

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 [23].

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) [27]. 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 [26]. 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 a more accurate 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 [3, 4, 5].

But one of the most important problems remains: the development of methods for choosing the regularization parameter. While the problem of regularization parameter selection has been discussed in very general settings (even encompassing the Poisson noise case), here we present a concrete (theoretical) tie between the negative-log Poisson likelihood function and a certain weighted least squares function. Using this link, we are able to motivate the use of the several standard regularization parameter selection techniques for least squares problems for use on regularized negative-log Poisson reconstruction problems. Our approach is different from other schemes for estimating the value of the regularization parameter in the case of non-quadratic likelihood (see e.g., [9, 13]).

The existing regularization parameter selection methods of interest to us in this paper are the discrepancy principle (DP), generalized cross validation (GCV), and unbiased predictive risk estimate (UPRE). DP uses an approximation which implies that the optimal value of the parameter should yield a solution for which the sum of squares of the weighted residuals is equal to the mean of a Chi-square distribution [17, 21, 27]. GCV selects the value of the parameter that minimizes GCV functional, which is an adaptation of the leave-one-out cross-validation functional to large-scale problems [27, 28]. UPRE chooses the value of the regularization parameter that minimizes the

predictive risk [27].

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

$$\mathbf{z} = \mathbf{A}\mathbf{u} + \boldsymbol{\gamma}, \quad (1)$$

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 [19, 23] is

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

where $\text{Pois}(\boldsymbol{\lambda})$ is an independent and identically distributed (iid) Poisson random vector with Poisson parameter vector $\boldsymbol{\lambda}$. (See [23] 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 $-\ln p(\mathbf{z}; \mathbf{u})$.

Similar to 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 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, $\mathbf{C} \in \mathbb{R}^{N \times N}$ is symmetric, positive semi-definite, and the regularization parameter α controls the magnitude of the eigenvalues of the covariance. Note in particular that larger α yields larger eigenvalues of $\alpha\mathbf{C}$, which yields a more restrictive (or informative) prior, since $p(\mathbf{u})$ 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 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, the presence of the regularization α is more familiar. 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, we test our methods on two synthetic data examples at a number of different noise levels and find that effective recommendations can be obtained. We end the paper with conclusions.

2. Regularization Parameter Choice Methods

In this section, we motivate the use of three existing regularization parameter selection methods on problems of the type (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, and $\text{diag}(\mathbf{v})$ indicates the diagonal matrix with the non-zero entries given by the components of \mathbf{v} . 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 $\mathbf{z}_e = \mathbf{A}\mathbf{u}_e + \gamma$ 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)

$$\begin{aligned} T_0(\mathbf{u}; \mathbf{z}) &= T_0(\mathbf{u}_e + \mathbf{h}; \mathbf{z}_e + \mathbf{k}), \\ &= T_0(\mathbf{u}_e; \mathbf{z}_e) + \mathbf{k}^T \nabla_{\mathbf{z}} T_0(\mathbf{u}_e; \mathbf{z}_e) + \frac{1}{2} \mathbf{h}^T \nabla_{\mathbf{u}\mathbf{u}}^2 T_0(\mathbf{u}_e; \mathbf{z}_e) \mathbf{h} \\ &\quad + \frac{1}{2} \mathbf{h}^T \nabla_{\mathbf{u}\mathbf{z}}^2 T_0(\mathbf{u}_e; \mathbf{z}_e) \mathbf{k} + \frac{1}{2} \mathbf{k}^T \nabla_{\mathbf{z}\mathbf{u}}^2 T_0(\mathbf{u}_e; \mathbf{z}_e) \mathbf{h} \\ &\quad + \mathcal{O}(\|\mathbf{h}\|_2^3, \|\mathbf{h}\|_2^2 \|\mathbf{k}\|_2, \|\mathbf{h}\|_2 \|\mathbf{k}\|_2^2, \|\mathbf{k}\|_2^3) \end{aligned} \quad (15)$$

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

$$\begin{aligned} &= T_0(\mathbf{u}_e; \mathbf{z}) + \frac{1}{2} (\mathbf{A}\mathbf{u} - (\mathbf{z} - \gamma))^T \text{diag} \left(\frac{\mathbf{1}}{\mathbf{z}_e} \right) (\mathbf{A}\mathbf{u} - (\mathbf{z} - \gamma)) \\ &\quad + \mathcal{O}(\|\mathbf{h}\|_2^3, \|\mathbf{h}\|_2^2 \|\mathbf{k}\|_2, \|\mathbf{h}\|_2 \|\mathbf{k}\|_2^2, \|\mathbf{k}\|_2^2). \end{aligned} \quad (17)$$

The final equality was obtained using the linearity of the inner-product and the fact that $\mathbf{z}_e = \mathbf{A}\mathbf{u}_e + \boldsymbol{\gamma}$. Thus the quadratic Taylor series approximation of $T_0(\mathbf{u}; \mathbf{z})$ about the points $(\mathbf{u}_e; \mathbf{z}_e)$ —modulo an additive constant—is given by

$$\frac{1}{2}(\mathbf{A}\mathbf{u} - (\mathbf{z} - \boldsymbol{\gamma}))^T \text{diag} \left(\frac{\mathbf{1}}{\mathbf{z}_e} \right) (\mathbf{A}\mathbf{u} - (\mathbf{z} - \boldsymbol{\gamma})). \quad (18)$$

Note that this derivation differs from the one present in [7] in that an error present in the expression for $\frac{1}{2}\mathbf{h}^T \nabla_{\mathbf{u}\mathbf{u}}^2 T_0(\mathbf{u}_e; \mathbf{z}_e) \mathbf{h}$ has been corrected. Furthermore, the accuracy in (17) has been revised to account for the fact that obtaining (17) from (16) requires the addition and subtraction of a term of the form $(\mathbf{z} - \mathbf{z}_e)^T \text{diag} \left(\frac{\mathbf{1}}{\mathbf{z}_e} \right) (\mathbf{z} - \mathbf{z}_e)$.

