Convex Optimization and Modeling

(Un)constrained minimization

9th lecture, 09.06.2010

Jun.-Prof. Matthias Hein

Descent Methods:

- steepest decsent: $x^{k+1} = x^k - \alpha \nabla f(x^k)$,
- general descent: $x^{k+1} = x^k + \alpha\, d^k$ where $\langle d^k, \nabla f(x^k)\rangle < 0$,
- linear convergence (stepsize selection with Armijo rule),

Newton method:

- Newton's method: $x^{k+1} = x^k - \alpha\,(Hf(x^k))^{-1}\nabla f(x^k)$,
- Hessian is positive-(semi)-definite for convex functions $\Longrightarrow \langle (Hf(x^k))^{-1}\,\nabla f(x^k), \nabla f(x^k)\rangle \geq 0$,
- quadratic convergence

Convergence analysis: involves possibly unknown properties of the function, bound is not affinely invariant

Unconstrained Minimization (continued)

- **Self-concordant functions:**
  - convergence analysis directly in terms of Newton decrement
- **Subgradient Methods**

Constrained Minimization:

- **Equality constrained minimization:**
  - Newton method with equality constraints
  - Newton method with infeasible start
- **Interior point methods:**
  - barrier method

Problems of classical convergence analysis

- depends on unknown constants $(m, L, \cdots)$,
- Newtons method is affine invariant but not the bound.

Convergence analysis via self-concordance (Nesterov and Nemirovski)

- does not depend on any unknown constants
- gives affine-invariant bound
- applies to special class of convex functions (self-concordant functions)
- developed to analyze polynomial-time interior-point methods for convex optimization

Self-concordant functions:

Definition 1. *A function $f : \mathbb{R} \to \mathbb{R}$ is* ***self-concordant*** *if*

$$|f'''(x)| \leq 2 f''(x)^{\frac{3}{2}}.$$

A function $f : \mathbb{R}^n \to \mathbb{R}$ is ***self-concordant*** *if $t \mapsto f(x + tv)$ is self-concordant for every $x, v \in \mathbb{R}^n$.*

Examples:

- linear and quadratic functions,
- negative logarithm $f(x) = -\log x$.

Properties:

- If f self-concordant, then also γf where $\gamma > 0$.
- If f is self-concordant then $f(Ax + b)$ is also self-concordant.

Convergence analysis for a strictly convex self-concordant function:

Two phases: $0 < \eta < \frac{1}{4}$, $\gamma > 0$,

- **damped Newton phase:** $\lambda(x^k) > \eta$,

$$\gamma > 0, \qquad f(x^{k+1}) - f(x^k) \leq -\gamma.$$

- **pure Newton phase:** $\lambda(x^k) \leq \eta$,

$$2\lambda(x^{k+1}) \leq \Big(2\lambda(x^k)\Big)^2.$$

stepsize $\alpha^k = 1 \Rightarrow$ pure Newton step for $l \geq k$

$$f(x^l) - p^* \leq \lambda(x^l)^2 \leq \left(\frac{1}{2}\right)^{2^{l-k+1}}.$$

$\Longrightarrow$ **complexity bound only depends on known constants !**

$\Longrightarrow$ **does not imply that Newton's method works better for self-concordant functions !**

Steepest Descent Method - f differentiable

Minimize linear approximation: $f(x + \alpha d) = f(x) + \alpha \langle \nabla f, d \rangle$,

$$\min_{\|d\| \leq 1} \langle \nabla f, d \rangle = -\|\nabla f\|, \qquad d^* = -\frac{\nabla f}{\|\nabla f\|}.$$

Steepest Descent Method - f non-differentiable but convex

Definition of subgradient/subdifferential,

$$f(y) \geq f(x) + \langle g, y - x \rangle, \quad \forall\, y \in \mathbb{R}^d,\ g \in \partial f(x).$$

Directional derivative $f'(x, d)$ of f into direction d,

$$f'(x, d) = \max_{g \in \partial f(x)} \langle g, d \rangle,$$

Direction with steepest descent

$$\min_{\|d\| \leq 1} \max_{g \in \partial f(x)} \langle g, d \rangle.$$

Min-Max Equality - Saddle-Point Theorems

Theorem 1. *Let $f(x, y)$ be convex in $x \in X$ and concave in $y \in Y$ and suppose* $\operatorname{dom} f \in X \times Y$, *$f$ is continuous and X, Y are compact. Then,*

$$\sup_{y \in Y} \inf_{x \in X} f(x, y) = \inf_{x \in X} \sup_{y \in Y} f(x, y).$$

Application to steepest descent problem yields

$$\min_{\|d\| \leq 1} \max_{g \in \partial f(x)} \langle g, d \rangle = \max_{g \in \partial f(x)} \min_{\|d\| \leq 1} \langle g, d \rangle = \max_{g \in \partial f(x)} - \|g\|,$$

where $d^* = -\frac{g^*}{\|g^*\|}$ if $\|g^*\| > 0$, otherwise $d^* = 0$.

