Sommaire

Généralités

On considère des fonctions f définies dans $\mathbb{R}^n$ et à valeurs dans $\mathbb{R}$. On cherche les points où f atteint des extrema locaux ou globaux.

Si x est un élément de $\mathbb{R}^n$, on pose $x = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}$.

Définition du gradient

$f : \mathbb{R}^n \to \mathbb{R}$. On suppose que f admet en un point x des dérivées partielles du premier ordre.

$$\textit{gradient de } f = \nabla f(x) = \begin{pmatrix} \dfrac{\partial f}{\partial x_1}(x) \\ \dots \\ \dfrac{\partial f}{\partial x_n}(x) \end{pmatrix}.$$

Généralités, suite

Définition du gradient d'un vecteur-ligne de fonctions

$F(x) = (f_1(x), \ldots, f_p(x))$, vecteur-ligne, avec $f_j : \mathbb{R}^n \to \mathbb{R}$ dérivable au point x.
$\nabla F(x)$ est la matrice dont la j^{e} colonne est $\nabla f_j(x)$.

A, matrice carrée constante d'ordre n, $u(x)$ et $v(x)$ deux vecteurs-colonnes :

$$\nabla\left(u^{\mathrm{t}}.A\right) = \nabla\left(u^{\mathrm{t}}\right).A \text{ et } \nabla\left(A.u\right) = A.\nabla\left(u\right).$$

$$\nabla\left(u^{\mathrm{t}}.v\right) = \nabla\left(u^{\mathrm{t}}\right).v + \nabla\left(v^{\mathrm{t}}\right).u$$

Si f admet en x^0 des dérivées partielles continues, formule de Taylor à l'ordre 1 :

$$\boxed{f(x) = f(x^0) + (x - x^0)^{\mathrm{t}}.\nabla f(x^0) + ||x - x^0||.\epsilon(x)}$$

où $\epsilon(x)$ est une fonction qui tend vers 0 quand x tend vers x^0.

Généralités, suite

Remarques

• Si f de classe C^1 :
$x_{n+1} = f(x^0) + (x - x^0)^t.\nabla f(x^0)$ est l'équation de l'hyperplan tangent à la surface S de $\mathbb{R}^{n+1}$ d'équation $x_{n+1} = f(x_1, ..., x_n)$ au point $(x^0, f(x^0))$.

• Variations de f dans une direction d de $\mathbb{R}^n$ donnée, à partir d'un point x^0 de $\mathbb{R}^n$ donné :

pour $s \in \mathbb{R}$, on pose $g(s) = f(x^0 + s.d)$.

On a :

$g'(s) = d^t.\nabla f(x^0 + s.d)$ et $g'(0) = d^t.\nabla f(x^0)$.

Matrice hessienne

Définition de la matrice hessienne

Si f admet des dérivées partielles d'ordre 2 en x :

$$\nabla^2 f(x) = \nabla\left(\nabla f(x)^{\mathrm{t}}\right) = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1^2}(x) & \dots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \dfrac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \dots & \dfrac{\partial^2 f}{\partial x_2 \partial x_n}(x) \\ \dots & \dots & \dots \\ \dfrac{\partial^2 f}{\partial x_n \partial x_1}(x) & \dots & \dfrac{\partial^2 f}{\partial x_n^2}(x) \end{pmatrix};$$

$\nabla^2 f$ s'appelle la *matrice hessienne* de f.

Si f est une fonction de classe C^2 (autrement dit, f admet des dérivées partielles à l'ordre 2 continues), la matrice hessienne de f est une matrice symétrique (théorème de Schwarz).

Formule de Taylor d'ordre 2

Si f est de classe C^2 en x^0, formule de Taylor d'ordre 2 :

$$\boxed{\begin{array}{c} f(x) = f(x^0) + (x - x^0)^t.\nabla f(x^0) + \dfrac{1}{2}(x - x^0)^t \nabla^2 f(x^0).(x - x^0) \\ + ||x - x^0||^2.\epsilon(x) \end{array}}$$

où $\epsilon(x)$ est une fonction qui tend vers 0 quand x tend vers x^0.

Optimisation non linéaire sans contraintes

Rappels

Définitions

Soit A une matrice réelle symétrique d'ordre n.

A est positive $\Leftrightarrow \forall h \in \mathbb{R}^n, h^t.A.h \geqslant 0.$

A est définie positive $\Leftrightarrow \forall h \in \mathbb{R}^n \setminus \{0\}, h^t.A.h > 0.$

Une matrice symétrique réelle est diagonalisable sur $\mathbb{R}$.

La matrice A est positive $\Leftrightarrow$ les valeurs propres de A sont positives ou nulles.

La matrice A est définie positive $\Leftrightarrow$ les valeurs propres de A sont strictement positives.

Condition nécessaire d'optimalité

On suppose désormais que f est de classe C^2.

Théorème

Si f admet un minimum local en x^*, alors :

1) $\nabla f(x^*) = 0$

2) $\nabla^2 f(x^*)$ est une matrice positive.

Preuve de 1). $f(x) = f(x^*) + (x - x^*)^t.\nabla f(x^*) + ||x - x^*||.\epsilon(x)$,
où $\epsilon(x)$ est une fonction qui tend vers 0 quand x tend vers x^*.

Avec $x = x^* - \theta.\nabla f(x^*)$, avec $\theta \in \mathbb{R}^+$:

$$f(x) - f(x^*) = -\theta.||\nabla f(x^*)||^2 + \theta.\epsilon_1(\theta)$$

$$f(x) - f(x^*) = \theta\left(-||\nabla f(x^*)||^2 + \epsilon_1(\theta)\right),$$

où $\epsilon_1(\theta)$ est une fonction qui tend vers 0 quand θ tend vers 0.

Si $\nabla f(x^*) \neq 0$: contradiction avec la minimalité en x^*.

Preuve de 2)

On suppose qu'il existe $h \in \mathbb{R}^n$ tel qu'on ait : $h^{\mathrm{t}}.\nabla^2 f(x^*).h < 0$.

D'après le développement de Taylor d'ordre 2 :

$$f(x^* + \theta h) - f(x^*) = \theta^2.\left(\frac{1}{2}h^t.\nabla^2 f(x^*).h + \epsilon_2(\theta)\right),$$

où $\epsilon_2(\theta)$ est une fonction qui tend vers 0 quand θ tend vers 0.

Pour θ assez petit, la différence $f(x^* + \theta h) - f(x^*)$ serait négative : contradiction.

Condition suffisante d'optimalité

Théorème

Si une fonction f vérifie en x^* :
• $\nabla f(x^*) = 0$,
• $\nabla^2 f(x^*)$ est une matrice définie positive,
alors f admet un minimum local en x^*.

Preuve. $S =$ sphère de centre 0 et de rayon 1.
$a = \inf\{h^{\mathrm{t}}.\nabla^2 f(x^*).h \text{ pour } h \in S\}$:

$$\forall h \in \mathbb{R}^n, h^t.\nabla^2 f(x^*).h \geqslant a||h||^2.$$

S est un compact $\Rightarrow$ la valeur a est atteinte et
$\nabla^2 f(x^*)$ définie positive $\Rightarrow a > 0$.

Formule de Taylor à l'ordre 2 :
$$f(x^*+h)-f(x^*) = \frac{1}{2}h^t.\nabla^2 f(x^*).h+||h||^2\epsilon(h) \geqslant ||h||^2.\left(\frac{a}{2}+\epsilon(h)\right),$$
où $\epsilon(h)$ tend vers 0 quand h tend vers 0. Ce qui montre le théorème.

Optimisation non linéaire sans contraintes

Fonctions quadratiques

A : matrice symétrique d'ordre n, b : vecteur-colonne d'ordre n, $c \in \mathbb{R}$.

Définition

L'application q de $\mathbb{R}^n$ dans $\mathbb{R}$ définie par :

$$q(x) = c + b^{\mathrm{t}}.x + \frac{1}{2}x^{\mathrm{t}}.A.x$$

s'appelle *fonction quadratique*.

