Chapitre 3 : Factorisation QR et systèmes surdéterminés

Factorisation QR : généralités

Une matrice Q est orthogonale si elle est carrée et $Q^T Q = I$.

Une matrice $R = (r_{ij})$ est trapézoïdale supérieure si $r_{ij} = 0$ pour tout $i > j$.

Une factorisation QR d'une matrice A (instruction `qr(A)`) de dimensions $m \times n$ ($m \geq n$) est une combinaison des matrices Q orthogonale (de dimensions $m \times m$) et R trapézoïdale supérieure (de dimensions $m \times n$) telles que

$$\boxed{A = QR}.$$

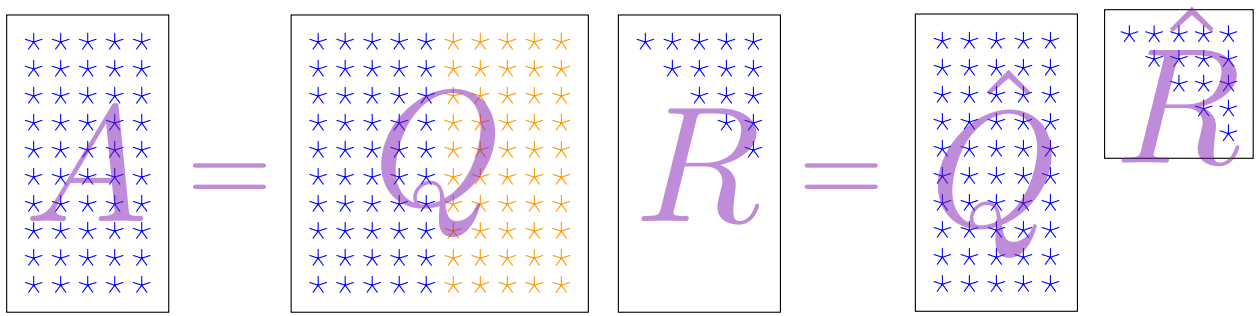
Comme les $m - n$ dernières lignes de R sont nulles, on a

$$R = \begin{pmatrix} \hat{R} \\ \end{pmatrix}, \quad \text{avec } \hat{R} \text{ de dimensions } n \times n,$$

et en subdivisant $Q = \begin{pmatrix} \hat{Q} & \hat{Q}_\perp \end{pmatrix}$ avec $\hat{Q}$ de dimensions $m \times n$ on a aussi une factorisation QR réduite (instruction `qr(A,0)`)

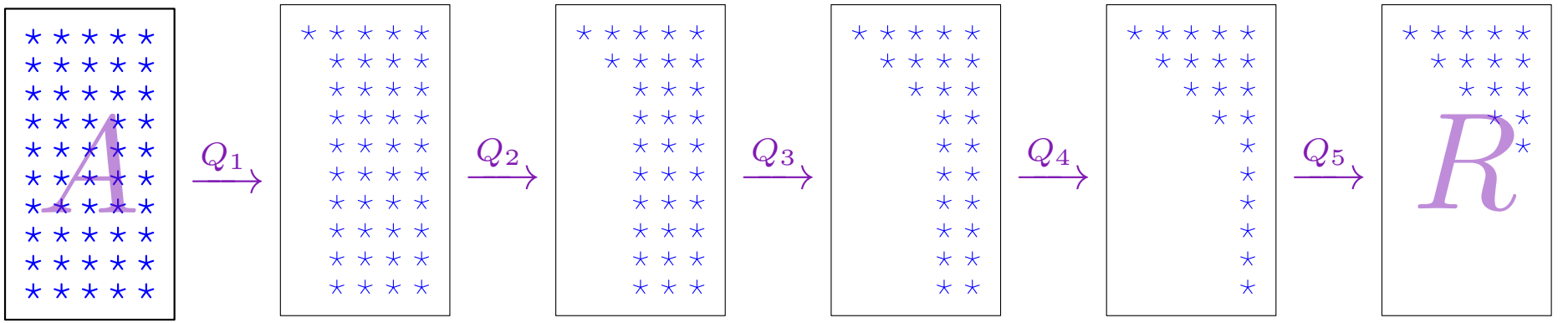
$$\boxed{A = \hat{Q}\hat{R}}.$$

Exemple :
($m = 10$, $n = 5$)



Méthode de Householder : principe

Principe de base :



Si Q_i, $i = 1, ..., 5$, sont des matrices orthogonales symétriques, alors

$$Q \;=\; Q_1 \cdots Q_5$$

et

$$Q^T \;=\; Q_5 \cdots Q_1$$

sont également orthogonales (pourquoi ? et symétriques ?).

D'autre part $R \;=\; Q_5 \cdots Q_1 A \;=\; Q^T A$ est bien trapézoïdale supérieure.

