Towards Solution of Large Scale Signal Restoration Problems with Multi-Parameter Estimates

Rosemary Renaut
Joint work with Jodi Mead, Boise State University
Anne Gelb, Aditya Viswanathan, Hongbin Guo, Doug Cochran, Youzuo Lin,
Arizona State University
Wolfgang Stefan now at Rice University

ARIZONA STATE UNIVERSITY

December 7, 2009
Outline

1 Motivation
   - Quick Review
   - Statistical Results for Least Squares
     - Summary of LS Statistical Results
     - Implications of Statistical Results for Regularized Least Squares

2 Newton algorithm
   - Algorithm with LSQR (Paige and Saunders)
   - Results

3 Conclusions and Future Work

4 Other Related Developments in progress
   - Large Scale Problems: Domain Decomposition
     - Application in Image Restoration
   - Variable Order Total Variation Regularization Using Edge Detection - No Noise
   - Edge Detection for PSF Estimation - With Noise
**Signal/Image Restoration:**

### Integral Model of Signal Degradation

\[ b(t) = \int K(t, s)x(s)ds \]

- \( K(t, s) \) describes blur of the signal.
- Convolutional model: invariant \( K(t, s) = K(t - s) \) is Point Spread Function (PSF).
- Typically sampling includes noise \( e(t) \), model is

\[ b(t) = \int K(t - s)x(s)ds + e(t) \]

### Discrete model: given discrete samples \( b \), find samples \( x \) of \( x \)

- Let \( A \) discretize \( K \), assume known, model is given by

\[ b = Ax + e. \]

- Naïvely invert (or use pseudoinverse) the system to find \( x \)!
Example 1-D Original and Blurred Noisy Signal

Original signal $x$.

Blurred and noisy signal $b$.

Gaussian PSF.
The Solution: Regularization is needed

Naïve Solution

A Regularized Solution
Least Squares for $Ax = b$: A Quick Review

Background

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

- **Classical Approach** Linear Least Squares ($A$ full rank)
  
  $$x_{LS} = \arg \min_x ||Ax - b||_2^2$$

- **Difficulty** $x_{LS}$ sensitive to changes in right hand side $b$ when $A$ is ill-conditioned.

  For convolutional models system is ill-posed.
Introduce Regularization to Find *Acceptable Solution*

**Weighted Fidelity with Regularization**

- Regularize
  \[ x_{RLS}(\lambda) = \arg \min_x \{ ||b - Ax||_{W_b}^2 + \lambda^2 R(x) \}, \]
- Weighting matrix \( W_b \)
- \( R(x) \) is a regularization term
- \( \lambda \) is a regularization parameter which is unknown.
- Solution \( x_{RLS}(\lambda) \)
  - depends on \( \lambda \).
  - depends on regularization operator \( R \)
  - depends on the weighting matrix \( W_b \)
The Weighting Matrix: $W_b$

Some Assumptions for Multiple Data Measurements

Given multiple measurements of data $b$:

- Usually error in $b$, $e$ is an $m -$ vector of **random measurement errors** with mean 0 and **positive definite covariance** matrix $C_b = \mathbf{E}(ee^T)$.
- For **uncorrelated heteroskedastic** measurements $C_b$ is **diagonal** matrix of **standard deviations** of the errors. (Colored noise)
- For **white noise** $C_b = \sigma^2 I$.
- Weighting by $W_b = C_b^{-1}$ in data fit term, theoretically, $\tilde{e} = W_b^{1/2}e$ are uncorrelated.
- Difficulty if $W_b$ increases ill-conditioning of $A$!
- For images find $W_b$ from the image data
Formulation: Generalized Tikhonov Regularization With Weighting

Use $R(x) = \|D(x - x_0)\|^2_{W_x}$

$$\hat{x} = \text{argmin } J(x) = \text{argmin}\{\|Ax - b\|^2_{W_b} + \|D(x - x_0)\|^2_{W_x}\}. \quad (1)$$