Remarque. La partie polynomiale du développement de Taylor d'ordre 2 d'une fonction f est la fonction quadratique q telle que la surface d'équation $x_{n+1} = q(x)$ soit « la plus proche » de la surface d'équation $x_{n+1} = f(x)$ au voisinage du point considéré.

Dérivation

$$q(x) = c + b^t.x + \frac{1}{2}x^t.A.x$$

$$\nabla q(x) = \nabla(x^t).b + \frac{1}{2}[\nabla(x^t).A.x + \nabla((A.x)^t).x].$$

$\nabla(x^t)$ est la matrice identité.

$\nabla(A.x)^t = \nabla(x^t.A^t) = \nabla(x^t).A^t = A^t = A.$

D'où : $\boxed{\nabla q(x) = b + A.x}$.

De plus : $\nabla^2 q(x) = \nabla((\nabla q(x))^t) = \nabla(b^t + x^t.A^t) = A^t.$

On a donc : $\boxed{\nabla^2 q(x) = A}$.

Les dérivées d'ordre au moins 3 de q sont nulles.

Une fonction quadratique coïncide avec son développement de Taylor à l'ordre 2.

Optimisation non linéaire sans contraintes

Fonctions convexes

Définition

On dit qu'une fonction f de $\mathbb{R}^n$ dans $\mathbb{R}$ est convexe si, pour tout x et tout y de $\mathbb{R}^n$ et pour tout λ de $]0,1[$, on a :
$f(\lambda x + (1-\lambda)y) \leqslant \lambda f(x) + (1-\lambda) f(y)$.

Si cette inégalité est stricte, on dit que f est strictement convexe.

Théorème

Théorème

Si f est une fonction convexe et admet des dérivées partielles, alors f admet un minimum global en x^* si et seulement si on a la relation $\nabla f(x^*) = 0$.

Preuve. x^* est un minimum local $\Rightarrow \nabla f(x^*) = 0$.

Réciproque : on suppose $\nabla f(x^*) = 0$.

Soit $x \in \mathbb{R}^n$. On pose, pour $s \in \mathbb{R}$: $g(s) = f(x^* + s(x - x^*))$.

$g(0) = f(x^*)$, $g(1) = f(x)$, $g'(0) = (x - x^*)^{\mathrm{t}}.\nabla f(x^*) = 0$.

g est une fonction convexe de $\mathbb{R}$ dans $\mathbb{R}$; sa dérivée est croissante : pour $s \geqslant 0, g'(s) \geqslant 0$; g est croissante pour $s \geqslant 0$.

$g(1) \geqslant g(0)$: f admet bien un minimum global en x^*.

Théorème

Théorème

Si f est convexe et admet un minimum local en x^*, alors f admet un minimum global en x^*.

Preuve.

Si f admet un minimum local en x^*, alors $\nabla f(x^*) = 0$.

Si de plus f est convexe, le théorème précédent montre qu'elle admet un minimum global en x^*.

Théorème

Théorème admis

Si f est deux fois continûment dérivable, les propositions suivantes sont équivalentes :

a. f est convexe.

b. Pour tout x et tout x^0 de $\mathbb{R}^n$,
$f(x) \geqslant f(x^0) + (\nabla f(x^0))^{\mathrm{t}}.(x - x^0)$ (la surface de $\mathbb{R}^{n+1}$d'équation $x_{n+1} = f(x)$ est au-dessus de ses hyperplans tangents).

c. Pour tout x, $\nabla^2 f(x)$ est positive.

Une fonction quadratique $q(x) = \dfrac{1}{2}x^{\mathrm{t}}.A.x + b^{\mathrm{t}}.x + c$ est convexe si et seulement si A est positive.

Si A est définie positive, alors q est strictement convexe et admet un minimum global unique.

Optimisation non linéaire sans contraintes

Généralités sur les méthodes d'optimisation sans contrainte

Même si on s'intéresse le plus souvent à des extrema globaux, on cherche en général des extrema locaux.

On indique des méthodes qui concernent des minima.

Pour déterminer un point où une fonction f atteint un minimum local, les méthodes consistent très souvent à construire une suite $x^0, x^1, ..., x^k, ...$ qui doit converger vers un point x^* vérifiant une condition nécessaire d'optimalité, par exemple $\nabla f(x^*) = 0$.

Méthodes de descente

On appelle *méthode de descente* toute méthode où, à chaque étape, on pose $x^{k+1} = x^k + s_k d^k$, où $s_k \in \mathbb{R}^+$ et d^k est une direction de $\mathbb{R}^n$ qui vérifie $(d^k)^{\mathrm{t}}.\nabla f(x^k) < 0$.

$(d^k)^{\mathrm{t}}.\nabla f(x^k) < 0$ signifie que $f(x^k + sd^k)$ a une dérivée négative pour $s = 0$: partant de x^k dans la direction d^k, f décroît (« on descend »).

La différence entre les diverses méthodes de descente porte sur le choix de s_k et de d^k, choix qui doit au minimum assurer $f(x^{k+1}) \leqslant f(x^k)$.

Vitesse de convergence

Lorsque la convergence d'un algorithme a été établie, une qualité importante de cet algorithme est sa *vitesse de convergence*.

• Si on a $\dfrac{||x^{k+1} - x^*||}{||x^k - x^*||} \leqslant \alpha < 1$ pour k assez grand, on dit que la convergence est *linéaire* de taux α.

• Si $\dfrac{||x^{k+1} - x^*||}{||x^k - x^*||}$ tend vers 0 quand k tend vers l'infini, on dit que la *convergence* est *superlinéaire*.

• Si $\dfrac{||x^{k+1} - x^*||}{||x^k - x^*||^\gamma}$ est borné, avec $\gamma > 1$, on dit que la *convergence* est *superlinéaire* d'ordre γ.

Dans le cas $\gamma = 2$, on dit que la convergence est *quadratique*.

Optimisation non linéaire sans contraintes

Méthodes de gradient, principe

Famille de méthodes itératives pour des fonctions dérivables.

Soient d un vecteur de $\mathbb{R}^n$ et x^k un point de $\mathbb{R}^n$ avec $\nabla f(x^k) \neq 0$.

On pose, pour $s \in \mathbb{R}$: $g(s) = f(x^k + sd)$.

On dit que d est une *direction de descente* si $g'(0) < 0$.

$g'(0) = d^{\mathrm{t}}.\nabla f(x^k)$. En notant θ l'angle entre $\nabla f(x^k)$ et d :

$g'(0) = ||\nabla f(x^k)||.||d||.\cos\theta$.

En supposant d unitaire, $g'(0)$ est minimum si $\cos\theta = -1$, *i.e.* si d est donnée par l'opposé du gradient : $d = -\dfrac{\nabla f(x^k)}{||\nabla f(x^k)||}$.

Cette dernière direction est la *direction de plus grande pente descendante*. C'est ce choix qui est fait dans les méthodes de gradient.

Méthode de la plus forte pente à pas optimal

L'algorithme de la plus forte pente peut s'écrire de la façon suivante :

- Choisir un point de départ x^0 ;
- $k \leftarrow 0$
- répéter
  - $\star$ $d^k \leftarrow -\nabla f(x^k)$
  - $\star$ déterminer s_k vérifiant $f(x^k + s_k d^k) = \mathrm{Min}_{s \geqslant 0} f(x^k + s d^k)$
  - $\star$ $x^{k+1} \leftarrow x^k + s_k d^k$
  - $\star$ $k \leftarrow k + 1$

tant qu'un test d'arrêt donné n'est pas vérifié.

Test d'arrêt de la méthode de la plus forte pente à pas optimal

Le test d'arrêt peut être par exemple :

• on a épuisé un nombre d'itérations fixé à l'avance ;

• le gradient est très petit : $\sum_{i=1}^{n}(\frac{\partial f}{\partial x_i}(x^k))^2 \leqslant \epsilon$, où ϵ est un paramètre donné ;

• la suite x^k est « presque » stationnaire : $|f(x^{k+1}) - f(x^k)| \leqslant \epsilon$, où ϵ est un paramètre donné.

