Statistical Tests for Total Variation
Regularization Parameter Selection

Jodi L. Mead

Abstract

Total Variation (TV) is an effective method of removing noise in digital image processing while preserving edges [23]. The choice of scaling or regularization parameter in the TV process defines the amount of denoising, with value of zero giving a result equivalent to the input signal. Here we explore three algorithms for specifying this parameter based on the statistics of the signal in the total variation process. The Discrepancy Principle, a new algorithm based on the $\chi^2$ method for Tikhonov regularization [17]–[21], and an "empirically Bayesian" approach suggested in [9]. In all three algorithms TV regularization is viewed as an M-estimator [3] and it is assumed to converge to a well defined limit even if the probability model is not correctly specified. These regularization parameter selection algorithms are implemented in such a way that they can supplement any TV optimization algorithm. The algorithms are useful for computationally large problems because a single regularization parameter is found that satisfies an appropriate statistical test, and the regularization parameter does not need to be manually adjusted, or iterated to zero. This is especially useful for nonlinear problems where an underlying linear problem is solved iteratively, taking the guesswork out of choosing the regularization parameter in each iterate.

Index Terms

discrepancy principle, regularization parameter, total variation (TV), parameter estimation, statistical distributions

I. INTRODUCTION

Most image processing problems require regularization due to noise in images. Regularization alleviates ill-posedness in inverse problems by adding additional assumptions about the solution. The choice of

J. Mead is with the Mathematics Department, Boise State University, Boise, ID, 83725 USA e-mail:jmead@boisestate.edu. This work was supported by National Science Foundation, DMS10431047
regularization parameter is an important issue in this process because the parameter controls the amount of regularization. One approach to parameter selection is to solve a sequence of regularized problems iteratively with the regularization parameter decreasing to zero. In this case the initial inverse estimate is updated at each iterate so that regularization is no longer necessary at the end of the iterative procedure. Alternatively, one could solve a single regularized problem and appropriately weight information contained in the regularization term. The latter view is taken here because it can be computationally less expensive than iterating the inversion. However, the difficulty with finding an appropriate weight is that the regularization term is then treated as prior information and there is potential to add bias to the solution. This is addressed in Bayesian methods by identifying prior probability models. Alternatively, in this work we use simple statistics of the problem to find appropriate weights.

Choice of regularization parameter for Tikhonov regularization is well studied and common methods include the Discrepancy Principle, L-curve and generalized cross-validation (GCV) [12]. More recently, some of these approaches have been extended to TV regularization, however, the theory behind the L-curve and GCV do not readily transfer to TV functionals. In [14] they exploit GCV to determine how to iteratively update the regularization parameter, while in [15] they explore a similar functional, the Unbiased Predictive Risk Estimator (UPRE), to choose the regularization parameter. More closely related to the work described here, the Discrepancy Principle was used in [24] to find the regularization parameter. In that work the authors represent the TV norm by the dual formulation and change the minimization problem into a minimax problem. This is similar to an approach to the optimization problem in [5] where the regularization parameter represents the threshold for the duality gap in the constrained optimization problem. In [5] the bound is chosen to be proportional to the relative noise level for denoising while for deblurring it is dominated by smallest eigenvalue of the rank-deficient approximation. This is similar to the Discrepancy Principle, and in [5] they find that the value of this parameter does not effect the number of iterations needed to solve the problem.

In [6] they use a spatially dependent regularization parameter in the ROF model [5] model for image restoration. The strategy for choosing the regularization parameter is automated whereby the data fit a constraint satisfying a Discrepancy Principle. All of these strategies rely the variance estimators of the data, and hence empirically give a Bayesian framework, at least for the observations. In this work, we also suggest a strategy that in addition uses the variance estimators of the total variation functional or prior. The difference between this approach and that in [5], [6], [24] is analogous to the differences between the Discrepancy Principle and the $\chi^2$ method [17]–[21] for Tikhonov regularization. We are able to extend the $\chi^2$ method to TV regularization by using results in [9] that give the statistics for the
total variation functional. The difference between the $\chi^2$ method and the Discrepancy Principle is subtle, but can be significant in inverse problems where the number of data and parameters are both large.

In Section IV we show results from TV regularization and compare regularization parameter selection using the Discrepancy Principle, the $\chi^2$ method extended to TV regularization which we call the statistical test method, and a similar approach in [9] for image processing. Implementation of the Discrepancy Principle differs from that in [24] in that the computed inverse estimate is used implicitly to estimate the regularization parameter, rather than using an analytical approximation of the inverse estimate. In addition, none of these implementations require a particular TV optimization algorithm. Rather, we suggest a strategy that can be used to find the regularization parameter with any TV optimization algorithm. Our goal is to provide simple algorithms to effectively compute a regularization parameter, and can be used in nonlinear or computationally expensive problems.

