# Medical Imaging

Prof. Dr. Tobias Knopp

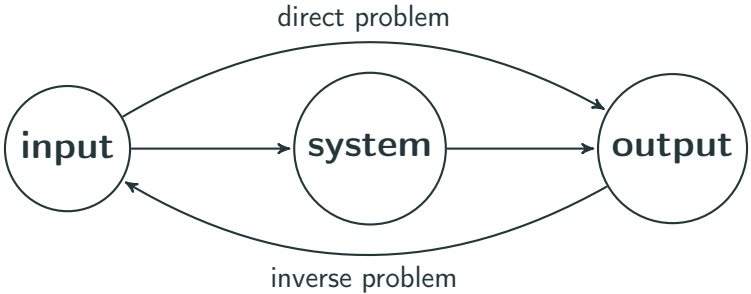
15. November 2022

Institut für Biomedizinische Bildgebung

# Inverse Problems

# Inverse Problems

**Recall**




**Remarks**

- So far we have learned that the existence and the uniqueness are important properties of an ill-posed problem.
- Next we will discuss what can further impact the solution negatively.
- In particular we discuss the *ill-posedness* of an inverse problem

# Definition

Hadamard: An inverse problem is *well-posed* if the following conditions are fulfilled:

- **Existence:** The problem must have a solution.
- **Uniqueness:** There must be only one solution to the problem.
- **Stability:** The solution must depend continuously on the data.

Otherwise, the problem is *ill-posed*.

**Remark**

The last condition of Hadamard is tailored towards continuous inverse problems. For discrete ill-posed problems one usually considers the condition number of the system matrix as a metric for the ill-posedness.

# Example - Existence

Consider the problem

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix} x = \begin{pmatrix} 1 \\ 2.2 \end{pmatrix}.$$

It has no solution since $x$ cannot be equal to 1 and 1.1 at the same time. But what we can do is to solve the **existence** issue by considering the least squares problem

$$\text{argmin}_x \left\| \begin{pmatrix} 1 \\ 2 \end{pmatrix} x - \begin{pmatrix} 1 \\ 2.2 \end{pmatrix} \right\|_2^2.$$

It has a unique solution $x = 1.08$.

# Example - Uniqueness

An example of an inverse problem where the uniqueness is not given is the underdetermined system

$$x_1 + x_2 = 1.$$

It has infinite solutions. One can enforce a solution by adding the additional restriction that the solution $\boldsymbol{x} = (x_1, x_2)^\mathsf{T}$ should have minimum 2-norm, i.e. $\|\boldsymbol{x}\|_2^2 = (x_1^2 + x_2^2)$ is minimum. In this case the unique solution is given by $x_1 = x_2 = \frac{1}{2}$.

# Example - Stability

The stability issue essentially means that small changes in the measurements can lead to very large changes in the solution.

Example:

$$\boldsymbol{A} = \begin{pmatrix} 0.16 & 0.1 \\ 0.17 & 0.11 \\ 2.02 & 1.29 \end{pmatrix} \quad \boldsymbol{A} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0.26 \\ 0.28 \\ 3.31 \end{pmatrix} =: \boldsymbol{b}$$

# Example - Stability

Now lets assume the measurement $\boldsymbol{b}$ is disturbed by noise, i.e. $\tilde{\boldsymbol{b}} = \boldsymbol{b} + \begin{pmatrix} 0.01 \\ -0.03 \\ 0.02 \end{pmatrix}$,

which is about 1% noise.

The least squares solution $\text{argmin}_{\boldsymbol{x}} \|\boldsymbol{A}\boldsymbol{x} - \tilde{\boldsymbol{b}}\|_2$ is in this case given by $\begin{pmatrix} 7.01 \\ -8.4 \end{pmatrix}$. It has nothing in common with the true solution $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$. The reason for this is that the matrix $\boldsymbol{A}$ is *ill-conditioned*. Ill-conditioned problems are *effectively underdetermined*.

# Example - Stability

For instance the vector $(-1.00, 1.57)^\mathsf{T}$ is almost in the nullspace of $\boldsymbol{A}$ since

$$\boldsymbol{A} \begin{pmatrix} -1.00 \\ 1.57 \end{pmatrix} = \begin{pmatrix} -0.0030 \\ 0.0027 \\ 0.0053 \end{pmatrix}$$

Hence one can add large amounts of this vector to a potential solution, with only minor impact on the right hand side. Minor perturbations of the right hand side (due to noise) thus can lead to large differences in the calculated solution.

# Discrete Linear Inverse Problems

# Discrete Linear Inverse Problems

A discrete and linear inverse problem can be formulated as a linear system of equations

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$

where $\boldsymbol{A} \in \mathbb{C}^{M \times N}$ is the system matrix, $\boldsymbol{x} \in \mathbb{C}^{N}$ is the input vector and $\boldsymbol{b} \in \mathbb{C}^{M}$ is the output vector.

