SOLUTION OF ILL-POSED INVERSE PROBLEMS PERTAINING TO SIGNAL RESTORATION: TOTAL VARIATION RESTORATION

Rosemary Renaut
http://math.asu.edu/~rosie

MANCHESTER
MARCH 23, 2012
Outline

Background
- Examples of Ill-posed problems
- SVD for examining the solution
- Importance of the Basis
- Picard Condition for Ill-Posed Problems

Generalized regularization
- GSVD for examining the solution

Alternative Regularizers: Total Variation
- Algorithms for TV

Regularization Parameter Estimation for TV

Conclusions and Future
Illustration: Blurred Signal Restoration

\[ b(t) = \int_{-\pi}^{\pi} h(t, s)x(s)ds \]

\[ h(s) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(\frac{-s^2}{2\sigma^2}\right), \sigma > 0. \]
Forward Problem: Given $x$, $A$ calculate $b$

- $x$ the source is sampled at $s_j$, $b(t)$ is measured at $t_i$.
- The kernel $h$ describes matrix $A$: integral is approximated using numerical quadrature (weights $w_j$) yielding a matrix $A$.
- Matrix equation
  \[ \mathbf{b} = A\mathbf{x} \]
  describes the linear relationship
- Generally $\int_{a}^{b} h(x) dx = 1$ (for PSF no energy is introduced, the signal is spread) leads to the normalization $\sum_j w_j h(x_j) = 1$.
- When $h(s, t) = h(t - s)$: kernel is **spatially invariant**.
- Typical choice of $h(x)$ is the Gaussian blur - a low pass filter. Also interested in the general integral equation problem.
Inverse Problem given $A, b$ find $x$

The solution depends on the conditioning of $A$ and on the noise in measurements $b$. Condition of $A$ is $1.8679e + 05$ Restoration with noise .0001: $x = A^{-1}b$

(Almost committing the inverse crime)
Example of restoration of 2D image with low noise

Figure: Original
Example of restoration of 2D image with low noise

Figure: Noisy and Blurred $10^{-5}$
Example of restoration of 2D image with low noise

Figure: Restored $10^{-5}$ inverse crime
Example of restoration of 2D image with low noise

**Figure:** Noisy and Blurred $10^{-3}$
Example of restoration of 2D image with low noise

**Figure:** Restored $10^{-3}$
Goals  Given data $b$ and kernel $h$ find $x$

Features  One may wish to find features from $x$, for example the edges in the signal

Difficulties  The solution is very sensitive to the data

Ill-Posed (according to Hadamard) A problem is ill-posed if it does not satisfy conditions for well-posedness, OR

1. $b \notin \text{range}(A)$
2. inverse is not unique because more than one image is mapped to the same data, or
3. an arbitrarily small change in $b$ can cause an arbitrarily large change in $x$.

Solution  Analyse the formulation.
Goals  Given data $b$ and kernel $h$ find $x$

Features One may wish to find features from $x$, for example the edges in the signal

Difficulties The solution is very sensitive to the data

Ill-Posed (according to Hadamard) A problem is ill-posed if it does not satisfy conditions for well-posedness, OR

1. $b \notin \text{range}(A)$
2. inverse is not unique because more than one image is mapped to the same data, or
3. an arbitrarily small change in $b$ can cause an arbitrarily large change in $x$.

Solution Analyse the formulation.
Summary

Goals  Given data $b$ and kernel $h$ find $x$

Features  One may wish to find features from $x$, for example the edges in the signal

Difficulties  The solution is very sensitive to the data

Ill-Posed  (according to Hadamard) A problem is ill-posed if it does not satisfy conditions for well-posedness, OR

1. $b \notin \text{range}(A)$
2. inverse is not unique because more than one image is mapped to the same data, or
3. an arbitrarily small change in $b$ can cause an arbitrarily large change in $x$.

Solution  Analyse the formulation.
Summary

Goals  Given data $b$ and kernel $h$ find $x$

Features  One may wish to find features from $x$, for example the edges in the signal

Difficulties  The solution is very sensitive to the data

Ill-Posed  (according to Hadamard) A problem is ill-posed if it does not satisfy conditions for well-posedness, OR

1. $b \notin \text{range}(A)$
2. inverse is not unique because more than one image is mapped to the same data, or
3. an arbitrarily small change in $b$ can cause an arbitrarily large change in $x$.

Solution  Analyse the formulation.
Summary

**Goals**  Given data \( \mathbf{b} \) and kernel \( h \) find \( \mathbf{x} \)

**Features**  One may wish to find features from \( \mathbf{x} \), for example the edges in the signal

**Difficulties**  The solution is very sensitive to the data

**Ill-Posed**  (according to Hadamard) A problem is **ill-posed** if it does not satisfy conditions for well-posedness, OR

1. \( \mathbf{b} \notin \text{range}(A) \)
2. inverse is not unique because more than one image is mapped to the same data, or
3. an arbitrarily small change in \( \mathbf{b} \) can cause an arbitrarily large change in \( \mathbf{x} \).

**Solution**  Analyse the formulation.
Consider general overdetermined discrete problem

\[ Ax = b, \quad A \in \mathbb{R}^{m \times n}, \quad b \in \mathbb{R}^m, \quad x \in \mathbb{R}^n, \quad m \geq n. \]

Singular value decomposition (SVD) of \( A \) (full column rank)

\[ A = U \Sigma V^T = \sum_{i=1}^{n} u_i \sigma_i v_i^T, \quad \Sigma = \text{diag}(\sigma_1, \ldots, \sigma_n). \]

gives expansion for the solution

\[ x = \sum_{i=1}^{n} \frac{u_i^T b}{\sigma_i} v_i \]

\( u_i, v_i \) are left and right singular vectors for \( A \)

Solution is a weighted linear combination of the basis vectors \( v_i \)
Spectral Decomposition of the Solution: The SVD

Consider general overdetermined discrete problem

\[ Ax = b, \quad A \in \mathbb{R}^{m \times n}, \quad b \in \mathbb{R}^m, \quad x \in \mathbb{R}^n, \quad m \geq n. \]

Singular value decomposition (SVD) of \( A \) (full column rank)

\[ A = U \Sigma V^T = \sum_{i=1}^{n} u_i \sigma_i v_i^T, \quad \Sigma = \text{diag}(\sigma_1, \ldots, \sigma_n). \]

gives expansion for the solution

\[ x = \sum_{i=1}^{n} \frac{u_i^T b}{\sigma_i} v_i \]

\( u_i, v_i \) are left and right singular vectors for \( A \)
Solution is a weighted linear combination of the basis vectors \( v_i \)
Consider general overdetermined discrete problem

\[ A\mathbf{x} = \mathbf{b}, \quad A \in \mathbb{R}^{m \times n}, \quad \mathbf{b} \in \mathbb{R}^{m}, \quad \mathbf{x} \in \mathbb{R}^{n}, \quad m \geq n. \]

Singular value decomposition (SVD) of \( A \) (full column rank)

\[ A = U \Sigma V^T = \sum_{i=1}^{n} u_i \sigma_i v_i^T, \quad \Sigma = \text{diag}(\sigma_1, \ldots, \sigma_n). \]

gives expansion for the solution

\[ \mathbf{x} = \sum_{i=1}^{n} \frac{u_i^T \mathbf{b}}{\sigma_i} v_i \]

\( u_i, v_i \) are left and right singular vectors for \( A \)

Solution is a weighted linear combination of the basis vectors \( v_i \)
Consider general overdetermined discrete problem

\[ A \mathbf{x} = \mathbf{b}, \quad A \in \mathbb{R}^{m \times n}, \quad \mathbf{b} \in \mathbb{R}^{m}, \quad \mathbf{x} \in \mathbb{R}^{n}, \quad m \geq n. \]

Singular value decomposition (SVD) of \( A \) (full column rank)

\[ A = U \Sigma V^T = \sum_{i=1}^{n} u_i \sigma_i v_i^T, \quad \Sigma = \text{diag}(\sigma_1, \ldots, \sigma_n). \]

gives expansion for the solution

\[ \mathbf{x} = \sum_{i=1}^{n} \frac{u_i^T \mathbf{b}}{\sigma_i} v_i \]

\( u_i, v_i \) are left and right singular vectors for \( A \)

