Two Approaches for Nonlinear Programming

- Richard Waltz, Northwestern University

  with

- Richard Byrd, University of Colorado
- Nick Gould, Rutherford Appleton Laboratory, UK
- Jorge Nocedal, Northwestern University

Motivation

- Diversity of nonlinear optimization probs. requires diversity of algorithms
- Interior-point methods are great but not a cure all
- Need for large-scale active-set method

Goal

- Develop comprehensive nonlinear optimization package offering a variety of algorithms/options
  - Interior-point and Active-set
  - Iterative and Direct approaches
  - 1st and 2nd derivative options
  - Large-scale capabilities
  - Choose appropriate algorithm/options based on problem characteristics
  - Adaptive techniques

Two Approaches

- Interior Point
  - Iterative (CG) & Direct versions
  - Well-established
  - **KNITRO**
- SLPEQP
  - Active-set approach
  - LP and EQP subprobs
  - Relatively new
  - **SLIQUE**

Why these two?

- Suitable for large-scale problems
- Approaches are complementary

Problem

NLP

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

- Functions twice continuously differentiable

SLPEQP Overview

- Given: x
- Solve LP to get working set $\mathcal{W}$
- Compute a step, d, by solving the EQP:

$$
\begin{aligned}
\min_{d} \quad & \nabla f(x)^{T} d + \frac{1}{2} d^{T} H d \\
\text{s.t.} \quad & h_i(x) + \nabla h_i(x)^{T} d = 0, \quad i \in \mathcal{W} \cap E \\
& g_i(x) + \nabla g_i(x)^{T} d = 0, \quad i \in \mathcal{W} \cap I \\
& \|d\|_2 \le \Delta_{\text{EQP}}
\end{aligned}
$$

- Set: $x_T = x+d$

LP Solution

- Estimate active set by solving an LP

$$\begin{aligned} \min_{d_{\mathrm{LP}}} \quad & \nabla f(x)^{\mathrm{T}} d_{\mathrm{LP}} \\ \text{s.t.} \quad & h_i(x) + \nabla h_i(x)^{\mathrm{T}} d_{\mathrm{LP}} = 0, \quad i \in E \\ & g_i(x) + \nabla g_i(x)^{\mathrm{T}} d_{\mathrm{LP}} \geq 0, \quad i \in I \\ & \|d_{\mathrm{LP}}\|_\infty \leq \Delta_{\mathrm{LP}} \end{aligned}$$

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

l_1 LP formulation

- Constraints may be inconsistent!
- L_1 relaxation

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

Cauchy Point

$$m(d) = \nabla f(x)^{\mathrm{T}} d + \tfrac{1}{2} d^{\mathrm{T}} H d$$
$$+ \nu \sum_{i \in E} | h_i(x) + \nabla h_i(x)^{\mathrm{T}} d |$$
$$+ \nu \sum_{i \in I} \max(0, -g_i(x) - \nabla g_i(x)^{\mathrm{T}} d)$$

$$\min_{\alpha} \; m(\alpha d^*_{\mathrm{LP}}), \quad \alpha \in [0,1]$$

$$x_{\mathrm{Cauchy}} = x + \alpha_1 d^*_{\mathrm{LP}}$$

EQP solution

- Start EQP step from $x_{LP} = x + d^*_{LP}$

$$\min_{d_{EQP}} \quad g_{EQP}^T d_{EQP} + \frac{1}{2} d_{EQP}^T H_{EQP} d_{EQP}$$