In Section II we introduce an extension of the $\chi^2$ method to TV regularization by discussing Tikhonov regularization in the context of prior information. Then in Section III we describe the statistical test approach with TV regularization, in addition to implementation of the Discrepancy Principle and the approach suggested by [9]. Pseudocode for each algorithm are also given in Section III. We conclude in Section V with a discussion of the strengths and weaknesses of each algorithm, and of the general approach to regularization parameter selection in this work.

II. LEAST SQUARES ESTIMATORS

Regularization parameter selection is well studied for Tikhonov regularization [12] in linear problems. We will briefly review it in a framework that introduces the statistical test strategy proposed here for TV.

Linear inverse problems are often stated as

$$\mathbf{d} = \mathbf{A}\mathbf{x} + \mathbf{\epsilon}$$  \hspace{1cm} (1)

where for two dimensional image reconstruction problems $\mathbf{d} \in \mathbb{R}^{m^2}$ is a vector of observations, $\mathbf{A} \in \mathbb{R}^{m^2 \times n^2}$ is a known model, and $\mathbf{x} \in \mathbb{R}^{n^2}$ contains the unknown information about the true image. The vector $\mathbf{\epsilon}$ represents noise in the observations and it is assumed it has zero mean with variance $\sigma^2$. In image processing, these vectors and matrix are specified by concatenating in a column- or row-wise fashion pixel values for the image. The matrix $\mathbf{A}$ can use used to represent a shift in the image, in which case the image is noisy, or $\mathbf{A}$ may represent a blur filter induced by motion in the image. In either case, $a_{ij}$ acts on the value of the image pixel at the point indexed by $(i, j)$. For example an image represented by $256 \times 256$ pixels, has $m = n = 256$. 
A least squares fit of the unknowns to the observation solves the minimization problem

$$\min_x \|d - Ax\|_2^2.$$  \hspace{1cm} (2)

Typically this problem is ill-posed so a regularization term, or constraints, are added and in a least squares context this takes the form:

$$\hat{x} = \arg\min_x \|d - Ax\|_2^2 + \alpha^2 \|L(x - x_p)\|_2^2.$$  \hspace{1cm} (3)

The matrix $L$ is a shaping matrix and $x_p$ is an initial guess that is often taken to be zero. If $\alpha^2$ chosen large, the regularization term dominates and more weight will be placed on the second misfit, hence the estimate $\hat{x}$ will be "close" to $x_p$. If $L$ is chosen to be the identity, there can be significant smoothing of the solution which may not be desirable in imaging applications. If $L$ is diagonal or dense, discontinuous least squares estimates result [20], and applying $L$ has the effect of applying a spatially dependent regularization parameter [6]. This regularization process can also be viewed as constrained linear inversion [5] or ridge regression [16].

In Bayesian statistics $x$ is treated as a random variable $x$. If $x_p$ is the mean of $x$ and $\alpha^2 L^T L$ is the inverse error covariance for this estimate, the optimal parameters found by Tikhonov regularization can be viewed as the maximum a posteriori estimate when error in the initial estimate and data are normally distributed. Using Bayes theorem, the optimal parameters are described by the posterior distribution of $x$ given by

$$p(x|d) = p(d|x)p(x).$$

It is not reasonable to assume the error distributions are always normal. However, if errors are not necessarily normal, reasonable estimates can still be found with Tikhonov regularization as long as bias in the data and initial guess are removed, and the misfits are weighted by their inverse covariances. In practice this can be done with rough initial estimates of the parameters, and data weighted with a diagonal matrix containing inverse variances of their errors. The difficulty then lies in estimating variances for errors in initial parameter estimates. Estimating this prior information is the focus of many Bayesian methods, for example in [2] they use a hierarchal Bayesian model for TV in blind deconvolution.

Most regularization methods focus on choice of parameter $\alpha^2$, and this can be viewed as finding a scalar estimate of the inverse variance of the error in the initial parameter estimate. For example, the Discrepancy Principle is a method whereby a value of $\alpha^2$ is found so that the optimal estimate $\hat{x}$ defined by (3) satisfies

$$\|d - A\hat{x}\|_2^2 = \tau m^2 \sigma^2$$  \hspace{1cm} (4)
where $\tau$ is a predetermined real number. This can be viewed as a $\chi^2$ test but the degrees of freedom in the test statistic are not correct, and this often leads to an over smoothed solutions [20]. The following $\chi^2$ test on the regularized problem gives the $\chi^2$ method [17], [18] for choice of $\alpha^2$ in Tikhonov regularization:

\[
\frac{1}{\sigma^2} \|d - Ax\|_2^2 + \alpha^2 \|L(\hat{x} - x_p)\|_2^2 \approx m^2. \tag{5}
\]

Even though the Discrepancy Principle and the $\chi^2$ method can both be viewed as finding a regularization parameter based on a $\chi^2$ test, they differ in the degrees of freedom, and that can be significant when there are a large number of data and parameters. The degrees of freedom in the $\chi^2$ statistic $\|d - Ax\|_2^2$ is $m^2 - n^2$ [1], while the Discrepancy Principle in (4) assumes the degrees of freedom is $m^2$. This can be accommodated by choosing $\tau \approx 0$.

