Infeasibility Detection in Nonlinear Programming

Frank E. Curtis
New York University

joint work with
Richard H. Byrd and Jorge Nocedal

Informs Conference on Optimization, March 2008

Outline

Infeasible Nonlinear Programming

We consider the optimization problems

$$(OPT) \triangleq \left\{ \begin{array}{l} \min \ f(x) \\ \text{ s.t. } \ c(x) \geq 0 \end{array} \right\} \quad \text{and} \quad (FEAS) \triangleq \left\{ \min \sum_{i=1}^{t} \max\{-c^i(x), 0\} \right\}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $c : \mathbb{R}^n \to \mathbb{R}^t$ are smooth functions

- We want to solve (OPT) when a *feasible* point exists (i.e., $\exists x \in \mathbb{R}^n$ s.t. $c(x) \geq 0$)
- Otherwise, the algorithm should solve $(FEAS)$ when (OPT) is *infeasible*
- Many optimization methods focus on the efficient solution of (OPT), often with guarantees toward solutions of $(FEAS)$ if the problem is infeasible
- ... however, this latter feature is often treated as an afterthought and the rate at which the method converges can be exceedingly slow

Focus on active set methods

- Interior-point methods are known to behave poorly on infeasible problems:

$$\left\{ \begin{array}{l} \min\ f(x) - \mu \sum_{i=1}^{t} \ln s^i \\ \text{s.t. } c(x) - s = 0,\ \ s > 0 \end{array} \right\} \quad \Leftarrow \quad \text{true interior is empty}$$

- Active-set methods present another option:
  Running `SNOPT` and `KNITRO` on NEOS:

Problem	`SNOPT`	`KNITRO`
`optprloc1`	11 itrs	10 itrs
`optprloc2`	14 itrs	44 itrs
`optprloc3`	30 itrs	29 itrs
`c-reload-14c`	37 itrs	1000+ itrs
`batch`	1000+ itrs	37 itrs

A single algorithm for an entire problem family

Our goal is to design a *single* optimization algorithm designed for the fast solution of (OPT), or the fast solution of $(FEAS)$ when (OPT) is infeasible, that does not *switch* between two separate techniques (e.g., no feasibility restoration as in Fletcher and Leyffer, 1997)

$$(OPT) \triangleq \left\{ \begin{array}{l} \min \ f(x) \\ \text{ s.t. } \ c(x) \geq 0 \end{array} \right\} \quad \text{and} \quad (FEAS) \triangleq \left\{ \begin{array}{c} \min \ e^T r \\ \text{s.t. } \ c(x) + r \geq 0 \\ r \geq 0 \end{array} \right\}$$

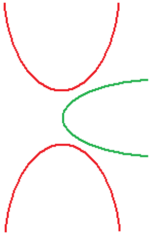
We combine (OPT) and $(FEAS)$ to define

$$(P) \triangleq \left\{ \begin{array}{c} \min \ \frac{1}{\pi} f(x) + e^T r \\ \text{s.t. } \ c(x) + r \geq 0 \\ r \geq 0 \end{array} \right\}$$

where $\pi > 0$ is a penalty parameter to be updated dynamically

An ideal run of KNITRO

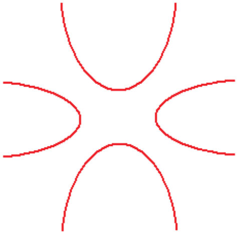
$$\min x_1$$
$$\text{s.t.} \quad -x_1^2 + x_2 - 1 \geq 0$$
$$-x_1^2 - x_2 - 1 \geq 0$$
$$x_1 - x_2^2 \geq 0$$
$$-x_1 + x_2^2 \geq 0$$



Iter	Objective	Feas err	Opt err	||Step||	pi
13	1.061997e-03	1.034e+00	1.000e+00	6.192e-02	1.000e+02
14	-6.689357e-05	1.000e+00	9.097e-01	3.379e-02	1.000e+02
15	-4.474151e-09	1.000e+00	9.999e-01	9.460e-05	1.000e+02
16	-2.001803e-17	1.000e+00	1.000e+00	6.327e-09	1.000e+02

