Levenberg-Marquardt and other regularisations for ill-posed nonlinear systems

Stefania Bellavia

Dipartimento di Ingegneria Industriale
Università degli Studi di Firenze

Joint work with

Benedetta Morini,
Università degli Studi di Firenze

Recent Advances on Optimization, July 24-26 2013, CERFACS-Toulose

Outline

The problem

We want to solve

$$F(x) = y$$

where $F : \Re^n \rightarrow \Re^n$ is a given vector-valued continuously differentiable function. We consider ill-posed problems:

- no finite bounds on the norm of the inverse of $F'(x)$ can be used in the analysis;
- the solution does not depend continuously on the data.

Noisy case

In realistic situation only noisy data y^δ are given:

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

where δ is the noise level.

Assume that for the exact data y a solution x^+ exists.

Need for regularization

Regularizing methods

Iterative processes that employ a regularization term and generate a sequence $\{x_k^\delta\}$.

For noisy data: the stopping criterion and tolerance proportional to the noise level δ. Assume that iterations are stopped at index k^*.

1. $x_{k^*}^\delta$ is an approximation to x^+.
2. $x_{k^*}^\delta \to x^+$ as δ goes to zero.
3. Noise-free case: Local convergence to x^+.

All the above results must hold without assuming boundness of the inverse of F'.

Regularizing methods

Iterative processes that employ a regularization term and generate a sequence $\{x_k^\delta\}$.

For noisy data: the stopping criterion and tolerance proportional to the noise level δ. Assume that iterations are stopped at index k^*.

1. $x_{k^*}^\delta$ is an approximation to x^+.
2. $x_{k^*}^\delta \to x^+$ as δ goes to zero.
3. Noise-free case: Local convergence to x^+.

All the above results must hold without assuming boundness of the inverse of F'.

Existing approaches

- Landweber iterations (gradient-type methods)[Hanke, Neubauer, Scherzer, 1995,Kaltenbacher, Neubauer, Scherzer, 2008]
- Truncated Newton-CG [Hanke,1997, Rieder, 2005]
- Levenberg-Marquardt scheme [Hanke,1997,2010,Kaltenbacher, Neubauer, Scherzer, 2008]
- Iteratively Regularized Gauss-Newton methods [Bakushinsky, 1992, Blaschke, Neubauer, Scherzer, 1997]

The analysis of these approaches is exclusively local (even in the noise free-case). The definition of global methods is an open task. We are aware only of [Kaltenbacher, 2006, Wang, Yuan, 2002, 2005]

Existing approaches

- Landweber iterations (gradient-type methods)[Hanke, Neubauer, Scherzer, 1995,Kaltenbacher, Neubauer, Scherzer, 2008]
- Truncated Newton-CG [Hanke,1997, Rieder, 2005]
- Levenberg-Marquardt scheme [Hanke,1997,2010,Kaltenbacher, Neubauer, Scherzer, 2008]
- Iteratively Regularized Gauss-Newton methods [Bakushinsky, 1992, Blaschke, Neubauer, Scherzer, 1997]

The analysis of these approaches is exclusively local (even in the noise free-case). The definition of global methods is an open task. We are aware only of [Kaltenbacher, 2006, Wang, Yuan, 2002, 2005]

Levenberg-Marquardt (LM) Approaches

Given x_k^δ, the step $p_k = p(\lambda_k)$ used to form the new iterate solves:

$$(F'(x_k^\delta)^T F'(x_k^\delta) + \lambda_k I)p = -F'(x_k^\delta)^T (F(x_k^\delta) - y^\delta)$$

for a specific $\lambda_k \geq 0$.