The $\chi^2$ test still holds even if the errors are not normally distributed [17], as long as there is a large number of data. In this work we assume there are good estimates of $\sigma$, or that the model $A$ is premultiplied by a matrix representing the square root of the data error covariance, and the goal is to find the regularization parameter $\alpha^2$ or weight on the TV regularization term. Alternatively in [18] it shown how to find $\sigma$ given the regularization weight $\alpha^2$ for Tikhonov regularization. Furthermore, in [20], multiple $\chi^2$ tests are derived to find multiple parameters or matrix weights. Future work involves extending these ideas to TV regularization.

Here we extend the $\chi^2$ test which finds the regularization parameter $\alpha^2$ from Tikhonov regularization to TV. The TV functional does not follow a $\chi^2$ distribution, therefore we call the approach a “statistical test” to find the regularization parameter rather than a $\chi^2$ test. We will motivate the test by using results in [9] that state that the TV functional follows a Laplacian distribution.

### III. Least Squares with Total Variation Regularization

Ill-conditioning of the image reconstruction problem defined by (2) is often alleviated by TV regularization, because the method results in reconstructions that maintain sharp edges in the image. With this type of regularization, the prior information is a gradient vector containing the finite difference approximation to partial derivatives with appropriate boundary conditions. The discrete total variation functional is denoted as

\[
TV(x) = \sum_{i=1}^{m} \sum_{j=1}^{n} \|Dx\|_1
\]

where $D$ represents the stacking of the difference approximations. The TV image reconstruction problem is typically written as

\[
\hat{x} = \arg\min_x \frac{\lambda}{2} \|d - Ax\|_2^2 + TV(x). \tag{6}
\]
For TV regularization, it is possible to do a $\chi^2$ test on the data misfit, as in (4), to find $\lambda$ using the Discrepancy Principle. This approach is implemented in [5], [6], [24]. However, as just described in Section II, the degrees of freedom in the Discrepancy Principle $\chi^2$ test should be reduced by the number of parameters, due to their dependence on the data. Therefore we develop a statistical test for TV regularization using the regularized residual, rather than solely the data misfit, and this is similar to the $\chi^2$ [18] method for Tikhonov regularization. First, we explore results by Green in [9] which suggest the TV functional follows a Laplacian distribution, and this will be used to determine the appropriate statistical test.

A. TV Functional and Laplacian distribution

The statistical assumptions underlying the TV functional are developed in [9] by Green by first introducing the joint distribution of pixel intensities in naturally occurring images. He shows analytically, and through histograms, that differences in intensities between adjacent pixels tends to a Laplacian distribution. This verified previous observations in [13] that pixel intensities in naturally occurring images are differentially Laplacian. Green states that this a good first approximation to reality, even though the joint Laplacian density function model requires the assumption that the elements in the vector are independent. Since the TV functional represents a discretization of linear differential operator, this statistical model was then extended to TV regularization. An important consideration in modeling the TV functional as differentially Laplacian, is the assumption that coefficients sum to 0, which is the case with pixel intensities.

One of the key assumptions is that the initial parameter estimate is independent of the data. Therefore we consider the data fit and TV regularized term separately. For the TV term let $y_i = x_{i+1} - x_i$, and if it follows a Laplacian distribution the probability density function is

$$f(y_i | \bar{y}_i, \beta) = \frac{\beta}{2} e^{-\beta(y_i - \bar{y}_i)}$$

where $\bar{y}_i$ is the mean of $y_i$ and $\frac{\beta}{2}$ is the variance. In [9] they suggest that the mean estimate is zero and the standard deviation can be estimated by the pixel differences. For the data fit if we consider the data are estimates of the mean of $Ax$, and the errors follow a normal distribution, then the probability density function is

$$f((Ax)_i | b_i, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2}((Ax)_i - b_i)^2}$$

where $\sigma$ is the standard deviation of the data errors. TV regularization results if we maximize the probability that the data were observed given that the pixel differences follow a Laplacian distribution.
There is an analogous view of Tikhonov regularization, where both the data fit and regularization terms are assumed to follow a normal distribution. In Tikhonov regularization, the data errors are often not represented explicitly and the regularization parameter is viewed as a ratio of the standard deviation of errors in data to those in initial parameter estimates. This suggests that for TV, the regularization parameter can be chosen with a similar ratio. Using $\lambda$ defined by (6) the proposition in [9] suggests choosing $\lambda = 1/2\beta\sigma^2$, and the corresponding algorithm is given in Algorithm 1. Numerical results using this algorithm are given in Section IV.