Steepest Descent: $x^{k+1} = x^k - \alpha \, (g^*)^k$.

Problem: Does not always converge to optimum with exact line search (Exercise).

Subgradient Method I

Alternative approach:

Instead of descent in $f(x^k) - p^* \implies$ descent in $\|x^k - x^*\|$.

Warning: Later approach does not necessarily lead to a monotonically decreasing sequence $f(x^k)$!

Subgradient method:

$$x^{k+1} = x^k - \alpha^k g^k, \qquad \alpha^k > 0,\ g_k \in \partial f(x^k).$$

- No stepsize selection ! α^k will be fixed initially.
- **Any** subgradient is o.k. ! Do not have to know $\partial f(x^k)$ - one subgradient for each point is sufficient.

Main Theorem

- For all $y \in X$, $k \geq 0$,

$$\left\| x^{k+1} - y \right\|^2 \leq \left\| x^k - y \right\|^2 - 2\alpha^k (f(x^k) - f(y)) + \alpha^k \left\| g^k \right\|^2.$$

- If $f(y) < f(x^k)$ and $0 < \alpha^k < \frac{2(f(x^k) - f(y))}{\left\| g^k \right\|^2}$, then

$$\left\| x^{k+1} - y \right\|^2 \leq \left\| x^k - y \right\|^2.$$

Proof: using $f(y) - f(x^k) \geq \left\langle g^k, y - x^k \right\rangle$.

$$\begin{aligned}\left\| x^{k+1} - y \right\|^2 = \left\| x^k - \alpha^k g^k - y \right\|^2 &= \left\| x^k - y \right\|^2 - 2\alpha^k \left\langle g^k, x^k - y \right\rangle + (\alpha^k)^2 \left\| g^k \right\|^2 \\ &\leq \left\| x^k - y \right\|^2 - 2\alpha^k (f(x^k) - f(y)) + (\alpha^k)^2 \left\| g^k \right\|^2,\end{aligned}$$

$\Longrightarrow$ we are interested in $y = x^*$!

Subgradient Method III

Assumptions: Optimum is attained and unique, $p^* = f(x^*)$

$$C := \sup_k \{\|g\| \mid g \in \partial f(x^k)\} < \infty.$$

This holds if f is Lipschitz continuous with Lipschitz constant $L < \infty$,

$$|f(y) - f(x)| \leq L \|y - x\|.$$

as $\|g\| \leq L$ for any x and $g \in \partial f(x)$.

Recursive application:

$$\left\|x^{k+1} - x^*\right\|^2 \leq \left\|x^0 - x^*\right\|^2 - 2\sum_{s=0}^{k} \alpha^s (f(x^s) - f(x^*)) + C^2 \sum_{s=0}^{k} (\alpha^s)^2.$$

$$\Longrightarrow \quad \min_{s=1,\ldots,k} f(x^s) - f(x^*) \leq \frac{\left\|x^0 - x^*\right\|^2 + C^2 \sum_{s=0}^k (\alpha^s)^2}{2\sum_{s=0}^k \alpha^s}.$$

Question: Choice of α^k ?

Subgradient Method - Choice of stepsize α^k

- **Constant stepsize:** $\alpha^k = \alpha$,

$$\min_{s=1,\ldots,k} f(x^s) - f(x^*) \leq \frac{R^2 + k\, C^2\, \alpha^2}{2\, k\, \alpha}.$$

As $k \to \infty$, $\min_{s=1,\ldots,k} f(x^s) - f(x^*) \leq \frac{C^2 \alpha}{2}$ - **no convergence !**

For desired accuracy ε set $\alpha = \frac{\varepsilon}{C^2}$, then $k = \left(\frac{RC}{\varepsilon}\right)^2$.

- **Square summable but not summable stepsizes α^k**

$$\sum_{s=0}^{\infty} (\alpha_s)^2 < \infty, \qquad \sum_{s=0}^{\infty} \alpha_s = \infty,$$

$$\Longrightarrow \quad \lim_{k \to \infty} \min_{s=1,\ldots,k} f(x^s) = f(x^*).$$

Example: $\alpha^s = \frac{1}{s^p}$ diverges for $p \leq 1$ and converges for $p > 1$. Use $\frac{1}{2} < p \leq 1$.

