Version continue de l'algorithme d'Uzawa

Continuous version of the Uzawa algorithm

Bertrand Maury [a]

[a] *Laboratoire Jacques–Louis Lions, Université Pierre et Marie Curie, Boîte courrier 187, 75252 Paris Cedex 05 France*

Abstract

In [3], we proposed an algorithm to approximate the projection of a function $f \in H_0^1(\Omega)$ (where Ω is a convex domain) onto the cone of convex functions. This algorithm is based on a dual expression of the constraint, which leads to a saddle–point problem which has no solution in general. We show here that the Uzawa algorithm for this saddle–point problem can be seen as the semi–discretization of an evolution equation

$$\frac{d\lambda}{dt} + \partial\Psi(\lambda) \ni 0,$$

where Ψ is a convex, l.s.c., proper function. In case the saddle–point problem has no solution, one has $0 \in \overline{R\,(\partial\Psi)}$ but $\partial\Psi^{-1}(0) = \emptyset$. We establish that $\lambda(t)$ is then divergent, and that a subsequence of the associated trajectory in the primal space converges weakly to the solution of the initial projection problem.

Résumé

Nous avons proposé dans [3] un algorithme permettant d'approximer la projection d'une fonction $f \in H_0^1(\Omega)$ (où Ω est un domaine convexe) sur le cône des fonctions convexes. Cet algorithme est basé sur une expression duale de la contrainte de convexité, qui conduit à un problème de point–selle qui n'a pas de solution en général. Nous montrons ici que l'algorithme d'Uzawa appliqué à cette situation peut être vu comme une discrétisation semi–implicite d'une équation d'évolution du type

$$\frac{d\lambda}{dt} + \partial\Psi(\lambda) \ni 0,$$

où Ψ est une fonction convexe, propre, et s.c.i. Dans le cas où le problème de point–selle n'admet pas de solution, on a

$$0 \in \overline{R\,(\partial\Psi)} \text{ mais } \partial\Psi^{-1}(0) = \emptyset.$$

Nous établissons que $\lambda(t)$ diverge alors, mais qu'une sous–suite de la composante primale de la trajectoire converge faiblement vers la solution du problème de projection initial.

Email address: `maury@ann.jussieu.fr` (Bertrand Maury).

Preprint submitted to Elsevier Science *7th September 2005*

Abridged English version

Let H and Λ be two Hilbert spaces, and let $K \subset H$ be defined as

$$K = \{u \in H\,,\ (\mu, Bu) \leq 0 \quad \forall \mu \in C\}\,,$$

where $C \subset \Lambda$ is a close convexe cone, and $B \in \mathcal{L}(H, \Lambda)$. We call $(\mathcal{P})$ the problem which consists in finding the projection of a vector $f \in H$ onto K, and $(\mathcal{P}')$ its saddle–point formulation: $(u, \lambda) \in H \times C$,

$$L(u, \mu) \leq L(u, \lambda) \leq L(v, \lambda) \quad \forall v \in H\,,\ \forall \mu \in C\,,\ L(v, \mu) = \frac{1}{2}\,|v - f|^2 + (v, B^\star \mu).$$

The Uzawa algorithm is a recursive process to build a sequence of Lagrange multipliers λ^k which is such that the associated sequence in the primal space $u^k = f - B^\star \lambda^k$ converges strongly to the solution of $(\mathcal{P})$. This algorithm can be seen as the semi–discretization of the evolution equation

$$\frac{d\lambda}{dt} + \partial\Psi(\lambda) \ \ni\ 0 \quad \text{with} \quad \Psi(\lambda) = I_C(\lambda) + \frac{1}{2}\,|f - B^\star \lambda|^2\,.$$

Our aim is to study the asymptotic behaviour of the primal trajectory $u(t) = f - B^\star \lambda(t)$ when problem $(\mathcal{P}')$ has no solution. Our main result (proposition 8) states that, in case $(\mathcal{P}')$ has no solution (which is equivalent to $0 \notin R(\partial\Psi)$), $\lambda(t)$ is not bounded, and there exists a subsequence $t^n \nearrow +\infty$ such that $u(t^n) = f - B^\star \lambda^n$ converges weakly to the solution $u = P_K f$ of problem $(\mathcal{P})$.