On a donc bien une factorisation

$$A \;=\; QR\,.$$

Transformation de Householder

Une manière d'obtenir la factorisation QR est d'utiliser la transformation de Householder, définie pour un vecteur $\mathbf{v} = (v_i)$ comme

$$\boxed{H \;=\; I - 2\frac{\mathbf{v}\mathbf{v}^T}{\|\mathbf{v}\|_2^2}}\,.$$

La transformation est :

- symétrique

  $(\mathbf{v}\mathbf{v}^T)^T \;=\; \mathbf{v}\mathbf{v}^T$;

  et donc H est combinaison linéaire de matrices symétriques ;
- orthogonale

  car $H^T H \;=\; HH \;=\; I - 4\dfrac{\mathbf{v}\mathbf{v}^T}{\|\mathbf{v}\|_2^2} + 4\dfrac{\mathbf{v}(\mathbf{v}^T\mathbf{v})\mathbf{v}^T}{\|\mathbf{v}\|_2^4} \;=\; I$;

Transformation de Householder

Une manière d'obtenir la factorisation QR est d'utiliser la transformation de Householder, définie pour un vecteur $\mathbf{v} = (v_i)$ comme

$$\boxed{H = I - 2\frac{\mathbf{v}\mathbf{v}^T}{\|\mathbf{v}\|_2^2}}.$$

La transformation permet de :

- introduire des zéros

  pour un vecteur $\mathbf{x} = (x_i)$ donné on choisit $\mathbf{v} = \mathbf{x} \pm \|\mathbf{x}\|_2 \mathbf{e}_j$
  où $\mathbf{e}_j$ est le $j$ème vecteur de la base canonique ; alors

  $$\boxed{H\mathbf{x} = \mp\|\mathbf{x}\|_2 \mathbf{e}_j}.$$

  En effet, $2\mathbf{v}^T\mathbf{x} = 2\|\mathbf{x}\|_2^2 \pm 2\|\mathbf{x}\|_2 x_j = \|\mathbf{v}\|_2^2$ et donc

  $$H\mathbf{x} = \mathbf{x} - 2\frac{\mathbf{v}\mathbf{v}^T}{\|\mathbf{v}\|_2^2}\mathbf{x} = \mathbf{x} - \mathbf{v} = \mp\|\mathbf{x}\|_2 \mathbf{e}_j\,.$$

  Note : Pour éviter le phénomène d'annulation on choisit le signe qui mène à l'addition de deux nombres de même signe ;
  en l'occurrence $\mathbf{v} = \mathbf{x} + \text{signe}(x_j)\|\mathbf{x}\|_2 \mathbf{e}_j$.

Factorisation QR : exemple

Exemple :

$$\begin{pmatrix} \boxed{1} & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \\ 1 & 3 & 9 & 27 \\ 1 & 5 & 25 & 125 \\ 1 & 6 & 36 & 216 \\ 1 & 7 & 49 & 343 \end{pmatrix} \quad \text{avec } \mathbf{v}^{(1)} = \begin{pmatrix} 1+\sqrt{6} \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \quad \text{et } Q_1 := I - 2\frac{\mathbf{v}^{(1)}\ \mathbf{v}^{(1)\,T}}{\|\mathbf{v}^{(1)}\|_2^2}$$

$$\rightarrow Q_1 A = \begin{pmatrix} -\sqrt{6} & -9.7... & -50.6... & -293.9... \\ & -1.1... & -10.9... & -77.5... \\ & -0.1... & -5.9... & -58.5... \\ & 1.8... & 10.0... & -39.4... \\ & 2.8... & 21.0... & 130.5... \\ & 3.8... & 34.0... & 257.5... \end{pmatrix}$$

Factorisation QR : exemple

Exemple :

$$\left(\begin{array}{|c|ccc} 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \\ 1 & 3 & 9 & 27 \\ 1 & 5 & 25 & 125 \\ 1 & 6 & 36 & 216 \\ 1 & 7 & 49 & 343 \end{array}\right) \text{ avec } \mathbf{v}^{(1)} = \begin{pmatrix} 1+\sqrt{6} \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \text{ et } Q_1 := I - 2\frac{\mathbf{v}^{(1)}\ \mathbf{v}^{(1)\,T}}{\|\mathbf{v}^{(1)}\|_2^2}$$

$$\rightarrow Q_1 A = \begin{pmatrix} -\sqrt{6} & \begin{array}{|c} -9.7... \\ -1.1... \\ -0.1... \\ 1.8... \\ 2.8... \\ 3.8... \end{array} & \begin{array}{c} -50.6... \\ -10.9... \\ -5.9... \\ 10.0... \\ 21.0... \\ 34.0... \end{array} & \begin{array}{c} -293.9... \\ -77.5... \\ -58.5... \\ -39.4... \\ 130.5... \\ 257.5... \end{array} \end{pmatrix} \text{ avec } \mathbf{v}^{(2)} = \begin{pmatrix} -9.7... \\ -1.1... - 11.1... \\ -0.1... \\ 1.8... \\ 2.8... \\ 3.8... \end{pmatrix}$$

et $Q_2 := I - 2\frac{\mathbf{v}^{(2)}\ \mathbf{v}^{(2)\,T}}{\|\mathbf{v}^{(2)}\|_2^2}$ $\rightarrow$ $Q_2 Q_1 A = \begin{pmatrix} -0.7... & 9.7... & 85.0... \\ 2.1... & 11.1... & 64.6... & 396.9... \\ 0.02... & -5.1... & -53.4... \\ -0.3... & -1.4... & -32.8... \\ -0.5... & 3.3... & 19.5... \\ -0.6... & 10.1... & 107.8... \end{pmatrix}$

$\rightarrow$ ce n'est pas ce qu'on voulait...

