Inexact Newton Methods for Large-Scale Nonlinear Optimization

Frank E. Curtis, Lehigh University

involving joint work with

Johannes Huber, University of Basel
Travis Johnson, Northwestern University
Daniel P. Robinson, Johns Hopkins University
Olaf Schenk, University of Lugano
Andreas Wächter, Northwestern University

IBM T. J. Watson Research Center, Math Seminar

11 November 2013

Outline

Introduction

Interior-Point Method

Sequential Quadratic Optimization Method

Summary

Outline

Introduction

Interior-Point Method

Sequential Quadratic Optimization Method

Summary

Problem formulation

Consider a constrained nonlinear optimization problem:

$$\begin{aligned} \min_x \ & f(x) \\ \text{s.t. } & c(x) = 0, \ \bar{c}(x) \le 0 \end{aligned} \tag{NLP}$$

Want a solver with the following features:

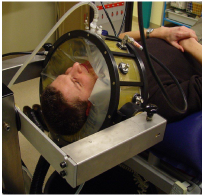
- Scalable step computation
- Superlinear convergence in primal-dual space
- Handles negative curvature

In addition, in real-time optimization, we want:

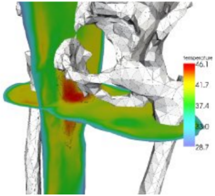
- Monotonicity in a measure of progress
- Active-set detection (for warm-starting)

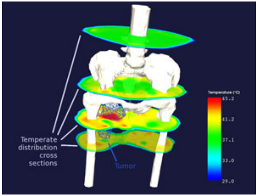
Motivating example 1: Hyperthermia treatment planning

- Cancer therapy to heat large, deeply seated tumors by radio wave adsorption



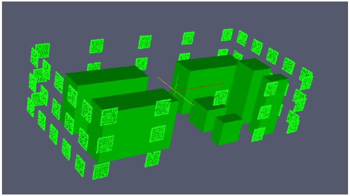
- Heat the tumor to a target temperature with minimal damage to nearby cells





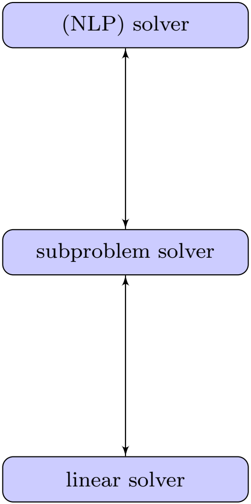
Motivating example 2: Server room cooling

- Heat generating equipment in a server room must be cooled

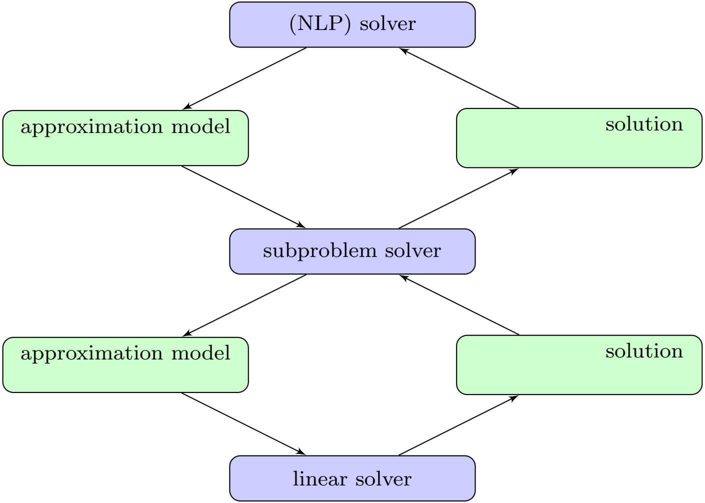


- Place and control ACs to satisfy cooling demands while minimizing costs
- Suggested by Michael Henderson, Vanessa Lopez, and Ulisses Mello (IBM)

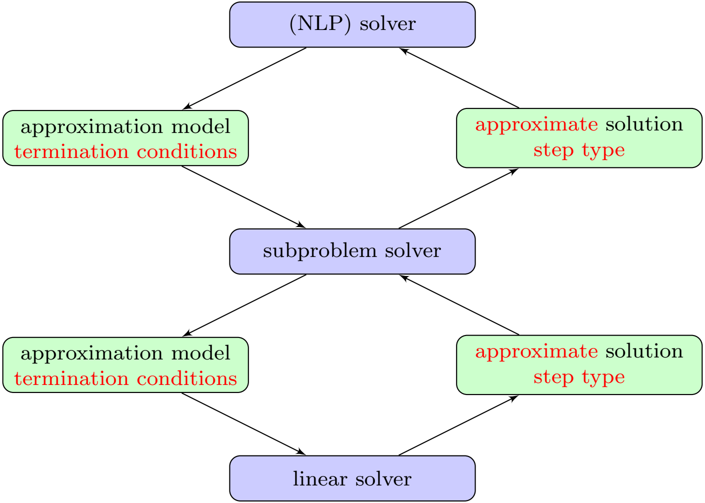
Algorithmic framework: Basic



Algorithmic framework: Detailed



Algorithmic framework: Inexact



Newton's method for nonlinear equations

Newton's method for solving the nonlinear system

$$F(x) = 0$$

involves solving the linear system (approximation model)

$$F(x_k) + \nabla F(x_k) d_k = 0.$$

If we measure progress with the merit function

$$\phi(x) := \tfrac{1}{2}\|F(x)\|_2^2,$$

then there is consistency between the step and merit function:

$$\nabla \phi(x_k)^T d_k = F(x_k)^T \nabla F(x_k) d_k = -\|F(x_k)\|_2^2 \leq 0.$$

Inexact Newton for nonlinear equations

Inexact Newton for solving the nonlinear system

$$F(x) = 0$$

involves approximately solving the linear system (approximation model)

$$F(x_k) + \nabla F(x_k) d_k = r_k.$$

If we measure progress with the merit function

$$\phi(x) := \tfrac{1}{2}\|F(x)\|_2^2 \quad \text{and require}^1 \quad \|r_k\|_2 \leq \kappa \|F(x_k)\|_2 \quad \text{for } \kappa \in (0,1),$$

then there is consistency between the step and merit function:

$$\nabla \phi(x_k)^T d_k = F(x_k)^T \nabla F(x_k) d_k = -\|F(x_k)\|_2^2 + F(x_k)^T r_k \leq (\kappa - 1)\|F(x_k)\|_2^2 \leq 0.$$

[1] Dembo, Eisenstat, Steihaug (1982)

Newton's method for unconstrained optimization

Newton's method for solving the unconstrained optimization problem

$$\min_x f(x) \implies \nabla f(x) = 0$$

involves solving the linear system (approximation model)

$$\nabla f(x_k) + \nabla^2 f(x_k) d_k = 0.$$

If we measure progress with the merit function

$$\phi(x) := f(x) \quad (\text{NOT } \tfrac{1}{2}\|\nabla f(x)\|_2^2)$$

then there MIGHT NOT BE consistency between the step and merit function:

$$\nabla \phi(x_k)^T d_k = \nabla f(x_k)^T d_k = -d_k^T \nabla^2 f(x_k) d_k \leq 0$$

