Optimization

Lusine Poghosyan

YSU

November 30, 2020

Line Search in Multidimensional Optimization

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function that we wish to minimize.

Iterative algorithms for finding a minimizer of f are of the form

$$x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}, \quad k = 0, 1, \ldots,$$

where $x^{(0)}$ is the initial approximation.

α_k is called the step-size and $d^{(k)} \in \mathbb{R}^n$ is called the search direction.

Let $\varepsilon \in (0,1)$, $\gamma > 1$ and $\eta \in (\varepsilon, 1)$.

The *Armijo condition* ensures that α_k is not too large by requiring that

$$\Phi_k(\alpha_k) \leq \Phi_k(0) + \varepsilon \alpha_k \Phi_k'(0)$$

or

$$f\left(x^{(k)} + \alpha_k d^{(k)}\right) \leq f\left(x^{(k)}\right) + \varepsilon \alpha_k \nabla f\left(x^{(k)}\right)^T d^{(k)}.$$

It also ensures that α_k is not too small by requiring that

$$\Phi_k(\gamma\alpha_k) \geq \Phi_k(0) + \varepsilon \gamma \alpha_k \Phi_k'(0)$$

or

$$f\left(x^{(k)} + \gamma\alpha_k d^{(k)}\right) \geq f\left(x^{(k)}\right) + \varepsilon \gamma \alpha_k \nabla f\left(x^{(k)}\right)^T d^{(k)}.$$

The *Goldstein condition* (Armijo-Goldstein):

$$\Phi_k(\alpha_k) \leq \Phi_k(0) + \varepsilon \alpha_k \Phi_k'(0),$$

$$\Phi_k(\alpha_k) \geq \Phi_k(0) + \eta \alpha_k \Phi'(0).$$

The *Wolfe condition* (Armijo-Wolfe):

$$\Phi_k(\alpha_k) \le \Phi_k(0) + \varepsilon \alpha_k \Phi_k'(0),$$

$$\Phi_k'(\alpha_k) \ge \eta \Phi_k'(0).$$

The *strong Wolfe condition*:

$$\Phi_k(\alpha_k) \leq \Phi_k(0) + \varepsilon \alpha_k \Phi_k'(0),$$

$$\left| \Phi_k'(\alpha_k) \right| \leq \eta \left| \Phi_k'(0) \right| .$$

Armijo backtracking algorithm to chose the step size α_k

- Step 1: We start with some candidate value $\alpha_k^{(0)}$ for the step size α_k. Take a contraction (backtracking) factor $\tau \in (0, 1)$ and $\ell = 0$.
- Step 2: If $\alpha_k^{(\ell)}$ satisfies a prespecified termination condition (usually the first Armijo inequality) then return $\alpha_k^{(\ell)}$ for α_k. If the condition is not satisfied, then take

$$\alpha_k^{(\ell+1)} = \tau \alpha_k^{(\ell)},$$

$$\ell \longmapsto \ell + 1$$

and do the Step 2.

Example

Assume we want to find the minimizer of

$$f(x_1, x_2) = 2x_1^2 + x_2^2,$$

using the line search method. We start with $\left(x_1^{(0)}, x_2^{(0)}\right) = (1, 1)^T$ and as search direction we take $d^{(0)} = -\nabla f\left(x^{(0)}\right)$. In order to calculate the next approximation $x^{(1)}$ we need a step size α_0 which we are going to find by using Armijo backtracking algorithm. In Armijo backtracking algorithm let's take $\alpha_0^{(0)} = 2$, $\tau = 0.5$ and $\varepsilon = 0.1$. Then we calculate $\left(x_1^{(1)}, x_2^{(1)}\right)$.

Theorem (Zoutendijk)

Consider any iteration of the form

$$x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}, \quad k = 0, 1, \ldots,$$

where $d^{(k)}$ is a descent direction and α_k satisfies the Wolfe (Armijo-Wolfe) conditions. Suppose that f is bounded below and that f is continuously differentiable in $\mathbb{R}^n$. Assume also that the gradient ∇f is Lipschitz continuous on $\mathbb{R}^n$, that is, there exists a constant $L > 0$ such that

$$||\nabla f(x) - \nabla f(y)|| \leq L||x - y||, \quad \forall x, y \in \mathbb{R}^n.$$

Then

$$\sum_{k=0}^{\infty} \cos^2(\theta_k) ||\nabla f\left(x^{(k)}\right)||^2 < \infty,$$

where $\cos(\theta_k) = -\dfrac{\nabla f\left(x^{(k)}\right)^T d^{(k)}}{||\nabla f\left(x^{(k)}\right)|| \cdot ||d^{(k)}||}$.

The theorem implies

$$\cos(\theta_k) \, ||\nabla f\left(x^{(k)}\right)|| \to 0 \quad \text{as} \quad k \to \infty.$$

If we can ensure that $\cos(\theta_k) \geq \delta > 0$, then $||\nabla f\left(x^{(k)}\right)|| \to 0$.