Cette dernière condition n'est que la définition de $\overline{v}(x)$ comme projection orthogonale de $-\overline{p}(x)/c$ sur le segment $[v_{min}(x), v_{max}(x)]$ (voir le Théorème 12.1.10). Finalement, on obtient (10.89) en remarquant que le support des fonctions de K est restreint à ω. □

Remarque 10.4.12 Comme dans la Remarque 10.4.8 nous expliquons comment trouver la forme de (10.86) qui définit l'état adjoint. On introduit le Lagrangien associé au problème de minimisation (10.85) sous la contrainte que l'équation d'état (10.82) (qui relie les deux variables indépendantes v et u) soit satisfaite

$$\mathcal{L}(v,u,p) = \frac{1}{2}\int_\Omega \left(|u-u_0|^2 + c|v|^2\right) dx + \int_\Omega p(\Delta u + f + v)\, dx,$$

où p est le multiplicateur de Lagrange pour la contrainte (10.82) entre v et u. Formellement, les conditions d'optimalité s'obtiennent en disant que le Lagrangien est stationnaire, c'est-à-dire que

$$\frac{\partial \mathcal{L}}{\partial v} = \frac{\partial \mathcal{L}}{\partial u} = \frac{\partial \mathcal{L}}{\partial p} = 0.$$

La première dérivée donne la condition d'optimalité (10.89), la seconde donne l'équation vérifiée par l'état adjoint p, et la troisième l'équation vérifiée par l'état u. •

10.5 Algorithmes numériques

10.5.1 Introduction

L'objet de cette section est de présenter et analyser quelques algorithmes permettant de calculer, ou plus exactement d'**approcher** la solution des problèmes d'optimisation étudiés précédemment. Tous les algorithmes étudiés ici sont effectivement utilisés en pratique pour résoudre sur ordinateur des problèmes concrets d'optimisation.

Ces algorithmes sont aussi tous de nature itérative : à partir d'une donnée initiale u^0, chaque méthode construit une suite $(u^n)_{n\in\mathbb{N}}$ dont nous montrerons qu'elle converge, sous certaines hypothèses, vers la solution u du problème d'optimisation considéré. Après avoir montré la **convergence de ces algorithmes** (c'est-à-dire, la convergence de la suite (u^n) vers u quel que soit le choix de la donnée initiale u^0), nous dirons aussi un mot de leur vitesse de convergence.

Dans toute cette section nous supposerons que la fonction objectif à minimiser J est α-convexe différentiable. Cette hypothèse d'α-convexité est assez forte, mais nous verrons plus loin qu'elle est cruciale pour les démonstrations de convergence des algorithmes. L'application des algorithmes présentés ici à la minimisation de fonctions convexes qui ne sont pas fortement convexes peut soulever quelques petites difficultés, sans parler des **grosses** difficultés qui apparaissent lorsque l'on cherche à approcher le minimum d'une fonction non convexe ! Typiquement, ces algorithmes peuvent ne pas converger et osciller entre plusieurs points de minimum, ou bien pire ils peuvent converger vers un minimum local, très loin d'un minimum global (dans le cas non convexe, cf. la Proposition 9.2.3).

Remarque 10.5.1 Nous nous limitons aux seuls algorithmes déterministes et nous ne disons rien des algorithmes de type stochastique (recuit simulé, algorithmes génétiques, etc.). Outre le fait que leur analyse fait appel à la théorie des probabilités (que nous n'abordons pas dans ce cours), leur utilisation est très différente. Pour schématiser simplement, disons que les algorithmes déterministes sont les plus efficaces pour la minimisation de fonctions convexes, tandis que les algorithmes stochastiques permettent d'approcher des minima **globaux** (et pas seulement locaux) de fonctions non convexes (à un prix toutefois assez élevé en pratique). •

