Optimization

Lusine Poghosyan

AUA

February 15, 2019

Numerical Methods for Unconstrained Optimization

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & x \in \Omega, \end{aligned}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $\Omega \subset \mathbb{R}^n$, with $n \geq 1$.

One Dimensional Search Methods

Here we consider the minimization of univariate function $f : [a, b] \to \mathbb{R}$.

In an iterative algorithm we start with an initial candidate solution x_0 and generate sequence of points $x_1, x_2, \ldots$. Each x_{k+1} iteration depends on previous points $x_0, x_1, \ldots, x_k$. The algorithm may also use the value of f or f' or even f'' at some points:

- Golden section Search (uses only f);
- Fibonacci method (uses only f);
- Bisection method (uses only f');
- Newton's method (uses f' and f'').

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.

At each iteration we face two problems:

- first, we need to choose the search direction
- second, we need to choose the step size α_k when $d^{(k)}$ is fixed

Assume we use a descent direction $d^{(k)}$, i.e., $\nabla f\left(x^{(k)}\right)^T d^{(k)} < 0$.

Choosing the step size

In order to choose the step size we need to consider the following univariate function

$$\Phi_k(\alpha) = f\left(x^{(k)} + \alpha d^{(k)}\right), \quad \alpha \geq 0.$$

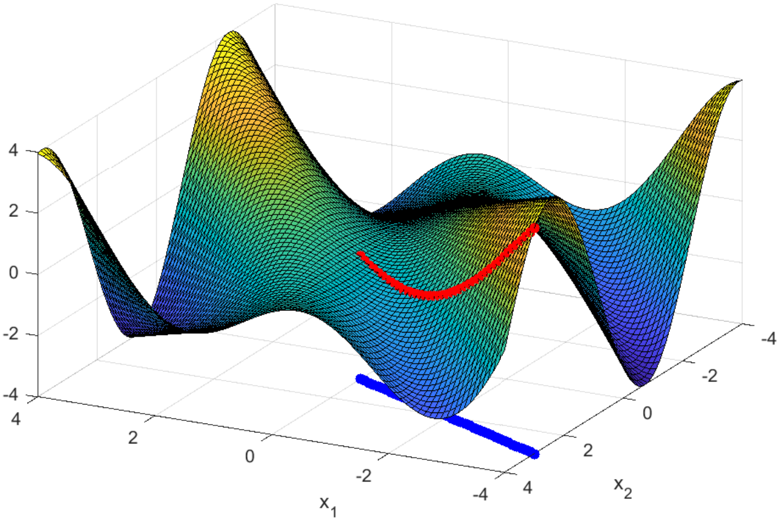
Example

Let $f(x_1, x_2) = x_1 \sin(x_1 + x_2)$.

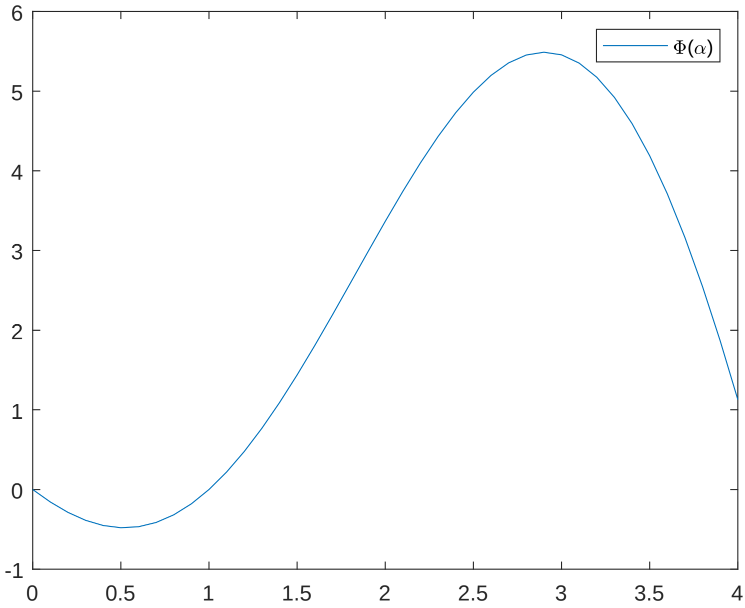
a. Show that $d = [-2, 1]^T$ is a descent direction at $x^* = [0, 1]^T$.

b. Construct the function $\Phi(\alpha) = f(x^* + \alpha d)$, $\alpha \geq 0$ and calculate $\Phi'(0)$.

c. Plot the graphs of $f(x_1, x_2)$ and $\Phi(\alpha)$, $\alpha \geq 0$.



4
2
0
-2
-4
4
2
0
-2
-4
x1
-4
-2
0
2
4
x2



6
5
4
3
2
1
0
-1
0
0.5
1
1.5
2
2.5
3
3.5
4
$\Phi(\alpha)$

$$\Phi_k'(\alpha) = \nabla f\left(x^{(k)} + \nabla d^{(k)}\right)^T d^{(k)}$$

$$\Phi_k'(0) = \nabla f\left(x^{(k)}\right)^T d^{(k)}$$

$\Phi_k'(0) < 0$ as $d^{(k)}$ is a descent direction at $x^{(k)}$.

There are two methods for choosing α_k:

- exact line search, i.e., find the minimum point of $\Phi_k(\alpha)$
- inexact line search, i.e., we need to choose α_k to ensure that $f(x^{(k+1)}) < f(x^{(k)})$ but α_k shouldn't be too small or too large.

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):

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

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

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 $\tau \in (0, 1)$ (typically $\tau = 0.5$) 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 $\left(x_1^{(1)}, x_2^{(1)}\right)$.