Solving ill-posed nonlinear systems with noisy data: a regularizing trust-region approach

Elisa Riccietti
Università degli Studi di Firenze
Dipartimento di Matematica e Informatica 'Ulisse Dini'

Joint work with Stefania Bellavia, Benedetta Morini

UNIVERSITÀ
DEGLI STUDI
FIRENZE

Opening Meeting for the Research Project GNCS 2016 PING - Inverse Problems in Geophysics Florence, April 6, 2016.

Ill-posed problems

Let us consider the following inverse problem: given $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ with $m \geq n$, nonlinear, continuously differentiable and $y \in \mathbb{R}^m$, find $x \in \mathbb{R}^n$ such that

$$F(x) = y.$$

Definition

The problem is well-posed if:

1. $\forall y \in \mathbb{R}^m \ \exists x \in \mathbb{R}^n$ such that $F(x) = y$ (existence),
2. F is an injective function (uniqueness),
3. F^{-1} is a continuous function (stability).

The problem is ill-posed if one or more of the previous properties do not hold.

Ill-posed problems

- Let us consider problems of the form $F(x) = y$ for $x \in (\mathbb{R}^n, \|\cdot\|_2)$ and $y \in (\mathbb{R}^m, \|\cdot\|_2)$, arising from the discretization of a system modeling an ill-posed problem, such that:
  - it exists a solution $x^\dagger$, but is not unique,
  - stability does not hold.
- In a realistic situation the data y are affected by noise, we have at disposal only y^δ such that:

$$\|y - y^\delta\| \leq \delta$$

  for some positive δ .
- We can handle only a noisy problem:

$$F(x) = y^\delta.$$

Need for regularization

- As stability does not hold, the solutions of the original problem do not depend continuously on the data.
  $\Longrightarrow$ The solutions of the noisy problem may not be meaningful approximations of the original problem solutions.

Need for regularization

- As stability does not hold, the solutions of the original problem do not depend continuously on the data.
  $\Longrightarrow$ The solutions of the noisy problem may not be meaningful approximations of the original problem solutions.
- For ill-posed problems there are no finite bounds on the inverse of the Jacobian of F around a solution of the original problem.

Need for regularization

- As stability does not hold, the solutions of the original problem do not depend continuously on the data.
  $\Longrightarrow$ The solutions of the noisy problem may not be meaningful approximations of the original problem solutions.
- For ill-posed problems there are no finite bounds on the inverse of the Jacobian of F around a solution of the original problem.
- Classical methods used for well-posed systems are not suitable in this contest.

$$\Downarrow$$

Need for regularization.

Outline

- Introduction to iterative regularization methods.
- Description of Levenberg-Marquardt method and of its regularizing variant.
- Description of a new regularizing trust-region approach, obtained by a suitable choice of the trust region radius .
- Regularization and convergence properties of the new approach.
- Numerical tests: we compare the new trust-region approach to the regularizing Levenberg-Marquardt and standard trust-region methods.
- Open issues and future developments.

Iterative regularization methods

Hypothesis: it exists $x^\dagger$ solution of $F(x) = y$.

Iterative regularization methods generate a sequence $\{x_k^\delta\}$. If the process is stopped at iteration $k^*(\delta)$ the method is supposed to guarantee the following properties:

- $x_{k^*(\delta)}^\delta$ is an approximation of $x^\dagger$;
- $\{x_{k^*(\delta)}^\delta\}$ tends to $x^\dagger$ if δ tends to zero;
- local convergence to $x^\dagger$ in the noise-free case.

Existing methods

- Landweber (gradient-type method)[Hanke, Neubauer, Scherzer, 1995,Kaltenbacher, Neubauer, Scherzer, 2008]
- Truncated Newton - Conjugate Gradients [Hanke,1997, Rieder, 2005]
- Iterative Regularizing Gauss-Newton [Bakushinsky, 1992, Blaschke, Neubauer, Scherzer, 1997]
- Levenberg-Marquardt [Hanke,1997,2010,Vogel 1990, Kaltenbacher, Neubauer, Scherzer, 2008]

These methods are analyzed only under local assumptions, the definition of globally convergent approaches is still an open task.

Levenberg-Marquardt method

- Given $x_k^\delta \in \mathbb{R}^n$ and $\lambda_k > 0$, we denote with $J \in \mathbb{R}^{m\times n}$ the Jacobian matrix of F. The step $p_k \in \mathbb{R}^n$ is the minimizer of

$$m_k^{LM}(p) = \frac{1}{2}\|F(x_k^\delta) - y^\delta + J(x_k^\delta)p\|^2 + \frac{1}{2}\lambda_k\|p\|^2;$$

- p_k is the solution of

$$(B_k + \lambda_k I)p_k = -g_k$$

with $B_k = J(x_k^\delta)^T J(x_k^\delta)$, $g_k = J(x_k^\delta)^T(F(x_k^\delta) - y^\delta)$;

- The step is then used to compute the new iterate

$$x_{k+1}^\delta = x_k^\delta + p_k.$$

Regularizing Levenberg-Marquardt method

- The parameter $\lambda_k > 0$ is chosen as the solution of:

$$\|F(x_k^\delta) - y^\delta + J(x_k^\delta)p\| = q\|F(x_k^\delta) - y^\delta\|$$

  with $q \in (0, 1)$;
- With noisy data the process is stopped at iteration $k^*(\delta)$ such that $x^\delta_{k^*(\delta)}$ satisfies the discrepancy principle:

$$\|F(x^\delta_{k^*(\delta)}) - y^\delta\| \leq \tau\delta < \|F(x_k^\delta) - y^\delta\|$$

  for $0 \leq k < k^*(\delta)$ and $\tau > 1$ suitable parameter.