Factorisation QR : exemple

Exemple :

$$\begin{pmatrix} \boxed{1} & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \\ 1 & 3 & 9 & 27 \\ 1 & 5 & 25 & 125 \\ 1 & 6 & 36 & 216 \\ 1 & 7 & 49 & 343 \end{pmatrix} \quad \text{avec } \mathbf{v}^{(1)} = \begin{pmatrix} 1+\sqrt{6} \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \quad \text{et } Q_1 := I - 2\frac{\mathbf{v}^{(1)}\ \mathbf{v}^{(1)\,T}}{\|\mathbf{v}^{(1)}\|_2^2}$$

$$\rightarrow Q_1 A = \begin{pmatrix} -\sqrt{6} & \boxed{-9.7\ldots} & -50.6\ldots & -293.9\ldots \\ & -1.1\ldots & -10.9\ldots & -77.5\ldots \\ & -0.1\ldots & -5.9\ldots & -58.5\ldots \\ & 1.8\ldots & 10.0\ldots & -39.4\ldots \\ & 2.8\ldots & 21.0\ldots & 130.5\ldots \\ & 3.8\ldots & 34.0\ldots & 257.5\ldots \end{pmatrix} \quad \text{avec } \mathbf{v}^{(2)} = \begin{pmatrix} -9.7\ldots \\ -1.1\ldots - 11.1\ldots \\ -0.1\ldots \\ 1.8\ldots \\ 2.8\ldots \\ 3.8\ldots \end{pmatrix}$$

$$\text{et } Q_2 := I - 2\frac{\mathbf{v}^{(2)}\ \mathbf{v}^{(2)\,T}}{\|\mathbf{v}^{(2)}\|_2^2} \quad \rightarrow \quad Q_2 Q_1 A = \begin{pmatrix} -0.7\ldots & & 9.7\ldots & 85.0\ldots \\ 2.1\ldots & 11.1\ldots & 64.6\ldots & 396.9\ldots \\ 0.02\ldots & & -5.1\ldots & -53.4\ldots \\ -0.3\ldots & & -1.4\ldots & -32.8\ldots \\ -0.5\ldots & & 3.3\ldots & 19.5\ldots \\ -0.6\ldots & & 10.1\ldots & 107.8\ldots \end{pmatrix}$$

$\rightarrow$ ce n'est pas ce qu'on voulait...

Factorisation QR : exemple

Exemple :

$$\begin{pmatrix} \boxed{1} & 1 & 1 & 1 \\ \boxed{1} & 2 & 4 & 8 \\ \boxed{1} & 3 & 9 & 27 \\ \boxed{1} & 5 & 25 & 125 \\ \boxed{1} & 6 & 36 & 216 \\ \boxed{1} & 7 & 49 & 343 \end{pmatrix} \text{ avec } \mathbf{v}^{(1)} = \begin{pmatrix} 1+\sqrt{6} \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \text{ et } Q_1 := I - 2\frac{\mathbf{v}^{(1)}\ \mathbf{v}^{(1)\,T}}{\|\mathbf{v}^{(1)}\|_2^2}$$

$$\rightarrow Q_1 A = \begin{pmatrix} -\sqrt{6} & -9.7... & -50.6... & -293.9... \\ & \boxed{-1.1...} & -10.9... & -77.5... \\ & \boxed{-0.1...} & -5.9... & -58.5... \\ & \boxed{1.8...} & 10.0... & -39.4... \\ & \boxed{2.8...} & 21.0... & 130.5... \\ & \boxed{3.8...} & 34.0... & 257.5... \end{pmatrix} \text{ avec } \mathbf{v}^{(2)} = \begin{pmatrix} -1.1... - 5.2... \\ -0.1... \\ 1.8... \\ 2.8... \\ 3.8... \end{pmatrix}$$

$$\text{et } Q_2 := \begin{pmatrix} 1 & \\ & I - 2\frac{\mathbf{v}^{(2)}\ \mathbf{v}^{(2)\,T}}{\|\mathbf{v}^{(2)}\|_2^2} \end{pmatrix} \rightarrow Q_2 Q_1 A = \begin{pmatrix} -\sqrt{6} & -9.7... & -50.6... & -293.9... \\ & 5.2... & 42.3... & 291.0... \\ & & -4.8... & -51.0... \\ & & -5.4... & -67.8... \\ & & -2.7... & -34.1... \\ & & 1.9... & 35.4... \end{pmatrix}$$

Factorisation QR : algorithme

L'algorithme retourne R et une séquence des vecteurs $\mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(n)}$ (laquelle définit Q implicitement)

Algorithme (QR de Householder) :

$R = A$

pour $k = 1, \ldots, n$

$\mathbf{x} = R(k:m, k)$ % vecteur de composantes k à m de la colonne k de R

$\mathbf{v}^{(k)} = \mathbf{x} + \text{sign}(x_1)\|\mathbf{x}\|_2 \mathbf{e}_1$

$\mathbf{v}^{(k)} = \mathbf{v}^{(k)} / \|\mathbf{v}^{(k)}\|_2$ % normaliser $\mathbf{v}^{(k)}$

% multiplier $I - 2\mathbf{v}^{(k)}\ \mathbf{v}^{(k)\,T}$ avec $R(k:m, k:n)$

% coût de l'application $\approx 4(m-k+1)(n-k+1)$

$R(k:m, k:n) = R(k:m, k:n) - \mathbf{v}^{(k)}(2(\mathbf{v}^{(k)\,T} R(k:m, k:n)))$

Coût : $\sum_{k=1}^{n} \underbrace{(\text{coût dernière ligne algorithme})}_{4(m-k+1)(n-k+1) \text{ flops}} + \mathcal{O}(mn) = 2n^2(m - n/3) + \mathcal{O}(mn)$ flops

Factorisation QR : algorithme (suite)

L'algorithme précédent ne forme pas la matrice

$$Q \;=\; Q_1 \cdots Q_n$$

explicitement ; elle est connue implicitement via les vecteurs $\mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(n)}$. Dans certains cas, seul le produit de Q avec un vecteur de dimension m (disons $\mathbf{w}$) est nécessaire.

