Part 4: Interior-point methods
for inequality constrained optimization

Nick Gould (RAL)

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x) \text{ subject to } c(x) \geq 0$$

CONSTRAINED MINIMIZATION

$$\operatorname*{minimize}_{x \in \mathrm{IR}^n} \ f(x) \ \text{subject to} \ c(x) \left\{ \begin{matrix} \geq \\ = \end{matrix} \right\} 0$$

where the **objective function** $f : \mathrm{IR}^n \longrightarrow \mathrm{IR}$
and the **constraints** $c : \mathrm{IR}^n \longrightarrow \mathrm{IR}^m$

⊙ assume that f, $c \in C^1$ (sometimes C^2) and Lipschitz

⊙ often in practice this assumption violated, but not necessary

CONSTRAINTS AND MERIT FUNCTIONS

Two conflicting goals:

⊙ minimize the objective function $f(x)$

⊙ satisfy the constraints

Overcome this by minimizing a composite **merit function** $\Phi(x, p)$ for which

⊙ p are parameters

⊙ (some) minimizers of $\Phi(x, p)$ wrt x approach those of $f(x)$ subject to the constraints as p approaches some set $\mathcal{P}$

⊙ only uses **unconstrained** minimization methods

AN EXAMPLE FOR EQUALITY CONSTRAINTS

$$\underset{x \in \mathrm{I\!R}^n}{\text{minimize}} \ f(x) \ \text{ subject to } \ c(x) = 0$$

Merit function **(quadratic penalty function)**:

$$\Phi(x, \mu) = f(x) + \frac{1}{2\mu}\|c(x)\|_2^2$$

⊙ required solution as μ approaches $\{0\}$ from above

⊙ may have other useless stationary points
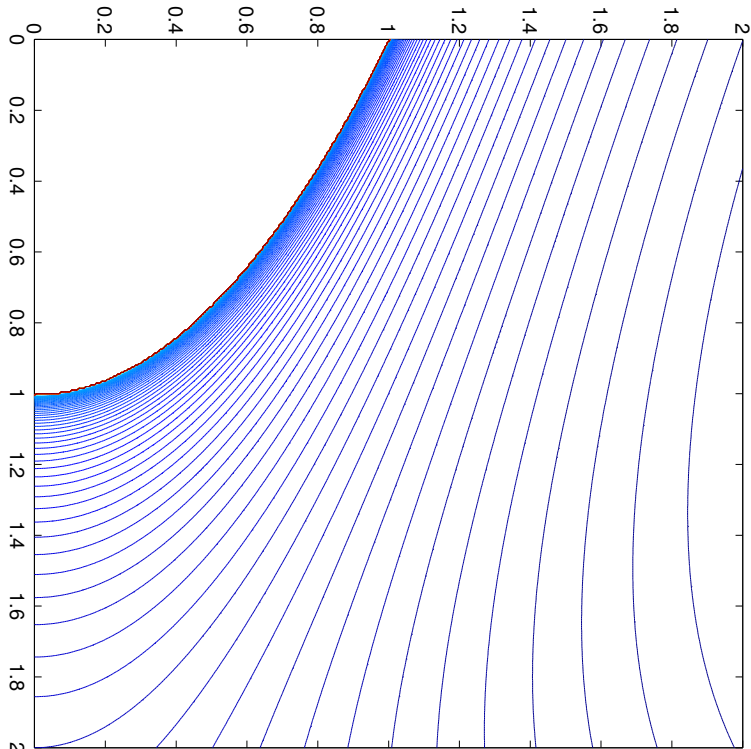
A MERIT Fⁿ FOR INEQUALITY CONSTRAINTS

$$\text{minimize}_{x \in \mathrm{I\!R}^n} \ f(x) \ \text{subject to} \ c(x) \geq 0$$

