Line Search Methods Analysis

Saurav Samantaray

Department of Mathematics

Indian Institute of Technology Madras

January 28, 2024

Step Length

- In computing the step length we face a trade-off.
- We want to choose α_k to give a substantial reduction of f, but we don't want to spend too mush time making the choice.
- Offcourse the ideal choice would be the global minimiser of the univariate function $\phi(.)$ defined by

$$\phi(\alpha) = f(x_k + \alpha p_k), \;\; \alpha > 0.$$

- But in general, it is too expensive to identify this value.
- It requires too many evaluations of the objective function and/or the gradient to even find a local minimiser to moderate precision.

Step Length

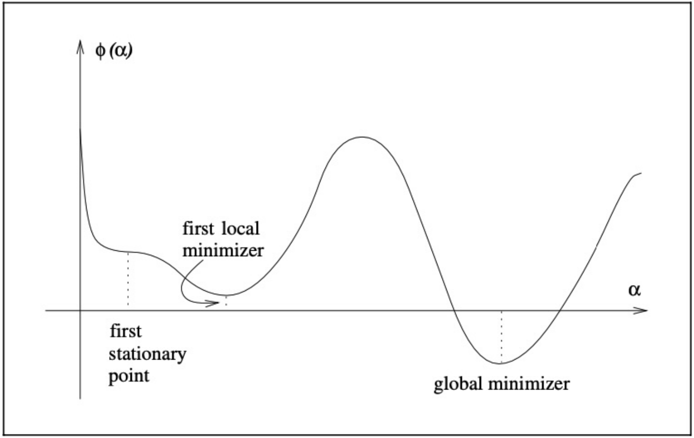

Figure: The ideal step length is the global minimiser

Step Length

- Practically, strategies perform an inexact line search to identify a step length that achieves adequate reductions in f at minimal cost.
- We will discuss these search strategies a little later.
- we will now discuss various termination conditions for line search algorithms and show that effective step lengths need not lie near minimisers of the univariate function $\phi(\alpha)$.
- Is $f(x_k + \alpha_k p_k) < f(x_k)$ good enough to get convergence??
- for example consider the function

$$f(x) = x^2 - 1$$

it has the global minima at $x = 0$, $f = -1$.

Step Length

- Consider a sequence $\{x_k\}$ s.t.

$$f(x_k) = \frac{5}{k}, \quad k = 1, 2, 3, \ldots$$
$$\implies f(x_k) > f(x_{k+1})$$

- The reduction in f at each step is not enough to get it to converge to the minimiser.

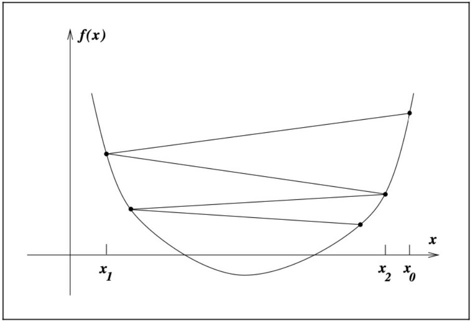

Figure: Insufficient reduction

The Wolfe Condition

Armijo Condition (Sufficient Decrease Condition):

α_k should be chosen such that

$$f(x_k + \alpha p_k) \leq f(x_k) + c_1 \alpha \nabla f_k^T p_k \tag{1}$$

for some constant $c_1 \in (0, 1)$.

- Since p_k is a descent direction and $c_1 > 0$ and $\alpha > 0$ the first thing that the Armijo condition asserts that there is a reduction in f from x_k to $x_{k+1} = x_k + \alpha p_k$.
- The reduction in f is atleast

  $$c_1 \alpha \nabla f_k^T p_k$$

  therefore it also says the reduction in f must be proportional to both the step length α_k and the directional derivative $\nabla f_k^T p_k$

The Wolfe Condition

- The right hand side of (1) is a linear function in α (say) $l(\alpha)$.

