Méthodes itératives pour la résolution d'équations

On souhaite résoudre une équation du type $F(x)=0$ où $x \in \mathbb{R}^r$ et F une fonction donnée. On se donne une méthode dépendant de n pour avoir une solution x_n proche de x_0 lorsque n croît.

Le principe général consiste à rechercher à travers une fonction f donnée les points fixe de f :

$$F(x) \iff f(x) = x$$

On définit ensuite $x_{n+1} = f(x_n)$ et, lorsque f est continue, si la suite converge vers x, $f(x) = x$. La notion de vitesse de convergence donnera notamment la précision de la méthode.

I) Théorème du point fixe.

Théorème : Soit $f : E \longrightarrow E$ une application k-contractante de E dans E où (E, d) est un espace métrique complet :

$$\forall x, y \in E, \; d(f(x), f(y)) \leq k\, d(f(x), f(y)).$$

On suppose que $k < 1$. Alors, pour tout $x_0 \in E$, la suite définie par :

$$\begin{cases} x_0 \in E \\ x_{n+1} = f(x_n) \end{cases}$$

est convergente vers l'unique point fixe de f et, si on note a ce point fixe :

$$d(x_n, a) \leq \frac{k^n}{1-k}\, d(x_0, x_1).$$

Démonstration : Soit $x_0 \in E$. On a, pour $n \geq 1$:

$$d(x_n, x_{n+1}) \leq d\big(f(x_n), f(x_{n-1})\big) \leq k\, d(x_{n-1}, x_n).$$

puis, par récurrence $d(x_n, x_{n+1}) \leq k^n d(x_0, x_1)$.

Soient $1 \leq p \leq q$, par l'inégalité triangulaire, on a :

$$d(x_p, x_q) \leq d(x_p, x_{p+1}) + \cdots + d(x_{q-1}, x_q)$$

$$\leq \sum_{l=p}^{q-1} d(x_l, x_{l+1}) \leq \sum_{l=p}^{q-1} k^l\, d(x_0, x_1)$$

Or $\displaystyle\sum_{l=p}^{q-1} k^l \leq \sum_{l=p}^{+\infty} k^l \leq \frac{k^p}{1-k}$ d'où :

$$d(x_p, x_q) \leq \frac{k^p}{1-k}\, d(x_0, x_1) \quad (*)$$

La suite x_n est de Cauchy et converge donc, par complétude de E, vers $x \in E$ et, f étant k-contractant, elle est lipschitzienne donc continue et on a, en passant à la limite dans $x_{n+1} = f(x_n)$, $f(x) = x$.

Supposons qu'un certain $\tilde{x} \in E$ soit un point fixe :

$$d(x, \tilde{x}) = d\big(f(x), f(\tilde{x})\big) \leq k\, d(x, \tilde{x})$$

ce qui implique $d(x, \tilde{x}) = 0$ puisque $k < 1$.

En passant à la limite sur $q \longrightarrow +\infty$ dans (*), on obtient :

$$d(x_n, x) \leq \frac{k^n}{1-k} d(x_0, x_1).$$

Remarque : On a également :

$$d(x_n, a) \leq k^n d(x_0, a).$$

II) Méthodes pour les fonctions de $\mathbb{R}$ dans $\mathbb{R}$.

1) Attractivité, répulsivité.

Définition : Soit I un intervalle fermé de $\mathbb{R}$ et $f : I \longrightarrow I$ une application de classe $\mathcal{C}^1$ dont a est un point fixe.
On dit que a est un point :

(i) attractif si $|f'(a)| < 1$; (cf. Fig 1)
(ii) répulsif si $|f'(a)| > 1$. (cf. Fig 2)

Propriété : Si a est attractif (resp. répulsif), il existe $h \in \mathbb{R}$ tel que :

$$\forall x_0 \in [a-h, a+h] \cap I, \lim_{n \to +\infty} x_n = a$$

(resp. $\forall x_0 \in (a-h, a+h] \cap I \setminus \{a\}, \forall n \ |x_n - a| > |x - a|$).

Démonstration : Comme f est $\mathcal{C}^1$, f' est continue donc il existe h tel que :

$$k := \sup_{x \in [a-h, a+h]} |f'(x)| < 1$$

Par l'inégalité des accroissements finis,

$\forall x, y \in [a-h, a+h] \cap I$, $|f(x)-f(y)| \leq k|x-y|$

et on peut appliquer le théorème du point fixe à $f_{|[a-h, a+h]\cap I}$.

Si $|f'(a)| > 1$, à nouveau il existe $h > 0$ tel que :

$\inf_{x \in [a-h, a+h]\cap I} |f'(x)| > 1$ donc

$\forall x \in [a-h, a+h] \cap I \setminus \{a\}$, $|f(x) - f(a)| > |x-a|$.

puis $|x_n - \underbrace{a}_{=f(a)}| > |x-a|$.

Remarque : Si a est répulsif, comme $(f^{-1})' = \frac{1}{f'}$, on peut transformer l'équation en $f^{-1}(x) = x$ et transformer a en point attractif pour $\tilde{f} = f^{-1}$.

Remarque : Si f est $\mathcal{C}^2$ et que $f'(a) = 0$, on a :

$\exists c \in [x, a]$ tel que :

$$f(x) - f(a) = (x-a)\underbrace{f'(a)}_{=0} + \frac{(x-a)^2}{2}\varphi''(c)$$

par l'égalité des accroissements finis.

Ainsi $|f(x) - \underbrace{a}_{=f(a)}| \leq \frac{1}{2} \underbrace{\sup_{x \in I} |\varphi''(x)|}_{=M>0} |x-a|^2$ puis

par récurrence :

$$|x_n - a| \leq \frac{2}{M}\left[\frac{M}{2}|x_0 - a|\right]^{2^n}.$$

On choisissant $|x_0 - a| < \frac{2}{\pi}$ par exemple, on aura convergence.

Remarque: Cas où $f'(a) = 1$. Toutes les situations sont possibles.

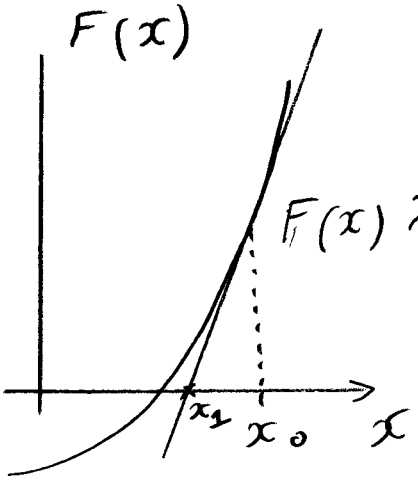
(cf. Fig 3)(i) $f(x) = \sin(x)$, $x \in [0, \frac{\pi}{2}]$; On a : $\sin(x) < x$ pour $x \in]0, \frac{\pi}{2}]$. La suite x_n est décroissante et minorée puis vérifie $\lim_{n \to +\infty} x_n = a$ avec $\sin(a) = a$ puis $a = 0$ avec $|\sin'(a)| = 1$

On a même : $x_n \underset{n \to +\infty}{\sim} \sqrt{\frac{3}{n}}$.

(cf. Fig 4)(ii) $f(x) = \mathrm{sh}(x)$ avec $x \in [0, +\infty[$. On a $\mathrm{sh}(x) > x$ pour tout x. Or $\mathrm{sh}(x)$ admet pour point fixe $a = 0$ et $|\mathrm{sh}'(a)| = 1$. Soustant la suite $x_0 > 0$, $x_{n+1} = \mathrm{sh}(x_n)$ est strictement croissante.

2) Méthode de Newton.

On va remplacer, en partant $x_0 \simeq x$ avec $F(x) = 0$, le remplacer localement F par sa tangente.



La fonction à itérer sera:

$$f(x) = x - \frac{F(x)}{F'(x)}.$$

Un point fixe a de f vérifiera $F(a) = 0$.

Théorème: Soit F une fonction de classe $\mathcal{C}^2$ s'annulant en a. On suppose qu'il existe $r > 0$ tel que $f' \neq 0$ sur $I = [a-r, a+r]$. On pose $M = \sup_{x \in I} \left| \frac{F''(x)}{F'(x)} \right|$ et $h = \min\left(r, \frac{1}{M}\right)$. Alors pour tout $x \in [a-h, a+h]$, $|f(x) - a| \leq M |x-a|^2$ et:

$$\forall x_0 \in [a-h, a+h], \ \forall n \in \mathbb{N}, \ |x_n - a| \leq \frac{1}{M} \left(M |x_0 - a|\right)^{2^n}.$$

Démonstration: Comme F est de classe $\mathcal{C}^2$ et que F' ne s'annule pas sur I, F' garde un signe constant sur I. On supposera sans perte de généralité (sinon on applique le théorème à $-F$) que $F' > 0$ sur I. La fonction F, par le théorème de Rolle, ne s'annule donc qu'en a sur I et est du signe de $(x-a)$.

