Towards a Regularized Funnel Algorithm for Constrained Optimization

Stefania Bellavia Coralia Cartis Nick Gould Benedetta Morini
Philippe Toint

Department of Mathematics,University of Namur, Belgium

(philippe.toint@fundp.ac.be)

SIAM Conference on Optimization - Boston, May 2008

Outline

1. A trust-funnel approach
2. Regularization techniques
   - Cubic
   - Quadratic
3. A regularized funnel method?
4. Conclusions

Outline

1. A trust-funnel approach
2. Regularization techniques
   - Cubic
   - Quadratic
3. A regularized funnel method?
4. Conclusions

Outline

1. A trust-funnel approach
2. Regularization techniques
   - Cubic
   - Quadratic
3. A regularized funnel method?
4. Conclusions

Outline

The problem

We consider the equality constrained nonlinear programming problem:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & c(x) = 0 \end{array}$$

for $x \in \mathbb{R}^n$, f and c smooth.

Work in progress. . .

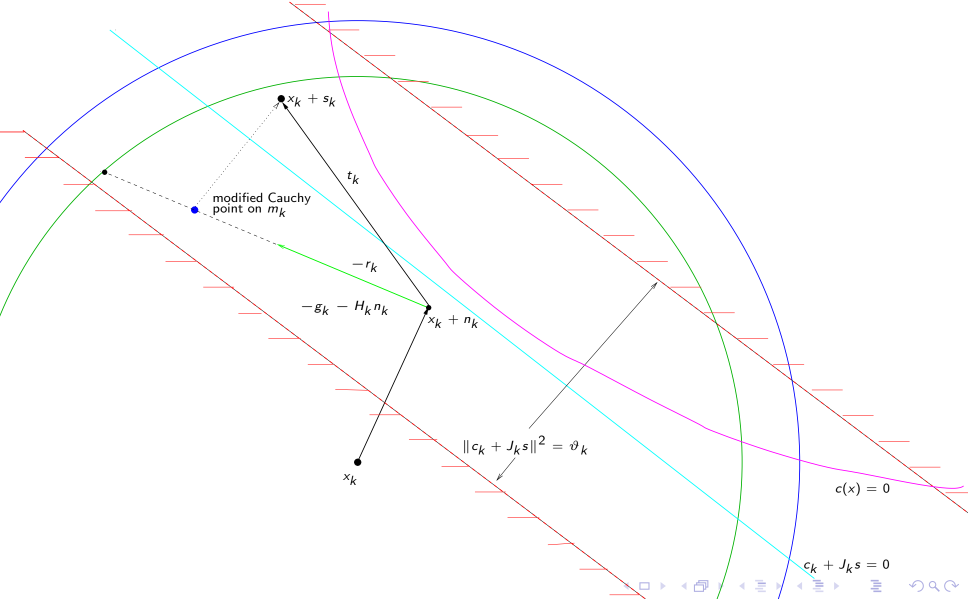
A Trust-Funnel Approach

What is the trust-funnel approach?

An inexact SQP algorithm for equality constrained problems

- two distinct trust-regions
  (constraint violation, objective function)
- normal + tangential steps
  (separate Cauchy conditions)
- no penalty/barrier parameter, no filter
- asymptotic feasiblity (shrinking funnel)
- promising numerical performance

The complete step



Regularization Techniques

Unconstrained optimization — a “mature” area?

$$\min_{x\in\mathbf{R}^n} f(x) \quad \text{where} \quad f \in C^1 \quad (\text{maybe} \quad C^2 \ )$$

Currently two main competing (but similar) methodologies

- **Linesearch methods**
  - compute a descent direction s_k from x_k
  - set $x_{k+1} = x_k + \alpha_k s_k$ to improve f
- **Trust-region methods**
  - compute a step s_k from x_k to improve a model m_k of f within the trust-region $\|s\| \leq \Delta$
  - set $x_{k+1} = x_k + s_k$ if m_k and f “agree” at $x_k + s_k$
  - otherwise set $x_{k+1} = x_k$ and reduce the radius Δ

Is there anything more to say?

Observe the following: if

- f has gradient g and globally Lipschitz continuous Hessian H with constant $2L$

Taylor, Cauchy-Schwarz and Lipschitz imply

$$\begin{aligned} f(x+s) &= f(x) + \langle s, g(x)\rangle + \tfrac{1}{2}\langle s, H(x)s\rangle \\ &\quad + \int_0^1 (1-\alpha)\langle s, [H(x+\alpha s) - H(x)]s\rangle \, d\alpha \\ &\le \underbrace{f(x) + \langle s, g(x)\rangle + \tfrac{1}{2}\langle s, H(x)s\rangle + \tfrac{1}{3}L\|s\|_2^3}_{m(s)} \end{aligned}$$

$\Longrightarrow$ reducing m from $s = 0$ improves f since $m(0) = f(x)$.

