The Chi-squared Distribution of the Regularized Least Squares Functional for Regularization Parameter Estimation

Rosemary Renaut

ARIZONA STATE UNIVERSITY
DEPARTMENT OF MATHEMATICS AND STATISTICS

GAMM Workshop 2008

Outline

Least Squares for $A\mathbf{x} = \mathbf{b}$, (Weighted)

- Consider discrete systems: $A \in \mathcal{R}^{m\times n}$, $\mathbf{b} \in \mathcal{R}^m$, $\mathbf{x} \in \mathcal{R}^n$

$$A\mathbf{x} = \mathbf{b} + \mathbf{e},$$

  - $\mathbf{e}$ is the $m-$vector of random measurement errors with mean 0 and **positive definite covariance** matrix

$$C_{\mathbf{b}} = \mathbf{E}(\mathbf{e}\mathbf{e}^T).$$

  - Assume that $C_{\mathbf{b}}$ is known. (Calculate if given multiple $\mathbf{b}$)
- For **uncorrelated** measurements $C_{\mathbf{b}}$ is **diagonal** matrix of **standard deviations** of the errors. (Colored noise)
- For **correlated** measurements, let $W_{\mathbf{b}} = {C_{\mathbf{b}}}^{-1}$ and $L_{\mathbf{b}}{L_{\mathbf{b}}}^T = W_{\mathbf{b}}$ be the Choleski factorization of $W_{\mathbf{b}}$ and weight the equation:

$$L_{\mathbf{b}}A\mathbf{x} = L_{\mathbf{b}}\mathbf{b} + \tilde{\mathbf{e}},$$

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

Weighted Regularized Least Squares for numerically ill-posed systems

Formulation:

$$\hat{\mathbf{x}} = \operatorname{argmin} J(\mathbf{x}) = \operatorname{argmin}\{\|A\mathbf{x} - \mathbf{b}\|^2_{W_{\mathbf{b}}} + \|\mathbf{x} - \mathbf{x}_0\|^2_{W_{\mathbf{x}}}\}. \quad (1)$$

$\mathbf{x}_0$ is a reference solution, often $\mathbf{x}_0 = 0$.

- **Standard:** $W_{\mathbf{x}} = \lambda^2 I$, λ unknown penalty parameter.
- Statistically, $W_{\mathbf{x}}$ is inverse covariance matrix for the model $\mathbf{x}$ i.e. $\lambda = 1/\sigma_{\mathbf{x}}$, $\sigma^2_{\mathbf{x}}$ the common variance in $\mathbf{x}$.
- Assumes the resulting estimates for $\mathbf{x}$ uncorrelated.
- $\hat{\mathbf{x}}$ is the standard **maximum a posteriori** (MAP) estimate of the solution, when all *a priori* information is provided.

Question: The Problem

How do we find an *appropriate* regularization parameter λ?
More generally, what is the correct $W_{\mathbf{x}}$?

Weighted Regularized Least Squares for numerically ill-posed systems

Formulation:

$$\hat{\mathbf{x}} = \operatorname{argmin} J(\mathbf{x}) = \operatorname{argmin}\{\|A\mathbf{x} - \mathbf{b}\|^2_{W_{\mathbf{b}}} + \|\mathbf{x} - \mathbf{x}_0\|^2_{W_{\mathbf{x}}}\}. \quad (1)$$

$\mathbf{x}_0$ is a reference solution, often $\mathbf{x}_0 = 0$.

- **Standard:** $W_{\mathbf{x}} = \lambda^2 I$, λ unknown penalty parameter.
- Statistically, $W_{\mathbf{x}}$ is inverse covariance matrix for the model $\mathbf{x}$ i.e. $\lambda = 1/\sigma_{\mathbf{x}}$, $\sigma^2_{\mathbf{x}}$ the common variance in $\mathbf{x}$.
- Assumes the resulting estimates for $\mathbf{x}$ uncorrelated.
- $\hat{\mathbf{x}}$ is the standard **maximum a posteriori** (MAP) estimate of the solution, when all *a priori* information is provided.

Question: The Problem

How do we find an *appropriate* regularization parameter λ?
More generally, what is the correct $W_{\mathbf{x}}$?

