Advanced Mathematical Models & Applications
Vol.5, No.1, 2020, pp.53-69

PARAMETER LEARNING FOR COMBINED FIRST AND SECOND ORDER TOTAL VARIATION FOR IMAGE RECONSTRUCTION

Mourad Nachaoui*

Sultan Moulay Slimane University, Faculty of Sciences and Technologies, Beni-Mellal, Morocco

Abstract. This work deals with the study of ill-posed inverse problems involved in many signal processing areas and image analysis. Indeed, the image degradation during the data acquisition process is inevitable. The degradation can be introduced by the imaging process, the image recording, the image transmission, etc. To restore the original image, we should provide supplementary information. This can be done by adding a regularization term. In this context, several regularization techniques have been developed. In particular, Tikhonov regularization and the total variation TV are popular, and their success is confirmed. In this paper a comparative study concerns Tikhonov regularization, first and second order Total Variation TV2 and combined TV+TV2. A particular attention is devoted to learn to regularization parameters via machine learning approach.

Keywords: inverse problems, variational methods, image restoration, total variation, parameter learning.
AMS Subject Classification: 65J22, 65F22, 68U10, 68A50, 68A55, 65K10.
Corresponding author: Nachaoui Mourad, Sultan Moulay Slimane University, Meghila, Beni-Mellal, Morocco, e-mail: nachaoui@gmail.com
Received: 18 December 2019; Revised: 25 March 2020; Accepted: 17 April 2020; Published: 29 April 2020.

1 Introduction

Inverse problem consists of using the observation of a response of an object to infer the values of the parameters that characterize this object. Examples of such problems appear in wide variety of scientific problems. Unlike to the direct problem which is well posed, the inverse problem is ill-posed in the sens of Hadamard (1923). This explains the extensive theoretical an numerical studies of inverse problems appear in many realistic applications fields. We cite some recent papers for different fields (Afraites & Atlas, 2015; Chakib et al., 2012, 2013; Begram et al., 2019; Nachaoui et al., 2016) and some monograph (Borcea et al., 2015; Kabanikhin, 2012; Ito & Jin, 2015; Kern, 2016; Gockenbach, 2012).

A general inverse problem in imaging reads as follows. Let $o \in H$ be a measured data in a suitable of functions defined on a domain Ω. We seek for the original or reconstructed image f that fulfills the model:

$$o = \mathcal{A}f + b, \tag{1}$$

where $\mathcal{A}$ is a linear operator defined in the space H, and b denotes a possible noise component. The operator $\mathcal{A}$ represent the forward operator of this problem. Examples are blurring operators (in which case $\mathcal{A}f$ denotes the convolution of f with a blurring kernel κ). To reconstruct f one has to invert the operator $\mathcal{A}$. This is not always possible since in many applications the problem can be highly ill posed and further complicated by interference like noise. Indeed, we know that by using the Fourier transformation, the convolution becomes a simple product (term by term). Thus, the equation (1) becomes:

$$\mathcal{F}(o) = \mathcal{F}(\kappa)\mathcal{F}(f) + \mathcal{F}(b), \tag{2}$$

where $\mathcal{F}$ denotes the Fourier transformation.

Insofar as the spectrum of $\mathcal{F}(\kappa)$ has no zero, there is existence and uniqueness of the solution $\mathcal{F}(f)$. This can be calculated by simple inversion:

$$\mathcal{F}(f) = \mathcal{F}(\kappa)^{-1}\mathcal{F}(o) - \mathcal{F}(\kappa)^{-1}\mathcal{F}(b). \tag{3}$$

Solving the problem (1) by this (naive) method gives the result presented in the Figure 3.

(a) (b) (c)

Figure 1: Naive Method: (a)original image ,(b) image degraded by a motion blur and noise by an additive Gaussian noise of variance $\sigma = 0.0001$,(c) restored image with the equation (3)

There is great noise amplification. Unfortunately, the term $\mathcal{F}(\kappa)^{-1}\mathcal{F}(b)$ dominates in this equation (3). Even if the noise is slight, it is amplified when reversing $\mathcal{F}(\kappa)$. Indeed, κ modeling a low-pass filter, $\mathcal{F}(\kappa)$ has very low values at high frequencies. So in this case, a common procedure in inverse problems is to add apriori information to the model. This information can be given in general by certain regularity assumption on the image f. Hence, instead of solving (1) one computes f as a minimiser of the following regularized problem

$$arg\min_{f} \|Af - o\|_2^2 + \underbrace{\mathfrak{R}(f)}_{\text{reguralization}} . \tag{4}$$

Regularization methods are a key tool in the solution of inverse problems Engl et al. (1996). Indeed, they are used to introduce prior knowledge and make the approximation of ill-posed problems feasible. In the case of imaging problem the term $\mathfrak{R}(f)$ must be able to fulfill the lack of information caused by the fact that the noise b is unknown. This explains the wide related works investigate the form of the regularization term $\mathfrak{R}(f)$ without claiming to be exaustive we cite some recent works (Alahyane et al., 2019; Dabrock & Van Gennip, 2018; De los Reyes & Schünlieb, 2013; El Mourabit et al., 2017; Laghrib et al., 2016, 2019, 2015, 2018; Papafitsoros & Schünlieb, 2014; Yehu, 2020). One of the most used regularization for inverse problems in general is the so called Tikhonov regularization (Tikhonov & Arsenin, 1978). This technique of regularization have attracted particular interest in the last years and link to other fields like image processing. It is formulated as follows: for all $\alpha > 0$.

$$arg\min_{f} \|Af - o\|_2^2 + \alpha\|\nabla f\|_2^2. \tag{5}$$

The goal is not only to adjust f to approach o, but also impose that the gradient to be "small enough" (depending on the parameter α). Thanks to this regularization, solutions will be regulars (in the space H) and will have small variations and therefore poorly defined edges. This variational formulation used to address the shortcomings of the approach of the inverse filter and overcome his sensitivity to noise, gave birth to several algorithms like

Wiener filtering (Lim & Oppenheim, 1979), SECB method (Carasso, 1999), Richardson-Lucy algorithm (Richardson, 1972; Lucy, 1974), and many others.

An alternative to Tikhonov regularization (that is too brutal), it's the ROF (Rudun, Osher and Fatemi) model Rudin et al. (1992), it involves replacing the regularization term $\|\nabla f\|_2^2$ by a less restrictive and regularizing standard. In the following, we will consider a first and second order total variation (TV) and (TV^2) regularization framework. The main point of this combination mainly consists of preserving the essential features of the image such as boundaries and corners that are degraded, using other approaches.

