Matrix basics

A matrix is an array of numbers. $A \in \mathbb{R}^{m\times n}$ means that:

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \qquad (m \text{ rows and } n \text{ columns})$$

Two matrices can be multiplied if inner dimensions agree:

$$\underset{(m\times p)}{C} = \underset{(m\times n)}{A}\underset{(n\times p)}{B} \quad \text{where} \quad c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

Transpose: The transpose operator A^T swaps rows and columns. If $A \in \mathbb{R}^{m\times n}$ then $A^\mathsf{T} \in \mathbb{R}^{n\times m}$ and $(A^\mathsf{T})_{ij} = A_{ji}$.

- $(A^\mathsf{T})^\mathsf{T} = A$.
- $(AB)^\mathsf{T} = B^\mathsf{T} A^\mathsf{T}$.

Matrix basics (cont'd)

Vector products. If $x, y \in \mathbb{R}^n$ are column vectors,

- The **inner product** is $x^\mathsf{T} y \in \mathbb{R}$ (a.k.a. dot product)
- The **outer product** is $xy^\mathsf{T} \in \mathbb{R}^{n \times n}$.

These are just ordinary matrix multiplications!

Inverse. Let $A \in \mathbb{R}^{n \times n}$ (square). If there exists $B \in \mathbb{R}^{n \times n}$ with $AB = I$ or $BA = I$ (if one holds, then the other holds with the same B) then B is called the *inverse* of A, denoted $B = A^{-1}$.

Some properties of the matrix inverse:

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

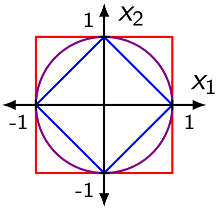
Vector norms

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)
- $\|x + y\| \leq \|x\| + \|y\|$ (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)
- $\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$ (max-norm)



norm ball: $\{x \mid \|x\| = 1\}$

Properties of the 2-norm (Euclidean norm)

- If you see $\|x\|$, think $\|x\|_2$ (it's the default)
- $x^\mathsf{T} x = \|x\|^2$
- $x^\mathsf{T} y \leq \|x\|\|y\|$ (Cauchy-Schwarz inequality)

Linear independence

A set of vectors $\{x_1, \ldots, x_n\} \in \mathbb{R}^m$ is **linearly independent** if

$$c_1 x_1 + \cdots + c_n x_n = 0 \quad \text{if and only if} \quad c_1 = \cdots = c_n = 0$$

If we define the matrix $A = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \in \mathbb{R}^{m \times n}$ then the columns of A are linearly independent if

$$Aw = 0 \quad \text{if and only if} \quad w = 0$$

If the vectors are not linearly independent, then they are **linearly dependent**. In this case, at least one of the vectors is redundant (can be expressed as a linear combination of the others). i.e. there exists a k and real numbers c_i such that

$$x_k = \sum_{i \neq k} c_i x_i$$

The rank of a matrix

$$\begin{aligned}\mathrm{rank}(A) &= \text{maximum number of linearly independent columns}\\ &= \text{maximum number of linearly independent rows}\end{aligned}$$

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

- $\mathrm{rank}(A) \le \min(m, n)$
- $\mathrm{rank}(AB) \le \min(\mathrm{rank}(A), \mathrm{rank}(B)) \le \min(m, n, p)$
- 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:

- $\mathrm{rank}(A + B) \le \mathrm{rank}(A) + \mathrm{rank}(B)$
- $\mathrm{rank}(A) = \mathrm{rank}(A^\mathsf{T}) = \mathrm{rank}(A^\mathsf{T}A) = \mathrm{rank}(AA^\mathsf{T})$
- $A \in \mathbb{R}^{n\times n}$ is invertible if and only if $\mathrm{rank}(A) = n$.

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 $x \in \mathbb{R}^n$. Three possibilities:

- No solutions. Example: $x_1 + x_2 = 1$ and $x_1 + x_2 = 0$.
- Exactly one solution. Example: $x_1 = 1$ and $x_2 = 0$.
- Infinitely many solutions. Example: $x_1 + x_2 = 0$.

Two common cases:

- **Overdetermined**: $m > n$. Typically no solutions. One approach is **least-squares**; find x to minimize $\|Ax - b\|_2$.
- **Underdetermined**: $m < n$. Typically infinitely many solutions. One approach is **regularization**; find the solution to $Ax = b$ such that $\|x\|_2$ is as small as possible.

