

6.5. La méthode des moindres carrés.

Motivation.

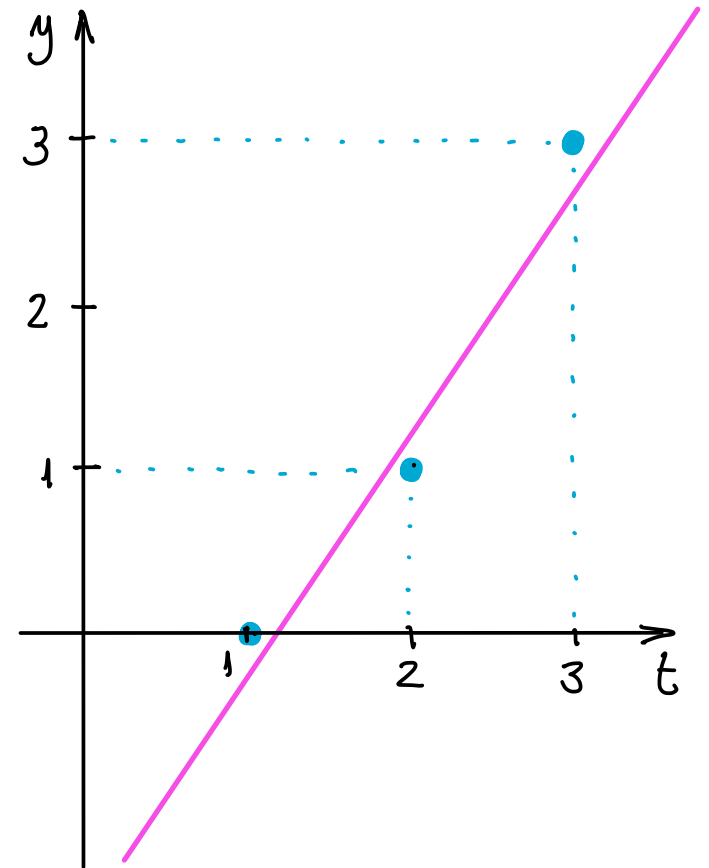
Considérons trois points du plan :

$$(1,0), (2,1), (3,3)$$

On constate qu'ils *ne sont pas* alignés !

Question.

Quelle est l'équation de la droite du plan la plus "proche" de ces points ?



Soit $y = x_1 + x_2 t$ l'équation cherchée, où x_1, x_2 sont à déterminer.

On aimerait que les trois points donnés soient sur la droite:

$$\begin{aligned} (t_1, y_1) &= (1, 0): \\ (t_2, y_2) &= (2, 1): \\ (t_3, y_3) &= (3, 3): \end{aligned} \quad \begin{cases} 0 = x_1 + 1 x_2 \\ 1 = x_1 + 2 x_2 \\ 3 = x_1 + 3 x_2 \end{cases}$$

Equation matricielle associée: $A\vec{x} = \vec{b}$ où $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$ et $\vec{b} = \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}$

matrice augmentée:

$$\left[\begin{array}{cc|c} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 1 & 3 & 3 \end{array} \right] \xrightarrow[\substack{L_2 \rightarrow L_2 - L_1 \\ L_3 \rightarrow L_3 - L_1}]{\sim} \left[\begin{array}{cc|c} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 2 & 3 \end{array} \right] \xrightarrow{L_3 \rightarrow L_3 - 2L_2} \left[\begin{array}{cc|c} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{array} \right]$$

ce qui confirme que les points ne sont pas alignés.

Considérons maintenant le cas général:

Soit A une matrice de taille $m \times n$ et $\vec{b} \in \mathbb{R}^m$.

Nous avons vu que le système $A\vec{x} = \vec{b}$ possède des solutions si et seulement si $\vec{b} \in \text{Col}(A)$.

Si $m > n$, alors il est fort probable que cela n'arrive pas.

