Infeasibility Detection and an Inexact Active-Set Method for Large-Scale Nonlinear Optimization

Frank E. Curtis, Lehigh University

involving joint work with

James V. Burke, University of Washington
Daniel P. Robinson, Johns Hopkins University
Andreas Wächter, Northwestern University
Hao Wang, Lehigh University

ISMP 2012, Berlin, Germany

20 August 2012

Outline

Outline

Motivation

Exact SQP

Inexact SQP

Summary

Two topics of this talk

1. Fill in a gap in algorithmic development for smaller-scale problems.

2. Propose a method for allowing/controlling inexactness for larger-scale problems.

Two contributions

1. Sequential quadratic optimization method with fast infeasibility detection
   - Guaranteed progress towards (linearized) feasibility each iteration
   - Carefully constructed update of a penalty parameter
   - Separate multipliers for feasibility and optimality
   - Global and fast local convergence for both feasible and infeasible problems
2. Active-set method with inexact step computations for large-scale problems
   - Framework extended from method above
   - Allows generic inexact solution of subproblems
   - Careful control of semi-smooth residual functions
   - Global and fast local convergence (at least, that's the plan!)

Outline

Constrained optimization

(OP):

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

(FP):

$$\min_x \ v(x)$$

where

$$v(x) := \left\| \begin{bmatrix} c(x) \\ \min\{x, 0\} \end{bmatrix} \right\|_1$$

(PP):

$$\min_x \ \phi(x, \mu)$$

where

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

(FJ):

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

Suppose that our optimization problem (OP) is infeasible:

- modeling errors
- data inconsistency
- perturbed/added constraints

Then, we want to solve a feasibility problem (FP).

Many algorithms/codes do this already, either by

- switching back-and-forth
- transitioning (via penalization)

Main contribution (Part 1)

Active-set method that completes the convergence picture for nonlinear optimization:

Problem type	Global convergence	Fast local convergence
Feasible	✓	✓
Infeasible	✓	?

Sequential quadratic optimization (SQO)

(OP):

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

(FP):

$$\min_x \ v(x)$$

where

$$v(x) := \left\| \begin{bmatrix} c(x) \\ \min\{x, 0\} \end{bmatrix} \right\|_1$$

(PP):

$$\min_x \ \phi(x, \mu)$$

where

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

(FJ):

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

Compute direction d and multipliers (y, z) for (OP) by solving

$$\min_d \ f_k + g_k^T d + \tfrac{1}{2} d^T H_k d$$
$$\text{s.t. } c_k + J_k^T d = 0,\ x_k + d \geq 0.$$

where

$$g(x) := \nabla f(x),$$
$$J(x) := \nabla c(x),$$
$$\text{and} \quad H(x, y, z, \mu) :\approx \nabla^2_{xx} \mathcal{F}(x, y, z, \mu).$$

Observe:

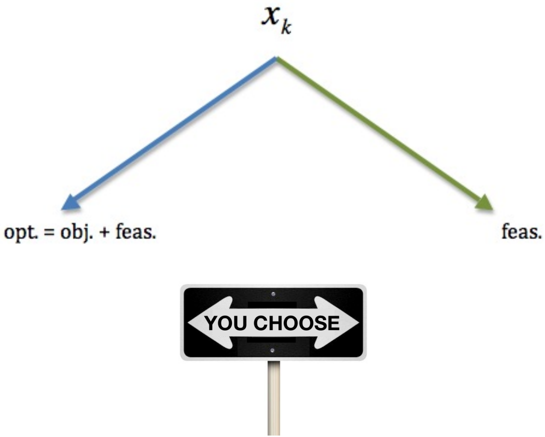
- May reduce to Newton's method if active set is identified.
- However, a globalization mechanism is needed.
- Moreover, this subproblem may be infeasible!

Recent literature

- (Rich history of SQO methods)
- Fletcher, Leyffer (2002)
- Byrd, Gould, Nocedal (2005)
- Byrd, Nocedal, Waltz (2008)
- Byrd, Curtis, Nocedal (2010)
- Byrd, López-Calva, Nocedal (2010)
- Gould, Robinson (2010)
- Morales, Nocedal, Wu (2010)

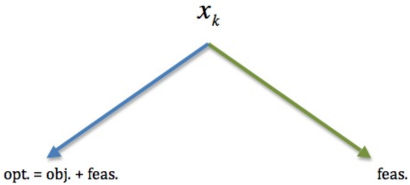
Issue faced by all optimization solvers

Move towards feasibility and/or objective decrease?