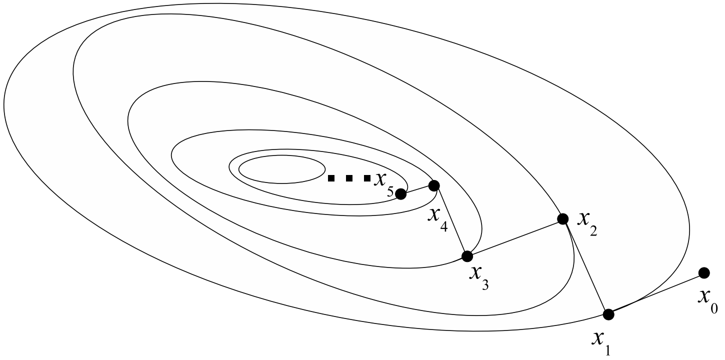
On peut montrer que si $f(x)$ est une fonction de classe C^1 qui tend vers l'infini quand $||x||$ tend vers l'infini, cet algorithme converge vers un point stationnaire (point où le gradient s'annule).

Vitesse de convergence

La vitesse de convergence peut être très faible (linéaire avec un taux proche de 1).

$$\frac{d}{ds}[f(x^k - s.\nabla f(x^k))](s_k) = 0 \text{ s'écrit : } [\nabla f(x^k)]^{\mathrm{t}}.\nabla f(x^{k+1}) = 0\,;$$

les directions de déplacement successives sont orthogonales. Il y a convergence en zig-zag.



Méthode de la plus forte pente accélérée

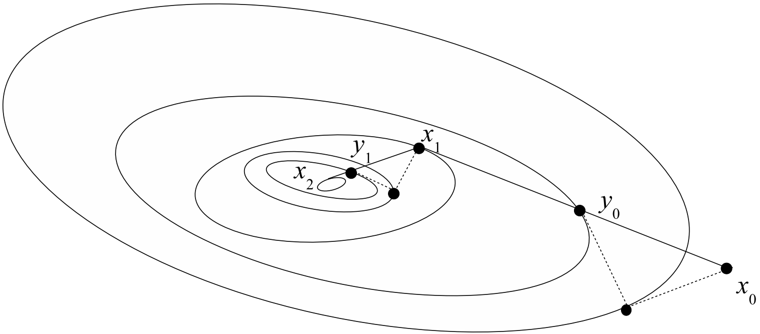
Méthode de descente fondée sur la méthode de la plus forte pente.

Soit p un entier fixé. À partir d'un point x^k, on effectue p itérations de la méthode de la plus forte pente $\rightarrow$ un point y^k.

On pose $d^k = y^k - x^k$.

x^{k+1} = un point où $f(x^k + sd^k)$ admet un minimum pour $s > 0$.

Ci-dessous, $p = 2$.



Optimisation non linéaire sans contraintes

Méthode de Newton, non prouvée

On suppose f de classe C^2.

Au voisinage d'un point x^k, on approche f par la fonction quadratique donnée par la formule de Taylor d'ordre 2 :

$$q(x) = f(x^k) + (x - x^k)^{\mathrm{t}}.\nabla f(x^k) + \frac{1}{2}(x - x^k)^{\mathrm{t}}\nabla^2 f(x^k)(x - x^k).$$

Soit x^{k+1} le point, s'il existe, qui minimise q.

Si $\nabla^2 f(x^k)$ est définie positive, on détermine x^{k+1} par l'équation

$\nabla q(x) = 0$ *i.e.* $\nabla f(x^k) + \nabla^2 f(x^k).(x - x^k) = 0$:

$x^{k+1} = x^k - [\nabla^2 f(x^k)]^{-1}.\nabla f(x^k)$.

Proposition

On suppose la fonction f de classe C^3. Si x^0 est choisi suffisamment proche d'un minimum local x^* où la matrice hessienne de f est définie positive, alors la suite (x^k) a une convergence quadratique vers x^*.

Les méthodes d'optimisation unidimensionnelle ont d'autant plus d'importance qu'elles servent d'outils pour l'optimisation de fonctions de plusieurs variables.

Méthode de Newton, cas unidimensionnel

Soit $f : \mathbb{R} \mapsto \mathbb{R}$.

À partir d'un point x_k, on approche f par :

$$q(x) = f(x_k) + f'(x_k)(x - x_k) + \frac{1}{2} f''(x_k)(x - x_k)^2.$$

On remarque : $q'(x) = f'(x_k) + f''(x_k)(x - x_k)$.

Si $f''(x_k) > 0$ (cas où f est strictement convexe autour de x_k) :

$$\boxed{x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}}$$

qui est le point où q atteint son minimum ($q'(x_{k+1}) = 0$).

Si $f''(x_k) \leqslant 0$, la méthode échoue.

Dichotomie pour une fonction dérivable

Définition

On dit qu'une fonction est *unimodale* s'il existe un réel x^* pour lequel la fonction est strictement décroissante sur $]-\infty, x^*]$ et strictement croissante sur $[x^*, +\infty[$.

x^* est alors minimum global de f.

Soit f dérivable et unimodale. On cherche x^* qui vérifie $f'(x^*) = 0$.

1) Recherche de x_{min} et x_{max} avec $f'(x_{min}) < 0$ et $f'(x_{max}) > 0$.

2) • $x \leftarrow \dfrac{1}{2}(x_{min} + x_{max})$.

- Si $f'(x) > 0$, $x_{max} \leftarrow x$, sinon $x_{min} \leftarrow x$.

On répète l'opération jusqu'à un critère d'arrêt à préciser.

La longueur de l'intervalle est à chaque itération divisée par 2, la convergence est linéaire de taux 0,5.

Détermination de x_{min} et x_{max}

Une bonne méthode est la suivante (on suppose $f'(0) \neq 0$) :

- définir un pas de déplacement $h > 0$.
- si $f'(0) < 0$, faire :
  - ⋆ $x_{min} \leftarrow 0$
  - ⋆ tant que $f'(h) < 0$, faire
    - ▶ $x_{min} \leftarrow h$
    - ▶ $h \leftarrow 2h$
  - ⋆ $x_{max} \leftarrow h$
- sinon si $f'(0) > 0$, faire :
  - ⋆ $h \leftarrow -h$
  - ⋆ $x_{max} \leftarrow 0$
  - ⋆ tant que $f'(h) > 0$, faire
    - ▶ $x_{max} \leftarrow h$
    - ▶ $h \leftarrow 2h$
  - ⋆ $x_{min} \leftarrow h$

Sommaire

Optimisation non linéaire avec contraintes

Position du problème

Dans toute la section, les fonctions f, g_i $(1 \leqslant i \leqslant m)$ et h_j $(1 \leqslant j \leqslant p)$ sont de $\mathbb{R}^n$ dans $\mathbb{R}$ et **de classe $\mathbf{C}^1$**.

$$\begin{array}{ll} & \text{Minimiser } f(x) \\ \text{avec} & \left\{ \begin{array}{ll} \text{pour } i \in \{1, ..., m\}, & g_i(x) \leqslant 0 \\ \text{pour } j \in \{1, ..., p\}, & h_j(x) = 0 \end{array} \right. \\ \text{et} & x \in \mathbb{R}^n. \end{array} \qquad (P)$$

On pose $I = \{1, ..., m\}$ et $J = \{1, ..., p\}$.

Les conditions $g_i(x) \leqslant 0$ et $h_j(x) = 0$ s'appellent les *contraintes*.

Tout x appartenant à $\mathbb{R}^n$ qui vérifie le système des contraintes s'appelle *solution réalisable*.

L'ensemble des solutions réalisables est le *domaine réalisable*. On note X le domaine réalisable.