Properties of the LM step

- Let ℓ_k be the rank of $F'(x_k^\delta)$
- Let $U_k \Sigma_k V_k^T$ be its s.v.d. with singular values:

$$(\varsigma_k)_1 \geq (\varsigma_k)_2 \geq \ldots \geq (\varsigma_k)_{\ell_k} > (\varsigma_k)_{\ell_k+1} = \ldots = (\varsigma_k)_n = 0.$$

- Let $r = U_k^T(F(x_k^\delta) - y^\delta)$. Then,

$$p_k = p(\lambda_k) = \sum_{i=1}^{\ell_k} \left(\frac{(\varsigma_k)_i r_i}{(\varsigma_k)_i^2 + \lambda_k} \right) (V_k)_i \quad \Rightarrow \quad p_k \in \mathcal{R}(F'(x_k^\delta)^T)$$

Levenberg-Marquardt approaches c.ed

LM approaches include:

- Trust-region:
  - p_k solution to:

$$\min_{\|p\|\leq\Delta_k} m_k^{TR}(p) = \|F(x_k) - y^\delta + F'(x_k)p\|^2.$$

$p_k = p(\lambda_k)$. When $\lambda_k > 0$, $\|p(\lambda_k)\| = \Delta_k$.

- Adaptive Regularized methods: p_k obtained minimizing quadratic/cubic regularization of the Gauss-Newton model.

Adaptive Quadratic Regularized Method (ARQ):

- Assume $F'(x)$ is Lipschitz continuous.
  - p_k is the minimizer of

$$m_k^Q(p) = \sqrt{\|F'(x_k)p + F(x_k) - y\|^2 + \mu_k\|p\|^2} + \frac{1}{2}\sigma_k\|p\|^2$$

    with $\sigma_k > 0$, $\mu_k \geq 0$.
  - $p_k = p(\lambda_k)$ is a LM step with

$$\lambda_k = \mu_k + 2\sigma_k\sqrt{\|F'(x_k)p_k + F(x_k) - y\|^2 + \mu_k\|p_k\|^2}$$

  - m_k^Q is a model for $\|F(x) - y\|$.

[Nesterov, 2007], [B.-Cartis-Gould-Morini-Toint, 2010].

Adaptive Cubic Regularized Method (ARC):

- Assume the Hessian of $\|F\|^2$ is Lipschitz continuous.
  - p_k is the minimizer of

$$m_k^C(p) = \|F'(x_k)p + F(x_k) - y\|^2 + \frac{1}{3}\sigma_k\|p\|^3,$$

    with $\sigma_k > 0$.
  - $p_k = p(\lambda_k)$ is a LM step with

$$\lambda_k = \sigma_k\|p(\lambda_k)\|$$

- m_k^C is a model for $\|F(x) - y\|^2$.

[Griewank, 1981], [Nesterov-Polyak, 2006], [Cartis-Gould-Toint, 2011], [Gould, Porcelli, Toint, 2012].

- The parameter λ_k is adaptively chosen and it is the result of the minimization process.
- The trust-region radius Δ_k and the regularization term σ_k are adaptatively chosen in order to satisfy the classical decrease condition.
- Given a model $m_k(p)$ for a merit function $f_m(x)$:

$$\rho = \frac{ared}{pred} \geq \eta_1 \qquad \eta_1 \in (0,1)$$

where

$$ared = f_m(x_k^\delta) - f_m(x_k^\delta + p_k) \quad pred = f_m(x_k^\delta) - m_k(p_k)$$

Globally convergent Levenberg-Marquardt approaches

Given the model m_k (m_k^{TR}, m_k^Q, m_k^C), the corresponding regularization term β_k ($1/\Delta_k$, σ_k) and $0 < \eta_1 < \eta_2 < 1$, $\gamma_1 > 1$:

Step 1: Compute p_k the minimizer of the model $m_k(p)$(possibly constrained in case of TR). The minimizer p_k is a LM step.

Step 2: Step acceptance. Set $\rho_k = \frac{ared}{pred}$.

If $\rho_k < \eta_1$,
  set $\beta_k = \gamma_1 \beta_k$
  go to Step 1
else set $x_{k+1}^\delta = x_k^\delta + p_k$

Step 3: Parameter update:Set

$$\begin{array}{lll} \beta_{k+1} \in (0, \beta_k] & \text{if } \rho_k \geq \eta_2 & \text{(very successful)}, \\ \beta_{k+1} = \beta_k & \text{if } \eta_1 \leq \rho_k < \eta_2 & \text{(successful)}. \end{array}$$

TR, ARQ, ARC and ill-posed problems

- These methods are extensively studied for well-posed problems.
- Can we prove convergence results for ill-posed problems in the noise-free case?
- Can we see these methods as regularizing methods for noisy problems?