The General Case : Generalized Tikhonov Regularization

Formulation: Regularization with Solution Mapping

Generalized Tikhonov regularization, operator D acts on $\mathbf{x}$.

$$\hat{\mathbf{x}} = \operatorname{argmin} J_D(\mathbf{x}) = \operatorname{argmin}\{\|A\mathbf{x} - \mathbf{b}\|^2_{W_{\mathbf{b}}} + \|(\mathbf{x} - \mathbf{x}_0)\|^2_{W_{\mathbf{D}}}\}. \quad (2)$$

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

$$\hat{\mathbf{x}}(\lambda) = \operatorname{argmin} J_D(\mathbf{x}) = \operatorname{argmin}\{\|A\mathbf{x} - \mathbf{b}\|^2_{W_{\mathbf{b}}} + \lambda^2\|D(\mathbf{x} - \mathbf{x}_0)\|^2\}. \quad (3)$$

GOAL

- **Can we estimate λ efficiently when $W_{\mathbf{b}}$ is known?**
- **Use statistics of the solution to find λ.**

The General Case : Generalized Tikhonov Regularization

Formulation: Regularization with Solution Mapping

Generalized Tikhonov regularization, operator D acts on $\mathbf{x}$.

$$\hat{\mathbf{x}} = \operatorname{argmin} J_D(\mathbf{x}) = \operatorname{argmin}\{\|A\mathbf{x} - \mathbf{b}\|^2_{W_{\mathbf{b}}} + \|(\mathbf{x} - \mathbf{x}_0)\|^2_{W_{\mathbf{D}}}\}. \quad (2)$$

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

$$\hat{\mathbf{x}}(\lambda) = \operatorname{argmin} J_D(\mathbf{x}) = \operatorname{argmin}\{\|A\mathbf{x} - \mathbf{b}\|^2_{W_{\mathbf{b}}} + \lambda^2\|D(\mathbf{x} - \mathbf{x}_0)\|^2\}. \quad (3)$$

GOAL

- **Can we estimate λ efficiently when $W_{\mathbf{b}}$ is known?**
- **Use statistics of the solution to find λ.**

Background: Statistics of the Least Squares Problem

Theorem (Rao73: First Fundamental Theorem)

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

$$J = \min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}\|^2 \sim \sigma_{\mathbf{b}}^2 \chi^2(m-r).$$

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

Corollary (Weighted Least Squares)

For $\mathbf{b} \sim N(A\mathbf{x}, C_{\mathbf{b}})$, and $W_{\mathbf{b}} = {C_{\mathbf{b}}}^{-1}$ then

$$J = \min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}\|^2_{W_{\mathbf{b}}} \sim \chi^2(m-r).$$

Background: Statistics of the Least Squares Problem

Theorem (Rao73: First Fundamental Theorem)

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

$$J = \min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}\|^2 \sim \sigma_{\mathbf{b}}^2 \chi^2(m-r).$$

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

Corollary (Weighted Least Squares)

For $\mathbf{b} \sim N(A\mathbf{x}, C_{\mathbf{b}})$, and $W_{\mathbf{b}} = {C_{\mathbf{b}}}^{-1}$ then

$$J = \min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}\|_{W_{\mathbf{b}}}^2 \sim \chi^2(m-r).$$

Extension: Statistics of the Regularized Least Squares Problem

Theorem: χ^2 distribution of the regularized functional

$$\hat{\mathbf{x}} = \operatorname{argmin} J_D(\mathbf{x}) = \operatorname{argmin}\{\|A\mathbf{x}-\mathbf{b}\|^2_{W_{\mathbf{b}}} + \|(\mathbf{x}-\mathbf{x}_0)\|^2_{W_{\mathbf{D}}}\}, \quad W_{\mathbf{D}} = D^T W_{\mathbf{x}} D. \tag{4}$$

Assume

- $W_{\mathbf{b}}$ and $W_{\mathbf{x}}$ are symmetric positive definite.
- Problem is uniquely solvable $\mathcal{N}(A) \cap \mathcal{N}(D) \neq 0$.
- Moore-Penrose generalized inverse of $W_{\mathbf{D}}$ is $C_{\mathbf{D}}$
- Statistics: $(\mathbf{b}-A\mathbf{x}) = \mathbf{e} \sim N(0, C_{\mathbf{b}})$, $(\mathbf{x}-\mathbf{x}_0) = \mathbf{f} \sim N(0, C_{\mathbf{D}})$,
  - $\mathbf{x}_0$ is the mean vector of the model parameters.

