A convergence rates result for an iteratively regularized Gauss–Newton–Halley method in Banach space

B Kaltenbacher

Alpen-Adria Universität Klagenfurt, Universitätsstraße 65-67, A-9020 Klagenfurt, Austria

E-mail: barbara.kaltenbacher@aau.at

Received 19 September 2014, revised 3 December 2014
Accepted for publication 4 December 2014
Published 6 January 2015

Abstract
The use of second order information on the forward operator often comes at a very moderate additional computational price in the context of parameter identification problems for differential equation models. On the other hand the use of general (non-Hilbert) Banach spaces has recently found much interest due to its usefulness in many applications. This motivates us to extend the second order method from Kaltenbacher (2014 Numer. Math. at press), (see also Hettlich and Rundell 2000 SIAM J. Numer. Anal. 37 587620) to a Banach space setting and analyze its convergence. We here show rates results for a particular source condition and different exponents in the formulation of Tikhonov regularization in each step. This includes a complementary result on the (first order) iteratively regularized Gauss–Newton method in case of a one-homogeneous data misfit term, which corresponds to exact penalization. The results clearly show the possible advantages of using second order information, which get most pronounced in this exact penalization case. Numerical simulations for an inverse source problem for a nonlinear elliptic PDE illustrate the theoretical findings.

Keywords: regularization, Halley’s method, parameter identification

1. Introduction
Identification of parameters in ordinary or partial differential equations by Newton methods usually requires repeated solution of the model equation (the PDE or ODE) and its linearization, since these methods rely on a first order Taylor expansion of the parameter-to-state
It has already been observed in [10, 16] that also higher derivative evaluation for this forward operator typically leads to the same linear differential equation as the one arising for the first order derivative, and only the right-hand sides differ. Let us illustrate this by means of some examples.

**Example 1** Consider identification of the pair \((a, c)\) of possibly spatially varying coefficients in the nonlinear elliptic boundary value problem

\[
\begin{aligned}
\Delta f(a, \nabla u) + g(c, u) &= 0 \text{ in } \Omega \\
u &= h \text{ on } \partial \Omega
\end{aligned}
\]  

(1)

from measurements \(y = Cu\) of the state \(u\), where \(C\) is some linear operator (e.g., a trace operator in case of boundary measurements). Problems of this kind arise, e.g., in stationary inverse groundwater filtration, as well as in the characterization or nondestructive inspection of (non)linearly elastic or magnetic materials. Here \(\Omega \subseteq \mathbb{R}^d\), and the functions \( f, \mathbb{R}^{d+1} \to \mathbb{R}, \ g: \mathbb{R}^d \to \mathbb{R}, h \in H^{1/2}(\partial \Omega) \) are given. Note that linear growth of \( f, g \) and monotonicity (uniform one in case of \( f \)) with respect to their second argument allow to show well-posedness of (1) by means of the Lax–Milgram lemma. Then, using the parameter-to-state map \( S: (a, c) \mapsto u \), the forward operator and its derivatives at some point \((a, c)\) in parameter space can be written as

\[
F(a, c) = CS(a, c), \quad F'(a, c) = CS'(a, c), \quad F''(a, c) = CS''(a, c),
\]

where the derivatives of \( S \) at \((a, c)\) in certain directions can be recovered as solutions of the same linearized elliptic PDE with different right-hand sides: For parameter increments \((\alpha, \gamma), (\tilde{\alpha}, \tilde{\gamma})\) we have that

\[
\begin{aligned}
\alpha \gamma &= \partial v^1(a, c) \\
\alpha \gamma \gamma &= \partial v^2(a, c)
\end{aligned}
\]

solve the same linear system

\[
\begin{aligned}
-\nabla (\partial_1 f(a, \nabla u) \nabla v^i) + \partial_2 g(c, u) v^i &= b^i \text{ in } \Omega \\
v^i &= 0 \text{ on } \partial \Omega
\end{aligned}
\]  

(2)

where

\[
\begin{aligned}
b^1 &= \nabla (\partial_1 f(a, \nabla u) a) - \partial_1 g(c, u) \gamma \\
b^2 &= -\partial_1^2 g(c, u) (\gamma, \gamma) - \partial_1 \partial_2 g(c, u) (\gamma, v^i) - \partial_2 \partial_2 g(c, u) (v^i, v^i) \\
&+ \nabla (\partial_1^2 f(a, \nabla u) (a, \tilde{\alpha}) + \partial_1 \partial_2 f(a, \nabla u) (a, \nabla v^i) + \partial_2 \partial_2 f(a, \nabla u) (\tilde{\alpha}, \nabla v^i)
\end{aligned}
\]

and \( u = S(a, c), \ v^1 = S'(a, c)(\alpha, \gamma), \ v^2 = S''(a, c)(\alpha, \gamma, (\tilde{\alpha}, \tilde{\gamma})) \), i.e., the same linear elliptic boundary value problem (2), only with different right-hand sides.

**Example 2** For modeling time dependent problems, consider the state space model

\[
\begin{aligned}
\dot{u}(t) + f(t, u(t), c) &= 0, \quad t > 0, \\
u(0) &= u_0
\end{aligned}
\]

(3)

where the dot denotes the time derivative, which includes systems of ODEs but also (thinking of \( u(t) \) as an element of a function space over some spatial domain \( \Omega \)) time dependent PDEs. Given \( f, u_0 \), we seek to identify the parameter \( c \)—possibly element of a finite or infinite dimensional Banach space—from measurements \( y = Cu \) of the state \( u \), where \( C \) is some linear operator. Again, using the parameter-to-state-map \( S: c \mapsto u \), the forward operator and its derivatives at some point \( c \) in parameter space can be written as

\[
F(c) = CS(c), \quad F'(c) = CS'(c), \quad F''(c) = CS''(c),
\]

where \( v^1 = S'(c) \gamma, \ v^2 = S''(c)(\gamma, \tilde{\gamma}) \) solve the same linear system

\[
\begin{aligned}
\dot{v}^i(t) + \partial_1 f(t, u(t), c)v^i(t) &= b^i, \quad t > 0, \\
v^i(0) &= 0
\end{aligned}
\]  

(4)
with different right-hand sides

\[ b^1 = -\partial g(t, u(t), c) \gamma \]
\[ b^2 = -\partial^2 g(t, u(t), c)(\gamma, \gamma) - \partial_2 \partial g(t, u(t), c)\left(\gamma, \gamma^1(t)\right) - \partial_2 \partial g(t, u(t), c)\left(\gamma, \gamma^1(t)\right) \]

where \( u = S(c), \gamma = S'(c) \gamma \).

More generally, for a forward operator \( F \) involving a parameter-to-state map \( S \) for a PDE \( D(c, u) = 0 \) and a measurement operator \( M \)

\[ F(c) = M(S(c)) \] with \( S(c) = u \) satisfying \( D(c, u) = 0 \) (5)

we get \( F'(c)h = M'(S(c))S'(c)h \), \( F''(c)(h, l) = M'(S(c))S'(c)h, S'(c)l + M'(S(c))S''(c)(h, l) \), \( v^1 = S'(c)h \), \( v^2 = S''(c)h \) satisfy

\[ D_{\iota}(c, S(c))v^i = b^i \]

with

\[ b^1 = -D_{\iota}(c, S(c))h, \]
\[ b^2 = -D_{\iota}(c, S(c))(h, l) - D_{\mu}(c, S(c))(h, S'(c)l) \]
\[ -D_{\mu}(c, S(c))(S'(c)h, l) - D_{\mu\mu}(c, S(c))(S'(c)h, S'(c)l) \]

i.e. the same linear PDE with different right-hand sides for the first, second and actually all higher derivatives.

