PROSEMINAIRE 2009
Théorie de l'optimisation avec contraintes
Introduction, Chapitre 12.1, 12.2

Céline Schneider

3 décembre 2009

Table des matières

1 Introduction

Nous allons ici nous intéresser à la minimisation (pour la maximisation, prendre l'inverse de la fonction à maximiser) de fonctions soumises à des contraintes sur les variables. Une formulation générale de ces problèmes est :

$$\min_{x\in\mathbb{R}^n} f(x) \qquad \text{sous contraintes} \begin{cases} c_i(x) = 0, i \in \mathcal{E}, \\ c_i(x) \geq 0, i \in \mathcal{I} \end{cases} \tag{1}$$

avec f et c_i des fonctions différentiables dont leurs différentielles sont continues, évaluées réellement dans un sous-ensemble de $\mathbb{R}^n$ et où $\mathcal{E}$ et $\mathcal{I}$ sont deux ensembles finis d'indices. Nous appelons f la fonction *objectif*. Les c_i avec $i \in \mathcal{E}$ sont des *contraintes d'égalité* alors que les c_i avec $i \in \mathcal{I}$ sont des *contraintes d'inégalité*. L'ensemble Ω de tous les points x satisfaisant les contraintes est défini par :

$$\Omega = \{x \mid c_i(x) = 0, i \in \mathcal{E}; c_i(x) \geq 0, i \in \mathcal{I}\} \tag{2}$$

Nous pouvons alors réécrire (1) plus simplement par

$$\min_{x\in\Omega} f(x) \tag{3}$$

Dans ce chapitre, nous allons essayer de trouver et de caractériser (3). Dans le cas d'une optimisation sans contrainte, nous avons que la solution x^* est caractérisée par des conditions nécessaires et des conditions suffisantes qui sont :

- Conditions nécessaires : les problèmes de minimums locaux ont $\nabla f(x^*) = 0$ et $\nabla^2 f(x^*)$ définie semi positive ;
- Conditions suffisantes : chaque point x^* avec $\nabla f(x^*) = 0$ et $\nabla^2 f(x^*)$ définie positive est un strict minimum local de f.

Notre but dans ce chapitre est d'obtenir des conditions similaires afin de caractériser les solutions de problèmes d'obtimisation avec contraintes.

1.1 Solutions locales et globales

Nous en avons déjà fait l'expérience à maintes reprises, trouver une solution globale minimale est très souvent difficile. Comme nous pouvons le sentir intuitivement, la recherche de ce minimun peut être améliorée par l'ajout de contraintes. En effet, l'ajout de contraintes permet bien souvent de diminuer l'ensemble des minimums locaux d'une fonction et ainsi faciliter la sélection de la bonne solution. Malheureusement, les contraintes peuvent aussi rendre les choses plus difficiles. Prenons un exemple simple pour nous en convaincre :

$$\min_{x\in\mathbb{R}^n} \|x\|_2^2, \qquad \text{sous contrainte } \|x\|_2^2 \geq 1$$

Sans la contrainte, ce problème possède la solution minimale unique $x^* = 0$. En ajoutant la contrainte, n'importe quel vecteur avec $\|x\|_2 = 1$ résoud le problème.Nous avons donc une infinité de solutions, à partir du moment où $n \geq 2$.
Les définitions des solutions locales dans le cas de l'optimisation avec contraintes sont de simples extensions des définitions pour le cas sans contrainte, sauf que maintenant nous

nous intéressons aux points admissibles dans le voisinage de x^* pour obtenir les définitions suivantes :

> Un vecteur x^* est une *solution locale* du problème (3) si $x^* \in \Omega$ et s'il existe un voisinage $\mathcal{N}$ de x^* tel que $f(x) \geq f(x^*)$ pour $x \in \mathcal{N} \cap \Omega$.
>
> Un vecteur x^* est une *solution locale stricte* si $x^* \in \Omega$ et s'il existe un voisinage $\mathcal{N}$ de x^* tel que $f(x) > f(x^*)$ pour $x \in \mathcal{N} \cap \Omega$.
>
> Un point x^* est *une solution locale isolée* si $x^* \in \Omega$ et s'il existe un voisinage $\mathcal{N}$ de x^* tel que x^* est la seule solution locale dans $\mathcal{N} \cap \Omega$.

1.2 La régularité des fonctions

Par *lisseur*, on entend que la fonction soit continue, différentiable et que ses différentielles soient continues. La lisseur des fonctions objectif ainsi que des contraintes est importante pour la caractérisation des solutions. Elle assure qu'elles se comportent d'une manière attendue et permet l'utilisation d'algorithmes pour la recherche de bonnes directions.
Des fonctions non lisses ont des sauts ou alors des trous. Si nous dessinons la région admissible d'un problème d'optimisation sous contraintes, nous aurons donc des sauts pour lesquels il ne sera pas possible d'appliquer les règles habituelles. Pourtant, cela ne veut pas dire que les contraintes qui décrivent ces régions seront forcément non lisses. En effet, une partie non lisse pourra souvent être décrite par un ensemble de contraintes lisses.

2 Exemples

Pour introduire les principes de base de l'optimisation avec contraintes, nous allons procéder par étapes. Nous allons tout d'abord voir le cas d'une seule contrainte d'égalité, puis d'une seule contrainte d'inégalité et enfin généraliser le tout pour plusieurs contraintes d'inégalité.
Notons tout d'abord qu'à un point x admissible comme solution, la contrainte d'inégalité $i \in \mathcal{I}$ est dite *active* si $c_i(x) = 0$ et *inactive* si la stricte inégalité $c_i(x) > 0$ est satisfaite.

2.1 Une seule contrainte d'égalité

Pour notre premier cas, prenons un exemple simple avec deux variables :

$$\min x_1 + x_2 \qquad \text{sous contrainte} \qquad x_1^2 + x_2^2 - 2 = 0 \tag{4}$$

> Si nous voulons exprimer cela à la manière de (1), nous avons que $f(x) = x_1 + x_2$, $\mathcal{I} = \emptyset$, $\mathcal{E} = \{1\}$, et $c_1(x) = x_1^2 + x_2^2 - 2$. Nous pouvons donc voir, à l'aide d'un dessin, que les x admissibles pour ce problème sont sur un cercle de rayon $\sqrt{2}$ centré à l'origine. La solution x^* est $(-1, -1)^T$. Partout ailleurs sur le cercle, il est possible de diminuer f tout en restant sur le cercle, donc en gardant la fonction de contrainte active.
>
> En reprenant éventuellement le graph de la contrainte, nous voyons qu'à la solution x^* la *normale à la contrainte* $\nabla c_1(x^*)$ est parallèle à $\nabla f(x^*)$. Ce qui veut dire qu'il existe un scalaire λ_1^* tel que

$$\nabla f(x^*) = \lambda_1^* \nabla c_1(x^*). \tag{5}$$

Nous avons donc, dans cet exemple, notre $\lambda_1^* = -\frac{1}{2}$.

On peut dériver (1) en examinant l'approximation par la série de Taylor de premier ordre de la fonction objectif ainsi que des fonctions de contraintes. On demande donc que $c_1(x+d) = 0$ et on obtient :

$$0 = c_1(x+d) \approx c_1(x) + \nabla c_1(x)^T d = \nabla c_1(x)^T d \tag{6}$$

puisque $c_1(x) = 0$. De là, la direction d garde la faisabilité tout en repectant la contrainte c_1 lorsque

$$\nabla c_1(x)^T d = 0 \tag{7}$$

est satisfait.
Identiquement, une direction d'amélioration doit produire une diminution de f, tel que

$$0 > f(x+d) - f(x) \approx f(x) + \nabla f(x)^T d - f(x) = \nabla f(x)^T d$$

donc, du premier ordre,

$$\nabla f(x)^T d < 0. \tag{8}$$

Si une telle direction d satisfaisant (7) et (8), cela veut dire que nous avons une amélioration de notre point x qui est possible. Il en suit qu'une *condition nécessaire* pour une solution optimale dans un problème de minimisation est qu'il n'existe pas de telle direction d.
La seule façon d'éviter une telle direction est quand $\nabla f(x)$ et $\nabla c_1(x)$ sont parallèles, i.e. si la condition $\nabla f(x) = \lambda_1 \nabla c_1(x)$ en x pour un scalaire λ_1. Dans le cas contraire, si ce d existe, il satisfait les équations (7) et (8) et est défini par

$$d = -\left(I - \frac{\nabla c_1(x) \nabla c_1(x)^T}{\|\nabla c_1(x)\|^2} \right) \nabla f(x) \tag{9}$$

Introduisons maintenant la *fonction de Lagrange* qui est

$$\mathcal{L}(x, \lambda_1) = f(x) - \lambda_1 c_1(x). \tag{10}$$

Nous avons également que $\nabla_x \mathcal{L}(x, \lambda_1) = \nabla f(x) - \lambda_1 \nabla c_1(x)$. Nous pouvons donc reformuler la condition (5) optimale comme suit :

$$\nabla_x \mathcal{L}(x^*, \lambda_1^*) = 0, \tag{11}$$

avec x^* qui est la solution et λ_1^* qui est un scalaire appelé multiplicateur de Lagrange. En d'autre mots, nous pouvons chercher les minimums d'une fonction avec contrainte d'égalité en cherchant les points stationnaires de la fonction de Lagrange.
La condition (5) est nécessaire mais pas suffisante pour trouver la solution du problème (4). En effet, la condition (5) est également satisfaite au point $x = (1, 1)$ avec un $\lambda_1 = \frac{1}{2}$. Mais ce point ne minimise pas f, il la maximise. Malheureusement, pour régler ce problème, il ne suffit pas d'imposer le signe de λ_1 comme étant négatif. Pour le voir, remplaçons simplement la contrainte $x_1^2 + x_2^2 - 2 = 0$ par sa négative $2 - x_1^2 - x_2^2 = 0$. La solution reste la même, mais $\lambda_1^* = -\frac{1}{2}$ change en $\lambda_1^* = \frac{1}{2}$

2.2 Une seule contrainte d'inégalité

Reprenons l'exemple précédent, mais on modifiant le signe d'égalité en un signe d'inégalité $\geq$. Remarquons que si nous avons une contrainte de forme $c_i(x) \leq 0$, il est facile d'obtenir une inégalité de forme $c_i(x) \geq 0$ par un réarrangement de l'inégalité. Nous considérerons donc dans les pages à venir, des inégalités $c_i(x) \geq 0$:

$$\min x_1 + x_2 \qquad \text{sous contrainte} \qquad x_1^2 + x_2^2 - 2 \geq 0 \tag{12}$$

A la manière de (1), nous avons que $f(x) = x_1 + x_2$, $\mathcal{E} = \emptyset$, $\mathcal{I} = \{1\}$, et $c_1(x) = x_1^2 + x_2^2 - 2$. Si nous dessinons la contrainte, nous obtenons à nouveau un cercle, mais cette fois l'intérieur fait également parti des x admissibles. Par intuition, nous remarquons que la solution optimale est toujours $(-1, -1)$ pour un $\lambda_1 = \frac{1}{2}$. La différence avec avant, c'est qu'ici le signe de λ_1 est important, comme nous allons le voir maintenant.

Comme avant, le point x n'est pas optimal si nous pouvons trouver une direction d qui garde x admissible et qui diminue f. Toujours comme précédemment, la direction d décroit la fonction objectif f si $\nabla f(x)^T d < 0$, i.e.

$$0 \leq c_1(x + d) \approx c_1(x) + \nabla c_1(x)^T d,$$

donc, avec le premier ordre, si

$$c_1(x) + \nabla c_1(x)^T d \geq 0 \tag{13}$$

Pour déterminer si une direction d telle que décrite ci-dessus existe, considérons deux cas :

1. x est strictement dans le cercle : nous avons l'inégalité stricte $c_1(x) > 0$. Dans ce cas, n'importe quel vecteur d satisfait la condition (13), pour autant que sa longueur soit suffisamment petite. Ce d est obtenu en posant :

   $$d = -c_1(x) \frac{\nabla f(x)}{\|\nabla f(x)\| \, \|\nabla c_1(x)\|}$$

   La seule situation pour laquelle une telle direction ne peut pas exister est quand

   $$\nabla f(x) = 0 \tag{14}$$

2. x est sur le périmètre du cercle : nous avons une contrainte d'égalité $c_1(x) = 0$. Là, les conditions (8) et (13) deviennent

   $$\nabla f(x)^T d < 0, \qquad \nabla c_1(x)^T d \geq 0.$$

   Ces conditions ne sont pas réunies que lorsque $\nabla f(x)$ et $\nabla c_1(x)$ pointent dans la même direction, donc quand

   $$\nabla f(x) = \lambda_1 \nabla c_1(x), \qquad \text{pour un } \lambda_1 \geq 0 \tag{15}$$