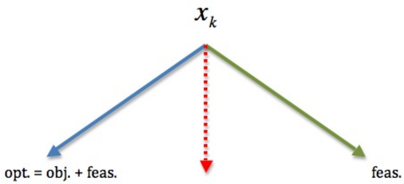


FILTER and steering strategies

- FILTER: "... we make use of a property of the phase I algorithm in our QP solver. If an infeasible QP is detected, [a feasibility restoration phase is entered]."

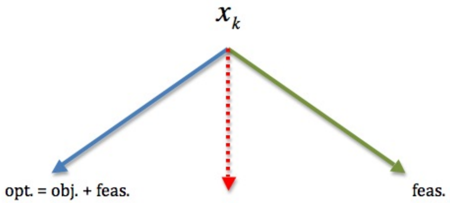


- Steering methods solve a sequence of constrained subproblems:



Our approach: Two-phase strategy

- "Feasibility step" to determine possible progress toward feasibility.
- "Optimality step" exploits information obtained by "feasibility step".



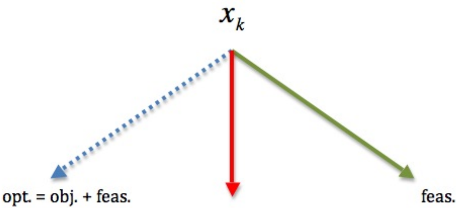
- Objective function never ignored (unlike `FILTER`).
- At most two subproblems solved per iteration (unlike steering).
- Reduces to SQO for optimization problem in feasible cases.
- Reduces to (perturbed) SQO for feasibility problem in infeasible cases.

Ensuring global convergence

(1) Compute feasibility step $\overline{d}_k$ to determine highest level of feasibility improvement.

(2) Compute optimality step $\widehat{d}_k$.

(3) Set $d_k \leftarrow \tau_k \widehat{d}_k + (1 - \tau_k)\overline{d}_k$ to obtain proportional feasibility improvement.



(4) Update penalty parameter to ensure sufficient decrease in a merit function.

Feasibility step

(OP):

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

(FP):

$$\min_x \ v(x)$$

where

$$v(x) := \left\| \begin{bmatrix} c(x) \\ \min\{x, 0\} \end{bmatrix} \right\|_1$$

(PP):

$$\min_x \ \phi(x, \mu)$$

where

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

(FJ):

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

(1) Compute feasibility step $\overline{d}_k$.

- Solve for $(\overline{d}_k, \overline{r}_k, \overline{s}_k, \overline{t}_k)$ and $(\overline{y}^{\mathcal{E}}_{k+1}, \overline{z}^{\mathcal{I}}_{k+1})$:

$$\min_{d,r,s,t} \ e^T(r+s) + e^T t + \tfrac{1}{2} d^T \overline{H}_k d$$

$$\text{s.t.} \begin{cases} c_k + J_k^T d = r - s \\ x_k + d \geq -t \\ (r, s, t) \geq 0. \end{cases} \qquad \text{(QO1)}$$

- Resulting $\overline{d}_k$ yields a reduction in a local model of v:

$$l_k(d) := \left\| \begin{bmatrix} c_k + J_k^T d \\ \min\{x_k + d, 0\} \end{bmatrix} \right\|_1 .$$

That is, it yields

$$\Delta l_k(\overline{d}_k) = l_k(0) - l_k(\overline{d}_k) \geq 0.$$

Optimality step

(OP):

$$\min_x f(x)$$

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

(FP):

$$\min_x \; v(x)$$

where

$$v(x) := \left\| \begin{bmatrix} c(x) \\ \min\{x, 0\} \end{bmatrix} \right\|_1$$

(PP):

$$\min_x \; \phi(x, \mu)$$

where

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

(FJ):

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

(2) Compute optimality step $\widehat{d}_k$.

