An Active-Set Algorithm for Large-Scale Nonlinear Optimization

Richard A. Waltz, Northwestern University

rwaltz@ece.northwestern.edu

joint with: Richard Byrd, Nick Gould and Jorge Nocedal

EUCCO

March 29-31, 2004

Motivation

Why develop a new active-set code?

- Demand for solving larger NLPs
- Active-set methods have nice properties
  - warm starts
  - good active-set and sensitivity info
  - more stable
- SQP ineffective when large reduced space

GOAL: Develop large-scale active-set method

Weaknesses of SQP

- Costly when many degrees of freedom
  - Form and factorize a (dense) reduced Hessian
- Infeasible QP subproblems
- Tradeoff: Exact 2nd derivatives vs. indefinite QPs
  - TR: Exact Hessian but indefinite QPs (FilterSQP)
  - LS: BFGS Hessian but convex QPs (SNOPT)
- Can you get a cheap Cauchy point for SQP???

SEQP Methods

Investigating class of SEQP methods:

- Estimate active-set at each iteration
- Only solve equality constrained QPs

SEQP Methods

Investigating class of SEQP methods:

- Estimate active-set at each iteration
- Only solve equality constrained QPs
- Can use iterative (projected CG) approach

SEQP Methods

Investigating class of SEQP methods:

- Estimate active-set at each iteration
- Only solve equality constrained QPs
- Can use iterative (projected CG) approach
- Can solve EQP subproblems inexactly

SEQP Methods

Investigating class of SEQP methods:

- Estimate active-set at each iteration
- Only solve equality constrained QPs
- Can use iterative (projected CG) approach
- Can solve EQP subproblems inexactly
- No need to factorize Hessian/reduced Hessian

SEQP Methods

Investigating class of SEQP methods:

- Estimate active-set at each iteration
- Only solve equality constrained QPs
- Can use iterative (projected CG) approach
- Can solve EQP subproblems inexactly
- No need to factorize Hessian/reduced Hessian
- Can work with exact Hessians and handle nonconvexities

SEQP Methods

Investigating class of SEQP methods:

- Estimate active-set at each iteration
- Only solve equality constrained QPs
- Can use iterative (projected CG) approach
- Can solve EQP subproblems inexactly
- No need to factorize Hessian/reduced Hessian
- Can work with exact Hessians and handle nonconvexities

BIG QUESTION:

How to estimate the active-set???

Estimating the Active-Set

1) Gradient Projection

- Estimate active-set by performing a line search along a piece-wise linear gradient projection path

Estimating the Active-Set

1) Gradient Projection

- Estimate active-set by performing a line search along a piece-wise linear gradient projection path
- Very effective for bound-constrained problems
  - LANCELOT (Conn, Gould, Toint)
  - L-BFGS-B (Zhu, Byrd, Nocedal)
  - TRON (Lin, Moré)

Estimating the Active-Set

1) Gradient Projection

- Estimate active-set by performing a line search along a piece-wise linear gradient projection path
- Very effective for bound-constrained problems
  - LANCELOT (Conn, Gould, Toint)
  - L-BFGS-B (Zhu, Byrd, Nocedal)
  - TRON (Lin, Moré)
- Projection too expensive for more general geometries

Estimating the Active-Set

1) Gradient Projection

- Estimate active-set by performing a line search along a piece-wise linear gradient projection path
- Very effective for bound-constrained problems
  - LANCELOT (Conn, Gould, Toint)
  - L-BFGS-B (Zhu, Byrd, Nocedal)
  - TRON (Lin, Moré)
- Projection too expensive for more general geometries

2) Solve LP to estimate active-set (SLQP)

Overview of SLQP

1. Solve LP to estimate active-set
2. Step given by LP solution determines Cauchy point
3. Solve equality constrained QP (use CG)
4. Trial step is a convex combination of LP and EQP steps (achieves at least as much decrease as Cauchy step).

- Fletcher, Sainz de la Maza (1989)
- Chin, Fletcher (1999,2003)
- Byrd, Gould, Nocedal, W. [SLIQUE] (2004)

Overview of SLIQUE

Nonlinear Problem:

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & h_i(x) = 0, \quad i \in \mathcal{E} \\ & g_i(x) \geq 0, \quad i \in \mathcal{I} \\ & x \in I\!\!R^n \end{aligned}$$

Assume all functions twice-continuously differentiable.

