

Regularization Parameter Selection Methods for Ill-Posed Poisson Maximum Likelihood Estimation

Johnathan M. Bardsley and John Goldes

Department of Mathematical Sciences

University of Montana

Missoula, Montana 59812

USA

E-mail: bardsleyj@mso.umt.edu, john.goldes@umontana.edu

Abstract. In image processing applications, image intensity is often measured via the counting of incident photons emitted by the object of interest. In such cases, image data-noise is accurately modeled by a Poisson distribution. This motivates the use of Poisson maximum likelihood estimation for image reconstruction. However, when the underlying model equation is ill-posed, regularization is needed. Regularized Poisson likelihood estimation has been studied extensively by the authors, though a problem of high importance remains: the choice of the regularization parameter. We will present three statistically motivated methods for choosing the regularization parameter, and numerical examples will be presented to illustrate their effectiveness.

MSC numbers: 15A29, 65F22, 94A08

Keywords: regularization parameter selection methods, Poisson maximum likelihood estimation, ill-posed problems, image reconstruction.

1. Introduction

A Poisson data-noise model arises in many inverse problems applications. In image processing, in particular, a CCD camera is often used to measure image intensity via the counting of incident photons. Counting processes are known to have error that is well-modeled by a Poisson distribution, and so a Poisson data-noise model is natural in this setting [16].

If the correct noise model is taken into account, the maximum likelihood estimator of the underlying true image is the minimizer of a negative-log Poisson likelihood function (with a penalty term when the underlying problem is sufficiently ill-posed). The computation of this estimator is made difficult by the fact that a constraint is needed and that the cost function is non-quadratic.

The approach taken by many researchers is to ignore the Poisson data-noise model and compute instead the minimizer of a least squares function (with penalty if needed) [18]. This approach is attractive for a number of reasons. First, for images with high and smoothly varying intensities, the results are as good as when the correct noise model is assumed [17]. Perhaps more importantly, however, is that fact that regularized least squares problems have been studied extensively, both from the theoretical and computational points of view. So the user has many tools at his or her disposal.

On the other hand, there is sometimes benefit to be gained by using the correct noise model, and recent work of the authors has shown that efficient and convergent computational methods exist for the resulting minimization problem [8, 6]. Moreover, a rigorous theoretical (functional analytic) justification of the approach, akin to that developed for least squares problems, has been developed [2, 3, 4].

But one of the most important problems remains: the development of methods for choosing the regularization parameter. This problem has been extensively studied in the context of least squares fit-to-data. However, it has not been addressed for negative-log Poisson likelihood estimation. That is our focus in this paper.

1.1. Preliminary Mathematical Development

In this subsection, we present the mathematical, statistical, and computational models of focus. We begin with the mathematical model.

The following problem is very common in imaging science: given a blurred, noisy $N \times N$ image array \mathbf{z} , obtain an estimate of the underlying $N \times N$ true object array \mathbf{u}_e by approximately solving a linear system of the form

$$\mathbf{z} = \mathbf{A}\mathbf{u} + \boldsymbol{\gamma}, \tag{1}$$

where \mathbf{z} has been column-stacked so that it is $N^2 \times 1$; \mathbf{A} is a known, and often ill-conditioned, $N^2 \times N^2$ matrix; and $\boldsymbol{\gamma}$ corresponds to the known, positive background intensity of the image being viewed. Moreover, we assume that $\mathbf{A}\mathbf{u} \geq \mathbf{0}$ for all $\mathbf{u} \geq \mathbf{0}$. For the remainder of the manuscript, we will use the convention $n \stackrel{\text{def}}{=} N^2$.

In both astronomical and medical imaging applications, the elements of \mathbf{z} correspond to noisy photon counts. A statistical model that is used to model stochastic errors in both astronomical and medical imaging [14, 16] is

$$\mathbf{z} = \text{Pois}(\mathbf{A}\mathbf{u}_e + \boldsymbol{\gamma}), \quad (2)$$

where $\text{Pois}(\boldsymbol{\lambda})$ is an i.i.d. Poisson random vector with Poisson parameter vector $\boldsymbol{\lambda}$. (See [5, 16] for a detailed discussion of this statistical model in the astronomical imaging case.)

The probability density function for the data \mathbf{z} given (2) has the form

$$p(\mathbf{z}; \mathbf{u}) := \prod_{i=1}^n \frac{([\mathbf{A}\mathbf{u}]_i + \gamma_i)^{z_i} \exp[-([\mathbf{A}\mathbf{u}]_i + \gamma_i)]}{z_i!}, \quad (3)$$

where $[\mathbf{A}\mathbf{u}]_i$, γ_i , and z_i are the i th components of $\mathbf{A}\mathbf{u}$, $\boldsymbol{\gamma}$, and \mathbf{z} , respectively.

Given image data \mathbf{z} arising from model (2), the maximum likelihood estimate of \mathbf{u}_e is obtained by maximizing $p(\mathbf{z}; \mathbf{u})$ with respect to \mathbf{u} , subject to the constraint $\mathbf{u} \geq \mathbf{0}$, or more commonly, by solving

$$\mathbf{u}_{\text{ML}} = \arg \min_{\mathbf{u} \geq \mathbf{0}} \left\{ T_0(\mathbf{u}; \mathbf{z}) \stackrel{\text{def}}{=} \sum_{i=1}^n \{([\mathbf{A}\mathbf{u}]_i + \gamma_i) - z_i \ln([\mathbf{A}\mathbf{u}]_i + \gamma_i)\} \right\}. \quad (4)$$

Note that T_0 is equal, up to an additive constant, to $-\ln p(\mathbf{z}; \mathbf{u})$.

As is the case for least squares estimation, solutions of (4) can be noise-corrupted when \mathbf{A} is ill-conditioned, in which case regularization is needed. This can be motivated statistically within a Bayesian setting. In particular, a prior probability density $p(\mathbf{u})$ for \mathbf{u} is specified and the posterior density

$$p(\mathbf{u}; \mathbf{z}) := \frac{p(\mathbf{z}; \mathbf{u})p(\mathbf{u})}{p(\mathbf{z})}, \quad (5)$$

given by Bayes' Law, is maximized with respect to \mathbf{u} . The maximizer of the posterior density function $p(\mathbf{u}; \mathbf{z})$ is called the maximum a posteriori (MAP) estimator. We note that maximizing (5) with respect to \mathbf{u} is equivalent to minimizing

$$T(\mathbf{u}) = T_0(\mathbf{u}; \mathbf{z}) - \ln p(\mathbf{u}). \quad (6)$$

The function $-\ln p(\mathbf{u})$ is the regularization term from classical inverse problems. Thus we see that in using the Bayesian formulation above, a statistically rigorous interpretation of regularization follows. Note, in particular, that $p(\mathbf{u})$ is the probability density function – known as the *prior* – from which the unknown \mathbf{u} is assumed to arise.

In this paper, we will consider regularization functions (negative-log priors) of the form

$$-\ln p(\mathbf{u}) = \frac{\alpha}{2} \langle \mathbf{C}\mathbf{u}, \mathbf{u} \rangle, \quad (7)$$

where $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product, and $\mathbf{C} \in \mathbb{R}^{N \times N}$ is symmetric, positive semi-definite. Equation (7) corresponds to the statistical assumption that the prior probability density $p(\mathbf{u})$ is a (possibly degenerate) Gaussian with mean $\mathbf{0}$ and covariance matrix $\alpha^{-1}\mathbf{C}^\dagger$, where “ \dagger ” denotes pseudo-inverse.

