NTNU
Kunnskap for en bedre verden

Department of Mathematical Sciences

Heuristic Parameter Choice for Ill Posed Problems

Author:
Petter Linberg

13.01.2025

Table of Contents

List of Figures ii

List of Tables ii

1 Introduction 1

2 Background 2

2.1 Regularization methods . . . 4

2.2 Singular value decomposition . . . 4

2.2.1 Compact operators . . . 5

2.2.2 SVD of compact operators . . . 5

2.2.3 SVD for finite dimensional linear transformations . . . 6

2.3 Truncated SVD . . . 7

2.4 Tikhonov Regularization . . . 7

2.4.1 Choice of the Regularization Parameter α . . . 8

2.4.2 Solution via Normal Equations . . . 8

2.4.3 Solution via Singular Value Decomposition (SVD) . . . 9

3 Parameter choice 10

3.1 Heuristic choice rules . . . 10

3.2 L-curve method . . . 11

3.2.1 L-curve criterion from solution norm . . . 11

3.2.2 Condition L-curve criterion . . . 11

3.2.3 Maximal curvature . . . 12

3.2.4 Regińska's method . . . 13

3.2.5 Quasi-optimality criterion . . . 13

3.2.6 Generalized Cross Validation (GCV) . . . 14

4 The discrete problems 15

4.1 Poisson problem . . . 15

4.2 Deconvolution problem with Gaussian kernel . . . 16

5 Results 17

5.1 Clustering of solutions . . . 17

5.2 Heavyside signal convolved with Gaussian kernel . . . 17

5.2.1 Tikhonov regularization results for the deconvolution problem . . . 18

5.2.2 TSVD results for the deconvolution problem 20
5.3 The Poisson problem with homogeneous boundary conditions 21
5.3.1 Tikhonov regularization results for the Poisson problem 21
5.3.2 TSVD results for the Poisson problem 23

6 Discussion **24**
6.1 Insights from the Deconvolution Problem 24
6.2 Insights from the Poisson Problem 24
6.3 General Observations 24
6.4 Limitations and Future Directions 25

Bibliography **26**

List of Figures

1 Noise sensitivity example 3
2 L-curve example 0.1% noise 12
3 L-curve a example 1% noise 12
4 Max curvature example on L-curve 13
5 L-curve cluster example for the deconvolution problem 17
6 L-curve cluster example for the Poisson problem 17
7 L-curve and curvature for the deconvolution problem 18
8 L-curve for the deconvolution problem zoomed 18
9 Recovered function for the deconvolution problem 18
10 Max curvature on condition L-curve 20
11 L-curve and curvature for the Poisson problem 21
12 L-curve and curvature for the Poisson problem zoomed 21
13 Recovered function for the Poisson problem 21

List of Tables

1 L-curve choice rules for Tikhonov regularization of the deconvolution problem . . . 19
2 Other choice rules for Tikhonov regularization of deconvolution problem 19
3 L-curve choice rules for TSVD regularization of the deconvolution problem 20
4 Other choice rules for TSVD regularization of the deconvolution problem 20
5 L-curve choice rules for Tikhonov regularization of the Poisson problem 22
6 Other choice rules for Tikhonov regularization of the Poisson problem 22

7 L-curve choice rules for TSVD regularization of the Poisson problem 23
8 Other choice rules for TSVD regularization of the Poisson problem 23
9 Errors for optimal choice with smaller step size 23

1 Introduction

The study of ill-posed problems and their regularization plays a crucial role in applied mathematics, with applications spanning diverse fields such as physics, engineering, and data science. In this project, we focus on the mathematical framework and computational techniques for solving ill-posed problems. This work examines two classical formulations of ill-posed problems: the inverse Poisson problem and the deconvolution problem with Gaussian kernels. These problems exemplify the challenges associated with instability and sensitivity to noise that arise when addressing inverse problems.

Regularization techniques provide a way to mitigate these challenges, enabling the recovery of stable and meaningful solutions from noisy or incomplete data. Two widely used methods, Tikhonov regularization and truncated singular value decomposition (TSVD), are analyzed in depth. These methods highlight the correlation between solution stability and fidelity, as well as the critical role of parameter selection in achieving optimal regularization.

The main focus of this project is to explore how heuristic choice rules can be applied to ill posed problems to achieve optimal solutions. Special attention is given to the L-curve and its variations, as these tools help visualize the balance between smooth solutions and fitting the data.

The sections in this project are organized as follows:

2. **Background and Regularization Techniques**: The first section introduces the mathematical concepts of well-posedness and ill-posedness, providing context for regularization methods. Theoretical insights are supplemented with illustrative examples to build an intuitive understanding of problem structures. Tikhonov regularization and TSVD are also introduced, with discussions on their respective mathematical formulations and practical implementations.
3. **Parameter Choice Rules**: Heuristic and systematic approaches to parameter selection are analyzed here, including the L-curve method, generalized cross-validation (GCV), quasi-optimality, and Reginska's method.
4. **Discrete Problem Analysis**: The Poisson and deconvolution problems are translated into their discrete forms for computational study in this section.
5. **Results**: Numerical experiments illustrating the performance of regularization methods and parameter choice rules across two problem domains are presented. Observations focus on the stability, accuracy, and practicality of the methods.
6. **Discussion and Future Work**: The final section summarizes the findings, highlights limitations, and suggests directions for further exploration in the regularization of inverse problems.

This project aims to provide both a theoretical foundation and practical insights into the use of heuristic choice rules for regularization of ill-posed problems. The aim is to investigate how the choice rules perform across different types of noise levels and problems.

2 Background

Proofs and definitions in the following chapter are taken from the lecture notes of Christian Clason [2].

We want to solve the problem $Tx = y$ given y where T is a bounded linear operator. The problem is said to be well-posed when the following three criteria are fulfilled:

- The problem has a solution.
- The solution is unique.
- The solution depend continuously on the data.

If either of these fails to hold we are dealing with an ill-posed problem.

For further reading, we denote the range of T as $\mathcal{R}(T)$ and the nullspace or kernel of T as $\mathcal{N}(T)$.

Consider two Hilbert spaces X and Y and a bounded linear operator $T : X \to Y$. When the first condition for well posedness fails to hold we can more formally define this as solving $Tx = y$ when $y \notin \mathcal{R}(T)$ meaning no solution exists. The second condition fails if $\mathcal{N}(T) \neq \{0\}$ meaning there exists infinitely many solutions.

An example of an ill-posed problem is the deconvolution problem on $X = Y = L^2(\mathbb{R})$. Consider the problem below for some smooth kernel $k(x)$:

$$(k * u)(x) = \int_{-\infty}^{\infty} k(x - y)u(x)dy = f(x). \tag{1}$$

By the convolution theorem

$$u(x) = \int_{-\infty}^{\infty} \frac{\hat{f}(\omega)}{\hat{k}(\omega)} e^{2\pi i\omega} d\omega \tag{2}$$

where $\hat{f}$ and $\hat{k}$ are the Fourier transforms of f and k.

If $f(x)$ is not exact, but also contains additional noise $\delta(x)$ such that $f(x)^\delta = f(x) + \delta(x)$ we encounter problems. Since noise tends to dominate in high frequencies combined with the fact that $\hat{k}(\omega) \to 0$ as $|\omega| \to \infty$ resulting in

$$\frac{\hat{f}(\omega)}{\hat{k}(\omega)} \ll \frac{\hat{\delta}(\omega)}{\hat{k}(\omega)}$$

for high frequencies making the problem severely ill-conditioned.

Below is an illustration of a deconvolution of a step function convolved with a Gaussian kernel. A noise level of $0,1\%$ is added to the convolved signal before we try to reconstruct it. Throughout this paper, when referring to added noise, the noise is assumed to be normally distributed, and the added noise level refers to scaling of the noise as a scaling of the L^2-norm, i.e.

$$\delta \sim N(1, 0) \times \|f\|_2 \cdot (\textit{noise level})$$

The right plot shows the attempt to deconvolve the noisy signal back to the step function.

When $y \notin \mathcal{R}(T)$ we can try to find the best approximation $x^\dagger$ that minimizes the distance also called the *least squares solution* defined as

$$\|Tx^\dagger - y\|_Y = \min_z \|Tz - y\|_Y. \tag{3}$$

For cases where $\mathcal{N}(T) \neq \{0\}$ we choose the *minimum norm solution* $x^\dagger$ defined as

$$x^\dagger = \operatorname{argmin}\{\|z\|_X \,\big|\, z \text{ solves } \min_z \|Tz - y\|\}. \tag{4}$$

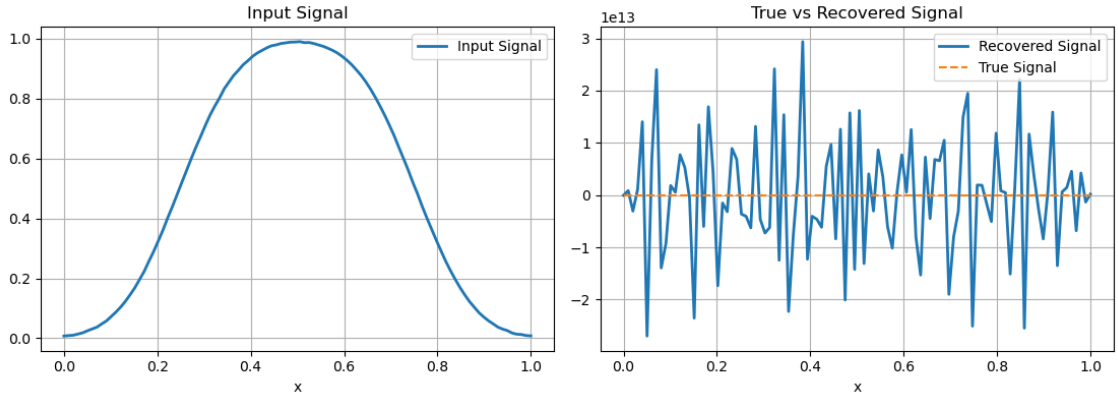

Figure 1: The left plot shows an input signal with 0,1% noise level. To the right is the numerically deconvolved signal where we can see how the noise dominates.