First step in proving regularity of TR methods: [Wang, Yuan 2002, 2005].

Standard Assumption for the local analysis of regularizing methods:

ASS. Given an initial guess x_0:

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

for $c > 0$ and $x, \tilde{x} \in B_\rho(x_0)$. If $F'(x)$ is singular this implies:

F constant along the affine subspace $x + \mathcal{N}(F'(x)) \cap B_\rho(x_0)$, for any $x \in B_\rho(x_0)$.

$$\boxed{\mathcal{N}(F'(x)) = \mathcal{N}(F'(\tilde{x})) \quad \text{for any} \quad x, \tilde{x} \ \text{ s.t. } \ F(x) = F(\tilde{x})}$$

[Kaltenbacher, Neubauer, Scherzer, 2008]

Standard Assumption for the local analysis of regularizing methods:

ASS. Given an initial guess x_0:

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

for $c > 0$ and $x, \tilde{x} \in B_\rho(x_0)$. If $F'(x)$ is singular this implies:

F constant along the affine subspace $x + \mathcal{N}(F'(x)) \cap B_\rho(x_0)$, for any $x \in B_\rho(x_0)$.

$\mathcal{N}(F'(x)) = \mathcal{N}(F'(\tilde{x}))$ for any $x, \tilde{x}$ s.t. $F(x) = F(\tilde{x})$

[Kaltenbacher, Neubauer, Scherzer, 2008]

Standard Assumption for the local analysis of regularizing methods:

ASS. Given an initial guess x_0:

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

for $c > 0$ and $x, \tilde{x} \in B_\rho(x_0)$. If $F'(x)$ is singular this implies:

F constant along the affine subspace $x + \mathcal{N}(F'(x)) \cap B_\rho(x_0)$, for any $x \in B_\rho(x_0)$.

$$\boxed{\mathcal{N}(F'(x)) = \mathcal{N}(F'(\tilde{x})) \quad \text{for any} \quad x, \tilde{x} \text{ s.t. } F(x) = F(\tilde{x})}$$

[Kaltenbacher, Neubauer, Scherzer, 2008]

Consequence of Ass. 1

If $F(x) = y$ is solvable in $B_\rho(x_0)$, then there exists a solution x^+ such that

$$x_0 - x^+ \in \mathcal{N}(F'(x^+))^\perp \qquad x_0\text{-minimum norm solution}$$

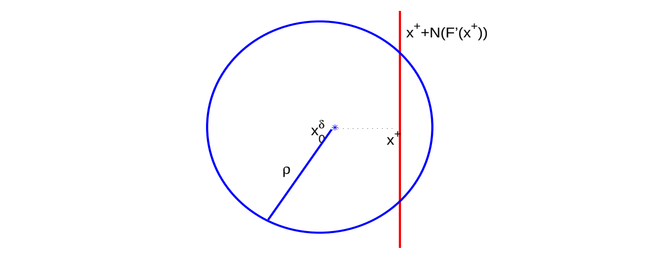


Figure : The x_0-minimum norm solution and the affine space of points x s.t. $F(x) = F(x^+) = y$

The q-condition [Hanke,1997,2010]

The decrease of the norm of the linear model is checked out.

- The LM step must be s.t.

$$\|F(x_k^\delta) - y^\delta + F'(x_k^\delta)p_k\| = q\|F(x_k^\delta) - y^\delta\|,$$

where $q \in (0,1)$ is a fixed constant.

- This calls for the solution of a secular equation.
- Discrepancy principle: the iterations are stopped at index k^* if

$$\|y^\delta - F(x_{k^*}^\delta)\| \le \tau\delta < \|y^\delta - F(x_k^\delta)\| \quad 0 \le k < k^*,$$

where τ is an appropriately chosen positive number.