1. Introduction

La prise en compte de la contrainte de convexité est un problème crucial en économie (voir par exemple [7] pour une application en tarification optimale). Il n'existe pas à l'heure actuelle d'algorithme pleinement satisfaisant pour gérer cette contrainte. Nous avons proposé dans [3] un algorithme basé sur une expression duale de la convexité, mais le problème de point–selle obtenu n'admet pas de solution. Nous montrons ici que, malgré la non–existence d'un point–selle, l'algorithme d'Uzawa permet d'approximer dans un certain sens (voir section 5) la solution du problème primal. Nous nous limitons dans cette première approche à l'étude d'une équation d'évolution dont l'algorithme d'Uzawa est une discrétisation semi–implicite. Dans la formulation abstraite qui suit, H joue le rôle d'un espace fonctionnel du type $H^1_0(\Omega)$, et K est cône des fonctions de cet espaces qui sont convexes, exprimé comme image réciproque par une application linéaire continue B du cône polaire d'un cône convexe fermé dans l'espace des multiplicateurs de Lagrange Λ.

2. Problème primal et formulation point–selle

Soient H et Λ deux espaces de Hilbert, et $K \subset H$ un cône convexe fermé défini de la façon suivante :

$$K = \{v \in H\,,\ (Bv, \mu) \leq 0 \quad \forall \mu \in C\}\,,$$

où $C \subset \Lambda$ est un cône convexe fermé de sommet 0, et $B \in \mathcal{L}(H, \Lambda)$. On considère le problème suivant :

$$(\mathcal{P}) \quad \begin{cases} u \in K, \\ J(u) = \inf\limits_{v \in K} J(v) \ \text{ avec } J(v) = \dfrac{1}{2}\,|v - f|^2 \end{cases}$$

dont la solution est la projection $u = P_K f$ de f sur K. On introduit le Lagrangien L défini sur l'espace produit $H \times C$ par $L(v, \mu) = J(v) + (Bv, \mu)$.

Definition 1 *On appelle point–selle de L un couple $(u,\lambda) \in H \times C$ tel que*

$$L(u,\mu) \leq L(u,\lambda) \leq L(v,\lambda) \quad \forall v \in H\ ,\ \forall \mu \in C. \tag{1}$$

Le lien avec le problème de projection initial est assuré par les deux propriétés élémentaires suivantes :
Proposition 2 *Le couple $(u,\lambda) \in H \times C$ est point–selle du Lagrangien L si et seulement s'il est solution du problème*

$$(\mathcal{P}') \quad \left\{ \begin{array}{lll} u & + B^\star\lambda & = f \\ Bu & & \geq 0 \\ (Bu,\lambda) & & = 0 \end{array} \right.$$

où $\geq$ est ici utilisé au sens de l'ordre partiel associé au cône convexe C°, cône polaire de C :

$$\xi_2 \geq \xi_1 \Longleftrightarrow \xi_2 - \xi_1 \in C^o = \{\xi \in \Lambda\ ,\ (\mu,\xi) \leq 0 \quad \forall \mu \in C\}\,.$$

Proposition 3 *Pour tout $(u,\lambda) \in H \times C$ solution de $(\mathcal{P}')$, u est solution du problème $(\mathcal{P})$.*

L'existence d'une solution au problème $(\mathcal{P})$ ne garantit pas l'existence d'un point–selle, mais le problème $(\mathcal{P}')$ peut être résolu de façon approchée au sens suivant :
Proposition 4 *Soit u la solution de $(\mathcal{P})$. Pour tout $\varepsilon > 0$, il existe $\lambda \in C$ tel que $|u + B^\star\lambda - f| < \varepsilon$.*