Algorithme $\mathbf{w} \;:=\; Q\mathbf{w}$:

pour $k = n, \ldots, 1$

$\quad \mathbf{w}(k:m) \;:=\; \mathbf{w}(k:m) - \mathbf{v}^{(k)}(2\,{\mathbf{v}^{(k)}}^T\,\mathbf{w}(k:m))$

Coût : $\displaystyle\sum_{k=1}^{n} 4(m-k+1)$

$= \; 2n(2m-n+1)$

Le produit $Q^T\mathbf{w} \;=\; Q_n \cdots Q_1\mathbf{w}$ est similaire.

Algorithme $\mathbf{w} \;:=\; Q^T\mathbf{w}$:

pour $k = 1, \ldots, n$

$\quad \mathbf{w}(k:m) \;:=\; \mathbf{w}(k:m) - \mathbf{v}^{(k)}(2\,{\mathbf{v}^{(k)}}^T\,\mathbf{w}(k:m))$

Coût : $\displaystyle\sum_{k=1}^{n} 4(m-k+1)$

$= \; 2n(2m-n+1)$

Factorisation QR : algorithme (suite)

Comme

$$Q = \begin{pmatrix} \hat{Q} & * \end{pmatrix}$$

avec $\hat{Q}$ une matrice $m \times n$, on peut utiliser les algorithmes précédents pour évaluer le produit de $\hat{Q}$ avec un vecteur $\hat{\mathbf{w}}$ de dimension n via

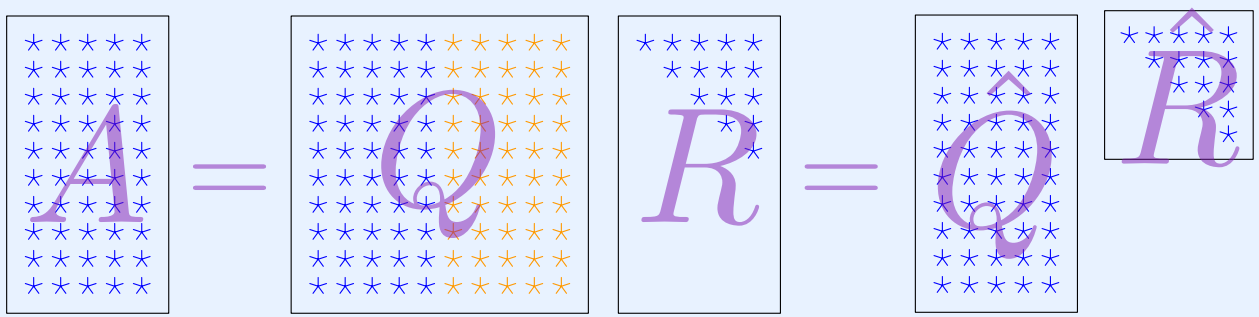
$$\hat{Q}\hat{\mathbf{w}} = Q \begin{pmatrix} \hat{\mathbf{w}} \\ \end{pmatrix};$$

de manière similaire, le produit $\hat{Q}^T$ avec un vecteur $\mathbf{w}$ de dimension m s'obtient avec

$$\begin{pmatrix} \hat{Q}^T\mathbf{w} \\ * \end{pmatrix} = Q^T\mathbf{w};$$

Finalement, les matrices Q et $\hat{Q}$ peuvent aussi être formées explicitement, leur $j$ème colonne étant $Q\mathbf{e}_j$.

Exemple p.2 : ($m = 10$, $n = 5$)



Interlude : propriétés de la norme euclidienne

Propriété 1 : soient A une matrice symétrique et λ_i, $i = 1, ..., n$, ses valeurs propres. Alors

$$\|A\|_2 = \max_i |\lambda_i| = \max_{\mathbf{v}} \frac{|\mathbf{v}^T A \mathbf{v}|}{\mathbf{v}^T \mathbf{v}} .$$

Soit $\mathbf{p}_i$ le vecteur propre normalisé associé à λ_i, $i = 1, ..., n$. Comme la matrice A est symétrique, l'ensemble de ces vecteurs forment une base orthonormale et tout vecteur $\mathbf{v}$ a une représentation $\mathbf{v} = \sum_{i=1}^n \alpha_i \mathbf{p}_i$ dans cette base. On a alors

$$\|A\|_2^2 = \max_{\mathbf{v}} \frac{\|A\mathbf{v}\|_2^2}{\|\mathbf{v}\|_2^2} = \max_{\mathbf{v}} \frac{\mathbf{v}^T A^2 \mathbf{v}}{\mathbf{v}^T \mathbf{v}} = \max_{\alpha_1,...,\alpha_n} \frac{\sum_{i=1}^n \lambda_i^2 \alpha_i^2}{\sum_{i=1}^n \alpha_i^2} = \max_i |\lambda_i|^2 .$$

Par analogie avec les deux dernières égalités on a aussi

$$\max_{\mathbf{v}} \frac{|\mathbf{v}^T A \mathbf{v}|}{\mathbf{v}^T \mathbf{v}} = \max_{\alpha_1,...,\alpha_n} \frac{\left|\sum_{i=1}^n \lambda_i \alpha_i^2\right|}{\sum_{i=1}^n \alpha_i^2} = \max_i |\lambda_i| .$$

Propriété 2 : pour toute matrice A rectangulaire

$$\|A^T A\|_2 = \|A\|_2^2 .$$

$$\|A^T A\|_2 = \max_{\mathbf{v}} \frac{\left|\mathbf{v}^T A^T A \mathbf{v}\right|}{\mathbf{v}^T \mathbf{v}} = \max_{\mathbf{v}} \frac{\|A\mathbf{v}\|_2^2}{\|\mathbf{v}\|_2^2} .$$

