Approximation au sens des moindres carrés

Grégory Vial

5 mars 2003

Une manière simple d'approcher une fonction f par une suite de polynômes p_n sur un intervalle I est d'utiliser l'interpolation de Lagrange. Cependant, il n'y a pas nécessairement convergence de la suite (p_n) vers la fonction d'origine quand le degré des polynômes tend vers l'infini (c'est ce que l'on nomme le *phénomène de Runge*, voir [2]). Une solution consiste à fixer le degré et à effectuer une interpolation par morceaux (il y a alors convergence de (p_n) vers f quand la taille des sous-intervalles tend vers 0).

Il arrive cependant qu'on préfère utiliser une représentation sur l'intervalle I tout entier. On adopte alors un point de vue différent de l'interpolation : plutôt que d'imposer que p_n et f coïncident en certains nœuds, on demande qu'ils soient proches d'une manière plus globale. Précisément, soit f une fonction continue sur l'intervalle I, on s'intéressera aux deux cas suivants.

Exemple 1. On recherche un polynôme p_n de degré inférieur ou égal à n tel que la quantité

$$\int_I |f(x) - p_n(x)|^2 \,\mathrm{d}x \quad \text{soit minimale.}$$

Exemple 2. On recherche un polynôme p_n de degré inférieur ou égal à n tel que la quantité

$$\sum_{i=0}^{N} |f(x_i) - p_n(x_i)|^2 \quad \text{soit minimale,}$$

où $(x_i)_{0\leq i\leq N}$ est une subdivision de I et $N \geq n$.

Point de vue théorique

Soit F un espace vectoriel muni d'une forme bilinéaire symétrique positive $\langle\cdot,\cdot\rangle$; la semi-norme associée est notée $\|\cdot\|$. On considère $\mathbb{P}$ un sous-espace de F de dimension finie et on suppose que $\|\cdot\|$ est une norme sur $\mathbb{P}$.

Pour $f \in F$, on recherche $p \in \mathbb{P}$ tel que la quantité $\|f - p\|$ soit minimale. Les deux exemples cités plus haut rentrent dans ce cadre :

Exemple 1. $F = \mathscr{C}(I, \mathbb{R})$, $\mathbb{P} = \mathbb{R}_n[X]$ et $\langle f, g\rangle_c = \displaystyle\int_I f(x)g(x)\,\mathrm{d}x$.

Exemple 2. $F = \mathscr{C}(I, \mathbb{R})$, $\mathbb{P} = \mathbb{R}_n[X]$ et $\langle f, g\rangle_d = \displaystyle\sum_{i=0}^{N} f(x_i)g(x_i)$, où (x_i) est une subdivision de I et $N \geq n$.

Proposition 1 *Soit $f \in F$. Il existe un unique élément $p \in \mathbb{P}$ tel que la quantité $\|f - p\|$ soit minimale.*

Démonstration. Introduisons l'application $\Phi : q \mapsto \|f-q\|^2 = \|q\|^2 - 2\langle f,q\rangle + \|f\|^2$. La fonctionelle Φ est continue et vérifie

$$\lim_{\|q\|\to+\infty} \Phi(q) = +\infty.$$

Donc Φ admet un minimum sur $\mathbb{P}$ (qui est de dimension finie). Tout polynôme p, en lequel il est atteint, est un point critique de Φ :

$$\forall q \in \mathbb{P}, \quad \langle f-p, q\rangle = 0.$$

Si on suppose que Φ atteint son minimum en deux polynômes p_1 et p_2, alors pour le choix $q = p_1 - p_2$, on obtient $\|p_1 - p_2\| = 0$, d'où $p_1 = p_2$, car $\langle\cdot,\cdot\rangle$ est définie sur $\mathbb{P}$. ■

Remarque. L'application Φ est quadratique, cela explique le terme *moindres carrés*. Si on ne s'intéresse pas à l'approximation au sens des moindres cubes, par exemple, c'est parce-que dans ce cas la différentielle de Φ n'est plus linéaire.

