Optimization Methods for Image Deblurring

April 2009

NASA Goddard 2009

Dianne P. O'Leary

Optimization Methods for Image Deblurring

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/~oleary/

Joint work with
Bert W. Rust
US National Institute of Standards and Technology (NIST)
James Nagy
Emory University

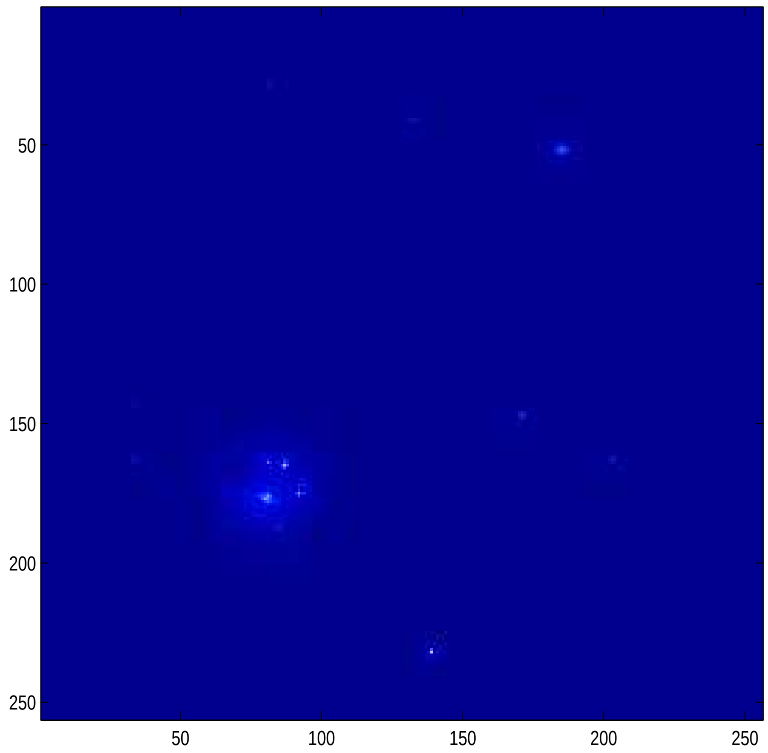
My work was supported by NIST and the National Science Foundation.

The Plan:

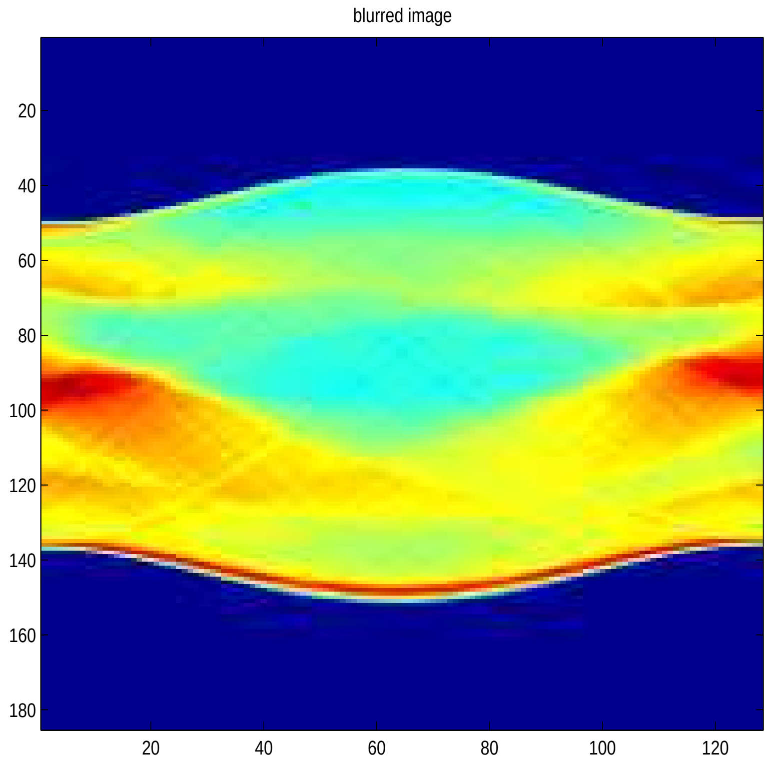
- Background: What is the image reconstruction problem?
- Using optimization to improve the quality of the reconstructions.
- Using *confidence images* to provide information on the reliability of the reconstructions.

The image reconstruction problem

Suppose we have taken an image of some scientific data ...



... Or suppose we have taken an image of some medical data...

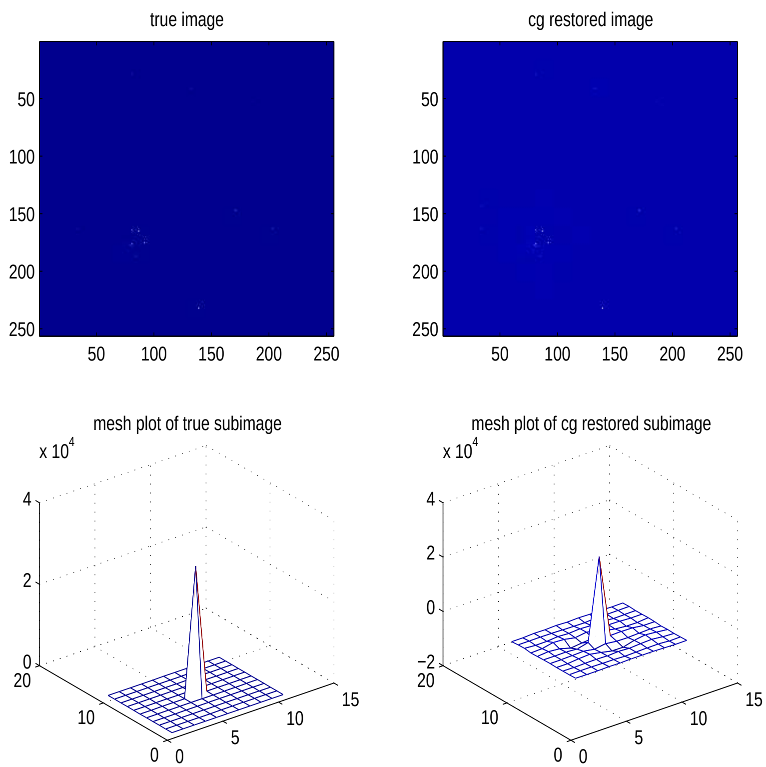


Goals

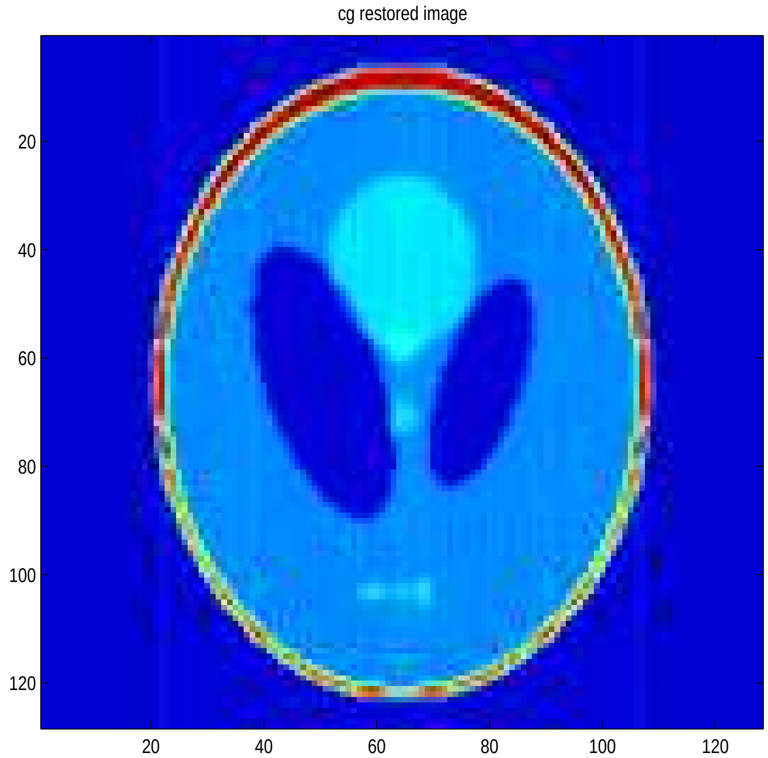
Our goals are to

- reconstruct the image as well as possible.
- give the scientist or clinician useful information about the reliability of the reconstruction.

For example, can we obtain an estimate and confidence limits on the brightness of a star...



... or can we give the clinician some confidence in the reality of a suspected tumor...



The underlying difficulty

The problem

Systems of first kind integral equations:

$$y_i \equiv y(t_i) = \int_a^b K(t_i, \xi) x(\xi)\, d\xi + \epsilon_i \ , \quad i = 1, 2, \ldots, m$$

- $K_i(\xi) \equiv K(t_i, \xi)$ are known (previously measured or calculated) response functions of the instrument,
- the y_i are measurements made on a discrete mesh $t_1, t_2, \ldots, t_m$,
- the ϵ_i are random, zero-mean measuring errors.

Discretizing gives a linear regression model

$$\boldsymbol{y} = \boldsymbol{K}\boldsymbol{x}^* + \boldsymbol{\epsilon} \ ,$$

Where $\boldsymbol{y}$ is the m-vector of measurements, $\boldsymbol{K}$ is a known $m \times n$ matrix, with $m \geq n$, and $\boldsymbol{x}^*$ is an unknown n-vector whose components are either discrete values of $x(\xi)$ on some mesh $\xi_1, \xi_2, \ldots, \xi_n$, or are the unknown coefficients in a truncated expansion for $x(\xi)$.

The assumption

The vector $\boldsymbol{\epsilon}$ is an m-vector of random measuring errors satisfying

$$\mathcal{E}(\boldsymbol{\epsilon}) = \mathbf{0} \ , \qquad \mathcal{E}\left(\boldsymbol{\epsilon}\, \boldsymbol{\epsilon}^T\right) = \boldsymbol{S}^2 \ ,$$

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

- The variances can be estimated as the squares of the half-lengths of the ± 1-sigma error bars on the measurements.
- An analyst who fails to use this information implicitly assumes that $\boldsymbol{S}^2 = s^2 \boldsymbol{I}_m$ where $\boldsymbol{I}_m$ 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 $\boldsymbol{x}$.

We assume that $\boldsymbol{S}$ is a known matrix and that the errors are samples from a multivariate normal distribution, i.e., that $\boldsymbol{\epsilon} \sim N(\ \mathbf{0}\ ,\ \boldsymbol{S}^2\ )$.

Scaling the model

Original linear regression model:

$$\boldsymbol{y} = \boldsymbol{K}\boldsymbol{x}^* + \boldsymbol{\epsilon} \,, \quad \boldsymbol{\epsilon} \sim N(\, \mathbf{0} \,,\, \boldsymbol{S}^2 \,) \,.$$

Scaled linear regression model:

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

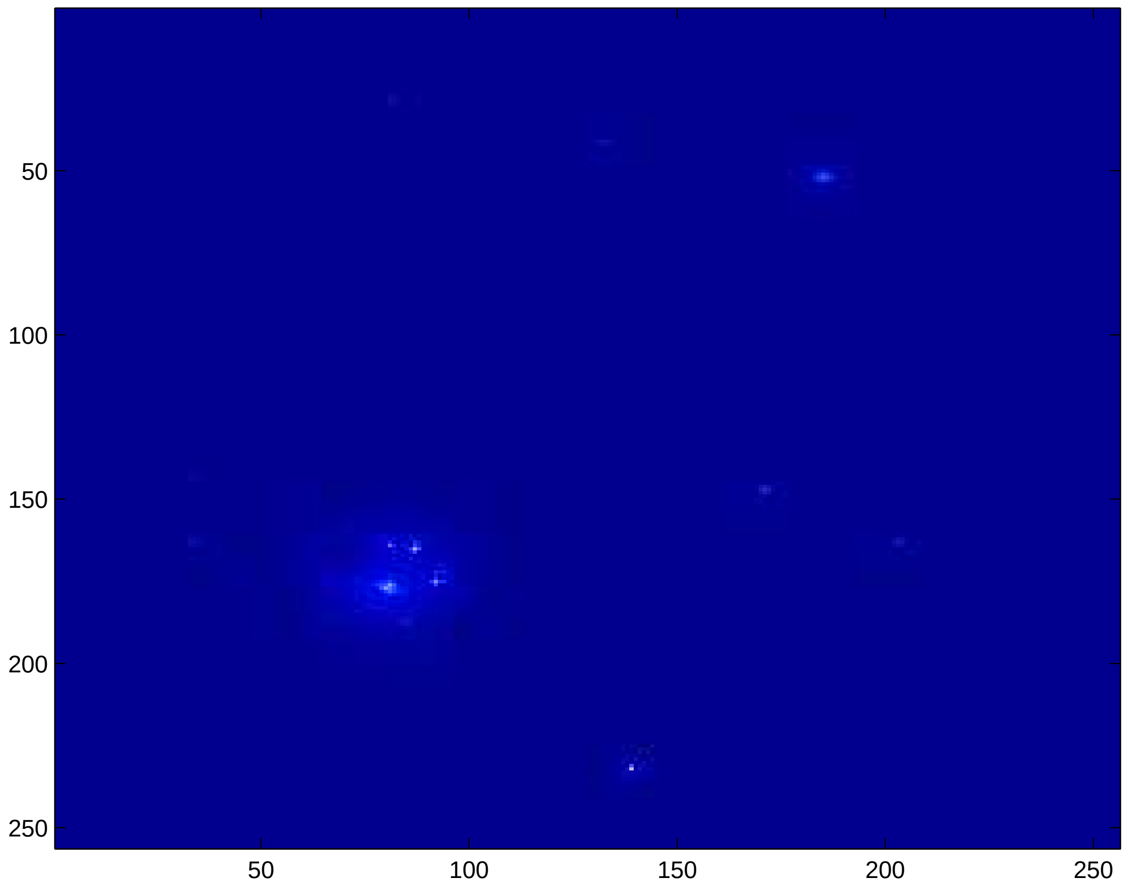
Then $\boldsymbol{\eta} \sim N(\, \boldsymbol{S}^{-1}\mathbf{0} \,,\, \boldsymbol{S}^{-1}\boldsymbol{S}^2[\boldsymbol{S}^{-1}]^T \,)$, so the scaled model can be written

$$\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x}^* + \boldsymbol{\eta} \,, \quad \boldsymbol{\eta} \sim N(\, \mathbf{0} \,,\, \boldsymbol{I}_m \,) \,.$$

As a consequence,

$$\|\boldsymbol{\eta}\|^2 \sim \chi^2(m).$$

An example: The image reconstruction problem



The problem fits our model:

$$\boldsymbol{y} = \boldsymbol{K}\boldsymbol{x}^* + \boldsymbol{\epsilon} \ ,$$

- $\boldsymbol{y}$: the observed image
- $\boldsymbol{K}$: the “blurring matrix”
- $\boldsymbol{x}$: the true image

Solutions to our problem

$$r_{\min}^2 = \min_{\boldsymbol{x}} \left\{ (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x})^T (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) \right\} \ .$$

Linear regression estimate:

$$\hat{\boldsymbol{x}} = \left(\boldsymbol{A}^T \boldsymbol{A} \right)^{-1} \boldsymbol{A}^T \boldsymbol{b}$$

is the best linear unbiased estimate of $\boldsymbol{x}^*$.