Si, pour $i \in \{1, 2, ..., m\}$ et pour $x \in X$, on a $g_i(x) = 0$, on dit que la contrainte g_i est *saturée* en x.

Théorème et définition

Le domaine réalisable X est un fermé de $\mathbb{R}^n$.

On supposera X non vide.

Théorème

Si X est borné, alors le problème (P) admet au moins une solution.

Preuve. Le domaine réalisable est un fermé, borné non vide de $\mathbb{R}^n$. Le résultat résulte d'un théorème de topologie : « toute fonction continue sur un fermé borné non vide de $\mathbb{R}^n$ est bornée et atteint ses bornes ».

Définition

Soit f une fonction de $\mathbb{R}^n$ dans $\mathbb{R}$. On dit que f est coercive si elle tend vers $+\infty$ quand $||x||$ tend vers $+\infty$.

Théorème

Théorème

Si la fonction f est coercive, le problème (P) admet au moins une solution.

Preuve. Soit x une solution réalisable.

f coercive $\Rightarrow \exists$ une boule B de $\mathbb{R}^n$ centrée sur l'origine telle que, pour tout x' non dans B, $f(x') > f(x)$.

L'ensemble $B \cap X$ étant un ensemble fermé et borné de $\mathbb{R}^n$, la fonction f atteint son minimum sur $B \cap X$. Ce minimum est aussi un minimum global pour le problème (P).

Définition

Définition

On dit qu'une direction d est admissible en $x^0 \in X$ s'il existe une fonction ϕ de $\mathbb{R}$ dans $\mathbb{R}^n$ telle que :

1. $\phi(0) = x^0$
2. pour tout $t > 0$ assez petit, $\phi(t) \in X$
3. la dérivée à droite de ϕ en 0 est d.

Soit $x^0 \in X$.

On note $A(x^0)$ l'ensemble des directions admissibles en x^0.

On pose $I_0(x^0) = \{i \in I \text{ vérifiant } g_i(x^0) = 0\}$.

Proposition

Si d est une direction admissible en x^0, alors :

(i) pour $i \in I_0(x^0)$, $d^{\mathrm{t}}.\nabla g_i(x^0) \leqslant 0$

(ii) pour $j \in J, d^{\mathrm{t}}.\nabla h_j(x^0) = 0$.

Preuve. Soit ϕ correspondant à la définition de d admissible.

(i) Si $g_i(x^0) = 0$, alors $g_i(\phi(t)) = td^{\mathrm{t}}.\nabla g_i(x^0) + t\epsilon(t)$ où $\epsilon(t) \to 0$ quand $t \to 0^+$ (formule de Taylor pour $t \mapsto g_i(\phi(t))$).

Pour $t > 0$ assez petit, $g_i(\phi(t)) \leqslant 0$: $d^{\mathrm{t}}.\nabla g_i(x^0) + \epsilon(t) \leqslant 0$, ce qui donne (i).

(ii) $h_j(\phi(t)) = h_j(x^0) + td^{\mathrm{t}}.\nabla h_j(x^0) + t\epsilon(t)$ où $\epsilon(t) \to 0$ quand $t \to 0^+$.

Pour $t > 0$ assez petit, $h_j(\phi(t)) = 0$ et $h_j(x^0) = 0$; on a donc pour $t > 0$ assez petit :

$$d^{\mathrm{t}}.\nabla h_j(x^0) + \epsilon(t) = 0, \text{ ce qui donne (ii).}$$

Qualification des contraintes

On note $B(x^0)$ l'ensemble des directions d vérifiant :

- pour $i \in I_0(x^0)$, $d^{\mathrm{t}}.\nabla g_i(x^0) \leqslant 0$

- pour $j \in J$, $d^{\mathrm{t}}.\nabla h_j(x^0) = 0$.

$A(x^0) \subseteq B(x^0)$.

Définition

On dit que les contraintes sont *qualifiées* en $x^0 \in X$ si toute direction dans $B(x^0)$ est limite d'une suite de directions de $A(x^0)$.

Lemme 1

On suppose que, pour $j \in J$, les fonctions h_j sont affines.

Soient $x^0 \in X$ et d une direction vérifiant :

- pour $i \in I_0(x^0), d^{\mathrm{t}}.\nabla g_i(x^0) < 0$
- pour $j \in J$, $d^{\mathrm{t}}.\nabla h_j(x^0) = 0$.

Alors d est une direction admissible en x^0.

Preuve. Pour $t \geqslant 0$, on pose : $\phi(t) = x^0 + td$.

$\phi(0) = x^0$ et $\phi'(0) = d$.

Pour $j \in J$: $h_j(\phi(t)) = h_j(x^0) + td^{\mathrm{t}}.\nabla h_j(x^0)$ (h_j affine).

$h_j(x^0) = 0$ et $d^{\mathrm{t}}.\nabla h_j(x^0) = 0$. D'où $h_j(\phi(t)) = 0$.

Pour $i \in I_0(x^0)$, $g_i(\phi(t)) = g_i(x^0) + t(d^{\mathrm{t}}.\nabla g_i(x^0) + \epsilon(t))$, où $\epsilon(t) \to 0$ quand $t \to 0$.

$g_i(x^0) = 0$ et $d^{\mathrm{t}}.\nabla g_i(x^0) < 0 \Rightarrow g_i(\phi(t)) \leqslant 0$ pour $t > 0$ petit.

Ainsi, la direction d est admissible.

Lemme 2

On suppose que, pour $j \in J$, les fonctions h_j sont affines.

Soit $x^0 \in X$. S'il existe une direction $\tilde{d}$ telle que :

- pour $i \in I_0(x^0)$, $\tilde{d}^{\mathrm{t}}.\nabla g_i(x^0) < 0$
- pour $j \in J, \tilde{d}^{\mathrm{t}}.\nabla h_j(x^0) = 0$,

alors les contraintes sont qualifiées en x^0.

Preuve. Soient $d \in B(x^0)$ et $\tilde{d}$ vérifiant les hypothèses du lemme.

Pour $\lambda \in [0, 1[$, soit $d_\lambda = \lambda d + (1 - \lambda)\tilde{d}$.

Si $i \in I_0(x^0)$: $d_\lambda^{\mathrm{t}}.\nabla g_i(x^0) = \lambda d^{\mathrm{t}}.\nabla g_i(x^0) + (1 - \lambda)\tilde{d}^{\mathrm{t}}.\nabla g_i(x^0) < 0$.

Si $j \in J$: $d_\lambda^{\mathrm{t}}.\nabla h_j(x^0) = \lambda d^{\mathrm{t}}.\nabla h_j(x^0) + (1 - \lambda)\tilde{d}^{\mathrm{t}}.\nabla h_j(x^0) = 0$.

Lemme 1 $\Rightarrow d_\lambda$ est une direction admissible.

λ_n : suite de nombres tendant vers 1 par valeurs inférieures $\rightarrow$ suite d_{λ_n} de directions admissibles qui tend vers d.

Les contraintes sont qualifiées en x^0.

Proposition

Si :

- les fonctions g_i sont convexes,
- les fonctions h_j sont affines,
- et il existe $\tilde{x} \in X$ avec,

pour tout $i \in I, g_i(\tilde{x}) < 0$

pour tout $j \in J, h_j(\tilde{x}) = 0$

alors les contraintes sont qualifiées en tout point de X.

Preuve. Soit $x^0 \in X$ et soit $\tilde{x}$ vérifiant les hypothèses.

$i \in I_0(x_0) : 0 > g_i(\tilde{x}) \geqslant g_i(x^0) + (\tilde{x} - x^0)^{\mathrm{t}}.\nabla g_i(x^0)$ (g_i convexe).

Comme $g_i(x^0) = 0 : (\tilde{x} - x^0)^{\mathrm{t}}.\nabla g_i(x^0) < 0$.

On pose $\tilde{d} = \tilde{x} - x^0$; on a $\tilde{d}^{\mathrm{t}}.\nabla g_i(x^0) < 0$.