Solution is a \textit{weighted} linear combination of the basis vectors \( v_i \).
Consider general overdetermined discrete problem

\[ A \mathbf{x} = \mathbf{b}, \quad A \in \mathbb{R}^{m \times n}, \quad \mathbf{b} \in \mathbb{R}^{m}, \quad \mathbf{x} \in \mathbb{R}^{n}, \quad m \geq n. \]

Singular value decomposition (SVD) of \( A \) (full column rank)

\[ A = U \Sigma V^T = \sum_{i=1}^{n} u_i \sigma_i v_i^T, \quad \Sigma = \text{diag}(\sigma_1, \ldots, \sigma_n). \]

gives expansion for the solution

\[ \mathbf{x} = \sum_{i=1}^{n} \frac{u_i^T \mathbf{b}}{\sigma_i} v_i \]

\( u_i, v_i \) are left and right singular vectors for \( A \)

Solution is a \textit{weighted} linear combination of the basis vectors \( v_i \)
Illustration: Example of the basis vectors

Problem wing \( n = 84 \): Evaluating impact of higher precision. Cond\((A) = 2.2056e + 19 \). Hansen’s Regularization Toolbox.

\[
    h(s, t) = se^{-ts^2}, \quad x(s) = 1, \quad \frac{1}{3} < t < \frac{2}{3} \quad b(t) = \frac{e^{-t/9} - e^{-4t/9}}{2t}
\]
Figure: The first few left singular vectors $u_i$ and basis vectors $v_i$. Are the results correct?
Use Matlab High Precision to examine the SVD

- Matlab digits allows high precision. Standard is 32.
- Symbolic toolbox allows operations on high precision variables with vpa.
- SVD for vpa variables calculates the singular values symbolically, but not the singular vectors.
- Higher accuracy for the SVs generates higher accuracy singular vectors.
- Solutions with high precision can take advantage of Matlab’s symbolic toolbox.
Left Singular Vectors and Basis Calculated in High Precision

**Figure:** The first few left singular vectors $u_i$ and basis vectors $v_i$. Apparently with higher precision we preserve the frequency content of the basis. How many can we use in the solution for $x$?
Suppose for a given vector \( y \) that it is a time series indexed by position, i.e. index \( i \).

**Diagnostic 1** Does the histogram of entries of \( y \) generate histogram consistent with \( y \sim \mathcal{N}(0, 1) \)? (i.e. independent normally distributed with mean 0 and variance 1) Not practical to automatically look at a histogram and make an assessment

**Diagnostic 2** Test the expectation that \( y_i \) are selected from a white noise time series. Take the Fourier transform of \( y \) and form cumulative periodogram \( z \) from power spectrum \( c \)

\[
c_j = |(\text{dft}(y)_j|^2, \quad z_j = \frac{\sum_{i=1}^{j} c_j}{\sum_{i=1}^{q} c_i}, \quad j = 1, \ldots, q,
\]

**Automatic:** Test is the line \((z_j, j/q)\) close to a straight line with slope 1 and length \(\sqrt{5}/2\)?
Suppose for a given vector $y$ that it is a time series indexed by position, i.e. index $i$.

**Diagnostic 1** Does the histogram of entries of $y$ generate histogram consistent with $y \sim \mathcal{N}(0, 1)$? (i.e. independent normally distributed with mean 0 and variance 1) Not practical to automatically look at a histogram and make an assessment

**Diagnostic 2** Test the expectation that $y_i$ are selected from a white noise time series. Take the Fourier transform of $y$ and form cumulative periodogram $z$ from power spectrum $c$

$$
c_j = |(\text{dft}(y)_j|^2, \quad z_j = \frac{\sum_{i=1}^{j} c_j}{\sum_{i=1}^{q} c_i}, \quad j = 1, \ldots, q,
$$

**Automatic:** Test is the line $(z_j, j/q)$ close to a straight line with slope 1 and length $\sqrt{5}/2$?
Power Spectrum for detecting white noise: a time series analysis technique

Suppose for a given vector $y$ that it is a time series indexed by position, i.e. index $i$.

**Diagnostic 1** Does the histogram of entries of $y$ generate histogram consistent with $y \sim \mathcal{N}(0, 1)$? (i.e. independent normally distributed with mean 0 and variance 1) Not practical to automatically look at a histogram and make an assessment.

**Diagnostic 2** Test the expectation that $y_i$ are selected from a white noise time series. Take the Fourier transform of $y$ and form cumulative periodogram $z$ from power spectrum $c$

$$c_j = |(\text{dft}(y)_j|^2, \quad z_j = \frac{\sum_{i=1}^{j} c_j}{\sum_{i=1}^{q} c_i}, \quad j = 1, \ldots, q,$$

**Automatic:** Test is the line $(z_j, j/q)$ close to a straight line with slope 1 and length $\sqrt{5}/2$?
Cumulative Periodogram for the left singular vectors

**Figure:** Standard precision on the left and high precision on the right: On left more vectors are close to white than on the right - vectors are white if the CP is close to the diagonal on the plot.
**Figure:** Standard precision on the left and high precision on the right: On left more vectors are close to white than on the right - if you count there are 9 vectors with true frequency content on the left.
Figure: Testing for white noise for the standard precision vectors: Calculate the cumulative periodogram and measure the deviation from the “white noise” line. In this case it suggests that about 9 vectors are noise free.

Cannot expect to use more than 9 vectors in the expansion for $x$. Additional terms are contaminated by noise - independent of noise in $b$. 

Measure Deviation from Straight Line: Basis Vectors
Figure: Testing for white noise for the standard precision vectors: Calculate the cumulative periodogram and measure the deviation from the “white noise” line. In this case it suggests that about 9 vectors are noise free.

Cannot expect to use more than 9 vectors in the expansion for $x$. Additional terms are contaminated by noise - independent of noise in $b$. 
Discrete Picard condition: examine the weights of the expansion

Recall

\[ x = \sum_{i=1}^{n} \frac{u_i^T b}{\sigma_i} v_i \]

Here

\[ |u_i^T b| / \sigma_i = \mathcal{O}(1) \]

Ratios are not large but are the values correct? Considering only the discrete Picard condition does not tell us whether the expansion for the solution is correct.
Discrete Picard condition: examine the weights of the expansion

- From high precision calculation of $\sigma_i$ shows they decay exponentially to zero (down to machine precision)
- The Picard condition considers ratios $|u_i^T b|/\sigma_i$ for data $b$; they decay exponentially (down to the machine precision).
- Note ratios in this case are $O(1)$ - hence noise contaminated basis vectors are not ignored. i.e. to approximate a solution with a discontinuity we need all the basis vectors.
- We may obtain solutions by truncating the SVD

$$ x = \sum_{i=1}^{k} \frac{u_i^T b}{\sigma_i} v_i $$

- Now parameter $k$ is a regularization parameter
- For given example we know $k < 10$ independent of the measured data $b$. We cannot see this from the Picard condition.
The Truncated Solutions (Noise free data b)

Figure: Truncated SVD Solutions: Standard precision $|u_i^T b|$. Error in the basis contaminates the solution
Figure: Truncated SVD Solutions: VPA calculation $|u_i^Tb|$. Number of terms not sufficient to represent the solution discontinuity
Observations

- Even when committing the inverse crime we will not achieve the solution if we cannot approximate the basis correctly.
- We need all basis vectors which contain the high frequency terms in order to approximate a solution with high frequency components - e.g. edges.
- Reminder - this is independent of the data.
- But is an indication of an ill-posed problem. In this case the data that is modified exhibits in the matrix $A$ decomposition.
- We look at a problem with a smoother solution - what are the issues?

Problem shaw from the regularization toolbox defined on interval $[-\pi/2, \pi/2]$

$$h(s,t) = (\cos(s) + \cos(t))(\sin(u)/u)^2, \quad u = \pi(\sin(s) + \sin(t))$$

$$x(t) = 2 \exp(-6(t - .8)^2) + \exp(-2(t + .5)^2)$$
Observations

- Even when committing the inverse crime we will not achieve the solution if we cannot approximate the basis correctly.
- We need all basis vectors which contain the high frequency terms in order to approximate a solution with high frequency components - e.g. edges.
- Reminder - this is independent of the data.
- But is an indication of an ill-posed problem. In this case the data that is modified exhibits in the matrix $A$ decomposition.
- We look at a problem with a smoother solution - what are the issues?

