

## Chapitre 3 : FACTORISATION QR ET SYSTÈMES SURDÉTERMINÉS

### 1 Factorisation QR

- Factorisation QR : généralités
- Méthode de Householder : principe
- Transformation de Householder
- Factorisation QR : exemple
- Factorisation QR : algorithme

### 2 Systèmes surdéterminés

- Interlude : propriétés de la norme euclidienne
- Systèmes surdéterminés : généralités
- Systèmes surdéterminés : solution formelle
- Interprétation géométrique
- Equations normales
- Conditionnement
- Méthode des équations normales (version LU)
- Méthode de la factorisation QR
- Comparaison des méthodes : exemple
- Annexe : conditionnement

# 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

$$A = QR.$$

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

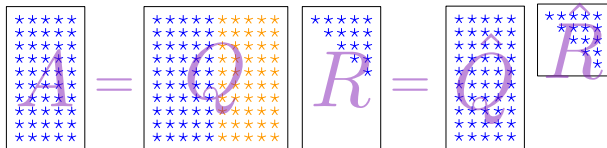
$$R = \begin{pmatrix} \hat{R} \\ 0 \end{pmatrix}, \quad \text{avec } \hat{R} \text{ de dimensions } n \times n,$$

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)`)

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

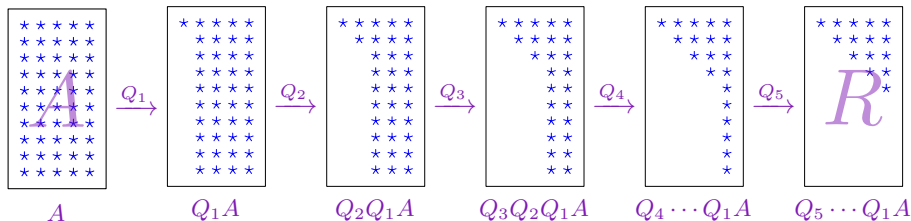
EXEMPLE :

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



# MÉTHODE DE HOUSEHOLDER : PRINCIPE

## PRINCIPE DE BASE :



Si  $Q_i$ ,  $i = 1, \dots, 5$ , 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  $R = Q_5 \cdots Q_1A = Q^TA$  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

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

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

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

# 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

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

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

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

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 : (matrice de l'Exercice 4 de la Séance 1)

$$\begin{pmatrix} \boxed{1} & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \\ 1 & 3 & 9 & 27 \\ 1 & 5 & 25 & 125 \\ 1 & 6 & 36 & 216 \\ 1 & 7 & 49 & 343 \end{pmatrix} \text{ avec } \mathbf{v}^{(1)} = \begin{pmatrix} 1 + \sqrt{6} \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \text{ et } Q_1 := I - 2 \frac{\mathbf{v}^{(1)} \mathbf{v}^{(1)T}}{\|\mathbf{v}^{(1)}\|_2^2}$$
$$\rightarrow Q_1 A = \begin{pmatrix} -\sqrt{6} & -9.7... & -50.6... & -293.9... \\ -1.1... & -10.9... & -77.5... & ... \\ -0.1... & -5.9... & -58.5... & ... \\ 1.8... & 10.0... & -39.4... & ... \\ 2.8... & 21.0... & 130.5... & ... \\ 3.8... & 34.0... & 257.5... & ... \end{pmatrix}$$

# FACTORISATION QR : EXEMPLE

EXEMPLE : (matrice de l'Exercice 4 de la Séance 1)