**Algorithm 1** Regularization parameter selection algorithm for TV from Green [9].

1: input $d$, $\sigma$ (data std. dev.)
2: $D=\text{diff}(d)$;
3: $\beta = \sqrt{2/\text{std}(D)}$;
4: $\lambda = 1/(2\beta\sigma^2)$

**B. Statistical tests for regularization parameter**

The choice of regularization parameter suggested by Green in [9] uses the maximum likelihood estimate as the inverse estimate when the data fit is assumed normal and the TV functional Laplacian. Even if these are not correct assumptions, and the probability models are unknown, the following estimator converges to a well defined limit:

$$\hat{x} = \arg\min_x \frac{1}{2\sigma^2} \|d - Ax\|_2^2 + \beta \text{TV}(x - x_p)$$

with $\beta = \sqrt{2}/\sigma_{TV}$, and $\sigma_{TV}$ representing the standard deviation of the adjacent pixel differences. This estimator is a first approximation to reality, as long as the data $d$ and initial estimate $x_p$ are independent. In addition, it is intuitive that the $L_1$ norm should be weighted by a proportion of the standard deviation, just as the $L_2$ norm has a proportional variance weight in Tikhonov regularization.

In image processing applications it is possible to estimate the standard deviation of adjacent pixel differences using the data, and reasonable results are shown in Section IV with Algorithm 1. When reliable statistical information is not available, we suggest estimating it by using a statistical test analogous to the $\chi^2$ test in (5) for Tikhonov regularization. That is, given good estimates of the standard deviation of data error $\sigma$, the standard deviation from the Laplacian distribution for the TV function, $\sigma_{TV}$, can be estimated by solving

$$\frac{1}{2\sigma^2} \|d - A\hat{x}\|_2^2 + \frac{\sqrt{2}}{\sigma_{TV}} \text{TV}(\hat{x} - x_p) \approx m^2,$$  \quad (7)
from which a value for the regularization parameter $\lambda$ can be inferred. Note that inferring $\lambda$ from this equation requires an estimate of $\hat{x}$.

For Tikhonov regularization, there is an analytical expression for the optimal value $\hat{x}$, and thus results do not depend on the optimization algorithm. The challenge with TV regularization is that it is computationally complex to estimate $\hat{x}$, and a good summary of algorithms is found in [5]. A similar approach to choosing the regularization parameter using the Discrepancy Principle with TV regularization was recently implemented in [24]. The main difference between that approach and the one described here is that we use a numerical estimate $\hat{x}$ rather than an approximate closed form solution of it to ensure the appropriate Principle, or test, is satisfied.

In [24] they use the following Discrepancy Principle

$$\frac{1}{\sigma^2} \|d - A\hat{x}(\lambda)\|_2^2 = DF$$

where DF is the effective degrees of freedom. They remark that an analytical derivation of DF is an open problem and suggest an estimation procedure for it. In the $\chi^2$ method, DF is taken to be $m^2 - n^2$ since $\hat{x}$ depends on the data [1]. However, the open question now is not the degrees of freedom, but rather how to weight residuals with uncertainty estimates based on standard deviations of errors. Using one standard deviation implies that the estimate lies within the confidence interval 68.2% of the time, while using two standard deviations will be 95.4%. In addition, since the solution estimate depends on the regularization parameter we write (7) as

$$\frac{1}{2(\tau\sigma)^2} \|d - A\hat{x}(\lambda)\|_2^2 + \frac{\sqrt{2}}{\sigma_{TV}}\text{TV}(\hat{x} - x_p) \approx m^2,$$

with $\tau$ chosen to be 1, 2 or 3. Note that for TV regularization as written in (6) we have $\lambda = \sigma_{TV}/(\sqrt{2}\tau\sigma^2)$. Therefore, we find a $\lambda$ that is the root of the function

$$f(\lambda) = \frac{\|d - A\hat{x}(\lambda)\|_2^2}{2(\tau\sigma)^2} + \frac{\text{TV}(\hat{x}(\lambda) - x_p)}{\lambda(\tau\sigma)^2} - m^2\lambda(\tau\sigma)^2.$$ (8)

An algorithm for this method is outlined in Algorithm 2 where a Newton type method is used to solve $f(\lambda) = 0$. The function $f(\lambda)$ depends on $\hat{x}$ and therefore a TV algorithm must be employed within each Newton iterate.

**Theorem 1:** Define

$$r_1(\lambda) = \|d - A\hat{x}(\lambda)\|_2^2$$

$$r_2(\lambda) = \text{TV}(\hat{x}(\lambda) - x_p).$$
If \( r_1 \) and \( r_2 \) are Lipschitz continuous, with corresponding Lipschitz constants
\[
K_1 = c_1 m^2 (\tau \sigma)^4 \\
K_2 = d_1 m^2 (\tau \sigma)^4 \lambda^0
\]
where \( \lambda \geq \lambda^0 \) and \( c_1 + 2d_1 \leq 2 \), then (8) is monotonically decreasing and any positive root is unique.