Interlude : propriétés de la norme euclidienne

Propriété 3 : pour toute matrice A rectangulaire

$$\|A^T\|_2 \;=\; \|A\|_2\,.$$

Considérons d'abord le cas particulier d'un vecteur $\mathbf{w}$. En notant avec θ l'angle entre les vecteurs $\mathbf{w}$ et $\mathbf{v}$, on a

$$\|\mathbf{w}^T\|_2 \;=\; \max_{\mathbf{v}} \frac{|\mathbf{w}^T\mathbf{v}|}{\|\mathbf{v}\|_2} \;=\; \max_{\theta} \frac{\|\mathbf{w}\|_2\|\mathbf{v}\|_2|\cos(\theta)|}{\|\mathbf{v}\|_2} \;=\|\mathbf{w}\|_2\,.$$

Maintenant, comme pour deux matrices A, B la norme du produit satisfait

$$\|AB\|_2 \;\leq\; \|A\|_2\|B\|_2\,,$$

on a

$$\|A^T\|_2 \;=\; \max_{\mathbf{v}} \frac{\|A^T\mathbf{v}\|_2}{\|\mathbf{v}\|_2} \;=\; \max_{\mathbf{v}} \frac{\|\mathbf{v}^TA\|_2}{\|\mathbf{v}^T\|_2} \;\leq\; \max_{\mathbf{v}} \frac{\|\mathbf{v}^T\|_2\|A\|_2}{\|\mathbf{v}^T\|_2} \;=\; \|A\|_2\,.$$

En même temps $A \;=\; A^{T\,T}$ et donc $\|A\|_2 \;\leq\; \|A^T\|_2$, d'où l'égalité.

Systèmes surdéterminés : généralités

On considère les systèmes surdéterminés

$$\begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + \cdots + a_{2n}x_n &= b_2 \\ \cdots & \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m \end{cases}$$

avec

- au moins autant d'équations que d'inconnues ($m \geq n$)
- le rang maximal (la matrice du système est de rang n).

Sous forme matricielle :

$$A\mathbf{x} = \mathbf{b}\,.$$

En règle générale, le système n'a pas de solution, et donc, le résidu

$$\mathbf{r} = \mathbf{b} - A\mathbf{x}$$

n'est pas nul. On parle alors de la solution $\mathbf{x}$ au sens des moindres carrés si elle minimise la norme euclidienne du résidu ; c'est-à-dire si

$$\|\mathbf{b} - A\mathbf{x}\|_2 = \min_{\mathbf{y}} \|\mathbf{b} - A\mathbf{y}\|_2\,.$$

Systèmes surdéterminés : solution formelle

On procède en deux étapes.

1. On détermine les points critiques de

$$f(\mathbf{x}) \;=\; \|\mathbf{b} - A\mathbf{x}\|_2^2\,,$$

c'est-à-dire tous les $\mathbf{x}$ tels que $\nabla f(\mathbf{x}) = 0$, ou encore

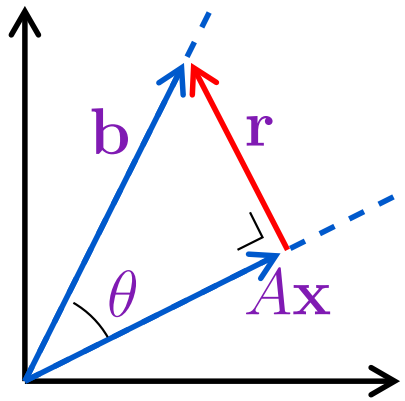
$$\boxed{A^T A\mathbf{x} \;=\; A^T\mathbf{b}}\,. \tag{1}$$

C'est le système d'équations normales. Si le rang de A est maximal (égal à n), alors le rang de $A^T A$ est maximal (égal à n) et $A^T A$ est inversible ; en particulier, il y a un et un seul point critique.

2. On montre que ce $\mathbf{x}$ minimise bien f (et minimise globalement !) :

$$\begin{aligned}
\|\mathbf{b} - A\mathbf{y}\|_2^2 &= \|\mathbf{b} - A\mathbf{x} + A(\mathbf{x} - \mathbf{y})\|_2^2 \\
&= \|\mathbf{b} - A\mathbf{x}\|_2^2 + 2(\mathbf{x} - \mathbf{y})^T \underbrace{A^T(\mathbf{b} - A\mathbf{x})}_{=0} + \underbrace{\|A(\mathbf{x} - \mathbf{y})\|_2^2}_{\geq 0} \\
&\geq \|\mathbf{b} - A\mathbf{x}\|_2^2\,.
\end{aligned}$$

Interprétation géométrique



Une autre manière d'interpréter la solution :

$\mathbf{r} = \mathbf{b} - A\mathbf{x}$ est minimisé si il est orthogonal à $A\mathbf{w}$ pour tout $\mathbf{w} \in \mathbb{R}^n$;

et donc si

$$A^T\mathbf{r} \;=\; A^T(\mathbf{b} - A\mathbf{x}) \;=\; \mathbf{0}\,.$$

L'angle θ entre le vecteur $\mathbf{b}$ et l'hyperplan $A\mathbf{w}$, $\mathbf{w} \in \mathbb{R}^n$ mesure la capacité de $A\mathbf{x}$ à approcher $\mathbf{b}$.