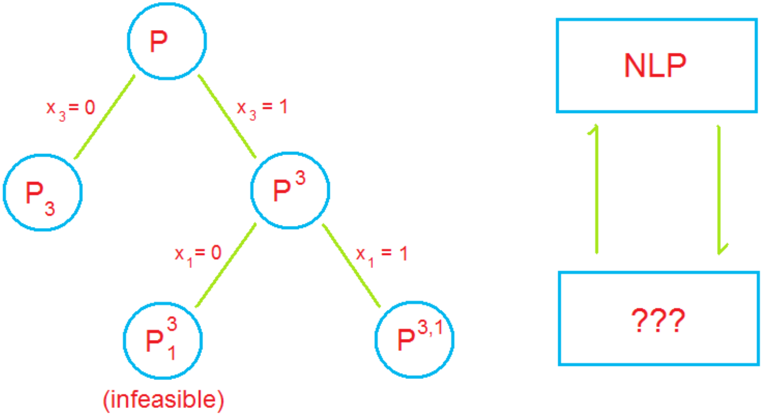
A *less than* ideal run of KNITRO

$$\begin{aligned} &\min\ x_1 + x_2 \\ &\text{s.t.}\ -x_1^2 + x_2 - 1 \geq 0 \\ &\quad -x_1^2 - x_2 - 1 \geq 0 \\ &\quad x_1 - x_2^2 - 1 \geq 0 \\ &\quad -x_1 - x_2^2 - 1 \geq 0 \end{aligned}$$

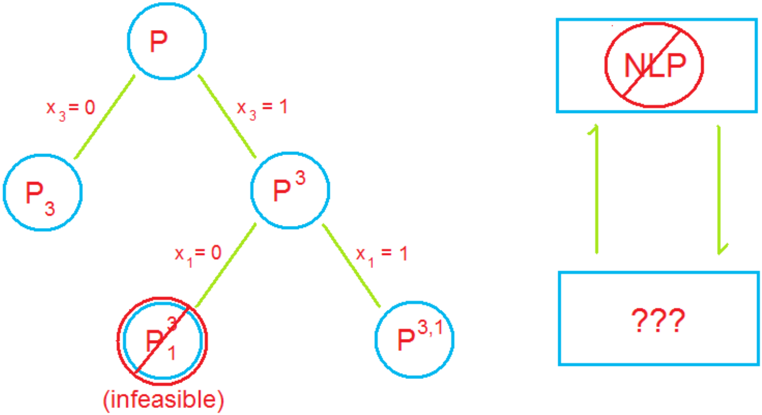


Iter	Objective	Feas err	Opt err	||Step||	pi
13	-5.000000e-07	1.000e+00	1.000e+00	0.000e+00	1.000e+06
14	-5.000000e-08	1.000e+00	1.000e+00	3.182e-07	1.000e+07
15	-5.000000e-08	1.000e+00	1.000e+00	0.000e+00	1.000e+07
16	-5.000000e-09	1.000e+00	1.000e+00	3.182e-08	1.000e+08

Effects compounded in MINLP methods



Effects compounded in MINLP methods



Summary

- There is a need for algorithms that converge quickly, regardless of whether the problem is feasible or infeasible
- Interior-point methods are known to perform poorly in infeasible cases, but active set methods seem promising
- Room for improvement in active set methods, too
- Feasibility restoration techniques are an option, but we prefer a smooth transition between solving (OPT) and solving $(FEAS)$
- When π remains finite, convergence can be fast since, after a point, we are solving a single problem
- However, we need to analyze the $\pi \to \infty$ case as well...

Our method for step computation and acceptance

We generate a step via the quadratic subproblem

$$(Q) \triangleq \begin{array}{l} \min\ q_k(d;\pi) \triangleq \frac{1}{\pi}\nabla f_k^T d + \frac{1}{2} d^T W_k d + e^T s \\ \text{s.t. } c_k + \nabla c_k^T d + s \geq 0, \quad s \geq 0 \end{array}$$

where W_k is an approximation for the Hessian of the Lagrangian of (P), and we measure progress with the exact penalty function

$$\phi(x;\pi) \triangleq \frac{1}{\pi} f(x) + \sum_{i=1}^{t} \max\{-c^i(x), 0\}$$