$$\begin{pmatrix} \boxed{1} & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \\ 1 & 3 & 9 & 27 \\ 1 & 5 & 25 & 125 \\ 1 & 6 & 36 & 216 \\ 1 & 7 & 49 & 343 \end{pmatrix} \text{ avec } \mathbf{v}^{(1)} = \begin{pmatrix} 1 + \sqrt{6} \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \text{ et } Q_1 := I - 2 \frac{\mathbf{v}^{(1)} \mathbf{v}^{(1)T}}{\|\mathbf{v}^{(1)}\|_2^2}$$
$$\rightarrow Q_1 A = \begin{pmatrix} -\sqrt{6} & \boxed{-9.7\dots} & -50.6\dots & -293.9\dots \\ -1.1\dots & -10.9\dots & -77.5\dots & \\ -0.1\dots & -5.9\dots & -58.5\dots & \\ 1.8\dots & 10.0\dots & -39.4\dots & \\ 2.8\dots & 21.0\dots & 130.5\dots & \\ 3.8\dots & 34.0\dots & 257.5\dots & \end{pmatrix} \text{ avec } \mathbf{v}^{(2)} = \begin{pmatrix} -9.7\dots \\ -1.1\dots - 11.1\dots \\ -0.1\dots \\ 1.8\dots \\ 2.8\dots \\ 3.8\dots \end{pmatrix}$$
$$\text{ et } Q_2 := I - 2 \frac{\mathbf{v}^{(2)} \mathbf{v}^{(2)T}}{\|\mathbf{v}^{(2)}\|_2^2} \rightarrow Q_2 Q_1 A = \begin{pmatrix} -0.7\dots & 9.7\dots & 85.0\dots \\ 2.1\dots - 11.1\dots & 64.6\dots & 396.9\dots \\ 0.02\dots & -5.1\dots & -53.4\dots \\ -0.3\dots & -1.4\dots & -32.8\dots \\ -0.5\dots & 3.3\dots & 19.5\dots \\ -0.6\dots & 10.1\dots & 107.8\dots \end{pmatrix}$$

→ ce n'est pas ce qu'on voulait...

# FACTORISATION QR : EXEMPLE

EXEMPLE : (matrice de l'Exercice 4 de la Séance 1)

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

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

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

→ ce n'est pas ce qu'on voulait...



# FACTORISATION QR : EXEMPLE

EXEMPLE : (matrice de l'Exercice 4 de la Séance 1)

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

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

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

# FACTORISATION QR : ALGORITHME

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

ALGORITHME (QR DE HOUSEHOLDER) :

$R = A$

pour  $k = 1, \dots, n$

$\mathbf{x} = R(k : m, k)$     % vecteur de composantes  $k$  à  $n$  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  $\mathbf{v}^{(k)}$

% appliquer  $I - 2\mathbf{v}^{(k)} \mathbf{v}^{(k)T}$

$R(k : m, k : n) = R(k : m, k : n) - \mathbf{v}^{(k)} \underbrace{(2 (\mathbf{v}^{(k)})^T R(k : m, k : n))}_{\text{vecteur ligne, } (2m-2k+1)(n-k+1) \text{ flops}}$

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

# 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  $\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(n)}$ . 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  $k = n, \dots, 1$

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

$$\begin{aligned} \text{Coût} : \sum_{k=1}^n 4(m-k+1) \\ = 2n(2m-n+1) \end{aligned}$$

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

ALGORITHME  $\mathbf{w} := Q^T \mathbf{w}$  :

pour  $k = 1, \dots, n$

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

$$\begin{aligned} \text{Coût} : \sum_{k=1}^n 4(m-k+1) \\ = 2n(2m-n+1) \end{aligned}$$

# FACTORISATION QR : ALGORITHME (SUITE)

Comme

$$Q = ( \hat{Q} \quad * )$$

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 de  $\mathbf{w}$  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 :  
( $m = 10, n = 5$ )

$$\begin{bmatrix} \text{*****} \\ \text{*****} \\ \text{*****} \\ \text{*****} \\ \text{*****} \\ \text{*****} \\ \text{*****} \\ \text{*****} \\ \text{*****} \\ \text{*****} \end{bmatrix} = \begin{bmatrix} \text{*****} & \text{*****} \\ \text{*****} & \text{*****} \\ \text{*****} & \text{*****} \\ \text{*****} & \text{*****} \\ \text{*****} & \text{*****} \\ \text{*****} & \text{*****} \\ \text{*****} & \text{*****} \\ \text{*****} & \text{*****} \\ \text{*****} & \text{*****} \\ \text{*****} & \text{*****} \end{bmatrix} = \begin{bmatrix} \text{*****} \\ \text{*****} \\ \text{*****} \\ \text{*****} \\ \text{*****} \\ \text{*****} \\ \text{*****} \\ \text{*****} \\ \text{*****} \\ \text{*****} \end{bmatrix} = \begin{bmatrix} \text{*****} \\ \text{*****} \\ \text{*****} \\ \text{*****} \\ \text{*****} \\ \text{*****} \\ \text{*****} \\ \text{*****} \\ \text{*****} \\ \text{*****} \end{bmatrix}$$

