Mise à jour du 5-12-2007 .

L3 ANALYSE NUMERIQUE, RESUME DE COURS, G. BOURGEOIS, 2007-2008

BIBLIOGRAPHIE :

i) Analyse numérique, une approche mathématique,2ème édition / Schatzman / Dunod.
ii) Introduction à l'analyse numérique . Applications sous Matlab./ Bastien et Martin / Dunod
iii) Numerical analysis in modern scientific computing / Deuflhard et Hohmann / Springer

NOTATIONS :

$f \in C^0(I)$ ssi f est continue sur l'intervalle I ; si I est un segment $\| f \|_\infty = \sup_{x\in I} | f(x) |$

$f \in C^k$ ssi f est k fois continûment dérivable.

$f = o(g)$ au $\mathcal{V}(a)$ ssi $\lim_{x\to a}(\frac{f}{g}) = 0$

$f = O(g)$ au $\mathcal{V}(a)$ ssi $\exists A > 0$ tel qu'au $\mathcal{V}(a)$ $| f | \leq A | g|$

$f \sim g$ au $\mathcal{V}(a)$ ssi $\lim_{x\to a}(\frac{f}{g}) = 1$

CHAPITRE 1 : **Approximation des restes de séries et d'intégrales convergentes.**

I- Séries. Théorème 1 :

Soient $(a_n)_n$ et $(b_n)_n$ des séries positives telles que: $a_n = o(b_n)$ (resp. $a_n \sim b_n$) et $\sum_{n=1}^{\infty} b_n$ converge. Alors $\sum_{n=1}^{\infty} a_n$ converge et $\sum_{p=n}^{\infty} a_p = o(\sum_{p=n}^{\infty} b_p)$ (resp. $\sum_{p=n}^{\infty} a_p \sim \sum_{p=n}^{\infty} b_p$).

II- Intégrales. Théorème 2 :

$b \in [a,\infty]$, f, g sont des fonctions positives continues sur [a,b[telles que $f = o(g)$ (resp. $f \sim g$) au $\mathcal{V}(b-)$ et $\int_a^b g(t)dt$ converge. Alors $\int_a^b f(t)dt$ converge et $\int_a^x f(t)dt = o(\int_a^x g(t)dt)$ (resp. $\int_a^x f(t)dt \sim \int_a^x g(t)dt$) au $\mathcal{V}(b-)$).

III- Comparaison séries-intégrales :

1) f est croissante (resp. décroissante) sur $[a, b+1]$;

alors $\int_a^{b+1} f(t)dt \leq \sum_{k=a}^{b} f(k) \leq f(a) + \int_a^b f(t)dt$ (resp. $\int_a^{b+1} f(t)dt \geq \sum_{k=a}^{b} f(k) \geq f(a) + \int_a^b f(t)dt$)

2) Si f est décroissante et $\int_a^{\infty} f(t)dt$ converge, alors $\int_{n+1}^{\infty} f(t)dt \leq \sum_{k=n}^{\infty} f(k) \leq \int_n^{\infty} f(t)dt$.

3) Exemple : si $\alpha > 1$, $\sum_{k=n}^{\infty} \frac{1}{k^\alpha \ln^\beta(k)} \sim \int_n^{\infty} \frac{dt}{t^\alpha \ln^\beta(t)} \sim \frac{1}{(\alpha-1)n^{\alpha-1}\ln^\beta(n)}$ et

$$\sum_{k=2}^{\infty} \frac{1}{k^\alpha \ln^\beta(k)} \approx \sum_{k=2}^{n-1} \frac{1}{k^\alpha \ln^\beta(k)} + \frac{1}{(\alpha-1)n^{\alpha-1}\ln^\beta(n)}$$

