Ill–Posed Inverse Problems in Image Processing

Introduction, Structured matrices, Spectral filtering, Regularization, Noise revealing

I. Hnětynková[1], M. Plešinger[2], Z. Strakoš[3]

hnetynko@karlin.mff.cuni.cz, martin.plesinger@sam.math.ethz.ch, strakos@cs.cas.cz

[1,3]Faculty of Mathematics and Phycics, Charles University, Prague
[2]Seminar of Applied Mathematics, Dept. of Math., ETH Zürich
[1,2,3]Institute of Computer Science, Academy of Sciences of the Czech Republic

SNA '11, January 24—28

Recapitulation of Lecture I

Linear system

Consider the problem

$$Ax = b, \quad b = b^{\text{exact}} + b^{\text{noise}}, \quad A \in \mathbb{R}^{N\times N}, \quad x,\, b \in \mathbb{R}^N,$$

where

- A is a discretization of a smoothing operator,
- singular values of A decay,
- singular vectors of A represent increasing frequencies,
- b^{exact} is smooth and satisfies the discrete Picard condition,
- b^{noise} is unknown white noise,

$$\|b^{\text{exact}}\| \gg \|b^{\text{noise}}\|, \qquad \text{but} \qquad \|A^{-1}b^{\text{exact}}\| \ll \|A^{-1}b^{\text{noise}}\|.$$

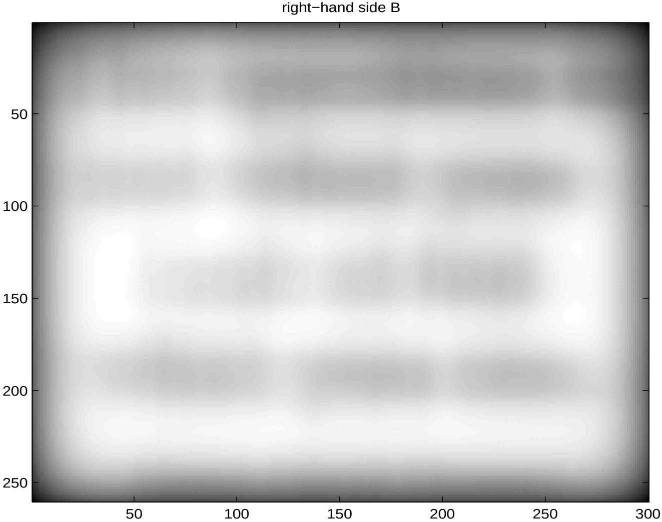
We want to approximate

$$x^{\text{exact}} = A^{-1}b^{\text{exact}}.$$

Recapitulation of Lecture I

Right-hand side

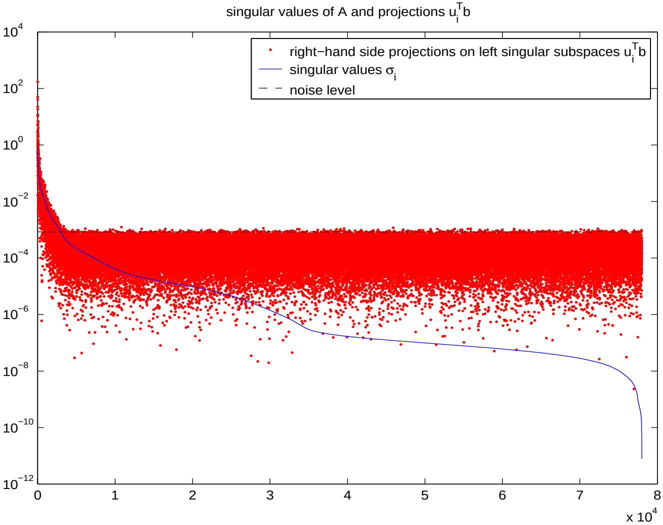
Smooth right-hand side (including noise):



Recapitulation of Lecture I

Violation of the discrete Picard condition

Violation of the dicrete Picard condition in the noisy b:



Recapitulation of Lecture I

Solution

Using SVD $A = U\Sigma V^T$ the **filtered** solution is

$$x^{\text{filtered}} = \sum_{j=1}^{N} \phi_j \frac{u_j^T b}{\sigma_j} v_j, \qquad x^{\text{filtered}} = V\Phi\Sigma^{-1}U^T b,$$

where $\Phi = \mathrm{diag}(\phi_1, \ldots, \phi_N)$. Particularly in the image deblurring problem

$$X^{\text{filtered}} = \sum_{j=1}^{N} \phi_j \frac{u_j^T \operatorname{vec}(B)}{\sigma_j} V_j, \qquad \text{where } V_j \text{ are singular images.}$$

The filter factors ϕ_j are given by some filter function

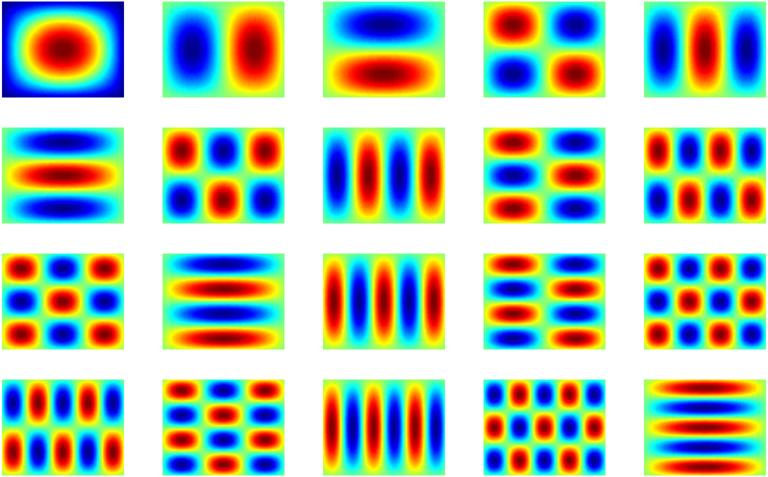
$$\phi_j = \phi(j, A, b, \ldots),$$

for $\phi_j = 1$, $j = 1, \ldots, N$, we get the **naive solution**.

