COMP 9602: Convex Optimization

Interior-Point Methods

Dr. C Wu

Department of Computer Science
The University of Hong Kong

Roadmap

Theory	convex set convex function standard forms of optimization problems, quasi-convex optimization linear program, integer linear program quadratic program geometric program semidefinite program vector optimization duality
Algorithm	unconstrained optimization equality constrained optimization interior-point method localization methods subgradient method decomposition methods

Inequality constrained minimization

$$
\begin{array}{ll}
\text{minimize} & f_0(x) \\
\text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& Ax = b
\end{array}
$$

- f_i convex, twice continuously differentiable
- $A \in \mathbf{R}^{p \times n}$ with $\mathbf{rank}\, A = p$
- we assume $p^\star$ is finite and attained
- we assume problem is strictly feasible: there exists $\tilde{x}$ with

$$
\tilde{x} \in \mathbf{dom}\, f_0, \qquad f_i(\tilde{x}) < 0, \quad i = 1, \ldots, m, \qquad A\tilde{x} = b
$$

hence, strong duality holds and dual optimum is attained

Interior-point method

- Basic idea
  - move inequality constraint to objective function via indicator functions

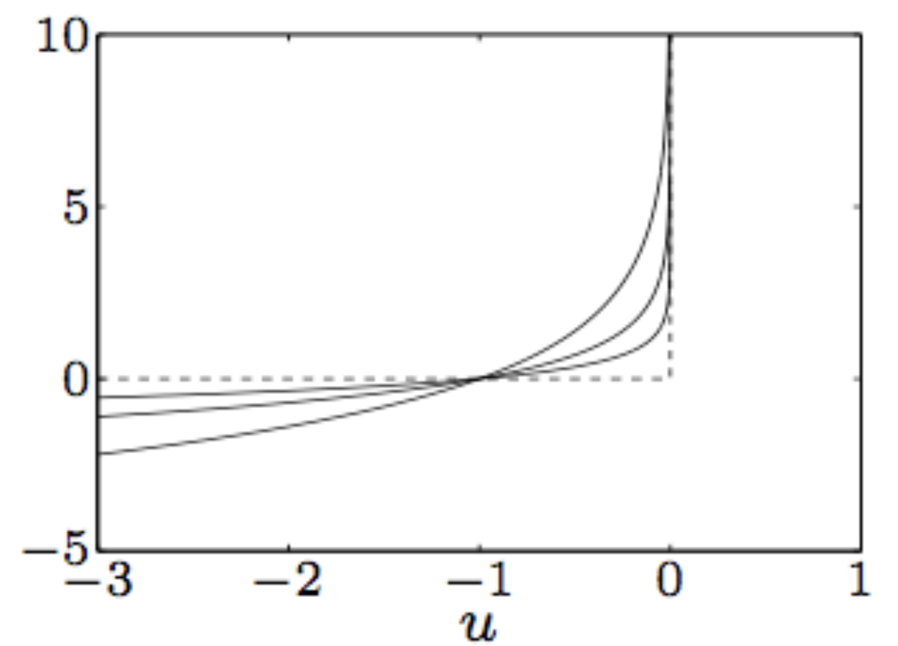
$$\begin{array}{ll} \text{minimize} & f_0(x) + \sum_{i=1}^m I_-(f_i(x)) \\ \text{subject to} & Ax = b \end{array}$$

where $I_-(u) = 0$ if $u \leq 0$, $I_-(u) = \infty$ otherwise (indicator function of $\mathbf{R}_-$)

  - approximation via logarithmic barrier: fix some t>0

$$\begin{array}{ll} \text{minimize} & f_0(x) - (1/t) \sum_{i=1}^m \log(-f_i(x)) \\ \text{subject to} & Ax = b \end{array}$$

- an equality constrained problem
- for $t > 0$, $-(1/t)\log(-u)$ is a smooth approximation of I_-
- approximation improves as $t \to \infty$



Logarithmic barrier function

$$\phi(x) = -\sum_{i=1}^{m} \log(-f_i(x)), \quad \mathbf{dom}\,\phi = \{x \mid f_1(x) < 0, \ldots, f_m(x) < 0\}$$

- convex (follows from composition rules)

