

Part 4: Interior-point methods for inequality constrained optimization

Nick Gould (RAL)

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

MSc course on nonlinear optimization

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

CONSTRAINED MINIMIZATION

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

where the **objective function** $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and the **constraints** $c : \mathbb{R}^n \rightarrow \mathbb{R}^m$

- assume that $f, c \in C^1$ (sometimes C^2) and Lipschitz
- often in practice this assumption violated, but not necessary

AN EXAMPLE FOR EQUALITY CONSTRAINTS

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

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 F^m FOR INEQUALITY CONSTRAINTS

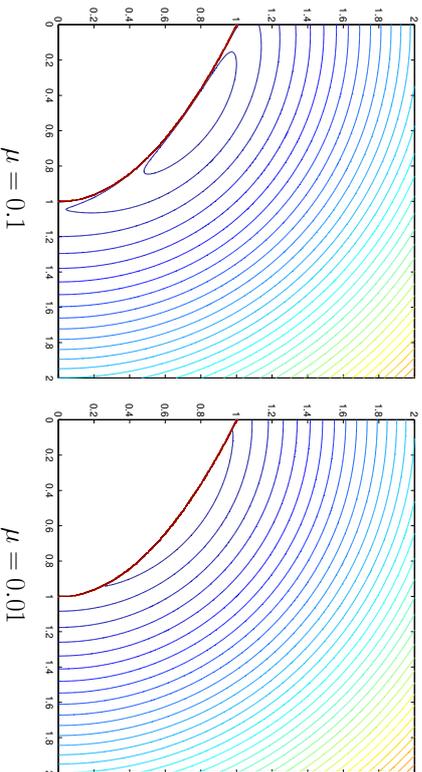
minimize $f(x)$ subject to $c(x) \geq 0$
 $x \in \mathbb{R}^n$

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

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

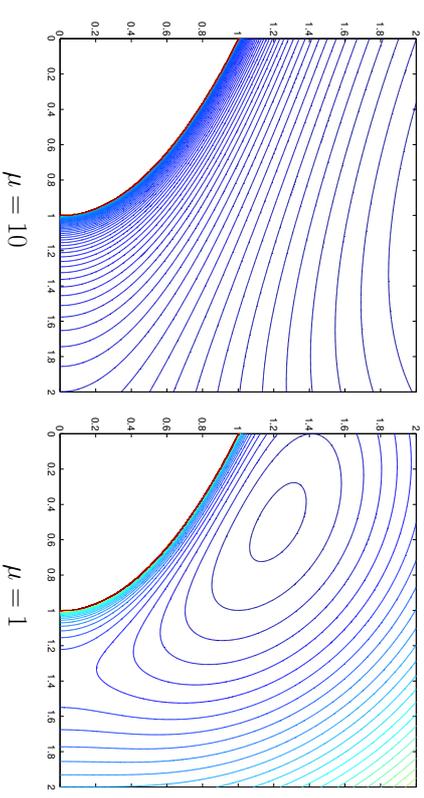
- required solution as μ approaches $\{0\}$ from above
- may have other useless stationary points
- requires a strictly interior point to start
- consequent points are interior

CONTOURS OF THE BARRIER FUNCTION (cont.)



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

CONTOURS OF THE BARRIER FUNCTION



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 $\lim_{k \rightarrow \infty} \mu_{k+1} = 0$ 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 $i = 1, \dots, m$, that

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

where ϵ_k converges to zero as $k \rightarrow \infty$, 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

$$\text{minimize } f(x) \text{ subject to } c(x) \geq 0$$

$$x \in \mathbb{R}^n$$

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

+ (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 $y_* + c_i(x_k)(y_k)_i = \mu_k \implies$

$$c(x_*) \geq 0, y_* \geq 0 \text{ and } c_i(x_*)(y_*)_i = 0.$$

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

PROOF OF THEOREM 4.1

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

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

Define

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

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

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

$$\begin{aligned} \|g(x_k) - A_{\mathcal{I}}^T(x_k)(y_k)_{\mathcal{A}}\|_2 &\leq \|g(x_k) - A^T(x_k)y_k\|_2 + \|A_{\mathcal{I}}^T(x_k)(y_k)_{\mathcal{I}}\|_2 \\ &\leq \bar{\epsilon}_k \stackrel{\text{def}}{=} \epsilon_k + \mu_k \frac{\|A_{\mathcal{I}}\|_2}{\min_{i \in \mathcal{I}} |c_i(x_*)|} \end{aligned} \quad (2)$$

$$\begin{aligned} \implies \|A_{\mathcal{I}}^+(x_k)g(x_k) - (y_k)_{\mathcal{A}}\|_2 &= \|A_{\mathcal{I}}^+(x_k)(g(x_k) - A_{\mathcal{I}}^T(x_k)(y_k)_{\mathcal{A}})\|_2 \\ &\leq 2\|A_{\mathcal{I}}^+(x_*)\|_2 \bar{\epsilon}_k \end{aligned}$$

$$\begin{aligned} \implies \|(y_k)_{\mathcal{A}} - (y_*)_{\mathcal{A}}\|_2 &\leq \|A_{\mathcal{I}}^+(x_*)g(x_*) - A_{\mathcal{I}}^+(x_k)g(x_k)\|_2 + \|A_{\mathcal{I}}^+(x_k)g(x_k) - (y_k)_{\mathcal{A}}\|_2 \\ &\leq \|A_{\mathcal{I}}^+(x_*)g(x_*) - A_{\mathcal{I}}^+(x_k)g(x_k)\|_2 + \|A_{\mathcal{I}}^+(x_k)g(x_k) - (y_k)_{\mathcal{A}}\|_2 \end{aligned}$$

