

ANNALES DE L'I. H. P., SECTION B

CÉSAR MACCHI

ODILE MACCHI

Un théorème d'itération stochastique multidimensionnelle

Annales de l'I. H. P., section B, tome 7, n° 3 (1971), p. 193-204

http://www.numdam.org/item?id=AIHPB_1971__7_3_193_0

© Gauthier-Villars, 1971, tous droits réservés.

L'accès aux archives de la revue « Annales de l'I. H. P., section B » (<http://www.elsevier.com/locate/anihpb>) 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/>

Un théorème d'itération stochastique multidimensionnelle

par

César MACCHI et Odile MACCHI

SOMMAIRE. — On propose un théorème d'itération stochastique pour estimer la racine d'une fonction de régression multidimensionnelle. Sous des conditions peu exigeantes, on détermine la solution de l'équation $\vec{M}(\vec{H}) = 0$ comme limite presque sûre d'une suite d'itérations. La fonction $\vec{M}(\vec{H})$ est inconnue. On dispose toutefois de réalisations successives d'un vecteur aléatoire $\vec{Y}(\vec{H})$, de moyenne $\vec{M}(\vec{H})$.

SUMMARY. — A stochastic approximation theorem is presented, to estimate the root of a multidimensional regression function. Under very weak conditions, the solution of the vectorial equation $\vec{M}(\vec{H}) = 0$ is obtained as an almost sure iterative limit. The function $\vec{M}(\vec{H})$ is not known. However, the successive samples of the random vector $\vec{Y}(\vec{H})$, with mean $\vec{M}(\vec{H})$, are available.

INTRODUCTION

L'étude qui suit a été poursuivie en vue d'applications à des problèmes de communications. Tous les vecteurs considérés appartiennent à $\mathbb{R}^L$. Très fréquemment la clef de ces problèmes, donnée par la théorie statistique de la décision, est contenue dans un vecteur $\vec{H}_*$ solution d'une équation

$$\vec{S} = R\vec{H}_*,$$

où $\vec{S}$ est un vecteur connu et où R est la matrice de corrélation d'un vecteur $\vec{B}$

aléatoire centré, stationnaire. Cependant, dans la pratique la matrice R est inconnue, et n'est appréhendée que par une suite d'échantillons $\tilde{B}_j$ du vecteur $\tilde{B}$. Pour estimer le vecteur $\tilde{H}_*$, plusieurs auteurs [1] [2] ont repris la méthode d'itération stochastique de Robbins et Monro [3], préalablement généralisée au cas multidimensionnel par Blum [4]. Considérant l'algorithme

$$(0) \quad \tilde{H}_{j+1} = \tilde{H}_j + \mu_j(-\tilde{B}_j \tilde{B}_j^T \tilde{H}_j + \tilde{S}),$$

ils ont montré que la suite $\tilde{H}_j$ converge en moyenne quadratique [1], ou presque sûrement [4] vers $\tilde{H}_*$, sous un certain nombre de conditions; celles-ci nous ont paru inacceptables dans la pratique, parce qu'elles conduisent toujours à supposer que la norme de $\tilde{H}_j$ est bornée :

$$\forall j, \quad E(\|\tilde{H}_j - \tilde{H}_*\|^2) < M.$$

L'application à l'algorithme (0) du théorème d'itération stochastique que nous présentons ici permet de démontrer sa convergence en s'affranchissant de la condition précédente. Nous le montrerons dans une étude ultérieure.

Plus généralement l'algorithme que nous étudions est le suivant, pour tout entier positif j :

$$(1) \quad \tilde{H}_{j+1} = \tilde{H}_j + \mu_j \tilde{Y}_j(\tilde{H}_j)$$

dans lequel :

- $\tilde{H}_1$ est un vecteur quelconque,
- μ_j est un nombre réel positif,
- la suite $\tilde{Y}_j(\tilde{H})$ est une suite infinie de fonctions aléatoires vectorielles indépendantes, de même loi, définies sur $\mathbb{R}^L$, de moyenne :

$$(2) \quad E\{\tilde{Y}_j\}(\tilde{H}) = \tilde{M}(\tilde{H}).$$

Nous supposons que $\tilde{M}(\tilde{H})$ est une application de $\mathbb{R}^L$ dans lui-même telle qu'il existe un vecteur $\tilde{H}_*$ solution de l'équation

$$(3) \quad \tilde{M}(\tilde{H}) = 0$$

Dans le cas où (3) admet plusieurs solutions, nous verrons plus loin que les conditions du théorème proposé définissent $\tilde{H}_*$ de manière unique parmi elles.

