Numerical Nonlinear Optimization Part IV

Andreas Wächter

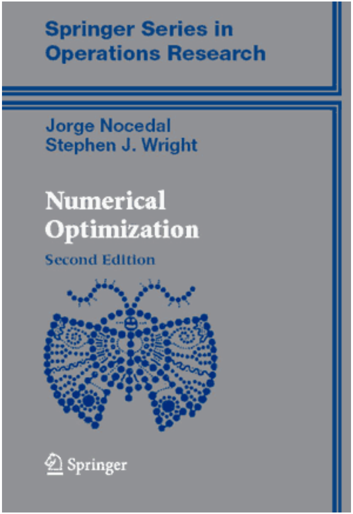
Center for Nonlinear Studies

July 13, 2020

waechter@iems.northwestern.edu

Managed by Triad National Security, LLC for the U.S. Department of Energy's NNSA

Book Recommendation



Outline

Last week:

- Optimality conditions for constrained optimization.
- Solving quadratic problems with equality constraints
- Solving quadratic problems with inequality constraints

Today:

- Sequential Quadratic Programming
- Interior-Point Methods

Equality-Constrained Nonlinear Problems

$$\boxed{\begin{aligned} \min_{x\in\mathbb{R}^n} \ & f(x) \\ \text{s.t. } & c(x) = 0 \end{aligned}} \quad \longrightarrow \quad \boxed{\begin{aligned} \nabla f(x) + \nabla c(x)\lambda &= 0 \\ c(x) &= 0 \end{aligned}}$$

- KKT conditions:
  - System of nonlinear equations in $(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^{n_E}$.
  - Apply Newton's method: Fast local convergence!
- Issues:
  - We would like to find local minima and not just any stationary point.
  - Newton's method guarantees only local convergence.
  - Need globalization technique.

Newton's Method

$$\begin{aligned} \nabla f(x) + \nabla c(x)\lambda &= 0 \\ c(x) &= 0 \end{aligned}$$

At iterate (x_k, λ_k) compute step p_k, p_k^λ from

Newton's Method

$$\begin{aligned} \nabla f(x) + \nabla c(x)\lambda &= 0 \\ c(x) &= 0 \end{aligned}$$

At iterate (x_k, λ_k) compute step p_k, p_k^λ from

$$\begin{bmatrix} H_k & \nabla c_k \\ \nabla c_k^T & 0 \end{bmatrix} \begin{pmatrix} p_k \\ p_k^\lambda \end{pmatrix} = - \begin{pmatrix} \nabla f_k + \nabla c_k \lambda_k \\ c_k \end{pmatrix}$$

$\nabla f_k := \nabla f(x_k)$ $\nabla c_k := \nabla c(x_k)$ $c_k := c(x_k)$

Newton's Method

$$\begin{aligned} \nabla f(x) + \nabla c(x)\lambda &= 0 \\ c(x) &= 0 \end{aligned}$$

At iterate (x_k, λ_k) compute step p_k, p_k^λ from

$$\begin{bmatrix} H_k & \nabla c_k \\ \nabla c_k^T & 0 \end{bmatrix} \begin{pmatrix} p_k \\ p_k^\lambda \end{pmatrix} = - \begin{pmatrix} \nabla f_k + \nabla c_k \lambda_k \\ c_k \end{pmatrix}$$

$$\nabla f_k := \nabla f(x_k) \qquad \nabla c_k := \nabla c(x_k) \qquad c_k := c(x_k)$$

$$\nabla_x \mathcal{L}(x,\lambda) := \nabla f(x) + \nabla c(x)\lambda \qquad H_k := \nabla^2_{xx}\mathcal{L}(x_k, \lambda_k)$$

Newton's Method

$$\begin{aligned} \nabla f(x) + \nabla c(x)\lambda &= 0 \\ c(x) &= 0 \end{aligned}$$

At iterate (x_k, λ_k) compute step p_k, p_k^λ from

$$\begin{bmatrix} H_k & \nabla c_k \\ \nabla c_k^T & 0 \end{bmatrix} \begin{pmatrix} p_k \\ p_k^\lambda \end{pmatrix} = -\begin{pmatrix} \nabla f_k + \nabla c_k \lambda_k \\ c_k \end{pmatrix}$$

- Update iterate $(x_{k+1}, \lambda_{k+1}) = (x_k, \lambda_k) + (p_k, p_k^\lambda)$

$$\nabla f_k := \nabla f(x_k) \qquad \nabla c_k := \nabla c(x_k) \qquad c_k := c(x_k)$$

$$\nabla_x \mathcal{L}(x, \lambda) := \nabla f(x) + \nabla c(x)\lambda \qquad H_k := \nabla^2_{xx} \mathcal{L}(x_k, \lambda_k)$$

Sequential Quadratic Programming

$$\begin{bmatrix} H_k & \nabla c_k \\ \nabla c_k^T & 0 \end{bmatrix} \begin{pmatrix} p_k \\ p_k^\lambda \end{pmatrix} = - \begin{pmatrix} \nabla f_k + \nabla c_k \lambda_k \\ c_k \end{pmatrix}$$

Sequential Quadratic Programming

$$\begin{bmatrix} H_k & \nabla c_k \\ \nabla c_k^T & 0 \end{bmatrix} \begin{pmatrix} p_k \\ \lambda_k + p_k^\lambda \end{pmatrix} = - \begin{pmatrix} \nabla f_k \cancel{+ \nabla c_k \lambda_k} \\ c_k \end{pmatrix}$$

Sequential Quadratic Programming

$$\begin{bmatrix} H_k & \nabla c_k \\ \nabla c_k^T & 0 \end{bmatrix} \begin{pmatrix} p_k \\ \tilde{\lambda}_{k+1} \end{pmatrix} = - \begin{pmatrix} \nabla f_k \\ c_k \end{pmatrix}$$

$$\tilde{\lambda}_{k+1} = \lambda_k + p_k^\lambda$$

Sequential Quadratic Programming

$$\begin{bmatrix} H_k & \nabla c_k \\ \nabla c_k^T & 0 \end{bmatrix} \begin{pmatrix} p_k \\ \tilde{\lambda}_{k+1} \end{pmatrix} = - \begin{pmatrix} \nabla f_k \\ c_k \end{pmatrix}$$

These are the optimality conditions of the QP

$$\min_{p \in \mathbb{R}^n} \ \frac{1}{2} p^T H_k p + \nabla f_k^T p + f_k$$
$$\text{s.t.} \ \nabla c_k^T p + c_k = 0$$

with multipliers $\tilde{\lambda}_{k+1} = \lambda_k + p_k^\lambda$

Sequential Quadratic Programming

$$\begin{bmatrix} H_k & \nabla c_k \\ \nabla c_k^T & 0 \end{bmatrix} \begin{pmatrix} p_k \\ \tilde{\lambda}_{k+1} \end{pmatrix} = - \begin{pmatrix} \nabla f_k \\ c_k \end{pmatrix}$$