Problem shaw from the regularization toolbox defined on interval $[-\pi/2, \pi/2]$

$$h(s, t) = (\cos(s) + \cos(t))(\sin(u)/u)^2, \ u = \pi(\sin(s) + \sin(t))$$

$$x(t) = 2 \exp(-6(t - .8)^2) + \exp(-2(t + .5)^2)$$
Observations

- Even when committing the inverse crime we will not achieve the solution if we cannot approximate the basis correctly.
- We need all basis vectors which contain the high frequency terms in order to approximate a solution with high frequency components - e.g. edges.
- Reminder - this is independent of the data.
- But is an indication of an ill-posed problem. In this case the data that is modified exhibits in the matrix $A$ decomposition.
- We look at a problem with a smoother solution - what are the issues?

Problem shaw from the regularization toolbox defined on interval $[-\pi/2, \pi/2]$

$$h(s, t) = (\cos(s) + \cos(t))(\sin(u)/u)^2, \ u = \pi(\sin(s) + \sin(t))$$

$$x(t) = 2 \exp(-6(t - .8)^2) + \exp(-2(t + .5)^2)$$
Observations

▶ Even when committing the inverse crime we will not achieve the solution if we cannot approximate the basis correctly.
▶ We need all basis vectors which contain the high frequency terms in order to approximate a solution with high frequency components - e.g. edges.
▶ Reminder - this is independent of the data.
▶ But is an indication of an ill-posed problem. In this case the data that is modified exhibits in the matrix $A$ decomposition.
▶ We look at a problem with a smoother solution - what are the issues?

Problem shaw from the regularization toolbox defined on interval $[-\pi/2, \pi/2]$

$$h(s, t) = (\cos(s) + \cos(t))(\sin(u)/u)^2, \ u = \pi(\sin(s) + \sin(t))$$

$$x(t) = 2 \exp(-6(t - .8)^2) + \exp(-2(t + .5)^2)$$
Observations

- Even when committing the inverse crime we will not achieve the solution if we cannot approximate the basis correctly.
- We need all basis vectors which contain the high frequency terms in order to approximate a solution with high frequency components - e.g. edges.
- Reminder - this is independent of the data.
- But is an indication of an ill-posed problem. In this case the data that is modified exhibits in the matrix $A$ decomposition.
- We look at a problem with a smoother solution - what are the issues?

Problem shaw from the regularization toolbox defined on interval $[-\pi/2, \pi/2]$

$$h(s, t) = (\cos(s) + \cos(t))(\sin(u)/u)^2, \ u = \pi(\sin(s) + \sin(t))$$

$$x(t) = 2 \exp(-6(t - .8)^2) + \exp(-2(t + .5)^2)$$
Basis vectors for problem shaw

**Figure:** The first few left singular vectors \( \mathbf{v}_i \) (left) and their white noise content on the right.
The Solutions with truncated SVD- problem shaw

Figure: Truncated SVD Solutions: data enters through coefficients $|u_i^T b|$. On the left no noise in $b$ and on the right with noise $10^{-4}$.

In this case the low frequency vectors can represent the solution but we need to know the regularization parameter $k$. 
Observations from the SVD analysis in presence of noise

- Number of terms $k$ in TSVD depends on $v_i$.
- Practically measured data also contaminated by noise $e$.

\[ x = \sum_{i=1}^{n} \left( \frac{u_i^T(b_{\text{exact}} + e)}{\sigma_i} \right) v_i = x_{\text{exact}} + \sum_{i=1}^{n} \left( \frac{u_i^T e}{\sigma_i} \right) v_i \]

- If $e$ is uniform, we expect $|u_i^T e|$ to be similar magnitude $\forall i$.
- When $\sigma_i << |u_i^T e|$ contribution of the high frequency error is magnified.
- Impact of basis vector $v_i$ is magnified.
- The truncated SVD is a special case of spectral filtering

\[ x_{\text{filt}} = \sum_{i=1}^{n} \gamma_i \left( \frac{u_i^T b}{\sigma_i} \right) v_i \]

- Spectral filtering is used to filter the components in the spectral basis, such that noise in signal is damped.
Observations from the SVD analysis in presence of noise

- Number of terms $k$ in TSVD depends on $v_i$.
- Practically measured data also contaminated by noise $e$.

$$
x = \sum_{i=1}^{n} \left( \frac{u_i^T (b_{\text{exact}} + e)}{\sigma_i} \right) v_i = x_{\text{exact}} + \sum_{i=1}^{n} \left( \frac{u_i^T e}{\sigma_i} \right) v_i
$$

- If $e$ is uniform, we expect $|u_i^T e|$ to be similar magnitude $\forall i$.
- When $\sigma_i \ll |u_i^T e|$ contribution of the high frequency error is magnified.
- Impact of basis vector $v_i$ is magnified.
- The truncated SVD is a special case of spectral filtering

$$
x_{\text{filt}} = \sum_{i=1}^{n} \gamma_i \left( \frac{u_i^T b}{\sigma_i} \right) v_i
$$

- Spectral filtering is used to filter the components in the spectral basis, such that noise in signal is damped.
Observations from the SVD analysis in presence of noise

- Number of terms $k$ in TSVD depends on $v_i$.
- Practically measured data also contaminated by noise $e$.

$$
\begin{align*}
\mathbf{x} &= \sum_{i=1}^{n} \left( \frac{\mathbf{u}_i^T \left( \mathbf{b}_{\text{exact}} + e \right)}{\sigma_i} \right) \mathbf{v}_i = \mathbf{x}_{\text{exact}} + \sum_{i=1}^{n} \left( \frac{\mathbf{u}_i^T e}{\sigma_i} \right) \mathbf{v}_i \\
\end{align*}
$$

- If $e$ is uniform, we expect $|\mathbf{u}_i^T e|$ to be similar magnitude $\forall i$.
- When $\sigma_i \ll |\mathbf{u}_i^T e|$ contribution of the high frequency error is magnified.
- Impact of basis vector $\mathbf{v}_i$ is magnified.
- The truncated SVD is a special case of spectral filtering

$$
\mathbf{x}_{\text{filt}} = \sum_{i=1}^{n} \gamma_i \left( \frac{\mathbf{u}_i^T \mathbf{b}}{\sigma_i} \right) \mathbf{v}_i
$$

- Spectral filtering is used to filter the components in the spectral basis, such that noise in signal is damped.
Observations from the SVD analysis in presence of noise

- Number of terms $k$ in TSVD depends on $v_i$.
- Practically measured data also contaminated by noise $e$.

\[
x = \sum_{i=1}^{n} \left( \frac{u_i^T(b_{\text{exact}} + e)}{\sigma_i} \right) v_i = x_{\text{exact}} + \sum_{i=1}^{n} \left( \frac{u_i^T e}{\sigma_i} \right) v_i
\]

- If $e$ is uniform, we expect $|u_i^T e|$ to be similar magnitude $\forall i$.
- When $\sigma_i \ll |u_i^T e|$ contribution of the high frequency error is magnified.
- Impact of basis vector $v_i$ is magnified.
- The truncated SVD is a special case of spectral filtering

\[
x_{\text{filt}} = \sum_{i=1}^{n} \gamma_i \left( \frac{u_i^T b}{\sigma_i} \right) v_i
\]

- Spectral filtering is used to filter the components in the spectral basis, such that noise in signal is damped.
Observations from the SVD analysis in presence of noise

- Number of terms $k$ in TSVD depends on $v_i$.
- Practically measured data also contaminated by noise $e$.

$$x = \sum_{i=1}^{n} \left( \frac{u_i^T(b_{\text{exact}} + e)}{\sigma_i} \right) v_i = x_{\text{exact}} + \sum_{i=1}^{n} \left( \frac{u_i^T e}{\sigma_i} \right) v_i$$

- If $e$ is uniform, we expect $|u_i^T e|$ to be similar magnitude $\forall i$.
- When $\sigma_i \ll |u_i^T e|$ contribution of the high frequency error is magnified.
- Impact of basis vector $v_i$ is magnified.
- The truncated SVD is a special case of spectral filtering

$$x_{\text{filt}} = \sum_{i=1}^{n} \gamma_i \left( \frac{u_i^T b}{\sigma_i} \right) v_i$$

