Optimization

Lusine Poghosyan

YSU

November 18, 2020

Newton's Method

Assume $f : \mathbb{R} \to \mathbb{R}$ is a twice differentiable function on $\mathbb{R}$.

Let x_0 be the initial approximation of the minimum point. Then we construct a quadratic function that matches its first and second derivatives at x_0 with that of the function f. This quadratic function has the form

$$q(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2.$$

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

The first-order necessary condition for a minimizer of q yields

$$0 = q'(x) = f'(x_0) + f''(x_0)(x - x_0).$$

The solution of this equation will be the next approximation

$$x_1 = x_0 - \frac{f'(x_0)}{f''(x_0)}.$$

Reapplying this procedure we get the sequence defined by Newton's Method

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}, k = 0, 1, \ldots.$$

Theorem

Let x^* be a local minimizer of f. Assume that for some $\varepsilon > 0$,

- *the initial point x_0 lies in the interval $[x^* - \varepsilon, x^* + \varepsilon]$,*
- *x^* is the unique local minimizer of f in $[x^* - \varepsilon, x^* + \varepsilon]$ and f is $m-$strongly convex in $[x^* - \varepsilon, x^* + \varepsilon]$, that is, $f(x) \geq f(y) + f'(y)(x - y) + \frac{m}{2}(x - y)^2, \forall x, y \in [x^* - \varepsilon, x^* + \varepsilon]$,*
- *$f'''(x)$ is bounded in $[x^* - \varepsilon, x^* + \varepsilon]$ by some constant L, that is $|f'''(x)| \leq L$,*
- *$L\varepsilon < 2m$,*

then the iterates of Newton's method are in ε neighborhood of x^* and

$$|x_n - x^*| \leq \frac{2m}{L}\left(\frac{L\varepsilon}{2m}\right)^{2^n}, \quad n = 0, 1 \ldots .$$

Let's prove the claims by induction, assume those are correct for $n-1$, we must prove for n.
From Taylor formula

$$|x^* - x_n| = \left| x^* - x_{n-1} + \frac{f'(x_{n-1})}{f''(x_{n-1})} \right| = \frac{|f'''(t)|}{2f''(x_{n-1})} (x^* - x_{n-1})^2 .$$

Since f is has continuous second-order derivative and is $m-$strongly convex, then $f''(x) \geq m > 0$. Using this and the fact that $f'''(x)$ is bounded

$$|x^* - x_n| \leq \frac{L}{2m} (x^* - x_{n-1})^2 .$$

Using the assumption about x_{n-1}

$$|x^* - x_n| \leq \frac{L}{2m} \varepsilon^2 < \varepsilon$$

and

$$|x^* - x_n| \leq \frac{L}{2m} \left(\frac{2m}{L} \left(\frac{L\varepsilon}{2m} \right)^{2^{n-1}} \right)^2 .$$

Stopping conditions

- $|x_n - x_{n-1}| < \varepsilon$ or $\frac{|x_n - x_{n-1}|}{|x_{n-1}|} < \varepsilon$, if $x_{n-1} \neq 0$
- $|f'(x_n)| < \varepsilon$
- $|f(x_n) - f(x_{n-1})| < \varepsilon$ or $\frac{|f(x_n) - f(x_{n-1})|}{|f(x_{n-1})|} < \varepsilon$, if $f(x_{n-1}) \neq 0$

Secant Method

Secant method mimics Newton's method but avoids the calculation of second derivative.

Newton's method

Let x_0 be an initial approximation of the root.

$$x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)}, \quad n = 0, 1 \ldots .$$

The definition of second derivative

$$f''(x) = \lim_{h \to 0} \frac{f'(x+h) - f'(x)}{h}.$$

For small h

$$f''(x) \approx \frac{f'(x+h) - f'(x)}{h}.$$

In particular, if $x = x_n$ and $h = x_{n-1} - x_n$

$$f''(x_n) \approx \frac{f'(x_{n-1}) - f'(x_n)}{x_{n-1} - x_n} = \frac{f'(x_n) - f'(x_{n-1})}{x_n - x_{n-1}}.$$

$$x_{n+1} = x_n - \left(\frac{x_n - x_{n-1}}{f'(x_n) - f'(x_{n-1})} \right) f'(x_n), \quad n = 1, 2, \ldots .$$

Let's note that we can start calculations if we have two initial approximations x_0 and x_1.