It is well known that for ill-posed problems, the elements of $\hat{\boldsymbol{x}}$ are pathologically sensitive to small variations in the elements of $\boldsymbol{b}$, so despite the desirable statistical properties of the estimator $\hat{\boldsymbol{x}}$, the measuring errors generally make it totally unphysical and wildly oscillating.

We'll see an example of this shortly.

Regularized Solution Estimates

Insight into the failure of the least squares method is obtained by use of the singular value decomposition (SVD) of $\boldsymbol{A}$:

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T = \boldsymbol{U}\begin{bmatrix} \boldsymbol{\Sigma}_1 \\ \boldsymbol{0} \end{bmatrix}\boldsymbol{V}^T \; , \qquad \boldsymbol{\Sigma}_1 = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n) \; .$$

Here $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$, and

$$\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}_m = \boldsymbol{U}\boldsymbol{U}^T \; , \qquad \boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I}_n = \boldsymbol{V}\boldsymbol{V}^T \; .$$

Least squares solution:

$$\hat{\boldsymbol{x}} \; = \; \sum_{i=1}^{n} \frac{\boldsymbol{u}_i{}^T\boldsymbol{b}}{\sigma_i}\boldsymbol{v}_i,$$

$$r_{\min}^2 \; = \; \|\boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}}\|^2 = \sum_{i=n+1}^{m} (\boldsymbol{u}_i{}^T\boldsymbol{b})^2 \; .$$

The Tikhonov Family of Regularized Solutions:

Rather than setting small singular values to zero, we can increase them so that they lead to a smaller contribution to the solution estimate.

The Tikhonov regularization estimate is given by

$$\tilde{\boldsymbol{x}}_\lambda = \left(\boldsymbol{A}^T\boldsymbol{A} + \lambda^2\boldsymbol{I}_n\right)^{-1}\boldsymbol{A}^T\boldsymbol{b}\ ,$$

where λ is a parameter chosen to balance the competing demands of fidelity to the measurements and insensitivity to measurement errors.

This interpretation of λ comes from the fact that the vector $\tilde{\boldsymbol{x}}_\lambda$ is the solution to the minimization problem

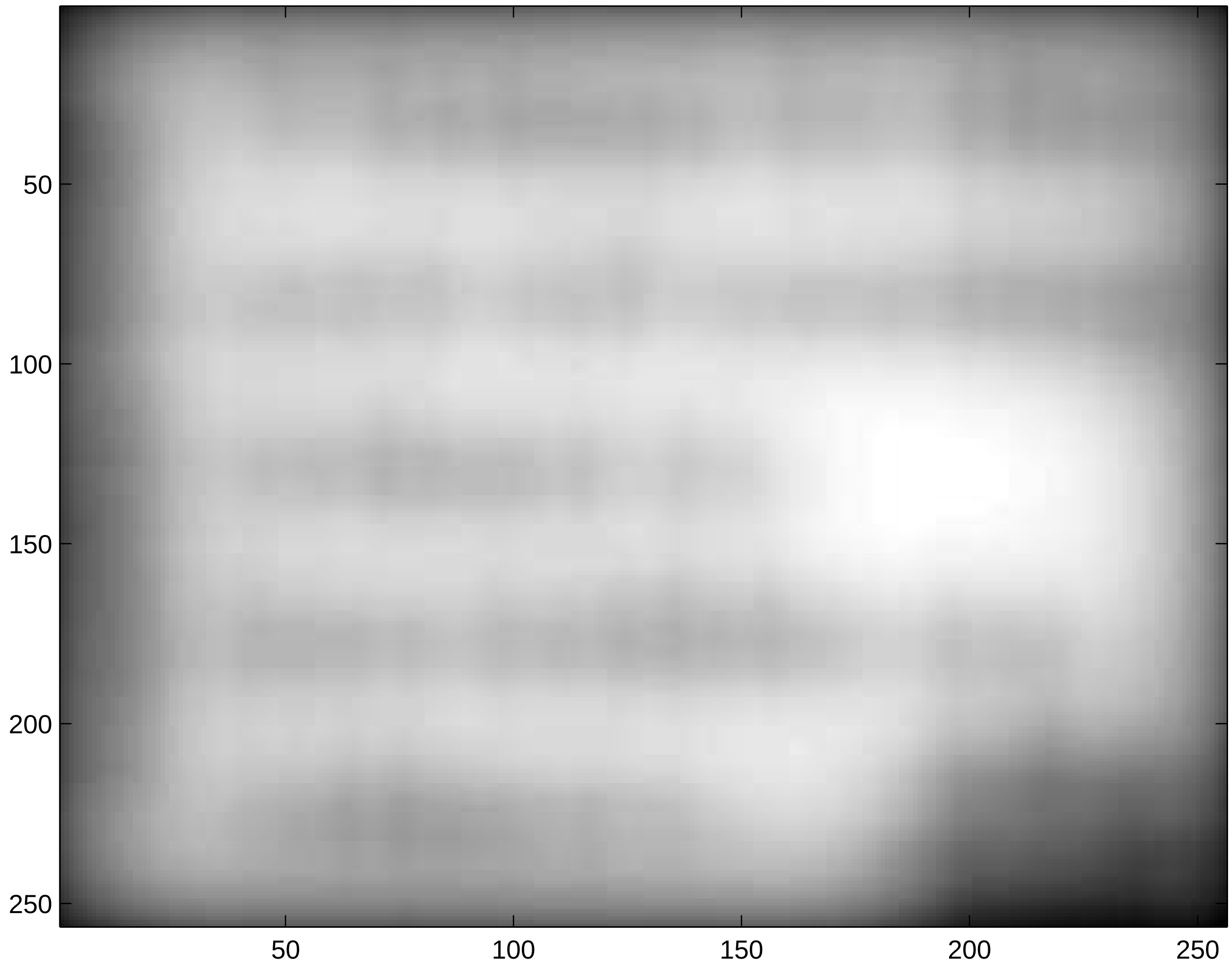
$$\min_{\boldsymbol{x}} \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|^2 + \lambda^2\|\boldsymbol{x}\|^2.$$

Some history of Tikhonov regularization

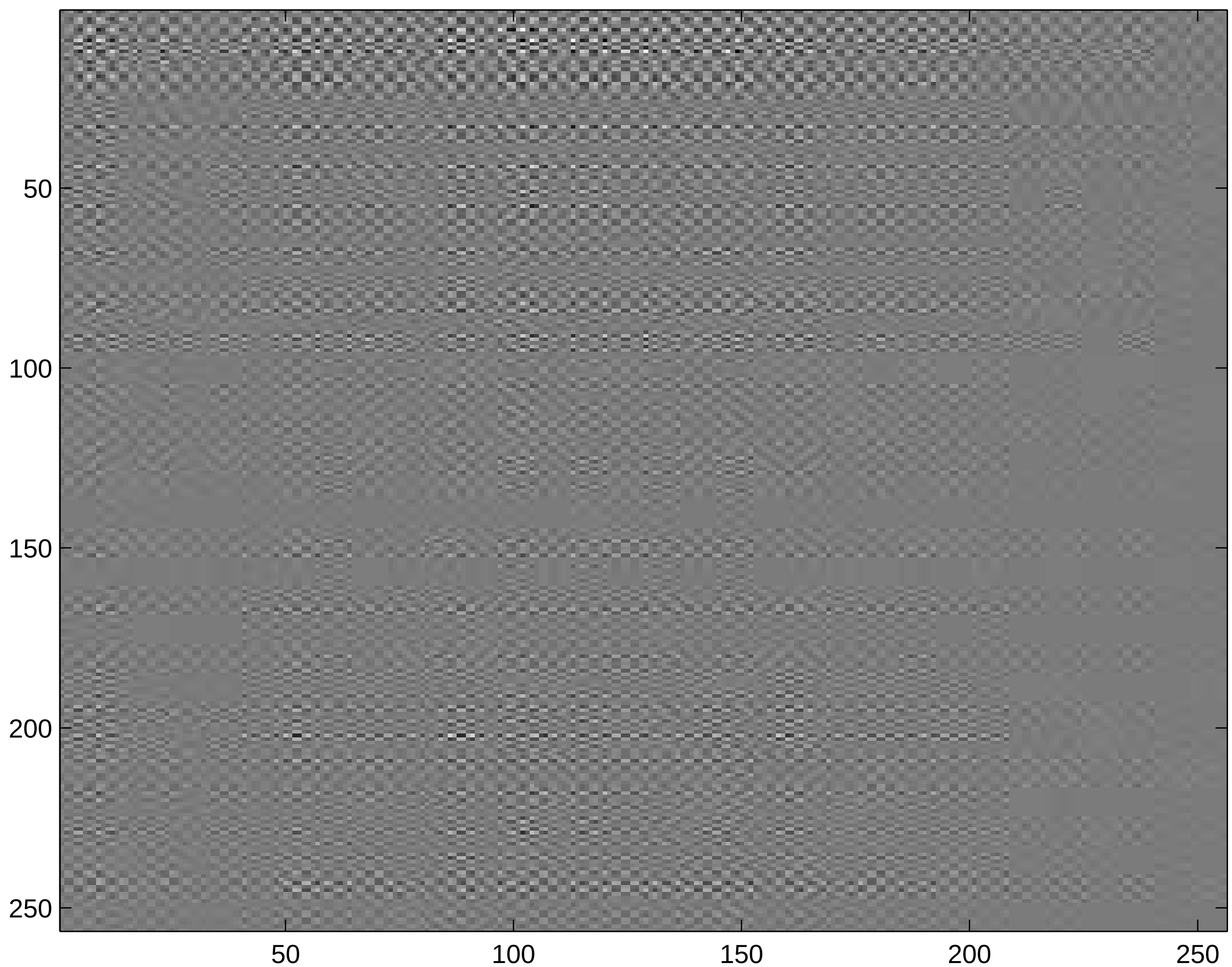
- This strategy is generally attributed to Tikhonov (1963).
- It was also pioneered by Phillips (1962) and Twomey (1963).
- Golub and Kahan (1965) noted its relationship to the theoretical treatment given by Smithies (1958) for the singular functions and singular values of first kind integral equations.
- Foster (1961) relates it to Wiener-Kolmogorov filtering and refers to Riley (1955) for the formula.
- A statistical variant, called ridge regression, was independently developed by Hoerl and Kennard (1970).

