Revue française d'informatique et de recherche opérationnelle

P. J. Laurent

Approximation uniforme de fonctions continues sur un compact avec contraintes de type inégalité

Revue française d'informatique et de recherche opérationnelle, tome 1, n° 5 (1967), p. 81-95
<http://www.numdam.org/item?id=M2AN_1967__1_5_81_0>

R.I.R.O.
(1re année, N° 5, 1967, p. 81-95)

APPROXIMATION UNIFORME DE FONCTIONS CONTINUES SUR UN COMPACT AVEC CONTRAINTES DE TYPE INEGALITE (1)

par P. J. LAURENT (2)

Résumé. — *On considère l'approximation au sens de Tchébycheff d'une fonction continue sur un compact par un élément d'une variété linéaire de dimension finie soumis à des contraintes continues de type inégalité. Les théorèmes d'existence, de caractérisation et d'unicité du meilleur approximant sont établis. Une généralisation de l'algorithme de Rémès est proposée et sa convergence est démontrée sous certaines hypothèses.*

Problème. — Soit K et H compacts (pas nécessairement disjoints) d'un espace métrique, $S = K \cup H$ et $C(S)$ l'ensemble des fonctions continues à valeurs réelles sur S. Soit V un sous-espace de dimension n de $C(S)$ engendré par $f_1, f_2, \ldots, f_n$ linéairement indépendants sur K. Pour $f \in C(S)$ donné, on appelle $\overline{V}$ l'ensemble des $g \in V$ qui vérifient $g(P) \leqslant f(P)$ pour tout $P \in H$.

Posons $\|h\| = \max\limits_{P \in K} |h(P)|$.

On cherche $g^* \in \overline{V}$ tel que :

$$\|f - g^*\| = \operatorname*{Inf}_{g \in \overline{V}} \|f - g\| = \rho$$

A. EXISTENCE, CARACTERISATION ET UNICITE DE LA MEILLEURE APPROXIMATION

Théorème 1 : (existence).

Si $\overline{V}$ n'est pas vide pour un certain $f \in C(S)$, alors il existe toujours au moins un meilleur approximant g^* de f dans $\overline{V}$.

---

(1) Cet article est le texte détaillé d'une communication présentée au 5e congrès de l'A.F.I.R.O. à Lille (juin 1966).

(2) Maître de conférences au service de Mathématiques Appliquées de la Faculté des Sciences de Grenoble.

En effet, $\overline{V}$ est un sous-ensemble convexe fermé de V.

Si $\overline{g} \in \overline{V}$ n'est pas meilleure approximation de f, $(\eta = \|f - \overline{g}\| > \rho)$, comme $\|g - \overline{g}\| > 2\eta$ entraîne $\|f - g\| \geqslant |\|g - \overline{g}\| - \|\overline{g} - f\|| > \eta$, on a :

$$\rho = \operatorname*{Inf}_{g \in B} \|f - g\| \qquad \text{avec} \qquad B = BF(\overline{g}, 2\eta) \cap \overline{V}.$$

B étant un compact, la borne inférieure est atteinte.

NOTATIONS : Étant donné $g \in \overline{V}$ et $e = f - g$ l'erreur correspondante, on notera :

$$\begin{aligned}
E^+ &= < P \in K : e(P) = \|e\| > \\
E^- &= < P \in K : e(P) = -\|e\| > \\
N &= < P \in H : e(P) = 0 > \\
E &= E^+ \cup E^- \text{ (ensemble des points extrémaux),} \\
C &= E \cup N \text{ (ensemble des points critiques).}
\end{aligned}$$

Théorème 2 :

Si pour un élément $g \in \overline{V}$ (auquel correspondent les ensembles de points critiques E^+, E^-, et N) il existe $h \in V$ vérifiant :

a) $h(P) > 0$ pour tout $P \in E^+$

b) $h(P) < 0$ pour tout $P \in E^- \cup N$

alors g n'est pas meilleure approximation de f dans $\overline{V}$.

On va démontrer que pour $\delta > 0$ suffisamment petit, $g = g + \delta \cdot h$ appartient à $\overline{V}$ et est meilleur que g.

Notons $\tilde{e} = e - \delta \cdot h$ l'erreur correspondante.

E et N sont des compacts disjoints.

Posons

$$a_1 = \min_{P \in E} |h(P)| \quad ; \quad \text{on a } a_1 > 0.$$

$$a_2 = \max_{P \in N} h(P) \quad ; \quad \text{on a } a_2 < 0.$$

En utilisant la continuité uniforme de h et e on peut trouver $\sigma > 0$ tel que :

1° $|h(Q)| > \dfrac{a_1}{2}$ et $|e(Q)| > q \cdot R$ (avec $q < 1$ et $R = \|e\|$)

pour tout $Q \in BO(P, \sigma)$ et tout $P \in E$.

2° $h(Q) < 0$ pour tout $Q \in BO(P, \sigma)$ et tout $P \in N$.

