A new principle for choosing regularization parameter in certain inverse problems

Hans Rullgård
April 15, 2008

Abstract
A new parameter choice rule for inverse problems is introduced. This parameter choice rule was developed for total variation regularization in electron tomography and might in general be useful for $L^1$ regularization of inverse problems with high levels of noise in the data.

1 Introduction

A standard procedure when solving an ill-posed inverse problem $Tf = g$, where $T : X \rightarrow Y$ is an operator which lacks a continuous inverse, is to define a class of continuous operators $S_\lambda : Y \rightarrow X$ depending on a parameter $\lambda$ and approximating an inverse of $T$ as $\lambda \rightarrow 0$. An approximate solution of the inverse problem is then given by $S_\lambda(g)$ where the regularization parameter $\lambda$ must be selected by some parameter choice rule. Numerous parameter choice rules have been proposed, which are suitable for different inverse problems. For a nice survey, see [2, Chapter 4].

The parameter choice rule proposed in this paper is inspired by the application of $L^1$-type regularization methods, particularly total variation (TV) regularization, to the inverse problem in electron tomography (ET). For a survey of total variation image restoration methods, see [1], and for a comprehensive account of the mathematics in ET, see [3]. It turns out that the well-known parameter choice rules are difficult to apply to this problem. One reason for this is that the regularized inverses $S_\lambda$ are non-linear even if the forward operator $T$ is linear. A second reason is that this inverse problem has unusually high noise level in the data, the norm of the noise component by far exceeding the norm of the signal. For this reason, in order to recover anything at all, it is necessary to make use of knowledge about the statistical properties of the noise, not just its magnitude.

These issues are taken into account by the proposed parameter choice rule. It is specifically designed for regularization functionals of $L^1$ type, and for inverse problems aiming at reconstructing sparse objects. As it is applied here to ET data, the method is a heuristic, or error free, parameter choice rule in the terminology of [2]. By this one understands a parameter choice rule which does not require an explicit estimate of the noise level to be made. Instead, the noise level is estimated directly from the data, relying on certain assumptions about the nature of the noise.

It turns out that the method developed for ET can be broken down in several components, which can probably be applied independently in various other inverse problems. Here an attempt is made to present these components independently. The paper is organized as follows. In Section 2, the most general part of the method, requiring the least structure of the inverse problem, is described. In Section 3, the framework is applied to total variation regularization. Finally, in Section 4, the application to ET...
is developed, and numerical examples are presented. A general discussion is given in Section 5.

2 Basic principle

Consider the following inverse problem. Let $T : X \rightarrow Y$ be a linear operator between two linear spaces, with $Y$ a Hilbert space. Let $f^{true} \in X$ be an unknown element, which it is our goal to estimate. What is known is an element $g^{data} = T f^{true} + g^{noise} \in Y$ where $g^{noise}$ is a sample of a random vector $G^{noise}$. Let us assume that $E[G^{noise}] = 0$. However, the probability distribution of $G^{noise}$ may not be completely known, and might to some extent depend on $f^{true}$.

As a regularized inverse of $T$ we consider a class of (non-linear) operators $S_\lambda : Y \rightarrow X$ depending on a regularization parameter $\lambda$ and defined by

$$S_\lambda(g) := \arg \min_{f \in X} R_\lambda(f) + \frac{1}{2} \|Tf - g\|^2$$

where $R_\lambda : X \rightarrow \mathbb{R}$ is a regularization functional parameterized by $\lambda$. The idea is that the reconstruction $f^{rec} = S_\lambda(g^{data})$ might be a good approximation to $f^{true}$ for suitable choice of the regularization parameter.

Let us assume the existence of a unique minimizer of the optimization problem in (1). Let us also assume that $R_\lambda$ is convex, and that

$$R_\lambda(\alpha f) = |\alpha| R_\lambda(f), \quad \forall \alpha \in \mathbb{R}. \quad (2)$$

The problem considered here is how to choose the regularization parameter $\lambda$. Various methods for choosing regularization parameters are of course known. The method proposed here is motivated by observing the solution of optimization problems similar to (1), but where $f$ is restricted to vary along a line in $X$. Explicitly, if $f \in X$ and $g \in Y$, define

$$\alpha_\lambda(f, g) := \arg \min_{\alpha \in \mathbb{R}} R_\lambda(\alpha f) + \frac{1}{2} \|T(\alpha f) - g\|^2. \quad (3)$$

Contrary to (1), the solution of (3) can be computed explicitly.

