Optimality conditions

V. Leclère (ENPC)

March 28th, 2025

Why should I bother to learn this stuff?

- Optimality conditions enable to solve exactly some easy optimization problems (e.g. in microeconomics, some mechanical problems...)
- Optimality conditions are used to derive algorithms for complex problems
- KKT is a certificate of optimality (or non optimality) used by optimization solvers (more than a solving method per se).
- $\Longrightarrow$ fundamental both for studying optimization as well as other science

Contents

Optimization problem: vocabulary

♡

Generically speaking, an optimization problem is

$$\underset{x \in X}{\mathrm{Min}} \qquad f(x) \qquad (P)$$

where

- $f : \mathbb{R}^n \to \mathbb{R}$ is the objective function (a.k.a. cost function),
- X is the feasible set,
- $x \in X$ is an admissible decision variable or a solution,
- $x^\sharp \in X$ such that $val(P) = f(x^\sharp) = \inf_{x \in X} f(x)$ is an optimal solution,
- if $X = \mathbb{R}^n$ the problem is unconstrained,
- if X and f are convex, then the problem is convex,
- if X is a polyhedron and f linear then the problem is linear,
- if X is a convex cone (optionally intersected with affine space) and f linear then the problem is conic.

Optimization problem: explicit formulation ♡

The previous optimization problem is often defined explicitly in the following standard form

$$
\begin{aligned}
\underset{x\in\mathbb{R}^n}{\mathrm{Min}} \quad & f(x) && (P)\\
\text{s.t.} \quad & g_i(x) = 0 && \forall i \in [n_E]\\
& h_j(x) \leq 0 && \forall j \in [n_I]
\end{aligned}
$$

with

$$X := \big\{x \in \mathbb{R}^n \mid \forall i \in [n_E], \quad g_i(x) = 0, \quad \forall j \in [n_I], \quad h_j(x) \leq 0\big\}.$$

- (P) is a differentiable optimization problem if f and $\{g_i\}_{i\in[n_E]}$ and $\{h_j\}_{j\in[n_I]}$ are differentiable.
- (P) is a convex differentiable optimization problem if f, and h_j (for $j \in [n_I]$) are convex differentiable and g_i (for $i \in [n_E]$) are affine.
  ♣ Exercise: Show that in this case X is convex.

Optimization problem: explicit formulation ♡

The previous optimization problem is often defined explicitly in the following standard form

$$
\begin{aligned}
\underset{x\in\mathbb{R}^n}{\mathrm{Min}} \quad & f(x) && (P)\\
\text{s.t.} \quad & g_i(x) = 0 && \forall i \in [n_E]\\
& h_j(x) \leq 0 && \forall j \in [n_I]
\end{aligned}
$$

with

$$X := \{x \in \mathbb{R}^n \mid \forall i \in [n_E], \quad g_i(x) = 0, \quad \forall j \in [n_I], \quad h_j(x) \leq 0\}.$$

- (P) is a differentiable optimization problem if f and $\{g_i\}_{i\in[n_E]}$ and $\{h_j\}_{j\in[n_I]}$ are differentiable.
- (P) is a convex differentiable optimization problem if f, and h_j (for $j \in [n_I]$) are convex differentiable and g_i (for $i \in [n_E]$) are affine.
  ♣ Exercise: Show that in this case X is convex.

A few remarks and tricks

- We can always write an abstract optimization problem in standard form (exercise!)
- For a given optimization problem there is an infinite number of possible standard forms (exercise!)
- We can always find an equivalent problem in dimension $\mathbb{R}^{n+1}$ with linear cost (exercise!)
- A minimization problem with $X = \emptyset$ has value $+\infty$ (by convention)
- A minimization problem has value $-\infty$ iff there exists a sequence $x_n \in X$ such that $f(x_n) \to -\infty$
- Maximizing f is just minimizing $-f$

Contents

Differentiable case

$\heartsuit$

Theorem

Assume that $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is differentiable at $x^\sharp$.

