Orthogonal projection algorithm for first and second order total variation denoising

Idriss El Mourabit, Mohammed El-Rhabi, and Abdelilah Hakim

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.

Key words and phrases. Projection algorithm, denoising, total variation, second order total variation, differential equation.

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

\[
u_0 = u + n,
\]

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

\[
\inf_{u \in BV(\Omega)} \lambda TV(u) + \frac{1}{2} \|u - u_0\|_{L^2(\Omega)}^2
\]

(ROF1)

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

\[
TV(u) = \int_{\Omega} |Du| dx
\]

\[
= \sup \left\{ \int_{\Omega} u \text{ div } \varphi \ dx \mid \varphi \in C^1(\Omega, \mathbb{R}), \|\varphi\|_{\infty} \leq 1 \right\},
\]

and \( BV(\Omega) \) is the space of integrable functions with bounded variations on a bounded domain \( \Omega \). 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 (ROF1) is a regularization part, controlled by a positive weighting parameter \( \lambda \), the second one, measures the difference between the original image \( u \) and the noised one using the \( L^2(\Omega) \) fidelity norm. The existence and uniqueness of the solution of (ROF1) has been 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

This paper has been presented at Congrès MOCASIM, Marrakech, 19-22 November 2014.
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

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

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

$$TV_2(u) = \int_{\Omega} |D^2u| \, dx$$

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

$$\inf_u \int_{\Omega} |u - u_0|^2 \, dx \quad (3)$$

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

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

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:

$$\inf_{u \in BV(\Omega)} \lambda TV(u) + \frac{1}{2} \int_{\Omega} |u - u_0|^2 \, dx. \quad (ROF'_1)$$

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

**Theorem 2.1.** The problem $(ROF_1)$ has an unique solution in $BV(\Omega)$ space.

The Euler-Lagrange equation associated to $(ROF_1)$ is

$$\lambda \partial TV(u) + (u - u_0) \geq 0 \quad (5)$$

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

$$w \in \partial TV(u) \iff TV(v) \geq TV(u) + \langle w, v - u \rangle_{L^2(\Omega)} \quad \forall v$$

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,
\[ u_0 \in \frac{u_0 - u}{\lambda} + \frac{1}{\lambda} \partial TV^* \left( \frac{u_0 - u}{\lambda} \right) \]
we get that \( w = \frac{(u_0 - u)}{\lambda} \) is the minimizer of
\[ \frac{1}{2} \| w - \frac{u_0}{\lambda} \|^2 + \frac{u_0}{\lambda} TV^*(w). \]  
(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
\[ K = \{ u \in L^2(\Omega) : \langle u, v \rangle_{L^2(\Omega)} \geq TV(v) \ \forall v \in L^2(\Omega) \} \]
from the definition of total variation (2), \( K \) can be seen as the closure of the following set
\[ \{ \text{div } \xi \in C^1_c(\Omega, \mathbb{R}), \| \xi \|_\infty \leq 1 \}. \]
Since the solution \( w \) of (6) is given by an orthogonal projection of \( \frac{u_0}{\lambda} \) on \( K \), and then the solution of \((\text{ROF}_1)\) can be computed by
\[ u = u_0 - \pi_{\lambda K}(u_0) \]  
(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 \( \langle ., . \rangle_X \) and \( \langle ., . \rangle_Y \) where
\[ \forall u, v \in X, \langle u, v \rangle_X = \sum_{i=1}^{N} \sum_{j=1}^{M} u_{i,j} v_{i,j} \]
and
\[ \forall p = (p^1, p^2), q = (q^1, q^2) \in Y, \langle p, q \rangle_Y = \sum_{i=1}^{N} \sum_{j=1}^{M} p^1_{i,j} q^1_{i,j} + p^2_{i,j} q^2_{i,j} \]
The gradient of an element \( u \) written \( \nabla u \) belongs to \( Y \) and could be defined by several manners. One of them consists to set
\[
(\nabla u)_{i,j}^1 = \begin{cases} u_{i+1,j} - u_{i,j} & \text{if } i < N \\ 0 & \text{if } i = N \end{cases} \quad (\nabla u)_{i,j}^2 = \begin{cases} u_{i,j+1} - u_{i,j} & \text{if } j < M \\ 0 & \text{if } j = M \end{cases}
\] (8)

