Continuous Optimization

Chapter 3: Constrained Optimization

1 Definitions

In this chapter we will consider constrained optimization problems with the following shape

$$\begin{aligned} \min \;& f(x) \\ \text{s.t.} \;& x \in C \end{aligned} \tag{1}$$

Definition 1.1 (Convex Set)**.** *A set C is said to be convex if given $x_1, x_2 \in C$ and $\lambda \in [0,1]$, then $\lambda x_1 + (1-\lambda)x_2 \in C$.*

Definition 1.2 (Convex Function)**.** *A function $f : C \to \mathbb{R}$ defined on a convex set C is said to be convex if given $x_1, x_2 \in C$ and $\lambda \in [0,1]$, then*

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2).$$

Definition 1.3 (Strictly Convex Function)**.** *A function $f : C \to \mathbb{R}$ defined on a convex set C is said to be strictly convex if given $x_1, x_2 \in C$ and $\lambda \in [0,1]$, then*

$$f(\lambda x_1 + (1-\lambda)x_2) < \lambda f(x_1) + (1-\lambda) f(x_2).$$

A function is called concave if $-f$ is convex and strictly concave if $-f$ is strictly convex.
Now, given Δ_k the unit-simplex, that is the subset of $\mathbb{R}^k$ comprising all nonnegative vectors whose sum is 1, i.e.,

$$\{\lambda \in \mathbb{R}^k : \lambda \geq 0, e^t\lambda = 1\},$$

we can provide the following very useful result by Jensen's.

Theorem 1.1 (Jensen's Inequality)**.** *Let $f : C \to \mathbb{R}$ be a convex function over a convex set C. Then for any $x_1, x_2, \ldots, x_k \in C$ and $\lambda \in \Delta_k$ we have*

$$f\left(\sum_{i=1}^{k} \lambda_i x_i\right) \leq \sum_{i=1}^{k} \lambda_i f(x_i). \tag{2}$$

Proof. We will prove (2) by induction on k. For $k = 1$ the result is obvious ($f(x_1) \leq f(x_1) \;\; \forall x_1 \in C$). We now assume that (2) holds for k and we will prove that is also holds for $k+1$. Suppose we have $x_1, x_2, \ldots, x_{k+1} \in C$ and $\lambda \in \Delta_{k+1}$, we will show that $f(z) \leq \sum_{i=1}^{k+1} \lambda_i f(x_i)$ with $z = \sum_{i=1}^{k+1} \lambda_i x_i$. If $\lambda_{k+1} = 1$, then $z = x_{k+1}$ and (2) is obvious. If $\lambda_{k+1} < 1$, then

$$\begin{aligned} f(z) &= f\left(\sum_{i=1}^{k} \lambda_i x_i + \lambda_{k+1} x_{k+1}\right) \\ &= f\left((1-\lambda_{k+1}) \sum_{i=1}^{k} \frac{\lambda_i}{1-\lambda_{k+1}} x_i + \lambda_{k+1} x_{k+1}\right) \\ &\leq (1-\lambda_{k+1}) f(v) + \lambda_{k+1} f(x_{k+1}), \end{aligned}$$

with $v = \frac{\lambda_i}{1-\lambda_{k+1}} x_i$. Since $\sum_{i=1}^{k} \frac{\lambda_i}{1-\lambda_{k+1}} = \frac{1-\lambda_{k+1}}{1-\lambda_{k+1}} = 1$, it follows that v is a convex combination of k points from C, hence by the induction hypotesis we have that $f(v) \leq \sum_{i=1}^{k} \frac{\lambda_i}{1-\lambda_{k+1}} f(x_i)$, which combined with the equality above yields

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

□

2 Characterizations of Convex Functions

Theorem 2.1 (Gradient characterization of convex functions)**.** *Let $f \in C^1(C)$, where C is convex. Then f is convex over C if and only if*

$$f(x) + \nabla f(x)^T (y - x) \leq f(y) \quad \forall x, y \in C. \tag{3}$$

Proof. Exercise. □

Proposition 2.1 (Sufficiency of stationarity under convexity)**.** *Let $f \in C^1(C)$, where $C \subseteq \mathbb{R}^n$ is convex. Suppose that $\nabla f(x^*) = 0$ for some $x^* \in C$. Then x^* is a global minimizer of f over C.*

