Confidence in Image Reconstruction
June 2011
IMA
Dianne P. O’Leary
©2011
Confidence in Image Reconstruction

Dianne P. O’Leary

Computer Science Dept. and Institute for Advanced Computer Studies
University of Maryland
oleary@cs.umd.edu

http://www.cs.umd.edu/users/oleary

Work supported by NIST and NSF.
Includes joint work with

Julianne Chung  Matthias Chung  Glenn Easley.

Builds upon joint work with

Bert Rust  James Nagy.
The Plan

The talk is organized around 4 case studies that we consider in turn.

- Quantitative diagnostics for solution quality → Case Study 1
- Confidence intervals for individual pixels of an image
  - General blurs, but especially 1D blurs → Case Study 2a and Case Study 2b
  - Given calibration data → Case Study 3
Case Study 1

Suppose you are a planetary scientist presented with these possible versions of an image taken by a spacecraft:

Which is “correct” and how confident are we of the correctness?
Case Study 2a and 2b

Given: An estimate of a 1-d blurring function, and an estimate of the SNR,
Deblur an image and give statistical confidence intervals on each pixel value.
Case Study 3

**Given:** A set of pairs of training images with noisy blurred images and an estimated blurring operator,

Deblur a new noisy blurred image and give statistical confidence intervals on each pixel value.

Note that we do not know any properties of the noise.
Basic facts and notation
The model

We are given

- a blurred picture \( Y \), specified by \( m \) pixel values.
- information about a blurring function \( k \).
- information about noise in the measurement of \( Y \).

Our model:

\[
y = Kx^* + \epsilon,
\]

Where

- \( y = \text{vec}(Y) \) is the \( m \)-vector of measurements.
- \( K \) is a known \( m \times n \) matrix derived from \( k \), with \( m \geq n \).
- \( x^* \) is an unknown \( n \)-vector whose components are “true” pixel values for the image (or a subimage).
The assumption

The vector $\mathbf{e}$ is an $m$-vector of random measuring errors satisfying

$$\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \mathbf{S}^2).$$

where $\mathcal{E}$ is the expectation operator and $\mathbf{S}^2$ is the positive definite variance matrix.

- The variances can be estimated by repeatedly taking a picture of a uniform surface.
- An analyst who fails to use this information implicitly assumes that $\mathbf{S}^2 = s^2 \mathbf{I}_m$ where $\mathbf{I}_m$ is the $m$-th order identity matrix and $s$ is an unknown scalar that can be, but usually is not, estimated from the sum of squared residuals for the least squares solution.
- Using variance information can greatly improve estimates of $\mathbf{x}$. 
Warning!

- Examples in this talk are engineered to exactly satisfy the assumption.
- We don’t have that luxury in real life, so results shown here are better than we should expect!
Scaling the model

**Original** linear regression model:

\[ y = Kx^* + \epsilon, \quad \epsilon \sim N(0, S^2). \]

**Scaled** linear regression model:

\[ b = Ax + \eta, \]

where

\[ b \equiv S^{-1}y, \quad A \equiv S^{-1}K, \quad \eta \equiv S^{-1}\epsilon. \]

Then the scaled model can be written

\[ b = Ax^* + \eta, \quad \eta \sim N(0, I_m). \]

As a consequence,

\[ \|\eta\|^2 \sim \chi^2(m). \]
Ill-Posedness

The underlying continuous problem is ill-posed, so we have a discrete ill-posed problem in which small changes in the data make extraordinarily large changes in the solution.

Practical methods change the ill-posed problem to a ”nearby” well-posed problem by using regularization.
How We View Image Reconstruction

Regularization
What Image Reconstruction Really Is
A Desirable Version of Image Reconstruction
Constraints used to regularize the solution
Data constraints

There are two kinds:

• Exact information
• Statistical information
Exact Data Constraints

There are some side-conditions that we know with certainty:

- **Nonnegativity**: present when data represents counts.
- **Upper and lower bounds**: present for pixel values: perhaps the range is $[0, 255]$.
- **Distribution of errors**: Poisson error for counts? White noise?
Statistical Data Constraints

There are some side-conditions that arise from models of the process that generates the noise.

- **Uncertainty estimate:** The experimentalist usually has an estimate of the noise level in the data.
- **Approx. distributional information:** “The world is normal.”
Bias constraints

Some constraints arise from our preference for how the solution should look. We may judge a solution to be bad if it isn’t what we expect to see!

- **Sparse representation**: We might prefer a solution that has
  - only a few nonzeros (e.g., for recovering a spectrum).
    Mathematical constraint: $\|x\|_0 \leq c$.
  - only a few edges, or oscillations, or ....
    Mathematical constraint: $\|Wx\|_0 \leq c$, where $W$ converts to a reasonable basis.