Then

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

Key Aspects of the Proof I: The Functional J

Algebraic Simplifications: Rewrite functional as quadratic form

- Regularized solution given in terms of **resolution** matrix $R(W_{\mathbf{D}})$

$$\begin{aligned}
\hat{\mathbf{x}} &= \mathbf{x}_0 + (A^T W_{\mathbf{b}} A + D^T W_{\mathbf{x}} D)^{-1} A^T W_{\mathbf{b}} \mathbf{r}, && (5)\\
&= \mathbf{x}_0 + R(W_{\mathbf{D}}) {W_{\mathbf{b}}}^{1/2} \mathbf{r}, \quad \mathbf{r} = \mathbf{b} - A\mathbf{x}_0 \\
&= \mathbf{x}_0 + \mathbf{y}(W_{\mathbf{D}}). && (6)\\
R(W_{\mathbf{D}}) &= (A^T W_{\mathbf{b}} A + D^T W_{\mathbf{x}} D)^{-1} A^T {W_{\mathbf{b}}}^{1/2} && (7)
\end{aligned}$$

- Functional is given in terms of **influence matrix** $A(W_{\mathbf{D}})$

$$\begin{aligned}
A(W_{\mathbf{D}}) &= {W_{\mathbf{b}}}^{1/2} A R(W_{\mathbf{D}}) && (8)\\
J_D(\hat{\mathbf{x}}) &= \mathbf{r}^T {W_{\mathbf{b}}}^{1/2} (I_m - A(W_{\mathbf{D}})) {W_{\mathbf{b}}}^{1/2} \mathbf{r}, \quad \text{let} \quad \tilde{\mathbf{r}} = {W_{\mathbf{b}}}^{1/2} \mathbf{r} && (9)\\
&= \tilde{\mathbf{r}}^T (I_m - A(W_{\mathbf{D}})) \tilde{\mathbf{r}}. && (10)
\end{aligned}$$

Key Aspects of the Proof II : Properties of a Quadratic Form

χ^2 distribution of Quadratic Forms $\mathbf{x}^T P\mathbf{x}$ for normal variables (Fisher-Cochran Theorem)

- Components x_i are independent normal variables $x_i \sim N(0,1)$, $i = 1 : n$.
- A necessary and sufficient condition that $\mathbf{x}^T P\mathbf{x}$ has a **central** χ^2 distribution is that P is **idempotent**, $P^2 = P$. In which case the degrees of freedom of χ^2 is rank(P) =trace$(P) = n$. .
- When the means of x_i are $\mu_i \neq 0$, $\mathbf{x}^T P\mathbf{x}$ has a **non-central** χ^2 distribution, with **non-centrality parameter** $c = \mu^T P\mu$
- A χ^2 random variable with n degrees of freedom and centrality parameter c has **mean** $n + c$ and **variance** $2(n + 2c)$.

Key Aspects of the Proof III: Requires the GSVD

Lemma

Assume invertibility and $m \geq n \geq p$. There exist unitary matrices $U \in \mathcal{R}^{m\times m}$, $V \in \mathcal{R}^{p\times p}$, and a nonsingular matrix $X \in \mathcal{R}^{n\times n}$ such that

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

$$\Upsilon = \mathrm{diag}(\upsilon_1, \dots, \upsilon_p, 1, \dots, 1) \in \mathcal{R}^{n\times n}, \quad M = \mathrm{diag}(\mu_1, \dots, \mu_p) \in \mathcal{R}^{p\times p},$$

$$\begin{gathered} 0 \leq \upsilon_1 \leq \cdots \leq \upsilon_p \leq 1, \quad 1 \geq \mu_1 \geq \cdots \geq \mu_p > 0, \\ \upsilon_i^2 + \mu_i^2 = 1, \quad i = 1, \dots p. \end{gathered} \tag{12}$$