Thus the computational problem of interest to us has the form

$$\mathbf{u}_\alpha = \arg \min_{\mathbf{u} \geq \mathbf{0}} \left\{ T_\alpha(\mathbf{u}) \stackrel{\text{def}}{=} T_0(\mathbf{u}; \mathbf{z}) + \frac{\alpha}{2} \langle \mathbf{C}\mathbf{u}, \mathbf{u} \rangle \right\}. \quad (8)$$

In this setting, α is known as the regularization parameter. We make the additional assumption that the null-spaces of \mathbf{A} and \mathbf{C} intersect only trivially so that T_α is a strictly convex function. In this case, \mathbf{u}_α exists and is unique [8]. Methods for estimating the value of α in (8) have not been developed and are the focus of this paper.

The remainder of the paper is organized as follows. In Section 2, three regularization parameter selection methods are presented. Then in Section 3 they are tested, and we end with conclusions.

2. Regularization Parameter Choice Methods

In this section, we motivate three new regularization parameter selection methods for use in conjunction with (8).

2.1. A Weighted Least Squares Approximation of $T_0(\mathbf{u}, \mathbf{z})$

Existing methods for estimating the regularization parameter in the least squares case can not be directly applied to our problem. However, in [7], a Taylor series argument is used to obtain a quadratic approximation of T_0 . Using this approximation, existing regularization parameter selection methods can be extended to (8). We present a corrected and extended derivation of this approximation now.

First, we compute various derivatives of T_0 . The gradient and Hessian of T_0 with respect to \mathbf{u} are given by

$$\nabla_{\mathbf{u}} T_0(\mathbf{u}; \mathbf{z}) = \mathbf{A}^T \left(\frac{\mathbf{A}\mathbf{u} - (\mathbf{z} - \gamma)}{\mathbf{A}\mathbf{u} + \gamma} \right), \quad (9)$$

$$\nabla_{\mathbf{u}\mathbf{u}}^2 T_0(\mathbf{u}; \mathbf{z}) = \mathbf{A}^T \text{diag} \left(\frac{\mathbf{z}}{(\mathbf{A}\mathbf{u} + \gamma)^2} \right) \mathbf{A}, \quad (10)$$

where division – here and for the remainder of the manuscript – is computed component-wise. The gradient and Hessian of T_0 with respect to \mathbf{z} are given by

$$\nabla_{\mathbf{z}} T_0(\mathbf{u}; \mathbf{z}) = -\ln(\mathbf{A}\mathbf{u} + \gamma), \quad (11)$$

$$\nabla_{\mathbf{z}\mathbf{z}}^2 T_0(\mathbf{u}; \mathbf{z}) = \mathbf{0}. \quad (12)$$

The second order mixed partial derivatives of T_0 are given by

$$\nabla_{\mathbf{u}\mathbf{z}}^2 T_0(\mathbf{u}; \mathbf{z}) = -\mathbf{A}^T \text{diag} \left(\frac{\mathbf{1}}{\mathbf{A}\mathbf{u} + \gamma} \right), \quad (13)$$

$$\nabla_{\mathbf{z}\mathbf{u}}^2 T_0(\mathbf{u}; \mathbf{z}) = -\text{diag} \left(\frac{\mathbf{1}}{\mathbf{A}\mathbf{u} + \gamma} \right) \mathbf{A}. \quad (14)$$

Now, let \mathbf{u}_e be the exact object and $\mathbf{z}_e \stackrel{\text{def}}{=} \mathbf{A}\mathbf{u}_e + \gamma$ the background shifted exact data. Then, letting $\mathbf{k} = \mathbf{z} - \mathbf{z}_e$ and $\mathbf{h} = \mathbf{u} - \mathbf{u}_e$ and expanding T_0 in a Taylor series

about \mathbf{u}_e and \mathbf{z}_e , we obtain from (9)-(14)

$$\begin{aligned}
T_0(\mathbf{u}; \mathbf{z}) &= T_0(\mathbf{u}_e + \mathbf{h}; \mathbf{z}_e + \mathbf{k}), \\
&= T_0(\mathbf{u}_e; \mathbf{z}_e) + \mathbf{k}^T \nabla_{\mathbf{z}} T_0(\mathbf{u}_e; \mathbf{z}_e) + \frac{1}{2} \mathbf{h}^T \nabla_{\mathbf{u}}^2 T_0(\mathbf{u}_e; \mathbf{z}_e) \mathbf{h} \\
&\quad + \frac{1}{2} \mathbf{h}^T \nabla_{\mathbf{u} \mathbf{z}}^2 T_0(\mathbf{u}_e; \mathbf{z}_e) \mathbf{k} + \frac{1}{2} \mathbf{k}^T \nabla_{\mathbf{z} \mathbf{u}}^2 T_0(\mathbf{u}_e; \mathbf{z}_e) \mathbf{h} \\
&\quad + \mathcal{O}(\|\mathbf{h}\|_2^3, \|\mathbf{h}\|_2^2 \|\mathbf{k}\|_2, \|\mathbf{h}\|_2 \|\mathbf{k}\|_2^2, \|\mathbf{k}\|_2^3) \\
&= \sum_{i=1}^n \{[\mathbf{A} \mathbf{u}_e]_i - [\mathbf{z}_e]_i \ln[\mathbf{A} \mathbf{u}_e]_i\} - (\mathbf{z} - \mathbf{z}_e)^T \ln(\mathbf{A} \mathbf{u}_e) \\
&\quad + \frac{1}{2} (\mathbf{A} \mathbf{u} - \mathbf{A} \mathbf{u}_e)^T \text{diag} \left(\frac{\mathbf{1}}{\mathbf{A} \mathbf{u}_e + \gamma} \right) (\mathbf{A} \mathbf{u} - \mathbf{A} \mathbf{u}_e) \\
&\quad - \frac{1}{2} (\mathbf{z} - \mathbf{z}_e)^T \text{diag} \left(\frac{\mathbf{1}}{\mathbf{A} \mathbf{u}_e + \gamma} \right) (\mathbf{A} \mathbf{u} - \mathbf{A} \mathbf{u}_e) \\
&\quad - \frac{1}{2} (\mathbf{A} \mathbf{u} - \mathbf{A} \mathbf{u}_e)^T \text{diag} \left(\frac{\mathbf{1}}{\mathbf{A} \mathbf{u}_e + \gamma} \right) (\mathbf{z} - \mathbf{z}_e) \\
&\quad + \mathcal{O}(\|\mathbf{h}\|_2^3, \|\mathbf{h}\|_2^2 \|\mathbf{k}\|_2, \|\mathbf{h}\|_2 \|\mathbf{k}\|_2^2, \|\mathbf{k}\|_2^3) \\
&= T(\mathbf{u}_e; \mathbf{z}) + \frac{1}{2} (\mathbf{A} \mathbf{u} - (\mathbf{z} - \gamma))^T \text{diag} \left(\frac{\mathbf{1}}{\mathbf{z}_e} \right) (\mathbf{A} \mathbf{u} - (\mathbf{z} - \gamma)) \\
&\quad + \mathcal{O}(\|\mathbf{h}\|_2^3, \|\mathbf{h}\|_2^2 \|\mathbf{k}\|_2, \|\mathbf{h}\|_2 \|\mathbf{k}\|_2^2, \|\mathbf{k}\|_2^3). \tag{16}
\end{aligned}$$

The finally equality was obtained using the linearity of the inner-product and the fact that $\mathbf{z}_e = \mathbf{A} \mathbf{u}_e + \gamma$. Thus the quadratic Taylor series approximation of $T(\mathbf{u}; \mathbf{z})$ about the points $(\mathbf{u}_e; \mathbf{z}_e)$ —modulo an additive constant—is given by

$$\frac{1}{2} (\mathbf{A} \mathbf{u} - (\mathbf{z} - \gamma))^T \text{diag} \left(\frac{\mathbf{1}}{\mathbf{z}_e} \right) (\mathbf{A} \mathbf{u} - (\mathbf{z} - \gamma)). \tag{17}$$

