Lecture 1 : Linear Algebra - Reminder

Summer Semester

Stéphane Marchand-Maillet
Stephane.Marchand-Maillet@cui.unige.ch

Outline

- Introduction
- Vector spaces
- Matrices
- Matrix calculus
- Eigen things
- Orthogonalisation
- SVD
- Pseudo-inverse
- Linear systems

Vector spaces (1)

- **Vector**: n-uple x (also noted $\vec{x}$ or $\underline{x}$ or $\mathbf{x}$) of elements of the field K (*e.g.*, $\mathbb{R}$ or $\mathbb{C}$):

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{pmatrix}$$

Vector spaces (2)

- **Vector space**: A set of vectors E on the field K supplemented with an internal addition operation $(+)$ and an external product operation $(\times)$ on $E \times K$ such that $(E, +)$ is a group and for all scalars $\lambda, \mu \in K$,
  - $\lambda \times x = x \times \lambda \in E$ (E closed under mult. with scalar),
  - $\lambda \times (x + y) = \lambda \times x + \lambda \times y$ (λ distributes over vector add.),
  - $(\lambda + \mu) \times x = \lambda \times x + \lambda \times x$ (vector distributes over λ add.),
  - $\lambda \times (\mu \times x) = (\lambda\mu) \times x$ (associative law of mult. by scalar),
  - $1 \times x = x$

Vector spaces (3)

- E is a **vector subspace** of E if for all $x, y \in F$ and all $\lambda, \mu \in K$, $\lambda \times x + \mu \times y \in F$.
- Examples (Vector spaces)
  - Polynomials with coefficients from K, $\mathcal{P}(K)$,
  - $n \times n$ matrices with elements in K, $\mathcal{M}_{n \times n}(K)$,
  - $\mathcal{C}^0(I)$: Continuous functions on an interval $I \subset K$
- Examples (Vector subspaces)
  - $\mathcal{P}_n(K)$, $n < n^*$
  - $\mathcal{C}^1(I)$: Continuous functions on an interval $I \subset K$ whose first derivative is also continuous.

Vector spaces (4)

- $x_1, \ldots, x_n \in E$ and $\lambda_1, \ldots \lambda_n \in \mathbb{R}$,

$$y \triangleq \sum_{i=1}^{n} \lambda_i \times x_i$$

is called a **linear combination** of vectors in E.

- $F = \{\text{linear combinations of } n \text{ vectors } x_1, \ldots, x_n \text{ of } E\}$
  F is a vector subspace of E.

- $\{e_1, \ldots, e_n\}$ are **linearly independent** if and only if

$$\sum_{i=1}^{n} \lambda_i e_i = 0_{\mathrm{E}} \text{ implies } \lambda_i = 0 \text{ for all } i = 1 \ldots n$$

otherwise, $\{e_1, \ldots, e_n\}$ are **linearly dependent**.

Vector spaces (5)

- The **rank** of F is the maximal number of linearly independent vectors that one can extract from F.
- $\mathcal{B} = \{e_1, \ldots\} \subset E$ is a **basis** of $E \Leftrightarrow$ any vector from E can uniquely be written as a linear combination of elements from $\mathcal{B}$

$\mathcal{B}$ basis of $E \Leftrightarrow e_1, \ldots e_n$ linearly independent $\Leftrightarrow \mathcal{B}$ **generates** E

Vector spaces (6)

A **scalar product** (or **dot product** or **inner product**) is a symmetric positive definite bilinear form from $E \times E$ to K: $x, y \mapsto < x, y >$ (also noted $x.y$) such that:

- $< x, y >=< y, x >$ for all x and y in E (commutativity)
- $< \lambda x + \mu y, z >= \lambda < x, z > + \mu < y, z >$ (bilinearity)
- $< x, x >> 0$ if $x \neq 0_{\mathrm{E}}$ (positive definiteness)

The scalar product induces a **norm** noted $\|x\|$ such that

$$\|x\|^2 =< x, x >$$

E added with the scalar product is said to be a **pre-Hilbert** space.