CHAPITRE 2 : **Polynômes orthogonaux et approximation dans L^2 .**

I- Espaces préhilbertiens :
1)Définition : Soit E un EV.
$(f,g)\in E\times E\to <f,g>\in\mathbb{R}$ est un produit scalaire sur E ssi $<f,g>$ est bilinéaire, symétrique, $\forall f$ $<f,f>\geq 0$ et $<f,f>=0\Rightarrow f=0$.
Remarque : E est dit alors EV préhilbertien et $f\in E\to <f,f>=\|f\|^2$ définit une norme sur E. Si E est de dimension finie, il est dit EV euclidien.
2) Procédé de Schmidt :
Soit $(e_p)_{p\leq n}$ une base de l'EV euclidien E de dimension n. Alors il existe une base orthonormée (BON) de E : $(f_p)_{p\leq n}$ « unique » (chaque vecteur est défini à un facteur -1 près) telle que $\forall k\leq n$ $\left[f_1,...,f_k\right]=\left[e_1,...,e_k\right]$.
La matrice de passage de $(e_p)_p$ à $(f_p)_p$ est triangulaire.
3) Projection orthogonale sur un sous EV de dimension finie.
Soient E hilbertien et F un sous EV de dim. n de E.
Proposition : il existe un projecteur unique $x\in E\to p(x)\in F$ tel que $\forall x,\ p(x)-x$ soit orthogonal à F. De plus
i) $p(x)$ réalise le minimum de la fonction $y\in F\to\|y-x\|$.
ii) Si $(f_i)_i$ est une BON de F alors $p(x)=\sum_{i=1}^{n}<x,f_i>f_i$.

II-Polynômes orthogonaux :

1) Définition 1 : I est un intervalle fermé de $\overline{\mathbb{R}}=\left[-\infty,\infty\right]$; $w:\overset{0}{I}\to\mathbb{R}^+$ est continue et $\forall n,\ x^n w(x)$ est intégrable sur I ; w ne s'annule sur aucun sous intervalle de I.
$\mathbb{R}\left[x\right]$ est un sous EV de $L^2(w)=\{f\in C^0(\overset{0}{I})\mid\int_I f(t)^2w(t)dt<\infty\}$.
Proposition 1: $(f,g)\to\int_I f(t)g(t)w(t)dt$ fait de $L^2(w)$ un espace préhilbertien de dimension infinie.
2) Définition 2: Soit $(P_k)_{k\in\mathbb{N}}$ une suite de polynômes. C'est une suite de polynômes orthogonaux relativement à w ssi :
$\forall k\in\mathbb{N},\ P_k(x)=x^k+.....$ et P_k est orthogonal à $\mathbb{R}_{k-1}\left[x\right]$.
Une telle suite s'obtient de façon « unique » par I-2).
3) Exemples : $w(x)=1$ donne les polynômes de Legendre [1752-1833] ($\left[(x^2-1)^n\right]^{(n)}$ à un facteur près).
$w(x)=\dfrac{1}{\sqrt{1-x^2}}$ donne les polynômes de Chebyshev [1821-1894] :
$\dfrac{1}{2^{n-1}}T_n(x)$ où $T_n(x)=\cos(n\ Arc\cos(x))$

4) Théorème minimax : si $P\in\mathbb{R}_n\left[x\right]$ est normalisé, alors $\sup_{x\in[-1,1]}|P(x)|\geq\dfrac{1}{2^{n-1}}$ et l'égalité n'est réalisée que pour $\dfrac{1}{2^{n-1}}T_n(x)$.

III- Approximation dans $L^2(w)$ (au sens des moindres carrés).

Soient $f \in L^2(w)$ et $n \in \mathbb{N}$. La projection orthogonale Q_n de f sur $\mathbb{R}_n[x]$ est la meilleure approximation de f au sens de $L^2(w)$ par un polynôme de degré n.

Proposition 2 : si $f \in C^2$ et $w = 1$ alors $\| f - Q_n \|_\infty = o(\frac{1}{\sqrt{n}})$, soit une convergence uniforme assez lente.

CHAPITRE 3 : **Interpolation et différences divisées.**

I- Différences divisées (Newton [1642-1727]):

Soit $f \in C^n$ et $t_0,...,t_n$ des réels non forcément distincts d'un intervalle *I*.

On définit $f[t_0,....,t_n]$ par l'algorithme récursif :

i) Tant que $\exists i, j,$ tels que $t_i \neq t_j$ alors $f[t_0,....,t_n] = \dfrac{f[t_0,..,\widehat{t_i},..,t_n] - f[t_0,..,\widehat{t_j},..,t_n]}{t_j - t_i}$.

