KKT conditions
Background material

Fabian Bastin
fabian.bastin@umontreal.ca
Université de Montréal – CIRRELT – IVADO – Fin-ML

Basic notions

Consider the problem

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

The feasible set is

$$\mathcal{X} = \{x \,|\, g_i(x) \leq 0, i = 1, \ldots, m, h_j(x) = 0, j = 1, \ldots, r\}.$$

x^* is a global minimizer of $f(\cdot)$ if $\forall x \in \mathcal{X}$, $f(x^*) \leq f(x)$.
x^* is a local minimizer of $f(\cdot)$ if $\exists \epsilon > 0$ such that
$\forall x \in \mathcal{B}(x^*, \epsilon) \cap \mathcal{X}$, $f(x^*) \leq f(x)$.

Lagrangian and Lagrangian dual function

We define the Lagrangian as

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

and the dual Lagrangian function

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

Lagrange multipliers: equality constraints

Consider the mathematical program

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

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

The Lagrangian of this problem is obtained by associating a Lagrange multiplier λ_i to each constraint function g_i:

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

We can obtain very general conditions under which x^* is an optimal solution to the optimization problem, while only basic assumptions are made over $\mathcal{X}$ and the functions f and g_i, $i = 1, \dots, m$.

Optimality

Theorem

Assume that the Lagrangian associated to the problem

$$\min_{x \in \mathcal{X}} f(x)$$
$$s.t.\ g_i(x) = 0,\ i = 1, \ldots, m,$$

has a local minimizer $x^* \in \mathcal{X}$ *when the multiplier vector* λ *is equal to* λ^*. *If* $g_i(x^*) = 0$, $i = 1, \ldots, m$, *then* x^* *if a local minimizer of* $f(x)$.

Optimality

Proof.

Assume by contradiction that x^* is not a local minimizer of $f(x)$. Then $\forall \epsilon > 0$, $\exists \overline{x} \in \mathcal{B}(x^*, \epsilon)$ such that $g_i(\overline{x}) = 0$, $i = 1, \ldots, m$, and $f(\overline{x}) < f(x^*)$.
Thus, $\forall \lambda$,

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

and

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

Taking $\lambda = \lambda^*$, the previous inequality contradicts that x^* is a local minimizer of the Lagrangian when $\lambda = \lambda^*$. $\square$

Lagrange multipliers: inequality constraints

Consider the mathematical program

$$\min_{x \in \mathcal{X}} f(x)$$
$$\text{subject to } g_i(x) \leq 0, \; i = 1, \ldots, m.$$

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

Theorem

Assume that the Lagrangian associated to the problem

$$\min_{x \in \mathcal{X}} f(x)$$
$$s.t. \; g_i(x) \leq 0, \; i = 1, \ldots, m,$$

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 $\lambda_i^* g_i(x^*) = 0$, $i = 1, \ldots, m$, then x^* is a local minimum of $f(x)$.

Lagrange multipliers: inequality constraints

Proof.

As previously, assume by contradiction that x^* is not a local minimizer of $f(x)$. Then $\forall \epsilon > 0$, $\exists\, \overline{x} \in \mathcal{B}(x^*, \epsilon)$ such that $g_i(\overline{x}) \leq 0$, $i = 1, \ldots, m$ and $f(\overline{x}) < f(x^*)$. Therefore, for $\lambda = \lambda^* \geq 0$,

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

Consequently,

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

contradicting that x^* is a local minimizer of the Lagrangian when $\lambda = \lambda^*$. $\square$

Dual problem

The dual problem is

$$\max_{\lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^r} \mathcal{L}(\lambda, \mu)$$
$$\text{such that } \lambda \geq 0.$$

Important properties:

- The dual problem is always convex, i.e. $\mathcal{L}$ is always concave (even if the primal problem is not convex).
- The primal and dual (global) optimal values, f^* and $\mathcal{L}^*$, always satisfy the weak duality: $f^* \geq \mathcal{L}^*$.
- Strong duality: under some conditions (constraint qualifications), $f^* = \mathcal{L}^*$.

Duality gap

Given a primal feasible solution x and a dual feasible solution (λ, μ), the quantity $f(x) - \mathcal{L}(\lambda, \mu)$ is called the duality gap between x and (λ, μ). Note that

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

Therefore, if the duality gap is equal to 0, then x is primal-optimal (and similarly, λ and μ are dual-optimal).