**Proof 1:** To see that \( f(\lambda) \) is monotonically decreasing
\[
\begin{align*}
\frac{df}{d\lambda} &= \frac{r'_1(\lambda)}{2(\tau \sigma)^2} + \frac{r'_2(\lambda)}{\lambda (\tau \sigma)^2} - \frac{r_2(\lambda)}{\lambda^2 (\tau \sigma)^2} - m^2 (\tau \sigma)^2 \\
&< \frac{c_1 m^2 (\tau \sigma)^4}{2(\tau \sigma)^2} + d_1 m^2 (\tau \sigma)^4 \lambda^0 - \frac{r_2(\lambda)}{\lambda^2 (\tau \sigma)^2} - m^2 (\tau \sigma)^2 \\
&< (\frac{c_1}{2} + d_1 - 1)m^2 \sigma^2 - \frac{r_2(\lambda)}{\lambda^2 \sigma^2}.
\end{align*}
\]
Thus, given the Lipschitz constants, a convergent TV algorithm for \( \hat{x} \), with \( \lambda > 0 \) and \( \sigma \neq 0 \), \( f(\lambda) \) is continuous and monotonically decreasing. There are three possibilities: \( f(\lambda) > 0 \), \( f(\lambda) < 0 \) or \( f(\lambda) = 0 \). In the first instance there is no solution because for increasing \( \lambda \) the second term decreases to zero and so in the limit we have
\[
\frac{\|d - A\hat{x}(\lambda)\|_2^2}{2(\tau \sigma)^2} > m^2 \lambda(\tau \sigma)^2.
\]
In this case, a larger value of \( \sigma \) is needed for there to exist a root, which means the data are not as accurate as indicated by the standard deviation estimate \( \sigma \). In the second case a smaller value of \( \sigma \) is needed for there to exist a root, meaning the data are more accurate than anticipated. In the third case there exists a solution and it is unique. If the Lipschitz assumption is violated, it is possible to increase \( \sigma \) and give less weight to the data. ■

**Algorithm 2** Regularization parameter selection algorithm for TV using the statistical test.

1: input \( d, A, \sigma, \tau, \lambda^0, \) and \( x_p \)
2: Use nonlinear root finding to solve \( f(\lambda) = 0 \) in (8) with
\[
\hat{x}(\lambda) = \arg\min_x \frac{1}{2}\|d - Ax\|_2^2 + TV(x);
\]
3: If not converged with \( f(\lambda) > 0 \), increase \( \sigma \)
4: If not converged with \( f(\lambda) < 0 \), decrease \( \sigma \)
5: End root finding

The Discrepancy Principle can be implemented in a manner similar to the algorithm in Algorithm 2. The optimal value of \( \lambda \) in this case is the root of
\[
g(\lambda) = \frac{\|d - A\hat{x}(\lambda)\|_2^2}{\sigma^2} - m^2
\]
(9)
where $\hat{x}$ is found by solving (6). However, it may be the case that no such root exists. Therefore, $\hat{x}$ is found for many values of $\lambda$, and the one that comes closest to satisfying $g(\lambda) = 0$ is chosen. This is similar to Occam’s inversion for Tikhonov regularization [1]. The corresponding algorithm is given in Algorithm 3.

In the Section IV we give numerical results comparing the accuracy of results when the regularization parameter $\lambda$ is chosen according to Green’s results in [9], the Discrepancy Principle and the statistical test proposed here using the regularized residual, i.e. Algorithms 1 - 3.

**Algorithm 3** Regularization parameter selection algorithm for TV using Discrepancy Principle.

1: input $d$, $A$, $\sigma$, $\lambda^{\text{min}}$, $\lambda^{\text{max}}$
2: For $\lambda = \lambda^{\text{min}} \ldots \lambda^{\text{max}}$
   \hspace{1cm} $\hat{x}(\lambda) = \arg\min_x \frac{1}{2}\|d - Ax\|_2^2 + TV(x)$;
3: Calculate $g(\lambda)$ in (9)
4: End for
5: $\lambda = \arg\min_\lambda g(\lambda)$

![Original images](image1.png)

**Fig. 1.** Original images

**IV. NUMERICAL RESULTS**

The goal of this work is to investigate methods for choosing the regularization parameter in TV regularization that can be used with any TV optimization algorithm. We chose to test Algorithm 1-3 on total variation image deblurring using the `tvreg` package in Matlab, authored by Pascal Getreuer. It uses the split Bregman method [10] and is very efficient so it is possible to execute the problem many
times with different values of $\lambda$ to determine the one that creates the best image. In this package the parameter $\lambda$ is viewed as balancing deblurring accuracy and noise reduction, where a smaller values of $\lambda$ imply stronger noise reduction, but at the cost of deblurring accuracy.

