Regularization Schemes for Linear Inverse Problems

Haley Rosso & Andreas Mang
Department of Mathematics, University of Houston, Houston, TX, USA

UNIVERSITY of HOUSTON

Introduction

Teaser: Our primary goal is the design of effective regularization schemes for inverse problems.

Mathematical Problem Formulation

In a first step, we consider linear inverse problems [1]. That is, we assume that the observed data $\boldsymbol{y}_{\text{obs}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\eta}$, where $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{A} \in \mathbb{R}^{m,n}$, $\boldsymbol{y}_{\text{obs}} \in \mathbb{R}^m$, and $\boldsymbol{\eta} \propto \mathcal{N}(0, \boldsymbol{I}_n)$ is a random perturbation. In the inverse problem, we seek $\boldsymbol{x}$ given $\boldsymbol{y}_{\text{obs}}$ and $\boldsymbol{A}$. In general, $\boldsymbol{A}$ will not be invertible and $\boldsymbol{y}_{\text{obs}} \notin \operatorname{col} \boldsymbol{A}$. Consequently, we will formulate the solution of the linear system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}_{\text{obs}}$ as a **regularized least squares problem** of the form

$$\underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}}\; f(\boldsymbol{x}), \quad \text{where } f(\boldsymbol{x}) := \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}_{\text{obs}}\|_2^2 + \frac{\alpha}{2}\|\boldsymbol{L}\boldsymbol{x}\|_2^2. \quad (1)$$

The first term of f measures the discrepancy between the model prediction $\boldsymbol{A}\boldsymbol{x} =: \boldsymbol{y}_{\text{pred}}$ and the observed data $\boldsymbol{y}_{\text{obs}}$. The second term $\|\boldsymbol{L}\boldsymbol{x}\|_2^2 = \langle \boldsymbol{L}^\mathsf{T}\boldsymbol{L}\boldsymbol{x}, \boldsymbol{x}\rangle$, $\boldsymbol{L} \in \mathbb{R}^{n,n}$, is a **regularization functional**; it is introduced to alleviate mathematical issues with the **ill-posedness** of the inverse problem, with regularization operator $\boldsymbol{L}$ and regularization parameter $\alpha > 0$. We will see that the choices for $\boldsymbol{L}$ and α greatly affect the computed solution $\boldsymbol{x}$ of (1). We consider the following regularization operators: (i) $\boldsymbol{L}^\mathsf{T}\boldsymbol{L} = \boldsymbol{I}_n$ (identity matrix), (ii) $\boldsymbol{L}^\mathsf{T}\boldsymbol{L} = -\Delta$ (Laplace operator), and (iii) $\boldsymbol{L}^\mathsf{T}\boldsymbol{L} = \boldsymbol{I}_n - \boldsymbol{V}\boldsymbol{V}^\mathsf{T}$, where $\boldsymbol{V}$ are the right-singular values of $\boldsymbol{A}$ (see below) [2].

Optimality Conditions

For an admissible solution $\boldsymbol{x}^\star \in \mathbb{R}^n$ of (1) we require that the gradient $\nabla f(\boldsymbol{x}^\star)$, $f : \mathbb{R}^n \to \mathbb{R}$ of the objective function f vanishes. The gradient of f is given by

$$\nabla f(\boldsymbol{x}) = \nabla\left(\frac{1}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2 + \frac{\alpha}{2}\|\boldsymbol{L}\boldsymbol{x}\|_2^2\right) = \boldsymbol{A}^\mathsf{T}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) + \alpha\boldsymbol{L}^\mathsf{T}\boldsymbol{L}\boldsymbol{x}.$$

Consequently, at optimality we have

$$\boldsymbol{A}^\mathsf{T}(\boldsymbol{A}\boldsymbol{x}^\star - \boldsymbol{b}) + \alpha\boldsymbol{L}^\mathsf{T}\boldsymbol{L}\boldsymbol{x}^\star \overset{!}{=} 0, \quad (2)$$

the so-called normal equation.

Numerical Methods

We consider different approaches to solve $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}_{\text{obs}}$ for $\boldsymbol{x}$. We consider two classes of approaches—direct solvers and iterative methods.