When $\mathcal{R}(T)$ is not closed, a minimum norm solution is not guaranteed to exist and it is therefore interesting to see for which $y \in Y$ we have a corresponding minimum norm solution. We now introduce the *Moore-Penrose inverse* also called the *pseudoinverse* or the *generalized inverse*.

We first start by restricting the domain and range of T such that an inverse exists and then extending this inverse to its maximal domain.

For a linear operator $T : \mathcal{D}(T) \to Y$ the domain of T is the linear subspace $\mathcal{D}(T) \subseteq X$ on which T is defined. $\mathcal{D}(T)$ does not need to be the whole space of X.

Theorem 1 (Moore-Penrose inverse)**.** *Let T be a linear operator between two Hilbert spaces $(X, \langle\cdot,\cdot\rangle_X)$ and $(Y, \langle\cdot,\cdot\rangle_Y)$, and denote by $\tilde{T}$ the restriction*

$$\tilde{T} = \big|_{\mathcal{N}(T)^\perp} : \mathcal{N}(T)^\perp \to \mathcal{R}(T). \tag{5}$$

Then there exists a unique linear extension of $\tilde{T}^{-1}$ called the Moore-Penrose inverse denoted $T^\dagger$ with properties

(i) $\mathcal{D}(T^\dagger) = \mathcal{R}(T) \oplus \mathcal{R}(T)^\perp$

(ii) $\mathcal{N}(T^\dagger) = \mathcal{R}(T)^\perp$.

Proof. The restriction of $\tilde{T}$ leads to a surjective and injevtive operator, thus $\tilde{T}^{-1}$ exists. This leads to $T^\dagger$ being well defined on $\mathcal{R}(T)$. Further if $y \in \mathcal{R}(T) \oplus \mathcal{R}(T)^\perp$ we have a unique orthogonal decomposition $y = y_1 + y_2$ with $y_1 \in \mathcal{R}(T)$ and $y_2 \in \mathcal{R}(T)^\perp$.Then

$$T^\dagger y := T^\dagger y_1 + T^\dagger y_2 = T^\dagger y_1 = \tilde{T}^{-1} y_1 \tag{6}$$

defines a unique linear extension satisfying $\mathcal{N}(T^\dagger) = \mathcal{R}(T)^\perp$. Hence $T^\dagger$ is well defined on the whole domain $\mathcal{D}(T^\dagger) = \mathcal{R}(T) \oplus \mathcal{R}(T)^\perp$. □

For $y \notin \mathcal{R}(T)$, the equation $Tx = y$ has no exact solution. The least squares solution minimizes the residual $\|Tx - y\|_Y$. Starting with:

$$\|Tx - y\|_Y^2 = \langle Tx - y, Tx - y\rangle_Y,$$

we differentiate to obtain the *normal equation*:

$$T^*Tx = T^*y, \tag{7}$$

where T^* is the adjoint of T. The normal equation ensures Tx is the closest point in $\mathcal{R}(T)$ to y, since solving the normal equation is equivalent to the least squares solution.

The Moore-Penrose pseudoinverse $T^\dagger$ provides the unique least squares solution:

$$x^\dagger = T^\dagger y.$$

While the Moore-Penrose inverse provides a framework for obtaining solutions to the problem $Tx = y$, it is often not practical in real life problems. When the problem is ill-posed, often due to instability or noise in the data, directly using the pseudoinverse can lead to large errors. If T is an ill-posed operator or $\mathcal{R}(T)$ is not closed we end up with an unbounded and therefore a non-continuous operator. The consequence of this is that small perturbations in the data often lead to large errors in the solution. We then introduce *regularization methods* to account for this.

2.1 Regularization methods

For real world measurements the data very often contain noise which can result in unstable numerical solutions, especially for ill-conditioned problems. We denote the noisy data as $y^\delta \in B_\delta(y)$. That is

$$\|y - y^\delta\| \leq \delta \text{ for some } \delta > 0. \tag{8}$$

For an ill-conditioned problem $T^\dagger y = x^\dagger$ we are not guaranteed that $x^\delta := T^\dagger y^\delta$ will be close to the true solution even for small δ. It is therefore necessary to adjust $T^\dagger$ such that the solution depends continuously on y^δ while still keeping the solution as close to $x^\dagger$ as possible. We do this by constructing a regularization of $T^\dagger$.

Definition 2.1. Let X and Y be two Hilbert spaces and $T \in \mathcal{L}(X, Y)$. A family $\{R_\alpha\}_{\alpha>0}$ of linear operators $R_\alpha : Y \to X$ is called a regularization of $T^\dagger$ if:

(i) $R_\alpha \in \mathcal{L}(X, Y) \ \ \forall\, \alpha > 0.$

(ii) $R_\alpha y \to T^\dagger y$ as $\alpha \to 0$ for all $y \in \mathcal{D}(T^\dagger)$.

Here $\mathcal{L}(X, Y)$ denotes the space of bounded linear operators between X and Y.

Ill-conditioned problems are characterized by being very sensitive to small changes in input data that lead to large changes in the solution. Examples of ill-conditioned problems which we will discuss later are discretizations of the Poisson problem and convolution with smooth kernels.

Given a problem $Tx = y$ we can get a picture of how ill-conditioned the problem is by looking at the condition number of T. In our case we work in the L^2-norm where the condition number is defined to be

$$\kappa(T) = \frac{\sigma_{max}(T)}{\sigma_{min}(T)} \tag{9}$$

where $\sigma(T)$ are the singular values of T. This brings us over to singular value decomposition, which plays a central part in regularization methods.

2.2 Singular value decomposition

The two problems that we are aiming to solve in this apper are the inverse one dimensional *Poisson equation*:

$$\Delta^{-1} f(x) = u(x), \tag{10}$$

and the *convolution problem*:

$$\int_X k(x, y)u(x)dx = f(x). \tag{11}$$

The inverse Laplace operator $\Delta^{-1} = (\nabla \cdot \nabla)^{-1}$ is a *compact operator*. This is easy to see if we look at the spectrum of the operator. The eigenfunctions of the Laplace operator on $[0, 1]$ with homogeneous Dirichlet boundary conditions are the distinct sine functions $\sin(n\pi)$ with the

corresponding eigenvalues $\lambda_n = \sqrt{n\pi}$. Furthermore, for homogeneous boundary conditions the operator is self adjoint resulting in *singular values* $\sigma(\Delta) = \lambda_n$.

We will also restrict the convolution problem to *Gaussian kernels* making the integral operator $(Tu)(x) = \int_X k(x,y)u(x)dx$ a compact operator. This property will later be shown useful for finding the Moore-Penrose inverse.

2.2.1 Compact operators

Let $T : X \to Y$ be a linear operator between the two Hilbert spaces X and Y. T is a compact operator if and only if T maps weakly convergent sequences in X to strongly convergent sequences in Y. Equivalently, if for every bounded sequence $\{x_n\}_{n\in\mathbb{N}} \subset X$ there exists a convergent subsequence $\{Tx_{n_k}\}_{n\in\mathbb{N}} \subset Y$, then T is compact. Every bounded linear operator with finite-dimensional range is compact. We denote the set of compact operators from X to Y as $\mathcal{K}(X,Y)$.

2.2.2 SVD of compact operators

An approach to finding the Moore-Penrose inverse of T is by use of spectral decomposition. The spectral theorem states that for a self adjoint compact operator T there exist an orthonormal system $\{u_n\}_{n\in\mathbb{N}}$ and a sequence $\{\lambda_n\}_{n\in\mathbb{N}} \subset [0,\infty)$ converging to 0 with

$$Tx = \sum_{n\in\mathbb{N}} \lambda_n \langle x, u_n\rangle_X u_n. \tag{12}$$

For proof and further information see [2, Theorem 2.5].

A given $T \in \mathcal{K}(X,Y)$ are not guaranteed to be self-adjoint. Instead we consider the Moore-Penrose inverse of T^*T, which is self adjoint, as in (7) leading to the *singular value decomposition*.

Theorem 2. *For every $T \in \mathcal{K}(X,Y)$ there exist:*

(i) a null sequence $\{\sigma_n\}_{n\in\mathbb{N}}$ with $\sigma_1 \geq \sigma_2 \geq \cdots > 0$

(ii) an orthonormal basis $\{u_n\}_{n\in\mathbb{N}} \subset Y$ of $\overline{\mathcal{R}(T)}$

(iii) an orthonormal basis $\{v_n\}_{n\in\mathbb{N}} \subset Y$ of $\overline{\mathcal{R}(T^*)}$

$$\text{with} \quad Tv_n = \sigma_n u_n \quad \text{and} \quad T^*u_n = \sigma_n v_n \text{ for all } n \in \mathbb{N} \tag{13}$$

and

$$Tx = \sum_{n\in\mathbb{N}} \sigma_n \langle x, v_n\rangle_X \; u_n \text{ for all } x \in X \tag{14}$$

The system (σ_n, u_n, v_n) is called a *singular system* and (14) is the *singular value decomposition* of T.

Proof. Since $T^*T : X \to X$ is compact and self-adjoint, the Spectral Theorem (Theorem 2.5) yields a null sequence $\{\lambda_n\}_{n\in\mathbb{N}} \subset [0,\infty)$ (ordered by decreasing magnitude) and an orthonormal system $\{v_n\}_{n\in\mathbb{N}} \subset X$ of corresponding eigenvectors such that

$$T^*Tx = \sum_{n\in\mathbb{N}} \lambda_n \langle x, v_n\rangle_X v_n \quad \text{for all } x \in X.$$

Since

$$\lambda_n = \lambda_n \|v_n\|_X^2 = \langle \lambda_n v_n, v_n\rangle_X = \langle T^*Tv_n, v_n\rangle_X = \|Tv_n\|_Y^2 > 0,$$

we can define for all $n \in \mathbb{N}$:

$$\sigma_n := \sqrt{\lambda_n} > 0 \quad \text{and} \quad u_n := \sigma_n^{-1} Tv_n \in Y.$$

(If the sequence $\{v_n\}_{n\in\mathbb{N}}$ terminates finitely, we set $\sigma_n := 0$.) The latter forms an orthonormal system because

$$\langle u_i, u_j\rangle_Y = \frac{1}{\sigma_i\sigma_j}\langle Tv_i, Tv_j\rangle_Y = \frac{1}{\sigma_i\sigma_j}\langle T^*Tv_i, v_j\rangle_X = \frac{\lambda_i}{\sigma_i\sigma_j}\langle v_i, v_j\rangle_X = \begin{cases}1, & \text{if } i = j,\\ 0, & \text{otherwise.}\end{cases}$$

Furthermore, for all $n \in \mathbb{N}$, we have

$$T^*u_n = \sigma_n^{-1}T^*Tv_n = \sigma_n^{-1}\lambda_n v_n = \sigma_n v_n.$$

□

This result will help us finding the Moore-Penrose inverse.

