

# MATH REVIEW

(A)

\*  $f: \mathbb{R}^n \rightarrow \mathbb{R}$

Gradient of  $f$  at  $x$ :  $\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$ , Hessian:  $\nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_i \partial x_j} \end{pmatrix}$   
( $n \times 1$  vector) (  $n \times n$  symmetric matrix)

\* Directional derivative along direction  $p \in \mathbb{R}^n$ :

$$\lim_{\varepsilon \rightarrow 0} \frac{f(x + \varepsilon p) - f(x)}{\varepsilon} = \nabla f(x)^T p \quad [\text{Pf: Taylor's th.}]$$

\* Mean value th.:

•  $\phi: \mathbb{R} \rightarrow \mathbb{R}$  continuously differentiable,  $\alpha_1 > \alpha_0$ :

$$\phi(\alpha_1) - \phi(\alpha_0) = \phi'(\xi)(\alpha_1 - \alpha_0) \quad \text{for some } \xi \in (\alpha_0, \alpha_1)$$

•  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  cont. diff., for any  $p \in \mathbb{R}^n$ :

$$f(x+p) - f(x) = \nabla f(x + \alpha p)^T p \quad \text{for some } \alpha \in (0, 1)$$

\* Taylor's th.:

•  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  cont. diff.,  $p \in \mathbb{R}^n$ . Then:

$$f(x+p) = f(x) + \nabla f(x + tp)^T p \quad \text{for some } t \in (0, 1) \quad [\text{Mean value th.}]$$

• If  $f$  is twice cont. diff. then:

$$f(x+p) = f(x) + \nabla f(x)^T p + \frac{1}{2} p^T \nabla^2 f(x + tp) p \quad \text{for some } t \in (0, 1).$$

\* Positive definite matrix  $B \Leftrightarrow p^T B p > 0 \quad \forall p \neq 0$  (pd)

positive semidefinite " if  $\geq 0$ . (psd)

\* Condition number of a nonsingular matrix  $A$ :  $\kappa(A) = \|A\| \|A^{-1}\|$   
where  $\|\cdot\|$  is any matrix norm. For  $\|\cdot\|_2 = |\text{largest eigenvalue of } A|$ ,

$$\kappa(A) = |\lambda_{\max}| / |\lambda_{\min}|.$$

Matrix norm induced by a vector norm:  $\|A\| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|}$ .

\* Linear independence; subspaces

•  $v_1, \dots, v_k \in \mathbb{R}^n$  are l.i. if  $\forall \lambda_1, \dots, \lambda_k \in \mathbb{R}: \lambda_1 v_1 + \dots + \lambda_k v_k = 0 \Rightarrow \lambda_1 = \dots = \lambda_k = 0$ .

• Vectors  $u, v \neq 0$  are orthogonal ( $u \perp v$ ) iff  $u^T v = 0$  ( $= \|u\| \|v\| \cos(u, v)$ )

Orthogonal matrix:  $U^{-1} = U^T$ . For the Euclidean norm:  $\|Ux\| = \|x\|$ .

•  $\text{span}(v_1, \dots, v_k) = \{x: x = \lambda_1 v_1 + \dots + \lambda_k v_k \text{ for } \lambda_1, \dots, \lambda_k \in \mathbb{R}\}$

is the set of all vectors that are linear combination of  $v_1, \dots, v_k$ , or the linear subspace spanned by  $v_1, \dots, v_k$ .

• If  $v_1, \dots, v_k$  are l.i. then they are a basis of  $\text{span}(v_1, \dots, v_k)$ , which has dimension  $k$ .

\* Null space of a matrix:  $\text{null}(A) = \{u: Au = 0\}$ , i.e., the subspace associated with eigenvalue 0.

Range space of a matrix:  $\text{range}(A) = \{u: u = Av \text{ for some vector } v\}$

\* Eigenvalues and eigenvectors of a <sup>real</sup> matrix:  $Au = \lambda u$   
 $\lambda \in \mathbb{C}$  (eigenvalue)  $u \in \mathbb{R}^n$  (eigenvector)

For all eigenvalues  $\lambda$

|                         |  |
|-------------------------|--|
| For symmetric matrices: | $\lambda \in \mathbb{R}$ ; eigenvectors of different eigenvalues are $\perp$ |
| For nonsingular "       | $\lambda \neq 0$   |
| For pd "                | $\lambda > 0$  |
| For psd "               | $\lambda \geq 0$   |
| For nd "                | $\lambda < 0$  |
| For nsd "               | $\lambda \leq 0$   |

\* Spectral theorem: A symmetric, real matrix with eigenvectors  $u_1, \dots, u_n \in \mathbb{R}^n$  associated with eigenvalues  $\lambda_1, \dots, \lambda_n \in \mathbb{R} \Rightarrow A = U \Lambda U^T = \sum_{i=1}^n \lambda_i u_i u_i^T$  where  $U = (u_1 \dots u_n)$  is orthogonal and  $\Lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$  diagonal.

