Improving the Efficiency of the Chi-squared Method for Regularization Parameter Estimation

Rosemary Renaut, Iveta Hnetynkova, Jodi Mead

Arizona State, Charles University and Boise State

SIAM Annual Meeting 2008
Least Squares for $Ax = b$, (Weighted)

- Consider discrete systems: $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $x \in \mathbb{R}^n$
  
  $$Ax = b + e,$$

- $e$ is the $m$–vector of random measurement errors with mean 0 and positive definite covariance matrix
  
  $$C_b = E(ee^T).$$

- Assume that $C_b$ is known. (Calculate if given multiple $b$)

- For uncorrelated measurements $C_b$ is diagonal matrix of standard deviations of the errors. (Colored noise)

- For correlated measurements, let $W_b = C_b^{-1}$ and $L_b L_b^T = W_b$ be the Choleski factorization of $W_b$ and weight the equation:
  
  $$L_b Ax = L_b b + \tilde{e},$$

- $\tilde{e}$ are uncorrelated. (White noise).

- $\tilde{e} \sim N(0, I)$, normally distributed mean 0 and variance $I$. 

Renaut, Hnetynkova, Mead (USA/CR) Large Scale Parameter Estimation July 8 3 / 27
Weighted Regularized Least Squares for numerically ill-posed systems

Formulation:

\[ \hat{x} = \arg\min_x J(x) = \arg\min_x \left\{ \| Ax - b \|^2_{W_b} + \| x - x_0 \|^2_{W_x} \right\}. \quad (1) \]

\( x_0 \) is a reference solution, often \( x_0 = 0 \).

- **Standard**: \( W_x = \lambda^2 I \), \( \lambda \) unknown penalty parameter.
- Statistically, \( W_x \) is inverse covariance matrix for the model \( x \) i.e. \( \lambda = 1/\sigma_x \), \( \sigma_x^2 \) the common variance in \( x \).
- Assumes the resulting estimates for \( x \) uncorrelated.
- \( \hat{x} \) is the standard **maximum a posteriori** (MAP) estimate of the solution

**Question: The Problem**

How do we find an *appropriate* regularization parameter \( \lambda \)?

More generally, what is the correct \( W_x \)?
Weighted Regularized Least Squares for numerically ill-posed systems

Formulation:

\[ \hat{x} = \arg\min J(x) = \arg\min \{ \|Ax - b\|_{W_b}^2 + \|x - x_0\|_{W_x}^2 \} . \] (1)

\( x_0 \) is a reference solution, often \( x_0 = 0 \).

- **Standard:** \( W_x = \lambda^2 I \), \( \lambda \) unknown penalty parameter.
- Statistically, \( W_x \) is inverse covariance matrix for the model \( x \) i.e. \( \lambda = 1/\sigma_x \), \( \sigma_x^2 \) the common variance in \( x \).
- Assumes the resulting estimates for \( x \) uncorrelated.
- \( \hat{x} \) is the standard maximum a posteriori (MAP) estimate of the solution

Question: The Problem

How do we find an *appropriate* regularization parameter \( \lambda \)?
More generally, what is the correct \( W_x \)?
Generalized Tikhonov regularization, operator $D$ acts on $x$.

$$\hat{x} = \arg\min J_D(x) = \arg\min \{ \|Ax - b\|^2_{W_b} + \|x - x_0\|^2_{W_D} \}. \quad (2)$$

- Assume invertibility $\mathcal{N}(A) \cap \mathcal{N}(D) = \emptyset$
- Then solutions depend on $W_D = \lambda^2 D^T D$:

$$\hat{x}(\lambda) = \arg\min J_D(x) = \arg\min \{ \|Ax - b\|^2_{W_b} + \lambda^2 \|D(x - x_0)\|^2 \}. \quad (3)$$

**GOAL**

- Can we estimate $\lambda$ efficiently when $W_b$ is known?
- Use statistics of the solution to find $\lambda$. 
Formulation: Regularization with Solution Mapping

