Image Deblurring

Fall 2005

Notes on Chapter 6

Dianne P. O'Leary

Regularization by Spectral Filtering

We know already that filtering is needed when noise is present, since the solution $\mathbf{x}_{naive} = \mathbf{A}^{-1}\mathbf{b}$ is typically too contaminated by noise to be useful.

Now we take a closer look at filtering.

Filtering is also called regularization because it can be interpreted as enforcing certain regularity conditions on the solution.

The degree of regularization is governed by a regularization parameter which should be chosen carefully.

We focus on two candidate regularization methods

- TSVD
- Tikhonov

and three candidate ways to compute the regularization parameter

- the discrepancy principle,
- generalized cross validation,
- the L-curve criterion.

Reference: Chapter 6 of Deblurring Images.

Two Important Regularization Methods

Recall the equation for our filtered solution

$$\mathbf{x}_{\text{filt}} = \sum_{i=1}^{N} \phi_i \frac{\mathbf{u}_i^T \mathbf{b}}{\sigma_i} \mathbf{v}_i$$

We need to choose the filter factors ϕ_i to control the spectral contents of the deblurred images.

The spectral coordinate system

$$\mathbf{x}_{\text{filt}} = \sum_{i=1}^{N} \phi_i \frac{\mathbf{u}_i^T \mathbf{b}}{\sigma_i} \mathbf{v}_i$$

Note that we have a coordinate system determined by the matrix $\mathbf{A}$:

- The data $\mathbf{b}$ is expressed in the coordinates $\mathbf{u}_i^T \mathbf{b}$ determined by the vectors $\mathbf{u}_i$ $(i = 1, \ldots, N)$.
- The solution $\mathbf{x}_{\text{filt}}$ is expressed in coordinates $\mathbf{v}_i^T \mathbf{x}$ determined by the vectors $\mathbf{v}_i$ $(i = 1, \ldots, N)$.

This is the spectral coordinate system, since these vectors are the eigenvectors of $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}\mathbf{A}^T$ respectively.

Our goal is to scale the solution component in the direction $\mathbf{v}_i$ by the filter factor ϕ_i in order to reduce the effect of error in the component $\mathbf{u}_i^T \mathbf{b}$.

Method 1: The Truncated SVD (TSVD) Method.

For this method, we define the filter factors to be

- one for large singular values,
- and zero for the rest.

More precisely,

$$\phi_i \equiv \begin{cases} 1, & i = 1, \ldots, k \\ 0, & i = k+1, \ldots, N. \end{cases}$$

The parameter k is called the truncation parameter and it determines the number of SVD components maintained in the regularized solution. Note that k always satisfies $1 \leq k \leq N$.

This is the method we used in Chapter 1 to improve upon the naïve method.

Method 2: The Tikhonov Method.

For this method we define the filter factors to be

$$\phi_i = \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2}, \qquad i = 1, \ldots, N,$$

where $\lambda > 0$ is called the regularization parameter.

This choice of filter factors yields the solution vector $\mathbf{x}_\lambda$ for the minimization problem

$$\min_{\mathbf{x}} \left\{ \|\mathbf{b} - \mathbf{A}\,\mathbf{x}\|_2^2 + \lambda^2 \|\mathbf{x}\|_2^2 \right\}.$$

This choice keeps $\|\mathbf{b} - \mathbf{A}\,\mathbf{x}\|_2$ small, but not so small that

$$\|\mathbf{x}\|_2^2 = \sum_{i=1}^{N} \phi_i^2 \frac{(\mathbf{u}_i^T \mathbf{b})^2}{\sigma_i^2}$$

is too big.

Thus, our minimization problem ensures that both the norm of the residual $\mathbf{b} - \mathbf{A}\,\mathbf{x}_\lambda$ and the norm of the solution $\mathbf{x}_\lambda$ are somewhat small.

How does λ affect the Tikhonov solution?

Affect on components for large singular values

Suppose $\sigma_i \gg \lambda$ (which is the case for some of the first filter factors).

Then, using the Taylor expansion $(1+\epsilon)^{-1} = 1 - \epsilon + \frac{1}{2}\epsilon^2 + O(\epsilon^3)$, we obtain

$$\phi_i = \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2} = \frac{1}{1 + \lambda^2/\sigma_i^2} = 1 - \frac{\lambda^2}{\sigma_i^2} + \frac{1}{2}\frac{\lambda^4}{\sigma_i^4} + \ldots \approx 1.$$

Affect on components for small singular values

Suppose $\sigma_i \ll \lambda$ (which is the case for some of the last filter factors). Again using the Taylor expansion of $(1+\epsilon)^{-1}$, we obtain

$$\phi_i = \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2} = \frac{\sigma_i^2}{\lambda^2}\frac{1}{1 + \sigma_i^2/\lambda^2} = \frac{\sigma_i^2}{\lambda^2}\left(1 - \frac{\sigma_i^2}{\lambda^2} + \frac{1}{2}\frac{\sigma_i^4}{\lambda^4} + \ldots\right).$$