- Determine $\mathcal{E}_k$ and $\mathcal{I}_k$ for which $\overline{d}_k$ is linearly feasible:

$$c_k^{\mathcal{E}_k} + {J_k^{\mathcal{E}_k}}^T \overline{d}_k = 0, \; x_k^{\mathcal{I}_k} + \overline{d}_k^{\mathcal{I}_k} \geq 0.$$

- Solve for $(\widehat{d}_k, \widehat{r}_k^{\mathcal{E}_k^c}, \widehat{s}_k^{\mathcal{E}_k^c}, \widehat{t}_k^{\mathcal{I}_k^c})$ and $(\widehat{y}_{k+1}^{\mathcal{E}}, \widehat{z}_{k+1}^{\mathcal{I}})$:

$$\min_{d, r^{\mathcal{E}_k^c}, s^{\mathcal{E}_k^c}, t^{\mathcal{I}_k^c}} \; \mu_k g_k^T d + e^T (r^{\mathcal{E}_k^c} + s^{\mathcal{E}_k^c}) + e^T t^{\mathcal{I}_k^c} + \tfrac{1}{2} d^T \widehat{H}_k d$$

$$\text{s.t. } \begin{cases} c_k^{\mathcal{E}_k} + {J_k^{\mathcal{E}_k}}^T d = 0 \\ c_k^{\mathcal{E}_k^c} + {J_k^{\mathcal{E}_k^c}}^T d = r^{\mathcal{E}_k^c} - s^{\mathcal{E}_k^c} \\ x_k^{\mathcal{I}_k} + d^{\mathcal{I}_k} \geq 0 \\ x_k^{\mathcal{I}_k^c} + d^{\mathcal{I}_k^c} \geq -t^{\mathcal{I}_k^c} \\ (r^{\mathcal{E}_k^c}, s^{\mathcal{E}_k^c}, t^{\mathcal{I}_k^c}) \geq 0. \end{cases} \qquad \text{(QO2)}$$

Search direction

(OP):

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

(FP):

$$\min_x \; v(x)$$

where

$$v(x) := \left\| \begin{bmatrix} c(x) \\ \min\{x, 0\} \end{bmatrix} \right\|_1$$

(PP):

$$\min_x \; \phi(x, \mu)$$

where

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

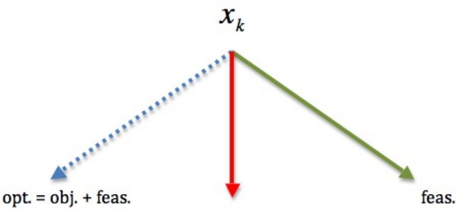
(FJ):

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

(3) Set $d_k \leftarrow \tau_k \widehat{d}_k + (1 - \tau_k)\overline{d}_k$.

- Choose $\beta \in (0, 1)$.
- Find largest $\tau_k \in [0, 1]$ such that d_k satisfies

$$\Delta l_k(d_k) \geq \beta \Delta l_k(\overline{d}_k).$$



μ update

(OP):

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

(FP):

$$\min_x\ v(x)$$

where

$$v(x) := \left\| \begin{bmatrix} c(x) \\ \min\{x, 0\} \end{bmatrix} \right\|_1$$

(PP):

$$\min_x\ \phi(x, \mu)$$

where

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

(FJ):

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

(4) Update penalty parameter to ensure sufficient decrease.

- Set μ_{k+1} so that

$$\mu_{k+1} \leq \frac{1}{\|(\widehat{y}_{k+1}, \widehat{z}_{k+1})\|_\infty}.$$

- Let

$$m_k(d, \mu) = \mu(f(x_k) + g_k^T d) + l_k(d),$$

then set μ_{k+1} so that d_k yields

$$\Delta m_k(d_k, \mu_{k+1}) \geq \epsilon \Delta l_k(d_k).$$

This ensures sufficient decrease in $\phi(\cdot, \mu_{k+1})$ from x_k.

Ensuring global convergence

(1) Determine potential level of feasibility improvement:

$$\Delta l_k(\overline{d}_k) \geq 0.$$

(2) Compute search direction satisfying

$$\Delta l_k(d_k) \geq \beta \Delta l_k(\overline{d}_k).$$

(3) Update penalty parameter (if necessary) so that

$$\Delta m_k(d_k, \mu_{k+1}) \geq \epsilon \Delta l_k(d_k).$$

(4) Perform line search on $\phi(\cdot, \mu_{k+1})$ at x_k along the descent direction d_k.