Recapitulation of Lecture I

Singular images

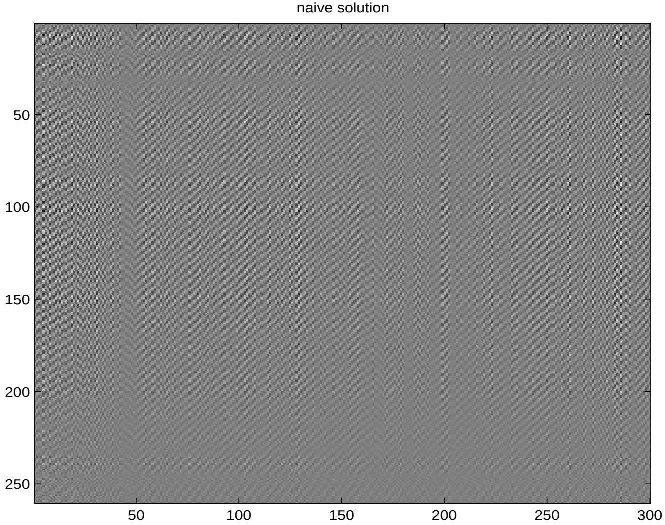
Singular images V_j (Gaußian blur, zero BC, artificial colors):



Recapitulation of Lecture I

Naive solution

The naive solution is dominated by high-frequency noise:



Outline of the tutorial

- **Lecture I—Problem formulation:**

  Mathematical model of blurring, System of linear algebraic equations, Properties of the problem, Impact of noise.

- **Lecture II—Regularization:**

  Basic regularization techniques (TSVD, Tikhonov), Criteria for choosing regularization parameters, Iterative regularization, Hybrid methods.

- **Lecture III—Noise revealing:**

  Golub-Kahan iteratie bidiagonalization and its properties, Propagation of noise, Determination of the noise level, Noise vector approximation, Open problems.

Outline of Lecture II

- **5. Basic regularization techniques:**

  Truncated SVD, Selective SVD, Tikhonov regularization.
- **6. Choosing regularization parameters:**

  Discrepancy principle, Generalized cross validation, L-curve, Normalized cumulative periodogram.
- **7. Iterative regularziation:**

  Landweber iteration, Cimmino iteration, Kaczmarz's method, Projection methods, Regularizing Krylov subspace iterations.
- **8. Hybrid methods:**

  Introduction, Projection methods with inner Tikhonov regularization.

5. Basic regularization techniques

5. Basic regularization techniques

Truncated SVD

The simplest regularization technique is the **truncated SVD (TSVD)**. Noise affects x^{naive} through the components corresponding to the smalest singular values,

$$x^{\text{naive}} = \underbrace{\sum_{j=1}^{k} \frac{u_j^T b}{\sigma_j} v_j}_{\text{data dominated}} + \underbrace{\sum_{j=k+1}^{N} \frac{u_j^T b}{\sigma_j} v_j}_{\text{noise dominated}}$$

$$= \sum_{j=1}^{k} \frac{u_j^T b^{\text{exact}}}{\sigma_j} v_j + \sum_{j=1}^{k} \frac{u_j^T b^{\text{noise}}}{\sigma_j} v_j + \sum_{j=k+1}^{N} \frac{u_j^T b^{\text{exact}}}{\sigma_j} v_j + \sum_{j=k+1}^{N} \frac{u_j^T b^{\text{noise}}}{\sigma_j} v_j.$$

Idea: Omit the noise dominated part. Define

$$x^{\mathrm{TSVD}(k)} \equiv \sum_{j=1}^{k} \frac{u_j^T b}{\sigma_j} v_j = \sum_{j=1}^{N} \phi_j \frac{u_j^T b}{\sigma_j} v_j,$$

where

$$\phi_j = \begin{cases} 1 & \text{for} \quad j \leq k \\ 0 & \text{for} \quad j > k \end{cases}.$$

A part of noise is still in the solution $\sum_{j=1}^{k} \frac{u_j^T b^{\mathrm{noise}}}{\sigma_j} v_j$,

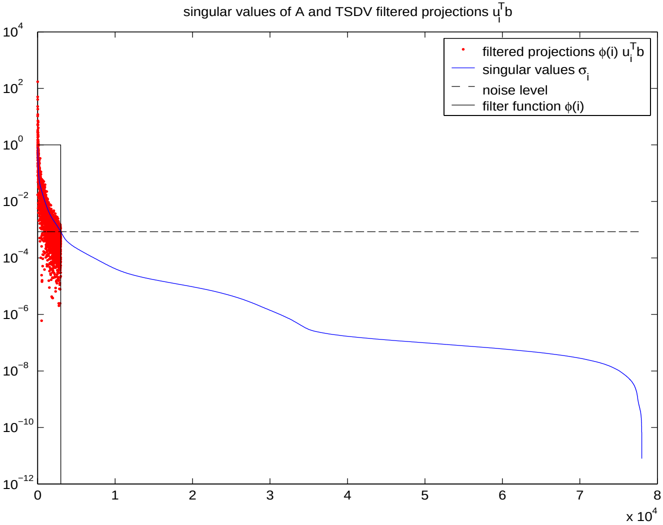
a part of useful information is lost $\sum_{j=k+1}^{N} \frac{u_j^T b^{\mathrm{exact}}}{\sigma_j} v_j$.

If k is too small $x^{\mathrm{TSVD}(k)}$ is **overregularized** (too smooth),
if k is too large $x^{\mathrm{TSVD}(k)}$ is **underregularized** (noisy).

5. Basic regularization techniques

Truncated SVD

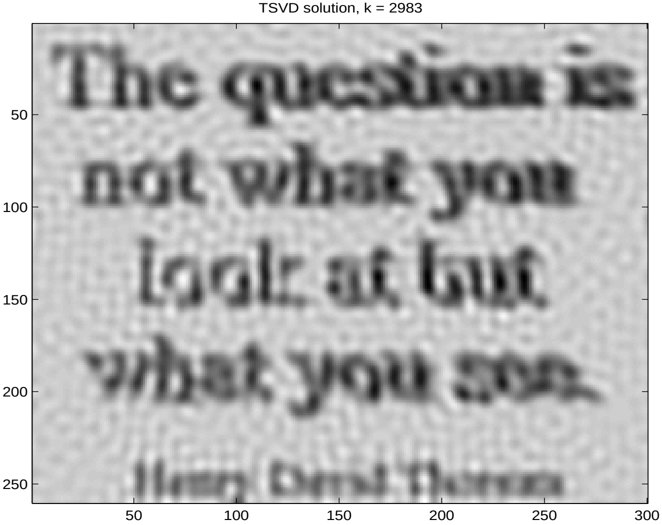
The TSVD filter function, $k = 2\,983$:



5. Basic regularization techniques

Truncated SVD

The TSVD solution, $k = 2\,983$:



5. Basic regularization techniques

Truncated SVD

Advantages:

- Simple idea, simple implementation, simple analysis,

$$A \quad \text{is replaced by} \quad U\Phi^{\dagger}\Sigma V^T, \quad \Phi = \mathrm{diag}(I_k, 0_{N-k}),$$

  i.e. the rank-k approximation of A.