And finally, in [7], \mathbf{z}_e is simply replaced by \mathbf{z} motivated by the fact that $E(\mathbf{z}) = \mathbf{z}_e$. However, we now go one step further and obtain the same approximation from an application of the mean value theorem. The mean value theorem is used by considering the i th component of $\frac{1}{\mathbf{z}-k}$ as a function of k_i . Let $\mathbf{r} = \mathbf{A}\mathbf{u} - (\mathbf{z} - \boldsymbol{\gamma})$, then (18) can be rewritten

$$\begin{aligned} \frac{1}{2}\mathbf{r}^T \left[\mathbf{r} \odot \left(\frac{\mathbf{1}}{\mathbf{z} - \hat{\mathbf{k}}} \right) \right] &= \frac{1}{2}\mathbf{r}^T \left[\mathbf{r} \odot \left(\frac{\mathbf{1}}{\mathbf{z}} + \text{diag} \left(\frac{\mathbf{1}}{(\mathbf{z} - \hat{\mathbf{k}})^2} \right) \mathbf{k} \right) \right] \\ &= \frac{1}{2}\mathbf{r}^T \left(\frac{\mathbf{r}}{\mathbf{z}} \right) + \frac{1}{2}\mathbf{r}^T \left(\frac{\mathbf{r} \odot \mathbf{k}}{(\mathbf{z} - \hat{\mathbf{k}})^2} \right), \end{aligned} \quad (19)$$

where “ \odot ” indicates component-wise multiplication and $0 < |\hat{k}_i| < |k_i|$ for all i . Noting that $\mathbf{r} = \mathbf{A}\mathbf{h} - \mathbf{k}$ and that \mathbf{z} is bounded away from zero, we obtain

$$\frac{1}{2}\mathbf{r}^T \left(\frac{\mathbf{r} \odot \mathbf{k}}{(\mathbf{z} - \hat{\mathbf{k}})^2} \right) = \mathcal{O}(\|\mathbf{h}\|_2^2 \|\mathbf{k}\|_2, \|\mathbf{h}\|_2 \|\mathbf{k}\|_2^2, \|\mathbf{k}\|_2^3), \quad (20)$$

Thus finally, recalling (17), we have the approximation

$$T_0(\mathbf{u}; \mathbf{z}) = T_0(\mathbf{u}_e; \mathbf{z}) + T_0^{\text{wls}}(\mathbf{u}; \mathbf{z}) + \mathcal{O}(\|\mathbf{h}\|_2^3, \|\mathbf{h}\|_2^2 \|\mathbf{k}\|_2, \|\mathbf{h}\|_2 \|\mathbf{k}\|_2^2, \|\mathbf{k}\|_2^2), \quad (21)$$

where

$$T_0^{\text{wls}}(\mathbf{u}; \mathbf{z}) \stackrel{\text{def}}{=} \frac{1}{2} \|\mathbf{Z}^{-1/2}(\mathbf{A}\mathbf{u} - (\mathbf{z} - \boldsymbol{\gamma}))\|^2, \quad \mathbf{Z} = \text{diag}(\mathbf{z}). \quad (22)$$

By exploiting the connection present in (21) between T_0 and T_0^{wls} , we now derive some regularization parameter selection methods for the Poisson likelihood problem. The least squares form of (22) provides a means of extending the standard methods of the discrepancy principle, generalized cross validation, and unbiased predictive risk [27] to (8). We begin with the discrepancy principle.

2.2. The Discrepancy Principle

From (21) and (22) we have

$$E(T_0(\mathbf{u}; \mathbf{z})) \approx T_0(\mathbf{u}_e; \mathbf{z}_e) + E(T_0^{\text{wls}}(\mathbf{u}; \mathbf{z})), \quad (23)$$

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

$$T_0^{\text{wls}}(\mathbf{u}_\alpha; \mathbf{z}) \approx E(T_0^{\text{wls}}(\mathbf{u}_e; \mathbf{z})). \quad (24)$$

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., [12, 21]):

$$\mathbf{z} - \boldsymbol{\gamma} = \mathbf{A}\mathbf{u}_e + \mathbf{e}, \quad (25)$$

where \mathbf{e} is a Gaussian random variable with mean $\mathbf{0}$ and covariance matrix \mathbf{Z} (i.e. $\mathbf{e} \sim N(\mathbf{0}, \mathbf{Z})$) with \mathbf{Z} defined in (22). This in turn implies

$$\mathbf{r}(\mathbf{u}_e) \sim N(\mathbf{0}, \mathbf{I}_n), \quad (26)$$

where

$$\mathbf{r}(\mathbf{u}) \stackrel{\text{def}}{=} \mathbf{Z}^{-1/2}(\mathbf{A}\mathbf{u} - (\mathbf{z} - \boldsymbol{\gamma})). \quad (27)$$

A standard statistical result then tells us that given (26),

$$\|\mathbf{r}(\mathbf{u}_e)\|_2^2 \sim \chi^2(n), \quad (28)$$

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

$$2T_0^{\text{wls}}(\mathbf{u}_\alpha; \mathbf{z}) \approx n. \quad (29)$$

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

$$n - 2\sqrt{2n} \leq 2T_0^{\text{wls}}(\mathbf{u}_\alpha; \mathbf{z}) \leq n + 2\sqrt{2n}. \quad (30)$$

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 [27] applied to the problem of finding the optimal regularization parameter for the minimization problem

$$\arg \min_{\mathbf{u} \geq \mathbf{0}} \left\{ T_0^{\text{wls}}(\mathbf{u}; \mathbf{z}) + \frac{\alpha}{2} \langle \mathbf{u}, \mathbf{C}\mathbf{u} \rangle \right\}. \quad (31)$$

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

$$\alpha_{\text{DP}} = \arg \min_{\alpha > 0} \left(2\tilde{T}_0^{\text{wls}}(\mathbf{u}_\alpha; \mathbf{z}) - n \right)^2, \quad (32)$$

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

$$\tilde{T}_0^{\text{wls}}(\mathbf{u}_\alpha; \mathbf{z}) = \frac{1}{2} \|\mathbf{Z}_\alpha^{-1/2}(\mathbf{A}\mathbf{u}_\alpha - (\mathbf{z} - \boldsymbol{\gamma}))\|, \quad (33)$$

and $\mathbf{Z}_\alpha = \text{diag}(\mathbf{A}\mathbf{u}_\alpha + \boldsymbol{\gamma})$. Note the change in the weighting term \mathbf{Z}_α . This change was implemented because we found that it yielded a better result than using \mathbf{Z} .

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

$$\mathbf{u}_\alpha^k \stackrel{\text{def}}{=} \arg \min_{\mathbf{u} \geq \mathbf{0}} \left\{ \sum_{i \neq k} ([\mathbf{A}\mathbf{u}]_i + \gamma) - z_i \ln([\mathbf{A}\mathbf{u}]_i + \gamma) + \frac{\alpha}{2} \langle \mathbf{u}, \mathbf{C}\mathbf{u} \rangle \right\}, \quad (34)$$

and then choose α to be the minimizer of

$$\text{CV}(\alpha) = \frac{1}{n} \sum_{k=1}^n \{ ([\mathbf{A}\mathbf{u}_\alpha^k]_k + \gamma) - z_k \ln([\mathbf{A}\mathbf{u}_\alpha^k]_k + \gamma) \}. \quad (35)$$