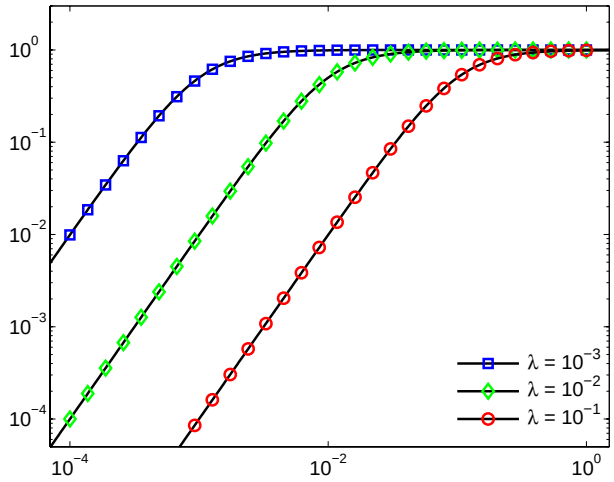
Thus we can conclude that the Tikhonov filter factors satisfy

$$\phi_i = \begin{cases} 1 - \left(\dfrac{\lambda}{\sigma_i}\right)^2 + \mathcal{O}\left(\left(\dfrac{\lambda}{\sigma_i}\right)^4\right), & \sigma_i \gg \lambda \\ \left(\dfrac{\sigma_i}{\lambda}\right)^2 + \mathcal{O}\left(\left(\dfrac{\sigma_i}{\lambda}\right)^4\right), & \sigma_i \ll \lambda. \end{cases}$$

Answer to how λ affects the Tikhonov solution

- $\phi_i \approx 1$ when $\lambda << \sigma_i$
- $\phi_i \approx \sigma_i^2/\lambda^2$ when $\lambda \geq \sigma_i$.

Therefore, λ determines the breakpoint at which the filter factors change nature: the point at which $\sigma_i \approx \lambda$.



The Tikhonov filter factors $\phi_i = \sigma_i^2/(\sigma_i^2 + \lambda^2)$ versus σ_i for three different values of the regularization parameter λ.

Relation between TSVD and Tikhonov

The parameter λ in Tikhonov's method acts in the same way as the parameter k in the TSVD method: it controls which SVD components we want to dampen or filter.

We also see that there is no point in choosing λ outside the interval $[\sigma_N, \sigma_1]$.

Implementation Issues

Implementation of Filtering Methods

If all of the singular values of $\mathbf{A}$ are nonzero, then the inverse solution can be written as

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} = \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^T\mathbf{b}\,.$$

Similarly, the spectral filter solution can be written as

$$\mathbf{x}_{\text{filt}} = \mathbf{V}\mathbf{\Phi}\mathbf{\Sigma}^{-1}\mathbf{U}^T\mathbf{b}$$

where $\mathbf{\Phi}$ is a diagonal matrix consisting of the filter factors ϕ_i for the particular method:

- 1's and 0's for TSVD and
- $\sigma_i^2/(\sigma_i^2 + \lambda^2)$ for Tikhonov.

Computational issues include:

- Exploiting structure in $\mathbf{A}$.
- Specifying the regularization parameter.
- Avoiding divide-by-zero.

Exploiting structure in $\mathbf{A}$

Recall:

- Structured matrices arise in image deblurring problems: e.g., Kronecker products, BTTB, etc.
- The SVD or spectral decomposition of such matrices can be computed efficiently.
- The (naïve) inverse solution is also easy.

Similarly, the TSVD and Tikhonov solutions can be computed efficiently.

A rewrite

Old equation:

$$\mathbf{x}_{\text{filt}} = \mathbf{V}\mathbf{\Phi}\mathbf{\Sigma}^{-1}\mathbf{U}^T\mathbf{b}$$

New equation:

$$\mathbf{x}_{\text{filt}} = \mathbf{V}\mathbf{\Sigma}_{\text{filt}}^{-1}\mathbf{U}^T\mathbf{b}$$

where $\mathbf{\Sigma}_{\text{filt}}^{-1} = \mathbf{\Phi}\mathbf{\Sigma}^{-1}$.

Thus, given the filter factors, it is simple to to compute $\mathbf{x}_{\text{filt}}$.

Filtered solutions for structured matrices

- Given:

  ```
  P = PSF
  center = [row, col] = center of PSF
  B = blurred image
  BC = string denoting boundary condition (e.g., 'zero')
  Phi = filter factors
  ```

- For periodic boundary conditions, use:

  ```
  S = fft2( circshift(P, 1 - center) );
  Sfilt = Phi ./ S;
  Xfilt = real( ifft2( fft2(B) .* Sfilt ) );
  ```

- For reflexive boundary conditions, with strongly symmetric PSF, use:

  ```
  e1 = zeros(size(P));, e1(1,1) = 1;
  S = dct2( dctshift(P, center) ) ./ dct2(e1);
  Sfilt = Phi ./ S;
  Xfilt = idct2( dct2(B) .* Sfilt ) ;
  ```

- For a separable PSF, use:

  ```
  [Ar, Ac] = kronDecomp(P, center, BC);
  [Uc, Sc, Vc] = svd(Ac);
  [Ur, Sr, Vr] = svd(Ar);
  S = diag(Sc) * diag(Sr)';
  Sfilt = Phi ./ S;
  Xfilt = Vc * ( (Uc' * B * Ur) .* Sfilt ) * Vr';
  ```

Specifying the regularization parameter

- The TSVD truncation index should statisfy $1 \leq k \leq N$.
- The Tikhonov parameter should satisfy $\sigma_N \leq \lambda \leq \sigma_1$.

Later we discuss automatic methods for estimating good choices for these parameters, but for now we can try to choose them experimentally.

Specifying the TSVD parameter

In the case of TSVD, we might specify a tolerance below which all singular (spectral) values are truncated. In this case the filter factors can be computed very easily as:

```
Phi = ( abs(S) >= tol );
```

By experimenting with various values of `tol`, and displaying the computed filter solution, `Xfilt`, we can see the effects of regularization.

