ITERATIVE CHOICE OF THE OPTIMAL REGULARIZATION PARAMETER IN TV IMAGE RESTORATION

ALINA TOMA AND BRUNO SIXOU
CREATIS, CNRS UMR 5220; INSERM U1044; INSA de Lyon; Université de Lyon 1
Université de Lyon, 69621, Villeurbanne Cedex, France

FRANÇOISE PEYRIN
CREATIS, CNRS UMR 5220; INSERM U1044; INSA de Lyon; Université de Lyon 1
Université de Lyon, 69621, Villeurbanne Cedex, France
and
ESRF, Imaging Group, 38043, Grenoble Cedex, France

(Communicated by Antonin Chambolle)

ABSTRACT. We present iterative methods for choosing the optimal regularization parameter for linear inverse problems with Total Variation regularization. This approach is based on the Morozov discrepancy principle or on a damped version of this principle and on an approximating model function for the data term. The theoretical convergence of the method of choice of the regularization parameter is demonstrated. The choice of the optimal parameter is refined with a Newton method. The efficiency of the method is illustrated on deconvolution and super-resolution experiments on different types of images. Results are provided for different levels of blur, noise and loss of spatial resolution. The damped Morozov discrepancy principle often outperforms the approaches based on the classical Morozov principle and on the Unbiased Predictive Risk Estimator. Moreover, the proposed methods are fast schemes to select the best parameter for TV regularization.

1. Introduction. Linear inverse problems are encountered in many engineering applications and they play a major role in image processing. For example, image deconvolution and restoration problems are crucial in medical imaging or image and video coding. Such problems can be formulated as:

\( g^\delta = Af + n, \)

where \( g^\delta \) represents the measured data, \( f \) the ideal image, \( n \) an additive noise corrupting the measurements and \( A \) a linear transform like a blurring kernel or the composition of a smoothing kernel and of an undersampling operator. This problem is ill-posed and small changes in the data can have large effects on the solutions. To obtain stable solutions in the presence of noise, a well-known approach is to include some prior information and to minimize a regularization functional including a data fidelity term and a regularization term \( R(f) \) that imposes an a priori constraint on the solution. A regularization parameter is balancing the contribution of the two terms. The effectiveness of the regularization approach depends on the choice of the optimal regularization parameter. A well-known method to control the noise

2010 Mathematics Subject Classification. Primary: 65J22, 65J20, 65K10; Secondary: 52A41.
Key words and phrases. Linear inverse problems, total variation regularization, regularization parameter, Morozov principle, UPRE.
is the Tikhonov regularization [35] with \( R(f) = \|Df\|^2 \), where \( D \) is a differential operator. More recently, the Total Variation regularization (TV) [32, 38] has been studied and applied to various image processing problems [41, 6, 27, 30]. For an image \( f \) in the first order Sobolev space \( H_1(\Omega) \), this regularization term is given by:

\[
R_{\text{TV}}(f) = \int_\Omega |\nabla f(r)| \, dr
\]

This regularization is known to preserve well edges. Convergence rates for TV regularization have been obtained under additional source conditions with quantitative estimates for the Bregman distance induced by the regularization functional [10]. Various numerical methods have been used to solved the TV regularized deconvolution problem [36, 7]. Yet, extensive numerical experiments show that algorithms based on the Alternating Direction Method of Multipliers (ADMM) are the state-of-the-art methods [41, 27, 43, 14, 1, 2].

In order to obtain a good reconstructed image, it is necessary to choose an optimal regularization parameter. A very common approach is to try various values until the reconstruction solution is satisfactory. For large scale problems, an exhaustive search with a grid of points of regularization parameters values is very slow. Several methods have been studied for the optimal selection of the regularization parameter for Tikhonov regularization. Methods like the discrepancy principle and the Unbiased Predictive Risk Estimator are based on a knowledge of the noise level \( \delta \) [38, 26, 8, 12, 28, 37, 11, 29, 40, 39]. The L-curve method and the Generalized Cross-Validation can be used without the knowledge of this noise level [17, 19]. Moreover, an iterative choice of the regularization parameters based on the Morozov discrepancy principle applied with a model function has been proposed in [22].

For TV regularization, fewer methods have been investigated. For denoising problems, Aujol et al. [4] have proposed a method where the denoised result is close to the optimal, in the SNR sense. An extension of the Unbiased Predictive Risk Estimator to TV regularization has been shown to give a good estimate of the regularization parameter [25]. A method exploiting the Generalized Cross-Validation technique has been proposed in [24]. Babacan et al. [5, 6] have studied a bayesian variational method for solving TV deconvolution and super-resolution problems. A new approach based on a proximal point method applied to the dual formulation of the TV regularization problem and the discrepancy principle has been proposed recently [42].

The goal of this paper is to propose a fast and simple method to select the optimal regularization parameter for the restoration with TV regularization of an image \( f \) from a blurred, noisy image \( g^\delta \) with various sub-sampling rates. Our contribution in this field is to design an iterative method following the idea of the model function for Tikhonov regularization of Kunisch et al. [22]. The outline of this paper is the following. In Section 2, we present the linear inverse problem studied, the Morozov discrepancy principle, a damped version of this principle, and some convergence results associated to our method of choice of the regularization parameter and the ADMM/TV algorithm chosen to estimate the regularized solutions. In Section 3, the new method for the selection of the regularization parameter is presented. In Section 4, we detail the UPRE method that will be used for comparison. In Section 5, numerical results are given to demonstrate the effectiveness of the approach for deconvolution or super-resolution problems. Concluding remarks are given in the last section.
2. TV regularization, Morozov based principles and ADMM approach.

2.1. Problem formulation. Let $\Omega$ be a bounded open domain with Lipschitz boundary. In the following, we consider a linear operator $A : L^2(\Omega) \rightarrow L^2(\Omega)$. We assume that $A$ is injective and continuous. We consider the regularization functional:

\[ J_{\mu,g^\delta}(f) = \mu \|Af - g^\delta\|^2 + R_{TV}(f) \]

where $\mu$ is a regularization parameter. The Total Variation regularization has been extensively studied in [34, 3]. The following theorem summarizes the main convergence and stability results.

**Theorem 2.1.** Let $A$ be a linear and bounded operator $A : L^2(\Omega) \rightarrow L^2(\Omega)$, satisfying $Aq \neq 0$ for all constants $q \in L^2(\Omega)$. The space $L^2(\Omega)$ is assumed to be endowed with its weak topology. Then:

- the TV semi-norm is convex and lower semicontinuous with respect to the weak topology on $L^2(\Omega)$.
- the level sets \( \{ f \in L^2(\Omega), \mu \|Af - g\|^2 + R_{TV}(f) \leq M \} \) are sequentially pre-compact with respect to the weak topology on $L^2(\Omega)$, for any positive numbers $M$ and $\mu$.
- consequently, the regularization with the TV semi-norm is well-defined, stable and convergent.

Let \( (g_k) \) be a sequence converging toward $g^\delta$, and $f_k$ be a minimizer of the regularization functional $J_{\mu,g_k}$. The stability of the regularization method means that every convergent subsequence of $(f_k)$ converges to a minimizer of $J_{\mu,g}$. Let $(\delta_k)$ be a sequence of positive numbers converging to 0, and $(g_k)$ a sequence such that $\|g - g_k\| \leq \delta_k$, the TV regularization method is convergent and thus for a good choice of the parameter $\mu(\delta_k)$, the sequence $(f_k)$ has a convergent subsequence. The limit is a minimal norm solution [34].

2.2. Morozov and modified Morozov principle, basic properties. In the following, we denote as $f(\mu)$ the regularized solution obtained with the regularization parameter $\mu$. The noise free data $g$ is assumed to be corrupted by noise in such a way that:

\[ \|g - g^\delta\| \leq \delta \]

where $\delta$ is the noise level. The unique solution of $Af = g$ will be denoted $f^\ast$. With the Morozov discrepancy principle [13], the regularization parameter is chosen such that the discrepancy term is equal to the known noise level $\delta$ on the observed data, so that the following equation for $\mu$ holds:

\[ \|Af(\mu) - g^\delta\| = \delta \]

In some cases, the use of the Morozov discrepancy principle may not lead to an optimal regularization parameter. As suggested in [22, 23], a damped Morozov equation may be more satisfactory. We have thus considered the damped equation:

\[ \|g^\delta - Af(\mu)\|^2 + \mu^\gamma R_{TV}(f(\mu)) = \delta^2 \]

where the exponent $\gamma$ has to be chosen.