10.5.2 Algorithmes de type gradient (cas sans contraintes)

Commençons par étudier la résolution pratique de problèmes d'optimisation en l'absence de contraintes. Soit J une fonction α-convexe différentiable définie sur l'espace de Hilbert réel V, on considère le problème sans contrainte

$$\inf_{v\in V} J(v) . \tag{10.92}$$

D'après le Théorème 9.2.6 il existe une unique solution u, caractérisée d'après la Remarque 10.2.2 par l'équation d'Euler

$$J'(u) = 0 .$$

Algorithme de gradient à pas optimal

L'algorithme de gradient consiste à "se déplacer" d'une itérée u^n en suivant la ligne de plus grande pente associée à la fonction coût $J(v)$. La direction de descente correspondant à cette ligne de plus grande pente issue de u^n est donnée par le gradient $J'(u^n)$. En effet, si l'on cherche u^{n+1} sous la forme

$$u^{n+1} = u^n - \mu^n w^n , \tag{10.93}$$

avec $\mu^n > 0$ petit et w^n unitaire dans V, c'est avec le choix de la direction $w_n = \dfrac{J'(u^n)}{\|J'(u^n)\|}$ que l'on peut espérer trouver la plus petite valeur de $J(u^{n+1})$ (en l'absence d'autres informations comme les dérivées supérieures ou les itérées antérieures).

Cette remarque simple nous conduit, parmi les méthodes du type (10.93) qui sont appelées "méthodes de descente", à l'algorithme de **gradient à pas optimal**, dans lequel on résout une succession de problème de minimisation à une seule variable réelle (même si V n'est pas de dimension finie). A partir de u^0 quelconque dans V, on construit la suite (u^n) définie par

$$u^{n+1} = u^n - \mu^n J'(u^n) , \tag{10.94}$$

où $\mu^n \in \mathbb{R}$ est choisi à chaque étape tel que

$$J(u^{n+1}) = \inf_{\mu\in R} J\big(u^n - \mu J'(u^n)\big) . \tag{10.95}$$

Cet algorithme converge comme l'indique le résultat suivant.

Théorème 10.5.2 *On suppose que J est α-convexe différentiable et que J' est Lipschitzien sur tout borné de V, c'est-à-dire que*

$$\forall M > 0,\ \exists C_M > 0,\ \|v\| + \|w\| \leq M \ \Rightarrow\ \|J'(v) - J'(w)\| \leq C_M \|v - w\| . \tag{10.96}$$

Alors l'algorithme de gradient à pas optimal converge : quel que soit u^0, la suite (u^n) définie par (10.94) et (10.95) converge vers la solution u de (10.92).

Démonstration. La fonction $f(\mu) = J\big(u^n - \mu J'(u^n)\big)$ est fortement convexe et dérivable sur $\mathbb{R}$ (si $J'(u^n) \neq 0$; sinon, on a déjà convergé, $u^n = u$!). Le problème de minimisation (10.95) a donc bien une solution unique, caractérisée par la condition $f'(\mu) = 0$, ce qui s'écrit aussi

$$\langle J'(u^{n+1}),\, J'(u^n)\rangle = 0 . \tag{10.97}$$

Ceci montre que deux "directions de descente" consécutives sont orthogonales.

Puisque (10.97) implique que $\langle J'(u^{n+1}), u^{n+1} - u^n\rangle = 0$, on déduit de l'$\alpha$-convexité de J que

$$J(u^n) - J(u^{n+1}) \geq \frac{\alpha}{2} \|u^n - u^{n+1}\|^2 . \tag{10.98}$$

Puisque la suite $J(u^n)$ est décroissante minorée (par $J(u)$), elle converge et l'inégalité (10.98) montre que $u^{n+1} - u^n$ tend vers 0. D'autre part, l'α-convexité de J et le fait que la suite $J(u^n)$ est bornée montrent que la suite (u^n) est bornée : il existe une constante M telle que

$$\|u^n\| \leq M .$$

Écrivant (10.96) pour $v = u^n$ et $w = u^{n+1}$ et utilisant (10.97), on voit facilement que $\|J'(u^n)\| \leq C_M \|u^{n+1} - u^n\|$, ce qui montre que $J'(u^n)$ tend vers 0. L'α-convexité de J donne alors

$$\alpha \|u^n - u\|^2 \leq \langle J'(u^n) - J'(u), u^n - u\rangle = \langle J'(u^n), u^n - u\rangle \leq \|J'(u^n)\|\ \|u^n - u\| ,$$

qui implique $\alpha \|u^n - u\| \leq \|J'(u^n)\|$, d'où l'on déduit la convergence de l'algorithme. □

Remarque 10.5.3 Il est utile de noter l'intérêt pratique de la dernière inégalité de cette démonstration : outre la preuve de la convergence, elle donne une majoration aisément calculable de l'erreur $u^n - u$. •

Algorithme de gradient à pas fixe

L'algorithme de gradient à pas fixe consiste simplement en la construction d'une suite u^n définie par

$$u^{n+1} = u^n - \mu J'(u^n) , \tag{10.99}$$

où μ est un paramètre positif fixé. Cette méthode est donc plus simple que l'algorithme de gradient à pas optimal, puisqu'on fait à chaque étape l'économie de la résolution de (10.95). Le résultat suivant montre sous quelles hypothèses on peut choisir le paramètre μ pour assurer la convergence.

Théorème 10.5.4 *On suppose que J est α-convexe différentiable et que J' est Lipschitzien sur V, c'est-à-dire qu'il existe une constante $C > 0$ telle que*

$$\|J'(v) - J'(w)\| \leq C\|v - w\| \quad \forall\, v, w \in V\ . \tag{10.100}$$

Alors, si $0 < \mu < 2\alpha/C^2$, l'algorithme de gradient à pas fixe converge : quel que soit u^0, la suite (u^n) définie par (10.95) converge vers la solution u de (10.92).

Démonstration. Posons $v^n = u^n - u$. Comme $J'(u) = 0$, on a $v^{n+1} = v^n - \mu\big(J'(u^n) - J'(u)\big)$, d'où il vient

$$\begin{aligned} \|v^{n+1}\|^2 &= \|v^n\|^2 - 2\mu\langle J'(u^n) - J'(u), u^n - u\rangle + \mu^2\big\|J'(u^n) - J'(u)\big\|^2 \\ &\leq \left(1 - 2\alpha\mu + C^2\mu^2\right)\ \|v^n\|^2\ , \end{aligned} \tag{10.101}$$

d'après (10.100) et l'α-convexité. Si $0 < \mu < 2\alpha/C^2$, il est facile de voir que $1 - 2\alpha\mu + C^2\mu^2 \in]0, 1[$, et la convergence se déduit de (10.101). □

Remarque 10.5.5 Une adaptation simple de la démonstration précédente, laissée au lecteur en guise d'exercice, permet de montrer la convergence en remplaçant (10.100) par l'hypothèse plus faible (10.96). Il faut noter aussi que, pour l'algorithme de gradient à pas fixe, à la différence du gradient à pas optimal, la suite $J(u^n)$ n'est pas nécessairement monotone. •

Remarque 10.5.6 Il existe de nombreux autres algorithmes de descente du type (10.93) que nous ne décrirons pas ici. On rencontre notamment dans cette classe d'algorithmes la méthode du gradient conjugué dans laquelle la direction de descente w^n dépend non seulement du gradient $J'(u^n)$ mais aussi des directions de descente utilisées aux itérations précédentes. Nous présentons cette méthode dans la Sous-section 13.1.5, pour le cas particulier d'une fonctionnelle quadratique du type $\frac{1}{2}Ax \cdot x - b \cdot x$. •

Remarque 10.5.7 Comment choisir entre les deux algorithmes de gradient que nous venons de voir, et plus généralement entre les différentes méthodes de minimisation numérique qui existent ? Un premier critère de choix concerne le coût de chaque itération. Par exemple, nous l'avons dit, chaque itération de l'algorithme de gradient à pas fixe est moins chère qu'une itération de gradient à pas optimal. Évidemment, si l'on part de la même itérée u^n, une itération du gradient à pas optimal décroît plus la fonction coût qu'une itération du gradient à pas fixe. On en arrive donc au deuxième critère de choix, souvent plus déterminant, qui est celui de la **vitesse de convergence** de l'algorithme, qui fixe le nombre d'itérations nécessaires pour rendre l'erreur $\|u^n - u\|$ inférieure à une tolérance ϵ fixée a priori.