Posons $u(x) = \frac{F(x)}{F'(x)}$ (qui est également du signe de $x-a$). On a, pour $x \in I$:

$$u'(x) = 1 - \frac{F(x)\,F''(x)}{(F'(x))^2} = 1 - \frac{F''(x)}{F'(x)}\,u(x) \quad \text{ce qui}$$

donne :

$$|u'(x)| \leq 1 + M|u(x)|.$$

On majore la dérivée de u par u il faut donc

Lemme de Gronwall :

$$|u'(x)| \leq 1 + M|u(x)| \text{ donne}$$

$$|u(x)| \leq \frac{1}{M}\left(e^{M|x-a|} - 1\right).$$

Démonstration : Supposons $x \geq a$, l'autre cas se traitant de la même manière.

On pose $v(x) = u(x)\, e^{-Mx}$. On a :

$$v'(x) = \left(u'(x) - M\underbrace{u(x)}_{=|u(x)| \text{ puisque } u \text{ est du signe de } x-a.}\right) e^{-Mx}$$

donc $v'(x) \leq e^{-Mx}$ ce qui donne en intégrant :

$\forall\, x \geq a$:

$$v(x) - v(a) \leq \frac{1}{M}\left(e^{-Ma} - e^{-Mx}\right).$$

Or $v(a) = \underbrace{u(a)}_{=0}\, e^{-Ma} = 0$ donc :

$$u(x)\, e^{-Mx} \leq \frac{1}{M}\left(e^{-Ma} - e^{-Mx}\right)$$

ce qui donne puis $x - a = |x-a|$ et $u(x) \geq 0$,

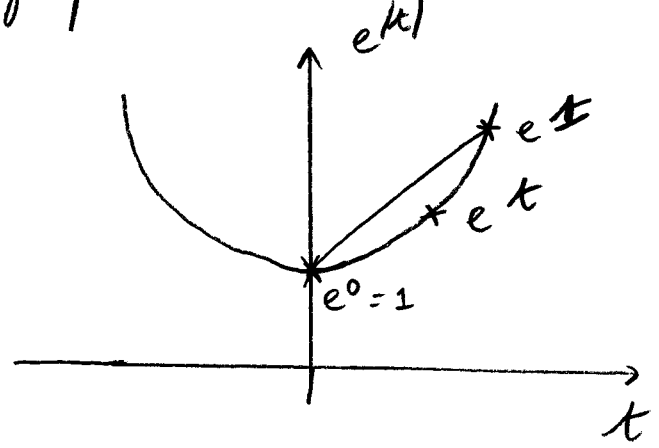
$$|u(x)| \leq \frac{1}{M}\left(e^{M|x-a|} - 1\right).$$

Lemme 2 :

$$\forall\, |t| \leq 1 \qquad e^{|t|} - 1 \leq 2|t|$$

Démonstration :

On peut étudier $t \mapsto e^{|t|} - 1 - 2|t|$ pour $t \geq 0$ et $t \leq 0$ ou remarquer que exp est convexe donc le graphe de $t \mapsto e^t$ est situé sous sa corde :



Pour $t \in [0,1]$, par exemple :

$$e^t \leq 1 + \underbrace{(e-1)}_{<2} t \leq 1 + 2t$$

On peut désormais conclure :

$$f'(x) = 1 - \left(1 - \frac{F(x)F''(x)}{(F'(x))^2}\right) = u(x)\frac{f''(x)}{F'(x)}.$$

donc les deux lemmes nous donnent, pour $|x-a| \leq h$ $(h = \min(r, \frac{1}{F}))$

$$|f'(x)| \leq M\left(\frac{1}{M} e^{M|x-a|} - 1\right) \leq 2M|x-a|$$

puis en intégrant, en distinguant $x \geq a$ et $x \leq a$:

$$|f(x) - \underbrace{f(a)}_{=a}| \leq M|x-a|^2.$$

Par récurrence, puisque cette inégalité donne $M|x_{n+1} - a| \leq M^2|x_n - a|^2$, on a :

$$|x_n - a| \leq \frac{1}{M}\left(M|x_0 - a|\right)^{2^n}.$$

3) Cas des polynômes réels de la variable réelle.

Considérons $P(x) = a_0 + a_1 x + \cdots + a_d x^d$ un polynôme à coefficients réels de de degré d. On recherche les racines de ce polynôme. L'équation $P(x) = 0$ ayant au plus d racines, la méthode de Newton précédemment étudiée ne s'appliquera pas directement puisqu'il pourrait exister plusieurs points fixes pour une éventuelle fonction f. On commence par chercher des intervalles où il n'y aura qu'une seule racine. La méthode de Sturm permet notamment de compter le nombre de racines dans un intervalle $]\alpha, \beta[$ donné. Lorsqu'on choisit bien α, β pour avoir une seule racine, on pourra appliquer la méthode de Newton. Commençons par un lemme réduisant la recherche des racines à un intervalle borné :

Lemme : Si λ est une racine de P, alors :

$$|\lambda| \leq 1 + \max_{0 \leq k \leq d-1} \frac{|a_k|}{|a_d|}.$$

Démonstration : Si $|\lambda| \leq 1$, on a l'inégalité. Supposons $|\lambda| > 1$.

On a :

$$P(\lambda) = 0 \Rightarrow \lambda^d = -\sum_{k=0}^{d-1} \frac{a_k}{a_d} \lambda^k \text{ puis,}$$

$$|\lambda|^d \leq \sum_{k=0}^{d-1} |\lambda|^k \frac{|a_k|}{|a_d|} \leq \left(\max_{0 \leq k \leq d-1} \frac{|a_k|}{|a_d|} \right) \frac{|\lambda|^d - 1}{|\lambda| - 1}$$

et finalement l'inégalité puisque :

$$|\lambda| - 1 \leq \underbrace{\frac{|\lambda|^d - 1}{|\lambda|^d}}_{\leq 1 \text{ puisque } |\lambda|^d > 1} \max_{0 \leq k \leq d-1} \frac{|a_k|}{|a_d|}.$$

Les racines sont donc dans un intervalle $[-M, M]$ où

$$M := 1 + \max_{0 \leq k \leq d-1} \frac{|a_k|}{|a_d|}.$$

a) Méthode de Sturm.

Théorème de Sturm :

Soit $P \in \mathbb{R}[X]$. On considère la suite $(S_i)_{1 \leq i \leq p+1}$ définie par $S_0 = P$, $S_1 = -P'$ et pour $i \in [\![1, p]\!]$, S_{i+1} est le reste dans la division euclidienne de S_{i-1} par S_i :

$$\begin{cases} \exists ! \ Q_i \in \mathbb{R}[X], \ S_{i-1} = Q_i S_i - S_{i+1} \\ \deg(S_{i+1}) < \deg(S_i). \end{cases}$$

On suppose que $S_{p+1} = 0$. Pour $x \in \mathbb{R}$, on note :

$$V(x) = \text{Card}\left(\{(i,j) ; 0 \leq i < j \leq p ; S_i(x) S_j(x) < 0 \text{ et } S_k(x) = 0 \text{ si } i < k < j\}\right)$$

la fonction mesurant le nombre de changements de signe de la suite $(S_i(x))_{0 \leq i \leq p}$ (0 n'étant pas un changement de signe).

Alors, pour $a < b$, le nombre de racines réelles distinctes dans $[a, b]$ de P est $V(b) - V(a)$.

Remarque :

La suite $(S_i)_{1 \leq i \leq p+1}$ est bien définie par l'algorithme d'Euclide. Seul l'entier p peut varier. On a $p+1 \leq d+1$ par la condition $\deg(S_{i+1}) < \deg(S_i)$.

Démonstration :

a) On se ramène au cas où les racines sont simples.

On a : $\text{pgcd}(S_0, S_1) = \text{pgcd}(q_1 S_1 - S_2, S_1) = \text{pgcd}(S_2, S_1) = \dots = \text{pgcd}(S_i, S_{i+1})$ pour $0 \leq i \leq p$. Or $S_{p+1} = 0$ donc :

$$\text{pgcd}(S_i, S_{i+1}) = S_p .$$

Or $S_0 = P$, $S_1 = P'$ donc $S_p = \text{pgcd}(P, P')$ également.