Vector spaces (6)

E pre-Hilbert space and F subset of E. The **orthogonal space** of F in E is

$$F^{\perp} \triangleq \{x \in E \text{ such that } < x, y >= 0 \text{ for all } y \in F\}$$

E is equivalently characterised by **Hilbert decomposition**

$$E = F \oplus F^{\perp}$$

In other words,

For any $x \in E$ there exist a unique pair (x_1, x_2)

with $x_1 \in F$ and $x_2 \in F^{\perp}$ such that $x = x_1 + x_2$

Matrix of a linear transform

E and F vector spaces with $\mathcal{B}$ and $\mathcal{B}'$ as bases of E and F

$f : E \to F$ a linear transform, the matrix

$$A = (a_{ij}) = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & a_{ij} & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}$$

defined **with respect to $\mathcal{B}$ and $\mathcal{B}'$**.

In other words, if $\mathcal{B} = \{e_1, \dots, e_n\}$ and $\mathcal{B}' = \{e'_1, \dots, e'_m\}$,

$$f(e_i) = \sum_{j=1}^{m} a_{ij} e'_j \qquad y = f(x) \Leftrightarrow Y = AX$$

- $\text{range}(A) = \{b \text{ such that } \exists x \text{ such that } b = Ax\}$
- $Ker(A) = \{x \text{ such that } Ax = 0\}$

Specific matrices (1)

- **Null** matrix $O_n = (a_{ij})_{ij}$ with $a_{ij} = 0$ for all $i, j = 1 \dots n$,
- **Identity** $I_n = (\delta_{ij})_{ij=1\dots n}$ ($\delta_{ij} = 1$ if $i = j$, $\delta_{ij} = 0$ otherwise),
- **Upper triangular** matrix $U = (u_{ij})_{ij}$ $u_{ij} = 0$ if $i > j$,
- **Lower triangular** matrix $L = (l_{ij})_{ij}$ such that $l_{ij} = 0$ if $i < j$,
- **Diagonal** matrix $D = (d_{ij}\delta_{ij})_{ij}$,
- **Symmetric** matrix $A = A^{\mathsf{T}}$,

Specific matrices (2)

- A **block matrix** can be divided into similar parts:

$$A = \begin{pmatrix} A_{11} & \dots & A_{1b} \\ \vdots & A_{ij} & \vdots \\ A_{b'1} & \dots & A_{b'b} \end{pmatrix}.$$

- A **block diagonal matrix**: $A_{ij} = 0$ if $i \neq j$,
- **Positive definite** matrix: $x^\mathsf{T} A x > 0$ for all $x \neq 0_\mathrm{E}$,
- **Orthogonal** matrix $A^\mathsf{T} A = I$ (A is real),
- **Unitary** matrix $A^\mathsf{H} A = I$ (A is complex),
- **Nilpotent** matrix $\exists k_0$ such that $A^k = 0$ for all $k \geq k_0$

Specific matrices (3)

- If $\det(A) = 0$ then A is a **singular**, in that case $\text{rank}(A) < n$.
- A and B **equivalent** then there exist P and Q such that $B = Q^{-1}AP$ (note that A may not be a square matrix),
- A and B are **similar** then there exists P such that $B = P^{-1}AP$ (note that this assumes that A is a square matrix),
- A and B are **congruent** then there exists P non-singular such that $B = P^{\mathsf{T}}AP$ ($B = P^{\mathsf{H}}AP$, if B is complex),

Specific matrices (4)

- $A = (a_{ij})_{ij}$ is a **Toepliz** matrix if there exist $2n-1$ scalar r_k, $k = -n+1, \ldots, n-1$ such that $a_{ij} = r_{j-i}$. In that case:

$$A = \begin{pmatrix} r_0 & r_1 & \ldots & r_{n-2} & r_{n-1} \\ r_{-1} & r_0 & \ldots & r_{n-3} & r_{n-2} \\ \vdots & \vdots & & \vdots & \vdots \\ r_{-n+1} & r_{-n+2} & \ldots & r_{-1} & r_0 \end{pmatrix}$$

  Each diagonal or sub-diagonal is formed by only one value.

- A **dyadic** matrix is such that $A = xy^\mathsf{T}$
- A pos. def. mat., B **matrix square root** of A if $A = BB^\mathsf{T}$,
- P is a **projection matrix** if $P^\mathsf{T} = P$ and $P^2 = P$.