On sort de cette boucle lors de la situation suivante:

ii) $f[t_0,....,t_n] = \dfrac{f^{(n)}(t_0)}{n!}$ si $t_0 = ... = t_n$.

On peut aussi en déduire un algorithme à la Pascal.

Propriétés : $f[t_0,....,t_n]$ est invariant par permutation des t_i.

Si *I* est un intervalle contenant les t_i alors $\exists \zeta \in I$ tel que $f[t_0,....,t_n] = \dfrac{f^{(n)}(\zeta)}{n!}$.

II-Interpolation de Lagrange [1736-1813]:

1) Définition :

Soit $f \in C^0$. Soient $x_0,....,x_n$ des réels 2 à 2 distincts et $L_i(x) = \dfrac{\prod_{j \neq i} (x - x_j)}{\prod_{j \neq i} (x_i - x_j)} \in \mathbb{R}_n[x]$.

Proposition 1: Il existe un unique polynôme P_n de degré n tel que $\forall i,\ P_n(x_i) = f(x_i)$; P_n est donné explicitement par $P_n = \sum_{i=0}^{n} f(x_i) L_i$.

Remarque : Si on choisit mal les x_i, l'approximation peut être mauvaise (phénomène de Runge).

2) Proposition 2 : $P_n(x) = \sum_{i=0}^{n} f[x_0,...,x_i] \prod_{j=0}^{i-1} (x - x_j)$ avec la convention que le produit $\prod$ vaut 1 si $i = 0$.

L'intérêt de cette présentation est que si on ajoute un point d'interpolation $x_{n+1} \in I$, alors le nouveau polynôme d'interpolation est $P_n(x) + f[x_0,...,x_{n+1}] \prod_{j=0}^{n} (x - x_j)$.

3) Erreur : $e_n(t) = f(t) - P_n(t) = f[x_0,...,x_n,t] \prod_{j=0}^{n} (t - x_j)$**.**

Proposition 3 : $|e_n(t)| \le \frac{\|f^{(n+1)}\|_\infty}{(n+1)!}\prod_{j=0}^{n}|t-x_j|$. Si $I=[-1,1]$ alors l'erreur est minimale (et on évite le phénomène de Runge) si on choisit les x_i comme zéros du polynôme de Chebyshev de degré $n+1$ (en raison de II-4)).

III- Interpolation d'Hermite [1822-1901]: Soit $f \in C^\infty$.

Soient $t_1,...,t_k$ des réels 2 à 2 distincts . Les α_i sont des entiers tels que $\sum_{i=1}^{k}\alpha_i = n+1$.

Proposition 4 : Il existe un polynôme H_n de degré n unique tel que :
$\forall i \le k, \forall \lambda_i \in \{0,...,\alpha_i - 1\},\ H_n^{(\lambda_i)}(t_i) = f^{(\lambda_i)}(t_i)$. De plus si on écrit α_i fois chaque t_i alors on obtient la suite $u_0,....,u_n$ et $H_n(x) = \sum_{i=0}^{n} f[u_0,...,u_i]\prod_{j=0}^{i-1}(x-u_j)$. L'erreur $f - H_n$ est donnée par la même formule qu'au II-3).

CHAPITRE 4 : **Approximation de l'intégrale par la méthode de Gauss [1777-1855].**
$(P_n)_n$ est la suite de polynômes orthogonaux (normalisés) associée à w (cf. chapitre II).

I- Proposition 1: si $n \in \mathbb{N},\ \exists! A_0,...,A_n \in \mathbb{R}$ et $x_0,...,x_n \in I$ tels que $\forall P \in \mathbb{R}_{2n+1}[x]$
$\int_I w(x)P(x)dx = \sum_{i=0}^{n} A_i P(x_i)$; de plus les x_i sont les racines de P_{n+1}.

II- Formule de Gauss :

1) Proposition 2: soit $f \in C^{2n+2}(I)$ **;** alors $\int_I w(x)f(x)dx$ admet pour valeur approchée $\sum_{i=0}^{n} A_i f(x_i)$ avec une erreur majorée par $\frac{\|f^{(2n+2)}\|_\infty}{(2n+2)!} < P_{n+1}, P_{n+1} >$.

