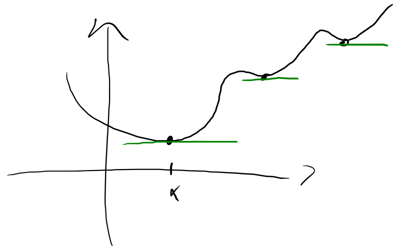
Optimierung

$f: \mathbb{R}^n \to \mathbb{R}$



$\exists \delta: \; f(x^*) < f(x), \quad x \in [x^* - \delta, x^* + \delta] \setminus x^*$ lokales Minimum

$\Omega = [v^-, v^+] \quad f(x^*) < f(x), \quad x \in \Omega \setminus x^*$ globales Minimum

$f: \mathbb{R} \to \mathbb{R}$, stetig diff-bar $\to$

notwendig $f'(x^*) = 0$

$f''(x^*) > 0 \to$ Minimum

konvex um x^*

nD

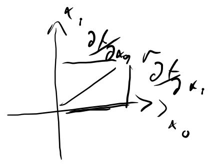
Minimierung mit Nebenbdg.

$\underset{x \in \Omega}{\operatorname{argmin}} f(x)$, s.d. $g(x) = c$

$f: \mathbb{R}^n \to \mathbb{R} \qquad \frac{\partial f}{\partial x_i} \to$ beschreibt wie sich $f(x) = f(x_0, x_1, \ldots, x_{n-1})$ verändert, wenn man an x_i „wackelt"

$r \qquad D_r f = \frac{\partial f}{\partial x_0} r_0 + \ldots + \frac{\partial f}{\partial x_{n-1}} r_{n-1} = r^T \nabla f$

$\nabla f = \left(\frac{\partial f}{\partial x_0}, \ldots, \frac{\partial f}{\partial x_{n-1}} \right)^T$



Eigenschaften von ∇f:

1. $\nabla f(x^*) = 0$ notwendige Bdg f. Minimum
2. $\underset{\text{argmax}}{\|r^T \nabla f\|}, \|r\| = 1 \to r = \frac{\nabla f}{\|\nabla f\|}$

$\to \nabla f(x)$ ist die Richtung des steilsten Anstieges

Hesse-Matrix

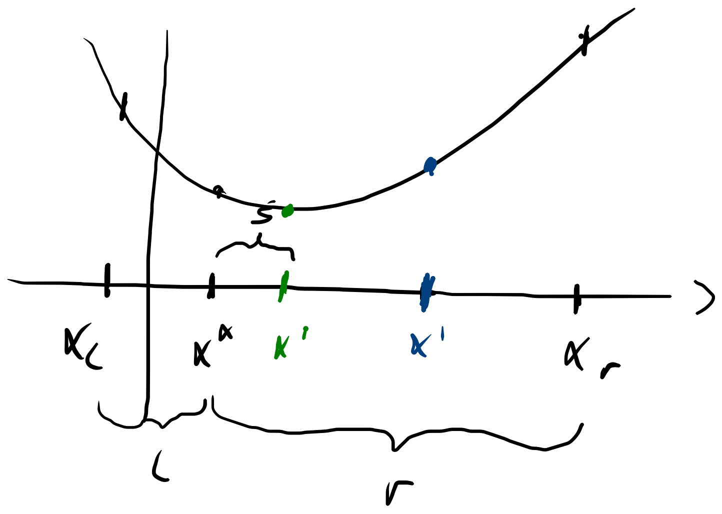
$$H_f = \begin{pmatrix} \frac{\partial^2 f}{\partial x_0^2} & \frac{\partial^2 f}{\partial x_0 \partial x_1} & \frac{\partial^2 f}{\partial x_0 \partial x_2} \\ \frac{\partial^2 f}{\partial x_1 \partial x_0} & \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \vdots & & \end{pmatrix}$$

$$f(x) \approx f(a) + (x-a)^T \nabla f(a) + \frac{1}{2} (x-a)^T H_f(a) (x-a)$$

$$f(x^* + d) \approx f(x^*) + d^T \underbrace{\nabla f(x^*)}_{0} + \frac{1}{2} \underbrace{d^T H_f(x^*) d}_{> 0}$$

Minimum $\Leftrightarrow$ H_f ist PD

Minimierung $f: \mathbb{R} \to \mathbb{R}$ ohne Ableitungen (Bisektion)



$$x_l < x^* < x_r$$

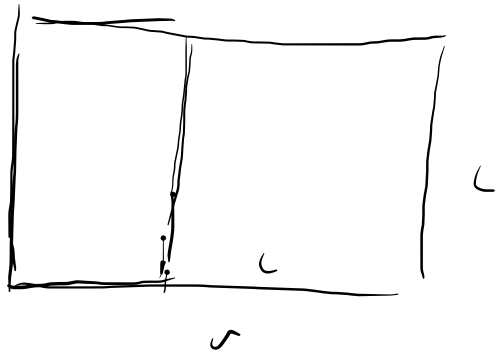
$$f(x^*) < f(x_l),\ f(x_r)$$

- $f(x') < f(x^*)$ $\quad x_l \leftarrow x^*$, $\quad x^* \leftarrow x'$
- $f(x') \geq f(x^*)$ $\quad x_r \leftarrow x'$

$$\frac{l}{r} = \frac{s}{r-s} = \frac{l}{s}$$

$$\to \frac{l}{r} = \frac{r-l}{l}$$

$\to$ Goldene Schnitt



Newton-Verfahren

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

Ges. $f(x) = 0$

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

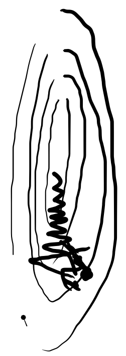
$$H_f(x_i)\,\Delta x = -\nabla f(x_i)$$ $(\nabla f : \mathbb{R}^n \to \mathbb{R}^n)$