Inexact Newton for unconstrained optimization

Inexact Newton for solving the unconstrained optimization problem

$$\min_x f(x) \implies \nabla f(x) = 0$$

involves approximately solving the linear system (approximation model)

$$\nabla f(x_k) + \nabla^2 f(x_k) d_k = r_k.$$

If we measure progress with the merit function

$$\phi(x) := f(x) \quad \text{and require} \quad \|r_k\| \leq \kappa \|\nabla f(x_k)\| \quad \text{for } \kappa \in (0,1),$$

then there MIGHT NOT BE consistency between the step and merit function:

$$\nabla \phi(x_k)^T d_k = \nabla f(x_k)^T d_k = -d_k^T \nabla^2 f(x_k) d_k + r_k^T d_k \overset{?}{\leq} 0$$

Newton's method for constrained optimization

Newton's method for solving the constrained optimization problem

$$\left\{ \begin{array}{l} \min_x \; f(x) \\ \text{ s.t. } c(x) = 0 \end{array} \right\} \Longrightarrow \begin{bmatrix} \nabla f(x) + \nabla c(x) y \\ c(x) \end{bmatrix} = 0$$

involves solving the linear system (approximation model)

$$\begin{bmatrix} \nabla_x L(x_k, y_k) \\ c(x_k) \end{bmatrix} + \begin{bmatrix} \nabla^2_{xx} L(x_k, y_k) & \nabla c(x_k) \\ \nabla c(x_k)^T & 0 \end{bmatrix} \begin{bmatrix} d_k \\ \delta_k \end{bmatrix} = 0.$$

If we measure progress with the merit function

$$\phi(x, \mu) := \mu f(x) + \|c(x)\|_1,$$

then there MIGHT NOT BE consistency between the step and merit function:

$$D\phi(x_k, d_k) \leq -\mu \nabla f(x_k)^T d_k - \|c(x_k)\|_1 \overset{?}{\leq} 0$$

Problem formulation

Consider a constrained nonlinear optimization problem:

$$\begin{aligned} &\min_x \ f(x) \\ &\text{s.t. } c(x) = 0, \ \bar{c}(x) \leq 0 \end{aligned} \tag{NLP}$$

Even if (NLP) is feasible, we need to balance progress between

- Minimizing the objective function
- Minimizing constraint violation

If (NLP) is infeasible, then we simply want to minimize constraint violation:

$$\min_x \ v(x), \ \text{where } v(x) := \|c(x)\|_1 + \|[\bar{c}(x)]^+\|_1. \tag{FP}$$

(A minimizer of (NLP) is always a minimizer of (FP).)

Outline

Introduction

Interior-Point Method

Sequential Quadratic Optimization Method

Summary

Interior-point method with inexact step computations

We propose an interior-point method for large-scale nonlinear optimization:

- Handles rank deficiency, ill-conditioning, nonconvexity
- Inexactness allowed/controlled with implementable conditions
- Global convergence guarantees
- Encouraging numerical results
- Important remaining issues...

Algorithm highlights

Formulate a logarithmic barrier subproblem:

$$\begin{aligned} \min_x \ & f(x) - \gamma e^T(\ln s) \\ \text{s.t. } & c(x) = 0, \ \bar{c}(x) + s = 0 \end{aligned} \tag{BP}$$

Apply an iterative method to solve the linear system:

$$\begin{bmatrix} H_k & 0 & \nabla c_k & \nabla \bar{c}_k \\ 0 & \Omega_k & 0 & S_k \\ \nabla c_k^T & 0 & 0 & 0 \\ \nabla \bar{c}_k^T & S_k & 0 & 0 \end{bmatrix} \begin{bmatrix} d_k^x \\ d_k^s \\ \delta_k \\ \bar{\delta}_k \end{bmatrix} = - \begin{bmatrix} \nabla f_k + \nabla c_k y_k + \nabla \bar{c}_k \bar{y}_k \\ -\gamma e + S_k \bar{y}_k \\ c_k \\ \bar{c}_k + s_k \end{bmatrix}$$

- Slack reset, fraction-to-the-boundary rules
- Step decomposition into normal and tangential components[2]
- Matrix modifications for handling nonconvexity
- Inexact step must satisfy 1 of 3 termination tests
- Dynamic update for penalty parameter

---

[2] Byrd, Omojokun (1989)

Convergence theory for (BP)[3]

Assumption

The sequence $\{(x_k, s_k, y_k, \bar{y}_k)\}$ is contained in a convex set over which f, c, and $\bar{c}$ and their first derivatives are bounded and Lipschitz continuous.

Theorem

If all limit points of the constraint Jacobian have full row rank, then

$$\lim_{k\to\infty} \left\| \begin{bmatrix} \nabla f_k + \nabla c_k y_k + \nabla \bar{c}_k \bar{y}_k \\ -\gamma e + S_k \bar{y}_k \\ c_k \\ \bar{c}_k + s_k \end{bmatrix} \right\| = 0.$$

Otherwise, we at least have

$$\lim_{k\to\infty} \left\| \begin{bmatrix} \nabla c_k & \nabla \bar{c}_k \\ 0 & S_k \end{bmatrix} \begin{bmatrix} c_k \\ \bar{c}_k + s_k \end{bmatrix} \right\| = 0.$$

and, if $\{\mu_k\}$ is bounded away from zero, then

$$\lim_{k\to\infty} \left\| \begin{bmatrix} \nabla f_k + \nabla c_k y_k + \nabla \bar{c}_k \bar{y}_k \\ -\gamma e + S_k \bar{y}_k \end{bmatrix} \right\| = 0.$$

---

[3] Curtis, Schenk, Wächter (2010)

Convergence theory for (NLP)[3]

Theorem

If the algorithm yields a sufficiently accurate solution of (BP) *for each* $\{\gamma_j\} \to 0$ *and if the linear independence constraint qualification (LICQ) holds at a limit point* x *of* $\{x_j\}$, *then there exist Lagrange multipliers* $(y, \bar{y})$ *such that the first-order optimality conditions of* (NLP) *are satisfied.*

Implementation details

- Incorporated in IPOPT software package[4]
  - interior-point algorithm with inexact step computations
  - flexible penalty function for promoting faster convergence[5]
  - tests on ~ 700 CUTEr problems (almost) on par with original IPOPT
- Linear systems solved with PARDISO[6]
  - includes iterative linear system solvers, e.g., SQMR[7]
  - incomplete multilevel factorization with inverse-based pivoting
  - stabilized by symmetric-weighted matchings
- Server cooling room example coded w/ libmesh[8]

[4] Wächter, Laird, Biegler
[5] Curtis, Nocedal (2008)
[6] Schenk, Gärtner
[7] Freund (1997)
[8] Kirk, Peterson, Stogner, Carey

Hyperthermia treatment planning

Let $u_j = a_j e^{i\phi_j}$ and $M_{jk}(x) = \langle E_j(x), E_k(x)\rangle$ where $E_j = \sin(jx_1x_2x_3\pi)$:

$$
\begin{array}{ll}
\min & \frac{1}{2}\int_\Omega (y(x) - y_t(x))^2 dx \\
\text{s.t.} & \left\{ \begin{array}{rcll}
-\Delta y(x) - 10(y(x) - 37) & = & u^* M(x) u & \text{in } \Omega \\
37.0 \leq y(x) & \leq & 37.5 & \text{on } \partial\Omega \\
42.0 \leq y(x) & \leq & 44.0 & \text{in } \Omega_0
\end{array} \right.
\end{array}
$$

Original `IPOPT` with $N = 32$ requires 408 seconds per iteration.

