Iterative choice of the optimal regularization parameter in TV image deconvolution

B Sixou¹, A Toma¹, L Denis² and F Peyrin¹,³

¹ CREATIS, INSA-Lyon; Inserm, U1044; CNRS UMR 5220; Université Lyon 1; Université de Lyon; 69621, Lyon, France
² LaHC, CNRS UMR 5516, Université de Saint-Etienne, 42000 Saint-Etienne, France
³ ESRF, Imaging Group, 38043 Grenoble Cedex, France

E-mail: bruno.sixou@insa-lyon.fr, alina.toma@creatis.insa-lyon.fr, loic.denis@univ-st-etienne.fr, peyrin@esrf.fr

Abstract. We present an iterative method for choosing the optimal regularization parameter for the linear inverse problem of Total Variation image deconvolution. This approach is based on the Morozov discrepancy principle and on an exponential model function for the data term. The Total Variation image deconvolution is performed with the Alternating Direction Method of Multipliers (ADMM). With a smoothed $l_2$ norm, the differentiability of the value of the Lagrangian at the saddle point can be shown and an approximate model function obtained. The choice of the optimal parameter can be refined with a Newton method. The efficiency of the method is demonstrated on a blurred and noisy bone CT cross section.

1. Introduction

Linear inverse problems play a major role in image processing applications. For example, image deconvolution and reconstruction problems are encountered in medical imaging or image and video coding. Such problems can be described by the following linear relationship:

$$g^\delta = Af + w,$$

where $g^\delta$ represents the measured data for the noise level $\delta$, $f$ the ideal image, $w$ an additive noise corrupting the measurements and $A$ the blurring kernel. To solve this ill-posed problem and to obtain stable solutions in the presence of noise, a possible 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. In the classical Tikhonov regularization, $R(f) = \|Df\|^2$, where $D$ is a differential operator. More recently, the Total Variation regularization (TV) [1, 2] has been studied and applied to various image processing problems [3, 4]. For an image $f \in H_1(\Omega)$, this regularization is given by:

$$R_{TV}(f) = \int_\Omega |\nabla f(r)|dr$$

(2)

This regularization is known to preserve well edges. Various numerical methods have been used to solved the TV regularized deconvolution problem [5, 6].
show that algorithms based on the Alternating Direction Method of Multipliers (ADMM) are the state-of-the-art methods [3, 4, 7, 8]. 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 $\delta$ [2, 9]. The L-curve method and the Generalized Cross-Validation can be used without the knowledge of this noise level [10, 11]. Moreover, an iterative choice of the regularization parameters based on the Morozov discrepancy principle applied with a model function has been proposed in [12].

For TV regularization, fewer methods have been investigated. For denoising problems, Aujol et al. [13] have proposed a method where the denoised result is closed to the optimal, in the SNR sense. An extension of the Unbiased Predictive Risk Estimator has been shown to give a good estimate of the regularization parameter [14]. A method exploiting the Generalized Cross-Validation technique has been proposed in [15]. Babacan et al. [16] have studied a bayesian variational method for solving TV deconvolution 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 [17].

The goal of this paper is to propose a fast and simple method to select the optimal regularization parameter for the restoration of an image $f$ from a blurred and noisy image $g^\delta$ with TV regularization. Our contribution in this field is to design an iterative method following the idea of the model function of Kunish et al. [12]. The outline of this paper is the following. In Section 2, we present the linear inverse problem studied and the ADMM/TV algorithm chosen. In Section 3, the new method for the selection of the regularization parameter is presented. In Section 4, numerical results are given to demonstrate the effectiveness of the approach. Concluding remarks are given in Section 5.

2. Problem formulation, TV regularization and ADMM approach

2.1. Problem formulation

In this paper, we consider a linear problem of the form $g^\delta = Af + w$, where $A$ is a linear and bounded convolution operator. We assume that the images $f$ and $g^\delta$ are defined on a bounded open subset $\Omega \subset \mathbb{R}^2$. The noisy data for the noise level $\delta$ is denoted as $g^\delta$. This problem is ill-posed and to obtain a stable solution, we use a total variation regularization. The recent ADMM methodology is the state-of-the-art method to obtain a stable solution with the TV regularization [3, 4].