2.2.3 SVD for finite dimensional linear transformations

Since every finite rank operator is compact all matrices with finite rank are compact. When we solve the problems numerically we discretize our problem to a finite number of grid points making our linear transformations representable by finite dimensional matrices. This allows us to apply singular value decomposition to our linear operators.

Let T be a linear transformation from $\mathbb{R}^N$ to $\mathbb{R}^M$. Then there exists a decomposition

$$T = U\Sigma V^T$$

where U and V are orthogonal matrices where the column vectors form an orthonormal basis for $\mathbb{R}^M$ and $\mathbb{R}^N$ respectively. The matrix Σ is an $M \times N$ matrix with singular values $\sigma(T) = \{\sigma_1, ..., \sigma_p\}$, $p = \min\{M, N\}$, where $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_r > \sigma_{r+1} = ... = \sigma_p = 0$, r being the rank of T. Beneath is an illustration of Σ with $M > N$.

$$\Sigma = \begin{pmatrix}\sigma_1 & 0 & 0 & \dots & 0\\ 0 & \sigma_2 & 0 & \dots & 0\\ 0 & 0 & \sigma_3 & \dots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \dots & \sigma_p\\ 0 & 0 & 0 & \dots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \dots & 0\end{pmatrix}_{M\times N}$$

This decomposition of T admits the Moore-Penrose inverse $T^\dagger = V\Sigma^\dagger U^T$ with

$$\Sigma^\dagger = \begin{pmatrix}\frac{1}{\sigma_1} & 0 & \dots & 0 & 0 & \dots & 0 & \dots & 0\\ 0 & \frac{1}{\sigma_2} & \dots & 0 & 0 & \dots & 0 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & \frac{1}{\sigma_r} & 0 & \dots & 0 & \dots & 0\\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & \dots & 0\end{pmatrix}_{N\times M}$$

only keeping the non-negative singular values.

One way of intuitively thinking of SVD is to imagine that the column vectors of U and V represent the characteristics of the linear transformation and Σ is a scaling of these characteristics. The singular values are particularly interesting when working with inverse problems since for ill-conditioned matrices the singular values decay rapidly towards zero. This will result in a condition number being extremely large or even undefined thus making T sensitive to small perturbations. A way to compensate for this is to use the *truncated singular value decomposition*.

2.3 Truncated SVD

Let $T = U\Sigma V^*$ be the singular value decomposition of T with

$$\Sigma = \text{diag}\{\sigma_1, \sigma_2, ..., \sigma_p\} \in \mathbb{R}^{M\times N}.$$

A way to regularize a problem is to truncate the SVD by retaining the first $k \leq r$ singular values of Σ where $r = \text{rank}(A)$. The TSVD approximation of T becomes

$$T_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T = U_k \Sigma_k V_k^T. \tag{15}$$

This formulation admits the Moore-Penrose inverse

$$T_k^\dagger y = \sum_{i=1}^{r} \frac{v_i u_i^T}{\sigma_i} = V_k \Sigma^\dagger U_k^T \tag{16}$$

We know by the Eckart-Young-Mirsky theorem [3] that this is the best low-rank approximation of T, i.e.

$$T_k = \min_M \{\|T - M\|_2 \mid \text{rank}(M) = k\}. \tag{17}$$

The theorem also states that

$$\|T - T_k\|_2 = \sigma_{k+1}. \tag{18}$$

We use this property to reduce the condition number without losing too much information from T. This approach will be central when choosing optimal parameters. This way we can keep the main characteristics of T, but reduce the condition number drastically.

Let's have a look at how this method reduces error when encountering noise. Consider the problem $Tx = y$ but we only have noisy data $y^\delta = y + \delta s$ where s is some normalized random vector. Using the direct solution with Moore-Penrose inverse will be as follows

$$x^\delta = T^\dagger y^\delta = \sum_{i=1}^{r} \frac{u_i^T y^\delta}{\sigma_i} v_i = \sum_{i=1}^{r} \frac{u_i^T y}{\sigma_i} v_i + \delta \sum_{i=1}^{r} \frac{u_i^T s}{\sigma_i} v_i$$

where r is the rank of T. Here we can see that for ill-conditioned problems with rapidly decaying singular values the error term will explode for small values of σ. With proper choice of k we can prevent

$$\delta \sum_{i=1}^{k} \frac{u_i^T s}{\sigma_i} v_i$$

from dominating the solution. We will later look at methods to choose an optimal k for reconstruction.

2.4 Tikhonov Regularization

Another common way to handle instability and noise sensitivity in ill-conditioned problems is a method called Tikhonov regularization. This method adds a penalty term to balance the fit of

the solution to the data with its "smoothness" or "regularity." This approach will, similarly to TSVD, handle cases where direct inversion would amplify noise due to small singular values in the operator T.

The Tikhonov-regularized solution x_α is defined as the minimizer of the following functional:

$$x_\alpha = \arg\min_x \left\{ \frac{1}{2}\|Tx - y^\delta\|^2 + \frac{\alpha}{2}\|Lx\|^2 \right\}, \tag{19}$$

where:

- $\|Tx - y^\delta\|^2$ is the fidelity term, measuring how well the solution x fits the noisy data y^δ.
- $\|Lx\|^2$ is the regularization term, which penalizes "large" or "irregular" solutions.
- $\alpha > 0$ is the regularization parameter that controls the trade-off between the fidelity term and the regularization term.

For this project we use $L = I$, the identity matrix, which is known as *standard Tikhonov regularization*. However, we can generalize the approach by choosing L as a differential operator or another matrix that imposes desired characteristics on x, such as smoothness.

2.4.1 Choice of the Regularization Parameter α

The regularization parameter α plays a central role in determining the balance between data fidelity and stability of the solution. A very small α results in a solution that may fit the noisy data closely but risks being unstable. Conversely, a large α overly smooths the solution, potentially losing significant details of the true solution.

Choosing an optimal value for α can be challenging. Based on our initial knowledge about our problem we can use different methods to choose α.

1. **Discrepancy Principle**: Choose α such that the residual norm $\|Tx_\alpha - y^\delta\|$ is approximately equal to the noise level δ.
2. **L-curve Method**: Plot $\|Tx_\alpha - y^\delta\|$ versus $\|Lx_\alpha\|$ for varying α, and choose the value of α corresponding to the "corner" of the resulting L-curve. In our case L is the identity matrix.
3. **Other choice rules**: Choice rules balancing the trade-off between overfitting and underfitting of the solution that only depend on the noisy data.

Each of these methods has its strengths and can be selected based on the properties of the data and the problem at hand. For this project we will look more closely at the L-curve method and other data-dependent choice rules as we are working with *heuristic choice rules*.

2.4.2 Solution via Normal Equations

The minimization problem in Eq. (7) can be solved by setting the gradient to zero, leading to the so-called Tikhonov normal equations:

$$(T^*T + \alpha L^*L)x_\alpha = T^*y^\delta. \tag{20}$$

For $L = I$, this simplifies to

$$(T^*T + \alpha I)x_\alpha = T^*y^\delta. \tag{21}$$

These equations can then be solved using numerical methods, such as iterative solvers or singular value decomposition (SVD). Tikhonov regularization is particularly effective because it smooths out small singular values in T, resulting in a more stable solution than the Moore-Penrose inverse.

2.4.3 Solution via Singular Value Decomposition (SVD)

For a clearer insight into the effect of Tikhonov regularization, we can use the singular value decomposition of T. Let

$$T = U\Sigma V^*$$

be the SVD of T, where U and V are orthogonal matrices and Σ is the diagonal matrix of singular values σ_i of T. In this case, the Tikhonov regularized solution x_α can be expressed as:

$$x_\alpha = (T^*T + \alpha I)^{-1}T^*y^\delta. \tag{22}$$

Using the SVD of T, we can substitute $T = U\Sigma V^*$ and $T^* = V\Sigma U^*$ into the expression:

$$\begin{aligned} x_\alpha &= V(\Sigma^*\Sigma + \alpha I)^{-1}\Sigma^* U^* y^\delta \\ &= V \operatorname{diag}\left(\frac{\sigma_i}{\sigma_i^2 + \alpha}\right) U^* y^\delta, \end{aligned} \tag{23}$$

where $\frac{\sigma_i}{\sigma_i^2+\alpha}$ acts as a filter on each singular component.

The regularization parameter α serves to dampen the contributions from components corresponding to smaller singular values σ_i. When α is large relative to σ_i^2, the factor $\frac{\sigma_i}{\sigma_i^2+\alpha}$ approaches zero, effectively filtering out components associated with small singular values, which are typically noise-sensitive. Conversely, when σ_i^2 is large compared to α, the factor $\frac{\sigma_i}{\sigma_i^2+\alpha}$ which is close to $\frac{1}{\sigma_i}$, closely approximating the solution without regularization.

This filtering effect helps to stabilize the solution x_α, making Tikhonov regularization particularly powerful for ill-conditioned problems.

3 Parameter choice

The choice rules described in the following part are taken from [7]. The discrete problems are also mappings from $\mathbb{R}^N \to \mathbb{R}^N$ resulting in square matrices.