From an algorithmic point of view, if strong duality holds, this provides a stopping criterion: if $f(x) - \mathcal{L}(\lambda, \mu) \leq \epsilon$, then $f(x) - f^* \leq \epsilon$.

Duality gap: local case

We can also define the dual Lagrangian function restricted to the ball $\mathcal{B}(x^*, \epsilon)$:

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

The weak duality still holds locally:

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

Under some conditions, the strong duality also holds:

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

Duality gap: local case

Note however that

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

so

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

Therefore, if x^* is a local minimizer and the strong duality locally holds,

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

The inequality can be strict.

Karush-Kuhn-Tucker (KKT) conditions

Consider $f, g_i, h_j \in C^1$, $i = 1, \dots, m$, $j = 1, \dots, r$, and the problem

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

Karush-Kuhn-Tucker (KKT) conditions:

$$\begin{aligned} \nabla_x L(x, \lambda, \mu) &= 0 && \text{(stationarity)} \\ \lambda_i g_i(x) &= 0 && \text{(complementarity)} \\ g_i(x) \le 0, \ h_j(x) &= 0 \ \forall i, j && \text{(primal feasibility)} \\ \lambda_i &\ge 0 \ \forall i && \text{(dual feasibility)} \end{aligned}$$

Necessary conditions

Let x^* be a minimizer of $f(\cdot)$ in $\mathcal{B}(x^*, \epsilon)$, $\epsilon > 0$, and (λ^*, μ^*) be a dual solution if x is restricted to $\mathcal{B}(x^*, \epsilon)$, with a zero duality gap (the strong duality holds). Then

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

Thus, x^* is a minimizer of $L(x, \lambda^*, \mu^*)$ in $\mathcal{B}(x^*, \epsilon)$, and $\nabla_x L(x^*, \lambda^*, \mu^*) = 0$.

We have obtained the stationarity conditions.

Necessary conditions

The previous inequalities also imply $\sum_{i=1}^{m} \lambda_i^* g_i(x^*) = 0$ as $\sum_{i=1}^{m} \lambda_i^* g_i(x^*) \leq 0$, and consequently, $\lambda_i^* g_i(x^*) = 0, \ \forall i$.

This establishes the complementarity conditions.

If x^* is a global minimizer, we can replace $\mathcal{B}(x^*, \epsilon)$ by $\mathbb{R}^n$.

We can summarize our findings in the theorem below.

Theorem (KKT necessary conditions)

If x^*, (λ^*, μ^*) are primal and dual solutions with a null duality gap, then x^*, (λ^*, μ^*) satisfy the KKT conditions.

Strong duality

The strong duality assumption often plays a key role. How to ensure that it holds?

- Linear programming. It always holds.
- Convex programming. Slater condition: $\exists x$ such that $g_i(x) < 0$, $i = 1, \dots, m$ et $h_i(x) = 0$, $i = 1, \dots, r$.
- Nonconvex programming. Constraint qualification hypothesis. The most common, while the most restrictive, is the linear independence constraint qualification (LICQ).

Nonconvex programming

Theorem (Necessary conditions)

If x^* is a local solution of

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

where f, g_i et h_i, $i = 1, \ldots, m$, $\in C^1$, and a constraint qualification condition holds at x^*, then $\exists(\lambda^*, \mu^*)$ such that the KKT conditions hold at (x^*, λ^*, μ^*).

Proof.

See Nocedal & Wright, "Numerical Optimization", Section 12.4. □

Sufficiency of KKT conditions

If $\exists\, x^*$, (λ^*, μ^*) satisfying the KKT conditions, then

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

Thus, the duality gap is null (strong duality).

In the convex case, this implies that x^* and (λ^*, μ^*) are global primal and dual optimal solutions, respectively.

In the nonconvex case, x^* is a local minimizer, not necessarily global, or a saddle point.

Active set

Definition (Active set)

The active set $\mathcal{A}(x)$ of the optimization problem

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

in a feasible point x is the index set of the equality constraints and the active inequality constraints at that point:

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

LICQ

The most popular constraint qualification is the LICQ.

Definition (LICQ)

Given the point x *and the active set* $\mathcal{A}(x)$*, the linear independence constraint qualification (LICQ) holds if the gradients of the active constraints,* $\{\nabla_x c_i(x), i \in \mathcal{A}(x)\}$*, are linearly independent.*