- $D$ is a suitable operator, often derivative approximation.
- Assume $\mathcal{N}(W_b^{1/2}A) \cap \mathcal{N}(W_x^{1/2}D) = \{0\}$
- Regularized solution given in terms of regularized inverse matrix $R(W_D)$
  $$\hat{x} = x_0 + (A^T W_b A + D^T W_x D)^{-1} A^T W_b r, \quad (2)\quad \hat{x} = x_0 + R(W_D) W_b^{1/2} r, \quad r = b - Ax_0, \quad W_D = D^T W_x D$$
  $$= x_0 + y(W_D). \quad (3)$$
  $$R(W_D) = (A^T W_b A + D^T W_x D)^{-1} A^T W_b^{1/2} \quad (4)$$

- $x_0$ is a reference solution, often $x_0 = 0$, might need to be average solution.
- Having found $W_x$, the posterior inverse covariance matrix is
  $$\tilde{W}_x = A^T W_b A + W_D$$

- Posterior information can give some confidence on parameter estimates.
Typically $W_x = \lambda^2 I$ and choice of $\lambda$ is crucial: an example with $D = I$. 
Typically $W_\lambda = \lambda^2 I$ and choice of $\lambda$ is crucial: an example with $D = I$. 
Typically $W_x = \lambda^2 I$ and choice of $\lambda$ is crucial: an example with $D = I$. 
Typically $W_x = \lambda^2 I$ and choice of $\lambda$ is crucial: an example with $D = I$. 
Typically $W_x = \lambda^2 I$ and choice of $\lambda$ is crucial: an example with $D = I$. 
Typically $W_x = \lambda^2 I$ and choice of $\lambda$ is crucial: an example with $D = I$. 
Typically $W_x = \lambda^2 I$ and choice of $\lambda$ is crucial: an example with $D = I$. 
Typically $W_x = \lambda^2 I$ and choice of $\lambda$ is crucial: an example with $D = I$. 
Typically $W_x = \lambda^2 I$ and choice of $\lambda$ is crucial: an example with $D = I$. 
Typically $W_x = \lambda^2 I$ and choice of $\lambda$ is crucial: an example with $D = I$. 

![Graph](image-url)
Typically $W_x = \lambda^2 I$ and choice of $\lambda$ is crucial: an example with $D = I$. 
Typically $W_x = \lambda^2 I$ and choice of $\lambda$ is crucial: an example with $D = I$. 
Typically $W_x = \lambda^2 I$ and choice of $\lambda$ is crucial: an example with $D = I$. 
Solution in terms of the GSVD - see for example Hansen for all details

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

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

where

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

and

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

$$\upsilon_i^2 + \mu_i^2 = 1, \quad i = 1, \ldots p.$$  \hspace{1cm} (6)

The Solution with the GSVD

- Let $\tilde{P} = \text{diag}(\mu_1, \ldots, \mu_p, I_{n-p}, O_{m-n})$ and $\tilde{r} = W_b^{1/2}r$. then
  $$y(W_D) = (X^T)^{-1} \tilde{P} U^T \tilde{r}$$

- The GSVD is for matrix pair $[W_b^{1/2} A, W_x^{1/2} D]$.

- Solution depends on $W_D$ through the $\mu_i$, $X$ and $U$. 
Choice of $\lambda$ crucial: Different algorithms - different solutions.

**Discrepancy Principle**

- Suppose noise is white: $C_b = \sigma_b^2 I$.
- Find $\lambda$ such that the regularized residual satisfies

$$\sigma_b^2 = \frac{1}{m} \| b - Ax(\lambda) \|^2_2.$$

- Can be implemented by a Newton root finding algorithm.
- But discrepancy principle typically oversmooths.

