Linear Algebra and Calculus refresher

General notations

Definitions

Vector — We note $x \in \mathbb{R}^n$ a vector with n entries, where $x_i \in \mathbb{R}$ is the i^{th} entry:

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$$

Matrix — We note $A \in \mathbb{R}^{m \times n}$ a matrix with n rows and m, where $A_{i,j} \in \mathbb{R}$ is the entry located in the i^{th} row and j^{th} column:

$$A = \begin{pmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & & \vdots \\ A_{m,1} & \cdots & A_{m,n} \end{pmatrix} \in \mathbb{R}^{m \times n}$$

Remark: the vector x defined above can be viewed as a $n \times 1$ matrix and is more particularly called a column-vector.

Main matrices

Identity matrix — The identity matrix $I \in \mathbb{R}^{n \times n}$ is a square matrix with ones in its diagonal and zero everywhere else:

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

Remark: for all matrices $A\in\mathbb{R}^{n\times n}$, we have $A\times I=I\times A=A$.

Diagonal matrix — A diagonal matrix $D\in\mathbb{R}^{n\times n}$ is a square matrix with nonzero values in its diagonal and zero everywhere else:

$$D=\begin{pmatrix}d_1&0&\cdots&0\\0&\ddots&\ddots&\vdots\\\vdots&\ddots&\ddots&0\\0&\cdots&0&d_n\end{pmatrix}$$

Remark: we also note D as $\textrm{diag}(d_1,...,d_n)$.

Matrix operations

Multiplication

Vector-vector — There are two types of vector-vector products:

- inner product: for $x,y\in\mathbb{R}^n$, we have:

$$\boxed{x^Ty=\sum_{i=1}^nx_iy_i\in\mathbb{R}}$$

- outer product: for $x\in\mathbb{R}^m,y\in\mathbb{R}^n$, we have:

$$\boxed{xy^T=\begin{pmatrix}x_1y_1&\cdots&x_1y_n\\\vdots&&\vdots\\x_my_1&\cdots&x_my_n\end{pmatrix}\in\mathbb{R}^{m\times n}}$$

Matrix-vector — The product of matrix $A \in \mathbb{R}^{m \times n}$ and vector $x \in \mathbb{R}^n$ is a vector of size $\mathbb{R}^n$, such that:

$$\boxed{Ax = \begin{pmatrix} a_{r,1}^T x \\ \vdots \\ a_{r,m}^T x \end{pmatrix} = \sum_{i=1}^{n} a_{c,i} x_i \in \mathbb{R}^n}$$

where $a_{r,i}^T$ are the vector rows and $a_{c,j}$ are the vector columns of A, and x_i are the entries of x.

Matrix-matrix — The product of matrices $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$ is a matrix of size $\mathbb{R}^{n \times p}$, such that:

$$\boxed{AB = \begin{pmatrix} a_{r,1}^T b_{c,1} & \cdots & a_{r,1}^T b_{c,p} \\ \vdots & & \vdots \\ a_{r,m}^T b_{c,1} & \cdots & a_{r,m}^T b_{c,p} \end{pmatrix} = \sum_{i=1}^{n} a_{c,i} b_{r,i}^T \in \mathbb{R}^{n \times p}}$$

where $a_{r,i}^T, b_{r,i}^T$ are the vector rows and $a_{c,j}, b_{c,j}$ are the vector columns of A and B respectively.

Other operations

Transpose — The transpose of a matrix $A \in \mathbb{R}^{m \times n}$, noted A^T, is such that its entries are flipped:

$$\boxed{\forall i, j, \qquad A_{i,j}^T = A_{j,i}}$$

Remark: for matrices A, B, we have $(AB)^T = B^T A^T$.

Inverse — The inverse of an invertible square matrix A is noted A^{-1} and is the only matrix such that:

$$\boxed{AA^{-1} = A^{-1}A = I}$$

Remark: not all square matrices are invertible. Also, for matrices A, B, we have $(AB)^{-1} = B^{-1}A^{-1}$

Trace — The trace of a square matrix A, noted $\textrm{tr}(A)$, is the sum of its diagonal entries:

$$\boxed{\textrm{tr}(A) = \sum_{i=1}^{n} A_{i,i}}$$

Remark: for matrices A, B, we have $\textrm{tr}(A^T) = \textrm{tr}(A)$ and $\textrm{tr}(AB) = \textrm{tr}(BA)$

Determinant — The determinant of a square matrix $A \in \mathbb{R}^{n \times n}$, noted $|A|$ or $\textrm{det}(A)$ is expressed recursively in terms of $A_{\backslash i, \backslash j}$, which is the matrix A without its i^{th} row and j^{th} column, as follows:

$$\boxed{\textrm{det}(A) = |A| = \sum_{j=1}^{n} (-1)^{i+j} A_{i,j} |A_{\backslash i, \backslash j}|}$$

Remark: A is invertible if and only if $|A| \neq 0$. Also, $|AB| = |A||B|$ and $|A^T| = |A|$.