For large-scale problems, minimizing (35) is intractable. The method of generalized cross validation—first introduced by Wahba [28] 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 (21) must be called upon once again:

$$\begin{aligned} \text{CV}(\alpha) &\approx \frac{1}{n} T_0(\mathbf{u}_e; \mathbf{z}) + \frac{1}{2n} \sum_{k=1}^n [\mathbf{r}(\mathbf{u}_\alpha^k)]_k^2 \\ &= \frac{1}{n} T_0(\mathbf{u}_e; \mathbf{z}) + \frac{1}{2n} \sum_{k=1}^n \left(\frac{[\mathbf{r}(\mathbf{u}_\alpha)]_k}{1 - [\mathbf{Z}^{-1/2} \mathbf{A} \mathbf{A}_\alpha]_{kk}} \right)^2. \end{aligned} \quad (36)$$

Here $\mathbf{r}(\mathbf{u})$ is defined as in (27), and (36) 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 [18])

$$\mathbf{D}_\alpha \nabla T_\alpha(\mathbf{D}_\alpha \mathbf{u}; \mathbf{z}) = \mathbf{0}. \quad (37)$$

Now, using the approximation (31) for T_α (motivated by (21)) we can approximate (37) by

$$\mathbf{D}_\alpha \mathbf{A}^T \mathbf{Z}^{-1} (\mathbf{A} \mathbf{D}_\alpha \mathbf{u} - (\mathbf{z} - \gamma)) + \alpha \mathbf{D}_\alpha \mathbf{C} \mathbf{D}_\alpha \mathbf{u} = \mathbf{0}, \quad (38)$$

which has minimum norm solution

$$(\mathbf{D}_\alpha (\mathbf{A}^T \mathbf{Z}^{-1} \mathbf{A} + \alpha \mathbf{C}) \mathbf{D}_\alpha)^\dagger \mathbf{D}_\alpha \mathbf{A}^T \mathbf{Z}^{-1} (\mathbf{z} - \gamma). \quad (39)$$

Thus we define

$$\mathbf{A}_\alpha = (\mathbf{D}_\alpha (\mathbf{A}^T \mathbf{Z}^{-1} \mathbf{A} + \alpha \mathbf{C}) \mathbf{D}_\alpha)^\dagger \mathbf{D}_\alpha \mathbf{A}^T \mathbf{Z}^{-1/2}. \quad (40)$$

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

GCV requires still another approximation, as even with (40), minimizing (36) remains impractical for large-scale problems. The approximation is given by

$$1 - [\mathbf{Z}^{-1/2} \mathbf{A} \mathbf{A}_\alpha]_{kk} \approx \text{trace}(\mathbf{I}_n - \mathbf{Z}^{-1/2} \mathbf{A} \mathbf{A}_\alpha) / n. \quad (41)$$

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

$$\text{GCV}(\alpha) = n T_0^{\text{wls}}(\mathbf{u}_\alpha; \mathbf{z}) / \text{trace}(\mathbf{I}_n - \mathbf{Z}^{-1/2} \mathbf{A} \mathbf{A}_\alpha)^2, \quad (42)$$

where \mathbf{A}_α is given by (40). We note that this is very similar to the GCV function for the weighted least squares problem (31). The difference is that \mathbf{u}_α is computed from (8), which is also used to define \mathbf{D}_α in (40).

Due to the size of the matrix \mathbf{A}_α and the presence of the matrices \mathbf{D}_α and \mathbf{Z} in (40), 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 [27]: given $\mathbf{B} \in \mathbb{R}^{n \times n}$,

$$\mathbf{v} \sim N(\mathbf{0}, \mathbf{I}_n) \quad \text{implies} \quad E(\mathbf{v}^T \mathbf{B} \mathbf{v}) = \text{trace}(\mathbf{B}). \quad (43)$$

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

$$\text{trace}(\mathbf{I}_n - \mathbf{Z}^{-1/2} \mathbf{A} \mathbf{A}_\alpha) \approx \mathbf{v}^T \mathbf{v} - \mathbf{v}^T \mathbf{Z}^{-1/2} \mathbf{A} \mathbf{A}_\alpha \mathbf{v}. \quad (44)$$

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

$$\tilde{\text{GCV}}(\alpha) \approx n T_0^{\text{wls}}(\mathbf{u}_\alpha; \mathbf{z}) / (\mathbf{v}^T \mathbf{v} - \mathbf{v}^T \mathbf{Z}^{-1/2} \mathbf{A} \mathbf{A}_\alpha \mathbf{v})^2, \quad (45)$$

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, $\mathbf{A}_\alpha \mathbf{v}$ in (45) must be approximated by applying a truncated conjugate gradient iteration with $\mathbf{x}_0 = \mathbf{0}$ to

$$\mathbf{D}_\alpha (\mathbf{A}^T \mathbf{Z}^{-1} \mathbf{A} + \alpha \mathbf{C}) \mathbf{D}_\alpha \mathbf{x} = \mathbf{D}_\alpha \mathbf{A}^T \mathbf{Z}^{-1/2} \mathbf{v}. \quad (46)$$

In the case of large-scale least squares estimation problems, an efficient method for approximating the GCV when randomized trace estimation is used is presented in [11]. This method does not require finding the solution of (46), something which for large-scale problems can be very computationally expensive.

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 UPRE [27] 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 (21). From it, we have

$$T_0(\mathbf{u}_\alpha; \mathbf{z}_e) \approx T_0(\mathbf{u}_e; \mathbf{z}_e) + T_0^{\text{wls}}(\mathbf{u}_\alpha, \mathbf{z}_e), \quad (47)$$

which we will call the predictive error, whereas

$$T_0(\mathbf{u}_\alpha; \mathbf{z}) \approx T_0(\mathbf{u}_e; \mathbf{z}) + T_0^{\text{wls}}(\mathbf{u}_\alpha, \mathbf{z}). \quad (48)$$

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

$$E(T_0(\mathbf{u}_\alpha; \mathbf{z}_e)) \approx T_0(\mathbf{u}_e; \mathbf{z}_e) + E(T_0^{\text{wls}}(\mathbf{u}_\alpha, \mathbf{z}_e)), \quad (49)$$

$$E(T_0(\mathbf{u}_\alpha; \mathbf{z})) \approx T_0(\mathbf{u}_e; \mathbf{z}_e) + E(T_0^{\text{wls}}(\mathbf{u}_\alpha, \mathbf{z})). \quad (50)$$

Following arguments in [27, Section 7.1] and using the approximate regularization operator \mathbf{A}_α defined by (40), we find that

$$E(T_0^{\text{wls}}(\mathbf{u}_\alpha, \mathbf{z}_e)) = \frac{1}{2} \left\| (\mathbf{Z}^{-1/2} \mathbf{A} \mathbf{A}_\alpha - \mathbf{I}_n) \mathbf{Z}^{-1/2} \mathbf{A} \mathbf{u}_e \right\|^2 + \frac{1}{2} \text{trace}((\mathbf{Z}^{-1/2} \mathbf{A} \mathbf{A}_\alpha)^2), \quad (51)$$

$$\begin{aligned} E(T_0^{\text{wls}}(\mathbf{u}_\alpha, \mathbf{z})) &= \frac{1}{2} \left\| (\mathbf{Z}^{-1/2} \mathbf{A} \mathbf{A}_\alpha - \mathbf{I}_n) \mathbf{Z}^{-1/2} \mathbf{A} \mathbf{u}_e \right\|^2 + \frac{1}{2} \text{trace}((\mathbf{Z}^{-1/2} \mathbf{A} \mathbf{A}_\alpha)^2) \\ &\quad - \text{trace}(\mathbf{Z}^{-1/2} \mathbf{A} \mathbf{A}_\alpha) + \frac{n}{2}. \end{aligned} \quad (52)$$