Notons :

$$BE = \bigcup_{P \in E} BO(P, \sigma) \qquad BN = \bigcup_{P \in N} BO(P, \sigma)$$

$$K^* = K - BE \qquad H^* = H - BN$$

K^* et H^* sont des compacts.

Pour déterminer le $\delta > 0$ convenable, nous allons nous placer successivement sur BE, BN, K^* et H^* qui ne sont d'ailleurs pas en général disjoints :

α) $Q \in BE$

$$\tilde{e}(Q) = e(Q) \cdot \left(1 - \delta \cdot \frac{h(Q)}{e(Q)}\right)$$

Pour $\delta < \dfrac{q \cdot R}{M}$ (avec $M = \max\limits_{P \in S} |h(P)|$) on aura :

$$|\tilde{e}(Q)| \leqslant R - \frac{\delta \cdot a_1}{2} < R$$

β) $Q \in BN$

$$\tilde{e}(Q) = e(Q) - \delta \cdot h(Q) \geqslant 0 \quad \text{car} \quad e(Q) \geqslant 0 \quad \text{et} \quad -\delta \cdot h(Q) > 0$$

γ) $Q \in K^*$

Posons : $\operatorname*{Max}\limits_{Q \in K^*} |e(Q)| = R^* < R.$

$$|\tilde{e}(Q)| \leqslant R^* + \delta \cdot M < R \quad \text{si} \quad \delta < \frac{R - R^*}{M}.$$

δ) $Q \in H^*$

Posons : $\operatorname*{Min}\limits_{Q \in H^*} e(Q) = \eta > 0$

$$\tilde{e}(Q) \geqslant \eta - \delta \cdot M > 0 \quad \text{si} \quad \delta < \frac{\eta}{M}.$$

En résumé, si $\delta < \dfrac{1}{M} \min(qR, R - R^*, \eta)$ alors $\tilde{g}$ appartient à $\overline{V}$ et est meilleur que g.

Théorème 3 (caractérisation) :

On fait l'hypothèse suivante :

$H1$: Il existe $g \in V$ vérifiant $g(P) < f(P)$ pour tout $P \in H$ (inégalité stricte). Une condition nécessaire et suffisante pour que $g^* \in \overline{V}$ soit meilleure approximation de f dans $\overline{V}$ est que l'on puisse trouver, pour

$e^* = f - g^*$, k points critiques $(k \leqslant n+1) P_1, P_2, \ldots, P_k$ (dont l'un au moins est extrémal) et une fonctionnelle L basée sur ces points

$$L(h) = \sum_{i=1}^{k} \lambda_i \cdot h(P_i)$$

s'annulant sur V tels que :

$$\lambda_i > 0 \quad \text{si} \quad P_i \in E^+$$

$$\lambda_i < 0 \quad \text{si} \quad P_i \in E^- \cup N.$$

Condition suffisante

Supposons que g^* vérifie les conditions du théorème sans être meilleure approximation. Appelons $g^{**} \in \overline{V}$ une meilleure approximation :

$$\rho = \|f - g^{**}\| < \|f - g^*\|$$

On aurait dans ces conditions, en posant :

$$d = g^{**} - g^* = (f - g^*) - (f - g^{**}) \quad ; \quad d \in V :$$

$$d(P_i) > 0 \quad \text{si} \quad P_i \in E^+ \quad (\lambda_i > 0)$$

$$d(P_i) < 0 \quad \text{si} \quad P_i \in E^- \quad (\lambda_i < 0)$$

$$d(P_i) \leqslant 0 \quad \text{si} \quad P_i \in N \quad (\lambda_i < 0)$$

Comme l'un au moins des P_i est extrémal (appartient à $E^+ \cup E^-$) on aurait $L(d) = \sum_{i=1}^{k} \lambda_i \cdot d(P_i) > 0$ d'où la contradiction.

Condition nécessaire

Soit $g^* \in \overline{V}$ une meilleure approximation de f et E^+, E^- et N les ensembles de points critiques correspondants.

Appelons Γ l'ensemble des points de R^n dont les composantes sont :

$$f_i(P) \quad (i = 1 \text{ à } n) \quad \text{pour} \quad P \in E^+$$

$$-f_i(P) \quad (i = 1 \text{ à } n) \quad \text{pour} \quad P \in E^- \cup N$$

Montrons que l'origine θ de R^n appartient à l'enveloppe convexe $co(\Gamma)$ de Γ : en effet, si l'on avait $\theta \notin co(\Gamma)$, comme Γ est compact, $co(\Gamma)$ l'est aussi (Eggleston [2], p. 22) et il existerait un hyperplan strictement séparateur (Eggleston, p. 20) : c'est-à-dire des constantes $c_j (j = 1 \text{ à } n)$ et $\tau > 0$ telles que :

$$\sum_{j=1}^{n} c_j z_j \geqslant \tau \qquad \text{pour tout } z \in co(\Gamma)$$

Donc en particulier pour les points de Γ :

$$\begin{cases} \sum_{j=1}^{n} c_j f_j(P) \geqslant \tau > 0 & \text{pour tout } P \in E^+ \\ -\sum_{j=1}^{n} c_j f_j(P) \geqslant \tau > 0 & \text{pour tout } P \in E^- \cup N \end{cases}$$

En posant $h = \sum_{j=1}^{n} c_j f_j \in V$, d'après le théorème 2, g^* ne pourrait être meilleure approximation, d'où la contradiction. Donc $\theta \in co(\Gamma)$.