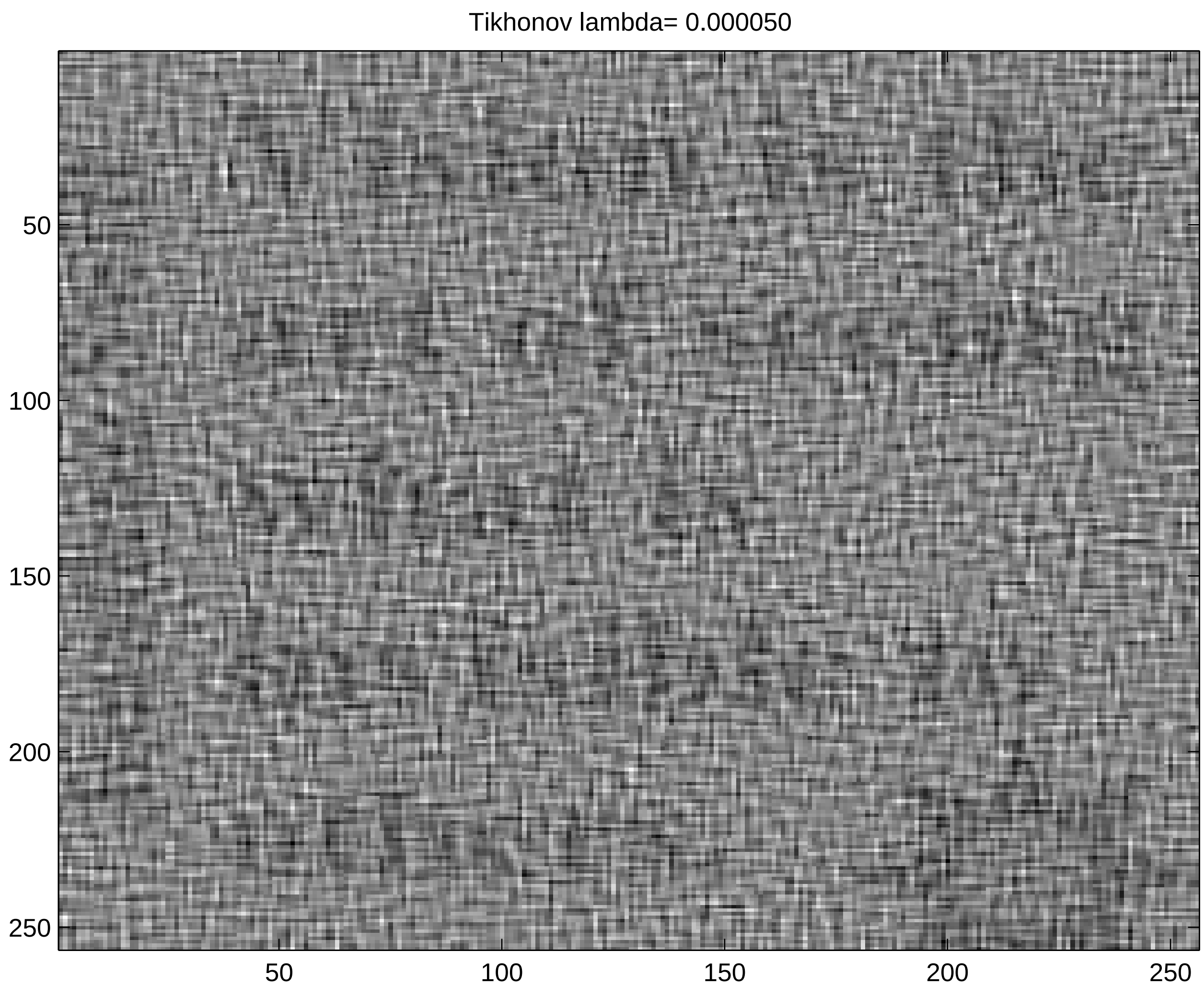
Optimization and Validation using the “eye” norm

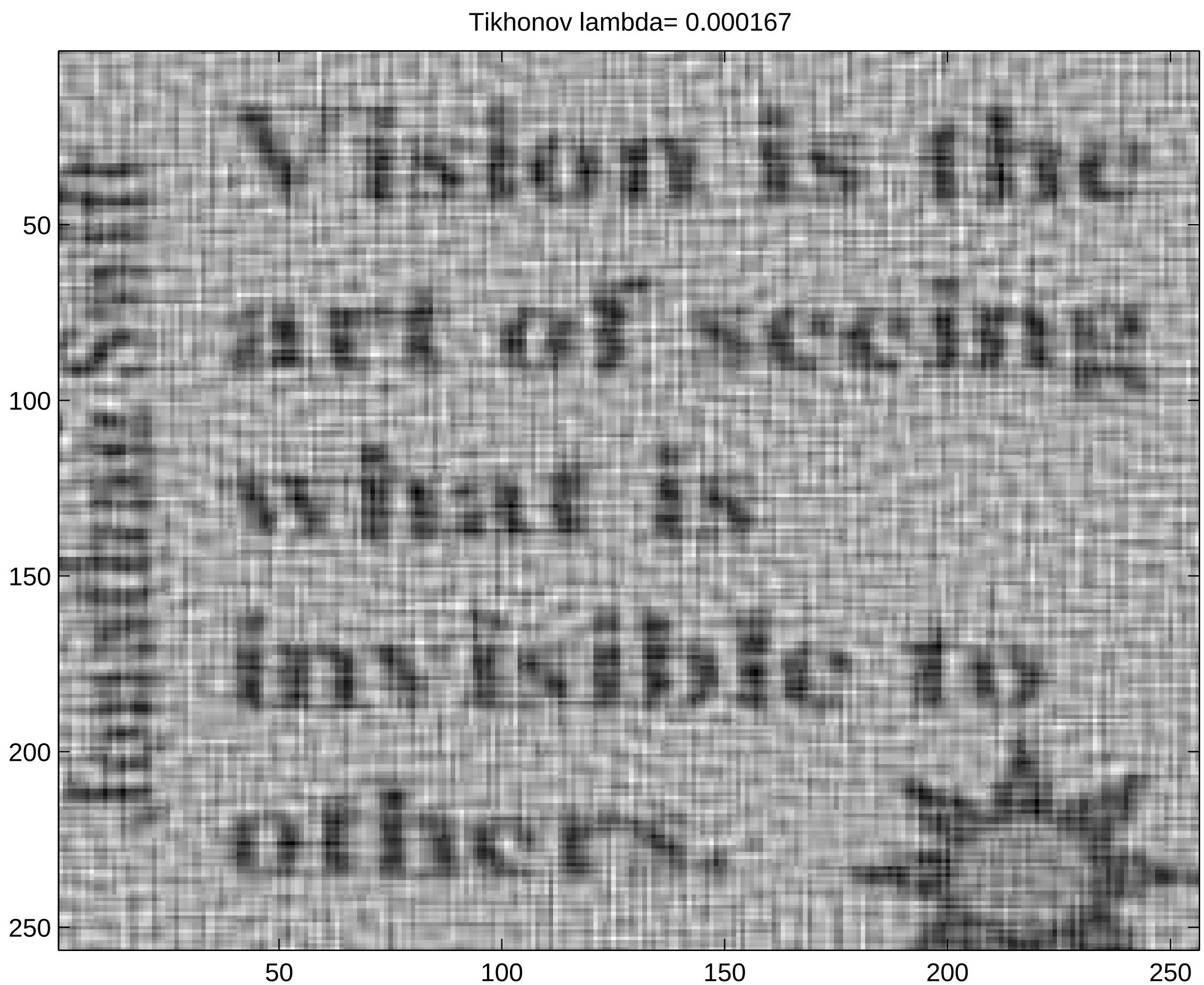
An example: Can we deblur this image?