LP Subproblem (SLIQUE)

$$\min_{d} \quad \nabla f(x)^T d$$

$$\text{s.t.} \quad h_i(x) + \nabla h_i(x)^T d = 0, \quad i \in \mathcal{E}$$

$$g_i(x) + \nabla g_i(x)^T d \geq 0, \quad i \in \mathcal{I}$$

LP Subproblem (SLIQUE)

$$\min_{d} \quad \nabla f(x)^T d$$

$$\text{s.t.} \quad h_i(x) + \nabla h_i(x)^T d = 0, \quad i \in \mathcal{E}$$

$$g_i(x) + \nabla g_i(x)^T d \geq 0, \quad i \in \mathcal{I}$$

$$\|d\|_\infty \leq \Delta^{\mathsf{LP}}$$

ℓ_1 LP Formulation (SLIQUE)

$$\begin{aligned} \ell(d) &= \nabla f(x)^T d + \nu \sum_{i \in \mathcal{E}} |h_i(x) + \nabla h_i(x)^T d| \\ &\quad + \nu \sum_{i \in \mathcal{I}} \max(0, -g_i(x) - \nabla g_i(x)^T d) \end{aligned}$$

ℓ_1 LP Formulation (SLIQUE)

$$\ell(d) = \nabla f(x)^T d + \nu \sum_{i \in \mathcal{E}} |h_i(x) + \nabla h_i(x)^T d|$$
$$+\nu \sum_{i \in \mathcal{I}} \text{max}(0, -g_i(x) - \nabla g_i(x)^T d)$$

$$\min_d \quad \ell(d)$$
$$\text{s.t.} \quad \|d\|_\infty \leq \Delta^{\text{LP}}$$

ℓ_1 LP Formulation (SLIQUE)

$$\ell(d) = \nabla f(x)^T d + \nu \sum_{i \in \mathcal{E}} |h_i(x) + \nabla h_i(x)^T d|$$
$$+\nu \sum_{i \in \mathcal{I}} \textsf{max}(0, -g_i(x) - \nabla g_i(x)^T d)$$

$$\min_d \quad \ell(d)$$
$$\textsf{s.t.} \quad \|d\|_\infty \le \Delta^{\textsf{LP}}$$

$$\mathcal{W}(x) = \{i \in \mathcal{E} \mid h_i(x) + \nabla h_i(x)^T d^{\textsf{LP}} = 0\} \cup$$
$$\{i \in \mathcal{I} \mid g_i(x) + \nabla g_i(x)^T d^{\textsf{LP}} = 0\}$$

Cauchy Point

Minimize quadratic model along LP direction

$$q(d) = \ell(d) + \frac{1}{2} d^T H(x,\lambda) d,$$

Cauchy Point

Minimize quadratic model along LP direction

$$q(d) = \ell(d) + \frac{1}{2} d^T H(x, \lambda) d,$$

$$\min_{\alpha} q(\alpha d^{\mathsf{LP}}), \qquad \alpha \in (0, 1]$$

Cauchy Point

Minimize quadratic model along LP direction

$$q(d) = \ell(d) + \frac{1}{2} d^T H(x, \lambda) d,$$

$$\min_{\alpha} q(\alpha d^{\mathsf{LP}}), \qquad \alpha \in (0, 1]$$

Global convergence theory based on Cauchy step.