As an alternative to running the problem with multiple $\lambda$ to find the image that looks best, here we test Algorithms 1-3 which each compute a single value of $\lambda$ based on the statistics of the noise model and total variational functional. The images on which we chose to test the Algorithms are gray-scale with intensity values normalized to the range $[0, 1]$. They are the frequently used test images, Einstein, cameraman and MRI, all with $256 \times 256$ resolution and plotted in Figure 1. The images were artificially blurred with a discrete linear blur filter $A$ that is a convolution of the image with a point spread function. We blurred them with Gaussian and uniform lowpass filters with variance 5 an 9, and a filter to approximate the linear motion of a camera with horizontal motion of 9 pixels. These are done with the MATLAB commands `fspecial('Gaussian',[9,9],3) `fspecial('average',9)` for variance 9, and `fspecial('motion',9,0)` to simulate linear motion.

In each case the noise was quantified using the Blurred Signal to Noise Ratio (BSNR):
\[
\text{BSNR} = 20 \log_{10} \frac{\|Ax\|_2}{m\sigma}.
\]

We compared BSNRs of 20, 30 and 40 dB, where the lower levels represent more noise. The effect of these noise levels on the cameraman image for Gaussian noise and motion is illustrated in Figures 2-4. The images for uniform noise are similar to those for Gaussian. Algorithm performance was measured in terms of the frequently used Improvement in Signal-to-Noise Ratio (ISNR) metric:
\[
\text{ISNR} = 20 \log_{10} \frac{\|d - \hat{x}\|_2}{\|\hat{x} - x\|_2}.
\]

The ISNR is a ratio of the restored image over the corrupted image and is an estimate of the perceptual quality of an image with larger values indicating better quality.

In Tables I-III we give resulting ISNR for Algorithms 1-3. Since different realizations of noise give slightly different ISNR values, we simulated the problem 100 times and reported the average. Algorithm 1, as suggested by Green [9], relies on the Laplace distribution parameter $\beta$. The distribution parameter is estimated using the difference in pixel values of the blurred image. In all examples, this algorithm gave the lowest ISNR values.

Algorithm 2 requires more information about the problem, in particular the point spread function $A$, $\tau$, and initial estimates of $\lambda$ and the image. We ran tests with one and two standard deviations, i.e. $\tau = 1, 2$, and used the corrupted image for $x_p$. The value of $\hat{x}$ used to find the root of (8) was output from the
TV algorithm tvreg, and it was called within the nonlinear root finding algorithm fzero as it searched for \( \lambda \). There was no need to adjust the data error standard deviation \( \sigma \) in these examples (as indicated in Algorithm 2) because a root was always found.

Algorithm 3 requires similar information as Algorithm 2. However, rather than an initial estimate of \( \lambda \), a range of \( \lambda \) is given over which to find the smallest value of the cost function \( g(\lambda) \). Similar to Algorithm 2, the value of \( \hat{x} \) used to find the root of (9) was output from the TV algorithm tvreg. Algorithm 3 took the longest computational time due to the slow search for the root.

In Table I we give ISNR values for Algorithms 1-3 after convolving the image with a Gaussian function with a variance of 5. This is the clearest blurred image of all of the tests. For the high noise level, BSNR=20, the statistical test with \( \tau = 1 \) and the discrepancy principle have the best ISNR values.
(a) Gaussian blur $\sigma = 5$

(b) Gaussian blur $\sigma = 9$

(c) Motion blur

Fig. 4. Blurred with BSNR= 40

### TABLE I

**AVERAGE ISNR OVER 100 SIMULATIONS, GAUSSIAN BLUR WITH VARIANCE 5**

<table>
<thead>
<tr>
<th>BSNR</th>
<th>Green [9] Discrepancy</th>
<th>Stat Test $\tau = 1$</th>
<th>Stat Test $\tau = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Einstein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.97       2.82       2.72       1.76</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>1.55       2.86       2.07       2.59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>3.71       4.81       3.71       4.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cameraman</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>1.90       3.21       3.37       1.94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>1.12       2.43       3.14       1.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>2.42       4.77       4.86       4.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MRI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.04       2.20       2.35       1.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>1.79       3.23       2.60       3.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>4.72       5.72       4.95       6.19</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For the lower noise levels, in particular BSNR=40, the statistical test with $\tau = 2$ and the method suggested by Green in [9] have competitive values of ISNR.

