HAL open science

Total Variation Projection with First Order Schemes

Jalal M. Fadili, Gabriel Peyré

To cite this version:

Jalal M. Fadili, Gabriel Peyré. Total Variation Projection with First Order Schemes. 2009. hal-00380491v2

HAL Id: hal-00380491

https://hal.science/hal-00380491v2

Preprint submitted on 6 May 2009 (v2), last revised 11 May 2009 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Total Variation Projection with First Order Schemes

Jalal M. Fadili and Gabriel Peyré

***Abstract*—This article proposes a new algorithm to compute the projection on the set of images whose total variation is bounded by a constant. The projection is computed through a dual formulation that is solved by first order non-smooth optimization methods. This yields an iterative algorithm that computes iterative soft thresholding of the dual vector fields. This projection algorithm can then be used as a building block in a variety of applications such as solving inverse problems under a total variation constraint, or for texture synthesis. Numerical results show that our algorithm competes favorably with state-of-the-art TV projection methods to solve denoising, texture synthesis, inpainting and deconvolution problems.**

***Index Terms*—Total variation, projection, duality, proximal operator, forward-backward splitting, Nesterov scheme, inverse problems.**

I. Introduction

TOTAL variation is a well known image prior introduced by Rudin, Osher and Fatemi [1]. For a differentiable function $f : \Omega = [0,1]^2 \to \mathbb{R}$, it is computed as $\|f\|_{\mathrm{TV}} = \int_\Omega |\nabla f|$, and can be extended to the space $\mathrm{BV}([0,1]^2)$ that contains functions with discontinuities.

The total variation is used as a regularization to denoise an image f_0 by solving the strictly convex problem

$$\min_{f \in \mathrm{BV}([0,1]^2)} \frac{1}{2}\|f - f_0\|^2 + \lambda \|f\|_{\mathrm{TV}}, \tag{1}$$

as originally proposed in [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], [4], [5], [6], [7], [8], [9]. Such primal, dual, or primal-dual schemes for denoising are often a building block for solving more complex inverse problems; see e.g. [10], [11].

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 $\tau < \|f_0\|_{\mathrm{TV}}$, which requires to solve[1]

$$\min_{\|f\|_{\mathrm{TV}} \leqslant \tau} \|f - f_0\|. \tag{2}$$

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 denoising application as has been advocated in [12].

Jalal M. Fadili is with GREYC CNRS-ENSICAEN-Université de Caen, 14050 Caen Cedex, France., email: jalal.fadili@greyc.ensicaen.fr

Gabriel Peyré is with the CEREMADE CNRS-Université Paris-Dauphine, 75775 Paris Cedex 16 France, email: gabriel.peyre@ceremade.dauphine.fr

[1]This constraint set is obviously a closed convex set.

An iterative projected sub-gradient method was introduced in [13], [12]. We propose in this paper a different algorithm that is based on a dual regularization of the primal projection problem. This bears similarities with Chambolle's algorithm [6] that solves the primal TV regularization (1) using a dual projection. Our dual problem is then solved using two first-order iterative schemes: *one-step* forward-backward splitting which dates back to [14], [15] and recently revitalized in [16], and accelerated *multi-step* Nesterov scheme [17]. Both of these algorithms require only the computation of a soft thresholding applied to the dual vector fields at each iteration.

Total variation projection also have far-reaching applications beyond denoising. For example, the extraction of Cheeger sets in landslides modeling [18] can be relaxed as a TV projection problem with boundary constraints, that has been recently solved using sub-gradient projection [19].

TV projection for synthesis. Classical synthesis methods constrain wavelet coefficients [20], [21] and are suitable to model some natural phenomena. Computer graphics methods do not rely on statistical modeling and generate a texture through a consistent copy of pixels from an example image [22], [23]. Higher order statistical models [24], [25] improve the visual quality of synthesis by capturing geometric singularities. Section VI-B shows that a total variation constraint can also be used to enhance the sharpness of edges in texture synthesis.

TV projection for inverse problems. Total variation projection might be useful as a proxy to solve more challenging inverse problems. Popular linear inverse problems such as inpainting and deconvolution have been the subject of a flurry of research activity where TV has been extensively used to regularize them.

Classical methods for inpainting use partial differential equations that propagate the information from the boundary of the missing region to its interior, see for instance [26], [27], [28]. Tensor diffusion makes use of a geometric layer to drive the diffusion, see [29]. TV regularization was proposed in [30] for inpainting. Sparsity over a redundant frame has also been successfully used to regularize the inpainting problem [31], [32].

There is an extensive literature on the deconvolution problem image processing, and many algorithms have been developed to tackle it. For instance, TV penalty has been used as a regularization term within a variational framework, see e.g. [33], [10], [11], [34], [9]. Wavelet-based or more generally sparsity-based deconvolution methods have also received considerable attention over the last decade, see e.g. [35], [36], [37], [38].

Inverse problems such as inpainting or deconvolution can be regularized with a total variation constraint [12]. Sections VI-C and VI-D are devoted to show how these TV-constrained

inverse problems can be solved efficiently using a projected gradient descent iteration, whose projector is computed with our algorithm.

II. Notation and elements of convex analysis

Throughout the paper, an image f of $N = n \times n$ pixels is a vector in $\mathbb{R}^N$. We denote by $\|.\|$ the norm induced by the inner product $\langle ., . \rangle$ in $\mathbb{R}^N$. We let $\mathcal{U} = \mathbb{R}^N \times \mathbb{R}^N$ be the space of vector fields with associated inner product $\langle u, v \rangle_{\mathcal{U}} = \sum_{0 \leqslant i,j \leqslant n-1} (u_1[i,j]v_1[i,j] + u_2[i,j]v_2[i,j])$, $\forall u = (u_1, u_2)$ and $v = (v_1, v_2) \in \mathcal{U}$. The ℓ^1 and ℓ^∞ norms of a vector field $u = (u_1, u_2) \in \mathcal{U}$ are respectively $\|u\|_1 = \sum_{0 \leqslant i,j \leqslant n-1} |u[i,j]|$ and $\|u\|_\infty = \max_{i,j} |u[i,j]|$, where $|u[i,j]| := \sqrt{u_1[i,j]^2 + u_2[i,j]^2}$.

Let C a nonempty convex set. The indicator function $\imath_C$ of C is

$$\imath_C(z) = \begin{cases} 0, & \text{if } z \in C \ , \\ +\infty, & \text{otherwise.} \end{cases}$$

The domain of a function φ is defined by $\mathrm{dom}(\varphi) = \{z : \varphi(z) < +\infty\}$ and φ is proper if $\mathrm{dom}(\varphi) \neq \emptyset$. The conjugate of a proper, lower-semicontinuous and convex function φ is the proper closed convex function φ^* defined by

$$\varphi^*(z) = \sup_{z \in \mathrm{dom}(\varphi)} \langle w, z \rangle - \varphi(z) \ ,$$

and we have the bi-conjugate $\varphi^{**} = \varphi$.

We also define the notion of a proximity operator, which was introduced in [39] as a generalization of convex projection operator. For every z, the function $w \mapsto \frac{1}{2}\|w - z\|^2 + \varphi(w)$ achieves its infimum at a unique point denoted by $\mathrm{prox}_\varphi z$. The uniquely-valued operator prox_φ thus defined is the *proximity operator* of φ.

III. Discrete Dual Formulation

The optimization (2) is computed numerically by discretizing the gradient operator and the total variation to project an image of $N = n \times n$ pixels.

A. Discrete Total Variation

A discretized gradient for an image $f \in \mathbb{R}^N$ is defined as $\nabla f[i,j] = (\partial_x f[i,j], \partial_y f[i,j])$, where

$$\partial_x f[i,j] = \begin{cases} f[i+1,j] - f[i,j] & \text{if} \quad 0 \leqslant i < n-1, \\ 0 \text{ otherwise,} \end{cases}$$

$$\partial_y f[i,j] = \begin{cases} f[i,j+1] - f[i,j] & \text{if} \quad 0 \leqslant j < n-1, \\ 0 \text{ otherwise.} \end{cases}$$

The gradient is thus a vector field $\nabla f \in \mathcal{U}$. The discrete total variation is $\|f\|_{\mathrm{TV}} = \|\nabla f\|_1$ where the ℓ^1 norm of a vector field in $\mathcal{U}$ is defined in Section II.

The adjoint of the gradient is $\nabla^* = -\mathrm{div}$, where the divergence of a vector field $u = (u_1, u_2) \in \mathcal{U}$ is $\mathrm{div}(u) = \partial_x^* u_1 + \partial_x^* u_2$, with

$$\partial_x^* f[i,j] = \begin{cases} f[i,j] - f[i-1,j] & \text{if} \quad 0 < i < n, \\ 0 \text{ otherwise,} \end{cases}$$

$$\partial_y^* f[i,j] = \begin{cases} f[i,j] - f[i,j-1] & \text{if} \quad 0 < j < n, \\ 0 \text{ otherwise.} \end{cases}$$

B. Total Variation Projection

The goal of this paper is to compute the projection $f^\star$ of an image f_0 on a set of images with bounded variation

$$f^\star = \underset{f \in \mathbb{R}^N, \, \|f\|_{\mathrm{TV}} \leqslant \tau}{\mathrm{argmin}} \|f - f_0\|^2. \tag{3}$$

where $0 < \tau < \|f_0\|_{\mathrm{TV}}$ to avoid the trivial solution $f^\star = f_0$.

The following proposition shows that the primal constrained optimization problem (3) is recast into a penalized form.

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

$$\inf_{u \in \mathbb{R}^{N \times 2}} \frac{1}{2}\|f_0 - \mathrm{div}(u)\|^2 + \tau \|u\|_\infty \ . \tag{4}$$

Proof: Let's introduce the dual variable $u \in \mathcal{U}$. First, it is easy to show that the Fenchel conjugate of the ℓ^∞ norm is the indicator function of the ℓ^1 norm. Indeed, by the Hölder inequality we have $\langle u, v \rangle_{\mathcal{U}} \leqslant \|u\|_\infty \|v\|_1$ [2], we then get

$$\begin{aligned} (\tau\|.\|_\infty)^*(v) &= \sup_{u \in \mathcal{U}} \langle u, v \rangle_{\mathcal{U}} - \tau\|u\|_\infty \\ &= \sup_{\rho > 0} \sup_{\|u\|_\infty = \rho} \langle u, v \rangle_{\mathcal{U}} - \tau\rho \\ &= \sup_{\rho > 0} \rho\left(\|v\|_1 - \tau\right) = \imath_{\{\|v\|_1 \leqslant \tau\}}(u) \ . \end{aligned}$$

Thus, using the bi-conjugate relation, one can write the TV-ball indicator function as

$$\imath_{\{\|\cdot\|_{\mathrm{TV}} \leqslant \tau\}}(f) = \sup_{u \in \mathcal{U}} \langle u, \nabla f \rangle_{\mathcal{U}} - \tau\|u\|_\infty.$$

This allows to rewrite (3) as This allows to rewrite (3) as

$$\min_{f \in \mathbb{R}^N} \tfrac{1}{2}\|f - f_0\|^2 + \imath_{\{\|\cdot\|_{\mathrm{TV}} \leqslant \tau\}}(f) \tag{5}$$

$$= \sup_{u \in \mathcal{U}} -\tau\|u\|_\infty + \min_{f \in \mathbb{R}^N} \langle u, \nabla f \rangle_{\mathcal{U}} + \tfrac{1}{2}\|f - f_0\|^2$$

$$= -\inf_{u \in \mathcal{U}} \tau\|u\|_\infty - \min_{f \in \mathbb{R}^N} -\langle \mathrm{div}(u), f \rangle + \tfrac{1}{2}\|f - f_0\|^2 \tag{6}$$

The inner minimization involves a strongly convex function, see e.g. [40] for the definition of strong convexity, whose unique primal solution $f^\star$ is recovered from the dual variable u as

$$f^\star = \underset{f \in \mathbb{R}^N}{\mathrm{argmin}} -\langle \mathrm{div}(u), f \rangle + \frac{1}{2}\|f - f_0\|^2 = f_0 + \mathrm{div}(u), \tag{7}$$

and

$$-\langle \mathrm{div}(u), f^\star \rangle + \frac{1}{2}\|f^\star - f_0\|^2 = -\frac{1}{2}\|f_0 + \mathrm{div}(u)\|^2 + \frac{1}{2}\|f_0\|^2 \ . \tag{8}$$

Combining (5) and (8) with the obvious change of sign on u leads to the optimization problem (4).

■

Remark 1. *Proposition 1 can be alternatively proved using Fenchel-Rockafellar duality formula well-known in convex analysis, see e.g. [41, Section III.4]. We have chosen this line of proof for convenience and the sake of accessibility to the reader.*

[2] An alternative proof can be found in e.g. [40, Proposition V.3.2.1].

IV. First-Order Schemes on the Dual Problem

In going from the primal problem (3) to the dual formulation (4), we have replaced a constrained problem with an unconstrained penalized form. The primal solution is easily recovered from the dual solution as $f^\star = f_0 - \mathrm{div}(u^\star)$. Furthermore, it turns out that the dual problem (4) is easier to solve with various first order non-smooth optimization schemes. Indeed, (4) involves a quadratic form which has a Lipschitz continuous gradient, and the non-differentiable ℓ^∞ term. Section IV-B presents a *one-step* forward-backward splitting iteration, as explained for instance in [15], [16], while Section IV-C proposes an accelerated *multi-step* scheme due to Nesterov, see [17] and also [7] for some applications to image processing. It turns out that both these algorithms involve the computation of the proximity operator associated to the ℓ^∞ norm that can be computed explicitly as explained in the following section.

A. ℓ^∞ Proximity Operator

Recall from Section II that the proximity operator $\mathrm{prox}_{\kappa\|\cdot\|_\infty}(u)$ associated to $\kappa\|\cdot\|_\infty$ amounts to solving

$$\mathrm{prox}_{\kappa\|\cdot\|_\infty}(u) = \underset{v\in\mathcal{U}}{\mathrm{argmin}}\ \frac{1}{2}\|u-v\|^2 + \kappa\|v\|_\infty. \tag{9}$$

Proposition 2 hereafter shows that $\mathrm{prox}_{\kappa\|\cdot\|_\infty}(u)$ is computed explicitly through a soft thresholding S_λ applied on the dual vector field u for a properly chosen value of λ.

To get the precise value of λ for a given vector field $u \in \mathcal{U}$, we need to compute $d[0] \leqslant d[1] \leqslant \ldots \leqslant d[N-1]$ that orders the set of norms

$$\{d[t]\}_{t=0}^{N-1} = \{|u[i,j]|\}_{i,j=0}^{n-1}, \tag{10}$$

and also the cumulated ordered norms

$$D[s] = \sum_{t=s+1}^{N-1} d[t]. \tag{11}$$

Proposition 2. *For $u \in \mathcal{U}$, with d and D as defined in* (10)-(11)*, we have $\mathrm{prox}_{\kappa\|\cdot\|_\infty}(u) = 0$ if $\|u\|_1 \leqslant \kappa$ and $\mathrm{prox}_{\kappa\|\cdot\|_\infty}(u) = u - S_\lambda(u)$ otherwise, where*

$$S_\lambda(u)[i,j] = \max\left(1 - \frac{\lambda}{|u[i,j]|}, 0\right) u[i,j] \tag{12}$$

and $\lambda > 0$ is given by

$$\lambda = d[t] + (d[t+1] - d[t])\frac{D[t+1]-\kappa}{D[t+1]-D[t]} \tag{13}$$

where t is such that $D[t+1] \leqslant \kappa < D[t]$.

Proof: Using a similar proof to that of Proposition 1, we have the relation between the primal and dual minimization problems

$$\begin{aligned} &\mathrm{prox}_{\kappa\|\cdot\|_\infty}(u) = \underset{v\in\mathcal{U}}{\mathrm{argmin}}\ \tfrac{1}{2}\|u-v\|^2 + \kappa\|v\|_\infty \\ \iff\quad &\mathrm{Proj}_{\{\|\cdot\|_1\leqslant\kappa\}}(u) = \underset{w\in\mathcal{U}}{\mathrm{argmin}}\ \tfrac{1}{2}\|u-w\|^2 + \imath_{\{\|\cdot\|_1\leqslant\kappa\}}(w)\ . \end{aligned}$$

where $\mathrm{Proj}_{\{\|\cdot\|_1\leqslant\kappa\}}(u)$ is the orthogonal projection of u onto the closed ℓ^1 ball in $\mathcal{U}$ of radius κ. In the same vein as (7), the relation between the primal and dual solutions (both are unique here) yields

$$\mathrm{prox}_{\kappa\|\cdot\|_\infty}(u) = u - \mathrm{Proj}_{\{\|\cdot\|_1\leqslant\kappa\}}(u)\ . \tag{14}$$

If $\|u\|_1 \leqslant \kappa$, then $\mathrm{prox}_{\kappa\|\cdot\|_\infty}(u) = 0$. Otherwise, the projection $\mathrm{Proj}_{\{\|\cdot\|_1\leqslant\kappa\}}(u)$ is computed by finding the Lagrange multiplier $\lambda(\kappa)$ such that[3]

$$\mathrm{Proj}_{\{\|\cdot\|_1\leqslant\kappa\}}(u) = \underset{v}{\mathrm{argmin}}\ \frac{1}{2}\|u-v\|^2 + \lambda(\kappa)\|v\|_1. \tag{15}$$

As noticed for instance in [42] for wavelet thresholding, the solution of (15) has a closed-form known as soft-thresholding extended to vector fields

$$\mathrm{Proj}_{\{\|\cdot\|_1\leqslant\kappa\}}(u) = S_{\lambda(\kappa)}(u).$$

Observe that $\|S_\lambda(u)\|_1 = \sum_{|u[i,j]|>\lambda}(|u[i,j]| - \lambda)$ is a piecewise affine and decreasing function of λ with slope changing at the ordered values $d[t]$. One can then check that the value of λ that meets the constraint $\|S_\lambda(u)\|_1 = \kappa$ is the one given by (13). ■

Remark 2. *Equation* (14) *can be proved alternatively using Moreau decomposition which allows to conclude that $\mathrm{prox}_{\varphi^*}(u) = u - \mathrm{prox}_\varphi(u)$, see for instance [43] and [16, Lemma 2.10].*

B. Forward-backward Projection Algorithm

The projection $f^\star = f_0 - \mathrm{div}(u^\star)$ is computed by solving the dual unconstrained optimization problem (4). Owing to Lipschitz differentiability of the quadratic term, the dual problem verifies the necessary prerequisite to be solved with the one-step forward-backward splitting recursion.

The forward-backward scheme can be written in a compact form with descent step-size $\mu > 0$ [4]

$$u^{(k+1)} = \mathrm{prox}_{\mu\tau\|\cdot\|_\infty}\left(u^{(k)} - \mu\nabla\left(f_0 - \mathrm{div}(u^{(k)})\right)\right), \tag{16}$$

where the proximity operator for $\kappa = \mu\tau$ is defined in (9). The overall algorithm to minimize (3) is summarized in Algorithm 1.

1) Convergence analysis: Theorem 1 ensures that the sequence $(f^{(k)})_{k\in\mathbb{N}}$ obtained from Algorithm 1 converges to the solution of (3) with a convergence rate $O(1/k)$.

Theorem 1. *Suppose that $\mu \in (0, 1/4)$. Let $u^{(0)} \in \mathbb{R}^{N\times 2}$. The sequence of iterates $f^{(k)} = f_0 - \mathrm{div}(u^{(k)})$, where $u^{(k)}$ is the dual sequence in* (16)*, converges to $f^\star$. More precisely, there exists a $C > 0$ such that*

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

Proof: By coercivity, the set of solutions of (4) is non empty. Moreover, let $|||\,\mathrm{div}\,|||$ be the operator norm of the discrete divergence. The term $\|f_0 - \mathrm{div}(u)\|^2/2$ is differentiable whose gradient is Lipschitz continuous with Lipschitz constant $|||\,\mathrm{div}\,|||^2 \leqslant 8$ [6]. Thus, by continuity of the mapping

[3] Actually, there is a bijection between λ and κ such that the constrained and Lagrangian problems share the same solution.

[4] The descent step-size can be even a sequence $\mu_t > 0$ that varies from one iteration to another.

Algorithm 1: Forward-backward total variation projection.

Initialization: choose some $u^{(0)} \in \mathcal{U}$, $\mu \in (0, 1/4)$, set $k = 0$.
Main iteration:
while $\|u^{(k+1)} - u^{(k)}\| > \eta$ **do**

1) *Gradient descent step:* compute
$$\tilde{u}^{(k)} = u^{(k)} - \mu \nabla \left(f_0 - \operatorname{div}(u^{(k)}) \right).$$
2) *TV correction step:* compute d and D as defined in (10) and (11) and set λ as in (13) with $\kappa = \mu\tau$. Define
$$u^{(k+1)} = \tilde{u}^{(k)} - S_\lambda(\tilde{u}^{(k)}),$$
using the soft thresholding operator (12).
3) $k = k + 1$.

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

$u \mapsto f = f_0 - \operatorname{div}(u)$, and since the projection $f^\star$ is unique, applying [44, Corollary 6.5], we deduce that the sequence of iterates $f^{(k)}$ converges to $f^\star$ provided that the step-size satisfies $0 < \underline{\mu} \leqslant \mu \leqslant \overline{\mu} < 1/4 \leqslant 2/|||\operatorname{div}|||^2$.

Let's now turn to the convergence rate. We let $J(f)$ and $\tilde{J}(u)$ be the primal and the dual objectives given in (3) and (4), namely

$$\begin{aligned} J(f) &= \underbrace{\frac{1}{2}\|f - f_0\|^2}_{H(f)} + \underbrace{\imath_{\{\|\cdot\|_{\mathrm{TV}} \leqslant \tau\}}(f)}_{G(\nabla f)}, \\ \tilde{J}(u) &= \underbrace{\frac{1}{2}\|f_0 - \operatorname{div}(u)\|^2 - \frac{1}{2}\|f_0\|^2}_{H^*(\operatorname{div}(u))} + \underbrace{\tau\|u\|_\infty}_{G^*(u)}, \end{aligned} \tag{17}$$

where H^* and G^* are the conjugates of H and G respectively. Recall that from Proposition 1, we have

$$\min_f \; J(f) = -\inf_u \; \tilde{J}(u) \iff \forall\, (f, u) \in \mathbb{R}^N \times \mathcal{U} \tag{18}$$
$$J(f) \geqslant J(f^\star) = -\tilde{J}(u^\star) \geqslant -\tilde{J}(u)\ , \tag{19}$$

and from the primal-dual solutions relationship

$$H(f^\star) + H^*(\operatorname{div}(u^\star)) = \langle f^\star, \operatorname{div}(u^\star)\rangle \tag{20}$$
$$\iff \qquad f^\star = (\mathrm{D_u}\, H^*)(\operatorname{div}(u^\star))\ , \tag{21}$$

where $\mathrm{D_u}\, H^*$ is the gradient of H^* with respect to u.

Relation (16) is equivalent to the following inclusion

$$u^{(k)} - \mu \nabla(\mathrm{D_u}\, H^*)(\operatorname{div}(u^{(k)})) - u^{(k+1)} \in \partial(\mu G^*)(u^{(k+1)})\ , \tag{22}$$

where ∂G^* is the subdifferential of G^*; see [40] for an introduction to subdifferential calculus. Since H^* is differentiable, for all $(u, v) \in \mathcal{U} \times \mathcal{U}$

$$\begin{aligned} \tilde{J}(v) = \; & H^*(\operatorname{div}(u)) + G^*(v) \\ & + \langle -\nabla(\mathrm{D_u}\, H^*)(\operatorname{div}(u)),\, v - u\rangle_{\mathcal{U}} \\ & + \Big(H^*(\operatorname{div}(v)) - H^*(\operatorname{div}(u)) \\ & + \langle -\nabla(\mathrm{D_u}\, H^*)(\operatorname{div}(u)),\, u - v\rangle_{\mathcal{U}} \Big)\ . \end{aligned}$$

By the Cauchy-Schwartz inequality, and as $\mathrm{D_u}\, H^*$ is 1-Lipschitz continuous

$$\begin{aligned} \langle -\nabla(\mathrm{D_u}\, H^*)(\operatorname{div}(u)),\, u - v\rangle_{\mathcal{U}} = \; & \int_0^1 \big\langle -\nabla(\mathrm{D_u}\, H^*)(\operatorname{div}(u + \tau(v - u)) \\ & + \nabla(\mathrm{D_u}\, H^*)(\operatorname{div}(u)), v - u \big\rangle_{\mathcal{U}}\, d\tau \\ \leqslant \; & \frac{|||\operatorname{div}|||^2}{2}\|v - u\|^2\ . \end{aligned}$$

Applying this inequality with $v = u^{(k+1)}$ and $u = u^{(k)}$, we obtain

$$\begin{aligned} \tilde{J}(u^{(k+1)}) \leqslant \; & H^*(\operatorname{div}(u^{(k)})) + G^*(u^{(k+1)}) \\ & + \langle -\nabla(\mathrm{D_u}\, H^*)(\operatorname{div}(u^{(k)})),\, u^{(k+1)} - u^{(k)}\rangle_{\mathcal{U}} \\ & + \frac{|||\operatorname{div}|||^2}{2}\|u^{(k+1)} - u^{(k)}\|^2\ . \end{aligned}$$

From (20), $\mu \in (0, 1/4) \subseteq \left(0, 2/|||\operatorname{div}|||^2\right)$ and the definition of the conjugate G^*, we obtain

$$\begin{aligned} \tilde{J}(u^{(k+1)}) \; & \leqslant \; -H(f^{(k)}) + G^*(u^{(k+1)}) - \langle \nabla f^{(k)},\, u^{(k+1)}\rangle_{\mathcal{U}} \\ & \quad + \frac{1}{\mu}\|u^{(k+1)} - u^{(k)}\|^2 \\ & \leqslant \; -H(f^{(k)}) - G(\nabla f^{(k)}) + \frac{1}{\mu}\|u^{(k+1)} - u^{(k)}\|^2 + \\ & \max_{u \in \operatorname{dom}(G^*)} \; \langle -\nabla f^{(k)},\, u^{(k+1)} - u\rangle_{\mathcal{U}} - G^*(u) + G^*(u^{(k+1)})\ . \end{aligned}$$

From the inclusion (22) (see also [45, Lemma 1]), we have the inequality

$$\begin{aligned} & \langle -\nabla f^{(k)},\, u^{(k+1)} - u\rangle_{\mathcal{U}} - G^*(u) + G^*(u^{(k+1)}) \\ \leqslant \; & \tfrac{1}{\mu}\langle u^{(k)} - u^{(k+1)},\, u^{(k+1)} - u\rangle_{\mathcal{U}},\ \forall u \in \mathcal{U}\ . \end{aligned} \tag{23}$$

Hence using (19) and the Cauchy-Schwartz inequality, we get

$$\begin{aligned} -J(f^\star) \; & \leqslant \; \tilde{J}(u^{(k+1)}) \leqslant -J(f^{(k)}) + \frac{1}{\mu}\|u^{(k+1)} - u^{(k)}\|^2 \\ & + \; \frac{1}{\mu} \max_{u \in \operatorname{dom}(G^*)} \langle u^{(k)} - u^{(k+1)},\, u^{(k+1)} - u\rangle_{\mathcal{U}} \\ & = \; -J(f^{(k)}) + \frac{c+1}{\mu}\|u^{(k+1)} - u^{(k)}\|^2\ , \end{aligned} \tag{24}$$

where c is a finite constant for which the Cauchy-Schwartz inequality becomes an equality for all $k \geqslant 0$.

From [17, Theorem 1] and (19), we show that

$$\begin{aligned} \|u^{(k+1)} - u^{(k)}\|^2 \; & \leqslant \; \left(1/\mu - |||\operatorname{div}|||^2/2\right)^{-1} \left(\tilde{J}(u^{(k)}) - \tilde{J}(u^{(k+1)})\right) \\ & \leqslant \; \left(1/\mu - |||\operatorname{div}|||^2/2\right)^{-1} \left(\tilde{J}(u^{(k)}) - \tilde{J}(u^\star)\right) \end{aligned} \tag{25}$$

and using [17, Theorem 4], the convergence rate over $\tilde{J}$ is such that

$$\|u^{(k+1)} - u^{(k)}\|^2 \; \leqslant \; \left(1/\mu - |||\operatorname{div}|||^2/2\right)^{-1} \frac{2|||\operatorname{div}|||^2 R^2}{k+2}, \forall k \geqslant 0 \tag{26}$$

where R is the radius of the sublevel sets of $\tilde{J}$, i.e. $R = \max_{u : \tilde{J}(u) \leqslant \tilde{J}(u^{(0)})} \|u - u^\star\| < +\infty$.

Thus, since $J(f)$ is strongly convex with modulus 1, we obtain using (24) and (26)

$$\begin{aligned}\|f^{(k)} - f^\star\|^2 &\leqslant 2\left(J(f^{(k)}) - J(f^\star)\right) \\ &\leqslant \frac{2\left(\left\|\!\left|\operatorname{div}\right|\!\right\|^{-2} - \mu/2\right)^{-1}(c+1)R^2}{k+2}, \forall k \geqslant 0 \end{aligned} \tag{27}$$

which gives the desired rate. ■

Remark 3.
- *It is important to note that the above proof is general and applies to any problem in the form* (17) *provided that $J(f)$ is strongly convex, beyond the TV projection problem considered here. Strong convexity is of course important to derive the rate* (27) *over the iterates, but the rate on the primal objective in* (27) *remains valid even if $J(f)$ is not strongly convex.*
- *We point out that another proof of the convergence rate of the forward-backward splitting in terms of the objective has recently appeared in [45, Proposition 2]. The rate of [17] is nevertheless sharper.*

This result asserts that Algorithm 1 necessitates as large as $O(1/\eta)$ iterations to reach a η convergence tolerance on the iterates.

C. Multi-step Projection Algorithm

In [17], [46], Nesterov proposes an accelerated multi-step gradient algorithm to solve functionals formed as a sum of two convex terms: a smooth one with Lipschitz-continuous gradient and a non-necessarily smooth term whose structure is simple. By simple, we intend in our context that the corresponding proximity operator is accessible. The dual problem (4) falls within the scope of the Nesterov algorithm. Unlike the forward-backward iteration (16), Nesterov accelerated version uses explicitly all previous iterates $u^{(i)}$, $i < k$ to compute $u^{(k)}$, hence the name multi-step algorithm.

Algorithm 2 details the steps of Nesterov scheme to minimize (4). It is formulated using proximal operators, as described in [7].

1) Convergence analysis: The following results gives the convergence rate of the primal sequence $f^{(k)} = f_0 - \operatorname{div}(u^{(k)})$ obtained from Algorithm 2.

Theorem 2. *Suppose that $\mu \in (0, 1/4)$. Let $u^{(0)} \in \mathbb{R}^{N\times 2}$. Define*

$$\bar{f}^{(k)} = \frac{\sum_{i=0}^{k} a_i f^{(i)}}{A_k} \ .$$

Then, after k iterations, the sequence of iterates $(f^{(k)})_{k\geqslant 1}$ is such that $\exists\, C > 0$

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

Proof: The proof of the convergence rate on the primal objective is patterned after that of [46, Theorem 3]. Indeed, using (20), convexity of H and (19), we get

$$-J(f^\star) \leqslant \tilde{J}(u^{(k)}) \leqslant -J(\bar{f}^{(k)}) + \frac{1}{2A_k}R^2 \ ,$$

Algorithm 2: Nesterov total variation projection.

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

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

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

where $R^2 = \max_{u\in\operatorname{dom} G^*} \|u - u^{(0)}\|^2$, which is bounded by definition. Thus, we arrive at

$$J(\bar{f}^{(k)}) - J(f^\star) \leqslant \frac{1}{2A_k}R^2 \ . \tag{28}$$

By virtue of the lower-bound estimate of $A_k \geqslant k^2/(8\left\|\!\left|\operatorname{div}\right|\!\right\|^2)$, see [17, Lemma 7], and $\left\|\!\left|\operatorname{div}\right|\!\right\|^2 \leqslant 8$, we deduce that

$$J(\bar{f}^{(k)}) - J(f^\star) \leqslant \frac{32R^2}{k^2} \ , \quad k \geqslant 1 \ . \tag{29}$$

Finally, by strong convexity of $J(f)$, we obtain the desired rate

$$\|\bar{f}^{(k)} - f^\star\|^2 \leqslant \frac{64R^2}{k^2} \ , \quad k \geqslant 1 \ . \tag{30}$$

■

Remark 4. *Again, in the same vein as the proof of Theorem 1, the above proof is general and applies straightforwardly to any problem in the form* (17) *beyond TV projection, with the proviso that $J(f)$ is strongly convex to get the rate on the iterates. Strong convexity is however not necessary to get the convergence rate* (29) *of the primal objective $J(f)$. This has been also noted in [47] who followed the lines of proof of [46, Theorem 3] in his convergence analysis on the objective.*

This result asserts that Algorithm 2 necessitates only $O(1/\sqrt{\eta})$ iterations to reach a η-convergence tolerance both on the primal iterates and objective. This is much faster than $O(1/\eta)$ iterations required by the forward-backward scheme.

V. Inverse Problems

Image acquisition devices compute $P \leqslant N$ noisy measurements

$$y = \Phi f_0 + \varepsilon \in \mathbb{R}^p. \tag{31}$$

of a high resolution image $f_0 \in \mathbb{R}^N$, where ε is an additive noise. The linear operator Φ typically accounts for blurring, sub-sampling or missing pixels so that the measured data y only captures a small fraction of the original image f one wishes to recover.

A. Regularization with TV Constraint

A total variation prior allows to to regularize the solution by reducing the space of candidate solutions of the inverse problem to those belonging to a total variation ball of finite radius. Thus solving such a constrained inverse problem provides an image that both matches approximately the forward measurements y and that has a low total variation. If the noise is assumed to be zero-mean of finite variance, it is devised to solve the constrained optimization problem

$$\min_{f\in\mathbb{R}^N} \|f\|_{\text{TV}} \quad \text{subject to} \quad \|\Phi f - y\| \leqslant \gamma \tag{32}$$

where γ is related to the noise standard deviation supposedly known a priori.

On the contrary, if a little is known about the noise ε, but one has some guess τ on the total variation of the image, it is better to consider the following problem

$$f^\star = \operatorname*{argmin}_{f\in\mathbb{R}^N} \frac{1}{2}\|\Phi f - y\|^2 \quad \text{subject to} \quad \|f\|_{\text{TV}} \leqslant \tau, \tag{33}$$

where the minimum is not necessarily unique depending on the kernel of the operator Φ.

B. Projected Gradient Descent

The solution to (33) requires the minimization of the gradient Lipschitz functional $\frac{1}{2}\|\Phi f - y\|^2$ under a convex constraint. It is thus possible to use a projected gradient descent

$$f^{(\ell+1)} = \text{Proj}_{\{\|\cdot\|_{\text{TV}}\leqslant\tau\}}\left(f^{(\ell)} + \nu\Phi^*(y - \Phi f^{(\ell)})\right), \tag{34}$$

where $\text{Proj}_{\{\|\cdot\|_{\text{TV}}\leqslant\tau\}}$ is the projector on the TV ball defined in (3).

The following Theorem ensures the convergence of the iteration.

Theorem 3. *If $\nu \in (0, 2/|||\Phi|||^2)$, where $|||\cdot|||$ is the operator spectral norm, then $f^{(\ell)}$ converges to a minimizer $f^\star$ of (33) with the rate $O(1/\ell)$ on the objective.*

Proof: We use the same arguments as in the first part in the proof of Theorem 1. The convergence rate is a classical result that can be found in [48]. ∎

The projector in (34) cannot be computed exactly but is rather computed via a nested inner iteration using either the forward-backward projection detailed in Algorithm 1 or the multi-step Nesterov projection detailed in Algorithm 2. As these projection algorithms are ran a finite number of inner iterations N_ℓ at each outer iteration ℓ, they yield an estimate of the TV-ball projector up to an error term a_ℓ. Put formally, we have

$$f^{(\ell+1)} = \tilde{f}^{(\ell)} - \text{div}\left(u^{[\ell+1]}\right) = \text{Proj}_{\{\|\cdot\|_{\text{TV}}\leqslant\tau\}}(\tilde{f}^{(\ell)}) + a_\ell\ , \tag{35}$$

where $\tilde{f}^{(\ell)} = f^{(\ell)} + \nu\Phi^*(y - \Phi f^{(\ell)})$, and $u^{[\ell+1]}$ is the dual vector solution provided by applying Algorithm 1 or 2 to $\tilde{f}^{(\ell)}$ with a convergence tolerance $\eta_{\text{proj},\ell}$. At this stage, we advise to use $u^{[\ell]}$ as an initialization in the next call of Algorithm 1 or 2 at the outer iteration $\ell+1$. This initialization is intended to make the constant smaller in the convergence rate of the projection algorithms; see the constants R in the proofs of Theorem 1 and 2. This leads to the TV projection algorithm to solve inverse problems summarized in Algorithm 3.

Algorithm 3: TV-constrained inverse problem resolution algorithm.

Initialization: set $f^{(0)} = 0$, $u^{[0]} = 0$, $\ell = 0$ and $\nu \in (0, 2/|||\Phi|||^2)$.
Main iteration:
while $\|f^{(\ell+1)} - f^{(\ell)}\| > \eta$ **do**

1) *Gradient descent step:*
$$\tilde{f}^{(\ell)} = f^{(\ell)} + \nu\Phi^*(y - \Phi f^{(\ell)}).$$
2) *Projection step:* use $u^{[\ell]}$ as an initialization in Algorithm 1 or Algorithm 2, and apply them to $\tilde{f}^{(\ell)}$ with convergence tolerance $\eta_{\text{proj},\ell}$. Get the new dual solution $u^{[\ell+1]}$ given by these algorithms.
3) *Update:* Set $f^{(\ell+1)} = \tilde{f}^{(\ell)} - \text{div}\left(u^{[\ell+1]}\right)$.
4) $\ell = \ell + 1$.

The errors a_ℓ are inevitable and may prevent the above algorithm from converging. But fortunately, a distinctive property of the projected gradient (which is actually a special instance of the forward-backward scheme), is its robustness to these errors under appropriate sufficient conditions to be made precise in the next proposition; see also [44], [16]. Denote C as the constant appearing either in the rate of Theorem 1 or 2.

Proposition 3. *Suppose that the sequence $(\eta_{proj,\ell})_{\ell\in\mathbb{N}}$ is summable. If Algorithm 1 or 2 are ran $N_\ell \geqslant (2C)^{1/\alpha}\eta_{proj,\ell}^{-1/\alpha}$ iterations, where $\alpha = 1/2$ for Algorithm 1 and $\alpha = 1$ for Algorithm 2, then Algorithm 3 converges.*

Proof: In view of Theorem 1 and 2, as well as the triangle inequality, it is sufficient to take $N_\ell \geqslant (2C)^{1/\alpha}\eta_{\text{proj},\ell}^{-1/\alpha}$ in Algorithm 1 and 2 to reach a $\eta_{\text{proj},\ell}$-tolerance on the successive primal iterates. From the definition of a_ℓ in (35), this implies that $\|a_\ell\| \leqslant CN_\ell^{-\alpha} \leqslant \eta_{\text{proj},\ell}/2$. For the projected gradient outer iteration to converge, the error sequence a_ℓ must obey $\sum_{\ell\in\mathbb{N}}\|a_\ell\| < +\infty$ [44], [16]. It is then sufficient to require that the sequence $(\eta_{\text{proj},\ell})_{\ell\in\mathbb{N}}$ is summable. ∎

This result provides a useful guideline on the way the sequence of tolerances $\eta_{\text{proj},\ell}$ should be chosen. But putting exactly this choice into practice is somewhat delicate as we have to circumvent two main difficulties: (i) the estimation of the constant C, more precisely the abstract constants R and c in (27)-(30), and (ii) the choice of a summable sequence $\eta_{\text{proj},\ell}$ such that the dual projection algorithms converge in a reasonable number of iterations given their predicted convergence rates. We here propose to take $\eta_{\text{proj},\ell} \propto \rho^\ell$ (or equivalently N_ℓ

at least $\propto \rho^{-\ell/\alpha}$ or larger), for $0 < \rho < 1$, a parameter that can be adjusted to the application at hand.

Remark 5. *A this stage, the reader may think of using the multi-step Nesterov algorithm to solve* (33) *instead of the one-step projected gradient descent. However, one must be aware that while the projected gradient is robust to errors in the computation of the projection operator as we discuss above, we do not have any proof of robustness of the multi-step Nesterov scheme to such errors. This is the main reason we did not use it here to solve TV-constrained inverse problems.*

VI. Numerical Examples

A. Denoising

We have first tested our TV projection algorithms for denoising, i.e. $\Phi = \mathrm{Id}$ in (31). In our experiment $y = f_0 + \varepsilon$ is an observed image of $N = 512^2$ pixels contaminated by a zero-mean additive white Gaussian noise (AWGN) ε of standard deviation $0.06\|f_0\|_\infty$ (PSNR=12.2 dB).

Figure 1 shows the projections $f^\star$ of the noisy image computed with our dual projection algorithms described either in Algorithm 1 or Algorithm 2 for a decreasing value of the constraint τ, so that only the strongest edges are present in the projected images.

Figure 2 compares the convergence speed of our one-step and multi-step projection algorithms summarized in Algorithm 1 and Algorithm 2, with the sub-gradient projection method proposed in [12], [13]. Since an iteration of the multi-step Nesterov projection is approximately twice the computational cost of one iteration of the two other algorithms, we displayed the errors generated by $f^{(k/2)}$ instead of $f^{(k)}$ for the curve of the multi-step algorithm. 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. Figure 3 shows the evolution of the total variation of the iterates.

B. Total Variation Texture Synthesis

Statistical approaches to texture synthesis draw an image at random from a set of potential textures defined by constraints that can be learned from an exemplar texture. A standard procedure considers the outputs of a filter-bank, such as wavelets, and constrains their marginal distributions to match those of the exemplar [21], [24]. Instead of considering the uniform distribution inside the set of constraints–assumed to have a non-empty intersection– and sample from this intersection, a simpler approach is to use alternating projections onto the constraints of an initial random white noise image [25]. Although the constrained sets are often non-convex, hopefully, this scheme converges to a point within the constraints that is expected to be visually similar to the exemplar image.

We propose here to consider a total variation constraint to better preserve sharp edges during the synthesis

$$\mathcal{C}_{\mathrm{TV}} = \left\{ f \in \mathbb{R}^N \setminus \|f\|_{\mathrm{TV}} \leqslant \tau \right\} \ ,$$

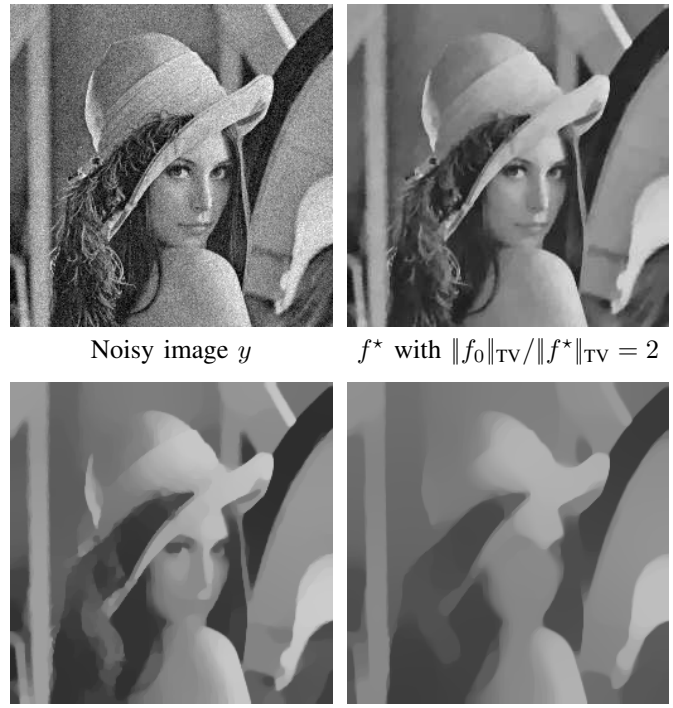

Fig. 1. *Examples of total variation projections computed with our algorithm.*

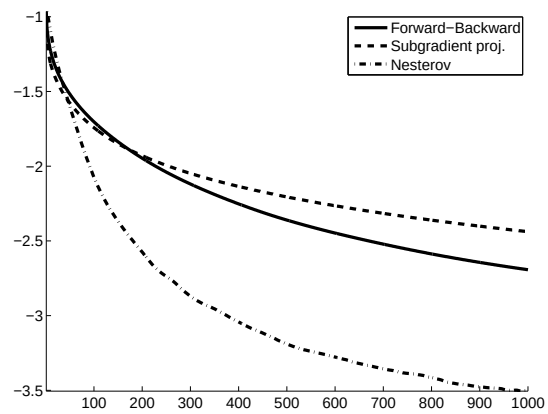

Fig. 2. *Decay with k of the error* $\log_{10}(\|f^{(k)} - f^\star\|/\|f^\star\|)$ *for the one-step forward-backward dual projection in Algorithm 1 (solid line), for the multi-step dual projection in Algorithm 2 (dashed-dotted line), and the sub-gradient projection [12] (dashed line). Here* $\tau = \|f_0\|_{TV}/4$.

where τ might be computed from an exemplar texture $\tau = \|f_0\|_{\mathrm{TV}}$. This constraint is especially suitable for piecewise-smooth textures representing objects occluding each other.

To enforce other statistical constraints as well, we set up a simple texture model where only the histogram of grayscale intensity is imposed. In a discrete setting, this amounts to selecting a set $V = \{v_i\}_{i=0}^{N-1}$ of N values $v_i \in \mathbb{R}$, and consider the convex compact constraint set

$$\mathcal{C}_V = \left\{ f \in \mathbb{R}^N \setminus \{f[i,j]\}_{i,j=0}^{n-1} = V \right\} .$$

In other words, the constraint $f \in \mathcal{C}_V$ imposes that the contrast of the texture remains the same during the synthesis. We note that more complicated statistical constraints can be imposed as well, see for instance [25].

The synthesis algorithm, detailed in Algorithm 4, corresponds to the alternating projection onto the two (convex)

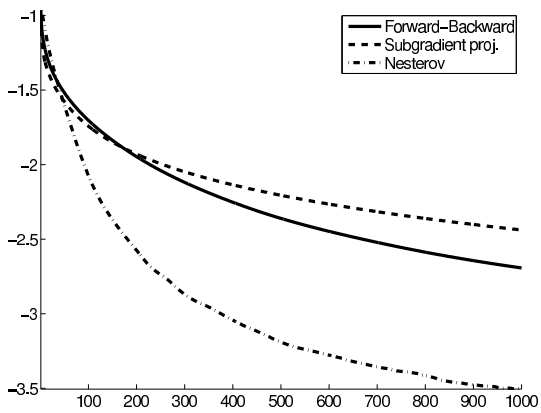

Fig. 3. *Decay with k of the total variation error* $\log_{10}(\|f^{(k)}\|_{TV}/\tau - 1)$ *for the same algorithms as in Figure 2. Here* $\tau = \|f_0\|_{TV}/4$.

constraint sets

$$f^{(\ell+1)} = \mathrm{Proj}_{\mathcal{C}_V}\left(\mathrm{Proj}_{\|\cdot\|_{\mathrm{TV}}\leqslant\tau}(f^{(\ell)})\right) \ ,$$

starting from an initial noise texture $f^{(0)}$.

Algorithm 4: Total variation texture synthesis algorithm.

Initialization: set $f^{(0)}$ = random white noise and set $\ell = 0$.
Main iteration:
while $\|f^{(\ell+1)} - f^{(\ell)}\| > \eta$ **do**

1) *Impose TV constraint:* compute
$$\tilde{f}^{(\ell)} = \mathrm{Proj}_{\|\cdot\|_{\mathrm{TV}}\leqslant\tau}\, f^{(\ell)}$$
using our dual projection Algorithm 1 or 2.
2) *Impose histogram constraint:* compute
$$f^{(\ell+1)} = \mathrm{Proj}_{\mathcal{C}_V}\, \tilde{f}^{(\ell)}$$
using histogram equalization (36).
3) $\ell \leftarrow \ell + 1$.

The orthogonal projection $\mathrm{Proj}_{\mathcal{C}_V}(f)$ of an image onto $\mathcal{C}_V$ corresponds to an histogram equalization. We assume that the values of V are ordered by increasing values $v_i \leqslant v_{i+1}$, and we denote by $f[r(i)]$ the ordered values of f, where r is a permutation of the pixel indices. The orthogonal projection is then

$$\mathrm{Proj}_{\mathcal{C}_V}(f) = \tilde{f} \quad \text{with} \quad \tilde{f}[r(i)] = v_i. \tag{36}$$

Figure 4 depicts examples of texture synthesis where the pixel values V are equally spaced in $[0, 1]$ so that the textures have a uniform distribution of gray values. When the total variation constraint τ decreases, short edges are removed. This allows to interpolate between a noisy texture and a cartoon image with sharp edges.

C. Inpainting

Inpainting aims at restoring an image f_0 from which a set $\Omega \subset \{0, \ldots, n-1\}^2$ of pixels is missing. It corresponds to

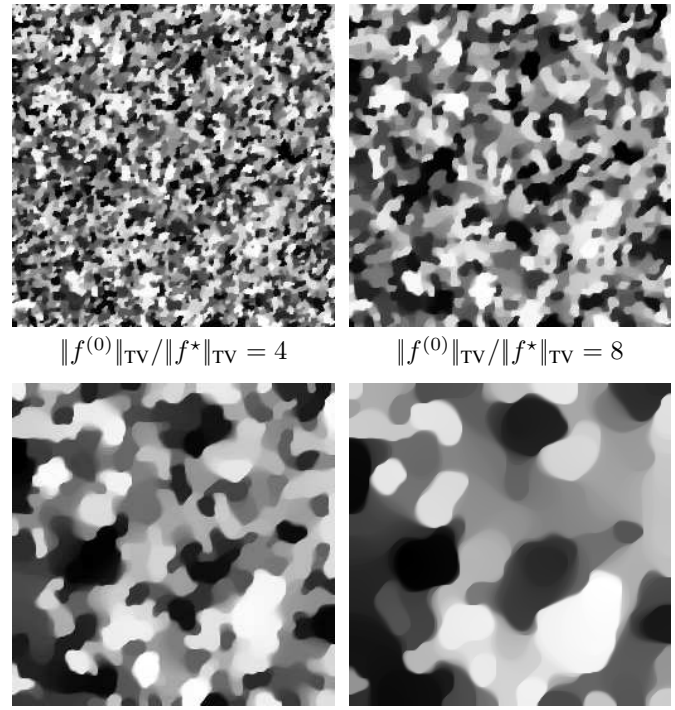

Fig. 4. *Texture synthesis with TV projection and histogram equalization.*

the inversion of the ill-posed linear problem (31) where Φ is defined as

$$(\Phi f)[i,j] = \begin{cases} 0 & \text{if} \quad (i,j) \in \Omega, \\ f[i,j] & \text{if} \quad (i,j) \notin \Omega. \end{cases} \tag{37}$$

In this case, it is obvious that $|||\Phi||| = 1$, and we use $\nu = 1$ for the projected gradient descent Algorithm 3. The recursion (34) amounts to first imposing the known values outside Ω

$$\tilde{f}^{(\ell)}[i,j] = \begin{cases} f^{(\ell)}[i,j] & \text{if} \quad (i,j) \in \Omega, \\ y[i,j] & \text{otherwise.} \end{cases}$$

and then projecting onto the TV ball

$$f^{(\ell+1)} = \mathrm{Proj}_{\{\|\cdot\|_{\mathrm{TV}}\leqslant\tau\}}\, \tilde{f}^{(\ell)}.$$

The top of Figure 5, exemplifies a damaged image y of $N = 512^2$ pixels, with $|\Omega|/N = 0.7$ of randomly removed pixels. The noise is AWGN with standard deviation $0.05\|f_0\|_\infty$ (PSNR=13 dB).

The inpainted image $f^\star$ is computed by solving (33) with the projected gradient descent, Algorithm 3, with a total variation constraint size $\tau = 0.6\|f_0\|_{\mathrm{TV}}$. The number of iterations for the projection step 3) is controlled by setting $\eta_{\mathrm{proj}} = 10^{-2}$. Roughly between 10 to 20 iterations of dual projections are required to maintain the total variation constraints during the gradient descent. Figure 6, top, depicts the decay in log scale of the iterates error, that exhibits a roughly linear convergence speed for large ℓ. This rate is likely to be a consequence of the special structure of the masking operator Φ. We think that this can justified in the light of compressed sensing arguments (the mask is random here), and the convergence analysis in [45]. We leave these aspects, which are beyond the scope of the paper, to a future work.

D. Deconvolution

An optical system produces a blur that is modeled by convolution with a low pass point spread function (PSF) φ. In such a case, the operator Φ in (31) represents the circular convolution with φ. The convolution by φ removes high frequency details and the total variation constraint (33) helps to recover sharp edges of the original image.

Figure 6, bottom, gives an example of blurred image y of $N = 512^2$ pixels. The PSF is a Gaussian kernel of standard deviation $s = 4$ pixels, normalized to a unit mass. Thus $|||\Phi||| = 1$. The noise is AWGN with standard deviation $0.02\|f_0\|_\infty$ (PSNR=17 dB). We use a gradient descent step-size $\nu = 1.9$ in Algorithm 3.

The deconvolved image $f^\star$ is computed by applying Algorithm 3 to y, with a total variation constraint size $\tau = 0.6\|f_0\|_{\mathrm{TV}}$. Figure 6, bottom, shows the decay in log scale of the iterates error. This is consistent with the predicted convergence rate of Theorem 3.

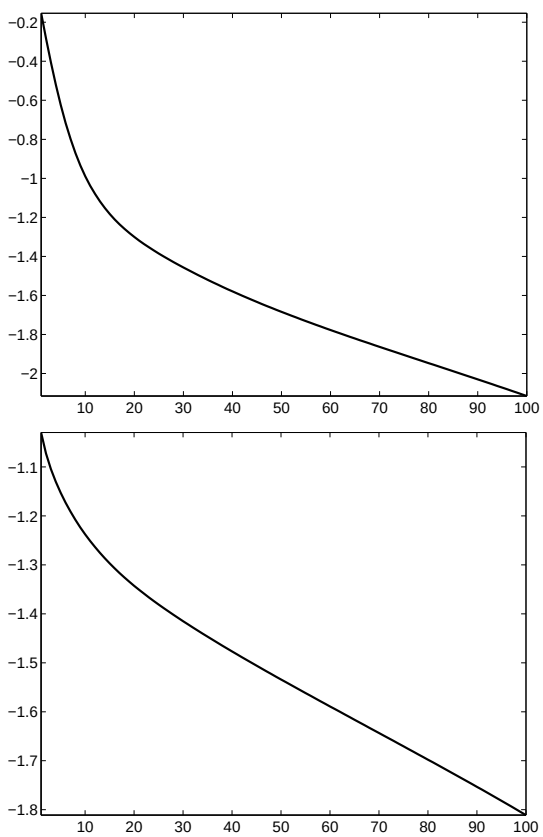

Fig. 6. *Decay of the error* $\log_{10}(\|f^{(\ell)} - f^\star\|/\|f_0\|)$ *for inpainting (top) and deconvolution (bottom).*

Conclusion

This paper proposes a new approach to compute the projection of an image on a total variation ball. This approach solves an unconstrained dual formulation of the primal problem, and boils down to an iterative soft thresholding on the gradient field. Two algorithms for solving the dual minimization problem were proposed. We also studied their convergence behavior and established their convergence rates.

Even though we only focused on the total variation norm in the constraint, our dualization-based projection approach is quite general and extends to any positively 1-homogeneous functional for which the conjugate can be easily computed. This includes constraints involving functionals of the form $\|Gf\|_1$ for any linear operator G (of explicit adjoint) such as the analysis operator of a frame. The scheme also generalizes very easily to arbitrary dimension.

We have illustrated the projection algorithm over several applications such as denoising when little is known about the noise statistics, or to enforce total variation constraint in texture synthesis.

A projected gradient descent was also proposed that uses this projection to solve linear inverse problems under a total variation constraint. Our projector can be advantageously used when additional constraints are involved, such as for example contrast bounds $a \leqslant f \leqslant b$ or affine constraints $Af = g$. Such a problem with compound constraints can be solved for instance using e.g. Douglas-Rachford splitting and its extensions.