**Others (Vogel, Hansen)**

- L-Curve
- Generalized Cross Validation (GCV)
- Unbiased Predictive Risk (UPRE)
**Some standard approaches I: L-curve - *Find the corner***

- Introduce the **influence** matrix
  \[ A(\lambda) W_b^{1/2} A(A^T W_b A + \lambda^2 D^T D)^{-1} A^T W_b^{1/2} \]

- Weighted residual is
  \[ r(\lambda) = (I_m - A(\lambda)) W_b^{1/2} r \]

- Plot
  \[ \log(\|Dx\|), \log(\|r(\lambda)\|) \]

  Trade off contributions.

- **Expensive** - requires range of \( \lambda \).

- GSVD makes calculations **efficient**.

- Not statistically based
Generalized Cross-Validation (GCV)

- Minimize GCV function (statistically based on leave one out analysis)
- Let $\tilde{r} = W_b^{1/2}r$

$$\frac{\| (r - Ay(\lambda)) \|^2_{W_b}}{\text{trace}(I_m - A(\lambda))^2} = \frac{\| (I_m - A(\lambda))\tilde{r} \|^2}{\text{trace}(I_m - A(\lambda))^2},$$

which estimates predictive risk.

- Expensive - requires range of $\lambda$.
- GSVD makes calculations efficient.
- Requires minimum

Multiple minima

Sometimes flat
Unbiased Predictive Risk Estimation (UPRE)

- Minimize expected value of predictive risk: Minimize UPRE function
  \[
  \| \mathbf{r} - A\mathbf{y}(\lambda) \|_{W_b}^2 + 2 \text{trace}(A(\lambda)) - m
  = \|(I_m - A(\lambda))\tilde{\mathbf{r}}\|^2 - 2 \text{trace}(I_m - A(\lambda)) + m
  \]

  - Expensive - requires range of $\lambda$.
  - GSVD makes calculations efficient.
  - Need estimate of trace
  - Minimum needed


**Background: Statistics of the Least Squares Problem**

**Theorem (χ² distribution of the Residual - well known)**

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

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

J follows a $\chi^2$ distribution with $m - \rho$ degrees of freedom:

*Basically the Discrepancy Principle*

**Corollary (Weighted Least Squares)**

For $\mathbf{b} \sim N(A\mathbf{x}, C_b)$, $W_b = C_b^{-1}$ then

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

A rule of thumb is that a *typical* value of the residual for a *moderately* good fit is $J \approx m - \rho$. More precisely is the statement when regarding $J$ as a statistic is that it has a mean $m - \rho$ and a standard deviation $\sqrt{2(m - \rho)}$, and, asymptotically for large $m - \rho$, $J$ becomes normally distributed even when data errors are not normally distributed.
Extension: Statistics of the Regularized Least Squares Problem

**Thm:** $\chi^2$ distribution of the regularized functional [MR09b] (Weighted Regularization)

\[
\hat{x} = \arg\min J_D(x) = \arg\min \left\{ \|Ax - b\|_{W_b}^2 + \|(x - x_0)\|_{W_D}^2 \right\}, \quad W_D = D^T W_x D. \quad (7)
\]

Assume
- $W_b$ and $W_x$ are symmetric positive definite and $\mathcal{N}(W_b^{1/2} A) \cap \mathcal{N}(W_x^{1/2} D) = \{0\}$.
- Moore-Penrose generalized inverse of $W_D$ is $C_D$.
- Statistics: Errors in the right hand side $e \sim N(0, C_b)$, and $x_0$ is the known mean so that $(x - x_0) = f \sim N(0, C_D)$.