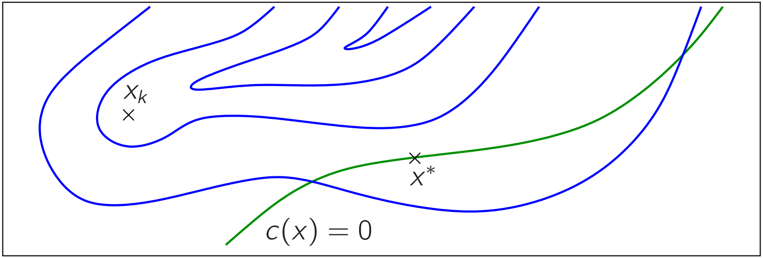
These are the optimality conditions of the QP

$$\begin{aligned} \min_{p \in \mathbb{R}^n} \ & \tfrac{1}{2} p^T H_k p + \nabla f_k^T p + f_k \\ \text{s.t.} \ & \nabla c_k^T p + c_k = 0 \end{aligned}$$

with multipliers $\tilde{\lambda}_{k+1} = \lambda_k + p_k^\lambda$

- Newton step can be interpreted as solution of a local QP model of the original problem!
- "Sequential Quadratic Programming" (SQP)

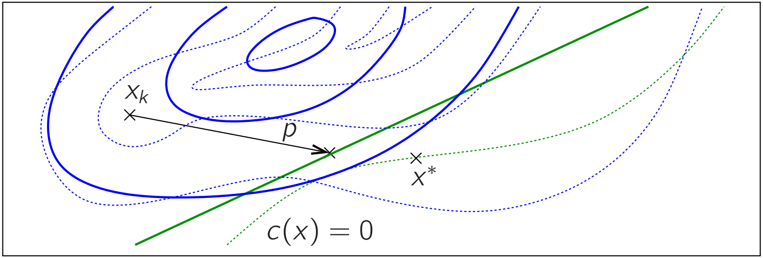
Local QP Model



Original Problem (NLP)

$$\min_{x \in \mathbb{R}^n} f(x)$$
$$\text{s.t. } c(x) = 0$$

Local QP Model



Original Problem (NLP)

$$\min_{x \in \mathbb{R}^n} f(x)$$
$$\text{s.t. } c(x) = 0$$

Local QP model (QP_k)

$$\min_{p \in \mathbb{R}^n} \frac{1}{2} p^T H_k p + \nabla f_k^T p + f_k$$
$$\text{s.t. } \nabla c_k^T p + c_k = 0$$

Regularization

$$\min_{p\in\mathbb{R}^n} \tfrac{1}{2}p^T(\nabla^2_{xx}\mathcal{L}_k \qquad)p + \nabla f_k^T p \quad \text{s.t. } \nabla c_k^T p + c_k = 0 \qquad (\mathrm{QP}_k)$$

Newton steps: $\underbrace{\begin{bmatrix} \nabla^2_{xx}\mathcal{L}_k & \nabla c_k \\ \nabla c_k^T & 0 \end{bmatrix}}_{=:K_k} \begin{pmatrix} p_k \\ \tilde{\lambda}_{k+1} \end{pmatrix} = -\begin{pmatrix} \nabla f_k \\ c_k \end{pmatrix}$

- Want ensure that QP has a minimizer.

Regularization

$$\min_{p\in\mathbb{R}^n} \tfrac{1}{2}p^T(\nabla^2_{xx}\mathcal{L}_k+\gamma I)p + \nabla f_k^T p \quad \text{s.t. } \nabla c_k^T p + c_k = 0 \qquad (\text{QP}_k)$$

Newton steps: $\underbrace{\begin{bmatrix} \nabla^2_{xx}\mathcal{L}_k+\gamma I & \nabla c_k \\ \nabla c_k^T & 0 \end{bmatrix}}_{=:K_k} \begin{pmatrix} p_k \\ \tilde{\lambda}_{k+1} \end{pmatrix} = -\begin{pmatrix} \nabla f_k \\ c_k \end{pmatrix}$

- Want ensure that QP has a minimizer.
- Choose $\gamma \geq 0$ so that K_k has inertia $(n, n_E, 0)$.
  - E.g.: Trial and error, computing inertia via factorization.
- Incentive: Avoid convergence to non-minimizers of (NLP).
- No regularization required close to 2^{nd}-order sufficient minimum.
  - At the end we have unmodified fast Newton steps.
- Other choices than $H_k = \nabla_{xx}\mathcal{L}_k$ possible, e.g., quasi-Newton.

Exact Penalty Function

- Need tool to facilitate convergence from any starting point.
- Here, we have two (usually competing) goals:

Optimality	Feasibility
$\min f(x)$	$\min \|c(x)\|$

- Combined in (non-differentiable) exact penalty function:

$$\phi_\rho(x) = f(x) + \rho \, \|c(x)\|_1$$

(penalty parameter $\rho > 0$)

Lemma

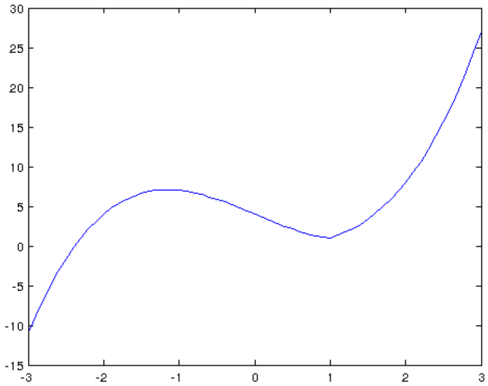
Suppose, x^* is a local minimizer of (NLP) with multipliers λ^* and LICQ holds. Then x^* is a local minimizer of ϕ_ρ if $\rho > \|\lambda^*\|_\infty$.