Le problème traité est la recherche de cette solution sans connaître la fonction $\tilde{M}_j(\cdot)$, mais en disposant d'une suite d'épreuves des fonctions aléatoires $\tilde{Y}_j(\cdot)$. Le théorème d'itération stochastique proposé établit que, sous certaines conditions portant sur $\tilde{Y}_j(\cdot)$ et $\tilde{M}(\cdot)$, la suite des vecteurs $\tilde{H}_j$ définie par (1) converge presque sûrement vers $\tilde{H}_*$.

GÉNÉRALITÉS

La suite $\vec{H}_j$ constitue un processus aléatoire (discret) dont l'espace des épreuves, noté Ω , est $(\mathbb{R}^L)^\infty$. Nous appelons F_n , pour n entier positif, (resp. F) la σ -algèbre engendrée dans Ω par la classe des ensembles $B_{j,A}$ pour lesquels la coordonnée $\vec{H}_j$ appartient à A , avec A borélien quelconque de $\mathbb{R}^L$ et $j \leq n$ (resp. j quelconque). Il est clair que les tribus F_n sont emboîtées.

Lorsque le vecteur $\vec{H}_1$ est donné, ainsi que la suite μ_j , la loi de probabilité du processus $\dots, \vec{H}_j, \dots$ est définie sur l'espace probabilisable (Ω, F) par la loi des fonctions aléatoires $\vec{Y}_j(\cdot)$.

Notre but est d'étudier la convergence de la suite $\vec{H}_j$, en en précisant le sens et les conditions. Pour cela, nous introduisons une fonction numérique certaine $f(\vec{H})$ définie et continue sur $\mathbb{R}^L$, possédant des dérivées partielles premières et secondes continues. Soit $\vec{D}(\vec{H})$ son gradient et $A(\vec{H})$ la matrice (L, L) de ses dérivées secondes.

Considérons $f(\vec{H}_{j+1})$; conformément à l'algorithme (1), la fonction numérique $f(\vec{H}_j + \mu_j \vec{Y}_j(\vec{H}_j))$ de la variable réelle μ_j admet le développement de Taylor

$$(4) \quad f(\vec{H}_{j+1}) = f(\vec{H}_j) + \mu_j \langle \vec{D}(\vec{H}_j), \vec{Y}_j(\vec{H}_j) \rangle + \frac{1}{2} \mu_j^2 \langle \vec{Y}_j(\vec{H}_j), A(\vec{H}_j + \mu_j \theta \vec{Y}_j(\vec{H}_j)) \vec{Y}_j(\vec{H}_j) \rangle$$

où θ est un nombre de $]0, 1]$, qui dépend de $\vec{Y}_j(\vec{H}_j)$, de $\vec{H}_j$ et de μ_j . Prenons l'espérance de ce développement, conditionnellement à la tribu F_j . Elle ne porte donc que sur le tirage de $\vec{Y}_j$. Or, cette fonction aléatoire est indépendante de $\vec{Y}_1, \dots, \vec{Y}_{j-1}$, donc de $\vec{H}_1, \dots, \vec{H}_j$, et par conséquent, indépendante de la tribu F_j des événements antérieurs à $(j+1)$. Ainsi, il vient par exemple

$$(5) \quad E \{ \langle \vec{D}(\vec{H}_j), \vec{Y}_j(\vec{H}_j) \rangle / F_j \} = \langle \vec{D}(\vec{H}_j), E \{ \vec{Y}_j \}(\vec{H}_j) \rangle,$$

soit d'après (2)

$$(6) \quad E \{ \langle \vec{D}(\vec{H}_j), \vec{Y}_j(\vec{H}_j) \rangle / F_j \} = \langle \vec{D}(\vec{H}_j), \vec{M}(\vec{H}_j) \rangle.$$

