

KKT conditions

Background material

Fabian Bastin

`fabian.bastin@umontreal.ca`

Université de Montréal – CIRRELT – IVADO – Fin-ML

Lagrangian and Lagrangian dual function

Consider the problem

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \leq 0, \quad i = 1, \dots, m, \\ & h_j(x) = 0, \quad j = 1, \dots, r. \end{aligned}$$

We define the Lagrangian as

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^m \lambda_i g_i(x) + \sum_{j=1}^r \mu_j h_j(x),$$

et la fonction duale lagrangienne

$$\mathcal{L}(\lambda, \mu) = \min_{x \in \mathbb{R}^n} L(x, \lambda, \mu)$$

Lagrange multipliers: equality constraints

Consider the mathematical program

$$\begin{aligned} \min_{x \in \mathcal{X}} f(x) \\ \text{subject to } g_i(x) = 0, \quad i = 1, \dots, m, \end{aligned}$$

where $\mathcal{X} \subset \mathbb{R}^n$, $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $g_i : \mathbb{R}^n \rightarrow \mathbb{R}$, $i = 1, \dots, m$.

Le **Lagrangien** associé à ce problème est obtenu en associant un multiplicateur de Lagrange λ_i à chaque fonction de contrainte g_i :

$$L(x, \lambda) = f(x) + \sum_{i=1}^m \lambda_i g_i(x).$$

Sans faire d'hypothèse particulière sur \mathcal{X} ou sur les fonctions f et g_i , nous pouvons obtenir des conditions très générales pour qu'un point x^* soit une solution optimale du problème.

Optimality

Theorem

Supposons que le Lagrangien associé au problème

$$\begin{aligned} \min_{x \in \mathcal{X}} f(x) \\ \text{t.q. } g_i(x) = 0, \quad i = 1, \dots, m. \end{aligned}$$

possède un minimum local $x^ \in \mathcal{X}$ lorsque le vecteur de multiplicateurs λ vaut λ^* . Si $g_i(x^*) = 0, i = 1, \dots, m$, alors x^* est un minimum local de $f(x)$.*

Optimality

Proof.

La preuve se fait par contradiction en supposant que x^* n'est pas un minimum local de $f(x)$. Alors $\forall \epsilon > 0, \exists \bar{x} \in \mathcal{B}(x^*, \epsilon)$ tel que $g_i(\bar{x}) = 0, i = 1, \dots, m$ et $f(\bar{x}) < f(x^*)$.

Par conséquent, pour tout λ ,

$$\sum_{i=1}^m \lambda_i g_i(x^*) = \sum_{i=1}^m \lambda_i g_i(\bar{x}) = 0.$$

Dès lors,

$$f(\bar{x}) + \sum_{i=1}^m \lambda_i g_i(\bar{x}) < f(x^*) + \sum_{i=1}^m \lambda_i g_i(x^*).$$

En prenant $\lambda = \lambda^*$, la relation précédente contredit le fait que est un minimum local du Lagrangien lorsque $\lambda = \lambda^*$.

Lagrange multipliers: inequality constraints

Consider the mathematical program

$$\min_{x \in \mathcal{X}} f(x)$$

$$\text{subject to } g_i(x) \leq 0, \quad i = 1, \dots, m.$$

where $\mathcal{X} \subset \mathbb{R}^n$, $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $g_i : \mathbb{R}^n \rightarrow \mathbb{R}$, $i = 1, \dots, m$.

Theorem

Assume that the Lagrangian associated to the problem

$$\min_{x \in \mathcal{X}} f(x)$$

$$\text{s.t. } g_i(x) \leq 0, \quad i = 1, \dots, m,$$

has a local minimum $x^ \in \mathcal{X}$ when the multipliers vector λ is equal to λ^* . If $g_i(x^*) \leq 0$, $\lambda_i^* \geq 0$, and $\lambda_i^* g_i(x^*) = 0$, $i = 1, \dots, m$, then x^* is a local minimum of $f(x)$.*

Lagrange multipliers: inequality constraints

Proof.

Comme précédemment, la preuve se fait par contradiction en supposant que x^* n'est pas un minimum local de $f(x)$. Alors $\forall \epsilon > 0, \exists \bar{x} \in \mathcal{B}(x^*, \epsilon)$ tel que $g_i(\bar{x}) \leq 0, i = 1, \dots, m$ et $f(\bar{x}) < f(x^*)$. Par conséquent, pour $\lambda = \lambda^* \geq 0$,

$$\sum_{i=1}^m \lambda_i g_i(\bar{x}) \leq 0 \text{ et } \sum_{i=1}^m \lambda_i g_i(x^*) = 0.$$

Dès lors,

$$f(\bar{x}) + \sum_{i=1}^m \lambda_i g_i(\bar{x}) < f(x^*) + \sum_{i=1}^m \lambda_i g_i(x^*).$$

La relation précédente contredit le fait que x^* est un minimum local du Lagrangien lorsque $\lambda = \lambda^*$.