The ISNR values in Table II show similar behavior to those in Table I. The ISNR values resulting from the Gaussian blur with variance 9 were highest with the statistical test, $\tau = 1$, except for the low noise level BSNR=40. This is also typical of the $\chi^2$ method [18], which is the least squares regularization version of the statistical test with $\tau = 1$ for TV regularization. That is, the method performs better with
TABLE II
AVERAGE ISNR OVER 100 SIMULATIONS, GAUSSIAN BLUR WITH VARIANCE 9

<table>
<thead>
<tr>
<th>BSNR</th>
<th>Green [9]</th>
<th>Discrepancy</th>
<th>Stat Test $\tau = 1$</th>
<th>Stat Test $\tau = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Einstein</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.60</td>
<td>2.27</td>
<td>2.67</td>
<td>1.25</td>
</tr>
<tr>
<td>30</td>
<td>1.13</td>
<td>2.59</td>
<td>2.89</td>
<td>2.06</td>
</tr>
<tr>
<td>40</td>
<td>2.50</td>
<td>4.58</td>
<td>2.91</td>
<td>3.83</td>
</tr>
<tr>
<td>Camerman</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>1.72</td>
<td>2.73</td>
<td>3.03</td>
<td>1.73</td>
</tr>
<tr>
<td>30</td>
<td>1.00</td>
<td>2.13</td>
<td>2.56</td>
<td>1.59</td>
</tr>
<tr>
<td>40</td>
<td>1.79</td>
<td>3.44</td>
<td>4.01</td>
<td>2.75</td>
</tr>
<tr>
<td>MRI</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.29</td>
<td>1.90</td>
<td>2.51</td>
<td>1.05</td>
</tr>
<tr>
<td>30</td>
<td>1.15</td>
<td>2.95</td>
<td>3.73</td>
<td>2.23</td>
</tr>
<tr>
<td>40</td>
<td>2.80</td>
<td>5.22</td>
<td>4.74</td>
<td>4.92</td>
</tr>
</tbody>
</table>

higher noise levels. It is interesting to note that when the statistical test with $\tau = 1$ has lower ISNR than discrepancy, the better choice is $\tau = 2$. This is true with the Einstein and MRI images for BSNR=40, while $\tau = 1$ gives the best ISNR for Camerman.

The motion blur case in Table III resulted in the highest ISNR values over I-III. The trend of algorithm performance is similar to that in I-II where the statistical test with $\tau = 1$ and the discrepancy principle give the highest ISNR values, with the $\tau = 1$ statistical test preferable for the highest noise level 20. In addition, the discrepancy principle has the highest ISNR overall for the Einstein image. This may be due to the fact that the original image is not as sharp as the other two images.

In Table IV we compare our results to other methods that calculate the regularization parameter rather than tune it to get the best images. We also show results from Algorithms 1-3 with uniform noise, variance 9. We do not show all of the results when using uniform noise with variances 5 and 9 because they are very similar to those from Gaussian noise.

The method given in [24] also uses the discrepancy principle, similar to Algorithm 2. In [24] the regularization parameter is adjusted adaptively at each iteration to satisfy the discrepancy principle. We see in Table IV the ISNR values from Algorithm 2 are consistently lower than those resulting from methods in [24] and [4]. Therefore, we calculated the maximum ISNR that could be obtained using $tvreg$.
### TABLE III
**IAverage ISNR over 100 simulations, Motion blur**

<table>
<thead>
<tr>
<th>BSNR</th>
<th>Green [9] Discrepancy</th>
<th>Stat Test $\tau = 1$</th>
<th>Stat Test $\tau = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Einstein</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>1.49</td>
<td>3.37</td>
<td>3.15</td>
</tr>
<tr>
<td>30</td>
<td>3.20</td>
<td>4.44</td>
<td>4.17</td>
</tr>
<tr>
<td>40</td>
<td>6.39</td>
<td>7.49</td>
<td>6.35</td>
</tr>
<tr>
<td><strong>Cameraman</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>2.11</td>
<td>3.43</td>
<td>3.71</td>
</tr>
<tr>
<td>30</td>
<td>2.44</td>
<td>4.01</td>
<td>4.78</td>
</tr>
<tr>
<td>40</td>
<td>5.44</td>
<td>7.82</td>
<td>8.43</td>
</tr>
<tr>
<td><strong>MRI</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>1.43</td>
<td>3.57</td>
<td>4.02</td>
</tr>
<tr>
<td>30</td>
<td>4.65</td>
<td>6.70</td>
<td>6.34</td>
</tr>
<tr>
<td>40</td>
<td>10.85</td>
<td>11.23</td>
<td>10.75</td>
</tr>
</tbody>
</table>

and reported it in the last column. This value was determined by systematically substituting different values of the regularization parameter into the algorithm. This is the "tuning" of the regularization parameter that the algorithms in this work, and in [24] and [4] are designed to avoid.

