Convex Optimization and Modeling

Interior Point Methods

10th lecture, 16.06.2010

Jun.-Prof. Matthias Hein

Constrained Minimization:

- **Equality constrained minimization:**
  - Newton method with infeasible start
- **Interior point methods:**
  - barrier method
  - How to obtain a feasible starting point
  - primal-dual barrier method

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:

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

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)

Other Interpretation

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: $ASy = b$

$$\text{Newton step: } S\, d^y_{NT} = d^x_{NT}.$$

- 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:

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

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 1. *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:

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

We have

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

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

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

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

Stepsize selection for primal-dual Newton step

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

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

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

  $\alpha^k = t$

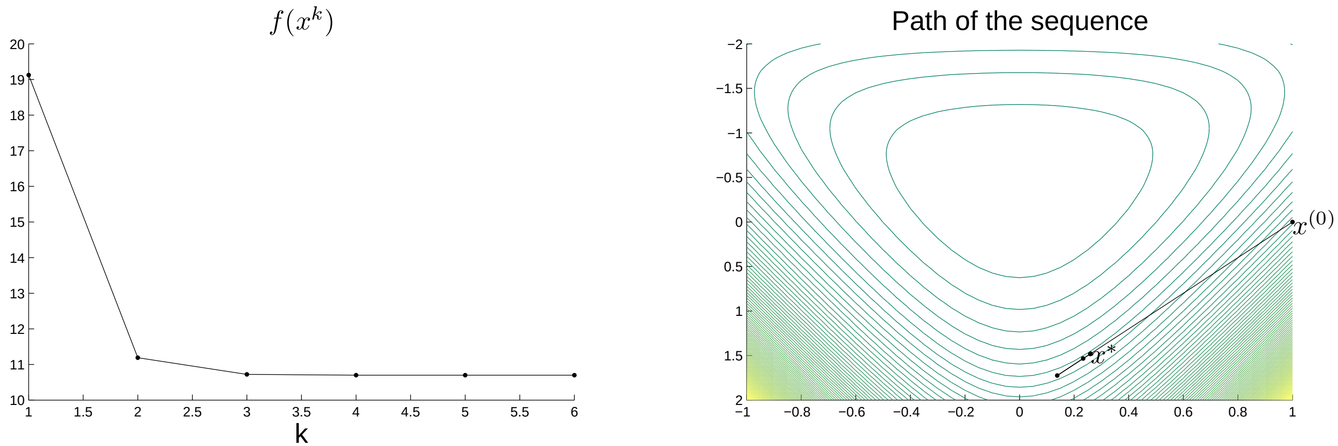
  UPDATE: $x^{k+1} = x^k + \alpha^k d_x^k$ and $\mu^{k+1} = \mu^k + \alpha^k d_\mu^k$.

