IV. - Algorithmes itératifs

Jean-Philippe Préaux

EOAA - 2009/10

Soit le problème d'optimisation :

$$\min_{\mathbf{x}\in\mathcal{D}} f(\mathbf{x})$$

pour $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ et $\mathcal{D} \subset \mathbb{R}^n$.

Copyright : Jean-Philippe Préaux

Soit le problème d'optimisation :

$$\min_{\mathbf{x}\in\mathcal{D}} f(\mathbf{x})$$

pour $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ et $\mathcal{D} \subset \mathbb{R}^n$.

Objectif de ce chapitre : se rapprocher algorithmiquement d'une solution en construisant une suite $(\mathbf{u}_n)_{n\in\mathbb{N}}$ de $\mathbb{R}^n$ qui converge vers un minimum global de f sur $\mathcal{D}$.

Copyright : Jean-Philippe Préaux

Soit le problème d'optimisation :

$$\min_{\mathbf{x}\in\mathcal{D}} f(\mathbf{x})$$

pour $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ et $\mathcal{D} \subset \mathbb{R}^n$.

Objectif de ce chapitre : se rapprocher algorithmiquement d'une solution en construisant une suite $(\mathbf{u}_n)_{n\in\mathbb{N}}$ de $\mathbb{R}^n$ qui converge vers un minimum global de f sur $\mathcal{D}$.
On établit pour cela plusieurs méthodes, et on s'intéresse à :

Copyright : Jean-Philippe Préaux

Soit le problème d'optimisation :

$$\min_{\mathbf{x}\in\mathcal{D}} f(\mathbf{x})$$

pour $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ et $\mathcal{D} \subset \mathbb{R}^n$.

Objectif de ce chapitre : se rapprocher algorithmiquement d'une solution en construisant une suite $(\mathbf{u}_n)_{n\in\mathbb{N}}$ de $\mathbb{R}^n$ qui converge vers un minimum global de f sur $\mathcal{D}$.
On établit pour cela plusieurs méthodes, et on s'intéresse à :

- des conditions suffisantes sur f et $\mathcal{D}$ pour que

$$\mathbf{u}_n \underset{n\to\infty}{\longrightarrow} \min_{\mathbf{x}\in\mathcal{D}} f(\mathbf{x}),$$

Copyright : Jean-Philippe Préaux

Soit le problème d'optimisation :

$$\min_{\mathbf{x}\in\mathcal{D}} f(\mathbf{x})$$

pour $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ et $\mathcal{D} \subset \mathbb{R}^n$.

Objectif de ce chapitre : se rapprocher algorithmiquement d'une solution en construisant une suite $(\mathbf{u}_n)_{n\in\mathbb{N}}$ de $\mathbb{R}^n$ qui converge vers un minimum global de f sur $\mathcal{D}$.
On établit pour cela plusieurs méthodes, et on s'intéresse à :

- des conditions suffisantes sur f et $\mathcal{D}$ pour que

$$\mathbf{u}_n \underset{n\to\infty}{\longrightarrow} \min_{\mathbf{x}\in\mathcal{D}} f(\mathbf{x}),$$

- sa vitesse de convergence.

Copyright : Jean-Philippe Préaux

Nous établirons d'abord des algorithmes dans le cas sans contrainte, *i.e.* $\mathcal{D} = \mathbb{R}^n$, puis dans le cas sous contraintes.

Copyright : Jean-Philippe Préaux

Nous établirons d'abord des algorithmes dans le cas sans contrainte, *i.e.* $\mathcal{D} = \mathbb{R}^n$, puis dans le cas sous contraintes.

Nous verrons plusieurs types de méthodes ; on les classifie parfois en :

Copyright : Jean-Philippe Préaux

Nous établirons d'abord des algorithmes dans le cas sans contrainte, *i.e.* $\mathcal{D} = \mathbb{R}^n$, puis dans le cas sous contraintes.

Nous verrons plusieurs types de méthodes ; on les classifie parfois en :

- *Les méthodes directes*. Elles n'utilisent pas les dérivées de l'application.

Copyright : Jean-Philippe Préaux

Nous établirons d'abord des algorithmes dans le cas sans contrainte, *i.e.* $\mathcal{D} = \mathbb{R}^n$, puis dans le cas sous contraintes.

Nous verrons plusieurs types de méthodes ; on les classifie parfois en :

• *Les méthodes directes*. Elles n'utilisent pas les dérivées de l'application. Nous n'en verrons aucune, mais on peut citer comme exemple la *méthode de Hooke-Jeeves* qui consiste à fixer un pas $\rho > 0$ puis à construire $\mathbf{u}_{k+1}$ à partir de $\mathbf{u}_k$, en choisissant parmi tous les points $\mathbf{u}_k \pm \rho\,\mathbf{e}_i$ (où $\mathbf{e}_i$, $i = 1, \ldots, n$ désignent les vecteurs de la base canonique) celui dont la valeur prise par f est minimale. Si $\mathbf{u}_{k+1} = \mathbf{u}_k$ on dimimue le pas.

Copyright : Jean-Philippe Préaux

Nous établirons d'abord des algorithmes dans le cas sans contrainte, *i.e.* $\mathcal{D} = \mathbb{R}^n$, puis dans le cas sous contraintes.

Nous verrons plusieurs types de méthodes ; on les classifie parfois en :

• *Les méthodes directes*. Elles n'utilisent pas les dérivées de l'application. Nous n'en verrons aucune, mais on peut citer comme exemple la *méthode de Hooke-Jeeves* qui consiste à fixer un pas $\rho > 0$ puis à construire $\mathbf{u}_{k+1}$ à partir de $\mathbf{u}_k$, en choisissant parmi tous les points $\mathbf{u}_k \pm \rho\,\mathbf{e}_i$ (où $\mathbf{e}_i$, $i = 1, \ldots, n$ désignent les vecteurs de la base canonique) celui dont la valeur prise par f est minimale. Si $\mathbf{u}_{k+1} = \mathbf{u}_k$ on dimimue le pas.
Cette méthode est employée pour minimiser une application non différentiable ; sa convergence, lorsqu'elle a lieu, est très lente.

Copyright : Jean-Philippe Préaux

• *Les méthodes de descente*. Elles utilisent les dérivées d'ordre 1. Ici $\mathbf{u}_{k+1}$ est construit à partir de $\mathbf{u}_k$ en choisissant une *direction de descente* $\mathbf{d}_k$ et un *pas de descente* $\rho_k > 0$, tels que $\mathbf{u}_{k+1} = \mathbf{u}_k + \rho_k \mathbf{d}_k$.

Copyright : Jean-Philippe Préaux

• *Les méthodes de descente*. Elles utilisent les dérivées d'ordre 1. Ici $\mathbf{u}_{k+1}$ est construit à partir de $\mathbf{u}_k$ en choisissant une *direction de descente* $\mathbf{d}_k$ et un *pas de descente* $\rho_k > 0$, tels que $\mathbf{u}_{k+1} = \mathbf{u}_k + \rho_k \mathbf{d}_k$.
Nous en verrons plusieurs :
– la *méthode de relaxation* qui prend pour direction de descente la direction des axes de façon cyclique et calcule le pas de descente en se ramenant à un problème de minimisation à une variable ;

Copyright : Jean-Philippe Préaux

• *Les méthodes de descente*. Elles utilisent les dérivées d'ordre 1. Ici $\mathbf{u}_{k+1}$ est construit à partir de $\mathbf{u}_k$ en choisissant une *direction de descente* $\mathbf{d}_k$ et un *pas de descente* $\rho_k > 0$, tels que $\mathbf{u}_{k+1} = \mathbf{u}_k + \rho_k \mathbf{d}_k$.
Nous en verrons plusieurs :
– la *méthode de relaxation* qui prend pour direction de descente la direction des axes de façon cyclique et calcule le pas de descente en se ramenant à un problème de minimisation à une variable ;
– la *méthode du gradient à pas optimal* qui prend comme direction de descente la direction locale de plus grande pente, l'opposé du vecteur gradient, et détermine le pas optimal en se ramenant à un problème à une variable ; ces deux dernières méthodes ont pour désavantage la résolution à chaque pas d'un problème à une variable ;

Copyright : Jean-Philippe Préaux

• *Les méthodes de descente*. Elles utilisent les dérivées d'ordre 1. Ici $\mathbf{u}_{k+1}$ est construit à partir de $\mathbf{u}_k$ en choisissant une *direction de descente* $\mathbf{d}_k$ et un *pas de descente* $\rho_k > 0$, tels que $\mathbf{u}_{k+1} = \mathbf{u}_k + \rho_k \mathbf{d}_k$.
Nous en verrons plusieurs :
– la *méthode de relaxation* qui prend pour direction de descente la direction des axes de façon cyclique et calcule le pas de descente en se ramenant à un problème de minimisation à une variable ;
– la *méthode du gradient à pas optimal* qui prend comme direction de descente la direction locale de plus grande pente, l'opposé du vecteur gradient, et détermine le pas optimal en se ramenant à un problème à une variable ; ces deux dernières méthodes ont pour désavantage la résolution à chaque pas d'un problème à une variable ;
– la *méthode du gradient à pas fixe* s'en démarque en fixant le pas ;

Copyright : Jean-Philippe Préaux

• *Les méthodes de descente*. Elles utilisent les dérivées d'ordre 1. Ici $\mathbf{u}_{k+1}$ est construit à partir de $\mathbf{u}_k$ en choisissant une *direction de descente* $\mathbf{d}_k$ et un *pas de descente* $\rho_k > 0$, tels que $\mathbf{u}_{k+1} = \mathbf{u}_k + \rho_k \mathbf{d}_k$.
Nous en verrons plusieurs :
– la *méthode de relaxation* qui prend pour direction de descente la direction des axes de façon cyclique et calcule le pas de descente en se ramenant à un problème de minimisation à une variable ;
– la *méthode du gradient à pas optimal* qui prend comme direction de descente la direction locale de plus grande pente, l'opposé du vecteur gradient, et détermine le pas optimal en se ramenant à un problème à une variable ; ces deux dernières méthodes ont pour désavantage la résolution à chaque pas d'un problème à une variable ;
– la *méthode du gradient à pas fixe* s'en démarque en fixant le pas ;
– enfin nous verrons la *méthode du gradient conjugué*, très ingénieuse, qui pour une fonction quadratique elliptique trouve le minimum en au plus n-itérations : on parle d'une *méthode exacte* par opposition aux *méthodes approchées* qui ne peuvent qu'approcher une solution.

Copyright : Jean-Philippe Préaux

• *Les méthodes utilisant les dérivées secondes.* Il s'agit essentiellement de la *méthode de Newton* (et de ses variantes) qui plus généralement détermine les zéros d'une application. Sous des hypothèses suffisantes sa convergence est très rapide ; elle a cependant pour désavantage qu'il faut choisir le point initial suffisamment proche du minimum.

Copyright : Jean-Philippe Préaux

• *Les méthodes utilisant les dérivées secondes.* Il s'agit essentiellement de la *méthode de Newton* (et de ses variantes) qui plus généralement détermine les zéros d'une application. Sous des hypothèses suffisantes sa convergence est très rapide ; elle a cependant pour désavantage qu'il faut choisir le point initial suffisamment proche du minimum.

Dans le cas sous contraintes, les méthodes que nous verrons sont déduites des méthodes ci-dessus en utilisant le théorème de projection convexe : *méthode de relaxation sous contraintes*, *méthode de gradient projeté*.

Copyright : Jean-Philippe Préaux

• *Les méthodes utilisant les dérivées secondes.* Il s'agit essentiellement de la *méthode de Newton* (et de ses variantes) qui plus généralement détermine les zéros d'une application. Sous des hypothèses suffisantes sa convergence est très rapide ; elle a cependant pour désavantage qu'il faut choisir le point initial suffisamment proche du minimum.

Dans le cas sous contraintes, les méthodes que nous verrons sont déduites des méthodes ci-dessus en utilisant le théorème de projection convexe : *méthode de relaxation sous contraintes*, *méthode de gradient projeté*.
Elles nécessitent cependant d'exprimer l'opérateur de projection convexe, ce qui n'est possible que dans des cas très simples.

Copyright : Jean-Philippe Préaux

• *Les méthodes utilisant les dérivées secondes.* Il s'agit essentiellement de la *méthode de Newton* (et de ses variantes) qui plus généralement détermine les zéros d'une application. Sous des hypothèses suffisantes sa convergence est très rapide ; elle a cependant pour désavantage qu'il faut choisir le point initial suffisamment proche du minimum.

Dans le cas sous contraintes, les méthodes que nous verrons sont déduites des méthodes ci-dessus en utilisant le théorème de projection convexe : *méthode de relaxation sous contraintes*, *méthode de gradient projeté*.
Elles nécessitent cependant d'exprimer l'opérateur de projection convexe, ce qui n'est possible que dans des cas très simples. La *méthode d'Uzawa* quant à elle contourne cette difficulté en mettant à profit la théorie de la dualité convexe pour rechercher un point-selle du lagrangien.

Copyright : Jean-Philippe Préaux

La convergence de toutes ces méthodes ne peut s'établir que sous de fortes hypothèses, tout au moins (hormis pour la méthode de Newton) l'ellipticité de f et la convexité du domaine $\mathcal{D}$.

Copyright : Jean-Philippe Préaux

La convergence de toutes ces méthodes ne peut s'établir que sous de fortes hypothèses, tout au moins (hormis pour la méthode de Newton) l'ellipticité de f et la convexité du domaine $\mathcal{D}$. En fait ce sont les méthodes algorithmiques à utiliser en programmation convexe, et uniquement dans ce cadre.

Copyright : Jean-Philippe Préaux

La convergence de toutes ces méthodes ne peut s'établir que sous de fortes hypothèses, tout au moins (hormis pour la méthode de Newton) l'ellipticité de f et la convexité du domaine $\mathcal{D}$. En fait ce sont les méthodes algorithmiques à utiliser en programmation convexe, et uniquement dans ce cadre. En programmation non convexe on utilise d'autres méthodes, stochastiques, programmation dynamique *etc...*, de la recherche opérationnelle qui sortent du cadre de ce cours (étudiées dans le cours de Recherche Opérationnelle EA3).

Copyright : Jean-Philippe Préaux

Vitesse de convergence.

Définitions. Soit $(\mathbf{u}_n)_{n\in\mathbb{N}}$ une suite de $\mathbb{R}^n$ qui converge vers $\mathbf{u} \in \mathbb{R}^n$. On note $\mathbf{r}_n = \mathbf{u} - \mathbf{u}_n$.

