This is a summary of Theorem 11.7 from Griva, Nash, and Sofer.

**ASSUMPTIONS**

\[ f : \mathbb{R}^n \rightarrow \mathbb{R} \]

\( \mathbf{x}_0 \) is given

\( \mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{p}_k \) is the iteration

each \( \alpha_k > 0 \) is chosen by backtracking line search for a sufficient decrease condition, i.e.

\[ f(\mathbf{x}_k + \alpha_k \mathbf{p}_k) \leq f(\mathbf{x}_k) + \mu \alpha_k \mathbf{p}_k^T \mathbf{g}(\mathbf{x}_k), \quad \mu < 1 \]

and \( \alpha_k \) is the first element of the sequence \( 1, \frac{1}{2}, \frac{1}{4}, \ldots \) to satisfy this bound

the set \( S = \{ x : f(x) \leq f(x_0) \} \) is bounded

\( \mathbf{g}(\mathbf{x}) \) is Lipschitz continuous, i.e.

\[ \| \mathbf{g}(\mathbf{y}) - \mathbf{g}(\mathbf{x}) \| \leq L \| \mathbf{y} - \mathbf{x} \| \quad L < \infty \]

the search directions \( \mathbf{p}_k \) satisfy sufficient descent, i.e.

\[ -\frac{\mathbf{p}_k^T \mathbf{g}(\mathbf{x})}{\| \mathbf{p}_k \| \| \mathbf{g}(\mathbf{x}_k) \|} \geq \varepsilon > 0 \]

the search directions are gradient related and bounded, i.e.

\[ \| \mathbf{p}_k \| \geq m \| \mathbf{g}(\mathbf{x}_k) \| \quad \text{and} \quad \| \mathbf{p}_k \| \leq M \]

each scalar \( m, M, \mu \) is fixed.

**CONCLUSION**

\[ \lim_{k \to \infty} \| \mathbf{g}(\mathbf{x}_k) \| = 0 \]

**PROOF**

There are five steps to the proof.

1. Show that \( f \) is bounded below. (i.e. Won't run forever ... )
2. Show that \( \lim_{k \to \infty} f(\mathbf{x}_k) \) exists (i.e. we converge in one sense ... )
3. Show that \( \lim_{k \to \infty} \alpha_k \| \mathbf{g}(\mathbf{x}_k) \|^2 = 0 \)
4. Show that \( \alpha_k < 1 \) implies \( \alpha_k \geq \gamma \| \mathbf{g}(\mathbf{x}_k) \|^2 \) (i.e. that small steps aren't too small...)
5. Finally, we conclude \( \lim_{k \to \infty} \| \mathbf{g}(\mathbf{x}_k) \| = 0 \)
**Step 1** We know that $f$ is continuous, so the set $S$ is closed. Because we assume that $S$ is bounded, then a closed bounded set must take on a minimum somewhere. Hence,

$$f(x) \geq C.$$

**Step 2** At each step $f(x_{k+1}) < f(x_k)$ and we have that $f$ is bounded from below, so $\lim_{k \to \infty} f(x_k)$ must converge (but may not be a minimizer.) Let $\bar{f}$ be the limit.

**Step 3** Things get a little tricker here. Note that

$$f(x_0) - \bar{f} = \sum_{k=0}^{\infty} [f(x_k) - f(x_{k+1})]$$

by a telescoping series.

Let’s use our conditions.

By the line search, $f(x_k) - f(x_{k+1}) \geq -\mu \alpha_k p_k^T g(x_k)$.

By sufficient descent, $p_k^T g(x_k) \geq -\epsilon \|p_k\| \|g(x_k)\|$.

By gradient relatedness, $\|p_k\| \geq m \|g_k\|$.

Thus

$$f(x_0) - \bar{f} \geq \sum_{k=0}^{\infty} \mu \alpha_k \epsilon m \|g(x_k)\|^2.$$

Because $f(x_0) - \bar{f} \leq f(x_0) - C < \infty$, this sum must converge, and thus

$$\lim_{k \to \infty} \alpha_k \|g(x_k)\|^2 = 0$$

(All the other terms in the limit were positive constants.)

**Step 4** At this point, we haven’t used the “backtracking” piece of the line-search algorithm yet. So we’ll see that here to show that small steps aren’t too small.

If $\alpha_k < 1$, then we know that $2\alpha_k$ violated sufficient decrease:

$$f(x_k + 2\alpha_k p_k) - f(x_k) > 2\mu \alpha_k p_k^T g(x).$$

By a theorem about Lipschitz functions,

$$f(x_k + 2\alpha_k p_k) - f(x_k) - 2\alpha_k p_k^T g(x) \leq \frac{1}{2} |L| 2\alpha_k p_k^T g(x).$$

By rearrangement, we have:

$$f(x_k) - f(x_k + 2\alpha_k p_k) \geq -2\alpha_k p_k^T g(x) - 2L \|\alpha_k p_k\|^2.$$

If we add this to our starting inequality:

$$f(x_k + 2\alpha_k p_k) - f(x_k) > 2\mu \alpha_k p_k^T g(x)$$

then the left hand side cancels and

$$0 \geq -2\alpha_k p_k^T g(x) - 2L \|\alpha_k p_k\|^2 + 2\mu \alpha_k p_k^T g(x)$$

or

$$\alpha_k L \|p_k\|^2 \geq -(1 - \mu) p_k^T g(k).$$

Sufficient descent and gradient relatedness let us play the same tricks with $p_k^T g(k)$, so we have

$$\alpha_k L \|p_k\|^2 \geq (1 - \mu) \epsilon m \|g(x_k)\|^2.$$

Consequently,

$$\alpha_k \geq \frac{\|g(x_k)\|^2}{\gamma} = \frac{(1 - \mu) \epsilon m}{M^2 L} > 0.$$

**Step 5** Because $\lim_{k \to \infty} \alpha_k \|g(x_k)\|^2 = 0$ and $\alpha_k$ doesn’t get too small, i.e.

$$\alpha_k \geq \min(1, \gamma \|g(x_k)\|^2).$$

then the norm must go to zero for this limit to exist.