SOLUTION OF ILL-POSED INVERSE PROBLEMS PERTAINING TO SIGNAL RESTORATION: UNDERSTANDING NOISE IN THE BASIS

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

EDINBURGH
APRIL 12, 2012
Outline

Background
Examples of Ill-posed problems
Motivation: Total Variation Solutions

The SVD Solution
Importance of the Basis
Picard Condition for Ill-Posed Problems

Generalized regularization
GSVD for examining the solution
Revealing the Noise in the GSVD Basis
Stabilizing the GSVD Solution

Conclusions and Future

Regularization Parameter Estimation for TV
Illustration: Blurred Signal Restoration

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

\[ h(s) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{s^2}{2\sigma^2}\right), \quad \sigma > 0. \]
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
  \[ b = Ax \]
  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.
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
  \[ b = Ax \]
  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.
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
  \[ b = Ax \]
  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.
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
  \[ b = Ax \]
  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.
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
  $$b = Ax$$
  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.
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
  \[ b = Ax \]
  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 $0.0001$ : $x = A^{-1}b$

(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 $x$.

**Required**  Analyse the formulation / adapt solution techniques

**Tikhonov Regularization**

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

Required
Analyse the formulation / adapt solution techniques

Tikhonov Regularization

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

Required  Analyse the formulation / adapt solution techniques

Tikhonov Regularization

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

Required
Analyse the formulation / adapt solution techniques

Tikhonov Regularization

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

Required  Analyse the formulation / adapt solution techniques

Tikhonov Regularization

\[
x_{\text{Tik}}(\lambda) = \arg \min_x \{ \| b - Ax \|^2 + \lambda^2 \| Lx \|^2 \}
\]
Tikhonov Regularized Solutions $x(\lambda)$ and derivative $Lx$ for changing $\lambda$

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

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

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

\[
x(\lambda) = \arg \min_x \{ \|Ax - b\|_2^2 + \lambda^2 \|Lx\|_1 \}
\]

Problematic for large scale problems.
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

$$x(\lambda) = \arg \min_x \{ \|Ax - b\|_2^2 + \lambda^2 \|Lx\|_1 \}$$

Problematic for large scale problems.
Total Variation Regularization

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

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

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

\[
x(\lambda) = \arg\min_x \{ \| Ax - b \|^2_2 + \lambda^2 \| Lx \|_1 \}
\]

Problematic for large scale problems.
The main reference is here with software:
T. GOLDSSTEIN, Split Bregman Software Page, 


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 $L \in \mathbb{R}^{q \times n}$: Introduce $d = Lx$

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

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

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

Various versions of the iteration can be defined.

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

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

S3 : $g^{(k+1)} = g^{(k)} + Lx^{(k+1)} - d^{(k+1)}$. (3)
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

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

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

Various versions of the iteration can be defined.

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

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

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

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

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

Various versions of the iteration can be defined.

1. $S1: \quad x^{(k+1)} = \arg \min_x \left\{ \frac{1}{2} \|Ax - b\|_2^2 + \frac{\lambda^2}{2} \|Lx - (d^{(k+1)} - g^{(k)})\|_2^2 \right\}$
2. $S2: \quad d^{(k+1)} = \arg \min_d \left\{ \frac{\lambda^2}{2} \|d - (Lx^{(k+1)} + g^{(k)})\|_2^2 + \mu \|d\|_1 \right\}$
3. $S3: \quad g^{(k+1)} = g^{(k)} + Lx^{(k+1)} - d^{(k+1)}$. 
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.

\begin{align*}
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\} \quad (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\} \quad (2) \\
S3 : g^{(k+1)} &= g^{(k)} + Lx^{(k+1)} - d^{(k+1)}. \quad (3)
\end{align*}
SB Unconstrained algorithm:

Update for \(\mathbf{x}\):

\[
\mathbf{x} = \arg\min_{\mathbf{x}} \left\{ \frac{1}{2} \|A\mathbf{x} - \mathbf{b}\|_2^2 + \frac{\lambda^2}{2} \|L\mathbf{x} - \mathbf{h}\|_2^2 \right\}, \quad \mathbf{h} = \mathbf{d} - \mathbf{g}
\]

