Methods for Inverse Problems:
IV. Newton type methods

Barbara Kaltenbacher, University of Klagenfurt, Austria

overview

1. Newton's method
2. Levenberg-Marquardt
   - Monotonicity of the errors
   - Convergence
   - Convergence rates
3. Iteratively regularized Gauss-Newton method (IRGNM)
   - Convergence and convergence rates

Newton's method

$$F'(x_k^\delta)(x_{k+1}^\delta - x_k^\delta) = y^\delta - F(x_k^\delta)\,. \qquad (1)$$

formulation as least squares problem

$$\min_{x \in \mathcal{D}(F)} \|y^\delta - F(x_k^\delta) - F'(x_k^\delta)(x - x_k^\delta)\|^2$$

$\rightsquigarrow$ ill-posedness $\rightsquigarrow$ apply Tikhonov regularization:
Levenberg-Marquardt method:

$$\min_{x \in \mathcal{D}(F)} \|y^\delta - F(x_k^\delta) - F'(x_k^\delta)(x - x_k^\delta)\|^2 + \alpha_k \|x - x_k^\delta\|^2\,, \qquad (2)$$

Iteratively regularized Gauss-Newton method (IRGNM)

$$\min_{x \in \mathcal{D}(F)} \|y^\delta - F(x_k^\delta) - F'(x_k^\delta)(x - x_k^\delta)\|^2 + \alpha_k \|x - x_0\|^2 \qquad (3)$$

choice of sequence α_k and convergence anaylsis different for both methods.

Levenberg-Marquardt

$$x_{k+1}^\delta = x_k^\delta + (F'(x_k^\delta)^* F'(x_k^\delta) + \alpha_k I)^{-1} F'(x_k^\delta)^* (y^\delta - F(x_k^\delta)) \,, \quad (4)$$

Choice of α_k:

$$\|y^\delta - F(x_k^\delta) - F'(x_k^\delta)(x_{k+1}^\delta(\alpha_k) - x_k^\delta)\| = q \|y^\delta - F(x_k^\delta)\| \quad (5)$$

for some $q \in (0,1) \rightsquigarrow$ inexact Newton method.
(5) has a unique solution α_k provided that for some $\gamma > 1$

$$\|y^\delta - F(x_k^\delta) - F'(x_k^\delta)(x^\dagger - x_k^\delta)\| \leq \frac{q}{\gamma} \|y^\delta - F(x_k^\delta)\| \quad (6)$$

which can be guaranteed by a condition on F: $\forall x, \tilde{x} \in \mathcal{B}_{2\rho}(x_0) \subseteq \mathcal{D}(F)$

$$\|F(x) - F(\tilde{x}) - F'(x)(x - \tilde{x})\| \leq c \|x - \tilde{x}\| \, \|F(x) - F(\tilde{x})\| \,, \quad (7)$$

Choice of stopping index k_*: discrepancy principle:

$$\|y^\delta - F(x_{k_*}^\delta)\| \leq \tau\delta < \|y^\delta - F(x_k^\delta)\| \,, \qquad 0 \leq k < k_* \,, \quad (8)$$

[Hanke 1996]

Levenberg-Marquardt: Monotonicity of the errors

Theorem

Let $0 < q < 1 < \gamma$ and assume that $F(x) = y$ has a solution and that (6) holds so that α_k can be defined via (5). Then, the following estimates hold:

$$\|x_k^\delta - x^\dagger\|^2 - \|x_{k+1}^\delta - x^\dagger\|^2 \geq \|x_{k+1}^\delta - x_k^\delta\|^2 , \tag{9}$$

$$\begin{aligned} &\|x_k^\delta - x^\dagger\|^2 - \|x_{k+1}^\delta - x^\dagger\|^2 \\ &\quad \geq \frac{2(\gamma - 1)}{\gamma \alpha_k} \|y^\delta - F(x_k^\delta) - F'(x_k^\delta)(x_{k+1}^\delta - x_k^\delta)\|^2 \qquad (10) \\ &\quad \geq \frac{2(\gamma - 1)(1 - q)q}{\gamma \|F'(x_k^\delta)\|^2} \|y^\delta - F(x_k^\delta)\|^2 . \qquad (11) \end{aligned}$$