# Discrete Linear Inverse Problems

In reality the measurement vector $\boldsymbol{b}$ is affected by (additive) noise $\boldsymbol{\varepsilon}$ such that one actually measures

$$\tilde{\boldsymbol{b}} = \boldsymbol{b} + \boldsymbol{\varepsilon}.$$

Hence the actual inverse problem reads

$$\boldsymbol{A}\boldsymbol{x} \approx \tilde{\boldsymbol{b}} = \boldsymbol{b} + \boldsymbol{\varepsilon}.$$

Note that that the relation between the model $\boldsymbol{A}\boldsymbol{x}$ and the measurement $\tilde{\boldsymbol{b}}$ is thus only a approximation.

# Least Squares Approach

The standard approach to solve an inconsistent linear system of equation is to minimize the difference between $\boldsymbol{A}\boldsymbol{x}$ and $\tilde{\boldsymbol{b}}$. This is named the residual vector

$$\boldsymbol{r} := \boldsymbol{A}\boldsymbol{x} - \tilde{\boldsymbol{b}}$$

and the least squares approach is to minimized the norm of the residual vector:

$$\boldsymbol{x}_{\mathsf{LS}} = \underset{\boldsymbol{x}}{\mathsf{argmin}} \|\boldsymbol{r}\|_2^2 = \underset{\boldsymbol{x}}{\mathsf{argmin}} \|\boldsymbol{A}\boldsymbol{x} - \tilde{\boldsymbol{b}}\|_2^2. \tag{1}$$

# Least Squares Approach

The least squares problem can be reformulated as a linear system of equations:

**Theorem**
*The least squares problem* (1) *is equivalent to the normal equation of first kind*

$$\boldsymbol{A}^{\mathsf{H}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathsf{H}}\tilde{\boldsymbol{b}}.$$

*and always has a solution.*

# Least Squares Approach - Proof Equivalence

To simplify the proof we assume that $\boldsymbol{A}$, $\boldsymbol{x}$, and $\tilde{\boldsymbol{b}}$ are real. Let $\boldsymbol{x}_*$ satisfy $\boldsymbol{A}^{\mathsf{H}}\boldsymbol{A}\boldsymbol{x}_* = \boldsymbol{A}^{\mathsf{H}}\tilde{\boldsymbol{b}}$ then for any $\boldsymbol{x} \in \mathbb{R}^N$

$$
\begin{aligned}
\|\boldsymbol{A}\boldsymbol{x} - \tilde{\boldsymbol{b}}\|_2^2 =& \|\boldsymbol{A}(\boldsymbol{x}_* + \boldsymbol{x} - \boldsymbol{x}_*) - \tilde{\boldsymbol{b}}\|_2^2 \\
=& (\boldsymbol{A}(\boldsymbol{x}_* + \boldsymbol{x} - \boldsymbol{x}_*) - \tilde{\boldsymbol{b}})^{\mathsf{T}}(\boldsymbol{A}(\boldsymbol{x}_* + \boldsymbol{x} - \boldsymbol{x}_*) - \tilde{\boldsymbol{b}}) \\
=& ((\boldsymbol{A}\boldsymbol{x}_* - \tilde{\boldsymbol{b}}) + \boldsymbol{A}(\boldsymbol{x} - \boldsymbol{x}_*))^{\mathsf{T}}((\boldsymbol{A}\boldsymbol{x}_* - \tilde{\boldsymbol{b}}) + \boldsymbol{A}(\boldsymbol{x} - \boldsymbol{x}_*)) \\
=& (\boldsymbol{A}\boldsymbol{x}_* - \tilde{\boldsymbol{b}})^{\mathsf{T}}(\boldsymbol{A}\boldsymbol{x}_* - \tilde{\boldsymbol{b}}) + (\boldsymbol{A}(\boldsymbol{x} - \boldsymbol{x}_*))^{\mathsf{T}}(\boldsymbol{A}(\boldsymbol{x} - \boldsymbol{x}_*)) \\
& + 2(\boldsymbol{x} - \boldsymbol{x}_*)^{\mathsf{T}} \underbrace{(\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{x}_* - \boldsymbol{A}^{\mathsf{T}}\tilde{\boldsymbol{b}})}_{=\boldsymbol{0}} \\
=& \|\boldsymbol{A}\boldsymbol{x}_* - \tilde{\boldsymbol{b}}\|_2^2 + \|\boldsymbol{A}(\boldsymbol{x} - \boldsymbol{x}_*)\|_2^2 \\
\geq& \|\boldsymbol{A}\boldsymbol{x}_* - \tilde{\boldsymbol{b}}\|_2^2
\end{aligned}
$$

Thus $\boldsymbol{x}_*$ is a solution to the least squares problem (1).