Nous introduisons, pour faciliter l'écriture, les fonctions de $\vec{H}$ ainsi définies :

$$(7) \quad U(\vec{H}) = \langle \vec{D}(\vec{H}), \vec{M}(\vec{H}) \rangle,$$

$$(8) \quad V_a(\vec{H}) = E \{ \langle \vec{Y}_j(\vec{H}), A(\vec{H} + \theta a \vec{Y}_j(\vec{H})) \vec{Y}_j(\vec{H}) \rangle \},$$

$$(9) \quad \Phi_a(\vec{H}) = U(\vec{H}) + \frac{1}{2} a V_a(\vec{H}),$$

qui sont des moyennes calculées avec la mesure de probabilité de la fonction aléatoire $\vec{Y}_j(\cdot)$.

Affectons simplement d'un indice j les valeurs prises par ces fonctions à la $j^{\text{ième}}$ itération, c'est-à-dire pour $\vec{H} = \vec{H}_j$ et $a = \mu_j$. Posons encore

$$(10) \quad \begin{cases} Z_* = f(\vec{H}_*), \\ Z_j = f(\vec{H}_j). \end{cases}$$

Avec ces notations, l'espérance conditionnelle de (4) s'écrit :

$$(11) \quad E(Z_{j+1} - Z_j / F_j) = \left[\mu_j U_j + \frac{1}{2} \mu_j^2 V_j \right] = \mu_j \Phi_j.$$

INTRODUCTION DES CONDITIONS DE CONVERGENCE

Avant d'aborder la démonstration de la convergence de la suite $\vec{H}_j$, nous allons introduire, en essayant de les légitimer, les conditions de convergence que nous utiliserons afin de mettre en évidence leur signification intuitive, que la formulation mathématique ne fournit pas.

La première idée directrice est de lier la convergence de la suite des vecteurs $\vec{H}_j$ à une convergence plus simple, celle d'une suite de nombres réels Z_j , par exemple :

$$(12) \quad Z_j = f(\vec{H}_j) = \|\vec{H}_j - \vec{H}_*\|^2.$$

C'est pour cela qu'a été introduite plus haut la fonction $f(\vec{H})$. Pour lier les deux convergences, on impose que $f(\vec{H})$ ne puisse tendre vers Z_* (limite présumée de Z_j) sans que $\vec{H}$ tende vers $\vec{H}_*$. En notant A_ε (resp. B_ε) l'ensemble des vecteurs $\vec{H}$ tels que $\|\vec{H} - \vec{H}_*\| \geq \varepsilon$ (resp. $\|\vec{H} - \vec{H}_*\| < \varepsilon$), nous introduisons donc la condition

$$(13) \quad \forall \varepsilon > 0, \quad \inf_{\vec{H} \in A_\varepsilon} |f(\vec{H}) - Z_*| > 0.$$

En conséquence, $f(\vec{H})$ ne prend qu'une fois (en $\vec{H}_*$), la valeur Z_* . On montre sans peine, en utilisant la continuité de $f(\vec{H})$, que, pour $L > 1$, ceci entraîne

a) soit que :

$$(14) \quad \vec{H} \neq \vec{H}_* \quad \text{implique} \quad f(\vec{H}) > f(\vec{H}_*);$$

$f(\vec{H})$ atteint alors son minimum absolu en un unique point $\vec{H}_*$;

b) soit le contraire, à savoir que Z_* est le maximum absolu de $f(\vec{H})$ atteint seulement en $\vec{H}_*$. Dans la suite nous supposons que $f(\vec{H})$ admet un minimum Z_* en $\vec{H}_*$. La transposition au cas b) des conditions que nous allons maintenant introduire et de la démonstration du théorème est immédiate.

Il est donc clair, puisque (13) entraîne — par exemple — (14), que la condition (13) peut aussi s'énoncer

$$(C-1) \quad \forall \varepsilon > 0 \quad \inf_{H \in A_\varepsilon} f(\vec{H}) - Z_* > 0.$$

Sur la figure 1, nous avons représenté, dans le cas $L = 2$, une fonction $f(\vec{H})$ qui ne convient pas parce qu'elle atteint son minimum dans tout un domaine.