Si $\mathbf{r}$ est le résidu minimal, alors

$$\sin(\theta) \;=\; \frac{\|\mathbf{r}\|_2}{\|\mathbf{b}\|_2}\,.$$

Equations normales

La résolution au sens des moindres carrés est donc équivalente à

$$\boxed{A^T A\mathbf{x} \;=\; A^T\mathbf{b}}\,. \tag{1}$$

La matrice $A^T A$ du système est :

- symétrique, car
$$(A^T A)^T \;= A^T A$$
- définie positive, car
$$\mathbf{v}^T A^T A\mathbf{v} \;=\; \|A\mathbf{v}\|_2^2 \;\geq\; 0$$
et
$$\mathbf{v}^T A^T A\mathbf{v} \;=\; 0 \quad\Leftrightarrow\quad A\mathbf{v} \;=\; \mathbf{0} \quad\overset{A \text{ de rang maximal}}{\Longleftrightarrow}\quad \mathbf{v} \;=\; \mathbf{0}\,.$$

Rappel : une matrice M est définie positive si pour tout $\mathbf{v} \neq \mathbf{0}$ on a

$$\mathbf{v}^T M\mathbf{v} \;>\; 0\,.$$

Conditionnement matriciel : généralisation

La résolution au sens des moindres carrés est donc équivalente à

$$\boxed{A^T A\mathbf{x} \;=\; A^T\mathbf{b}}\,. \tag{1}$$

Pour une matrice A régulière, le conditionnement (en norme 2) est défini par

$$\kappa(A) \;=\; \|A^{-1}\|\|A\|\,.$$

Si le nombre de lignes est supérieur au nombre de colonnes, la matrice A n'est pas inversible ; néanmoins, si $\mathrm{rang}(A) = n$, elle possède un pseudo-inverse

$$A^\dagger \;=\; (A^T A)^{-1} A^T\,;$$

notez que la solution de (1) est $\mathbf{x} = A^\dagger\mathbf{b}$. Le conditionnement de A (en norme euclidienne) est alors défini par

$$\boxed{\kappa(A) \;=\; \|A^\dagger\|_2\|A\|_2}\,.$$

Pour ce qui est du pseudo-inverse, l'identité suivante nous sera utile

$$\|(A^T A)^{-1}\|_2 \;=\; \left\|A^\dagger\, A^{\dagger\,T}\right\|_2 \;=\; \left\|A^{\dagger\,T}\right\|_2^2 \;=\; \|A^\dagger\|_2^2\,. \tag{2}$$

Systèmes surdéterminés : conditionnement

La résolution au sens des moindres carrés est donc équivalente à

$$A^T A\mathbf{x} \;=\; A^T\mathbf{b}\,.$$

Le conditionnement d'un problème est

$$\kappa \;=\; \sup \frac{\text{erreurs relatives résultat}}{\text{erreurs relatives données}}\,,$$

avec les erreurs données $\to 0$.

Pour un problème aux moindres carrés :

- données : A, $\mathbf{b}$ (avec perturbations $A + \delta A$, $\mathbf{b} + \delta\mathbf{b}$)
- résultat : $\mathbf{x}$ ($\mathbf{x} + \delta\mathbf{x}$ pour le système perturbé)

Systèmes surdéterminés : conditionnement

La résolution au sens des moindres carrés est donc équivalente à

$$A^T A\mathbf{x} \;=\; A^T\mathbf{b}\,.$$

Le conditionnement d'un problème est

$$\kappa \;=\; \sup \frac{\text{erreurs relatives résultat}}{\text{erreurs relatives données}}\,,$$

avec les erreurs données $\to 0$.

On peut montrer (cf. Annexe) que

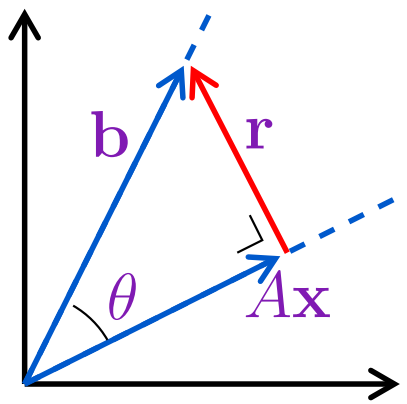
pour erreurs δA sur A : $\kappa_{MC,A} \;\leq \kappa(A) + \dfrac{\kappa(A)^2\|\mathbf{r}\|_2}{\|A\|_2\|\mathbf{x}\|_2} \leq\; \kappa(A) + \kappa(A)^2\tan(\theta)$

pour erreurs $\delta\mathbf{b}$ sur $\mathbf{b}$: $\kappa_{MC,\mathbf{b}} \;\leq\; \kappa(A)/\cos(\theta)$

Conclusion :

- $\theta \approx 0 \quad \Rightarrow \quad \kappa_{MC,A} \approx \kappa_{MC,\mathbf{b}} \approx \kappa(A)$
- $0 \ll \theta \ll \frac{\pi}{2} \quad \Rightarrow \quad \kappa_{MC,A} \approx \kappa(A)^2$

  $\kappa_{MC,\mathbf{b}} \approx \kappa(A)$
- $\theta \to \frac{\pi}{2} \quad \Rightarrow \quad \kappa_{MC,A}\,,\ \kappa_{MC,\mathbf{b}} \to \infty$ (car $\mathbf{x} \approx \mathbf{0}$)



Méthode des équations normales (version LU)

La méthode consiste à former le système

$$A^T A\mathbf{x} \;=\; A^T\mathbf{b} \tag{1}$$

explicitement et en calculer sa factorisation LU avec pivotage.