Par exemple, l'inégalité (10.101) montre que la convergence de l'algorithme de gradient à pas fixe est au moins géométrique, puisque

$$\|u^n - u\| \leq \gamma^n\|u^0 - u\| \quad \text{avec} \quad \gamma = \sqrt{1 - 2\alpha\mu + \mu^2 C^2}.$$

Cette remarque conduit d'ailleurs, sous réserve d'une analyse plus poussée, à préférer pour le paramètre μ la valeur médiane α/C^2 dans l'intervalle $]0, 2\alpha/C^2[$, valeur qui minimise celle de γ. En fait, on peut montrer que la convergence des deux algorithmes étudiés ci-dessus est effectivement géométrique dans certains cas particulier (ce qui signifie que la quantité $\|u^n - u\|^{1/n}$ a une limite finie, comprise strictement entre 0 et 1, lorsque n tend vers $+\infty$). •

Exercice 10.5.1 Pour $V = \mathbb{R}^2$ et $J(x, y) = ax^2 + by^2$ avec $a, b > 0$, montrer que l'algorithme de gradient à pas optimal converge en une seule itération si $a = b$ ou si $x^0 y^0 = 0$, et que la convergence est géométrique dans les autres cas. Étudier aussi la convergence de l'algorithme de gradient à pas fixe : pour quelles valeurs du paramètre μ la convergence se produit-elle, pour quelle valeur est-elle la plus rapide ?

10.5.3 Algorithmes de type gradient (cas avec contraintes)

On étudie maintenant la résolution de problèmes d'optimisation avec contraintes

$$\inf_{v \in K} J(v) , \tag{10.102}$$

où J est une fonction α-convexe différentiable définie sur K, sous-ensemble convexe fermé non vide de l'espace de Hilbert réel V. Le Théorème 9.2.6 assure alors l'existence et l'unicité de la solution u de (10.102), caractérisée d'après le Théorème 10.2.1 par la condition

$$\langle J'(u), v - u \rangle \geq 0 \quad \forall v \in K . \tag{10.103}$$

Selon les algorithmes étudiés ci-dessous, nous serons parfois amenés à préciser des hypothèses supplémentaires sur l'ensemble K.

Algorithme de gradient à pas fixe avec projection

L'algorithme de gradient à pas fixe s'adapte au cas du problème (10.102) avec contraintes à partir de la remarque suivante. Pour tout réel $\mu > 0$, (10.103) s'écrit

$$\langle u - \big(u - \mu J'(u)\big), v - u \rangle \geq 0 \quad \forall v \in K . \tag{10.104}$$

Notons P_K l'opérateur de projection sur l'ensemble convexe K, défini au Théorème 12.1.10 de projection sur un convexe (voir la Remarque 12.1.11). Alors, d'après ce théorème, (10.104) n'est rien d'autre que la caractérisation de u comme la projection orthogonale de $u - \mu J'(u)$ sur K. Autrement dit,

$$u = P_K\big(u - \mu J'(u)\big) \quad \forall \mu > 0 . \tag{10.105}$$

Il est facile de voir que (10.105) est en fait équivalent à (10.103), et caractérise donc la solution u de (10.102). L'algorithme de **gradient à pas fixe avec projection** (ou plus simplement de gradient projeté) est alors défini par l'itération

$$u^{n+1} = P_K\big(u^n - \mu J'(u^n)\big) , \tag{10.106}$$

où μ est un paramètre positif fixé.