Disadvantages:

- We have to compute the SVD of A (or the first k singular triplets).
- Choice of the **regularization parameter** k is usualy based on a knowledge of the norm of b^{noise} which is

  either revealed from the SVD analysis,

  or given explictly as an additional information.
- The noise dominated part still contains some information useful for reconstruction which is lost (step filter function).

5. Basic regularization techniques

Selective SVD

Similar approach to TSVD is the **selective SVD (SSVD)**.
Consider $\|b^{\text{noise}}\|$ is known. Then

$$\|b^{\text{noise}}\| = \left(\sum_{j=1}^{N} (u_j^T b^{\text{noise}})^2\right)^{1/2} \equiv \Delta^{\text{noise}}, \quad |u_j^T b^{\text{noise}}| \approx \varepsilon \equiv \frac{\Delta^{\text{noise}}}{N^{1/2}},$$

because u_j represent frequencies and b^{noise} represents white noise.

We define

$$x^{\text{SSVD}(\varepsilon)} \equiv \sum_{|u_j^T b| > \varepsilon} \frac{u_j^T b}{\sigma_j} v_j = \sum_{j=1}^{N} \phi_j \frac{u_j^T b}{\sigma_j} v_j,$$

where

$$\phi_j = \begin{cases} 1 & \text{for} \quad |u_j^T b| > \varepsilon \\ 0 & \text{for} \quad |u_j^T b| \le \varepsilon \end{cases}.$$

5. Basic regularization techniques

Tikhonov approach

Classical **Tikhonov approach** is based on penalizing the norm of the solution

$$x^{\mathrm{Tikhonov}(\lambda)} \equiv \arg\min_{x} \{\|b - Ax\|^2 + \lambda^2 \|Lx\|^2\},$$

where

- $\|b - Ax\|$ represents the residual norm,
- $\|Lx\|$ represents $(L^T L)$–(semi)norm of the solution,

  often $L = I_N$ (we restrict to this case),

  or it is a discretized 1st or 2nd order derivative operator,
- λ is the (positive) penalty parameter; clearly

$$\lim_{\lambda \longrightarrow 0} x^{\mathrm{Tikhonov}(\lambda)} = x^{\mathrm{naive}}.$$

The Tikhonov minimization problem can be rewritten as

$$x^{\mathrm{Tikhonov}(\lambda)} = \arg\min_x \{\|b - Ax\|^2 + \lambda^2 \|Lx\|^2\}$$
$$= \arg\min_x \left\{ \left\| \begin{bmatrix} b \\ 0 \end{bmatrix} - \begin{bmatrix} A \\ -\lambda L \end{bmatrix} x \right\|^2 \right\},$$

i.e. to get the Tikhonov solution we solve a **least squares (LS) problem**

$$\begin{bmatrix} A \\ -\lambda L \end{bmatrix} x = \begin{bmatrix} b \\ 0 \end{bmatrix}.$$

In particular, we do not have to compute the SVD of A.

5. Basic regularization techniques

Tikhonov approach

A solution of the Tikhonov LS problem

$$\begin{bmatrix} A \\ -\lambda L \end{bmatrix} x = \begin{bmatrix} b \\ 0 \end{bmatrix}$$

can be analyzed through the system of normal equations

$$\begin{bmatrix} A \\ -\lambda L \end{bmatrix}^T \begin{bmatrix} A \\ -\lambda L \end{bmatrix} x = \begin{bmatrix} A \\ -\lambda L \end{bmatrix}^T \begin{bmatrix} b \\ 0 \end{bmatrix},$$

$$(A^T A + \lambda^2 L^T L)x = A^T b.$$

With the SVD of A, $A = U\Sigma V^T$, and $L = I_N = VV^T$ we get

$$(\Sigma^2 + \lambda^2 I_N)y = \Sigma U^T b,$$

where $y = V^T x$ and $x = Vy$.

Thus

$$x^{\mathrm{Tikhonov}(\lambda)} = V(\Sigma^2 + \lambda^2 I_N)^{-1} \Sigma U^T b,$$

which gives

$$\begin{aligned} x^{\mathrm{Tikhonov}(\lambda)} &= \sum_{j=1}^{N} \frac{\sigma_j}{\sigma_j^2 + \lambda^2} (u_j^T b) v_j \\ &= \sum_{j=1}^{N} \frac{\sigma_j^2}{\sigma_j^2 + \lambda^2} \frac{u_j^T b}{\sigma_j} v_j = \sum_{j=1}^{N} \phi_j \frac{u_j^T b}{\sigma_j} v_j, \end{aligned}$$

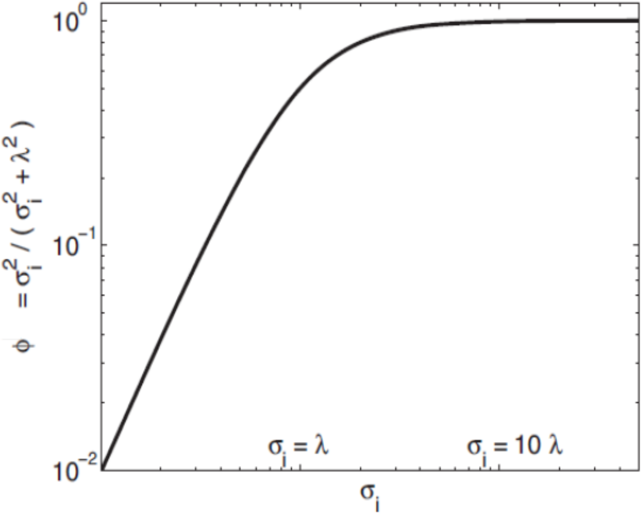
where

$$\phi_j = \frac{\sigma_j^2}{\sigma_j^2 + \lambda^2} \approx \begin{cases} 1 & \text{for} \quad \sigma_j \gg \lambda \\ \sigma_j^2/\lambda^2 & \text{for} \quad \sigma_j \ll \lambda \end{cases}, \qquad 0 < \phi_j < 1.$$

