Mathematical background

Inner product of vectors

standard inner product on $\mathbf{R}^n$

$x, y \in \mathbf{R}^n$ ⇒ $$\boxed{\langle x, y\rangle = x^T y = \sum_{i=1}^{n} x_i y_i,}$$

Euclidean norm, or ℓ_2-norm, of a vector $x \in \mathbf{R}^n$

$$\boxed{\|x\|_2 = (x^T x)^{1/2} = (x_1^2 + \cdots + x_n^2)^{1/2}.}$$

$$\boxed{|x^T y| \le \|x\|_2 \|y\|_2 \text{ for any } x, y \in \mathbf{R}^n}$$

Angle between vectors

$$x, y \in \mathbf{R}^n \quad \Rightarrow \quad \angle(x, y) = \cos^{-1}\left(\frac{x^T y}{||x||_2 ||y||_2}\right)$$

x and y are *orthogonal* if $x^T y = 0$.

Inner product of matrices

trace of a matrix: the sum of its diagonal elements.

standard inner product on $\mathbf{R}^{m\times n}$

$X, Y \in \mathbf{R}^{m\times n}$

$$\boxed{\langle X, Y\rangle = \mathbf{tr}(X^T Y) = \sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij}Y_{ij},}$$

Frobenius norm of a matrix $X \in \mathbf{R}^{m\times n}$

$$\boxed{\|X\|_F = \left(\mathbf{tr}(X^T X)\right)^{1/2} = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij}^2\right)^{1/2}.}$$

Norms

A function $f : \mathbf{R}^n \to \mathbf{R}$ with $\mathbf{dom}\, f = \mathbf{R}^n$ is called a *norm* if

- f is nonnegative: $f(x) \geq 0$ for all $x \in \mathbf{R}^n$
- f is definite: $f(x) = 0$ only if $x = 0$
- f is homogeneous: $f(tx) = |t| f(x)$, for all $x \in \mathbf{R}^n$ and $t \in \mathbf{R}$
- f satisfies the triangle inequality: $f(x + y) \leq f(x) + f(y)$, for all $x, y \in \mathbf{R}^n$

$f(x) \;=\; \|x\|$ $\qquad \|x\|_{\mathrm{symb}}$ $\qquad \mathbf{dist}(x, y) = \|x - y\|$

$\boxed{\textit{unit ball} \text{ of the norm } \|\cdot\|}$

$$\mathcal{B} = \{x \in \mathbf{R}^n \mid \|x\| \leq 1\},$$

Norms (vectors)

$\boxed{\textit{sum-absolute-value}, \text{ or } \ell_1\textit{-norm},}$

$$\|x\|_1 = |x_1| + \cdots + |x_n|,$$

$\boxed{\textit{Chebyshev} \text{ or } \ell_\infty\textit{-norm},}$

$$\|x\|_\infty = \max\{|x_1|, \ldots, |x_n|\}.$$

$\boxed{\ell_p\textit{-norm}}$ A family of norms $p \geq 1$

$$\|x\|_p = (|x_1|^p + \cdots + |x_n|^p)^{1/p}$$

Norms (vectors)

$\boxed{quadratic\ norms.}$

Another important family of norms

$P \in \mathbf{S}^n_{++}$ define the P-quadratic norm as

$$||x||_P = (x^T P x)^{1/2}$$

Norms (Matrices)

common norms on $\mathbf{R}^{m \times n}$

$\boxed{\text{Frobenius norm}}$

$\boxed{\text{sum-absolute-value norm}}$

$$\|X\|_{\mathrm{sav}} = \sum_{i=1}^{m} \sum_{j=1}^{n} |X_{ij}|$$

$\boxed{\text{maximum-absolute-value norm}}$

$$\|X\|_{\mathrm{mav}} = \max\{|X_{ij}| \mid i = 1, \ldots, m, \ j = 1, \ldots, n\}$$

Norms (Matrices)

max- column-sum norm

$$\|X\|_1 = \max_{j=1,\ldots,n} \sum_{i=1}^{m} |X_{ij}|$$

max- row-sum norm

$$\|X\|_\infty = \max_{i=1,\ldots,m} \sum_{j=1}^{n} |X_{ij}|$$

Norms (Matrices)