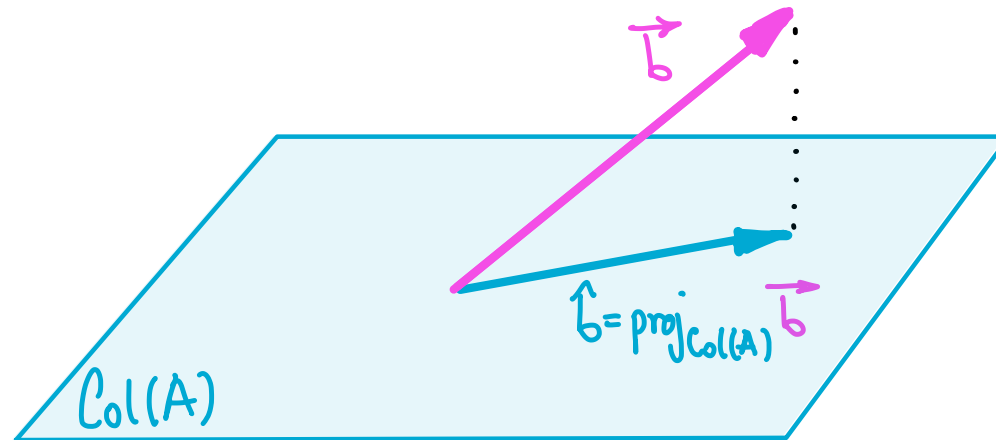
En pratique, on s'intéresse à des situations où $m \gg n$ (m beaucoup plus grand que n).

En particulier, si $\vec{b} \notin \text{Col}(A)$ alors $A\vec{x} = \vec{b}$ ne possède pas de solution!

On prend

$$\hat{\vec{b}} = \text{proj}_{\text{Col}(A)} \vec{b}$$

Géométriquement:



Comme par construction $\hat{\vec{b}} \in \text{Col}(A)$, le système $A\vec{x} = \hat{\vec{b}}$ est consistant et il existe au moins un $\hat{\vec{x}} \in \mathbb{R}^n$ tel que

$$A\hat{\vec{x}} = \hat{\vec{b}}.$$

Définition.

Soit A une matrice de taille $m \times n$ et $\vec{b} \in \mathbb{R}^m$.

On dit que $\hat{x} \in \mathbb{R}^n$ est une solution au sens des moindres carrés de l'équation $A\vec{x} = \vec{b}$ si

$$\|\vec{b} - A\hat{x}\| \leq \|\vec{b} - A\vec{y}\| \quad \text{pour tout } \vec{y} \in \mathbb{R}^n.$$

La distance $\|\vec{b} - A\hat{x}\|$ est appelée erreur de l'approximation.

Remarque.

Si $\vec{b} \in \text{Col}(A)$ alors $\hat{x} \in \mathbb{R}^n$ n'est rien d'autre qu'une solution de l'équation $A\vec{x} = \vec{b}$

Si $\vec{b} \notin \text{Col}(A)$ alors $\hat{x} \in \mathbb{R}^n$ n'est pas une solution de $A\vec{x} = \vec{b}$.

Question. Comment déterminer $\hat{x}$?

Par construction, $\vec{b} - \hat{b} \in (\text{Col}(A))^\perp = \text{Nul}(A^T)$.

Par conséquent, $A^T(\vec{b} - \hat{b}) = \vec{0} \iff A^T\vec{b} = A^T\hat{b}$

Comme $\hat{x} \in \mathbb{R}^n$ est tel que $A\hat{x} = \hat{b}$ nous avons donc :

$$(A^T A) \hat{x} = A^T \vec{b}$$

Définition.

Soit A une matrice de taille $m \times n$ et $\vec{b} \in \mathbb{R}^m$.

Le système $(A^T A) \vec{x} = A^T \vec{b}$ est appelé *système normal* associé au système $A \vec{x} = \vec{b}$.

Remarque.