Penalty Function Example

$$\min_{x \in \mathbb{R}} x^3$$
$$\text{s.t. } x \geq 1$$

$x^* = 1$, $\lambda^* = 3$

$$\phi_\rho(x) = x^3 + \rho \max(1 - x, 0)$$



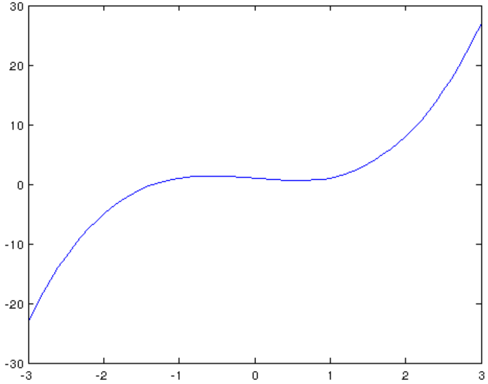
$$\rho = 4 > \|\lambda^*\|_\infty$$

Penalty Function Example

$$\min_{x \in \mathbb{R}} x^3 \quad \text{s.t. } x \geq 1$$

$x^* = 1$, $\lambda^* = 3$

$$\phi_\rho(x) = x^3 + \rho \max(1 - x, 0)$$



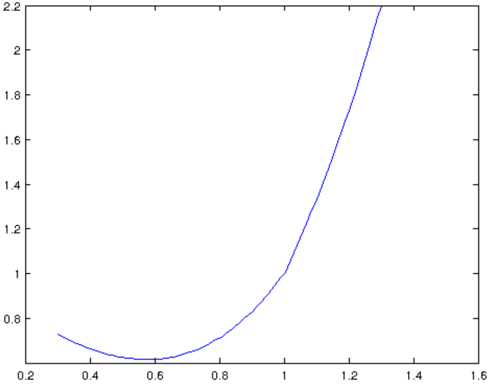
$\rho = 1 < \|\lambda^*\|_\infty$

Penalty Function Example

$$\min_{x \in \mathbb{R}} x^3$$
$$\text{s.t. } x \geq 1$$

$x^* = 1$, $\lambda^* = 3$

$$\phi_\rho(x) = x^3 + \rho \max(1 - x, 0)$$



$$\rho = 1 < \|\lambda^*\|_\infty$$

Descent Direction for Exact Penalty Function

$$\phi_\rho(x) = f(x) + \rho \, \|c(x)\|_1$$

- We can use reduction of ϕ_ρ as a measure of progress towards a local minimizer of (NLP).
- Directional derivative of ϕ_ρ at x into direction p:

$$D\phi_\rho(x;p) = \lim_{t \to 0, t > 0} \frac{\phi_\rho(x + t \cdot p) - \phi_\rho(x)}{t}$$

- p is a <u>descent direction</u> at x when $D\phi_\rho(x;p) < 0$.
- One (strong) sufficient condition for p_k being descent direction:

H_k is positive definite and $\rho > \|\tilde{\lambda}_{k+1}\|_\infty$.

Basic SQP Algorithm with Line Search

Given: Stopping tolerance $\epsilon > 0$ and parameters $\beta > 0, \eta \in (0,1)$

Choose x_0, λ_0, and $\rho_{-1} > 0$, and set $k \leftarrow 0$.
while $\|\text{KKT error}\| > \epsilon$ **do**
    Solve (QP$_k$) to get p_k and $\tilde{\lambda}_{k+1}$.
    Update penalty parameter:

$$\rho_k = \begin{cases} \rho_{k-1} & \text{if } \rho_{k-1} \geq \|\tilde{\lambda}_{k+1}\|_\infty + \beta \\ \|\tilde{\lambda}_{k+1}\|_\infty + 2\beta & \text{otherwise.} \end{cases}$$

    Find largest $\alpha_k \in \{1, \frac{1}{2}, \frac{1}{4}, \ldots\}$ with

$$\phi_{\rho_k}(x_k + \alpha_k p_k) \leq \phi_{\rho_k}(x_k) + \eta \alpha_k D\phi_{\rho_k}(x_k; p_k).$$

    Update iterate $x_{k+1} = x_k + \alpha_k p_k$ and $\lambda_{k+1} = \tilde{\lambda}_{k+1}$.
    Increase iteration counter $k \leftarrow k + 1$.
end while

Convergence Result for Basic SQP Algorithm

Assumptions

- *f and c are twice continuously differentiable.*
- *The matrices H_k are bounded and their smallest eigenvalues are uniformly bounded away from zero.*
- *The smallest singular value of ∇c_k is uniformly bounded away from zero.*

Theorem

Under these assumptions, we have

$$\lim_{k\to\infty} \left\| \begin{pmatrix} \nabla f_k + \nabla c_k \tilde{\lambda}_{k+1} \\ c_k \end{pmatrix} \right\| = 0.$$

So, each limit point of $\{x_k\}$ is a stationary point for (NLP).

Trust-Region SQP Method

$$\min_{p \in \mathbb{R}^n} \tfrac{1}{2} p^T H_k p + \nabla f_k^T p \qquad \text{s.t. } \nabla c_k^T p + c_k = 0, \quad \|p\| \le \Delta_k \qquad (\mathrm{QP}_k)$$

- No positive-definiteness requirements for H_k
- Piece-wise quadratic model of $\phi_\rho(x) = f(x) + \rho\|c(x)\|_1$:

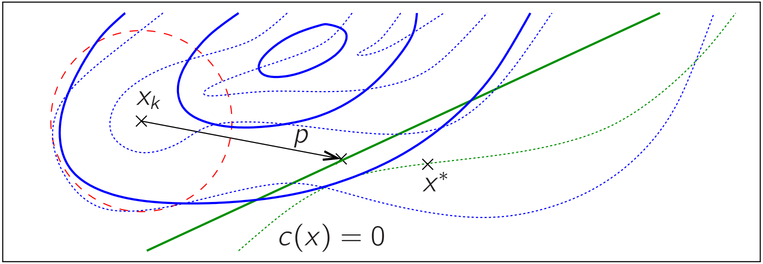
$$q_k(p) = f_k + \nabla f_k^T p + \tfrac{1}{2} p^T H_k p + \rho\|c_k + \nabla c_k^T p\|_1$$

- Step p_k is accepted if $\mathrm{ared}_k \ge \eta\, \mathrm{pred}_k$ with $(\eta \in (0,1))$

$$\mathrm{pred}_k = q_k(0) - q_k(p_k), \qquad \mathrm{ared}_k = \phi_\rho(x_k) - \phi_\rho(x_k + p_k)$$

- Otherwise, decrease Δ_k

Inconsistent QPs



- If x_k is not feasible and Δ_k small, (QP$_k$) might not be feasible
- One remedy: Penalize constraint violation in QP objective

$$
\begin{aligned}
\min_{p \in \mathbb{R}^n; t,s \in \mathbb{R}^{n_E}} \quad & \frac{1}{2} p^T H_k p + \nabla f_k^T p + \rho \sum_{j=1}^{n_E} (s_j + t_j) \\
\text{s.t.} \quad & \nabla c_k^T p + c_k = s - t \\
& \|p\| \le \Delta_k, \qquad s, t \ge 0
\end{aligned}
$$

Fletcher's Sℓ_1QP

$$\begin{aligned} \min_{p\in\mathbb{R}^n; t,s\in\mathbb{R}^{n_E}} \quad & \tfrac{1}{2}p^T H_k p + \nabla f_k^T p + \rho \sum_{j=1}^{n_E} (s_j + t_j) \\ \text{s.t.} \quad & \nabla c_k^T p + c_k = s - t \\ & \|p\| \le \Delta_k, \quad s, t \ge 0 \end{aligned}$$

is equivalent to

$$\min_{p\in\mathbb{R}^n} \; q_k(p) = f_k + \nabla f_k^T p + \tfrac{1}{2}p^T H_k p + \rho \|c_k + \nabla c_k^T p\|_1$$

- Natural algorithm for minimizing $\phi_\rho(x)$:
  - Compute steps that minimize piece-wise quadratic model q_k.
- Difficulty: Selecting sufficiently large value of ρ.
  - This motivated the invention of *filter methods*.