Thus we have the following estimation of the predictive risk:

$$\begin{aligned} E(T_0(\mathbf{u}_\alpha, \mathbf{z}_e)) &\approx T_0(\mathbf{u}_e; \mathbf{z}_e) + E(T_0^{\text{wls}}(\mathbf{u}_\alpha, \mathbf{z}_e)) \\ &= E(T_0^{\text{wls}}(\mathbf{u}_\alpha, \mathbf{z})) + \text{trace}(\mathbf{Z}^{-1/2} \mathbf{A} \mathbf{A}_\alpha) - \frac{n}{2}. \end{aligned} \quad (53)$$

This motivates our choice of the UPRE function

$$\text{UPRE}(\alpha) = T_0^{\text{wls}}(\mathbf{u}_\alpha; \mathbf{z}) + \text{trace}(\mathbf{Z}^{-1/2} \mathbf{A} \mathbf{A}_\alpha) - \frac{n}{2}. \quad (54)$$

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 simulate star field data, and due to the fact that we could not visualize it otherwise, plot it on the right on a log scale in Fig. 1.

Generating corresponding blurred, noisy data requires a discrete PSF \mathbf{a} , which we compute using the Fourier optics PSF physical model [10, 20, 27]

$$\mathbf{a} = \left| \mathbf{F}^{-1} \left\{ \mathbf{p} \odot e^{i\phi} \right\} \right|^2, \quad (55)$$

where “ \odot ” denotes component-wise multiplication; \mathbf{p} is the $\sqrt{n} \times \sqrt{n}$ indicator array over the telescopes pupil (usually an annulus); ϕ is the $\sqrt{n} \times \sqrt{n}$ array that represents the aberrations in the incoming wavefronts of light (obtained using the Kolmogorov turbulence model [20]); $i = \sqrt{-1}$; 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

$$\mathbf{A} \mathbf{u} = \text{vec}(\text{ifft2}(\hat{\mathbf{a}} \odot (\text{fft2}(\mathbf{u}))))), \quad \hat{\mathbf{a}} = \text{fft2}(\text{fftshift}(\mathbf{a})), \quad (56)$$

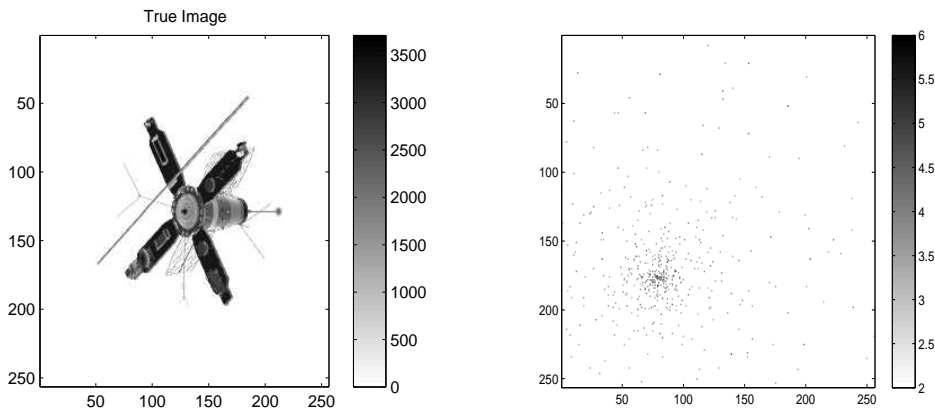


Figure 1. True image of the satellite on the left and the true image of the star field, plotted on a log scale, with entries less than 100 set to 0, on the right.

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} [16].

In order to generate data that is as accurate as possible, we use the full statistical model of [23] (rather than (2)) for CCD camera noise for generating our data:

$$\mathbf{z} = \text{Pois}(\mathbf{A}\mathbf{u}_e + \boldsymbol{\gamma}) + N(\mathbf{0}, \sigma^2 \mathbf{I}_n), \quad (57)$$

where the iid Gaussian term $N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, with \mathbf{I}_n the $n \times n$ identity matrix, models instrument noise. We assume pixel-wise independence as well as independence among the two random variable on the right in (57).

We generate the noisy signal in MATLAB using the `poissrnd` and `randn` functions, and we choose the values $\gamma = 10$ and $\sigma = 5$, which are realistic values for these parameters. Corresponding blurred, noisy data is plotted in Figure 2. Again, note that the star field data is plotted on a log scale (as in Figure 1) so that it look much worse than it is.

In solving the inverse problem, we approximate (57) by (2) because the resulting maximum likelihood computational problem is much more tractable [24]. This is accomplished using the approximation $N(\sigma^2, \sigma^2) \approx \text{Pois}(\sigma^2)$. Then using independence properties (57) becomes by

$$\mathbf{z} + \boldsymbol{\sigma}^2 = \text{Pois}(\mathbf{A}\mathbf{u}_e + \boldsymbol{\gamma} + \boldsymbol{\sigma}^2), \quad (58)$$

where $\boldsymbol{\sigma}^2 = \sigma^2 \mathbf{1}$. This approximation is used in the implementation of the methods given in section 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 SNR, which for

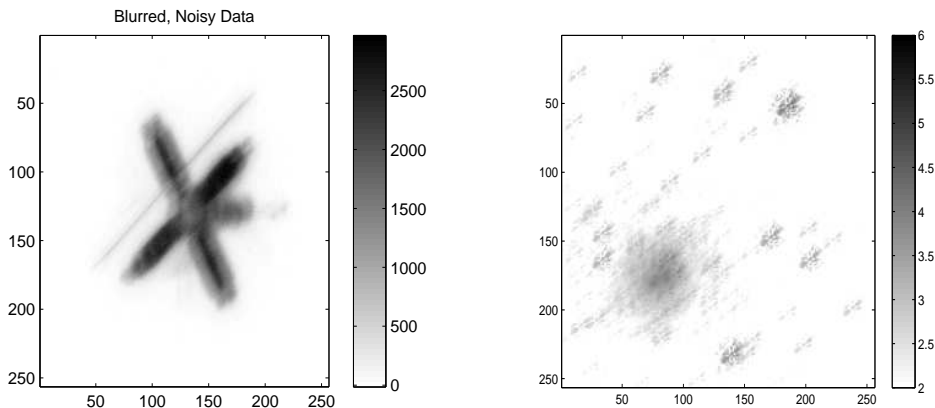


Figure 2. Blurred, noisy images of the satellite on the left, and the star field, plotted on a log scale, with entries less than 100 set to 0, on the right, both with $SNR = 30$.