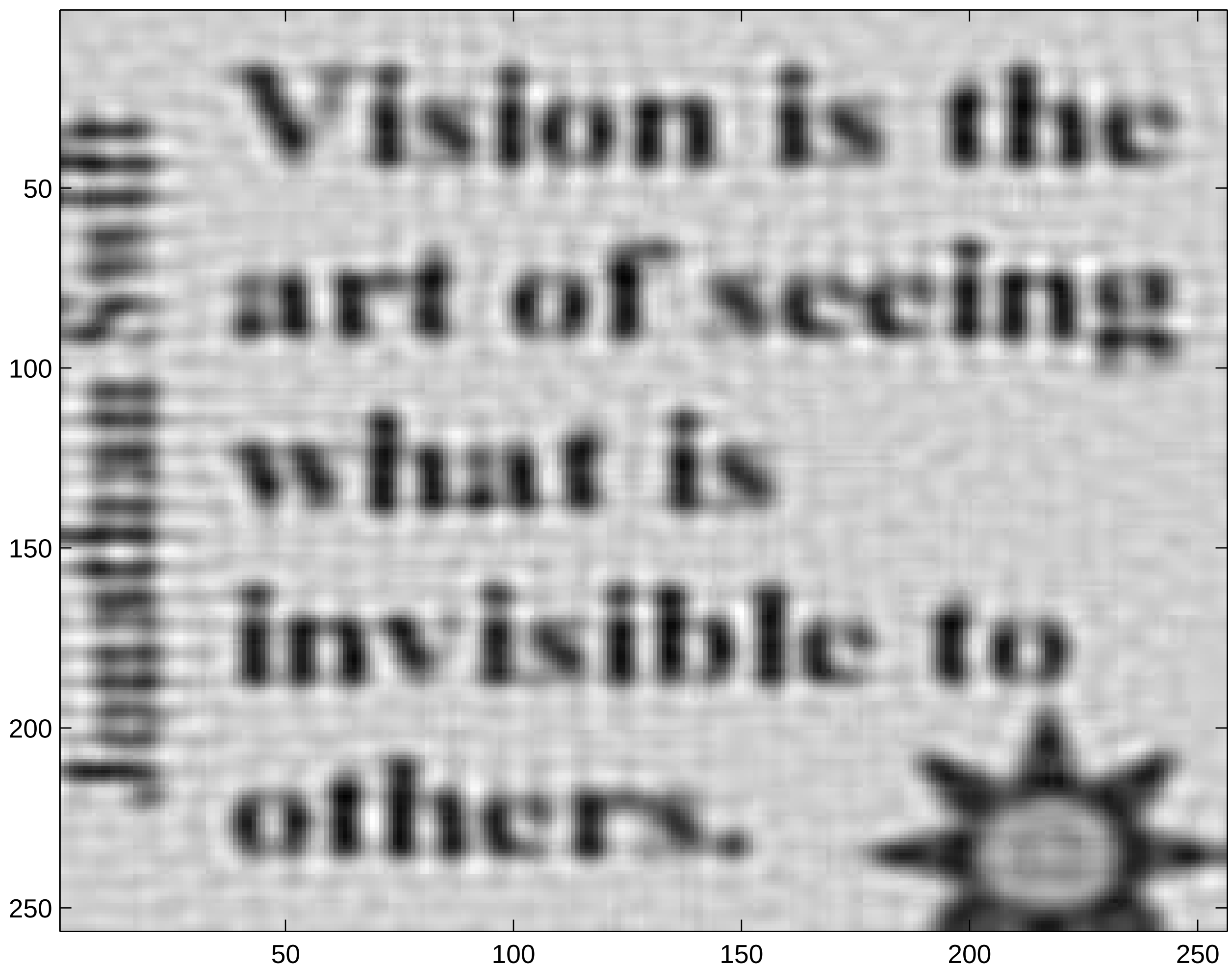
Tikhonov lambda= 0.000000



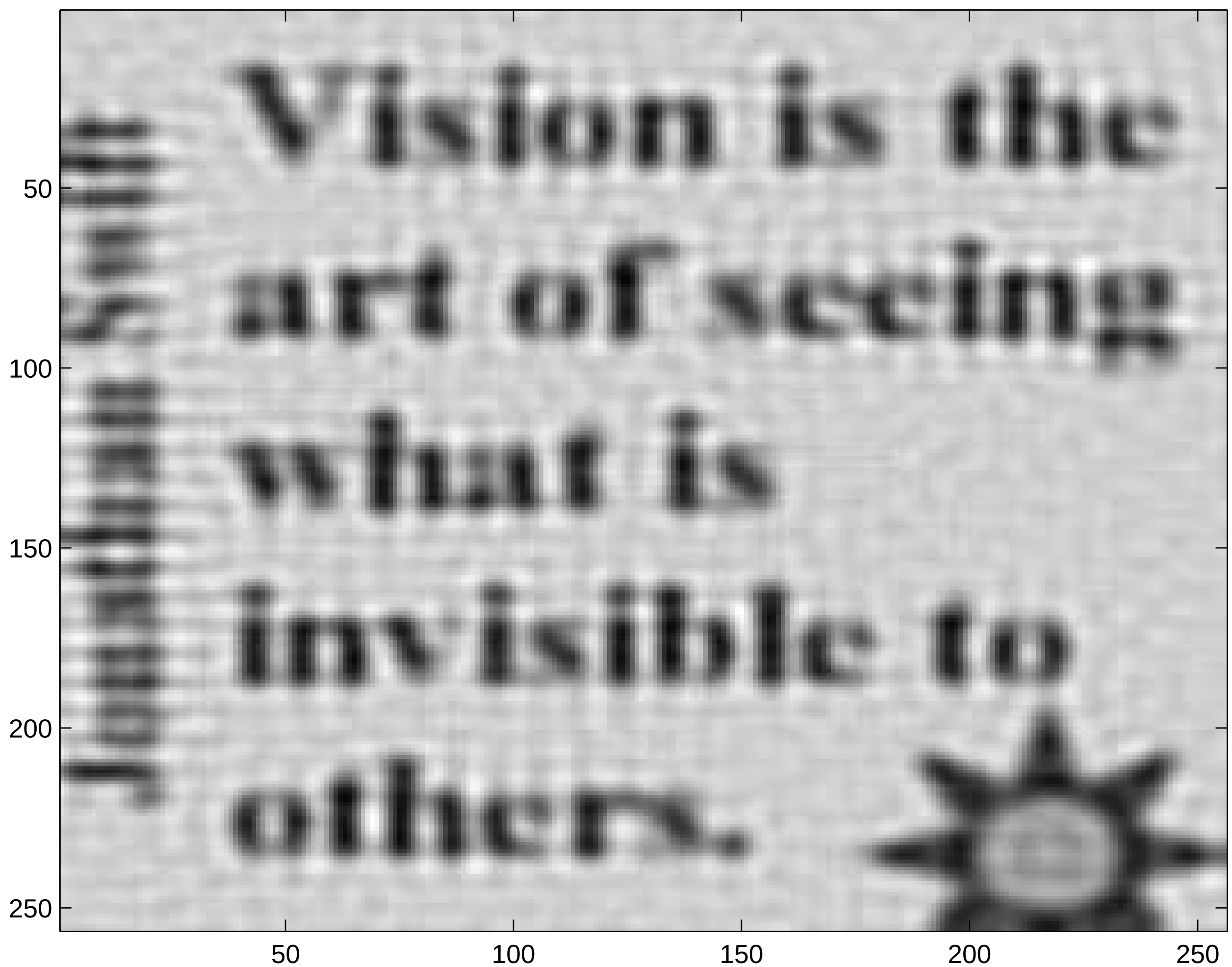

Tikhonov lambda= 0.000050
50
100
150
200
250
50
100
150
200
250

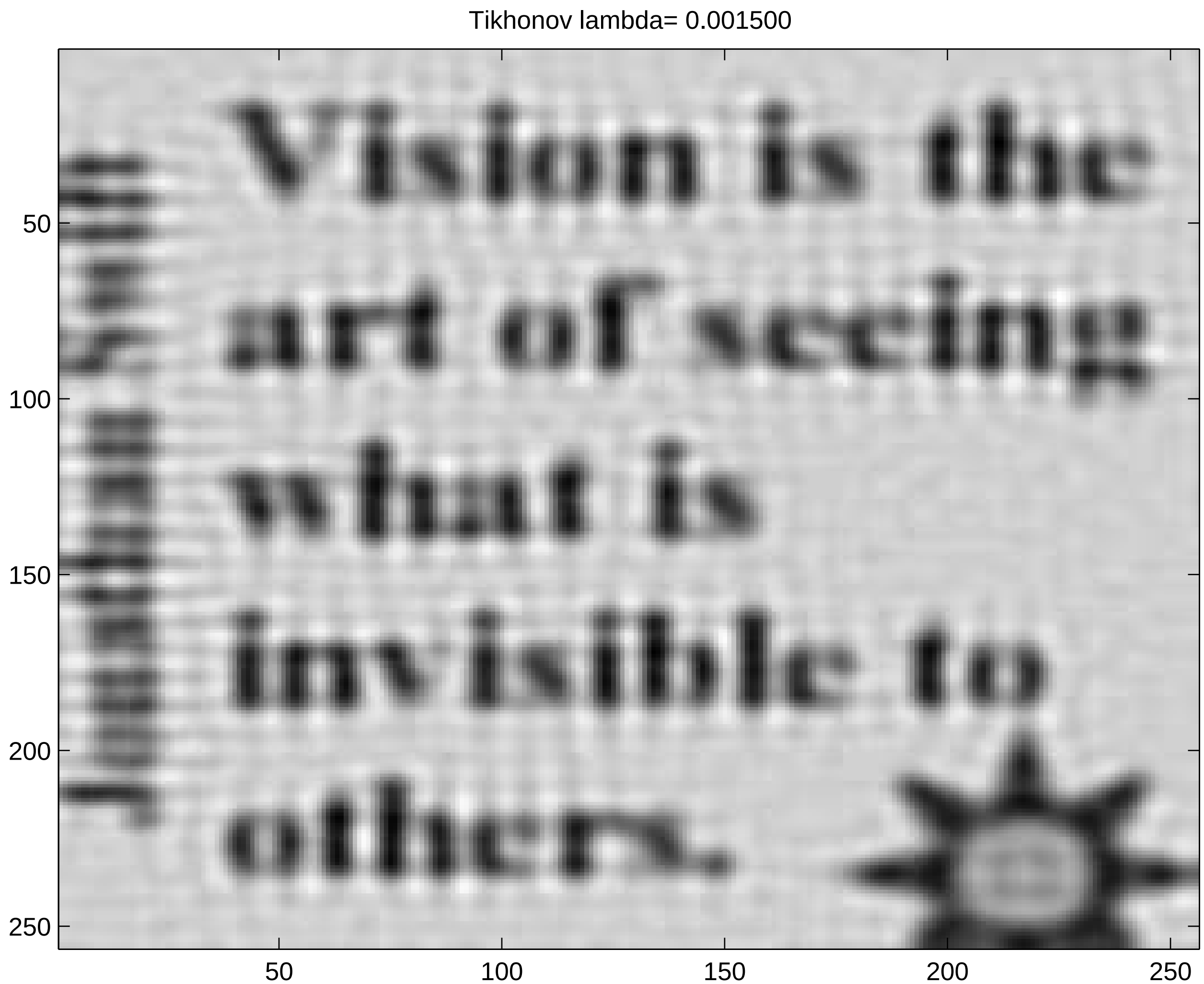

Tikhonov lambda= 0.000167
Vision is the
art of seeing
what is
invisible to
others.
50
100
150
200
250
50
100
150
200
250