Convex optimization problem with equality constraint:

$$\min_{x \in \mathbb{R}^n} f(x)$$
$$\text{subject to: } Ax = b.$$

Assumptions:

- $f : \mathbb{R}^n \to \mathbb{R}$ is convex and twice differentiable,
- $A \in \mathbb{R}^{p \times n}$ with $\operatorname{rank} A = p < n$,
- optimal solution x^* exists and $p^* = \inf\{f(x) \,|\, Ax = b\}$.

Reminder: A pair (x^*, μ^*) is primal-dual optimal if and only if

$$Ax^* = b, \qquad \nabla f(x^*) + A^T \mu^* = 0, \qquad \textbf{(KKT-conditions)}.$$

Primal and **dual feasibility** equations.

How to solve an equality constrained minimization problem ?

- elimination of equality constraint - unconstrained optimization over

$$\{\hat{x} + z \,|\, z \in \ker(A)\},$$

where $A\hat{x} = b$.

- solve the unconstrained dual problem,

$$\max_{\mu \in \mathbb{R}^p} q(\mu).$$

- direct extension of Newton's method for equality constrained minimization.

Quadratic function with linear equality constraints - $P \in S_+^n$

$$\min \frac{1}{2}\langle x, Px\rangle + \langle q, x\rangle + r \ ,$$
$$\text{subject to: } Ax = b.$$

KKT conditions: $Ax^* = b, \qquad Px^* + q + A^T\mu^* = 0.$

$$\Longrightarrow \quad \textbf{KKT-system:} \quad \begin{pmatrix} P & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x^* \\ \mu^* \end{pmatrix} = \begin{pmatrix} -q \\ b \end{pmatrix}.$$

Cases:

- KKT-matrix nonsingular $\Longrightarrow$ unique primal-dual optimal pair (x^*, μ^*),
- KKT-matrix singular:
  - no solution: quadratic objective is unbounded from below,
  - a whole subspace of possible solutions.

Nonsingularity of the KKT matrix:

- P and A have no (non-trivial) common nullspace,

$$\ker(A) \cap \ker(P) = \{0\}.$$

- P is positive definite on the nullspace of A ($\ker(A)$),

$$Ax = 0, \; x \neq 0 \quad \Longrightarrow \quad \langle x, Px \rangle > 0.$$

If $P \succ 0$ the KKT-matrix is always non-singular.

Assumptions:

- initial point $x^{(0)}$ is feasible, that is $Ax^{(0)} = b$.

Newton direction - second order approximation:

$$\min_{d \in \mathbb{R}^n} \hat{f}(x+d) = f(x) + \langle \nabla f(x), d \rangle + \frac{1}{2} \langle d, Hf(x)\, d \rangle \,,$$

$$\text{subject to: } A(x+d) = b.$$

Newton step d_{NT} is the minimizer of this quadratic optimization problem:

$$\begin{pmatrix} Hf(x) & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} d_{NT} \\ w \end{pmatrix} = \begin{pmatrix} -\nabla f(x) \\ 0 \end{pmatrix}.$$

- x is feasible $\Rightarrow$ $Ad = 0$.
- Newton step lies in the null-space of A.
- $x + \alpha d$ is feasible (stepsize selection by Armijo rule)

Necessary and sufficient condition for optimality:

$$Ax^* = b, \qquad \nabla f(x^*) + A^T \mu^* = 0.$$

Linearized optimality condition:

Next point $x' = x + d$ solves linearized optimality condition:

$$A(x+d) = b, \qquad \nabla f(x+d) + A^T w \;\approx\; \nabla f(x) + Hf(x)d + A^T w = 0.$$

With $Ax = b$ (initial condition) this leads again to:

$$\begin{pmatrix} Hf(x) & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} d_{NT} \\ w \end{pmatrix} = \begin{pmatrix} -\nabla f(x) \\ 0 \end{pmatrix}.$$

Properties:

- Newton step is affine invariant, $x = Sy$ $\bar{f}(y) = f(Sy)$.

$$\nabla \bar{f}(y) = S^T \nabla f(Sy), \qquad H\bar{f}(y) = S^T H f(Ty)S,$$

feasibility: $ASy = b$

$$\text{Newton step: } S\, d_{NT}^y = d_{NT}^x.$$

- Newton decrement: $\lambda(x)^2 = \langle d_{NT}, Hf(x)d_{NT}\rangle$.

1. **Stopping criterion:** $\hat{f}(x+d) = f(x) + \langle \nabla f(x), d\rangle + \frac{1}{2}\langle d, Hf(x)d\rangle$

$$f(x) - \inf\{\hat{f}(x+v) \,|\, Ax = b\} = \frac{1}{2}\lambda^2(x).$$

$\Longrightarrow$ estimate of the difference $f(x) - p^*$.