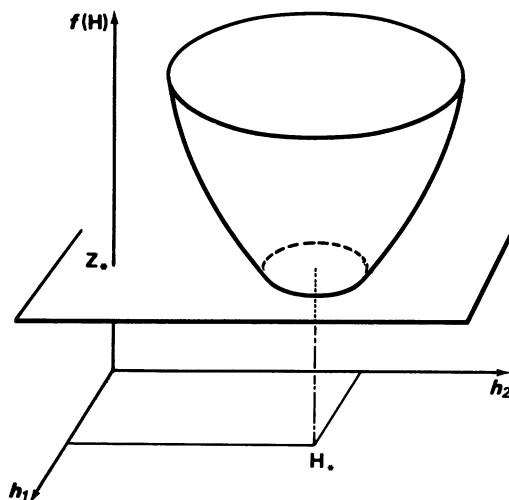


FIG. 1.

Pour assurer la convergence de la suite Z_j , la deuxième idée directrice est de choisir la fonction $f(\vec{H})$ de telle façon que la suite $Z_j = f(\vec{H}_j)$ qui est bornée inférieurement par Z_* soit décroissante.

Naturellement, cela n'est possible qu'en moyenne (en un certain sens) et (11) nous incite à choisir $f(\vec{H})$ vérifiant

$$(15) \quad E(Z_{j+1}/F_j) - Z_j < 0.$$

La suite Z_j serait alors presque sûrement convergente, comme surmar-

tingale bornée inférieurement. D'après (11), nous sommes donc conduits à imposer :

$$(16) \quad \forall a > 0, \quad \forall \vec{H} \neq \vec{H}_* \quad \Phi_a(\vec{H}) < 0.$$

Cependant, nous verrons au cours de la démonstration (lemme 2, (35)) que ceci est insuffisant, et qu'il faut introduire une uniformité dans le caractère négatif de $\Phi_a(\vec{H})$.

On pourrait alors penser à

$$(17) \quad \left\{ \begin{array}{l} \forall a > 0, \quad \forall \varepsilon > 0 \\ \sup_{\vec{H} \in A_\varepsilon} \Phi_a(\vec{H}) < 0. \end{array} \right.$$

L'ennui est que dans beaucoup des applications, il est très difficile d'exhiber des fonctions $f(\vec{H})$ satisfaisant à (17). Essayant de trouver une condition moins stricte, Blum [4] a proposé une condition qui ressemble à (17) pour les a d'un intervalle borné :

$$(18) \quad \left\{ \begin{array}{l} \exists \lambda > 0 \text{ tel que } \forall 0 < a \leq 2\lambda \text{ et } \forall \varepsilon > 0 \\ \sup_{\vec{H} \in A_\varepsilon} \Phi_a(\vec{H}) < 0. \end{array} \right.$$

Cette condition nous est apparue encore trop exigeante, parce que c'est précisément elle qui n'est pas satisfaite dans les applications fondamentales de l'algorithme (1) à la théorie des communications, tel que le filtrage numérique auto-adapté à la détection d'un signal certain dans un bruit de corrélation inconnue.

Nous proposons de la transformer ainsi :

$$(C-2) \quad \left\{ \begin{array}{l} \forall \varepsilon > 0, \quad \exists \lambda(\varepsilon) > 0 \text{ et } \eta(\varepsilon) > 0 \text{ tels que } \forall 0 < a \leq 2\lambda(\varepsilon) \text{ et } \forall \vec{H} \in A_\varepsilon \\ \Phi_a(\vec{H}) \leq -\eta(\varepsilon). \end{array} \right.$$

Aucune des deux conditions (18) et (C.2) n'implique l'autre. Remarquons simplement qu'à la différence de (18) la condition (C-2) permet que $\lambda(\varepsilon)$ tende vers zéro avec ε , et ainsi que pour tout a donné non nul, $\sup \Phi_a(\vec{H}) < 0$ à l'extérieur d'une boule de diamètre $\varepsilon(a)$, mais non à l'extérieur d'une boule plus petite.

