Chapter 4

Optimization of differentiable functions

4.1 Optimality conditions

In this chapter we are interested in solving the problem

$$\min_{x \in K} f(x) \tag{4.1.1}$$

where $f : E \to \mathbb{R}$ is a differentiable convex function such that

$$\lim_{|x| \to +\infty} f(x) = +\infty \tag{4.1.2}$$

and K is a closed convex subset of E.

Theorem 4.1.5 *Under the previous assumptions, the optimisation problem (4.1.1) has at least one solution. If moreover f is strictly convex then (4.1.1) has a unique solution.*

Proof. Since f is differentiable and convex, it is upper-semicontinuous by Proposition 3.2.16. Since K is closed we get from (4.1.2) that (4.1.1) has at least one solution.

Suppose in addition that f is strictly convex. Then if x_1^* and x_2^* are two different solutions, we have

$$f(\lambda x_1^* + (1-\lambda)x_2^*) < \lambda f(x_1^*) + (1-\lambda) f(x_2^*) \leq \min_K f$$

for $\lambda \in (0,1)$, which contradicts the convexity of K. □

We next have the following the following result.

Theorem 4.1.6 *Suppose that f is convex and differentiable and K is closed and convex. Then, x^* is solution of (4.1.1) if and only if*

$$\langle \nabla f(x^*), v - x^* \rangle \quad \geq \quad 0$$

for all $v \in K$.

Proof. This is a consequence of Proposition 3.2.16. □

Lagrange multipliers We consider the case where the set K is defined by $m \geq 1$ functions $F_1, \ldots, F_m$ from E to $\mathbb{R}$ as follows

$$K \quad = \quad \{x \in E \ : \ F_i(x) = 0 \text{ for } i = 1, \ldots, m\} \ .$$

Theorem 4.1.7 *Suppose that f is differentiable and that the functions $F_1, \ldots, F_m$ are C^1. Let $x^* \in E$ be solution to (4.1.1) such that $(\nabla F_i(x^*))_{i=1}^m$ is a family of linearly independent vectors of E. Then there exists $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ such that*

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla F_i(x^*) \quad = \quad 0 \ .$$

The coefficients $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ are called Lagrange multipliers.

We next turn to the case where the set K is defined by inequalities instead of equalities. More precisely, we suppose that K is given by

$$K \quad = \quad \{x \in E \ : \ F_i(x) \leq 0 \text{ for } i = 1, \ldots, m\} \ . \qquad (4.1.3)$$

before stating the main result, we need to introduce the following definition.

Definition 4.1.15 (Qualified constraints) *The constraints defining K in (4.1.3) are said to be qualified at $x^* \in K$ if there exists $w \in E$ such that*

$$\langle \nabla F_i(x^*), w \rangle \quad < \quad 0,$$

or

$$\langle \nabla Fi(x^*), w \rangle \quad = \quad 0 \textit{ and } F_i \textit{ is affine}$$

for any $i = 1, \dots, m$ such that $F_i(x^*) = 0$.

Theorem 4.1.8 *Suppose that K is given by (4.1.3). Let $x^* \in K$ such that the functions f and $(F_i)_{i=1}^m$ are differentiable at x^* and the constraints are qualified at x^*. Then, if x^* is a local minimizer of f over K, there exists $\lambda_1, \dots, \lambda_M \in \mathbb{R}_+$ called Lagrange multipliers, such that*

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla F_i(x^*) \quad = \quad 0\,,$$

and $\lambda_i = 0$ if $F_i(u) < 0$ for $i = 1, \dots, m$.

The main difference with the previous results lies in the constraints defining K which are inequalities instead of equalities. There exists another version of this theorem allowing to consider equality and inequality constraints. For the sake of clarity, this version is not presented.

Karush, Kuhn and Tucker conditions We go back to the convex framework and we suppose that the function f and the contraints are convex. We notice that if the functions $(F_i)_{i=1}^m$ are convex and differentiable, the set K defined by the (4.1.3) is closed and convex.

We also notice that an equality constraint given by an affine function F_i can be expressed as a double convex inequality constraint as follows:

$$F_i(x) \;=\; 0 \quad \Leftrightarrow \quad F_i(x) \;\leq\; 0 \text{ and } -F_i(x) \;\leq\; 0\,.$$

In this convex framework, we can state a necessary and sufficient condition of optimality as follows.

Theorem 4.1.9 *Suppose that the functions f and $(F_i)_{i=1}^m$ are convex, continuous on E and differentiable on K defined by*(4.1.3)*. Let $x^* \in K$ be such that the constraints are qualified at x^*. Then, x^* is a global minimum of f over K if and only if there exists $\lambda_1, \ldots, \lambda_M \in \mathbb{R}_+$ called Lagrange multipliers, such that*

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla F_i(x^*) \quad = \quad 0 \,,$$

and $\lambda_i = 0$ if $F_i(x^*) < 0$ for $i = 1, \ldots, m$.