There are different ways of approaching the choice of regularization parameter depending on prior and posterior knowledge. The three ways of choosing the parameter are as follows

Definition 3.1. A parameter choice rule is a function $\alpha : \mathbb{R}^+ \times Y \to \mathbb{R}^+$ $(\delta, y^\delta) \mapsto \alpha(\delta, y^\delta)$. There are three ways of defining such a function

(i) *a priori* choice, only dependent on δ

(ii) *posterior* choice, dependent on both δ and y^δ

(iii) *heuristic* choice, only dependent on y^δ

If $\{R_\alpha\}_{\alpha>0}$ is a regularization of $T^\dagger$ and α is a parameter choice rule, the pair (R_α, α) is called a (convergent) regularization method if

$$\lim_{\delta\to 0} \sup_{y_\delta \in B_\delta(y)} \|R_{\alpha(\delta,y^\delta)} y^\delta - T^\dagger y\|_X = 0 \quad \text{for all } y \in D(T^\dagger). \tag{24}$$

We thus demand that the regularization error vanishes for all noisy measurements y_δ that are compatible with the noise level $\delta \to 0$.

This project focuses on heuristic choice rules as these are the ones applied in the numerical methods.

3.1 Heuristic choice rules

As stated above, heuristic choice rules depend only on the data y^δ. These rules are used when we do not possess sufficient knowledge of the noise level δ. Even though heuristic choice rules are widely use there is a profound challenge in relation to regularization methods called *Bakushinski's veto* [1].

Theorem 3 (Bukushinski's veto)**.** *Let $\{R_\alpha\}_{\alpha>0}$ be a regularization of $T^\dagger$. If there exists a heuristic choice rule α such that (R_α, α) is a regularization method, then $T^\dagger$ is continuous.*

Proof. Assuming to the contrary that such a parameter choice rule $\alpha : Y \to \mathbb{R}^+$ exists, we can define the (possibly nonlinear) mapping

$$R : Y \to X, \quad y \mapsto R_{\alpha(y)} y.$$

Let now $y \in D(T^\dagger)$ be arbitrary and consider any sequence $\{y_n\}_{n\in\mathbb{N}} \subset D(T^\dagger)$ with $y_n \to y$.

On the one hand, then naturally $y_n \in B_\delta(y_n)$ for all $\delta > 0$ and $n \in \mathbb{N}$, and the assumption (24) for fixed $y_\delta = y = y_n$ and $\delta \to 0$ yields that $Ry_n = T^\dagger y_n$ for all $n \in \mathbb{N}$ (and hence that R is in fact linear on $D(T^\dagger)$).

On the other hand, for $\delta_n := \|y_n - y\|_Y$, we also have $y_n \in B_{\delta_n}(y)$, and in this case, passing to the limit $n \to \infty$ in (24) shows that

$$T^\dagger y_n = Ry_n = R_\alpha(y_n) y_n \to T^\dagger y,$$

i.e., $T^\dagger$ is continuous on $D(T^\dagger)$.

□

The Bakushinskii veto highlights that heuristic choice rules cannot guarantee regularization methods for infinite-dimensional ill-posed problems due to the inherent discontinuity of the operator $T^\dagger$ in such settings. This limitation arises because heuristic rules lack precise knowledge of the noise level δ, a core factor in achieving stability. However, in practical numerical computations, many problems are finite-dimensional or involve smoothing operators where noise does not typically align with $R(T)$. This is advantageous because when $y^\delta \notin R(T)$, the noise is smoothed out by the regularization process, reducing its impact on the solution and improving stability. In these "usual" scenarios, heuristic choice rules, combined with appropriate regularization operators, often perform effectively, stabilizing solutions while avoiding overfitting to noise. Examples include popular rules like the discrepancy principle or L-curve criterion, which adaptively balance fidelity to noisy data and solution smoothness. Thus, while theoretical constraints exist, heuristic rules remain valuable tools in applied numerical analysis.

3.2 L-curve method

The L-curve is a graphical tool to visualize the trade-off between the size of the solution and how well the solution is fit to the data. More formally the L-curve is a plot based on a decreasing sequence $\{\alpha_n\}_{n\in\{1,\dots,N\}}$ with points corresponding to each α_n. The reason for the name L-curve is the way the plot tends to take the form of an L. The idea is that the optimal parameter is found around the vertex of the curve. Figure 2 shows a *residual* L-curve with a clear bend making the L-shape. The idea of this vertex is to find a solution at the point where the residual is not too large while simultaneously keeping the norm of the solution small. Keeping the solution norm small is a way to keep the solution from being too contaminated by noise. For the residual we also want to keep it fairly small to ensure that our solution is accurate. Determining the vertex gives us a parameter that results in a good trade-off between the desired properties.

For cases where the L-curve takes an unambiguous shape as in Figure 2 it is quite easy to determine the vertex visually. However, finding this vertex by a computer program is not equally trivial. For discrete points connected by line segments small kinks can occur making the curve non convex and may also make small vertices "fooling" our program to determine a false vertex. Measures can be made to avoid such problems, for instance by choosing the largest convex subset of consecutive points.

There are different ways to determine the set of points that draw out the L-curve. The most general way is by looking directly at the residual $\|y^\delta - Tx\|$ against the solution norm $\|x\|$.

3.2.1 L-curve criterion from solution norm

The residual L-curve is a plot consisting of the points

$$\{\log \|x_n\|, \log \|y^\delta - Tx_n\|\} \tag{25}$$

where $x_n = R_{\alpha_n}(y^\delta)$ and R_{α_n} is a regularization of $T^\dagger$. If we only base the parameter choice upon visual inspection, we are reliant on an easily interpretable shape as in Figure 2. If the noise is too big we can end up with a plot that is harder to interpret. In Figure 3 the noise is 1% as opposed to 0,1% in Figure 2 resulting in the bend being wider making it harder to distinguish where the best parameter choice lies.

3.2.2 Condition L-curve criterion

The condition L-curve uses the condition number of the regularization matrix rather than the solution norm. In the same way that a low solution norm ensures that our solution is not too contaminated by noise, a low condition number corresponds to the same.

If the SVD of our regularized matrix is available the condition number is easily accessible. Recall that $\kappa(T) = \frac{\sigma_{max}}{\sigma_{min}}$ thus the TSVD this number becomes $\kappa_k = \frac{\sigma_1}{\sigma_k}$ where k is the truncation index.

Similarly for Tikhonov regularization we have $\kappa_\alpha = \frac{\sqrt{\sigma_1^2+\alpha}}{\sqrt{\sigma_N^2+\alpha}}$. Using these properties we construct the L-curve from the point set

$$\{\log(\kappa), \log \|y^\delta - Tx_n\|\}. \tag{26}$$

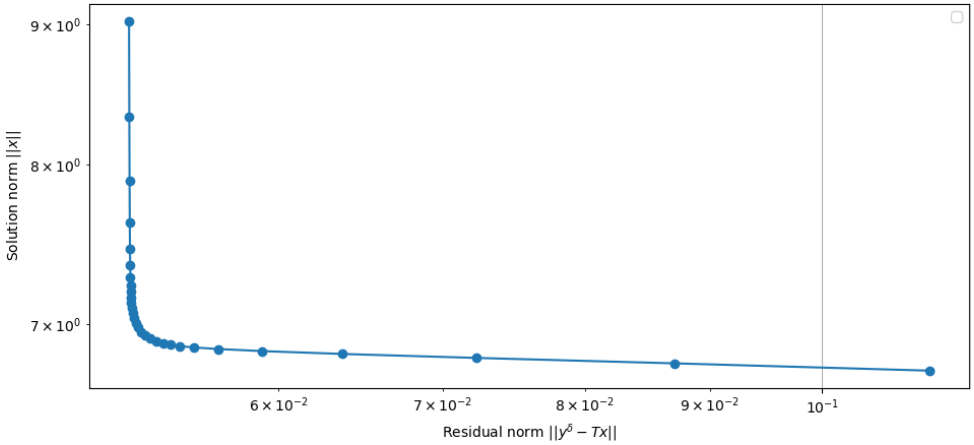

Figure 2: L-curve for solutions of the deconvolution problem with Tikhonov regularization and 0.1% noise level. Each point corresponds to a distinct value for α.

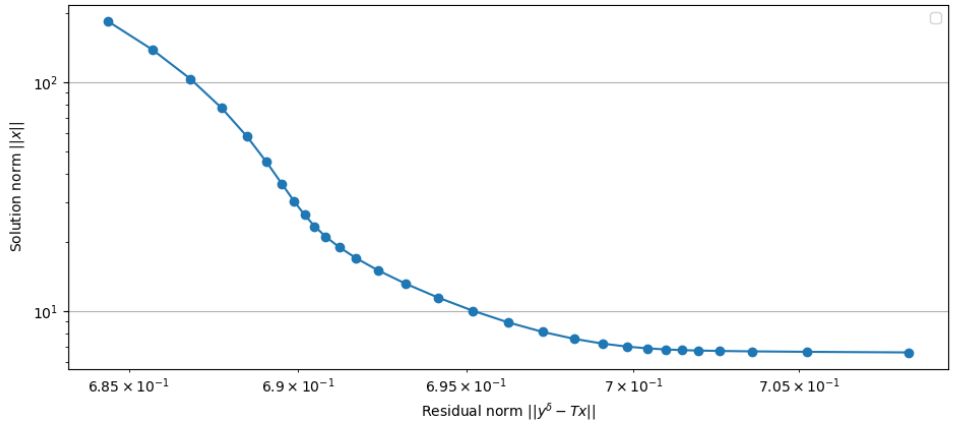

Figure 3: L-curve for solutions of the deconvolution problem with Tikhonov regularization and 1% noise level.

3.2.3 Maximal curvature

Hansen and O'Leary [5] suggested a way to find the optimal parameter by looking at the point on the L-curve with maximal curvature.

The curvature $\kappa(\alpha)$ for the L-curve is defined as:

$$\kappa(\alpha) = \frac{\rho'\eta'' - \rho''\eta'}{((\rho')^2 + (\eta')^2)^{3/2}},$$

where:

- $\rho =$ is the residual norm

- $\eta = \|Lx\|$ is the solution norm
- $'$ denotes the derivative with respect to the regularization parameter α.