[Hanke, 1997,2010]

Local analysis

Hypothesis for the local analysis:
Given the starting guess x_0, it exist positive ρ and c such that

- the system $F(x) = y$ is solvable in $B_\rho(x_0)$;
- for $x, \tilde{x} \in B_{2\rho}(x_0)$

$$\|F(x) - F(\tilde{x}) - J(x)(x - \tilde{x})\| \leq c\|x - \tilde{x}\|\|F(x) - F(\tilde{x})\|.$$

[Hanke, 1997,2010]

Due to the ill-posedness of the problem it is not possible to assume that a finite bound on the inverse of the Jacobian matrix exists.

Regularizing properties of the Levenberg-Marquardt method

Choosing λ_k as the solution of

$$\|F(x_k^\delta) - y^\delta + J(x_k^\delta)p\| = q\|F(x_k^\delta) - y^\delta\|$$

and stopping the process when the discrepancy principle

$$\|F(x_{k^*(\delta)}^\delta) - y^\delta\| \leq \tau\delta < \|F(x_k^\delta) - y^\delta\|$$

is satisfied, Hanke proves that:

- With exact data ($\delta = 0$): local convergence to $x^\dagger$,
- With noisy data ($\delta > 0$): if $\tau > \frac{1}{q}$, choosing x_0 close to $x^\dagger$ the discrepancy principle is satisfied after a finite number of iterations $k^*(\delta)$ and $\{x_{k^*(\delta)}^\delta\}$ converges to a solution of $F(x) = y$ if δ tends to zero.

This is a regularizing method

Trust-region methods

- Given $x_k^\delta \in \mathbb{R}^n$, the step $p_k \in \mathbb{R}^n$ is the minimizer of

$$\min_p m_k^{TR}(p) = \frac{1}{2}\|F(x_k^\delta) - y^\delta + J(x_k^\delta)p\|^2,$$
$$\text{s.t. } \|p\| \leq \Delta_k,$$

  with $\Delta_k > 0$ trust-region radius.
- Set $\Phi(x) = \frac{1}{2}\|F(x) - y^\delta\|^2$, and compute

$$\pi_k(p_k) = \frac{\Phi(x_k) - \Phi(x_k + p_k)}{m_k^{TR}(0) - m_k^{TR}(p_k)}.$$

- Given $\eta \in (0, 1)$:
  - If $\pi_k < \eta$ then set $\Delta_{k+1} < \Delta_k$ and $x_{k+1} = x_k$.
  - If $\pi_k \geq \eta$ then set $\Delta_{k+1} \geq \Delta_k$ and $x_{k+1} = x_k + p_k$.

Trust-region methods

It is possible to prove that p_k solves

$$(B_k + \lambda_k I)p_k = -g_k$$

for some $\lambda_k \geq 0$ such that

$$\lambda_k(\|p_k\| - \Delta_k) = 0,$$

where we have set $B_k = J(x_k^\delta)^T J(x_k^\delta)$ and $g_k = J(x_k^\delta)^T (F(x_k^\delta) - y^\delta)$.

Trust-region methods

From $\lambda_k(\|p_k\| - \Delta_k) = 0$ it follows that:

- If the minimum norm solution p^* of $B_k p = -g_k$ satisfies $\|p^*\| \leq \Delta_k$ then $\lambda_k = 0$ and $p_k = p(0)$;
- otherwise $\lambda_k \neq 0$, $\|p_k\| = \Delta_k$ and $p_k = p(\lambda_k)$ is a Levenberg-Marquardt step.

$$\Downarrow$$

- The standard trust-region does not ensure regularizing properties.
- Trust-region should be active to have a regularizing method:

$$\|p_k\| = \Delta_k.$$

Regularizing trust-region

- Levenberg-Marquardt and trust-region methods are strictly connected, due to the form of the step.
- As Hanke did, can we introduce a trust-region method with regularizing properties and still globally convergent?

Goals

We modify the standard trust-region to have:

- monotone decay of the function

$$\Phi(x) = \frac{1}{2}\|F(x) - y^\delta\|^2,$$

- the q-condition to hold:

$$\|F(x_k^\delta) - y^\delta + J(x_k^\delta)p\| \geq q\|F(x_k^\delta) - y^\delta\|.$$

The q-condition is a relaxed reformulation of

$$\|F(x_k^\delta) - y^\delta + J(x_k^\delta)p\| = q\|F(x_k^\delta) - y^\delta\|.$$

Regularizing trust-region

We now describe the new trust-region approach that thanks to a suitable trust-region radius update ensures:

- the q-condition to hold,
- the same regularizing properties of Levenberg-Marquardt method.