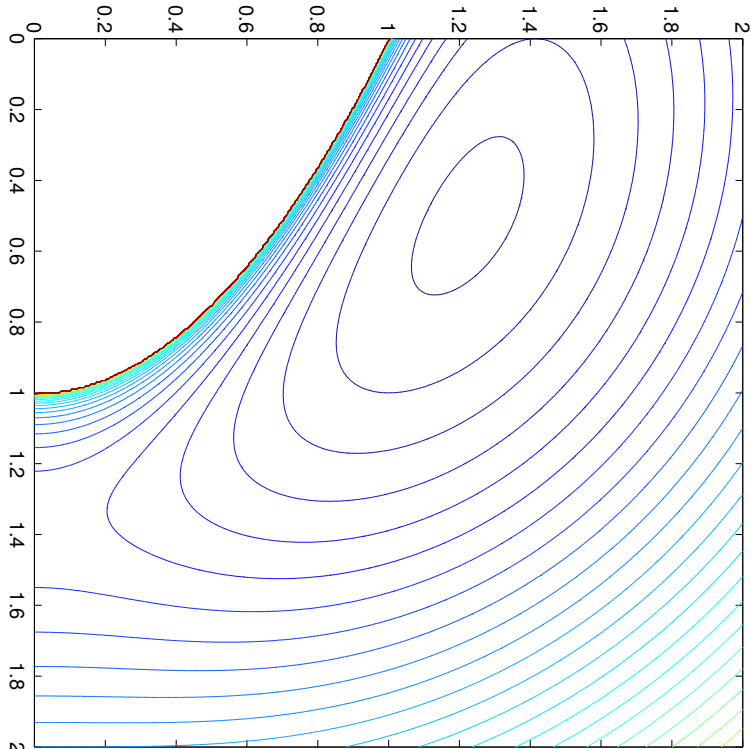Merit function **(logarithmic barrier function)**:

$$\Phi(x, \mu) = f(x) - \mu \sum_{i=1}^{m} \log c_i(x)$$

⊙ required solution as μ approaches $\{0\}$ from above

⊙ may have other useless stationary points

⊙ requires a strictly interior point to start

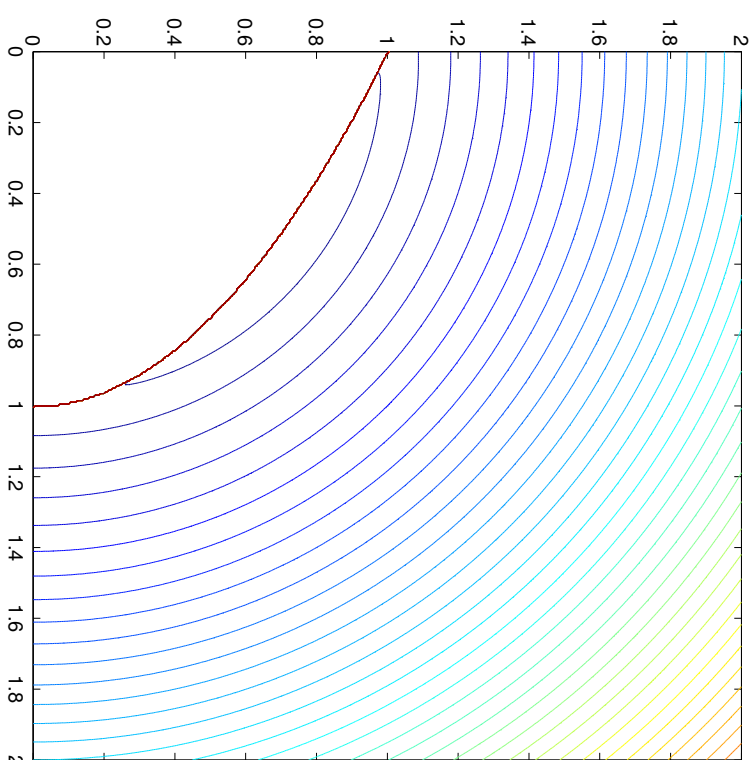⊙ consequent points are interior
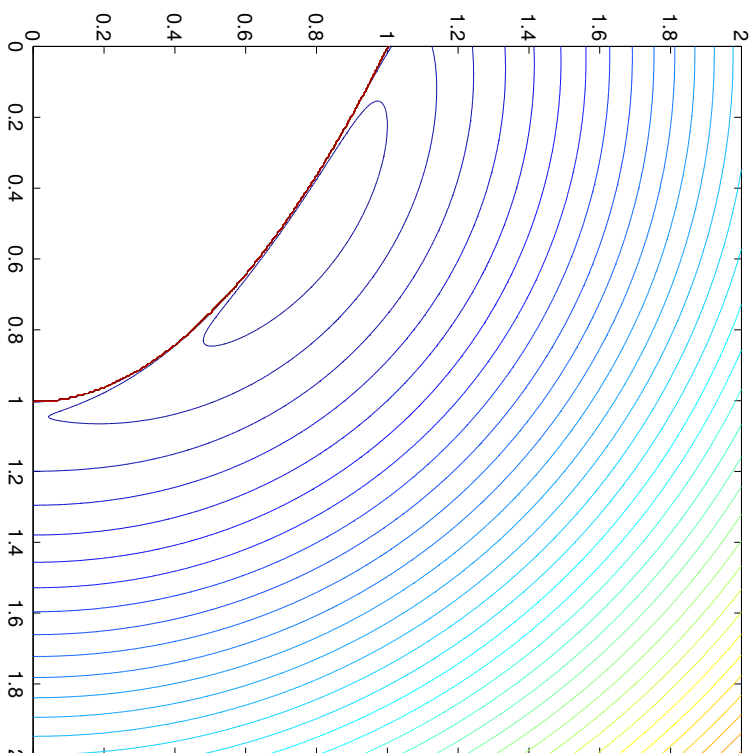
CONTOURS OF THE BARRIER FUNCTION

