Adaptive Methods for Large-Scale Nonlinear Optimization

Frank E. Curtis, Lehigh University

involving joint work with

James V. Burke (U. of Washington), **Richard H. Byrd** (U. of Colorado),
Nicholas I. M. Gould (RAL), **Johannes Huber** (SAFEmine),
Hao Jiang (Johns Hopkins), **Travis Johnson** (Square, Inc.),
Jorge Nocedal (Northwestern), **Daniel P. Robinson** (Johns Hopkins),
Olaf Schenk (U. della Svizzera Italiana), **Philippe L. Toint**, (U. of Namur),
Andreas Wächter (Northwestern), **Hao Wang** (Lehigh)

SIAM Conference on CSE — Salt Lake City, UT

17 March 2015

Outline

Outline

Constrained nonlinear optimization

Consider the constrained nonlinear optimization problem

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \ & f(x) \\ \text{s.t. } & c(x) \le 0, \end{aligned} \tag{NLP}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $c : \mathbb{R}^n \to \mathbb{R}^m$ are continuously differentiable.

(Equality constraints also OK; just suppressed for simplicity.)

We are interested in algorithms such that if (NLP) is infeasible, then there will be an automatic transition to solving the feasibility problem

$$\min_{x\in\mathbb{R}^n} \ v(x), \quad \text{where} \quad v(x) = \text{dist}(c(x)|\mathbb{R}^m_-). \tag{FP}$$

Any feasible point for (NLP) is an optimal solution of (FP).

Inefficiencies of traditional approaches

The traditional NLP algorithm classes, i.e.,

- augmented Lagrangian (AL) methods
- sequential quadratic optimization (SQP) methods
- interior-point (IP) methods

may fail or be inefficient when

- exact subproblem solves are expensive
- ...or inexact solves are not computed intelligently
- algorithmic parameters are initialized poorly
- ...or are updated too slowly or inappropriately
- a globalization mechanism inhibits productive early steps
- ...or blocks superlinear local convergence

This is especially important when your subproblems are NLPs!

Contributions

A variety of algorithms and algorithmic tools incorporating/allowing

- inexact subproblem solves
- flexible step acceptance strategies
- adaptive parameter updates
- global convergence guarantees
- superlinear local convergence guarantees
- efficient handling of nonconvexity

This talk provides an overview of these tools within

- AL methods with adaptive penalty parameter updates
- SQP methods with inexact subproblem solves
- IP methods with inexact linear system solves
- a penalty-IP method with adaptive parameter updates

Outline

AL methods

A traditional augmented Lagrangian (AL) method for solving

$$\min_{x\in\mathbb{R}^n, s\in\mathbb{R}^m_+} f(x) \quad \text{s.t.} \quad c(x)+s=0,$$

observes the following strategy:

- Given $\rho \in \mathbb{R}_+$ and $y \in \mathbb{R}^m$, approximately solve

$$\min_{x\in\mathbb{R}^n, s\in\mathbb{R}^m_+} \rho f(x) + (c(x)+s)^T y + \tfrac{1}{2}\|c(x)+s\|_2^2$$

- Update ρ and y to drive (global and local) convergence

Potential inefficiencies:

- Poor initial (ρ, y) may ruin good initial (x, s)
- Slow/poor update for (ρ, y) may lead to poor performance

Ideas: Steering rules and adaptive multiplier updates

"Steering" rules for exact penalty methods: Byrd, Nocedal, Waltz (2008)

- Do not fix ρ during minimization
- Rather, before accepting any step, reduce ρ until the step yields progress in a model of constraint violation proportional to that yielded by a feasibility step
- Cannot be applied directly in an AL method

"Steering" rules for AL methods: Curtis, Jiang, Robinson (2014) w/ Gould (2015)

- Modified steering rules that may allow constraint violation increase
- Simultaneously incorporate adaptive multiplier updates
- Gains in performance in trust region and line search contexts
- Implementation (partially) in LANCELOT

(Also, "steering" rules in penalty-IP method: Curtis (2012))

SQP methods

A traditional sequential quadratic optimization (SQP) method follows:

- Given $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$, solve

$$\begin{aligned} \min_{d \in \mathbb{R}^n} \ & f(x) + g(x)^T d + \tfrac{1}{2} d^T H(x,y) d \\ \text{s.t. } & c(x) + J(x) d \leq 0 \end{aligned} \tag{QP}$$

- Set $\rho \in \mathbb{R}_+$ to ensure d yields sufficient descent in

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

Potential inefficiencies/issues:

- (QP) may be infeasible
- (QP) expensive to solve exactly
- inexact solves might not ensure d yields descent in $\phi(x;\rho)$
- "steering" not viable with inexact solves

Ideas (equalities only): Step decomposition and SMART tests

Step decomposition in trust region (TR) framework: Celis, Dennis, Tapia (1985)

- Normal step toward constraint satisfaction
- Tangential step toward optimality in null space of constraints
- Requires projections during tangential computation