Si $j \in \{1, ..., p\} : 0 = h_j(\tilde{x}) = h_j(x^0) + \tilde{d}^{\mathrm{t}}.\nabla h_j(x^0)$ (h_j affines)

d'où : $\tilde{d}^{\mathrm{t}}.\nabla h_j(x^0) = 0$.

Lemme 2 : les contraintes sont qualifiées en x^0.

Proposition

On suppose que, pour $j \in J$, les fonctions h_j sont affines.

Si, en $x^0 \in X$, l'ensemble des gradients

- $\nabla g_i(x^0)$ pour $i \in I_0(x^0)$

- $\nabla h_j(x^0)$ pour $j \in J$

sont linéairement indépendants,

alors les contraintes sont qualifiées en x^0.

Début de la preuve

Preuve. On considère le problème (Q) d'optimisation linéaire ci-dessous :

Maximiser $z = \sum_{i \in I_0(x^0)} \lambda_i$

$$\text{avec } \begin{cases} \sum_{j \in J} \mu_j \nabla h_j(x^0) - \sum_{i \in I_0(x^0)} \lambda_i \nabla g_i(x^0) = 0 \quad (Q) \\ \text{pour } i \in I_0(x^0), \lambda_i \geqslant 0 \end{cases}$$

Le problème (Q) est réalisable puisque la solution nulle l'est.

Supposons qu'il puisse prendre une valeur strictement positive.

Dans cette solution, au moins un λ_i $(i \in I_0(x^0))$ est non nul : les vecteurs $\nabla g_i(x^0)$ $(i \in I_0(x^0))$ et $\nabla h_j(x^0)$ $(j \in J)$ sont linéairement dépendants, contrairement à l'hypothèse.

Le maximum du problème (Q) vaut donc 0 et le problème (Q) admet une solution optimale.

Suite de la preuve

On considère le problème (R) d'optimisation linéaire ci-dessous :

$$\text{Minimiser } w = 0$$
$$\text{avec } \begin{cases} \text{pour } i \in I_0(x^0), d^{\mathrm{t}}.\nabla g_i(x^0) \leqslant -1 & . \\ \text{pour } j \in J, d^{\mathrm{t}}.\nabla h_j(x^0) = 0. & (R) \end{cases}$$

On peut facilement vérifier que les problèmes (Q) et (R) sont duaux l'un de l'autre.

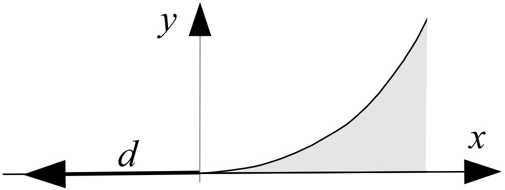
On utilise le théorème de la dualité pour l'optimisation linéaire : le problème (R) est réalisable. On note $\tilde{d}$ une solution réalisable de (R). Le lemme 2 permet de conclure.

Remarque. On peut montrer que cette proposition est encore exacte si on retire l'hypothèse que les fonctions h_j sont affines.

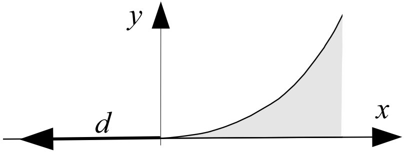
Un exemple de points où les contraintes ne sont pas qualifiées

On considère dans $\mathbb{R}^2$ le domaine défini par :

$$\begin{cases} y \leqslant x^3 \\ y \geqslant 0. \end{cases}$$



Exemple, suite



$$g_1(x,y) = y - x^3 \quad , \quad \left\{ \begin{array}{l} g_1(x,y) \leqslant 0 \\ g_2(x,y) \leqslant 0 \end{array} \right. .$$
$$g_2(x,y) = -y$$

Les contraintes g_1 et g_2 sont saturées en $(0,\ 0)$.

$\nabla g_1(0,0) = (0,1)^t$ et $\nabla g_2(0,0) = (0,-1)^t$.

Pour $d = (-1,0)^t$: $d^t.\nabla g_1(0,0) = 0$ et $d^t.\nabla g_2(0,0) = 0$

La direction d appartient à $B(0,0)$.

Or, la seule direction admissible en $(0,\ 0)$ est la direction $(1,0)^t$.
d n'est pas limite d'une suite de directions admissibles.

Théorème

On suppose que le problème admet un minimum local en un point x^* où les contraintes sont qualifiées. Alors, si $d \in B(x^*)$:
$d^{\mathrm{t}}.\nabla f(x^*) \geqslant 0$ (en particulier, aucune direction admissible en x^* n'est de descente).

Preuve. Soit (d_k) une suite de directions admissibles tendant vers d et soit ϕ_k la fonction associée à d_k.

Soit $t > 0$.

$f[\phi_k(t)] = f(x^*) + td_k^{\mathrm{t}}.\nabla f(x^*) + t\epsilon(t)$ où $\epsilon(t) \to 0$ quand $t \to 0^+$.

Si t est assez petit pour que $\phi_k(t)$ appartienne à X :
$f[\phi_k(t)] \geqslant f(x^*)$.

On a alors : $t[d_k^{\mathrm{t}}.\nabla f(x^*) + \epsilon(t)] \geqslant 0$ et donc $d_k^{\mathrm{t}}.\nabla f(x^*) + \epsilon(t) \geqslant 0$.

Par passage à la limite quand t tend vers 0, on obtient
$d_k^{\mathrm{t}}.\nabla f(x^*) \geqslant 0$.

Optimisation non linéaire avec contraintes

Condition nécessaire de Lagrange

On suppose qu'il n'y a **que des contraintes d'égalité** :

$$\begin{aligned}
&\text{Minimiser } f(x) \\
&\text{avec } x \in \mathbb{R}^n \\
&\text{et, pour } 1 \leqslant j \leqslant p, \quad h_j(x) = 0
\end{aligned}$$

où les fonctions f et h_j $(1 \leqslant j \leqslant p)$ sont de classe C^1.
La condition de Lagrange fournit une condition nécessaire pour qu'un élément de $\mathbb{R}^n$ soit un minimum local de (P).

Condition de Lagrange

Soit x^* un minimum local du problème. On suppose que les contraintes sont qualifiées en x^*. Alors il existe des réels $\mu_1, \mu_2, \ldots, \mu_p$ tels que :

$$\nabla f(x^*) = \sum_{j=1}^{p} \mu_j \nabla h_j(x^*)$$

Preuve de la condition de Lagrange

Condition de Lagrange : si x^* est un minimum local du problème où les contraintes sont qualifiées, il existe des réels $\mu_1, \mu_2, \ldots, \mu_p$ tels que : $\nabla f(x^*) = \sum_{j=1}^{p} \mu_j \nabla h_j(x^*)$.

Preuve.

E : le sous-espace de $\mathbb{R}^n$ engendré par les vecteurs $\nabla h_j(x^*)$.

$E^\perp$: le sous-espace orthogonal à E.

$\nabla f(x^*) = y + z$ avec $y \in E$ et $z \in E^\perp$.

Pour $1 \leqslant i \leqslant p$, $(-z)^{\mathrm{t}}.\nabla h_j(x^*) = 0$ puisque $-z$ appartient à $E^\perp$; d'où $-z \in B(x^*)$.

D'après le théorème précédent : $(-z)^{\mathrm{t}}.\nabla f(x^*) \geqslant 0$.

Or : $(-z)^{\mathrm{t}}.\nabla f(x^*) = (-z)^{\mathrm{t}}.y + (-z)^{\mathrm{t}}.z = (-z)^{\mathrm{t}}.z = -||z||^2$.

$-||z||^2 \geqslant 0 \Rightarrow z = 0$. D'où le théorème.

Condition de Lagrange suffisante ?

Le théorème suivant, démontré plus loin, donne des hypothèses pour lesquelles la condition de Lagrange est suffisante.

Théorème

La condition de Lagrange est suffisante lorsque f est convexe dans un ouvert contenant X et que les h_j $(1 \leqslant i \leqslant p)$ sont affines.