$\mu = 10$

$\mu = 1$

Barrier function for $\min x_1^2 + x_2^2$ subject to $x_1 + x_2^2 \geq 1$

$\mu = 0.1$

$\mu = 0.01$

Barrier function for $\min x_1^2 + x_2^2$ subject to $x_1 + x_2^2 \geq 1$

BASIC BARRIER FUNCTION ALGORITHM

Given $\mu_0 > 0$, set $k = 0$

Until "convergence" iterate:

Find x_k^s for which $c(x_k^s) > 0$

Starting from x_k^s, use an unconstrained

minimization algorithm to find an

"approximate" minimizer x_k of $\Phi(x, \mu_k)$

Compute $\mu_{k+1} > 0$ smaller than μ_k such

that $\lim_{k \to \infty} \mu_{k+1} = 0$ and increase k by 1

⊙ often choose $\mu_{k+1} = 0.1\mu_k$ or even $\mu_{k+1} = \mu_k^2$

⊙ might choose $x_{k+1}^s = x_k$

MAIN CONVERGENCE RESULT

The **active set** $\mathcal{A}(x) = \{i \mid c_i(x) = 0\}$

Theorem 4.1. Suppose that f, $c \in \mathcal{C}^2$, that $(y_k)_i \stackrel{\text{def}}{=} \mu_k / c_i(x_k)$ for $i = 1, \ldots, m$, that

$$\|\nabla_x \Phi(x_k, \mu_k)\|_2 \leq \epsilon_k$$

where ϵ_k converges to zero as $k \to \infty$, and that x_k converges to x_* for which $\{a_i(x_*)\}_{i \in \mathcal{A}(x_*)}$ are linearly independent. Then x_* satisfies the first-order necessary optimality conditions for the problem

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \; f(x) \; \text{subject to} \; c(x) \geq 0$$

and $\{y_k\}$ converge to the associated Lagrange multipliers y_*.