1. *If $x^\sharp$ is an unconstrained local minimizer of f then $\nabla f(x^\sharp) = 0$.*
2. *If in addition f is convex, then $\nabla f(x^\sharp) = 0$ iff $x^\sharp$ is a global minimizer.*

Proof:

1. Assume $\nabla f(x^\sharp) \neq 0$. DL of order 1 at $x^\sharp$ show that $f(x^\sharp - t\nabla f(x^\sharp)) < f(x^\sharp)$ for $t > 0$ small enough.
2. $f(y) \geq f(x^\sharp) + \langle \nabla f(x^\sharp), y - x^\sharp \rangle$.

Differentiable case

Theorem

Assume that $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is differentiable at $x^\sharp$.

1. *If $x^\sharp$ is an unconstrained local minimizer of f then $\nabla f(x^\sharp) = 0$.*
2. *If in addition f is convex, then $\nabla f(x^\sharp) = 0$ iff $x^\sharp$ is a global minimizer.*

Proof:

1. Assume $\nabla f(x^\sharp) \neq 0$. DL of order 1 at $x^\sharp$ show that $f(x^\sharp - t\nabla f(x^\sharp)) < f(x^\sharp)$ for $t > 0$ small enough.
2. $f(y) \geq f(x^\sharp) + \langle \nabla f(x^\sharp), y - x^\sharp \rangle$.

Convex case ♡

Theorem

Consider $f : \mathbb{R}^n \to \bar{\mathbb{R}}$. Then $x^\sharp$ is a global minimum iff

$$0 \in \partial f(x^\sharp)$$

Convex case

♡

Theorem

Consider $f : \mathbb{R}^n \to \bar{\mathbb{R}}$. Then $x^\sharp$ is a global minimum iff

$$0 \in \partial f(x^\sharp)$$

Theorem

Consider a proper convex function $f : \mathbb{R}^n \to \bar{\mathbb{R}}$, and X a closed convex set, such that $\mathrm{ri}(\mathrm{dom}(f)) \cap \mathrm{ri}(X) \neq \emptyset$.
Then $x^\sharp$ is a minimizer of f on X iff there exists $g \in \partial f(x^\sharp)$ such that $-g \in N_X(x^\sharp)$.

Convex case

♡

Theorem

Consider $f : \mathbb{R}^n \to \bar{\mathbb{R}}$. Then $x^\sharp$ is a global minimum iff

$$0 \in \partial f(x^\sharp)$$

Theorem

Consider a proper convex function $f : \mathbb{R}^n \to \bar{\mathbb{R}}$, and X a closed convex set, such that $\mathrm{ri}(\mathrm{dom}(f)) \cap \mathrm{ri}(X) \neq \emptyset$.
Then $x^\sharp$ is a minimizer of f on X iff there exists $g \in \partial f(x^\sharp)$ such that $-g \in N_X(x^\sharp)$.

proof : The technical assumption ensures that $\partial(f + \mathbb{I}_X) = \partial f + \partial(\mathbb{I}_X)$.
As $\partial(\mathbb{I}_X) = N_X$, we have, $0 \in \partial(f + \mathbb{I}_X)(x^\sharp)$ iff there exists $g \in \partial f(x^\sharp)$ such that $-g \in N_X(x^\sharp)$.

Contents

From local minimality to a tangent-direction condition

Consider

$$(P) \qquad \min_{x \in X} f(x),$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable and $X \subset \mathbb{R}^n$ is closed. Let $x \in X$ be a local minimizer.

Take any sequence of unit vectors $d_k \to d$ such that

$$x + \frac{1}{k} d_k \in X \qquad \text{for } k \text{ large enough.}$$

Then $x + \frac{1}{k} d_k \to x$, hence by local minimality,

$$f(x + \tfrac{1}{k} d_k) \;\geq\; f(x) \qquad \text{for } k \text{ large enough.}$$

Since f is differentiable at x,

$$f(x + \tfrac{1}{k} d_k) = f(x) + \tfrac{1}{k}\langle \nabla f(x), d_k \rangle + o(\tfrac{1}{k}).$$

