TOTAL VARIATION PROJECTION WITH FIRST ORDER SCHEMES

Jalal M. Fadili

GREYC CNRS,
ENSICAEN-Université de Caen
14050 Caen France

Gabriel Peyré

CEREMADE CNRS,
Université Paris-Dauphine
75016 Paris France

ABSTRACT

This paper proposes a new class of algorithms to compute the projection onto the set of images with a total variation bounded by a constant. The projection is computed on a dual formulation of the problem that is minimized using either a one-step gradient descent method or a multi-step Nesterov scheme. This yields iterative algorithms that compute soft thresholding of the dual vector fields. We show the convergence of the method with a convergence rate of $O(1/k)$ for the one step method and $O(1/k^2)$ for the multi-step one, where k is the iteration number. The projection algorithm can be used as a building block in several applications, and we illusrtate it by solving linear inverse problems under total variation constraint. Numerical results show that our algorithm competes favorably with state-of-the-art TV projection methods to solve denoising, inpainting and deblurring problems.

Index Terms— Total variation, projection, duality, forward-backward splitting, inverse problems.

1. INTRODUCTION

Total variation is a well known image prior introduced by Rudin, Osher and Fatemi [1]. For a discrete image $f \in \mathbb{R}^N$ of $N = n \times n$ pixels, the discrete total variation reads $\|f\|_{\text{TV}} = \|\nabla f\|_1$, where the ℓ^1 of a vector field $u = (u_1, u_2) \in \mathbb{R}^{N\times 2}$ is defined as

$$\|u\|_1 = \sum_{i,j} \sqrt{u_1[i,j]^2 + u_2[i,j]^2},$$

and the discrete gradient

$$\nabla f[i,j] = (f[i+1,j] - f[i,j], f[i,j+1] - f[i,j]),$$

with Neumann boundary conditions. The adjoint of the gradient is $\nabla^* = -\operatorname{div}$, where the discrete divergence of a vector field $u \in \mathbb{R}^{N\times 2}$ is $\operatorname{div}(u) = u_1[i,j] - u_1[i-1,j] + u_2[i,j] - u_2[i,j-1]$, with Neumann boundary conditions.

The total variation is used as a regularization to denoise an image f_0 by solving the strongly convex problem [1]

$$\min_{f\in\mathbb{R}^N} \frac{1}{2}\|f - f_0\|^2 + \lambda\|f\|_{\text{TV}} \, . \qquad (1)$$

The regularization weight λ should be tuned to match the noise level contaminating f_0. Several algorithms have been proposed to solve this problem, see for instance [2, 3].

TV projection for denoising. Much less work has focused on denoising an image f_0 by projecting it on a total variation ball of radius $\tau < \|f_0\|_{\text{TV}}$, which requires to solve

$$f^\star = \underset{f\in\mathbb{R}^N,\, \|f\|_{\text{TV}}\leqslant\tau}{\operatorname{argmin}} \|f - f_0\|. \qquad (2)$$

Such a formulation might be preferable over (1) when little is known about the noise level perturbing f_0, but when an estimate τ of the total variation of the clean image is known. Computing the solution of (2) with a fast algorithm is thus important for a denoising application [4].

An iterative projected sub-gradient method was introduced in [4]. We here propose a different algorithm that is based on a dual formulation of the primal projection problem. This bears similarities with Chambolle's algorithm [2] that solves the primal regularization (1) using a dual projection. Our dual problem is then solved using two first-order iterative schemes: *one-step* forward-backward splitting revitalized in [6], and accelerated *multi-step* Nesterov scheme [7]. These two algorithms require only the computation of a soft thresholding applied to the dual vector fields at each iteration.

TV projection for inverse problems. Total variation projection might be useful as a proxy to solve more challenging inverse problems. Popular inverse problems such as inpainting and deblurring have been the subject of a flurry of research activity. These inverse problems can be regularized with a total variation constraint [4]. Sections 4.2 and 4.3 are devoted to show how these constrained inverse problems can be solved efficiently using a projected gradient descent iteration, whose projector is computed with our algorithm.