(57) is defined as

$$SNR = \sqrt{\frac{\|\mathbf{A}\mathbf{u}_e + \boldsymbol{\gamma}\|^2}{E(\|\mathbf{z} - (\mathbf{A}\mathbf{u}_e + \boldsymbol{\gamma})\|^2)}}. \quad (59)$$

Note that under the square root, the noise-free signal (mean of \mathbf{z}) is in the numerator, whereas the noise power (variance of \mathbf{z}) is in the denominator. For model (57), we have

$$E(\|\mathbf{z} - (\mathbf{A}\mathbf{u}_e + \boldsymbol{\gamma})\|^2) = \sum_{i=1}^n ([\mathbf{A}\mathbf{u}_e]_i + \gamma_i + \sigma^2). \quad (60)$$

We note that this is the traditional definition of signal-to-noise ratio. However, in the context of image processing this definition has its flaws and a more meaningful definition is possible (see e.g. [22]).

3.1. Regularization Functions

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 $\boldsymbol{\Theta}$. Like \mathbf{L} , $\boldsymbol{\Theta}$ is a discretization of a diffusion operator. However, unlike \mathbf{L} , it allows for the formation of edges in reconstructions when the data/inverse model suggest it and should **only** be used in the case that the user knows *a priori* that the object contains edges. It is defined as follows:

$$\boldsymbol{\Theta} = \mathbf{D}_x^T \boldsymbol{\Lambda} \mathbf{D}_x + \mathbf{D}_y^T \boldsymbol{\Lambda} \mathbf{D}_y, \quad (61)$$

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 less than 1 for pixels near an edge.

The diagonal matrix $\mathbf{\Lambda}$ in (61) is defined as follows. First, let

$$\mathbf{v} = [\mathbf{D}_x \mathbf{u}_{\text{approx}}]^2 + [\mathbf{D}_y \mathbf{u}_{\text{approx}}]^2, \quad (62)$$

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

$$[\mathbf{v}_\epsilon]_i = \begin{cases} v_i & v_i > \epsilon \|\mathbf{v}\|_\infty \\ 0 & \text{otherwise,} \end{cases}, \quad (63)$$

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

$$\mathbf{\Lambda} = \text{diag} \left(\max \left\{ \frac{\mathbf{1}}{\mathbf{1} + \mathbf{v}_\epsilon}, \frac{1}{10} \right\} \right). \quad (64)$$

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 (and hence decreasing smoothing), whereas when $[\mathbf{v}_\epsilon]_i \approx 0$ the regularization parameter remains approximately the same. The result is that for regions in which the existence of an edge is indicated, sharp changes in intensity will incur a relatively lesser penalty than in regions in which the image is assumed to be smooth. We note that the value of ϵ in (63) and, in particular, the value $\frac{1}{10}$ in (64) may need to be modified for different applications.

3.2. Regularization Parameter Selection and Results

In each of DP, GCV, and UPRE, a minimization problem must be solved: for DP it's (32); for GCV, $\text{GCV}(\alpha)$ defined by (45) must be minimized subject to $\alpha > 0$; whereas for UPRE, $\text{UPRE}(\alpha)$ defined by (54) 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. We used a lower bound of zero and an upper bound of 0.01. The `Tolx` (step tolerance) parameter was set to 10^{-8} for the satellite example and 10^{-10} for the star field example.

Moreover, when DP is used in the examples below and (32) approximately solved, the minimum value computed by `fminbnd` is 10^{-8} in all cases but one, in which case it is 10^{-6} . Thus we can say that the weighted residuals for the reconstructions computed using DP are, in all cases, very near to the noise level; that is, assuming that (27) is correct.

Solving the minimization problems stemming from the three methods requires the computation of \mathbf{u}_α . Recall that \mathbf{u}_α is the solution to (8), which is itself a minimization problem. The solution of (8) is found by applying the gradient-projection reduced-Newton method outlined in [27]. This is an iterative method and the iterations are

stopped if either the step-norm tolerance is satisfied, the gradient-norm tolerance is satisfied, or the number of iterations grows past a certain value.

For $\mathbf{C} = \mathbf{I}_n$, we perform experiments on both the satellite and star field data with SNR = 5, 10, 30 and 100. For $\mathbf{C} = \mathbf{L}$ or $\mathbf{C} = \mathbf{\Theta}$, we perform experiments on the satellite data with SNR = 10 and 30. In each case, we calculate the relative error,

$$\frac{\|\mathbf{u}_e - \mathbf{u}_\alpha\|}{\|\mathbf{u}_e\|}, \quad (65)$$

over a range of values of α and the points corresponding to the recommendations by the various selection methods. We use relative error as a measure of the goodness of the regularized approximation because this is standard (see e.g. [27])

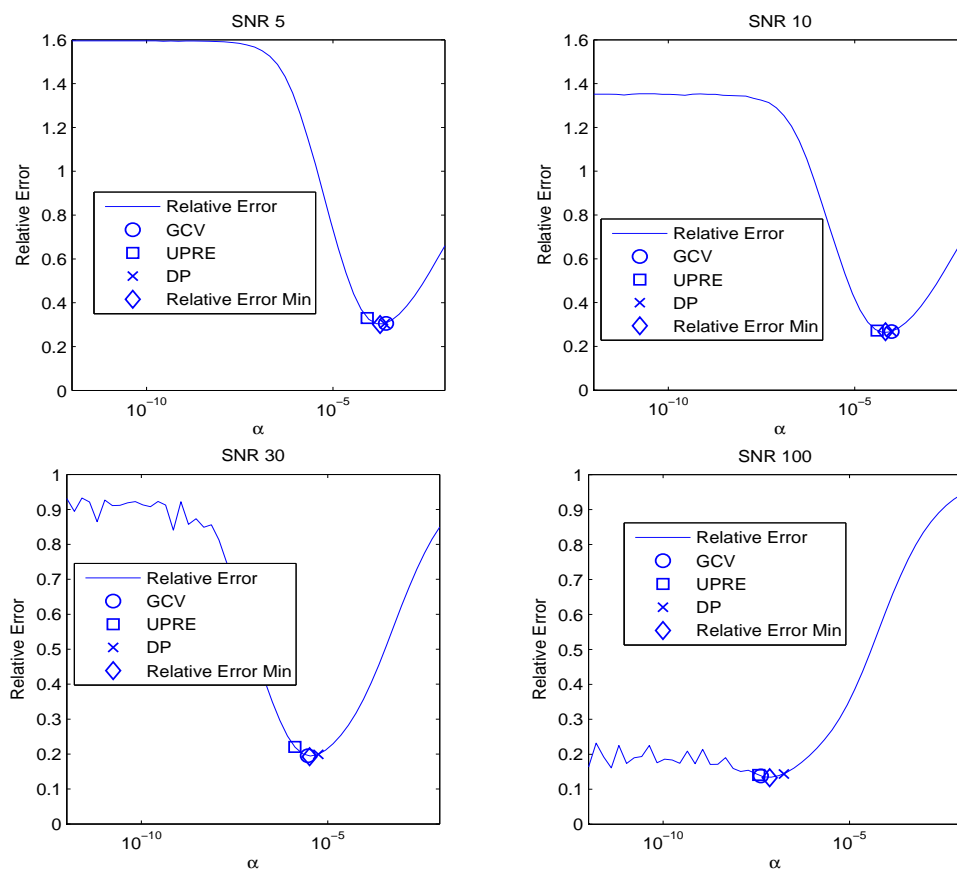


Figure 3. *Satellite Test Case, with $\mathbf{C} = \mathbf{I}_n$. Plot of relative error together with the values of α chosen by the regularization parameter selection methods GCV, UPRE, DP.*

