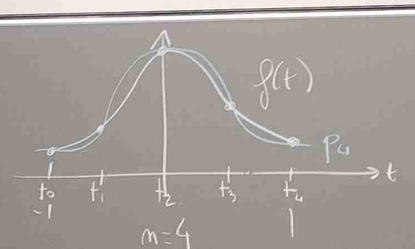
Chap 1 Line. Interpolation

lagrange: $f:[a,b]\to\mathbb{R}$ donnée

$t_j = a + \frac{b-a}{n} j$, $j=0,1,\dots,n$

Ex: $f(t) = \frac{1}{1+25t^2}$ $[a,b]=[-1,1]$



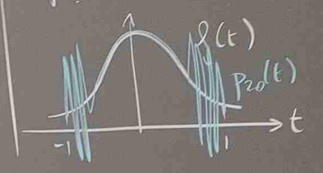
On cherche $p_n \in \mathbb{P}_n$ tq

$p_n(t_j) = f(t_j)$ $j=0,1,\dots,n$

$p_n(t) = \sum_{j=0}^{n} f(t_j)\,\varphi_j(t)$

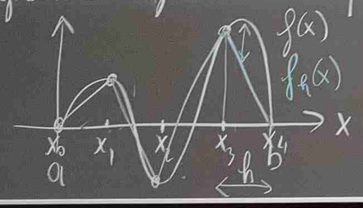
$\varphi_0(t) = \frac{(t-t_2)(t-t_n)\cdots(t-t_n)}{(t_0-t_1)(t_0-t_2)\cdots(t_0-t_n)}$

$p_n \xrightarrow{n\to\infty} f$? pas toujours



Interp. par intervalles:

polynôme de degré 1 sur chaque intervalle



$x_i = a + \underbrace{\frac{b-a}{N}}_{h} i$, $i=0,1,\dots,N$

$f_h \in \mathcal{C}^0[a,b]$

$f_h|_{[x_i,x_{i+1}]} \in \mathbb{P}_1$

$f_h(x_i) = f(x_i)$ $i=0,\dots,N$

Théorème (1,2 ligne n=1):

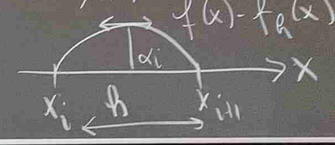
$\forall f \in \mathcal{C}^2[a,b]\ \exists C>0\ \forall 0<h<b-a$

$\max_{a\le x\le b} |f(x) - f_h(x)| \le C h^2$

Interprétation: • C dépend de f mais pas de h

• Exp. num: on choisit $f \in \mathcal{C}^2$ l'erreur est (au mieux) divisée par $2^2 = 4$ chaque fois que h est divisé par 2

Dem: Soit $f \in \mathcal{C}^2[a,b]$, soit $i=0,1,\dots,N-1$



Rolle: $\exists x_i < \alpha_i < x_{i+1}$ tq $w'(\alpha_i) = 0$

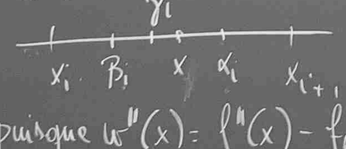
Soit $x_i < x < x_{i+1}$, on a: $w(x) = w(x_i) + (x-x_i)\,w'(\beta_i)$ ($w(x_i) = 0$)

où $x_i < \beta_i < x$

Puisque $w = f - f_h \in \mathcal{C}^2\ \exists x_i, x_{i+1}[$, on a

$w'(\beta_i) = \underbrace{w'(\alpha_i)}_{0} + (\beta_i - \alpha_i)\,w''(\gamma_i)$

où $\min(\alpha_i,\beta_i) < \gamma_i < \max(\alpha_i,\beta_i)$



puisque $w''(x) = f''(x) - f_h''(x)$ ($f_h'' = 0$)

on a $|f(x) - f_h(x)| = |w(x)| = |(x-x_i)(\beta_i-\alpha_i)\,f''(\gamma_i)| \le h\,h \max_{a\le y\le b} |f''(y)|$ (C)

On a démontré que $\forall i=0,\dots,N-1\ \forall\ x_i < x < x_{i+1}$,

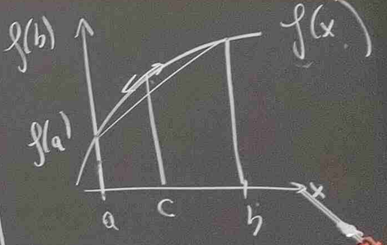
$|f(x) - f_h(x)| \le C h^2$ avec $C = \max_{a\le y\le b} |f''(y)|$

D'autre part si $x = x_i$ $f(x) - f_h(x) = 0$

Et donc $\max_{a\le x\le b} |f(x) - f_h(x)| \le C h^2$

$x_i < x < x_{i+1}$

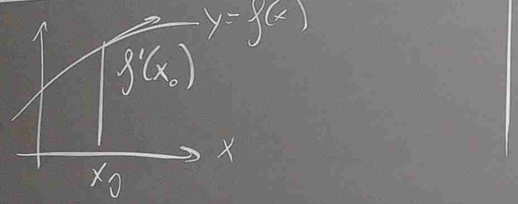
$f(b) - f(a) = (b-a) f'(a) + \frac{1}{2}(b-a)^2 f''(c)$



$\frac{f(b)-f(a)}{b-a} = f'(c)$ $c \in]a,b[$

Chap 2 Formules de différences finies (FDF)

$f : \mathbb{R} \to \mathbb{R}$, $x \to f(x)$



$$\left| f'(x_0) - \frac{f(x_0+h) - f(x_0)}{h} \right| = O(h) \quad (\text{si } f \text{ est } \mathcal{C}^2)$$ progressive

$$\left| f'(x_0) - \frac{f(x_0) - f(x_0-h)}{h} \right| = O(h)$$ rétrograde

$$\left| f'(x_0) - \frac{f(x_0+h/2) - f(x_0-h/2)}{h} \right| = O(h^2) \quad (\text{si } f \text{ est } \mathcal{C}^3)$$

$$\left| f''(x_0) - \frac{f(x_0-h) - 2f(x_0) + f(x_0+h)}{h^2} \right| = O(h^2)$$

FDF centrée (si f est $\mathcal{C}^4$)

Il existe un grand nombre de FDF

Auj.

$$\left| f'(x_0) - \frac{3f(x_0) - 4f(x_0-h) + f(x_0-2h)}{2h} \right| = O(h^2)$$

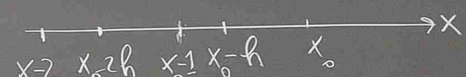
(si f est $\mathcal{C}^3$)

Thm : $\forall x_0 \in \mathbb{R}$ $\forall f \in \mathcal{C}^3[x_0-2, x_0]$ $\exists C > 0$ $\forall 0 < h \leq 1$ on a

$$\left| f'(x_0) - \frac{3f(x_0) - 4f(x_0-h) + f(x_0-2h)}{2h} \right| \leq C h^2$$

Interpr :
- C dépend de x_0 et f mais pas de h
- fixe x_0 et f, l'erreur est divisée par $2^2 = 4$ si h est divisé par 2.

Dem : Soit $x_0 \in \mathbb{R}$, soit $f \in \mathcal{C}^3[x_0-2, x_0]$, soit $0 < h \leq 1$



On a :

$$4\left(f(x_0-h) = f(x_0) - h f'(x_0) + \frac{h^2}{2} f''(x_0) - \frac{h^3}{6} f'''(\alpha)\right) \quad \text{où } x_0-h \leq \alpha \leq x_0$$

$$-\left(f(x_0-2h) = f(x_0) - 2h f'(x_0) + \frac{4h^2}{2} f''(x_0) - \frac{8h^3}{6} f'''(\beta)\right) \quad \text{où } x_0-2h \leq \beta \leq x_0$$

$$4f(x_0-h) - f(x_0-2h) = 3f(x_0) - 2h f'(x_0) - \frac{4h^3}{6} f'''(\alpha) + \frac{8h^3}{6} f'''(\beta)$$

$$\left| f'(x_0) - \frac{3f(x_0) - 4f(x_0-h) + f(x_0-2h)}{2h} \right| = \left| -\frac{h^2}{3} f'''(\alpha) + \frac{2h^3}{3} f'''(\beta) \right| \leq h^2 \left(\frac{1}{3}|f'''(\alpha)| + \frac{2}{3}|f'''(\beta)| \right)$$

attention α, β dép. de h !

$$\leq h^2 \left(\frac{1}{3} + \frac{2}{3}\right) \max_{x_0-2 \leq x \leq x_0} |f'''(x)|$$

On a donc le thm avec $C = \max_{x_0-2 \leq x \leq x_0} |f'''(x)|$ dépend de x_0 et f mais pas de h.

Série $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2} + \dots + \frac{x^n}{n!} + \dots$

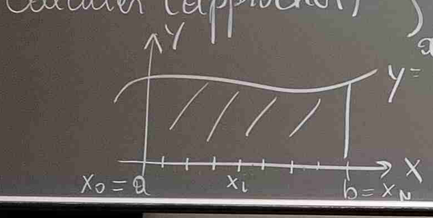
dév. limité $e^x = 1 + x + \frac{x^2}{2} e^{\alpha}$ $0 < \alpha < x$ en $x_0 = 0$, $x > 0$

$|a+b| \leq |a| + |b|$

$|a-b| \leq |a| + |b|$

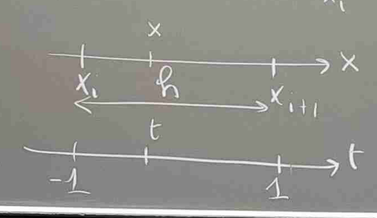
Chap 3 Formules de quadrature

Calculer (approcher) $\int_a^b f(x)dx$



$x_i = a + \frac{b-a}{N} i \quad i = 0, \dots, N$ ($h = \frac{b-a}{N}$)

$\int_a^b f(x)dx = \sum_{i=0}^{N-1} \int_{x_i}^{x_{i+1}} f(x)dx$



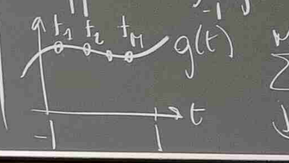
$x = x_i + h\,\frac{t+1}{2}$

$dx = \frac{h}{2}\,dt$

$\int_a^b f(x)dx = \frac{h}{2} \sum_{i=0}^{N-1} \int_{-1}^{1} \underbrace{f\left(x_i + h\,\frac{t+1}{2}\right)}_{g_i(t)} dt$

Etant donné $g : [-1,1] \to \mathbb{R}$, $t \to g(t)$

on approche $\int_{-1}^{1} g(t)dt$ par



$\sum_{j=1}^{M} w_j\, g(t_j)$ — $J(g)$ form. quadr.

On approche $\int_a^b f(x)dx$ par

$\frac{h}{2} \sum_{i=0}^{N-1} J(g_i) = \frac{h}{2} \sum_{i=0}^{N-1} \sum_{j=1}^{M} w_j\, g_i(t_j)$

$= \frac{h}{2} \sum_{i=0}^{N-1} \sum_{j=1}^{M} w_j\, f\left(x_i + h\,\frac{t_j+1}{2}\right)$ $\quad L_h(f)$

Aujourd'hui : Formule de Simpson

$\int_{-1}^{1} g(t)dt$ approché par $J(g) = \frac{1}{3}g(-1) + \frac{4}{3}g(0) + \frac{1}{3}g(1)$

On a

$\forall p \in \mathbb{P}_3 \quad \int_{-1}^{1} p(t)dt = J(p)$

$= \frac{1}{3}p(-1) + \frac{4}{3}p(0) + \frac{1}{3}p(1)$

et on peut dém.

Thm : (3, ligne 2=3)

$\forall f \in \mathcal{C}^4[a,b] \ \exists C > 0 \ \forall 0 < h < b-a$

$\left| \int_a^b f(x)dx - L_h(f) \right| \leq C h^4$

Interp : • C dépend de f mais pas de h

• On choisit a, b fixés, l'erreur est (au moins) divisée par $2^4 = 16$ si h est divisé par 2

Remarque : $L_h(f) = \frac{h}{2} \sum_{i=0}^{N-1} \left(\frac{1}{3} f(x_i) + \frac{4}{3} f\left(x_i + \frac{h}{2}\right) + \frac{1}{3} f(x_{i+1}) \right)$

Dem : $E = \int_a^b f(x)dx - L_h(f) = \frac{h}{2} \sum_{i=0}^{N-1} \left(\int_{-1}^{1} g_i(t)dt - J(g_i) \right)$

Si on peut écrire : $g_i(t) = p_i(t) + r_i(t)$ où $p_i \in \mathbb{P}_3$ (*)

$-1 \leq t \leq 1$, $i = 0, \dots, N-1$

alors on a :

$E = \frac{h}{2} \sum_{i=0}^{N-1} \left(\int_{-1}^{1} p_i(t) + r_i(t)dt - J(p_i + r_i) \right)$

puisque $\int_{-1}^{1} p_i(t)dt = J(p_i)$ on a :

$E = \frac{h}{2} \sum_{i=0}^{N-1} \left(\int_{-1}^{1} r_i(t)dt - J(r_i) \right)$

dem de (*) :

$g_i(t) = g_i(0) + t\,g_i'(0) + \frac{t^2}{2} g_i''(0) + \frac{t^3}{6} g_i'''(0) + \frac{t^4}{24} g_i^{(4)}(\alpha_i)$

où $\min(0,t) \leq \alpha_i \leq \max(0,t)$ $\quad (-1 \leq t \leq 1)$

$g_i'(t) = f\left(x_i + h\,\frac{t+1}{2}\right), \quad g_i''(t) = f'\left(x_i + h\,\frac{t+1}{2}\right)\frac{h}{2}$

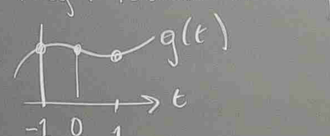
$g_i''(t) = f''\left(x_i + h\,\frac{t+1}{2}\right)\frac{h^2}{4}, \quad g_i'''(t) = f'''\left(x_i + h\,\frac{t+1}{2}\right)\frac{h^3}{8}$

On approche $\int_a^b f(x)dx$ par

$\frac{h}{2} \sum_{i=0}^{N-1} J(g_i) = \frac{h}{2} \sum_{i=0}^{N-1} \sum_{j=1}^{M} w_j\, g_i(t_j)$

$= \frac{h}{2} \sum_{i=0}^{N-1} \sum_{j=1}^{M} w_j\, f\left(x_i + h\,\frac{t_j+1}{2}\right)$ $\quad L_h(f)$

Aujourd'hui : Formule de Simpson



$\int_{-1}^{1} g(t)dt$ approché par $J(g) = \frac{1}{3}g(-1) + \frac{4}{3}g(0) + \frac{1}{3}g(1)$

On a

$\forall p \in \mathbb{P}_3 \quad \int_{-1}^{1} p(t)dt = J(p)$

$= \frac{1}{3}p(-1) + \frac{4}{3}p(0) + \frac{1}{3}p(1)$

et on peut dém.

Thm : (3, ligne 2=3)

$\forall f \in \mathcal{C}^4[a,b] \ \exists C > 0 \ \forall 0 < h < b-a$

$\left| \int_a^b f(x)dx - L_h(f) \right| \leq C h^4$

Interp : • C dépend de f mais pas de h

• On choisit a, b fixés, l'erreur est (au moins) divisée par $2^4 = 16$ si h est divisé par 2

Remarque : $L_h(f) = \frac{h}{2} \sum_{i=0}^{N-1} \left(\frac{1}{3} f(x_i) + \frac{4}{3} f\left(x_i + \frac{h}{2}\right) + \frac{1}{3} f(x_{i+1}) \right)$

Dem : $E = \int_a^b f(x)dx - L_h(f) = \frac{h}{2} \sum_{i=0}^{N-1} \left(\int_{-1}^{1} g_i(t)dt - J(g_i) \right)$

Si on peut écrire : $g_i(t) = p_i(t) + r_i(t)$ où $p_i \in \mathbb{P}_3$ (*)

$-1 \leq t \leq 1$, $i = 0, \dots, N-1$

alors on a :

$E = \frac{h}{2} \sum_{i=0}^{N-1} \left(\int_{-1}^{1} p_i(t) + r_i(t)dt - J(p_i + r_i) \right)$

puisque $\int_{-1}^{1} p_i(t)dt = J(p_i)$ on a :

$E = \frac{h}{2} \sum_{i=0}^{N-1} \left(\int_{-1}^{1} r_i(t)dt - J(r_i) \right)$

dem de (*) :

$g_i(t) = g_i(0) + t\,g_i'(0) + \frac{t^2}{2} g_i''(0) + \frac{t^3}{6} g_i'''(0) + \frac{t^4}{24} g_i^{(4)}(\alpha_i)$

où $\min(0,t) \leq \alpha_i \leq \max(0,t)$ $\quad (-1 \leq t \leq 1)$

$g_i'(t) = f\left(x_i + h\,\frac{t+1}{2}\right), \quad g_i''(t) = f'\left(x_i + h\,\frac{t+1}{2}\right)\frac{h}{2}$

$g_i''(t) = f''\left(x_i + h\,\frac{t+1}{2}\right)\frac{h^2}{4}, \quad g_i'''(t) = f'''\left(x_i + h\,\frac{t+1}{2}\right)\frac{h^3}{8}$

Donc $g_i(t) = \underbrace{f\left(x_i + \frac{h}{2}\right) + \frac{t}{2}h\, f'\left(x_i + \frac{h}{2}\right) + \frac{t^2}{2}\frac{h^2}{4} f''\left(x_i + \frac{h}{2}\right) + \frac{t^3}{6}\frac{h^3}{8} f'''\left(x_i + \frac{h}{2}\right)}_{p_i(t)} + \underbrace{\frac{t^4}{24}\frac{h^4}{16} f^{(4)}(x_i + h\alpha_i)}_{r_i(t)}$

(attention α_i dépend de t)

Finalement $|E| = \left| \frac{h}{2} \sum_{i=0}^{N-1} \left(\int_{-1}^{1} r_i(t)dt - J(r_i) \right) \right| \leq \frac{h}{2} \sum_{i=0}^{N-1} \left| \int_{-1}^{1} r_i(t)dt - J(r_i) \right| \leq \frac{h}{2} \sum_{i=0}^{N-1} \left(\left| \int_{-1}^{1} r_i(t)dt \right| + |J(r_i)| \right)$

On a : $\left| \int_{-1}^{1} r_i(t)dt \right| \leq \int_{-1}^{1} |r_i(t)|dt$

$\leq 2 \max_{-1 \leq t \leq 1} |r_i(t)|$

$\leq 2\frac{1}{24.16} h^4 \max_{a \leq x \leq b} |f^{(4)}(x)|$

$|J(r_i)| = \left| \frac{1}{3} r_i(-1) + \frac{4}{3} \underbrace{r_i(0)}_{0} + \frac{1}{3} r_i(1) \right| \leq \frac{1}{3}|r_i(-1)| + \frac{1}{3}|r_i(1)|$

$\leq \frac{2}{3}\frac{1}{24.16} h^4 \max_{a \leq x \leq b} |f^{(4)}(x)|$

$|a+b| \leq |a|+|b|$

$|a-b| \leq |a|+|b|$

$|E| \leq \frac{h}{2} \sum_{i=0}^{N-1} \frac{h^4}{24.16} \max_{a \leq x \leq b} |f^{(4)}(x)| \left(2 + \frac{2}{3}\right)$

$= \frac{h^4}{24.16} \max_{a \leq x \leq b} |f^{(4)}(x)| \left(2 + \frac{2}{3}\right) \frac{h}{2} \sum_{i=0}^{N-1} 1$

$C = \frac{b-a}{2.24.16}\left(2 + \frac{2}{3}\right) \max_{a \leq x \leq b} |f^{(4)}(x)|$ $\quad \frac{hN}{2} = \frac{b-a}{2}$

Chap 4, 5, 6 Lime

Ce qu'il faut absolument savoir faire

Etant donné $\vec{b} \in \mathbb{R}^N$ $A \in \mathbb{R}^{N \times N}$

trouver $\vec{x} \in \mathbb{R}^N$ tel que $A\vec{x} = \vec{b}$

$$\begin{pmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{N1} & \cdots & a_{NN} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_N \end{pmatrix}$$

Ex :

$$A = \begin{pmatrix} 2 & -1 & (0) \\ -1 & \ddots & \ddots \\ (0) & \ddots & 2 \end{pmatrix}$$

Lien entre résol de syst lin et minimisation d'une fonctionnelle

On suppose que A est symétrique ($A = A^T$) définie positive $\forall \vec{x} \in \mathbb{R}^N$ $\vec{x}^T A \vec{x} \geq 0$ et $(\vec{x}^T A \vec{x} = 0) \Leftrightarrow (\vec{x} = \vec{0})$

$$= \sum_{i,j=1}^{N} a_{ij} x_i x_j$$

Notation : $\vec{x} \in \mathbb{R}^N$ $\vec{y} \in \mathbb{R}^N$ produit scalaire

$$\langle \vec{x}, \vec{y} \rangle = \vec{x}^T \vec{y} = \vec{y}^T \vec{x}$$

$$= (x_1 \cdots x_N) \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} = \sum_{i=1}^{N} x_i y_i = (y_1 \cdots y_N) \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix}$$

et donc

$$\vec{x}^T A \vec{x} = \langle \vec{x}, A\vec{x} \rangle = \sum_{i=1}^{N} x_i (A\vec{x})_i = \sum_{i=1}^{N} x_i \sum_{j=1}^{N} a_{ij} x_j$$

Remarque : si A est s.d.p alors A est inversible ($\ker A = \vec{0}$)

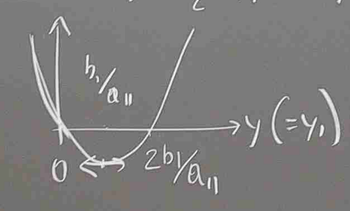
On introduit la fonctionnelle $\mathcal{L}$ définie par

$$\mathcal{L} : \mathbb{R}^N \to \mathbb{R}$$

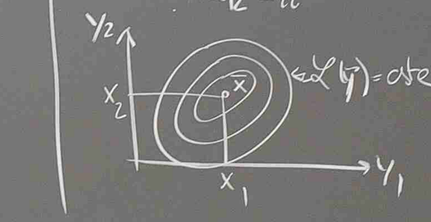
$$\vec{y} \to \mathcal{L}(\vec{y}) = \frac{1}{2} \vec{y}^T A \vec{y} - \vec{b}^T \vec{y}$$

$$= \frac{1}{2} \sum_{i,j=1}^{N} a_{ij} y_i y_j - \sum_{i=1}^{N} b_i y_i$$

$N = 1$ $\mathcal{L}(y) = \frac{1}{2} a_{11} y^2 - b_1 y$ $a_{11} > 0$



$N = 2$ $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}$



On veut démontrer :

$$\left(\text{Trouver } \vec{x} \in \mathbb{R}^N \text{ tel que } A\vec{x} = \vec{b} \right)$$

$$\Leftrightarrow$$

$$\left(\text{Trouver } \vec{x} \in \mathbb{R}^N \text{ tq } \mathcal{L}(\vec{x}) \leq \mathcal{L}(\vec{y}) \ \forall \vec{y} \in \mathbb{R}^N \right)$$

---

opérateur gradient $\overrightarrow{\text{grad}}\, \mathcal{L}(\vec{y})$ (noté $\vec{\nabla}$ nabla)

$\forall \vec{y} \in \mathbb{R}^N$ on note $\overrightarrow{\text{grad}}\, \mathcal{L}(\vec{y}) = \begin{pmatrix} \frac{\partial}{\partial y_1} \mathcal{L}(\vec{y}) \\ \frac{\partial}{\partial y_2} \mathcal{L}(\vec{y}) \\ \vdots \\ \frac{\partial}{\partial y_N} \mathcal{L}(\vec{y}) \end{pmatrix}$

(On a : $\overrightarrow{\text{grad}}\, \mathcal{L}(\vec{y}) = A\vec{y} - \vec{b}$)

En effet montrons $k = 1, \ldots, N$

$$\frac{\partial}{\partial y_k} \mathcal{L}(\vec{y}) = (A\vec{y} - \vec{b})_k = (A\vec{y})_k - b_k = \sum_{i=1}^{N} a_{ki} y_i - b_k$$

$$\frac{\partial}{\partial y_k} \mathcal{L}(\vec{y}) = \frac{\partial}{\partial y_k} \left(\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} y_i y_j - \sum_{i=1}^{N} b_i y_i \right)$$

k fixé, $k = 1, \ldots, N$

$$= \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} \left(\underbrace{\frac{\partial y_i}{\partial y_k}}_{\substack{1 \text{ si } i=k \\ 0 \text{ sinon}}} y_j + \underbrace{\frac{\partial y_j}{\partial y_k}}_{\substack{1 \text{ si } j=k \\ 0 \text{ sinon}}} y_i \right) - \sum_{i=1}^{N} b_i \underbrace{\frac{\partial y_i}{\partial y_k}}_{\substack{1 \text{ si } i=k, 0 \text{ sinon}}}$$

$$= \frac{1}{2} \left(\sum_{j=1}^{N} a_{kj} y_j + \sum_{i=1}^{N} \underbrace{a_{ik}}_{=a_{ki}} y_i \right) - b_k = \sum_{i=1}^{N} a_{ki} y_i - b_k = (A\vec{y} - \vec{b})_k$$

Montrons le thm :

$\Leftarrow$: Soit $\bar{x}$ tq $\mathcal{L}(\bar{x}) \leq \mathcal{L}(\vec{y})$ $\forall \vec{y} \in \mathbb{R}^N$

D'après le cours d'analyse II

$\overrightarrow{\text{grad}}\, \mathcal{L}(\bar{x}) = \vec{0} = A\bar{x} - \vec{b}$

$\Rightarrow$: Soit $\bar{x} \in \mathbb{R}^N$ tq $A\bar{x} = \vec{b}$ montrons

$\mathcal{L}(\bar{x}) \leq \mathcal{L}(\vec{y})$ $\forall \vec{y} \in \mathbb{R}^N$

On a $\forall \vec{y} \in \mathbb{R}^N$

$$\mathcal{L}(\vec{y}) = \mathcal{L}(\bar{x} + \vec{y} - \bar{x}) = \frac{1}{2} (\bar{x} + \vec{y} - \bar{x})^T A (\bar{x} + \vec{y} - \bar{x}) - \vec{b}^T (\bar{x} + \vec{y} - \bar{x})$$

$$= \frac{1}{2} \bar{x}^T A \bar{x} + \frac{1}{2} (\vec{y} - \bar{x})^T A \bar{x} + \frac{1}{2} \bar{x}^T A (\vec{y} - \bar{x}) + \frac{1}{2} (\vec{y} - \bar{x})^T A (\vec{y} - \bar{x}) - \vec{b}^T \bar{x} - \vec{b}^T (\vec{y} - \bar{x})$$

Or $\bar{x}^T A (\vec{y} - \bar{x}) = \left(A(\vec{y} - \bar{x}) \right)^T \bar{x} = (\vec{y} - \bar{x})^T A^T \bar{x} = (\vec{y} - \bar{x})^T A \bar{x}$

Donc $\mathcal{L}(\vec{y}) = \mathcal{L}(\bar{x}) + (\vec{y} - \bar{x})^T \underbrace{(A\bar{x} - \vec{b})}_{\vec{0}} + \frac{1}{2} \underbrace{(\vec{y} - \bar{x})^T A (\vec{y} - \bar{x})}_{\geq 0}$

et donc $\mathcal{L}(\vec{y}) \geq \mathcal{L}(\bar{x})$

Chap 8. Système d'éq. non linéaires

On cherche $x_1, x_2, \dots, x_N \in \mathbb{R}$ tels que

$$\begin{cases} 2x_1 - x_2 + e^{x_1} = 0 \\ -x_1 + 2x_2 - x_3 + e^{x_2} = 0 \\ \vdots \\ -x_{N-2} + 2x_{N-1} - x_N + e^{x_{N-1}} = 0 \\ -x_{N-1} + 2x_N + e^{x_N} = 0 \end{cases}$$

On écrit ce problème sous la forme

trouver $\vec{x} \in \mathbb{R}^N$ tel que $\vec{F}(\vec{x}) = \vec{0}$

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} \qquad \vec{F}(\vec{x}) = \begin{pmatrix} F_1(x_1, x_2, \dots, x_N) \\ \vdots \\ F_N(x_1, x_2, \dots, x_N) \end{pmatrix} = A\vec{x} + e^{\vec{x}}$$

où $A = \begin{pmatrix} 2 & -1 & & (0) \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ (0) & & -1 & 2 \end{pmatrix}$ et $e^{\vec{x}} = \begin{pmatrix} e^{x_1} \\ \vdots \\ e^{x_N} \end{pmatrix}$

Méthode de Newton : étant donné $\vec{x}^k$ on cherche $\vec{x}^{k+1}$ tel que

$$\underbrace{DF(\vec{x}^k)}_{\text{connu}\ \in \mathbb{R}^{N\times N}} \underbrace{(\vec{x}^{k+1} - \vec{x}^k)}_{\text{inconnu}\ \in \mathbb{R}^N} = \underbrace{-\vec{F}(\vec{x}^k)}_{\text{connu}\ \in \mathbb{R}^N}$$

où $\left(DF(\vec{x})\right)_{ij} = \dfrac{\partial F_i(\vec{x})}{\partial x_j}$ $i,j = 1, \dots, N$

$$DF(\vec{x}) = \begin{pmatrix} 2+e^{x_1} & -1 & & (0) \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ (0) & & -1 & 2+e^{x_N} \end{pmatrix}$$

