Chapter 2

Optimization, convex functions and sets

The basic ingredients of an optimization problem are a real-valued function and a subset of its domain over which looking for the minima and/or maxima of the function. Given any $f : \mathbb{R}^n \to \mathbb{R}$ and a set $D \subseteq \mathbb{R}^n$, the minimization problem

$$(P) \qquad \min\{f(x) \ : \ x \in D\}$$

amounts to finding $\bar{x} \in D$ such that $f(\bar{x}) \leq f(x)$ for any $x \in D$. The infimum of the set of real numbers $\{f(x) \ : \ x \in D\}$ is called the *optimal value* of (P) if it is finite. If the infimum is $-\infty$, then the minimization problem (P) is called *unbounded by below*. The corresponding definitions for maximization problems are given just recalling that maximizing f over a set D is equivalent to minimizing $-f$ over the same set. Optimization problems are often called programs, so that optimization and programming are synonyms in this framework.

Optimization problems can be classified according to the kind of function f and set D which are involved. *Unconstrained optimization* deals with the case $D = \mathbb{R}^n$, while *constrained optimization* with the case $D \neq \mathbb{R}^n$. If D is finite or countable, then *discrete optimization* comes into play: *combinatorial optimization* deals specifically with the case $D \subseteq \{0, 1\}^n$, while *integer programming* with the generic case $D \subseteq \mathbb{Z}^n$. If D is uncountable (and f is continuous), the most used term is *continuous optimization.* If f is linear and D is a polyhedron, then *linear programming* is used, otherwise *nonlinear optimization or nonlinear programming.*

These lecture notes focus on continuous nonlinear optimization both in the unconstrained and constrained case.

2.1 Optimization and convexity

Definition 2.1. $\bar{x} \in \mathbb{R}^n$ is called a *global minimum point* of (P) if it is feasible and $f(\bar{x})$ is the optimal value of (P), namely if

(i) $\bar{x} \in D$ (feasibility)

(ii) $f(\bar{x}) \leq f(x)$ for any $x \in D$ (optimality).

A global minimum point $\bar{x}$ is called *strict* if the strict inequality $f(\bar{x}) < f(x)$ holds for any $x \in D$ with $x \neq \bar{x}$.

Example 2.1. Take $n = 1$, $f(x) = -x^2$ and $D = \mathbb{R}$: the optimal value does not exist since $f(x) \to -\infty$ as $x \to \pm\infty$, hence no global mimimum point may exist.

Take (P) with f and D as in Example 1.6: the optimal value is 0 but no global minimum point exists all the same.

Definition 2.2. $\bar{x} \in \mathbb{R}^n$ is called a *local minimum point* of (P) if

(i) $\bar{x} \in D$ (feasibility)

(ii) $\exists\, \varepsilon > 0$ such that $f(\bar{x}) \leq f(x)$ for any $x \in D \cap B(\bar{x}, \varepsilon)$ (local optimality).

A local minimum point $\bar{x}$ is called *strict* if there exists $\varepsilon' > 0$ such that $f(\bar{x}) < f(x)$ holds for any $x \in D \cap B(\bar{x}, \varepsilon')$ with $x \neq \bar{x}$.

Clearly, any global minimum point is also a local minimum point but not vice versa. Indeed, finding a global minimum may be much harder than finding a local minimum: the distinction between *global optimization* and *local optimization* is generally very meaningful.

Definition 2.3. $D \subseteq \mathbb{R}^n$ is called *convex* if $\lambda x + (1-\lambda)y \in D$ for any $x, y \in D$ and any $\lambda \in [0, 1]$.

Example 2.2. $\mathbb{R}^n$, $\emptyset$, $B(x, \varepsilon)$, $[\ell, u] = \{x \in \mathbb{R}^n \ : \ \ell_i \leq x_i \leq u_i, \quad i = 1, ..., n\}$ with $\ell, u \in \mathbb{R}^n$ are convex sets in $\mathbb{R}^n$.

Definition 2.4. Let $D \subseteq \mathbb{R}^n$ be convex. Then, $f : \mathbb{R}^n \to \mathbb{R}$ is called *convex on D* if

$$\forall\, x, y \in D,\ \lambda \in [0, 1] \ : \ f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y).$$

f is called *strictly convex on D* if the strict inequality

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y)$$

holds whenever $x, y \in D$ and $\lambda \in [0, 1]$ satisfy $x \neq y$ and $\lambda \neq 0, 1$.

f is called *[strictly] concave on D* if $-f$ is [strictly] convex on D. If $D = \mathbb{R}^n$, then f is simply called [strictly] convex/concave, omitting the (convex) domain of convexity/concavity.