Copyright : Jean-Philippe Préaux

Vitesse de convergence.

Définitions. Soit $(\mathbf{u}_n)_{n\in\mathbb{N}}$ une suite de $\mathbb{R}^n$ qui converge vers $\mathbf{u} \in \mathbb{R}^n$. On note $\mathbf{r}_n = \mathbf{u} - \mathbf{u}_n$.

– Si $\exists\, k < 1$ et $n_0 \in \mathbb{N}$ tel que $\forall\, n > n_0$, $\|\mathbf{r}_{n+1}\| \leqslant k \|\mathbf{r}_n\|$ la convergence est dite *géométrique*.

Copyright : Jean-Philippe Préaux

Vitesse de convergence.

Définitions. Soit $(\mathbf{u}_n)_{n\in\mathbb{N}}$ une suite de $\mathbb{R}^n$ qui converge vers $\mathbf{u} \in \mathbb{R}^n$. On note $\mathbf{r}_n = \mathbf{u} - \mathbf{u}_n$.

– Si $\exists\, k < 1$ et $n_0 \in \mathbb{N}$ tel que $\forall\, n > n_0$, $\|\mathbf{r}_{n+1}\| \leqslant k \|\mathbf{r}_n\|$ la convergence est dite *géométrique*.

– Si $\exists\, p \in \mathbb{N}^*$ et $n_0 \in \mathbb{N}$ tel que $\forall\, n > n_0$, $\|\mathbf{r}_{n+1}\| \leqslant \|\mathbf{r}_n\|^p$ la convergence est dite *d'ordre p* ; si $p = 1$ elle est dite *linéaire* et si $p = 2$ elle est dite *quadratique*.

Copyright : Jean-Philippe Préaux

Vitesse de convergence.

Définitions. Soit $(\mathbf{u}_n)_{n\in\mathbb{N}}$ une suite de $\mathbb{R}^n$ qui converge vers $\mathbf{u} \in \mathbb{R}^n$. On note $\mathbf{r}_n = \mathbf{u} - \mathbf{u}_n$.

– Si $\exists\, k < 1$ et $n_0 \in \mathbb{N}$ tel que $\forall\, n > n_0$, $\|\mathbf{r}_{n+1}\| \leqslant k \|\mathbf{r}_n\|$ la convergence est dite *géométrique*.

– Si $\exists\, p \in \mathbb{N}^*$ et $n_0 \in \mathbb{N}$ tel que $\forall\, n > n_0$, $\|\mathbf{r}_{n+1}\| \leqslant \|\mathbf{r}_n\|^p$ la convergence est dite *d'ordre p* ; si $p = 1$ elle est dite *linéaire* et si $p = 2$ elle est dite *quadratique*.

– Si $\exists\, N_0$ tel que $\forall n \geqslant N_0$, $\|\mathbf{r}_n\| = 0$, la convergence est dite *finie*. C'est ce qui caractérise les *méthodes exactes*.

Copyright : Jean-Philippe Préaux

Algorithmes itératifs en optimisation sans contrainte
- Méthode de Newton
- Méthode de relaxation
- Méthode du gradient à pas optimal
- Méthode du gradient à pas fixe.
- Méthode du gradient conjugué

Algorithmes itératifs en optimisation sous contraintes
- Théorème de projection convexe
- Méthode de relaxation sur un produit d'intervalles
- Méthode du gradient projeté
- Méthode d'Uzawa

Exercices
- Exercice 1
- Exercice 2
- Exercice 3

Copyright . Jean-Philippe Préaux

Méthode de Newton.

Pour chercher un extremum $\mathbf{u}$ d'une fonction différentiable, on peut se ramener à chercher ses points critiques, $\nabla f(\mathbf{u}) = 0$.

Copyright : Jean-Philippe Préaux

Méthode de Newton.

Pour chercher un extremum $\mathbf{u}$ d'une fonction différentiable, on peut se ramener à chercher ses points critiques, $\nabla f(\mathbf{u}) = 0$.

Résoudre cette équation n'est pas toujours facile, ni même faisable.
Il est utile de considérer une méthode de calcul approché.

Copyright : Jean-Philippe Préaux

Méthode de Newton.

Pour chercher un extremum $\mathbf{u}$ d'une fonction différentiable, on peut se ramener à chercher ses points critiques, $\nabla f(\mathbf{u}) = 0$.

Résoudre cette équation n'est pas toujours facile, ni même faisable. Il est utile de considérer une méthode de calcul approché.

Nous voyons ici la (célèbre) méthode de Newton, qui d'une façon plus générale permet d'approcher les zéros d'une fonction (sous certaines hypothèses).

Copyright : Jean-Philippe Préaux

Méthode de Newton pour une application dérivable $f : \mathbb{R} \longrightarrow \mathbb{R}$.

Copyright : Jean-Philippe Préaux

Méthode de Newton pour une application dérivable $f : \mathbb{R} \longrightarrow \mathbb{R}$.

On cherche $u \in \mathbb{R}$, tel que $f(u) = 0$.

Copyright : Jean-Philippe Préaux

Méthode de Newton pour une application dérivable $f : \mathbb{R} \longrightarrow \mathbb{R}$.

On cherche $u \in \mathbb{R}$, tel que $f(u) = 0$.

Au voisinage de x_0, $f(x) \approx f(x_0) + f'(x_0)(x - x_0)$.

Copyright : Jean-Philippe Préaux

Méthode de Newton pour une application dérivable $f : \mathbb{R} \longrightarrow \mathbb{R}$.

On cherche $u \in \mathbb{R}$, tel que $f(u) = 0$.

Au voisinage de x_0, $f(x) \approx f(x_0) + f'(x_0)(x - x_0)$.

$\Longrightarrow$ Si $f'(x_0) \neq 0$, on considère :

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$

Copyright : Jean-Philippe Préaux

Méthode de Newton pour une application dérivable $f : \mathbb{R} \longrightarrow \mathbb{R}$.

On cherche $u \in \mathbb{R}$, tel que $f(u) = 0$.

Au voisinage de x_0, $f(x) \approx f(x_0) + f'(x_0)(x - x_0)$.

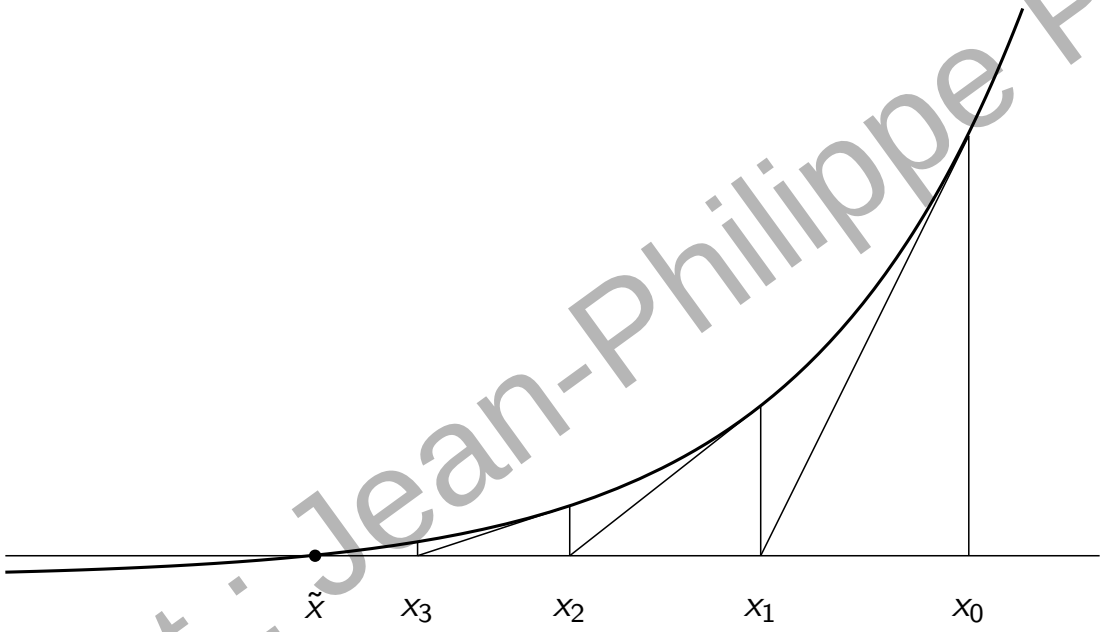
$\Longrightarrow$ Si $f'(x_0) \neq 0$, on considère :

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$

et on construit une suite par récurrence, par :

$$\text{si } f'(x_n) \neq 0, \quad x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

Copyright : Jean-Philippe Préaux



La méthode de descente de Newton pour la recherche du zéro d'une application $f : \mathbb{R} \longrightarrow \mathbb{R}$.

Copyright : Jean-Philippe Préaux

Convergence de la méthode de newton

Sous certaines hypothèses, la suite $(x_n)_{n\in\mathbb{N}}$ converge vers un zéro de f.

Copyright : Jean-Philippe Préaux

Convergence de la méthode de newton

Sous certaines hypothèses, la suite $(x_n)_{n\in\mathbb{N}}$ converge vers un zéro de f.

Plus précisément et plus généralement :

Théorème (Convergence de la méthode de Newton.)

Soit $F : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ de classe C^1, et $\mathbf{u}$ un zéro isolé de F.

Copyright : Jean-Philippe Préaux

Convergence de la méthode de newton

Sous certaines hypothèses, la suite $(x_n)_{n\in\mathbb{N}}$ converge vers un zéro de f.

Plus précisément et plus généralement :

Théorème (Convergence de la méthode de Newton.)

Soit $F : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ de classe C^1, et $\mathbf{u}$ un zéro isolé de F.
Si la matrice jacobienne $DF(\mathbf{u})$ de F en $\mathbf{u}$ est inversible, alors il existe une boule $B(\mathbf{u})$ centrée en $\mathbf{u}$, telle que $\forall\, \mathbf{u}_0 \in B(\mathbf{u})$, la suite :

$$\mathbf{u}_{n+1} = \mathbf{u}_n - DF(\mathbf{u}_n)^{-1} F(\mathbf{u}_n)$$

soit contenue dans $B(\mathbf{u})$ et converge vers $\mathbf{u}$ seul zéro de F dans $B(\mathbf{u})$.

Copyright : Jean-Philippe Préaux

Convergence de la méthode de newton

Sous certaines hypothèses, la suite $(x_n)_{n\in\mathbb{N}}$ converge vers un zéro de f.

Plus précisément et plus généralement :

Théorème (Convergence de la méthode de Newton.)

Soit $F : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ de classe C^1, et $\mathbf{u}$ un zéro isolé de F.
Si la matrice jacobienne $DF(\mathbf{u})$ de F en $\mathbf{u}$ est inversible, alors il existe une boule $B(\mathbf{u})$ centrée en $\mathbf{u}$, telle que $\forall\, \mathbf{u}_0 \in B(\mathbf{u})$, la suite :

$$\mathbf{u}_{n+1} = \mathbf{u}_n - DF(\mathbf{u}_n)^{-1} F(\mathbf{u}_n)$$

soit contenue dans $B(\mathbf{u})$ et converge vers $\mathbf{u}$ seul zéro de F dans $B(\mathbf{u})$. De plus la convergence est géométrique.

Copyright : Jean-Philippe Préaux

Avantage : La convergence est rapide. Si F est supposée de classe C^2 la convergence est même quadratique

Copyright : Jean-Philippe Préaux

Avantage : La convergence est rapide. Si F est supposée de classe C^2 la convergence est même quadratique

Désavantages : – Il faut prendre $\mathbf{u}_0$ suffisamment proche du zéro $\mathbf{u}$.

Copyright : Jean-Philippe Préaux

Avantage : La convergence est rapide. Si F est supposée de classe C^2 la convergence est même quadratique

Désavantages : – Il faut prendre $\mathbf{u}_0$ suffisamment proche du zéro $\mathbf{u}$.

– Le calcul de la matrice jacobienne et son inversion sont coûteux en temps de calcul. Pour cette raison ont été développées des méthodes dites 'quasi-Newton' où $DF(\mathbf{u}_n)$ est remplacée par une matrice moins coûteuse à inverser ; par exemple la matrice identité : c'est la *méthode des approximations successives*.

Copyright : Jean-Philippe Préaux

Avantage : La convergence est rapide. Si F est supposée de classe C^2 la convergence est même quadratique

Désavantages : – Il faut prendre $\mathbf{u}_0$ suffisamment proche du zéro $\mathbf{u}$.

– Le calcul de la matrice jacobienne et son inversion sont coûteux en temps de calcul. Pour cette raison ont été développées des méthodes dites 'quasi-Newton' où $DF(\mathbf{u}_n)$ est remplacée par une matrice moins coûteuse à inverser ; par exemple la matrice identité : c'est la *méthode des approximations successives*.

– Cette méthode est amplement employée en informatique, du fait de sa rapidité de convergence, par exemple pour le calcul approché de $\sqrt{\alpha}$ (avec $f(x) = x^2 - \alpha$) ou de $1/\alpha$ (avec $f(x) = \alpha x - 1$).

Voir l'exercice 1.

Copyright : Jean-Philippe Préaux

Exemple : calcul approché de $\sqrt{2}$.

Copyright : Jean-Philippe Préaux

Exemple : calcul approché de $\sqrt{2}$.

On applique la méthode de pour la recherche de zéro de l'application $f(x) = x^2 - 2$.

Copyright : Jean-Philippe Préaux

Exemple : calcul approché de $\sqrt{2}$.

On applique la méthode de pour la recherche de zéro de l'application $f(x) = x^2 - 2$.
Elle s'écrit :

$$\mathbf{u}_{k+1} = \mathbf{u}_k - f'(\mathbf{u}_k)^{-1} f(\mathbf{u}_k) = \mathbf{u}_k - (2\,\mathbf{u}_k)^{-1}(\mathbf{u}_k^2 - 2) = \frac{\mathbf{u}_k}{2} + \frac{1}{\mathbf{u}_k}\ .$$

Copyright : Jean-Philippe Préaux

Exemple : calcul approché de $\sqrt{2}$.