As α is a continuous parameter for Tikhonov regularization the resulting L-curve takes a smooth form which makes this method more likely to give a reasonable suggestion. For TSVD however, we only have a discrete set of points corresponding to our truncation parameter making the L-curve non-differentiable. Trying to make a discrete approximation to the curvature can be misleading, as we often have clustering of singular values. This results in fine-grained details disturbing the interpretation of the overall shape of the curve.

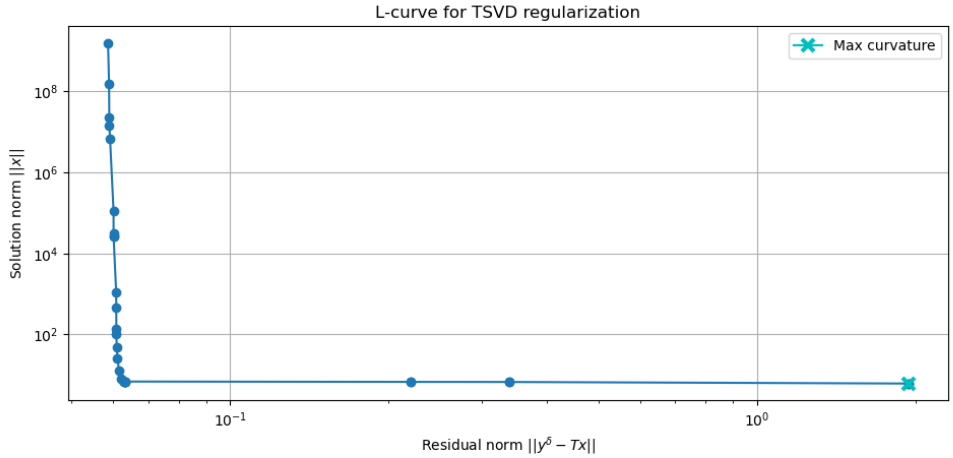

Figure 4: The computed maximum curvature of the L-curve TSVD is marked with a cross. By visual inspection it is clear that the maximal curvature method fail to determine the vertex of the L-curve where our desired optimal parameter lies.

3.2.4 Regińska's method

Another parameter choice rule suggested by Regińska [6] where she showed how it relates to the L-curve criterion. This was originally meant for Tikhonov regularization with a continuous set of parameter, but can easily be adapted to a discrete set. Regińska's method applied to the discrete set of parameters $\{\alpha_j\}_{j=1}^N$ with the corresponding set of solution norms and residuals

$$\{\|x_j\|, \|r_j\|\}_{j=1}^N,$$

is done by finding the index j that minimizes

$$\phi(j) = \|r_j\|\|x_j\|^\nu. \tag{27}$$

Here $\nu > 0$ is some tuning parameter often set to $\nu = 2$ and j corresponds to either the truncation index k for TSVD or α_j for Tikhonov regularization.

3.2.5 Quasi-optimality criterion

This criterion utilizes the assumption that we have a clustering of points near the optimal parameter. For a discrete set of parameters choose the parameter of index j such that

$$j_q = \underset{1 \le j \le N}{\arg\min} \|x_{j+1} - x_j\|. \tag{28}$$

This method is quite straightforward for TSVD since our discretization of parameters is quite trivial. However, for Tikhonov it is not equally trivial to use this method. In theory the discrete

set $\{\alpha_j\}_{1\leq j\leq N}$ can have two parameters arbitrarily close making $\|x_{j+1} - x_j\|$ arbitrarily small for a continuous solution. Thus, we need to be careful when applying this method on Tikhonov regularization and choose a suitable spacing of the parameters in the set.

We have used logarithmic spacing for the set of parameters for all methods, giving the opportunity for comparison across different methods. By looking at different methods for the same set of parameters we can more easily determine whether the optimal parameter suggested by a given method is feasible.

3.2.6 Generalized Cross Validation (GCV)

Generalized Cross Validation is a statistical method for selecting the regularization parameter, based on the idea of leaving out parts of the data and evaluating how well the regularized solution predicts the missing data. The error in the data, denoted by e, is assumed to follow a random distribution with zero mean and covariance matrix $\sigma^2 I$, where σ^2 is unknown.

GCV for TSVD For Truncated Singular Value Decomposition (TSVD), the GCV method identifies the truncation index k that minimizes the function:

$$G(j) = \frac{\|r_j\|^2}{(m-j)^2},$$

where:

- $r_j = \|y^\delta - Tx_j\|$ is the residual for the solution x_j corresponding to the truncation index j,
- m is the number of data points,
- j is the current truncation index.

GCV for Tikhonov For Tikhonov we have a matrix dependent choice rule where the optimal parameter α_j is found by minimizing

$$G(\alpha_j) = \frac{\|r_j\|^2}{\text{trace}(I - T(\alpha_j))^2},$$

where

$$T(\alpha_j) = T(T^T T + \alpha_j I)^{-1} T^T.$$

This is quite computationally expensive, especially for large scale problems. However, if the SVD of T is available this becomes considerably easier. It is shown in [4] that if $T \in \mathbb{R}^{M\times N}$

$$\text{trace}(I - T(\alpha_j)) = M - \sum_{k=1}^{N} \frac{\sigma_k^2}{\sigma_k^2 + \alpha_j}. \tag{29}$$

This formula balances the residual norm $\|r_j\|$ with the number of parameters used in the regularization (via $m - j$), ensuring a good trade-off between overfitting and underfitting.

4 The discrete problems

We now reintroduce our two problems in terms of the discrete form.

4.1 Poisson problem

The Poisson problem with homogeneous Dirichlet boundary conditions can be discretized with a central difference matrix $L \in \mathbb{R}^{N\times N}$ and vectors $x, y \in \mathbb{R}^N$

$$Lx = y, \quad x_1 = x_N = 0 \tag{30}$$

with

$$L = \frac{1}{h^2}\begin{bmatrix} -2 & 1 & 0 & \cdots & 0 & 0 \\ 1 & -2 & 1 & \cdots & 0 & 0 \\ 0 & 1 & -2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -2 & 1 \\ 0 & 0 & 0 & \cdots & 1 & -2 \end{bmatrix} \tag{31}$$

where h is the step size of the discretization. The ill-posedness of the problem comes to show when noise is added to x, since the Laplace is sesnitive to rapid changes in the data. Thus the actual problem we want to regularize is the inverse Poisson problem

$$L^{-1}y = x^\delta. \tag{32}$$

Since L is symmetric and *positive semi-definite* we know that the singular values corresponds to the absolute value of the eigenvalues of L. The symmetry also yields that $U = V$ in the SVD of L and that all eigenvalues are real. Thus the SVD of L becomes

$$L^{-1} = V\Sigma^{-1}U^T = U\Lambda^{-1}U^T, \quad \Lambda = \operatorname{diag}(\lambda_1, ..., \lambda_N). \tag{33}$$

The eigenvalues and corresponding eigenfunctions to the discrete Poisson problem with homogeneous Dirichlet boundary conditions equal to zero are [8, (13.8)]

$$-\Delta \sin(n\pi) = \lambda_n \sin(n\pi) \tag{34}$$

where

$$\lambda_n = \frac{4}{h^2}\sin^2\Big(\frac{n\pi}{2(N+1)}\Big). \tag{35}$$

This tells us that all eigenvalues for a finite N is non zero implying that Σ^{-1} exists and is then equal to $\operatorname{diag}\Big(\frac{1}{\lambda_1}, ..., \frac{1}{\lambda_N}\Big)$.

Tikhonov regularization The solution to the Tikhonov regularization can then be expressed as

$$y^\dagger = (L^{-T}T^{-1} + \alpha I)^{-1}T^{-T}x^\delta = (U\Sigma^{-2}U^T + \alpha I)T^{-T}x^\delta \tag{36}$$

where L^{-T} denotes the transpose of L inverse. Inverting matrices is often computationally expensive. Instead we split the problem into two equations and use a linear solver on the two equations:

$$L^T v = x^\delta,$$
$$(U\Sigma^{-2}U^T + \alpha I)y = v.$$

Solving $L^T v = x^\delta$ is the same as using $v = L^{-T}x^\delta$ and similarly solving $(U\Sigma^{-2}U^T + \alpha I)y = v$ corresponds to $y = (U\Sigma^{-2}U^T + \alpha I)^{-1}v$. Putting these two equations together results in (36).

TSVD For the TSVD we have to make one consideration before applying the method, namely that we are solving for L^{-1} and not L directly. This makes the singular values "inverted" in the sense that $\frac{1}{\sigma_N} \geq \frac{1}{\sigma_{N-1}} \geq ... \geq \frac{1}{\sigma_1}$. Thus when we truncate our singular values we truncate from the smallest singular value giving

$$L_k^{-1} = U\Sigma_k^{-1}U^T, \quad \text{with } \Sigma_k^{-1} = \text{diag}\Big(\frac{1}{\sigma_N}, \frac{1}{\sigma_{N-1}}, ..., \frac{1}{\sigma_{N-k}}\Big).$$

This makes for a pretty straightforward regularization giving

$$\begin{aligned} y^\dagger &= R_k(x^\delta) = (L_k^{-1})^\dagger x^\delta \\ &= U(\Sigma_k^{-1})^{-1}U^T x^\delta = U\Sigma_k' U^T x^\delta, \quad \Sigma_k' = \text{diag}(\sigma_N, \sigma_{N-1}, ..., \sigma_{N-k}). \end{aligned}$$

4.2 Deconvolution problem with Gaussian kernel

Contrary to the Poisson problem, we now have a more straightforward solution where we can apply the methods described in Section 2 directly. The discrete deconvolution problem consists of solving

$$\frac{1}{N}Kx = y^\delta \tag{37}$$

where K is a symmetric matrix with rows equal to a normalized Gaussian vector with zero mean and some parameter σ determining the width of the kernel.