Trust-region radius choice

Lemma

Let p_k the solution of trust-region problem. If

$$\Delta_k \leq \frac{1-q}{\|B_k\|}\|g_k\|$$

then p_k satisfies the q-condition.

Consequence: Δ_k's choice

$$\Delta_k \in \left[C_{\min}\|g_k\|, \min\left\{ C_{\max}, \frac{1-q}{\|B_k\|}\|g_k\| \right\} \right],$$

with $C_{\min}, C_{\max}$ suitable constant, $B_k = J(x_k^\delta)^T J(x_k^\delta)$ e $g_k = J(x_k^\delta)^T(F(x_k^\delta) - y^\delta)$.

Algorithm : k**-th iteration of regularizing trust-region**

Given x_k^δ, $\eta \in (0,1)$, $\gamma \in (0,1)$, $0 < C_{\min} < C_{\max}$.

Exact data: y, $q \in (0,1)$.

Noisy data: y^δ, $q \in (0,1)$, $\tau > 1/q$.

1. Compute $B_k = J(x_k^\delta)^T J(x_k^\delta)$ and $g_k = J(x_k^\delta)^T (F(x_k^\delta) - y^\delta)$.
2. Choose $\Delta_k \in \left[C_{\min}\|g_k\|, \min\left\{ C_{\max}, \frac{1-q}{\|B_k\|} \right\} \|g_k\| \right]$
3. Repeat
   3.1 Compute the solution p_k of trust-region problem.
   3.2 Compute
   $$\pi_k(p_k) = \frac{\Phi(x_k^\delta) - \Phi(x_k^\delta + p_k)}{m_k^{TR}(0) - m_k^{TR}(p_k)}$$
   with $\Phi(x) = \frac{1}{2}\|F(x) - y^\delta\|^2$, $m_k^{TR}(p) = \frac{1}{2}\|F(x_k^\delta) + J(x_k^\delta)p\|^2$.
   3.3 If $\pi_k(p_k) < \eta$,set $\Delta_k = \gamma\Delta_k$.

   Until $\pi_k(p_k) \geq \eta$.
4. Set $x_{k+1}^\delta = x_k^\delta + p_k$.

Local analysis

Hypothesis 1: the same as for Levenberg-Marquardt method.
We assume that for index $\bar{k}$ it exist positive ρ and c such that

1. the system $F(x) = y$ is solvable in $B_\rho(x^\delta_{\bar{k}})$;
2. for $x, \tilde{x} \in B_{2\rho}(x^\delta_{\bar{k}})$

$$\|F(x) - F(\tilde{x}) - J(x)(x - \tilde{x})\| \leq c\|x - \tilde{x}\|\|F(x) - F(\tilde{x})\|.$$

Hypothesis 2: It exists positive K_J such that

$$\|J(x)\| \leq K_J$$

for all $x \in \mathcal{L} = \{x \in \mathbb{R}^n \ \ s.t. \ \ \Phi(x) \leq \Phi(x_0)\}$.

Results for $\delta = 0$

Lemma

The method generates a sequence $\{x_k\}$ such that for $k \geq \bar{k}$

- *trust-region is active, i.e. $\lambda_k > 0$;*
- *x_k belongs to $B_{2\rho}(x_{\bar{k}})$ and to $B_\rho(x^\dagger)$;*
- $\|x_{k+1} - x^\dagger\| < \|x_k - x^\dagger\|$;
- *it exists $\bar{\lambda} > 0$ such that $\lambda_k \leq \bar{\lambda}$.*

Theorem

The sequence $\{x_k\}$ converges to a solution x^* of $F(x) = y$ such that $\|x^* - x^\dagger\| \leq \rho$.

It holds $\lim_{k\to\infty} \|g_k\| = 0$ so the trust-region radius tends to zero.

Results for $\delta > 0$

Lemma

Let $\bar{k} < k^*(\delta)$. The method generates a sequence $\{x_k^\delta\}$ such that for $\bar{k} \leq k < k^*(\delta)$

- *the trust-region is active, i.e. $\lambda_k > 0$;*
- *x_k^δ belongs to $B_{2\rho}(x_k^\delta)$ and to $B_\rho(x^\dagger)$;*
- *$\|x_{k+1}^\delta - x^\dagger\| < \|x_k^\delta - x^\dagger\|$;*
- *it exists $\bar{\lambda} > 0$ such that $\lambda_k \leq \bar{\lambda}$.*

Theorem

The discrepancy principle is satisfied after a finite number of iterations $k^*(\delta)$ and the sequence $\{x_{k^*(\delta)}^\delta\}$ converges to a solution of $F(x) = y$ if δ tends to zero.

This is a regularizing method.

