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 $F(\mathbf{x})$ be a nonlinear Frechét differentiable function

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

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

$$\min_{\mathbf{x}} \|\mathbf{r}(\mathbf{x})\|^2, \qquad \mathbf{r}(\mathbf{x}) = F(\mathbf{x}) - \mathbf{b},$$

where $\|\cdot\|$ denotes the Euclidean norm.

The Gauss-Newton method

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

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

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

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

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

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

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

$$\min_{\mathbf{s}} \|\mathbf{r}(\mathbf{x}^{(k)}) + J(\mathbf{x}^{(k)})\mathbf{s}\|^2.$$

The damped Gauss-Newton method

The iteration of the damped Gauss–Newton method is

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

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

$$\|\mathbf{r}(\mathbf{x}^{(k)})\|^2 - \|\mathbf{r}(\mathbf{x}^{(k)} + \alpha^{(k)}\mathbf{s}^{(k)})\|^2 \geqslant \frac{1}{2}\alpha^{(k)}\|J(\mathbf{x}^{(k)})\mathbf{s}^{(k)}\|^2.$$

Minimal norm least squares / regularization

When $\min(m, \operatorname{rank}(J)) < n$, the solution of $\min_{\mathbf{s}} \|\mathbf{r}(\mathbf{x}^{(k)}) + J(\mathbf{x}^{(k)})\mathbf{s}\|^2$ is not unique.
To make it unique, the new iterate $\mathbf{x}^{(k+1)}$ can be obtained by solving the minimal norm least squares problem

$$\begin{cases} \min_{\mathbf{s}} \|\mathbf{s}\|^2 \\ \text{s. t. } \min_{\mathbf{s}} \|J(\mathbf{x}^{(k)})\mathbf{s} + \mathbf{r}(\mathbf{x}^{(k)})\|^2. \end{cases}$$

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

In this situation, it is common to apply a regularization method to each step of the Gauss–Newton method.

Tikhonov regularization

A classical approach is Tikhonov regularization, which consists of minimizing the functional

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

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

In the following we denote $J^{(k)} = J(\mathbf{x}^{(k)}), \quad \mathbf{r}^{(k)} = \mathbf{r}(\mathbf{x}^{(k)})$.

Tikhonov regularization

In

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

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

$$\min_{\mathbf{x}} \|\mathbf{r}(\mathbf{x})\|^2, \quad \mathbf{r}(\mathbf{x}) = F(\mathbf{x}) - \mathbf{b}.$$

Nonlinear Tikhonov regularization

We add a regularizing term to the least squares problem $\min_{\mathbf{x}} \|F(\mathbf{x}) - \mathbf{b}\|^2$, turning it to the minimization of the nonlinear Tikhonov functional

$$\min_{\mathbf{x}} \{\|F(\mathbf{x}) - \mathbf{b}\|^2 + \lambda^2 \|L\mathbf{x}\|^2\}.$$

Linearizing it we get

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

Nonlinear Tikhonov regularization

We compare

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

Normal equations:

$$(J^TJ+\lambda^2L^TL)\mathbf{s}=-J^T\mathbf{r}^{(k)} \qquad (J^TJ+\lambda^2L^TL)\mathbf{s}=-J^T\mathbf{r}^{(k)}-\lambda^2L^TL\mathbf{x}^{(k)}$$

We analyze the case $L=I_n$.

Nonlinear Tikhonov regularization

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

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

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

The solution of the diagonal normal equations

$$y_i = \begin{cases} -\dfrac{\sigma_i c_i^{(k)}}{\sigma_i^2 + \lambda^2} \\ 0 \end{cases} \qquad y_i = \begin{cases} -\dfrac{\sigma_i c_i^{(k)} + \lambda^2 z_i^{(k)}}{\sigma_i^2 + \lambda^2} & i = 1, \dots, p \\ -z_i^{(k)} & i = p+1, \dots, n \end{cases}$$

Nonlinear Tikhonov regularization

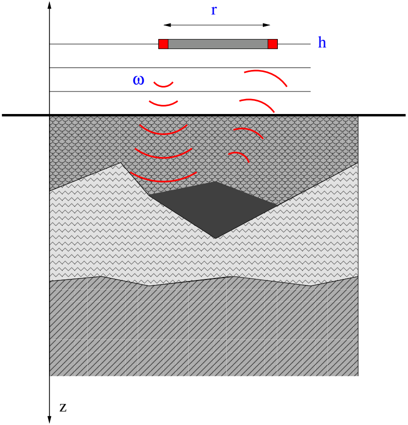
The resulting iterations for the two different approaches are