   Le signe du multiplicateur est important ici. Si le signe était négatif, $\nabla c_1(x)\ et \nabla f(x)$ pointeraient dans des directions opposées.

Les conditions d'optimalité tirées des cas 1 et 2 peuvent à nouveau être résumées par la fonction de Lagrange. Lorsqu'il n'existe pas de direction d améliorant f à un certain point x^*, nous avons

$$\nabla_x \mathcal{L}(x^*, \lambda_1^*) = 0, \qquad \text{pour un } \lambda_1^* \geq 0 \tag{16}$$

où nous demandons encore que

$$\lambda_1^* c_1(x^*) = 0 \tag{17}$$

Cette condition est appelée *condition complémentaire* et implique que le multiplicateur de Lagrange peut être strictement positif seulement lorsque la contrainte correspondante et active. Des conditions de ce genre jouent un rôle central dans l'optimisation avec contrainte, comme nous le verrons dans les pages à venir. Dans le cas 1, nous avons que $c_1(x^*) > 0$, nous devons donc avoir un $\lambda_1^* = 0$. Par conséquent, la fonction de Lagrange en est réduite à $\nabla f(x^*) = 0$ comme demandé en (14). Dans le cas 2, la condition $\lambda_1^* c_1(x^*) = 0$ fait que $\lambda_1^* \geq 0$ donc la fonction de Lagrange devient la condition (15).
Nous pouvons faire de même avec une fonction soumise à deux, trois, quatre, n contraintes d'(in-)égalité. Mais passons directement à la généralisation.

3 Conditions d'optimalité du premier ordre

3.1 Etablissement des conditions du premier ordre

Les exemples vus précédemment nous ont inspiré un certain nombre de conditions utiles pour l'optimisation avec contraintes, dont la fonction de Lagrange, la non négativité de λ_i et la condition complémentaire $\lambda_i c_i(x) = 0$ dans le cas d'inégalités. Généralisons maintenant ces conditions de manière rigoureuse.
En général, la fonction de Lagrange, aussi appelée lagrangien, pour les problèmes d'optimisation avec contrainte est définie ainsi :

$$\mathcal{L}(x, \lambda) = f(x) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i c_i(x). \tag{18}$$

L'ensemble actif $\mathcal{A}(x)$ à chaque point admissible x est l'union de l'ensemble $\mathcal{E}$ et des indices des contraintes d'inégalités actives, donc

$$\mathcal{A}(x) = \mathcal{E} \cup \{i \in \mathcal{I} \mid c_i(x) = 0\} \tag{19}$$

Attardons-nous maintenant un peu plus sur les gradients des contraintes. Le vecteur $\nabla c_i(x)$ est souvent appelé la *normale* à la contrainte c_i au point x, car le gradient est souvent perpendiculaire au contour de la contrainte c_i. De plus, dans le cas d'inégalités, ce vecteur pointe à l'intérieur de la région possible. Il arrive cependant que ∇c_i disparaisse à cause de la représentation algébrique de c_i, et ainsi le terme $\lambda_i c_i(x) = 0$ pour toute valeur de λ_i et ne puisse ainsi jouer aucun rôle dans le gradient du lagrangien.
En général, nous émettons l'hypothèse appelée *qualification de contrainte* pour s'assurer que de tels comportements n'ont pas lieu au point x en question. Une de ces qualifications de contrainte est celle-ci :

Définition 1 (LICQ)**.** Soient le point x^* et l'ensemble actif $\mathcal{A}(x^*)$. On dit que la *qualification de contraintes d'indépendance linéaire* (LICQ) tient si l'ensemble actif des gradients des contraintes $\{\nabla c_i(x^*) \mid i \in \mathcal{A}(x^*)\}$ est linéairement indépendant.

