MT09-Analyse numérique élémentaire

Chapitre 3 : Résolution des problèmes de moindres carrés

Équipe de Mathématiques Appliquées

UTC

Juin 2007

Chapitre III

Résolution des problèmes de moindres carrés

III.1 Formulation générale des problèmes de moindres carrés 3
III.2 Approche algébrique du problème de moindres carrés 10
III.3 Résolution des problèmes de moindres carrés par "QR". 17
Documents du chapitre III 24
Exercices du chapitre III 31

Sommaire
Concepts

Exemples
Exercices
Documents

III.1 Formulation générale des problèmes de moindres carrés

III.1.1 Un exemple : un problème de lissage. 4
III.1.2 Formulation matricielle 6
III.1.3 Les équations normales 8

III.1.1 Un exemple : un problème de lissage.

Exercices :
Exercice III.1

Si on se donne une famille de points du plan $(t_i, b_i)_{1\leq i\leq m}$ (les t_i étant distincts) alors il existe un unique polynôme $P(t)$ de degré inférieur ou égal à $m-1$ tel que

$$P(t_i) = b_i, \quad i = 1, \dots, m,$$

qui est le polynôme d'interpolation des points (t_i, b_i). Si le nombre de points m est trop grand, ou si les ordonnées sont bruitées, on préfère en général chercher une fonction $f(t)$ qui, dans une

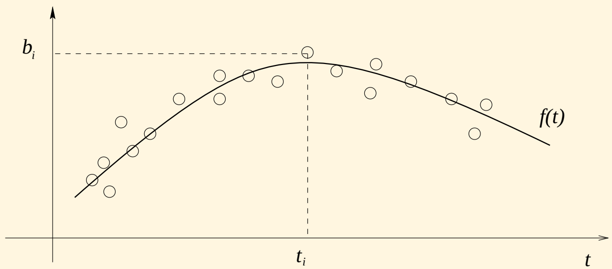

FIG. III.1.1: un problème de lissage

classe donnée (polynômes, fractions rationnelles, polynômes trigonométriques, exponentielles

Un exemple : un problème de lissage.

...), approche “au mieux” les points (t_i, b_i), on parle alors d’approximation, de lissage ou bien de *régression* (voir la figure III.1.1).

Soit donc une famille de fonctions

$$f_1(t), f_2(t), \ldots, f_n(t),\ n \leq m,$$

linéairement indépendantes. Etant donné n nombres réels $x_1, x_2, \ldots, x_n$ on peut introduire le nombre $E(x)$

$$E(x) = \sum_{i=1}^{m} [f(t_i) - b_i]^2,$$

avec $f(t) = \sum_{k=1}^{n} x_k f_k(t)$. La quantité $E(x)$ représente la somme des erreurs quadratiques entre les valeurs données et celles prises par f aux points t_i. Le problème d’approximation se formule alors de la façon suivante :

Trouver $\widehat{x} \in \mathbb{R}^n$*, tel que* $E(\widehat{x}) \leq E(x),\ \forall x \in \mathbb{R}^n.$

Par exemple si l’on désire faire de la régression polynômiale, c’est-à-dire prendre pour $f(t)$ un polynôme de degré $\leq n - 1$, on a

$$f_k(t) = t^{k-1},\ k = 1, \ldots, n,\quad f(t) = x_1 + x_2 t + \ldots + x_n t^{n-1}.$$

L’écriture inhabituelle du polynôme avec x_k comme coefficients permet de mettre en évidence que les inconnues du problème de moindres carrés sont ces coefficients.

III.1.2 Formulation matricielle

Exercices :
Exercice III.2

Cours :
Lissage

On suppose que l'on a à résoudre un système linéaire $Ax = b$ ($A \in \mathcal{M}_{mn}$), avec un second membre b non nul et on suppose que le nombre d'équations est supérieur strictement au nombre d'inconnues ($m > n$). Dans la plupart des cas, ce système n'a pas de solution. On cherche alors une approximation de la solution qui réduise la différence $Ax - b$. Un des choix possible est de minimiser la norme euclidienne de cette différence. Dans tout ce chapitre, nous n'utiliserons que la norme euclidienne.

Définition III.1.1. *Soit $A \in M_{mn}(\mathbb{R})$ et $b \in \mathbb{R}^m$ donnés. On appelle problème de moindres carrés le problème*

$$\min_{x \in \mathbf{R}^n} \|Ax - b\|_2^2 . \tag{III.1.1}$$

On notera $\widehat{x}$ la solution de ce problème. On montrera dans la suite qu'elle existe, et qu'elle est unique sous l'hypothèse fondamentale suivante :

$$\textit{Les colonnes de A sont linéairement indépendantes}. \tag{III.1.2}$$

ou, de façon équivalente

$$\text{rang } A = n.$$

Un cas particulier est le cas $m = n$ et A inversible, alors $\widehat{x}$ est la solution unique de $Ax = b$, mais ce n'est pas ce cas particulier qui nous intéresse dans ce chapitre.

Un autre exemple, pour le problème de lissage, est introduit dans le paragraphe référencé. Il n'est pas possible en général de faire passer une fonction $f(t)$ dépendant de n inconnues (par exemple un polynôme de degré $n-1$) par m points ($m > n$). Il n'existe donc sans doute pas de $x \in \mathbb{R}^n$ qui annule les quantités

$$\sum_{k=1}^{n} x_k f_k(t_i) - b_i, \quad 1 \leq i \leq m\,.$$

Cette expression s'écrit bien $Ax - b$ où

$$a_{ij} = f_j(t_i), \quad 1 \leq i \leq m, \quad 1 \leq j \leq n,$$

et l'on a bien à minimiser

$$E(x) = \sum_{i=1}^{m} \left[\sum_{k=1}^{n} x_k f_k(t_i) - b_i \right]^2 .$$

Cette fois-ci le minimum ne sera pas nul.

III.1.3 Les équations normales

Exercices :
Exercice III.3

Documents :
Document III.2

Nous ferons plus tard une résolution purement algébrique du problème des moindres carrés. Cependant nous allons en donner une approche analytique qui permet de voir ce problème sous un angle différent. Tout d'abord, explicitons la fonction à minimiser. On a

$$\begin{aligned} E(x) &= \|Ax-b\|_2^2 = (Ax-b)^T(Ax-b) = x^TA^TAx - b^TAx - x^TA^Tb + \|b\|_2^2 \\ &= x^TA^TAx - 2(A^Tb)^Tx + \|b\|_2^2. \end{aligned}$$

Le problème de moindres carrés peut donc se reformuler en

$$\min_{x\in\mathbb{R}^n} J(x) = x^TGx - 2h^Tx,$$

où $G = A^TA$ est symétrique et $h = A^Tb$ est un vecteur donné.

Définition III.1.2. *On appelle* fonction quadratique *une fonction* $J : \mathbb{R}^n \to \mathbb{R}$ *, de la forme*

$$J(x) = x^TGx - 2h^Tx,$$

où G *est une matrice* $n \times n$ *symétrique et* h *est un vecteur donné de* $\mathbb{R}^n$.

Rappelons que si $J : \mathbb{R} \to \mathbb{R}$, continûment dérivable, admet un minimum $\widehat{x} \in \mathbb{R}$, alors $J'(\widehat{x}) = 0$. De même soit $J : \mathbb{R}^n \to \mathbb{R}$ continûment dérivable, alors

$$J(\widehat{x}) \le J(x)\ \forall x \in \mathbb{R}^n \ \Rightarrow\ \nabla J(\widehat{x}) = 0,\ \nabla \text{ opérateur gradient .}$$

Sommaire
Concepts

Exemples
Exercices
Documents

Ici le calcul du gradient de J donne (voir le document référencé)

$$\nabla J(x) = 2(Gx - h)$$

et la solution $\widehat{x}$ du problème de moindre carrés vérifie donc nécessairement
$G\widehat{x} = h$, soit

$$A^T A\widehat{x} = A^T b.$$

Ces relations sont appelées *équations normales* du problème, ce sont des conditions nécessaires pour que $\widehat{x}$ soit minimum, on montre de plus que ces conditions sont suffisantes d'où le théorème.

Théorème III.1.1. *Soit $A \in M_{mn}(\mathbb{R})$ et $b \in \mathbb{R}^m$ donnés, si l'on suppose que la matrice A est de rang n, le problème de moindres carrés $\min\limits_{x\in\mathbb{R}^n} \|Ax - b\|_2^2$, admet une solution unique $\widehat{x}$ donnée par*

$$A^T A\widehat{x} = A^T b. \tag{III.1.3}$$

Démonstration

Sommaire
Concepts

Exemples
Exercices
Documents

III.2 Approche algébrique du problème de moindres carrés

III.2.1 Idée intuitive de l'approche algébrique 11
III.2.2 Espaces orthogonaux - Rappels 12
III.2.3 Problèmes de projection 14
III.2.4 Utilisation de l'orthogonalisation de Schmidt pour résoudre les équations normales 16

III.2.1 Idée intuitive de l'approche algébrique

Soit $\mathrm{Im}\,(A) = \{y \in \mathbb{R}^m \mid \exists x \in \mathbb{R}^n,\ y = Ax\}$. Alors le problème de moindres carrés

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2\,,$$

signifie que l'on cherche dans l'image de A l'élément le plus "proche" de b. Il se formule donc comme un problème de projection orthogonale de b sur le sous-espace vectoriel $\mathrm{Im}\,(A)$ (voir la figure III.2.2). Si on appelle $\widehat{x}$ la solution de ce problème, on s'attend donc à ce que le *résidu* $r = b - A\widehat{x}$ soit orthogonal à $\mathrm{Im}\,(A)$ (ceci sera évidemment revu dans la suite de ce cours).

