Methods for inverse problems: VI. Using higher derivatives: Halley's method

Barbara Kaltenbacher, University of Klagenfurt, Austria

nonlinear inverse problem

$$F(x) = y$$

nonlinear inverse problem

$$F(x) = y$$

such as parameter identification in PDE

$$A(x, u) = 0 \quad y = Cu$$

nonlinear inverse problem

$$F(x) = y$$

such as parameter identification in PDE

$$A(x, u) = 0 \quad y = Cu$$

e.g., identification of space dependent diffusion a in elliptic PDE

$$\begin{cases} -\nabla(a\nabla u) = f \text{ in } \Omega \\ u = g \text{ on } \partial\Omega \end{cases} \qquad y = u \text{ or } y = u|_{\omega} \text{ or } y = u|_{\partial\Omega}$$

nonlinear inverse problem

$$F(x) = y$$

such as parameter identification in PDE

$$A(x, u) = 0 \quad y = Cu$$

e.g., identification of space dependent diffusion a in elliptic PDE

$$\begin{cases} -\nabla(a\nabla u) = f \text{ in } \Omega \\ u = g \text{ on } \partial\Omega \end{cases} \qquad y = u \text{ or } y = u|_\omega \text{ or } y = u|_{\partial\Omega}$$

e.g., identification of space dependent potential c in elliptic PDE

$$\begin{cases} -\Delta u + cu = f \text{ in } \Omega \\ u = g \text{ on } \partial\Omega \end{cases} \qquad y = u \text{ or } y = u|_\omega \text{ or } y = u|_{\partial\Omega}$$

Newton's method relies on 1st order Taylor expansion:

$$y = F(x) \approx F(x_k) + F'(x_k)(x - x_k)$$

Newton's method relies on 1st order Taylor expansion:

$$y = F(x) \approx F(x_k) + F'(x_k)(x - x_k)$$
$$x_{k+1} = x_k + F'(x_k)^{-1}(y - F(x_k))$$

Newton's method relies on 1st order Taylor expansion:

$$y = F(x) \approx F(x_k) + F'(x_k)(x - x_k)$$
$$x_{k+1} = x_k + F'(x_k)^{-1}(y - F(x_k))$$

F smooth $\rightsquigarrow$ 2nd order Taylor expansion:

$$y = F(x) \approx F(x_k) + F'(x_k)(x - x_k) + \frac{1}{2}F''(x_k)(x - x_k)^2$$

Newton's method relies on 1st order Taylor expansion:

$$y = F(x) \approx F(x_k) + F'(x_k)(x - x_k)$$
$$x_{k+1} = x_k + F'(x_k)^{-1}(y - F(x_k))$$

F smooth $\rightsquigarrow$ 2nd order Taylor expansion:

$$y = F(x) \approx F(x_k) + F'(x_k)(x - x_k) + \tfrac{1}{2}F''(x_k)(x - x_k)^2$$
$$x_{k+1} = {\color{red}?} \qquad \text{quadratic operator equation}$$

Newton's method relies on 1st order Taylor expansion:

$$y = F(x) \approx F(x_k) + F'(x_k)(x - x_k)$$
$$x_{k+1} = x_k + F'(x_k)^{-1}(y - F(x_k))$$

F smooth $\rightsquigarrow$ 2nd order Taylor expansion:

$$y = F(x) \approx F(x_k) + F'(x_k)(x - x_k) + \tfrac{1}{2}F''(x_k)(x - x_k)^2$$
$$x_{k+1} = {\color{red}?} \qquad \text{quadratic operator equation}$$

alternative: intermediate step x_{k+} and Taylor expansion of F':

$$x_{k+} = x_k + F'(x_k)^{-1}(y - F(x_k)) \text{ Newton step}$$
$$F'(x_{k+}) \approx F'(x_k) + F''(x_k)(x_{k+} - x_k) =: S_k$$
$$x_{k+1} = x_k + S_k^{-1}(y - F(x_k)) \text{ enhanced Newton step}$$

Newton's method relies on 1st order Taylor expansion:

$$y = F(x) \approx F(x_k) + F'(x_k)(x - x_k)$$
$$x_{k+1} = x_k + F'(x_k)^{-1}(y - F(x_k))$$