2. TV PROJECTION ALGORITHM

2.1. Dual Total Variation Projection

The goal of this paper is to solve the primal projection problem (2). Obviously $f^\star = f_0$ if $\tau \geqslant \|f_0\|_{\text{TV}}$. The following proposition, whose proof can be found in [5], shows that the primal constraint problem (2) is recast as a penalized dual optimization.

Proposition 1. *For any $f \in \mathbb{R}^N$, the primal solution is recovered as $f^\star = f_0 - \operatorname{div}(u^\star)$ where $u^\star$ is the solution of*

$$\min_{u \in \mathbb{R}^{N\times 2}} \frac{1}{2}\|f_0 - \operatorname{div}(u)\|^2 + \tau \|u\|_\infty \ . \tag{3}$$

$$\textit{where} \quad \|u\|_\infty = \max_{i,j} \sqrt{u_1[i,j]^2 + u_2[i,j]^2}.$$

2.2. First Order Schemes and Proximal Operator

As $\|f_0 - \operatorname{div}(u)\|^2$ is differentiable with Lipschitz-continuous gradient, and the set of solutions is not empty (by coercivity), (3) can be solved using first order algorithms, which are extension to non-differentiable functionals of the gradient descent. This paper considers a forward-backward iteration [6] (that is a special case of splitting schemes for maximal monotone operators), and a multi-step algorithm due to Nesterov [7].

Both algorithms requires the resolution of the Moreau-Yosida regularization of the functional $\kappa\|\cdot\|_\infty$, which is the unique minimizer of the convex problem

$$\operatorname{prox}_{\kappa\|\cdot\|_\infty}(u) = \underset{v \in \mathbb{R}^{N\times 2}}{\operatorname{argmin}} \frac{1}{2}\|u - v\|^2 + \kappa \|v\|_\infty. \tag{4}$$

Proposition 2, whose proof can be found in [5], shows that this proximal operator can be computed explicitly using a soft thresholding S_λ applied to the dual vector field u for a well chosen value of λ. Computing the precise value of λ for a given vector field $u \in \mathbb{R}^{N\times 2}$ requires the computation of ordered norms $d[0] \leqslant d[1] \leqslant \ldots \leqslant d[N-1]$, and of cumulated ordered norms

$$\{d[t]\}_{t=0}^{N-1} = \{\|u[i,j]\|\}_{i,j=0}^{n-1}, \quad D[s] = \sum_{t=s+1}^{N-1} d^r[t]. \tag{5}$$

Proposition 2. *For $u \in \mathbb{R}^{N\times 2}$, we have $\operatorname{prox}_{\kappa\|\cdot\|_\infty}(u) = u - S_\lambda(u)$ where*

$$S_\lambda(u)[i,j] = \max\left(1 - \frac{\lambda}{\|u[i,j]\|}, 0\right) u[i,j] \tag{6}$$

and $\lambda > 0$ is computed as

$$\lambda = d[t] + (d[t+1] - d[t]) \frac{D[t+1] - \kappa}{D[t+1] - D[t]} \tag{7}$$

where d and D are defined in (5)*, and where t is such that $D[t+1] \leqslant \kappa < D[t]$.*

2.3. Forward-backward Total Variation Projection

An iteration of the forward-backward projection algorithm (with descent step-size sequence $\mu > 0$) to solve (3) reads

$$u^{(k+1)} = \operatorname{prox}_{\mu\tau\|\cdot\|_\infty}\left(u^{(k)} + \mu \nabla \left(f_0 - \operatorname{div}(u^{(k)})\right)\right) \tag{8}$$

where the proximal operator for $\kappa = \mu\tau$ is computed thanks to Proposition 2.