Maratos Effect

- Even arbitrarily close to solution, full step $\alpha = 1$ might be rejected because the non-smooth merit function ϕ_ρ increases.
- Degrades fast local convergence.
- Remedies: Second-order correction steps or "watchdog" method.

SQP For Inequality-Constrained Nonlinear Problems

$$\min_{x \in \mathbb{R}^n} \; f(x) \quad \text{s.t. } c_E(x) = 0$$

Compute p_k from local QP model

$$\min_{p \in \mathbb{R}^n} \; \tfrac{1}{2} p^T H_k p + \nabla f(x_k)^T p \quad \text{s.t. } \nabla c_E(x_k)^T p + c_E(x_k) = 0 \qquad (\mathrm{QP}_k)$$

SQP For Inequality-Constrained Nonlinear Problems

$$
\begin{aligned}
\min_{x\in\mathbb{R}^n} \;& f(x) \\
\text{s.t. } & c_E(x) = 0 \\
& c_I(x) \le 0
\end{aligned}
$$

Compute p_k from local QP model

$$
\begin{aligned}
\min_{p\in\mathbb{R}^n} \;& \tfrac{1}{2} p^T H_k p + \nabla f(x_k)^T p \\
\text{s.t. } & \nabla c_E(x_k)^T p + c_E(x_k) = 0
\end{aligned}
\qquad (\text{QP}_k)
$$

SQP For Inequality-Constrained Nonlinear Problems

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \ & f(x) \\ \text{s.t. } & c_E(x) = 0 \\ & c_I(x) \le 0 \end{aligned}$$

Compute p_k from local QP model

$$\begin{aligned} \min_{p\in\mathbb{R}^n} \ & \tfrac{1}{2} p^T H_k p + \nabla f(x_k)^T p \\ \text{s.t. } & \nabla c_E(x_k)^T p + c_E(x_k) = 0 \\ & \nabla c_I(x_k)^T p + c_I(x_k) \le 0 \end{aligned} \qquad (\text{QP}_k)$$

Local Behavior

$$\min_{x \in \mathbb{R}^n} f(x)$$
$$\text{s.t. } c_i(x) = 0 \quad i \in \mathcal{E}$$
$$c_i(x) \leq 0 \quad i \in \mathcal{I}$$

Local Behavior

$$\min_{x \in \mathbb{R}^n} f(x)$$
$$\text{s.t. } c_i(x) = 0 \quad i \in \mathcal{E}$$
$$c_i(x) \le 0 \quad i \in \mathcal{A}_*^{\text{NLP}}$$
$$c_i(x) \le 0 \quad i \in \overline{\mathcal{A}}_*^{\text{NLP}}$$

$$\mathcal{A}_*^{\text{NLP}} = \{i \in \mathcal{I} : c_i(x^*) = 0\} \qquad \overline{\mathcal{A}}_*^{\text{NLP}} = \{i \in \mathcal{I} : c_i(x^*) < 0\}$$

Local Behavior

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

$$\text{s.t. } c_i(x) = 0 \quad i \in \mathcal{E}$$

$$c_i(x) = 0 \quad i \in \mathcal{A}_*^{\text{NLP}}$$

$$\cancel{c_i(x) \le 0} \quad i \in \overline{\mathcal{A}}_*^{\text{NLP}}$$

$$\mathcal{A}_*^{\text{NLP}} = \{i \in \mathcal{I} : c_i(x^*) = 0\} \qquad \overline{\mathcal{A}}_*^{\text{NLP}} = \{i \in \mathcal{I} : c_i(x^*) < 0\}$$

- Same solution if treat active c_i as equality and ignore inactive c_i.

Local Behavior

$$\min_{x\in\mathbb{R}^n} f(x)$$
$$\text{s.t. } c_i(x) = 0 \quad i \in \mathcal{E}$$
$$c_i(x) = 0 \quad i \in \mathcal{A}_*^{\text{NLP}}$$
$$\cancel{c_i(x) \le 0} \quad i \in \overline{\mathcal{A}}_*^{\text{NLP}}$$

$$\min_{p\in\mathbb{R}^n} \tfrac{1}{2}p^T H_k p + \nabla f_k^T p$$
$$\text{s.t. } \nabla c_{k,i}^T p + c_{k,i} = 0 \quad i \in \mathcal{E}$$
$$\nabla c_{k,i}^T p + c_{k,i} = 0 \quad i \in \mathcal{A}_*^{\text{QP}_k}$$
$$\cancel{\nabla c_{k,i}^T p + c_{k,i} \le 0} \quad i \in \overline{\mathcal{A}}_*^{\text{QP}_k}$$

$$\mathcal{A}_*^{\text{NLP}} = \{i \in \mathcal{I} : c_i(x^*) = 0\} \qquad \overline{\mathcal{A}}_*^{\text{NLP}} = \{i \in \mathcal{I} : c_i(x^*) < 0\}$$

- Same solution if treat active c_i as equality and ignore inactive c_i.

Local Behavior

$$\min_{x \in \mathbb{R}^n} f(x)$$
$$\text{s.t. } c_i(x) = 0 \quad i \in \mathcal{E}$$
$$c_i(x) = 0 \quad i \in \mathcal{A}_*^{\text{NLP}}$$
$$\sout{c_i(x) \le 0} \quad i \in \overline{\mathcal{A}}_*^{\text{NLP}}$$

$$\min_{p \in \mathbb{R}^n} \tfrac{1}{2} p^T H_k p + \nabla f_k^T p$$
$$\text{s.t. } \nabla c_{k,i}^T p + c_{k,i} = 0 \quad i \in \mathcal{E}$$
$$\nabla c_{k,i}^T p + c_{k,i} = 0 \quad i \in \mathcal{A}_*^{\text{QP}_k}$$
$$\sout{\nabla c_{k,i}^T p + c_{k,i} \le 0} \quad i \in \overline{\mathcal{A}}_*^{\text{QP}_k}$$

$$\mathcal{A}_*^{\text{NLP}} = \{ i \in \mathcal{I} : c_i(x^*) = 0 \} \qquad \overline{\mathcal{A}}_*^{\text{NLP}} = \{ i \in \mathcal{I} : c_i(x^*) < 0 \}$$

- Same solution if treat active c_i as equality and ignore inactive c_i.