But (17) is unusable due the the presence of the exact data \mathbf{z}_e . However, one more application of Taylor's theorem will give us a useable approximation, and without a decrease in the order of accuracy. Let $\mathbf{r} = \mathbf{A} \mathbf{u} - (\mathbf{z} - \gamma)$, then (17) can be rewritten

$$\begin{aligned}
\frac{1}{2} \mathbf{r}^T \left[\mathbf{r} \odot \left(\frac{\mathbf{1}}{\mathbf{z} - \mathbf{k}} \right) \right] &= \frac{1}{2} \mathbf{r}^T \left[\mathbf{r} \odot \left(\frac{\mathbf{1}}{\mathbf{z}} + \text{diag} \left(\frac{\mathbf{1}}{\mathbf{z}^2} \right) \hat{\mathbf{k}} \right) \right] \\
&= \frac{1}{2} \mathbf{r}^T \left(\frac{\mathbf{r}}{\mathbf{z}} \right) + \frac{1}{2} \mathbf{r}^T \left(\frac{\mathbf{r} \odot \hat{\mathbf{k}}}{\mathbf{z}^2} \right),
\end{aligned}$$

where, via Taylor's theorem, $0 < |\hat{k}_i| < |k_i|$ for all i . Noting that $\mathbf{r} = \mathbf{A} \mathbf{h} - \mathbf{k}$ and that \mathbf{z} is bounded away from zero, we obtain

$$\frac{1}{2} \mathbf{r}^T \left(\frac{\mathbf{r} \odot \hat{\mathbf{k}}}{\mathbf{z}^2} \right) = \mathcal{O}(\|\mathbf{h}\|_2^2 \|\mathbf{k}\|_2, \|\mathbf{h}\|_2 \|\mathbf{k}\|_2^2, \|\mathbf{k}\|_2^3),$$

Thus finally, recalling (16), we have the approximation

$$T_0(\mathbf{u}; \mathbf{z}) = T_0(\mathbf{u}_e; \mathbf{z}) + T_0^{\text{wls}}(\mathbf{u}; \mathbf{z}) + \mathcal{O}(\|\mathbf{h}\|_2^3, \|\mathbf{h}\|_2^2 \|\mathbf{k}\|_2, \|\mathbf{h}\|_2 \|\mathbf{k}\|_2^2, \|\mathbf{k}\|_2^3), \tag{18}$$

where

$$T_0^{\text{wls}}(\mathbf{u}; \mathbf{z}) \stackrel{\text{def}}{=} \frac{1}{2} \|\mathbf{Z}^{-1/2}(\mathbf{A}\mathbf{u} - (\mathbf{z} - \boldsymbol{\gamma}))\|^2, \quad \mathbf{Z} = \text{diag}(\mathbf{z}). \quad (19)$$

We note that approximating (17) by (19) can also be motivated (as in [7]) by the fact that $E(\mathbf{z}) = \mathbf{z}_e$, however the order of accuracy of the approximation is not addressed.

We can now present our regularization parameter selection methods. Each method exploits the connection between T_0 and T_0^{wls} evident in (18). The least squares form of (19) provides a means of extending the standard methods of the discrepancy principle, generalized cross validation, and unbiased predictive risk [18] to (8). We begin with the discrepancy principle.

2.2. The Discrepancy Principle

From (18) and (19) we have

$$E(T_0(\mathbf{u}; \mathbf{z})) \approx T_0(\mathbf{u}_e; \mathbf{z}_e) + E(T_0^{\text{wls}}(\mathbf{u}; \mathbf{z})),$$

where E is the expected value function. Thus it is reasonable to say that acceptable values of the regularization parameter α in (8) will be those for which

$$T_0^{\text{wls}}(\mathbf{u}_\alpha; \mathbf{z}) \approx E(T_0^{\text{wls}}(\mathbf{u}_e; \mathbf{z})).$$

In order to obtain an estimate of $E(T_0^{\text{wls}}(\mathbf{u}_e; \mathbf{z}))$ we first approximate (2) as is often done in practice (see, e.g., [9, 15]):

$$\mathbf{z} - \boldsymbol{\gamma} = \mathbf{A}\mathbf{u}_e + \mathbf{e}, \quad (20)$$

where $\mathbf{e} \sim N(\mathbf{0}, \mathbf{Z})$, with \mathbf{Z} defined in (19). This in turn implies

$$\mathbf{r}(\mathbf{u}_e) \sim N(\mathbf{0}, \mathbf{I}_n), \quad (21)$$

where

$$\mathbf{r}(\mathbf{u}) \stackrel{\text{def}}{=} \mathbf{Z}^{-1/2}(\mathbf{A}\mathbf{u} - (\mathbf{z} - \boldsymbol{\gamma})). \quad (22)$$

A standard statistical result then tells us that given (21),

$$\|\mathbf{r}(\mathbf{u}_e)\|_2^2 \sim \chi^2(n), \quad (23)$$

where “ \sim ” means “is distributed as”, and $\chi^2(n)$ denotes the chi-squared distribution with n degrees of freedom, which has mean and variance n and $2n$, respectively. And hence we have $E(2T_0^{\text{wls}}(\mathbf{u}_e; \mathbf{z})) \approx n$.

Our acceptability criterion for a given regularization parameter α is then that

$$2T_0^{\text{wls}}(\mathbf{u}_\alpha; \mathbf{z}) \approx n.$$

Or more specifically, we can say that the appropriate choices of α are those values for which $\|\mathbf{r}(\mathbf{u}_\alpha)\|_2^2$ lies within two standard deviations of n ; that is, if

$$n - 2\sqrt{2n} \leq 2T_0^{\text{wls}}(\mathbf{u}_\alpha; \mathbf{z}) \leq n + 2\sqrt{2n}. \quad (24)$$

If a specific value of α is desired, which is typical, it is natural to choose the value that maximizes the likelihood that $\|\mathbf{r}(\mathbf{u}_\alpha)\|^2$ has arisen from $\chi^2(n)$, i.e. for which $\|\mathbf{r}(\mathbf{u}_\alpha)\|^2 = E(\chi^2(n)) = n$. We note that this is equivalent to Morozov's discrepancy principle [18] applied to the problem of finding the optimal regularization parameter for the the minimization problem

$$\arg \min_{\mathbf{u} \geq \mathbf{0}} \left\{ T_0^{\text{wls}}(\mathbf{u}; \mathbf{z}) + \frac{\alpha}{2} \langle \mathbf{u}, \mathbf{C}\mathbf{u} \rangle \right\}. \quad (25)$$

The difference here is that we are computing \mathbf{u}_α from (8).

In order to automate this approach, as well as to handle the case in which there is no α for which $\|\mathbf{r}(\mathbf{u}_\alpha)\|^2 = n$, we define the desired regularization parameter to be the one that solves

$$\alpha_{\text{DP}} = \arg \min_{\alpha > 0} \left(2T_0^{\text{wls}}(\mathbf{u}_\alpha; \mathbf{z}) - n \right)^2, \quad (26)$$

where \mathbf{u}_α is computed from (8).

2.3. Generalized Cross Validation

The method of leave-one-out cross validation for regularization parameter selection applied to (8) can be described as follows. Define

$$\mathbf{u}_\alpha^k \stackrel{\text{def}}{=} \arg \min_{\mathbf{u} \geq \mathbf{0}} \left\{ \sum_{i \neq k} ([\mathbf{A}\mathbf{u}]_i + \gamma) - z_i \ln([\mathbf{A}\mathbf{u}]_i + \gamma) + \frac{\alpha}{2} \langle \mathbf{u}, \mathbf{C}\mathbf{u} \rangle \right\},$$

and then choose α to be the minimizer of

$$\text{CV}(\alpha) = \frac{1}{n} \sum_{k=1}^n \left\{ ([\mathbf{A}\mathbf{u}_\alpha^k]_k + \gamma) - z_k \ln([\mathbf{A}\mathbf{u}_\alpha^k]_k + \gamma) \right\}. \quad (27)$$

For large-scale problems, minimizing (27) is intractable. The method of generalized cross validation—first introduced by Wahba [19] for regularized least squares problems—approximates $\text{CV}(\alpha)$ by a function that can be much more efficiently minimized.