Proof. Let $z \in C$. Plugging $x = x^*$ and $y = z$ in Theorem 2.1 we obtain that

$$f(z) \geq f(x^*) + \nabla f(x^*)^T (z - x^*),$$

which implies that $f(z) \geq f(x^*)$ because $\nabla f(x^*) = 0$. □

We note that Proposition 2.1 establishes only the sufficiency of the stationarity condition $\nabla f(x^*) = 0$ for guaranteeing that x^* is a global optimal solution. When $C$is not the entire space, this condition is not necessary, in fact it might be that the points for which $\nabla f(\cdot) = 0$ are not in C. On the other hand, when $C = \mathbb{R}^n$ and f is convex, $\nabla f(x^*) = 0$ is both sufficient and necessary condition for x^* to be a global minimum. We can now establish the conditions under which a twice continuously differentiable function f is convex.

Theorem 2.2 (Second order characterization of convexity)**.** *Let $f \in C^2(C)$, where $C \subseteq \mathbb{R}^n$ is convex and open. Thus, we have that f is convex iff $\nabla^2 f(x) \succcurlyeq 0 \quad \forall x \in C$.*

Proof. Suppose that $\nabla^2 f(x) \succcurlyeq 0$ for all $x \in C$. We will prove (3) which is enough to establish convexity. Let $x, y \in C$, then by the Mean Value Theorem[2] (Theorem 2.6 from Chapter 1) we get that there exists $z \in [x, y]$ (and hence $z \in C$) for which

$$f(y) = f(x) + \nabla f(x)^T (y - x) + \frac{1}{2}(y - x)^T \nabla^2 f(z)(y - x). \tag{4}$$

Since $\nabla^2 f(z) \succcurlyeq 0$, it follows that $(y - x)^T \nabla^2 f(z)(y - x) \geq 0$, which implies (3). To prove the opposite direction, assume that f is convex over C. Let $x \in C$ and $y \in \mathbb{R}^n$. Since C is open, it follows that $x + \lambda y \in C$, for $0 < \lambda < \epsilon$, where ϵ is a small enough positive constant. Using now the gradient characterization of convex functions (3) we get

$$f(x + \lambda y) \geq f(x) + \lambda \nabla f(x)^T y.$$

In addition, by the quadratic approximation theorem (Theorem 2.4 from Chapter 1), we have that

$$f(x + \lambda y) = f(x) + \lambda \nabla f(x)^T y + \frac{\lambda^2}{2} y^t \nabla^2 f(x) y + o(\lambda^2 ||y||^2),$$

which combined with the above inequality gives

$$\frac{\lambda^2}{2} y^t \nabla^2 f(x) y + o(\lambda^2 ||y||^2) \geq 0 \quad \forall \lambda \in (0, \epsilon).$$

Dividing the latter inequality by λ^2 and taking the limit for $\lambda \to 0^+$, we have

$$\frac{\lambda^2}{2} y^t \nabla^2 f(x) y \geq 0 \quad \forall y \in \mathbb{R}^n,$$

which concludes the proof. □

The same theorem works with positive definiteness and strict convexity, meaning also that the minimum in this case is unique.

3 Optimization over convex problems

From now on, we consider (1) where f and C are convex. As a direct consequence of the convexity of f we have the following two theorems.