5. Basic regularization techniques

Tikhonov approach

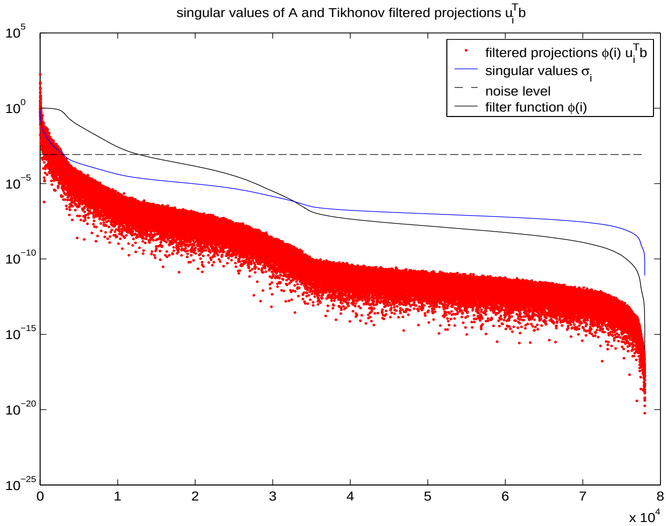
The behavior of the Tikhonov filter function:



5. Basic regularization techniques

Tikhonov approach

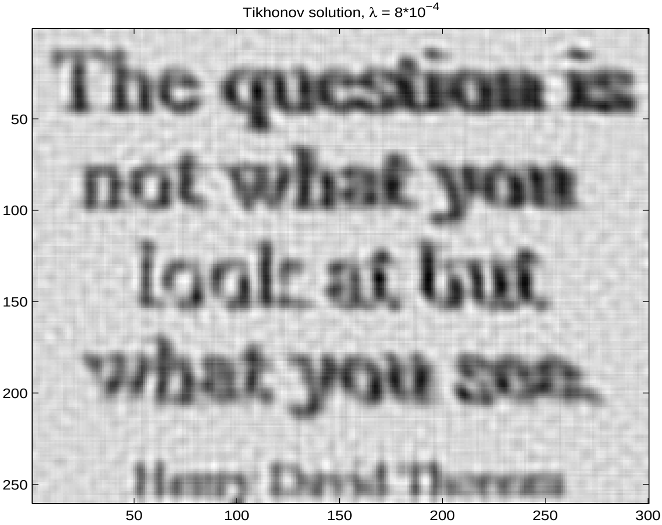
The Tikhonov filter function, $\lambda = 8 \times 10^{-4}$:



5. Basic regularization techniques

Tikhonov approach

The Tikhonov solution, $\lambda = 8 \times 10^{-4}$:



Advantages:

- Simple idea, with $L = I_N$ simple analysis,

$$A \quad \text{is replaced by} \quad U\Phi^{-1}\Sigma V^T, \quad \Phi = (\Sigma^2 + \lambda^2 I_N)^{-1}\Sigma^2.$$

- We do not have to compute SVD of A (compare with TSVD).
- The solution is given by some LS problem.
- The filter function is smooth (compare with TSVD).

Disadvantages:

- With $L \neq I_N$ the analysis is more complicated.
- We have to chose the **penalty parameter** λ (at this moment it is not clear how to do it).

5. Basic regularization techniques

Summary

We have two basic approaches:

- **Truncated SVD** (requires a part of the SVD of A)

$$x^{\mathrm{TSVD}(k)} = V\Phi\Sigma^{-1}U^T b, \quad \Phi = \mathrm{diag}(I_k, 0_{N-k}),$$

  where k is a truncation (regularization) parameter.

- **Tikhonov regularization** (leads to a LS problem)

$$x^{\mathrm{Tikhonov}(\lambda)} = V\Phi\Sigma^{-1}U^T b, \quad \Phi = (\Sigma^2 + \lambda^2 I_n)^{-1}\Sigma^2,$$

  where λ is a penalty (regularization) parameter.

The question is:

How to choose the regularization parameters?

5. Basic regularization techniques

Note on monotonicity (TSVD)

The **norms of the TSVD solution and the residual**

$$\|x^{\mathrm{TSVD}(k)}\|, \quad \|b - Ax^{\mathrm{TSVD}(k)}\|$$

are **nondecreasing** and **nonincreasing**, respectively, with k.

Simply, using SVD,

$$\|x^{\mathrm{TSVD}(k)}\|^2 = \sum_{j=1}^{k} \frac{(u_j^T b)^2}{\sigma_j^2}$$

is nondecreasing with k;

$$\|b - Ax^{\mathrm{TSVD}(k)}\|^2 = \|(I - \Phi)U^T b\|^2 = \sum_{j=k+1}^{N} \frac{(u_j^T b)^2}{\sigma_j^2}$$

is nonincreasing with k (here $\Phi = \mathrm{diag}(I_k, 0_{N-k})$).

5. Basic regularization techniques

Note on monotonicity (Tikhonov)

Similarly the **norms of the Tikhonov solution and the residual**

$$\xi(\lambda) \equiv \|x^{\mathrm{Tikhonov}(\lambda)}\|^2 = \sum_{j=1}^{N} \phi_j^2 \frac{(u_j^T b)^2}{\sigma_j^2},$$

$$\rho(\lambda) \equiv \|b - Ax^{\mathrm{Tikhonov}(\lambda)}\|^2 = \sum_{j=1}^{N} (1-\phi_j)^2 (u_j^T b)^2$$

are **increasing** and **decreasing**, respectively, with λ.