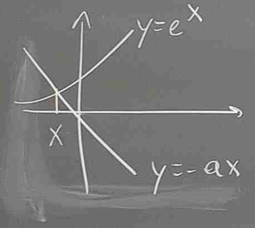
Intuition formule : $\vec{0} = \vec{F}(\vec{x}^{k+1}) = \vec{F}(\vec{x}^k) + DF(\vec{x}^k)(\vec{x}^{k+1} - \vec{x}^k) + \dots$

2 questions :

- existence : existe-t-il $\vec{x}$ tq $\vec{F}(\vec{x}) = \vec{0}$

  ($N=1$ $ax + e^x = 0$ $x \in \mathbb{R}$, $a > 0$)

  $e^x = -ax$



- unicité : si la solution existe, alors elle est unique

  Supp. $\vec{x}$ et $\vec{y} \in \mathbb{R}^N$ tq $\vec{F}(\vec{x}) = \vec{0}$ et $\vec{F}(\vec{y}) = \vec{0}$

  et montrons $\vec{x} = \vec{y}$. En effet

$$\vec{0} = \vec{F}(\vec{x}) - \vec{F}(\vec{y}) = A(\vec{x} - \vec{y}) + e^{\vec{x}} - e^{\vec{y}}$$

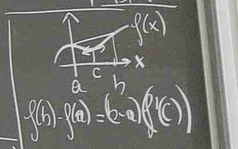
On prend le produit scalaire avec $\vec{x} - \vec{y}$:

$$(\vec{x} - \vec{y})^T A(\vec{x} - \vec{y}) + \underbrace{(\vec{x} - \vec{y})^T (e^{\vec{x}} - e^{\vec{y}})}_{I} = 0$$

$$I = \sum_{i=1}^{N} (x_i - y_i)(e^{x_i} - e^{y_i}) \qquad \boxed{\text{donc } I \geq 0}$$

$$= \sum_{i=1}^{N} (x_i - y_i)(x_i - y_i) e^{z_i} \quad \text{où } \min(x_i, y_i) \leq z_i \leq \max(x_i, y_i)$$

$\vec{a}^T \vec{b} = (a_1 \dots a_N)\begin{pmatrix} b_1 \\ \vdots \\ b_N \end{pmatrix} = \sum_{i=1}^{N} a_i b_i$



$f(b) - f(a) = (b-a) f'(c)$

On a donc

$$(\vec{x} - \vec{y})^T A(\vec{x} - \vec{y}) \leq (\vec{x} - \vec{y})^T A(\vec{x} - \vec{y}) + (\vec{x} - \vec{y})^T (e^{\vec{x}} - e^{\vec{y}}) = 0$$

Mais A est sym. déf. positive donc

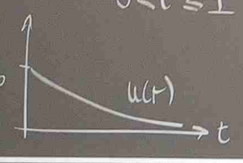
$$(\vec{x} - \vec{y})^T A(\vec{x} - \vec{y}) = 0 \iff (\vec{x} - \vec{y} = \vec{0})$$

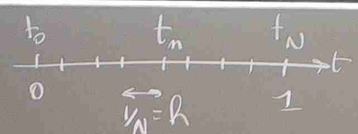
$$\iff (\vec{x} = \vec{y})$$

Chap 9. Eq. différentielles

$$\begin{cases} u'(t) = -\beta u(t) & \beta > 0 \\ u(0) = u_0 & 0 < t \leq 1 \end{cases}$$

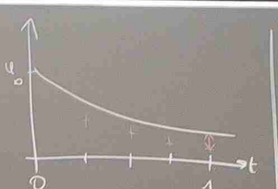
$u(t) = u_0 e^{-\beta t}$





$t_n = nh \quad n = 0, \dots, N$

u^n approx de $u(t_n)$, schéma d'Euler rétrograde (implicite)



$$\frac{u^{n+1} - u^n}{h} = -\beta u^{n+1}$$

$$u^{n+1} = \frac{1}{1+\beta h} u^n = \frac{1}{(1+\beta h)^{n+1}} u^0$$

$$u(t_N) - u^N = u_0 \left(e^{-\beta N h} - \frac{1}{(1+\beta h)^N} \right)$$

On va dém :

Thm : $\exists C > 0 \ \forall u_0 \in \mathbb{R} \ \forall \beta > 0 \ \forall N > 0$ on a

$$\left| u(t_N) - u^N \right| \leq \underbrace{C |u_0| \beta^2 \frac{1}{N}}_{O(h)}$$

Interpr : On choisit u_0, β, l'erreur est /2 si N est mult par 2 ou N est /2

Dem : $a, b \in \mathbb{R}$, $a^N - b^N = (a-b)\left(a^{N-1} + a^{N-2} b + \dots + a b^{N-2} + b^{N-1}\right)$

donc $u(t_N) - u^N = u_0 \left(e^{-\beta h} - \frac{1}{1+\beta h} \right) \left(e^{-\beta h(N-1)} + \frac{e^{-\beta h(N-2)}}{1+\beta h} + \dots + \frac{e^{-\beta h}}{(1+\beta h)^{N-2}} + \frac{1}{(1+\beta h)^{N-1}} \right)$

Lemme : $\forall x > 0 \quad \left| e^{-x} - \frac{1}{1+x} \right| \leq \frac{3}{2} x^2$

Si on admet mom. ce lemme on obtient :

$$\left| u(t_N) - u^N \right| \leq |u_0| \left| e^{-\beta h} - \frac{1}{1+\beta h} \right| \cdot \left(e^{-\beta h(N-1)} + \frac{e^{-\beta h(N-2)}}{1+\beta h} + \dots + \frac{1}{(1+\beta h)^{N-1}} \right)$$

ici on a utilisé l'ineg triang et on a supprimé les val (car $a = a$ si $a \geq 0$)

On utilise le lemme :

$$\left| u(t_N) - u^N \right| \leq |u_0| \frac{3}{2} \beta^2 h^2 N$$

$\left(\frac{e^{-\beta h(N-2)}}{1+\beta h} \leq 1.1, \ N \text{ termes} \right)$

mais $h = 1/N$:

$$\left| u(t_N) - u^N \right| \leq \frac{3}{2} |u_0| \beta^2 h = \frac{3}{2} |u_0| \frac{\beta^2}{N}$$

On a bien dem. le Thm avec $C = 3/2$

Dem lemme : $\forall x > 0$ dev. limité $x = 0$

$$e^{-x} = 1 - x + \frac{x^2}{2} e^{-\alpha} \quad \text{où } 0 < \alpha < x$$

$$\frac{1}{1+x} = 1 - x + \frac{x^2}{2} \frac{2}{(1+\gamma)^3} \quad \text{où } 0 < \gamma < x$$

$\frac{1}{1+x}$, $\frac{-1}{(1+x)^2}$, $\frac{2}{(1+x)^3}$

$$e^{-x} - \frac{1}{1+x} = \frac{x^2}{2} \left(e^{-\alpha} - \frac{2}{(1+\gamma)^3} \right)$$

$$\left| e^{-x} - \frac{1}{1+x} \right| \leq \frac{x^2}{2} \left| e^{-\alpha} - \frac{2}{(1+\gamma)^3} \right| \leq \frac{x^2}{2} \left(\left| e^{-\alpha} \right| + \left| \frac{2}{(1+\gamma)^3} \right| \right)$$

donc $\left| e^{-x} - \frac{1}{1+x} \right| \leq \frac{x^2}{2} (1+2) = \frac{3}{2} x^2$

Ce qui change avec le schéma d'Euler explicite

$$\frac{u^{n+1} - u^n}{h} = -\beta u^n$$

$$u^{n+1} = (1-\beta h) u^n = (1-\beta h)^{n+1} u^0$$

$$u(t_N) - u^N = u_0 \left(e^{-\beta N h} - (1-\beta h)^N \right)$$

$$= u_0 \left(e^{-\beta h} - (1-\beta h) \right) \left(e^{-\beta h(N-1)} + \dots + (1-\beta h)^{N-1} \right)$$

Lemme : $x > 0 \quad \left| e^{-x} - (1-x) \right| \leq \frac{1}{2} x^2$

N termes $\left(|(1-\beta h)^n| \leq 1 \text{ si } |1-\beta h| \leq 1 \right.$ i.e. $\left. -1 \leq 1-\beta h \leq 1 \right)$

Thm mais Si $h \leq \frac{2}{\beta}$ et $C = 1/2$ si $h \leq \frac{2}{\beta}$

Chap 9 : pbm à valeur initiale

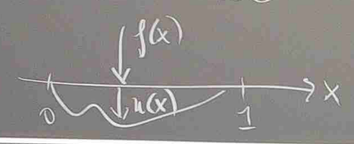
$m X''(t) = \text{Forces}(X(t), X'(t), t)$

$$\begin{cases} m V'(t) = \text{——} \\ X'(t) = V(t) \end{cases}$$

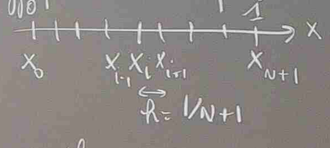
$X(0)$ $V(0)$ connus

Chap 10 : pbm aux limites (1d)

$$\begin{cases} -u''(x) = f(x) & 0<x<1 \\ u(0)=0 \quad u(1)=0 \end{cases}$$



Diff. finies (chap 2)



$x_i = ih \quad i=0,1,\dots,N+1$

u_i approx de $u(x_i)$ $\quad i=1,\dots,N$

$$-u''(x_i) = f(x_i) \quad i=1,\dots,N$$

$$-\frac{u(x_{i-1}) - 2u(x_i) + u(x_{i+1})}{h^2} = f(x_i) + r_i$$

$$|r_i| \leq \frac{1}{12} \max_{0\leq x\leq 1} |u^{(4)}(x)| h^2 \text{ si } u \in \mathcal{C}^4[0,1]$$

Schéma : $-\dfrac{u_{i-1} - 2u_i + u_{i+1}}{h^2} = f(x_i)$ $\quad i=1,\dots,N$ $\quad u_0 = 0$ $\quad u_{N+1} = 0$

$$A = \frac{1}{h^2}\begin{pmatrix} 2 & -1 & & \\ -1 & 2 & -1 & \\ & \ddots & \ddots & \ddots \\ & & -1 & 2 \end{pmatrix} \qquad \vec{f} = \begin{pmatrix} f(x_1) \\ \vdots \\ f(x_N) \end{pmatrix}$$

$$\vec{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_N \end{pmatrix} \qquad \vec{U} = \begin{pmatrix} u(x_1) \\ \vdots \\ u(x_N) \end{pmatrix}$$

Schéma $\quad A\vec{u} = \vec{f}$

Sol. Exacte $\quad A\vec{U} = \vec{f} + \vec{r}$

$$A(\vec{U} - \vec{u}) = \vec{r}$$

Lemme (stabilité) : Soit $\vec{g} \in \mathbb{R}^N$ et soit $\vec{v} \in \mathbb{R}^N$ tel que $A\vec{v} = \vec{g}$

On a : $\|\vec{v}\|_\infty \leq \frac{1}{8} \|\vec{g}\|_\infty$

où $\|\vec{v}\|_\infty = \max_{1\leq i\leq N} |v_i|$

Si on admet momentanément ce lemme, on obtient :

$$\max_{1\leq i\leq N} |u(x_i) - u_i| = \|\vec{U} - \vec{u}\|_\infty \leq \frac{1}{8} \|\vec{r}\|_\infty \leq \underbrace{\frac{1}{8} \frac{1}{12} \max_{1\leq i\leq N} |u^{(4)}(x)|}_{C} h^2$$

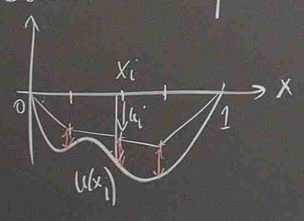
On a dem le thm suivant :

Thm : $\forall u \in \mathcal{C}^4[0,1]$ $\exists C>0$ $\forall 0<h\leq 1$ on a

$$\max_{1\leq i\leq N} |u(x_i) - u_i| \leq C h^2$$

Interpr : On choisit f (et donc $u \in \mathcal{C}^4$)

l'erreur est div par $2^2 = 4$ si h est div par 2



Dem lemme :

le lemme est facile à montrer si on admet

1_ $A\vec{w} = \vec{1}$ $\quad w_i = x_i \frac{(1-x_i)}{2}$ $\quad i=1,\dots,N$

2_ $(A\vec{z} \geqslant \vec{0}) \Rightarrow (\vec{z} \geqslant \vec{0})$

forces $\geqslant 0$ $\quad$ def $\geqslant 0$

En effet, soit $\vec{g} \in \mathbb{R}^N$ et soit $\vec{v}$ tq $A\vec{v} = \vec{g}$

On a : $i=1,\dots,N$, $|g_i| \leq \|\vec{g}\|_\infty$

ie $\quad -\|\vec{g}\|_\infty \leq g_i \leq \|\vec{g}\|_\infty$

$$-\vec{1}\|\vec{g}\|_\infty \leq \vec{g} \leq \vec{1}\|\vec{g}\|_\infty$$

$$\left(A\vec{w} = \vec{1}\right) \times \|\vec{g}\|_\infty$$

$$A\left(\vec{w}\|\vec{g}\|_\infty\right) = \vec{1}\|\vec{g}\|_\infty$$

$$A\vec{v} = \vec{g}$$

$$A\left(\vec{w}\|\vec{g}\|_\infty + \vec{v}\right) = \vec{1}\|\vec{g}\|_\infty + \vec{g} \geqslant \vec{0}$$

$$A\left(\quad - \quad\right) = \quad - \quad \geqslant \vec{0}$$

On utilise 2_ $\quad \vec{w}\|g\|_\infty + \vec{v} \geqslant \vec{0}$

$\vec{w}\|\vec{g}\|_\infty - \vec{v} \geqslant \vec{0}$

ie $\quad -\vec{w}\|g\|_\infty \leqslant \vec{v} \leqslant \vec{w}\|\vec{g}\|_\infty$

$-w_i\|\vec{g}\|_\infty \leqslant v_i \leqslant w_i\|\vec{g}\|_\infty \quad i=1,\dots,N$

$|v_i| \leqslant |w_i|\,\|\vec{g}\|_\infty = \frac{x_i(1-x_i)}{2}\|\vec{g}\|_\infty$

et donc $|v_i| \leq \frac{1}{2}\frac{1}{4}\|\vec{g}\|_\infty \quad i=1,\dots,N$

ie $\quad \|\vec{v}\|_\infty \leq \frac{1}{8}\|\vec{g}\|_\infty$

Dem de 1_ Vérification

$$(A\vec{w})_1 = 2w_1 - w_2 = 2\frac{x_1(1-x_1)}{2} - \frac{x_2(1-x_2)}{2}$$

$x_1 = h \quad x_2 = 2h$

intuition : Soit $w(x)$ tq $\begin{cases} -w''(x) = 1 & 0<x<1 \\ w(0)=0 \quad w(1)=0 \end{cases}$

$$w(x) = \frac{x(1-x)}{2}$$

$$-\frac{w(x_{i-1}) - 2w(x_i) + w(x_{i+1})}{h^2} = 1 + \text{reste}$$

le reste est nul car $w^{(4)}(x) = 0$

$$A\begin{pmatrix} w(x_1) \\ \vdots \\ w(x_N) \end{pmatrix} = \vec{1}$$

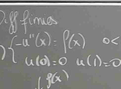
Chap 10 suite

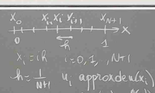
Lundi : 10.0 et 10.2 et 10.6 différences finies

Aujourd'hui : 10.3 et 10.4 éléments finis

Diff. finies

(1) $\begin{cases} -u''(x) = f(x) & 0<x<1 \\ u(0)=0 \quad u(1)=0 \end{cases}$





$\frac{1}{h^2}\begin{pmatrix} 2 & -1 & (0) \\ -1 & \ddots & \ddots \\ (0) & \ddots & -1 \\ & & -1 \quad 2 \end{pmatrix}\begin{pmatrix} u_1 \\ \vdots \\ u_N \end{pmatrix} = \begin{pmatrix} f(x_1) \\ \vdots \\ f(x_N) \end{pmatrix}$

$A \qquad u = f$

Autres :

1. Formulation variationnelle
2. Méthode de Galerkin
3. —— des éléments finis

pour résoudre à $Au = f$ plus générale

1. Soit $v : [0,1] \to \mathbb{R}$ quelconque

on multiplie (1) par $v(x)$ et on $\int_0^1 dx$:

$\int_0^1 -u''(x)v(x) = \int_0^1 f(x)v(x)\,dx$

On intègre par parties

$(uv)' = u'v + uv'$

$[uv]_0^1 = \int u'v + \int uv'$

$u \to -u'$

$[-u'v]_0^1 = \int -u''v + \int -u'v'$

$\int_0^1 u'(x)v'(x)\,dx - [u'(x)v(x)]_{x=0}^{x=1} = \int_0^1 f(x)v(x)\,dx$

$u'(1)v(1) - u'(0)v(0)$

On choisit v telle que $v(1) = v(0) = 0$ et on obtient la formulation variationnelle

$u \in V$ telle que $\int_0^1 u'(x)v'(x)\,dx = \int_0^1 f(x)v(x)\,dx \quad \forall v \in V$ (2)

(forme bilinéaire, symétrique, positive (travaux virtuels))

$a(\cdot,\cdot)$: déformation virtuelle

$\int_0^1 f(x)v(x)\,dx$ travail virtuel

V ? espace des fonctions admissibles

$V = \{v : [0,1] \to \mathbb{R},\ v \text{ continue, } v' \text{ continue par morceaux, } v(0) = 0, v(1) = 0\}$

Remarques :

- Si u est sol de (1) alors u est solution de (2)
- Il existe des sol de (2) qui ne sont pas sol de (1)

Ex :

- Si on choisit $v = u$ dans (2) on obtient :

$\int_0^1 (u'(x))^2\,dx = \int_0^1 f(x)u(x)\,dx$

énergie de déformation — travail des forces

On (2), on en déduit que la solution, si elle existe, est unique. En effet, soit u_1 et u_2 2 sol de (2), on a :

$\int_0^1 (u_1'(x) - u_2'(x))v'(x)\,dx = 0 \quad \forall v \in V$

On choisit $v = u_1 - u_2 \in V$ et on obtient

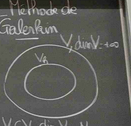
$\int_0^1 (u_1'(x) - u_2'(x))^2\,dx = 0$

et donc $u_1'(x) - u_2'(x) = 0 \quad \forall x \in]0,1[$

$u_1(x) - u_2(x) =$ constante

mais $u_1(0) - u_2(0) = 0$ donc la constante est nulle, $u_1 = u_2$

2 Méthode de Galerkin



Soit $\varphi_1, \varphi_2, \dots, \varphi_N$ N fonctions $\in V$ linéairement indép.

Soit $V_N = \text{span}(\varphi_1, \varphi_2, \dots, \varphi_N)$

$\dim V_N = N$. On approche la solution de (2) en cherchant

$u_N \in V_N$ tel que $\int_0^1 u_N'(x)v_N'(x)\,dx = \int_0^1 f(x)v_N(x)\,dx \quad \forall v_N \in V_N$ (3)

On va montrer que trouver la solution de (3) (u_N) est équivalent à rechercher un système linéaire à N éq et N inconnues. En effet

$u_N \in V_N$: $u_N(x) = \sum_{j=1}^N u_j \varphi_j(x)$ N inconnues

on choisit $v_N(x) = \varphi_i(x)$ $i = 1, \dots, N$ N équations

$u_N(x) = \sum_{j=1}^N u_j \varphi_j'(x)$, (3) devient :

$\sum_{j=1}^N u_j \int_0^1 \varphi_j'(x)\varphi_i'(x)\,dx = \int_0^1 f(x)\varphi_i(x)\,dx \quad i = 1, \dots, N$

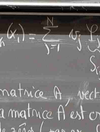
$\sum_{j=1}^N u_j \int_0^1 \varphi_j'(x)\varphi_i'(x)\,dx = \int_0^1 f(x)\varphi_i(x)\,dx$

$\underbrace{A_{ij}}_{A} \underbrace{u_j}_{u} = \underbrace{f_i}_{f} \quad i = 1, \dots, N$

Reste à choisir $\varphi_1, \varphi_2, \dots, \varphi_N$:

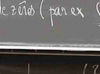
3 possibilités

- $\varphi_i(x) = \sin(i\pi x)$ $i = 1, \dots, N$ → séries de Fourier
- $\varphi_i(x) = x^i(1-x)$ ——
- $\varphi_i(x)$ éléments finis continus, linéaires par morceaux



$x_i = ih \quad i = 0, 1, \dots, N+1$

$v_N(x) = \sum_{j=1}^N v_j \varphi_j(x) \in V_N$



$v_N(x_i) = \sum_{j=1}^N v_j \varphi_j(x_i) = v_i \quad i = 1, \dots, N$

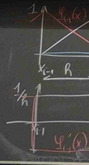
$\varphi_j(x_i) = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{sinon} \end{cases}$

matrice A, vecteur f ?

la matrice A est creuse (beaucoup de zéros) (par ex $\int_0^1 \varphi_i(x)\varphi_j(x)\,dx = 0$)

La matrice A est tridiagonale symétrique ($A_{ij} = A_{ji}$)

Calculons A_{ii} et $A_{i,i-1}$ et f_i $\quad i = 1, \dots, N$



$A_{ii} = \int_0^1 (\varphi_i'(x))^2\,dx = \int_{x_{i-1}}^{x_i} (\varphi_i'(x))^2\,dx + \int_{x_i}^{x_{i+1}} (\varphi_i'(x))^2\,dx$

$= h\left(-\frac{1}{h}\right)^2 + h\left(\frac{1}{h}\right)^2 = \frac{2}{h}$

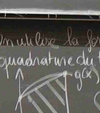
$A_{i,i-1} = \int_0^1 \varphi_{i-1}'(x)\varphi_i'(x)\,dx = \int_{x_{i-1}}^{x_i} \varphi_{i-1}'(x)\varphi_i'(x)\,dx = h\left(\frac{1}{h}\right)\left(\frac{1}{h}\right) = -\frac{1}{h}$

$A = \frac{1}{h}\begin{pmatrix} 2 & -1 & (0) \\ -1 & \ddots & \ddots \\ (0) & \ddots & -1 \\ & & -1 \quad 2 \end{pmatrix}$

$f_i = \int_0^1 f(x)\varphi_i(x)\,dx$

$= \int_{x_{i-1}}^{x_i} f(x)\varphi_i(x)\,dx + \int_{x_i}^{x_{i+1}} f(x)\varphi_i(x)\,dx$

On utilise la formule de quadrature du trapèze



$\int_{x_{i-1}}^{x_i} g(x)\,dx \simeq \frac{h}{2}\left(g(x_{i-1}) + g(x_i)\right)$

Reste à choisir $\varphi_1, \varphi_2, \dots, \varphi_N$:

3 possibilités

- $\varphi_i(x) = \sin(i\pi x)$ $i = 1, \dots, N$ → séries de Fourier
- $\varphi_i(x) = x^i(1-x)$ ——
- $\varphi_i(x)$ éléments finis continus, linéaires par morceaux

$x_i = ih \quad i = 0, 1, \dots, N+1$

$v_N(x) = \sum_{j=1}^N v_j \varphi_j(x) \in V_N$

$v_N(x_i) = \sum_{j=1}^N v_j \varphi_j(x_i) = v_i \quad i = 1, \dots, N$

$\varphi_j(x_i) = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{sinon} \end{cases}$

matrice A, vecteur f ?

la matrice A est creuse (beaucoup de zéros) (par ex $\int_0^1 \varphi_i(x)\varphi_j(x)\,dx = 0$)

La matrice A est tridiagonale symétrique ($A_{ij} = A_{ji}$)

Calculons A_{ii} et $A_{i,i-1}$ et f_i $\quad i = 1, \dots, N$

$A_{ii} = \int_0^1 (\varphi_i'(x))^2\,dx = \int_{x_{i-1}}^{x_i} (\varphi_i'(x))^2\,dx + \int_{x_i}^{x_{i+1}} (\varphi_i'(x))^2\,dx$

$= h\left(-\frac{1}{h}\right)^2 + h\left(\frac{1}{h}\right)^2 = \frac{2}{h}$

$A_{i,i-1} = \int_0^1 \varphi_{i-1}'(x)\varphi_i'(x)\,dx = \int_{x_{i-1}}^{x_i} \varphi_{i-1}'(x)\varphi_i'(x)\,dx = h\left(\frac{1}{h}\right)\left(\frac{1}{h}\right) = -\frac{1}{h}$

$A = \frac{1}{h}\begin{pmatrix} 2 & -1 & (0) \\ -1 & \ddots & \ddots \\ (0) & \ddots & -1 \\ & & -1 \quad 2 \end{pmatrix}$

$f_i = \int_0^1 f(x)\varphi_i(x)\,dx$

$= \int_{x_{i-1}}^{x_i} f(x)\varphi_i(x)\,dx + \int_{x_i}^{x_{i+1}} f(x)\varphi_i(x)\,dx$

On utilise la formule de quadrature du trapèze

$\int_{x_{i-1}}^{x_i} g(x)\,dx \simeq \frac{h}{2}\left(g(x_{i-1}) + g(x_i)\right)$

$f_i \simeq \frac{h}{2}\left(\underbrace{g(x_{i-1})}_{0}\underbrace{\varphi_i(x_{i-1})}_{0} + g(x_i)\underbrace{\varphi_i(x_i)}_{1} + g(x_i)\underbrace{\varphi_i(x_i)}_{1} + g(x_{i+1})\underbrace{\varphi_i(x_{i+1})}_{0}\right)$

$= h\,g(x_i)$

Système linéaire :

$\left(u_N(x) = \sum_{j=1}^N u_j \varphi_j(x),\ u_N(x_i) = u_i\right)$

$\frac{1}{h}\begin{pmatrix} 2 & -1 & (0) \\ -1 & \ddots & \ddots \\ (0) & \ddots & -1 \\ & & -1 \quad 2 \end{pmatrix}\begin{pmatrix} u_1 \\ \vdots \\ u_N \end{pmatrix} = h\begin{pmatrix} g(x_1) \\ \vdots \\ g(x_N) \end{pmatrix}$

On obtient le même système linéaire que celui de la méthode des diff. finies

2D

1D

2D

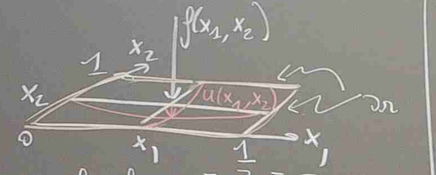
3D

Chap II (remarque II.2)

Pbm aux limites 2d

Membrane élastique (tambour)

Etant donné $f : [0,1]\times[0,1] \to \mathbb{R}$, $(x_1, x_2) \to f(x_1, x_2)$



on cherche $u : [0,1]\times[0,1] \to \mathbb{R}$, $(x_1, x_2) \to u(x_1, x_2)$

telle que

$$\begin{cases} -\Delta u(x_1,x_2) = f(x_1,x_2) & (x_1,x_2) \in \overbrace{[0,1]\times[0,1]}^{\Omega} \\ u(x_1,x_2) = 0 & (x_1,x_2) \in \partial\Omega \end{cases}$$

$$\Delta u(x_1,x_2) = \operatorname{div}(\overrightarrow{\operatorname{grad}}\, u(x_1,x_2)) = \vec{\nabla}\cdot(\vec{\nabla} u), \quad \nabla = \begin{pmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \end{pmatrix}$$

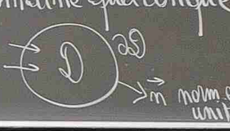
$$= \frac{\partial^2 u}{\partial x_1^2}(x_1,x_2) + \frac{\partial^2 u}{\partial x_2^2}(x_1,x_2)$$

Modélisation (électromagn.)

$\operatorname{div} \vec{E} = \rho$ densité charge

Chp élec.