The q-condition c.ed

- A sufficiently small step is needed in order to prevent to approach the solution of the noisy problem and to leave the region around x^+
- The q-condition prevents to take too long steps

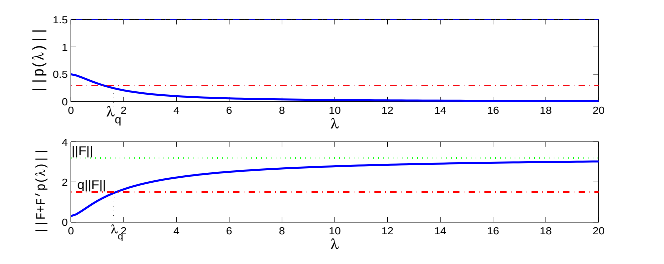


Figure : $\|p(\lambda)\|$ and $\|F - y + F'(p(\lambda))\|$ varying λ.

The q-condition c.ed

Imposing the q-condition guarantees:

- Local convergence to x^+ in the noise-free case ($\delta = 0$)
- For x_0^δ sufficiently close to x^+, the discrepancy principle is satisfied after a finite number of iterations k^* and

$$\|x_{k+1}^\delta - x^+\| < \|x_k^\delta - x^+\| \quad k = 0, 1, \ldots, k^*$$

- $\|x_{k^*}^\delta - x^+\|$ converges to zero whenever δ goes to zero.

$\Rightarrow$ REGULARIZING METHOD

Regularizing TR and ARC approaches

Given the model m_k (m_k^{TR}, m_k^C), the corresponding regularization term β ($1/\Delta_k$, σ_k), $0 < \eta_1 < \eta_2 < 1$, $\gamma_1 > 1$, $q \in (0, 1)$:

Step 1: Compute p_k the minimizer of the model $m_k(p)$(possibly constrained in case of TR).

Step 2: Step acceptance. Set $\rho_k = \frac{ared}{pred}$ and
$q_k = \|F(x_k^\delta) - y^\delta + F'(x_k^\delta)p_k\| / \|F(x_k^\delta) - y^\delta\|$
If $\rho_k < \eta_1$ or $q_k < q$
set $\beta_k = \gamma_1 \beta_k$
go to Step 1
else set $x_{k+1}^\delta = x_k^\delta + p_k$

Step 3: Parameter update:Set

$$\begin{array}{lll} \beta_{k+1} \in (0, \beta_k] & \text{if } \rho_k \geq \eta_2 & \text{(very successful)}, \\ \beta_{k+1} = \beta_k & \text{if } \eta_1 \leq \rho_k < \eta_2 & \text{(successful)}. \end{array}$$

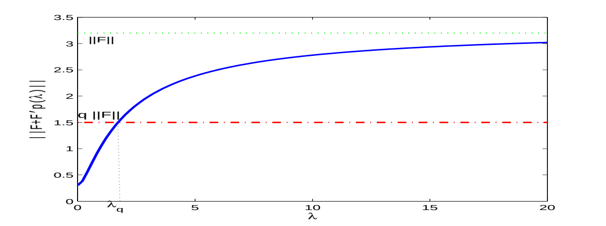
Enforcing the q-condition

The condition

$$\|F(x_k^\delta) - y^\delta + F'(x_k^\delta)p_k\| \geq q\|F(x_k^\delta) - y^\delta\|,$$

is enough to obtain a regularizing method.

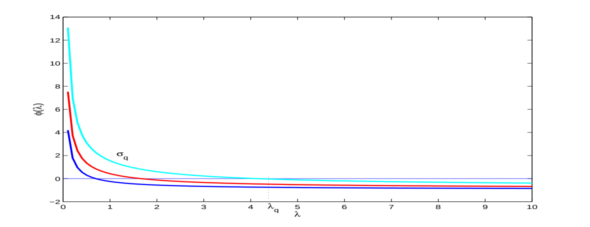
Let λ_q s.t $\|F(x_k^\delta) - y^\delta + F'(x_k\delta)p_k\| = q\|F(x_k^\delta) - y^\delta\|$, the q-condition is satisfied for $\lambda_k \geq \lambda_q$.



After a finite number of increases of β_k the q condition is satisfied.

Enforcing the q-condition c.ed

- TR: Let $\Delta_q = \|p(\lambda_q)\|$, due to the monotonicity of $\|p(\lambda)\|$, the q-condition is satisfied for $\Delta_k > (1/\gamma_1)\Delta_q$.
- ARC: Let σ_q such that $p_k = p(\lambda_q)$ (we can prove that σ_q exists).
  Increasing σ_k, λ_k increases and therefore for $\sigma_k < \gamma_1\sigma_q$ the q-condition is satisfied.