- Spectral filtering is used to filter the components in the spectral basis, such that noise in signal is damped.
Observations from the SVD analysis in presence of noise

- Number of terms $k$ in TSVD depends on $v_i$.
- Practically measured data also contaminated by noise $e$.

$$
\mathbf{x} = \sum_{i=1}^{n} \left( \frac{\mathbf{u}_i^T (\mathbf{b}_{\text{exact}} + \mathbf{e})}{\sigma_i} \right) \mathbf{v}_i = \mathbf{x}_{\text{exact}} + \sum_{i=1}^{n} \left( \frac{\mathbf{u}_i^T \mathbf{e}}{\sigma_i} \right) \mathbf{v}_i
$$

- If $e$ is uniform, we expect $|\mathbf{u}_i^T \mathbf{e}|$ to be similar magnitude $\forall i$.
- When $\sigma_i << |\mathbf{u}_i^T \mathbf{e}|$ contribution of the high frequency error is magnified.
- Impact of basis vector $\mathbf{v}_i$ is magnified.
- The truncated SVD is a special case of spectral filtering

$$
\mathbf{x}_{\text{filt}} = \sum_{i=1}^{n} \gamma_i \left( \frac{\mathbf{u}_i^T \mathbf{b}}{\sigma_i} \right) \mathbf{v}_i
$$

- Spectral filtering is used to filter the components in the spectral basis, such that noise in signal is damped.
Observations from the SVD analysis in presence of noise

- Number of terms $k$ in TSVD depends on $v_i$.
- Practically measured data also contaminated by noise $e$.

$$
x = \sum_{i=1}^{n} \left( \frac{u_i^T(b_{\text{exact}} + e)}{\sigma_i} \right) v_i = x_{\text{exact}} + \sum_{i=1}^{n} \left( \frac{u_i^T e}{\sigma_i} \right) v_i
$$

- If $e$ is uniform, we expect $|u_i^T e|$ to be similar magnitude $\forall i$.
- When $\sigma_i \ll |u_i^T e|$ contribution of the high frequency error is magnified.
- Impact of basis vector $v_i$ is magnified.
- The truncated SVD is a special case of spectral filtering

$$
x_{\text{filt}} = \sum_{i=1}^{n} \gamma_i \left( \frac{u_i^T b}{\sigma_i} \right) v_i
$$

- Spectral filtering is used to filter the components in the spectral basis, such that noise in signal is damped.
Regularization by Spectral Filtering: This is Tikhonov regularization

\[ x_{\text{Tik}} = \sum_{i=1}^{n} \gamma_i \left( \frac{u_i^T b}{\sigma_i} \right) v_i \]

- Tikhonov Regularization \( \gamma_i = \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2} \), \( i = 1 \ldots n \), \( \lambda \) is the regularization parameter, and solution is

\[ x_{\text{Tik}}(\lambda) = \arg \min_x \{ \| b - Ax \|^2 + \lambda^2 \| x \|^2 \} \]

- Choice of \( \lambda^2 \) impacts the solution.
1-D Interesting but Noisy Signal

Blur with Gaussian and add noise - can we find the solution?
Solution for Increasing $\lambda$

Regularized Solution $\lambda = 0.001$

Solutions $x(\lambda)$
Solution for Increasing $\lambda$

Regularized Solution $\lambda = 0.0021544$

Solutions $x(\lambda)$
Solution for Increasing $\lambda$

Regularized Solution $\lambda = 0.0046416$

Solutions $x(\lambda)$
Solution for Increasing $\lambda$

Regularized Solution $\lambda = 0.01$

Solutions $x(\lambda)$
Solution for Increasing $\lambda$

Regularized Solution $\lambda = 0.021544$

Solutions $x(\lambda)$
Solution for Increasing $\lambda$

Regularized Solution $\lambda = 0.046416$

Solutions $x(\lambda)$
Solution for Increasing $\lambda$

Regularized Solution $\lambda = 0.1$

Solutions $x(\lambda)$
Solution for Increasing $\lambda$

Regularized Solution $\lambda = 0.21544$

Solutions $x(\lambda)$
Solution for Increasing $\lambda$

Regularized Solution $\lambda = 0.46416$

Solutions $x(\lambda)$
Solution for Increasing $\lambda$

Regularized Solution $\lambda = 1$

Solutions $x(\lambda)$
Solution for Increasing $\lambda$

Regularized Solution $\lambda = 2.1544$

Solutions $x(\lambda)$
Solution for Increasing $\lambda$

Regularized Solution $\hat{\lambda} = 4.6416$

Solutions $x(\lambda)$
Solution for Increasing $\lambda$
Notice gradients in the solution are smoothed

Consider the more general weighting $\|Lx\|^2$

$$x(\lambda) = \arg \min_x \{ \|Ax - b\|^2 + \lambda^2\|Lx\|^2 \}$$

Typical $L$ approximates the first or second order derivative

$$L_1 = \begin{pmatrix} -1 & 1 & \cdots & \cdots & -1 & 1 \\ -1 & 1 & & & & \end{pmatrix}, \quad L_2 = \begin{pmatrix} 1 & -2 & 1 & & & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ & & & & & 1 & -2 & 1 \end{pmatrix}$$

$L_1 \in \mathbb{R}^{(n-1) \times n}$ and $L_2 \in \mathbb{R}^{(n-2) \times n}$. Note that neither $L_1$ nor $L_2$ are invertible.
Notice gradients in the solution are smoothed
Consider the more general weighting $\|Lx\|^2$

$$x(\lambda) = \arg\min_x \{\|Ax - b\|^2 + \lambda^2\|Lx\|^2\}$$

Typical $L$ approximates the first or second order derivative

$$L_1 = \begin{pmatrix} -1 & 1 & \cdots & \cdots & -1 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ -1 & 1 \end{pmatrix} \quad L_2 = \begin{pmatrix} 1 & -2 & 1 \\ \cdots & \ddots & \ddots & \cdots \\ 1 & -2 & 1 \end{pmatrix}$$

$L_1 \in \mathbb{R}^{(n-1) \times n}$ and $L_2 \in \mathbb{R}^{(n-2) \times n}$. Note that neither $L_1$ nor $L_2$ are invertible.
Extending the Regularization - Change the basis

Notice gradients in the solution are smoothed
Consider the more general weighting $\|Lx\|^2$

$$x(\lambda) = \arg \min_x \{\|Ax - b\|^2 + \lambda^2 \|Lx\|^2\}$$

Typical $L$ approximates the first or second order derivative

$$L_1 = \begin{pmatrix} -1 & 1 \\ \\ \\ \vdots & \ddots & \ddots \\ \\ -1 & 1 \end{pmatrix} \quad L_2 = \begin{pmatrix} 1 & -2 & 1 \\ \vdots & \ddots & \ddots & \ddots \\ 1 & -2 & 1 \end{pmatrix}$$

$L_1 \in \mathbb{R}^{(n-1)\times n}$ and $L_2 \in \mathbb{R}^{(n-2)\times n}$. Note that neither $L_1$ nor $L_2$ are invertible.
Notice gradients in the solution are smoothed
Consider the more general weighting $\|Lx\|^2$

$$x(\lambda) = \arg \min_x \{ \|Ax - b\|^2 + \lambda^2 \|Lx\|^2 \}$$

Typical $L$ approximates the first or second order derivative

$$L_1 = \begin{pmatrix} -1 & 1 \\ \vdots & \vdots \\ -1 & 1 \end{pmatrix} \quad L_2 = \begin{pmatrix} 1 & -2 & 1 \\ \vdots & \vdots & \vdots \\ 1 & -2 & 1 \end{pmatrix}$$

$L_1 \in \mathbb{R}^{(n-1)\times n}$ and $L_2 \in \mathbb{R}^{(n-2)\times n}$. Note that neither $L_1$ nor $L_2$ are invertible.
The Generalized Singular Value Decomposition

Introduce generalization of the SVD to obtain a expansion for

\[
x(\lambda) = \arg \min_x \{ \| Ax - b \|^2 + \lambda^2 \| L(x - x_0) \|^2 \}
\]