2) Exemple : Si $w(x) = \frac{1}{\sqrt{1-x^2}}$, $\int_{-1}^{1}\frac{f(t)}{\sqrt{1-t^2}}dt \approx \frac{\pi}{n+1}\sum_{i=0}^{n} f(\cos(\frac{2i+1}{2n+2}\pi))$.

CHAPITRE 5 : **Approximation uniforme ; polynômes de Bernstein [1880-1968]:**

1) Définition 1 et propriété: Soit $f \in C^0(I)$ où I est un *segment* ; alors f est uniformément continue et le module de continuité de f $\omega : h \in \mathbb{R}^+ \to \sup_{\substack{x,y\in I \\ |x-y|\le h}} |f(x) - f(y)|$ est une fonction continue croissante.

Exemple : si $f \in C^1(I)$ alors $|\omega(h)| \le h\|f'\|_\infty$.

Dans la suite $I = [0,1]$ et $f \in C^0(I)$.

Définition 2 (polynômes de Bernstein) : $\beta_{n,j}(x) = C_n^j x^j(1-x)^{n-j}$; $B_n(x) = \sum_{j=0}^{n}\beta_{n,j}(x) f(\frac{j}{n})$.

2) Théorème de Weierstrass [1815-1897] (par Bernstein) :
$\|f - B_n(f)\|_\infty \le \frac{9}{4}\omega(\frac{1}{\sqrt{n}})$ (la suite $(B_n(f))_n$ converge uniformément sur I vers f).

Les polynômes sont denses dans $C^0([a,b], \|\ \|_\infty)$.

Remarque 0 : le résultat est faux si I n'est pas borné.

Remarque 1 : si $f \in C^1(I)$ alors $\| f - B_n(f) \|_\infty = O(\frac{\| f' \|_\infty}{\sqrt{n}})$ **.**

Remarque 2 : si $f \in C^\infty(I)$ alors $\| f - B_n(f) \|_\infty = O(\frac{\| f'' \|_\infty}{n^2})$ et on ne peut améliorer cette estimation.

CHAPITRE 6 : **Les courbes de Bézier [1910-1999].**
Méthode utilisée en CAO, initiée dans les années 1960 dans l'industrie automobile par Bézier (Renault), Casteljau (Citroën), Birkhoff et De Boor (General Motors).
I- Rappel de géométrie : soient $x_0, ..., x_n$ des points de $\mathbb{R}^d$ et $(\alpha_i)_{0 \le i \le n}$ des réels tels que $\sum_{i=0}^{n} \alpha_i = 1$; alors le point $G = \sum_{i=0}^{n} \alpha_i x_i$ est bien défini et est dit barycentre des $x_i(\alpha_i)$.
Si de plus les $(\alpha_i)_i$ sont positifs alors G est dans l'enveloppe convexe des $(x_i)_i$.
Réciproquement si un point est dans cette enveloppe convexe, alors il est barycentre (à poids positifs) de $d+1$ des $(x_i)_i$ (Caratheodory).
II- 1) Définition : Soient $x_0, ..., x_n$ des points du plan ; la courbe de Bézier associée est $t \in [0,1] \to X(t; x_0, ..., x_n) = \sum_{j=0}^{n} \beta_{n,j}(t) x_j$ (notations du chapitre 5)..
2) Propriété 1: X est un chemin orienté de x_0 à x_n admettant pour vecteurs tangents en ces points : $n(x_1 - x_0)$ et $n(x_n - x_{n-1})$.
3) Propriété 2 : 1er) $X(t; x_0, ..., x_n) = (1-t) X(t; x_0, ..., x_{n-1}) + t\, X(t; x_1, ..., x_n)$
2ème) d'où une construction géométrique de $X(t; x_0, ..., x_n)$ par une récurrence à la Pascal (Algorithme de Casteljau).
a) En voici le diagramme pour $n = 3$ (cf. cours):

x_0

$x_1 \quad y_0$

$x_2 \quad y_1 \quad z_0$

$x_3 \quad y_2 \quad z_1 \quad X(t)$

b) Alors
$X(I; x_0, y_0, z_0, X(t)) = X([0,t]; x_0, x_1, x_2, x_3)$, $X(I; X(t), z_1, y_2, x_3) = X([t,1]; x_0, x_1, x_2, x_3)$
et $X'(t; x_0, x_1, x_2, x_3)$ est dirigé par $z_1 - z_0$.
c) Si on itère la transformation $x_0, x_1, x_2, x_3 \to x_0, y_0, z_0, X(\frac{1}{2}), z_1, y_2, x_3$ on obtient une ligne polygonale qui devient rapidement très proche de $X(I; x_0, x_1, x_2, x_3)$.