On applique la méthode de pour la recherche de zéro de l'application $f(x) = x^2 - 2$.
Elle s'écrit :

$$\mathbf{u}_{k+1} = \mathbf{u}_k - f'(\mathbf{u}_k)^{-1} f(\mathbf{u}_k) = \mathbf{u}_k - (2\,\mathbf{u}_k)^{-1}(\mathbf{u}_k^2 - 2) = \frac{\mathbf{u}_k}{2} + \frac{1}{\mathbf{u}_k}\ .$$

Code matlab (en prenant $u_0 = 1$) :

```
%% Méthode de Newton pour sqrt(2) %%
format long;
u=1;
N=5;
for i=1:N
    u=u/2+1/u
end
```

Copyright : Jean-Philippe Préaux

En prenant $\mathbf{u}_0 = 1$, on obtient :
Avec 10 chiffres significatifs : $(\sqrt{2} = 1.414213562373 09)$.

Copyright : Jean-Philippe Préaux

En prenant $\mathbf{u}_0 = 1$, on obtient :
Avec 10 chiffres significatifs : ($\sqrt{2} = 1.41421356237309$).

$$\begin{aligned}
\mathbf{u}_1 &= 1.50000000000000,\\
\mathbf{u}_2 &= 1.41666666666667,\\
\mathbf{u}_3 &= 1.41421568627451,\\
\mathbf{u}_4 &= 1.41421356237469,\\
\mathbf{u}_5 &= 1.41421356237309,
\end{aligned}$$

Copyright : Jean Philippe Préaux

En prenant $\mathbf{u}_0 = 1$, on obtient :
Avec 10 chiffres significatifs : $(\sqrt{2} = 1.41421356237309)$.

$$\begin{aligned}
\mathbf{u}_1 &= 1.50000000000000,\\
\mathbf{u}_2 &= 1.41666666666667,\\
\mathbf{u}_3 &= 1.41421568627451,\\
\mathbf{u}_4 &= 1.41421356237469,\\
\mathbf{u}_5 &= 1.41421356237309,
\end{aligned}$$

La convergence est particulièrement rapide ! On trouve une valeur approchée à 10^{-10} près en 5 itérations.

Copyright : Jean Philippe Préaux

En prenant $\mathbf{u}_0 = 1$, on obtient :
Avec 10 chiffres significatifs : ($\sqrt{2} = 1.41421356237309$).

$$\begin{aligned}
\mathbf{u}_1 &= 1.50000000000000,\\
\mathbf{u}_2 &= 1.41666666666667,\\
\mathbf{u}_3 &= 1.41421568627451,\\
\mathbf{u}_4 &= 1.41421356237469,\\
\mathbf{u}_5 &= 1.41421356237309,
\end{aligned}$$

La convergence est particulièrement rapide ! On trouve une valeur approchée à 10^{-10} près en 5 itérations.
Elle dépend ici peu du point base. Avec un point base négatif elle converge néanmoins vers $-\sqrt{2}$, c'est à dire vers l'autre zéro de f.

Copyright : Jean Philippe Préaux

En prenant $\mathbf{u}_0 = 1$, on obtient :
Avec 10 chiffres significatifs : ($\sqrt{2} = 1.41421356237309$).

$$\begin{aligned}
\mathbf{u}_1 &= 1.50000000000000,\\
\mathbf{u}_2 &= 1.41666666666667,\\
\mathbf{u}_3 &= 1.41421568627451,\\
\mathbf{u}_4 &= 1.41421356237469,\\
\mathbf{u}_5 &= 1.41421356237309,
\end{aligned}$$

La convergence est particulièrement rapide ! On trouve une valeur approchée à 10^{-10} près en 5 itérations.
Elle dépend ici peu du point base. Avec un point base négatif elle converge néanmoins vers $-\sqrt{2}$, c'est à dire vers l'autre zéro de f.
Pour des valeurs initiales s'éloignant de la solution le nombre d'itération nécessaire est plus important. On peut choisir u_0 grâce à $\sqrt{x} \approx \sqrt{1} - \frac{1}{2}(x-1)$.

Copyright : Jean Philippe Préaux

En prenant $\mathbf{u}_0 = 1$, on obtient :
Avec 10 chiffres significatifs : ($\sqrt{2} = 1.41421356237309$).

$$\begin{aligned}
\mathbf{u}_1 &= 1.50000000000000,\\
\mathbf{u}_2 &= 1.41666666666667,\\
\mathbf{u}_3 &= 1.41421568627451,\\
\mathbf{u}_4 &= 1.41421356237469,\\
\mathbf{u}_5 &= 1.41421356237309,
\end{aligned}$$

La convergence est particulièrement rapide ! On trouve une valeur approchée à 10^{-10} près en 5 itérations.
Elle dépend ici peu du point base. Avec un point base négatif elle converge néanmoins vers $-\sqrt{2}$, c'est à dire vers l'autre zéro de f.
Pour des valeurs initiales s'éloignant de la solution le nombre d'itération nécessaire est plus important. On peut choisir u_0 grâce à $\sqrt{x} \approx \sqrt{1} - \frac{1}{2}(x-1)$.
Voir à ce sujet l'exercice 1 du TP4...

Copyright : Jean Philippe Préaux

Application à la recherche de minimum.

Copyright : Jean-Philippe Préaux

Application à la recherche de minimum.

Pour résoudre

$$\min \ f(\mathbf{x})$$

avec $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, on applique la méthode de Newton à $\nabla f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$.

Copyright : Jean-Philippe Préaux

Application à la recherche de minimum.

Pour résoudre

$$\min \ f(\mathbf{x})$$

avec $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, on applique la méthode de Newton à $\nabla f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$.

Théorème (Application de la méthode de Newton à l'optimisation.)

Soit
$f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *de classe* C^2 *et* $\mathbf{u}$ *un minimum local de* f *isolé.*

Copyright : Jean Philippe Préaux

Application à la recherche de minimum.

Pour résoudre

$$\min \ f(\mathbf{x})$$

avec $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, on applique la méthode de Newton à $\nabla f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$.

Théorème (Application de la méthode de Newton à l'optimisation.)

Soit

$f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *de classe* C^2 *et* $\mathbf{u}$ *un minimum local de* f *isolé. Si* $\nabla^2 f(\mathbf{u})$ *est définie positive, alors* $\exists\, B(\mathbf{u})$ *une boule centrée en* $\mathbf{u}$, *tel que* $\forall\, \mathbf{u}_0 \in B(\mathbf{u})$, *la suite* $\mathbf{u}_n$ *définie par :*

$$\mathbf{u}_{n+1} = \mathbf{u}_n - \nabla^2 f(\mathbf{u}_n)^{-1} \nabla f(\mathbf{u}_n)$$

converge géométriquement vers le minimum $\mathbf{u}$.

Copyright : Jean Philippe Préaux

Remarques.

Copyright : Jean-Philippe Préaux

Remarques.

– Si f est de classe C^3 la convergence est même quadratique !

Copyright : Jean-Philippe Préaux

Remarques.

– Si f est de classe C^3 la convergence est même quadratique !

– Il s'agit d'une *méthode locale* dans le sens où il faut être suffisamment proche d'un extremum (que justement l'on cherche) pour converger. On peut raffiner le résultat pour majorer cette distance.

Copyright : Jean-Philippe Préaux

Remarques.

– Si f est de classe C^3 la convergence est même quadratique !

– Il s'agit d'une *méthode locale* dans le sens où il faut être suffisamment proche d'un extremum (que justement l'on cherche) pour converger. On peut raffiner le résultat pour majorer cette distance.

– Notons que si la méthode de Newton a pour désavantage de ne converger que localement, elle a par contre l'avantage d'être la seule méthode de ce chapitre qui ne nécessite aucune hypothèse de convexité ; elle a donc un champ d'applications très large.

Copyright : Jean-Philippe Préaux

Méthode de relaxation.

Copyright : Jean-Philippe Préaux

Méthode de relaxation.

Dans une méthode de descente, on construit une suite $(\mathbf{u}_n)_n$ en choisissant en chaque point $\mathbf{u}_n$ une direction de descente $\mathbf{d}_n$ et un pas de descente ρ_n.

Copyright : Jean-Philippe Préaux

Méthode de relaxation.

Dans une méthode de descente, on construit une suite $(\mathbf{u}_n)_n$ en choisissant en chaque point $\mathbf{u}_n$ une direction de descente $\mathbf{d}_n$ et un pas de descente ρ_n.
Une méthode pour construire ρ_n peut consister à se ramener à un problème d'optimisation en dimension 1 par :

$$f(\mathbf{u}_n + \rho_n \mathbf{d}_n) = \inf_{\rho \in \mathbb{R}} f(\mathbf{u}_n + \rho \mathbf{d}_n)$$

Copyright : Jean-Philippe Préaux

Méthode de relaxation.

Dans une méthode de descente, on construit une suite $(\mathbf{u}_n)_n$ en choisissant en chaque point $\mathbf{u}_n$ une direction de descente $\mathbf{d}_n$ et un pas de descente ρ_n.
Une méthode pour construire ρ_n peut consister à se ramener à un problème d'optimisation en dimension 1 par :

$$f(\mathbf{u}_n + \rho_n \mathbf{d}_n) = \inf_{\rho \in \mathbb{R}} f(\mathbf{u}_n + \rho \mathbf{d}_n)$$

La méthode de relaxation consiste en une telle méthode, où l'on prend pour direction de descente successivement chacun des axes.

Copyright : Jean-Philippe Préaux

Plus formellement la suite $(\mathbf{u}_k)_{k\in\mathbb{N}}$ est construite ainsi : $\mathbf{u}_0$ est choisi arbitrairement, en pratique, si possible, proche d'un minimum, et si

$$\mathbf{u}_k = (x_1^{(k)}, x_2^{(k)}, \ldots, x_n^{(k)})$$

$\mathbf{u}_{k+1} = (x_1^{(k+1)}, x_2^{(k+1)}, \ldots, x_n^{(k+1)})$ est construit par :

Copyright : Jean-Philippe Préaux

Plus formellement la suite $(\mathbf{u}_k)_{k\in\mathbb{N}}$ est construite ainsi : $\mathbf{u}_0$ est choisi arbitrairement, en pratique, si possible, proche d'un minimum, et si

$$\mathbf{u}_k = (x_1^{(k)}, x_2^{(k)}, \ldots, x_n^{(k)})$$

$\mathbf{u}_{k+1} = (x_1^{(k+1)}, x_2^{(k+1)}, \ldots, x_n^{(k+1)})$ est construit par :

$$f(x_1^{(k+1)}, x_2^{(k)}, \ldots, x_n^{(k)}) = \inf_{x\in\mathbb{R}} f(x, x_2^{(k)}, \ldots, x_n^{(k)})$$

$$f(x_1^{(k+1)}, x_2^{(k+1)}, \ldots, x_n^{(k)}) = \inf_{x\in\mathbb{R}} f(x_1^{(k+1)}, x, \ldots, x_n^{(k)})$$

$$\vdots$$

$$f(x_1^{(k+1)}, \ldots, x_{n-1}^{(k+1)}, x_n^{(k+1)}) = \inf_{x\in\mathbb{R}} f(x_1^{(k+1)}, \ldots, x_{n-1}^{(k+1)}, x)$$

Copyright : Jean-Philippe Préaux

Plus formellement la suite $(\mathbf{u}_k)_{k\in\mathbb{N}}$ est construite ainsi : $\mathbf{u}_0$ est choisi arbitrairement, en pratique, si possible, proche d'un minimum, et si

$$\mathbf{u}_k = (x_1^{(k)}, x_2^{(k)}, \ldots, x_n^{(k)})$$

$\mathbf{u}_{k+1} = (x_1^{(k+1)}, x_2^{(k+1)}, \ldots, x_n^{(k+1)})$ est construit par :

$$f(x_1^{(k+1)}, x_2^{(k)}, \ldots, x_n^{(k)}) = \inf_{x\in\mathbb{R}} f(x, x_2^{(k)}, \ldots, x_n^{(k)})$$

$$f(x_1^{(k+1)}, x_2^{(k+1)}, \ldots, x_n^{(k)}) = \inf_{x\in\mathbb{R}} f(x_1^{(k+1)}, x, \ldots, x_n^{(k)})$$

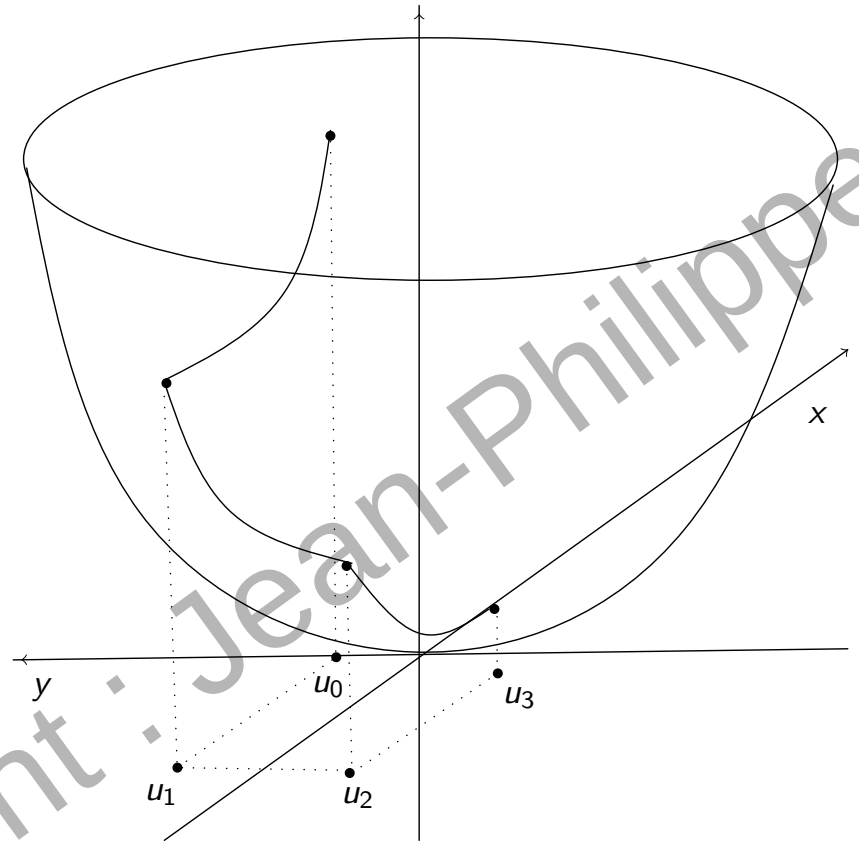
$$\vdots$$

$$f(x_1^{(k+1)}, \ldots, x_{n-1}^{(k+1)}, x_n^{(k+1)}) = \inf_{x\in\mathbb{R}} f(x_1^{(k+1)}, \ldots, x_{n-1}^{(k+1)}, x)$$

Théorème (Convergence de la méthode de relaxation.)