To extend this approach to the case where the fit-to-data function is T_0 , the Taylor series approximation (18) must be called upon once again:

$$\begin{aligned} \text{CV}(\alpha) &\approx \frac{1}{n} T_0(\mathbf{u}_e; \mathbf{z}) + \frac{1}{2n} \sum_{k=1}^n [\mathbf{r}(\mathbf{u}_\alpha^k)]_k^2 \\ &= \frac{1}{n} T_0(\mathbf{u}_e; \mathbf{z}) + \frac{1}{2n} \sum_{k=1}^n \left(\frac{[\mathbf{r}(\mathbf{u}_\alpha)]_k}{1 - [\mathbf{Z}^{-1/2} \mathbf{A} \mathbf{A}_\alpha]_{kk}} \right)^2. \end{aligned} \quad (28)$$

Here $\mathbf{r}(\mathbf{u})$ is defined as in (22), and (28) follows from arguments found in [1] assuming \mathbf{A}_α is a matrix satisfying $\mathbf{u}_\alpha = \mathbf{A}_\alpha \mathbf{Z}^{-1/2} (\mathbf{z} - \gamma)$.

However, for (8), the data-to-regularized-solution (or regularization) operator is nonlinear, and so \mathbf{A}_α must be a linear approximation satisfying $\mathbf{u}_\alpha \approx \mathbf{A}_\alpha \mathbf{Z}^{-1/2} (\mathbf{z} - \gamma)$. To derive such an approximation, we define \mathbf{D}_α to be a diagonal matrix with diagonal

entries $[\mathbf{D}_\alpha]_{ii} = 1$ if $[\mathbf{u}_\alpha]_i > 0$ and $[\mathbf{D}_\alpha]_{ii} = 0$ otherwise, then since T_α is a strictly convex function, \mathbf{u}_α will be the minimum norm solution of (see [13])

$$\mathbf{D}_\alpha \nabla T_\alpha(\mathbf{D}_\alpha \mathbf{u}; \mathbf{z}) = \mathbf{0}. \quad (29)$$

Now, using the approximation (25) for T_α (motivated by (18)) we can approximate (29) by

$$\mathbf{D}_\alpha \mathbf{A}^T \mathbf{Z}^{-1} (\mathbf{A} \mathbf{D}_\alpha \mathbf{u} - (\mathbf{z} - \boldsymbol{\gamma})) + \alpha \mathbf{D}_\alpha \mathbf{C} \mathbf{D}_\alpha \mathbf{u} = \mathbf{0},$$

which has minimum norm solution

$$(\mathbf{D}_\alpha (\mathbf{A}^T \mathbf{Z}^{-1} \mathbf{A} + \alpha \mathbf{C}) \mathbf{D}_\alpha)^\dagger \mathbf{D}_\alpha \mathbf{A}^T \mathbf{Z}^{-1} (\mathbf{z} - \boldsymbol{\gamma}).$$

Thus we define

$$\mathbf{A}_\alpha = (\mathbf{D}_\alpha (\mathbf{A}^T \mathbf{Z}^{-1} \mathbf{A} + \alpha \mathbf{C}) \mathbf{D}_\alpha)^\dagger \mathbf{D}_\alpha \mathbf{A}^T \mathbf{Z}^{-1/2}. \quad (30)$$

Recall that we wanted $\mathbf{u}_\alpha \approx \mathbf{A}_\alpha \mathbf{Z}^{-1/2} (\mathbf{z} - \boldsymbol{\gamma})$.

Generalized cross validation (GCV) requires still another approximation, as even with (30), minimizing (28) remains impractical for large-scale problems. The approximation is given by

$$1 - [\mathbf{A} \mathbf{A}_\alpha]_{kk} \approx \text{trace}(\mathbf{I}_n - \mathbf{Z}^{-1/2} \mathbf{A} \mathbf{A}_\alpha) / n.$$

Then, finally, we have the following (approximate) GCV function for (8):

$$\text{GCV}(\alpha) = n T_0^{\text{wls}}(\mathbf{u}_\alpha; \mathbf{z}) / \text{trace}(\mathbf{I}_n - \mathbf{Z}^{-1/2} \mathbf{A} \mathbf{A}_\alpha)^2, \quad (31)$$

where \mathbf{A}_α is given by (30). We note that this is very similar to the GCV function for the weighted least squares problem (25). The difference is that \mathbf{u}_α is computed from (8), which is also used to define \mathbf{D}_α in (30).

Due to the size of the matrix \mathbf{A}_α and the presence of the matrices \mathbf{D}_α and \mathbf{Z} in (30), the Trace Lemma must be used to approximate the value of $\text{trace}(\mathbf{I}_n - \mathbf{Z}^{-1/2} \mathbf{A} \mathbf{A}_\alpha)$. The Trace Lemma has the form [18]: given $\mathbf{B} \in \mathbb{R}^{n \times n}$,

$$\mathbf{v} \sim N(\mathbf{0}, \mathbf{I}_n) \quad \text{implies} \quad E(\mathbf{v}^T \mathbf{B} \mathbf{v}) = \text{trace}(\mathbf{B}). \quad (32)$$

Thus given a realization \mathbf{v} from $N(\mathbf{0}, \mathbf{I}_n)$,

$$\text{trace}(\mathbf{I}_n - \mathbf{Z}^{-1/2} \mathbf{A} \mathbf{A}_\alpha) \approx \mathbf{v}^T \mathbf{v} - \mathbf{v}^T \mathbf{Z}^{-1/2} \mathbf{A} \mathbf{A}_\alpha \mathbf{v}.$$

Finally, the GCV method for choosing α in (8) is defined as follows: choose the α that minimizes

$$\text{GCV}(\alpha) \approx n T_0^{\text{wls}}(\mathbf{u}_\alpha; \mathbf{z}) / (\mathbf{v}^T \mathbf{v} - \mathbf{v}^T \mathbf{Z}^{-1/2} \mathbf{A} \mathbf{A}_\alpha \mathbf{v})^2, \quad (33)$$

where \mathbf{v} is a realization from $N(\mathbf{0}, \mathbf{I}_n)$.

In practice, due to the large-scale nature of the problems of interest to us, $\mathbf{A}_\alpha \mathbf{v}$ in (33) must be approximated by applying a truncated conjugate gradient iteration with $\mathbf{x}_0 = \mathbf{0}$ to

$$\mathbf{D}_\alpha (\mathbf{A}^T \mathbf{Z}^{-1} \mathbf{A} + \alpha \mathbf{C}) \mathbf{D}_\alpha \mathbf{x} = \mathbf{D}_\alpha \mathbf{A}^T \mathbf{Z}^{-1/2} \mathbf{v}.$$

Before continuing, we acknowledge that the number of approximations used above is bordering on the ridiculous. However, we will see later that the approach is effective nonetheless.

2.4. The Unbiased Predictive Risk Estimate

The motivation behind the unbiased predictive risk estimate (UPRE) [18] of α is straightforward: we seek the value of α that minimizes the *predictive risk* $E(T_0(\mathbf{u}_\alpha; \mathbf{z}_e))$. However since \mathbf{z}_e is unknown, the execution of the method requires some work.