Problème dual

Le problème dual est

$$\begin{aligned} \max_{\lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^r} \quad & \mathcal{L}(\lambda, \mu) \\ \text{tel que } & \lambda \geq 0. \end{aligned}$$

Propriétés importantes:

- Le problème dual est toujours convexe, i.e. \mathcal{L} est toujours concave (même si la problème primal n'est pas convexe).
- Les valeurs optimales (globales) primale et duale, f^* et \mathcal{L}^* , satisfont toujours la dualité faible: $f^* \geq \mathcal{L}^*$.
- **Dualité forte**: sous certaines conditions (qualifications de contraintes), $f^* = \mathcal{L}^*$.

Duality gap

Étant donné une solution primale réalisable x et une solution duale réalisable (λ, μ) , la quantité $f(x) - \mathcal{L}(\lambda, \mu)$ est appelé le saut de dualité entre x et (λ, μ) . Notons que

$$f(x) - f^* \leq f(x) - \mathcal{L}(\lambda, \mu)$$

de sorte que si le saut de dualité est nul, alors x est optimal primal (et similairement, λ et μ sont optimaux duaux).

D'un point de vue algorithmique, si la dualité forte tient, ceci fournit un critère d'arrêt: si $f(x) - \mathcal{L}(\lambda, \mu) \leq \epsilon$, nous avons alors la garantie que $f(x) - f^* \leq \epsilon$.

Saut de dualité: cas local

Désignons l'ensemble réalisable par

$$\mathcal{X} = \{x \mid g_i(x), i = 1, \dots, m, h_j(x), j = 1, \dots, r\}.$$

Considérons x^* un minimum local de $f(\cdot)$, i.e.

$$\exists \epsilon > 0 \text{ t.q. } \forall x \in \mathcal{B}(x^*, \epsilon) \cap \mathcal{X}, f(x^*) \leq f(x).$$

Nous pouvons également définir la fonction duale lagrangienne restreinte à la boule $\mathcal{B}(x^*, \epsilon)$:

$$\mathcal{L}_{\mathcal{B}(x^*, \epsilon)}(\lambda, \mu) = \min_{x \in \mathcal{B}(x^*, \epsilon)} L(x, \lambda, \mu).$$

Dans ce cas, la dualité faible tient toujours localement:

$$\mathcal{L}_{\mathcal{B}(x^*, \epsilon)}^* \leq f(x^*).$$

Sous certaines conditions, la dualité forte tient également:

$$\mathcal{L}_{\mathcal{B}(x^*, \epsilon)}^* = f(x^*).$$

Saut de dualité: cas local

Remarquons cependant que

$$\min_{x \in \mathbb{R}^n} L(x, \lambda, \mu) \leq \min_{x \in \mathcal{B}(x^*, \epsilon)} L(x, \lambda, \mu)$$

et donc

$$\mathcal{L}^* \leq \mathcal{L}_{\mathcal{B}(x^*, \epsilon)}^*.$$

Dès lors, si x^* est un minimum local et si la dualité forte tient localement,

$$\mathcal{L}^* \leq f(x^*),$$

l'inégalité pouvant être stricte.

Karush-Kuhn-Tucker (KKT) conditions

Soient $f, g_i, h_j \in C^1$, $i = 1, \dots, m$, $j = 1, \dots, r$, et le problème

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{t.q.} \quad & g_i(x) \leq 0, \quad i = 1, \dots, m, \\ & h_j(x) = 0, \quad j = 1, \dots, r. \end{aligned}$$

Conditions de Karush-Kuhn-Tucker (KKT):

$\nabla_x L(x, \lambda, \mu) = 0$	(stationarité)
$\lambda_i g_i(x) = 0$	(écarts de complémentarités)
$g_i(x) \leq 0, h_j(x) = 0 \quad \forall i, j$	(faisabilité primale)
$\lambda_i \geq 0 \quad \forall i$	(faisabilité duale)


Nécessité

Soient x^* minimum pour $x \in \mathcal{B}(x^*, \epsilon)$, $\epsilon > 0$, et (λ^*, μ^*) solution du dual si x est restreint à $\mathcal{B}(x^*, \epsilon)$, avec un saut de dualité nul (la dualité forte tient). Alors

$$\begin{aligned} f(x^*) &= \mathcal{L}(\lambda^*, \mu^*) \\ &= \min_{x \in \mathcal{B}(x^*, \epsilon)} \left(f(x) + \sum_{i=1}^m \lambda_i^* g_i(x) + \sum_{i=1}^r \mu_i^* h_i(x) \right) \\ &\leq f(x^*) + \sum_{i=1}^m \lambda_i^* g_i(x^*) + \sum_{i=1}^r \mu_i^* h_i(x^*) \\ &\leq f(x^*) \end{aligned}$$

Dès lors, x^* est un minimum de $L(x, \lambda^*, \mu^*)$ sur $\mathcal{B}(x^*, \epsilon)$.

