

ANNALES SCIENTIFIQUES
DE L'UNIVERSITÉ DE CLERMONT-FERRAND 2
Série Mathématiques

P. DAUBEZE

Application de l'approximation stochastique à la détermination des masques pour la lecture automatique de caractères imprimés

Annales scientifiques de l'Université de Clermont-Ferrand 2, tome 58, série *Mathématiques*, n° 12 (1976), p. 64-93

http://www.numdam.org/item?id=ASCFM_1976__58_12_64_0

© Université de Clermont-Ferrand 2, 1976, tous droits réservés.

L'accès aux archives de la revue « Annales scientifiques de l'Université de Clermont-Ferrand 2 » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

APPLICATION DE L'APPROXIMATION STOCHASTIQUE
A LA DETERMINATION DES MASQUES
POUR LA LECTURE AUTOMATIQUE DE CARACTERES IMPRIMES

P. DAUBEZE , Université de Toulouse

I N T R O D U C T I O N

Nous nous proposons de résoudre un problème de détermination de masques en vue de la lecture automatique de caractères imprimés dans un style donné.

Ce problème peut se formuler de la manière suivante :

soit $Y(x)$ une variable aléatoire réelle dépendant du paramètre $x \in \mathbb{R}^d$ ($d \geq 1$), dont la distribution de probabilité est inconnue, mais sur laquelle on peut effectuer, quelque soit $x \in \mathbb{R}^d$, des observations indépendantes.

Considérons la fonction de régression $M(x) = E(Y(x))$, le problème consiste alors, étant donné un convexe fermé C de \mathbb{R}^d , à minimiser $M(x)$ sur C . Nous sommes donc amenés à résoudre un problème d'optimum avec contrainte .

Le point de vue adopté pour la résolution de ce problème a été l'introduction d'une suite de fonctions de pénalité convexes, différentiables, telle qu'à la limite l'optimum soit dans le domaine des contraintes cependant que la fonction de pénalité limite n'est pas différentiable sur la frontière du convexe.

CHAPITRE 1

UNE EXTENSION D'UN THEOREME DE C. MACCHI

Introduction

Dans le cadre des processus d'itération stochastique multi-dimensionnels, les résultats présentés dans ce chapitre constituent une extension d'un théorème de C. MACCHI (11). Cette extension nous permettra de résoudre un problème d'optimisation avec contrainte qui sera présenté dans le chapitre 2.

Avant de présenter cette extension nous allons décrire le problème qui fut à l'origine de ce travail de recherche, ensuite nous donnerons quelques généralités et notations utilisées dans la démonstration du théorème puis, précédant le théorème, nous énoncerons les hypothèses assurant sa démonstration.

1.1. - Position du problème

Soit $\{Y_x | x \in \mathbb{R}^d\}$ une famille de variables aléatoires de \mathbb{R}^d , ayant pour loi de probabilité $\{P_x | x \in \mathbb{R}^d\}$, chacune dépendant du paramètre $x \in \mathbb{R}^d$.

Définissons $M(x) = E(Y_x)$ comme l'espérance mathématique de Y_x lorsque x est fixé.

La fonction $M(x)$ est supposée inconnue, mais il est possible d'effectuer quel que soit $x \in \mathbb{R}^d$, des observations indépendantes sur la variable aléatoire Y_x . Le problème consiste alors à rechercher la solution x_* de l'équation $M(x) = 0$ comme limite d'un processus d'itération stochastique à temps discret construit comme suit :

X_1 variable aléatoire constante de \mathbb{R}^d choisie arbitrairement,
 $X_{n+1} = X_n + a_n Y_n$ $n = 1, 2, \dots$, où $\{Y_n | n \geq 1\}$ est une suite de variables aléatoires de \mathbb{R}^d ayant respectivement pour loi de probabilité P_{X_n} et $\{a_n | n \geq 1\}$ une suite de nombres réels positifs.

1.2. - Généralités et notations

Soit $\{f_n | n \geq 1\}$ une famille d'applications de \mathbb{R}^d dans l'ensemble des réels positifs \mathbb{R}_+ , deux fois continûment différentiables.

Pour tout n entier ≥ 1 nous noterons :

$$\text{grad } f_n(x) = \left(\frac{\partial f_n(x)}{\partial x_1}, \dots, \frac{\partial f_n(x)}{\partial x_d} \right) \quad \text{le vecteur gradient de } f_n$$

au point $x = (x_1, \dots, x_d) \in \mathbb{R}^d$.

$$A_n(x) = \left(\frac{\partial^2 f_n(x)}{\partial x_i \partial x_j} \right)_{\substack{1 \leq i \leq d \\ 1 \leq j \leq d}} \quad \text{La matrice carrée } d \times d \text{ dont les éléments sont les dérivées partielles secondes de } f_n \text{ au point } x.$$

Pour $n \geq 1$ nous pouvons écrire :

$$f_n(x+ay) = f_n(x) + a \langle \text{grad } f_n(x), y \rangle + \frac{a^2}{2} \langle A_n(x+\theta_n ay), y, y \rangle$$

où θ_n désigne une valeur réelle de l'intervalle $(0,1)$.

Si nous substituons à y la variable aléatoire de \mathbb{R}^d Y_x de distribution de probabilité P_x , nous obtenons en prenant l'espérance mathématique :

$$E(f_n(x+ay)) = f_n(x) + a \langle \text{grad } f_n(x), M(x) \rangle + \frac{a^2}{2} E(\langle A_n(x+\theta_n a Y_x) Y_x, Y_x \rangle).$$

Par suite conformément à BLUM (3) et C. MACCHI (11) nous adopterons les notations suivantes :

$$\begin{aligned} U(n,x) &= \langle \text{grad } f_n(x), M(x) \rangle, & U_n &= U(n, X_n); \\ V_a(n,x) &= E(\langle A_n(x+\theta_n a Y_x) Y_x, Y_x \rangle), & V_n &= V_{a_n}(n, X_n); \\ \phi_a(n,x) &= U(n,x) + \frac{a}{2} V_a(n,x), & \phi_n &= \phi_{a_n}(n, X_n); \\ Z_n(x) &= f_n(x) & Z_n &= f_n(X_n). \end{aligned}$$

Enfin, nous noterons $B(x,\epsilon) = \{y \in \mathbb{R}^d \mid \|x-y\| < \epsilon\}$ la boule ouverte de centre x et de rayon ϵ et par $B^c(x,\epsilon)$ le complémentaire par rapport à \mathbb{R}^d de $B(x,\epsilon)$.

1.3. - Les hypothèses considérées

Nous ferons d'abord une hypothèse de continuité uniforme en n de la suite de fonctions $\{f_n | n \geq 1\}$.

$$C'0 \forall \epsilon > 0, \forall x \in \mathbb{R}^d, \exists \eta(\epsilon, x) > 0 : \|x-y\| \leq \eta(\epsilon, x) \Rightarrow \forall n \geq 1 : |f_n(x) - f_n(y)| < \epsilon.$$

Ensuite les hypothèses sont celles de C. MACCHI sauf en ce qui concerne

la substitution de la famille de fonctions $\{f_n | n \geq 1\}$ à une seule fonction f .

C0 chaque fonction f_n atteint son minimum en un point x_n^* .

C1. $\forall \epsilon > 0, \exists v(\epsilon) > 0 : \forall n \geq 1 : \inf_{x \in B^c(x_n^*, \epsilon)} (|f_n(x) - f_n(x_n^*)|) \geq v(\epsilon)$.

C2 $\forall \epsilon > 0, \exists \lambda(\epsilon) > 0, \exists \eta(\epsilon) > 0 : 0 < a \leq 2\lambda(\epsilon)$ et pour tout $n \geq 1 :$
 $\sup_{x \in B^c(x_n^*, \epsilon)} \phi_a(n, x) \leq -\eta(\epsilon)$.

C3 $\forall n \geq 1, \forall x \in \mathbb{R}^d : U(n, x) \leq 0$.

C4 $\forall n \geq 1, \exists k_n > 0$ et $\exists \epsilon_1 > 0$ tel que pour tout $x \in B(x_n^*, \epsilon_1)$ et pour tout $0 < a \leq 2\lambda(\epsilon_1) :$

$$V_a(n, x) \leq K_n.$$

C5 $\{a_n | n \geq 1\}$ suite de nombres réels non négatifs telle que :

$$\lim_{n \rightarrow \infty} a_n = 0, \sum_{n \geq 1} a_n = +\infty \text{ et } \sum_{n \geq 1} a_n^2 k_n < +\infty.$$

C6 $\forall n \geq 1, \forall x \in \mathbb{R}^d : |f_{n+1}(x) - f_n(x)| \leq \delta_n f_n(x) + \gamma_n$.

$$\text{où } \delta_n > 0, \gamma_n > 0, \sum_{n \geq 1} \delta_n < +\infty, \sum_{n \geq 1} \gamma_n < +\infty.$$

C7 $x_n^* \xrightarrow{n \rightarrow +\infty} x_*$.

1.4. - Théorème de convergence presque sûre

Sous les hypothèses C'0 jusqu'à C7, le processus d'itération stochastique $\{X_n | n \geq 1\}$ défini en 1.1. converge presque sûrement vers la solution x_* de l'équation $M(x) = 0$.

Démonstration

La démonstration comporte quatre parties :

1. La convergence presque sûre de Z_n .
2. L'existence de $f = \lim_{n \rightarrow \infty} f_n$.
3. La convergence presque sûre de Z_n vers $f(x_*)$.
4. La convergence presque sûre de X_n vers x_* .

