Calculus and Linear Algebra

Kaiwen Zhou

Contents

1 Calculus

Theorem 1.1. *Taylor Expansion*

$$f(x) = T_n(x) + R_n(x) = \sum_{i=0}^{n} \frac{f^{(i)}(a)}{i!}(x-a)^i + \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1}, \quad \textit{for some} \quad \xi \in [a,x]$$

where we must have $\lim_{n\to\infty} R_n(x) = 0$.

We use $T_n(x) = \sum_{i=0}^{n} \frac{f^{(i)}(a)}{i!}(x-a)^i$ *as the approximate for* $f(x)$, *and* $R_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1}$ *as the error of approximation.*

The bound for the remainder $R_n(x)$ *is* $|R_n(x)| \le \frac{M}{(n+1)!}(x-a)^{n+1}$, $M = \max_{\xi\in[a,x]} |f^{(n+1)}(\xi)|$.

Lemma 1.2. ***(Useful Lemma for Differentiating Integrals)*** *Let* $p(x) = \int_x^C f(u)du$ *and* $q(x) = \int_C^x f(u)du$, *where* C *a constant, it is easy to verify the following.*

$$\frac{d}{dx}p(x) = -f(x), \quad \frac{d}{dx}q(x) = f(x)$$

1.1 Leibniz integral rule

Theorem 1.3. ***(Leibniz integral rule)*** *In calculus, the Leibniz integral rule for differentiation under the integral sign states that for an integral of the form*

$$\int_{a(x)}^{b(x)} f(x,t)dt$$

where $-\infty < a(x), b(x) < \infty$ *and the integrands are functions dependent on* x, *the derivative of this integral is expressible as*

$$\frac{d}{dx}\left(\int_{a(x)}^{b(x)} f(x,t)dt\right) = f(x,b(x))\cdot\frac{d}{dx}b(x) - f(x,a(x))\cdot\frac{d}{dx}a(x) + \int_{a(x)}^{b(x)} \frac{\partial}{\partial x}f(x,t)dt$$

where the partial derivative $\frac{\partial}{\partial x}$ *indicates that inside the integral, only the variation of* $f(x,t)$ *with* x *is considered in taking the derivative.*

Solution: This comes straightaway from Leibniz rule and Chain rule. Let

$$g(x,a(x),b(x)) = \int_{a(x)}^{b(x)} f(t,x)dt$$

Using the Chain rule of integration

$$\begin{aligned}\frac{d}{dx}g(x,a(x),b(x)) &= \frac{\partial g}{\partial x}\frac{dx}{dx} + \frac{\partial g}{\partial a}\frac{da}{dx} + \frac{\partial g}{\partial b}g\frac{db}{dx} \\ \textit{Lemma 1.2} \Longrightarrow &= \int_{a(x)}^{b(x)} \frac{\partial}{\partial x}f(t,x)dt - f(a(x),x)\frac{d}{dx}a(x) + f(b(x),x)\frac{d}{dx}b(x)\end{aligned}$$

Remark 1.4.

1. When applying the Lemma 1.2 in the above proof, we treat $b(x)$ and $a(x)$ as constants when doing $\frac{\partial g}{\partial a}$ and $\frac{\partial g}{\partial b}$ respectively.
2. In the special case where the functions $a(x)$ and $b(x)$ are constants $a(x) = a$ and $b(x) = b$ with values that do not depend on x, this simplifies to:

$$\frac{d}{dx}\left(\int_a^b f(x,t)dt\right) = \int_a^b \frac{\partial}{\partial x}f(x,t)dt$$

3. If $a(x) = a$ is constant and $b(x) = x$, which is another common situation (for example, in the proof of Cauchy's repeated integration formula), the Leibniz integral rule becomes:

$$\frac{d}{dx}\left(\int_a^x f(x,t)dt\right) = f(x,x) + \int_a^x \frac{\partial}{\partial x}f(x,t)dt$$

△

Problem 1.5. We have the integral equation

$$\int_0^x \left(1 + x + e^{x-t}\right) y(t)\mathrm{d}t = g(x), \quad 0 \le x \le 1$$

By using Leibniz integral rule, we have that

$$g'(x) = (2+x)y(x) + \int_0^x \left(1 + e^{x-t}\right) y(t)\mathrm{d}t$$

Where y and g are chosen to satisfy the condition of Leibniz integral rule.

Solution: In the above example, $a(x) = 0, b(x) = x, f(t,x) = \left(1 + x + e^x e^{-t}\right) y(t)$, we will have the above equation simplifies to,

$$\begin{aligned}\frac{d}{dx}g\left(x,a(x),b(x)\right) &= \int_{a(x)}^{b(x)} \frac{\partial}{\partial x}f(t,x)dt - f\left(a(x),x\right)\frac{d}{dx}a(x) + f\left(b(x),x\right)\frac{d}{dx}b(x) \\ &= \int_0^x \frac{\partial}{\partial x}\left(1 + x + e^x e^{-t}\right) y(t)dt - \left(1 + x + e^x e^0\right) y(0)\frac{d}{dx}0 \\ &\quad + \left(1 + x + e^x e^{-x}\right) y(x)\frac{d}{dx}x \\ &= \int_0^x \left(1 + e^x e^{-t}\right) y(t)dt + (1 + x + 1)\, y(x) \\ &= (2+x)\, y(x) + \int_0^x \left(1 + e^{x-t}\right) y(t)dt.\end{aligned}$$

1.2 Gradient

1.3 Jacobian Matrix

Suppose $\mathbf{f}:\mathbf{R}^n \to \mathbf{R}^m$ is a function such that each of its first-order partial derivatives exist on $\mathbf{R}^n$. This function takes a point $\mathbf{x} \in \mathbf{R}^n$ as input and produces the vector $\mathbf{f}(\mathbf{x}) \in \mathbf{R}^m$ as output. Then the Jacobian matrix of $\mathbf{f}$ is defined to be an $m \times n$ matrix, denoted by $\mathbf{J}$, whose (i,j) th entry is $\mathbf{J}_{ij} = \frac{\partial f_i}{\partial x_j}$, or explicitly

$$\mathbf{J} = \begin{bmatrix} \frac{\partial \mathbf{f}}{\partial x_1} & \cdots & \frac{\partial \mathbf{f}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \nabla^{\mathrm{T}} f_1 \\ \vdots \\ \nabla^{\mathrm{T}} f_m \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

where $\nabla^{\mathrm{T}} f_i$ is the transpose (row vector) of the gradient of the i component. The Jacobian matrix, whose entries are functions of $\mathbf{x}$, is denoted in various ways; common notations include $D\mathbf{f}$, $\mathbf{J_f}$, $\nabla \mathbf{f}$, and $\frac{\partial(f_1,\ldots,f_m)}{\partial(x_1,\ldots,x_n)}$. Some authors define the Jacobian as the transpose of the form given above.

Example 1 Consider the function $\mathbf{f}:\mathbf{R}^2 \to \mathbf{R}^2$, with $(x,y) \mapsto (f_1(x,y), f_2(x,y))$, given by

$$\mathbf{f}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} f_1(x,y) \\ f_2(x,y) \end{bmatrix} = \begin{bmatrix} x^2 y \\ 5x + \sin y \end{bmatrix}.$$

Then we have

$$f_1(x,y) = x^2 y$$

and

$$f_2(x,y) = 5x + \sin y$$

and the Jacobian matrix of $\mathbf{f}$ is

$$\mathbf{J_f}(x,y) = \begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{bmatrix} = \begin{bmatrix} 2xy & x^2 \\ 5 & \cos y \end{bmatrix}$$

and the Jacobian determinant is

$$det\left(\mathbf{J_f}(x,y)\right) = 2xy\cos y - 5x^2$$

Example 2: polar-Cartesian transformation [edit] The transformation from polar coordinates (r,φ) to Cartesian coordinates (x,y), is given by the function $\mathbf{F}:\mathbf{R}^+ \times [0,2\pi) \to \mathbf{R}^2$ with components:

$$\begin{aligned} x &= r\cos\varphi \\ y &= r\sin\varphi \\ \mathbf{J_F}(r,\varphi) &= \begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \varphi} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \varphi} \end{bmatrix} = \begin{bmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{bmatrix} \end{aligned}$$

The Jacobian determinant is equal to r. This can be used to transform integrals between the two coordinate systems:

$$\iint_{\mathbf{F}(A)} f(x,y)dxdy = \iint_A f(r\cos\varphi, r\sin\varphi) r dr d\varphi.$$

Example 3: spherical-Cartesian transformation [edit] The transformation from spherical coordinates (r,φ,θ)[6] to Cartesian coordinates (x,y,z), is given by the function $\mathbf{F}:\mathbf{R}^+ \times [0,\pi) \times [0,2\pi) \to \mathbf{R}^3$ with components:

$$\begin{aligned} x &= r\sin\varphi\cos\theta; \\ y &= r\sin\varphi\sin\theta; \\ z &= r\cos\varphi. \end{aligned}$$