$$l(\alpha) = f(x_\alpha) + c_1 \alpha \nabla f_k^T p_k$$

- The function $l(.)$ has a negative slope $c_1 \nabla f_k^T p_k$ but $c_1 \in (0, 1)$.
- therefore it lies above the graph of ϕ for small positive values of α.
- The sufficient decrease condition states that α is acceptable only if
$$\phi(\alpha) \leq l(\alpha).$$
- In practice, c_1 is chosen to be quite small, say

$$c_1 = 10^{-4}$$

The Wolfe Condition

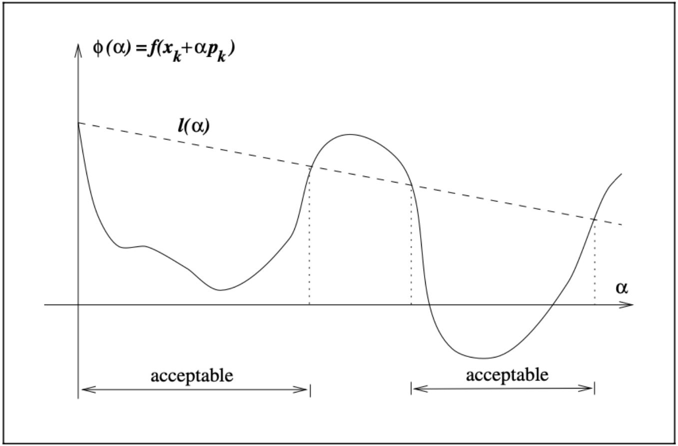

Figure: The intervals on which the Armijo condition is satisfied is shown

The Wolfe Condition

- The sufficient decrease condition is not enough by itself to ensure that the algorithm makes reasonable progress.
- As it is satisfied for all sufficiently small values of α

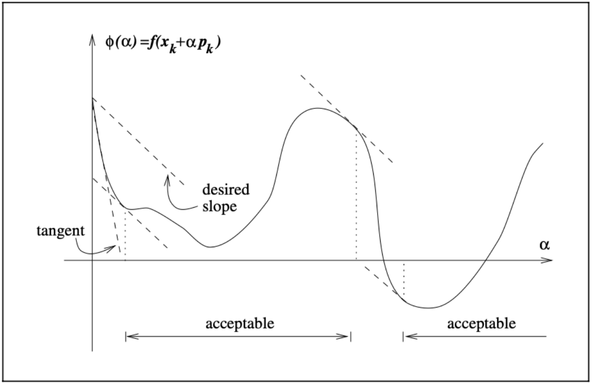

Figure: Insufficient Reduction

The Wolfe Condition

- To rule out unacceptable short steps we introduce a second requirement.

Curvature Conditions

α_k should satisfy

$$\nabla f(x_k + \alpha_k p_k)^T p_k \geq c_2 \nabla f_k^T p_k \tag{2}$$

for some constant $c_2 \in (c_1, 1)$.

- The left-hand side is simply the derivative $\phi'(\alpha_k)$.
- So the curvature condition ensures that the slope of ϕ at α_k is greater than c_2 times the initial slope $\phi'(0)$.
- If the slope $\phi'(\alpha)$ is strongly negative, we have an indication that we can reduce f significantly by moving further along the chosen direction.

The Wolfe Condition

- On, the other hand if $\phi'(\alpha)$ is only slightly negative or even positive, it is a sign that we cannot expect much more decrease in f in this direction.
- So it makes sense to terminate the line search. (See Figure 5)
- Typical values of c_2 are 0.9 when the search direction p_k is chosen by a Newton or quasi-Newton method, and 0.1 when p_k is obtained from a non-linear conjugate gradient method.
- The sufficient decrease and curvature conditions are known collectively as the Wolfe conditions.

The Wolfe Condition

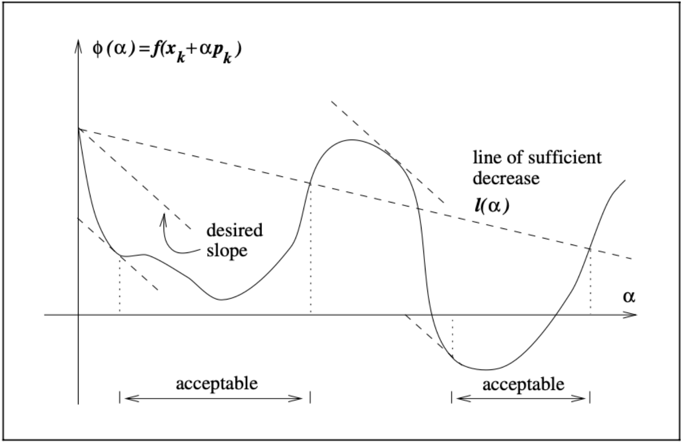

Figure: Step Lengths satisfying the Wolfe conditions.