Then

\[
J_D(\hat{x}(W_D)) \sim \chi^2(m + p - n)
\]

**Significance**

For sufficiently large $\tilde{m} = m + p - n$, $E(J(x(W_D)))) = m + p - n$, $E(JJ^T) = 2(m + p - n)$

Moreover

\[
\tilde{m} - \sqrt{2\tilde{m}} z_{\alpha/2} < J(\hat{x}(W_D)) < \tilde{m} + \sqrt{2\tilde{m}} z_{\alpha/2}. \quad (8)
\]

$z_{\alpha/2}$ is the relevant $z$-value for a $\chi^2$-distribution with $\tilde{m} = m + p - n$ degrees.
Key Aspects of the Proof

Algebraic Simplifications: Rewrite functional as quadratic form

- Functional in terms of $A(W_D) J = \tilde{r}^T(I_m - A(W_D))\tilde{r}$ A Quadratic Form
- Using GSVD and defining $\tilde{Q} = \text{diag}(\mu_1, \ldots, \mu_p, 0_{n-p}, I_{m-n}) J = \|\tilde{Q}U^T \tilde{r}\|_2^2 = \|k\|_2^2$

$\chi^2$ distribution of Quadratic Forms $x^T P x$ 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 $x^T P x$ has a central $\chi^2$ distribution is that $P$ is idempotent, $P^2 = P$. In which case the degrees of freedom of $\chi^2$ is rank($P$) = trace($P$) = $n$.
- When the means of $x_i$ are $\mu_i \neq 0$, $x^T P x$ has a non-central $\chi^2$ distribution, with non-centrality parameter $c = \mu^T P \mu$
- A $\chi^2$ random variable with $n$ degrees of freedom and centrality parameter $c$ has mean $n + c$ and variance $2(n + 2c)$. 
Proof: use statistics of the data and the quadratic form

**Covariance Structure**

- Errors in \( b \) are \( e \sim N(0, C_b) \). Now \( b \) depends on \( x \), \( b = Ax \) hence we can show \( b \sim N(Ax_0, C_b + AC_DA^T) \) (\( x_0 \) is mean of \( x \))
- Residual \( r = b - Ax \sim N(0, C_b + AC_DA^T) \).
- \( \tilde{r} = W_b^{1/2}r \sim N(0, I + \tilde{A}C_D\tilde{A}^T) \), \( \tilde{A} = W_b^{1/2}A \).
- Use the GSVD
  
  \[
  I + \tilde{A}C_D\tilde{A}^T = UQ^{-2}U^T, \quad Q = \text{diag}(\mu_1, \ldots, \mu_p, I_{n-p}, I_{m-n})
  \]
- Now \( k = QU^T\tilde{r} \) then \( k \sim N(0, QU^T(UQ^{-2}U^T)UQ) \sim N(0, I_m) \)
- But \( J = \|\tilde{Q}U^T\tilde{r}\|^2 = \|\tilde{k}\|^2 \), where \( \tilde{k} \) is the vector \( k \) excluding components \( p + 1 : n \). Thus
  
  \[
  J_D \sim \chi^2(m + p - n).
  \]
When mean of the parameters is not known, or $x_0 = 0$ is not the mean

**Corollary: non-central $\chi^2$ distribution of the regularized functional**

Recall

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

Assume all assumptions as before, but $\bar{x} \neq x_0$ is the mean vector of the model parameters. Let

$$c = \|c\|^2_2 = \|\tilde{Q} U^T W_b^{1/2} A(\bar{x} - x_0)\|^2_2$$

Then

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

The functional at optimum follows a non-central $\chi^2$ distribution

**The Cost Functional follows a $\chi^2$ Statistical Distribution**

- Suppose degrees of freedom $\tilde{m}$ and centrality parameter $c$ then

  $$E(J_D) = \tilde{m} + c \quad E(J_D J_D^T) = 2(\tilde{m}) + 4c$$

- Suggests: Try to find $W_D$ so that $E(J) = \tilde{m} + c$

- First find $\lambda$ only. $W_x = \lambda^2 I$
What do we need to apply the Theory?

