf{A}^{i+1} = \mathbf{I} - \mathbf{A}^{t+1}$$

Simplify to the case where $\mathbf{A}$ is diagonalizable:

$$\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}$$

where $\mathbf{\Lambda}$ is a diagonal matrix containing the eigenvalues $\lambda_1, \ldots, \lambda_N$ of $\mathbf{A}$ on its principal diagonal

Now, use fact 3.3.5,

$$\mathbf{A}^t = \mathbf{P} \begin{pmatrix} \lambda_1^t & 0 & \cdots & 0 \\ 0 & \lambda_2^t & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_N^t \end{pmatrix} \mathbf{P}^{-1}$$

If $\varrho(\mathbf{A}) < 1$, then $|\lambda_n| < 1$ for all $n$, and hence $\lambda_n^t \to 0$ as $t \to \infty$. It follows that $\mathbf{A}^t \to \mathbf{0}$