Travaux pratiques N°2
Méthodes de gradient

Objectif : blabla

1 Ensembles de niveau

Soit $f : \mathcal{X} \to \mathbb{R}$ une fonction. Soit $t \in \mathbb{R}$. On appelle *ensemble de niveau inférieur t* l'ensemble des points suivant :

$$\mathrm{niv}_{\leq t} f = \Big\{ x \in \mathcal{X} \mid f(x) \leq t \Big\} = f^{-1}(]-\infty\,; t\,])$$

et *ensemble de niveau t* l'ensemble des points suivant :

$$\mathrm{niv}_{t} f = \Big\{ x \in \mathcal{X} \mid f(x) = t \Big\} = f^{-1}(t)$$

Exercice 1 – Ensembles de niveau

(a) Montrer que les ensembles de niveau inférieur d'une fonction $f : \mathcal{X} \to \mathbb{R}$ sont emboîtés, c'est-à-dire que, pour tous réels $y \leq y'$, on a

$$\Big\{ x \in \mathcal{X} \mid f(x) \leq y \Big\} \subset \Big\{ x \in \mathcal{X} \mid f(x) \leq y' \Big\}$$

(b) Montrer que les ensembles de niveau d'une fonction convexe sont convexes.

Exercice 2 – Fonctions affines

Soit $(a, b) \in \mathbb{R}^n \times \mathbb{R}$. On considère la fonction affine :

$$f : \left\{ \begin{array}{ccc} \mathbb{R}^n & \to & \mathbb{R} \\ x & \mapsto & \langle a, x \rangle + b \end{array} \right.$$

(a) Montrer que les ensembles de niveau de f sont des droites parallèles.

(b) Vérifier que $\qquad \forall\, x \in \mathbb{R}^n, \qquad \nabla f(x) = a$

En déduire que, pour tout $x_0 \in \mathbb{R}^n$, le vecteur $\nabla J(x_0)$ est orthogonal à $\mathrm{niv}_{f(x_0)} f$ en x_0. Autrement dit,

$$\forall\, x \in \mathbb{R}^n, \qquad \langle \nabla f(x_0), x - x_0 \rangle = 0$$

On peut généraliser ce résultat à toute fonction différentiable : pour tout $x_0 \in \mathbb{R}^n$, le vecteur $\nabla J(x_0)$ est orthogonal à $\mathrm{niv}_{f(x_0)} f$ en x_0. Autrement dit,

$$\forall\, x \in \mathbb{R}^n, \qquad \langle \nabla f(x_0), x - x_0 \rangle = 0$$

2 Méthode du gradient à pas fixe

Soit $(a, b) \in \mathbb{R}^2$. On considère la fonction quadratique :

$$J : \left\{ \begin{array}{ccc} \mathbb{R}^2 & \to & \mathbb{R} \\ (x, y) & \mapsto & a\,x^2 + b\,y^2 \end{array} \right.$$

Exercice 3 – Propriétés de J

(a) Justifier que la fonction J est de classe $\mathcal{C}^\infty$ et montrer que

$$\forall\,(x, y) \in \mathbb{R}^2, \qquad \nabla J(x, y) = \begin{pmatrix} a\,x \\ b\,y \end{pmatrix}$$

(b) En déduire que ∇J est $\max(|a|, |b|)$-lipschitzien.

(c) Montrer que J est convexe si et seulement si a et b sont positifs.

(d) Justifier que $(0, 0)$ est l'unique minimiseur de J.

(e) On suppose que a et b sont strictement positifs. Montrer que J est une fonction fortement convexe.

Exercice 4 – Fonctions quadratiques

(a) Pour différentes valeurs de $(a, b) \in \mathbb{R}$, tracer des ensembles de niveau de J.

(b) Soit $(x_0, y_0) \in \mathbb{R}^2$. Tracer l'ensemble de niveau $J(x_0, y_0)$, puis le vecteur $\nabla J(x_0, y_0)$ au point (x_0, y_0).

Lorsque J est une fonction convexe à gradient lipschitzien, de constante de Lipschitz L, et admet un minimiseur, alors la méthode du gradient à pas constant suivant

$$x_0 \in \mathcal{X} \qquad \text{et} \qquad \forall\,k \in \mathbb{N}, \quad \begin{cases} \tau & < \dfrac{2}{L} \\ x_{k+1} = x_k - \tau\,\nabla J(x_k) \end{cases}$$

converge dans le sens où

- la suite $(x_k)_{k\in\mathbb{N}}$ converge vers un minimiseur de J ;
- la suite $(J(x_k))_{k\in\mathbb{N}}$ converge vers le minimum de J ;
- la suite $(\nabla J(x_k))_{k\in\mathbb{N}}$ converge vers 0.