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

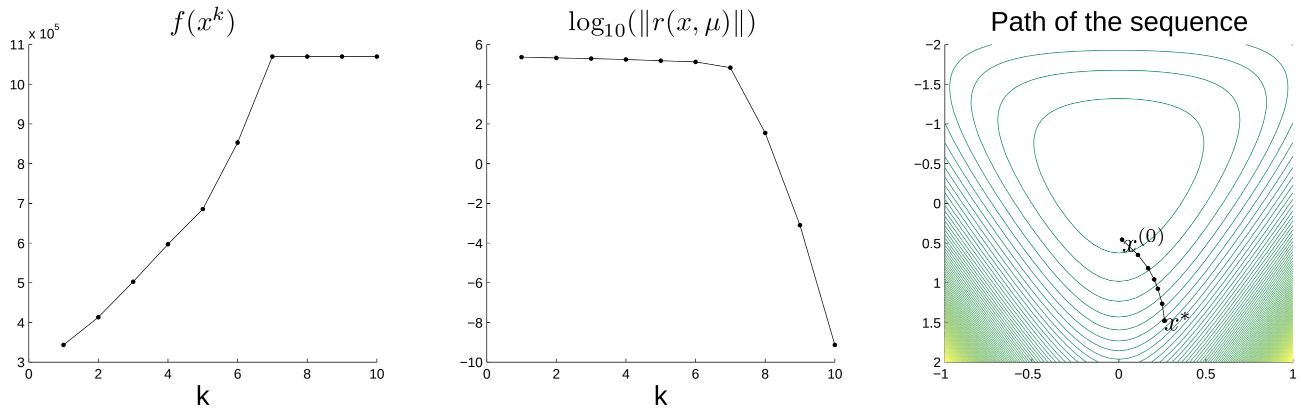
Comparison of both methods

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

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



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).

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

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

Interior point methods II

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 of S.

Hierarchy of convex optimization algorithms:

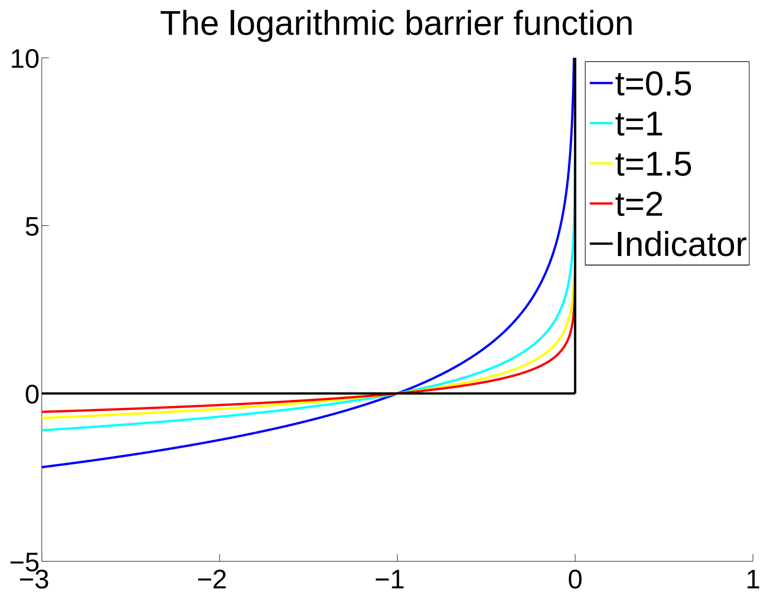
- quadratic objective with linear equality constraints $\Rightarrow$ analytic solution,
- general objective with linear eq. const. $\Rightarrow$ solve sequence of problems with quadratic objective and linear equality constraints,
- general convex optimization problem $\Rightarrow$ solve a sequence of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

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

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

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

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 2. *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