Standard least squares update using a Tikhonov regularizer. Uses the standard basis but perturbed weights through \(\mathbf{h}\)

\[
\mathbf{x}^{(k+1)} = \sum_{i=1}^{p} \frac{\nu_i \mathbf{u}_i^T \mathbf{b} + \lambda^2 \mu_i \mathbf{v}_i^T \mathbf{h}^{(k)}}{\nu_i^2 + \lambda^2 \mu_i^2} \mathbf{z}_i + \sum_{i=p+1}^{n} (\mathbf{u}_i^T \mathbf{b}) \mathbf{z}_i
\]

Update for \(\mathbf{d}\):

\[
\mathbf{d} = \arg\min_{\mathbf{d}} \left\{ \mu \|\mathbf{d}\|_1 + \frac{\lambda^2}{2} \|\mathbf{d} - \mathbf{c}\|_2^2 \right\}, \quad \mathbf{c} = L\mathbf{x} + \mathbf{g}
\]

\[
= \arg\min_{\mathbf{d}} \left\{ \|\mathbf{d}\|_1 + \frac{\gamma}{2} \|\mathbf{d} - \mathbf{c}\|_2^2 \right\}, \quad \gamma = \frac{\lambda^2}{\mu}
\]

To obtain an effective solution we need to find the parameters \(\lambda, \mu\).
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. Uses the standard 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 \(d\):**

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

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

To obtain an effective solution we need to find the parameters \(\lambda, \mu\).
The Situation

1. Inverse problem we need regularization
2. For feature extraction we need more than Tikhonov Regularization - e.g. TV
3. Both techniques are parameter dependent
4. Notice the dependence of the TV solution on Tikhonov solutions
5. The TV iterates over many Tikhonov solutions
6. Moreover the parameters are needed
7. We need to fully understand the Tikhonov and ill-posed problems
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

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

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

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

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

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

gives expansion for the solution

\[ \mathbf{x} = \sum_{i=1}^{n} \frac{\mathbf{u}_i^T\mathbf{b}}{\sigma_i} \mathbf{v}_i \]

\( \mathbf{u}_i, \mathbf{v}_i \) are left and right singular vectors for \( \mathbf{A} \)

Solution is a weighted linear combination of the basis vectors \( \mathbf{v}_i \)
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

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

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 \( v_i \)

$$h(s,t) = se^{-ts^2}, \quad x(s) = 1, \quad \frac{1}{3} < t < \frac{2}{3} \quad b(t) = \frac{e^{-t/9} - e^{-4t/9}}{2t}$$
Left Singular Vectors and Basis Depend on the kernel (matrix $A$)

**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.
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 $y \sim \mathbb{N}(0, 1)$? (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 $c$

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

**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 $c$

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

**Automatic:** Test is the line $(z_j, j/q)$ close to a straight line with slope 1 and length $\sqrt{5}/2$?
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: Low frequency vectors lie above the diagonal and high frequency below the diagonal. White noise follows the diagonal
Cumulative Periodogram for the basis 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 - if you count there are 9 vectors with true frequency content on the left
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 or assess proportion of the vector outside the Kolmogorov Smirnov test at a 5% confidence level for white noise lines. 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 \( b \)
Figure: Testing for white noise for the standard precision vectors: Calculate the cumulative periodogram and measure the deviation from the “white noise” line or assess proportion of the vector outside the Kolmogorov Smirnov test at a 5% confidence level for white noise lines. 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

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

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 $|u_i^T b|$. 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?

Problemshaw 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 $[-\pi/2, \pi/2]$

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

$$x(t) = 2 \exp(-6(t - .8)^2) + \exp(-2(t + .5)^2)$$
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 $[-\pi/2, \pi/2]$

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

$$x(t) = 2 \exp(-6(t - .8)^2) + \exp(-2(t + .5)^2)$$
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 $[-\pi/2, \pi/2]$

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

$x(t) = 2 \exp(-6(t - .8)^2) + \exp(-2(t + .5)^2)$
Figure: The first few left singular vectors $v_i$ (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 $e$.

