

MODULE B2

Méthodes de gradient explicite

Sauf mention contraire, \mathcal{X} (resp. \mathcal{Y}) est un espace de HILBERT, muni du produit scalaire noté $\langle \cdot, \cdot \rangle$ et on note $\| \cdot \|$ la norme qui découle du produit scalaire.

Dans ce module, on s'intéresse au problème de minimisation suivant

$$\min_{x \in \mathcal{X}} J(x) \quad (\mathcal{P})$$

où $J : \mathcal{X} \rightarrow \mathbb{R}$ est une fonction différentiable (sur tout \mathcal{X}). On va introduire un algorithme d'optimisation pour résoudre ce problème et étudier le comportement (et en particulier la convergence) des suites générées par cet algorithme.

1 Direction de descente

1.1 Définition et exemple

Définition 1 (Direction de descente)

Soit $J : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ une fonction. Soit $x^0 \in \text{dom } J$ et $d \in \mathcal{X}$. On dit que d est une *direction de descente* pour J au point x^0 si $J'(x^0; d)$ existe et que

$$J'(x^0; d) < 0$$

où $J'(x^0; d)$ désigne la dérivée directionnelle de J en x^0 dans la direction d .

Autrement dit, si d est une direction de descente pour J en x^0 , alors il existe $\alpha > 0$ tel que

$$\forall t \in]0; \alpha], \quad J(x^0 + t d) < J(x^0)$$

(on notera que la réciproque est fausse). On peut alors énoncer le résultat suivant :

Proposition 1

Soit $J : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ une fonction. Soit $x^0 \in \text{dom } J$. Si x^0 est un minimiseur de J , alors J n'admet aucune direction de descente en x^0 . De manière équivalente, si J admet une direction de descente en x^0 , alors x^0 n'est pas un minimiseur de J .

DÉMONSTRATION : Démontrons cette affirmation par l'absurde. On suppose donc que x^0 est un minimiseur de J et que J admet d comme direction de descente en x^0 . Par définition de d , il existe $\alpha > 0$ tel que

$$\forall t \in]0; \alpha], \quad J(x^0 + t d) < J(x^0)$$

Cela implique notamment que le point $x = x^0 + \alpha d/2$ vérifie $J(x) < J(x^0)$, ce qui contredit le fait que x^0 soit un **minimiseur** de J . ■

Lorsque J est différentiable en x^0 , la proposition suivante permet de caractériser de manière plus pratique les directions de descente :

Proposition 2

Soit $J : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ une fonction. Soit $x^0 \in \text{dom } J$. On suppose que J est différentiable au voisinage de x^0 . Alors d est une direction de descente pour J en x^0 si et seulement si

$$\langle \nabla J(x^0), d \rangle < 0$$

DÉMONSTRATION : Ce résultat découle de l'identité $J'(x^0; d) = \langle \nabla J(x^0), d \rangle$ valable lorsque J est différentiable. ■

Il est à noter que les directions de descente, lorsqu'elles existent, ne sont pas uniques en un même point. L'exemple suivant est à la base de la méthode de NEWTON.

EXEMPLE

Méthode de NEWTON. On suppose que J est deux fois différentiable et que la matrice hessienne $\text{Hess}J(x^0)$ de J en $x^0 \in \mathcal{X}$ est définie positive. Posons

$$d = -(\text{Hess}J(x^0))^{-1}(\nabla J(x^0))$$

Alors d est une direction de descente pour J en x^0 . En effet, $\text{Hess}J(x)$ est diagonalisable et définie positive pour tout $x \in \mathcal{X}$. Notons λ_{\max} la plus grande valeur propre (positive) de $\text{Hess}J(x^0)$. On a alors

$$\langle \nabla J(x^0), -(\text{Hess}J(x^0))^{-1} \nabla J(x^0) \rangle = -\frac{\|\nabla J(x^0)\|^2}{\lambda_{\max}} < 0$$

ce qui assure que d est bien une direction de descente pour J d'après la proposition 2. Ce vecteur apparaît lorsque l'on applique la méthode de NEWTON pour trouver les zéros du gradient ∇J . Les itérations de cette méthode pour rechercher les zéros d'une fonction $G : \mathcal{X} \rightarrow \mathcal{Y}$ s'écrivent

$$\forall n \in \mathbb{N}, \quad x_{k+1} = x_k - (\nabla G(x_k))^{-1}(G(x_k))$$

ce qui donne, lorsque $G = \nabla J$,

$$x_{k+1} = x_k - (\text{Hess}J(x_k))^{-1}(\nabla J(x_k))$$

Pour plus de détails sur la méthode de NEWTON, le lecteur est invité à consulter le module **B3 : Méthode de NEWTON. Méthode du gradient** (3MA261).

1.2 Méthodes de descente

Les méthodes dites de descente génèrent une suite d'itérées $(x_k)_{k \in \mathbb{N}}$ satisfaisant la propriété de descente suivante :

$$\forall k \in \mathbb{N}, \quad J(x_{k+1}) \leq J(x_k)$$

le point x_{k+1} pouvant toujours s'exprimer de manière arbitraire à l'aide du point précédent x_k en posant

$$x_{k+1} = x_k + \tau_k d_k \quad \text{avec } \tau_k > 0 \text{ et } d_k \in \mathcal{X}$$

(pour tout couple de points successifs (x_k, x_{k+1}) , cette décomposition n'est pas unique). Ainsi, d'après ce qui précède, le vecteur d_k peut être choisi parmi les direction de descente de J en x_k , auquel cas l'existence d'un pas de temps τ_k assurant la décroissance de J est assurée tant que x_k n'est pas un minimiseur. Il est à noter que d_k n'est pas nécessairement une direction de descente au sens de la définition 1. La question réside donc dans le choix de ce vecteur et des valeurs admissibles pour les pas de temps τ_k .

2 Méthode du gradient explicite

Il s'agit d'une classe de méthodes très étudiée et bien connue, généralement l'une des premières qui sont abordées en optimisation numérique car elle est facile à mettre en œuvre dès lors que le problème d'optimisation étudié possède les bonnes propriétés.

2.1 Description de l'algorithme

Les méthodes du gradient explicite sont des méthodes de descentes (on parle parfois de *méthode de descente du gradient*, l'adjectif *explicite* étant souvent omis), dans lesquelles le vecteur d_k introduit à la fin de la section précédente est une direction de descente, définie à l'aide du gradient de la fonction J au point courant.

Plus précisément, ce vecteur est l'opposé du gradient de J au point courant x_k , car selon la proposition suivante, ce vecteur constitue une direction de descente pour la fonction J au point x_k :

Proposition 3

Soit $J : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ une fonction. Soit $x \in \text{dom } J$. On suppose que J est différentiable au voisinage de x . On suppose par ailleurs que x n'est pas un point critique de J . Alors $d = -\nabla J(x)$ est une direction de descente pour J en x .

