ECE 580 Lecture

Regularization by Linear Filtering

October 10, 2002

Goal of Lecture

- Review the SVD and pseudo-inverse, and define ill-conditioning.
- Introduce regularization by linear filtering
- Examples
  - Tikhonov Regularization
  - Truncated Singular Value Decomposition (TSVD)
  - Landweber Iteration
- Illustrate the importance of correctly choosing the regularization (filtering) parameter.

Review of the SVD

Linear operator $A : \mathbb{R}^n \to \mathbb{R}^m$ with $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y} = \sum_i x_i y_i$. The singular value decomposition (SVD) of A is

$$A = \underbrace{[\mathbf{u}_1, \ldots, \mathbf{u}_m]}_{m \times m \;\; U} \; \underbrace{\text{diag}(s_1, \ldots, s_n)}_{m \times n \;\; S} \; \underbrace{[\mathbf{v}_1, \ldots, \mathbf{v}_n]^T}_{n \times n \;\; V^T}.$$

with singular values s_i and singular vectors $\mathbf{u}_i, \mathbf{v}_i$ satisfying

$$s_1 \geq s_2 \geq \ldots \geq s_n \geq 0, \;\; \langle \mathbf{u}_i, \mathbf{u}_j \rangle = \delta_{ij}, \;\; \langle \mathbf{v}_i, \mathbf{v}_j \rangle = \delta_{ij}$$

The action of A on a vector $\mathbf{x} \in \mathbb{R}^n$ is given by

$$A\mathbf{x} = U[s_1 \, \mathbf{v}_1^T \mathbf{x}, \ldots, s_n \, \mathbf{v}_n^T \mathbf{x}, 0, \ldots, 0]^T = \sum_{s_i > 0} s_i \, \langle \mathbf{v}_i, \mathbf{x} \rangle \, \mathbf{u}_i$$

The summation representation extends to linear operators on infinite dimensional Hilbert spaces.

The Pseudo-Inverse

Let $A : \mathcal{F} \to \mathcal{G}$ be a linear operator on (possibly infinite dimensional) Hilbert spaces. If $g \in$ Range(A), then

$$g = Af = \sum_{s_i > 0} s_i \langle f, v_i \rangle u_i, \qquad f \in \mathcal{F}.$$

Take inner product with u_j and use orthonormality to get $\langle g, u_j \rangle = s_i \langle f, v_i \rangle \, \delta_{ij}$. Then

$$f = \sum_i \langle f, v_i \rangle v_i = \sum_{s_i = 0} \langle f, v_i \rangle v_i \; + \sum_{s_i > 0} \frac{\langle g, u_i \rangle}{s_i} v_i.$$

The first sum on the r.h.s. gives the projection of f onto Null(A); the second gives $A^\dagger g$.

Pseudo-Inverse, Continued

Formally define $A^\dagger : \mathcal{G} \to \mathcal{F}$ by

$$A^\dagger g = \sum_{s_i > 0} \frac{\langle g, u_i \rangle}{s_i} v_i .$$

This is well-defined iff the Picard condition holds,

$$||A^\dagger g||^2 = \sum_{s_i > 0} \frac{|\langle g, u_i \rangle|^2}{s_i^2} < \infty .$$

This condition will fail for some $g \in \mathcal{G}$ if $s_i \to 0$. This condition on singular values characterizes compact operators A on infinite dimensional separable Hilbert spaces, and it means that existence and continuous dependence fail to hold for the operator equation $Af = g$.

Ill-Conditioning

In finite dimensions, $s_i \to 0$ makes no sense, and the pseudo-inverse operator $A^\dagger$ is continuous, or bounded. (we can show $||A^\dagger|| = 1/\min\{s_i | s_i > 0\} \ < \infty$).
But this isn't the whole story.

The (spectral, or ℓ^2) condition number of a finite dimensional linear operator (matrix) A is given by

$$\text{cond}(A) = \frac{\max\{s_i | s_i > 0\}}{\min\{s_i | s_i > 0\}}$$

A is ill-conditioned iff cond(A) is large iff A has a relatively broad range of singular values. Then the pseudo-inverse solution to $Af = g$ can be unstable with respect to perturbations in g. In particular, error components corresponding to small singular values are greatly amplified.

