Solution to Problem 3 on Assignment 5

Ken Jackson

20 April 2014

In Problem 3 on Assignment 5, you are asked to prove the following result.

**Theorem 1** If

1. \( f : \mathbb{R}^n \to \mathbb{R} \) is twice continuously differentiable (i.e., \( \nabla^2 f(x) \) exists and is continuous for all \( x \in \mathbb{R}^n \)),

2. the level set \( \mathcal{L} = \{ x \in \mathbb{R}^n : f(x) \leq f(x_0) \} \) is convex and there exist positive constants \( m \) and \( M \) such that
   \[
   m \| z \|^2 \leq z^T \nabla^2 f(x) z \leq M \| z \|^2
   \]
   for all \( z \in \mathbb{R}^n \) and all \( x \in \mathcal{L} \),

3. the sequence of points \( x_0, x_1, x_2, \ldots \) generated by the minimization algorithm satisfies \( f(x_{k+1}) \leq f(x_k) \) for all \( k = 0, 1, 2, \ldots \),

4. the sequence of points \( x_0, x_1, x_2, \ldots \) generated by the minimization algorithm also satisfies
   \[
   \liminf_{k \to \infty} \| \nabla f(x_k) \| = 0
   \]

then \( f \) has a unique strict minimizer \( x^* \in \mathcal{L} \) and

\[
\lim_{k \to \infty} x_k = x^*
\]

By \( f \) having a unique strict minimizer \( x^* \in \mathcal{L} \), we mean that there exists a point \( x^* \in \mathcal{L} \) that satisfies \( f(x^*) < f(x) \) for all \( x \in \mathcal{L} \) for which \( x \neq x^* \).

Of course, there cannot be an \( \hat{x} \notin \mathcal{L} \) such that \( f(\hat{x}) \leq f(x^*) \), since \( \hat{x} \notin \mathcal{L} \) implies that \( f(x_0) < f(\hat{x}) \) and \( x^* \in \mathcal{L} \) implies that \( f(x^*) \leq f(x_0) \), whence \( f(x^*) \leq f(x_0) < f(\hat{x}) \). So, if \( x^* \) is the unique strict minimizer of \( f \) in \( \mathcal{L} \), then \( x^* \) is also the unique strict minimizer of \( f \) in all of \( \mathbb{R}^n \).

There are many ways to prove Theorem 1 above. Here’s one. If you notice any errors, let me know.
Note that the proof below does not use the inequality \( z^T \nabla^2 f(x) z \leq M \|z\|_2^2 \) in Assumption 3, but it does rely heavily on the inequality \( m \|z\|_2^2 \leq z^T \nabla^2 f(x) z \). So, I believe we can weaken Assumption 3 to

the level set \( \mathcal{L} = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\} \) is convex and there exists a positive constant \( m \) such that

\[
m \|z\|_2^2 \leq z^T \nabla^2 f(x) z
\]

for all \( z \in \mathbb{R}^n \) and all \( x \in \mathcal{L} \).

If you can see why we need the stronger Assumption 3, let me know.

**Proof:** To begin, note that Assumption 1 ensures that \( f : \mathbb{R}^n \to \mathbb{R} \) is twice continuously differentiable. That is, \( \nabla^2 f(x) \) exists and is continuous for all \( x \in \mathbb{R}^n \). This also implies that \( \nabla f(x) \) exists and is continuous for all \( x \in \mathbb{R}^n \) and \( f(x) \) is continuous for all \( x \in \mathbb{R}^n \). We use these properties without further comment throughout the proof.

We first show that \( f \) has a unique minimizer \( x^* \in \mathcal{L} \). That is, there is an \( x^* \in \mathcal{L} \) such that \( f(x^*) < f(x) \) for all \( x \in \mathcal{L} \) for which \( x \neq x^* \).

To this end, we begin by showing that \( \mathcal{L} \) is compact (i.e., closed and bounded). To show that \( \mathcal{L} \) is closed, it is sufficient to show that, if \( y_n \to y \) as \( n \to \infty \) and all \( y_n \in \mathcal{L} \), for \( n = 0, 1, 2, \ldots \), then \( y \in \mathcal{L} \). We prove this by first noting that, \( y_n \in \mathcal{L} \) implies that \( f(y_n) \leq f(x_0) \), since

\[
\mathcal{L} = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}
\]

Next note that \( f(y_n) \leq f(x_0) \), for \( n = 0, 1, 2, \ldots \), and \( y_n \to y \) as \( n \to \infty \) implies that \( f(y) \leq f(x_0) \), since \( f \) is continuous. Therefore, \( y \in \mathcal{L} \). Hence, \( \mathcal{L} \) is closed.