Posons $T_i = \dfrac{S_i}{S_p}$. On a $T_0 = \dfrac{P}{\text{pgcd}(P, P')}$ qui est à racines simples car sinon, par la relation de Bézout, pour $\tilde{x}$ une racine de P et P' :

$$1 = U(\tilde{x}) \frac{P}{\text{pgcd}(P,P')}(\tilde{x}) + V(\tilde{x}) \frac{P'}{\text{pgcd}(P,P')}(\tilde{x}) = 0 .$$

La suite T_i vérifie également pour $i \in [\![0, p]\!]$:

$$T_{i-1} = q_i T_i - T_{i+1} .$$

Or, pour tout x, le nombre de changements de signe dans la suite $(T_i(x))$ est le même que pour la suite $(S_i(x))$ puisque $T_i(x) = \dfrac{S_i(x)}{S_p(x)}$ et $S_p(x)$ est constant, le cas $S_p(x) = 0$ étant évident car dans ce cas $S_i(x) = 0$ pour tout i car $S_p \mid S_i$.

On se ramène donc au cas où les racines de P sont simples. Supposons que P est à racines simples.
Notons $\mathcal{R}_0$ l'ensemble des racines de $S_0, \dots, S_p$. Cet ensemble est fini.

b) Etude de la fonction V sur $\mathbb{R} \setminus \mathcal{R}_0$ puis sur $\mathcal{R}_0$.

Soit $]\alpha, \beta[$ un intervalle de $\mathbb{R}$ ne contenant aucun point de $\mathcal{R}_0$. Les signes des S_i sont constants sur $]\alpha, \beta[$, donc V est constante sur $]\alpha, \beta[$. V est une fonction en escalier sur $\mathbb{R} \setminus \mathcal{R}_0$.

Etudions son comportement en un point $x \in \mathcal{R}_0$.

(i) x est racine de S_i pour $i \in [\![1, p]\!]$.

Comme P est à racines simple $S_p = pgcd(P, P') = 1$ donc x est racine S_i pour $i \in [\![1, p-1]\!]$. Comme $pgcd(S_{i-1}, S_i) = pgcd(S_i, S_{i+1}) = S_p = 1$, $S_{i-1}(x), S_{i+1}(x) \neq 0$ donc par la relation:

$$S_{i-1}(x) = q_i(x) \underbrace{S_i(x)}_{=0} - S_{i+1}(x),$$

$$S_{i-1}(x) S_{i+1}(x) = -(S_{i+1}(x))^2 < 0$$

Par continuité, il existe $\delta > 0$ tel que $S_{i-1} S_{i+1} < 0$ sur $[x-\delta, x+\delta]$.
Ainsi sur $[x-\delta, x+\delta]$, la suite

$$\big(S_0(x), S_1(x), \dots, S_{i-1}(x), S_i(x), S_{i+1}(x), \dots, S_p(x)\big)$$

α changements de signe | un changement de signe | β changements de signe

change le même nombre de fois de signe, c'est à dire que V est constante sur $[x-\delta, x+\delta]$.

(ii) x est racine de $S_0 = P$

Dans ce cas, on a les deux tableaux de signe suivants au voisinage de x (sur un intervalle où x est le seul point de R) :

	x		
$S_0 = P$	$+$	0	$-$
$-S_1 = P'$	$-$		$-$
S_1	$+$		$+$

ou

x		
$-$	0	$+$
$+$		$+$
$-$		$-$

donc, au voisinage de x on a deux situations :

1) $\tilde{x} < x$

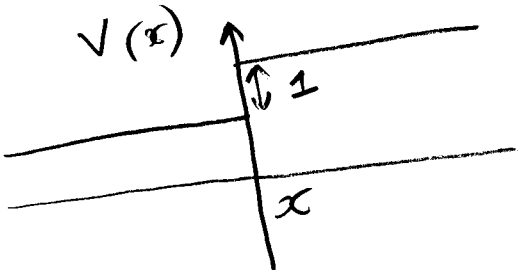
$(\underbrace{S_0, S_1}_{\text{pas de changement de signe}}, S_2, \dots, S_p)$

2) $\tilde{x} > x$

$(\underbrace{S_0, S_1}_{\text{changement de signe}}, S_2, \dots, S_p)$

Ainsi sur $[x-\delta, x+\delta]$ pour $[x-\delta, x+\delta] \cap R_0 = \{x\}$,

$V(\lambda) = V(\mu) + 1$ où $\lambda \in [x+\delta, x[$ et $\mu \in [x-\delta, x[$,



Finalement V croît de 1 si et seulement si elle traverse une racine de P.

En conclusion, le nombre de racines de P dans $[a, b]$ est donc $V(b) - V(a)$ pour $a < b$.

La méthode de Sturm permet donc, par exemple par dichotomie de connaître des intervalles ne contenant qu'une seule racine de P.
On va donc chercher à maintenant appliquer la méthode de Newton par exemple même si on pourrait continuer la dichotomie de la méthode de Sturm.
(cf. annexe pour une autre démonstration).

Exemple pour la méthode de Sturm

Pour $P(x) = x^4 - 1$, on obtient :

$S_0(x) = x^4 - 1$, $S_1(x) = -4x^3$, $S_2(x) = 1$.

La borne de l'introduction donne les racines dans $[-2, 2]$.

On a $P(-2) = S_0(-2) = 15$
$S_1(-2) = 32$
$S_2(-2) = 1$

donc $V(-2) = 0$.
De la même manière $V(2) = 2$ donc P possède bien que 2 racines réelles dans $[-2, 2]$.

b) Méthode de Newton pour les polynômes.

On propose une méthode de calcul de la plus grande racine de P.

DÉVELOPPEMENT 24

MÉTHODE DE NEWTON POUR LES POLYNÔMES

On considère un polynôme

$$P(x) = (x-\xi_1)^{m_1} \cdots (x-\xi_r)^{m_r}$$

où $\xi_1 < \cdots < \xi_r$ sont des réels et les m_i sont des entiers non nuls. On pose

$$x_{n+1} = x_n - \frac{P(x_n)}{P'(x_n)}.$$

Proposition. — *Si $x_0 > \xi_r$ alors la suite $(x_n)_{n\geq 0}$ décroît strictement et converge vers ξ_r.*

Démonstration. — Pour tout x, on a

$$\frac{P'(x)}{P(x)} = (\log|P(x)|)' = \left(\sum_{i=1}^{r} m_i \log|(x-\xi_i)|\right)' = \sum_{i=1}^{r} \frac{m_i}{x-\xi_i}$$

d'où pour tout $n \geq 0$

$$x_{n+1} = x_n - \left(\sum_{i=1}^{r} \frac{m_i}{x_n - \xi_i}\right)^{-1}.$$

En particulier, si $x_n > \xi_r$ alors $x_{n+1} < x_n$. La fonction f définie par $f(x) = x - \dfrac{P(x)}{P'(x)}$ se prolonge par continuité à $[\xi_r, +\infty[$ en posant $f(\xi_r) = \xi_r$ et sa dérivée est donnée par

$$f'(x) = 1 - \frac{(P'(x))^2 - P(x)P''(x)}{(P'(x))^2} = \frac{P(x)P''(x)}{(P'(x))^2}.$$

Mais d'après le théorème de Gauss-Lucas, les racines de P' sont dans l'enveloppe convexe de celles de P donc sont dans $[\xi_1, \xi_r]$ et de même pour les racines de P''. Puisque P est unitaire, P, P' et P'' sont strictement positifs sur $[\xi_r, +\infty[$ donc $f' > 0$ et il s'ensuit que f est strictement croissante. Le fait que $\xi_r < x_n$ implique donc que $\xi_r = f(\xi_r) < f(x_n) = x_{n+1}$. La condition $x_0 > \xi_r$ implique donc par récurrence que $x_n > \xi_r$ pour tout $n \geq 0$ et que la suite $(x_n)_{n\geq 0}$ décroît strictement. Enfin, comme cette suite est minorée par ξ_r, elle converge vers un élément qui annule $\frac{P}{P'}$ *i.e.* vers ξ_r. □

Proposition. — *Si $m_r = 1$ alors pour tout $c > 0$, on a $|x_n - \xi_r| = o(c^n)$.*

Démonstration. — Comme $\dfrac{P'(x)}{P(x)} = \sum_{i=1}^{r} \dfrac{m_i}{x-\xi_i}$, on a en dérivant

$$\frac{P''(x)(x)P(x) - (P'(x))^2}{(P(x))^2} = -\sum_{i=1}^{r} \frac{m_i}{(x-\xi_i)^2}$$

i.e.

$$\frac{P''(x)(x)P(x)}{(P(x))^2} = \frac{(P'(x))^2}{(P(x))^2} - \sum_{i=1}^{r} \frac{m_i}{(x-\xi_i)^2}$$

d'où

$$f'(x) = \frac{P(x)P''(x)}{(P'(x))^2} = \frac{P(x)P''(x)}{(P(x))^2}\frac{(P(x))^2}{(P'(x))^2} = 1 - \left(\sum_{i=1}^{r} \frac{m_i}{x-\xi_i}\right)^{-2} \sum_{i=1}^{r} \frac{m_i}{(x-\xi_i)^2}$$

d'où $\lim\limits_{x\to\xi_r} f'(x) = 1 - \dfrac{1}{m_r}$.