Direct Solvers

In our first approach, we consider direct solvers for computing a solution to the regularized least squares problem (1). Here, we directly solve the optimality system (2) using MATLAB's backslash command; the numerical solution is given by

$$\boldsymbol{x}_{\text{sol}} = (\boldsymbol{A}^\mathsf{T}\boldsymbol{A} + \alpha\boldsymbol{L}^\mathsf{T}\boldsymbol{L})^{-1}\boldsymbol{A}^\mathsf{T}\boldsymbol{b}.$$

In our second approach, we compute the pseudo-inverse $\boldsymbol{A}^+$ of $\boldsymbol{A}$. We use a **truncated singular value decomposition** (**TSVD**). That is, we compute the factorization $\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\mathsf{T}$ of $\boldsymbol{A}$, where $\boldsymbol{U} \in \mathbb{R}^{m,m}$ is an orthogonal matrix for the left-singular vectors, $\boldsymbol{S} \in \mathbb{R}^{m,n}$ diagonal matrix for the singular values, and $\boldsymbol{V} \in \mathbb{R}^{n,n}$ is an orthogonal matrix for the right-singular values of $\boldsymbol{A}$. Suppose that $\boldsymbol{A}$ has rank $r \ll n$. Under this assumption, $\boldsymbol{U} = [\boldsymbol{u}_1 \ldots \boldsymbol{u}_r \;\; \boldsymbol{u}_{r+1} \ldots \boldsymbol{u}_m] \in \mathbb{R}^{m,m}$, $\boldsymbol{S} = \operatorname{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0) \in R^{m,n}$, and $\boldsymbol{V} = [\boldsymbol{v}_1 \ldots \boldsymbol{v}_r \;\; \boldsymbol{v}_{r+1} \ldots \boldsymbol{v}_n] \in \mathbb{R}^{n,n}$. Consequently, we can decompose $\boldsymbol{A}$ into

$$\boldsymbol{A} \approx \boldsymbol{U}_r\boldsymbol{S}_r\boldsymbol{V}_r^\mathsf{T},$$

where $\boldsymbol{U}_r = [\boldsymbol{u}_1 \ldots \boldsymbol{u}_r] \in \mathbb{R}^{n,r}$ and $\boldsymbol{V}_r = [\boldsymbol{v}_1 \ldots \boldsymbol{v}_r] \in \mathbb{R}^{n,r}$ are the left- and right singular values and $\boldsymbol{S}_r = \operatorname{diag}(\sigma_1, \ldots, \sigma_r) \in \mathbb{R}^{r,r}$. The pseudo-inverse is given by $\boldsymbol{A}^+ = \boldsymbol{V}\boldsymbol{S}^+\boldsymbol{U}^\mathsf{T}$, where $\boldsymbol{S}^+ \in \mathbb{R}^{n,m}$ is computed by taking the reciprocal of each non-zero element on the diagonal (leaving the zeros in place) and then transposing the matrix. In numerical computation, we either select a small tolerance $\epsilon > 0$ and set all $\sigma_i \leq \epsilon$ to 0, or we select a target rank r. The choice of the target rank depends on how fast the spectrum decays.

Iterative Solvers

We use an iterative scheme of the form

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \gamma_k\boldsymbol{B}_k\nabla f(\boldsymbol{x}_k), \quad k = 1, 2, \ldots$$

Here, $k \in \mathbb{N}$ is the iteration index and $\gamma_k \in [0, 1]$ is determined using a backtracking line search [3,4]. The search direction is given by $\boldsymbol{s}_k := -\boldsymbol{B}_k\nabla f(\boldsymbol{x}_k)$. We consider a gradient descent approach for which $\boldsymbol{B}_k = \boldsymbol{I}_n$ and Newton's method with $\boldsymbol{B}_k = (\nabla^2 f(\boldsymbol{x}_k))^{-1}$, where $\nabla^2 f(\boldsymbol{x}_k) = \boldsymbol{A}^\mathsf{T}\boldsymbol{A} + \alpha\boldsymbol{L}^\mathsf{T}\boldsymbol{L}$ is the Hessian matrix. In our prototype implementation we invert the Hessian matrix using MATLAB's backslash. We terminate our solver if (i) we reach a user defined number of iterations, or (ii) the ℓ^2-norm of the gradient is reduced by a user defined tolerance $\kappa > 0$, or (iii) the ℓ^2-norm of $\nabla f(\boldsymbol{x}_k)$ is smaller or equal to 1e−6. Our numerical scheme is as follows.

```
x_0 ← 0,    k ← 0
g_0 ← objfun(x_0)
while not converged do
    s_k ← compSearchDir(objfun, x_k, g_k)
    γ_k ← doLineSearch(objfun, x_k, s_k)
    x_k ← x_k + γ_k s_k,    k ← k + 1
    converged ← checkConvergence(k, g_k, g_0)
```

Regularization Parameter Selection

To identify an optimal regularization parameter α, we consider the L-curve method [1], i.e., a log-log plot of the norm of a regularized solution versus the norm of the corresponding residual norm.