We first study the range of the functions $P(f(\mu)) = \|g^\delta - Af(\mu)\|^2$ and $P_\gamma(f(\mu)) = \|g^\delta - Af(\mu)\|^2 + \mu^\gamma R_{TV}(f(\mu))$. We define:

\[ \eta = \inf\{\|Af - g^\delta\|, f \in L^2(\Omega)\} \]
and
\begin{equation}
ν = \inf \{ \|Af_C - g^δ\|, f_C = C \text{ on } Ω, C ∈ ℝ \}
\end{equation}

We denote \( f₀ \) the value of \( f \) for which the minimum \( ν \) is obtained. We are first interested in some basic properties of \( P_γ \) and \( P \). The proofs are similar to the one given in Kunisch et al [23], but some changes are necessary because of the use of the TV regularization term.

**Lemma 2.2.** For \( γ ≤ -1 \), \( P_γ(f(μ)) ≥ ν^2 \) as \( μ → 0 \). As \( μ → ∞ \), \( P_γ(f(μ)) → η^2 \). The same results hold for \( P \).

**Proof.** The minimum of \( J_μ(f) \) is obtained for \( f(μ) \) and \( R_TV(f₀) = 0 \) thus:
\[
\frac{\|Af(μ) - g^δ\|^2}{2} + μ⁻¹R_TV(f(μ)) ≤ \frac{\|Af₀ - g^δ\|^2}{2} + μ⁻¹R_TV(f₀)
\]
\[
≤ \frac{ν^2}{2} + μ⁻¹R_TV(f(μ))
\]
Thus \( ∥Af(μ) - g^δ∥ ≤ ν \) and \( R_TV(f(μ)) ≤ μ ∥Af₀ - g^δ∥^2 \).

Therefore \( R_TV(f(μ)) → 0 \) as \( μ → 0 \). Let \( f > \) be the average value of \( f(μ) \) on \( Ω, (R_TV(f(μ)) - f) → 0 \) as \( μ → 0 \). As the TV semi-norm is equivalent to the BV norm for functions with zero average value [34], this result implies \( ∥f(μ) - f∥_BV → 0 \) as \( μ → 0 \). Therefore \( f(μ) → f > \) in \( L₂(Ω) \) since BV is continuously embedded in \( L₂(Ω) \) [34]. The operator \( A \) is continuous on \( L₂(Ω) \) and thus \( ∥Af(μ) - g^δ∥^2 → ∥A < f > - g^δ∥^2 \) as \( μ → 0 \). Consequently \( P_γ(f(μ)) ≥ ν^2 \) as \( μ → 0 \).

For \( ε > 0 \), we consider \( f_ε \) such that \( ∥Af_ε - g^δ∥ ≤ η + ε \).
\[
\frac{∥Af(μ) - g^δ∥^2}{2} + μ⁻¹R_TV(f(μ)) ≤ \frac{∥Af_ε - g^δ∥^2}{2} + μ⁻¹R_TV(f_ε)
\]
\[
≤ \frac{(η + ε)^2}{2} + μ⁻¹R_TV(f_ε)
\]
and thus for \( μ ≥ 1 \) and \( γ ≤ -1, μ^γ < \frac{2}{μ} \) and
\begin{equation}
∥Af(μ) - g^δ∥^2 + μ^γR_TV(f(μ)) ≤ (η + ε)^2 + 2μ⁻¹R_TV(f_ε)
\end{equation}

On the other hand, \( ∥Af(μ) - g^δ∥^2 + μ^γR_TV(f(μ)) ≥ η^2 \).

Let \( μ₁ = 2R_TV(f_ε)/ε^2 \), then it follows from (9) that for \( μ ≥ μ₁, √P_γ(f(μ)) ≤ η + 2ε \) and thus \( P_γ(f(μ)) → η^2 \) as \( μ → ∞ \). From the former results, it follows that \( P(f(μ)) ≥ ν^2 \) as \( μ → 0 \) and \( P(f(μ)) → η^2 \) as \( μ → ∞ \).

**2.3. Convergence results.** It is well-known that for ill-posed operator equations, the convergence rates of regularized solutions require additional properties of the true solution \( f^* \) of the equation \( Af = g \) such as source conditions. Convergence rates results can be obtained in terms of Bregman distance between the true solution and the regularized solution [10, 31, 34]. The Bregman distance is a measure of the difference between two elements in a Banach space [21, 9]. The use of this distance has now evolved as a standard tool for error measure. For a element \( ξ ∈ ∂R_TV(v) \), where \( ∂R_TV(v) \) denotes the subdifferential of TV at \( v \), the Bregman distance with respect to TV between \( u \) and \( v \) at \( ξ \) is defined by: \( D_{TV}^ξ(u, v) = R_TV(u) - R_TV(v) - (ξ, u - v) \). In the following, we assume the standard source condition holds and that there exist \( τ ∈ L₂(Ω) \) such that:
\begin{equation}
ξ = A(f^*)^*τ ∈ ∂R_TV(f^*)
\end{equation}
Some convergence results have been obtained for a priori choices of the regularization parameter \([34, 18]\) given by the following proposition:

**Proposition 2.3.** Assume that the former source condition holds. Then for the choice \(1/\mu \sim \delta\), the Bregman distance between the regularized solution and the ground truth is \(D_{TV}^\xi(f(\mu), f^*) = O(\delta)\).

These control of the error in terms of Bregman divergence was extended by a \(l_1\) control in the discrete setting \([15]\). We present here some convergence rates results as the error level \(\delta \to 0\). We develop an error estimation when the regularization parameter \(\mu(g^\delta, \delta)\) is chosen according to the modified Morozov principle (6). We first show that the TV norm of \(f(\mu)\) is uniformly bounded if the modified Morozov principle is satisfied.

**Lemma 2.4.** If the regularization parameter is chosen according to the modified Morozov principle, \(\mu > 2^{\frac{1}{\gamma+1}}\) and \(\gamma < -1\), the TV semi-norm of the regularized solution \(R_{TV}(f(\mu))\) is uniformly bounded.

**Proof.** From the definition of \(f(\mu)\), we have:

\[
\|Af(\mu) - g^\delta\|^2 + \mu^{-1}R_{TV}(f(\mu)) \leq \frac{\delta^2}{2} + \mu^{-1}R_{TV}(f^*)
\]

Combining the modified Morozov Principle and (11), we obtain:

\[
\|Af(\mu) - g^\delta\|^2 + \mu^{-1}R_{TV}(f(\mu)) = \frac{\|Af(\mu) - g^\delta\|^2}{2} + (\mu^{-1} - \frac{\mu^\gamma}{2})R_{TV}(f(\mu)) + \frac{\mu^\gamma}{2}R_{TV}(f(\mu))
\]

\[
= \frac{\delta^2}{2} + (\mu^{-1} - \frac{\mu^\gamma}{2})R_{TV}(f(\mu))
\]

\[
\leq \frac{\delta^2}{2} + \mu^{-1}R_{TV}(f^*)
\]

and consequently \(R_{TV}(f(\mu)) \leq \frac{\frac{1}{1 - \frac{\mu_0^\gamma}{2}}R_{TV}(f^*)}{\mu_0}\). For \(\mu \geq \mu_0\) with \(\mu_0 > 2^{\frac{1}{\gamma+1}}\) and \(\gamma + 1 < 0\)

\[
R_{TV}(f(\mu)) \leq \frac{1}{1 - \frac{\mu_0^\gamma}{2}}R_{TV}(f^*),
\]

\(R_{TV}(f(\mu))\) is thus uniformly bounded. \(\square\)

**Proposition 2.5.** If the regularization parameter is chosen according to the modified Morozov principle, the Bregman distance between the true solution and the regularized solution is given by \(D_{TV}^\xi(f(\mu), f^*) = O(\delta^{\min(1,2(\gamma+1)/\gamma)})\).