Matrix calculus (1)

- The **trace** of A is defined as $\mathrm{Tr}(A) \triangleq \sum_i a_{ii}$
- $\lambda A + \mu B \triangleq (\lambda a_{ij} + \mu b_{ij})_{ij}$
- $AB \triangleq (\sum_k a_{ik} b_{kj})_{ij}$ (only the number of columns of A needs to be the same as the number of lines of B for the matrix multiplication to be possible),
- Other operations are defined from the elements:

$$A^{\mathsf{T}} = (a_{ji})_{ji} \text{ (transpose) ; } \bar{A} = (\bar{a}_{ij})_{ij} \text{ (conjugate) ;}$$

$$A^{\mathsf{H}} = (\bar{a}_{ji})_{ji} = \bar{A}^{\mathsf{T}} \text{ (Hermitian),}$$

Matrix calculus (2)

- Let A_{ij} be the sub-matrix created from A when removing the column and the line containing a_{ij} then, the determinant of A is defined recursively as

$$\det(A) = |A| \triangleq \sum_{j} (-1)^{i+j} a_{ij} \det(A_{ij})$$

A_{ij} is a **minor** and $(-1)^{i+j} \det(A_{ij})$ is a **cofactor** of A.

Properties (1)

- $AB \neq BA$ in general
- $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$
- $\det(AB) = \det(A)\det(B) = \det(BA)$
- $\det(A^{\mathsf{T}}) = \det(A)$
- $\det(A^{-1}) = (\det(A))^{-1}$. A^{-1} exists iff A non-singular.
- A $n \times n$ matrix, $\det(\lambda A) = \lambda^n \det(A)$, $\forall \lambda \in K$.
- $\det(A^{\mathsf{H}}) = \overline{\det(A)}$
- A $n \times m$ matrix and B $m \times n$ matrix then $\det(I_n - AB) = \det(I_m - BA)$

Properties (2)

- M block matrix
$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix},$$
then $\det(M) = \det(A)\det(D - CA^{-1}B)$.
If $B = C = 0$ (M block diag.), $\det(M) = \det(A)\det(D)$.

- If A is a dyadic matrix then $\mathrm{rank}(A) =?$.

- **Matrix inversion lemma**: Given A, B, C, D
$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(DA^{-1}B + C^{-1})^{-1}DA^{-1},$$
if A^{-1}, C^{-1} and $(DA^{-1}B + C^{-1})^{-1}$ exist.

- B matrix square root of A, then $\mathrm{range}(B) = \mathrm{range}(A)$.

Eigenspaces (1)

Eigenvalue, eigenvector:

- $f : E \to F$ linear transform. λ is an **eigenvalue** of f if there exists $u \neq 0_{\mathrm{E}}$ such that $f(u) = \lambda u$. In that case u is an **eigenvector** of f.
  The subspace generated by $\{u \text{ such that } f(u) = \lambda u\}$ is called the **eigensubspace** associated with the eigenvalue λ.
- Given $A \in \mathcal{M}_{n \times n}(\mathbb{R})$, λ is an eigenvalue of A associated with the eigenvector u if and only if

$$Au = \lambda u$$

.

Eigenspaces (2)

A a $n \times n$ matrix its **characteristic polynomial** is given by

$$P_A(\lambda) \stackrel{\Delta}{=} \det(A - \lambda I)$$

Its is a polynomial in λ of order n and is of the form

$$P_A(\lambda) = (-1)^n \lambda^n + (-1)^{n-1} \mathrm{Tr}(A) \lambda^{n-1} + \ldots + \det(A)$$

Eigenspaces (3)

- **Characteristic equation**:

$$P_A(\lambda) = 0$$

  and the set $\{\lambda_i\}$ of all eigenvalues of A is defined by the set of solution of the characteristic polynomial.

- In other words, the eigenvalues of A are the roots of the characteristic polynomial.

- **Cayley Hamilton theorem**: Any matrix satisfies its own characteristic equation,

$$P_A(A) = 0.$$

Eigenspaces (4)

- The equation $Ax = \lambda x$ can be generalised as $AX = X\Lambda$, where the column of X are the eigenvectors of A. This therefore shows that $\Lambda = X^{-1}AX$ so that A is similar to the diagonal matrix Λ formed by its eigenvalues. The transition matrix is such that its columns are the eigenvectors of A.