Recall that $0 < \phi_j < 1$,

$$\phi_j = \frac{\sigma_j^2}{\sigma_j^2 + \lambda^2}, \qquad \text{thus} \qquad (1-\phi_j) = \frac{\lambda^2}{\sigma_j^2 + \lambda^2}.$$

Look at

$$\xi'(\lambda) = \frac{d\xi(\lambda)}{d\lambda}, \qquad \rho'(\lambda) = \frac{d\rho(\lambda)}{d\lambda}.$$

5. Basic regularization techniques

Note on monotonicity (Tikhonov)

First

$$\frac{d}{d\lambda}\phi_j^2 = -\frac{4}{\lambda}(1-\phi_j)\phi_j^2, \quad \frac{d}{d\lambda}(1-\phi_j)^2 = \frac{4}{\lambda}(1-\phi_j)^2\phi_j.$$

Then

$$\xi'(\lambda) = -\frac{4}{\lambda}\sum_{j=1}^{N}(1-\phi_j)\phi_j^2\,\frac{(u_j^T b)^2}{\sigma_j^2},$$

$\xi'(\lambda) < 0$ for $\lambda > 0$, i.e. $\xi(\lambda)$ is decreasing with λ.

Analogously

$$\rho'(\lambda) = \frac{4}{\lambda}\sum_{j=1}^{N}(1-\phi_j)^2\phi_j(u_j^T b)^2,$$

$\rho'(\lambda) > 0$ for $\lambda > 0$, i.e. $\rho(\lambda)$ is increasing with λ.

6. Choosing regularization parameters

6. Choosing regularization parameters

Spectral filtering, Error analysis

In general

$$\begin{aligned} x^{\text{filtered}} &= V\Phi\Sigma^{-1}U^T b \\ &= V\Phi\Sigma^{-1}U^T b^{\text{exact}} + V\Phi\Sigma^{-1}U^T b^{\text{noise}} \\ &= V\Phi\Sigma^{-1}U^T A x^{\text{exact}} + V\Phi\Sigma^{-1}U^T b^{\text{noise}} \\ &= (V\Phi V^T) x^{\text{exact}} + V\Phi\Sigma^{-1}U^T b^{\text{noise}}, \end{aligned}$$

where $V\Phi V^T$ is called the **resolution matrix**.

The **absolute error** is

$$x^{\text{exact}} - x^{\text{filtered}} = \underbrace{(I_N - V\Phi V^T)x^{\text{exact}}}_{\text{regularization error}} - \underbrace{V\Phi\Sigma^{-1}U^T b^{\text{noise}}}_{\text{perturbation error}},$$

regularization error is caused by using filtered inverse,
perturbation error consists of the inverted and filtered noise.

There is **no universal approach** for chosing the regularization parameter (k or λ), the choice is always **problem dependent!**
In general:

- If $\Phi \approx I_N$ ($V \Phi V^T \approx I_N$), the regularization error is small, but the perturbation error (caused by noise) is large.

  The solution is **undersmoothed**.

- If $\Phi \approx 0_N$ ($V \Phi V^T$ is far from the identity), inverted noise is heavily damped, but we lose a lot of original data.

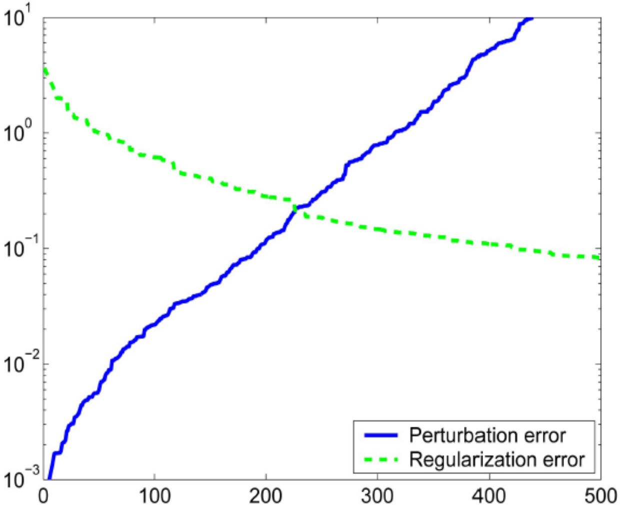
  The solution is **oversmoothed**.

A proper choice of the regularization parameter balances these two types of errors.

6. Choosing regularization parameters

Spectral filtering, A proper choice of the parameter

Regularization and perturbation error for TSVD method:



6. Choosing regularization parameters

Discrepancy principle

The **discrepancy principle**: Let

$$\|b^{\text{noise}}\| = \Delta^{\text{noise}}$$

be known either from the nature of the problem, or we have some **approximation** of it, see ⟨Lecture III⟩.

We look for a regularization parameter such that

$$\|b - Ax^{\text{filtered}}\| = \tau \Delta^{\text{noise}},$$

for some fixed τ.

Recall that for TSVD and Tikhonov regularization the norms of the residuals are **monotonic** in k and λ, respectively.

[Morozov: '66], [Morozov: '84].

Using $x^{\text{filtered}} = V\Phi\Sigma^{-1}U^T b$ the residual satisfies

$$b - Ax^{\text{filtered}} = \left(I_N - AV\Phi\Sigma^{-1}U^T\right) b = \left(I_N - U\Phi U^T\right) b.$$

Defining the **generalized cross validation (GCV)** functional

$$G^{\text{filtered}}(\cdot) \equiv \frac{\|b - Ax^{\text{filtered}}\|^2}{\text{trace}(I_N - AV\Phi\Sigma^{-1}U^T)^2} = \frac{\|(I_N - \Phi)U^T b\|^2}{(N - \sum_{j=1}^{N} \phi_j)^2},$$

we look for its minimum.

(Note: The GCV functional depends on ordering of equations.)

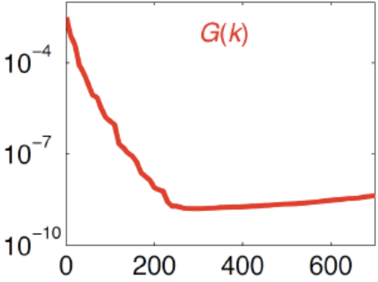
[Chung, Nagy, O'Leary: '04], [Golub, Von Matt: '97], [Nguyen, Milanfar, Golub: '01].