Si $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ est elliptique la méthode de relaxation converge vers son unique minimum.

Copyright : Jean-Philippe Préaux



Dans la méthode de relaxation on prend comme direction de descente successivement les directions des axes et on détermine le pas en résolvant un problème de minimisation à une variable.

Copyright : Jean-Philippe Préaux

Méthode de gradient.

Copyright : Jean-Philippe Préaux

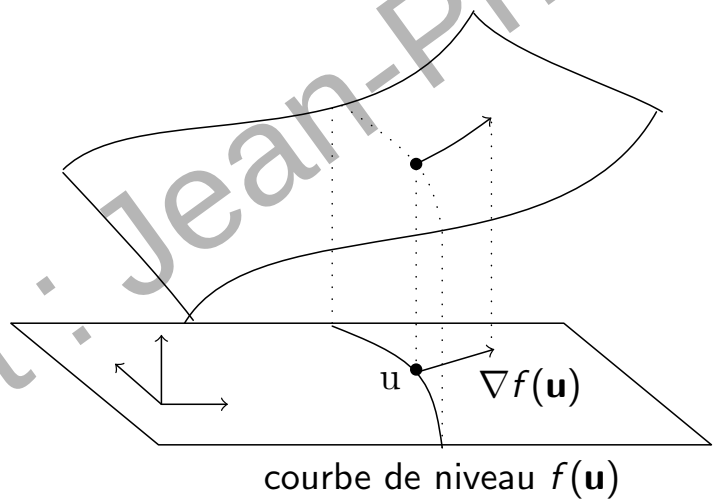
Méthode de gradient.

Une méthode de type gradient est une méthode de descente où la direction choisie en chaque point $\mathbf{x}$ est celle de plus grande pente, c'est à dire : $-\nabla f(\mathbf{x})$.

Copyright : Jean-Philippe Préaux

Méthode de gradient.

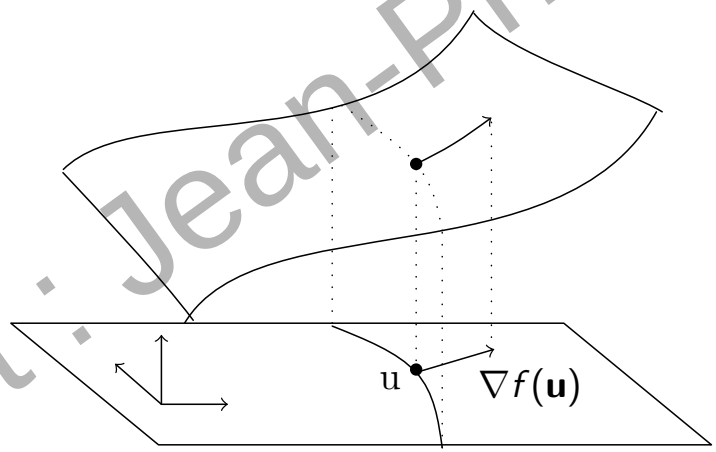
Une méthode de type gradient est une méthode de descente où la direction choisie en chaque point $\mathbf{x}$ est celle de plus grande pente, c'est à dire : $-\nabla f(\mathbf{x})$.



Copyright : Jean-Philippe Préaux

Méthode de gradient.

Une méthode de type gradient est une méthode de descente où la direction choisie en chaque point $\mathbf{x}$ est celle de plus grande pente, c'est à dire : $-\nabla f(\mathbf{x})$.



On construit par récurrence une suite de points $(\mathbf{u}_k)_{k\in\mathbb{N}}$, par la formule : $\boxed{\mathbf{u}_{k+1} = \mathbf{u}_k - \rho_k \nabla f(\mathbf{u}_k)}$

Copyright : Jean-Philippe Préaux

Méthode du gradient à pas optimal.

Copyright : Jean-Philippe Préaux

Méthode du gradient à pas optimal.

La *méthode du gradient à pas optimal* détermine à chaque itération le pas ρ_k par :

$$f(u_k - \rho_k \nabla f(\mathbf{u}_k)) = \inf_{\rho \in \mathbb{R}} (\mathbf{u}_k - \rho \nabla f(\mathbf{u}_k))$$

c'est à dire en se ramenant à un problème à une seule variable.

Copyright : Jean-Philippe Préaux

Méthode du gradient à pas optimal.

La *méthode du gradient à pas optimal* détermine à chaque itération le pas ρ_k par :

$$f(u_k - \rho_k \nabla f(\mathbf{u}_k)) = \inf_{\rho \in \mathbb{R}} (\mathbf{u}_k - \rho \nabla f(\mathbf{u}_k))$$

c'est à dire en se ramenant à un problème à une seule variable.

Théorème (Convergence de la méthode du gradient à pas optimal.)

Si $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ est α-elliptique, la méthode du gradient à pas optimal converge vers l'unique minimum de f. L'erreur à l'étape k est majorée par :

$$\|\mathbf{r}_k\| = \|\mathbf{u}_k - \mathbf{u}\| \leqslant \frac{1}{\alpha} \|\nabla f(\mathbf{u}_k)\| \ .$$

Copyright : Jean-Philippe Préaux

Remarque. Un point essentiel de la preuve réside dans le fait que $\nabla f(\mathbf{u}_k)$ et $\nabla f(\mathbf{u}_{k+1})$ sont orthogonaux.

Copyright : Jean-Philippe Préaux

Remarque. Un point essentiel de la preuve réside dans le fait que $\nabla f(\mathbf{u}_k)$ et $\nabla f(\mathbf{u}_{k+1})$ sont orthogonaux. On peut mettre à profit cela pour s'abstenir de résoudre à chaque étape un problème d'optimisation à une variable dans le cas d'une fonction quadratique elliptique.

Copyright : Jean-Philippe Préaux

Remarque. Un point essentiel de la preuve réside dans le fait que $\nabla f(\mathbf{u}_k)$ et $\nabla f(\mathbf{u}_{k+1})$ sont orthogonaux. On peut mettre à profit cela pour s'abstenir de résoudre à chaque étape un problème d'optimisation à une variable dans le cas d'une fonction quadratique elliptique.

Soit $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top A\mathbf{x} - \mathbf{b}^\top \mathbf{x} + c$ une fonction quadratique elliptique.

On peut ici donner une formule explicite pour le pas optimal ρ_k.

Copyright : Jean-Philippe Préaux

Remarque. Un point essentiel de la preuve réside dans le fait que $\nabla f(\mathbf{u}_k)$ et $\nabla f(\mathbf{u}_{k+1})$ sont orthogonaux. On peut mettre à profit cela pour s'abstenir de résoudre à chaque étape un problème d'optimisation à une variable dans le cas d'une fonction quadratique elliptique.
Soit $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top A\mathbf{x} - \mathbf{b}^\top \mathbf{x} + c$ une fonction quadratique elliptique.
On peut ici donner une formule explicite pour le pas optimal ρ_k.

Théorème (Pas optimal en programmation quadratique elliptique.)

Dans le cas de la fonction quadratique elliptique $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top A\mathbf{x} - \mathbf{b}^\top \mathbf{x} + c$, le pas optimal ρ_k est donné par :

$$\rho_k = \frac{\|A\mathbf{u}_k - \mathbf{b}\|^2}{\langle A(A\mathbf{u}_k - \mathbf{b}), A\mathbf{u}_k - \mathbf{b}\rangle} = \frac{\|\nabla f(\mathbf{u}_k)\|^2}{\langle A\nabla f(\mathbf{u}_k), \nabla f(\mathbf{u}_k)\rangle}.$$

Copyright : Jean-Philippe Préaux

Preuve.

Copyright : Jean-Philippe Préaux

Preuve.

Puisque :

$$\begin{aligned}\langle \nabla f(\mathbf{u}_{k+1}), \nabla f(\mathbf{u}_k)\rangle &= 0 \\ &= \langle A(\mathbf{u}_k - \rho_k(A\mathbf{u}_k - b)) - b, A\mathbf{u}_k - b\rangle.\end{aligned}$$

Copyright : Jean-Philippe Préaux

Preuve.

Puisque :

$$\begin{aligned}\langle \nabla f(\mathbf{u}_{k+1}), \nabla f(\mathbf{u}_k)\rangle &= 0 \\ &= \langle A(\mathbf{u}_k - \rho_k(A\mathbf{u}_k - b)) - b, A\mathbf{u}_k - b\rangle.\end{aligned}$$

Par bilinéarité du produit scalaire :

$$\begin{aligned}\Longrightarrow \quad & \rho_k\langle A(A\mathbf{u}_k - b), A\mathbf{u}_k - b\rangle = \langle A\mathbf{u}_k - b, A\mathbf{u}_k - b\rangle \\ \Longrightarrow \quad & \rho_k = \frac{\|A\mathbf{u}_k - b\|^2}{\langle A(A\mathbf{u}_k - b), A\mathbf{u}_k - b\rangle} = \frac{\|\mathbf{d}_k\|^2}{\langle A\mathbf{d}_k, \mathbf{d}_k\rangle}\ .\end{aligned}$$

□

Copyright : Jean-Philippe Préaux

Méthode du gradient à pas fixe

Les méthodes de relaxation et de gradient à pas optimal ont en commun la recherche à chaque pas d'un pas de descente optimal, en se ramenant à un problème uni-dimensionnel.

Copyright : Jean-Philippe Préaux

Méthode du gradient à pas fixe

Les méthodes de relaxation et de gradient à pas optimal ont en commun la recherche à chaque pas d'un pas de descente optimal, en se ramenant à un problème uni-dimensionnel.
C'est pour s'abstraire de cette recherche du pas qu'on développe la méthode du gradient à pas fixe.

Copyright : Jean-Philippe Préaux

Méthode du gradient à pas fixe

Les méthodes de relaxation et de gradient à pas optimal ont en commun la recherche à chaque pas d'un pas de descente optimal, en se ramenant à un problème uni-dimensionnel.
C'est pour s'abstraire de cette recherche du pas qu'on développe la méthode du gradient à pas fixe.
Il s'agit d'une méthode de gradient où le pas de descente est fixé à $\rho > 0$:

$$\mathbf{u}_{k+1} = \mathbf{u}_k - \rho \nabla f(\mathbf{u}_k) \ .$$

Copyright : Jean-Philippe Préaux

Méthode du gradient à pas fixe

Les méthodes de relaxation et de gradient à pas optimal ont en commun la recherche à chaque pas d'un pas de descente optimal, en se ramenant à un problème uni-dimensionnel.
C'est pour s'abstraire de cette recherche du pas qu'on développe la méthode du gradient à pas fixe.
Il s'agit d'une méthode de gradient où le pas de descente est fixé à $\rho > 0$:

$$\mathbf{u}_{k+1} = \mathbf{u}_k - \rho \nabla f(\mathbf{u}_k) \ .$$

Sous des hypothèses suffisantes, on peut choisir le pas pour s'assurer de la convergence.

Copyright : Jean-Philippe Préaux

Théorème (Convergence de la méthode du gradient à pas fixe.)

Copyright : Jean-Philippe Préaux

Théorème (Convergence de la méthode du gradient à pas fixe.)

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *une application* α*-elliptique dont la différentielle est lipschitzienne, c'est à dire qu'il existe* $M > 0$ *telle que* $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leqslant M\|\mathbf{x} - \mathbf{y}\| \ .$$

Copyright : Jean-Philippe Préaux

Théorème (Convergence de la méthode du gradient à pas fixe.)

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une application α-elliptique dont la différentielle est lipschitzienne, c'est à dire qu'il existe $M > 0$ telle que $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leqslant M\|\mathbf{x} - \mathbf{y}\| \ .$$

Si le pas ρ est choisi tel que :

$$0 < \rho < \frac{2\alpha}{M^2}$$

alors la méthode du gradient à pas fixe converge géométriquement vers l'unique minimum global de f.

Copyright : Jean-Philippe Préaux

Remarques.

Copyright : Jean-Philippe Préaux

Remarques. – Lorsque f est deux fois différentiable les hypothèses du théorème reviennent à l'existence de deux réels strictement positifs $\alpha \leqslant M$ tels que pour tout $\mathbf{x} \in \mathbb{R}^n$, toutes les valeurs propres de $\nabla^2 f(\mathbf{x})$ sont dans $[\alpha, M]$.

Copyright : Jean-Philippe Préaux

Remarques. – Lorsque f est deux fois différentiable les hypothèses du théorème reviennent à l'existence de deux réels strictement positifs $\alpha \leqslant M$ tels que pour tout $\mathbf{x} \in \mathbb{R}^n$, toutes les valeurs propres de $\nabla^2 f(\mathbf{x})$ sont dans $[\alpha, M]$.

– En général le meilleur pas de descente donné par la preuve est $\rho = \alpha/M^2 = \min_\rho (1 - 2\alpha\rho + M^2\rho^2)$.

Copyright : Jean-Philippe Préaux

Remarques. – Lorsque f est deux fois différentiable les hypothèses du théorème reviennent à l'existence de deux réels strictement positifs $\alpha \leqslant M$ tels que pour tout $\mathbf{x} \in \mathbb{R}^n$, toutes les valeurs propres de $\nabla^2 f(\mathbf{x})$ sont dans $[\alpha, M]$.

– En général le meilleur pas de descente donné par la preuve est $\rho = \alpha/M^2 = \min_\rho (1 - 2\alpha\rho + M^2\rho^2)$.

– Pour $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top A\mathbf{x} - \mathbf{b}^\top \mathbf{x}$ une fonction quadratique elliptique, α et M sont respectivement donnés par la plus petite (la plus grande) valeur propre de A. On peut vérifier que dans ce cas le meilleur pas de descente est $\frac{2}{\lambda_1+\lambda_n}$ où λ_1, λ_n désignent la plus petite et la plus grande valeur propre de A.

Copyright : Jean-Philippe Préaux

Remarques. – Lorsque f est deux fois différentiable les hypothèses du théorème reviennent à l'existence de deux réels strictement positifs $\alpha \leqslant M$ tels que pour tout $\mathbf{x} \in \mathbb{R}^n$, toutes les valeurs propres de $\nabla^2 f(\mathbf{x})$ sont dans $[\alpha, M]$.