D'après le théorème de Carathéodory sur les ensembles convexes de R^n (Eggleston, p. 35) on peut trouver k points $U_1, U_2, \ldots, U_k$ de Γ (avec $k \leqslant n+1$) tels que θ appartienne encore à l'enveloppe convexe de ces points. Appelons $P_1, P_2, \ldots, P_k$ les points de S qui correspondent à ces U_i.

On a donc :

$$\theta = \sum_{i=1}^{k} \rho_i U_i \qquad \text{avec} \qquad \sum_{i=1}^{k} \rho_i = 1, \rho_i > 0,$$

ce qui donne en projection :

$$0 = \sum_{i=1}^{k} \rho_i \varepsilon_i f_j(P_i) \qquad (j = 1 \text{ à } n)$$

$$\text{avec } \varepsilon_i = \begin{cases} +1 & \text{si} \quad P_i \in E^+ \\ -1 & \text{si} \quad P_i \in E^- \cup N \end{cases}$$

En posant $\lambda_i = \rho_i \varepsilon_i$ on a la fonctionnelle $L(h) = \sum_{i=1}^{k} \lambda_i h(P_i)$ avec les conditions demandées dans le théorème.

Il reste seulement à démontrer que l'un des P_i au moins est extrémal : Supposons le contraire : $P_i \in H - K$, $i = 1$ à k. Pour tout $g \in \overline{V}$ on aurait $g(P_i) = f(P_i)$ car si pour i_0 :

$g(P_{i_0}) < f(P_i) = g^*(P_{i_0})$ alors $L(g^* - g) < 0$ d'où la contradiction. Si $g(P_i) = f(P_i)$ pour tout $g \in \overline{V}$, on est en contradiction avec l'hypothèse $H1$. Donc, l'un des P_i est extrémal.

Remarque. — L'hypothèse $H1$ ne sert pas pour démontrer que la condition est suffisante.

Théorème 4 (unicité)

Une condition nécessaire et suffisante d'unicité du meilleur approximant pour tout f tel que le $\overline{V}$ associé vérifie H_1 est qu'il n'existe pas de fonctionnelle L de la forme $L(h) = \sum_{i=1}^{k} \lambda_i \cdot h(P_i)$, avec $k \leqslant n$, s'annulant sur V, les coefficients λ_i correspondants aux P_i de $H - K$ étant de signe constant et l'un au moins des P_i appartenant à K.

Condition suffisante

Si la condition du théorème 4 est vérifiée, il est évident que la fonctionnelle L du théorème 3 servant à caractériser une meilleure approximation est forcément basée sur exactement $n + 1$ points.

Supposons que g^* et g^{**} soient deux meilleures approximations de f dans $\overline{V}$. Formons $\bar{g} = (g^* + g^{**})/2$ qui est aussi meilleure approximation ; soit $P_1, P_2, \ldots, P_{n+1}$ les points critiques de $\bar{g}$. On montre facilement qu'ils sont aussi points critiques de g^* et g^{**} et de même nature.

On a pour $d = g^* - g^{**} = \sum_{j=1}^{n} \alpha_j f_j$:

$$d(P_i) = 0 \qquad i = 1 \text{ à } n + 1$$

Montrons que cela entraîne $d = 0$, c'est-à-dire $\alpha_j = 0$, $j = 1$ à n : l'un au moins des points critiques, soit P_μ, appartient à K.

Les α_j sont solution du système homogène :

$$\sum_{j=1}^{n} \alpha_j f_j(P_i) = 0 \qquad i = 1 \text{ à } n + 1, \qquad i \neq \mu$$

Pour que l'on ait la solution unique $\alpha_j = 0$, il suffit que $D = \det [f_j(P_i),\ j = 1 \text{ à } n,\ i = 1 \text{ à } n + 1,\ i \neq \mu] \neq 0$.

Appelons λ_i ($i = 1$ à $n + 1$) les coefficients de la fonctionnelle qui a servi à caractériser $\bar{g}$ (théorème 3).

Si l'on avait $D = 0$, le système

$$\sum_{\substack{i=1 \\ i \neq \mu}}^{n+1} \lambda_i \cdot f_j(P_i) = - \lambda_\mu \cdot f_j(P_\mu) \qquad (j = 1 \text{ à } n)$$

n'admettrait pas une solution unique : en particulier, on pourrait trouver (avec le même λ_μ, $P_\mu \in K$) une solution λ'_i de même signe que λ_i, avec $\lambda'_{i_0} = 0$: Ce serait une fonctionnelle basée sur n points, l'un au moins appartenant à K et les λ'_i correspondant à des $P_i \in H - K$ étant de signe constant (comme c'est le cas pour les λ_i, théorème de caractérisation) d'où la contradiction.