The first row of K is constructed by

$$K_1 = \frac{1}{\sqrt{2\pi}\sigma}\exp\Big(\frac{1}{2}\frac{x^2}{\sigma^2}\Big)$$

We then input this vector into `scipy.linalg.toeplitz` to create the full matrix K. For $N = 7$ and $\sigma = 0.1$ the matrix will look like this:

$$K = \begin{bmatrix} 0.25330 & 0.06316 & 0.00098 & 0.00000 & 0.00000 & 0.00000 & 0.00000 \\ 0.06316 & 0.25330 & 0.06316 & 0.00098 & 0.00000 & 0.00000 & 0.00000 \\ 0.00098 & 0.06316 & 0.25330 & 0.06316 & 0.00098 & 0.00000 & 0.00000 \\ 0.00000 & 0.00098 & 0.06316 & 0.25330 & 0.06316 & 0.00098 & 0.00000 \\ 0.00000 & 0.00000 & 0.00098 & 0.06316 & 0.25330 & 0.06316 & 0.00098 \\ 0.00000 & 0.00000 & 0.00000 & 0.00098 & 0.06316 & 0.25330 & 0.06316 \\ 0.00000 & 0.00000 & 0.00000 & 0.00000 & 0.00098 & 0.06316 & 0.25330 \end{bmatrix}.$$

Tikhonov regularization For this problem we calculate the SVD of K and use the regularization as described in (23).

TSVD Similarly for TSVD we apply the method as described in (16), without the need for further complications.

5 Results

In this chapter, we present the results of numerical experiments conducted to evaluate the effectiveness of the parameter choice rules described in Chapter 2 applied to the problems described in Chapter 4. The aim is to demonstrate how these techniques perform on different types of ill-posed problems and under various noise conditions. The focus is on analyzing the stability and accuracy of the different choice rules.

We will be using the true signal without noise to compare with the recovered signal. We will use all parameter choice rules on the same set of parameters to get a comparison between the choice rules. For all cases, we have used gridsize $N = 100$.

5.1 Clustering of solutions

An early observation made for the TSVD L-curves is that the residual norms and solution norms seem to cluster together in pairs. Recall from (34) that the eigenfunctions of the Laplace operator are a set of even and odd functions. We know that the basis for the discrete Gaussian convolution also relates to some basis dependent on $e^{i2\pi\frac{n}{N}}$ arising from the Fourier transform which also gives a set of even and odd functions. Since the operators in our case are symmetric and the problems we have chosen are as well, it makes sense that we only get contributions from the even eigenfunctions. Thus, for each odd eigenfunction we do not get a large contribution, resulting in a pairwise clustering of the solution.

We use this observation to precondition the parameter set to make the choice rules more precise and exclude the truncation indexes corresponding to the odd eigenfunctions.

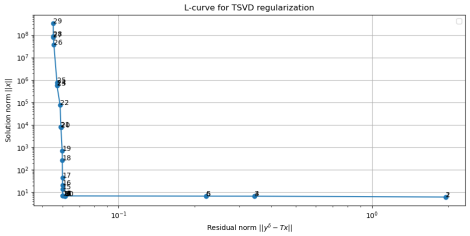

Figure 5: Example of TSVD clustering for the deconvolution problem.

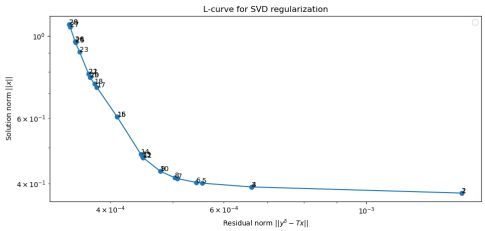

Figure 6: Example of TSVD clustering for the Poisson problem.

5.2 Heavyside signal convolved with Gaussian kernel

For the convolution problem we look at the classical Heavyside function

$$H(x-a) = \begin{cases} 0 & \text{if } x < a, \\ 1 & \text{if } x \geq a. \end{cases} \tag{38}$$

We use the function $f(x) = H(x - 0,25) - H(x - 0,75)$ with $x \in [0, 1]$ as our true solution and convolve with the Gaussian kernel. After this we add noise to the convolved function, which we then try to regularize to recover $f(x)$. The problem will be to solve

$$Tx = y^\delta, \quad x, y \in \mathbb{R}^N, \quad T \in \mathbb{R}^{N\times N}, \tag{39}$$

where T is the kernel matrix described in section 4, x is the unknown function and y^δ is the convolved Heavyside function with added noise.

5.2.1 Tikhonov regularization results for the deconvolution problem

In this section we will look at how the choice rules perform for Tikhonov regularization of (39).

Let us start by looking at the problem with $\delta = 0.1\%$.

In Figure 7 we can see an L-curve for the Tikhonov regularization on the left, and on the right is the curvature plotted as a function of α. We observe that the L-curve has a very sharp vertex which is favorable for both visual inspection and the max curvature choice rule. In Figure 8 we can see that the choice rules have resulted in different choices, but all close to the vertex. The Regińska optimal is not visible, as it coincided with the max curvature choice.

In Figure 9 we can see the recovered solution. The parameter choice used was the choices of both of Regińska and max curvature of the standard L-curve, as they coincided.

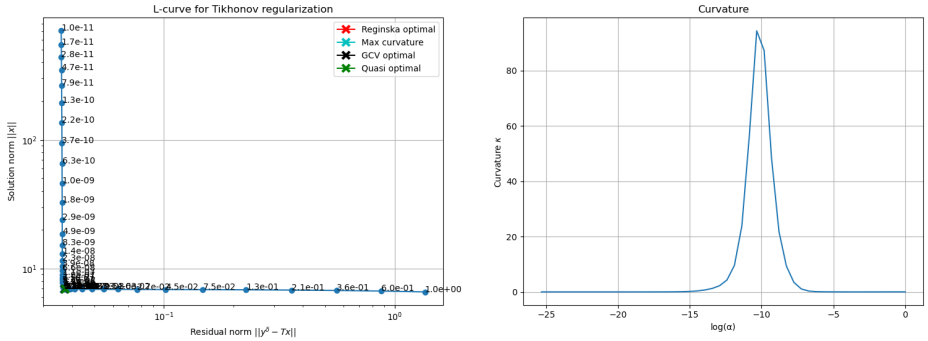

Figure 7: The deconvolution problem with 0.1% noise results in an L-curve seen in the plot to the left. To the right is the curvature of the L-curve plotted as a function of the regularization parameter α.

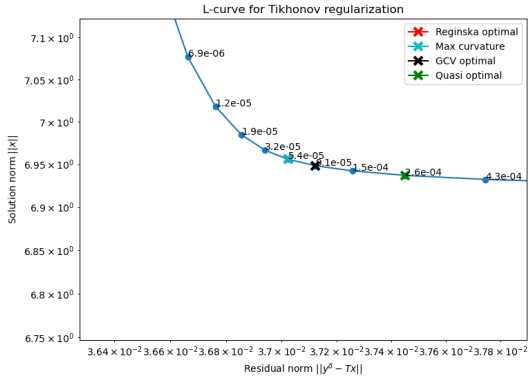

Figure 8: L curve from Figure 7 zoomed in at the vertex. The Regińska optimal coincided with max curvature for this particular case.

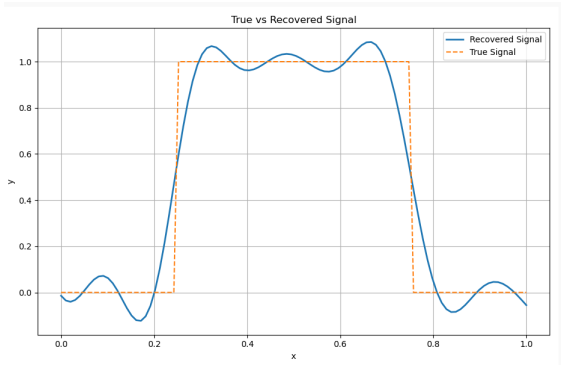

Figure 9: The recovered function with the max curvature as parameter choice where $\alpha = 5.4 \times 10^{-5}$.

In Table 1 and 2 we can see the performance of the different choice rules. For each noise level the reconstruction error for the respective parameter choice is listed. The table presents the normalized reconstruction error $\frac{\|y_{\text{true}} - y_{\text{rec}}\|_2}{\|y_{\text{true}}\|_2}$ for Tikhonov regularization of the deconvolution problem.

As the noise level decreases, the residual error reduces in general across all three methods. This is expected since lower noise allows for a more accurate reconstruction of the original signal.

Overall, no immediate large deviations from the true optimal appears. All methods except the max curvature from the condition L-curve admits quite consistent results with strictly decreasing errors implying good stability. No method can consistently choose the optimal parameter, but

Noise Level (%)	Standard L-curve	Condition L-curve	True Optimal
10%	**0.2517**	**0.2517**	0.2507
1%	**0.2028**	0.2228	0.2028
0.1%	**0.1839**	0.2141	0.1713
0.01%	**0.1739**	0.1849	0.1713
0.001%	**0.1559**	0.2141	0.1541
0.0001%	**0.1532**	0.1568	0.1427

Table 1: Tikhonov regularization: Errors $\frac{\|y_{\text{true}}-y_{\text{rec}}\|_2}{\|y_{\text{true}}\|_2}$ for the max curvature choice rule applied to both standard L-curve and the condition L-curve. The best performance for each noise level is in bold.

Noise Level (%)	Regińska	Quasi Optimal	GCV Optimal	True Optimal
10%	**0.2507**	0.2740	0.2740	0.2507
1%	**0.2050**	0.2085	0.2085	0.2028
0.1%	0.1778	0.1839	**0.1734**	0.1713
0.01%	**0.1561**	0.1658	0.1578	0.1541
0.001%	**0.1490**	0.1543	0.1558	0.1427
0.0001%	**0.1490**	0.1543	0.1558	0.1427

Table 2: Tikhonov regularization: Errors $\frac{\|y_{\text{true}}-y_{\text{rec}}\|_2}{\|y_{\text{true}}\|_2}$ with Regińska, Quasi Optimal and GCV as choice rules. The best performance for each noise level is in bold.

the Regińska choice stands out as the most consistent and accurate choice for this problem and regularization method.