The adjoint operator of \( -\nabla \) is defined as \( \text{div} : X^2 \to X \) with, for all \( p = (p^1, p^2) \in X^2 \) we have
\[
\forall w \in X, \quad \langle \text{div} p, w \rangle = -\langle p, \nabla w \rangle
\]
When the gradient is given by (8) then
\[
(\text{div} p)_{i,j} = (\text{div} p)_{i,j}^1 + (\text{div} p)_{i,j}^2
\] (9)

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
\[
J(u) = \sup_{p \in \mathbb{Y}} \{ \langle p, \nabla u \rangle_{\mathbb{Y}}, p \in \mathbb{Y}, |p_{i,j}| \leq 1 \quad i = 1, \ldots, N \quad \text{and} \quad j = 1, \ldots, M \}
\]
and the set \( G \) is defined by
\[
G = \{ -\text{div} p \mid p \in \mathbb{Y}, |p_{i,j}| \leq 1 \quad i = 1, \ldots, N \quad \text{and} \quad j = 1, \ldots, M \}
\]
from the previous notation the ROF is equivalent to solve the following problem
\[
\min_{u \in \mathbb{X}} \frac{1}{2} |u - u_0|_{\mathbb{Y}} + \lambda J(u)
\] (10)

According to the previous the solution of (10) is
\[
u = u_0 - \pi_{\lambda G}(u_0)
\] (11)

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

Thus, to compute \( u \) we are lead to compute the projection operator \( \pi_{\lambda G} \) on the convex set \( \lambda G \), i.e., to solve the problem
\[
\min_{p \in \lambda G} F(u)
\] (12)

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

Let \( \rho > 0 \), the optimality condition for the problem (12) could be written as follows
\[
p = \Pi_D(p - \rho \nabla F(p))
\] (13)

where \( D = \{ p \in \mathbb{Y} \mid \forall i, j \mid p_{i,j} \leq 1 \} \), \( \nabla F(p) = -2\lambda \nabla (\lambda \text{div} p - u_0) \) and \( \Pi_D \) is the orthogonal projection on \( D \). It is straightforward to see that
\[
\Pi_D(q)_{i,j} = \begin{cases} \frac{q_{i,j}}{|q_{i,j}|} & \text{if } |q_{i,j}| > 1 \\ q_{i,j} & \text{otherwise} \end{cases}
\]
the problem can be solved by a classical projection algorithm which gives 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
\[
\|p^{n+1} - p\| = \|\Pi_D(p^n - \rho \nabla F(p^n)) - \Pi_D(p - \rho \nabla F(p))\|
\leq \|p^n - p + 2\rho \lambda^2 \nabla \text{div} (p^n - p)\|
\leq \|(I + 2\rho \lambda^2 \nabla \text{div})(p^n - p)\|
\leq \|I + 2\rho \lambda^2 \nabla \text{div}\| \|p^n - p\|
\]
Since the eigenvalues of \( \nabla \text{div} \) are negative the sequence \( \|p^n - p\| \) is necessary decreasing. Then, it converges when
\[
\lambda^2 \rho < \frac{1}{|\mu|} = \frac{1}{\|\nabla \text{div}\|_2}
\]
where \( \mu \) is the eigenvalue of \( \nabla \text{div} \) 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 Cambole algorithm [5] (see numerical implementation section).