The Jacobian matrix for this coordinate change is

$$\mathbf{J_F}(r,\varphi,\theta) = \begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \varphi} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \varphi} & \frac{\partial y}{\partial \theta} \\ \frac{\partial z}{\partial r} & \frac{\partial z}{\partial \varphi} & \frac{\partial z}{\partial \theta} \end{bmatrix} = \begin{bmatrix} \sin\varphi\cos\theta & r\cos\varphi\cos\theta & -r\sin\varphi\sin\theta \\ \sin\varphi\sin\theta & r\cos\varphi\sin\theta & r\sin\varphi\cos\theta \\ \cos\varphi & -r\sin\varphi & 0 \end{bmatrix}$$

The determinant is $r^2\sin\varphi$. Since $dV = dxdydz$ is the volume for a rectangular differential volume element (because the volume of a rectangular prism is the product of its sides), we can interpret $dV = r^2\sin\varphi dr d\varphi d\theta$ as the volume of the spherical differential volume element. Unlike rectangular differential volume element's volume, this differential volume element's volume is not a constant, and varies with coordinates (r and φ). It can be used to transform integrals between the two coordinate systems:

$$\iiint_{\mathbf{F}(U)} f(x,y,z)dxdydz = \iiint_U f(r\sin\varphi\cos\theta, r\sin\varphi\sin\theta, r\cos\varphi) r^2 \sin\varphi dr d\varphi d\theta$$

Example 4: The Jacobian matrix of the function $\mathbf{F}:\mathbf{R}^3 \to \mathbf{R}^4$ with components

$$\begin{aligned} y_1 &= x_1 \\ y_2 &= 5x_3 \\ y_3 &= 4x_2^2 - 2x_3 \\ y_4 &= x_3\sin x_1 \end{aligned}$$

is

$$\mathbf{J_F}(x_1,x_2,x_3) = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \frac{\partial y_1}{\partial x_3} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \frac{\partial y_2}{\partial x_3} \\ \frac{\partial y_3}{\partial x_1} & \frac{\partial y_3}{\partial x_2} & \frac{\partial y_3}{\partial x_3} \\ \frac{\partial y_4}{\partial x_1} & \frac{\partial y_4}{\partial x_2} & \frac{\partial y_4}{\partial x_3} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 5 \\ 0 & 8x_2 & -2 \\ x_3\cos x_1 & 0 & \sin x_1 \end{bmatrix}.$$

This example shows that the Jacobian matrix need not be a square matrix.

Example 5: The Jacobian determinant of the function $\mathbf{F}:\mathbf{R}^3 \to \mathbf{R}^3$ with components

$$\begin{aligned} y_1 &= 5x_2 \\ y_2 &= 4x_1^2 - 2\sin(x_2x_3) \\ y_3 &= x_2x_3 \end{aligned}$$

is

$$\begin{vmatrix} 0 & 5 & 0 \\ 8x_1 & -2x_3\cos(x_2x_3) & -2x_2\cos(x_2x_3) \\ 0 & x_3 & x_2 \end{vmatrix} = -8x_1 \begin{vmatrix} 5 & 0 \\ x_3 & x_2 \end{vmatrix} = -40x_1x_2.$$

From this we see that $\mathbf{F}$ reverses orientation near those points where x_1 and x_2 have the same sign; the function is locally invertible everywhere except near points where $x_1 = 0$ or $x_2 = 0$. Intuitively, if one starts with a tiny object around the point $(1, 2, 3)$ and apply $\mathbf{F}$ to that object, one will get a resulting object with approximately $40 \times 1 \times 2 = 80$ times the volume of the original one, with orientation reversed.

1.4 Hessian Matrix

1.5 Taylor Expansion

2 Linear Algebra

3 Basics

Proposition 3.1. *1. $\alpha\vec{v}+\beta\vec{u}$ - linear combination.*

2. $\alpha\vec{a}$ - line; $\alpha\vec{a}+\beta b-$ plane; $\alpha\vec{a}+\beta\vec{b}+8\vec{c}-$ space.

Proposition 3.2. $\vec{a}=(a_1,a_2,a_3), \vec{b}=(b_1,b_2,b_3)$.

1. $\vec{a}\cdot\vec{b}=a_1b_1+a_2b_2+a_3b_3=\vec{b}\cdot\vec{a}=\vec{a}^\top b=\vec{b}^\top\vec{a}$
2. $\|\vec{a}\|=\sqrt{a_1^2+a_2^2+a_3^2}$
3. $\frac{\vec{a}\cdot\vec{b}}{\|a\|\|b\|}=\cos\theta$
4. $|\vec{a}\cdot\vec{b}|\le\|\vec{a}\|\|\vec{b}\|$ *(Cauchy-Schwartz)*
5. *if* $\vec{a}\cdot\vec{b}=0\Rightarrow\vec{a},\vec{b}$ *perpendicular* $\Rightarrow\|\vec{a}\|^2+\|\vec{b}\|^2=\|\vec{a}-\vec{b}\|^2$.
6. *unit vector:* $\|\vec{a}\|=1=\frac{\vec{b}}{\|\vec{b}\|}$
7. $\vec{a}^\top\vec{b}$ *- inner product;* $\vec{a}\vec{b}^\top$ *- outer product*
8. $e_i^\top A=Row_i(A)$.
9. *lower triangular matrix · lower triangular matrix = lower triangular matrix.*

Proposition 3.3.

1. *Linearly independence:* $\alpha\vec{a}+\beta\vec{b}+\gamma\vec{c}=\vec{0}$ *only if* $\alpha=\beta=\gamma=0$
2. *linearly dependence:* $\exists(\alpha,\beta,\gamma)\neq(0,0,0)$ *sit.* $\alpha\vec{a}+\beta\vec{b}+\gamma\vec{c}=\vec{0}$.
3. $\begin{bmatrix} | & | & | \\ \vec{a} & \vec{b} & \vec{c} \\ | & | & | \end{bmatrix}\begin{bmatrix}\alpha\\ \beta\\ \gamma\end{bmatrix}=\vec{0}\Longrightarrow\begin{cases}\textit{linearly independence}\Longrightarrow A-invertible.\\ \textit{linearly dependence } A \textit{ - singular (not invertible)}\end{cases}$

Matix Multiplication

- $AB=A\begin{bmatrix}\vec{b_1} & \vec{b_2} & \vec{b_3}\end{bmatrix}=\begin{bmatrix}\vec{Ab_1} & \vec{Ab_2} & \vec{Ab_3}\end{bmatrix}$.
- $Row_i(AB)=A\cdot Row_i(B)$.
- $Row_i(AB)=Row_i(A)\cdot B$.

4 Vector norms

A norm $\|\cdot\|:\mathbb{R}^n\to\mathbb{R}$ is a function satisfying the properties:

- $\|x\|=0$ if and only if $x=0$ (definiteness)
- $\|cx\|=|c|\|x\|$ for all $c\in\mathbb{R}$ (homogeneity)
- $\|x+y\|\le\|x\|+\|y\|$ (triangle inequality)

Common examples of norms:

- $\|x\|_1=|x_1|+\cdots+|x_n|$ (the 1-norm)
- $\|x\|_2=\sqrt{x_1^2+\cdots+x_n^2}$ (the 2-norm / Euclidean norm) -> default norm
- $\|x\|_\infty=\max_{1\le i\le n}|x_i|$ (max-norm)

5 Linear Equations, Elimination, Permutation

Linear Equations Given $A\in\mathbb{R}^{m\times n}$ and $b\in\mathbb{R}^m$, linear equations take the form

$$Ax=b$$

where we must solve for $x\in\mathbb{R}^n$.

Understandings of $Ax=b$

- $A\vec{x}$: matrix A acts on $\vec{x}$.
- General Thinking: The column picture of $A\vec{x}=\vec{b}$.
  A combination of n columns of A produces the vector $\vec{b}$.
- Geometric Thinking: The row picture of $A\vec{x}=\vec{b}$ coefficient matrix
  Matrix A can be view as a coefficient matrix, then m equations from m rows give m planes ($P_i:\text{row}_i\cdot\vec{x}-b_i=0$) meeting at $\vec{x}$.

Four possibilities for solutions

- $r=m=n\Longrightarrow R=\begin{bmatrix}I\end{bmatrix}$, $Ax=b$ has exactly one solution.
- $r=m, r<n\Longrightarrow R=\begin{bmatrix}I & F\end{bmatrix}$, $Ax=b$ has ∞ solution.
- $r<m, r=n\Longrightarrow R=\begin{bmatrix}I\\ \mathbf{0}\end{bmatrix}$, $Ax=b$ has 0 or 1 solution.
- $r<m, r<n\Longrightarrow R=\begin{bmatrix}I & F\\ \mathbf{0} & \mathbf{0}\end{bmatrix}$, $Ax=b$ has 0 or ∞ solution.

