Infeasibility Detection in Nonlinear Optimization

Frank E. Curtis, Lehigh University

involving joint work with

James V. Burke, University of Washington
Hao Wang, Lehigh University

Twelfth Copper Mountain Conference on Iterative Methods

29 March 2012

Outline

Outline

Motivation

Active-set Method

Interior-Point Method

Summary

Constrained minimization

Is this how we should formulate nonlinear optimization (NLO) problems?

$$\min_{x} \; f(x) \quad \text{s.t.} \; \begin{cases} c^{\mathcal{E}}(x) = 0 \\ c^{\mathcal{I}}(x) \leq 0 \end{cases} \tag{OP}$$

Constraint violation minimization

What if the constraints are infeasible?

- modeling errors
- data inconsistency
- branch-and-bound for mixed-integer optimization

Then, we want to solve

$$\min_x \; v(x) := \left\| \begin{bmatrix} c^{\mathcal{E}}(x) \\ \max\{c^{\mathcal{I}}(x), 0\} \end{bmatrix} \right\|. \tag{FP}$$

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

Active-set and interior-point method that complete the convergence picture for NLO:

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

Outline

Sequential quadratic optimization (SQO)

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

$$\min_{d} \ f(x_k) + \nabla f(x_k)^T d + \tfrac{1}{2} d^T H(x_k, \lambda_k) d$$
$$\text{s.t.} \ \begin{cases} c^{\mathcal{E}}(x_k) + \nabla c^{\mathcal{E}}(x_k)^T d = 0 \\ c^{\mathcal{I}}(x_k) + \nabla c^{\mathcal{I}}(x_k)^T d \leq 0. \end{cases}$$

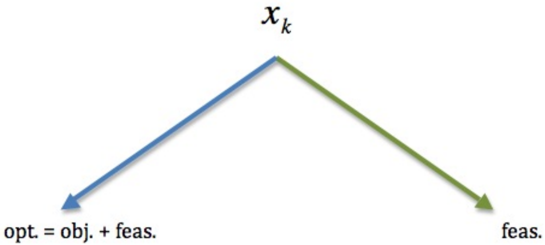
- May reduce to Newton's method once the active set is identified.
- However, a globalization mechanism is needed.
- Moreover, this subproblem may be infeasible!

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)

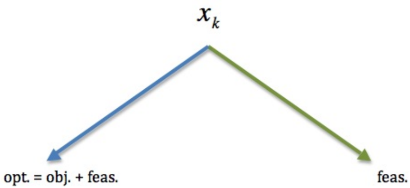
Issue faced by all NLO solvers

Move towards feasibility and/or objective decrease?

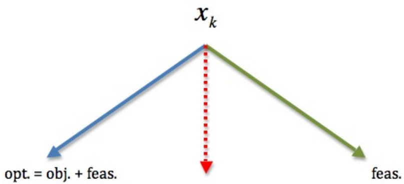


FILTER and steering strategies

- 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]."

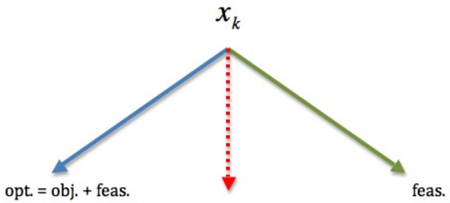


- Steering methods solve a sequence of constrained subproblems:



Our approach: Two-phase strategy

- Exploratory step to determine possible progress toward feasibility.
- Formulation of optimality step exploits information obtained by exploratory step.



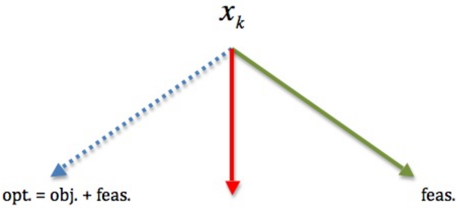
- Objective function never ignored (unlike `FILTER`).
- At most two subproblems solved per iteration (unlike steering).
- Reduces to SQO for optimization problem in feasible cases.
- Reduces to (perturbed) SQO for feasibility problem in infeasible cases.

