Gradient Methods

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function that we wish to minimize.

Here we consider algorithms of the form

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

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.

The Steepest Descent Method

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

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

- If $f : \mathbb{R}^n \to \mathbb{R}$ is bounded below, f is continuously differentiable and $\nabla f(x)$ is Lipschitz continuous, then The Steepest Descent Method is globally convergent i.e. $||\nabla f\left(x^{(k)}\right)|| \to 0$, as $k \to \infty$ for any initial approximation $x^{(0)}$.
- Slow convergence, even if f is strictly convex quadratic function The Steepest Descent Method has 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 $x^{(0)} = (1, 2)^T$. Calculate $x^{(2)} = \left(x_1^{(2)}, x_2^{(2)}\right)^T$ by using the Steepest Descent Method.

Example

Assume we want to use the Steepest Descent Method to minimize

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

We start with $x^{(0)} = (1, 2)^T$. Calculate $x^{(1)} = \left(x_1^{(1)}, x_2^{(1)}\right)^T$ by using the Steepest Descent Method.

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

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}, k = 1, 2, \ldots.$$

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 $f\left(x_{n-1}\right) \neq 0$