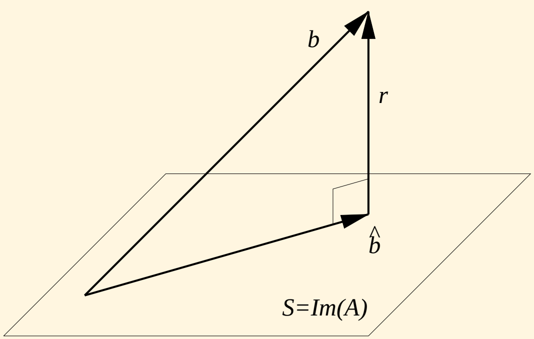

FIG. III.2.2: projection de b sur $\mathrm{Im}\,(A)$

Quel est le lien avec les équations normales ? Pour le retrouver nous allons rappeler quelques résultats sur les espaces euclidiens.

III.2.2 Espaces orthogonaux - Rappels

Exercices :
Exercice III.4
Exercice III.5

Documents :
Document III.3

Définition III.2.1. *On rappelle la définition du produit scalaire usuel sur* $\mathbb{R}^n$ *:*

$$x, y \in \mathbb{R}^n,\ \langle x, y\rangle = x^T y = y^T x.$$

Cette définition généralise le produit scalaire usuel que vous connaissez dans $\mathbb{R}^2$ ou $\mathbb{R}^3$.

Définition III.2.2. *Soit S un sous-espace vectoriel de $\mathbb{R}^n$, on appelle* orthogonal *de S le sous espace vectoriel de $\mathbb{R}^n$ noté $S^\perp$ défini par :*

$$x \in S^\perp \Leftrightarrow \forall y \in S, \langle x, y\rangle = 0.$$

Proposition III.2.3. *Soit S un sous-espace vectoriel de $\mathbb{R}^n$, $S^\perp$ l'orthogonal de S alors*

$$S \cap S^\perp = \{0\},\ \ \mathbb{R}^n = S \oplus S^\perp.$$

Démonstration -

$$x \in S \cap S^\perp \Rightarrow \langle x, x\rangle = 0 \Rightarrow \sum_{i=1}^{n} x_i^2 = 0 \Rightarrow x = 0.$$

Pour montrer la somme directe, on utilise l'orthogonalisation de Schmidt (voir le document référencé) pour montrer que tout sous-espace vectoriel S admet une base orthogonale $\mathcal{B}$. On

Sommaire
Concepts

Exemples
Exercices
Documents

rappelle qu'une famille libre, par exemple $\mathcal{B}$, de E peut être complétée par une famille $\mathcal{C}$ telle que $\mathcal{B} \cup \mathcal{C}$ soit une base de E (théorème de la base incomplète). On continue alors le processus d'orthogonalisation de Schmidt sur $\mathcal{C}$ et on obtient ainsi une base orthogonale de la forme $\mathcal{B} \cup \mathcal{B}'$. Alors il est clair, par construction, que $\mathcal{B}'$ engendre un sous espace vectoriel qui est $S^\perp$.

Corollaire III.2.1. *Soit $y \in \mathbb{R}^n$ alors il existe un unique $\widehat{y} \in S$ tel que*

$$y - \widehat{y} \in S^\perp,$$

le vecteur $\widehat{y}$ étant appelé projection orthogonale *de y sur S.*

Démonstration - D'après le théorème précédent, tout vecteur y s'écrit de manière unique sous la forme : $y = \widehat{y} + z$, où $\widehat{y} \in S$ et $z \in S^\perp$.

Proposition III.2.4. *Soit $A \in \mathcal{M}_{mn}$*

$$(\mathbf{Im}\, A)^\perp = \mathbf{Ker}\,(A^T).$$

Démonstration -

$$\begin{array}{rcl} x \in \mathbf{Im}\,(A)^\perp & \Leftrightarrow & x^T y = 0,\ \forall y \in \mathbf{Im}\,(A), \\ & \Leftrightarrow & x^T(Az) = 0,\ \forall z \in \mathbb{R}^n, \\ & \Leftrightarrow & (A^T x)^T z = 0, \forall z \in \mathbb{R}^n, \\ & \Leftrightarrow & A^T x = 0. \end{array}$$

Pour la démonstration du dernier point voir l'exercice III.4

Espaces orthogonaux - Rappels

Sommaire
Concepts

Exemples
Exercices
Documents

III.2.3 Problèmes de projection

Proposition III.2.5. *Etant donnés S un sous-espace vectoriel de E et $y \in E$, le problème*

$$\min_{z \in S} \|z-y\|_2$$

admet pour solution unique $\widehat{y} \in S$ projection orthogonale de y sur S.

Démonstration - Soit z un élément quelconque de S, alors

$$\begin{aligned} \|z-y\|_2^2 &= \|z-\widehat{y}-(y-\widehat{y})\|_2^2, \\ &= \|z-\widehat{y}\|_2^2 + \|y-\widehat{y}\|_2^2 - 2, \langle z-\widehat{y}, y-\widehat{y}\rangle , \\ &= \|z-\widehat{y}\|_2^2 + \|y-\widehat{y}\|_2^2, \end{aligned}$$

puisque $(z-\widehat{y}) \in S$ et $(y-\widehat{y}) \in S^\perp$. Alors

$$\|z-y\|_2^2 > \|\widehat{y}-y\|_2^2 \; \forall z \neq \widehat{y}$$

ce qui montre que $\min_{z\in S} \|z-y\|_2 = \|\widehat{y}-y\|_2$.

Corollaire III.2.2. *Soit $A \in M_{mn}(\mathbb{R})$ et $b \in \mathbb{R}^m$ donnés, le problème :*
trouver $\widehat{x} \in \mathbb{R}^n$ tel que

$$\|A\widehat{x}-b\|_2 \leq \|Ax-b\|_2 \quad \forall x \in \mathbb{R}^n,$$

est équivalent à : trouver $\widehat{x} \in \mathbb{R}^n$ tel que :

$$A^T A\widehat{x} = A^T b.$$

Démonstration -
On notera $\widehat{b}$ la projection orthogonale de b sur Im A, on a donc par définition

$$\widehat{b} \in \text{Im } A, \quad b-\widehat{b} \in (\text{Im } A)^\perp.$$

Sommaire
Concepts

Exemples
Exercices
Documents

Problèmes de projection

– On utilise la proposition précédente avec $E = \mathbb{R}^m, S = \text{Im}\,(A), y = b$.

$$\|A\widehat{x} - b\|_2 \leq \|Ax - b\|_2 \;\; \forall x \in \mathbb{R}^n \Leftrightarrow \|A\widehat{x} - b\|_2 \leq \|z - b\|_2 \;\; \forall z \in \text{Im}\ A \Leftrightarrow A\widehat{x} = \widehat{b}.$$

– Montrons que

$$A\widehat{x} = \widehat{b} \Leftrightarrow A^T A\widehat{x} = A^T\widehat{b}.$$

l'implication de gauche à droite est évidente. montrons la réciproque :

$$A^T A\widehat{x} = A^T\widehat{b} \Leftrightarrow A^T(A\widehat{x} - \widehat{b}) = 0 \Leftrightarrow A\widehat{x} - \widehat{b} \in \text{Ker}\ A^T = (\text{Im}\ A)^\perp$$

or $\widehat{b} \in \text{Im}\ A$, donc $A\widehat{x} - \widehat{b} \in \text{Im}\ A$, donc

$$A\widehat{x} - \widehat{b} \in \text{Im}\ A \cap (\text{Im}\ A)^\perp \Rightarrow A\widehat{x} - \widehat{b} = 0 \Rightarrow A\widehat{x} = \widehat{b}.$$

Ce qui termine de montrer l'équivalence.

– Montrons maintenant que

$$A^T A\widehat{x} = A^T\widehat{b} \Leftrightarrow A^T A\widehat{x} = A^T b.$$

Par définition de $\widehat{b}$, on a $b - \widehat{b} \in (\text{Im}\ A)^\perp = \text{Ker}\ A^T$, donc $A^T(b - \widehat{b}) = 0$ donc $A^T b = A^T\widehat{b}$.
Ce qui termine de démontrer cette dernière équivalence.

On retrouve donc par un raisonnement faisant intervenir les projections que les équations normales sont des conditions nécessaires et suffisantes d'optimalité pour le problème de minimisation.

Sommaire
Concepts

Exemples
Exercices
Documents

III.2.4 Utilisation de l'orthogonalisation de Schmidt pour résoudre les équations normales

Exercices :
Exercice III.7
Exercice III.6

A l'aide de l'orthogonalisation de Schmidt, on calcule une base orthogonale $\{E_1, E_2, \ldots, E_n\}$ de Im (A), on obtient (démontré en exercice)

$$A = ET, \text{ où } E = [E_1 E_2 \ldots E_n],$$

$E \in \mathcal{M}_{mn}$ est une matrice rectangulaire et $T \in \mathcal{M}_{nn}$ est une matrice triangulaire supérieure inversible. Puisque les colonnes de E sont des vecteurs orthonormés, on a $E^T E = I$, montrer ce résultat en exercice.