PROOF OF THEOREM 4.1

Let $\mathcal{M} \stackrel{\text{def}}{=} \{1, \ldots, m\}$, $\mathcal{A} \stackrel{\text{def}}{=} \{i \mid c_i(x_*) = 0\}$ and $\mathcal{I} \stackrel{\text{def}}{=} \mathcal{M} \setminus \mathcal{A}$.

Generalized inv. $A_{\mathcal{A}}^+(x) \stackrel{\text{def}}{=} \left(A_{\mathcal{A}}(x) A_{\mathcal{A}}^T(x)\right)^{-1} A_{\mathcal{A}}(x)$ bounded near x_*.

Define

$$(y_k)_i = \frac{\mu_k}{c_i(x_k)}, \ i \in \mathcal{M}, \ (y_*)_{\mathcal{A}} = A_{\mathcal{A}}^+(x_*) g(x_*) \ \text{ and } \ (y_*)_{\mathcal{I}} = 0.$$

$$\|(y_k)_{\mathcal{I}}\|_2 \leq 2\mu_k \sqrt{|\mathcal{I}|} / \min_{i \in \mathcal{I}} |c_i(x_*)| \qquad (1)$$

(if $\mathcal{I} \neq \emptyset$) for all sufficiently large k. (1) + inner-it. termination $\Longrightarrow$

$$\|g(x_k) - A_{\mathcal{A}}^T(x_k)(y_k)_{\mathcal{A}}\|_2 \leq \|g(x_k) - A^T(x_k) y_k\|_2 + \|A_{\mathcal{I}}^T(x_k)(y_k)_{\mathcal{I}}\|_2$$

$$\leq \epsilon_k \stackrel{\text{def}}{=} \epsilon_k + \mu_k \frac{2\sqrt{|\mathcal{I}|} \|A_{\mathcal{I}}\|_2}{\min_{i \in \mathcal{I}} |c_i(x_*)|} \qquad (2)$$

$$\Longrightarrow \|A_{\mathcal{A}}^+(x_k) g(x_k) - (y_k)_{\mathcal{A}}\|_2 = \|A_{\mathcal{A}}^+(x_k)(g(x_k) - A_{\mathcal{A}}^T(x_k)(y_k)_{\mathcal{A}})\|_2$$

$$\leq 2\|A_{\mathcal{A}}^+(x_*)\|_2 \bar{\epsilon}_k$$

$$\Longrightarrow \|(y_k)_{\mathcal{A}} - (y_*)_{\mathcal{A}}\|_2$$

$$\leq \|A_{\mathcal{A}}^+(x_*) g(x_*) - A_{\mathcal{A}}^+(x_k) g(x_k)\|_2 + \|A_{\mathcal{A}}^+(x_k) g(x_k) - (y_k)_{\mathcal{A}}\|_2$$

$+ \; (1) \implies \{y_k\} \longrightarrow y_*.$ Continuity of gradients $+ \; (2) \implies$

$$g(x_*) - A^T(x_*)y_* = 0$$

$c(x_k) > 0,$ defs. of y_k and $y_* + c_i(x_k)(y_k)_i = \mu_k \implies$
$c(x_*) \geq 0, \; y_* \geq 0$ and $c_i(x_*)(y_*)_i = 0.$

$\implies (x_*, y_*)$ satisfies the first-order optimality conditions.

ALGORITHMS TO MINIMIZE $\Phi(x, \mu)$

Can use

⊙ linesearch methods

◇ should use specialized linesearch to cope with singularity of log

⊙ trust-region methods

◇ need to reject points for which $c(x_k + s_k) \not> 0$

◇ (ideally) need to "shape" trust region to cope with contours of the singularity