**Examples:**

* $W =$ Fourier basis. This would filter out high-frequency terms.
* $W =$ SVD basis. This results in Truncated-SVD.
* $W =$ wavelet, or ridgelet, or curvelet basis.
Bias Constraints, Continued

• **Bound on norm:** We might prefer a solution that is small in some (semi-)norm, to avoid the growth caused by using $A^{-1}$.

  Examples:
  - **Bound on solution norm:** $\|x\| \leq c$,
  - **Bound on derivative:** $\|Lx\| \leq c$.

  This results in **Tikhonov regularization**.
More bias constraints: Boundary Conditions

What is happening immediately outside the frame of the image?

- **Zero boundary conditions** are used for images of the sky.
- **Reflexive boundary conditions** are used when the image is rather flat near the edges.
- **Periodic boundary conditions** are used for convenience.
Evaluating the correctness of a solution

- Statistical properties of the residual
- Confidence intervals: useful if data constraints are strong enough.
Statistical properties of the residual

For our model \( b = Ax + \eta \), we assume that \( \eta \sim N(0, I) \).

As a consequence,
\[
\|\eta\|^2 \sim \chi^2(m).
\]

The advantage of our scaling is that the characteristics of a reasonable residual to the model are now obvious.

Let \( \tilde{x} \) be an estimate of \( x^* \) and let
\[
\tilde{r} = b - A\tilde{x}
\]
be the corresponding residual vector. Since the regression model can also be written
\[
\eta = b - Ax^*,
\]
it is clear that \( \tilde{x} \) is acceptable only if \( \tilde{r} \) is a plausible sample from the distribution from which \( \eta \) is drawn.
Since

\[ \mathcal{E}\left\{ \|\eta\|^2 \right\} = m, \quad \text{Var}\left\{ \|\eta\|^2 \right\} = 2m, \]

these two quantities provide rough bounds for the \( \|\tilde{r}\|^2 \) that might be expected from a reasonable estimate of \( x^* \): an estimate that gives

\[ m - \sqrt{2m} \leq \|b - A\tilde{x}\|^2 \leq m + \sqrt{2m} \]

would be reasonable.

These indicators can be sharpened and quantified by using percentiles of the cumulative distribution function for \( \chi^2(m) \).
Useful statistical diagnostics

Rust (1998, 2000) suggested several diagnostics for judging a regularized solution $\tilde{x}$ with residual $\tilde{r}$.

**Diagnostic 1.** The residual norm-squared should be within two standard deviations of the expected value of $\|\eta\|^2$. This quantifies the Morozov discrepancy principle.

**Diagnostic 2.** The elements of $\eta$ are drawn from a $N(0, 1)$ distribution, and a graph of the elements of $\tilde{r}$ should look like samples from this distribution. (In fact, a histogram of the entries of $\tilde{r}$ should look like a bell curve.)

**Diagnostic 3.** We consider the elements of both $\eta$ and $\tilde{r}$ as time series, with the index $i$ ($i = 1, \ldots, m$) taken to be the time variable. Since $\eta \sim N(0, I_m)$, the $\eta_i$ form a white noise series. Therefore the residuals $\tilde{r}$ for an acceptable estimate should constitute a realization of such a series.
Diagnostics: Periodogram

A formal test of Diagnostics 2 and 3, used by Rust (2000), is based on a plot of the periodogram, which is an estimate of the power spectrum on the frequency interval \(0 \leq f \leq \frac{1}{2T}\), where \(T\) is the sample spacing for the time variable.

Here, the time variable is the element number \(i\), so \(T = 1\). The periodogram is formed by

- zero-padding the time series to length \(N\) (e.g., a power of 2),
- taking the discrete Fourier transform of this zero-padded series,
- taking the squares of the absolute values of the first half of the coefficients.

This gives us the periodogram \(z\) of values corresponding to the frequencies \(k/NT, k = 0, \ldots, N/2\).
Diagnostics: Cumulative Periodogram

The cumulative periodogram $c$ is the vector of partial sums of the periodogram, normalized by the sum of all of the elements.

- For an ideal residual, periodogram ordinates are multiples of independent $\chi^2(2)$ samples and hence the vector elements are distributed like an ordered sample of size $N/2$ from a uniform (0,1) distribution.

- Therefore, the ideal cumulative periodogram is a straight line between 0 and 1 as the frequency varies between 0 and 0.5, so we expect its length to be close to 1.118 (taking $T = 1$).