# Least Squares Approach - Proof Existence

Now suppose $\boldsymbol{x}_*$ is a solution to the least squares problem (1). The function

$$f(\boldsymbol{x}) = \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2 \tag{2}$$

thus has an extremum at $\boldsymbol{x}_*$. Since $f$ is differentiable, the gradient is zero at $\boldsymbol{x}_*$, i.e. $\nabla f(\boldsymbol{x}_*) = \boldsymbol{0}$. Lets calculate the gradient

$$\begin{aligned}\frac{\partial f}{\partial x_j} f(\boldsymbol{x}) &= \frac{\partial f}{\partial x_j} \sum_{m=1}^{M} \left( \sum_{n=1}^{N} A_{m,n} x_n - b_m \right)^2 \\ &= 2 \sum_{m=1}^{M} \left( \sum_{n=1}^{N} A_{m,n} x_n - b_m \right) A_{m,j}\end{aligned}$$

and thus

$$\nabla f(\boldsymbol{x}) = \boldsymbol{A}^{\mathsf{T}}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})$$

# Least Squares Approach - Proof Equivalence

Consequently we have

$$\nabla f(\boldsymbol{x}_*) = \boldsymbol{A}^\mathsf{T}(\boldsymbol{A}\boldsymbol{x}_* - \boldsymbol{b}) = \boldsymbol{0}$$

and thus the minimizer of the least squares problem is a solution of the normal equation.

# Least Squares Approach - Proof Existence

The existence of a solution is guaranteed since

$$\lim_{\|\boldsymbol{x}\| \to \infty} f(\boldsymbol{x}) = \infty \tag{3}$$

and we can thus find a compact subset $U \subset \mathbb{R}^N$ where

$$f(\boldsymbol{x}) \leq f(\boldsymbol{0}) = \|\boldsymbol{b}\|_2^2$$

Since $f$ is continuous and $U$ is compact, the function will have a minimum in $U$ according to the extreme value theorem.

# Least Squares Approach - Unique Solution

**Lemma**
*If* $rank(\boldsymbol{A}) = N$ *the solution of the normal equation*

$$\boldsymbol{A}^{\mathsf{H}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathsf{H}}\tilde{\boldsymbol{b}}.$$

*is unique.*

For real matrices we have $rank(\boldsymbol{A}) = rank(\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A})$. Hence $\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}$ is a square $N \times N$ matrix with full rank. For such matrices an inverse always exists. The unique solution is given by

$$\boldsymbol{x}_* = (\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathsf{T}}\tilde{\boldsymbol{b}} \tag{4}$$

The matrix $\boldsymbol{A}^{+} := (\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathsf{T}}$ is the so-called *pseudoinverse*.

# Why does the Least Squares Approach often fail?

Let us consider the residual vector of the true solution $\boldsymbol{x}$:

$$\boldsymbol{r} = \boldsymbol{A}\boldsymbol{x} - \tilde{\boldsymbol{b}} = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} - \boldsymbol{\varepsilon} = -\boldsymbol{\varepsilon}.$$

Thus the true solution has a **non-zero** residual. The least squares approach does, however, minimize the residual even below the "optimal" residual of $\|\boldsymbol{\varepsilon}\|_2$. It will in particular find solutions with

$$\|\boldsymbol{r}\|_2 < \|\boldsymbol{\varepsilon}\|_2$$

In this case the solution is fitted to the noise leading to undesired results.

# How can the noise amplification be quantified?

**Theorem**
*Let* $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ *and* $\boldsymbol{A}\tilde{\boldsymbol{x}} = \boldsymbol{b} + \varepsilon =: \tilde{\boldsymbol{b}}$. *Then the following inequality holds*

$$\frac{\|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|_2}{\|\boldsymbol{x}\|_2} \leq \mathit{cond}(\boldsymbol{A}) \frac{\|\boldsymbol{b} - \tilde{\boldsymbol{b}}\|_2}{\|\boldsymbol{b}\|_2}$$

Thus an error in the measurement data can be amplified by the factor $\mathrm{cond}(\boldsymbol{A})$, which is the condition number of $\boldsymbol{A}$.

# Regularization

# Regularization

In order to handle ill-posed problems where the least squares method fails one applies so-called *regularization techniques* that stabilize the solution. In particular the linear system is exchanged with a similar system that is better conditioned.