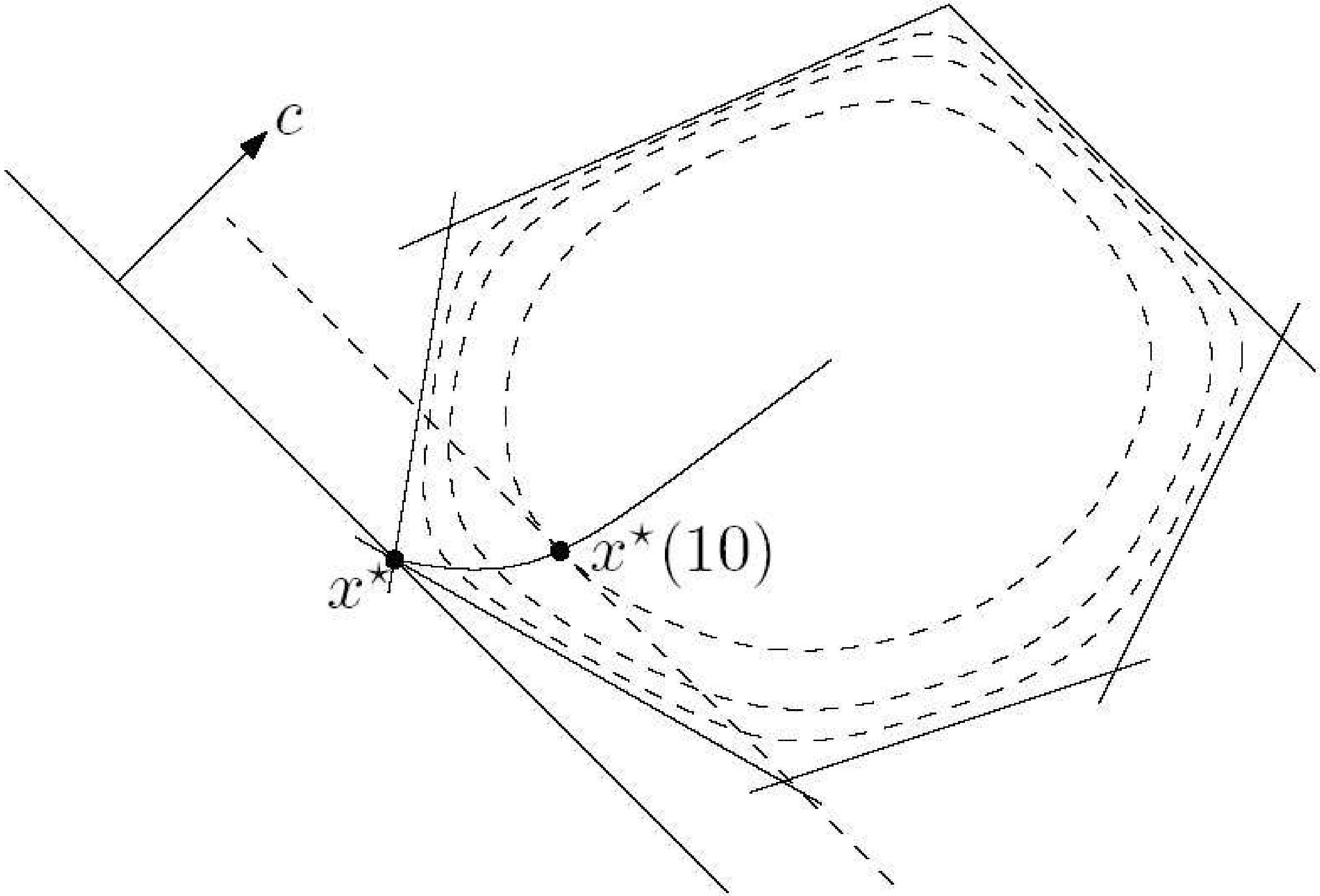


Figure 1: The central path for an LP. The dashed lines are the the contour lines of ϕ. The central path converges to x^* as $t \to \infty$.

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: $\lambda_i^*(t) = -\frac{1}{t g_i(x^*(t))}$ 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: $F_i(x) = -\nabla(-\log(-g_i(x))) = \frac{1}{g_i(x)} \nabla g_i(x)$,

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

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

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

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

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.

The full barrier method

Two step process:

- **Phase I:** find strictly feasible initial point $x^{(0)}$ or determine that no feasible point exists.
- **Phase II:** barrier method.

Strictly feasible point:

$$g_i(x) < 0, \; i = 1, \ldots, m, \qquad Ax = b.$$

Basic phase I method:

$$\begin{aligned} \min_{s \in \mathbb{R},\, x \in \mathbb{R}^n} \quad & s \\ \text{subject to:} \quad & g_i(x) \leq s, \quad i = 1, \ldots, m \\ & Ax = b. \end{aligned}$$

Choose $x^{(0)}$ such that $Ax^{(0)} = b$ and use $s^{(0)} = \max_{i=1,\ldots,m} g_i(x^{(0)})$.

Three cases:

1. $p^* < 0$: there exists a strictly feasible solution $\Longrightarrow$ as soon as $s < 0$ the optimization procedure can be stopped.
2. $p^* > 0$: there exists no feasible solution $\Longrightarrow$ one can terminate when a dual feasible point has been found which proves $p^* > 0$.
3. $p^* = 0$:
   - a minimum is attained at x^* and $s^* = 0 \Longrightarrow$ the set of inequalities is feasible but not strictly feasible.
   - the minimum is not attained $\Longrightarrow$ the inequalities are infeasible.