On obtient $A^T A = T^T E^T E T = T^T T$, $A^T b = T^T E^T b$.

En reprenant les équations normales, on a donc :

$$A^T A\hat{x} = A^T b \Leftrightarrow T^T T\hat{x} = T^T E^T b \Leftrightarrow T\hat{x} = E^T b.$$

En effet la matrice T donc la matrice T^T est inversible, on peut donc simplifier la deuxième équation, il n'était évidemment pas possible de simplifier la première équation par A^T puisque la matrice A^T n'est pas carrée donc pas inversible.

Il s'avère que le système $T\hat{x} = E^T b$ est mieux conditionné que celui donné par les équations normales. Cependant l'orthogonalisation de Schmidt est très sujette à l'accumulation des erreurs d'arrondi. C'est pourquoi on lui préfère une factorisation de A du même type, basée cette fois sur la transformation de *Householder*, qui fait l'objet du paragraphe suivant.

III.3 Résolution des problèmes de moindres carrés par "QR".

III.3.1 La transformation de *Householder* 18
III.3.2 La factorisation "QR" 20
III.3.3 Application à la résolution du problème de moindres carrés 22

Sommaire
Concepts

Exemples
Exercices
Documents

III.3.1 La transformation de *Householder*

Dans l'espace euclidien $\mathbb{R}^n$, les transformations orthogonales (c'est à dire les applications linéaires représentées par des matrices orthogonales) conservent la norme euclidienne $\|.\|_2$. En effet si H est orthogonale on a $H^T = H^{-1}$ et donc

$$\|Hx\|_2^2 = (Hx)^T Hx = x^T H^T Hx = x^T x = \|x\|_2^2.$$

Le but de ce paragraphe est d'introduire une transformation orthogonale particulière : la transformation de *Householder*, qui est une symétrie plane.

Définition III.3.1. *On appelle* transformation de Householder, *une transformation dont la matrice est de la forme*

$$H = I - 2yy^T,$$

où $y \in \mathbb{R}^n$ *et* $\|y\|_2 = 1$.

Il est clair que H est symétrique et on vérifie sans difficulté que H est orthogonale :

$$HH^T = HH = (I - 2yy^T)(I - 2yy^T) = I - 2yy^T - 2yy^T + 4yy^T = I.$$

Remarque III.3.2. *La transformation de Householder conserve donc la norme. Par d'ailleurs on a :*

$$Hy = y - 2yy^T y = -y$$

et si z *est orthogonal à* y *alors*

$$Hz = z - 2yy^T z = z$$

ce qui montre que H *correspond à une symétrie par rapport au "plan" perpendiculaire à* y.

Théorème III.3.3. *Soit* $x \in \mathbb{R}^n$, *avec* $\|x\|_2 = 1$ *et* $x \neq e = \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix}^T$. *Alors il existe une transformation de Householder* H *telle que* $Hx = e$.

Sommaire
Concepts

Exemples
Exercices
Documents

Démonstration - Posons $y = \alpha(x-e)$ avec $\alpha = (\|x-e\|_2)^{-1}$, alors la matrice $H = I - 2yy^T$ répond à la question. En effet

$$Hx = [I - 2\alpha^2(x-e)(x-e)^T]x = x - 2\alpha^2\left((x-e)^T x\right)(x-e).$$

En effet $(x-e)^T x$ est un scalaire, on peut donc commuter.

On a de plus $\alpha^{-2} = \|x-e\|_2^2 = (x-e)^T(x-e) = \|x\|_2^2 - 2e^T x + 1 = 2(1-e^T x)$ et $(x-e)^T x = 1 - e^T x$, d'où $2\alpha^2\left((x-e)^T x\right) = 1$.

Ce qui donne : $Hx = x - (x-e) = e$.

Sommaire
Concepts

Exemples
Exercices
Documents

III.3.2 La factorisation "QR"

Exercices :
Exercice III.8

Définition III.3.1. *Soit Q une matrice carrée, on dit que Q est orthogonale si :*

$$Q^T = Q^{-1} \Leftrightarrow Q^T Q = I.$$

Théorème III.3.4. *Soit $A \in \mathcal{M}_{mn}(\mathbb{R})$ avec $m \geq n$, alors il existe une matrice orthogonale $Q \in \mathcal{M}_{mm}(\mathbb{R})$ et une matrice triangulaire supérieure $\widetilde{R} \in \mathcal{M}_{nn}(\mathbb{R})$ telles que*

$$A = Q \begin{pmatrix} \widetilde{R} \\ 0 \end{pmatrix}. \tag{III.3.1}$$

Démonstration

On note $R = \begin{pmatrix} \widetilde{R} \\ 0 \end{pmatrix}$.

On peut remarquer que la formule (III.3.1) correspond à une orthogonalisation des colonnes de A. En effet, la $j^{\text{ème}}$ colonne de A, soit A_j, est donnée par

$$A_j = QR_j = \sum_{i=1}^{j} r_{ij} Q_i,$$

Sommaire
Concepts

Exemples
Exercices
Documents

ceci compte tenu de la structure de R. Ceci signifie que $\forall\ k = 1...n$, les colonnes $[Q_j]_{j=1...k}$ (qui constituent une famille orthonormée) et $[A_j]_{j=1...k}$ engendrent le même sous-espace. La factorisation QR est donc une façon d'orthogonaliser une famille de vecteurs, au même titre que *l'orthogonalisation de Schmidt*. En fait on utilise toujours la factorisation QR car elle est moins sujette aux erreurs d'arrondi que l'orthogonalisation de Schmidt.

Sommaire
Concepts

Exemples
Exercices
Documents

III.3.3 Application à la résolution du problème de moindres carrés

Exercices :
Exercice III.9
Exercice III.10

On part donc de la factorisation QR précédente, ce qui permet d'écrire A sous la forme

$$A = Q \begin{pmatrix} \widetilde{R} \\ 0 \end{pmatrix},$$

où $\widetilde{R} \in \mathcal{M}_{nn}(\mathbb{R})$ est triangulaire supérieure. Notons que pour tout vecteur y de $\mathbb{R}^m$ on a

$$\|Q^T y\|_2^2 = \|y\|_2^2,$$

puisque Q^T, comme Q, est orthogonale. On a donc, en particulier :

$$\|Ax - b\|_2^2 = \|Q^T(Ax - b)\|_2^2 = \|Q^T Ax - Q^T b\|_2^2.$$

Définissons $c \in \mathbb{R}^n$ et $d \in \mathbb{R}^{m-n}$ par

$$Q^T b = \begin{pmatrix} c \\ d \end{pmatrix}.$$

Alors

$$Q^T Ax - Q^T b = \begin{pmatrix} \widetilde{R} \\ 0 \end{pmatrix} x - Q^T b = \begin{pmatrix} \widetilde{R}x - c \\ d \end{pmatrix}$$

et donc

$$\|Ax - b\|_2^2 = \|\widetilde{R}x - c\|_2^2 + \|d\|_2^2.$$

Sommaire
Concepts

Exemples
Exercices
Documents

Le vecteur $\widehat{x}$ minimisant la norme de $\|Ax - b\|_2$ est donné par

$$\widetilde{R}\widehat{x} = c, \qquad \text{(III.3.2)}$$

puisque $\|d\|_2^2$ est une constante qui ne joue aucun rôle dans la minimisation. Le vecteur $\widehat{x}$ est unique si $\widetilde{R}$ est inversible, ce qui est le cas si rang $(A) = n$.

Remarquons que la factorisation QR est présente dans tous les logiciels d'analyse numérique, comme par exemple MATLAB ou SCILAB.

Documents du chapitre III

III.1 Démonstration du théorème III.1.1 25
III.2 Le gradient d'une forme quadratique 26
III.3 L'orthogonalisation de Schmidt 27
III.4 Démonstration du théorème III.3.4 29

Document III.1 Démonstration du théorème III.1.1

On remarque tout d'abord que si A est de rang n, alors $G = A^T A$ est définie positive en effet :

$$x^T G x = \|Ax\|_2^2 \geq 0 \; \forall x \in \mathbb{R}^n,$$

d'autre part puisque le rang de A vaut n, alors la dimension de Ker A est nulle donc

$$x^T G x = 0 \Leftrightarrow \|Ax\|_2^2 = 0 \Leftrightarrow Ax = 0 \Leftrightarrow x = 0.$$

La matrice G étant définie positive, elle est inversible, donc il existe une unique solution $\widehat{x}$ vérifiant $A^T A\widehat{x} = A^T b \Leftrightarrow G\widehat{x} = h$.