Test problems

- Four nonlinear ill-posed systems arising from the discretization of nonlinear first-kind Fredholm integral equation are considered, they model gravimetric and geophysics problems:

$$\int_0^1 k(t,s,x(s))ds = y(t), \qquad t \in [0,1],$$

  P1,**P2**, [Vogel, 1990], **P3**,**P4** [Kaltenbacher,2007];
- Their kernel is of the form

$$\begin{aligned} k(t,s,x(s)) &= log\left(\frac{(t-s)^2+H^2}{(t-s)^2+(H-x(s))^2}\right); \\ k(t,s,x(s)) &= \frac{1}{\sqrt{1+(t-s)^2+x(s)^2}}; \end{aligned}$$

Test problems: discretization

- We chose $n = m$, interval $[0,1]$ was discretized using n=64 equidistant grid points $t_i = (i-1)h,\ h = 1/(n-1),\ i = 1, \dots, n$;
- $x(s)$ was approximated by piecewise linear functions $\Phi_j(s)$ on the grid $s_j = t_j,\ j = 1, \dots, n$; $x(s) \sim \hat{x}_n(s) = \sum_{j=1}^{n} \Phi_j(s) x_j$

Test problems: discretization

- The integrals $\int_0^1 k(t_i, s, \hat{x}(s))ds$, $i = 1, \dots, n$ were approximated by the composite trapezoidal rule on the points s_j $j = 1, \dots, n$.
- The resulting nonlinear system is

$$\sum_{i=1}^{n} w_j k(t_i, s_j, \hat{x}(s_j)) = y(t_i) \qquad j = 1, \dots, n.$$

with $w_1 = w_n = \frac{1}{2}$, $w_i = 1$ for all $i \neq 1, n$.

Choice of parameters λ_k

- Parameters λ_k were computed to have an active trust-region:

$$\|p(\lambda)\| = \Delta_k.$$

- We used Newton method to solve this reformulation of the condition:

$$\psi(\lambda) = \frac{1}{\|p(\lambda)\|} - \frac{1}{\Delta_k} = 0.$$

  that is more suitable to the application of Newton method.
- Each Newton iteration requires Cholesky factorization of $B_k + \lambda_k I$.

Regularizing trust-region implementation

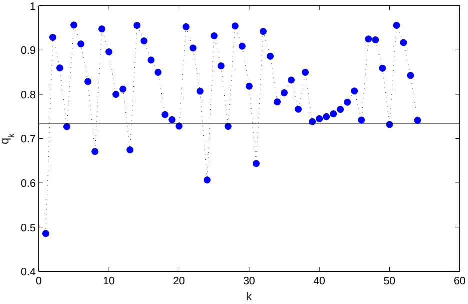
Trust-region radius update:

$$\Delta_k = \mu_k \|F(x_k^\delta) - y^\delta\|, \qquad \mu_k = \begin{cases} \frac{1}{6}\mu_{k-1} & \text{if } q_{k-1} < q \\ 2\mu_{k-1} & \text{if } q_{k-1} > \nu q \\ \mu_{k-1} & \text{otherwise} \end{cases}$$

with $q_k = \frac{\|F(x_k^\delta) - y^\delta + J(x_k^\delta)p_k\|}{\|F(x_k^\delta) - y^\delta\|}$, and $\nu = 1.1$.

- Δ_k is less expensive to compute if compared to $\frac{1-q}{\|B_k\|}\|g_k\|$ but preserves convergence to zero if $\delta = 0$.
- In the update the fulfillment of q-condition is considered.

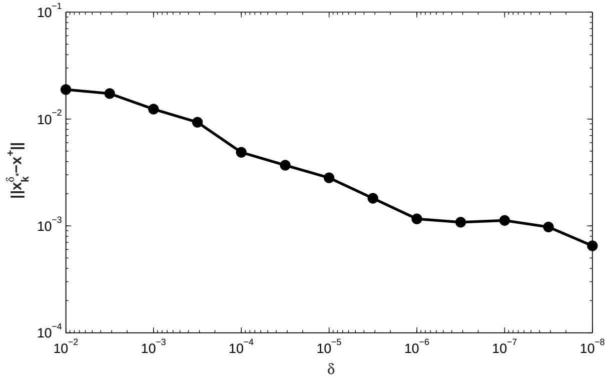
Regularizing properties



- • = Values of $q_k = \frac{\|F(x_k^\delta)-y^\delta+J(x_k^\delta)p_k\|}{\|F(x_k^\delta)-y^\delta\|}$, solid line: $q = 1.1/\tau$.

The q-condition is satisfied in most of the iterations even if not esplicitly imposed.

Regularizing properties of the method.



Logarithmic plot of the error $||x^{\delta}_{k^*(\delta)} - x^{\dagger}||$ as a function of the noise level δ.

Comparison between regularizing TR-LM, $\delta = 10^{-2}$.