The convergence rates for the Augmented Lagrangian method when Morozov's discrepancy principle is chosen as stopping rule have been investigated in [18]. These type of algorithms have attracted much attention and have also been proposed to solve a number of image processing tasks, such as image inpainting and deblurring. In the framework of the ADMM method, an augmented Lagrangian is considered given by:

$$
\mathcal{L}(f, g, \lambda) = \sum ||g_i||_2 - \lambda_i (g_i - D_i f) + \frac{\beta}{2} ||g_i - \lambda_i||_2^2 + \frac{\mu}{2} ||g^\delta - Af||_L^2
$$

where $\mu$ is the regularization parameter, $\beta$ the Lagrangian parameter, and $||.||_2$ is the Euclidean norm in $\mathbb{R}^2$. For the ith pixel at the location $(j, k)$, the first order finite difference operator $D_i$ is defined by $D_i : \mathbb{R}^n \rightarrow \mathbb{R}^2$, $D_i(f) = (D^x_i(f), D^y_i(f))$, where $D^x_i$ and $D^y_i$ are the classical horizontal and vertical difference operators. In order to provide differentiability at the origin, the $l_2$ norm of $g_i = (g_i^1, g_i^2)$ is approximated by $\sqrt{(g_i^1)^2 + (g_i^2)^2 + \xi}$. The dual variable associated to the constraint $g_i = D_i(f)$ is denoted $\lambda_i \in \mathbb{R}^2$.  

3rd International Workshop on New Computational Methods for Inverse Problems IOP Publishing
Journal of Physics: Conference Series 464 (2013) 012005 doi:10.1088/1742-6596/464/1/012005
2.2. The ADMM algorithm and optimality conditions
The ADMM algorithm search for the saddle point of the augmented Lagrangian by iterating the following equations:

\[
\begin{align*}
    f^{k+1} &= \arg\min_x L(x, g^k, \lambda^k) \\
    g^{k+1}_i &= \arg\min_y L(f^{k+1}, y, \lambda^k) \\
    \lambda^{k+1}_i &= \lambda^k_i - \beta (g^{k+1}_i - D_i f^{k+1})
\end{align*}
\]  

(4)

In this work, we have used the isotropic TV and the $L_2$ norm of the gradient. With the alternating minimization algorithm, three sequences $(f^k, g^k, \lambda^k)$ are constructed with the following iterative scheme: For each pixel \(i\)

\[
g^{k+1}_i = \max\{\|D_i f^k + \frac{1}{\beta} (\lambda^k_i) - \frac{1}{\beta} 0\|, 0\} \frac{D_i f^k + \frac{1}{\beta} (\lambda^k_i)}{\|D_i f^k + \frac{1}{\beta} (\lambda^k_i)\|}
\]  

(5)

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

\[
(\sum D_i^t D_i + \frac{\mu}{\beta} A^t A) f = \sum D_i^t (g^{k+1}_i - \frac{1}{\beta} \lambda^k_i) + \frac{\mu}{\beta} A^t g^k.
\]  

(6)

The Lagrange multipliers are updated with, \(\lambda^{k+1}_i = \lambda^k_i - \beta (g^{k+1}_i - D_i f^{k+1})\). The sequence $(f^k, g^k, \lambda^k)$ generated by the ADMM algorithm converges to a Kuhn-Tucker point $(f^*, g^*_i, \lambda^*_i)$ of the problem, which correspond to a saddle point of the Lagrangian. With the augmented Lagrangian given in (3), the Karush-Kuhn-Tucker (KKT) conditions can be written as:

\[
\mu A^t (Af(\mu) - g^*) = -\sum D_i^t (\lambda^*_i), \quad \lambda^*_i \in \partial \|D_i f^*(\mu)\|_2, \quad g^*_i = D_i f^*
\]  

(7)

3. Iterative choice of the regularization parameter
In the following, we assume that the solution $f^*(\mu)$ obtained with the ADMM algorithm is unique. We first show the differentiability of $f^*(\mu)$ and of the value function $E(\mu)$ of the Lagrangian at the saddle point.

3.1. Lipschitz continuity and differentiability of $f^*(\mu)$
We first show that $f^*(\mu)$ is Lipschitz continuous. Subtracting the KKT condition for $\mu + t$ and $\mu$ yields:

\[
\mu A^t A(f^*(\mu + t) - f^*(\mu)) \in t A^t A f^*(\mu + t) + t A^t g^\delta - \sum D_i (\partial \|D_i f^*(\mu + t)\|_2 - \partial \|D_i f^*(\mu)\|_2).
\]  

(8)

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:

\[
\mu \|A(f^*(\mu + t) - f^*(\mu))\|^2 \leq C \|f^*(\mu + t) - f^*(\mu)\|_2
\]  

(9)

where $C$ is a constant. 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. We now show that the function $f^*(\mu)$ is differentiable and that its derivative is the unique solution $\omega$ of:

\[
\mu A^t A \omega + \sum D_i \frac{D_i \omega}{\|D_i f^*(\mu)\|} = A^t g^\delta - A^t A f^*(\mu).
\]  

(10)

Let $\epsilon = (f^*(\mu + t) - f^*(\mu))/t - \omega$. Dividing (8) by $t$ and substracting (10), we obtain:

\[
\langle \mu A^t A \epsilon, \epsilon \rangle = -\langle A^t A(f^*(\mu + t) - f^*(\mu)), \epsilon \rangle + \sum \langle \gamma_i, \epsilon \rangle.
\]  

(11)
with $\langle \gamma_i, \epsilon \rangle = - \langle (\partial f^* - \partial \lambda^* f^* + t - \frac{D_i f^*}{\| D_i f^* \|_2}, D_i \epsilon \rangle$.