Idea of Elimination

The core idea is to convert A to an upper Triangular matrix A' (Elimination), then solve for $\vec{x}$ from x_n to x_1 (Back substitution). E.g.

$$\underbrace{\begin{bmatrix} * & * & * & * \\ & * & * & * \\ & & * & * \\ & & & * \end{bmatrix}}_{A}\underbrace{\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix}}_{x}=\underbrace{\begin{bmatrix}b_1\\b_2\\b_3\\b_4\end{bmatrix}}_{b}$$

Elimination matrix $E_{ij}=\begin{bmatrix}1 & & & \\ & 1 & & \\ -l & & \ddots & \\ & & & 1\end{bmatrix}$ where $e_{ij}=-l$ is use to reduce the (i,j) entry of A, a_{ij}, to zero.

$$E_{ij}A=\begin{bmatrix}\vdots\\ Row_i-l\cdot Row_1\\ \vdots\end{bmatrix},\quad E_{ij}^{-1}=\begin{bmatrix}1 & & & \\ & 1 & & \\ l & & \ddots & \\ & & & 1\end{bmatrix}$$

Permutation Matrix

$$P_{23}=\begin{bmatrix}1 & 0 & 0\\ 0 & 0 & 1\\ 0 & 1 & 0\end{bmatrix}=\begin{bmatrix}-\vec{e_1}^\top-\\ -\vec{e_3}^\top-\\ -\vec{e_2}^\top-\end{bmatrix}$$

$P_{23}A$ exchanges Row 2 & Row 3 of matrix A.

- Permutation matrix is orthogonal matrix, i.e., $P^\top=P^{-1}, PP^\top=I$.
- Sometimes we need to exchange some rows of A so it can be reduced to a valid rref R. In this case, we need permutation matrix P.
  E.g. $\begin{bmatrix}0 & 2\\ 3 & -2\end{bmatrix}$ can be fixed though has 0 as the first pivot $\Rightarrow\begin{bmatrix}3 & -2\\ 0 & 2\end{bmatrix}\to$ a row exchange produces an upper triangular matrix.
- For the ease of Elimination & Permutation for both sides of $A\vec{x}=\vec{b}$ we create an augmented matrix $L=\begin{bmatrix}A & \vec{b}\end{bmatrix}$ and let elimination and permutation matrices act on L.
  $$(E_{ij}P_{ij}\cdots)\cdot\begin{bmatrix}A & \vec{b}\end{bmatrix}$$

Pivots

- Pivots are on the diagonal of the triangle after elimination. We need n pivots to solve for n unknowns.

6 CR Decomposition

7 LU Decomposition

8 Inverse & Transpose

Proposition 8.1. ***(Basic properties of the matrix inverse and transpose):***

A^{-1} *is unique if it exists.*	$(A^{-1})^{-1}=A$	$(A^{-1})^\top=(A^\top)^{-1}$
$(AB)^{-1}=B^{-1}A^{-1}$	$(A^\top)^\top=A$	$(AB)^\top=B^\top A^\top$

left-inverse = right inverse = two-sided inverse

Suppose $BA=I$ *&* $AC=I\Rightarrow B=B(AC)=(BA)C=C$.

Gauss - Jordan Elimination for computing A^{-1} *By using* $A^{-1}[A\mid I]=\begin{bmatrix}I\mid A^{-1}\end{bmatrix}$ *where* $[A\mid I]$ *is the augmented matrix.* $\Rightarrow$ *convert* $[A\mid I]\to[I\mid A]$ *using elimination matrix.*

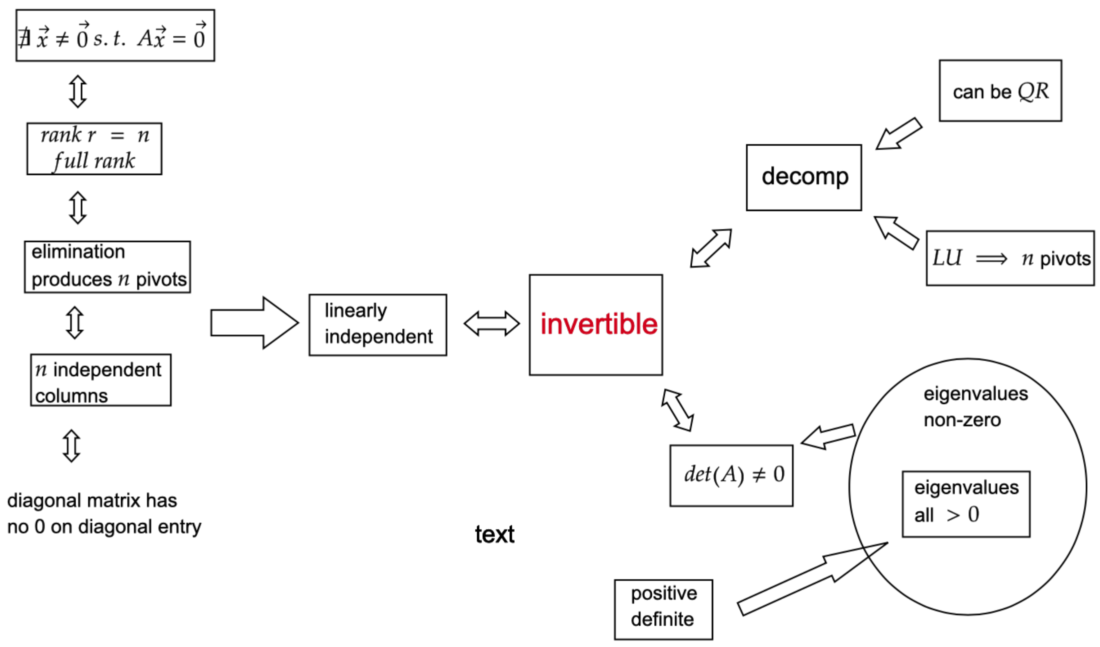

Figure 1: inverse all in one

9 The rank of a matrix

Note: When considering rank, think about the rref (the row reduced echelon form) of a matrix.

$\mathrm{rank}(A) =$ maximum number of linearly independent columns $=$ maximum number of linearly independent rows $=$ number of pivots

If $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times p}$ then

- $\mathrm{rank}(A) \leq \min(m, n)$
- $\mathrm{rank}(AB) \leq \min(\mathrm{rank}(A), \mathrm{rank}(B)) \leq \min(m, n, p)$
- if $\mathrm{rank}(A) = n$ then $\mathrm{rank}(AB) = \mathrm{rank}(B)$
- if $\mathrm{rank}(B) = n$ then $\mathrm{rank}(AB) = \mathrm{rank}(A)$

So multiplying by an invertible matrix does not alter the rank.

General properties of the matrix rank:

- $\mathrm{rank}(A + B) \leq \mathrm{rank}(A) + \mathrm{rank}(B)$
- $\mathrm{rank}(A) = \mathrm{rank}\left(A^\top\right) = \mathrm{rank}\left(A^\top A\right) = \mathrm{rank}\left(AA^\top\right)$
- $A \in \mathbb{R}^{n\times n}$ is invertible if and only if $\mathrm{rank}(A) = n$.
- rank-1 matrix $A = cuv^\top$ where $u \in \mathbb{R}^m$, $v \in \mathbb{R}^n$ and $A \in \mathbb{R}^{m\times n}$.
- rank-2 matrix $A = au_1v_1^\top + bu_2v_2^\top$ where $u_i \in \mathbb{R}^m$, $v_i \in \mathbb{R}^n$ and $A \in \mathbb{R}^{m\times n}$.
- rank-k matrix $A = \sum_{i=1}^k a_iu_iv_i^\top$ where $u_i \in \mathbb{R}^m$, $v_i \in \mathbb{R}^n$ and $A \in \mathbb{R}^{m\times n}$.

10 Four Spaces

Vector Space:

Subspace: A is a subspace, $\subseteq$, of S if $\forall u, v \in S, a, b$-constant, we have $au + bv \in S$.

Basis: Vector spaces are linearly independent & span the space. (e.g. The columns of every invertible matrix give a basis for $\mathbb{R}^n$.) The dimension of the space is the number of basis in the set of basis.

Four Space: Given $A \in \mathbb{R}^{m\times n}$, we have the definitions:

- Range/Column/Image space: $R(A) = C(A) = Im(A) = \{Ax \mid x \in \mathbb{R}^n\} \subseteq \mathbb{R}^m$.
- Row space: $C(A^\top) \subseteq \mathbb{R}^n$.
- Null/Kernel space: $N(A) = Ker(A) = \{x \in \mathbb{R}^n \mid Ax = 0\} \subseteq \mathbb{R}^n$.
- Left-Null space: $N(A^\top) = Ker(A^\top) = \left\{x \in \mathbb{R}^m \mid A^\top x = 0\right\} \subseteq \mathbb{R}^m$.

