Optimality Conditions

1 Local and global minimum

Let $f : C \subseteq \mathbb{R}^n \to \mathbb{R}$ be a function. We say that $\mathbf{x} \in C$ is its *global minimizer* over the set C if $f(\mathbf{x}) \leq f(\mathbf{y})$ for every $\mathbf{y} \in C$. The value $f(\mathbf{x})$ is called the *global minimum* of f. This is the lower bound on the values f can attain on C. We will often encounter loose terminology, in which the word "minimum" refers both to the minimum value and the minimizer.

In practice, it is very difficult to know any global properties about a function; in sharp contrast, local properties are readily available for most well-behaved (e.g., C^n) functions through their first- and higher-order derivatives. We say that $\mathbf{x} \in C$ is a *local minimizer* of f over the set C if there exists $\epsilon > 0$ such that $\mathbf{x}$ is a global minimizer of f over $B_\epsilon(\mathbf{x}) \cap C$, where $B_\epsilon(\mathbf{x}) = \{\mathbf{y} : \|\mathbf{x} - \mathbf{y}\| \leq \epsilon\}$ is the ϵ-ball. The value of $f(\mathbf{x})$ is called the *local minimum.*

We will see that many practical optimization algorithms are guaranteed to find a local minimum of a function. However, without additional assumptions, no practical (i.e., polynomial time) algorithm capable of finding the global minimum is known (that is why we can say that general optimization problems are *unsolvable*).

2 Minima of convex functions

Convex functions are of paramount importance in optimization due to the following important property.

Theorem 1. *A local minimizer of a convex function is also a global minimizer.*

The property means, in simple words, that if the (unconstrained) optimization problem involves a convex objective, local optimization methods are guaranteed to find its global minimum. To prove it, let $f : C \subseteq \mathbb{R}^n \to \mathbb{R}$ be a convex function, and let $\mathbf{x}$ be its local minimum. We assume by contradiction that $\mathbf{x}$ is not a global minimum, meaning that there exists $\mathbf{y} \in C$ such that $f(\mathbf{y}) < f(\mathbf{x})$. By definition of a convex function, $f((1-\alpha)\mathbf{x} + \alpha\mathbf{y}) \leq (1-\alpha)f(\mathbf{x}) + \alpha f(\mathbf{y})$, or

$$f(\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x})) \leq f(\mathbf{x}) + \alpha(f(\mathbf{y}) - f(\mathbf{x})).$$

By our assumption, $f(\mathbf{y}) - f(\mathbf{x}) < 0$, so for every $\alpha \in (0, 1)$,

$$f(\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x})) < f(\mathbf{x}).$$

This implies, however, that for every $\epsilon > 0$, we can choose $\alpha = \epsilon/2$ such that $\mathbf{z} = \mathbf{x} + \alpha(\mathbf{y} - \mathbf{x})$ is in the ball $B_\epsilon(\mathbf{x})$ and $f(\mathbf{z}) < f(\mathbf{x})$. This leads to the contradiction of the fact that $\mathbf{x}$ is a local minimizer of f.

Theorem 2. *A strictly convex function has at most one global minimizer.*

To prove this result, let $\mathbf{x}$ and $\mathbf{y}$ be two global minimizers of f; we will denote $f(\mathbf{x}) = f(\mathbf{y}) = q$. Denote $\mathbf{z} = \frac{1}{2}(\mathbf{x} + \mathbf{y})$. By strong convexity,

$$f(\mathbf{z}) = f\left(\frac{\mathbf{x} + \mathbf{y}}{2}\right) < \frac{f(\mathbf{x}) + f(\mathbf{y})}{2} = q,$$

implying that $f(\mathbf{z}) < q$ in contradiction to the fact that q is global minimum.

This result means that a strictly convex function can have either one or no global minimizers. As an example of the latter case, consider the strictly convex function $f(x) = 1/x$ on $[1, \infty)$. The function has the infimum value 0, but there is no point at which it is achieved; hence, f has no minimizer.

3 Optimality conditions for convex functions

As we will see, unconstrained minimization algorithms will use local information (the gradient and, sometimes, the Hessian) in order to find the (local) minimum of the objective. The same local information will be used to stop the algorithm, i.e., to determine when the point at the current iteration is (sufficiently close to) a local minimum. In the sequel, we derive the conditions a local minimum has to satisfy, expressed in terms of the function derivatives.