Problem: in practice $|f(x^{(end)}) - p^*| < \varepsilon \Longrightarrow$ with $f(x^{(end)}) \approx 0$ we get $|p^*| \leq \varepsilon$.

$\Longrightarrow \quad g_i(x) \leq -\varepsilon$ infeasible, $\quad g_i(x) \leq \varepsilon$ feasible.

Variant of phase I method:

$$
\begin{aligned}
\min_{s \in \mathbb{R}^m,\, x \in \mathbb{R}^n} \quad & \sum_{i=1}^{m} s_i \\
\text{subject to:} \quad & g_i(x) \leq s_i, \quad i = 1, \ldots, m \\
& Ax = b, \\
& s_i \geq 0, \quad i = 1, \ldots, m.
\end{aligned}
$$

Feasibility:

$$p^* = 0 \quad \Longleftrightarrow \quad \text{inequalities feasible.}$$

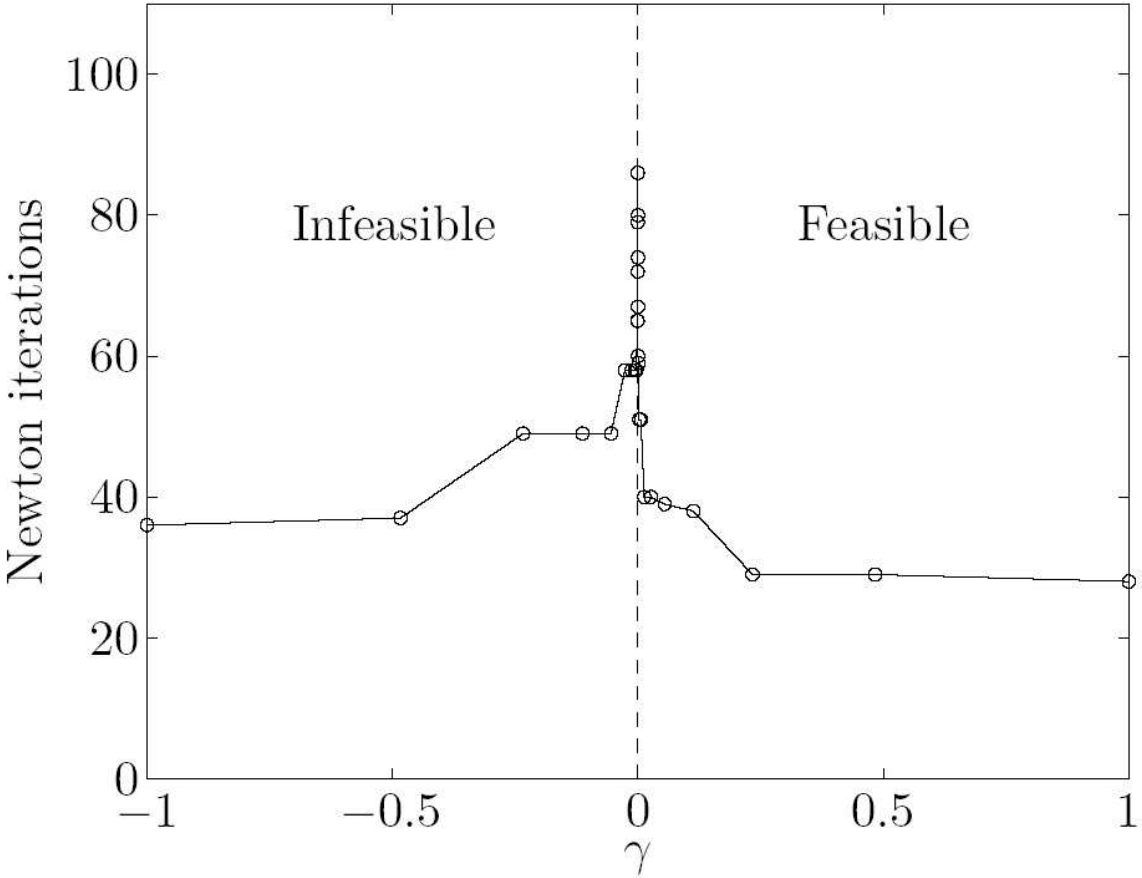
Advantage: identifies the set of feasible inequalities.