We point out that in these examples, evaluating \( F''(c_k) \) at some iterate \( c_k \) leads to the same linear problem as evaluating \( F'(c_k) \), just with some different inhomogeneities. In case of elliptic or parabolic PDEs this means that the stiffness matrix remains unchanged and therefore, once we have done the computations for \( F'(c_k) \), the additional effort for evaluating \( F''(c_k) \) can be kept quite moderate, usually much lower than the effort for doing an additional Newton step that requires \( F'(c_{k+1}) \), thus setting up a new stiffness matrix, at some different coefficient \( c_{k+1} \). This cheap evaluation of the second derivative at the same iterate is just what Halley’s method (see [3, 8, 19] for the well-posed setting) does, which for ill-posed problems in Hilbert spaces can be formulated as follows (the coefficient iterates are now denoted by \( x_\iota^\delta \) instead of \( c_k \)):

\[ x_0^\delta = x_0, \]
for \( k = 1, 2, \ldots, \]
\[ T_k = F'(x_k^\delta); \quad \eta_k = F\left(x_k^\delta\right) - y^\delta, \]
\[ x_{k+1}^\delta = x_k^\delta - \left(T_k^* T_k + \beta_k I\right)^{-1}\left\{ T_k^* \eta_k + \beta_k \left(x_k^\delta - x_0\right)\right\}, \]
\[ S_k = T_k + \frac{1}{2}F''\left(x_k^\delta\right)\left(x_k^\delta - x_k^\delta, \cdot \right), \]
\[ x_{k+1}^\delta = x_k^\delta - \left(S_k^* S_k + \alpha_k I\right)^{-1}\left\{ S_k^* \eta_k + \alpha_k \left(x_k^\delta - x_0\right)\right\} \] (6)
with two a priori fixed sequences \((\alpha_k)_{k \in \mathbb{N}}, (\beta_k)_{k \in \mathbb{N}}\) satisfying
\[
\alpha_k \not\in 0, \quad \beta_k \not\in 0, \quad 1 \leq \frac{\alpha_k}{\alpha_{k+1}} \leq q, \quad 1 \leq \frac{\beta_k}{\beta_{k+1}} \leq q,
\]
see [16]. In here, \(F: X \rightarrow Y\) is the forward operator in the operator equation formulation
\[
F(x) = y
\]
of the coefficient identification problem and the superscript \(\delta\) indicates the presence of noise in the given data \(y^\delta\), whose deterministic level we assume to be known, i.e.,
\[
\|y - y^\delta\| \leq \delta.
\]
Here we use a fixed reference point \(x_0\), i.e., the first order version of this method (skipping the step with \(S_k\)) would be the iteratively regularized Gauss–Newton method (IRGNM), see, e.g., [1, 12, 15, 17, 20]. If we would replace \(x_0\) by the current iterate \(x_k^\delta\) in each step, we would arrive at the Levenberg–Marquardt type version of Halley’s method considered by Hettlich and Rundell in [10]. While [10, 16] concentrate on the case of \(X, Y\) being Hilbert spaces, it is often desirable to also work in Banach spaces that to not possess Hilbert space structure, such as \(L^1\) or the space of Radon measures for obtaining sparse solutions or modeling impulsive noise, or \(L^\infty\) for guaranteeing essential bounds (e.g., nonnegativity) of coefficients or modeling uniform noise, see, e.g., [5–7, 13, 14]. Method (6) can indeed be extended to the more general setting of \(X, Y\) being Banach spaces in a straightforward manner:
\[
x_k^\delta = x_0, \quad \text{for } k = 1, 2, \ldots,
\]
\[
T_k = F'(x_k^\delta); \quad n_k = F(x_k^\delta) - y^\delta,
\]
\[
x_k^\delta + \in \arg \min_{x^\delta} \frac{1}{r} \| T_k(x - x_k^\delta) + r_k \| + \frac{\beta_k}{p} \| x - x_0 \|^p,
\]
\[
S_k = T_k + \frac{1}{2} F''(x_k^\delta)(x_k^\delta - x_k^\delta, \cdot),
\]
\[
x_{k+1}^\delta \in \arg \min_{x^\delta} \frac{1}{r} \| S_k(x - x_k^\delta) + r_k \| + \frac{\alpha_k}{p} \| x - x_0 \|^p
\]
with \(p, r \in [1, \infty)\) and (7) in case \(r > 1\), as well as
\[
\alpha_k \geq \alpha, \quad \beta_k \geq \beta
\]
for some constants \(\alpha, \beta > 0\) in case \(r = 1\). The paradigm of regularization with general powers of Banach space norms, which we follow here, has been put forward in the seminal paper [9] and several subsequent publications. For the exact penalization case \(r = 1\), it has already been observed in [4] that the regularization parameters should not tend to zero.

Below we will prove a convergence result under the source condition
\[
T^\gamma_v \in J_p(x^\gamma - x_0)
\]
for some \(v \in X\). Here \(J_p = \vartheta^p\) \(\cdot \| \cdot \|^p\) denotes the duality mapping. Our analysis will make use of the shifted Bregman distance
\[
D_{p, \delta}(\tilde{x}, x) = \frac{1}{p} \| x - x_0 \|^p - \frac{1}{p} \| \tilde{x} - x_0 \|^p - \langle \delta, x - \tilde{x} \rangle\] for \(\xi \in J_p(\tilde{x} - x_0)\).
which, if \( X \) is \( p \)-convex, satisfies the coercivity estimate
\[
D_{\rho, \xi}^\alpha(\tilde{x}, x) \geq \xi \| \tilde{x} - x \|_p^p \quad \text{for all} \quad \tilde{x}, x \in X
\]  
for some constant \( \xi > 0 \) depending on \( p \) (see, e.g., [2 lemma 2.7]). The stopping index \( k^\ast \) will be the first one such that \( \alpha_{k^\ast} \leq \tau \delta \), i.e.,
\[
\alpha_{k^\ast} \leq \tau \delta < \alpha_{k^\ast - 1} \quad \forall k \in \{0, \ldots, k^\ast - 1\},
\]
if \( r > 1 \). In case \( r = 1 \) we can choose \( \beta_i \) and \( \alpha_i \) constant (i.e., the Tikhonov regularization parameter is independent of \( \delta \), as typical for this case, see, e.g., [4]) and
\[
k^\ast \geq \left\lfloor \log_\sigma \left( \frac{\log_\sigma(\delta^{-1/p})}{\log_\sigma(\delta^{-1/p})} \right) \right\rfloor,
\]
where \( \sigma = \min \{[\rho^n, p - 1], p - 1\} \), which is larger than one for \( p < 2 \).

Concerning smoothness of the forward operator, we will make the following assumption.

**Assumption 1.** Let \( F \) be twice Fréchet differentiable with \( F^\ast \) bounded
\[
\forall x \in B_p(\chi^\ast) \quad \| F^\ast(x) \| \leq C_2
\]
and Lipschitz continuous
\[
\forall x, \tilde{x} \in B_p(\chi^\ast) \quad \| F^\ast(x) - F^\ast(\tilde{x}) \| \leq L_2 \| x - \tilde{x} \|
\]
in a ball \( B_p(\chi^\ast) \) containing \( x_0 \).

We first of all consider the case \( r > 1 \).

**Theorem 1.** Let \( X, Y \) be Banach spaces with additionally \( X \) being \( p \)-convex so that (13) holds, where the exponents satisfy
\[
1 \leq p \leq 2 \quad \text{and} \quad 1 \leq r \leq p < 2r.
\]
Moreover, let a source condition (12) with \( \| v \| \) sufficiently small hold, let assumption 1 be satisfied and let \( x_0 \) be sufficiently close to \( \chi^\ast \). Assume that with \( \alpha_k \) satisfying (7), \( \beta_k \) is chosen according to
\[
\beta_k = s \alpha_k
\]
with some fixed \( s > 0 \) and \( \alpha_0 \) sufficiently small (in case \( p = 2 \) also \( s \) has to be sufficiently small), and let \( k^\ast \) be chosen according to (14) with \( \tau \) sufficiently large.