In Figure 10 we can see the condition L-curve plotted for 0.001% noise level which is where the we observe the largest inconsistency in our results. By visual inspection we can see the max curvature choice rule chooses a parameter far from the vertex. This highlights why choosing a suitable set of points for the choice rule is important for optimal result. For this case, one possible approach can be to choose the largest convex subset of consecutive points in the L-curve, as presented for the restricted Regińska in [7].

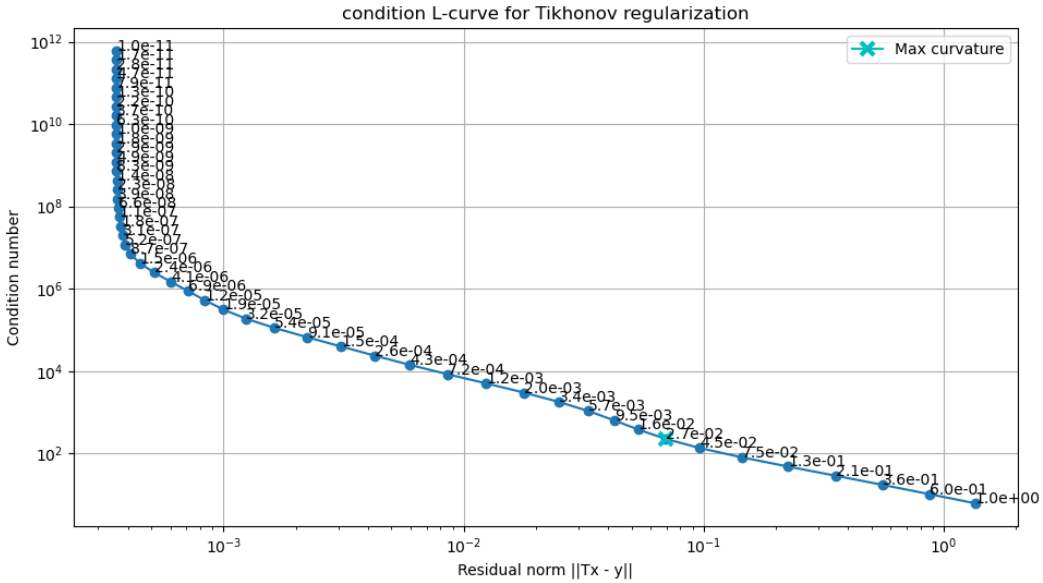

Figure 10: The condition L-curve corresponding to 0.001% noise level with max curvature choice rule.

5.2.2 TSVD results for the deconvolution problem

Noise Level (%)	Standard L-curve	Condition L-curve	True Optimal
10%	**0.2299**	0.3080	0.2299
1%	**0.1981**	0.2233	0.1981
0.1%	**0.1846**	0.2232	0.1846
0.01%	**0.1773**	0.1823	0.1669
0.001%	**0.1563**	0.1724	0.1563
0.0001%	**0.1563**	**0.1563**	0.1563

Table 3: TSVD: Errors $\frac{\|y_{\text{true}}-y_{\text{rec}}\|_2}{\|y_{\text{true}}\|_2}$ for the max curvature choice rule applied to both standard L-curve and the condition L-curve. The best performance for each noise level is in bold.

Noise Level (%)	Regińska	Quasi Optimal	GCV Optimal	True Optimal
10%	**0.2299**	0.3080	**0.2299**	0.2299
1%	**0.1981**	0.2233	0.2233	0.1981
0.1%	**0.1846**	0.2232	0.1903	0.1846
0.01%	**0.1773**	0.2233	**0.1773**	0.1669
0.001%	**0.1563**	0.2233	**0.1563**	0.1563
0.0001%	**0.1563**	0.2233	**0.1563**	0.1563

Table 4: TSVD: Errors $\frac{\|y_{\text{true}}-y_{\text{rec}}\|_2}{\|y_{\text{true}}\|_2}$ with Regińska, Quasi optimal and GCV as choice rule. The best performance for each noise level is in bold.

In Table 3 and 4 the errors obtained from the TSVD method applied to the deconvolution problem are presented. For this case we can see that max curvature applied to the standard L-curve together with Regińska obtains the true optimal for each noise level except 0.01%. GCV also performs quite well, but without the same consistency.

The max curvature applied to the condition L-curve and the quasi optimal do not preform as well as the other three. For the condition L-curve we have a strictly decreasing error which is good,

but compared to the best performing choice rules it is only for 0.0001% noise level where the method performs on the same level as the others. The quasi optimal choice rule is neither strictly decreasing nor close to the true optimal.

5.3 The Poisson problem with homogeneous boundary conditions

For the Poisson problem $Tx = y$ with $T = L$ being the Laplace operator, we have used the following function

$$x(t) = \frac{t^4}{12} - \frac{t^5}{10} + \frac{t^6}{30} - \frac{t}{60}. \tag{40}$$

Then the Laplace operator gives us the analytical solution for y to be

$$y(t) = t^2(1-t)^2. \tag{41}$$

We construct $x(t)$ on the discretized interval [0,1], add noise and try to regularize

$$T^{-1}y = x^\delta$$

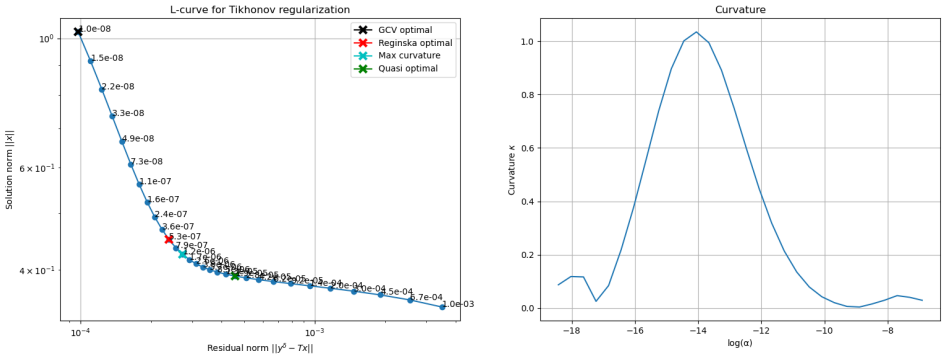

Figure 11: The Poisson problem with 0.1% noise results in an L-curve seen in the plot to the left. To the right is the curvature of the L-curve plotted as a function of the regularization parameter α.

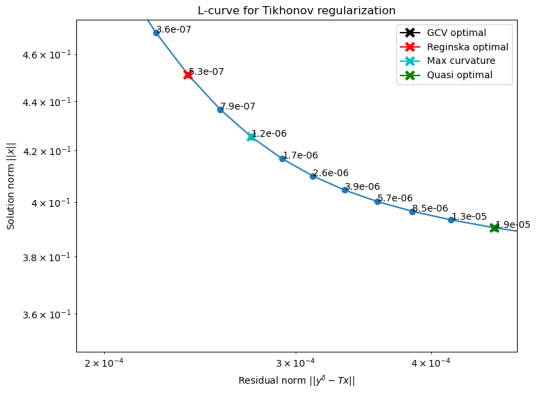

Figure 12: The L-curve from Figure 11 zoomed in at the vertex. The parameter choices are quite spread out.

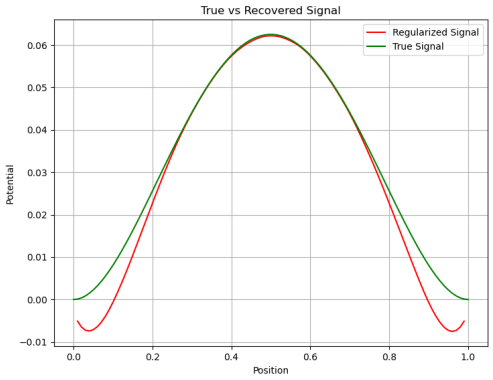

Figure 13: Recovered function for the Poisson problem with 1% noise with quasi optimal choice $\alpha = 1.9 \times 10^{-5}$.

5.3.1 Tikhonov regularization results for the Poisson problem

In Figure 11 we can see the resulting L-curve from Tikhonov regularization with 0.1% noise. The first obvious observation is that the GCV optimal choice is at the edge of our parameter interval,

implying that it is not very accurate. For the other three, we observe that they are quite far apart, with almost two orders of size difference.

The general observation for the L-curves resulting from different noise levels was that the curves did not have very sharp vertices. This might indicate that the choice rules will have a harder time selecting the optimal parameter. If we look at Table 5 and 6 we can see a large deviation in performance of the choice rules confirming this.

Noise Level (%)	**Standard L-curve**	**Condition L-curve**	**True Optimal**
10%	**0.1115**	0.1412	0.1112
1%	0.2981	**0.1044**	0.0463
0.1%	0.4192	**0.1039**	0.0445
0.01%	0.4133	**0.1039**	0.0444
0.001%	0.4127	**0.1039**	0.0444
0.0001%	0.4127	**0.1039**	0.0444

Table 5: Tikhonov regularization: Errors $\frac{\|y_{\text{true}} - y_{\text{rec}}\|_2}{\|y_{\text{true}}\|_2}$ for the max curvature choice rule applied to both standard L-curve and the condition L-curve. The best performance for each noise level is in bold.

Noise Level (%)	**Regińska**	**Quasi Optimal**	**GCV Optimal**	**True Optimal**
10%	0.1235	**0.1115**	1.9742	0.1112
1%	0.3560	**0.0758**	1.7509	0.04628
0.1%	0.5672	**0.0930**	2.3518	0.04448
0.01%	0.5592	**0.1190**	2.3229	0.04445
0.001%	0.5584	**0.1189**	2.3219	0.04445
0.0001%	0.5584	**0.1189**	2.3218	0.04445

Table 6: Tikhonov regularization for the Poisson problem: Errors $\frac{\|y_{\text{true}} - y_{\text{rec}}\|_2}{\|y_{\text{true}}\|_2}$ with Regińska, Quasi Optimal and GCV as choice rules. The best performance for each noise level is in bold.

Right away we observe that the condition L-curve and the quasi optimal choice stands out as the best performing choice rules. Still, no choice rule is close to the true optimal. We also see that all choice rules except from the condition L-curve results in an increase in the error for smaller noise levels. This contradicts the general assumption that a lower noise level leads to more accurate solutions.