**Definition**
The Tikhonov regularization technique considers the following optimization problem

$$\boldsymbol{x}_{\mathsf{LS}}^{\lambda} = \underset{\boldsymbol{x}}{\mathsf{argmin}} \|\boldsymbol{A}\boldsymbol{x} - \tilde{\boldsymbol{b}}\|_2^2 + \lambda \|\boldsymbol{x}\|_2^2 \tag{5}$$

Here, $\lambda$ is the so-called regularization parameter

**Theorem**
*The regularized least squares problem can be equivalently solved by the regularized normal equation*

$$(\boldsymbol{A}^{\mathsf{H}}\boldsymbol{A} + \lambda \boldsymbol{I})\boldsymbol{x} = \boldsymbol{A}^{\mathsf{H}}\tilde{\boldsymbol{b}}.$$

# Regularization - Proof Equivalence

Reformulate the minimization problem

$$\begin{aligned}\underset{\boldsymbol{x}}{\operatorname{argmin}} \|\boldsymbol{A}\boldsymbol{x} - \tilde{\boldsymbol{b}}\|_2^2 + \lambda\|\boldsymbol{x}\|_2^2 &= \underset{\boldsymbol{x}}{\operatorname{argmin}} \left\| \begin{pmatrix} \boldsymbol{A}\boldsymbol{x} - \tilde{\boldsymbol{b}} \\ \sqrt{\lambda}\boldsymbol{I}\boldsymbol{x} \end{pmatrix} \right\|_2^2 \\ &= \underset{\boldsymbol{x}}{\operatorname{argmin}} \left\| \begin{pmatrix} \boldsymbol{A} \\ \sqrt{\lambda}\boldsymbol{I} \end{pmatrix} \boldsymbol{x} - \begin{pmatrix} \tilde{\boldsymbol{b}} \\ \boldsymbol{0} \end{pmatrix} \right\|_2^2\end{aligned}$$

into the standard least squares form. The normal equation of this least squares problem is given by

$$\begin{pmatrix} \boldsymbol{A} \\ \sqrt{\lambda}\boldsymbol{I} \end{pmatrix}^{\mathsf{T}} \begin{pmatrix} \boldsymbol{A} \\ \sqrt{\lambda}\boldsymbol{I} \end{pmatrix} \boldsymbol{x} = \begin{pmatrix} \boldsymbol{A} \\ \sqrt{\lambda}\boldsymbol{I} \end{pmatrix}^{\mathsf{T}} \begin{pmatrix} \tilde{\boldsymbol{b}} \\ \boldsymbol{0} \end{pmatrix}$$

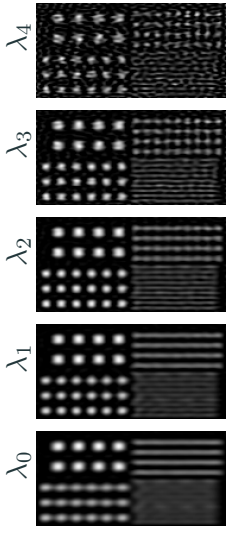
which is equivalent to

$$(\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A} + \lambda\boldsymbol{I})\boldsymbol{x} = \boldsymbol{A}^{\mathsf{T}}\tilde{\boldsymbol{b}}.$$

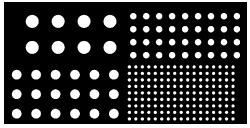
# Regularization

- The regularization term (or penalty term) $\lambda \|\boldsymbol{x}\|_2^2$ controls that the solution gets not too large.
- Regularization reduces noise in the calculated solution, i.e. it smoothes $\boldsymbol{x}_{\mathrm{LS}}^{\lambda}$.
- Regularization does, however, also introduces a *bias* (i.e. systematic "global" deviation) between $\boldsymbol{x}$ and $\boldsymbol{x}_{\mathrm{LS}}^{\lambda}$.
- In practice one has to trade of between a too noise and a too smooth solution by appropriate choice of lambda.

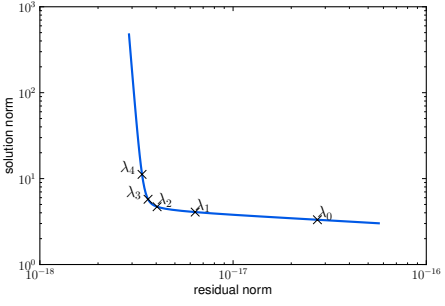
# Influence of the regularization parameter