Then the iterates defined by (10) converge at the rate
\[
\| x^\ast_n - \chi^\ast \| = O(\delta^{-\beta}) \quad \text{as} \quad \delta \to 0.
\]
If \( \delta = 0 \) we have convergence
\[
\| x^\ast_k - \chi^\ast \| = O\left( \alpha_k^{-\frac{1}{p-1}} \right) \quad \text{as} \quad k \to \infty.
\]
The case $r = 1$ corresponding to exact penalization of the data misfit (see, e.g., [4]) is treated separately. Since the existing results on the IRGNM from [12, 17] do not seem to be applicable, we also prove the corresponding result for the IRGNM.

**Theorem 2.** Let $X, Y$ be Banach spaces with additionally $X$ being $p$-convex so that (13) holds, where the exponents satisfy

$$1 \leq p < 2 \text{ and } r = 1.$$  

Moreover, let a source condition (12) with $\|v\|$ sufficiently small hold, let assumption 1 be satisfied and let $x_0$ be sufficiently close to $x^\dagger$. Assume that the regularization parameters $\alpha_k, \beta_k$ are chosen to be bounded away from zero, i.e., such that (11) holds for some constants $\alpha, \beta > 0$, and let $k^*$ be chosen according to (15) with $\sigma = \frac{p+2}{p^2} > 1$.

Then the iterates defined by (10) converge at the rate

$$\|x^\delta_k - x^\dagger\| = O\left(\delta^\sigma\right) \text{ as } \delta \to 0. \quad (18)$$

If $\delta = 0$ we have convergence of order $\sigma = \frac{p+2}{p^2}$, i.e., $x^\delta_k \to x^\dagger$ as $k \to \infty$ and there exists a constant $C > 0$ such that for all $k \in \mathbb{N}$

$$\|x^\delta_{k+1} - x^\dagger\| \leq C \|x^\delta_k - x^\dagger\|^\sigma. \quad (19)$$

In particular, we have cubic convergence in case $p = 1$.

For the IRGNM defined by setting $x^\delta_{k+1} = x^\delta_k$, under the same assumptions (except those on $\alpha_k$, which are not needed, and assumption 1, which can be replaced by once Lipschitz continuous Fréchet differentiability of $F$ in $B_p(x^\dagger)$) with $k^*$ chosen according to (15) with $\sigma = \frac{2}{p} > 1$, the iterates converge at the rate (18). If $\delta = 0$, we have convergence (19) of order $\sigma = \frac{2}{p}$ for the IRGNM iterates (i.e., quadratic for $p = 1$).

**Remark 1.** The difference between first and second order IRGNM becomes even clearer here than in the Hilbert space case with quadratic penalties from [16], especially in the case $r = 1$ of theorem 2: in the exact data case, the order of convergence is always better for Halley than for IRGNM, since

$$\forall p \in [1, 2): \quad \sigma_{\text{Halley}} = \frac{p+2}{p^2} > \frac{2}{p} = \sigma_{\text{IRGNM}},$$

which becomes most obvious in the case $p = 1$, where we get cubic convergence for Halley’s method and quadratic one for the IRGNM. This faster convergence is also reflected in the number of iterates according to (15) in case of noisy data, since the logarithm in (15) is taken with respect to a larger basis for Halley than for IRGNM.

1.1. Proof of theorems 1, 2

By minimality in the definition (10) of the iterates we have

$$\frac{1}{r} \|K(x - x^\delta_k) + n_k\| + \frac{\kappa}{p} \|x - x_0\|^p \leq \frac{1}{r} \|K(x^\dagger - x^\delta_k) + n_k\| + \frac{\kappa}{p} \|x^\dagger - x_0\|^p \quad (20)$$

\footnote{Note that [17] requires $r > 1$, whereas in [12], the multiplicative source condition (12) does not contain the case (12) if $r = 1$ and the additive one (35) with theorem 4.2 does not give the desired rate in this case.}
for

\[(x, K, \kappa) \in \left\{ \left( x^k, T_k, \beta_k \right), \left( x^k_{k+1}, S_k, \alpha_k \right) \right\} \,.
\]

On the other hand, we use the fact that by definition of the Bregman distance and the source condition (12) in both cases we have

\[
\frac{\kappa}{p} \|x - x_0\|^p - \frac{\kappa}{p} \|x^\dagger - x_0\|^p = \kappa D^{\kappa_0}_{p, T_v^\kappa} (x^\dagger, x) + \kappa \left\{ T^\kappa v, x - x^\dagger \right\} \\
\geq \kappa D^{\kappa_0}_{p, T_v^\kappa} (x^\dagger, x) - \kappa \|v\| \|T (x - x^\dagger)\|.
\]  

(21)

Note that we have

\[
\|K (x - x^\delta_d) + n\| \geq \left( \|T (x - x^\dagger)\| - \|\tilde{t} - \delta\| \right)^r, \quad \|K (x^\dagger - x^\delta_d) + n\| \leq (\|t\| + \delta)^r
\]  

with Taylor remainder terms

\[
t = F (x^\dagger) - F (x^\delta_d) - K (x^\dagger - x^\delta_d),
\]

\[
\tilde{t} = F (x^\dagger) - F (x^\delta_d) - T (x^\dagger - x) - K (x - x^\delta_d) = t + (K - T) (x^\dagger - x),
\]

(to be estimated using assumption 1,) so we will correspondingly dominate the term \(\kappa \|v\| \|T (x - x^\dagger)\|\) from (21) by a small multiple of \(\|T (x - x^\dagger)\|^r\), which is obvious in case \(r = 1\) with the smallness assumption

\[
\kappa \|v\| \leq \frac{1}{2}
\]

and for \(r \in (1, \infty)\) follows from Young's inequality in the form

\[
ab \leq \epsilon a^r + C (\epsilon, r) b^{r^*}
\]  

(23)

with

\[
r^* = \frac{r}{r - 1}, \quad C (\epsilon, r) = \frac{r - 1}{\epsilon^{1/(r - 1)} r^{r^*}}
\]

setting \(\epsilon = 2^{-r'/r}, a = \|T (x - x^\dagger)\|, b = \kappa \|v\|\). Putting estimates (20)–(23) together and using the simple inequalities \((a - b)^r + b^r \geq 2^{-(r - 1)} a^r, (a + b)^r \leq 2^{r - 1} (a^r + b^r)\) we obtain two estimates of the form

\[
2^{-(r - 1)} \|T (x - x^\dagger)\|^r + \kappa D^{\kappa_0}_{p, T_v^\kappa} (x^\dagger, x) \\
\leq \kappa \|v\| \|T (x - x^\dagger)\| + \frac{1}{r} (\|t\| + \delta)^r + \frac{1}{r} (\|\tilde{t} - \delta\|) \\
\leq c^0 \|T (x - x^\dagger)\|^r + C^0 \|v\| \|\tilde{t} - \delta\|^r + \frac{1}{r} (2 \delta + \|t\| + \|\tilde{t}\|)^r,
\]

where \(0 < c^0 < 2^{-(r - 1)}\) in case \(r > 1\) and

\[
\|T (x - x^\dagger)\| + \kappa D^{\kappa_0}_{p, T_v^\kappa} (x^\dagger, x) \leq \frac{1}{2} \|T (x - x^\dagger)\| + 2 \delta + \|t\| + \|\tilde{t}\|
\]  