$\mathcal{D}$ domaine quelconque



$$\int_{\mathcal{D}} \operatorname{div}\vec{E}\, dV = \int_{\mathcal{D}} \rho\, dV$$

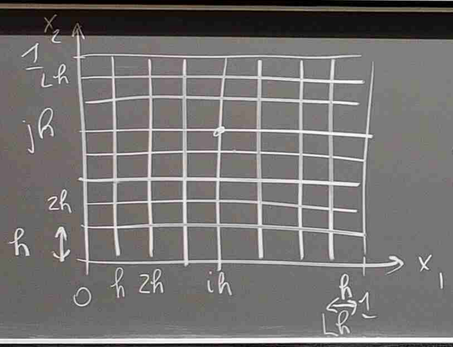
$$\int_{\partial\mathcal{D}} \vec{E}\cdot\vec{n}\, dS = \int_{\mathcal{D}} \rho\, dV$$

$\vec{\operatorname{rot}}\, \vec{E} = \vec{0} \to \vec{E} = \vec{\operatorname{grad}}\, \varphi$ (φ potentiel)

$-\operatorname{div}(\vec{\operatorname{grad}}\, \varphi) = \rho$

$-\Delta\varphi = \rho$

Résol. numérique : grille différences finies



L entier pas (grand)

$h = \frac{1}{L+1}$ pas d'espace $\xrightarrow{L\to\infty} 0$

But : calculer $U_{i,j}$ approx de $u(ih, jh)$

On écrit l'eq aux pts ih, jh :

$$-\Delta u(ih,jh) = f(ih,jh) \quad i,j = 1,\dots,L$$

$$\frac{\partial^2 u}{\partial x_1^2}(ih,jh) = \frac{u((i-1)h,jh) - 2u(ih,jh) + u((i+1)h,jh)}{h^2} + O(h^2)$$

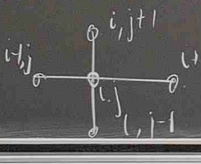
$$\frac{\partial^2 u}{\partial x_2^2}(ih,jh) = \frac{u(ih,(j-1)h) - 2u(ih,jh) + u(ih,(j+1)h)}{h^2} + O(h^2)$$

Pour obtenir le schéma, on remplace $u(ih,jh)$ par $U_{i,j}$ et on oublie les $O(h^2)$

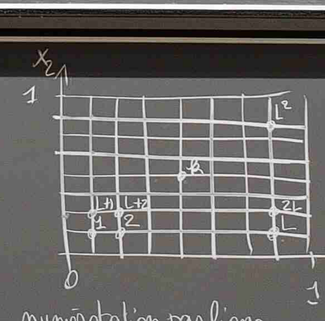
Schéma : $$\frac{-U_{i-1,j} - U_{i,j-1} + 4U_{i,j} - U_{i+1,j} - U_{i,j+1}}{h^2} = f(ih,jh) \quad i,j = 1,\dots,L$$

$U_{i,j} = 0$ si $i = 0$ ou $i = L+1$ ou $j = 0$ ou $j = L+1$

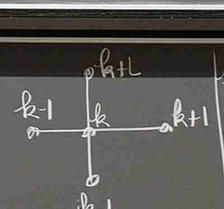
Schéma à 5 points



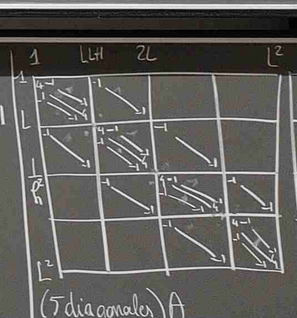
Ce schéma corresp à la résol d'un système linéaire



numérotation par ligne



$U_{i,j} \to u_k$



($\bar{}$ diagonales) A

$A \in \mathbb{R}^{L^2\times L^2}$

$\vec{u} \in \mathbb{R}^{L^2}$ $\vec{f} \in \mathbb{R}^{L^2}$

On peut montrer que si $u \in \mathcal{C}^4([0,1]^2)$ alors

$$\max_{1\le i,j\le L} |u(ih,jh) - U_{i,j}| = O(h^2)$$

Implémentation avec matlab (octave)

la matrice A peut s'écrire $A = \frac{1}{h^2}\left(D_L \otimes I_L + I_L \otimes D_L\right)$

où $D_L \in \mathbb{R}^{L\times L}$ $D_L = \begin{pmatrix} 2 & -1 & & (0) \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ (0) & & -1 & 2 \end{pmatrix}$ et $I_L \in \mathbb{R}^{L\times L}$ $I_L = \begin{pmatrix} 1 & & (0) \\ & \ddots & \\ (0) & & 1 \end{pmatrix}$

et $\otimes$ kronecker product "kron"

$B, C \in \mathbb{R}^{L\times L}$

$$B\otimes C = \begin{pmatrix} b_{11}C & \cdots & b_{1L}C \\ \vdots & \ddots & \vdots \\ b_{L1}C & \cdots & b_{LL}C \end{pmatrix} \begin{matrix} \} L \\ \\ \} L^2 \end{matrix}$$

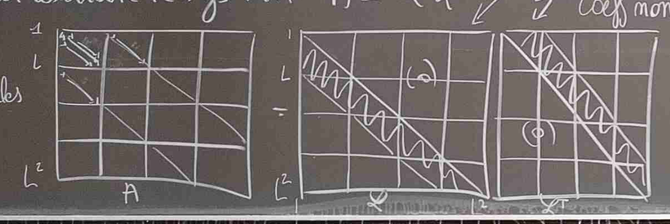
$$A = \frac{1}{h^2}\left(\begin{array}{cccc} 2I_L & -I_L & & \\ -I_L & 2I_L & -I_L & \\ & -I_L & 2I_L & -I_L \\ & & -I_L & 2I_L \end{array} + \begin{array}{cccc} D_L & & & \\ & D_L & & \\ & & D_L & \\ & & & D_L \end{array} \right)$$

tridiag par bloc

Mémoire ? Nbre d'opérations ? (complexité)

Pour résoudre le syst lin $A = \mathcal{L}\mathcal{L}^T$ ← partie hachurée ∑ coeff non nuls de $\mathcal{L}$ et $\mathcal{L}^T$

A. 5 diagonales

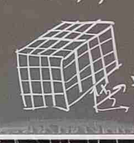


		$L=100$	$L=1000$
nombre d'inconnues	L^2	10^4	10^6
mémoire (RAM)	L^3	10^6	10^9
nbre d'opérations (CPU)	L^4	10^8	10^{12} → 1000 s !

Algorithme décomp. Cholesky $\left.\begin{matrix} i=1,L\times L \\ j=1,L \\ k=1,L \end{matrix}\right\}$ $L^2 L L = L^4$

Calcul à 3 dimensions

$$\begin{cases} -\Delta u(x_1,x_2,x_3) = f(x_1,x_2,x_3) & (x_1,x_2,x_3) \in \Omega = [0,1]^3 \\ u(x_1,x_2,x_3) = 0 & (x_1,x_2,x_3) \in \partial\Omega \end{cases}$$



L points dans chaque direction

L^3 inconnues

Schéma : $$\frac{-U_{i-1,j,k} - U_{i,j-1,k} - U_{i,j,k-1} + 6U_{i,j,k} - U_{i+1,j,k} - U_{i,j+1,k} - U_{i,j,k+1}}{h^2} = f(ih,jh,kh)$$

$i,j,k = 1,\dots,L$

$U_{i,j,k} = 0$ si $i = 0$ ou $i = L+1$

j — j —

k — k —

		$L=100$
nombre d'inconnues	L^3	
mem	$L^3 L^2 = L^5$	$L = 10^{10}$
nbre d'op (CPU)	$L^3 L^2 L^2 = L^7$	$L = 10^{14}$

$i = 1, L\times L\times L$

$j = 1, L\times L$

$k = 1, L\times L$

Conclusion : on ne peut plus faire la déc. de Cholesky

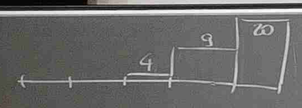
il faut utiliser des méth. itératives (ex la méthode du gradient)

Question : nombre d'itérations ?

30/150
20%

Chap 12 Eq de la chaleur

Exp : matériau ρc_p, $k>0$; $0°C$ en $x=0$, $0°C$ en $x=1$; à $t=0$ temp $w(x)$; $f(x,t)$

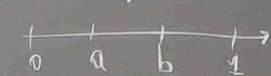


On cherche $u(x,t)$ telle que

$$\begin{cases} \rho c_p \frac{\partial u}{\partial t}(x,t) - k \frac{\partial^2 u}{\partial x^2}(x,t) = f(x,t) & 0<x<1 \quad t>0 \\ u(0,t)=0 \quad u(1,t)=0 & \text{cond limites} \quad t>0 \\ u(x,0)=w(x) & \text{—— initiale} \quad 0<x<1 \end{cases}$$

Remarques :

- L'eq de la chaleur traduit un principe thermodyn.



On intègre entre $x=a$ et $x=b$:

$$\int_a^b \rho c_p \frac{\partial u}{\partial t}(x,t)\,dx - \int_a^b k \frac{\partial^2 u}{\partial x^2}(x,t)\,dx = \int_a^b f(x,t)\,dx$$

$$\frac{d}{dt}\int_a^b \rho c_p u(x,t)\,dx - \left[k \frac{\partial u}{\partial x}(x,t)\right]_{x=a}^{x=b} = \int_a^b f(x,t)\,dx$$

- $\frac{d}{dt}$: la variation au cours du temps de
- $\int_a^b \rho c_p u(x,t)\,dx$: Energie thermique
- $\left[k \frac{\partial u}{\partial x}(x,t)\right]_{x=a}^{x=b}$: flux de Chaleur entrant ou sortant
- $\int_a^b f(x,t)\,dx$: puissance thermique source

- Dans la suite $\rho c_p = 1$.

Si $f(x,t)=0$ $w(x)=\sin(\pi x)$ alors $u(x,t) = \sin(\pi x)\, e^{-k\pi^2 t}$

- Si $f(x,t)=0$, alors $u(x,t)$ satisfait

* positivité : si $w(x)\geq 0$ $0\leq x\leq 1$ alors $u(x,t)\geq 0$ ——

* principe du max :

$$\max_{0\leq x\leq 1} |u(x,t)| \leq \max_{0\leq x\leq 1} |w(x)|$$

Méthode numérique

Grille espace temps

$h = \frac{1}{N+1}$ $x_i = ih$ $i=0,1,\dots,N+1$

$t_n = n\tau$ $n=0,1,2,\dots$

h : pas d'espace, τ pas de temps

But : calculer u_i^n approx de $u(x_i,t_n)$

$t=0$: $u(x_i,0) = w(x_i) = u_i^0$

n fixé, étant donné les u_i^n, $i=1,\dots,N$ comment calculer u_i^{n+1}, $i=1,\dots,N$?

Marche en temps

Dans le linéaire :

- on discrétise en espace et on obtient un syst diff

$$\vec{u}'(t) + A\vec{u}(t) = \vec{f}(t) \quad t>0$$

- on discrétise en temps, sch. Euler progr

$$\frac{\vec{u}^{n+1} - \vec{u}^n}{\tau} + A\vec{u}^n = \vec{f}(t_n)$$

Ici on discret en temps et en espace ;

On écrit l'eq de la chaleur en x_i, t_n

$$\underbrace{\frac{\partial u}{\partial t}(x_i,t_n)}_{\text{F.D.F. progressive}} - k \underbrace{\frac{\partial^2 u}{\partial x^2}(x_i,t_n)}_{\text{FDF centrée}} = f(x_i,t_n) \quad i=1,\dots,N$$

$$\frac{u(x_i,t_{n+1}) - u(x_i,t_n)}{\tau} - k\,\frac{u(x_{i-1},t_n) - 2u(x_i,t_n) + u(x_{i+1},t_n)}{h^2} = f(x_i,t_n) + O(\tau) + O(h^2)$$

Schéma d'Euler progressif

on remplace $u(x_i,t_n)$ par u_i^n et on oublie les $O(\tau)$ et $O(h^2)$

$$\begin{cases} \frac{u_i^{n+1} - u_i^n}{\tau} - k\,\frac{u_{i-1}^n - 2u_i^n + u_{i+1}^n}{h^2} = f(x_i,t_n) & i=1,\dots,N \\ u_0^n = 0 \quad u_{N+1}^n = 0 \end{cases}$$

Schéma explicite :

$$u_i^{n+1} = u_i^n\left(1 - \frac{2\tau k}{h^2}\right) + \frac{\tau k}{h^2}\left(u_{i-1}^n + u_{i+1}^n\right) + \tau f(x_i,t_n) \quad i=1,\dots,N$$

- avantage. Facile à programmer
- inconvénient : cond. de stabilité

On peut écrire ce schéma sous la forme

$$\frac{\vec{u}^{n+1} - \vec{u}^n}{\tau} + A\vec{u}^n = \vec{f}(t_n)$$

ou $\vec{u}^{n+1} = (I - \tau A)\vec{u}^n + \tau \vec{f}(t_n)$

où on a noté

$$\vec{u}^n = \begin{pmatrix} u_1^n \\ \vdots \\ u_N^n \end{pmatrix} \quad A = \frac{k}{h^2}\begin{pmatrix} 2 & -1 & & (0) \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ (0) & & -1 & 2 \end{pmatrix}$$

$$\vec{f}(t_n) = \begin{pmatrix} f(x_1,t_n) \\ \vdots \\ f(x_N,t_n) \end{pmatrix}$$

Schéma d'Euler rétrograde

on écrit l'eq au point x_i temps t_{n+1}

$$\underbrace{\frac{\partial u}{\partial t}(x_i,t_{n+1})}_{\text{FDF rétrograde}} - k\,\underbrace{\frac{\partial^2 u}{\partial x^2}(x_i,t_{n+1})}_{\text{centrée}} = f(x_i,t_{n+1})$$

$$\frac{u(x_i,t_{n+1}) - u(x_i,t_n)}{\tau} - k\,\frac{u(x_{i-1},t_{n+1}) - 2u(x_i,t_{n+1}) + u(x_{i+1},t_{n+1})}{h^2} = f(x_i,t_{n+1})$$

Schéma :

$$\begin{cases} \frac{u_i^{n+1} - u_i^n}{\tau} - k\,\frac{u_{i-1}^{n+1} - 2u_i^{n+1} + u_{i+1}^{n+1}}{h^2} = f(x_i,t_{n+1}) & i=1,\dots,N \\ u_0^{n+1} = 0 \quad u_{N+1}^{n+1} = 0 \end{cases}$$

$\frac{\vec{u}^{n+1} - \vec{u}^n}{\tau} + A\vec{u}^{n+1} = \vec{f}(t_{n+1})$ i.e $(I+\tau A)\vec{u}^{n+1} = \vec{u}^n + \tau \vec{f}(t_{n+1})$

Inconv : schéma impliciter résoudre un syst. lin. plus diff à programmer

Avantage : pas de cond. de stabilité

Stabilité : ($f(x,t)=0$)

Def : on dira que le schéma est stable si les 2 ppties suivantes sont satisfaites

- Si $u_i^n \geq 0$ $i=1,\dots,N$ alors $u_i^{n+1} \geq 0$ ——
- $\max_{1\leq i\leq N} |u_i^{n+1}| \leq \max_{1\leq i\leq N} |u_i^n|$

Lemme :

- le sch. d'Euler progressif (expl.) est stable si $\tau \leq \frac{h^2}{2k}$
- le sch. d'Euler rétrograde (impl.) est stable $\forall \tau>0 \; \forall h>0$

Remarques :

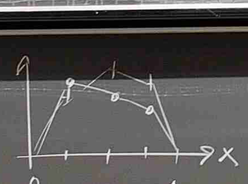
- la cond $\tau \leq \frac{h^2}{2k}$ est pénalisante ex : $k=1$ $N=9$ $h=0,1$
  $\tau \leq 0,005$ $t=1s$: 200 pas de temps
  $N=19$ $h=0,05$ $\tau \leq 0,00125$ $t=1s$: 800 pas de temps
- les 2 schémas convergent à l'ordre $O(\tau)+O(h^2)$ Si $\tau/4$ et $h/2$ l'erreur au temps final est divisée par 4.

Dem lemme Euler progressif (expl.)

Supp $\tau \leq \frac{h^2}{2k}$ et montrons

1 - Si $u_i^n \geq 0$ $i=1,\dots,N$ alors $u_i^{n+1} \geq 0$ $i=1,\dots,N$

2 - $\max_{1\leq i\leq N} |u_i^{n+1}| \leq \max_{1\leq i\leq N} |u_i^n|$



1 - $u_i^{n+1} = \underbrace{\left(1 - \frac{2\tau k}{h^2}\right)}_{\geq 0} u_i^n + \underbrace{\frac{\tau k}{h^2}}_{\geq 0}\left(\underbrace{u_{i-1}^n}_{\geq 0} + \underbrace{u_{i+1}^n}_{\geq 0}\right)$ $i=1,\dots,N$

Donc $u_i^{n+1} \geq 0$ $i=1,\dots,N$

2 - $|u_i^{n+1}| = \left|\left(1 - \frac{2\tau k}{h^2}\right)u_i^n + \frac{\tau k}{h^2}\left(u_{i-1}^n + u_{i+1}^n\right)\right|$

$\leq \left|1 - \frac{2\tau k}{h^2}\right| \cdot |u_i^n| + \left|\frac{\tau k}{h^2}\right|\left(|u_{i-1}^n| + |u_{i+1}^n|\right)$

$= \left(1 - \frac{2\tau k}{h^2}\right)|u_i^n| + \frac{\tau k}{h^2}\left(|u_{i-1}^n| + |u_{i+1}^n|\right)$

$\leq \left(1 - \frac{2\tau k}{h^2} + \frac{\tau k}{h^2} + \frac{\tau k}{h^2}\right)\max_{1\leq j\leq N}|u_j^n| = \max_{1\leq j\leq N}|u_j^n|$

(si $i=1$ $u_0^n = 0$; si $i=N$ $u_{N+1}^n = 0$)

$\leq \max_{1\leq j\leq N}|u_j^n|$

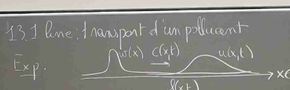
Chap12 $\frac{\partial u}{\partial t} - k\frac{\partial^2 u}{\partial x^2} = f$

Chap13 transport $\frac{\partial u}{\partial t} + c\frac{\partial u}{\partial x} = f$

ondes $\frac{\partial^2 u}{\partial t^2} - c^2\frac{\partial^2 u}{\partial x^2} = f$

13.1 ligne : transport d'un polluant

Exp :



cas initiale de polluant $u_0(x)$
vitesse de transp. du poll. $c(x,t)$
source ou puits de polluant

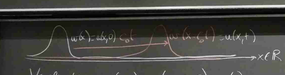
inconnue : $u(x,t)$ conc de polluant x, t

$$\begin{cases} \frac{\partial u}{\partial t}(x,t) + c(x,t)\frac{\partial u}{\partial x}(x,t) = f(x,t) & x \in \mathbb{R},\ t>0 \\ u(x,0) = u_0(x) & x \in \mathbb{R} \end{cases}$$

pas de cond. aux x limites pour l'instant $x \in \mathbb{R}$

Remarques :

- Si $f(x,t)=0$, si $c(x,t)=c_0=$cste alors $u(x,t) = u_0(x-c_0 t)$



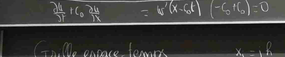
Vérification : $u(x,0) = u_0(x-c_0\,0) = u_0(x)$

$\frac{\partial u}{\partial t}(x,t) = u_0'(x-c_0 t)(-c_0)$ (règle $(u\circ v)'$)

$c_0\left(\frac{\partial u}{\partial x}(x,t) = u_0'(x-c_0 t)\cdot 1\right)$

$\frac{\partial u}{\partial t} + c_0\frac{\partial u}{\partial x} = u_0'(x-c_0 t)(-c_0+c_0) = 0$

- Si $c(x,t) = c_0 =$ cste l'eq de transport satisfait un principe de conservation



on $\int_a^b dx$

$\int_a^b \frac{\partial u}{\partial t}(x,t)dx + \int_a^b c_0\frac{\partial u}{\partial x}(x,t)dx = \int_a^b f(x,t)dx$

$\frac{d}{dt}\int_a^b u(x,t)dx + \left[c_0\,u(x,t)\right]_{x=a}^{x=b} = \int_a^b f(x,t)dx$

Variation au cours du temps | quantité de polluant dans $[a,b]$ | flux entrant ou sortant du polluant | source ou puits de poll.

- Si $f(x,t)=0$ $c(x,t) = c_0 =$ cste on a $x \mapsto u(x,t) \geq 0$ $x \in \mathbb{R}$ alors $u(x,t) \geq 0$ $x \in \mathbb{R}$
- $\sup_{x\in\mathbb{R}} |u(x,t)| \leq \sup_{x\in\mathbb{R}} |u_0(x)|$

Grille espace-temps

$x_j = jh$ $j \in \mathbb{Z}$ h pas d'espace

$t_n = n\tau$ $n=0,1,2,\dots$ τ pas de temps

But : calculer u_j^n une approx de $u(x_j, t_n)$

marche en temps : $u_j^0 = u_0(jh)$ $j \in \mathbb{Z}$

comment calculer u_j^{n+1} $j \in \mathbb{Z}$?

Il faut donner u_j^n $j \in \mathbb{Z}$ comment calculer u_j^{n+1} $j \in \mathbb{Z}$?

$\frac{\partial u}{\partial t}(x_j,t_n) + c(x_j,t_n)\frac{\partial u}{\partial x}(x_j,t_n) = f(x_j,t_n)$

FDF progressive | FDF centrée | $j \in \mathbb{Z}$

$\frac{u(x_j,t_{n+1}) - u(x_j,t_n)}{\tau} + c(x_j,t_n)\frac{u(x_{j+1},t_n) - u(x_{j-1},t_n)}{2h} = f(x_j,t_n) + O(\tau) + O(h^2)$

2 schémas : centré, décentré

1. Schéma explicit centré

on écrit l'eq au point x_j temps t_n

le schéma explicite centré est

$\frac{u_j^{n+1} - u_j^n}{\tau} + c(x_j,t_n)\frac{u_{j+1}^n - u_{j-1}^n}{2h} =: f(x_j,t_n)$, $j \in \mathbb{Z}$

On peut montrer (cf plus) que ce schéma est instable $\forall h>0$ $\forall \tau>0$

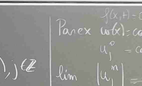
Par ex $u_0(x) =$ cos(mx) $m \in \mathbb{N}$

$u_j^0 = \cos(mjh)$

$\lim_{n\to\infty} |u_j^n| = +\infty$!

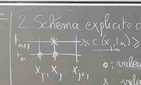
et $|\alpha(x,\tau)| = |\cos(m\alpha - c_0\tau)| \leq 1$

2. Schéma explicite décentré



si $c(x_j,t_n) \geq 0$

$\frac{u_j^{n+1} - u_j^n}{\tau} + c(x_j,t_n)\frac{u_j^n - u_{j-1}^n}{h} =: f(x_j,t_n)$ $j \in \mathbb{Z}$



si $c(x_j,t_n) \leq 0$

$\frac{u_j^{n+1} - u_j^n}{\tau} + c(x_j,t_n)\frac{u_{j+1}^n - u_j^n}{h} = f(x_j,t_n)$ $j \in \mathbb{Z}$

Ce schéma est explicite $(u_j^{n+1} = \dots$ $j \in \mathbb{Z})$

stable sous condition et converge $O(\tau) + O(h)$

Stabilité du schéma décentré

Déf : Cas $f(x,t)=0$ $c(x,t) = c_0 =$ cste $(u(x,t) = u_0(x-c_0 t))$

On dira que le schéma est stable si les deux propriétés suivantes sont satisfaites

- si $(u_j^n \geq 0$ $j \in \mathbb{Z})$ alors $(u_j^{n+1} \geq 0$ $j \in \mathbb{Z})$
- $\sup_{j\in\mathbb{Z}} |u_j^{n+1}| \leq \sup_{j\in\mathbb{Z}} |u_j^n|$

Lemme : Cas $f(x,t)=0$, $c(x,t) = c_0 =$ cste $\neq 0$

Si $\tau \leq \frac{h}{|c_0|}$ alors le schéma explicite décentré est stable

Rem cas $c_0 > 0$

le schéma s'écrit

$\frac{u_j^{n+1} - u_j^n}{\tau} + c_0\frac{u_j^n - u_{j-1}^n}{h} = 0$

ie $u_j^{n+1} = \left(1 - c_0\frac{\tau}{h}\right)u_j^n + c_0\frac{\tau}{h}u_{j-1}^n$ $j \in \mathbb{Z}$

Supposons $\tau \leq \frac{h}{c_0}$ et montrons

- si $(u_j^n \geq 0$ $j \in \mathbb{Z})$ alors $(u_j^{n+1} \geq 0$ $j \in \mathbb{Z})$
- $\sup_{j\in\mathbb{Z}} |u_j^{n+1}| \leq \sup_{j\in\mathbb{Z}} |u_j^n|$

- si $u_j^n \geq 0$ $j \in \mathbb{Z}$, $u_j^{n+1} \geq 0$ $j \in \mathbb{Z}$ et montrons $u_j^{n+1} \geq 0$ $j \in \mathbb{Z}$

Or $\tau \leq \frac{h}{c_0} \Leftrightarrow c_0\frac{\tau}{h} \leq 1 \Leftrightarrow 1 - c_0\frac{\tau}{h} \geq 0$

u_j^{n+1} est donc la somme et le produit de termes ≥ 0 et donc $u_j^{n+1} \geq 0$ $\forall$ $j \in \mathbb{Z}$

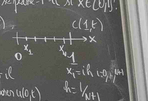
- $\left|u_j^{n+1}\right| = \left|\left(1 - c_0\frac{\tau}{h}\right)u_j^n + c_0\frac{\tau}{h}u_{j-1}^n\right|$ $j \in \mathbb{Z}$

$\leq \left|1 - c_0\frac{\tau}{h}\right| \cdot |u_j^n| + \left|c_0\frac{\tau}{h}\right||u_{j-1}^n|$

$\left(1 - c_0\frac{\tau}{h}\right)\left|u_j^n\right| + c_0\frac{\tau}{h}\left|u_{j-1}^n\right|$

$\leq \left(1 - c_0\frac{\tau}{h} + c_0\frac{\tau}{h}\right)\sup_{k\in\mathbb{Z}}|u_k^n| = \sup_{k\in\mathbb{Z}}|u_k^n|$

Que se passe-t-il si $x \in]0,1]$?



$x_1 = ih$ i.e. $j=i$ $h = \frac{1}{N+1}$

Faut-il imposer $u(0,t)$ ou $u(1,t)$?

Si $c(0,t) > 0$ il faut imposer $u(0,t) =$ quelque chose

Si $c(1,t) < 0$ ——— $u(1,t) =$ autre chose

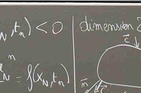
Schéma explicite décentré si $c(x_j,t_n) > 0$:

$\frac{u_j^{n+1} - u_j^n}{\tau} + c(x_j,t_n)\frac{u_j^n - u_{j-1}^n}{h} = f(x_j,t_n)$ — il faut imposer u_0^n

Schéma expl décentré si $c(x_N,t_n) < 0$

$\frac{u_N^{n+1} - u_N^n}{\tau} + c(x_N,t_n)\frac{u_{N+1}^n - u_N^n}{h} = f(x_N,t_n)$ — il faut imposer u_{N+1}^n

dimension 2 d'espace



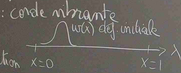
$\frac{\partial u}{\partial t}(x_1,x_2,t) + \vec{c}\cdot\vec{\nabla}u(x_1,x_2,t) = f(x_1,x_2,t)$

il faut imposer u sur le bord si $\vec{c}\cdot\vec{n} < 0$

Chap 13. (13.2 linéaire)

Eq. des ondes

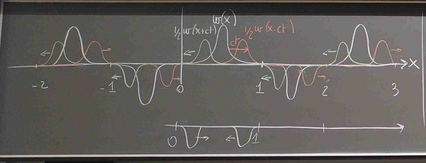
Exp : corde vibrante

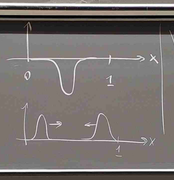


$t=0$ flèche corde interne $w(x)$
$t>0$ on applique force $f(x,t)$
déformation $u(x,t)$?
u satisfait l'éq des ondes

$$\begin{cases} \frac{\partial^2 u}{\partial t^2}(x,t) - c^2 \frac{\partial^2 u}{\partial x^2}(x,t) = f(x,t) & 0<x<1,\ t>0 \\ u(0,t) = 0 \quad u(1,t) = 0 & \text{cond. limites} \quad t>0 \\ u(x,0) = w(x) \\ \frac{\partial u}{\partial t}(x,0) = v(x) \end{cases} \text{ conds initiales } 0<x<1$$

Remarques :
- Si $f(x,t)=0$ et $v(x)=0$ alors
$u(x,t) = \frac{1}{2}\left(w(x-ct) + w(x+ct)\right)$
où w a été prolongée par impaire sur $[-1,0]$ et par 2-périodicité pour $x \in \mathbb{R}$





Vérif

$$\frac{\partial u}{\partial t}(x,t) = \frac{1}{2}\left(w'(x-ct)(-c) + w'(x+ct)(c)\right)$$
$$\frac{\partial^2 u}{\partial t^2}(x,t) = \frac{1}{2}\left(w''(x-ct)c^2 + w''(x+ct)c^2\right)$$
$$\frac{\partial u}{\partial x}(x,t) = \frac{1}{2}\left(w'(x-ct)1 + w'(x+ct)1\right)$$
$$\frac{\partial^2 u}{\partial x^2} = \frac{1}{2}\left(w''(x-ct) + w''(x+ct)\right)$$

$$= c^2 \frac{\partial^2 u}{\partial x^2}$$
$$u(x,0) = \frac{1}{2}\left(w(x)+w(x)\right) = w(x)$$
$$\frac{\partial u}{\partial t}(x,0) = \frac{1}{2}\left(w'(x)(c) + w'(x)(c)\right) = 0$$

$u(0,t) = \frac{1}{2}\left(w(-ct) + w(ct)\right)$
$= \frac{1}{2}\left(-w(ct) + w(ct)\right) = 0$
w impaire
$u(1,t) = \frac{1}{2}\left(w(1-ct) + w(1+ct)\right)$
w 2-périod. $= \frac{1}{2}\left(w(1-ct-2) + w(1+ct)\right)$

$u(1,t) = \frac{1}{2}\left(w(-1-ct) + w(1+ct)\right)$
w impaire $= \frac{1}{2}\left(-w(1+ct) + w(1+ct)\right) = 0$

- L'éq des ondes traduit un principe de conservation (énergie cinétique - déformation)

On multiplie l'éq des ondes par $\frac{\partial u}{\partial t}(x,t)$
et on $\int_0^1 dx$, on obtient

$$\int_0^1 \frac{\partial^2 u}{\partial t^2}(x,t)\frac{\partial u}{\partial t}(x,t)\,dx - \int_0^1 c^2 \frac{\partial^2 u}{\partial x^2}(x,t)\frac{\partial u}{\partial t}(x,t)\,dx = \int_0^1 f(x,t)\frac{\partial u}{\partial t}(x,t)\,dx$$

$$\frac{d}{dt}\left(\frac{1}{2}\int_0^1 \left(\frac{\partial u}{\partial t}(x,t)\right)^2 dx\right)$$

on intègre par parties (en x)

$$(uv)' = u'v + uv' \qquad u(x)\,w(x)$$
$$\left[u\,v\right]_{x=0}^{x=1} = \int_0^1 u'v\,dx + \int_0^1 uv'\,dx$$
$$u \to u' \qquad \left[u'v\right]_{x=0}^{x=1} = \int_0^1 u''v\,dx + \int_0^1 u'v'\,dx$$
$$u' \to \frac{\partial u}{\partial x} \qquad v \to \frac{\partial u}{\partial t}$$

$$\frac{d}{dt}\,\frac{1}{2}\int_0^1 \left(\frac{\partial u}{\partial t}(x,t)\right)^2 dx + c^2 \int_0^1 \frac{\partial u}{\partial x}(x,t)\,\frac{\partial^2 u}{\partial x\,\partial t}(x,t)\,dx = c^2\left[\frac{\partial u}{\partial x}(x,t)\frac{\partial u}{\partial t}(x,t)\right]_{x=0}^{x=1} = \int_0^1 f(x,t)\frac{\partial u}{\partial t}(x,t)\,dx$$

$$\frac{d}{dt}\,\frac{c^2}{2}\int_0^1 \left(\frac{\partial u}{\partial x}(x,t)\right)^2 dx$$

Variation au cours du temps (l'énergie cinétique , l'énergie de déformation) = puissance forces

$$\frac{\partial u}{\partial t}(0,t) = \lim_{\varepsilon \to 0} \frac{u(0,t+\varepsilon) - u(0,t)}{\varepsilon} = \lim_{\varepsilon \to 0} \frac{0-0}{\varepsilon} = 0 = \frac{\partial u}{\partial t}(1,t)$$

Exp : $f(x,t) = 0 \quad v(x) = 0$
$w(x) = \sin(\pi x)$, $u(x,t) = \sin(\pi x)\cos(\pi c t)$

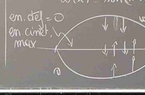
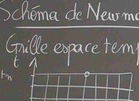


Schéma de Newmark

Grille espace temps



$x_i = ih \quad i = 0,1,\ldots,N,N+1$
$h = \frac{1}{N+1}$
$t_n = n\tau \quad n = 0,1,2,\ldots$
But : calculer u_i^n approx de $u(x_i,t_n)$

On écrit l'éq. pour x_i, temps t_n
$\frac{\partial^2 u}{\partial t^2}(x_i,t_n) - c^2 \frac{\partial^2 u}{\partial x^2}(x_i,t_n) = f(x_i,t_n)$
On utilise des formules de diff. finies centrées

$$\frac{u(x_i,t_{n+1}) - 2u(x_i,t_n) + u(x_i,t_{n-1})}{\tau^2} - c^2 \frac{u(x_{i+1},t_n) - 2u(x_i,t_n) + u(x_{i-1},t_n)}{h^2} = f(x_i,t_n) + O(\tau^2) + O(h^2)$$

$$\begin{cases} \frac{u_i^{n-1} - 2u_i^n + u_i^{n+1}}{\tau^2} - c^2 \frac{u_{i-1}^n - 2u_i^n + u_{i+1}^n}{h^2} = f(x_i,t_n) \\ u_0^n = 0 \quad u_{N+1}^n = 0 \end{cases} \quad i = 1,\ldots,N$$

---

u_i^{n+1} et u_i^n, $i=1,\ldots,N$ on peut calculer u_i^{n+1} $i=1,\ldots,N$

$$u_i^{n+1} = u_i^{n-1} + u_i^n\left(2 - 2\frac{\tau^2 c^2}{h^2}\right) + \frac{\tau^2 c^2}{h^2}\left(u_{i-1}^n + u_{i+1}^n\right) + \tau^2 f(x_i,t_n) \quad i=1,\ldots,N$$

schéma explicite

Etant donné u_i^0 et u_i^1 on peut calculer u_i^2 $i=1,\ldots,N$

Comment calculer u_i^0 et u_i^1, $i=1,\ldots,N$?