Theorem 1, whose proof can be found in [5], ensures that the primal sequence $(f^{(k)})_{k\in\mathbb{N}}$ obtained from the dual $(u^{(k)})_{k\in\mathbb{N}}$ one converges to a solution of (2) at a rate $O(1/k)$.

Theorem 1. *Suppose that $\mu \in (0, 1/4)$ and $u^{(0)} \in \mathbb{R}^{N\times 2}$. There exists $C > 0$ such that*

$$\|f^{(k)} - f^\star\|^2 \leqslant C/k,$$

where $f^{(k)} = f_0 - \operatorname{div}(u^{(k)})$ and $u^{(k)}$ is defined in (8).

2.4. Nesterov Total Variation Projection

Y. Nesterov proposed in [7, 8] a multi-step scheme to optimize the sum of a smooth convex functional with Lipschitz gradient and a convex non-differentiable functional whose structure is simple (which means that its proximal operator can be computed explicitly).

The dual problem (3) fits into this framework. Algorithm 1 details the step of Nesterov scheme to minimize (3). Following [3], the algorithm uses the proximal operator of $\|\cdot\|_\infty$.

Theorem 2, whose proof can be found in [5], ensures that the sequence $(f^{(k)})_{k\in\mathbb{N}}$ converges to a solution of (2) at a rate $O(1/k^2)$.

Theorem 2. *Let $\mu \in]0, 1/4[$ and $u^{(0)} \in \mathbb{R}^{N\times 2}$. There exists $C > 0$ such that the sequence $f^{(k)}$ defined in Algorithm 1 satisfies*

$$\|f^{(k)} - f^\star\|^2 \leqslant C/k^2.$$

A major distinction between our work and the one of [3, 8] is that the non-differentiable part of our dual problem does not have a bounded domain, which is required in [3] to exhibit the convergence speed on the primal iterates.

3. INVERSE PROBLEMS

Image acquisition devices provide $p \leqslant N$ noisy and degraded measurements

$$y = \Phi f_0 + \varepsilon \in \mathbb{R}^p \ , \tag{9}$$

of an image $f_0 \in \mathbb{R}^N$, where ε is an additive noise. The linear operator Φ typically accounts for some blurring, sub-sampling or missing pixels so that the measured data y only captures a fraction of the information originally contained in

Algorithm 1: Nesterov total variation projection.

Initialization : $u^{(0)} \in \mathbb{R}^{N\times 2}$, $A_0 = 0$, $\xi^{(0)} = 0$, $\mu < 1/4$.
Main iteration:
while $\|u^{(k+1)} - u^{(k)}\| > \eta$ **do**

1. *First proximal computation:*
$$\upsilon^{(k)} = \text{prox}_{A_k\tau\|\cdot\|_\infty}(u^{(0)} - \xi^{(k)}) ,$$
where the proximal operator is computed as detailed in Proposition 2 with $\kappa = A_k\tau$.
2. Let $a_k = \left(\mu + \sqrt{\mu^2 + 4\mu A_k}\right)/2$ and $\omega^{(k)} = \frac{A_k u^{(k)} + a_k \upsilon^{(k)}}{A_k + a_k}$.
3. *Second proximation computation:*
$$\begin{aligned}\tilde{\omega}^{(k)} &= \omega^{(k)} - \frac{\mu}{2}\nabla\left(f_0 - \text{div}(\omega^{(k)})\right),\\ u^{(k+1)} &= \text{prox}_{\mu\tau/2\|\cdot\|_\infty}(\tilde{\omega}^{(k)}) ,\end{aligned}$$
where the proximal operator is computed as detailed in Proposition 2 with $\kappa = \mu\tau/2$.
4. Update $A_{k+1} = A_k + a_k$ and $\xi^{(k+1)} = \xi^{(k)} + a_k\nabla\left(f_0 - \text{div}(u^{(k+1)})\right)$.
5. $k \leftarrow k+1$, define $f^{(k)} = f_0 - \text{div}(u^{(k)})$.