The Functional with the GSVD

$$\begin{aligned} \text{Let} \quad \tilde{Q} &= \mathrm{diag}(\mu_1, \dots, \mu_p, 0_{n-p}, I_{m-n}) \\ \text{then} \quad J &= \tilde{\mathbf{r}}^T (I_m - A(W_{\mathbf{D}}))\tilde{\mathbf{r}} = \|\tilde{Q}U^T\tilde{\mathbf{r}}\|_2^2, \end{aligned}$$

Key Aspects of the Proof III: Requires the GSVD

Lemma

Assume invertibility and $m \geq n \geq p$. There exist unitary matrices $U \in \mathcal{R}^{m\times m}$, $V \in \mathcal{R}^{p\times p}$, and a nonsingular matrix $X \in \mathcal{R}^{n\times n}$ such that

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

$$\Upsilon = \mathrm{diag}(\upsilon_1, \ldots, \upsilon_p, 1, \ldots, 1) \in \mathcal{R}^{n\times n}, \quad M = \mathrm{diag}(\mu_1, \ldots, \mu_p) \in \mathcal{R}^{p\times p},$$

$$\begin{gathered} 0 \leq \upsilon_1 \leq \cdots \leq \upsilon_p \leq 1, \quad 1 \geq \mu_1 \geq \cdots \geq \mu_p > 0, \\ \upsilon_i^2 + \mu_i^2 = 1, \quad i = 1, \ldots p. \end{gathered} \tag{12}$$

The Functional with the GSVD

$$\begin{aligned} \text{Let} \quad \tilde{Q} &= \mathrm{diag}(\mu_1, \ldots, \mu_p, 0_{n-p}, I_{m-n}) \\ \text{then} \quad J &= \tilde{\mathbf{r}}^T (I_m - A(W_{\mathbf{D}}))\tilde{\mathbf{r}} = \|\tilde{Q} U^T \tilde{\mathbf{r}}\|_2^2, \end{aligned}$$

Key Aspects of the Proof IV: Statistical Distribution of the Weighted Residual

Covariance Structure

- $\mathbf{e} = A\mathbf{x} - \mathbf{b} \sim N(0, C_{\mathbf{b}})$ hence we can show $\mathbf{b} \sim N(A\mathbf{x}_0, C_{\mathbf{b}} + AC_{\mathbf{D}}A^T)$ **Note that b depends on x.**
- $\mathbf{r} \sim N(0, C_{\mathbf{b}} + AC_{\mathbf{D}}A^T)$, and $\tilde{\mathbf{r}} \sim N(0, I + \tilde{A}C_{\mathbf{D}}\tilde{A}^T)$, $\tilde{A} = {W_{\mathbf{b}}}^{1/2}A$.
- Use the GSVD

$$I + \tilde{A}C_{\mathbf{D}}\tilde{A}^T = UQ^{-2}U^T,$$

$$Q = \mathrm{diag}(\mu_1, \dots, \mu_p, I_{n-p}, I_{m-n})$$

The Functional is a rv

- Let $\mathbf{k} = QU^T\tilde{\mathbf{r}}$, then $\mathbf{k} \sim N(0, QU^T(UQ^{-2}U^T)UQ) \sim N(0, I_m)$
- But $J = \|\tilde{Q}U^T\tilde{\mathbf{r}}\|^2 = \|\tilde{\mathbf{k}}\|^2$, where $\tilde{\mathbf{k}}$ is the vector $\mathbf{k}$ excluding components $p+1:n$. Thus

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

Key Aspects of the Proof IV: Statistical Distribution of the Weighted Residual

Covariance Structure

- $\mathbf{e} = A\mathbf{x} - \mathbf{b} \sim N(0, C_{\mathbf{b}})$ hence we can show $\mathbf{b} \sim N(A\mathbf{x}_0, C_{\mathbf{b}} + AC_{\mathbf{D}}A^T)$ **Note that b depends on x**.
- $\mathbf{r} \sim N(0, C_{\mathbf{b}} + AC_{\mathbf{D}}A^T)$, and $\tilde{\mathbf{r}} \sim N(0, I + \tilde{A}C_{\mathbf{D}}\tilde{A}^T)$, $\tilde{A} = W_{\mathbf{b}}{}^{1/2}A$.
- Use the GSVD

$$I + \tilde{A}C_{\mathbf{D}}\tilde{A}^T = UQ^{-2}U^T,$$

$$Q = \mathrm{diag}(\mu_1, \ldots, \mu_p, I_{n-p}, I_{m-n})$$

The Functional is a rv

- Let $\mathbf{k} = QU^T\tilde{\mathbf{r}}$, then $\mathbf{k} \sim N(0, QU^T(UQ^{-2}U^T)UQ) \sim N(0, I_m)$
- But $J = \|\tilde{Q}U^T\tilde{\mathbf{r}}\|^2 = \|\tilde{\mathbf{k}}\|^2$, where $\tilde{\mathbf{k}}$ is the vector $\mathbf{k}$ excluding components $p+1:n$. Thus