Démonstration. – On introduit l'ensemble

$$K^\circ = \{v \in H\ ,\ (v,w) \leq 0 \quad \forall w \in K\}\,,$$

cône polaire de K. Comme $u = P_K f$, on a (voir Moreau [6]) $f = u + f^\circ$, avec $f^\circ = P_{K^\circ} f$. Il suffit donc de montrer que $B^\star(C)$ est dense dans K° pour établir le résultat. L'inclusion de $\overline{B^\star(C)}$ dans K° est évidente. Supposons que cette inclusion soit stricte, et considérons $v \in K^\circ$, $v \notin \overline{B^\star(C)}$. Il existe alors, d'après le théorème de Hahn–Banach, $w \in H$ tel que $(w, B^\star\lambda) < (w,v)$ pour tout $\lambda \in C$. Comme C est un cône de sommet 0, on a $(w,B^\star\lambda) \leq 0$ pour tout $\lambda \in C$, d'où $w \in K$, et $(w,v) > 0$, ce qui est absurde car $v \in K^\circ$. □

3. Algorithme d'Uzawa

Pour λ^0 et $\rho > 0$ donnés, l'algorithme d'Uzawa consiste en la construction d'une suite (λ^k) dans Λ selon le procédé de récurrence suivant :

$$\lambda^{k+1} = P_C\left(\lambda^k + \rho B(f - B^\star\lambda^k)\right), \tag{2}$$

où P_C est l'opérateur de projection sur le convexe fermé C, et $B^\star \in \mathcal{L}(\Lambda, H)$ est l'adjoint de B.
Proposition 5 *On suppose qu'il existe un point–selle (u,λ) de L. On a alors convergence de la suite $u^k = f - B^\star\lambda^k$ vers la solution u du problème $(\mathcal{P})$ initial dès que $\rho \leq 2/\|BB^\star\|$.*

Démonstration. – La démonstration donnée dans Ciarlet [4] en dimension finie s'applique sans modification au cas général (aucun argument de compacité n'est utilisé). □

Si de nombreux travaux portent sur la convergence éventuelle de la suite (λ^k), notamment en lien avec le problème de Stokes (K est alors le sous–espace vectoriel des champs de vitesse à divergence nulle, et λ est le champ de pression), le comportement asymptotique de cet algorithme dans le cas où l'on n'a pas existence d'un point–selle n'a pas été étudié à notre connaissance.

4. Equation d'évolution

Soit I_C la fonction indicatrice de C ($I_C(\mu) = 0$ si $\mu \in C$, $I_C(\mu) = +\infty$ sinon). Comme C est un ensemble convexe et fermé, on a

$$\partial I_C(\lambda) = \{\mu - \lambda\,,\ \lambda = P_C \mu\}\,. \tag{3}$$

L'algorithme d'Uzawa peut donc s'écrire $(\lambda^{k+1} - \lambda^k)/\rho + \partial I_C(\lambda^{k+1}) + BB^\star \lambda^k \ \ni\ Bf$. Il apparaît ainsi comme une discrétisation semi–implicite de l'équation d'évolution

$$\frac{d\lambda}{dt} + (\partial I_C + BB^\star)(\lambda) \ \ni\ Bf. \tag{4}$$

Proposition 6 *Pour toute condition initiale $\lambda^0 \in C$, il existe une unique fonction $t \mapsto \lambda(t)$ solution de (4) avec $\lambda(0) = \lambda^0$.*

Démonstration. – Il suffit (voir Brezis [1]) de montrer que l'opérateur multivoque $\partial I_C + BB^\star$ est maximal monotone. Le premier terme ∂I_C est maximal monotone car C est convexe fermé, et d'autre part $I + BB^\star$ est un isomorphisme bicontinu d'après le théorème de Lax–Milgram, donc $BB^\star$ est maximal monotone. Leur somme est monotone car l'un des deux (l'opérateur univoque $BB^\star$) est Lipschitzien. □

5. Comportement asymptotique de la trajectoire

Cas parabolique

On établit ici une première propriété de convergence dans le cadre d'hypothèses qui assurent l'existence et l'unicité d'un point–selle.