Once again, we use the Taylor series approximation (18). From it, we have

$$T_0(\mathbf{u}_\alpha; \mathbf{z}_e) \approx T_0(\mathbf{u}_e; \mathbf{z}_e) + T_0^{\text{wls}}(\mathbf{u}_\alpha, \mathbf{z}_e),$$

which we will call the predictive error, whereas

$$T_0(\mathbf{u}_\alpha; \mathbf{z}) \approx T_0(\mathbf{u}_e; \mathbf{z}) + T_0^{\text{wls}}(\mathbf{u}_\alpha, \mathbf{z}).$$

Taking the expected value of these two equations yields, respectively,

$$\begin{aligned} E(T_0(\mathbf{u}_\alpha; \mathbf{z}_e)) &\approx T_0(\mathbf{u}_e; \mathbf{z}_e) + E(T_0^{\text{wls}}(\mathbf{u}_\alpha, \mathbf{z}_e)), \\ E(T_0(\mathbf{u}_\alpha; \mathbf{z})) &\approx T_0(\mathbf{u}_e; \mathbf{z}) + E(T_0^{\text{wls}}(\mathbf{u}_\alpha, \mathbf{z})). \end{aligned}$$

Following arguments in [18, Section 7.1] and using the approximate regularization operator \mathbf{A}_α defined by (30), we find that

$$\begin{aligned} E(T_0^{\text{wls}}(\mathbf{u}_\alpha, \mathbf{z}_e)) &= \frac{1}{2} \left\| \mathbf{Z}^{-1/2}(\mathbf{A}\mathbf{A}_\alpha - \mathbf{I}_n)\mathbf{A}\mathbf{u}_e \right\|^2 + \frac{1}{2} \text{trace}((\mathbf{Z}^{-1/2}\mathbf{A}\mathbf{A}_\alpha)^2), \\ E(T_0^{\text{wls}}(\mathbf{u}_\alpha, \mathbf{z})) &= \frac{1}{2} \left\| \mathbf{Z}^{-1/2}((\mathbf{A}\mathbf{A}_\alpha - \mathbf{I}_n)\mathbf{A}\mathbf{u}_e) \right\|^2 + \frac{1}{2} \text{trace}((\mathbf{Z}^{-1/2}\mathbf{A}\mathbf{A}_\alpha)^2) \\ &\quad - \text{trace}(\mathbf{Z}^{-1/2}\mathbf{A}\mathbf{A}_\alpha) + \frac{n}{2}. \end{aligned}$$

Thus we have the following estimation of the predictive risk:

$$\begin{aligned} E(T_0(\mathbf{u}_\alpha, \mathbf{z}_e)) &\approx T_0(\mathbf{u}_e; \mathbf{z}_e) + E(T_0^{\text{wls}}(\mathbf{u}_\alpha, \mathbf{z}_e)) \\ &= E(T_0^{\text{wls}}(\mathbf{u}_\alpha, \mathbf{z})) + \text{trace}(\mathbf{Z}^{-1/2}\mathbf{A}\mathbf{A}_\alpha) - \frac{n}{2}. \end{aligned}$$

This motivates our choice of the UPRE function

$$\text{UPRE}(\alpha) = T_0^{\text{wls}}(\mathbf{u}_\alpha; \mathbf{z}) + \text{trace}(\mathbf{Z}^{-1/2}\mathbf{A}\mathbf{A}_\alpha) - \frac{n}{2}. \quad (34)$$

We then choose the regularization parameter α in (8) that minimizes $\text{UPRE}(\alpha)$. As in the case of GCV, $\text{trace}(\mathbf{Z}^{-1/2}\mathbf{A}\mathbf{A}_\alpha)$ must be approximated using the Trace Lemma and a truncated conjugate gradient iteration.

3. Numerical Tests

In this section we present numerical results using a standard synthetic data example developed at the US Air Force Phillips Laboratory, Lasers and Imaging Directorate, Kirtland Air Force Base, New Mexico. The image is a computer simulation of a field experiment showing a satellite as taken from a ground based telescope. The true image has 256×256 pixels and is shown on the left in Fig. 1. In addition, we imported the simulated star field data seen on the right in Fig. 1.

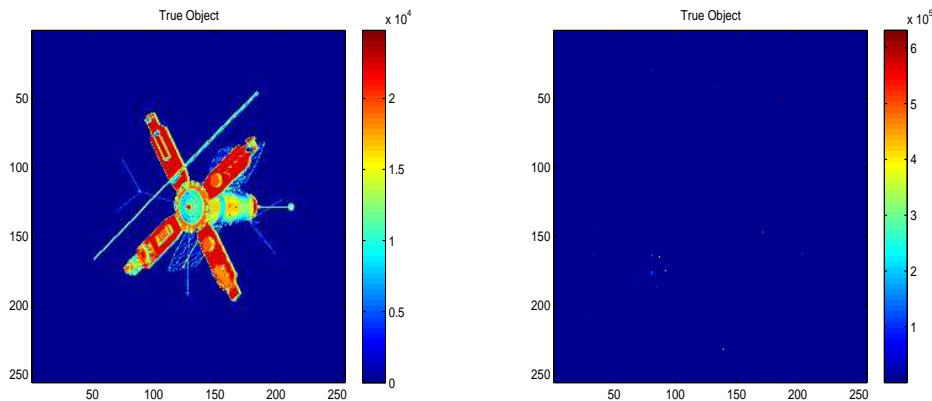


Figure 1. True image of the satellite on the left and the true image of the star field on the right.

Generating corresponding blurred, noisy data requires a discrete PSF \mathbf{a} , which we compute using the Fourier optics PSF model [21, 18]

$$\mathbf{a} = \left| \mathbf{F}^{-1} \left\{ \mathbf{p} \odot e^{i\phi} \right\} \right|^2,$$

where \mathbf{p} is the $\sqrt{n} \times \sqrt{n}$ indicator array for the telescopes pupil; ϕ is the $\sqrt{n} \times \sqrt{n}$ array that represents the aberrations in the incoming wavefronts of light; $i = \sqrt{-1}$; and \mathbf{F} is the two-dimensional discrete Fourier transform matrix. If periodic boundary conditions are assumed in the image, the $n \times n$ blurring matrix \mathbf{A} takes the form

$$\mathbf{A}\mathbf{u} = \text{vec}(\text{ifft2}(\hat{\mathbf{a}} \odot (\text{fft2}(\mathbf{u}))))), \quad \hat{\mathbf{a}} = \text{fft2}(\text{fftshift}(\mathbf{a})),$$

where vec takes $\sqrt{n} \times \sqrt{n}$ arrays to $n \times 1$ vectors by stacking columns; fft2 and ifft2 are the two-dimensional discrete Fourier transform and inverse discrete Fourier transform, respectively; and fftshift swaps the first and third and the second and fourth quadrants of the array \mathbf{a} . A variety of other boundary conditions can be assumed in the image, resulting in slightly different formulations for \mathbf{A} [12].

To generate our data, we use the statistical model of [16], which has the form

$$\mathbf{z} = \text{Pois}(\mathbf{A}\mathbf{u}_e) + \text{Pois}(\boldsymbol{\gamma}) + N(\mathbf{0}, \sigma^2 \mathbf{I}_n), \quad (35)$$

Using the approximation $N(\sigma^2, \sigma^2) \approx \text{Pois}(\sigma^2)$, and assuming independence, (35) can be approximated by

$$\mathbf{z} + \boldsymbol{\sigma}^2 = \text{Pois}(\mathbf{A}\mathbf{u}_e + \boldsymbol{\gamma} + \boldsymbol{\sigma}^2),$$

where $\boldsymbol{\sigma}^2 = \sigma^2 \mathbf{1}$, which has the form of (2). We generate Poisson noise using MATLAB's `poissrnd` function and choose the values $\gamma = 10$ and $\sigma = 5$, which are realistic values for these parameters. The blurred, noisy data with SNR approximately 30 is plotted on the right in Figure 2.