DÉMONSTRATION : Il s'agit d'une conséquence directe de la proposition 2, puisque

$$\langle \nabla J(x), d \rangle = -\langle \nabla J(x), \nabla J(x) \rangle = -\|\nabla J(x)\|^2 \quad \blacksquare$$

Les itérations de la méthode du gradient explicite sont alors données par

$$x_0 \in \mathcal{X} \quad \text{et} \quad \forall k \in \mathbb{N}, \quad x_{k+1} = x_k - \tau_k \nabla J(x_k) \quad \text{avec } \tau_k > 0.$$

L'hypothèse de différentiabilité sur J assure que, quelle que soit l'initialisation x_0 et le choix des pas de temps τ_k , la suite des itérées est bien définie.

Ce qui distingue les différentes versions de la méthode du gradient explicite entre elles, c'est le choix du pas de temps τ_k , qui est détaillé dans les paragraphes suivants, de même que le choix du critère d'arrêt.

2.2 Choix du pas de temps

Une fois le choix de la direction de descente arrêté, celui du pas de temps est crucial. Il doit répondre à plusieurs exigences :

- **descente** : pour rester dans le cadre des méthodes de descente, le pas de temps doit être choisi de sorte que l'algorithme génère une suite d'itérées dont la valeur par J décroît (cette contrainte pourra être levée dans le cas général) ;

- **convergence** : le choix du pas de temps doit permettre d'assurer la convergence du schéma itératif, autrement dit qu'en un nombre fini d'itérations, la suite générée approche raisonnablement bien une solution du problème (\mathcal{P}) ;
- **vitesse de convergence** : comme on le verra par la suite, le choix du pas de temps peut grandement influencer la vitesse de convergence de l'algorithme, c'est-à-dire le nombre d'itérations nécessaire avant d'activer le critère d'arrêt ; il faut donc veiller à en sélectionner un qui permet de converger en un nombre raisonnable d'itérations ;
- **complexité** : outre le nombre d'itérations, c'est le temps de calcul pour atteindre la solution (à une précision souhaitée) qui importe ; or, celui-ci est fonction du nombre d'itérations que l'algorithme doit calculer avant d'activer le critère d'arrêt, mais également du temps de calcul d'une itération isolée. Pour la méthode du gradient explicite, celui-ci repose essentiellement sur deux calculs : l'évaluation du gradient au point courant, et le choix du pas de temps. Il apparaît donc essentiellement que le choix de ce dernier ne mène pas à des calculs trop complexes qui serait trop gourmands en temps (ou en mémoire).

De manière générale (mais il existe de nombreuses exceptions), on peut dire que la convergence du schéma nécessite des pas de temps relativement faibles, mais qu'obtenir une convergence en un nombre d'itérations petit demande de choisir des pas de temps relativement grands. La première difficulté réside dans la conciliation de ces deux exigences contradictoires. Il existe alors deux possibilités : soit on choisit un pas de temps fixe, qui a le mérite de satisfaire le critère sur la complexité (puisque'il n'est au plus calculé qu'une seule fois), mais avec l'inconvénient d'être très rigide, soit on choisit un pas de temps qui s'adapte localement au point courant (permettant parfois d'économiser un nombre intéressant d'itérations), mais qui peut demander un grand nombre de calculs.

Voici quelques exemples de règles de choix de pas de temps classiques.

- **Pas optimal.** Cette règle est la plus naturelle : à chaque itération, on cherche le point suivant en sélectionnant celui qui, dans la direction de descente, minimise la fonction objectif :

$$x_{k+1} = \underset{\substack{x = x_k - \tau \nabla J(x_k) \\ \tau > 0}}{\operatorname{argmin}} J(x)$$

Le pas de temps associé est alors défini de la manière suivante :

$$\tau_k = \underset{\tau > 0}{\operatorname{argmin}} J(x_k - \tau \nabla J(x_k))$$

La recherche de ce pas optimal est appelée *recherche linéaire* (*line search* en anglais). Il existe des contre-exemples connus démontrant que cette stratégie n'est pas toujours optimale, ou ne permet pas de faire converger l'algorithme ; par ailleurs, elle est en général coûteuse en calcul, puisqu'il s'agit de résoudre un problème d'optimisation. Supposons que x_k n'est pas un point critique. Puisque la fonction

$$\tau \mapsto J(x_k - \tau \nabla J(x_k))$$

est différentiable et que $-\nabla J(x_k)$ est une direction de descente, alors (on suppose que la fonction J admet un minimiseur), la fonction précédente admet au moins un minimiseur local sur $]0; +\infty[$, parmi les points τ^* vérifiant

$$\langle \nabla J(x_k), \nabla J(x_k - \tau^* \nabla J(x_k)) \rangle = 0$$

de sorte que le nouveau point x_{k+1} vérifie $\langle \nabla J(x_k), \nabla J(x_{k+1}) \rangle = 0$. On peut interpréter cette dernière égalité de manière géométrique : les gradients en deux points consécutifs sont orthogonaux (cf. figure 1).

- **Pas constant.** Enfin, le cas des pas de temps fixes

$$\forall k \in \mathbb{N}, \quad \tau_k = \tau$$

où $\tau > 0$, est le plus simple, puisqu'il suffit de choisir une seule fois la valeur du pas. L'inconvénient le plus visible de cette stratégie est le fait qu'un pas fixe ne tenant *a priori* pas compte de l'information locale (c'est-à-dire au point courant x_k), il peut être trop restrictif et conduire à des convergences plus lentes. Cependant, on verra dans la section suivante que, sous certaines hypothèses sur le problème, il est possible d'ajuster correctement la valeur τ pour aboutir à une convergence de l'algorithme.

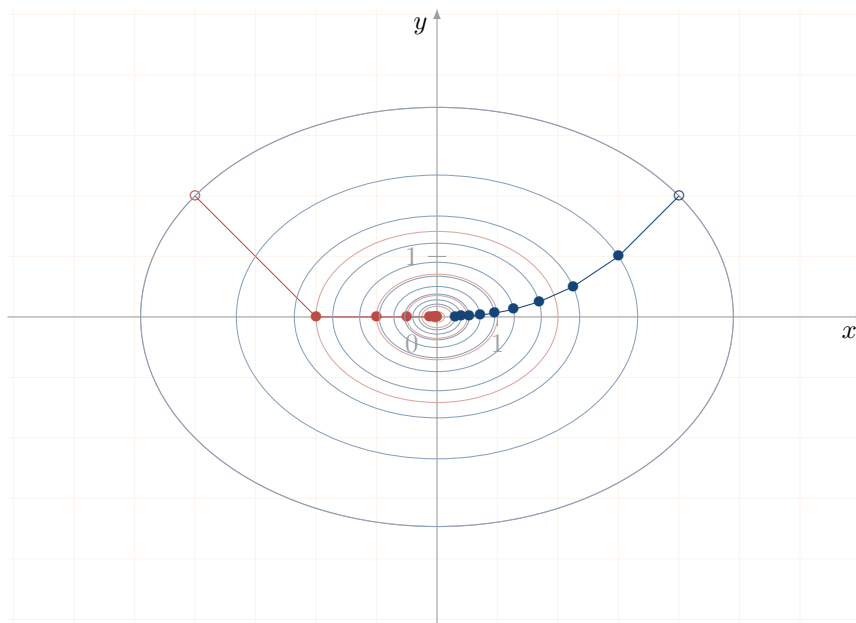


FIGURE 1 – Illustration de la méthode du gradient à pas optimal dans le cas d'une fonction convexe quadratique pour deux choix de pas de temps et d'initialisation différents. Les ellipses correspondent à des lignes de niveau de la fonction objectif.

On pourra noter que toutes ces considérations sur le choix des pas de temps, ainsi que les différentes stratégies proposées pour résoudre un tel problème, ne sont pas spécifiques aux méthodes du gradient explicite, et doivent être soulevées pour d'autres méthodes d'optimisation.

CONTRE-EXEMPLE

Oscillations dans le cas des pas constants. Considérons le problème de minimisation convexe suivant :

$$\min_{x \in \mathbb{R}} \frac{1}{2} x^2$$

Les itérations de la méthode du gradient à pas constant s'écrivent

$$\forall k \in \mathbb{N}, \quad x_{k+1} = x_k - \tau x_k = (1 - \tau) x_k$$

Si $\tau = 2$, alors on constate que $x_{k+1} = -x_k$, de sorte que la suite des itérées oscille entre deux points x_0 et $-x_0$, qui sont différents si $x_0 \neq 0$. Notons que, dans cet exemple, la suite des valeurs $J(x_k)$ reste constante. Il n'y a donc pas de décroissance stricte. Le prochain exemple montre que rajouter cette seconde condition n'est pas non plus suffisant pour assurer la convergence vers un point critique de la fonction objectif.

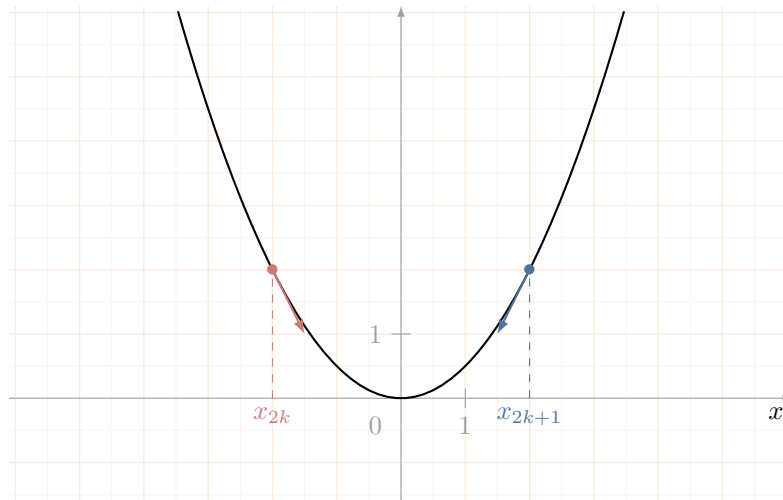


FIGURE 2 – Oscillations dans le cas des pas constants.

CONTRE-EXEMPLE

Oscillations dans le cas des pas optimaux. Considérons le problème de minimisation convexe quadratique suivant :

$$\min_{(x,y) \in \mathbb{R}^2} \frac{1}{2} (x^2 + \lambda y^2)$$

avec $\lambda > 0$. Les itérations de la méthode du gradient à pas optimal s'écrivent

$$\forall k \in \mathbb{N}, \quad \begin{pmatrix} x_{k+1} \\ y_{k+1} \end{pmatrix} = \begin{pmatrix} x_k \\ y_k \end{pmatrix} - \tau_k \begin{pmatrix} x_k \\ \lambda y_k \end{pmatrix} = \begin{pmatrix} (1 - \tau_k) x_k \\ (1 - \lambda \tau_k) y_k \end{pmatrix}$$

où $\tau_k > 0$ est choisi de sorte de minimiser la quantité

$$\frac{1}{2} ((1 - \tau_k)^2 x_k^2 + \lambda (1 - \lambda \tau_k)^2 y_k^2) = \frac{1}{2} (x_k^2 + \lambda y_k^2 - 2(x_k^2 + \lambda^2 y_k^2) \tau_k + (x_k^2 + \lambda^3 y_k^2) \tau_k^2)$$

Il s'agit d'un polynôme du second degré en τ_k ; on vérifie donc aisément que

$$\tau_k = \frac{x_k^2 + \lambda^2 y_k^2}{x_k^2 + \lambda^3 y_k^2} \quad \text{donc} \quad \frac{1}{2} (x_{k+1}^2 + \lambda y_{k+1}^2) = \frac{1}{2} \frac{(\lambda - 1)^2 \lambda x_k^2 y_k^2}{(x_k^2 + \lambda^3 y_k^2)}$$

Notons que, lorsque λ est grand, les croissances comparées assurent que τ_k est proche de zéro. Dans la figure 3, on montre deux exemples de suites d'itérées dans le cas $\lambda = 100$, générées à partir de deux points initiaux différents. On voit que, dans l'un de deux cas, il faut un grand nombre d'itérations pour approcher

raisonnablement le minimiseur $(0,0)$. Cela est dû au fait que les lignes de niveau de la fonction objectif (ici, des ellipses) sont très allongées. Formellement, cette propriété se traduit par un grand (donc mauvais) conditionnement (il vaut ici $\kappa = \sqrt{\lambda}$).