N	n	p	q	# iter	CPU sec (per iter)
16	4116	2744	2994	68	22.893 (0.3367)
32	32788	27000	13034	51	3055.9 (59.920)

Server room cooling

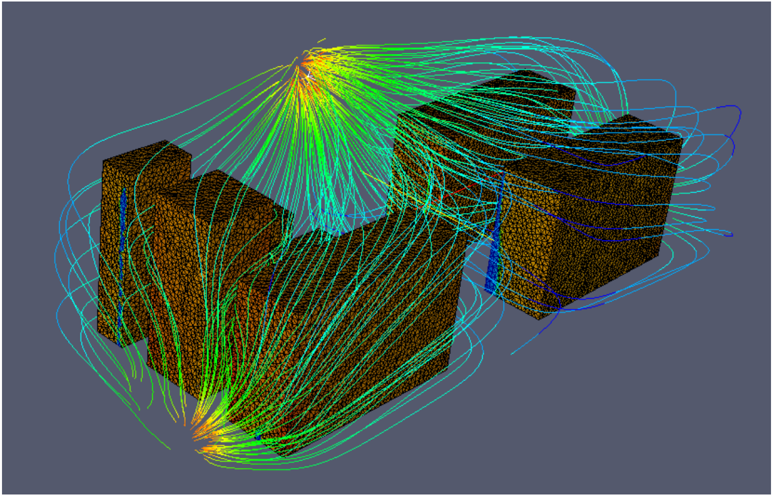
Let $\phi(x)$ be the air flow velocity potential:

$$\min \sum c_i v_{AC_i}$$
$$\text{s.t.} \begin{cases} \nabla \phi(x) = 0 & \text{in } \Omega \\ \partial_n \phi(x) = 0 & \text{on } \partial\Omega_{wall} \\ \partial_n \phi(x) = -v_{AC_i} & \text{on } \partial\Omega_{AC_i} \\ \phi(x) = 0 & \text{in } \Omega_{Exh_j} \\ \|\nabla \phi(x)\|_2^2 \geq v_{min}^2 & \text{on } \partial\Omega_{hot} \\ v_{AC_i} \geq 0 & \end{cases}$$

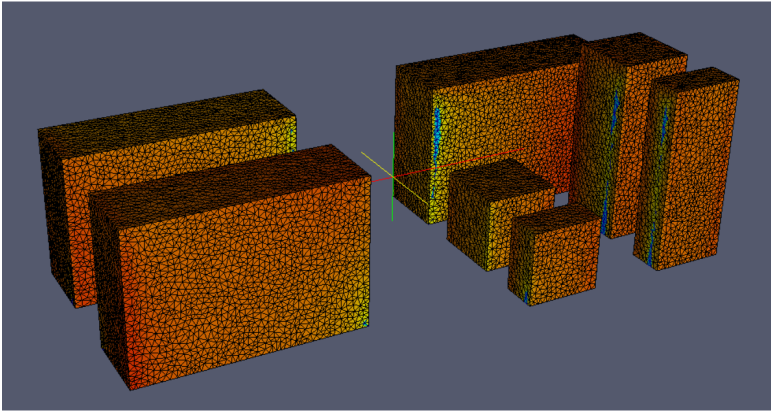
Original `IPOPT` with $h = 0.05$ requires 2390.09 seconds per iteration.

h	n	p	q	# iter	CPU sec (per iter)
0.10	43816	43759	4793	47	1697.47 (36.1164)
0.05	323191	323134	19128	54	28518.4 (528.119)

Server room cooling solution



Server room cooling solution (multiplier values)



Summary and challenges

Interior point method with inexact step computations:

- Handles rank deficiency, ill-conditioning, nonconvexity
- Inexactness allowed/controlled with implementable conditions
- Global convergence guarantees
- Encouraging numerical results

Challenges ahead:

- PRECONDITIONING
- Handling lack of strict complementarity
- Mesh refinement and warm-starting

Outline

Introduction

Interior-Point Method

Sequential Quadratic Optimization Method

Summary

Sequential quadratic optimization

Advantages:

- "Parameter free" search direction computation (ideally)
- Strong global convergence properties and behavior
- Active-set identification $\Longrightarrow$ Newton-like local convergence

Disadvantage:

- Quadratic subproblems (QPs) are expensive to solve exactly

Fritz John and penalty functions

(NLP):

$$\min_x f(x)$$
$$\text{s.t. } c(x) = 0,\ \bar{c}(x) \leq 0$$

(FP):

$$\min_x v(x) := \left\| \begin{bmatrix} c(x) \\ [\bar{c}(x)]^+ \end{bmatrix} \right\|_1$$

Define the Fritz John (FJ) function

$$\mathcal{F}(x, y, \bar{y}, \mu) := \mu f(x) + c(x)^T y + \bar{c}(x)^T \bar{y}$$

and the ℓ_1-norm exact penalty function

$$\phi(x, \mu) := \mu f(x) + v(x).$$

$\mu \geq 0$ acts as objective multiplier/penalty parameter.

Optimality conditions

(NLP):

$$\min_x \; f(x)$$
$$\text{s.t. } c(x) = 0, \;\; \bar{c}(x) \leq 0$$

(FP):

$$\min_x \;\; v(x) := \left\| \begin{bmatrix} c(x) \\ [\bar{c}(x)]^+ \end{bmatrix} \right\|_1$$

(PP):

$$\min_x \;\; \phi(x, \mu) := \mu f(x) + v(x)$$

(FJ):

$$\mathcal{F}(x, y, \bar{y}, \mu) :=$$
$$\mu f(x) + c(x)^T y + \bar{c}(x)^T \bar{y}$$

KKT conditions for (FP) and (PP) expressed with residual

$$\rho(x, y, \bar{y}, \mu) := \begin{bmatrix} \mu g(x) + J(x)y + \bar{J}(x)\bar{y} \\ \min\{[c(x)]^+, e - y\} \\ \min\{[c(x)]^-, e + y\} \\ \min\{[\bar{c}(x)]^+, e - \bar{y}\} \\ \min\{[\bar{c}(x)]^-, \quad \bar{y}\} \end{bmatrix}$$

- FJ point:

$$\rho(x, y, \bar{y}, \mu) = 0, \;\; v(x) = 0, \;\; (y, \bar{y}, \mu) \neq 0$$

- KKT point:

$$\rho(x, y, \bar{y}, \mu) = 0, \;\; v(x) = 0, \;\; \mu > 0$$

- Infeasible stationary point:

$$\rho(x, y, \bar{y}, 0) = 0, \;\; v(x) > 0$$

Penalty function model and QP subproblems

(NLP):

$$\min_x f(x)$$
$$\text{s.t. } c(x) = 0, \;\; \bar{c}(x) \leq 0$$

(FP):

$$\min_x \; v(x) := \left\| \begin{bmatrix} c(x) \\ [\bar{c}(x)]^+ \end{bmatrix} \right\|_1$$