(24)
in case $r = 1$, hence

$$
\begin{align*}
    c^1 \| T(x - x^\dagger) \| + \kappa D_{\alpha, \beta, \gamma}^x (x^\dagger, x)
    \leq \begin{cases} 
        C^1 \| v \|^{r^*} \kappa^r + \frac{1}{r} (2\delta + \| h \| + \| \bar{h} \|) \| \bar{f} \| & \text{in case } r > 1, \\
        (2\delta + \| h \| + \| \bar{h} \|)^r & \text{in case } r = 1
    \end{cases}
\end{align*}
$$

for some constants $c^1, C^1 > 0$ depending only on $r$. Using assumption 1, the Taylor remainders $t = t_0, \bar{t} = \bar{t}_0$ if $K = K_0$, $x = x_{1,0}^\text{dagger}$ and $t = t_1, \bar{t} = \bar{t}_1$ if $K = S_k$, $x = x_{1,1}^\text{dagger}$ in (22) can be estimated as follows

$$
\| t_1 \| = \| \int_0^1 \int_0^1 \frac{1}{\sqrt{C}} \nu \theta (x^\dagger - x^\text{dagger}) \nu \, d\theta \, d\tau \| \leq \frac{1}{2} C_2 \| x^\text{dagger} - x^\dagger \|^2,
$$

$$
\| \bar{t}_1 \| = \| t_1 - \int_0^1 F^\nu (x^\text{dagger} + \theta (x^\dagger - x^\text{dagger})) \nu \, d\theta \| \leq \| t_1 \| + \| C_2 \| x^\text{dagger} - x^\dagger \| \| x^\text{dagger} - x^\dagger \|.
$$

$$
\| t_2 \| = \| \int_0^1 \int_0^1 \frac{1}{\sqrt{C}} \nu \theta (x^\dagger - x^\text{dagger}) \nu \, d\theta \, d\tau \| \leq \frac{1}{2} L_2 \| x^\text{dagger} - x^\dagger \|^2 + \frac{1}{2} C_2 \| x^\text{dagger} - x^\dagger \| \| x^\text{dagger} - x^\dagger \|.
$$

$$
\| \bar{t}_2 \| = \| t_2 - \int_0^1 \frac{1}{\sqrt{C}} \nu \theta (x^\dagger - x^\text{dagger}) \nu \, d\theta \| \leq \| t_2 \| + \frac{1}{2} (L_2 \| x^\text{dagger} - x^\dagger \|^2 + C_2 \| x^\text{dagger} - x^\dagger \| + \| x^\text{dagger} - x^\dagger \|) \| x^\text{dagger} - x^\dagger \|. \quad (25)
$$

Hence, considering first of all the case $r > 1$, we end up with

$$
\begin{align*}
    c \| T(x^\text{dagger} - x^\dagger) \| + \beta_k D_{\alpha, \beta, \gamma}^x (x^\dagger, x^\text{dagger}) \\
    \leq C \left( \| v \|^{r^*} \beta_k^x + \delta^r + \| x^\text{dagger} - x^\dagger \|^2 + \| x^\text{dagger} - x^\dagger \| \| x^\text{dagger} - x^\dagger \| \right) \quad (26)
\end{align*}
$$

and

$$
\begin{align*}
    c \| T(x^{\text{dagger}, 1} - x^\dagger) \| + \alpha_k D_{\alpha, \beta, \gamma}^x (x^\dagger, x^{\text{dagger}, 1}) \\
    \leq C \left( \| v \|^{r^*} \alpha_k^x + \delta^r + \| x^\text{dagger} - x^\dagger \|^3 + \| x^\text{dagger} - x^\dagger \| \| x^\text{dagger} - x^\dagger \| \right) \quad (27)
\end{align*}
$$

with some constants $c, C > 0$ depending only on $r, C_2, L_2$. Thus, using the coercivity estimate (13), we expect to obtain the rates...
\[
\| x_{k+1}^\xi - x^\xi \| = O\left( \alpha_k^{\frac{1}{p-1}} \right), \quad \| x_k^\zeta - x^\zeta \| = O\left( \beta_k^{\frac{1}{p-1}} \right)
\]

and hence consider the quantities
\[
\gamma_k = \frac{\| x_k^\delta - x^\delta \|}{\alpha_k^{\frac{1}{p-1}}}, \quad k \in \{0, \ldots, k_*\},
\]
\[
\Gamma_{k+1} = \frac{\| x_k^\zeta - x^\zeta \|}{\beta_k^{\frac{1}{p-1}}}, \quad k \in \{1, \ldots, k_* - 1\}, \quad \Gamma_0 = 0,
\]

for which, dividing by \( \beta_k^\xi \) and \( \alpha_k^{\zeta+1} \), respectively, in (26), (27), using (13) and the stopping rule
\[
\alpha_k \leq (\tau \delta)^{-1} < \alpha_k \beta_{k+1} \leq (\tau \delta)^{-1} < \beta_k \quad \forall k \in \{0, \ldots, k_* - 1\},
\]

(see (14)) we obtain, for \( k \leq k_* - 1 \)
\[
\Gamma_{k+1}^\rho \leq \frac{C}{\xi} \left( \| v \|^{\tau^*} + \frac{1}{\tau^*} + \alpha_k \beta_k r^{-\tau^*} \right) + \frac{\alpha_k \beta_k r^{-\tau^*}}{\xi} \left( \Gamma_{k+1}^\rho \right)
\]

and
\[
\gamma_{k+1}^\rho \leq \frac{C}{\xi} q^{\tau^*} \left( \| v \|^{\tau^*} + \frac{1}{\tau^*} + \alpha_k \beta_k r^{-\tau^*} \right) + \frac{\alpha_k \beta_k r^{-\tau^*}}{\xi} \left( \Gamma_{k+1}^\rho \right)
\]

Since the sequences \( \alpha_k, \beta_k \) tend to zero, the desired boundedness of the right-hand sides in (30), (31) imposes some restrictions to the exponents. Namely, in view of term VI in (31) we need
\[
p \leq 2
\]
and from terms I, II, IV, using the fact that by (32) \( \frac{1}{p-1} \geq \frac{2}{p} \) we infer condition
\[
ma_k \leq \beta_k \leq Ma_k^{\frac{1}{p-1}}
\]
for some \( m, M > 0 \) independent of \( k \). For instance, \( \beta_k = sa_k \) with some \( s > 0 \) is an admissible choice satisfying (33), and setting \( \tau = s^{\frac{1}{p-1}} \tau \) in (29) guarantees well-definedness of \( k_* \). On the other hand, conditions (32), (33) imply boundedness of all the terms I–VII. Therewith we end up with estimates
\[
\Gamma_{k+1}^\rho \leq a + b r_{k+1}^{\xi^*} + c q_{k+1}^{\xi^*} \Gamma_{k+1}^\rho =: \Phi(\gamma_k, \Gamma_{k+1}^\rho)
\]
and

\[ \gamma_{k+1}^\rho \leq d + \varepsilon \gamma_{k}^\rho + f_{k}^\rho \Gamma_{k+1}^\rho + \left( h_{k}^\rho \gamma_{k}^\rho + j_{k}^\rho + j_{k+1}^\rho \right) \gamma_{k+1}^\rho \]

\[ = \Phi(\gamma_k, \Gamma_{k+1}), \tag{35} \]

where with (16), we have

\[ a = \frac{C}{\varepsilon} \left( \| v \|^p + \frac{1}{\varepsilon^p} \right), \quad b = \frac{C}{\varepsilon} \alpha_0^{\frac{\rho + p + 1}{\rho}} s^{-\frac{p}{\rho}}, \]

\[ c = \frac{C}{\varepsilon} \alpha_0 \left( 1 - \frac{1}{p} \right)^{\frac{p}{\rho}} s^{-\frac{p-1}{\rho}}, \quad d = \frac{C}{\varepsilon} q^{-\frac{1}{p}} \left( \| v \|^p + \frac{1}{\varepsilon^p} \right), \]

\[ e = \frac{C}{\varepsilon} q^{-\frac{1}{p}} \alpha_0 \left( 1 - \frac{1}{p} \right)^{\frac{p}{\rho}} s^{-\frac{p-1}{\rho}}, \quad f = \frac{C}{\varepsilon} q^{-\frac{1}{p}} \alpha_0 \left( 1 - \frac{1}{p} \right)^{\frac{p}{\rho}} s^{-\frac{p-1}{\rho}}, \]

\[ h = \frac{C}{\varepsilon} \alpha_0 \left( 1 - \frac{1}{p} \right)^{\frac{p}{\rho}} s^{-\frac{p-1}{\rho}}, \quad i = \frac{C}{\varepsilon} \alpha_0 \left( 1 - \frac{1}{p} \right)^{\frac{p}{\rho}} s^{-\frac{p-1}{\rho}}, \]

\[ j = \frac{C}{\varepsilon} \alpha_0 \left( 1 - \frac{1}{p} \right)^{\frac{p}{\rho}} s^{-\frac{p-1}{\rho}}. \]

