CONVERGENCE OF BACKTRACKING LINE SEARCH

David F. Gleich
February 25, 2025

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

ASSUMPTIONS

$f : \mathbb{R}^n \mapsto \mathbb{R}$

$\mathbf{x}_0$ 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.

$$f(\mathbf{x}_k + \alpha_k \mathbf{p}_k) \le f(\mathbf{x}_k) + \mu \alpha_k \mathbf{p}_k^T \mathbf{g}(\mathbf{x}_k) \qquad \mu < 1$$

and α_k is the first element of the sequence $1, 1/2, 1/4, \ldots$ to satisfy this bound

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

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

$$\|\mathbf{g}(\mathbf{y}) - \mathbf{g}(\mathbf{x})\| \le L\|\mathbf{y} - \mathbf{x}\| \qquad L < \infty$$

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

$$-\frac{\mathbf{p}_k^T \mathbf{g}(\mathbf{x})}{\|\mathbf{p}_k\| \|\mathbf{g}(\mathbf{x}_k)\|} \ge \varepsilon > 0$$

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

$$\|\mathbf{p}_k\| \ge m\|\mathbf{g}(\mathbf{x}_k)\| \text{ and } \|\mathbf{p}_k\| \le M$$

each scalar m, M, μ 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 $\alpha_k \ge \gamma \|\mathbf{g}(\mathbf{x}_k)\|^2$ (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,

$$f(\mathbf{x}) \geq C.$$

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

Step 3 Things get a little tricker here. Note that

$$f(\mathbf{x}_0) - \bar{f} = \sum_{k=0}^{\infty} [f(\mathbf{x}_k) - f(\mathbf{x}_{k+1})]$$

by a telescoping series.

Let's use our conditions.

By the line search, $f(\mathbf{x}_k) - f(\mathbf{x}_{k+1}) \geq -\mu \alpha_k \mathbf{p}_k^T \mathbf{g}(\mathbf{x}_k)$.
By sufficient descent, $\mathbf{p}_k^T \mathbf{g}(\mathbf{x}_k) \geq -\varepsilon \|\mathbf{p}_k\| \|\mathbf{g}(\mathbf{x}_k)\|$.
By gradient relatedness, $\|\mathbf{p}_k\| \geq m \|\mathbf{g}_k\|$.
Thus

$$f(\mathbf{x}_0) - \bar{f} \geq \sum_{k=0}^{\infty} \mu \alpha_k \varepsilon m \|\mathbf{g}(\mathbf{x}_k)\|^2.$$

Because $f(\mathbf{x}_0) - \bar{f} \leq f(\mathbf{x}_0) - C < \infty$, this sum must converge, and thus

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

(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:

$$f(\mathbf{x}_k + 2\alpha_k \mathbf{p}_k) - f(\mathbf{x}_k) > 2\mu \alpha_k \mathbf{p}_k^T \mathbf{g}(\mathbf{x}).$$

By a theorem about Lipschitz functions,

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

Here's the result, if f has Lipschitz gradients, with constant L, then $|f(\mathbf{y}) - f(\mathbf{x}) - \mathbf{g}(x)^T(\mathbf{y} - \mathbf{x})| \leq 1/2L\|\mathbf{y} - \mathbf{x}\|^2$. You can prove this using a Taylor series and Cauchy Schwartz without the extra factor of 1/2, but to get that factor, you need to do a path-integral of the gradient from $x \to y$.

By rearrangement, we have:

$$f(\mathbf{x}_k) - f(\mathbf{x}_k + 2\alpha_k \mathbf{p}_k) \geq -2\alpha_k \mathbf{p}_k^T \mathbf{g}(\mathbf{x}) - 2L \|\alpha_k \mathbf{p}_k\|^2.$$

If we add this to our starting inequality:

$$f(\mathbf{x}_k + 2\alpha_k \mathbf{p}_k) - f(\mathbf{x}_k) > 2\mu \alpha_k \mathbf{p}_k^T \mathbf{g}(\mathbf{x})$$

then the left hand side cancels and

$$0 \geq -2\alpha_k \mathbf{p}_k^T \mathbf{g}(\mathbf{x}) - 2L \|\alpha_k \mathbf{p}_k\|^2 + 2\mu \alpha_k \mathbf{p}_k^T \mathbf{g}(\mathbf{x})$$

or

$$\alpha_k L \|\mathbf{p}_k\|^2 \geq -(1-\mu) \mathbf{p}_k^T \mathbf{g}(\mathbf{k}).$$

Sufficient descent and gradient relatedness let us play the same tricks with $\mathbf{p}_k^T \mathbf{g}(\mathbf{k})$, so we have

$$\alpha_k L \|\mathbf{p}_k\|^2 \geq (1-\mu) \varepsilon m \|\mathbf{g}(\mathbf{x}_k)\|^2.$$

Consequently,

$$\alpha_k \geq \gamma \|\mathbf{g}(\mathbf{x}_k)\|^2 \qquad \gamma = \frac{(1-\mu)\varepsilon m}{M^2 L} > 0.$$

Step 5 Because $\lim_{k\to\infty} \alpha_k \|\mathbf{g}(\mathbf{x}_k)\|^2 = 0$ and α_k doesn't get too small, i.e.

$$\alpha_k \geq \min(1, \gamma \|\mathbf{g}(\mathbf{x}_k)\|^2).$$

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