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

Corollary: a-priori information not mean value, e.g. $\mathbf{x}_0 = 0$

Corollary: non-central χ^2 distribution of the regularized functional

$$\hat{\mathbf{x}} = \operatorname{argmin} J_D(\mathbf{x}) = \operatorname{argmin}\{\|A\mathbf{x} - \mathbf{b}\|^2_{W_{\mathbf{b}}} + \|(\mathbf{x} - \mathbf{x}_0)\|^2_{W_{\mathbf{D}}}\}, \quad W_{\mathbf{D}} = D^T W_{\mathbf{x}} D. \tag{13}$$

Assume all assumptions as before, but $\mathbf{x}_1 \neq \mathbf{x}_0$ is the mean vector of the model parameters.

Let

$$c = \|\mathbf{c}\|_2^2 = \|\tilde{Q} U^T W_{\mathbf{b}}{}^{1/2} A(\mathbf{x}_1 - \mathbf{x}_0)\|_2^2$$

Then

$$J_D \sim \chi^2(m + p - n, c)$$

$$E(J_D) = m + p - n + c \quad E(J_D J_D^T) = 2(m + p - n) + 4c$$

Requirements of the Theory

To apply the theory we require

- Covariance information $C_{\mathbf{b}}$ on data parameters $\mathbf{b}$ (or on model parameters $\mathbf{x}$!)
- A priori information either $\mathbf{x}_0$ is the mean, or mean value $\mathbf{x}_1$.
  - $\mathbf{x}_1$ and $\mathbf{x}_0$ are not known.
  - Assume $C_{\mathbf{b}}$ is calculated from measurement values. Then we can calculate $\mathbf{b}_1$ the mean of $\mathbf{b}$, and $E(\mathbf{b}) = AE(\mathbf{x})$ implies $\mathbf{b}_1 = A\mathbf{x}_1$. Hence

$$c = \|\mathbf{c}\|_2^2 = \|\tilde{Q}U^T W_{\mathbf{b}}{}^{1/2}(\mathbf{b}_1 - A\mathbf{x}_0)\|_2^2$$

  - $E(J_D) = E(\|\tilde{Q}U^T W_{\mathbf{b}}{}^{1/2}(\mathbf{b}-A\mathbf{x}_0)\|_2^2) = m+p-n+\|\tilde{Q}U^T W_{\mathbf{b}}{}^{1/2}(\mathbf{b}_1-A\mathbf{x}_0)\|_2^2$

Assume $\mathbf{x}_0$ is the mean

DESIGNING THE ALGORITHM: I

- If $C_{\mathbf{b}}$ and $C_{\mathbf{x}}$ are good estimates of the covariance matrices

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

  should be **small.**

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

$$\tilde{m} - \sqrt{2\tilde{m}} z_{\alpha/2} < \mathbf{r}^T W_{\mathbf{b}}{}^{1/2}(I_m - A(W_{\mathbf{D}}))W_{\mathbf{b}}{}^{1/2}\mathbf{r} < \tilde{m} + \sqrt{2\tilde{m}} z_{\alpha/2}. \quad (14)$$

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

GOAL

Find $W_{\mathbf{x}}$ to make (14) tight: Single Variable case find λ

$$J_D(\hat{\mathbf{x}}(\lambda)) \approx \tilde{m}$$

Assume $\mathbf{x}_0$ is the mean

DESIGNING THE ALGORITHM: I

- If $C_{\mathbf{b}}$ and $C_{\mathbf{x}}$ are good estimates of the covariance matrices

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

  should be **small.**

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

$$\tilde{m} - \sqrt{2\tilde{m}}z_{\alpha/2} < \mathbf{r}^T {W_{\mathbf{b}}}^{1/2}(I_m - A(W_{\mathbf{D}})){W_{\mathbf{b}}}^{1/2}\mathbf{r} < \tilde{m} + \sqrt{2\tilde{m}}z_{\alpha/2}. \quad (14)$$

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

GOAL

Find $W_{\mathbf{x}}$ to make (14) tight: Single Variable case find λ

$$J_D(\hat{\mathbf{x}}(\lambda)) \approx \tilde{m}$$

A Newton-line search Algorithm to find λ.