Si $\hat{x} \in \mathbb{R}^n$ est une solution au sens des moindres carrés de $A\vec{x} = \vec{b}$, alors $\hat{x}$ est une solution du système normal $(A^T A)\vec{x} = A^T \vec{b}$ associé.

Théorème.

L'ensemble des solutions au sens des moindres carrés de l'équation $A\vec{x} = \vec{b}$ est égal à l'ensemble (non-vide) des solutions du système normal $(A^T A)\vec{x} = A^T \vec{b}$ associé.

Preuve.

Il reste à montrer que si $\hat{x} \in \mathbb{R}^n$ est une solution de $(A^T A)\vec{x} = A^T \vec{b}$, alors $\hat{x}$ est une solution au sens des moindres carrés de $A\vec{x} = \vec{b}$.

Si on a $A^T A \hat{x} = A^T \vec{b}$, alors $A^T (A \hat{x} - \vec{b}) = \vec{0}$ et

Par conséquent, $\vec{b} - A \hat{x} \in \text{Nul}(A^T) = (\text{Col}(A))^\perp$.

Comme $\vec{b} = \underbrace{(\vec{b} - A \hat{x})}_{\in (\text{Col}(A))^\perp} + \underbrace{A \hat{x}}_{\in \text{Col}(A)}$

l'unicité de la décomposition orthogonale nous dit que

$$A \hat{x} = \text{proj}_{\text{Col}(A)} \vec{b} = \hat{b}$$

autrement dit, $\hat{x}$ est une solution au sens des moindres carrés de $A\vec{x} = \vec{b}$. ■

Exemple.

$$\text{Soient } A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \text{ et } \vec{b} = \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}$$

Trouver une solution au sens des moindres carrés du système $A\vec{x} = \vec{b}$.

$$\text{On a } A^T A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}$$

$$\text{et } A^T \vec{b} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 4 \\ 11 \end{bmatrix}$$

matrice augmentée associée au système normal:

$$\left[\begin{array}{cc|c} 3 & 6 & 4 \\ 6 & 14 & 11 \end{array} \right] \xrightarrow{L_2 \rightarrow L_2 - 2L_1} \left[\begin{array}{cc|c} 3 & 6 & 4 \\ 0 & 2 & 3 \end{array} \right] \xrightarrow{L_1 \rightarrow L_1 - 3L_2} \left[\begin{array}{cc|c} 3 & 0 & -5 \\ 0 & 2 & 3 \end{array} \right] \sim \left[\begin{array}{cc|c} 1 & 0 & -5/3 \\ 0 & 1 & 3/2 \end{array} \right]$$