– En général le meilleur pas de descente donné par la preuve est $\rho = \alpha/M^2 = \min_\rho (1 - 2\alpha\rho + M^2\rho^2)$.

– Pour $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top A\mathbf{x} - \mathbf{b}^\top \mathbf{x}$ une fonction quadratique elliptique, α et M sont respectivement donnés par la plus petite (la plus grande) valeur propre de A. On peut vérifier que dans ce cas le meilleur pas de descente est $\frac{2}{\lambda_1+\lambda_n}$ où λ_1, λ_n désignent la plus petite et la plus grande valeur propre de A.

– Il faut noter que contrairement à la méthode de relaxation ou du gradient à pas optimal on peut avoir $f(\mathbf{u}_{k+1}) > f(\mathbf{u}_k)$.

Copyright : Jean-Philippe Préaux

Méthode du gradient conjugué.

Copyright : Jean-Philippe Préaux

Méthode du gradient conjugué.

Même si la direction opposée au gradient est localement la direction de plus grande descente locale, ce n'est pas en appliquant une méthode de descente du type gradient que l'on converge le plus rapidement vers un minimum. Et ce n'est pas ce que l'on peut faire de mieux à l'ordre 1.

Copyright : Jean-Philippe Préaux

Méthode du gradient conjugué.

Même si la direction opposée au gradient est localement la direction de plus grande descente locale, ce n'est pas en appliquant une méthode de descente du type gradient que l'on converge le plus rapidement vers un minimum. Et ce n'est pas ce que l'on peut faire de mieux à l'ordre 1.

L'idée de la méthode du gradient conjugué est de construire $\mathbf{u}_{k+1}$ comme le minimum de la fonction sur l'espace affine $\mathbf{u}_k + < \mathbf{d}_0, \ldots, \mathbf{d}_k >$. Dans le cas d'une fonction quadratique elliptique la méthode converge en au plus n-itérations : c'est une méthode exacte particulièrement rapide.

Copyright : Jean Philippe Préaux

Méthode du gradient conjugué.

Même si la direction opposée au gradient est localement la direction de plus grande descente locale, ce n'est pas en appliquant une méthode de descente du type gradient que l'on converge le plus rapidement vers un minimum. Et ce n'est pas ce que l'on peut faire de mieux à l'ordre 1.

L'idée de la méthode du gradient conjugué est de construire $\mathbf{u}_{k+1}$ comme le minimum de la fonction sur l'espace affine $\mathbf{u}_k + < \mathbf{d}_0, \ldots, \mathbf{d}_k >$. Dans le cas d'une fonction quadratique elliptique la méthode converge en au plus n-itérations : c'est une méthode exacte particulièrement rapide.

Cette méthode ingénieuse prend en compte la géométrie globale de la nappe représentative de la fonction. Cette propriété n'est plus vérifiée pour une fonction quelconque.

Copyright : Jean Philippe Préaux

Méthode du gradient conjugué.

Même si la direction opposée au gradient est localement la direction de plus grande descente locale, ce n'est pas en appliquant une méthode de descente du type gradient que l'on converge le plus rapidement vers un minimum. Et ce n'est pas ce que l'on peut faire de mieux à l'ordre 1.

L'idée de la méthode du gradient conjugué est de construire $\mathbf{u}_{k+1}$ comme le minimum de la fonction sur l'espace affine $\mathbf{u}_k + < \mathbf{d}_0, \ldots, \mathbf{d}_k >$. Dans le cas d'une fonction quadratique elliptique la méthode converge en au plus n-itérations : c'est une méthode exacte particulièrement rapide.

Cette méthode ingénieuse prend en compte la géométrie globale de la nappe représentative de la fonction. Cette propriété n'est plus vérifiée pour une fonction quelconque.

Elle se généralise cependant à des fonctions non quadratiques par des méthodes telles que *Fletcher-Reeves* ou *Polak-Ribière*.

Copyright : Jean Philippe Préaux

Soit f une fonction quadratique elliptique, $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top A\mathbf{x} - \mathbf{b}^\top\mathbf{x} + c$. La méthode du gradient conjugué est la méthode de descente définie par :

Copyright : Jean-Philippe Préaux

Soit f une fonction quadratique elliptique, $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top A\mathbf{x} - \mathbf{b}^\top\mathbf{x} + c$. La méthode du gradient conjugué est la méthode de descente définie par :

- **Etape 1 :**

$$\mathbf{d}_0 = \nabla f(\mathbf{u}_0) = A\mathbf{u}_0 - \mathbf{b}$$

$$\rho_0 = \frac{\|\mathbf{d}_0\|^2}{\langle A\mathbf{d}_0, \mathbf{d}_0\rangle}$$

$$\mathbf{u}_1 = \mathbf{u}_0 - \rho_0\mathbf{d}_0$$

Copyright : Jean-Philippe Préaux

Soit f une fonction quadratique elliptique, $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top A\mathbf{x} - \mathbf{b}^\top\mathbf{x} + c$. La méthode du gradient conjugué est la méthode de descente définie par :

- **Etape 1 :**

$$\mathbf{d}_0 = \nabla f(\mathbf{u}_0) = A\mathbf{u}_0 - \mathbf{b}$$

$$\rho_0 = \frac{\|\mathbf{d}_0\|^2}{\langle A\mathbf{d}_0, \mathbf{d}_0\rangle}$$

$$\mathbf{u}_1 = \mathbf{u}_0 - \rho_0\mathbf{d}_0$$

- **Etape $k+1$:**

$$\mathbf{d}_k = \nabla f(\mathbf{u}_k) + \frac{\|\nabla f(\mathbf{u}_k)\|^2}{\|\nabla f(\mathbf{u}_{k-1})\|^2}\mathbf{d}_{k-1} \quad \text{soit :} \quad \mathbf{d}_k = A\mathbf{u}_k - \mathbf{b} + \frac{\|A\mathbf{u}_k - \mathbf{b}\|^2}{\|A\mathbf{u}_{k-1} - \mathbf{b}\|^2}\mathbf{d}_{k-1}$$

$$\rho_k = \frac{\langle \nabla f(\mathbf{u}_k), \mathbf{d}_k\rangle}{\langle A\mathbf{d}_k, \mathbf{d}_k\rangle} \qquad \rho_k = \frac{\langle A\mathbf{u}_k - \mathbf{b}, \mathbf{d}_k\rangle}{\langle A\mathbf{d}_k, \mathbf{d}_k\rangle}$$

$$\mathbf{u}_{k+1} = \mathbf{u}_k - \rho_k\mathbf{d}_k$$

Et ce tant que $\nabla f(\mathbf{u}_k) = A\mathbf{u}_k - \mathbf{b} \neq \mathbf{0}$.

Copyright : Jean-Philippe Préaux

Théorème (Convergence de la méthode du gradient conjugué.)

La méthode du gradient conjugué appliquée à une fonction quadratique elliptique de $\mathbb{R}^n$ converge en au plus n itérations.

Copyright : Jean-Philippe Préaux

Théorème (Convergence de la méthode du gradient conjugué.)

La méthode du gradient conjugué appliquée à une fonction quadratique elliptique de $\mathbb{R}^n$ converge en au plus n itérations.

Esquisse de preuve. Le point essentiel réside dans le fait que dans ce cas la famille des directions successives $\mathbf{d}_0, \ldots, \mathbf{d}_k$ est orthogonale pour le produit scalaire

$$(\mathbf{x}, \mathbf{y}) \longrightarrow \mathbf{x}^\top A\mathbf{y}$$

associé à la matrice hessienne A de la fonction quadratique elliptique.

Copyright : Jean-Philippe Préaux

Copyright : Jean-Philippe Préaux

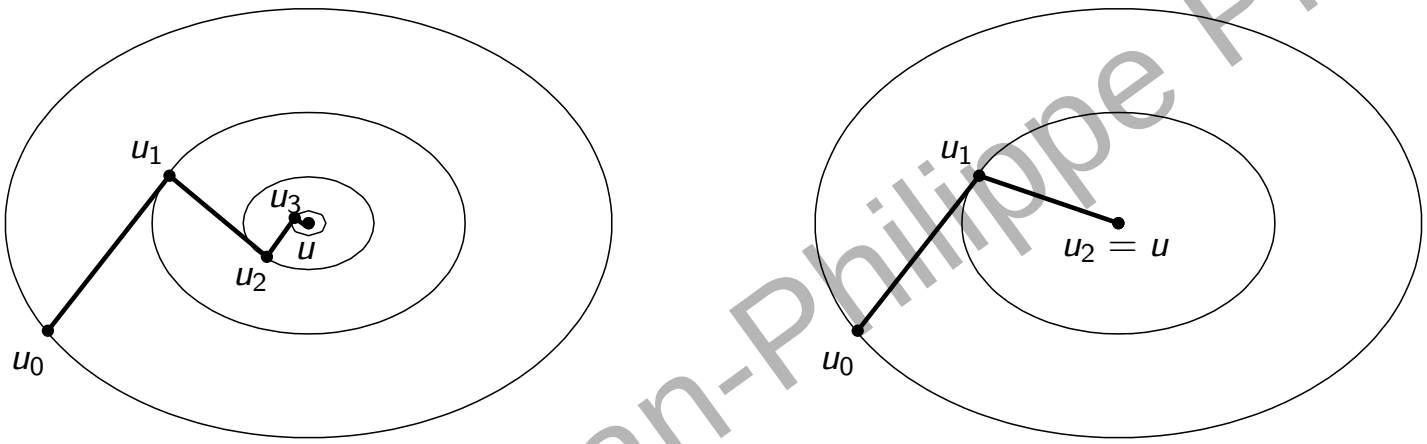
Théorème (Convergence de la méthode du gradient conjugué.)

La méthode du gradient conjugué appliquée à une fonction quadratique elliptique de $\mathbb{R}^n$ converge en au plus n itérations.

Esquisse de preuve. Le point essentiel réside dans le fait que dans ce cas la famille des directions successives $\mathbf{d}_0, \ldots, \mathbf{d}_k$ est orthogonale pour le produit scalaire

$$(\mathbf{x}, \mathbf{y}) \longrightarrow \mathbf{x}^\top A \mathbf{y}$$

associé à la matrice hessienne A de la fonction quadratique elliptique. En particulier si $\mathbf{d}_0, \ldots, \mathbf{d}_k \neq \mathbf{0}$ c'est une famille libre dans un espace vectoriel de dimension finie n. □



Comparaison de la méthode du gradient à pas optimal (à gauche) et de la méthode du gradient conjugué (à droite) pour minimiser $f(x, y) = \frac{x^2}{a^2} + \frac{y^2}{b^2}$. Lorsque $a^2 \neq b^2$ la méthode du gradient conjugué converge en deux étapes, tandis que la méthode du gradient à pas optimal ne converge pas en un nombre fini d'étapes.

Copyright : Jean-Philippe Préaux

Exemple

Minimisation de la forme quadratique :

$$f(x,y,z) = x^2 + y^2 + z^2 + xy + xz + yz + x - y + 3z \ .$$

Le tableau qui suit permet de comparer quatre méthodes (gradient conjugué, gradient optimal, gradient à pas fixe, relaxation) de minimisation de l'application quadratique f.
Son minimum est $(0.25, -1.75, 2.25)$ en lequel la fonction vaut -4.375.

Copyright : Jean-Philippe Préaux

Itérations	Gradient conjugué	Gradient optimal	Gradient fixe	Relaxation
point initial	$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$
1	$\begin{pmatrix} -0.5454 \\ -0.3181 \\ 1.4545 \end{pmatrix}$	$\begin{pmatrix} -0.5454 \\ -0.3181 \\ 1.4545 \end{pmatrix}$	$\begin{pmatrix} -1.4000 \\ -1.6000 \\ 0.6000 \end{pmatrix}$	$\begin{pmatrix} -2.0000 \\ -1.0000 \\ 3.0000 \end{pmatrix}$
2	$\begin{pmatrix} 0.2500 \\ -1.7500 \\ 2.2500 \end{pmatrix}$	$\begin{pmatrix} 0.3086 \\ -1.4569 \\ 2.3086 \end{pmatrix}$	$\begin{pmatrix} 0.5200 \\ -0.4000 \\ 2.5200 \end{pmatrix}$	$\begin{pmatrix} -0.5000 \\ -1.7500 \\ 2.6250 \end{pmatrix}$
3		$\begin{pmatrix} 0.1878 \\ -1.6381 \\ 2.1878 \end{pmatrix}$	$\begin{pmatrix} -0.3440 \\ -1.6960 \\ 1.6560 \end{pmatrix}$	$\begin{pmatrix} 0.0625 \\ -1.8437 \\ 2.3906 \end{pmatrix}$
5		$\begin{pmatrix} 0.2451 \\ -1.7412 \\ 2.2451 \end{pmatrix}$	$\begin{pmatrix} 0.0361 \\ -1.7305 \\ 2.0361 \end{pmatrix}$	$\begin{pmatrix} 0.2587 \\ -1.7749 \\ 2.2580 \end{pmatrix}$
10		$\begin{pmatrix} 0.2500 \\ -1.7499 \\ 2.2500 \end{pmatrix}$	$\begin{pmatrix} 0.2545 \\ -1.7273 \\ 2.2545 \end{pmatrix}$	$\begin{pmatrix} 0.2500 \\ -1.7499 \\ 2.2499 \end{pmatrix}$
20		$\begin{pmatrix} 0.2500 \\ -1.7499 \\ 2.2500 \end{pmatrix}$	$\begin{pmatrix} 0.2500 \\ -1.7498 \\ 2.2500 \end{pmatrix}$	$\begin{pmatrix} 0.2500 \\ -1.7500 \\ 2.2499 \end{pmatrix}$
30		$\begin{pmatrix} 0.2500 \\ -1.7500 \\ 0.2500 \end{pmatrix}$	$\begin{pmatrix} 0.2500 \\ -1.7499 \\ 2.2500 \end{pmatrix}$	$\begin{pmatrix} 0.2499 \\ -1.7500 \\ 2.2500 \end{pmatrix}$
50			$\begin{pmatrix} 0.2500 \\ -1.7499 \\ 2.2500 \end{pmatrix}$	$\begin{pmatrix} 0.2500 \\ -1.7500 \\ 2.2500 \end{pmatrix}$
80			$\begin{pmatrix} 0.2500 \\ -1.7500 \\ 2.2500 \end{pmatrix}$	

Copyright : Jean-Philippe Préaux

Algorithmes itératifs dans le cas sous contraintes.