**Proof.** From the proof of Proposition 3.41 in \([34]\), it can be shown that if assumption (10) holds, then there exist \(\beta_1 \in [0,1]\), \(\beta_2 \geq 0\) and \(\xi \in \partial R_{TV}(f^*)\) such that:

\[
\|Af(\mu) - g^\delta\|^2 + \frac{1}{\mu}D_{TV}^\xi(f(\mu), f^*) \leq \delta^2 + \frac{1}{\mu}(\beta_1 D_{TV}^\xi(f(\mu), f^*) + \beta_2(\|Af(\mu) - g^\delta\| + \delta))
\]

With the choice of \(\mu(g^\delta, \delta)\) according to the modified Morozov principle:

\[
\frac{1}{\mu}D_{TV}^\xi(f(\mu), f^*) \leq \frac{1}{\mu}(\beta_1 D_{TV}^\xi(f(\mu), f^*) + \beta_2(\|Af(\mu) - g^\delta\| + \delta)) + \mu^\gamma R_{TV}(f(\mu))
\]
and thus
\[
D^2_{TV}(f(\mu), f^*)(1 - \beta_1) \leq \beta_2(\|Af(\mu) - g^δ\| + \delta) + \mu^{\gamma + 1}R_{TV}(f(\mu)).
\]

The modified Morozov principle implies \(\mu^\gamma R_{TV}(f(\mu)) \leq \delta^2\) and we obtain
\[
D^2_{TV}(f(\mu), f^*)(1 - \beta_1) \leq \beta_2(\|Af(\mu) - g^δ\| + \delta) + \frac{1}{1 - \frac{\mu^{\gamma + 1}}{2}}\delta^{2(\gamma + 1)/\gamma}R_{TV}(f^*)^{-1/\gamma}
\]
and with (12), it follows that:
\[
D^2_{TV}(f(\mu), f^*)(1 - \beta_1) \leq \beta_2(\|Af(\mu) - g^δ\| + \delta) + \frac{1}{1 - \frac{\mu^{\gamma + 1}}{2}}\delta^{2(\gamma + 1)/\gamma}R_{TV}(f^*)^{-1/\gamma}
\]
and therefore \(D^2_{TV}(f(\mu), f^*) = O(\delta^{\min(1,2(\gamma + 1)/\gamma)})\).

2.4. The ADMM approach. The iterative method developed in this paper to select the regularization parameter is independent of the algorithm used to minimize the regularization functional. Many approaches have been proposed to solve the TV regularization problem. The recent ADMM methodology is the state-of-the-art method to obtain a stable solution with the TV regularization [27, 41]. This type of algorithms has attracted much attention and has also been proposed to solve a number of image processing tasks, such as image inpainting and deblurring. The convergence rates for the Augmented Lagrangian method when the Morozov’s discrepancy principle is chosen as parameter selection method has been investigated in [16].

For discrete data, in the framework of the ADMM method, an augmented Lagrangian \(\mathcal{L}(f, \{g_i\}_{1 \leq i \leq N^2}, \{\lambda_i\}_{1 \leq i \leq N^2})\) is considered given for \(f \in \mathbb{R}^{N^2}\) by:
\[
\mathcal{L}(f, \{g_i\}, \{\lambda_i\}) = \sum(\|g_i\|_2 - \lambda_i^t(g_i - D_i f)) + \frac{\beta}{2}\|g_i - D_i f\|^2_2 + \frac{\gamma}{2}\|g^δ - Af\|^2_2
\]
where \(\mu\) is the regularization parameter, \(\beta\) the Lagrangian parameter, and \(\|\cdot\|_2\) is the Euclidean norm in \(\mathbb{R}^{2}\). For the ith pixel at the location \((ix, iy)\), the first order finite difference operator \(D_i\) is defined by \(D_i : \mathbb{R}^{N^2} \rightarrow \mathbb{R}^2\), \(D_i(f) = (D^x_i(f), D^y_i(f))\). The horizontal difference operator is given by:
\[
D^x_i(f) = \begin{cases} f_{ix+1, iy} - f_{ix, iy} & i x < N \\ 0 & i x = N \end{cases}
\]
A similar definition holds for the vertical difference operator \(D^y_i(f)\). The augmented Lagrangian depends on the auxiliary unknown \(g_i = (g^x_i, g^y_i) \in \mathbb{R}^2\) and on the dual variable associated to the constraint \(g_i = D_i(f)\), denoted \(\lambda_i \in \mathbb{R}^2\). The ADMM algorithm is searching for the saddle point of the augmented Lagrangian that is obtained by iterating the following equations:
\[
f^{k+1} = \arg\min_x \mathcal{L}(x, g^k_i, \lambda^k_i)
g^{k+1}_i = \arg\min_y \mathcal{L}(f^{k+1}, y, \lambda^k_i)
\lambda^{k+1}_i = \lambda_i^k - \beta(g^{k+1}_i - D_i f^{k+1})
\]
With the alternating minimization algorithm, three sequences \((f^k, g^k_i, \{\lambda^k_i\})\) are constructed with the following iterative scheme:
For each pixel \( i \), \( g_i \) is updated with:

\[
g_i^{k+1} = \max \{ \| D_i f^k + \frac{1}{\beta} (\lambda_i^k) \| - \frac{1}{\beta}, 0 \} \frac{D_i f^k + \frac{1}{\beta} (\lambda_i^k)}{\| D_i f^k + \frac{1}{\beta} (\lambda_i^k) \|} \]

The new iterate \( f^{k+1} \) is obtained from the following linear system:

\[
(\sum D_i^t D_i + \frac{\mu}{\beta} A^t A) f = \sum D_i^t (g_i^{k+1} - \frac{1}{\beta} \lambda_i^k) + \frac{\mu}{\beta} A^t g^\delta.
\]

The Lagrange multipliers are updated with, \( \lambda_i^{k+1} = \lambda_i^k - \beta (g_i^{k+1} - D_i f^{k+1}) \). The convergence of the algorithm is summarized by the following theorem [27, 41, 16]:

**Theorem 2.6.** The sequence \( (f^k, \{g_i^k\}, \{\lambda_i^k\}) \) generated by the ADMM algorithm converges to a Kuhn-Tucker point (Theorem 2.6). The sequence convergence of the algorithm is summarized by the following theorem [27, 41, 16]:

**Theorem 2.6.** The sequence \( (f^k, \{g_i^k\}, \{\lambda_i^k\}) \) generated by the ADMM algorithm converges to a Kuhn-Tucker point (KKT) conditions can be written as:

\[
(22) \quad \mu A^t (Af(\mu) - g^\delta) = - \sum D_i^t (\lambda_i^*), \quad \lambda_i^* \in \partial \|D_i f(\mu)\|_2, \quad g_i^* = D_i f(\mu)
\]

### 3. Proposed Morozov based methods to choose the regularization parameter

In this section, we detail two numerical schemes to choose the regularization parameter. We assume that \( f(\mu) \) is a minimum \( R_{TV} \) norm solution and that it is unique. We first show, under restrictive conditions, the differentiability of the function \( f \) and of the value function \( E \) around \( \mu \), which are functional properties on which the iterative scheme relies.

#### 3.1. Lipschitz continuity and differentiability of \( f(\mu) \)

We consider here the function \( f : \mathbb{R} \to \mathbb{R}^N \) of the regularization parameter. The results in this section will be formulated in a discrete setting. They can be generalized to the continuous setting if the subdifferential of the discrete TV semi-norm \( \sum D_i^t (\partial \|D_i f(\mu)\|) \) is replaced by its continuous version.

**Proposition 3.1.** Under the assumptions of Theorem 2.1, the regularized solution \( f \) is continuous around the regularization parameter \( \mu \).