Les conditions (17) (18) et (C-2) visent toutes à ce que lorsque $\vec{H}_{j+1}$ « n'est pas très proche de $\vec{H}_*$ » Z_{j+1} soit plus petit que Z_j (en moyenne). Il faut donc rajouter une condition qui empêche que Z_{j+1} soit beaucoup plus grand que Z_j lorsque $\vec{H}_j$ est « proche de $\vec{H}_*$ ». Pour cela, nous supposons que, dans un voisinage de $\vec{H}_*$, $E\{Z_{j+1} - Z_j/F_j\}$ est majoré par une série convergente ; une méthode (qui s'avèrera efficace), est de supposer $U(\vec{H}) \leq 0$ pour tout $\vec{H}$, les fonctions $V_{\mu_j}(\vec{H})$ uniformément bornées dans ce voisinage, et la série des μ_j^2 convergente.

Ceci nous amène à poser les deux conditions

$$(C-3) \quad \begin{cases} U(\vec{H}) \leq 0 & \forall \vec{H}, \\ \exists V \text{ et } \exists \varepsilon_1 > 0 \text{ tels que } \forall \vec{H} \in B_{\varepsilon_1} \text{ et } \forall 0 < a \leq 2\lambda(\varepsilon_1) \\ V_a(\vec{H}) \leq V; \end{cases}$$

$$(C-4)' \quad \sum_{j=1}^{\infty} \mu_j^2 < \infty.$$

THÉORÈME

La démonstration se décompose comme suit :

- Z_j admet une limite presque sûre,
- cette limite ne peut être que Z_* ,
- $\vec{H}_j$ converge presque sûrement vers $\vec{H}_*$.

Dans le lemme qui suit, nous étudions le premier point. C'est pourquoi la condition (C-1), introduite pour assurer la convergence des vecteurs $\vec{H}_j$ n'y est pas nécessaire, mais seulement une version affaiblie permettant de minorer les Z_j

$$(C-1)' \quad \forall \vec{H}_1, \exists Z_0 \text{ réel tel que } \forall j \text{ entier positif} \\ Z_j \geq Z_0.$$

Énonçons alors :

LEMME 1. — La suite $\{Z_j\}$ de variables aléatoires définies par (1) et (10) converge presque sûrement vers une variable aléatoire intégrable pourvu que soient remplies les conditions (C-1)', (C-2), (C-3) et (C-4)'.

Démonstration du lemme 1. — Elle consiste à prouver que Z_j est « presque » une surmartingale, plus précisément qu'il existe un entier positif J_1 tel que :

$$(19) \quad E \{ Z_{j+1} - Z_j / F_j \} \leq \frac{1}{2} \mu_j^2 V, \text{ pour } j > J_1.$$

En définissant alors les variables aléatoires ξ_j par

$$(20) \quad \xi_j = Z_j - \sum_{k=1}^{j-1} \frac{1}{2} \mu_k^2 V,$$

l'inégalité (19) montre que la suite ξ_j ($j > J_1$) constitue une surmartingale, qui, grâce à (C-1)' et (C-4)' est bornée inférieurement par

$$Z_0 - \sum_{k=1}^{\infty} \frac{1}{2} \mu_k^2 V.$$

D'après un théorème dû à Doob [5], la suite ξ_j converge presque sûrement vers une variable aléatoire intégrable ; d'après (20) il en est de même pour la suite Z_j .

Nous réécrivons l'équation tout à fait fondamentale

$$(11) \quad E(Z_{j+1} - Z_j / F_j) = \mu_j \left[U_j + \frac{1}{2} \mu_j V_j \right] = \mu_j \Phi_j.$$

Pour démontrer (19), nous considérons deux cas, suivant que dans l'espérance (11), qui est conditionnelle à $\vec{H}_1, \dots, \vec{H}_j$, le vecteur $\vec{H}_j$ appartient ou non à l'ensemble A_{ε_1} .