F smooth $\rightsquigarrow$ 2nd order Taylor expansion:

$$y = F(x) \approx F(x_k) + F'(x_k)(x - x_k) + \frac{1}{2}F''(x_k)(x - x_k)^2$$
$$x_{k+1} = ? \qquad \text{quadratic operator equation}$$

alternative: intermediate step x_{k+} and Taylor expansion of F':

$$x_{k+} = x_k + F'(x_k)^{-1}(y - F(x_k)) \text{ Newton step}$$
$$F'(x_{k+}) \approx F'(x_k) + F''(x_k)(x_{k+} - x_k) =: S_k$$
$$x_{k+1} = x_k + S_k^{-1}(y - F(x_k)) \text{ enhanced Newton step}$$

uses F'', converges cubically in well-posed case $\rightsquigarrow$ 2nd order method

Halley's method, method of tangent hyperbolas
[Brown'77], [Döring'70], [Ren&Argyros'12] well-posed
[Hettlich&Rundell'00] ill-posed problems

Why not just do a second Newton step?

two Newton steps:

$$x_{k+} = x_k + F'(x_k)^{-1}(y - F(x_k)) \text{ 1st Newton step}$$
$$x_{k+1} = x_{k+} + F'(x_{k+})^{-1}(y - F(x_{k+})) \text{ 2nd Newton step}$$

intermediate step x_{k+} and Taylor expansion of F':

$$x_{k+} = x_k + F'(x_k)^{-1}(y - F(x_k)) \text{ 1st Newton step}$$
$$F'(x_{k+}) \approx F'(x_k) + F''(x_k)(x_{k+} - x_k, \cdot) =: S_k$$
$$x_{k+1} = x_k + S_k^{-1}(y - F(x_k)) \text{ enhanced Newton step}$$

two Newton steps:

$$x_{k+} = x_k + F'(x_k)^{-1}(y - F(x_k)) \text{ 1st Newton step}$$
$$x_{k+1} = x_{k+} + F'(x_{k+})^{-1}(y - F(x_{k+})) \text{ 2nd Newton step}$$

intermediate step x_{k+} and Taylor expansion of F':

$$x_{k+} = x_k + F'(x_k)^{-1}(y - F(x_k)) \text{ 1st Newton step}$$
$$F'(x_{k+}) \approx F'(x_k) + F''(x_k)(x_{k+} - x_k, \cdot) =: S_k$$
$$x_{k+1} = x_k + S_k^{-1}(y - F(x_k)) \text{ enhanced Newton step}$$

two Newton steps:

$$x_{k+} = x_k + F'(x_k)^{-1}(y - F(x_k)) \text{ 1st Newton step}$$
$$x_{k+1} = x_{k+} + F'(x_{k+})^{-1}(y - F(x_{k+})) \text{ 2nd Newton step}$$

intermediate step x_{k+} and Taylor expansion of F':

$$x_{k+} = x_k + F'(x_k)^{-1}(y - F(x_k)) \text{ 1st Newton step}$$
$$F'(x_{k+}) \approx F'(x_k) + F''(x_k)(x_{k+} - x_k, \cdot) =: S_k$$
$$x_{k+1} = x_k + S_k^{-1}(y - F(x_k)) \text{ enhanced Newton step}$$

After 1st Newton step $F''(x_k)(x_{k+} - x_k, \cdot)$ is cheaper to evaulate than $F'(x_{k+})$, $F(x_{k+})$!

e.g., parameter identification in PDEs:

$$F(x_k) = C\mathcal{S}(x_k) \text{ with } \mathcal{S}(x_k) = u \text{ satisfying } A(x_k, u) = 0$$

derivatives:

$$\begin{aligned} F'(x_k)h &= C\tilde{u}(h) & &\text{with } \tilde{u}(h) \text{ satisfying } A_u(x_k, u)\tilde{u} = f_1(h) \\ F''(x_k)(x_{k+} - x_k, h) &= C\tilde{\tilde{u}}(h) & &\text{with } \tilde{\tilde{u}}(h) \text{ satisfying } A_u(x_k, u)\tilde{\tilde{u}} = f_2(h) \end{aligned}$$

with

$$\begin{aligned} f_1(h) = &- A_x(x_k, u)h \\ f_2(h) = &- A_{xx}(x_k, u)(x_{k+} - x_k, h) - A_{xu}(x, u)(h, \tilde{u}(x_{k+} - x_k)) \\ &- A_{ux}(x, u)(\tilde{u}(x_{k+} - x_k), h) - A_{uu}(x, u)(\tilde{u}(x_{k+} - x_k), \tilde{u}) \end{aligned}$$

Same PDE with same parameter for $F'(x_k)$ and $F''(x_k)(x_{k+} - x_k)$ with different right hand sides e.g., identification of diffusion in elliptic PDE

$$\begin{aligned} -\nabla(a\nabla u) &= f \quad \text{in } \Omega \\ u &= g \quad \text{on } \partial\Omega \end{aligned}$$

from measurements $y = Cu$ of u.