Properties:

- The column space of $A \iff$ the row space of $A^\top$.
- $0 \in N(A) = Ker(A)$.
- Rank-Nullity Theorem: $\mathrm{rank}(A) + \dim(N(A)) = n$
- $\dim(Im(A)) = \dim(Im(A^\top)) = rank(A) = r$
- $\dim(N(A)) = n - r, \dim(N(A^\top)) = m - r$.

Space Orthogonality: The orthogonal complement of a subspace V, $V^\perp$, contains every vector that is perpendicular to V.

- $C(A^\top) = N(A)^\perp$, i.e. The row space is perpendicular to the null space. (Think $Ax = 0$.)
- $C(A) = N(A^\top)^\perp$, i.e. The column space is perpendicular to the null space of $A^\top$ (left-null space).(Think $A^\top x = 0$.)
- Suppose $A \in \mathbb{R}^{m\times n}$, $\forall x \in \mathbb{R}^n$, it can be represented as $x = x_r + x_n$ where $x_r \in C(A^\top)$ and $x_n \in N(A)$.

rref of A: Suppose $R = rref(A)$, then

- $C(A^\top) = C(R^\top)$, i.e. same row space.
- $C(A) \neq C(R)$, the lase few entries of $C(R)$ could only be zero.
- $\dim(C(A)) = \dim(C(R))$, $(Ax = 0 \implies Rx = 0)$.
- $N(A) = N(R)$.

Link to equation $Ax = b$

The following statements are equivalent:

- There exists a solution to the equation $Ax = b$.
- $b \in R(A)$.
- $\mathrm{rank}(A) = \mathrm{rank}\left(\begin{bmatrix} A & b \end{bmatrix}\right)$

The following statements are equivalent:

- Solutions to the equation $Ax = b$ are unique.
- $N(A) = \{0\}$.
- $\mathrm{rank}(A) = n$.

11 Determinant

Proposition 11.1. ***(Basic 3)***

1. *swapping two columns:* $\left|\begin{bmatrix}\cdots & \boldsymbol{a}_i & \cdots & \boldsymbol{a}_j & \cdots\end{bmatrix}\right| = -\left|\begin{bmatrix}\cdots & \boldsymbol{a}_j & \cdots & \boldsymbol{a}_i & \cdots\end{bmatrix}\right|$
2. $\left|\begin{bmatrix}\cdots & \alpha\boldsymbol{u} + \beta\boldsymbol{v} & \cdots\end{bmatrix}\right| = \alpha\left|\begin{bmatrix}\cdots & \boldsymbol{u} & \cdots\end{bmatrix}\right| + \beta\left|\begin{bmatrix}\cdots & \boldsymbol{v} & \cdots\end{bmatrix}\right|$
3. *duplicates:* $\left|\begin{bmatrix}\cdots & \boldsymbol{a}_i & \cdots & \boldsymbol{a}_i & \cdots\end{bmatrix}\right| = 0$

Proposition 11.2. *If A is singular $\iff \{\boldsymbol{a}_1 \cdots \boldsymbol{a}_n\}$ linearly dependent $\iff |A| = 0$.*

Hint: singular -> one of the columns / rows linearly depends on the rest -> use Basic 2 and 3.

Proposition 11.3.

1. $|I| = 1$, $|P^{permu}| = 1/-1$ *(use Basic 1)*, $|P^{ortho}| = 1/-1$.
2. $|A| = |A^\top|$
3. $|AB| = |A| \cdot |B|$
4. $|A^{-1}| = \frac{1}{|A|}$
5. *Orthogonal matrix* $|Q| = \pm|1|$ *because* $Q^\top Q = I$ *gives* $|Q|^2 = 1$.

6. *Triangular matrix* $|U| = u_{11}u_{22}\cdots u_{nn}$
7. $|A| = |LU| = |L|\cdot|U| =$ *product of the pivots* u_{ii}
8. $\left|\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}\right| = |A|\cdot|B|$, $\left|\begin{bmatrix} A & C \\ 0 & B \end{bmatrix}\right| = |A|\cdot|B|$.

Proposition 11.4. ***(Cramer's Rule)*** *Cramer's Rule to Solve* $Ax = b$ *Start from*

$$\left[\ \boldsymbol{A}\ \right]\begin{bmatrix} \boldsymbol{x_1} & 0 & 0 \\ \boldsymbol{x_2} & 1 & 0 \\ \boldsymbol{x_3} & 0 & 1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{b_1} & a_{12} & a_{13} \\ \boldsymbol{b_2} & a_{22} & a_{23} \\ \boldsymbol{b_3} & a_{32} & a_{33} \end{bmatrix} = \boldsymbol{B_1}$$

Use $(\det \boldsymbol{A})(\boldsymbol{x}_1) = (\det \boldsymbol{B}_1)$ *to find* $\boldsymbol{x}_1$ *Same idea* $[A]\begin{bmatrix} 1 & x_1 & 0 \\ 0 & x_2 & 0 \\ 0 & x_3 & 1 \end{bmatrix} = \begin{bmatrix} a_1 & b & a_3 \end{bmatrix} = B_2$ $\quad x_1 = \dfrac{\det B_1}{\det A}$ $\quad x_2 = \dfrac{\det B_2}{\det A}$ *Cramer's Rule is usually not efficient! Too many determinants*

$$\begin{bmatrix} 3 & 2 \\ 5 & 4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 12 \\ 22 \end{bmatrix} \quad \boldsymbol{B}_1 = \begin{bmatrix} 12 & 2 \\ 22 & 4 \end{bmatrix} \quad \boldsymbol{B}_2 = \begin{bmatrix} 3 & 12 \\ 5 & 22 \end{bmatrix} \quad x_1 = \frac{\det B_1}{\det A} = \frac{4}{2} \quad x_2 = \frac{2}{2}$$

12 Least squares

Suppose $A \in \mathbb{R}^{n\times p}$. When the linear equations $Ax = b$ are overdetermined ($n > p$) and there is no solution, we minimize the 2-norm of the residual:

$$\underset{x}{\text{minimize}}\|Ax - b\|_2 \Longrightarrow A^\top A\hat{x} = A^\top b \quad \to \text{Normal Equation}$$

The normal equations have a unique solution iff $rank(A) = p$ (full column rank / the columns of A are linearly independent). Why? Because $A^\top A \in \mathbb{R}^{p\times p}$ and $Im(A^\top A) = Im(A) \Longrightarrow rank(A^\top A) = rank(A) = p \Longrightarrow A^\top A$ is full rank.

Then,

$$\hat{x} = \left(A^\top A\right)^{-1} A^\top b = A^\dagger b$$

where $A^\dagger = (A^\top A)^{-1}A^\top$ is the Moore-Penrose inverse (or pseudo-inverse).

13 Orthogonal matrices

A square matrix U is orthogonal if $U^\top U = I$.

Some properties of orthogonal U and V:

- **Orthogonal Columns and Rows**: $u_i \cdot u_j = 0$, $\|u_i\| = \|u_j\| = 1$.
- **Orthogonal Basis**: The columns (or rows) of an orthogonal matrix form an orthogonal basis for the vector space.
- **Orthogonal transformations preserve angles & length**: $(Ux)^\top(Uz) = x^\top z$ and $\|Ux\|_2 = \|x\|_2$.
- Certain matrix norms are also invariant: $\|UAV\|_2 = \|A\|_2$ and $\|UAV\|_F = \|A\|_F$
- If U is square, $U^\top U = UU^\top = I$ and $U^{-1} = U^\top$.
- UV is orthogonal.

Every subspace has an orthonormal basis: For any $A \in \mathbb{R}^{m\times n}$, there exists an orthogonal $U \in \mathbb{R}^{m\times r}$ such that $R(A) = R(U)$ and $r = \text{rank}(A)$. One way to find U is using Gram-Schmidt.

14 Projections

If $P \in \mathbb{R}^{n\times n}$ satisfies $P^2 = P$, $P^\top = P$ it's called a projection matrix.

In general, $P : \mathbb{R}^n \to S$, where $S \subseteq \mathbb{R}^n$ is a subspace.

If P is a projection matrix, so is $(I - P)$. We can uniquely decompose:

$$x = u + v = Px + (I-P)x \quad \text{where } u \in S, v \in S^\perp$$

Pythagorean theorem: $\|x\|_2^2 = \|u\|_2^2 + \|v\|_2^2$

If $A \in \mathbb{R}^{m\times n}$ has linearly independent columns, then the projection onto $R(A)$ is given by $P = A\left(A^\top A\right)^{-1}A^\top$.

$$e \perp R(A) \Longrightarrow (b - Ax)^\top A = 0 \Longrightarrow x = \left(A^\top A\right)^{-1}A^\top b \Longrightarrow p = Ax = A\left(A^\top A\right)^{-1}A^\top b$$

Least-squares: decompose b using the projection above:

$$\begin{aligned} b &= A\left(A^\top A\right)^{-1}A^\top b + \left(I - A\left(A^\top A\right)^{-1}A^\top\right)b \\ &= A\hat{x} + (b - A\hat{x}) \end{aligned}$$

where $\hat{x} = \left(A^\top A\right)^{-1}A^\top b$ is the LS estimate. Therefore the optimal residual is orthogonal to $A\hat{x}$.