Problem	x_0	Regularizing TR it	nf	cf	Regularizing LM it	nf	cf
P1	$0\ e$	20	21	6	17	18	4
	$-0.5\ e$	29	30	6	22	23	4
	$-1\ e$	35	36	5	24	25	4
	$-2\ e$	40	41	5	25	26	4
P2	$0\ e$	30	31	5	*	*	*
	$0.5\ e$	25	26	5	*	*	*
	$1\ e$	29	30	5	22	23	5
	$2\ e$	37	39	5	25	26	5
P3	$x_0(1.25)$	15	16	4	12	13	4
	$x_0(1.5)$	17	18	4	14	15	4
	$x_0(1.75)$	19	20	4	15	16	4
	$x_0(2)$	22	23	4	16	17	4
P4	$x_0(1,1)$	17	18	5	10	11	4
	$x_0(0.5,0)$	20	21	4	*	*	*
	$x_0(1.5,1)$	22	23	4	15	16	4
	$x_0(1.5,0)$	26	27	4	*	*	*

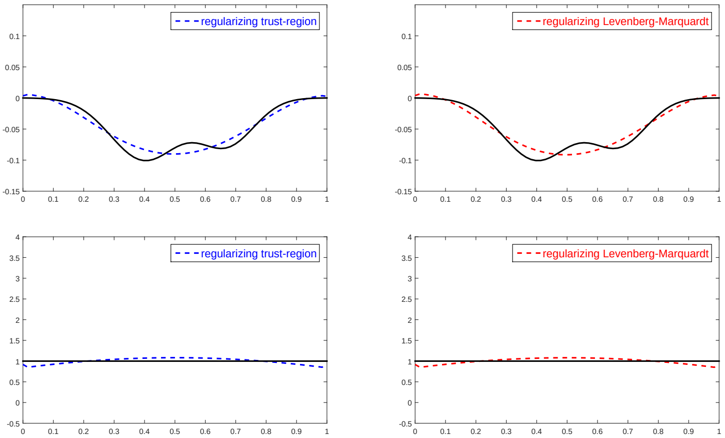
it=iterations, **nf**=function evaluations,

cf=mean number of Cholesky factorizations.

$*$=failure, reached maximum number of iterations or convergence to a solution of the noisy problem

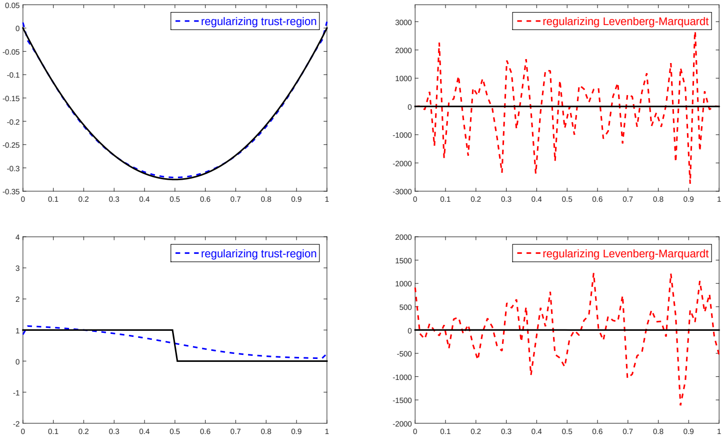
$e = (1,\dots,1)^T$, **P3:** $(x_0(\alpha))_j = (-4\alpha+4)s_j^2 + (4\alpha-4)s_j + 1$, **P4:** $x_0(\beta,\chi) = \beta - \chi s_j$, s_j grid points, $j = 1,\dots,n$.

Comparison between regularizing TR and LM



Left: regularizing TR, **Right**: regularizing LM , **Solid line**: solution of the original problem.

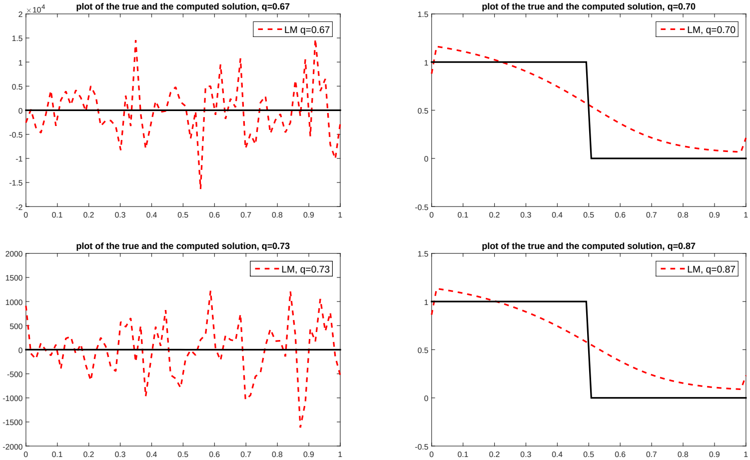
Comparison between regularizing TR e LM



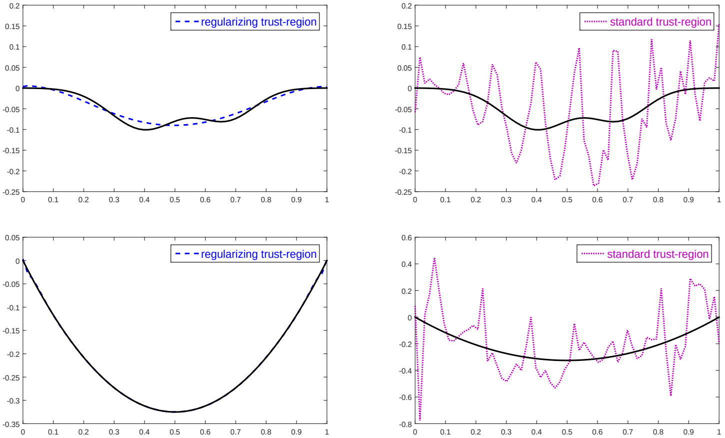
Left: regularizing TR , **Right**: regularizing LM , **Solid line**: solution of the original problem.