We use proof by contradiction to show that \( \mathcal{L} \) is bounded. To this end, suppose that \( \mathcal{L} \) is not bounded. Thus, there is an \( \hat{x} \in \mathcal{L} \) such that \( \|\hat{x} - x_0\| > \frac{2}{m} \|\nabla f(x_0)\| \), where \( m \) is the positive constant from Assumption 2. Since \( f \) is twice continuously differentiable, we get from Taylor’s Theorem (equation (2.6) on page 14 of your textbook) that

\[
f(\hat{x}) = f(x_0) + (\hat{x} - x_0)^T \nabla f(x_0) + \frac{1}{2} (\hat{x} - x_0)^T \nabla^2 f(x_0 + t(\hat{x} - x_0)) (\hat{x} - x_0)
\]

for some \( t \in [0, 1] \). Since \( \hat{x} \in \mathcal{L} \), \( x_0 \in \mathcal{L} \) and \( \mathcal{L} \) is convex, \( x_0 + t(\hat{x} - x_0) = (1 - t)x_0 + t\hat{x} \in \mathcal{L} \). Therefore, by Assumption 2,

\[
\frac{1}{2} (\hat{x} - x_0)^T \nabla^2 f(x_0 + t(\hat{x} - x_0)) (\hat{x} - x_0) \geq \frac{m}{2} \|\hat{x} - x_0\|^2
\]

Also, by the Cauchy-Schwartz inequality (inequality (A.5) on page 600 of your textbook),

\[
|(\hat{x} - x_0)^T \nabla f(x_0)| \leq \|\hat{x} - x_0\| \|\nabla f(x_0)\|
\]

Therefore,

\[
(\hat{x} - x_0)^T \nabla f(x_0) \geq -\|\hat{x} - x_0\| \|\nabla f(x_0)\|
\]

2
Combining (1), (2), (3) and the assumption that \( \|\hat{x} - x_0\| > \frac{2}{m}\|\nabla f(x_0)\| \), which implies that
\[
-\frac{2}{m}\|\nabla f(x_0)\| + \|\hat{x} - x_0\| > 0,
\]
we get
\[
f(\hat{x}) - f(x_0) \geq -\|\hat{x} - x_0\|\|\nabla f(x_0)\| + \frac{m}{2}\|\hat{x} - x_0\|^2
\]
\[
= \frac{m}{2}\|\hat{x} - x_0\| \left(-\frac{2}{m}\|\nabla f(x_0)\| + \|\hat{x} - x_0\|\right)
\]
\[
> 0
\]
whence \( f(\hat{x}) > f(x_0) \). However, this contradicts, \( \hat{x} \in \mathcal{L} \). Therefore, \( \mathcal{L} \) must be bounded.

Since \( \mathcal{L} \) is closed and bounded, it is compact. Moreover, \( \mathcal{L} \) is not empty, since \( x_0 \in \mathcal{L} \). A key theorem in analysis tells us that any continuous function obtains its minimum on a compact set. Therefore, there is an \( x^* \in \mathcal{L} \) such that \( f(x^*) \leq f(x) \) for all \( x \in \mathcal{L} \).

To prove that \( x^* \) is the unique minimizer of \( f(x) \) in \( \mathcal{L} \), we first show that \( \nabla f(x^*) = 0 \). We have to be a little careful about this, since \( x^* \) could be on the boundary of \( \mathcal{L} \). Thus, there may not be an open neighbourhood of \( x^* \) contained in \( \mathcal{L} \). Nevertheless, we can prove \( \nabla f(x^*) = 0 \) by contradiction.