(PP):

$$\min_x \; \phi(x,\mu) := \mu f(x) + v(x)$$

(FJ):

$$\mathcal{F}(x,y,\bar{y},\mu) :=$$
$$\mu f(x) + c(x)^T y + \bar{c}(x)^T \bar{y}$$

KKT residual:

$$\rho(x,y,\bar{y},\mu)$$

Define a local model of $\phi(\cdot,\mu)$ at x_k:

$$l_k(d,\mu) := \mu(f_k + g_k^T d) + \|c_k + J_k^T d\|_1 + \|[\bar{c}_k + \bar{J}_k^T d]^+\|_1$$

Reduction in this model yielded by a given d:

$$\Delta l_k(d,\mu) := \Delta l(0,\mu) - \Delta l(d,\mu)$$

Two subproblems of interest:

$$\min_d \; -\Delta l_k(d,\mu_k) + \tfrac{1}{2} d^T H_k d \qquad \text{(PQP)}$$

$$\min_d \; -\Delta l_k(d,0) + \tfrac{1}{2} d^T H_k d \qquad \text{(FQP)}$$

$\Delta l_k(d,\mu) > 0$ implies d is a direction of strict descent for $\phi(\cdot,\mu)$ from x_k

Optimality conditions (for QPs)

(NLP):

$$\min_x f(x)$$

$$\text{s.t. } c(x) = 0, \ \bar{c}(x) \leq 0$$

(FP):

$$\min_x v(x) := \left\| \begin{bmatrix} c(x) \\ [\bar{c}(x)]^+ \end{bmatrix} \right\|_1$$

(PP):

$$\min_x \phi(x,\mu) := \mu f(x) + v(x)$$

(FJ):

$$\mathcal{F}(x,y,\bar{y},\mu) :=$$

$$\mu f(x) + c(x)^T y + \bar{c}(x)^T \bar{y}$$

KKT residual:

$$\rho(x,y,\bar{y},\mu)$$

Local model of ϕ at x_k:

$$l_k(d,\mu)$$

KKT conditions for (PQP) and (FQP) expressed with

$$\rho_k(d,y,\bar{y},\mu,H) := \begin{bmatrix} \mu g_k + Hd + J_k y + \bar{J}_k \bar{y} \\ \min\{[c_k + J_k^T d]^+, e - y\} \\ \min\{[c_k + J_k^T d]^-, e + y\} \\ \min\{[\bar{c}_k + \bar{J}_k^T d]^+, e - \bar{y}\} \\ \min\{[\bar{c}_k + \bar{J}_k^T d]^-, \bar{y}\} \end{bmatrix}$$

- "Exact" solution of (PQP):

$$\rho_k(d,y,\bar{y},\mu_k,H_k) = 0$$

- "Exact" solution of (FQP):

$$\rho_k(d,y,\bar{y},0,H_k) = 0$$

"Exact" SQO: Scenario A

(NLP):

$$\min_x f(x)$$
$$\text{s.t. } c(x) = 0, \; \bar{c}(x) \leq 0$$

(FP):

$$\min_x v(x) := \left\| \begin{bmatrix} c(x) \\ [\bar{c}(x)]^+ \end{bmatrix} \right\|_1$$

(PP):

$$\min_x \phi(x,\mu) := \mu f(x) + v(x)$$

(FJ):

$$\mathcal{F}(x,y,\bar{y},\mu) :=$$
$$\mu f(x) + c(x)^T y + \bar{c}(x)^T \bar{y}$$

KKT residuals:

$$\rho(x,y,\bar{y},\mu)$$
$$\rho_k(d,y,\bar{y},\mu,H)$$

Local model of ϕ at x_k:

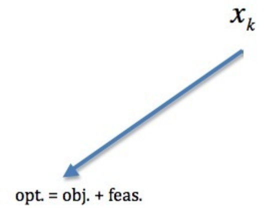
$$l_k(d,\mu)$$

Solution $(d_k, y_{k+1}, \bar{y}_{k+1})$ of (PQP) satisfies

- $\rho(d_k, y_{k+1}, \bar{y}_{k+1}, \mu_k, H_k) = 0$
- $\Delta l_k(d_k, \mu_k) \geq \epsilon v_k$ for $\epsilon \in (0,1)$

then

- $d_k \leftarrow d_k$ is the search direction
- $\mu_{k+1} \leftarrow \mu_k$



"Exact" SQO: Scenario B

(NLP):

$$\min_x f(x)$$
$$\text{s.t. } c(x) = 0,\ \bar{c}(x) \le 0$$

(FP):

$$\min_x v(x) := \left\| \begin{bmatrix} c(x) \\ [\bar{c}(x)]^+ \end{bmatrix} \right\|_1$$

(PP):

$$\min_x \phi(x,\mu) := \mu f(x) + v(x)$$

(FJ):

$$\mathcal{F}(x,y,\bar{y},\mu) :=$$
$$\mu f(x) + c(x)^T y + \bar{c}(x)^T \bar{y}$$

KKT residuals:

$$\rho(x,y,\bar{y},\mu)$$
$$\rho_k(d,y,\bar{y},\mu,H)$$

Local model of ϕ at x_k:

$$l_k(d,\mu)$$

Solution $(d_k, y_{k+1}, \bar{y}_{k+1})$ of (PQP) satisfies

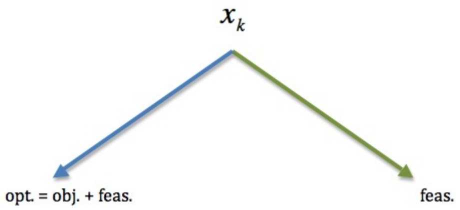
- $\rho(d_k, y_{k+1}, \bar{y}_{k+1}, \mu_k, H_k) = 0$
- $\Delta l_k(d_k, \mu_k) \ge \epsilon \Delta l_k(d_k, 0)$ for $\epsilon \in (0,1)$

where solution $(d_k, y_{k+1}, \bar{y}_{k+1})$ of (FQP) satisfies

- $\rho(d_k, y_{k+1}, \bar{y}_{k+1}, 0, H_k) = 0$

then

- $d_k \leftarrow d_k$ is the search direction
- $\mu_{k+1} \leftarrow \mu_k$



"Exact" SQO: Scenario C

(NLP):

$$\min_x f(x)$$
$$\text{s.t. } c(x) = 0, \; \bar{c}(x) \leq 0$$

(FP):

$$\min_x v(x) := \left\| \begin{bmatrix} c(x) \\ [\bar{c}(x)]^+ \end{bmatrix} \right\|_1$$

(PP):

$$\min_x \phi(x, \mu) := \mu f(x) + v(x)$$

(FJ):

$$\mathcal{F}(x, y, \bar{y}, \mu) :=$$
$$\mu f(x) + c(x)^T y + \bar{c}(x)^T \bar{y}$$