In any regularization method the choice of parameter α is crucial for the design of an image reconstruction model fitting appropriately the given data. Recently, some authors determine an optimal regularization procedure introducing particular knowledge of the noise distribution into the learning approach (Van Chung et al., 2017; De los Reyes et al., 2013; Dong et al., 2011). In this context, we will combine machine learning and variational regularization techniques to learn the parameters realization of the first and second total variation. In particular we will use the idea proposed in Lyaqini et al. (2020) to learn the optimal regularization parameter.

The outline of the paper is as follows. In Section 2, we present the total variational model and we detail its dicretization. In section 3 we derive the proposed algorithm, and in Section 4, we present the second order total variation model and its dicretization. Section 5 is devoted to the combined total variation $(TV) + (TV^2)$. In section 6 we present some experimental results; in addition, we present our methodology to learn optimal regularization parameters. We finally end the paper by a conclusion.

2 Total variation model

Variational methods have proven to be particularly useful to solve a number of ill-posed inverse imaging problems.

The introduction of the total variation had a lasting impact in imaging sciences and was used for various tasks including denoising, deblurring and segmentation. The major advantage of the total variation is that it allows for sharp discontinuities in the solution. This is of vital interest for many imaging problems, since edges represent important features, e.g. object boundaries or motion boundaries. As a prototype for total variation methods in imaging we recall the total variation based image denoising model proposed by Rudin et al. (1992). The ROF model is defined as the variational problem

$$arg \min_{f \in BV(\Omega)} \|Af - o\|_2^2 + \alpha TV(f). \tag{6}$$

Here the regularization term is $TV(f)$ total variation of f is given by

$$TV(f;\Omega) := \sup \left\{ \int_\Omega f \operatorname{div} \varphi \, \mathrm{dx}; \varphi \in \mathbb{C}^1(\Omega), \|\varphi\| \leq 1 \right\}. \tag{7}$$

Note that when the function f is in

$$W^{1,1}(\Omega) = \{f \in L^1(\Omega), \nabla f \in L^1(\Omega)\} \subset BV(\Omega).$$

The total variation is none other than $TV(f) = \|\nabla f\|_1$. Here $\|\nabla f\|_p$ denotes the usual norm of $L^p(\Omega)$.

Theorem 1. *Assuming that the operator $A : L^2(\Omega) \mapsto L^2(\Omega)$ is continuous and A does not cancel the constants, especially $(A.1 \neq 0)$. The minimization problem (7) has a unique solution $f \in BV(\Omega)$.*

Proof. see Laghrib et al. (2015). □

2.1 Discretization of the problem TV

The discretization is an essential step to solve the minimization problem (7). Lets gives some notations. We denote by $X_{ij}, i = 1, \dots, N, j = 1,_l dots, M$ the discrete image, and $X = \mathbb{R}^{N\times M}$ all discrete images of size $N \times M$ and $\mathcal{Y} = \mathcal{X} \times \mathcal{X}$. The spaces $\mathcal{X}$ and $\mathcal{Y}$ are respectively provided the scalar product $< .,. >_{\mathcal{X}}$ and $< .,. >_{\mathcal{Y}}$ where

$$\forall X, Y \in \mathcal{X},\ < X, Y >_{\mathcal{X}} = \sum_{i=1}^{N}\sum_{j=1}^{M} X_{i,j}Y_{i,j},$$

and

$$\forall p = (p^1, p^2),\ q = (q^1, q^2) \in \mathcal{Y},\ < p, q >_{\mathcal{Y}} = \sum_{i=1}^{N}\sum_{j=1}^{M} p^1_{i,j}q^1_{i,j} + p^2_{i,j}q^2_{i,j}.$$

The gradient vector of X denoted by ∇X is defined on Y as follows $\nabla X = ((\nabla X)^1, (\nabla X)^2)$, with

$$(\nabla X)^1_{i,j} = \begin{cases} X_{i+1,j} - X_{i,j} & if \;\; i < N \\ 0 & i = N \end{cases}, (\nabla X)^2_{i,j} = \begin{cases} X_{i,j+1} - X_{i,j} & if \;\; j < M \\ 0 & j = M \end{cases} \tag{8}$$

The associated operator of $-\nabla$ is defined as follows
div : $\mathcal{X}^2 \to \mathcal{X}$ such that for $p = (p^1, p^2) \in \mathcal{X}^2$, we have

$$\forall \omega \in \mathcal{X},\ < \text{div}\ \ p, \omega > = - < p, \nabla\omega >,$$

where the gradient is defined by the equation (8), then

$$(\text{div}\ \ p)_{i,j} = (\text{div}\ \ p)^1_{i,j} + (\text{div}\ \ p)^2_{i,j},$$

where

$$(\text{div}\ \ p)^1_{i,j} = \begin{cases} p^1_{i,j} - p^1_{i-1,j} & si \;\; 1 < i < N \\ p^1_{i,j} & si \;\; i = 1 \\ -p^1_{i-1,j} & si \;\; i = N \end{cases}, (\text{div}\ \ p)^2_{i,j} = \begin{cases} p^2_{i,j} - p^2_{i,j-1} & si \;\; 1 < j < M \\ p^2_{i,j} & si \;\; j = 1 \\ -p^2_{i,j-1} & si \;\; j = M \end{cases} \tag{9}$$

The discrete version of the regularization TV, represented by J reads as follows

$$J(X) = \sum_{i=1}^{N}\sum_{j=1}^{M} |(\nabla X)_{i,j}|,$$

where $|.|$ is the Euclidean norm of $\mathbb{R}^2$, defined as follows

$$|(\nabla X)_{i,j}| = |\left((\nabla X)^1_{i,j}, (\nabla X)^2_{i,j}\right)| = \sqrt{\left((\nabla X)^1_{i,j}\right)^2 + \left((\nabla X)^2_{i,j}\right)^2}.$$

To solve the problem with the TV regularization (7), several numerical optimization algorithms were developed especially for the model where it is assumed that the noise is Gaussian. These algorithms are too many to list them all here. One can for example refer to Zhu et al. (2010); Goldstein & Osher (2009); Chambolle (2004). In this work we use the gradient descent algorithm Nocedal & Wright (2006).