Tikhonov Regularization

Tikhonov regularized solution to $Af = g$ is

$$\begin{aligned} f_\alpha &= \arg\min_{f \in \mathcal{F}} ||Af - g||^2 + \alpha ||f||^2 \\ &= (A^*A + \alpha I)^{-1} A^* g \\ &= \sum_i \frac{1}{s_i^2 + \alpha}\, s_i \,\langle u_i, g\rangle\, v_i \\ &= \sum_{s_i > 0} \underbrace{\frac{s_i^2}{s_i^2 + \alpha}}_{w_\alpha(s_i^2)} \frac{\langle u_i, g\rangle}{s_i}\, v_i . \end{aligned}$$

Tikhonov filter function $w_\alpha(s^2) = s^2/(s^2 + \alpha)$.

Truncated SVD Regularization

Take

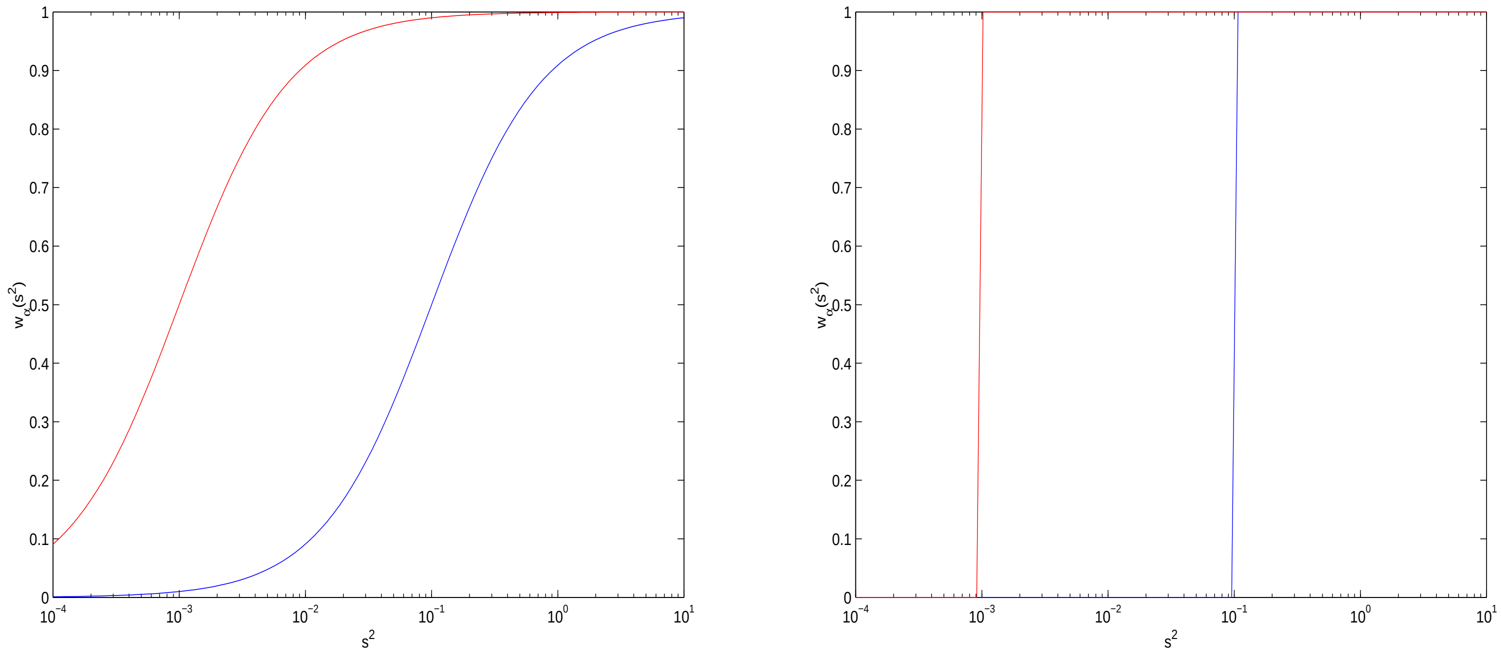
$$f_\alpha = \sum_{s_i>0} w_\alpha(s_i^2)\frac{\langle u_i, g\rangle}{s_i}\, v_i,$$

with truncated singular value decomposition (TSVD) filter

$$w_\alpha(s^2) = \begin{cases} 0, & s^2 < \alpha, \\ 1, & s^2 \geq \alpha. \end{cases}$$

The regularization parameter is the cut-off level α.