Optimisation non linéaire avec contraintes

Condition nécessaire de Karush, Kuhn et Tucker

On reprend le problème (P) initial :

$$(P) \quad \begin{array}{ll} & \text{Minimiser } f(x) \\ \text{avec} & \begin{cases} \text{pour } 1 \leqslant i \leqslant m, & g_i(x) \leqslant 0 \\ \text{pour } 1 \leqslant j \leqslant p, & h_j(x) = 0 \end{cases} \\ \text{et} & x \in \mathbb{R}^n \end{array}$$

f, g_i $(1 \leqslant i \leqslant m)$ et h_j $(1 \leqslant j \leqslant p)$ de classe C^1.

Théorème (Condition de Karush, Kuhn et Tucker)

On suppose que les contraintes sont qualifiées en x^* et que x^* est un minimum local du problème ; alors il existe :

- m nombres réels positifs ou nuls $\lambda_1, \lambda_2, ..., \lambda_m$
- p nombres réels $\mu_1, \mu_2, ..., \mu_p$

tels que : $\nabla f(x^*) = \sum_{j \in J} \mu_j \nabla h_j(x^*) - \sum_{i \in I_0(x^*)} \lambda_i \nabla g_i(x^*)$.

Première illustration, $n = 2, p = 0$, une seule contrainte d'inégalité saturée

x et y : les coordonnées d'un point.

On suppose que seule la contrainte $g(x, y) \leqslant 0$ est saturée en (x^*, y^*) : $g(x^*, y^*) = 0$.

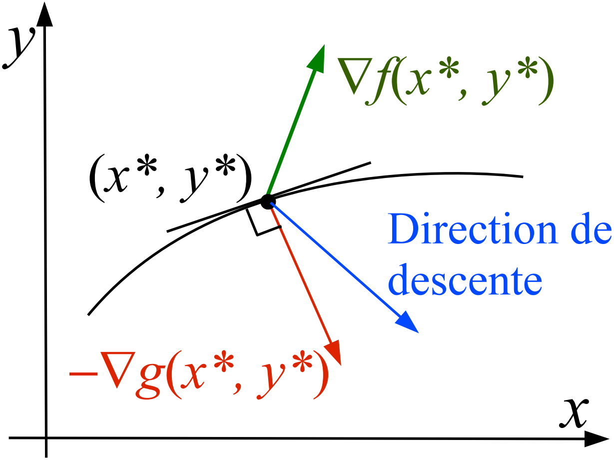
Le vecteur $-\nabla g(x^*, y^*)$ est perpendiculaire à la courbe d'équation $g(x, y) = 0$ et dirigée vers l'intérieur du domaine.

On rappelle que :

• une direction est une direction de descente si elle fait un angle obtus avec $\nabla f(x^*, y^*)$

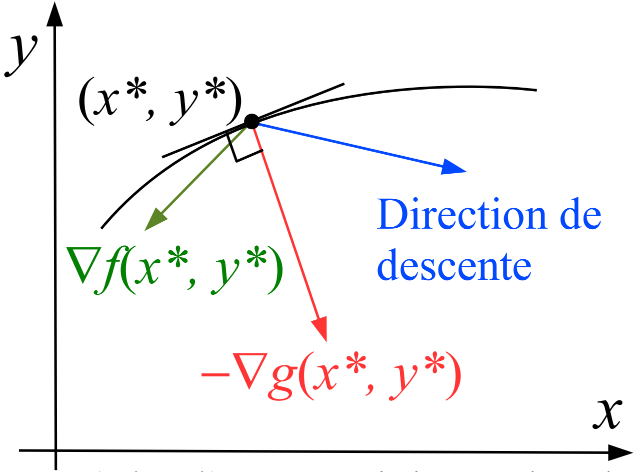
• une direction admissible fait un angle aigu avec $-\nabla g(x^*, y^*)$.

Première illustration, cas 1



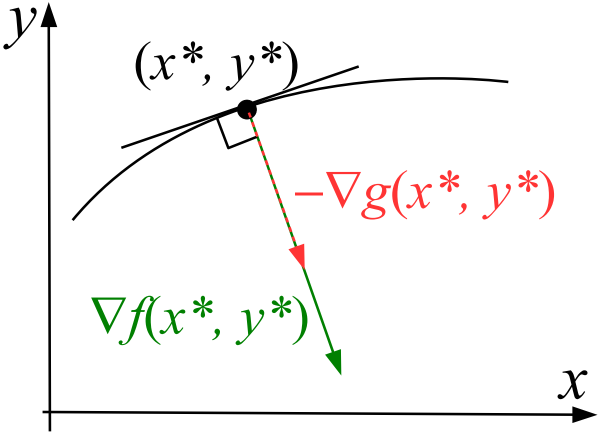
(x^*, y^*) : non minimum local

Première illustration, cas 2



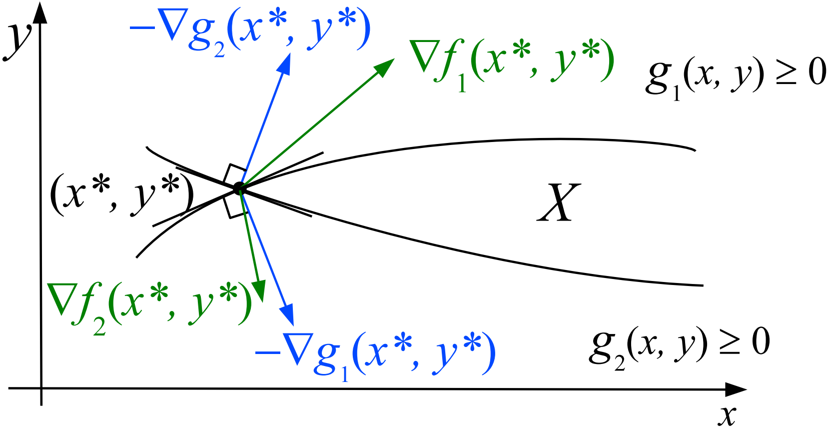
(x^*, y^*) : non minimum local

Première illustration, cas 3



(x^*, y^*) : condition satisfaite

Seconde illustration, $n = 2, p = 0$, deux contraintes d'inégalité saturées



Pour f_1, (x^*, y^*) peut être un minimum local mais pas pour f_2.

Condition de Karush, Kuhn et Tucker suffisante ?

Théorème

On suppose que les contraintes sont qualifiées en un point x^*. La condition de Karush, Kuhn et Tucker en x^* est suffisante pour avoir un minimum local lorsque simultanément f est convexe dans un voisinage de x^*, les g_i ($i \in I_0(x^*)$) sont convexes dans un voisinage de x^* et les h_j ($1 \leqslant j \leqslant p$) sont affines dans un voisinage de x^*.

Preuve. On suppose qu'il existe des nombres réels positifs ou nuls λ_i ($i \in I_0(x^*)$) et des nombres réels μ_j ($1 \leqslant j \leqslant p$) tels que :

$\nabla f(x^*) = \sum_{j \in J} \mu_j \nabla h_j(x^*) - \sum_{i \in I_0(x^*)} \lambda_i \nabla g_i(x^*)$.

B : boule de centre x^* dans laquelle f et les g_i sont convexes et les h_j affines.

$x \in B \cap X$; on va montrer que $f(x) \geqslant f(x^*)$, ce qui prouvera le théorème.

Preuve suite

Rappel : $\nabla f(x^*) = \sum_{j\in J} \mu_j \nabla h_j(x^*) - \sum_{i\in I_0(x^*)} \lambda_i \nabla g_i(x^*)$.

La convexité de f dans B induit : $f(x) \geqslant f(x^*) + (x - x^*)^{\mathrm{t}}.\nabla f(x^*)$.

$$KKT \Rightarrow f(x) \geqslant f(x^*) + \sum_{j\in J} \mu_j (x - x^*)^{\mathrm{t}}.\nabla h_j(x^*) \\ - \sum_{i\in I_0(x^*)} \lambda_i (x - x^*)^{\mathrm{t}}.\nabla g_i(x^*).$$

Pour $j \in J, (x - x^*)^{\mathrm{t}}.\nabla h_j(x^*) = h_j(x) - h_j(x^*) = 0$.