Point 1 Convergence presque sûre de Z_n .

$$\begin{aligned} Z_{n+1} &= f_{n+1}(X_{n+1}) \leq (1 + \delta_n) f_n(X_n + a_n Y_n) + \gamma_n \\ &= (1 + \delta_n) f_n(X_n) + (1 + \delta_n) a_n \langle \text{grad} f_n(X_n), Y_n \rangle + (1 + \delta_n) \frac{a_n^2}{2} \langle A_n(X_n + \theta_n a_n Y_n) Y_n, Y_n \rangle + \gamma_n. \end{aligned}$$

Désignons par \mathcal{O}_n la tribu engendrée par les variables aléatoires

$X_1, X_2, \dots, X_n (n \geq 1)$ et prenons l'espérance conditionnelle à \mathcal{G}_n des deux côtés de l'inégalité précédente, il vient alors compte tenu des notations :

$$\mathbb{Q}_{E^n}(Z_{n+1}) \leq (1+\delta_n)Z_n + (1+\delta_n)a_n\phi_n + \gamma_n.$$

D'après l'hypothèse C5, pour $\epsilon_1 > 0$ fixé, il existe un entier $N_1 = N(\epsilon_1)$ tel que pour tout $n \geq N_1$ on ait : $0 < a_n \leq 2\lambda(\epsilon_1)$.

Montrons que pour $n \geq N_1$, nous avons :

$$\mathbb{Q}_{E^n}(Z_{n+1}) \leq (1+\delta_n)Z_n + (1+\delta_n)\frac{a_n^2}{2}K_n + \gamma_n.$$

Si $X_n \in B^c(x_n^*, \epsilon_1)$ alors d'après C2 : $\phi_n \leq 0$ et par suite le résultat annoncé.

Si $X_n \in B(x_n^*, \epsilon_1)$ alors d'après C3 et C4 :

$$\left. \begin{array}{l} U_n \leq 0 \\ V_n \leq K_n \end{array} \right\} \Rightarrow \phi_n = U_n + \frac{a_n}{2}V_n \leq U_n + \frac{a_n}{2}K_n \leq \frac{a_n}{2}K_n$$

d'où également le résultat.

Par conséquent la suite de variables aléatoires $\{Z_n | n \geq 1\}$ est "presque" une surmartingale et la limite $\lim_{n \rightarrow \infty} Z_n$ existe p.s. et est finie p.s sur tout l'espace.

Point 2. Existence de $f = \lim_n f_n$

Soit x fixé appartenant à \mathbb{R}^d .

Pour tout $\epsilon > 0$, il existe un entier positif $N(\epsilon)$ tel que si $n \geq N(\epsilon)$ alors $\sum_{n \geq N(\epsilon)} \delta_n \leq \epsilon$.

Soient p et $q \geq N(\epsilon)$ nous avons en supposant $p < q$:

$$|f_p(x) - f_q(x)| \leq \sum_{j=p}^{q-1} \delta_j f_j(x).$$

Or pour tout $k \geq 1$, nous avons d'après C6

$$f_{k+1}(x) \leq (1+\delta_k) f_k(x) + \gamma_k$$

et après itérations de cette inégalité jusqu'au rang 1, il vient :

$$f_{k+1}(x) \leq \prod_{j=1}^k (1+\delta_j) f_1(x) + \sum_{i=1}^k \prod_{j=i+1}^k (1+\delta_j) \gamma_i \leq B f_1(x) + C$$

avec $B = \lim_k \left(\prod_{j=1}^k (1+\delta_j) \right)$ et $C = \lim_k \left(\sum_{i=1}^k \prod_{j=i+1}^k (1+\delta_j) \gamma_i \right)$.

Par suite, en posant $F(x) = B f_1(x) + C$ on obtient :

$$|f_p(x) - f_q(x)| \leq \epsilon F(x).$$

et par conséquent $f_n(x)$ tend vers une limite $f(x)$.

Propriété

Nous allons montrer que si $u = \lim_n u_n$ alors $f(u) = \lim_n f_n(u_n)$.

Soit $\epsilon > 0$ fixé.

D'après C'0 : $\exists \eta(u, \epsilon) > 0 : \|u - y\| \leq \eta(u, \epsilon) \Rightarrow \forall n \geq 1 : |f_n(u) - f_n(y)| \leq \epsilon$.

$u = \lim_n u_n \Rightarrow \exists N_1(\epsilon) > 0 : n \geq N_1(\epsilon) : \|u - u_n\| \leq \eta(u, \epsilon)$.

Par suite pour $n \geq N_1(\epsilon) : |f_n(u) - f_n(u_n)| \leq \epsilon$.

$f = \lim_n f_n \Rightarrow \exists N_2(\epsilon, u) > 0 : n \geq N_2(\epsilon, u) : |f(u) - f_n(u)| \leq \epsilon$.

Finalement pour $n \geq N(\epsilon, u) = \max\{N_1(\epsilon), N_2(\epsilon, u)\}$ nous avons :

$$|f(u) - f_n(u_n)| \leq |f(u) - f_n(u)| + |f_n(u) - f_n(u_n)| \leq 2\epsilon.$$

Puisque ϵ peut être choisi arbitrairement petit, la propriété est démontrée.

Point 3. Convergence presque sûre de Z_n vers $f(x_*)$.

Pour tout $\epsilon > 0$ introduisons la variable aléatoire

$$W_n = \begin{cases} U_n & \text{si } X_n \in B(x_n^*, \epsilon) \\ \phi_n = U_n + \frac{a_n}{2} V_n & \text{si } X_n \in B^c(x_n^*, \epsilon). \end{cases}$$

D'après C2, C3 et C5, il existe pour $\epsilon > 0$ fixé un entier $N(\epsilon) > 0$ tel que pour $n \geq N(\epsilon)$ on ait $W_n \leq 0$ et par suite, à partir de ce rang $N(\epsilon)$ le terme général de la série $\sum_n a_n \prod_{j \geq n} (1 + \delta_j) E(W_n)$ est négatif ou nul.

Soit donc $\epsilon_1 > 0$ fixé et $N_1 = N(\epsilon_1)$ tel que pour $n \geq N_1$ on ait $W_n \leq 0$.

Nous allons montrer que pour $0 < \epsilon \leq \epsilon_1$ nous avons :

$$\sum_{i \geq N_1} a_i \prod_{j \geq i} (1 + \delta_j) E(W_i) > -\infty.$$

Soit donc $0 < \epsilon \leq \epsilon_1$. Nous pouvons écrire :

$$W_n = \phi_n - \frac{a_n}{2} V_n \mathbb{1}_{\{X_n \in B(x_n^*, \epsilon)\}}.$$

Après avoir pris l'espérance mathématique des deux membres de l'inégalité

$$E(Z_{n+1}) \leq (1 + \delta_n) Z_n + (1 + \delta_n) a_n \phi_n + \gamma_n$$

et itérer de N_1 jusqu'à $n \geq N_1$ nous obtenons

$$E(Z_{n+1}) \leq \prod_{j=N_1}^n (1 + \delta_j) E(Z_{N_1}) + \sum_{i=N_1}^n \prod_{j=i}^n (1 + \delta_j) a_i E(\phi_i) + \sum_{i=N_1}^n \prod_{j=i+1}^n (1 + \delta_j) \gamma_i.$$

Par suite :

$$\sum_{i=N_1}^n \prod_{j=i}^n (1+\delta_j) a_i E(W_i) \geq E(Z_{n+1}) - \prod_{j=N_1}^n (1+\delta_j) E(Z_{N_1}) - \sum_{i=N_1}^n \prod_{j=i+1}^n (1+\delta_j) \gamma_i - \sum_{i=N_1}^n \prod_{j=i}^n (1+\delta_j) \frac{a_i^2}{2} E(V_{i1} \{X_i \in B(x_i^*, \epsilon)\})$$

Or compte tenu de C4 et C5, nous avons :

$$\sum_{i=N_1}^n \prod_{j=i}^n (1+\delta_j) \frac{a_i^2}{2} E(V_{i1} \{X_i \in B(x_i^*, \epsilon)\}) \leq \sum_{i=N_1}^n \prod_{j=i}^n (1+\delta_j) \frac{a_i^2}{2} K_i \leq \sum_{i \geq N_1} \prod_{j \geq i} (1+\delta_j) \frac{a_i^2}{2} K_i = L(N_1) < \infty,$$

et en remarquant que Z_{n+1} est une variable aléatoire positive et en posant $B = \lim_n (\prod_{j=1}^n (1+\delta_j))$ il vient pour $0 < \epsilon \leq \epsilon_1$:

$$\sum_{i=N_1}^n \prod_{j=i}^n (1+\delta_j) a_i E(W_i) \geq -BE(Z_{N_1}) - B \sum_i \gamma_i - L(N_1) > -\infty.$$

Comme $\sum_{i \geq N_1} \prod_{j \geq i} (1+\delta_j) a_i = +\infty$ nous en concluons que pour $0 < \epsilon \leq \epsilon_1$

$$\liminf_n E(-W_n) = 0,$$

soit encore

$$\liminf_n (E(-U_n \{X_n \in B(x_n^*, \epsilon)\}) + E(-\phi_n \{X_n \in B^c(x_n^*, \epsilon)\})) = 0$$

et par conséquent :

$$\liminf_n E(-U_n \{X_n \in B(x_n^*, \epsilon)\}) + \liminf_n E(-\phi_n \{X_n \in B^c(x_n^*, \epsilon)\}) \leq 0.$$

Par suite, puisque pour $n \geq N(\epsilon) \geq N(\epsilon_1)$: $W_n \leq 0$, il y a compatibilité entre ces inégalités que si, en particulier :

$$\liminf_n E(-\phi_n \{X_n \in B^c(x_n^*, \epsilon)\}) = 0 \text{ si } 0 < \epsilon \leq \epsilon_1.$$