The “tvreg max” value given in Table IV allows us to compare the ISNR value from regularization parameters chosen by Algorithms 2 and 3 with the best possible ISNR value that can be obtained. We see that the for BSNR equal to 20 or 30 the regularization parameter chosen with Algorithm 3 gives nearly the highest ISNR value possible. We also see that higher ISNR values necessarily result when the regularization parameter selection is integrated into the TV optimization algorithm. If this is not an option, the near optimal ISNR values with tvreg imply that Algorithms 2 and 3 are a good alternative for an efficient method of choosing the regularization parameter.

### V. SUMMARY AND CONCLUSIONS

Regularization parameter selection plays an important role in TV regularization. If the regularization parameter $\lambda$ is too small the image does not fit the data well, while if it is too large, the image is under smoothed. A suitable choice of $\lambda$ is often found by trying different values until one is found that results in a clear image. This is not the most efficient way of choosing the parameter, and may not be an option
TABLE IV
ISNR, Camerman

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian blur, σ = 9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>2.59</td>
<td>2.21</td>
<td>2.73</td>
</tr>
<tr>
<td>30</td>
<td>4.05</td>
<td>3.59</td>
<td>2.13</td>
</tr>
<tr>
<td>40</td>
<td>6.21</td>
<td>5.78</td>
<td>3.44</td>
</tr>
<tr>
<td>9 x 9 uniform blur</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>3.85</td>
<td>3.27</td>
<td>3.20</td>
</tr>
<tr>
<td>30</td>
<td>5.86</td>
<td>5.69</td>
<td>3.45</td>
</tr>
<tr>
<td>40</td>
<td>8.46</td>
<td>8.46</td>
<td>5.62</td>
</tr>
</tbody>
</table>

if the linear TV optimization problem is part of a greater iteration scheme, such as that for nonlinear problems.

Successful, automated methods for choosing the regularization parameter in TV regularization can be found in [5], [6], [24]. These methods find regularization parameters that satisfy the discrepancy principle. When statistical information on the measurement errors is available, the discrepancy principle takes on significance as a $\chi^2$ test on the data fit. However, the degrees of freedom in the discrepancy principle is taken to be the number of data, which is not correct. The degrees of freedom should be reduced by the number of parameters, since the optimal estimate depends on the data [1]. Alternatively, the $\chi^2$ method [17] is a discrepancy principle applied to the regularized residual in Tikhonov regularization, where the degrees of freedom is equal to the number of data. Here we extend the $\chi^2$ method to TV regularization. There is no appropriate $\chi^2$ test for TV regularization since the TV functional is not an $L^2$ norm of the parameter vector. Therefore, we developed the approach more generally as a statistical test method in Algorithm 2. The statistics of the TV regularized residual were taken from results in [9] that state the TV function follows a Laplace distribution.

In this work we examined not just the statistical test method for finding a regularization parameter in TV regularization, but in addition the algorithm given in [9] (Algorithm 1) and an approach to the discrepancy principle that can be used with any TV optimization algorithm (Algorithm 2). Experiments were run with three test images, three different levels of noise, and three types of noise.

We found that estimates of the Laplace distribution parameter $\beta$ given in Algorithm 1 give regularization
parameters that result in images with the lowest ISNR values of all methods tested. The new approach developed here, Algorithm 2, gives regularization parameters that consistently result in images with the highest ISNR values when the original image contains the most noise (BSNR=20). Algorithm 3, the discrepancy principle, consistently resulted in the highest ISRN when the image contains the least noise (BSNR=40). This observation that when using the regularized residual to find a regularization parameter (as is the case with Algorithm 2 and the $\chi^2$ method) works best for more noisy problems has also been observed with Tikhonov regularization.

We also developed the statistical test method as a function of the parameter $\tau$, representing the number of standard deviations in the confidence interval that the statistical test must satisfy. For high noise images, a value of $\tau = 1$ is the preferred choice, while the less noisy images have a higher ISNR with $\tau = 2$. This added parameter makes the statistical test more competitive with the discrepancy principle for low noise images, and we therefore suggest Algorithm 2 is the best choice of the three algorithms.

The algorithms were all used as a supplement to $tvreg$, a Matlab package for Total Variation deconvolution that requires the regularization parameter as input. The statistical test approach consistently found images with the highest ISNR values possible with this algorithm. For low noise images, the highest ISNR values possible with $tvreg$ were not competitive with those found in [4], [24]. This suggests that Algorithms 2-3 are best used within a TV optimization algorithm, rather than as a supplement. However, we still do recommend using Algorithm 2 if the TV optimization algorithm does not have a choice of $\lambda$ built into it, especially for high noise images. In addition, we suggest implementing the statistical test using the regularized residual in Algorithm 2 rather than the discrepancy principle which does not have the correct degrees of freedom.

VI. ACKNOWLEDGMENTS

I would like to thank the authors of $tvreg$ for making their source codes freely available. In addition, I am thankful to Prof. Renaut and Anna Nelson for stimulating discussions that led to this work.
REFERENCES