Condition nécessaire

Supposons qu'il existe une fonctionnelle L vérifiant les conditions du théorème basée sur n points $P_1, P_2, \dots P_n$; supposons par exemple $\lambda_i < 0$ pour $P_i \in H - K$. Les λ_i forment une solution non identiquement nulle du système homogène :

$$\sum_{i=1}^{n} \lambda_i \cdot f_j(P_i) = 0 \qquad (j = 1 \text{ à } n)$$

donc det $(f_j(P_i)) = 0$. Cela entraîne qu'il existe $g = \sum_{j=1}^{n} \alpha_j \cdot f_j$ avec des α_j non identiquement nuls telle que :

$$g(P_i) = \sum_{j=1}^{n} \alpha_j f_j(P_i) = 0 \qquad (i = 1 \text{ à } n)$$

On peut prendre g telle que : $\underset{P \in S}{\operatorname{Max}} |g(P)| < 1/2$

Choisissons $\tilde{f} \in C(S)$ vérifiant :

$$1)\ \tilde{f}(P_i) = \begin{cases} +1 & \text{si} \quad \lambda_i > 0,\ P_i \in K \\ -1 & \text{si} \quad \lambda_i < 0,\ P_i \in K - H \\ \ \ 0 & \text{si} \quad \lambda_i < 0,\ P_i \in H \end{cases}$$

2) $\|\tilde{f}\| = 1$

3) $\dfrac{|g(P)|}{1 - |g(P)|} \leqslant \tilde{f}(P)$ pour tout $P \in H$.

Il est facile de voir que ces conditions sont compatibles.

Posons

$$f(P) = f(P) \cdot (1 - |g(P)|)$$
$$g^*(P) = \ \varepsilon\ \cdot g(P)$$

et montrons que pour tout $\varepsilon \in [0, 1]$, g^* est meilleure approximation de f dans $\overline{V}$: On a : $\|f - g^*\| \leqslant 1$; en effet, pour tout $P \in K$ et tout $\varepsilon \in [0, 1]$:

$$\begin{aligned} |f(P) - \varepsilon \cdot g(P)| &\leqslant |f(P)| + \varepsilon \cdot |g(P)| \\ &= |\tilde{f}(P)| \cdot (1 - |g(P)|) + \varepsilon \cdot |g(P)| \\ &\leqslant 1 - (1 - \varepsilon) \cdot |g(P)| \leqslant 1 \end{aligned}$$

D'après la condition 3) $g^* = \varepsilon \cdot \tilde{g}$ appartient à $\overline{V}$ pour tout $\varepsilon \in [0, 1]$. Comme $e^*(P_i) = f(P_i) - g^*(P_i) = \tilde{f}(P_i)$, d'après la condition 1), les points P_i sont critiques et leur nature est en accord avec le signe des λ_i : d'après le théorème 3, g^* est meilleure approximation de f dans $\overline{V}$ quel que soit $\varepsilon \in [0, 1]$; il n'y a donc pas unicité.

Pour l'obtention numérique de la meilleure approximation, nous ferons deux hypothèses supplémentaires qui rendent les énoncés plus simples et qui suffisent en pratique.

Théorème 5 :

Avec les hypothèses :

$H2$: le sous-espace V vérifie la condition de Haar sur $K \cup H$: c'est-à-dire tout élément $g \in V$ admet au plus $n - 1$ zéros sur $K \cup H$.

$H3$: il existe dans V un élément de signe constant sur H, la meilleure approximation de f dans $\overline{V}$ existe et est unique quel que soit f. Une condition nécessaire et suffisante pour que $g^* \in \overline{V}$ soit cette meilleure approximation est que l'on puisse trouver $n + 1$ points critiques dans S : $P_1, P_2, \ldots, P_{n+1}$ pour l'erreur $e^* = f - g^*$ et une fonctionnelle L basée sur ces points $\left(L(h) = \sum_{i=1}^{n+1} \lambda_i \cdot h(P_i)\right)$ s'annulant sur V tels que : $\lambda_i > 0$ pour $P_i \in E^+$ et $\lambda_i < 0$ pour $P_i \in E^- \cup N$.

L'hypothèse $H2$ est suffisante pour assurer l'unicité, puisqu'elle entraîne qu'une fonctionnelle L s'annulant sur V est forcément basée sur au moins $n + 1$ points. Par contre elle n'est pas nécessaire (cf. th. 4). L'hypothèse $H3$ assure l'existence et entraîne $H1$. En plus, elle sera utile pour la démonstration de la convergence du procédé numérique.

B. OBTENTION NUMERIQUE DE LA MEILLEURE APPROXIMATION

Dans toute la suite nous ferons les hypothèses $H2$ et $H3$ et nous utiliserons le théorème 5.

Le principe de la méthode est de trouver la meilleure approximation sur un ensemble de $n + 1$ points de $H \cup K$ et de faire évoluer ces points de façon à converger vers la meilleure approximation g^*.