15 QR Decomposition

16 Eigenvalues and Eigenvectors

Intuition: The whole idea is to avoid the complexity presented by matrix A. It's generally more convenient to deal with λx instead of Ax.

Basics: Suppose $A \in \mathbb{R}^{n\times n}$ is a square matrix.

- $A^k x = \lambda^k x$.
- $A^{-1}x = \lambda^{-1}x$.
- $(A + cI)x = (\lambda + c)x$.
- If α is an eigenvalue of A and β is an eigenvalue of B, then $\alpha\beta$ is NOT an eigenvalue of AB; and $\alpha + \beta$ is NOT an eigenvalue of $A + B$.
- $Ax = \lambda x \Longrightarrow (A - \lambda I)x$ to have non-zero solutions for $x \Longrightarrow (A - \lambda I)$ is singular $\Longrightarrow \det(A - \lambda I) = 0$.
- The mapping between eigenvalue and eigenvector is 1-to-1.

  But why are there cases where one eigenvalue maps to two eigenvector?

  This is because there could be two eigenvalues having the same value.
- If $A \in \mathbb{R}^{n\times n}$, then A has at most n different eigenvalues.
- If $A \in \mathbb{R}^{n\times n}$ has $d \le n$ distinct eigenvalues $\lambda_1, \ldots, \lambda_d$, then we have at least d independent eigenvectors. (E.g. Identity matrix I has only 1 distinct eigenvalue, but it has $e_1, e_2, \ldots, e_n$, in total, n independent eigenvectors.)
- Eigenvectors $x_1, \ldots, x_j$ correspond to distinct eigenvalues are linearly independent.
- One eigenvector (unit) of A cannot correspond to two or more different eigenvalues.

  Otherwise, $A\vec{x} = \lambda_1\vec{x}$, $A\vec{x} = \lambda_2\vec{x}$, $\lambda_1 \neq \lambda_2 \Rightarrow 0 = (\lambda_1 - \lambda_2)\vec{x} \Rightarrow \vec{x} = 0$; and that's a contradiction!
- Elementary matrices E, P, row-exchange/permutation matrices DOES NOT preserve eigenvalues.

Characteristic Polynomial:

$A - \lambda I$: characteristic polunomial of A; $|A - \lambda I|$ is the characteristic equation of A.

- Vieta's Formula: $\sum_{i=1}^n \lambda_i = trace(A) = \sum_{i=1}^n a_{ii}$ and $\prod_{i=1}^n \lambda_i = |A|$.
- Eigenvalues of a triangular matrix lie along its diagonal. (Because $\det(A) = \prod_{i=1}^n a_{ii}$ if $A \in \mathbb{R}^{n\times n}$ is triangular. $\Longrightarrow \det(A - \lambda I) = \prod_{i=1}^n (a_{ii} - \lambda)$.)

17 Diagonalizing a Matrix

Suppose $A \in \mathbb{R}^{n\times n}$ has n linearly independent eigenvectors $x_1, x_2, \ldots, x_n$. Set $X = \begin{bmatrix} | & & | \\ x_1 & \cdots & x_n \\ | & & | \end{bmatrix}$ and $\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$, then we have $AX = X\Lambda$. Since X is formed by n linearly independent vectors, X is invertible. Hence

$$A = X\Lambda X^{-1}$$

Properties:

- Any matrix with no repeated eigenvalues, i.e. n eigenvalues, can be diagonalized.

  If $\lambda_1, \ldots, \lambda_n$ are all different $\Longrightarrow$ we have $\geq n$ independent eigenvectors. Since $\dim(A) = n$, we have exactly n eigenvectors. Hence, X is invertible.
- $A^k = (X\Lambda X^{-1})^k = X\Lambda^k X^{-1}$.
- If all eigenvalues of A has $|\lambda| < 1$, then $\lim_{k\to\infty} \Lambda^k = 0$. Since matrix $A^k = XA^kX^{-1}$, $\lim_{k\to\infty} A^k = 0$ matrix.
- If $A = S\Lambda S^{-1}$, then $A^{-1} = S\Lambda^{-1}S^{-1}$ (same eigenvectors, inverse eigenvalues).
- If A has 0 as its eigenvalue, then $Ax = 0x \Longrightarrow Ax = 0$ has non-zero eigenvectors, $\Longrightarrow A$ is singular.

GM, AM

- Geometric Multiplicity = GM = dim of Null space of $(A - \lambda I)$: counts the independent eigenvectors for λ.
- Algebraic Mutciplicity =AM $\rightarrow$ look at roots of $\det(A - \lambda I)$: counts the repetition of λ among the eigenvalues.

E.g. If A has $\lambda = 4, 4, 4 \Rightarrow AM = 3, GM = 1, 2,$ or 3.

If for A, $GM \leq AM$. That means we have an eigenvalue repeated AM times but have only GM lines of eigenvectors correspond to it $\Rightarrow$ lack of independent eigenvectors for $\mathbb{R}^{n\times n}$ eigenvector matrix X. $\Rightarrow A$ is not diagonalizable.

Similar Matrices:

If $B \in \mathbb{R}^{n\times n}$ - invertible, $C \in \mathbb{R}^{n\times n}$ - constant matrix, then $A = BCB^{-1}$ are similar matrices (one for each choice of invertible matrix B).

- All matrices $A = BCB^{-1}$ are "similar". They all share the eigenvalues of C.

  Proof:

  $$(A - \lambda I) = BCB^{-1} - \lambda I = BCB^{-1} - B\lambda IB^{-1} = B(C - \lambda I)B^{-1}$$

  Then, we have

  $$\det(C - \lambda I) = \det(A - \lambda I)$$

  Hence, they share the same set of eigenvalues as C's.

18 Symmetric Matrix

$S = S^\top \in \mathbb{R}^{n\times n}$.

Properties:

- S has only real eigenvalues.
- **(Important!)** S always has n independent, mutually orthogonal eigenvectors (a set of orthonormal basis).

Spectral Theorem Every real symmetric matrix has factorization $S = Q\Lambda Q^\top$, with n (countingmultiplicities) real eigenvalues in Λ and orthonormal eigenvectors as columns of Q:

$$S = Q\Lambda Q^{-1} = Q\Lambda Q^\top = \lambda_1 q_1 q_1^\top + \cdots + \lambda_n q_n q_n^\top$$

where q_i's are orthonormal eigenvectors of S.

Symmetric Matrix Decompositions

- **LU:**

  $$A = LDU = LDL^\top = LD^{1/2}(LD^{1/2})^\top$$

  where the 4th term is the *square-root-free* Cholesky Decomposition of A which is only valid if A is positive definite, since you want eigenvalues to be all positive so that you could take the square root.

  Note: It is reminiscent of the eigen-decomposition of real symmetric matrices, $A = Q\Lambda Q^\top$, but is quite different in practice because Λ and D are not similar matrices.
- **SVD:**

  $$S = Q\Lambda Q^{-1} = Q\Lambda Q^\top \xlongequal{S \text{ is positive-definite}} U\Sigma V^\top$$

Proposition 18.1. *For any matrix $A \in \mathbb{R}^{m\times n}$, the matrix $A^\top A \in \mathbb{R}^{n\times n}$ is always square, symmetric, and positive semi-definite($\forall x \neq 0, x^\top A^\top Ax = \|Ax\|^2 \geq 0$). If in addition A has linearly independent columns, then A^TA is positive definite.*

Skew-symmetric Matrix $A^\top = -A$.

- Eigenvalues are purely imaginary.
- A always has n independent, mutually orthogonal eigenvectors (a set of orthonormal basis).

Quadratic Form $q = x^\top Qx$, where Q is a symmetic matrix.

E.g. Consider the case of quadratic forms in three variables x, y, z. The matrix Q has the form

$$Q = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & k \end{bmatrix}$$

The above formula gives

$$q = ax^2 + ey^2 + kz^2 + (b + d)xy + (c + g)xz + (f + h)yz.$$

So, two different matrices define the same quadratic form if and only if they have the same elements on the diagonal and the same values for the sums $b + d$, $c + g$ and $f + h$. In particular, the quadratic form q is defined by a unique symmetric matrix

$$Q = \begin{bmatrix} a & \frac{b+d}{2} & \frac{c+g}{2} \\ \frac{b+d}{2} & e & \frac{f+h}{2} \\ \frac{c+g}{2} & \frac{f+h}{2} & k \end{bmatrix}$$

19 Positive-Definite Matrix

A positive definite if

1. A is symmetric.
2. Eigenvalues of A are all positive.
3. All the upper-left determinants are positive
4. $x^TAx > 0$ unless $x = 0$.