Soit $i \in I_0(x^*)$.

$g_i(x) \geqslant g_i(x^*) + (x - x^*)^{\mathrm{t}}.\nabla g_i(x^*)$ (convexité de g_i).

$(x - x^*)^{\mathrm{t}}.\nabla g_i(x^*) \leqslant g_i(x) - g_i(x^*)$.

Or, on a $\lambda_i \geqslant 0$; de plus, par hypothèse, $g_i(x^*) = 0$ et $g_i(x) \leqslant 0$.

$\lambda_i (x - x^*)^{\mathrm{t}}.\nabla g_i(x^*) \leqslant 0$.

On obtient $f(x) \geqslant f(x^*)$: f admet un minimum local en x^*.

Optimisation non linéaire avec contraintes

Méthode de descente

On considère le problème suivant :

$$\begin{array}{rl} & \text{Minimiser } f(x) \\ \text{avec,} & \text{pour } 1 \leqslant i \leqslant m,\ g_i(x) \leqslant 0. \end{array}$$

On choisit un point de départ $x^0 \in X$ et on construit de façon itérative une suite x^k de X vérifiant $f(x^{k+1}) < f(x^k)$, jusqu'à ce qu'on estime avoir obtenu une approximation satisfaisante.

À partir de x^k, on recherche une direction de descente d qui ne fasse pas sortir « immédiatement » de X.

On cherche alors, en se déplaçant dans la direction d, un point x^{k+1} de X meilleur que x^k.

On recommence à partir de x^{k+1} tant qu'un certain critère d'arrêt n'est pas vérifié.

Méthode de descente, choix de d

On peut résoudre le problème :

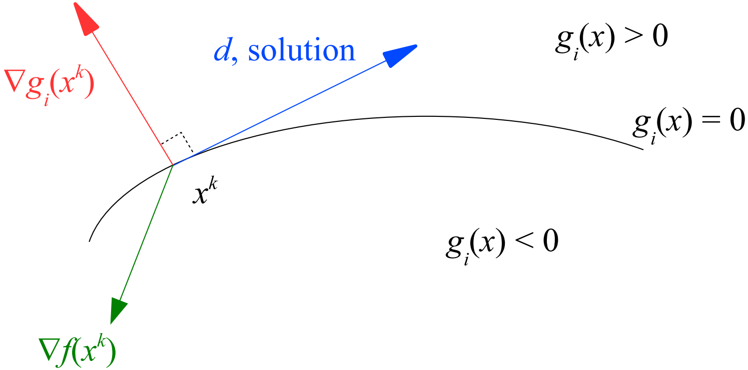
$$\begin{array}{ll} & \text{Minimiser } d^{\mathrm{t}}.f(x) \\ \text{avec} & d^{\mathrm{t}}.\nabla g_i(x^k) \leqslant 0 \text{ pour tout } i \text{ tel que } g_i(x^k) = 0 \\ \text{et} & d^{\mathrm{t}}.d = 1. \end{array}$$

On obtient alors pour d la direction dans $B(x^k)$ de plus grande descente.

On peut choisir de remplacer la condition $d^{\mathrm{t}}.d = 1$ par la condition : $-1 \leqslant d_i \leqslant 1$ $(1 \leqslant i \leqslant n)$ pour avoir un problème linéaire (la direction retenue ne sera pas exactement la direction de plus grande descente).

Méthode de descente, un problème

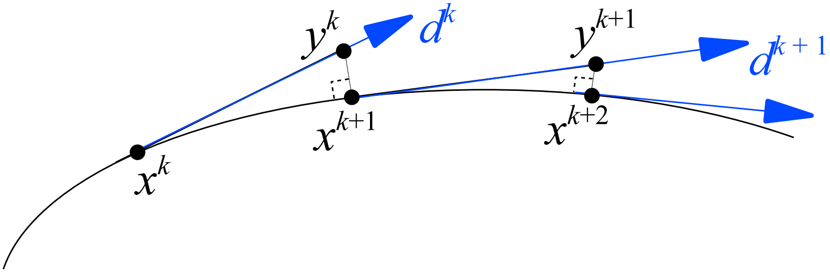
On considére l'exemple représenté ci-dessous.



Tout déplacement dans la direction d fait sortir de X.

Méthode de descente, projection

Il faut une procédure de projection pour que x^{k+1} soit dans X.



Optimisation non linéaire avec contraintes

Généralités

On suppose dans toute cette partie que :

- les fonctions g_i $(1 \leqslant i \leqslant m)$ sont convexes,

- les fonctions h_j $(1 \leqslant j \leqslant p)$ sont affines,

- la fonction f est convexe.

Définition

Soit C une partie de $\mathbb{R}^n$. On dit que C est convexe si, pour tout couple (x, x') de points de C, le segment d'extrémités x et x' est contenu dans C.

Remarque

Le domaine réalisable du problème (P), c'est-à-dire :

$$\{x \in \mathbb{R}^n | g_i(x) \leqslant 0 \text{ pour } 1 \leqslant i \leqslant m \text{ et } h_j(x) = 0 \text{ pour } 1 \leqslant j \leqslant p\}$$

est convexe.

Théorème 1

Si f est strictement convexe, le problème (P) admet au plus une solution optimale.

Preuve

On suppose qu'il existe dans C deux solutions optimales x et y.

On pose : $z = \dfrac{x+y}{2}$.

Convexité du domaine $\Rightarrow z \in C$

Stricte convexité de $f \Rightarrow f(z) < \dfrac{f(x)+f(y)}{2}$.

On a donc soit $f(z) < f(x)$, soit $f(z) < f(y)$, contradiction avec l'optimalité supposée de x et y.

Conséquences

Rappel : domaine borné $\Rightarrow$ au moins une solution optimale.

Théorème 2

Si le domaine réalisable est borné et si f est strictement convexe, le problème (P) admet une unique solution.

Rappel : f coercive $\Rightarrow$ au moins une solution optimale.

Théorème 3

Si f est strictement convexe et coercive, le problème (P) admet une unique solution.

Théorème 4

On suppose qu'on a un minimum local en un point x^* où les contraintes sont qualifiées. Alors le problème (P) admet un minimum global en x^*.

Preuve. Soit x une solution réalisable du problème (P).

Soit ψ la fonction définie sur l'intervalle $[0, 1]$ par :
$\psi(t) = f[x^* + t(x - x^*)]$.

On a : $\psi(0) = f(x^*)$ et $\psi(1) = f(x)$.

$\psi'(0) = (x - x^*)^t.\nabla f(x^*)$.

La direction $d = x - x^*$ est admissible (car X est convexe).

Donc $\psi'(0) \geqslant 0$ (aucune direction admissible n'est de descente).

Convexité de $f \Rightarrow$ convexité de ψ sur $[0, 1]$.

$\psi'(0) \geqslant 0 \Rightarrow \psi(1) \geqslant \psi(0) \Rightarrow f(x) \geqslant f(x^*)$.

Théorème 5

On suppose que la condition de Karush, Kuhn et Tucker est vérifiée en un point x^* où les contraintes sont qualifiées. Avec les hypothèses de cette partie, le point x^* est un minimum global de (P). De plus, si f est strictement convexe, x^* est l'unique point où (P) atteint le minimum global.

Preuve. D'après le théorème donnant les hypothèses pour que la condition KKT soit suffisante, le problème (P) atteint un minimum local en x^*.

Théorème 4 $\Rightarrow$ le problème (P) admet un minimum global en x^*.

Si, de plus, f est strictement convexe, le théorème 1 $\Rightarrow$ (P) atteint son minimum global uniquement en x^*.