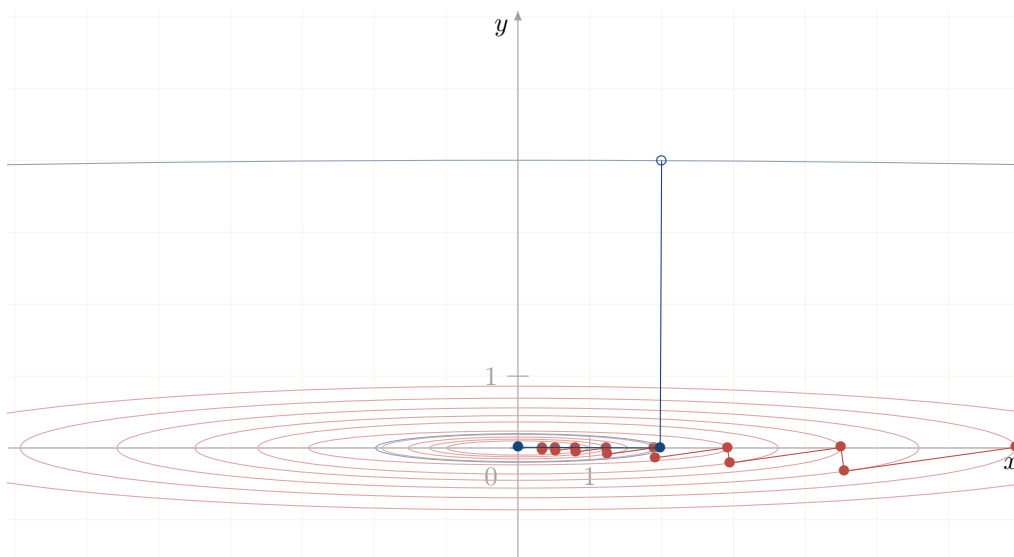


FIGURE 3 – Oscillations dans le cas des pas optimaux. Les ellipses sont des lignes de niveau de la fonction objectif. En **bleu** et en **rouge**, deux trajectoires correspondant à deux initialisations différentes.

3 Convergence de l'algorithme dans le cas régulier

On considère dans cette section l'algorithme du gradient explicite à pas constant pour la minimisation d'une fonction **régulière**.

3.1 Convergence du critère et des gradients

Dans ce paragraphe, on commence par établir quelques résultats valables lorsque le problème est régulier et que le pas de temps τ est correctement choisi en fonction de la constante de LIPSCHITZ du gradient de la fonction objectif.

Proposition 4 (Convergence du critère)

Soit $J : \mathcal{X} \rightarrow \mathbb{R}$ une fonction **L -régulière**. On suppose que J est **minorée**. Soit $(x_k)_{k \in \mathbb{N}}$ une suite générée par l'algorithme suivant

$$x_0 \in \mathcal{X} \quad \text{et} \quad \forall k \in \mathbb{N}, \quad x_{k+1} = x_k - \tau \nabla J(x_k)$$

Si $\tau \in]0; 2/L[$, alors la suite $(J(x_k))_{k \in \mathbb{N}}$ est décroissante et converge vers une valeur J^* . De plus, on a

$$J(x_{k+1}) \leq J(x_k) - \left(\tau - \frac{\tau^2}{2} L \right) \|\nabla J(x_k)\|^2$$

DÉMONSTRATION : Soit $k \in \mathbb{N}$. Commençons par écrire que

$$J(x_{k+1}) - J(x_k) = J(x_k - \tau \nabla J(x_k)) - J(x_k) = \left[J(x_k - t \nabla J(x_k)) \right]_0^\tau$$

En intégrant, on obtient donc que

$$\left[J(x_k - t \nabla J(x_k)) \right]_0^\tau = - \int_0^\tau \langle \nabla J(x_k), \nabla J(x_k - t \nabla J(x_k)) \rangle dt$$

En ajoutant le vecteur nul $\nabla J(x_k) - \nabla J(x_k)$ dans le second membre du produit scalaire, on obtient que (attention au signe du terme intégral)

$$J(x_{k+1}) - J(x_k) = -\tau \|\nabla J(x_k)\|^2 + \int_0^\tau \langle \nabla J(x_k), \nabla J(x_k) - \nabla J(x_k - t \nabla J(x_k)) \rangle dt$$

Majorons le produit scalaire à l'aide de l'inégalité de CAUCHY-SCHWARZ, puis de l'hypothèse de **régularité** sur J ; il s'ensuit que

$$J(x_{k+1}) - J(x_k) \leq -\tau \|\nabla J(x_k)\|^2 + \int_0^\tau t L \|\nabla J(x_k)\|^2 dt = -\left(\tau - \frac{\tau^2}{2} L \right) \|\nabla J(x_k)\|^2$$

Par hypothèse sur le pas de temps τ , on a $\tau - \tau^2 L/2 > 0$. On prouve ainsi que la suite des $J(x_k)$ est décroissante. La convergence suit car cette suite est minorée par $\inf J$ par hypothèse. ■

Proposition 5 (Convergence du critère d'optimalité)

Soit $J : \mathcal{X} \rightarrow \mathbb{R}$ une fonction **L -régulière**. On suppose que J est **minorée**. Soit $(x_k)_{k \in \mathbb{N}}$ une suite générée par l'algorithme suivant

$$x_0 \in \mathcal{X} \quad \text{et} \quad \forall k \in \mathbb{N}, \quad x_{k+1} = x_k - \tau \nabla J(x_k)$$

Si $\tau \in]0; 2/L[$, alors $\lim_{k \rightarrow +\infty} \|\nabla J(x_k)\| = 0$

DÉMONSTRATION : Dans la proposition 4, on a établi que

$$J(x_{k+1}) \leq J(x_k) - \left(\tau - \frac{\tau^2}{2} L \right) \|\nabla J(x_k)\|^2 \leq J(x_k)$$

Par encadrement, la suite des

$$J(x_k) - \left(\tau - \frac{\tau^2}{2} L \right) \|\nabla J(x_k)\|^2$$

converge vers J^* . Il s'ensuit que la suite $(\|\nabla J(x_k)\|)_{k \in \mathbb{N}}$ converge vers 0. ■

3.2 Convergence vers un minimum dans le cas convexe

Pour obtenir des résultats plus intéressants, il faut ajouter des hypothèses supplémentaires. Parmi les options possibles, on peut supposer la convexité du problème.

Lemme 1

Soit $J : \mathcal{X} \rightarrow \mathbb{R}$ une fonction **convexe L -régulière**. On suppose que J admet un minimiseur x^* . Soit $(x_k)_{k \in \mathbb{N}}$ une suite générée par l'algorithme suivant

$$x_0 \in \mathcal{X} \quad \text{et} \quad \forall k \in \mathbb{N}, \quad x_{k+1} = x_k - \tau \nabla J(x_k)$$

Si $\tau \in]0; 2/L[$, alors la suite $(\|x_k - x^*\|)_{k \in \mathbb{N}}$ est décroissante.

REMARQUE : Il s'ensuit immédiatement que cette suite converge, puisqu'elle est réelle minorée par 0. Cependant, tant que l'on n'a pas établi que la limite de cette suite est nulle, on ne peut pas conclure à la convergence des itérées.