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 $d_k = w_k\overline{d}_k + (1 - w_k)\widehat{d}_k$ to obtain proportional feasibility improvement.



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

$$\phi(x; \rho) := \rho f(x) + v(x).$$

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 $(\overline{\lambda}^{\mathcal{E}}_{k+1}, \overline{\lambda}^{\mathcal{I}}_{k+1})$:

$$\min_{d,r,s,t} \; e^T(r+s) + e^T t + \tfrac{1}{2} d^T H(x_k, 0, \overline{\lambda}_k) d$$
$$\text{s.t.} \begin{cases} c^{\mathcal{E}}(x_k) + \nabla c^{\mathcal{E}}(x_k)^T d = r - s \\ c^{\mathcal{I}}(x_k) + \nabla c^{\mathcal{I}}(x_k)^T d \leq t \\ (r, s, t) \geq 0. \end{cases} \tag{QO1}$$

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

$$l_k(d) := \|c^{\mathcal{E}}(x_k) + \nabla c^{\mathcal{E}}(x_k)^T d\|_1 + \|\max\{c^{\mathcal{I}}(x_k) + \nabla c^{\mathcal{I}}(x_k)^T d, 0\}\|_1.$$

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:

$$\begin{aligned} c^{\mathcal{E}_k}(x_k) + \nabla c^{\mathcal{E}_k}(x_k)^T \overline{d}_k &= 0 \\ c^{\mathcal{I}_k}(x_k) + \nabla c^{\mathcal{I}_k}(x_k)^T \overline{d}_k &\leq 0. \end{aligned}$$

- Solve for $(d_k, r_k^{\mathcal{E}_k^c}, s_k^{\mathcal{E}_k^c}, t_k^{\mathcal{I}_k^c})$ and $(\widehat{\lambda}_{k+1}^{\mathcal{E}}, \widehat{\lambda}_{k+1}^{\mathcal{I}})$:

$$\min_{d, r^{\mathcal{E}_k^c}, s^{\mathcal{E}_k^c}, t^{\mathcal{I}_k^c}} \rho_k \nabla f(x_k)^T d + e^T (r^{\mathcal{E}_k^c} + s^{\mathcal{E}_k^c}) + e^T t^{\mathcal{I}_k^c} + \tfrac{1}{2} d^T H(x_k, \rho_k, \widehat{\lambda}_k) d$$

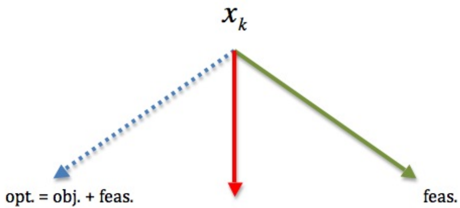
$$\text{s.t.} \begin{cases} c^{\mathcal{E}_k^c}(x_k) + \nabla c^{\mathcal{E}_k^c}(x_k)^T d = 0 \\ c^{\mathcal{E}_k^c}(x_k) + \nabla c^{\mathcal{E}_k^c}(x_k)^T d = r^{\mathcal{E}_k^c} - s^{\mathcal{E}_k^c} \\ c^{\mathcal{I}_k}(x_k) + \nabla c^{\mathcal{I}_k}(x_k)^T d \leq 0 \\ c^{\mathcal{I}_k^c}(x_k)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

Search direction

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

- Find the smallest w_k such that, for $\beta \in (0,1)$, d_k satisfies

$$v(x_k) - l_k(d_k) \geq \beta(v(x_k) - l_k(\overline{d}_k)).$$



ρ update

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

$$\phi(x;\rho) := \rho f(x) + v(x).$$

- Set ρ_{k+1} so that
$$\rho_{k+1} \leq \frac{1}{\|\widehat{\lambda}_{k+1}\|_\infty}.$$
- Set ρ_{k+1} so that d_k yields
$$\phi(x_k;\rho_{k+1}) - \rho_{k+1}\nabla f(x_k)^T d - l_k(d_k) \geq \epsilon(v(x_k) - l_k(d_k)).$$

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

Ensuring fast local convergence

(1) (OP) feasible: (QO2) reduces to standard SQO subproblem.
(2) (OP) infeasible: Rapidly reduce ρ so that (QO2) reduces to (QO1).