Ensuring fast local convergence: Feasible case

If (OP) is feasible, then:

(a) (QO1) eventually produces linearly feasible directions

(b) μ eventually remains constant

(c) (QO2) reduces to a standard SQO subproblem

$$\min_{d, r^{\mathcal{E}_k^c}, s^{\mathcal{E}_k^c}, t^{\mathcal{I}_k^c}} \mu_k g_k^T d + e^T (r^{\mathcal{E}_k^c} + s^{\mathcal{E}_k^c}) + e^T t^{\mathcal{I}_k^c} + \tfrac{1}{2} d^T \widehat{H}_k d$$

$$\text{s.t.} \begin{cases} c_k^{\mathcal{E}_k} + J_k^{\mathcal{E}_k\,T} d = 0 \\ c_k^{\mathcal{E}_k^c} + J_k^{\mathcal{E}_k^c\,T} d = r^{\mathcal{E}_k^c} - s^{\mathcal{E}_k^c} \\ x_k^{\mathcal{I}_k} + d^{\mathcal{I}_k} \geq 0 \\ x_k^{\mathcal{I}_k^c} + d^{\mathcal{I}_k^c} \geq -t^{\mathcal{I}_k^c} \\ (r^{\mathcal{E}_k^c}, s^{\mathcal{E}_k^c}, t^{\mathcal{I}_k^c}) \geq 0. \end{cases} \tag{QO2}$$

Ensuring fast local convergence: Infeasible case

If (OP) is infeasible, then:

(a) (QO1) eventually yields small improvement towards feasibility

(b) $\mu \to 0$ and $(\widehat{y}_k, \widehat{z}_k) \to (\overline{y}_k, \overline{z}_k)$:

- If $v_k \neq 0$ and $\Delta l_k(\overline{d}_k) \leq \theta v_k$, then

$$\begin{aligned} \mu_k &\leq \text{KKT}_{inf}(x_k, \overline{y}_{k+1}, \overline{z}_{k+1})^2 \\ \|(\widehat{y}_k, \widehat{z}_k) - (\overline{y}_k, \overline{z}_k)\| &\leq \text{KKT}_{inf}(x_k, \overline{y}_{k+1}, \overline{z}_{k+1})^2. \end{aligned}$$

(c) (QO2) reduces to (QO1) (i.e., SQO for feasibility)

SQuID

Sequential Quadratic Optimization with Fast Infeasibility Detection

(1) Compute feasibility step via (QO1).
(2) Check whether infeasible stationary point has been obtained.
(3) Update μ_k, $\widehat{y}_k$, and $\widehat{z}_k$, if necessary (for fast local convergence).
(4) Compute optimality step via (QO2).
(5) Check whether optimal solution has been obtained.
(6) Compute combination of feasibility and optimality steps (for global convergence).
(7) Update μ_k, if necessary (for global convergence).
(8) Perform line search to obtain decrease in merit function.

Global convergence

Assumptions:

(1) The problem functions f and c and their first-order derivatives are bounded and Lipschitz continuous in a convex set containing $\{x_k\}$.

(2) There exist $\mu_{\max} \geq \mu_{\min} > 0$ such that, for any d,

$$\mu_{\min}\|d\|^2 \leq d^T \overline{H}_k d \leq \mu_{\max}\|d\|^2$$
$$\mu_{\min}\|d\|^2 \leq d^T \widehat{H}_k d \leq \mu_{\max}\|d\|^2.$$

Theorem

Either all limit points of $\{x_k\}$ are feasible or all are infeasible stationary.

Theorem

If $\mu_k \geq \mu_*$ for some $\mu_* > 0$ for all k, then every feasible limit point is a KKT point.

Theorem

Suppose $\mu_k \to 0$ and let K_μ be the subsequence of iterations during which the penalty parameter μ_k is decreased. Then, if all limit points of $\{x_k\}$ are feasible, then all limit points of $\{x_k\}_{k \in K_\mu}$ correspond to Fritz John points for (OP) where MFCQ fails.

Local convergence: Assumptions

(1) f and c and their first and second derivatives are bounded and Lipschitz continuous in an open convex set containing a given point of interest x_*.