Il suffit de remarquer que d'après C2, pour $n \geq N(\epsilon) \geq N_1$ $\phi_n \leq -\eta(\epsilon)$ d'où

$$E(-\phi_n \{X_n \in B^c(x_n^*, \epsilon)\}) \geq \eta(\epsilon) P\{X_n \in B^c(x_n^*, \epsilon)\}$$

et par conséquent :

$$0 \leq \eta(\epsilon) \liminf_n P\{X_n \in B^c(x_n^*, \epsilon)\} \leq \liminf_n E(-\phi_n \{X_n \in B^c(x_n^*, \epsilon)\}) = 0.$$

Considérons maintenant le cas de $\epsilon_1 < \epsilon$.

$$\text{Si } X_n \in B^C(x_n^*, \epsilon) \Leftrightarrow \|X_n - x_n^*\| \geq \epsilon \Rightarrow \|X_n - x_n^*\| \geq \epsilon_1 \Leftrightarrow X_n \in B^C(x_n^*, \epsilon_1).$$

Par suite :

$$P\{X_n \in B^C(x_n^*, \epsilon)\} \leq P\{X_n \in B^C(x_n^*, \epsilon_1)\}.$$

Finalement nous venons de montrer :

$$\forall \epsilon > 0 : \liminf_n P\{X_n \in B^C(x_n^*, \epsilon)\} = \liminf_n P\{\|X_n - x_n^*\| \geq \epsilon\} = 0.$$

Il existe donc une sous-suite $\{n_k | k \geq 1\} \subset \{n | n \geq 1\}$ telle que :

$$\lim_k P\{\|X_{n_k} - x_{n_k}^*\| \geq \epsilon\} = 0,$$

et par suite, il existe une sous-suite $\{n_{k_j} | j \geq 1\} \subset \{n_k | k \geq 1\}$

telle que la sous-suite $\{X_{n_{k_j}} - x_{n_{k_j}}^* | j \geq 1\}$ converge p.s vers 0.

Puisque d'après C7 la sous-suite $\{(x_{n_{k_j}}^* - x_{n_{k_j}}^*) | j \geq 1\}$ converge vers 0, nous

en déduisons que la sous-suite de variables aléatoires $\{(X_{n_{k_j}} - x_{n_{k_j}}^*) | j \geq 1\}$ converge p.s vers 0.

D'après la propriété démontrée au point 2, pour tout ω appartenant à un ensemble de probabilité 1, on en déduit que la sous-suite

$\{(f_{n_{k_j}}(X_{n_{k_j}}(\omega)) - f(x_{n_{k_j}}^*)) | j \geq 1\}$ converge vers 0, ce qui est équivalent à

la convergence presque sûre vers $f(x_{n_{k_j}}^*)$ de la sous-suite de variables aléatoires $\{f_{n_{k_j}}(X_{n_{k_j}}) | j \geq 1\}$ et par conséquent d'après le point 1 à la

convergence presque sûre de $Z_n = f_n(X_n)$ vers $f(x_{n_{k_j}}^*)$.

Point 4

REMARQUE

Compte tenu de l'inégalité

$$|f_n(x_n^*) - f_n(X_n)| \leq |f_n(X_n) - f_{n_{k_j}}(X_{n_{k_j}})| + |f_{n_{k_j}}(X_{n_{k_j}}) - f(x_{n_{k_j}}^*)| + |f(x_{n_{k_j}}^*) - f_n(x_n^*)|$$

nous pouvons déterminer, d'après les points 2 et 3, pour $\epsilon > 0$ fixé et pour ω appartenant à un ensemble de probabilité 1 un rang $N(\omega, \epsilon)$ tel que, à partir de ce rang, chacun des trois termes du second membre de l'inégalité soit inférieur à $\frac{\epsilon}{3}$ et, par suite, nous concluons à la convergence presque sûre vers 0 de la suite

$$\{(f_n(X_n) - f_n(x_n^*)) | n \geq 1\}.$$

Convergence presque sûre de X_n vers $x_{n_{k_j}}^*$.

Supposons que $\{(X_n - x_n^*) | n \geq 1\}$ ne converge pas presque sûrement vers 0.

Désignons par Ω_1 un ensemble de probabilité 1 tel que pour tout $\omega \in \Omega_1$: $\{f_n(X_n(\omega)) - f_n(x_n^*) | n \geq 1\}$ converge vers 0 conformément à la remarque ci-dessus.

Il existe donc $\omega \in \Omega_1$, $\epsilon > 0$ et une sous-suite $\{n_k | k \geq 1\} \subset \{n | n \geq 1\}$ tels que pour tout k : $\|X_{n_k}(\omega) - x_{n_k}^*\| > \epsilon$.

Mais d'après C1 il existe $\nu(\epsilon) > 0$ tel que :

$$|f_{n_k}(X_{n_k}(\omega)) - f_{n_k}(x_{n_k}^*)| \geq \inf_{x \in B^c(x_{n_k}^*, \epsilon)} (|f_{n_k}(x) - f_{n_k}(x_{n_k}^*)|) \geq \nu(\epsilon) > 0,$$

et par conséquent la sous-suite $\{f_{n_k}(X_{n_k}(\omega)) - f_{n_k}(x_{n_k}^*) | k \geq 1\}$ ne converge pas vers 0, d'où la contradiction.

La convergence presque sûre vers 0 de la suite $\{(X_n - x_n^*) | n \geq 1\}$ et la condition C7 nous permettent alors de conclure à la convergence presque sûre vers x_* de la suite $\{X_n | n \geq 1\}$.

CHAPITRE 2

UN ALGORITHME D'ITERATION STOCHASTIQUE

POUR LES PROBLEMES D'OPTIMUM AVEC CONTRAINTE

Introduction

Les problèmes d'optimum avec contrainte se rencontrent très souvent en pratique, et notamment en Reconnaissance des Formes dans le domaine de la lecture automatique de caractères imprimés.

Dans ce chapitre nous nous proposons de résoudre un de ces problèmes. A cette fin, les résultats du chapitre 1 nous permettront de donner un algorithme d'itération stochastique convergeant presque sûrement vers l'optimum cherché.

2.1. - Minimum d'une fonction de regression $M(x)$ sur un ensemble convexe fermé C de \mathbb{R}^d .

2.1.1. - Position du problème

Nous supposons donné un espace probabilisé (Ω, \mathcal{Q}, P) et soit U une variable aléatoire de $\mathbb{R}^{d'}$ ($d' \geq 1$) dont nous désignerons la distribution de probabilité par P_U .

2.1.1.1. La fonction φ : hypothèses.

Soit φ une application de $\mathbb{R}^d \times \mathbb{R}^{d'}$ ($d, d' \geq 1$) dans \mathbb{R}_+ satisfaisant aux hypothèses suivantes :

- pour tout $u \in \mathbb{R}^{d'}$: $\varphi(\cdot, u)$ est convexe
- pour tout $x \in \mathbb{R}^d$: $\varphi(x, \cdot)$ est lebesgue-mesurable
- pour tout $u \in \mathbb{R}^{d'}$: $\varphi(\cdot, u)$ est différentiable et on notera $\text{grad } \varphi(x, u) = \left(\frac{\partial \varphi(x, u)}{\partial x_1}, \dots, \frac{\partial \varphi(x, u)}{\partial x_d} \right)$ le gradient de φ au point $x = (x_1, \dots, x_d) \in \mathbb{R}^d$.

Considérons alors la famille de variables aléatoires réelles positives $\{\varphi(x, U) \mid x \in \mathbb{R}^d\}$ dont nous supposons que pour tout $x \in \mathbb{R}^d$ $\|\text{grad } \varphi(x, U)\|^2$ et $\varphi(x, U)$ sont P_U -intégrables.

2.1.1.2. La fonction de régression $M(x)$:

Définissons la fonction de régression $M(x) = E(\varphi(x,U))$ comme l'espérance mathématique de $\varphi(x,U)$.

Il résulte des conditions imposées à φ que $M(x)$ est une application convexe de \mathbb{R}^d dans \mathbb{R}_+ que nous supposerons différentiable.

En tout point $x \in \mathbb{R}^d$ la relation $\text{grad } M(x) = E(\text{grad } \varphi(x,U))$ est vérifiée.

La distribution de probabilité P_U associée à U est inconnue, mais il est possible d'effectuer sur U des observations indépendantes.

Etant donné un sous-ensemble convexe fermé C contenu dans \mathbb{R}^d , le problème consiste à minimiser $M(x)$ sur C .

Nous supposerons donc qu'il existe un point x_* appartenant à C tel que pour tout $x \in C$ on ait $M(x_*) \leq M(x)$ et nous désignerons par $\pi_C(x)$ la projection d'un point x de \mathbb{R}^d sur C .

2.1.2. Condition d'unicité du minimum x_* de la fonction de régression $M(x)$

L'hypothèse supplémentaire P1 que nous introduisons maintenant assure l'unicité du minimum x_* de $M(x)$ sur C .

$$P1: \forall \rho > 0, \exists \delta(\rho) > 0 : \inf\{M(x) - M(x_*) \mid x \in B^C(x_*, \rho) \cap C\} \geq \delta(\rho).$$

Proposition 1

La condition P1 est équivalente à la condition P2 suivante :

$$P2 : \forall \nu > 0, \exists \eta(\nu) > 0, \exists \varepsilon(\nu) > 0 : \langle x - x_*, \gamma_* \rangle \geq -\eta(\nu), \|\sigma_*(x)\| \geq \nu \implies M(x) - M(x_*) + \|\gamma_*\| \|x - \pi_C(x)\| \geq \varepsilon(\nu) ;$$

dans laquelle nous avons posé :

$\gamma_* = \text{grad } M(x_*) \neq 0$, $\gamma_* = \frac{G_*}{\|G_*\|}$, $\sigma_*(x)$ désignant la projection de $x - x_*$ sur une droite passant par x_* et perpendiculaire à G_* .