- twice continuously differentiable, with **gradient and Hessian**

$$\begin{aligned}
\nabla\phi(x) &= \sum_{i=1}^{m} \frac{1}{-f_i(x)} \nabla f_i(x) \\
\nabla^2\phi(x) &= \sum_{i=1}^{m} \frac{1}{f_i(x)^2} \nabla f_i(x) \nabla f_i(x)^T + \sum_{i=1}^{m} \frac{1}{-f_i(x)} \nabla^2 f_i(x)
\end{aligned}$$

Interior-point method (cont'd)

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{array}$$

$$\Downarrow$$

$$\begin{array}{ll} \text{minimize} & f_0(x) - (1/t) \sum_{i=1}^{m} \log(-f_i(x)) \\ \text{subject to} & Ax = b \end{array}$$

with $t \to \infty$

- difficult to minimize using Newton's method (from a random starting point) when t is large

  because Hessian varies rapidly near boundary of feasibility set

- can be circumvented by solving a sequence of problems with increasing t

  starting each Newton minimization from the solution to the problem with previous t

Central path

for $t > 0$, define $x^\star(t)$ as the solution of

$$\begin{array}{ll} \text{minimize} & t f_0(x) + \phi(x) \\ \text{subject to} & Ax = b \end{array}$$

$$\boxed{\phi(x) = -\sum_{i=1}^{m} \log(-f_i(x))}$$

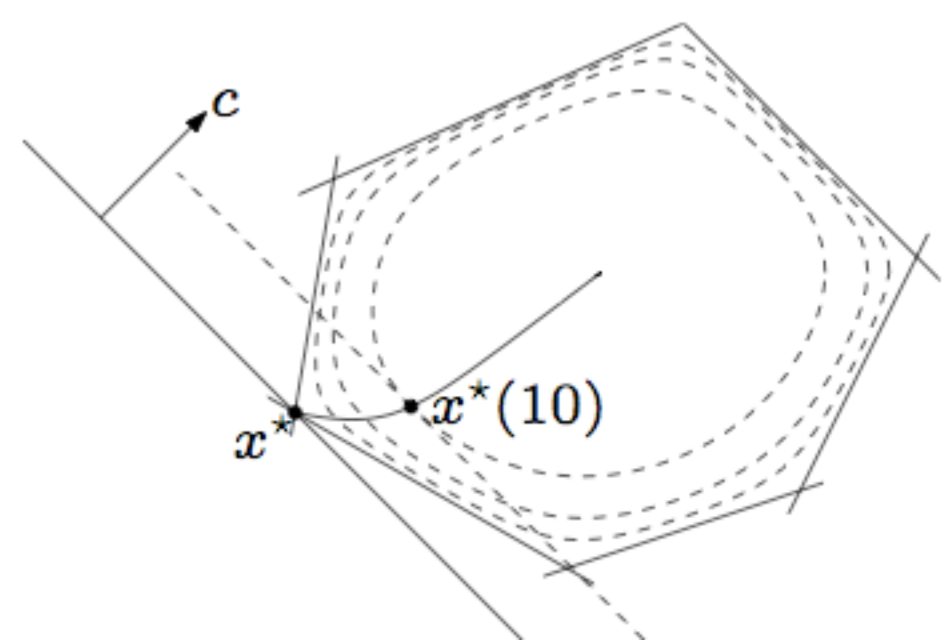
(for now, assume $x^\star(t)$ exists and is unique for each $t > 0$)

- Central paths: {x*(t) | t>0}
- x*(t): central points

example: central path for an LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \le b_i, \quad i = 1, \ldots, 6 \end{array}$$

hyperplane $c^T x = c^T x^\star(t)$ is tangent to level curve of ϕ through $x^\star(t)$



- Take the central path through interior of the feasible set

Barrier method (one interior-point method)

given strictly feasible x, $t := t^{(0)} > 0$, $\mu > 1$, tolerance $\epsilon > 0$.

repeat

1. *Centering step.* Compute $x^\star(t)$ by minimizing $tf_0 + \phi$, subject to $Ax = b$.
2. *Update.* $x := x^\star(t)$.
3. *Stopping criterion.* **quit** if $m/t < \epsilon$.
4. *Increase* t. $t := \mu t$.