Levenberg-Marquardt: Monotonicity proof

$$K_k := F'(x_k^\delta)$$
$$x_{k+1}^\delta - x_k^\delta = K_k^*(K_k K_k^* + \alpha_k I)^{-1}(y^\delta - F(x_k^\delta))$$
$$\alpha_k(K_k K_k^* + \alpha_k I)^{-1}(y^\delta - F(x_k^\delta)) = y^\delta - F(x_k^\delta) - K_k(x_{k+1}^\delta - x_k^\delta)\,,$$

$$\begin{aligned}
&\|x_{k+1}^\delta - x^\dagger\|^2 - \|x_k^\delta - x^\dagger\|^2 \\
&\quad = \; 2\langle\, x_{k+1}^\delta - x_k^\delta, x_k^\delta - x^\dagger\,\rangle + \|x_{k+1}^\delta - x_k^\delta\|^2 \\
&\quad = \; \langle\, (K_k K_k^* + \alpha_k I)^{-1}(y^\delta - F(x_k^\delta)), \\
&\qquad\quad 2K_k(x_k^\delta - x^\dagger) + (K_k K_k^* + \alpha_k I)^{-1} K_k K_k^*(y^\delta - F(x_k^\delta))\,\rangle \\
&\quad = \; -2\alpha_k \|(K_k K_k^* + \alpha_k I)^{-1}(y^\delta - F(x_k^\delta))\|^2 \\
&\qquad - \|(K_k^* K_k + \alpha_k I)^{-1} K_k^*(y^\delta - F(x_k^\delta))\|^2 \\
&\qquad + 2\langle\, (K_k K_k^* + \alpha_k I)^{-1}(y^\delta - F(x_k^\delta)), y^\delta - F(x_k^\delta) - K_k(x^\dagger - x_k^\delta)\,\rangle \\
&\quad \le \; -\|x_{k+1}^\delta - x_k^\delta\|^2 - 2\alpha_k^{-1}\|y^\delta - F(x_k^\delta) - K_k(x_{k+1}^\delta - x_k^\delta)\| \cdot \\
&\qquad \Big(\|y^\delta - F(x_k^\delta) - K_k(x_{k+1}^\delta - x_k^\delta)\| - \|y^\delta - F(x_k^\delta) - K_k(x^\dagger - x_k^\delta)\|\Big)\,.
\end{aligned}$$

$$\|y^\delta - F(x_k^\delta) - K_k(x^\dagger - x_k^\delta)\| \le \gamma^{-1}\|y^\delta - F(x_k^\delta) - K_k(x_{k+1}^\delta - x_k^\delta)\|\,.$$

Levenberg-Marquardt method: Convergence

Theorem

Let $0 < q < 1$ and assume that $F(x) = y$ is solvable in $\mathcal{B}_\rho(x_0)$, that F' is uniformly bounded in $\mathcal{B}_\rho(x^\dagger)$, and that the Taylor remainder of F satisfies (7) for some $c > 0$. Then the Levenberg-Marquardt method with exact data $y^\delta = y$, $\|x_0 - x^\dagger\| < q/c$ and α_k determined from (5), converges to a solution of $F(x) = y$ as $k \to \infty$.

Theorem

Let the assumptions of Theorem 2 hold. Additionally let $k_* = k_*(\delta, y^\delta)$ be chosen according to the stopping rule (8) with $\tau > 1/q$ and let $\|x_0 - x^\dagger\|$ be sufficiently small. Then for some solution x_* of $F(x) = y$

$$k_*(\delta, y^\delta) = O(1 + |\ln \delta|) \text{ and } \|x^\delta_{k_*} - x_*\| \to 0 \text{ as } \delta \to 0$$

Levenberg-Marquardt method: Convergence rates

Theorem

Let a solution $x^\dagger$ of $F(x) = y$ exist and let

$$F'(x) = R_x F'(x^\dagger) \text{ and } \|I - R_x\| \leq c_R \|x - x^\dagger\| \,,\; x \in \mathcal{B}_\rho(x_0) \subseteq \mathcal{D}(F)\,, \tag{12}$$

$$x^\dagger - x_0 = (F'(x^\dagger)^* F'(x^\dagger))^\mu v\,, \quad v \in \mathcal{N}(F'(x^\dagger))^\perp \tag{13}$$

hold with some $0 < \mu \leq 1/2$ and $\|v\|$ sufficiently small. Moreover, let α_k and k_* be chosen according to (5) and (8), respectively with $\tau > 2$ and $1 > q > 1/\tau$. Then the Levenberg-Marquardt iterates defined by (4) remain in $\mathcal{B}_\rho(x_0)$ and converge with the rate

$$\|x_{k_*}^\delta - x^\dagger\| = O\left(\delta^{\frac{2\mu}{2\mu+1}}\right).$$

[Hanke 2009]

