# An Inexact Active-Set Method for Large-Scale 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

University of Edinburgh, ERGO Seminar

17 August 2012

# Outline

# Outline

# Nonlinear constrained optimization

Decades of algorithmic and theoretical development:

- Sequential quadratic, interior-point, etc., with numerous variations
- Global and fast local convergence guarantees
- Multiple popular, successful software packages

So what's left for smooth constrained nonlinear optimization?

- PDE-constrained optimization
- Large-scale network problems
- Mixed-integer optimization
- Parametric optimization
- Real-time optimization

# 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!)

# Prior work

Infeasibility detection:

- 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, 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.

# Outline

## Constrained optimization

This is how we formulate constrained optimization problems:

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

We also like to think that the Karush-Kuhn-Tucker (KKT) conditions are necessary.

# Constraint violation minimization

What if the constraints are infeasible?

- modeling errors
- data inconsistency
- perturbed/added constraints

Then, we want to solve

$$\min_x \; v(x) := \|c(x)\|_1 + \|\min\{x, 0\}\|_1. \tag{FP}$$

Many algorithms/codes do this already, either by

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

But are they doing it efficiently?

# Numerical experiments: Infeasible optimization problems

Iterations and evaluations for 8 infeasible optimization problems (2-3 variables):

| Prob. | Ipopt | | Knitro | | Filter | |
|---|---|---|---|---|---|---|
| | Iter. | Eval. | Iter. | Eval. | Iter. | Eval. |
| 1 | 48 | 281 | 38 | 135 | 16 | 16 |
| 2 | 109 | 170 | *10000 | *40544 | 12 | 12 |
| 3 | 788 | 3129 | 12 | 83 | 10 | 10 |
| 4 | 46 | 105 | 25 | 61 | 11 | 11 |
| 5 | 72 | 266 | *1060 | *3401 | 26 | 26 |
| 6 | 63 | 141 | *76 | *264 | 27 | 27 |
| 7 | 87 | 152 | *10000 | *43652 | 30 | 30 |
| 8 | 104 | 206 | 33 | 97 | 28 | 28 |

Problems also run with `SNOPT` and `LOQO`, but they failed every time.

# Numerical experiments: Feasibility problems (solved directly)

Iterations and evaluations for 8 feasibility problems (2-3 variables):

| Problem | Ipopt | | Knitro | | Filter | |
|---|---|---|---|---|---|---|
| | Iter. | Eval. | Iter. | Eval. | Iter. | Eval. |
| 1 | 28 | 29 | 14 | 15 | 17 | 21 |
| 2 | 31 | 32 | 31 | 33 | 12 | 13 |
| 3 | 50 | 131 | 10 | 11 | 12 | 13 |
| 4 | 24 | 79 | 18 | 29 | 10 | 12 |
| 5 | 166 | 786 | 29 | 40 | 30 | 32 |
| 6 | 37 | 48 | 20 | 21 | 26 | 27 |
| 7 | 59 | 65 | 31 | 34 | 25 | 28 |
| 8 | 46 | 71 | 19 | 20 | 26 | 29 |

$\Longrightarrow$ If we can switch/transition efficiently, then our current tools work well.

# 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)

Define the Fritz John function

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

Compute search direction $d$ and multiplier $y$ for (OP) by solving

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

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

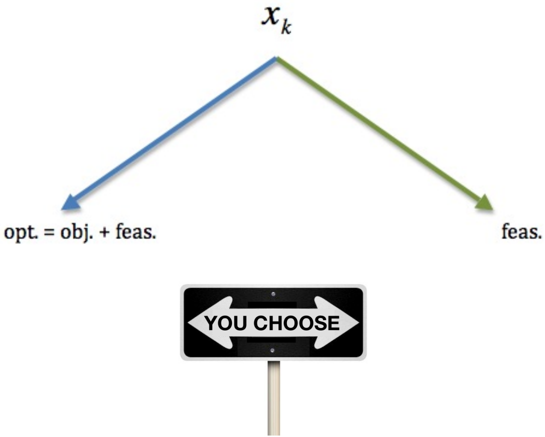
- May reduce to Newton's method once the active set is identified.
- However, a globalization mechanism is needed.
- Moreover, this subproblem may be infeasible!

# 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)

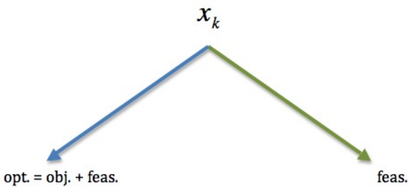
## Issue faced by all optimization solvers

Move towards feasibility and/or objective decrease?

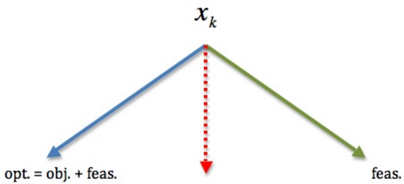


# FILTER and steering strategies

- 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]."