Letting $k \to \infty$ (so $k\, o(1/k) \to 0$ and $d_k \to d$) yields

$$\langle \nabla f(x), d \rangle \geq 0.$$

From local minimality to a tangent-direction condition ♡

Consider

$$(P) \qquad \min_{x\in X} f(x),$$

where $f:\mathbb{R}^n\to\mathbb{R}$ is differentiable and $X\subset\mathbb{R}^n$ is closed. Let $x\in X$ be a local minimizer.
Take any sequence of unit vectors $d_k\to d$ such that

$$x+\frac{1}{k}d_k\in X \qquad \text{for } k \text{ large enough.}$$

Then $x+\frac{1}{k}d_k\to x$, hence by local minimality,

$$f(x+\tfrac{1}{k}d_k)\ \geq\ f(x) \qquad \text{for } k \text{ large enough.}$$

Since f is differentiable at x,

$$f(x+\tfrac{1}{k}d_k)=f(x)+\tfrac{1}{k}\langle\nabla f(x),d_k\rangle+o(\tfrac{1}{k}).$$

Letting $k\to\infty$ (so $k\,o(1/k)\to 0$ and $d_k\to d$) yields

$$\langle\nabla f(x),d\rangle\geq 0.$$

From local minimality to a tangent-direction condition ♡

Consider

$$(P) \qquad \min_{x \in X} f(x),$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable and $X \subset \mathbb{R}^n$ is closed. Let $x \in X$ be a local minimizer.
Take any sequence of unit vectors $d_k \to d$ such that

$$x + \tfrac{1}{k} d_k \in X \qquad \text{for } k \text{ large enough.}$$

Then $x + \frac{1}{k} d_k \to x$, hence by local minimality,

$$f(x + \tfrac{1}{k} d_k) \;\geq\; f(x) \qquad \text{for } k \text{ large enough.}$$

Since f is differentiable at x,

$$f(x + \tfrac{1}{k} d_k) = f(x) + \tfrac{1}{k}\langle \nabla f(x), d_k \rangle + o(\tfrac{1}{k}).$$

Letting $k \to \infty$ (so $k\, o(1/k) \to 0$ and $d_k \to d$) yields

$$\langle \nabla f(x), d \rangle \geq 0.$$

Tangent cones

For $f : \mathbb{R}^n \to \mathbb{R}$, we consider an optimization problem of the form

$$\underset{x \in X}{\text{Min}} \qquad f(x).$$

Definition (Bouligand cone)

We say that $d \in \mathbb{R}^n$ is tangent to X at $x \in X$ if there exists a sequence $x_k \in X$ converging to x and a sequence $t_k \searrow 0$ such that

$$d = \lim_k \frac{x_k - x}{t_k}.$$

Tangent cones

For $f : \mathbb{R}^n \to \mathbb{R}$, we consider an optimization problem of the form

$$\underset{x \in X}{\text{Min}} \qquad f(x).$$

Definition (Bouligand cone)

We say that $d \in \mathbb{R}^n$ is tangent to X at $x \in X$ if there exists a sequence $x_k \in X$ converging to x and a sequence $t_k \searrow 0$ such that

$$d = \lim_k \frac{x_k - x}{t_k}.$$

Let $T_X(x)$ be the tangent cone of X at x, that is, the set of all tangent vectors to X at x.

Tangent cones

For $f : \mathbb{R}^n \to \mathbb{R}$, we consider an optimization problem of the form

$$\underset{x \in X}{\text{Min}} \qquad f(x).$$

Definition (Bouligand cone)

We say that $d \in \mathbb{R}^n$ is tangent to X at $x \in X$ if there exists a sequence $x_k \in X$ converging to x and a sequence $t_k \searrow 0$ such that

$$d = \lim_k \frac{x_k - x}{t_k}.$$

Let $T_X(x)$ be the tangent cone of X at x, that is, the set of all tangent vectors to X at x.

Equivalently,

$$T_X(x) = \{ \, d \in \mathbb{R}^n \ \mid \exists t_k \searrow 0, \ \exists d_k \to d, \ x + t_k d_k \in X \}$$

Optimality conditions - differentiable case

Consider a function $f : \mathbb{R}^n \to \mathbb{R}$ and the optimization problem

$$(P) \qquad \underset{x \in X}{\mathrm{Min}} \qquad f(x).$$

If $x^\sharp \notin \mathrm{int}(X)$ we do not necessarily need to have $\nabla f(x^\sharp) = 0$, indeed we just to have $\langle d, \nabla f(x^\sharp) \rangle \geq 0$ for all "admissible" direction d.

