Matrix Operations

In what follows, $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times p}$ and $x, y \in \mathbb{R}^n$.

1. Compute the following, giving conditions on m, n, p for the formulas to be well-defined. When the result is a matrix, give its value at position (i, j). When it is a vector, give its value at position i.

- AB
- $x^T A$
- $x^T y$
- xy^T
- $y^T Ax$
- $Tr(A^T A)$

2. Show that

- $(AB)^T = B^T A^T$
- $(AB)^{-1} = B^{-1}A^{-1}$
- $(A^T)^{-1} = (A^{-1})^T$
- The *trace* operator $A \to Tr(A)$ is a *linear operator* on $\mathbb{R}^{m\times n}$.
- $Tr(AB) = Tr(BA)$.

3. Given a matrix $A \in \mathbb{R}^{n\times n}$ and a vector $r \in \mathbb{R}^n$, how to multiply each row of A by the elements of r?

4. Given $x_i \in \mathbb{R}^d$, $i = 1, ..., n$, how to compute the mean of $\{||x_i - x_j||_2;\ i, j = 1, ..., n\}$ in Python, without a loop? *Solution:* `np.sqrt(((X[None,:]-X[:,None])**2).sum(-1)).mean()` *with* $X = [x_1, ..., x_n]^T \in \mathbb{R}^{n\times d}$.

Gradients

1. Let $A \in \mathbb{R}^{m\times n}$. In the following, give the function's output dimension and compute the gradient:

- $\nabla_x\ (xA), \quad x \in \mathbb{R}$
- $\nabla_{\mathrm{x}}\ (A\mathrm{x}), \quad \mathrm{x} \in \mathbb{R}^n$
- $\nabla_{\mathrm{x}}\ (\mathrm{y}^T A\mathrm{x})$, $\nabla_{\mathrm{y}}\ (\mathrm{y}^T A\mathrm{x}), \quad \mathrm{x}, \mathrm{y} \in \mathbb{R}^n$
- $\nabla_{\mathrm{x}}\ (\mathrm{x}^T A\mathrm{x}), \quad \mathrm{x} \in \mathbb{R}^n$ *Solution*: $(x+h)^T A(x+h) = x^T Ax + h^T Ax + x^T Ah + h^T Ah$.* *Since the scalar product is symmetric,* $h^T Ax = (Ax)^T h = x^T A^T h$. *Therefore,* $(\nabla f)^T = x^T(A^T + A)$, *and* $\nabla f = (A + A^T)x$.
- $\nabla_{\mathrm{x}}\ ||\mathrm{y} - A\mathrm{x}||_2^2, \quad \mathrm{x} \in \mathbb{R}^n$ *Solution:* $||y - A(x+h)||_2^2 = ||y - Ax||_2^2 - 2(y - Ax)^T(Ah) + ||Ah||_2^2$. *Therefore* $(\nabla f)^T = -2(y - Ax)^T A$, *and* $\nabla f = -2A^T(y - Ax)$.
- $\nabla_X\ Tr(A^T X), \quad X \in \mathbb{R}^{m\times n}$ *Solution:* $Tr(A^T X) = \sum_{i,j} A_{ij}X_{ij}$, *hence* $\dfrac{\partial Tr(A^T X)}{\partial X_{kl}} = A_{kl}$, *meaning that* $\nabla_X\, Tr(A^T X) = A$.

2. Compute the following gradients using the chain rule:

- $\nabla_z \, Ax(z)$
- $\nabla_z \, y(z)^T Ax(z)$

Symmetric matrices

Reminders

Def (Eigenvector Equation). Let $A \in \mathbb{R}^{n \times n}$. If $Au = \lambda u$ with $\lambda \in \mathbb{R}$ and $u \in \mathbb{R}^n$, then λ is called an *eigenvalue* of A and u an *eigenvector*.

Def (Orthogonal matrix). $U \in \mathbb{R}^n$ is orthogonal $\iff$ U is inversible and $U^{-1} = U^T$.

Spectral theorem. Any (real) symmetric matrix $A \in \mathbb{R}^{n \times n}$ can be decomposed as

$$A = U \Lambda U^T$$

where U is an orthogonal matrix and $\lambda = \mathrm{diag}(\lambda_1, ..., \lambda_n)$.