**Proof.** We first show that \( f \) is Lipschitz continuous around \( \mu \). Subtracting the KKT condition for \( \mu + t \) and \( \mu \) yields:

\[
(23) \quad \mu A^t A(f(\mu + t) - f(\mu)) \in t A^t A f(\mu + t) + t A^t g^\delta - \sum D_i^t (\partial \|D_i f(\mu + t)\| - \partial \|D_i f(\mu)\|).
\]

Taking the inner product with \( f(\mu + t) - f(\mu) \), using the Schwarz inequality, that \( f \) is bounded and the monotonicity of the subdifferential, we obtain:

\[
(24) \quad \mu \|A(f(\mu + t) - f(\mu))\|^2 \leq C_1 |t| \|f(\mu + t) - f(\mu)\|
\]

where \( C_1 \) is a positive constant and

\[
(25) \quad \sum < \partial \|D_i f(\mu + t)\|_2 - \partial \|D_i f(\mu)\|_2, D_i (f(\mu + t) - f(\mu)) > \leq C_1 |t| \|f(\mu + t) - f(\mu)\|
\]

If \( A \) is injective, \( \|A(f(\mu + t) - f(\mu))\|^2 \geq \alpha_{\min} \|f(\mu + t) - f(\mu)\|^2 \), where \( \alpha_{\min} \) is the minimum singular value of \( A^t A \) and thus \( f \) is Lipschitz continuous around \( \mu \).
The regularization functional for the regularization parameter $\mu + t$ is minimized for $f(\mu + t)$ and thus:

\[
(\mu + t) \frac{\|Af(\mu + t) - g^\delta\|^2}{2} + R_{TV}(f(\mu + t)) \leq (\mu + t) \frac{\|Af(\mu) - g^\delta\|^2}{2} + R_{TV}(f(\mu))
\]

and thus, there are constants $C_2, C_3$ such that:

\[
0 \leq R_{TV}(f(\mu + t)) - R_{TV}(f(\mu)) \\
\leq t \frac{\|Af(\mu) - g^\delta\|^2}{2} \\
+ \mu \left( \frac{\|Af(\mu) - g^\delta\|^2}{2} - \frac{\|Af(\mu + t) - g^\delta\|^2}{2} \right) \\
\leq t \frac{\|Af(\mu) - g^\delta\|^2}{2} + C_2 \|Af(\mu + t) - Af(\mu)\| \\
\leq C_3 \sqrt{t}
\]

where we have used (24).

From (25), we can deduce that, for all $i$, as $t \to 0$, $D_i f(\mu + t) = 0$ if and only if $D_i f(\mu) = 0$, and that the components of $D_1 f(\mu)$ and $D_1 f(\mu + t)$ have the same sign. Therefore, as $t \to 0$, we have, $R_{TV}(f(\mu + t)) - R_{TV}(f(\mu)) \sim R_{TV}(f(\mu + t) - f(\mu))$, and there is a positive constant $C_4$ such that:

\[
R_{TV}(f(\mu + t) - f(\mu)) \leq C_4 \sqrt{t}
\]

Let $BV(\Omega)_\circ$, the set of all functions of bounded variation with zero mean, then on $BV(\Omega)_\circ$ the TV and BV norm are equivalent [34]:

\[
R_{TV}(f) \leq \|f\|_{BV} \leq C_2 R_{TV}(f) , \forall f \in BV(\Omega)_\circ
\]

This property is also verified for the discrete TV semi-norm. Let us denote $\Delta f = f(\mu + t) - f(\mu)$, and $M : L_1(\Omega) \to L_1(\Omega)$ the averaging operator of a function $f$. The operator $M$ is linear and continuous. In a continuous setting, the operator $M$ is defined by:

\[
M(f) = 1_\Omega \int_\Omega f(x), \forall f \in L_1(\Omega)
\]

where $1_\Omega$ is the indicator function of the set $\Omega$. Using the operator $M$ we have:

\[
\|f(\mu + t) - f(\mu)\|_1 \leq \|M(\Delta f)\|_1 + \|\Delta f - M(\Delta f)\|_1
\]

Under the assumption that the operator $A$ is injective on the set of constant functions, we obtain that there is a positive constant $C_5$ such that:

\[
\|M(\Delta f)\|_1 \leq C_5 \|AM(\Delta f)\|_2
\]

Using (29) and (32), we get:

\[
\|f(\mu + t) - f(\mu)\|_1 \leq C_5 \|AM(\Delta f)\|_2 + C_2 R_{TV}(\Delta f - M(\Delta f)) \\
\leq C_5 \|AM(\Delta f)\|_2 + C_2 R_{TV}(\Delta f)
\]

We now assume that the operator $A$ commute with $M$. It can be shown easily that it is the case for the convolution integral operator investigated in the following.
Using the continuity of $M$, (24) and (28), we thus obtain that, as $t \to 0$ there is positive constant $D$ such that:

$$
\|f(\mu + t) - f(\mu)\|_1 \leq C_5 \|MA(\Delta f)\|_2 + C_2 R_{TV}(\Delta f) \leq D\sqrt{t}
$$

(34)

We now demonstrate that under some restricted injectivity condition, the regularized solution is a differentiable function of the regularization parameter.

**Proposition 3.2.** Let us assume there exists a constant $E > 0$ such that the operator $A$ is injective on the subset $C$ of $BV(\Omega)$, defined by:

$$
C = \{ g \in BV(\Omega), \| \sum D_i^f D_i g - \sum D_i^f f(\mu) \|_1 \leq E \}
$$

Then, the regularized solution $f$ is differentiable around the regularization parameter $\mu$.

**Proof.** We now show that the function $f$ is differentiable around $\mu$ and that its derivative is the unique solution of:

$$
\mu A^\dagger A \omega = A^\dagger g^\delta - A^\dagger Af(\mu) \ a.e
$$

(35)

We assume that the operator $A$ is injective on the subset $C$. Let $\epsilon = (f(\mu + t) - f(\mu)) - t\omega$. Dividing (23) by $t$ and substracting (36), we obtain:

$$
\langle \mu A^\dagger A \epsilon, \epsilon \rangle = -\langle A^\dagger A(f(\mu + t) - f(\mu)), \epsilon \rangle + \sum \langle \gamma_i, \epsilon \rangle.
$$

(37)

with $\langle \gamma_i, \epsilon \rangle = -\langle \partial \|D_i f(\mu + t)\|_2 - \partial \|D_i f(\mu)\|_2, D_i \epsilon \rangle$.

The optimality condition shows that as $t \to 0$, $f(\mu + t)$ belongs to the set $C$. It is also the case for $\epsilon$. By the Schwarz inequality:

$$
\mu \alpha_{min} \| \epsilon \| \leq \|f(\mu + t) - f(\mu)\| + \sum \| \gamma_i \|
$$

(38)

where $\alpha_{min}$ is the minimal singular value of $A^\dagger A$ restricted to $C$. When $t \to 0$, $\gamma_i \to 0$, since $\frac{D_i f(\mu)}{\|D_i f(\mu)\|_2}$ and $\frac{D_i f(\mu + t)}{\|D_i f(\mu + t)\|_2}$ are equal for all $i$ as $t \to 0$. Therefore $\epsilon \to 0$ and $f$ is thus differentiable around $\mu$. □

### 3.2. Differentiability of the value function $E(\mu)$.

In the following, we denote $E(\mu)$ the value obtained at the saddle point for $\mathcal{L}(f(\mu), \{g^\gamma(\mu)\}, \{\lambda^\gamma(\mu)\}, \mu)$.

**Proposition 3.3.** The value function $E(\mu)$ is twice differentiable with $E'(\mu) = \frac{\|g^\delta - Af(\mu)\|^2}{2}$ and $E''(\mu) = \langle Af(\mu) - g^\delta, Af'(\mu) \rangle$, and $E'(\mu)$ is a decreasing function.

**Proof.** The differentiability of $E$ follows from its definition and from the differentiability of $f(\mu)$. Classical differentiability calculation rules gives:

$$
\partial E(\mu) = \mu \langle Af(\mu) - g^\delta, Af'(\mu) \rangle + \frac{\|g^\delta - Af(\mu)\|^2}{2} + \sum \langle D_i \partial \|D_i f(\mu)\|_2, f'(\mu) \rangle.
$$

Taking into account the optimality condition, the first derivative of $E(\mu)$ is given by:

$$
E'(\mu) = \frac{\|g^\delta - Af(\mu)\|^2}{2}.
$$

(39)
Thus, the second derivative of $E(\mu)$ is given by:

$$E''(\mu) = \langle Af(\mu) - g\delta, Af'(\mu) \rangle.$$  \hspace{1cm} (40)

Using equation (36), $E''(\mu)$ can be rewritten:

$$E''(\mu) = -\langle \omega, \mu A^t A \omega \rangle = -\langle A\omega, \mu A\omega \rangle.$$  \hspace{1cm} (41)

We obtain that $E''(\mu) \leq 0$. The function $E'(\mu)$ is decreasing and the value function $E(\mu)$ is monotonically increasing and concave.

3.3. Iterative schemes for the proposed Morozov based regularization parameter selection methods. We present in this section the iterative schemes to choose the regularization parameter according to the Morozov or to the damped Morozov principle.