In order to test the selection methods on multiple data sets, we vary the intensity of the true image \mathbf{u}_e in order to obtain noise at four different levels of signal-to-noise

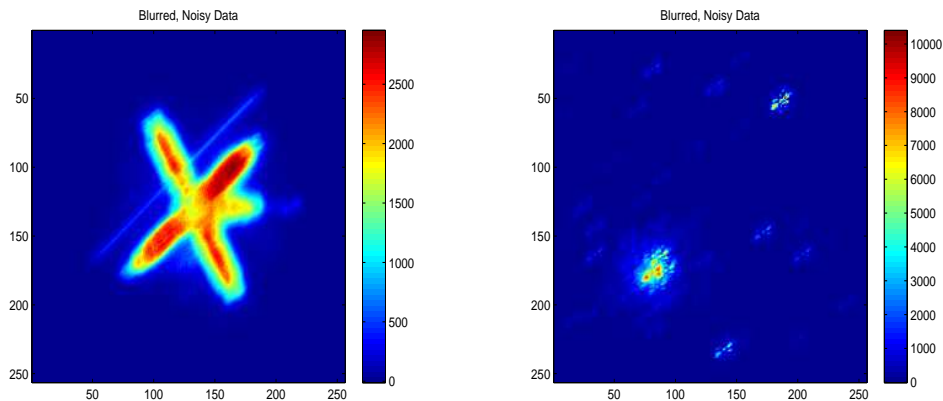


Figure 2. *Blurred, noisy images of the satellite on the left, and the star field on the right, both with $SNR = 30$.*

ratio (SNR), which for (35) is defined as

$$SNR = \sqrt{\frac{\|\mathbf{A}\mathbf{u}_e\|^2}{\sum_{j=1}^n ([\mathbf{A}\mathbf{u}_e]_j + \gamma + \gamma^2 + \sigma^2)}}.$$

Three choices of regularization matrix \mathbf{C} are examined. The most standard is \mathbf{I}_n , yielding standard Tikhonov regularization, which penalizes reconstructions with large ℓ^2 -norm. We tested this regularization matrix on both the satellite and star field examples.

Another standard choice for \mathbf{C} is the negative Laplacian, referred to here as \mathbf{L} , which penalizes non-smooth reconstructions. Since the star field is non-smooth, we tested this regularization only on the satellite example.

The third choice for \mathbf{C} is described in [8]. We refer to it as Θ . Like \mathbf{L} , Θ is a discretization of a diffusion operator. However, unlike \mathbf{L} , it encourages the formation of edges in reconstructions. It is defined as follows:

$$\Theta = \mathbf{D}_x^T \mathbf{\Lambda} \mathbf{D}_x + \mathbf{D}_y^T \mathbf{\Lambda} \mathbf{D}_y, \quad (36)$$

where \mathbf{D}_x and \mathbf{D}_y are discretizations of the x and y partial derivative operators, respectively, and $\mathbf{\Lambda}$ is a diagonal matrix with entries near 1 corresponding to pixels away from an edge and significantly less than 1 for pixels near an edge.

The diagonal matrix $\mathbf{\Lambda}$ in (36) is defined as follows. First, let

$$\mathbf{v} = [\mathbf{D}_x \mathbf{u}_{\text{approx}}]^2 + [\mathbf{D}_y \mathbf{u}_{\text{approx}}]^2, \quad (37)$$

where the square is computed component-wise; $\mathbf{u}_{\text{approx}}$ is computed from (8) with $\mathbf{C} = \mathbf{L}$ and α chosen using GCV; and MATLAB's `gradient` function is used for computing $\mathbf{D}_x \mathbf{u}_{\text{approx}}$ and $\mathbf{D}_y \mathbf{u}_{\text{approx}}$. We then define

$$[\mathbf{v}_\epsilon]_i = \begin{cases} v_i & v_i > \epsilon \|\mathbf{v}\|_\infty \\ 0 & \text{otherwise,} \end{cases}, \quad (38)$$

where $0 < \epsilon < 1$ (we chose $\epsilon = 0.01$ for our experiments), and $\|\cdot\|_\infty$ denotes the ℓ^∞ -norm on \mathbb{R}^N . $\mathbf{\Lambda}$ is then given by

$$\mathbf{\Lambda} = \text{diag} \left(\max \left\{ \frac{\mathbf{1}}{\mathbf{1} + \mathbf{v}_\epsilon}, \frac{1}{10} \right\} \right). \quad (39)$$

Note that when $[\mathbf{v}_\epsilon]_i$ is large, i.e. at or near an edge, $\mathbf{\Lambda}$ has the effect of decreasing the regularization parameter by an order of magnitude, whereas when $[\mathbf{v}_\epsilon]_i \approx 0$ the regularization parameter remains approximately the same.

In each of DP, GCV, and UPRE, a minimization problem must be solved: for DP its (26); for GCV, $\text{GCV}(\alpha)$ defined by (33) must be minimized subject to $\alpha > 0$; whereas for UPRE, $\text{UPRE}(\alpha)$ defined by (34) must be minimized subject to $\alpha > 0$. For this we use MATLAB's `fminbnd` function. The input for this function includes lower and upper bounds for the minimizer. For the satellite example, the lower bound was 0 and the upper bound was 0.01. For the star field example, the lower bound was 0 and the upper bound was 10^{-7} . A drawback of `fminbnd` is that it can converge to a local minimizer; this is the reason that two different upper bounds are used.

For $\mathbf{C} = \mathbf{I}_n$, we perform experiments on both the satellite and star field data with $\text{SNR} = 1, 5, 10$, and 30. For $\mathbf{C} = \mathbf{L}$ or $\mathbf{C} = \mathbf{\Theta}$, we perform experiments on the satellite data with $\text{SNR} = 10$ and 30. In each case, we calculate the relative error,

$$\frac{\|\mathbf{u}_e - \mathbf{u}_\alpha\|}{\|\mathbf{u}_e\|},$$

over a range of values of α and the points corresponding to the recommendations by the various selection methods.

The results of the satellite test case with $\mathbf{C} = \mathbf{I}_n$ are displayed in Fig. 3. The three methods appear to be quite effective, yielding good recommendations for the cases of $\text{SNR} = 1, 5$ and 10, with DP giving a slightly better value of α than GCV and UPRE. The recommendations in the case of $\text{SNR} = 30$ are somewhat farther away from the minimizer of the relative error, however a quality reconstruction is still obtained.

The results of the star field test case are displayed in Fig. 4. Again, the three methods yield good recommendations. We note, though, that the methods yielded better recommendations in the SNR cases of 1 and 5 than in the cases of 10 and 30.

The results of the satellite test case with $\mathbf{C} = \mathbf{L}$ are displayed in the upper plots in Fig. 5. The three methods yielded good recommendations. Again, a quality reconstruction was obtained in the case of $\text{SNR} = 30$, though the recommendation was somewhat far from the minimizer of the relative error.

The results of the satellite test case with $\mathbf{C} = \mathbf{\Theta}$ are given in the lower plots in Fig. 5. In order to generate $\mathbf{\Theta}$, \mathbf{u}_{init} was obtained by setting $\mathbf{C} = \mathbf{L}$ and using the GCV recommendation for the value of α . In the case of $\text{SNR} = 10$, the GCV and UPRE methods yielded a slightly better value of α than DP.

Reconstructions of the satellite and the star field are given in Fig. 6. Both reconstructions are from data with $\text{SNR} = 30$. For the satellite, the regularization matrix was chosen to be $\mathbf{\Theta}$.