Optimisation non linéaire avec contraintes

Algorithme basique

Recherche du minimum d'une fonction $f : \mathbb{R}^n \mapsto \mathbb{R}$ de classe C^1 sur un domaine fermé convexe X.

On va « linéariser » f, en l'approchant par son développement de Taylor à l'ordre 1. Algorithme :

- $x^0 \leftarrow$ un point quelconque
- $k \leftarrow 0$
- répéter

  $x^{k+1} \leftarrow$ point qui minimise $f(x^k) + (x - x^k)^{\mathrm{t}}.\nabla f(x^k)$ sur X.

  $k \leftarrow k+1$

jusqu'à ce qu'un test d'arrêt à préciser soit vérifié.

Remarques.

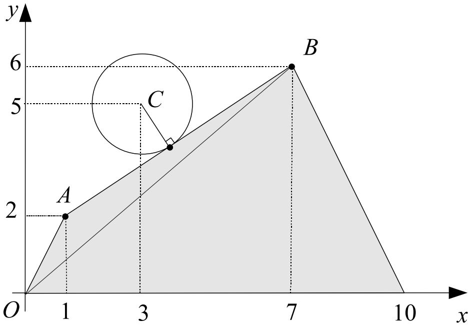
1) x^{k+1} est aussi un point qui minimise $x^{\mathrm{t}}.\nabla f(x^k)$ sur X.

2) Si le domaine est un polyèdre, la détermination de x^{k+1} est un problème d'optimisation linéaire.

Algorithme basique : exemple et difficulté

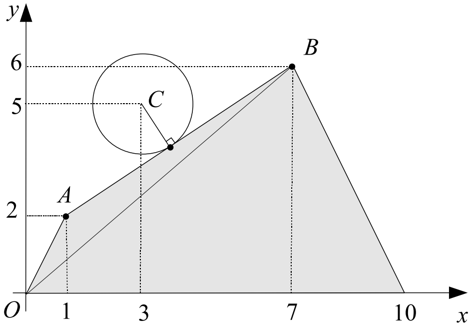
$$(P_0) \qquad \text{Minimiser } f(x,y) = (x-3)^2 + (y-5)^2$$

$$\text{avec les contraintes : } \begin{cases} y - 2x \leqslant 0 \\ 2x + y - 20 \leqslant 0 \\ -2x + 3y - 4 \leqslant 0 \\ y \geqslant 0. \end{cases}$$



f est constante sur des cercles centrés sur le point $C = (3,\ 5)$.
f est minimum pour le point le plus proche de C : $(\frac{49}{13}, \frac{50}{13})$.

Algorithme basique : exemple et difficulté, suite



$\nabla f(x,y) = (2x-6, 2y-10)^t$.
On démarre à partir de O, $\nabla f(0,0) = (-6,-10)^t$.
Le minimum de $-6x-10y$ est atteint en $B=(7,6)$.
À partir du sommet B : $\nabla f(7,6) = (8,2)^t$. Le minimum de $8x+2y$ est atteint en $A=(1,2)$.
À partir de A : $\nabla f(1,2) = (-4,-6)$. Le minimum de $-4x-6y$ est atteint en B.
La méthode ne converge pas.

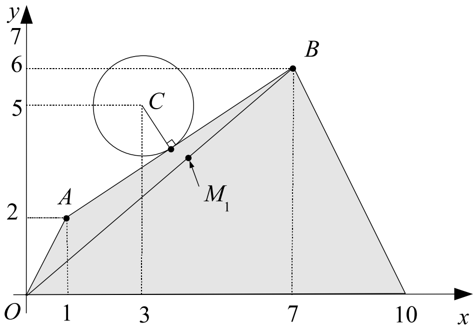
Linéarisation : méthode de Frank et Wolfe

La méthode de Frank et Wolfe s'applique à la minimisation d'**une fonction de classe C^1 convexe sur un domaine X convexe et compact**.

- $x^0 \leftarrow$ un point quelconque
- $k \leftarrow 0$
- répéter
  - $\star$ $\tilde{x^k} \leftarrow$ un point qui minimise $x^t.\nabla f(x^k)$ sur X
  - $\star$ $x^{k+1} \leftarrow$ un point qui minimise f sur le segment $[x^k, \tilde{x^k}]$
  - $\star$ $k \leftarrow k+1$

jusqu'à ce qu'un test d'arrêt à préciser soit vérifié.

Frank et Wolfe pour le problème précédent, étape 1



$\nabla f(x, y) = (2x - 6, 2y - 10)^{t}$.

Étape 1 : à partir du point (0, 0).

Minimum de $-6x - 10y$ sur $X \rightarrow (\tilde{x_0}, \tilde{y_0}) = B = (7, 6)$.

Minimum de f sur le segment $[O, B]$?

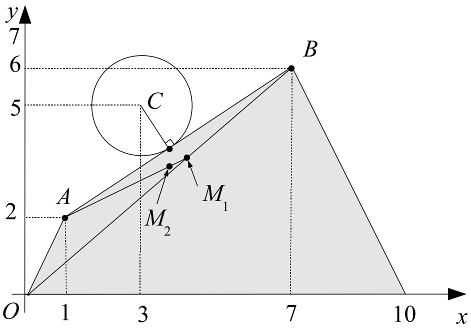
$x = 7t, y = 6t$ $(0 \leqslant t \leqslant 1)$, $\phi_1(t) = f(7t, 6t)$.

Après calcul : le minimum de ϕ_1 est obtenu pour $t_1 = \frac{3}{5} = 0,6$.

La fonction f atteint son minimum sur le segment $[O, B]$ au point :

$M_1 = (x_1, y_1) = (7 \times 0,6\ ; 6 \times 0,6) = (4,2\ ; 3,6)$.

Frank et Wolfe pour le problème précédent, étape 2



$\nabla f(x, y) = (2x - 6, 2y - 10)^{\mathrm{t}}$.

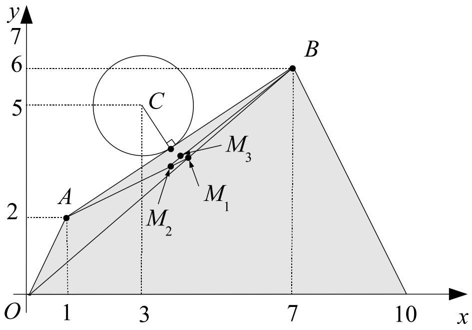
Étape 2 : à partir du point $M_1 = (4,2\,;3,6)$.

Minimum de $(2 \times 4,2 - 6)x + (2 \times 3,6 - 10)y = 2,4x - 2,8y$ sur $X \to (\tilde{x}_1, \tilde{y}_1) = A = (1, 2)$.

Minimum de f sur le segment $[M_1, A]$?

Après paramétrisation et calcul : la fonction f atteint son minimum sur le segment $[M_1, A]$ au point : $M_2 = (x_2, y_2) = (3,8\,;3,4)$.

Frank et Wolfe pour le problème précédent, étape 3



$\nabla f(x, y) = (2x - 6, 2y - 10)^t$.

Étape 3 : à partir du point $M_2 = (3,8\ ; 3,4)$.

Minimum de $(2 \times 3,8 - 6)x + (2 \times 3,4 - 10)y = 1,6x - 3,2y$ sur $X \to (\tilde{x_2}, \tilde{y_2}) = B = (7, 6)$.

Minimum de f sur le segment $[M_2, B]$?

Après paramétrisation et calcul : la fonction f atteint son minimum sur le segment $[M_2, B]$ au point : $M_3 = (x_3, y_3) \approx (4,101\ ; 3,64)$.

Rappel : point optimal $= (49/13; 50/13) \approx (3,769; 3,846)$.

Convergence de la méthode de Frank et Wolfe

Théorème

Soit f une fonction de $\mathbb{R}^n$ dans $\mathbb{R}$, de classe C^1, strictement convexe. Soit X un polyèdre convexe compact de $\mathbb{R}^n$. La méthode de Frank et Wolfe appliquée au problème (P) de la minimisation de f sur X converge vers le minimum global.