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

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

$$\mathbf{x}(\sigma^{(k)}) = \mathbf{y}(\sigma^{(k)}) + \mathbf{x}_0$$

  Then

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

- Hence we have a basic Newton Iteration

$$\sigma^{(k+1)} = \sigma^{(k)}(1 + \frac{1}{2}(\frac{\sigma^{(k)}}{\|D\mathbf{y}\|})^2(J_D(\sigma^{(k)}) - \tilde{m})).$$

Algorithm Using the GSVD

GSVD

- Use GSVD of $[{W_{\mathbf{b}}}^{1/2}A, D]$
- For γ_i the generalized singular values, and $\mathbf{s} = U^T {W_{\mathbf{b}}}^{1/2}\mathbf{r}$
- $\tilde{m} = m - n + p - \sum_{i=1}^{p} s_i^2 \delta_{\gamma_i 0} - \sum_{i=n+1}^{m} s_i^2$,
- $\tilde{s}_i = s_i/(\gamma_i^2 \sigma_{\mathbf{x}}^2 + 1)$, $i = 1, \ldots, p \quad t_i = \tilde{s}_i \gamma_i$.
- Find root of

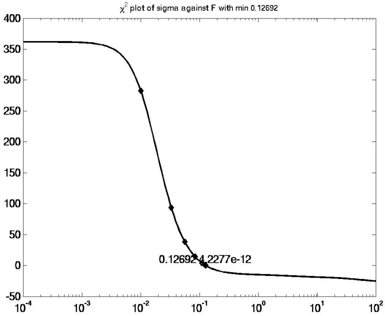
$$F(\sigma_{\mathbf{x}}) = \sum_{i=1}^{p} (\frac{1}{\gamma_i^2 \sigma_{\mathbf{x}}^2 + 1}) s_i^2 + \sum_{i=n+1}^{m} s_i^2 - \tilde{m} = 0$$

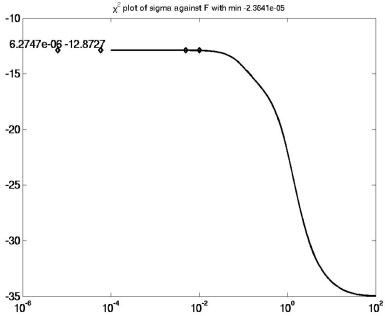
- Equivalently solve $F = 0$, where

$$F(\sigma_{\mathbf{x}}) = \mathbf{s}^T \tilde{\mathbf{s}} - \tilde{m} \quad \text{and } F'(\sigma_{\mathbf{x}}) = -2\sigma_{\mathbf{x}} \|\mathbf{t}\|_2^2.$$

Discussion on Convergence

- F is **monotonic decreasing** $(F'(\sigma_{\mathbf{x}}) = -2\sigma_{\mathbf{x}}\|\mathbf{t}\|_2^2)$
- Solution either exists and is **unique** for positive σ **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

Find the parameter

- **Step 1:** Bracket the root by logarithmic search on σ to handle the asymptotes: yields sigmamax and sigmamin
- **Step 2:** Calculate step, with steepness controlled by tolD. Let $\mathbf{t} = D\mathbf{y}/\sigma^{(k)}$, where $\mathbf{y}$ is the current update, given from the GSVD, then

$$\text{step} = \frac{1}{2}\left(\frac{1}{\max\{\|\mathbf{t}\|, \text{tolD}\}}\right)^2 (J_D(\sigma^{(k)}) - \tilde{m})$$

- **Step 3:** Introduce line search $\alpha^{(k)}$ in Newton

$$\text{sigmanew} = \sigma^{(k)}(1 + \alpha^{(k)}\text{step})$$

$\alpha^{(k)}$ chosen such that sigmanew within bracket.