Lemma

Suppose x^* is a local minimizer satisfying the sufficient second-order optimality conditions, at which LICQ and strict complementarity hold. Then $\mathcal{A}_*^{\text{NLP}} = \mathcal{A}_*^{\text{QP}_k}$ for all x_k sufficiently close to x_*.

Back to Newton's Method

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \; & f(x) \\ \text{s.t.} \; & c_i(x) = 0 \quad i \in \mathcal{E} \\ & c_i(x) = 0 \quad i \in \mathcal{A}_*^{\text{NLP}} \\ & \cancel{c_i(x) \le 0} \quad i \in \overline{\mathcal{A}}_*^{\text{NLP}} \end{aligned}$$

$$\begin{aligned} \min_{p\in\mathbb{R}^n} \; & \tfrac{1}{2} p^T H_k p + \nabla f_k^T p \\ \text{s.t.} \; & \nabla c_{k,i}^T p + c_{k,i} = 0 \quad i \in \mathcal{E} \\ & \nabla c_{k,i}^T p + c_{k,i} = 0 \quad i \in \mathcal{A}_*^{\text{NLP}} \\ & \cancel{\nabla c_{k,i}^T p + c_{k,i} \le 0} \quad i \in \overline{\mathcal{A}}_*^{\text{NLP}} \end{aligned}$$

- When x_k is close to x^*, (QP$_k$) produces the same steps as SQP for equality-constrained NLP.
- We are back to Newton's method...
- Fast local convergence!

Global Convergence

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \ & f(x) \\ \text{s.t. } & c_E(x) = 0 \\ & c_I(x) \le 0 \end{aligned}$$

Globalization methods for equality constraints can be generalized.

- For example, penalty function

$$\phi_\rho(x) = f(x) + \rho\|c_E(x)\|_1 + \rho\|\max\{c_I(x), 0\}\|_1 .$$

$$\begin{aligned} \min_{p\in\mathbb{R}^n;t,s\in\mathbb{R}^{n_E};r\in\mathbb{R}^{n_I}} \ & \tfrac{1}{2}p^T H_k p + \nabla f_k^T p + \rho\sum_{j=1}^{n_E}(s_j + t_j) + \rho\sum_{j=1}^{n_I} r_j \\ \text{s.t. } & \nabla c_{E,k}^T p + c_{E,k} = s - t \\ & \nabla c_{E,k}^T p + c_{E,k} \le r \\ & \|p\| \le \Delta_k, \quad s, t, r \ge 0 \end{aligned}$$

Conclusion SQP Methods

- Solve inequality-constrained QP in each iteration
- Active-set QP solver needs to identify optimal active set.
  - Combinatorial problem.
  - Too time consuming for large-scale problems.

- Exact penalty function ("merit function") measures progress.
- Several globalization techniques:
  - Line search
  - Trust region

- Can exploit good initial guess of active set and optimal solution.
  - SQP methods can be warm-started well.
  - For example: Important for real-time optimal control.

- Many variants:
  - Several choices for Hessian matrix H_k possible, incl. quasi-Newton.
  - Decomposition-based versions can exploit problem structure.

Interior Point Methods: Barrier Problem

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \ & f(x) \\ \text{s.t. } & c(x) = 0 \\ & x \geq 0 \end{aligned}$$

Interior Point Methods: Barrier Problem

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t. } c(x) = 0, \quad \cancel{x \geq 0}$$

$$\longrightarrow$$

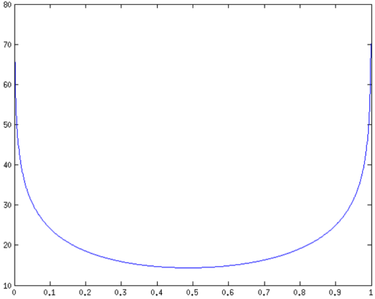
$$\min_{x \in \mathbb{R}^n} f(x) - \mu \sum_{i=1}^{n} \log(x_i) \quad \text{s.t. } c(x) = 0$$

- $\mu > 0$: Barrier parameter.

Interior Point Methods: Barrier Problem

$$\min_{x\in\mathbb{R}} x \quad \text{s.t.} \quad \sout{x \geq 0}, \quad \sout{x \leq 10}$$

$$\longrightarrow \quad \min_{x\in\mathbb{R}} x - \mu\log(x) - \mu\log(10-x)$$

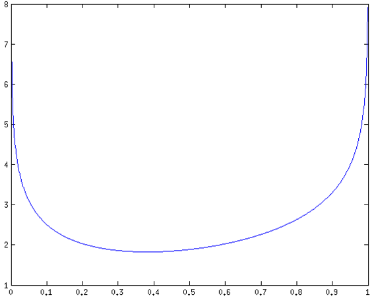


$\mu = 10$

Interior Point Methods: Barrier Problem

$$\min_{x \in \mathbb{R}} x \quad \text{s.t. } \cancel{x \geq 0}, \ \cancel{x \leq 10}$$

$$\longrightarrow \quad \min_{x \in \mathbb{R}} x - \mu \log(x) - \mu \log(10 - x)$$

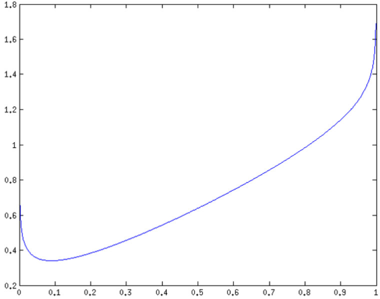


$\mu = 1$

Interior Point Methods: Barrier Problem

$$\min_{x \in \mathbb{R}} x \quad \text{s.t. } \cancel{x \geq 0},\ \cancel{x \leq 10}$$

$$\longrightarrow \quad \min_{x \in \mathbb{R}} x - \mu \log(x) - \mu \log(10 - x)$$



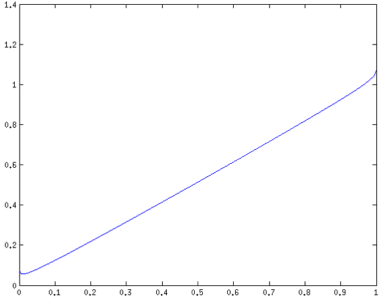
$\mu = 0.1$

Interior Point Methods: Barrier Problem

$$\min_{x \in \mathbb{R}} x \quad \text{s.t. } \cancel{x \geq 0}, \ \cancel{x \leq 10}$$