Démonstration

Remarque : Nous avons supposé $G_* \neq 0$ sinon nous serions ramenés à la recherche du minimum de $M(x)$ sur \mathbb{R}^d .

Point 1 : $P_2 \Rightarrow P_1$.

Soit $\rho > 0$ fixé et $x \in B^C(x_*, \rho) \cap C$; nous avons alors $x \in C$ et $\rho \leq \|x - x_*\|$.

Choisissons $\eta > 0$ et $\nu > 0$ tels que $\nu^2 + \eta^2 \leq \rho^2$ et vérifiant P_2 . Cela étant toujours possible. D'après la définition de γ_* et de $\sigma_*(x)$ nous pouvons écrire :

$$x - x_* = \langle x - x_*, \gamma_* \rangle \gamma_* + \sigma_*(x),$$

soit

$$\|x - x_*\|^2 = (\langle x - x_*, \gamma_* \rangle)^2 + \|\sigma_*(x)\|^2.$$

Si $\|\sigma_*(x)\| \leq \nu$ alors nécessairement $\eta \leq \langle x - x_*, \gamma_* \rangle$ et par suite $M(x)$ étant convexe

$$M(x) \geq M(x_*) + \langle x - x_*, G_* \rangle \geq M(x_*) + \eta \|G_*\|.$$

Si $\|\sigma_*(x)\| > \nu$ alors d'après P_2

$$M(x) - M(x_*) + \|G_*\| \|x - \Pi_C(x)\| \geq \epsilon(\nu)$$

et puisque $x \in C$, nous obtenons $M(x) - M(x_*) \geq \epsilon(\nu)$.

Il nous suffit donc de prendre $\delta(\rho) = \inf\{\eta \|G_*\|, \epsilon(\nu)\}$.

Point 2 : $P_1 \Rightarrow P_2$.

Supposons que $P_1 \not\Rightarrow P_2$. Si P_2 n'est pas vraie alors :

$\exists \nu > 0, \forall \eta > 0, \forall \epsilon > 0, \exists x \in \mathbb{R}^d : \langle x - x_*, \gamma_* \rangle \geq -\eta, \|\sigma_*(x)\| \geq \nu$ et $M(x) - M(x_*) + \|G_*\| \|x - \Pi_C(x)\| < \epsilon$.

Mais alors, $M(x)$ étant convexe, nous avons

$$\epsilon > M(x) - M(x_*) \geq \langle x - x_*, G_* \rangle \geq -\eta \|G_*\|$$

et par suite : $\|x - \Pi_C(x)\| < \frac{\epsilon + \eta \|G_*\|}{\|G_*\|}$.

Posons $\xi = \frac{\epsilon + \eta \|G_*\|}{\|G_*\|}$ et remarquons que ϵ et η pouvant être pris arbitrairement petits, ξ peut être rendu inférieur à tout nombre $\delta > 0$ donné à l'avance et par suite $x = \Pi_C(x)$.

Posons alors $\|x - x_*\| = \rho$.

Si $\rho > 0$ alors d'après P_1 il existe $\delta(\rho) > 0$ tel que $M(x) - M(x_*) \geq \delta(\rho) > 0$ et par conséquent on obtient $0 < \delta(\rho) \leq M(x) - M(x_*) < \epsilon$ ce qui est absurde puisque ξ peut être pris aussi petit que l'on veut.

2.1.3. Problème équivalent : minimum d'une fonction $h(x)$ sur \mathbb{R}^d .

Nous allons montrer que le problème du minimum de $M(x)$ sur C (problème 1) est équivalent au problème du minimum, sur \mathbb{R}^d , d'une fonction $h(x)$ (problème 2).

Proposition 2

Nous pouvons trouver une application $\Psi_{C,\alpha}(x)$ de la forme $\alpha \|x - \Pi_C(x)\|$ ($\alpha \in \mathbb{R}_+$) telle que les deux problèmes suivants soient équivalents :

Problème 1 : Trouver $x_* \in C$ tel que pour tout $x \in C$ on ait $M(x_*) \leq M(x)$.

Problème 2 : Trouver $x'_* \in \mathbb{R}^d$ tel que pour tout $x \in \mathbb{R}^d$ on ait, en posant $h(x) = M(x) + \Psi_{C,\alpha}(x)$, $h(x'_*) \leq h(x)$.

Démonstration

Il nous suffit de prouver que si $x \notin C$ alors

(1) $M(x_*) < h(x)$ pour un choix convenable de α .

En effet $\Psi_{C,\alpha}(x) = \alpha \|x - \Pi_C(x)\| = 0$ sur C et par suite l'optimum de $h(x)$ s'obtiendra sur C où cette application coïncide avec $M(x)$.

M étant différentiable et convexe l'inégalité (1) est impliquée par

(2) $M_* < \Psi_{C,\alpha}(x) + \sup\{0, M_* + \langle G_*, x - x_* \rangle\}$
avec $M_* = M(x_*)$ et $G_* = \text{grad } M(x_*)$.

L'inégalité (2) est équivalente à

(3) $\frac{M_* - \sup\{0, M_* + \langle G_*, x - x_* \rangle\}}{\|x - \Pi_C(x)\|} < \alpha$.

Mais la définition de x_* entraîne que G_* est orthogonal à un plan d'appui à C en x_* et est orienté vers C . Il en résulte l'inégalité :

(4) $\langle G_*, x - x_* \rangle - \langle G_*, x - \Pi_C(x) \rangle = \langle G_*, \Pi_C(x) - x_* \rangle \geq 0$

qui entraîne : $\langle G_*, x - \Pi_C(x) \rangle \leq \langle G_*, x - x_* \rangle$

et par suite

$$\frac{M_* - \sup\{0, M_* + \langle G_*, x - x_* \rangle\}}{\|x - \Pi_C(x)\|} \leq \frac{M_* - \sup\{0, M_* + \langle G_*, x - \Pi_C(x) \rangle\}}{\|x - \Pi_C(x)\|}$$

L'inégalité (3) est donc impliquée par :

(5) $\frac{M_* - \sup\{0, M_* + \langle G_*, x - \Pi_C(x) \rangle\}}{\|x - \Pi_C(x)\|} < \alpha$

qui est équivalent à

(6) $\inf \left\{ \frac{M_*}{\|x - \Pi_C(x)\|}, - \frac{\langle G_*, x - \Pi_C(x) \rangle}{\|x - \Pi_C(x)\|} \right\} < \alpha$.

Enfin (6) est impliquée par l'inégalité

(7) $\inf \left\{ \frac{M_*}{\|x - \Pi_C(x)\|}, \|G_*\| \right\} < \alpha$.

Un choix de α est donc donné par l'inégalité : $\|G_*\| < \alpha$.

2.1.3.2. Unicité du minimum x_* de $h(x)$ sur \mathbb{R}^d .

Compte tenu de l'hypothèse d'unicité du minimum x_* de $M(x)$ sur C formulée en 2.1.2. nous allons montrer dans la proposition 3 suivante que x_* est aussi le minimum unique de $h(x)$ sur \mathbb{R}^d en démontrant que $h(x)$ vérifie une condition analogue à la condition P1 exposée en 2.1.2..

Proposition 3

$$\forall \rho > 0, \exists \delta(\rho) > 0 : \inf\{h(x) - h(x_*) \mid x \in B^C(x_*, \rho)\} \geq \delta(\rho).$$

Démonstration

Nous savons d'après la proposition 2 que x_* est un minimum de $h(x)$, par suite, pour tout $x \in \mathbb{R}^d$, nous avons : $h(x) - h(x_*) \geq 0$.

Soit $\rho > 0$ fixé et $x \in B^C(x_*, \rho)$. Choisissons v et η conformément à la proposition 1(2.1.2.) et vérifiant $v^2 + \eta^2 \leq \rho^2$.

Par hypothèse $x \in B^C(x_*, \rho) \Leftrightarrow \rho \leq \|x - x_*\| \Rightarrow v^2 + \eta^2 \leq \|x - x_*\|^2$.

Si $|\langle x - x_*, \gamma_* \rangle| \leq \eta$ alors $\|\sigma_*(x)\| \geq v$.

De la relation $h(x) - h(x_*) = M(x) - M(x_*) + \alpha \|x - \Pi_C(x)\|$ en tenant compte de la convexité de $M(x)$, du choix de α et de la proposition 1 nous obtenons

$$\begin{aligned} h(x) - h(x_*) &\geq \sup\{\epsilon(v) - \|\sigma_*\| \cdot \|x - \Pi_C(x)\|, -\eta(v)\|\sigma_*\| + \|\sigma_*\| \cdot \|x - \Pi_C(x)\|\} \\ &= \epsilon(v) 1_{\{\|x - \Pi_C(x)\| \leq \xi\}} + (-\eta(v)\|\sigma_*\| + \|\sigma_*\| \cdot \|x - \Pi_C(x)\|) 1_{\{\|x - \Pi_C(x)\| > \xi\}} \end{aligned}$$

avec $\xi = \frac{\epsilon + \eta \|\sigma_*\|}{\|\sigma_*\|}$

et par suite

$$h(x) - h(x_*) \geq \epsilon(v).$$

Si $\langle x - x_*, \gamma_* \rangle < -\eta$ alors de la relation

$$\langle \gamma_*, x - \Pi_C(x) \rangle = \langle \gamma_*, x - x_* \rangle + \langle \gamma_*, x_* - \Pi_C(x) \rangle$$

compte tenu de l'inégalité $\langle \gamma_*, x_* - \Pi_C(x) \rangle \leq 0$

nous obtenons $\langle \gamma_*, x - \Pi_C(x) \rangle \leq \langle \gamma_*, x - x_* \rangle$.

