Unconstrained convex optimization problems

Huanle Xu *

April 1, 2017

1 Basic Elements of Iterative Algorithms

To fix ideas, let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function. Consider the unconstrained optimization problem:

$$\min_{x \in \mathbb{R}^n} f(x). \tag{1}$$

In general, it may be too ambitious to find a global minimum of f. Hence, we will just look for a **stationary point** of f; i.e., a point $\bar{x} \in \mathbb{R}^n$ that satisfies $\nabla f(\bar{x}) = \mathbf{0}$. To begin, let $x^0 \in \mathbb{R}^n$ be an initial iterate with $\nabla f(x^0) \neq \mathbf{0}$. In order to achieve progress, we need to proceed in some **search direction** $d^k \in \mathbb{R}^n$. For instance, we can update the iterates according to the following rule:

$$x^{k+1} = x^k + \alpha_k d^k \qquad \text{for } k = 0, 1, \ldots. \tag{2}$$

Here, $\alpha_k > 0$ is called the **step size** and controls how far we proceed in the direction d^k. Note that (2) actually defines a *family* of update rules that are parametrized by the search directions $\{d^k\}_{k \geq 0}$ and step sizes $\{\alpha_k\}_{k \geq 0}$. There are many possibilities in choosing the search directions and step sizes. Below are some common choices.

*Huanle Xu is with the College of Computer Science and Technology, Dongguan University of Technology. E-mail: {xuhl}@dgut.edu.cn.

1.1 Choosing the search directions

Roughly speaking, the method of steepest descent is based on minimizing a linear approximation of f at the current iterate $x^k \in \mathbb{R}^n$. Specifically, suppose that the current iterate x^k satisfies $\nabla f(x^k) \neq \mathbf{0}$. Then, we may construct a *linear approximation* of f at x^k, which is given by

$$f^k(x) \equiv f(x^k) + \nabla f(x^k)^T(x - x^k).$$

Now, recall that if there exists a $d \in \mathbb{R}^n$ such that $\nabla f(x^k)^T d < 0$, then there exists an $\alpha_0 > 0$ such that $f(x^k + \alpha d) < f(x^k)$ for all $\alpha \in (0, \alpha_0)$ (Proposition 1 of Handout 7). Thus, in order to guarantee descent, we need to choose $x \in \mathbb{R}^n$ such that $\nabla f(x^k)^T(x - x^k) < 0$. Of course, if $\nabla f(x^k) \neq \mathbf{0}$, then we can make $\nabla f(x^k)^T(x - x^k)$ as negative as possible. Thus, we need to restrict the length of the direction $d = x - x^k$. In particular, we may consider the following:

$$\begin{array}{ll} \text{minimize} & \nabla f(x^k)^T d \\ \text{subject to} & \|d\|_2^2 \leq \|\nabla f(x^k)\|_2^2. \end{array}$$

By the Cauchy–Schwarz inequality, we see that the optimal solution to the above problem is $d^k = -\nabla f(x^k)$. The direction d^k is called the **direction of steepest descent**, and the resulting iterative algorithm

$$x^{k+1} = x^k - \alpha_k \nabla f(x^k) \qquad \text{for } k = 0, 1, \ldots \tag{3}$$

is called the **method of steepest descent** or simply the **gradient method**.