Optimization

Lusine Poghosyan

YSU

December 3, 2020

Gradient Methods

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function. Assume $d \in \mathbb{R}^n$. Using Taylor's Theorem, we can write

$$f(x + \alpha d) = f(x) + \alpha \nabla f(x)^T d + o(\alpha), \quad \alpha \geq 0.$$

If $||d|| = 1$, then $\nabla f(x)^T d$ is called the rate of increase of f in the direction d at the point x.
By the Cauchy-Schwarz inequality

$$\nabla f(x)^T d \leq ||\nabla f(x)||.$$

We will have equality when $d = \frac{\nabla f(x)}{||\nabla f(x)||}$ and the direction in which $\nabla f(x)$ points is the direction of maximum rate of increase of f at x. The direction in which $-\nabla f(x)$ points is the direction of maximum rate of decrease of f at x. $-\nabla f(x)$ is also called steepest descent direction.

The level set of a function $f : \mathbb{R}^n \to \mathbb{R}$ at level c is the set of points

$$S = \{x : f(x) = c\}.$$

Example

Plot the level curves of f at level c if

a. $f(x_1, x_2) = x_2 - x_1^2$, $c = -1$, $c = 0$, $c = 2$;

b. $f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2$, $c = 0$, $c = 1$;

Theorem

The vector $\nabla f(x_0)$ is orthogonal to the tangent vector to an arbitrary smooth curve passing through x_0 on the level set determined by $f(x) = f(x_0)$.

Here we consider algorithms of the form

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

where x_0 is the initial approximation,

$$d^{(k)} = -\frac{\nabla f\left(x^{(k)}\right)}{||\nabla f\left(x^{(k)}\right)||}$$

and $\alpha_k \geq 0$ is the step size.

Stopping conditions

- $||\nabla f\left(x^{(k)}\right)|| < \varepsilon$
- $||x^{(k+1)} - x^{(k)}|| < \varepsilon$ or $\frac{||x^{(k+1)} - x^{(k)}||}{||x^{(k)}||} < \varepsilon$ if $||x^{(k)}|| \neq 0$
- $\left|f\left(x^{(k+1)}\right) - f\left(x^{(k)}\right)\right| < \varepsilon$ or $\frac{\left|f\left(x^{(k+1)}\right) - f\left(x^{(k)}\right)\right|}{\left|f\left(x^{(k)}\right)\right|} < \varepsilon$ if $f\left(x^{(k)}\right) \neq 0$.

Example

Assume we want to minimize numerically

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

Our initial approximation is $x^{(0)} = [3, 2]^T$.

a. Use gradient method and calculate $x^{(2)}$. Take $\alpha_0 = 5$ and $\alpha_1 = 4.5$.

b. Is this algorithm usable for our problem? Explain!

The Steepest Descent Method

The Steepest Descent Method is a gradient algorithm where α_k is chosen to be the global minimizer of $\Phi_k(\alpha)$

$$\alpha_k = \text{arg min}_{\alpha \geq 0} \Phi_k(\alpha) = \text{arg min}_{\alpha \geq 0} f\left(x^{(k)} - \alpha \nabla f(x^{(k)})\right),$$

$$x^{(k+1)} = x^{(k)} - \alpha_k \nabla f(x^{(k)}), \quad k = 0, 1, \ldots.$$

Proposition

If $\left\{x^{(k)}\right\}_{k=0}^{\infty}$ is a steepest descent sequence for a given function $f : \mathbb{R}^n \to \mathbb{R}$, then for each k the vector $x^{(k+1)} - x^{(k)}$ is orthogonal to the vector $x^{(k+2)} - x^{(k+1)}$.

- The Steepest Descent Method is globally convergent, i.e. $||\nabla f(x^{(k)})|| \to 0$, as $k \to \infty$ for any initial approximation $x^{(0)}$.
- Slow convergence, generally linear rate of convergence.

Stopping conditions

- $||\nabla f\left(x^{(k)}\right)|| < \varepsilon$
- $||x^{(k+1)} - x^{(k)}|| < \varepsilon$ or $\frac{||x^{(k+1)}-x^{(k)}||}{||x^{(k)}||} < \varepsilon$ if $||x^{(k)}|| \neq 0$
- $\left|f\left(x^{(k+1)}\right) - f\left(x^{(k)}\right)\right| < \varepsilon$ or $\frac{\left|f\left(x^{(k+1)}\right)-f\left(x^{(k)}\right)\right|}{\left|f\left(x^{(k)}\right)\right|} < \varepsilon$ if $f\left(x^{(k)}\right) \neq 0$.

Example

Assume we want to use the Steepest Descent Method to minimize

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

We start with $x^{(0)} = (1, 1)^T$. Calculate $x^{(2)}$ by using the Steepest Descent Method.