Optimization

Lusine Poghosyan

YSU

December 4, 2020

Newton's Method

Let $f \in \mathbb{C}^2(\mathbb{R}^n)$ and our aim is to find the minimizer of f.

Let $x^{(0)} \in \mathbb{R}^n$ be the starting point. Then we construct a quadratic function that matches its value, first and second derivatives at $x^{(0)}$ with that of the function f. This quadratic function has the form

$$q(x) = f\left(x^{(0)}\right) + \nabla f\left(x^{(0)}\right)^T \left(x - x^{(0)}\right) + \frac{1}{2}\left(x - x^{(0)}\right)^T \nabla^2 f\left(x^{(0)}\right)\left(x - x^0\right).$$

Then, instead of minimizing f, we minimize its approximation q.

The FONC for q yields

$$\nabla q(x) = \nabla f\left(x^{(0)}\right) + \nabla^2 f\left(x^{(0)}\right)\left(x - x^{(0)}\right) = 0.$$

The solution of this system

$$x^{(1)} = x^{(0)} - \left[\nabla^2 f\left(x^{(0)}\right)\right]^{-1} \nabla f\left(x^{(0)}\right)$$

will be our next approximation. Reapplying this procedure we get the sequence defined by Newton's Method

$$x^{(k+1)} = x^{(k)} - \left[\nabla^2 f\left(x^{(k)}\right)\right]^{-1} \nabla f\left(x^{(k)}\right), \quad k = 0, 1, \ldots.$$

The k-th iteration can be written in two steps:

1. Solve $\nabla^2 f\left(x^{(k)}\right) d^{(k)} = -\nabla f\left(x^{(k)}\right)$.
2. Set $x^{(k+1)} = x^{(k)} + d^{(k)}$.

- The convergence is local.
- Suppose that $f \in C^3$ and $x^* \in \mathbb{R}^n$ is a point such that $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is invertible. Then, for all $x^{(0)}$ sufficiently close to x^*, Newton's method is well-defined for all k and converges to x^* with an order of convergence at least 2.
- The direction of search is

$$d^{(k)} = -\left[\nabla^2 f\left(x^{(k)}\right)\right]^{-1} \nabla f\left(x^{(k)}\right).$$

If $\nabla^2 f\left(x^{(k)}\right)$ is positive definite, then $d^{(k)}$ is a descent direction.

Modification of Newton's method

The step size is usually $\alpha_k = 1$ but sometimes one takes other step size and gets

$$x^{(k+1)} = x^{(k)} - \alpha_k \left[\nabla^2 f\left(x^{(k)}\right)\right]^{-1} \nabla f\left(x^{(k)}\right), \quad k = 0, 1, \ldots.$$

For example we can take

$$\alpha_k = \arg\min_{\alpha \geq 0} f\left(x^{(k)} - \alpha \left[\nabla^2 f\left(x^{(k)}\right)\right]^{-1} \nabla f\left(x^{(k)}\right)\right)$$

to ensure that $f\left(x^{(k+1)}\right) < f\left(x^{(k)}\right)$.

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 Newton's Method to minimize

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

We start with $x^{(0)} = (1, 1)^T$. Calculate $x^{(2)}$ by using the Newton's Method. Explain why after one iteration we have that $\nabla f\left(x^{(1)}\right) = 0$.

Nonlinear Constrained Optimization

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & x \in \Omega, \end{aligned} \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $\Omega \subset \mathbb{R}^n$. Here, we are going to consider minimization problems, for which the constraint set Ω is given by

$$\Omega = \left\{ x \in \mathbb{R}^n : \; h_i(x) = 0, \text{ for } i = 1, \ldots m, \; g_j(x) \leq 0, \text{ for } j = 1, \ldots p \right\},$$

where $h_i : \mathbb{R}^n \to \mathbb{R}$ for $i = 1, \ldots m$, $m \leq n$ and $g_j : \mathbb{R}^n \to \mathbb{R}$ for $j = 1, \ldots p$ are given functions.

$$
\begin{aligned}
\text{minimize} \quad & f(x) \\
\text{subject to} \quad & h_i(x) = 0, \quad i = 1, \dots m, \\
& g_j(x) \leq 0, \quad j = 1, \dots p
\end{aligned}
$$

or

$$
\begin{aligned}
\text{minimize} \quad & f(x) \\
\text{subject to} \quad & h(x) = 0, \\
& g(x) \leq 0,
\end{aligned}
$$

where $h : \mathbb{R}^n \to \mathbb{R}^m$, $m \leq n$ and $g : \mathbb{R}^n \to \mathbb{R}^p$.

Example

Consider the following optimization problem:

$$\begin{aligned} \text{minimize} \quad & (x_1 - 1)^2 + x_2 - 2 \\ \text{subject to} \quad & x_2 - x_1 = 1, \\ & x_1 + x_2 \leq 2. \end{aligned}$$

Problems with equality constraints

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & h_i(x) = 0, \quad i = 1, \dots m. \end{aligned}$$

We will assume that f, h_i for $i = 1, \dots m$ are continuously differentiable functions on $\mathbb{R}^n$.

Definition

A point x^* satisfying the constraints $h_i(x^*) = 0$, $i = 1, \dots m$ is said to be a regular point of the constraints, if the gradient vectors $\nabla h_1(x^*), \dots, \nabla h_1(x^*)$ are linearly independent. When $m = 1$, this means $\nabla h_1(x^*) \neq 0$

Example

Consider following constraints $h_1(x) = x_1$ and $h_2(x) = x_2 - x_3^2$ on $\mathbb{R}^3$. Show that all feasible points are regular points.

Theorem (Lagrange's Theorem, First Order Necessary Condition)

Let x^* be a local minimizer (maximizer) of $f : \mathbb{R}^n \to \mathbb{R}$ subject to $h_i(x) = 0$, $h_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots m$, $m \leq n$. Assume f, h_i for $i = 1, \ldots m$ are continuously differentiable functions and x^* is a regular point. Then, there exists $\lambda^* \in \mathbb{R}^m$ such that

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) = 0.$$

We refer to the vector λ^* as the Lagrange multiplier vector, and its components as Lagrange multipliers.

It's convenient to introduce the Lagrangian function $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, given by

$$\mathcal{L}(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x).$$

The necessary condition for x^* to be a local minimizer will be

$$\nabla \mathcal{L}(x^*, \lambda^*) = 0$$

for some $\lambda^* \in \mathbb{R}^m$.

Example

Consider the following optimization problem:

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & h(x) = 0, \end{aligned}$$

where $f(x) = x$ and $h(x) = \begin{cases} x^2 & \text{if } x < 0, \\ 0 & \text{if } 0 \le x \le 1, \\ (x-1)^2 & \text{if } x > 1. \end{cases}$

Example

Assume we want to find the extremum points of $f(x_1, x_2) = x_1^2 + x_2^2$ subject to $x_1^2 + 2x_2^2 = 2$. Use Lagrange's theorem to find all possible local extremum points.