4.2 Gradient descent

For a convex and differentiable function f, there exist several gradient descent methods allowing to construct minimizing sequences. In this section, we shall present a classical method using the gradient ∇f of f to approximate solutions of the minimization problem

$$\min_{x \in E} f(x). \tag{4.2.4}$$

We consider $\gamma \in (0, +\infty)$ called step size, $x_0 \in E$ and we define the sequence $(x_n)_{n \geq 0}$ by

$$x_{n+1} \quad = \quad x_n - \gamma \nabla f(x_n) \,, \quad n \geq 0 \,. \tag{4.2.5}$$

Denote by T the map from E to E defined by

$$T(x) \quad = \quad x - \gamma \nabla f(x) \quad = \quad (I - \gamma \nabla f)(x)$$

for $x \in E$, the sequence $(x_n)_{n \in \mathbb{N}}$ is given by

$$x_{n+1} = T(x_n) \,, \quad n \geq 0.$$

Without any assumption on ∇f, this algorithm may diverge, especially if γ is taken too large. However, if we suppose that ∇f is Lipschitz continuous, we are able to make this algorithm converge to a minimizer by choosing correctly γ as the following result shows.

Theorem 4.2.10 *Suppose that f is convex differentiable with grandient L-Lipschitz. Supoose also that f has a global minimum at $x^* \in E$. Then for any $x_0 \in E$, t and any $\gamma \in (0, \frac{2}{L})$ the sequence $(x_n)_{n\geq 0}$ defined by (4.2.5) converges to x^*.*

To prove this result, we first need the following lemma. In the sequel, for a given map T from E to E, a point $x \in E$ such that $T(x) = x$ is called a fixed point for the map T

Lemma 4.2.3 *Let T be a map from E to E, 1-Lipschitz continuous and admitting at least one fixed point. Let $x_0 \in E$ and $(x_n)_{n\geq 0}$ the sequence defined by (4.2.5). If $\lim_{n\to+\infty} |x_{n+1} - x_n| = 0$, $(x_n)_{n\geq 0}$ converges to a fixed point of T.*

Proof. Let y be a fixed point for T. Since T is 1-Lipschitz continuous, the sequence $(|x_n - y|)_{n\in\mathbb{N}}$ is nonincreasing and hence bounded. Therefore $(x_n)_{n\in\mathbb{N}}$ is bounded. Since E is finite a dimension space, the sequence $(x_n)_{n\in\mathbb{N}}$ admits a subsequence converging to $z \in E$. Since $\lim_{n\infty} |x_{n+1} - x_n| = 0$, we have $Tz = z$ and z is a fixed point for T. The sequence $|z - x_n|$ is then nonincreasing and admits a subsequence converging to 0, therefore, it converges to 0. □

To prove Theorem 4.2.10, we show that the operator T satisfies the assumptions of Lemma 4.2.3.

Lemma 4.2.4 *Let f be a convex differentiable function whose gradient ∇f is L-Lipschitz continuous. For $\gamma \in (0, \frac{2}{L}]$, the map $T = Id - \gamma\nabla f$ is 1-Lipschitz continuous.*

Proof. Define the operator $S = Id - T = \gamma\nabla f$. We then have

$$|x - y|^2 - |Tx - Ty|^2 \;=\; 2\langle Sx - Sy, x - y\rangle - |Sx - Sy|^2.$$

From Proposition 3.2.18 we have

$$\langle Sx - Sy, x - y\rangle \;\geq\; \frac{1}{\gamma L}|Sx - Sy|^2.$$

Therefore we get for $\gamma \leq \frac{2}{L}$

$$|x-y|^2 - |Tx - Ty|^2 \ \geq \ 0$$

and T is 1-Lipschitz continuous. $\square$

Proof of Theorem 4.2.10 We first notice that being a minimizer of f is equivalent to being a fixed point for T, since $T(x) = x$ is equivalent to $\nabla f(x) = 0$ and f is convex.

From Lemmata 4.2.3 and 4.2.4, it suffices to prove that $\lim_{n\infty} |x_{n+1} - x_n| = 0$.

For that we have by an application of first order Taylor formula and Lipschitz property of ∇f

$$f(z) \ \leq \ f(y) + \langle \nabla f(y), z - y\rangle + \frac{L}{2}|z-y|^2$$

for $y, z \in E$. We then apply this inequality with $y = x_n$ and $z = x_{n+1}$. Since $x_{n+1} - x_n = -\gamma \nabla f(x_n)$ we have

$$f(x_{n+1}) \ \leq \ f(x_n) - \frac{1}{\gamma}|x_{n+1} - x_n|^2 + \frac{L}{2}|x_{n+1} - x_n|^2 \ .$$

Since $\gamma < \frac{2}{L}$ we get

$$f(x_{n+1}) + \Big(\frac{1}{\gamma} - \frac{L}{2}\Big)|x_{n+1} - x_n|^2 \ \leq \ f(x_n) \ .$$

Hence, the sequence $(f(x_n))_{n\in\mathbb{N}}$ is nondecreasing. As it is lower bounded, it converges and hence

$$\lim_{n\to+\infty} |x_{n+1} - x_n| \ = \ 0 \ .$$