Numerical Experiments

Synthetic Test Problem

We consider a synthetic test problem to study the performance of the proposed methodology. The operators in (1) are as follows: For $\boldsymbol{A}$, we consider a Helmholtz-type operator of the general form $\boldsymbol{A} = (-\Delta + k^2\boldsymbol{I}_n)^{-1}$, where $-\Delta$ is a Laplace operator, $k > 0$, and $\boldsymbol{I}_n$ is an $n \times n$ identity matrix. The structure of the matrix $\boldsymbol{A}$ is shown in **Figure 1**.

We compute $\boldsymbol{y}_{\text{obs}}$ by applying the forward operator $\boldsymbol{A}$. That is, $\boldsymbol{y}_{\text{obs}} = \boldsymbol{A}\boldsymbol{x}_{\text{true}} + \kappa\boldsymbol{\eta}$, $\boldsymbol{\eta} \propto \mathcal{N}(0, \boldsymbol{I}_n)$, $\kappa = \theta\|\boldsymbol{x}_{\text{true}}\|_2^2$, $\theta \in [0, 1]$. We select

$$\boldsymbol{x}_{\text{true}} := (\sin(\boldsymbol{z}) + \gamma\boldsymbol{z} \odot \sin(4\boldsymbol{z})) \odot \exp(-\|\boldsymbol{z} - \pi\|_2^2/2\kappa), \quad (3)$$

with $\kappa = 9/10$ and $\gamma = 7/2$ and $z_i = hi$, $h = 2\pi/n$, $i = 1, \ldots, n$.

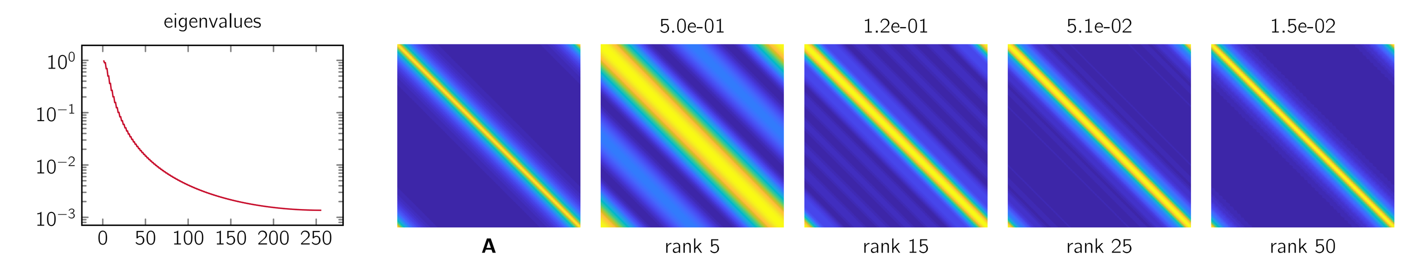

Figure 1: *Visualization of the structure of the Helmholtz matrix $\boldsymbol{A}$. We show the decay of the eigenvalues on the left. The remaining figures (from left to right) show the entries of the matrix $\boldsymbol{A}$ and low rank approximations $\boldsymbol{A}_r$ for $r \in \{5, 15, 25, 50\}$ built using TSVD. Above each visualization of the low rank approximations we report the relative error $\|\boldsymbol{A} - \boldsymbol{A}_r\|_F^2/\|\boldsymbol{A}\|_F^2$.*

Numerical Results

We report numerical results for different strategies to solve the considered inverse problem for the test problem described in the former section in **Figure 2**.