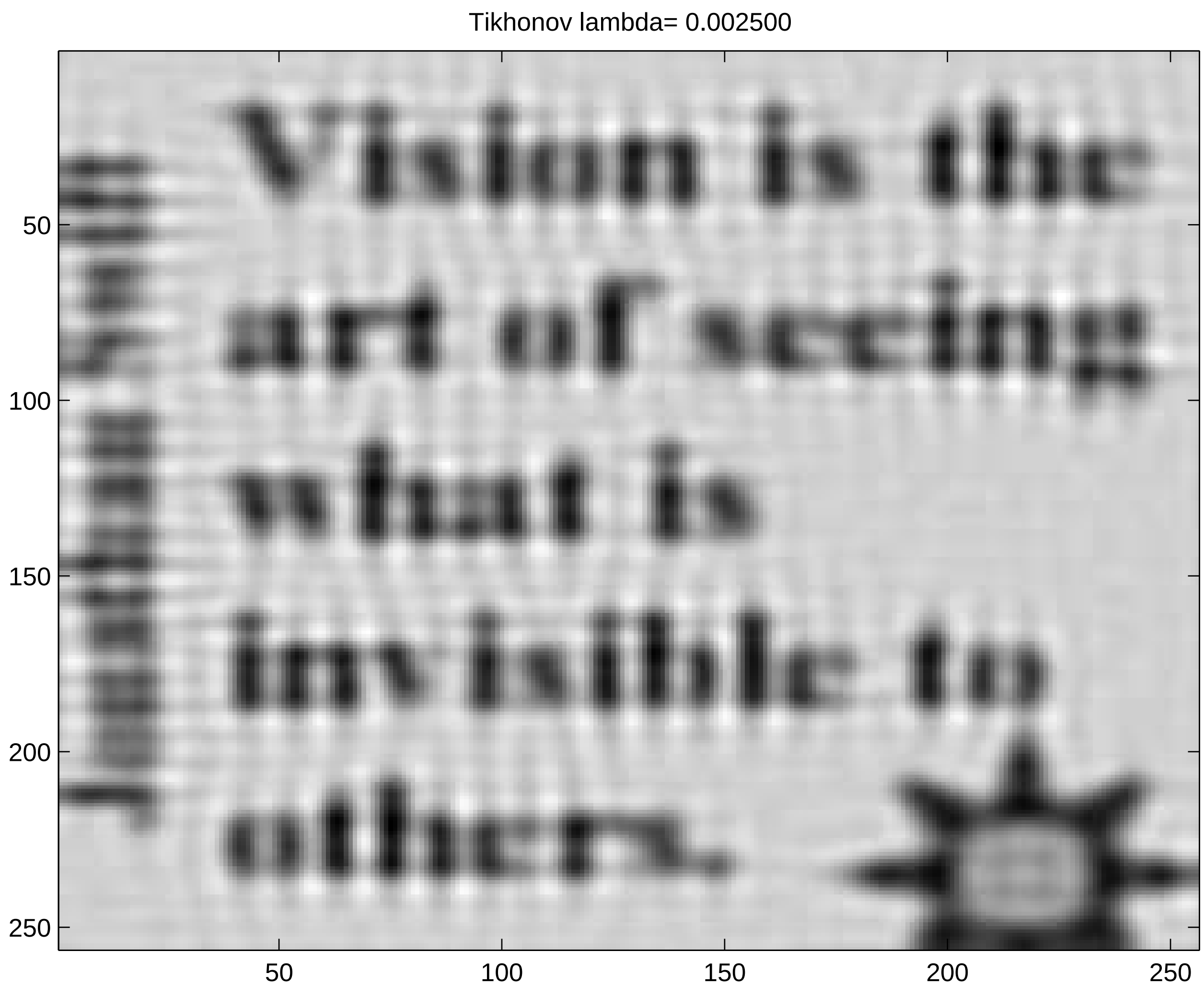
Tikhonov lambda= 0.000950

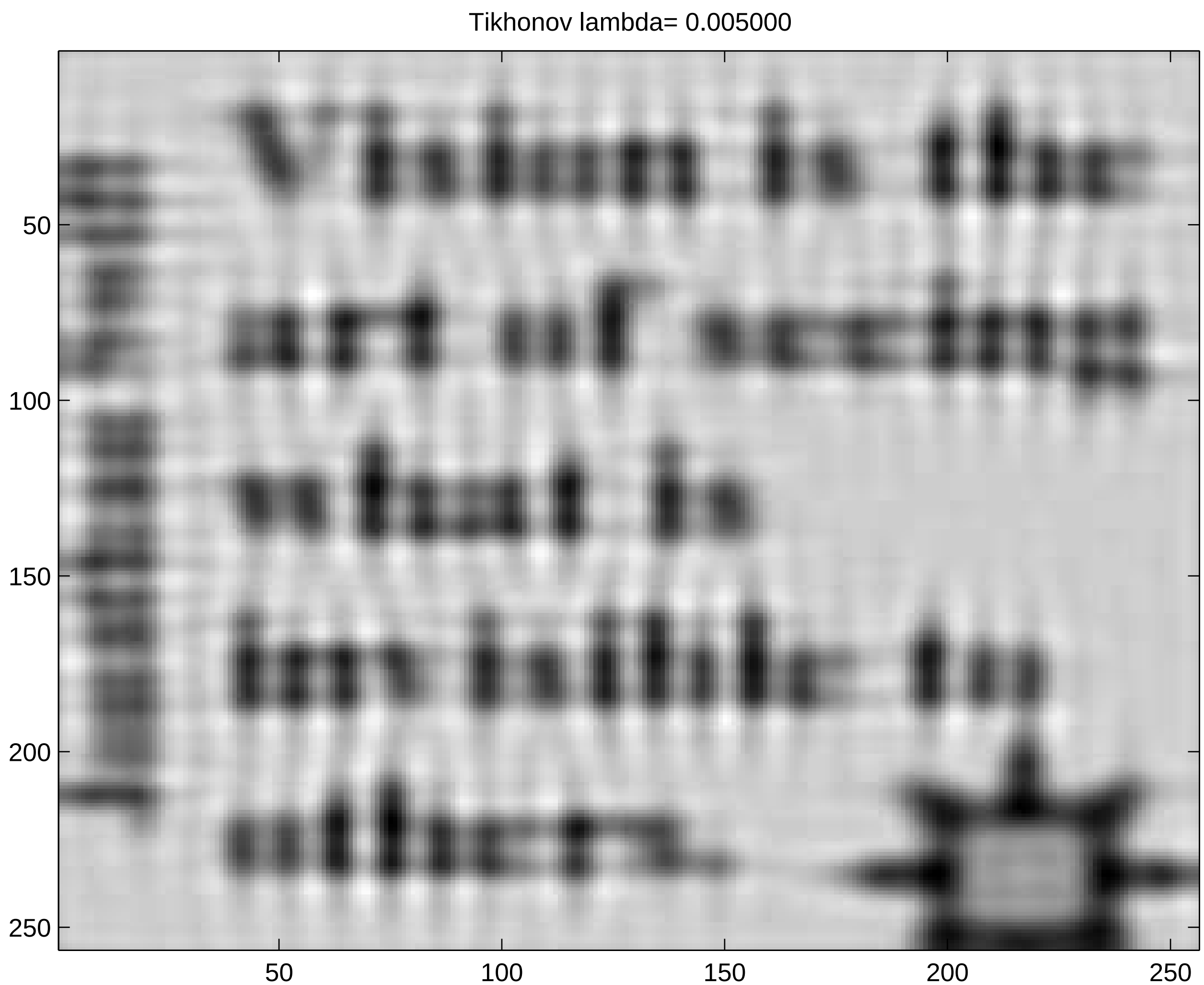


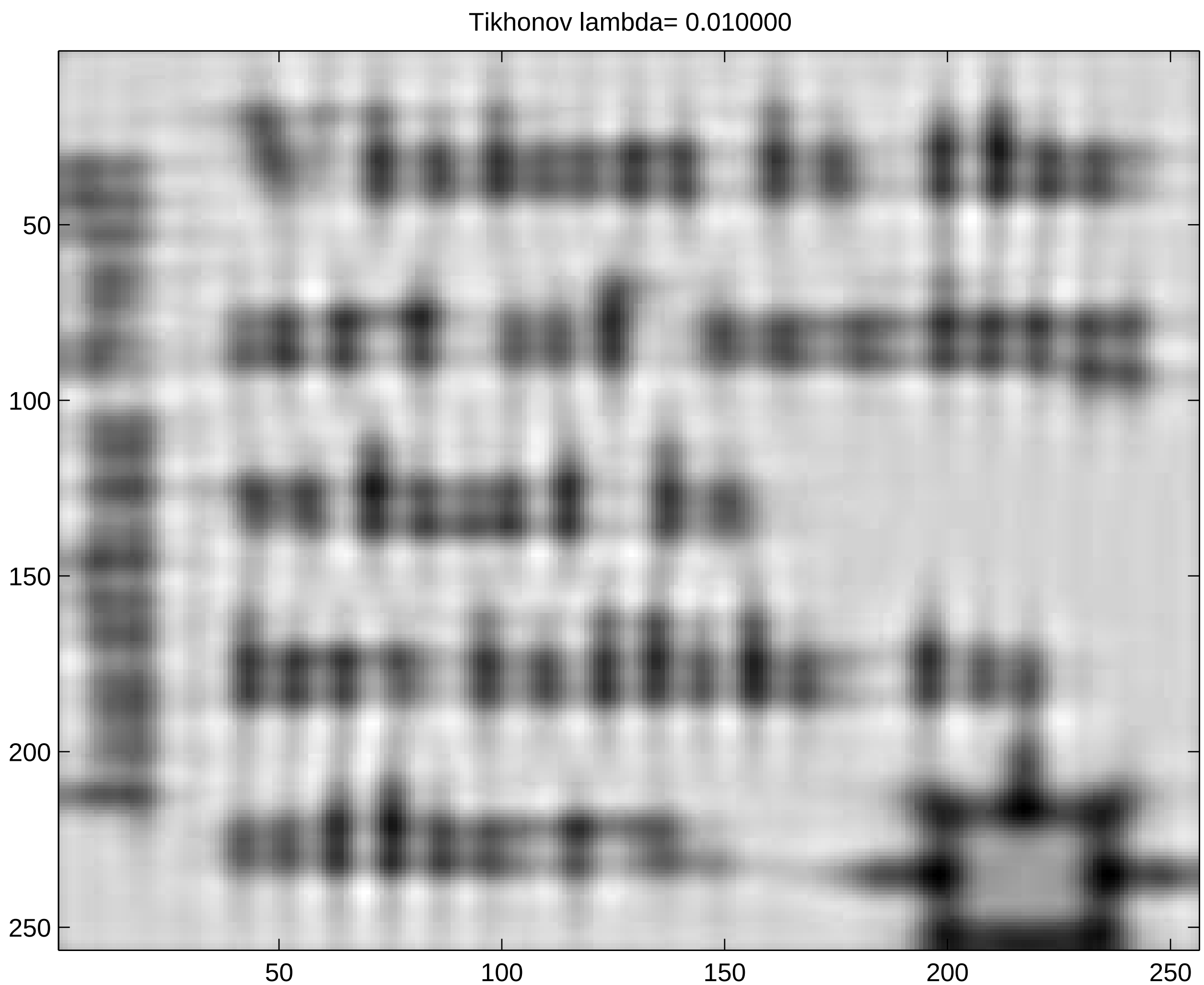
Tikhonov lambda= 0.001000



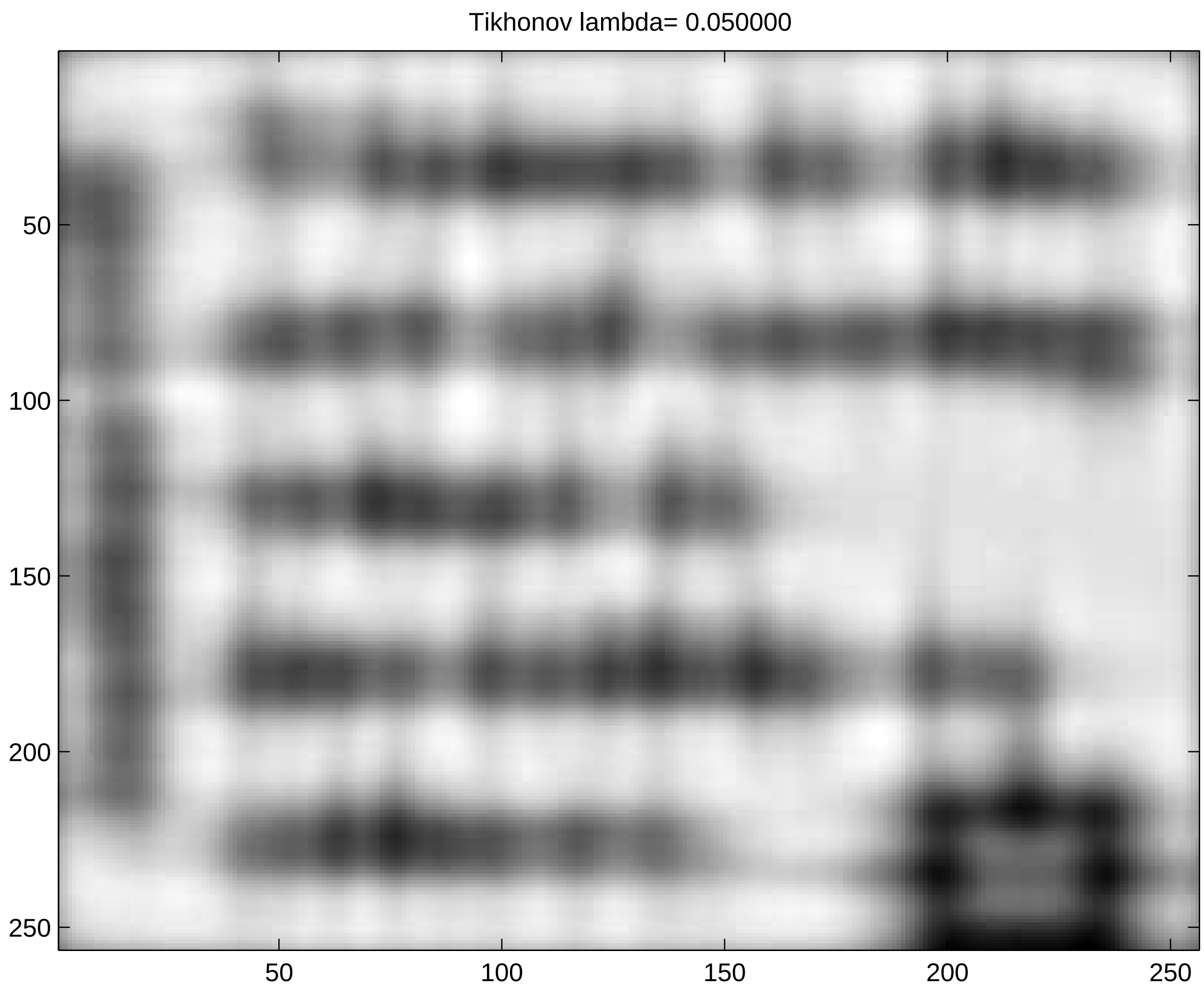

Tikhonov lambda= 0.001500
Vision is the
art of seeing
what is
invisible to
others.
50
100
150
200
250
50
100
150
200
250


Tikhonov lambda= 0.002500
Vision is the art of seeing what is invisible to others.
50
100
150
200
250
50
100
150
200
250


Tikhonov lambda= 0.005000
50
100
150
200
250
50
100
150
200
250



Tikhonov lambda= 0.010000
50
100
150
200
250
50
100
150
200
250


Tikhonov lambda= 0.050000
50
100
150
200
250
50
100
150
200
250

What we have demonstrated:

- The ill-conditioning of the matrix and the noise in the data make image deblurring very difficult.
- We can deblur well using Tikhonov regularization.
- Choosing the regularization parameter is easy if the “eye” norm can be used.
- In general, a good regularization parameter is hard to find.

Estimators, means, and variances

These estimates can be written as

$$\boldsymbol{V}^T \boldsymbol{x} = \boldsymbol{F}\boldsymbol{\Sigma}^\dagger \boldsymbol{U}^T \boldsymbol{b},$$

where $\boldsymbol{F}$ is an $n \times n$ diagonal matrix specific to each method.

The entries of $\boldsymbol{F}$ are called filter factors.

- Least squares: $\boldsymbol{F} = \boldsymbol{I}$.
- Tikhonov regularization:
$$f_j = \frac{\sigma_j^2}{\sigma_j^2 + \lambda^2}.$$
- TSVD: the first p diagonal entries of $\boldsymbol{F}$ are 1 and the others are zero.

The statistical distributions of our estimators

Estimator	$\boldsymbol{x} = \boldsymbol{V}\boldsymbol{F}\boldsymbol{\Sigma}^{\dagger}\boldsymbol{U}^T\boldsymbol{b}$	$\boldsymbol{r} = \boldsymbol{U}_2\boldsymbol{U}_2^T\boldsymbol{b} + \boldsymbol{U}_1(\boldsymbol{I}_n - \boldsymbol{F})\boldsymbol{U}_1^T\boldsymbol{b}$
Mean	$\boldsymbol{x}^* - \boldsymbol{V}(\boldsymbol{I}_n - \boldsymbol{F})\boldsymbol{V}^T\boldsymbol{x}^*$	$\boldsymbol{U}_1(\boldsymbol{I}_n - \boldsymbol{F})\boldsymbol{\Sigma}_1\boldsymbol{V}^T\boldsymbol{x}^*$
Variance	$\boldsymbol{V}(\boldsymbol{F}\boldsymbol{\Sigma}_1^{-1})^2\boldsymbol{V}^T$	$\boldsymbol{U}_2\boldsymbol{U}_2^T + \boldsymbol{U}_1(\boldsymbol{I}_n - \boldsymbol{F})^2\boldsymbol{U}_1^T$

Each of the estimators is normally distributed.

Regularization reduces variance in $\boldsymbol{x}$, increases bias.

The statistical properties of computed solutions

Behavior of the residual

Let $\tilde{\boldsymbol{x}}$ be an estimate of $\boldsymbol{x}^*$ and let

$$\tilde{\boldsymbol{r}} = \boldsymbol{b} - \boldsymbol{A}\tilde{\boldsymbol{x}}$$

be the corresponding residual vector. Since the regression model can also be written

$$\boldsymbol{\eta} = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^* ,$$

it is clear that $\tilde{\boldsymbol{x}}$ is acceptable only if $\tilde{\boldsymbol{r}}$ is a plausible sample from the distribution from which $\boldsymbol{\eta}$ is drawn.

Since

$$\mathcal{E}\left\{\|\boldsymbol{\eta}\|^2\right\} = m\ , \quad \mathsf{Var}\left\{\|\boldsymbol{\eta}\|^2\right\} = 2m\,,$$

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

$$m - \sqrt{2m} \le \|\boldsymbol{b} - \boldsymbol{A}\tilde{\boldsymbol{x}}\|^2 \le m + \sqrt{2m}$$

would be reasonable.

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

We will see how to use this information in choosing among a family of possible regularized estimates for $\boldsymbol{x}^*$.

Choosing a Regularization Parameter

The residual norm is

- a monotonically decreasing function of the regularization parameter p for the TSVD method.
- a monotonically increasing function of the regularization parameter λ for Tikhonov's method.

Similarly, the norm of the estimated solution is

- a monotonically increasing function of p
- a monotonically decreasing function of λ.

We discuss various methods for choosing the regularization parameter.

Well-Known Methods

- The most basic choice of regularization parameter is via Morozov's Discrepancy Principle (1966). This method chooses the regularization parameter so that the norm of the residual is approximately equal to its expected value.

- An alternate strategy is to choose λ or p to minimize the generalized cross-validation (GCV) function

$$G(\lambda) = \sum_{k=1}^{m} \left[b_k - \left(\boldsymbol{A} \tilde{\boldsymbol{x}}_\lambda^{(k)} \right)_k \right]^2 ,$$

where $\tilde{\boldsymbol{x}}_\lambda^{(k)}$ is the estimate when the kth measurement b_k is omitted. The basic idea, first introduced by Wahba (1977), is to choose λ to make $\boldsymbol{A}\tilde{\boldsymbol{x}}_\lambda$ the best overall predictor for missing data values.

- A third strategy, often used when the error cannot be assumed to be normally distributed, is based on the L-curve first introduced by Hanson and Lawson (1974) and further developed by P. C. Hansen (1992).

  In current use, the L-curve for Tikhonov regularization is a plot of $\log \| \tilde{\boldsymbol{x}}_\lambda \|$ versus $\log \| \boldsymbol{b} - \boldsymbol{A}\tilde{\boldsymbol{x}}_\lambda \|$ for a range of positive values of λ. The λ chosen corresponds to the point in the corner where the curvature of the curve is maximized.

  The L-curve for TSVD is defined in a similar way, but consists of discrete points for various values of p; curvature can be defined as the curvature of a smoothing cubic spline.

Useful statistical diagnostics

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

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

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

Diagnostic 3. We consider the elements of both $\boldsymbol{\eta}$ and $\tilde{\boldsymbol{r}}$ as time series, with the index i ($i = 1, \ldots, m$) taken to be the time variable. Since $\boldsymbol{\eta} \sim N(\mathbf{0}, \boldsymbol{I}_m)$, the η_i form a white noise series. Therefore the residuals $\tilde{\boldsymbol{r}}$ 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 $0 \le f \le \frac{1}{2T}$, 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 $\boldsymbol{z}$ of values corresponding to the frequencies k/NT, $k = 0, \ldots, N/2$.

Diagnostics: Cumulative Periodogram

The cumulative periodogram $\boldsymbol{c}$ 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.