In other words, a symmetric real matrix can be diagonalised in terms of its eigenvalues and eigenvectors.

spectrum of  $A =$  eigenvalues of  $A$ .

\* Sherman-Morrison-Woodbury formula: given  $A_{p \times p}$ ,  $B_{p \times q}$ ,  $C_{q \times p}$ ,  $D_{q \times q}$ , if  $A, C$  invertible:  $(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$ .

\* Order notation: consider  $f(n), g(n) \geq 0$  for  $n=1, 2, 3, \dots$

• Asymptotic upper bound  $O()$ :

$f$  is  $O(g)$  iff  $f(n) \leq c g(n)$  for  $c > 0$  and all  $n > n_0$

$f$  is of order  $g$  at most.

Ex:  $3n+5$  is  $O(n)$  and  $O(n^2)$  but not  $O(\log n)$  or  $O(\sqrt{n})$ .

• Asymptotic upper bound  $o()$ :

$f$  is  $o(g)$  iff  $\lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} = 0$

$f$  becomes insignificant relative to  $g$  as  $n$  grows

Ex:  $3n+5$  is  $o(n^2)$  and  $o(n^{1.3})$  but not  $o(n)$ .

• Asymptotic tight bound  $\Omega()$ :

$f$  is  $\Omega(g)$  iff  $c_0 g(n) \leq f(n) \leq c_1 g(n)$  for  $c_1, c_0 > 0$  and all  $n > n_0$ .

$\Leftrightarrow f$  is  $O(g)$  and  $g$  is  $O(f)$ .

Ex:  $3n+5$  is  $\Omega(n)$  but not  $\Omega(n^2)$ ,  $\Omega(\log n)$ ,  $\Omega(\sqrt{n})$ .

\* Cost of operations: assume  $n \times 1$  vectors and  $n \times n$  matrices.

• Space: vectors are  $O(n)$ , matrices are  $O(n^2)$ . Less if sparse or structured, eg. a diagonal matrix or a circulant matrix can be stored as  $O(n)$ .

• Time: we count scalar multiplications only.

- vector  $\times$  vector:  $O(n)$

- matrix  $\times$  vector:  $O(n^2)$

- matrix  $\times$  matrix:  $O(n^3)$

- eigenvalues and eigenvectors:  $O(n^3)$

- inversion and linear system solution:  $O(n^3)$

Less if sparse or structured, eg. matrix  $\times$  vector is  $O(n)$  for a diagonal matrix and  $O(n \log n)$  for a circulant matrix.

\* Cone: set  $F$  verifying:  $x \in F \Rightarrow \alpha x \in F \quad \forall \alpha > 0$ .

Ex.  $\left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : x_1 > 0, x_2 \geq 0 \right\}$ .

\* Subspaces of a real matrix  $A_{m \times n} \Leftrightarrow$  linear mapping  $\mathbb{R}^n \rightarrow \mathbb{R}^m$

Null space:  $\text{null}(A) = \{x \in \mathbb{R}^n : Ax = 0\}$

Range space:  $\text{range}(A) = \{y \in \mathbb{R}^m : y = Ax \text{ for some } x \in \mathbb{R}^n\}$

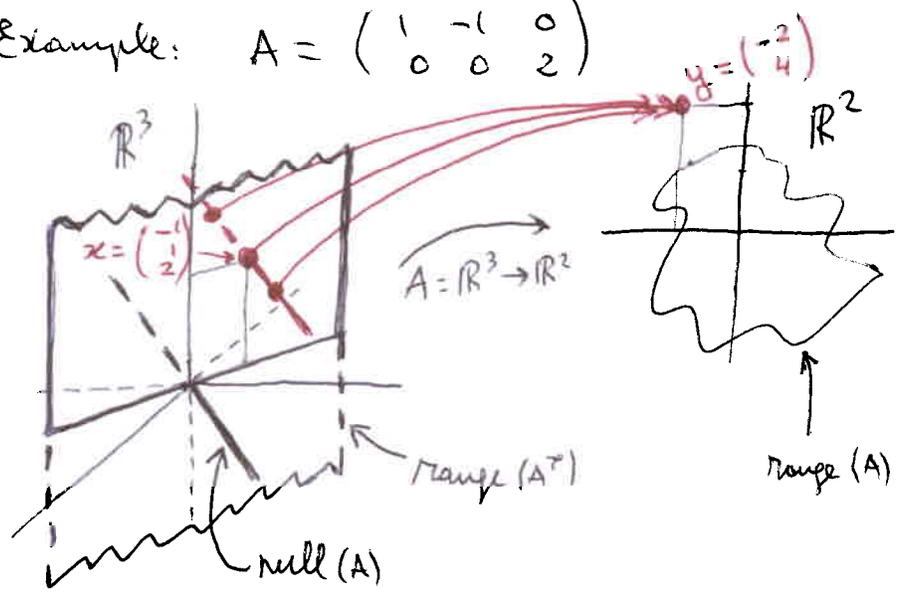
Fundamental theorem of linear algebra:  $\text{null}(A) \oplus \text{range}(A^T) = \mathbb{R}^n$

