PRINCIPLES OF OPTIMIZATION

CS6140

Predrag Radivojac
Khoury College of Computer and Information Sciences
Northeastern University

Spring, 2019
NEWTON-RAPHSON OPTIMIZATION

Setting: $f : \mathbb{R}^d \rightarrow \mathbb{R}$

Objective: solve the following optimization problem

$$x^* = \arg \max_x \{f(x)\}$$
Newton-Raphson Optimization

Suppose $d = 1$. A function $f(x)$ in the neighborhood of point $x_0$

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n,$$

Taylor approximation

where $f^{(n)}(x_0)$ is the $n$-th derivative of function $f(x)$ evaluated at point $x_0$.

Consider a second order approximation:

$$f(x) \approx f(x_0) + (x - x_0)f'(x_0) + \frac{1}{2}(x - x_0)^2f''(x_0).$$
NEWTON-RAPHSON OPTIMIZATION

Find the first derivative and make it equal to zero:

\[ f'(x) \approx f'(x_0) + (x^* - x_0)f''(x_0) = 0. \]

Solving this equation for \( x^* \) gives us:

\[ x^* = x_0 - \frac{f'(x_0)}{f''(x_0)}. \]

**Idea:** Iterative optimization.

Let \( t \) be the current iteration and \( x^{(0)} \) an initial solution.

\[ x^{(t+1)} = x^{(t)} - \frac{f'(x^{(t)})}{f''(x^{(t)})}. \]
Newton-Raphson Optimization

Take \( \mathbf{x} = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d \)

\[
f(\mathbf{x}) \approx f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)^T \cdot (\mathbf{x} - \mathbf{x}_0) + \frac{1}{2} (\mathbf{x} - \mathbf{x}_0)^T \cdot H_{f(\mathbf{x}_0)} \cdot (\mathbf{x} - \mathbf{x}_0),
\]

where

\[
\nabla f(\mathbf{x}) = \left( \frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_d} \right) \quad \text{Gradient}
\]

and

\[
H_{f(\mathbf{x})} = \begin{bmatrix}
\frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_d} \\
\frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_d} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial^2 f}{\partial x_d \partial x_1} & \frac{\partial^2 f}{\partial x_d \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_d^2}
\end{bmatrix} \quad \text{Hessian}
\]
NEWTON-RAPHSON OPTIMIZATION

New update rule:

\[ \mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} - (H_{f(\mathbf{x}^{(t)})})^{-1} \cdot \nabla f(\mathbf{x}^{(t)}) \]

Both gradient and Hessian are evaluated at point \( \mathbf{x}^{(t)} \)

\[ H_{f(\mathbf{x}^{(t)})} = I \quad \rightarrow \text{gradient descent (minimization)} \quad \mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} - \eta \cdot \nabla f(\mathbf{x}^{(t)}) \]

\[ H_{f(\mathbf{x}^{(t)})} = -I \quad \rightarrow \text{gradient ascent (maximization)} \quad \mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} + \eta \cdot \nabla f(\mathbf{x}^{(t)}) \]

\[ \eta \in (0, 1) \]
Newton-Raphson Optimization

Let $\mathbf{e}^{(t)} = \mathbf{x}^{(t)} - \mathbf{x}^*$ be an error, where $\mathbf{x}^*$ is the optimum.

$$||\mathbf{e}^{(t+1)}|| = O \left( ||\mathbf{e}^{(t)}||^p \right)$$

convergence of $p$-th order

Theorem. Assume Hessian satisfies the following conditions in the neighborhood of $\mathbf{x}^*$

$$\left| \left| H(\mathbf{x}^{(t+1)}) - H(\mathbf{x}^{(t)}) \right| \right| \leq \lambda \left| \left| \mathbf{x}^{(t+1)} - \mathbf{x}^{(t)} \right| \right|$$

If $\mathbf{x}^{(t)}$ is sufficiently close to $\mathbf{x}^*$ for some $t$ and if Hessian is positive definite, then the Newton-Raphson technique is well defined and converges at second order.
CONSTRUANED OPTIMIZATION

Objective: solve the following optimization problem

\[ x^* = \arg \max_x \{ f(x) \} \]

Subject to:
\[ g_i(x) = 0 \quad \forall i \in \{1, 2, \ldots, m\} \]
\[ h_j(x) \geq 0 \quad \forall j \in \{1, 2, \ldots, n\} \]

Or, in a shorter notation, to:
\[ g(x) = 0 \]
\[ h(x) \geq 0 \]
LAGRANGE MULTIPLIERS

Taylor’s expansion for $g(x)$, where $x + \epsilon$ is on the surface of $g(x)$

$$g(x + \epsilon) \approx g(x) + \epsilon^T \nabla g(x)$$

We know that $g(x) = g(x + \epsilon)$

$$\epsilon^T \nabla g(x) \approx 0$$

when $\epsilon \to 0$

$$\epsilon^T \nabla g(x) = 0 \implies \nabla g(x) \text{ is orthogonal to the surface}$$

$g(x) = 0$ \quad $\nabla g(x)$ and $\nabla f(x)$ are parallel!

$$\nabla f(x) + \alpha \nabla g(x) = 0 \quad \alpha \neq 0$$

$L(x, \alpha) = f(x) + \alpha g(x)$
**LAGRANGE MULTIPLIERS**

Inactive constraint

\[ h(x) > 0 \]

\[ \nabla f(x) = 0 \]

Active constraint

\[ h(x) > 0 \]

\[ \nabla f(x) = -\mu \nabla h(x) \quad \mu > 0 \]

It holds that:

\[ h(x) \geq 0 \]
\[ \mu \geq 0 \]
\[ \mu \cdot h(x) = 0 \]  

**Karush-Kuhn-Tucker (KKT) conditions**

\[ L(x, \alpha, \mu) = f(x) + \alpha^T g(x) + \mu^T h(x) \]