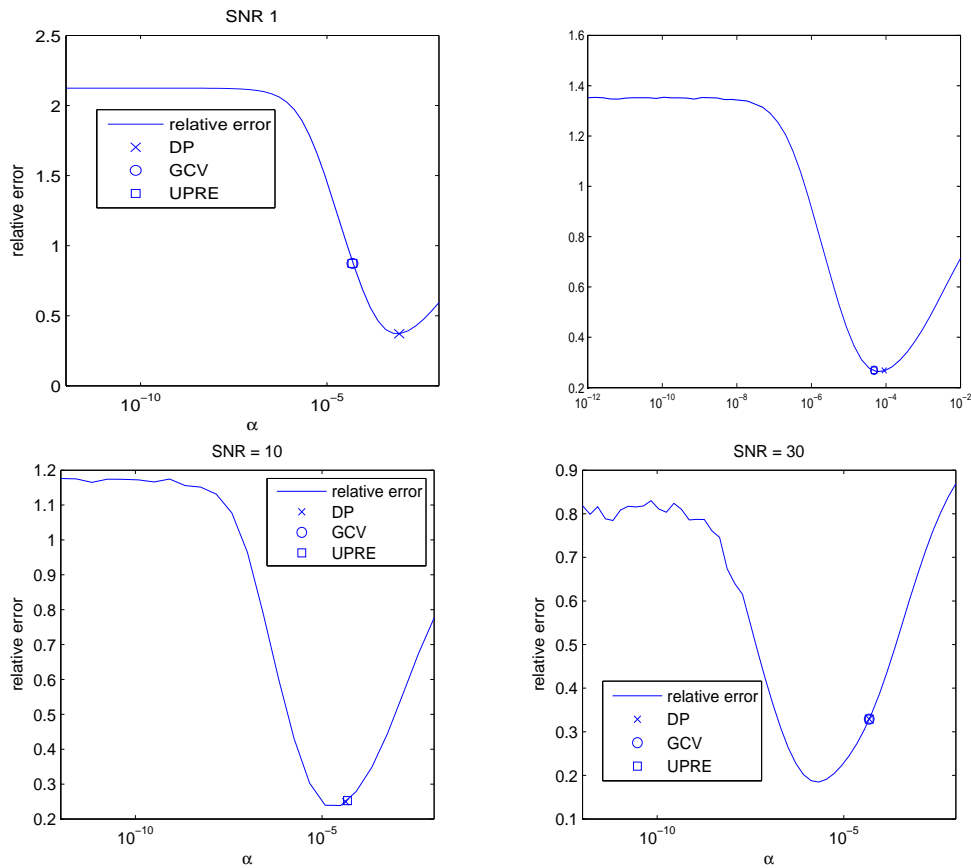


Figure 3. *Satellite Test Case, with $\mathbf{C} = \mathbf{I}_n$. Plot of relative error with stopping rules GCV, UPRE, DP.*

These results indicate that the three methods give useful recommendations for the value of the regularization parameter. The three methods yielded values for α that were very similar for each of the four data sets. In particular, the recommendations from GCV and UPRE were identical. This suggests that the methods are similar, and this is supported by the analysis found in [18]. GCV is slightly more flexible than UPRE, in that for regular least squares problems it does not require knowledge of the noise variance, while UPRE does require such knowledge.

An aspect of the UPRE and GCV methods that should be pointed out is that the recommended values for α can change with different realizations of \mathbf{v} from $N(0, \mathbf{I}_n)$ in (33) and (34). A better approach might be to take several realizations from $N(0, \mathbf{I}_n)$, and then average the resulting trace estimates. The downside of this is the increased computational cost.

4. Conclusion

Data collected by a CCD camera array follow Poisson statistics. Obtaining the maximum likelihood estimate therefore requires solving the problem given in (3). Due to the fact that solutions of (3) are typically noise corrupted, regularization is required.

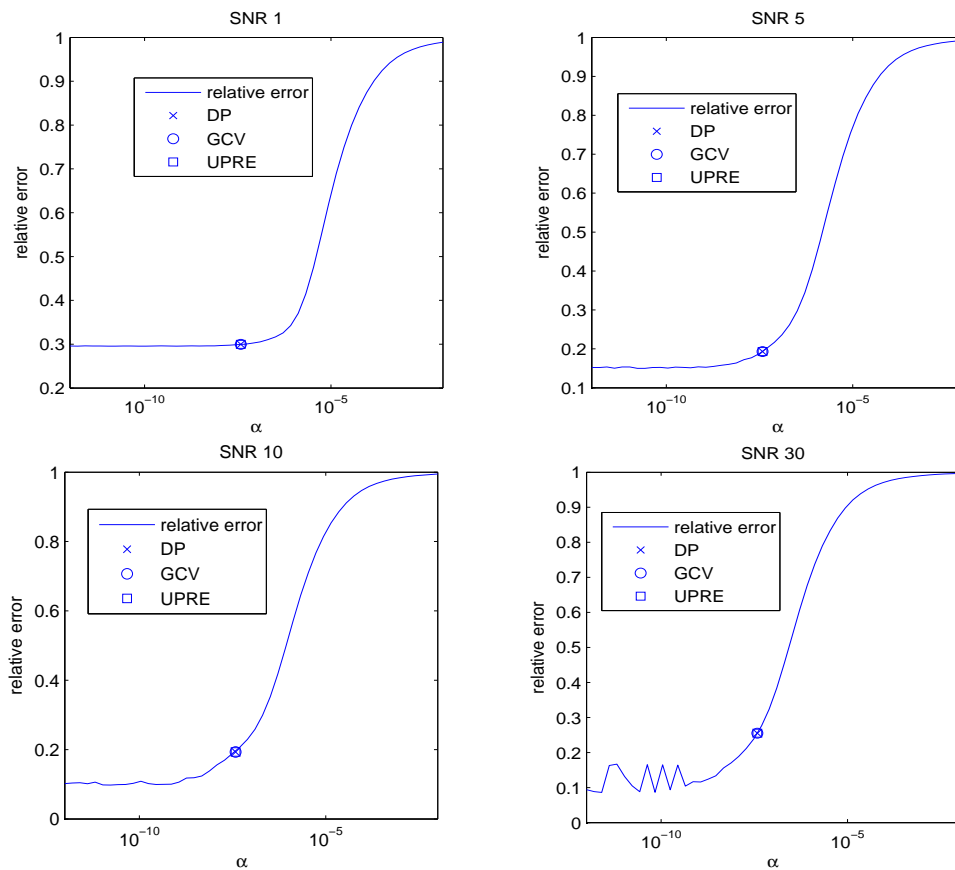


Figure 4. *Star Field Test Case. Plot of relative error with stopping rules GCV, UPRE, DP.*

General Tikhonov regularization is applied, and the resulting estimate is the solution of (8).

Existing methods for selecting the value of the regularization parameter depend on (8) having a regular least squares fit-to-data function and hence cannot be directly applied. However, a Taylor series argument shows that $T_0(\mathbf{u}; \mathbf{z})$ can be well approximated by a weighted least squares function, $T_0^{\text{wls}}(\mathbf{u}; \mathbf{z})$ (given in (19)). Using this approximation, we have applied the discrepancy principle (DP), generalized cross validation (GCV) and the unbiased predictive risk estimator (UPRE) methods for selecting the regularization parameter in (8).

We performed tests on synthetically generated images of a satellite and a star field. The data was generated with the statistical model given in (35). Multiple data sets were generated by varying the intensity of the true images to yield four different signal to noise ratios: 1, 5, 10 and 30. Tests of the methods on these data sets using various regularization operators indicated that all three methods gave good recommendations for the value of the regularization parameter.