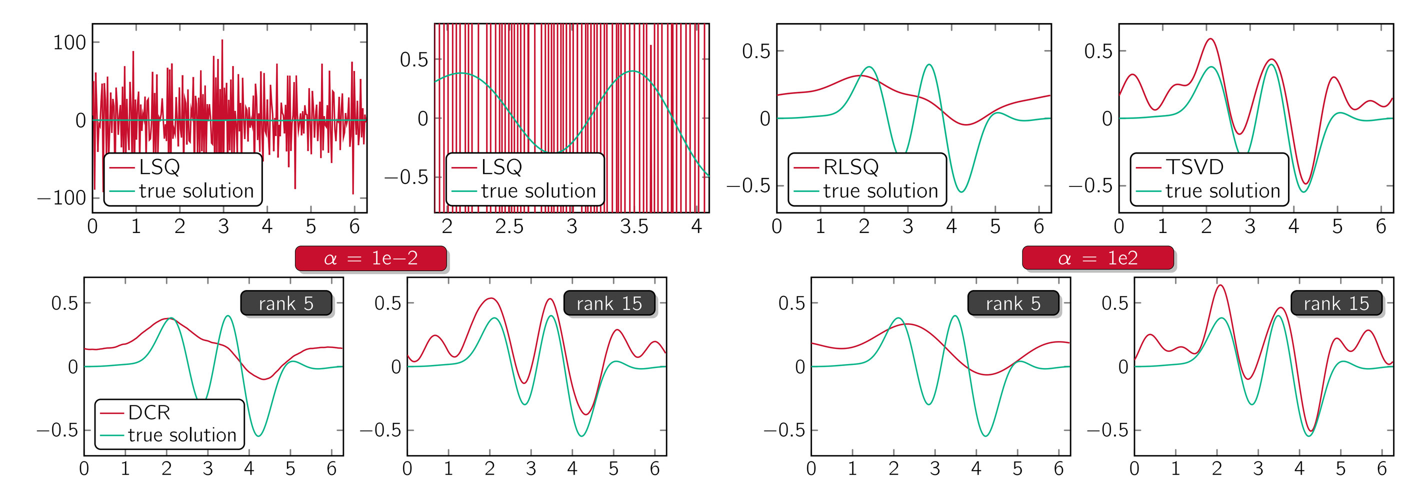

Figure 2: *We report solutions for the least squares problem in (1) for different numerical strategies. The numerical solution is shown in red and the true solution $\boldsymbol{x}_{true}$ is shown in green. Top row: For the unregularized case ($\alpha = 0$) we can observe that the noise is amplified; the computed solution has nothing to do with the true solution. For the regularized case (with regularization parameter α = 1e−2 and regularization operator $\boldsymbol{L}^\mathsf{T}\boldsymbol{L} = -\Delta$), we can see that we underfit the data. The best result is obtained by computing the solution through a low rank approximation $\boldsymbol{U}_r\boldsymbol{S}_r\boldsymbol{V}_r^\mathsf{T} \approx \boldsymbol{A}$ (truncated SVD; we consider rank $r = 15$). Bottom row: We report results for the regularization operators $\boldsymbol{L}^\mathsf{T}\boldsymbol{L} = \boldsymbol{I} - \boldsymbol{V}_r\boldsymbol{V}_r^\mathsf{T}$ with regularization parameters $\alpha \in \{$1e−2, 1e2$\}$ and ranks $r \in \{5, 15\}$.*

Conclusions

We have developed and tested a computational framework for solving and regularizing linear inverse problems [1]. We have tested a Thikonov-type regularization operator that is motivated from SVD and yields results that are consistent with the TSVD [2]. This regularization scheme outperforms standard regularization approaches based on the identity and Laplace operator. In our future work we aim at extending the proposed methodology to nonlinear inverse problems [5,6].

References

1. C. R. Vogel, Computational methods for inverse problems, SIAM, 2002.
2. J. Wittmer, B. Marin, & T. Bui-Thanh, A data-oriented statistical framework for inversion and imaging, AMS Madison, 2019.
3. J. Nocedal & S. Wright, Numerical Optimization, Springer Science, 1999.
4. A. Beck, Introduction to nonlinear optimization: Theory, algorithms, and applications with MATLAB, SIAM, 2014.
5. A. Mang & G. Biros: An inexact Newton–Krylov algorithm for constrained diffeomorphic image registration. SIAM Journal on Imaging Sciences 8(2):1030–1069, 2015.
6. A. Mang et al.: CLAIRE: A distributed-memory solver for constrained large deformation diffeomorphic image registration. SIAM Journal on Scientific Computing 41(5):C548–C584, 2019.