The q-condition

The condition imposed by Hanke is strongly dependent on the choice of the value of free parameter q. Values of $q = 0.67,\ 0.70,\ 0.73,\ 0.87$.



Comparison between regularizing and standard trust-region



Left: regularizing TR, **Right**: standard TR , **Solid line**: solution of the original problem.

Future developments: nonlinear least squares problems

- Consider the following least squares problem: given $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ with $m \geq n$, nonlinear, continuously differentiable and $y \in \mathbb{R}^m$, solve

$$\min_x f(x) = \frac{1}{2}\|F(x) - y\|^2.$$

- Non-zero residual problem: let x^* be a solution of the problem and assume that $\|F(x^*) - y\| > 0$.
- **Newton**: given the current iterate x_k, at each iteration the step p_k is computed as:

$$H(x_k)p_k = -J(x_k)^T(F(x_k) - y)$$

where

$$H(x_k) = J(x_k)^T J(x_k) + S(x_k),$$
$$S(x_k) = \sum_{i=1}^{m}(F_i(x) - y_i)\nabla^2 F_i(x).$$

Future developments: nonlinear least squares problems

- **Gauss Newton**: given the current iterate x_k, at each iteration the step p_k is computed as:

$$J(x_k)^T J(x_k) p_k = -J(x_k)^T (F(x_k) - y).$$

- The Gauss Newton method converges if $\|S(x^*)\| < \lambda_*$ with λ_* the smallest eigenvalue of $J(x^*)^T J(x^*)$.
- This hypothesis is rather restrictive when dealing with ill-posed problems (the Jacobian matrix should be invertible).
- We want to design a trust region approach to solve least squares problems, converging under less restrictive hypotheses on λ_* compared to the Gauss- Newton method.

Future developments: nonlinear least squares problems

- We want our trust-region method to be able to deal also with ill-posed noisy least-squares problems:

$$\min_x f(x) = \frac{1}{2}\|F(x) - y\|^2,$$

  and only noisy data y^δ are at disposal: $\|y - y^\delta\| \leq \delta$.
- Non-zero residual problem: if x^* is a solution of the problem we assume that $\|F(x^*) - y\| > 0$.

THANK YOU FOR YOUR ATTENTION!

Open issues: Convergence to the infinite dimensional solution.

Let $\mathcal{X}, \mathcal{Y}$ be Hilbert spaces, $F_\infty : \mathcal{X} \rightarrow \mathcal{Y}$, $y_\infty \in \mathcal{Y}$. The nonlinear system is the discretization of a infinite dimensional problem: find $x_\infty \in \mathcal{X}$ such that $F_\infty(x_\infty) = y_\infty$. We are interested in the convergence of the discrete solution $\hat{x}_n(s) = \sum_{j=1}^{n} \Phi_j(s) x_j$ to a solution of the infinite dimensional problem as $n \rightarrow \infty$.

Theorem

The sequence $\{\hat{x}_n\}$ has a weakly convergent subsequence $\{\hat{x}_k\}$.

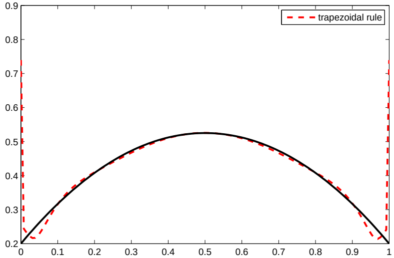
Theorem

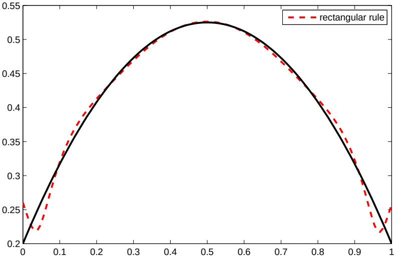
The sequence $\{\|F_\infty(\hat{x}_k) - y_\infty\|\}$ converges to zero as k tends to infinite, i.e. the weak limit x^* of sequence $\{\hat{x}_k\}$ is a solution of the original problem, $F_\infty(x^*) = y_\infty$.

Open issues: peaks

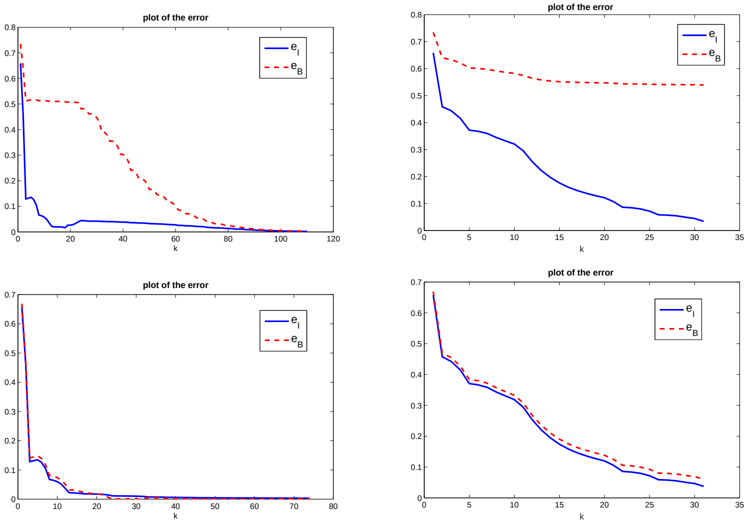
- Problem: when solving the nonlinear system obtained computing the integral by the trapezoidal rule, the approximated solution shows peaks at the end points of the interval. Peaks are higher and higher as the starting guess moves away from the solution and the noise increases.
- When solving the nonlinear system obtained computing the integral by the rectangular rule, the approximated solution does not show peaks at the end points of the interval.