Normal step with TR, but TR-free tangential: Curtis, Nocedal, Wächter (2009)

- Incorporate "SMART" tests: Byrd, Curtis, Nocedal (2008, 2010)
- Normal and tangential steps can be computed approximately
- Consider various types of inexact solutions
- Prescribed inexactness conditions based on penalty function model reduction

Ideas (inequalities, too): Feasibility and optimality steps w/ scenarios

Inexact SQP method: Curtis, Johnson, Robinson, Wächter (2014)

- Similar to steering, compute approximate(!) feasibility step for reference
- Also given $\rho \in \mathbb{R}_+$, solve

$$\begin{aligned} \min_{d \in \mathbb{R}^n} \ & \rho(f(x) + g(x)^T d) + e^T s + \tfrac{1}{2} d^T H(\rho, x, y) d \\ \text{s.t.} \ & c(x) + J(x) d \leq s \end{aligned} \tag{PQP}$$

- Consider various types of inexact solutions
  - Approximate Sℓ_1QP step
  - Multiplier-only step
  - Convex combination of feasibility and Sℓ_1QP step

IP methods

A traditional interior-point (IP) method for solving

$$\min_{x \in \mathbb{R}^n, s \in \mathbb{R}^m_+} f(x) \quad \text{s.t.} \quad c(x) + s = 0,$$

observes the following strategy:

- Given $\mu \in \mathbb{R}_+$, approximately solve

$$\min_{x \in \mathbb{R}^n, s \in \mathbb{R}^m_+} f(x) - \mu \sum_{i=1}^{m} \ln s^i \quad \text{s.t.} \quad c(x) + s = 0$$

- Update μ to drive (global and local) convergence

Potential inefficiences:

- Direct factorizations for Newton's in subproblem can be expensive
- Slack bounds can block long steps ("jamming")
- Slow/poor update for μ may lead to poor performance

Ideas: Inexact IP method

Step decomposition with scaled trust region: Byrd, Hribar, Nocedal (1999)

$$\left\| \begin{bmatrix} d_k^x \\ S_k^{-1} d_k^s \end{bmatrix} \right\|_2 \leq \Delta_k$$

- Allow inexact linear system solves: Curtis, Schenk, Wächter (2010)
- Normal step with TR, but TR-free tangential
- Incorporate (modified) “SMART” tests
- Implemented in IPOPT (optimizer) with inexact, iterative linear system solves by PARDISO (SQMR): Curtis, Huber, Schenk, Wächter (2011)

Ideas: Penalty-IP method

"Jamming" can be avoided by relaxation via penalty methods

- Two parameters to juggle: ρ and μ
- Simultaneous update motivated by "steering" penalty methods...
- ... and "adaptive barrier" method: Nocedal, Wächter, Waltz (2009)
- ... leading to penalty-interior-point method: Curtis (2012)
- More reliable than IPOPT on infeasible and degenerate problems

Ideas: Trust-funnel IP method

Trust-funnel method for equality constrained problems: Gould, Toint (2010)

- Does not require a penalty function or a filter
- Drives convergence by a monotonically decreasing sequence with

$$v(x_k) \leq v_k^{max} \quad \text{for all} \quad k$$

- Normal and tangential step decomposition, but allows flexible computation that may allow skipping certain subproblems in some iterations
- IP method with scaled trust regions: Curtis, Gould, Robinson, Toint (2015)

Outline

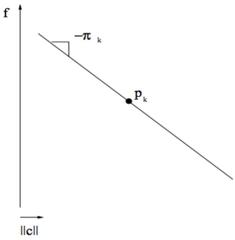
Idea: Flexible penalty function

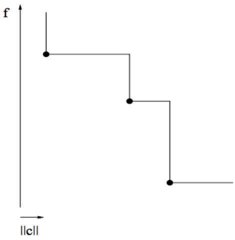
A traditional penalty/merit function—Han (1977); Powell (1978)—requires

$$f(x_k + \alpha_k d_k) + \pi_k v(x_k + \alpha_k d_k) \ll f(x_k) + \pi_k v(x_k)$$

while a filter—Fletcher, Leyffer (1998)—requires

$$f(x_k + \alpha_k d_k) \ll f_j \quad \text{or} \quad v(x_k + \alpha_k d_k) \ll v_j \quad \text{for all} \quad j \in \mathcal{F}_k$$



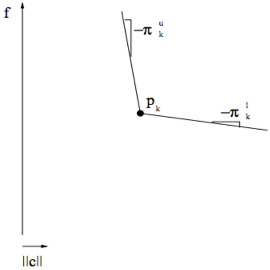


Either can result in “blocking” of certain steps

Idea: Flexible penalty function

A flexible penalty function—Curtis, Nocedal (2008)—requires

$$f(x_k + \alpha_k d_k) + \pi v(x_k + \alpha_k d_k) \ll f(x_k) + \pi v(x_k) \quad \text{for some} \quad \pi \in [\pi_k^l, \pi_k^u]$$