KKT residuals:

$$\rho(x, y, \bar{y}, \mu)$$
$$\rho_k(d, y, \bar{y}, \mu, H)$$

Local model of ϕ at x_k:

$$l_k(d, \mu)$$

Solution $(d_k, y_{k+1}, \bar{y}_{k+1})$ of (PQP) satisfies

- $\rho(d_k, y_{k+1}, \bar{y}_{k+1}, \mu_k, H_k) = 0$

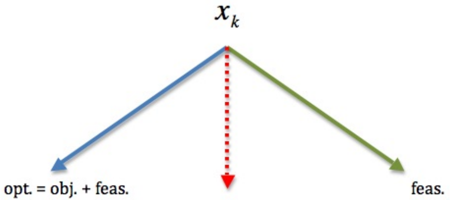
and solution $(d_k, y_{k+1}, \bar{y}_{k+1})$ of (FQP) satisfies

- $\rho(d_k, y_{k+1}, \bar{y}_{k+1}, 0, H_k) = 0$

then

- $d_k \leftarrow \tau d_k + (1 - \tau) d_k$ so $\Delta l_k(d_k, 0) \geq \epsilon \Delta l_k(d_k, 0)$
- $\mu_{k+1} < \mu_k$ so $\Delta l_k(d_k, \mu_k) \geq \beta \Delta l_k(d_k, 0)$ for $\beta \in (0, 1)$
- Overall:

$$\Delta l_k(d_k, \mu_k) \geq \beta \Delta l_k(d_k, 0) \geq \beta \epsilon \Delta l_k(d_k, 0)$$



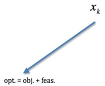
"Exact" SQO

repeat

(1) Check whether KKT point or infeasible stationary point has been obtained.

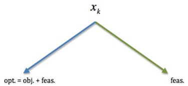
(2) Compute an exact solution of (PQP).

  (a) If Scenario A occurs, then go to step 4.

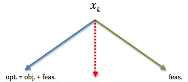


(3) Compute an exact solution of (FQP).

  (a) If Scenario B occurs, then go to step 4.



  (b) Otherwise, Scenario C occurs.



(4) Perform a backtracking line search to reduce $\phi(\cdot, \mu_{k+1})$.

endrepeat

Inexact algorithm preview

Our inexact algorithm involves 6 (!) scenarios:

- 3 are derived from scenarios in the "exact" algorithm
- 3 focus on penalty parameter and Lagrange multiplier updates

Two critical questions to answer:

- When can we terminate the QP solver?
- When is a given inexact solution good enough?

Terminating the QP solver: Test P1

(NLP):

$$\min_x f(x)$$
$$\text{s.t. } c(x) = 0, \ \bar{c}(x) \leq 0$$

(FP):

$$\min_x v(x) := \left\| \begin{bmatrix} c(x) \\ [\bar{c}(x)]^+ \end{bmatrix} \right\|_1$$

(PP):

$$\min_x \phi(x, \mu) := \mu f(x) + v(x)$$

(FJ):

$$\mathcal{F}(x, y, \bar{y}, \mu) :=$$
$$\mu f(x) + c(x)^T y + \bar{c}(x)^T \bar{y}$$

KKT residuals:

$$\rho(x, y, \bar{y}, \mu)$$
$$\rho_k(d, y, \bar{y}, \mu, H)$$

Local model of ϕ at x_k:

$$l_k(d, \mu)$$

Solution $(d_k, y_{k+1}, \bar{y}_{k+1})$ of (PQP) satisfies

- $y_{k+1} \in [-e, e]$, $\bar{y}_{k+1} \in [0, e]$
- $\Delta l_k(d_k, \mu_k) \geq \theta \|d_k\|^2 > 0$ for $\theta \in (0, 1)$
- $\|\rho_k(d_k, y_{k+1}, \bar{y}_{k+1}, \mu_k, H_k)\| \leq \kappa \left\| \begin{bmatrix} \rho(x_k, y_k, \bar{y}_k, \mu_k) \\ \rho(x_k, y_k, \bar{y}_k, 0) \end{bmatrix} \right\|$

Terminating the QP solver: Test P2

(NLP):

$$\min_x f(x)$$
$$\text{s.t. } c(x) = 0, \ \bar{c}(x) \leq 0$$

(FP):

$$\min_x v(x) := \left\| \begin{bmatrix} c(x) \\ [\bar{c}(x)]^+ \end{bmatrix} \right\|_1$$

(PP):

$$\min_x \phi(x,\mu) := \mu f(x) + v(x)$$

(FJ):

$$\mathcal{F}(x,y,\bar{y},\mu) :=$$
$$\mu f(x) + c(x)^T y + \bar{c}(x)^T \bar{y}$$

KKT residuals:

$$\rho(x,y,\bar{y},\mu)$$
$$\rho_k(d,y,\bar{y},\mu,H)$$

Local model of ϕ at x_k:

$$l_k(d,\mu)$$

Solution $(d_k, y_{k+1}, \bar{y}_{k+1})$ of (PQP) satisfies

- $y_{k+1} \in [-e, e]$, $\bar{y}_{k+1} \in [0, e]$
- $\Delta l_k(d_k, \mu_k) \geq \theta \|d_k\|^2 > 0$ for $\theta \in (0,1)$
- $\|\rho_k(d_k, y_{k+1}, \bar{y}_{k+1}, \mu_k, H_k)\| \leq \kappa \|\rho(x_k, y_k, \bar{y}_k, 0)\|$

Furthermore,

- If $\Delta l_k(d_k, 0) < \epsilon v_k$, then

$$\|(y_{k+1}, \bar{y}_{k+1})\|_\infty \gg 0$$

- If $\Delta l_k(d_k, \mu_k) < \beta \Delta l_k(d_k, 0)$, then

$$\|(y_{k+1}, \bar{y}_{k+1})\|_\infty \gg 0$$

Terminating the QP solver: Test P3

(NLP):

$$\min_x \ f(x)$$
$$\text{s.t. } c(x) = 0, \ \bar{c}(x) \leq 0$$

(FP):

$$\min_x \ v(x) := \left\| \begin{bmatrix} c(x) \\ [\bar{c}(x)]^+ \end{bmatrix} \right\|_1$$

(PP):

$$\min_x \ \phi(x, \mu) := \mu f(x) + v(x)$$

(FJ):

$$\mathcal{F}(x, y, \bar{y}, \mu) :=$$
$$\mu f(x) + c(x)^T y + \bar{c}(x)^T \bar{y}$$

KKT residuals:

$$\rho(x, y, \bar{y}, \mu)$$
$$\rho_k(d, y, \bar{y}, \mu, H)$$

Local model of ϕ at x_k:

$$l_k(d, \mu)$$

Solution $(d_k, y_{k+1}, \bar{y}_{k+1})$ of (PQP) satisfies

- $y_{k+1} \in [-e, e]$, $\bar{y}_{k+1} \in [0, e]$
- $\|\rho_k(0, y_{k+1}, \bar{y}_{k+1}, \mu_k, H_k)\| \leq \kappa \|\rho(x_k, y_k, \bar{y}_k, \mu_k)\|$

Furthermore, for $\zeta \in (0, 1)$,

- If $\|\rho(x_k, y_k, \bar{y}_k, \mu_k)\| < \zeta \|\rho(x_k, y_k, \bar{y}_k, 0)\|$, then

$$\|(y_{k+1}, \bar{y}_{k+1})\|_\infty \gg 0$$

Terminating the QP solver: Test F

(NLP):

$$\min_x f(x)$$
$$\text{s.t. } c(x) = 0,\ \bar{c}(x) \le 0$$

(FP):

$$\min_x v(x) := \left\| \begin{bmatrix} c(x) \\ [\bar{c}(x)]^+ \end{bmatrix} \right\|_1$$