Si $m_r = 1$ alors $f'(\xi_r) = 0$ mais la formule de Taylor-Lagrange donne $y_n \in]\xi_r, x_n[$ tel que

$$x_{n+1} - \xi_r = f(x_n) - f(\xi_r) = (x_n - \xi_r)f'(y_n).$$

Soit $c > 0$, comme x_n tend vers ξ_r, il existe un rang n_0 à partir duquel $|f'(y_n)| < c$ d'où

$$|x_{n+1} - \xi_r| \leq c\,|x_n - \xi_r|$$

puis pour tout $n \geq n_0$

$$|x_n - \xi_r| \leq c^{n-n_0}\,|x_{n_0} - \xi_r| = O(c^n)$$

et quitte à prendre $0 < d < c$, il vient $|x_n - \xi_r| = o(d^n)$. □

Proposition. — *Si $m_r \geq 2$ alors il existe $c > 0$ tel que $|x_n - \xi_r| \sim c\left(1 - \frac{1}{m_r}\right)^n$.*

Démonstration. — Comme plus haut, on a $f'(\xi_r) = 1 - \frac{1}{m_r}$ et la formule de Taylor-Lagrange donne $y_n \in]\xi_r, x_n[$ tel que $x_{n+1} - \xi_r = (x_n - \xi_r)f'(y_n)$ d'où

$$\log(x_{n+1} - \xi_r) - \log(x_n - \xi_r) = \log f'(y_n)$$

qui converge vers $\log f'(\xi_r)$ quand n tend vers l'infini donc, d'après le théorème de Cesáro, $\log(x_n - \xi_r)$ est équivalent à $n \log f'(\xi_r)$ quand n tend vers l'infini ; en particulier, pour tout $1 > d > f'(\xi_r)$, on a $|x_n - \xi_r| = O(d^n)$. D'après la formule de Taylor-Lagrange, il existe $z_n \in]\xi_r, x_n[$ tel que

$$x_{n+1} - \xi_r = f'(\xi_r)(x_n - \xi_r) + \frac{f''(z_n)}{2}(x_n - \xi_r)^2$$

d'où

$$\varepsilon_n = \frac{x_{n+1} - \xi_r}{f'(\xi_r)(x_n - \xi_r)} - 1 = O(x_n - \xi_r)$$

et il s'ensuit que la série de terme général

$$\log(x_{n+1} - \xi_r) - \log(x_n - \xi_r) - \log f'(\xi_r) = \log(1 + \varepsilon_n) = O(d^n)$$

converge. Par conséquent, $\log(x_n - \xi_r) - n \log f'(\xi_r)$ converge vers un réel λ donc $x_n - \xi_r \simeq e^{\lambda} f'(\xi_r)^n$ quand n tend vers l'infini. □

Leçons concernées

15 Différentiabilité d'une fonction définie sur un ouvert de $\mathbb{R}^n$. Exemples et applications
18 Application des formules de Taylor et des développements limits
23 Convergence des suites numériques. Exemples et applications
24a Comportement asymptotique des suites numériques. Exemples
24b Rapidité de convergence d'un suite. Exemples
25 Comportement d'une suite réelle ou vectorielle définie par une itération $u_{n+1} = f(u_n)$. Exemples
27 Continuité et dérivabilité des fonctions réelles d'une variable réelle. Exemples et contre-exemples
28 Fonctions monotones. Fonctions convexes. Exemples et applications.
31 Méthodes d'approximation des solutions d'une quation $F(x) = 0$. Exemples

Sébastien Pellerin

Complément

Localisation des racines de P'. —
Le théorème de Gauss-Lucas affirme que les racines de P' sont dans (l'intérieur de) l'enveloppe convexe des racines (dans $\mathbb{C}$) de P. En effet, écrivons $P = (X-\xi_1)^{m_1}\cdots(X-\xi_r)^{m_r}$ alors

$$\frac{P'}{P} = \sum_{i=1}^{r} \frac{m_i}{X-\xi_i}.$$

Soit ζ une racine de P'. S'il s'agit aussi d'une racine de P alors le résultat est clair, sinon on a

$$\sum_{i=1}^{r} m_i \frac{\overline{\zeta}-\overline{\xi_i}}{|\zeta-\xi_i|^2} = \sum_{i=1}^{r} \frac{m_i}{\zeta-\xi_i} = 0$$

d'où

$$\sum_{i=1}^{r} m_i \frac{\zeta-\xi_i}{|\zeta-\xi_i|^2} = 0 \quad i.e. \quad \zeta \sum_{i=1}^{r} \frac{m_i}{|\zeta-\xi_i|^2} = \sum_{i=1}^{r} \frac{m_i}{|\zeta-\xi_i|^2}\xi_i.$$

Dans le cas où les racines sont toutes réelles, le résultat s'obtient aussi en remarquant que chaque ξ_i est une racine de P' de multiplicité $m_i - 1$ et en appliquant le théorème de Rolle sur chaque intervalle $]\xi_i, \xi_{i+1}[$, on obtient $r-1$ autres racines et on a bien $n-1$ racines, toutes comprises entre ξ_1 et ξ_r.

Pour trouver les autres racines. —
Tout d'abord, pour trouver un $x_0 > \xi_r$, on commence par écrire $P = x^n + a_1 x^{n-1} + \cdots + a_{n-1}$ alors, si $P(\xi) = 0$, on a $|\xi|^n = \left|\sum_{i=1}^{n} a_i \xi^{n-i}\right| \leq \sum_{i=1}^{n} |a_i|\,|\xi|^{n-i}$ et si $|\xi| \geq 1$, on a donc $|\xi| \leq \sum_{i=1}^{n} |a_i|$. D'où

$$|\xi_r| \leq \max\left(1; \sum_{i=1}^{n} |a_i|\right).$$

Par ailleurs, il est conseillé de diviser P par le pgcd de P et P' de sorte que ξ_r soit une racine simple ce qui donne une convergence plus rapide. Pour trouver les autres racines, on applique la méthode de Newton à $\frac{P(x)}{x-\xi_r}$ *i.e.* on considère la suite définie par

$$x_{n+1} = x_n - \frac{P(x_n)}{P'(x_n) - \frac{P(x_n)}{x_n-\xi_r}}.$$

Référence

A. Chambert-Loir et S. Fermigier, *Exercices d'analyse 2*, Dunod, 1999.

III) Cas des fonctions de $\mathbb{R}^d$ dans $\mathbb{R}^d$.

1) Méthode de Newton-Raphson.

Soit F une fonction de classe $\mathcal{C}^2$ de $\mathbb{R}^d$ dans $\mathbb{R}^d$ et on souhaite résoudre :

$$F(x) = 0 \quad \text{pour } x \in \mathbb{R}^d.$$

On modifie la méthode en une dimension pour ajouter à remplacer "f'" par DF la différentielle de F.

Théorème : Soit F une fonction de classe $\mathcal{C}^2 : \mathbb{R}^d \longrightarrow \mathbb{R}^d$. Soit a un zéro de F. On suppose que $DF(a)$ est inversible. Alors, il existe $\varepsilon > 0$ tel que pour toute condition initiale $x_0 \in B(a, \varepsilon)$, la suite :

$$x_{n+1} = x_n - \left(DF(x_n)\right)^{-1} \left(F(x_n)\right)$$

est bien définie et, il existe $C, \Pi > 0$ tels que :

$$\|x_n - a\| \leqslant \frac{1}{\Pi} \left[\Pi \|x_0 - a\|\right]^{2^n}.$$

Démonstration : (i) Comme $\det\left(DF(a)\right) \neq 0$ et que la fonction $x \longmapsto \det\left(DF(x)\right)$ est continue, il existe $r > 0$ tel que :

$$x \in B(a, r) \Rightarrow DF(x) \text{ inversible.}$$

(ii) Montrons que les zéros de F sont isolés. Supposons qu'il existe une suite de zéros de F notés x_n tant pour n :

$$u_n = \frac{x_n - a}{\|x_n - a\|} \quad \text{avec} \quad \lim_{n \to +\infty} x_n = a.$$

On a : $\|u_n\| = 1$ donc il existe une sous suite de u_n qui converge vers u avec $\|u\| = 1$:

$$\lim_{n \to +\infty} u_{\varphi(n)} = u \neq 0.$$

On a :

$$\underbrace{F(x_n)}_{=0} = F(a + (x_n - a)) = \underbrace{F(a)}_{=0} + DF(a)(x_n - a) + \underset{\|x_n - a\| \to 0}{o(\|x_n - a\|)}$$

donc : $DF(a)\left(\frac{x_n - a}{\|x_n - a\|}\right) = DF(a)(\vec{u}_n) + \underset{\|x_n - a\| \to 0}{o(1)}$ qui donne en passant à la sous suite $\varphi(n)$, par continuité de DF

$$DF(a)(u) = 0 \quad \text{avec } u \neq 0.$$

On a donc $\ker(DF(a)) \neq \{0\}$ ce qui est faux par hypothèse. Ainsi, il existe $r' < r$ tel que F ne s'annule qu'en a sur $B(a, r')$.