Montrons maintenant que

$$\forall x \in \mathbb{R}^n, x \neq \widehat{x} \Rightarrow J(x) > J(\widehat{x}).$$

Posons $y = x - \widehat{x}$, on a donc $y \neq 0$.

$$\begin{aligned} J(\widehat{x}+y) &= (\widehat{x}+y)^T G(\widehat{x}+y) - 2h^T(\widehat{x}+y), \\ &= \widehat{x}^T G\widehat{x} + y^T G y + 2\widehat{x}^T G y - 2h^T y - 2h^T\widehat{x}, \\ &= \widehat{x}^T G\widehat{x} + y^T G y + 2(G\widehat{x} - h)^T y - 2h^T\widehat{x}, \\ &= y^T G y + \widehat{x}^T G\widehat{x} - 2h^T\widehat{x}, \end{aligned}$$

puisque $G\widehat{x} = h$. d'où

$$J(\widehat{x}+y) = J(\widehat{x}) + y^T G y.$$

Puisque G est une matrice définie positive, et que $y \neq 0$, on a $y^T G y > 0$, et donc

$$J(\widehat{x}+y) > J(\widehat{x})\,,\ \forall y \in \mathbb{R}^n,\ y \neq 0\,,$$

ce qui montre que $\widehat{x}$ réalise le minimum de J.

Retour au théorème III.1.1 ▲

Document III.2 Le gradient d'une forme quadratique

Soit la fonction quadratique

$$J(x) = x^T G x - 2h^T x,$$

où G est une matrice symétrique. On veut calculer son gradient, soit

$$\nabla J(x) = \left(\frac{\partial J}{\partial x_1}, \frac{\partial J}{\partial x_2}, \ldots, \frac{\partial J}{\partial x_n} \right)$$

Développons la fonction J

$$J(x) = \sum_{i=1}^{n} x_i (Gx)_i - 2 \sum_{i=1}^{n} h_i x_i.$$

Alors

$$\frac{\partial J}{\partial x_k} = (Gx)_k + \sum_{i=1}^{n} x_i \frac{\partial}{\partial x_k} (Gx)_i - 2h_k,$$

et

$$\frac{\partial}{\partial x_k} (Gx)_i = \frac{\partial}{\partial x_k} \left(\sum_{j=1}^{n} g_{ij} x_j \right) = g_{ik} = g_{ki}.$$

Donc

$$\frac{\partial J}{\partial x_k} = (Gx)_k + \sum_{i=1}^{n} x_i g_{ki} - 2h_k = (Gx)_k + (Gx)_k - 2h_k = 2(Gx)_k - 2h_k,$$

d'où le résultat

$$\nabla J(x) = 2(Gx - h).$$

retour au cours

Sommaire
Concepts

Exemples
Exercices
Documents

Document III.3 L'orthogonalisation de Schmidt

Théorème III.3.5 (Orthogonalisation de *Schmidt*). *Dans tout espace euclidien de dimension finie il existe des bases orthonormées.*

Démonstration - Elle est constructive : on part d'une base quelconque de E soit $\{B_1, B_2, \ldots, B_n\}$ et on construit par récurrence une base orhonormée $\{E_1, E_2, \ldots, E_n\}$. Première étape, on pose

$$E_1 = \frac{B_1}{\|B_1\|_2},$$

en effet $\|B_1\|_2 \neq 0$ car la famille $\{B_1, B_2, ..., B_n\}$ est libre.

Deuxième étape, on pose

$$\widetilde{E}_2 = B_2 - \langle B_2, E_1 \rangle E_1,$$

où on a noté $\langle , \rangle$ le produit scalaire dans E.

On a bien

$$\left\langle \widetilde{E}_2, E_1 \right\rangle = \langle B_2 - \langle B_2, E_1 \rangle E_1, E_1 \rangle = \langle B_2, E_1 \rangle - \langle B_2, E_1 \rangle \langle E_1, E_1 \rangle = 0.$$

$\|\widetilde{E}_2\|_2 \neq 0$ car sinon B_2 serait proportionnel à E_1 donc à B_1 ce qui est impossible car la famille $\{B_1, B_2, ..., B_n\}$ est libre.

On peut donc définir

$$E_2 = \frac{\widetilde{E}_2}{\|\widetilde{E}_2\|_2}.$$

Etape k : supposons que les vecteurs $\{E_1, E_2, \ldots, E_{k-1}\}$ sont construits et orthogonaux deux à deux. On définit alors $\widetilde{E}_k$ par

$$\widetilde{E}_k = B_k - \sum_{j=1}^{k-1} \langle B_k, E_j \rangle E_j,$$

Sommaire
Concepts

Exemples
Exercices
Documents

Vérifions que le vecteur $\widetilde{E}_k$ est orthogonal aux vecteurs $E_1, E_2, ..., E_{k-1}$

$$\begin{aligned}
\left\langle \widetilde{E}_k, E_i \right\rangle &= \left\langle B_k - \sum_{j=1}^{k-1} \langle B_k, E_j \rangle E_j, E_i \right\rangle, \\
&= \langle B_k, E_i \rangle - \sum_{j=1}^{k-1} \langle B_k, E_j \rangle \langle E_j, E_i \rangle, \\
&= \langle B_k, E_i \rangle - \langle B_k, E_i \rangle \langle E_i, E_i \rangle, \\
&= \langle B_k, E_i \rangle - \langle B_k, E_i \rangle = 0.
\end{aligned}$$

$\|\widetilde{E}_k\|_2 \neq 0$ car sinon B_k serait une combinaison linéaire de $E_1, E_2, ..., E_{k-1}$, donc une combinaison linéaire de $B_1, B_2, ..., B_{k-1}$ ce qui est impossible car la famille $\{B_1, B_2, ..., B_n\}$ est libre.

On peut donc définir

$$E_k = \frac{\widetilde{E}_k}{\|\widetilde{E}_k\|_2}.$$

Il faut remarquer que, à chaque étape k de la méthode, l'espace engendré par $\{E_1, E_2, \ldots, E_k\}$ est le même que celui engendré par $\{B_1, B_2, \ldots, B_k\}$.

retour au cours

Sommaire
Concepts

Exemples
Exercices
Documents

Document III.4 Démonstration du théorème III.3.4

Il est important de donner ici la démonstration complète de ce théorème, car elle est *constructive*. C'est à dire qu'elle donne l'algorithme pour obtenir la factorisation. On va chercher à obtenir la matrice R comme le résultat de n transformations orthogonales successives $U^{(k)}$, soit

$$\begin{pmatrix} \widetilde{R} \\ 0 \end{pmatrix} = A^{(n+1)} = U^{(n)}U^{(n-1)}\ldots U^{(1)}A$$

les matrices $U^{(k)}$ étant construites à l'aide de transformations de Householder.

Si la première colonne de A s'écrit $(\alpha_1 0 ... 0)^T$, il n'y a rien à faire et on pose donc $U^{(1)} = I$. Sinon on sait qu'il existe une transformation de Householder $H^{(1)}$ qui transforme A_1 en $(\alpha_1 0...0)^T$, avec $\alpha_1 = \|A_1\|_2$. En posant $U^{(1)} = H^{(1)}$ on a donc :

$$A^{(2)} = U^{(1)}A = \begin{pmatrix} \alpha_1 & \times & \ldots & \times \\ 0 & \times & \ldots & \times \\ \vdots & \vdots & & \vdots \\ 0 & \times & \ldots & \times \end{pmatrix}.$$

Soit $v^{(2)} \in \mathbb{R}^{m-1}$ le vecteur dont les éléments sont $[a^{(2)}_{i2}]_{i=2\ldots m}$; il s'agit de la partie de la deuxième colonne de $A^{(2)}$ qui commence à l'élément diagonal. On sait qu'il existe une transformation de Householder $H^{(2)}$ qui transforme $v^{(2)}$ en $(\alpha_2\ 0\ \ldots 0)^T \in \mathbb{R}^{m-1}$, avec $\alpha_2 = \|v^{(2)}\|_2$. On définit alors $U^{(2)}$ comme

$$U^{(2)} = \begin{pmatrix} 1 & 0 \\ 0 & H^{(2)} \end{pmatrix}$$

Sommaire
Concepts

Exemples
Exercices
Documents

et on obtient

$$A^{(3)} = U^{(2)}A^{(2)} = \begin{pmatrix} \alpha_1 & \times & \times & \dots & \times \\ 0 & \alpha_2 & \times & \dots & \times \\ 0 & 0 & \times & \dots & \times \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \times & \dots & \times \end{pmatrix}.$$

On peut remarquer que, par définition de $U^{(2)}$, la première colonne de $A^{(2)}$ n'a pas été modifiée.

On peut aisément généraliser ce procédé : supposons que l'on a obtenu $A^{(k)}$ dont les $k-1$ premières colonnes forment une matrice trapézoïdale supérieure (les éléments en dessous de la diagonale sont nuls). Si on note $v^{(k)} \in \mathbb{R}^{m-k+1}$ le vecteur dont les éléments sont $[a_{ik}^{(k)}]_{i=k\dots m}$, alors il existe aussi une transformation de Householder $H^{(k)}$ qui transforme $v^{(k)}$ en $[\alpha_k \ 0 \ \dots 0]^T \in \mathbb{R}^{m-k+1}$, avec $\alpha_k = \|v^{(k)}\|_2$. On définit alors $U^{(k)}$ comme

$$U^{(k)} = \begin{pmatrix} I_{k-1} & 0 \\ 0 & H^{(k)} \end{pmatrix},$$

et on obtient $A^{(k+1)} = U^{(k)}A^{(k)}$. On continue ce procédé jusqu'à obtenir une matrice $A^{(n+1)}$

$$A^{(n+1)} = U^{(n)}A^{(n)} = U^{(n)}U^{(n-1)}\dots U^{(1)}A,$$

qui, par construction, a la structure désirée. On a donc bien

$$\begin{pmatrix} \widetilde{R} \\ 0 \end{pmatrix} = UA,$$

avec $U = U^{(n)}U^{(n-1)}\dots U^{(1)}$. On obtient la factorisation QR en remarquant que le produit de matrices orthogonales reste une matrice orthogonale et en posant $Q = U^T$.