3 Optimization algorithm

We consider the general problem:
find $f^* \in BV(\mathbb{R}^{mn})$ such that

$$J(f^*) = \min_{f \in BV(\mathbb{R}^{mn})} J(f), \tag{10}$$

where J is continuously differentiable functional on $\mathbb{R}^{nm}$. To solve the TV problem we use the classical Steepest descent algorithm Nocedal & Wright (2006). This method is summarized in in the following algorithm

Algorithm 1: Optimal gradient steepest

Require: f^0: We start by random f^0,
Ensure: f^*: solution of the problem 10
while the stopping criterion is not satisfied **do**
- at the step k

$$f^{k+1} = f^k - \delta t_k \nabla J(f^k)$$

where $\delta t_k \in \mathbb{R}$ is: $arg \min_{\delta t} J(f^k - \delta t \nabla J(f^k))$.
end while

The algorithm is stopped when convergence test is true, or if a maximum number of iterations is exceeded. A convergence test could depend on an error (relative or absolute). Let

$$\min_{f \in \mathbb{R}^{nm}} \frac{1}{2} \sum_{i=0}^{n} \sum_{j=0}^{m} |(Af)_{i,j} - o_{i,j}|^2 + \alpha \sqrt{\left((\nabla f)^1_{i,j}\right)^2 + \left((\nabla f)^2_{i,j}\right)^2}, \tag{11}$$

where $(\nabla f)^1_{i,j}$ and $(\nabla f)^2_{i,j}$ are defined in the equation (8).

To apply the algorithm 1, we must compute $\nabla J(f^k)$ After standard computations we obtain

$$\nabla J(f) = A^T(Af - o) + \left(- \operatorname{div} \left(\frac{\nabla f}{\sqrt{((\nabla f)^1)^2 + ((\nabla f)^2)^2 + \varepsilon^2}} \right) \right)_{i,j}, \tag{12}$$

where the division should be understood as a term by term division. We will see in the numerical results, the regularization TV type is very effective to remove the blur and noise. However, it still suffers from the effect called staircaising. For this, the introduction of higher order model would be better.

4 Second order Regularization

The placing of the two order will allow us to sufficiently smooth the regular part of the image, to get rid of staircasing effect, while generating an acceptable blur, preserving information about the contours of the image. The image restoration problem f with TV^2 regularization is defined as follows

$$\hat{f}_{MAP} = \underset{f \in HB(\Omega)}{\arg\min} \quad \frac{1}{2} \parallel Af - o \parallel_2^2 + \alpha TV^2(f). \tag{13}$$

The natural space for which we are seeking a solution is the space $W(\Omega)$, defined as

$$W(\Omega) = \left\{ f \in L^2(\Omega); \nabla^2 f \in \left[L^1(\Omega) \right]^4 \right\}.$$

As for the problem with the legalization TV, this space is not reflexive since the space $L^1(\Omega)$ is not. However, an interesting point is that bounded sequences in $W(\Omega)$ are bounded in space bounded Hessian $HB(\Omega)$ Demengel & Temam (1984) defined by

$$HB(\Omega) = \{ f \in W^{1,1}(\Omega) / \nabla f \in (BV(\Omega))^2 \}. \tag{14}$$

Then we could use the compactness properties in this space. We then have the following theorem

Theorem 2. *Suppose the operator* $A : L^2(\Omega) \mapsto L^2(\Omega)$ *is continuous and injective, with* $(A.1 \neq 0)$. *The minimization problem (13) has a unique solution* $f \in HB(\Omega)$.

Proof. see Laghrib et al. (2015). □

4.1 Discretization and optimization of the problem TV^2:

Let

$$R(f) = \alpha \parallel \nabla^2 f \parallel_1$$

As for the problem with TV, the discretization of the problem with TV^2 is the same way, using finite difference schemes. We use the alternative approach proposed in Rudin et al. (1992), based on the discretization of the PDE obtained by the method of gradient descent Nocedal & Wright (2006). Using variations of computational techniques, we find the PDE associated with the problem (13) with periodic conditions on δ (we denote $M = n \times m$ the image size). Similarly the problem TV after discretization of the problem, one must calculate $\delta_t R$ to apply the gradient descent method.

$$\begin{cases} \delta_t R = \operatorname{div}^2\left(\dfrac{\nabla^2 f}{|\nabla^2 f|}\right) & in\ \Omega, \\ f(0,y) = f(m,y), f(x,0) = f(x,n) & in\ \delta\Omega, \end{cases} \tag{15}$$

where div^2 is the second order divergence operator, satisfying the following properties

$$\operatorname{div}^2 X \cdot Y = X \cdot \nabla^2 Y \quad \forall Y \in \mathbb{R}^M,\ X \in (\mathbb{R}^M)^4,$$

where "$\cdot$" is the Euclidean product. Since the problem (13) is convex, the solution obtained by the gradient descent method coincides with that of the associated PDE. Let

$f_{i,j}$ the discrete image of f , such that $f_{i,j} = f(i,j), i = 1...m, j = 1...n$. We define the second order denoted ∇^2 as follows

$$\nabla^2 X = (\nabla_{xx} X\ \nabla_{xy} X\ \nabla_{xy}\ X \nabla_{yy} X),$$

where

$$\nabla_{xx} X_{i,j} = \begin{cases} X_{i,n} - 2X_{i,1} + X_{i,2} & si\ \ 1 \le i \le m,\ j = 1 \\ X_{i,j-1} - 2X_{i,j} + X_{i,j+1} & si\ \ 1 \le i \le m,\ 1 < j < n \\ X_{i,n-1} - 2X_{i,n} + X_{i,1} & si\ \ 1 \le i \le m,\ j = n, \end{cases}$$

$$\nabla_{yy} X_{i,j} = \begin{cases} X_{m,j} - 2X_{1,j} + X_{2,j} & si\ \ 1 \le j \le n,\ i = 1 \\ X_{i-1,j} - 2X_{i,j} + X_{i+1,j} & si\ \ 1 \le j \le n,\ 1 < i < m \\ X_{m-1,j} - 2X_{m,j} + X_{1,j} & si\ \ 1 \le j \le n,\ i = m, \end{cases}$$

on the other hand we have

$$\nabla_{xy} X_{i,j} = \begin{cases} X_{i,j} - X_{i+1,j} - X_{i,j+1} + X_{i+1,j+1} & si\ \ 1 \le i < m,\ 1 \le j < n \\ X_{i,n} - X_{i+1,n} - X_{i,1} + X_{i+1,1} & si\ \ 1 \le i < m,\ j = n \\ X_{m,j} - X_{1,j} - X_{m,j+1} + X_{1,j+1} & si\ \ i = m,\ 1 \le j < n \\ X_{m,n} - X_{1,n} - X_{m,1} + X_{1,1} & si\ \ i = m,\ j = n. \end{cases}$$