- A test of the hypothesis that the residuals are white noise can be obtained using Kolomogorov-Smirnov statistics.
Quantitative measures for the diagnostics

- **Diagnostics 2 and 3:** the length of the cumulative periodogram and the number of samples outside the 95% confidence band.
- **Diagnostic 1:** residual norm-squared.

These quantitative diagnostics are used in conjunction with plots of the residual vector, its periodogram, and its cumulative periodogram.
Some history

• **Rust** (2000) suggested choosing a parameter that passed all three of the diagnostics given above.

• Later **Hansen, Kilmer and Kjeldsen** (2006) proposed choosing the parameter by either of two methods:
  – Choosing the most regularized solution estimate for which the cumulative periodogram lies within the 95% confidence interval.
  – Minimizing the sum of the absolute values of the difference between components of $c$ and the straight line passing through the origin with slope $2/NT$.

  Both methods tend to undersmooth.

• **Mead and Renaut** (2007) used distribution properties of the residual to choose regularization parameters.
Back to Case Study 1

Noisy image, SNR = 40dB

Original blurred image, $9 \times 9$ boxcar blur, SNR = 40
Wiener, ISNR = -16.21dB, SNR = 27.9dB

Ridgelet, reflexive b.c. (ISNR = 50.77dB, SNR = 43.77dB)
Cumulative Periodogram for True Solution

Length = 1.197

95% confidence interval
Cumulative Periodograms for Computed Solutions

- **Wavelet (forWard)**
  - Length = 1.212
  - 95% confidence interval

- **Curvelet**
  - Length = 0.517
  - 95% confidence interval

- **Wiener (a default param)**
  - Length = 1.347
  - 95% confidence interval

- **Ridgelet**
  - Length = 0.518
  - 95% confidence interval
Ridgelets ($64 \times 64$ subimage)  

Tikhonov

True + noise  

True + $A^{-1}$ noise
Cumulative Periodograms and the Corresponding Images
Wavelet (forWard)

ResidTarget = .21

Estimated image (ISNR =50.77dB, SNR =43.77dB)

Curvelet

ResidTarget = 4.6
Wiener (a default param)  

Ridgelets  

ResidTarget = .003  

ResidTarget = 4.6
Ridgelets \((64 \times 64 \text{ subimage})\)

\(\text{ResidTarget} = 5.1\)

Tikhonov

\(\text{ResidTarget} = 1.0\)
True + noise

ResidTarget = 6.0

True + $A^{-1}$ noise

ResidTarget = .40
The implicit constraints

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelet</td>
<td>Solution represented as a small number of <strong>square bumps</strong>.</td>
</tr>
<tr>
<td>Ridgelet</td>
<td>Solution represented as a small number of <strong>straight ridges</strong>.</td>
</tr>
<tr>
<td>Curvelet</td>
<td>Solution represented as a small number of <strong>curved ridges</strong>.</td>
</tr>
<tr>
<td>Wiener</td>
<td>The norm of the solution is constrained to be small (<strong>SVD basis</strong>).</td>
</tr>
<tr>
<td>Postprocessing</td>
<td>(Applied to four methods above.) Solution is represented as a small number of low-frequency functions (<strong>Fourier basis</strong>).</td>
</tr>
<tr>
<td>Tikhonov</td>
<td>The norm of the solution is constrained to be small (<strong>SVD basis</strong>).</td>
</tr>
</tbody>
</table>
Validation of these methods

- Theoretical properties derived based on assumptions about the class of images to be encountered: e.g., they are sparse in the chosen basis.
- Numerical validation on test images satisfying these assumptions.

Our biases are showing!
Does the “best” result represent the truth well?

ForWaRD, ISNR = 8.37dB, SNR = 52.49dB

Wavelet

Truth
False color for highlighting

ForWaRD, ISNR = 8.37dB, SNR = 52.49dB

Wavelet

Truth

Original image
Even our best reconstruction misses important features without giving indication of trouble
Confidence in Image Reconstructions
Confidence in Image Reconstructions

Usually, when we are given a problem $Ax \approx b$, we return a point estimate, our best approximate solution.

If careful, a numerical analyst might also return information such as the residual norm $\|b - Ax_{\text{comp}}\|$ and $\kappa(A) = \|A\|\|A^\dagger\|$.