Original phantom

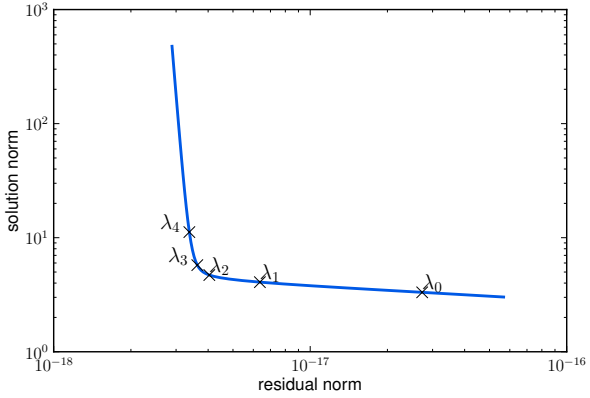


L curve



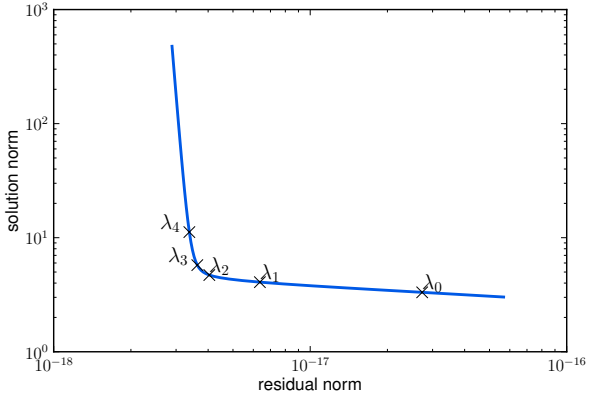
# How to Choose the Regularization Parameter?

Choosing $\lambda$ is a challenging problem in practice. What is often done is to compute various solutions for different $\lambda$ and plot the solution norm $\|\boldsymbol{x}_{\mathsf{LS}}^{\lambda}\|_2$ versus the residual norm $\|\boldsymbol{A}\boldsymbol{x}_{\mathsf{LS}}^{\lambda} - \tilde{\boldsymbol{b}}\|_2^2$



# How to Choose the Regularization Parameter?

This curve typically has an L-curved shape and the best promise is usually found in the corner of the L. The corner is that point where the residual is close to minimal but the solution norm $\|\boldsymbol{x}_{\mathrm{LS}}^{\lambda}\|_2$ is still not "blown up" due noise amplification.



# Singular Value Decomposition

# Singular Value Decomposition

The singular value decomposition is a matrix decomposition that allows to

- solve a linear system of equations
- apply regularization efficiently
- understand ill-posed problems, i.e. do a fine grained analysis of the ill-posedness of an inverse problem

# Singular Value Decomposition

As before we consider the linear system

$$\boldsymbol{A}\boldsymbol{x} \approx \tilde{\boldsymbol{b}} = \boldsymbol{b} + \boldsymbol{\varepsilon} \quad \text{where} \quad \boldsymbol{A} \in \mathbb{C}^{M \times N}$$

**Theorem**
Any $M \times N$ matrix can be decomposed into

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathsf{H}}$$

where $\boldsymbol{U} \in \mathbb{C}^{M \times r}$ and $\boldsymbol{V} \in \mathbb{C}^{N \times r}$ are rectangular with orthogonal columns and $r = \mathsf{rank}(\boldsymbol{A}) \leq \min(M, N)$. The diagonal matrix

$$\boldsymbol{\Sigma} = \mathsf{diag}(\boldsymbol{\sigma}) \in \mathbb{R}_{+}^{r \times r}$$

contains the *singular values* $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_r)^{\mathsf{T}}$ in descending order.

(without proof)

### Remarks

- The SVD is not unique but at least one SVD exists
- Calculating the SVD has a complexity of $\mathcal{O}(N^3)$ if $M \approx N$. It thus has the same complexity as Gaussian Elimination. The constant infront of the $N^3$ factor is large though
- We have defined the *compact* SVD. There is also a non-compat version

$$\boldsymbol{A} = \tilde{\boldsymbol{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\boldsymbol{V}}^{\mathsf{H}}$$

where $\tilde{\boldsymbol{U}} \in \mathbb{C}^{M \times M}$ and $\tilde{\boldsymbol{V}} \in \mathbb{C}^{N \times N}$ are unitary ($\tilde{\boldsymbol{U}}^{\mathsf{H}}\tilde{\boldsymbol{U}} = \boldsymbol{I}$, $\tilde{\boldsymbol{V}}^{\mathsf{H}}\tilde{\boldsymbol{V}} = \boldsymbol{I}$) and $\boldsymbol{\Sigma} = \mathbb{R}^{M \times N}$ is a rectangular diagonal matrix that may contain zero entries at its main diagonal.