Furthermore, for $X = (X_1, X_2, X_3, X_4) \in (\mathbb{R}^M)^4$, we define the operator div^2 as follows

$$(\operatorname{div}^2 X)_{i,j} = \overline{\nabla_{xx}}(X_1)_{i,j} + \overline{\nabla_{yy}}(X_2)_{i,j} + \overline{\nabla_{xy}}(X_3)_{i,j} + \overline{\nabla_{xy}}(X_4)_{i,j}.$$

where

$$\overline{\nabla_{xx}} = \nabla_{xx},\ \overline{\nabla_{yy}} = \nabla_{yy}$$

and

$$\overline{\nabla_{xy}} X_{i,j} = \begin{cases} X_{1,1} - X_{1,n} - X_{m,1} + X_{m,n} & si\ \ i=1,\ j=1 \\ X_{1,j} - X_{1,j-1} - X_{m,j} + X_{m,j-1} & si\ \ i=1,\ 1<j\leq n \\ X_{i,1} - X_{i-1,n} - X_{i-1,1} + X_{i-1,n} & si\ \ 1<i\leq m,\ j=1 \\ X_{i,j} - X_{i,j-1} - X_{i-1,j} + X_{i-1,j-1} & si\ \ 1<i\leq m,\ 1<j\leq n \end{cases}$$

Finally, we all determined, it remains to solve the equation (15) by applying the algorithm 1. We will see later, in the part of numerical results, the type of regularization model TV^2 provides a very effective solution to staircaising. However, the restored image is slightly degraded by the appearance of the blur. This leads us to look for other adjustments where we can balance between removing the staircasing effect and the appearance of the blur. Hence the introduction of mixed models we present an example, the model where we combine the regularization TV with TV^2.

5 Combination of regularization TV and TV^2

To fight against the blur problem that appears in the regularization TV^2, one idea is to exploit the benefits of regularization TV^2 (reduced scairtasing) and also preserve image details by exploiting benefits of TV regularization. We define the problem of restoration of the image f from o in the following way:

$$\hat{f}_{MAP} = \arg\min_{f} \quad \frac{1}{2} \parallel Af - o \parallel_2^2 + \alpha TV(f) + \beta TV^2(f)$$

$$\hat{f}_{MAP} = \arg\min_{f} \quad \frac{1}{2} \parallel Af - o \parallel_2^2 + \alpha \int_{\Omega} F(\nabla f)\,\mathrm{dx} + \beta \int_{\Omega} G(\nabla^2 f)\,\mathrm{dx} \tag{16}$$

with α et β are positive parameters, and $F : \mathbb{R}^2 \to \mathbb{R}^+$, $G : \mathbb{R}^4 \to \mathbb{R}^+$ are convex functions with at least one linear growth at infinity, ie

$$\exists k_1, k_1' > 0 / \forall X \in \mathbb{R}^2, k_1|X|_2 - k_1' \leq F(X) \leq k_1|X|_2 + k_1'$$

and

$$\exists k_2, k_2' > 0 / \forall X \in \mathbb{R}^2, k_2|X|_2 - k_2' \leq G(X) \leq k_2|X|_2 + k_2',$$

where $|.|_2$ is the Euclidean norm on $\mathbb{R}^2$ and $\mathbb{R}^4$ respectively. These two conditions ensure that if $\nabla f \in [L^1(\Omega)]^2$ then $\parallel F(\nabla f) \parallel_1$ is well defined, and and that if $\nabla^2 f \in [L^1(\Omega)]^4$, thus $\parallel G(\nabla^2 f) \parallel_1$,is also well defined. Moreover, these two conditions are also needed to show the coercivity. We note with J function to minimize, defined as

$$J(f) = \frac{1}{2} \parallel Af - o \parallel_2^2 + \alpha \int_{\Omega} F(\nabla f)\,\mathrm{dx} + \beta \int_{\Omega} G(\nabla^2 f)\,\mathrm{dx}\,. \tag{17}$$

This function is well defined on the space $W^{2,1}(\Omega)$. Since this space is not reflexive, we can not use direct methods of calculating variations to prove the existence of a solution. However, relaxation techniques we use: we will extend the function J on a space larger than the space $W^{2,1}(\Omega)$, with interesting properties compactness compared with a well defined topology. As in the case of regularization of order 2, we will work in space $HB(\Omega)$). We now extends the function J on space $HB(\Omega)$ as follows

$$J_{et}(f) = \begin{cases} \frac{1}{2} \parallel Af - o \parallel_2^2 + \alpha \int\limits_{\Omega} F(\nabla f)\,\mathrm{dx} + \beta \int\limits_{\Omega} G(\nabla^2 f)\,\mathrm{dx} & si \;\; f \in W^{2,1}(\Omega) \\ +\infty & si \;\; f \in HB(\Omega) \setminus W^{2,1}(\Omega). \end{cases} \tag{18}$$

We denote by χ_Ω the characteristic function of Ω.

Theorem 3. *Assume that $A(\chi_\Omega \neq 0),\ \alpha > 0,\ \beta > 0$, the minimization problem reads as follows*

$$\min_{f \in HB(\Omega)} J(f), \tag{19}$$

admits a unique solution unique in $HB(\Omega)$.

Proof. see (Papafitsoros & Schünlieb, 2014). □

5.1 Discretization and optimization of the problem $TV + TV^2$:

We consider for each element $X = (X_1, X_2) \in (\mathbb{R}^{m\times n})^2$

$$\parallel X \parallel_1 = \sum_{i=1}^{m} \sum_{j=1}^{n} \sqrt{(X_1(i,j))^2 + (X_2(i,j))^2}.$$

We first begin with the discretization of the gradient operator $\nabla X = ((\nabla X)^1, (\nabla X)^2)$ such that $X_{i,j},\ i,j = 1, ...(m,n)$ is the discrete image, and $\mathbb{R}^{n\times m}$ the set of the discrete images.

$$(\nabla X)^1_{i,j} = \begin{cases} X_{i+1,j} - X_{i,j} & si \;\; i < m \\ 0 & si \;\; i = m, \end{cases}$$
$$(\nabla X)^2_{i,j} = \begin{cases} X_{i,j+1} - X_{i,j} & si \;\; j < n \\ 0 & si \;\; j = n. \end{cases}$$

