Orthogonal projection algorithm for first and second order total variation denoising

Idriss El Mourabit, Mohammed El Rhabi, Abdelilah Hakim

Orthogonal projection algorithm for first and second order total variation denoising

Idriss El Mourabit[a,*], Mohammed El Rhabi [a,b], Abdelilah Hakim[a]

[a] *University Cadi Ayyad, LAMAI FST Marrakech, Morocco;[*] Correspondig author : z.driss.88@gmail.com*

[b] *Applied Mathematics and Computer Science department, Ecole des Ponts ParisTech (ENPC), Paris, France.*

Abstract

The denoising problem is the process of removing the noise from a degraded image. As we know, the Rodin Osher Fatemi (ROF) denoising model based on total variation is a robust approach for solving the ill-posed problem. To avoid the staircasing effects caused by the first order total variation, the second order one is proposed. In this work, we present an orthogonal projection algorithm for solving the ROF model with first and second order total variation. The efficiency and robustness against noise of the proposed model are illustrated and compared with the classical methods through numerical simulations.

Keywords: Projection algorithm, Denoising, Total variation, Second order total variation, Partial differential equations.

1. Introduction

The general idea behind variational denoising models is to consider an observed image u_0 as a noiseless version of an image one u. In denoising problems, u is the solution of an ill-posed inverse problem which is presented as follow

$$u_0 = u + n, \tag{1}$$

where u_0 is noisy image and n is an additive Gaussian noise.

The ROF denoising model introduced in [8] is one of the most successful variational algorithms, which consists of minimizing the first order total variation. To reconstruct the true image u from u_0, Rodin Osher Fatemi proposed to minimize the following function

$$\inf_{u\in BV(\Omega)} \lambda TV(u) + \frac{1}{2}\|u-u_0\|^2_{\mathrm{L}^2(\Omega)} \qquad (ROF_1)$$

where $TV(u)$ represents the total variation of a $u \in BV(\Omega)$ defined as follows

$$\begin{aligned} TV(u) &= \int_\Omega |Du|dx \\ &= \sup\left\{\int_\Omega u \;\mathrm{div}\,\varphi\, dx \;\; \varphi \in \mathcal{C}^1(\Omega,\mathbb{R}),\; \|\varphi\|_\infty \le 1\right\}, \end{aligned}$$

and $BV(\Omega)$ is the space of integrable functions with bounded variations on a bounded domain Ω. This space appears to be a suitable functional space for image analysis, since it contains functions that can be discontinuous across edges. The first term in (ROF_1) is a regularization part, controlled by a positive weighting parameter λ, the second one, measures the difference between the original image u and the noised one using the $\mathrm{L}^2(\Omega)$ fidelity norm. The existence and uniqueness of the solution of (ROF_1) has being studied in [11, 1]. To treat the very noised images, Chambolle proposes in [5] an efficient algorithm to approximate this solution. In fact, the use of the TV in the (ROF_1) model favors piecewise constant structures more than smooth structures, this makes staircasing effect. One of the solutions of this problem has been proposed in [4], the authors propose to use the second order total variation in ROF_1, then the problem writes as follows

$$\inf_{u\in BV^2(\Omega)} \lambda TV_2(u) + \frac{1}{2}\|u-u_0\|^2_{\mathrm{L}^2(\Omega)} \qquad (ROF_2)$$

where $BV^2(\Omega)$ is the space of bounded hessian functions and TV_2 is the second order total variation defined as

$$TV_2(u) = \int_\Omega \mid D^2u \mid dx \qquad (2)$$

with D^2u is the second distributional derivative of the function u. In the same reference [4] the authors present an other version Chombolle algorithm for resolving the problem (ROF_2). On the other hand, the (ROF_1) can be resolved with another projection orthogonal algorithm which has been presented in [7] in dimension one and is based on a projected gradient method. In this work we generalize this algorithm for minimizing the first and second total variation of an image of dimension two.