At each iteration we have only one evaluation of derivative.

Stopping conditions

- $|x_n - x_{n-1}| < \varepsilon$ or $\frac{|x_n - x_{n-1}|}{|x_{n-1}|} < \varepsilon$, if $x_{n-1} \neq 0$
- $|f'(x_n)| < \varepsilon$
- $|f(x_n) - f(x_{n-1})| < \varepsilon$ or $\frac{|f(x_n) - f(x_{n-1})|}{|f(x_{n-1})|} < \varepsilon$, if $f(x_{n-1}) \neq 0$

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

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

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

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

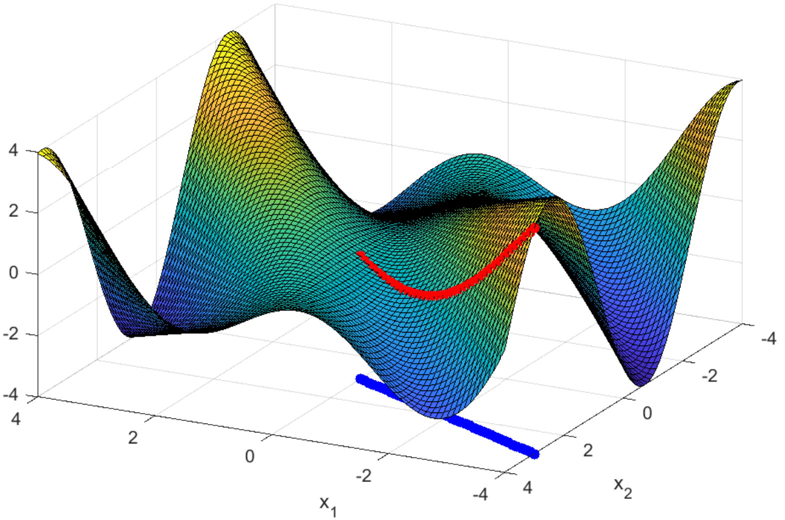
Example

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

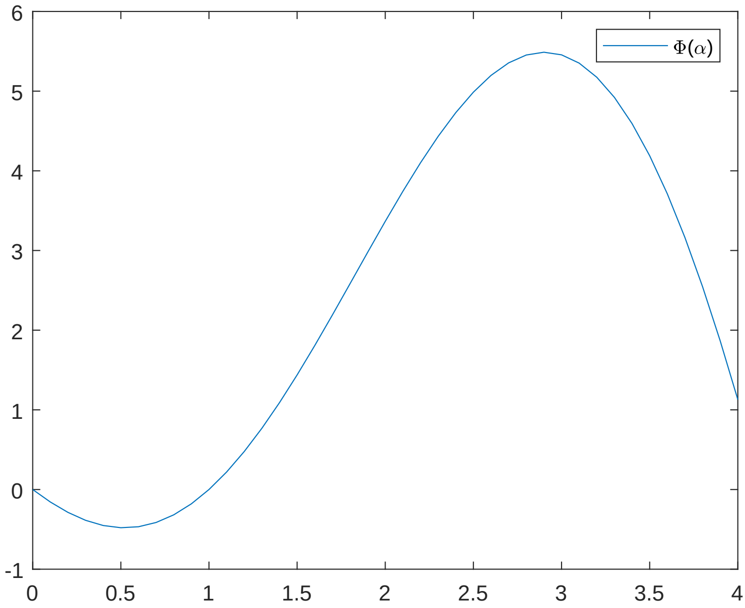
a. Show that $d = [-2, 1]^T$ is a descent direction at $x^* = [0, 1]^T$.

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

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

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