Algorithme :

1. former $A^T A$, $A^T\mathbf{b}$ $\quad (\approx mn^2$ (symétrie) $+\ 2mn$ flops)
2. calculer la factorisation LU avec piv. $LU \;=\; P(A^T A)$ $\quad (\approx 2n^3/3$ flops)
3. résoudre $L\mathbf{y} = P(A^T\mathbf{b})$ et $U\mathbf{x} = \mathbf{y}$ $\quad (\approx 2n^2$ flops)

- dans la situation (habituelle) où $m \gg n$ le coût est dominé par mn^2
- le conditionnement du système (1) est (en utilisant (2) et Propriété 2)

$$\kappa(A^T A) \;=\; \|(A^T A)^{-1}\|_2 \|A^T A\|_2 \;=\; \|A^\dagger\|_2^2 \|A\|_2^2 \;=\; \kappa(A)^2\,;$$

pour rappel, si l'angle θ entre $\mathbf{b}$ et $A\mathbf{x}$ est proche de 0, $\kappa_{MC,A} \approx \kappa(A)$ et une perte de précision est possible si on utilise cette méthode lorsque la matrice est mal conditionnée (c'est-à-dire si $\kappa(A) \gg 1$) ; dans ce cas la méthode est instable.

Méthode de la factorisation QR

La méthode utilise la factorisation $\hat{Q}\hat{R}$ réduite de A ; elle est basée sur le fait que la solution $\mathbf{x}$ du système

$$A^T A\mathbf{x} \;=\; A^T\mathbf{b}$$

satisfait aussi

$$\mathbf{x} \;=\; (A^T A)^{-1}A^T\mathbf{b} \;=\; (\hat{R}^T\cancel{\hat{Q}^T\hat{Q}}\hat{R})^{-1}\hat{R}^T\hat{Q}^T\mathbf{b} \;=\; \hat{R}^{-1}\cancel{\hat{R}^{-T}\hat{R}^T}\hat{Q}^T\mathbf{b} \;=\; \hat{R}^{-1}\hat{Q}^T\mathbf{b}\,;$$

Algorithme :

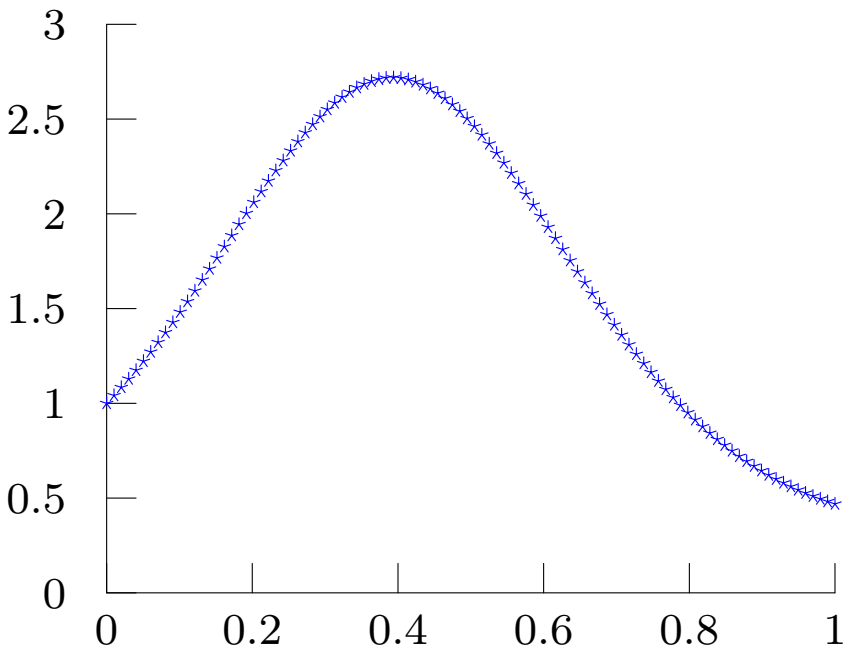
1. calculer la factorisation $\hat{Q}\hat{R} = A$ $\quad(\approx 2n^2(m - n/3)$ flops$)$
2. former $\hat{Q}^T\mathbf{b}$ $\quad(\approx 4mn$ flops$)$
3. résoudre $\hat{R}\mathbf{x} = \hat{Q}^T\mathbf{b}$ $\quad(\approx n^2$ flops$)$

- $A^T A$ ne doit pas être formée ;
- dans la situation (habituelle) où $m \gg n$ le coût est dominé par $2mn^2$
- si la factorisation $\hat{Q}\hat{R}$ réduite est bien calculée par la méthode de Householder, le présent algorithme a la stabilité inverse

Comparaison des méthodes : exemple

Exemple[1] : interpolation de la fonction `exp(sin(4*t))` aux points repartis uniformément sur l'intervalle $[0, 1]$ par un polynôme de degré 14

```
m = 100; n = 15;
t = 0:1/(m-1):1; t = t';
A = [];
for i=1:n
  A = [A t.^(i-1)];
end
b = exp(sin(4*t));
  % valeur exacte de x(15)
  % obtenue via des outils
  % de précision étendue
x15ex = 2006.787453080206
```



1. Repris de L. N. Trefethen, D. Bau, *Numerical Linear Algebra*, SIAM, Philadelphia, PA, 1997

Comparaison des méthodes : exemple

Exemple [1] : interpolation de la fonction `exp(sin(4*t))` aux points repartis uniformément sur l'intervalle $[0, 1]$ par un polynôme de degré 14

Instruction Octave

```
x = A\b;
x(15)/x15ex
```

```
ans = 1.00000007318865
```

Méthode QR (Householder)

```
[Q,R] = qr(A,0);
x = R\(Q'*b);
x(15)/x15ex
```

```
ans = 1.00000007318102
```

