Trust-region methods for rectangular systems of nonlinear equations

Margherita Porcelli

Dipartimento di Matematica "U.Dini"

Università degli Studi di Firenze

Joint work with Maria Macconi and Benedetta Morini

Università degli Studi di Firenze

CERFACS 20th Anniversary Meeting
Oct 10 - 12 2007, Toulouse

The problem

Bound-constrained nonlinear system (BCNS)

$$\Theta(x) = 0, \qquad x \in \Omega$$

- $\Theta : \mathbb{R}^n \rightarrow \mathbb{R}^m$ continuously differentiable;
- $\Omega = \{x \in \mathbb{R}^n \quad \text{t.c.} \quad l \leq x \leq u\}$;
- $l \in (\mathbb{R} \cup -\infty)^n$, $u \in (\mathbb{R} \cup +\infty)^n$;
- any relationship between m and n is allowed.

• Nonlinear systems arise in several applications: differential problems, equilibrium problems, data-fitting problems.

• Bound constraints on the variables prevent computing solutions without physical meaning.

System of Nonlinear Equalities and Inequalities (NEI)

$$C_E(v) = 0,$$
$$C_I(v) \leq 0,$$
$$l \leq v \leq u,$$

- $C_E : \mathbb{R}^p \to \mathbb{R}^{m_E}$, $C_I : \mathbb{R}^p \to \mathbb{R}^{m_I}$, continuously differentiable.
- Two alternative reformulations:
  1. Add a slack variable $s \in \mathbb{R}^{m_I}$: let $x = (v^T, s^T)^T$

$$\Theta(x) = \begin{pmatrix} C_E(v) \\ s - C_I(v) \end{pmatrix} = 0, \qquad l \leq v \leq u, \; s \leq 0.$$

  2. Use the function $[t]_+ = \max\{t, 0\}^2/2$: let $x = v$

$$\Theta(x) = \begin{pmatrix} C_E(x) \\ \dfrac{\max\{C_I(x), 0\}^2}{2} \end{pmatrix} = 0, \qquad l \leq x \leq u.$$

Trust-region methods

Trust-region methods for BCNS

- Square systems: Bellavia, Macconi and Morini, 2003-2006, Kanzow and Klug, COAP 2007,
- Underdetermined systems: Francisco, Krejić and Martínez, JCAM 2005,
- Overdetermined systems: Kanzow and Petra, OMS 2007.

Trust-region methods for NEI

- Fletcher and Leyffer, Kluwer 2003,
- Dennis, El-Alem and Williamson, SIOPT 1999,
- Gould and Toint, ACM Trans. Math. Software 2007.

BCNS as a bound-constrained least-squares problem

- We address the solution of BCNS considering the following bound-constrained least-squares problem

$$\min_{x \in \Omega} \theta(x) = \frac{1}{2}\|\Theta(x)\|_2^2.$$

  - If x^* is a solution to the BCNS problem $\Rightarrow$ x^* is a global minimizer of θ.
  - If x^* is a zero-residual solution of $\min_{x \in \Omega} \theta(x) \Rightarrow x^*$ is a solution to the BCNS problem.

- Recent algorithms for bound-constrained optimization are reliable and comparable in efficiency with unconstrained algorithms.

The scaling matrix [Coleman and Li, SIOPT 1996]

The first-order optimality conditions for a minimizer of

$$\min_{x \in \Omega} \theta(x) = \frac{1}{2}\|\Theta(x)\|_2^2$$

are equivalent to the system of nonlinear equations

$$D(x)\nabla\theta(x) = D(x)\Theta'(x)^T\Theta(x) = 0,$$

where $D(x) = diag(|v_1(x)|, \ldots, |v_n(x)|)$, and

$$v_i(x) = \begin{cases} x_i - u_i & \text{if } (\nabla\theta(x))_i < 0, \; u_i < \infty \\ x_i - l_i & \text{if } (\nabla\theta(x))_i \geq 0, \; l_i > -\infty \\ 1 & \text{if } (\nabla\theta(x))_i < 0, \; u_i = \infty \\ & \text{or if } (\nabla\theta(x))_i \geq 0, \; l_i = -\infty \end{cases}$$

for $i = 1, \ldots, n$.

Outline of the methods

- Two trust-region methods for $\min_{x \in \Omega} \theta(x) = \frac{1}{2}\|\Theta(x)\|_2^2$ are presented.

  The globalization strategy employs the Coleman and Li scaling matrix.

  The sequence $\{x_k\}$ generated is feasible, i.e. $x_k \in \Omega$, $k \geq 0$.

- The methods differ in the choice of the quadratic model used throughout the iterations.