(2) If $(x_*, \overline{y}_*, \overline{z}_*)$ is a KKT point for (FP), then

- (a) ${J_*^{\mathcal{Z}_* \cup \mathcal{A}_*}}^T$ has full row rank.
- (b) $-e < \overline{y}_*^{\mathcal{Z}_*} < e$ and $0 < \overline{z}_*^{\mathcal{A}_*} < e$.
- (c) $d^T \overline{H}_* d > 0$ for all $d \neq 0$ such that ${J_*^{\mathcal{Z}_* \cup \mathcal{A}_*}}^T d = 0$.

(3) If $(x_*, \widehat{y}_*, \widehat{z}_*, \mu_*)$ is a KKT point for (OP), then (2) holds, $\mu_k \to \mu_* > 0$, and

- (a) $\widehat{z}_*^{\mathcal{A}_*} + c_*^{\mathcal{A}_*} > 0$.
- (b) $d^T \widehat{H}_* d > 0$ for all $d \neq 0$ such that ${J_*^{\mathcal{E}_* \cup \mathcal{A}_*}}^T d = 0$.

Local convergence

Theorem

If $v_* > 0$, and $(x_k, \overline{y}_k, \overline{z}_k)$ and $(x_k, \widehat{y}_k, \widehat{z}_k)$ are each close to $(x_*, \overline{y}_*, \overline{z}_*)$, then

$$\left\| \begin{bmatrix} x_{k+1} - x_* \\ \overline{y}_{k+1} - \overline{y}_* \\ \overline{z}_{k+1} - \overline{z}_* \end{bmatrix} \right\| \leq C \left\| \begin{bmatrix} x_k - x_* \\ \overline{y}_k - \overline{y}_* \\ \overline{z}_k - \overline{z}_* \end{bmatrix} \right\|^2 + O\left(\left\| \begin{bmatrix} \widehat{y}_k - \overline{y}_k \\ \widehat{z}_k - \overline{z}_k \end{bmatrix} \right\| \right) + O(\mu)$$

for some constant $C > 0$ independent of k.

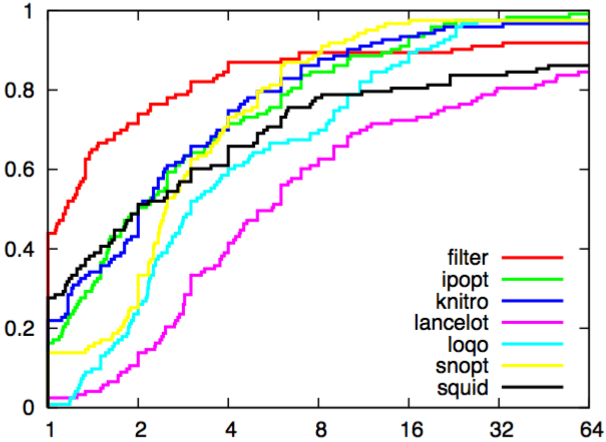
Theorem

If $(x_k, \overline{y}_k, \overline{z}_k)$ is close to $(x_*, \overline{y}_*, \overline{z}_*)$ and $(x_k, \widehat{y}_k, \widehat{z}_k)$ is close to $(x_*, \widehat{y}_*, \widehat{z}_*)$, then

$$\left\| \begin{bmatrix} x_{k+1} - x_* \\ \widehat{y}_{k+1} - \widehat{y}_* \\ \widehat{z}_{k+1} - \widehat{z}_* \end{bmatrix} \right\| \leq C \left\| \begin{bmatrix} x_k - x_* \\ \widehat{y}_k - \widehat{y}_* \\ \widehat{z}_k - \widehat{z}_* \end{bmatrix} \right\|^2$$