3.3.1. Iterative scheme for the Morozov principle. In the framework of the Morozov discrepancy principle, the regularization parameter is chosen such that the discrepancy term is equal to the noise level $\delta$, that is $\|g - Af(\mu)\| = \delta$. This condition can be written $2E'(\mu) = \delta^2$. In order to determine the regularization parameter and to solve the Morozov equation, we have used a two parameters model function. The approximate value of $\mu$ obtained is refined with a Newton method. Following the model function methodology in [22], a model function can be calculated that preserves the main properties of $E'(\mu)$.

Proposition 3.4. The discrepancy term $E'(\mu)$ can be approximated by the decreasing function, $\mu \rightarrow a\mu^{-2} + b$, with $(a,b) \in \mathbb{R}^2$.

Proof. Upon differentiating the optimality condition with respect to $\mu$, we obtain:

$$A'(Af(\mu) - g\delta) + \mu A'(Af'(\mu)) = -\frac{d}{d\mu} \sum D_i^t \frac{D_i f(\mu)}{\|D_i f(\mu)\|_2}.$$  \hspace{1cm} (42)

Taking the inner product of this equation with $f(\mu)$ and neglecting the second member, we obtain:

$$\langle (Af(\mu) - g\delta), Af(\mu) \rangle + \mu \langle Af'(\mu), Af(\mu) \rangle = 0$$  \hspace{1cm} (43)

This equation will be rewritten:

$$\|Af(\mu) - g\delta\|^2 + \mu \|Af'(\mu) - g\delta\|^2 = \mu \langle Af'(\mu), Af(\mu) - g\delta \rangle + \langle Af(\mu) - g\delta, -g\delta \rangle.$$  \hspace{1cm} (44)

If the right member is assumed constant and using (39) and (40), we get:

$$2E'(\mu) + \mu E''(\mu) = constant$$  \hspace{1cm} (45)

Solving this differential equation, we obtain a decreasing and convex model function depending on two real parameters $a$ and $b$:

$$E'(\mu) = a\mu^{-2} + b$$  \hspace{1cm} (46)

which describes the main properties of the function $E'(\mu)$. \hspace{1cm} \Box

As expected, this model function is decreasing and the value function $E(\mu)$ is concave. This model function yields an approximate value, $\mu_{approx}$, of the regularization parameter obtained with the equation:

$$2E'_{approx}(\mu) = \delta^2$$  \hspace{1cm} (47)

In a second step, we define the function $G(\mu)$ by:

$$G(\mu) = 2E'(\mu) - \delta^2$$  \hspace{1cm} (48)
In order to solve $G(\mu) = 0$, given two initial guess values close to $\mu_{\text{approx}}$, $\mu_0$ and $\mu_1$, a sequence of iterates $(\mu_k)_{k \geq 0}$ is generated for $k \geq 1$ by:

$$\mu_{k+1} = \mu_k - \frac{2E'(\mu_k) - \delta^2}{E''(\mu_k)}$$

(49)

where the value $f'(\mu_k)$ in $E''(\mu_k)$ is approximated with $\frac{f(\mu_k) - f(\mu_{k-1})}{\mu_k - \mu_{k-1}}$.

### 3.3.2. Iterative scheme for the damped Morozov principle.

With the damped Morozov principle [22, 23], we have to consider the modified equation:

$$\|g^\delta - Af(\mu)\|^2 + \mu \gamma R_{TV}(f(\mu)) = \delta^2$$

(50)

for a parameter $\gamma$. The first step of the iterative scheme is the same. For the Newton refinement step, we can redefine the function $G(\mu)$ as:

$$G(\mu) = \|g^\delta - Af(\mu)\|^2 + \mu \gamma R_{TV}(f(\mu)) - \delta^2$$

(51)

From an approximate value calculated with the model function of the classical Morozov principle, the optimal regularization parameter can be obtained with the Newton method by defining a sequence $(\mu_k)_{k \geq 0}$, given $\mu_0$ and $\mu_1$, two initial guess values close to $\mu_{\text{approx}}$ and applying:

$$\mu_{k+1} = \mu_k - \frac{G(\mu_k)}{G'(\mu_k)}$$

(52)

Taking into account the differentiability of the value function $E(\mu)$ for a smoothed version of the gradient, it is possible to demonstrate that $G(\mu)$ is decreasing for some values of $\mu$ and $\gamma$:

**Proposition 3.5.** The function $\mu \rightarrow G(\mu)$ is strictly decreasing for $\gamma = -1$ and $\mu > 0$ and for $\gamma < -1$ and $\mu > 2^{1/(\gamma+1)}$.

**Proof.** With the value function $E(\mu) = \frac{\mu \|g^\delta - Af(\mu)\|^2}{2} + R_{TV}(f(\mu))$, and $E'(\mu)$ given by (39), $G(\mu)$ can be rewritten as:

$$G(\mu) = 2E'(\mu)(1 - \frac{\mu^{\gamma+1}}{2}) + E(\mu)\mu^\gamma - \delta^2$$

(53)

The first derivative of $G(\mu)$ is given by:

$$G'(\mu) = 2E''(\mu)(1 - \frac{\mu^{\gamma+1}}{2}) + \gamma \mu^{\gamma-1}R_{TV}(f(\mu))$$

(54)

If $\gamma$ is large and negative, and $\mu$ large enough, $G'(\mu) = 2E''(\mu)$, and the exact Morozov principle is recovered. We know that $E''(\mu) \leq 0$ and that $E'(\mu)$ is decreasing. It follows from (54), that for $\gamma = -1$, the function $G$ is strictly decreasing. For $\gamma < -1$ and $\mu > 2^{1/(\gamma+1)}$, $G'(\mu) < 0$ and the function $G$ is also strictly decreasing. □

This result implies that, if it exists, the solution to $G(\mu) = 0$ that solves the damped Morozov principle is unique.
4. **UPRE method for Total Variation parameter selection.** In this section, we first briefly summarize the Unbiased Predictive Risk Estimator (UPRE) method for the TV regularization that will be used for comparison with our iterative methods. The UPRE approach has been studied in detail for optimal parameter selection for Tikhonov problems [38, 8, 12, 28, 37]. The predictive error for the parameter \( \mu \) is defined as \( p(\mu) = Af(\mu) - Af^* \). In the framework of the UPRE method, the optimal parameter \( \mu \) is the minimizer of the predictive risk \( \frac{\|p(\mu)\|^2}{N^2} \), where \( N^2 \) is the total number of pixels in the image. Recently, the UPRE method has been extended to TV regularization and to large-scale problems [25]. As detailed in [25], the optimal parameter is the minimizer of the Unbiased Predictive Risk Estimator which has the same functional form as the one for the Tikhonov regularization:

**Theorem 4.1.** The Unbiased Predictive Risk Estimator of the predictive risk for TV regularization is given by:

\[
UPRE(\mu) = \frac{\|p(\mu)\|^2}{N^2} + \frac{2\sigma^2}{N^2} \text{tr}(C_{TV}(\mu)) - \sigma^2
\]

where \( \sigma^2 \) is the noise variance and \( \text{tr}(\cdot) \) denotes the trace operator. The influence matrix \( C_{TV}(\mu) \) in the TV problem can be written as:

\[
C_{TV}(\mu) = A(\mu A^t A + L(f(\mu)))^{-1} A^t
\]

where the matrix \( L(f(\mu)) \) is given by:

\[
L(f(\mu)) = D_x^t \text{diag}(\frac{1}{\|\text{grad}(f(\mu))\|_2}) D_x + D_y^t \text{diag}(\frac{1}{\|\text{grad}(f(\mu))\|_2}) D_y
\]

In order to avoid divergencies in the numerical tests, the norm of the gradient of the image at pixel \( i \) is approximated by: \( \sqrt{(D_x^t(f))^2 + (D_y^t(f))^2 + \epsilon} \) where \( \epsilon \) is a smoothing constant.

