SW8

Der Gradient ∇f of Funktion $f: \mathbb{R}^n \to \mathbb{R}$ $\mathcal{S}$ the vector consisting of the n partial derivatives:

$$\nabla f(\vec{x}) = \begin{pmatrix} f_{x_1}(\vec{x}) \\ f_{x_2}(\vec{x}) \\ \vdots \\ f_{x_n}(\vec{x}) \end{pmatrix}$$

// Ableitung von allen Funktionen im Vektor

#1 Jeder Punkt $\vec{x}$ zeigt in die Richtung des deepest descent.

#2 Dessen Norm $\|\nabla f(\vec{x})\|$ zeigt die Steigung dieser Richtung

$\Rightarrow$ Wenn alle $\vec{x}$ in $\nabla f(\vec{x})$ 0 sind, haben wir ein Stationären Punkt.

Algorithmn $\to$ Gradient descent algorithm

To find a local minimum, we start at a point x_0 and go into the negativ direction of the gradient.

$x^{i+1} = x^i - \beta \nabla f(x^i)$ $\Rightarrow$ choose a good step size β.

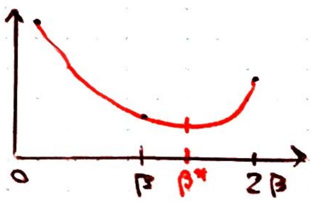
Step size rule 1:

- Set β to 1 and check if you are getting smaller if you go in direction of the gradient.
- If you didn't get smaller, make β half of it and repeat the first step. If you found a smaller point with β in the direction of the gradient, you double the step size β.

Step size rule 2

- Improvement of β after applying rule 1: Approximate f near x^i in direction of the negativ gradient by a quadratic parabola. Consider choosing x^{i+1} according to the minimum of the parabola.

Compute $P(t) = at^2 + bt + c$ such that



$$\left.\begin{aligned} P(0) &= f(x^i) \\ P(\beta) &= f(x^i - \beta\nabla f(x^i)) \\ P(2\beta) &= f(x^i - 2\beta\nabla f(x)) \end{aligned}\right\} \begin{aligned} &= c \\ &= a\beta^2 + b\beta + c \\ &= 4a\beta^2 + 2b\beta + c \end{aligned}$$

$$\Rightarrow a = \frac{1}{2\beta^2}\left[f(x^i) - 2f(x^i - \beta\nabla f(x^i)) + f(x^i - 2\beta\nabla f(x^i))\right]$$

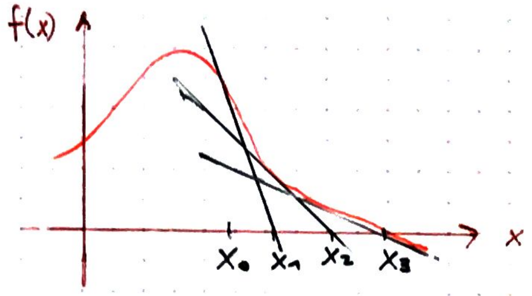
$$b = \frac{1}{2\beta}\left[-3f(x^i) + 4f(x^i - \beta\nabla f(x^i)) - f(x^i - 2\beta\nabla f(x^i))\right]$$

$$\Rightarrow \beta^* = \frac{-b}{2a} = \frac{\beta}{2} \cdot \frac{3f(x^i) - 4f(x^i - \beta\nabla f(x^i)) + f(x^i - 2\beta\nabla f(x^i))}{f(x^i) - 2f(x^i - \beta\nabla f(x^i)) + f(x^i - 2\beta\nabla f(x^i))}$$

Newton's method

In it's original form, it finds zeros of a function.

We want to find a point $\vec{x}$ for which $f(\vec{x}) = 0$



$$x^{i+1} = x^i - \frac{f(x^i)}{f'(x^i)}$$

Newton's method applied to the derivative f'

$$x^{i+1} = x^i - \frac{f'(x^i)}{f''(x^i)}$$

approximates the zeros of f', $\Rightarrow$ i.e. the stationary points of f.

Equivalently:
$$x^{i+1} = x^i - (f''(x^i))^{-1} \cdot f'(x^i)$$

For a multidimensional function:
$$x^{i+1} = x^i - (H_f(x^i))^{-1} \cdot \nabla f(x^i)$$

$H_f(x^i)$ = Hessian Matrix = Ableitung vom $\nabla f(x^i)$

$$H_f(x^i) = \begin{bmatrix} \frac{\partial^2 f(x^i)}{\partial x_1 \partial x_1} & \frac{\partial^2 f(x^i)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x^i)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x^i)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x^i)}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f(x^i)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x^i)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x^i)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x^i)}{\partial x_n \partial x_n} \end{bmatrix}$$