Remarquons que si cette condition est réalisée, aucun des gradients ne peut être zéro.
Cette condition supplémentaire nous permet, additionnée des conditions précédentes, de formuler le théorème suivant.

Theorème 2 (Conditions nécessaires de premier ordre)**.** Soit x^* une solution locale de (1) et supposons que la LICQ tienne en x^*.
Alors, il existe un vecteur multiplicateur de Lagrange λ^*, avec composantes λ_i^*, $i \in \mathcal{E} \cup \mathcal{I}$, tel que les conditions suivantes soient satisfaites en (x^*, λ^*)

$$\begin{aligned}
\nabla_x \mathcal{L}(x^*, \lambda^*) &= 0, && && \text{(20a)}\\
c_i(x^*) &= 0, && \text{pour tout } i \in \mathcal{E}, && \text{(20b)}\\
c_i(x^*) &\geq 0, && \text{pour tout } i \in \mathcal{I}, && \text{(20c)}\\
\lambda_i^* &\geq 0, && \text{pour tout } i \in \mathcal{I}, && \text{(20d)}\\
\lambda_i^* c_i(x^*) &= 0, && \text{pour tout } i \in \mathcal{I} \cup \mathcal{E}, && \text{(20e)}
\end{aligned}$$

Ces conditions sont appelées *conditions de premier ordre* car elles concernes les gradients (donc les dérivées du premier ordre) des fonctions objectif et des contraintes ou encore *conditions de Karush-Kuhn-Tucker.* La condition complémentaire impose aux λ_i correspondant aux contraintes d'inégalité d'être zéro. Nous pouvons donc ne pas nous occuper des termes dont l'indice $i \notin \mathcal{A}(x^*)$ puisque son poids est nul et écrire la condition (20a) comme

$$0 = \nabla_x \mathcal{L}(x^*, \lambda^*) = \nabla f(x^*) - \sum_{i \in \mathcal{A}(x^*)} \lambda_i^* \nabla c_i(x^*). \tag{21}$$

Comme dit précédemment, la condition (20e) est appelée condition de complémentarité puisqu'elle implique que pour $i \in \mathcal{I} \cup \mathcal{E}, \lambda_i^* = 0$ ou $c_i(x^*) = 0$. Un cas particulier est :

Définition 3 (La complémentarité stricte)**.** Etant donné une solution x^* associée au vecteur de Lagrange λ^*, la *condition de complémentarité stricte* tient si exactement un des λ_i^* et $c_i(x^*)$ est nul, pour chaque index $i \in \mathcal{I}$.
Autrement dit, nous avons $\lambda_i^* > 0$, pour tout $i \in \mathcal{I} \cap \mathcal{A}(x^*)$.

Nous avons donc que, pour un problème de minimisation sous contrainte donné et une solution x^*, il peut y avoir beaucoup de vecteur λ^* qui satisfont les conditions de Karush-Kuhn-Tucker. Cependant, avec la LICQ, la solution optimale λ^* est unique.
La preuve du théorème ci-dessus est assez complexe et ne sera pas exposée dans ce chapitre.

3.2 Le multiplicateur de Lagrange

Dans cette dernière partie, nous allons nous intéresser d'un peu plus près à la signification du multiplicateur de Lagrange. En effet, son utilité doit maintenant être assez évidente, mais sa représentation l'est peut-être un peu moins.
En fait, les valeurs prises à l'optimum par les multiplicateurs de Lagrange λ_i^* mesurent la sensibilité de la fonction objectif par rapport aux différentes contraintes c_i. Avec une petite analyse, nous pouvons distinguer deux cas :

1. Contrainte inactive : On a donc $c_i > 0$, et $i \notin \mathcal{A}(x^*)$. La solution x^* et la valeur de $f(x^*)$ sont presques identiques,que la contrainte soit présente ou pas. Si nous perturbons un tout petit peu c_i, elle sera toujours active et x^* sera toujours une solution locale du problème. Ceci est confirmé par la condition (20e) qui nous dit que pour de telles contraintes, le multiplicateur de Lagrange est nul et indique donc que la contrainte n'a pas d'influence notable sur la fonction f.