Feasible case

(1) (OP) feasible: (QO2) reduces to standard SQO subproblem.

$$\min_{d, r^{\mathcal{E}_k^c}, s^{\mathcal{E}_k^c}, t^{\mathcal{I}_k^c}} \rho_k \nabla f(x_k)^T d + e^T (r^{\mathcal{E}_k^c} + s^{\mathcal{E}_k^c}) + e^T t^{\mathcal{I}_k^c} + \frac{1}{2} d^T H(x_k, \rho_k, \widehat{\lambda}_k) d$$

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

Infeasible case

(2) (OP) infeasible: Rapidly reduce ρ so that (QO2) reduces to (QO1).

$$\min_{d,r,s,t} \; e^T(r+s) + e^T t + \frac{1}{2} d^T H(x_k, 0, \overline{\lambda}_k) d$$

$$\text{s.t.} \begin{cases} c^{\mathcal{E}}(x_k) + \nabla c^{\mathcal{E}}(x_k)^T d = r - s \\ c^{\mathcal{I}}(x_k) + \nabla c^{\mathcal{I}}(x_k)^T d \leq t \\ \qquad\qquad (r, s, t) \geq 0. \end{cases} \tag{QO1}$$

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

$$\rho_k \leq \text{KKT}_{inf}(x_k, \overline{\lambda}_{k+1})^2$$

$$\|\widehat{\lambda}_k - \overline{\lambda}_k\| \leq \text{KKT}_{inf}(x_k, \overline{\lambda}_{k+1})^2.$$

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 ρ_k and $\widehat{\lambda}_k$, 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 ρ_k, if necessary (for global convergence).
(8) Perform line search to obtain decrease in merit function.

Global convergence: Assumptions

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

$$\mu_{\min}\|d\|^2 \leq d^T H(x_k, 0, \overline{\lambda}_k) d \leq \mu_{\max}\|d\|^2$$
$$\mu_{\min}\|d\|^2 \leq d^T H(x_k, \rho_k, \widehat{\lambda}_k) d \leq \mu_{\max}\|d\|^2.$$

(2) Continuity and boundedness: f, $c^{\mathcal{E}}$, $c^{\mathcal{I}}$ and their first-order derivatives are bounded and Lipschitz continuous in a convex set containing $\{x_k\}$.

Global convergence

Theorem

All limit points of $\{x_k\}$ are either feasible or infeasible stationary.

Theorem

If $\rho_k \geq \rho_*$ for some constant $\rho_* > 0$ for all k, then every limit point $\{(x_*, \rho_*, \lambda_*)\}$ of $\{(x_k, \rho_{k+1}, \lambda_{k+1})\}$ with $v(x_*) = 0$ is a KKT point for (OP).

Theorem

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

Local convergence: Assumptions

(1) f, $c^{\mathcal{E}}$ and $c^{\mathcal{I}}$ 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 $(x_*, \overline{\lambda}_*)$ is a KKT point for (FP), then

  (a) $\nabla c^{\mathcal{Z}_* \cup \mathcal{A}_*}(x_*)^T$ has full row rank.

  (b) $-e < \overline{\lambda}_*^{\mathcal{Z}_*} < e$ and $0 < \overline{\lambda}_*^{\mathcal{A}_*} < e$.

  (c) $d^T H(x_*, 0, \overline{\lambda}_*) d > 0$ for all $d \neq 0$ such that $\nabla c^{\mathcal{Z}_* \cup \mathcal{A}_*}(x_*)^T d = 0$.