# INTERLUDE : PROPRIÉTÉS DE LA NORME EUCLIDIENNE

**PROPRIÉTÉ 1** : soient  $A$  une matrice symétrique et  $\lambda_i, i = 1, \dots, n$ , 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  $\lambda_i, i = 1, \dots, n$ . 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

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

Par analogie avec les deux dernières égalités on a aussi

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

**PROPRIÉTÉ 2** : pour toute matrice  $A$  rectangulaire

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

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

# 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  $\mathbf{w}$  :

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

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

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

on a

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

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

# SYSTÈMES SURDÉTERMINÉS : GÉNÉRALITÉS

On considère les systèmes **surdéterminés**

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

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 :

$$\mathbf{Ax} = \mathbf{b}.$$

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

$$\mathbf{r} = \mathbf{b} - \mathbf{Ax}$$

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

$$\|\mathbf{b} - \mathbf{Ax}\|_2 = \min_{\mathbf{y}} \|\mathbf{b} - \mathbf{Ay}\|_2.$$

# 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

$$\boxed{A^T A \mathbf{x} = A^T \mathbf{b}}. \quad (1)$$

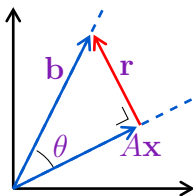
C'est le système d'équations normales. Si le rang de  $A$  est maximal (égal à  $n$ ), alors le rang de  $A^T A$  est maximal (égal à  $n$ ) et  $A^T A$  est inversible ; en particulier, il y a un et un seul point critique.

- 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  $\theta$  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

$$\boxed{A^T A \mathbf{x} = A^T \mathbf{b}}. \quad (1)$$

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

$$\mathbf{v}^T A^T A \mathbf{v} = 0 \quad \Leftrightarrow \quad A \mathbf{v} = \mathbf{0} \quad \xleftrightarrow{A \text{ de rang maximal}} \quad \mathbf{v} = \mathbf{0}.$$

**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

$$\boxed{A^T A \mathbf{x} = A^T \mathbf{b}}. \quad (1)$$

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  $\text{rang}(A) = n$ , 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

$$\|(A^T A)^{-1}\|_2 = \|A^\dagger A^{\dagger T}\|_2 = \|A^{\dagger T}\|_2^2 = \|A^\dagger\|_2^2. \quad (2)$$

# 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  $\rightarrow 0$ .

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  $\rightarrow 0$ .

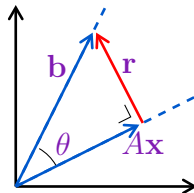
On peut montrer (cf. Annexe) que

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

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

CONCLUSION :

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



# MÉTHODE DES ÉQUATIONS NORMALES (VERSION LU)

La méthode consiste à former le système

$$A^T A \mathbf{x} = A^T \mathbf{b} \quad (1)$$

explicitement et en calculer sa factorisation LU avec pivotage.

## ALGORITHME :

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

- 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  $\theta$  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

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

## ALGORITHME :

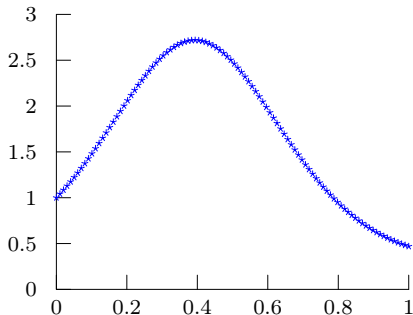
- ❶ calculer la factorisation  $\hat{Q}\hat{R} = A$  ( $\approx 2n^2(m - n/3)$  flops)
- ❷ former  $\hat{Q}^T \mathbf{b}$  ( $\approx 4mn$  flops)
- ❸ résoudre  $\hat{R}\mathbf{x} = \hat{Q}^T \mathbf{b}$  ( $\approx n^2$  flops)

- $A^T A$  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**<sup>1</sup> : 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**<sup>1</sup> : 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<sup>1</sup>** : 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

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

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

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

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  $\rightarrow 0$ .

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

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

$$\Rightarrow A^T A \delta \mathbf{x} = A^T \delta \mathbf{b} \quad (\text{soustraction de (1)})$$

$$\Rightarrow \|\delta \mathbf{x}\|_2 = \|A^\dagger \delta \mathbf{b}\|_2 \leq \|A^\dagger\|_2 \|\delta \mathbf{b}\|_2 \quad (\text{norme matricielle})$$

et donc

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

## 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  $\rightarrow 0$ .

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

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

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

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

et donc

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