Matrix properties

Definitions

Symmetric decomposition — A given matrix A can be expressed in terms of its symmetric and antisymmetric parts as follows:

$$\boxed{A = \underbrace{\frac{A + A^T}{2}}_{\textrm{Symmetric}} + \underbrace{\frac{A - A^T}{2}}_{\textrm{Antisymmetric}}}$$

Norm — A norm is a function $N : V \longrightarrow [0, +\infty[$ where V is a vector space, and such that for all $x, y \in V$, we have:

- $N(x+y) \leqslant N(x)+N(y)$
- $N(ax)=|a|N(x)$ for a scalar
- if $N(x)=0$, then $x=0$

For $x \in V$, the most commonly used norms are summed up in the table below:

Norm	Notation	Definition	Use case
Manhattan, L^1	$\|\|x\|\|_1$	$\displaystyle\sum_{i=1}^n \|x_i\|$	LASSO regularization
Euclidean, L^2	$\|\|x\|\|_2$	$\displaystyle\sqrt{\sum_{i=1}^n x_i^2}$	Ridge regularization
p-norm, L^p	$\|\|x\|\|_p$	$\displaystyle\left(\sum_{i=1}^n x_i^p\right)^{\frac{1}{p}}$	Hölder inequality
Infinity, L^∞	$\|\|x\|\|_\infty$	$\underset{i}{\max}\ \|x_i\|$	Uniform convergence

Linearly dependence — A set of vectors is said to be linearly dependent if one of the vectors in the set can be defined as a linear combination of the others.

Remark: if no vector can be written this way, then the vectors are said to be linearly independent.

Matrix rank — The rank of a given matrix A is noted $\textrm{rank}(A)$ and is the dimension of the vector space generated by its columns. This is equivalent to the maximum number of linearly independent columns of A.

Positive semi-definite matrix — A matrix $A \in \mathbb{R}^{n\times n}$ is positive semi-definite (PSD) and is noted $A \succeq 0$ if we have:

$$\boxed{A=A^T}\quad \textrm{and}\quad \boxed{\forall x\in\mathbb{R}^n,\quad x^TAx\geqslant 0}$$

Remark: similarly, a matrix A is said to be positive definite, and is noted $A \succ 0$, if it is a PSD matrix which satisfies for all non-zero vector x, $x^TAx > 0$.

Eigenvalue, eigenvector — Given a matrix $A \in \mathbb{R}^{n \times n}$, λ is said to be an eigenvalue of A if there exists a vector $z \in \mathbb{R}^n \backslash \{0\}$, called eigenvector, such that we have:

$$\boxed{Az = \lambda z}$$

Spectral theorem — Let $A \in \mathbb{R}^{n \times n}$. If A is symmetric, then A is diagonalizable by a real orthogonal matrix $U \in \mathbb{R}^{n \times n}$. By noting $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, we have:

$$\boxed{\exists \Lambda \text{ diagonal}, \quad A = U \Lambda U^T}$$

Singular-value decomposition — For a given matrix A of dimensions $m \times n$, the singular-value decomposition (SVD) is a factorization technique that guarantees the existence of U $m \times m$ unitary, Σ $m \times n$ diagonal and V $n \times n$ unitary matrices, such that:

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

Matrix calculus

Gradient — Let $f : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ be a function and $A \in \mathbb{R}^{m \times n}$ be a matrix. The gradient of f with respect to A is a $m \times n$ matrix, noted $\nabla_A f(A)$, such that:

$$\boxed{\Big(\nabla_A f(A)\Big)_{i,j} = \frac{\partial f(A)}{\partial A_{i,j}}}$$

Remark: the gradient of f is only defined when f is a function that returns a scalar.

Hessian — Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a function and $x \in \mathbb{R}^n$ be a vector. The hessian of f with respect to x is a $n \times n$ symmetric matrix, noted $\nabla_x^2 f(x)$, such that:

$$\boxed{\Big(\nabla_x^2 f(x)\Big)_{i,j} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}}$$

Remark: the hessian of f is only defined when f is a function that returns a scalar.

Gradient operations — For matrices A, B, C, the following gradient properties are worth having in mind:

$$\boxed{\nabla_A \textrm{tr}(AB) = B^T} \qquad \boxed{\nabla_{A^T} f(A) = \left(\nabla_A f(A)\right)^T}$$

$$\boxed{\nabla_A \textrm{tr}(ABA^TC) = CAB + C^TAB^T} \qquad \boxed{\nabla_A |A| = |A|(A^{-1})^T}$$

(https://twitter.com/shervinea) (https://linkedin.com/in/shervineamidi) (https://github.com/shervinea) (https://scholar.google.com/citations?user=nMnMTm8AAAAJ)