Optimality conditions - differentiable case

Consider a function $f : \mathbb{R}^n \to \mathbb{R}$ and the optimization problem

$$(P) \qquad \underset{x \in X}{\text{Min}} \qquad f(x).$$

If $x^\sharp \notin \text{int}(X)$ we do not necessarily need to have $\nabla f(x^\sharp) = 0$, indeed we just to have $\langle d, \nabla f(x^\sharp)\rangle \geq 0$ for all "admissible" direction d.

Theorem

Assume that f is differentiable at $x^\sharp$.

1. *If $x^\sharp$ is a local minimizer of (P) we have*

$$\nabla f(x^\sharp) \in \left[T_X(x^\sharp)\right]^{\oplus}. \qquad (*)$$

2. *If f and X are both convex, and $(*)$ holds, then $x^\sharp$ is an optimal solution of (P)*

Optimality conditions - differentiable case

Consider a function $f : \mathbb{R}^n \to \mathbb{R}$ and the optimization problem

$$(P) \qquad \underset{x \in X}{\text{Min}} \qquad f(x).$$

If $x^\sharp \notin \mathrm{int}(X)$ we do not necessarily need to have $\nabla f(x^\sharp) = 0$, indeed we just to have $\langle d, \nabla f(x^\sharp)\rangle \geq 0$ for all "admissible" direction d.

Theorem

Assume that f is differentiable at $x^\sharp$.

1. *If $x^\sharp$ is a local minimizer of (P) we have*

$$\nabla f(x^\sharp) \in \left[T_X(x^\sharp)\right]^{\oplus}. \qquad (*)$$

2. *If f and X are both convex, and $(*)$ holds, then $x^\sharp$ is an optimal solution of (P)*

♠ Exercise: Prove this result.

Convex case

Let $K_X^{ad}(x)$ be the cone of admissible direction

$$K_X^{ad}(x) := \left\{ t(y-x) \in \mathbb{R}^n \quad | \quad y \in X, \quad t \geq 0 \right\}$$

Lemma

If $X \subset \mathbb{R}^n$ is convex, and $x \in X$, we have

$$T_X(x) = \overline{K_X^{ad}(x)}.$$

Recall that

$$T_X(x) = \{ \ d \in \mathbb{R}^n \ | \ \exists t_k \searrow 0, \ \exists d_k \to d, \ x + t_k d_k \in X \}$$

♠ Exercise: Prove this lemma

Differentiable constraints

We consider the following set of admissible solution

$$X = \Big\{x \in \mathbb{R}^n \quad | \quad g_i(x) = 0,\ i \in [n_E] \quad h_j(x) \leq 0,\ j \in [n_I]\Big\},$$

where g and h are differentiable functions.

Recall that the tangent cone is given by

$$T_X(x) = \{\ d \in \mathbb{R}^n\ \mid \exists t_k \searrow 0,\ \exists d_k \to d,\ g(x + t_k d_k) = 0,\ h(x + t_k d_k) \leq 0\}$$

We define the linearized tangent cone

$$\begin{aligned} T_X^\ell(x) := \{\ d \in \mathbb{R}^n\ \mid & \left\langle \nabla g_i(x)\,, d \right\rangle = 0,\ \forall i \in [n_E] \\ & \left\langle \nabla h_j(x)\,, d \right\rangle \leq 0,\ \forall j \in I_0(x)\} \end{aligned}$$

where

$$I_0(x) := \Big\{j \in [n_I] \quad | \quad h_j(x) = 0\Big\}.$$

Constraint qualifications

We always have

$$T_X(x) \subset T_X^\ell(x).$$

♣ Exercise: Prove it.

We say that the constraints are qualified[1] at x if

$$T_X(x) = T_X^\ell(x).$$

Constraint qualifications ensure the existence of Lagrange multipliers at local minima.

[1] This is Abadie CQ, there is a weaker version called Guignard CQ: $(T_X(x))^\oplus = (T_X^\ell(x))^\oplus$.

Constraint qualifications

We always have

$$T_X(x) \subset T_X^\ell(x).$$

♣ Exercise: Prove it.

We say that the constraints are qualified[1] at x if

$$T_X(x) = T_X^\ell(x).$$

Constraint qualifications ensure the existence of Lagrange multipliers at local minima.