Proposition 7 *On suppose que $B \in \mathcal{L}(H, \Lambda)$ est surjectif. Pour toute condition initiale $\lambda^0 \in C$, on a convergence de $\lambda(t)$, solution de (4), vers l'unique λ^∞ tel que $(f - B^\star\lambda^\infty, \lambda^\infty)$ est solution de $(\mathcal{P}')$.*

Démonstration. – L'opérateur $BB^\star$ est elliptique, le second membre Bf est constant (donc dans L^1_{loc}) : on peut en déduire (voir Brezis [1]) la convergence de $\lambda(t)$ vers

$$\lambda^\infty = (\partial I_C + BB^\star)^{-1} (Bf)\,. \tag{5}$$

On pose $u^\infty = f - B^\star\lambda^\infty$, de telle sorte que $u^\infty + B^\star\lambda = f$. L'identité (5) traduit, conformément à la caractérisation (3) de ∂I_C, que λ^∞ est la projection sur C de $\lambda^\infty + Bu^\infty$. Or tout élément de Λ se décompose en la somme de sa projection sur C et sa projection sur C° (voir Moreau [6]). On a donc $Bu^\infty \in C^\circ$, c'est à dire $Bu^\infty \geq 0$. Le couple $(u^\infty, \lambda^\infty)$ est donc solution de $(\mathcal{P}')$. □

Cas général

On s'intéresse maintenant au cas général : B est simplement supposé linéaire continu.

Proposition 8 *On note $\lambda(t)$ la solution de (4) associé à la condition initiale $\lambda^0 \in C$. On a l'alternative suivante :*

(i) Si le problème $(\mathcal{P}')$ admet une solution, alors $\lambda(t)$ et $u(t) = f - B^\star\lambda(t)$ convergent faiblement vers λ^∞ et u^∞, respectivement, tels que le couple $(u^\infty, \lambda^\infty)$ est solution de $(\mathcal{P}')$.

(ii) Si le problème $(\mathcal{P}')$ n'admet pas de solution, alors $\lambda(t)$ n'est pas borné, et il existe une sous–suite (t^n) de temps croissants vers $+\infty$ tels que $u(t^n)$ converge faiblement vers $P_K f$, solution de $(\mathcal{P})$.

Démonstration. – L'équation d'évolution peut se mettre sous la forme $d\lambda/dt + \partial\Psi(\lambda) \ \ni\ 0$, où Ψ est la fonction convexe s.c.i.

$$\Psi(\lambda) = I_C + \frac{1}{2}\left|B^\star\lambda - f\right|^2 = I_C + \frac{1}{2}\left|u\right|^2,$$

avec $u = u(\lambda) = f - B^\star(\lambda)$.

Si $(\mathcal{P}')$ admet une solution (u^∞, μ^∞), alors Ψ atteint son minimum en μ^∞. Le premier terme de l'alternative entre donc dans le cadre des résultats établis dans [2] (théorèmes 2 et 3), qui assurent la convergence faible de λ vers un λ^∞ tel que $(f - B^\star\lambda^\infty, \lambda^\infty)$ est point–selle de L, donc solution de $(\mathcal{P}')$. La convergence faible de $u = f - B^\star\lambda$ vers $u^\infty = f - B^\star\lambda^\infty$ est immédiate.

On se place maintenant dans le cas où $(\mathcal{P}')$ n'admet pas de solution. On utilisera les propriétés suivantes (voir Brezis [1] ou Haraux [5]) :

(i) La fonction $t \mapsto \lambda(t)$ admet une dérivée à droite $d^+\lambda/dt$ pour tout $t > 0$, égale à la projection de 0 sur $\partial\Psi(\lambda)$, notée $\partial\Psi^0(\lambda)$. On a de plus

$$\left|\partial\Psi^0(\lambda(t))\right| \le \left|\partial\Psi^0(\mu)\right| + \frac{1}{t}\left|\mu\right| \quad \forall\mu \in C. \tag{6}$$

(ii) La fonction $t \mapsto \Psi(\lambda(t))$ est décroissante.