(iii) On a :

$$x_{n+1} - a = x_n - a - (DF(x_n))^{-1}(F(x_n))$$

$$= (DF(x_n))^{-1}\left[DF(x_n)(x_n - a) - F(x_n) + \underbrace{F(a)}_{=0}\right]$$

La fonction $y \mapsto (DF(y))^{-1}$ est continue sur $B(a, r')$ car (i) $\det u \mapsto u^{-1}$ est continue sur $GL_d(\mathbb{R})$. Ainsi, il existe $M > 0$ tel que :

$$\forall y \in B(a, r') \Rightarrow \|DF(y)^{-1}\| \leq M.$$

La formule de Taylor avec reste intégral nous donne pour tous $x, y \in B(a, r')$:

$$F(x) - F(y) - DF(y)(x - y) = \int_0^1 (1 - t) D^2 F(x + t(y - x))((y - x), (y - x))\, dt$$

Or, $z \mapsto D^2F(z)$ est continue sur $B(a,r')$ (qui est convexe) et pour tout z, $D^2F(z)$ est une forme bilinéaire continue. Ainsi,

$$\left\| D^2F\left(\underbrace{x + t(y-x)}_{\in B(a,r')}\right)(y-x, y-x)\right\| \leq \underbrace{\sup_{z \in B(a,r')} \|D^2F(z)\|}_{:= C'} \|y-x\|^2$$

ce qui donne :

$$\| DF(x_n)(x_n - a) + F(x_n) - F(a)\| \leq C' \|x_n - a\|^2.$$

Ainsi :

$$\|x_{n+1} - a\| \leq M'C' \|x_n - a\|^2.$$

Posons $M := M'C'$. On a : $\frac{1}{M}$

$$(M \|x_{n+1} - a\|) \leq (M\|x - a\|)^2$$ ce qui donne par récurrence immédiate :

$$(M\|x_n - a\|) \leq (M\|x_0 - a\|)^{2^n}$$ qui est le résultat.

Remarque : Ainsi, pour avoir la convergence, la condition :

$$\sup_{z \in B(a,r')} \|D^2F(z)\| \sup_{z \in B(a,r')} \|(DF(z))^{-1}\| \, \|x_0 - a\| < 1$$

est suffisante.

2) Méthode du gradient à pas optimal.

Soit A une matrice symétrique définie positive $p \times p$ et $b \in \mathbb{R}^p$.

On pose toujours :

$$f(x) = \frac{1}{2}(Ax, x) + (b, x).$$

La fonction f est $\mathcal{C}^\infty$ de $\mathbb{R}^p$ dans $\mathbb{R}$.

Notons $\lambda_1 \leq \dots \leq \lambda_p$ les valeurs propres de A.

On a, $P^T A P = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_p \end{pmatrix}$ pour une matrice

P orthogonale et :

$$f(x) = \frac{1}{2}(D(Px), Px) + (b, Px)$$
$$\geq \frac{1}{2} \lambda_1 \|Px\|^2 + (b, Px)$$

Comme P est orthogonale, $\|Px\|^2 = \|x\|^2$.

On a : $|(b, Px)| \leq \|b\| \, \|x\|$ donc, pour $\|x\|$ suffisamment grande, par exemple $\|x\| \geq C$

$$f(x) \geq \frac{1}{4} \lambda_1 \|x\|^2.$$

Ainsi, si on recherche le minimum de f, on constate que $f(0) = 0$ et pour $\|x\| \geq \sqrt{\frac{4}{\lambda_1} + C^2}$, $f(x) \geq 1$. On a donc :

$$\min_{x \in \mathbb{R}^n} f(x) = \min_{x \in B\left(0, C^2 + \sqrt{\frac{4}{\lambda_1}}\right)} f(x)$$

et $B\left(0, \sqrt{\frac{4}{\lambda_1}}\right)$ est un compact. f atteint ② son minimum sur ce compact qui est un point critique puisque f est $\mathcal{C}^\infty$:

$$\min_{x \in \mathbb{R}^n} f(x) = f(\overline{x})$$

avec $\nabla f(\overline{x}) = 0$.

On a : $\nabla f(x) = Ax + b$. Le minimum de f est donc l'unique solution du système linéaire :

$$\boxed{A\overline{x} + b = 0}.$$

<u>Théorème</u> : (méthode du gradient à pas optimal)

Soit $x_0 \in \mathbb{R}^n$ et :

$$\begin{cases} x_{n+1} = x_n + t_n d_n \\ \text{avec } d_n = -\nabla f(x_n) \text{ et } t_n = \dfrac{\|d_n\|^2}{(Ad_n, d_n)} \end{cases},$$

on a : $\|x_n - \overline{x}\| \leqslant \sqrt{\dfrac{2f(x_0) - f(\overline{x})}{\lambda_1}} \left(\dfrac{\lambda_n - \lambda_1}{\lambda_1 + \lambda_n}\right)^n$.

On commence par un lemme.

Lemme : Inégalité de Kantorovitch.

$$\min_{x \in \mathbb{R}^n} \frac{\|x\|^4}{(Ax, x)(A^{-1}x, x)} = \frac{4\lambda_1 \lambda_n}{(\lambda_1 + \lambda_n)^2}$$

Démonstration :

On a : $\frac{(Ax, x)}{\|x\|^2} = A\left(\frac{x}{\|x\|}, \frac{x}{\|x\|}\right)$ donc on peut traiter l'inégalité précédente avec $\|x\| = 1$.

On prend ensuite $y = P^T x$, on a $\|y\| = \|Px\| = \|x\|$ car P est orthogonale puis :

$$\min_{x \in \mathbb{R}^n} \frac{\|x\|^4}{(Ax, x)(A^{-1}x, x)} = \max_{\substack{y \in \mathbb{R}^n \\ \|y\| = 1}} (Dy, y)(D^{-1}y, y)$$

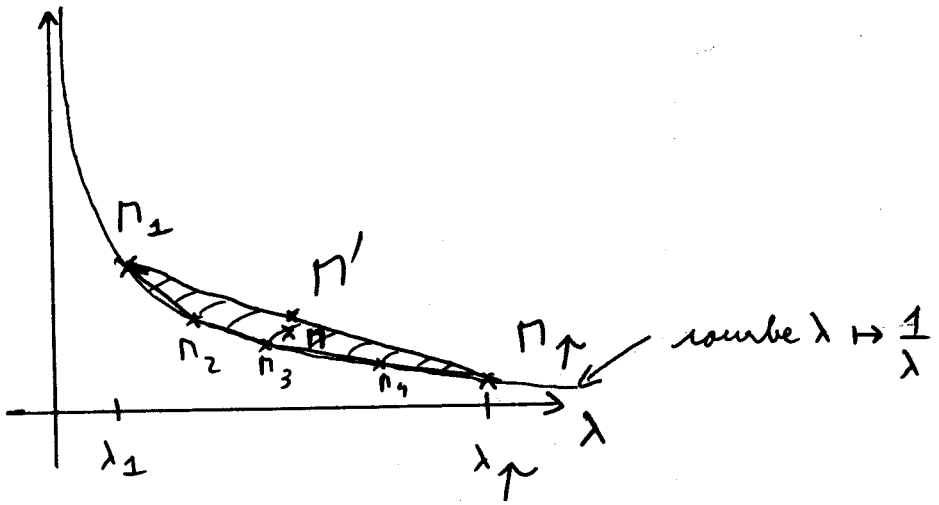
car $P^T A^{(-1)} P = D^{(-1)}$ et $(D^{(-1)}y, y) = (A^{(-1)}x, x)$.

On cherche donc le maximum de :

$$\begin{cases} \left(\sum_{i=1}^{n} \lambda_i y_i^2\right)\left(\sum \frac{y_i^2}{\lambda_i}\right) \\ \sum_{i=1}^{n} y_i^2 = 1 \end{cases}$$

Notons M_i les points de coordonnées $\left(\lambda_i, \frac{1}{\lambda_i}\right)$ et M leur barycentre avec les poids α_i^2.