We also need to define the discretization of the adjoint operator of the gradient " div " : $(\mathbb{R}^{m\times n})^2 \to \mathbb{R}^{m\times n}$", satisfying the following relationship

$$-\operatorname{div} Y.X = Y.\nabla X, \forall X \in \mathbb{R}^{m\times n}, Y \in (\mathbb{R}^{m\times n})^2$$

with

$$\operatorname{div}(Y^1, Y^2)_{i,j} = (\operatorname{div}(Y^1, Y^2))^1_{i,j} + (\operatorname{div}(Y^1, Y^2))^2_{i,j}$$

and

$$(\operatorname{div}(Y^1, Y^2))^1_{i,j} = \begin{cases} Y^1_{i,j} - Y^1_{i-1,j} & si \;\; 1 < i < m \\ Y^1_{i,j} & si \;\; i = 1 \\ 0 & si \;\; i = m \end{cases}$$
$$(\operatorname{div}(Y^1, Y^2))^2_{i,j} = \begin{cases} Y^2_{i,j} - Y^2_{i,j-1} & si \;\; 1 < j < n \\ p^2_{i,j} & si \;\; j = 1 \\ -p^2_{i,j-1} & si \;\; j = M. \end{cases}$$

We define also in same way the Hessian operator "∇^2", as follows

$$\nabla^2 X = (\nabla_{xx} X \;\; \nabla_{xy} X \;\; \nabla_{xy} \;\; X \nabla_{yy} X),$$

where

$$\nabla_{xx}X_{i,j} = \begin{cases} X_{i,n} - 2X_{i,1} + X_{i,2} & si \;\; 1 \le i \le m, \; j = 1 \\ X_{i,j-1} - 2X_{i,j} + X_{i,j+1} & si \;\; 1 \le i \le m, \; 1 < j < n \\ X_{i,n-1} - 2X_{i,n} + X_{i,1} & si \;\; 1 \le i \le m, \; j = n \end{cases}$$

$$\nabla_{yy}X_{i,j} = \begin{cases} X_{m,j} - 2X_{1,j} + X_{2,j} & si \;\; 1 \le j \le n, \; i = 1 \\ X_{i-1,j} - 2X_{i,j} + X_{i+1,j} & si \;\; 1 \le j \le n, \; 1 < i < m \\ X_{m-1,j} - 2X_{m,j} + X_{1,j} & si \;\; 1 \le j \le n, \; i = m \end{cases}$$

in the other hand we have

$$\nabla_{xy}X_{i,j} = \begin{cases} X_{i,j} - X_{i+1,j} - X_{i,j+1} + X_{i+1,j+1} & si \;\; 1 \le i < m, \; 1 \le j < n \\ X_{i,n} - X_{i+1,n} - X_{i,1} + X_{i+1,1} & si \;\; 1 \le i < m, \; j = n \\ X_{m,j} - X_{1,j} - X_{m,j+1} + X_{1,j+1} & si \;\; i = m, \; 1 \le j < n \\ X_{m,n} - X_{1,n} - X_{m,1} + X_{1,1} & si \;\; i = m, \; j = n. \end{cases}$$

Moreover, for $X = (X_1, X_2, X_3, X_4) \in (\mathbb{R}^M)^4$, we define the operator div^2 as follows

$$(\operatorname{div}^2 X)_{i,j} = \overline{\nabla_{xx}}(X_1)_{i,j} + \overline{\nabla_{yy}}(X_2)_{i,j} + \overline{\nabla_{xy}}(X_3)_{i,j} + \overline{\nabla_{xy}}(X_4)_{i,j},$$

where

$$\overline{\nabla_{xx}} = \nabla_{xx}, \; \overline{\nabla_{yy}} = \nabla_{yy}$$

and

$$\overline{\nabla_{xy}}X_{i,j} = \begin{cases} X_{1,1} - X_{1,n} - X_{m,1} + X_{m,n} & si \;\; i = 1, \; j = 1 \\ X_{1,j} - X_{1,j-1} - X_{m,j} + X_{m,j-1} & si \;\; i = 1, \; 1 < j \le n \\ X_{i,1} - X_{i-1,n} - X_{i-1,1} + X_{i-1,n} & si \;\; 1 < i \le m, \; j = 1 \\ X_{i,j} - X_{i,j-1} - X_{i-1,j} + X_{i-1,j-1} & si \;\; 1 < i \le m, \; 1 < j \le n. \end{cases}$$

Having defined all the discretized operators needed to solve the problem $TV + TV^2$, it remains only to apply an optimization algorithm such as gradient descent algorithm 1.

6 Parameter learning and numerical results

To show the performance of the approaches mentioned above, we choose to do a synthetic test on the image of the cameraman (figure 2), we destroy the original image by adding a blur using a convolution kernel that simulates a blur circular with a radius r defined. Can we disrupt the image blurred by adding a white stone additive Gaussian noise variance of $\sigma = 0.0001$.

Note that the image of the cameraman suffered severe damage and so many details that have been destroyed and so more information is lost. Obviously, the choice of algorithm parameters affects the quality of the restored image. And especially for the α of TV model, the β model TV^2 and couple (α, β) for the mix model $TV + TV^2$. To do this, we will learn these parameters to find the optimal choice.

(a) Cameraman **psnr=20.5834 ssim=0.1030**

(b) Lenna **psnr=25.4621 ssim=0.2681**

Figure 2: `Image test`

6.1 Bi-level optimization for parameter learning

A well-known example of parameter optimization problems is supervised machine learning. In our general learning problem, we look for (α, β) solving the problem

$$\min_{(\alpha,\beta)\in[\underline{\alpha},\overline{\alpha}]\times[\underline{\beta},\overline{\beta}]} \Psi(f_{\alpha,\beta}) \tag{20a}$$

subject to

$$f_{\alpha,\beta} \in \operatorname*{arg\,min}_{f\in HB(\Omega)} J(f, \alpha, \beta). \tag{20b}$$

The problem (20a)(20b) is the so called bi-level problem. For training data $\mathcal{I}$ with desired reconstruction $\tilde{f} \in \mathcal{I}$, we consider a loss function Ψ that estimates the discrepancy between $\tilde{f}$ and the reconstruction $f_{\alpha,\beta}$. Generally, Ψ is expressed in L^2 norm. However, recently it was proved in Lyaqini et al. (2020) that the non-smooth loss functions as L^1 norm for supervised learning problem give more consistent results. In this case we define our loss function as

$$\Psi(f_{\alpha,\beta}) := \sum_{k\in\mathcal{I}} \|\tilde{f}_k - f^k_{\alpha,\beta}\|_1. \tag{21}$$

Note that the lower problem (20b) is non-smooth and the same for the upper problem (20a). For the lower problem we use the smoother C^2-Huber regularized version of the total variation De los Reyes et al. (2013). More precisely, we consider for $\gamma >> 1$ the C^2 Huber-type regularization of the