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \sum_{i=1}^{p} \frac{\sigma_i c_i^{(k)}}{\sigma_i^2 + \lambda^2} \mathbf{v}_i$$

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \sum_{i=1}^{p} \frac{\sigma_i c_i^{(k)} + \lambda^2 z_i^{(k)}}{\sigma_i^2 + \lambda^2} \mathbf{v}_i - V_2 V_2^T \mathbf{x}^{(k)}$$

where $V_2 = [\mathbf{v}_{p+1}, \dots, \mathbf{v}_n]$.

Nonlinear model



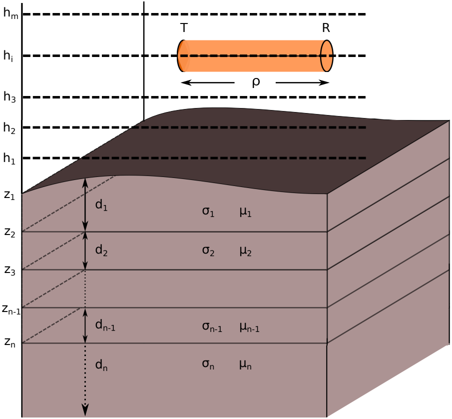
The following parameters

- orientation (vertical/horizontal)
- height h over the ground
- angular frequency $\omega = 2\pi f$
- inter-coil distance ρ

can be varied in order to generate multiple measurements and realize data inversion, that is, approximate $\sigma(z)$ and/or $\mu(z)$.

Nonlinear model

We assume the soil has a layered structure.



For each layer ($k = 1, \ldots, n$)

- depth z_k
- width d_k
- conductivity σ_k
- permeability μ_k

Nonlinear model

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

$\Longrightarrow$ 20 measurements

Model: $\sigma(z) = \mathrm{e}^{-(z-1.2)^2}$, $\mu(z) = \mu_0 = 4\pi 10^{-7}\, H/m$

20 layers, noise 10^{-3}, $L = D_2$

Inversion of the nonlinear model

We consider the residual vector

$$\mathbf{r}(\boldsymbol{\sigma}) = F(\boldsymbol{\sigma}) - \mathbf{b},$$

with $F(\boldsymbol{\sigma}) = \mathbf{M}(\boldsymbol{\sigma}; \boldsymbol{\mu}_0, \mathbf{h}, \omega, \rho)$, $\mathbf{h} = (h_1, \dots, h_{m_h})$, $\boldsymbol{\omega} = (\omega_1, \dots, \omega_{m_\omega})$, $\boldsymbol{\rho} = (\rho_1, \dots, \rho_{m_\rho})$, as a function of the conductivities σ_i, $i = 1, \dots, n$; $\mathbf{b}$ is a vector containing the sensed data.

We perform a nonlinear least squares fitting

$$\min_{\boldsymbol{\sigma} \in \mathbb{R}^n} \|\mathbf{r}(\boldsymbol{\sigma})\|^2.$$

Inversion algorithm

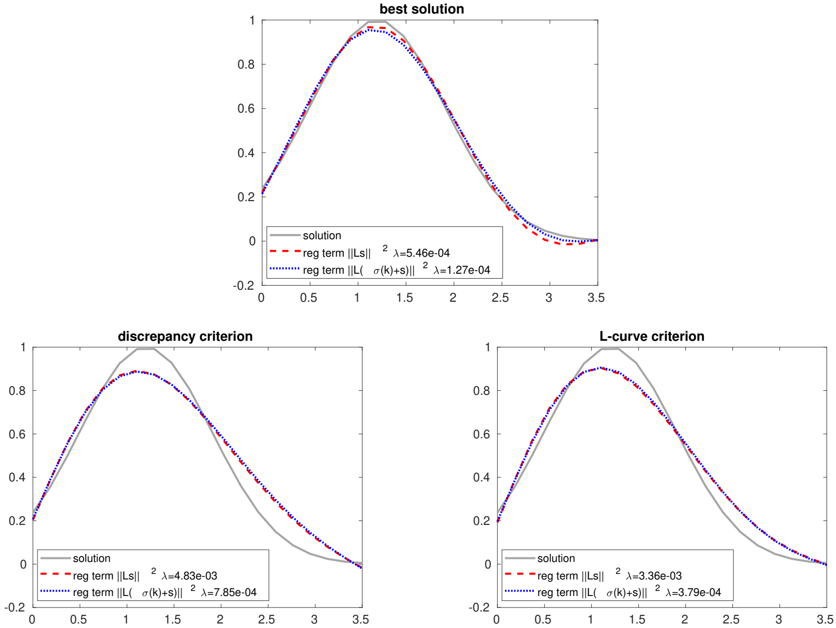
The solution is computed by following the two different approaches to regularization.