Parameters π_k^l and π_k^u updated separately

Of course, "blocking" may still occur, but (hopefully) less often

Idea: Rapid infeasibility detection

Suppose your algorithm is "driven" by a penalty/merit function such as

$$\phi(x) = \rho f(x) + v(x)$$

and, at an iterate x_k, the following occur:

- x_k is infeasible for (NLP) in that $v(x_k) \gg 0$
- you have computed a feasibility step that does not reduce a model of v sufficiently compared to $v(x_k)$

Then, there is reason to believe that (NLP) is (locally) infeasible, and you may obtain superlinear convergence to an infeasible stationary point by setting

$$\rho \leq \|\text{KKT for (FP)}\|^2$$

Byrd, Curtis, Nocedal (2010); Burke, Curtis, Wang (2014)

Idea: Dynamic Hessian modifications

Inexact SQP and IP methods involve iterative solves of systems of the form

$$\begin{bmatrix} H_k & J_k^T \\ J_k & 0 \end{bmatrix} \begin{bmatrix} \Delta x_k \\ \Delta y_k \end{bmatrix} = \begin{bmatrix} g_k \\ c_k \end{bmatrix}$$

If H_k is not positive definite in the null space of J_k, then even an exact solution may not be productive in terms of optimization

- Apply a symmetric indefinite iterative solver
- Under certain conditions, modify H_k (say, adding some multiple of a positive definite matrix) and restart the solve (hopefully not entirely from scratch)

Outline

Contributions

A variety of algorithms and algorithmic tools incorporating/allowing

- inexact subproblem solves
- flexible step acceptance strategies
- adaptive parameter updates
- global convergence guarantees
- superlinear local convergence guarantees
- efficient handling of nonconvexity

We have outlined these tools within

- an AL method with adaptive penalty parameter updates
- SQP methods with inexact subproblem solves
- an IP method with inexact linear system solves
- a penalty-IP method with dynamic parameter updates

Bibliography I

[1] J. V. Burke, F. E. Curtis, and H. Wang.
A Sequential Quadratic Optimization Algorithm with Rapid Infeasibility Detection.
SIAM Journal on Optimization, 24(2):839–872, 2014.

[2] R. H. Byrd, F. E. Curtis, and J. Nocedal.
An Inexact SQP Method for Equality Constrained Optimization.
SIAM Journal on Optimization, 19(1):351–369, 2008.

[3] R. H. Byrd, F. E. Curtis, and J. Nocedal.
An Inexact Newton Method for Nonconvex Equality Constrained Optimization.
Mathematical Programming, 122(2):273–299, 2010.

[4] R. H. Byrd, F. E. Curtis, and J. Nocedal.
Infeasibility Detection and SQP Methods for Nonlinear Optimization.
SIAM Journal on Optimization, 20(5):2281–2299, 2010.

[5] F. E. Curtis.
A Penalty-Interior-Point Algorithm for Nonlinear Constrained Optimization.
Mathematical Programming Computation, 4(2):181–209, 2012.

[6] F. E. Curtis, N. I. M. Gould, H. Jiang, and D. P. Robinson.
Adaptive Augmented Lagrangian Methods: Algorithms and Practical Numerical Experience.
Technical Report 14T-006, COR@L Laboratory, Department of ISE, Lehigh University, 2014. In first review for *Optimization Methods and Software*.

[7] F. E. Curtis, N. I. M. Gould, D. P. Robinson, and Ph. L. Toint.
An Interior-Point Trust-Funnel Algorithm for Nonlinear Optimization.
Technical Report 13T-010, COR@L Laboratory, Department of ISE, Lehigh University, 2013. In second review for *Mathematical Programming*.

Bibliography II

[8] F. E. Curtis, J. Huber, O. Schenk, and A. Wächter.
A Note on the Implementation of an Interior-Point Algorithm for Nonlinear Optimization with Inexact Step Computations.
Mathematical Programming, Series B, 136(1):209–227, 2012.

[9] F. E. Curtis, H. Jiang, and D. P. Robinson.
An Adaptive Augmented Lagrangian Method for Large-Scale Constrained Optimization.
Mathematical Programming, DOI: 10.1007/s10107-014-0784-y, 2014.

[10] F. E. Curtis, T. Johnson, D. P. Robinson, and A. Wächter.
An Inexact Sequential Quadratic Optimization Algorithm for Large-Scale Nonlinear Optimization.
SIAM Journal on Optimization, 24(3):1041–1074, 2014.

[11] F. E. Curtis and J. Nocedal.
Flexible Penalty Functions for Nonlinear Constrained Optimization.
IMA Journal of Numerical Analysis, 28(4):749–769, 2008.

[12] 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, 20(3):1224–1249, 2009.

[13] 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, 32(6):3447–3475, 2010.