2. **Stepsize selection:** $\frac{d}{dt} f(x + t d_{NT}) = \langle \nabla f(x), d_{NT}\rangle = -\lambda(x)^2$.

Assumption replacing $Hf(x) \succeq m\mathbb{1}$:

$$\left\| \begin{pmatrix} Hf(x) & A^T \\ A & 0 \end{pmatrix}^{-1} \right\|_2 \leq K.$$

Result: Elimination yields the same Newton step.

$\Longrightarrow$ convergence analysis of unconstrained problem applies.

- linear convergence (damped Newton phase),
- quadratic convergence (pure Newton phase).

Self-concordant Objectives - required steps bounded by:

$$\frac{20 - 8\sigma}{\sigma\beta(1 - 2\sigma)^2}\big(f(x^{(0)}) - p^*\big) + \log_2 \log_2\left(\frac{1}{\varepsilon}\right),$$

where α, β are the backtracking parameters (Armijo rule: σ is α).

Do we have to ensure feasibility of x ?

Necessary and sufficient condition for optimality:

$$Ax^* = b, \qquad \nabla f(x^*) + A^T \mu^* = 0.$$

Linearized optimality condition:

Next point $x' = x + d$ solves linearized optimality condition:

$$A(x+d) = b, \qquad \nabla f(x+d) + A^T w \;\approx\; \nabla f(x) + Hf(x)d + A^T w = 0.$$

This results in

$$\begin{pmatrix} Hf(x) & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} d_{IFNT} \\ w \end{pmatrix} = - \begin{pmatrix} \nabla f(x) \\ Ax - b \end{pmatrix}.$$

Interpretation as primal-dual Newton step

Definition 2. *In a **primal-dual** method both the primal variable x and the dual variable μ are updated.*

- **Primal residual:** $r_{\text{pri}}(x, \mu) = Ax - b,$
- **Dual residual:** $r_{\text{dual}}(x, \mu) = \nabla f(x) + A^T \mu,$
- **Residual:** $r(x, \mu) = \big(r_{\text{dual}}(x, \mu), r_{\text{pri}}(x, \mu)\big).$

  Primal-dual optimal point: $(x^*, \mu^*) \iff r(x^*, \mu^*) = 0.$

Primal-dual Newton step minimizes first-order Taylor approx. of $r(x, \mu)$:

$$r(x + d_x, \mu + d_\mu) \;\approx\; r(x, \mu) + Dr\big|_{(x,\mu)} \begin{pmatrix} d_x \\ d_\mu \end{pmatrix} = 0$$

$$\Longrightarrow \quad Dr\big|_{(x,\mu)} \begin{pmatrix} d_x \\ d_\mu \end{pmatrix} = -r(x, \mu).$$

Primal-dual Newton step:

$$Dr|_{(x,\mu)} \begin{pmatrix} d_x \\ d_\mu \end{pmatrix} = -r(x,\mu).$$

We have

$$Dr|_{(x,\mu)} = \begin{pmatrix} \nabla_x r_{\mathrm{dual}} & \nabla_\mu r_{\mathrm{dual}} \\ \nabla_x r_{\mathrm{pri}} & \nabla_\mu r_{\mathrm{pri}} \end{pmatrix} = \begin{pmatrix} Hf(x) & A^T \\ A & 0 \end{pmatrix}$$

$$\Longrightarrow \quad \begin{pmatrix} Hf(x) & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} d_x \\ d_\mu \end{pmatrix} = - \begin{pmatrix} r_{\mathrm{dual}}(x,\mu) \\ r_{\mathrm{pri}}(x,\mu) \end{pmatrix} = - \begin{pmatrix} \nabla f(x) + A^T \mu \\ Ax - b \end{pmatrix}.$$

and get with $\mu^+ = \mu + d_\mu$

$$\begin{pmatrix} Hf(x) & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} d_x \\ \mu^+ \end{pmatrix} = - \begin{pmatrix} \nabla f(x) \\ Ax - b \end{pmatrix}.$$

The primal-dual step is not necessarily a descent direction:

$$\frac{d}{dt} f(x + t d_x)\Big|_{t=0} = \langle \nabla f(x), d_x \rangle = - \big\langle Hf(x) d_x + A^T w \, , \, d_x \big\rangle$$
$$= - \langle d_x, Hf(x) d_x \rangle + \langle w, Ax - b \rangle \, .$$

where we have used, $\nabla f(x) + Hf(x) d_x + A^T w = 0$, and, $A d_x = b - Ax$.

