Optimization and Computational Linear Algebra for Data Science
Lecture 10: Optimality conditions

Léo Miolane · leo.miolane@gmail.com

July 9, 2019

Warning: *This material is not meant to be lecture notes. It only gathers the main concepts and results from the lecture, without any additional explanation, motivation, examples, figures...*

1 Local and global minimizers

We aim at minimizing a function $f : \mathbb{R}^n \to \mathbb{R}$. We say that $x^* \in \mathbb{R}^n$ is

- a *global* minimizer of f if for all $x \in \mathbb{R}^n$, $f(x^*) \leq f(x)$.
- a *local* minimizer of f if there exists $\delta > 0$ such that for all $x \in B(x^*, \delta)$, $f(x^*) \leq f(x)$.

Of course, a global minimizer if also a local minimizer but the converse is not true.

Proposition 1.1

Let $x \in \mathbb{R}^n$ be a point at which f is differentiable. Then

$$x \text{ is a local minimizer of } f \quad \Longrightarrow \quad \nabla f(x) = 0.$$

If f is convex then the converse is true:

Proposition 1.2

Assume that f is convex. Let $x \in \mathbb{R}^n$ be a point at which f is differentiable. Then

$$\nabla f(x) = 0 \quad \Longrightarrow \quad x \text{ is a global minimizer of } f.$$

2 Constrained optimization

We would now like to investigate constrained optimization problems:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g_i(x) \leq 0, \quad i = 1, \dots, m \\ & h_i(x) = 0, \quad i = 1, \dots, p, \end{array} \tag{1}$$

with variable $x \in \mathbb{R}^n$. Here we have m inequality constraints $g_1(x) \leq 0, \dots, g_m(x) \leq 0$ and p equality constraints $h_1(x) = 0, \dots, h_p(x) = 0$ to satisfy.

Definition 2.1 (*Feasible point*)

A point $x \in \mathbb{R}^n$ is feasible *if it satisfies all the constraints: $g_1(x) \leq 0, \dots, g_m(x) \leq 0$ and $h_1(x) = 0, \dots, h_p(x) = 0$. We will denote by F the set of feasible points.*

We would now get for the problem (1) the analog of Proposition 1.1. Since an equality constraint $h_i(x) = 0$ can be equivalently written in two inequality constraints $h_i(x) \leq 0$ and

$-h_i(x) \leq 0$, we can assume to have only inequality constraints. For simplicity, we first assume to have only one inequality constraint $g(x) \leq 0$ so that (1) reduces to

$$\text{minimize } f(x) \text{ subject to } g(x) \leq 0. \tag{2}$$

Let x be a solution of (2), i.e. $g(x) \leq 0$ and $f(x) \leq f(x')$ for all x' such that $g(x') \leq 0$ We distinguish two cases:

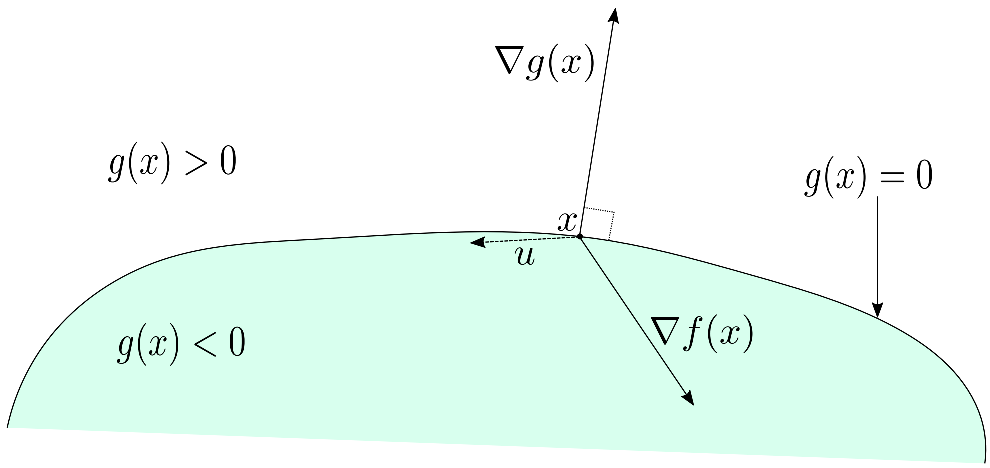
Case 1: x is "strictly feasible" $g(x) < 0$. In that case x is in the interior of F: one can find $\delta > 0$ such that $B(x, \delta) \subset F$. Since x is a solution of (2) we have for all $x' \in B(0, \delta)$, $f(x) \leq f(x')$. One can therefore apply Proposition 1.1 to get that $\nabla f(x) = 0$.

We conclude that in the case where the constraint is not active, the constraint does not play any role and one gets the same optimality condition as in the unconstrained setting.