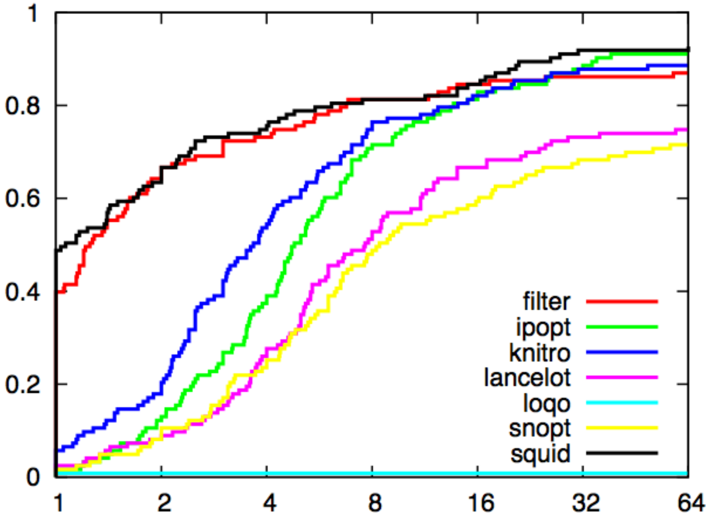
for some constant $C > 0$ independent of k.

Numerical experiments: Feasible optimization problems



`FILTER`:	9 "failures" due to declaration of infeasibility
`Others`:	≤ 2 "failures" due to declaration of infeasibility
`SQuID`:	Lack of robustness due to lack of SOC and simplistic handling of nonconvexity

Numerical experiments: Infeasible optimization problems



Outline

Motivation

Exact SQP

Inexact SQP

Summary

Large-scale optimization

You're thinking to yourself:

- SQO is expensive enough with one subproblem to solve.
- You want me to solve two?!

Our response:

- Allowing inexactness gives flexibility in the subproblem solves.
- Any work to solve one subproblem is better split between our two.

Optimality conditions

(OP):

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

(FP):

$$\min_x \; v(x) \; \text{ s.t. } \; x \geq 0$$

where

$$v(x) := \|c(x)\|_1$$

(PP):

$$\min_x \; \phi(x, \mu) \; \text{ s.t. } \; x \geq 0$$

where

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

(FJ):

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

Optimality conditions for both (PP) and (FP) involve

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

where

$$[c(x)]^+ := \max\{c(x), 0\}$$

$$\text{and } \; [c(x)]^- := \max\{-c(x), 0\}.$$

Observe:

- If $\rho(x, y, \mu) = 0$, $v(x) = 0$, and $\mu > 0$, then

  $(x, y/\mu)$ is a KKT point for (OP).

- If $\rho(x, y, \mu) = 0$, $v(x) > 0$, and $\mu = 0$, then

  x is an infeasible stationary point.

Two subproblems

(OP):

$$\min_x f(x)$$

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

(FP):

$$\min_x \ v(x) \ \text{ s.t. } \ x \geq 0$$

where

$$v(x) := \|c(x)\|_1$$

(PP):

$$\min_x \ \phi(x,\mu) \ \text{ s.t. } \ x \geq 0$$

where

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

(FJ):

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

$$\mu f(x) + y^T c(x) - z^T x$$

Define the model of $\phi(\cdot,\mu)$ at x_k given by

$$m_k(d,\mu) := \mu(f_k + g_k^T d) + \|c_k + J_k d\|_1.$$

We iterate by (approximately) solving two subproblems:

- Feasibility subproblem:

$$\min_d \ -\Delta m_k(d,0) + \tfrac{1}{2} d^T \overline{H}_k d \ \text{ s.t. } \ x_k + d \geq 0. \quad \text{(QO1)}$$

- Optimality subproblem:

$$\min_d \ -\Delta m_k(d,\mu_k) + \tfrac{1}{2} d^T \widehat{H}_k d \ \text{ s.t. } \ x_k + d \geq 0. \quad \text{(QO2)}$$

Optimality conditions for both (QO1) and (QO2) involve

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

Main contribution (Part 2)

Active-set method that allows generic inexact subproblem solves:

- "Termination tests" dictate when an inexact solution is acceptable.
- Tests primarily monitor the reductions

$$\Delta m_k(\overline{d}_k, 0) \text{ and } \Delta m_k(\widehat{d}_k, \mu_k)$$

  and the residuals

$$\rho_k(\overline{d}_k, \overline{y}_{k+1}, 0, \overline{H}_k) \text{ and } \rho_k(\widehat{d}_k, \widehat{y}_{k+1}, \mu_k, \widehat{H}_k).$$

- Essentially, these ensure primal and dual convergence, respectively.
- ("Exact case": residuals are always zero, so we focused on model reductions.)