BUT: it reduces the residual,

$$\frac{d}{dt} \left\| r(x + t d_x, \mu + t d_\mu) \right\| \Big|_{t=0} = - \left\| r(x, \mu) \right\| .$$

Towards feasibility: we have $A d_x = b - Ax$

$$r_{\mathrm{pri}}^{+} = A(x + t d_x) - b = (1-t)(Ax - b) = (1-t) r_{\mathrm{pri}} \quad \Longrightarrow \quad r_{\mathrm{pri}}^{(k)} = \Big(\prod_{i=0}^{k-1} (1 - t^{(i)}) \Big) r^{(0)}.$$

Infeasible start Newton method

Require: an initial starting point x^0 and μ^0,

repeat

  compute the primal and dual Newton step d_x^k and d_μ^k

  Backtracking Line Search:

  $t = 1$

  while $\left\| r(x + t d_x^k, \mu + t d_\mu^k) \right\| > (1 - \sigma) t \left\| r(x, \mu) \right\|$ **do**

    $t = \beta t$

  end while

  $\alpha^k = t$

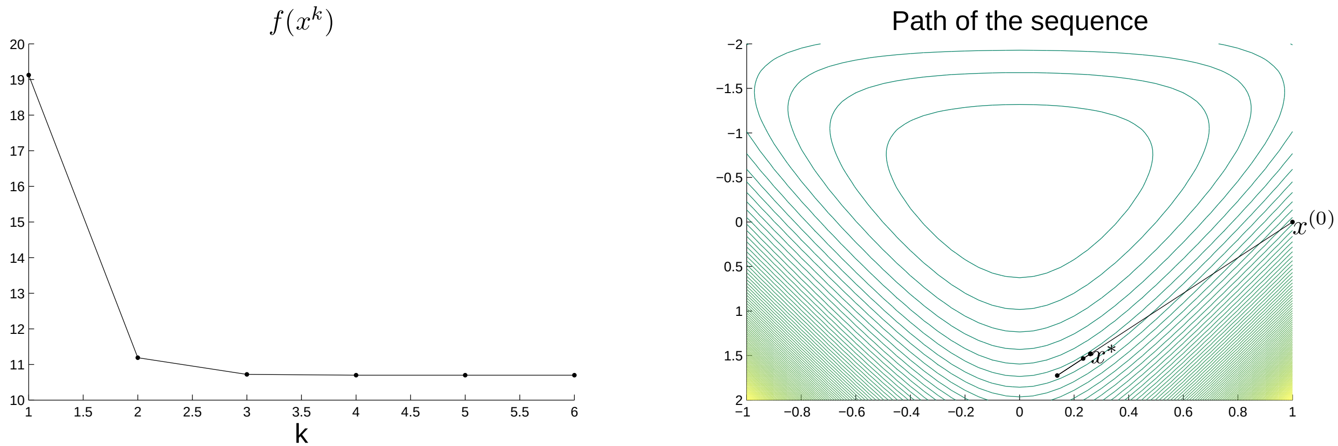
  UPDATE: $x^{k+1} = x^k + \alpha^k d_x^k$ and $\mu^{k+1} = \mu^k + \alpha^k d_\mu^k$.

until $A x^k = b$ and $\left\| r(x^k, \mu^k) \right\| \leq \varepsilon$

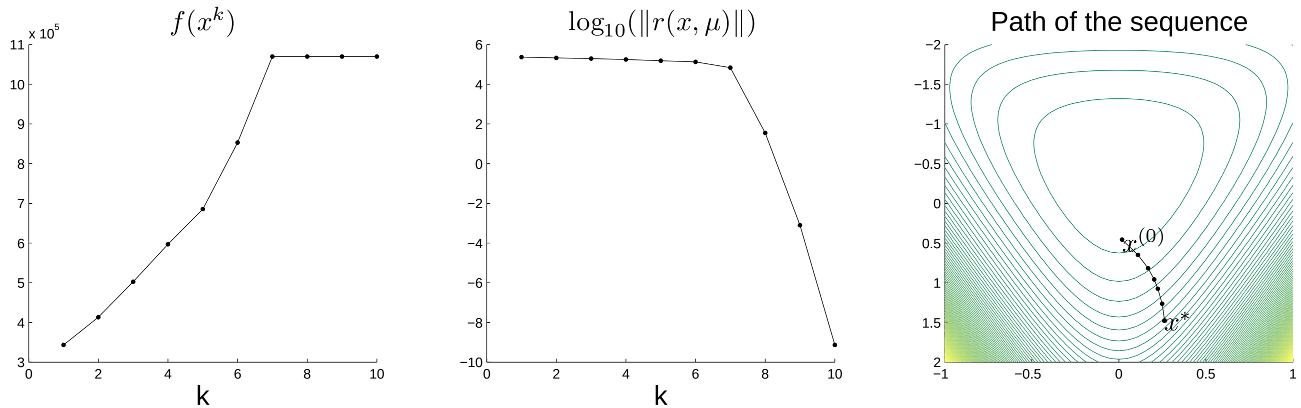
Comparison of both methods

$$\min_{x \in \mathbb{R}^2} f(x_1, x_2) = e^{x_1 + 3x_2 - 0.1} + e^{x_1 - 3x_2 - 0.1} + e^{-x_1 + 0.1}$$

subject to: $\dfrac{x_1}{2} + x_2 = 1.$



The constrained Newton method with feasible starting point.