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, with DP giving a slightly better value of α than GCV and UPRE. Note that as α tends toward zero, \mathbf{u}_α tends toward the unregularized solution, which is far from the true solution since our problem is ill-posed. On the other hand, when α is too large, the reconstruction is also far from the true solution because too little emphasis is given to the data/inverse

model. Such relative error plots vs. regularization parameter are standard for ill-posed problems (see e.g. [27]).

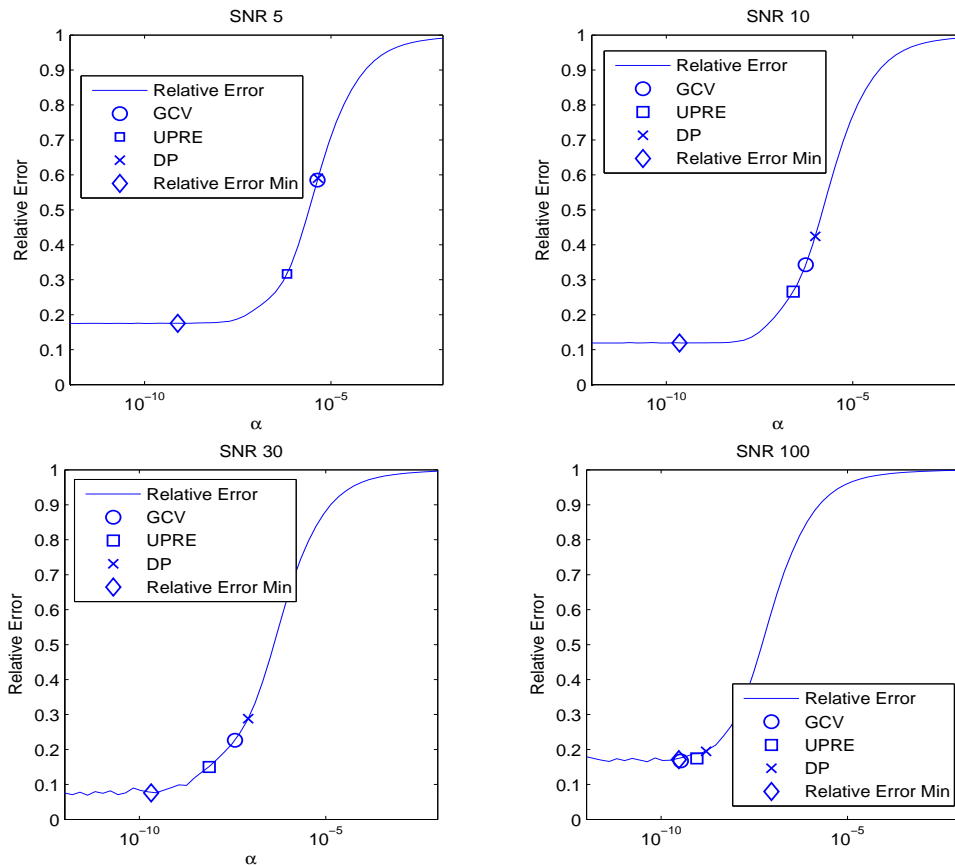


Figure 4. *Star Field Test Case. Plot of relative error together with the values of α chosen by the regularization parameter selection methods GCV, UPRE, DP.*

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 30 and 100 than in the cases of 5 and 10. Here, UPRE gave a better recommendation than GCV and DP. The plots indicate that the optimal value of α is very small and so regularization might not be necessary in this case. One reason for this is that the nonnegativity constraints coupled with the point source nature of the object have been shown to have a stabilizing effect on inverse solution methods [2]. This stabilization shows itself in the relatively flat nature of the relative error curves as $\alpha \rightarrow 0$.

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 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

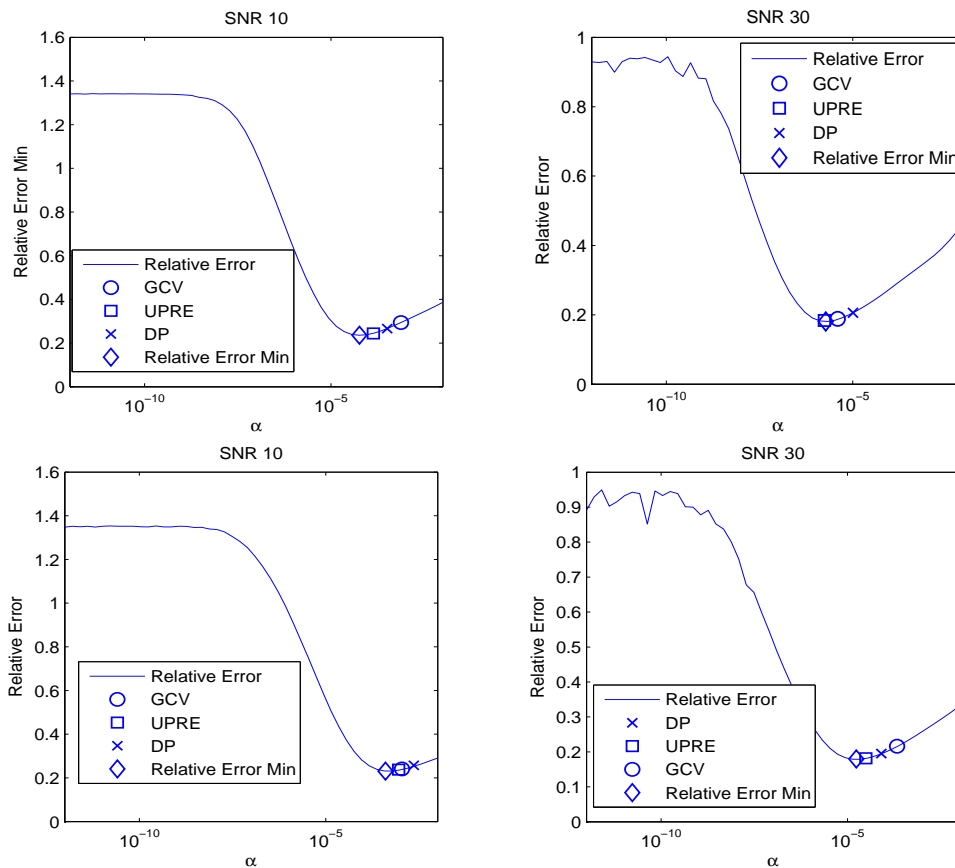


Figure 5. In the top row, Satellite Test Case, with $\mathbf{C} = L_n$. In the bottom row, Satellite Test Case, with $\mathbf{C} = \Theta$. Plot of relative error together with the values of α chosen by the regularization parameter selection methods GCV, UPRE, DP.