# Solving Linear Systems

In the following we consider $r = N$, $M \geq N$, i.e. an overdetermined system, where $\boldsymbol{V}$ is square and unitary, i.e. the inverse exists.

$$
\begin{aligned}
&\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\
\Rightarrow &\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathsf{H}}\boldsymbol{x} = \boldsymbol{b} \\
\Rightarrow &\underbrace{\boldsymbol{U}^{\mathsf{H}}\boldsymbol{U}}_{\boldsymbol{I}}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathsf{H}}\boldsymbol{x} = \boldsymbol{U}^{\mathsf{H}}\boldsymbol{b} \\
\Rightarrow &\underbrace{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}}_{\boldsymbol{I}}\boldsymbol{V}^{\mathsf{H}}\boldsymbol{x} = \boldsymbol{\Sigma}^{-1}\boldsymbol{U}^{\mathsf{H}}\boldsymbol{b} \\
\Rightarrow &\boldsymbol{x} = \boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^{\mathsf{H}}\boldsymbol{b}
\end{aligned}
$$

**Remark:** The matrix $\boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^{\mathsf{H}}$ is the same as the pseudo inverse $\boldsymbol{A}^{+} = (\boldsymbol{A}^{\mathsf{H}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathsf{H}}$.

$$
\begin{aligned}
\boldsymbol{A}^{+} &= (\boldsymbol{A}^{\mathsf{H}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathsf{H}} \\
&= (\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}^{\mathsf{H}}\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathsf{H}})^{-1}\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}^{\mathsf{H}} \\
&= (\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathsf{H}})^{-1}\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}^{\mathsf{H}} \\
&= (\boldsymbol{V}\boldsymbol{\Sigma}^{2}\boldsymbol{V}^{\mathsf{H}})^{-1}\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}^{\mathsf{H}} \\
&= \underbrace{(\boldsymbol{V}^{\mathsf{H}})^{-1}}_{\boldsymbol{V}}\boldsymbol{\Sigma}^{-2}\boldsymbol{V}^{-1}\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}^{\mathsf{H}} \\
&= \boldsymbol{V}\boldsymbol{\Sigma}^{-2}\boldsymbol{\Sigma}\boldsymbol{U}^{\mathsf{H}} \\
&= \boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^{\mathsf{H}}
\end{aligned}
$$

# Solving Linear Systems

**Theorem:** The solution $\boldsymbol{x} = \boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^{\mathsf{H}}$ can also be expressed as

$$\boldsymbol{x} = \sum_{i=1}^{r} \frac{\boldsymbol{U}_{\cdot,i}^{\mathsf{H}}\boldsymbol{b}}{\sigma_i}\boldsymbol{V}_{\cdot,i}$$

**Proof:**

$$\begin{aligned}
&\begin{pmatrix} \boldsymbol{V}_{\cdot,1} & \cdots & \boldsymbol{V}_{\cdot,r} \end{pmatrix} \begin{pmatrix} \frac{1}{\sigma_1} & & \\ & \ddots & \\ & & \frac{1}{\sigma_r} \end{pmatrix} \begin{pmatrix} \boldsymbol{U}_{\cdot,1}^{\mathsf{H}} \\ \vdots \\ \boldsymbol{U}_{\cdot,r}^{\mathsf{H}} \end{pmatrix} \boldsymbol{b} \\
&= \begin{pmatrix} \boldsymbol{V}_{\cdot,1} & \cdots & \boldsymbol{V}_{\cdot,r} \end{pmatrix} \begin{pmatrix} \frac{1}{\sigma_1}\boldsymbol{U}_{\cdot,1}^{\mathsf{H}}\boldsymbol{b} \\ \vdots \\ \frac{1}{\sigma_r}\boldsymbol{U}_{\cdot,r}^{\mathsf{H}}\boldsymbol{b} \end{pmatrix} \\
&= \sum_{i=1}^{r} \frac{\boldsymbol{U}_{\cdot,i}^{\mathsf{H}}\boldsymbol{b}}{\sigma_i}\boldsymbol{V}_{\cdot,i}
\end{aligned}$$

# SVD for Ill-Posed Problems

Let us have a look at the SVD based solution in case of a noisy linear system. If we insert $\tilde{\boldsymbol{b}} = \boldsymbol{b} + \boldsymbol{\varepsilon}$ then we have

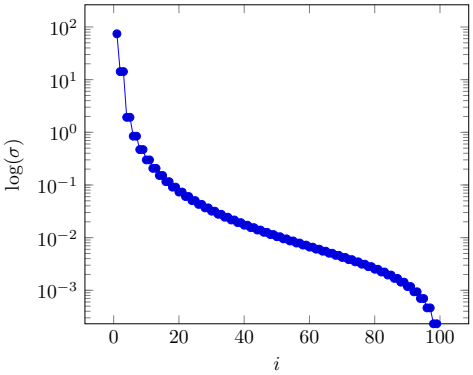
$$\begin{aligned}\tilde{\boldsymbol{x}} &= \sum_{i=1}^{r} \frac{\boldsymbol{U}_{\cdot,i}^{\mathsf{H}}(\boldsymbol{b} + \boldsymbol{\varepsilon})}{\sigma_i} \boldsymbol{V}_{\cdot,i} \\ &= \sum_{i=1}^{r} \frac{\boldsymbol{U}_{\cdot,i}^{\mathsf{H}}\boldsymbol{b}}{\sigma_i} \boldsymbol{V}_{\cdot,i} + \sum_{i=1}^{r} \frac{\boldsymbol{U}_{\cdot,i}^{\mathsf{H}}\boldsymbol{\varepsilon}}{\sigma_i} \boldsymbol{V}_{\cdot,i}\end{aligned}$$

The numerator in the right sum has a constant standard deviation independently of $r$. The denominator, however, decreases with increasing $i$ for ill-posed problems.