Retour au théorème III.3.4 ▲

Sommaire
Concepts

Exemples
Exercices
Documents

Exercices du chapitre III

III.1 32
III.2 33
III.3 34
III.4 35
III.5 36
III.6 37
III.7 38
III.8 39
III.9 40
III.10 41

Exercice III.1

1. Ecrire le problème de la régression linéaire comme un problème de moindres carrés : plus précisément on se donne une famille de points $(t_i, b_i)_{1\leq i\leq m}$ (les t_i étant distincts) et on cherche à faire passer une droite le plus près possible de ces points.
2. La question précédente conduit à une fonction de deux variables à minimiser. On admet que ce minimum est donné en annulant les deux dérivées partielles. Donner le système linéaire de deux équations à deux inconnues ainsi obtenu.

retour au cours

Solution

Sommaire
Concepts

Exemples
Exercices
Documents

Exercice III.2

On cherche à approcher les données $(t_i, b_i)_{1\leq i\leq m}$ (les t_i étant distincts) à l'aide d'un polynôme de degré inférieur ou égal à $n-1$, on suppose $m \geq n$, on note

$$p(t) = \alpha_1 + \alpha_2 t + \ldots + \alpha_n t^{n-1},$$

et on cherche les coefficients $\alpha_1, \alpha_2, ..., \alpha_n$ qui minimisent

$$E(\alpha_1, \alpha_2, ..., \alpha_n) = \sum_{i=1}^{m} (p(t_i) - b_i)^2.$$

Ecrire le problème de moindres carrés sous forme matricielle. Montrer alors que la matrice A est bien de rang n. (On rappelle que la matrice carrée de Van der Monde V, $v_{ij} = t_i^{j-1}$, est inversible si tous les t_i sont distincts).

retour au cours

Solution

Sommaire
Concepts

Exemples
Exercices
Documents

Exercice III.3

Donner les équations normales du problème de moindres carrés associé à la régression linéaire. Montrer que l'on retrouve les équations de l'exercice III.1 .

retour au cours

Solution

Sommaire
Concepts

Exemples
Exercices
Documents

Exercice III.4

Soit $b \in \mathbb{R}^n$, montrer que

$$b^T z = 0, \forall z \in \mathbb{R}^n \;\Leftrightarrow\; b = 0.$$

retour au cours

Solution

Sommaire
Concepts

Exemples
Exercices
Documents

Exercice III.5

Ecrire l'algorithme de l'orthogononalisation de Schmidt, donnée dans le document III.3 .

retour au cours

Solution

Sommaire
Concepts

Exemples
Exercices
Documents

Exercice III.6

Soit $E \in \mathcal{M}_{mn}$, une matrice dont les colonnes sont des vecteurs orthonormés de $\mathbb{R}^n$, montrer que $E^T E = I$

retour au cours

Solution

Sommaire
Concepts

Exemples
Exercices
Documents

Exercice III.7

On applique l'orthogonalisation de Schmidt (voir le document III.3), sur les n colonnes $A_1, A_2, ..., A_n$ d'une matrice $A \in \mathcal{M}_{m,n}$ de rang n ($m \geq n$), on obtient les vecteurs $E_1, E_2, ..., E_n$ qui seront les colonnes d'une matrice E.

1. Montrer que les colonnes A_k de A peuvent s'écrire

$$A_k = \sum_{j=1}^{k} \alpha_{jk} E_j$$

   sans expliciter les scalaires α_{jk}.

2. En déduire que $A = ET$, où T est une matrice triangulaire supérieure inversible.

retour au cours

Solution

Sommaire
Concepts

Exemples
Exercices
Documents

Exercice III.8

1. Montrer que si Q est une matrice orthogonale, alors Q^T est orthogonale.
2. Montrer que si Q est orthogonale, alors $\|Qy\|_2 = \|y\|_2$ pour tout vecteur y de $\mathbb{R}^n$.

retour au cours

Solution

Sommaire
Concepts

Exemples
Exercices
Documents

Exercice III.9

On veut effectuer une régression linéaire sur les points suivants : $(-1, 0.5)$, $(0.5, 1)$, $(2, 2.5)$. Appliquer la méthode "QR" pour résoudre ce problème et utiliser SCILAB, en particulier la procédure "qr", pour faire les calculs.

retour au cours

Solution

Sommaire
Concepts

Exemples
Exercices
Documents

Exercice III.10

Soit A une matrice $m \times n$ de rang $n \leq m$. Soient Q une matrice orthogonale et $\widetilde{R}$ une matrice carrée triangulaire supérieure telles que $A = Q \begin{pmatrix} \widetilde{R} \\ 0 \end{pmatrix}$.

Montrer que, si χ_2 désigne le conditionnement calculé à partir de la norme matricielle subordonnée à la norme 2,

$$\chi_2(\widetilde{R}) = \sqrt{\chi_2(A^T A)}.$$

retour au cours

Solution

Sommaire
Concepts

Exemples
Exercices
Documents

Annexe A
Exercices

A.1 Exercices de TD du chapitre 3 43

A.1 Exercices de TD du chapitre 3

A.1.1 44
A.1.2 46
A.1.3 Factorisation $A = QR$ par la méthode de Householder 47

Exercice A.1.1

1. On cherche à approcher des données expérimentales (t_i, y_i), $i = 1 \ldots m$, par une fonction $f_{a,b,c,d}$ définie par morceaux de la façon suivante :

$$f_{a,b,c,d}(t) = \begin{cases} a + bt + ct^2 & \text{si} \quad t \leq 0, \\ a + dt & \text{si} \quad t > 0. \end{cases}$$

   On suppose qu'il existe p, $3 \leq p < m$, tel que

$$t_1 < t_2 < \cdots < t_p \leq 0 < t_{p+1} < \cdots < t_m.$$

   Pour calculer les quatre paramètres a, b, c et d on cherche à minimiser par rapport à ces quatre paramètres la fonction erreur suivante :

$$E(a,b,c,d) = \sum_{i=1}^{m} [f_{a,b,c,d}(t_i) - y_i]^2.$$

   (a) Montrer que minimiser cette fonction erreur revient à résoudre un problème de moindres carrés

$$\min_x \|Ax - y\|_2^2,$$

   dont on précisera l'inconnue x, la matrice A et le vecteur y.

   (b) Quel est le rang de A ? Ecrire les équations normales, montrer que ces équations ont une solution unique.

   (c) Que se passe-t-il si $p = 3, m = 4$?

2. On cherche à approcher les données par une fonction $f_{a,b,c}$ définie par :

Sommaire
Concepts

Exemples
Exercices
Documents

Exercice A.1.1

$$f_{a,b,c}(t) = \begin{cases} a\ln(t+1) + ct & \text{si} \quad t > 0, \\ b\,(e^t - 1) + ct & \text{si} \quad t \leq 0. \end{cases}$$

On suppose qu'il existe p, $2 \leq p < m$, tel que

$$t_1 < t_2 < \cdots < t_p < 0 < t_{p+1} < \cdots < t_m.$$

Pour calculer les trois paramètres a, b et c on cherche à minimiser par rapport à ces trois paramètres la fonction erreur suivante :

$$E(a,b,c) = \sum_{i=1}^{m} [f_{a,b,c}(t_i) - y_i]^2.$$

(a) Montrer que minimiser cette fonction erreur revient à résoudre un problème de moindres carrés

$$\min_x \|Ax - y\|_2^2,$$

dont on précisera l'inconnue x, la matrice A et le vecteur y.

(b) Quel est le rang de A ?

Question 2a Aide 1 Aide 2
Question 2b Aide 1 Aide 2 Aide 3

Sommaire
Concepts

Exemples
Exercices
Documents

Exercice A.1.2

1. On définit $\tau_i = i, 0 \leq i \leq 5$, on définit g la fonction telle que la courbe d'équation $y = g(t)$ soit une ligne brisée qui joint les points de coordonnés $(\tau_i, z_i), 0 \leq i \leq 5$. Donner l'expression de $g(t)$ pour $t \in [0, 5]$ à l'aide des z_i.
2. On définit $t_k = 0.5k, 1 \leq k \leq 10$, on cherche à approcher le nuage de points $(t_k, y_k), 1 \leq k \leq 10$ par $g(t)$ au sens des moindres carrés. Mettre ce problème sous la forme : $\min\limits_{z \in \mathbf{R}^n} \|Az - y\|_2^2$, que vaut n ? Quelle est la taille de la matrice A ? Expliciter ses termes. Quel est le rang de A ?