L2 norm

$$||X||_2 = \sigma_{\max}(X) = (\lambda_{\max}(X^T X))^{1/2}$$

Functions

Function notation and continuity

$$f : A \to B$$

$$f : \mathbf{R}^n \to \mathbf{R}^m$$

f maps (some) n-vectors into m-vectors

Continuity

whenever the sequence $x_1, x_2, \ldots$

in $\mathbf{dom}\, f$ converges to a point $x \in \mathbf{dom}\, f$, the sequence $f(x_1), f(x_2), \ldots$ converges to $f(x)$

A function f is *continuous* if it is continuous at every point in its domain

Derivatives

$f : \mathbf{R}^n \to \mathbf{R}^m$

$\boxed{\textit{derivative} \text{ (or } \textit{Jacobian}\text{) of } f \text{ at } x}$

is the matrix $Df(x) \in \mathbf{R}^{m \times n}$, given by

$$Df(x)_{ij} = \frac{\partial f_i(x)}{\partial x_j}, \qquad i = 1, \ldots, m, \quad j = 1, \ldots, n,$$

If the partial derivatives exist, we say f is differentiable at x.

Derivatives

Gradient

f is real-valued

$$f : \mathbf{R}^n \to \mathbf{R}$$

derivative $Df(x)$ is a $1 \times n$ matrix

gradient of the function:

$$\boxed{\nabla f(x) = Df(x)^T}$$

$$\nabla f(x)_i = \frac{\partial f(x)}{\partial x_i}, \quad i = 1, \ldots, n.$$

The first-order approximation of f at a point x

$$f(x) + \nabla f(x)^T (z - x)$$

Derivatives

Example:

quadratic function

$$f : \mathbf{R}^n \to \mathbf{R}$$

$$f(x) = (1/2)x^T P x + q^T x + r$$

$P \in \mathbf{S}^n$, $q \in \mathbf{R}^n$, and $r \in \mathbf{R}$

$$Df(x) = x^T P + q^T$$

$$\nabla f(x) = Px + q$$

Derivatives

Chain rule

$f : \mathbf{R}^n \to \mathbf{R}^m$ is differentiable at x

$g : \mathbf{R}^m \to \mathbf{R}^p$ is differentiable at $f(x)$

$h : \mathbf{R}^n \to \mathbf{R}^p$ $\qquad h(z) = g(f(z))$

$$\boxed{Dh(x) = Dg(f(x))Df(x)}$$

Special case:

$f : \mathbf{R}^n \to \mathbf{R}$, $g : \mathbf{R} \to \mathbf{R}$, and $h(x) = g(f(x))$

$$\nabla h(x) = g'(f(x))\nabla f(x)$$

Second derivative

$$f : \mathbf{R}^n \to \mathbf{R}$$

second derivative or Hessian matrix $\nabla^2 f(x)$

$$\boxed{\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \qquad i = 1, \ldots n, \quad j = 1, \ldots, n,}$$

If f is differentiable gradient mapping is the function

$$\nabla f : \mathbf{R}^n \to \mathbf{R}^n$$

The derivative of this mapping is

$$D\nabla f(x) = \nabla^2 f(x)$$

Second derivative

$$f(x) + \nabla f(x)^T(z-x)$$

second-order approximation of f, at or near x

$$\widehat{f}(z) = f(x) + \nabla f(x)^T(z-x) + (1/2)(z-x)^T\nabla^2 f(x)(z-x)$$

Example:

$$f(x) = (1/2)x^TPx + q^Tx + r$$

$P \in \mathbf{S}^n$, $q \in \mathbf{R}^n$, and $r \in \mathbf{R}$

$$\nabla f(x) = Px + q$$

$$\nabla^2 f(x) = P$$

Second derivative

Chain rule for second derivative

$f : \mathbf{R}^n \to \mathbf{R}$, $g : \mathbf{R} \to \mathbf{R}$, and $h(x) = g(f(x))$

$$\nabla^2 h(x) = g'(f(x))\nabla^2 f(x) + g''(f(x))\nabla f(x)\nabla f(x)^T$$

Linear Algebra

Vector space

Linear Independence of vectors

$$v_1, v_2, \cdots, v_n \Leftrightarrow \sum_{i=1}^{n} \alpha_i v_i = 0 \Rightarrow \alpha_i = 0$$