By the Schwarz inequality, $\mu \delta_{\min} \| \epsilon \| \leq \| f^* - f^* + t \| + \sum \gamma_i$. When $t \to 0$, $\gamma_i \to 0$, and thus $\epsilon \to 0$ and $f^* - f^* + t$ is thus differentiable.

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

In the following, we denote $E(\mu)$ the value obtained at the saddle point for $L(f^*, g^*, \lambda^*, \mu)$. The differentiability of $E$ follows from its definition and from the differentiability of $f^* - f^* + t$. Classical differentiability calculation rules give:

$$\partial E(\mu) = \mu (Af^* - g^* + A f^*(\mu)) + \frac{\| g^* - Af^* \|^2}{2} + \sum_i \langle D_i \partial f^*, f^*(\mu) \rangle,$$

(12)

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

$$E'(\mu) = \frac{\| g^* - Af^* \|^2}{2}, \quad E''(\mu) = \langle Af^* - g^*, Af^*(\mu) \rangle,$$

(13)

It can be shown that $E''(\mu) \geq 0$. The function $E(\mu)$ is monotonically increasing and convex.

3.3. The Morozov discrepancy principle and iterative choice of the regularization parameter

The Morozov discrepancy principle has been much studied in linear inverse problems [19]. In the Morozov discrepancy principle, the regularization parameter is chosen such that the discrepancy term is equal to the known noise level on the observed data $\delta$, that is $\| g^* - Af^* \| = c \delta$, for a constant $c \geq 1$. This condition can be written $2E'(\mu) = \delta^2$ for $c=1$.

In order to determine the regularization parameter and to solve the Morozov equation, we have used a two parameters model function. This model function and the approach detailed below for the choice of $\mu$ are independent of the method used to solve the Total Variation optimization problem. The approximate value obtained for the regularization parameter is refined with a Newton method. Following the model function methodology in [12], a model function can be calculated that preserve the main properties of $E'(\mu)$. Upon differentiating the optimality condition with respect to $\mu$, we obtain:

$$A'(Af^* - g^*) + \mu A'(Af^* - g^*) = - \frac{d}{d\mu} \sum_i \langle D_i f^*, f^*(\mu) \rangle,$$

(14)

Taking the inner product of the optimality condition with $f^*$ and neglecting the second term, we obtain:

$$\langle (Af^* - g^*), Af^*(\mu) \rangle, \quad \mu A'(Af^* - g^*) + \mu Af^* \rangle = 0,$$

(15)

This equation will be approximated with $E'(\mu) + \mu E''(\mu) = constant$. Solving this differential equation, we obtain a decreasing and convex two parameters model function $2E'(\mu) = aexp(-\mu) + b$ which is expected to describe the main properties of the value function $E'(\mu)$. This model function yields an approximate value of the regularization parameter. In a second step, we define the function $G(\mu)$ by $G(\mu) = 2E'(\mu) - \delta^2$. In order to solve $G(\mu) = 0$, given two initial guess values, $\mu_0, \mu_1$, a sequence of iterates $(\mu_k)_{k \geq 0}$ is generated with:

$$\mu_{k+1} = \mu_k - 0.1 \frac{2E'(\mu_k) - \delta^2}{E''(\mu_k)},$$

(16)