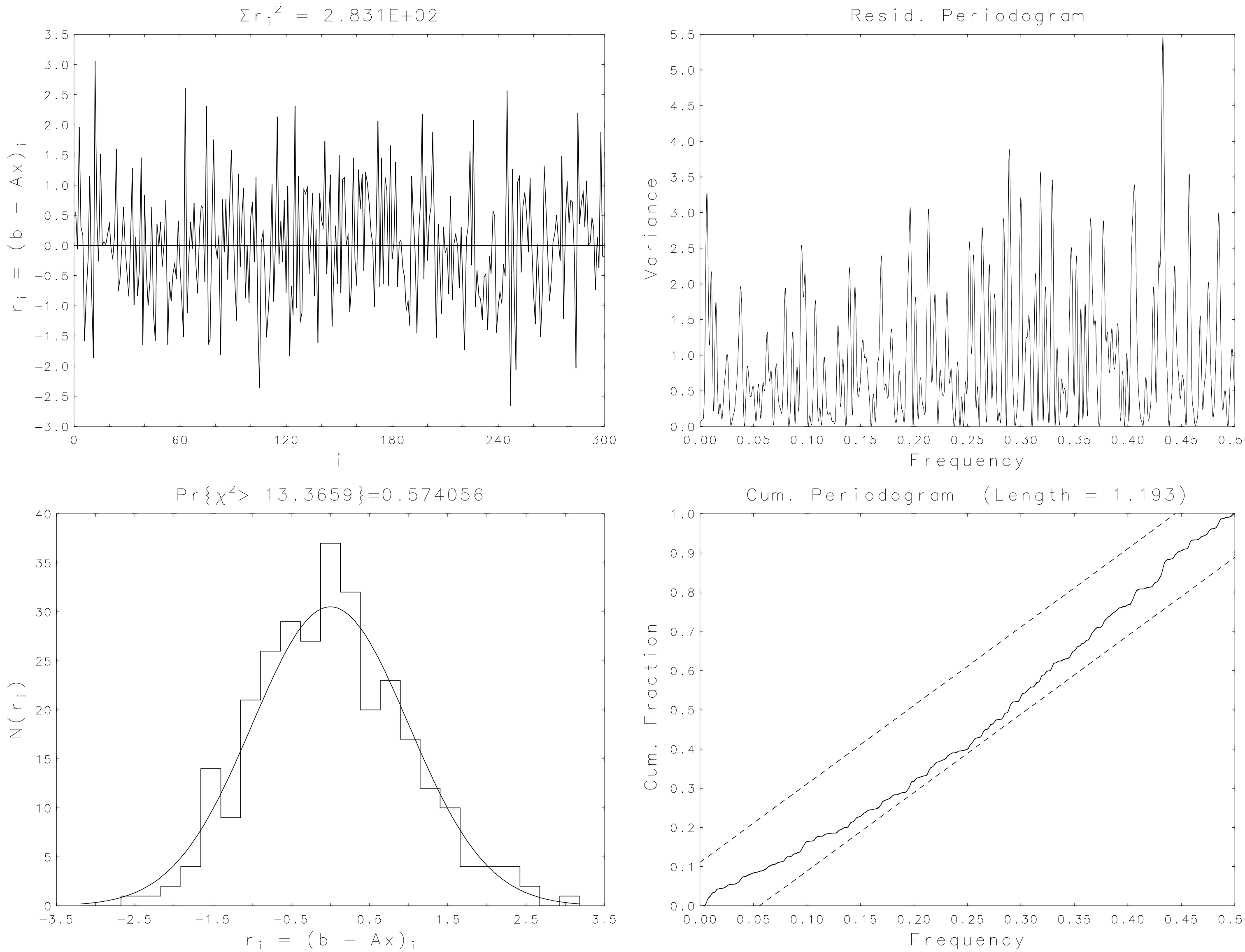
Some history

- 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 $\boldsymbol{c}$ 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.

Optimization and Validation using periodograms

Applying the diagnostics to the true solution of the modified Phillips problem



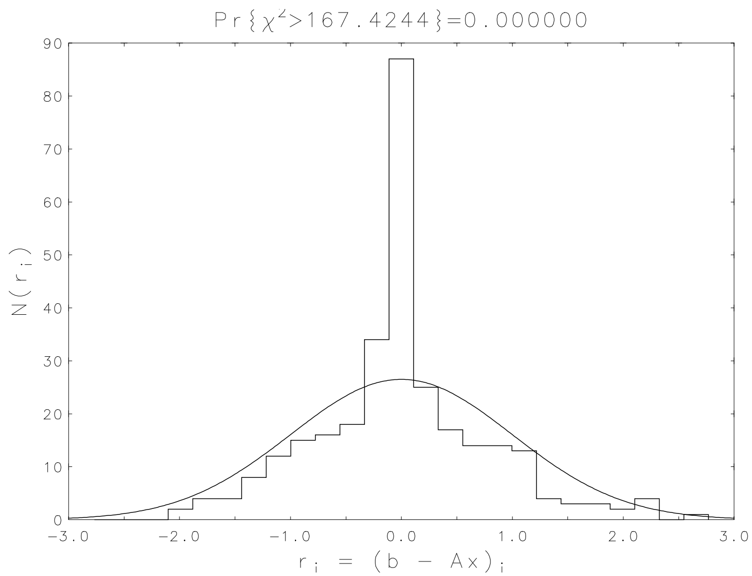
Σr_i² = 2.831E+02
r_i = (b − Ax)_i
i
Resid. Periodogram
Variance
Frequency
Pr{χ²> 13.3659}=0.574056
N(r_i)
r_i = (b − Ax)_i
Cum. Periodogram (Length = 1.193)
Cum. Fraction
Frequency

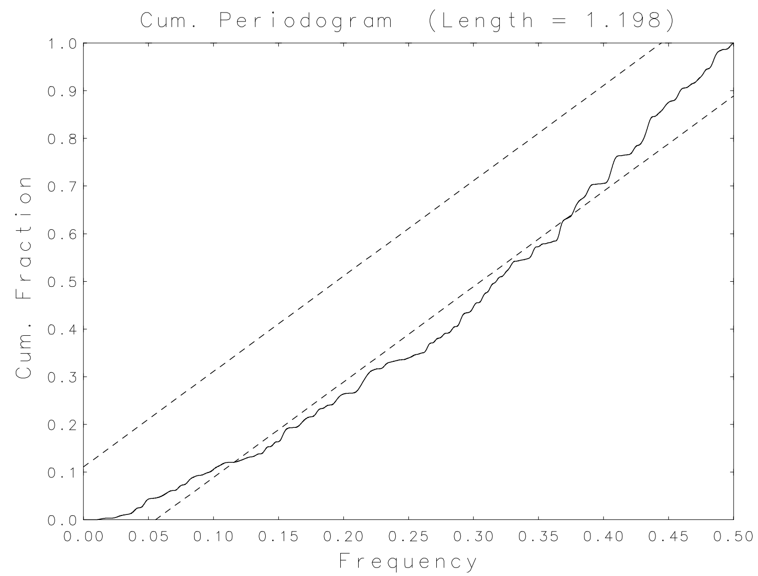
Applying the diagnostics to the least squares solution of the modified Phillips problem

The least squares solution is not reasonable:

- $\|\hat{\boldsymbol{r}}\|^2 = 43.01$, which is outside the $\pm 2\sigma$ confidence interval $[251.02, 348.98]$, so Diagnostic 1 is violated.
- The components of the residual $\hat{\boldsymbol{r}}$ do not look like samples from a $N(0,1)$ distribution, so Diagnostic 2 is violated.
- By considering plots of the periodogram and cumulative periodogram we see that Diagnostic 3 is violated.

Results for Tikhonov Regularization, L-curve parameter choice

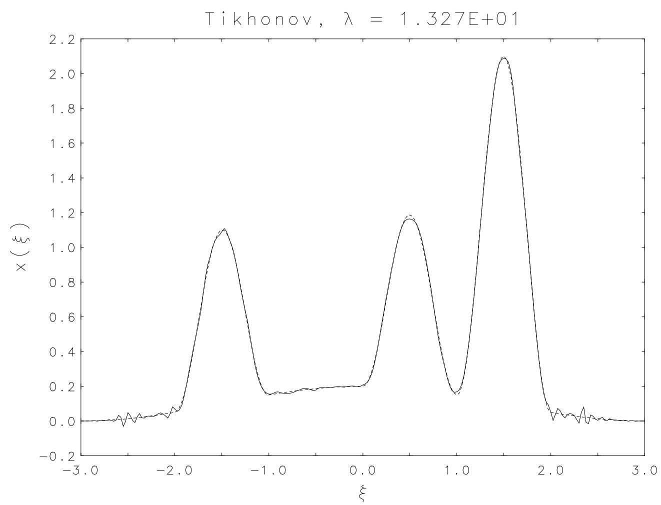


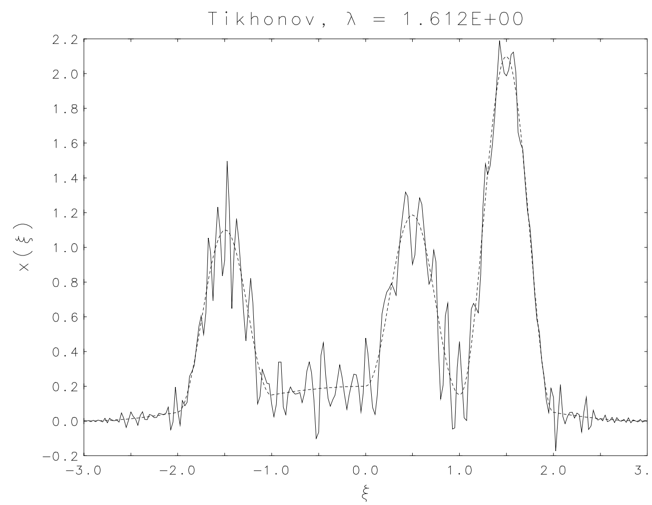


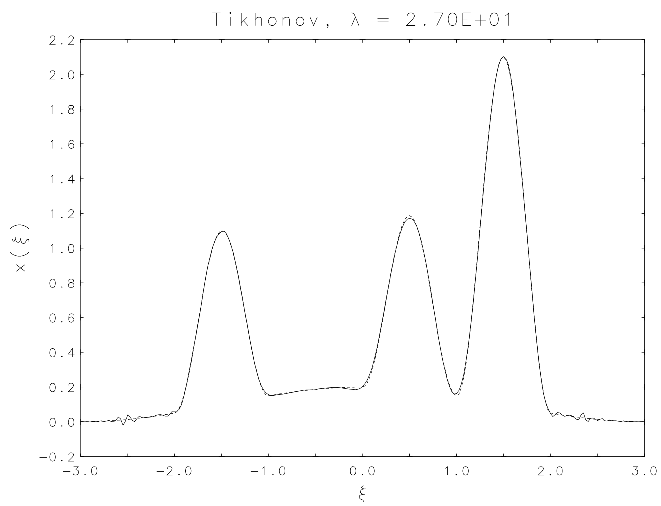
Diagnostics for various choices of λ

	Method	λ	$\sum(\boldsymbol{b} - \boldsymbol{A}\tilde{\boldsymbol{x}})_i^2$	$\%(c_k \in \mathcal{B}_{95})$	Length$\{\boldsymbol{c}\}$	$\lvert\Delta x\rvert_{rms}$
1	L-curve	1.61	177.3	48.4	1.1982	0.12131
2	GCV	13.27	214.1	62.3	1.2004	0.01271
3	Trial λ	19.20	233.1	95.3	1.1991	0.00948
4	Trial λ	20.00	236.8	99.1	1.1992	0.00925
5	Trial λ	22.00	247.5	100.0	1.1998	0.00880
6	Trial λ	24.00	260.7	100.0	1.2014	0.00853
7	Trial λ	26.00	276.9	100.0	1.2040	0.00841
8	Trial λ	27.00	286.3	96.9	1.2057	0.00839
9	Trial λ	28.00	296.6	87.5	1.2076	0.00841
10	Trial λ	29.00	307.9	83.9	1.2098	0.00844
11	Trial λ	30.00	320.2	76.8	1.2123	0.00850

Solution plots: GCV, L-curve, Ours







Results similar (but not as good) for TSVD.

Automating the Residual Periodogram Diagnostics

We have seen that the plots of the residual vector, its periodogram, and its cumulative periodogram, in conjunction with the three diagnostics lead to good choices of a regularization parameter, and we advise their use whenever possible.

However, in some circumstances (e.g., real-time applications), it is not feasible to choose the parameter manually, so it is also important to have automatic procedures.

There are several ways in which the periodogram can be used:

- Minimize the length of the cumulative periodogram.
- Minimize the number of points of the cumulative periodogram that lie outside the 95% confidence limits.
- Choose the smallest amount of regularization that places 95% of the points of the cumulative periodogram inside the confidence limits.
- Minimize the distance between the cumulative periodogram and the line through the origin with slope $2/NT$, using either the 1-, 2-, or ∞-norm to measure distance.
- Choose a value for which the maximum magnitude element in the periodogram is not significantly bigger than expected.

It is also a good idea to restrict our choice to one which puts the norm of the residual within 2 standard deviations of its expected value m (Diagnostic 1).

A sample automation

We experiment with last of the options for using the periodogram, by computing the ratio of the largest value to the average value.

The distribution of this quantity has been studied by Fisher and we chose the least amount of regularization for which the probability of a sample from this distribution being larger than our ratio was at least 5%.

Results on Some Standard Test Problems

We tested several parameter choice methods on some standard test problems taken from the MATLAB Regularization Tools: Baart, Foxgood, Heat, Ilaplace, Phillips, and Shaw.

We generated 100 white noise samples for each of the six problems with standard deviation equal to 0.001 times the norm of the right-hand side, and used $m = n = 256$.

The table indicates the number of problems for which the relative error in the regularized solution $\boldsymbol{x}_\lambda$ was less than 20%.

PFT	Discr.	GCV	L-Curve	HKK	Optimal
579	419	495	409	594	600
(97%)	(70%)	(83%)	(68%)	(99%)	(100%)

Validation using Confidence Images

Statistical framework:

- Assume that the noise η is normally distributed, mean zero, standard deviation $\boldsymbol{S}$.
- Note that the pixel values are constrained to be nonnegative, and also have upper bounds.

Theorem: The probability that $\boldsymbol{x}_i$, a component of $\boldsymbol{x}$, is contained in the interval $[\ell_i, u_i]$ is greater than or equal to α, where

$$\begin{aligned}
\ell_i &= \min\{\boldsymbol{x}_i : \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_{\boldsymbol{S}} \leq \mu,\ \boldsymbol{x} \geq 0\}\,, \\
u_i &= \max\{\boldsymbol{x}_i : \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_{\boldsymbol{S}} \leq \mu,\ \boldsymbol{x} \geq 0\}\,, \\
\beta &= \min_{\boldsymbol{x}} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2_{\boldsymbol{S}}\,, \\
\mu^2 &= \beta + \gamma^2\,, \\
\alpha &= \int_0^{\gamma^2} \chi^2(\rho)d\rho\,,
\end{aligned}$$

and χ^2 is the probability density function for the chi-squared distribution with the number of degrees of freedom equal to the rank of $\boldsymbol{A}$.