$u_i^0 = u(x_i,0) = w(x_i)$
u_i^1 ? on va utiliser $\frac{\partial u}{\partial t}(x_i,0) = v(x_i)$, $i=1,\ldots,N$
en gardant $O(\tau^2)$

$$\frac{u(x_i,\tau) - u(x_i,-\tau)}{2\tau} = v(x_i) + O(\tau^2)$$

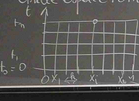
schéma : $\frac{u_i^1 - u_i^{-1}}{2\tau} = v(x_i)$

schéma $n=0$: $\frac{u_i^{-1} - 2u_i^0 + u_i^1}{\tau^2} - c^2\frac{u_{i-1}^0 - 2u_i^0 + u_{i+1}^0}{h^2} = f(x_i,0)$

On élimine u_i^{-1}

Schéma de Newmark

Grille espace temps



$x_i = ih \quad i = 0,1,\ldots,N,N+1$
$h = \frac{1}{N+1}$
$t_n = n\tau \quad n = 0,1,2,\ldots$
But : calculer u_i^n approx de $u(x_i,t_n)$

On écrit l'éq. pour x_i, temps t_n
$\frac{\partial^2 u}{\partial t^2}(x_i,t_n) - c^2 \frac{\partial^2 u}{\partial x^2}(x_i,t_n) = f(x_i,t_n)$
On utilise des formules de diff. finies centrées

$$\frac{u(x_i,t_{n+1}) - 2u(x_i,t_n) + u(x_i,t_{n-1})}{\tau^2} - c^2 \frac{u(x_{i+1},t_n) - 2u(x_i,t_n) + u(x_{i-1},t_n)}{h^2} = f(x_i,t_n) + O(\tau^2) + O(h^2)$$

$$\begin{cases} \frac{u_i^{n-1} - 2u_i^n + u_i^{n+1}}{\tau^2} - c^2 \frac{u_{i-1}^n - 2u_i^n + u_{i+1}^n}{h^2} = f(x_i,t_n) \\ u_0^n = 0 \quad u_{N+1}^n = 0 \end{cases} \quad i = 1,\ldots,N$$

u_i^{n+1} et u_i^n, $i=1,\ldots,N$ on peut calculer u_i^{n+1} $i=1,\ldots,N$

$$u_i^{n+1} = u_i^{n-1} + u_i^n\left(2 - 2\frac{\tau^2 c^2}{h^2}\right) + \frac{\tau^2 c^2}{h^2}\left(u_{i-1}^n + u_{i+1}^n\right) + \tau^2 f(x_i,t_n) \quad i=1,\ldots,N$$

schéma explicite

Etant donné u_i^0 et u_i^1 on peut calculer u_i^2 $i=1,\ldots,N$

Comment calculer u_i^0 et u_i^1, $i=1,\ldots,N$?

$u_i^0 = u(x_i,0) = w(x_i)$
u_i^1 ? on va utiliser $\frac{\partial u}{\partial t}(x_i,0) = v(x_i)$, $i=1,\ldots,N$
en gardant $O(\tau^2)$

$$\frac{u(x_i,\tau) - u(x_i,-\tau)}{2\tau} = v(x_i) + O(\tau^2)$$

schéma : $\frac{u_i^1 - u_i^{-1}}{2\tau} = v(x_i)$

schéma $n=0$: $\frac{u_i^{-1} - 2u_i^0 + u_i^1}{\tau^2} - c^2\frac{u_{i-1}^0 - 2u_i^0 + u_{i+1}^0}{h^2} = f(x_i,0)$

On élimine u_i^{-1}

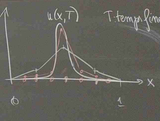
$u_i^{-1} = u_i^1 - 2\tau v(x_i)$ et on injecte dans le schéma $n=0$

$$\frac{2u_i^1 - 2u_i^0 - 2\tau v(x_i)}{\tau^2} - c^2\frac{u_{i-1}^0 - 2u_i^0 + u_{i+1}^0}{h^2} = f(x_i,0)$$

on en déduit $u_i^1 = \ldots$

On implémente ce schéma et on observe que

- le schéma est stable si $\tau \leq \frac{h}{c}$ (identransport)
- Sous cette condition, le schéma converge en $O(\tau^2) + O(h^2)$



Livre
MOOC : outils
Eq dérivées partielles
Alg valeurs propres, SVD
Optim

Chap 7 : calcul de valeurs propres (matrice symmétrique)

$A \in \mathbb{R}^{N\times N}$, on cherche $\lambda \in \mathbb{C}$ et $\vec{\varphi} \in \mathbb{C}^N$ $\vec{\varphi} \neq \vec{0}$ tels que $A\vec{\varphi} = \lambda\vec{\varphi}$ ← vecteur propre, ← valeur propre

Remarques

- Apriori $\lambda \in \mathbb{C}$ et $\vec{\varphi} \in \mathbb{C}^N$
- Ex $A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$ $\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ -i \end{pmatrix} = (1+i)\begin{pmatrix} 1 \\ -i \end{pmatrix}$
- Si λ et $\vec{\varphi}$ sont tq $A\vec{\varphi} = \lambda\vec{\varphi}$ alors on a $(A-\lambda I)\vec{\varphi} = \vec{0}$ et $\vec{\varphi} \neq \vec{0}$ Ker $(A-\lambda I) \neq \vec{0}$ ($A-\lambda I$ n'est pas inversible) $\det(A-\lambda I) = 0$

On cherche $\lambda \in \mathbb{C}$ zéro du polyn charact de $(A-\lambda I)$

On n'existe pas de formule explicite pour trouver les zéros d'un polyn de degré $N \geq 4$. Donc trouver les zéros de ce polyn ou de manière équiv les valeurs propres de $A(\lambda)$ nécessite une méthode itérative

Si $A \in \mathbb{R}^{N\times N}$ est symétrique, $\lambda \in \mathbb{R}$, $\vec{\varphi} \in \mathbb{R}^N$

et il existe une base orthonormée de vecteurs propres.

$A\vec{\varphi}_i = \lambda_i\vec{\varphi}_i$ $i = 1, \dots, N$ $\lambda_i \in \mathbb{R}$

$\vec{\varphi}_i^T\vec{\varphi}_j = \delta_{ij}$ ($1 \leftrightarrow i=j$; $0 \leftrightarrow i\neq j$)

Ex $A = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$

$\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = 1\begin{pmatrix} 1 \\ 1 \end{pmatrix}$

$\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = 3\begin{pmatrix} 1 \\ -1 \end{pmatrix}$

$\lambda_1 = 3$ $\lambda_2 = 1$ $\vec{\varphi}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ $\vec{\varphi}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$

- Si A est sym définie positive alors les valeurs propres sont réelles positives (strictement)

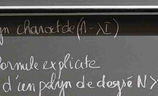
Utilité des valeurs propres

Si on connaît les valeurs propres on peut résoudre les problèmes suivants.

1- Etant donné $\vec{v}_0 \in \mathbb{R}^N$ et $A \in \mathbb{R}^{N\times N}$ sym définie positive, on cherche $\vec{u}(t) \in \mathbb{R}^N$ tq

(1) $\begin{cases} \vec{u}''(t) + A\vec{u}(t) = \vec{0} & t > 0 \\ \vec{u}(0) = \vec{u}_0 \end{cases}$

(chaple après discrétisation en espace,



on obtient (1) avec $u_i(t) \approx u(x_i, t)$

$A = \frac{1}{h^2}\begin{pmatrix} 2 & -1 & & \\ -1 & 2 & -1 & \\ & \ddots & \ddots & \ddots \\ & & -1 & 2 \end{pmatrix}$

A est déf p $A\vec{\varphi}_i = \lambda_i\vec{\varphi}_i$ $i = 1, \dots, N$

On décompose $\vec{u}_0$ dans la base des vecteurs propres

$\vec{u}_0 = \sum_{j=1}^N \alpha_j \vec{\varphi}_j$, $\alpha_j = \vec{u}_0^T\vec{\varphi}_j$

(N=2 $\vec{e}_2$ $\vec{e}_1$ $\vec{u} \in \mathbb{R}^2$)

$\vec{a} = (\vec{a}\cdot\vec{e}_1)\,\vec{e}_1 + (\vec{a}\cdot\vec{e}_2)\,\vec{e}_2$

La sol de (1) est donnée par

$\vec{u}(t) = \sum_{j=1}^N \alpha_j e^{\lambda_j t}\vec{\varphi}_j$

Vérif. $\vec{u}''(t) = \sum_{j=1}^N \alpha_j(\lambda_j)^2 e^{\lambda_j t}\vec{\varphi}_j$

$A\vec{u}(t) = \sum_{j=1}^N \alpha_j e^{\lambda_j t} A\vec{\varphi}_j = \sum_{j=1}^N \alpha_j e^{\lambda_j t}\lambda_j\vec{\varphi}_j$

$\vec{u}''(t) + A\vec{u}(t) = \vec{0}$

$\vec{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_N \end{pmatrix}$ $\vec{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_N \end{pmatrix}$

$\vec{a}^T\vec{b} = (a_1 \cdots a_N)\begin{pmatrix} b_1 \\ \vdots \\ b_N \end{pmatrix}$

2- (2) $\begin{cases} \vec{u}''(t) + A\vec{u}(t) = \vec{0} & t > 0 \\ \vec{u}(0) = \vec{u}_0 \\ \vec{u}'(0) = \vec{0} \end{cases}$

$\vec{u}_0 \in \mathbb{R}^N$ $A \in \mathbb{R}^{N\times N}$ s.d.p. $\vec{u}_0 = \sum_{j=1}^N \alpha_j\vec{\varphi}_j$

$\vec{u}(t)$ est donnée par

$\vec{u}(t) = \sum_{j=1}^N \alpha_j \cos(\sqrt{\lambda_j}\,t)\,\vec{\varphi}_j$

(Chap B après discrétisation en espace)

$A \in \mathbb{R}^{N\times N}$ symétrique

On ordonne les valeurs propres de la manière suivante

$0 \leq |\lambda_1| \leq |\lambda_2| \leq \cdots \leq |\lambda_N|$

Algorithme méthode de la puissance (power)

$\vec{x}^0 \in \mathbb{R}^N$ donné

$n = 1, 2, \dots$

$\vec{x}^n = A\vec{x}^{n-1}$

$\mu^n = \frac{\vec{x}^{n\,T}A\vec{x}^n}{\vec{x}^{n\,T}\vec{x}^n}$ (quotient de Rayleigh)

Thm Hyp : A sym $A\vec{\varphi}_j = \lambda_j\vec{\varphi}_j$ $\vec{\varphi}_i^T\vec{\varphi}_j = \delta_{ij}$

$0 \leq |\lambda_1| \leq |\lambda_2| \leq \cdots \leq |\lambda_N|$

$\vec{x}^{0\,T}\vec{\varphi}_1 \neq 0$

Concl : $\frac{\vec{x}^n}{\|\vec{x}^n\|} \xrightarrow{n\to\infty} \pm\vec{\varphi}_1$ et $\mu^n \xrightarrow{n\to\infty} \lambda_1$

Dém on décompose $\vec{x}^0$ dans la base des vecteurs propres

$\vec{x}^0 = \sum_{j=1}^N \alpha_j\vec{\varphi}_j$ $\alpha_j = \vec{x}^{0\,T}\vec{\varphi}_j$ ($\alpha_1 \neq 0$ par hyp)

$\vec{x}^1 = A\vec{x}^0 = \sum_{j=1}^N \alpha_j A\vec{\varphi}_j = \sum_{j=1}^N \alpha_j\lambda_j\vec{\varphi}_j$

par induction $\vec{x}^n = \sum_{j=1}^N \alpha_j\lambda_j^n\vec{\varphi}_j$

$\vec{x}^n - \vec{x}^{n-1} = \sum_{j=1}^N \alpha_j\lambda_j^{n-1}(\lambda_j - 1)\vec{\varphi}_j$

$\|\vec{x}^n - \vec{x}^{n-1}\|^2 = \sum_{j=1}^N \alpha_j^2\lambda_j^{2n-2}(\lambda_j-1)^2$

$= \left(\sum_{j=1}^N \alpha_j\lambda_j^{n-1}\vec{\varphi}_j^T\right)\left(\sum_{k=1}^N \alpha_k\lambda_k^{n-1}\vec{\varphi}_k\right)$

$= \sum_{j=1}^N\sum_{k=1}^N \alpha_j\alpha_k\lambda_j^{n-1}\lambda_k^{n-1}\vec{\varphi}_j^T\vec{\varphi}_k$ ($\vec{\varphi}_j^T\vec{\varphi}_k = \delta_{jk}$ = 1 si $j=k$, 0 sinon)

$= \sum_{j=1}^N \alpha_j^2\lambda_j^{2n-2}$

$\frac{\vec{x}^n}{\|\vec{x}^n\|} = \frac{\sum_{j=1}^N \alpha_j\lambda_j^n\vec{\varphi}_j}{\sqrt{\sum_{j=1}^N \alpha_j^2\lambda_j^{2n}}}$

Or $|\lambda_j| < |\lambda_1|$ $j = 2, \dots, N$ $\left|\frac{\lambda_j}{\lambda_1}\right| < 1$ et donc $\left(\frac{\lambda_j}{\lambda_1}\right)^n \xrightarrow{n\to\infty} 0$

$\frac{\vec{x}^n}{\|\vec{x}^n\|} = \frac{\sum_{j=1}^N \alpha_j\left(\frac{\lambda_j}{\lambda_1}\right)^n\vec{\varphi}_j}{\sqrt{\sum_{j=1}^N \alpha_j^2\left(\frac{\lambda_j}{\lambda_1}\right)^{2n}}} \xrightarrow{n\to\infty} \frac{\alpha_1\vec{\varphi}_1}{\sqrt{\alpha_1^2}} = \begin{cases} +\vec{\varphi}_1 & \text{si } \alpha_1 > 0 \\ -\vec{\varphi}_1 & \text{si } \alpha_1 < 0 \end{cases}$

$A\vec{x}^n = A\sum_{j=1}^N \alpha_j(\lambda_j)^n\vec{\varphi}_j = \sum_{j=1}^N \alpha_j(\lambda_j)^{n+1}\vec{\varphi}_j$, $\vec{x}^{n\,T}A\vec{x}^n = \left(\sum_{j=1}^N \alpha_j\lambda_j^n\vec{\varphi}_j^T\right)\left(\sum_{k=1}^N \alpha_k\lambda_k^{n+1}\vec{\varphi}_k\right)$

$\mu^n = \frac{\sum_{j=1}^N \alpha_j^2\lambda_j^{2n+1}}{\sum_{j=1}^N \alpha_j^2\lambda_j^{2n}} = \frac{\sum_{j=1}^N \alpha_j^2\lambda_j^{2n+1}}{\sum_{j=1}^N \alpha_j^2\lambda_j^{2n+1}}\lambda_1 \xrightarrow{n\to\infty} \alpha_1^2\lambda_1 = \alpha_1^2\lambda_1$

$= \frac{\sum_{j=1}^N \alpha_j^2\lambda_j^{2n+1}}{\sum_{j=1}^N \alpha_j^2\lambda_j^{2n}} \xrightarrow{n\to\infty} \alpha_1^2\lambda_1 = \alpha_1^2\lambda_1$

2- Remarque : Soit $\lambda, \vec{\varphi} \neq \vec{0}$ tq $A\vec{\varphi} = \lambda\vec{\varphi}$

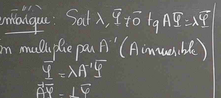
On multiplie par A^{-1} (A inversible)

$\vec{\varphi} = \lambda A^{-1}\vec{\varphi}$

$A^{-1}\vec{\varphi} = \frac{1}{\lambda}\vec{\varphi}$

les vecteurs propres de A^{-1} sont ceux de A

les valeurs propres sont $\frac{1}{\lambda}$



On applique la méthode de la puissance à A^{-1}

Alg puissance inverse

$\vec{x}^0 \in \mathbb{R}^N$

$n = 1, 2, \dots$

$\vec{x}^n = A^{-1}\vec{x}^{n-1}$ (on résout le syst lin $A\vec{x}^n = \vec{x}^{n-1}$)

$\mu^n = \frac{\vec{x}^{n\,T}A\vec{x}^n}{\vec{x}^{n\,T}\vec{x}^n}$ (intuition : si $\frac{\vec{x}^n}{\|\vec{x}^n\|} \to \vec{\varphi}_N$, $\mu^n \xrightarrow{n\to\infty} \frac{\vec{\varphi}_N^T A\vec{\varphi}_N}{\vec{\varphi}_N^T\vec{\varphi}_N} = \lambda_N$)

Thm Hyp sym $A\vec{\varphi}_j = \lambda_j\vec{\varphi}_j$ $j = 1, \dots, N$

$0 < |\lambda_1| < |\lambda_2| \leq \cdots \leq |\lambda_N|$

$\vec{x}^{0\,T}\vec{\varphi}_N \neq 0$

Concl : $\frac{\vec{x}^n}{\|\vec{x}^n\|} \xrightarrow{n\to\infty} \pm\vec{\varphi}_N$

$\mu^n \xrightarrow{n\to\infty} \lambda_N$

Méthode de la puissance inverse avec shift

Soit μ tel que $|\lambda_k - \mu| < |\lambda_j - \mu|$ $j = 1, \dots, N$ $j \neq k$

Pour trouver λ_k on applique l'alg suivant

$\vec{x}^0 \in \mathbb{R}^N$

$n = 1, 2, \dots$

$(A - \mu I)\vec{x}^n = \vec{x}^{n-1}$

$\mu^n = \frac{\vec{x}^{n\,T}A\vec{x}^n}{\vec{x}^{n\,T}\vec{x}^n}$

On peut montrer que (sous certaines cond.)

$\frac{\vec{x}^n}{\|\vec{x}^n\|} \xrightarrow{n\to\infty} \pm\vec{\varphi}_k$

$\mu^n \xrightarrow{n\to\infty} \lambda_k$

SVD (Singular Value decomposition)

But : Compresser l'info

Ex 1 : image — on mémorise les pixels dans $A \in \mathbb{R}^{m \times n}$

Ex 2 : matrice expérience — écoulement derrière un cylindre — $\vec{q}_i$: calcul

Champ de vitesse aux m points grille — $A = \begin{pmatrix} | & & | \\ \vec{q}_1 & \cdots & \vec{q}_n \\ | & & | \end{pmatrix}$ (m lignes, n colonnes)

Thm : $A \in \mathbb{R}^{m \times n}$. Il existe une matrice $U \in \mathbb{R}^{m \times m}$ orthogonale ($U^T U = U U^T = I_m$), une matrice $V \in \mathbb{R}^{n \times n}$ orthogonale ($V V^T = V^T V = I_n$) et une matrice diagonale $\Sigma \in \mathbb{R}^{m \times n}$ de coeff diagonaux $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \cdots \geq \sigma_{\min(m,n)} \geq 0$ (positifs)

Idée de la preuve : $A = U \Sigma V^T$

Si $m \geq n$: $A = U \Sigma V^T$ avec $\Sigma = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{pmatrix}$

Si $m \leq n$: $A = U \Sigma V^T$ avec $\Sigma = \begin{pmatrix} \sigma_1 & & & 0 & \cdots & 0 \\ & \ddots & & \vdots & & \vdots \\ & & \sigma_m & 0 & \cdots & 0 \end{pmatrix}$

Ex : $A \in \mathbb{R}^{n \times n}$ sym def positive — $A \varphi_i = \lambda_i \varphi_i$ $i = 1, \dots, n$ $\lambda > 0$ $\varphi_i^T \varphi_j = \delta_{ij}$

$A = Q \Lambda Q^T$

Puisque les φ_i sont orthonormés : $Q^T Q = I_n$

la matrice Q est orthogonale ($AQ = QA$) Q^T

$A = Q \Lambda Q^T$ $I \Sigma U = Q$ $\Sigma = \Lambda$ $V = Q$

matlab : svd

$A = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \overset{\text{svd}}{=} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \frac{1}{2} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$

$A = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ n'est pas diagonalisable (pas de base de vecteurs propre)

$\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ \alpha \end{pmatrix} = 0 \begin{pmatrix} 0 \\ \alpha \end{pmatrix}$ $\alpha \in \mathbb{R}$

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

Def : Soit r tq $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0 = \sigma_{r+1} = \cdots = \sigma_{\min(m,n)}$. On dit que la matrice est de rang r et on a :

$$A = U \Sigma V^T = \sigma_1 \vec{u}_1 \vec{v}_1^T + \cdots + \sigma_r \vec{u}_r \vec{v}_r^T \quad (*)$$

où $U = \begin{pmatrix} | & & | \\ \vec{u}_1 & \cdots & \vec{u}_m \\ | & & | \end{pmatrix}$ $V = \begin{pmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_n \\ | & & | \end{pmatrix}$

Notation : $\vec{u} \in \mathbb{R}^m$ $\vec{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_m \end{pmatrix}$ $\vec{v} \in \mathbb{R}^n$ $\vec{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$

$\vec{u}\vec{v}^T = \begin{pmatrix} u_1 \\ \vdots \\ u_m \end{pmatrix} \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}$ $\left(\vec{u}\vec{v}^T\right)_{ij} = u_i v_j$

Compression de l'info :

$A \in \mathbb{R}^{m \times n}$ en mémoire $m \times n$ réels

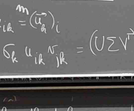
$(*)$ ——— $r \times (m+n)$ réels

Si r est petit devant m ou n alors $r \times (m+n)$ sera petit devant $m \times n$ et $(*)$ est économique

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

Compression d'image

$A = U \Sigma V^T$



Si les σ_i décroissent rapidement il suffit de quelques termes dans $(*)$ pour représenter correctement l'image

Dém de $(*)$: montrons $\left(U \Sigma V^T\right)_{ij} = \left(\sum_{k=1}^{r} \sigma_k \vec{u}_k \vec{v}_k^T\right)_{ij}$ $i = 1, \dots, m$ $j = 1, \dots, n$

En effet :

$$\left(U \Sigma V^T\right)_{ij} = \sum_{k=1}^{m} (U\Sigma)_{ik} \left(V^T\right)_{kj} = \sum_{k=1}^{m} \sum_{l=1}^{n} u_{ik} \Sigma_{kl} v_{jk} = \sum_{k=1}^{r} u_{ik} \sigma_k v_{jk}$$

$1 \leq l \leq k \leq r$

D'autre part :

$$\left(\sum_{k=1}^{r} \sigma_k \vec{u}_k \vec{v}_k^T\right)_{ij} = \sum_{k=1}^{r} \sigma_k \left(\vec{u}_k \vec{v}_k^T\right)_{ij} = \sum_{k=1}^{r} \sigma_k \left(\vec{u}_k\right)_i \left(\vec{v}_k\right)_j$$

$U = \begin{pmatrix} | & & | \\ \vec{u}_1 & \cdots & \vec{u}_m \\ | & & | \end{pmatrix}$ $u_{ik} = \left(\vec{u}_k\right)_i$

$$\longleftarrow = \sum_{k=1}^{r} \sigma_k u_{ik} v_{jk} = \left(U \Sigma V^T\right)_{ij}$$

Remarque :

$A^T A = \left(U \Sigma V^T\right)^T U \Sigma V^T$

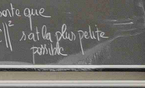
$= V \Sigma^T U^T U \Sigma V^T$

$= V \Sigma^T \Sigma V^T$

$A^T A V = V \Sigma^T \Sigma$

Donc $A^T A V = V \Sigma^T \Sigma$ si $m \geq n$

Analogue $AQ = Q\Lambda$



$A^T A \vec{v}_i = \sigma_i^2 \vec{v}_i$ $i = 1, \dots, n$

Il s'agit de calculer les valeurs propres σ_i^2 et les vecteurs propres $\vec{v}_i$ de $A^T A$

On peut utiliser la méthode de la puissance puissance inverse, shift

$\vec{v}_i$? $AV = U \Sigma \underbrace{V^T V}_{I} = U \Sigma$

$A \vec{v}_i = \vec{u}_i \sigma_i$ $i = 1, \dots, n$

on obtient $\vec{u}_i$ $i = 1, \dots, n$

Application à la matrice d'expérience

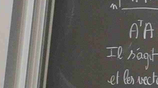
$A = \begin{pmatrix} | & & | \\ \vec{q}_1 & \cdots & \vec{q}_n \\ | & & | \end{pmatrix}$

Comment représenter l'info de manière concise ?

$m = 2$

On cherche $\vec{u}$, $\|\vec{u}\| = 1$ de sorte que $\sum_{i=1}^{n} \left\| \vec{q}_i - \left(\vec{q}_i^T \vec{u}\right) \vec{u} \right\|^2$ soit la plus petite possible

---



$A^T A \vec{v}_i = \sigma_i^2 \vec{v}_i$ $i = 1, \dots, n$

Il s'agit de calculer les valeurs propres σ_i^2 et les vecteurs propres $\vec{v}_i$ de $A^T A$

On peut utiliser la méthode de la puissance puissance inverse, shift

$\vec{v}_i$? $AV = U \Sigma \underbrace{V^T V}_{I} = U \Sigma$

$A \vec{v}_i = \vec{u}_i \sigma_i$ $i = 1, \dots, n$

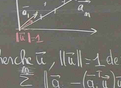
on obtient $\vec{u}_i$ $i = 1, \dots, n$

Application à la matrice d'expérience

$A = \begin{pmatrix} | & & | \\ \vec{q}_1 & \cdots & \vec{q}_n \\ | & & | \end{pmatrix}$

Comment représenter l'info de manière concise ?