Specifying the Tikhonov parameter

In the case of Tikhonov regularization, we could specify a value for λ, and compute the filter factors from the singular (spectral) values as follows:

```
Phi = abs(S).^2 ./ (abs(S).^2 + lambda^2);
```

Note that the use of `abs` is necessary in the case when FFTs are used.

Again, we can experiment with various values of `lambda` and display the filtered solution to see the effects of regularization.

Avoiding divide-by-zero

In computing the quantity

```
Sfilt = Phi ./ S
```

we will commit divide-by-zero if any singular (spectral) value is zero.

This will cause some values of `Sfilt` to be set to `Inf` or to `NaN`.

To avoid this, perform the computation only for nonzero values of `S`, and set all other `Sfilt` values to 0.

```
idx = (S ~= 0);
Sfilt = zeros(size(Phi));
Sfilt(idx) = Phi(idx) ./ S(idx);
```

Regularization Errors and Perturbation Errors

Regularization Errors and Perturbation Errors

Recall: $\mathbf{x}_{\text{filt}}$ can always be written in the SVD framework as

$$\mathbf{x}_{\text{filt}} = \mathbf{V}\,\mathbf{\Phi}\,\mathbf{\Sigma}^{-1}\mathbf{U}^T\mathbf{b},$$

where $\mathbf{\Phi}$ is a diagonal matrix consisting of the spectral filters ϕ_i for the particular method

- 0s and 1s for TSVD,
- $\sigma_i^2/(\sigma_i^2 + \lambda^2)$ for Tikhonov,
- etc.

Equipped with this formulation, we can now easily separate the two different types of errors in a regularized solution

$$\begin{aligned}
\mathbf{x}_{\text{filt}} &= \mathbf{V}\,\mathbf{\Phi}\,\mathbf{\Sigma}^{-1}\mathbf{U}^T\mathbf{b} \\
&= \mathbf{V}\,\mathbf{\Phi}\,\mathbf{\Sigma}^{-1}\mathbf{U}^T\mathbf{b}^{\text{exact}} + \mathbf{V}\,\mathbf{\Phi}\,\mathbf{\Sigma}^{-1}\mathbf{U}^T\mathbf{e} \\
&= \mathbf{V}\,\mathbf{\Phi}\,\mathbf{\Sigma}^{-1}\mathbf{U}^T A\,\mathbf{x}^{\text{exact}} + \mathbf{V}\,\mathbf{\Phi}\,\mathbf{\Sigma}^{-1}\mathbf{U}^T\mathbf{e} \\
&= \mathbf{V}\,\mathbf{\Phi}\,\mathbf{V}^T\mathbf{x}^{\text{exact}} + \mathbf{V}\,\mathbf{\Phi}\,\mathbf{\Sigma}^{-1}\mathbf{U}^T\mathbf{e}
\end{aligned}$$

and therefore the error in $\mathbf{x}_{\text{filt}}$ is given by

$$\mathbf{x}^{\text{exact}} - \mathbf{x}_{\text{filt}} = (\mathbf{I}_N - \mathbf{V}\,\mathbf{\Phi}\,\mathbf{V}^T)\mathbf{x}^{\text{exact}} - \mathbf{V}\,\mathbf{\Phi}\,\mathbf{\Sigma}^{-1}\mathbf{U}^T\mathbf{e}.$$

Two contributions to the errors:

- Regularization error $(\mathbf{I}_N - \mathbf{V}\,\mathbf{\Phi}\,\mathbf{V}^T)\mathbf{x}^{\text{exact}}$, caused by using a regularized inverse $\mathbf{V}\,\mathbf{\Phi}\,\mathbf{\Sigma}^{-1}\mathbf{U}^T$ (instead of the inverse $\mathbf{A}^{-1} = \mathbf{V}\,\mathbf{\Sigma}^{-1}\mathbf{U}^T$) in order to obtain the filtering.
- Perturbation error $\mathbf{V}\,\mathbf{\Phi}\,\mathbf{\Sigma}^{-1}\mathbf{U}^T\mathbf{e}$, which consists of the inverted and filtered noise.

Changing the regularization parameter changes the size of these errors.

- When too many filter factors ϕ_i are close to one, then
  - the regularization error is small,
  - the perturbation error is large.

The solution is undersmoothed.

- When too few filter factors are close to one, then
  - the regularization error is large,
  - the perturbation error is small.

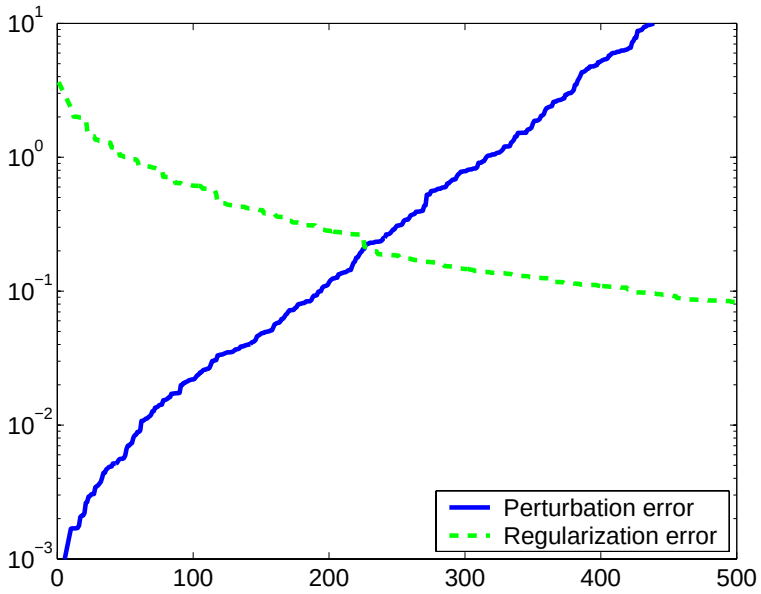
  The solution is oversmoothed.