We see later on that this SQP approach has the benefit that it can identify the correct *active set* near a solution point for π sufficiently large

A Penalty-SQP algorithm

Step 0. Initialize x_0 and set $\eta \in (0, 1)$, $\tau \in (0, 1)$ and $k \leftarrow 0$

Step 1. If x_k solves (OPT) or $(FEAS)$, then stop

Step 2. Compute a value for the penalty parameter, call it π_k

Step 3. Compute d_k by solving (Q) with $\pi \leftarrow \pi_k$

Step 4. Let α_k be the first member of the sequence $\{1, \tau, \tau^2, \ldots\}$ s.t.

$$\phi(x_k; \pi_k) - \phi(x_k + \alpha_k d_k; \pi_k) \geq \eta\alpha_k[q_k(0; \pi_k) - q_k(d_k; \pi_k)]$$

Step 5. Update $x_{k+1} \leftarrow x_k + \alpha_k d_k$, go to Step 1

A Penalty-SQP algorithm

Step 0. Initialize x_0 and set $\eta \in (0,1)$, $\tau \in (0,1)$ and $k \leftarrow 0$

Step 1. If x_k solves (OPT) or $(FEAS)$, then stop

Step 2. Compute a value for the penalty parameter, call it π_k

Step 3. Compute d_k by solving (Q) with $\pi \leftarrow \pi_k$

Step 4. Let α_k be the first member of the sequence $\{1, \tau, \tau^2, \ldots\}$ s.t.

$$\phi(x_k; \pi_k) - \phi(x_k + \alpha_k d_k; \pi_k) \geq \eta \alpha_k [q_k(0; \pi_k) - q_k(d_k; \pi_k)]$$

Step 5. Update $x_{k+1} \leftarrow x_k + \alpha_k d_k$, go to Step 1

Strategy for fast convergence

Hitting a moving target:

$$x_k \longrightarrow x_\pi \longrightarrow \hat{x}$$

where

$$\begin{aligned} x_k &\triangleq \ k\text{th iterate of the algorithm} \\ x_\pi &\triangleq \ \text{solution of penalty problem } (P) \\ \hat{x} &\triangleq \ \text{infeasible stationary point of } (OPT)\text{, solution of } (FEAS) \end{aligned}$$

We aim to show, for some $C, C' > 0$,

$$\begin{aligned} \|x_{k+1} - \hat{x}\| &\leq \ \|x_{k+1} - x_\pi\| + \|x_\pi - \hat{x}\| \\ &\leq \ C\|x_k - x_\pi\|^2 + O(1/\pi) \\ &\leq \ C'\|x_k - \hat{x}\|^2 + O(1/\pi), \end{aligned}$$

so convergence is quadratic if $(1/\pi) \propto \|x_k - \hat{x}\|^2$

Optimality conditions for problem (P)

First-order optimality conditions for

$$(P) \triangleq \left\{ \min \ \tfrac{1}{\pi} f(x) + e^T r, \quad \text{s.t. } c(x) + r \geq 0, \ r \geq 0 \right\} :$$

$$\left\{ \begin{array}{rll} \frac{1}{\pi}\nabla f(x) - \sum_{i\in\mathcal{I}} \lambda^i \nabla c^i(x) = & 0 & \\ 1 - \lambda^i - \sigma^i = & 0, & i \in \mathcal{I} \\ \lambda^i(c^i(x) + r^i) = & 0, & i \in \mathcal{I} \\ \sigma^i r^i = & 0, & i \in \mathcal{I} \\ c^i(x) + r^i \geq & 0, & i \in \mathcal{I} \\ r, \lambda, \sigma \geq & 0 & \end{array} \right\}$$

At an infeasible stationary point $\hat{x}$ we define

$$\hat{\mathcal{A}} = \{i : c^i(\hat{x}) = 0\}, \quad \hat{\mathcal{V}} = \{i : c^i(\hat{x}) < 0\}, \quad \hat{\mathcal{S}} = \{i : c^i(\hat{x}) > 0\}$$

as the sets of *active*, *violated*, and *strictly satisfied* constraints

Assumptions

The point $(\hat{x}, \hat{r}, \hat{\lambda}, \hat{\sigma})$ is a first-order optimal solution of (P) at which the following conditions hold:

- (Regularity) $\nabla c(\hat{x})^T$ has full row rank;
- (Strict Complementarity) $\hat{\lambda}^i > 0$ for all $i \in \hat{\mathcal{A}}$;
- (Second Order Sufficiency) The Hessian of the Lagrangian for problem (P) with $\pi = \infty$, denoted by $\hat{W}$, satisfies $d^T \hat{W} d > 0$ for all $d \neq 0$ such that $\nabla c(\hat{x})^T d = 0$

The optimality conditions now reduce to: (define $\rho = 1/\pi$)

$$F(x, \lambda_{\hat{\mathcal{A}}}, \rho) = \begin{bmatrix} \rho \nabla f(x) - \sum_{i \in \hat{\mathcal{A}}} \lambda^i \nabla c^i(x) - \sum_{i \in \hat{\mathcal{V}}} \nabla c^i(x) \\ c_{\hat{\mathcal{A}}}(x) \end{bmatrix} = 0$$

$$\lambda_{\hat{\mathcal{A}}} \in (0, 1)$$

(all other values can be determined uniquely)

Lemma 1: $x_\pi \longrightarrow \hat{x}$

For all π sufficiently large the penalty problem (P) has a solution x_π with the same sets of active, violated, and strictly satisfied constraints as $\hat{x}$. Moreover,

$$\|x_\pi - \hat{x}\| = O(1/\pi)$$

Proof.

We have $F(\hat{x}, \hat{\lambda}_{\hat{\mathcal{A}}}, 0) = 0$. Differentiating F yields:

$$\frac{\partial F(x, \lambda_{\hat{\mathcal{A}}}, \rho)}{\partial(x, \lambda_{\hat{\mathcal{A}}})} = \begin{bmatrix} W(x, \lambda_{\hat{\mathcal{A}}}, \rho) & -\nabla c_{\hat{\mathcal{A}}}(x) \\ \nabla c_{\hat{\mathcal{A}}}(x)^T & 0 \end{bmatrix},$$

which is nonsingular under our assumptions. The implicit function theorem then implies that there is an open neighborhood $\mathcal{N} \in \mathbb{R}$ containing $\rho = 0$ such that

$$F(x(\rho), \lambda_{\hat{\mathcal{A}}}(\rho), \rho) = 0 \;\text{ for all }\; \rho \in \mathcal{N}.$$

Then, since $\hat{\lambda}_{\hat{\mathcal{A}}} \in (0, 1)$, $(x(\rho), \lambda_{\hat{\mathcal{A}}}(\rho), \rho)$ satisfies the first-order optimality conditions for ρ sufficiently small (π large) □

Lemma 1: $x_\pi \to \hat{x}$

For all π sufficiently large the penalty problem (P) has a solution x_π with the same sets of active, violated, and strictly satisfied constraints as $\hat{x}$. Moreover,

$$\|x_\pi - \hat{x}\| = O(1/\pi)$$

Example: (recall $\rho = 1/\pi$)

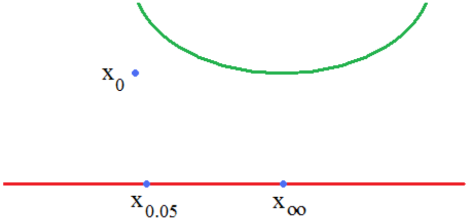
$$\begin{aligned}
\min\ & \rho\left((x_1+1)^2 + (x_2-1)^2\right) + r_1 + r_2 \\
\text{s.t.}\ & -x_1^2 + x_2 - 1 + r_1 \geq 0 \\
& -100x_2 + r_2 \geq 0 \\
& (r_1, r_2) \geq 0
\end{aligned}$$

Lemma 1: $x_\pi \rightarrow \hat{x}$

For all π sufficiently large the penalty problem (P) has a solution x_π with the same sets of active, violated, and strictly satisfied constraints as $\hat{x}$. Moreover,

$$\|x_\pi - \hat{x}\| = O(1/\pi)$$

Example:

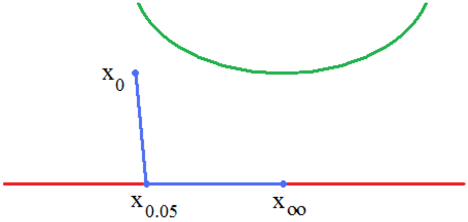


Lemma 1: $x_\pi \longrightarrow \hat{x}$

For all π sufficiently large the penalty problem (P) has a solution x_π with the same sets of active, violated, and strictly satisfied constraints as $\hat{x}$. Moreover,

$$\|x_\pi - \hat{x}\| = O(1/\pi)$$

Example:



Lemma 2: $x_k \longrightarrow x_\pi \longrightarrow \hat{x}$

For π sufficiently large and for x_k sufficiently close to x_π, the solution of the SQP subproblem identifies the same sets of active, violated, and strictly satisfied constraints as x_π (and $\hat{x}$). Then, standard Newton analysis for equality constrained optimization yields for some $C > 0$:

$$\|x_{k+1} - x_\pi\| \leq C\|x_k - x_\pi\|^2$$

Proof.

Similar to before, at $(x, \lambda_{\hat{\mathcal{A}}}, \rho) = (\hat{x}, \hat{\lambda}_{\hat{\mathcal{A}}}, 0)$ the SQP step is the solution $(d, \delta_{\hat{\mathcal{A}}}) = (0, \hat{\lambda}_{\hat{\mathcal{A}}})$ to:

$$\begin{bmatrix} W(x, \lambda_{\hat{\mathcal{A}}}, \rho) & -\nabla c_{\hat{\mathcal{A}}}(x) \\ \nabla c_{\hat{\mathcal{A}}}^T(x) & 0 \end{bmatrix} \begin{bmatrix} d \\ \delta_{\hat{\mathcal{A}}} \end{bmatrix} = - \begin{bmatrix} \rho \nabla f(x) - \sum_{i \in \hat{\mathcal{V}}} \nabla c^i(x) \\ c_{\hat{\mathcal{A}}}(x) \end{bmatrix}$$

This matrix is nonsingular and the solution varies continuously with $(x, \lambda_{\hat{\mathcal{A}}}, \rho)$ near $(\hat{x}, \hat{\lambda}_{\hat{\mathcal{A}}}, 0)$, so since $\hat{\lambda}^i \in (0, 1)$ for $i \in \hat{\mathcal{A}}$ the solution of the SQP subproblem can be obtained via this linear system (setting $\delta_{\hat{\mathcal{V}}} = 1$ and $\delta_{\hat{\mathcal{S}}} = 0$) for $(x, \lambda_{\hat{\mathcal{A}}})$ near $(\hat{x}, \hat{\lambda}_{\hat{\mathcal{A}}})$ and ρ small (π large) □

Main result

Thus, we find:

$$
\begin{aligned}
\|x_{k+1} - \hat{x}\| &\leq \|x_{k+1} - x_\pi\| + \|x_\pi - \hat{x}\| \text{ (triangle inequality)} \\
&\leq C\|x_k - x_\pi\|^2 + O(1/\pi) \text{ (Lemmas 1 and 2)} \\
&\vdots \\
&\leq C'\|x_k - \hat{x}\|^2 + O(1/\pi),
\end{aligned}
$$

so convergence is quadratic if $(1/\pi) \propto \|x_k - \hat{x}\|^2$; e.g., $1/\pi$ proportional to the squared optimality error of the problem $(FEAS)$

Summary

- We have discussed methods for the fast solution of infeasible optimization problems
- We have analyzed a penalty-SQP approach that transitions smoothly between solving an optimization problem and its feasibility problem counterpart
- We have shown that the approach can converge quadratically if the penalty parameter is handled correctly

Future work

- How can we construct a practical method for updating π that satisfies our condition? e.g., consider the auxiliary problem

$$\begin{aligned} &\min \sum s^i \\ &\text{s.t. } c_k + \nabla c_k^T d + s \geq 0, \quad s \geq 0 \end{aligned}$$

  and set π_k so that the reduction in linearized feasibility of the SQP problem is proportional to that achieved by the solution of this problem – can this do the trick?
- Can we relax our assumptions? For example, for many infeasible problems, the Hessian of the Lagrangian is not positive definite at $\hat{x}$