GENERIC BARRIER NEWTON SYSTEM

Newton correction s from x for barrier function is

$$\left(H(x, y(x)) + \mu A^T(x) C^{-2}(x) A(x)\right) s = -g(x, y(x))$$

where

⊙ $C(x) = \text{diag}(c_1(x), \cdots, c_m(x))$

⊙ **Lagrange multiplier estimates**: $y(x) = \mu C^{-1}(x) e$

  where e is the vector of ones

⊙ $g(x, y(x)) = g(x) - A^T(x) y(x)$

⊙ $H(x, y(x)) = H(x) - \sum\limits_{i=1}^{m} y_i(x) H_i(x)$

Sometimes written as

$$\left(H(x, y) + A^T(x) C^{-1}(x) Y(x) A(x)\right) s = -g(x, y(x))$$
$$\text{or} \quad \left(H(x, y) + \frac{1}{\mu} A^T(x) Y^2(x) A(x)\right) s = -g(x, y(x))$$

where

⊙ $Y(x) = \text{diag}(y_1(x), \cdots, y_m(x))$

POTENTIAL DIFFICULTIES I

Ill-conditioning of the Hessian of the barrier function:

roughly speaking (non-degenerate case)

⊙ m_a eigenvalues $\approx \lambda_i(A_A^T Y^2 A_A)/\mu_k$

⊙ $n - m_a$ eigenvalues $\approx \lambda_i(N_A^T H(x_*, y_*) N_A)$

where

m_a = number of active constraints

A = active set at x_*

Y = diagonal matrix of Lagrange multipliers

N_A = orthogonal basis for null-space of A_A

$\implies$ condition number of $\nabla_{xx}\Phi(x_k, \mu_k) = O(1/\mu_k)$

$\implies$ may not be able to find minimizer easily

POTENTIAL DIFFICULTIES II

Value $x_{k+1}^s = x_k$ is a poor starting point: Suppose

$$0 \approx \nabla_x \Phi(x_k, \mu_k) = g(x_k) - \mu_k A^T(x_k) C^{-1}(x_k) e$$
$$\approx g(x_k) - \mu_k A_A^T(x_k) C_A^{-1}(x_k) e$$

Roughly speaking (non-degenerate case) Newton correction satisfies

$$\mu_{k+1} A_A^T(x_k) C_A^{-2}(x_k) A_A(x_k) s \approx (\mu_{k+1} - \mu_k) A_A^T(x_k) C_A^{-1}(x_k) e$$

$\implies$ (full rank)

$$A_A(x_k) s \approx \left(1 - \frac{\mu_k}{\mu_{k+1}}\right) c_A(x_k)$$

$\implies$ (Taylor expansion)

$$c_A(x_k + s) \approx c_A(x_k) + A_A(x_k) s \approx \left(2 - \frac{\mu_k}{\mu_{k+1}}\right) c_A(x_k) < 0$$

if $\mu_{k+1} < \frac{1}{2}\mu_k \implies$ Newton step infeasible $\implies$ slow convergence

PERTURBED OPTIMALITY CONDITIONS

First order optimality conditions for

$$\underset{x \in \mathrm{I\!R}^n}{\text{minimize}} \ \ f(x) \ \ \text{subject to} \ \ c(x) \geq 0$$

are:

$$g(x) - A^T(x)y = 0 \qquad \text{dual feasibility}$$
$$C(x)y = 0 \qquad \text{complementary slackness}$$
$$c(x) \geq 0 \ \text{and} \ y \geq 0$$

Consider the "perturbed" problem

$$g(x) - A^T(x)y = 0 \qquad \text{dual feasibility}$$
$$C(x)y = \mu e \qquad \textbf{\textcolor{red}{perturbed}} \ \text{comp. slkns.}$$
$$c(x) > 0 \ \text{and} \ y > 0$$

where $\mu > 0$

PRIMAL-DUAL PATH-FOLLOWING METHODS

Track roots of

$$g(x) - A^T(x)y = 0 \text{ and } C(x)y - \mu e = 0$$

as $0 < \mu \to 0$, while maintaining $c(x) > 0$ and $y > 0$

$\odot$ nonlinear system $\implies$ use Newton's method

Newton correction (s, u) to (x, y) satisfies

$$\begin{pmatrix} H(x, y) & -A^T(x) \\ YA(x) & C(x) \end{pmatrix} \begin{pmatrix} s \\ u \end{pmatrix} = - \begin{pmatrix} g(x) - A^T(x)y \\ C(x)y - \mu e \end{pmatrix}$$

Eliminate $u \implies$

$$\left(H(x, y) + A^T(x)C^{-1}(x)YA(x)\right)s = -\left(g(x) - \mu A^T(x)C^{-1}e\right)$$

c.f. Newton method for barrier minimization!