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

- Note $\|U^T e\| = \|e\|$ and $\|U^T b\| = \|b\|$ is dominated by low frequency terms when $A$ smooths $x$ to give $b$.
- 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 and that of basis vector $v_i$.
- 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 $e$.

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

- Note $\|U^T e\| = \|e\|$ and $\|U^T b\| = \|b\|$ is dominated by low frequency terms when $A$ smooths $x$ to give $b$.
- 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 and that of basis vector $v_i$.
- 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 $e$.

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

- Note $\|U^T e\| = \|e\|$ and $\|U^T b\| = \|b\|$ is dominated by low frequency terms when $A$ smooths $x$ to give $b$.
- 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 and that of basis vector $v_i$.
- 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 $e$.

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

- Note $\|U^T e\| = \|e\|$ and $\|U^T b\| = \|b\|$ is dominated by low frequency terms when $A$ smooths $x$ to give $b$.
- 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 and that of basis vector $v_i$
- 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 $e$.

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

- Note $\|U^T e\| = \|e\|$ and $\|U^T b\| = \|b\|$ is dominated by low frequency terms when $A$ *smooths* $x$ to give $b$.
- 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 and that of basis vector $v_i$
- 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 $e$.

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

- Note $\|U^T e\| = \|e\|$ and $\|U^T b\| = \|b\|$ is dominated by low frequency terms when $A$ smooths $x$ to give $b$.
- 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 and that of basis vector $v_i$
- The truncated SVD is a special case of spectral filtering

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

- 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 $e$.

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

- Note $\|U^T e\| = \|e\|$ and $\|U^T b\| = \|b\|$ is dominated by low frequency terms when $A$ smooths $x$ to give $b$.
- 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 and that of basis vector $v_i$.
- 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 $\hat{\lambda} = 0.46416$

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 $\lambda = 4.6416$

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

Regularized Solution $\lambda = 10$

Solutions $x(\lambda)$
Finding the Optimal Parameter

1. Many options exist
2. Debatable as to best approach
3. Should use information on noise if available
4. In general somewhat costly
5. Usually involves finding several solutions
6. We consider Unbiased Predictive Risk and $\chi^2$ (the latter is very cheap)
7. But it is important to reduce cost
Regularizing the TSVD: But we can just use the truncated expansion

Inaccuracy of the basis $v_i$ suggest that one should not just seek the filtered SVD solution but use the **stabilized TSVD**

**Figure**: Two Example Solutions: Calculating the solution is robust for fewer basis vectors 1% noise
Regularizing the TSVD: But we can just use the truncated expansion

Inaccuracy of the basis $v_i$ suggest that one should not just seek the filtered SVD solution but use the stabilized TSVD

Figure: Two Example Solutions: Calculating the solution is robust for fewer basis vectors 1% noise
Further Observations

1. We can use a limited basis to obtain the solution.
2. We can stabilize the TSVD if needed and parameter choice methods are cheaper.
3. Calculating the basis to use is data independent: depends only on the ill-posedness of the system.
4. Removing the noisy vectors still does not permit all good solutions.
5. Clearly high frequency cannot be represented with low frequency basis.
6. We need to alter the basis.
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

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

\( L_1 \in \mathbb{R}^{(n-1) \times n} \) 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

$L_1 = \begin{pmatrix} -1 & 1 \\ \\ \cdots & \cdots \\ \\ -1 & 1 \end{pmatrix}$ \hspace{1cm} $L_2 = \begin{pmatrix} 1 & -2 & 1 \\ \\ \cdots & \cdots & \cdots \\ \\ 1 & -2 & 1 \end{pmatrix}$

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

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

$L_1 \in \mathbb{R}^{(n-1) \times n}$ 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