**Lemma 1.** Let $f \in X$ and $g \in Y$, and suppose that $R_\lambda$ satisfies (2). If $Tf \neq 0$ or $R_\lambda(f) > 0$ then (3) has a unique solution given by

$$\alpha_\lambda(f, g) = \begin{cases} \frac{\langle Tf, g \rangle - R_\lambda(f)}{\|Tf\|^2}, & \langle Tf, g \rangle > R_\lambda(f) \\ 0, & \|\langle Tf, g \rangle\| \leq R_\lambda(f) \\ \frac{\langle Tf, g \rangle + R_\lambda(f)}{\|Tf\|^2}, & \langle Tf, g \rangle < -R_\lambda(f). \end{cases}$$

If on the other hand $Tf = 0$ and $R_\lambda(f) = 0$, the solution is not unique.

The proof is a straightforward computation.

Now we want to look at the random variable $\alpha_\lambda(f, G^{noise})$. The intuitive idea is that if $\alpha_\lambda(f, G^{noise})$ is close to 0 with high probability, this is an indication that $S_\lambda(g^{data})$ is not heavily influenced by noise. This intuition may or may not be correct, depending on the nature of the regularization functional and the forward operator. The precise conditions needed for the idea to be valid are not yet clear.
Let us for any \( f \in X \) define
\[
\sigma(f) := \text{Var}[(Tf, G_{\text{noise}})]^{1/2}
\] (5)
and
\[
s_{\lambda}(f) := \frac{R_{\lambda}(f)}{\sigma(f)}.
\] (6)

Lemma 4 indicates that if \( s_{\lambda}(f) \gg 1 \), then \( \alpha_{\lambda}(f, G_{\text{noise}}) = 0 \) with high probability. On the other hand, if \( s_{\lambda}(f) \ll 1 \), then \( \alpha_{\lambda}(f, G_{\text{noise}}) \) is not strongly affected by the regularization functional.

The conclusion is that the values of \( s_{\lambda}(f) \) for different \( f \in X \) might serve as quantitative measures of the strength of the regularization as compared to the noise level. This leads us up to the formulation of the basic form of the proposed parameter choice rule:

1. Choose a finite set \( F \) of elements in \( X \).
2. For each \( f \in F \), choose \( s_{\min}(f) \in \mathbb{R} \). The choice of \( s_{\min}(f) \) determines how strongly the inverse problem is regularized. As a rule of thumb, \( s_{\min}(f) \geq 5 \) corresponds to very strong regularization, while \( s_{\min}(f) \leq 1 \) corresponds to weak regularization.
3. Choose the smallest regularization parameter \( \lambda \) such that \( s_{\lambda}(f) \geq s_{\min}(f) \) for all \( f \in F \).

In Section 3 a suitable choice of the set \( F \) and a model for computing \( s_{\min}(f) \) is given for the special case of TV regularization. In Section 4 a method for estimating \( \sigma(f) \) for ET data is provided. With these ingredients we will then be ready to apply the parameter choice rule in TV regularized ET.

### 3 Application to TV regularization

Let us specialize the setting as follows. Let \( \Omega \subset \mathbb{R}^n \) be a bounded open set and let \( X \) be the space of functions of bounded variation with support in \( \Omega \), with the total variation norm
\[
\|f\|_{TV} := \sup \left\{ \int_{\mathbb{R}^n} f\nabla \cdot h \, dx : h \in C^1_0(\mathbb{R}^n, \mathbb{R}^n), \ |h(x)| \leq 1 \right\}
\] (7)
where \( C^1_0(\mathbb{R}^n, \mathbb{R}^n) \) denotes the space of continuously differentiable functions from \( \mathbb{R}^n \) to \( \mathbb{R}^n \) with compact support and \( \nabla \cdot h \) is the divergence of \( h \). For continuously differentiable functions \( f \) (among others) the total variation is given by
\[
\|f\|_{TV} = \int_{\mathbb{R}^n} |\nabla f(x)| \, dx.
\] (8)
However, the space \( X \) contains many non-differentiable functions, including for example the characteristic functions of certain sets, known as Caccioppoli sets. Caccioppoli sets include all sets with \( C^2 \) boundary, see [4] for details.