(PP):

$$\min_x \phi(x, \mu) := \mu f(x) + v(x)$$

(FJ):

$$\mathcal{F}(x, y, \bar{y}, \mu) :=$$
$$\mu f(x) + c(x)^T y + \bar{c}(x)^T \bar{y}$$

KKT residuals:

$$\rho(x, y, \bar{y}, \mu)$$
$$\rho_k(d, y, \bar{y}, \mu, H)$$

Local model of ϕ at x_k:

$$l_k(d, \mu)$$

Solution $(d_k, y_{k+1}, \bar{y}_{k+1})$ of (FQP) satisfies

- $y_{k+1} \in [-e, e]$, $\bar{y}_{k+1} \in [0, e]$
- $\max\{\Delta l_k(d_k, \mu_k), \Delta l_k(d_k, 0)\} \ge \theta \|d_k\|^2$ for $\theta \in (0, 1)$
- $\|\rho_k(d_k, y_{k+1}, \bar{y}_{k+1}, 0, H_k)\| \le \kappa \|\rho(x_k, y_k, \bar{y}_k, 0)\|$

"Inexact" SQO: Scenario 1

(NLP):

$$\min_x \; f(x)$$
$$\text{s.t. } c(x) = 0, \;\; \bar{c}(x) \le 0$$

(FP):

$$\min_x \; v(x) := \left\| \begin{bmatrix} c(x) \\ [\bar{c}(x)]^+ \end{bmatrix} \right\|_1$$

(PP):

$$\min_x \; \phi(x,\mu) := \mu f(x) + v(x)$$

(FJ):

$$\mathcal{F}(x,y,\bar{y},\mu) :=$$
$$\mu f(x) + c(x)^T y + \bar{c}(x)^T \bar{y}$$

KKT residuals:

$$\rho(x,y,\bar{y},\mu)$$
$$\rho_k(d,y,\bar{y},\mu,H)$$

Local model of ϕ at x_k:

$$l_k(d,\mu)$$

The current iterate satisfies

- $\|\rho(x_k, y_k, \bar{y}_k, \mu_k)\| = 0$
- $v(x_k) > 0$

then

- $d_k \leftarrow 0$ is the search direction
- $\mu_{k+1} < \delta \mu_k$ for $\delta \in (0,1)$

"Inexact" SQO: Scenario 2

(NLP):

$$\min_x f(x)$$
$$\text{s.t. } c(x) = 0, \ \bar{c}(x) \le 0$$

(FP):

$$\min_x v(x) := \left\| \begin{bmatrix} c(x) \\ [\bar{c}(x)]^+ \end{bmatrix} \right\|_1$$

(PP):

$$\min_x \phi(x,\mu) := \mu f(x) + v(x)$$

(FJ):

$$\mathcal{F}(x,y,\bar{y},\mu) :=$$
$$\mu f(x) + c(x)^T y + \bar{c}(x)^T \bar{y}$$

KKT residuals:

$$\rho(x,y,\bar{y},\mu)$$
$$\rho_k(d,y,\bar{y},\mu,H)$$

Local model of ϕ at x_k:

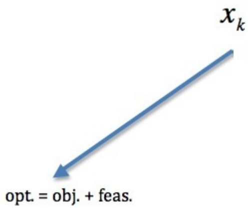
$$l_k(d,\mu)$$

Solution $(d_k, y_{k+1}, \bar{y}_{k+1})$ of (PQP) satisfies

- Test P1
- $\Delta l_k(d_k,\mu_k) \ge \epsilon v_k$ for $\epsilon \in (0,1)$

then

- $d_k \leftarrow d_k$ is the search direction
- $\mu_{k+1} \leftarrow \mu_k$



"Inexact" SQO: Scenario 3

(NLP):

$$\min_x f(x)$$
$$\text{s.t. } c(x) = 0, \ \bar{c}(x) \le 0$$

(FP):

$$\min_x v(x) := \left\| \begin{bmatrix} c(x) \\ [\bar{c}(x)]^+ \end{bmatrix} \right\|_1$$

(PP):

$$\min_x \phi(x,\mu) := \mu f(x) + v(x)$$

(FJ):

$$\mathcal{F}(x,y,\bar{y},\mu) :=$$
$$\mu f(x) + c(x)^T y + \bar{c}(x)^T \bar{y}$$

KKT residuals:

$$\rho(x,y,\bar{y},\mu)$$
$$\rho_k(d,y,\bar{y},\mu,H)$$

Local model of ϕ at x_k:

$$l_k(d,\mu)$$

Solution $(d_k, y_{k+1}, \bar{y}_{k+1})$ of (PQP) satisfies

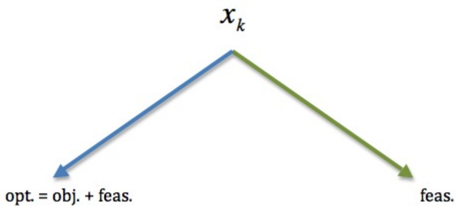
- Test P1
- $\Delta l_k(d_k,\mu_k) \ge \epsilon \Delta l_k(d_k,0)$ for $\epsilon \in (0,1)$

where solution $(d_k, y_{k+1}, \bar{y}_{k+1})$ of (FQP) satisfies

- Test F

then

- $d_k \leftarrow d_k$ is the search direction
- $\mu_{k+1} \leftarrow \mu_k$



"Inexact" SQO: Scenario 4

(NLP):

$$\min_x f(x)$$

$$\text{s.t. } c(x) = 0, \; \bar{c}(x) \leq 0$$

(FP):

$$\min_x v(x) := \left\| \begin{bmatrix} c(x) \\ [\bar{c}(x)]^+ \end{bmatrix} \right\|_1$$

(PP):

$$\min_x \phi(x,\mu) := \mu f(x) + v(x)$$

(FJ):

$$\mathcal{F}(x,y,\bar{y},\mu) :=$$

$$\mu f(x) + c(x)^T y + \bar{c}(x)^T \bar{y}$$

KKT residuals:

$$\rho(x,y,\bar{y},\mu)$$

$$\rho_k(d,y,\bar{y},\mu,H)$$

Local model of ϕ at x_k:

$$l_k(d,\mu)$$

Solution $(d_k, y_{k+1}, \bar{y}_{k+1})$ of (PQP) satisfies

- Test P2

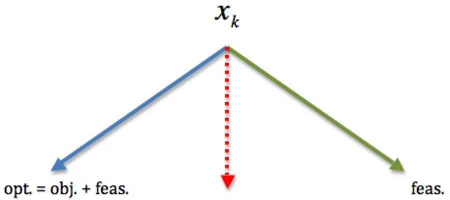
and solution $(d_k, y_{k+1}, \bar{y}_{k+1})$ of (FQP) satisfies