- This generalises to the base change case. When **changing the basis** $\mathcal{B}$ for another basis $\mathcal{B}'$ then the matrix is mapped onto a similar matrix A' by a **transition matrix** P as follows

$$A' = P^{-1}AP,$$

where the columns of P are the coordinates of the new basis vectors (vectors of $\mathcal{B}'$) in the former basis $\mathcal{B}$. Moreover, if A is symmetric then, P is orthogonal (or unitary).

Gram-Schmidt orthogonalisation (1)

Gram determinant

n vectors $\{x_1, \ldots, x_n\}$ are linearly independent if and only if their corresponding Gram determinant $\det(G)$ is non-zero, where

$$G \triangleq \begin{pmatrix} < x_1, x_1 > & \ldots & < x_1, x_n > \\ \vdots & \ddots & \vdots \\ < x_n, x_1 > & \ldots & < x_n, x_n > \end{pmatrix}$$

G is called the **Gram matrix** associated with $\{x_1, \ldots, x_n\}$.
Clearly, G is symmetric positive definite.

Gram-Schmidt orthogonalisation (2)

- Normalise x_1 to define e_1.
- Given $\{e_1, \ldots, e_{i-1}\}$, e_i is defined by subtracting from x_i its projection on the space spanned by $\{e_1, \ldots, e_{i-1}\}$ and normalising the result

$$p_i = \sum_{k=1}^{i-1} < x_i, e_k > e_k \text{ and } e_i = \frac{x_i - p_i}{\|x_i - p_i\|}$$

Gram-Schmidt orthogonalisation (3)

1. $C \leftarrow \{x_1, \ldots, x_n\}$
2. $m \leftarrow 1$, $t \leftarrow 0$, $y \leftarrow x_1$, $C \leftarrow C - \{x_1\}$
3. $t \leftarrow n + 1$, $e_t \leftarrow y/\|y\|$
4. If C is empty, then end.
5. Else $m \leftarrow m + 1$, $y \leftarrow x_m - \sum_{k=1}^{t} < x_m, e_k > e_k$,
   $C \leftarrow C - \{x_m\}$
6. If $y = 0$ repeat 4. (This may happen if one of the initial vectors is null).
7. Else go to 3.

m gives the dimension of the space spanned by $\{x_1, \ldots, x_n\}$ (and therefore the number of orthonormal vectors e_i found).

QR decomposition

- Any $n \times m$ matrix A whose columns are linearly independent can be written as $A = QR$ where the columns of Q are orthonormal and R is an upper triangular invertible matrix. Moreover, if $m = n$, then Q is an orthogonal matrix.
- The idea here is to use this scheme for solving a system $Ax = b$. In the case where matrices are square, Q is orthogonal so that $Q^{\mathsf{T}} = Q^{-1}$. Therefore

$$Ax = b \Leftrightarrow QRx = b \Leftrightarrow Rx = Q^{\mathsf{T}} b$$

  The latter system is straightforward to solve since R is an upper triangular matrix.

Singular Value Decomposition (1)

If A is a real $m \times n$ matrix, there exist two square orthogonal matrices U $(m \times m)$ and V $(n \times n)$ such that

$$U^{\mathsf{T}} A V = \Sigma \triangleq \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_p \\ 0 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & 0 \end{pmatrix},$$

where $p = \min(m, n)$ and $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p \geq 0$ are called the **singular values** of A and the columns u_i of U and v_i of V are the left and right **singular vectors** of A.

SVD (2)

If we define r such that

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > \sigma_{r+1} = \ldots = \sigma_p = 0$$

then

- $\operatorname{rank}(A) = r$,
- $\ker(A)$ is the subspace generated by $\{v_{r+1}, \ldots, v_n\}$,
- $\operatorname{range}(A)$ is the subspace generated by $\{u_1, \ldots, u_r\}$,
- and the SVD expansion of A is defined as

$$A \triangleq \sum_{i=1}^{r} \sigma_i u_i v_i^{\mathsf{T}}.$$

SVD (3)

If $U^\mathsf{T} A V = \Sigma$ then,

$$(U^\mathsf{T} A V)^\mathsf{T} U^\mathsf{T} A V = \Sigma^\mathsf{T} \Sigma = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_p^2 \end{pmatrix} = (V^\mathsf{T} A^\mathsf{T} U)(U^\mathsf{T} A V)$$