Soit finalement

$$0 < \eta < -\langle \gamma_*, x - x_* \rangle \leq -\langle \gamma_*, x - \Pi_C(x) \rangle \leq \|x - \Pi_C(x)\|.$$

De plus α peut être mis sous la forme $\alpha = \|\sigma_*\| + \lambda$ avec $\lambda > 0$.

Il vient alors

$$h(x) - h(x_*) = M(x) - M(x_*) + \alpha \|x - \Pi_C(x)\| \geq \langle \sigma_*, x - x_* \rangle + (\|\sigma_*\| + \lambda) \|x - \Pi_C(x)\| \geq \lambda \eta.$$

Enfin si $\langle \gamma_*, x - x_* \rangle > \eta$ alors

$$h(x) - h(x_*) \geq M(x) - M(x_*) \geq \langle \sigma_*, x - x_* \rangle > \eta \|\sigma_*\|.$$

Nous pouvons donc prendre $\epsilon(\rho) = \inf\{\epsilon(v), \lambda \eta, \eta \|\sigma_*\|\}$.

2.1.3.3. Remarques sur $h(x)$

Nous venons de voir, dans ce qui précède, que la recherche du minimum x_* de $M(x)$ sur C est équivalente à celle du minimum x_* de $h(x)$ sur \mathbb{R}^d . Afin d'obtenir ce minimum x_* notre but est de mettre en oeuvre une procédure d'itération stochastique ; or pour ce faire, il faudrait que $h(x)$ soit différentiable en tous les points de \mathbb{R}^d , ce qui n'est pas le cas puisque la fonction $\Psi_{C,\alpha}(x) = \alpha \|x - \Pi_C(x)\|$ n'est pas différentiable sur la frontière du convexe C et, par suite, le gradient de $h(x)$ n'est pas défini en ces points.

Nous allons surmonter cette difficulté en introduisant une suite de fonctions de pénalité compatibles avec le théorème d'itération stochastique démontré dans le chapitre 1.

2.2. Nouvelle approche du problème d'optimum de la fonction de régression $M(x)$

2.2.1. Généralités

Puisque la fonction $h(x)$ n'est pas différentiable sur la frontière de C , nous allons lui substituer une suite de fonctions $\{h_n | n \geq 1\}$ qui seront différentiables et sur lesquelles nous feront des hypothèses appropriées pour que la suite des optima auxquels elles donneront lieu, converge vers la solution x_* du problème initial.

Comme c'est la fonction de pénalité $\Psi_{C,\alpha}(x) = \alpha \|x - \Pi_C(x)\|$ qui n'est pas différentiable sur la frontière du convexe C , nous allons lui substituer une suite de fonctions de pénalité $\{g_{n,\alpha} | n \geq 1\}$ que nous allons maintenant définir et dont nous indiquerons les hypothèses qu'elles devront satisfaire afin de générer la suite de fonctions $\{h_n | n \geq 1\}$.

2.2.2. Définition et propriétés de la suite de fonctions de pénalité $\{g_{n,\alpha} | n \geq 1\}$

Nous définissons une suite de fonctions de pénalité $g_{n,\alpha}$ telle que pour tout n entier ≥ 1 :

- $g_{n,\alpha}$ est une application de \mathbb{R}^d dans \mathbb{R}_+ ; convexe et différentiable.
- $g_{n,\alpha}(x) \leq g_{n+1,\alpha}(x)$ pour tout $x \in \mathbb{R}^d$.
- $0 \leq \Psi_{C,\alpha}(x) - g_{n,\alpha}(x) \leq \epsilon_n$ pour tout $x \in \mathbb{R}^d$, où $\{\epsilon_n | n \geq 1\}$ est une suite de nombres réels positifs décroissante vers 0.

En tout point x de \mathbb{R}^d nous pouvons donc écrire $\Psi_{C,\alpha}(x) = \liminf_n g_{n,\alpha}(x)$.

2.2.3. Exemple de suite de fonctions de pénalité

Proposition 4

La suite de fonctions $g_{n,\alpha}$ définie pour n entier ≥ 1 et pour tout $x \in \mathbb{R}^d$ par :

$$g_{n,\alpha}(x) = \begin{cases} 0 & \text{si } x \in C \\ \frac{\alpha}{2t_n} \|x - \Pi_C(x)\|^2 & \text{si } \|x - \Pi_C(x)\| \leq t_n \\ \alpha \left(\|x - \Pi_C(x)\| - \frac{t_n}{2} \right) & \text{si } \|x - \Pi_C(x)\| > t_n \end{cases}$$

où $\{t_n | n \geq 1\}$ est une suite décroissante de nombres réels positifs dont la limite est 0 satisfait aux conditions exigées en 2.2.2..

Démonstration

Les fonctions $g_{n,\alpha}(x)$ sont manifestement positives et convexes. Elles sont différentiables et un calcul simple montre que :

$$\text{grad} g_{n,\alpha}(x) = \alpha \frac{(x - \Pi_C(x))}{\max(\|x - \Pi_C(x)\|, t_n)} .$$

Il est clair que pour tout $x \in \mathbb{R}^d$ nous avons $g_{n,\alpha}(x) \leq g_{n+1,\alpha}(x)$.

Montrons enfin que si $x \in \mathbb{R}^d$ nous avons

$$0 \leq \Psi_{C,\alpha}(x) - g_{n,\alpha}(x) < \varepsilon_n .$$

Si $x \in C$: $\Psi_{C,\alpha}(x) = g_{n,\alpha}(x) = 0$.

Si $\|x - \Pi_C(x)\| \leq t_n$ alors
 $0 \leq \Psi_{C,\alpha}(x) - g_{n,\alpha}(x) < \frac{\alpha t_n}{2}$.

Si $\|x - \Pi_C(x)\| > t_n$ alors
 $0 \leq \Psi_{C,\alpha}(x) - g_{n,\alpha}(x) = \frac{\alpha t_n}{2}$.

Il nous suffit donc de prendre $\varepsilon_n = \frac{\alpha}{2} t_n$ pour obtenir $\lim_{n \rightarrow \infty} \varepsilon_n = 0$.

2.2.4. La suite de fonctions $\{h_n | n \geq 1\}$

2.2.4.1. Définition et propriétés

Pour chaque entier $n \geq 1$ et pour tout point x de \mathbb{R}^d nous définissons

$$h_n(x) = M(x) + g_{n,\alpha}(x) .$$

Compte tenu des propriétés vérifiées par les fonctions de

pénalité $g_{n,\alpha}$ et des hypothèses faites sur $M(x)$ nous en déduisons que la suite $\{h_n | n \geq 1\}$ possède les propriétés suivantes :

Pour tout entier $n \geq 1$:

- h_n est une application de \mathbb{R}^d dans \mathbb{R}_+ , convexe et différentiable.
- $h_n(x) \leq h_{n+1}(x)$ pour tout $x \in \mathbb{R}^d$.
- $h(x) = \lim_{n \rightarrow \infty} h_n(x)$ pour tout $x \in \mathbb{R}^d$.

2.2.4.2. La suite $\{x_n^* | n \geq 1\}$ des minima de la suite $\{h_n | n \geq 1\}$

Chacune des fonctions $h_n(x)$ tendant vers $+\infty$ avec $\|x\|$ nous savons que l'ensemble \mathcal{M}_n des minima associés à $h_n(x)$ n'est pas vide.

Nous choisissons pour chaque entier $n \geq 1$ un élément x_n^* appartenant à \mathcal{M}_n et considérons la suite $\{x_n^* | n \geq 1\}$ des minima ainsi choisis.

2.2.4.3. Relation entre une suite $\{x_n^* | n \geq 1\}$ et $h(x)$.

Proposition 5

Une suite $\{x_n^* | n \geq 1\}$ appartenant à $\prod_{n \geq 1} \mathcal{M}_n$, la suite $\{\epsilon_n | n \geq 1\}$ de nombres réels introduite dans la proposition 4, la fonction $h(x)$ et son minimum x_* vérifient la condition

$$0 \leq h(x_n^*) - h(x_*) \leq \epsilon_n \quad n = 1, 2, \dots$$

Démonstration

D'après la proposition 4 nous savons que pour tout n entier ≥ 1 et pour tout $x \in \mathbb{R}^d$ nous avons :

$$0 \leq \Psi_{C,\alpha}(x) - g_{n,\alpha}(x) \leq \epsilon_n$$

Par suite

$$0 \leq h(x_n^*) - h(x_*) = h(x_n^*) - M(x_*)$$

et puisque x_n^* est un minimum de $h_n(x)$ nous obtenons

$$0 \leq h(x_n^*) - h(x_*) \leq h(x_n^*) - h_n(x_n^*) = \Psi_{C,\alpha}(x_n^*) - g_{n,\alpha}(x_n^*) \leq \epsilon_n$$

En particulier puisque $\lim_{n \rightarrow \infty} \epsilon_n = 0$, nous obtenons

$$h(x_*) = \lim_{n \rightarrow \infty} h(x_n^*)$$

2.2.4.4. Convergence de la suite $\{x_n^* | n \geq 1\}$ vers x_* .

Proposition 6

Une suite $\{x_n^* | n \geq 1\} \in \prod_{n \geq 1} \mathcal{M}_n$ converge vers x_* .

Démonstration

Supposons que la suite $\{x_n^* | n \geq 1\} \in \prod_{n \geq 1} \mathcal{M}_n$ ne converge pas vers x_* . Alors il existe $\rho > 0$ et une sous-suite d'entiers $\{n_\ell | \ell \geq 1\} \subset \{n | n \geq 1\}$ telle que pour tout $\ell \geq 1$ on ait $\|x_{n_\ell}^* - x_*\| > \rho$.