Méthode équations normales (version LU)

```
[L U P] = lu(A'*A);
x = U\(L\(P*(A'*b)));
x(15)/x15ex
```

```
ans = 1.56194368637586
```

1. Repris de L. N. Trefethen, D. Bau, *Numerical Linear Algebra*, SIAM, Philadelphia, PA, 1997

Comparaison des méthodes : exemple

Exemple [1] : interpolation de la fonction `exp(sin(4*t))` aux points repartis uniformément sur l'intervalle $[0, 1]$ par un polynôme de degré 14

Explication : pour ce problème

- $\kappa_{MC,A} \leq 3.2\,10^{10}$
- $\kappa_{MC,\mathbf{b}} \leq 2.3\,10^{10}$

et donc en double précision une méthode qui a la stabilité directe doit fournir un résultat avec une précision relative d'environ $u\kappa_{MC} \approx 10^{-6}$. Par contre

- $\kappa(A)^2 = 2.3\,10^{20}$.

et donc la réponse obtenue par la méthode des équations normales risque de ne pas être (et n'est en effet pas !) précise.

```
kA = cond(A)
x=A\b; r = A*x-b;
theta = asin(norm(r)/norm(b));
c = norm(r)/norm(A)/norm(x);
K_MCA_max = kA + kA^2 * c
K_MCB_max = kA/cos(theta)
```

```
K_MCA_max = 3.1909e+10
K_MCB_max = 2.2718e+10
```

1. Repris de L. N. Trefethen, D. Bau, *Numerical Linear Algebra*, SIAM, Philadelphia, PA, 1997

Annexe : Conditionnement (dérivation)

Le problème aux moindres carrés est équivalent à

$$A^T A\mathbf{x} \;=\; A^T\mathbf{b}\,.$$

Le conditionnement d'un problème est

$$\kappa \;=\; \sup \frac{\text{erreurs relatives résultat}}{\text{erreurs relatives données}}\,,$$

avec les erreurs données $\to 0$.

Cas 1 : perturbations $\delta\mathbf{b}$ de $\mathbf{b}$ ($\mathbf{x}+\delta\mathbf{x}$ est la solution du système perturbé)

$$\begin{aligned}
& A^T A(x+\delta\mathbf{x}) \;=\; A^T(\mathbf{b}+\delta\mathbf{b}) \\
\Rightarrow\quad & A^T A\delta\mathbf{x} \;=\; A^T\delta\mathbf{b} && \text{(soustraction de (1))} \\
\Rightarrow\quad & \|\delta\mathbf{x}\|_2 \;=\; \|A^\dagger\delta\mathbf{b}\|_2 \;\le\; \|A^\dagger\|_2\,\|\delta\mathbf{b}\|_2 && \text{(norme matricielle)}
\end{aligned}$$

et donc

$$\kappa_{MC,\mathbf{b}} \;=\; \sup_{\|\delta\mathbf{b}\|} \frac{\|\delta\mathbf{x}\|_2/\|\mathbf{x}\|_2}{\|\delta\mathbf{b}\|_2/\|\mathbf{b}\|_2} \;\le\; \kappa(A)\frac{\|\mathbf{b}\|_2}{\|A\|_2\|\mathbf{x}\|_2} \;\le\; \kappa(A)\frac{\|\mathbf{b}\|_2}{\|A\mathbf{x}\|_2} \;=\; \frac{\kappa(A)}{\cos(\theta)}$$

Annexe : Conditionnement (dérivation)

Le problème aux moindres carrés est équivalent à

$$A^T A\mathbf{x} \;=\; A^T\mathbf{b}\,.$$

Le conditionnement d'un problème est

$$\kappa \;=\; \sup \frac{\text{erreurs relatives résultat}}{\text{erreurs relatives données}}\,,$$

avec les erreurs données $\to 0$.

Cas 2 : perturbations δA de A ($\mathbf{x}+\delta\mathbf{x}$ est la solution du système perturbé)

$$(A+\delta A)^T(A+\delta A)(x+\delta\mathbf{x}) \;=\; (A+\delta A)^T\mathbf{b}$$

$$\Rightarrow \quad A^T A\delta\mathbf{x} \;+A^T\delta A\;\mathbf{x}+\delta A^T A\mathbf{x}+\underbrace{\cancel{\text{le reste}}}_{\text{second ordre}} \;=\; \delta A^T\mathbf{b} \quad \text{(soustraction de (1))}$$

$$\Rightarrow \quad \|\delta\mathbf{x}\|_2 \;\le\; \|A^\dagger\|_2\|\delta A\|_2\|\mathbf{x}\|_2 + \underbrace{\|(A^TA)^{-1}\|_2}_{\|A^\dagger\|_2^2}\underbrace{\|\delta A^T\|_2}_{\|\delta A\|_2}\underbrace{\|\mathbf{b}-A\mathbf{x}\|_2}_{\|\mathbf{r}\|}$$

et donc

$$\kappa_{MC,A} \;=\; \sup_{\|\delta A\|} \frac{\|\delta\mathbf{x}\|_2/\|\mathbf{x}\|_2}{\|\delta A\|_2/\|A\|_2} \;\le\; \kappa(A)+\kappa(A)^2\frac{\|\mathbf{r}\|_2}{\|A\|_2\|\mathbf{x}\|_2} \;\le\; \kappa(A)+\kappa(A)^2\,\overbrace{\frac{\|\mathbf{r}\|_2}{\|A\mathbf{x}\|_2}}^{\tan(\theta)}$$