L'identité $\overline{B^\star(C)} = K^\circ$ (voir prop. 4) assure que $\partial\Psi^0(\mu)$ peut être rendu arbitrairement petit sur C. Cette remarque et la propriété (i) permettent d'établir que $\partial\Psi^0(\lambda(t))$ tend vers 0 quand t tend vers l'infini. Comme $\partial\Psi = \partial I_C - Bu$, il existe $h = h(t) \in \partial I_C(\lambda)$ tel que $\varepsilon(t) = Bu - h(t)$ tend vers 0. L'ensemble C étant un cône convexe, $\lambda + C$ est inclus dans C pour tout $\lambda \in C$, donc nécessairement $h \in C^\circ$. La distance de Bu à C° tend donc vers 0.

On suppose dans un premier temps que $\lambda(t)$ est non borné. Il existe alors une suite croissante (t^n) tendant vers $+\infty$ telle que $(\lambda, d^+\lambda/dt) \ge 0$ en t^n pour tout n. D'après (ii), $|u(t)| = |f - B^\star\lambda(t)|$ est décroissant, donc borné. Il existe donc une suite extraite, que nous noterons toujours (t^n), telle que $u(t^n)$ converge faiblement vers $u^\infty \in H$. Montrons que $u^\infty \in K$. Pour tout $\mu \in C$, on a

$$(Bu^\infty, \mu) = (u^\infty, B^\star\mu) = \lim(u(t^n), B^\star\mu) = \lim(Bu(t^n), \mu) \le 0, \tag{7}$$

car $Bu(t^n) = \varepsilon(t^n) + h(t^n)$, avec $\varepsilon(t^n) \to 0$, et $h(t^n) \in C^\circ$ pour tout n. La suite $B^\star\lambda(t^n)$ converge faiblement vers un élément $f^\circ \in H$ tel que $u^\infty + f^\circ = f$, et f° appartient à $\overline{B^\star(C)}$ car cet ensemble est faiblement séquentiellement fermé comme convexe fermé. On a donc

$$f = u^\infty + f^\circ\,,\ u^\infty \in K\,,\ f^\circ \in K^\circ. \tag{8}$$

Pour établir que u^∞ est bien la projection de f sur K, on montre que f° est la projection de f sur K°. Comme $K^\circ = \overline{B^\star(C)}$, il suffit de montrer

$$(f - f^\circ, v - f^\circ) \le 0 \quad \forall v \in B^\star(C).$$

Soit $\mu \in C$. Le produit scalaire ci–dessus étant convexe s.c.i. par rapport à f°, on a

$$\begin{aligned}(f - f^\circ, B^\star\mu - f^\circ) &\le \liminf(f - B^\star\lambda(t^n), B^\star\mu - B^\star\lambda(t^n)) && (9)\\ &= \liminf(u(t^n), B^\star\mu - B^\star\lambda(t^n)) && (10)\\ &= \liminf(Bu(t^n), \mu - \lambda(t^n)). && (11)\end{aligned}$$

Or $Bu(t^n)$ s'écrit $h(t^n) + d^+\lambda/dt(t^n)$, avec $h(t^n) \in \partial I_C(\lambda(t^n))$. On a donc

$$(Bu(t^n), \mu - \lambda(t^n)) = (h(t^n), \mu - \lambda(t^n)) + \left(\frac{d^+\lambda}{dt}(t^n), \mu - \lambda(t^n)\right). \tag{12}$$

Le premier terme du membre de droite est négatif par définition du sous–différentiel de I_C en $\lambda(t^n)$. Pour le second terme, on a

$$\left(\frac{d^+\lambda}{dt}(t^n),\mu\right) \longrightarrow 0 \quad \text{et} \quad -\left(\frac{d^+\lambda}{dt}(t^n),\lambda(t^n)\right) \leq 0. \tag{13}$$

La lim inf de (11) est donc négative. On a donc $f^\circ = P_{K^\circ} f$, et donc $u^\infty = P_K f$ est la solution de $(\mathcal{P})$.