Étudions le cas où $\vec{H}_j \in A_{\varepsilon_1}$. Puisque d'après (C-4)' μ_j tend vers zéro, il existe un entier $J_1 > 0$ tel que, pour $j > J_1$, on ait $\mu_j < 2\lambda(\varepsilon_1)$; d'après (C-2) ceci entraîne

$$\sup_{\vec{H} \in A_{\varepsilon_1}} \Phi_{\mu_j}(\vec{H}) < 0,$$

et par suite, $\Phi_j = \Phi_{\mu_j}(\vec{H}_j) < 0$ qui conduit, d'après (11) à

$$(19)' \quad E(Z_{j+1} - Z_j / F_j) < 0 \quad \text{pour } j > J_1.$$

Regardons maintenant le cas où $\vec{H}_j \in B_{\varepsilon_1}$. Alors les majorations (C-3) valables dans la boule B_{ε_1} prouvent que pour $j > J_1$

$$(21) \quad U_j = U(\vec{H}_j) \leq 0,$$

$$(22) \quad V_j = V_{\mu_j}(\vec{H}_j) \leq V,$$

et donc, toujours d'après (11)

$$(19)'' \quad E(Z_{j+1} - Z_j / F_j) \leq \frac{1}{2} \mu_j^2 V.$$

Avec (19)' ceci achève la démonstration de (19), et par là, du lemme 1.

LEMME 2. — Supposons que la série μ_j soit divergente. Alors, dans les conditions du lemme 1, la suite de vecteurs $\vec{H}_j$ définie par (1) vérifie la propriété

$$(23) \quad \forall \varepsilon > 0, \quad \liminf P \{ \vec{H}_j \in A_\varepsilon \} = 0.$$

COROLLAIRE. — Dans les conditions du lemme 2, la suite Z_j définie par (1) et (10) converge presque sûrement vers Z_* .

Appelons X la limite presque sûre de Z_j (lemme 1). Pour démontrer le corollaire, nous prouvons que $X = Z_*$ presque sûrement. A l'aide du lemme 2 (23), et de la continuité de $f(\vec{H})$, on voit sans peine que

$$(24) \quad \forall \varepsilon, \quad \liminf P \{ |Z_j - Z_*| \geq \varepsilon \} = 0.$$

Ainsi, pour tout ε , il existe une sous-suite j_k^ε telle que :

$$(25) \quad P \{ |Z_{j_k^\varepsilon} - Z_*| < \varepsilon \} \rightarrow 1 \quad \text{lorsque } k \rightarrow \infty.$$

Or, puisque $Z_j \rightarrow X$ presque sûrement,

$$(26) \quad P \{ |Z_{j_k^\varepsilon} - X| < \varepsilon \} \rightarrow 1 \quad \text{lorsque } k \rightarrow \infty.$$

En outre

$$(27) \quad P \{ |X - Z_*| < \varepsilon \} \geq P \left\{ \left(|Z_{j_k^\varepsilon} - Z_*| < \frac{\varepsilon}{2} \right) \times \left(|Z_{j_k^\varepsilon} - X| < \frac{\varepsilon}{2} \right) \right\}.$$

En reportant dans (27) l'inégalité ensembliste

$$P \{ D \times E \} \geq P \{ D \} + P \{ E \} - 1$$

il vient :

$$(28) \quad P \{ |X - Z_*| < \varepsilon \} \geq P \left\{ |Z_{j_k^\varepsilon} - Z_*| < \frac{\varepsilon}{2} \right\} + P \left\{ |Z_{j_k^\varepsilon} - X| < \frac{\varepsilon}{2} \right\} - 1$$

D'après (25) et (26) ceci prouve que

$$(29) \quad \forall \varepsilon > 0, \quad P \{ |X - Z_*| < \varepsilon \} = 1,$$

soit $P \{ X = Z_* \} = 1$. Ceci démontre le corollaire.

Démonstration du lemme 2. — Pour prouver (23) introduisons pour tout $\varepsilon > 0$, la variable aléatoire auxiliaire W_j , définie par :

$$(30) \quad W_j = \begin{cases} U_j \dots & \text{si } \vec{H}_j \in B_\varepsilon, \\ U_j + \frac{1}{2} \mu_j V_j \dots & \text{si } \vec{H}_j \in A_\varepsilon. \end{cases}$$

Il est clair d'après (C-2) et (C-3), qu'il existe un entier $J(\varepsilon)$ tel que :

$$(31) \quad W_j \leq 0 \quad \text{si } j > J(\varepsilon).$$

A partir d'un certain rang, le terme général de la série $\sum \mu_j E \{ W_j \}$ est donc négatif ou nul.