- choice of μ involves a trade-off: large μ means fewer outer iterations, more inner (Newton) iterations; typical values: $\mu = 10$–20

For more practical choices of parameters, pp. 570, textbook

Dual points from central path

Every x*(t) corresponds to a dual feasible point (of the original inequality constrained problem)

$$\lambda_i^\star(t) = 1/(-tf_i(x^\star(t)) \text{ and } \nu^\star(t) = w/t$$

Verification:

x*(t) solves minimize $tf_0(x) + \phi(x)$
subject to $Ax = b$

=> $Ax^* = b \qquad f_i(x^*) < 0, i = 1, \ldots, m$

$$\exists w, \quad t\nabla f_0(x^*) + \sum_{i=1}^{m} \frac{1}{-f_i(x^*)} \nabla f_i(x^*) + A^T w = 0$$

=> $x^\star(t)$ minimizes the Lagrangian (of the original problem)

$$L(x, \lambda^\star(t), \nu^\star(t)) = f_0(x) + \sum_{i=1}^{m} \lambda_i^\star(t) f_i(x) + \nu^\star(t)^T (Ax - b)$$

at: $\lambda_i^\star(t) = 1/(-tf_i(x^\star(t))$ and $\nu^\star(t) = w/t$

dual feasible since $\lambda_i^\star(t)$ >0

Dual points from central path (cont'd)

Duality gap m/t:

$$
\begin{aligned}
f_0(x^\star(t)) \geq p^* \geq d^* \geq \quad & g(\lambda^\star(t), \nu^\star(t)) \\
= \quad & L(x^\star(t), \lambda^\star(t), \nu^\star(t)) \\
= \quad & f_0(x^\star(t)) - m/t
\end{aligned}
$$

=> $f_0(x^\star(t)) - p^\star \leq m/t$

=> $f_0(x^\star(t)) \rightarrow p^\star$ if $t \rightarrow \infty$

Interpretation via KKT conditions

$x = x^\star(t)$, $\lambda = \lambda^\star(t)$, $\nu = \nu^\star(t)$ satisfy

1. primal constraints: $f_i(x) \le 0$, $i = 1, \ldots, m$, $Ax = b$
2. dual constraints: $\lambda \succeq 0$
3. approximate complementary slackness: $-\lambda_i f_i(x) = 1/t$, $i = 1, \ldots, m$
4. gradient of Lagrangian with respect to x vanishes:

$$\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + A^T \nu = 0$$

- also shows $f_0(x^\star(t)) \to p^\star$ if $t \to \infty$

Convergence

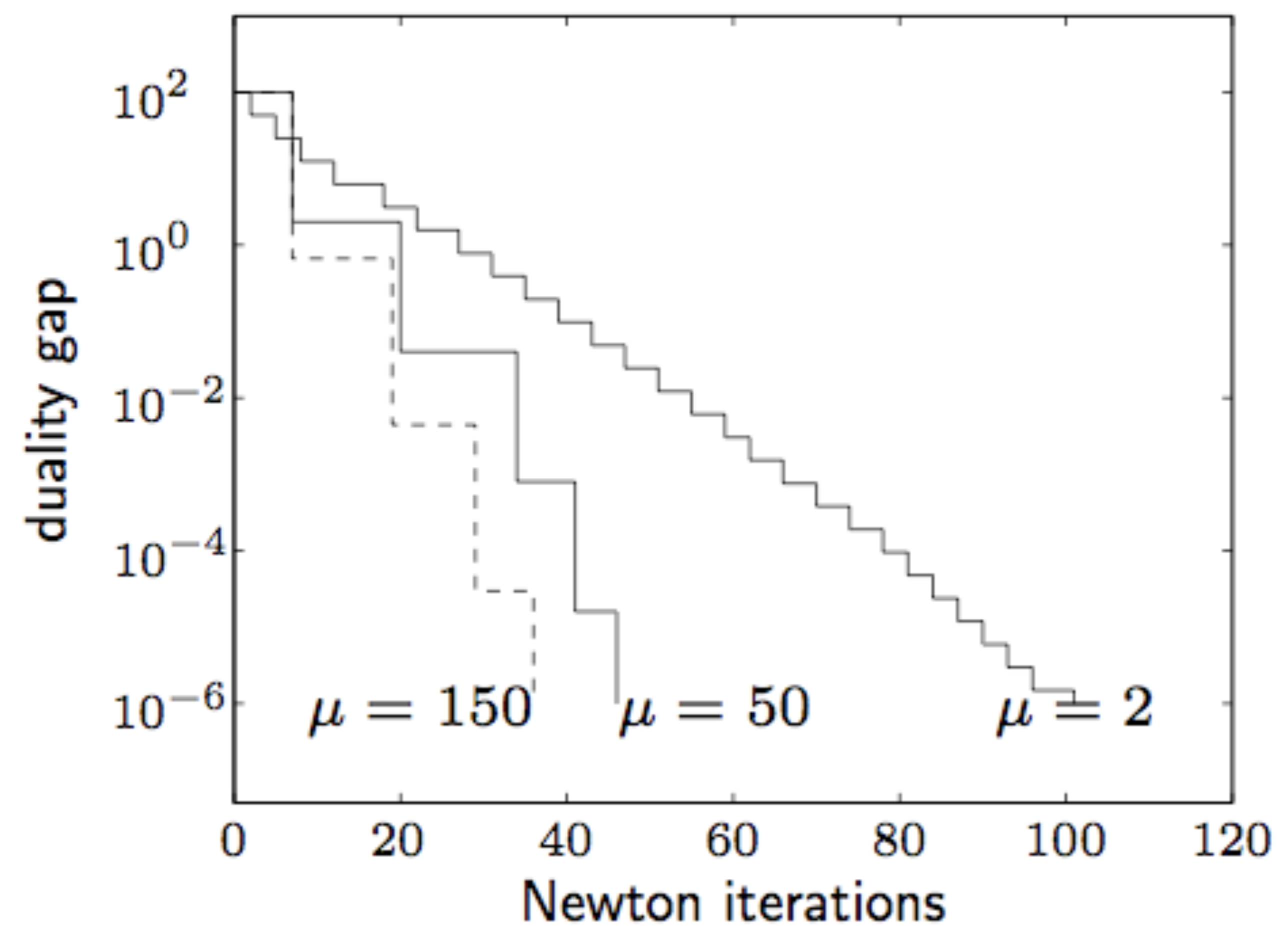
The number of steps to converge within tolerance ϵ :

$$\left\lceil \frac{\log(m/(\epsilon t^{(0)}))}{\log \mu} \right\rceil$$ outer iterations (centering steps)

plus the initial centering step (to compute $x^\star(t^{(0)})$)

Example: geometric program

$$\begin{array}{ll} \text{minimize} & \log\left(\sum_{k=1}^{5} \exp(a_{0k}^T x + b_{0k})\right) \\ \text{subject to} & \log\left(\sum_{k=1}^{5} \exp(a_{ik}^T x + b_{ik})\right) \leq 0, \\ & i = 1, \ldots, m \end{array}$$



($m = 100$ inequalities and $n = 50$ variables)

Feasibility and phase I methods

- The barrier method requires a strictly feasible starting point

feasibility problem: find x such that

$$f_i(x) \leq 0, \quad i = 1, \ldots, m, \qquad Ax = b \tag{1}$$

phase I: computes strictly feasible starting point for barrier method

Basic phase I method

$$
\begin{array}{ll}
\text{minimize (over } x,\ s) & s \\
\text{subject to} & f_i(x) \leq s, \quad i = 1, \ldots, m \\
& Ax = b
\end{array}
\qquad (2)
$$

- if x, s feasible, with $s < 0$, then x is strictly feasible for (1)
- if optimal value $\bar{p}^\star$ of (2) is positive, then problem (1) is infeasible
- if $\bar{p}^\star = 0$ and attained, then problem (1) is feasible (but not strictly);
  if $\bar{p}^\star = 0$ and not attained, then problem (1) is infeasible

Sum of infeasibilities phase I method

sum of infeasibilities phase I method

$$
\begin{array}{ll}
\text{minimize} & \mathbf{1}^T s \\
\text{subject to} & s \succeq 0, \quad f_i(x) \leq s_i, \quad i = 1, \ldots, m \\
& Ax = b
\end{array}
\tag{3}
$$