- Formally the two methods can be stated by the following TREI framework.

TREI framework (Trust-Region method for systems of nonlinear Equalities and Inequalities)

- Given $x_k \in \Omega$, let m_k be a quadratic model for $\theta = \frac{\|\Theta\|_2^2}{2}$ at x_k

$$m_k(p) = \frac{1}{2}\|\Theta'_k p + \Theta_k\|_2^2 + \frac{1}{2}\mu_k\|p\|_2^2, \quad \mu_k \geq 0$$

If $\mu_k = 0$ for all $k \geq 0 \Rightarrow$ Gauss-Newton method (TREI-GN).

If $\mu_k > 0$ for all $k \geq 0 \Rightarrow$ Levenberg-Marquardt method (TREI-LM).

- The trust-region problem (TRP) has the form:

$$\min_p\{m_k(p) : \ \|p\|_2 \leq \Delta_k\}$$

- The generalized Cauchy step p_k^C along the scaled steepest descent direction $d_k = -D_k\nabla\theta_k$ of θ is defined as

$$p_k^C = \operatorname*{argmin}_{p \in span\{d_k\}} m_k(p) \quad \text{s.t.} \quad \|p\|_2 \leq \Delta_k, \ \ x_k + p \in \Omega.$$

kth iteration

- **Step 0 - Input** : given $x_k \in \Omega$, $\Delta_k > 0$, $\mu_k \geq 0$.
- **Step 1 - Test for optimality** : check if $D_k \nabla \theta_k = 0$.
- **Step 2 - Solve the TRP** : compute a solution p_{tr} to

$$\min_p \{ m_k(p) : \quad \|p\|_2 \leq \Delta_k \}.$$

- **Step 3 - Enforce feasibility** : let $\bar{p}_{tr}$ be s.t.

$$\bar{p}_{tr} = P_\Omega(x_k + p_{tr}) - x_k,$$

where $P_\Omega(x)$ is the projection of x onto Ω.

kth iteration

- **Step 4 - Test for global convergence**:
  If
  $$\rho_c(\bar{p}_{tr}) = \frac{m_k(0) - m_k(\bar{p}_{tr})}{m_k(0) - m_k(p_k^C)} \geq \beta_1, \quad \beta_1 \in (0,1),$$
  Set $p_k = \bar{p}_{tr}$;
  Else find $p_k = t\, p_k^C + (1-t)\bar{p}_{tr}$, $t \in (0,1]$, s.t. $\rho_c(p_k) = \beta_1$.

- **Step 5 - Test to accept the trial step**:
  If
  $$\rho_\theta(p_k) = \frac{\theta(x_k) - \theta(x_k + p_k)}{m_k(0) - m_k(p_k)} \geq \beta_2, \quad \beta_2 \in (0,1),$$
  then p_k is accepted;
  Otherwise, p_k is rejected and the trust-region size Δ_k is reduced.

The solution of the TRP

$$\min_p \{m_k(p) : \quad \|p\|_2 \leq \Delta_k\}$$

In the general case the unconstrained minimizer of m_k is non-unique.

- Due to the trust-region constraints, it would be convenient to use the minimum norm step.

  Compute an *approximate* solution of TRP by
  - the dogleg algorithm using the minimun norm step
  - the truncated CG method
    (appropriate for large and sparse problems).
- The Moré and Sorensen algorithm has a pitfall: in the hard case, e.g. if $m < n$, this strategy steps to the boundary of the trust-region even when an unconstrained minimizer of m_k is safely inside.

The minimum norm minimizer p_k^N of m_k

The quadratic model m_k for $\theta = \frac{\|\Theta\|_2^2}{2}$ at x_k is

$$m_k(p) = \frac{1}{2}\|\Theta'_k p + \Theta_k\|_2^2 + \frac{1}{2}\mu_k\|p\|_2^2, \quad \mu_k \geq 0.$$

- TREI-GN method ($\mu_k = 0$): p_k^N has the form

$$p_k^N = -{\Theta'_k}^+ \Theta_k.$$

- TREI-LM method ($\mu_k > 0$): p_k^N solves the positive definite system of equations

$$({\Theta'_k}^T \Theta'_k + \mu_k I_n)\, p_k^N = -{\Theta'_k}^T \Theta_k,$$

or equivalently

$$\begin{pmatrix} I_m & -\Theta'_k \\ -{\Theta'_k}^T & -\mu_k I_n \end{pmatrix} \begin{pmatrix} z \\ p_k^N \end{pmatrix} = \begin{pmatrix} \Theta_k \\ 0 \end{pmatrix}, \quad \text{where } z \in \mathbb{R}^m.$$

Convergence analysis

Assumptions

Let L be an open, bounded and convex set containing $\{x_k\}$.