PRIMAL VS. PRIMAL-DUAL

Primal:

$$(H(x, y(x)) + A^T(x)C^{-1}(x)Y(x)A(x)) s^P = -g(x, y(x))$$

Primal-dual:

$$(H(x, y) + A^T(x)C^{-1}(x)YA(x)) s^{PD} = -g(x, y(x))$$

where

$$y(x) = \mu C^{-1}(x)e$$

What is the difference?

⊙ freedom to choose y in $H(x, y) + A^T(x)C^{-1}(x)YA(x)$ for primal-dual . . . vital

POTENTIAL DIFFICULTY II ...REVISITED

Value $x_{k+1}^s = x_k$ can be a good starting point:

$\odot$ primal method has to choose $y = y(x_k^s) = \mu_{k+1} C^{-1}(x_k) e$

$\diamond$ factor μ_{k+1} / μ_k too small for a good Lagrange multiplier estimate

$\odot$ primal-dual method can choose $y = \mu_k C^{-1}(x_k) e \rightarrow y_*$

Advantage: roughly (non-degenerate case) correction s^{PD} satisfies

$$\mu_k A_{\mathcal{A}}^T(x_k) C_{\mathcal{A}}^{-2}(x_k) A_{\mathcal{A}}(x_k) s^{\text{PD}} \approx (\mu_{k+1} - \mu_k) A_{\mathcal{A}}^T(x_k) C_{\mathcal{A}}^{-1}(x_k) e$$

$\implies$ (full rank)

$$A_{\mathcal{A}}(x_k) s^{\text{PD}} \approx \left(\frac{\mu_{k+1}}{\mu_k} - 1 \right) c_{\mathcal{A}}(x_k)$$

$\implies$ (Taylor expansion)

$$c_{\mathcal{A}}(x_k + s^{\text{PD}}) \approx c_{\mathcal{A}}(x_k) + A_{\mathcal{A}}(x_k) s^{\text{PD}} \approx \frac{\mu_{k+1}}{\mu_k} c_{\mathcal{A}}(x_k) > 0$$

$\implies$ Newton step allowed $\implies$ fast convergence

PRIMAL-DUAL BARRIER METHODS

Choose a search direction s for $\Phi(x, \mu_k)$ by
(approximately) solving the problem

$$\underset{s \in \mathbb{R}^n}{\text{minimize}} \ g(x, y(x))^T s + \tfrac{1}{2} s^T \left(H(x, y) + A^T(x) C^{-1}(x) Y A(x)\right) s$$

possibly subject to a trust-region constraint

$\odot$ $y(x) = \mu C^{-1}(x) e \implies g(x, y(x)) = \nabla_x \Phi(x, \mu)$

$\odot$ $y = \cdots$

$\diamond$ $y(x) \implies$ primal Newton method

$\diamond$ occasionally $(\mu_{k-1}/\mu_k) y(x) \implies$ good starting point

$\diamond$ $y^{\text{OLD}} + w^{\text{OLD}} \implies$ primal-dual Newton method

$\diamond$ $\max(y^{\text{OLD}} + w^{\text{OLD}}, \epsilon(\mu_k) e)$ for "small" $\epsilon(\mu_k) > 0$
(e.g., $\epsilon(\mu_k) = \mu_k^{1.5}$) $\implies$ practical primal-dual method