1) Meilleure approximation sur un ensemble de n + 1 points :

$$M_k = \{ Q_i^k, i = 1 \text{ à } n + 1 \}$$

Les Q_i^k sont pris dans $K \cup H$. Appelons $\overline{V}_k$ l'ensemble des $g \in V$ qui vérifient $g(P) \leqslant f(P)$ pour tout $P \in M_k \cap H$. Soit $g_k \in \overline{V}_k$ la meilleure approximation sur M_k, c'est-à-dire :

$$\underset{P \in M_k \cap K}{\text{Max}} |f(P) - g_k(P)| = \underset{g \in \overline{V}_k}{\text{Inf}} [\underset{P \in M_k \cap K}{\text{Max}} |f(P) - g(P)|] = \rho_k$$

Remarquons qu'il n'y a aucune raison pour que g_k appartienne à $\overline{V}$. Il n'est pas toujours possible d'utiliser le théorème 5 pour caractériser g_k. Appelons λ_i^k ($i = 1$ à $n + 1$) les coefficients de la fonctionnelle L_k basée sur M_k s'annulant sur V : les λ_i^k sont définis à une constante multiplicative près (car V vérifie la condition de Haar *a fortiori* sur M_k) : il faut que les coefficients λ_i^k correspondant à des Q_i^k appartenant à $H - K$ soient de *signe constant* ; nous supposerons cette propriété réalisée pour M_k ; nous montrerons comment la maintenir pour M_{k+1}. Pour M_0 il suffit de prendre tous les Q_i^0 dans K.

Pour la commodité du calcul nous imposerons aux coefficients λ_i^k deux conditions supplémentaires qui les détermineront complètement :

$$\sum_{i=1}^{n+1} \lambda_i^k \cdot f(Q_i^k) > 0 \tag{1}$$

Les λ_i^k sont maintenant définis à une constante multiplicative positive près. Soit $J_k = < i : (Q_i^k \in K - H)$ ou $(Q_i^k \in H \cap K$ et $\lambda_i^k > 0) >$

$$\sum_{i \in J_k} |\lambda_i^k| = 1 \tag{2}$$

J_k représente l'ensemble des indices qui correspondent à des points extrémaux, les autres étant des points nuls.

La meilleure approximation g_k est alors obtenue en résolvant le système en $\alpha_j \left(g_k = \sum_{j=1}^{n} \alpha_j \cdot f_j \right)$:

$$\begin{cases} g_k(Q_i^k) = f(Q_i^k) \text{ pour } i \notin J_k \\ g_k(Q_i^k) = f(Q_i^k) - \rho_k \cdot \text{signe } (\lambda_i^k) \text{ pour } i \in J_k \end{cases}$$

On a :

$$L_k(f) = L_k(f - g_k) = \sum_{i \in J_k} \lambda_i^k \cdot (f(Q_i^k) - g_k(Q_i^k)) = \rho_k$$

Propriété : Si g_k n'est pas la meilleure approximation g^* on a :

$$\rho_k < \rho$$

En effet, formons $d = g^* - g_k = (f - g_k) - (f - g^*) \in V$. Si l'on avait $\rho_k > \rho$, $d(Q_i^k)$ serait du signe de $f(Q_i^k) - g_k(Q_i^k)$ pour $i \in J_k$:

$$\left.\begin{array}{lll} d(Q_i^k) > 0 & \text{pour} & \lambda_i^k > 0 \\ d(Q_i^k) < 0 & \text{pour} & \lambda_i^k < 0 \end{array}\right\} i \in J_k$$

et par ailleurs, puisque $g^* \in \overline{V}$:

$$d(Q_i^k) \leqslant 0 \qquad \text{pour} \qquad i \notin J_k(\lambda_i^k < 0) ;$$

donc on aurait $L_k(d) > 0$ d'où la contradiction.

En plus, si $\rho_k = \rho$ alors g_k ne peut être que la meilleure approximation de f dans $\overline{V}$ d'après le théorème de caractérisation.

2) Itération

On va remplacer M_k par $M_{k+1} = \{ Q_i^{k+1}, i = 1 \text{ à } n+1 \}$ qui ne diffère de M_k que par un point, de telle façon que l'écart ρ_{k+1} correspondant soit supérieur à ρ_k (dans le cas où g_k n'est pas déjà la meilleure approximation).

a) On choisit d'abord le point nouveau $Q \in S$ qui va être introduit :
Soit $\xi \in K$ tel que $|e_k(\xi)| = \|e_k\|$ avec $e_k = f - g_k$

$$\eta \in H \text{ tel que } e_k(\eta) = \min_{t \in H} e_k(t) = - m_k$$

On choisit comme nouveau point :

$$Q = \xi \text{ si } \|e_k\| - \rho_k > m_k$$
$$Q = \eta \text{ dans le cas contraire.}$$

b) On choisit le point Q_μ^k qui est supprimé et on pose :

$$\begin{cases} Q_i^{k+1} = Q_i^k & i \neq \mu \\ Q_\mu^{k+1} = Q & \end{cases}$$