Mais d'après la proposition 3 il existe $\delta(\rho) > 0$ tel que pour tout $x \in B^C(x_*, \rho)$ on ait : $0 < \delta(\rho) \leq h(x) - h(x_*)$.

En particulier nous aurions pour tout $l \geq 1$:

$0 < \delta(\rho) \leq h(x_{n_l}^*) - h(x_*)$ et par suite $\{h(x_{n_l}^*) \mid l \geq 1\}$ ne convergerait pas vers $h(x_*)$ ce qui est contraire au résultat de la proposition 5.

2.3. Un algorithme d'itération stochastique convergeant presque sûrement vers x_*

Nous sommes maintenant en mesure d'utiliser les résultats du chapitre 1 pour construire un algorithme d'itération stochastique convergeant presque sûrement vers x_* .

2.3.1. Définition de l'algorithme d'itération stochastique

Définissons le processus d'itération stochastique à temps discret suivant :

X_1 point arbitraire de \mathbb{R}^d ,

$$X_{n+1} = X_n + a_n Y_n \quad n = 1, 2, \dots$$

où $\{a_n \mid n \geq 1\}$ est une suite de nombres réels non négatifs vérifiant

$$\sum_{n \geq 1} a_n = +\infty, \quad \sum_{n \geq 1} a_n^2 < +\infty$$

et pour $n \geq 1$ $Y_n(x) = -(\text{grad } \varphi(x, U) + \text{grad } g_{n, \alpha}(x))$ est une variable aléatoire de \mathbb{R}^d .

Afin d'être dans les conditions du théorème 1.4., chapitre 1, nous prendrons comme famille d'applications $\{f_n \mid n \geq 1\}$ la famille définie par : pour tout $x \in \mathbb{R}^d$ et pour $n \geq 1$: $f_n(x) = \|x - x_n^*\|^2$ où $x_n^* \in \mathcal{M}_n$.

Nous donnons dans le théorème suivant les hypothèses permettant de conclure à la convergence presque sûre du processus ainsi défini vers le minimum x_* de la fonction de régression $M(x)$ sur l'ensemble convexe fermé C .

2.3.2. Théorème

Si L et L_i ($i=0, 1, 2$) sont des constantes positives telles que pour tout $x \in \mathbb{R}^d$ et pour tout entier $n \geq 1$

$$L \|x - x_n^*\|^2 \leq \langle x - x_n^*, \text{grad } h_n(x) \rangle$$

$$E \|Y_n(x)\|^2 \leq \sum_{i=0}^2 L_i \|x - x_n^*\|^i$$

et si $\sum_{n \geq 1} \|x_n^* - x_{n+1}^*\| < +\infty$

alors le processus stochastique à temps discret défini en 2.3.1. converge presque sûrement vers x_* .

Démonstration

Elle consiste à vérifier que nous sommes dans les conditions d'application du théorème 1.4. - Chapitre 1.

Pour tout $n \geq 1$ et tout $x \in \mathbb{R}^d$ nous avons pris $f_n(x) = \|x - x_n^*\|^2$ par suite en utilisant les notations du chapitre 1-

$$\text{grad} f_n(x) = 2(x - x_n^*),$$

$$A_n(x) = 2 I_{(d,d)} \quad (I_{(d,d)} \text{ matrice carrée unité d'ordre } d).$$

Vérifions maintenant les conditions C'0 jusqu'à C7 du chapitre 1.

C'0 soient $\epsilon > 0$, x et y deux points de \mathbb{R}^d . Nous avons

$$|f_n(x) - f_n(y)| = |\langle x - x_n^* + y - x_n^*, x - y \rangle| \leq \|x - y\|^2 + 2\|x - y\| (\|x\| + \sup \|x_n^*\|)$$

et par suite il existe $\eta(\epsilon, x) > 0$ tel que $\|x - y\| \leq \eta(\epsilon, x) \Rightarrow |f_n(x) - f_n(y)| \leq \epsilon$.

C0 Il est évident que chaque fonction f_n atteint son minimum en x_n^* .

$$C1 \quad \forall \epsilon > 0, \exists v(\epsilon) > 0 : \forall n \geq 1 \quad \inf \{ |f_n(x) - f_n(x_n^*)| \mid x \in B^C(x_n^*, \epsilon) \} \geq v(\epsilon).$$

Soient $\epsilon > 0$ fixé et $x \in B^C(x_n^*, \epsilon)$. Pour tout $n \geq 1$ nous avons

$$|f_n(x) - f_n(x_n^*)| = \|x - x_n^*\|^2 \geq \epsilon^2$$

et il nous suffit de prendre $v(\epsilon) = \epsilon^2 > 0$.

$$C3 \quad \forall n \geq 1, \forall x \in \mathbb{R}^d : U(n, x) \leq 0.$$

Par définition

$$U(n, x) = E(\langle \text{grad} f_n(x), Y_n(x) \rangle) = 2 \langle x - x_n^*, E(Y_n(x)) \rangle$$

$$\text{or } E(Y_n(x)) = -E(\text{grad } \varphi(x, U) + \text{grad} g_{n,d}(x)) = -\text{grad } h_n(x)$$

$$\text{d'où } U(n, x) = -2 \langle x - x_n^*, \text{grad } h_n(x) \rangle.$$

Mais h_n étant convexe et admettant x_n^* pour minimum nous avons

$$0 \leq h_n(x) - h_n(x_n^*) \leq \langle x - x_n^*, \text{grad } h_n(x) \rangle$$

et par suite $U(n, x) \leq 0$.

$$C4. \quad \forall n \geq 1, \exists K_n > 0 \text{ et } \exists \epsilon_1 > 0 \text{ tel que pour tout } x \in B(x_n^*, \epsilon_1) \text{ et pour tout}$$

$$0 < a \leq 2\lambda(\epsilon_1) : V_a(n, x) \leq K_n.$$

Nous avons

$$V_a(n, x) = E(\langle A_n(x + \theta a y_n(x)) Y_n(x), Y_n(x) \rangle) = E(\|Y_n(x)\|^2) \leq \sum_{i=0}^2 L_i \|x - x_n^*\|^i.$$

Soient $\epsilon_1 > 0$ fixé, pour tout $x \in B(x_n^*, \epsilon_1)$ nous avons pour tout a

$$V_a(n, x) \leq \sum_{i=0}^2 L_i \epsilon_1^2 = K_n.$$

$$C2 \quad \forall \epsilon > 0, \exists \lambda(\epsilon) > 0, \exists \eta(\epsilon) > 0 : 0 < a \leq 2\lambda(\epsilon_1) \text{ et pour tout } n \geq 1$$

$$\sup_{x \in B^C(x_n^*, \epsilon)} \phi_a(n, x) \leq -\eta(\epsilon).$$

$$x \in B^C(x_n^*, \epsilon)$$

$$\text{Par définition } \phi_a(n, x) = U(n, x) + \frac{a}{2} V_a(n, x).$$

Compte tenu des hypothèses sur $U(n, x)$ et $V_a(n, x)$ il vient

$$\phi_a(n, x) \leq -\frac{L}{2} \|x - x_n^*\|^2 + \frac{a}{2} \sum_{i=0}^2 L_i \|x - x_n^*\|^i.$$

Soit maintenant $x \in B^C(x_n^*, \epsilon)$. Nous obtenons

$$\phi_a(n, x) \leq \|x - x_n^*\|^2 \left(\frac{a}{2} \left(L_2 + \frac{L_1}{\epsilon} + \frac{L_0}{\epsilon^2} \right) - \frac{L}{2} \right).$$

Prenons par exemple $\lambda(\epsilon) = \frac{L}{2 \left(L_2 + \frac{L_1}{\epsilon} + \frac{L_0}{\epsilon^2} \right)}$

nous avons alors

$$\phi_a(n, x) \leq \|x - x_n^*\|^2 \frac{L}{2} \left(\frac{a}{2\lambda(\epsilon)} - 1 \right)$$

d'où pour $\eta(\epsilon) = \frac{L}{2} \epsilon^2$

$$\phi_a(n, x) \leq -\eta(\epsilon).$$

C5 Cette condition C5 est vérifiée en prenant par exemple

$$a_n = \frac{1}{n} \quad n=1, 2, \dots$$

C6 $\forall n \geq 1, \forall x \in \mathbb{R}^d : |f_{n+1}(x) - f_n(x)| \leq \delta_n f_n(x) + \gamma_n$,

où $0 < \delta_n, 0 < \gamma_n$ et $\sum_{n \geq 1} \delta_n < +\infty, \sum_{n \geq 1} \gamma_n < +\infty$.

Nous avons

$$|f_{n+1}(x) - f_n(x)| \leq 2 \|x - x_n^*\| \|x_n^* - x_{n+1}^*\| + \|x_n^* - x_{n+1}^*\|^2.$$

Si $\|x - x_n^*\| \geq 1$ alors

$$|f_{n+1}(x) - f_n(x)| \leq 2 \|x_n^* - x_{n+1}^*\| \|x - x_n^*\|^2 + \|x_n^* - x_{n+1}^*\|^2.$$

Si $\|x - x_n^*\| < 1$ alors

$$|f_{n+1}(x) - f_n(x)| \leq 2 \|x_n^* - x_{n+1}^*\| \|x - x_n^*\|^2 + 2 \|x_n^* - x_{n+1}^*\| + \|x_n^* - x_{n+1}^*\|^2.$$

Dans les deux cas nous obtenons :

$$|f_{n+1}(x) - f_n(x)| \leq \delta_n f_n(x) + \gamma_n$$

avec $\delta_n = 2 \|x_n^* - x_{n+1}^*\|$ et $\gamma_n = 2 \|x_n^* - x_{n+1}^*\| + \|x_n^* - x_{n+1}^*\|^2$.