Comparison of Tikhonov, TSVD Filters



On the left are Tikhonov filter functions; on the right are TSVD filter functions. Blue corresponds to regularization parameter $\alpha = 10^{-1}$; red corresponds to $\alpha = 10^{-3}$.

Landweber Iteration

Minimize the least squares fit-to-data functional

$$J(f) = \frac{1}{2}||Af - g||^2$$

using gradient descent iteration, initial guess $f^0 = 0$, and fixed step length parameter $0 < \tau < 1/||A||^2$.

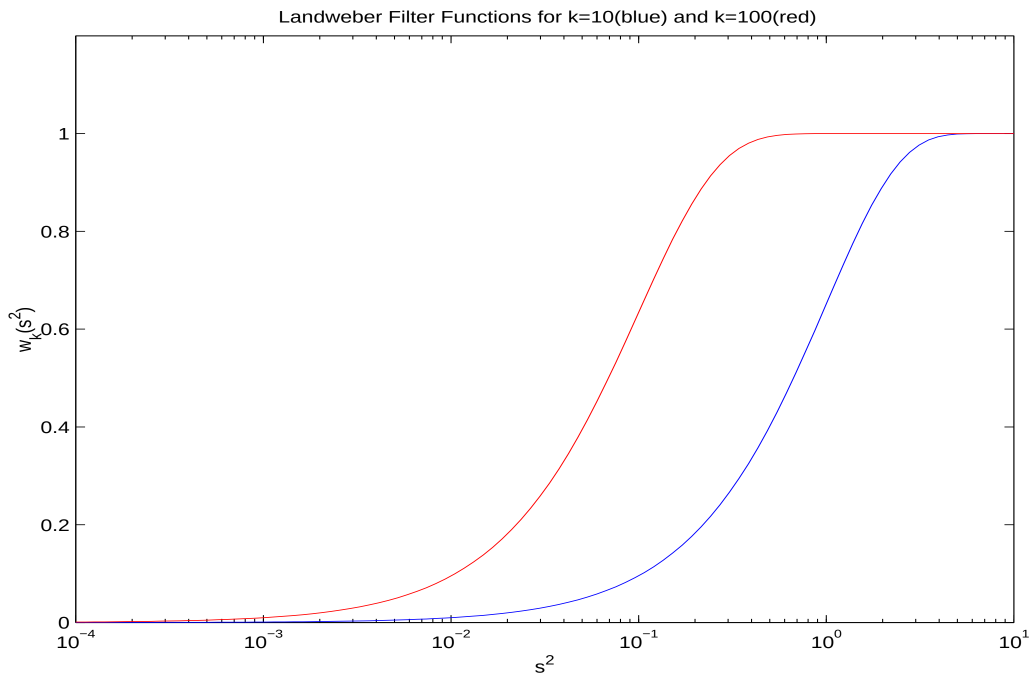
$$\begin{aligned} f^{k+1} &= f^k - \tau \,\text{grad}\, J(f^k), \quad k = 0, 1, 2, \ldots \\ &= f^k - \tau A^*(Af^k - g) \\ &= (I - \tau A^* A) f^k + \tau A^* g. \end{aligned}$$

Landweber Iteration, Continued

Using the SVD, one can derive a filter representation

$$f^{k+1} = \sum_{s_i > 0} \underbrace{[1 - (1 - \tau s_i^2)^k]}_{w_k(s_i^2)} \frac{\langle u_i, g \rangle}{s_i} \, v_i$$