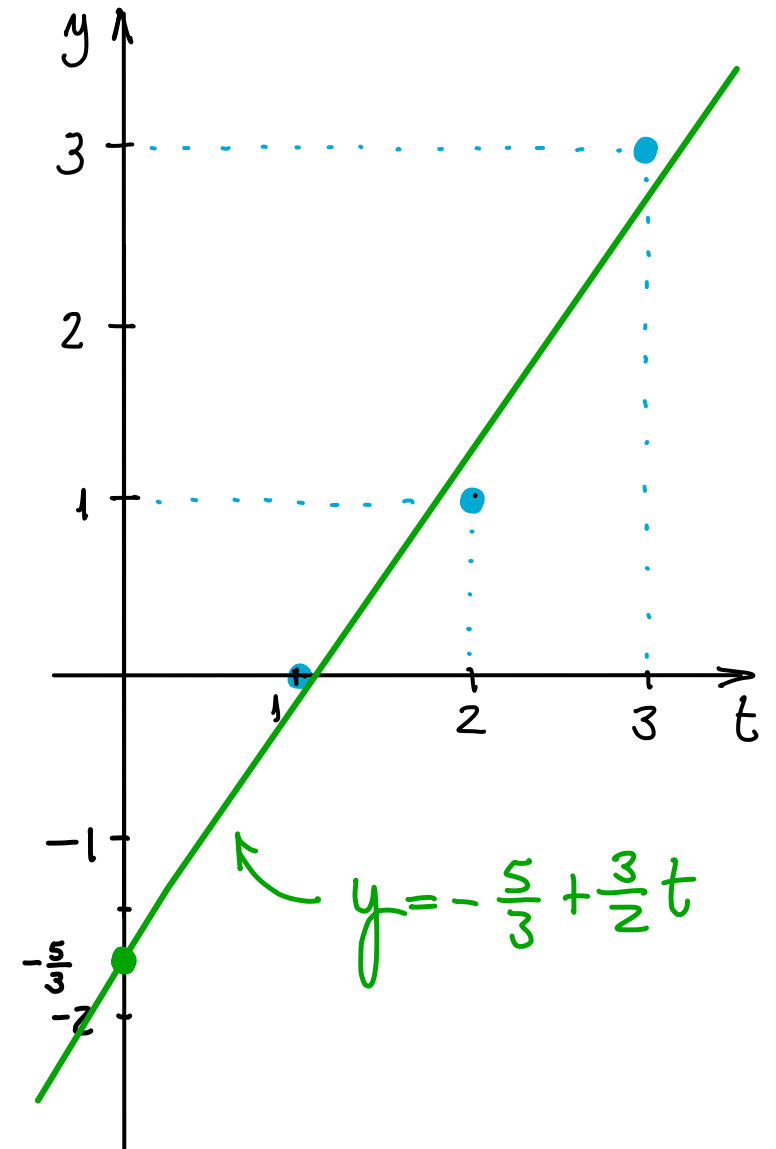
$$\text{solution: } \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} -5/3 \\ 3/2 \end{bmatrix}$$

Remarque.

La matrice A et le vecteur $\vec{b}$ de l'exemple sont associés au problème de recherche de la droite la plus "proche" des points $(1,0), (2,1), (3,3)$. Ainsi,

$$y = \hat{x}_1 + \hat{x}_2 t = -\frac{5}{3} + \frac{3}{2} t$$

est la droite la plus "proche" des points $(1,0), (2,1), (3,3)$, appelée droite de régression linéaire ou droite des moindres carrés.



De manière générale, la droite de régression linéaire pour les points

$$(t_1, y_1), (t_2, y_2), \dots, (t_N, y_N) \in \mathbb{R}^2$$

peut être trouvée à l'aide de la solution au sens des moindres carrés du système $A\vec{x} = \vec{b}$ où

$$A = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_N \end{bmatrix} \in M_{N,2}(\mathbb{R}) \quad \text{et} \quad \vec{b} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \in \mathbb{R}^N$$

Si $\hat{x} = \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix}$ est la solution au sens des moindres carrés, alors l'équation de la droite de régression linéaire est:

$$y = \hat{x}_1 + \hat{x}_2 t$$

Exemple.

$$\text{Soit } A = \begin{bmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \text{ et } \vec{b} = \begin{bmatrix} 4 \\ 4 \\ 5 \\ 5 \\ 7 \end{bmatrix}$$

On constate que le système $A\vec{x} = \vec{b}$ n'a pas de solution.

Trouver une solution au sens des moindres carrés du système $A\vec{x} = \vec{b}$.

$$\text{On a } A^T A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 5 & 0 \\ 0 & 10 \end{bmatrix}$$

$$\text{et } A^T \vec{b} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ 4 \\ 5 \\ 5 \\ 7 \end{bmatrix} = \begin{bmatrix} 25 \\ 7 \end{bmatrix}$$

$$\text{d'où } \left[\begin{array}{cc|c} 5 & 0 & 25 \\ 0 & 10 & 7 \end{array} \right] \sim \left[\begin{array}{cc|c} 1 & 0 & 5 \\ 0 & 1 & 7/10 \end{array} \right]$$

$$\Rightarrow \text{solution: } \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 5 \\ 7/10 \end{bmatrix}$$

Conséquence.