UPRE recommendation for the value of α . In the case of SNR = 10, the GCV and UPRE methods yielded a slightly better value of α than DP, while for SNR = 30, UPRE gave the best value of α .

Reconstructions of the satellite at a SNR of 30, and each of the three regularization functions, are given in Fig. 6. In Fig. 6, we have also plotted the reconstruction of the star field data with an SNR or 30 with Tikhonov regularization. In addition, in Fig. 7, we have plotted the weighted residuals (27), which should, ideally, approximate white noise. As far as that goes, for the satellite case it seems that Tikhonov regularization yields the least desirable reconstruction, while $\mathbf{C} = \Theta$ yields the most desirable reconstruction.

These results indicate that the three methods give useful recommendations for the value of the regularization parameter as well as good corresponding reconstructions of the unknown object. The three methods yielded values for α that were very similar for each of the four data sets.

For least squares problems, it is well-known that GCV is slightly more flexible than UPRE and DP due to the fact that it does not require knowledge of the noise variance [27]. However this advantage does not extend to the Poisson case if we assume

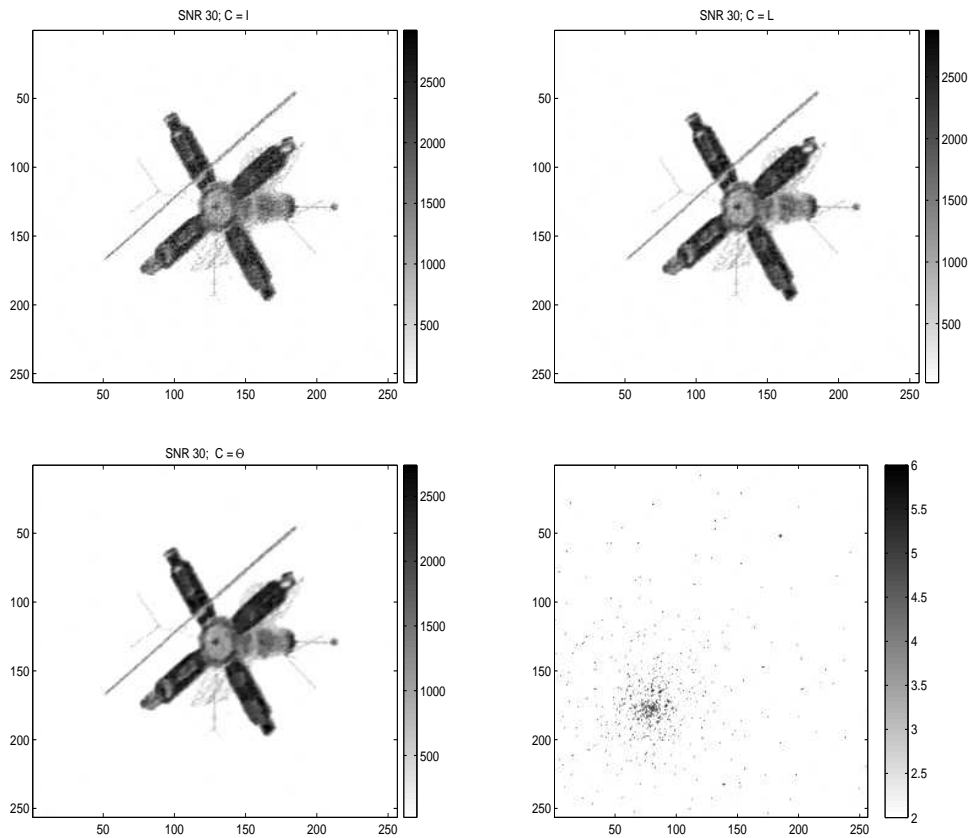


Figure 6. Reconstructions of the satellite: on the upper-left with Tikhonov regularization ($\mathbf{C} = \mathbf{I}$), on the upper-right with Laplacian regularization ($\mathbf{C} = \mathbf{L}$), and on the lower-left with $\mathbf{C} = \Theta$. Reconstruction of the star field with Tikhonov regularization is given (on a log scale and with entries less than 100 set to 0) on the lower-right.

a statistical model of the form (2) and use the approximation (21),(22).

The motivation for DP is the simplest, as is its implementation. Moreover, its connection to the χ^2 test and variance matching of the weighted residuals with the (approximate) true variances is likely appealing to the applied scientist.

On the other hand, theoretical analyses of these three methods in the least squares case [27] indicate that UPRE has the best asymptotic convergence properties and that GCV has better asymptotic convergence properties than DP. In order for us to draw the same conclusions in the Poisson case, we would have to extend these analyses, which we feel is possible using the Taylor series arguments of this paper. However, even if this is the case, the implementations of GCV and UPRE require two additional approximations: that of the (nonlinear) regularization operator by (40), and the use of (43) for the trace approximation.

Finally, considerations of other statistical properties, such as the bias, of the estimator \mathbf{u}_α of \mathbf{u}_e may lead to different conclusions; see [25] for a discussion of this subject when the noise is additive Gaussian.