We wish to carry out an induction proof of the claim

\[ x_k^\rho, x_{k+1}^\rho \in \mathcal{B}_\rho(x^\ast), \quad \gamma_k \leq \bar{\gamma}, \quad \Gamma_k \leq \bar{\Gamma} \quad \forall k \in \{0, \ldots, k_\ast\}, \tag{37} \]

(for all \( k \in \mathbb{N}_0 \) in case \( \delta = 0 \)) with appropriately chosen constants \( \bar{\gamma}, \bar{\Gamma} > 0 \). For this purpose, it suffices to do the induction step, since the induction beginning can be easily established by imposing the closeness condition \( \| x_k^\rho - x^\ast \| \leq \bar{\gamma} \alpha_0^{\frac{\rho - 1}{\rho}} \) and using the convention \( \Gamma_0 = 0 \), \( \| x_{k+1}^\rho - x_0 \| = 0 \). The induction step can be carried out by means of the following lemma.

**Lemma 1.** Let \( \phi, \Phi \) be defined as in (34), (35) with

\[ r \leq p \leq 2r. \]

Then there exist \( \bar{\gamma}, \bar{\Gamma} \) such that for \( d, e, f, h, j > 0 \) sufficiently small the implication

\[ \forall \gamma, \Gamma > 0 : \quad (\Gamma^\rho \leq \Phi(\bar{\gamma}, \bar{\Gamma}) \text{ and } \gamma^\rho \leq \Phi(\bar{\gamma}, \gamma, \bar{\Gamma})) \Rightarrow (\Gamma \leq \bar{\Gamma} \text{ and } \gamma \leq \bar{\gamma}) \]

holds.

**Proof.** See the appendix. \( \square \)

Note that the coefficients \( d, e, f, h, j \) according to (36) can indeed be made small: \( d \) by imposing \( \| v \| \) small and choosing \( \tau \) large; \( e, f, h, j \) by choosing \( \alpha_0 \) small and in case \( p = 2 \) also \( s \) small. Also note that the choice of all bounds in this lemma is independent of \( \delta \) and of \( k \).

Thus we have established (37). This immediately implies the claimed rate in the exact data case. The stopping rule (14) then implies the rate \( \| x_k^\rho - x^\ast \| = O(\delta^\rho) \) in case of noisy data.

Finally, we consider the special case \( r = 1 \), where in place of (26), (27), from estimates (24), (25) and (13) we get
\[ c \| T(x^k - x^\dagger) \| + \beta_k \| x^k - x^\dagger \|^p \leq C \left( \delta + \| x^k - x^\dagger \|^2 + \| x^k - x^\dagger \| \| x^k - x^\dagger \| \right) \]  
\( (38) \)

and

\[ c \| T(x^k+1 - x^\dagger) \| + \alpha_k \| x^k+1 - x^\dagger \|^p \leq C \left( \delta + \| x^k - x^\dagger \|^3 + \| x^k - x^\dagger \| \| x^k - x^\dagger \| \right) \]  
\( + \left( \| x^k - x^\dagger \|^2 + \| x^k - x^\dagger \| + \| x^k - x^\dagger \| \right) \| x^k+1 - x^\dagger \|. \]  
\( (39) \)

In case \( p > 1 \), the elementary estimate \( (23) \) with \( p \) instead of \( r \) and \( c = 1 \),

\[ a = \left( \frac{\beta k}{2c} \right)^{\frac{1}{p}} \| x^k - x^\dagger \|, \quad b = \left( \frac{2c}{\beta k} \right)^{\frac{1}{p}} \| x^k - x^\dagger \|, \quad d = \delta + \| x^k - x^\dagger \|^2 \]  
in \( (38) \) implies

\[ 2C a^p \leq C (d + ab) \leq C \left( d + a^p + C(1, p)b^{p^*} \right) \] hence \( a^p \leq d + C(1, p)b^{p^*} \)
i.e.,

\[ \| x^k - x^\dagger \|^p \leq \frac{C(1, p)}{\beta^{p^*}} \left( \frac{2C}{\xi} \right)^{p^*} \| x^k - x^\dagger \|^p + \frac{2C}{\xi \beta} \left( \delta + \| x^k - x^\dagger \|^2 \right) \]  
\( (40) \)

and similarly

\[ \| x^k+1 - x^\dagger \|^p \leq \frac{C(1, p)}{a^{p^*}} \left( \frac{2C}{\xi} \right)^{p^*} \left( \| x^k - x^\dagger \|^2 + \| x^k - x^\dagger \| + \| x^k+1 - x^\dagger \| \right) \]  
\( + \frac{2C}{\xi a} \left( \delta + \| x^k - x^\dagger \|^3 + \| x^k+1 - x^\dagger \| \| x^k - x^\dagger \| \right), \]  
\( (41) \)

where we have used \( \alpha_k \geq \alpha, \beta_k \geq \beta \). If \( p = 1 \), we have

\[ \| x^k+1 - x^\dagger \| \leq \frac{C \left( \delta + \| x^k - x^\dagger \|^2 \right)}{\xi \beta} = \frac{C}{\xi \beta} \| x^k - x^\dagger \| \]  
\( (42) \)

and

\[ \| x^k+1 - x^\dagger \| \leq \frac{C \left( \delta + \| x^k - x^\dagger \|^3 + \| x^k+1 - x^\dagger \| \| x^k - x^\dagger \| \right)}{\xi a} = \frac{C}{\xi a} \left( \| x^k+1 - x^\dagger \|^2 + \| x^k - x^\dagger \| + \| x^k+1 - x^\dagger \| \right). \]  
\( (43) \)

Inserting \( (40) \) into \( (41) \) and \( (42) \) into \( (43) \) we conclude that for all \( k \leq k^*, \), the iterates \( x^k, x^k+1 \) remain in \( \beta_p(x^\dagger) \), provided \( \| x_0 - x^\dagger \| \) is sufficiently small, and that the sequence of errors can be bounded by a sequence \( (\mu_k)_{k \in \mathbb{N}} \)

\[ \| x^k - x^\dagger \| \leq \mu_k \leq \mu, \quad k \leq k^*, \]

where \( \mu_0 = \| x_0 - x^\dagger \| \), and \( \mu \) is sufficiently small so that \( \mu \leq \rho \), \( \xi \beta - C \mu > 0 \), \( \xi a - C (\mu^2 + \bar{\mu} + \frac{C(1+\mu^2)}{\xi \beta - C \mu}) > 0 \), and \( (\mu_k)_{k \in \mathbb{N}} \) satisfies the recursion
\[ \mu_{k+1} = \hat{C}\left(\mu_k^\sigma + \delta\right), \] (44)

with \( \hat{C} \) sufficiently large and

\[
\sigma = \begin{cases} 
\frac{1}{p} \min \left\{ 2p^\sigma, p^\sigma, \left(\frac{p^\sigma}{p}\right)^2, 3, 1 + \frac{p^\sigma}{p}, 1 + \frac{2}{p} \right\} & \text{if } p > 1, \\
3 & \text{if } p = 1.
\end{cases}
\] (45)

The requirement \( \sigma > 1 \) resulting from the need of proving boundedness of \( \mu_{k+1} \) according to (44) with possibly large \( \hat{C} \) translates to the condition \( p < 2 \), in which case \( \sigma \) according to (45) becomes

\[ \sigma = \frac{p + 2}{p^2} \]

in both cases \( p > 1 \) and \( p = 1 \). Now we make use of an elementary consequence of the recursion (44).