The infeasible Newton method - note that the function value does not decrease.

Solution of the KKT system: $\begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} v \\ w \end{pmatrix} = - \begin{pmatrix} g \\ h \end{pmatrix}.$

- **Direct solution:** symmetric, but not positive definite.
  LDL^T-factorization costs $\frac{1}{3}(n+p)^3$.
- **Elimination:** $Hv + A^T w = -g \implies v = -H^{-1}[g + A^T w]$.
  and $AH^{-1}A^T w + AH^{-1}g = h \implies w = (AH^{-1}A^T)[h - AH^{-1}g]$.
  1. build $H^{-1}A^T$ and $H^{-1}g$, factorization of H and p+1 rhs
     $\Rightarrow$ cost: $f + (p+1)s$,
  2. form $S = AH^{-1}A^T$, matrix multiplication $\Rightarrow$ cost: $p^2 n$,
  3. solve $Sw = [h - AH^{-1}g]$, factorization of S $\Rightarrow$ cost $\frac{1}{3}p^3 + p^2$,
  4. solve $Hv = g + A^T w$, cost: $2np + s$.

  Total cost: $f + ps + p^2 n + \frac{1}{3}p^3$ (leading terms).

General convex optimization problem:

$$\min_{x \in \mathbb{R}^n} f(x)$$

$$\text{subject to: } g_i(x) \leq 0, \quad i = 1, \ldots, m,$$

$$Ax = b.$$

Assumptions:

- $f, g_1, \ldots, g_m$ are convex and twice differentiable,
- $A \in \mathbb{R}^{p \times n}$ with $\operatorname{rank}(A) = p$,
- there exists an optimal x^* such that $f(x^*) = p^*$,
- the problem is strictly feasible (Slater's constraint qualification holds).

$$Ax^* = b, \qquad g_i(x^*) \leq 0, \; i = 1, \ldots, m, \qquad \lambda \succeq 0,$$

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* g_i(x^*) + A^T \mu^* = 0, \qquad \lambda_i^* g_i(x^*) = 0.$$

What are interior point methods ?

- solve a sequence of equality constrained problem using Newton's method,
- solution is always strictly feasible $\Rightarrow$ lies in the **interior** of the constraint set $S = \{x \,|\, g_i(x) \leq 0, \; i = 1, \ldots, m\}$.
- basically the inequality constraints are added to the objective such that the solution is forced to be away from the boundary.

Hierarchy of convex optimization algorithms:

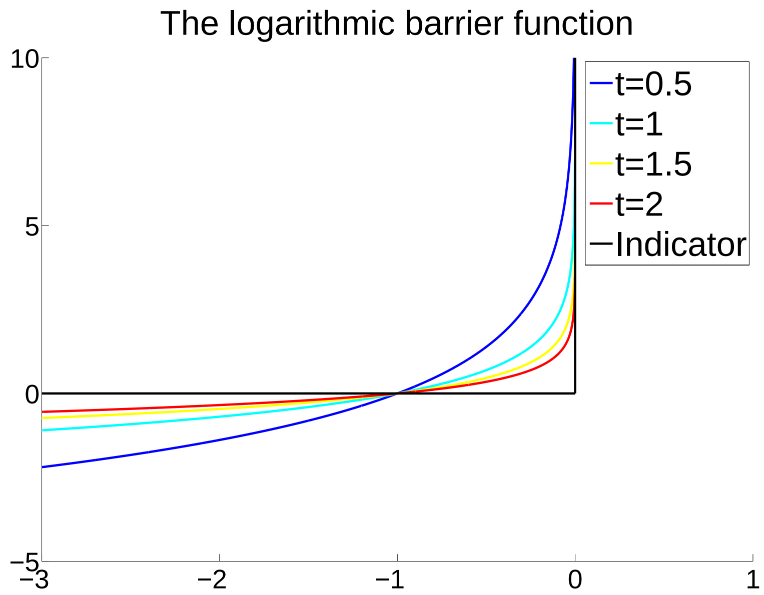
- quadratic objective with linear equality constraints $\Rightarrow$ analytic solution,
- general objective with linear eq. const. $\Rightarrow$ solve sequence of problems with quadratic objective and linear equality constraints,
- general convex optimization problem $\Rightarrow$ solve a sequence of problems with general objective and linear equality constraints.