DÉMONSTRATION : La règle de FERMAT ($\nabla J(x^*) = 0$) entraîne que

$$x^* = x^* - \tau \nabla J(x^*)$$

(Autrement dit, x^* est un point fixe de la méthode du gradient explicite). Aussi, pour tout $k \in \mathbb{N}$, on a

$$\begin{aligned} \|x_{k+1} - x^*\|^2 &= \|x_k - \tau \nabla J(x_k) - x^* + \tau \nabla J(x^*)\|^2 \\ &= \|x_k - x^*\|^2 - 2\tau \langle \nabla J(x_k) - \nabla J(x^*), x_k - x^* \rangle \\ &\quad + \tau^2 \|\nabla J(x_k) - \nabla J(x^*)\|^2 \end{aligned}$$

Il suffit alors d'utiliser le lemme de BAILLON-HADDAD (module **A4 : Fonctions régulières**) et la régularité de J pour obtenir la majoration

$$\begin{aligned} \|x_{k+1} - x^*\|^2 &\leq \|x_k - x^*\|^2 - 2\frac{\tau}{L} \|\nabla J(x_k) - \nabla J(x^*)\|^2 + \tau^2 \|\nabla J(x_k) - \nabla J(x^*)\|^2 \\ &\leq \|x_k - x^*\|^2 - \frac{2}{L} \left(\tau - \frac{\tau^2}{2} L \right) \|\nabla J(x_k) - \nabla J(x^*)\|^2 \end{aligned}$$

Puisque $\tau - \tau^2 L/2 > 0$, on en déduit que $\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2$. ■

Une conséquence importante de ce lemme est le fait que, si le problème est convexe et régulier, qu'il admet une solution, et que le pas de temps est correctement choisi, alors la suite générée par la méthode du gradient explicite est bornée. On verra dans le paragraphe suivant que cette propriété est déterminante pour démontrer la convergence de cet algorithme.

Démontrons en effet que, dans le cas **convexe**, la suite des $J(x_k)$ converge vers le minimum de J lorsque le pas de temps τ est correctement choisi.

Proposition 6 (Convergence du critère vers le minimum)

Soit $J : \mathcal{X} \rightarrow \mathbb{R}$ une fonction **convexe** L -régulière. On suppose que J admet un minimiseur x^* . Soit $(x_k)_{k \in \mathbb{N}}$ une suite générée par l'algorithme suivant

$$x_0 \in \mathcal{X} \quad \text{et} \quad \forall k \in \mathbb{N}, \quad x_{k+1} = x_k - \tau \nabla J(x_k)$$

Si $\tau \in]0; 2/L[$, alors la suite $(J(x_k))_{k \in \mathbb{N}}$ converge vers $J(x^*)$ et

$$\forall k \in \mathbb{N}, \quad 0 \leq J(x_k) - J(x^*) \leq \frac{\|x_0 - x^*\|^2}{\tau - \tau^2 L/2} \frac{1}{k+1}$$

REMARQUE : Par définition de x^* , on a toujours la positivité de la quantité $J(x_k) - J(x^*)$.

DÉMONSTRATION : Soit $k \in \mathbb{N}$. La convexité de J assure que

$$J(x^*) \geq J(x_k) + \langle \nabla J(x_k), x^* - x_k \rangle$$

L'inégalité de CAUCHY-SCHWARZ permet d'obtenir la majoration suivante :

$$J(x_k) - J(x^*) \leq \|\nabla J(x_k)\| \|x_k - x^*\| \leq \|\nabla J(x_k)\| \|x_0 - x^*\|$$

où la deuxième inégalité est obtenue grâce au lemme 2. En utilisant la proposition 4, on peut alors écrire que

$$J(x_{k+1}) - J(x^*) \leq J(x_k) - J(x^*) - \left(\tau - \frac{\tau^2}{2} L \right) \frac{(J(x_k) - J(x^*))^2}{\|x_0 - x^*\|^2}$$

On conclut la preuve par un raisonnement par récurrence. Commençons par l'initialisation ($k = 0$). Pour cela, il suffit de remarquer que l'inégalité précédente s'écrit

$$\frac{J(x_{k+1}) - J(x^*)}{\|x_0 - x^*\|^2} \leq \frac{J(x_k) - J(x^*)}{\|x_0 - x^*\|^2} \left(1 - (\tau - \tau^2 L/2) \frac{J(x_k) - J(x^*)}{\|x_0 - x^*\|^2} \right)$$

Le membre de droite étant positif, on en déduit que, pour tout $k \in \mathbb{N}$,

$$0 \leq J(x_k) - J(x^*) \leq \frac{\|x_0 - x^*\|^2}{\tau - \tau^2 L/2}$$

Supposons la majoration de la proposition 6 vraie au rang k . En multipliant la troisième inégalité de cette preuve par la quantité adéquate, on obtient alors

$$(k+2) \frac{\tau - \tau^2 L/2}{\|x_0 - x^*\|^2} (J(x_{k+1}) - J(x^*)) \leq 1 + \frac{\tau - \tau^2 L/2}{\|x_0 - x^*\|^2} (J(x_k) - J(x^*)) - (k+2) \frac{(\tau - \tau^2 L/2)^2}{\|x_0 - x^*\|^2} \frac{(J(x_k) - J(x^*))^2}{\|x_0 - x^*\|^2}$$

où le premier terme du membre de droite est obtenu à l'aide de l'hypothèse de récurrence. Intéressons-nous aux deux termes restants. En utilisant la décroissance de la suite $(J(x_k) - J(x^*))_{k \in \mathbb{N}}$, on peut majorer leur somme par

$$\frac{\tau - \tau^2 L/2}{\|x_0 - x^*\|^2} (J(x_k) - J(x^*)) \left(1 - (k+2) (\tau - \tau^2 L/2) \frac{J(x_{k+1}) - J(x^*)}{\|x_0 - x^*\|^2} \right)$$

On en déduit que l'hypothèse de récurrence est vérifiée au rang $k+1$, c'est-à-dire que la quantité ci-dessus est positive. Sinon, elle serait négative et l'inégalité précédente serait majorée par 1, ce qui serait l'hypothèse de récurrence au rang $k+1$. On en déduit que par récurrence le résultat souhaité. ■

La proposition 6 donne, dans le cas convexe, une vitesse de convergence / décroissance pour le critère, qui est en $O(1/k)$, avec k le nombre d'itérations. Ce taux de convergence dépend de l'initialisation (en particulier, de la distance du premier point à l'ensemble des minimiseurs de J), mais également du choix du pas de temps. Il apparaît que le taux de convergence est d'autant plus intéressant que la quantité

$$\frac{\|x_0 - x^*\|^2}{\tau - \tau^2 L/2}$$

est faible, donc que la quantité $\tau - \tau^2 L/2$ est grande. Si $\tau \in]0; 2/L[$, alors la valeur optimale τ^* du pas de temps selon ce critère est

$$\tau^* = \frac{1}{L}$$

auquel cas $\forall k \in \mathbb{N}, \quad 0 \leq J(x_k) - J(x^*) \leq \frac{2L \|x_0 - x^*\|^2}{k+1}$

Par ailleurs, il est facile de voir que le taux établi dans la proposition 6 est croissant en fonction de la constante de LIPSCHITZ L . Ainsi, plus L est petit, meilleure est la convergence de la méthode du gradient explicite à pas constant.