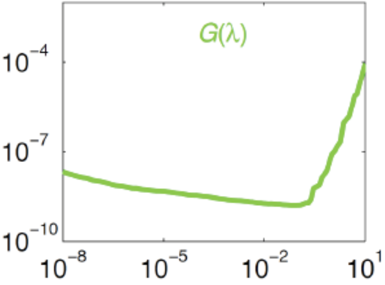
In particular for the TSVD and Tikhonov method we have

$$G^{\mathrm{TSVD}}(k) = \frac{\sum_{j=k+1}^{N} (u_j^T b)^2}{(N-k)^2},$$

$$G^{\mathrm{Tikhonov}}(\lambda) = \frac{\sum_{j=1}^{N} \left(\frac{u_j^T b}{\sigma_j^2 + \lambda^2} \right)^2}{\left(\sum_{j=1}^{N} \frac{1}{\sigma_j^2 + \lambda^2} \right)^2}.$$

The GCV functional for TSVD (left) and Tikhonov (right) methods:





Note: The GCV functional is often flat close to the minimum.

6. Choosing regularization parameters

L-curve

Both norms

$$\|x^{\text{filtered}}\|, \qquad \|b - Ax^{\text{filtered}}\|$$

are monotonic with respect to the regularization parameter k, λ in TSVD and Tikhonov regularization, respectively.

We plot the norm of the regularized solution agains the norm of the residual. For emphasizing the point where both norms are **balanced**, we use the **log-log** scale.

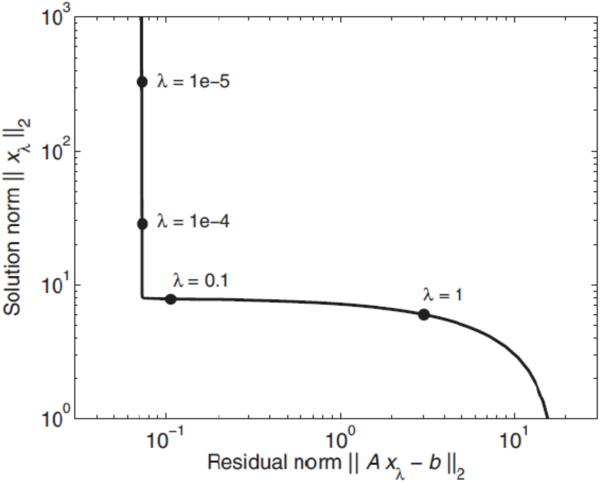
Criterion based on this approach is called the **L-curve**. The L-curve-optimal parameter then corresponds to the point with maximal curvature.

Note that for TSDV we have only discrete set of points (parameter k is discrete). The curvature is defined using an interpolation.

[Calvetti, Golub, Reichel: '99], [Calvetti, Morigi, Reichel, Sgallari: '00], [Calvetti, Reichel: '04].

6. Choosing regularization parameters

Ideal L-curve for Tikhonov method (often the corner is not sharp). Here λ grows from the upper left to the bottom right corner along the curve:



6. Choosing regularization parameters

Normalized cumulative periodogram (NCP)

The last criterion is based on the assumption that the residual corresponding to the true solution

$$b^{\mathrm{noise}} = b - Ax^{\mathrm{exact}}$$

represents white noise. We try to choose a regularization parameter such that the residual

$$r^{\mathrm{filtered}} = b - Ax^{\mathrm{filtered}}$$

resembles white noise. See also ⟨Lecture III⟩.

The **normalized cumulative periodogram (NCP)** uses the statistical properties of Fourier spectrum of white noise.

[Rust: '98], [Rust: '00], [Rust, O'Leary: '08] (FFT-based), [Hansen, Kilmer, Kjeldsen: '06] (DCT-based).

6. Choosing regularization parameters

Normalized cumulative periodogram (NCP)

The NCP transforms the residual $r^{\text{filtered}} \in \mathbb{R}^N$ using the discrete Fourier transform (DFT/FFT algorithm) to get its spectrum

$$p^{\text{filtered}} = \mathcal{F}(r^{\text{filtered}}) = (p_1, p_2, \ldots, p_{\nu+1})^T, \qquad \nu = \lfloor N/2 \rfloor.$$

The periodogram is a vector C^{filtered} with coefficients

$$c_j = \frac{|p_2| + \ldots |p_{j+1}|}{|p_2| + \ldots |p_{\nu+1}|}, \qquad j = 1, \ldots, \nu.$$

If the residual consists only of white noise, then by the definiton of white noise the mean values satisfy

$$E[|p_2|] = E[|p_3|] = \ldots = E[|p_\nu|],$$

and by linearity of $E[\,\cdot\,]$, points $(j, E[c_j])$ would be on a straight line from $(0, 0)$ to $(\nu, 1)$.

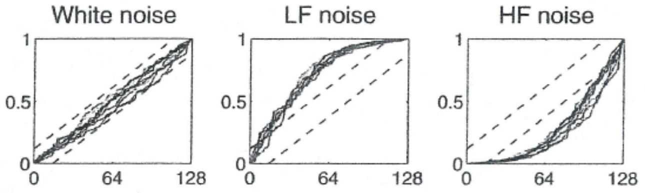
6. Choosing regularization parameters

Normalized cumulative periodogram (NCP)

Thus we look for the regularization parameter (k or λ) such that the coefficients of the periodogram c^{filtered} lie (moreorless) on a straight line,

$$\min_{k \text{ or } \lambda} \|C^{\text{filtered}} - C^{\text{white noise}}\|_2, \quad C^{\text{white noise}} = \frac{1}{\nu}(1, 2, \ldots, \nu)^T.$$

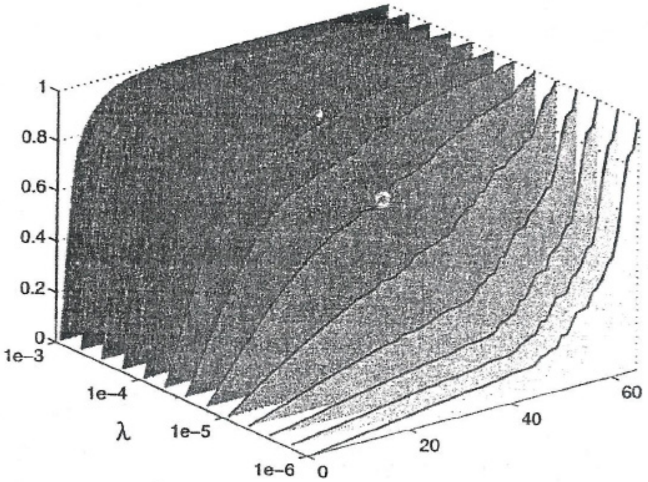
Note that the periodogram is normalized, i.e. $c_\nu = 1$.