**Requirements**

- **Covariance** $C_b$ on data parameters $b$ (or on model parameters $x$!)
- **A priori** information $x_0$, mean $\bar{x}$.
- But $\bar{x}$ (and hence $x_0$) are not known.
- If not known use repeated data measurements calculate $C_b$ and mean $\bar{b}$.
- Hence estimate the **centrality** parameter $E(b) = AE(x)$ implies $\bar{b} = A\bar{x}$. Hence

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

- $E(J_D) = E(\|\tilde{Q}U^T W_b^{1/2} (b - Ax_0)\|_2^2) = m + p - n + \|c\|_2^2$
- Given the GSVD estimate the degrees of freedom $\tilde{m}$.

Then we can use $E(J)$ to find $\lambda$
Assume $x_0$ is the mean (experimentalists know something about model parameters)

**DESIGNING THE ALGORITHM: I**

- Recall: if $C_b$ and $C_x$ are good estimates of covariance

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

  should be small.

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

  $$\tilde{m} - \sqrt{2\tilde{m}z_{\alpha/2}} < J(x(W_D)) < \tilde{m} + \sqrt{2\tilde{m}z_{\alpha/2}}.$$  

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

**GOAL**

Find $W_x$ to make (8) tight: Single Variable case find $\lambda$

$$J_D(\hat{x}(\lambda)) \approx \tilde{m}$$
A Newton-line search Algorithm to find $\lambda = 1/\sigma$. (Basic algebra)

**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

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

- We can use the GSVD easily for small problems let $s = U^T W_b^{1/2} r$

- Find root of

$$F(\sigma) = \sum_{i=1}^{p} \left( \frac{1}{\gamma_i^2 \sigma_i^2 + 1} \right) s_i^2 + \sum_{i=n+1}^{m} s_i^2 - \tilde{m} = 0$$

- Take care with the line search.
Discussion on Convergence

- \( F \) is **monotonic decreasing** \( (F'(\sigma_x) = -2\sigma_x \|Dy\|_2^2) \)
- 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.
**Algorithm**

**Initialization**
- Convert generalized Tikhonov problem to standard form. (if $L$ is not invertible you just need to know how to find $Ax$ and $A^T x$, and the null space of $L$)
- Use LSQR (Paige and Saunders) algorithm to find the bidiagonal matrix for the projected problem.
- Obtain a solution of the bidiagonal problem for given initial $\sigma$.

**Subsequent Steps**
- Increase dimension of space if needed with reuse of existing bidiagonalization. May also use smaller size system if appropriate.
- Each $\sigma$ calculation of algorithm reuses information from Golub-Kahan bidiagonalization.

**Advantages : Costs**
- Needs only cost of standard LSQR algorithm with some updates for solution solves for iterated $\sigma$.
- The regularization introduced by LSQR projection may be useful for preventing problems with GSVD expansion.
- Makes algorithm viable for large scale problems.
Illustrating the Results for Problem Size 512: Two Standard Test Problems

Comparison for noise level 10%. On left $D = I$ and on right $D$ is first derivative

- Notice L-curve and $\chi^2$-LSQR (here denoted by LSQR) perform well.
- UPRE does not perform well.
- Results are illustrative of statistical testing over many experiments.
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 $x$ from $Ax = b$, where $A$ is PSF matrix and $b$ is given blurred signal.
- Method of Comparison- no exact solution known: use convergence with respect to downsampling.
Comparison High Resolution White noise

Greater contrast with $\chi^2$. UPRE is insufficiently regularized. L-curve severely undersmooths (not shown). Parameters not consistent across resolutions. Here GSVD denotes $\chi^2$ algorithm with the GSVD and central denotes $\chi^2$ approach for central distribution.
THE UPRE SOLUTION: $x_0 = 0$ White Noise

Regularization Parameters are consistent: $\sigma = 0.01005$ all resolutions
Regularization quite consistent resolution 2 to 100

\[ \sigma = 0.0000029, 0.0000029, 0.0000029, 0.0000057, 0.0000057 \]