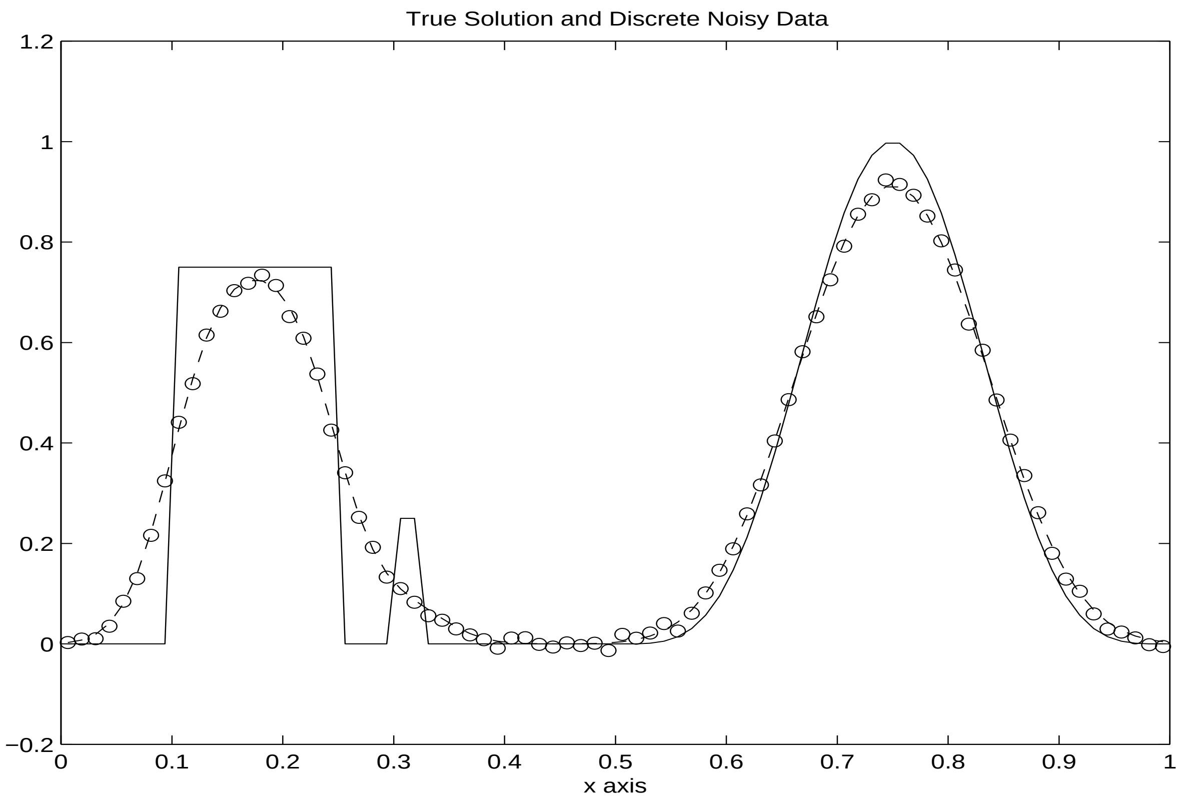
Below are filter functions for $k = 10$ and $k = 100$ iterations. Iteration k plays the role of regularization parameter.



Effect of Regularization Parameter

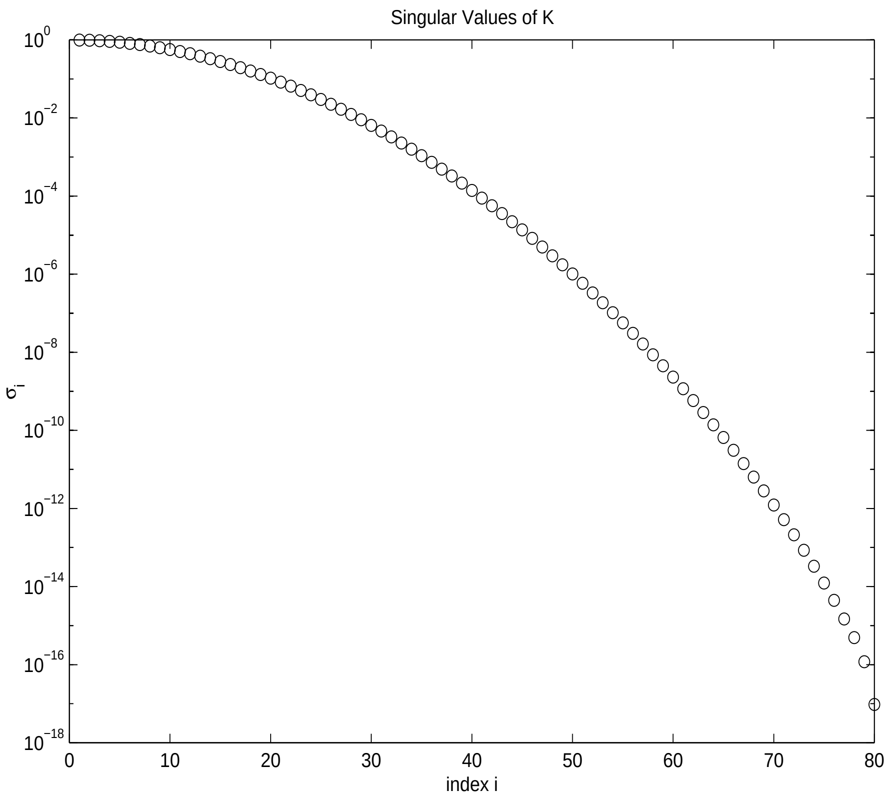
Illustrative Example: 1-D deconvolution with Gaussian kernel $a(t) = C \exp(-t^2/2\gamma^2)$ and discrete data