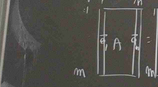
$m = 2$



On cherche $\vec{u}$, $\|\vec{u}\| = 1$ de sorte que $\sum_{i=1}^{n} \left\| \vec{q}_i - \left(\vec{q}_i^T \vec{u}\right) \vec{u} \right\|^2$ soit la plus petite possible

---

On peut montrer que le meilleurs choix est $\vec{u} = \vec{u}_1$



cf PCA statistiques

On cherche deux vecteurs normés à 1 qui représentent au mieux les $\vec{q}_i$ $i = 1, \dots, n$

Le meilleur choix est $\vec{u}_1, \vec{u}_2$

Remarque : puisque $\vec{u}_1, \vec{u}_2, \dots, \vec{u}_m \in \mathbb{R}^m$ sont orthonormés, il constituent une base de $\mathbb{R}^m$

$i = 1, \dots, n$ $\vec{q}_i = \left(\vec{q}_i^T \vec{u}_1\right) \vec{u}_1 + \left(\vec{q}_i^T \vec{u}_2\right) \vec{u}_2 + \cdots + \left(\vec{q}_i^T \vec{u}_m\right) \vec{u}_m$

Partie 2 : optimisation

Optimisation sans contraintes, théorie

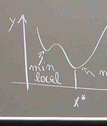
$f : \mathbb{R}^n \to \mathbb{R}$ continue

$\vec{x} \to f(\vec{x})$

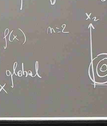
$(x_1, x_2, \dots, x_n) \to f(x_1, x_2, \dots, x_n)$

On cherche $\vec{x}^* \in \mathbb{R}^n$ tel que $f(\vec{x}^*) \leq f(\vec{x}) \ \forall \vec{x} \in \mathbb{R}^n$ (1)

$n=1$



$n=2$



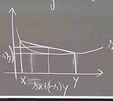
On note aussi $\min_{x \in \mathbb{R}^n} f(x)$ ou $x^* = \operatorname{argmin}_{x \in \mathbb{R}^n} f(x)$

Remarque Si $f : \mathbb{R}^n \to \mathbb{R}$ et si $x^* \in \mathbb{R}^n$ tq $f(x^*) \geq f(x) \ \forall x \in \mathbb{R}^n$ on a $-f(x^*) \leq -f(x)$ — et on ne ramène au pb (1) en posant $f(x) = -f(x)$

Def et notations

- $f : \mathbb{R}^n \to \mathbb{R}$ $\mathcal{C}^1$ on note $\vec{\nabla} f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\vec{x}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\vec{x}) \end{pmatrix}$ gradient
- $f : \mathbb{R}^n \to \mathbb{R}$ $\mathcal{C}^2$ on note $H(\vec{x}) \in \mathbb{R}^{n \times n}$ matrice Hessienne $\left(H(\vec{x})\right)_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}(\vec{x})$
- f est convexe si $\forall \vec{x}, \vec{y} \in \mathbb{R}^n \ \forall 0 \leq s \leq 1$ $f(s\vec{x} + (1-s)\vec{y}) \leq s f(\vec{x}) + (1-s) f(\vec{y})$ ↑ strictement si $<$

$n=1$



Norme euclidienne

$\vec{x} \in \mathbb{R}^n \quad \vec{x} = \begin{pmatrix} x_1 \\ x_n \end{pmatrix} \quad \vec{x}^T \vec{x} = (x_1 \dots x_n) \begin{pmatrix} x_1 \\ x_n \end{pmatrix} = \sum_{i=1}^n x_i^2$

$\|\vec{x}\| = \sqrt{\vec{x}^T \vec{x}} = \sqrt{\sum_{i=1}^n x_i^2}$

Théorème

a. Si $f : \mathbb{R}^n \to \mathbb{R}$ continue, infinie à l'infini ($f(x) \to +\infty$ si $\|\vec{x}\| \to +\infty$) alors il existe $\vec{x}^*$ sol de (1)

b. Si $f : \mathbb{R}^n \to \mathbb{R}$ $\mathcal{C}^1$, si $\vec{x}^*$ sol de (1), alors $\vec{\nabla} f(\vec{x}^*) = \vec{0}$

c. Si $f : \mathbb{R}^n \to \mathbb{R}$ $\mathcal{C}^2$, si $\vec{\nabla} f(\vec{x}^*) = \vec{0}$, si $H(\vec{x}^*)$ sym def positive alors $f(\vec{x}^*) \leq f(\vec{x})$ $\forall \vec{x}$ dans un voisinage de $\vec{x}^*$ (minimum local)

d. Si $f : \mathbb{R}^n \to \mathbb{R}$ $\mathcal{C}^1$ ◦ convexe alors ◦ strictement

(Trouver $\vec{x} \in \mathbb{R}^n$ tq $f(\vec{x}) \leq f(\vec{x}) \ \forall \vec{x} \in \mathbb{R}^n$) ⟺ (Trouver $\vec{x}^*$ tq $\vec{\nabla} f(\vec{x}^*) = \vec{0}$)

Dém :

b. dans $\mathbb{R}^n$ les 2 notions sont égales.

$\vec{\nabla} f(\vec{x}) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\vec{x}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\vec{x}) \end{pmatrix}$ $\frac{\partial f}{\partial x_j}(\vec{x}) =$

G $\lim_{\varepsilon \to 0} \frac{f(x_1 \dots x_j + \varepsilon \dots x_n) - f(x_1 \dots x_n)}{\varepsilon}$

Preuve dans chaque direction

$\forall \vec{d} \in \mathbb{R}^n \ \forall \vec{x} \in \mathbb{R}^n$

$\vec{\nabla} f(\vec{x})^T \vec{d} = \lim_{\varepsilon \to 0} \frac{f(\vec{x} + \varepsilon \vec{d}) - f(\vec{x})}{\varepsilon}$

ex : $\vec{d} = \vec{e}_j = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$ $\vec{\nabla} f(\vec{x}) \vec{e}_j = \lim_{\varepsilon \to 0} \frac{f(x_1 + 0, \dots, x_j + \varepsilon, \dots, x_n) - f(x_1 \dots x_n)}{\varepsilon} = \frac{\partial f}{\partial x_j}(\vec{x})$

Soit $\vec{x}^* \in \mathbb{R}^n$ tq $f(\vec{x}^*) \leq f(\vec{x}) \ \forall \vec{x} \in \mathbb{R}^n$

Montrons $\vec{\nabla} f(\vec{x}^*) = \vec{0}$

On choisit $\vec{x} = \vec{x}^* + \varepsilon \vec{d}$ $\vec{d} \in \mathbb{R}^n$

$f(\vec{x}^*) \leq f(\vec{x}^* + \varepsilon \vec{d})$

$\frac{f(\vec{x}^* + \varepsilon \vec{d}) - f(\vec{x}^*)}{\varepsilon} \geq 0 \quad \forall \varepsilon > 0$

et donc en prenant $\lim_{\varepsilon \to 0}$

$\vec{\nabla} f(\vec{x}^*)^T \vec{d} \geqslant 0 \quad \forall \vec{d} \in \mathbb{R}^n$

On choisit $\vec{d} = -\vec{\nabla} f(\vec{x}^*)$:

$-\|\vec{\nabla} f(\vec{x}^*)\|^2 \geqslant 0$

ie $\|\vec{\nabla} f(\vec{x}^*)\|^2 \leqslant 0$

et donc $\|\vec{\nabla} f(\vec{x}^*)\|^2 = 0$

$\|\vec{\nabla} f(\vec{x}^*)\| = 0$

norme.

$\vec{\nabla} f(\vec{x}^*) = \vec{0}$

c. Supposons $\vec{\nabla} f(\vec{x}^*) = 0$, $H(\vec{x}^*)$ sym def pos, $f \in \mathcal{C}^2$ et montrons $f(\vec{x}^*) \leq f(\vec{x}) \ \forall \vec{x}$ dans un voisinage de $\vec{x}^*$

En effet (Taylor) $f(\vec{x}) = f(\vec{x}^*) + (\vec{x} - \vec{x}^*)^T \vec{\nabla} f(\vec{x}^*) + \frac{1}{2} (\vec{x} - \vec{x}^*)^T H(s\vec{x} + (1-s)\vec{x}^*)(\vec{x} - \vec{x}^*)$

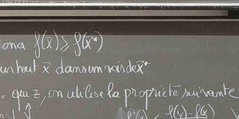
$(n=1 \quad f(x) = f(x^*) + (x - x^*) f'(x^*) + \frac{1}{2} (x - x^*)^2 f''(sx + (1-s)x^*))$

où $0 < s < 1$

Soit $H(\vec{x}^*)$ est s.d.p. alors $H(s\vec{x}^* + (1-s)\vec{x}^*)$ reste s.d.p pour autant que $\vec{x}$ soit suffisamment proche de $\vec{x}^*$ et donc $(\vec{x} - \vec{x}^*)^T H(s\vec{x} + (1-s)\vec{x}^*)(\vec{x} - \vec{x}^*) > 0$

et on a $f(\vec{x}) \geqslant f(\vec{x}^*)$ pourvu que $\vec{x}$ soit dans un voisinage de $\vec{x}^*$

d. Soit $\vec{z}$, on calcule la propriété suivante



$f(\vec{x}) \leq \frac{f(\vec{y}) - f(\vec{x})}{\|\vec{y} - \vec{x}\|} \quad \forall \vec{x}, \vec{y} \in \mathbb{R}^n$

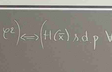
$n \geq 1 \quad \vec{\nabla} f(\vec{x})^T (\vec{y} - \vec{x}) \leq f(\vec{y}) - f(\vec{x}) \quad \forall \vec{x}, \vec{y} \in \mathbb{R}^n$

Remarque $\left(f \text{ convexe } \mathcal{C}^2\right) \Leftrightarrow \left(H(\vec{x}) \text{ s.d.p } \forall \vec{x} \in \mathbb{R}^n\right)$

Exemples :

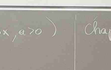
- $f(\vec{x}) = \frac{1}{2} \vec{x}^T A \vec{x} - \vec{b}^T \vec{x} \quad \forall \vec{x} \in \mathbb{R}^n$ où A est s.d.p $A \in \mathbb{R}^{n \times n}$ $\vec{b} \in \mathbb{R}^n$

$(n=1 \quad f(x) = \frac{1}{2} a x^2 - bx \ , a > 0)$



MOOC chap 4, 5, 6 on a montré $\left(\vec{x}^* \in \mathbb{R}^n \text{ tq } f(\vec{x}^*) \leqslant f(\vec{x}) \forall \vec{x} \in \mathbb{R}^n\right) \Rightarrow \left(\vec{x}^* \in \mathbb{R}^n \text{ tq } A\vec{x}^* = \vec{b}\right)$

Chap 10 conde élastique



$A\vec{x}^* = \vec{b}$

$f(\vec{x})$: énergie de déf $\vec{x}$

On utilise cl. du théorème

$f(\vec{x}) = \frac{1}{2} \vec{x}^T A \vec{x} - \vec{b}^T \vec{x}$

$\vec{\nabla} f(\vec{x}) = A\vec{x} - \vec{b}$

$H(\vec{x}) = A$

$\left(H(\vec{x})\right)_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}(\vec{x}) = \frac{\partial}{\partial x_i} \frac{\partial f}{\partial x_j}(\vec{x})$

$= \frac{\partial}{\partial x_i} \left(A\vec{x} - \vec{b}\right)_j$

$= \frac{\partial}{\partial x_i} \left(\sum_{k=1}^n a_{jk} x_k - b_j\right)$

$= a_{ji} = a_{ij}$

$H(\vec{x}) = A$ s.d.p et donc f est str. convexe

Remarque $\left(f \text{ convexe } \mathcal{C}^2\right) \Leftrightarrow \left(H(\vec{x}) \text{ s.d.p } \forall \vec{x} \in \mathbb{R}^n\right)$

Exemples :

- $f(\vec{x}) = \frac{1}{2} \vec{x}^T A \vec{x} - \vec{b}^T \vec{x} \quad \forall \vec{x} \in \mathbb{R}^n$ où A est s.d.p $A \in \mathbb{R}^{n \times n}$ $\vec{b} \in \mathbb{R}^n$

$(n=1 \quad f(x) = \frac{1}{2} a x^2 - bx \ , a > 0)$

MOOC chap 4, 5, 6 on a montré $\left(\vec{x}^* \in \mathbb{R}^n \text{ tq } f(\vec{x}^*) \leqslant f(\vec{x}) \forall \vec{x} \in \mathbb{R}^n\right) \Rightarrow \left(\vec{x}^* \in \mathbb{R}^n \text{ tq } A\vec{x}^* = \vec{b}\right)$

Chap 10 conde élastique

$A\vec{x}^* = \vec{b}$

$f(\vec{x})$: énergie de déf $\vec{x}$

On utilise cl. du théorème

$f(\vec{x}) = \frac{1}{2} \vec{x}^T A \vec{x} - \vec{b}^T \vec{x}$

$\vec{\nabla} f(\vec{x}) = A\vec{x} - \vec{b}$

$H(\vec{x}) = A$

$\left(H(\vec{x})\right)_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}(\vec{x}) = \frac{\partial}{\partial x_i} \frac{\partial f}{\partial x_j}(\vec{x})$

$= \frac{\partial}{\partial x_i} \left(A\vec{x} - \vec{b}\right)_j$

$= \frac{\partial}{\partial x_i} \left(\sum_{k=1}^n a_{jk} x_k - b_j\right)$

$= a_{ji} = a_{ij}$

$H(\vec{x}) = A$ s.d.p et donc f est str. convexe

- $A \in \mathbb{R}^{m \times n}$ $\vec{b} \in \mathbb{R}^m$ $\vec{x} \in \mathbb{R}^n$

$f(\vec{x}) = \frac{1}{2} \|A\vec{x} - \vec{b}\|^2 = \frac{1}{2} \sum_{i=1}^m \left((A\vec{x} - \vec{b})_i\right)^2$

Ex :

$\min \frac{1}{2} \sum_{i=1}^m (\alpha x_i + \beta - y_i)^2$

$A = \begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_m & 1 \end{pmatrix}$ $\vec{x} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ $\vec{b} = \begin{pmatrix} y_1 \\ y_m \end{pmatrix}$

On cherche $\vec{x}^* \in \mathbb{R}^n$ tq $f(\vec{x}^*) \leq f(\vec{x}) \ \forall \vec{x} \in \mathbb{R}^n$

D'après Thm b on a $\vec{\nabla} f(\vec{x}^*) = \vec{0}$

On a $\vec{\nabla} f(\vec{x}) = A^T (A\vec{x} - \vec{b}) \quad \forall \vec{x} \in \mathbb{R}^n$ (2)

et donc $\vec{0} = \vec{\nabla} f(\vec{x}^*) = A^T \left(A\vec{x}^* - \vec{b}\right)$

$A^T A \vec{x}^* = A^T \vec{b}$

A^T A $\vec{x}^*$ $A^T \vec{b}$ $\vec{x} \in \mathbb{R}^n$

Démo de (2) : montrons que $\frac{\partial f}{\partial x_j}(\vec{x}) = \left(A^T (A\vec{x} - \vec{b})\right)_j$

$= \sum_{k=1}^m A^T_{jk} (A\vec{x} - \vec{b})_k$

$A^T_{jk} = A_{kj}$

Ex :

$\min_{a_j, A_n} \frac{1}{2} \sum_{i=1}^n \left(\sum_{j=1}^m a_j \varphi_j(x_i) - y_i\right)^2$ $\sum_{j=1}^m a_j \varphi_j(x)$

Fin cours lundi:

Optim sans contraintes

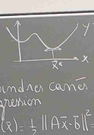
$f : \mathbb{R}^n \to \mathbb{R}$ C^2

$\vec{x} \mapsto f(\vec{x})$

On cherche $\vec{x}^* \in \mathbb{R}^n$ tq $f(\vec{x}^*) \leq f(\vec{x})\ \forall \vec{x} \in \mathbb{R}^n$

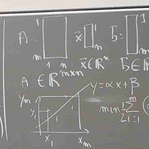
La solution satisfait $\vec{\nabla} f(\vec{x}^*) = \vec{0}$

$n=1$



Ex: moindres carrés régression

$f(\vec{x}) = \frac{1}{2} \| A\vec{x} - \vec{b} \|^2 = \frac{1}{2} \sum_{i=1}^{m} \left((A\vec{x} - \vec{b})_i \right)^2$



$\min \frac{1}{2} \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} x_j(a) - y_i \right)^2$

On a $\vec{\nabla} f(\vec{x}) = A^T (A\vec{x} - \vec{b})$ (à dém)

et donc $\vec{0} = \vec{\nabla} f(\vec{x}^*) = A^T (A\vec{x}^* - \vec{b})$

$A^T A \vec{x}^* = A^T \vec{b}$

On veut dém que:

$\frac{\partial f}{\partial x_k}(\vec{x}) = \left(A^T (A\vec{x} - \vec{b}) \right)_k \quad k = 1, \dots, n$

On a $\left(A^T (A\vec{x} - \vec{b}) \right)_k = \sum_{i=1}^{m} A^T_{ki} \left(A\vec{x} - \vec{b} \right)_i = \sum_{i=1}^{m} A_{ik} \left(A\vec{x} - \vec{b} \right)_i$

$\frac{\partial f}{\partial x_k}(\vec{x}) = \frac{\partial}{\partial x_k} \frac{1}{2} \sum_{i=1}^{m} \left((A\vec{x} - \vec{b})_i \right)^2$

$= \frac{1}{2} \sum_{i=1}^{m} 2 (A\vec{x} - \vec{b})_i \frac{\partial}{\partial x_k} (A\vec{x} - \vec{b})_i$

$= \sum_{i=1}^{m} (A\vec{x} - \vec{b})_i \frac{\partial}{\partial x_k} \left(\sum_{j=1}^{n} A_{ij} x_j - b_i \right)$

$= \sum_{i=1}^{m} (A\vec{x} - \vec{b})_i A_{ik}$

Optimisation sans contraintes: Méthode du gradient

On cherche $\vec{x}^* \in \mathbb{R}^n$ tq $\vec{\nabla} f(\vec{x}^*) = \vec{0}$

$n=2$

$\vec{\nabla} f(\vec{x}_k) \perp f(\vec{x}) = \text{cste}$ qui passe par $\vec{x}^k$

paramétrisation $f(\vec{x}) = \text{cste}$

$f(\vec{x}(t)) = \text{cste}$ t variable

$\frac{d}{dt} f(\vec{x}(t)) = 0 = \vec{\nabla} f(\vec{x}(t)) \cdot \vec{x}'(t)$

$\frac{d}{dt} f(x_1(t), x_2(t)) = 0$

$= \frac{\partial f}{\partial x_1} (x_1(t), x_2(t)) x_1'(t) + \frac{\partial f}{\partial x_2} (\dots) x_2'(t)$

$\alpha > 0$ paramètre (pas, step)

$\vec{x}^{k+1} = \vec{x}^k - \alpha \vec{\nabla} f(\vec{x}^k)$

Si la méthode converge, si $\lim_{k \to \infty} \vec{x}^k = \vec{x}^*$ alors

$\vec{x}^* = \vec{x}^* - \alpha \vec{\nabla} f(\vec{x}^*)$

ie $\vec{\nabla} f(\vec{x}^*) = \vec{0}$

Alg: $\vec{x} \in \mathbb{R}^n$, $\alpha > 0$, $\varepsilon > 0$ petit (10^{-3})

$k = 0, 1, 2, \dots$

$\vec{p}^k = -\vec{\nabla} f(\vec{x}^k)$

Si $\|\vec{p}^k\| < \varepsilon$ stop

Sinon $\vec{x}^{k+1} = \vec{x}^k + \alpha \vec{p}^k$

Question: pour quelle valeurs de α la méthode converge?

Réponse cas où $f(\vec{x}) = \frac{1}{2} \vec{x}^T A \vec{x} - \vec{b}^T \vec{x}$ où $A \in \mathbb{R}^{n \times n}$ s.d.p $\vec{b} \in \mathbb{R}^n$

$n=2$

$f(\vec{x}) = \frac{1}{2} \vec{x}^T A \vec{x} - \vec{b}^T \vec{x}$

$\vec{\nabla} f(\vec{x}) = A\vec{x} - \vec{b}$

$f(\vec{x}) = \text{cste}$ ellipse d'axes les vecteurs propres de A

ex. $A = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$ $\vec{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ $\vec{v}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

Thm: cas quadratique, la méthode converge si $0 < \alpha < \frac{2}{\lambda_n}$ où λ_n est la plus grande valeur propre de A

$\forall \vec{b} \in \mathbb{R}^n \ \forall \vec{x}^0 \in \mathbb{R}^n$, si $\alpha < \frac{2}{\lambda_n}$

où $A \vec{v}_i = \lambda_i \vec{v}_i$, $i = 1, \dots, n$ $\vec{v}_i^T \vec{v}_j = \delta_{ij}$

$0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$

Dém

$\vec{x}^{k+1} - \vec{x}^k = \vec{x}^k - \alpha \vec{\nabla} f(\vec{x}^k) - \vec{x}^k = -\alpha (A\vec{x}^k - \vec{b})$

$\vec{x}^* = \vec{x}^* - \alpha \vec{\nabla} f(\vec{x}^*) = \vec{x}^* - \alpha (A\vec{x}^* - \vec{b})$

$\vec{x}^{k+1} - \vec{x}^* = \vec{x}^k - \vec{x}^* - \alpha A (\vec{x}^k - \vec{x}^*) = (I - \alpha A)(\vec{x}^k - \vec{x}^*)$

Si on peut montrer que $\|\vec{x}^{k+1} - \vec{x}^*\| \leq C \|\vec{x}^k - \vec{x}^*\|$ avec $C < 1$ alors $\|\vec{x}^k - \vec{x}^*\| \leq C^k \|\vec{x}^0 - \vec{x}^*\| \xrightarrow[k \to +\infty]{} 0$ donc la suite converge

On décompose $\vec{x}^0 - \vec{x}^*$ dans la base $(\vec{v}_1, \dots, \vec{v}_n)$

$\vec{x}^0 - \vec{x}^* = \sum_{j=1}^{n} \beta_j \vec{v}_j \quad \left(\beta_j = (\vec{x}^0 - \vec{x}^*)^T \vec{v}_j \right)$

$\|\vec{x}^0 - \vec{x}^*\|^2 = (\vec{x}^0 - \vec{x}^*)^T (\vec{x}^0 - \vec{x}^*) = \left(\sum_{i=1}^{n} \beta_i \vec{v}_i^T \right) \left(\sum_{j=1}^{n} \beta_j \vec{v}_j \right) = \sum_{i=1}^{n} \sum_{j=1}^{n} \beta_i \beta_j \underbrace{\vec{v}_i^T \vec{v}_j}_{\delta_{ij}} = \sum_{i=1}^{n} \beta_i^2$

$\vec{x}^k - \vec{x}^{k+1} = (I - \alpha A)(\vec{x}^k - \vec{x}^*) = \sum_{j=1}^{n} \beta_j (I - \alpha A) \vec{v}_j = \sum_{j=1}^{n} \beta_j (1 - \alpha \lambda_j) \vec{v}_j$

$\|\vec{x}^k - \vec{x}^{k+1}\|^2 = \sum_{j=1}^{n} \beta_j^2 (1 - \alpha \lambda_j)^2 \leq \max_{1 \leq j \leq n} (1 - \alpha \lambda_j)^2 \sum_{j=1}^{n} \beta_j^2 = \max_{1 \leq j \leq n} (1 - \alpha \lambda_j)^2 \|\vec{x}^0 - \vec{x}^*\|^2$

On a montré que $\|\vec{x}^k - \vec{x}^{k+1}\|^2 \leq \max_{1 \leq j \leq n} (1 - \alpha \lambda_j)^2 \|\vec{x}^0 - \vec{x}^*\|^2$

On prend la $\sqrt{\ }$: $\|\vec{x}^k - \vec{x}^{k+1}\| \leq \max_{1 \leq j \leq n} |1 - \alpha \lambda_j| \, \|\vec{x}^0 - \vec{x}^*\|$

La méthode converge si < 1

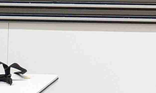
$|1 - \alpha \lambda_j| < 1 \quad i = 1, \dots, n$ si

$-1 < 1 - \alpha \lambda_j < 1$

$\alpha \lambda_j < 2$ (toujours vrai car $\alpha > 0$ et $\lambda_j > 0$)

$\alpha < \frac{2}{\lambda_j}$ pour tout $j = 1, \dots, n$

$(A \ s.d.p)$



et donc si $\alpha < \frac{2}{\lambda_n}$ alors $\max_{1 \leq j \leq n} |1 - \alpha \lambda_j| < 1$

Cas $f(\vec{x}) = \frac{1}{2} \vec{x}^T A \vec{x} - \vec{b}^T \vec{x}$, on peut calculer un pas optimal.

$\vec{p}^k = -\vec{\nabla} f(\vec{x}^k) = \vec{b} - A\vec{x}^k$

On cherche α^k tel que $f(\vec{x}^k + \alpha^k \vec{p}^k) \leq f(\vec{x}^k + \alpha \vec{p}^k) \ \forall \alpha \in \mathbb{R}$

$g(\alpha^k) \leq g(\alpha) \ \forall \alpha \in \mathbb{R}$

On a $g'(\alpha^k) = 0$

Calculons $g'(\alpha) \ \forall \alpha \in \mathbb{R}$

$g(\alpha) = f(\vec{x}^k + \alpha \vec{p}^k) = \frac{1}{2} (\vec{x}^k + \alpha \vec{p}^k)^T A (\vec{x}^k + \alpha \vec{p}^k) - \vec{b}^T (\vec{x}^k + \alpha \vec{p}^k)$

$= \alpha^2 \frac{1}{2} \vec{p}^T A \vec{p}^k + \alpha \left(\frac{1}{2} \vec{p}^{kT} A \vec{x}^k + \frac{1}{2} \vec{x}^{kT} A \vec{p}^k - \vec{b}^T \vec{p}^k \right) + \frac{1}{2} \vec{x}^{kT} A \vec{x}^k - \vec{b}^T \vec{x}^k$

$(A\vec{x}^k - \vec{b})^T \vec{p}^k$

$g'(\alpha) = \alpha \vec{p}^{kT} A \vec{p}^k + (A\vec{x}^k - \vec{b})^T \vec{p}^k$

$= \alpha \vec{p}^{kT} A \vec{p}^k - \vec{p}^{kT} \vec{p}^k$

α^k pas optimal tel que $g'(\alpha^k) = 0$

et donc $\alpha^k = \frac{\vec{p}^{kT} \vec{p}^k}{\vec{p}^{kT} A \vec{p}^k} \ (> 0 \text{ si } \vec{p}^k \neq \vec{0})$

Alg gradient à pas optimal

$\vec{x}^0 \in \mathbb{R}^n$, $\varepsilon > 0$

$k = 0, 1, 2, \dots$

$\vec{p}^k = \vec{b} - A\vec{x}^k$

Si $\|\vec{p}^k\| < \varepsilon$ stop

Sinon $\alpha^k = \frac{\vec{p}^{kT} \vec{p}^k}{\vec{p}^{kT} A \vec{p}^k}$, $\vec{x}^{k+1} = \vec{x}^k + \alpha^k \vec{p}^k$

On peut montrer que le nombre d'itérations est $O\left(\frac{\lambda_n}{\lambda_1} \right)$

Par ex si $A = \begin{pmatrix} 1 & & & 1 \\ & \ddots & & \\ & & \ddots & \\ 1 & & & 1 \end{pmatrix} \in \mathbb{R}^{n \times n}$ $\lambda_1 = 2 - 2\cos\frac{\pi}{n+1}$ $i = 1, \dots, n$

$\lambda_n \xrightarrow[n \to \infty]{} 4$

$\lambda_1 \xrightarrow[n \to \infty]{} 0$

$\frac{\lambda_n}{\lambda_1} = O(n^2)$

Si on est multiplié par 2 le nombre d'iter est multiplié par $2^2 = 4$

Si on est multiplié par 10 le nombre d'iter est multiplié par $10^2 = 100$

$A\vec{x}^* = \vec{b}$. Décomp LL^T ou Cholesky

nombre d'opérations $O(n)$

Méthode du gradient (pas constant ou pas optimal)

$O(n^2)$ itérations

A chaque itération $\vec{p}^k = \vec{b} - A\vec{x}^k$ $O(n)$ opérations

En tout $O(n \cdot n^2) = O(n^3)$ opérations

Idem $-\Delta u = f$ dans $[0,1]^d$ $d = 1, 2, 3$

$d = 10^6$ la méthode du gradient est efficace

Très utilisée en machine learning

Optimisation avec contraintes (1/2)

Rappel : sans contraintes — $f : \mathbb{R}^n \to \mathbb{R}$, $\mathcal{C}^2$

On cherche $\vec{x}^* \in \mathbb{R}^n$ tq $f(\vec{x}^*) \leq f(\vec{x}) \quad \forall \vec{x} \in \mathbb{R}^n$

$\vec{x}^*$ tq $\vec{\nabla} f(\vec{x}^*) = \vec{0}$

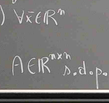
Ex : $f(\vec{x}) = \frac{1}{2}\vec{x}^T A \vec{x} - \vec{b}^T \vec{x} \quad A \in \mathbb{R}^{n \times n}$ s.d.p.

et $\vec{b} \in \mathbb{R}^n$

$\vec{\nabla} f(\vec{x}) = A\vec{x} - \vec{b} \quad \forall \vec{x} \in \mathbb{R}^n$

et donc $A\vec{x}^* = \vec{b}$

Ex : $n=2$ $A = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$ $\vec{b} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ $\vec{x}^* = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$



Avec contraintes

On cherche $\vec{x}^* \in \Omega$ tq $f(\vec{x}^*) \leq f(\vec{x}) \; \forall \vec{x} \in \Omega$

$\Omega \subset \mathbb{R}^n$ est défini par deux fonctions $\vec{g}$ et $\vec{h}$

$\vec{g} : \mathbb{R}^n \to \mathbb{R}^m$, m nombre de contraintes d'inégalité

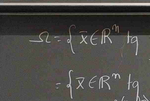
$\vec{x} \to \vec{g}(\vec{x}) = \begin{pmatrix} g_1(x_1, \dots, x_n) \\ \vdots \\ g_m(x_1, \dots, x_n) \end{pmatrix}$

$\vec{h} : \mathbb{R}^n \to \mathbb{R}^p$, p nombre de contraintes d'égalité

$\vec{x} \to \vec{h}(\vec{x}) = \begin{pmatrix} h_1(x_1, \dots, x_n) \\ \vdots \\ h_p(x_1, \dots, x_n) \end{pmatrix}$