Given our noise assumptions, a statistician might also return confidence intervals, bounds $x_{lo}$ and $x_{hi}$ so that (for example) if the measurement is repeated many times, we expect that 95% of the time the computed solution satisfies

$$x_{lo} \leq x_{\text{comp}} \leq x_{hi}.$$
A problem with this:

For ill-posed problems, confidence intervals are so wide that they are useless unless we make use of other constraints on the problem.
What We Can Achieve with Exact Data Constraints

The probability that the true value of pixel $k$ is in the interval $[\ell_k, u_k]$ is at least 95% if

$$
\ell_k = \min\{x_k : \|Ax - b\|_2^2 \leq r_{\min} + \gamma, 0 \leq x \leq 255\},
$$

$$
u_k = \max\{x_k : \|Ax - b\|_2^2 \leq r_{\min} + \gamma, 0 \leq x \leq 255\},
$$

where

$$r_{\min} = \min\{\|Ax - b\|_2^2 : 0 \leq x \leq 255\},$$

and $\gamma$ is determined from the $\chi$-squared distribution.

So, for each pixel we need to solve 2 optimization problems with a quadratic constraint and bounds on the $n^2$ variables (for an $n \times n$ image).
Examples Where the Cost is (Relatively) Low

One-dimensional blurs of images:

- **Vertical or horizontal motion blur.** Then each optimization problem involves only $n$ variables, not $n^2$.

- **Radial blur.** After transformation to polar coordinates, each optimization problem only involves pixels along a radial line.

- **Spin blur.** Polar coordinates give problems with the number of pixels equal to the number of radial lines.

The radial and spin blur work is joint with Julianne Chung and Glenn Easley.
Sample Results

- Spin blur through an angle of $\pi/12$.
- MRI image (convenient since the boundary conditions are zero outside the circle of reconstructed pixels)
- SNR $= 100$. 
How might we illustrate a confidence interval?

Easy for signals, and it yields important information, depending on the width of the intervals.

Actually just as easy for images: Twinkle (joint work with Jim Nagy).
More Sample Results

Consider a camera mounted over a lane of traffic, imaging license plates. The blur is motion blur in the vertical direction.

How well can it capture data from a car traveling at 60 mph?
Work in progress

• Improve the reliability of the optimization.
• Make it more efficient, so that the optimization algorithm takes advantage of previous solutions.
• Take advantage of structure in $\mathcal{A}$. 
Case Study 3: Another way to determine confidence intervals

This is joint work with Julianne Chung and Matthias Chung. See Julianne’s poster for more information on this method for deblurring.
Case Study 3

Assume that we have calibration data for our imaging device, pairs of true images \( x_k \) and the corresponding noisy blurred images \( b_k, k = 1, \ldots, \hat{k} \).

And assume that these are a statistically good set of samples from the space of true images and the space of noise and that \( \hat{k} \) is sufficiently large.

We are given the blurring operator but know nothing about the noise characteristics.

Now, determine a reconstruction algorithm optimal in some sense.

Apply the reconstruction to each \( b_k \) to obtain \( y_k \), and study the distribution of the error in each pixel: \( (y_k - x_k)_{ij}, k = 1, \ldots, \hat{k} \).

This gives an indication of the expected error in each pixel of new noisy blurred images produced by the imaging device.
Experimental Results: How well are the errors predicted?

Histogram of Training Images mean error pixel-by-pixel

Histogram of Test Images mean error pixel-by-pixel
Conclusions

- Without constraints, finding the “solution” to an ill-posed problem is not possible.
- Some data constraints are usually evident from the physical problem, but these are not usually enough to stabilize the solution.
- With strong bias constraints, the solution becomes well-determined and unique and reproducible and often quite visually attractive. But looks can be deceiving.
- Diagnostics such as residual periodograms and confidence images should be reported along with the solution.
- Data constraints may give a less attractive solution but one in which we can be confident.
- Confidence intervals, from solving optimization problems or from training data, can be very useful in distinguishing real features from apparent ones.
References / Acknowledgements


• The Mars image: a modification of one supplied by Tim McClanahan, NASA.

• The curvelet and ridgelet software: Glenn Easley and Julianne Chung.

• The wavelet software was written by R. Neelamani, H. Choi, and R. G. Baraniuk: http://www.dsp.rice.edu/


Specs for Case Study 1

• The psf was “boxcar”: equally-weighted blur over a $9 \times 9$ square of pixels.

• The problems were generated to accommodate the software:
  – Ridgelet and curvelet required dimensions equal to a power of 2 and used reflexive b.c.
  – Wavelet required a power of 2 plus a border.
  – Tikhonov used Kronecker product structure and required zero boundary conditions.

• SNR = 40.

• Parameter choice for Tikhonov was to produce the best cumulative periodogram for the residual.

• Parameters for wavelet, ridgelets, and curvelets were chosen by the software.
Specs for Case Study 3

- 8 images of satellites with 100 rotations, translations, and magnification, for a total of 800 training images.
- 800 test images.
- Gaussian blur, reflexive boundary conditions.
- Gaussian noise, with noise level chosen uniformly between 0.1 and 0.15.
- “Optimal error” filter