$F'(a)h = C\mathcal{S}'(a)h$, $F''(a)(h, l) = C\mathcal{S}''(a)(h, \ell)$
with $\tilde{u} = \mathcal{S}'(a)h$, $\tilde{\tilde{u}} = \mathcal{S}''(a)(h, \ell)$ defined by

$$\begin{aligned} -\nabla(a\nabla \tilde{u}) &= \nabla(h\nabla\mathcal{S}(a)) \quad \text{in } \Omega \\ \tilde{u} &= 0 \quad \text{on } \partial\Omega \end{aligned}$$

$$\begin{aligned} -\nabla(a\nabla \tilde{\tilde{u}}) &= \nabla(h\nabla\mathcal{S}'(a)\ell) + \nabla(\ell\nabla\mathcal{S}'(a)h) \quad \text{in } \Omega \\ \tilde{\tilde{u}} &= 0 \quad \text{on } \partial\Omega \end{aligned}$$

Systems for $F(a)$, $F'(a)$, $F''(a)$ contain the same stiffness matrix.
Evaluation of $F'(a_+)$ would require new stiffness matrix.

e.g., identification of potential in elliptic PDE

$$\begin{aligned} -\Delta u + cu &= f \quad \text{in } \Omega \\ u &= g \quad \text{on } \partial\Omega \end{aligned}$$

from measurements $y = Cu$ of u.

$F'(c)h = C\mathcal{S}'(c)h$, $F''(c)(h,\ell) = C\mathcal{S}''(c)(h,\ell)$
with $\tilde{u} = \mathcal{S}'(c)h$, $\tilde{\tilde{u}} = \mathcal{S}''(c)(h,\ell)$

$$\begin{aligned} -\Delta \tilde{u} + c\tilde{u} &= h\mathcal{S}(c) \quad \text{in } \Omega \\ \tilde{u} &= 0 \quad \text{on } \partial\Omega \end{aligned}$$

$$\begin{aligned} -\Delta \tilde{\tilde{u}} + c\tilde{\tilde{u}} &= h\mathcal{S}'(c)\ell + \ell\mathcal{S}'(c)h \quad \text{in } \Omega \\ \tilde{\tilde{u}} &= 0 \quad \text{on } \partial\Omega \end{aligned}$$

Systems for $F(c)$, $F'(c)$, $F''(c)$ contain the same stiffness matrix.
Evaluation of $F'(c_+)$ would require new stiffness matrix.

Halley's method for ill-posed problems

Halley's method for ill-posed problems in Hilbert spaces

$$T_k = F'(x_k^\delta); \quad r_k = F(x_k^\delta) - y^\delta$$
$$x_{k+}^\delta = x_k^\delta - (T_k^* T_k + \beta_k I)^{-1}\{T_k^* r_k + \beta_k(x_k^\delta - x_0)\}$$
$$S_k = T_k + \frac{1}{2}F''(x_k^\delta)(x_{k+}^\delta - x_k^\delta, \cdot)$$
$$x_{k+1}^\delta = x_k - (S_k^* S_k + \alpha_k I)^{-1}\{S_k^* r_k + \alpha_k(x_k^\delta - x_0)\}$$

with a priori fixed sequences of regularization parameters $(\alpha_k)_{k\in\mathbb{N}}$, $(\beta_k)_{k\in\mathbb{N}}$ satisfying

$$\alpha_k \searrow 0\,, \quad \beta_k \searrow 0\,, \quad 1 \le \frac{\alpha_k}{\alpha_{k+1}} \le q\,, \quad 1 \le \frac{\beta_k}{\beta_{k+1}} \le q\,,$$

and a priori stopping rule $k_* = k_*(\delta)$ depending on noise level

$$\delta \ge \|y^\delta - y\|\,.$$

[Hettlich&Rundell'00]: Levenberg-Marquardt type
a posteriori regularization parameter choice,
Hilbert space setting, convergence without rates