To this end, suppose that \( \nabla f(x^*) \neq 0 \). Let \( p = -\alpha \nabla f(x^*) \) for some \( \alpha > 0 \) to be determined below and let \( x^+ = x^* + p \). By Taylor’s Theorem (equation (2.4) on page 14 of your textbook),
\[
f(x^+) = f(x^* + p) = f(x^*) + p^T \nabla f(x^* + tp) = f(x^*) - \alpha \nabla f(x^*)^T \nabla f(x^* - t\alpha \nabla f(x^*))
\]
for some \( t \in [0, 1] \). Now let
\[
\phi(s) = \nabla f(x^*)^T \nabla f(x^* - s \nabla f(x^*))
\]
Note that \( \nabla f(x^*) \neq 0 \) implies that \( \phi(0) > 0 \). Moreover, \( \phi(s) \) is continuous, since \( \nabla f(x) \) is continuous. Therefore, there is a \( \hat{s} > 0 \) such that \( \phi(s) > 0 \) for all \( s \in [0, \hat{s}] \). Moreover, if we choose \( \alpha \in (0, \hat{s}] \), then \( t\alpha \in [0, \hat{s}] \) for all \( t \in [0, 1] \). Hence,
\[
\nabla f(x^*)^T \nabla f(x^* - t\alpha \nabla f(x^*)) > 0
\]
for all \( t \in [0, 1] \). Also, recall \( \alpha > 0 \). Therefore,
\[
f(x^+) = f(x^* + p) = f(x^*) - \alpha \nabla f(x^*)^T \nabla f(x^* - t\alpha \nabla f(x^*)) < f(x^*)
\]
Moreover, \( x^* \in \mathcal{L} \) implies \( f(x^*) \leq f(x_0) \), whence \( f(x^+) \leq f(x_0) \) too. Thus, \( x^+ \in \mathcal{L} \), since \( \mathcal{L} = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\} \). However, \( f(x^+) \) < \( f(x^*) \) and \( x^+ \in \mathcal{L} \) contradicts our claim that \( x^* \) is a minimizer for \( f(x) \) in \( \mathcal{L} \). Therefore, the assumption that \( \nabla f(x^*) \neq 0 \) must be false. Thus, we have proven that \( \nabla f(x^*) = 0 \).

We want to show that \( x^* \) is the unique strict minimizer of \( f(x) \) in \( \mathcal{L} \). That is, \( f(x^*) < f(x) \) for all \( x \in \mathcal{L} \) for which \( x \neq x^* \). To this end, choose any \( x \in \mathcal{L} \) for which \( x \neq x^* \). Since \( f \) is twice continuously differentiable (by Assumption 1), we have from Taylor’s Theorem (equation (2.6) on page 14) that
\[
f(x) = f(x^*) + \nabla f(x^*)^T (x - x^*) + \frac{1}{2} (x - x^*)^T \nabla^2 f(x^* + t(x - x^*)) (x - x^*) \quad (4)
\]
for some $t \in [0, 1]$. Since $x^* \in \mathcal{L}$, $x \in \mathcal{L}$ and $\mathcal{L}$ is convex, $x^* + t(x - x^*) = (1 - t)x^* + tx \in \mathcal{L}$. Therefore, it follows from Assumption 2 that
\[(x - x^*)^T \nabla^2 f(x^* + t(x - x^*)) (x - x^*) \geq m\|x - x^*\|^2\]
for some $m > 0$. Moreover, we showed above that $\nabla f(x^*) = 0$. Therefore, (4) reduces to
\[f(x) - f(x^*) \geq \frac{m}{2}\|x - x^*\|^2 > 0\]
since $m > 0$ and $x \neq x^*$. Therefore, $f(x^*) < f(x)$, establishing the result that $x^*$ is the unique strict minimizer of $f(x)$ in $\mathcal{L}$.

Since, by Assumption 4,
\[\lim_{k \to \infty} \|\nabla f(x_k)\| = 0\]
there is a subsequence $\{x_{k_j}\}$ of the full sequence $\{x_k\}$, such that
\[\lim_{k_j \to \infty} \|\nabla f(x_{k_j})\| = 0\]
We show next that $x_{k_j} \to x^*$ as $k_j \to \infty$.