**Lemma 2.** For any \( \hat{C} > 0, \sigma > 1, p \in [1, 2), \mu \in (0, 1) \), there exist \( \bar{\mu}_0, \bar{\delta} > 0 \) such that for any \( \delta \in [0, \bar{\delta}] \) and any \( k_\delta \in \mathbb{N} \) we have the following:

Any sequence starting with \( \mu_0 \in [0, \bar{\mu}_0] \) and satisfying (44) for all \( k \in \{1, \ldots, k_\delta - 1\} \) obeys the bound

\[ \mu_{k+1} \leq 2^{-\alpha k+1} \mu + C(\sigma)\delta^r \leq \bar{\mu} \text{ for all } k \leq k_\delta - 1, \]

where \( C(\sigma) := \sum_{m=0}^{\infty} 2^{-\alpha^{m+1}} \).

**Proof.** See the appendix. \( \square \)

So by setting \( k = k_\delta - 1 \) according to (15) we get

\[ \mu_{k_\delta} \leq (C(\sigma) + 1)\delta^r, \]

i.e., the stated convergence rate with noisy data.

If \( \delta = 0 \) then (42), (43) provides us with an estimate of the form (19), i.e., convergence order \( \sigma \).

In the same manner the respective convergence result for the IRGNM in case \( r = 1 \) can be seen: namely, since the IRGNM corresponds to setting \( x_{k+1} = x_{k}^\delta \), by (40), (42) we have

\[ \| x_{k+1}^{\text{IRGNM}} - x^r \| \leq \tilde{\mu}_{k+1}, \]

where \( \tilde{\mu}_0 = \| x_0^\delta - x^r \|, \)

\[ \tilde{\mu}_{k+1} = \hat{C}\left(\tilde{\mu}_k^\delta + \delta\right), \]

with \( \tilde{\sigma} = \frac{1}{p} \min \{p^\sigma, 2\} \). The requirement \( \tilde{\sigma} > 1 \) again translates to \( p < 2 \), which entails that in fact \( \tilde{\sigma} = \frac{2}{p} \) and lemma 2 yields the claimed result.
2. Numerical experiments

We now show results of numerical tests with a Matlab implementation of method (10) for the test example of identifying $c$ in

$$-\Delta u + \xi(u) = c, \quad \text{in } \Omega,$$

$$u = 0, \quad \text{on } \partial \Omega$$

$\Omega \subseteq \mathbb{R}^d$, from measurements $y = Cu$ of $u$, i.e., an inverse source problem for a nonlinear PDE, which is a special case of example 1 with $f(a, z) = z$, $g(c, u) = \xi(u) - c$, $h = 0$. Here $\xi$ is assumed to be a monotone and twice Lipschitz continuously differentiable function; in our computational experiments we will use the simple choice $\xi(u) = u^3$. For the forward operator $F = C \circ S$ with $S: L^1(\Omega) \to W^{2, s}(\Omega)$ and $C: W^{2, s}(\Omega) \to Z$ some linear observation operator mapping in to some Banach space $Z$, $s \geq 1$ we get twice differentiability and Lipschitz continuity of $F$ analogously to the example in section 4 of [18], as long as $s > \frac{d}{2}$ so that $W^{2, s}(\Omega)$ continuously embeds into $L^\infty(\Omega)$.

Numerical tests are here done for $C$ being just the embedding operator $W^{2, s}(\Omega) \to L^1(\Omega)$ or $W^{2, s}(\Omega) \to L^2(\Omega)$, i.e., full observations, $\Omega = (0, 1)^2$, $d = 2$, $s = 2$,

$$u(x_1, x_2) = \sin (\pi x_1) \sin (\pi x_2),$$

and correspondingly $c(x_1, x_2) = (1 + 2\pi^2) \sin (\pi x_1) \sin (\pi x_2)$, see figure 1. As a starting value we use $c_0 \equiv 1$. The regularization parameters were set to $\alpha_k = 10^{-4} 2^{-k}$, $\beta_k = 0.01 \alpha_k$, and the PDE is discretized by finite differences on a $30 \times 30$ grid.

We particularly study the situation of impulsive noise as described in [7], i.e., the data noise is generated by randomly (uniformly distributed) picking measurement points and perturbing their values by normally distributed random numbers with a variance of one. For such noise, the use of the $L^1$ norm for penalizing the data misfit is particularly appropriate as detailed, e.g., in [6, 7]. Thus we will use $Y = L^1(\Omega)$ and correspondingly $r = 1$. In preimage space for simplicity (and since we deal with a smooth source here) we stay with a quadratic Hilbert space penalty and choose $X = L^2(\Omega)$, $p = 2$. For Tikhonov regularization of linear problems in this setting a semismooth Newton method has been devised in [7], and a Matlab implementation has been made available on the authors’ homepages. We utilize this code for solving the nonsmooth minimization problems on lines 4 and 6 of (10).

A comparison of Halley’s method with the IRGNM first of all with exact data shows that indeed the speed of convergence can be considerably increased by the use of second order information, see figure 2. Note that for both methods, each iteration corresponds to one nonlinear PDE solve for evaluating $F$ and one stiffness matrix inversion for evaluating $F'(c_k)$, (and $F''(c_k)$).

To demonstrate the usefulness of using $L^1$ in place of $L^2$ in data space we carry out tests with $Y = L^1(\Omega)$, $r = 1$ as compared to $Y = L^2$, $r = 2$ (figure 3) for different amounts of impulsive noise, i.e., different numbers of corrupted data points (note that the variance of the perturbation in each of these points is equally high in all tests, namely one).

In table 1, we list the relative errors and residuals to demonstrate that the expected rate $O(\sqrt{k})$ for the errors $O(\delta)$ can indeed be reached. Here $k_*$ denotes the index of the best iterate in terms of smallness of the relative error. Note that theorems 1, 2 and the stopping rules (14), (15) do not exactly cover the case $p = 2$, $r = 1$ considered here. We observe that $k_*$ appears to be close to constant as expected from the respective limiting case in (14) and (15).
3. Conclusions and remarks

In this paper we have extended the IRGNM-Halley method from [16] to the general Banach space setting with possibly nonquadratic penalties and proven convergence rates under a particular source condition and with a priori regularization parameter choice for a variety of exponents in the data misfit and regularization terms.

More general convergence rates, including convergence without rates, have yet to be shown. Such results might be obtained using approximate or variational source conditions, see, e.g., [11]. As soon as (12) is violated, certainly stronger structural assumptions on $F$ will be needed to still establish convergence. It is not yet clear, though, how such conditions should be formulated to enable convergence proofs and still be satisfied for relevant applications. In [10] a tangential cone type condition was successfully used for proving convergence without source conditions. However, for the Levenberg–Marquardt type approach taken there, a monotonicity argument can be used, which does not apply to the IRGNM-type version considered here. In the Hilbert space setting of [16], we have proven convergence without (or with weaker) source conditions under a range invariance condition on $F'$, $F''$, (which is in some sense dual to the tangential cone condition). However, it is not yet clear how to carry out proofs under such conditions in a Banach space setting, where the classical
Hilbert space spectral calculus is not available. Even for the first order version, i.e., the original IRGNM, proofs under range invariance conditions on $F$ have yet to be done in non-Hilbert spaces.
Further research will therefore be concerned with providing answers to these open questions.