$\Leftrightarrow x \in \mathbb{R}^n \Rightarrow \exists! u, v \in \mathbb{R}^n : u \in \text{null}(A), v \in \text{range}(A^T), x = u + v$

$\text{null}(A) \cap \text{range}(A^T) = \{0\}, \text{null}(A) \perp \text{range}(A^T)$

$$\dim(\text{null}(A)) + \underbrace{\dim(\text{range}(A^T))}_{= \dim(\text{range}(A))} = n = \text{rank}(A)$$

Example:  $A = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$



| Subspace            | Basis   | Dimension |
|---------------------|---|-----------|
| $\text{null}(A)$    | $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$   | 1         |
| $\text{range}(A)$   | $\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}$            | 2         |
| $\text{range}(A^T)$ | $\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}$ | 2         |

# \* Matrix identities

- Inverse of a sum of matrices: given  $A_{p \times p}, B_{p \times q}, C_{q \times q}, D_{q \times p}$ , if  $A, C$  invertible:

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$$

(Sherman-Morrison-Woodbury formula)

- Inverse of a matrix by blocks:  $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$ , if  $A_{11}, A_{22}$  invertible:

$$A^{-1} = \begin{pmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{pmatrix} : \quad A^{11} = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} \quad A^{12} = -A^{11}A_{12}A_{22}^{-1} = -A_{11}^{-1}A_{12}A^{22}$$

$$A^{22} = (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} \quad A^{21} = -A^{22}A_{21}A_{11}^{-1} = -A_{22}^{-1}A_{21}A^{11}$$

- Derivatives:

$$\frac{d(a^T x)}{dx} = \nabla_x(a^T x) = a \quad \text{if } a, x \in \mathbb{R}^n, a \text{ independent of } x$$

$$\frac{d(x^T A x)}{dx} = \nabla_x(x^T A x) = (A + A^T)x \quad \underline{\underline{\text{A symmetric}}} \quad 2Ax \quad \text{if } A_{nn} \text{ indep. of } x$$

$$x_{m \times 1}, y_{n \times 1} : \frac{dy^T}{dx} = m \times n \text{ Jacobian matrix } J(x) = \left( \frac{\partial y_i}{\partial x_j} \right)_{ij}$$

$$f_{1 \times 1}, x_{n \times 1} : \frac{d^2 f}{dx dx^T} = n \times n \text{ Hessian matrix } \nabla^2 f(x) = \left( \frac{\partial^2 f}{\partial x_i \partial x_j} \right)_{ij}$$

$$\frac{d(x^T C)}{dx} = C_{n \times m}, \quad \frac{d(Bx)}{dx^T} = B_{m \times n} \quad \text{if } B, C \text{ indep. of } x.$$

$$\text{Product rule: } \frac{d(u^T v)}{dx} = \nabla_x(u^T v) = \frac{du^T}{dx} v + \frac{dv^T}{dx} u, \quad x, u, v \in \mathbb{R}^n$$

$$\text{Chain rule: } \frac{dy(x(t))}{dt} = \left( \nabla_x y \right)^T \frac{dx}{dt} = \sum_{i=1}^n \frac{\partial y}{\partial x_i} \frac{dx_i}{dt}, \quad x, y \in \mathbb{R}^n, t \in \mathbb{R}$$

# LEAST SQUARES, PSEUDOINVERSE AND SINGULAR VALUE DECOMPOSITION

(F)

\* Linear system  $Ax=b$  with  $m$  eqs.,  $n$  unknowns, assume  $A$  is full-rank:

- $m=n$ : unique solution  $x = A^{-1}b$
- $m < n$ : underconstrained, infinite solutions  $x = x_0 + u$   $\begin{cases} x_0 = \text{particular solution} \\ u \in \text{null}(A), \text{ with } \text{rank}(\text{null}(A)) = n-m \end{cases}$

Minimum-norm solution (min  $\|x\|^2$  s.t.  $Ax=b$ ):  $x = A^+b = A^T(AA^T)^{-1}b$ .