Question 1 Aide 1 Aide 2 Aide 3
Question 2 Aide 1 Aide 2 Aide 3 Aide 4 Aide 5 Aide 6 Aide 7 Aide 8

Sommaire
Concepts

Exemples
Exercices
Documents

Exercice A.1.3 Factorisation $A = QR$ par la méthode de Householder

$$\text{Soit } A = \begin{pmatrix} 1 & 3 \\ 1 & -1 \\ 1 & 1 \end{pmatrix},\ b = \begin{pmatrix} 1 \\ 5 \\ -6 \end{pmatrix}.$$

1. On veut résoudre $\min_{x \in \mathbf{R}^2} \|Ax - b\|_2^2$ à l'aide de la factorisation QR.

   (a) i. Déterminer la matrice de Householder, que l'on notera $H^{(1)}$, qui transforme $\dfrac{A_1}{\|A_1\|_2}$ en e_1.

   ii. Calculer $A^{(2)} = H^{(1)} A$.
   Réponse ;

$$H^{(1)} = \begin{pmatrix} a & a & a \\ a & b & c \\ a & c & b \end{pmatrix},\ A^{(2)} = \begin{pmatrix} \sqrt{3} & \sqrt{3} \\ 0 & \sqrt{3}-1 \\ 0 & \sqrt{3}+1 \end{pmatrix},$$

$$\text{avec } a = \frac{1}{\sqrt{3}},\ b = \frac{1}{2}\left(1 - \frac{1}{\sqrt{3}}\right),\ c = -\frac{1}{2}\left(1 + \frac{1}{\sqrt{3}}\right).$$

   iii. Comment obtient-on la factorisation QR de la matrice A ? Calculer alors les matrices $H^{(2)}$ et $A^{(3)} = R$, en déduire $\widetilde{R}$ puis Q.
   Réponse ;

$$H^{(2)} = \frac{1}{4 - \sqrt{6} + \sqrt{2}} \begin{pmatrix} a & b \\ b & -a \end{pmatrix},\ R = A^{(3)} = \begin{pmatrix} \sqrt{3} & \sqrt{3} \\ 0 & 2\sqrt{2} \\ 0 & 0 \end{pmatrix},$$

Sommaire
Concepts

Exemples
Exercices
Documents

$$Q = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & -\frac{2}{\sqrt{6}} \end{pmatrix},$$

avec $a = -2 + \sqrt{6} - \sqrt{2} + \sqrt{3},\ b = -1 + \sqrt{6} + \sqrt{2}$

(b) i. Calculer $Q^\top b$, on note $\widetilde{c}$ le vecteur constitué des deux premières composantes de $Q^\top b$.

ii. Montrer que $A^\top A\widetilde{x} = A^\top b \Leftrightarrow \widetilde{R}\widetilde{x} = \widetilde{c}$

iii. En déduire $\widetilde{x}$ solution du problème de minimisation.

2. On veut utiliser l'orthogonalisation de Schmidt. En orthogonalisant la famille $\{A_1, A_2\}$ à l'aide de l'orthogonalisation de Schmidt, on obtient la matrice :

$$\hat{A} = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & 0 \end{pmatrix}.$$

(a) Montrer que $A = \hat{A}\widetilde{R}$. Comparer $\hat{A}$ et Q, R et $\widetilde{R}$.

(b) On a montré, dans l'exercice précédent, que la solution $\hat{x}$ du problème de minimisation vérifiait $\widetilde{R}\hat{x} = \hat{A}^\top b$. Montrer que $\hat{A}^\top b = \widetilde{c}$. En déduire que l'on retrouve le système $\widetilde{R}\widetilde{x} = \widetilde{c}$.

Question 1(a)i Aide 1
Question 1(a)ii Aide 1
Question 1(a)iii Aide 1

Sommaire
Concepts

Exemples
Exercices
Documents

Index des concepts

Le gras indique un grain où le concept est défini ; l'italique indique un renvoi à un exercice ou un exemple, le gras italique à un document, et le romain à un grain où le concept est mentionné.

E
Equations normales **8**
Equations normales - projection orthogonale **14**
Equations normales - utilisation de l'orthogonalisation de Schmidt **16**

F
Factorisation QR **20**
Factorisation QR
application aux moindres carrés **22**

H
Householder-transformation **18**

L
Lissage **4**, 6

M
Moindres carrés
formulation matricielle **6**
intuition géométrique **11**

O
Orthogonaux-espaces **12**

Solution de l'exercice III.1

1. Soit $y = \alpha + \beta t$, l'équation de la droite considérée. Le problème de régression linéaire s'écrit

$$\min_{(\alpha,\beta)\in\mathbb{R}^2} E(\alpha,\beta)$$

où

$$E(\alpha,\beta) = \sum_{i=1}^{m} (\alpha + \beta t_i - b_i)^2.$$

La solution (α^*, β^*) donne les coefficients de la droite solution du problème de régression linéaire et elle vérifie

$$E(\alpha^*,\beta^*) = \min_{(\alpha,\beta)\in\mathbb{R}^2} E(\alpha,\beta).$$

2. Calculons les deux dérivées partielles

$$\begin{cases} \frac{\partial E}{\partial \alpha}(\alpha,\beta) = \sum_{i=1}^{m} 2(\alpha + \beta t_i - b_i) \\ \\ \frac{\partial E}{\partial \beta}(\alpha,\beta) = \sum_{i=1}^{m} 2(\alpha + \beta t_i - b_i) t_i \end{cases}$$

La solution (α^*, β^*) du problème de régression linéaire est donc donnée par la solution des deux équations linéaires obtenues en regroupant les termes

$$\begin{cases} \alpha^* m + \beta^* \sum_{i=1}^{m} t_i = \sum_{i=1}^{m} b_i \\ \\ \alpha^* \sum_{i=1}^{m} t_i + \beta^* \sum_{i=1}^{m} t_i^2 = \sum_{i=1}^{m} b_i t_i \end{cases}$$

Retour à l'exercice ▲

Soit

$$p(t) = \alpha_1 + \alpha_2 t + \ldots + \alpha_n t^{n-1},$$

un polynôme de degré $n-1$. Pour que ce polynôme approche les données $(t_i, b_i)_{i=1,\ldots,m}$ le plus près possible, il doit minimiser la quantité suivante

$$E(\alpha_1, \alpha_2, \ldots, \alpha_n) = \sum_{i=1}^{m} (p(t_i) - b_i)^2 = \|Ax - b\|_2^2.$$

En effet on peut écrire

$$\begin{pmatrix} p(t_1) - b_1 \\ p(t_2) - b_2 \\ \ldots \\ p(t_m) - b_m \end{pmatrix} = \begin{pmatrix} \alpha_1 + \alpha_2 t_1 + \ldots + \alpha_n t_1^{n-1} - b_1 \\ \alpha_1 + \alpha_2 t_2 + \ldots + \alpha_n t_2^{n-1} - b_2 \\ \ldots \\ \alpha_1 + \alpha_2 t_m + \ldots + \alpha_n t_m^{n-1} - b_m \end{pmatrix} = A \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \ldots \\ \alpha_n \end{pmatrix} - \begin{pmatrix} b_1 \\ b_2 \\ \ldots \\ b_m \end{pmatrix} = Ax - b,$$

où les matrices A et x sont définies par :

$$A = \begin{pmatrix} 1 & t_1 & \ldots & t_1^{n-1} \\ 1 & t_2 & \ldots & t_2^{n-1} \\ \ldots & \ldots & \ldots & \ldots \\ 1 & t_m & \ldots & t_m^{n-1} \end{pmatrix}, \ x = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}.$$

Dans les données, les points t_i sont tous distincts, ce qui implique que la matrice V, constituée des n premières lignes de A est inversible, d'où la matrice A est de rang n.

Retour à l'exercice ▲

Solution de l'exercice III.3

La matrice A du problème de régression linéaire s'écrit (voir la correction de l'exercice III.2) :

$$A = \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \dots & \dots \\ 1 & t_m \end{pmatrix}.$$

Les équations normales sont

$$A^T A\hat{x} = A^T b$$

ce qui donne en effectuant les produits matriciels

$$\begin{pmatrix} m & \sum_{i=1}^m t_i \\ \sum_{i=1}^m t_i & \sum_{i=1}^m t_i^2 \end{pmatrix} \hat{x} = \begin{pmatrix} \sum_{i=1}^m b_i \\ \sum_{i=1}^m t_i b_i \end{pmatrix}.$$

On retrouve bien ainsi le système de deux équations à deux inconnues de l'exercice III.1 .

Retour à l'exercice ▲

L'implication $\Leftarrow$ est évidente puisque l'on multiplie 0 par le vecteur z.

Supposons maintenant que

$$b^T z = 0, \forall z \in \mathbb{R}^n,$$

alors cette égalité étant vraie pour tout z l'est en particulier pour $z = b$, ce qui donne

$$b^T b = \|b\|^2|_2 = 0.$$

Or la norme d'un vecteur est nulle si et seulement si ce vecteur est nul, ce qui donne

$$b = 0.$$

Retour à l'exercice ▲

Solution de l'exercice III.5

Cet algorithme est très simple et suppose connues des fonctions telles que norme, produit scalaire ... ce qui est le cas de Scilab.