Par hypothèse $\sum_{n \geq 1} \delta_n < +\infty$ et par suite $\sum_{n \geq 1} \gamma_n < +\infty$.

CHAPITRE 3

CALCUL DES MASQUES POUR L'IDENTIFICATION

AUTOMATIQUE DE CARACTERES IMPRIMES

Introduction

Nous nous proposons dans ce chapitre de résoudre un problème de calcul de masques pour la lecture automatique de caractères imprimés dans un style donné.

La résolution de ce problème fera appel aux résultats obtenus dans le chapitre 2.

3.1. - Généralités et notations

Soit $\mathcal{A} = \{U_1, U_2, \dots, U_m\}$ un alphabet comportant m classes de lettres. Nous supposons que chaque lettre de chacune des m classes est décrite par un point de \mathbb{R}^d . Par exemple, une lettre $U_{i,j}$ d'une classe donnée i est inscrite dans un carré quadrillé et est repérée par rapport au centre du quadrillage par l'un quelconque des déplacements suivants : N, N.E, E, S.E, S, S.O, O, N.O, où N désigne le Nord etc..., étant bien entendu que la lettre peut être centrée dans le quadrillage ce qui correspond au déplacement nul.

Nous supposons qu'à la description brute de la lettre s'ajoute dans chaque composante du quadrillage un bruit aléatoire.

Chaque composante de la description d'une lettre est donnée par le taux de blanc dans un carré élémentaire du quadrillage.

Une lettre dépend donc aléatoirement

- de la classe à laquelle elle appartient et qui est caractérisée par sa description brute.
- d'un déplacement à l'intérieur du quadrillage dans l'une des directions données précédemment.
- d'un bruit qui vient perturber chacune des cases de la description.

Ainsi une lettre sera considérée comme une variable aléatoire de \mathbb{R}^d que nous noterons U , et dont nous désignerons la distribution de probabilité par P_U .

Le problème de décision lié à l'identification des lettres de l'alphabet introduit un problème d'optimisation que nous pouvons formuler par exemple de la manière suivante :

On dispose d'éléments X_i ($i = 1, 2, \dots, m$) de \mathbb{R}_+^d ($d \geq 1$) tels que $\langle X_i, U_{i,0} \rangle = 1$ où $U_{i,0}$ désigne la lettre de la classe i correspondant au déplacement nul et on minimise l'expression $E(\langle X_i, U \rangle^2)$ où U représente une lettre n'appartenant pas à la classe i .

Les éléments X_i sont appelés masques ce qui se justifie par le fait que technologiquement on peut envisager de les réaliser par un rectangle quadrillé.

Le taux de transmission lumineuse de chaque carré élémentaire sera donné par la composante de X_i correspondante de manière que la quantité de lumière traversant le masque déterminé par X_i soit proportionnelle à $\langle X_i, U \rangle$ quand est présentée la lettre U . Après focalisation par une lentille cette quantité est mesurée par une cellule photoélectrique.

Nous supposons que les lettres U sont des variables aléatoires à valeurs dans \mathbb{R}_+^d et nous pouvons toujours prendre $\|U_{i,0}\| = 1$ pour le calcul du masque X_i .

3.2. Position du problème

3.2.1. La fonction φ

Nous définissons l'application φ de $\mathbb{R}^d \times \mathbb{R}^d$ dans \mathbb{R}_+ par : pour tout $x \in \mathbb{R}^d$, $u \in \mathbb{R}^d$: $\varphi(x, u) = \langle x, u \rangle^2$.

Lorsque on remplace u par une variable aléatoire U de \mathbb{R}^d on notera $\varphi(x, U) = \langle x, U \rangle^2$.

3.2.2. L'ensemble convexe fermé C de \mathbb{R}_+^d

Pour le calcul du masque X_i l'ensemble convexe fermé C de \mathbb{R}_+^d est défini comme l'intersection de l'orthant positif \mathbb{R}_+^d et de l'hyperplan $\mathcal{H} = \{y \mid \langle y, U_{i,0} \rangle = 1\}$.

C est un ensemble convexe fermé non vide car il contient $U_{i,0}$ et puisque C est borné c'est un compact de \mathbb{R}_+^d .

3.2.3. La fonction de régression M(x)

La fonction de régression M(x) est définie comme l'espérance mathématique de $\varphi(x,U)$ où U est une lettre n'appartenant pas à la classe i lorsqu'on calcule le masque correspondant X_i .

Pour le calcul du masque X_i nous introduisons donc le problème d'optimisation suivant :

Minimiser $M(x) = E(\langle x, U \rangle^2)$ sur l'ensemble convexe C défini en 3.2.2. où U désigne une lettre n'appartenant pas à la classe i et dont la distribution de probabilité P_U est inconnue.

Nous supposons que $E(\|U\|^4) < +\infty$ et que les matrices $E(UU^t)$ et $E(UU^t UU^t)$ sont définies positives.

3.3. Vérification des hypothèses

On vérifie facilement que les fonctions φ et M ainsi définies vérifient les hypothèses du chapitre 2 (2.1.1.1. et 2.1.1.2.).

3.3.1. Unicité du minimum de M(x) sur C

C étant un compact de \mathbb{R}^d il existe un point $x_* \in C$ tel que pour tout $x \in C$ on ait $M(x_*) \leq M(x)$.

Montrons maintenant que M(x) vérifie l'hypothèse P1 du chapitre 2 (2.1.2) c'est-à-dire :

$$\forall \rho > 0, \exists \epsilon(\rho) > 0 : \inf\{M(x) - M(x_*) \mid x \in B^C(x_*, \rho) \cap C\} \geq \epsilon(\rho).$$

Soit donc $x \in B^C(x_*, \rho) \cap C$. Nous avons
 $M(x) - M(x_*) = x^t A x - x_*^t A x_* = (x - x_*)^t A (x - x_*) + 2x_*^t A (x - x_*)$
 où $A = E(UU^t)$.

Or puisque $x \in C : 2x_*^t A (x - x_*) = \langle x - x_*, \text{grad } M(x_*) \rangle \geq 0$
 et par suite

$$M(x) - M(x_*) \geq (x - x_*)^t A (x - x_*) \geq \lambda \|x - x_*\|^2$$

où λ désigne la plus petite valeur propre de A ; ce qui implique $M(x) - M(x_*) \geq \lambda \rho^2$
 et il nous suffit de prendre $\epsilon(\rho) = \lambda \rho^2$.

D'après les résultats obtenus dans le chapitre 2 nous concluons que le problème du minimum de la fonction de régression M(x) sur l'ensemble convexe fermé C est équivalent à celui du minimum de h(x) sur \mathbb{R}^d , où h(x) est la fonction introduite en 2.1.3. chapitre 2 et, par conséquent, en utilisant les suites de fonctions $\{g_{n,\alpha} \mid n \geq 1\}$ et $\{h_n \mid n \geq 1\}$ définies au chapitre 2 en 2.2.2. et 2.2.4. respectivement il nous reste à vérifier que nous sommes

dans les conditions d'application du théorème de convergence presque sûre 2.3.2., chapitre 2 pour obtenir un algorithme d'itération stochastique convergeant presque sûrement vers le minimum x_* de $M(x)$ sur C .

3.3.3. - Les hypothèses du chapitre 2

Ce sont les hypothèses considérées dans le théorème de convergence presque sûre énoncé en 2.3.2.

$$Y_n(x) = - (\text{grad } \varphi(x, U) + g_{n,\alpha}(x)) = - (2\langle x, U \rangle U + \alpha \frac{(x - \Pi_C(x))}{\max\{\|x - \Pi_C(x)\|, t_n\}})$$

et en utilisant pour $n = 1, 2, \dots$

$$f_n(x) = \|x - x_n^*\|^2$$

où $x_n^* \in \mathcal{M}_n$ (\mathcal{M}_n ensemble des minima de $h_n(x)$ sur \mathbb{R}^d),

nous obtenons pour tout x de \mathbb{R}^d

$$\begin{aligned} \langle x - x_n^*, \text{grad} h_n(x) \rangle &= \langle x - x_n^*, \text{grad} M(x) - \text{grad} M(x_n^*) \rangle + \langle x - x_n^*, \text{grad} g_{n,\alpha}(x) \rangle \\ &\quad + \langle x - x_n^*, \text{grad} M(x_n^*) \rangle \end{aligned}$$

or, $x_n^* \in \mathcal{M}_n \Leftrightarrow \text{grad} h_n(x_n^*) = 0 \Leftrightarrow \text{grad} M(x_n^*) = - \text{grad} g_{n,\alpha}(x_n^*)$

et compte tenu de la convexité de $g_{n,\alpha}(x)$

$$\begin{aligned} \langle x - x_n^*, \text{grad} h_n(x) \rangle &\geq \langle x - x_n^*, \text{grad} M(x) - \text{grad} M(x_n^*) \rangle \\ &= 2\langle x - x_n^*, A(x - x_n^*) \rangle \geq 2\lambda \|x - x_n^*\|^2 \end{aligned}$$

où λ désigne toujours la plus petite valeur propre de $A = E(UU^t)$.

Il nous suffit donc de prendre $L = 2\lambda > 0$ pour satisfaire la première condition du théorème 2.3.2..