An interesting observation for this particular case is that here the max curvature choice rule applied to the condition L-curve outperforms the choice from the standard L-curve. Compared to the deconvolution problem we now have an opposite performance. It is hard to draw any obvious conclusions from this, but one observation is that the max condition number for the Poisson problem is $\sim 10^7$, while the max condition number for the deconvolution problem is $\sim 10^9$ for the same parameter set. However, for the TSVD we see the same tendency for the condition numbers with $\sim 10^3$ for the Poisson problem and $\sim 10^{11}$ for the deconvolution problem, but the performance of the max curvature choice rule on the condition L-curve is very inaccurate. Further investigation of this behavior would call for more observations.

5.3.2 TSVD results for the Poisson problem

Noise Level (%)	Max Curvature	Max Curvature Condition	True Optimal
10%	**0.1739**	17.3044	0.1279
1%	**1.0756**	1.0529	0.1208
0.1%	**0.3309**	1.2678	0.1222
0.01%	**0.4642**	1.8473	0.1223
0.001%	**0.4636**	1.8460	0.1223
0.0001%	**0.4636**	1.8459	0.1223

Table 7: TSVD for the Poisson problem: Errors $\frac{\|y_{\text{true}}-y_{\text{rec}}\|_2}{\|y_{\text{true}}\|_2}$ for the max curvature choice rule applied to both standard L-curve and the condition L-curve. The best performance for each noise level is in bold.

Noise Level (%)	Regińska	Quasi Optimal	GCV Optimal	True Optimal
10%	**0.1739**	0.6992	3.4610	0.1279
1%	0.3289	**1.0529**	1.0529	0.1208
0.1%	0.4704	**0.1222**	2.2560	0.1222
0.01%	0.4642	**0.1223**	2.2400	0.1223
0.001%	0.4636	**0.1223**	2.2390	0.1223
0.0001%	0.4636	**0.1223**	2.2390	0.1223

Table 8: TSVD for the Poisson problem: Errors $\frac{\|y_{\text{true}}-y_{\text{rec}}\|_2}{\|y_{\text{true}}\|_2}$ with Regińska, Quasi Optimal and GCV as choice rules. The best performance for each noise level is in bold..

In Table 7 and 8 we see that the quasi optimal choice rule is the only choice rule that performs close to the true optimal, but only for noise levels less than or equal to 0.1%. For this particular case, not even the true optimal choice gives a strictly decreasing error and we observe few signs of stability for all other choice rules except the quasi optimal.

We observe for both the Tikhonov regularization and the TSVD of the Poisson problem that the true optimal error flattens out at 0.1% noise level. This problem might arise from a discretization error since we for each noise level compare with the analytical solution. In Table 9 we see results of the true optimal choice for both regularization methods applied to the Poisson problem with grid size increased from $N = 100$ to $N = 1000$. We can see that the result gets slightly better for the TSVD where the error decreases all the way down to 0.001% noise level. The precision of the Tikhonov solution does not increase although one could argue that the decrease from 0.1% to 0.01% noise level is slightly more significant.

Noise Level (%)	True Optimal Tikhonov	True Optimal TSVD
10%	0.06206	0.06656
1%	0.01189	0.03116
0.1%	0.004731	0.02715
0.01%	0.004666	0.02678
0.001%	0.004672	0.02674
0.0001%	0.004673	0.02715

Table 9: Errors $\frac{\|y_{\text{true}}-y_{\text{rec}}\|_2}{\|y_{\text{true}}\|_2}$ for the True Optimal choice rule with grid size increased from $N = 100$ to $N = 1000$.

6 Discussion

The results in Section 5 gives insight to how the parameter choice rules work on the deconvolution and the Poisson problem. The approach made in this project can be expanded to an arbitrary number of problems to get a larger basis for evaluating each choice rules performance. For now lets dicuss the perfomance observed for the problems we have presented.

6.1 Insights from the Deconvolution Problem

For the deconvolution problem, the standard L-curves generated by both regularization methods showed sharp vertex, which was favorable for identifying the optimal regularization parameter. The max curvature and Regińska choice rules often aligned, yielding residual errors close to the true optimal value, particularly at lower noise levels. This consistency demonstrates the robustness of these methods for problems where the L-curve vertex is well-defined.

The max curvature method on the condition L-curve did not produce equally good results. We saw that the method can produce good results

The clustering behavior observed in TSVD solutions (Figure 5 and 6) highlights the symmetry of the problem and the operator. By leveraging this symmetry to exclude odd eigenfunctions, we enhanced the precision of the parameter choices. Across all noise levels, the standard L-curve approach consistently outperformed the condition L-curve for TSVD. This suggests that the condition number-based method may not be as reliable for problems with higher condition numbers, such as the deconvolution problem.

6.2 Insights from the Poisson Problem

The results for the Poisson problem reveal a contrasting behavior. The L-curves for Tikhonov regularization often lacked sharp vertices, leading to a broader spread of parameter choices. Among the methods, the quasi-optimal and condition L-curve rules performed best, but even these were far from the true optimal for higher noise levels. Interestingly, the max curvature applied to the condition L-curve performed better for the Poisson problem than for the deconvolution problem. This discrepancy might be attributed to the lower condition number of the Laplace operator relative to the kernel matrix in the deconvolution problem.

For TSVD regularization of the Poisson problem, the quasi-optimal rule emerged as the most reliable, especially at low noise levels, consistently achieving errors close to the true optimal for TSVD. It also performed well compared to other methods for Tikhonov regularization, although it was quite far from the optimal choice. However, other methods, including GCV and Regińska, exhibited significant instability, with errors increasing at certain noise levels. This lack of stability underscores why the heuristic parameter choice rules need to be carefully considered.

We observed for both regularization methods that the error stopped decreasing for smaller noise levels. After decreasing the step size we saw that the TSVD method performed slightly better in terms of decreasing error with noise level, implying that the step size for our calculations were too big. However, the Tikhonov regularization did not show very significant signs of improvement after the change of step size. The performance improvement in TSVD with smaller step sizes suggests it is better suited for the given problem when the noise level decreases. It can adapt more effectively to the cleaner data, whereas Tikhonov regularization might not exploit the reduced noise to the same extent due to its inherent smoothing effect.

6.3 General Observations

A recurring theme in the results is the sensitivity of parameter choice rules to problem characteristics, noise levels, and the sharpness of the L-curve vertex. While methods like Regińska and

max curvature of the standard L-curve showed robust performance for well-defined problems with sharp vertices, they also showed to be less reliable for cases where the L-curves took a less distinct shape, such as those encountered in the Poisson problem.

Another key observation is the importance of preconditioning and point selection in enhancing the performance of parameter choice rules. For example, the restricted parameter set for TSVD, excluding odd eigenfunctions, improved the clustering and precision of solutions. This suggests that problem-specific preprocessing can significantly influence the accuracy of the choice rules.

6.4 Limitations and Future Directions

The results also highlight some limitations. No single parameter choice rule consistently achieved the true optimal across all scenarios, highlighting the need for hybrid approaches to get a wider perspective when choosing a parameter.

Additionally, the performance of condition L-curves varied significantly between problems, indicating that their use may be limited to specific problem classes or regularization methods.

Future work could explore adaptive strategies that combine multiple choice rules based on problem characteristics. Further investigation into the relationship between condition numbers, noise levels, and choice rule performance could provide deeper insights into the stability and accuracy of regularization methods. Finally, extending these analyses to other ill-posed problems and regularization techniques could broaden the applicability of these findings.

Bibliography

[1] A. B. Bakushinskiı. 'Remarks on choosing a regularization parameter using the quasioptimality and ration criterion'. In: *Zh. Vychisl. Mat. i Mat. Fiz.* 24 (1984), pp. 181–182.

[2] Christian Clason. *Regularization of Inverse Problems*. 2021. arXiv: 2001.00617 [math.FA]. URL: https://arxiv.org/abs/2001.00617.

[3] Carl Eckart and Gale Young. 'The Approximation of One Matrix by Another of Lower Rank'. In: *Psychometrika* 1.3 (1936), pp. 211–218. DOI: 10.1007/BF02288367. URL: https://doi.org/10.1007/BF02288367.

[4] Caterina Fenu, Lothar Reichel and Giuseppe Rodriguez. 'GCV for Tikhonov regularization via global Golub-Kahan decomposition'. In: *Numer. Linear Algebra Appl.* 23.3 (2016), pp. 467–484. ISSN: 1070-5325,1099-1506. DOI: 10.1002/nla.2034. URL: https://doi.org/10.1002/nla.2034.

[5] Per Christian Hansen and Dianne Prost O'Leary. 'The Use of the L-Curve in the Regularization of Discrete Ill-Posed Problems'. In: *SIAM Journal on Scientific Computing* 14.6 (1993), pp. 1487–1503. DOI: 10.1137/0914086. eprint: https://doi.org/10.1137/0914086. URL: https://doi.org/10.1137/0914086.

[6] Teresa Regińska. 'A Regularization Parameter in Discrete Ill-Posed Problems'. In: *SIAM Journal on Scientific Computing* 17.3 (1996), pp. 740–749. DOI: 10.1137/S1064827593252672. eprint: https://doi.org/10.1137/S1064827593252672. URL: https://doi.org/10.1137/S1064827593252672.

[7] Lothar Reichel and Giuseppe Rodriguez. 'Old and new parameter choice rules for discrete ill-posed problems'. In: *Numer. Algorithms* 63.1 (2013), pp. 65–87. ISSN: 1017-1398,1572-9265. DOI: 10.1007/s11075-012-9612-8. URL: https://doi.org/10.1007/s11075-012-9612-8.

[8] Yousef Saad. *Iterative Methods for Sparse Linear Systems*. Second. Society for Industrial and Applied Mathematics, 2003. DOI: 10.1137/1.9780898718003. eprint: https://epubs.siam.org/doi/pdf/10.1137/1.9780898718003. URL: https://epubs.siam.org/doi/abs/10.1137/1.9780898718003.