Our Assumptions:

Noise-free case:

- Global convergence to a stationary point: assume that $\lim_{k\to\infty} \|F'(x_k)^T(F(x_k) - y)\| = 0$.
- Assume that there exists an accumulation point x^+ s.t. $F(x^+) = y$. This implies

$$\|y - F(x_k)\| \to 0.$$

Noisy case:

- Assume that the initial guess is enough close to x^+. Otherwise we can only show the monotonic decrease of the noisy residual $\|y^\delta - F(x_k^\delta)\|$ and the satisfaction of the discrepancy principle.

Local convergence results

- λ_k must be greater than zero: i.e. the trust-region must be forced to be active.
- λ_k must be uniformly bounded from above:
  - We can prove that this is true for TR, ARC

$\Rightarrow$ TR, ARC are regularizing methods for $\tau > \frac{1}{q}$.

LM methods for a class of problems

Assumptions:

- Source condition:

  $$x^+ - x_0^\delta = (F'(x^+)^T F'(x^+))^\nu v, \quad \nu \in (0, 1/2] \quad v \in \mathcal{R}(F'(x^+)^T),$$

  $\|v\|$ sufficiently small.
- Restriction on the "nonlinearity" of F:

  $$F'(x) = R_x F'(x^+) \quad R_x \in \mathbb{R}^{n\times n} \quad \|I - R_x\| \leq c_R \|x - x^+\|$$

  for any $x \in B_\rho(x^+)$.

[Kaltenbacher, Neubauer, Scherzer, 2008]

LM approach [Kaltenbacher, Neubauer, Scherzer, 2008]

A priori choice of λ_k:

$$\lambda_k = \lambda_0 q^k, \text{ for some } \lambda_0 > 0, q \in (0,1).$$

Iterations stopped at index k^* s.t.

$$\lambda_k^* < C\delta \qquad C > 0$$

The decrease of λ_k is explicitly imposed.

Less expensive
but

the above strong conditions on x_0^δ and F are needed to prove its regularizing properties.
It is also possible to give complexity results:

$$k^* = O(1 + \ln(\delta)).$$

LM approach [Kaltenbacher, Neubauer, Scherzer, 2008]

A priori choice of λ_k:

$$\lambda_k = \lambda_0 q^k, \text{ for some } \lambda_0 > 0, q \in (0,1).$$

Iterations stopped at index k^* s.t.

$$\lambda_k^* < C\delta \qquad C > 0$$

The decrease of λ_k is explicitly imposed.

Less expensive
but

the above strong conditions on x_0^δ and F are needed to prove its regularizing properties.
It is also possible to give complexity results:

$$k^* = O(1 + \ln(\delta)).$$

Comments on the source condition

Observation: the local regularizing behaviour of this LM approach under the source condition

$$x^+ - x_0^\delta = (F'(x^+)^T F'(x^+))^{1/2} v, \quad v \in \mathcal{R}(F'(x^+)^T), \quad \|v\| \text{ small}$$

is equivalent to measure the error $x^+ - x_k^\delta$ in the weighted norm

$$\|x^+ - x_k^\delta\|_+^2 = \|(F'(x^+)^T F'(x^+))^{1/2}(x^+ - x_k^\delta)\|_2^2;$$

Note that $\|x^+ - x_k^\delta\| \in \mathcal{R}((F'(x^+)^T))$ so that $\|x^+ - x_k^\delta\|_+^2 = 0$ iff $\|x^+ - x_k^\delta\| = 0$.

Comments on the source condition c.ed

In other words, the source condition can be replaced by the following assumption:

Assume:

$$\|x^+ - x_0^\delta\|_+^2 \quad \text{and} \quad \|x^+ - x_0^\delta\|_2^2$$

sufficiently small.

Then, the errors

$$\|x^+ - x_k^\delta\|_+^2 \qquad \|x^+ - x_k^\delta\|_2^2$$

remain small and decrease until the stopping criterion is met.