Case 2: the constraint is active in x, $g(x) = 0$. In that case, there exists $\lambda \geq 0$ such that

$$\nabla f(x) = -\lambda \nabla g(x). \tag{3}$$

To see that, assume that (3) does not hold. Then we are in the following situation:



As we can see on the figure, we can find a vector u such that

$$\langle u, \nabla g(x) \rangle < 0 \quad \text{and} \quad \langle u, \nabla f(x) \rangle < 0.$$

Starting from x and following the direction u one remains in the feasible set because for small $\delta > 0$

$$g(x + \delta u) \simeq g(x) + \delta \langle u, \nabla g(x) \rangle \leq 0.$$

Moreover, f decreases locally on the direction u:

$$f(x + \delta u) \simeq f(x) + \delta \langle u, \nabla f(x) \rangle < f(x).$$

This means that one can find $\delta > 0$ such that $x + \delta u$ is feasible and such that $f(x + \delta u) < f(x)$. This contradicts the assumption that x is solution of (2). We conclude that (3) holds, i.e. that there exists $\lambda \geq 0$ such that

$$\nabla f(x) + \lambda \nabla g(x) = 0.$$

We will only cover the case where the equality constraints are linear, i.e. $h_i(x) = \langle a_i, x \rangle + b_i$ for from $a_i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$.

This generalize to the case (1) where we have multiple constraints:

Definition 2.2

We say that the constraints are qualified *at $x \in F$ if there exists a vector $v \in \mathbb{R}^n$ such that*

- $\langle v, \nabla g_i \rangle < 0$ *for all* $i = 1, \dots, m$.
- $\langle v, \nabla h_i \rangle = 0$ *for all* $i = 1, \dots, p$.

Theorem 2.1 (*Karush-Kuhn-Tucker conditions*)

Assume that the functions f, $g_1, \dots, g_m$ are differentiable and that $h_1, \dots, h_p$ are linear. If x is solution of (1) and if the constraints are qualified at x then there exists $\lambda_1, \dots, \lambda_m \geq 0$ and $\nu_1, \dots, \nu_p \in \mathbb{R}$ such that:

$$\nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0. \tag{4}$$

Moreover $\lambda_i = 0$ if $g_i(x) < 0$.

The scalars λ_i, ν_i are often called *Lagrange multipliers*.

3 The Lagrangian and the dual problem

We define the Lagrange dual function L associated with the problem (1) by

$$L(x, \lambda, \nu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{i=1}^{p} \nu_i h_i(x), \tag{5}$$

where $x \in \mathbb{R}^n$, $\lambda \in \mathbb{R}^m_{\geq 0}$ and $\nu \in \mathbb{R}^p$. We define the Lagrange dual function by

$$\ell(\lambda, \nu) = \inf_{x \in \mathbb{R}^n} L(x, \lambda, \nu).$$

Notice that for all feasible point x,

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

because $h_i(x) = 0$ and $\lambda_i g_i(x) \leq 0$. By taking the infimum in x on both sides of the inequality we get a lower bound on the value of the optimization problem (1):

Proposition 3.1

Let p^* be the optimal value of the problem (1). For all $\lambda_1, \dots, \lambda_m \geq 0$ and all $\nu_1, \dots, \nu_p \in \mathbb{R}$ we have:

$$\ell(\lambda, \nu) \leq p^*. \tag{6}$$

3.1 Dual problem

We would like to make the lower bound (6) as tight as possible: one would like therefore to solve the so-called *dual problem*:

$$\begin{array}{ll} \text{maximize} & \ell(\lambda, \nu) \\ \text{subject to} & \lambda_i \geq 0, \quad i = 1, \dots, m \\ & \nu_i \in \mathbb{R}, \quad i = 1, \dots, p. \end{array} \tag{7}$$

Notice that the Lagrange dual function is always concave, as an infimum of linear functions. Hence, the dual problem might be easier to solve than the original problem.

From (6) we deduce that the optimal value of the primal problem is greater or equal than the one of the dual problem:

$$\sup_{\lambda \geq 0, \nu} \ell(\lambda, \nu) \leq p^*. \tag{8}$$

This is known as *weak duality*.

Notice that $p^* = \inf_{x \in \mathbb{R}^n} F(x)$ where

$$F(x) \overset{\text{def}}{=} \sup_{\lambda \geq 0, \nu} L(x, \lambda, \nu) = \begin{cases} f(x) & \text{if } x \text{ is feasible,} \\ +\infty & \text{otherwise.} \end{cases}$$

Hence, the weak duality inequality can be rewritten as:

$$\sup_{\lambda \geq 0, \nu} \inf_{x \in \mathbb{R}^n} L(x, \lambda, \nu) \leq \inf_{x \in \mathbb{R}^n} \sup_{\lambda \geq 0, \nu} L(x, \lambda, \nu). \tag{9}$$

When there is equality in (8) (or equivalently in (9)) we say that there is *strong duality*. We will see in the next section that strong duality holds for convex problems under mild assumptions.