$E_1 = B_1 / \|B_1\|_2$
pour $k = 2, \ldots, n$ **faire**
  $\widetilde{E}_k = B_k - \sum_{j=1}^{k-1} \langle B_k, E_j \rangle E_j$
  $E_k = \widetilde{E}_k / \|\widetilde{E}_k\|_2$
fin pour

Retour à l'exercice ▲

On a donc

$$\langle E_i, E_i \rangle = 1,\ \langle E_i, E_j \rangle = 0 \text{ pour } i \neq j,$$

ou ce qui est équivalent

$$E_i^T E_i = 1,\ E_i^T E_j = 0 \text{ pour } i \neq j.$$

Les termes de la matrice (carrée) $C = E^T E$ sont donc

$$c_{ii} = E_i^T E_i = 1,\quad c_{ij} = E_i^T E_j = 0 \text{ pour } i \neq j,$$

C est donc la matrice identité.

Retour à l'exercice ▲

Solution de l'exercice III.7

1. Reprenons l'algorithme d'orthogonalisation de Schmidt, alors

$$E_1 = A_1/\|A_1\| \quad \Rightarrow \quad A_1 = \alpha_{11} E_1,$$

puis

$$\widetilde{E}_2 = A_2 - \langle A_1, E_1 \rangle E_1 \text{ et } E_2 = \widetilde{E}_2/\|\widetilde{E}_2\| \quad \Rightarrow \quad A_2 = \langle A_1, E_1 \rangle E_1 + \|\widetilde{E}_2\| E_2,$$

ce qui donne

$$A_2 = \alpha_{12} E_1 + \alpha_{22} E_2.$$

De manière générale

$$\widetilde{E}_k = A_k - \sum_{j=1}^{k-1} \langle A_k, E_j \rangle E_j \text{ et } E_k = \widetilde{E}_k/\|\widetilde{E}_k\| \quad \Rightarrow \quad A_k = \sum_{j=1}^{k-1} \langle A_k, E_j \rangle E_j + \|\widetilde{E}_k\| E_k,$$

ce qui donne

$$A_k = \sum_{j=1}^{k} \alpha_{jk} E_j.$$

2. Considérons le produit $C = ET$ de deux matrices, $E \in \mathcal{M}_{m,n}$ et $T \in \mathcal{M}_{n,n}$, alors

$$c_{ik} = \sum_{j=1}^{n} e_{ij} t_{jk}.$$

On peut aussi considérer c_{ik} comme le ième élément de la kième colonne de C. Alors cette colonne est donnée par

$$C_k = ET_k = \sum_{j=1}^{n} t_{jk} E_j.$$

Si l'on compare avec le résultat de la question précédente :

$$A_k = \sum_{j=1}^{k} \alpha_{jk} E_j,$$

on voit que $t_{jk} = \alpha_{jk}$ pour $j = 1, \ldots, k$ et que $t_{jk} = 0$ pour $j = k+1, \ldots, n$, ce qui correspond à une matrice triangulaires supérieure

$$T = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \ldots & \ldots & \alpha_{1n} \\ 0 & \alpha_{22} & \ldots & \ldots & \alpha_{2n} \\ 0 & \ldots & \vdots & \ldots & \ldots \\ 0 & \ldots & \ldots & \ldots & \alpha_{nn} \end{pmatrix}$$

La matrice T est inversible car $\alpha_{ii} = \|\widetilde{E}_i\|$.

Retour à l'exercice ▲

Solution de l'exercice III.8

1. On a
$$Q \text{ orthogonale} \quad \Leftrightarrow \quad Q^{-1} = Q^T \quad \Leftrightarrow \quad (Q^T)^{-1} = Q \quad \Leftrightarrow \quad Q^T \text{ orthogonale} .$$

2. On va démontrer le résultat pour le carré de l'expression, ce qui est équivalent pour des réels positifs. Dans ces équivalences, on utilise le fait que $Q^TQ = I$.
$$\|Qy\|_2^2 = (Qy)^TQy = y^TQ^TQy = y^Ty = \|y\|_2^2.$$

Retour à l'exercice ▲

Solution de l'exercice III.9

La matrice A et le vecteur b correspondants à ce problème sont

$$A = \begin{pmatrix} 1 & -1 \\ 1 & 0.5 \\ 1 & 2 \end{pmatrix} \quad , \quad b = \begin{pmatrix} 0.5 \\ 1 \\ 2.5 \end{pmatrix} .$$

Les étapes du calcul sont alors les suivantes :

– calcul de la décomposition QR par "qr" ce qui donne $A = QR$, où $R = \begin{pmatrix} \widetilde{R} \\ 0 \end{pmatrix}$,

– calcul de $Q^T b = \begin{pmatrix} c \\ d \end{pmatrix}$,

– résolution de $\widetilde{R}x = c$, ce qui donne $x = \begin{pmatrix} 1 \\ 0.666667 \end{pmatrix}$

– calcul de l'erreur $\|d\|_2^2 = 0.1666667$.

Retour à l'exercice ▲

Revoyez le lien entre la norme $\|.\|_2$ et le rayon spectral vu au chapitre 2.

$$(\chi_2(\widetilde{R}))^2 = \|\widetilde{R}\|_2^2 \|\widetilde{R}^{-1}\|_2^2 = \rho(\widetilde{R}^T\widetilde{R})\rho((\widetilde{R}^{-1})^T\widetilde{R}^{-1}).$$

On remarque que $A^TA = R^TQ^TQR = R^TR = \widetilde{R}^T\widetilde{R}$ donc

$$\chi_2(A^TA) = \chi_2(\widetilde{R}^T\widetilde{R}) = \|\widetilde{R}^T\widetilde{R}\|_2 \|\|(\widetilde{R}^T\widetilde{R})^{-1}\|_2,$$

or $\widetilde{R}^T\widetilde{R}$ et son inverse sont des matrices symétriques, toujours dans le chapitre 2, on a montré

$$\|R^TR\|_2 = \rho(R^TR),$$

on a donc également

$$\|(\widetilde{R}^T\widetilde{R})^{-1}\|_2 = \rho((\widetilde{R}^T\widetilde{R})^{-1}) = \rho(\widetilde{R}^{-1}(\widetilde{R}^T)^{-1}) = \rho((\widetilde{R}^T)^{-1}\widetilde{R}^{-1})$$

On a utilisé le résultat montré dans le chapitre 2 : $\rho(AB) = \rho(BA)$ on sait d'autre part que $(\widetilde{R}^T)^{-1} = (\widetilde{R}^{-1})^T$, ce qui permet de terminer la démonstration.

Ce résultat est important car dans le cas des équations normales on est conduit à résoudre un système dont la matrice est A^TA, dans le cas de la factorisation QR on est amené à résoudre un système dont la matrice est $\widetilde{R}$, comme vous le savez le conditionnement est toujours supérieur à 1 donc la matrice $\widetilde{R}$ a un conditionnement plus faible que la matrice A^TA, ce qui est intéressant numériquement.

Retour à l'exercice ▲

Le vecteur inconnu x est $x = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$.

On doit avoir par ailleurs

$$\begin{pmatrix} f_{a,b,c}(t_1) \\ f_{a,b,c}(t_2) \\ \ldots \\ f_{a,b,c}(t_p) \\ f_{a,b,c}(t_{p+1}) \\ \ldots \\ f_{a,b,c}(t_m) \end{pmatrix} = Ax,$$

déterminez la matrice A.

Retour à l'exercice ▲

Aide 2, Question 2a, Exercice A.1.1

$$\begin{pmatrix} f_{a,b,c}(t_1) \\ f_{a,b,c}(t_2) \\ \dots \\ f_{a,b,c}(t_p) \\ f_{a,b,c}(t_{p+1}) \\ \dots \\ f_{a,b,c}(t_m) \end{pmatrix} = \begin{pmatrix} b(e^{t_1}-1)+ct_1 \\ b(e^{t_2}-1)+ct_2 \\ \dots \\ b(e^{t_p}-1)+ct_p \\ a\ln(t_{p+1}+1)+ct_{p+1} \\ \dots \\ a\ln(t_m+1)+ct_m \end{pmatrix} = \begin{pmatrix} 0 & e^{t_1}-1 & t_1 \\ 0 & e^{t_2}-1 & t_2 \\ \dots & \dots & \dots \\ 0 & e^{t_p}-1 & t_p \\ \ln(t_{p+1}+1) & 0 & t_{p+1} \\ \dots & \dots & \dots \\ \ln(t_m+1) & 0 & t_m \end{pmatrix} x$$

d'où la matrice A.

Retour à l'exercice ▲

On sait que $p \geq 2$ et $m > p$, donc il y a au moins deux lignes du "premier type" et une ligne du "deuxième type", on peut donc extraire par exemple la matrice

$$\widehat{A} = \begin{pmatrix} 0 & e^{t_1} - 1 & t_1 \\ 0 & e^{t_2} - 1 & t_2 \\ \ln(t_m + 1) & 0 & t_m \end{pmatrix}$$

Montrez que cette matrice est inversible.

Retour à l'exercice ▲

$$\det \widehat{A} = \ln(t_m+1)(t_2(e^{t_1}-1) - t_1(e^{t_2}-1)) = \ln(t_m+1)t_1t_2\left(\frac{e^{t_1}-1}{t_1} - \frac{e^{t_2}-1}{t_2}\right)$$

Montrez que les quatre termes du produit sont non nuls.