- Test F

then

- $d_k \leftarrow \tau d_k + (1-\tau) d_k$ so $\Delta l_k(d_k, 0) \geq \epsilon \Delta l_k(d_k, 0)$
- $\mu_{k+1} < \mu_k$ so $\Delta l_k(d_k, \mu_k) \geq \beta \Delta l_k(d_k, 0)$ for $\beta \in (0,1)$
- Overall:

$$\Delta l_k(d_k, \mu_k) \geq \beta \Delta l_k(d_k, 0) \geq \beta \epsilon \Delta l_k(d_k, 0)$$



"Inexact" SQO: Scenarios 5 & 6

(NLP):

$$\min_x f(x) \quad \text{s.t. } c(x) = 0,\ \bar{c}(x) \leq 0$$

(FP):

$$\min_x v(x) := \left\| \begin{bmatrix} c(x) \\ [\bar{c}(x)]^+ \end{bmatrix} \right\|_1$$

(PP):

$$\min_x \phi(x,\mu) := \mu f(x) + v(x)$$

(FJ):

$$\mathcal{F}(x,y,\bar{y},\mu) := \mu f(x) + c(x)^T y + \bar{c}(x)^T \bar{y}$$

KKT residuals:

$$\rho(x,y,\bar{y},\mu)$$

$$\rho_k(d,y,\bar{y},\mu,H)$$

Local model of ϕ at x_k:

$$l_k(d,\mu)$$

Solution $(d_k, y_{k+1}, \bar{y}_{k+1})$ of (PQP) satisfies

- Test P3

If $\|\rho(x_k, y_k, \bar{y}_k, \mu_k)\| < \zeta \|\rho(x_k, y_k, \bar{y}_k, 0)\|$, then

- $\mu_{k+1} \leftarrow \mu_k$ (Scenario 5)

else

- $\mu_{k+1} < \mu_k$ (Scenario 6)

"Inexact" SQO (iSQO)

repeat

(1) Check whether KKT point or infeasible stationary point has been obtained.

(2) Check for a trivial iteration.
  - (a) If Scenario 1 occurs, then go to step 6.

(3) Compute an inexact solution of (PQP) satisfying Test P1 or P3.
  - (a) If Scenario 2 occurs, then go to step 6.

(4) Compute an inexact solution of (FQP) satisfying Test F.
  - (a) If Scenario 3 occurs, then go to step 6.
  - (b) If Scenario 5 occurs, then go to step 6.

(5) Compute an inexact solution of (PQP) satisfying Test P2 or P3.
If Test 1 holds, then compute an inexact solution of (FQP) satisfying Test F.
  - (a) If Scenario 3 occurs, then go to step 6.
  - (b) If Scenario 4 occurs, then go to step 6.
  - (c) If Scenario 5 occurs, then go to step 6.
  - (d) Otherwise, Scenario 6 occurs.

(6) Perform a backtracking line search to reduce $\phi(\cdot, \mu_{k+1})$.

endrepeat

Well-posedness

Assumption

The following hold:

(1) *The functions f, c, and $\bar{c}$ are continuously differentiable in an open convex set Ω containing the sequences $\{x_k\}$ and $\{x_k + d_k\}$.*

(2) *The QP solver can solve* (PQP) *arbitrarily accurately.*

(3) *The QP solver can solve* (FQP) *arbitrarily accurately.*

Lemma

In iteration k, either **iSQO** *terminates or exactly one of Scenario 1–6 will occur.*

Theorem

One of the following holds:

1. **iSQO** *terminates with a KKT point or infeasible stationary point.*
2. **iSQO** *generates an infinite sequence of iterates*

$$\left(x_k, \begin{bmatrix} y_k \\ \bar{y}_k \end{bmatrix}, \begin{bmatrix} y_k \\ \bar{y}_k \end{bmatrix}, \mu_k\right) \textit{ where } \begin{bmatrix} y_k \\ y_k \end{bmatrix} \in [-e, e], \ \begin{bmatrix} \bar{y}_k \\ \bar{y}_k \end{bmatrix} \in [0, e], \ \textit{and } \mu_k > 0.$$

Global convergence

Assumption

(1) *The well-posedness assumptions still apply.*

(2) *The functions f, c, and $\bar{c}$ and their first derivatives are bounded and Lipschitz continuous in Ω (the open convex set containing $\{x_k\}$ and $\{x_k + d_k\}$).*

(3) *The sequences $\{H_k\}$ and $\{H_k\}$—including the initial and any modified values of H_k and H_k—are bounded.*

Lemma

The KKT residual for (FP) *vanishes:*

$$\lim_{k\to\infty} \|\rho(x_k, y_k, \bar{y}_k, 0)\| = 0.$$

Lemma

The KKT residual for (PP) vanishes:

$$\lim_{k\to\infty} \|\rho(x_k, y_k, \bar{y}_k, \mu_k)\| = 0.$$

Global convergence: Vanishing penalty parameter

(If $\{\mu_k\} \geq \underline{\mu}$ for some $\underline{\mu} > 0$, then we know the KKT residual for (PP) vanishes.)

Lemma

If $\mu_k \to 0$, then either all limit points of $\{x_k\}$ are feasible or all are infeasible.

Lemma

If $\mu_k \to 0$ and all limit points of $\{x_k\}$ are feasible, then, with

$$K_\mu := \{k : \mu_{k+1} < \mu_k\},$$

all limit points of $\{x_k\}_{k \in K_\mu}$ are FJ points.

Lemma

If $\rho(x_*, y_*, \bar{y}_*, 0) = 0$, x_* is feasible, and $(y_*, \bar{y}_*) \neq 0$, then the MFCQ fails at x_*.

Global convergence

Theorem

One of the following holds:

(a) $\mu_k = \underline{\mu}$ *for some* $\underline{\mu} > 0$ *for all large* k *and either every limit point of* $\{x_k\}$ *corresponds to a KKT point or is an infeasible stationary point;*

(b) $\mu_k \to 0$ *and every limit point of* $\{x_k\}$ *is an infeasible stationary point;*

(c) $\mu_k \to 0$*, all limit points of* $\{x_k\}$ *are feasible, and, with*

$$K_\mu := \{k : \mu_{k+1} < \mu_k\},$$

every limit point of $\{x_k\}_{k \in K_\mu}$ *corresponds to an FJ point where the MFCQ fails.*

Corollary

If $\{x_k\}$ *is bounded and every limit point of this sequence is a feasible point at which the MFCQ holds, then* $\mu_k = \underline{\mu}$ *for some* $\underline{\mu} > 0$ *for all large* k *and every limit point of* $\{x_k\}$ *corresponds to a KKT point.*

Implementation details

- Matlab implementation
- BQPD for QP solves with indefinite Hessians[10]
- *Simulated* inexactness by perturbing QP solutions
- Test set involves 307 CUTEr problems with
  - at least one free variable
  - at least one general (non-bound) constraint
  - at most 200 variables and constraints
- Termination conditions ($\epsilon_{tol} = 10^{-6}$ and $\epsilon_{\mu} = 10^{-8}$):

$$\|\rho(x_k, y_k, \bar{y}_k, \mu_k)\| \leq \epsilon_{tol} \quad \text{and} \quad v_k \leq \epsilon_{tol}; \qquad \text{(Optimal)}$$
$$\|\rho(x_k, y_k, \bar{y}_k, 0)\| = 0 \quad \text{and} \quad v_k > 0; \qquad \text{(Infeasible)}$$
$$\|\rho(x_k, y_k, \bar{y}_k, 0)\| \leq \epsilon_{tol} \quad \text{and} \quad v_k > \epsilon_{tol} \quad \text{and} \quad \mu_k \leq \epsilon_{\mu}. \qquad \text{(Infeasible)}$$

- Investigate performance of inexact algorithm with $\kappa = 0.01$, 0.1, and 0.5.