Une manière de traduire la proposition 6 est la suivante : la suite des itérées $(x_k)_{k \in \mathbb{N}}$ est une suite minimisante de J . Ainsi, si J est fortement convexe et qu'on est en dimension finie, alors on en déduit immédiatement que cette suite converge vers l'unique minimiseur de J (d'après la proposition 15 du module **B1 : Méthodes d'optimisation du premier ordre**), avec (si $\tau = \tau^* = 1/L$)

$$\forall k \in \mathbb{N}, \quad 0 \leq \|x_k - x^*\|^2 \leq \frac{8L \|x_0 - x^*\|^2}{\alpha(k+1)} = \frac{8\kappa \|x_0 - x^*\|^2}{k+1}$$

où κ est le conditionnement de la fonction J . On peut donc déjà observer que ce taux est d'autant plus intéressant que le conditionnement est bon (κ proche de 0). On verra plus loin qu'il est en réalité possible d'obtenir un taux de convergence beaucoup plus intéressant.

On termine ce paragraphe avec le résultat central sur la convergence de la méthode du gradient explicite dans le cas convexe.

Proposition 7 (Convergence des itérés dans le cas convexe)

Soit $J : \mathcal{X} \rightarrow \mathbb{R}$ une fonction **convexe** L -régulière. On suppose que J admet un minimiseur. Soit $(x_k)_{k \in \mathbb{N}}$ une suite générée par l'algorithme suivant

$$x_0 \in \mathcal{X} \quad \text{et} \quad \forall k \in \mathbb{N}, \quad x_{k+1} = x_k - \tau \nabla J(x_k)$$

Si $\tau \in]0; 2/L[$, alors la suite $(x_k)_{k \in \mathbb{N}}$ converge vers un minimiseur de J .

DÉMONSTRATION : Décomposons cette preuve en deux étapes.

- **La suite des itérées est bornée.** Il s'agit d'une conséquence du lemme 2.
- **La suite des itérées converge.** La proposition 8 (qui se démontre de manière indépendante) assure l'existence d'une sous-suite $(x_{k_j})_{j \in \mathbb{N}}$ qui converge vers un point critique (donc un minimiseur) x^* de J . D'après le lemme 2, il s'ensuit que, par unicité de la limite,

$$\lim_{k \rightarrow +\infty} \|x_k - x^*\| = \lim_{j \rightarrow +\infty} \|x_{k_j} - x^*\| = 0$$

Autrement dit, la suite $(x_k)_{k \in \mathbb{N}}$ converge vers x^* . ■

3.3 Convergence vers un point critique dans le cas coercif et KL

Lorsque la fonction objectif n'est pas convexe, on peut obtenir des résultats intéressants en ajoutant deux autres hypothèses, qui sont la coercivité et la propriété KL (voir module **a3 : Propriété de KURDYKA-ŁOJASIEWICZ**). La première hypothèse assure que la suite des itérées est bornée :

Lemme 2

On suppose que \mathcal{X} est de dimension finie. Soit $J : \mathcal{X} \rightarrow \mathbb{R}$ une fonction **coercive** et L -régulière. Soit $(x_k)_{k \in \mathbb{N}}$ une suite générée par l'algorithme suivant

$$x_0 \in \mathcal{X} \quad \text{et} \quad \forall k \in \mathbb{N}, \quad x_{k+1} = x_k - \tau \nabla J(x_k)$$

Si $\tau \in]0; 2/L[$, alors la suite $(x_k)_{k \in \mathbb{N}}$ est bornée.

DÉMONSTRATION : La propriété de décroissance du critère assure que

$$\forall k \in \mathbb{N}, \quad x_k \in \text{niv}_{\leq J(x_0)} J$$

Or, on a démontré dans la preuve de la proposition 5 du module **B1 : Méthodes d'optimisation du premier ordre** que cet ensemble est borné, ce qui conclut la preuve. ■

Cette hypothèse n'est pas nécessaire s'il existe un autre moyen pour assurer que la suite générée par la méthode du gradient explicite est bornée.

Proposition 8

Soit $J : \mathcal{X} \rightarrow \mathbb{R}$ une fonction L -régulière. On suppose que J est minorée. Soit $(x_k)_{k \in \mathbb{N}}$ une suite générée par l'algorithme suivant

$$x_0 \in \mathcal{X} \quad \text{et} \quad \forall k \in \mathbb{N}, \quad x_{k+1} = x_k - \tau \nabla J(x_k)$$

Si la suite $(x_k)_{k \in \mathbb{N}}$ est **bornée** et que $\tau \in]0; 2/L[$, alors il existe une sous-suite $(x_{k_j})_{j \in \mathbb{N}}$ qui converge vers un point critique de J .

DÉMONSTRATION : Puisque la suite des x_k est supposée **bornée**, elle admet une sous-suite $(x_{k_j})_{j \in \mathbb{N}}$ convergente. Notons \bar{x} sa limite. D'après la proposition 5, la sous-suite des gradients $\nabla J(x_{k_j})$ converge vers 0. Or, par continuité, on a

$$0 = \lim_{j \rightarrow +\infty} \nabla J(x_{k_j}) = \nabla J \left(\lim_{j \rightarrow +\infty} x_{k_j} \right) = \nabla J(\bar{x})$$

D'où le résultat annoncé. ■

Une conséquence immédiate de ce résultat est le suivant :

Corollaire 1 (Convergence des itérés dans le cas borné)

Soit $J : \mathcal{X} \rightarrow \mathbb{R}$ une fonction L -régulière. On suppose que J admet un unique minimiseur x^* . Soit $(x_k)_{k \in \mathbb{N}}$ une suite générée par l'algorithme suivant

$$x_0 \in \mathcal{X} \quad \text{et} \quad \forall k \in \mathbb{N}, \quad x_{k+1} = x_k - \tau \nabla J(x_k)$$

Si la suite $(x_k)_{k \in \mathbb{N}}$ est **bornée** et que $\tau \in]0; 2/L[$, alors la suite $(x_k)_{k \in \mathbb{N}}$ converge vers x^* .

