Epigraphical splitting for solving constrained convex optimization problems with proximal tools

Giovanni Chierchia, Nelly Pustelnik, Jean-Christophe Pesquet, Béatrice Pesquet-Popescu

HAL Id: hal-01372568

https://hal.archives-ouvertes.fr/hal-01372568

Submitted on 27 Sep 2016

Epigraphical splitting for solving constrained convex optimization problems with proximal tools – *extended version*

G. Chierchia · N. Pustelnik · J.-C. Pesquet · B. Pesquet-Popescu

Abstract We propose a proximal approach to deal with a class of convex variational problems involving nonlinear constraints. A large family of constraints, proven to be effective in the solution of inverse problems, can be expressed as the lower level set of a sum of convex functions evaluated over different blocks of the linearly-transformed signal. For such constraints, the associated projection operator generally does not have a simple form. We circumvent this difficulty by splitting the lower level set into as many epigraphs as functions involved in the sum. In particular, we focus on constraints involving ℓ_q-norms with $q \geq 1$, distance functions to a convex set, and $\ell_{1,p}$-norms with $p \in \{2, +\infty\}$. The proposed approach is validated in the context of image restoration by making use of constraints based on Non-Local Total Variation. Experiments show that our method leads to significant improvements in term of convergence speed over existing algorithms for solving similar constrained problems. A second application to a pulse shape design problem is provided in order to illustrate the flexibility of the proposed approach.

Keywords Iterative methods · Optimization · Epigraph · Projection · Proximal algorithms · Restoration · Total variation · Non-local regularization · Patch-based processing · Parallel algorithms

G. Chierchia · B. Pesquet-Popescu
Institut Mines-Télécom, Télécom ParisTech, CNRS LTCI, 75014 Paris, France. E-mail: {chierchi, beatrice.pesquet}@telecom-paristech.fr

N. Pustelnik
ENS Lyon, Laboratoire de Physique, CNRS-UMR 5672, F69007 Lyon, France. E-mail: nelly.pustelnik@ens-lyon.fr

J.-C. Pesquet
Université Paris-Est, LIGM, CNRS-UMR 8049, 77454 Marne-la-Vallée, France. E-mail: jean-christophe.pesquet@univ-paris-est.fr

1 Introduction

As an offspring of the wide interest in frame representations and sparsity promoting techniques for data recovery, proximal methods have become popular for solving large-size non-smooth convex optimization problems [1, 2]. The efficiency of these methods in the solution of inverse problems has been widely studied in the recent signal and image processing literature (see e.g. [3–8] and references therein). For instance, in the context of image restoration, when the noise is assumed to be zero-mean additive white Gaussian, the target signal can be recovered from the blurred and noisy observations $z \in \mathbb{R}^N$ by solving the following non-smooth optimization problem:

$$\underset{x \in \mathbb{R}^N}{\text{minimize}} \ \|Ax - z\|_2^2 + \lambda \|Fx\|_1,$$

where $A \in \mathbb{R}^{N \times N}$ is the matrix associated with the degradation blur, $F \in \mathbb{R}^{M \times N}$ (with $M \geq N$) is an analysis frame operator that sparsifies the target signal [9, 10], and λ is a positive regularization parameter. A constrained formulation of the above problem may be also considered [11], yielding

$$\underset{x \in \mathbb{R}^N}{\text{minimize}} \ \|Ax - z\|_2^2 \quad \text{s.t.} \quad \|Fx\|_1 \leq \eta,$$

where η is a positive constraint bound. Both optimization problems are equivalent for some specific values of λ and η, since there exist some conceptual Lagrangian equivalences between regularized and constrained formulations [12]. However, it has been recognized for a long time that incorporating constraints directly on the solution, instead of considering regularized functions, may often facilitate the choice of the involved parameters [13–21]. Indeed, the bound η may be easier to set than the parameter λ, as the former may be related to some physical properties of the target solution. This is one of the primary motivations in considering constrained formulations of convex optimization problems.