The cubic regularization

Change from

$$\min_s \quad f(x) + \langle s, g(x)\rangle + \tfrac{1}{2}\langle s, H(x)s\rangle \;\text{ s.t. }\; \|s\| \le \Delta$$

to

$$\min_s \quad f(x) + \langle s, g(x)\rangle + \tfrac{1}{2}\langle s, H(x)s\rangle + \tfrac{1}{3}\sigma\|s\|^3$$

σ is the (adaptive) regularization parameter

(ideas from Griewank, Weiser/Deuflhard/Erdmann, Nesterov/Polyak, Cartis/Gould/T)

The main features of adaptive cubic regularization

And the result is. . .

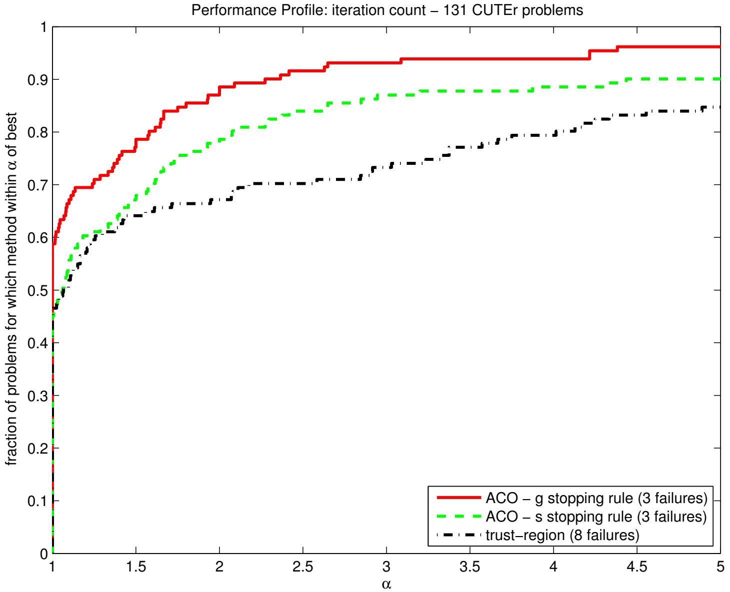
longer steps on ill-conditioned problems

similar (very satisfactory) convergence analysis

best known worst-case complexity for nonconvex problems

excellent performance and reliability

Numerical experience — small problems using Matlab



The quadratic regularization for NLS

Change from

$$\min_s \quad \tfrac{1}{2}\|c(x)\|^2 + \langle s, J(x)^T c(x)\rangle + \tfrac{1}{2}\langle s, J(x)^T J(x) s\rangle \ \text{ s.t. } \ \|s\| \leq \Delta$$

to

$$\min_s \quad \|c(x) + J(x)s\| + \tfrac{1}{2}\sigma\|s\|^2$$

σ is the (adaptive) regularization parameter

(idea by Nesterov)

Quadratic regularization: reformulation

Note that

$$\min_s \quad \|c(x) + J(x)s\| + \tfrac{1}{2}\sigma\|s\|^2$$

$$\Leftrightarrow$$

$$\min_{\nu,s} \quad \nu + \tfrac{1}{2}\sigma\|s\|^2$$

such that

$$\|c(x) + J(x)s\|^2 = \nu^2$$

exact penalty function for the problem of minimizing $\|s\|$ subject to $c(x) + J(x)s = 0$.

Convergence theory for the quadratic regularization

Cauchy-point condition:

$$m(x_k) - m(x_k + s_k) \geq \frac{\|J_k^T c_k\|^2}{4\|c_k\|} \min\left[\frac{1}{\sigma_k \|c_k\|}, \frac{1}{1 + \|J_k^T J_k\|}\right]$$

. . . and hence. . .

Global convergence to first-order critical points

(and more. . .)

Computing regularization steps

Iterative techniques. . .

solve the problem in nested Krylov subspaces

- Lanczos $\longrightarrow$ basis of the Krylov subspace
- $\longrightarrow$ factorization of tridiagonal matrices
- different scalar secular equation (solution by Newton's method)

Approach valid for

- trust-region (GLTR),
- cubic and quadratic regularizations

(details in CGT techreport)

A Regularized Funnel Method?

Next on the "todo" list:

- which regularization for the normal step? (cubic, quadratic)

$$m_c(x_k) - m_c(x_k + s_k) \geq \kappa \|J_k^T c_k\| \min \left[\frac{\|J_k^T c_k\|}{1 + \|J_k^T J_k\|}, \sqrt{\frac{\|J_k^T c_k\|}{\sigma_k^n}} \right]$$

$$m_q(x_k) - m_q(x_k + s_k) \geq \frac{\|J_k^T c_k\|}{4\|c_k\|} \min \left[\frac{\|J_k^T c_k\|}{1 + \|J_k^T J_k\|}, \frac{\|J_k^T c_k\|}{\sigma_k^n \|c_k\|} \right]$$

$\rightarrow$ ongoing numerical experiments (with M. Porcelli)

- coordination of the two regularization parameters
- convergence/complexity analysis
- extension to inequality constraints (e.g. see Nick's talk)
- software and extensive testing
- . . .

Conclusions

- Much left to do... but very interesting
- Could lead to a very untypical method
- Many issues regarding regularizations still unresolved
- ... more detail later!

Thank you for your attention !

(see `http://perso.fundp.ac.be/~phtoint/publications.html` for references)