Moreover, the fact that computation also of higher order derivatives of the forward operator involves the same linearized PDE (with different right-hand sides), motivates to consider higher order \((n\) stage) versions of Halley’s method of the form
\[
\alpha_l = -\frac{\delta_l}{\|\Delta_l^n\|} = \ldots = -\frac{\delta_2}{\|\Delta_2\|} = \frac{\delta_0}{\|\Delta_0\|},
\]

for \(k = 1, 2, \ldots\),
\[
T_k^0 = 0, \quad n_0 = F(x_k^\delta) - y^\delta,
\]
for \(j = 1, \ldots n\) do ,
\[
T_k^j = T_k^{j-1} + \frac{1}{j!} F^{(j)}(x_k^\delta) \left( (x_{k+1}^\delta - x_k^\delta)^{j-1} \right),
\]
\[
x_{k+1}^\delta \in \text{argmin} \frac{1}{p} \| T_k^j (x - x_k^\delta) + n_k \| + \frac{\alpha_j}{p} \| x - x_0 \|^p.
\]

**Acknowledgments**

The author wishes to thank both referees for their detailed reports and fruitful suggestions, leading to an improved version of the manuscript. Moreover, we are much obliged to the authors of [7] for allowing us to use the Matlab implementation of their semismooth Newton method, which strongly enhanced our numerical results. Financial support by the Austrian Science Fund FWF under grant P24970 is gratefully acknowledged.

**Appendix**

**Proof.** (Lemma 1)

For any fixed tentative constant \(\bar{p} > 0\) we ask for existence of a \(\bar{\Gamma} > 0\) such that the implication
\[
\Gamma_{\bar{p}} \leq \psi \left( \bar{p}, \bar{\Gamma} \right) \Rightarrow \Gamma \leq \bar{\Gamma}
\]
(A.1)
holds. By inspection of the function \( \phi(\gamma, \cdot) : \Gamma \mapsto a + b\gamma^2 + c\gamma^3 \), it is readily checked that

\[
p \geq r \quad \text{with} \quad c\gamma^r < 1 \quad \text{in case} \quad p = r
\]
is necessary for implication (A.1) to hold (counterexamples are given by the unbounded sequences \( \Gamma_n = (c\gamma^r)^{\frac{1}{n}} \) in case \( p < r \) and \( \Gamma_n = n \) in case \( p = r, \quad c\gamma^r \geq 1 \)).

Thus from now on we assume that

\[
p \geq r \quad \text{(A.2)}
\]
and return to the inequality

\[
\Gamma^p \leq \phi(\gamma, \Gamma)
\]
to compute a resulting explicit upper estimate of \( \Gamma \) in terms of \( \gamma \) by distinction between the cases \( \Gamma \leq 1 \) and \( \Gamma > 1 \), the latter by (A.3) resulting in

\[
a + b\gamma^2 \geq \psi(\Gamma^p) = \psi(1) + \psi'(1 + \theta(\Gamma^p - 1))(\Gamma^p - 1)
\]
for some \( \theta \in [0, 1] \), with

\[
\psi(\lambda) := \lambda - c\gamma^3 \lambda^2, \quad \psi'(\lambda) = 1 - c\gamma^3 \frac{\gamma^r}{\lambda^{1-r}} \geq 1 - c\gamma^3 \gamma^r \quad \text{for} \quad \lambda \geq 1,
\]
(where we have used (A.2)), hence

\[
\Gamma^p \leq 1 + \frac{a + b\gamma^2 - 1 + c\gamma^r}{1 - c\gamma^r/p^r},
\]
provided \( c\gamma^r/p^r < 1 \), which altogether gives

\[
\Gamma \leq \left(1 + \frac{a + b\gamma^2 - 1 + c\gamma^r}{1 - c\gamma^r/p^r}\right)^{\frac{1}{r}}
\]
in either of the two cases \( \Gamma \leq / > 1 \). Inserting this into

\[
\gamma^p \leq \Phi(\gamma, \gamma, \Gamma)
\]
yields

\[
A(\gamma) \geq \Psi(\gamma^p) = \Psi(\lambda_0) + \Psi'(\lambda_0 + \theta(\Gamma^p - \lambda_0))(\gamma^p - \lambda_0) \quad \text{(A.4)}
\]
for any fixed \( \lambda_0 > 0 \) (to be chosen appropriately below) with

\[
A(\gamma) := d + e\gamma^3 + f\gamma^r \left(1 + \frac{a + b\gamma^2 - 1 + c\gamma^r}{1 - c\gamma^r/p^r}\right)^\gamma,
\]

\[
B(\gamma) := h\gamma^2 + i\gamma^r + j \left(1 + \frac{a + b\gamma^2 - 1 + c\gamma^r}{1 - c\gamma^r/p^r}\right)^\gamma,
\]

\[
\Psi(\lambda) := \lambda - B(\gamma)\lambda^2, \quad \Psi'(\lambda) = 1 - \frac{\gamma^r B(\gamma)}{p^2 \lambda^{2-r}}.
\]
Thus, similarly to above, by distinction between the cases $\gamma^p < \bar{\gamma} \geq \lambda_0$ we can estimate
\[
\gamma^p < \lambda_0 \text{ or } \lambda_0 \leq \gamma^p \leq \lambda_0 + \frac{A(\bar{\gamma}) - \lambda_0 + B(\bar{\gamma})\lambda_0}{1 - \frac{rB(\bar{\gamma})}{\bar{\gamma}^p \lambda_0^{p-1}}}.
\]

It remains to show that the right hand side of this inequality can be bounded by $\bar{\gamma}^p$, using a proper choice of $\bar{\gamma} > 0$ and $\lambda_0 > 0$. We do so by setting $\lambda_0 = \left(\frac{\bar{\gamma}}{3}\right)^p$, so that it remains to show that
\[
\gamma^p < \left(\frac{\bar{\gamma}}{3}\right)^p \text{ or }
0 \leq \gamma^p - \left(\frac{\bar{\gamma}}{3}\right)^p \leq \frac{A(\bar{\gamma}) - \left(\frac{\bar{\gamma}}{3}\right)^p + B(\bar{\gamma})\left(\frac{\bar{\gamma}}{3}\right)^p}{1 - \frac{rB(\bar{\gamma})}{\bar{\gamma}^p \lambda_0^{p-1}}} \leq (3^p - 1)\left(\frac{\bar{\gamma}}{3}\right)^p.
\]
i.e., unless $\gamma^p \leq \left(\frac{\bar{\gamma}}{3}\right)^p$ happens to hold (in which case we would already be finished)
\[
0 \leq A(\bar{\gamma}) - \left(\frac{\bar{\gamma}}{3}\right)^p + B(\bar{\gamma})\left(\frac{\bar{\gamma}}{3}\right)^p \leq (3^p - 1)\left(\frac{\bar{\gamma}}{3}\right)^p - \frac{rB(\bar{\gamma})\left(\frac{\bar{\gamma}}{3}\right)^p}{\bar{\gamma}^p\lambda_0^{p-1}}
\]
must be shown. Considering the asymptotic behavior as $\bar{\gamma} \to 0$ yields the requirement
\[
0 \leq d + e\gamma^{3\gamma} + f\gamma + \gamma^p\left(1 + a \pm O(\gamma^p)\right)^\bar{\gamma} - \left(\frac{\bar{\gamma}}{3}\right)^p
\]
\[
+ \left(h\gamma^{2\gamma} + i\gamma + j\left(1 + a \pm O(\gamma^p)\right)^\bar{\gamma}\right)\left(\frac{\bar{\gamma}}{3}\right)^p
\]
\[
\leq (3^p - 1)\left(\frac{\bar{\gamma}}{3}\right)^p - \frac{r}{\bar{\gamma}^p}\left(h\gamma^{2\gamma} + i\gamma + j\left(1 + a + O(\gamma^p)\right)^\bar{\gamma}\right)\left(\frac{\bar{\gamma}}{3}\right)^p.
\]
(A.5)