Generalized Tikhonov regularization, operator $D$ acts on $x$.

\[ \hat{x} = \text{argmin } J_D(x) = \text{argmin } \{ \|Ax - b\|_{W_b}^2 + \|x - x_0\|_{W_D}^2 \}. \] (2)

- Assume invertibility $\mathcal{N}(A) \cap \mathcal{N}(D) = \emptyset$
- Then solutions depend on $W_D = \lambda^2 D^T D$

\[ \hat{x}(\lambda) = \text{argmin } J_D(x) = \text{argmin } \{ \|Ax - b\|_{W_b}^2 + \lambda^2 \|D(x - x_0)\|^2 \}. \] (3)

GOAL

- Can we estimate $\lambda$ efficiently when $W_b$ is known?
- Use statistics of the solution to find $\lambda$. 
Background: Statistics of the Least Squares Problem

Theorem (Rao73: First Fundamental Theorem)

Let $r$ be the rank of $A$ and for $b \sim N(Ax, \sigma_b^2 I)$, (errors in measurements are normally distributed with mean 0 and covariance $\sigma_b^2 I$), then

$$J = \min_x \|Ax - b\|^2 \sim \sigma_b^2 \chi^2(m - r).$$

$J$ follows a $\chi^2$ distribution with $m - r$ degrees of freedom.

Corollary (Weighted Least Squares)

For $b \sim N(Ax, C_b)$, and $W_b = C_b^{-1}$ then

$$J = \min_x \|Ax - b\|_{W_b}^2 \sim \chi^2(m - r).$$
Theorem (Rao73: First Fundamental Theorem)

Let $r$ be the rank of $A$ and for $b \sim N(Ax, \sigma_b^2 I)$, (errors in measurements are normally distributed with mean 0 and covariance $\sigma_b^2 I$), then

$$J = \min_x \|Ax - b\|^2 \sim \sigma_b^2 \chi^2(m - r).$$

$J$ follows a \(\chi^2\) distribution with \(m - r\) degrees of freedom.

Corollary (Weighted Least Squares)

For $b \sim N(Ax, C_b)$, and $W_b = C_b^{-1}$ then

$$J = \min_x \|Ax - b\|_{W_b}^2 \sim \chi^2(m - r).$$
Theorem: $\chi^2$ distribution of the regularized functional

$$\hat{x} = \arg\min J_D(x) = \arg\min \{ \| Ax - b \|_{W_b}^2 + \| (x - x_0) \|_{W_D}^2 \}, \quad W_D = D^T W_x D.$$  

(4)

Assume

- $W_b$ and $W_x$ are symmetric positive definite.
- Problem is uniquely solvable $\mathcal{N}(A) \cap \mathcal{N}(D) \neq \emptyset$.
- Moore-Penrose generalized inverse of $W_D$ is $C_D$.
- Statistics: $(b - Ax) = e \sim N(0, C_b), (x - x_0) = f \sim N(0, C_D)$,  
  - $e$ and $f$ are i.i.d. random variables.

Then $J_D \sim \chi^2(m + p - n)$:

$J_D$ is a random variable which follows a $\chi^2$ distribution with $m + p - n$ degrees of freedom.
Key Aspects of the Proof I: The Functional $J$

**Algebraic Simplifications**

- Regularized solution given in terms of **resolution** matrix $R(W_D)$

\[
\hat{x} = x_0 + (A^T W_b A + D^T W_x D)^{-1} A^T W_b r, \quad (5)
\]

\[
= x_0 + R(W_D) W_b^{1/2} r, \quad r = b - A x_0
\]

\[
= x_0 + y(W_D). \quad (6)
\]

\[
R(W_D) = (A^T W_b A + D^T W_x D)^{-1} A^T W_b^{1/2} \quad (7)
\]