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

Example

Consider TSVD as the regularization method.

We see in the next plot that the two types of errors are balanced for $k \approx 200$:



The 2-norms of the regularization error $(\mathbf{I}_N - \mathbf{V}\,\mathbf{\Phi}\,\mathbf{V}^T)\mathbf{x}^{\text{exact}}$ and the perturbation error $\mathbf{V}\,\mathbf{\Phi}\,\mathbf{\Sigma}^{-1}\mathbf{U}^T\mathbf{e}$ versus the truncation parameter k for the TSVD method.

The Resolution matrix

The matrix $\mathbf{V}\,\mathbf{\Phi}\,\mathbf{V}^T$ is called the resolution matrix for the regularized solution; it describes the mapping between the exact solution and the filtered component in $\mathbf{x}_{\text{filt}}$.

- The closer the resolution matrix is to the identity, the smaller the regularization error, but the inverted noise will dominate.
- On the other hand, when most of the filter factors are small (or zero), then the inverted noise is heavily damped (the perturbation error is small) – but the resolution matrix is far from the identity and the regularization error is large.

The importance of the Discrete Picard Condition

The reason why we are able to compute regularized approximations to the exact solution, in spite of the large condition number, is that spectral filtering suppresses much of the inverted noise while – at the same time – keeping the regularization error small.

This is possible because the image deblurring problem satisfies the discrete Picard condition – the exact right-hand side exhibits decaying expansion coefficients when expressed in the spectral basis.

As a consequence, the noise affects primarily the high-frequency components which are associated with the smaller singular values, and which are damped by the spectral filtering method.

What is left in the regularized solution is primarily the low-frequency SVD components associated with the larger singular values, and these components are dominated by the contributions for the exact right-hand side.

Consider the norm of the regularization error:
$\mathbf{x}^{\text{exact}} = \mathbf{A}^{-1}\mathbf{b}^{\text{exact}} = \mathbf{V}\,\mathbf{\Sigma}^{-1}\mathbf{U}^T\mathbf{b}^{\text{exact}}$, so

$$\begin{aligned} \|(\mathbf{I}_N - \mathbf{V}\,\mathbf{\Phi}\,\mathbf{V}^T)\mathbf{x}^{\text{exact}}\|_2 &= \|(\mathbf{I}_N - \mathbf{\Phi})\,\mathbf{V}^T\mathbf{x}^{\text{exact}}\|_2 \\ &= \|(\mathbf{I}_N - \mathbf{\Phi})\,\mathbf{\Sigma}^{-1}\mathbf{U}^T\mathbf{b}^{\text{exact}}\|_2 \\ &= \left(\sum_{i=1}^{N}\left((1-\phi_i)\,\mathbf{u}_i^T\mathbf{b}^{\text{exact}}/\sigma_i\right)^2\right)^{1/2}. \end{aligned}$$

- Due to the discrete Picard condition, the coefficients $|\mathbf{u}_i^T\mathbf{b}^{\text{exact}}/\sigma_i|$ decay (on average).
- Since the first filter factors ϕ_i (for $i = 1, 2, \ldots$) are close to one, the factors $(1-\phi_i)$ dampen the contributions from the larger coefficients $\mathbf{u}_i^T\mathbf{b}^{\text{exact}}/\sigma_i$.
- Moreover, the small filter factors ϕ_i (for $i = N, N-1, \ldots$) correspond to factors $(1-\phi_i)$ close to one, which are multiplied by small coefficients $\mathbf{u}_i^T\mathbf{b}^{\text{exact}}/\sigma_i$.
- Hence we conclude that if the filters are suitably chosen, then the norm of the regularization error cannot be large.

Parameter Choice Methods

We describe three important parameter choice methods:

- the discrepancy principle,
- generalized cross-validation,
- the L-curve criterion.

Effects of parameter choice

- The spectral filtering solution

$$\|\mathbf{x}_{\text{filt}}\|_2^2 = \sum_{i=1}^{N} \left(\phi_i \frac{\mathbf{u}_i^T \mathbf{b}}{\sigma_i} \right)^2$$

- The norm of the residual

$$\|\mathbf{b} - \mathbf{A}\,\mathbf{x}_{\text{filt}}\|_2^2 = \sum_{i=1}^{N} \left((1 - \phi_i) \mathbf{u}_i^T \mathbf{b} \right)^2 .$$

- TSVD method:
  - the norm of the solution is a monotonically nondecreasing function of k,
  - the residual norm is monotonically nonincreasing.

- For the Tikhonov method,
  - the norm of the solution $\mathbf{x}$ is a monotonically nonincreasing function of λ
  - the residual norm is monotonically nondecreasing.

The Discrepancy Principle.

Required information: a good estimate of δ, the expected value of $\|\mathbf{e}\|_2$ (the error in the observations $\mathbf{b}$).

This is powerful information, but often unreliable.