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] D. Goldfarb and W. Yin, "Second-order cone programming methods for total variation-based image restoration," *SIAM J. Scientific Computing*, vol. 27, no. 2, pp. 622–645, 2005.

[3] T. F. Chan, G. H. Golub, and P. Mulet, "A nonlinear primal-dual method for total variation-based image restoration," *SIAM Journal on Scientific Computing*, vol. 20, no. 6, pp. 1964–1977, Nov. 1999.

[4] M. Zhu and T. Chan, "An efficient primal-dual hybrid gradient algorithm for total variation image restoration," *UCLA CAM Report 08-34*, 2007.

[5] C. R. Vogel and M. E. Oman, "Iterative methods for total variation denoising," *SIAM Journal on Scientific Computing*, vol. 17, no. 1, pp. 227–238, Jan. 1996.

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

[7] 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.

[8] J.-F. Aujol, "Some first-order algorithms for total variation based image restoration," *Journal of Mathematical Imaging and Vision*, 2009, to appear.

[9] P. Rodríguez and B. Wohlberg, "fficient minimization method for a generalized total variation functional," *IEEE Trans. Image Proc.*, vol. 18, no. 2, pp. 322–332, 2009.

[10] Y. Wang, W. Yin, and Y. Zhang, "A fast algorithm for image deblurring with total variation regularization," *Rice University CAAM Technical Report TR07-10*, 2007.

[11] J. Bect, L. Blanc Féraud, G. Aubert, and A. Chambolle, "A ℓ_1-unified variational framework for image restoration," in *Proc. of ECCV04*. Springer-Verlag, 2004, pp. Vol IV: 1–13.

[12] 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.

[13] P.-L. Combettes, "A block-iterative surrogate constraint splitting method for quadratic signal recovery," *IEEE Trans. Signal Proc.*, vol. 51, no. 7, pp. 1771–1782, 2003.

[14] D. Gabay, "Applications of the method of multipliers to variational inequalities," in *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems*, M. Fortin and R. Glowinski, Eds. North-Holland, Amsterdam, 1983.

[15] P. Tseng, "Applications of a splitting algorithm to decomposition in convex programming and variational inequalities," *SIAM Journal on Control and Optimization*, vol. 29, no. 1, pp. 119–138, 1991.

[16] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Modeling & Simulation*, vol. 4, no. 4, pp. 1168–1200, 2005.

f_0 $y = \Phi f_0 + \varepsilon$ $f^\star$

Fig. 5. *Examples of inpainting (first row) and deconvolution (bottom row) using the TV constraint.*

[17] Y. Nesterov, "Gradient methods for minimizing composite objective function," Universit catholique de Louvain, Center for Operations Research and Econometrics (CORE), CORE Discussion Papers 2007076, Sep. 2007.

[18] G. Buttazzo, G. Carlier, and M. Comte, "On the selection of maximal cheeger sets," *Differential and Integral Equations*, vol. 20, no. 9, pp. 991–1004, 2007.

[19] G. Carlier, M. Comte, and G. Peyré, "Approximation of maximal cheeger sets by projection," *to appear in ESAIM: Mathematical Modelling and Numerical Analysis*, 2008.

[20] K. Perlin, "An image synthesizer," in *Proc. Siggraph '85*. New York, NY, USA: ACM Press, 1985, pp. 287–296.

[21] D. J. Heeger and J. R. Bergen, "Pyramid-Based texture analysis/synthesis," in *Proc. Siggraph '95*, ser. Annual Conference Series, R. Cook, Ed., ACM SIGGRAPH. Addison Wesley, Aug. 1995, pp. 229–238.

[22] A. A. Efros and T. K. Leung, "Texture synthesis by non-parametric sampling," in *ICCV '99: Proceedings of the International Conference on Computer Vision-Volume 2*. IEEE Computer Society, 1999, p. 1033.

[23] L.-Y. Wei and M. Levoy, "Fast texture synthesis using tree-structured vector quantization," in *Proc. Siggraph '00*. ACM Press/Addison-Wesley Publishing Co., 2000, pp. 479–488.

[24] S. C. Zhu, Y. Wu, and D. Mumford, "Filters, random fields and maximum entropy (FRAME): Towards a unified theory for texture modeling," *Int. J. Comput. Vision*, vol. 27, no. 2, pp. 107–126, 1998.

[25] J. Portilla and E. P. Simoncelli, "A parametric texture model based on joint statistics of complex wavelet coefficients," *Int. J. Comput. Vision*, vol. 40, no. 1, pp. 49–70, 2000.

[26] S. Masnou, "Disocclusion: a variational approach using level lines," *IEEE Trans. Image Processing*, vol. 11, no. 2, pp. 68–76, Feb. 2002.

[27] C. Ballester, M. Bertalmio, V. Caselles, G. Sapiro, and J. Verdera, "Filling-in by joint interpolation of vector fields and gray levels," *IEEE Trans. Image Processing*, vol. 10, no. 8, pp. 1200–1211, Aug. 2001.

[28] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proc. Siggraph 2000*, 2000, pp. 417–424.

[29] D. Tschumperlé and R. Deriche, "Vector-valued image regularization with PDEs: A common framework for different applications," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 27, no. 4, pp. 506–517, 2005.

[30] T. Chan and J. Shen, "Mathematical models for local non-texture inpainting," *SIAM J. Appl. Math.*, vol. 62, no. 3, pp. 1019–1043, 2001.

[31] M. Elad, J.-L. Starck, D. Donoho, and P. Querre, "Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)," *Journal on Applied and Computational Harmonic Analysis*, vol. 19, pp. 340–358, 2005.

[32] M. J. Fadili, J.-L. Starck, and F. Murtagh, "Inpainting and zooming using sparse representations," *The Computer Journal*, vol. 52, no. 1, pp. 64–79, 2007.

[33] C. R. Vogel and M. E. Oman, "Fast, robust total variation-based reconstruction of noisy, blurred images," *IEEE Trans. Image Proc.*, vol. 7, no. 6, pp. 813–824, Jun. 1998.

[34] T. Chan, S. Esedoglu, F. Park, and A. Yip, *Recent developments in total variation image restoration*. New York: Springer, 2005.

[35] M. Figueiredo and R. Nowak, "An EM Algorithm for Wavelet-Based Image Restoration," *IEEE Trans. Image Proc.*, vol. 12, no. 8, pp. 906–916, 2003.

[36] I. Daubechies, M. Defrise, and C. D. Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. on Pure and Appl. Math.*, vol. 57, pp. 1413–1541, 2004.

[37] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs, "A variational formulation for frame-based inverse problems," *Inv. Prob.*, vol. 23, pp. 1495–1518, 2007.

[38] M. J. Fadili and J.-L. Starck, "Sparse representation-based image deconvolution by iterative thresholding," in *ADA IV*. France: Elsevier, 2006.

[39] J.-J. Moreau, "Fonctions convexes duales et points proximaux dans un espace hilbertien," *CRAS Sér. A Math.*, vol. 255, pp. 2897–2899, 1962.

[40] C. Lemaréchal and J.-B. Hiriart-Urruty, *Convex Analysis and Minimization Algorithms I and II*, 2nd ed. Springer, 1996.

[41] I. Ekeland and R. Temam, *Analyse convexe et problèmes variationnels*. Dunod, 1974.

[42] A. Chambolle, R. D. Vore, N. Lee, and B. Lucier, "Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage," *IEEE Trans. Image Proc.*, vol. 7, no. 3, pp. 319–335, March 1998.

[43] J.-J. Moreau, "Proximité et dualité dans un espace hilbertien," *Bulletin de la S.M.F*, vol. 93, pp. 273–299, 1965.

[44] P. L. Combettes, "Solving monotone inclusions via compositions of nonexpansive averaged operators," *Optimization*, vol. 53, no. 5-6, pp. 475–504, 2004.

[45] K. Bredies and D. Lorenz, "Linear convergence of iterative soft-thresholding," *J. Fourier Analysis and Applications*, vol. 14, pp. 813–837, Dec 2008.

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

[47] P. Weiss, "Algorithmes rapides d'optimisation convexe: application à la restauration d'images et à la détection de changements," Ph.D. dissertation, Université Sophia-Antipolis, Nice, France, 2008.

[48] B. Polyak, *Introduction to Optimization*. New York: Optimization Software, Inc., Publications Division, 1987.