**Lemma (GSVD)**

Assume invertibility and \( m \geq n \geq p \). There exist unitary matrices \( U \in \mathbb{R}^{m \times m} \), \( V \in \mathbb{R}^{p \times p} \), and a nonsingular matrix \( Z \in \mathbb{R}^{n \times n} \) such that

\[
A = U \begin{bmatrix}
\Upsilon \\
0_{(m-n) \times n}
\end{bmatrix} Z^T, \quad L = V[M, 0_{p \times (n-p)}] Z^T,
\]

\[
\Upsilon = \text{diag}(\nu_1, \ldots, \nu_p, 1, \ldots, 1) \in \mathbb{R}^{n \times n}, \quad M = \text{diag}(\mu_1, \ldots, \mu_p) \in \mathbb{R}^{p \times p},
\]

with

\[
0 \leq \nu_1 \leq \cdots \leq \nu_p \leq 1, \quad 1 \geq \mu_1 \geq \cdots \geq \mu_p > 0, \quad \nu_i^2 + \mu_i^2 = 1, \quad i = 1, \ldots, p.
\]

Use \( \tilde{\Upsilon} \) and \( \tilde{M} \) to denote the rectangular matrices containing \( \Upsilon \) and \( M \).
Solution of the Generalized Problem using the GSVD

We can use the GSVD to write down the solution for the generalized problem:

\[
x(\lambda) = \sum_{i=1}^{p} \frac{\nu_i}{\nu_i^2 + \lambda^2 \mu_i^2} (u_i^T b) \tilde{z}_i + \sum_{i=p+1}^{n} (u_i^T b) \tilde{z}_i
\]

where \( \tilde{z}_i \) is the \( i^{th} \) column of \( (Z^T)^{-1} \).

With generalized singular value \( \rho_i = \nu_i / \mu_i, i = 1, \ldots, p \)

\[
x(\lambda) = \sum_{i=1}^{p} \gamma_i \frac{u_i^T b}{\nu_i} \tilde{z}_i + \sum_{i=p+1}^{n} (u_i^T b) \tilde{z}_i,
\]

\[
Lx(\lambda) = \sum_{i=1}^{p} \gamma_i \frac{u_i^T b}{\rho_i} v_i, \quad \gamma_i = \frac{\rho_i^2}{\rho_i^2 + \lambda^2},
\]

Notice the similarity with the filtered SVD solution

\[
x(\lambda) = \sum_{i=1}^{n} \gamma_i \frac{u_i^T b}{\sigma_i} v_i, \quad \gamma_i = \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2}.
\]
Solution of the Generalized Problem using the GSVD

We can use the GSVD to write down the solution for the generalized problem:

\[
x(\lambda) = \sum_{i=1}^{p} \frac{\nu_i}{\nu_i^2 + \lambda^2 \mu_i^2} (u_i^T b) \tilde{z}_i + \sum_{i=p+1}^{n} (u_i^T b) \tilde{z}_i
\]

where \( \tilde{z}_i \) is the \( i \)th column of \((Z^T)^{-1}\).

With generalized singular value \( \rho_i = \nu_i/\mu_i \), \( i = 1, \ldots, p \)

\[
x(\lambda) = \sum_{i=1}^{p} \gamma_i \frac{u_i^T b}{\nu_i} \tilde{z}_i + \sum_{i=p+1}^{n} (u_i^T b) \tilde{z}_i,
\]

\[
Lx(\lambda) = \sum_{i=1}^{p} \gamma_i \frac{u_i^T b}{\rho_i} v_i, \quad \gamma_i = \frac{\rho_i^2}{\rho_i^2 + \lambda^2},
\]

Notice the similarity with the filtered SVD solution

\[
x(\lambda) = \sum_{i=1}^{n} \gamma_i \frac{u_i^T b}{\sigma_i} v_i, \quad \gamma_i = \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2}.
\]
Solution of the Generalized Problem using the GSVD

We can use the GSVD to write down the solution for the generalized problem:

\[
x(\lambda) = \sum_{i=1}^{p} \frac{\nu_i}{\nu_i^2 + \lambda^2 \mu_i^2} (u_i^T b) \tilde{z}_i + \sum_{i=p+1}^{n} (u_i^T b) \tilde{z}_i
\]

where \( \tilde{z}_i \) is the \( i \)th column of \((Z^T)^{-1}\).

With generalized singular value \( \rho_i = \nu_i/\mu_i \), \( i = 1, \ldots, p \)

\[
x(\lambda) = \sum_{i=1}^{p} \gamma_i \frac{u_i^T b}{\nu_i} \tilde{z}_i + \sum_{i=p+1}^{n} (u_i^T b) \tilde{z}_i,
\]

\[
Lx(\lambda) = \sum_{i=1}^{p} \gamma_i \frac{u_i^T b}{\rho_i} v_i, \quad \gamma_i = \frac{\rho_i^2}{\rho_i^2 + \lambda^2},
\]

Notice the similarity with the filtered SVD solution

\[
x(\lambda) = \sum_{i=1}^{n} \gamma_i \frac{u_i^T b}{\sigma_i} v_i, \quad \gamma_i = \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2}.
\]
Figure: The basis vectors $\tilde{z}_i$ for $x$ in reverse order - Observe frequency increases in reverse order; not orthogonal.
Figure: Basis vectors $v_i$ in reverse order for $Lx$. Observe frequency increases in reverse order, vectors are not noise contaminated. $L$ acts as a smoother for the basis.
Picard condition for the GSVD: for $x$ and $Lx$ examine the weights in the expansion

**Figure:** Weights for the expansion - $\lambda = .0001$ - blow up together
Picard condition for the GSVD: for $x$ and $Lx$ examine the weights in the expansion.

*Figure:* Weights for the expansion - $\lambda = 0.05$ separate for low frequency.
Picard condition for the GSVD: for $x$ and $Lx$ examine the weights in the expansion

**Figure:** Weights for the expansion - $\lambda = 5$

Notice that for $L_1$, $\nu_i$ are small except for large $i$, i.e. $\mu_i \approx 1$
Solutions $x$ and the derivative $Lx$ for changing $\lambda$

**Figure:** We cannot capture $Lx$ (red) from the solution (green): Notice that $\|Lx\|$ decreases as $\lambda$ increases
Solutions $x$ and the derivative $Lx$ for changing $\lambda$

**Figure:** We cannot capture $Lx$ (red) from the solution (green): Notice that $\|Lx\|$ decreases as $\lambda$ increases.
Solutions $x$ and the derivative $Lx$ for changing $\lambda$

Figure: We cannot capture $Lx$ (red) from the solution (green): Notice that $\|Lx\|$ decreases as $\lambda$ increases
Solutions $\mathbf{x}$ and the derivative $L\mathbf{x}$ for changing $\lambda$

**Figure:** We cannot capture $L\mathbf{x}$ (red) from the solution (green): Notice that $\|L\mathbf{x}\|$ decreases as $\lambda$ increases
Without noise the basis vectors are still subject to noise and contaminate the solution.

TSVD cannot capture features of solutions.

Solutions require finding $k$.

Using the GSVD we see that Tikhonov regularization yields a basis with smoothed basis vectors.

We cannot represent high frequency information in the solutions.

We cannot represent edges or steep gradients.

Regularizing with respect to $Lx$ does not give any realistic estimate of $Lx$!
Without noise the basis vectors are still subject to noise and contaminate the solution.

TSVD cannot capture features of solutions

Solutions require finding $k$.

Using the GSVD we see that Tikhonov regularization yields a basis with smoothed basis vectors

We cannot represent high frequency information in the solutions.

We cannot represent edges or steep gradients

Regularizing with respect to $Lx$ does not give any realistic estimate of $Lx$!
Without noise the basis vectors are still subject to noise and contaminate the solution.

TSVD cannot capture features of solutions

Solutions require finding $k$.

Using the GSVD we see that Tikhonov regularization yields a basis with smoothed basis vectors.

We cannot represent high frequency information in the solutions.

We cannot represent edges or steep gradients.

Regularizing with respect to $Lx$ does not give any realistic estimate of $Lx$!
Without noise the basis vectors are still subject to noise and contaminate the solution.

TSVD cannot capture features of solutions

Solutions require finding $k$.

Using the GSVD we see that Tikhonov regularization yields a basis with smoothed basis vectors

We cannot represent high frequency information in the solutions.