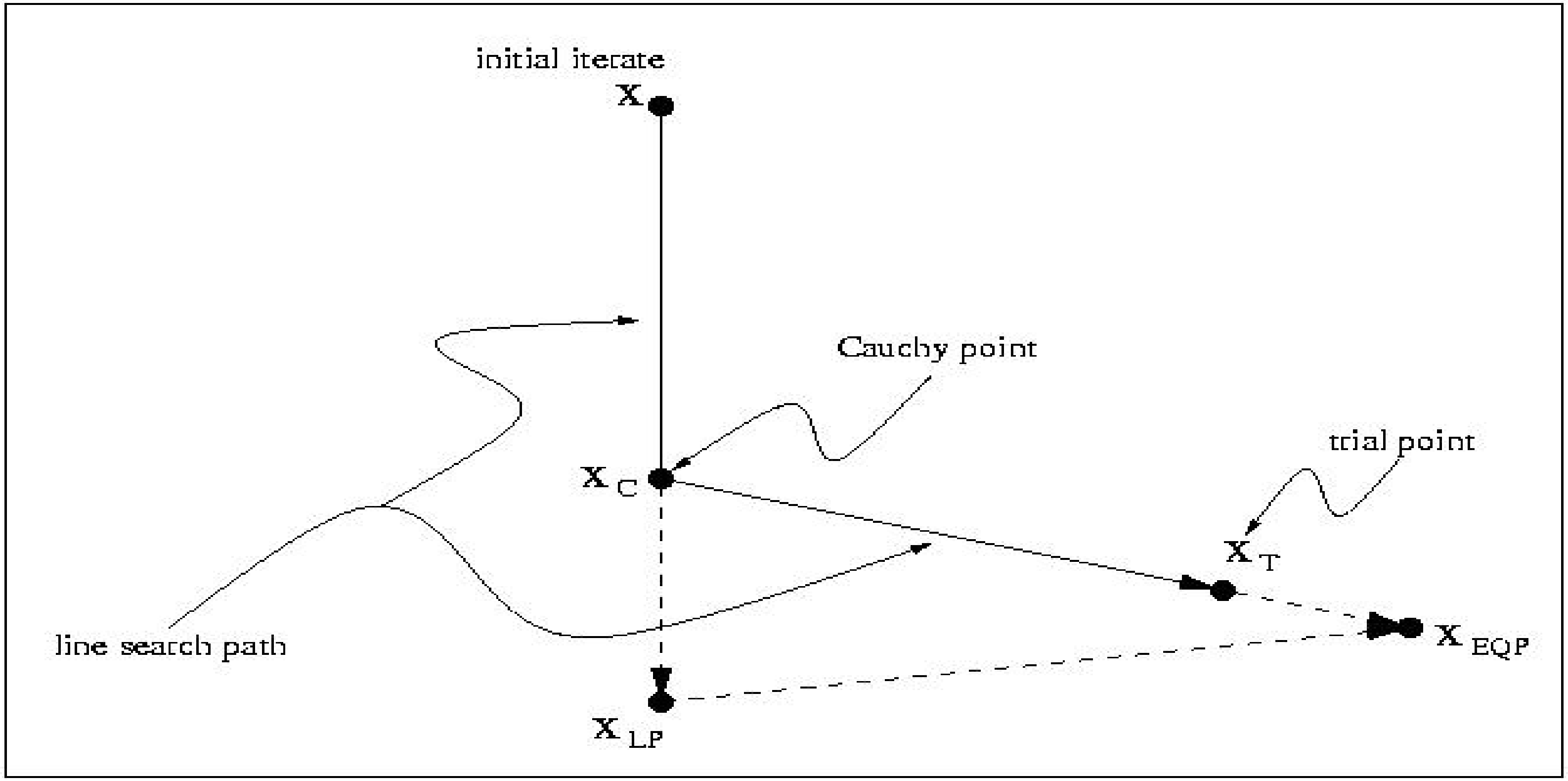
$$\text{s.t.} \quad A_{\mathcal{W}} d_{EQP} = 0$$

$$\|d_{EQP}\|_2 \le \Delta_{EQP}$$

- Solve using Projected Conjugate Gradient

Trial Step

Define: $$x_{\mathrm{T}} = x_{\mathrm{Cauchy}} + \alpha_2 (x_{\mathrm{EQP}} - x_{\mathrm{Cauchy}})$$



Strengths/Weaknesses SLPEQP

- Strengths:
  - Only LP and EQP subproblems
  - Good active-set identification
  - Good warm start properties
  - Only factor systems including active constraints

- Weaknesses:
  - Combinatorial difficulties (can slow things down)
  - Large-scale?

Software Comparison

- Compare

  - SLIQUE (SLPEQP)
  - KNITRO-CG (Interior Point-CG)
  - KNITRO-Dir (Interior Point-Direct)

Costs per iter (SLIQUE)

- Solution of LP
  - Simplex vs. IP
  - Cheaper near the solution because of warm starts

- Solution of EQP (via projected CG)
  - Factor

$$\begin{pmatrix} \mathbf{I} & \mathbf{A}_{\mathcal{W}}^{T} \\ \mathbf{A}_{\mathcal{W}} & 0 \end{pmatrix}$$

  - Hessian-vector products
  - Backsolves

Costs per iteration (KNITRO)

KNITRO-CG

- Factor

$$\begin{pmatrix} \mathbf{I} & \mathbf{A}^T \\ \mathbf{A} & 0 \end{pmatrix}$$

- Solve EQP via projected CG
  - Hessian-vector products
  - Backsolves

KNITRO-Direct

- Factor

$$\begin{pmatrix} \mathbf{H} & \mathbf{A}^T \\ \mathbf{A} & 0 \end{pmatrix}$$

Tradeoffs (Summary)

Prob char.	**SLIQUE**	Kn(CG)	Kn(Dir)
H expensive	+	+	--
H ill-conditioned	--	--	+
$\mathbf{A}_{\mathcal{W}} \ll \mathbf{A}$	+	--	--
Warm starts	+	--	--
Active set info	+	--	--
Large-scale	?	+	+

+ favorable; --unfavorable

Example 1

- Problem CVXQP2
  - n=10,000, m=2,500 + bounds
  - nnzH = 40,000
  - 99.6% time spent on factor in KNITRO-Dir

Code	iters	time	time/it
SLIQUE	24	612	25.5
KNITRO-CG	11	401	36.5
KNITRO-DIR	14	2638	188.4

Example 2

- Problem BQPGAUSS
  - n = 2003, bound-constrained
  - H ill-conditioned but not expensive