For large scale problems the Singular Value Decomposition of \( A \) is not available and it is necessary to approximate the trace value \( \text{tr}(C_{TV}(\mu)) \) with an algorithm requiring only matrix-vector products. With the approach of Hutchinson [20], the trace of a smooth function \( F(M) \) of a symmetric positive definite matrix \( M \) can be evaluated with the unbiased trace estimator: \( E(u^t F(M) u) = \text{tr}(F(M)) \) where \( u \) is white noise vector with standard deviation equal to \( \sigma_u = 1 \) and \( E \) denotes the expected value function. The choice of \( u \) that minimizes the variance in this estimator is the one for which the components of the vector are independent and take on the values of 1 and -1 with equal probability [38]. The expectation value is evaluated with a number \( n_u \) of realization of the random variable. In order to approximate the trace of \( C_{TV}(\mu) \), we have to solve the linear system: \( (\mu A^t A + L(f(\mu)))^{-1} A^t u = y \) for a given vector \( y \) which can be done with a conjugate gradient algorithm. In this work, in order to obtain the best reconstruction results, the \( UPRE(N_{iter}, \mu) \) was calculated as a function of the number of iterates \( N_{iter} \) and as a function of the regularization parameter \( \mu \).

5. **Numerical experiments.** In order to show the effectiveness of our method for choosing the regularization parameter, we have considered two test problems: deconvolution and super-resolution. In the deconvolution case, the test image is blurred with a Gaussian kernel of standard deviation \( \sigma_{blur} \) and corrupted by additive Gaussian noise of standard deviation \( \sigma \). In the super-resolution case, after the blur and the noise are applied, the image is also undersampled by a factor \( p \), increasing the ill-posedness of the problem.

---

1182 Alina Toma, Bruno Sixou and Françoise Peyrin
5.1. Practical implementation of the classical and damped Morozov methods. We start by finding the optimal regularization parameter that maximizes the PSNR of the resulting images defined as:

$$PSNR(\mu) = 10 \log_{10} \left( \frac{N^2(\max(f^*) - \min(f^*))^2}{\|f^* - f(\mu)\|^2} \right)$$

where $f^*$, $f(\mu)$ represent the ground-truth image and the reconstructed image for the regularization parameter $\mu$ respectively and $N^2$ is the total number of pixels. Thus, an extensive sweeping of the regularization parameter values is performed. The deteriorated image is restored with the ADMM algorithm for each value of the regularization parameter $\mu$. The ADMM iterations are stopped when $\|f^{k+1} - f^*\|_2 < \text{stop}_{\text{ADMM}}$, where $\text{stop}_{\text{ADMM}}$ is a small positive constant.

In order to apply the classical and damped Morozov schemes described in section 3.3, we need to determine approximate values of the parameters $a$ and $b$ of the model function $E'(\mu)$, by first choosing two values $\mu_*$ and $\mu_{**}$ with $\mu_* < \mu_{**}$. For each of these regularization parameters, $\mu_*$ and $\mu_{**}$, the blurred noisy image is restored with the ADMM algorithm and the values $E'(\mu_*)$ and $E'(\mu_{**})$ are calculated. If $2E'(\mu_*) \geq \delta^2$, $\mu_*$ is not changed, otherwise it is decreased. If $2E'(\mu_{**}) \leq \delta^2$, $\mu_{**}$ is not changed, otherwise it is increased. An approximate value for the regularization parameter, $\mu_{\text{approx}}$, is obtained from the equation $\delta^2 = 2(a(\mu_{\text{approx}})^{-2} + b)$, where the noise level $\delta$ is evaluated with the standard deviation $\sigma$ as $\delta^2 = N^2\sigma^2$. Starting from this approximate solution, the Newton method is then applied to refine the solution. Considering $\mu_0 = \mu_{\text{approx}} + 0.1$ and $\mu_1 = \mu_{\text{approx}}$, the iterative formula given by (49) or (52) are applied for the classical or damped Morozov principles. For each iteration of the Newton method, the ADMM algorithm is applied to obtain the solution $f(\mu)$ with the same value $\text{stop}_{\text{ADMM}}$. The Newton iterations are stopped when $\| Af(\mu) - g \|_2^2 < \text{stop}_{\text{Newton}}$.

5.2. Deconvolution problem. The test image $f \in \mathbb{R}^{N^2}$ ($N=330$) to be reconstructed is displayed in Figure 1.(a). It is a slice of a bone synchrotron micro-CT sample at a voxel size of 20 $\mu$m [33]. Trabecular bone is organized in a complex network made of struts of about 100 $\mu$m (white patterns in the image), requiring images at high spatial resolution that are still difficult to acquire in-vivo.

We first consider the application of the methods to the noisy and blurred image shown in Figure 1.(b) where $\sigma_{\text{blur}} = 1.2$ and $\sigma = 4.3$. The noise level $\delta$ is evaluated as $\delta^2 = N^2\sigma^2 = 2 \cdot 10^6$. The noisy blurred image has a PSNR=24.89. The evolution of the PSNR of the image restored with the ADMM algorithm as a function of $\mu$ is displayed on Figure 2.(b). From this curve, the optimal value of the regularization parameter is $\mu_{\text{OPT}} = 0.7$ and the optimal PSNR of the reconstructed image is $PSNR(\mu_{\text{OPT}}) = 32.32$. A value $\text{stop}_{\text{ADMM}} = 5 \cdot 10^{-5}$ was considered to ensure the convergence of the regularization method.

We applied the Morozov principles for selecting the regularization parameter as described in section 5.1. With $\mu_* = 0.1$ and $\mu_{**} = 10$, the values obtained with (46) for $a$ and $b$ are $5.36 \cdot 10^3$ and $7.26 \cdot 10^5$ respectively. The approximate value for the regularization parameter was found as $\mu_{\text{approx}}=0.14$. By applying the Newton scheme for the classical Morozov principle, we observed that the convergence was fast and only four iterations were necessary. A value $\text{stop}_{\text{Newton}} = 0.001$ was considered. The final value of the regularization parameter was found to be $\mu_M = 0.34$. The final image obtained with $\mu_M$ is displayed in Figure 1.(c). Its PSNR value

Inverse Problems and Imaging Volume 9, No. 4 (2015), 1171–1191
Figure 1. (a). Ground truth - Bone image; (b). Blurred and noisy image ($\sigma = 4.3, \sigma_{\text{blur}} = 1.2.$); (c). Restored image with $\mu_M$; (d). Restored image with $\mu_{\text{UPRE}}$.

Figure 2. Deconvolution applied to the bone image for $\sigma = 4.3$ and $\sigma_{\text{blur}} = 1.2$: (a). The evolution of the data term for the classical Morozov method; (c). The evolution of the modified data term for the damped Morozov method; (b). and (d). The evolution of PSNR and UPRE as function of the regularization parameter $\mu$.

is $PSNR(\mu_M) = 31.71$ and it is close to the optimal value $PSNR(\mu_{\text{OPT}}) = 32.32$. The Figure 2.(a) displays the evolution of the data term $\|Af(\mu) - g^\delta\|^2$ as a function of the regularization parameter $\mu$. The model function, the noise level and the iterates $\mu_k$ have been reported on the same Figure.

The damped Morozov method with $\gamma = -1$, was then applied by following the same scheme. A value $\mu_{DM} = 1.64$ was obtained after six Newton iterations (Figure 2.(c)), with a PSNR value equal to 31.43.

For comparison, the UPRE method was also implemented for selecting the regularization parameter. The trace of the influence matrix was calculated with 50 white
noise vectors \( u \). The smallest \( \epsilon \) value that gives a positive UPRE curve was chosen. The UPRE curves are rather noisy thus the mean and the standard deviation of \( \mu \) and PSNR where estimated from five UPRE curves for the chosen \( \epsilon \). The evolution of the UPRE(\( \mu \)) for the value of the smoothing parameter \( \epsilon = 0.05 \) as a function of the regularization parameter is displayed in Figure 2.(d) for \( \text{stop}_{\text{ADMM}} = 5 \times 10^{-5} \).

The UPRE(\( N_{\text{iter}}, \mu \)) was also calculated as a function of the number of ADMM iterations, \( N_{\text{iter}} \), for different \( \mu \) values. Figure 3 shows the evolution of UPRE as a function of \( N_{\text{iter}} \) and \( \mu \).

**Figure 3.** Evolution of UPRE as a function of \( N_{\text{iter}} \) and \( \mu \) for deconvolution case, \( \sigma = 4.3, \sigma_{\text{blur}} = 1.2 \) - Bone image.

The best mean regularization parameter obtained is \( \mu_{\text{UPRE}} = 1.44 \) with a PSNR=31.73 and the image obtained is displayed in Figure 1.(d).