- $\Theta'(x)$ bounded in norm and Lipschitz continuous for $x \in L$.

(for NEI we assume $C_E(x)$, $C_I(x)$, $C_E'(x)$, $C_I'(x)$ bounded in norm for $x \in L$, $C_E'(x)$ and $C_I'(x)$ Lipschitz continuous for $x \in L$.)

Global convergence results

Let $\{x_k\}$ be the sequence generated by either the TREI-GN method or the TREI-LM method.

- Every limit point of the sequence $\{x_k\}$ is a first-order stationary point for the problem $\min_{x\in\Omega} \theta(x) = \frac{1}{2}\|\Theta(x)\|_2^2$.
- If x^* is a limit point of $\{x_k\}$ and $\Theta(x^*) = 0$, then all the limit points of $\{x_k\}$ solve the BCNS problem.

Quadratic local convergence

Let x^* be a limit point of the sequence $\{x_k\}$ s.t. $\Theta(x^*) = 0$.

- The sequence generated by the TREI-GN method converges quadratically to x^* if

$$\Theta'(x^*) \quad \text{is full rank}$$

- The sequence generated by the TREI-LM method converges quadratically to x^* if

$$\mu_k = O(\|\Theta_k\|_2^2),$$

and

$$\|\Theta\|_2 \text{ provides a local error bound near } x^*.$$

Error bound condition

[Kanzow, Yamashita, Fukushima, JCAM 2004] There exist α, $\rho > 0$ s.t.

$$d(x,S) \le \alpha \|\Theta(x)\|_2 \quad \forall x \text{ s.t. } \|x - x^*\|_2 \le \rho$$

where

$$S = \{y \in \Omega : \Theta(y) = 0\}, \quad d(x,S) = \inf\{\|x - y\|_2, y \in S\}$$

- The error bound condition depends on x^*.
- If $m \ge n$ and $\Theta'(x^*)$ is full rank $\Rightarrow$ the error bound condition is guaranteed on some neighbourhood of x^*.

  The converse is not true: the local error bound is a weaker condition than the full rank condition of $\Theta'(x^*)^T \Theta'(x^*)$ ($m \ge n$ and $\Theta'(x^*)$ is full rank).

Test problems

- The TREI-GN and TREI-LM were tested on 32 problems from the *Hock-Schittkowsky* test collection and from *Handbook of test problems in local and global optimization* by Floudas et al., Kluwer 1999.

- The dimensions are such that $2 \leq m \leq 500$, $2 \leq n \leq 1000$.

- If the problems are NEI problems, the reformulation used is

$$\Theta(x) = \begin{pmatrix} C_E(x) \\ \dfrac{\max\{0, C_I(x)\}^2}{2} \end{pmatrix} \qquad l \leq x \leq u.$$

Implementation issues

- The TREI-GN and TREI-LM methods were run under `Matlab 7.0` with $\epsilon_m \simeq 2 \cdot 10^{-16}$.
- In the TREI-LM method we set

$$\mu_0 = 10^{-8}\|\Theta_0\|_2^2, \quad \mu_k = \min\{\mu_{k-1}, \|\Theta_k\|_2^2\}, \; k > 0$$

[Kanzow, Yamashita, Fukushima, JCAM 2004]

- The strategy for updating the trust-region radius is the standard one.
- Successful termination: $\|\Theta_k\|_2 \le 10^{-6}$.
- Failures:

$\Delta_k \le \sqrt{\epsilon_m}$;

300 nonlinear iterations were performed;

$\|D_k \nabla\theta_k\|_2 \le 100\,\epsilon_m$ or $\|\Theta_{k+1} - \Theta_k\|_2 \le 100\,\epsilon_m\|\Theta_k\|_2$.

Implementation issues

- The TREI-GN and TREI-LM methods were run under `Matlab 7.0` with $\epsilon_m \simeq 2 \cdot 10^{-16}$.
- In the TREI-LM method we set

$$\mu_0 = 10^{-8}\|\Theta_0\|_2^2, \quad \mu_k = \min\{\mu_{k-1}, \|\Theta_k\|_2^2\}, \; k > 0$$

  [Kanzow, Yamashita, Fukushima, JCAM 2004]
- The strategy for updating the trust-region radius is the standard one.
- Successful termination: $\|\Theta_k\|_2 \leq 10^{-6}$.
- Failures:

  $\Delta_k \leq \sqrt{\epsilon_m}$;

  300 nonlinear iterations were performed;

  $\|D_k \nabla \theta_k\|_2 \leq 100\, \epsilon_m$ or $\|\Theta_{k+1} - \Theta_k\|_2 \leq 100\, \epsilon_m \|\Theta_k\|_2$.