We cannot represent edges or steep gradients

Regularizing with respect to $Lx$ does not give any realistic estimate of $Lx$!
Without noise the basis vectors are still subject to noise and contaminate the solution.

TSVD cannot capture features of solutions

Solutions require finding $k$.

Using the GSVD we see that Tikhonov regularization yields a basis with smoothed basis vectors

We cannot represent high frequency information in the solutions.

We cannot represent edges or steep gradients

Regularizing with respect to $L_x$ does not give any realistic estimate of $L_x$!
Without noise the basis vectors are still subject to noise and contaminate the solution.

TSVD cannot capture features of solutions

Solutions require finding $k$.

Using the GSVD we see that Tikhonov regularization yields a basis with smoothed basis vectors

We cannot represent high frequency information in the solutions.

We cannot represent edges or steep gradients

Regularizing with respect to $Lx$ does not give any realistic estimate of $Lx$!
Without noise the basis vectors are still subject to noise and contaminate the solution.

TSVD cannot capture features of solutions.

Solutions require finding $k$.

Using the GSVD we see that Tikhonov regularization yields a basis with smoothed basis vectors.

We cannot represent high frequency information in the solutions.

We cannot represent edges or steep gradients.

Regularizing with respect to $Lx$ does not give any realistic estimate of $Lx$!
Another approach: Total Variation Regularization

A more general regularization term $R(x)$ may be considered to better preserve properties of the solution $x$:

$$x(\lambda) = \arg \min \{ \|Ax - b\|_W^2 + \lambda^2 R(x) \}$$

Suppose $R$ is total variation of $x$ (general options are possible). The total variation for a function $x$ defined on a discrete grid is

$$TV(x(s)) = \sum_i |x(s_i) - x(s_{i-1})| \approx \Delta \sum_i |dx(s_i)/ds|$$

$TV$ approximates a scaled sum of the magnitude of jumps in $x$. $\Delta$ is a scale factor dependent on the grid size. Notice $TV(x(s)) \approx \Delta \|Lx\|_1$ so we solve

$$x(\lambda) = \arg \min \{ \|Ax - b\|_2^2 + \lambda^2 \|Lx\|_1 \}$$
Another approach: Total Variation Regularization

A more general regularization term $R(x)$ may be considered to better preserve properties of the solution $x$:

$$
x(\lambda) = \arg \min_x \{ \|Ax - b\|_W^2 + \lambda^2 R(x) \}
$$

Suppose $R$ is total variation of $x$ (general options are possible). The total variation for a function $x$ defined on a discrete grid is

$$
TV(x(s)) = \sum_i |x(s_i) - x(s_{i-1})| \approx \Delta \sum_i |dx(s_i)/ds|
$$

$TV$ approximates a scaled sum of the magnitude of jumps in $x$. $\Delta$ is a scale factor dependent on the grid size. Notice $TV(x(s)) \approx \Delta \|Lx\|_1$ so we solve

$$
x(\lambda) = \arg \min_x \{ \|Ax - b\|_2^2 + \lambda^2 \|Lx\|_1 \}$$
Another approach: Total Variation Regularization

A more general regularization term $R(x)$ may be considered to better preserve properties of the solution $x$:

$$x(\lambda) = \arg \min_x \{ \|Ax - b\|_W^2 + \lambda^2 R(x) \}$$

Suppose $R$ is total variation of $x$ (general options are possible). The total variation for a function $x$ defined on a discrete grid is

$$TV(x(s)) = \sum_i |x(s_i) - x(s_{i-1})| \approx \Delta \sum_i |dx(s_i)/ds|$$

$TV$ approximates a scaled sum of the magnitude of jumps in $x$. $\Delta$ is a scale factor dependent on the grid size. Notice $TV(x(s)) \approx \Delta \|Lx\|_1$ so we solve

$$x(\lambda) = \arg \min_x \{ \|Ax - b\|_2^2 + \lambda^2 \|Lx\|_1 \}$$
Newton’s Method  Use $|x| \approx \sqrt{|x|^2 + \beta^2} = \psi(|x|), 1 \gg \beta > 0$.

$$TV(x) \approx \Delta \sum \sqrt{|\nabla x|^2 + \beta^2} = \sqrt{||\nabla x||^2 + \beta^2}$$
$$\approx \frac{1}{2} \sum_{i=2}^{n} \psi(|l_i^T x|^2), \quad l_i \text{ row of } L.$$

Requires update of gradient and Hessian of the functional each step - costly for large $n$.

$$\nabla R(x) = L^T \text{diag}(\psi'(x)) L x = \Psi(x)x, \quad \Psi(x) = L^T \text{diag}(\psi'(x)) L$$
$$\nabla^2 R(x) = \Psi(x) + \Psi'(x)x, \quad \Psi'(x)x = L^T \text{diag}(2(Lx)^2 \psi''(x)) L.$$
Newton's Method  Use $|x| \approx \sqrt{|x|^2 + \beta^2} = \psi(|x|)$, $1 \gg \beta > 0$.

$$TV(x) \approx \Delta \sum \sqrt{|\nabla x|^2 + \beta^2} = \sqrt{||\nabla x||^2 + \beta^2}$$

$$\approx \frac{1}{2} \sum_{i=2}^{n} \psi(|l_i^T x|^2), \quad l_i \text{ row of } L.$$  

Requires update of gradient and Hessian of the functional each step - costly for large $n$.

$$\nabla R(x) = L^T \text{diag}(\psi'(x))Lx = \Psi(x)x, \quad \Psi(x) = L^T \text{diag}(\psi'(x))L$$

$$\nabla^2 R(x) = \Psi(x) + \Psi'(x)x, \quad \Psi'(x)x = L^T \text{diag}(2(Lx)^2\psi''(x))L.$$
Iteratively Reweighted Norm (Rodriguez and Wohlberg 2007 and 2009)

**Goal:** replace $l_1$ norm by quadratic $l_2$ and iterate.

Consider approximations for $Lx \approx d$.

\[
R(x) = \| \sqrt{\|\nabla x\|^2} \|_q^q = \sum_l (\|d_l\|_2)^q \quad 0 < q \leq 2
\]

Let $W$ be a diagonal matrix with entries $W_{ll} = (\|d_l\|_2)^2(q-2)$, $R(x) \approx \|W^{1/2}Lx\|^2_2$,

and update each iteration

\[
R^{(k)}(x) = \|Lx\|_{W^{(k)}}^2
\]

To avoid singularity again introduce scaling parameter:

\[
W_{ll}^k = \tau(\|d_l^k\|_2^2)(\|d_l^k\|_2^2), \quad \tau(z) = 1, \quad z \geq \epsilon \quad \tau(z) = 0, \quad z < \epsilon.
\]
Iteratively Reweighted Norm (Rodriguez and Wohlberg 2007 and 2009)

**Goal:** replace $l_1$ norm by quadratic $l_2$ and iterate.

Consider approximations for $Lx \approx d$.

$$ R(x) = \| \sqrt{\| \nabla x \|^2} \|^q = \sum_l (\| d_l \|_2^2)^q \quad 0 < q \leq 2 $$

Let $W$ be a diagonal matrix with entries $W_{ll} = (\| d_l \|_2^2)^{(q-2)}$, $W(x) \approx \| W^{1/2} Lx \|_2^2$,

and update each iteration

$$ R^{(k)}(x) = \| Lx \|_{W^{(k)}}^2 $$

To avoid singularity again introduce scaling parameter:

$$ W_{ll}^k = \tau(\| d_l^k \|_2^2)(\| d_l^k \|_2^2), \quad \tau(z) = 1, \quad z \geq \epsilon \quad \tau(z) = 0, \quad z < \epsilon. $$
Iteratively Reweighted Norm (Rodriguez and Wohlberg 2007 and 2009)

**Goal:** replace $l_1$ norm by quadratic $l_2$ and iterate.

Consider approximations for $Lx \approx d$.

$$R(x) = \| \sqrt{\| \nabla x \|^2} \|_q^q = \sum_l (\| d_l \|_2)^q \quad 0 < q \leq 2$$

Let $W$ be a diagonal matrix with entries

$$W_{ll} = (\| d_l \|_2^2)^{(q-2)}$$

and update each iteration

$$R^{(k)}(x) = \| Lx \|_{W^{(k)}}^2$$

To avoid singularity again introduce scaling parameter:

$$W^{(k)}_{ll} = \tau(\| d_l^{(k)} \|_2^2)(\| d_l^{(k)} \|_2^2), \quad \tau(z) = 1, \quad z \geq \epsilon \quad \tau(z) = 0, \quad z < \epsilon.$$
Iteratively Reweighted Norm (Rodriguez and Wohlberg 2007 and 2009)

Goal: replace $l_1$ norm by quadratic $l_2$ and iterate.