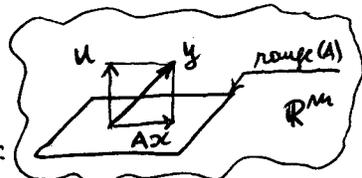
Homogeneous system:  $Ax=0$ :

- Unit-norm ~~solution~~ <sup>best approximation</sup> (min  $\|Ax\|^2$  s.t.  $\|x\|^2=1$ ):  $x = \text{minor eigenvector of } A^T A$ .
- min  $\|Ax\|^2$  s.t.  $\|Bx\|^2=1$ :  $x = \text{minor generalised eigenvector of } A^T A, B^T B$ .

•  $m > n$ : overconstrained, no solution in general; instead, define LSQ solution as

$$\min \|Ax-b\|^2 \Rightarrow x = A^+b = (A^T A)^{-1} A^T b.$$

$AA^+ = A(A^T A)^{-1} A^T$  is the orthogonal projection on  $\text{range}(A)$ :



(Pf: write  $y \in \mathbb{R}^m$  as  $y = Ax + u$  where  $x \in \mathbb{R}^n$  and  $u \perp Ax \forall x \in \mathbb{R}^n$  (i.e.,  $u \in \text{null}(A^T)$ ).

Then  $A(A^T A)^{-1} A^T y = Ax$ .

Likewise,  $A^+A = A^T(AA^T)^{-1}A$  is the orthogonal projection on  $\text{range}(A^T)$ .

\* Pseudoinverse of  $A$  is a matrix  $A^+$   $\begin{matrix} m \times n \\ n \times m \end{matrix}$  satisfying the Moore-Penrose conditions:

$$A^+AA^+ = A^+; AA^+A = A; AA^+, A^+A \text{ symmetric (uniquely defined)}.$$

Given the SVD of  $A$  as  $USV^T$ :  $A = \begin{matrix} m \times n \\ n \times n \end{matrix} \begin{matrix} n \times n \\ n \times n \end{matrix} \begin{matrix} n \times n \\ n \times n \end{matrix}$  with  $s_i^+ = \begin{cases} s_i^{-1} & \text{if } s_i > 0 \\ 0 & \text{if } s_i = 0 \end{cases}$

Particular cases:  $\begin{cases} m > n \Rightarrow A^+ = (A^T A)^{-1} A^T \\ m < n \Rightarrow A^+ = A^T (AA^T)^{-1} \\ m = n \Rightarrow A^+ = A^{-1} \end{cases}$

\* Singular value decomposition of  $A$ ,  $m \geq n$ :  $A = USV^T = \sum_{i=1}^n s_i u_i v_i^T$  with

$U = (u_1 \dots u_n)$  and  $V = (v_1 \dots v_n)$  orthonormal,  $S = \text{diag}(s_1, \dots, s_n)$  and singular

values  $s_1 \geq \dots \geq s_n \geq 0$ . Unique up to permutations/multiplicity of s.v.

•  $\text{rank}(A) = p \leq n \Leftrightarrow s_{p+1} = \dots = s_n = 0$ :  $A = (U_p \ U_{n-p}) \begin{pmatrix} S_p & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_p^T \\ V_{n-p}^T \end{pmatrix} = U_p S_p V_p^T$  with  $U_p = (u_1 \dots u_p)$  and  $V_{n-p} = (v_{p+1} \dots v_n)$  orthonormal bases of  $\text{range}(A)$  and  $\text{null}(A)$ , resp.

•  $\text{rank}(A) \geq p \Rightarrow U_p S_p V_p^T$  is the best rank- $p$  approximation to  $A$  in the sense of the Frobenius norm ( $\|A\|_F^2 = \text{tr} AA^T = \sum_{i,j} a_{ij}^2$ ) and the 2-norm ( $\|A\|_2 = \text{largest s.v.}$ ).

•  $A_{m \times n}, B_{n \times p}$ :  $\text{rank}(A) + \text{rank}(B) - n \leq \text{rank}(AB) \leq \min(\text{rank}(A), \text{rank}(B))$ .

$A_{m \times n}, B_{m \times n}$ :  $\text{rank}(A+B) \leq \text{rank}(A) + \text{rank}(B)$ .