- If A and B are positive definite then so is $A + B$.

Example: For symmetric positive-definite matrix A, we have

$$A = \begin{bmatrix} 4 & 12 & -16 \\ 12 & 37 & -43 \\ -16 & -43 & 98 \end{bmatrix} = Q\Lambda Q^\top = \begin{bmatrix} -0.16 & 0.21 & 0.96 \\ -0.45 & 0.84 & -0.26 \\ 0.87 & 0.48 & 0.04 \end{bmatrix} \begin{bmatrix} 123.47 & 0 & 0 \\ 0 & 15.50 & 0 \\ 0 & 0 & 0.018 \end{bmatrix} \begin{bmatrix} -0.16 & 0.21 & 0.96 \\ -0.45 & 0.84 & -0.26 \\ 0.87 & 0.48 & 0.04 \end{bmatrix} = U\Sigma V^\top$$

$$= LL^\top = \begin{bmatrix} 2 & 0 & 0 \\ 6 & 1 & 0 \\ -8 & 5 & 3 \end{bmatrix} \begin{bmatrix} 2 & 6 & -8 \\ 0 & 1 & 5 \\ 0 & 0 & 3 \end{bmatrix} = LDL^\top \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ -4 & 5 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 9 \end{bmatrix} \begin{bmatrix} 1 & 3 & -4 \\ 0 & 1 & 5 \\ 0 & 0 & 1 \end{bmatrix}$$

20 Cholesky Decomposition

21 The Singular Value Decomposition

For any matrix $\mathbf{X} \in \mathbb{R}^{m\times n}$, there exist orthogonal matrices $\mathbf{U}, \mathbf{V}$ and a $m \times n$ diagonal matrix $\mathbf{\Sigma}$

$$\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_m] \in \mathbb{R}^{m\times m}$$
$$\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_n] \in \mathbb{R}^{n\times n}$$
$$\mathbf{\Sigma} = \operatorname{diag}(\sigma_1, \ldots, \sigma_p) \in \mathbb{R}^{m\times n}, p := \min\{m, n\}$$

with $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p \geq 0$, such that

$$\mathbf{X} = \mathbf{U\Sigma V}'$$

To be more specific, matrix $\mathbf{X} \in \mathbb{C}_r^{m\times n}$ can be decomposed as

$$\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^* = \begin{bmatrix}\mathbf{U}_{\mathcal{R}} & \mathbf{U}_{\mathcal{N}}\end{bmatrix} \left[\begin{array}{cccc|cc} \sigma_1 & 0 & \cdots & & \cdots & 0 \\ 0 & \sigma_2 & & & & \\ \vdots & & \ddots & & & \\ & & & \sigma_r & & \\ \hline & & & & 0 & \\ \vdots & & & & & \ddots \\ 0 & & & & & 0 \end{array}\right] \begin{bmatrix}\mathbf{V}_{\mathcal{R}}{}^* \\ \mathbf{V}_{\mathcal{N}}{}^*\end{bmatrix} = \begin{bmatrix}u_1 & \cdots & u_r & u_{r+1} & \cdots & u_m\end{bmatrix} \begin{bmatrix}\mathbf{S}_{r\times r} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix} \begin{bmatrix} v_1^* \\ \vdots \\ v_r^* \\ v_{r+1}^* \\ \vdots \\ v_n^* \end{bmatrix} \longrightarrow \textbf{full SVD}$$

$$\xlongequal{\text{spectral decomposition}} \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top + \sum_{i=r+1}^{p} 0 \cdot \mathbf{u}_i \mathbf{v}_i^\top = \mathbf{U}_{\mathcal{R}} \mathbf{S}_{r\times r} \mathbf{V}_{\mathcal{R}}{}^* \longrightarrow \textbf{Compact SVD}$$

$$\approx \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top, k \le t \longrightarrow \textbf{Truncated SVD}$$

The r positive singular values are real and ordered (descending) and are the square root of non-zero eigenvalues of the product matrices $\mathbf{A}^*\mathbf{A}$ and $\mathbf{A}\mathbf{A}^*$. Please note that the singular values only correspond to range space vectors, i.e., $\mathbf{U}_{\mathcal{R}}$ and $\mathbf{V}_{\mathcal{R}}$.

Remark 21.1.

1. This decomposition is not unique: the singular values part Σ is unique; however the signs in the left and right singular vectors can be interchanged. Besides when at least one singular value is zero, there are many possible corresponding singular vectors.
2. if $\mathbf{A}$ is **real symmetric** then (spectral theorem) it is diagonalizable and therefore has at least one eigen-decomposition $\mathbf{A} = \mathbf{Q}\Lambda\mathbf{Q}^{-1} = \mathbf{Q}\Lambda\mathbf{Q}^\top$. In general this decomposition is not unique: the eigenvalues part Λ is unique; however the eigenvectors part $\mathbf{Q}$ is only unique if no eigenvalue is zero.
3. If $\mathbf{A}$ is a **real symmetric and at least positive semi-definite**, then Σ is a diagonal matrix containing the eigenvalues, and $U = V$. That is, the SVD

$$\mathbf{A} \xlongequal{\text{SVD}} \mathbf{U}\Sigma\mathbf{V}^\top \xlongequal{\mathbf{U}=\mathbf{V}=\mathbf{Q}, \Sigma=\Lambda} \mathbf{Q}\Lambda\mathbf{Q}^\top \longrightarrow \text{eigen-decomposition}$$

△

Fundamental Subspaces

The column vectors of $\mathbf{U}$ are called **left singular vectors** and are an orthonormal span of $\mathbb{C}^m$ (column space) $\mathbf{C}^n = \mathcal{R}(\mathbf{X}^*) \oplus \mathcal{N}(\mathbf{X})$.

The column vectors of $\mathbf{V}$ are called **right singular vectors** are an orthonormal span of $\mathbb{C}^n$ (row space) $\mathbf{C}^m = \mathcal{R}(\mathbf{X}) \oplus \mathcal{N}(\mathbf{X}^*)$.

matrix	subspace		orthogonal projection matrix
$\mathbf{U}_{\mathcal{R}} \in \mathbb{C}^{m\times r}$	$\mathcal{R}(\mathbf{X})$	$=\text{span}\{u_1,\ldots,u_r\}$	$\mathbf{U}_{\mathcal{R}}\mathbf{U}_{\mathcal{R}}^\top$ orth proj onto $\mathcal{R}(\mathbf{X})$
$\mathbf{V}_{\mathcal{R}} \in \mathbb{C}^{n\times r}$	$\mathcal{R}(\mathbf{X}^*)$	$=\text{span}\{v_1,\ldots,v_r\}$	$\mathbf{V}_{\mathcal{R}}\mathbf{V}_{\mathcal{R}}^\top$ orth proj onto $\mathcal{R}(\mathbf{X}^*)$
$\mathbf{U}_{\mathcal{N}} \in \mathbb{C}^{m\times m-r}$	$\mathcal{N}(\mathbf{X}^*)$	$=\text{span}\{u_{r+1},\ldots,u_m\}$	$\mathbf{U}_{\mathcal{N}}\mathbf{U}_{\mathcal{N}}^\top$ orth proj onto $\mathcal{N}(\mathbf{X}^*)$
$\mathbf{V}_{\mathcal{N}} \in \mathbb{C}^{n\times n-r}$	$\mathcal{N}(\mathbf{X})$	$=\text{span}\{v_{r+1},\ldots,v_n\}$	$\mathbf{V}_{\mathcal{N}}\mathbf{V}_{\mathcal{N}}^\top$ orth proj onto $\mathcal{N}(\mathbf{X})$

The singular value decomposition provides an orthonormal basis for the four fundamental subspaces.