Montrons pour finir que, dans le cadre du second terme de l'alternative, $\lambda(t)$ est nécessairement non borné. On raisonne par l'absurde en supposant $\lambda(t)$ borné. Le résultat obtenu précédemment est toujours valable. En effet, la convergence de $d^+\lambda/dt$ vers 0 assure que le second produit scalaire de (13) tend vers 0, ce qui assure que la lim inf est bien négative. Mais comme $\lambda(t^n)$ est borné, on peut extraire une sous–suite (toujours notée t^n) telle que $\lambda(t^n)$ converge faiblement vers un $\lambda^\infty \in C$. On a $u^\infty = f - B^\star\lambda^\infty$, d'où $(u^\infty, \lambda^\infty)$ solution de $(\mathcal{P}')$, ce qui est absurde. □

6. Exemple

Comme annoncé en introduction, ce travail est motivé par la recherche d'algorithmes permettant d'approximer la projection d'une fonction sur le cône des fonctions convexes. Différentes manières de construire l'opérateur B sont décrites dans [3]. Dans tous les cas envisageables, B n'est pas à image fermée. Nous proposons ici d'illustrer le problème par un exemple plus simple. On considère $H = \Lambda = \ell^2$, et l'on s'intéresse au problème de la projection sur l'ensemble K des suites de H décroissantes. Le cône convexe K peut s'écrire

$$K = \{u \in H\,,\ (Bu,\mu) \leq 0\,,\ \forall \mu \in C\}\,,$$

où $B \in \mathcal{L}(H,\Lambda)$ est défini par

$$B\ :\ u = (u_n)_{n\geq 0} \longmapsto (u_{n+1} - u_n)_{n\geq 0},$$

et C est le cône convexe fermé des suites dont tous les termes sont positifs. Là encore le problème initial admet trivialement une solution unique, mais le problème de point–selle associé est en général mal posé, car $B^\star(C)$ est strictement inclus dans K°. On vérifie notamment que la suite des multiplicateurs de Lagrange construits par l'algorithme d'Uzawa n'est pas bornée lorsque la distance de f à $B^\star C$ n'est pas atteinte. Les tests numériques effectués suggèrent en revanche que l'on a convergence (forte dans les cas traités) de la suite des composantes primales vers la projection recherchée.

Références

[1] H. Brezis, *Opérateurs maximaux monotones et semi–groupes de contraction dans les espaces de Hilbert*, North Holland publishing company 1973.

[2] R. E. Bruck Jr, *Asymptotic Convergence if Nonlinear Contraction Semigroups in Hilbert Space*, Journal of Functional Analysis, **18**, 15–26, 1975.

[3] G. Carlier, T. Lachand-Robert, B. Maury, *H^1–projection into the set of convex functions : a saddle–point formulation*, ESAIM Proceedings, CEMRACS 1999.

[4] P.G. Ciarlet, *Introduction à l'Analyse Numérique Matricielle et à l'Optimisation*, ed. Masson, Paris.

[5] A. Haraux, *Nonlinear Evolution Equations – Global begaviour of Solutions*, Lecture Notes in Mathematics, Springer 1981.

[6] J.J. Moreau, *Décomposition orthogonale d'un espace Hilbertien selon deux cônes mutuellement polaires*, C. R. Acad. Sci. Paris, t. **255**, Série I, pp. 238–240, 1962.

[7] J.-C. Rochet, P. Choné, *Ironing, Sweeping and Multidimensional screening*, Econometrica, vol. 66 (1998), pp. 783–826.