This result also assumes \( x_0 \neq 0 \) and requires \( c \) which is estimated from average signals.
Illustrating the Deblurring Result: Problem Size 65536

Example taken from RESTORE TOOLS Nagy et al 2007-8: 15% Noise

Computational Cost is Minimal: Projected Problem Size is $15, \lambda = .58$
Problem Grain noise 15% added: increasing subproblem size to validate against increasing subproblem size

(a) Signal to noise ratio $10 \log_{10} \left( \frac{1}{e} \right)$ relative error $e$

(b) Regularization Parameter Against Problem Size
Problem Grain noise 15% added for increasing subproblem size

Figure: Signal to noise ratio $10 \log_{10}(1/e)$ relative error e
Problem Grain noise 15% added using central algorithm

![SNR 2.8396](image1)
(a) Blurred

![SNR 2.8595](image2)
(b) LSQR Size 25

![SNR 3.4899 ± 0.5](image3)
(c) Hybrid Central

**Figure:** Modifying the central algorithm with hybrid which is more robust than using the zero background solution
Future Work Combining Approaches

- Extend the parameter selection methods to the domain decomposition problems for large scale.
- Use efficient schemes for large scale problems - eg right hand side updates
- Extend to edge detection approaches
- Use tensor product of the PSF for extension to 2D - is it feasible
- Use parameter estimation techniques for the 2D problem
- Further development of statistical techniques for estimating acceptable solutions.
- Work is described in a number of papers [MR09a], [MR09b], [RHM09], [RLG09], [SGR09], [CR09]
An Alternative Direction For Large Scale Problems: Domain Decomposition (Renaut, Lin and Guo)

- Domain decomposition of $\mathbf{x}$ into several domains:
  \[
  \mathbf{x} = (\mathbf{x}_1^T, \mathbf{x}_2^T, \ldots, \mathbf{x}_p^T)^T.
  \]

- Corresponding to different splitting of image $\mathbf{x}$, kernel operator $A$ is split
  \[
  A = (A_1, A_2, \ldots, A_p).
  \]

- The linear system $A\mathbf{x} \approx \mathbf{b}$ is replaced with the split systems
  \[
  A_i\mathbf{y}_i \approx \mathbf{b}_i(\mathbf{x}), \quad \mathbf{b}_i(\mathbf{x}) = \mathbf{b} - \sum_{j \neq i} A_j \mathbf{x}_j = \mathbf{b} - A\mathbf{x} + A_i \mathbf{x}_i.
  \]

- Locally solve $A\mathbf{x} \approx \mathbf{b}$
  \[
  \min_{\mathbf{y}_i \in \mathbb{R}^{n_i}} \|A_i\mathbf{y}_i - \mathbf{b}_i(\mathbf{x})\|_2, \quad 1 \leq i \leq p.
  \]

- If the problem is ill-posed we have the regularized problem and apply similar splitting
  \[
  \min_{\mathbf{x}} \left\{ \|A\mathbf{x} - \mathbf{b}\|_2^2 + \|\Lambda D\mathbf{x}\|_2^2 \right\}.
  \]
Feasibility Signal Restoration Variable Noise

Figure: (a): Degraded 1-D signal, length 1024, Gaussian kernel with variance of 144, different levels of the Gaussian noise are added to the first and second portions of the signal; (b): Restored 1D signal using optimal global $\lambda$; (c): Restored 1D signal. Local $\lambda$ values

Will need to obtain information on how to do problem decomposition, find edges, and large scale
Implementation Details: Using the Krylov subspace update effectively

- LSQR to solve the local problems - bidiagonalization is based on the local matrix $A_i$
- Complication: right hand side $b_i$ changes each iteration - so require Krylov update.
- Instead use a CGLS-based algorithm with a seeded Krylov subspace.
- Solve with seed system, initial $b_i$. Then project for later steps onto seeded Krylov subspace.
- Algorithm is improved also by adding new CG directions to the CGLS.