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

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

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

\[
\Upsilon = \text{diag}(\upsilon_1, \ldots, \upsilon_p, 1, \ldots, 1) \in \mathbb{R}^{n \times n}, \quad M = \text{diag}(\mu_1, \ldots, \mu_p) \in \mathbb{R}^{p \times p},
\]

with

\[
0 \leq \upsilon_1 \leq \cdots \leq \upsilon_p \leq 1, \quad 1 \geq \mu_1 \geq \cdots \geq \mu_p > 0, \quad \upsilon_i^2 + \mu_i^2 = 1, \quad i = 1, \ldots, p.
\]

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 \( \rho_i = \nu_i / \mu_i \), \( i = 1, \ldots, p \)

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

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

Notice the similarity with the filtered SVD solution

\[
x(\lambda) = \sum_{i=1}^{n} \frac{\gamma_i}{\sigma_i} u_i^T b \nu_i, \quad \gamma_i = \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2}.
\]
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

\[
x(\lambda) = \sum_{i=1}^{n} \gamma_i \frac{u_i^T b}{\sigma_i} v_i, \quad \gamma_i = \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2}.
\]
What matrix to use for the GSVD? Form the truncated matrix $A_k$ using only $k$ basis vectors $v_i$: Above $A$ and below $A_k$. Problem ilaplace

Figure: Contrast GSVD basis $u$ (left) $z$ (right). Trade off $U$ and $Z$
What matrix to use for the GSVD? Form the truncated matrix $A_k$ using only $k$ basis vectors $v_i$.

**Figure:** This is confirmed by examining the KS test for white noise.
Example Solution: problem $ilaplace \ A$ above and $A_k$ below

**Figure:** Solution is robust to using the reduced matrix $A_k$. 
Example Solution: problem ilaplace $A$ above and $A_k$ below

Figure: Solution is robust to truncating the GSVD
Figure: Solution is robust to using the reduced matrix $A_k$.
Figure: Solution is robust to truncating the GSVD
Basis vectors are subject to noise and contaminate the solution independent of the data.

Solutions require finding $k$.

We can determine $k$ by examining noise in the basis.

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

But we can apply the same techniques.

Stabilizing both TSVD and TGSVD is robust to parameter estimation.

Parameter estimation for truncated expansions is cheaper.

Reminder truncation $k$ is independent of the data.
- Basis vectors are subject to noise and contaminate the solution independent of the data.
- Solutions require finding $k$.
- We can determine $k$ by examining noise in the basis.
- Using the GSVD we see that Tikhonov regularization yields a basis with smoothed basis vectors.
- But we can apply the same techniques.
- Stabilizing both TSVD and TGSVD is robust to parameter estimation.
- Parameter estimation for truncated expansions is cheaper.
- Reminder truncation $k$ is independent of the data $b$. 
Basis vectors are subject to noise and contaminate the solution independent of the data.

Solutions require finding $k$.

We can determine $k$ by examining noise in the basis.

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

But we can apply the same techniques:

Stabilizing both TSVD and TGSVD is robust to parameter estimation.

Parameter estimation for truncated expansions is cheaper.

Reminder truncation $k$ is independent of the data $b$. 
Basis vectors are subject to noise and contaminate the solution independent of the data.

Solutions require finding \( k \).

We can determine \( k \) by examining noise in the basis.

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

But we can apply the same techniques.

Stabilizing both TSVD and TGSVD is robust to parameter estimation.

Parameter estimation for truncated expansions is cheaper.

Reminder truncation \( k \) is independent of the data \( b \).
Basis vectors are subject to noise and contaminate the solution independent of the data.

Solutions require finding $k$.

We can determine $k$ by examining noise in the basis.

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

But we can apply the same techniques.

Stabilizing both TSVD and TGSVD is robust to parameter estimation.

Parameter estimation for truncated expansions is cheaper.

Reminder truncation $k$ is independent of the data $b$. 
Basis vectors are subject to noise and contaminate the solution independent of the data.

