Regularization methods for nonlinear ill-posed problems

Federica Pes, Giuseppe Rodriguez

University of Cagliari, Italy

Due Giorni di Algebra Lineare Numerica
Roma - 18-19 Febbraio 2019
Nonlinear least squares problem

Let $F(x)$ be a nonlinear Frechét differentiable function

$$F(x) = \begin{bmatrix} F_1(x) \\ F_2(x) \\ \vdots \\ F_m(x) \end{bmatrix} \in \mathbb{R}^m, \quad x \in \mathbb{R}^n.$$  

For a given $b \in \mathbb{R}^m$ we want to solve the least squares data fitting problem

$$\min_x \|r(x)\|^2, \quad r(x) = F(x) - b,$$

where $\| \cdot \|$ denotes the Euclidean norm.
The Gauss-Newton method

Chosen an initial point $x^{(0)}$, we consider the iterative method

$$x^{(k+1)} = x^{(k)} + s^{(k)}$$

where the step $s^{(k)}$ is computed minimizing, at each step, the linearization of the residual

$$\| r(x^{(k+1)}) \|^2 \simeq \| r(x^{(k)}) + J(x^{(k)})s \|^2,$$

where $J(x^{(k)})$ is the evaluation of the Jacobian matrix of $r(x)$ at the point $x^{(k)}$

$$J(x^{(k)})_{ij} = \frac{\partial r_i}{\partial x_j}(x^{(k)}), \quad i = 1, \ldots, m, \ j = 1, \ldots, n.$$ 

So, $s^{(k)}$ is computed as a solution to the linear least squares problem

$$\min_s \| r(x^{(k)}) + J(x^{(k)})s \|^2.$$
The iteration of the damped Gauss–Newton method is

\[ x^{(k+1)} = x^{(k)} + \alpha^{(k)} s^{(k)} \]

where the scalar \( \alpha^{(k)} \) is a step length.

To choose it, we can use the Armijo-Goldstein principle, which selects \( \alpha^{(k)} \) as the largest number in the sequence \( 2^{-i}, \ i = 0, 1, ... \), for which the following inequality holds

\[ \| r(x^{(k)}) \|^2 - \| r(x^{(k)} + \alpha^{(k)} s^{(k)}) \|^2 \geq \frac{1}{2} \alpha^{(k)} \| J(x^{(k)}) s^{(k)} \|^2. \]
When \( \min(m, \text{rank}(J)) < n \), the solution of \( \min_s \| r(x^{(k)}) + J(x^{(k)})s \|^2 \) is not unique.

To make it unique, the new iterate \( x^{(k+1)} \) can be obtained by solving the minimal norm least squares problem

\[
\begin{align*}
\min_s & \quad \|s\|^2 \\
\text{s. t.} & \quad \min_s \| J(x^{(k)})s + r(x^{(k)}) \|^2.
\end{align*}
\]

The nonlinear function \( F(x) \) is considered ill-conditioned in a domain \( \mathcal{D} \subset \mathbb{R}^n \) when the condition number \( \kappa(J) \) of the Jacobian matrix \( J = J(x) \) is very large for any \( x \in \mathcal{D} \).

In this situation, it is common to apply a regularization method to each step of the Gauss–Newton method.
A classical approach is Tikhonov regularization, which consists of minimizing the functional

$$\|J(x^{(k)})s + r(x^{(k)})\|^2 + \lambda^2 \|s\|^2$$

for a fixed value of the parameter $\lambda > 0$.

In the following we denote $J^{(k)} = J(x^{(k)})$, $r^{(k)} = r(x^{(k)})$. 
In
\[ \min_{\mathbf{s}} \{ \| J^{(k)} \mathbf{s} + r^{(k)} \|^2 + \lambda^2 \| \mathbf{s} \|^2 \} \]
the term \( \| \mathbf{s} \|^2 \) is often substituted by \( \| L \mathbf{s} \|^2 \),
where \( L \in \mathbb{R}^{q \times n} (q \leq n) \) is a regularization matrix which incorporates available a priori information on the solution.

It is important to remark that it imposes some kind of regularity on the update vector \( \mathbf{s} \) for the solution \( \mathbf{x}^{(k)} \), and not on the solution itself.

We will explore which is the consequence of imposing a regularity constraint directly on the solution of the problem
\[ \min_{\mathbf{x}} \| \mathbf{r}(\mathbf{x}) \|^2, \quad \mathbf{r}(\mathbf{x}) = \mathbf{F}(\mathbf{x}) - \mathbf{b}. \]
Nonlinear Tikhonov regularization

We add a regularizing term to the least squares problem
\[ \min_x \| F(x) - b \|^2, \]
turning it to the minimization of the nonlinear Tikhonov functional
\[ \min_x \{ \| F(x) - b \|^2 + \lambda^2 \| Lx \|^2 \}. \]

Linearizing it we get
\[ \min_s \{ \| J^{(k)} s + r^{(k)} \|^2 + \lambda^2 \| L(x^{(k)} + s) \|^2 \}. \]
Nonlinear Tikhonov regularization

We compare

$$\min_s \{ \| Js + r^{(k)} \|^2 + \lambda^2 \| Ls \|^2 \}$$

$$\min_s \{ \| Js + r^{(k)} \|^2 + \lambda^2 \| L(x^{(k)} + s) \|^2 \}$$

Normal equations:

$$\begin{align*}
(J^T J + \lambda^2 L^T L) s &= -J^T r^{(k)} \\
(J^T J + \lambda^2 L^T L) s &= -J^T r^{(k)} - \lambda^2 L^T L x^{(k)}
\end{align*}$$

We analyze the case $L = I_n$. 
Nonlinear Tikhonov regularization

By using the SVD of $J = U\Sigma V^T$, assuming $\text{rank}(J) = p$, the normal equations become, respectively

$$(\Sigma^T \Sigma + \lambda^2 I_n) y = -\Sigma^T c^{(k)}$$

$$(\Sigma^T \Sigma + \lambda^2 I_n) y = -\Sigma^T c^{(k)} - \lambda^2 z^{(k)}$$

with $y = V^T s$, $c^{(k)} = U^T r^{(k)}$, $z^{(k)} = V^T x^{(k)}$.