Halley's method for ill-posed problems in Banach spaces

$$T_k = F'(x_k^\delta); \quad r_k = F(x_k^\delta) - y^\delta$$
$$x_{k+}^\delta \in \text{argmin}_x \frac{1}{\mathrm{r}} \| T_k(x - x_k^\delta) + r_k\|^{\mathrm{r}} + \frac{\beta_k}{\mathrm{p}}\|x - x_0\|^{\mathrm{p}}$$
$$S_k = T_k + \frac{1}{2} F''(x_k^\delta)(x_{k+}^\delta - x_k^\delta, \cdot)$$
$$x_{k+1}^\delta \in \text{argmin}_x \frac{1}{\mathrm{r}} \| S_k(x - x_k^\delta) + r_k\|^{\mathrm{r}} + \frac{\alpha_k}{\mathrm{p}}\|x - x_k^\delta\|^{\mathrm{p}}$$

with $\mathrm{p}, \mathrm{r} \in [1, \infty)$
a priori fixed sequences of regularization parameters $(\alpha_k)_{k\in\mathbb{N}}$, $(\beta_k)_{k\in\mathbb{N}}$ satisfying

$$\alpha_k \searrow 0\,, \quad \beta_k \searrow 0\,, \quad 1 \le \frac{\alpha_k}{\alpha_{k+1}} \le q\,, \quad 1 \le \frac{\beta_k}{\beta_{k+1}} \le q\,, \qquad (1)$$

and a priori stopping rule $k_* = k_*(\delta)$ depending on noise level

$$\delta \ge \|y^\delta - y\|\,.$$

convergence results

Theorem

Assume that a source condition

$$x^\dagger - x_0 = (T^*T)^\mu v$$

with $\mu \in [\frac{1}{2}, 1]$, *and* $\|v\|$ *sufficiently small holds.*
Let F'' *be bounded and Lipschitz continuous in a neighborhood of* $x^\dagger$ *and let* x_0 *be suffiently close to* $x^\dagger$.
Assume that $\beta_k = \alpha_k$ *is chosen so that* (1) *holds and let* k_* *be chosen as the first index such that*

$$\alpha_{k_*}^{\mu+\frac{1}{2}} \leq \tau\delta$$

with τ *sufficiently large.*
Then

$$\|x_{k_*}^\delta - x^\dagger\| = O(\delta^{\frac{2\mu}{2\mu+1}}) \text{ as } \delta \to 0 \,.$$

If $\delta = 0$

$$\|x_k^\delta - x^\dagger\| = O(\alpha_k^\mu) \text{ as } k \to \infty \,.$$

Convergence with weak or no regularity of $x^\dagger - x_0$

source condition

$$x^\dagger - x_0 = f(T^*T)v \tag{2}$$

with $f : (0,\infty) \to (0,\infty)$ continuous and strictly increasing with

$$f(\lambda) \to 0\,, \quad \frac{\lambda}{f(\lambda)} \le C \text{ as } \lambda \to 0\,, \quad \mu_f := \sup_{\alpha\in(0,\alpha_0]} \frac{f'(\alpha)\alpha}{f(\alpha)} < \infty \tag{3}$$

e.g., $f(\lambda) = \log(1/\lambda)^p$, $f(\lambda) = \lambda^\mu$, $\mu \le 1$
or no source condition

$$x^\dagger - x_0 \in \mathcal{N}(T)^\perp \tag{4}$$

a priori stopping rule: k_* is chosen as the first index such that

$$\sqrt{\alpha_{k_*}} f(\alpha_{k_*}) \le \tau\delta$$

with τ sufficiently large and – if we only have (know) (4) – just

$$k_* \to \infty \quad \frac{\delta}{\sqrt{\alpha_{k_*}}} \to 0 \text{ as } \delta \to 0 \tag{5}$$

Structural condition on F: range invariance condition

$$F'(\tilde{x}) = F'(x)R(x,\tilde{x}) \text{ with}$$
$$R(x,\tilde{x}) \in \mathcal{L}(X,X)\,,\ \|R(\tilde{x},x) - I\| \leq M_1\|\tilde{x} - x\| \quad \forall x,\tilde{x} \in \mathcal{B}_\rho(x^\dagger)$$
$$F''(x)(h,l) = F'(x)R_2^x(h,l) \text{ with}$$
$$R_2^x(h,\cdot)\,,\ R_2^x(\cdot,l) \in \mathcal{L}(X,X)\,,\ \|R_2^x\| \leq M_2 \quad \forall x \in \mathcal{B}_\rho(x^\dagger)\,,\ h,l \in X \tag{6}$$

e.g., diffusion identification from (complete, partial, boundary) measurements of u in 1-d
e.g., potential identification from (complete, partial, boundary) measurements of u in 3-d