2. The ROF1 model

2.1. The theoretical framework

The image restoration problem consists of finding an original image $u : \Omega \to \mathbb{R}$ from a degraded one u_0 and the image u_0 is related to u by the degradation model (1) with a Gaussian noise n. According to the maximum likelihood principle u can be approximated by solving the following least-square problem

$$\inf_u \int_\Omega |u - u_0|^2 dx \tag{3}$$

but this problem is ill-posed in the sense of Hadamard and it has been regularized in [10] by Tickonov and Arsenin by adding a regularization term to (3). The authors proposed to consider the following problem

$$\inf_u \lambda \|\nabla u\|^2_{\mathrm{L}^2(\Omega)} + \frac{1}{2} \int_\Omega |u - u_0|^2 dx. \tag{4}$$

The associated Euler-Lagrange equation of (4) is very strong isotropic and then the edges are not preserved. The Rodin Osher Fatemi (ROF_1) based on variation total is an efficient model for removing the noise and preserve the image edges. Solving this problem leads to minimizing the following expression:

$$\inf_{u \in BV(\Omega)} \lambda TV(u) + \frac{1}{2} \int_\Omega |u - u_0|^2 dx. \tag{ROF_1'}$$

The following theorem has been proved in [11, 1].

Theorem 1. *The problem (ROF_1) has an unique solution in $BV(\Omega)$ space.*

The Euler-Lagrange equation associated to (ROF_1) is

$$\lambda \partial TV(u) + (u - u_0) \ni 0 \tag{5}$$

where $\partial TV(u)$ is the sub-differential of the function $TV(u)$ defined as

$$w \in \partial TV(u) \Leftrightarrow TV(v) \geq TV(u) + < w, v - u >_{\mathrm{L}^2(\Omega)} \qquad \forall v$$

the equation (5) can be written

$$(u_0 - u) \in \lambda \partial TV(u)$$

and according to convex analysis [9] we have

$$u \in \partial TV^* \left(\frac{u_0 - u}{\lambda} \right)$$

with TV^* is the Legendre-Fenchel transform of TV.

Equivalently,

$$\frac{u_0}{\lambda} \in \frac{(u_0 - u)}{\lambda} + \frac{1}{\lambda} \partial TV^* \left(\frac{u_0 - u}{\lambda} \right)$$

we get that $w = \dfrac{(u_0 - u)}{\lambda}$ is the minimizer of

$$\frac{1}{2} \parallel w - \frac{u_0}{\lambda} \parallel^2 + \frac{u_0}{\lambda} TV^*(w). \tag{6}$$

Since TV is a convex and one-homogeneous ($TV(\lambda u) = \lambda TV(u)$ for all u and $\lambda > 0$) function and according to convex analysis results [9] TV^* is the indicator function of a closed convex set K:

$$TV^*(u) = \begin{cases} 0 & u \in K \\ \infty & \text{otherwise} \end{cases}$$

where K is given by

$$K = \left\{ u \in \mathrm{L}^2(\Omega) : \langle u, v \rangle_{\mathrm{L}^2(\Omega)} \geq TV(v) \ \forall v \in \mathrm{L}^2(\Omega) \right\}$$

from the definition of total variation (2), K can be seen as the closure of the following set

$$\left\{ div\ \xi \ \ \xi \in \mathcal{C}_c^1(\Omega, \mathbb{R}), \|\xi\|_\infty \leq 1 \right\}.$$

Since the solution w of (6) is given by an orthogonal projection of $\frac{u_0}{\lambda}$ on K, and then the solution of (ROF_1) can be computed by

$$u = u_0 - \pi_{\lambda K}(u_0) \tag{7}$$

where $\pi_{\lambda K}$ is the projection operator onto K.

2.2. Computing the discrete solution with the orthogonal projection algorithm

Form the formula (7), resolving total variation problem is equivalent to computing the projection $\pi_{\lambda K}(u_0)$, to this end Chambolle (2004) has been proposed an efficient algorithm for estimating this projection based on a fixed point method. In other way, it could be computed by a projected gradient method, this idea was proposed in [7] applied in the case of a 1-D noisy signal, in this section we will generalize this for an image of dimension two.

Now we will give some notation, we denote by $u_{i,j},\ i = 1, ...N,\ j = 1, ..., M$ a discrete image and $X = \mathbb{R}^{N\times M}$ the set of all discrete images of size $N \times M$ and $Y = X \times X$.