6. Choosing regularization parameters

Normalized cumulative periodogram (NCP)

NCP for Tikhonov regularization:



[Hansen: SIAM, FA07, 2010].

Discrepancy principle: Converges as noise tends to zero, requires an explicite information about the norm of noise component of b, the solution tends to be oversmooth.

Generalized cross validation (GCV): No convergence when noise tends to zero, functional is flat close to the minimum, various adaptations for structured matrices (BCCB, etc.).

L-curve: No convergence when noise tends to zero, various adaptations (L-ribbon, etc.), well numerically tracktable (it is sufficient to compute only a few points of the L-curve), troubles when using with TSVD because k is a discrete parameter.

Usually we need to solve one system with several values of the regularization parameter to choose the optimal one.

See also [Björk: '88], [Björk, Grimme, Van Dooren: '94].
For comparison see [Hansen: 98], [Kilmer, O'Leary: '01].

7. Iterative regularization

7. Iterative regularization

Introduction

Up to now we have considered **direct** regularization methods suitable for small problems (SVD-based methods, Tikhonov regularization leading to a LS problem which can be solved directly only in small dimensions).

For solving **large ill-posed problems**, it is advatagous to use **iterative regularization methods**. We briefly introduce several of them:

- stationary iterative methods (Landweber iteration, Cimmino iteration, Kaczmarz's method (ART)),
- projection methods (regularizing Krylov subspace iterations).

In all iterative methods the **number of iterations** plays the role of the **regularization parameter**.

Simultaneous iterative reconstruction techniques (SIRT) is a class of stationary iterative methods with a general scheme

$$x^{[\ell]} := x^{[\ell-1]} + \omega A^T M(b - Ax^{[\ell-1]}), \qquad \ell = 1, 2, \dots, k,$$

where M is a symmetric positive definite (SPD) matrix and ω is a relaxation parameter.

For example often used methods are:

- the **Landweber iteration** with $M = I_N$, and
- the **Cimminio iteration** with $M = D = \mathrm{diag}(d_1, \dots, d_N)$,

$$d_j = \frac{1}{N} \frac{1}{\|\underline{a}_j\|^2},$$

where $\underline{a}_j$ is the (transposed) jth row of A (column vector).

7. Iterative regularization

Stationary iterative methods, Landweber iteration

The Landweber method

$$x^{[\ell]} := x^{[\ell-1]} + \omega A^T(b - Ax^{[\ell-1]}), \qquad \ell = 1, 2, \ldots, k,$$

with $0 < \omega < 2\sigma_1^{-2}(A) = 2\|A^T A\|^{-1}$ gives the approximation

$$x^{[k]} = V\Phi^{[k]}\Sigma^{-1}U^T b, \qquad \Phi^{[k]} = I_N - (I_N - \omega\Sigma^2)^k,$$

i.e. $\phi_j^{[k]} = 1 - (1 - \omega\sigma_j^2)^k$.

Using the Taylor expansion for small σ_j's we get $\phi_j^{[k]} \approx k\omega\sigma_j^2$.

Thus the Landweber filters decay with the same rate as the Tikhonov filters ($\phi_j \approx \sigma_j^2\lambda^{-2}$).

7. Iterative regularization

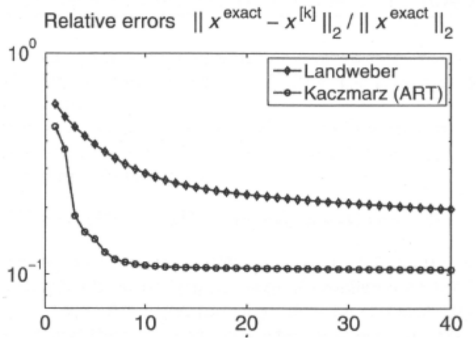
Stationary iterative methods, Kaczmarz's method (ART)

Kaczmarz's method or **algebraic reconstruction technique (ART)** is an iterative algorithm given by the following scheme

$$
\begin{aligned}
& x^{[\ell-1,0]} := x^{[\ell-1]}, \\
& \textbf{for} \quad j = 1, \dots, N \\
& \qquad x^{[\ell-1,j]} := x^{[\ell-1,j-1]} + \omega\, \underline{a}_j \frac{1}{\|\underline{a}_j\|^2} \left(b_j - \underline{a}_j^T x^{[\ell-1,j-1]}\right), \\
& \textbf{end} \\
& x^{[\ell]} := x^{[\ell-1,N]}, \qquad \ell = 1, 2, \dots, k.
\end{aligned}
$$

The ART method converges quite quickly in the first few iterations, after this the convergence may become very slow.

Comparison of relative error decay for Landweber and Kaczmarz's (ART) method:



[Hansen: SIAM, FA07, 2010].

In direct techniques we have looked for an approximation of x^{exact} which lies in a low dimensional subspace of $\mathbb{R}^N$ spanned by the first k right singular vectors of A (TSVD); or which is dominated by several first right singular vectors of A (Tikhonov).

Thus the approximation is always dominated by the low frequencies and the high frequecies are dumped.

We try to look for an approximation in an a-priori given low dimensional subspace $\mathcal{W}_k$ dominated by low frequencies.

7. Iterative regularization

Projection methods

Consider a subspace

$$\mathcal{W}_k = \mathrm{span}(w_1, \dots, w_k) \subset \mathbb{R}^N, \qquad W_k = [w_1, \dots, w_k] \in \mathbb{R}^{N\times k},$$

such that $W_k^T W_k = I_k$ and w_j are dominated by low frequecies.
Then we solve

$$\min_{x\in\mathcal{W}_k} \|b - Ax\|.$$

This can be reformulated as a **projected problem**

$$\min_{y\in\mathbb{R}^k} \|b - (AW_k)y\|,$$

or, equivalently,

$$W_k^T(A^T A)W_k y = W_k^T A^T b.$$

The question is, how to choose the basis w_j?