CHAPITRE 7 : **Résolution approchée des systèmes non linéaires de *p* équations à *p* inconnues.**
I- Rappels sur les normes :
1) Définition : Si ||.|| est une norme sur $\mathbb{R}^p$, alors la norme subordonnée sur $\mathfrak{M}_p(\mathbb{R})$ associée est : $A \to \sup_{x \in \mathbb{R}^p \setminus \{0\}} \frac{\|Ax\|}{\|x\|}$.

2) Exemples : Les normes sur $\mathfrak{M}_p(\mathbb{R})$ subordonnées à l_1 et l_∞ sont
$\| A\|_1 = \underset{j}{Max}\sum_i |a_{ij}|$ et $\| A\|_\infty = \underset{i}{Max}\sum_j |a_{ij}|$

II- Théorème du point fixe de Banach [1892-1945] (TPF) :

Ici *E* est un espace de Banach (par exemple $\mathbb{R}^p$) et *F* un fermé de *E*.
Les normes d'applications linéaires sont ici des normes subordonnées.

1) Enoncé du théorème :
$f : F \to F$ est telle que $\exists k < 1, \forall x, y \in F, \| f(x) - f(y)\| \leq k \| x - y\|$.
Alors i) $\exists! a \in F,\ f(a) = a$.
ii) La suite définie par $\{x_0 \in F, x_{n+1} = f(x_n)\}$ converge vers *a*.
iii) $\| x_n - a\| = O(k^n)$ quand $n \to \infty$.

2) Remarque : Si l'erreur $\| x_n - a\|$ est majorée par Δ, alors l'erreur $\| x_{n+r} - a\|$ est majorée par Δk^r.

3) Conditions d'application du TPF : Si *F* est convexe**,** $f : F \to F$ est C^1 et $\exists k < 1$ tel que $\forall x \in F, \| Df(x)\| \leq k$ alors on peut appliquer le TPF.

III- Méthode de Newton :

1) Formule de Taylor : *B* est un compact convexe de $\mathbb{R}^p$ et $f : B \to \mathbb{R}^p$ est C^2 ; si $a, b \in B$:
$f(b) = f(a) + Df(a)\left[b - a\right] + O(\| b - a\|^2)$.

2) Théorème de l'itération de Newton-Raphson : *B* est une boule fermée de $\mathbb{R}^p$ et $f : B \to \mathbb{R}^p$ est C^2 ; $\exists x \in \overset{0}{B}$ tel que $f(x) = 0$ et $Df(x) \in GL_p(\mathbb{R})$.

Alors i) $\exists \varepsilon > 0$ tel que si $\| y_0 - x\| < \varepsilon$ alors la suite $\{y_0, y_{j+1} = y_j - \left[Df(y_j)\right]^{-1}(f(y_j))\}$ est bien définie et converge vers *x*.
ii) $\| y_{j+1} - x\| = O(\| y_j - x\|^2)$.

a) Remarque : il y a doublement du nombre de chiffres significatifs à chaque pas.
b) Avantage de la méthode : elle converge beaucoup plus vite que l'itération associée au TPF.
Désavantage de la méthode : pour être sûr de la convergence vers *x*, il faut débuter l'itération près de *x*.

3) Applications aux fonctions holomorphes :

a) Définition : U est un ouvert de $\mathbb{C}$ **;** $f : U \to \mathbb{C}$ est dite holomorphe sur U ssi $\forall u \in U, \exists\ r > 0$ tel que f soit développable en série entière dans $B(u, r)$. En particulier f est dérivable . Exemple : $z \in \mathbb{C} \to e^z$.

b) Théorème : si $f : U \to V \subset \mathbb{C}$ est holomorphe bijective et si $\forall z, f'(z) \neq 0$ alors f admet pour réciproque une fonction holomorphe g telle que $g'(z) = \dfrac{1}{f'(f^{-1}(z))}$.