Weaker Assumptions

Analysis carried out assuming:

Let x^+ be the x_0-minimum norm solution.

$$\mathcal{N}(F'(x^+)) \subseteq \mathcal{N}(F'(x)) \quad \text{for any} \quad x \in B_\rho(x^+).$$

- It is a weaker condition: it is implied by the "restriction on the nonlinearity of F" ($F'(x) = R_x F'(x^+)$)
- Let $\varsigma_1 \geq \varsigma_2 \geq \ldots \geq \varsigma_\ell > \varsigma_{\ell+1} = \ldots = \varsigma_n = 0$ be the singular values of $F'(x)$. Then:

  ς_ℓ bounded away from zero for any $x \in B_\rho(x^+)$.

Weaker Assumptions c.ed

Let x^+ be the x_0-minimum norm solution. As in a LM approach

$$p_k \in \mathcal{R}(F'(x_k^\delta)^T)$$

the assumption

$$\mathcal{N}(F'(x^+)) \subseteq \mathcal{N}(F'(x)) \quad \text{for any} \quad x \in B_\rho(x^+).$$

implies:

- $p_k \in \mathcal{R}(F'(x^+)^T)$ for any $x_k^\delta \in B_\rho(x^+)$.
- $x_k^\delta - x^+ \in \mathcal{R}(F'(x^+)^T)$ for any $x_k^\delta \in B_\rho(x^+)$.

Noise Free case

Can we prove local convergence of TR, ARC, ARQ under the Assumption $\mathcal{N}(F'(x^+)) \subseteq \mathcal{N}(F'(x_k))$?

- Global convergence to a stationary point: assume that $\lim_{k\to\infty} \|F'(x_k)^T(F(x_k) - y)\| = 0$.
- Assume that there exists an accumulation point x^+ s.t. $F(x^+) = y$ and

$$\mathcal{N}(F'(x^+)) \subseteq \mathcal{N}(F'(x)) \qquad x \in B_\rho(x^+).$$

Then, we can prove that x^+ is an isolated limit point of $\{x_k\}$.

Noise Free case c.ed

- Note that $r_k = U_k^T(F(x_k) - y) \to 0$.
- Then, as $\ell_k \leq \ell_+$, with $\ell_+ = rank(F'(x^+))$ it follows:
  - $(\varsigma_k)_{\ell_k}$ is bounded away from zero.
  - $\|p_k\| \to 0$ as

$$\|p_k\|^2 = \|p(\lambda_k)\|^2 = \sum_{i=1}^{\ell_k} \left(\frac{(\varsigma_k)_i r_i}{(\varsigma_k)_i^2 + \lambda} \right)^2 \leq \frac{(\varsigma_k)_1}{(\varsigma_k)_{\ell_k}} ||r_k||^2$$

$\Longrightarrow \quad x_k \to x^+$ [Moré, Sorensen, 1983]

This proves the convergence to x^+ of the sequence generated by any globally convergent LM approach

Noise Free case c.ed

- Note that $r_k = U_k^T(F(x_k) - y) \to 0$.
- Then, as $\ell_k \leq \ell_+$, with $\ell_+ = rank(F'(x^+))$ it follows:
  - $(\varsigma_k)_{\ell_k}$ is bounded away from zero.
  - $\|p_k\| \to 0$ as

$$\|p_k\|^2 = \|p(\lambda_k)\|^2 = \sum_{i=1}^{\ell_k} \left(\frac{(\varsigma_k)_i r_i}{(\varsigma_k)_i^2 + \lambda} \right)^2 \leq \frac{(\varsigma_k)_1}{(\varsigma_k)_{\ell_k}} ||r_k||^2$$

$\Rightarrow \quad x_k \to x^+$ [Moré, Sorensen, 1983]

This proves the convergence to x^+ of the sequence generated by any globally convergent LM approach

Noisy-case

Given $C > 0, \tilde{\gamma} > 0$ assume that:

- $\lambda_0 > C\delta$.
- $\|x_0^\delta - x^+\| \leq \tilde{\gamma}\lambda_0^{1/2}$. This condition is ensured whenever

$$\|x_0^\delta - x^+\| \leq \tilde{\gamma}\sqrt{C\delta}.$$

- Stopping criteria: the iterations are stopped at index k^* if

$$\lambda_{k^*} \leq C\delta \quad \text{or} \quad \|y^\delta - F(x_{k^*}^\delta)\| \leq C\delta \quad C > 0$$