3 Méthode du gradient à pas optimal

On rappelle que la méthode du gradient à pas optimal consiste à calculer les itérations suivantes :

$$x_0 \in \mathcal{X} \qquad \text{et} \qquad \forall\,k \in \mathbb{N}, \quad \begin{cases} \tau_k & \in \arg\min\limits_{\tau\in\mathbb{R}} J\big(x_k - \tau\,\nabla J(x_k)\big) \\ x_{k+1} = x_k - \tau_k\,\nabla J(x_k) \end{cases}$$

Il n'est pas prouvé que cette méthode converge dans le cas général. La principale difficulté de mise en œuvre est évidemment le cas du pas optimal τ_k.

Exercice 5 – Théorie

(a) Montrer que, si le pas optimal τ_k existe, alors il vérifie

$$\langle \nabla J(x_k), \nabla J\big(x_k - \tau_k \nabla J(x_k)\big)\rangle = 0$$

Comment interpréter géométriquement ce résultat ?

(b) Montrer que si J est infinie à l'infini alors le pas optimal τ_k existe.

On commence par considérer un cas où le pas optimal est meilleur que le pas fixe. On s'intéresse au cas particulier suivant

$$\min_{(x,y)\in\mathbb{R}^2} J(x,y) = x^2 + 100\, y^2$$

> Dans la feuille d'exercices ? ? ?, on montre que, dans le cas considéré, de même que dans le cas général ? ? ?, la méthode du gradient à pas optimal converge. Autrement dit, les suites $(J(x_k, y_k))_{k\in\mathbb{N}}$ et $(\|(x_k, y_k)\|)_{k\in\mathbb{N}}$ convergent vers $\min J$ et 0 respectivement.

Exercice 6 – Calcul du pas optimal

Montrer que le pas optimal vaut pour tout $k \in \mathbb{N}$

$$\tau_k = \frac{x_k^2 + 10^4\, y_k^2}{2\,(x_k^2 + 10^6\, y_k^2)}$$

Exercice 7 – Vitesse de convergence

(a) Tracer quelques lignes de niveau de la fonction J.

(b) Mettre en œuvre la méthode du gradient pour $\tau = 1.9999/100$ (pas fixe) et le pas optimal. On choisira comme initialisation $(x_0, y_0) = (1, 1)$.

(c) Sur le même graphique, afficher les suites $(x_k, y_k)_{k\in\mathbb{N}}$.

(d) Sur un deuxième graphique, afficher les courbes $(J(x_k, y_k))_{k\in\mathbb{N}}$.

(e) Sur un troisième graphique, afficher les courbes $(\|(x_k, y_k)\|)_{k\in\mathbb{N}}$.

> Lorsque J est une fonction α-convexe et que ∇J est une fonction L-lipschitzien, le rapport suivant
>
> $$\kappa = \frac{L}{\alpha}$$
>
> est appelé *conditionnement* de J. Cette valeur est nécessairement supérieure à 1. Lorsque cette valeur est grande, on dit que la fonction est mal conditionnée.

Parfois, le pas optimal est également optimal dans le sens où il s'agit du pas conduisant au meilleur taux de convergence. On s'intéresse au cas particulier suivant

$$\min_{(x,y)\in\mathbb{R}^2} J(x,y) = x^2 + 2\, y^2$$

Exercice 8 – Calcul du pas optimal

On suppose que $(x_0, y_0) = (2\,a, a)$ avec $a \neq 0$. Montrer que le pas optimal est constant et vaut $1/3$, et que pour tout $k \in \mathbb{N}$, on a

$$x_k^2 = 4\,y_k^2$$

On pourra raisonner par récurrence.

Exercice 9 – Vitesse de convergence

(a) Tracer quelques lignes de niveau de la fonction J.

(b) Mettre en œuvre la méthode du gradient pour $\tau \in \{1/3 \pm 0.01, 0.9999\}$ (pas fixe) et $\tau = 1/3$ (pas optimal). On choisira comme initialisation $(x_0, y_0) = (20, 10)$.

(c) Sur le même graphique, afficher les suites $(x_k, y_k)_{k\in\mathbb{N}}$.

(d) Sur un deuxième graphique, afficher les courbes $(J(x_k, y_k))_{k\in\mathbb{N}}$.

(e) Sur un troisième graphique, afficher les courbes $(\|(x_k, y_k)\|)_{k\in\mathbb{N}}$.

> On peut en réalité montrer que le pas optimal conduit dans ce cas au taux de convergence optimal. En d'autres termes, le choix du pas optimal permet d'approcher le minimiseur de J à une précision donnée en un minimum d'itérations.