[1] This is Abadie CQ, there is a weaker version called Guignard CQ: $(T_X(x))^\oplus = (T_X^\ell(x))^\oplus$.

Sufficient qualification conditions

Recall that g and h are assumed differentiable.
We denote the index set of active constraints at x

$$I_0(x) := \{j \in [n_I] \quad | \quad h_j(x) = 0\}.$$

The following conditions are sufficient qualification conditions at x:

1. g and h_i for $i \in I_0(x)$ are affine in a neighborhood of x;
2. (Slater) g is affine, h_j are convex, and there exists x_S such that $g(x_S) = 0$ and $h_j(x_S) < 0$;
3. (Mangasarian-Fromowitz) For all $\alpha \in \mathbb{R}^{n_E}$ and $\beta \in \mathbb{R}_+^{n_{I_0(x)}}$,

$$\sum_{i \in [n_E]} \alpha_i \nabla g_i(x) + \sum_{j \in I_0(x)} \beta_j \nabla h_j(x) = 0 \qquad \Longrightarrow \qquad \alpha = 0 \text{ and } \beta = 0$$

Expliciting the optimality condition

Under constraint qualification, the optimality condition reads

$$\nabla f(x) \in \left[T_X^\ell(x)\right]^\oplus$$

where

$$T_X^\ell(x) = \{\ d \in \mathbb{R}^n \ \mid \underbrace{\langle \nabla g_i(x) , d \rangle = 0, i \in [n_E] \quad \langle \nabla h_j(x) , d \rangle \leq 0, j \in I_0(x)}_{= A_x d \in C} \}.$$

with $A_x = \begin{pmatrix} \left((\nabla g_i(x))^\top\right)_{i \in [n_E]} \\ \left((\nabla h_j(x))^\top\right)_{j \in I_0(x)} \end{pmatrix}$ and $C = \{0\}^{n_E} \times (\mathbb{R}_-)^{n_I}$.

♣ Exercise: Show that $C^\oplus = \mathbb{R}^{n_E} \times (\mathbb{R}_-)^{|I_0(x)|}$

Expliciting the optimality condition II ◇

Recall that the positive dual cone of a set K is

$$K^{\oplus} := \{ d \in \mathbb{R}^n \quad | \quad \langle d, x \rangle \geq 0, \forall x \in K \}.$$

Let C be a closed convex cone. Consider

$$K = A^{-1}C := \{ x \in \mathbb{R}^n \mid Ax \in C \},$$

then

$$K^{\oplus} = \{ A^\top \lambda \mid \lambda \in C^{\oplus} \}.$$

♣ Exercise: prove it.

Hence,

$$\nabla f(x) \in \big[\underbrace{T_X^{\ell}(x)}_{A_x^{-1}C} \big]^{\oplus}$$

$$\iff \qquad \exists \lambda \in C^{\oplus}, \quad \nabla f(x) + A_x^\top \lambda = 0$$

$$\iff \quad \exists \lambda \in \mathbb{R}^{n_E}, \ \exists \mu \in \mathbb{R}_+^{I_0(x)} \quad \nabla f(x) + \sum^{n_E} \lambda_i \nabla g_i(x) + \sum \mu_j \nabla h_j(x) = 0.$$

Expliciting the optimality condition II ♢

Recall that the positive dual cone of a set K is

$$K^{\oplus} := \{d \in \mathbb{R}^n \quad | \quad \langle d, x \rangle \geq 0, \forall x \in K\}.$$

Let C be a closed convex cone. Consider

$$K = A^{-1}C := \{x \in \mathbb{R}^n \mid Ax \in C \},$$

then

$$K^{\oplus} = \{ A^{\top}\lambda \mid \lambda \in C^{\oplus}\}.$$

♣ Exercise: prove it.