$\Omega = \{ \vec{x} \in \mathbb{R}^n \text{ tq } g_i(\vec{x}) \geq 0 \; i=1,\dots,m \; ; \; h_i(\vec{x}) = 0 \; i=1,\dots,p \}$

$= \{ \vec{x} \in \mathbb{R}^n \text{ tq } \vec{g}(\vec{x}) \geqslant \vec{0}, \; \vec{h}(\vec{x}) = \vec{0} \}$



Thm : KKT (Karush Kuhn Tucker) — 1939 — 1953

$f, g, h \in \mathcal{C}^1$, soit $\vec{x}^* \in \Omega$ tq $f(\vec{x}^*) \leq f(\vec{x}) \; \forall \vec{x} \in \Omega$

Sous des hypothèses raisonnables entre $\vec{x}^*$ et g et h

Il existe $\vec{\lambda}^* \in \mathbb{R}^m$ et $\vec{\mu}^* \in \mathbb{R}^p$ (appelés multiplicateurs de lagrange) tels que

$$\vec{\nabla} f(\vec{x}^*) = \sum_{i=1}^{m} \lambda_i^* \vec{\nabla} g_i(\vec{x}^*) + \sum_{i=1}^{p} \mu_i^* \vec{\nabla} h_i(\vec{x}^*)$$

$h_i(\vec{x}^*) = 0 \quad i = 1, \dots, p$

$g_i(\vec{x}^*) \geqslant 0 \quad i = 1, \dots, m$

$\lambda_i^* \geqslant 0$ ——

$\lambda_i^* g_i(\vec{x}^*) = 0$ ——

} conditions de complémentarité

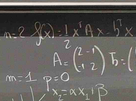
$\lambda_i^* = 0$ ou $g_i(\vec{x}^*) = 0$

Ex : $n=2$ $f(x) = \frac{1}{2} x^T A x - b^T x$

$A = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$ $b = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

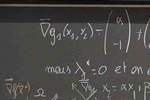
$m=1$ $p=0$

$\Omega = \{ (x_1, x_2) \in \mathbb{R}^2 \text{ tq } x_2 \leqslant \alpha x_1 + \beta \}$



$\vec{\nabla} g_1(x_1, x_2) = \begin{pmatrix} \alpha \\ -1 \end{pmatrix}$ et $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ $\quad \vec{\nabla} f(\vec{x}^*) - A\vec{x}^* + \vec{b} = \vec{0}$

mais $\lambda_1^* = 0$ et on a bien $\vec{\nabla} f(\vec{x}^*) = \lambda_1^* \vec{\nabla} g_1(\vec{x}^*)$



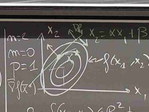
et $\lambda_1^* = g_1(\vec{x}^*) = 0$

$\vec{\nabla} g_1(\vec{x}^*) = \begin{pmatrix} \alpha \\ -1 \end{pmatrix}$

$\vec{\nabla} f(\vec{x}^*) = \lambda_1^* \vec{\nabla} g_1(\vec{x}^*)$ (vect. col. dans la même dir.)

et $\lambda_1^* > 0$

et $g_1(\vec{x}^*) = 0$



$\vec{\nabla} h_1(\vec{x}^*) = \begin{pmatrix} -\alpha \\ 1 \end{pmatrix}$ ou $\begin{pmatrix} \alpha \\ -1 \end{pmatrix}$

$\vec{\nabla} f(\vec{x}^*) = \mu_1^* \vec{\nabla} h_1(\vec{x}^*)$

μ_1^* peut être positif ou négatif, selon la déf$^{\text{n}}$ de $h_1(x_1, x_2)$

La première égalité de cond. KKT peut s'écrire plus facilement en introduisant le lagrangien

---

Remarque : dans les cond KKT, les inconnues sont $\vec{x}^*$, $\vec{\lambda}^*$ et $\vec{\mu}^*$ et on proposera une méthode assez efficace pour rechercher ce point

Soit $\mathcal{L}$ le lagrangien défini par

$\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$

$(\vec{x}, \vec{\lambda}, \vec{\mu}) \to \mathcal{L}(\vec{x}, \vec{\lambda}, \vec{\mu})$

$$\mathcal{L}(\vec{x}, \vec{\lambda}, \vec{\mu}) = f(\vec{x}) - \sum_{i=1}^{m} \lambda_i g_i(\vec{x}) - \sum_{i=1}^{p} \mu_i h_i(\vec{x})$$

On a $\partial_x \mathcal{L}(\vec{x}, \vec{\lambda}, \vec{\mu}) = \vec{\nabla} f(\vec{x}) - \sum_{i=1}^{m} \lambda_i \vec{\nabla} g_i(\vec{x}) - \sum_{i=1}^{p} \mu_i \vec{\nabla} h_i(\vec{x})$

ie $\frac{\partial}{\partial x_k} \mathcal{L}(\vec{x}, \vec{\lambda}, \vec{\mu}) = \frac{\partial}{\partial x_k} f(\vec{x}) - \sum_{i=1}^{m} \lambda_i \frac{\partial g_i}{\partial x_k}(\vec{x}) - \sum_{i=1}^{p} \mu_i \frac{\partial h_i}{\partial x_k}(\vec{x})$

pour $k = 1, \dots, n$ et la première condition KKT s'écrit $\partial_x \mathcal{L}(\vec{x}^*, \vec{\lambda}^*, \vec{\mu}^*) = \vec{0}$

Ex 1 : contraintes d'égalité

$f(\vec{x}) = \frac{1}{2}\vec{x}^T A \vec{x} - \vec{b}^T \vec{x}$ A s.d.p. $\quad \vec{b} \in \mathbb{R}^n$

$m=0$ $\quad \vec{h}(\vec{x}) = B\vec{x} - \vec{c} = \begin{matrix} \uparrow p \\ \downarrow \end{matrix} \boxed{B} \boxed{\vec{x}} = \boxed{\vec{c}}_p$

On cherche $\vec{x}^* \in \Omega$ tq $f(\vec{x}^*) \leq f(\vec{x}) \; \forall \vec{x} \in \Omega$

$\Omega = \{ \vec{x} \in \mathbb{R}^n, \; B\vec{x} - \vec{c} = \vec{0} \} = \{ \vec{x} \in \mathbb{R}^n, \; \vec{h}(\vec{x}) = \vec{0} \}$

On introduit le lagrangien

$\mathcal{L}(\vec{x}, \vec{\mu}) = f(\vec{x}) - \sum_{i=1}^{p} \mu_i h_i(\vec{x})$

$\forall \vec{x} \in \mathbb{R}^n \; \forall \vec{\mu} \in \mathbb{R}^p \quad \mathcal{L}(\vec{x}, \vec{\mu}) = g(\vec{x}) - \sum_{i=1}^{p} \mu_i (B\vec{x} - \vec{c})_i$

$\partial_x \mathcal{L}(\vec{x}, \vec{\mu}) = A\vec{x} - \vec{b} - B^T \vec{\mu}$

$\boxed{B^T} \; \boxed{\vec{\mu}}_p$

En effet $\frac{\partial}{\partial x_k} \mathcal{L}(\vec{x}, \vec{\mu}) = (A\vec{x} - \vec{b})_k - \frac{\partial}{\partial x_k} \sum_{i=1}^{p} \mu_i \left(\sum_{j=1}^{m} B_{ij} x_j - c_i \right)$

$k = 1, \dots, n$

$= \sum_{i=1}^{p} \mu_i \sum_{j=1}^{m} B_{ij} \frac{\partial x_j}{\partial x_k}$

$= \sum_{i=1}^{p} \mu_i B_{ik} = (B^T \vec{\mu})_k \quad \left(= \sum_{i=1}^{p} B_{ki}^T \mu_i \right)$

Les conditions KKT s'écrivent

Trouver $\vec{x}^*$ et $\vec{\mu}^*$ tq

$\partial_x \mathcal{L}(\vec{x}^*, \vec{\mu}^*) = \vec{0} = A\vec{x}^* - \vec{b} - B^T \vec{\mu}^*$

$\vec{h}(\vec{x}^*) = \vec{0} = B\vec{x}^* - \vec{c}$

Il s'agit donc de résoudre le syst. lin

$$\begin{matrix} \uparrow m \\ \downarrow p \end{matrix} \begin{pmatrix} A & -B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} \vec{x}^* \\ \vec{\mu}^* \end{pmatrix} = \begin{pmatrix} \vec{b} \\ \vec{c} \end{pmatrix}$$

Ex 2 : contraintes d'inégalité

$p=0$ $\quad f(\vec{x}) = \frac{1}{2}\vec{x}^T A \vec{x} - \vec{b}^T \vec{x}$

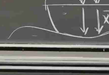
$m=n$ $\quad \vec{g}(\vec{x}) = \vec{x} - \vec{c} \quad g_i(\vec{x}) = x_i - c_i \geqslant 0 \; i = 1, \dots, n$

On cherche $\vec{x}^* \in \Omega$ tq $f(\vec{x}^*) \leq f(\vec{x}) \; \forall \vec{x} \in \Omega$

$\Omega = \{ \vec{x} \in \mathbb{R}^n \text{ tq } \vec{g}(\vec{x}) \geqslant \vec{0} \}$

$= \{ \text{——} \quad x_i \geqslant c_i \; i = 1, \dots, n \}$

Pbm de l'obstacle



On introduit le lagrangien

$\forall \vec{x} \in \mathbb{R}^n \; \forall \vec{\lambda} \in \mathbb{R}^m$

$\mathcal{L}(\vec{x}, \vec{\lambda}) = f(\vec{x}) - \sum_{i=1}^{m} \lambda_i g_i(\vec{x})$

$= \frac{1}{2}\vec{x}^T A \vec{x} - \vec{b}^T \vec{x} - \sum_{i=1}^{m} \lambda_i (x_i - c_i)$

$\partial_x \mathcal{L}(\vec{x}, \vec{\lambda}) = A\vec{x} - \vec{b} - \vec{\lambda}$

Les conditions KKT s'écrivent

Trouver $\vec{x}^*$ et $\vec{\mu}^*$ tq

$\partial_x \mathcal{L}(\vec{x}^*, \vec{\mu}^*) = \vec{0} = A\vec{x}^* - \vec{b} - B^T \vec{\mu}^*$

$\vec{h}(\vec{x}^*) = \vec{0} = B\vec{x}^* - \vec{c}$

Il s'agit donc de résoudre le syst. lin

$$\begin{matrix} \uparrow m \\ \downarrow p \end{matrix} \begin{pmatrix} A & -B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} \vec{x}^* \\ \vec{\mu}^* \end{pmatrix} = \begin{pmatrix} \vec{b} \\ \vec{c} \end{pmatrix}$$

Ex 2 : contraintes d'inégalité

$p=0$ $\quad f(\vec{x}) = \frac{1}{2}\vec{x}^T A \vec{x} - \vec{b}^T \vec{x}$

$m=n$ $\quad \vec{g}(\vec{x}) = \vec{x} - \vec{c} \quad g_i(\vec{x}) = x_i - c_i \geqslant 0 \; i = 1, \dots, n$

On cherche $\vec{x}^* \in \Omega$ tq $f(\vec{x}^*) \leq f(\vec{x}) \; \forall \vec{x} \in \Omega$

$\Omega = \{ \vec{x} \in \mathbb{R}^n \text{ tq } \vec{g}(\vec{x}) \geqslant \vec{0} \}$

$= \{ \text{——} \quad x_i \geqslant c_i \; i = 1, \dots, n \}$

Pbm de l'obstacle

On introduit le lagrangien

$\forall \vec{x} \in \mathbb{R}^n \; \forall \vec{\lambda} \in \mathbb{R}^m$

$\mathcal{L}(\vec{x}, \vec{\lambda}) = f(\vec{x}) - \sum_{i=1}^{m} \lambda_i g_i(\vec{x})$

$= \frac{1}{2}\vec{x}^T A \vec{x} - \vec{b}^T \vec{x} - \sum_{i=1}^{m} \lambda_i (x_i - c_i)$

$\partial_x \mathcal{L}(\vec{x}, \vec{\lambda}) = A\vec{x} - \vec{b} - \vec{\lambda}$

Les cond. KKT sont :

Trouver $\vec{x}^* \in \mathbb{R}^n$ et $\vec{\lambda}^* \in \mathbb{R}^m$ tq

$\partial_x \mathcal{L}(\vec{x}^*, \vec{\lambda}^*) = \vec{0} = A\vec{x}^* - \vec{b} - \vec{\lambda}^*$

$\vec{x}^* - \vec{c} \geqslant \vec{0}, \quad (x_i^* - c_i \geqslant 0 \; i = 1, \dots, n)$

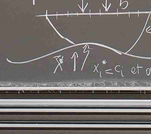
$\vec{\lambda}^* \geqslant \vec{0} \quad (\lambda_i^* \geqslant 0$ ——$)$

$\lambda_i^* (x_i^* - c_i) = 0 \quad i = 1, \dots, n$

Il s'agit d'un pbm non linéaire

On utilisera une méthode itérative

Interprétation pbm de l'obstacle



$\vec{x}^*$ est la force que il faut appliquer à la corde élastique pour qu'elle n'aille pas au delà de l'obstacle

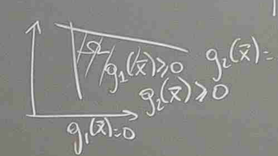
Optim avec contraintes (2/2)
Conditions KKT

Rappel : $f : \mathbb{R}^n \to \mathbb{R}$, $\vec{x} \to f(\vec{x})$

$\Omega = \{ \vec{x} \in \mathbb{R}^n \text{ tq } g_i(\vec{x}) \geq 0 \ i=1,\dots,m \quad h_i(\vec{x}) = 0 \ i=1,\dots,p \}$

On cherche $\vec{x}^* \in \Omega$ tq $f(\vec{x}^*) \leq f(\vec{x}) \ \forall \vec{x} \in \Omega$

$n=2$, $m=2$, $p=0$



les conditions d'optimalités (KKT) sont

trouver $\vec{x}^* \in \mathbb{R}^n$, $\vec{\lambda}^* \in \mathbb{R}^m$, $\vec{\mu}^* \in \mathbb{R}^p$

$$\vec{\nabla} f(\vec{x}^*) = \sum_{i=1}^{m} \lambda_i^* \vec{\nabla} g_i(\vec{x}^*) + \sum_{i=1}^{p} \mu_i^* \vec{\nabla} h_i(\vec{x}^*)$$

$h_i(\vec{x}^*) = 0 \quad i = 1, \dots, p$

$g_i(\vec{x}^*) \geq 0 \quad i = 1, \dots, m$

$\lambda_i^* \geq 0$ ——

$\lambda_i^* \, g_i(\vec{x}^*) = 0$ ——

peut s'écrire

$$\partial_{\vec{x}} \mathcal{L}(\vec{x}^*, \vec{\lambda}^*, \vec{\mu}^*) = 0$$

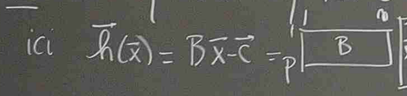
où $\mathcal{L}(\vec{x}, \vec{\lambda}, \vec{\mu})$

$$= f(\vec{x}) - \sum_{i=1}^{m} \lambda_i g_i(\vec{x}) - \sum_{i=1}^{p} \mu_i h_i(\vec{x})$$

Nous allons dém les conditions KKT sur les 2 ex suivants : $f(\vec{x}) = \frac{1}{2} \vec{x}^T A \vec{x} - \vec{b}^T \vec{x}$ A s.d.p $B \in \mathbb{R}^n$

Ex 1 : $\Omega = \{ \vec{x} \in \mathbb{R}^n \text{ tq } B\vec{x} = \vec{c} \}$ $m = 0$

ici $\vec{h}(\vec{x}) = B\vec{x} - \vec{c} = 0$



$$\mathcal{L}(\vec{x}, \vec{\mu}) = f(\vec{x}) - \sum_{i=1}^{p} \mu_i (B\vec{x} - \vec{c})_i$$

$\vec{x}^* \in \mathbb{R}^n$, $\vec{\mu}^* \in \mathbb{R}^p$ tq

$A\vec{x}^* - \vec{b} = B^T \vec{\mu}^*$

$B\vec{x}^* = \vec{c}$

$$\begin{pmatrix} A & -B^T \\ B & 0 \end{pmatrix}_{\substack{m+n \\ m+p}} \begin{pmatrix} \vec{x}^* \\ \vec{\mu}^* \end{pmatrix} = \begin{pmatrix} \vec{b} \\ \vec{c} \end{pmatrix}$$

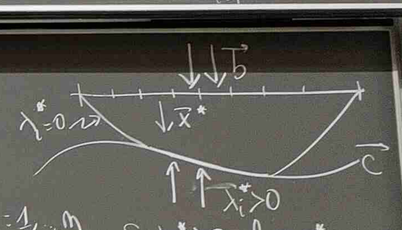
Ex 2 : $\Omega = \{ \vec{x} \in \mathbb{R}^n \text{ tq } \vec{x} \geq \vec{c} \}$ $\quad \vec{g}(\vec{x}) = \vec{x} - \vec{c}$

$m = n$ —— $x_i \geq c_i \ \ i = 1, \dots, n \}$

$p = 0$ $\quad \mathcal{L}(\vec{x}, \vec{\lambda}) = f(\vec{x}) - \sum_{i=1}^{n} \lambda_i (x_i - c_i)$

$\vec{x}^* \in \mathbb{R}^n$ $\vec{\lambda}^* \in \mathbb{R}^n$ tq

$$\begin{cases} A\vec{x}^* - \vec{b} = \vec{\lambda}^* \\ \vec{x}^* - \vec{c} \geq \vec{0} \\ \vec{\lambda}^* \geq \vec{0} \\ \lambda_i^* (x_i^* - c_i) = 0 \quad i = 1, \dots, n \end{cases}$$



Si $\lambda_i^* > 0$ alors $x_i^* = c_i$

Si $x_i^* > c_i$ alors $\lambda_i^* = 0$

On veut montrer que si $\vec{x}^* \in \Omega$ tq $f(\vec{x}^*) \leq f(\vec{x}) \ \forall \vec{x} \in \Omega$ alors les conditions KKT sont satisfaites pour ces 2 exemples

Rappel : dérivée directionnelle $\forall \vec{x} \in \mathbb{R}^n \ \forall \vec{d} \in \mathbb{R}^n$

$$\vec{\nabla} f(\vec{x})^T \vec{d} = \lim_{\varepsilon \to 0} \frac{f(\vec{x} + \varepsilon \vec{d}) - f(\vec{x})}{\varepsilon}$$

par ex $\vec{d} = \vec{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$

$$\vec{\nabla} f(\vec{x})^T \vec{e}_1 = \lim_{\varepsilon \to 0} \frac{f(\vec{x} + \varepsilon \vec{e}_1) - f(\vec{x})}{\varepsilon}$$

$$= \lim_{\varepsilon \to 0} \frac{f(x_1 + \varepsilon, x_2 \dots x_n) - f(x_1 x_2 \dots x_n)}{\varepsilon}$$

$$= \frac{\partial f}{\partial x_1}(\vec{x})$$

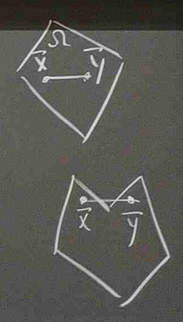
Lemme : Supp Ω convexe

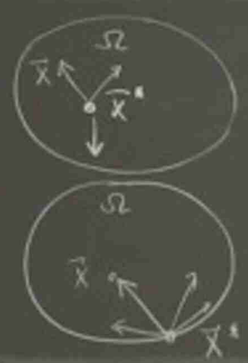
$(\vec{x}, \vec{y} \in \Omega, 0 \leq s \leq 1, s\vec{x} + (1-s)\vec{y} \in \Omega)$

Soit $\vec{x}^* \in \Omega$ tq $f(\vec{x}^*) \leq f(\vec{x}) \ \forall \vec{x} \in \Omega$

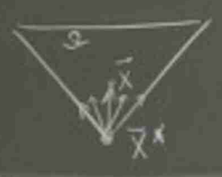
Alors on a

$$\vec{\nabla} f(\vec{x}^*)^T (\vec{x} - \vec{x}^*) \geq 0 \quad \forall \vec{x} \in \Omega$$





f augmente lorsqu'on passe de $\vec{x}^*$ à $\vec{x}$



Dem : Soit $\vec{x} \in \Omega$, soit $0 < s < 1$

Puisque Ω est convexe $s\vec{x} + (1-s)\vec{x}^* \in \Omega$

et donc $f(\vec{x}^*) \leq f(s\vec{x} + (1-s)\vec{x}^*)$

i.e. $f(\vec{x}^*) \leq f(\vec{x}^* + s(\vec{x} - \vec{x}^*))$

ie. $\dfrac{f(\vec{x}^* + s(\vec{x} - \vec{x}^*)) - f(\vec{x}^*)}{s} \geq 0$

On prend la limite quand $s \to 0$:

$$\vec{\nabla} f(\vec{x}^*)^T (\vec{x} - \vec{x}^*) \geq 0$$

Démontrons les conditions KKT pour l'ex 1

Soit $\Omega = \{ \vec{x} \in \mathbb{R}^n \text{ tq } B\vec{x} = \vec{c} \}$

Soit $\vec{x}^* \in \Omega$ tq $f(\vec{x}^*) \leq f(\vec{x}) \ \forall \vec{x} \in \Omega$

Montrons $\exists \vec{\mu}^* \in \mathbb{R}^p$ tq

$$A\vec{x}^* - \vec{b} = B^T \vec{\mu}^*$$

Utilisons le lemme

Ω est convexe : Soit $\vec{x}, \vec{y} \in \Omega$ ($B\vec{x} = \vec{c}$ et $B\vec{y} = \vec{c}$) soit $0 \leq s \leq 1$

$s\vec{x} + (1-s)\vec{y} \in \Omega$

$B(s\vec{x} + (1-s)\vec{y}) = sB\vec{x} + (1-s)B\vec{y} = s\vec{c} + (1-s)\vec{c} = \vec{c}$

D'après le lemme :

$$(A\vec{x}^* - \vec{b})^T (\vec{x} - \vec{x}^*) \geq 0 \quad \forall \vec{x} \text{ tq } B\vec{x} = \vec{c} \quad (1)$$

On choisit $\vec{x} = \vec{x}^* + \vec{y}$ où $B\vec{y} = \vec{0}$ $(\vec{y} \in \operatorname{Ker} B)$

On a bien $\vec{x} \in \Omega$: $B\vec{x} = B(\vec{x}^* + \vec{y}) = B\vec{x}^* + B\vec{y} = \vec{c} + \vec{0} = \vec{c}$

(1) devient

$$(A\vec{x}^* - \vec{b})^T \vec{y} \geq 0$$

On choisit $\vec{x} = \vec{x}^* - \vec{y}$ où $B\vec{y} = \vec{0}$

On a bien $\vec{x} \in \Omega$ et (1) devient

$$(A\vec{x}^* - \vec{b})^T (-\vec{y}) \geq 0$$

ie $(A\vec{x}^* - \vec{b})^T \vec{y} \leq 0$

et donc $(A\vec{x}^* - \vec{b})^T \vec{y} = 0$

ie $A\vec{x}^* - \vec{b}$ est orthogonal à $\vec{y}$

$A\vec{x}^* - \vec{b} \in (\operatorname{Ker} B)^\perp = \operatorname{Im} B^T$

$(\operatorname{Ker} B = (\operatorname{Im} B^T)^\perp)$

$(\forall \vec{x}, \vec{y} \in \mathbb{R}^n, \ \vec{x}^T B\vec{y} = (B\vec{y})^T \vec{x} = \vec{y}^T B^T \vec{x})$

et donc $\exists \vec{\mu}^* \in \mathbb{R}^p$ tq $A\vec{x}^* - \vec{b} = B^T \vec{\mu}^*$

Démontrons les cond. KKT pour l'ex 2

---

les cond. KKT sont $\exists \vec{\lambda}^* \in \mathbb{R}^n$ tq

$A\vec{x}^* - \vec{b} = \vec{\lambda}^*$

$\vec{x}^* - \vec{c} \geq \vec{0}$

$\vec{\lambda}^* \geq \vec{0}$

$\lambda_i^* (x_i^* - c_i) = 0 \quad i = 1, \dots, n$

Éliminons $\vec{\lambda}^*$

$$\begin{cases} A\vec{x}^* - \vec{b} \geq \vec{0} \\ \vec{x}^* - \vec{c} \geq \vec{0} \\ (A\vec{x}^* - \vec{b})_i (x_i^* - c_i) = 0 \quad i = 1, \dots, n \end{cases} \quad (2)$$

Soit $\Omega = \{ \vec{x} \in \mathbb{R}^n \text{ tq } x_i - c_i \geq 0 \ i = 1, \dots, n \}$

Soit $\vec{x}^* \in \Omega$ tq $f(\vec{x}^*) \leq f(\vec{x}) \ \forall \vec{x} \in \Omega$

Montrons les 3 conditions (2)

Montrons $A\vec{x}^* - \vec{b} \geq \vec{0}$

Utilisons le lemme :

Ω est convexe, soit $\vec{x}, \vec{y} \in \Omega$, $0 \leq s \leq 1$

(ie $\vec{x} \geq \vec{c}$ $\vec{y} \geq \vec{c}$)

$s\vec{x} + (1-s)\vec{y} \geq s\vec{c} + (1-s)\vec{c} = \vec{c}$

et donc $s\vec{x} + (1-s)\vec{y} \in \Omega$

D'après le lemme :

$$(A\vec{x}^* - \vec{b})^T (\vec{x} - \vec{x}^*) \geq 0 \quad (3)$$

$\forall \vec{x} \in \mathbb{R}^n$ tq $\vec{x} \geq \vec{c}$

Choisissons $\vec{x} = \vec{x}^* + \vec{e}_i$ $\quad \vec{e}_i = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$ i

$\vec{x} \in \Omega$ car $\vec{x} = \vec{x}^* + \vec{e}_i \geq \vec{c} + \vec{0} = \vec{c}$

---

(3) devient : f augmente lorsqu'on passe

$$(A\vec{x}^* - \vec{b})^T \vec{e}_i \geq 0$$

ie $(A\vec{x}^* - \vec{b})_i \geq 0 \quad i = 1, \dots, n$

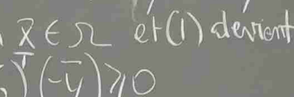
Reste à montrer que $(A\vec{x}^* - \vec{b})_i (x_i^* - c_i) = 0 \quad i = 1, \dots, n$ (4)

Soit $i = 1, \dots, n$ fixé

2 possibilités :

Si $x_i^* = c_i$ alors (4) est satisfaite

Si $x_i^* > c_i$



On choisit dans (3)

$\vec{x}$ tq $c_i < x_i < x_i^*$

$x_j = x_j^* \quad j \neq i$

(3) devient

$$(A\vec{x}^* - \vec{b})_i \underbrace{(x_i - x_i^*)}_{<0} \geq 0$$

et donc $(A\vec{x}^* - \vec{b})_i \leq 0$

On choisit dans (3)

$\vec{x}$ tq $c_i < x_i^* < x_i$

$x_j = x_j^* \quad j \neq i$

(3) devient

$$(A\vec{x}^* - \vec{b})_i \underbrace{(x_i - x_i^*)}_{>0} \geq 0$$

et donc $(A\vec{x}^* - \vec{b})_i \geq 0$

et donc $(A\vec{x}^* - \vec{b})_i = 0$

Optim avec contraintes

Conditions KKT

Algorithme des points intérieurs sur deux exemples

Ex 1 : quadratic programming (obstacle)

On cherche $\bar{x}^* \in \Omega$ tq $f(\bar{x}^*) \leq f(\bar{x})\ \forall \bar{x} \in \Omega$

$\forall \bar{x} \in \mathbb{R}^n \quad f(\bar{x}) = \frac{1}{2}\bar{x}^T A \bar{x} - \bar{b}^T \bar{x}$ $\quad A \in \mathbb{R}^{n\times n}$ s.d.p. $\quad \bar{b} \in \mathbb{R}^n$

et $\Omega = \{\bar{x} \in \mathbb{R}^n$ tq $B\bar{x} - \bar{c} \geqslant 0\}$

$\bar{g}(\bar{x}) = B\bar{x} - \bar{c}$

$p = 0$ pas de contr. d'éq.