Retour à l'exercice ▲

$$t_m > 0 \Rightarrow \ln(t_m + 1) > 0,$$

$$t_1 < t_2 < 0,$$

si l'on note $a(t) = \frac{(e^t - 1)}{t}$, si on appelle C la courbe d'équation $y = e^t$, $a(t)$ est égal à la pente de la droite qui joint les deux points de C $\Omega = (0, 1)$ et $M = (t, e^t)$, si $t_1 \neq t_2$ alors $a(t_1) \neq a(t_2)$, car si $a(t_1) = a(t_2)$ cela signifierait que les points M_1, M_2, Ω de C sont alignés, ce qui n'est pas possible, on aurait également pu montrer que la fonction $a(t)$ était strictement croissante. Donc

$$\frac{e^{t_1} - 1}{t_1} - \frac{e^{t_2} - 1}{t_2} \neq 0.$$

$\widehat{A}$ est donc inversible, A est de rang 3.

Retour à l'exercice ▲

Sur chaque intervalle $[\tau_i, \tau_{i+1}]$, $g(t)$ est un polynôme du premier degré en t dont les coefficients dépendent de i.
Ces coefficients doivent vérifier $g(\tau_i) = z_i, g(\tau_{i+1}) = z_{i+1}$.

Retour à l'exercice ▲

Sur l'intervalle $[\tau_i, \tau_{i+1}]$, on a donc

$$g(t) = \alpha_i(t - \tau_i) + z_i, \text{ avec } \quad \alpha_i = \frac{z_{i+1} - z_i}{\tau_{i+1} - \tau_i} = z_{i+1} - z_i.$$

On aurait pu également déterminer g à l'aide du polynôme d'interpolation de Lagrange, vérifier que l'on obtient bien la même chose.

Retour à l'exercice ▲

Avec le polynôme de Lagrange on obtient :

$$g(t) = z_i \frac{t - \tau_{i+1}}{\tau_i - \tau_{i+1}} + z_{i+1} \frac{t - \tau_i}{\tau_{i+1} - \tau_i} = -z_i \,(t - \tau_{i+1}) + z_{i+1} \,(t - \tau_i).$$

C'est bien sûr la même chose.

Retour à l'exercice ▲

Les τ_i, les t_k et y_k sont donnés.
Quels sont les paramètres inconnus ?
Quel est leur nombre ?

Retour à l'exercice ▲

Les six inconnues sont $z_0, z_1, ..., z_5$.

$$\text{On peut noter } z = \begin{pmatrix} z_0 \\ z_1 \\ z_2 \\ z_3 \\ z_4 \\ z_5 \end{pmatrix}.$$

Retour à l'exercice ▲

Ecrivez le vecteur $\begin{pmatrix} g(t_1) \\ g(t_2) \\ g(t_3) \\ g(t_4) \\ g(t_5) \\ g(t_6) \\ g(t_7) \\ g(t_8) \\ g(t_9) \\ g(t_{10}) \end{pmatrix}$ sous la forme Az. Quelle est la taille de la matrice A ?

Retour à l'exercice ▲

Avant tout calcul on sait que A doit avoir 10 lignes et 6 colonnes.

$$\begin{cases} t_1 = \frac{1}{2} \in [0,\, 1] = [\tau_0,\, \tau_1] \Longrightarrow i = 0 & \textbf{donc} \quad g(t_1) = z_0\left(\tau_1 - \frac{1}{2}\right) + z_1\left(\frac{1}{2} - \tau_0\right) = \frac{1}{2}z_0 + \frac{1}{2}z_1 \\ t_2 = 1 \in [0,\, 1] \Longrightarrow i = 0 & \textbf{donc} \quad g(t_2) = g(\tau_1) = z_0\,(1-1) + z_1\,(1-0) = z_1 \\ t_3 = \frac{3}{2} \in [1,\, 2] = [\tau_1,\, \tau_2] \Longrightarrow i = 1 & \textbf{donc} \quad g(t_3) = z_1\left(\tau_2 - \frac{3}{2}\right) + z_1\left(\frac{3}{2} - \tau_1\right) = \frac{1}{2}z_1 + \frac{1}{2}z_2 \\ t_4 = 2 \in [1,\, 2] & \textbf{donc} \quad g(t_4) = z_2 \\ \vdots & \\ \vdots & \\ t_{10} = 5 & \textbf{donc} \quad g(t_{10}) = z_5 \end{cases}$$

Que vaut A ?

Retour à l'exercice ▲

Aide 5, Question 2, Exercice A.1.2

Si on note

$$A = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

on a bien

$$Az = \begin{pmatrix} g(t_1) \\ g(t_2) \\ g(t_3) \\ g(t_4) \\ g(t_5) \\ g(t_6) \\ g(t_7) \\ g(t_8) \\ g(t_9) \\ g(t_{10}) \end{pmatrix}$$

Si l'on note

$$y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \\ y_{10} \end{pmatrix},$$

on cherche z qui minimise $\|Az - y\|_2$.

Retour à l'exercice ▲

On calcule maintenant le rang de A.

Retour à l'exercice ▲

Puisque A a 6 colonnes, on sait que rang $A \leq 6$.

Essayez d'extraire de A une matrice carrée $\tilde{A}$ inversible ayant 6 lignes, on aura alors rang $A \geq 6$, donc rang $A = 6$.

Retour à l'exercice ▲

On peut choisir

$$\tilde{A} = \begin{pmatrix} \underline{A_1} \\ \underline{A_2} \\ \underline{A_4} \\ \underline{A_6} \\ \underline{A_8} \\ \underline{A_{10}} \end{pmatrix}, \quad \text{on a } \det \tilde{A} = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{vmatrix} = \frac{1}{2} \neq 0$$

donc $\tilde{A}$ est inversible, donc rang $A = 6$.

Retour à l'exercice ▲

Si l'on note

$$f_1 = \frac{A_1}{\|A_1\|_2} = \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}$$

On a

$$H^{(1)} = I - 2uu^T, \text{ avec } u = \frac{f_1 - e_1}{\|f_1 - e_1\|_2}$$

$$f_1 - e_1 = \begin{pmatrix} \frac{1}{\sqrt{3}} - 1 \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}, \ \|f_1 - e_1\|_2^2 = 2 - \frac{2}{\sqrt{3}}, \ \frac{1}{\|f_1 - e_1\|_2^2} = \frac{3}{4}\left(1 + \frac{1}{\sqrt{3}}\right)$$

$$uu^T = \frac{1}{\|f_1 - e_1\|_2^2}(f_1 - e_1)(f_1 - e_1)^T.$$

Pour calculer $(f_1 - e_1)(f_1 - e_1)^T$ il y a seulement trois termes différents à calculer :

$$\alpha = \left(\frac{1}{\sqrt{3}} - 1\right)^2, \ \beta = \left(\frac{1}{\sqrt{3}} - 1\right) \times \frac{1}{\sqrt{3}} = \frac{1}{3} - \frac{1}{\sqrt{3}}, \ \gamma = \frac{1}{\sqrt{3}} \times \frac{1}{\sqrt{3}} = \frac{1}{3},$$

$$(f_1 - e_1)(f_1 - e_1)^T = \begin{pmatrix} \alpha & \beta & \beta \\ \beta & \gamma & \gamma \\ \beta & \gamma & \gamma \end{pmatrix}, \ uu^T = \begin{pmatrix} \hat{\alpha} & \hat{\beta} & \hat{\beta} \\ \hat{\beta} & \hat{\gamma} & \hat{\gamma} \\ \hat{\beta} & \hat{\gamma} & \hat{\gamma} \end{pmatrix}, \ \text{ où } \hat{\alpha} = \frac{1}{2}\left(1 - \frac{1}{\sqrt{3}}\right), \ \hat{\beta} = -\frac{1}{2\sqrt{3}}, \ \hat{\gamma} = \frac{1}{4}\left(1 + \frac{1}{\sqrt{3}}\right)$$

On écrit enfin

$$H^{(1)} = I - 2uu^T.$$

Retour à l'exercice ▲

Aide 1, Question 1(a)ii, Exercice A.1.3

Il suffit d'éffectuer le produit, il est cependant possible de prévoir la première colonne de $A^{(2)}$ sans calculs, on a en effet

$$A_1^{(2)} = H^{(1)} A_1 = \|A_1|_2 H^{(1)} f_1 = \|A_1\|_2 e_1 = \begin{pmatrix} \sqrt{3} \\ 0 \\ 0 \end{pmatrix}$$

Retour à l'exercice ▲

On note

$$v = \begin{pmatrix} \sqrt{3}-1 \\ \sqrt{3}-1 \end{pmatrix},$$

$$f_2 = \frac{v}{\|v\|_2}, e_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, u_2 = \frac{f_2 - e_2}{\|f_2 - e_2\|_2}, H^{(2)} = I - 2u_2u_2^T,\ U = \begin{pmatrix} 1 & 0 \\ 0 & H^{(2)} \end{pmatrix}, A^{(3)} = UA^{(2)}$$

On a

$$R = A^{(3)}, Q = H^{(1)}U.$$

Retour à l'exercice ▲