Theorem 3.1 (global=local in convex optimization)**.** *Let $f : C \to \mathbb{R}$ be a convex function over a convex set $C \subseteq \mathbb{R}^n$. Let $x^* \in C$ be a local minimum of f over C. Then x^* is a global minimum of f over C.*

Proof. Since x^* is a local minimum of f over C there exists r such that $f(x) \geq f(x^*)$ for any $x \in C \cap B[x^*, r]$. Now let $y \in C$ with $y \neq x^*$. We want to show that $f(y) \geq f(x^*)$. Let $\lambda \in (0, 1]$ be such that $x^* + \lambda(y - x^*)$. Let $\lambda \in (0, 1]$ be such that $x^* + \lambda(y - x^*) \in B[x^*, r]$, for instance $\lambda = \frac{r}{||y - x^*||}$. Now, since $x^* + \lambda(y - x^*) \in C$, it follows that $f(x^*) \leq f(x^* + \lambda(y - x^*))$, and hence, by convexity of f, also

$$f(x^*) \leq f(x^* + \lambda(y - x^*)) \leq (1 - \lambda)f(x^*) + \lambda f(y)$$

Thus, $\lambda f(x^*) \leq \lambda f(y)$, which concludes the proof. □

Theorem 3.2 (Convexity of the optimal set in convex optimization)**.** *Let $f : C \to \mathbb{R}$ be a convex function with $C \subseteq \mathbb{R}^n$ convex. Then, the set of optimal solutions of the problem* (1)*, which we denote by X^* is convex. Moreover, if f is strictly convex over C, then there exists at most one optimal solution.*

Proof. If $X^* = \emptyset$, the result follows trivially. Suppose that $X \neq \emptyset$ and denote the optimal value of f by f^*. Let $x, y \in C$ with $\lambda \in [0, 1]$. Then, by convexity $f(\lambda x + (1 - \lambda)y) \leq \lambda f^* + (1 - \lambda)f^* = f^*$, hence $\lambda x + (1 - \lambda)y$ is also optimal, i.e., it belongs to X^*, establishing the convexity of X^*. Suppose now that f is strictly convex and X^* is nonempty, and suppose by contradiction that there are 2 points x, y in X^*. Then $\lambda x + (1 - \lambda)y \in C$, and by the strict convexity of f we have

$$f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y) = f^*,$$

which is a contradiction to the fact that f^* is the optimal value. □

3.1 Stationarity

Note that the following definition and the following theorem are given also for the more general case in which f is not convex.

Definition 3.1 (Stationary points of convex constrained problems)**.** *Let $f \in \mathrm{C}^1(C)$, where C is closed and convex. Then x^* is a stationary point of* (1) *if $\nabla f(x^*)(x - x^*) \geq 0 \ \forall x \in C$.*

In words, this means that there are no feasible descent directions of f at x^*. This suggests that stationarity is in fact a necessary condition for a local minimum of (1).

Theorem 3.3 (Stationarity as necessary optimality condition of a convex constrained problem)**.** *Let $f \in \mathrm{C}^1(C)$, where C is closed and convex and let x^* be a local minimum of* (1)*. Then x^* is a stationary point of* (1)*.*

Proof. Let x^* be a local minimum of f and assume by contradiction that is not a stationary point of (1). Then there exists $x \in C$ such that $\nabla f(x^*)(x - x^*) < 0$. Therefore, $f'(x, d) < 0$, where $d = x - x^*$. Hence, by Lemma 1.1 of Chapter 2, there exists $\epsilon \in (0, 1)$ such that $f(x^* + td) < f(x^*) \ \forall t \in (0, \epsilon)$. Since C is convex, we have that $x + td = (1 - t)x^* + tx \in C$, leading to the conclusion that x^* is not a local optimum of (1), which is a contradiction. □

Theorem 3.4 (Stationarity as necessary and sufficient optimality condition for a convex problem)**.** *Let $f \in \mathrm{C}^1(C)$, where C is closed and convex and f is also convex. Let x^* be a local minimum of* (1)*. Then x^* is a stationary point of* (1) *iff x^* is a optimal solution of* (1)*.*

Proof. The necessity of the stationarity condition follows from Theorem 3.3. To prove the sufficiency, assume that x^* is a stationary point of (1) and let $x \in C$. Then, the gradient characterization of convex functions (3) and stationarity of x^*, we get

$$f(x) \geq f(x^*) + \nabla f(x^*)^T(x - x^*) \geq f(x^*),$$

which concludes the proof. □

References