$\rightarrow$ The small singular values ($\frac{1}{\sigma}$) amplify the noise.

# SVD for Ill-Posed Problems

Example singular values of a discrete convolution matrix




```
using PGFPlots, ToeplitzMatrices,
      LinearAlgebra
N = 100
t = range(-1,1,length=N)
sigma = 1
a = exp.(-(t.^2)./sigma)
A = Circulant(a) |> Matrix
U,S,V = svd(A)


p = Plots.Linear(1:N, S)
p = Axis(p, ymode="log", ymin=S[end-1],
            xlabel=L"$i$", ylabel=L"log($\sigma$)")
```

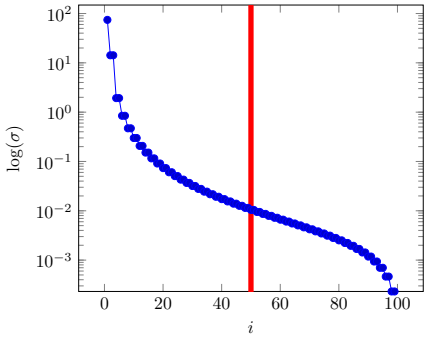
**Remark:** Recall that the noise amplification is bounded by the condition number $\text{cond}(\boldsymbol{A}) = \frac{\sigma_1}{\sigma_r}$.

# Truncated SVD

One regularization method is to neglect small singular values that are responsible for the noise amplification:

$$\tilde{\boldsymbol{x}} = \sum_{i=1}^{\alpha} \frac{\boldsymbol{U}_{\cdot,i}^{\mathsf{H}} \tilde{\boldsymbol{b}}}{\sigma_i} \boldsymbol{V}_{\cdot,i}$$

where $\alpha$ with $1 \leq \alpha \leq r$ is the truncation parameter that acts like a regularization parameter.



# Truncated SVD

**Remarks**

- The truncation suppresses the noise amplification but it smoothes the solution (like Tikhonov regularization).
- The truncation is a filter (rect function) that acts on the singular values. Since it lets small singular values pass, it is a *low-pass filter*.

# Convolution Matrix

Note that if $\boldsymbol{A}$ is a (periodic) convolution matrix, we have

$$\boldsymbol{A} = \boldsymbol{F}\boldsymbol{\Sigma}\boldsymbol{F}^{\mathsf{H}}$$

where $\boldsymbol{F} = \boldsymbol{U} = \boldsymbol{V}$ is the discrete Fourier matrix and the singular values in $\boldsymbol{\Sigma}$ contain the transfer function (Fourier coefficients of the convolution kernel).

Filtering of singular values is thus directly related to the filtering we considered for Fourier transformation. In fact, a Fourier transform can be defined for other bases than the trigonometric functions.

# Tikhonov Regularization with the SVD

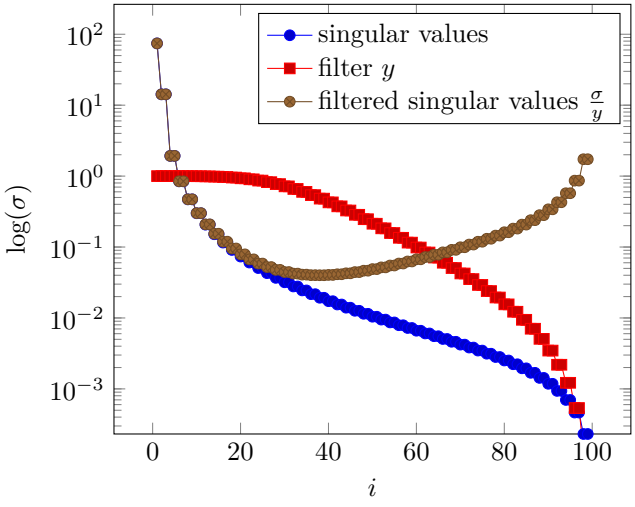
Next, let us investigate the relation between Tikhonov regularization and the SVD.