$\longrightarrow$

$$\min_{x \in \mathbb{R}} x - \mu \log(x) - \mu \log(10 - x)$$

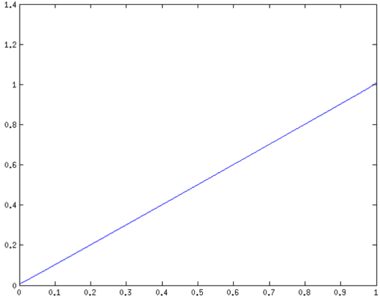


$\mu = 0.01$

Interior Point Methods: Barrier Problem

$$\min_{x\in\mathbb{R}} x \quad \text{s.t.} \quad \cancel{x \geq 0}, \quad \cancel{x \leq 10}$$

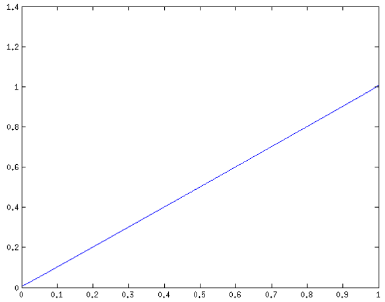
$$\longrightarrow \quad \min_{x\in\mathbb{R}} x - \mu\log(x) - \mu\log(10 - x)$$



$\mu = 0.001$

Interior Point Methods: Barrier Problem

$$\min_{x\in\mathbb{R}} x \quad \text{s.t.} \quad \sout{x \geq 0}, \quad \sout{x \leq 10} \quad \longrightarrow \quad \min_{x\in\mathbb{R}} x - \mu\log(x) - \mu\log(10-x)$$



$\mu = 0.001$

- Basic idea: $x^*(\mu) \to x^*$ as $\mu \to 0$.

Barrier Method

$$\min_{x\in\mathbb{R}^n} f(x) \quad \text{s.t. } c(x) = 0, \quad \cancel{x \geq 0}$$

$\longrightarrow$

$$\min_{x\in\mathbb{R}^n} f(x) - \mu \sum_{i=1}^{n} \log(x_i) \quad \text{s.t. } c(x) = 0 \qquad (\text{BP}_\mu)$$

Basic Interior Point Method:

Given: Final barrier parameter $\bar{\mu} > 0$.

1: Choose $x_0 \in \mathbb{R}^n$, $\mu_0 > 0$, $\epsilon_0 > 0$. Set $k \leftarrow 0$.
2: **while** $\mu_k > \bar{\mu}$ **do**
3: Starting from x_k, solve (BP_{μ_k}) to tolerance ϵ_k and obtain x_{k+1}.
4: Decrease $\mu_{k+1} < \mu_k$ and $\epsilon_{k+1} < \epsilon_k$.
5: Increase iteration counter $k \leftarrow k + 1$.
6: **end while**

Solving the Barrier Problem

$$\boxed{\begin{aligned} &\min_{x\in\mathbb{R}^n} f(x) - \mu \sum_{i=1}^{n} \log(x_i) \\ &\text{s.t. } c(x) = 0 \end{aligned}} \qquad (\mathrm{BP}_\mu)$$

- (BP_μ) is an equality constrained problem.
  - Can use SQP techniques!
- Step computation:
  - KKT system with regularization
  - Decomposition
- Globalization strategy:
  - Line search
  - Trust region

Barrier Term Considerations

$$\min_{x \in \mathbb{R}^n} f(x) - \mu \sum_{i=1}^{n} \log(x_i)$$
$$\text{s.t. } c(x) = 0$$

- For log, x_k must stay positive.
  - Fraction-to-the-boundary rule: $(\tau \in (0,1)$, e.g., $\tau = 0.99)$

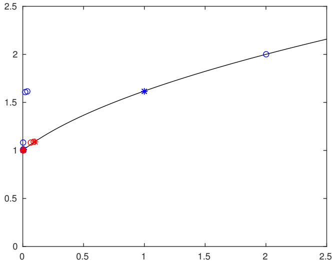
$$\alpha_k^{\max} = \arg\max \{\alpha \in (0,1] : x_k + \alpha p_k \geq (1-\tau) x_k\}$$

  - Largest step (with margin) without leaving $x > 0$.
- Ill-conditioning in SQP step computation: $(X_k = \text{diag}(x_k))$

$$\begin{bmatrix} \nabla^2_{xx}\mathcal{L}_k + \mu X_k^{-2} & \nabla c_k \\ \nabla c_k^T & 0 \end{bmatrix} \begin{pmatrix} p_k \\ \tilde{\lambda}_{k+1} \end{pmatrix} = - \begin{pmatrix} \nabla f_k - \mu X_k^{-1} e \\ c_k \end{pmatrix}$$

Example

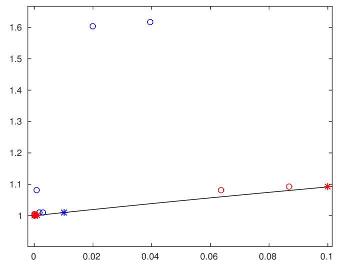
$$\min_{x\in\mathbb{R}^2} \; x_1 + (x_2 - 1)^2$$
$$\text{s.t. } x \geq 0$$




```
  k |      mu_k |       f_k | ( x_k(1),  x_k(2) ) | (  p_k(1),    p_k(2) ) |  bar_kkt |    alpha
-----------------------------------------------------------------------------------------------------
  0 | 1.00e+00 | 2.50e+00 | (2.00e+00,2.00e+00) | ( 0.00e+00, 0.00e+00) | 7.07e-01 | 0.00e+00
  1 | 1.00e+00 | 2.02e-01 | (2.00e-02,1.60e+00) | (-2.00e+00,-4.00e-01) | 2.04e+00 | 9.90e-01
  2 | 1.00e+00 | 2.31e-01 | (3.96e-02,1.62e+00) | ( 1.96e-02, 1.40e-02) | 2.41e-02 | 1.00e+00
  3 | 1.00e-01 | 6.69e-02 | (6.35e-02,1.08e+00) | ( 2.39e-02,-5.36e-01) | 5.36e-01 | 1.00e+00
  4 | 1.00e-01 | 9.09e-02 | (8.67e-02,1.09e+00) | ( 2.32e-02, 9.33e-03) | 2.50e-02 | 1.00e+00
  5 | 1.00e-02 | 4.15e-03 | (8.67e-04,1.08e+00) | (-6.65e-01,-8.18e-02) | 6.70e-01 | 1.29e-01
  6 | 1.00e-02 | 1.71e-03 | (1.66e-03,1.01e+00) | ( 7.92e-04,-7.12e-02) | 7.12e-02 | 1.00e+00
  ...
 12 | 3.16e-05 | 2.70e-05 | (2.70e-05,1.00e+00) | ( 7.45e-06, 2.95e-11) | 7.45e-06 | 1.00e+00
 13 | 1.78e-07 | 4.98e-10 | (4.81e-12,1.00e+00) | (-4.09e-03,-3.14e-05) | 4.09e-03 | 6.62e-03
Warning: Matrix is close to singular or badly scaled. Results may be inaccurate. RCOND =  1.300371e-16.
 14 | 1.78e-07 | 9.63e-12 | (9.62e-12,1.00e+00) | ( 4.81e-12,-3.12e-05) | 3.12e-05 | 1.00e+00
 15 | 1.78e-07 | 1.93e-11 | (1.92e-11,1.00e+00) | ( 9.62e-12, 9.44e-17) | 9.62e-12 | 1.00e+00
 16 | 7.50e-11 | 3.35e-11 | (3.35e-11,1.00e+00) | ( 1.43e-11,-1.78e-07) | 1.78e-07 | 1.00e+00
```