Exemple : $\exp : z \in U = \{u \in \mathbb{C};\ -\pi < \mathrm{Im}(u) < \pi\} \to e^z \in \mathbb{C} \setminus \mathbb{R}^- = \{\rho e^{i\theta}; \rho > 0, \theta \in \left]-\pi, \pi\right[\}$ admet pour réciproque la fonction holomorphe $\ln : \rho e^{i\theta} \in \mathbb{C} \setminus \mathbb{R}^- \to \ln(\rho) + i\theta \in U$ de dérivée $\ln'(z) = \dfrac{1}{z}$.

c) Le théorème de Newton et le TPF s'appliquent *mutatis mutandis* aux fonctions holomorphes, la différentielle $Df(x)$ devenant $f'(x)$, la norme étant le module dans $\mathbb{C}$.

CHAPITRE 8 : **Systèmes linéaires.**
I) Matrices symétriques définies positives (SDP **).**
1) Définition : Soit $A \in \mathfrak{M}_n(\mathbb{R})$ une matrice symétrique. A est SDP ssi
$\forall x \in \mathbb{R}^n \setminus \{0\}, {}^t xAx > 0 \Leftrightarrow spectre(A) \subset \mathbb{R}^{+*} \Leftrightarrow$ Les n déterminants principaux de A sont strictement positifs .
2) Exemple : Si $A \in GL_n(\mathbb{R})$, alors ${}^t AA$ est SDP.
II) Approximation des moindres carrés (Gauss).
Théorème : $m > n, A \in \mathfrak{M}_{m,n}(\mathbb{R}), rang(A) = n, b \in \mathbb{R}^m$. Alors $x \in \mathbb{R}^n \to \|Ax - b\|$ (norme euclidienne) présente un minimum absolu unique en le point x défini par ${}^t AAx = {}^t Ab$, soit un système matriciel d'ordre n de matrice SDP.
III) Résolution itérative.
Théorème : $A \in \mathfrak{M}_n(\mathbb{R})$ est SDP, $b \in \mathbb{R}^n$. $A = M - N$ où M est triangulaire inférieure et N est triangulaire supérieure stricte. Soit la suite définie par $x_0 \in \mathbb{R}^n$, $Mx_{k+1} = Nx_k + b$.
Alors $(x_k)_k$ converge vers la solution de $Ax = b$; on déduit x_{k+1} de x_k en $\sim 2n^2$ opérations.
IV) Complexité.
1) Produits : Si $A, B \in \mathfrak{M}_n(\mathbb{R})$, alors le calcul de AB se fait en $\sim 2n^3$ opérations et si $x \in \mathbb{R}^n$, le calcul de Ax se fait en $\sim 2n^2$ opérations.
2) La résolution d'un système triangulaire (n, n) se fait en $\sim 2n^2$ opérations.

CHAPITRE 9 : **Erreurs d'arrondis.**
1) Définitions : Si on écrit $a \approx 567.7865$, on sous-entend que a est connue avec une **erreur absolue** $\Delta a < 0.5 \times 10^{-4}$, c'est-à-dire $a \in \left]567.7865 - 0.5 \times 10^{-4}, 567.7865 + 0.5 \times 10^{-4}\right[$.
L'**erreur relative** commise sur a est $\dfrac{\Delta a}{|a|}$, soit pratiquement $\dfrac{0.5 \times 10^{-4}}{567.7865} \approx 10^{-7}$.
2) Si a est connu avec k chiffres significatifs (CS), alors $\dfrac{\Delta a}{|a|} \approx 10^{-k}$. (Ici k=7).
3) Addition/Soustraction : si a, b sont connus avec k CS, alors on n'a pas le contrôle du nombre de CS de $a + b$.
4) Multiplication/Division : si a, b sont connus avec k CS, alors ab et $\dfrac{a}{b}$ sont connus avec k CS (Utiliser la dérivée logarithmique).