SVD (4)

Therefore, since U and V are orthogonal,

$$V^{\mathsf{T}}(A^{\mathsf{T}}A)V = \Sigma^{\mathsf{T}}\Sigma = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_p^2 \end{pmatrix}$$

$$U^{\mathsf{T}}(AA^{\mathsf{T}})U = \Sigma\Sigma^{\mathsf{T}} = \begin{pmatrix} \sigma_1^2 & & & & & \\ & \ddots & & & 0 & \\ & & \sigma_p^2 & & & \\ & & & 0 & & \\ & 0 & & & \ddots & \\ & & & & & 0 \end{pmatrix}$$

SVD (5)

Finally,

$$A^\mathsf{T} A = V \Sigma^\mathsf{T} \Sigma V^\mathsf{T} = V \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_p^2 \end{pmatrix} V^\mathsf{T}$$

$$A A^\mathsf{T} = U \Sigma \Sigma^\mathsf{T} U^\mathsf{T}$$

The computation of the singular value decomposition is therefore simply done by calculating the eigenvalues and eigenvectors of $A^\mathsf{T} A$ $(n \times n)$. This gives V as eigenvector matrix and $\{\sigma_1^2, \ldots, \sigma_p^2\}$ as eigenvalues. U is similarly computed from the eigensystem corresponding to AA^T $(m \times m)$.

SVD (6)

It is often the case that $n \ll m$. In that case finding the eigenvalues of $A^\top A$is far easier than finding the eigensystem of $AA^\top$. Since often only the singular values are of interest, the calculation only concentrates on $A^\top A$.

Pseudo-inverse matrix (1)

Given the equation system $Ax = b$, one would ideally like to calculate $x = A^{-1}b$. In the case where A^{-1} does not exist, we define a matrix $A^{\#}$ (also noted A^{+}) called the **pseudo-inverse** of A, such that,

$$x = A^{\#}b = A^{+}b.$$

Pseudo-inverse matrix (2)

It should be an inverse in a weaker sense and we define it such that:

- $AA^{\#}A = A$ (as opposed to $AA^{-1} = I$)
- $A^{\#}AA^{\#} = A^{\#}$ (as opposed to $A^{-1}A = I$)

One can easily verify that if we define

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

The properties are satisfied.

Pseudo-inverse matrix (3)

It can be shown that the above definition derives from the solution of the system "projected" into the range of A (range(A)). In other words, the solution $x^* = A^+ b$ characterised is such that $b^* = Ax^*$ is the projection of b on range(A) (minimum square error (MSE) solution). The projection matrix such that $b^* = Pb$ is then

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

Solving linear systems (1)

Cramer system: A^{-1} exists

$\Rightarrow$ $x = A^{-1}b$

$\Rightarrow$ $b \in \text{range}(A)$

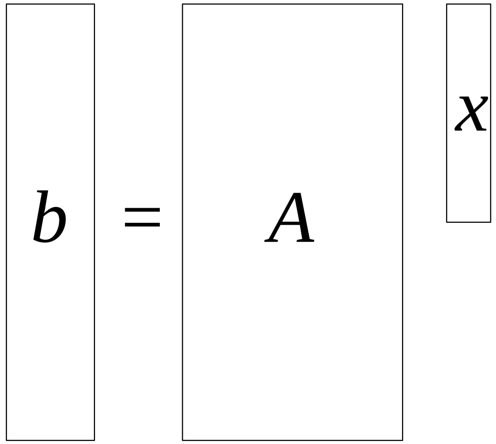
$\Rightarrow$ $Ker(A) = \{\vec{0}\}$

$\Rightarrow$ Unique solution

Solving linear systems (2)

Over determined system: n equations, m unknown, $n > m$

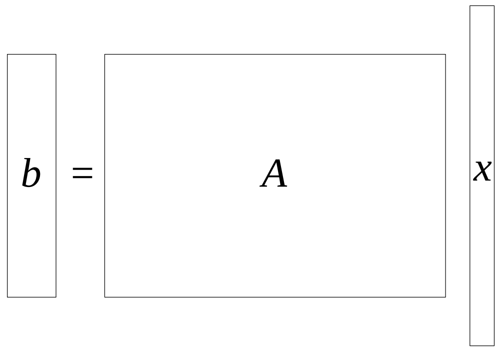
$\Rightarrow$ Unlikely to have $b \in \text{range}(A)$



Solving linear systems (3)

Under determined system: n equations, m unknown, $n < m$

$\Rightarrow$ Many possible solutions. Which to choose?