To this end, first note that, by Taylor’s Theorem (equation (2.5) on page 14 of your textbook)
\[\nabla f(x_{k_j}) = \nabla f(x^*) + \int_0^1 \nabla^2 f(x^* + t(x_{k_j} - x^*)) (x_{k_j} - x^*) \, dt\]
(5)
Using the fact that $\nabla f(x^*) = 0$ established above and multiplying (5) by $(x_{k_j} - x^*)^T$, we get
\[(x_{k_j} - x^*)^T \nabla f(x_{k_j}) = \int_0^1 (x_{k_j} - x^*)^T \nabla^2 f(x^* + t(x_{k_j} - x^*)) (x_{k_j} - x^*) \, dt\]
(6)
Now note that the sequence $\{x_k\}$ satisfies Assumption 3, whence $f(x_k) \leq f(x_0)$ for all $k \geq 0$. Moreover, since $\{x_{k_j}\} \subset \{x_k\}$, $f(x_{k_j}) \leq f(x_0)$ too for all $k_j$, whence all $x_{k_j} \in \mathcal{L}$. Furthermore, $x^* \in \mathcal{L}$ and $\mathcal{L}$ is convex by Assumption 2. Therefore, $x^* + t(x_{k_j} - x^*) = (1 - t)x^* + tx_{k_j} \in \mathcal{L}$ for all $t \in [0, 1]$. Hence, by Assumption 2, there is an $m > 0$ such that
\[m\|x_{k_j} - x^*\|^2 \leq (x_{k_j} - x^*)^T \nabla^2 f(x^* + t(x_{k_j} - x^*)) (x_{k_j} - x^*)\]
(7)
for all $t \in [0, 1]$. Combining (6) and (7), we get
\[m\|x_{k_j} - x^*\|^2 \leq (x_{k_j} - x^*)^T \nabla f(x_{k_j})\]
(8)
By the Cauchy-Schwartz inequality (inequality (A.5) on page 600 of your textbook),
\[(x_{k_j} - x^*)^T \nabla f(x_{k_j}) \leq \|x_{k_j} - x^*\| \|\nabla f(x_{k_j})\|\]
(9)
Combining (8) and (9), we get
\[m\|x_{k_j} - x^*\|^2 \leq \|x_{k_j} - x^*\| \|\nabla f(x_{k_j})\|\]
which implies
\[ \|x_{k_j} - x^*\| \leq \frac{1}{m} \|\nabla f(x_{k_j})\| \] (10)

Since
\[ \lim_{k_j \to \infty} \nabla f(x_{k_j}) = 0 \]

(10) implies that \( x_{k_j} \to x^* \) as \( k_j \to \infty \). That is, we have shown that the subsequence \( \{x_{k_j}\} \) converges to \( x^* \).

To see that the whole sequence \( \{x_k\} \) converges to \( x^* \), first note that, from Assumption 3,
\[ f(x_k) \leq f(x_{k_j}) \]
for all \( k \geq k_j \). Moreover, since \( x^* \) is the minimum of \( f(x) \), \( f(x^*) \leq f(x_k) \). Therefore,
\[ 0 \leq f(x_k) - f(x^*) \leq f(x_{k_j}) - f(x^*) \] (11)

However, we showed above that \( x_{k_j} \to x^* \) as \( k_j \to \infty \). Therefore, by the continuity of \( f(x) \), \( f(x_{k_j}) - f(x^*) \to 0 \) as \( k_j \to \infty \). Hence, by (11), \( f(x_k) - f(x^*) \to 0 \) as \( k \to \infty \).

By Taylor’s Theorem (equation (2.6) on page 14 of your textbook),
\[ f(x_k) = f(x^*) + (x_k - x^*)^T \nabla f(x^*) + \frac{1}{2} (x_k - x^*)^T \nabla^2 f(x^* + t(x_k - x^*)) (x_k - x^*) \] (12)

for some \( t \in [0, 1] \).

As noted above, the sequence \( \{x_k\} \) satisfies Assumption 3, whence \( f(x_k) \leq f(x_0) \) for all \( k \geq 0 \). Therefore, \( x_k \in \mathcal{L} \) for all \( k \geq 0 \). Moreover, \( x^* \in \mathcal{L} \) and \( \mathcal{L} \) is convex (by Assumption 2). Therefore, \( x^* + t(x_k - x^*) = (1 - t)x^* + tx_k \in \mathcal{L} \) for all \( t \in [0, 1] \). Hence, by Assumption 2, there is an \( m > 0 \) such that
\[ m\|x_k - x^*\|^2 \leq (x_k - x^*)^T \nabla^2 f(x^* + t(x_k - x^*)) (x_k - x^*) \] (13)

for all \( t \in [0, 1] \). Therefore, combining (12) and (13) and using the fact that \( \nabla f(x^*) = 0 \), we get
\[ \frac{1}{2} m\|x_k - x^*\|^2 \leq f(x_k) - f(x^*) \]
which implies
\[ \|x_k - x^*\| \leq \sqrt{\frac{2(f(x_k) - f(x^*))}{m}} \] (14)

Recall that we showed above that \( f(x_k) - f(x^*) \to 0 \) as \( k \to \infty \). This together with (14) implies that \( \|x_k - x^*\| \to 0 \) as \( k \to \infty \). That is, the whole sequence \( \{x_k\} \) converges to \( x^* \).