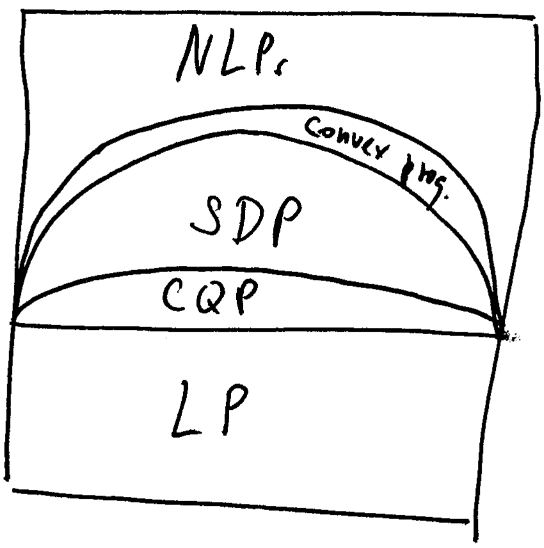
<u>Algorithmic consequence:</u> Every descent method in which intermediate steps <u>stay in C</u> will eventually find an optimal solution.

iii) Staying in C can be computationally difficult.

$\rightsquigarrow$ Convex program are still to general for a mathematical theory with efficient algorithm.

iv) many NLPs can be convexified ($\rightsquigarrow$ see end of this lecture)

<u>Landscape of NLPs</u>



main subject.

SDP = semidefinite programm

CQP = conic quadratic program.

Chapter II Minimizing a convex function

For solving convex programs up to high accuracy one employs Newton's method (also inside the interior point method).

<u>Here</u>: Analysis of Newton's method for minimizing an unconstrained convex function satisfying rather strong assumptions.

<u>given</u>: $f : \mathbb{R}^n \to \mathbb{R}$ convex

<u>goal</u>: find $p^* = \inf_{x \in \mathbb{R}^n} f(x)$ $\quad (*)$.

To solve $(*)$ one constructs a <u>minimizing sequence</u> $x^{(0)}, x^{(1)}, \ldots$ with $f(x^{(0)}) \geq f(x^{(1)}) \geq \ldots$ and $f(x^{(k)}) \to p^*$ as $k \to \infty$.

<u>Assumptions</u> a) f is convex and twice continuously differentiable.

b) $x^{(0)}$ is given. Define $S = \{ x \in \mathbb{R}^n : f(x) \leq f(x^{(0)}) \}$, a convex set.

c) f is <u>strongly convex</u> on S, i.e. $\exists m > 0 \; \forall x \in S$: smallest eigenvalue of Hessian $Hf(x)$ is at least m.

d) Hessian of f is Lipschitz continuous on S with constant L, i.e.

$$\| Hf(x) - Hf(y) \| \leq L \, \| x - y \| \quad \text{for all } x, y \in S.$$

e) $\| \nabla f(x^{(0)}) \| \leq \eta$ with $0 \leq \eta \leq \frac{m^2}{L}$ holds,

where $\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \ldots, \frac{\partial f}{\partial x_n}(x) \right)^T$ is the <u>gradient</u> of f at x.

<u>Algorithm 1</u> (Newton's method)

<u>input</u> f, $x^{(0)}$ satisfying assumptions above, $N \in \mathbb{N}$

<u>output</u> $x^{(N)}$ with $f(x^{(N)}) - p^* \leq \frac{2m^3}{L^2} \left(\frac{1}{2}\right)^{2^{N+1}}$

for $i = 1$ to N do

$$x^{(i)} = x^{(i-1)} - \left(Hf(x^{(i-1)}) \right)^{-1} \nabla f(x^{(i-1)}).$$

end for

Where is the formula for $x^{(i)}$ coming from?

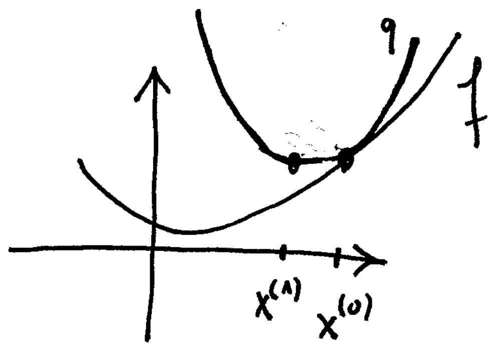
Taylor approximation of f at $x^{(i-1)}$:

$$f(x^{(i-1)} + x) = \overbrace{f(x^{(i-1)}) + \nabla f(x^{(i-1)})^T x + \frac{1}{2} x^T Hf(x^{(i-1)}) x}^{q(x)} + \ldots$$

$q(x)$: quadratic approximation of f at $x^{(i-1)}$.

By assumption: $Hf(x^{(i-1)})$ is positive definite. So q has a unique minimizer x^*:

$$0 = \nabla q(x^*) = \nabla f(x^{(i-1)}) + Hf(x^{(i-1)})x^*$$
$$\iff \quad x^* = -(Hf(x^{(i-1)}))^{-1} \nabla f(x^{(i-1)})$$



Analysis of Algorithm 1

Proposition 2 (Method of Lagrange multipliers ⇝ Analysis II)

Let $U \subseteq \mathbb{R}^n$ be an open set, let $f: U \to \mathbb{R}$, $g_1, \dots, g_N: U \to \mathbb{R}$ be continuously differentiable functions. Suppose $x^* \in U$ is a local extreme point of f satisfying the equality constraints $g_1(x^*) = \dots = g_N(x^*) = 0$. Then there are $\lambda_1, \dots, \lambda_N \in \mathbb{R}$ (Lagrange multipliers) so that

$$\nabla f(x^*) = \sum_{i=1}^{N} \lambda_i \nabla g_i(x^*)$$ holds.

Proposition 3 (Rayley-Ritz principle)

Let $C \in \mathbb{R}^{n\times n}$ be a symmetric matrix and let $\lambda_{max}(C)$, $\lambda_{min}(C)$ the largest, resp. the smallest eigenvalue of C.

Then,

$$\lambda_{max}(C) = \max_{x \in \mathbb{R}^n \setminus \{0\}} \frac{x^T C x}{x^T x} = \max_{x \in S^{n-1}} x^T C x,$$

$$\lambda_{min}(C) = \min_{x \in \mathbb{R}^n \setminus \{0\}} \frac{x^T C x}{x^T x} = \min_{x \in S^{n-1}} x^T C x,$$

where $S^{n-1} = \{x \in \mathbb{R}^n : x^T x = 1\}$ is the unit sphere.

<u>Proof</u>: Use Lagrange multipliers:

$$f(x) = x^T C x, \quad g_1(x) = x^T x - 1$$

Then $\nabla f(x) = 2Cx$, $\nabla g_1(x) = 2x$. If $x^* \in \mathbb{R}^n$ is a local extreme point of f under the constraint $(x^*)^T x^* = 1$, then there is a λ_1 with

$$2Cx^* = \lambda_1 2 x^*.$$

Hence, x^* is an eigenvector of C with eigenvalue λ_1.
Now the result follows immediately. $\boxtimes$

<u>Lemma 4</u> Under the above assumptions we have

$$f(x) - p^* \leq \frac{1}{2m} \|\nabla f(x)\|^2 \quad \text{for all } x \in S.$$

<u>Proof</u> Use Lagrange form of the Taylor expansion of f at x: