Proximity Operator of a Sum of Functions; Application to Depth Map Estimation

Nelly Pustelnik, *Member, IEEE* and Laurent Condat

***Abstract*—Proximal splitting algorithms for convex optimization are largely used in signal and image processing. They make possible to call the individual proximity operators of an arbitrary number of functions, whose sum is to be minimized. But the larger this number, the slower the convergence. In this work, we show how to compute the proximity operator of a sum of two functions, for a certain type of functions operating on objects having a graph structure. The gain provided by avoiding unnecessary splitting is illustrated by an application to depth map estimation.**

***Index Terms*—Proximal algorithms, proximity operator, splitting, convex optimization, support function, disparity map estimation.**

I. Introduction

Proximal algorithms – Optimization is a key step in solving data processing problems formulated as the minimization of an energy, like inverse problems or learning problems. Efficient convex optimization procedures were restricted to smooth cost functions for several decades, until the development of proximal methods, so-called because they make use of the proximity operator of the functions. This class of algorithms, which are well suited for a broad range of nonsmooth large-scale problems [1]–[3], includes the forward–backward algorithm (and its derivatives ISTA, FISTA…), the Douglas–Rachford algorithm (related to ADMM, PPXA and PPXA+ [4]), and several recent primal–dual algorithms [5]–[9].

Splitting: pros and cons – Let us consider a general template problem:

$$\underset{x\in\mathcal{H}}{\text{minimize}} \sum_{r=1}^{R} f_r(x) + \sum_{s=1}^{S} g_s(x), \tag{1}$$

where $\mathcal{H}$ is a real Hilbert space, the functions $(f_r)_{1\leq r\leq R}$ belong to $\Gamma_0(\mathcal{H})$, the set of proper, convex, lower semicontinuous functions from $\mathcal{H}$ to $]-\infty,+\infty]$ [1], and the functions $(g_s)_{1\leq s\leq S}$ in $\Gamma_0(\mathcal{H})$ are differentiable, with Lipschitz-continuous gradient. The computation of the gradient of the sum of functions g_s, which is the sum of their gradients, does not raise any difficulty. By contrast, the proximity operator[1] of the sum of functions f_r is generally intractable, so that the proximity operators of the individual functions f_r are called, instead. The flexibility of the splitting algorithms allows to do so, but at the price of a slower convergence and increased memory usage. It is then preferable to compute the proximity operator of a sum of functions, whenever possible.

Proximity operator of a sum of two functions – Several works have been dedicated to the computation of the proximity operator of a sum of two functions. For instance, we can refer to [11]–[13] for iterative solutions and the convergence guarantees when inner iterations are required. In this work, we focus on sufficient conditions for the following equality to hold:

$$\operatorname{prox}_{g+h} = \operatorname{prox}_h \circ \operatorname{prox}_g, \tag{2}$$

where g and h are functions in $\Gamma_0(\mathcal{H})$ and $\circ$ denotes the mapping composition. This desirable property is not satisfied, in general. Given a closed convex set C, we define its indicator function ι_C as the convex function which maps x to $\{0$ if $x \in C$, $+\infty$ else$\}$. Then if C is reduced to a singleton or is a non-empty subset of $\mathbb{R}$, the property (2) holds, with $h = \iota_C$ and any $g \in \Gamma_0(\mathbb{R})$. By extension, it also holds if C and g are separable in the same basis of $\mathcal{H}$ [11]. A more general sufficient condition is $\partial g(x) \subset \partial g(\operatorname{prox}_h(x))$, for every $x \in \mathcal{H}$ [14], [15]. When $\mathcal{H} = \mathbb{R}$, g is the support function σ_C of a closed, convex, non-empty subset C of $\mathcal{H}$ [1], and h is differentiable at 0 with $h'(0) = 0$, the equality (2) holds as well [16, Proposition 3.6].

Contribution – A new result is derived in Section II and particular cases of this result are related to the signal processing literature in Section III. The interest of our result compared to splitting is illustrated and discussed by an application to depth estimation in Section IV.

II. New Result

We place ourselves in a real Hilbert space $\mathcal{H} = \mathbb{R}^\Omega$ of objects $\boldsymbol{x} = (x_n)_{n\in\Omega}$, with real elements x_n and domain Ω. For instance, 1-D signals of size N correspond to $\Omega = \{1,\dots,N\}$; 2-D images or matrices of size $N_1 \times N_2$ correspond to $\Omega = \{1,\dots,N_1\} \times \{1,\dots,N_2\}$, with $n = (n_1, n_2)$ a 2-D index. We view $\boldsymbol{x} \in \mathcal{H}$ as a graph with values x_n at its edges. Let Υ be a subset of Ω^2, consisting of pairs of indexes, which can be viewed as the edges of the graph. Let g and h be two functions of $\Gamma_0(\mathcal{H})$, which penalize the edges and the vertices of the graph, respectively. More precisely, we suppose that the following holds.

This work is supported by DEFI Imag'In 2017 - Projet SIROCCO.

N. Pustelnik (Corresponding Author) is with Univ. Lyon, ENS de Lyon, Univ. Claude Bernard, CNRS, Laboratoire de Physique, F-69342 Lyon, France. E-mail: nelly.pustelnik@ens-lyon.fr

L. Condat is with CNRS, GIPSA-Lab, Univ. Grenoble Alpes, F-38000 Grenoble, France.

[1]The proximity operator prox_f of a function f maps an object $\boldsymbol{y}$ to the unique minimizer over $\boldsymbol{x}$ of $\|\boldsymbol{x}-\boldsymbol{y}\|^2/2 + f(\boldsymbol{x})$ [1], [10].

Assumption II.1 *(i) h is separable, with*

$$\big(\forall \boldsymbol{x} = (x_n)_{n\in\Omega}\big) \qquad h(\boldsymbol{x}) = \sum_{n\in\Omega} h_0(x_n), \tag{3}$$

for some function $h_0 \in \Gamma_0(\mathbb{R})$.

(ii) g has the following form:

$$\big(\forall \boldsymbol{x} = (x_n)_{n\in\Omega}\big) \quad g(\boldsymbol{x}) = \sum_{(m,m')\in\Upsilon} \sigma_{C_{m,m'}}(x_{m'} - x_m), \tag{4}$$

where $\sigma_{C_{m,m'}} : t \in \mathbb{R} \mapsto \sup\{tp, p \in C_{m,m'}\}$ is the support function of a closed real interval $C_{m,m'}$, such that $\inf C_{m,m'} = a_{m,m'}$ and $\sup C_{m,m'} = b_{m,m'}$, for some $a_{m,m'} \in \mathbb{R}\cup\{-\infty\}$ and $b_{m,m'} \in \mathbb{R}\cup\{+\infty\}$, with $a_{m,m'} \leq b_{m,m'}$. Thus, we have

$$(\forall t \in \mathbb{R}) \quad \sigma_{C_{m,m'}}(t) = \begin{cases} a_{m,m'}t & \text{if } t < 0, \\ 0 & \text{if } t = 0, \\ b_{m,m'}t & \text{if } t > 0, \end{cases} \tag{5}$$

with the convention $(-\infty)t = +\infty$ and $(+\infty)t = +\infty$.

It is straightforward to show that g and h belong to $\Gamma_0(\mathcal{H})$.

Proposition II.2 *Under Assumption II.1, it holds that*

$$\operatorname{prox}_{g+h} = \operatorname{prox}_h \circ \operatorname{prox}_g . \tag{6}$$

The proof is derived in Appendix V. This proposition extends several results of the literature, which are summarized in the next section.

III. Interest in Signal/Image Processing

The kind of functions considered in the previous Section are often encountered in signal/image processing applications. Here are some examples having practical interest.

Example III.1 *In the 1-D case, with $\Omega = \{1,\dots,N\}$ and $\Upsilon = \{(1,2),(2,3),\dots,(N-1,N)\}$, the weighted total variation of $\boldsymbol{x}$ corresponds to our setting with $b_{n,n+1} = -a_{n,n+1} = \omega_n \geq 0$; that is,*

$$\big(\forall \boldsymbol{x} \in \mathbb{R}^\Omega\big) \quad g(\boldsymbol{x}) = \sum_{n=1}^{N-1} \omega_n |x_{n+1} - x_n|. \tag{7}$$

There exist very efficient algorithms to compute the proximity operator of this specific function [17], [18]. Its extension to 2-D images is known as the anisotropic total variation [19] and its proximity operator can be computed efficiently using graph cuts [20].

Example III.2 *The fused lasso [21] corresponds to the case where g is defined as in Example III.1 and h is an ℓ_1-norm; that is, in* (3)*, $h_0 = \lambda|\cdot|$, for some $\lambda > 0$.*

The proximity operator of the fused lasso is obtained by applying the proximity operator of the total variation, followed by soft-thresholding, which is the proximity operator of the ℓ_1-norm. This particular case of Proposition II.2 was already known [22].

Example III.3 *In the 1-D case, with $\Omega = \{1,\dots,N\}$ and $\Upsilon = \{(1,2),(2,3),\dots,(N-1,N)\}$, isotonic regression corresponds to*

$$g(\boldsymbol{x}) = \iota_E(\boldsymbol{x}), \quad \text{with } E = \{\boldsymbol{x} \in \mathbb{R}^N \,:\, x_1 \geq \ldots \geq x_N\}. \tag{8}$$

This is a particular case of our setting, with $a_{n,n+1} = 0$ and $b_{n,n+1} = +\infty$.

The Pool Adjacent Violators Algorithm (PAVA) [23] allows us to compute the proximity operator of g with complexity $O(N)$.

When, in addition, $h_0 = \iota_{[0,1]}$, the property of Proposition II.2 was already known [24]. It was also shown for signals defined on a tree, instead of a 1-D chain [25].

IV. Application to Depth Estimation in Stereovision

We now focus on a practical application of Proposition II.2. We consider the problem of depth map estimation for noisy stereovision images. An efficient strategy for this purpose has been designed in [26] and involves the same functions as in Example III.3. In this section, we first detail the convex optimization problem to solve and we provide several strategies with more or less splitting, in order to evaluate the impact of our theoretical result in Proposition II.2.

A. Minimal-Partition Type Formulation

Given a pair of images with domain $\Omega = \{1,\dots,N_1\}\times\{1,\dots,N_2\}$, we want to estimate the depth of the scene at every pixel $n \in \Omega$. A variational formulation of depth reconstruction consists in estimating a partition $(\Omega^{(q)})_{1\leq q\leq Q}$ of Ω, such that the depth is constant in each region $\Omega^{(q)}$ [26], [27]. Spatial homogeneity is obtained by penalizing the perimeter of the regions. Given a set of images $\boldsymbol{\eta}^{(q)} \in \mathbb{R}^\Omega$, $q = 1,\dots,Q$, so that $\eta_n^{(q)}$ is the cost of assigning the pixel n to the region $\Omega^{(q)}$, the minimization problem reads

$$\begin{aligned} \underset{\Omega^{(1)},\dots,\Omega^{(Q)}}{\text{minimize}} \; & \sum_{q=1}^{Q} \sum_{n\in\Omega^{(q)}} \eta_n^{(q)} + \lambda \sum_{q=1}^{Q} \operatorname{Per}(\Omega^{(q)}) \\ \text{s.t.} \; & \begin{cases} \bigcup_{q=1}^{Q} \Omega^{(q)} = \Omega, \\ (\forall q \neq p),\ \Omega^{(q)} \cap \Omega^{(p)} = \emptyset, \end{cases} \end{aligned} \tag{9}$$

where $\operatorname{Per}(\cdot)$ denotes the perimeter and $\lambda > 0$ controls the amount of spatial regularization. A convex relaxation of this nonconvex problem is [26]:

$$\begin{aligned} \underset{\Theta=(\boldsymbol{\theta}^{(q)})_{1\leq q\leq Q-1}}{\text{minimize}} \; & \sum_{q=1}^{Q} \sum_{n\in\Omega} \big(\theta_n^{(q-1)} - \theta_n^{(q)}\big)\eta_n^{(q)} \\ & + \lambda \sum_{q=1}^{Q} \operatorname{TV}(\boldsymbol{\theta}^{(q-1)} - \boldsymbol{\theta}^{(q)}) \\ \text{s.t.} \; & \boldsymbol{\theta}^{(0)} \geq \boldsymbol{\theta}^{(1)} \geq \ldots \geq \boldsymbol{\theta}^{(Q-1)} \geq \boldsymbol{\theta}^{(Q)}, \end{aligned} \tag{10}$$

where $\boldsymbol{\theta}^{(q)} \in \mathbb{R}^\Omega$, $\boldsymbol{\theta}^{(0)} \equiv 1$, $\boldsymbol{\theta}^{(Q)} \equiv 0$, the inequalities over the $\boldsymbol{\theta}^{(q)}$ are meant pixelwise, and TV denotes some form of the 2-D discrete total variation [19]. In this paper, we use

the classical isotropic total variation [19]; that is, for every $u \in \mathbb{R}^\Omega$,

$$\begin{aligned}\mathrm{TV}(\boldsymbol{u}) &= \sum_{n_1=1}^{N_1}\sum_{n_2=1}^{N_2}\sqrt{\begin{array}{c}\frac{1}{8}(u_{n_1+1,n_2}-u_{n_1,n_2})^2\\ +\frac{1}{8}(u_{n_1,n_2+1}-u_{n_1,n_2})^2\end{array}}\\ &= \|D\boldsymbol{u}\|_{2,1},\end{aligned} \tag{11}$$

where $\|\cdot\|_{2,1}$ is the $\ell_{1,2}$ norm and D is the linear operator taking horizontal and vertical finite differences, scaled so that $\|D\| \leq 1$. Hence, using indicator functions, the minimization problem (10) can be rewritten as

$$\widehat{\Theta} \in \operatorname*{Argmin}_{\Theta} \mathcal{F}(\Theta) + \iota_{[0,1]^{Q\times\Omega}}(\Theta) + \iota_E(\Theta), \tag{12}$$

where

$$\mathcal{F}(\Theta) = \sum_{q=1}^{Q-1}\langle \boldsymbol{\alpha}^{(q)}, \boldsymbol{\theta}^{(q)}\rangle + \lambda\sum_{q=1}^{Q}\|DH_q\Theta\|_{2,1}, \tag{13}$$

$\boldsymbol{\alpha}^{(q)} = \boldsymbol{\eta}^{(q+1)} - \boldsymbol{\eta}^{(q)}$, $H_q\colon \mathbb{R}^{Q\times\Omega} \to \mathbb{R}^\Omega\colon \Theta \mapsto \boldsymbol{\theta}^{(q-1)} - \boldsymbol{\theta}^{(q)}$, and $E = \{\Theta \in \mathbb{R}^{Q\times\Omega} \;:\; \boldsymbol{\theta}^{(1)} \geq \ldots \geq \boldsymbol{\theta}^{(Q-1)}\}$.

This generic formulation can be applied to many partitioning problems [26]. In the case of depth estimation in stereovision [28], [29], we are given two multicomponent images $\boldsymbol{y}^L \in \mathbb{R}^{N_1\times N_2\times C}$ and $\boldsymbol{y}^R \in \mathbb{R}^{N_1\times N_2\times C}$, corresponding to the views of the same scene by the left and right eyes. $C = 3$ in the case of color images. The principle used for depth estimation is that the pixel values $y^L_{n_1,n_2}$ and $y^R_{n_1,n_2'}$ tend to be similar, with the depth of the imaged object inversely proportional to $n_2 - n_2'$. Therefore, we fix the set of possible values of $n_2 - n_2'$ as the integer range $q-1 = 0,\ldots,Q-1$ and we consider the data fidelity term

$$\big(\forall n = (n_1,n_2) \in \Omega\big) \quad \eta_n^{(q)} = \sum_{c=1}^{C}|y^L_{n_1,n_2,c} - y^R_{n_1,n_2-q+1,c}|^p, \tag{14}$$

where an usual choice for p is 1 or 2. Finally, the estimated disparity map is obtained from $\widehat{\Theta}$ as

$$\widehat{\mathcal{M}} = \sum_{q=1}^{Q}(q-1)\big(\widehat{\boldsymbol{\theta}}^{(q-1)} - \widehat{\boldsymbol{\theta}}^{(q)}\big). \tag{15}$$

B. Algorithmic Strategies

In order to estimate $\widehat{\Theta}$, we need a generic algorithmic scheme able to deal with non-smooth functions (hard constraints and $\ell_{1,2}$-norm), with a number of functions to minimize greater than two, and with the presence of linear operators (due to the presence of DH_q). Thus, we adopt the Chambolle–Pock primal–dual algorithm [5], [7].

We recall below the form of this algorithm in a general setting, which consists in finding a minimizer $\widehat{\Theta}$ of $\sum_{r=1}^{R-1} f_r(L_r\Theta) + f_R(\Theta)$ where the $L_r\colon \mathcal{H} \to \mathcal{H}_r$ (resp. $L_r^*\colon \mathcal{H}_r \to \mathcal{H}$) are linear operators (resp. their adjoints) and the functions belong to $\Gamma_0(\mathcal{H})$ or $\Gamma_0(\mathcal{H}_r)$. We assume an upper bound K of the operator norm $\|\sum_{r=1}^{R-1} L_r^*L_r\|$ is known.

In the next Section, we evaluate the impact of three strategies with different levels of splitting:

Algorithm 1 Chambolle–Pock Proximal Splitting Algorithm

Require: $Q \geq 2$, $(\boldsymbol{\alpha}^{(q)})_{1\leq q\leq Q-1}$, $\lambda > 0$.
Choose the parameter $\tau > 0$. Set $\sigma = 1/(K\tau)$.
Initialize $\Theta^{[0]} \in \mathcal{H}$.
Initialize $\Lambda_r^{[0]} \in \mathcal{H}_i$ for every $r \in \{1,\ldots,R-1\}$.
for $i = 0, 1, \ldots$ **do**
  $\Theta^{[i+1]} = \mathrm{prox}_{\tau f_R}(\Theta^{[i]} - \tau\sum_{r=1}^{R-1} L_r^*\Lambda_r^{[i]})$
  for $r = 1,\ldots,R-1$
    $\Lambda_r^{[i+1]} = \mathrm{prox}_{\sigma f_r^*}(\Lambda_r^{[i]} + \sigma L_r(2\Theta^{[i+1]} - \Theta^{[i]}))$
  end for
end for

Ensure: The sequence $(\Theta^{[i]})_{i\in\mathbb{N}}$ converges to a solution $\widehat{\Theta}$.

- Full splitting – The first strategy provides a splitting in $R = 4$ functions:

$$\begin{cases} f_1(\Xi) = \lambda\sum_{q=1}^{Q}\|\Xi\|_{2,1} \quad \text{and} \quad L_1 = DH_q,\\ f_2(\Theta) = \iota_{E_1}(\Theta) \quad \text{and} \quad L_2 = \mathrm{Id},\\ f_3(\Theta) = \iota_{E_2}(\Theta) \quad \text{and} \quad L_3 = \mathrm{Id},\\ f_4(\Theta) = \sum_{q=1}^{Q-1}\langle\boldsymbol{\alpha}^{(q)},\boldsymbol{\theta}^{(q)}\rangle + \iota_{[0,1]^{Q\times\Omega}}(\Theta),\end{cases} \tag{16}$$

  where E_1 and E_2 are defined so as to avoid overlapping variables: $E_1 = \{\Theta \in \mathbb{R}^{Q\times\Omega} \;:\; \boldsymbol{\theta}^{(1)} \geq \boldsymbol{\theta}^{(2)},\ \boldsymbol{\theta}^{(3)} \geq \boldsymbol{\theta}^{(4)},\ldots\}$ and $E_2 = \{\Theta \in \mathbb{R}^{Q\times\Omega} \;:\; \boldsymbol{\theta}^{(2)} \geq \boldsymbol{\theta}^{(3)},\ \boldsymbol{\theta}^{(4)} \geq \boldsymbol{\theta}^{(5)},\ldots\}$. We remark that $E = E_1 \cap E_2$ and that $\iota_E = \iota_{E_1} + \iota_{E_2}$. The proximity operators of ι_{E_1}, ι_{E_2}, f_1 and f_4 have simple closed-form expressions. $K = 4+1+1 = 6$.

- Intermediate splitting – The second strategy consists in splitting as follows:

$$\begin{cases} f_1(\Xi) = \lambda\sum_{q=1}^{Q}\|\Xi\|_{2,1} \quad \text{and} \quad L_1 = DH_q,\\ f_2(\Theta) = \iota_E(\Theta) \quad \text{and} \quad L_2 = \mathrm{Id},\\ f_3(\Theta) = \sum_{q=1}^{Q-1}\langle\boldsymbol{\alpha}^{(q)},\boldsymbol{\theta}^{(q)}\rangle + \iota_{[0,1]^{Q\times\Omega}}(\Theta).\end{cases} \tag{17}$$

  The proximity operator of f_2, which is the projection onto E, can be computed by the exact and fast PAVA algorithm, applied pixelwise, with complexity $O(N_1N_2Q)$ [23]. $K = 4+1 = 5$.

- Proposed minimal splitting: The last strategy is:

$$\begin{cases} f_1(\Xi) = \lambda\sum_{q=1}^{Q}\|\Xi\|_{2,1} \quad \text{and} \quad L_1 = DH_q,\\ f_2(\Theta) = \sum_{q=1}^{Q-1}\langle\boldsymbol{\alpha}^{(q)},\boldsymbol{\theta}^{(q)}\rangle + \iota_{[0,1]^{Q\times\Omega}\cap E}(\Theta).\end{cases} \tag{18}$$

  This solution is made possible by Proposition II.2. Indeed, the proximity operator of $\iota_{[0,1]^{Q\times\Omega}\cap E} = \iota_E + \iota_{[0,1]^{Q\times\Omega}}$ is the projection onto E followed by the projection onto $[0,1]^{Q\times\Omega}$. $K = 4$.

C. Experiments

We apply the depth estimation procedure to two image pairs 'Teddy' and 'Corridor' of the classical 2003 Middlebury database [30], whose sizes are $N_1\times N_2\times C = 375\times 450\times 3$ and $N_1\times N_2\times C = 128\times 128\times 1$. The original images were corrupted by additive white Gaussian noise of standard deviation 10. These left and right images $\boldsymbol{y}^L$ and $\boldsymbol{y}^R$ are shown in Figure 1. We set $Q = 50$ (resp. $Q = 20$), $\lambda = 150$

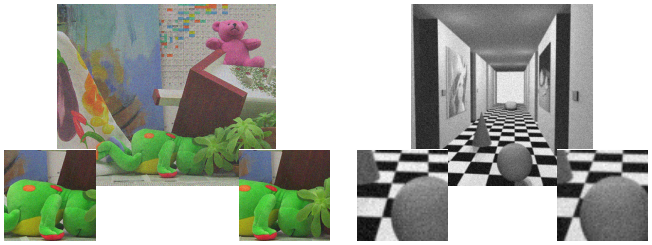

Fig. 1. First column: noisy left 'Teddy' image used for the depth estimation experiment $\boldsymbol{y}^L$ and cropped versions of the noisy left $\boldsymbol{y}^L$ and noisy right $\boldsymbol{y}^R$ images. Second column: 'Corridor' dataset.

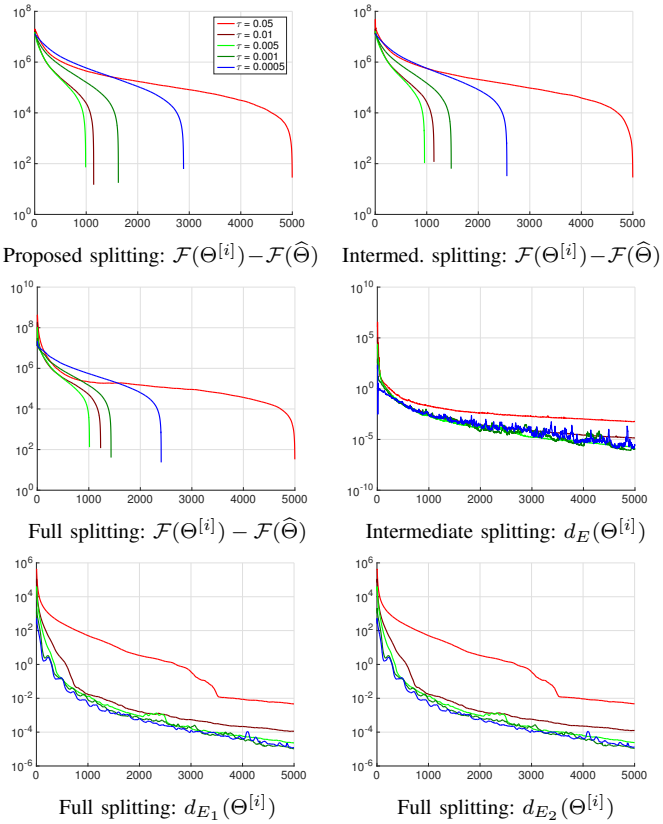

Fig. 2. Evolution of different criteria with respect to the iteration number, for the dataset 'Teddy' and for different values of the parameter τ. $d_\cdot$ denotes the distance to a convex set. d_E is zero for the proposed method.

(resp. $\lambda = 50$), as these values lead to the best results. Note that the value of Q can be deduced from the translation of the foreground items on the noisy data. The methods were implemented in MATLAB code, which was run on a Apple Mac Pro desktop computer with a 2x2.4GHz Quad-Core Intel Xeon CPU and 16GB RAM. We show in Figure 2 the influence of the parameter τ on the objective value and on the distance to the convex set E, or E_1 and E_2, along the iterations on 'Teddy' dataset. For this, the solution $\widehat{\Theta}$ was computed with a very large number of iterations of the proposed method with $\tau = 0.005$. We can observe that with the full splitting and intermediate splitting, the constraints take time to be satisfied, whereas with the proposed minimal splitting, the constraint is satisfied at every iteration. For all three methods, the value $\tau = 0.005$ yields the fastest convergence. Similar results have been observed for 'Corridor' dataset.

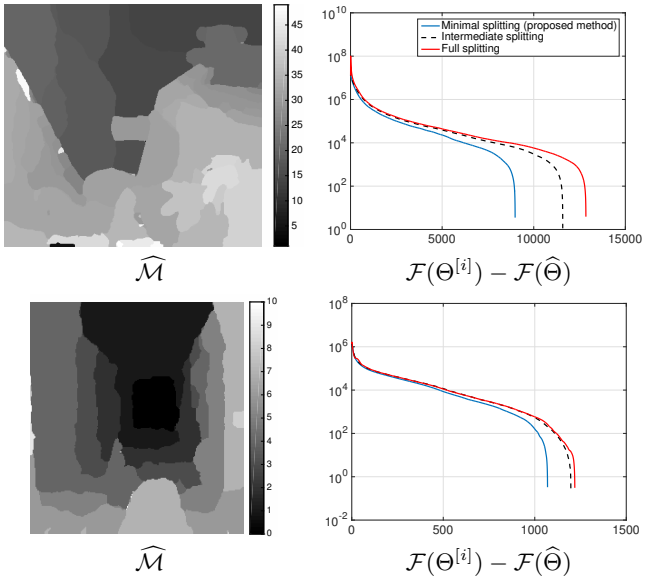

Fig. 3. First row: results for "Teddy" dataset. Second row: results for "Corridor dataset". Left: estimated disparity map. Right: convergence behavior with respect to computation time for $\tau = 0.005$.

In Figure 3, we show the estimated map and the evolution of the objective function with respect to computation time. We observe that, for both datasets, the proposed minimal splitting is the fastest. This demonstrates the interest of avoiding unnecessary splitting whenever possible.

V. Appendix – Proof of Proposition II.2

Let $\boldsymbol{y} \in \mathbb{R}^\Omega$. We set $\tilde{\boldsymbol{z}} = \mathrm{prox}_g(\boldsymbol{y})$, $\tilde{\boldsymbol{x}} = \mathrm{prox}_h(\tilde{\boldsymbol{z}})$. According to the subdifferential characterization of the proximity operator, there exists $u_{m,m'} \in \partial\sigma_{C_{m,m'}}(\tilde{z}_{m'} - \tilde{z}_m)$, for every $(m,m') \in \Upsilon$, such that, for every $n \in \Omega$,

$$\begin{cases} 0 \in \tilde{z}_n - y_n - \sum\limits_{\substack{m' : \\ (n,m') \in \Upsilon}} u_{n,m'} + \sum\limits_{\substack{m : \\ (m,n) \in \Upsilon}} u_{m,n}\,, \\ 0 \in \tilde{x}_n - \tilde{z}_n + \partial h_0(\tilde{x}_n)\,. \end{cases} \tag{19}$$

For $t < 0$, $\partial\sigma_{C_{m,m'}}(t)$ is $\{a_{m,m'}\}$ if $a_{m,m'}$ is real, $\emptyset$ else; for $t > 0$, $\partial\sigma_{C_{m,m'}}(t)$ is $\{b_{m,m'}\}$ if $b_{m,m'}$ is real, $\emptyset$ else. Moreover, $\partial\sigma_{C_{m,m'}}(0) = C_{m,m'}$. Then we can remark that the proximity operator of a function of $\Gamma_0(\mathbb{R})$ is a nondecreasing function; therefore, for every $(m,m') \in \Upsilon$,

$$\begin{cases} \text{if} \quad \tilde{z}_{m'} - \tilde{z}_m > 0 & \text{then} \quad \tilde{x}_{m'} - \tilde{x}_m \geq 0, \\ \text{if} \quad \tilde{z}_{m'} - \tilde{z}_m < 0 & \text{then} \quad \tilde{x}_{m'} - \tilde{x}_m \leq 0, \\ \text{if} \quad \tilde{z}_{m'} - \tilde{z}_m = 0 & \text{then} \quad \tilde{x}_{m'} - \tilde{x}_m = 0. \end{cases} \tag{20}$$

Hence, $\partial\sigma_{C_{m,m'}}(\tilde{z}_{m'} - \tilde{z}_m) \subset \partial\sigma_{C_{m,m'}}(\tilde{x}_{m'} - \tilde{x}_m)$. Together with (19), this implies that there exists $u_{m,m'} \in \partial\sigma_{C_{m,m'}}(\tilde{x}_{m'} - \tilde{x}_m)$, for every $(m,m') \in \Upsilon$, such that, for every $n \in \Omega$,

$$0 = \tilde{x}_n - y_n - \sum_{\substack{m' : \\ (n,m') \in \Upsilon}} u_{n,m'} + \sum_{\substack{m : \\ (m,n) \in \Upsilon}} u_{m,n} + \partial h_0(\tilde{x}_n). \tag{21}$$

This is exactly the subdifferential characterization of prox_{g+h} applied to $\boldsymbol{y}$, so that $\tilde{\boldsymbol{x}} = \mathrm{prox}_{g+h}(\boldsymbol{y})$.

REFERENCES

[1] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, New York, 2011.

[2] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds., pp. 185–212. Springer-Verlag, New York, 2010.

[3] M. Burger, A. Sawatzky, and G. Steidl, "First order algorithms in variational image processing," in *Splitting Methods in Communication, Imaging, Science, and Engineering*, R. Glowinski, S. J. Osher, and W. Yin, Eds., pp. 345–407. Springer, 2016.

[4] J.-C. Pesquet and N. Pustelnik, "A parallel inertial proximal optimization method," *Pacific journal of Optimization*, vol. 8, no. 2, pp. 273–305, Apr. 2012.

[5] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imag. Vis.*, vol. 40, no. 1, pp. 120–145, 2011.

[6] P. L. Combettes and J.-C. Pesquet, "Primal–dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators," *Set-Valued and Variational Analysis*, vol. 20, no. 2, pp. 307–330, 2012.

[7] L. Condat, "A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms," *J. Optim. Theory Appl.*, vol. 158, no. 2, pp. 460–479, 2013.

[8] B. C. Vũ, "A splitting algorithm for dual monotone inclusions involving cocoercive operators," *Adv. Comput. Math.*, vol. 38, no. 3, pp. 667–681, Apr. 2013.

[9] N. Komodakis and J.-C. Pesquet, "Playing with duality: An overview of recent primal–dual approaches for solving large-scale optimization problems," vol. 32, no. 6, pp. 31–54, Nov. 2015.

[10] J. J. Moreau, "Proximité et dualité dans un espace hilbertien," *Bull. Soc. Math. France*, vol. 93, pp. 273–299, 1965.

[11] C. Chaux, J.-C. Pesquet, and N. Pustelnik, "Nested iterative algorithms for convex constrained image recovery problems," *SIAM J. Imaging Sci.*, vol. 2, no. 2, pp. 730–762, June 2009.

[12] P.L. Combettes, Dinh Dũng, and B.C. Vũ, "Dualization of signal recovery problems," *Set-Valued and Variational Analysis*, vol. 18, pp. 373–404, 2010.

[13] F. Abboud, E. Chouzenoux, J.-C. Pesquet, J.-H. Chenot, and L. Laborelli, "Dual block-coordinate forward-backward algorithm with application to deconvolution and deinterlacing of video sequences," *J. Math. Imag. Vis.*, pp. 1–17, 2017.

[14] Y. Yu, "On decomposing the proximal map," in *Proc. of the 26th International Conference on Neural Information Processing Systems (NIPS)*, Dec. 2013, pp. 91–99.

[15] H.-J. M. Shi, S. Tu, Y. Xu, and W. Yin, "A primer on coordinate descent algorithms," 2017, preprint arXiv:1610.00040v2.

[16] P. L. Combettes and J.-C Pesquet, "Proximal thresholding algorithm for minimization over orthonormal bases," *SIAM J. Opt.*, vol. 18, no. 4, pp. 1351–1376, Nov. 2007.

[17] P.L. Davies and A. Kovac, "Local extremes, runs, strings and multiresolution," *Ann. Stat.*, vol. 29, no. 1, pp. 1–65, 2001.

[18] L. Condat, "A direct algorithm for 1D total variation denoising," *IEEE Signal Process. Lett.*, vol. 20, no. 11, pp. 1054–1057, 2013.

[19] L. Condat, "Discrete total variation: New definition and minimization," *SIAM J. Imaging Sciences*, vol. 10, no. 3, pp. 1258–1290, 2017.

[20] A. Chambolle and J. Darbon, "A parametric maximum flow approach for discrete total variation regularization," in *Image Processing and Analysis with Graphs: Theory and Practice*, R. Lukac, Ed., Digital Imaging and Computer Vision. CRC Press / Taylor and Francis, 2012.

[21] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, "Sparsity and smoothness via the fused lasso," *J. Royal Stat. Soc. B*, vol. 67, no. 1, pp. 91–108, June 2005.

[22] J. Friedman, T. Hastie, H. Höfling, and R. Tibshirani, "Pathwise coordinate optimization," *Ann. Appl. Statist.*, vol. 1, no. 2, pp. 302–332, 2007.

[23] M. Ayer, H.D. Brunk, G.M. Ewing, W.T. Reid, and E. Silverman, "An empirical distribution function for sampling with incomplete information," *Ann. Math. Statist.*, vol. 26, no. 4, pp. 641–647, 1995.

[24] S.J. Grotzinger and C. Witzgall, "Projections onto order simplexes," *Applied Mathematics and Optimization*, vol. 12, no. 1, pp. 247–270, June 1984.

[25] D. Cremers and J. Stühmer, "A fast projection method for connectivity constraints in image segmentation," in *Proc. of the 10th International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*, Hong Kong, China, Jan. 2015, vol. 8932, pp. 183–196.

[26] A. Chambolle, D. Cremers, and T. Pock, "A convex approach to minimal partitions," *SIAM J. Imaging Sci.*, vol. 5, no. 4, pp. 1113–1158, 2012.

[27] L. Condat, "A convex approach to K-means clustering and image segmentation," preprint hal-01504799, 2017.

[28] W. Miled, J.-C. Pesquet, and M. Parent, "A convex optimization approach for depth estimation under illumination variation," *IEEE Trans. Image Process.*, vol. 18, no. 4, pp. 813–830, 2009.

[29] C. Chaux, M. El Gheche, J. Farah, J.-C. Pesquet, and B. Pesquet-Popescu, "Disparity estimation from multicomponent images under illumination variation," *J. Math. Imag. Vis.*, vol. 47, no. 3, pp. 167–178, 2013.

[30] D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light," in *Proc. of the Conference on Computer Vision and Pattern Recognition (CVPR)*, Madison, WI, June 2003, vol. 1, pp. 195–202.