L'initialisation est importante dans le cas du pas optimal. Revenons en effet au cas du problème mal conditionné

$$\min_{(x,y)\in\mathbb{R}^2} J(x, y) = x^2 + 100\,y^2$$

Si la méthode du gradient à pas optimal est initialisée de manière inappropriée, alors sa convergence peut être plus lente que celle du gradient à pas fixe.

Exercice 10 – Influence de l'initialisation

(a) Tracer quelques lignes de niveau de la fonction J.

(b) Mettre en œuvre la méthode du gradient à pas optimal. On choisira comme initialisation $(x_0, y_0) = (1, 1)$ et $(x_0, y_0) = (0, \sqrt{2})$.

(c) Sur le même graphique, afficher les suites $(x_k, y_k)_{k\in\mathbb{N}}$.

(d) Sur un deuxième graphique, afficher les courbes $(J(x_k, y_k))_{k\in\mathbb{N}}$.

(e) Sur un troisième graphique, afficher les courbes $(\|(x_k, y_k)\|)_{k\in\mathbb{N}}$.

> Ce n'est pas la distance au minimiseur de l'initialisation qui est en cause ici, puisqu'elle vaut $\sqrt{2}$ dans les deux cas.

Dans le cas général, il est très difficile de calculer le pas optimal. Lorsque la fonction objectif J est deux fois différentiable, la formule de Taylor permet d'écrire au voisinage de x_k

$$J(x) \approx \underbrace{J(x_k) + \langle \nabla J(x_k), x - x_k \rangle + \frac{1}{2} \langle \mathrm{Hess}\, J(x_k)\,(x - x_k), x - x_k \rangle}_{\tilde{J}_{x_k}(x)}$$

On notera que la fonction $\tilde{J}_{x_k}$ est une fonction quadratique. Dans ce cas, au lieu de calculer le pas optimal pour la fonction J, on se contente de calculer le pas optimal pour la fonction $\tilde{J}_{x_k}$, de sorte que la méthode du gradient à pas optimal devient

$$x_0 \in \mathcal{X} \qquad \text{et} \qquad \forall\, k \in \mathbb{N}, \qquad \begin{cases} \tilde{\tau}_k & \in \arg\min\limits_{\tau \in \mathbb{R}} \tilde{J}_{x_k}\big(x_k - \tau\, \nabla J(x_k)\big) \\ x_{k+1} = x_k - \tilde{\tau}_k\, \nabla J(x_k) \end{cases}$$

Exercice 11 – Pas optimal approché

Montrer que $\qquad \forall\, k \in \mathbb{N}, \qquad \tau_k = \dfrac{\|\nabla J(x_k)\|_2^2}{2\,\langle \mathrm{Hess} J(x_k)\, \nabla J(x_k), \nabla J(x_k)\rangle}$

On considère le problème suivant :

$$\min_{(x,y)\in\mathbb{R}^2} J(x,y) = (x + 2\,y)^4$$

Exercice 12 – Calcul des pas de temps

(a) Montrer que pour tout $(x,y) \in \mathbb{R}^2$,

$$\nabla J(x,y) = \begin{pmatrix} 4\,(x+2\,y)^3 \\ 8\,(x+2\,y)^3 \end{pmatrix} \qquad \text{et} \qquad \mathrm{Hess}\, J(x,y) = \begin{pmatrix} 12\,(x+2\,y)^2 & 24\,(x+2\,y)^2 \\ 24\,(x+2\,y)^2 & 48\,(x+2\,y)^2 \end{pmatrix}$$

(b) Montrer que le pas optimal vaut pour tout $k \in \mathbb{N}$

$$\tau_k = \frac{1}{20\,(x_k + 2\,y_k)^2}$$

(c) Montrer que le pas optimal approché vaut pour tout $k \in \mathbb{N}$

$$\tilde{\tau}_k = \frac{1}{120\,(x_k + 2\,y_k)^2}$$

Exercice 13 – Cas non quadratique

(a) Tracer quelques lignes de niveau de la fonction J.

(b) Mettre en œuvre la méthode du gradient à pas optimal avec les pas exact τ_k et approché $\tilde{\tau}_k$. On choisira comme initialisation $(x_0, y_0) = (1, 1)$.

(c) Sur le même graphique, afficher les suites $(x_k, y_k)_{k\in\mathbb{N}}$.

(d) Sur un deuxième graphique, afficher les courbes $(J(x_k, y_k))_{k\in\mathbb{N}}$.

(e) Sur un troisième graphique, afficher les courbes $(\|(x_k, y_k)\|)_{k\in\mathbb{N}}$.