- The optimal value of (3) is zero and achieved iff problem (1) is feasible

Phase I via infeasible start Newton method

Express the original problem as:

$$\min f_0(x)$$

subject to:

$$f_i(x) \leq s, i = 1, \ldots, m$$

$$Ax = b, s = 0$$

Use an infeasible start Newton method to solve:

$$\min t^{(0)} f_0(x) - \sum_{i=1}^{m} \log(s - f_i(x))$$

subject to:

$$Ax = b, s = 0$$

initiate with any $x \in dom f_0 \cap dom f_1 \cap \ldots \cap dom f_m, s > \max_i f_i(x)$

Primal-dual interior-point method

- Update both primal and dual variables in the Newton method
  - to solve the modified KKT conditions

$$f_i(x) \leq 0,\ i = 1, \ldots, m,\ Ax = b$$

$$\lambda \succeq 0$$

$$-\lambda_i f_i(x) = 1/t,\ i = 1, \ldots, m$$

$$\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + A^T \nu = 0$$

=>

$$r_t(x, \lambda, \nu) = \begin{bmatrix} \nabla f_0(x) + Df(x)^T \lambda + A^T \nu \\ -\mathbf{diag}(\lambda) f(x) - (1/t)\mathbf{1} \\ Ax - b \end{bmatrix}$$

$$= 0$$

$$r_{\text{pri}} = Ax - b$$

$$r_{\text{dual}} = \nabla f_0(x) + Df(x)^T \lambda + A^T \nu$$

$$r_{\text{cent}} = -\mathbf{diag}(\lambda) f(x) - (1/t)\mathbf{1}$$

Primal-dual interior-point method (cont'd)

Derive the Newton step that solves $r_t(x, \lambda, \nu) = 0$ for fixed t, at (x, λ, ν):

(satisfying $f(x) \prec 0, \lambda \succ 0$)

Let $y = (x, \lambda, \nu), \triangle y = (\triangle x, \triangle \lambda, \triangle \nu)$

Linearize $r_t(y) = 0$ gives $r_t(y + \Delta y) \approx r_t(y) + Dr_t(y)\Delta y = 0$

$\Updownarrow$

$$\begin{bmatrix} \nabla^2 f_0(x) + \sum_{i=1}^m \lambda_i \nabla^2 f_i(x) & Df(x)^T & A^T \\ -\mathbf{diag}(\lambda) Df(x) & -\mathbf{diag}(f(x)) & 0 \\ A & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta \lambda \\ \Delta \nu \end{bmatrix} = - \begin{bmatrix} r_{\mathrm{dual}} \\ r_{\mathrm{cent}} \\ r_{\mathrm{pri}} \end{bmatrix}$$

Solve the above for $(\triangle x, \triangle \lambda, \triangle \nu)$

Primal-dual interior-point method (cont'd)

Algorithm 11.2 *Primal-dual interior-point method.*

given x that satisfies $f_1(x) < 0, \ldots, f_m(x) < 0$, $\lambda \succ 0$, $\mu > 1$, $\epsilon_{\text{feas}} > 0$, $\epsilon > 0$.

repeat

1. *Determine* t. Set $t := \mu m / \hat{\eta}$.
2. Compute primal-dual search direction Δy_{pd}.
3. *Line search and update.*
   Determine step length $s > 0$ and set $y := y + s \Delta y_{\text{pd}}$.

until $\|r_{\text{pri}}\|_2 \le \epsilon_{\text{feas}}$, $\|r_{\text{dual}}\|_2 \le \epsilon_{\text{feas}}$, and $\hat{\eta} \le \epsilon$.

- No distinction between inner and outer iterations
- The primal iterates are not necessarily feasible
- $\hat{\eta}(x, \lambda) = -f(x)^T \lambda$: surrogate duality gap
- Backtracking line search based on the norm of the residuals

Backtracking line search

- Based on the norm of the residuals, modified to ensure $f(x) \prec 0, \lambda \succeq 0$

denote $x^+ = x + s\Delta x_{\text{pd}}, \qquad \lambda^+ = \lambda + s\Delta\lambda_{\text{pd}}, \qquad \nu^+ = \nu + s\Delta\nu_{\text{pd}}$