Implementation issues: the solution of TRP

The Dogleg algorithm

1. Compute the minimum norm minimizer p_k^N of m_k:
    - if $\mu_k = 0$ by the complete orthogonal factorization of Θ'_k;
    - if $\mu_k > 0$ by the QR decomposition to the least-squares pb

$$\min_p \; m_k(p) = \min_p \frac{1}{2} \left\| \begin{pmatrix} \Theta'_k \\ \sqrt{\mu_k}\, I_n \end{pmatrix} p + \begin{pmatrix} \Theta_k \\ 0 \end{pmatrix} \right\|^2 .$$

   If p_k^N lies inside the trust-region, stop.

2. Compute the Cauchy step

$$p_{k,c} = \mathop{\mathrm{argmin}}_{p \in span\{-\nabla\theta_k\}} m_k(p) \quad \text{s.t.} \quad \|p\|_2 \le \Delta_k .$$

3. Dogleg between $p_{k,c}$ and p_k^N.

Results

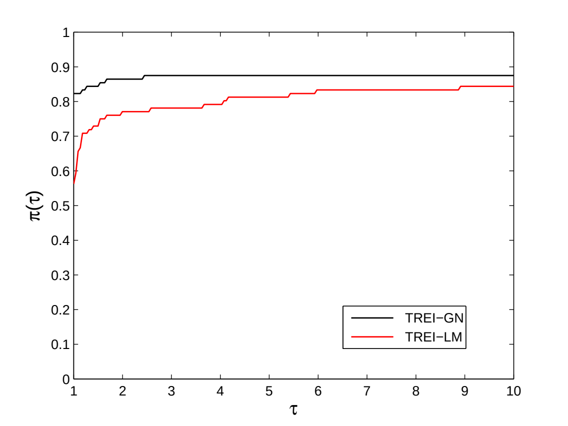
- All problems were run starting from three different initial guesses x_0:

$$(x_0)_i = \begin{cases} l_i + c(u_i - l_i)/2 & \text{if } l_i > -\infty \text{ and } u_i < \infty, \\ l_i + c\,10^c & \text{if } l_i > -\infty \text{ and } u_i = \infty, \\ u_i - c\,10^c & \text{if } l_i = -\infty \text{ and } u_i < \infty, \\ 10(c-1) & \text{if } l_i = -\infty \text{ and } u_i = \infty, \end{cases}$$

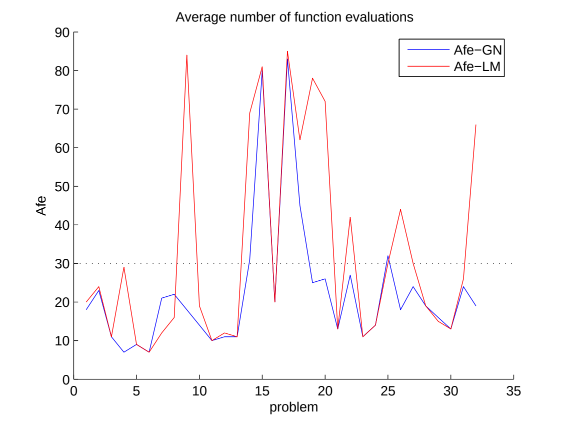
for $c = 0, 1, 2$. Then 96 runs were performed.

- The method TREI-GN and TREI-LM converged to a first-order stationary point for $\min_{x \in \Omega} \theta(x) = \frac{1}{2}\|\Theta(x)\|_2^2$ for about 91% of the runs.

Performance profile



We summarize the overall computational effort of the methods plotting the function Θ-evaluation performance profile, see [Dolan, Moré, Math. Program. 2002]



Most problems were solved with a low number of Θ-function evaluations and this number is, on average, favorable for the TREI-GN method.

Conclusions

- We have presented two trust-region methods for solving BCNS problems based on recent developments in nonlinear optimization with simple bounds.
- At each iteration:
  - The TREI-GN method requires the minimum norm solution of a linear least-squares problem.
    For large-scale problems the use of iterative methods can reduce the cost of this task.
  - The TREI-LM method requires the solution of a positive definite system and the choice of the parameter μ_k. This choice is critical for its numerical performances.

Future work

- Enhancement of the global strategy: use of the filter technique in our framework;
- TRP solution: use of the truncated CG algorithm instead of the dogleg method $\Rightarrow$ efficient solution of large problems;
- Development of software;
- Development of an efficient procedure in computing non-zero residual stationary point.

Thanks for your attention