Noisy-case: assumptions on λ_k

- The sequence λ_k must go to zero slowly. A sufficiently large regularization is needed in order to prevent to approach the solution of the noisy problem and to leave the region around x^+ :

$$\frac{\lambda_k}{\lambda_{k+1}} \leq \Omega \qquad \Omega > 0.$$

- $\sum_{k=1}^{k^*} \lambda_k^{1/2} \leq M$ with $M > 0$ independent of k^*.
- Decrease of λ_k must be related to the decrease of the nonlinear residual:

$$\lambda_k > \gamma \|F(x_k^\delta) - y^\delta\| \qquad \gamma > 0.$$

Noisy case: local properties

Under the above assumptions we can prove that :

$$\|x_k^\delta - x^+\| \le \Lambda \lambda_{k-1}^{1/2}$$

whenever $\lambda_k \ge C\delta$ and $\|y^\delta - F(x_k^\delta)\| \ge C\delta$, with Λ independent of k.

- If $\delta > 0$, the error $\|x_k^\delta - x^+\|$ remains bounded and decreases with λ_k until the stopping criterion

$$\lambda_{k_*} < C\delta \quad \text{or} \quad \|y^\delta - F(x_{k^*}^\delta)\| < C\delta$$

  at iteration k^* is met.
- $x_{k^*}^\delta \to x^+$ as δ goes to zero.

$\Longrightarrow$ REGULARIZING METHOD

Noisy case: local properties

Under the above assumptions we can prove that :

$$\|x_k^\delta - x^+\| \leq \Lambda \lambda_{k-1}^{1/2}$$

whenever $\lambda_k \geq C\delta$ and $\|y^\delta - F(x_k^\delta)\| \geq C\delta$, with Λ independent of k.

- If $\delta > 0$, the error $\|x_k^\delta - x^+\|$ remains bounded and decreases with λ_k until the stopping criterion

$$\lambda_{k_*} < C\delta \quad \text{or} \quad \|y^\delta - F(x_{k^*}^\delta)\| < C\delta$$

  at iteration k^* is met.
- $x_{k^*}^\delta \to x^+$ as δ goes to zero.

$\Longrightarrow$ REGULARIZING METHOD

Regularizing Trust-region and ARQ

Can we design TR and ARQ methods that fits in the above process?

- Can we choose in an appropriate way Δ_k (TR) and μ_k (ARQ) in order to get regularizing methods?
- Assume $F'(x)$ is Lipschitz continuous and $\|F'(x)\|$ bounded.

Regularizing Trust-region and ARQ c.ed

Further assumption to get $\sum_{k=1}^{k^*} \lambda_k^{1/2} \leq M$:

Let

$$F(x_k^\delta) - y^\delta = F_{\mathcal{R}} + F_{\mathcal{N}} \quad \text{with} \quad F_{\mathcal{R}} \in \mathcal{R}(F'(x_k^\delta)^T), \quad F_{\mathcal{N}} \in \mathcal{N}(F'(x_k^\delta))$$

Then

$$\|F_{\mathcal{R}}\|$$

does not have to reduce faster than

$$\|F_{\mathcal{N}}\|$$

Reasonable as, if $F(x) = y^\delta$ admits a solution x_δ^+,
$\|F_{\mathcal{N}}\| = o(\|x_k - x_\delta^+\|)^2$ and $\|F_{\mathcal{R}}\| = o(\|x_k - x_\delta^+\|)$

A regularizing TR

The scalar λ_k is chosen a priori:

$$\lambda_k = \tau_k \|F(x_k^\delta) - y^\delta\|$$

with an adaptive choice of τ_k. [Fan, 2003]

The step p_k is the solution of a TR subproblem with

$$\Delta_k = \|p(\lambda_k)\|$$

the trust region is always active.