Idea: The regularization parameter should be chosen so that the norm of the residual is approximately δ.

$$\|\mathbf{b} - \mathbf{A}\,\mathbf{x}_{\text{filt}}\|_2 = \tau\delta,$$

where $\tau > 1$ is some predetermined real number.

As $\delta \to 0$, the filtered solution satisfies $\mathbf{x}_{\text{filt}} \to \mathbf{x}^{\text{exact}}$.

How to compute the solution: Systematically try different values of k or λ to satisfy the equation.

Cost: Given the SVD, the filter factors and the vector $U^T\mathbf{b}$, the cost is $2N$ multiplications and additions for each trial to compute the residual norm.

Generalized Cross-Validation (GCV).

Required information: In constrast to the discrepancy principle, the parameter choice in GCV does not depend on *a priori* knowledge about the noise variance.

Idea: If we omit a data value, then a good value of the parameter should be able to predict the missing data point well.

How to compute the solution: Determine the parameter λ that minimizes the GCV functional

$$G(\lambda) = \frac{\|(\mathbf{I}_N - \mathbf{A}\,\mathbf{V}\,\mathbf{\Phi}\,\mathbf{\Sigma}^{-1}\mathbf{U}^T)\mathbf{b}\|_2^2}{(\text{trace}(\mathbf{I}_N - \mathbf{A}\,\mathbf{V}\,\mathbf{\Phi}\,\mathbf{\Sigma}^{-1}\mathbf{U}^T))^2},$$

where

- λ is the Tikhonov parameter or, abusing notation, $\lambda = 1/k$ where k is the TSVD cutoff.
- $\mathbf{V}\,\mathbf{\Phi}\,\mathbf{\Sigma}^{-1}\mathbf{U}^T$ is the matrix that maps the right hand side $\mathbf{b}$ onto the regularized solution $\mathbf{x}_\lambda$.

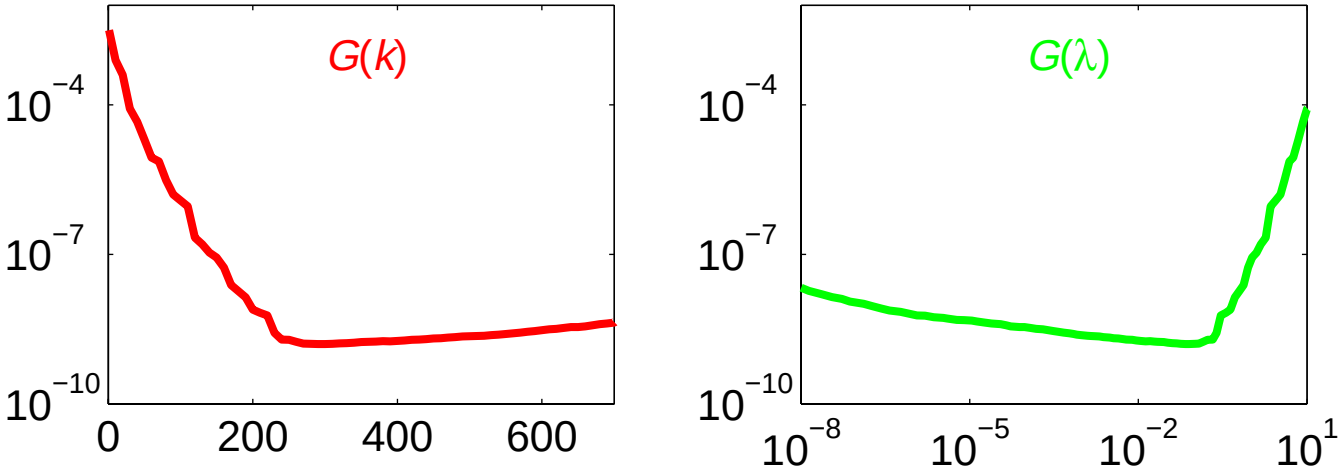
Cost:

- The numerator is just $\|\mathbf{b} - \mathbf{A}\,\mathbf{x}_{\text{filt}}\|_2^2$, for which we already have a formula.
- We evaluate the denominator by noting that the trace of a matrix is the sum of its main diagonal elements, and the trace is invariant under orthogonal transformation, so

$$\begin{aligned} \text{trace}(\mathbf{I}_N - \mathbf{A}\,\mathbf{V}\,\mathbf{\Phi}\,\mathbf{\Sigma}^{-1}\mathbf{U}^T) &= \text{trace}(\mathbf{U}(\mathbf{I}_N - \mathbf{\Phi})\mathbf{U}^T) \\ &= \text{trace}(\mathbf{I}_N - \mathbf{\Phi}) \\ &= N - \textstyle\sum_{i=1}^N \phi_i\,. \end{aligned}$$

In particular, for the TSVD method we have
$G(k) = \|\mathbf{b} - \mathbf{A}\,\mathbf{x}_k\|_2^2/(N-k)^2$.

- Given the SVD, the filters and $\mathbf{U}^T\mathbf{b}$ we can therefore compute $G(\lambda)$ in $2N$ multiplications and $3N$ additions.