DÉMONSTRATION : En effet, la proposition précédente assure que tout point d'adhérence de la suite des itérées est un point critique de J . Si J admet un unique minimiseur, alors l'adhérence de cette suite est réduite à ce point. Il s'ensuit qu'elle est nécessairement convergente. ■

En combinant le lemme 2 et la proposition 8, on prouve le corollaire suivant :

Corollaire 2

On suppose que \mathcal{X} est de dimension finie. Soit $J : \mathcal{X} \rightarrow \mathbb{R}$ une fonction **coercive** et L -régulière. Soit $(x_k)_{k \in \mathbb{N}}$ une suite générée par l'algorithme suivant

$$x_0 \in \mathcal{X} \quad \text{et} \quad \forall k \in \mathbb{N}, \quad x_{k+1} = x_k - \tau \nabla J(x_k)$$

Si $\tau \in]0; 2/L[$, alors il existe une sous-suite $(x_{k_j})_{j \in \mathbb{N}}$ qui converge vers un point critique de J .

Pour obtenir la convergence forte des itérées, il faut généralement, dans le cas non-convexe, utiliser la propriété KL.

Proposition 9 (Convergence des itérés dans le cas KL et coercif)

Soit $J : \mathcal{X} \rightarrow \mathbb{R}$ une fonction KL, coercive et L -régulière. Soit $(x_k)_{k \in \mathbb{N}}$ une suite générée par l'algorithme suivant

$$x_0 \in \mathcal{X} \quad \text{et} \quad \forall k \in \mathbb{N}, \quad x_{k+1} = x_k - \tau \nabla J(x_k)$$

Si $\tau \in]0; 2/L[$, alors la suite $(x_k)_{k \in \mathbb{N}}$ converge vers un minimiseur de J .

DÉMONSTRATION : Il s'agit d'appliquer le théorème d'ATTOUCH, BOLTE & SVAITER (module **A3 : Propriété de KURDYKA-ŁOJASIEWICZ**). La condition (C1) de décroissance est une conséquence directe de la régularité de J et du choix associé de τ (proposition 4). La condition (C2) d'erreur relative découle directement de la définition des itérées, puisque

$$x_{k+1} = x_k - \tau \nabla J(x_k) \quad \implies \quad \|\nabla J(x_k)\| = \frac{1}{\tau} \|x_k - x_{k+1}\|$$

Enfin, la condition (C3) de continuité est une conséquence de la coercivité qui, couplée avec le caractère décroissant du critère, assure l'existence d'un point d'adhérence pour la suite des itérées, tandis que la limite du critère est due à la continuité de J . ■

3.4 Convergence linéaire dans le cas fortement convexe

Dans le cas convexe, on ne peut *a priori* pas établir de taux de convergence pour la suite des itérées. Dans le cas fortement convexe en revanche, il est possible d'établir un taux de convergence linéaire :

Théorème 1 (Convergence linéaire dans le cas fortement convexe)

Soit $J : \mathcal{X} \rightarrow \mathbb{R}$ une fonction convexe L -régulière. On suppose que J est **fortement convexe** de module α . Soit $(x_k)_{k \in \mathbb{N}}$ une suite générée par l'algorithme suivant

$$x_0 \in \mathcal{X} \quad \text{et} \quad \forall k \in \mathbb{N}, \quad x_{k+1} = x_k - \tau \nabla J(x_k)$$

Si $\tau \in]0; 1/L]$, alors la suite $(x_k)_{k \in \mathbb{N}}$ converge vers le minimiseur x^* de J et

$$\forall k \in \mathbb{N}, \quad 0 \leq \|x_k - x^*\|^2 \leq (1 - \tau\alpha)^k \|x_0 - x^*\|^2$$

REMARQUE : Attention aux valeurs possibles pour le pas de temps τ , qui diffèrent des propositions précédentes !

DÉMONSTRATION : Dans la preuve du lemme 2, on a démontré que

$$\|x_{k+1} - x^*\|^2 = \|x_k - x^*\|^2 + 2\tau \langle \nabla J(x_k), x^* - x_k \rangle + \tau^2 \|\nabla J(x_k)\|^2$$

(en utilisant le fait que $\nabla J(x^*) = 0$). D'après la caractérisation de la forte convexité par les (sous-)gradients (corollaire 1 du module **A2 : Sous-différentiabilité**), on a

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - 2\tau \left(J(x_k) - J(x^*) + \frac{\alpha}{2} \|x_k - x^*\|^2 \right) + \tau^2 \|\nabla J(x_k)\|^2$$

Dans la proposition 4, on a par ailleurs démontré que

$$J(x_{k+1}) - J(x_k) \leq -\left(\tau - \frac{\tau^2}{2} L\right) \|\nabla J(x_k)\|^2$$

de sorte que, puisque $J(x^*) \leq J(x_{k+1})$ par optimalité, on a (après simplification)

$$\|x_{k+1} - x^*\|^2 \leq (1 - \tau\alpha) \|x_k - x^*\|^2 - \tau \frac{\tau - \tau^2 L}{\tau - \tau^2 L/2} (J(x_k) - J(x^*))$$

Dans le quotient qui apparaît dans l'inégalité ci-dessus, le dénominateur est strictement positif car $0 < \tau < 2/L$, tandis que le numérateur est positif car $0 < \tau \leq 1/L$. On peut donc ignorer ce terme et terminer la preuve par une récurrence. ■

Le taux de convergence établi dans cette dernière proposition est donc un taux de convergence linéaire, avec

$$0 < 1 - \tau\alpha < 1$$

car $L \geq \alpha$ d'après une remarque du module **A4 : Fonctions régulières**. Celui-ci est d'autant plus intéressant que la quantité ci-dessus est proche de zéro. On voit donc qu'il s'agit de choisir τ aussi grand que possible dans l'intervalle autorisé, c'est-à-dire

$$\tau^* = \frac{1}{L}$$

Le pas de temps optimal sous ces contraintes est donc le même que celui déterminé dans le cas convexe. Dans ce cas, on a

$$\forall k \in \mathbb{N}, \quad 0 \leq \|x_k - x^*\|^2 \leq \left(1 - \frac{\alpha}{L}\right)^k \|x_0 - x^*\|^2 = \left(1 - \frac{1}{\kappa}\right)^k \|x_0 - x^*\|^2$$

et on voit que la convergence de la méthode du gradient explicite à pas constant optimal est d'autant plus bonne que le conditionnement de la fonction objectif est bon.

Il est important de noter qu'en pratique, seul le résultat de convergence dans le cas fortement convexe est réellement exploitable. En effet, sans taux de convergence, il n'existe aucun indice sur le temps nécessaire (c'est-à-dire, le nombre d'itérations nécessaire) pour obtenir une estimation *suffisamment bonne* d'une solution du problème. Par ailleurs, on notera que, pour l'erreur sur le critère dans le cas convexe, par exemple (proposition 6), dépend de la distance entre le point initial x_0 et un minimiseur. Il est évident que cette distance est généralement inconnue et difficile à estimer. Or, si l'initialisation est faite arbitrairement, cette distance peut être importante. Dans le cas d'une convergence linéaire, la distance au minimiseur décroît de manière exponentielle vers 0, de sorte que l'initialisation a un poids moins important.