Example

$$\min_{x \in \mathbb{R}^2} \; x_1 + (x_2 - 1)^2$$
$$\text{s.t. } x \geq 0$$



k	mu_k	f_k	(x_k(1), x_k(2))	(p_k(1), p_k(2))	bar_kkt	alpha
0	1.00e+00	2.50e+00	(2.00e+00,2.00e+00)	(0.00e+00, 0.00e+00)	7.07e-01	0.00e+00
1	1.00e+00	2.02e-01	(2.00e-02,1.60e+00)	(-2.00e+00,-4.00e-01)	2.04e+00	9.90e-01
2	1.00e+00	2.31e-01	(3.96e-02,1.62e+00)	(1.96e-02, 1.40e-02)	2.41e-02	1.00e+00
3	1.00e-01	6.69e-02	(6.35e-02,1.08e+00)	(2.39e-02,-5.36e-01)	5.36e-01	1.00e+00
4	1.00e-01	9.09e-02	(8.67e-02,1.09e+00)	(2.32e-02, 9.33e-03)	2.50e-02	1.00e+00
5	1.00e-02	4.15e-03	(8.67e-04,1.08e+00)	(-6.65e-01,-8.18e-02)	6.70e-01	1.29e-01
6	1.00e-02	1.71e-03	(1.66e-03,1.01e+00)	(7.92e-04,-7.12e-02)	7.12e-02	1.00e+00
...						
12	3.16e-05	2.70e-05	(2.70e-05,1.00e+00)	(7.45e-06, 2.95e-11)	7.45e-06	1.00e+00
13	1.78e-07	4.98e-10	(4.81e-12,1.00e+00)	(-4.09e-03,-3.14e-05)	4.09e-03	6.62e-03
Warning: Matrix is close to singular or badly scaled. Results may be inaccurate. RCOND = 1.300371e-16.						
14	1.78e-07	9.63e-12	(9.62e-12,1.00e+00)	(4.81e-12,-3.12e-05)	3.12e-05	1.00e+00
15	1.78e-07	1.93e-11	(1.92e-11,1.00e+00)	(9.62e-12, 9.44e-17)	9.62e-12	1.00e+00
16	7.50e-11	3.35e-11	(3.35e-11,1.00e+00)	(1.43e-11,-1.78e-07)	1.78e-07	1.00e+00

- Several iterations per barrier parameter μ_k; slow convergence.

Primal-Dual System

$$\boxed{\begin{aligned} \min_{x\in\mathbb{R}^n} \ & f(x) \\ \text{s.t. } & c(x) = 0 \\ & x \geq 0 \end{aligned}} \quad \xrightarrow{\text{KKT}} \quad \boxed{\begin{aligned} \nabla f(x) + \nabla c(x)\lambda - z &= 0 \\ c(x) &= 0 \\ XZe &= 0 \\ x, z &\geq 0 \end{aligned}}$$

$X = \text{diag}(x)$, $Z = \text{diag}(z)$, $e = (1, \ldots, 1)^T$

Primal-Dual System

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \; & f(x) \\ \text{s.t. } & c(x) = 0 \\ & x \geq 0 \end{aligned} \quad \xrightarrow{\text{KKT}} \quad \begin{aligned} \nabla f(x) + \nabla c(x)\lambda - z &= 0 \\ c(x) &= 0 \\ XZe &= \mu e \\ (x, z &> 0) \end{aligned}$$

$X = \text{diag}(x)$, $Z = \text{diag}(z)$, $e = (1, \ldots, 1)^T$

Primal-Dual System

$$\boxed{\begin{aligned}\min_{x\in\mathbb{R}^n}\ & f(x)\\ \text{s.t. } & c(x)=0\\ & x\geq 0\end{aligned}} \quad \xrightarrow{\text{KKT}} \quad \boxed{\begin{aligned}\nabla f(x)+\nabla c(x)\lambda-z&=0\\ c(x)&=0\\ XZe&=\mu e\\ (x,z&>0)\end{aligned}}$$

Newton Steps for perturbed KKT conditions:

$$\begin{bmatrix}\nabla^2_{xx}\mathcal{L}_k & \nabla c_k & -I\\ \nabla c_k^T & 0 & 0\\ Z_k & 0 & X_k\end{bmatrix}\begin{pmatrix}p_k\\ p_k^\lambda\\ p_k^z\end{pmatrix}=-\begin{pmatrix}\nabla f_k+\nabla c_k\lambda_k-z_k\\ c_k\\ X_kZ_ke-\mu e\end{pmatrix}$$

Block elimination leads to

$$\begin{bmatrix}\nabla^2_{xx}\mathcal{L}_k+X_k^{-1}Z_k & \nabla c_k\\ \nabla c_k^T & 0\end{bmatrix}\begin{pmatrix}p_k\\ \tilde{\lambda}_{k+1}\end{pmatrix}=-\begin{pmatrix}\nabla f_k-\mu X_k^{-1}e\\ c_k\end{pmatrix}$$

Primal-Dual Steps

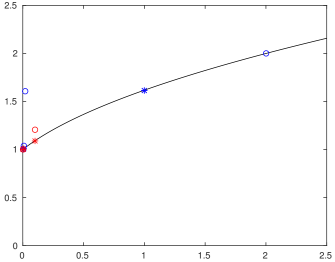
$$\begin{bmatrix} \nabla^2_{xx}\mathcal{L}_k + \Sigma_k & \nabla c_k \\ \nabla c_k^T & 0 \end{bmatrix} \begin{pmatrix} p_k \\ \tilde{\lambda}_{k+1} \end{pmatrix} = -\begin{pmatrix} \nabla f_k - \mu X_k^{-1} e \\ c_k \end{pmatrix}$$

$$p_k^z = \mu X_k^{-1} e - z_k - \Sigma_k p_k$$

- Primal and primal-dual version solve very similar linear system.
- Barrier Hessian term:
  - $\Sigma_k = X_k^{-2}$: primal
  - $\Sigma_k = X_k^{-1} Z_k$: primal-dual
- Primal-dual perspective: Homotopy method, follow “central path”.
  - Now fast local convergence.
- Primal perspective: Can use SQP-type globalization techniques.
- Ill-conditioning is benign for direct symmetric linear solvers.

