Optimality conditions for unconstrained optimization

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

September 3, 2019

Outline

1 The problem and definitions
2 First-order optimality conditions
3 Second-order optimality conditions
The basic problem

\[ \text{minimize } f(x) \quad x \in \mathbb{R}^n \]

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

\[ f(x^*) \leq f(x) \quad \text{for all } x \in B(x^*, \varepsilon) \]

where \( B(x^*, \varepsilon) := \{ x \in \mathbb{R}^n : \| x - x^* \| \leq \varepsilon \} \)

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

![Graph of a function with local and global minimizers]

- 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

\[
\text{lack of optimality} \iff \text{indication of improvement}
\]

Recall the following notation

- \( g(x) = \nabla f(x) \)
- \( H(x) = \nabla^2 f(x) \)

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

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

\[
g(x^*) = 0
\]

**Proof:**

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

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

Since \( f \) is continuously differentiable, so is \( \phi \), i.e., \( \phi' \) is continuous. Thus, there exists \( 0 < \delta < \epsilon \) such that \( \phi'(\xi) < 0 \) for all \( \xi \in (-\delta, \delta) \). 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 \( \|\delta s\| = \delta \|s\| = \delta < \epsilon \), contradicting the hypothesis that \( x^* \) is a local minimizer.
if \( g(x^*) \neq 0 \), then \( x^* \) is not a local minimizer
- we can limit our search to points \( x^* \) such that \( g(x^*) = 0 \)
- a stationary point is any point \( x \) that satisfies \( g(x) = 0 \)
- IMPORTANT: if \( g(x^*) = 0 \), it does not imply that we have found a local minimizer

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

\[
g(x) T \tilde{s} < 0
\]

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

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

**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

\[
f(x + \alpha s) < f(x) \quad \text{for all} \quad 0 < \alpha < \bar{\alpha}.
\]
Theorem (Second-order necessary conditions)

Suppose that \( f : \mathbb{R}^n \to \mathbb{R} \) is twice continuously differentiable. If \( x^* \) is a local minimizer of \( f \), then \( g(x^*) = 0 \) and \( H(x^*) \) is positive semi-definite, i.e.,

\[
s^T H(x^*) s \geq 0 \quad \text{for all } s \in \mathbb{R}^n
\]

Proof:

By definition of local minimizer, there exists \( \varepsilon > 0 \) such that \( f(x^*) \leq f(x) \) for all \( x \in B(x^*, \varepsilon) \). We know from the previous theorem that \( g(x^*) = 0 \). Suppose that \( s^T H(x^*) s < 0 \) for some \( s \in \mathbb{R}^n \) with \( ||s|| = 1 \). Consider the function \( \phi(\lambda) = f(x + \lambda s) \).

Then,

\[
\phi'(0) = g(x^*)^T s = 0, \quad \phi''(0) = s^T H(x^*) s < 0.
\]

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

\[
\phi(\delta) = \phi(0) + \phi'(0)(\delta - 0) + \phi''(\xi)(\delta - 0)^2
\]

for some \( \xi \in (0, \delta) \). Since \( \phi'(0) = 0 \) and \( \phi''(\xi) < 0 \), 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.

Note: these conditions are not sufficient for being a local minimizer

- \( f(x) = x^2 \implies x = 0 \) is a saddle point
- \( f(x) = -x^4 \implies x = 0 \) is a maximizer

Theorem (Second-order sufficient conditions)

If \( f : \mathbb{R}^n \to \mathbb{R} \) is twice continuously differentiable, the vector \( x^* \) satisfies \( g(x^*) = 0 \), and the matrix \( H(x^*) \) is positive definite, i.e.,

\[
s^T H(x^*) s > 0 \quad \text{for all } s \neq 0
\]

then \( x^* \) is a strict local minimizer of \( f \).

Proof:

Continuity implies that \( H(x) \) is positive definite for all \( x \) in an open ball \( B(x^*, \varepsilon) \) for some \( \varepsilon \). For any \( s \neq 0 \) satisfying

\[
x^* + s \in B(x^*, \varepsilon)
\]

we may use the fact that \( g(x^*) = 0 \) and the (higher dimensional) Mean Value Theorem to conclude that there exists \( z \) between \( x^* \) and \( x^* + s \) such that

\[
f(x^* + s) = f(x^*) + g(x^*)^T s + \frac{1}{2} s^T H(z) s
\]

\[
= f(x^*) + \frac{1}{2} s^T H(z) s
\]

\[
> f(x^*),
\]

which implies that \( x^* \) is a strict local minimizer.
In the previous theorem, the quantity $s^2H(x)s$ was important.

<table>
<thead>
<tr>
<th>Definition (direction of positive curvature)</th>
</tr>
</thead>
<tbody>
<tr>
<td>We say that the direction $d$ for a twice-continuously differentiable function $f$ is a <strong>direction of positive curvature</strong> at the point $x$ if $d^2H(x)d &gt; 0$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Definition (direction of negative curvature)</th>
</tr>
</thead>
<tbody>
<tr>
<td>We say that the direction $d$ for a twice-continuously differentiable function $f$ is a <strong>direction of negative curvature</strong> at the point $x$ if $d^2H(x)d &lt; 0$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Definition (direction of zero curvature)</th>
</tr>
</thead>
<tbody>
<tr>
<td>We say that the direction $d$ for a twice-continuously differentiable function $f$ is a <strong>direction of zero curvature</strong> at the point $x$ if $d^2H(x)d = 0$</td>
</tr>
</tbody>
</table>

*Note:* the quantity $d^2H(x)d$ provides second-order curvature information at the point $x$ along the direction $d$