- Byrd, Gould, Nocedal, W. (2004)

EQP Subproblem (SLIQUE)

$$
\begin{array}{ll}
\min\limits_{d} & \frac{1}{2} d^T H(x,\lambda) d + \nabla f(x)^T d \\
\text{s.t.} & h_i(x) + \nabla h_i(x)^T d = 0, \quad i \in \mathcal{E} \cap \mathcal{W} \\
 & g_i(x) + \nabla g_i(x)^T d = 0, \quad i \in \mathcal{I} \cap \mathcal{W} \\
 & \|d\|_2 \leq \Delta
\end{array}
$$

- Solve using projected CG (e.g. KNITRO PCG, GLTR)

Total Step

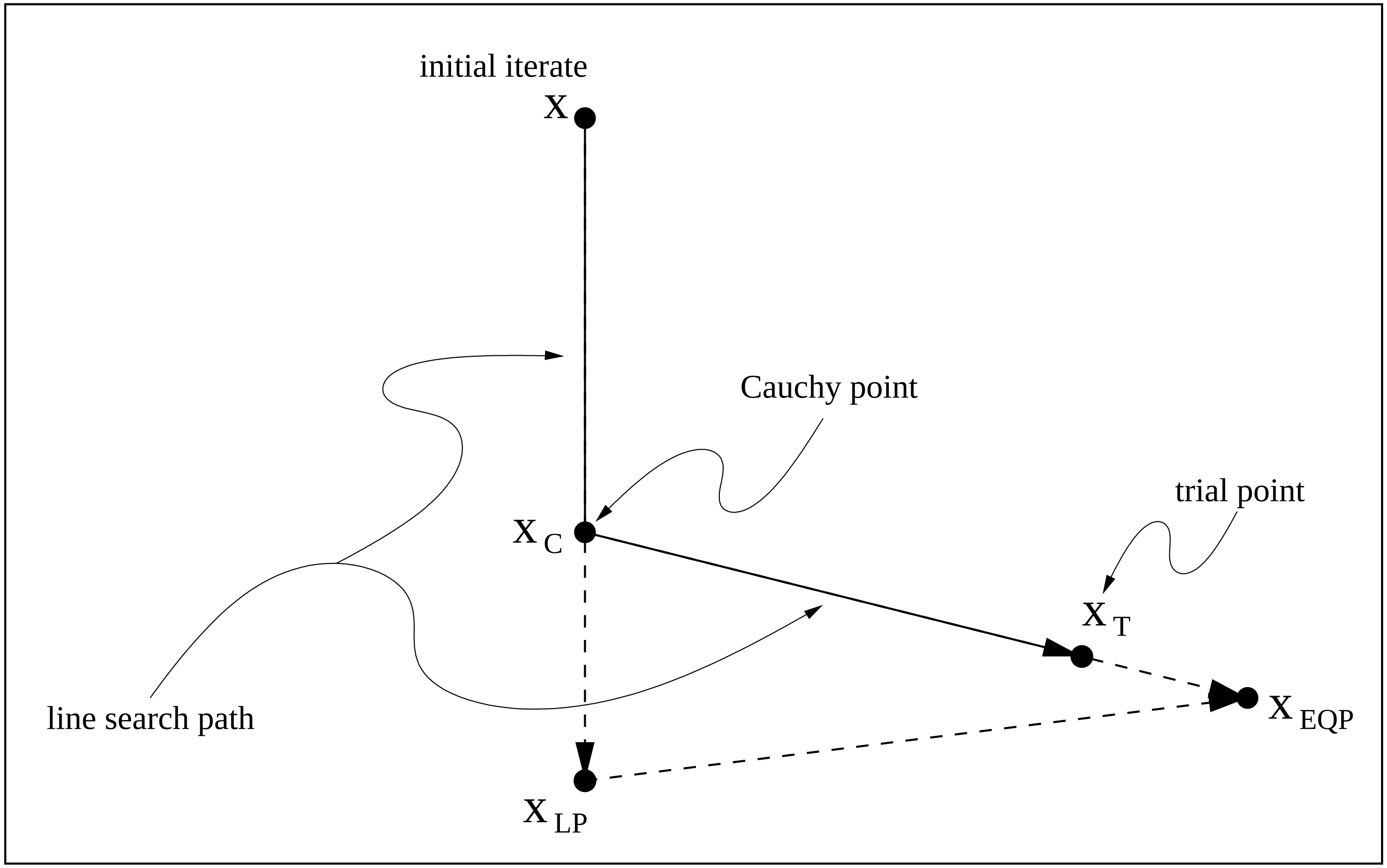


Figure 0: Line search path for step computation.