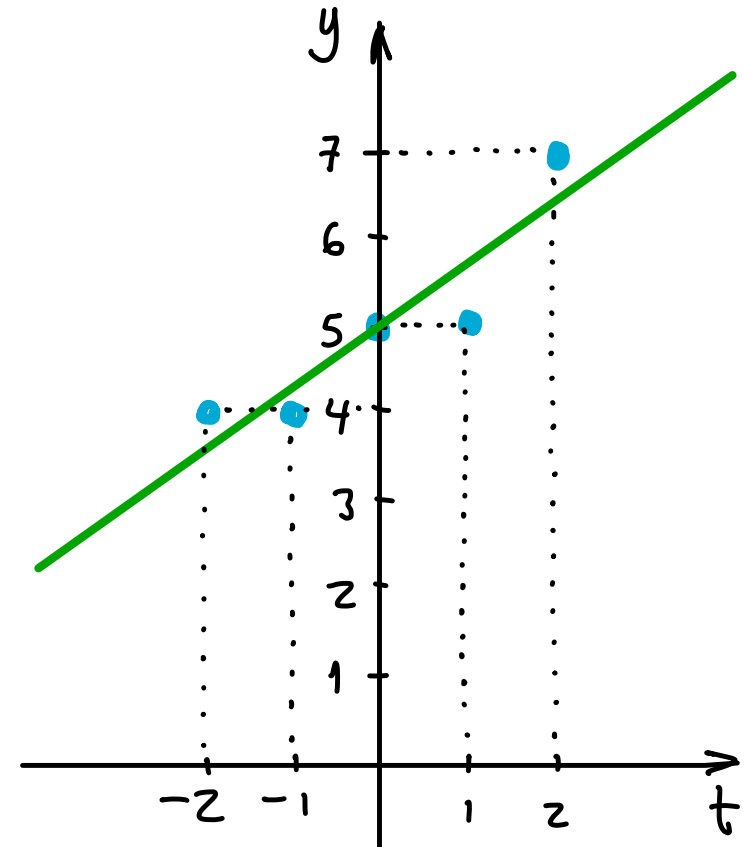
La droite

$$y = \hat{x}_1 + \hat{x}_2 t = 5 + \frac{7}{10} t$$

est la droite de régression linéaire

pour les points

$$(-2, 4), (-1, 4), (0, 5), (1, 5), (2, 7)$$



Théorème. Soit A une matrice de taille $m \times n$.

Les affirmations suivantes sont équivalentes:

1. $A\vec{x} = \vec{b}$ admet une unique solution au sens des moindres carrés pour tout $\vec{b} \in \mathbb{R}^m$.

2. Les colonnes de A sont linéairement indépendantes

3. $\text{rang}(A) = n$

4. $\text{rang}(A^T A) = n$

5. La matrice $A^T A$ est inversible

Dans ce cas, la solution au sens des moindres carrés s'écrit

$$\hat{\vec{x}} = (A^T A)^{-1} A^T \vec{b}$$

Remarque.

Pour montrer l'équivalence entre 3 et 4 on peut utiliser :

Si A est une matrice de taille $m \times n$, alors $\text{rang}(A^T A) = \text{rang}(A)$
(voir exercice 16, série 13).

Théorème.

Soit A une matrice de taille $m \times n$ dont les colonnes sont linéairement indépendantes.

Soit $A = QR$ une factorisation QR de A . La solution au sens des moindres carrés de $A\vec{x} = \vec{b}$ est unique et s'écrit :

$$\hat{\vec{x}} = R^{-1} Q^T \vec{b}$$

Preuve.

Comme les colonnes de A sont linéairement indépendantes, la solution au sens des moindres carrés est unique :

$$\hat{x} = (A^T A)^{-1} A^T \vec{b}$$

Par hypothèse, A peut s'écrire sous la forme $A = QR$

où Q est une matrice de taille $m \times n$ telle que $Q^T Q = I_n$

et R est une matrice de taille $n \times n$ triangulaire supérieure inversible.