[10] Fletcher (2000)

Success statistics

Counts of termination messages for exact and three variants of inexact algorithm:

Termination message	Exact	Inexact		
		$\kappa = 0.01$	$\kappa = 0.1$	$\kappa = 0.5$
Optimal solution found	271	269	272	275
Infeasible stationary point found	4	3	2	2
Iteration limit reached	12	10	11	9
Subproblem solver failure	18	23	20	19

Termination statistics and reliability do not degrade with inexactness.

Inexactness levels

Observe relative residuals for QP solves:

$$\kappa_I := \frac{\|\rho_k(d, y, \bar{y}, \mu_k, H_k)\|}{\|\rho(x_k, y_k, \bar{y}_k, \mu_k)\|} \quad \text{or} \quad \kappa_I := \frac{\|\rho_k(d, y, \bar{y}, 0, H_k)\|}{\|\rho(x_k, y_k, \bar{y}_k, 0)\|},$$

For problem j, we compute minimum ($\kappa_I(j)$) and mean ($\bar{\kappa}_I(j)$) values over run:

min	κ	$\kappa_{I,\text{mean}}$	$[0, 10^{-8})$	$[10^{-8}, 10^{-6})$	$[10^{-6}, 10^{-4})$	$[10^{-4}, 10^{-3})$	$[10^{-3}, 0.01)$	$[0.01, 0.1)$	$[0.1, 0.5)$	$[0.5, 1)$	$[1, \infty)$
$\kappa_I(j)$	0.01	3.5e-03	0	2	10	7	253	0	0	0	0
	0.1	2.8e-02	0	0	2	10	30	232	0	0	0
	0.5	8.8e-02	0	0	2	4	23	69	179	0	0
mean	κ	$\bar{\kappa}_{I,\text{mean}}$									
$\bar{\kappa}_I(j)$	0.01	7.3e-03	0	0	0	0	254	18	0	0	0
	0.1	6.9e-02	0	0	0	0	0	261	13	0	0
	0.5	3.5e-01	0	0	0	0	0	1	264	12	0

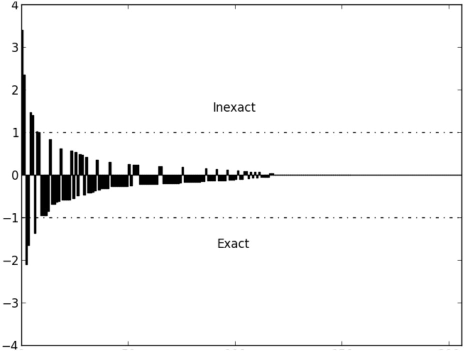
Relative residuals generally need only be moderately smaller than parameter κ.

Iteration comparison

Considering the logarithmic outperforming factor

$$r^j := -\log_2(\text{iter}^j_{\text{inexact}}/\text{iter}^j_{\text{exact}}),$$

we compare iteration counts of our inexact ($\kappa = 0.01$) and exact algorithms:



Iteration counts do not degrade significantly with inexactness.

Summary and future work

Inexact SQO method with inexact step computations:

- Handles rank deficiency, ill-conditioning, nonconvexity
- Allows generic inexactness in QP subproblem solves
- No specific QP solver required
- Global convergence guarantees
- Numerical experiments suggest inexact algorithm is reliable
- Inexact solutions do not degrade performance

Challenges ahead:

- What QP solver(s) to use? Active-set, interior-point, ADMM, IRWA[11], ...
- Monotonicity in QP solver?

[11] Burke, Curtis, Wang, Wang (2013)

Outline

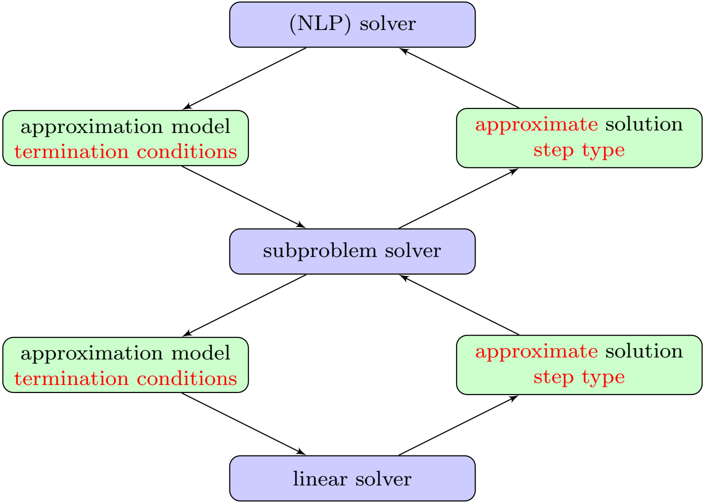
Introduction

Interior-Point Method

Sequential Quadratic Optimization Method

Summary

Algorithmic framework: Inexact



Thanks!

- J. V. Burke, F. E. Curtis, H. Wang, J. Wang, "Matrix-free Solvers for Exact Penalty Subproblems," in preparation.
- J. V. Burke, F. E. Curtis, and H. Wang, "A Sequential Quadratic Optimization Algorithm with Rapid Infeasibility Detection," second review for *SIAM Journal on Optimization*, submitted 2012.
- F. E. Curtis, T. C. Johnson, D. P. Robinson, and A. Wächter, "An Inexact Sequential Quadratic Optimization Algorithm for Large-Scale Nonlinear Optimization," first revision for *SIAM Journal on Optimization*, submitted 2013.
- F. E. Curtis, J. Huber, O. Schenk, and A. Wächter, "A Note on the Implementation of an Interior-Point Algorithm for Nonlinear Optimization with Inexact Step Computations," *Mathematical Programming, Series B*, Volume 136, Issue 1, pg. 209–227, 2012.
- F. E. Curtis, O. Schenk, and A. Wächter, "An Interior-Point Algorithm for Large-Scale Nonlinear Optimization with Inexact Step Computations," *SIAM Journal on Scientific Computing*, Volume 32, Issue 6, pg. 3447-3475, 2010.
- R. H. Byrd, F. E. Curtis, and J. Nocedal, "An Inexact Newton Method for Nonconvex Equality Constrained Optimization," *Mathematical Programming*, Volume 122, Issue 2, pg. 273-299, 2010.
- F. E. Curtis, J. Nocedal, and A. Wächter, "A Matrix-free Algorithm for Equality Constrained Optimization Problems with Rank-Deficient Jacobians," *SIAM Journal on Optimization*, Volume 20, Issue 3, pg. 1224-1249, 2009.
- R. H. Byrd, F. E. Curtis, and J. Nocedal, "An Inexact SQP Method for Equality Constrained Optimization," *SIAM Journal on Optimization*, Volume 19, Issue 1, pg. 351-369, 2008.