Dans le cas sous contraintes de domaine convexe fermé, on établit des méthodes itératives en appliquant les méthodes sans contraintes vues précédemment tout en projetant à chaque itération le point obtenu sur le domaine.

Copyright : Jean-Philippe Préaux

Algorithmes itératifs dans le cas sous contraintes.

Dans le cas sous contraintes de domaine convexe fermé, on établit des méthodes itératives en appliquant les méthodes sans contraintes vues précédemment tout en projetant à chaque itération le point obtenu sur le domaine.

On utilise pour cela le *théorème de projection convexe* que nous rappelons ci-après. C'est un grand classique dont on peut trouver une preuve dans l'exercice 4 du chapitre II.

Copyright : Jean-Philippe Préaux

Théorème de projection convexe.

Théorème (Théorème de projection convexe.)

Soit $\mathcal{C}$ un sous-ensemble non vide fermé, convexe de $\mathbb{R}^n$.

Copyright : Jean-Philippe Préaux

Théorème de projection convexe.

Théorème (Théorème de projection convexe.)

Soit $\mathcal{C}$ un sous-ensemble non vide fermé, convexe de $\mathbb{R}^n$.
Donné $\mathbf{u} \in \mathbb{R}^n$ il existe un unique $P_{\mathcal{C}}(\mathbf{u}) \in \mathcal{C}$ tel que

$$\|\mathbf{u} - P_{\mathcal{C}}(\mathbf{u})\| = \inf_{\mathbf{v} \in \mathcal{D}} \|\mathbf{u} - \mathbf{v}\| ,$$

et $P_{\mathcal{C}}(\mathbf{u})$ est caractérisé par l'inégalité :

$$\forall \mathbf{v} \in \mathcal{C}, \quad \langle P_{\mathcal{C}}(\mathbf{u}) - \mathbf{u}, \mathbf{v} - P_{\mathcal{C}}(\mathbf{u}) \rangle \geqslant 0 .$$

Copyright : Jean-Philippe Préaux

Théorème de projection convexe.

Théorème (Théorème de projection convexe.)

Soit $\mathcal{C}$ un sous-ensemble non vide fermé, convexe de $\mathbb{R}^n$.
Donné $\mathbf{u} \in \mathbb{R}^n$ il existe un unique $P_{\mathcal{C}}(\mathbf{u}) \in \mathcal{C}$ tel que

$$\|\mathbf{u} - P_{\mathcal{C}}(\mathbf{u})\| = \inf_{\mathbf{v} \in \mathcal{D}} \|\mathbf{u} - \mathbf{v}\| \ ,$$

et $P_{\mathcal{C}}(\mathbf{u})$ est caractérisé par l'inégalité :

$$\forall \mathbf{v} \in \mathcal{C}, \quad \langle P_{\mathcal{C}}(\mathbf{u}) - \mathbf{u}, \mathbf{v} - P_{\mathcal{C}}(\mathbf{u}) \rangle \geqslant 0 \ .$$

L'application $P_{\mathcal{C}} : \mathbb{R}^n \longrightarrow \mathcal{C}$ ainsi définie est appelée l'opérateur de projection sur $\mathcal{C}$. C'est une application contractante, i.e. :

$$\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \quad \|P_{\mathcal{C}}(\mathbf{x}) - P_{\mathcal{C}}(\mathbf{y})\| \leqslant \|\mathbf{x} - \mathbf{y}\| \ .$$

Copyright : Jean-Philippe Préaux

Preuve.

Copyright : Jean-Philippe Préaux

Preuve. Le problème de minimisation $\min_{\mathbf{v}\in\mathcal{C}} \|\mathbf{u}-\mathbf{v}\|$ est équivalent au problème $\min_{\mathbf{v}\in\mathcal{C}} \|\mathbf{u}-\mathbf{v}\|^2$.

Copyright : Jean-Philippe Préaux

Preuve. Le problème de minimisation $\min_{\mathbf{v}\in\mathcal{C}} \|\mathbf{u}-\mathbf{v}\|$ est équivalent au problème $\min_{\mathbf{v}\in\mathcal{C}} \|\mathbf{u}-\mathbf{v}\|^2$. Or l'application

$$f : \mathbf{x} \mapsto \|\mathbf{u}-\mathbf{x}\|^2 = \sum_{i=1}^{n} (u_i - x_i)^2 = \mathbf{x}^\top \mathrm{Id}\, \mathbf{x} - 2\mathbf{u}^\top \mathbf{x} + \|\mathbf{u}\|^2$$

est une application quadratique de matrice hessienne $2\,\mathrm{Id}$.

Copyright : Jean-Philippe Préaux

Preuve. Le problème de minimisation $\min_{\mathbf{v}\in\mathcal{C}} \|\mathbf{u}-\mathbf{v}\|$ est équivalent au problème $\min_{\mathbf{v}\in\mathcal{C}} \|\mathbf{u}-\mathbf{v}\|^2$. Or l'application

$$f : \mathbf{x} \mapsto \|\mathbf{u}-\mathbf{x}\|^2 = \sum_{i=1}^{n} (u_i - x_i)^2 = \mathbf{x}^\top \mathrm{Id}\, \mathbf{x} - 2\mathbf{u}^\top \mathbf{x} + \|\mathbf{u}\|^2$$

est une application quadratique de matrice hessienne $2\,\mathrm{Id}$. Ainsi f est une application elliptique, et donc strictement convexe et coercive.

Copyright : Jean-Philippe Préaux

Preuve. Le problème de minimisation $\min_{\mathbf{v}\in\mathcal{C}} \|\mathbf{u}-\mathbf{v}\|$ est équivalent au problème $\min_{\mathbf{v}\in\mathcal{C}} \|\mathbf{u}-\mathbf{v}\|^2$. Or l'application

$$f : \mathbf{x} \mapsto \|\mathbf{u}-\mathbf{x}\|^2 = \sum_{i=1}^{n} (u_i - x_i)^2 = \mathbf{x}^\top \mathrm{Id}\, \mathbf{x} - 2\mathbf{u}^\top \mathbf{x} + \|\mathbf{u}\|^2$$

est une application quadratique de matrice hessienne $2\,\mathrm{Id}$. Ainsi f est une application elliptique, et donc strictement convexe et coercive. Le domaine $\mathcal{C}$ étant convexe fermé et non vide elle y admet un unique minimum, $P_{\mathcal{C}}(\mathbf{u})$.

Copyright : Jean-Philippe Préaux

Preuve. Le problème de minimisation $\min_{\mathbf{v}\in\mathcal{C}} \|\mathbf{u}-\mathbf{v}\|$ est équivalent au problème $\min_{\mathbf{v}\in\mathcal{C}} \|\mathbf{u}-\mathbf{v}\|^2$. Or l'application

$$f : \mathbf{x} \mapsto \|\mathbf{u}-\mathbf{x}\|^2 = \sum_{i=1}^{n} (u_i - x_i)^2 = \mathbf{x}^\top \mathrm{Id}\, \mathbf{x} - 2\mathbf{u}^\top \mathbf{x} + \|\mathbf{u}\|^2$$

est une application quadratique de matrice hessienne $2\,\mathrm{Id}$. Ainsi f est une application elliptique, et donc strictement convexe et coercive. Le domaine $\mathcal{C}$ étant convexe fermé et non vide elle y admet un unique minimum, $P_{\mathcal{C}}(\mathbf{u})$.

On est dans le cadre de la programmation convexe, et f est différentiable. On peut caractériser $P_{\mathcal{C}}(\mathbf{u})$ par : $\forall\, \mathbf{v} \in \mathcal{C}$, $\langle \nabla f(P_{\mathcal{C}}(\mathbf{u})), \mathbf{v} - P_{\mathcal{C}}(\mathbf{u})\rangle \geqslant 0$.

Copyright - Jean-Philippe Préaux

Preuve. Le problème de minimisation $\min_{\mathbf{v}\in\mathcal{C}} \|\mathbf{u}-\mathbf{v}\|$ est équivalent au problème $\min_{\mathbf{v}\in\mathcal{C}} \|\mathbf{u}-\mathbf{v}\|^2$. Or l'application

$$f : \mathbf{x} \mapsto \|\mathbf{u}-\mathbf{x}\|^2 = \sum_{i=1}^{n} (u_i - x_i)^2 = \mathbf{x}^\top \mathrm{Id}\, \mathbf{x} - 2\mathbf{u}^\top \mathbf{x} + \|\mathbf{u}\|^2$$

est une application quadratique de matrice hessienne $2\,\mathrm{Id}$. Ainsi f est une application elliptique, et donc strictement convexe et coercive. Le domaine $\mathcal{C}$ étant convexe fermé et non vide elle y admet un unique minimum, $P_{\mathcal{C}}(\mathbf{u})$.

On est dans le cadre de la programmation convexe, et f est différentiable. On peut caractériser $P_{\mathcal{C}}(\mathbf{u})$ par : $\forall\, \mathbf{v} \in \mathcal{C}$, $\langle \nabla f(P_{\mathcal{C}}(\mathbf{u})), \mathbf{v} - P_{\mathcal{C}}(\mathbf{u})\rangle \geqslant 0$. Or $\nabla f(P_{\mathcal{C}}(\mathbf{u})) = 2\,(P_{\mathcal{C}}(\mathbf{u}) - \mathbf{u})$.

Copyright - Jean-Philippe Préaux

Preuve. Le problème de minimisation $\min_{\mathbf{v}\in\mathcal{C}} \|\mathbf{u}-\mathbf{v}\|$ est équivalent au problème $\min_{\mathbf{v}\in\mathcal{C}} \|\mathbf{u}-\mathbf{v}\|^2$. Or l'application

$$f : \mathbf{x} \mapsto \|\mathbf{u}-\mathbf{x}\|^2 = \sum_{i=1}^{n} (u_i - x_i)^2 = \mathbf{x}^\top \mathrm{Id}\, \mathbf{x} - 2\mathbf{u}^\top \mathbf{x} + \|\mathbf{u}\|^2$$

est une application quadratique de matrice hessienne $2\,\mathrm{Id}$. Ainsi f est une application elliptique, et donc strictement convexe et coercive. Le domaine $\mathcal{C}$ étant convexe fermé et non vide elle y admet un unique minimum, $P_{\mathcal{C}}(\mathbf{u})$.

On est dans le cadre de la programmation convexe, et f est différentiable. On peut caractériser $P_{\mathcal{C}}(\mathbf{u})$ par : $\forall\, \mathbf{v} \in \mathcal{C}$, $\langle \nabla f(P_{\mathcal{C}}(\mathbf{u})), \mathbf{v} - P_{\mathcal{C}}(\mathbf{u})\rangle \geqslant 0$. Or $\nabla f(P_{\mathcal{C}}(\mathbf{u})) = 2\,(P_{\mathcal{C}}(\mathbf{u}) - \mathbf{u})$. On obtient donc :

$$\forall\, \mathbf{v} \in \mathcal{C},\ 2\langle P_{\mathcal{C}}(\mathbf{u}) - \mathbf{u}, \mathbf{v} - P_{\mathcal{C}}(\mathbf{u})\rangle \geqslant 0.$$

On a la caractérisation recherchée.

Copyright - Jean-Philippe Préaux

Preuve. Le problème de minimisation $\min_{\mathbf{v}\in\mathcal{C}} \|\mathbf{u}-\mathbf{v}\|$ est équivalent au problème $\min_{\mathbf{v}\in\mathcal{C}} \|\mathbf{u}-\mathbf{v}\|^2$. Or l'application

$$f : \mathbf{x} \mapsto \|\mathbf{u}-\mathbf{x}\|^2 = \sum_{i=1}^{n} (u_i - x_i)^2 = \mathbf{x}^\top \mathrm{Id}\, \mathbf{x} - 2\mathbf{u}^\top \mathbf{x} + \|\mathbf{u}\|^2$$

est une application quadratique de matrice hessienne $2\,\mathrm{Id}$. Ainsi f est une application elliptique, et donc strictement convexe et coercive. Le domaine $\mathcal{C}$ étant convexe fermé et non vide elle y admet un unique minimum, $P_{\mathcal{C}}(\mathbf{u})$.

On est dans le cadre de la programmation convexe, et f est différentiable. On peut caractériser $P_{\mathcal{C}}(\mathbf{u})$ par : $\forall\, \mathbf{v} \in \mathcal{C}$, $\langle \nabla f(P_{\mathcal{C}}(\mathbf{u})), \mathbf{v} - P_{\mathcal{C}}(\mathbf{u})\rangle \geqslant 0$. Or $\nabla f(P_{\mathcal{C}}(\mathbf{u})) = 2\,(P_{\mathcal{C}}(\mathbf{u}) - \mathbf{u})$. On obtient donc :

$$\forall\, \mathbf{v} \in \mathcal{C},\ 2\langle P_{\mathcal{C}}(\mathbf{u}) - \mathbf{u}, \mathbf{v} - P_{\mathcal{C}}(\mathbf{u})\rangle \geqslant 0.$$

On a la caractérisation recherchée. (Voir exercice II.4 pour montrer que $P_{\mathcal{C}}$ est contractante.) □

Copyright - Jean-Philippe Préaux

Méthode de relaxation sur $\prod_{i=1}^{n}[a_i, b_i]$.

Considérons une application $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ elliptique que l'on souhaite minimiser sur un domaine de la forme :

$$\mathcal{D} = \prod_{i=1}^{n}[a_i, b_i] \qquad \text{où } a_i, b_i \in \overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}.$$

Copyright : Jean-Philippe Préaux

Méthode de relaxation sur $\prod_{i=1}^{n}[a_i, b_i]$.

Considérons une application $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ elliptique que l'on souhaite minimiser sur un domaine de la forme :

$$\mathcal{D} = \prod_{i=1}^{n}[a_i, b_i] \qquad \text{où } a_i, b_i \in \overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}.$$

Dans ce cas $\mathcal{D}$ est un fermé convexe de $\mathbb{R}^n$ (car produit direct de fermés convexes de $\mathbb{R}$), et puisque f est elliptique elle y admet un unique minimum.