Implementation Assumptions

Covariance of Error: Statistics of Measurement Errors

- Information on the covariance structure of errors in **b** needed.
- Use $C_{\mathbf{b}} = \sigma_{\mathbf{b}}^2 I$ for common covariance, white noise.
- Use $C_{\mathbf{b}} = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2)$ for colored uncorrelated noise.
- With no noise information $C_{\mathbf{b}} = I$.
- Use $\mathbf{b}_1$ as the mean of measured **b**, when implemented as central case.

Tolerance on Convergence

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

Implementation Assumptions

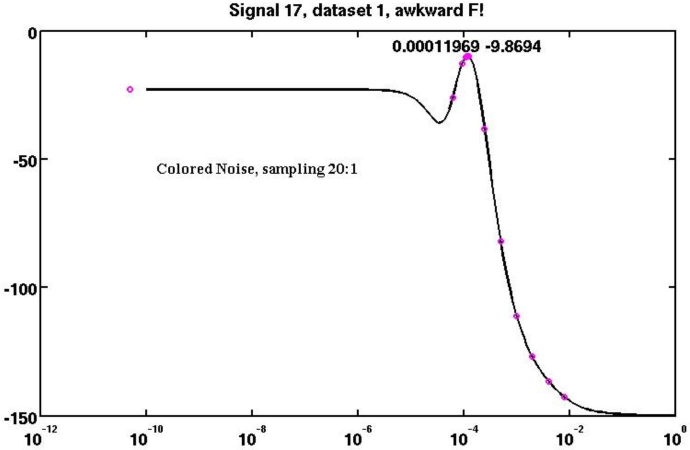
Covariance of Error: Statistics of Measurement Errors

- Information on the covariance structure of errors in **b** needed.
- Use $C_{\mathbf{b}} = \sigma_{\mathbf{b}}^2 I$ for common covariance, white noise.
- Use $C_{\mathbf{b}} = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2)$ for colored uncorrelated noise.
- With no noise information $C_{\mathbf{b}} = I$.
- Use $\mathbf{b}_1$ as the mean of measured **b**, when implemented as central case.

Tolerance on Convergence

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

Difficulties with the Central Case



Functional with centrality parameter need not be monotonic: modify algorithm to solve min $F(\sigma)^2$.

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 known

- 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? Compare with UPRE and L-Curve.

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.

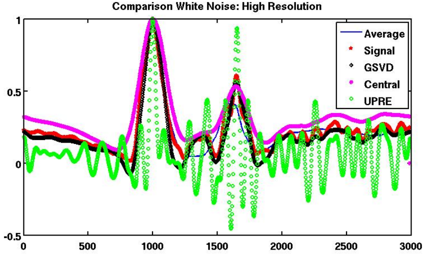
Method of Comparison- no exact solution known

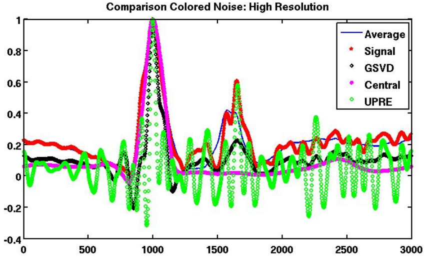
- No exact solution.
- Downsample the signal and restore for different resolutions

Resolution	$2:1$	$5:1$	$10:1$	$20:1$	$100:1$
Points	1500	600	300	150	30

- Do results converge? Compare with UPRE and L-Curve.

Comparison High Resolution White noise (left) and Colored Noise (right)





Greater contrast with χ^2 . UPRE is insufficiently regularized.
L-curve severely undersmooths (not shown). Parameters not consistent across resolutions

Conclusions

Conclusions

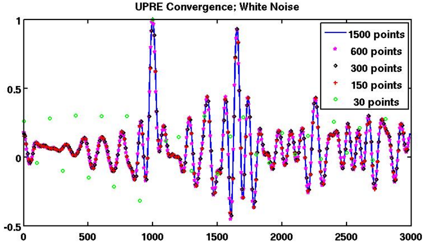
- A new statistical method for estimating regularization parameter
  - Compares favorably with UPRE with respect to performance
- Method can be used for large scale problems, without GSVD (not shown)
- Method is very efficient, Newton method is robust and fast.
- $\mathbf{x}_0$ is the mean.
  - More problematic for Central version with $\mathbf{x}_0$ not the mean.
  - σ can be bounded by result of non-central case.
  - Range of σ given by range of γ_i.
  - Appears to oversmooth the solution.
  - Function need not be monotonic