Example Revisited with Primal-Dual Method

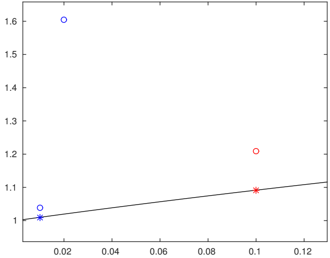
$$\min_{x \in \mathbb{R}^2} \; x_1 + (x_2 - 1)^2$$
$$\text{s.t. } x \geq 0$$



k	mu_k	f_k	(x_k(1), x_k(2))	(p_k(1), p_k(2))	bar_kkt	alpha
0	1.00e+00	2.50e+00	(2.00e+00,2.00e+00)	(0.00e+00, 0.00e+00)	7.07e-01	0.00e+00
1	1.00e+00	2.02e-01	(2.00e-02,1.60e+00)	(-2.00e+00,-4.00e-01)	4.00e-03	9.90e-01
2	1.00e-01	1.22e-01	(1.00e-01,1.21e+00)	(8.00e-02,-3.94e-01)	4.74e-16	1.00e+00
3	1.00e-02	1.07e-02	(1.00e-02,1.04e+00)	(-9.00e-02,-1.72e-01)	5.55e-17	1.00e+00
4	1.00e-03	1.00e-03	(1.00e-03,1.00e+00)	(-9.00e-03,-3.58e-02)	5.55e-17	1.00e+00
5	3.16e-05	3.16e-05	(3.16e-05,1.00e+00)	(-9.68e-04,-2.24e-03)	3.64e-17	1.00e+00
6	1.78e-07	1.78e-07	(1.78e-07,1.00e+00)	(-3.14e-05,-3.65e-05)	1.10e-16	1.00e+00
7	7.50e-11	7.50e-11	(7.50e-11,1.00e+00)	(-1.78e-07,-1.79e-07)	5.34e-17	1.00e+00

Example Revisited with Primal-Dual Method

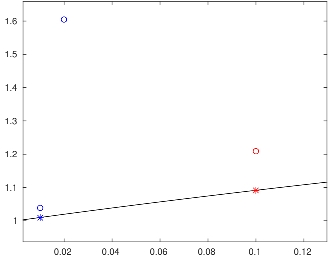
$$\min_{x \in \mathbb{R}^2} \; x_1 + (x_2 - 1)^2$$
$$\text{s.t. } x \geq 0$$



k	mu_k	f_k	(x_k(1), x_k(2))	(p_k(1), p_k(2))	bar_kkt	alpha
0	1.00e+00	2.50e+00	(2.00e+00,2.00e+00)	(0.00e+00, 0.00e+00)	7.07e-01	0.00e+00
1	1.00e+00	2.02e-01	(2.00e-02,1.60e+00)	(-2.00e+00,-4.00e-01)	4.00e-03	9.90e-01
2	1.00e-01	1.22e-01	(1.00e-01,1.21e+00)	(8.00e-02,-3.94e-01)	4.74e-16	1.00e+00
3	1.00e-02	1.07e-02	(1.00e-02,1.04e+00)	(-9.00e-02,-1.72e-01)	5.55e-17	1.00e+00
4	1.00e-03	1.00e-03	(1.00e-03,1.00e+00)	(-9.00e-03,-3.58e-02)	5.55e-17	1.00e+00
5	3.16e-05	3.16e-05	(3.16e-05,1.00e+00)	(-9.68e-04,-2.24e-03)	3.64e-17	1.00e+00
6	1.78e-07	1.78e-07	(1.78e-07,1.00e+00)	(-3.14e-05,-3.65e-05)	1.10e-16	1.00e+00
7	7.50e-11	7.50e-11	(7.50e-11,1.00e+00)	(-1.78e-07,-1.79e-07)	5.34e-17	1.00e+00

Example Revisited with Primal-Dual Method

$$\min_{x \in \mathbb{R}^2} \; x_1 + (x_2 - 1)^2$$
$$\text{s.t. } x \geq 0$$



k	mu_k	f_k	(x_k(1), x_k(2))	(p_k(1), p_k(2))	bar_kkt	alpha
0	1.00e+00	2.50e+00	(2.00e+00,2.00e+00)	(0.00e+00, 0.00e+00)	7.07e-01	0.00e+00
1	1.00e+00	2.02e-01	(2.00e-02,1.60e+00)	(-2.00e+00,-4.00e-01)	4.00e-03	9.90e-01
2	1.00e-01	1.22e-01	(1.00e-01,1.21e+00)	(8.00e-02,-3.94e-01)	4.74e-16	1.00e+00
3	1.00e-02	1.07e-02	(1.00e-02,1.04e+00)	(-9.00e-02,-1.72e-01)	5.55e-17	1.00e+00
4	1.00e-03	1.00e-03	(1.00e-03,1.00e+00)	(-9.00e-03,-3.58e-02)	5.55e-17	1.00e+00
5	3.16e-05	3.16e-05	(3.16e-05,1.00e+00)	(-9.68e-04,-2.24e-03)	3.64e-17	1.00e+00
6	1.78e-07	1.78e-07	(1.78e-07,1.00e+00)	(-3.14e-05,-3.65e-05)	1.10e-16	1.00e+00
7	7.50e-11	7.50e-11	(7.50e-11,1.00e+00)	(-1.78e-07,-1.79e-07)	5.34e-17	1.00e+00

- One iteration per μ_k, superlinear convergence.

Conclusion Interior-Point Methods

- Two perspectives:
  - Primal: Solve sequence of barrier problems with SQP method.
    - Exploit existing globalization techniques.
  - Primal-dual: Homotopy following central path.
    - Fast local convergence.
- Avoids combinatorial complexity of identifying active set.
- Can solve very large-scale problems (billions of variables).
- Difficult to warm-start:
  - Starting point has to be in the interior.
  - Need to move starting point away from boundary.
  - Cannot exploit knowledge of starting point very close to solution.
  - Needs a good number of iterations even when starting point is optimal.

Thank You!