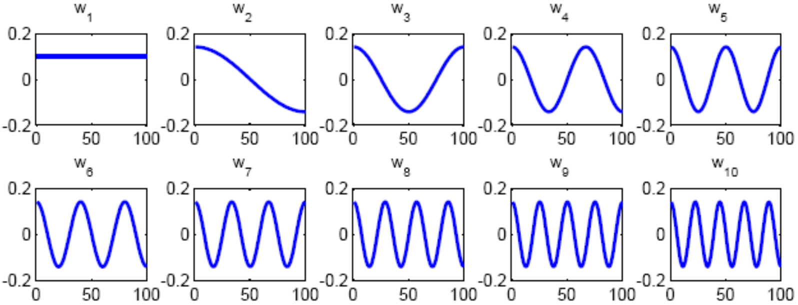
7. Iterative regularization

Projection methods, DCT basis

An example of a suitable basis is the DCT basis

$$w_1 = \sqrt{\frac{1}{N}}\,(1,1,\ldots,1)^T,$$
$$w_j = \sqrt{\frac{2}{N}}\left(\cos\left(\frac{(j-1)\pi}{2N}\right), \cos\left(\frac{3(j-1)\pi}{2N}\right), \ldots \cos\left(\frac{(2N-1)(j-1)\pi}{2N}\right)\right)^T,$$

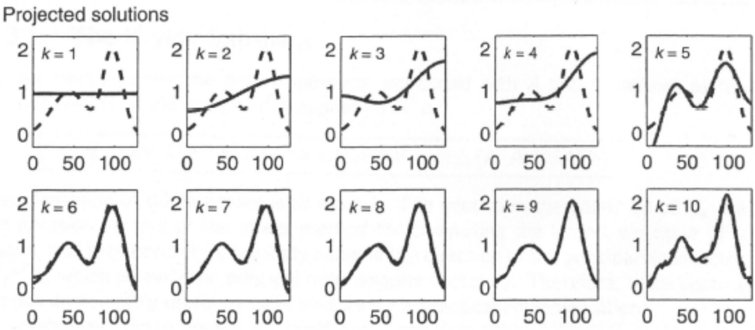
for $j > 1$.



7. Iterative regularization

Projection methods, DCT basis

Solutions computed using the DCT basis $w_1, \ldots, w_k$, $k = 1, \ldots, 10$ ($k = 9$ seems to be the optimal one):



Note: If there are a-priori known certain properties of the true solution (symmetry, periodicity, etc.), use this knowledge to choose basis vectors satisfying these properties.

Note that choosing $w_j = v_j$ (the right singular vectors of A), we get exactly the TSVD mehtod. Thus TSVD is an projection method.

Advantage: With fixed set of basis vectors, computations can be performed quickly. Using e.g. DCT basis we do not have to compute and store the basis vectors (we compute only the DCT and the inverse DCT (IDCT) of a vector).

Disadvantage: The basis vectors are not adapted to the particular problem.

To avoid this disadvatage we introduce the regularizing Krylov subspace iteration.

Specific projection methods are the **Krylov subspace methods**. Here the orthonormal basis of a Krylov subspace

$$\mathcal{K}_k(v, M) = \mathrm{span}(v, Mv, \ldots, M^{k-1}v),$$

is used for w_j, $j = 1, \ldots, k$, vectors. For example the choice

$$v = A^T b, \quad M = A^T A,$$

leads to very popular iterative (regularization) methods **CGLS**, **LSQR** or **CGNE**, which are mathematically equivalent to applying CG method on the normal equations $A^T Ax = A^T b$.

The regularizing properties of the Krylov subspace methods will be dicussed in ⟨Lecture III⟩ in more details, in particular in the context of hybrid methods.

7. Iterative regularization

Further remarks

In the iterative regularization (using stationary or projection methods), the number of computed iterations k plays the role of the regularization parameter.

As a stopping criterion for the iterative process any of the previously mentioned approaches can be used, e.g.:

- the discrepancy principle,
- the generalized cross validation (GCV),
- the L-curve criterion,
- the normalized cumulative periodograms (NCP).

8. Hybrid methods

The best of both worlds

8. Hybrid methods

Introduction

Hybrid methods combine both previous approaches. Here the regularization is realized in two steps.

First, the original problem is projeted onto a lower dimensional subspace using an iterative (projection) method, which by itself represents a form of regularization by projection, i.e. **outer regularization**.

The small projected problem, however, may inherit a part of the ill-posedness of the original problem and therefore some form of **inner regularization** is applied.

Stopping criteria for the whole process are then based on the regularization of the projected (small) problems.

[O'Leary, Simmons: '81], [Hansen: '98] or [Fiero, Golub, Hansen, O'Leary: '97], [Kilmer, O'Leary: '01], [Kilmer, Español: '06], [O'Leary, Simmnos: '81].

8. Hybrid methods

Projection methods with inner Tikhonov regularization

As an example we introduce the **Projection method with inner Tikhonov regularization**. Consider the ill-posed problem $Ax = b$ and a subspace $\mathcal{W}_k = \mathrm{span}(w_1, \ldots, w_k)$. Denote

$$M_k = W_k^T (A^T A) W_k \in \mathbb{R}^{k \times k}, \qquad \text{where} \quad W_k = [w_1, \ldots, w_k].$$

The system of normal equations $A^T A x = A^T b$ is projected on $\mathcal{W}_k$,

$$M_k y = W_k^T b, \qquad x = W_k y.$$

The inner Tikhonov regularization can be applied on this small problem

$$y^{\mathrm{Tikhonov}(\lambda)} \;=\; \arg\min_y \{ \| W_k^T b - M_k y \|^2 + \lambda^2 \| y \|^2 \}.$$

This leads to a small LS problem that can be easily solved directly for **many** values of λ.

Summary

We have described the following regularization methods:

- the **direct regularization** techniques (TSVD, Tikhonov regularization) suitable for solving **small** ill-posed problems;
- **stationary** regularization methods (Landweber and Cimmino iterations, Kaczmarz's (ART) method);
- **projection** regularization methods including **regularizing Krylov subspace iterations**;
- **hybrid** methods combining the previous techniques.

All regularization techniques require to **choose a good regularization parameter**, that can be find using, e.g., the discrepancy principle, the generalized cross validation, the L-curve criterion, or the normalized cumulative periodograms.