Remarks

- rates with a priori α_k, k_*:

$$\alpha_k = \alpha_0 q^k \,, \qquad \text{for some} \quad \alpha_0 > 0 \,, \quad q \in (0,1) \,,$$

$$c(k_*+1)^{-(1+\varepsilon)}\alpha_{k_*}^{\mu+\frac{1}{2}} \leq \delta < c(k+1)^{-(1+\varepsilon)}\alpha_k^{\mu+\frac{1}{2}} \,, \qquad 0 \leq k < k_* \,,$$

$$k_* = O(1+|\ln\delta|) \,, \quad \|x_{k_*}^\delta - x^\dagger\| = O\Big(\big(\delta\,(1+|\ln\delta|)^{(1+\varepsilon)}\big)^{\frac{2\mu}{2\mu+1}} \Big) \,.$$

[BK&Neubauer&Scherzer 2008]

- generalization to other regularization methods (e.g., CG) in place of Tikhonov [Hanke 1997], [Rieder 1999, 2001, 2005]

Iterative regularized Gauss-Newton method (IRGNM)

$$x_{k+1}^\delta = x_k^\delta + (F'(x_k^\delta)^* F'(x_k^\delta) + \alpha_k I)^{-1} (F'(x_k^\delta)^* (y^\delta - F(x_k^\delta)) + \alpha_k (x_0 - x_k^\delta)) \,. \quad (14)$$

a-priori choice of α_k:

$$\alpha_k > 0 \,, \qquad 1 \leq \frac{\alpha_k}{\alpha_{k+1}} \leq r \,, \qquad \lim_{k\to\infty} \alpha_k = 0 \,, \qquad (15)$$

for some $r > 1$.
a-priori or a posteriori choice of k_*

$$\|y^\delta - F(x_{k_*}^\delta)\| \leq \tau\delta < \|y^\delta - F(x_k^\delta)\| \,, \qquad 0 \leq k < k_* \,, \qquad (16)$$

[Bakushinski 1992], see also the book [Bakushinski&Kokurin 2004];
[BK&Neubauer&Scherzer 1997], see also the book [BK& Neubauer&Scherzer 2008

IRGNM: Convergence and convergence rates: idea of proof I

key idea:
$\|x_{k+1}^\delta - x^\dagger\| \approx \alpha_k^\mu w_k(\mu)$ with $w_k(s)$ as in the following lemma.

Lemma

Let $K \in \mathcal{L}(\mathcal{X},\mathcal{Y})$, $s \in [0,1]$, and let $\{\alpha_k\}$ be a sequence satisfying $\alpha_k > 0$ and $\alpha_k \to 0$ as $k \to \infty$. Then it holds that

$$w_k(s) := \alpha_k^{1-s} \|(K^*K+\alpha_k I)^{-1}(K^*K)^s v\| \le s^s(1-s)^{1-s} \|v\| \le \|v\| \tag{17}$$

and that

$$\lim_{k\to\infty} w_k(s) = \begin{cases} 0\,, & 0 \le s < 1\,, \\ \|v\|\,, & s = 1\,, \end{cases}$$

for any $v \in \mathcal{N}(A)^\perp$.