The spaces X and Y are both equipped respectively with scalar product $< .,. >_X$ and $< .,. >_Y$ where

$$\forall u, v \in X,\ < u, v >_X = \sum_{i=1}^{N}\sum_{j=1}^{M} u_{i,j}v_{i,j}$$

and

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

The gradient of an element u written ∇u belongs to Y and could be defined by several manners. One of them consists to set $\nabla u = ((\nabla u)^1, (\nabla u)^2)$ with

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

The adjoint operator of $-\nabla$ is defined as $div : X^2 \to X$ with, for all $p = (p^1, p^2) \in X^2$ we have

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

When the gradient is given by (8) then

$$(\text{div } p)_{i,j} = (\text{div } p)^1_{i,j} + (\text{div } p)^2_{i,j} \tag{9}$$

where

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

Then the discrete version of TV denoted by J is defined as

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

where $|.|$ is the Euclidean norm.

Then we can write the J as

$$\begin{aligned} J(u) &= \sup \{< p, \nabla u >_Y, p \in Y, |p_{i,j}| \leq 1 \ \ i = 1, ..., N \ \text{ and } j = 1, ..., M\} \\ &= \sup_{p \in G} < p, u >_X \end{aligned}$$

and the set G is defined by

$$G = \{- \text{div } p | \ p \in Y, |p_{i,j}| \leq 1 \ \ i = 1, ..., N \ \text{ and } j = 1, ..., M\}$$

from the previous notation the ROF_1 is equivalent to solve the following problem

$$\min_{u \in X} \frac{1}{2} |u - u_0|_Y + \lambda J(u) \tag{10}$$

According to the previous the solution of (10) is

$$u = u_0 - \pi_{\lambda G}(u_0) \tag{11}$$

where $\pi_{\lambda G}(u_0)$ is the orthogonal projection of u_0 on the convex set λG.

Thus, to compute u we are lead to compute the projection operator $\pi_{\lambda G}$ on the convex set λG, i.e., to solve the problem

$$\min_{p \in \lambda G} F(u) \tag{12}$$

where $F(p) = \|p - u_0\|^2$.

Let $\rho > 0$, the optimality condition for the problem (12) could be written as follows

$$p = \Pi_D(p - \rho \nabla F(p)) \tag{13}$$

where $D = \{p \in Y | \ \forall i,j \ \ |p_{i,j}| \leq 1 \ \}$, $\nabla F(p) = -2\lambda\nabla(\lambda \operatorname{div} \ p - u_0)$ and Π_D is the orthogonal projection on D. It is straightforward to see that

$$\Pi_D(q)_{i,j} = \begin{cases} \dfrac{q_{i,j}}{|q_{i,j}|} & \text{if} \quad |q_{i,j}| > 1 \\ q_{i,j} & \text{otherwise} \end{cases}$$

the problem can be solved by a classical projection algorithm which give as:

1. $p^0 = 0$.

2. for $n \geq 0$

$$p^{n+1} = \Pi_D(p^n - \rho\nabla F(p^n))$$

where $\rho > 0$ suitably chosen. Then

$$\begin{aligned} \|p^{n+1} - p\| &= \|\Pi_D(p^n - \rho\nabla F(p^n)) - \Pi_D(p - \rho\nabla F(p))\| \\ &\leq \|p^n - p + 2\rho\lambda^2\nabla \operatorname{div} \ (p^n - p)\| \\ &\leq \|(I + 2\rho\lambda^2\nabla \operatorname{div})(p^n - p)\| \\ &\leq |I + 2\rho\lambda^2\nabla \operatorname{div} | \|p^n - p\| \end{aligned}$$

Since the eigenvalues of $\nabla \operatorname{div}$ are negative the sequence $\|p^n - p\|$ is necessary decreasing. Then, it converges when

$$\lambda^2\rho < \frac{1}{|\mu|} = \frac{1}{\|\nabla \operatorname{div} \|_2}$$

where μ is the eigenvalue of ∇div having the largest absolute value.
In order to evaluate the performance of the proposed algorithm in sense of the convergence, we will compare it with the Cambolle algorithm [5] (see numerical implementation section).