Theorem

Let F satisfy the range invariance condition (6) and let x_0 be suffently close to $x^\dagger$ and satisfy (4). Assume that $\beta_k = \alpha_k$, is chosen so that (1) holds and let k_* be chosen according to (5). Then the iterates $x_{k_*}^\delta$ converge to $x^\dagger$ as $\delta \to 0$.
If a source condition (2) with (3) is satisfied, then the rate

$$\|x_{k_*}^\delta - x^\dagger\| = O\left(f(\Theta^{-1}(\delta))\right) = O\left(\frac{\delta}{\sqrt{\Theta^{-1}(\delta)}}\right) \text{ as } \delta \to 0\,.$$

is obtained, where $\Theta(\lambda) := f(\lambda)\sqrt{\lambda}$. If $\delta = 0$ we have convergence

$$\|x_k^\delta - x^\dagger\| = O(f(\alpha_k)) \text{ as } k \to \infty\,.$$

numerical results

Test problem: Potential identification

Identify c in

$$\begin{aligned} -\Delta u + \Phi(c)u &= f \quad \text{in } \Omega \\ u &= g \quad \text{on } \partial\Omega \end{aligned}$$

from measurements $y = Cu$ *of* u,

where $\Phi(\lambda) = \frac{1}{2}\lambda^2 \mathbb{1}_{[-\bar{c},\bar{c}]} + \frac{1}{2}\bar{c}(2|\lambda| - \bar{c})\mathbb{1}_{\mathbb{R}\setminus[-\bar{c},\bar{c}]}$, so that

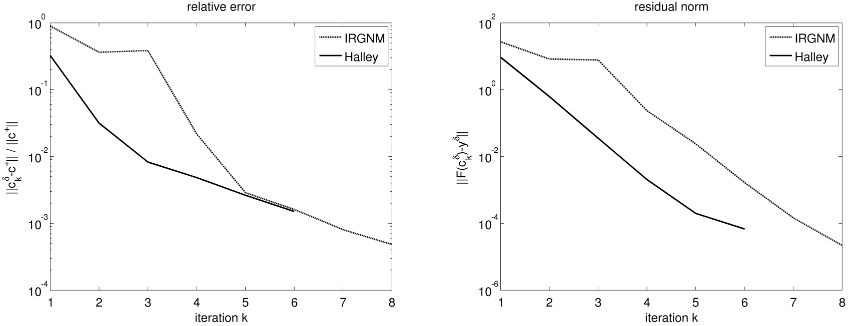
$$c \in L^2(\Omega) \quad \Rightarrow \quad \Phi(c) \geq 0 \text{ and } \Phi(c) \in L^2(\Omega)$$

$\Omega = (0,1)^2$

$$c(x_1, x_2) = 1 + \frac{1}{2}\xi(1 - \cos(4\pi x_1))(1 - \cos(4\pi x_2))\mathbb{1}_{(0,\frac{1}{2})^2}$$

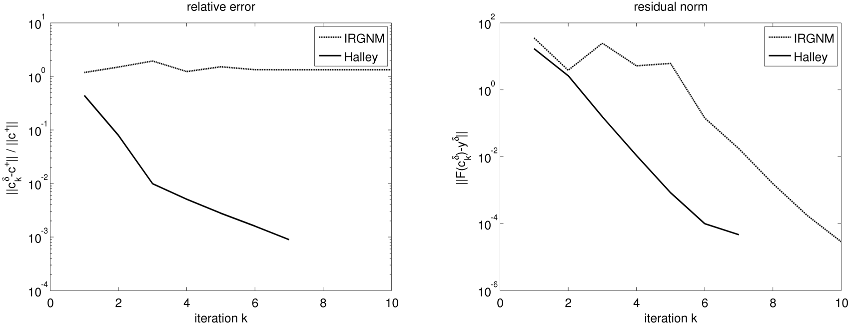
with $\xi \in \{5, 7, 10\}$,
starting value $c_0 \equiv 1$.