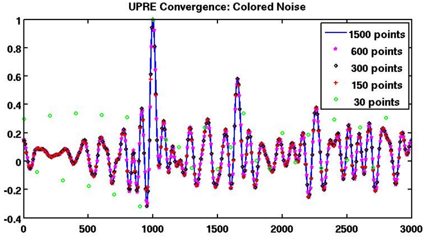
Future Work

Other Results and Future Work

- Degrees of freedom reduced when using the GSVD.
- How to apply Picard condition for GSVD to handle problems with robustness due to conditioning of $C_{\mathbf{b}}$
- Image deblurring. (Implementation to use minimal storage)
- Diagonal Weighting Schemes
- Edge preserving regularization
- Constraint implementation (with Mead).

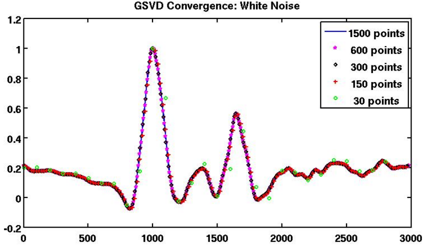
THE UPRE SOLUTION: White Noise and Colored Noise $\mathbf{x}_0 = 0$

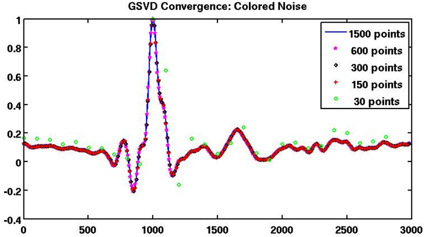




Regularization Parameters are consistent: $\sigma = 0.01005$ all resolutions

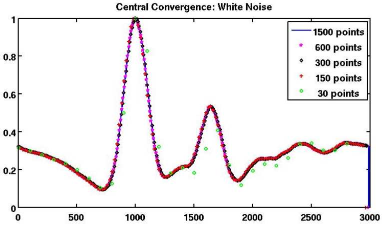
THE GSVD SOLUTION: White Noise (left) and Colored Noise (right) $\mathbf{x}_0 = 0$

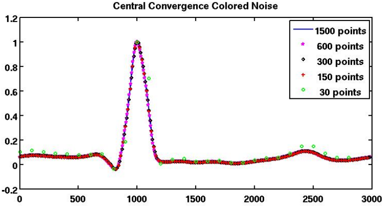




Regularization Parameters are consistent:
$\sigma = 0.00058$ (left), $\sigma = 0.00069$ (right) all resolutions

THE CENTRAL GSVD SOLUTION: White Noise (left) and Colored Noise (right) $\mathbf{x}_0 = 0$





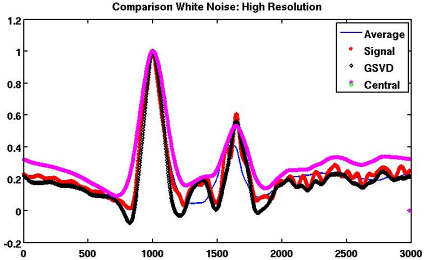
Regularization Parameters less smoothing for low resolution

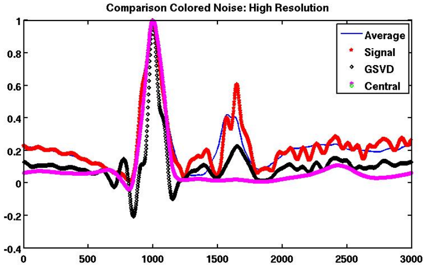
$\sigma = 0.0000029, .0000029, .0000029, .0000057, .0000057$ (left)

$\sigma = 0.00007, .00007, .00007, .00007, .00012$ (right).

, resolution 2 to 100

Comparison White noise (left) and Colored Noise (right) GSVD and Central GSVD





Non-central shows existence of secondary signal for colored noise
central scheme is oversmoothed, but white noise shows major second arrival