3. The ROF2 model

The computed (ROF_1) solution makes some numerical perturbations. In fact, the computed solution turns to be piecewise constant which is called the staircasing effect. In order to resolve this problem it has been proposed to use the second order functional space of bounded variation $BV^2(\Omega)$. This model leads to the minimisation of the following expression:

$$\inf_{u \in BV^2(\Omega)} TV^2(u) + \frac{\lambda}{2} \parallel u - u_0 \parallel^2_{\mathrm{L}^2(\Omega)} \qquad (ROF_2')$$

where TV^2 is the second total variation define by

$$TV^2(u) = \sup \left\{ \int_\Omega u H^* \varphi dx \ \ \varphi \in \mathcal{C}^2_c(\Omega, \mathbb{R}), \ \|\varphi\|_\infty \leq 1 \right\}$$

and $BV^2(\Omega)$ is space of bonded hessian function defined as

$$BV^2(\Omega) = \left\{ u \in W^{1,1}(\Omega), \ \ TV^2(u) < \infty \right\}$$

this space endowed with the norm $\|.\|_{BV^2(\Omega)} = \|.\|_{W^{1,1}(\Omega)} + TV^2(.)$ is a Banach space, for others properties of $BV^2(\Omega)$ see [6].
In the case of the finite dimension the following theorem has been proved in [4].

Theorem 2. *The ROF_2 model has an unique solution in $BV^2(\Omega)$ space.*

3.1. Compute the discrete ROF2 solution

In this section we are going to compute the ROF_2 numerical solution with the proposed orthogonal projection algorithm. We first present some recalls and notations. The we note by $X = \mathbb{R}^{N \times M}$ set of all discrete images and $Z = X^4$. Let u be an element of X, the Hessian matrix of u denoted by Hu is identified to a vector of Z and we denote by J_2 the discrete version of the TV^2 defined by

$$J_2(u) = \sum_{i=1}^{N} \sum_{j=1}^{M} \|(Hu)_{i,j}\|_{\mathbb{R}^4}.$$

Thus, the discretization of the ROF_2 model can be defined as

$$\min_{u \in X} \frac{1}{2} |u - u_0|_X + \lambda J_2(u) \qquad (14)$$

the associate Euler Lagrange equation of (12) is the following

$$u - u_0 + \lambda \partial J_2(u) \ni 0$$

then

$$\frac{u - u_0}{\lambda} \in J_2(u)$$

this equivalent [9] as

$$u \in J_2^*(\frac{u-u_0}{\lambda})$$

where J_2^* is the Legendre-Fenchel transform of J_2. Therefore

$$\frac{u_0}{\lambda} \in z + \frac{1}{\lambda} J_2^*(z)$$

with $z = \dfrac{u-u_0}{\lambda}$ solution the following problem

$$\min \frac{1}{2} \parallel z - \frac{u_0}{\lambda} \parallel^2 + \frac{u_0}{\lambda} J_2^*(z) \tag{15}$$

thus, the solution of ROF_2 is given as [4]

$$u = u_0 - \pi_{\lambda G_2}(u_0)$$

where $\pi_{\lambda G_2}(u_0)$ orthogonal projection of u_0 on G_2 with

$$G_2 = \{H^* q,\ q \in Z,\ |q_{i,j}\|_{\mathbb{R}^4} \leq 1\ \forall i,j\}\,.$$

In [4] the authors present a new version of Chambolle algorithm for computing the projection $\pi_{\lambda G_2}(u_0)$, but here we are going to compute it with the proposed orthogonal projection algorithm which can be more efficient in sense of convergence and restoration. Computing the projection $\pi_{\lambda G_2}(u_0)$ is equivalent to resolve the following problem

$$\min_{z \in G_2} \|z - u_0\|_X^2$$

or

$$\min_{q \in D_2} \|\lambda H^* q - u_0\|_X^2 \tag{16}$$

where D_2 is the following convex set

$$D_2 = \{q \in Z,\ |q_{i,j}\|_{\mathbb{R}^4} \leq 1\ \forall i,j\}\,.$$

We denote $G(q) = \|\lambda H^* q - u_0\|_X$ then $\nabla G(q) = 2\lambda H(\lambda H^* q - u_0)$.

Therefore, the numerical solution of (15) is given by

1. $q^0 = 0$.

2. for $n \geq 0$

$$q^{n+1} = \Pi_{D_2}(q^n - 2\tau\lambda H(\lambda H^* q^n - u_0)) \tag{17}$$

where $\tau > 0$ is a real number.