In order to have a more extensive comparison of the methods of choice of the regularization parameter, we have tested several noise and blur levels. The values of \( \mu_{M}, \mu_{DM} \) with \( \gamma = -1 \), \( \mu_{\text{UPRE}}, \mu_{\text{OPT}} \) and the PSNR values are summarized in Table 1. For low blur, the classical Morozov principle is better than the damped one. On the contrary, for high blur level, the damped Morozov principle improves the results obtained with the classical Morozov principle. The UPRE method is efficient except for the highest noise level case.

| Method            | \( \sigma = 4.3, \sigma_{\text{blur}} = 1.2 \) | \( \sigma = 6, \sigma_{\text{blur}} = 1.2 \) | \( \sigma = 4.3, \sigma_{\text{blur}} = 3.6 \) |
|-------------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|
| \( \mu \)         | PSNR                                         | \( \mu \)                                     | PSNR                                         |
| classical Morozov | 0.34                                         | 0.24                                         | 0.95                                         |
| damped Morozov    | 1.64                                         | 1.02                                         | 5.12                                         |
| \( \text{stop}_{\text{ADMM}} \) | 1.46 \( \pm 0.11 \) | 1.02 \( \pm 0.08 \) | 5.64 \( \pm 0.02 \) |
| \( \text{UPRE}(\mu) \) | \( \mu \)                                         | PSNR                                         | \( \mu \)                                     | PSNR                                         |
| UPRE(\( N_{\text{iter}}, \mu \)) | 0.70 \( \pm 0.08 \) | 0.46 \( \pm 0.06 \) | 4.71 \( \pm 0.02 \) |
| optimal           | 0.70                                         | 0.46                                         | 4.71                                         |

**Table 1.** Comparison of regularization parameter selection methods for deconvolution case - Bone image.
5.3. Super-resolution problem. In this section we investigate a more ill-posed problem, the super-resolution problem. The standard deviation of the Gaussian blurring kernel was $\sigma_{\text{blur}} = 4.85$ and a sub-sampling rate of $p = 4$ was applied to the images. An additional Gaussian noise with a standard deviation $\sigma$ was added to the low-resolution image. The aim of the regularization approach is to estimate the high resolution image.

Simulations were performed with several noise levels. To illustrate the method, we first detail the results obtained with $\sigma = 6$. Figure 4.(a) and 4.(b) display the high resolution and low resolution images. Figure 5.(b) displays the evolution of the PSNR for $\sigma = 6$. The best PSNR is 19.57. In this case, the optimal regularization parameter is $\mu_{\text{OPT}} = 6.2$. The regularization parameter estimated with the classical Morozov principle is $\mu_{\text{M}} = 2.22$, yielding a PSNR of 19.07, far from the optimal value. The evolution of the data term for this case is displayed in Figure 5.(a).

When applying the damped Morozov method with $\gamma = -1$, the obtained regularization parameter is $\mu_{\text{DM}} = 6.2$. The evolution of the data term weighted with the TV norm of the image $f$ is shown in Figure 5.(c) as a function of the regularization parameter. The corresponding PSNR value is 19.57 which is very close to the optimal one. We obtain thus a very good estimation of the optimal regularization parameter.

For comparison, the evolution of the UPRE($\mu$) term for $\sigma = 6$ with $\epsilon = 0.05$ is displayed in Figure 5.(d). The regularization parameter obtained is $\mu_{\text{UPRE}} = 13.48$ with a PSNR of 19.14, slightly lower than with the damped Morozov method. The images obtained with this two methods are displayed in Figures 4.(c)-(d).

Table 2 summarizes the values of the regularization parameters and the PSNR obtained for different noise levels. For all these noise levels, the UPRE method gives a better estimation of the optimal regularization parameter than the classical Morozov method. Yet, the damped Morozov method is more effective than the UPRE approach for this super-resolution problem.

5.4. Super-resolution test problem for other images. To assess the potential of the proposed method in other domains of application, we made additional experiments on the usual Cameraman ($N=512$) and Lena ($N=512$) images. The simulation parameters of the super-resolution problem where: $p=4$, $\sigma_{\text{blur}} = 4.85$ and two noise levels $\sigma = 4.3$ and $\sigma = 6$.

The results are summarized in Table 3 and Table 4. Resulting images are displayed in Figures 6 - 9. In most cases, the UPRE method provides better results.
Figure 5. Super-resolution applied to the bone image for $p = 4$, $\sigma = 6$ and $\sigma_{blur} = 4.85$: (a). The evolution of the data term for the classical Morozov method; (c). The evolution of the modified data term for the damped Morozov method; (b) and (d). The evolution of PSNR and UPRE as function of the regularization parameter $\mu$.

| Method          | $\mu$ | PSNR | $\mu$ | PSNR | $\mu$ | PSNR |
|-----------------|-------|------|-------|------|-------|------|
| classical Morozov | 25.43 | 21.38 | 3.45  | 19.49 | 2.22  | 19.07 |
| damped Morozov  | 141.5 | 22.28 | 10.94 | 20.06 | 6.2   | 19.57 |
| UPRE($\mu$)     | 198.4 | 22.25 | 18.26 | 19.93 | 13.2  | 19.17 |
| UPRE($N_{iter}, \mu$) | 199.6 | 22.25 | 18.36 | 19.92 | 13.48 | 19.14 |
| optimal         | 158.7 | 22.28 | 11.5  | 20.06 | 6.2   | 19.57 |

Table 2. Comparison of regularization parameter selection methods for super-resolution case - Bone image.

than the classical Morozov method. Yet, the damped Morozov method leads to a more accurate estimation of the optimal regularization parameter. It should be emphasized that the methods developed with the classical and damped Morozov principles are very fast and does not rely on an extensive sweeping of the regularization parameter values.

6. Conclusion. In this paper, we have proposed new iterative methods to determine the optimal regularization parameter in linear inverse problems with TV regularization. For a given regularization parameter, the TV regularization functional is minimized with the classical ADMM algorithm. Our approach is based on the classical Morozov principle or on a damped version of this principle. The convergence properties of the regularization method for this choice of the regularization parameter have been demonstrated. Based on the regularity properties of
Table 3. Comparison of regularization parameter selection methods for super-resolution case - Cameraman image.

| Method             | $\sigma = 4.3$ | $\sigma = 6$ |
|--------------------|----------------|--------------|
|                    | $\mu$ | PSNR | $\mu$ | PSNR |
| classical Morozov  | 3.15  | 23.74| 1.95  | 23.38 |
| damped Morozov     | 13.34 | 24.11| 7.58  | 23.72 |
| UPRE($\mu$)        | 19.91 | 23.99±0.05 | 12.24 | 23.50±0.06 |
| UPRE($N_{iter}$, $\mu$) | 18.94 | 24.01±0.04 | 10.91 | 23.58±0.06 |
| optimal            | 10.97 | 24.14| 5.83  | 23.74 |

Table 4. Comparison of regularization parameter selection methods for super-resolution case - Lena image.

| Method             | $\sigma = 4.3$ | $\sigma = 6$ |
|--------------------|----------------|--------------|
|                    | $\mu$ | PSNR | $\mu$ | PSNR |
| classical Morozov  | 2.48  | 23.51| 1.73  | 23.23 |
| damped Morozov     | 13.11 | 23.77| 7.89  | 23.42 |
| UPRE($\mu$)        | 17.55 | 23.63±0.03 | 11.33 | 23.17±0.18 |
| UPRE($N_{iter}$, $\mu$) | 17.19 | 23.65±0.04 | 11.34 | 23.18±0.15 |
| optimal            | 8.85  | 23.83| 4.63  | 23.53 |

the value function of the Augmented Lagrangian, it was shown that the data term can be approximated by a simple model function. This model function for the
data term of the regularization functional is first obtained from the reconstructed images with two different regularization parameters. An approximate value of the regularization parameter is calculated with the model function approximating the discrepancy term. Then the solution is refined with a Newton iterative method on the basis of the Morozov principle or its damped version. This new approach has been compared with the one based on the UPRE. The methods have been compared for deconvolution and super-resolution test problems for various noise and blur levels and for different images.

For the deconvolution case, the damped Morozov principle is the most efficient method for the highest blur level. Otherwise, the classical Morozov principle offers the best estimation of the optimal regularization parameter. For the super-resolution case, the UPRE method outperforms the classical Morozov method. Yet, a more accurate estimation of the optimal regularization parameter is obtained with the damped Morozov principle, where the data term is weighted by a TV term. The iterative schemes based on the new model function presented in this work are very fast and avoid an extensive sweeping of the values of the regularization parameter.