- Step acceptance: $\tau_k = 4\tau_k$ whenever $\rho_k < \eta_1$
- Parameter updating:

$$\tau_{k+1} = \begin{cases} \max(\tau_k/4, m) & \text{if } \rho_k \geq \eta_2 & \text{(very successful),} \\ \tau_k & \text{if } \eta_1 \leq \rho_k < \eta_2 & \text{(successful),} \end{cases}$$

A regularizing TR

The scalar λ_k is chosen a priori:

$$\lambda_k = \tau_k \|F(x_k^\delta) - y^\delta\|$$

with an adaptive choice of τ_k. [Fan, 2003]

The step p_k is the solution of a TR subproblem with

$$\Delta_k = \|p(\lambda_k)\|$$

the trust region is always active.

- Step acceptance: $\tau_k = 4\tau_k$ whenever $\rho_k < \eta_1$
- Parameter updating:

$$\tau_{k+1} = \begin{cases} \max(\tau_k/4, m) & \text{if } \rho_k \geq \eta_2 \quad \text{(very successful)}, \\ \tau_k & \text{if } \eta_1 \leq \rho_k < \eta_2 \quad \text{(successful)}, \end{cases}$$

A regularizing TR c.ed

The method has to be modified in order to be regularizing. It is not guaranteed that

$$\frac{\lambda_k}{\lambda_{k+1}} = \frac{\tau_k \|F(x_k^\delta) - y^\delta\|}{\tau_{k+1}\|F(x_{k+1}^\delta) - y^\delta\|} \leq \Omega \qquad \Omega > 0.$$

- Modification:

$$\tau_{k+1} = \bar{\tau}_{k+1} \max(1, \frac{1}{\Omega} \frac{\|F(x_k^\delta) - y^\delta\|}{\|F(x_{k+1}^\delta) - y^\delta\|})$$

with

$$\bar{\tau}_{k+1} = \begin{cases} \max(\tau_k/4, m) & \text{if } \rho_k \geq \eta_2 \quad \text{(very successful)}, \\ \tau_k & \text{if } \eta_1 \leq \rho_k < \eta_2 \quad \text{(successful)}, \end{cases}$$

A regularizing TR c.ed

In practice, when Fan's choice of λ_{k+1} is too small we set

$$\lambda_{k+1} = \frac{1}{4\Omega}\|F(x_k^\delta) - y^\delta\|.$$

- The condition $\lambda_k \geq \gamma\|F(x_k^\delta) - y^\delta\|$ is satisfied.

The assumptions on λ_k are fitted and the method is regularizing.

A regularizing ARQ

Properties of λ_k:

- $\lambda_k = \mu_k + \sigma_k \| F(x_k^\delta) - y^\delta + F'(x_k^\delta)^T p(\lambda_k) \|$.
- $\lambda_k \leq \mu^{up} + \sigma_{max} \| F(x_k^\delta) - y^\delta \|$

$$\lambda_k \geq \mu_k :$$

we can control how λ_k goes to zero choosing μ_k.

A regularizing ARQ c.ed

Choice of μ_k in order to get

$$\frac{\lambda_k}{\lambda_{k+1}} \le \Omega \qquad \Omega > 0 \quad \text{and } \lambda_k \ge \gamma \|F(x_k^\delta) - y^\delta\|.$$

We set

$$\mu_k = \max(\|F(x_k^\delta) - y^\delta\|^{1/2}, q\mu_{k-1}) \qquad q \in (0,1).$$

This yields

$$\frac{\lambda_k}{\lambda_{k+1}} \le \underbrace{\frac{1}{q} + \sigma_{max}\|F(x_0^\delta) - y^\delta\|^{1/2}}_{\Omega}.$$

and

$$\lambda_k \ge \|F(x_k^\delta) - y^\delta\|$$

The assumptions on λ_k are fitted and the method is regularizing.

Comments and open issues

- How do they work in practice?
- A step p_k statisfying the q-condition

$$\|F(x_k^\delta) - y^\delta + F'(x_k^\delta)p_k\| = q\|F(x_k^\delta) - y^\delta\|,$$

is a Inexact Newton step.

- Can we prove regularization properties using approximate minimizers of the model in TR,ARC,ARQ? The previous analysis does not apply. First step in this direction: [Wang, Yuan, 2005].
- Can we develop a variant of the TR method that gets rid of the q condition?

THANK YOU FOR YOUR ATTENTION