Par conséquent, $\nabla_x L(x^*, \lambda^*, \mu^*) = 0$.

Nous retrouvons les conditions de stationarité. 

Nécessité

Nous devons aussi avoir $\sum_{i=1}^m \lambda_i^* g_i(x^*) = 0$ puisque $\sum_{i=1}^m \lambda_i^* g_i(x^*) \leq 0$. Ceci implique que pour tout i , $\lambda_i^* g_i(x^*) = 0$.

Nous retrouvons les conditions de complémentarité.

Si x^* est un minimum global, nous pouvons remplacer $\mathcal{B}(x^*, \epsilon)$ par \mathbb{R}^n .

Nécessité

Theorem (Nécessité des conditions KKT)

Si x^ , (λ^*, μ^*) sont des solutions primale et duale avec un saut de dualité nul (i.e. la dualité forte tient), alors x^* , (λ^*, μ^*) satisfont les conditions KKT.*

Dès lors, l'hypothèse de dualité forte apparaît importante. Elle sera garantie sous certaines conditions.

- Programme linéaire. La dualité forte tient toujours.
- Programme convexe. Condition de Slater: $\exists x$ tel que $g_i(x) < 0$, $i = 1, \dots, m$ et $h_i(x) = 0$, $i = 1, \dots, r$.
- Programme non-convexe. Hypothèse de qualification de contraintes. La plus courante, mais aussi la plus forte, est la condition d'indépendance linéaire des gradients à la solution.

Nécessité (cas non convexe)

Theorem (Nécessité des conditions KKT)

Si x^ est une solution locale de*

$$\begin{aligned} \min_{x \in \mathcal{X}} \quad & f(x) \\ \text{t.q.} \quad & g_i(x) \leq 0, \quad i = 1, \dots, m \\ & h_i(x) = 0, \quad i = 1, \dots, r, \end{aligned}$$

où les fonctions f , g_i et h_i , $i = 1, \dots, m$, sont continûment différentiables, et qu'une condition de qualification de contrainte tient en x^ . Alors, il existe un vecteur de multiplicateurs de Lagrange (λ^*, μ^*) tel que les conditions KKT sont satisfaites en (x^*, λ^*, μ^*) .*

Proof.

Preuve: technique! Voir par exemple Nocedal & Wright, "Numerical Optimization", Section 12.4.

Suffisance des conditions KKT

S'il existent x^* , (λ^*, μ^*) satisfaisant les conditions KKT, alors

$$\begin{aligned} L(\lambda^*, \mu^*) &= f(x^*) + \sum_{i=1}^m \lambda_i^* g_i(x^*) + \sum_{i=1}^r \mu_i^* h_i(x^*) \\ &= f(x^*) \end{aligned}$$

Dès lors, le saut de dualité est nul (**dualité forte**).

Dans le cas convexe, cela implique que x^* et (λ^*, μ^*) sont des solutions globales primale et duale, respectivement.

Dans la cas non-convexe, x^* est un minimum local, pas nécessairement global, voire un point-selle.

Active set

Definition (Active set)

L'ensemble actif $\mathcal{A}(x)$ du problème d'optimisation

$$\begin{aligned} \min_{x \in \mathcal{X}} \quad & f(x) \\ \text{t.q.} \quad & g_i(x) \leq 0, \quad i \in \mathcal{I} \\ & h_i(x) = 0, \quad i \in \mathcal{E}, \end{aligned}$$

en un point réalisable x est l'ensemble des indices des contraintes d'égalité et l'ensemble des indices i des contraintes d'inégalité telles que $g_i(x) = 0$, c'est-à-dire

$$\mathcal{A}(x) = \mathcal{E} \cup \{i \mid g_i(x) = 0\}$$

LICQ

The most popular constraint qualification is the LICQ.

Definition (LICQ)

Étant donné le point x et l'ensemble actif $\mathcal{A}(x)$, nous disons que la qualification de contraintes d'indépendance linéaire (linear independence constraint qualification – LICQ) tient si l'ensemble des contraintes actives $\nabla_x c_i(x), i \in \mathcal{A}(x)$ est linéairement indépendant.