Exercises

1. Show that for an orthogonal matrix U,

$$u_i^T u_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

Solution: $I = U^T U = [u_i^T u_j]_{i,j}$.

2. Let $A \in \mathbb{R}^{n \times n}$ a (real) symmetric matrix and $A := U \Lambda U^T$ with U orthogonal and $\Lambda = \mathrm{diag}(\lambda_1, ..., \lambda_n)$. Show that the columns of U are eigenvectors of A and $\lambda_i, i = 1, ..., n$ are eigenvalues. *Solution:* $(U^T u_i)_j = \delta_{ij}$, *hence* $(\Lambda U^T u_i)_j = \lambda_i \delta_{ij}$ *and* $U \Lambda U^T u_i = \lambda_i u_i$.

3. Let $A = U \Lambda U^T$ a symmetric matrix. Show that:

a. A is inversible $\iff$ $\lambda_i \neq 0, \forall i$ *Solution:* $det(A) = det(U \Lambda U^T) = det(U^T U \Lambda) = det(\Lambda) = \prod_i \lambda_i$.
b. If $\lambda_i \neq 0, \forall i$, then $A^{-1} = U \Lambda^{-1} U^T$ with $\Lambda^{-1} = \mathrm{diag}(\lambda_1^{-1}, ..., \lambda_n^{-1})$ *Solution:* $(U \Lambda^{-1} U^T)(U \Lambda U^T) = U \Lambda^{-1} \Lambda U^T = U U^T = I$.
c. $\mathrm{Tr}(A) = \sum_i \lambda_i$, $\mathrm{Tr}(A^{-1}) = \sum_i \lambda_i^{-1}$ *Solution:* $Tr(A) = Tr(U \Lambda U^T) = Tr(U^T U \Lambda) = Tr(\Lambda) = \sum_i u_i$.

4. Let $\langle \cdot, \cdot \rangle$ be the standard scalar product on $\mathbb{R}^n$: $\langle a, b \rangle = a^T b$.

a. Show that for any square matrix A, and any $x, y \in \mathbb{R}^n$,

$$\langle Ax, y \rangle = \langle x, A^T y \rangle$$

b. Prove that an orthonormal matrix U preserves the inner product (hence preserves the *norms* and *angles*):

$$\forall a, b \in \mathbb{R}^n, \quad \langle Ua, Ub \rangle = \langle a, b \rangle$$

Positive semi-definte matrices

Reminders

Def (Positive (semi-)definite matrix). Let $A \in \mathbb{R}^{n \times n}$.

$$\begin{array}{lll} A \text{ p.s.d.} & \iff & x^T A x \geq 0, \forall x \in \mathbb{R}^n \\ A \text{ p.d.} & \iff & x^T A x > 0, \forall x \in \mathbb{R}^n,\ x \neq 0 \end{array}$$

Exercies

1. Let $A \in \mathbb{R}^{n \times n}$ symmetric. Show that

a. A p.s.d. $\iff$ A has n eigenvalues $\lambda_i \geq 0$ *Solution:* $\implies$ *: We assume A is p.s.d. Let λ s.t. $Au = \lambda u$. Then $0 \leq u^T A u = u^T.(\lambda u) = \lambda u^T u = \lambda$.* $\impliedby$ *: We assume that $\lambda_i \geq 0,\ \forall i$. For $x \in \mathbb{R}^n$, let $b = U^{-1}x$. Then $x^T A x = (Ub)^T U \Lambda U^T (Ub) = b^T \Lambda b = \sum_{i,j} \Lambda_{ij}^2 b_i b_j = \sum_i \Lambda_{ii}^2 b_i^2 \geq 0$*

b. A p.d. $\iff$ A has n eigenvalues $\lambda_i > 0$

2. Let $X \in \mathbb{R}^{n \times p}$, $\lambda > 0$. Prove that $M = X^T X + \lambda I_p$ is inversible. *Hint: Prove that the eigenvalues of M are larger than λ.*

3. For any general $X \in \mathbb{R}^{m \times n}$, show that

a. $X^T X$ and $X X^T$ are symmetric, p.s.d.

b. the non-zero eigenvalues of $X^T X$ and $X X^T$ are the same, that are $\{\sigma_i^2, \sigma_i \neq 0, i = 1, ..., \min(m, n)\}$ where σ_i's are singular values of X