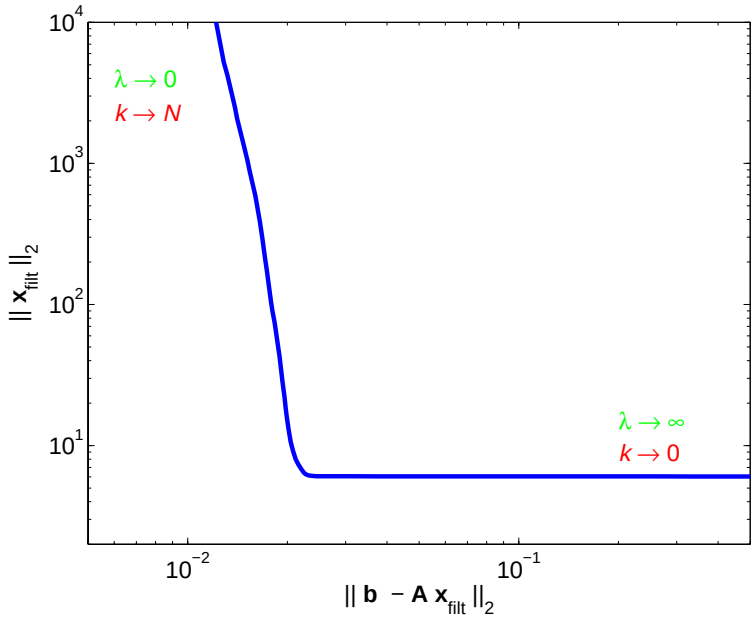


The GCV functions $G(k) = \|\mathbf{b} - \mathbf{A}\,\mathbf{x}_k\|_2^2/(N-k)^2$ for TSVD (left) and $G(\lambda)$ for Tikhonov regularization (right), applied to the same problem.

The L-Curve Criterion.

Required information: None.

Idea: The L-curve is a log-log plot of the norm of the regularized solution versus the corresponding residual norm for each of a set of regularization parameter values.



The L-curve for the TSVD method applied to the same problem. This plot often is in the shape of the letter L, from which it draws its name. The log-log scale emphasizes the L shape.

Intuitively, the best regularization parameter should lie at the corner of the L, since

- for values higher than this, the residual increases rapidly while the the norm of the solution decreases only slowly,
- for values smaller than this, the norm of the solution increases rapidly without much decrease in residual.

Hence, we expect a solution near the corner to balance the regularization and perturbation errors.

How to compute the solution: In practice, only a few points on the L-curve need to be computed, and the corner is located by estimating the point of maximum curvature.

Cost: Computing a point on the L-curve costs only $3N$ multiplications and additions and N divisions.

Which Choice Is Best?

Choosing an appropriate regularization parameter is very difficult.

Every parameter choice method, including the three we discussed, has severe flaws:

- either they require more information than is usually available,
- or they fail to converge to the true solution as the error norm goes to zero

---

Specific flaws

- The Discrepancy Principle is convergent as the noise goes to zero, but it relies on information that is often unavailable or erroneous. Even with a correct estimate of the variance, the solutions tend to be over-smoothed.
- For GCV, the solution estimates fail to converge to the true solution as the error norm goes to zero.
- Another noted difficulty with GCV is that the graph for G can be very flat near the minimizer, so that numerical methods have difficulty in determining a good value of λ.
- The L-Curve Criterion is usually more tractable numerically, but its limiting properties are far from ideal. The solution estimates fail to converge to the true solution as $N \to \infty$ or as the error norm goes to zero.

Implementation of GCV

(Similar details for the L-curve will be given in a later lecture by Per Christian.)

We want to minimize

$$G(\lambda) = \frac{\|(\mathbf{I}_N - \mathbf{A}\,\mathbf{V}\,\mathbf{\Phi}\,\mathbf{\Sigma}^{-1}\mathbf{U}^T)\mathbf{b}\|_2^2}{(\mathrm{trace}(\mathbf{I}_N - \mathbf{A}\,\mathbf{V}\,\mathbf{\Phi}\,\mathbf{\Sigma}^{-1}\mathbf{U}^T))^2},$$

so we need to evaluate it efficiently.

Specifically, in the case we are using the SVD, we obtain

$$G(\lambda) = \frac{||(\mathbf{I}_N - \mathbf{A}\mathbf{V}\mathbf{\Phi}\mathbf{\Sigma}^{-1}\mathbf{U}^T)\mathbf{b}||_2^2}{\mathrm{trace}(\mathbf{I}_N - \mathbf{A}\,\mathbf{V}\,\mathbf{\Phi}\,\mathbf{\Sigma}^{-1}\mathbf{U}^T)^2} = \frac{||(\mathbf{I}_N - \mathbf{\Phi})\mathbf{U}^T\mathbf{b}||_2^2}{\mathrm{trace}(\mathbf{I}_N - \mathbf{\Phi})^2}\,.$$

A similar simplification can be done for spectral decompositions.

Consider now specific regularization methods:

- **GCV for TSVD.**

$$G(k) = \frac{\sum_{i=k+1}^{N} \hat{b}_i^2}{(N-k)^2}$$

where $\hat{\mathbf{b}} = \mathbf{U}^T\mathbf{b}$. Note that this is a discrete function. The truncation index is found by evaluating $G(k)$ for $k = 1, 2, \ldots, N-1$, and finding the index at which $G(k)$ attains its minimum.