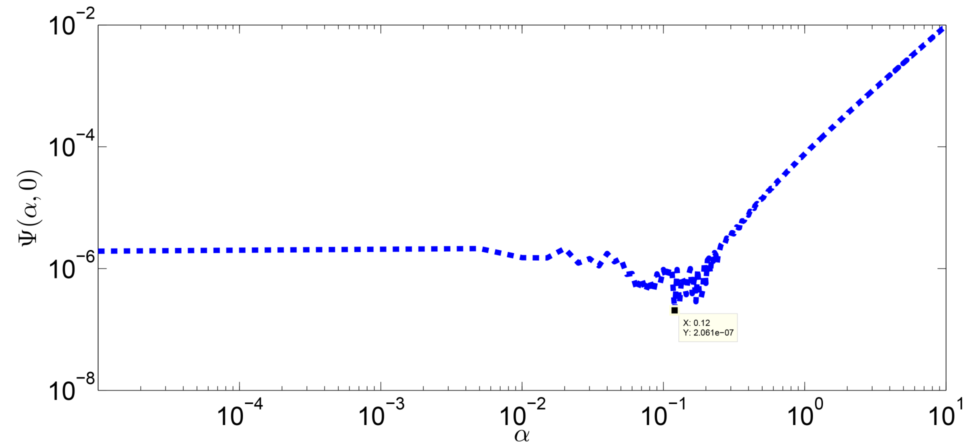
$$h_\gamma(z) := \begin{cases} \frac{z}{|z|} & \text{if } \gamma|z| - 1 \geq \frac{1}{2\gamma} \\ \frac{z}{|z|}(1 - \frac{\gamma}{2}(1 - \gamma|z| + \frac{1}{2\gamma})^2) & \text{if } \gamma|z| - 1 \in (-\frac{1}{2\gamma}, \frac{1}{2\gamma}) \\ \gamma z & \text{if } \gamma|z| - 1 \leq -\frac{1}{2\gamma}. \end{cases} \tag{22}$$

For the cost functional Ψ we will use the smoothed method used in Lyaqini et al. (2020),

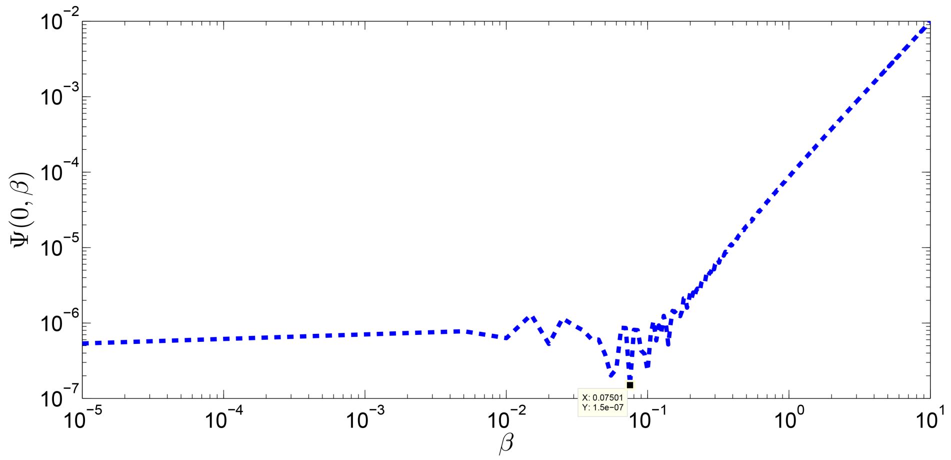
$$|z|_\varepsilon = \sqrt{z^2 + \varepsilon}. \tag{23}$$

As in classical bi-level optimization, sin the lower problem (20b) is convex, the problem (20a)-(20b) will changed equivalently to a single optimization where the lower problem is replaced by its first order optimality conditions. A necessary and sufficient optimality condition for (20b) is then given by the following equation

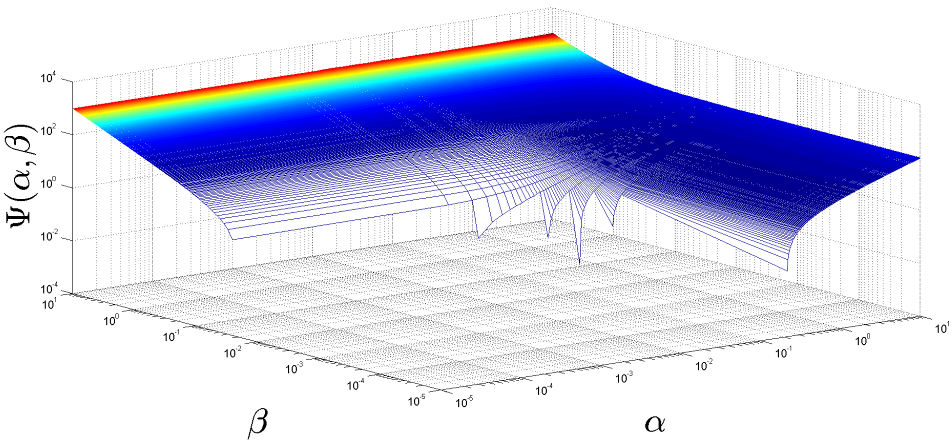
$$(Af - b, w) + \alpha(h_\gamma(\nabla f), \nabla w) + \beta(h_\gamma(\nabla^2 f), \nabla^2 w) = 0. \tag{24}$$



(a) Learning parameter α, $\beta = 0$



(b) Learning parameter β, $\alpha = 0$



(c) Learning parameter (α, β)

Figure 3: `The loss function with respect to the free parameters.`

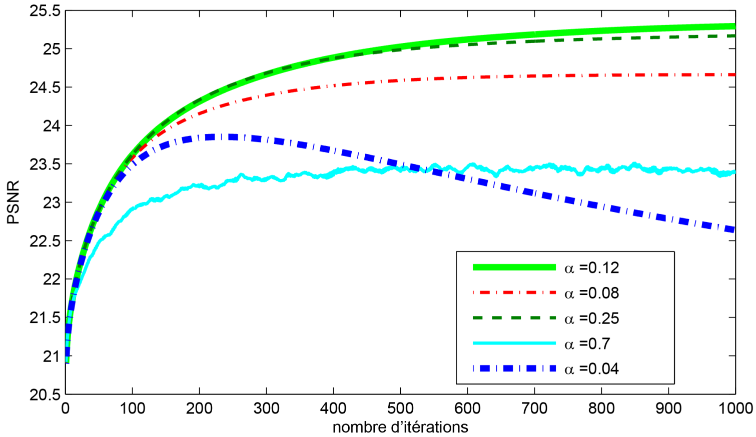


Figure 4: PSNR Variation with respect to the iterations number for different values of α for the TV model