Ill-conditioning $\not\Longrightarrow$ we can't solve equations accurately:

roughly (non-degenerate case, $\mathcal{I}$ = inactive set at x_*)

$$\begin{pmatrix} H & -A^T \\ YA & C \end{pmatrix} \begin{pmatrix} s \\ w \end{pmatrix} = -\begin{pmatrix} g - A^T y \\ Cy - \mu e \end{pmatrix} \Longrightarrow$$

$$\begin{pmatrix} H & -A_A^T & -A_I^T \\ Y_A A_A & C_A & 0 \\ Y_I A_I & 0 & C_I \end{pmatrix} \begin{pmatrix} s \\ w_A \\ w_I \end{pmatrix} = -\begin{pmatrix} g - A_A^T y_A - A_I^T y_I \\ C_A y_A - \mu e \\ C_I y_I - \mu e \end{pmatrix} \Longrightarrow$$

$$\begin{pmatrix} H + A_I^T C_I^{-1} Y_I A_I & -A_A^T \\ A_A & C_A Y_A^{-1} \end{pmatrix} \begin{pmatrix} s \\ w_A \end{pmatrix} = -\begin{pmatrix} g - A_A^T y_A - \mu A_I^T C_I^{-1} e \\ c_A - \mu Y_A^{-1} e \end{pmatrix}$$

$\odot$ potentially bad terms C_I^{-1} and Y_A^{-1} bounded

$\odot$ in the limit becomes well-behaved

$$\begin{pmatrix} H & -A_A^T \\ A_A & 0 \end{pmatrix} \begin{pmatrix} s \\ w_A \end{pmatrix} = -\begin{pmatrix} g - A_A^T y_A \\ 0 \end{pmatrix}$$

PRACTICAL PRIMAL-DUAL METHOD

Given $\mu_0 > 0$ and feasible $(x_0^{\rm s}, y_0^{\rm s})$, set $k = 0$

Until "convergence" iterate:

Inner minimization: starting from $(x_k^{\rm s}, y_k^{\rm s})$, use an
unconstrained minimization algorithm to find (x_k, y_k) for which

$$\|C(x_k)y_k - \mu_k e\| \leq \mu_k \text{ and } \|g(x_k) - A^T(x_k)y_k\| \leq \mu_k^{1.00005}$$

Set $\mu_{k+1} = \min(0.1\mu_k, \mu_k^{1.9999})$

Find $(x_{k+1}^{\rm s}, y_{k+1}^{\rm s})$ using a primal-dual Newton step from (x_k, y_k)

If $(x_{k+1}^{\rm s}, y_{k+1}^{\rm s})$ is infeasible, reset $(x_{k+1}^{\rm s}, y_{k+1}^{\rm s})$ to (x_k, y_k)

Increase k by 1

FAST ASYMPTOTIC CONVERGENCE

Theorem 4.2. Suppose that f, c $\in$ $\mathcal{C}^2$, that a subsequence $\{(x_k, y_k)\}$, $k \in \mathcal{K}$, of the practical primal-dual method converges to (x_*, y_*) satisfying second-order sufficiency conditions, that $A_{\mathcal{A}}(x_*)$ is full-rank, and that $(y_*)_{\mathcal{A}} > 0$. Then the starting point satisfies the inner-minimization termination test (i.e., $(x_k, y_k) = (x_k^{\mathrm{s}}, y_k^{\mathrm{s}})$) and the whole sequence $\{(x_k, y_k)\}$ converges to (x_*, y_*) at a superlinear rate (Q-factor 1.9998).

OTHER ISSUES

⊙ polynomial algorithms for many convex problems

    ⋄ linear programming

    ⋄ quadratic programming

    ⋄ semi-definite programming . . .

⊙ excellent practical performance

⊙ globally, need to keep away from constraint boundary until near convergence, otherwise very slow

⊙ initial interior point:

$$\underset{(x,c)}{\text{minimize}} \ \ e^T c \ \ \text{subject to} \ \ c(x) + c \geq 0$$