Acknowledgments. This work was supported by the LABEX PRIMES (ANR-11-LABX-0063) of Université de Lyon, within the program “Investissements d’Avenir” (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR).

REFERENCES

[1] M. Afonso, J. Bioucas-Dias and M. Figueiredo, Fast image recovery using variable splitting and constrained optimization, IEEE Trans. Image Process., 19 (2010), 2345–2356.
[2] M. Afonso, J. Bioucas-Dias and M. Figueiredo, An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems, *IEEE Trans. Image Process.*, 20 (2011), 681–695.

[3] L. Ambrosio, N. Fusco and D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems*, Clarendon Press Oxford, 2000.

[4] J. F. Aujol and G. Gilboa, Constrained and SNR-based solutions for TV-Hibert space image denoising, *J. Math. Imaging Vis.*, 26 (2006), 217–237.

[5] S. D. Babacan, R. Molina and A. K. Katsaggelos, Parameter estimation in TV image restoration using variational distribution approximation, *IEEE Trans. Image Process.*, 17 (2008), 326–339.

[6] S. D. Babacan, R. Molina and A. K. Katsaggelos, Total Variation Super Resolution Using A Variational Approach, *Proc. Int. Conf. Image Process.*, (2008), 641–644.

[7] S. Becker, J. Bobin and E. Candes, NESTA: A fast and accurate first-order method for sparse recovery, *SIAM Journal on Imaging Sciences*, 4 (2011), 1–39.

[8] T. Blu and F. Luisier, The SURE-LET approach to image denoising, *IEEE Trans. Image Process.*, 16 (2007), 2778–2786.

[9] L. M. Bregman, The relaxation method for finding common points of convex sets and its application to the solution of problems in convex programming, *USSR Comput. Math. Math+*, 7 (1967), 200–217.

[10] M. Burger and S. Osher, Convergence rates of convex variational regularization, *Inverse Probl.*, 20 (2004), 1411–1421.

[11] C. A. Deledalle, S. Vaiter, J. Fadili and G. Peyre, Stein Unbiased Gradient estimator for the risk (SUGAR) for multiple parameter selection, *SIAM Journal on Imaging Sciences*, 7 (2014), 2448–2487.

[12] Y. C. Eldar, Generalized SURE for exponential families: Application to regularization, *IEEE Trans. Signal Process.*, 57 (2009), 471–481.

[13] H. W. Engl, M. Hanke and A. Neubauer, *Regularization of Inverse Problems*, Mathematics and its Applications, 375, Kluwer Academic Publishers Group, Dordrecht, 1996.

[14] E. Esser, Applications of Lagrangian-based alternating direction methods and connections to split Bregman, CAM report, 2009.

[15] J. M. Fadili, G. Peyre, S. Vaiter, C. Deledalle and J. Salmon, Stable recovery with analysis decomposable priors, preprint, arXiv:1304.4407.

[16] K. Frick, D. A. Lorenz and E. Resmerita, Morozov’s principle for the augmented Lagrangian method applied to linear inverse problems, *Multiscale Model Simul.*, 9 (2011), 1528–1548.

[17] G. H. Golub, M. T. Heath and C. Wahba, Generalized cross-validation as a method for choosing a good ridge parameter, *Technometrics*, 21 (1979), 215–223.

[18] M. Grasmair, Linear convergence rates for Tikhonov regularization with positively homogeneous functionals, *Inverse Problems*, 27 (2011), 075014, 16pp.

[19] P. C. Hansen and D. P. O’Leary, The use of the L-curve in the regularization of discrete ill-posed problem, *SIAM Journal on Scientific Computing*, 14 (1993), 1487–1503.

[20] M. F. Hutchinson, A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines, *Commun. Stat. Simulat.*, 19 (1990), 433–450.

[21] K. Kiwiel, Proximal minimization methods with generalized Bregman functions, *SIAM J. Control Optim.*, 35 (1997), 1142–1168.

[22] K. Kunisch and J. Zou, Iterative choices of regularization parameters in linear inverse problems, *Inverse Probl.*, 14 (1998), 1247–1264.

[23] K. Kunisch, On a class of damped Morozov principles, *Computing*, 50 (1993), 185–198.

[24] H. Liao, F. Li and M. K. Ng, Selection of regularization parameter in total variation restoration, *JOSA A*, 26 (2009), 2311–2320.

[25] Y.Lin, B.Wohlberg and H.Guo, UPRE method for total variation parameter selection, *Signal Processing*, 90 (2010), 2546–2551.

[26] V. A. Morozov, *Methods for Solving Incorrectly Posed Problems*, Springer-Verlag, New York, 1984.

[27] M. K. Ng, P. Weiss and X. Yuan, Solving constrained total-variation image restoration and reconstruction problems via alternating direction methods, *SIAM Journal on Scientific Computing*, 32 (2010), 2170–2176.

[28] S. Ramani, T. Blu and M. Unser, Monte-Carlo SURE: A black-box optimization of regularization parameters for general denoising algorithms, *IEEE Trans. Image Process.*, 17 (2008), 1540–1554.

Inverse Problems and Imaging Volume 9, No. 4 (2015), 1171–1191
[29] S. Ramani, Z. Liu, J. Rosen, J. Nielsen and J. A. Fessler Regularization parameter selection for nonlinear iterative image restoration and MRI reconstruction using GCV and SURE-based methods, IEEE Transactions on Image Processing, 21 (2012), 3659–3672.

[30] Z. Ren, C. He and Q. Zhang, Fractional order total variation regularization for image super-resolution, Signal Process., 93 (2013), 2408–2421.

[31] E. Resmerita, Regularization of ill-posed problems in Banach spaces: Convergence rates, Inverse Probl., 21 (2005), 1303–1314.

[32] L. I. Rudin, S. Osher and E. Fatemi, Nonlinear total variation based noise removal algorithms, Phys. D, 60 (1992), 259–268.

[33] M. Salome, F. Peyrin, P. Cloetens, C. Odet, A. M. Laval-Jeantet, J. Baruchel and P. Spanne, A synchrotron radiation microtomography system for the analysis of trabecular bone samples, Medical Physics, 26 (1999), 2194–2204.

[34] O. Scherzer, M. Grassmair, H. Grossauer, M. Haltmaier and F. Lenzen, Variational Methods in Imaging, Springer Verlag, New York, 2008.

[35] A. N. Tikhonov and V. Y. Arsenin, Solutions to ill-posed problems, Winston-Wiley, New York, 1977.

[36] E. Van den Berg E and M. P. Friedlander, Probing the Pareto frontier for basis pursuit solutions, SIAM Journal on Scientific Computing, 31 (2008), 890–912.

[37] S. Vaiter, A. Deledalle, G. Peyre, C. Dossal and J. Fadili, Local behavior of sparse analysis regularization: Applications to risk estimation, Applied and Computational Harmonic Analysis, 35 (2013), 433–451.

[38] C. R. Vogel, Computational Methods for Inverse Problems, SIAM, Philadelphia, PA, 2002.

[39] C. Vonesch, S. Ramani and M. Unser, Risk estimation for non-linear image deconvolution with a wavelet-domain sparsity constraint, Inverse Problems, 27 (2011), 075014.

[40] C. Vonesch, S. Ramani and M. Unser, Recursive risk estimation for non-linear image deconvolution with a wavelet-domain sparsity constraint, ICIP 2008, 15th IEEE International Conference on Image Processing, (2008), 665–668.

[41] Y. Wang, J. Yang, W. Yin and Y. Zhang, A new alternating minimization algorithm for total variation image reconstruction, SIAM Journal on Imaging Sciences, 1 (2008), 248–272.

[42] Y. W. Wen and R. H. Chan, Parameter selection for total-variation-based image restoration using discrepancy principle, IEEE Trans. Image Process., 21 (2012), 1770–1781.

[43] J. Yang, W. Yin, Y. Zhang and Y. Wang, A fast algorithm for edge-preserving variational multichannel image restoration, SIAM Journal on Imaging Sciences, 2 (2009), 569–592.

Received June 2014; revised April 2015.

E-mail address: alina.toma@creatis.insa-lyon.fr
E-mail address: bruno.sixou@insa-lyon.fr
E-mail address: peyrin@esrf.fr