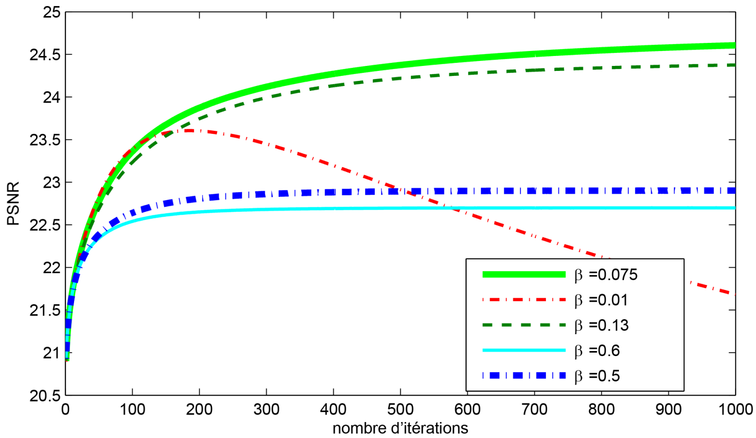


Figure 5: PSNR Variation with respect to the iterations number for different values of β for the TV^2 model

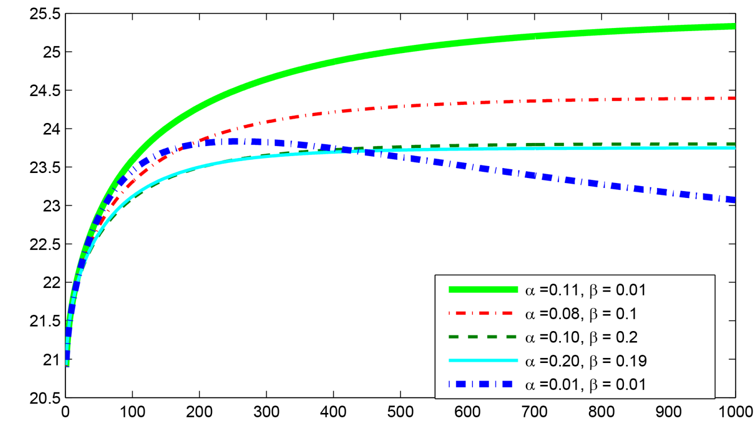


Figure 6: PSNR Variation with respect to the iterations number for different values of (α, β) for the $TV + TV^2$ model

To derive to optimality conditions for the problem (20a) subject to (24) we define the Lagrangian function

$$\mathcal{L}(\alpha, \beta, f, p) := \Psi(f) + (Af - b, p) + \alpha(h_\gamma(\nabla f), \nabla p) + \beta(h_\gamma(\nabla^2 f), \nabla^2 p). \tag{25}$$

By derivation of $\mathcal{L}$ with respect to α, β and f and make it equal to zero we obtain the following aptimality system
Adjoint

$$(A^* p, w) + \alpha(h'_\gamma(\nabla f)\nabla p, \nabla w) + \beta(h'_\gamma(\nabla^2 f)\nabla^2 p, \nabla^2 w) = 0, \text{ for all } w \in H^2(\Omega). \tag{26a}$$

Gradient

$$\begin{cases} \int_\Omega h_\gamma(\nabla f)\nabla p dx(\alpha - \alpha_1) \geq 0 \\ \int_\Omega h_\gamma(\nabla^2 f)\nabla^2 p(\beta - \beta_1) \geq 0 \end{cases} \tag{26b}$$

for all $\alpha_1 \in [\underline{\alpha}, \overline{\alpha}]$ and $\beta_1 \in [\underline{\beta}, \overline{\beta}]$.

For the determination of the optimal parameter values we consider a decent gradient method, together with an Armijo backtracking line search rule. We distinguish between three cases. Just TV which means that $\beta = 0$ and we look for the optimal α. Just TV^2,which means that $\alpha = 0$ and we look for the optimal β. Finally, for the combined $TV + TV^2$ we look for (α, β). In figure 3 we present the evolution of the loss function with respect of free parameter in each case.

The figures (4,6,16) show the influence of parameter regularization on the picture quality "cameraman" in terms of psnr for 1000 iterations.

6.2 Comparison results

For all algorithms, we took care to choose the best settings that make the walk the best possible algorithm. These parameters are in fact those who give the highest values in Figures 6, 5, 4. The figure 7 illustrates a comparison between different image restoration methods that we detailed, and other algorithms in the state of art.

Whether the vision or the values of PSNR and SSIM, the mixed model $TV + TV^2$ has proven effective (zoom on images to note the difference). Indeed, we note that other algorithms produce artifacts on the restored image, for example the Wiener algorithm, eliminating the noise was not really satisfactory, while the Richardson-Lucy algorithm produces the ringing artifact that can be observed on the edges of the image, the TV has given considerable results but it still suffers from the effect of scairtasing the madman has also infected the image restored by the model TV^2.

(a) Wiener **PSNR=20.4900 ssim =0.2103**

(b) Richardson-Lucy **PSNR=22.5998 ssim =0.2672**

(c) TV **PSNR=25.2933 ssim =0.3421**

(d) TV2 **PSNR=24.5915 ssim =0.3160**

(e) TV+TV2 **PSNR=25.3329 ssim =0.3488**

Figure 7: Restoration of the cameraman image using different deconvolution methods.

7 Conclusion

In this work we have investigate a variational approach to solve an inverse problem appears in image processing. More precisely we have deal with a restoration of blurred and degraded images. To solve this ill-posed problem we have opted of regularization techniques. In this case, tree type of regularization have been studied. Namely Thikhonv, total variational (TV) and second order total variational (TV^2) an the combined $(TV + TV^2)$ have been investigated. In particular, we have learned the regularization parameters by considered the bi-level optimization combined with machine learning. The effectiveness of our approach have been proved via the presentation of some numerical.

8 Acknowledgement

This work was initiated during an author's stay at the Jean Leray Mathematics Laboratory, Nantes University and was partially supported by the "Fédération de Recherche Mathématiques des Pays de Loire FR-CNRS2962, France.

References

Afraites, L., Atlas, A. (2015). Parameters identification in the mathematical model of immune competition cells. *Journal of Inverse and Ill-posed Problems, 23*(4), 323-337.

Afraites, L., Hadri, A. & Laghrib, A. (2020). A denoising model adapted for impulse and Gaussian noises using a constrained-PDE. *Inverse Problems, 36*(2), 025006.