Code	Iters	time	time/it
SLIQUE	135	1791	13.3
KNITRO-CG	27	1310	48.5
KNITRO-DIR	19	3	0.16

Example 3

- Problem SAWPATH
  - n=583, m=774, nnzA =3285
  - Only 3.6 second spent on matrix factorization in SLIQUE

Code	iters	time	Time/it
SLIQUE	25	15	0.6
KNITRO-CG	8	37	4.6
KNITRO-DIR	11	79	7.2

Robustness Results by Problem Size

Size	#	SLIQUE	KNITRO-CG
VS	554	526	514
S	190	155	163
M	135	105	113
L	66	32	52
TOT	**945**	**818**	**842**

VS: $n + m < 100$

S: $100 <= n + m < 1000$

M: $1000 <= n + m < 10000$

L: $10000 <= n + m$

Summary

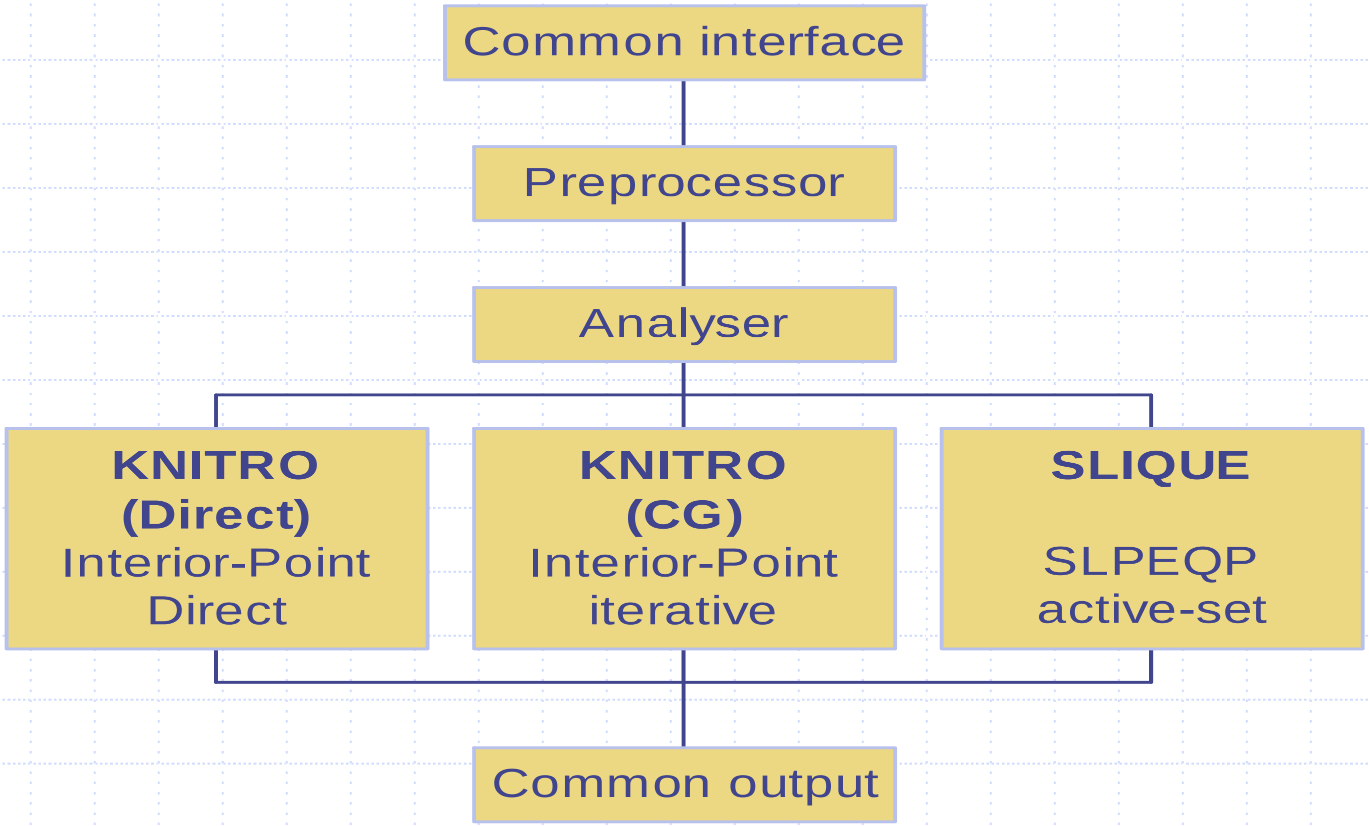
- Diversity of nonlinear problems calls for integrated approach
- Requires:
  - Characterization of problem
  - Match with appropriate algorithm
  - Good defaults
  - Adaptive (crossover) techniques

    Interior-Point ←→ Active-set

    Iterative ←→ Direct

KNITRO future developments

KNITRO Optimization Package



References

- www.ece.northwestern.edu/~rwaltz
- www.ziena.com/knitro/kindex.htm