Properties of the SVD

Basics

$$\begin{cases} \mathbf{A}^\top\mathbf{A} &= \mathbf{V}\left(\Sigma^\top\Sigma\right)\mathbf{V}^\top \Longrightarrow \mathbf{A}^\top\mathbf{A}\mathbf{V} = \mathbf{V}\left(\Sigma^\top\Sigma\right) \xlongequal{\Sigma^\top\Sigma \text{ diagonal}} \Sigma^\top\Sigma\mathbf{V} \\ \mathbf{A}\mathbf{A}^\top &= \mathbf{U}\left(\Sigma\Sigma^\top\right)\mathbf{U}^\top \Longrightarrow \mathbf{A}\mathbf{A}^\top\mathbf{U} = \mathbf{V}\left(\Sigma\Sigma^\top\right) \xlongequal{\Sigma\Sigma^\top \text{ diagonal}} \Sigma\Sigma^\top\mathbf{U} \end{cases}$$

- $A\mathbf{v}_i = \sigma_i\mathbf{u}_i$ and $A^\top\mathbf{u}_i = \sigma_i\mathbf{v}_i$.
- $\text{rank}(A) = \text{rank}\,\Sigma = r$.
- $\sigma_A^2 = \sigma_{AA^\top} = \sigma_{A^\top A}$. That is, The singular values of $A^\top A$ and $AA^\top$ are $\sigma_1^2,\ldots,\sigma_n^2$, and thus $\|A^\top A\|_2 = \|AA^\top\|_2 = \sigma_{\max}^2$.
- The right singular vectors (columns of V) are eigenvectors of $A^\top A$
- The left singular vectors (columns of U) are eigenvectors of $AA^\top$

Matrix Norm

If $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $p := \min\{m,n\}$, then the following identities hold

$$\|\mathbf{A}\|_F := \sqrt{\sum_{i=1,j=1}^{m,n} |a_{ij}|^2} = \sqrt{\sigma_1^2 + \ldots + \sigma_p^2},$$

$$\|\mathbf{A}\|_2 := \max\frac{\|\mathbf{A}\mathbf{v}\|_2}{\|\mathbf{v}\|_2} = \sigma_1,$$

$$\min\frac{\|\mathbf{A}\mathbf{v}\|_2}{\|\mathbf{v}\|_2} = \sigma_n (m \ge n).$$

Eckart-Young Theorem

Suppose $\mathbf{X} = \sum_{i=1}^r \sigma_i\mathbf{u}_i\mathbf{v}_i'$, $k < r = \text{rank}(\mathbf{X})$ and denote by $\mathbf{X}_k := \sum_{i=1}^k \sigma_i\mathbf{u}_i\mathbf{v}_i'$ the k-truncated SVD of $\mathbf{X}$. Then $\mathbf{X}_k$ is the best rank- k approximation of $\mathbf{X}$ both in the L^2-norm and the Frobenius norm, that is

$$\min_{\mathbf{Y} \text{ s.t. } \text{rank}(\mathbf{Y})=k} \|\mathbf{X}-\mathbf{Y}\|_2 = \|\mathbf{X}-\mathbf{X}_k\|_2 = \sigma_{k+1}$$

and

$$\min_{\mathbf{Y} \text{ s.t. } \text{rank}(\mathbf{Y})=k} \|\mathbf{X}-\mathbf{Y}\|_F^2 = \|\mathbf{X}-\mathbf{X}_k\|_F^2 = \sum_{k+1}^{N}\sigma_i^2$$

SVD and Linear Regression

Suppose $\mathbf{b} \in \mathbb{R}^m$ and $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}' \in \mathbb{R}^{m\times n}$ with $r = \text{rank}(\mathbf{A})$ and

$$\mathbf{U} = [\mathbf{u}_1,\ldots,\mathbf{u}_m]$$
$$\mathbf{V} = [\mathbf{v}_1,\ldots,\mathbf{v}_n]$$

Then

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2 = \sum_{i=1}^{r} \frac{\mathbf{u}_i'\mathbf{b}}{\sigma_i}\mathbf{v}_i$$

and

$$\|\mathbf{A}\mathbf{x}^*-\mathbf{b}\|_2^2 = \sum_{i=r+1}^{m} (\mathbf{u}_i'\mathbf{b})^2$$

The Moore-Penrose Generalized Inverse

Suppose $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}' = \sum_{i=1}^r \sigma_i\mathbf{u}_i\mathbf{v}_i$ with $r = \text{rank}(\mathbf{A})$. Then we define the Moore-Penrose inverse of $\mathbf{A}$ by

$$\mathbf{A}^\dagger := \mathbf{V}\Sigma^\dagger\mathbf{U}'$$

where

$$\Sigma^\dagger = \left(\Sigma^\top\Sigma\right)^{-1}\Sigma^\top = \text{diag}\left(\sigma_1^{-1},\ldots,\sigma_r^{-1},0,\ldots,0\right) \in \mathbb{R}^{n\times m}$$

It is easy to verify that $\mathbf{A}^+$satisfies the four Moore-Penrose conditions

$$\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}, \mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+, \left(\mathbf{A}\mathbf{A}^+\right)' = \mathbf{A}\mathbf{A}^+, \left(\mathbf{A}^+\mathbf{A}\right)' = \mathbf{A}^+\mathbf{A}$$

Remark 21.2. There is an infinite number of ways in which one can define a pseudo inverse. What makes the Moore-Penrose inverse useful is that it is unique. In other words, there is no other inverse that satisfies the Moore-Penrose conditions above.

Note that from the Moore-Penrose conditions it follows that $\mathbf{A}\mathbf{A}^{+}$ and $\mathbf{A}^{+}\mathbf{A}$ are orthogonal projections onto range $(\mathbf{A})$ and range $(\mathbf{A}')$, respectively.

△

22 Triangular Matrix

- The eigenvalues of a triangular matrix are the entries on its main diagonal.

23 Some nice problems

Proposition 23.1. *Diagonally dominant matrices are invertible. Diagonally dominant:* $|a_{ii}| > \sum_{j\neq i} |a_{ij}|$

Solution: If $\exists \vec{x} \neq \overrightarrow{0}$ s.t. $A\vec{x} = \overrightarrow{0}$, $A \in \mathbb{R}^{n\times n}$. then $\sum_{j=1}^{n} a_{ij}x_j = 0 \quad \forall i \in [1, n], i \in Z^+$ W LOG suppose $|\vec{x}_i| \geq |\vec{x}_k| \quad \forall k \neq i \quad k \in [1, n], k \in Z^+$.

$\Rightarrow |a_{ii}|\, x_i| = |\sum_{j\neq i} a_{ij}x_j \Big| \leqslant \sum_{j\neq i}\Big| a_{ij}||x_j\,|$ That's a contradiction since we have.

$$|a_{ii}|\,|x_i| > \left(\sum_{j\neq i} |a_{ij}|\right) |x_i| > \sum_{j\neq i} |a_{ij}|\,|x_j|$$

Proposition 23.2. *Prove that matrix* $A = au_1v_1^\top + bu_2v_2^\top$ *where* $u_i \in \mathbb{R}^m$, $v_i \in \mathbb{R}^n$ *and* $A \in \mathbb{R}^{m\times n}$ *is a rank-2 matrix.*

Solution: We have

$$A = u_1v_1^\top + u_2v_2^\top = \begin{bmatrix} | & | \\ u_1 & u_2 \\ | & | \end{bmatrix} \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \begin{bmatrix} - & v_1^\top & - \\ - & v_2^\top & - \end{bmatrix} = UV \Longrightarrow rankA =$$

Proposition 23.3. *Suppose* $A \in \mathbb{R}^{m\times n}$, $\forall x \in \mathbb{R}^n$, *it can be represented as* $x = x_r + x_n$ *where* $x_r \in C(A^\top)$ *and* $x_n \in N(A)$.

Solution:

Proposition 23.4. *Eigenvectors* $x_1, \ldots, x_j$ *that correspond to distinct eigenvalues are linearly independent.*

Solution: Assume dependent $c_i\vec{x}_i + c_j\vec{x}_j = 0 \Longrightarrow \lambda_i c_i\vec{x}_i + \lambda_i c_j\vec{x}_j = 0$. We have

$$\Rightarrow A\left(c_i\vec{x}_i + c_j\vec{x}_j\right) = \lambda_i c_i\vec{x}_i + \lambda_j c_j\overrightarrow{x_j} = \lambda_i c_i\vec{x}_i + \lambda_i c_j\vec{x}_j + (\lambda_j - \lambda_i)c_j\vec{x}_j = 0 \Longrightarrow (\lambda_j - \lambda_i)c_j\vec{x}_j = 0 \Longrightarrow c_j = 0 \Longrightarrow c_i = 0.$$

This proof can be extended to $c_1\vec{x}_i + c_2\vec{x}_i + \cdots + c_j\vec{x}_j = 0$

Proposition 23.5. A, B *share the same* n *independent eigenvectors if and only if* $AB = BA$.

Solution:

"$\Longrightarrow$"

Suppose A, B share same n independent eigenvectors. $\vec{v}_1, \ldots, \vec{V}_n$, then A, B have eigen-decompositions $A = S\Lambda_a S^{-1}$, $B = S\Lambda_b S^{-1}$.

$$\Rightarrow AB = S\Lambda_a S^{-1}S\Lambda_b S^{-1} = S\Lambda_a\Lambda_b S^{-1} = S\Lambda_b\Lambda_a S^{-1} = BA$$

"$\Longleftarrow$"

Suppose $AB = BA$, and $\vec{v}_1, \ldots, \vec{v}_n$ are n unit-length, independent eigenvectors of A, then

$$AB\overrightarrow{v_i} = BA\overrightarrow{v_i} = B\alpha_i\overrightarrow{v_i} = \alpha_i B\overrightarrow{v_i} \Rightarrow B\vec{v}_i \text{ is eigenvector of } A$$

Since the eigenvectors of A are independent, we have $B\overrightarrow{v_i} = \beta\overrightarrow{v_i} \Rightarrow \vec{v}_i$ is an eigenvector of B.