On a le dessin :



Ce barycentre appartient à l'enveloppe convexe de $(M_1, \ldots, M_p)$. Notons $\overline{\lambda} = \sum_{i=1}^{p} \lambda_i \alpha_i^2$,

M a pour coordonnées $\left(\overline{\lambda}, \sum_{i=1}^{p} \frac{\alpha_i^2}{\lambda_i}\right)$.

Considérons M' sur la droite $(M_1 M_p)$ d'abscisse $\overline{\lambda}$.

On a, puisque cette droite a pour équation :

$$\lambda \mapsto \frac{1}{\lambda_1} + \frac{1}{\lambda_p} - \frac{\lambda}{\lambda_1 \lambda_p},$$

$$M'\left(\overline{\lambda}, \frac{1}{\lambda_1} + \frac{1}{\lambda_p} - \frac{\overline{\lambda}}{\lambda_1 \lambda_p}\right).$$

Le dessin montre que l'ordonné de M est inférieure à l'ordonnée de M', ce qui donne en multipliant par λ

$$\underbrace{\lambda \sum_{i=1} \frac{y_i^2}{\lambda_i}}_{= \sum_{i=1} \lambda_i y_i^2} \leq \lambda \left(\frac{1}{\lambda_1} + \frac{1}{\lambda_p} - \frac{\lambda}{\lambda_1 \lambda_p} \right).$$

Etudions maintenant la fonction $\lambda \mapsto \lambda \left(\frac{1}{\lambda_1} + \frac{1}{\lambda_p} - \frac{\lambda}{\lambda_1 \lambda_p} \right)$.

Ce polynôme de degré 2 atteint son unique maximum en $\lambda = \frac{\lambda_1 + \lambda_p}{2}$ ce qui donne :

$$\left(\sum_{i=1}^{\uparrow} \frac{y_i^2}{\lambda_i} \right) \left(\sum_{i=1}^{\uparrow} \lambda_i y_i^2 \right) \leq \left(\frac{\lambda_1 + \lambda_p}{2} \right) \left(\frac{\lambda_1 + \lambda_p}{2 \lambda_1 \lambda_p} \right)$$

qui est l'inégalité souhaitée avec égalité lorsque

$y_1 = y_p = \frac{1}{\sqrt{2}}$ et $y_2 = \dots = y_{p-1} = 0$ car :

$$\left(\sum_{i=2}^{\uparrow} \frac{y_i^2}{\lambda_i} \right) \left(\sum_{i=1}^{\uparrow} \lambda_i y_i^2 \right) = \frac{1}{2} \left(\frac{1}{\lambda_1} + \frac{1}{\lambda_p} \right) \frac{1}{2} \left(\lambda_1 + \lambda_p \right) = \frac{(\lambda_1 + \lambda_p)^2}{4 \lambda_1 \lambda_p}.$$

Pour étudier la vitesse de convergence, on va étudier :

$$f(x_n) - f(\bar{x}) = \frac{1}{2}(Ax_n, x_n) + (b, x_n) - \frac{1}{2}(A\bar{x}, \bar{x}) - (b, \bar{x})$$

Or $\frac{1}{2}(A(x_n - \bar{x}), x_n - \bar{x})$

$$= \frac{1}{2}(Ax_n, x_n) - (\underbrace{A\bar{x}}_{\substack{A \text{ symétrique} \\ \text{et } A\bar{x} = -b}}, x_n) + \frac{1}{2}(A\bar{x}, \bar{x})$$

$$= \frac{1}{2}(Ax_n, x_n) + (b, x_n) \pm \frac{1}{2}(A\bar{x}, \bar{x})$$

$$= \frac{1}{2}(Ax_n, x_n) + (b, x_n) - \frac{1}{2}(A\bar{x}, \bar{x}) - (\underbrace{(-A\bar{x})}_{=b}, \bar{x})$$

$$= f(x_n) - f(\bar{x}).$$

Ainsi, $f(x_n) - f(\bar{x}) = \frac{1}{2}(A(x_n - \bar{x}), x_n - \bar{x})$

$$\geq \frac{1}{2}\lambda_1 \|x_n - \bar{x}\|^2$$

(cf : précédemment).

On va donc estimer $f(x_n) - f(\bar{x})$.

Le choix de t_n dans la méthode précédente est particulier puisqu'il minimise :

$\varphi : t \mapsto f(x_n + t d_n)$ car

$$\varphi'(t) = t(Ad_n, d_n) + (\underbrace{Ax_n + b}_{=-d_n}, d_n)$$

donc $t_n = \dfrac{\|d_n\|^2}{(Ad_n, d_n)}$.

Posons : $e_n = f(x_n) - f(\bar{x})$. On a :

$$e_{n+1} = f(x_{n+1}) - f(\bar{x}) \qquad = -d_n - b$$

$$= \frac{1}{2}(Ax_n, x_n) + t_n(A\overbrace{x_n}, d_n) + \frac{t_n^2}{2}(Ad_n, d_n)$$

$$+ (b, x_n) + t_n(b, d_n) - f(\bar{x})$$

$$= f(x_n) - f(\bar{x}) \underbrace{- t_n \|d_n\|^2 + t_n^2 (Ad_n, d_n)}_{= -\frac{1}{2}\frac{\|d_n\|^4}{A(d_n, d_n)}}$$

donc :

$$e_{n+1} = e_n - \frac{1}{2}\frac{\|d_n\|^4}{(Ad_n, d_n)} = e_n\left(1 - \frac{\|d_n\|^4}{2(Ad_n, d_n)e_n}\right).$$

On a vu précédemment que

$$e_n = f(x_n) - f(\bar{x}) = \frac{1}{2}(A(x_n - \bar{x}), (x_n - \bar{x})), \text{ or}$$

$A(x_n - \bar{x}) = -d_n - b - A\bar{x} = -d_n$ donc :

$$e_n = \frac{1}{2}(d_n, A^{-1}d_n)$$

On applique l'inégalité de Kantorovitch :

$$\frac{-\|d_n\|^4}{(Ad_n, d_n)\underbrace{(A^{-1}d_n d_n)}_{2e_n}} < -\frac{4\lambda_1\lambda_p}{(\lambda_1 + \lambda_p)^2}$$

Ceci donne:

$$e_{n+1} \leq e_n \left(1 - \frac{4\lambda_1\lambda_p}{(\lambda_1+\lambda_p)^2}\right) = e_n \frac{(\lambda_1-\lambda_p)^2}{(\lambda_1+\lambda_p)^2}$$

ainsi, pour tout n,

$$e_n \leq \left(\frac{\lambda_1-\lambda_p}{\lambda_1+\lambda_p}\right)^{2n} e_0 \quad \text{et avec } e_n \geq \frac{1}{2}\lambda_1\|x_n-\bar{x}\|^2,$$

on obtient:

$$\|x_n-\bar{x}\| \leq \sqrt{\frac{2e_0}{\lambda_1}} \left|\frac{\lambda_1-\lambda_p}{\lambda_1+\lambda_p}\right|^n \text{ c'est à dire:}$$

$$\|x_n-\bar{x}\| \leq \sqrt{\frac{A(x_0,x_0)+2(b,x_0)-(A\bar{x},\bar{x})-(b,x_0)}{\lambda_1}} \left|\frac{\lambda_1-\lambda_p}{\lambda_1+\lambda_p}\right|^n.$$

Remarque: Comme $\lambda_p-\lambda_1 < \lambda_p+\lambda_1$, la convergence a lieu.

3) Méthode de relaxation pour les systèmes linéaires.

Le principe de cette méthode pour résoudre

$$A x = b \qquad A \in \mathbb{R}^{n \times n}$$

consiste à écrire $A = M - N$ puis à définir à partir de x_0,

$$M x_{n+1} = N x_n + b$$

où M est "facile" à inverser. Et nouveau, on peut écrire $x_{n+1} = f(x_n)$ avec $f(x) = M^{-1} N x + M^{-1} b$.

1) Théorème élémentaire.

Théorème: Si $\|\cdot\|$ est une norme sur $\mathbb{R}^n$ dont la norme subordonnée matricielle

$$|||A||| = \sup_{\|x\| = 1} \|Ax\|,$$

alors $|||M^{-1} N||| < 1$ donne la convergence de la méthode pour n'importe quel x_0.

Démonstration: On peut appliquer le théorème du point fixe à f :

$$\|f(x) - f(y)\| = \|M^{-1} N (x - y)\| \leq \underbrace{|||M^{-1} N|||}_{< 1} \|x - y\|$$

—

2) Rappels d'algèbre linéaire:

a) Décomposition de Dunford.