Proposition 3. *The sequence q^n converges to q solution of* (15)*, when $n \to \infty$, if λ and τ satisfy the following condition:*

$$\tau\lambda^2 \leq \frac{1}{\|HH^*\|}$$

Proof. We have

$$\begin{aligned}
\|q^{n+1} - q\|_{\mathbb{R}^4} &= \|\Pi_{D_2}(q^n - 2\tau\lambda H(\lambda H^* q^n - u_0)) - \Pi_{D_2}(q - 2\tau\lambda H(\lambda H^* q - u_0))\|_{\mathbb{R}^4} \\
&\leq \|q^n - 2\tau\lambda H(\lambda H^* q^n - u_0) - q + 2\tau\lambda H(\lambda H^* q - u_0)\|_{\mathbb{R}^4} \\
&= \|q^n - q - 2\tau\lambda^2 HH^*(q^n - q)\|_{\mathbb{R}^4} \\
&\leq \|q^n - q\|_{\mathbb{R}^4} \|I - 2\tau\lambda^2 HH^*\|
\end{aligned}$$

this equivalent to

$$\|q^n - q\|_{\mathbb{R}^4} \leq \|q\|_{\mathbb{R}^4} \|I - 2\tau\lambda^2 HH^*\|^{n+1}$$

then, $\|q^n - q\|_{\mathbb{R}^4} \to 0$ if only if

$$\|I - 2\tau\lambda^2 HH^*\| < 1$$

we conclude that

$$\tau\lambda^2 \leq \frac{1}{\|HH^*\|}.$$

$\square$

4. Numerical implementation

In this section we will present some numerical examples of denoising process to illustrate the difference between the proposed algorithm (13, 17) and both the Chambolle 1 [5] and Chambolle 2 [4]. The comparison is about the speed of convergence.

First of all we consider two examples, the first one is the image of Lena degraded with an additive Gaussian noise of standard deviation $\sigma = 28$ and the second one is a color image degraded with the same noise. In the Figures 1 and 3 we display these images with corresponding results of each algorithm. As we know that the ROF_1 approach makes staircasing effect, since the resulting image is piecewise constant on smooth areas (Fig. 1 and Fig. 3: Chambolle 1 and projection 1), which is not the case with the ROF_2 model the staircasing effect disappears (Fig. 1 and Fig. 3: Chambolle 2 and projection 2).

Figure 1: From left to right and from top to bottom : true image, noisy image and denoised images provided by Chambolle 1, projection 1, Chambolle 2 and projection 2.

Figure 2: From left to right and from top to bottom: true image, noisy image and denoised images provided by Chambolle 1, projection 1, Chambolle 2 and projection 2.

In addition, we can evaluate the performance of each algorithm, by using the Peak Signal to Noise Ratio (PSNR) and Signal to Noise Ratio (SNR) which is defined by

$$PSNR = 10 \log 10 \left(\frac{255^2}{\|u - u^*\|_2^2} \right)$$

and

$$SNR = 10 \log 10 \left(\frac{\|u\|_2}{\|u - u^*\|_2} \right)$$

where u is the true image and u^* is the denoised image. Table 1 displays the averages of the PSNR and SNR of the four algorithms. We show that the projection 1 and Chambolle 1 have probably the same results, but the proposed algorithm (projection 2) provides the best results in terms of both PSNR and SNR as Chombolle 2. To illustrate the efficiency of the proposed algorithm in sense of convergence, we display in the Figures 2 and 4 the comparison of the criterion. The left-curve in the Figure 2 and 4 compare the convergence between the proposed algorithm and Chambolle 1 and the right-curve in the same figures present the comparison between projection 2 and Chambolle 2. We show that the proposed algorithm converge much faster to the solution than Chambolle 1 and 2. The parameter λ is selected appropriately for each algorithm, and the gradient descent parameter ρ and τ are respectively $\rho = 0.01$ and $\tau = 0.001$ for

	measures	Chambolle1	Projection1	Chambolle2	Projection2
21 cm Lena	PSNR	29.65	29.65	29.61	29.92
	SNR	23.97	23.96	23.95	24.25
21 cm Onion	PSNR	29.41	29.38	29.18	29.51
	SNR	22.35	22.32	22.17	22.49

Table 1: Comparison of measures PSNR and SNR.

the two examples.