These confidence intervals also have **joint** probability α.

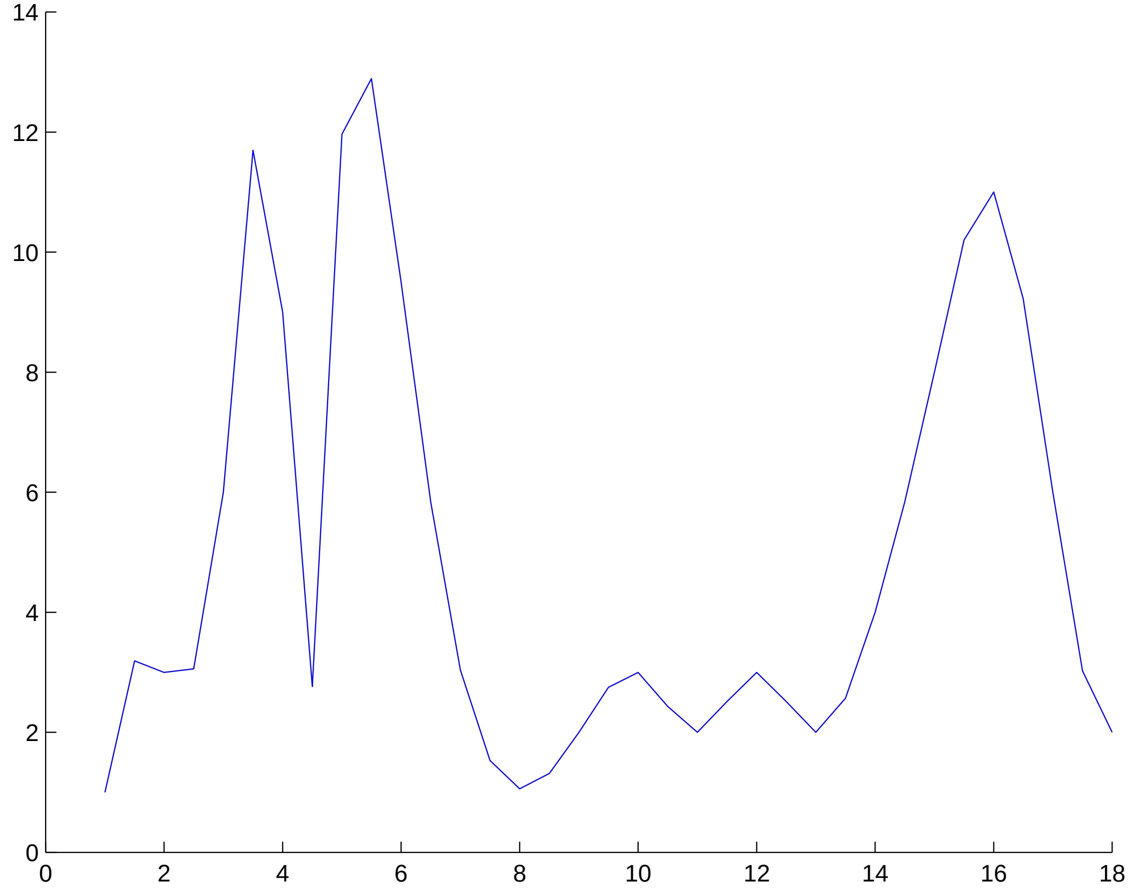
(Note: There is also a non-parametric form of the theorem.)

Therefore, confidence intervals for p pixel values can be computed, but at the cost of solving $2p$ constrained least squares problems.

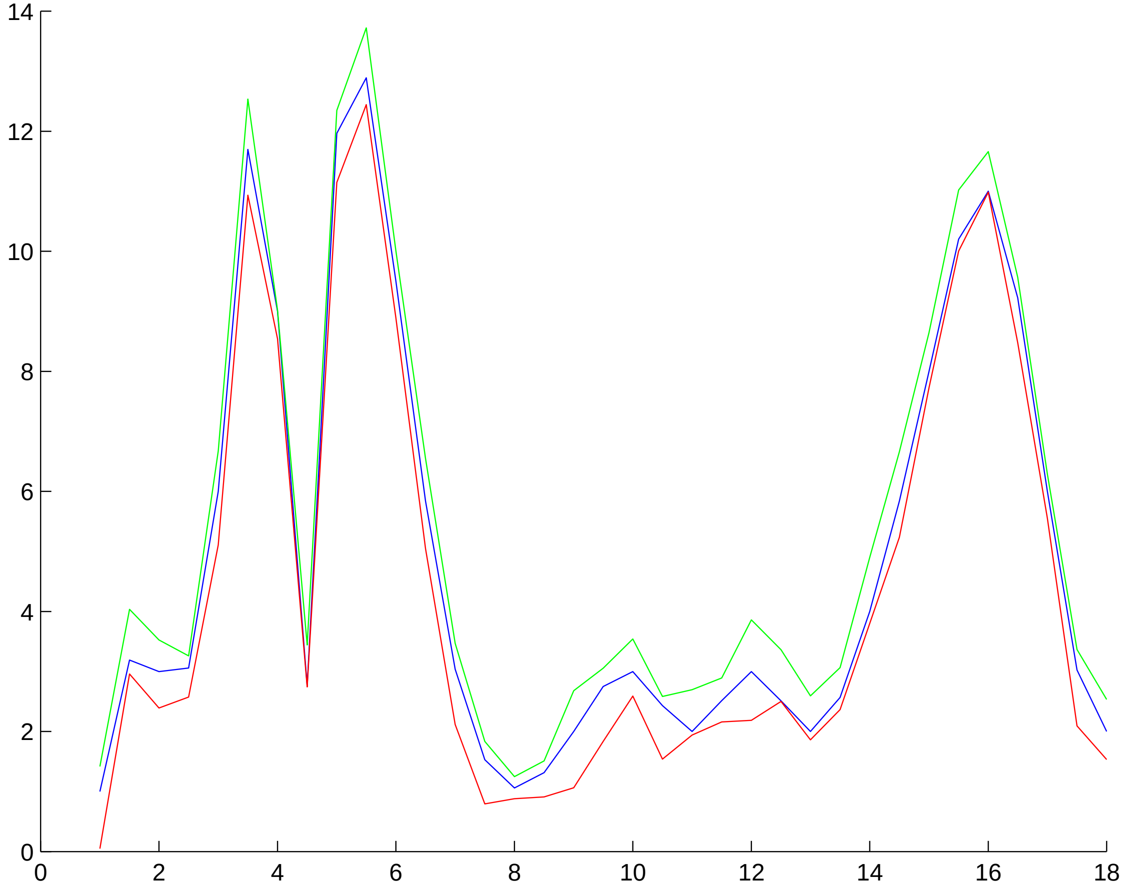
Expensive, but possible for modestly-sized subimages.

How to display confidence intervals

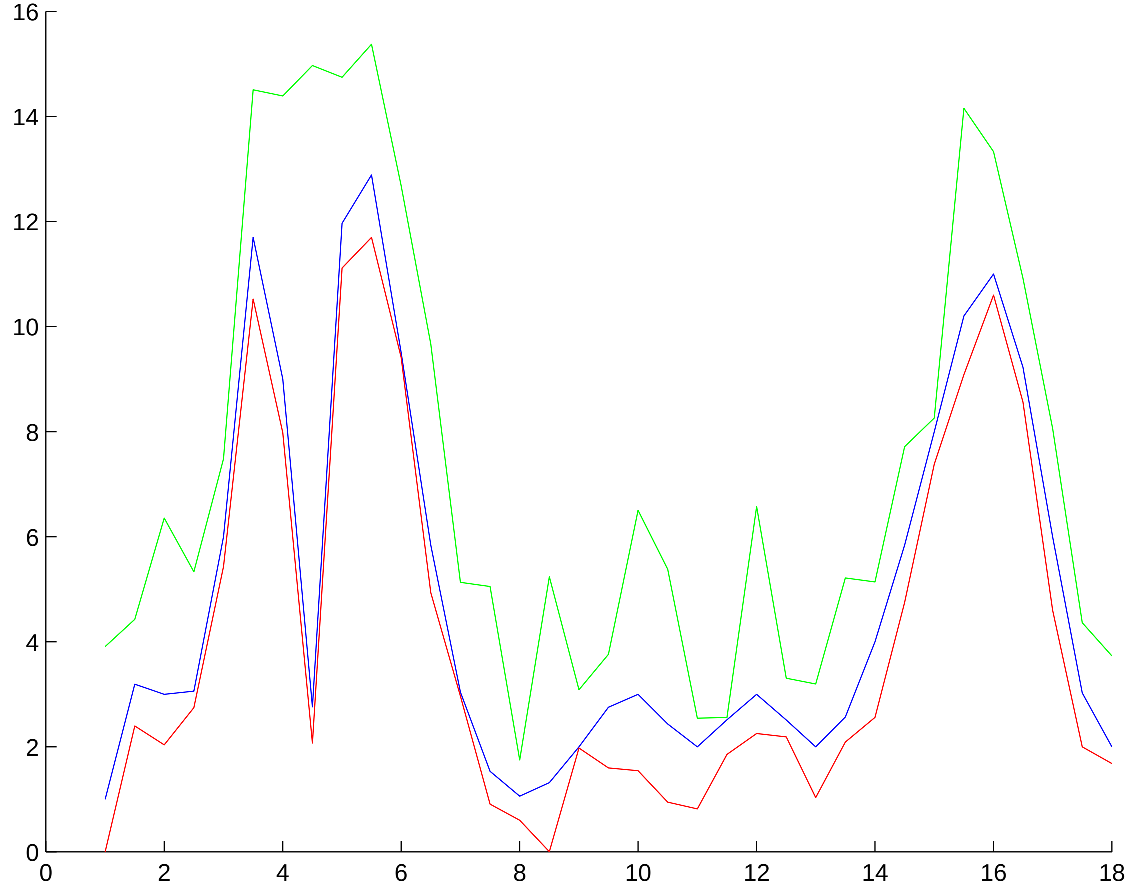
Suppose we solve a 1-dimensional problem and estimate the solution as:



Then we draw one conclusion if the confidence intervals are:



… and a different one if the intervals are:



How to display confidence images

But, once we compute upper and lower bounds on each pixel, how do we display that information in a useful way?

Idea: the **twinkle algorithm**.

Display a sequence of images, each one having a pixel value chosen uniformly from the interval $[\ell_i, u_i]$.

Features that are *persistent* throughout the display have probability at least α of being real.

Features that *flicker* due to wide confidence intervals quite possibly are artifacts.

An example of Confidence Images: Motion blur

Consider the image taken by a speed-limit-enforcing camera.

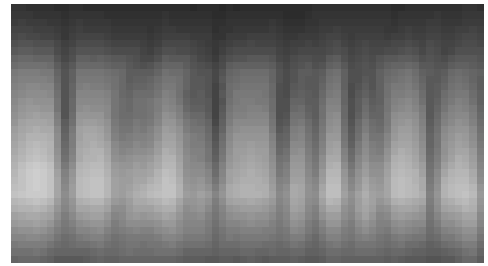
The image may be contaminated by motion blur, so we investigate the uncertainties involved in processing such images.

We used a digital camera to image a stationary white van. We applied a vertical blur, spreading each pixel to the adjacent 19 pixels, and then added a noise image with each component chosen from a normal distribution and with the norm of the noise equal to 6x10^{-5} times the norm of the image.

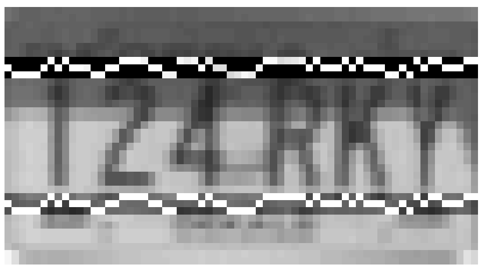
We then cropped the image to size 36 x 66, to isolate the license plate.