Termination test 1: High-level

The primal-dual pairs $(\overline{d}_k, \overline{y}_{k+1})$ and $(\widehat{d}_k, \widehat{y}_{k+1})$ are acceptable if

- the feasibility step
  - yields a nontrivial reduction in the feasibility model
  - the error in solving (QO1) is small compared to the error in solving (FP)
- the optimality step
  - yields a nontrivial reduction in the penalty function model
  - the error in solving (QO2) is small compared to the error in solving (PP)
- and, in addition,
  - the optimality step yields proportional improvement in the feasibility model.

Termination test 1: Low-level

The primal-dual pairs $(\overline{d}_k, \overline{y}_{k+1})$ and $(\widehat{d}_k, \widehat{y}_{k+1})$ are acceptable if

- the feasibility step satisfies

$$\Delta m_k(\overline{d}_k, 0) \geq \theta \|\overline{d}_k\|^2 > 0$$
$$\|\rho_k(\overline{d}_k, \overline{y}_{k+1}, 0, \overline{H}_k)\| \leq \kappa \|\rho(x_k, \overline{y}_k, 0)\|$$

- the optimality step satisfies

$$\Delta m_k(\widehat{d}_k, \mu_k) \geq \theta \|\widehat{d}_k\|^2 > 0$$
$$\|\rho_k(\widehat{d}_k, \widehat{y}_{k+1}, \mu_k, \widehat{H}_k)\| \leq \kappa \|\rho(x_k, \widehat{y}_k, \mu_k)\|$$

- and, in addition,

$$\Delta m_k(\widehat{d}_k, \mu_k) \geq \epsilon \Delta m_k(\overline{d}_k, 0)$$

where $\theta > 0$, $\kappa \in (0, 1)$, and $\epsilon \in (0, 1)$ are user-defined constants.

Termination test 2: High-level

The primal-dual pairs $(\overline{d}_k, \overline{y}_{k+1})$ and $(\widehat{d}_k, \widehat{y}_{k+1})$ are acceptable if

- the feasibility step
  - yields a nontrivial reduction in the feasibility model
  - the error in solving (QO1) is small compared to the error in solving (FP)
- the optimality step
  - yields a nontrivial reduction in the penalty function model
  - the error in solving (QO2) is small compared to the error in solving (FP)
- and, in addition, we
  - take a convex combination of the steps for proportional improvement toward feasibility
  - (potentially) decrease the penalty parameter.

Termination test 2: Low-level

The primal-dual pairs $(\overline{d}_k, \overline{y}_{k+1})$ and $(\widehat{d}_k, \widehat{y}_{k+1})$ are acceptable if

- the feasibility step satisfies

$$\Delta m_k(\overline{d}_k, 0) \geq \theta \|\overline{d}_k\|^2 > 0$$
$$\|\rho_k(\overline{d}_k, \overline{y}_{k+1}, 0, \overline{H}_k)\| \leq \kappa \|\rho(x_k, \overline{y}_k, 0)\|$$

- the optimality step satisfies

$$\Delta m_k(\widehat{d}_k, \mu_k) \geq \theta \|\widehat{d}_k\|^2 > 0$$
$$\|\rho_k(\widehat{d}_k, \widehat{y}_{k+1}, \mu_k, \widehat{H}_k)\| \leq \kappa \|\rho(x_k, \widehat{y}_k, \mu_k)\|$$

- and, in addition,

$$d_k \leftarrow \tau_k \widehat{d}_k + (1 - \tau_k)\overline{d}_k \text{ so that } \Delta m_k(d_k, 0) \geq \epsilon \Delta m_k(\overline{d}_k, 0)$$
$$\mu_{k+1} \leftarrow \begin{cases} \mu_k & \text{if } \Delta m_k(d_k, \mu_k) \geq \beta \Delta m_k(d_k, 0) \\ \min\left\{\delta \mu_k, \frac{(1-\beta)\Delta m_k(d_k,0)}{g_k^T d_k + \theta \|d_k\|^2}\right\} & \text{otherwise} \end{cases}$$

where $\theta > 0$, $\kappa \in (0,1)$, $\epsilon \in (0,1)$, $\beta \in (0,1)$, and $\delta \in (0,1)$ are constants.