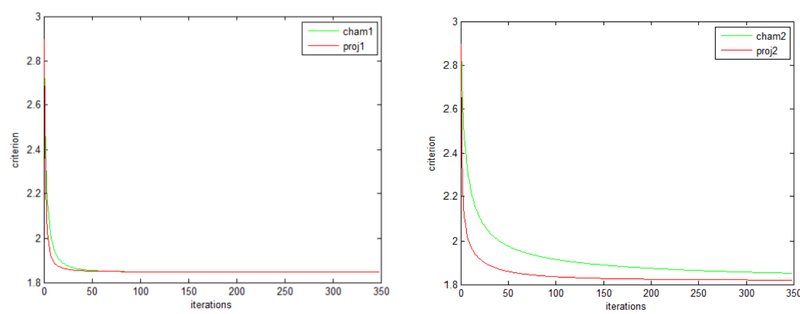

Figure 3: Comparative study of convergence for the image Lena, from left to right: comparison of criterion for Chambolle 1(cham1) and projection (proj1), Chambolle 2(cham2) and projection (proj2).

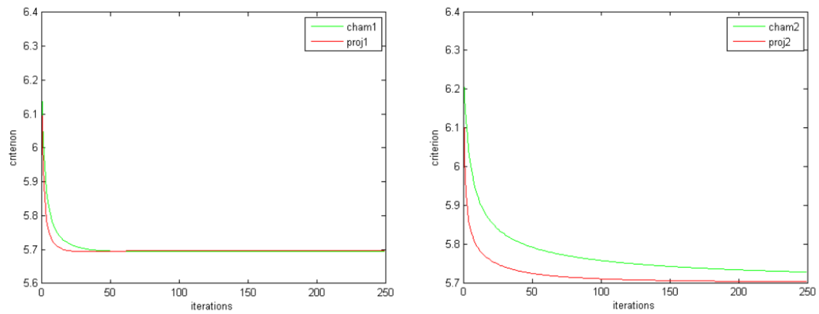

Figure 4: Comparative study of convergence for the image Onion, from left to right: comparison of criterion for Chambolle 1(cham1) and projection (proj1), Chambolle 2(cham2) and projection (proj2).

5. Conclusion

In this work, we have proposed an algorithm based on projected gradient method for solving the famous Rodin Osher Fatemi model (order one and two). The numerical implementation is described and experiment results for image denoising have been tested and compared with the classical algorithms. The efficiency of this paper is about the convergence, indeed, the proposed algorithm converges faster than Chambolle algorithms (order one and two), in addition it has a best denoising results using the PSNR and SNR criteria for the second order total variation model compared with the two Chambolle algorithm.

[1] R. Acart and C.R. Vogel, Analysis of bounded variation penalty methods for ill-posed problems, *Inverse Problems* **10** (1994), 1217–1229.

[2] H. Attouch, G. Buttazzo, and G. Michaille, *Variational analysis in sobolev and BV spaces: applications to PDEs and optimization*, MOS-SIAM series on optimization, Philadelphia, 2006.

[3] G. Aubert and P. Kornprobst, *Mathematical problems in image processing. Partial differential equations and the calculus of variations* (second edition), Applied Mathematical Sciences Seris **147**, Springer, New York, 2006.

[4] M. Bergounioux and L. Piffet, A second-order model for image denoising, *Set-Valued Var. Anal.* **18** (2010), no. 3-4, 277–306.

[5] A. Chambolle, An algorithm for total variation minimization and applications, *Journal of Mathematical Imaging and Vision* **20** (2004), 89–97.

[6] F. Demengel, Fonctions à hessien borné, *Annales de l'institut Fourier* **34** (1984), no. 2, 155–190.

[7] M. El Rhabi, H. Fenniri, A. Keziou, and E. Moreau, A robust algorithm for convolutive blind source separation in presence of noise, *Signal Processing* **93** (2013), no. 4, 818–827.

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

[9] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, N. J., 1970.

[10] A. N. Tikhonov and V. Y. Arsenin, *Solutions of ill-posed problems*, V.H. Winston and Sons, Washington D.C., John Wiley and Sons, New York. Translated from the Russian, 1977.

[11] L. Vese, A study in the BV space of a denoising-deblurring variational problem. *Applied Mathematics and Optimization*, **44**(2) (2001), 131–161.