Original subimage

Blurred subimage

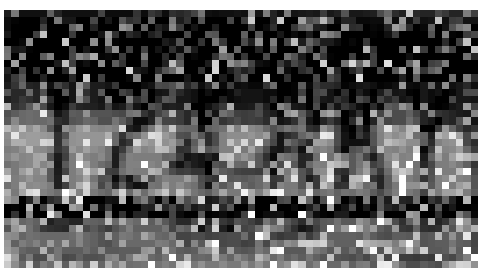


Reconstructed subimage, using nonnegativity

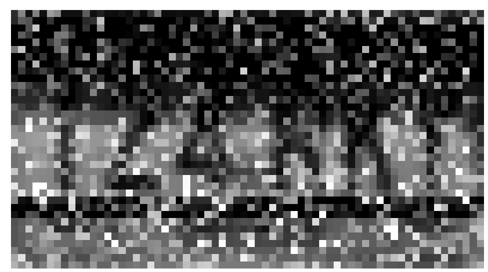


Reconstructed subimage, using [0,1] bounds

Confidence image, using [0,1] bounds



Confidence image, using [0,1] bounds



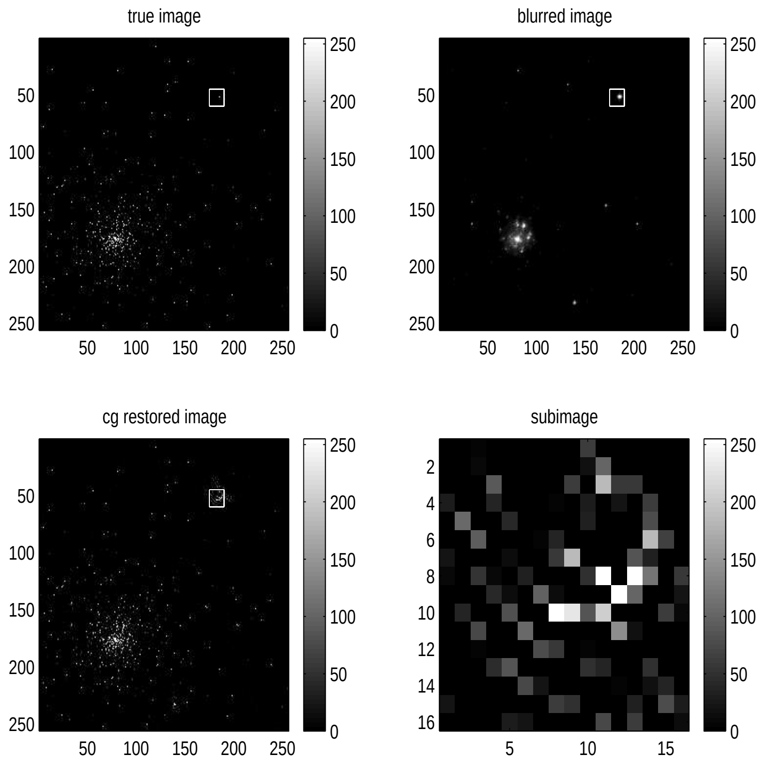
`licensedemo`

Example

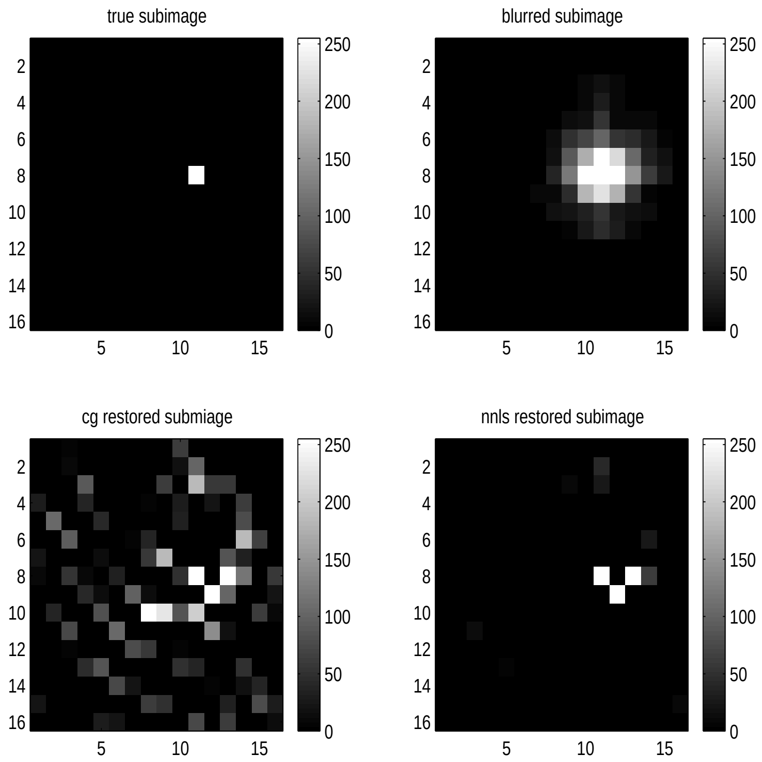
This example uses the astronomical image presented earlier.

- The blurring function A is spacially varying, but we don't have exact coefficients for it. Instead, we approximate it as either spacially constant over the entire picture, or spacially constant over each of the four quadrants of the picture, and use measured values for the coefficients.
- We reconstruct the entire image using preconditioned conjugate gradients.
- Then we focus on a subimage containing a bright object and use nonnegativity constraints to try to get a more reliable reconstruction.

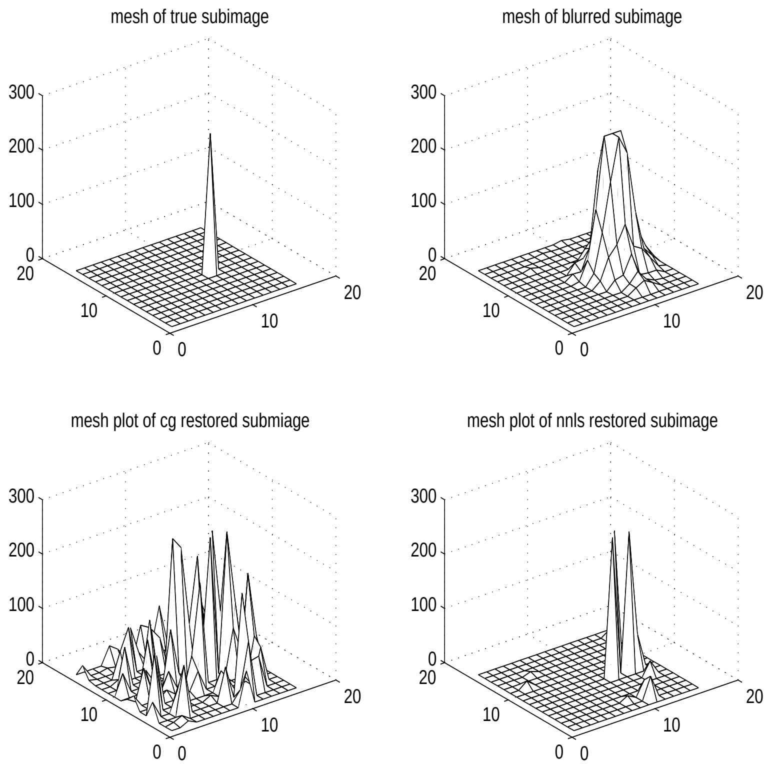
Satellite data, with restoration using one PSF in PCG



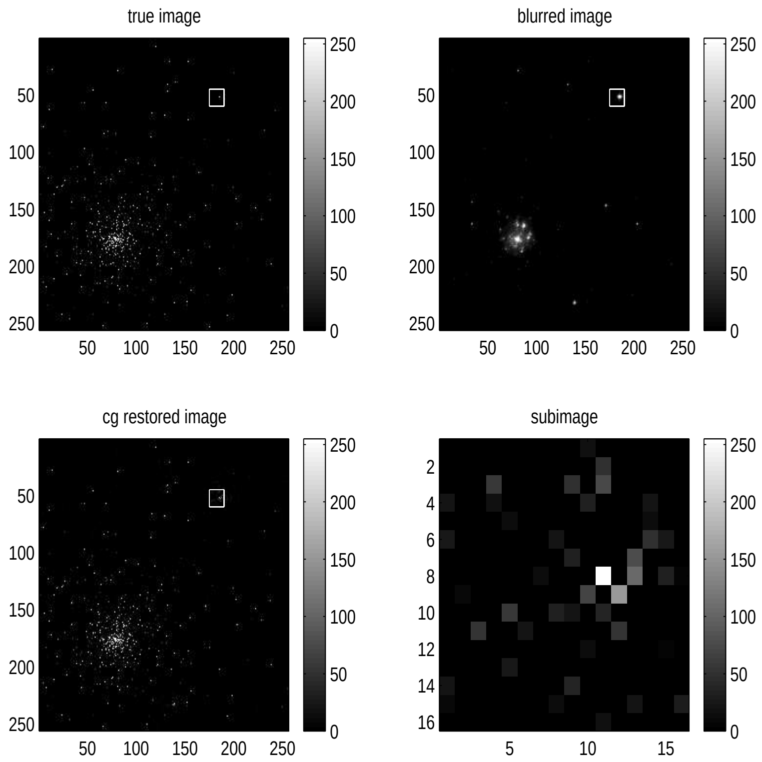
Subimages, using one PSF for PCG



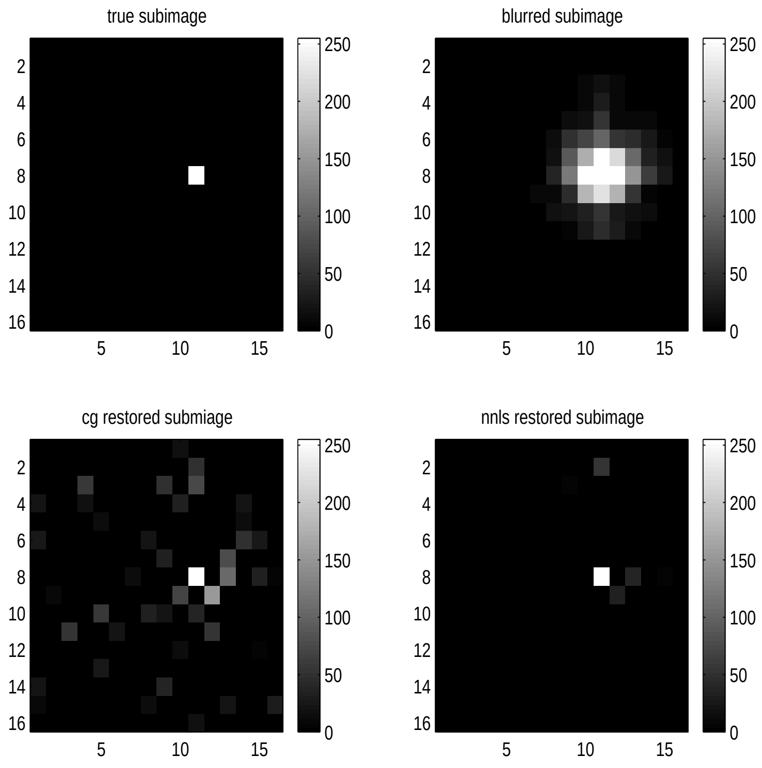
Mesh plot of subimages, using one PSF for PCG.



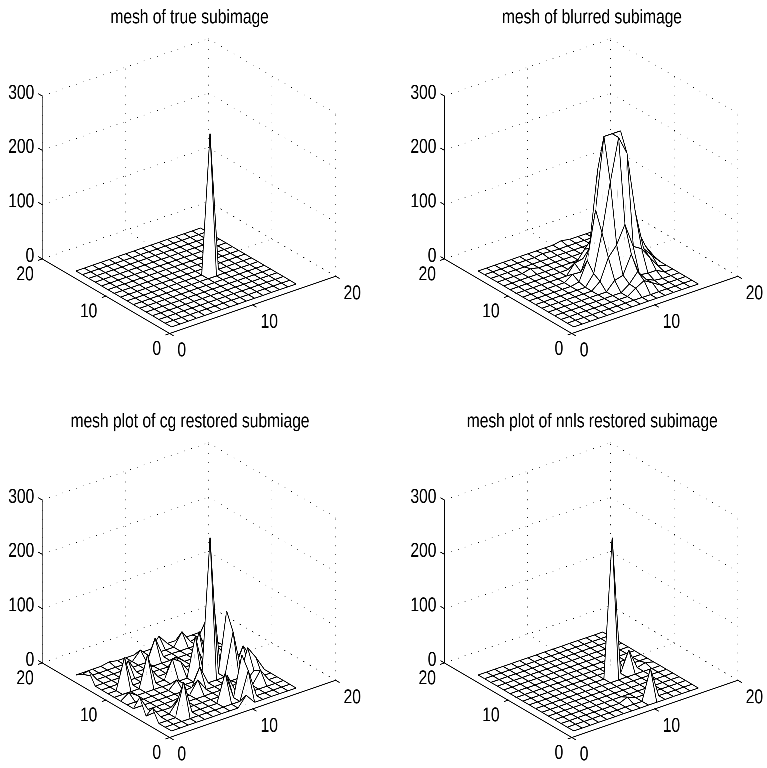
Satellite data, with restoration using four PSFs in PCG.



Subimages, using four PSFs for PCG



Mesh plot of subimages, using four PSFs for PCG.



Bottom line

By working with subimages, we can afford to do better reconstructions.

But, more importantly, we can afford to compute *confidence intervals* for the pixel values.

Conclusions

- We have shown how to use optimization in order to produce high-quality reconstructions.
- We have used *confidence images* to provide information on the reliability of the reconstructions.
- We have provided the means to compute confidence intervals using Braketls, and to display confidence images using Twinkle.

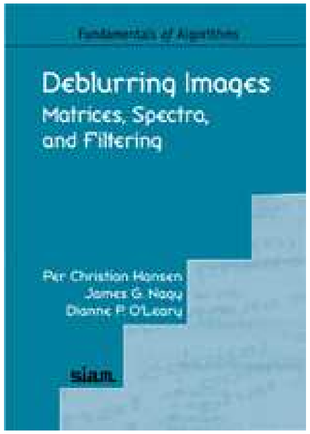
References

Bert W. Rust and Dianne P. O'Leary, "Residual Periodograms for Choosing Regularization Parameters for Ill-Posed Problems," *Inverse Problems,* Vol. 24, 034005 (30 pages) May 2008.

Dianne P. O'Leary and B. W. Rust, "Confidence intervals for inequality-constrained least squares problems, with applications to ill-posed problems," *SIAM Journal on Scientific and Statistical Computing* 7 (1986) 473-489.

James G. Nagy and Dianne P. O'Leary, "Image Restoration through Subimages and Confidence Images," *Electronic Transactions on Numerical Analysis*, 13 (2002) 22-37.

Per Christian Hansen, James G. Nagy, and Dianne P. O'Leary, Deblurring Images: Matrices, Spectra, and Filtering, SIAM Press, Philadelphia, 2006.



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