L'indice μ est choisi de façon à avoir les propriétés suivantes pour λ_i^{k+1} [coefficients de la fonctionnelle L_{k+1} basée sur M_{k+1}, déterminée de la même façon que L_k : c'est-à-dire s'annulant sur V avec en plus les conditions (1) et (2)] :

$$(c) \begin{cases} \text{signe } (\lambda_i^{k+1}) = \text{signe } (\lambda_i^k) \qquad i \neq \mu \\ \text{signe } (\lambda_\mu^{k+1}) = \begin{cases} +1 & \text{si} \quad e_k(Q) = \|e_k\| \\ -1 & \text{si} \quad e_k(Q) = -\|e_k\| \text{ ou } e_k(Q) = -m_k \end{cases} \end{cases}$$

Cela entraîne que les n points communs entre M_k et M_{k+1} seront de même nature (extrémaux positifs, extrémaux négatifs ou critiques nuls). D'autre part, dans l'approximation sur M_{k+1}, le nouveau point Q sera respectivement extrémal positif, extrémal négatif ou critique nul selon que l'on a $e_k(Q) = \|e_k\|$, $e_k(Q) = -\|e_k\|$ ou $e_k(Q) = -m_k$. Enfin, on sera assuré que les λ_i^{k+1} correspondant aux Q_i^{k+1} qui appartiennent à $H - K$ seront de signe constant (négatif).

Montrons qu'avec cet échange on a bien :

$$L_{k+1}(f) = \rho_{k+1} > L_k(f) = \rho_k$$

D'après les conditions (C) on a :

$$L_{k+1}(f) = L_{k+1}(f - g_k) = \sum_{i=1}^{n+1} |\lambda_i^{k+1}| \cdot |e_k(Q_i^{k+1})|$$

$$= \sum_{\substack{i \in J_k \\ i \neq \mu}} |\lambda_i^{k+1}| \cdot |e_k(Q_i^k)| + |\lambda_\mu^{k+1}| \cdot |e_k(Q)|$$

mais comme $|e_k(Q_i^k)| = \rho_k$ pour $i \in J_k$, $i \neq \mu$, on a :

$$\rho_{k+1} = \rho_k \cdot \sum_{\substack{i \in J_k \\ i \neq \mu}} |\lambda_i^{k+1}| + |\lambda_\mu^{k+1}| \cdot |e_k(Q)|$$

On appelle J_{k+1} l'ensemble d'indice analogue à J_k mais relatif à M_{k+1}

Deux cas sont à considérer :

— le nouveau point Q est du type ξ : $|e_k(Q)| = \|e_k\|$: $Q_\mu^{k+1} = Q$ sera alors du type extrémal, $\mu \in J_{k+1}$ donc

$$\sum_{\substack{i \in J_k \\ i \neq \mu}} |\lambda_i^{k+1}| = 1 - |\lambda_\mu^{k+1}|$$

d'où

$$\rho_{k+1} = \rho_k + |\lambda_\mu^{k+1}| \cdot (\|e_k\| - \rho_k)$$

— le nouveau point Q est du type η : $|e_k(Q)| = m_k$: $Q_\mu^{k+1} = Q$ sera alors de type critique nul, $\mu \notin J_{k+1}$ donc

$$\sum_{\substack{i \in J_k \\ i \neq \mu}} |\lambda_i^{k+1}| = 1 \qquad \text{d'où} \qquad \rho_{k+1} = \rho_k + |\lambda_\mu^{k+1}| \cdot (m_k)$$

D'après le critère utilisé pour choisir les deux types possibles, on aura finalement :

$$(R) \quad \boxed{\rho_{k+1} - \rho_k = |\lambda_\mu^{k+1}| \cdot \max\,[\|e_k\| - \rho_k, m_k]}$$

Indiquons maintenant comment obtenir en pratique les conditions (C) [7] : Soit M un ensemble à $n + 2$ points :

$$M = (Q_1^k, Q_2^k, \ldots, Q_{n+1}^k, Q)$$

auquel on fait correspondre le vecteur de R^{n+2} qui définit la fonctionnelle L_k sur M_k :

$$\Lambda^k = (\lambda_1^k, \lambda_2^k, \ldots, \lambda_{n+1}^k, 0)$$

Fixons-nous un indice arbitraire $i_0 \neq n+2$ (nous prendrons dans notre écriture $i_0 = n+1$) et soit Ω le vecteur de R^{n+2} qui représente une fonctionnelle (déterminée à une constante multiplicative près) s'annulant sur V et basée sur les points $(Q_i^k, i \neq i_0$, et $Q)$: $w_{i_0} = 0$:

$$\Omega = (w_1, w_2, \ldots, w_n, 0, w_{n+2})$$

La fonctionnelle L_{k+1}, basée sur $(Q_i^k, i \neq \mu$, et $Q)$ peut être représentée par un vecteur de R^{n+2} de la forme :

$$\bar{\Lambda}^k = \alpha \cdot \Lambda^k + \beta \cdot \Omega,$$

α et β étant choisis de façon à avoir $\bar{\lambda}_\mu^k = 0$:

$$\bar{\Lambda}^k = \beta\left(\Omega - \frac{w_\mu}{\lambda_\mu^k} \Lambda^k\right)$$