On a $A^T = R^T Q^T$ et $A^T A = (R^T Q^T)(QR) = R^T (Q^T Q) R = R^T R$,

d'où $(A^T A)^{-1} A^T = (R^T R)^{-1} R^T Q^T = R^{-1} (R^T)^{-1} R^T Q^T = R^{-1} Q^T$

et $\hat{x} = R^{-1} Q^T \vec{b}$



Calcul alternatif de la projection orthogonale.

Nous avons défini $\hat{\mathbf{b}} = \text{proj}_{\text{Col}(A)} \vec{\mathbf{b}}$

Comme $\hat{\mathbf{x}} \in \mathbb{R}^n$ est tel que $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$, pour trouver la

projection orthogonale de $\vec{\mathbf{b}}$ sur $\text{Col}(A)$ il suffit

de trouver une solution $\hat{\mathbf{x}}$ du système normal $(A^T A) \vec{\mathbf{x}} = A^T \vec{\mathbf{b}}$

et la multiplier par A :

$$\text{proj}_{\text{Col}(A)} \vec{\mathbf{b}} = A \hat{\mathbf{x}}$$

plutôt que de construire une base orthogonale à l'aide du procédé d'orthogonalisation de Gram-Schmidt.

(voir exercice 8, série 14).

Complément: (hors programme)

Théorème.

Soit W un sous-espace vectoriel de $\mathbb{R}^n$ de dimension $k > 0$.

Soit $\{\vec{w}_1, \dots, \vec{w}_k\}$ une base quelconque de W (pas forcément orthogonale). Soit A la matrice de taille $n \times k$ suivante.

$$A = [\vec{w}_1 \ \vec{w}_2 \ \dots \ \vec{w}_k]$$

alors la matrice canoniquement associée à l'application

linéaire $T = \text{proj}_W: \mathbb{R}^n \rightarrow \mathbb{R}^n$ est

$$A_T = A(A^T A)^{-1} A^T$$

Preuve.

Comme $\{\vec{w}_1, \dots, \vec{w}_k\}$ est une base de W et $A = [\vec{w}_1 \ \vec{w}_2 \ \dots \ \vec{w}_k]$, nous avons $W = \text{Col}(A)$.

Par conséquent, $W^\perp = (\text{Col}(A))^\perp = \text{Nul}(A^T)$

Le théorème de la décomposition spectrale nous dit que tout vecteur $\vec{y} \in \mathbb{R}^n$ s'écrit de manière unique

$$\vec{y} = \hat{y} + \vec{z}, \text{ où } \hat{y} \in W \text{ et } \vec{z} \in W^\perp.$$

Comme $\hat{y} \in W = \text{Col}(A)$, on a $\hat{y} = A \vec{x}$, avec $\vec{x} \in \mathbb{R}^k$

Comme $\vec{z} \in W^\perp = \text{Nul}(A^T)$, on a $A^T \vec{z} = \vec{0}$.

$$\text{Ainsi } A^T \vec{y} = A^T(\hat{y} + \vec{z}) = A^T \hat{y} + A^T \vec{z} = A^T A \vec{x} + \vec{0}$$

$$\text{d'où } A^T \vec{y} = A^T A \vec{x}$$

Comme $A^T A$ est une matrice carrée de taille $k \times k$

et $\text{rang}(A^T A) = \text{rang}(A)$ (voir exercice 16, série 13)

la matrice $A^T A$ est inversible car $\text{rang}(A^T A) = k$.

Nous avons donc $\vec{x} = (A^T A)^{-1} A^T \vec{y}$, d'où

$$\begin{aligned} \text{proj}_W \vec{y} &= \hat{y} = A \vec{x} \\ &= A(A^T A)^{-1} A^T \vec{y} \end{aligned}$$