![Convergence plots](image)

(a) Convergence: Canonical CG  (b) Convergence: Seeded System

**Figure:** Outer and inner iteration plot for reconstruction of Shepp-Logan phantom of size $64 \times 64$, noise variance 0.015 and mean 0. Horizontal axis denotes the outer (or global) iteration steps, and the vertical axis denotes the inner iteration steps. (a): canonical CG as solver for the inner loop; (b): projected CG with updates. The four different line styles in each plot stand for four subproblems, specifically, “–o–” is subproblem 1, “–*–” is subproblem 2, “–|–” is subproblem 3 and “–△–” is subproblem 4.
Figure: In (a) the piecewise smooth test function is contaminated by blurring with an out of focus PSF of 12 pixels width. (b) shows the result of the TV restoration with $\lambda = 10^{-2}$, which yields an $l^2$ error of .034. (c) shows the VOTV restoration with an $l^2$ error of .023. (d) estimation of jump function.
Blind Deconvolution using Edge Detection (Cochran, Gelb, Viswanathan, Renaut)

Given the blurring model (PSF convolution operator $K$) and $x \in L^2(-\pi, \pi)$ piecewise-smooth. We estimate the psf starting with $2N + 1$ blurred Fourier coefficients $\hat{b}(j), j = -N, \ldots, N$.

$$b = K \ast x + e$$

Principle:

Apply a linear edge detector, denote by $T$. We shall assume that the edge detector can be written as a convolution with an appropriate kernel

$$T \ast (K \ast x + e) = (K \ast x + e) \ast T$$
$$= x \ast K \ast T + e \ast T$$
$$= (x \ast T) \ast K + e \ast T$$
$$\approx [x] \ast K + \tilde{e}$$

Here $[x](s)$ is a jump function. For a jump discontinuity in a function the jump function at any point $s$ only depends on the values of $x$ at $s^+$ and $s^-$. 

$$[x](s) := x(s^+) - x(s^-)$$

Hence, we observe shifted and scaled replicates of the psf.
Example (No Noise)

Figure: Function subjected to motion blur, $N = 128$
Representative Examples: Gaussian PSF

(a) Noisy blur estimation

(b) After low-pass filtering

Figure: Function subjected to Gaussian blur, $N = 128$

- Complex noise distribution on Fourier coefficients – $\hat{e} \sim \mathcal{N}(0, \frac{1.5}{(2N+1)^2})$
- Second picture subjected to low-pass (Gaussian) filtering
- It is conceivable that parameter estimation for a Gaussian PSF can take into account the effect of Gaussian filtering
Representative Examples: Motion Blur

Figure: Function subjected to Motion blur, $N = 128$

- Cannot perform conventional low-pass filtering since blur is piecewise-smooth
- We compute the noisy blur estimate for Fourier expansion of blurred jump
  $S_N[b] \approx [x] \ast K + \tilde{e}$
- Denoising problem formulation
  $$\min_x \| x - S_N[b] \|^2_2 + \lambda^2 \| Dx \|_1.$$
Bibliography I

Aditya Viswanathan Anne Gelb Doug Cochran and Rosemary A Renaut.
On reconstruction from non-uniform spectral data.

J. Mead and R. A. Renaut.
Least squares problems with inequality constraints as quadratic constraints.

J Mead and R. A. Renaut.
Newton root-finding algorithm for estimating the regularization parameter for solving ill-conditioned least squares problems.

R. A. Renaut, I. Hnetynkova, and J. Mead.
Regularization parameter estimation for large scale Tikhonov regularization using a priori information.
*Computational Statistics and Data Analysis*, 54(1), 2009.

R. A. Renaut, Youzuo Lin, and H. Guo.
Multisplitting for regularized least squares.
*Numerical Linear Algebra with applications*, 2009.

Wolfgang Stefan, Anne Gelb, and Rosemary A Renaut.
Improved total variation-type regularization using higher-order edge detectors.