Solving for a feasible point

The less feasible the harder to identify:

$$\text{Inequalities:} \qquad Ax \preceq b + \gamma\, d.$$

where for $\gamma > 0$: feasible, $\gamma < 0$: infeasible.



Number of Newton steps versus the "grade" of feasibility .

Assumptions:

- $t\,f(x) + \phi(x)$ is self concordant for every $t \geq t^{(0)}$,
- the sublevel sets of the objective (subject to the constraints) are bounded.

Number of Newton steps for equality constrained problem:

$$N \leq \frac{f(x^{(0)}) - p^*}{\delta(\alpha, \beta)} + \log_2 \log_2 \left(\frac{1}{\varepsilon}\right),$$

where $\delta(\alpha, \beta) = \frac{\alpha\beta(1-2\alpha)^2}{20-8\alpha}$.

Number of Newton steps for one outer iteration of the barrier method:

$$N \leq \frac{m(\gamma - 1 - \log \gamma)}{\delta(\alpha, \beta)} + \log_2 \log_2 \left(\frac{1}{\varepsilon}\right),$$

Bound depends linearly on number of constraints m and roughly linear on μ.

Complexity analysis II

Total number of Newton steps for outer iterations:

$$N \leq \frac{\log\left(\frac{m}{t^{(0)}\varepsilon}\right)}{\log \gamma} \frac{m(\gamma - 1 - \log \gamma)}{\delta(\alpha, \beta)} + \log_2 \log_2 \left(\frac{1}{\varepsilon}\right),$$

$\Longrightarrow$ at least **linear convergence.**

Properties:

- independent of the dimension n of the optimization variable and the number of equality constraints.
- bound suggests $\gamma = 1 + \sqrt{m}$ - but not a good choice in practice.
- bound applies only to self-concordant functions but method still works fine for other convex functions.

Barrier for S_+^n

Generalized Inequalities: Can be integrated in the barrier method via **generalized logarithms Ψ.**

Example: positive semi-definite cone $K = S_+^n$.

$$\Psi(X) = \log \det X.$$

$\Rightarrow$ becomes infinite at the boundary of K (remember: the boundary are the matrices in S_+^n which have not full rank $\Longleftrightarrow$ positive semi-definite but not positive definite)

Properties:

- generalize primal-dual method for equality constrained minimization.
- no distinction between inner and outer iterations - at each step **primal** and **dual** variables are updated.
- in the primal-dual method, the primal and dual iterates need **not be feasible**.

Primal-Dual Interior point method:

$$\text{modified KKT equations satisfied} \quad \Longleftrightarrow \quad r_t(x, \lambda, \mu) = 0.$$

$$\begin{aligned}
r_{\text{dual}}(x, \lambda, \mu) &= \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) + A^T \mu \\
r_{\text{central},i}(x, \lambda, \mu) &= -\lambda_i \, g_i(x) - \frac{1}{t} \\
r_{\text{primal}}(x, \lambda, \mu) &= Ax - b.
\end{aligned}$$

$$r_t(x, \lambda, \mu) : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p, \qquad r_t(x, \lambda, \mu) = \begin{pmatrix} r_{\text{dual}}(x, \lambda, \mu) \\ r_{\text{central}}(x, \lambda, \mu) \\ r_{\text{primal}}(x, \lambda, \mu) \end{pmatrix}$$

Primal-Dual Interior point method III

Solving $r_t(x, \lambda, \mu) = 0$ via Newton:

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

which gives the descent directions:

$$\begin{pmatrix} Hf(x) + \sum_{i=1}^m \lambda_i Hg_i(x) & Dg(x)^T & A^T \\ -\mathrm{diag}(\lambda) Dg(x) & -\mathrm{diag}(g(x)) & 0 \\ A & 0 & 0 \end{pmatrix} \begin{pmatrix} d_x \\ d_\lambda \\ d_\mu \end{pmatrix} = - \begin{pmatrix} r_{\mathrm{dual}}(x, \lambda, \mu) \\ r_{\mathrm{central}}(x, \lambda, \mu) \\ r_{\mathrm{primal}}(x, \lambda, \mu) \end{pmatrix}$$

where

$$Dg(x) = \begin{pmatrix} \nabla g_1(x)^T \\ \vdots \\ \nabla g_m(x)^T \end{pmatrix}, \qquad Dg(x) \in \mathbb{R}^{m \times n}.$$

Primal-Dual Interior point method IV

Surrogate duality gap:

$x^{(k)}$, $\lambda^{(k)}$, $\nu^{(k)}$ need not be feasible $\Longrightarrow$ no computation of duality gap possible as in the barrier method.

Barrier method:

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

Pretend that x^k is primal feasible, λ^k, μ^k are dual feasible:

$$\text{Surrogate duality gap:} \quad -\sum_{i=1}^{m} \lambda_i^{(k)}(t) g_i(x^{(k)}(t)).$$

Associated parameter t

$$t = -\frac{m}{\left\langle \lambda^{(k)}, g(x^{(k)}) \right\rangle}.$$

Stopping condition:

$$\|r_{\text{dual}}\| \leq \varepsilon_{\text{feas}}, \quad \|r_{\text{primal}}\| \leq \varepsilon_{\text{feas}}, \quad -\left\langle \lambda^{(k)}, g(x^{(k)}) \right\rangle \leq \varepsilon.$$

Stepsize selection:

as usual but first set maximal stepsize s such that $\lambda + s d_\lambda \succ 0$ and ensure $g(x_{\text{new}}) \prec 0$ during stepsize selection.

Final algorithm:

Require: $x^{(0)}$ with $g_i(x^{(0)}) < 0$, $i = 1, \ldots, m$, $\lambda^{(0)} \succ 0$, and $\mu^{(0)}$, param: $\varepsilon_{\text{feas}}, \varepsilon, \gamma$.

1: **repeat**
2: Determine $t = -\gamma \frac{m}{\langle \lambda, g(x) \rangle}$,
3: Compute primal-dual descent direction,
4: Line search and update,
5: **until** $\|r_{\text{dual}}\| \leq \varepsilon_{\text{feas}}, \quad \|r_{\text{primal}}\| \leq \varepsilon_{\text{feas}}, \quad -\langle \lambda, g(x) \rangle \leq \varepsilon$

Non-negative Least Squares (NNLS):