For convex functions, the following necessary and sufficient optimality condition can be expressed using first-order derivatives:

Theorem 3. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a C^1 convex function. Then, $\mathbf{x}^* \in \mathbb{R}^n$ is its global minimum over $\mathbb{R}^n$ iff $\nabla f(\mathbf{x}^*) = 0$.*

We will prove the necessity of this condition in the sequel for general (not necessarily convex) functions. To prove the other direction, recall the gradient inequality

$$f(\mathbf{x}^* + \mathbf{r}) \geq f(\mathbf{x}^*) + \mathbf{r}^{\mathrm{T}} \nabla f(\mathbf{x}^*).$$

holding for every $\mathbf{r}$. Since the second term vanishes, we have

$$f(\mathbf{x}^* + \mathbf{r}) \geq f(\mathbf{x}^*),$$

meaning that $\mathbf{x}^*$ is a global minimizer of f.

4 Optimality conditions for non-convex functions

Unlike for convex functions, in general vanishing gradient is a necessary yet insufficient optimality condition (that is, if $\mathbf{x}^*$ is a local minimizer, then $\nabla f(\mathbf{x}^*) = 0$, but not necessarily vice versa – as an example, consider $f(x) = x^3$ which has $f'(0) = 0$ but no minimum at $x = 0$).

Theorem 4 (Necessary optimality conditions)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a C^1 function, and let $\mathbf{x}^*$ be its local minimizer. Then,*

1. $\nabla f(\mathbf{x}^*) = 0$.

2. *If f is C^2, then,* $\nabla^2 f(\mathbf{x}^*) \succeq 0$.

To prove the first (first-order) condition, we use the first-order Taylor expansion

$$f(\mathbf{x}^* + \mathbf{r}) = f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^{\mathrm{T}}\mathbf{r} + O(\|\mathbf{r}\|^2).$$

Setting $\mathbf{y} = \mathbf{x}^* - \alpha \nabla f(\mathbf{x}^*)$, and recalling that $\mathbf{x}^*$ is a local minimizer of f, we have for a sufficiently small $\alpha > 0$

$$\begin{aligned}
0 \le \frac{1}{\alpha}(f(\mathbf{y}) - f(\mathbf{x}^*)) &= \frac{1}{\alpha}(f(\mathbf{x}^* - \alpha\nabla f(\mathbf{x}^*)) - f(\mathbf{x}^*)) \\
&= \frac{1}{\alpha}\left(-\nabla f(\mathbf{x}^*)^{\mathrm{T}}(\alpha\nabla f(\mathbf{x}^*)) + O(\|\alpha\nabla f(\mathbf{x}^*)\|^2)\right) \\
&= -\|\nabla f(\mathbf{x}^*)\|^2 + \alpha O(\|\nabla f(\mathbf{x}^*)\|^2).
\end{aligned}$$

In the limit $\alpha \downarrow 0$, we have $0 \le -\|\nabla f(\mathbf{x}^*)\|^2 \le 0$. Hence, $\nabla f(\mathbf{x}^*) = 0$.

To prove the second (second-order) condition, let us fix some $\mathbf{r} \in \mathbb{R}^n$, $\alpha > 0$, and invoke the second-order Taylor expansion (into which we already substituted the former result, $\nabla f(\mathbf{x}^*) = 0$):

$$f(\mathbf{x}^* + \alpha\mathbf{r}) = f(\mathbf{x}^*) + \frac{1}{2}(\alpha\mathbf{r})^{\mathrm{T}}\nabla^2 f(\mathbf{x}^*)(\alpha\mathbf{r}) + O(\|\alpha\mathbf{r}\|^3).$$

Again, recalling that $\mathbf{x}^*$ is a local minimizer, for a sufficiently small α we have

$$0 \le \frac{1}{\alpha^2}(f(\mathbf{x}^* + \alpha\mathbf{r}) - f(\mathbf{x}^*)) = \frac{1}{2}\mathbf{r}^{\mathrm{T}}\nabla^2 f(\mathbf{x}^*)\mathbf{r} + \alpha O(\|\mathbf{r}\|^3).$$

Taking the limit $\alpha \downarrow 0$, we have $\mathbf{r}^{\mathrm{T}}\nabla^2 f(\mathbf{x}^*)\mathbf{r} \ge 0$. Since the latter holds for every $\mathbf{r}$, $\nabla^2 f(\mathbf{x}^*) \succeq 0$.

Theorem 5 (Sufficient optimality conditions)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a C^2 function. If*

1. $\nabla f(\mathbf{x}^*) = 0$, *and*

2. $\nabla^2 f(\mathbf{x}^*) \succ 0$,

then, $\mathbf{x}^*$ is a local minimizer of f.

Exercise 1. *Prove the theorem.*