  - compute the largest positive step length that does not exceed one and gives $\lambda^+ \succeq 0$:

$$\begin{aligned} s^{\max} &= \sup\{s \in [0,1] \mid \lambda + s\Delta\lambda \succeq 0\} \\ &= \min\{1, \ \min\{-\lambda_i/\Delta\lambda_i \mid \Delta\lambda_i < 0\}\} \end{aligned}$$

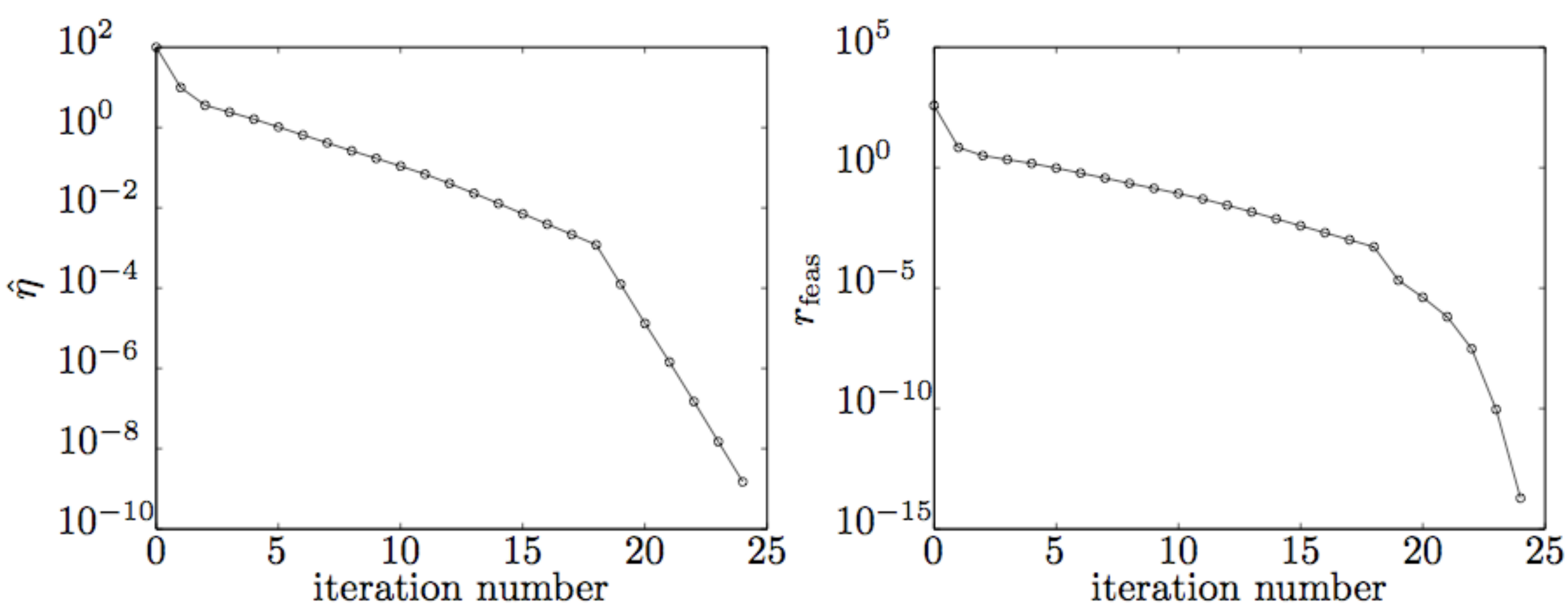
  - start with $s = 0.99 s^{\max}$, multiply s by $\beta \in (0,1)$, until $f(x^+) \prec 0$

  - continue multiplying s by β until

$$\|r_t(x^+, \lambda^+, \nu^+)\|_2 \leq (1 - \alpha s)\|r_t(x, \lambda, \nu)\|_2$$

Convergence

- More efficient than the barrier method, especially when high accuracy is needed
  - better than linear convergence
- Example
  - geometric program



$$r_{\text{feas}} = (||r_{\text{pri}}||_2^2 + ||r_{\text{dual}}||_2^2)^{\frac{1}{2}}$$

- Reference
  - Chapter 11.1-11.4, 11.7, Convex Optimization.
- Acknowledgement
  - Some materials are extracted from the slides created by Prof. Stephen Boyd for the textbook