1. $\min_{\mathbf{s}}\{\|J_k\mathbf{s} + \mathbf{r}(\boldsymbol{\sigma}_k)\|^2 + \lambda^2\|L\mathbf{s}\|^2\}$
2. $\min_{\mathbf{s}}\{\|J_k\mathbf{s} + \mathbf{r}(\boldsymbol{\sigma}_k)\|^2 + \lambda^2\|L(\boldsymbol{\sigma}_k + \mathbf{s})\|^2\}$

We iterate the damped Gauss-Newton method $\boldsymbol{\sigma}_{k+1} = \boldsymbol{\sigma}_k + \alpha^{(k)}\mathbf{s}_k$ until

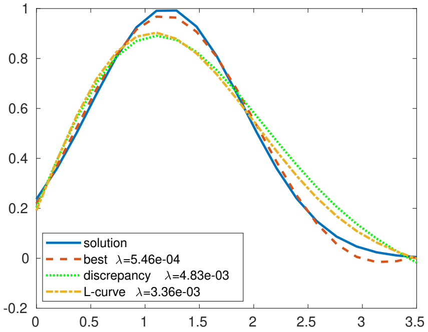
$$\|\boldsymbol{\sigma}_k - \boldsymbol{\sigma}_{k-1}\| < \tau\|\boldsymbol{\sigma}_k\| \quad \text{or} \quad k > K_{\max}.$$

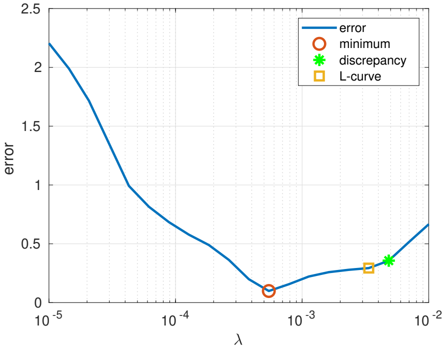
Tikhonov regularization



Tikhonov regularization

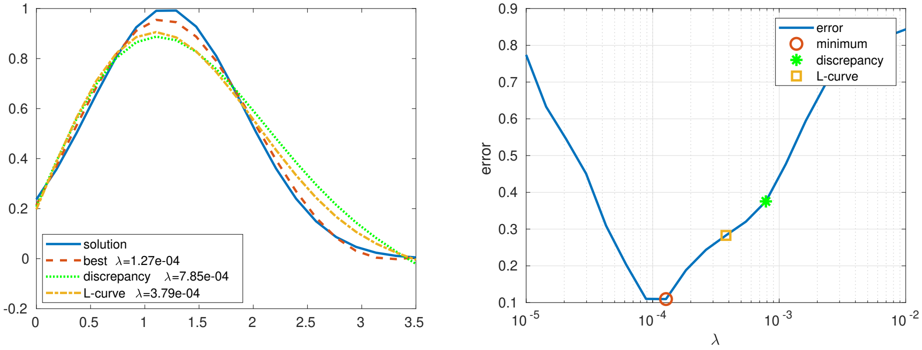
$$\min_{\mathbf{s}} \{ \| J_k \mathbf{s} + \mathbf{r}(\sigma_k) \|^2 + \lambda^2 \| L\mathbf{s} \|^2 \}$$





Tikhonov regularization

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



Observations

1. $\min_{\mathbf{s}} \{ \|J_k \mathbf{s} + \mathbf{r}(\boldsymbol{\sigma}_k)\|^2 + \lambda^2 \|L\mathbf{s}\|^2 \}$
2. $\min_{\mathbf{s}} \{ \|J_k \mathbf{s} + \mathbf{r}(\boldsymbol{\sigma}_k)\|^2 + \lambda^2 \|L(\boldsymbol{\sigma}_k + \mathbf{s})\|^2 \}$

In the 2nd approach the condition $\|\boldsymbol{\sigma}_k - \boldsymbol{\sigma}_{k-1}\| < \tau \|\boldsymbol{\sigma}_k\|$ 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!