Cases 2 and 3: characterise a solution. We define

$$\varepsilon = b - Ax$$

Solving linear systems (case 2)

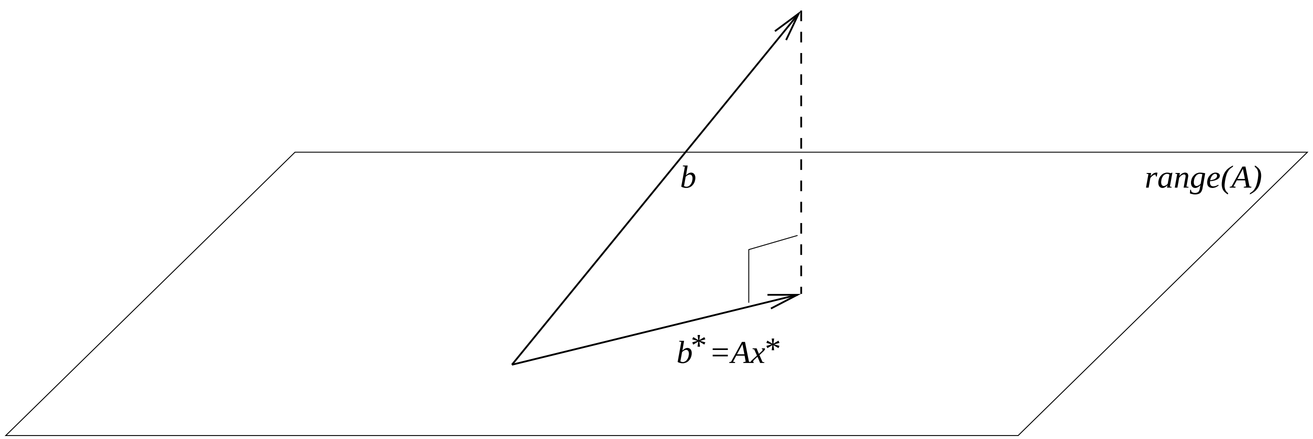
MSE criterion $\|\varepsilon\|^2 = \varepsilon^\mathsf{T}\varepsilon$

$$\|\varepsilon\|^2 = b^\mathsf{T} b - 2x^\mathsf{T} A^\mathsf{T} b + x^\mathsf{T} A^\mathsf{T} A x$$

$$\frac{\partial\|\varepsilon\|^2}{\partial x} = 0 - 2A^\mathsf{T} b + 2A^\mathsf{T} A x$$

$$\frac{\partial\|\varepsilon\|^2}{\partial x} = 0 \Leftrightarrow x^* = (A^\mathsf{T} A)^{-1} A^\mathsf{T} b$$

$$\Leftrightarrow b^* = Ax^* = A(A^\mathsf{T} A)^{-1} A^\mathsf{T} b$$



Solving linear systems (case 3)

We have plenty of solutions.

We choose $\min \|x\|^2$ as criterion (min norm least square MNLS)

$Ax = b$ is a constraint

Lagragian: $\mathcal{L} = \|x\|^2 + \lambda^\mathsf{T}(b - AX)$

Saddle point: $\frac{\partial \mathcal{L}}{\partial x} = 0; \frac{\partial \mathcal{L}}{\partial \lambda} = 0$

$$\frac{\partial \mathcal{L}}{\partial x} = 2x - A^\mathsf{T}\lambda \qquad \frac{\partial \mathcal{L}}{\partial \lambda} = b - Ax$$

$$\text{Saddle point} \Leftrightarrow 2Ax - AA^\mathsf{T}\lambda = 0$$

$$\Leftrightarrow \lambda = 2(AA^\mathsf{T})^{-1}b \qquad \Leftrightarrow x^* = A^\mathsf{T}(AA^\mathsf{T})^{-1}b$$

Duality: MSE $((A^\mathsf{T}A)^{-1}A^\mathsf{T}) \leftrightarrow$ MNLS $(A^\mathsf{T}(AA^\mathsf{T})^{-1})$