Notice that f is convex on D if and only if given any $k \in \mathbb{N}$ the inequality

$$f(\sum_{i=1}^{k} \lambda_i x_i) \leq \sum_{i=1}^{k} \lambda_i f(x_i)$$

holds for any $x_1, \ldots, x_k \in D$ and any $\lambda_1, \ldots, \lambda_k \in [0, 1]$ such that $\lambda_1 + \cdots + \lambda_k = 1$.

Theorem 2.1. *Let $D \subseteq \mathbb{R}^n$ be convex. Then,* *(i) If f is convex on D, then any local minimum point of (P) is also a global minimum point.*

(ii) If f is strictly convex on D, there exists at most one minimum point of (P).

Proof. (i) Ab absurdo, suppose there exists a local minimum point $\bar{x} \in D$ which is not a global minimum point. Thus, there exists $x \in D$ such that $f(x) < f(\bar{x})$. Consider $x(\lambda) = \lambda x + (1-\lambda)\bar{x}$: it belongs to D for any $\lambda \in [0,1]$ by the convexity of D. Furthermore, the convexity of f implies

$$f(x(\lambda)) \leq \lambda f(x) + (1-\lambda)f(\bar{x}) < \lambda f(\bar{x}) + (1-\lambda)f(\bar{x}) = f(\bar{x})$$

for any $\lambda \neq 0$. Since $x(\lambda) \to \bar{x}$ as $\lambda \to 0$, for any $\varepsilon > 0$ there exists $\bar{\lambda} \in [0,1[$ such that $x(\bar{\lambda}) \in B(\bar{x}, \varepsilon)$. Thus, $\bar{x}$ is not a local minimum point contradicting the assumption.

(ii) Suppose there exist $\bar{x}, \hat{x} \in D$ which are both minimum points of (P). Thus, $f(\hat{x}) = f(\bar{x})$. Take any $\lambda \in]0,1[$ and $x(\lambda) = \lambda\hat{x} + (1-\lambda)\bar{x}$. If $\bar{x} \neq \hat{x}$, then the strict convexity of f implies

$$f(x(\lambda)) < \lambda f(\hat{x}) + (1-\lambda)f(\bar{x}) = \lambda f(\bar{x}) + (1-\lambda)f(\bar{x}) = f(\bar{x}) = f(\hat{x})$$

so that neither $\bar{x}$ nor $\hat{x}$ is a minimum point of (P). □

The theorem shows that local and global optimization coincide if f and D are both convex, therefore the distinction between *convex optimization* and *nonconvex optimization* is very meaningful as well.

Proposition 2.1. *Let $D \subseteq \mathbb{R}^n$ be convex and f be convex on D. Then, the set of all the minimum points of (P) is convex*

Proof. Take any two minimum points $\bar{x}, \hat{x} \in D$ $(\hat{x} \neq \bar{x})$ and any $\lambda \in [0,1]$. Then, $x(\lambda) = \lambda\hat{x} + (1-\lambda)\bar{x} \in D$ by the convexity of D, while the convexity of f implies

$$f(x(\lambda)) \leq \lambda f(\hat{x}) + (1-\lambda)f(\bar{x}) = \lambda f(\bar{x}) + (1-\lambda)f(\bar{x}) = f(\bar{x}) = f(\hat{x}) \leq f(x(\lambda))$$

where the last inequality is due to the optimality of $\bar{x}$. Thus, $f(x(\lambda)) = f(\bar{x}) = f(\hat{x})$ and therefore $x(\lambda)$ is a minimum point as well. □

2.2 Properties of convex functions

Proposition 2.2. *Let $D \subseteq \mathbb{R}^n$ be convex. Then, f is convex on D if and only if the restriction of the epigraph of f to D, namely*

$$\mathrm{epi}_D(f) = \{(x,t) \ : \ x \in D, \ t \geq f(x)\}$$

is a convex set in $\mathbb{R}^{n+1}$.