Consider approximations for $Lx \approx d$.

$$R(x) = \| \sqrt{\| \nabla x \|_2^2} \|_q^q = \sum_l (\|d_l\|_2^2)^q \quad 0 < q \leq 2$$

Let $W$ be a diagonal matrix with entries $W_{ll} = (\|d_l\|_2^2)^{(q-2)}$, 

$$R(x) \approx \|W^{1/2} Lx\|_2^2,$$

and update each iteration 

$$R^{(k)}(x) = \|Lx\|_{W^{(k)}}^2$$

To avoid singularity again introduce scaling parameter:

$$W_{ll}^k = \tau(\|d_l^k\|_2^2)(\|d_l^k\|_2^2), \quad \tau(z) = 1, \quad z \geq \epsilon \quad \tau(z) = 0, \quad z < \epsilon.$$
The main reference is here with software:


Many developments have been made since that time:

Above reference discusses relationship of SB to Augmented Lagrangian and Peaceman-Rachford alternating direction
For $R(x) = Lx$ for $L \in \mathcal{R}^{q \times n}$: Introduce $d = Lx$

Rewrite $R(x) = \frac{\lambda^2}{2} \| d - Lx \|_2^2 + \mu \| d \|_1$ and restate as

$$(x, d) = \arg \min_{x, d} \left\{ \frac{1}{2} \| Ax - b \|_2^2 + \frac{\lambda^2}{2} \| d - Lx \|_2^2 + \mu \| d \|_1 \right\}$$

Solve using an alternating minimization which separates minimization for $d$ from $x$

Various versions of the iteration can be defined.

$$S1 : x^{(k+1)} = \arg \min_x \left\{ \frac{1}{2} \| Ax - b \|_2^2 + \frac{\lambda^2}{2} \| Lx - (d^{(k+1)} - g^{(k)}) \|_2^2 \right\}$$  \hspace{1cm} (1)

$$S2 : d^{(k+1)} = \arg \min_d \left\{ \frac{\lambda^2}{2} \| d - (Lx^{(k+1)} + g^{(k)}) \|_2^2 + \mu \| d \|_1 \right\}$$  \hspace{1cm} (2)

$$S3 : g^{(k+1)} = g^{(k)} + Lx^{(k+1)} - d^{(k+1)}.$$  \hspace{1cm} (3)
For $R(x) = Lx$ for $L \in \mathcal{R}^{q \times n}$: Introduce $d = Lx$

Rewrite $R(x) = \frac{\lambda^2}{2} \|d - Lx\|_2^2 + \mu \|d\|_1$ and restate as

$$(x, d) = \arg \min_{x, d} \left\{ \frac{1}{2} \|Ax - b\|_2^2 + \frac{\lambda^2}{2} \|d - Lx\|_2^2 + \mu \|d\|_1 \right\}$$

Solve using an alternating minimization which separates minimization for $d$ from $x$

Various versions of the iteration can be defined.

S1: $x^{(k+1)} = \arg \min_x \left\{ \frac{1}{2} \|Ax - b\|_2^2 + \frac{\lambda^2}{2} \|Lx - (d^{(k+1)} - g^{(k)})\|_2^2 \right\}$ \hspace{1cm} (1)

S2: $d^{(k+1)} = \arg \min_d \left\{ \frac{\lambda^2}{2} \|d - (Lx^{(k+1)} + g^{(k)})\|_2^2 + \mu \|d\|_1 \right\}$ \hspace{1cm} (2)

S3: $g^{(k+1)} = g^{(k)} + Lx^{(k+1)} - d^{(k+1)}$. \hspace{1cm} (3)
SB The Main Idea of GO Paper: for Regularization $R(x)$

For $R(x) = Lx$ for $L \in \mathbb{R}^{q \times n}$: Introduce $d = Lx$

Rewrite $R(x) = \frac{\lambda^2}{2} \|d - Lx\|_2^2 + \mu \|d\|_1$ and restate as

$$(x, d) = \underset{x,d}{\arg\min}\left\{\frac{1}{2}\|Ax - b\|_2^2 + \frac{\lambda^2}{2}\|d - Lx\|_2^2 + \mu \|d\|_1\right\}$$

Solve using an alternating minimization which separates minimization for $d$ from $x$

Various versions of the iteration can be defined.

S1: $x^{(k+1)} = \underset{x}{\arg\min}\left\{\frac{1}{2}\|Ax - b\|_2^2 + \frac{\lambda^2}{2}\|Lx - (d^{(k+1)} - g^{(k)})\|_2^2\right\}$ \hspace{1cm} (1)

S2: $d^{(k+1)} = \underset{d}{\arg\min}\left\{\frac{\lambda^2}{2}\|d - (Lx^{(k+1)} + g^{(k)})\|_2^2 + \mu \|d\|_1\right\}$ \hspace{1cm} (2)

S3: $g^{(k+1)} = g^{(k)} + Lx^{(k+1)} - d^{(k+1)}$. \hspace{1cm} (3)
For $R(x) = Lx$ for $L \in \mathcal{R}^{q \times n}$: Introduce $d = Lx$

Rewrite $R(x) = \frac{\lambda^2}{2} \|d - Lx\|_2^2 + \mu \|d\|_1$ and restate as

$$ (x, d) = \arg \min_{x, d} \{ \frac{1}{2} \|Ax - b\|_2^2 + \frac{\lambda^2}{2} \|d - Lx\|_2^2 + \mu \|d\|_1 \} $$

Solve using an alternating minimization which separates minimization for $d$ from $x$

Various versions of the iteration can be defined.

S1: $x^{(k+1)} = \arg \min_x \{ \frac{1}{2} \|Ax - b\|_2^2 + \frac{\lambda^2}{2} \|Lx - (d^{(k+1)} - g^{(k)})\|_2^2 \}$  

S2: $d^{(k+1)} = \arg \min_d \{ \frac{\lambda^2}{2} \|d - (Lx^{(k+1)} + g^{(k)})\|_2^2 + \mu \|d\|_1 \}$  

S3: $g^{(k+1)} = g^{(k)} + Lx^{(k+1)} - d^{(k+1)}$.  


SB Unconstrained algorithm:

**Update for $x$:**

$$ x = \arg \min_x \left\{ \frac{1}{2} \| Ax - b \|_2^2 + \frac{\lambda^2}{2} \| Lx - h \|_2^2 \right\}, \quad h = d - g $$

Standard least squares update using a Tikhonov regularizer. Still uses the same basis but perturbed weights through $h$

$$ x^{(k+1)} = \sum_{i=1}^{p} \frac{\nu_i u_i^T b + \lambda^2 \mu_i v_i^T h^{(k)}}{\nu_i^2 + \lambda^2 \mu_i^2} \tilde{z}_i + \sum_{i=p+1}^{n} (u_i^T b) \tilde{z}_i $$

**Update for $d$:**

$$ d = \arg \min_d \left\{ \mu \| d \|_1 + \frac{\lambda^2}{2} \| d - c \|_2^2 \right\}, \quad c = Lx + g $$

$$ = \arg \min_d \left\{ \| d \|_1 + \frac{\gamma}{2} \| d - c \|_2^2 \right\}, \quad \gamma = \frac{\lambda^2}{\mu} $$

This is achieved using *soft* thresholding.
SB Unconstrained algorithm:

Update for $x$:

$$ x = \text{arg min}_x \left\{ \frac{1}{2} \|Ax - b\|_2^2 + \frac{\lambda^2}{2} \|Lx - h\|_2^2 \right\}, \quad h = d - g $$

Standard least squares update using a Tikhonov regularizer. Still uses the same basis but perturbed weights through $h$

$$ x^{(k+1)} = \sum_{i=1}^{p} \frac{\nu_i u_i^T b + \lambda^2 \mu_i v_i^T h^{(k)}}{\nu_i^2 + \lambda^2 \mu_i^2} \tilde{z}_i + \sum_{i=p+1}^{n} (u_i^T b) \tilde{z}_i $$

