Optimality conditions for unconstrained optimization

Daniel P. Robinson
Department of Applied Mathematics and Statistics
Johns Hopkins University

September 2, 2020

The basic problem

\[
\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad x \in \mathbb{R}^n
\end{align*}
\]

Definition (global minimizer)
The vector \( x^* \) is a global minimizer if

\[ f(x^*) \leq f(x) \quad \text{for all } x \in \mathbb{R}^n \]

Definition (local minimizer)
The vector \( x^* \) is a local minimizer if there exists \( \varepsilon > 0 \) such that

\[ f(x^*) \leq f(x) \quad \text{for all } x \in B(x^*, \varepsilon) := \{ x \in \mathbb{R}^n : \| x - x^* \| \leq \varepsilon \} \]

Definition (strict local minimizer)
The vector \( x^* \) is a strict local minimizer if there exists \( \varepsilon > 0 \) such that

\[ f(x^*) < f(x) \quad \text{for all } x \neq x^* \text{ such that } x \in B(x^*, \varepsilon) \]

Definition (isolated local minimizer)
The vector \( x^* \) is an isolated local minimizer if there exists \( \varepsilon > 0 \) such that \( x^* \) is the only local minimizer in \( B(x^*, \varepsilon) \)

- If \( x^* \) is an isolated local minimizer then \( x^* \) is a strict local minimizer

One-dimensional example

If we assume that \( f \) is continuously differentiable, then we can derive verifiable local optimality conditions for determining whether a point is a local minimizer.

- We rarely can verify that a point is a global minimizer

Theorem
If \( x^* \) is a local minimizer of a convex function \( f \) defined on \( \mathbb{R}^n \), then \( x^* \) is a global minimizer of \( f \).
We are interested in optimality conditions because they
- provide a means of guaranteeing when a candidate solution $x$ is indeed optimal (sufficient conditions)
- indicate when a point is not optimal (necessary conditions)
- guide in the design of algorithms since lack of optimality $\iff$ indication of improvement

Recall the following notation
- $g(x) = \nabla_x f(x)$
- $H(x) = \nabla^2_x f(x)$

**Theorem (First-order necessary condition)**

Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable. If $x^*$ is a local minimizer of $f$, then

$$g(x^*) = 0$$

**Proof:**

By definition of local minimizer, there exists $\varepsilon > 0$ such that $f(x^*) \leq f(x)$ for all $x \in B(x^*, \varepsilon)$. Suppose to the contrary that $g(x^*) \neq 0$. Set $s = -\frac{g(x^*)}{\|g(x^*)\|}$ and consider the function

$$\phi(\lambda) = f(x + \lambda s).$$

Then

$$\phi'(0) = g(x^*)^T s = -\|g(x^*)\| < 0.$$

Since $f$ is continuously differentiable, so is $\phi$, i.e., $\phi'$ is continuous. Thus, there exists $0 < \delta < \varepsilon$ such that $\phi'(c) < 0$ for all $c \in (-\delta, \delta)$. By the Mean Value Theorem,

$$\phi(\delta) = \phi(0) + \phi'(<\xi)(\delta - 0)$$

for some $\xi \in (0, \delta)$. Since $\phi'(\xi) < 0$, we have

$$f(x + \delta s) = \phi(\delta) < \phi(0) = f(x^*)$$

and $\|\delta s\| = \delta \|s\| = \delta < \varepsilon$, contradicting the hypothesis that $x^*$ is a local minimizer. $lacksquare$

In the proof of the previous theorem, the direction $-g(x^*) \neq 0$ was a descent direction.

**Definition (descent direction)**

We say that the direction $s$ is a descent direction for the continuously differentiable function $f$ at the point $x$ if

$$g(x)^T s < 0$$

**Note:** when the directional derivative of $f$ at $x$ in the direction $d$ exists, then it equals $g(x)^T s$, i.e.,

$$f'(x; s) \overset{def}{=} \lim_{t \to 0} \frac{f(x + ts) - f(x)}{t} = g(x)^T s$$

**Question:** Why do we call them descent directions?

**Answer:** We can use the Mean Value Theorem based argument in the proof of the first order necessary condition to show that there exists $\bar{\alpha} > 0$ such that

$$f(x + \alpha < f(x) \quad \text{for all} \quad 0 < \alpha < \bar{\alpha}.$$
\[ \phi(x) = f(x + 2s) \]

\[ \phi'(0) = \nabla f(x^*)^T s \]

\[ = -g(x^*) \frac{g(x^*)}{\|g(x^*)\|} \]

\[ = -\|g(x^*)\| < 0 \]

\[ \phi(s) = \phi(0) + \int_0^s \phi'(t) \, dt < 0 \]

\[ \phi(s) < \phi(0) \]
In the previous theorem, the quantity $s^TH(x)s$ was important.

**Definition (direction of positive curvature)**

We say that the direction $d$ for a twice-continuously differentiable function $f$ is a direction of positive curvature at the point $x$ if

$$d^TH(x)d > 0$$

**Definition (direction of negative curvature)**

We say that the direction $d$ for a twice-continuously differentiable function $f$ is a direction of negative curvature at the point $x$ if

$$d^TH(x)d < 0$$

**Definition (direction of zero curvature)**

We say that the direction $d$ for a twice-continuously differentiable function $f$ is a direction of zero curvature at the point $x$ if

$$d^TH(x)d = 0$$

Note: the quantity $d^TH(x)d$ provides second-order curvature information at the point $x$ along the direction $d$.
\[ f(x^*) + \nabla f(x^*)^T s + \frac{1}{2} s^T H(x^*) s \]