- Functional is given in terms of **influence matrix** $A(W_D)$

\[
A(W_D) = W_b^{1/2} A R(W_D) \quad (8)
\]

\[
J_D(\hat{x}) = r^T W_b^{1/2} (I_m - A(W_D)) W_b^{1/2} r, \quad \text{let } \tilde{r} = W_b^{1/2} r \quad (9)
\]

\[
= \tilde{r}^T (I_m - A(W_D)) \tilde{r}. \quad (10)
\]
Key Aspects of the Proof II: Requires the GSVD

**Lemma**

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 \( X \in \mathbb{R}^{n \times n} \) such that

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

(11)

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

\[
0 \leq \upsilon_1 \leq \cdots \leq \upsilon_p \leq 1, \quad 1 \geq \mu_1 \geq \cdots \geq \mu_p > 0,
\]

(12)

\[
\upsilon_i^2 + \mu_i^2 = 1, \quad i = 1, \ldots, p.
\]

The Functional with the GSVD

Let \( Q = \text{diag}(\mu_1, \ldots, \mu_p, 0_{n-p}, I_{m-n}) \),

then 
\[
J = \tilde{r}^T(I_m - A(W_D))\tilde{r} = \|QU^T\tilde{r}\|_2^2,
\]
Lemma

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 \( X \in \mathbb{R}^{n \times n} \) such that

\[
A = U \begin{bmatrix} \Upsilon & \end{bmatrix} X^T, \quad D = V [M, 0_{p \times (n-p)}] X^T, \quad (11)
\]

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

\[
0 \leq \upsilon_1 \leq \cdots \leq \upsilon_p \leq 1, \quad 1 \geq \mu_1 \geq \cdots \geq \mu_p > 0,
\]

\[
\upsilon_i^2 + \mu_i^2 = 1, \quad i = 1, \ldots, p. \quad (12)
\]

The Functional with the GSVD

Let \( Q = \text{diag}(\mu_1, \ldots, \mu_p, 0_{n-p}, I_{m-n}) \)

then \( J = \tilde{r}^T (I_m - A(W_D))\tilde{r} = \|QU^T\tilde{r}\|_2^2 \),
Key Aspects of the Proof III: Statistical Distribution of the Weighted Residual

Covariance Structure

- \( e = Ax - b \sim N(0, C_b) \) hence we can show
- \( b \sim N(Ax_0, C_b + AC_D A^T) \) Note that \( b \) depends on \( x \).
- \( r \sim N(0, C_b + AC_D A^T) \), and \( \tilde{r} \sim N(0, I + \tilde{A}C_D \tilde{A}^T) \), \( \tilde{A} = W_b^{1/2} A \).
- Use the GSVD
  \[
  I + \tilde{A}C_D \tilde{A}^T = UP^2 U^T,
  \]
  where \( P = \text{diag}(\mu_1^2, \ldots, \mu_p^2, I_{n-p}, I_{m-n}) \).

The Functional is a rv

- Let \( k = P^{-1} U^T \tilde{r} \), then \( k \sim N(0, P^{-1} U^T (UP^2 U^T) UP^{-1}) \sim N(0, I_m) \)
- Now \( J = \tilde{r} UQU^T \tilde{r} = k^T PQPk \). Thus
  \[
  J_D = \sum_{i=1}^{p} k_i^2 + \sum_{i=n+1}^{m} k_i^2 \sim \chi^2(m + p - n).
  \]
Covariance Structure

- $e = Ax - b \sim N(0, C_b)$ hence we can show $b \sim N(Ax_0, C_b + AC_D A^T)$ Note that $b$ depends on $x$.
- $r \sim N(0, C_b + AC_D A^T)$, and $\tilde{r} \sim N(0, I + \tilde{A}C_D \tilde{A}^T)$, $\tilde{A} = W_b^{1/2} A$.
- Use the GSVD

\[ I + \tilde{A}C_D \tilde{A}^T = UP^2 U^T, \]
\[ P = \text{diag}(\mu_1^2, \ldots, \mu_p^2, I_{n-p}, I_{m-n}) \]