Update for $d$:

$$ d = \text{arg min}_d \left\{ \mu \|d\|_1 + \frac{\lambda^2}{2} \|d - c\|_2^2 \right\}, \quad c = Lx + g $$

$$ = \text{arg min}_d \left\{ \|d\|_1 + \frac{\gamma}{2} \|d - c\|_2^2 \right\}, \quad \gamma = \frac{\lambda^2}{\mu} $$

This is achieved using soft thresholding.
Thresholding for $d$

If $d$ has $q$ components $(d)_i$ componentwise solution:

$$(d)_i = \frac{c_i}{|c_i|} \max(|c_i| - \frac{1}{\gamma}, 0) \quad i = 1 : q$$

If $d$ is two dimensional it contains components $d_x$ and $d_y$. Threshold is applied for each component of $(d_x^T, d_y^T)^T$: we use

$$\|d\|_1 = \|d_x\|_1 + \|d_y\|_1$$

The TV norm for the two dimensional case can be written

$$\|d\|_{TV} = \left( \sum_{l=1}^{n} \|d_l\|_2 \right) \quad q = 2(n - 1)$$

Intrinsically TV is still local.

$$(d_{TV})_i = \frac{c_i}{\|c_i\|_2} \max(\|c_i\|_2 - \frac{1}{\gamma}, 0) \quad i = 1 : n$$
Illustrating Total Variation Solutions for noise level .1 in the data

Figure: $\lambda = .1, \gamma = .5$: We see that final $h$ is too large relative to $b$
Illustrating Total Variation Solutions for noise level 0.1 in the data

Figure: $\lambda = 1, \gamma = 0.5$: Final $h$ balances $b$
Illustrating Total Variation Solutions for noise level .1 in the data

Figure: $\lambda = 10, \gamma = .5$: $b$ dominates and the solution is over smooth
Illustrating Total Variation Solutions for noise level .1 in the data: The impact of $\gamma$

Figure: $\lambda = .1$, On the left $\gamma = .5$ and on the right $\gamma = 5$
Advantage of TV is clear for constant components. But solutions still depend on parameters. We need to find both $\lambda$ and $\mu$. We may use standard parameter estimation to find $\lambda$ - for updating $x$. We can use reweighting for $\mu$ - for updating $d$. 

Observations
Advantage of TV is clear for constant components. But solutions still depend on parameters. We need to find both $\lambda$ and $\mu$. We may use standard parameter estimation to find $\lambda$ - for updating $x$. We can use reweighting for $\mu$ - for updating $d$. 
Advantage of TV is clear for constant components. But solutions still depend on parameters. We need to find both $\lambda$ and $\mu$. We may use standard parameter estimation to find $\lambda$ - for updating $x$. We can use reweighting for $\mu$ - for updating $d$. 
Advantage of TV is clear for constant components.
But solutions still depend on parameters
We need to find both $\lambda$ and $\mu$
We may use standard parameter estimation to find $\lambda$ - for updating $x$
We can use reweighting for $\mu$ - for updating $d$
Advantage of TV is clear for constant components.
But solutions still depend on parameters.
We need to find both $\lambda$ and $\mu$.
We may use standard parameter estimation to find $\lambda$ - for updating $x$.
We can use reweighting for $\mu$ - for updating $d$. 
Consideration of the Update for $x$

$Lx$ update - examine update

$$x = \arg \min_x \left\{ \frac{1}{2} \| Ax - b \|_2^2 + \frac{\lambda^2}{2} \| Lx - h \|_2^2 \right\}, \quad h = d - g$$

Standard form  Introduce $\bar{A} = AL^\#$ where $L^\#$ is the oblique pseudoinverse of $L$,

Shift $y = Lx - h$ and solve with $b^{(k+1)} = \bar{b} - \bar{A}h^{(k)}$

$$y^{(k+1)} = \arg \min_y \left\{ \frac{1}{2} \| \bar{A}y - b^{(k+1)} \|_2^2 + \frac{\lambda^2}{2} \| y \|_2^2 \right\}$$

Independence  Updating does not require $x$

Final step $x = L^\#y^{(k+1)} + x_N$, $x_N = W(AW)^\dagger b$,

$\bar{b} = b - Ax_N$, $W$ spans the null space of $L$

Parameter Choice  Look at Picard condition for $y$ and not $x$. 
Exploiting the GSVD for analysis

\[ x^{(1)} = \sum_{i=1}^{p} \frac{\phi_i}{\nu_i} u_i^T b z_i + \sum_{i=p+1}^{n} (u_i^T b) z_i, \]

\[ x^{(k+1)} = \sum_{i=1}^{p} \left( \frac{\phi_i}{\nu_i} u_i^T b + \frac{1 - \phi_i}{\mu_i} v_i^T h^{(k)} \right) z_i + \sum_{i=p+1}^{n} (u_i^T b) z_i \]  \hspace{1cm} (4)

\[ L x^{(k+1)} = \sum_{i=1}^{p} \left( \frac{\phi_i \mu_i}{\nu_i} (u_i^T b) + (1 - \phi_i) v_i^T h^{(k)} \right) v_i. \]  \hspace{1cm} (5)

Notice in this case that we must also control the coefficients for terms with h. Relevant coefficients

\[ \frac{\phi_i}{\nu_i} \hspace{1cm} \frac{\phi_i \mu_i}{\nu_i} \hspace{1cm} \frac{(1 - \phi_i)}{\mu_i} \hspace{1cm} (1 - \phi_i) \]
GSVD coefficients for the Problem

Figure: $\lambda = .001$

Of course the coefficients are independent of the data.
GSVD coefficients for the Problem

Figure: $\lambda = .01$

Of course the coefficients are independent of the data
GSVD coefficients for the Problem

Figure: $\lambda = 1$

Of course the coefficients are independent of the data
GSVD coefficients for the Problem

Figure: $\lambda = 1$

Of course the coefficients are independent of the data
GSVD coefficients for the Problem

Of course the coefficients are independent of the data
Picard Condition using GSVD for noise level $\nu$.

$\lambda = 0.001$, $\gamma = 0.5$, noise level $= 0.1$, steps $= 100$

But $h$ changes with the iteration
Picard Condition using GSVD for noise level $0.1$

But $h$ changes with the iteration
Picard Condition using GSVD for noise level $\lambda = 0.1, \gamma = 0.5$ noise level = 0.1 steps = 100

But $h$ changes with the iteration
Picard Condition using GSVD for noise level $0.1$

But $h$ changes with the iteration
For regularization in general the choice of $\lambda$ depends on the right hand side vector.

In this case $h$ changes each step.

It is clear that we should update $\lambda$ each step.

We use standard approach - Unbiased Predictive Risk Estimation and L-Curve.

We also use iteratively reweighed norm approach for the $d$ update.
Figure: $\gamma = 200$ Low noise. Without updating $\lambda$ left and updated right. SB UPRE uses the estimated $\lambda$ from UPRE for all SB steps. Update SB, updates $\lambda$ each step. SB IRN updates and iteratively reweights $\|d\|_1$. 
Figure: $\gamma = 200$ High noise. Without updating $\lambda$ left and updated right. SB UPRE uses the estimated $\lambda$ from UPRE for all SB steps. Update SB, updates $\lambda$ each step. SB IRN updates and iteratively reweights $\|d\|_1$. 
Example Solution: 2D - similar blurring operator

Figure: $\gamma = 5$

SB with updated $\lambda$ is useful
Example Solution: 2D - similar blurring operator

Figure: $\gamma = 5$

SB with updated $\lambda$ is useful
Further Observations and Future Work

**Results** demonstrate basic analysis of problem is worthwhile

**Parameter estimation** from basic LS can be used to find appropriate parameter

**Questions** that may be raised - cost of finding optimal $\lambda$
  
  ★ Overhead of optimal $\lambda$ for the first step - reasonable
  
  ★ Overhead of subsequent steps - UPRE requires matrix trace - but for deblurring we can use results about spectrum of Toeplitz matrices

**Future** Implement using the Toeplitz operators

**Extensions** Implement using statistical estimation using $\chi^2$ approach. Takes account of covariance on $h$

**Convergence** testing is based on $h$. 
See me for extensive references to literature

THANK YOU