In TV regularization, the regularization functional is chosen to be \( R_{\lambda}(f) := \lambda \|f\|_{TV} \). Also, let us assume that \( \sigma(f) \) is translation invariant, meaning that if \( f_1, f_2 \in X \) are related by \( f_1(x) = f_2(x - x_0) \) for some \( x_0 \in \mathbb{R}^n \), then \( \sigma(f_1) = \sigma(f_2) \). Since \( \|f_1\|_{TV} = \|f_2\|_{TV} \), it then follows from the definition of \( s_{\lambda} \) that \( s_{\lambda}(f_1) = s_{\lambda}(f_2) \).
3.1 Choice of the set $F$

Here we must choose a finite subset $F \subset X$ which is somehow representative of all functions in $X$. Lemma 2 below suggests that it is reasonable to restrict attention to characteristic functions of Caccioppoli subsets of $\Omega$.

Suppose that $E \subset \Omega$ is a Caccioppoli set, and that $E = E_1 \cup E_2$ with $E_1 \cap E_2 = \emptyset$. Let $f_1$ and $f_2$ be the characteristic functions of $E_1$ and $E_2$, so $f := f_1 + f_2$ is the characteristic function of $E$. Then $\sigma(f) \leq \sigma(f_1) + \sigma(f_2)$ and $\|f\|_{TV} = \|f_1\|_{TV} + \|f_2\|_{TV}$ and it follows that

$$s_\lambda(f) \geq \frac{\lambda \|f_1\|_{TV} + \|f_2\|_{TV}}{\sigma(f_1) + \sigma(f_2)} = \frac{s_\lambda(f_1)\sigma(f_1) + s_\lambda(f_2)\sigma(f_2)}{\sigma(f_1) + \sigma(f_2)} \geq \min\{s_\lambda(f_1), s_\lambda(f_2)\}.$$ 

Hence, if $f_1$ and $f_2$ are both included in $F$, there is no need to include $f$ unless $s_{\min}(f) > \max\{s_{\min}(f_1), s_{\min}(f_2)\}$. From this observation it would be tempting to conclude that only characteristic functions of connected Caccioppoli sets need to be included in $F$. However, this is not strictly true, since there are very complicated Caccioppoli sets, which for example can have uncountably many connected components. Nevertheless, of all functions that are likely to be treated numerically in practice, only characteristic functions of connected sets need to be included in $F$. By similar reasoning, if $\Omega$ has connected complement, only characteristic functions of sets with connected complement need to be included in $F$.

I further suggest that in many cases it should be reasonable to choose $F$ as a set of characteristic functions of balls of different sizes. The precise arguments for this and the conditions under which they are valid remain to be clarified.

Suppose we make the choice to let $F$ consist of characteristic functions of balls. We would then choose a finite set $D = \{d_1, \ldots, d_k\}$ of diameters of these balls. By the assumption of translation invariance, it is sufficient to include one ball of each diameter in $F$, so we have $F = \{f_d : d \in D\}$ where $f_d$ denotes the characteristic function of an arbitrary ball of diameter $d$.

To conclude this section, we state and prove the lemma which was used above to motivate the restriction to characteristic functions of Caccioppoli sets.

**Lemma 2.** Suppose there exists a constant $C$ such that $\sigma(f) \leq C\|f\|_{TV}$ and $\sigma(f) \leq C\|f\|_{L^\infty}$ for all $f \in X$. If $s_\lambda(f) \geq s_0 > 0$ for every $f \in X$ which is the characteristic function of a Caccioppoli subset of $\Omega$, then the same inequality holds for every $f \in X$.

**Remark 1.** From the hypothesis of the lemma it is trivially true that $s_\lambda(f) \geq \lambda/C$. However, the constant $C$ could a priori be very large. The point of the lemma is that if a better estimate holds for characteristic functions of sets, then the same estimate necessarily holds for all functions of bounded variation.

**Proof.** The idea of the proof is simply that an arbitrary function is a superposition of characteristic functions defined by its level sets.

Note first that it is sufficient to prove the statement for positive functions. For assuming that this has been done, an arbitrary function can be written as $f = f_+ - f_-$ where $f_+$ and $f_-$ are positive and $\|f\|_{TV} = \|f_+\|_{TV} + \|f_-\|_{TV}$. Since $\sigma(f) \leq \sigma(f_+) + \sigma(f_-)$ it follows that

$$s_\lambda(f) \geq \frac{\lambda \|f_+\|_{TV} + \|f_-\|_{TV}}{\sigma(f_+) + \sigma(f_-)} = \frac{s_\lambda(f_+)\sigma(f_+) + s_\lambda(f_-)\sigma(f_-)}{\sigma(f_+) + \sigma(f_-)} \geq s_0.$$