$\forall \bar{x} \in \mathbb{R}^n \ \forall \bar{\lambda} \in \mathbb{R}^m$

$\mathcal{L}(\bar{x}, \bar{\lambda}) = \frac{1}{2}\bar{x}^T A \bar{x} - \bar{b}^T \bar{x} - \sum_{i=1}^{m} \lambda_i (B\bar{x} - \bar{c})_i$

$\partial_x \mathcal{L}(\bar{x}, \bar{\lambda}) = A\bar{x} - \bar{b} - B^T \bar{\lambda}$

KKT : $\bar{x}^* \in \mathbb{R}^n$ $\bar{\lambda}^* \in \mathbb{R}^m$ tq

$$\begin{cases} A\bar{x}^* - \bar{b} - B^T\bar{\lambda}^* = \bar{0} \\ B\bar{x}^* - \bar{c} \geqslant \bar{0} \\ \bar{\lambda}^* \geqslant \bar{0} \\ \lambda_i^* (B\bar{x}^* - \bar{c})_i = 0 \quad i = 1, \dots, m \end{cases}$$

On introduit une "slack variable"

$$\begin{cases} A\bar{x}^* - \bar{b} - B^T\bar{\lambda}^* = \bar{0} & \updownarrow n \text{ équations} \\ B\bar{x}^* - \bar{c} = \bar{s}^* & \updownarrow m \text{ éq.} \\ \lambda_i^* s_i^* = 0 \quad i = 1, \dots, m & \updownarrow m \text{ éq.} \\ \bar{x}^* \geqslant \bar{0} \\ \bar{s}^* \geqslant \bar{0} \end{cases}$$

inconnues $\bar{x}^* \in \mathbb{R}^n$, $\bar{\lambda}^* \in \mathbb{R}^m$, $\bar{s}^* \in \mathbb{R}^m$

$n + 2m$ éq. et $n + 2m$ inconnues

1 pas de la méthode de Newton

On écrit les trois premières cond. KKT sous la forme $\bar{F}(\bar{x}^*, \bar{\lambda}^*, \bar{s}^*) = \bar{0}$

où $\forall \bar{x} \in \mathbb{R}^n \ \forall \bar{\lambda} \in \mathbb{R}^m \ \forall \bar{s} \in \mathbb{R}^m$

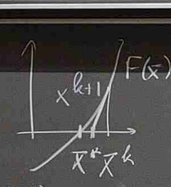
$\bar{F}(\bar{x}, \bar{\lambda}, \bar{s})$ est def. par

$$\bar{F}(\bar{x}, \bar{\lambda}, \bar{s}) = \begin{pmatrix} A\bar{x} - \bar{b} - B^T\bar{\lambda} \\ B\bar{x} - \bar{c} - \bar{s} \\ \lambda_1 s_1 \\ \lambda_i s_i \\ \lambda_m s_m \end{pmatrix}$$

Newton : étant donné $\bar{x}^k, \bar{\lambda}^k, \bar{s}^k$ des approx de $\bar{x}^*, \bar{\lambda}^*, \bar{s}^*$, on résout le syst. linéaire

$$DF(\bar{x}^k, \bar{\lambda}^k, \bar{s}^k) \begin{pmatrix} \Delta\bar{x} \\ \Delta\bar{\lambda} \\ \Delta\bar{s} \end{pmatrix} = \bar{F}(\bar{x}^k, \bar{\lambda}^k, \bar{s}^k)$$

$\bar{F}(\bar{x}^*) = \bar{0}$



$$\bar{0} = \bar{F}(\bar{x}^{k+1}) = \bar{F}(\bar{x}^k) + DF(\bar{x}^k)(\bar{x}^{k+1} - \bar{x}^k) + \text{ordre} > 1$$

$$DF(\bar{x}^k)\underbrace{(\bar{x}^{k+1} - \bar{x}^k)}_{-\Delta\bar{x}} = -\bar{F}(\bar{x}^k)$$

On pose

$\bar{x}^{k+1} = \bar{x}^k - \Delta\bar{x}$

$\bar{\lambda}^{k+1} = \bar{\lambda}^k - \Delta\bar{\lambda}$

$\bar{s}^{k+1} = \bar{s}^k - \Delta\bar{s}$

mais $\bar{\lambda}^* \geqslant \bar{0}$ et $\bar{s}^* \geqslant \bar{0}$

On pose plutôt

$\bar{x}^{k+1} = \bar{x}^k - \Delta\bar{x}$

$\bar{\lambda}^{k+1} = \max(\bar{0}, \bar{\lambda}^k - \Delta\bar{\lambda})$

$\bar{s}^{k+1} = \max(\bar{0}, \bar{s}^k - \Delta\bar{s})$

Reste à calculer $\forall \bar{x} \in \mathbb{R}^n, \forall \bar{\lambda} \in \mathbb{R}^m, \forall \bar{s} \in \mathbb{R}^m$

$DF(\bar{x}, \bar{\lambda}, \bar{s})$ = matrice jacobienne

$$\begin{pmatrix} A & -B^T & 0 \\ B & 0 & -I \\ \lambda_1 & 0 & s_1 \\ 0 & \ddots & 0 \\ 0 & \lambda_m & 0 \end{pmatrix} \quad \begin{matrix} n \\ m \\ m \end{matrix}$$

$\partial_x(A\bar{x} - \bar{b} - B^T\bar{\lambda}) = A$

$\frac{\partial}{\partial x_j}(A\bar{x})_i = \frac{\partial}{\partial x_j}\left(\sum_{k=1}^{n} A_{ik} x_k\right) = \sum_{k=1}^{n} A_{ik} \frac{\partial x_k}{\partial x_j} = A_{ij}$

1 pas de la méthode Newton s'écrit

$$\begin{pmatrix} A & -B^T & 0 \\ B & 0 & -I \\ 0 & \lambda_1^k & s_1^k \\ & & \ddots \\ & & \lambda_m^k \quad s_m^k \end{pmatrix} \begin{pmatrix} \Delta\bar{x} \\ \Delta\bar{\lambda} \\ \Delta\bar{s} \end{pmatrix} = \begin{pmatrix} A\bar{x}^k - \bar{b} - B^T\bar{\lambda}^k \\ B\bar{x}^k - \bar{c} - \bar{s}^k \\ \lambda_1^k s_1^k \\ \lambda_m^k s_m^k \end{pmatrix}$$

On peut montrer que

si $\lambda_i^k > 0$ et $s_i^k > 0$

la matrice $DF(\bar{x}^k, \bar{\lambda}^k, \bar{s}^k)$ (jacobienne) est inversible. Par conséquent on pose

$\bar{x}^{k+1} = \bar{x}^k - \Delta\bar{x}$

$\bar{\lambda}^{k+1} = \max(\bar{\varepsilon}, \bar{\lambda}^k - \Delta\bar{\lambda})$

$\bar{s}^{k+1} = \max(\bar{\varepsilon}, \bar{s}^k - \Delta\bar{s})$

où $\bar{\varepsilon} = \begin{pmatrix} \varepsilon \\ \vdots \\ \varepsilon \end{pmatrix}$ si $\varepsilon > 0$ est un paramètre à choisir

Ex 2 : linear programming

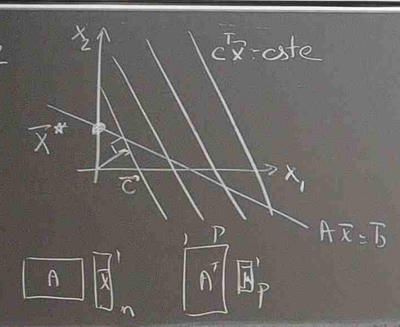
On cherche $\bar{x}^* \in \Omega$ tq $f(\bar{x}^*) \leqslant f(\bar{x})\ \forall \bar{x} \in \Omega$.

$f(\bar{x}) = \bar{c}^T\bar{x}$ $\quad \bar{c} \in \mathbb{R}^n$

$\Omega = \{\bar{x} \in \mathbb{R}^n ;\ \bar{x} \geqslant \bar{0}$ et $A\bar{x} = \bar{b}\}$ $\quad \bar{g}(\bar{x}) = \bar{x}$ $\ m = n$

où $A = B'$ $\quad \bar{h}(\bar{x}) = A\bar{x} - \bar{b}$

$n = 2$



$\forall \bar{x} \in \mathbb{R}^n \ \forall \bar{\lambda} \in \mathbb{R}^n \ \forall \bar{\mu} \in \mathbb{R}^p$

$\mathcal{L}(\bar{x}, \bar{\lambda}, \bar{\mu}) = \bar{c}^T\bar{x} - \sum_{i=1}^{n} \lambda_i x_i - \sum_{i=1}^{p} \mu_i (A\bar{x} - \bar{b})_i$

$\partial_x \mathcal{L}(\bar{x}, \bar{\lambda}, \bar{\mu}) = \bar{c} - \bar{\lambda} - A^T\bar{\mu}$

$\frac{\partial}{\partial x_k}\mathcal{L}(\bar{x}, \bar{\lambda}, \bar{\mu}) = \frac{\partial}{\partial x_k}\left(\sum_{i=1}^{n} c_i x_i\right) - \frac{\partial}{\partial x_k}\left(\sum_{i=1}^{n} \lambda_i x_i\right) - \frac{\partial}{\partial x_k}\sum_{i=1}^{p} \mu_i \left(\sum_{j=1}^{n} A_{ij} x_j\right)$

KKT : trouver $\bar{x}^* \in \mathbb{R}^n$, $\bar{\lambda}^* \in \mathbb{R}^n$, $\bar{\mu}^* \in \mathbb{R}^p$ tq

$$\begin{cases} \text{neq} \\ \text{peq} \\ \text{neq} \end{cases} \begin{cases} \bar{c} - \bar{\lambda}^* - A^T\bar{\mu}^* = \bar{0} \\ A\bar{x}^* - \bar{b} = \bar{0} \\ \lambda_i^* x_i^* = 0 \quad i = 1, \dots, n \\ \bar{x}^* \geqslant \bar{0} \\ \bar{\lambda}^* \geqslant \bar{0} \end{cases}$$ qu'on écrit $\bar{F}(\bar{x}^*, \bar{\lambda}^*, \bar{\mu}^*) = \bar{0}$

($\bar{x}^*$: n inconnues ; $\bar{\lambda}^*$: n inconnues ; $\bar{\mu}^*$: p inconnues)

$\forall \bar{x} \in \mathbb{R}^n \ \forall \bar{\lambda} \in \mathbb{R}^n \ \forall \bar{\mu} \in \mathbb{R}^p$

$$\bar{F}(\bar{x}, \bar{\lambda}, \bar{\mu}) = \begin{pmatrix} \bar{c} - \bar{\lambda} - A^T\bar{\mu} \\ A\bar{x} - \bar{b} \\ \lambda_1 x_1 \\ \lambda_i x_i \\ \lambda_n x_n \end{pmatrix}$$

$$DF(\bar{x}, \bar{\lambda}, \bar{\mu}) = \begin{matrix} n \\ n \\ p \end{matrix} \begin{matrix} \downarrow \\ \downarrow \\ \downarrow \end{matrix} \begin{pmatrix} 0 & -I & -A^T \\ A & 0 & 0 \\ \lambda_i & x_i & 0 \end{pmatrix}$$

1 pas de la méth de Newton

$$DF(\bar{x}^k, \bar{\lambda}^k, \bar{\mu}^k) \begin{pmatrix} \Delta\bar{x} \\ \Delta\bar{\lambda} \\ \Delta\bar{\mu} \end{pmatrix} = \bar{F}(\bar{x}^k, \bar{\lambda}^k, \bar{\mu}^k)$$

$$\begin{pmatrix} 0 & -I & -A^T \\ A & 0 & 0 \\ \lambda_i^k & x_i^k & 0 \end{pmatrix} \begin{pmatrix} \Delta\bar{x} \\ \Delta\bar{\lambda} \\ \Delta\bar{\mu} \end{pmatrix} = \begin{pmatrix} \bar{c} - \bar{\lambda}^k - A^T\mu^k \\ A\bar{x}^k - b \\ \lambda_i^k x_i^k \end{pmatrix}$$

On peut montrer que si $\lambda_i^k > 0$ et $x_i^k > 0$ alors la matrice jacobienne est inversible, par conséquent on pose

$\bar{x}^{k+1} = \max(\bar{\varepsilon}, \bar{x}^k - \Delta\bar{x})$

$\bar{\lambda}^{k+1} = \max(\bar{\varepsilon}, \bar{\lambda}^k - \Delta\bar{\lambda})$ (ε > 0 petit param.)

$\bar{\mu}^{k+1} = \bar{\mu}^k - \Delta\bar{\mu}$

1 pas de la méth de Newton ?

On va montrer que la matrice jacobienne de l'ex 1 est inversible sous conditions

$$\begin{pmatrix} A & -B^T & 0 \\ B & 0 & -I \\ 0 & s_1 \ddots s_m & \lambda_1 \ddots \lambda_m \end{pmatrix} \begin{pmatrix} \tilde{x} \\ \tilde{\lambda} \\ \tilde{s} \end{pmatrix} = \begin{pmatrix} \bar{0} \\ \bar{0} \\ \bar{0} \end{pmatrix} \overset{?}{\Rightarrow} \begin{pmatrix} \tilde{x} = 0 \\ \tilde{\lambda} = 0 \\ \tilde{s} = 0 \end{pmatrix}$$

$$\begin{cases} A\tilde{x} - B^T\tilde{\lambda} = \bar{0} & \text{neq} \\ B\tilde{x} - \tilde{s} = \bar{0} & \text{neq} \\ s_i\tilde{\lambda}_i + \lambda_i\tilde{s}_i = 0 \quad i = 1, \dots, m \end{cases}$$

et montrons $\tilde{x} = \bar{0}$, $\tilde{\lambda} = \bar{0}$, $\tilde{s} = \bar{0}$

$\tilde{x}^T(A\tilde{x} - B^T\tilde{\lambda}) = 0$

$\tilde{\lambda}^T(B\tilde{x} - \tilde{s}) = 0$

on exprime $\tilde{s}_i$ en fonction de $\tilde{\lambda}_i$

$\tilde{s}_i = -\frac{s_i}{\lambda_i}\tilde{\lambda}_i$ si $\lambda_i > 0$

On somme les 2 premières eq

$\tilde{x}^T A\tilde{x} - \tilde{x}^T B^T\tilde{\lambda} + \tilde{\lambda}^T B\tilde{x} - \sum_{i=1}^{m} \tilde{\lambda}_i \left(-\frac{s_i}{\lambda_i}\tilde{\lambda}_i\right) = 0$

$\tilde{x}^T A\tilde{x}$ $\geqslant 0$ car A est s.d.p

$+ \sum_{i=1}^{m} \tilde{\lambda}_i^2 \frac{s_i}{\lambda_i} = 0$ $\geqslant 0$ si $s_i > 0$

Donc $\tilde{x}^T A\tilde{x} = 0$

et donc $\tilde{x} = 0$

$B^T\tilde{\lambda} - A\tilde{x} = 0$

ceci implique $\tilde{x} = 0$ si Ker $B^T = 0$

$\tilde{s}_i = -\frac{s_i}{\lambda_i}\tilde{\lambda}_i = 0$

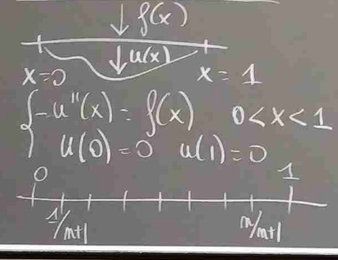
et donc $\tilde{\lambda} = 0$

Pbm de contrôle optimal (identification de param pbm inverse)

Modèle + Mesures

Trouver un comprimis

Modèle : $A\bar{x} = \bar{b}$

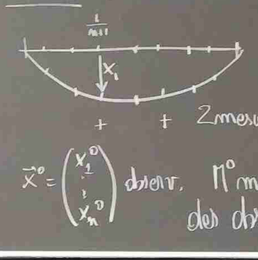


$A = (n+1)^2 \begin{pmatrix} 2 & -1 & & (0) \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ (0) & & -1 & 2 \end{pmatrix} \in \mathbb{R}^{n\times n}$

$b_i = f\left(\frac{i}{n+1}\right) \quad i = 1, \dots, n$

x_i approx de $u\left(\frac{i}{n+1}\right) \quad i = 1, \dots, n$

Mesures (observations)



$\vec{x}^0 = \begin{pmatrix} x_1^0 \\ \vdots \\ x_n^0 \end{pmatrix}$ observ, M^0 matrice des observ

1 observation au point $\frac{1}{n+1}$:

$M^0 = \underset{1}{(1}\ 0 \cdots 0\underset{n}{)} \quad M^0\vec{x} = x_1^0$

2 observations aux x points $\frac{1}{n+1}$ et $\frac{n}{n+1}$:

$M^0 = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & \cdots & & 0 & 1 \end{pmatrix} \quad M^0\vec{x} = \begin{pmatrix} x_1^0 \\ x_n^0 \end{pmatrix}$

Cas général : m observations ($m < n$)

$M^0 = \underset{m}{1}\ \boxed{}\ \overset{n}{1}$ on n'a accès qui à m composantes de $\vec{x}^0$

$M^0\vec{x}^0 = \begin{bmatrix} \\ \\ \\ \end{bmatrix}_m^1$

On se donne un paramètre $\alpha > 0$ qui pondère le modèle et les mesures

On introduit 3 pbms (qui sont équivalents)

Pbm 1 : (problème "physique" primal)

On cherche $\vec{x}^* \in \mathbb{R}^n$ tq $f(\vec{x}^*) \le f(\vec{x}) \ \forall x \in \mathbb{R}^n$

où $\forall \vec{x} \in \mathbb{R}^n \quad f(\vec{x}) = \alpha \underbrace{\frac{1}{2}\|A\vec{x} - \vec{b}\|^2}_{\text{Modèle}} + \underbrace{\frac{1}{2}\|M^0(\vec{x} - \vec{x}^0)\|^2}_{\text{mesures}}$

On introduit la variable $\vec{q} \in \mathbb{R}^n$ tq $A\vec{x} - \vec{b} = \vec{q}$, c'est la force qu'il faut ajouter au modèle de sorte à être proche des mesures. le pbm est le suivant :

Pbm 2 : ($\vec{x}$ variable physique, primale / $\vec{q}$ — de contrôle, duale)

On cherche $(\vec{x}^*, \vec{q}^*) \in \Omega$ tq $J(\vec{x}^*, \vec{q}^*) \le J(\vec{x}, \vec{q})$

où $J(\vec{x}, \vec{q}) = \frac{\alpha}{2}\|\vec{q}\|^2 + \frac{1}{2}\|M^0(\vec{x} - \vec{x}^0)\|^2 \quad \forall (\vec{x}, \vec{q}) \in \Omega$

où $\Omega = \{(\vec{x}, \vec{q}) \in \mathbb{R}^n \times \mathbb{R}^m \text{ tq } A\vec{x} - \vec{b} = \vec{q}\}$

On élimine $\vec{x}$: $A\vec{x} - \vec{b} = \vec{q}$

hyp : A est inversible

($\det A \neq 0$, $\ker A = \vec{0}$, $\operatorname{Im} A = \mathbb{R}^n$)

$\vec{x} = A^{-1}(\vec{b} + \vec{q})$

$M^0(\vec{x} - \vec{x}^0) = M^0\left(A^{-1}(\vec{b} + \vec{q}) - \vec{x}^0\right)$

Pbm 3 : (pbm dual)

On cherche $\vec{q}^* \in \mathbb{R}^n$ tq $f(\vec{q}^*) \le f(\vec{q}) \ \forall q \in \mathbb{R}^n$

où $\forall \vec{q} \in \mathbb{R}^n \quad f(\vec{q}) = \frac{\alpha}{2}\|\vec{q}\|^2 + \frac{1}{2}\underbrace{\left\|M^0\left(A^{-1}(\vec{b} + \vec{q}) - \vec{x}^0\right)\right\|^2}_{\text{qu'on note } B\vec{q} - \vec{c}}$

où $B = M^0A^{-1} = \underset{m}{\overset{n}{\boxed{M^0}}}\ \underset{n}{\overset{n}{\boxed{A^{-1}}}} = \underset{m}{\boxed{B}}$ $\quad \vec{c} = M^0(\vec{x}^0 - A^{-1}\vec{b}) = \begin{bmatrix} \\ \\ \\ \end{bmatrix}_m$

Solution du pbm 1 :

$\vec{x}^*$ satisfait $\vec{\nabla} f(\vec{x}^*) = \vec{0}$

$\forall \vec{x} \in \mathbb{R}^n \quad \vec{\nabla} f(\vec{x}) = \alpha A^T(A\vec{x} - \vec{b}) + M^{0T}M^0(\vec{x} - \vec{x}^0)$

$\vec{x}^*$ satisfait $\alpha A^T(A\vec{x}^* - \vec{b}) + M^{0T}M^0(\vec{x}^* - \vec{x}^0) = 0$

soit encore $\left(\alpha A^TA + M^{0T}M^0\right)\vec{x}^* = \alpha A^T\vec{b} + M^{0T}M^0\vec{x}^0 \quad (1)$

Remarque :

1 mesure :

$M^0 = \underset{1}{(1}\ 0 \cdots 0\underset{m}{)}$

$M^{0T}M^0 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}(1\ 0 \cdots 0) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & & 0 \\ \vdots & & & \\ 0 & 0 & & 0 \end{pmatrix}$

Solution du pbm 2 :

Ecrivons les conditions KKT

Lagrangien $\mathcal{L}(\vec{x}, \vec{q}, \vec{\mu}) = J(\vec{x}, \vec{q}) - \sum_{i=1}^{n} \mu_i(A\vec{x} - \vec{b} - \vec{q})_i$

$\in \mathbb{R}^n \quad \in \mathbb{R}^n \quad \in \mathbb{R}^n$

$= \frac{\alpha}{2}\|\vec{q}\|^2 + \frac{1}{2}\|M^0(\vec{x} - \vec{x}^0)\|^2 - \vec{\mu}^T(A\vec{x} - \vec{b} - \vec{q})$

$\partial_{\vec{x}}\mathcal{L}(\vec{x}, \vec{q}, \vec{\mu}) = M^{0T}M^0(\vec{x} - \vec{x}^0) - A^T\vec{\mu}$

$\partial_{\vec{q}}\mathcal{L}(\vec{x}, \vec{q}, \vec{\mu}) = \alpha\vec{q} + \vec{\mu}$

KKT : $\exists \vec{\mu}^* \in \mathbb{R}^n$ tq

$M^{0T}M^0(\vec{x}^* - \vec{x}^0) - A^T\vec{\mu}^* = \vec{0}$

$\alpha\vec{q}^* + \vec{\mu}^* = \vec{0}$

$A\vec{x}^* - \vec{b} = \vec{q}^*$

On élimine $\vec{\mu}^* = -\alpha\vec{q}^*$ et on a :

$\begin{cases} \alpha A^T\vec{q}^* + M^{0T}M^0(\vec{x}^* - \vec{x}^0) = \vec{0} \\ A\vec{x}^* - \vec{b} - \vec{q}^* = \vec{0} \end{cases} \quad (2)$

Solution du pbm 3

$\vec{q}^*$ satisfait $\vec{\nabla} f(\vec{q}^*) = \vec{0}$

$\forall \vec{q} \in \mathbb{R}^n \quad \vec{\nabla} f(\vec{q}) = \alpha\vec{q} + B^T(B\vec{q} - \vec{c})$

$\vec{q}^*$ satisfait : $\alpha\vec{q}^* + B^T(B\vec{q}^* - \vec{c}) = \vec{0}$

ie $\alpha\vec{q}^* + A^TM^{0T}\left(\underbrace{M^0\left(A^{-1}(\vec{b} + \vec{q}^*) - \vec{x}^0\right)}_{=\vec{0}}\right) \quad (3)$

les 3 systèmes linéaires (1) (2) (3) sont équivalents Δ

En effet, si $\vec{x}^*$ est sol de (1), on pose $\vec{q}^* = A\vec{x}^* - \vec{b}$ et on a

$\begin{cases} \alpha A^T\vec{q}^* + M^{0T}M^0(\vec{x}^* - \vec{x}^0) = \vec{0} \\ A\vec{x}^* - \vec{b} = \vec{q}^* \end{cases}$

qui est bien (2)

Si $(\vec{x}^*, \vec{q}^*)$ sont sol de (2), on élimine $\vec{x}^*$:

$\vec{x}^* = A^{-1}(\vec{b} + \vec{q}^*)$

la première eq de (2) devient

$\alpha A^T\vec{q}^* + M^{0T}M^0\left(A^{-1}(\vec{b} + \vec{q}^*) - \vec{x}^0\right) = \vec{0}$

On multiplie par A^{-T} :

$\alpha\vec{q}^* + A^{-T}M^{0T}M^0\left(A^{-1}(\vec{b} + \vec{q}^*) - \vec{x}^0\right) = \vec{0}$

qui est bien (3).

Si $\vec{q}^*$ est sol de (3), soit $\vec{x}^*$ tq

$\vec{x}^* = A^{-1}(\vec{b} + \vec{q}^*)$

On élimine $\vec{q}^* = A\vec{x}^* - \vec{b}$ et (3) devient

$\alpha\left(A\vec{x}^* - \vec{b}\right) + A^{-T}M^{0T}M^0\left(\vec{x}^* - \vec{x}^0\right) = \vec{0}$

Il suffit de multiplier par A^T pour obtenir (1).

Montrons que (1) a une solution

$\forall \vec{b} \in \mathbb{R}^n \ \forall \vec{x}^0 \in \mathbb{R}^n \ \forall M^0 \in \mathbb{R}^{m\times n}$

Montrons ker $\alpha A^TA + M^{0T}M^0 = \vec{0}$

En effet soit $\vec{x} \in \mathbb{R}^n$ tq $\left(\alpha A^TA + M^{0T}M^0\right)\vec{x} = \vec{0}$

montrons $\vec{x} = \vec{0}$. En effet, on fait le pdt scalaire avec $\vec{x}$:

$\vec{x}^T\left(\alpha A^TA + M^{0T}M^0\right)\vec{x} = 0$

i.e. $\alpha\ \vec{x}^TA^TA\vec{x} + \vec{x}^TM^{0T}M^0\vec{x} = 0$

$\alpha > 0 \rightsquigarrow \alpha \quad \|A\vec{x}\|^2 + \|M^0\vec{x}\|^2 = 0$

et donc $\|A\vec{x}\| = 0$ ie $A\vec{x} = \vec{0}$

et puisque A est inversible $\vec{x} = \vec{0}$

Exercice :

Pbm 3 : on cherche $\vec{q}^* \in \mathbb{R}^n$ qui minimise $\frac{\alpha}{2}\|\vec{q}\|^2 + \frac{1}{2}\|B\vec{q} - \vec{c}\|^2$

Pbm 4 : $\vec{q}^*$ qui minimise $\frac{1}{2}\|\vec{q}\|^2$ sous la contrainte $B\vec{q} = \vec{c}$ (le modèle satisfait exactement les mesures)

On cherche $\vec{q}^* \in \Omega$ tq $f(\vec{q}^*) \le f(\vec{q}) \ \forall \vec{q} \in \Omega$ où $f(\vec{q}) = \frac{1}{2}\|\vec{q}\|^2$ et $\Omega = \{\vec{q} \in \mathbb{R}^n \text{ tq } B\vec{q} = \vec{c}\}$

Aug Contrôle optimal
Me Neural network 1960
Me: ——— 2012
Dernière sem: la. chap 14, me. exam 23

Contrôle opt.: suite et fin

Pbm 1. $\min_{\vec{x}} \underbrace{\frac{\alpha}{2}\|A\vec{x}-\vec{b}\|^2 + \frac{1}{2}\|M^\circ(\vec{x}-\vec{x}^\circ)\|^2}_{f(\vec{x})}$

$\vec{x}^* \in \mathbb{R}^n$ tq $f(\vec{x}^*) \le f(\vec{x})\ \forall \vec{x} \in \mathbb{R}^n$

$\vec{\nabla} f(\vec{x}^*) = \vec{0}$

$\alpha A^T(A\vec{x}^*-\vec{b}) + M^{\circ T}M^\circ(\vec{x}^*-\vec{x}^\circ) = \vec{0}$

1 obs on x_1. $M^\circ = (1\ 0\ \dots\ 0)$

$M^\circ \vec{x}^\circ = x_1^\circ$

$M^{\circ T}M^\circ = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}$

m observations M° $\begin{matrix} 1 \\ m \end{matrix}\boxed{}\begin{matrix} \\ n \end{matrix}$ (1 … n)

Pbm 2: On introduit $\vec{q} = A\vec{x}-\vec{b}$

$\min_{\vec{x},\vec{q}} \underbrace{\frac{\alpha}{2}\|\vec{q}\|^2 + \frac{1}{2}\|M^\circ(\vec{x}-\vec{x}^\circ)\|^2}_{f(\vec{x},\vec{q})}$ sous la contrainte $A\vec{x}-\vec{b}-\vec{q} = \vec{0}$

Lagrangien $\mathcal{L}(\vec{x},\vec{q},\vec{\mu}) = f(\vec{x},\vec{q}) - \sum_{i=1}^m \mu_i (A\vec{x}-\vec{b}-\vec{q})_i$

KKT: tq $\alpha A^T\vec{q}^* + M^{\circ T}M^\circ(\vec{x}^*-\vec{x}^\circ) = \vec{0}$

$A\vec{x}^*-\vec{b} = \vec{q}^*$

Pbm 3. on élimine $\vec{x}$:

$A\vec{x}-\vec{b} = \vec{q}$ $\quad \vec{x} = A^{-1}(\vec{b}+\vec{q})$

$M^\circ(\vec{x}-\vec{x}^\circ) = M^\circ(A^{-1}(\vec{b}+\vec{q})-\vec{x}^\circ) = B\vec{q}-\vec{c}$ $\quad B = \begin{matrix} 1 \\ n \end{matrix}\boxed{}\begin{matrix} \\ m \end{matrix}$, $\vec{c} = \begin{matrix} 1 \\ n \end{matrix}\boxed{}\begin{matrix} \\ m \end{matrix}$

$\min_{\vec{q}} \underbrace{\frac{\alpha}{2}\|\vec{q}\|^2 + \frac{1}{2}\|B\vec{q}-\vec{c}\|^2}_{g(\vec{q})}$

$\vec{\nabla} g(\vec{q}^*) = 0$

$\alpha\vec{q}^* + B^T(B\vec{q}^*-\vec{c}) = \vec{0}$

Pbm 4: on veut que les mesures soient satisfaites exactement

$B\vec{q}-\vec{c} = \vec{0}$

$\min_{\vec{q}} \frac{1}{2}\|\vec{q}\|^2$ sous la contrainte $B\vec{q}-\vec{c}$

On cherche $\vec{q}^* \in \Omega$ tq $f(\vec{q}^*) \le f(\vec{q})\ \forall \vec{q} \in \Omega$

$f(\vec{q}) = \frac{1}{2}\|\vec{q}\|^2$

$\Omega = \{\vec{q} \in \mathbb{R}^m,\ \underbrace{B\vec{q}-\vec{c}}_{\vec{g}(\vec{q})} = \vec{0}\}$

On introduit le lagrangien $\forall \vec{q} \in \mathbb{R}^m$, $\forall \vec{\mu} \in \mathbb{R}^m$

$\mathcal{L}(\vec{q},\vec{\mu}) = \frac{1}{2}\|\vec{q}\|^2 - \sum_{i=1}^m \mu_i (B\vec{q}-\vec{c})_i$

$\vec{\nabla}_{\vec{q}}\mathcal{L}(\vec{q},\vec{\mu}) = \vec{q} - B^T\vec{\mu}$

$\vec{q}^*$ et $\vec{\mu}^*$ sont tels que

$\begin{cases} \vec{q}^* - B^T\vec{\mu}^* = 0 \\ B\vec{q}^* = \vec{c} \end{cases}$

$\begin{pmatrix} I & -B^T \\ B & 0 \end{pmatrix}\begin{pmatrix} \vec{q}^* \\ \vec{\mu}^* \end{pmatrix} = \begin{pmatrix} \vec{0} \\ \vec{c} \end{pmatrix}$

Matrice carrée inversible? $\ker = \vec{0}$?