This shows that we have to decrease $d$, $e$, $f$, $h$, $j$ depending on $\gamma$, i.e., we assume that we can choose
\[
d = \frac{3}{2}\left(\frac{\bar{\gamma}}{3}\right)^p, \quad e = o(\gamma^{p-3\gamma}), \quad f = o(\bar{\gamma}^{p-\gamma}), \quad h = o(\gamma^{p-3\gamma}), \quad j = o(\gamma^{p-\gamma}).
\]

Also for the $i$ term we need $i\gamma^{2\gamma} = o(\gamma^p)$, which can be achieved by assuming
\[
p < 2\gamma
\]
(note that by $r \geq 1$ this is less restrictive than assuming $p < 2$ in order to make
\[
i = \frac{C}{\bar{\gamma}^p\alpha_0^{p-2}} \text{ small}). These choices render (A.5) an asymptotic estimate of the form
\[
0 \leq \frac{1}{2}\left(\frac{\bar{\gamma}}{3}\right)^p \pm o(\gamma^p) \leq (3^p - 1)\left(\frac{\bar{\gamma}}{3}\right)^p - o(\gamma^p)
\]
which is obviously feasible, so that the desired estimate
\[
\gamma \leq \bar{\gamma}
\]
can be achieved by choosing $\bar{\gamma}$ sufficiently small. □
Proof. (Lemma 2) For any \( l \leq k \), we have the estimate
\[
\mu_{k+1} \leq \hat{C}^{d^{(r-1)-1}} \left( \hat{C}^{d^{(r-1)-1}} - \hat{C}^{d^{(r-1)-1}} \right) \mu_{k-l} + \sum_{m=0}^{l} \hat{C}^{d^{n+1}} \left( \hat{C}^{d^{(r-1)-1}} - \hat{C}^{d^{(r-1)-1}} \right) \mu_{k-l+m} \ltimes \delta_{\alpha},
\]
which can be seen by induction with respect to \( l \) and the elementary estimate \((a + b)^{\lambda} \leq 2^{-1}a^{\lambda} + 2^{-1}b^{\lambda}\) for \( a, b \geq 0, \lambda \geq 1 \). Namely, from (44) we have
\[
\mu_{k+1} = \hat{C} \left( \left( \hat{C} \left( \mu_{k-l-1} + \delta_{\alpha} \right) \right) \right)^{d^{(r-1)}} + \hat{C} \delta_{\alpha} \leq \hat{C} 2^{d^{(r-1)-1}} \delta_{\alpha} + \hat{C} \delta_{\alpha},
\]
which is just (A.6) with \( l = 1 \). To carry out the induction step \( l \to l + 1 \) we again use (44) with \( k \) replaced by \( k - l - 1 \) in (A.6) to obtain
\[
\mu_{k+1} \leq \hat{C}^{d^{(r-1)-1}} \left( \hat{C}^{d^{(r-1)-1}} - \hat{C}^{d^{(r-1)-1}} \right) \mu_{k-l+1} \ltimes \delta_{\alpha},
\]
which completes the proof of (A.6). We now use \( l = k \) in (A.6) to conclude that
\[
\mu_{k+1} \leq \hat{C}^{d^{(r-1)-1}} \left( \hat{C}^{d^{(r-1)-1}} - \hat{C}^{d^{(r-1)-1}} \right) \mu_{0}^{d^{(r-1)}} + \sum_{m=0}^{k} \hat{C}^{d^{n+1}} \left( \hat{C}^{d^{(r-1)-1}} - \hat{C}^{d^{(r-1)-1}} \right) \mu_{k-l+m} \ltimes \delta_{\alpha},
\]
\[
\leq \hat{C}^{d^{(r-1)-1}} \left( \left( 2 \hat{C}^{d^{(r-1)}} \right) \mu_{0}^{d^{(r-1)}} + \sum_{m=0}^{k} \left( 2 \hat{C}^{d^{(r-1)}} \delta_{\alpha} \right) \right) \delta_{\alpha},
\]
\[
\leq 2^{-d^{(r-1)}} \mu + \hat{C}^{d^{(r-1)}} C \left( \delta_{\alpha} \right) \delta_{\alpha} \leq \mu \leq 1.
\]
under the smallness assumptions
\[
\mu_0 \leq \mu_0 < \min \left\{ \frac{1}{\beta C^{-\frac{1}{2}}} \right\},
\delta \leq \delta := \min \left\{ \frac{1 - 2^{-\frac{\sigma^2}{C^2}}}{C^{-\frac{1}{2}}C(\sigma)}, \frac{1}{2} \left( \frac{2C}{\beta} \right)^{-\frac{\sigma'}{C^2}} \right\},
\]

where \( C(\sigma) := \sum_{n=0}^{\infty} 2^{-n^{\sigma+1}} \).

References

[1] Bakushinsky A B 1992 The problem of the convergence of the iteratively regularized Gauss–Newton method Comput. Math. Math. Phys. 32 1353–9
[2] Bonesky T, Kazimierski K S, Maass P, Schöpfer F and Schuster T 2008 Minimization of Tikhonov functionals in Banach spaces J. Abstr. Appl. Anal. 2008 192679
[3] Brown G H Jr 1977 On Halley’s variation of Newton’s method Am. Math. Mon. 84 726–7
[4] Burger M and Osher S 2004 Convergence rates of convex variational regularization Inverse Probl. 20 1411–21
[5] Clason C 2012 \( L^p \) fitting for inverse problems with uniform noise Inverse Probl. 28 104007
[6] Clason C and Jin B 2012 A semi-smooth Newton method for nonlinear parameter identification problems with impulsive noise SIAM J. Imaging Sci. 5 505–36
[7] Clason C, Jin B and Kunisch K 2010 A semismooth Newton method for \( L^1 \) data fitting with automatic choice of regularization parameters and noise calibration SIAM J. Imaging Sci. 3 199–231
[8] Döring B 1970 Einige Sätze über das Verfahren der tangierenden Hyperbeln in Banach-Räumen Apl. Mat. 15 418–64
[9] Grasmair M, Haltmeier M and Scherzer O 2008 Sparse regularization with \( \ell_0 \) penalty term Inverse Probl. 24 055020
[10] Hettlich F and Rundell W 2000 A second degree method for nonlinear inverse problems SIAM J. Numer. Anal. 37 587–620
[11] Hein T and Hofmann B 2009 Approximate source conditions for nonlinear ill-posed problems—chances and limitations Inverse Probl. 2 035003
[12] Hohage T and Werner F 2013 Iteratively regularized Newton-type methods for general data misfit functionals and applications to Poisson data Numer. Math. 123 745–79
[13] Hohage T and Werner F 2014 Convergence rates for inverse problems with impulsive noise SIAM J. Numer. Anal. 52 1203–21
[14] Jin B and Maass P 2012 Sparsity regularization for parameter identification problems Inverse Probl. 28 123001
[15] Jin Q and Zhong M 2013 On the iteratively regularized Gauss–Newton method in Banach spaces with applications to parameter identification problems Numer. Math. 124 647–83
[16] Kaltenbacher B 2014 An iteratively regularized Gauss–Newton–Halley method for solving nonlinear ill-posed problems Numer. Math. at press doi:10.1007/s00211-014-0682-5
[17] Kaltenbacher B and Hofmann B 2010 Convergence rates for the iteratively regularized Gauss–Newton method in Banach spaces Inverse Probl. 26 035007
[18] Kaltenbacher B, Kirchner A and Vexler B 2011 Adaptive discretizations for the choice of a Tikhonov regularization parameter in nonlinear inverse problems Inverse Probl. 27 125008
[19] Ren H and Argyros I K 2012 On the semi-local convergence of Halley’s method under a center-Lipschitz condition on the second Fréchet derivative Appl. Math. Comput. 218 11488–95
[20] Werner F 2014 On convergence rates for iteratively regularized Newton-type methods under a Lipschitz-type nonlinearity condition J. Inverse Ill-Posed Probl. at press doi:10.1515/jiip-2013-0074