So suppose that $f$ is a positive function. For $t > 0$ define $\chi_t$ to be the characteristic function of the set $\{x : f(x) \geq t\}$. By the coarea formula \cite[Theorem 1.23]{4}, it holds that

$$\|f\|_{TV} = \int_0^\infty \|\chi_t\|_{TV} dt.$$ 

(9)
If it can be shown that

$$\sigma(f) \leq \int_0^\infty \sigma(\chi_t) \, dt$$

(10)

the conclusion follows, since then

$$s_\lambda(f) \geq \int_0^\infty s_\lambda(\chi_t) \, dt = \int_0^\infty \sigma(\chi_t) \, dt \geq s_0.$$  

Let \(0 = t_0 < t_1 < t_2 < \cdots < t_m\) be a sequence of positive numbers, and define \(\delta_i = t_i - t_{i-1}\). For each \(i = 1, \ldots, m\) there is some \(\tau_i \in [t_{i-1}, t_i]\) such that

$$\delta_i \sigma(\chi_{\tau_i}) \leq \int_{t_{i-1}}^{t_i} \sigma(\chi_t) \, dt.$$  

If

$$f_1 = \sum_{i=1}^m \delta_i \chi_{\tau_i}$$

then it holds that

$$\sigma(f_1) \leq \int_{t_{i-1}}^{t_i} \sigma(\chi_t) \, dt.$$  

If we define \(f_2(x) = \max\{0, f(x) - t_m\}\), it follows from the coarea formula and the assumptions of the lemma that

$$\sigma(f_2) \leq C \|f\|_{TV} = C \int_0^\infty \|\chi_t\|_{TV} \, dt.$$  

Finally, we have that

$$\sigma(f - f_1 - f_2) \leq C \|f - f_1 - f_2\|_{L^\infty} \leq C \max_{1 \leq i \leq m} \delta_i.$$  

This shows that \(\sigma(f_2)\) and \(\sigma(f - f_1 - f_2)\) can be made arbitrarily small by making \(m\) and \(t_m\) large. Since \(\sigma(f) \leq \sigma(f_1) + \sigma(f_2) + \sigma(f - f_1 - f_2)\), (10) follows from (11) and this completes the proof.

### 3.2 Model for computing \(s_{\min}(f)\)

Here I will provide a model for computing \(s_{\min}(f)\). Let \(f_d \in F\) be a function whose support is a ball of diameter \(d\), for example a characteristic function as suggested in the previous section. Let us assume that we are given a real number \(a \geq 0\) and want to avoid that \(|f_{\text{rec}}(x)| > a\) in regions where \(f_{\text{true}}(x) = 0\). Let us also assume that the probability distribution of \(\langle Tf, G^{\text{noise}} \rangle\) is to a good approximation Gaussian.

Consider balls of diameter \(d\) in \(\Omega\). The maximum number \(N_d\) of such disjoint balls in \(\Omega\) is approximately \(N_d \approx |\Omega|d^{-n}\). Let \(f_{d_1}^1, \ldots, f_{d_1}^{N_d}\) be translations of \(f_d\) with support in these disjoint balls. A heuristic argument suggests that we should look at the probability that \(|\alpha_{f_d} G_{d}^{\text{noise}}| \cdot \|f_d\|_{L^\infty} > a\). By Lemma 1, that probability is Gaussian (assuming that \(\langle Tf_d, G_{d}^{\text{noise}} \rangle\) is Gaussian)

$$\text{erfc} \left( \frac{1}{\sqrt{2}} \left( s_\lambda(f_d) + \frac{a\|Tf_d\|_{L^\infty}}{\sigma(f_d)\|f_d\|_{L^\infty}} \right) \right).$$  

(12)
Let us, rather arbitrarily, choose the regularization parameter so that the expected number of $j \in \{1, \ldots, N_d\}$ with $|\alpha_\lambda(f^d_j, G_{\text{noise}})| > a$ is not more than 1. This is equivalent to the inequality

$$s_\lambda(f^d) \geq s_{\min}(f^d) = \sqrt{2} \operatorname{erfc}^{-1}\left(\frac{1}{N_d}\right) - \frac{a\|Tf^d\|_2}{\sigma(f^d)\|f^d\|_{L_\infty}},$$

$$\approx \sqrt{2} \operatorname{erfc}^{-1}\left(\frac{d^n}{|\Omega|}\right) - \frac{a\|Tf^d\|_2}{\sigma(f^d)\|f^d\|_{L_\infty}}.$$  \hspace{1cm} (13)

The formula is easily modified if some other restriction on the expected number of $j$ with $|\alpha_\lambda(f^d_j, G_{\text{noise}})| > a$ is desired.