- **GCV for Tikhonov.**

$$G(\lambda) = \frac{\sum_{i=1}^{N} \left(\frac{\hat{b}_i}{\sigma_i^2 + \lambda^2} \right)^2}{\left(\sum_{i=1}^{N} \frac{1}{\sigma_i^2 + \lambda^2} \right)^2}$$

where $\hat{\mathbf{b}} = \mathbf{U}^T\mathbf{b}$. To find the minimum of this continuous function we can use MATLAB's built-in routine `fminbnd`. For example, if we implement the GCV function as:

```
function G = GCV(lambda, bhat, s)
  t = 1 ./ (s.^2 + lambda^2);
  G = sum((bhat .* t).^2)/(sum(t)^2);
end
```

Then the "optimal" λ can be found using:

```
lambda = fminbnd(@GCV,min(s),max(s),[],bhat,s);
```

where `s`=diag(S) and `bhat`= $\mathbf{U}^T\mathbf{b}$.

If the spectral decomposition is used instead of the SVD, the values in `S` and `bhat` may be complex, and so absolute values must be included with the squaring operations.

See `gcv_tik` and `gcv_tsvd` for details on exploiting matrix structure in these computations.

Statistical Aspects

A few more details on the statistics of the error

Consider the SVD analysis of the noise and the inverted noise.

We first note that the coefficients $\mathbf{u}_i^T\mathbf{b}$ in the spectral expansion are the elements of the vector

$$\mathbf{U}^T\mathbf{b} = \mathbf{U}^T\mathbf{b}^{\text{exact}} + \mathbf{U}^T\mathbf{e}.$$

Assume that the elements of the vector $\mathbf{e}$ are statistically independent, with zero mean and identical standard deviation.

Then the expected value of $\mathbf{e}$ is the zero vector, while its covariance matrix is a scaled identity matrix,

$$\mathcal{E}(\mathbf{e}) = \mathbf{0}, \qquad \text{Cov}(\mathbf{e}) = \mathcal{E}(\mathbf{e}\,\mathbf{e}^T) = \eta^2\mathbf{I}_N,$$

where $\eta > 0$ is the standard deviation.

Then it follows that the expected value of the vector $\mathbf{U}^T\mathbf{e}$ is also the zero vector, $\mathcal{E}(\mathbf{U}^T\mathbf{e}) = \mathbf{0}$, and that the covariance matrix for $\mathbf{U}^T\mathbf{e}$ is given by

$$\text{Cov}(\mathbf{U}^T\mathbf{e}) = \mathbf{U}^T\text{Cov}(\mathbf{e})\,\mathbf{U} = \eta^2\mathbf{U}^T\mathbf{U} = \eta^2\mathbf{I}_N.$$

Hence the coefficients $\mathbf{u}_i^T\mathbf{e}$ behave, statistically, like the elements of the noise vector $\mathbf{e}$.

The expected value of $(\mathbf{u}_i^T\mathbf{b})^2$ is

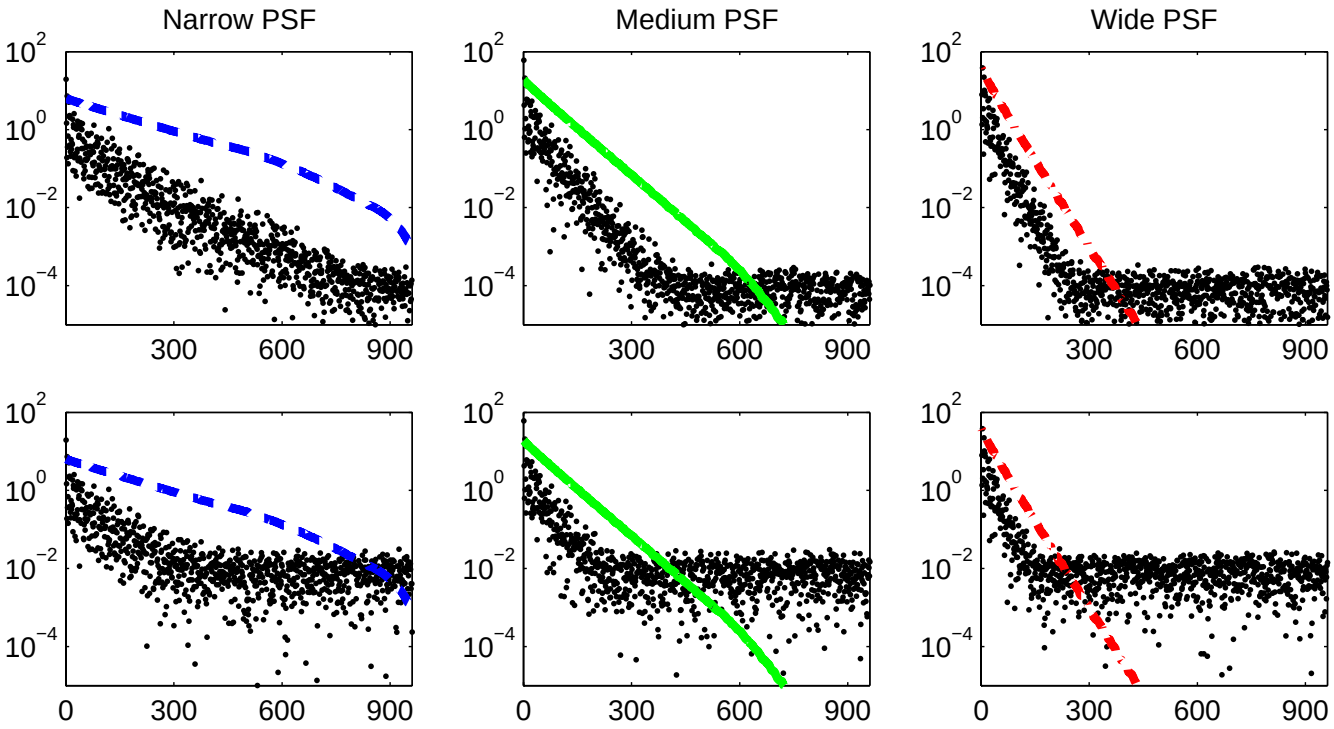
$$\begin{aligned}
\mathcal{E}\big((\mathbf{u}_i^T\mathbf{b})^2\big) &= \mathcal{E}\Big((\mathbf{u}_i^T\mathbf{b}^{\text{exact}} + \mathbf{u}_i^T\mathbf{e})^2\Big) \\
&= \mathcal{E}\Big((\mathbf{u}_i^T\mathbf{b}^{\text{exact}})^2 + 2\,\mathbf{u}_i^T\mathbf{b}^{\text{exact}}\,\mathbf{u}_i^T\mathbf{e} + (\mathbf{u}_i^T\mathbf{e})^2\Big) \\
&= (\mathbf{u}_i^T\mathbf{b}^{\text{exact}})^2 + \eta^2
\end{aligned}$$