Computed solution





Computed solution, $x_0 = 1e$, $\delta = 1.e - 2$. **Left**: trapezoidal rule, **Right**: rectangular rule, **Solid line**: solution of the original problem.



e_I= error computed on the points inside the interval, e_B=border error. **Upper part**: trapezoidal rule, *left*: $\delta = 0$, *right*: $\delta = 1.e - 2$. **Lower part**: rectangular rule, *left*: $\delta = 0$, *right*: $\delta = 1.e - 2$.

Comparison of the nonlinear systems

- Trapezoidal rule: the resulting nonlinear system is

$$\frac{1}{2}k(t_i, s_1, x_1) + 1k(t_i, s_2, x_2) + \cdots + 1k(t_i, s_{n-1}, x_{n-1}) + \frac{1}{2}k(t_i, s_n, x_n) = y(t_i),$$

$i = 1, \dots, n.$

- Rectangular rule: the resulting nonlinear system is

$$1k(t_i, s_1, x_1) + 1k(t_i, s_2, x_2) + \cdots + 1k(t_i, s_{n-1}, x_{n-1}) + 1k(t_i, s_n, x_n) = y(t_i),$$

$i = 1, \dots, n.$

Linear system: trapezoidal rule

We solve $(J^T J + \lambda I)p(\lambda) = -J^T(F - y^\delta)$. Let $n = 5$.

$$J = \begin{pmatrix} \frac{1}{2}\partial_1 k(t_1, s_1, x_1) & 1\partial_2 k(t_1, s_2, x_2) & . & 1\partial_4 k(t_1, s_4, x_4) & \frac{1}{2}\partial_5 k(t_1, s_5, x_5) \\ . & . & . & . & . \\ . & . & . & . & \\ . & . & . & . & . \\ \frac{1}{2}\partial_1 k(t_5, s_1, x_1) & 1\partial_2 k(t_5, s_2, x_2) & . & 1\partial_4 k(t_5, s_4, x_4) & \frac{1}{2}\partial_5 k(t_5, s_5, x_5) \end{pmatrix}$$

We denote $k_{i,j} = k(t_i, s_j, x_j)$ $i, j = 1, \dots, n$.

$$J^T J =$$

$$\begin{pmatrix} \frac{1}{4}\sum_{i=1}^5 \partial_1 k_{i,1}\partial_1 k_{i,1} & \frac{1}{2}\sum_{i=1}^5 \partial_1 k_{i,1}\partial_2 k_{i,2} & . & \frac{1}{2}\sum_{i=1}^5 \partial_1 k_{i,1}\partial_4 k_{i,4} & \frac{1}{4}\sum_{i=1}^5 \partial_1 k_{i,1}\partial_5 k_{i,5} \\ \frac{1}{2}\sum_{i=1}^5 \partial_2 k_{i,2}\partial_1 k_{i,1} & 1\sum_{i=1}^5 \partial_2 k_{i,2}\partial_2 k_{i,2} & . & 1\sum_{i=1}^5 \partial_2 k_{i,2}\partial_4 k_{i,4} & \frac{1}{2}\sum_{i=1}^5 \partial_2 k_{i,2}\partial_5 k_{i,5} \\ . & . & . & . & . \\ \frac{1}{2}\sum_{i=1}^5 \partial_4 k_{i,5}\partial_1 k_{i,1} & 1\sum_{i=1}^5 \partial_4 k_{i,5}\partial_2 k_{1,2} & . & 1\sum_{i=1}^5 \partial_4 k_{i,5}\partial_4 k_{i,4} & \frac{1}{2}\sum_{i=1}^5 \partial_4 k_{i,5}\partial_5 k_{i,5} \\ \frac{1}{4}\sum_{i=1}^5 \partial_4 k_{i,5}\partial_1 k_{i,1} & \frac{1}{2}\sum_{i=1}^5 \partial_5 k_{i,5}\partial_2 k_{1,2} & . & \frac{1}{2}\sum_{i=1}^5 \partial_5 k_{i,5}\partial_4 k_{i,4} & \frac{1}{4}\sum_{i=1}^5 \partial_5 k_{i,5}\partial_5 k_{i,5} \end{pmatrix}.$$

Linear system: rectangular rule

We solve $(J^TJ+\lambda I)p(\lambda) = -J^T(F-y^\delta)$. Let $n=5$.