$$
\begin{aligned}
\boldsymbol{x}_\lambda &= (\boldsymbol{A}^{\mathsf{H}}\boldsymbol{A} + \lambda \boldsymbol{I})^{-1}\boldsymbol{A}^{\mathsf{H}}\boldsymbol{b} \\
&= (\boldsymbol{V}\boldsymbol{\Sigma}^2\boldsymbol{V}^{\mathsf{H}} + \lambda \boldsymbol{I})^{-1}\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}^{\mathsf{H}}\boldsymbol{b} \\
&= (\boldsymbol{V}(\boldsymbol{\Sigma}^2 + \lambda \boldsymbol{I})\boldsymbol{V}^{\mathsf{H}})^{-1}\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}^{\mathsf{H}}\boldsymbol{b} \\
&= (\boldsymbol{V}^{\mathsf{H}})^{-1}(\boldsymbol{\Sigma}^2 + \lambda \boldsymbol{I})^{-1}\boldsymbol{V}^{-1}\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}^{\mathsf{H}}\boldsymbol{b} \\
&= \boldsymbol{V}\underbrace{(\boldsymbol{\Sigma}^2 + \lambda \boldsymbol{I})^{-1}\boldsymbol{\Sigma}}_{\mathsf{diag}\left(\left(\frac{\sigma_i}{\sigma_i^2+\lambda}\right)_{i=1}^{r}\right)}\boldsymbol{U}^{\mathsf{H}}\boldsymbol{b} \\
&= \sum_{i=1}^{r} \frac{\sigma_i \boldsymbol{U}_{\cdot,i}^{\mathsf{H}}\boldsymbol{b}}{\sigma_i^2 + \lambda}\boldsymbol{V}_{\cdot,i} = \sum_{i=1}^{r} y_i \frac{\boldsymbol{U}_{\cdot,i}^{\mathsf{H}}\boldsymbol{b}}{\sigma_i}\boldsymbol{V}_{\cdot,i}
\end{aligned}
$$

where $y_i = \frac{\sigma_i^2}{\sigma_i^2+\lambda}$ are the filter factors.

# Tikhonov Regularization with the SVD

The following figure showcases the effect of the filter on the singular values.




Compare this to the truncated SVD where the singular values were cutted.

Tikhonov regularization also has an effective cutoff (at about $i = 40$ in this example) but the transition is smoother.

# Tikhonov Regularization with the SVD

**Remarks**

- Explicit version of Tikhonov regularization
- SVD takes $\mathcal{O}(N^3)$ but solutions for different $\lambda$ can be calculated in $\mathcal{O}(N^2)$

$\rightarrow$ One time cost that pays off.

# L-Curve with the SVD

The L-curve needs for $\gamma = 1, \ldots, \Gamma$ the values

$$(\|\boldsymbol{A}\boldsymbol{x}_{\lambda_\gamma} - \tilde{\boldsymbol{b}}\|_2^2, \|\boldsymbol{x}_{\lambda_\gamma}\|_2^2)$$

Using a linear solver without matrix decomposition (e.g. Gaussian elimination) this requires $\mathcal{O}(N^3\Gamma)$ operations.

Using a matrix decomposition technique (e.g. SVD) this requires $\mathcal{O}(N^2\Gamma)$ operations.

But with the SVD we can obtain these values even faster.

# L-Curve with the SVD

Due to the orthogonality of $\boldsymbol{V}$ we have

$$\|\boldsymbol{A}\boldsymbol{x}_{\lambda_\gamma} - \tilde{\boldsymbol{b}}\|_2^2 = \sum_{i=1}^{r} \left| \frac{\lambda_\gamma}{\sigma_i^2 + \lambda_\gamma} \boldsymbol{U}_{\cdot,i}^{\mathsf{H}} \tilde{\boldsymbol{b}} \right|^2$$

$$\|\boldsymbol{x}_{\lambda_\gamma}\|_2^2 = \sum_{i=1}^{r} \left| \frac{\sigma_i}{\sigma_i^2 + \lambda_\gamma} \boldsymbol{U}_{\cdot,i}^{\mathsf{H}} \tilde{\boldsymbol{b}} \right|^2$$

One can observe that $\boldsymbol{U}_{\cdot,i}^{\mathsf{H}} \tilde{\boldsymbol{b}}$ is independent of $\lambda$ an thus can be precomputed once. In total, an L-curve using the SVD thus can be obtained in $\mathcal{O}(N^2 + \Gamma N)$ steps. For $\Gamma \in \mathcal{O}(N)$ this is the same complexity as calculating the solution $\boldsymbol{x}_\lambda$. Furthermore, the matrix-vector operation $\boldsymbol{U}^{\mathsf{H}} \tilde{\boldsymbol{b}}$ needs only to be computed once.

# Summary

- Inverse problems are hard
- They are prone to noise amplification
- They require special treatment to yield a satisfying solution
- One can reduce the noise amplification but has to live with a bias
- The come up in many real-live problems