Gradientenabstieg

$$H_f(x) \approx s \cdot I$$

$$x^{i+1} = x^i - \frac{1}{s}\nabla f(x_i) = x^i - \sigma \nabla f(x^i)$$

$f(x_0, x_1) = x_0^2 + \varepsilon x_1^2$



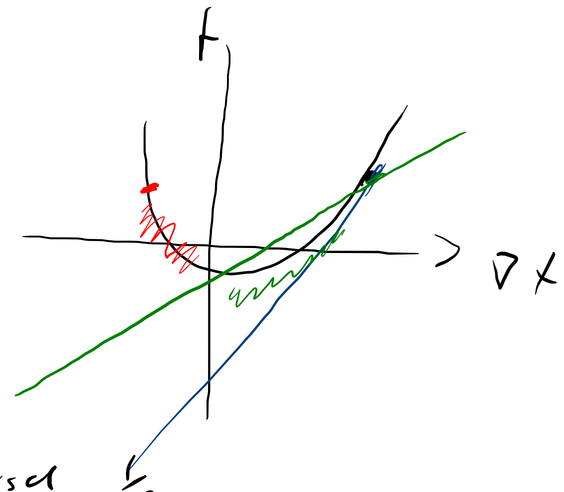
$$g(\sigma) = f(x^i - \sigma \nabla f(x^i))$$ $g : \mathbb{R} \to \mathbb{R}$

$$\arg\min_{\sigma} g(\sigma)$$

Backtracking

$\sigma = s \cdot \tau^k$ $\quad \tau \in]0,1[$ typisch $\frac{1}{2}$

$k = 0, 1 \ldots$

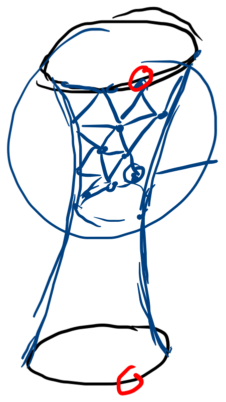


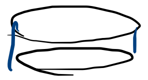
$$f(x^i - \sigma \nabla f(x^i)) < f(x^i) - \alpha \sigma \nabla f(x^i)^T \nabla f(x^i)$$

Armijo-Regel
Goldstein
1. Wolfe Condition

$\alpha \in]0,1[$ typisch $\frac{1}{2}$

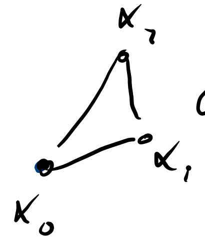
Anwendungsproblem





$X \in \mathbb{R}^{3n}$

$x_i \in \mathbb{R}^3$



$a_0 = \frac{1}{2}\|(x_1 - x_0) \times (x_2 - x_0)\|$

$$f(x) = \sum_i a_i(x)$$

$$\operatorname*{argmin}_x \; f(x)$$

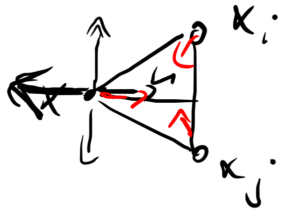
$$\nabla f(x) = \sum_i \nabla_x a_i(x)$$

$x_i = (a, b, 1)$ $\quad a^2 + b^2 = r^2$

$x_j = (c, d, -1)$ $\quad c^2 + d^2 = r^2$

$\vdots$

$g(x) = c$



$a_i = \frac{1}{2} \cdot h \cdot \|x_i - x_j\|$

$\|h\| = 1$

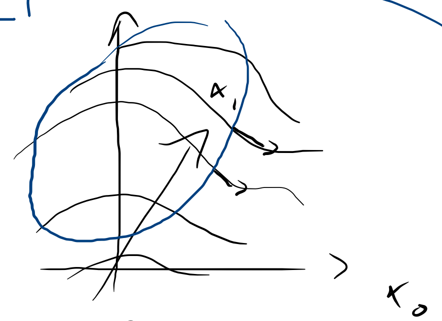
$\nabla_x a_i = \frac{1}{2}(x_i - x_j) \times h$

$h^T(x_0 - x_1) = 0$

$h^T(x_1 - x_2) = 0$

Lagrange - Multiplikatoren

$\arg\min_x f(x)$ s.d. $\boxed{g(x) = c}$



Annahme $f, g \in \mathbb{R}^2 \to \mathbb{R}$

Bewegung auf $g(x) = c$ in Richtung r

$r^T \nabla g = 0$

Minimum von f

$r^T \nabla f = 0$

→ Für $f(x^*)$ Minimum sind

$\nabla g(x^*)$ und $\nabla f(x^*)$ <u>parallel</u>

→ $\nabla f(x^*) = \lambda \nabla g(x^*)$

$$L(x, \lambda) = f(x) - \lambda(g(x) - c)$$

$$\nabla_x = \nabla f(x) - \lambda \nabla g(x) = 0$$

$$\frac{\partial L}{\partial \lambda} = g(x) - c = 0$$

→ Lagrange Funktion

Viele Nebenbdg $g_i(x) = c_i$

$$L(x, \lambda_0, \lambda_1, \dots) = f(x) - \lambda_0(g_0(x) - c_0) - \lambda_1(g_1(x) - c_1) - \dots$$