Comparison of IRGNM (dashed) and Halley (solid)



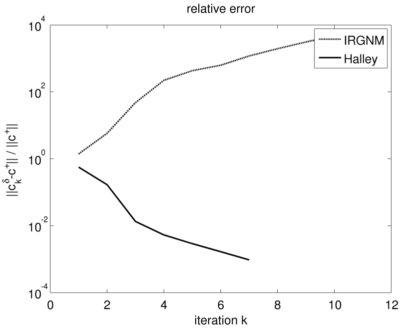
relative error (left) and residual (right) for $\xi = 5$

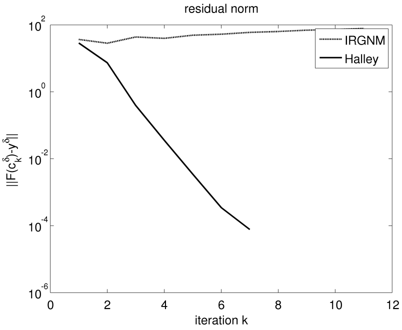
Comparison of IRGNM (dashed) and Halley (solid)



relative error (left) and residual (right) for $\xi = 7$

Comparison of IRGNM (dashed) and Halley (solid)

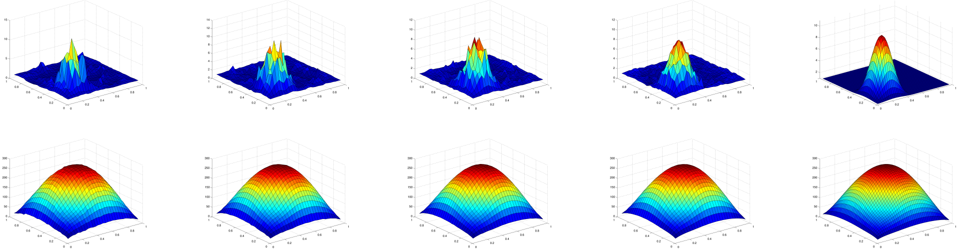




relative error (left) and residual (right) for $\xi = 10$

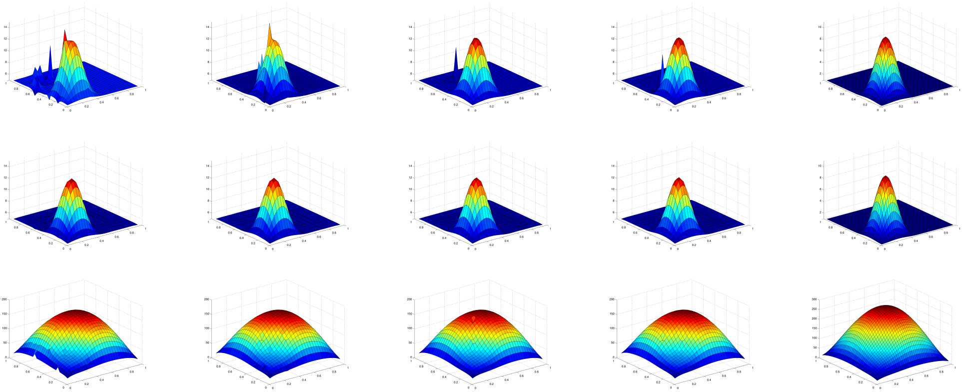
Reconstructions (top) from data (bottom) with Gaussian noise of decreasing level

$\delta = 1\%$ $\delta = 0.5\%$ $\delta = 0.25\%$ $\delta = 0.125\%$ $\delta = 0\%$



$\xi = 10$

Reconstructions with $Y = L^2$ (top) and $Y = L^{1.1}$ (middle) from data (bottom) with impulsive noise of decreasing amount from left to right



Conclusions and Outlook

- higher order methods seem to pay off in parameter identification for PDEs
- existing analysis:
  - convergence (rates) in high and low regularity (source condition) case
  - convergence rate under benchmark source condition in Banach spaces

→ several open questions in analysis (rates with a posteriori regularization parameter choice, general rates in Banach spaces,...)

→ n stage versions of Halley's method

$T_k^0 = 0 \quad r_k = F(x_k^\delta) - y^\delta$

for $j = 1, \ldots n$ do

$$T_k^j = T_k^{j-1} + \sum_{m=1}^{j} \frac{1}{m!} F^{(m)}(x_k^\delta)((x^\delta_{k+\frac{m-1}{n}} - x_k^\delta)^{m-1}, \cdot)$$

$$x^\delta_{k+\frac{j}{n}} = x_k^\delta - (T_k^{j*} T_k^j + \alpha_k^j I)^{-1} \left\{ T_k^{j*} r_k + \alpha_k^j (x_k^\delta - x_0) \right\}$$