Proof. *Only if)* Take any $(x,t),(y,\tau) \in \mathrm{epi}_D(f)$ and any $\lambda \in [0,1]$:

$$\lambda t + (1-\lambda)\tau \geq \lambda f(x) + (1-\lambda) f(y) \geq f(\lambda x + (1-\lambda)y)$$

where the last inequality is due to the convexity of f. Moreover, $\lambda x + (1-\lambda)y \in D$ since D is convex, and thus $(\lambda x + (1-\lambda)y, \lambda t + (1-\lambda)\tau) \in \mathrm{epi}_D(f)$.

If) Take any $x, y \in D$ and any $\lambda \in [0,1]$. Therefore, $(x, f(x)), (y, f(y)) \in \mathrm{epi}_D(f)$ and the the convexity of $\mathrm{epi}_D(f)$ imply $(\lambda x+(1-\lambda)y, \lambda f(x)+(1-\lambda)f(y)) \in \mathrm{epi}_D(f)$, which reads

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y).$$ □

Proposition 2.3. *Let $D \subseteq \mathbb{R}^n$ be convex and f be convex on D. Then, the intersection of the α-sublevel set of f with D, namely*

$$\{x \in \mathbb{R}^n \ : \ f(x) \leq \alpha\} \cap D$$

is a convex set for any $\alpha \in \mathbb{R}$.

Proof. Take any $x, y \in D$ and any $\lambda \in [0,1]$: $\lambda x + (1-\lambda)y \in D$ since D is convex and moreover

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y). \leq \lambda\alpha + (1-\lambda)\alpha = \alpha$$ □

Example 2.3. Take $n = 1$, $f(x) = x^3$ and $D = \mathbb{R}$. The α-sublevel set

$$\{x \in \mathbb{R}^n \ : \ x^3 \leq \alpha\} =]-\infty, \sqrt[3]{\alpha}]$$

is convex, while f is not convex. In fact, it is enough to take $x = -1$, $y = 1/2$ and $\lambda = 1/3$ to get $f(\lambda x + (1-\lambda)y) = f(0) = 0 > -1/4 = \lambda f(x) + (1-\lambda) f(y)$.

Proposition 2.4. *Let $D \subseteq \mathbb{R}^n$ be convex and $f_i : \mathbb{R}^n \to \mathbb{R}$ be convex on D for any $i \in I$ for some index set I.*

(i) If I is finite, then $\Big(\sum_{i \in I} f_i\Big)(x) = \sum_{i \in I} f_i(x)$ is convex on D;

(ii) $\Big(\sup_{i \in I} f_i\Big)(x) = \sup_{i \in I} f_i(x)$ is convex on D.

Proof. (i) It is enough to exploit the definition of convexity for each f_i summing all the inequalities.

(ii) Take any $x, y \in D$ and any $\lambda \in [0,1]$. Given any $\varepsilon > 0$, there exists $k = k(\varepsilon) \in I$ such that

$$\begin{aligned}
\Big(\sup_{i \in I} f_i\Big)(\lambda x + (1-\lambda)y) &\leq f_k(\lambda x + (1-\lambda)y) + \varepsilon \\
&\leq \lambda f_k(x) + (1-\lambda) f_k(y) + \varepsilon \\
&\leq \lambda \Big(\sup_{i \in I} f_i\Big)(x) + (1-\lambda)\Big(\sup_{i \in I} f_i\Big)(y) + \varepsilon
\end{aligned}$$

and therefore

$$\Big(\sup_{i\in I} f_i\Big)(\lambda x + (1-\lambda)y) \leq \lambda\Big(\sup_{i\in I} f_i\Big)(x) + (1-\lambda)\Big(\sup_{i\in I} f_i\Big)(y)$$

follows since ε is arbitrary. □

Theorem 2.2. *Let $D \subseteq \mathbb{R}^n$ be a convex set with a nonempty interior. If f is convex on D, then f is continuous on $\operatorname{int} D$.*

Theorem 2.3. *Let f be differentiable on D. Then, f is convex on D if and only if*

$$\forall\ x, y \in D\ :\ f(y) \geq f(x) + \nabla f(x)^T (y - x).$$

Proof. *Only if)* Take any $x, y \in D$ and any $\lambda \in [0,1]$: the definition of convexity can be equivalently stated as

$$f(y) - f(x) \geq [f(x + \lambda(y-x)) - f(x)]/\lambda.$$

By Proposition 1.6 $\lim\limits_{\lambda\to 0}\ [f(x+\lambda(y-x)) - f(x)]/\lambda = \nabla f(x)^T(y-x)$ Therefore, the required inequality follows from the above one just taking the limit as $\lambda \to 0^+$.