(3) If $(x_*, \rho_*, \widehat{\lambda}_*)$ is a KKT point for (OP), then (2) holds, $\rho_k \to \rho_* > 0$, and

  (a) $\widehat{\lambda}_*^{\mathcal{A}_*} + c^{\mathcal{A}_*}(x_*) > 0$.

  (b) $d^T H(x_*, \rho_*, \widehat{\lambda}_*) d > 0$ for all $d \neq 0$ such that $\nabla c^{\mathcal{E}_* \cup \mathcal{A}_*}(x_*)^T d = 0$.

Local convergence

Theorem

If $v(x_*) > 0$, and $(x_k, \overline{\lambda}_k)$ and $(x_k, \widehat{\lambda}_k)$ are each sufficiently close to $(x_*, \overline{\lambda}_*)$, then

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

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

Theorem

If $\|(x_k, \overline{\lambda}_k) - (x_*, \overline{\lambda}_*)\|$ and $\|(x_k, \widehat{\lambda}_k) - (x_*, \widehat{\lambda}_*)\|$ each sufficiently small, then

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

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

Feasible and infeasible test problems

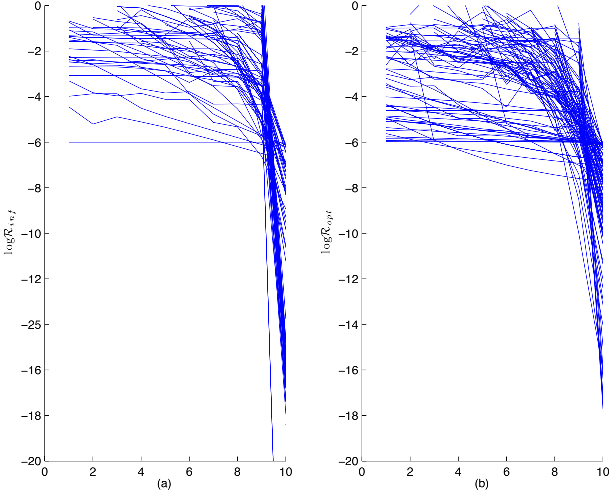
Table: Performance statistics of SQuID on feasible problems

Problem type	Succeed	Fail	Infeasible	Total
Feasible	110 (90.16%)	11 (9.02%)	1 (0.82%)	122

Table: Performance statistics of SQuID on infeasible problems

Problem type	Succeed	Fail	Feasible	Total
Infeasible	111 (90.24%)	12 (9.76%)	0 (0.0%)	123

Feasible and infeasible test problems



Outline

Penalty and interior-point methods

Constrained subproblems in penalty methods can be expensive:

$$\min_{x,r,s,t} \rho f(x) + e^T r + e^T s + e^T t$$
$$\text{s.t.} \begin{cases} c^{\mathcal{E}}(x) = & r - s \\ c^{\mathcal{I}}(x) \leq & t \\ (r,s,t) \geq & 0 \end{cases} \tag{PP}$$

Interior-point methods are more efficient for large-scale problems:

$$\min_{x,u} f(x) - \mu \sum \ln u^i$$
$$\text{s.t.} \begin{cases} c^{\mathcal{E}}(x) = & 0 \\ c^{\mathcal{I}}(x) = & -u \\ u \geq & 0 \end{cases} \tag{IP}$$

Penalty-interior-point method

Applying a penalty-interior-point reformulation to (OP):

$$\begin{aligned} \min_{x,r,s,t,u}\ & \rho f(x) - \mu \left(\sum (\ln r^i + \ln s^i) + \sum (\ln t^i + \ln u^i) \right) + e^T r + e^T s + e^T t \\ \text{s.t.}\ & \begin{cases} c^{\mathcal{E}}(x) = & r - s \\ c^{\mathcal{I}}(x) = & t - u \end{cases} \end{aligned} \tag{PIP}$$

The optimization problem (OP) and feasibility problem (FP) can be solved via (PIP):

- $\mu \to 0$ and $\rho \to \bar{\rho} > 0$ to solve (OP).
- $\mu \to 0$ and $\rho \to 0$ to solve (FP).