Théorème: Si $A \in \mathbb{R}^{n \times n}$, il existe un unique couple (D, N) tel que : (i) $A = D + N$,
(ii) D diagonalisable, N nilpotente,
(iii) $DN = ND$.

Éléments de démonstration : On raisonne sur les endomorphismes.

Si π_u est le polynôme minimal de u :

$$E = \bigoplus_{i=1}^{r} \ker\left[(u - \lambda_i \mathrm{id})^{m_i}\right] \quad \text{où} \quad \pi_u = \prod_{i=1}^{r} (X - \lambda_i)^{m_i}$$

On pose ensuite $d = \sum_{i=1}^{r} \lambda_i p_i$ où p_i est le projecteur sur $\ker\left[(u-\lambda_i)^{m_i}\right]$ puis $n = u - d$

On montre que ces projecteurs sont des polynômes en u ce qui donne $d \circ n = n \circ d$.

La diagonalisabilité de d se réduit du choix d'une base adapté à chaque projecteur de $\ker\left((u-\lambda_i \mathrm{id})^{m_i}\right)$ puis à sa concaténation.

On vérifie que $n_{|\ker(u-\lambda_i \mathrm{id})^{m_i}}$ est nilpotent pour tout i.

On montre l'unicité avec $u = d + n = d' + n'$ donc

$$d - d' = n' - n$$

d et d' sont co-diagonalisables car, d' commutant avec d et d' sont co-diagonalisables car, d' commutant avec u s'il commute avec u donc avec tout polynôme en u donc avec d. Ainsi $d - d'$ est diagonalisable. $n - n'$ est nilpotent. Le seul endomorphisme diagonalisable et nilpotent est l'endomorphisme nul.

b) Rayon spectral.

Définition : On pose :

$$\rho(A) = \max_{\lambda \in \mathrm{Sp}(A)} |\lambda| \quad \text{où } \mathrm{Sp}(A) \text{ est le spectre de } A.$$

Théorème : Sont équivalents :

(i) $\forall x$, $A^n x \xrightarrow[n \to +\infty]{} 0$.

(ii) $\rho(A) < 1$.

Eléments de démonstration :

Si x est un vecteur propre de A associé à λ :

$$A^n x = \lambda^n x \text{ donc } A^n x \xrightarrow[n\to+\infty]{} 0 \Rightarrow |\lambda| < 1$$

puis $\rho(A) < 1$.

On a : $\forall k \quad A^n = (D+N)^n = \sum_{i=0}^{k-1} \binom{n}{i} D^{n-i} N^{i-i}$ avec $N^k = 0$

car N est nilpotente.

Savons $\|D\| = \rho(D)$ (par exemple avec la norme euclidienne).

On a $\rho(A) = \rho(D) = \|D\|$ puis :

$$\|(D+N)^n\| \leq \sum_{i=0}^{k-1} \binom{n}{i} \rho(D)^{n-i} \|N\|^i \quad (*)$$

ainsi si $\rho(D) < 1$, $(D+N)^n \xrightarrow[n\to+\infty]{} 0$ pour la norme matricielle subordonnée et donc pour tout x, $A^n x \xrightarrow[n\to+\infty]{} 0$.

théorème : Pour toute norme matricielle (pas forcément subordonnée) sur $\mathbb{R}^{n\times n}$:

$$\lim_{n\to+\infty} \|A^n\|^{\frac{1}{n}} = \rho(A).$$

Eléments de démonstration :

On va se limiter au cas des normes subordonnées, par équivalence des normes en dimension finie,

$$\alpha^{\frac{1}{n}} \|A^n\|_1^{\frac{1}{n}} \leq \|A^n\|_2^{\frac{1}{n}} \leq \beta^{\frac{1}{n}} \|A^n\|_1^{\frac{1}{n}}$$

donne $\lim_{n\to+\infty} \|A^n\|_1^{\frac{1}{n}} = \lim_{n\to+\infty} \|A^n\|^{\frac{1}{n}}$

Pour une norme subordonnée, on a : si x est un vecteur propre associé à une valeur propre de module maximal :

$$\rho(A) = |\lambda| = \frac{\|\lambda x\|}{\|x\|} = \frac{\|Ax\|}{\|x\|} \leq |||A|||.$$

Puis, $\rho(A^n) = \rho(A)^n$ donne, puis $|||\ |||$ est une norme d'algèbre :

$$\rho(A)^n = \rho(A^n) \leq |||A^n||| \text{ donc } \rho(A) \leq \lim_{n\to+\infty} |||A^n|||^{\frac{1}{n}}.$$

Soit $\varepsilon > 0$.

Montrons qu'il existe une norme subordonnée telle que

$$|||A||| \leq \rho(A) + \varepsilon$$

Posons $B = \frac{1}{(\rho(A)+\varepsilon)} A$ et considérons $\|\cdot\|$ la norme euclidienne. On pose :

$$N(x) = \sum_{i=0}^{+\infty} \|B^i x\| \text{ qui existe par une majoration de}$$

type (*) ; comme $\rho(B) = \frac{\rho(A)}{\rho(A)+\varepsilon} < 1$, $\|B^i x\| < \alpha^i \|x\|$ $\alpha < 1$.

N est une norme et si $N(x) = 1$

$$N(Bx) = \sum_{i=1}^{\infty} \|B^i x\| = 1 - \|x\|.$$

Ainsi $\sup_{N(x)=1} N(Bx)$ est atteint en un point x_0 (sup sur un compact) où $N(x) = 1$ donc $x \neq 0$.

Ainsi $\sup_{N(x)=1} N(Bx) = 1 - \underbrace{\|x_0\|}_{\neq 0} < 1$.

On pose $|||B||| = \sup_{N(x)=1} N(Bx)$. On a :

$$|||A||| < \rho(A) + \varepsilon.$$

On considère la norme: $\sqrt{(Ax|x)} = \|x\|$. On a:

$M^{-1}N = I - M^{-1}A$

On a $|||M^{-1}N||| = \max_{\|x\|=1} \|x - M^{-1}Ax\|$.

On a, avec $y = M^{-1}Ax$, avec $(Ax|x) = 1$

$$(A(x-y)|(x-y)) = 1 - ({}^tMy|y) - (My|y) + (Ay|y)$$

$$= 1 - ((M + {}^tM - A)y|y)$$

$$= 1 - ((M + {}^tN)y|y)$$

Si $M + {}^tN$ est définie positive, pour tout $y \neq 0$,

$(A(x-y)|(x-y)) < 1$ et

$(A(x-y)|(x-y)) = \|x - M^{-1}Ax\|$.

Or $\max_{\|x\|} \|x - M^{-1}Ax\|$ est atteint en x_0, avec $y_0 = M^{-1}Ax_0 \neq 0$, on a:

$|||M^{-1}N||| = \|x_0 - M^{-1}Ax_0\| = 1 - ((M + {}^tN)y_0|y_0) < 1$

Ainsi $\rho(M^{-1}N) < 1$ ce qui donne le résultat.

4) Méthode de relaxation:

Cette fois on pose: $M_\omega = \left(\frac{D}{\omega} - E\right)$ $N_\omega = \left(\frac{(1-\omega)D}{\omega} + F\right)$

puis $\mathcal{L}_\omega = M_\omega^{-1} N_\omega$.

Théorème: Soit A symétrique définie positive, et $\omega \in]0, 2[$.

Pour tous x_0, b, la méthode converge.

Démonstration: $M + {}^tN = \frac{D}{\omega} - E + \frac{1-\omega}{\omega}D + {}^tF$

comme ${}^tA = A$, $-E = {}^tF$

Ceci donne:
$\|\|A^n\|\|^{\frac{1}{n}} \leq \rho(A) + \varepsilon$ pour tout ε et la conclusion.

3) Méthodes de Jacobi et Gauss-Seidel.

Méthode de Jacobi :

$$A = \begin{pmatrix} D & -F \\ -E & \end{pmatrix} \qquad M = D \ , \ N = E + F$$

Méthode de Gauss-Seidel : $M = D - E \quad N = F$.

On donne un théorème pour les méthodes $x_{n+1} = M^{-1}N x_n + M^{-1}b$.

Théorème : La méthode converge pour tous x_0, b ssi $\rho(M^{-1}N) < 1$.

Eléments de démonstration : Soyons $B = M^{-1}N$, $c = M^{-1}b$.

On a : $x_n = B^n x_0 + \sum_{i=0}^{n-1} B^i c$.

Si $\rho(B) < 1$, par la majoration (*), il existe $c < 1$ telle que $\|B\|^i \leq C_1 C^i$ donc on a convergence.