We conclude that DP is a more effective stopping rule than GCV and UPRE, however the recommendations were very close. The GCV and UPRE methods were

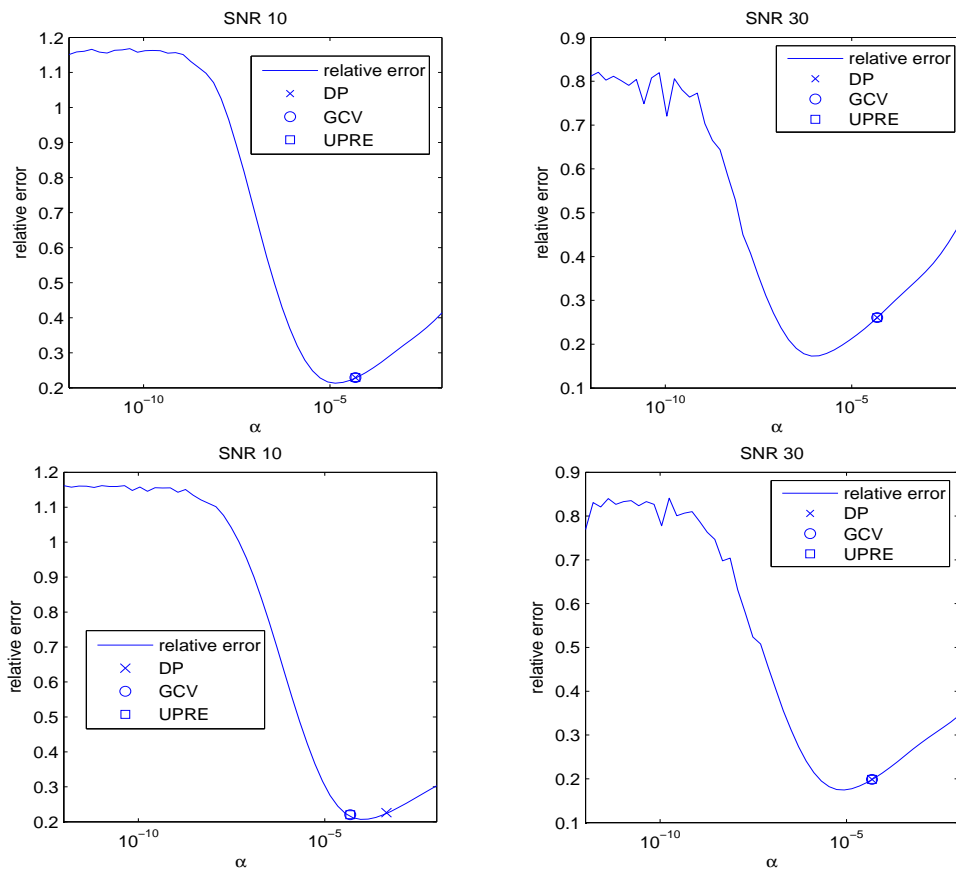


Figure 5. In the top row, Satellite Test Case, with $\mathbf{C} = L_n$. In the bottom row, Satellite Test Case, with $\mathbf{C} = \Theta$. Plot of relative error with stopping rules GCV, UPRE, DP.

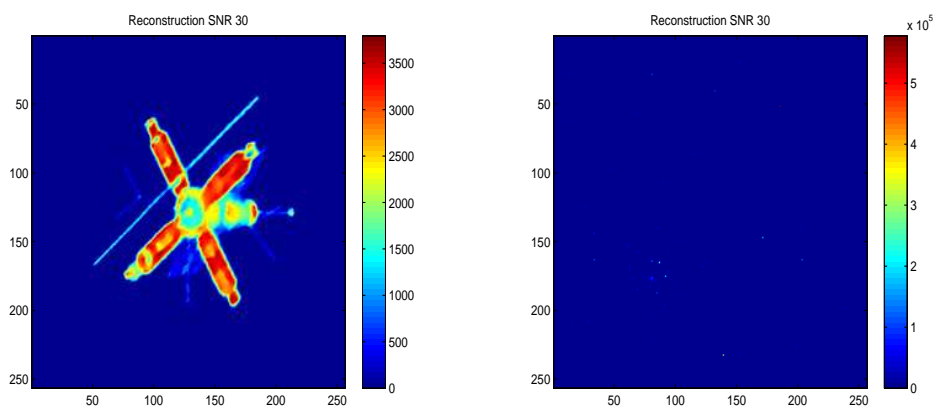


Figure 6. Reconstructions of the satellite, on the left, and of the star field, on the right. Both reconstructions are from data with SNR=30. For the satellite, $\mathbf{C} = \Theta$.

very similar for the examples under consideration.

References

- [1] Richard C. Aster, Brian Borchers, and Clifford H. Thurber, *Parameter Estimation and Inverse Problems*, Elsevier 2005.
- [2] Johnathan M. Bardsley and N'djekornom Laobeul, *Tikhonov Regularized Poisson Likelihood Estimation: Theoretical Justification and a Computational Method*, Inverse Problems in Science and Engineering, Volume 16, Issue 2 January 2008 , pages 199 - 215.
- [3] Johnathan M. Bardsley and N'djekornom Laobeul, *An Analysis of Regularization by Diffusion for Ill-Posed Poisson Likelihood Estimation*, to appear in Inverse Problems in Science and Engineering. University of Montana Technical Report #23, 2008.
- [4] Johnathan M. Bardsley and Aaron Luttman, *Total Variation-Penalized Poisson Likelihood Estimation for Ill-Posed Problems*, Advances in Computational Mathematics, DOI 10.1007/s10444-008-9081-8, to appear in print.
- [5] Johnathan M. Bardsley and James G. Nagy, *Covariance-Preconditioned Iterative Methods for Nonnegatively Constrained Image Reconstruction*, SIAM Journal on Matrix Analysis and Applications, 27 (4), 2006, pp. 1184-1198.
- [6] J. M. Bardsley and C. R. Vogel, *A Nonnegatively Constrained Convex Programming Method for Image Reconstruction*, SIAM Journal on Scientific Computing, 25(4), 2004, pp. 1326-1343.
- [7] Johnathan M. Bardsley, *Stopping Rules for a Nonnegatively Constrained Iterative Method for Ill-Posed Poisson Imaging Problems*, BIT Numerical Mathematics, Volume 48, Number 4, December, 2008, pp. 651-664.
- [8] Johnathan M. Bardsley and John Goldes, *An Iterative Method for Edge-Preserving MAP Estimation when Data-Noise is Poisson*, submitted to SIAM Journal on Scientific Computing. University of Montana Technical Report #26, 2008.
- [9] A. Grinvald and I. Steinberg, *On the Analysis of Fluorescence Decay Kinetics by the Method of Least-Squares*, Analytical Biochemistry, 59, 1974, pp.583-594.
- [10] Jari Kaipio and Erkki Somersalo, *Statistical and Computational Inverse Problems*, Springer 2005.
- [11] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems*, SIAM, Philadelphia, 1997.
- [12] P. C. Hansen, J. G. Nagy, and D. P. O'Leary, *Deblurring Images: Matrices, Spectra, and Filtering*, SIAM, Philadelphia, 2006.
- [13] Jorge Nocedal and Stephen J. Wright, *Numerical Optimization*, Springer-Verlag, New York, 1999.
- [14] John M. Ollinger and Jeffrey A. Fessler, *Positron-Emission Tomography*, IEEE Signal Processing Magazine, January 1997.
- [15] Bert Rust, *Truncating the Singular Value Decomposition for Ill-Posed Problems*, NISTIR 6131, U.S. Dept. of Commerce, July 1998.
- [16] D. L. Snyder, A. M. Hammoud, and R. L. White, *Image recovery from data acquired with a charge-coupled-device camera*, Journal of the Optical Society of America A, **10** (1993), pp. 1014-1023.
- [17] R. Vio, J. Bardsley, and W. Wamsteker, *Least-Squares methods with Poissonian noise: analysis and a comparison with the Richardson-Lucy algorithm*, Astronomy and Astrophysics, 436, 2005, pp. 741-755.
- [18] C. R. Vogel, *Computational Methods for Inverse Problems*, SIAM, Philadelphia, 2002.
- [19] G. Wahba, *Practical approximate solutions to linear operator equations when the data are noisy*, SIAM Journal on Numerical Analysis, vol. 14, 1977, pp. 651-667.
- [20] Daniel F. Yu and Jeffrey A. Fessler, *Edge-Preserving Tomographic Reconstruction with Nonlocal Regularization*, IEEE Trans. on Medical Imaging, 21(2), 2002, pp. 159-173.
- [21] J. W. Goodman, *Introduction to Fourier Optics, 2nd Edition*, McGraw-Hill, 1996.