IRGNM: Convergence and convergence rates: idea of proof I

Indeed, in the linear and noiseless case ($F(x) = Kx$, $\delta = 0$) we get from (14) using $Kx^\dagger = y$ and (13)

$$\begin{aligned} & x_{k+1} - x^\dagger \\ & \quad = x_k - x^\dagger + (K^*K + \alpha_k I)^{-1}(K^*K(x^\dagger - x_k) + \alpha_k(x_0 - x^\dagger + x^\dagger - x_k)) \\ & \quad = \; -\alpha_k(K^*K + \alpha_k I)^{-1}(K^*K)^\mu v \end{aligned}$$

To take into account noisy data and nonlinearity, we rewrite (14) as

$$\begin{aligned} x_{k+1}^\delta - x^\dagger \;=\; & -\alpha_k(K^*K + \alpha_k I)^{-1}(K^*K)^\mu v \\ & -\alpha_k(K_k^*K_k + \alpha_k I)^{-1}\Big(K^*K - K_k^*K_k\Big) \\ & \qquad (K^*K + \alpha_k I)^{-1}(K^*K)^\mu v \\ & + (K_k^*K_k + \alpha_k I)^{-1}K_k^*(y^\delta - F(x_k^\delta) + K_k(x_k^\delta - x^\dagger))\,. \end{aligned} \tag{18}$$

where we set $K_k := F'(x_k^\delta)$, $K := F'(x^\dagger)$.

IRGNM: Convergence and convergence rates

Theorem

Let $\mathcal{B}_{2\rho}(x_0) \subseteq \mathcal{D}(F)$ *for some* $\rho > 0$, (15),

$$
\begin{aligned}
F'(\tilde{x}) &= R(\tilde{x},x)F'(x) + Q(\tilde{x},x) \\
\|I - R(\tilde{x},x)\| &\leq c_R\,, \quad \|Q(\tilde{x},x)\| \leq c_Q \|F'(x^\dagger)(\tilde{x}-x)\|
\end{aligned}
$$

and

$$
x^\dagger - x_0 = (F'(x^\dagger)^* F'(x^\dagger))^\mu v\,, \quad v \in \mathcal{N}(F'(x^\dagger))^\perp
$$

for some $0 \leq \mu \leq 1/2$, *and let* $k_* = k_*(\delta)$ *be chosen according to the discrepancy principle* (16) *with* $\tau > 1$*. Moreover, we assume that* $\|x_0 - x^\dagger\|$, $\|v\|$, $1/\tau$, ρ, *and* c_R *are sufficiently small. Then we obtain the rates*

$$
\|x_{k_*}^\delta - x^\dagger\| = \begin{cases} o\left(\delta^{\frac{2\mu}{2\mu+1}}\right), & 0 \leq \mu < \frac{1}{2}\,, \\ O(\sqrt{\delta})\,, & \mu = \frac{1}{2}\,. \end{cases}
$$

For convergence (without rates) the tangential cone condition suffices.

Remarks

- The same convergence rates result can be shown with the a priori stopping rule

$$k_* \to \infty \qquad \text{and} \qquad \eta \geq \delta \alpha_{k_*}^{-\frac{1}{2}} \to 0 \ \text{ as } \ \delta \to 0\,. \tag{19}$$

  for $\mu = 0$ and

$$\eta \alpha_{k_*}^{\mu+\frac{1}{2}} \leq \delta < \eta \alpha_k^{\mu+\frac{1}{2}}\,, \qquad 0 \leq k < k_*\,, \tag{20}$$

  even for $0 < \mu \leq 1$.

- The a priori result remains valid under the alternative weak nonlinearity condition

$$F'(\tilde{x}) = F'(x) R(\tilde{x}, x) \qquad \text{and} \qquad \|I - R(\tilde{x}, x)\| \leq c_R \|\tilde{x} - x\| \tag{21}$$

  for $x, \tilde{x} \in \mathcal{B}_{2\rho}(x_0)$ and some positive constant c_R.

Further remarks

- logarithmic rates: [Hohage 1997]
- generalization to regularization methods $R_\alpha(F'(x)) \approx F'(x)^\dagger$ in place of Tikhonov [BK 1997]

$$x_{k+1}^\delta = x_0 + R_{\alpha_k}(F'(x_k^\delta))(y^\delta - F(x_k^\delta) - F'(x_k^\delta)(x_0 - x_k^\delta)) . \quad (22)$$

- continuous version [BK&Neubauer&Ramm 2002]
- projected version for constrained problems [BK&Neubauer 2006]
- analysis with stochastic noise [Bauer&Hohage&Munk 2009]
- analysis in Banach space [Bakushinski&Konkurin 2004], [BK& Schöpfer&Schuster 2009], [BK& Hofmann 2010]
- preconditioning [Egger 2007], [Langer 2007]
- quasi Newton methods [BK 1998]