4 A numerical example from electron tomography

In this section I will apply the proposed parameter choice method to a numerical example from electron tomography (ET). First I give a brief description of this inverse problem. For details, see for example [3].

Electron tomography is a method using a transmission electron microscope to construct three-dimensional models of biological macromolecules and similar structures. The data collected consists of a series of images, a tilt series, with the specimen tilted in different angles. Hence the data space can be decomposed as a direct sum $Y = Y_1 \oplus \cdots \oplus Y_m$, where $Y_j$ corresponds to the $j$th image in the tilt series, and each vector $g \in Y$ can be written as $g = g_1 + \cdots + g_m$ where $g_j \in Y_j$. The forward operator $T$ has a corresponding decomposition into components $T_j : X \rightarrow Y_j$.

A commonly used approximation of the forward operator is that each $T_j$ consists of a parallel beam transform in a direction depending on $j$, followed by convolution with a point spread function. The noise in the data comes mainly from the stochastic nature of the detection of the electrons, and has a Poisson distribution. Due to the necessity of using a very low electron dose in each image, the noise level is usually very high.

In this model, the assumption of translation invariance made in Section 3 is valid. Since $(Tf, G_{\text{noise}})$ is composed of noise from all the images, which can reasonably be assumed to be independent, the assumption that its probability distribution is Gaussian seems plausible according to the central limit theorem.

4.1 Estimation of $\sigma(f)$

In order to apply the proposed parameter choice method, it is necessary to estimate $\sigma(f)$ for a given $f \in X$. For tomographic data of the type encountered in ET, this estimate can be made directly from the data set, given the following very reasonable assumptions.

1. The noise components in separate images are uncorrelated.
2. The noise components in different parts of the same image are at most weakly correlated.
3. The probability distribution of $(T_j f, G_{j \text{noise}})$ is invariant under translation of $f$.
4. In each image, the signal to noise ratio is much lower than 1.

Now, by assumption [1] we have that

$$\operatorname{Var}[\langle Tf, G_{\text{noise}} \rangle] = \sum_{j=1}^{m} \operatorname{Var}[\langle T_j f, G_{j \text{noise}} \rangle].$$

(14)
To estimate the right hand side of this equality, take a number of random translations $f_1, \ldots, f_l$ of $f$ (this requires that the support of $f$ is small compared to $\Omega$). I claim that $\text{Var}[(T_j f, G_j^{\text{noise}})]$ can be approximated by a sample variance

$$\text{Var}[(T_j f, G_j^{\text{noise}})] \approx \frac{1}{l-1} \sum_{i=1}^{l} \left( (T_j f_i, G_j^{\text{noise}}) - \frac{1}{l} \sum_{i'=1}^{l} (T_j f_{i'}, G_j^{\text{noise}}) \right)^2. \quad (15)$$

This is justified by assumptions 2 and 3. Finally, assumption 4 justifies that we can replace $g^{\text{noise}}$ by $g^{\text{data}}$ in (15). Combining these steps leads to the following approximation:

$$\sigma(f)^2 \approx \sum_{j=1}^{m} \frac{1}{l-1} \sum_{i=1}^{l} \left( (T_j f_i, g_j^{\text{data}}) - \frac{1}{l} \sum_{i'=1}^{l} (T_j f_{i'}, g_j^{\text{data}}) \right)^2. \quad (16)$$

### 4.2 Numerical results

The numerical results presented in this section were obtained by approximately solving the minimization problem (1) with $T$ the forward operator from electron tomography described above and $g$ a real or simulated data set. The regularization functional is an approximation of the total variation norm, defined as follows. Let $x_{i,j,k}, (i, j, k) \in I \subset \mathbb{Z}^3$ be a rectangular lattice of points at which the function $f$ is sampled. For all $(i, j, k) \notin I$ we take $f(x_{i,j,k})$ to be 0. Let $\bar{I}$ be the subset of $\mathbb{Z}^3$ consisting of $I$ together with all points adjacent to $I$. Let

$$D^+_1 f(x_{i,j,k}) := f(x_{i+1,j,k}) - f(x_{i,j,k})$$
$$D^+_2 f(x_{i,j,k}) := f(x_{i,j+1,k}) - f(x_{i,j,k})$$
$$D^+_3 f(x_{i,j,k}) := f(x_{i,j,k+1}) - f(x_{i,j,k})$$

be discrete partial derivatives of $f$ in the forward direction, and similarly let