$$\min_{x \in \mathbb{R}^n} \; \|\Phi\, x - Y\|_2^2$$

$$\text{subject to:} \;\; x \succeq 0,$$

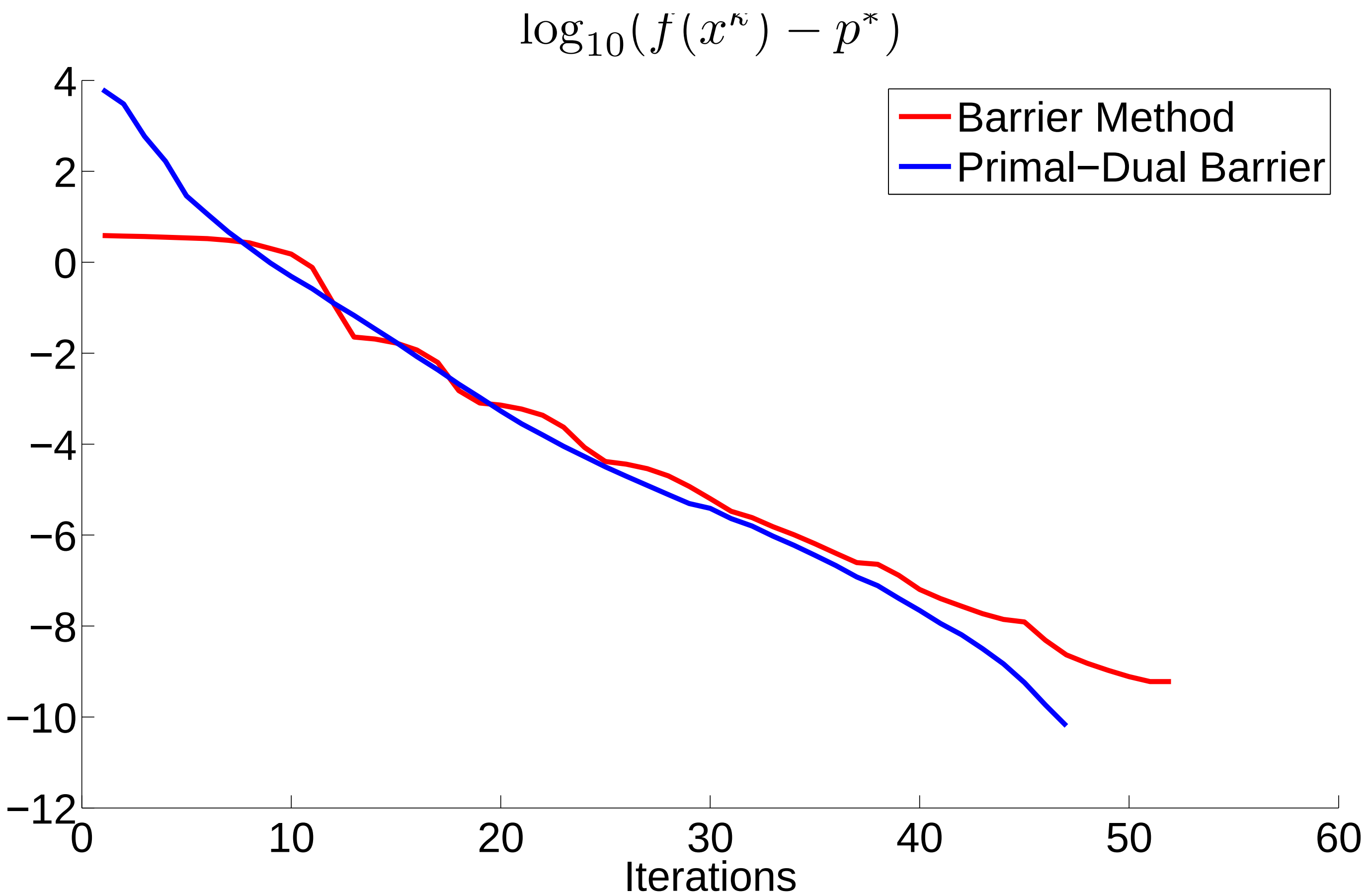
where $\Phi \in \mathbb{R}^{d \times n}$ and $Y \in \mathbb{R}^d$.

$$\begin{pmatrix} Hf(x) + \sum_{i=1}^m \lambda_i H g_i(x) & Dg(x)^T \\ -\mathrm{diag}(\lambda) Dg(x) & -\mathrm{diag}(g(x)) \end{pmatrix} = \begin{pmatrix} \Phi^T \Phi & -\mathbb{1} \\ \mathrm{diag}(\lambda) & \mathrm{diag}(x) \end{pmatrix}$$

- d_λ can be eliminated,
- Solve $\left(\Phi^T \Phi - \mathrm{diag}(\frac{\lambda}{x})\right) d_x = RHS$.

Computation time per iteration is roughly the same for the barrier and primal-dual method (dominated by the time for solving the linear system).

Comparison for NNLS



The primal-dual method is more robust against parameter changes than the barrier method (e.g. no choice of $t^{(0)}$).