3. The ROF2 model

The computed (ROF1) 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 minimization of the following expression:

\[
\inf_{u \in BV^2(\Omega)} TV^2(u) + \frac{\lambda}{2} \| u - u_0 \|_{L^2(\Omega)}^2
\]

(ROF2)

where \( TV^2 \) is the second total variation defined by
\[
TV^2(u) = \sup \left \{ \int_{\Omega} uH^* \varphi dx \mid \varphi \in C^2_c(\Omega, \mathbb{R}), \| \varphi \|_{\infty} \leq 1 \right \}
\]
and \( BV^2(\Omega) \) is space of bounded 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 3.1.** The ROF2 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 ROF2 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 defined by

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

Thus, the discretization of the ROF2 model can be defined as

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

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

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

then

\[
u - u_0 \over \lambda \ni J_2(u)
\]

this equivalent [9] as

\[
u \ni J_2^*(\frac{u - u_0}{\lambda})
\]

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

\[
u_0 \over \lambda \ni z + \frac{1}{\lambda} J_2^*(z)
\]

with \( z = \frac{u - u_0}{\lambda} \) solution the following problem

\[
\min_{z \in \mathbb{R}^4} \frac{1}{2} \| z - \frac{u_0}{\lambda} \|^2 + \frac{u_0}{\lambda} J_2^*(z) \tag{15}
\]

thus, the solution of ROF2 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

\[
\min_{z \in G^2} \| z - u_0 \|^2_X
\]

or

\[
\min_{q \in D^2} \| \lambda H^* q - u_0 \|^2_X \tag{16}
\]

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

\[
(1) \quad q^0 = 0.
\]
(2) for \( n \geq 0 \)

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

(17)

where \( \tau > 0 \) is a real number.

**Proposition 3.2.** The sequence \( q^n \) converges to \( q \) solution of (15), when \( n \to \infty \), if \( \lambda \) and \( \tau \) satisfy the following condition:

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

**Proof.** We have

\[
\| q^{n+1} - q \|_{\mathbb{R}^4} = \| \Pi_{D_2}(q^n - 2\tau \lambda H(\lambda H^* q^n - u_0)) - \Pi_{D_2}(q - 2\tau \lambda H(\lambda H^* q - u_0)) \|_{\mathbb{R}^4}
\]

\[
\leq \| q^n - 2\tau \lambda H(\lambda H^* q^n - u_0) - q + 2\tau \lambda H(\lambda H^* q - u_0) \|_{\mathbb{R}^4}
\]

\[
= \| q^n - q - 2\tau \lambda^2 H H^*(q^n - q) \|_{\mathbb{R}^4}
\]

\[
\leq \| q^n - q \|_{\mathbb{R}^4} \| I - 2\tau \lambda^2 H H^* \|
\]

this equivalent to

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

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

\[
\| I - 2\tau \lambda^2 H H^* \| < 1
\]

we conclude that

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

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 ROF1 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 ROF2 model the staircasing effect disappears (Fig. 1 and Fig. 3: 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 Chambolle 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 $\lambda$ is selected appropriately for each algorithm, and the gradient descent parameter $\rho$ and $\tau$ are respectively $\rho = 0.01$ and $\tau = 0.001$ for the two examples.

![Figure 1](image1.png)

**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.

### 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

<table>
<thead>
<tr>
<th></th>
<th>measures</th>
<th>Chambolle1</th>
<th>Projection1</th>
<th>Chambolle2</th>
<th>Projection2</th>
</tr>
</thead>
<tbody>
<tr>
<td>21 cm Lena</td>
<td>PSNR</td>
<td>29.65</td>
<td>29.65</td>
<td>29.61</td>
<td>29.92</td>
</tr>
<tr>
<td></td>
<td>SNR</td>
<td>23.97</td>
<td>23.96</td>
<td>23.95</td>
<td>24.25</td>
</tr>
<tr>
<td>21 cm Onion</td>
<td>PSNR</td>
<td>29.41</td>
<td>29.38</td>
<td>29.18</td>
<td>29.51</td>
</tr>
<tr>
<td></td>
<td>SNR</td>
<td>22.35</td>
<td>22.32</td>
<td>22.17</td>
<td>22.49</td>
</tr>
</tbody>
</table>

**Table 1.** Comparison of measures PSNR and SNR.
Figure 2. 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 3. 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 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).
the PSNR and SNR criteria for the second order total variation model compared with the two Chambolle algorithm.

References