$$D^-_1 f(x_{i,j,k}) := f(x_{i,j,k}) - f(x_{i-1,j,k})$$
$$D^-_2 f(x_{i,j,k}) := f(x_{i,j,k}) - f(x_{i,j-1,k})$$
$$D^-_3 f(x_{i,j,k}) := f(x_{i,j,k}) - f(x_{i,j,k-1})$$

be discrete partial derivatives in the backward direction for all $(i, j, k) \in \bar{I}$. Now we define

$$R_\lambda(f) = \lambda \sum_{l=1}^{3} \left( \beta^2 + \frac{1}{2} \sum_{i=1}^{3} \left( (D^+_l f(x_{i,j,k}))^2 + (D^-_l f(x_{i,j,k}))^2 \right) \right)^{1/2}. \quad (17)$$

The parameter $\beta$ is included in order to make the regularization functional smooth, which was necessary for the minimization algorithm used. (For alternative optimization algorithms which do not require this approximation, see for example [5].) It was set to a small positive value, $\beta = 3 \cdot 10^{-4}$. This is well below the level where changes in $\beta$ do not seem to have any noticeable effect on the solution. However, in the application of the parameter choice rule, $\beta$ was set to 0, so that the condition 2 is exactly satisfied.

The approximate solution of the minimization problem (1) was computed by iteratively searching for the minimum in 2-dimensional subspaces of $X$, where each subspace is spanned by the gradient of the objective functional and a vector in the direction of the previous update. This method seems to be considerably faster than a
gradient descent method minimizing over a 1-dimensional subspace in each iteration. The iteration was continued until no appreciable change occurred even after many iterations. The number of iterations used was in most cases between 500 and 1500, with a fairly good approximation of the end result occurring within 100 iterations.

4.2.1 Simulated data set

The first numerical example is reconstructed from a simulated data set. The phantom used in the simulation contains 30 Y-shaped objects of varying size and contrast. From this phantom an ET data set was simulated, consisting of 121 projections. The specimen was tilted about a single axis, with the tilt angle ranging from $-60^\circ$ to $60^\circ$.

The simulated electron dose was 15.7 electrons per pixel on average over the tilt series. A section through the phantom and the central projection in the data set are shown in figure 1.

![Figure 1](image)

Figure 1: Left: A section through the phantom. Right: The central projection in the data set.

Next the parameter choice rule from the previous sections was applied to the data set. Note that the only input needed is the data set, a model for the forward operator and the threshold parameter $a$. A reasonable choice for the threshold parameter, given the overall levels of contrast in the phantom, seems to be $a = 0.5$. With this choice we should hope that the reconstructed objects are well above the noise level, even if the contrast is somewhat reduced by the regularization. Figure 2 shows the dependence of $\lambda$ on $a$, and the dependence of $s_{\text{min}}(f_d)$ and $s_{\lambda}(f_d)$ on $f_d$ for the choice of $\lambda$ corresponding to $a = 0.5$. The diameter $d$ is measured in voxel units.

If the proposed parameter choice rule is applicable, $\lambda \approx 30$ should be a suitable choice of the regularization parameter. To test if this is the case, a series of TV regularized reconstructions were computed with the regularization parameter ranging from 12 to 48. The size of the reconstructions is $200 \times 200 \times 100$ voxels. A section through each of the reconstructions is shown in figures 3–7. Which one of these reconstructions would be considered optimal is of course strongly dependent on the type of further analysis it is intended for.

To further investigate the amount of undesirable noise in the reconstructions, the following analysis was applied. For a given threshold $a \geq 0$ and a given reconstruction $f_{\text{rec}}$, the set $\{x \in \Omega : f_{\text{rec}}(x) > a\}$ was computed and decomposed into its connected components. (In the discrete setting, a voxel was considered to be connected to each of its 8 nearest neighbors.) A connected component was classified as a true hit if it has nonempty intersection with some object in the phantom, otherwise it was classified as a false hit. Hence, the number of true and false hits can be counted, where the count...
Figure 2: Left: The solid line shows the dependence of the regularization parameter $\lambda$ on the threshold parameter $a$ when it is chosen according to the proposed parameter choice rule. The graph is composed of line segments since a discrete set of diameters $d$ were used in the computation. This graph is compared to the ideal parameter choice rule for this problem (stars), which can be determined by comparing a set of reconstructions with different regularization parameters to the true solution (see below). Right: The dependence of $s_{\min}(f_d)$ (circles) and $s_{\lambda}(f_d)$ (stars) on $d$ for $a = 0.5$ and the corresponding $\lambda = 31.4$. For $d$ greater than approximately 6, $s_{\min}(f_d)$ drops below 0, and does not impose any restricton on $\lambda$.