4 Application : débruitage de signal

4.1 Régularisation de TIKHONOV

Une application possible pour la méthode du gradient explicite est le problème de débruitage d'un signal bruité. Plus précisément, on a un signal (son, image par exemple) $g \in \mathcal{X}$ (où $\mathcal{X} = \mathbb{R}^n$ dans le cas d'un son, $\mathcal{X} = \mathbb{R}^{mn}$ ou \mathbb{R}^{3mn} dans le cas d'une image en niveaux de gris ou en couleur RGB), résultant d'une mesure (enregistrement, prise de vue). Cette mesure a introduit du bruit additif, c'est-à-dire que g est une version dégradée d'un signal idéal u , et que la relation entre la mesure et le signal idéal est modélisée par

$$g = u + n$$

avec n le bruit additif. Sans entrer dans les détails, si n est un bruit blanc gaussien, et que le signal est supposé *relativement régulier*¹, alors on peut essayer de trouver une

1. En réalité, ce n'est pas le cas pour une image, qui comporte des discontinuités. On le voit d'ailleurs dans les tests expérimentaux, où la solution trouvée est certes débruitée, mais floue, ce qui correspond à une image régulière.

estimation de u en résolvant le problème d'optimisation suivant

$$\min_{u \in \mathcal{X}} \left\{ \frac{1}{2} \|u - g\|_2^2 + \frac{\lambda}{2} \|\nabla^h u\|_2^2 \right\}$$

Le premier terme est appelé *terme d'attache aux données* ou *terme de fidélité* ; il sert à imposer à la solution u^* à rester proche des données bruitées g . Le second terme est le terme de régularisation : il permet d'injecter une forme d'informations supplémentaires sur la solution, ici de la régularité grâce à la régularisation dite de TIKHONOV. Elle est définie à l'aide d'une discrétisation du gradient spatial. Enfin, le réel positif λ est un paramètre qui permet de moduler l'importance relative des deux termes ; plus λ est grand, plus le terme de régularisation est prépondérant, et de manière équivalente moins le terme d'attache aux données est fort ; en particulier, cela signifie que la solution u^* sera régulier, au détriment de sa proximité avec la donnée bruitée g .

En posant J la fonction objectif de ce problème, on voit que J est convexe, et même fortement convexe (de module 1) grâce au terme d'attache aux données. Elle est également différentiable, de gradient

$$\forall u \in \mathcal{X}, \quad \nabla J(u) = u - g + \lambda (\nabla^h)^* \nabla^h u = (\text{Id} + \lambda (\nabla^h)^* \nabla^h) u - g$$

On voit donc que ∇J est lipschitzien, de constante de LIPSCHITZ

$$L = \|\text{Id} + \lambda (\nabla^h)^* \nabla^h\|$$

Les itérations de la méthode du gradient explicite s'écrivent

$$\forall k \in \mathbb{N}, \quad u_{k+1} = u_k - \tau (u_k - g + \lambda (\nabla^h)^* \nabla^h u_k)$$

et si le pas de temps est choisi strictement inférieur à $2/L$, on a la convergence de l'algorithme, tandis qu'un pas de temps strictement inférieur à $1/L$ assure une convergence linéaire.

4.2 Approximation du problème de RUDIN–OSHER–FATEMI

La méthode décrite dans le paragraphe précédent souffre d'un inconvénient important : elle a tendance à sur-lisser les images / signaux. Dans le cas d'une image qui comporte des discontinuités (par exemple des photographies usuelles), cela crée des artefacts gênants.

Une manière de restituer les discontinuités naturelles présentes dans le signal original (idéal), RUDIN, OSHER et FATEMI ont proposé de remplacer la régularisation de TIKHONOV par la régularisation suivante :

$$\text{TV}(u) = \|\nabla^h u\|_2$$

appelée *variation totale* de u . Ici, dans sa version discrétisée, il s'agit simplement de remplacer la norme au carré par la norme. En procédant de la sorte, on perd la régularité de la fonction objective, et la méthode du gradient explicite n'est plus applicable.

Pour être en mesure d'appliquer cette méthode, on peut remplacer la norme (non différentiable), par une approximation régulière, comme celles introduites dans le module **A4 : Fonctions régulières**, par exemple

$$\widetilde{\text{TV}}_\varepsilon(u) = \sqrt{\|\nabla^h u\|_2^2 + \varepsilon}$$

avec $\varepsilon > 0$. Notons que, plus ε est petit, plus l'approximation est précise, mais plus la constante de LIPSCHITZ du gradient de J est grand (et donc les pas de temps de la méthode du gradient petits).

Pour aller plus loin

Fonctions régulières. La méthode du gradient explicite est la méthode la plus simple pour optimiser une fonction régulière, mais n'est pas applicable qu'à une famille de problèmes restreinte. La différentiabilité est en particulier une propriété difficile à obtenir dans les problèmes pratiques. Une manière de l'obtenir est de remplacer le problème initial par un problème régulier en remplaçant tout ou partie de la fonction objectif par une approximation régulière. Lorsque la non-régularité de la fonction objectif provient de l'utilisation d'une norme, on peut utiliser l'approximation régulière de la valeur absolue introduite dans le module **A4 : Fonctions régulières**.

Algorithme du point proximal. Lorsque le problème est non différentiable, une idée naturelle est de remplacer le gradient par un sous-gradient. En pratique, le comportement de cette variante n'est pas intéressant (en particulier, le sous-gradient n'est pas nécessairement une direction de descente). En revanche, au lieu de prendre un sous-gradient au point courant x_k , si on le prend au point suivant x_{k+1} , alors on obtient un algorithme très adapté à l'optimisation de fonctions non différentiables, appelé *algorithme du point proximal* (module **B3 : Algorithme du point proximal**). La difficulté principale réside dans la possibilité de calculer l'itération suivante, alors que le sous-gradient (qui définit donc ce point) est lui-même pris en ce même point. On verra que ce calcul fait intervenir un opérateur appelé *proximal* (module **A5 : Opérateur proximal de MOREAU**).

Éclatement d'opérateurs. Les itérations de la méthode du gradient interviennent dans d'autres algorithmes, lorsque la fonction objectif est partiellement régulière. On pourra citer en particulier l'algorithme FBS (module **B4 : Éclatement primal d'opérateurs**), pour lequel l'étude de la convergence utilise des résultats établis pour la méthode du gradient.