The Functional is a rv

- Let $k = P^{-1} U^T \tilde{r}$, then $k \sim N(0, P^{-1} U^T (UP^2 U^T) UP^{-1}) \sim N(0, I_m)$
- Now $J = \tilde{r}UQU^T \tilde{r} = k^T PQP k$. Thus

\[ J_D = \sum_{i=1}^{p} k_i^2 + \sum_{i=n+1}^{m} k_i^2 \sim \chi^2(m + p - n). \]
Implication of $J_D \sim \chi^2(m + p - n)$

DESIGNING THE ALGORITHM: I

- If $C_b$ and $C_x$ are good estimates of the covariance matrices

\[ |J_D(\hat{x}) - (m + p - n)| \]

should be small.
- Thus, let $\tilde{m} = m + p - n$ then we want

\[ \tilde{m} - \sqrt{2\tilde{m}}z_{\alpha/2} < r^T W_b^{1/2}(I_m - A(W_D)) W_b^{1/2}r < \tilde{m} + \sqrt{2\tilde{m}}z_{\alpha/2}. \]  

(13)

- $z_{\alpha/2}$ is the relevant $z$-value for a $\chi^2$-distribution with $\tilde{m}$ degrees

GOAL

Find $W_x$ to make (13) tight: Single Variable case find $\lambda$

$J_D(\hat{x}(\lambda)) \approx \tilde{m}$
Implication of $J_D \sim \chi^2(m + p - n)$

**DESIGNING THE ALGORITHM: I**

- If $C_b$ and $C_x$ are good estimates of the covariance matrices

$$|J_D(\hat{x}) - (m + p - n)|$$

should be small.

Thus, let $\tilde{m} = m + p - n$ then we want

$$\tilde{m} - \sqrt{2\tilde{m}}z_{\alpha/2} < r^T W_b^{1/2}(I_m - A(W_D)) W_b^{1/2} r < \tilde{m} + \sqrt{2\tilde{m}}z_{\alpha/2}.$$  

(13)

$z_{\alpha/2}$ is the relevant $z$-value for a $\chi^2$-distribution with $\tilde{m}$ degrees

**GOAL**

Find $W_x$ to make (13) tight: Single Variable case find $\lambda$

$$J_D(\hat{x}(\lambda)) \approx \tilde{m}$$
A Newton-line search Algorithm to find $\lambda$.

### Newton to Solve $F(\sigma) = J_D(\sigma) - \tilde{m} = 0$

- We use $\sigma = 1/\lambda$, and $y(\sigma^{(k)})$ is the current solution for which $x(\sigma^{(k)}) = y(\sigma^{(k)}) + x_0$

Then

$$\frac{\partial}{\partial \sigma} J(\sigma) = -\frac{2}{\sigma^3} \|Dy(\sigma)\|^2 < 0$$

- Hence we have a basic Newton Iteration

$$\sigma^{(k+1)} = \sigma^{(k)} \left(1 + \frac{1}{2} \left(\frac{\sigma^{(k)}}{\|Dy\|}\right)^2 (J_D(\sigma^{(k)}) - \tilde{m})\right).$$
Convergence

- $F$ is monotonic decreasing
- Solution either exists and is **unique** for positive $\sigma$
- Or no solution exists $F(0) < 0$.
  - implies incorrect statistics of the model
- Theoretically, $\lim_{\sigma \to \infty} F > 0$ possible.
  - Equivalent to $\lambda = 0$. No regularization needed.
Practical Details of Algorithm

Initialization
- Convert generalized Tikhonov problem to standard form.
- Lancos-hybrid projects to smaller problem with bidiagonal matrix.
- Each $\sigma$ calculation of algorithm reuses saved information from the Lancos bidiagonalization. The system is augmented if needed.