Output: $f^\star = f_0 - \text{div}(u^{(k+1)})$.

the original image f_0 that one wishes to recover. This is in general an ill-posed inverse problem.

If the noise has a known bounded norm, it is possible to solve the constrained optimization problem

$$\min_{f\in\mathbb{R}^N} \|f\|_{\text{TV}} \quad \text{subject to} \quad \|\Phi f - y\| \leqslant \sigma , \qquad (10)$$

where σ is related to the noise norm $\|\varepsilon\|$. On the contrary, if little is known about the noise ε, but one has some prior guess τ of the total variation of the image, it is preferable to consider the (equivalent) problem

$$f^\star = \underset{f\in\mathbb{R}^N}{\text{argmin}} \|\Phi f - y\|^2 \quad \text{subject to} \quad \|f\|_{\text{TV}} \leqslant \tau, \qquad (11)$$

where the minimum is not necessarily unique unless Φ is injective.

3.1. Projected Gradient Descent

Given the structure of (11), we suggest a projected gradient descent iteration to solve it. Formally, the iteration is

$$f^{(\ell+1)} = \text{Proj}_{\|\cdot\|_{\text{TV}}\leqslant\tau}\left(f^{(\ell)} + \gamma_\ell\Phi^*(y - \Phi f^{(\ell)})\right), \qquad (12)$$

where $\text{Proj}_{\|\cdot\|_{\text{TV}}\leqslant\tau}$ is the projector on the TV ball of radius τ.

The following theorem, whose proof can be found in [5], ensures the convergence of this iteration even if $\text{Proj}_{\|\cdot\|_{\text{TV}}\leqslant\tau}$ is not computed exactly but using the inner-iteration in (8). Let $a_\ell \in \mathbb{R}^N$ be an error term that models the inexact computation of $\text{Proj}_{\|\cdot\|_{\text{TV}}\leqslant\tau}$ at iteration ℓ.

Theorem 3. *Suppose that* $0 < \inf_\ell \gamma_\ell \leqslant \sup_\ell \gamma_\ell < 2/\|\Phi\|^2$, *where* $\|\Phi\|$ *is the spectral norm of* Φ, *and* $\sum_{\ell\in\mathbb{N}}\|a_\ell\| < \infty$. *Then* $f^{(\ell)}$ *converges to a minimizer* $f^\star$ *of* (11).

Bringing together (12) and the TV projection iteration (8), leads to our TV projection algorithm to solve inverse problems.

Although problem (11) could be solved with a Nesterov scheme, there is no proof of stability for such a scheme. This is the main reason why this article considers only a projected gradient descent.

4. NUMERICAL EXAMPLES

4.1. Denoising

We first tested our total variation algorithm for denoising, where $y = f_0 + \varepsilon$ is an image of $N = 512^2$ pixels contaminated by an additive white Gaussian noise (AWGN) ε of standard deviation $0.06\|f_0\|_\infty$.

Fig. 1 shows projections $f^\star$ computed with our algorithm for a decreasing value of the constraint τ, so that only the strongest edges are present in the projected image. Fig. 2(a) compares the convergence speed of our one-step and multi-step dual projection algorithms summarized with the sub-gradient projection method of [4]. The one-step algorithm converges slightly faster compared to the sub-gradient projection. Moreover, and as predicted by our convergence analysis, the multi-step projection algorithm clearly outperforms the two other methods. Fig. 2(b) shows the evolution of the total variation of $f^{(k)}$.