Copyright : Jean-Philippe Préaux

On adapte alors la méthode de relaxation naturellement par :

$\mathbf{u}_k = (x_1^{(k)}, x_2^{(k)}, \ldots, x_n^{(k)}) \in \mathcal{D}$

$\mathbf{u}_{k+1} = (x_1^{(k+1)}, x_2^{(k+1)}, \ldots, x_n^{(k+1)}) \in \mathcal{D}$ est construit par :

Copyright : Jean-Philippe Préaux

On adapte alors la méthode de relaxation naturellement par :

$\mathbf{u}_k = (x_1^{(k)}, x_2^{(k)}, \ldots, x_n^{(k)}) \in \mathcal{D}$

$\mathbf{u}_{k+1} = (x_1^{(k+1)}, x_2^{(k+1)}, \ldots, x_n^{(k+1)}) \in \mathcal{D}$ est construit par :

$$f(x_1^{(k+1)}, x_2^{(k)}, \ldots, x_n^{(k)}) = \inf_{a_1 \leqslant x \leqslant b_1} f(x, x_2^{(k)}, \ldots, x_n^{(k)})$$

$$f(x_1^{(k+1)}, x_2^{(k+1)}, \ldots, x_n^{(k)}) = \inf_{a_2 \leqslant x \leqslant b_2} f(x_1^{(k+1)}, x, \ldots, x_n^{(k)})$$

$$\vdots$$

$$f(x_1^{(k+1)}, \ldots, x_{n-1}^{(k+1)}, x_n^{(k+1)}) = \inf_{a_n \leqslant x \leqslant b_n} f(x_1^{(k+1)}, \ldots, x_{n-1}^{(k+1)}, x)$$

(les inégalités ayant lieu dans $\overline{\mathbb{R}}$ et étant bien évidemment strictes lorsque $a_i, b_i = \pm\infty$.)

Copyright : Jean-Philippe Préaux

Théorème (Convergence de la méthode de relaxation sous contraintes.)

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *une application elliptique sur* $\mathcal{D} \subset \mathbb{R}^n$ *qui est un produit d'intervalles :*

$$\mathcal{D} = \prod_{i=1}^{n} [a_i, b_i] \qquad \text{où } a_i, b_i \in \overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\},\ a_i \leqslant b_i.$$

Copyright : Jean-Philippe Préaux

Théorème (Convergence de la méthode de relaxation sous contraintes.)

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *une application elliptique sur* $\mathcal{D} \subset \mathbb{R}^n$ *qui est un produit d'intervalles :*

$$\mathcal{D} = \prod_{i=1}^{n} [a_i, b_i] \qquad \text{où } a_i, b_i \in \overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\},\ a_i \leqslant b_i.$$

La méthode de relaxation converge vers le minimum de f *sur* $\mathcal{D}$.

Copyright : Jean-Philippe Préaux

Méthode du gradient projeté.

La méthode du gradient projeté consiste à projeter sur le domaine $\mathcal{C}$ (convexe, fermé, non vide) les points obtenus à chaque itération par la méthode du gradient à pas fixe.

Copyright : Jean-Philippe Préaux

Méthode du gradient projeté.

La méthode du gradient projeté consiste à projeter sur le domaine $\mathcal{C}$ (convexe, fermé, non vide) les points obtenus à chaque itération par la méthode du gradient à pas fixe. C'est-à-dire, soit $\rho > 0$ un pas de descente :

$$\mathbf{u}_{k+1} = P_{\mathcal{C}}(\mathbf{u}_k - \rho \nabla f(\mathbf{u}_k)) .$$

Copyright : Jean-Philippe Préaux

Sa convergence est assurée sous les mêmes hypothèses que pour la méthode du gradient à pas fixe par le théorème suivant :

Copyright : Jean-Philippe Préaux

Sa convergence est assurée sous les mêmes hypothèses que pour la méthode du gradient à pas fixe par le théorème suivant :

Théorème (Convergence de la méthode du gradient projeté.)

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une application α-elliptique et un domaine $\mathcal{C}$ non vide fermé et convexe. On suppose de plus que $\nabla f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ est M-lipschtzienne (c'est à dire $\exists\, M > 0$, $\forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leqslant M\|\mathbf{x} - \mathbf{y}\|$).

Copyright : Jean-Philippe Préaux

Sa convergence est assurée sous les mêmes hypothèses que pour la méthode du gradient à pas fixe par le théorème suivant :

Théorème (Convergence de la méthode du gradient projeté.)

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une application α-elliptique et un domaine $\mathcal{C}$ non vide fermé et convexe. On suppose de plus que $\nabla f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ est M-lipschtzienne (c'est à dire $\exists\, M > 0$, $\forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leqslant M\|\mathbf{x} - \mathbf{y}\|$). Si le pas de descente ρ est choisi tel que :

$$0 < \rho < \frac{2\alpha}{M^2}$$

alors la méthode du gradient projeté converge géométriquement vers le minimum de f sur $\mathcal{C}$.

Copyright : Jean-Philippe Préaux

Méthode d'Uzawa.

Cette méthode applique la théorie de la dualité convexe (vue au chap. III), et recherche dans un problème de programmation convexe un point-selle du lagrangien.

Copyright : Jean-Philippe Préaux

Méthode d'Uzawa.

Cette méthode applique la théorie de la dualité convexe (vue au chap. III), et recherche dans un problème de programmation convexe un point-selle du lagrangien.
Il s'agit en fait de la méthode du gradient projeté appliquée au problème dual.

Copyright : Jean-Philippe Préaux

Méthode d'Uzawa.

Cette méthode applique la théorie de la dualité convexe (vue au chap. III), et recherche dans un problème de programmation convexe un point-selle du lagrangien.
Il s'agit en fait de la méthode du gradient projeté appliquée au problème dual.
Mais dans ce cas l'opérateur de projection de $\mathbb{R}^q$ sur $(\mathbb{R}_+)^q$ est particulièrement simple à écrire ; c'est là qu'en réside tout l'intérêt.

Copyright : Jean-Philippe Préaux

Algorithme d'Uzawa.

Copyright : Jean-Philippe Préaux

Algorithme d'Uzawa.

On construit une suite $(\mathbf{x}_k)_{k\in\mathbb{N}}$ de $\mathbb{R}^n$ et deux suites $(\lambda_k)_{k\in\mathbb{N}}$ de $\mathbb{R}^p$ et $(\mu_k)_{k\in\mathbb{N}}$ de $(\mathbb{R}_+)^q$ de la façon suivante.

Copyright : Jean-Philippe Préaux

Algorithme d'Uzawa.

On construit une suite $(\mathbf{x}_k)_{k\in\mathbb{N}}$ de $\mathbb{R}^n$ et deux suites $(\lambda_k)_{k\in\mathbb{N}}$ de $\mathbb{R}^p$ et $(\mu_k)_{k\in\mathbb{N}}$ de $(\mathbb{R}_+)^q$ de la façon suivante.

- **Initialement.** On fixe $\rho > 0$ et on choisit arbitrairement $(\lambda_0, \mu_0) \in \mathbb{R}^p \times (\mathbb{R}_+)^q$.

Copyright : Jean-Philippe Préaux

Algorithme d'Uzawa.

On construit une suite $(\mathbf{x}_k)_{k\in\mathbb{N}}$ de $\mathbb{R}^n$ et deux suites $(\lambda_k)_{k\in\mathbb{N}}$ de $\mathbb{R}^p$ et $(\mu_k)_{k\in\mathbb{N}}$ de $(\mathbb{R}_+)^q$ de la façon suivante.

- **Initialement.** On fixe $\rho > 0$ et on choisit arbitrairement $(\lambda_0, \mu_0) \in \mathbb{R}^p \times (\mathbb{R}_+)^q$.
- **Itération** k. On détermine $\mathbf{x}_k \in \mathbb{R}^n$ par :

$$\mathbf{x}_k \text{ est solution de } \min_{\mathbf{x}\in\mathbb{R}^n} \mathcal{L}(\mathbf{x}, \lambda_k, \mu_k)$$

$$\text{soit encore : } \quad \nabla_x \mathcal{L}(\mathbf{x}_k, \lambda_k, \mu_k) = \mathbf{0} \ .$$

Copyright : Jean-Philippe Préaux

Algorithme d'Uzawa.

On construit une suite $(\mathbf{x}_k)_{k\in\mathbb{N}}$ de $\mathbb{R}^n$ et deux suites $(\lambda_k)_{k\in\mathbb{N}}$ de $\mathbb{R}^p$ et $(\mu_k)_{k\in\mathbb{N}}$ de $(\mathbb{R}_+)^q$ de la façon suivante.

- **Initialement.** On fixe $\rho > 0$ et on choisit arbitrairement $(\lambda_0, \mu_0) \in \mathbb{R}^p \times (\mathbb{R}_+)^q$.
- **Itération** k**.** On détermine $\mathbf{x}_k \in \mathbb{R}^n$ par :

$$\mathbf{x}_k \text{ est solution de } \min_{\mathbf{x}\in\mathbb{R}^n} \mathcal{L}(\mathbf{x}, \lambda_k, \mu_k)$$

$$\text{soit encore : } \quad \nabla_x \mathcal{L}(\mathbf{x}_k, \lambda_k, \mu_k) = \mathbf{0} \ .$$

On détermine λ_{k+1} et μ_{k+1} par :

$$\lambda_{k+1} = \lambda_k + \rho.(\varphi_1(\mathbf{x}_k), \ldots, \varphi_p(\mathbf{x}_k)) \ ,$$

$$\mu_{k+1} = P_{(\mathbb{R}_+)^q}(\mu_k + \rho.(\psi_1(\mathbf{x}_k), \ldots, \psi_q(\mathbf{x}_k)) \ .$$

Copyright : Jean-Philippe Préaux

Sa convergence est assurée sous certaines hypothèses par le résultat suivant.

Copyright : Jean-Philippe Préaux

Sa convergence est assurée sous certaines hypothèses par le résultat suivant.

Théorème (Convergence de la méthode d'Uzawa.)

On suppose que f est α-elliptique, que $\varphi_1, \ldots, \varphi_p$ et $\psi_1, \ldots, \psi_q$ sont affines, i.e.,

$$\mathcal{D} = \left\{ \mathbf{x} \in \mathbb{R}^n \,|\, A\mathbf{x} = \mathbf{b} \,;\, C\mathbf{x} \leqslant \mathbf{d} \right\}$$

avec $A \in \mathcal{M}_{p,n}(\mathbb{R})$, $\mathbf{b} \in \mathbb{R}^p$, $C \in \mathcal{M}_{q,n}(\mathbb{R})$ et $\mathbf{d} \in \mathbb{R}^q$.

Copyright : Jean-Philippe Préaux

Sa convergence est assurée sous certaines hypothèses par le résultat suivant.

Théorème (Convergence de la méthode d'Uzawa.)

On suppose que f est α-elliptique, que $\varphi_1, \ldots, \varphi_p$ et $\psi_1, \ldots, \psi_q$ sont affines, i.e.,

$$\mathcal{D} = \left\{ \mathbf{x} \in \mathbb{R}^n \,|\, A\mathbf{x} = \mathbf{b}\,;\; C\mathbf{x} \leqslant \mathbf{d} \right\}$$

avec $A \in \mathcal{M}_{p,n}(\mathbb{R})$, $\mathbf{b} \in \mathbb{R}^p$, $C \in \mathcal{M}_{q,n}(\mathbb{R})$ et $\mathbf{d} \in \mathbb{R}^q$.
Alors en choisissant ρ tel que

$$0 < \rho < \frac{2\alpha}{\|A\|^2 + \|C\|^2}$$

la suite $(\mathbf{u}_k)_{k\in\mathbb{N}}$ converge vers l'unique minimum de f sur $\mathcal{D}$.

Copyright : Jean-Philippe Préaux

Remarque.

Copyright : Jean-Philippe Préaux

Remarque.

On peut utiliser dans la conclusion n'importe quelle norme matricielle compatible avec la norme vectorielle usuelle ($\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$), c'est à dire telle que $\|A\mathbf{x}\| \leqslant \|A\| \, \|\mathbf{x}\|$. On a un large choix :

Copyright : Jean-Philippe Préaux

Remarque.

On peut utiliser dans la conclusion n'importe quelle norme matricielle compatible avec la norme vectorielle usuelle ($\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$), c'est à dire telle que $\|A\mathbf{x}\| \leqslant \|A\| \, \|\mathbf{x}\|$. On a un large choix :

– La *norme de Frobenius* :

$$\|A\|_f = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2} = tr(AA^\top) \, .$$

Copyright : Jean-Philippe Préaux

Remarque.

On peut utiliser dans la conclusion n'importe quelle norme matricielle compatible avec la norme vectorielle usuelle ($\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$), c'est à dire telle que $\|A\mathbf{x}\| \leqslant \|A\| \, \|\mathbf{x}\|$. On a un large choix :
– La *norme de Frobenius* :

$$\|A\|_f = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2} = tr(AA^\top) \; .$$

– La *norme induite par* $\|.\|_2$:

$$\|A\|_{\sup} = \sup_{\mathbf{x} \neq \mathbf{0}} \frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \; .$$

Copyright : Jean-Philippe Préaux

Remarque.

On peut utiliser dans la conclusion n'importe quelle norme matricielle compatible avec la norme vectorielle usuelle ($\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$), c'est à dire telle que $\|A\mathbf{x}\| \leqslant \|A\| \, \|\mathbf{x}\|$. On a un large choix :

– La *norme de Frobenius* :

$$\|A\|_f = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2} = tr(AA^\top) \ .$$

– La *norme induite par* $\|.\|_2$:

$$\|A\|_{\text{sup}} = \sup_{\mathbf{x} \neq \mathbf{0}} \frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \ .$$

– Si A est une matrice carrée diagonalisable, elle coïncide avec la *norme spectrale* :

$$\|A\|_s = \max \{|\lambda| \ | \ \lambda \text{ est une valeur propre de } A\} \ .$$

Copyright : Jean-Philippe Préaux

Exercices

- Exercice 1.
- Exercice 2.
- Exercice 3.

Copyright : Jean-Philippe Préaux

Exercice 1.

Considérons le cas d'un problème de programmation quadratique elliptique sans contraintes :

$$\min_{\mathbf{x}} \frac{1}{2}\mathbf{x}^\top A\mathbf{x} - \mathbf{b}^\top \mathbf{x} \ .$$

Comment s'exprime dans ce cas la méthode de Newton ?
Sa convergence dépend-elle du point initial ?
En combien d'itérations converge-t-elle ?
Réinterpréter la méthode de Newton.

Retour.

Copyright : Jean-Philippe Préaux

Dans ce cas $\forall \mathbf{x} \in \mathbb{R}^n$, $\nabla f(\mathbf{x}) = A\mathbf{x} - \mathbf{b}$ et $\nabla^2 f(\mathbf{x}) = A$.