Penalty Parameter Update

LP subproblem:

$$\begin{aligned} \min_{d} \quad & \nabla f(x)^T d + \nu \sum_{i \in \mathcal{E}} |h_i(x) + \nabla h_i(x)^T d| \\ & + \nu \sum_{i \in \mathcal{I}} \text{max}(0, -g_i(x) - \nabla g_i(x)^T d) \\ \text{s.t} \quad & \|d\|_\infty \le \Delta^{\text{LP}} \end{aligned}$$

Choose ν such that:

1. Satisfy linearized constraints if possible
2. Else achieve some sufficient reduction in linearized infeasibility

Results Summary (Robustness)

CUTEr problems with inequality constraints/bounds

Problem class	Sample size	KNITRO (IP) % optimal	SLIQUE (SLQP) % optimal	SNOPT (SQP) % optimal
$1 \leq n+m < 1000$	344	89.0	88.1	92.1
$1000 \leq n+m < 10000$	125	81.6	73.6	70.4
$10000 \leq n+m$	147	76.9	68.7	53.7
Total	616	84.6	80.5	79.2
DOF > 2000	171	82.5	78.9	53.8

Table 0: Robustness results by problem size

Results Summary (Timing)

SLIQUE Timing Statistics

Problem size	% LP	% EQP	% AugFact	% Eval	% Other
$1 \leq n+m < 1000$	30.8	13.5	21.0	17.8	16.8
$1000 \leq n+m < 10000$	41.5	13.5	18.3	15.8	10.8
$10000 \leq n+m$	45.0	9.3	29.6	9.5	6.6
Total	41.6	11.4	24.3	13.0	9.6

Table 0: SLIQUE timing results by problem size. Average percentage of time spent on various tasks.

Successes

- Estimating active-set by solving an LP works well
- Generally very efficient and robust on small and medium scale problems
- Better able to solve problems with a large reduced space compared to SQP
- Able to get a penalty method to work well
- ...

Difficulties - Ineffcencies in the LP

$$\min_d \quad \nabla f(x)^T d + \nu \sum_{i \in \mathcal{E}} |h_i(x) + \nabla h_i(x)^T d|$$

$$+\nu \sum_{i \in \mathcal{I}} \mathsf{max}(0, -g_i(x) - \nabla g_i(x)^T d)$$

$$\text{s.t} \quad \|d\|_\infty \leq \Delta^{\mathsf{LP}}$$

Difficulties - Inefficencies in the LP

$$\min_{d} \quad \nabla f(x)^T d + \nu \sum_{i \in \mathcal{E}} |h_i(x) + \nabla h_i(x)^T d|$$

$$+\nu \sum_{i \in \mathcal{I}} \text{max}(0, -g_i(x) - \nabla g_i(x)^T d)$$

$$\text{s.t} \quad \|d\|_\infty \leq \Delta^{\text{LP}}$$

1. Many LP TR constraints active at solution
2. Problem constraints stabilize, but TR constraints don't!

Difficulties - Ineffcencies in the LP

Attempts to reduce the LP cost:

- Give better initial basis to LP - improve warm starts

Difficulties - Ineffcencies in the LP

Attempts to reduce the LP cost:

- Give better initial basis to LP - improve warm starts
  1. Apply grad. proj. to estimate active TR constraints

Difficulties - Ineffcencies in the LP

Attempts to reduce the LP cost:

- Give better initial basis to LP - improve warm starts
  1. Apply grad. proj. to estimate active TR constraints
  2. Solve LP with initial basis given by:
     - active problem constraints from previous LP
     - active TR constraints from gradient projection

Difficulties - Ineffcencies in the LP

Attempts to reduce the LP cost:

- Give better initial basis to LP - improve warm starts
  1. Apply grad. proj. to estimate active TR constraints
  2. Solve LP with initial basis given by:
     - active problem constraints from previous LP
     - active TR constraints from gradient projection
- Truncate LP (stop with inexact solution)

Difficulties - Ineffcencies in the LP

Attempts to reduce the LP cost:

- Give better initial basis to LP - improve warm starts
  1. Apply grad. proj. to estimate active TR constraints
  2. Solve LP with initial basis given by:
     - active problem constraints from previous LP
     - active TR constraints from gradient projection