Si $\rho(B) \geq 1$, $Bx = \lambda x$ avec $|\lambda| = \rho(B)$ donne pour $x_0 = 0$ et $c = x$, $x_n = \left(\sum_{i=0}^{n-1} \lambda^i \right) x$ qui diverge.

Théorème : Soit A une matrice symétrique. Alors $M + {}^tN$ est symétrique. Si $M + {}^tN$ est définie positive la méthode converge pour tout x_0, b.

Démonstration :

$$M + {}^tN = M + {}^tM - \underbrace{A}_{\text{symétrique}}$$ est bien symétrique

$M + tN = \frac{2-\omega}{\omega} D$ qui est bien définie positive lorsque $\omega \in]0,2[$.

Théorème : Pour tout $\omega \neq 0$, $\rho(\mathcal{L}_\omega) \geq |\omega - 1|$.

Démonstration :

$$\det(\mathcal{L}_\omega) = \frac{\det\left(\frac{1-\omega}{\omega} D + F\right)}{\det\left(\frac{D}{\omega} - E\right)}$$

(triangulaire inférieure)

(triangulaire supérieure)

$$= \frac{\left(\frac{1-\omega}{\omega}\right)^n}{\left(\frac{1}{\omega}\right)^n} \frac{\det D}{\det D} = (1-\omega)^n.$$

Or le déterminant étant le produit des valeurs propres :

$$\rho(\mathcal{L}_\omega)^n \geq |\det(\mathcal{L}_\omega)| \geq |1-\omega|^n.$$

Remarque : Donc la méthode de relaxation ne converge pas si $|1-\omega| > 1$ c'est à dire $\omega \notin]0,2[$.

On peut calculer une valeur optimale ω_0 pour une matrice donné pour que $\rho(\mathcal{L}_\omega)$ soit minimal.

Par exemple avec $A = \begin{pmatrix} 2 & -1 & 0 & \\ -1 & \ddots & \ddots & -1 \\ & \ddots & \ddots & 2 \\ 0 & & -1 & 2 \end{pmatrix}$, on a, pour $n \to +\infty$:

$$-\log\left(\rho\left(D^{-1}(E+F)\right)\right) \sim \frac{\pi^2}{2(n+1)^2} \quad \text{(Jacobi)}$$

$$-\log\left(\rho\left(\mathcal{L}_1\right)\right) \sim \frac{\pi^2}{(n+1)^2} \quad \text{(Gauss-Seidel)}$$

$$-\log\left(\rho\left(\mathcal{L}_{\omega_0}\right)\right) \sim \frac{2\pi}{n+1} \quad \text{(relaxation } \omega_0 \text{ optimal)}$$

ANNEXE

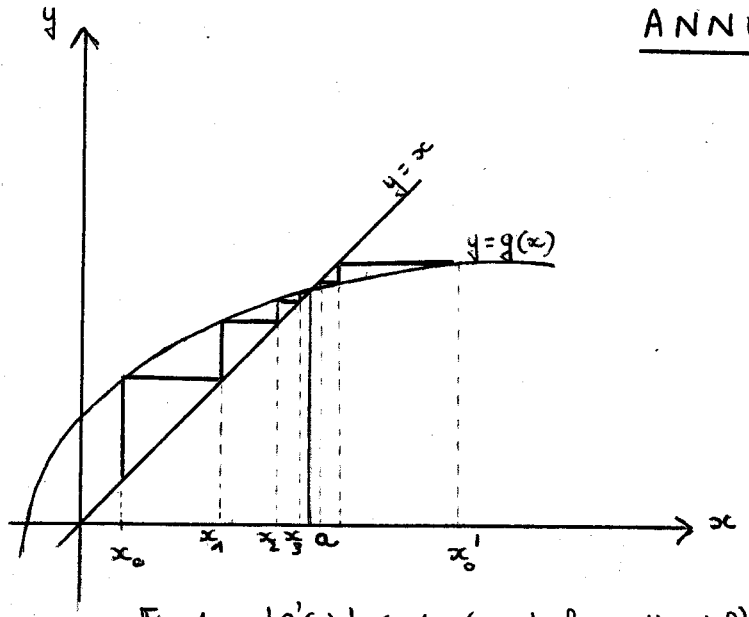


Fig 1 : $|g'(a)| < 1$ (point fixe attractif)
→ la méthode converge

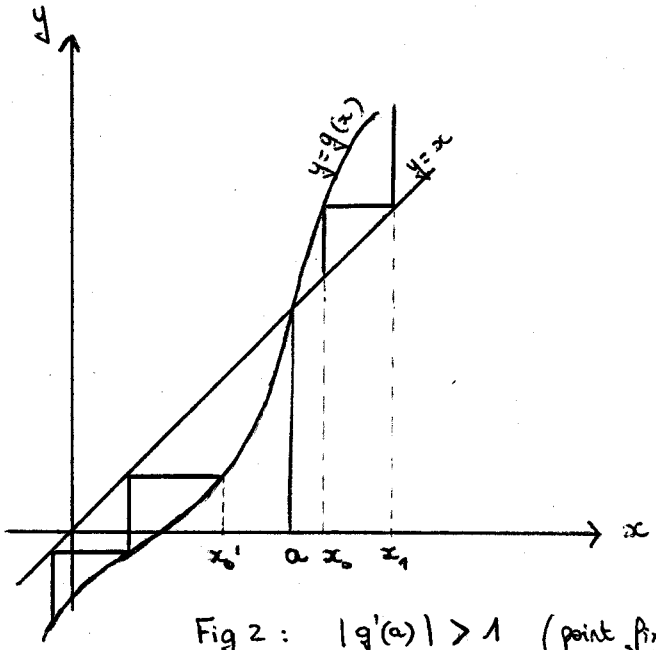


Fig 2 : $|g'(a)| > 1$ (point fixe répulsif)
→ la méthode ne converge pas.

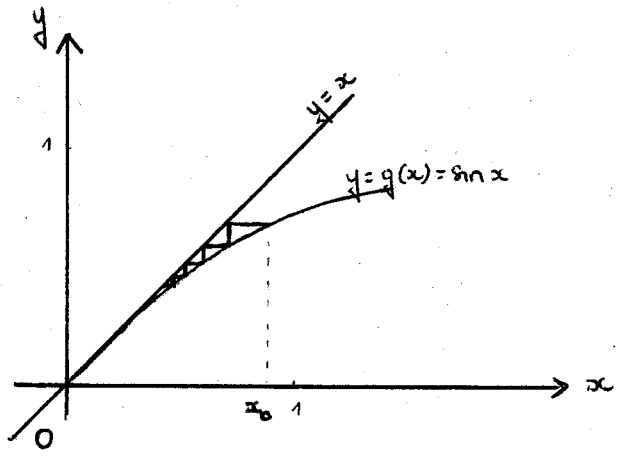


Fig 3 : $g'(0) = 1$
→ la méthode converge

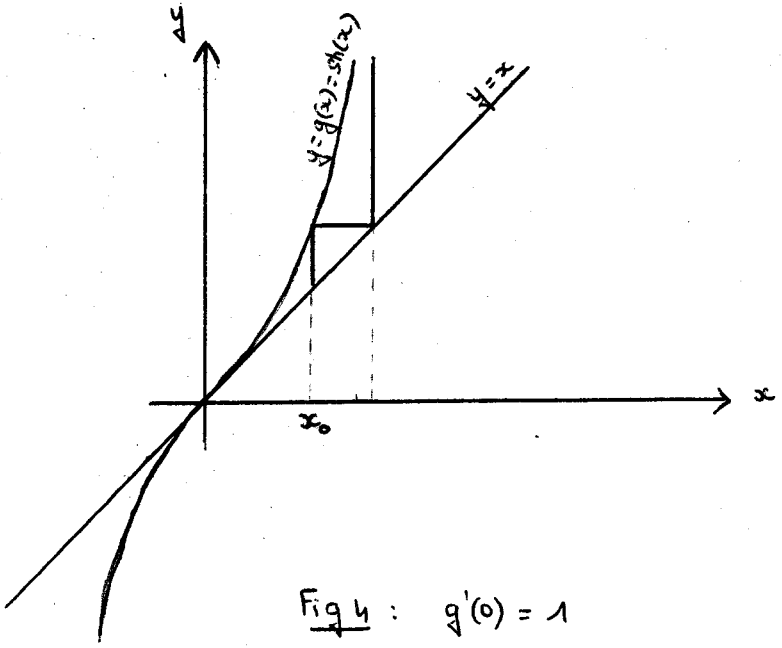


Fig 4 : $g'(0) = 1$
→ la méthode ne converge pas.