Find the parameter
- Step 1: Bracket the root by logarithmic search on $\sigma$ to handle the asymptotes: yields $\text{sigmamax}$ and $\text{sigmamin}$
- Step 2: Calculate step, with steepness controlled by $\text{tolD}$. Let $t = \frac{Dy}{\sigma^{(k)}}$ then
  \[
  \text{step} = \frac{1}{2} \left( \frac{1}{\max \{\|t\|, \text{tolD}\}} \right)^2 (J_D(\sigma^{(k)}) - \tilde{m})
  \]
- Step 3: Introduce line search $\alpha^{(k)}$ in Newton
  \[
  \text{sigmanew} = \sigma^{(k)} (1 + \alpha^{(k) \text{step}})
  \]
  $\alpha^{(k)}$ chosen such that $\text{sigmanew}$ within bracket.
Practical Details of Algorithm

Initialization
- Convert generalized Tikhonov problem to standard form.
- Lanczos-hybrid projects to smaller problem with bidiagonal matrix.
- Each $\sigma$ calculation of algorithm reuses saved information from the Lancos bidiagonalization. The system is augmented if needed.

Find the parameter
- **Step 1:** Bracket the root by logarithmic search on $\sigma$ to handle the asymptotes: yields $\sigma_{\text{max}}$ and $\sigma_{\text{min}}$
- **Step 2:** Calculate step, with steepness controlled by $\text{tolD}$. Let $t = D y / \sigma^{(k)}$ then
  \[
  \text{step} = \frac{1}{2} \left( \frac{1}{\max \{ \|t\|, \text{tolD} \}} \right)^2 \left( J_D(\sigma^{(k)}) - \tilde{m} \right)
  \]
- **Step 3:** Introduce line search $\alpha^{(k)}$ in Newton
  \[
  \sigma_{\text{new}} = \sigma^{(k)} (1 + \alpha^{(k)} \text{step})
  \]
  $\alpha^{(k)}$ chosen such that $\sigma_{\text{new}}$ within bracket.
Numerical Experiments and Assumptions

Covariance of Error: Statistics of Measurement Errors

- Information on the covariance structure of errors in $b$ needed.
- Use $C_b = \sigma_b^2 I$ for common covariance, white noise.
- Use $C_b = \text{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2)$ for colored uncorrelated noise.
- With no noise information $C_b = I$.

Tolerance on Convergence

- The convergence tolerance depends on the noise structure.
- Use $TOL = \sqrt{2\tilde{m}z_\alpha/2}$.
- No noise structure use $\alpha = .001$, generates large $TOL$.
- Good noise information use $\alpha = .95$, generates small $TOL$.
Numerical Experiments and Assumptions

**Covariance of Error: Statistics of Measurement Errors**

- Information on the covariance structure of errors in $\mathbf{b}$ needed.
- Use $\mathbf{C}_b = \sigma_b^2 \mathbf{I}$ for common covariance, **white noise**.
- Use $\mathbf{C}_b = \text{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2)$ for **colored uncorrelated noise**.
- With no noise information $\mathbf{C}_b = \mathbf{I}$.

**Tolerance on Convergence**

- The convergence tolerance depends on the noise structure.
- Use $\text{TOL} = \sqrt{2\tilde{m}z_{\alpha/2}}$.
- No noise structure use $\alpha = .001$, generates large TOL
- Good noise information use $\alpha = .95$, generates small TOL
The Data Set and Goal

- Real data set of 48 signals of length 3000.
- The point spread function is derived from the signals.
- Calculate the signal variance pointwise over all 48 signals.
- Goal: restore the signal $\mathbf{x}$ from $A\mathbf{x} = \mathbf{b}$, where $A$ is psf matrix and $\mathbf{b}$ is given blurred signal.

Method of Comparison

- No exact solution.
- Downsample the signal and restore for different resolutions:
  - Resolution: 2 : 1, 5 : 1, 10 : 1, 20 : 1, 100 : 1
  - Points: 1500, 600, 300, 150, 30
- Do results converge?
Real data: Seismic Signal Restoration

The Data Set and Goal
- Real data set of 48 signals of length 3000.
- The point spread function is derived from the signals.
- Calculate the signal variance pointwise over all 48 signals.
- Goal: restore the signal $\mathbf{x}$ from $A\mathbf{x} = \mathbf{b}$, where $A$ is psf matrix and $\mathbf{b}$ is given blurred signal.