Nous montrons plus loin que :

$$(32) \quad \sum_{j=1}^{\infty} \mu_j E \{ W_j \} > -\infty \quad \text{si} \quad 0 < \varepsilon \leq \varepsilon_1.$$

Puisque la série des μ_j diverge, il apparaît alors que :

$$(33) \quad \liminf E \{ -W_j \} = 0 \quad \text{si} \quad 0 < \varepsilon \leq \varepsilon_1.$$

Cette égalité s'écrit, à l'aide de la loi de probabilité $P_f(\vec{H})$ du vecteur $\vec{H}_j$

$$\liminf \left\{ \int_{B_\varepsilon} -U(\vec{H}) dP_f(\vec{H}) + \int_{A_\varepsilon} - \left(U(\vec{H}) + \frac{1}{2} \mu_j V_{\mu_j}(\vec{H}) \right) dP_f(\vec{H}) \right\} = 0$$

pour $0 < \varepsilon \leq \varepsilon_1$; par suite, dans les mêmes conditions

$$(34) \quad \liminf \left\{ \int_{B_\varepsilon} -U(\vec{H}) dP_f(\vec{H}) \right\} + \liminf \left\{ \int_{A_\varepsilon} - \left(U(\vec{H}) + \frac{1}{2} \mu_j V_{\mu_j}(\vec{H}) \right) dP_f(\vec{H}) \right\} \leq 0.$$

Les inégalités (31) et (34) ne sont compatibles que si, en particulier :

$$(35) \quad \liminf \int_{A_\varepsilon} - \left(U(\vec{H}) + \frac{1}{2} \mu_j V_{\mu_j}(\vec{H}) \right) dP_f(\vec{H}) = 0 \quad \text{si} \quad 0 < \varepsilon \leq \varepsilon_1.$$

Il suffit de remarquer qu'il existe $\eta > 0$ tel que si $j > J(\varepsilon)$,

$$- \left(U(\vec{H}) + \frac{1}{2} \mu_j V_{\mu_j}(\vec{H}) \right) > \eta \quad \text{pour} \quad \vec{H} \in A_\varepsilon,$$

(condition C-2), pour obtenir l'équation (23), dans le cas $0 < \varepsilon \leq \varepsilon_1$; et donc dans tous les cas.

Pour achever la démonstration du lemme 2, il nous reste à prouver (32).

Il découle immédiatement de la définition (30) que

$$(36) \quad E \{ W_j \} = E \left\{ U_j + \frac{1}{2} \mu_j V_j \right\} - \frac{1}{2} \mu_j \int_{B_\varepsilon} V_{\mu_j}(\vec{H}) dP_f(\vec{H}).$$

Or, en considérant à nouveau l'égalité (11) et en prenant l'espérance mathématique, il vient après itérations

$$(37) \quad E \{ Z_{n+1} - Z_1 \} = \sum_{j=1}^n \mu_j E \left\{ U_j + \frac{1}{2} \mu_j V_j \right\}.$$

Si nous rapprochons (36) et (37), nous trouvons :

$$(38) \quad \sum_{j=1}^n \mu_j E \{ W_j \} = E \{ Z_{n+1} - Z_1 \} - \sum_{j=1}^n \frac{1}{2} \mu_j^2 \int_{B_\varepsilon} V_{\mu_j}(\vec{H}) dP_j(\vec{H}).$$

Supposons maintenant $\varepsilon \leq \varepsilon_1$: dans ces conditions il vient d'après (C-3) et (C-4)'

$$(39) \quad \sum_{j=1}^n \frac{1}{2} \mu_j^2 \int_{B_\varepsilon} V_{\mu_j}(\vec{H}) dP_j(\vec{H}) \leq V \sum_{j=1}^n \frac{1}{2} \mu_j^2 < +\infty.$$

En reportant (39) dans (38), la condition (C-1)' entraîne le résultat (32).