4.2. Inpainting

Inpainting aims at restoring an image f_0 from which a set of pixels is missing. It corresponds to the inversion of (9) where Φ acts as a binary mask. In this case, $\|\Phi\| = 1$ and we use $\gamma_\ell \equiv 1$ for the projected gradient descent. Fig. 3, top, shows an example of damaged image y of $N = 512^2$ pixels, with 70% of randomly removed pixels. The noise is AWGN with $\|\varepsilon\| = 0.05\|f_0\|$. The total variation constraint is set to $\tau = 0.6\|f_0\|_{\text{TV}}$. The number of inner iterations for the projection step in (12) is controlled by setting the convergence tolerance of (8) to 10^{-2}. Roughly between 10 to 20 inner iterations of dual projections are required to maintain the total variation constraint at each outer iteration ℓ. Fig. 4(a) shows the decay in log scale of the residual error, that exhibits a linear convergence speed.

4.3. Deblurring

We finally illustrate our algorithm on a deblurring problem where Φ is a convolution by a unit-mass Gaussian kernel of width 4 pixels. Thus $\|\Phi\| = 1$. The noise is again AWGN with $\|\varepsilon\| = 0.02\|f_0\|$. We use a gradient step-size $\gamma_\ell \equiv 1.9$ in the iteration (12). The total variation constraint is set to

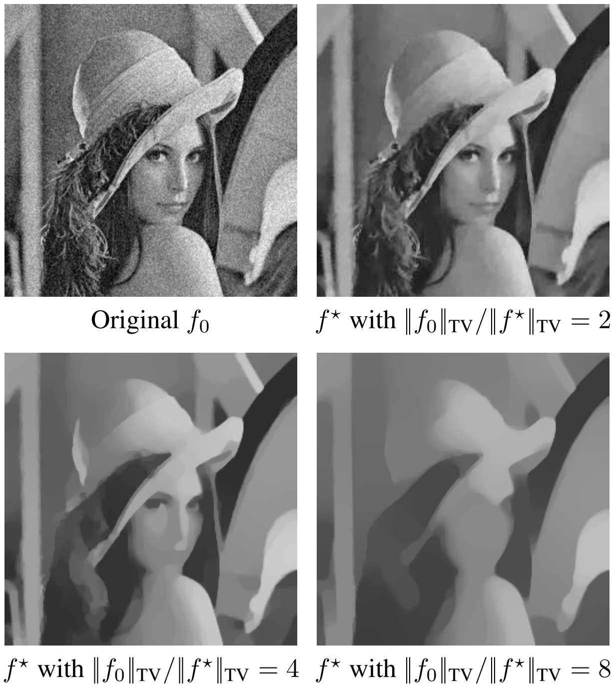

Fig. 1. *Examples of total variation projections computed with our algorithm.*

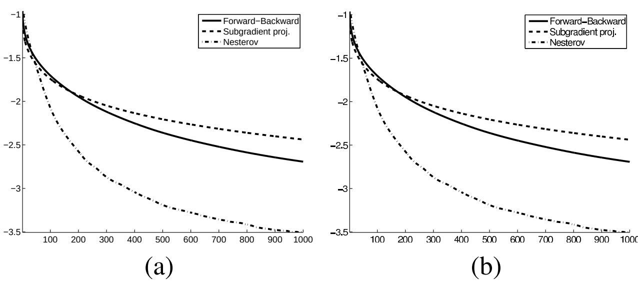

Fig. 2. *Decay with k of (a) the error $\log_{10}(\|f^{(k)} - f^\star\|/\|f^\star\|)$ and (b) of the total variation error $\|f^{(k)}\|_{TV}/\tau - 1$ for the dual projection* (8) *(solid line) and the sub-gradient projection [4] (dashed line). Here $\tau = \|f_0\|_{TV}/4$.*

$\tau = 0.6\|f_0\|_{\mathrm{TV}}$. Fig. 4(b) shows the decay in log scale of the residual error.