Figure 3: Left: Section through reconstruction with $\lambda = 12$. Right: Section through reconstruction with $\lambda = 16$.

Figure 4: Left: Section through reconstruction with $\lambda = 20$. Right: Section through reconstruction with $\lambda = 24$. 

9
Figure 5: *Left:* Section through reconstruction with $\lambda = 28$. *Right:* Section through reconstruction with $\lambda = 32$. These are in the range obtained by applying the proposed parameter selection rule, depending on the threshold parameter.

Figure 6: *Left:* Section through reconstruction with $\lambda = 36$. *Right:* Section through reconstruction with $\lambda = 40$.

Figure 7: *Left:* Section through reconstruction with $\lambda = 44$. *Right:* Section through reconstruction with $\lambda = 48$. 
depends both on the regularization parameter $\lambda$ and the threshold $a$. Only one true hit was counted for each of the objects in the phantom, so the number of true hits can never exceed 30. Table 1 shows how the number of true and false hits depend on the regularization parameter when $a = 0.5$. For $\lambda$ below a certain level, the number of false hits increases dramatically, and there is an obvious risk of misinterpretation. In this case, the proposed parameter choice rule makes a surprisingly accurate prediction of the point where false hits start to occur.

| $\lambda$ | 12 | 16 | 20 | 24 | 28 | 32 | 36 | 40 | 44 | 48 |
|-----------|----|----|----|----|----|----|----|----|----|----|
| True hits | 25 | 24 | 22 | 19 | 19 | 16 | 14 | 13 | 13 | 12 |
| False hits | 471 | 98 | 28 | 5  | 1  | 0  | 1  | 0  | 0  | 0  |

Table 1: The number of true and false hits at threshold level 0.5 as a function of the regularization parameter $\lambda$.

Another way to look at false hits is as follows. In a given reconstruction we can compute the smallest threshold $a$ which does not give rise to any false hits. This defines a relation between $a$ and $\lambda$ which can be considered as an ideal parameter choice rule. The only obvious way to determine this ideal parameter choice rule is to compute a set of reconstructions with different regularization parameters and compare them to the true solution $f^{\text{true}}$, in order to classify reconstructed objects as true or false hits. This, of course, impossible in real life problems, primarily because $f^{\text{true}}$ is not known. However, with simulated data, the ideal parameter choice rule can be compared to a practically applicable parameter choice rule.

Such a comparison is shown in figure 1 above. The plot shows that, while there is certainly a discrepancy, the proposed parameter choice rule provides at least a rough approximation of the ideal parameter choice rule for this particular problem. The regularization parameter selected by the proposed rule tends to be too small when $a$ is small and too large when $a$ is large. The reason for this trend is not yet clear.

### 4.2.2 Real data set of TMV specimen

The second example uses a real ET data set of a specimen containing Tobacco Mosaic Virus (TMV). The TMV is a long and fairly rigid cylindrical object, approximately 18 nm in diameter. The tilt series contains 61 projections, with the tilt angle varying in the range $-60^\circ$ to $60^\circ$. The electron dose was 64.5 electrons per pixel on average over the tilt series.

Application of the parameter choice rule to the data set yields the dependence of the regularization parameter $\lambda$ on the threshold $a$ shown in figure 8. This suggests that $\lambda \approx 70$ should be a suitable choice.

Sections through reconstructions with a range of regularization parameters are shown in figures 9–11. The reconstructed volume contains two TMV’s, which are clearly visible in the images.

When working with real data it is of course not possible to know for certain which features in a reconstruction correspond to real objects in the specimen. In this case, all that we can be reasonably sure about is that the large elongated objects in the reconstructions indicate the presence of TMV’s in the specimen. Smaller objects might be due to noise, but could also represent some contamination in the specimen. Table 2 lists the number of small objects above the threshold level 0.5 in the reconstructed volume for different values of the regularization parameter.

The number of small objects in the reconstruction grows rapidly when the value of the regularization parameter goes below the value provided by the parameter choice
Figure 8: Left: The dependence of the regularization parameter $\lambda$ on the threshold parameter $a$ when it is chosen according to the proposed parameter choice rule. Right: The dependence of $s_{\min}(f_d)$ (circles) and $s_{\lambda}(f_d)$ (stars) on $d$ for $a = 0.5$ and the corresponding $\lambda = 73.5$.