If) Take any $x, y \in D$ and any $\lambda \in [0,1]$: the following inequalities follow from the assumption just considering the pairs of points x, $\lambda x + (1-\lambda)y$ and y, $\lambda x + (1-\lambda)y$:

$$\begin{aligned} f(x) &\geq f(\lambda x + (1-\lambda)y) + (1-\lambda)\nabla f(\lambda x + (1-\lambda)y)^T(x-y),\\ f(y) &\geq f(\lambda x + (1-\lambda)y) + \lambda\nabla f(\lambda x + (1-\lambda)y)^T(x-y). \end{aligned}$$

Summing λ times the first inequality with $(1-\lambda)$ times the second inequality gives

$$\lambda f(x) + (1-\lambda)f(y) \geq f(\lambda x + (1-\lambda)y).$$ □

Roughly speaking, the theorem states that the convexity of f means exactly that the graph of f is "above" the tangent hyperplane to the graph at $(x, f(x))$ everywhere on D for any $x \in D$.

An analogous characterization holds for strict convexity.

Theorem 2.4. *Let f be differentiable on D. Then, f is strictly convex on D if and only if*

$$\forall\ x, y \in D \text{ s.t. } x \neq y\ :\ f(y) > f(x) + \nabla f(x)^T(y-x).$$

Theorem 2.5. *Let f be twice continuously differentiable (on $\mathbb{R}^n$). Then, f is convex if and only if $\nabla^2 f(x)$ is positive semidefinite for any $x \in \mathbb{R}^n$, i.e.,*

$$\forall\ x, y \in \mathbb{R}^n\ :\ y^T\nabla^2 f(x) y \geq 0.$$

Proof. *Only if)* Take any $x, y \in \mathbb{R}^n$ and any $t \in \mathbb{R}$: the second-order Taylor's formula and Theorem 2.3 guarantee

$$\frac{1}{2}t^2 y^T \nabla^2 f(x) y + r_{(f,x)}(ty) = f(x+ty) - f(x) - t\nabla f(x)^T y \geq 0$$

and therefore (supposing $y \neq 0$)

$$\frac{1}{2}y^T \nabla^2 f(x) y + \frac{r_{(f,x)}(ty)}{\|ty\|_2^2}\|y\|_2^2 \geq 0.$$

Taking the limit as $t \to 0$, the inequality $y^T \nabla^2 f(x) y \geq 0$ follows.

If) Take any $x, y \in \mathbb{R}^n$. By Theorem 1.8 (i) there exists $t \in [0,1]$ such that

$$f(y) - f(x) - \nabla f(x)^T (y-x) = \frac{1}{2}(y-x)^T \nabla^2 f(x + t(y-x))(y-x) \geq 0$$

where the inequality is due to the positive semidefiniteness of $\nabla^2 f(x + t(y-x))$. Therefore, Theorem 2.3 guarantees that f is convex. □

A similar sufficient condition for strict convexity can be proved in the same way.

Theorem 2.6. *Let f be twice continuously differentiable (on $\mathbb{R}^n$). If $\nabla^2 f(x)$ is positive definite for any $x \in \mathbb{R}^n$, i.e.,*

$$\forall\ x, y \in \mathbb{R}^n \ : \ y^T \nabla^2 f(x) y > 0,$$

then f is strictly convex.

Example 2.4. Take $n = 1$ and $f(x) = x^4$: f is strictly convex but $\nabla^2 f(0) = f''(0) = 0$.

Theorem 2.7. *Let $f(x) = \frac{1}{2}x^T Q x + b^T x + c$ (with $Q = Q^T$). Then,*

(i) f is convex if and only if Q is positive semidefinite;

(ii) f is strictly convex if and only if Q is positive definite.

Proof. (i) and the "if part" of (ii) are just Theorems 2.5 and 2.6 for the quadratic function f since $\nabla^2 f(x) = Q$ for any $x \in \mathbb{R}^n$.

(ii) *Only if)* The residual in the second-order Taylor's formula is zero, that is

$$\frac{1}{2}t^2 y^T Q y = f(x+ty) - f(x) - t\nabla f(x)^T y$$

for any $x, y \in \mathbb{R}^n$. By Theorem 2.4 the right-hand side is always positive, hence Q is positive definite. □