3.2 Saddle-points

Definition 3.1 (*Saddle-point*)

We say that $(x; \lambda, \nu) \in \mathbb{R}^n \times \mathbb{R}^m_{\geq 0} \times \mathbb{R}^p$ *is a* saddle-point *of* L *if*

$$\forall (\lambda', \nu') \in \mathbb{R}^m_{\geq 0} \times \mathbb{R}^p \quad L(x, \lambda', \nu') \leq L(x, \lambda, \nu) \leq L(x', \lambda, \nu) \quad \forall x' \in \mathbb{R}^n. \tag{10}$$

Notice that if $(x; \lambda, \nu)$ is a saddle point of L, then x is solution of (1). Indeed, by taking the supremum in (10) we get:

$$F(x) \leq L(x, \lambda, \nu) \leq L(x', \lambda, \nu) \leq F(x')$$

for all $x' \in \mathbb{R}^n$. By an analog argument one also get that (λ, ν) is a solution of the dual problem (7).

Theorem 3.1 (*Saddle-point theorem*)

$(x; \lambda, \nu) \in \mathbb{R}^n \times \mathbb{R}^m_{\geq 0} \times \mathbb{R}^p$ *is a saddle-point of* L *if and only if*

$$F(x) = \inf_{x'} \sup_{\lambda' \geq 0, \nu'} L(x', \lambda', \nu') = \sup_{\lambda' \geq 0, \nu'} \inf_{x'} L(x', \lambda', \nu') = \ell(\lambda, \nu), \tag{11}$$

i.e. if and only if x *is a solution of the primal problem* (1), (λ, ν) *is a solution of the dual problem* (7) *and strong duality holds.*

3.3 Solving the primal problem via the dual

From Theorem 3.1 we get that if strong duality holds and if (λ^*, ν^*) is a solution of the dual problem, then any solution of the primal problem is a minimizer of $x \mapsto L(x, \lambda^*, \nu^*)$.

Second, assume that strong duality holds and assume that we computed a solution (λ^*, ν^*) of the dual problem (7). Assume also that $x \mapsto L(x, \lambda^*, \nu^*)$ admits a unique minimizer x^*. Then if

x^* is feasible then it is a solution of the problem (1). If x^* is not feasible then the minimum of (1) is not attained. Indeed, one has by construction of x^* and definition of ℓ

$$L(x^*, \lambda^*, \nu^*) = \inf_x L(x, \lambda^*, \nu^*) = \ell(\lambda^*, \nu^*) = p^*$$

where the last equality comes from the definition of (λ^*, ν^*) and the strong duality. So if x^* is feasible, then for all $x \in \mathbb{R}^n$

$$L(x^*, \lambda^*, \nu^*) \leq L(x, \lambda^*, \nu^*) \leq F(x)$$

4 Kuhn Tucker Theorem

In this section, we assume that the functions f, $g_1, \ldots, g_m$ are convex and that $h_1, \ldots, h_p$ are linear. We say then that the optimization problem (1) is convex.

Definition 4.1 (*Slater's condition*)

We say that the problem (1) verifies Slater's condition if there exists a feasible point x such that $g_i(x) < 0$ for all $i \in \{1, \ldots, m\}$.

One can verify that is the problem (1) is convex and verifies Slater's condition, then the constraints are qualified at every feasible point.

Theorem 4.1 (*Kuhn Tucker*)

*Assume that the functions f, $g_1, \ldots, g_m$ are **convex**, differentiable and that $h_1, \ldots, h_p$ are linear. Assume that (1) verifies Slater's condition. Then $x \in \mathbb{R}^n$ is solution of (1) if and only if x is feasible and there exists $\lambda \in \mathbb{R}^m_{\geq 0}$ and $\nu \in \mathbb{R}^p$ such that*

$$\begin{cases} \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0 \\ \lambda_i g_i(x) = 0 \quad \text{for all} \quad i = 1, \ldots, m. \end{cases} \tag{12}$$

In other words, x is a solution of (1) if and only if there exists $\lambda \in \mathbb{R}^m_{\geq 0}$ and $\nu \in \mathbb{R}^p$ such that $(x; \lambda, \nu)$ is a saddle point of the Lagrangian L.

1. First we find a solution (λ, ν) of the dual problem (7).
2. Then we solve the **unconstrained** minimization problem $\min_x L(x, \lambda, \nu)$ to find a solution x.
3. Finally, Theorem 3.1 and Theorem 4.1 give us that x is a solution of the original problem (1). Indeed, since (λ, ν) is solution of the dual problem, the last equality of (11) holds.

Further reading

See [2] for a very nice introduction to spectral clustering and [1] for lecture notes on spectral graph theory.

References

[1] Daniel Spielman. Spectral graph theory. *Lecture Notes, Yale University, http://www.cs.yale.edu/homes/spielman/561/2012/*, 2012.

[2] Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.