Least squares

When the linear equations $Ax = b$ are overdetermined and there is no solution, one approach is to find an x that *almost* works by minimizing the 2-norm of the residual:

$$\underset{x}{\text{minimize}} \, \|Ax - b\|_2 \tag{1}$$

This problem always has a solution (not necessarily unique). $\hat{x}$ minimizes (1) iff $\hat{x}$ satisfies the **normal equations**:

$$A^\mathsf{T} A \hat{x} = A^\mathsf{T} b$$

The normal equations (and therefore (1)) have a unique solution iff the columns of A are linearly independent. Then,

$$\hat{x} = (A^\mathsf{T} A)^{-1} A^\mathsf{T} b$$

Range and nullspace

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

Range: $R(A) = \{Ax \mid x \in \mathbb{R}^n\}$. Note: $R(A) \subseteq \mathbb{R}^m$
Nullspace: $N(A) = \{x \in \mathbb{R}^n \mid Ax = 0\}$. Note: $N(A) \subseteq \mathbb{R}^n$

The following statements are equivalent:

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

The following statements are equivalent:

- Solutions to the equation $Ax = b$ are **unique**
- $N(A) = \{0\}$
- $\text{rank}(A) = n$ Remember: $\text{rank}(A) = \dim(R(A))$

Theorem: $\text{rank}(A) + \dim(N(A)) = n$

Orthogonal matrices

A matrix $U \in \mathbb{R}^{m \times n}$ is **orthogonal** if $U^\mathsf{T} U = I$. Note that we must have $m \geq n$. Some properties of orthogonal U and V:

- Columns are orthonormal: $u_i^\mathsf{T} u_j = \delta_{ij}$.
- Orthogonal transformations preserve angles & distances: $(Ux)^\mathsf{T}(Uz) = x^\mathsf{T} z$ and $\|Ux\|_2 = \|x\|_2$.
- Certain matrix norms are also invariant: $\|UAV^\mathsf{T}\|_2 = \|A\|_2$ and $\|UAV^\mathsf{T}\|_F = \|A\|_F$
- If U is square, $U^\mathsf{T} U = UU^\mathsf{T} = I$ and $U^{-1} = U^\mathsf{T}$.
- 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 $r = \operatorname{rank}(A)$. One way to find U is using **Gram-Schmidt**.

Projections

If $P \in \mathbb{R}^{n\times n}$ satisfies $P^2 = P$, 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:

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

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 $P = A(A^\mathsf{T} A)^{-1}A^\mathsf{T}$.

Least-squares: decompose b using the projection above:

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

Where $\hat{x} = (A^\mathsf{T} A)^{-1}A^\mathsf{T} b$ is the LS estimate. Therefore the optimal residual is orthogonal to $A\hat{x}$.

The singular value decomposition

Every $A \in \mathbb{R}^{m \times n}$ can be factored as

$$\underset{(m\times n)}{A} = \underset{(m\times r)}{U_1}\underset{(r\times r)}{\Sigma_1}\underset{(n\times r)^\mathsf{T}}{V_1^\mathsf{T}} \qquad \text{(economy SVD)}$$

U_1 is orthogonal, its columns are the *left singular vectors*
V_1 is orthogonal, its columns are the *right singular vectors*
Σ_1 is diagonal. $\sigma_1 \geq \cdots \geq \sigma_r > 0$ are the **singular values**

Complete the orthogonal matrices so they become square:

$$\underset{(m\times n)}{A} = \underset{(m\times m)}{\begin{bmatrix} U_1 & U_1 \end{bmatrix}} \underset{(m\times n)}{\begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix}} \underset{(n\times n)}{\begin{bmatrix} V_1^\mathsf{T} \\ V_2^\mathsf{T} \end{bmatrix}} = U\Sigma V^\mathsf{T} \quad \text{(full SVD)}$$

The singular values σ_i are an intrinsic property of A.
(the SVD is not unique, but every SVD of A has the same Σ_1)

Properties of the SVD

Singular vectors u_i, v_i and singular values σ_i satisfy

$$Av_i = \sigma_i u_i \qquad \text{and} \qquad A^\mathsf{T} u_i = \sigma_i v_i$$

Suppose $A = U\Sigma V^\mathsf{T}$ (full SVD) as in previous slide.

- rank: $\operatorname{rank}(A) = r$
- transpose: $A^\mathsf{T} = V\Sigma U^\mathsf{T}$
- pseudoinverse: $A^\dagger = V_1 \Sigma_1^{-1} U_1^\mathsf{T}$

Fundamental subspaces:

- $R(U_1) = R(A)$ and $R(U_2) = R(A)^\perp$ (range of A)
- $R(V_1) = N(A)^\perp$ and $R(V_2) = N(A)$ (nullspace of A)

Matrix norms:

- $\|A\|_2 = \sigma_1$ and $\|A\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2}$