Method of Comparison
- No exact solution.
- Downsample the signal and restore for different resolutions
  - Resolution: 2 : 1  5 : 1  10 : 1  20 : 1  100 : 1
  - Points: 1500  600  300  150  30
- Do results converge?
THE UPRE SOLUTION

Regularization Parameters are consistent: $\sigma = 0.01005$ all resolutions
Regularization Parameters are consistent:
\[ \sigma = 0.00069, 0.00069, 0.00069, 0.00065, 0.00065, \text{ resolution from 2 to 100} \]
Comparison

Greater contrast with $\chi^2$ parameter estimation than with UPRE. Regularization Parameters of L-curve (not shown) are not consistent across resolutions, solution is poor.
Some Small Scale Experiments:Verify Robustness of LSQR/GSVD

Details

- Take example from Hansen’s toolbox, eg shaw, philips, heat, ilaplace.
- Generate 500 copies for each noise level, here .005, .01, .05, .1.
- Solve for 500 cases using GSVD and LSQR Newton.
- Pairwise t test on obtained \( \sigma \): verify equivalence GSVD and LSQR.
- Compare results with statistical technique: UPRE
  - Errors - relative least squares, and max error. Calculate over all errors less than .5.
  - Regularization parameter (not given).
Example of the Colored Noise distribution

The pointwise variance for each noise level
Noise and Exact Right Hand Side Vector

Noise distribution for problem shaw

Actual Noise Levels

Renaut, Hnetynkova, Mead (USA/CR)
Large Scale Parameter Estimation
July 8 22 / 27
Noise and Exact Right Hand Side Vector

Noise distribution for problem iLaplace

Actual Noise Levels

Renaut, Hnetynkova, Mead (USA/CR)  Large Scale Parameter Estimation  July 8 22/27
Noise and Exact Right Hand Side Vector

Noise distribution for problem heat

Actual Noise Levels

Renaut, Hnetynkova, Mead (USA/CR)  Large Scale Parameter Estimation  July 8  22 / 27
Noise and Exact Right Hand Side Vector

Noise distribution for problem phillips

Actual Noise Levels

Renaut, Hnetynkova, Mead (USA/CR)
Large Scale Parameter Estimation
July 8
Table: Convergence characteristics and P-values comparing GSVD and LSQR iteration steps and $\sigma$ values

<table>
<thead>
<tr>
<th></th>
<th>Bidiagonal Average</th>
<th>Newton Steps</th>
<th>P value Iteration</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$5e - 02$</td>
<td>7.2</td>
<td>7.1</td>
<td>1.000e + 00</td>
<td>1.000e + 00</td>
</tr>
<tr>
<td>$1e - 01$</td>
<td>6.2</td>
<td>7.0</td>
<td>9.166e - 01</td>
<td>8.489e - 01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>GSVD</th>
<th>LSQR</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$5e - 02$</td>
<td>6.5</td>
<td>6.5</td>
<td>9.991e - 01</td>
</tr>
<tr>
<td>$1e - 01$</td>
<td>6.5</td>
<td>6.5</td>
<td>4.870e - 01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$5e - 02$</td>
<td>1.000e + 00</td>
</tr>
<tr>
<td>$1e - 01$</td>
<td>1.000e + 00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>GSVD</th>
<th>LSQR</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$5e - 02$</td>
<td>5.8</td>
<td>5.8</td>
<td>1.000e + 00</td>
</tr>
<tr>
<td>$1e - 01$</td>
<td>6.0</td>
<td>6.0</td>
<td>1.000e + 00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$5e - 02$</td>
<td>1.000e + 00</td>
</tr>
<tr>
<td>$1e - 01$</td>
<td>1.000e + 00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>GSVD</th>
<th>LSQR</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$5e - 02$</td>
<td>6.1</td>
<td>6.1</td>
<td>1.000e + 00</td>
</tr>
<tr>
<td>$1e - 01$</td>
<td>6.1</td>
<td>6.1</td>
<td>1.000e + 00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$5e - 02$</td>
<td>1.000e + 00</td>
</tr>
<tr>
<td>$1e - 01$</td>
<td>1.000e + 00</td>
</tr>
</tbody>
</table>
Problem size 64, Regularization First Order Derivative, Colored Noise