Termination test 3: High-level

The primal-dual pairs $(\overline{d}_k, \overline{y}_{k+1})$ and $(\widehat{d}_k, \widehat{y}_{k+1})$ are acceptable if

- the feasibility step is zero
- the optimality step is zero
- the feasibility multipliers are not changed
- and, in addition,
  - the error in solving (QO2) is small compared to the error in solving (PP)

Termination test 3: Low-level

The primal-dual pairs $(\overline{d}_k, \overline{y}_{k+1})$ and $(\widehat{d}_k, \widehat{y}_{k+1})$ are acceptable if

- $\overline{d}_k \leftarrow 0$
- $\widehat{d}_k \leftarrow 0$
- $\overline{y}_{k+1} \leftarrow \overline{y}_k$
- and, in addition,

$$\|\rho_k(0, \widehat{y}_{k+1}, \mu_k, \widehat{H}_k)\| \leq \kappa \|\rho(x_k, \widehat{y}_k, \mu_k)\|$$

where $\kappa \in (0, 1)$ is a constant.

Inexact SQO

(1) Check whether infeasible stationary point or optimal solution obtained.

(2) Compute optimality and feasibility steps so that TT1, TT2, or TT3 holds.
  - Hessian modifications performed within subproblems solves.

(3) Update $\overline{y}_k$, and $\widehat{y}_k$, if necessary (details suppressed).

(4) Compute combination of feasibility and optimality steps.

(5) Update μ_k, if necessary.

(5) Perform line search to obtain decrease in merit function.

Global convergence: Assumptions

(1) The problem functions f and c and their first-order derivatives are bounded and Lipschitz continuous in a convex set containing $\{x_k\}$.

(2) $\{\overline{H}_k\}$ and $\{\widehat{H}_k\}$ are bounded.

(3) If x_k is stationary for the penalty problem, then the subproblem solver eventually produces a direction of insufficient curvature or y such that $\rho_k(0, y, \mu_k, \widehat{H}_k)$ is arbitrarily small. Else, it eventually produces a direction of insufficient curvature or (d, y) such that $\Delta m_k(d, \mu_k) \geq \theta\|d\|^2$ with $\rho_k(d, y, \mu_k, \widehat{H}_k)$ arbitrarily small.

(4) For the feasibility subproblem, the subproblem solver eventually produces a direction of insufficient curvature or, for $\xi_k > 0$, it produces (d, y) satisfying $\max\{\xi_k, \Delta m_k(d, 0)\} \geq \theta\|d\|^2$ with $\rho_k(d, y, 0, \overline{H}_k)$ arbitrarily small.

Global convergence

Theorem

The following limit always holds:

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

Theorem

The following limit holds if $\mu_k = \underline{\mu} > 0$ *for all large* k*:*

$$\lim_{k\to\infty} \|\rho(x_k, \widehat{y}_k, \underline{\mu})\| = 0.$$

TO DO:

- Ensure that μ stays bounded below under “nice” conditions.
- Local convergence(?!)

Outline

Summary

- Developed an SQO method that completes the convergence picture for NLO:

Problem type	Global convergence	Fast local convergence
Feasible	✓	✓
Infeasible	✓	✓

- Proposed extensions for solving large-scale problems with inexact calculations.

Thanks!!

Infeasibility detection:

- J.V. Burke, F.E. Curtis, and H. Wang, "A Sequential Quadratic Optimization Algorithm with Rapid Infeasibility Detection," submitted to *SIAM Journal on Optimization*, 2012.
- F. E. Curtis, "A Penalty-Interior-Point Algorithm for Nonlinear Constrained Optimization," Mathematical Programming Computation, Volume 4, Issue 2, pg. 181-209, 2012.
- R. H. Byrd, F. E. Curtis, and J. Nocedal, "Infeasibility Detection and SQP Methods for Nonlinear Optimization," SIAM Journal on Optimization, Volume 20, Issue 5, pg. 2281-2299, 2010.

Inexactness and constrained optimization:

- F.E. Curtis, D.P. Robinson, and A. Wächter, "An Inexact Sequential Quadratic Optimization Algorithm for Large-Scale Nonlinear Optimization," in preparation.
- F. E. Curtis, J. Huber, O. Schenk, and A. Wchter, "On the Implementation of an Interior-Point Algorithm for Nonlinear Optimization with Inexact Step Computations," to appear in Mathematical Programming, Series B, 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.