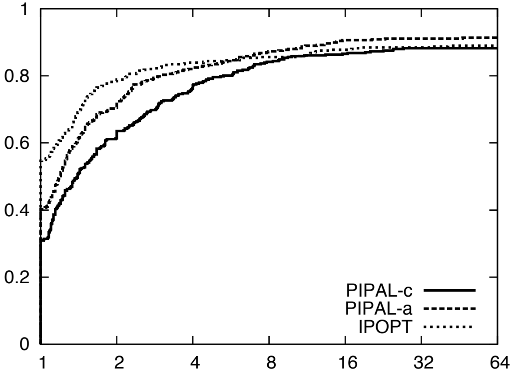
Literature

Previous work with similar motivations:

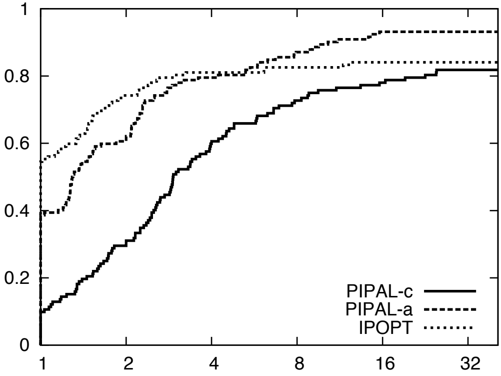
- Jittorntrum and Osborne (1980)
- Polyak (1982, 1992, 2008)
- Breitfeld and Shanno (1994, 1996)
- Goldfarb, Polyak, Scheinberg, and Yuzefovich (1999)
- Gould, Orban, and Toint (2003)
- Chen and Goldfarb (2006, 2006)
- Benson, Sen, and Shanno (2008)

Parameter updates are essential to have a practical algorithm.

Numerical results: Feasible problems (sample size = 417)

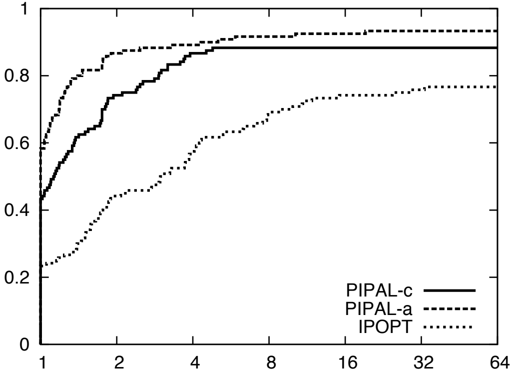


Numerical results: Feasible problems w/ ρ decrease (sample size $= 132$)



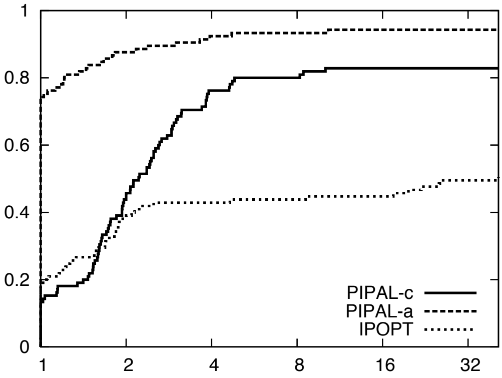
Numerical results: Degenerate problems (sample size = 120)

Added constraints: $c^i(x)^2 \leq 0$



Numerical results: Infeasible problems (sample size = 105)

Added constraints: $c^i(x)^2 \leq -1$



Outline

Summary

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

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

- Referred to a penalty-interior-point method with similar motivations.
- Numerical results for both algorithms are encouraging.

Thanks!!

References:

- J.V. Burke, F.E. Curtis, and H. Wang, "A Sequential Quadratic Optimization Algorithm with Rapid Infeasibility Detection," in preparation.
- F.E. Curtis, "A Penalty-Interior-Point Algorithm for Nonlinear Constrained Optimization," to appear in Mathematical Programming Computation.