On peut préciser la convergence du polynôme d'approximation pour les deux exemples considérés.

Proposition 2 *Soit $p_n \in \mathbb{R}_n[X]$ le polynôme obtenu dans le cadre de l'exemple 1 (resp. 2). Alors p_n converge vers f dans la norme $\|\cdot\|_c$ (resp. semi-norme $\|\cdot\|_d$) quand n tend vers l'infini.*

Démonstration. Les deux (semi-)normes $\|\cdot\|_c$ et $\|\cdot\|_d$ sont dominées par la norme uniforme :

$$\forall g \in F, \quad \|g\|_c \le \sqrt{|I|}\ \|g\|_\infty \quad \text{et} \quad \|g\|_d \le \sqrt{N}\ \|g\|_\infty .$$

Pour toute fonction continue f, il existe une suite $q_n \in \mathbb{R}_n[X]$ convergeant uniformément vers f. Par définition de p_n, pour la norme continue $\|\cdot\|_c$,

$$\|f-p_n\|_c \le \|f-q_n\|_c \le \sqrt{|I|}\ \|f-q_n\|_\infty \longrightarrow 0.$$

Il en est de même pour la norme discrète $\|\cdot\|_d$. ■

Mise en œuvre pratique

On s'intéresse dans ce paragraphe au calcul explicite du polynôme d'approximation au sens des moindres carrés.

Norme continue

Soit f une fonction continue sur l'intervalle I ; on recherche un polynôme p_n, de degré inférieur ou égal à n, tel que

$$\int_I |f-p_n|^2 \,\mathrm{d}x \quad \text{soit minimale.}$$

On a prouvé, dans la démonstration de la proposition 1, la relation de Pythagore

$$\forall q \in \mathbb{R}_n[X], \quad \int_I p_n q \,\mathrm{d}x = \int_I f q \,\mathrm{d}x. \tag{1}$$

Il s'agit d'un système linéaire carré : soit $(e_0, \dots, e_n)$ une base de $\mathbb{R}_n[X]$, si on recherche p_n sous la forme $p_n = \sum \alpha_j e_j$, alors (1) devient

$$\forall i = 0, \dots n, \quad \sum_{j=0}^{n} \underbrace{\left(\int_I e_i e_j \,\mathrm{d}x \right)}_{a_{ij}} \alpha_j = \underbrace{\int_I f e_i \,\mathrm{d}x}_{b_i},$$

qui n'est autre que le système $A\alpha = b$, avec des notations naturelles.

Le choix de la base canonique $e_j = X^j$ conduit à la matrice de Hilbert, qui est très mal conditionnée (voir [3]), et n'est donc jamais utilisé. Il est préférable de choisir les polynômes e_j tels que la matrice A soit l'identité, c'est-à-dire la famille orthogonale pour le produit scalaire de $\mathrm{L}^2(I)$. Le calcul des composantes du vecteur b nécessite la plupart du temps une intégration numérique.

Norme discrète

Soit f une fonction continue sur l'intervalle I ; on recherche un polynôme p_n, de degré inférieur ou égal à n, tel que

$$\sum_{i=0}^{N} |f(x_i) - p_n(x_i)|^2 \quad \text{soit minimale.} \tag{2}$$

On peut écrire matriciellement de problème de minimisation : soient A la matrice de Van-Der-Monde de coefficient générique $a_{ij} = x_i^j$ et b le vecteur de composantes $b_i = f(x_i)$. Alors (2) devient

$$\text{Chercher } \alpha \in \mathbb{R}^n \text{ tel que } \|A\alpha - b\|_2 \quad \text{soit minimale,} \tag{3}$$

les composantes du vecteur α sont celles du polynôme p_n dans la base canonique de $\mathbb{R}_n[X]$.