Copyright : Jean-Philippe Préaux

Dans ce cas $\forall \mathbf{x} \in \mathbb{R}^n$, $\nabla f(\mathbf{x}) = A\mathbf{x} - \mathbf{b}$ et $\nabla^2 f(\mathbf{x}) = A$.

La méthode de Newton s'exprime :

$$\mathbf{u}_{k+1} = \mathbf{u}_k - \nabla^2 f(\mathbf{u}_k)^{-1} \nabla f(\mathbf{u}_k) = \mathbf{u}_k - A^{-1}(A\mathbf{u}_k - \mathbf{b}) = A^{-1}\mathbf{b} \ .$$

Copyright : Jean-Philippe Préaux

Dans ce cas $\forall \mathbf{x} \in \mathbb{R}^n$, $\nabla f(\mathbf{x}) = A\mathbf{x} - \mathbf{b}$ et $\nabla^2 f(\mathbf{x}) = A$.

La méthode de Newton s'exprime :

$$\mathbf{u}_{k+1} = \mathbf{u}_k - \nabla^2 f(\mathbf{u}_k)^{-1} \nabla f(\mathbf{u}_k) = \mathbf{u}_k - A^{-1}(A\mathbf{u}_k - \mathbf{b}) = A^{-1}\mathbf{b} \ .$$

Puisque A est définie positive, $A^{-1}\mathbf{b}$ est l'unique minimum de f sur $\mathbb{R}^n$.

Copyright : Jean-Philippe Préaux

Ainsi la méthode de Newton équivaut à la résolution directe de ce problème.

Copyright : Jean-Philippe Préaux

Ainsi la méthode de Newton équivaut à la résolution directe de ce problème. Sa convergence se fait en une itération et ne dépend pas du point base.

Copyright : Jean-Philippe Préaux

Ainsi la méthode de Newton équivaut à la résolution directe de ce problème. Sa convergence se fait en une itération et ne dépend pas du point base. Elle n'apporte donc rien ici.

Copyright : Jean-Philippe Préaux

Ainsi la méthode de Newton équivaut à la résolution directe de ce problème. Sa convergence se fait en une itération et ne dépend pas du point base. Elle n'apporte donc rien ici.

En général, la méthode de Newton peut se réinterpréter de la façon suivante : elle revient à approcher au voisinage de $\mathbf{u}_k$ l'application à minimiser par une application quadratique.

Retour.

Copyright : Jean-Philippe Préaux

Exercice 2.

Considérons le cas d'un problème de programmation quadratique elliptique sans contraintes :

$$\min_{\mathbf{x}} \frac{1}{2}\mathbf{x}^\top A\mathbf{x} - \mathbf{b}^\top \mathbf{x} \ .$$

Comment s'exprime dans ce cas la méthode de relaxation ?

Retour.

Copyright : Jean-Philippe Préaux

Soit $t \in \{1, \ldots, n\}$.

$$f(x_1, x_2, \ldots, x_n)$$
$$= \frac{1}{2}\sum_{i=1}^{n} a_{ii}x_i^2 + \sum_{i<j} a_{ij}x_i x_j - \sum_{i=1}^{n} b_i x_i$$
$$= \underbrace{\frac{1}{2}a_{tt}x_t^2 + x_t \sum_{\substack{j=1 \\ j\neq t}}^{n} a_{tj}x_j - b_t x_t}_{\text{dépend de } x_t} + \underbrace{\frac{1}{2}\sum_{\substack{i=1 \\ i\neq t}}^{n} a_{ii}x_i^2 + \sum_{\substack{i<j \\ i,j\neq t}} a_{ij}x_i x_j - \sum_{\substack{i=1 \\ i\neq t}}^{n} b_i x_i}_{\text{ne dépend pas de } x_t}$$

Copyright : Jean-Philippe Préaux

Soit $t \in \{1, \ldots, n\}$.

$$f(x_1, x_2, \ldots, x_n)$$
$$= \frac{1}{2}\sum_{i=1}^{n} a_{ii}x_i^2 + \sum_{i<j} a_{ij}x_i x_j - \sum_{i=1}^{n} b_i x_i$$
$$= \underbrace{\frac{1}{2}a_{tt}x_t^2 + x_t \sum_{\substack{j=1 \\ j\neq t}}^{n} a_{tj}x_j - b_t x_t}_{\text{dépend de } x_t} + \underbrace{\frac{1}{2}\sum_{\substack{i=1 \\ i\neq t}}^{n} a_{ii}x_i^2 + \sum_{\substack{i<j \\ i,j\neq t}} a_{ij}x_i x_j - \sum_{\substack{i=1 \\ i\neq t}}^{n} b_i x_i}_{\text{ne dépend pas de } x_t}$$

Alors :

$$\frac{\partial f}{\partial x_t}(x_1, \ldots, x_n) = a_{tt}x_t + \sum_{\substack{j=1 \\ j\neq t}}^{n} a_{tj}x_j - \mathbf{b} = \sum_{j=1}^{n} a_{tj}x_j - \mathbf{b}$$

Copyright : Jean-Philippe Préaux

Puisque A est définie positive, f est elliptique.

Copyright : Jean-Philippe Préaux

Puisque A est définie positive, f est elliptique.
Cela implique qu'en tout point $\mathbf{u} \in \mathbb{R}^n$, et pour tout $t \in \{1, \ldots, n\}$ chacune des applications $x \mapsto f(\mathbf{u} + x\mathbf{e}_t)$ est strictement convexe et coercive.

Copyright : Jean-Philippe Préaux

Puisque A est définie positive, f est elliptique.
Cela implique qu'en tout point $\mathbf{u} \in \mathbb{R}^n$, et pour tout $t \in \{1, \ldots, n\}$ chacune des applications $x \mapsto f(\mathbf{u} + x\mathbf{e}_t)$ est strictement convexe et coercive.
Chacune admet donc un minimum global caractérisé par la condition d'Euler :

$$\frac{\partial f}{\partial x_t}(\mathbf{u} + x\mathbf{e}_t) = 0 \ .$$

Copyright : Jean-Philippe Préaux

Ainsi, si :

$$f(x_1^{(k+1)}, x_2^{(k)}, \ldots, x_n^{(k)}) = \inf_{x \in \mathbb{R}} f(x, x_2^{(k)}, \ldots, x_n^{(k)})$$
$$f(x_1^{(k+1)}, x_2^{(k+1)}, \ldots, x_n^{(k)}) = \inf_{x \in \mathbb{R}} f(x_1^{(k+1)}, x, \ldots, x_n^{(k)})$$
$$\vdots$$
$$f(x_1^{(k+1)}, \ldots, x_{n-1}^{(k+1)}, x_n^{(k+1)}) = \inf_{x \in \mathbb{R}} f(x_1^{(k+1)}, \ldots, x_{n-1}^{(k+1)}, x)$$

Alors donné $\mathbf{u}_k = (x_1^{(k)}, \ldots, x_n^{(k)})$, le point $\mathbf{u}_{k+1} = (x_1^{(k+1)}, \ldots, x_n^{(k+1)})$ construit par la méthode de relaxation est caractérisé par :

Copyright : Jean-Philippe Préaux

Ainsi, si :

$$f(x_1^{(k+1)}, x_2^{(k)}, \ldots, x_n^{(k)}) = \inf_{x\in\mathbb{R}} f(x, x_2^{(k)}, \ldots, x_n^{(k)})$$
$$f(x_1^{(k+1)}, x_2^{(k+1)}, \ldots, x_n^{(k)}) = \inf_{x\in\mathbb{R}} f(x_1^{(k+1)}, x, \ldots, x_n^{(k)})$$
$$\vdots$$
$$f(x_1^{(k+1)}, \ldots, x_{n-1}^{(k+1)}, x_n^{(k+1)}) = \inf_{x\in\mathbb{R}} f(x_1^{(k+1)}, \ldots, x_{n-1}^{(k+1)}, x)$$

Alors donné $\mathbf{u}_k = (x_1^{(k)}, \ldots, x_n^{(k)})$, le point $\mathbf{u}_{k+1} = (x_1^{(k+1)}, \ldots, x_n^{(k+1)})$ construit par la méthode de relaxation est caractérisé par :

$$a_{11}x_1^{(k+1)} + a_{12}x_2^{(k)} + \ldots + a_{1n}x_n^{(k)} = b_1$$
$$\vdots$$
$$\Longleftrightarrow$$
$$a_{21}x_1^{(k+1)} + a_{22}x_2^{(k+1)} + \ldots + a_{2n}x_n^{(k)} = b_2$$
$$a_{n1}x_1^{(k+1)} + a_{n2}x_2^{(k+1)} + \ldots + a_{nn}x_n^{(k+1)} = b_n \ .$$

Copyright : Jean-Philippe Préaux

On le construit donc grâce à :

$$\begin{aligned}
x_1^{(k+1)} &= \frac{1}{a_{11}}(b_1 - a_{12}x_2^{(k)} - \ldots - a_{1n}x_n^{(k)}) \\
x_2^{(k+1)} &= \frac{1}{a_{22}}(b_2 - a_{21}x_1^{(k+1)} - \ldots - a_{2n}x_n^{(k)}) \\
&\vdots \\
x_n^{(k+1)} &= \frac{1}{a_{nn}}(b_n - a_{n1}x_1^{(k+1)} - \ldots - a_{nn-1}x_{n-1}^{(k+1)}) \ .
\end{aligned}$$

en effet la matrice A étant définie positive, $a_{11}, a_{22}, \ldots, a_{nn} \neq 0$.

Retour.

Copyright : Jean-Philippe Préaux

Exercice 3.

Considérons le cas d'un problème de programmation quadratique elliptique sous contraintes :

$$\min_{\mathbf{x}} \frac{1}{2}\mathbf{x}^\top A\mathbf{x} - \mathbf{b}^\top \mathbf{x}$$
$$C\mathbf{x} = \mathbf{c}$$
$$D\mathbf{x} \leqslant \mathbf{d}$$

où $A \in \mathcal{M}_n(\mathbb{R})$ définie positive, $\mathbf{b} \in \mathbb{R}^n$, $C \in \mathcal{M}_{p,n}(\mathbb{R})$, $\mathbf{c} \in \mathbb{R}^p$, $D \in \mathcal{M}_{q,n}(\mathbb{R})$, $\mathbf{d} \in \mathbb{R}^q$.

a. Comment s'exprime le vecteur gradient du lagrangien $\mathcal{L}(\mathbf{x}, \lambda, \mu)$ de ce problème ?

b. Exprimer les conditions nécessaires de KKT pour ce problème sous forme matricielle.

c. Comment s'exprime ici la méthode d'Uzawa ?

Copyright : Jean-Philippe Préaux

Retour.

a.

Copyright : Jean-Philippe Préaux

a. Le vecteur gradient du lagrangien du problème est :

$$\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{u},\lambda,\mu) = \nabla f(\mathbf{u}) + \sum_{i=1}^{p} \lambda_i \nabla \varphi_i(\mathbf{u}) + \sum_{j=1}^{q} \mu_j \nabla \psi_j(\mathbf{u}) \ .$$

Copyright : Jean-Philippe Préaux

a. Le vecteur gradient du lagrangien du problème est :

$$\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{u},\lambda,\mu) = \nabla f(\mathbf{u}) + \sum_{i=1}^{p} \lambda_i \nabla\varphi_i(\mathbf{u}) + \sum_{j=1}^{q} \mu_j \nabla\psi_j(\mathbf{u}) \ .$$

En notant $\mathbf{c}_i$ et $\mathbf{d}_j$ la i^e ligne de la matrice C et la j^e ligne de la matrice D, $\nabla\varphi_i(\mathbf{u}) = \mathbf{c}_i^\top$ et $\nabla\psi_j(\mathbf{u}) = \mathbf{d}_j^\top$.

Copyright : Jean-Philippe Préaux

a. Le vecteur gradient du lagrangien du problème est :

$$\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{u},\lambda,\mu) = \nabla f(\mathbf{u}) + \sum_{i=1}^{p} \lambda_i \nabla\varphi_i(\mathbf{u}) + \sum_{j=1}^{q} \mu_j \nabla\psi_j(\mathbf{u}) \ .$$

En notant $\mathbf{c}_i$ et $\mathbf{d}_j$ la i^e ligne de la matrice C et la j^e ligne de la matrice D, $\nabla\varphi_i(\mathbf{u}) = \mathbf{c}_i^\top$ et $\nabla\psi_j(\mathbf{u}) = \mathbf{d}_j^\top$.
Aussi :

$$\begin{aligned}\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{u},\lambda,\mu) &= \nabla f(\mathbf{u}) + \sum_{i=1}^{p} \lambda_i \mathbf{c}_i^\top + \sum_{j=1}^{q} \mu_j \mathbf{d}_j^\top \\ &= A\mathbf{u} - \mathbf{b} + C^\top\lambda + D^\top\mu \ .\end{aligned}$$

Copyright : Jean-Philippe Préaux

b.

Copyright : Jean-Philippe Préaux

b. Les conditions de KKT s'écrivent ici :

$$\begin{aligned} &A\mathbf{u} - \mathbf{b} + C^\top \lambda + D^\top \mu = \mathbf{0}, \\ &\mu \geqslant \mathbf{0}\ , \\ &\mu^\top (D\mathbf{u} - \mathbf{d}) = \mathbf{0}\ . \end{aligned}$$

Copyright : Jean-Philippe Préaux

b. Les conditions de KKT s'écrivent ici :

$$\begin{aligned} &A\mathbf{u} - \mathbf{b} + C^\top \lambda + D^\top \mu = \mathbf{0} , \\ &\mu \geqslant \mathbf{0} , \\ &\mu^\top (D\mathbf{u} - \mathbf{d}) = \mathbf{0} . \end{aligned}$$

c. En utilisant l'expression du vecteur gradient du lagrangien obtenue en **a.** :

$$\begin{aligned} \mathbf{x}_k &= A^{-1}(\mathbf{b} - C^\top \lambda - D^\top \mu) , \\ \lambda_{k+1} &= \lambda_k + \rho(C\mathbf{x}_k - \mathbf{c}) , \\ \mu_{k+1} &= P_{(\mathbb{R}_+)^q}(\mu_k + \rho(D\mathbf{x}_k - \mathbf{d})) . \end{aligned}$$

Copyright : Jean-Philippe Préaux

Retour.