Alahyane, M., Hakim, A., Laghrib, A. & Raghay, S. (2019). A fast approach of nonparametric elastic image registration problem. *Mathematical Methods in the Applied Sciences, 42*(18), 7059-7075.

Bergam, A., Chakib, A., Nachaoui, A. & Nachaoui, M. (2019). Adaptive mesh techniques based on a posteriori error estimates for an inverse Cauchy problem. *Applied Mathematics and Computation*, 346, 865-878.

Borcea, L., Kang, H., Liu, H., Uhlmann, G. (2013). Inverse problems and imaging. Lectures from the workshop held at the Institut Henri Poincaré, Paris, February 20–22. Edited by H. Ammari and J. Garnier.

Carasso, A.S. (1999). Linear and nonlinear image deblurring: A documented study. *SIAM Journal on Numerical Analysis, 36*(6), 1659-1689.

Chakib, A., Nachaoui, A., & Nachaoui, M. (2013). Approximation and numerical realization of an optimal design welding problem. *Numerical Methods for Partial Differential Equations, 29*(5), 1563-1586.

Chakib, A., Ellabib, A., Nachaoui, A. & Nachaoui, M. (2012). A shape optimization formulation of weld pool determination. Applied *Mathematics Letters, 25*(3), 374-379.

Chambolle, A. (2004). An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision, 20*(1-2), 89-97.

Dabrock, N., Van Gennip, Y. (2018). A note on "Anisotropic total variation regularized L1-approximation and denoising/deblurring of 2D bar codes". *Inverse Problems & Imaging, 12*(2), 525-526.

De los Reyes, J.C., Schünlieb, C.B. (2013). Image denoising: learning the noise model via nonsmooth PDE-constrained optimization. *Inverse Problems & Imaging, 7*(4), 1183-1214.

De los Reyes, J.C., Schünlieb, C.B. & Valkonen, T. (2017). Bilevel parameter learning for higher-order total variation regularisation models. *Journal of Mathematical Imaging and Vision, 57*(1), 1-25.

Demengel, F., Temam, R. (1984). Convex functions of a measure and applications. *Indiana University Mathematics Journal, 33*(5), 673-709.

Dong, Y., Hintermüller, M., & Rincon-Camacho, M.M. (2011). Automated regularization parameter selection in multi-scale total variation models for image restoration. *Journal of Mathematical Imaging and Vision, 40*(1), 82-104.

El Mourabit, I., El Rhabi, M., Hakim, A., Laghrib, A. & Moreau, E. (2017). A new denoising model for multi-frame super-resolution image reconstruction. *Signal Processing,* 132, 51-65.

Engl, H.W., Hanke, M. & Neubauer, A. (1996). *Regularization of inverse problems* (Vol. 375). Springer Science & Business Media.

Gockenbach, M. (2016). *Linear Inverse Problems and Tikhonov Regularization* (No. 32). The Mathematical Association of America.

Goldstein, T., Osher, S. (2009). The split Bregman method for L1-regularized problems. *SIAM Journal on Imaging Sciences, 2*(2), 323-343.

Hadamard, J. (1923). Lecture on the Cauchy problem in Linear Partial Differential Equations, Oxford University Press, London.

Kabanikhin, S.I. (2011). Inverse and Ill-Posed Problems: Theory and Applications (Vol. 55). Walter De Gruyter.

Kern, M. (2016). *Numerical methods for inverse problems.* Mathematics and Statistics Series. ISTE, London; John Wiley & Sons, Inc., Hoboken, NJ.

Laghrib, A., Ghazdali, A., Hakim, A. & Raghay, S. (2016). A multi-frame super-resolution using diffusion registration and a nonlocal variational image restoration. *Computers & Mathematics with Applications, 72*(9), 2535-2548.

Laghrib, A., Hadri, A., & Hakim, A. (2019). An edge preserving high-order PDE for multiframe image super-resolution. *Journal of the Franklin Institute, 356*(11), 5834-5857.

Laghrib, A., Ezzaki, M., El Rhabi, M., Hakim, A., Monasse, P. & Raghay, S. (2018). Simultaneous deconvolution and denoising using a second order variational approach applied to image super resolution. *Computer Vision and Image Understanding,* 168, 50-63.

Laghrib, A., Hakim, A. & Raghay, S. (2015). A combined total variation and bilateral filter approach for image robust super resolution. *EURASIP Journal on Image and Video Processing, 2015*(1), 1-10.

Lim, J.S., Oppenheim, A.V. (1979). Enhancement and bandwidth compression of noisy speech. *Proceedings of the IEEE, 67*(12), 1586-1604.

Lucy, L.B. (1974). An iterative technique for the rectification of observed distributions. *Astronomical Journal,* 79, 745.

Lyaqini, S., Quafafou, M., Nachaoui, M. & Chakib, A. (2020). Supervised learning as an inverse problem based on non-smooth loss function. *Knowledge and Information Systems*, 1-20.

Ito, K., Jin, B. (2015). *Inverse problems: Tikhonov theory and algorithms.* Series on Applied Mathematics, 22. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ.

Nachaoui, A., Chakib, A., & Nachaoui, M. (2016). Existence analysis of an optimal shape design problem with non coercive state equation. *Nonlinear Analysis: Real World Applications*, 28, 171-183.

Nocedal, J., Wright, S. (2006). *Numerical Optimization.* Springer Science & Business Media.

Papafitsoros, K. & Schünlieb, C.B. (2014). A combined first and second order variational approach for image reconstruction. *Journal of Mathematical Imaging and Vision, 48*(2), 308-338.

Rudin, L.I., Osher, S. & Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, *60*(1-4), 259-268.

Richardson, W.H. (1972). Bayesian-based iterative method of image restoration. *J. Opt. Soc. Am.*, *62*(1), 55-59.

Tikhonov, A., Arsenin, V.Y. (1978). Solution of incorrectly formulated problems and the regularization method. *Mathematics of Computation*, *32*(144), 1320-1322.

Van Chung, C., De los Reyes, J.C. & Schünlieb, C.B. (2017). Learning optimal spatially-dependent regularization parameters in total variation image denoising. *Inverse Problems, 33*(7), 074005.

Yehu, Lv. (2020). Total generalized variation denoising of speckled images using a primal-dual algorithm. *Journal of Applied Mathematics and Computing, 62*(1-2), 489-509.

Zhu, M., Wright, S.J. & Chan, T.F. (2010). Duality-based algorithms for total-variation-regularized image restoration. *Computational Optimization and Applications, 47*(3), 377-400.