Equivalent formulation of general convex optimization problem:

$$\min_{x\in\mathbb{R}^n} f(x) + \sum_{i=1}^{m} I_-(g_i(x))$$

subject to: $Ax = b$,

where $I_-(u) = \begin{cases} 0, & u \leq 0 \\ \infty, & u > 0. \end{cases}$.



Basic idea: approximate indicator function with a differentiable function with closed level sets.

$$\hat{I}_-(u) = -\left(\frac{1}{t}\right)\log(-u), \qquad \operatorname{dom}\hat{I} = \{x \,|\, x < 0\}.$$

where t is a parameter controlling the accuracy of the approximation.

Definition: $\phi(x) = -\sum_{i=1}^{m} \log(-g_i(x))$.

Approximate formulation:

$$\min_{x \in \mathbb{R}^n} t\, f(x) + \phi(x)$$
$$\text{subject to: } Ax = b,$$

Derivatives of ϕ:

- $\nabla \phi(x) = -\sum_{i=1}^{m} \frac{1}{g_i(x)} \nabla g_i(x)$,
- $H\phi(x) = \sum_{i=1}^{m} \frac{1}{g_i(x)^2} \nabla g_i(x) \nabla g_i(x)^T - \sum_{i=1}^{m} \frac{1}{g_i(x)} H g_i(x)$.

Definition 3. *Let $x^*(t)$ be the optimal point of the above problem, which is called* ***central point****. The* ***central path*** *is the set of points $\{x^*(t) \,|\, t > 0\}$.*