Soit encore :

$$\bar{\lambda}_i^k = \begin{cases} \beta \cdot \lambda_i^k \cdot \left(\dfrac{w_i}{\lambda_i^k} - \dfrac{w_\mu}{\lambda_\mu^k}\right) & i \neq n+2 \\ \beta \cdot w_{n+2} & i = n+2 \end{cases}$$

On peut choisir Ω de façon à avoir (changement de signe éventuel) :

$$w_{n+2} \begin{cases} > 0 & \text{si} \quad e_k(Q) = \|e_k\| \\ < 0 & \text{si} \quad e_k(Q) = -\|e_k\| \quad \text{ou} \quad e_k(Q) = -m_k. \end{cases}$$

Pour que les conditions (C) soient satisfaites, il suffit alors d'avoir $\bar{\lambda}_i^k$ du même signe que λ_i^k ; on prend donc μ vérifiant :

$$\frac{w_\mu}{\lambda_\mu^k} = \min_{i=1}^{n+1} \frac{w_i}{\lambda_i^k}$$

β est un coefficient positif que l'on choisit de façon à avoir :

$$\sum_{i=1}^{n+2} |\bar{\lambda}_i^k| = 1$$

Pour avoir les coefficients $\lambda_i^{k+1} (i = 1$ à $n+1)$ de L_{k+1} correspondant aux Q_i^{k+1} il suffit de poser :

$$\lambda_i^{k+1} = \begin{cases} \bar{\lambda}_i^k & i \neq \mu \\ \bar{\lambda}_{n+2}^k & i = \mu \end{cases}$$

Remarque. — La condition (1) est automatiquement vérifiée pour L_{k+1}.

3) Convergence

Théorème 6 :

La suite g_k obtenue par le procédé décrit ci-dessus, converge uniformément vers la meilleure approximation g^* de f dans $\overline{V}$.

Comme $|\lambda_\mu^{k+1}|$ ne peut être nul, on a $\rho_{k+1} > \rho_k$ (sauf si $\|e_k\| = \rho_k$ et $m_k \leqslant 0$ ce qui entraîne $g_k = g^*$). Avant de montrer que ρ_k converge vers ρ, il faut démontrer que $|\lambda_\mu^{k+1}|$ ne peut devenir arbitrairement petit : nous obtiendrons ce résultat en deux lemmes :

Lemme 1 : Il existe une constante $r > 0$ telle que les Q_i^k obtenus dans le procédé vérifient :

$d(Q_i^k, Q_j^k) \geqslant r$ pour tout i, j (de 1 à $n+1, i \neq j$) et pour tout $k = 1, 2, \ldots$

On le montre pour i et j fixés et on prend pour r le minimum des r_{ij} trouvés : $r_{ij} = \operatorname{Inf} (d(Q_k^i, Q_k^j) \mid k = 1, 2, \ldots)$.

Si la propriété n'était pas vraie, il existerait une sous-suite telle que $\lim\limits_{r\to\infty} Q_l^{k_r} = \bar{Q}_l$ $(l = 1$ à $n+1)$ avec $\overline{Q}_i = \overline{Q}_j$.

On pourrait alors trouver $g \in V$ qui coïncide avec f aux n points distincts $\overline{Q}_l$: $f(\overline{Q}_l) = g(\overline{Q}_l)$.

Comme ρ_k est strictement croissant et différent de zéro (sauf peut-être pour $k = 0$), pour $k \geqslant 2$ on a :

$$L_k(f - g_k) = \rho_k > \rho_1 > 0$$

Si $h \in V$ est un élément de signe constant sur H (positif), en posant

$$a = \min_{P \in H} h(P) \qquad (a > 0)$$

$$A = \max_{P \in K} |h(P)|,$$

on a

$$\sum_{i=1}^{n+1} |\lambda_i^k| \leqslant \frac{A}{a} + 1 = \Lambda \qquad (J_k \text{ ne peut être vide})$$

Choisissons $\varepsilon > 0, \varepsilon < \dfrac{\rho_1}{\Lambda}$ et un voisinage convenable V_ε des $\overline{Q}_l$ tel que $P \in V_\varepsilon$ entraîne $|f(P) - g(P)| < \varepsilon$.

Pour r suffisamment grand (posons $k_r = m$) on a :

$$Q_i^m \in V_\varepsilon \qquad \text{pour} \qquad i = 1 \text{ à } n+1$$

Formons :

$$[f(Q_i^m) - g_m(Q_i^m)] - [f(Q_i^m) - g(Q_i^m)] = g(Q_i^m) - g_m(Q_i^m)$$

$d = g - g_m \in V$. On suppose $g_m \neq g$ (Sinon, il suffirait de prendre $g_{m'} = g_{k_{r+1}}$).

$$L_m(d) = L_m(f - g_m) - L_m(f - g)$$

$$= \rho_m - \sum_{i=1}^{n+1} \lambda_i^m \cdot (f(Q_i^m) - g(Q_i^m))$$

$\geqslant \rho_m - \varepsilon \sum_{i=1}^{n+1} |\lambda_i^m| \geqslant \rho_m - \varepsilon \cdot \Lambda > 0$ d'où la contradiction.

Lemme 2 : Il existe une constante $s > 0$ telle que les λ_j^k obtenus par le procédé vérifient :

$|\lambda_j^k| \geqslant s$ pour tout $j = 1$ à $n+1$ et tout $k = 1, 2, \ldots$

Les λ_j^k dépendent uniquement des $Q_j^k (j = 1 \text{ à } n+1)$, V étant fixé. Considérons le compact produit cartésien S^{n+1}. Les points Q^k de composantes $Q_j^k (j = 1 \text{ à } n+1)$ parcourent d'après le lemme 1 une partie compacte $\Delta \subset S^{n+1}$ définie par $d(Q_i, Q_j) \geqslant r$ pour tout i et tout $j (i \neq j)$.

Soit Q^k une suite quelconque de Δ et λ^k, de composantes λ_j^k, la suite correspondante : on notera $\lambda^k = \lambda(Q^k)$. Du fait de la normalisation que nous avons introduite $\left(\sum_{i \in J_k} |\lambda_i^k| = 1\right)$ la fonction λ n'est pas continue : toutefois, si l'on définit $\bar{\lambda}_i^k$ tels que $\sum_{i=1}^{n+1} |\bar{\lambda}_i^k| = 1 : \bar{\lambda}^k = \bar{\lambda}(Q^k)$, la fonction $\bar{\lambda}$ est continue (obtention par résolution d'un système linéaire de déterminant strictement positif (ou négatif) et normalisation). Or on a $\bar{\lambda}^k = \frac{1}{c_k} \lambda^k$ avec $c_k \in [1, \Lambda]$.

Posons

$$s_j = \operatorname*{Inf}_k |\lambda_j^k| = \operatorname*{Inf}_k |\lambda_j(Q^k)|.$$

Si l'on avait $s_j = 0$, on pourrait extraire une sous-suite convergente $\lim_{l \to \infty} Q^{kl} = \bar{Q}$ telle que $\lim_{l \to \infty} \lambda_j(Q^{kl}) = 0$ donc aussi $\lim_{l \to \infty} \bar{\lambda}_j(Q^{kl}) = 0$; on aurait donc $\bar{\lambda}_j(\bar{Q}) = 0$, ce qui est impossible.

On a donc :

$$\rho_{k+1} - \rho_k \geqslant s \cdot \max\left(\|e_k\| - \rho_k, m_k\right).$$

ρ_k est une suite croissante bornée par ρ : appelons β sa limite et montrons que $\beta = \rho$. Supposons $\beta < \rho$:

$$\lim_k \|e_k\| - \rho_k = 0 \qquad \text{donc} \qquad \lim_k \|e_k\| = \beta$$

$$\overline{\lim_{k\to\infty}}\, m_k \leqslant 0$$

De la suite g_k (suite bornée de V) on pourrait extraire une sous-suite convergente g_{k_l}, de limite $\hat{g}$ telle que

$$\hat{g} \in \overline{V} \qquad (\text{puisque} \lim_{l\to\infty} m_{k_l} = m \leqslant 0)$$

$$\|f - \hat{g}\| = \beta < \rho$$

d'où la contradiction : donc $\beta = \rho$.

Toute sous-suite convergente de g_k converge vers g^* (unique). Par ailleurs, g_k est une suite d'un compact. Si g_k ne convergeait pas vers g^* on pourrait extraire une sous-suite dont les éléments resteraient à une distance supérieure à un certain $\delta > 0$ de g^* ; mais de cette suite, on pourrait extraire à nouveau une sous-suite convergente dont la limite ne pourrait être que g^*, d'où la contradiction. *Donc g_k converge (uniformément) vers g^*.*

BIBLIOGRAPHIE

[1] L. Collatz, *Approximation von Funktionen bei einer und bei mehreren unabhängigen Veränderlichen*, Z. Angew. Math. u. Mech. 36, pp. 198-211 (1956).

[2] H. G. Eggleston, *Convexity*, Cambridge Univ. Press (1963).

[3] A. N. Kolmogoroff, *Eine Bemerkung zu den Polynomen von P.L. Tschebyscheff, die von einer gegebenen Funktion am wenigsten abweichen*, Usp. Math. Nauk 3, pp. 216-221 (1948).

[4] P. J. Laurent, *Cours de théorie de l'approximation*, fasc. 1, Faculté des Sciences de Grenoble (1966).

[5] G. Meinardus, *Approximation von Funktionen und ihre numerische Behandlung*, Springer Verlag (1964).

[6] T. J. Rivlin and H. S. Shapiro, *A unified approach to certain problems of approximation and minimization*, J. Soc. Ind. Appl. Math. (SIAM) 9, pp. 670-699 (1961).

[7] E. Stiefel, *Über diskrete und lineare Tschebyscheff-Approximation*, Num. Math. 1, pp. 1-28 (1959).