$$J = \begin{pmatrix} 1\partial_1 k(t_1,s_1,x_1) & 1\partial_2 k(t_1,s_2,x_2) & . & 1\partial_4 k(t_1,s_4,x_4) & 1\partial_5 k(t_1,s_5,x_5) \\ . & . & . & . & . \\ . & . & . & . & \\ . & . & . & . & . \\ 1\partial_1 k(t_5,s_1,x_1) & 1\partial_2 k(t_5,s_2,x_2) & . & 1\partial_4 k(t_5,s_4,x_4) & 1\partial_5 k(t_5,s_5,x_5) \end{pmatrix}$$

We denote $k_{i,j} = k(t_i,s_j,x_j)$ $i,j = 1,\dots,n$.

$$J^TJ =$$

$$\begin{pmatrix} 1\sum_{i=1}^5 \partial_1 k_{i,1}\partial_1 k_{i,1} & 1\sum_{i=1}^5 \partial_1 k_{i,1}\partial_2 k_{i,2} & . & 1\sum_{i=1}^5 \partial_1 k_{i,1}\partial_4 k_{i,4} & 1\sum_{i=1}^5 \partial_1 k_{i,1}\partial_5 k_{i,5} \\ 1\sum_{i=1}^5 \partial_2 k_{i,2}\partial_1 k_{i,1} & 1\sum_{i=1}^5 \partial_2 k_{i,2}\partial_2 k_{i,2} & . & 1\sum_{i=1}^5 \partial_2 k_{i,2}\partial_4 k_{i,4} & 1\sum_{i=1}^5 \partial_2 k_{i,2}\partial_5 k_{i,5} \\ . & . & . & . & . \\ 1\sum_{i=1}^5 \partial_4 k_{i,5}\partial_1 k_{i,1} & 1\sum_{i=1}^5 \partial_4 k_{i,5}\partial_2 k_{1,2} & . & 1\sum_{i=1}^5 \partial_4 k_{i,5}\partial_4 k_{i,4} & 1\sum_{i=1}^5 \partial_4 k_{i,5}\partial_5 k_{i,5} \\ 1\sum_{i=1}^5 \partial_4 k_{i,5}\partial_1 k_{i,1} & 1\sum_{i=1}^5 \partial_5 k_{i,5}\partial_2 k_{1,2} & . & 1\sum_{i=1}^5 \partial_5 k_{i,5}\partial_4 k_{i,4} & 1\sum_{i=1}^5 \partial_5 k_{i,5}\partial_5 k_{i,5} \end{pmatrix}.$$

SVD decomposition: trapezoidal rule

Let consider matrix J^TJ SVD decomposition.

- $J^TJ = U\Sigma U^T$
- $cond(J^TJ) = 10^6$, $\lambda = 15.7$, $cond(J^TJ + \lambda I) = 1.2\ 10^0$
- $\sigma = diag(\Sigma) = \begin{pmatrix} 3.8\ 10^0 \\ 8.5\ 10^{-2} \\ 2.3\ 10^{-3} \\ 7.1\ 10^{-5} \\ 1.6\ 10^{-6} \end{pmatrix}, \qquad p = \begin{pmatrix} -7.6\ 10^{-2} \\ -1.7\ 10^{-1} \\ -1.8\ 10^{-1} \\ -1.7\ 10^{-1} \\ -7.6\ 10^{-2} \end{pmatrix}$
- $U = \begin{pmatrix} -0.24 & -0.44 & 0.58 & 0.56 & 0.32 \\ -0.54 & -0.56 & 0.04 & -0.44 & -0.46 \\ -0.56 & 3.5\ 10^{-8} & -0.56 & -7.3\ 10^{-8} & 0.61 \\ -0.54 & 0.56 & 0.04 & 0.44 & -0.46 \\ -0.24 & 0.44 & 0.58 & -0.56 & 0.32 \end{pmatrix}$

SVD decomposition: rectangular rule

Let consider matrix J^TJ SVD decomposition.

- $J^TJ = U\Sigma U^T$
- $cond(J^TJ) = 10^6$, $\lambda = 17.4$, $cond(J^TJ + \lambda I) = 1.3\ 10^0$
- $\sigma = diag(\Sigma) = \begin{pmatrix} 5.1\ 10^0 \\ 1.8\ 10^{-1} \\ 5.8\ 10^{-3} \\ 1.3\ 10^{-4} \\ 1.8\ 10^{-6} \end{pmatrix}, \qquad p = \begin{pmatrix} -1.8\ 10^{-1} \\ -2.0\ 10^{-1} \\ -2.1\ 10^{-1} \\ -2.0\ 10^{-1} \\ -1.8\ 10^{-1} \end{pmatrix}$
- $U = \begin{pmatrix} -0.41 & -0.60 & 0.55 & -0.38 & -0.17 \\ -0.46 & -0.38 & -0.19 & 0.60 & 0.5 \\ -0.48 & -4.1\ 10^{-8} & -0.57 & -1.4\ 10^{-6} & -0.66 \\ -0.46 & 0.38 & -0.19 & -0.60 & 0.50 \\ -0.41 & 0.60 & 0.55 & 0.38 & -0.17 \end{pmatrix}$