Solutions require finding $k$.

We can determine $k$ by examining noise in the basis.

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

But we can apply the same techniques.

Stabilizing both TSVD and TGSVD is robust to parameter estimation.

Parameter estimation for truncated expansions is cheaper.

Reminder truncation $k$ is independent of the data $b$. 
Basis vectors are subject to noise and contaminate the solution independent of the data.

Solutions require finding $k$.

We can determine $k$ by examining noise in the basis.

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

But we can apply the same techniques.

Stabilizing both TSVD and TGSVD is robust to parameter estimation.

Parameter estimation for truncated expansions is cheaper.

Reminder truncation $k$ is independent of the data $b$. 
Summary/Conclusions

- Basis vectors are subject to noise and contaminate the solution independent of the data.
- Solutions require finding $k$.
- We can determine $k$ by examining noise in the basis.
- Using the GSVD we see that Tikhonov regularization yields a basis with smoothed basis vectors.
- But we can apply the same techniques.
- Stabilizing both TSVD and TGSVD is robust to parameter estimation.
- Parameter estimation for truncated expansions is cheaper.
- Reminder truncation $k$ is independent of the data $b$. 
Extensions

- Truncating the basis means that we will not see high frequency in the solutions
- We cannot represent edges or steep gradients
- Need to use Total Variation or alternate regularizers for feature enhancement
- Still will be important to analyze the noise
- Use the same ideas to further analyze TV solutions - work in progress
Extensions

- Truncating the basis means that we will not see high frequency in the solutions
- We cannot represent edges or steep gradients
- Need to use Total Variation or alternate regularizers for feature enhancement
- Still will be important to analyze the noise
- Use the same ideas to further analyze TV solutions - work in progress
Extensions

- Truncating the basis means that we will not see high frequency in the solutions
- We cannot represent edges or steep gradients
- Need to use Total Variation or alternate regularizers for feature enhancement
- Still will be important to analyze the noise
- Use the same ideas to further analyze TV solutions - work in progress
Extensions

- Truncating the basis means that we will not see high frequency in the solutions
- We cannot represent edges or steep gradients
- Need to use Total Variation or alternate regularizers for feature enhancement
- Still will be important to analyze the noise
- Use the same ideas to further analyze TV solutions - work in progress
Extensions

- Truncating the basis means that we will not see high frequency in the solutions
- We cannot represent edges or steep gradients
- Need to use Total Variation or alternate regularizers for feature enhancement
- Still will be important to analyze the noise
- Use the same ideas to further analyze TV solutions - work in progress
Picard condition for the GSVD: for $\mathbf{x}$ and $L\mathbf{x}$ examine the weights in the expansion.

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

Figure: Weights for the expansion - $\lambda = .05$ 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. $\mu_i \approx 1$
Illustrating Total Variation Solutions for noise level 0.1 in the data

Figure: \( \lambda = 0.1, \gamma = 0.5 \): We see that final \( \mathbf{h} \) is too large relative to \( \mathbf{b} \)
Illustrating Total Variation Solutions for noise level .1 in the data

Figure: $\lambda = 1$, $\gamma = .5$: 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 $d$. 
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 $d$. 
Exploiting the GSVD for analysis

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

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

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

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

\[ \frac{\phi_i}{\nu_i} \quad \frac{\phi_i \mu_i}{\nu_i} \quad \frac{(1 - \phi_i)}{\mu_i} \quad (1 - \phi_i) \]
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

Figure: $\lambda = 10$

Of course the coefficients are independent of the data.
Picard Condition using GSVD for noise level 0.1

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 .1

But $h$ changes with the iteration
But $h$ changes with the iteration
Further Observation

- 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

Noise level 0.099428 Updating $\gamma$
LS
SB with updated $\lambda$
SB Fixed $\gamma$

SB Update
SB IRN Update

Figure: $\gamma = 5$

SB with updated $\lambda$ is useful
**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 $h$. 

---

Further Observations and Future Work
See me for extensive references to literature

THANK YOU