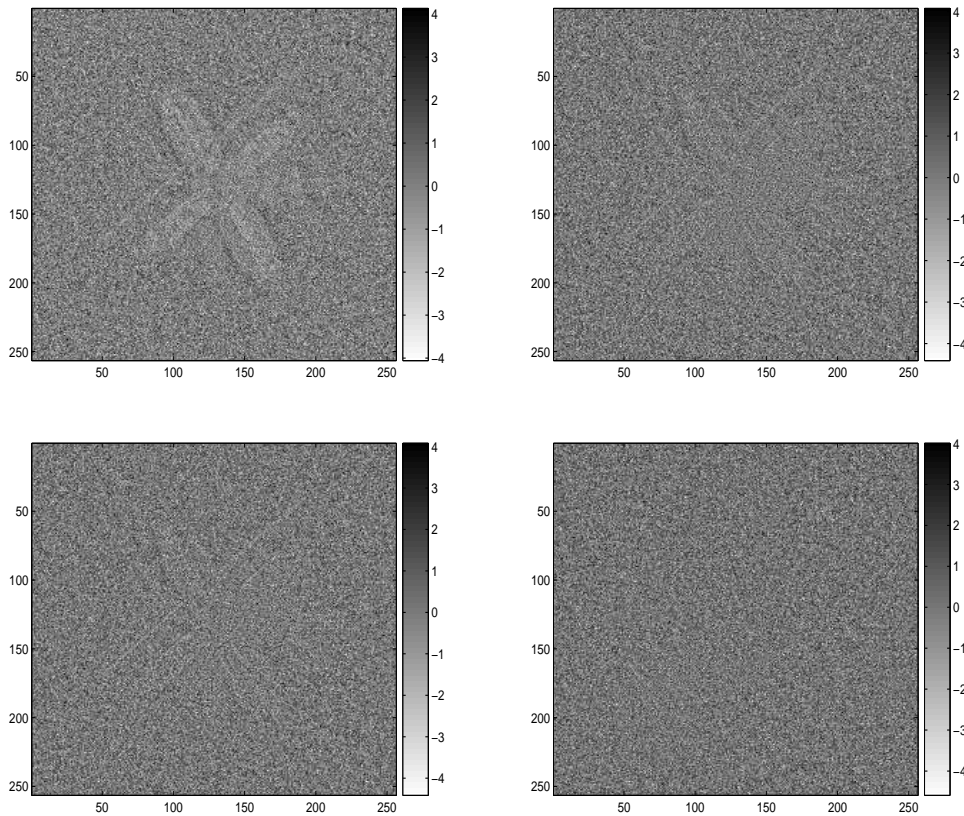


Figure 7. *Weighted residuals (27) of the satellite reconstruction in Fig. 6: on the upper-left with Tikhonov regularization ($\mathbf{C} = \mathbf{I}$), on the upper-right with Laplacian regularization ($\mathbf{C} = \mathbf{L}$), and on the lower-left with $\mathbf{C} = \Theta$. Weighted residuals (27) of the star field with Tikhonov regularization on the lower-right.*

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 the problem is ill-posed and the data is noise corrupted, regularization is required for computing stable solutions to (3). 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 (22)). Using this approximation, we have applied the DP, GCV and 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 (57). 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 regularization parameter selection method than GCV and UPRE for the examples considered here, however the recommendations were very close. The GCV and UPRE methods were very similar for the examples under consideration.

There are several possibilities for future work along the lines set forth in this paper. For one, an extension of the theoretical analyses of DP, GCV, and UPRE found in [27] seems possible using the Taylor series approximation of the negative-log Poisson likelihood presented here. Secondly, these methods should also be useable for non-quadratic regularization functions such as total variation, though this extension is immediate. Thirdly, a more efficient method than MATLAB's `fminbnd` for solving (32) and minimizing (45) and (54) could likely be developed. And finally, using these methods in the context of positron emission tomography, where (2) is also used, is a subject of our current work.

Acknowledgements

We would like to thank the referees for their many helpful comments. The paper is much improved because of your efforts.

References

- [1] R. C. Aster, B. Borchers, and Clifford H. Thurber, *Parameter Estimation and Inverse Problems*, Elsevier 2005.
- [2] J. M. Bardsley, J. Merikoski, and R. Vio, *The Stabilizing Properties of Nonnegativity Constraints in Least-Squares Image Reconstruction*, International Journal of Pure and Applied Mathematics, vol. 43, no. 1, 2008, pp. 95-109.
- [3] J. M. Bardsley and N. Laobeul, *Tikhonov Regularized Poisson Likelihood Estimation: Theoretical Justification and a Computational Method*, Inverse Problems in Science and Engineering, Volume 16, Issue 2, January 2008, pp. 199-215.
- [4] J. M. Bardsley and N. Laobeul, *An Analysis of Regularization by Diffusion for Ill-Posed Poisson Likelihood Estimation*, Inverse Problems in Science and Engineering, Volume 17, Issue 4 June 2009, pp. 537-550.
- [5] J. M. Bardsley and A. Luttman, *Total Variation-Penalized Poisson Likelihood Estimation for Ill-Posed Problems*, Advances in Computational Mathematics, Advances in Computational Mathematics, Volume 31, Issue 1, (2009), pp. 35-59.
- [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] J. 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] J. M. Bardsley and J. A. Goldes, *An Iterative Method for Edge-Preserving MAP Estimation when Data-Noise is Poisson*, to appear in the SIAM Journal on Scientific Computing.

- [9] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Kluwer Academic Press 1996.
- [10] J. W. Goodman, *Introduction to Fourier Optics, 2nd Edition*, McGraw-Hill, 1996.
- [11] G. Golub and U. von Matt, *Generalized Cross-Validation for Large-Scale Problems*, Journal of Computational and Graphical Statistics, Vol. 6, No. 1, 1997, pp. 1-34.
- [12] 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.
- [13] E. Haber and E. Oldenburg, *A GCV Based Method for Non-Linear Ill-Posed Problems*, Computational Geosciences, 4, 2000, pp. 41-63.
- [14] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, Springer 2005.
- [15] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems*, SIAM, Philadelphia, 1997.
- [16] P. C. Hansen, J. G. Nagy, and D. P. O’Leary, *Deblurring Images: Matrices, Spectra, and Filtering*, SIAM, Philadelphia, 2006.
- [17] V. A. Morozov, *On the Solution of Functional Equations by the Method of Regularization*, Soviet Mathematics Doklady, vol. 7, 1966, pp. 414-417.
- [18] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer-Verlag, New York, 1999.
- [19] J. M. Ollinger and J. A. Fessler, *Positron-Emission Tomography*, IEEE Signal Processing Magazine, January 1997.
- [20] M. C. Roggeman and Byron Welsh, *Imaging Through Turbulence*, CRC Press, 1996.
- [21] B. Rust, *Truncating the Singular Value Decomposition for Ill-Posed Problems*, NISTIR 6131, U.S. Dept. of Commerce, July 1998.
- [22] [http://en.wikipedia.org/wiki/Signal_to_noise_ratio_\(image_processing\)](http://en.wikipedia.org/wiki/Signal_to_noise_ratio_(image_processing))
- [23] 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.
- [24] D. L. Snyder, C. W. Helstrom, A. D. Lanterman, M. Faisal, R. L. White, *Compensation for Readout Noise in CCD Images*, Journal of the Optical Society of America A, **12** (1995), pp. 272-283.
- [25] S. N. Evans and P. B. Stark, *Inverse Problems as Statistics*, Technical Report 609, Department of Statistics, U.C., 2002.
- [26] 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.
- [27] C. R. Vogel, *Computational Methods for Inverse Problems*, SIAM, Philadelphia, 2002.
- [28] 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.
- [29] D. F. Yu and J. A. Fessler, *Edge-Preserving Tomographic Reconstruction with Nonlocal Regularization*, IEEE Trans. on Medical Imaging, 21(2), 2002, pp. 159-173.