Central Path

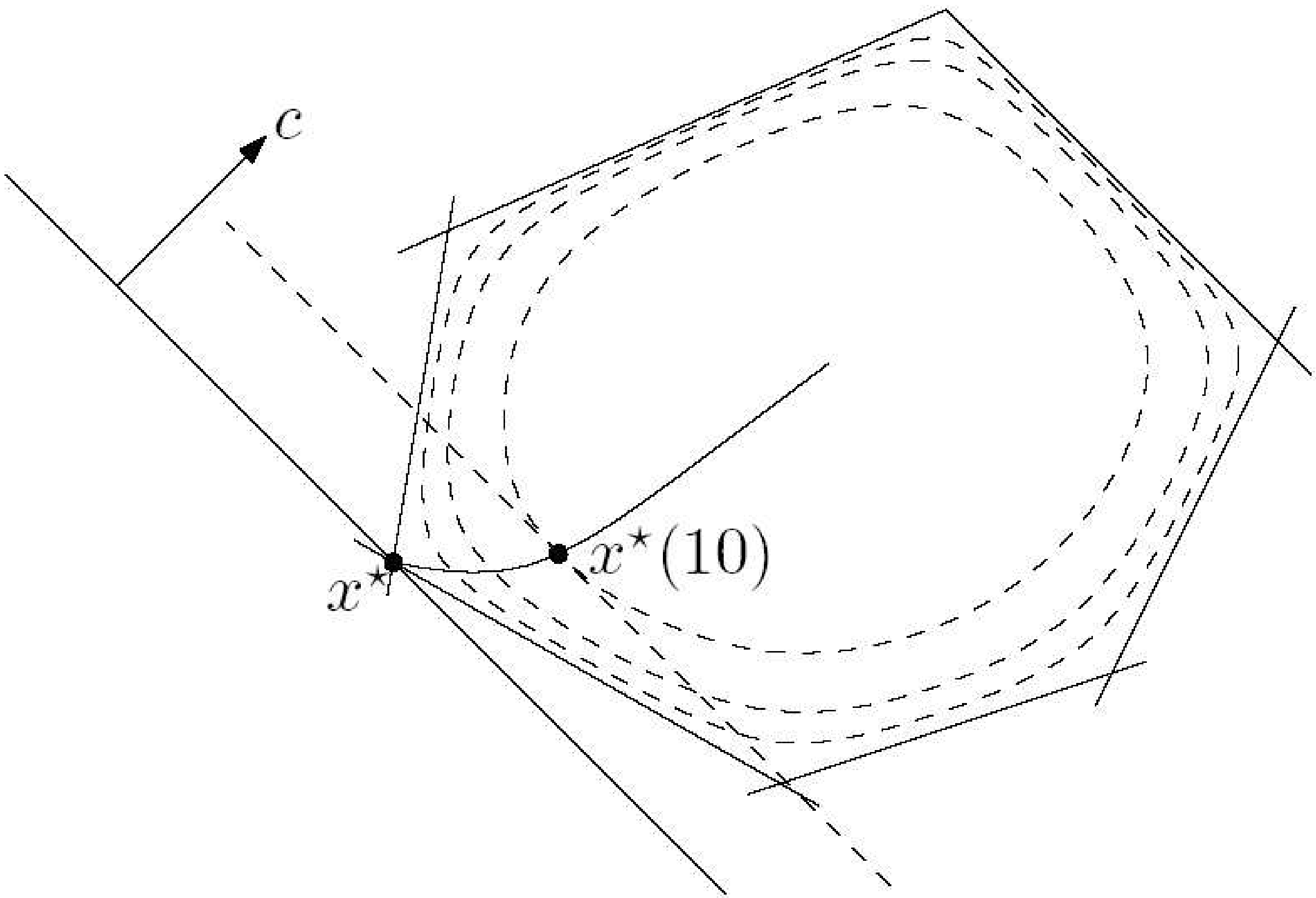


Figure 1: The central path for an LP. The dashed lines are the the contour lines of ϕ. The central path converges to x^* as $t \to \infty$.

Central points (opt. cond.): $Ax^*(t) = b, \quad g_i(x^*(t)) < 0, \; i = 1, \ldots, m,$

$$0 = t\nabla f(x^*(t)) + \nabla\phi(x^*(t)) + A^T\hat{\mu} = t\nabla f(x^*(t)) + \sum_{i=1}^{m} -\frac{1}{g_i(x^*(t))}\nabla g_i(x^*(t)) + A^T\hat{\mu}$$

Define: $\lambda_i^*(t) = -\frac{1}{tg_i(x^*(t))}$ and $\mu^*(t) = \frac{\hat{\mu}}{t}$.

$\Longrightarrow \quad (\lambda^*(t), \mu^*(t))$ are dual feasible for the original problem

and $x^*(t)$ is minimizer of Lagrangian !

- **Lagragian:** $L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \langle \mu, Ax - b\rangle$.
- **Dual function** evaluated at $(\lambda^*(t), \mu^*(t))$:

$$q(\lambda^*(t), \mu^*(t)) = f(x^*(t)) + \sum_{i=1}^{m} \lambda_i^*(t) g_i(x^*(t)) + \langle \mu^*, Ax^*(t) - b\rangle = f(x^*(t)) - \frac{m}{t}.$$

- **Weak duality:** $p^* \;\geq\; q(\lambda^*(t), \mu^*(t)) = f(x^*(t)) - \frac{m}{t}$.

$$f(x^*(t)) - p^* \leq \frac{m}{t}.$$

Interpretation via KKT conditions:

$$-\lambda_i^*(t) g_i(x^*(t)) = \frac{1}{t}.$$

$\Longrightarrow$ for t large the original KKT conditions are approximately satisfied.

Force field interpretation (no equality constraints):

Force for each constraint: $F_i(x) = -\nabla(-\log(-g_i(x))) = \dfrac{1}{g_i(x)} \nabla g_i(x)$,

generated by the potential ϕ: $F_i = -\nabla \phi(x)$.

- $F_i(x)$ is moving the particle away from the boundary,
- $F_0(x) = -t \nabla f(x)$ is moving particle towards smaller values of f.
- at the central point $x^*(t) \Longrightarrow$ forces are in equilibrium.

The barrier method (direct): set $t = \frac{\varepsilon}{m}$ then

$$f(x^*(t)) - p^* \le \varepsilon. \;\Rightarrow\; \textbf{generally does not work well.}$$

Barrier method or path-following method:

Require: strictly feasible x^0, γ, $t = t^{(0)} > 0$, tolerance $\varepsilon > 0$.

1: **repeat**

2: Centering step: compute $x^*(t)$ by minimizing

$$\min_{x \in \mathbb{R}^n} \; t\, f(x) + \phi(x)$$
$$\text{subject to: } \; Ax = b,$$

where **previous central point is taken as starting point.**

3: **UPDATE:** $x = x^*(t)$.

4: $t = \gamma t$.

5: **until** $\frac{m\gamma}{t} < \varepsilon$

The barrier method - Implementation

- **Accuracy of centering:** Exact centering (that is very accurate solution of the centering step) is not necessary but also does not harm.
- **Choice of γ:** for a small γ the last center point will be a good starting point for the new centering step, whereas for large γ the last center point is more or less an arbitrary initial point.

  trade-off between inner and outer iterations

  $\Longrightarrow$ turns out that for $3 < \gamma < 100$ the total number of Newton steps is almost constant.
- **Choice of $t^{(0)}$:** $\frac{m}{t^{(0)}} \approx f(x^{(0)}) - p^*$.
- **Infeasible Newton method:** start with $x^{(0)}$ which fulfills inequality constraints but not necessarily equality constraints. Then when feasible point is found continue with normal barrier method.