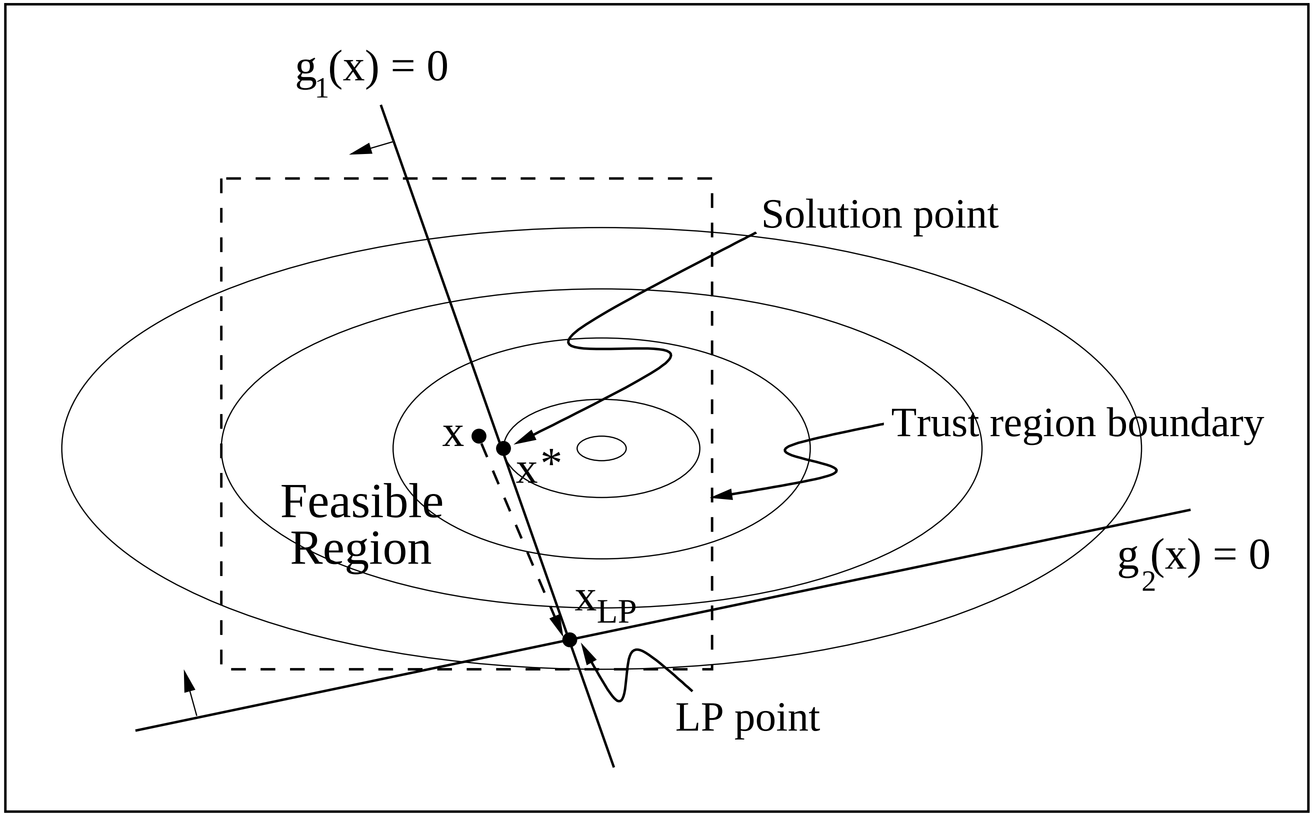
- Truncate LP (stop with inexact solution)

- Skip LP (when problem constraints stabilize)
  - When to skip?
  - How to recover if active-set is wrong?

Difficulties - Two Trust-Regions

How to manage 2 trust-regions?

- LP: box trust-region must not get “too big” near solution
- EQP: spherical trust-region (standard)



Summary/Conclusions

- Demonstrated viability of the SLQP method
- Improvement over SQP on large-scale problems
- Not yet competitive with IP on large-scale problems
- Demonstrated success of penalty method
- Still room for improvement...