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 $c(x_k + s_k) \not\geq 0$
 - (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

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

where

$$\circ C(x) = \text{diag}(c_1(x), \dots, c_m(x))$$

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

where e is the vector of ones

$$\circ g(x, y(x)) = g(x) - A^T(x)y(x): \text{gradient of the Lagrangian}$$

$$\circ H(x, y(x)) = H(x) - \sum_{i=1}^m y_i(x)H_i(x)$$

Sometimes written as

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

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

where

$$\circ Y(x) = \text{diag}(y_1(x), \dots, y_m(x))$$

POTENTIAL DIFFICULTIES II

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

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

Roughly speaking (non-degenerate case) Newton correction satisfies

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

\implies (full rank)

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

\implies (Taylor expansion)

$$c_{\mathcal{A}}(x_k + s) \approx c_{\mathcal{A}}(x_k) + A_{\mathcal{A}}(x_k)s \approx \left(2 - \frac{\mu_k}{\mu_{k+1}}\right)c_{\mathcal{A}}(x_k) < 0$$

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

POTENTIAL DIFFICULTIES I

Ill-conditioning of the Hessian of the barrier function:

roughly speaking (non-degenerate case)

$$\circ m_a \text{ eigenvalues } \approx \lambda_i(A_{\mathcal{A}}^T Y_{\mathcal{A}}^2 A_{\mathcal{A}}) / \mu_k$$

$$\circ n - m_a \text{ eigenvalues } \approx \lambda_i(N_{\mathcal{A}}^T H(x_*, y_*) N_{\mathcal{A}})$$

where

m_a = number of active constraints

\mathcal{A} = active set at x^*

Y = diagonal matrix of Lagrange multipliers

$N_{\mathcal{A}}$ = orthogonal basis for null-space of $A_{\mathcal{A}}$

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

\implies may not be able to find minimizer easily

PERTURBED OPTIMALITY CONDITIONS

First order optimality conditions for

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

are:

$$g(x) - A^T(x)y = 0 \quad \text{dual feasibility}$$

$$C(x)y = 0 \quad \text{complementary slackness}$$

$$c(x) \geq 0 \text{ and } y \geq 0$$

Consider the ‘‘perturbed’’ problem

$$g(x) - A^T(x)y = 0 \quad \text{dual feasibility}$$

$$C(x)y = \mu e \quad \text{perturbed comp. slacks.}$$

$$c(x) > 0 \text{ and } y > 0$$

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

- nonlinear system \implies use Newton's method

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

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

Eliminate $w \implies$

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

c.f. Newton method for barrier minimization!

POTENTIAL DIFFICULTY II... REVISITED

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

- primal method has to choose $y = y(x_k^s) = \mu_{k+1}C^{-1}(x_k)e$
 - factor μ_{k+1}/μ_k too small for a good Lagrange multiplier estimate
- primal-dual method can choose $y = \mu_k C^{-1}(x_k)e \rightarrow y_*$

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

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

\implies (full rank)

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

\implies (Taylor expansion)

$$c_A(x_k + s^{pd}) \approx c_A(x_k) + A_A(x_k)s^{pd} \approx \frac{\mu_{k+1}}{\mu_k} c_A(x_k) > 0$$

\implies Newton step allowed \implies fast convergence

PRIMAL VS. PRIMAL-DUAL

Primal:

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

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

PRIMAL-DUAL BARRIER METHODS

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

$$\text{minimize}_{s \in \mathbb{R}^n} g(x, y(x))^T s + \frac{1}{2} s^T (H(x, y) + A^T(x)C^{-1}(x)YA(x))s$$

possibly subject to a trust-region constraint

- $y(x) = \mu C^{-1}(x)e \implies g(x, y(x)) = \nabla_x \Phi(x, \mu)$
- $y = \dots$
 - $y(x) \implies$ primal Newton method
 - occasionally $(\mu_{k-1}/\mu_k)y(x) \implies$ good starting point
 - $y^{pd} + w^{pd} \implies$ primal-dual Newton method
 - $\max(y^{pd} + w^{pd}, \epsilon(\mu_k)e)$ for "small" $\epsilon(\mu_k) > 0$ (e.g., $\epsilon(\mu_k) = \mu_k^{1.5}$) \implies practical primal-dual method

POTENTIAL DIFFICULTY 1... REVISITED

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

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

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

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

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

◦ potentially bad terms C_I^{-1} and Y_A^{-1} bounded

◦ in the limit becomes well-behaved

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

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., $(x_k, y_k) = (x_k^s, y_k^s)$) and the whole sequence $\{(x_k, y_k)\}$ converges to (x_*, y_*) at a superlinear rate (Q-factor 1.9998).

PRACTICAL PRIMAL-DUAL METHOD

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

Until "convergence" iterate:

Inner minimization: starting from (x_k^s, y_k^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}^s, y_{k+1}^s) using a primal-dual Newton step from (x_k, y_k)

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

Increase k by 1

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:

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