Hence,

$$\nabla f(x) \in \big[\underbrace{T_X^{\ell}(x)}_{A_x^{-1}C}\big]^{\oplus}$$

$$\iff \qquad \exists \lambda \in C^{\oplus}, \quad \nabla f(x) + A_x^{\top}\lambda = 0$$

$$\iff \quad \exists \lambda \in \mathbb{R}^{n_E}, \ \exists \mu \in \mathbb{R}_+^{I_0(x)} \quad \nabla f(x) + \sum^{n_E} \lambda_i \nabla g_i(x) + \sum \mu_j \nabla h_j(x) = 0.$$

Expliciting the optimality condition II $\diamondsuit$

Recall that the positive dual cone of a set K is

$$K^{\oplus} := \{d \in \mathbb{R}^n \quad | \quad \langle d, x \rangle \geq 0, \forall x \in K\}.$$

Let C be a closed convex cone. Consider

$$K = A^{-1}C := \big\{x \in \mathbb{R}^n \mid Ax \in C \ \big\},$$

then

$$K^{\oplus} = \big\{ \ A^\top \lambda \mid \lambda \in C^{\oplus} \big\}.$$

♣ Exercise: prove it.

Hence,

$$\nabla f(x) \in \Big[\underbrace{T_X^{\ell}(x)}_{A_x^{-1}C} \Big]^{\oplus}$$

$$\iff \qquad \exists \lambda \in C^{\oplus}, \quad \nabla f(x) + A_x^\top \lambda = 0$$

$$\iff \quad \exists \lambda \in \mathbb{R}^{n_E}, \ \exists \mu \in \mathbb{R}_+^{I_0(x)} \quad \nabla f(x) + \sum^{n_E} \lambda_i \nabla g_i(x) + \sum \ \mu_j \nabla h_j(x) = 0.$$

Karush Kuhn Tucker condition

♡

Theorem (KKT)

Assume that the objective function f and the constraint function g_i and h_j are differentiable. Assume that the constraints are qualified at x.

Then if x is a local minimum of

$$\min_{\tilde{x}\in\mathbb{R}^n} \Big\{ f(\tilde{x}) \quad | \quad g_i(\tilde{x}) = 0,\ \forall i \in [n_E] \quad h_j(\tilde{x}) \leq 0,\ \forall j \in [n_I] \Big\}$$

then there exists dual variables λ, μ such that

$$(KKT) \quad \begin{cases} \nabla f(x) + \sum_{i=1}^{n_E} \lambda_i \nabla g_i(x) + \sum_{j=1}^{n_I} \mu_j \nabla h_j(x) = 0 & \nabla_x \mathcal{L} = 0 \\ g(x) = 0, \quad h(x) \leq 0 & \textit{Primal feasibility} \\ \lambda \in \mathbb{R}^{n_E}, \quad \mu \in \mathbb{R}_+^{n_I} & \textit{dual feasibility} \\ \mu_j h_j(x) = 0 \qquad \forall j \in [n_I] & \textit{complementarity constraint} \end{cases}$$

KKT conditions in the convex case

♡

Theorem (KKT is necessary and sufficient in convex case)

Assume that f and h_i are convex differentiable functions, and that g_j are affine functions. Moreover, assume that the Slater's condition holds: $\exists x_S \in \mathbb{R}^n$ such that

$$g_i(x_S) = 0 \ \forall i \in [n_E], \qquad h_j(x_S) < 0 \ \forall j \in [n_I].$$

Then x is a (global) minimum of

$$\min_{\tilde{x} \in \mathbb{R}^n} \left\{ f(\tilde{x}) \mid g_i(\tilde{x}) = 0 \ \forall i \in [n_E], \ \ h_j(\tilde{x}) \leq 0 \ \forall j \in [n_I] \right\}$$

if and only if there exist dual variables λ, μ such that (x, λ, μ) satisfy (KKT).

Exercise

Solve the following optimization problem

$$\begin{aligned} \underset{x,y\in\mathbb{R}^2}{\text{Min}} \quad & (x-1)^2 + (y-2)^2 \\ & x \leq y \\ & x + 2y \leq 2 \end{aligned}$$

Contents

What you have to know

- Basic vocabulary: objective, constraint, admissible solution, differentiable optimization problem
- First order necessary KKT conditions

What you really should know

- What is a tangent cone
- Sufficient qualification conditions (linear and Slater's)
- That KKT conditions are sufficient in the convex case

What you have to be able to do

- Write the KKT condition for a given explicit problem and use them to solve said problem

What you should be able to do

- Check that constraints are qualified