PRINCIPLES OF OPTIMIZATION

CS6140

Predrag Radivojac

KHOURY COLLEGE OF COMPUTER SCIENCES

NORTHEASTERN UNIVERSITY

Spring 2021

Newton-Raphson optimization

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

Objective: solve the following optimization problem

$$\boldsymbol{x}^* = \arg\max_{\boldsymbol{x}} \{f(\boldsymbol{x})\}$$

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 the second order approximation:

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

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:

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

Idea: Iterative optimization

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

$$x^{(t+1)} = x^{(t)} - \frac{f'(x^{(t)})}{f''(x^{(t)})} \qquad t = 1, 2, \ldots$$

NEWTON-RAPHSON OPTIMIZATION

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

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

where

$$\nabla f(\boldsymbol{x}) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_d}\right)$$

Gradient

and

$$H_{f(\boldsymbol{x})} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_d} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & & \\ \vdots & & \ddots & \\ \frac{\partial^2 f}{\partial x_d \partial x_1} & & & \frac{\partial^2 f}{\partial x_d^2} \end{bmatrix}$$

Hessian

Newton-Raphson optimization

New update rule:

$$\boldsymbol{x}^{(t+1)} = \boldsymbol{x}^{(t)} - \left(H_{f(\boldsymbol{x}^{(t)})}\right)^{-1} \cdot \nabla f(\boldsymbol{x}^{(t)})$$

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

$H_{f(\boldsymbol{x}^{(t)})} \leftarrow I$ $\rightarrow$ gradient descent (minimization) $\boldsymbol{x}^{(t+1)} = \boldsymbol{x}^{(t)} - \eta \cdot \nabla f(\boldsymbol{x}^{(t)})$

$H_{f(\boldsymbol{x}^{(t)})} \leftarrow -I$ $\rightarrow$ gradient ascent (maximization) $\boldsymbol{x}^{(t+1)} = \boldsymbol{x}^{(t)} + \eta \cdot \nabla f(\boldsymbol{x}^{(t)})$

$\eta \in (0, 1]$

NEWTON-RAPHSON OPTIMIZATION

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

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

convergence of p-th order

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

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

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

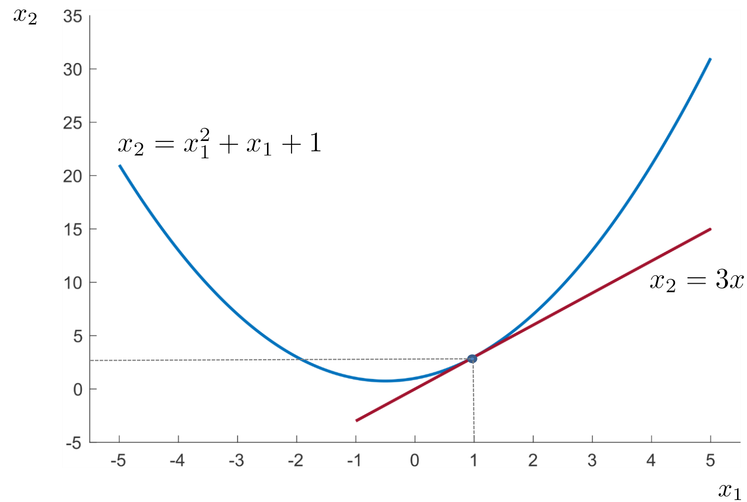
Preliminaries for constrained optimization

$$f(x_1, x_2) = x_1^2 + x_1 - x_2 + 1$$

Consider: level set $f(x_1, x_2) = 0$

point $(1, 3)$ from the level set

Find: tangent $x_2 = ax_1 + b$ at $(1, 3)$



CONSTRAINED OPTIMIZATION

Objective: solve the following optimization problem

$$\boldsymbol{x}^* = \arg\max_{\boldsymbol{x}} \{f(\boldsymbol{x})\}$$

Subject to:

$$g_i(\boldsymbol{x}) = 0 \qquad \forall i \in \{1, 2, \ldots, m\}$$

$$h_j(\boldsymbol{x}) \geq 0 \qquad \forall j \in \{1, 2, \ldots, n\}$$

Or, in a shorter notation, to:

$$\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{0}$$

$$\boldsymbol{h}(\boldsymbol{x}) \geq \boldsymbol{0}$$

Lagrange multipliers

Taylor's expansion for $g(\boldsymbol{x})$, where $\boldsymbol{x}+\boldsymbol{\epsilon}$ is on a level surface of $g(\boldsymbol{x})$

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

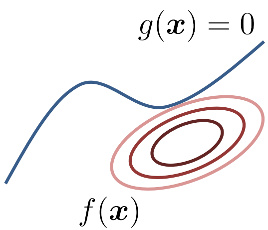
We know that $g(\boldsymbol{x}) = g(\boldsymbol{x}+\boldsymbol{\epsilon})$

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

when $\boldsymbol{\epsilon} \to \mathbf{0}$

$$\boldsymbol{\epsilon}^T \nabla g(\boldsymbol{x}) = 0$$

$\implies$ $\nabla g(\boldsymbol{x})$ is orthogonal to the level surface



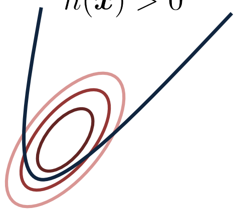
$\nabla g(\boldsymbol{x})$ and $\nabla f(\boldsymbol{x})$ are parallel!

$$\nabla f(\boldsymbol{x}) + \alpha \nabla g(\boldsymbol{x}) = 0 \qquad \alpha \neq 0$$

$$L(\boldsymbol{x}, \alpha) = f(\boldsymbol{x}) + \alpha g(\boldsymbol{x})$$

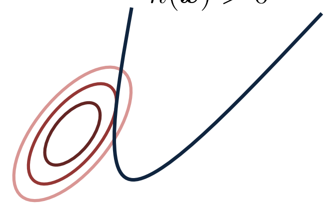
LAGRANGE MULTIPLIERS

Inactive constraint:



$\nabla f(\boldsymbol{x}) = 0$

Active constraint:



$\nabla f(\boldsymbol{x}) = -\mu \nabla h(\boldsymbol{x}) \qquad \mu > 0$

It holds that:

$$h(\boldsymbol{x}) \geq 0$$
$$\mu \geq 0$$
$$\mu \cdot h(\boldsymbol{x}) = 0$$

Karush-Kuhn-Tucker (KKT) conditions

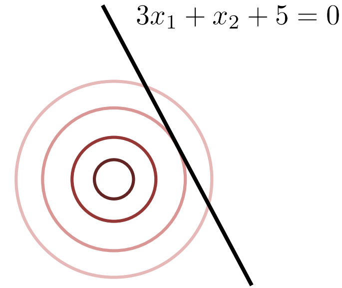
$$L(\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\mu}) = f(\boldsymbol{x}) + \boldsymbol{\alpha}^T \boldsymbol{g}(\boldsymbol{x}) + \boldsymbol{\mu}^T \boldsymbol{h}(\boldsymbol{x})$$

EXAMPLE: LAGRANGE MULTIPLIERS

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}} \left\{ x_1^2 + x_2^2 \right\}$$

Subject to:

$$3x_1 + x_2 + 5 = 0$$



Example: Lagrange multipliers

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \{||\mathbf{x}||\}$$

Subject to:

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m$, $m < n$