<table>
<thead>
<tr>
<th></th>
<th>shaw</th>
<th>Max Error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Least Squares Error</td>
<td></td>
</tr>
<tr>
<td>( \epsilon )</td>
<td>GSVD</td>
<td>LSQR</td>
</tr>
<tr>
<td>.05</td>
<td>.34(207)</td>
<td>.34(207)</td>
</tr>
<tr>
<td>.1</td>
<td>.39(105)</td>
<td>.39(105)</td>
</tr>
</tbody>
</table>

|        | ilaplace      |          |
|        | Least Squares Error |          |
| \( \epsilon \) | GSVD | LSQR | UPRE | GSVD | LSQR | UPRE |
| .05    | .23(361) | .23(361) | .16(398) | .12(495) | .12(495) | .07(425) |
| .1     | .25(317) | .25(317) | .21(370) | .15(490) | .15(490) | .10(428) |

|        | heat          |          |
|        | Least Squares Error |          |
| \( \epsilon \) | GSVD | LSQR | UPRE | GSVD | LSQR | UPRE |
| .05    | .27(500) | .27(500) | .28(479) | .19(500) | .19(500) | .20(489) |
| .1     | .35(497) | .35(497) | .36(461) | .27(500) | .27(500) | .27(489) |

|        | phillips      |          |
|        | Least Squares Error |          |
| \( \epsilon \) | GSVD | LSQR | UPRE | GSVD | LSQR | UPRE |
| .05    | .10(500) | .10(500) | .10(455) | .10(500) | .10(500) | .09(455) |
| .1     | .12(500) | .12(500) | .11(431) | .12(500) | .12(500) | .11(431) |

**Table:** Comparison with UPRE, Relative Least Squares and max error, in parentheses the number of accepted values, large values (\( > .5 \)) excluded from the average.
Summary of Results

Major Observations

**GSVD-LSQR**
- High correlation between $\sigma$ for both $\chi^2$ (GSVD and LSQR) obtained. $p$-values at or near 1.
- Algorithms converge with very few regularization calculations, average $\approx 7$.
- The bidiagonalized system is on average much smaller than the system size.
- Errors distributions equivalent.
- Total cost of LSQR is the cost of bidiagonalization plus use of bidiagonalization to obtain a solution, eg column 2 plus column 4 solves.

**UPRE**
- Errors of UPRE are similar to $\chi^2$ approaches.
- UPRE fails more often (more discarded solutions with high noise).
Conclusions and Future Work

Conclusions

- A new statistical method for estimating regularization parameter
  - Compares favorably with UPRE with respect to performance
- Method can be used for large scale problems, without GSVD
- Method is very efficient, Newton method is robust and fast.

Future Work

- Investigate robustness with respect to conditioning of $C_b$
- Image deblurring (with Hnetynkova in prep.)
- Diagonal Weighting Schemes
- Edge preserving regularization
Conclusions and Future Work

Conclusions

- A new statistical method for estimating regularization parameter
  - Compares favorably with UPRE with respect to performance
- Method can be used for large scale problems, without GSVD
- Method is very efficient, Newton method is robust and fast.

Future Work

- Investigate robustness with respect to conditioning of $C_b$
- Image deblurring (with Hnetynkova in prep.)
- Diagonal Weighting Schemes
- Edge preserving regularization
THANK YOU!