$\begin{pmatrix} I & -B^T \\ B & 0 \end{pmatrix}\begin{pmatrix} \vec{q} \\ \vec{\mu} \end{pmatrix} = \begin{pmatrix} \vec{0} \\ \vec{0} \end{pmatrix} \overset{?}{\Rightarrow} \begin{pmatrix} \vec{q} = \vec{0} \\ \vec{\mu} = \vec{0} \end{pmatrix}$

$\begin{cases} \vec{q} - B^T\vec{\mu} = \vec{0} \\ B\vec{q} = \vec{0} \end{cases} \overset{?}{\Rightarrow} \begin{pmatrix} \vec{q} = \vec{0} \\ \vec{\mu} = \vec{0} \end{pmatrix}$

On élimine $\vec{q}$.

$BB^T\vec{\mu} = \vec{0}$

On fait le pdt scal avec $\vec{\mu}$

$\vec{\mu}^T BB^T\vec{\mu} = 0$

$\|B^T\vec{\mu}\|^2 = 0$

et donc $B^T\vec{\mu} = \vec{0}$

ce qui implique $\mu = \vec{0}$

si $\ker B^T = \vec{0}$

$B = M^\circ A^{-1}$ $\quad \begin{matrix} 1 \\ m \end{matrix}\boxed{}\begin{matrix} \\ m \end{matrix}$, A^{-1} $\quad \begin{matrix} 1 \\ n \end{matrix}\boxed{}\begin{matrix} \\ n \end{matrix}$

$B^T = A^{-T}M^{\circ T}$

$B^T\vec{\mu} = \vec{0} = A^{-T}M^{\circ T}\vec{\mu} = \vec{0} \Rightarrow M^{\circ T}\vec{\mu} = \vec{0}$

A inversible $\Rightarrow \vec{\mu} = \vec{0}$ si $\ker M^{\circ T} = \vec{0}$

$M^{\circ T}$ ——

$\ker M^{\circ T} = \vec{0}$

par exemple $m = 1$

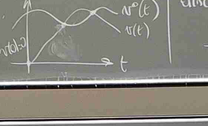
$M^\circ = (1\ 0\ \dots\ 0)$ $\quad M^{\circ T} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ $\quad \ker M^{\circ T} = \vec{0}$

Dans le résumé Pbm 5 et 6 (pas à l'examen)

$\min_{\vec{q}} \sum_{i=1}^m |q_i|$ contrainte $B\vec{q} = \vec{c}$

Pbm 7: particule soumise à une force

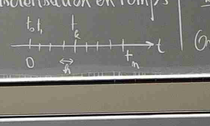
On considère une part de masse 1 dont la vitesse $v(t)$ soumise à une force $g(t)$. On souhaite déterminer $g(t)$ de sorte que la vitesse $v(t)$ soit proche d'une vitesse objectif $v^\circ(t)$ à certains temps t_i



Eq de Newton:

$\begin{cases} v'(t) = g(t) \\ v(0) = 0 \end{cases}$

discrétisation en temps



$t_k = kh$ $\quad k = 1, \dots, m$

schéma d'Euler v_k approche de $v(t_k)$ $\quad k = 1, \dots, m$

q_k —— $q(t_k)$ ——

$\frac{v_k - v_{k-1}}{h} = q_k$ $\quad k = 1, \dots, m$ $\quad v_0 = 0$

On écrit ces relations sous forme d'1 syst. lin

$\frac{1}{h}\begin{pmatrix} 1 & 0 & \dots & 0 \\ -1 & 1 & & \vdots \\ & \ddots & \ddots & \\ 0 & & -1 & 1 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{pmatrix} = \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_m \end{pmatrix}$

B $\quad \vec{v} = \vec{q}$

le pbm d'optimisation est

$\min_{\vec{v},\vec{q}} \frac{\alpha}{2}\|\vec{q}\|^2 + \frac{1}{2}\|M^\circ(\vec{v}-\vec{v}^\circ)\|^2$ sous la contrainte $B\vec{v}-\vec{q}$

par exemple si on a une vitesse objectif au temps t_m:

$M^\circ = (0\ \dots\ 0\ 1)$, $M^\circ\vec{v}^\circ = v_m^\circ$ vitesse au temps t_m

Cas général $M^\circ = \begin{matrix} 1 \\ m \end{matrix}\boxed{}\begin{matrix} \\ n \end{matrix}$ $\quad$ mi vitesses objectif

On cherche $(\vec{v}^*,\vec{q}^*) \in \Omega$ tq $f(\vec{v}^*,\vec{q}^*) \le f(\vec{v},\vec{q})\ \forall (\vec{v},\vec{q}) \in \Omega$

où $f(\vec{v},\vec{q}) = \frac{\alpha}{2}\|\vec{q}\|^2 + \frac{1}{2}\|M^\circ(\vec{v}-\vec{v}^\circ)\|^2$

et $\Omega = \{(\vec{v},\vec{q}) \in \mathbb{R}^n \times \mathbb{R}^n,\ B\vec{v}-\vec{q} = \vec{0}\}$

$\mathcal{L}(\underbrace{\vec{v},\vec{q}}_{\in \mathbb{R}^n \times \mathbb{R}^n}, \underbrace{\vec{\mu}}_{\in \mathbb{R}^m}) = f(\vec{v},\vec{q}) - \sum_{i=1}^m \mu_i (B\vec{v}-\vec{q})_i$

$\partial_{\vec{v}}\mathcal{L}(\vec{v},\vec{q},\vec{\mu}) = M^{\circ T}M^\circ(\vec{v}-\vec{v}^\circ) - B^T\vec{\mu}$

$\partial_{\vec{q}}\mathcal{L}(\vec{v},\vec{q},\vec{\mu}) = \alpha\vec{q} + \vec{\mu}$

KKT: trouver $(\vec{v}^*,\vec{q}^*,\vec{\mu}^*) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m$ tq

$\begin{cases} \partial_{\vec{v}}\mathcal{L}(\vec{v}^*,\vec{q}^*,\vec{\mu}^*) = \vec{0} \\ \partial_{\vec{q}}\mathcal{L}(\vec{v}^*,\vec{q}^*,\vec{\mu}^*) = \vec{0} \\ B\vec{v}^* = \vec{q}^* \end{cases} \Leftrightarrow \begin{cases} M^{\circ T}M^\circ(\vec{v}^*-\vec{v}^\circ) - B^T\vec{\mu}^* = \vec{0} \\ \alpha\vec{q}^* + \vec{\mu}^* = \vec{0} \\ B\vec{v}^* - \vec{q}^* = \vec{0} \end{cases}$

On élimine $\vec{\mu}^*$

$\begin{cases} \alpha B^T\vec{q}^* + M^{\circ T}M^\circ(\vec{v}^*-\vec{v}^\circ) = \vec{0} \\ B\vec{v}^* - \vec{q}^* = \vec{0} \end{cases}$

$\begin{pmatrix} M^{\circ T}M^\circ & \alpha B^T \\ B & -I \end{pmatrix}\begin{pmatrix} \vec{v}^* \\ \vec{q}^* \end{pmatrix} = \begin{pmatrix} M^{\circ T}M^\circ\vec{v}^\circ \\ \vec{0} \end{pmatrix}$

cf Pbm 2

On élimine $\vec{q}^*$:

$(\alpha B^TB + M^{\circ T}M^\circ)\vec{v}^* = M^{\circ T}M^\circ\vec{v}^\circ$

Si 1 vitesse objectif au temps t_m:

$M^{\circ T}M^\circ = \begin{pmatrix} 0 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix}$

Montrons $\ker(\alpha B^TB + M^{\circ T}M^\circ) = \vec{0}$

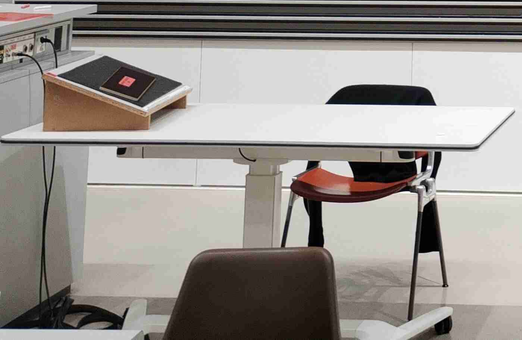
ie $(\alpha B^TB + M^{\circ T}M^\circ)\vec{v} = \vec{0} \Rightarrow (\vec{v} = \vec{0})$

produit scalaire avec $\vec{v}$:

$\alpha\vec{v}^TB^TB\vec{v} + \vec{v}^TM^{\circ T}M^\circ\vec{v} = 0$

$\alpha\|B\vec{v}\|^2 + \|M^\circ\vec{v}\|^2 = 0$

$\Rightarrow$ et donc $\|B\vec{v}\| = 0 \Leftrightarrow (B\vec{v} = \vec{0}) \Leftrightarrow (\vec{v} = \vec{0})$



Neural Networks

Machine à interpoler $\vec{f} : \mathbb{R}^{d_{in}} \to \mathbb{R}^{d_{out}}$
avec d_{in} et d_{out} grands (10^3-10^5)
Ex : classification d'images

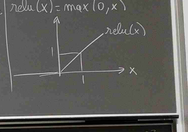
1 image : 1000×1000 = 10^6 pixels (x)
output : chat, chien, arbre, voiture ...

- Tous layers Neural Network (perceptron 1960)
- Deep Neural Neural Network (2012)

Auj : Tous layers NN

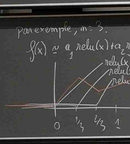
Soit $f : [0,1] \to \mathbb{R}$ cont donnée.
qui on veut approcher par
$\sum_{j=1}^{m} a_j \, \mathrm{relu}(w_j x + b_j)$
où $a_j, w_j, b_j \in \mathbb{R}$ et

$\mathrm{relu}(x) = \max(0, x)$



par exemple, $m = 3$:

$f(x) \simeq a_1 \, \mathrm{relu}(x) + a_2 \, \mathrm{relu}(x - \frac{1}{3}) + a_3 \, \mathrm{relu}(x - \frac{2}{3})$

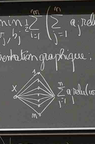


$\sum_{j=1}^{m} a_j \, \mathrm{relu}(w_j x + b_j)$ est continue
polynômiale de degré 1 par
morceaux x (dem
éléments finis 1d, chap 10)

Question : comment choisir a_j, w_j, b_j
de sorte que $\sum_{j=1}^{m} a_j \, \mathrm{relu}(w_j x + b_j)$
soit une bonne approx de $f(x)$?
(On se donne des points d'interpolation
$0 \le x_i \le 1$, $i = 1, \dots, m$ et on veut minimiser)

$\min_{a_j, w_j, b_j} \sum_{i=1}^{m} \left(\sum_{j=1}^{m} a_j \, \mathrm{relu}(w_j x_i + b_j) - f(x_i) \right)^2$

Représentation graphique :



2 layers NN
1 hidden layer

On incide $\vec{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix}$ $\vec{w} = \begin{pmatrix} w_1 \\ \vdots \\ w_m \end{pmatrix}$ $\vec{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$

On cherche $(\vec{a}^*, \vec{w}^*, \vec{b}^*) \in \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^m$ tq
$F(\vec{a}^*, \vec{w}^*, \vec{b}^*) \le F(\vec{a}, \vec{w}, \vec{b}) \quad \forall (\vec{a}, \vec{w}, \vec{b}) \in \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^m$

où $F(\vec{a}, \vec{w}, \vec{b}) = \sum_{i=1}^{m} \left(\sum_{j=1}^{m} a_j \, \mathrm{relu}(w_j x_i + b_j) - f(x_i) \right)^2$

La solution satisfait
$\begin{cases} \frac{\partial F}{\partial a_k}(\vec{a}^*, \vec{w}^*, \vec{b}^*) = 0 \\ \frac{\partial F}{\partial w_k}(\vec{a}^*, \vec{w}^*, \vec{b}^*) = 0 \\ \frac{\partial F}{\partial b_k}(\vec{a}^*, \vec{w}^*, \vec{b}^*) = 0 \end{cases}$ où on a noté

$\frac{\partial F}{\partial a_k}(\vec{a}, \vec{w}, \vec{b}) = \begin{pmatrix} \frac{\partial F}{\partial a_1}(\vec{a}, \vec{w}, \vec{b}) \\ \vdots \\ \frac{\partial F}{\partial a_m}(\vec{a}, \vec{w}, \vec{b}) \end{pmatrix} \in \mathbb{R}^m$

idem pour $\frac{\partial F}{\partial w_k}$ et $\frac{\partial F}{\partial b_k}$

Soit $k = 1, \dots, m$ on a :

$\frac{\partial F}{\partial a_k}(\vec{a}, \vec{w}, \vec{b}) = \frac{\partial}{\partial a_k} \sum_{i=1}^{m} \left(\sum_{j=1}^{m} a_j \, \mathrm{relu}(w_j x_i + b_j) - f(x_i) \right)^2$

$= \sum_{i=1}^{m} 2 \left(\sum_{j=1}^{m} a_j \, \mathrm{relu}(w_j x_i + b_j) - f(x_i) \right) \frac{\partial}{\partial a_k} \left(\sum_{j=1}^{m} a_j \, \mathrm{relu}(w_j x_i + b_j) \right)$

$\frac{\partial}{\partial a_k} F(\vec{a}, \vec{w}, \vec{b}) = \sum_{i=1}^{m} \left(\sum_{j=1}^{m} A_{ij} a_j - f_i \right) \underbrace{\mathrm{relu}(w_k x_i + b_k)}_{A_{ik}}$

$= \sum_{i=1}^{m} A_{ik} \left(A\vec{a} - \vec{f} \right)_i$

$= \left(A^T (A\vec{a} - \vec{f}) \right)_k$

Ici $A^T (A\vec{a} - \vec{f}) = \begin{pmatrix} \cdots & A_1 & \cdots \\ & \vdots & \\ \cdots & A_m & \cdots \end{pmatrix} \begin{pmatrix} | & & | \\ A_1 & \cdots & A_m \\ | & & | \end{pmatrix} \begin{pmatrix} \\ \vec{a} - \vec{f} \\ \\ \end{pmatrix}$

donc $\frac{\partial}{\partial a_k} F(\vec{a}, \vec{w}, \vec{b}) = A^T (A\vec{a} - \vec{f})$

A dépend de $\vec{w}$ et $\vec{b}$

Ça mandra carrés : $\min_{\vec{a}} \|A\vec{a} - \vec{f}\|^2$ $A^T (A\vec{a}^* - \vec{f}) = \vec{0}$

$\frac{\partial}{\partial a_k} \vec{F}(\vec{a}, \vec{w}, \vec{b}) = \sum_{i=1}^{m} \left(\sum_{j=1}^{m} a_j \, \mathrm{relu}(w_j x_i + b_j) - f(x_i) \right) \frac{\partial}{\partial a_k} \left(\sum_{j=1}^{m} a_j \, \mathrm{relu}(w_j x_i + b_j) \right)$

$\frac{\partial}{\partial a_k} F(\vec{a}, \vec{w}, \vec{b}) = \sum_{i=1}^{m} \left(\sum_{j=1}^{m} a_j \, \mathrm{relu}(w_j x_i + b_j) - f(x_i) \right) \underbrace{a_k \, \mathrm{relu}'(w_k x_i + b_k) x_i}_{B_{ik}}$

$= \sum_{i=1}^{m} B_{ik} \left(A\vec{a} - \vec{f} \right)_i$

$= \left(B^T (A\vec{a} - \vec{f}) \right)_k$ donc $\frac{\partial}{\partial b} F(\vec{a}, \vec{w}, \vec{b}) = B^T (A\vec{a} - \vec{f})$

B dépend de $\vec{a}, \vec{w}$ et $\vec{b}$

De même

$\frac{\partial}{\partial b_k} F(\vec{a}, \vec{w}, \vec{b}) = \sum_{i=1}^{m} \left(\sum_{j=1}^{m} a_j A_{ij} - f_i \right) \underbrace{a_k \, \mathrm{relu}'(w_k x_i + b_k) 1}_{C_{ij}}$

$\frac{\partial}{\partial b} F(\vec{a}, \vec{w}, \vec{b}) = C^T (A\vec{a} - \vec{f})$

2 questions mathématiques

- exponentiel : est ce que $\sum_{j=1}^{m} a_j^* \, \mathrm{relu}(w_j^* x + b_j^*)$ approche $f(x)$?
  question résolue

- optimisation : comment trouver $\vec{a}^*, \vec{w}^*, \vec{b}^*$?
  On utilise la méthode du gradient $k = 0, 1, 2, \dots$

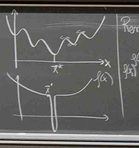
$\vec{a}^{k+1} = \vec{a}^k - \gamma \frac{\partial}{\partial a} F(\vec{a}^k, \vec{w}^k, \vec{b}^k)$

$\vec{w}^{k+1} = \vec{w}^k - \gamma \frac{\partial}{\partial w} F(\text{———})$

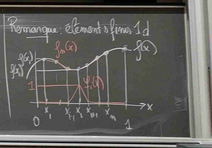
$\vec{b}^{k+1} = \vec{b}^k - \gamma \frac{\partial}{\partial b} F(\text{———})$

où $\gamma > 0$ est à choisir judicieusement
Cette méthode converge lentement
pour des γ petits. Peu de théorie
En pratique on utilise des variantes
(gradient stochastique)

$\min_{x \in \mathbb{R}} f(x)$



Remarque : éléments finis 1d



On a :

$f_m(x) = \sum_{j=1}^{m} f(x_j) \varphi_j(x)$ interpolant de degré 1 de f

En effet $f_m(x)$ cont, polyn de degré 1 sur
Chaque intervalle

de plus

$f_m(x_i) = \sum_{j=1}^{m} f(x_j) \underbrace{\varphi_j(x_i)}_{= 1 \text{ si } j = i \\ = 0 \text{ sinon}} = f(x_i) \underbrace{\varphi_i(x_i)}_{= 1}$ $i = 1, \dots, m$

On peut faire cette construction si $f : \mathbb{R}^2 \to \mathbb{R}$
$(x_1, x_2) \mapsto f(x_1, x_2)$

mais pas si $f : \mathbb{R}^{1000} \to \mathbb{R}$ où $f : \mathbb{R}^2 \to \mathbb{R}$

$x \in \mathbb{R}^2$

5 points dans chaque direction
$5^2 = 25$ points
$5^{100} = 7{,}9 \cdot 10^{69}$
pas cont ni on peut construire
$\sum_{j=1}^{m} a_j \, \mathrm{relu}(w_j x + b_j)$ pour
$x \in \mathbb{R}^{100}$

Neural Networks

- 2 layers NN (1960)

On veut approcher $f : [0,1] \to \mathbb{R}$

$\mathrm{relu}(x) = \max(x, 0)$

On approche $f(x)$ par $\sum_{j=1}^{n} a_j \, \mathrm{relu}(w_j x + b_j)$

On se donne des points x_i, $i = 1, \dots, m$, $0 \le x_i \le 1$

et on veut $\min_{\vec{a},\vec{w},\vec{b}} \frac{1}{2} \sum_{i=1}^{m} \underbrace{\left(\sum_{j=1}^{n} a_j \, \mathrm{relu}(w_j x_i + b_j) - f(x_i) \right)^2}_{F(\vec{a},\vec{w},\vec{b})}$

On cherche $\vec{a}^*, \vec{w}^*, \vec{b}^* \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n$ tels que

$F(\vec{a}^*, \vec{w}^*, \vec{b}^*) \le F(\vec{a}, \vec{w}, \vec{b}) \quad \forall \vec{a}, \vec{w}, \vec{b} \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n$

$\vec{\nabla} F(\vec{a}^*, \vec{w}^*, \vec{b}^*) = \vec{0}$

$\frac{\partial}{\partial a} F(\vec{a}^*, \vec{w}^*, \vec{b}^*) = 0$

$\frac{\partial}{\partial w} \quad = 0$

$\frac{\partial}{\partial b} \quad = 0$

Méthode du gradient

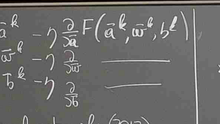
$\vec{a}^{k+1} = \vec{a}^k - \eta \frac{\partial}{\partial \vec{a}} F(\vec{a}^k, \vec{w}^k, \vec{b}^k)$

$\vec{w}^{k+1} = \vec{w}^k - \eta \frac{\partial}{\partial \vec{w}} \quad$

$\vec{b}^{k+1} = \vec{b}^k - \eta \frac{\partial}{\partial \vec{b}} \quad$

- deep neural networks (2012)

$\sum_{j=1}^{n} a_j \, \mathrm{relu}(w_j x + b_j)$



fully connected feed forward NN

L layers

$L-1$ hidden layers

approx de $f(x)$

Pour simplifier la prés., n neurones par couche

$x_1 = \mathrm{relu}(W_1 x + b_1)$

$x_2 = \mathrm{relu}(W_2 x_1 + b_2)$

$\vdots$

$x_{L-1} = \mathrm{relu}(W_{L-1} x_{L-2} + b_{L-1})$

$x_L = W_L x_{L-1} + b_{L-1}$

$L=2$: 1 couche cachée

$L=3$: 2 couches cachées

Rappel : éléments finis

$f(x) \simeq \sum_{i=1}^{n} f(x_i) \varphi_i(x)$

$\varphi_i(x) = \frac{1}{h} \mathrm{relu}(x - x_{i-1}) - \frac{2}{h} \mathrm{relu}(x - x_i) + \frac{1}{h} \mathrm{relu}(x - x_{i+1})$

$\bar{x} \in [0,1]^2$

$\mathrm{relu}(\bar{x} \cdot \bar{a}_0) - \mathrm{cste}$

$\mathrm{relu}(x_1) = \mathrm{cste}$

$\mathrm{relu}(x_2)$

$\mathrm{relu}(\bar{x} \cdot \bar{w}_0)$

$L=2$: 1 couche cachée

$L=3$: 2 couches cachées

Rappel : éléments finis

$f(x) \simeq \sum_{i=1}^{n} f(x_i) \varphi_i(x)$

$\varphi_i(x) = \frac{1}{h} \mathrm{relu}(x - x_{i-1}) - \frac{2}{h} \mathrm{relu}(x - x_i) + \frac{1}{h} \mathrm{relu}(x - x_{i+1})$

$\bar{x} \in [0,1]^2$

$\mathrm{relu}(\bar{x} \cdot \bar{a}_0) - \mathrm{cste}$

$\mathrm{relu}(x_1) = \mathrm{cste}$

$\mathrm{relu}(x_2)$

$\mathrm{relu}(\bar{x} \cdot \bar{w}_0)$

approx $\sum_{j=1}^{n} a_j \, \mathrm{relu}(\bar{x} \cdot \bar{w}_j)$

cont, polyn. degré 1 dans chaque polygone

$\mathrm{relu}(x - \bar{x}_i) \cdot \bar{w}_i$

Question : Comment trouver les $W_1, b_1, \dots, W_L, b_L$?

On note $x_L = f_{\mathrm{deep}}(x, \theta)$ où $\theta = (W_1, b_1, \dots, W_L, b_L)$ les poids

On se donne des points $\tilde{x}_i$, $i = 1, \dots, m$ (training set)

On cherche θ^* tq $F(\theta^*) \le F(\theta) \ \forall \theta$ où

$F(\theta) = \frac{1}{2} \sum_{i=1}^{m} \left(f_{\mathrm{deep}}(\tilde{x}_i, \theta) - f(\tilde{x}_i) \right)^2$

On cherche θ^* tq $\vec{\nabla} F(\theta^*) = \vec{0}$

Alg du gradient

$\theta^{k+1} = \theta^k - \eta \nabla F(\theta^k)$

Calcul de $\nabla F(\theta)$? backpropagation

Quel η ?

Google (keras)

Facebook (pytorch)

réseaux ce prm très efficacement

Exp 1 : $f : [0,1] \to \mathbb{R}$ par $\sum_{j=1}^{n} a_j \, \mathrm{relu}(w_j x + b_j)$ approche

x_i, $i = 1, \dots, m$ points d'interpolation

two layers norm. m

param n, m, η, nbre d'iter gradient

Exp 2 : $f : [0,1]^{10} \to \mathbb{R}$ qui m'approche par

$f_{\mathrm{deep}}(x, \theta)$

deepnn.py

$f(x_1, x_2, \dots, x_{10}) = g(x_1 + x_2 + \dots + x_{10})$

$\tilde{x}_j$, $j = 1, \dots, 10000$ points dans $[0,1]^{10}$

$\min \sum_{j=1}^{10000} \left(f_{\mathrm{deep}}(\tilde{x}_j, \theta) - f(\tilde{x}_j) \right)^2$

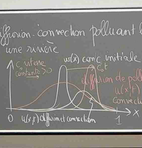
$f(x_1 + \dots + x_{10})$

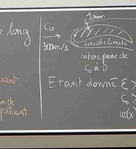
$x_1 + \dots + x_{10}$

Chap 14 : diffusion-convection transport

Chap 12 : eq chaleur (diffusion) 1-D, eq transport

Exp : diffusion-convection polluant le long d'une rivière



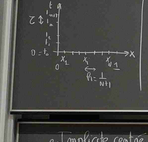


Etant donnée $\begin{cases} c \geqslant 0 \text{ (coefficient)} \\ v \geqslant 0 \text{ (vitesse)} \\ u(x) \text{ (cond. initiale)} \end{cases}$

On cherche $u(x,t)$ telle que

$$\begin{cases} \frac{\partial u}{\partial t}(x,t) - c\frac{\partial^2 u}{\partial x^2}(x,t) + v\frac{\partial u}{\partial x}(x,t) = 0 & \text{pour } x \in]-\infty,+\infty[\text{ et } t > 0 \\ u(x,0) = u(x) & \text{pour } x \in]-\infty,+\infty[\text{ cond. initiale} \\ u(0,t) = 0 \quad u(1,t) = 0 & t > 0 \text{ cond. limites} \end{cases}$$

2 schémas : - explicite centré (nommé $\frac{\partial u}{\partial x}$)
- implicite décentré (—— $\frac{\partial u}{\partial x}$)



$x_i = ih$ $i = 0, 1, \dots, N+1$

$t_n = n\tau$ $n = 0, 1, 2, \dots$

On note u_i^n une approx de $u(x_i, t_n)$

Manière entemps. $u_i^0 = u(x_i)$ $i = 1, \dots, N$

Etant donnée u_i^n $i = 1, \dots, N$ comment calc u_i^{n+1} ?

◦ Explicite centré

On écrit l'eq aux part x_i, temps t_n

- On utilise une bonn diff finie progressive pour $\frac{\partial u}{\partial t}(x_i, t_n)$
- ——— centrée pour $\frac{\partial u}{\partial x}(x_i, t_n)$
- ——— décentrée pour $\frac{\partial u}{\partial x}(x_i, t_n)$

schéma :

$$\begin{cases} \frac{u_i^{n+1} - u_i^n}{\tau} - c\frac{u_{i+1}^n - 2u_i^n + u_{i-1}^n}{h^2} + v\frac{u_{i+1}^n - u_{i-1}^n}{2h} = 0 & i = 1, \dots, N \\ u_0^{n+1} = 0 \quad u_{N+1}^{n+1} = 0 \end{cases}$$

schéma explicite ($u_i^{n+1} = \dots$)

stable sous condition entre h et τ

◦ Implicite centré

On écrit l'eq aux part x_i, temps t_{n+1}

- On utilise une FDF rétrograde pour $\frac{\partial u}{\partial t}(x_i, t_{n+1})$
- ——— centrée pour $\frac{\partial u}{\partial x}(x_i, t_{n+1})$
- ——— décentrée pour $\frac{\partial u}{\partial x}(x_i, t_{n+1})$

schéma :

$$\begin{cases} \frac{u_i^{n+1} - u_i^n}{\tau} - c\frac{u_{i+1}^{n+1} - 2u_i^{n+1} + u_{i-1}^{n+1}}{h^2} + v\frac{u_{i+1}^{n+1} - u_{i-1}^{n+1}}{2h} = 0 & i = 1, \dots, N \\ u_0^{n+1} = 0 \quad u_{N+1}^{n+1} = 0 \end{cases}$$

Il s'écrit de manière matricielle sous la forme linéaire

$$\begin{pmatrix} 1+2\frac{c\tau}{h^2} & -\frac{c\tau}{h^2} - \frac{v\tau}{2h} & & 0 \\ -\frac{c\tau}{h^2} + \frac{v\tau}{2h} & \ddots & \ddots & \\ & \ddots & \ddots & -\frac{c\tau}{h^2} - \frac{v\tau}{2h} \\ 0 & & -\frac{c\tau}{h^2} + \frac{v\tau}{2h} & 1+2\frac{c\tau}{h^2} \end{pmatrix} \begin{pmatrix} u_1^{n+1} \\ \vdots \\ \vdots \\ u_N^{n+1} \end{pmatrix} = \begin{pmatrix} u_1^n \\ \vdots \\ \vdots \\ u_N^n \end{pmatrix}$$

(a) (b)

pas de limitation sur $h > 0$ et $\tau > 0$