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

In order to show the effectiveness of our method for choosing the regularization parameter, we have considered a simple deconvolution test problem. The image $f \in \mathbb{R}^{N^2}$ ($N=328$) to be recovered is a real synchrotron bone micro-CT slice displayed in Figure 1A. The image was artificially blurred with a Gaussian kernel of variance $\sigma^2_{\text{blur}} = 1.46$ and a Gaussian noise with a variance $\sigma^2 = 18.4$ was added to the blurred image. The obtained image has the PSNR=24.87 and is displayed in Figure 1B. The noise level is thus set to $\delta^2 = N^2 \sigma^2 = 1.98 \cdot 10^6$. In order to determine an approximate values of the parameters $a$ and $b$ of the model function $E'(\mu)$, two values $\mu_{\text{approx}} = 0.2$ and $\mu_{\text{approx}} = 10$ are first chosen. For each of the regularization parameters, $\mu_*$ and $\mu_{\text{approx}}$, the blurred noisy image is deconvolved with the ADMM algorithm and the values $E'(\mu_*)$ and $E'(\mu_{\text{approx}})$ are calculated. The values obtained for $a$ and $b$ are $8.13 \cdot 10^5$ and $1.44 \cdot 10^6$ respectively. An approximate value for the regularization parameter, $\mu_{\text{approx}}=0.42$, is obtained from the equation $\delta^2 = a \exp(-\mu_{\text{approx}}) + b$. The Newton method is then applied to refine the solution. Starting from $\mu_0 = \mu_{\text{approx}} + 0.1$ and $\mu_1 = \mu_{\text{approx}}$, the iterative formula (16) is applied. For each iteration of the Newton method, the ADMM algorithm is applied, 10 iterations being sufficient to obtain the convergence of the regularization functional. The Newton iterations are stopped when $\frac{\|A f^*(\mu_k) - g\|^2 - \delta^2}{\delta^2} < 0.01$. The convergence is fast and only three iterations are necessary. The final value of the regularization parameter is $\mu_{\text{Morozov}} = 0.37$. The final image obtained with $\mu_{\text{Morozov}}$ has a PSNR=32.39 and is displayed in Figure 1C. The value obtained is a good estimate of the optimal value $\mu_{\text{opt}} = 0.8$ considering the PSNR of the reconstructed image, PSNR=33.03. The iterates $\mu_k$ have been reported on Figure 2-left, which displays the evolution of the data term as a function of the regularization parameter. An extensive comparison with other techniques to get the regularizing parameter will not be detailed here. Yet, the UPRE method for TV parameter selection [14] leads to the optimal regularization parameter $\mu_{\text{UPRE}}=1.4$ (Figure 2-right) with a PSNR=32.45. The results obtained are thus rather similar in both cases but the proposed method does not require the calculation of the reconstructed images for all the regularization parameters.

5. Conclusion

In this paper, we proposed a new iterative method to determine the optimal regularization parameter in deconvolution problems with TV regularization based on the Morozov discrepancy principle. For each value of the regularization parameter, the TV regularization functional is minimized with the ADMM algorithm. An approximate model function for the data term of the regularization functional is first obtained. Then the solution is refined with a Newton iterative method. This method is very fast and avoid an extensive sweeping of the values of the regularization parameter.
References

[1] Rudin L I, Osher S and Fatemi E 1992 Nonlinear total variation based noise removal algorithms Phys. D 60 259-68

[2] Vogel C R 2002 Computational Methods for Inverse Problems (Philadelphia, PA: SIAM)

[3] Wang Y, Yang J, Yin W and Zhang Y 2008 A new alternating minimization algorithm for total variation image reconstruction SIAM Journal on Imaging Sciences 1 248-72

[4] Ng M K, Weiss P and Yuan X 2010 Solving constrained total-variation image restoration and reconstruction problems via alternating direction methods SIAM Journal on Scientific Computing 32 2710-36

[5] Van den Berg E and Friedlander M P 2008 Probing the Pareto frontier for basis pursuit solutions SIAM Journal on Scientific Computing 31 890-912

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

[7] Yang J, Yin W, Zhang Y and Wang Y 2008 A fast algorithm for edge-preserving variational multichannel image restoration SIAM Journal on Imaging Sciences 2 569-92

[8] Afonso M, Bioucas-Dias J and Figueiredo M 2011 An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems IEEE Transactions on Image Processing 20 681-95

[9] Morozov V A 1984 Methods for Solving Incorrectly Posed Problems ed Z Nashed (New York: Springer-Verlag)

[10] Golub G H, Heath M T and Wahba G 1979 Generalized cross-validation as a method for choosing a good ridge parameter Technometrics 21 215-23

[11] Hansen P C and O’Leary D P 1993 The use of the L-curve in the regularization of discrete ill-posed problem SIAM Journal on Scientific Computing 14 1487-503

[12] Kunish K and Zou J 1998 Iterative choices of regularization parameters in linear inverse problems Inverse Problems 14 1247-64

[13] Aujol J F and Gilboa G 2006 Constrained and SNR-based solutions for TV-Hilbert space image denoising J. Math. Imaging Vis. 26 217-37

[14] Lin Y, Wohlberg B and Guo H 2010 UPRE method for total variation parameter selection Signal Processing 90 2546-51

[15] Liao H, Li F and Ng M K 2009 Selection of regularization parameter in total variation restoration JOSA A 26 2311-20

[16] Babacan S D, Molina R and Katsaggelos A K 2008 Parameter estimation in TV image restoration using variational distribution approximation IEEE Transactions on Image Processing 17 326-39

[17] Wen Y W and Chan R H 2012 Parameter selection for total-variation-based image restoration using discrepancy principle IEEE Transactions on Image Processing 21 1770-81

[18] Frick K, Lorenz D A and Resmerita E 2011 Morozov’s principle for the augmented Lagrangian method applied to linear inverse problems Multiscale Modeling and Simulation 1528-48

[19] Engl H W, Hanke M and Neubauer A 2000 Regularization of Inverse Problems (Dordrecht: Kluwer Academic)