Figure 9: Left: The central projection in the data set. Right: Section through reconstruction with $\lambda = 40$.

Figure 10: Left: Section through reconstruction with $\lambda = 50$. Right: Section through reconstruction with $\lambda = 60$. 
Table 2: The number of small objects, not classified as TMV’s, at threshold level \( a = 0.5 \) as a function of the regularization parameter \( \lambda \).

| Regularization parameter \( \lambda \) | 40 | 50 | 60 | 70 | 80 |
|---------------------------------------|----|----|----|----|----|
| Number of small objects               | 129| 40 | 17 | 5  | 1  |

Discussion

The numerical examples presented suggest that the parameter choice rule outlined in this paper might be useful for inverse problems of the type encountered in ET. To further investigate under what circumstances the method would be useful, it is indispensable to gain a better understanding of the underlying mathematics. Once this has been done it should be possible to devise a pertinent set of numerical test cases.

It might certainly be argued that in these examples, some interpretations of details could be more easily made from the reconstructions with a smaller regularization parameter than the one indicated by the proposed parameter choice rule. This suggests that perhaps the parameter choice rule should be modified so that the computed value of \( s_{\min} \) is somewhat smaller. This could be done by relaxing the restriction on the expected number of j with \( |\alpha_{\lambda}(f_{j,d}^{\text{noise}})| > a \), which was in the derivation above, rather arbitrarily, chosen as 1. However, if \( s_{\min} \) is decreased, both the heuristic argument and the numerical examples indicate that one should expect to have false objects appearing in the reconstructions, and the interpretations must be made with this in mind.

In the application to ET data exemplified in this paper, the proposed parameter choice rule is heuristic, or error free, the magnitude of the noise being estimated directly from the data. It is well known that such heuristic parameter choice rules, can not provide convergent regularization methods in the strict sense, that is, methods that converge to the true solution when the noise level approaches zero. Such heuristic parameter choice rules might nevertheless be very useful in practice, when the noise level certainly does not approach zero. As mentioned in the introduction, one of the motivations for this new parameter choice rule is the difficulties associated with very
high noise levels. A variant of the proposed parameter choice rule would be to estimate the quantities $\sigma(f)$ from some a priori estimate of the noise level. Used in this way, the parameter choice could provide a convergent regularization method.

In the derivations leading up to the proposed parameter choice rule, particularly in the estimation of $\sigma(f)$, rather restrictive, albeit very realistic, assumptions were made on the statistical properties of the noise. It is worth noting that the necessity of making such assumptions, explicitly or implicitly, is inherent to the inverse problem in ET; without them no reconstruction would be possible at all. A malicious demon, if allowed to choose the noise with only a restriction on its norm, could easily hide every trace of the signal we are trying to recover.

One advantage of the suggested parameter choice rule, as compared for example to the L-curve method, is its computational efficiency. In the numerical examples considered here, the application of the parameter choice rule has a lower computational cost than one single iteration of the algorithm subsequently used to compute the regularized solution.

The considerations leading up to the parameter choice rule can also be employed in an alternative way. Suppose a regularization parameter has been chosen by some other method and a reconstruction has been computed. In the interpretation of the reconstruction it is then desirable to know which features are reliable, and which are likely to be an effect of noise. A computation of the quantity $s_\lambda(f)$, where $f$ is a certain feature in the reconstruction, can then be used to estimate how likely it is for such features to occur solely as a consequence of random noise.

Acknowledgements
The author wants to thank Ozan Öktem for encouragement and support, and Sidec AB for supplying ET data.

References

[1] T. Chan, S. Esedoglu, F. Park, and A. Yip. Total variation image restoration: overview and recent developments. In Handbook of mathematical models in computer vision, pages 17–31. Springer, New York, 2006.

[2] Heinz W. Engl, Martin Hanke, and Andreas Neubauer. Regularization of inverse problems, volume 375 of Mathematics and its Applications. Kluwer Academic Publishers Group, Dordrecht, 1996.

[3] D. Fanelli and Ö. Öktem. Electron tomography: a short overview with an emphasis on the absorption potential model for the forward problem. Inverse Problems, 24(1):013001, 2008.

[4] Enrico Giusti. Minimal surfaces and functions of bounded variation, volume 80 of Monographs in Mathematics. Birkhäuser Verlag, Basel, 1984.

[5] M. Hintermüller and K. Kunisch. Total bounded variation regularization as a bilaterally constrained optimization problem. SIAM J. Appl. Math., 64(4):1311–1333 (electronic), 2004.