Le résultat du corollaire précédent entraîne la convergence presque sûre du vecteur $\vec{H}_j$ vers $\vec{H}_*$ pourvu que l'on remplace la condition (C-1)' par la condition plus forte (C-1). Celle-ci implique en effet que $f(\vec{H})$ ne peut tendre vers Z_* que si $\vec{H}$ tend vers $\vec{H}_*$.

Nous avons ainsi démontré le théorème suivant :

THÉORÈME. — S'il existe une fonction $f(\vec{H})$ à valeurs réelles, définie sur $\mathbb{R}^L$, deux fois continûment différentiable, satisfaisant (C-1), (C-2), (C-3), alors pour toute suite de nombres réels positifs μ_j tels que :

$$(C-4) \quad \sum_{j=1}^{\infty} \mu_j = \infty, \quad \sum_{j=1}^{\infty} \mu_j^2 < \infty,$$

la suite de vecteurs $\vec{H}_j$ définie, à partir d'un vecteur $\vec{H}_1$ quelconque, par l'algorithme

$$\vec{H}_{j+1} = \vec{H}_j + \mu_j \vec{Y}_j(\vec{H}_j),$$

converge presque sûrement vers le vecteur $\vec{H}_*$.

Remarque 1. — C'est au cours de la démonstration de ce lemme 2, pour le passage de (32) à (33), qu'apparaît la nécessité de l'hypothèse, un peu étonnante que la série des μ_j diverge. Ceci correspond à l'idée intuitive suivante : pour obtenir une bonne convergence de l'algorithme (1), il faut que le terme complémentaire $\mu_j \vec{Y}_j(\vec{H}_j)$ ne soit pas trop petit, et donc que μ_j soit assez grand.

Remarque 2. — C'est encore dans ce lemme, pour obtenir

$$(23) \quad \forall \varepsilon, \quad \liminf P \{ \vec{H}_j \in A_\varepsilon \} = 0$$

à partir de (35), qu'apparaît la nécessité de la condition (C-2) non complè-

tement exploitée dans le lemme 1 ; on ne peut pas affaiblir (C-2) sans changer toute la démonstration, donc aussi les autres conditions.

Remarque 3. — On peut se poser la question de l'unicité de la solution de (3). Dans le cas où il existe une fonction $f(\vec{H})$ satisfaisant aux conditions (C-1)-(C-3), et telle que $V_a(\vec{H}) \geq 0$ pour tout $\vec{H}$, la solution est unique. En particulier il en est ainsi lorsqu'on peut choisir

$$(40) \quad f(\vec{H}) = \|\vec{H} - \vec{H}_*\|^2.$$

De toutes façons, on voit facilement que s'il existe une autre solution $\vec{H}_0$ à l'équation (3) celle-ci présente une irrégularité pour la statistique de $\vec{Y}(\vec{H})$, en ce sens que

$$(41) \quad E \{ \|\vec{Y}(\vec{H}_0)\|^2 \} = +\infty,$$

alors que la statistique de $\vec{Y}(\vec{H})$ est régulière en $\vec{H}_*$

$$(42) \quad E \{ \|\vec{Y}(\vec{H}_*)\|^2 \} < \infty.$$

BIBLIOGRAPHIE

- [1] SAKRISON, Stochastic approximation, a recursive method for solving regression problems, in *Advances in Communication Systems*. A. V. Balakrishnan Ed., N. Y., Academic Press, 1966, p. 51.
- [2] J. CHANG and F. TUTEUR, Adaptive tapped delay line filters. *Proc. Second Annual conference on information sciences and systems*, p. 164-168, March 1968.
- [3] H. ROBBINS and S. MONRO, A stochastic approximation method. *The Annals of Mathematical Statistics*, vol. 22, 1951, p. 400.
- [4] J. R. BLUM, Multidimensional stochastic approximation methods. *The Annals of Mathematical Statistics*, vol. 25, 1954, p. 737.
- [5] P. A. MEYER, *Probabilités et Potentiel*. Publications de l'Institut de Mathématiques de l'Université de Strasbourg. Hermann, 1966.

Manuscrit reçu le 5 mars 1971.