Calculons maintenant $E(\|Y_n(x)\|^2)$.

$$\begin{aligned} \|Y_n(x)\|^2 &= \|\text{grad } \varphi(x, U) + \text{grad} g_{n,\alpha}(x)\|^2 \\ &\leq 8 \|\langle x, U \rangle U\|^2 + 2\alpha^2 (\|\text{grad} g_{n,\alpha}(x)\| \leq \alpha) \\ &= 8 \|UU^t x\|^2 + 2\alpha^2 \\ &\leq 16(x - x_n^*)^t UU^t (x - x_n^*) + 16x_n^{*t} UU^t x_n^* + 2\alpha^2 \end{aligned}$$

par suite

$$\begin{aligned} E(\|Y_n(x)\|^2) &\leq 16(x - x_n^*)^t (E(UU^t)) (x - x_n^*) + 16x_n^{*t} (E(UU^t)) x_n^* + 2\alpha^2 \\ &\leq 16\Lambda \|x - x_n^*\|^2 + 16\Lambda \sup_n (\|x_n^*\|^2) + 2\alpha^2 \end{aligned}$$

où Λ désigne la plus grande valeur propre de $E(UU^t)$ et en prenant

$$L_0 = 2\alpha^2 + 16\Lambda \sup_n \|x_n^*\|^2$$

$$L_1 = 0$$

$$L_2 = 16\Lambda$$

nous obtenons

$$E(\|Y_n(x)\|^2) \leq \sum_{i=0}^2 L_i \|x - x_n^*\|^i$$

qui est la deuxième condition du théorème 2.3.2.

Enfin la troisième condition $\sum_n \|x_n^* - x_{n+1}^*\| < +\infty$ est vérifiée si on prend une suite $\{t_n | n \geq 1\}$ telle que $\sum_n \sqrt{t_n} < +\infty$.

Nous concluons de ce qui précède que le processus d'itération stochastique défini comme en 2.3.1., chapitre 2 par

$$X_1 \text{ point arbitraire de } \mathbb{R}^d$$

$$X_{n+1} = X_n + \alpha Y_n(X_n) \quad n = 1, 2, \dots$$

avec $Y_n(x) = -(\text{grad } \varphi(x, U) + \text{grad } g_{n, \alpha}(x)) = -\{2\langle x, U \rangle U + \alpha \frac{(x - \Pi_C(x))}{\max\{\|x - \Pi_C(x)\|, t_n\}}\}$

converge presque sûrement vers le minimum x_* de la fonction de régression $M(x)$ sur C .

3.4. - Mise en oeuvre de l'algorithme

Nous allons maintenant donner des indications sur le choix des paramètres intervenant dans l'algorithme afin d'effectuer sa mise en oeuvre sur ordinateur.

3.4.1. - Choix des paramètres

Afin d'assurer une convergence assez rapide de l'algorithme ainsi défini il faut essayer de déterminer d'une manière aussi satisfaisante que possible ses différents paramètres.

Tout d'abord le choix du point de départ X_1 est arbitraire or, comme la solution cherchée doit ressembler à la lettre avec laquelle il est associé, il semble raisonnable de choisir pour X_1 un point de \mathbb{R}^d représentant cette lettre.

Le choix de α est assez délicat, nous savons qu'il doit être strictement plus grand que $\|\text{grad}M(x_*)\|$ ce qui nécessite une évaluation de cette quantité.

Lorsque $\|x - \Pi_C(x)\| \leq t_n$ c'est-à-dire lorsque le processus est suffisamment près du convexe on doit faire attention à ce que le rapport $\frac{\alpha \alpha_n}{t_n}$ ne soit pas trop grand sinon le processus pourrait osciller fréquemment de part et d'autre du convexe. En fait en début d'itérations il est prudent de surveiller l'évolution du processus afin de s'assurer que les paramètres choisis sont bien déterminés et au besoin procéder à différents essais avec des valeurs différentes des paramètres.

Notons qu'il faut vérifier fréquemment que les masques obtenus sont meilleurs que ceux précédemment calculés. Ainsi ayant obtenu à l'itération n le masque $X_i^{(n)}$ correspondant à la classe i on estime la quantité $E(\langle X_i^{(n)}, U \rangle^2)$ où U désigne une lettre n'appartenant pas à la classe i . Si les masques s'améliorent on doit observer une décroissance d'ensemble de ces différentes estimations au cours des itérations.

Enfin lorsqu'on décide d'arrêter les itérations il y a lieu de procéder à partir des masques retenus à un programme de reconnaissance des lettres de l'alphabet afin de s'assurer que le pourcentage d'erreurs commises lors de cette identification est raisonnable.

3.4.2. Simulation sur ordinateur

Afin de ne pas utiliser un temps considérable de calcul nous nous sommes restreints au cas d'un alphabet \mathcal{A} comportant cinq classes de lettres et par suite, nous avons calculé les cinq masques correspondants.

3.4.2.1. Description et choix des lettres

Pour chaque lettre de l'alphabet \mathcal{A} nous avons constitué un carré quadrillé comportant vingt lignes et vingt colonnes dans lequel, sur sept lignes et sept colonnes, de part et d'autre de la position centrale, s'inscrit la description brute de la lettre. Cette description est donnée par des 1 là où se trouve la lettre et des 0 ailleurs, ce qui correspond à une description en noir et blanc.

Les lettres choisies pour constituer l'alphabet \mathcal{A} sont :
A, D, I, J, L.

Nous supposons que les lettres sont équiprobables et qu'elles ne peuvent subir qu'une translation d'une ligne ou d'une colonne dans les différentes directions autorisées à savoir :

N, N.E, E, S.E, S, S.O, O et N.O

avec les probabilités suivantes :

$\frac{1}{4}$ pour le déplacement nul (position centrale)

$\frac{1}{8}$ pour les directions N, S, E et O,

$\frac{1}{16}$ pour les directions N.E., S.E, S.O et N.O. .

3.4.2.2. Description du bruit.

Nous supposons de plus qu'à la description brute de la lettre s'ajoute (resp. se retranche) dans chaque carré élémentaire du quadrillage lorsqu'il contient 0 (resp. lorsqu'il contient 1) un bruit aléatoire positif dont l'intensité est comprise entre 0 et 1 et la valeur moyenne égale à 0,25. Ce bruit aléatoire est obtenu avec une distribution Bêta de paramètres 2 et 6.

Pratiquement, l'addition du bruit dans les cases "blanches" correspond au fait que le "blanc" n'est pas parfait et comporte des salissures quant à sa soustraction dans les cases comportant la lettre cela veut dire que l'impression de la lettre n'est pas rigoureusement "noire" et comporte des imperfections.

3.4.2.3. Les résultats obtenus.

Les résultats numériques des masques obtenus nous ont permis de montrer que la reconnaissance des lettres imprimées auxquelles s'ajoute une perturbation aléatoire était possible et cela de façon satisfaisante.

Les détails des résultats obtenus sont donnés dans (4).

B I B L I O G R A P H I E

- (1) C. BERGE et A. GHOUILA-HOURI
Programmes, jeux et réseaux de transport. Dunod 1962
- (2) J.R. BLUM
Approximation methods which converge with probability one. Ann. Math. Stat. 25 (1954), 382.386.
- (3) J.R. BLUM
Multidimensional Stochastic approximation method. Ann. Math. Stat. 25 (1954), 737.744.
- (4) P. DAUBEZE
Contribution à l'étude du calcul des masques pour la lecture automatique de caractères imprimés.
Doctorat de Spécialité (1974) - Université Paul Sabatier - Toulouse - France - n° 1662.
- (5) C. DERMAN et J. SACKS
On DVORETZKY's stochastic approximation theorem.
Ann. Math. Stat. 30 (1959), 601.606.
- (6) J.L. DOOB
Stochastic Processes. John Wiley. New-York, (1953).
- (7) A. DVORETZKY
On stochastic approximation. Proceeding of the third Berkeley symposium on Mathematical Statistics and Probability I (1956), 39.55.
- (8) P.L. HENNEQUIN et A. TORTRAT
Théorie des Probabilités et quelques applications.
Masson et Cie (1965).
- (9) J. KIEFER and J. WOLFOWITZ
Stochastic estimation of the maximum of a regression function
Ann. Math. Stat. 23 (1952), 462.466.
- (10) M. LOEVE
Probability theory. van Nostrand. New-York (1960).

- (11) C. MACCHI
Thèse de Doctorat d'Etat es Sciences Mathématiques.
Université de Paris VI.
N° d'enregistrement au C.N.R.S. : AO 7150
- (12) M. METIVIER
Notions fondamentales de la théorie des Probabilités.
Dunod 1968.
- (13) P.A. MEYER
Probabilités et Potentiel. Hermann.
- (14) E. MOURIER
Eléments aléatoires dans un espace de Banach. Ann Inst
H. Poincaré 13 (1952-1953). 161.244.
- (15) J. NEVEU
Bases Mathématiques du calcul des Probabilités.
Masson et Cie.
- (16) J. NEVEU
Martingales à temps discret. Masson et Cie.
- (17) G. FERRENOU
Thèse de Doctorat d'Etat es Sciences Mathématiques.
Université Paul Sabatier - Toulouse - N° d'ordre 332.
- (18) A. RENYI
Calcul des Probabilités avec un appendice sur la théorie
de l'information. Dunod 1966.
- (19) H. ROBBINS and S. MONRO
A stochastic approximation method. Ann. Math. Stat.
22 (1951), 400.407.
- (20) L. SCHMETTERER
Stochastic approximation. Proceedings fourth Berkeley
symposium on Mathematical Statistics and Probability
1 (1960), 587.609.
- (21) L. SCHMETTERER
Sur l'itération stochastique. Le calcul des Probabilités
et ses applications. 87. 55. 63.
- (22) A. TORTRAT
Calcul des Probabilités et introduction aux processus
aléatoires. Masson et Cie.
- (23) J.H. VENTER
On DVORETZKY stochastic approximation theorems.
Ann. Math. Stat. 27 (1966) 1134.1144.

(24) J. WOLFOWITZ

On stochastic approximation methods.
Ann. Math. Stat. 27 (1956) 1151.1156.

(25) J. WOLFOWITZ

On the stochastic approximation method of ROBBINS and
MONRO.
Ann. Math. Stat. 23 (1952). 457.461.