Vector space

The set V is a vector space if

$$for\ \forall x, y \in V\ and\ \forall \alpha, \beta \in R \Rightarrow \alpha x + \beta y \in V$$

Base of a vector space

$\{v_1, v_2, \cdots, v_k\} \subseteq V$ is a base for vector space V if

- $v_1, v_2, \cdots, v_k$ are linear independent
- $\forall x \in V, \exists \alpha_1, \alpha_2, \cdots, \alpha_k \in R \Rightarrow x = \sum_{i=1}^{k} \alpha_i v_i$

$\Rightarrow$ **dim V =k**

Rank and range of matrices

Rank of A

$A \in \mathbf{R}^{m \times n}$

The number of independent columns or the number of independent rows

The rank of A can never be greater than the minimum of m and n.

A has full rank if rank A = min{m, n}.

Range of A

$$\boxed{\mathcal{R}(A) = \{Ax \mid x \in \mathbf{R}^n\}}$$

The range $\mathcal{R}(A)$ is a subspace of $\mathbf{R}^m$

Its dimension is the rank of A

Kernel of matrices

Nullspace or kernel of a matrix

The *nullspace* (or *kernel*) of A, denoted $\mathcal{N}(A)$, is the set of all vectors x mapped into zero by A:

$$\mathcal{N}(A) = \{x \mid Ax = 0\}.$$

The nullspace is a subspace of $\mathbf{R}^n$

Eigenvalue and vector

λ is an eigenvalue of matrix A if there exists a non-zero vector x such that

$A \in R^{n\times n}$

$$Ax = \lambda x$$

Eigenvectors are not unique

Eigenvalues can be real or complex

the roots of the characteristic polynomial

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

All eigenvalues of symmetric matrices are real

$for\ \lambda_i \neq \lambda_j, their\ corresponding\ eigenvectors\ x_i\ and\ x_j\ are\ orthogonal$

Eigenvalue and vector

$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ $\quad$ $\lambda_1(A) = \lambda_{\max}(A)$ $\quad$ $\lambda_n(A) = \lambda_{\min}(A)$

$$\boxed{\det A = \prod_{i=1}^{n} \lambda_i}$$

$$\boxed{\mathbf{tr}\, A = \sum_{i=1}^{n} \lambda_i}$$

Symmetric eigenvalue decomposition

$A \in \mathbf{S}^n$, *i.e.*, A is a real symmetric $n \times n$ matrix

$$\boxed{A = Q\Lambda Q^T}$$

$$\Lambda = \mathbf{diag}(\lambda_1, \dots, \lambda_n)$$

The columns of Q form an orthonormal set of eigenvectors of A. $Q^T Q = I$

spectral decomposition or eigenvalue decomposition of A.

Singular value decomposition(SVD)

$A \in \mathbf{R}^{m \times n}$ with $\mathbf{rank}\, A = r$

$$\boxed{A = U\Sigma V^T}$$

$U \in \mathbf{R}^{m \times r}$ satisfies $U^T U = I$, $V \in \mathbf{R}^{n \times r}$ satisfies $V^T V = I$

$\Sigma = \mathbf{diag}(\sigma_1, \ldots, \sigma_r)$ $\qquad \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$

- ✓columns of U are called left singular vectors of A
- ✓columns of V are right singular vectors
- ✓the numbers σi are the singular values.

SVD

singular value decomposition of a matrix A is closely related to the eigenvalue decomposition of A^TA.

$$A = U\Sigma V^T \quad \Rightarrow \quad A^TA = V\Sigma^2V^T = \begin{bmatrix} V & \tilde{V} \end{bmatrix} \begin{bmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V & \tilde{V} \end{bmatrix}^T$$

where $\tilde{V}$ is any matrix for which $[V \;\; \tilde{V}]$ is orthogonal

$\boxed{\text{eigenvalue decomposition of } A^TA}$

- ✓its nonzero eigenvalues are the singular values of A squared
- ✓associated eigenvectors of A^TA are the right singular vectors of A.
- ✓A similar analysis of AA^T