

LINE SEARCH AND THE WOLFE CONDITIONS

David F. Gleich

February 25, 2025

Consider a line search method on the unconstrained optimization problem:

$$\text{minimize } f(\mathbf{x})$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is twice continuously differentiable.

At a point $\mathbf{x}^{(k)}$, a line search method proceeds by choosing

a search direction \mathbf{p} and a step length α

such that

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha \mathbf{p}.$$

In this note, we'll work through what we need to guarantee in a line search in order to ensure convergence.

If \mathbf{p} is a descent direction, then we showed that we can always choose α sufficiently small such that $f_{k+1} \leq f_k$. So one thing we want to guarantee in a line search is that:

$$\mathbf{p} \text{ is a descent direction, i.e. } \mathbf{p}^T \mathbf{g} < 0$$

and

$$f_{k+1} < f_k$$

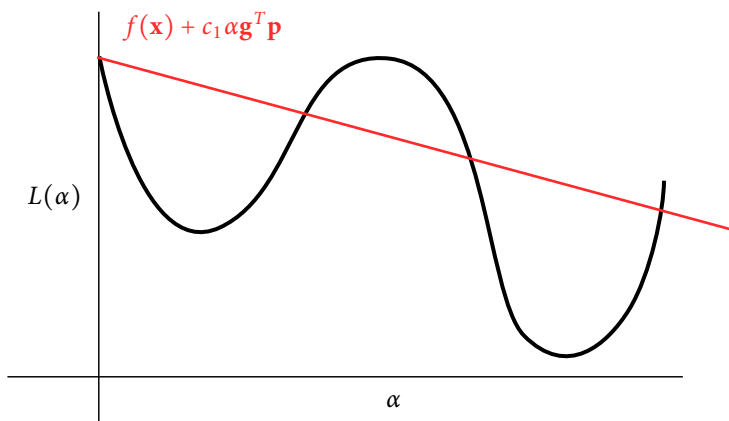
but we found that this wasn't good enough via an example.

SUFFICIENT DECREASE

The problem we encountered there was that you might not have to reduce the function by enough at each iteration. To combat this problem, we want to ensure you reduce the function by some reasonable amount every step. This idea is formalized as follows:

you should decrease the function by at least as much as a linear function.

In this case, a picture makes the statement clear:



Consider the function $L(\alpha) = f(\mathbf{x} + \alpha \mathbf{p}^T \mathbf{g})$ that describes the line search. Our first order Taylor approximation says that the behavior of $L(\alpha)$ should be described by the line:

$$L(\alpha) \approx f(\mathbf{x}) + \mathbf{p}^T \mathbf{g} \alpha.$$

We want to ensure that we take a step α , then we descend at least as much as some "shallower" version of this line. So we use the upper-bound $f(\mathbf{x}) + c_1 \alpha \mathbf{p}^T \mathbf{g}$, for some

constant $0 < c_1 < 1$.¹ Thus, we will reject any step that lies above this line! This ensures that we are taking a step reasonably well correlated with the Taylor approximations. For instance, this criteria ensures that if you take a big step, you get a big reward! So this breaks our example where we kept oscillating between the points ± 1 .

¹ Think c_1 really small, around 10^{-4} . It's just really hard to draw a picture of this choice as a line!

The condition

$$L(\alpha) \leq f(\mathbf{x}) + c_1 \alpha \mathbf{p}^T \mathbf{g}$$

is called *sufficient decrease* or the Armijo condition. Note that we'll satisfy $f_{k+1} < f_k$ with any such step.

CURVATURE CONDITION

However, this example doesn't prohibit another problem: we can always take a tiny step and still lie below the line $f(\mathbf{x}) + c_1 \alpha \mathbf{p}^T \mathbf{g}$. Thus, for the quadratic we could take $\alpha_k = 1/2^k$ and never make any headway to our minimizer.

The real problem here is that we don't want to take a step α_k if we expect that increasing α_k will decrease the function further! Put another way, we want to ensure we decrease the function as much as we can reasonably expect.

THE WOLFE CONDITIONS

These two conditions define the Wolfe conditions:

DEFINITION 1 (Wolfe conditions) *Let $f(\mathbf{x})$ be continuously differentiable. We say that a line search step α_k satisfies the Wolfe conditions if:*

$$\begin{aligned} f(\mathbf{x}_k + \alpha \mathbf{p}_k) &\leq f(\mathbf{x}_k) + \alpha c_1 \mathbf{p}_k^T \mathbf{g}_k \\ \mathbf{g}(\mathbf{x}_k + \alpha \mathbf{p}_k)^T \mathbf{p}_k &\geq c_2 \mathbf{g}_k^T \mathbf{p}_k \end{aligned}$$

with $c_1 \leq c_2 \leq 1$.

THEOREM 2 *Let f be continuously differentiable. Let \mathbf{p}_k be a descent direction and let $L(\alpha) = f(\mathbf{x}_k + \alpha \mathbf{p}_k)$. Suppose that $L(\alpha)$ is bounded from below. Then there exists α_k that satisfies the strong Wolfe conditions.*

Proof This is a perfect proof by picture case. ■