Notons que la matrice A comporte N lignes et n colonnes ; comme $N \geq n$, on parle de matrice sur-déterminée.

Proposition 3 *Le problème* (3) *est équivalent au système linéaire suivant, dit équations normales :*

$$A^\mathsf{T} A\alpha = A^\mathsf{T} b.$$

Démonstration. Il s'agit simplement d'exprimer que α annule la différentielle de la fonctionnelle quadratique $\|A\alpha - b\|_2^2$. ∎

La proposition précédente fournit un procédé de construction du polynôme d'approximation p_n : la matrice $A^\mathsf{T} A$ étant symétrique définie positive, on peut utiliser les méthodes de Choleski ou du gradient conjugué. Cependant, les équations normales sont souvent assez mal conditionnées ; on leur préfère une méthode basée sur la décomposition QR de la matrice A (voir [1]).

Proposition 4 *Soit Q, orthogonale, et R, triangulaire supérieure, deux matrices telles que $A = QR$. On écrit, par blocs,*

$$R = \left[\frac{R_1}{0} \right] \quad \text{et} \quad Q^\mathsf{T} b = \left[\frac{d_1}{d_2} \right].$$

Alors α est donné par $R_1 \alpha = d_1$.

Démonstration. Il suffit de remarquer que la norme euclidienne est stable par transformation orthogonale. ■

Remarque. En général, le nombre d'équations $N+1$ est beaucoup plus grand que la dimension $n+1$ de l'espace $\mathbb{R}_n[X]$; le cas particulier $n=1$ correspond à la *régression linéaire*.

La fonction `matlab` suivante calcule le polynôme d'approximation de `f` :

```
function valp=approx_mc(strf,x,n,t)
% Calcule l'évaluation du polynôme
% d'approximation de degré n de f
% sur l'intervalle x aux abscisses t.
 A=[];
 for j=0:n
   A=[A,x.^j];
 end
 [Q,R]=qr(A);
 d=Q'*feval(strf,x);
```

```
d1=d(1:n+1);
R1=R(1:n+1,1:n+1);
p=R1\d1;
valp=polyval(flipud(p),t);

  -------------

function y=f(x)
y=1./(1+x.^2);
```

Noter l'utilisation de la commande `feval` pour l'appel de la fonction `f`. Il est possible d'utiliser les équations normales :

```
p=(A'*A)\(A'*feval(strf,x));
```

ou encore la résolution de `matlab` pour le calcul du vecteur `p` :

```
p=A\feval(strf,x);
```

qui résout le système au sens des moindres carrés quand il est sur-déterminé. Il existe aussi une fonction `matlab` qui calcule directement les coefficients du polynôme d'approximation au sens des moindres carrés : `polyfit`.

À l'aide du script `main.m`, on a tracé la fonction `f` et son polynôme d'approximation au sens des moindes carrés pour $n=10$ et on a mis en évidence la convergence

$$\|f-p_n\|_d \xrightarrow[n\to+\infty]{} 0.$$

```
N=100;
M=5;
x=linspace(-M,M,N)';
t=linspace(-M,M,1000);

% Tracé pour n=10

n=10;
pt=approx_mc('f',x,n,t);
close all
plot(t,f(t),t,pt)
title('f et p pour n=10')
xlabel('x');
legend('f','p');
```

```
% Convergence en n

nn=1:30;
err=[];
for n=nn
  px=approx_mc('f',x,n,x);
  err=[err,norm(f(x)-px,'inf')];
end
figure
plot(nn,err)
title('Erreur en norme discrète
en fonction du degré')
xlabel('degré n')
ylabel('erreur ||f-p||_d')
```

Les graphes construits par le programme `main.m` sont consignés dans la figure 1. On a choisi la fonction f définie sur $[-5, 5]$ par

$$f(x) = \frac{1}{1+x^2},$$

qui est l'exemple habituellement choisi pour illustrer le phénomène de Runge dans l'interpolation de Lagrange. En revanche, l'approximation au sens des moindres carrés converge quand le degré augmente.