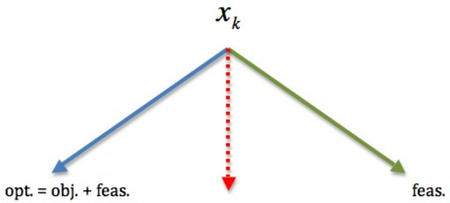



- Steering methods solve a sequence of constrained subproblems:



## Our approach: Two-phase strategy

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



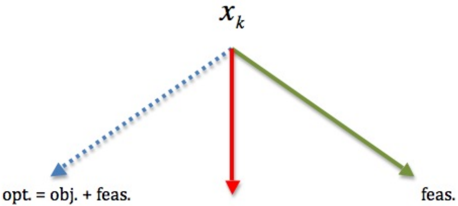

- Objective function never ignored (unlike FILTER).
- At most two subproblems solved per iteration (unlike steering).
- Reduces to SQO for optimization problem in feasible cases.
- Reduces to (perturbed) SQO for feasibility problem in infeasible cases.

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




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

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

# Feasibility step

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

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

$$\min_{d,r,s,t} \; e^T(r+s) + e^T t + \tfrac{1}{2} d^T \overline{H}_k d \quad \text{s.t.} \begin{cases} c_k + J_k^T d = r - s \\ x_k + d \geq -t \\ (r,s,t) \geq 0. \end{cases} \tag{QO1}$$

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

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

That is, it yields

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

## Optimality step

(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 $(\widehat{y}_{k+1}^{\mathcal{E}}, \widehat{z}_{k+1}^{\mathcal{I}})$:

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

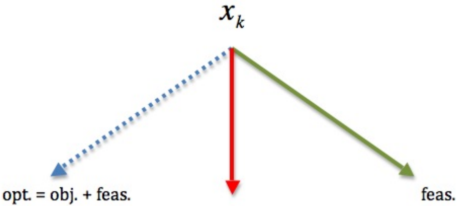
$$\text{s.t.} \; \begin{cases} c_k^{\mathcal{E}_k} + {J_k^{\mathcal{E}_k}}^T d = 0 \\ c_k^{\mathcal{E}_k^c} + {J_k^{\mathcal{E}_k^c}}^T d = r^{\mathcal{E}_k^c} - s^{\mathcal{E}_k^c} \\ x_k^{\mathcal{I}_k} + d^{\mathcal{I}_k} \geq 0 \\ x_k^{\mathcal{I}_k^c} + d^{\mathcal{I}_k^c} \geq -t^{\mathcal{I}_k^c}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

## Search direction

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

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

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



## $\mu$ update

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

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

- Set $\mu_{k+1}$ 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 $\mu_{k+1}$ 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 fast local convergence

Under "nice" conditions:

(1) If (OP) is feasible, then:
  (a) (QO1) eventually produces linearly feasible directions
  (b) $\mu$ eventually remains constant
  (c) (QO2) reduces to a standard SQO subproblem

(2) If (OP) is infeasible, then:
  (a) (QO1) eventually yields small improvement towards feasibility
  (b) $\mu$ is driven to zero
  (c) (QO2) reduces to (QO1) (i.e., SQO for feasibility)

# Feasible case

(1) If (OP) is feasible, then:
   (a) (QO1) eventually produces linearly feasible directions
   (b) $\mu$ eventually remains constant
   (c) (QO2) reduces to a standard SQO subproblem

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

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

# Infeasible case

(2) If (OP) is infeasible, then:
   (a) (QO1) eventually yields small improvement towards feasibility
   (b) $\mu$ is driven to zero
   (c) (QO2) reduces to (QO1) (i.e., SQO for feasibility)

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

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

# 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 $\mu_k$, $\widehat{y}_k$, 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 $\mu_k$, 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 $d$,

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

## Global convergence

**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* $\mu_k \rightarrow 0$ *and let* $K_\mu$ *be the subsequence of iterations during which the penalty parameter* $\mu_k$ *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 sufficiently 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 sufficiently close to $(x_*, \overline{y}_*, \overline{z}_*)$ and $(x_k, \widehat{y}_k, \widehat{z}_k)$ is sufficiently close to $(x_*, \widehat{y}_*, \widehat{z}_*)$, then*

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

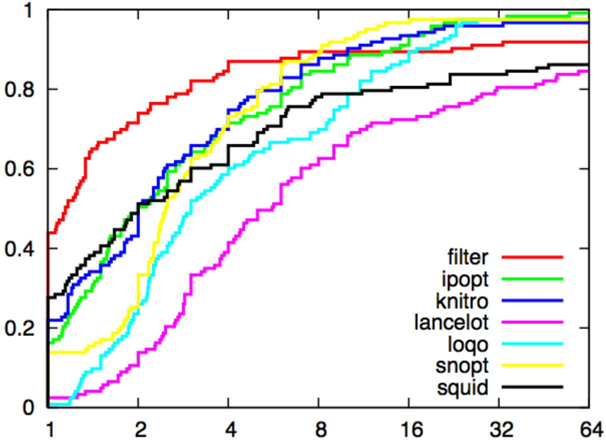
*for some constant $C > 0$ independent of $k$.*