(because $\mathcal{E}(\mathbf{u}_i^T\mathbf{e}) = 0$), and

$$\mathcal{E}\big(|\mathbf{u}_i^T\mathbf{b}|\big) \approx \sqrt{\mathcal{E}\big((\mathbf{u}_i^T\mathbf{b})^2\big)} = \sqrt{(\mathbf{u}_i^T\mathbf{b}^{\text{exact}})^2 + \eta^2}.$$

We conclude that for any index i where $|\mathbf{u}_i^T\mathbf{b}^{\text{exact}}|$ is somewhat larger than η we have $\mathbf{u}_i^T\mathbf{b} \approx \mathbf{u}_i^T\mathbf{b}^{\text{exact}}$, while $\mathcal{E}(|\mathbf{u}_i^T\mathbf{b}|) \approx \eta$ when $|\mathbf{u}_i^T\mathbf{b}^{\text{exact}}|$ is smaller than η.

Recall this plot:



Plots of singular values σ_i (colored lines) and coefficients $|u_i^T \mathbf{b}|$ (black dots) for the three matrices defined by various PSFs and two different noise levels in $\mathbf{B} = \mathbf{B}^{\text{exact}} + \mathbf{E}$. Top row: $\|\mathbf{E}\|_{\text{F}} = 3 \cdot 10^{-3}$; bottom row: $\|\mathbf{E}\|_{\text{F}} = 3 \cdot 10^{-1}$.

It is now evident that for small indices i the quantities $\mathbf{u}_i^T \mathbf{b}$ are indeed dominated by the component $\mathbf{u}_i^T \mathbf{b}^{\text{exact}}$ (which has an overall decreasing behavior), while for larger indices we have $\mathbf{u}_i^T \mathbf{b} \approx \mathbf{u}_i^T \mathbf{e} \approx \eta$ whose statistical behavior is identical to that of $\mathbf{e}$. We have thus explained the overall behavior of the plot.

We can also say something about the statistical properties of the regularized solution $\mathbf{x}_{\text{filt}}$. Assuming again that $\eta^2 \mathbf{I}_N$ is the covariance matrix for the errors $\mathbf{e}$ in the right-hand side, the covariance matrix for the errors in the naïve solution $\mathbf{x}^{\text{naive}} = \mathbf{A}^{-1}\mathbf{b}$ and the filtered solution $\mathbf{x}_{\text{filt}} = \mathbf{V}\,\mathbf{\Phi}\,\mathbf{\Sigma}^{-1}\mathbf{U}^T\mathbf{b}$ are

$$\begin{aligned} \text{Cov}(\mathbf{x}^{\text{naive}}) &= \eta^2\,\mathbf{A}^{-1}\mathbf{A}^{-T} = \eta^2\,\mathbf{V}\,\mathbf{\Sigma}^{-2}\mathbf{V}^T = \eta^2 \sum_{i=1}^{N} \frac{1}{\sigma_i^2}\,\mathbf{v}_i\,\mathbf{v}_i^T, \\ \text{Cov}(\mathbf{x}_{\text{filt}}) &= \eta^2\,\mathbf{V}\,\mathbf{\Phi}^2\,\mathbf{\Sigma}^{-2}\mathbf{V}^T = \eta^2 \sum_{i=1}^{N} \frac{\phi_i^2}{\sigma_i^2}\,\mathbf{v}_i\,\mathbf{v}_i^T, \end{aligned}$$

showing that the elements in the latter covariance matrix are much smaller in magnitude than those in the former.

Summary

- For TSVD regularization, we choose the truncation parameter k so that the residual $\|\mathbf{b} - \mathbf{A}\mathbf{x}\|$ is reasonably small but the solution $\mathbf{x}$ does not include components corresponding to small singular values $\sigma_{k+1}, \ldots, \sigma_N$.
- Exploit structure in $\mathbf{A}$ when computing the TSVD or Tikhonov solutions.
- Practical implementations of filtering methods should avoid possible division by zero.
- Regularization by means of spectral filtering requires
  - Choosing a suitable filter and a corresponding regularized inverse so that the resolution matrix $\mathbf{V}\,\mathbf{\Phi}\,\mathbf{V}^T$ is sufficiently close to the identity matrix, and
  - Finding a suitable balance between the regularization error and the perturbation error.
- No parameter choice method is perfect, and the choice between the Discrepancy Principle, GCV, the L-Curve, and other methods is dependent on what information is available about the problem.