Conclusion

This paper proposes a new class of algorithms to compute the projection of an image on a total variation ball. The algorithm solves an unconstrained dual formulation of the problem, and uses iterative soft thresholding on the gradient field. These schemes are quite general, and extends to any positively 1-homogeneous functional for which the dual can be easily computed. It also generalizes to arbitrary dimension. A projected gradient descent uses this projection to solve linear inverse problems under a total variation constraint. Additional structural constraints could also be incorporated as well using for instance Douglas-Rachford splitting iteration.

Acknowledgements. This works is supported by the NatImages ANR grant ANR-08-EMER-009. We warmly thank Pierre Weiss for discussions during preparation of this manuscript.

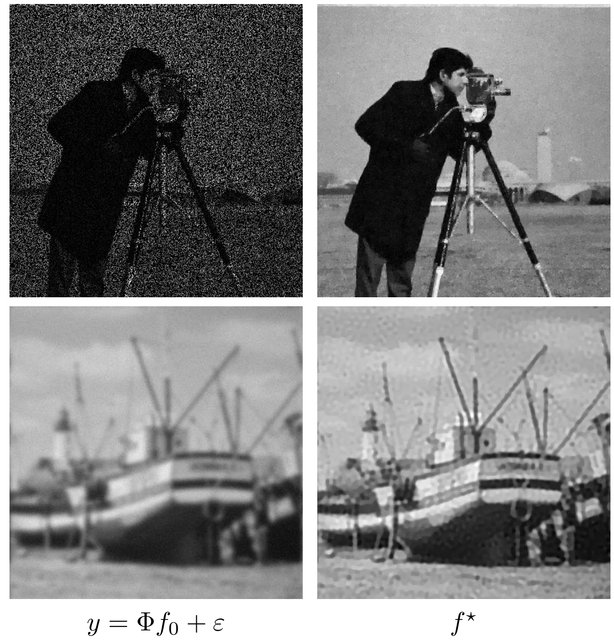

Fig. 3. *Examples of inpainting (first row) and deblurring (bottom row) resolution using TV constraints.*

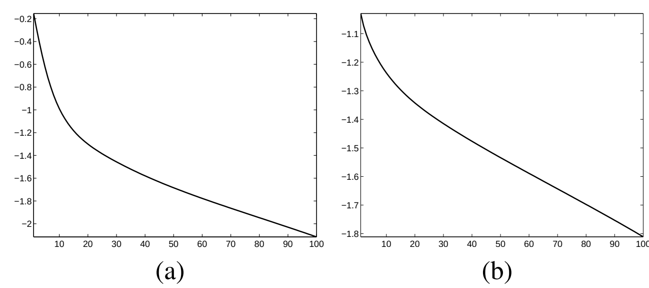

Fig. 4. *Decay of the error $\log_{10}(\|f^{(\ell)} - f^\star\|/\|f_0\|)$ for inpainting (a) and deblurring (b).*

5. REFERENCES

[1] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, no. 1-4, pp. 259–268, 1992.

[2] A. Chambolle, "An algorithm for total variation minimization and applications," *Journal of Mathematical Imaging and Vision*, vol. 20, pp. 89–97, 2004.

[3] P. Weiss, L. Blanc-Féraud, and G. Aubert, "Efficient schemes for total variation minimization under constraints in image processing," *SIAM Journal on Scientifigc Computing*, 2009, to appear.

[4] P.-L. Combettes and J.-C. Pesquet, "Image restoration subject to a total variation constraint," *IEEE Trans. Image Proc.*, vol. 13, no. 9, pp. 1213–1222, 2004.

[5] J. Fadili and G. Peyré, "Total variation projection with first order schemes," *Preprint Hal-00380491*, 2009.

[6] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *SIAM Journal on Multiscale Modeling and Simulation*, vol. 4, no. 4, 2005.

[7] Y. Nesterov, "Gradient methods for minimizing composite objective function," *Preprint UCL*, 2007.

[8] Y. Nesterov, "Smooth minimization of non-smooth functions," *Math. Program.*, vol. 103, no. 1, Ser. A, pp. 127–152, 2005.