Proposition 23.6. *How can you estimate the eigenvalues of any* A *? (Gershgorin)*

Solution: **Intuition:** Every eigenvalue of A must be "near" at least one of the entries a_{ii} on the main diagonal; i.e. every λ is in the circle around one or more diagonal entries a_{ii}:

$$|a_{ii} - \lambda| \leq R_i = \sum_{j\neq i} |a_{ij}|$$

If λ is eigenvalue $\Rightarrow (A - \lambda I)$ is singular $\Rightarrow \det(A - \lambda I) = 0 \Rightarrow (A - \lambda I)$ is not invertible. $\Rightarrow (A - \lambda I)$ is not diagonally dominant.

$$\Rightarrow \exists i \text{ sit } |a_{ii} - \lambda| \leq R_i$$

Proposition 23.7. ***(Spectral Theorem)*** *Every real symmetric matrix has factorization* $S = Q\Lambda Q^\top$, *with real eigenvalues in* Λ *and orthonormal eigenvectors as columns of* Q:

$$S = Q\Lambda Q^{-1} = Q\Lambda Q^\top$$

Solution:

1. All eigenvalues are real. Suppose $S\vec{x} = \lambda\vec{x}$ where $\lambda = a + bi, b \neq 0$. Take conjugate, we get
$$S\overline{\vec{x}} = \bar{\lambda}\overline{\vec{x}} \Rightarrow \overline{\vec{x}}^\top S^\top = \bar{\lambda}\overline{\vec{x}}^\top \Rightarrow \overline{\vec{x}}^\top S^\top \vec{x} = \bar{\lambda}\overline{\vec{x}}^\top \vec{x}$$
Since $S\vec{x} = \lambda\vec{x}$, we have $\overline{\vec{x}}^\top S^\top \vec{x} = \lambda\overline{\vec{x}}^\top \vec{x}$, then
$$\Rightarrow 0 = (\lambda - \bar{\lambda})\overline{\vec{x}}^\top \vec{x} \Rightarrow \lambda = \bar{\lambda} \Rightarrow b = 0 \quad \text{contradiction!}$$
Hence, $\lambda \in \mathbb{R}$.

2. Eigenvectors of a real symmetric matrix (when they correspond to different λ 's) are always perpendicular.
Suppose $S\vec{x}_1 = \lambda_1\vec{x}_1; S\vec{x}_2 = \lambda_2\vec{x}_2; \lambda_1 \neq \lambda_2$, then
$$\lambda_1\vec{x}_2^\top\vec{x}_1 = \vec{x}_2^\top S\vec{x}_1 \xlongequal{constant} (\vec{x}_2^\top S\vec{x}_1)^\top = \vec{x}_1^\top S^\top\vec{x}_2 \xlongequal{symmetric} \vec{x}_1^\top S\vec{x}_2 = \lambda_2\vec{x}_1^\top\vec{x}_2$$
Since $\vec{x}_2^\top\vec{x}_1 \xlongequal{constant} \vec{x}_1^\top\vec{x}_2$, we have
$$(\lambda_2 - \lambda_1)\vec{x}_2^\top\vec{x}_1 \Rightarrow \vec{x}_2^\top\vec{x}_1 = 0 \rightarrow \text{orthogonal.}$$

Problem 23.8. 2-norm of orthogonal transformation of a matrix is invariant: For any matrix A and an orthogonal matrix Q, we have
$$\|QA\|_2 = \|A\|_2 \text{ and } \|AQ\|_2 = \|A\|_2$$

Solution: For the first equation, we have
$$\|QA\|_2^2 = (QAx)^T(QAx) = (Ax)^T(Ax) = \|A\|_2^2$$
For the second equation, recall that the 2-norm for matrices is defined as
$$\|B\|_2 = \sup_{\|x\|=1} \|Bx\|.$$
But for any orthogonal matrix Q we have that $\|Qx\| = \|x\|$. Thus, we can write
$$\|AQ\|_2 = \sup_{\|x\|=1} \|AQx\| = \sup_{\|Qx\|=1} \|AQx\| = \sup_{\|y\|=1} \|Ay\| = \|A\|_2.$$
Together, if U, V are comformable and orthogonal, then we have
$$\|UAV\|_2 = \|A\|_2$$

Theorem 23.9. ***(Shur's Theorem)*** *If $\boldsymbol{A}$ is a square real matrix with real eigenvalues, then there is an orthogonal matrix $\boldsymbol{Q}$ and an upper triangular matrix $\boldsymbol{T}$ such that $\boldsymbol{A} = \boldsymbol{QTQ}^{\mathrm{T}}$.*

Solution: Note that $\boldsymbol{A} = \boldsymbol{QTQ}^{\mathrm{T}} \Leftrightarrow \boldsymbol{AQ} = \boldsymbol{QT}$. Let $\boldsymbol{q}_1$ be an eigenvector of norm 1, with eigenvalue λ_1. Let $\boldsymbol{q}_2, \ldots, \boldsymbol{q}_n$ be any orthonormal vectors orthogonal to $\boldsymbol{q}_1$. Let $\boldsymbol{Q}_1 = [\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n]$. Then $\boldsymbol{Q}_1^{\mathrm{T}}\boldsymbol{Q}_1 = \boldsymbol{I}$, and
$$\boldsymbol{Q}_1^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Q}_1 = \begin{pmatrix} \lambda_1 & \cdots \\ \underline{0} & \boldsymbol{A}_2 \end{pmatrix}$$
Now I claim that $\boldsymbol{A}_2$ has eigenvalues $\lambda_2, \ldots, \lambda_n$. This is true because
$$\begin{aligned} \det(\boldsymbol{A} - \lambda\boldsymbol{I}) &= \det \boldsymbol{Q}_1^{\mathrm{T}} \det(\boldsymbol{A} - \lambda\boldsymbol{I}) \det \boldsymbol{Q}_1 = \det\left(\boldsymbol{Q}_1^{\mathrm{T}}(\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{Q}_1\right) \\ &= \det\left(\boldsymbol{Q}_1^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Q}_1 - \lambda\boldsymbol{Q}_1^{\mathrm{T}}\boldsymbol{Q}_1\right) = \det\begin{pmatrix} (\lambda_1 - \lambda) & \cdots \\ \mathbf{0} & (\boldsymbol{A}_2 - \lambda\boldsymbol{I}) \end{pmatrix} \\ &= (\lambda_1 - \lambda) \det(\boldsymbol{A}_2 - \lambda\boldsymbol{I}). \end{aligned}$$
So $\boldsymbol{A}_2$ has real eigenvalues, namely $\lambda_2, \ldots, \lambda_n$. Now we proceed by induction. Suppose we have proved the theorem for $n = k$. Then we use this fact to prove the theorem is true for $n = k + 1$. Note that the theorem is trivial if $n = 1$.

So for $n = k + 1$, we proceed as above and then apply the known theorem to $\boldsymbol{A}_2$, which is $k \times k$. We find that $\boldsymbol{A}_2 = \boldsymbol{Q}_2\boldsymbol{T}_2\boldsymbol{Q}_2^{\mathrm{T}}$. Now this is the hard part. Let $\boldsymbol{Q}_1$ and $\boldsymbol{A}_2$ be as above, and let
$$\boldsymbol{Q} = \boldsymbol{Q}_1\begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Q}_2 \end{pmatrix}$$
Then
$$\begin{aligned} \boldsymbol{AQ} = &\boldsymbol{A}\boldsymbol{Q}_1\begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Q}_2 \end{pmatrix} = \boldsymbol{Q}_1\begin{pmatrix} \lambda_1 & \cdots \\ \mathbf{0} & \boldsymbol{A}_2 \end{pmatrix}\begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Q}_2 \end{pmatrix} \\ &= \boldsymbol{Q}_1\begin{pmatrix} \lambda_1 & \cdots \\ \mathbf{0} & \boldsymbol{A}_2\boldsymbol{Q}_2 \end{pmatrix} = \boldsymbol{Q}_1\begin{pmatrix} \lambda_1 & \cdots \\ \mathbf{0} & \boldsymbol{Q}_2\boldsymbol{T}_2 \end{pmatrix} \\ &= \boldsymbol{Q}_1\begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Q}_2 \end{pmatrix}\begin{pmatrix} \lambda_1 & \cdots \\ \mathbf{0} & \boldsymbol{T}_2 \end{pmatrix} = \boldsymbol{QT} \end{aligned}$$
where $\boldsymbol{T}$ is upper triangular. So $\boldsymbol{AQ} = \boldsymbol{QT}$, or $\boldsymbol{A} = \boldsymbol{QTQ}^{\mathrm{T}}$.