$$d_i = \int_0^1 a(s_i - t)\, x_{\text{true}}(t)\, dt + \text{noise}, \quad i = 1, \ldots, n.$$

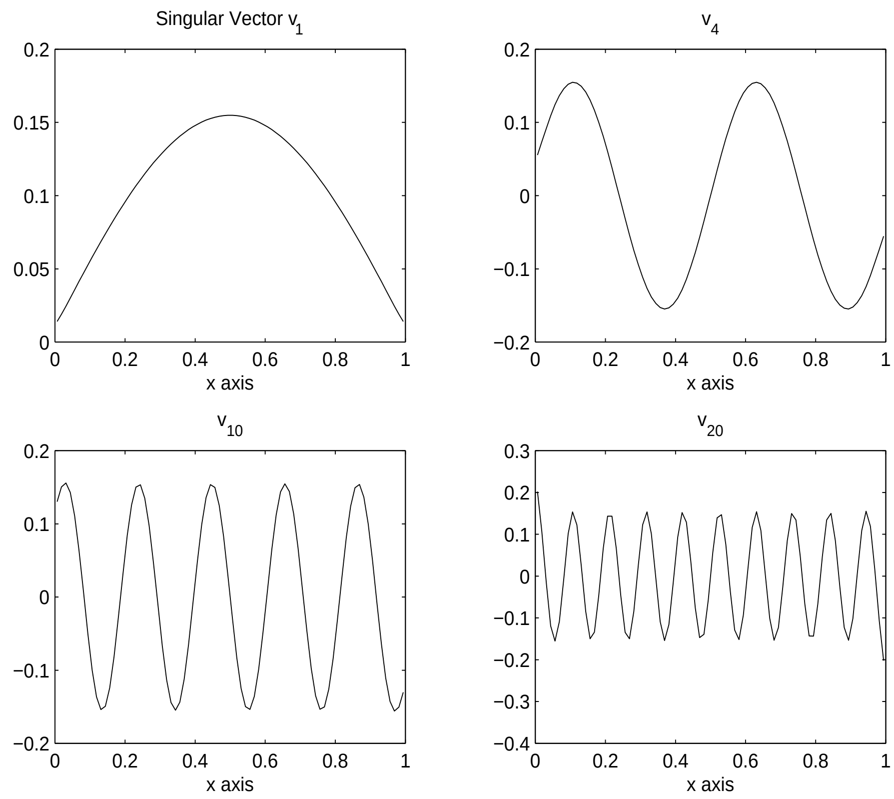


Graphical Representation of SVD

Singular Values

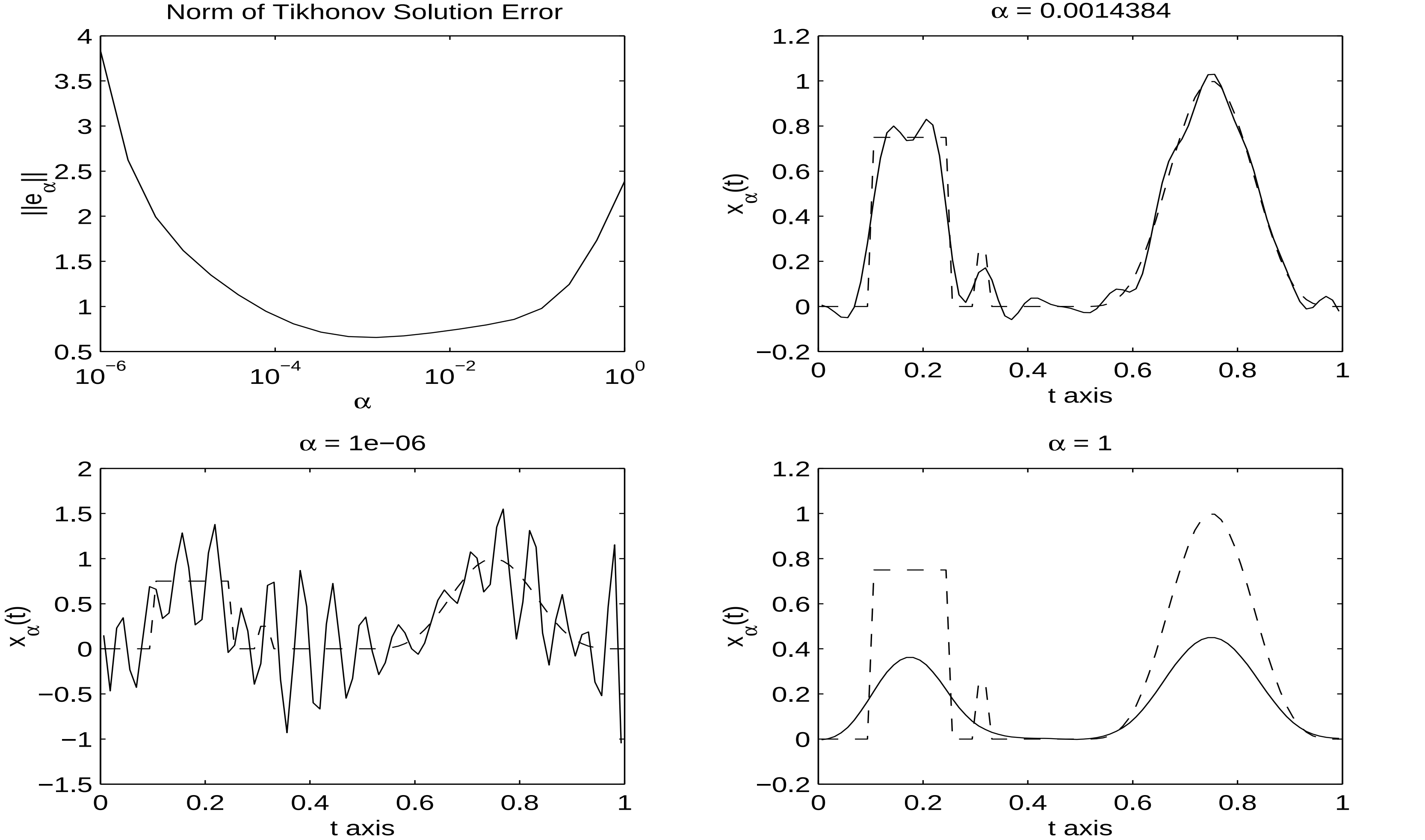


Singular Vectors



Tikhonov Solutions vs α

Tikhonov regularized solution is $\mathbf{x}_\alpha = (A^*A + \alpha I)^{-1} A^*\mathbf{d}$.
Solution error is $\mathbf{e}_\alpha = \mathbf{x}_\alpha - \mathbf{x}_{\text{true}}$.



Observations

- Instability is caused by amplification of error components corresponding to small singular values.
- Small singular values have corresponding singular vectors which are highly oscillatory (have "high spatial frequency").
- When regularization parameter α is too small, high frequency artifacts appear.
- When regularization parameter is too large, high and low frequency components of the solution are filtered out. Solution is too smooth.
- There seems to be an optimal value of α, at least in this example.

Exercises

1. The norm of a linear operator $A : \mathcal{F} \to \mathcal{G}$ is given by

$$||A|| \stackrel{\text{def}}{=} \sup_{||f||_{\mathcal{F}} \leq 1} ||Af||_{\mathcal{G}}$$

   Use the SVD of A to show that $||A|| = \sup_i s_i$.

2. Use the SVD to derive the linear filtering representation for Landweber iteration.
3. What is the effect of varying the step length parameter τ in Landweber iteration? Generate some plots to illustrate your answer to this question.
4. Implement TSVD regularization for the test problem presented above, and generate graphs similar to those on p. 14.

References

C. R. Vogel, *Computational Methods for Inverse Problems*, SIAM, 2002.