The solution of the diagonal normal equations

$$y_i = \begin{cases} -\frac{\sigma_i c_i^{(k)}}{\sigma_i^2 + \lambda^2} & \text{if } i = 1, \ldots, p \\ 0 & \text{if } i = p + 1, \ldots, n \end{cases}$$

$$y_i = \begin{cases} -\frac{\sigma_i c_i^{(k)} + \lambda^2 z_i^{(k)}}{\sigma_i^2 + \lambda^2} & \text{if } i = 1, \ldots, p \\ -z_i^{(k)} & \text{if } i = p + 1, \ldots, n \end{cases}$$
The resulting iterations for the two different approaches are

\[
x^{(k+1)} = x^{(k)} - \sum_{i=1}^{p} \frac{\sigma_i c_i^{(k)}}{\sigma_i^2 + \lambda^2} v_i
\]

\[
x^{(k+1)} = x^{(k)} - \sum_{i=1}^{p} \sigma_i c_i^{(k)} + \frac{\lambda^2 z_i^{(k)}}{\sigma_i^2 + \lambda^2} v_i - V_2 V_2^T x^{(k)}
\]

where \( V_2 = [v_{p+1}, \ldots, v_n] \).
The following parameters
- orientation (vertical/horizontal)
- height $h$ over the ground
- angular frequency $\omega = 2\pi f$
- inter-coil distance $\rho$

can be varied in order to generate multiple measurements and realize data inversion, that is, approximate $\sigma(z)$ and/or $\mu(z)$. 
We assume the soil has a layered structure.

For each layer $(k = 1, \ldots, n)$:
- depth $z_k$
- width $d_k$
- conductivity $\sigma_k$
- permeability $\mu_k$
We generate synthetic measurements corresponding to the following device/configuration:

**Geophex GEM-2** (single-coil, multi-frequency)
- \( \rho = 1.66 \, m \),
- \( f = 775, 1.175, 3.925, 9.825, 21.725 \, KHz \),
- \( h = 0.75, 1.5 \, m \)
- orientation: vertical - horizontal

\[ \implies \text{20 measurements} \]

**Model:**
- \( \sigma(z) = e^{-(z-1.2)^2} \), \( \mu(z) = \mu_0 = 4\pi10^{-7} \, H/m \)
- 20 layers, noise \( 10^{-3} \), \( L = D_2 \)
Inversion of the nonlinear model

We consider the residual vector

\[ r(\sigma) = F(\sigma) - b, \]

with \( F(\sigma) = M(\sigma; \mu_0, h, \omega, \rho) \), \( h = (h_1, \ldots, h_{m_h}) \), \( \omega = (\omega_1, \ldots, \omega_{m_\omega}) \), \( \rho = (\rho_1, \ldots, \rho_{m_\rho}) \), as a function of the conductivities \( \sigma_i, i = 1, \ldots, n \); \( b \) is a vector containing the sensed data.

We perform a nonlinear least squares fitting

\[ \min_{\sigma \in \mathbb{R}^n} \| r(\sigma) \|^2. \]
The solution is computed by following the two different approaches to regularization.

1. \[
\min_{s} \left\{ \| J_k s + r(\sigma_k) \|^2 + \lambda^2 \| Ls \|^2 \right\}
\]

2. \[
\min_{s} \left\{ \| J_k s + r(\sigma_k) \|^2 + \lambda^2 \| L(\sigma_k + s) \|^2 \right\}
\]

We iterate the damped Gauss-Newton method \( \sigma_{k+1} = \sigma_k + \alpha^{(k)} s_k \) until
\[
\| \sigma_k - \sigma_{k-1} \| < \tau \| \sigma_k \| \quad \text{or} \quad k > K_{\text{max}}.
\]
Tikhonov regularization

- **Best solution**
  - $\|Ls\|^2 = 5.46 \times 10^{-4}$
  - $\|L(k) + s\|^2 = 1.27 \times 10^{-4}$

- **Discrepancy criterion**
  - $\|Ls\|^2 = 4.83 \times 10^{-3}$
  - $\|L(k) + s\|^2 = 7.85 \times 10^{-4}$

- **L-curve criterion**
  - $\|Ls\|^2 = 3.36 \times 10^{-3}$
  - $\|L(k) + s\|^2 = 3.79 \times 10^{-4}$

Federica Pes

Regularization methods for nonlinear ill-posed problems
Tikhonov regularization

$$\min_{s} \left\{ \|J_k s + r(\sigma_k)\|^2 + \lambda^2 \|Ls\|^2 \right\}$$
Tikhonov regularization

$$\min_{\mathbf{s}} \{ \| J_k \mathbf{s} + r(\sigma_k) \|^2 + \lambda^2 \| L(\sigma_k + \mathbf{s}) \|^2 \}$$

solution best = 1.27e-04
discrepancy = 7.85e-04
L-curve = 3.79e-04
In the 2nd approach the condition $\|\sigma_k - \sigma_{k-1}\| < \tau \|\sigma_k\|$ is reached faster than the 1st approach, so less iterations of the damped Gauss–Newton method are needed.
Research directions

- Analyze the case with a regularization matrix different from the identity matrix
- Investigate the same approach to the TSVD regularization
- Apply to other nonlinear problems
- Use other norms that are different from the 2-norm
Thanks!