# Numerical experiments: Infeasible optimization problems

Iterations and $f$ evaluations for 8 infeasible optimization problems (2-3 variables):

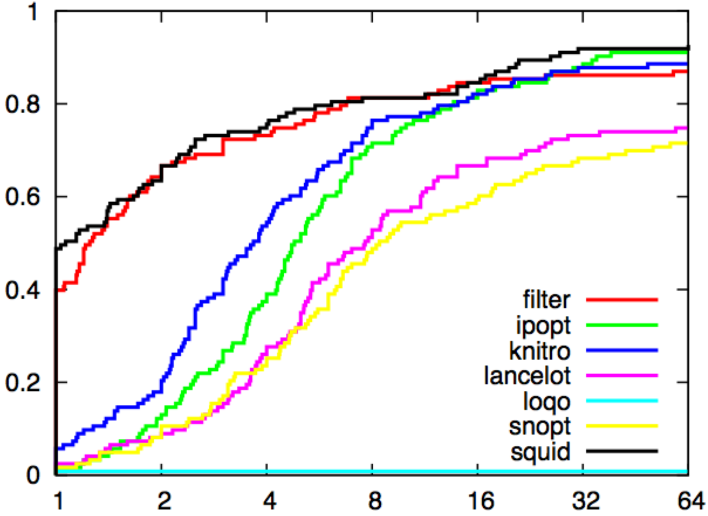
| Prob. | Filter | | SQuID | |
|---|---|---|---|---|
| | Iter. | Eval. | Iter. | Eval. |
| 1 | 16 | 16 | 16 | 18 |
| 2 | 12 | 12 | 16 | 55 |
| 3 | 10 | 10 | 37 | 41 |
| 4 | 11 | 11 | 21 | 28 |
| 5 | 26 | 26 | 21 | 78 |
| 6 | 27 | 27 | 33 | 121 |
| 7 | 30 | 30 | 17 | 32 |
| 8 | 28 | 28 | 47 | 59 |

# Numerical experiments: Feasible optimization problems

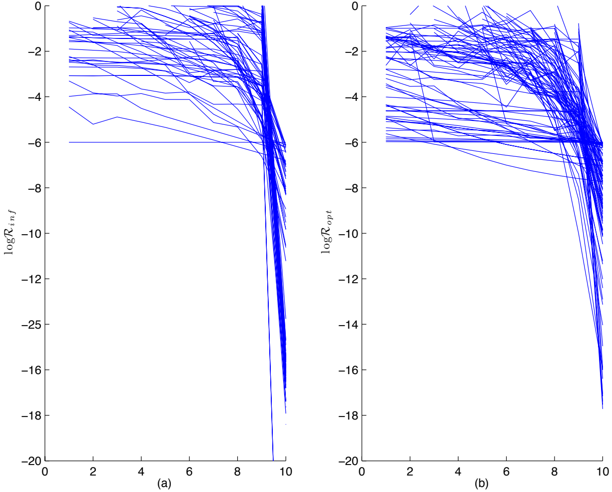



`FILTER`: 9 "failures" due to declaration of infeasibility
`Others`: $\leq 2$ "failures" due to declaration of infeasibility
`SQuID`: Lack of robustness due to lack of SOC and simplistic handling of nonconvexity

# Numerical experiments: Infeasible optimization problems



# Infeasible and feasible test problems



# Outline

# Large-scale optimization

A nonlinear optimizer would be saying to themselves:

- 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.


Inexactness also useful when:

- Starting from poor initial points
- Gradients and/or factorizations are inexact

# New notation (slightly different from before)

Write the optimization problem as

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

the feasibility problem as

$$\min_x \ v(x) \ \text{ s.t. } \ x \geq 0, \ \text{ where } v(x) := \|c(x)\|_1 \tag{FP}$$

and the penalty problem as

$$\min_x \ \phi(x, \mu) \ \text{ s.t. } \ x \geq 0, \ \text{ where } \phi(x, \mu) := \mu f(x) + v(x). \tag{PP}$$

The Fritz John function is now written as

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

# Optimality conditions

Optimality conditions for both (PP) and (FP) can be described with

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

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

- 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

Recall the penalty function

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

and define the corresponding model at $x_k$:

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

We iterate by (approximately) solving two subproblems:

- Feasibility subproblem:

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

- Optimality subproblem:

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

# Optimality conditions (for the subproblems)

We iterate by (approximately) solving two subproblems:

- Feasibility subproblem:

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

- Optimality subproblem:

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

Optimality conditions for both (QO1) and (QO2) can be described with

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

## 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 $\mu_k$, 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 $\mu$ 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!!

References:


- 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, D.P. Robinson, and A. Wächter, "An Inexact Sequential Quadratic Optimization Algorithm for Large-Scale Nonlinear Optimization," in preparation.