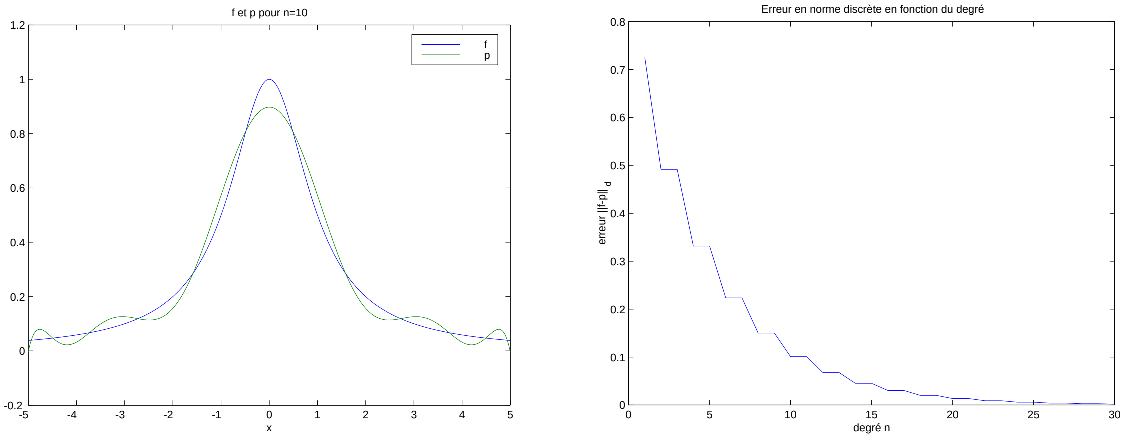

FIG. 1 – Résultats du script `main.m`.

Régression linéaire

Le cas particulier $n = 1$ est particulièrement utile quand on cherche à mettre en évidence une vitesse de convergence : il permet d'approcher un nuage de points par une droite "moyenne". Le script `reglin.m` illustre son utilisation :

```
% On clique avec le bouton
% de droite le dernier point.
figure
plot([0,1,1,0],[0,0,1,1],'k')
hold on
button=1;
xx=[];yy=[];
while (button~=3)
  [x,y,button]=ginput(1);
  plot(x,y,'ro')
  xx=[xx;x];
  yy=[yy;y];
end
p=polyfit(xx,yy,1);
t=0:.01:1;
plot(t,p(1)*t+p(2));
title('Regression linéaire')
```

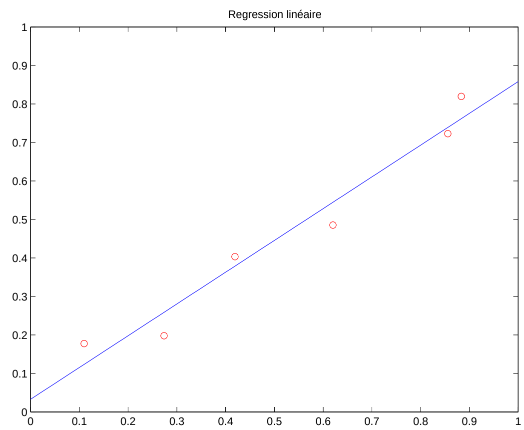

FIG. 2 – Résultat du script `reglin.m`.

Références

[1] P. G. Ciarlet. *Introduction à l'analyse numérique matricielle et à l'optimisation*. Collection Mathématiques Appliquées pour la Maîtrise. [Collection of Applied Mathematics for the Master's Degree]. Masson, Paris 1982.

[2] J.-P. Demailly. *Analyse numérique et équations différentielles*. PUG, Grenoble 1996.

[3] M. Schatzman. *Analyse numérique*. InterEditions, Paris 1991. Cours et exercices pour la licence. [Course and exercises for the bachelor's degree].