2. Conctrainte active : Au lieu de considérer $c_i(x^*) \geq 0$, perturbons un peu le coté droite de l'inégalité de manière à avoir $c_i(x) \geq -\epsilon \|\nabla c_i(x^*)\|$. Supposons que ϵ soit assez petit pour que la solution perturbée $x^*(\epsilon)$ possède toujours le même ensemble de contraintes actives et que le multiplicateur de Lagrange ne soit pas beaucoup touché par cette perturbation. On trouve que :

$$\begin{aligned} -\epsilon \|\nabla c_i(x^*)\| &= c_i(x^*(\epsilon)) - c_i(x^*) \approx (x^*(\epsilon) - x^*)^T \nabla c_i(x^*), \\ 0 &= c_j(x^*(\epsilon)) - c_j(x^*) \approx (x^*(\epsilon) - x^*)^T \nabla c_j(x^*), \quad \forall j \in \mathcal{A}(x^*) \text{ avec } j \neq i. \end{aligned}$$

   La valeur de $f(x^*(\epsilon))$ peut être approchée en s'aidant de l'équation (20a) :

$$\begin{aligned} f(x^*(\epsilon)) - f(x^*) &\approx (x^*(\epsilon) - x^*)^T \nabla f(x^*) \\ &= \sum_{j \in \mathcal{A}(x^*)} \lambda_j^* (x^*(\epsilon) - x^*)^T \nabla c_j(x^*) \\ &\approx -\epsilon \|\nabla c_i(x^*)\| \lambda_i^*. \end{aligned}$$

   En prenant les limites, nous obtenons que la famille de solutions $x^*(\epsilon)$ satisafait à

$$\frac{df(x^*(\epsilon))}{d\epsilon} = -\lambda_i^* \|\nabla c_i(x^*)\|. \tag{23}$$

   Après analyse, nous pouvons conclure que si $\lambda_i^* \|\nabla c_i(x^*)\|$ est grand, alors la valeur optimale est sensible à la $i$ème contrainte. Au contraire, si cette quantité est petite, la dépendance avec cette même contrainte n'est pas grande. Si λ_i^* est zéro pour une contrainte active, une petite perturbation de ci dans une direction perturbera à peine la valeur optimale.

Nous pouvons maintenant faire la définission suivante, qui classe les contraintes selon si leur multiplicateur de Lagrange est zéro ou non :

Définition 4. Soit x^* une solution du problème (1), et supposons les conditions de KKT satisfaites.
On dit qu'une contrainte d'inégalité c_i est *fortement active* si $i \in \mathcal{A}(x^*)$ et $\lambda_i^* > 0$ pour un multiplicateur de Lagrange λ^* respectant les conditions de KKT.
On dit que c_i est *faiblement active* si $i \in \mathcal{A}(x^*)$ et $\lambda_i^* = 0$ pour tout λ^* respectant les conditions de KKT.

Notons que l'analyse ci-dessus est indépendante de la multiplication, de l'extension de chaque contrainte. Nous pourrions changer la formulation du problème en multipliant la contrainte c_i par 10, on aurait $10c_i$ et le problème serait le même, il possèderait la même solution. Seul le multiplicateur λ_i^* serait changé et remplacé par λ_i^* 10. De plus, comme $\|\nabla c_i(x^*)\|$ serait remplacé par $10\,\|\nabla c_i(x^*)\|$, le produit $\lambda_i^*\,\|\nabla c_i(x^*)\|$ ne change pas. Si, d'un autre côté, nous remplaçons cette fois ci la fonction objectif f par $10f$, les composantes λ_i^* du multiplicateur λ^* seront toutes divisées par 10. Nous voyons en (23) que la sensibilité de f par rapport aux perturbations a été multiplié par 10.

Références

[1] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization.* Springer Series in Operations Research, New York, Springer 1999.

[2] Le tutoriel sur : `http://tex.loria.fr/general/apprends-latex.pdf`

[3] `http://mirror.switch.ch/ftp/mirror/tex/info/lshort/french/flshort-3.20.pdf`

[4] Le site internet du proséminaire : `http://perso.unifr.ch/ales.janka/numeroptim`

[5] Cours d'optimisation : `http://maths.insa-lyon.fr/~pousin/site/Cours/Opti2/Opti2.php`