1.1 Proximal algorithms

The wide class of proximal algorithms can efficiently deal with constrained convex optimization problems, as they provide a unifying framework that allows one to address nonsmooth functions as well as hard constraints [1, 22–38]. The key tool in these methods is the *proximity operator* [39] of a proper lower-semicontinuous convex function φ from a real Hilbert space $\mathscr{H}$ to $]-\infty, +\infty]$, defined as

$$(\forall y \in \mathscr{H}) \qquad \mathrm{prox}_\varphi(y) = \underset{u \in \mathscr{H}}{\mathrm{argmin}} \; \frac{1}{2}\|u - y\|^2 + \varphi(u).$$

The proximity operator can be interpreted as a sort of subgradient step for the function φ, as $p = \mathrm{prox}_\varphi(y)$ is uniquely defined through the inclusion $y - p \in \partial \varphi(p)$. When the function is smooth, some proximal algorithms allow one to replace the proximity operator with a simpler gradient-descent computation. When a hard constraint is involved, instead, the proximity operator reduces to the *orthogonal projection* onto a nonempty closed convex subset $C \subset \mathscr{H}$, in the sense that

$$(\forall y \in \mathscr{H}) \qquad \mathrm{prox}_{\iota_C}(y) = P_C(y) = \underset{u \in C}{\mathrm{argmin}} \; \frac{1}{2}\|u - y\|^2,$$

where ι_C denotes the indicator function of C, equal to 0 on C and $+\infty$ otherwise.

In order to solve a convex optimization problem, proximal algorithms iterate a sequence of steps in which the proximity operators of the involved functions are evaluated at every iteration (e.g. $\mathrm{prox}_{\|\cdot - z\|_2^2}$ and $P_{\{\|\cdot\|_1 \leq \eta\}}$ in the constrained formulation mentioned earlier). The efficient computation of these operators is thus essential for dealing with large-size problems. Unfortunately, it turns out that a closed-form expression of the projection onto a closed convex set is available in a limited number of instances. Some well-known examples are the projections onto an hypercube, a closed half-space, an hyperslab, an affine set, a simplex and an ℓ_2-norm ball [2, 7, 40]. However, more sophisticated convex sets are usually necessary in a large number of inverse problems (e.g. the ℓ_1-ball in the above example), raising the issue of computing the associated projection operators.

When an expression of the direct projection is not available, a possible solution is to approximate the convex set by a half-space, which leads to the concept of subgradient projection. The main limitation of this approach is that the objective function must be strictly convex [41]. Recent works have proposed specific numerical methods to deal with an ℓ_1-ball [11, 42] and an $\ell_{1,\infty}$-ball [43], but they may prove inefficient in large-size problems, due to inner iterations.

1.2 Contributions

The objective of this paper is to propose an efficient splitting technique for solving the following class of constrained convex optimization problems:

Problem 1

$$\underset{x \in \mathscr{H}}{\mathrm{minimize}} \; \sum_{r=1}^{R} g_r(T_r x) \quad \mathrm{s.t.} \quad h(Fx) \leq \eta,$$

where $\eta \in \mathbb{R}$ and, for every $r \in \{1, \ldots, R\}$,

- T_r (resp. F) is a bounded linear operator from $\mathscr{H}$ to $\mathbb{R}^{N_r}$ (resp. $\mathbb{R}^M$),
- g_r is a proper lower-semicontinuous convex function from $\mathbb{R}^{N_r}$ to $]-\infty, +\infty]$,
- h is a proper lower-semicontinuous function from $\mathbb{R}^M$ to $]-\infty, +\infty]$ having a *block-wise decomposable* structure:

$$(\forall y \in \mathbb{R}^M) \quad h(y) = \sum_{\ell=1}^{L} h^{(\ell)}(\mathrm{y}^{(\ell)}),$$

with $\mathrm{y}^{(\ell)} \in \mathbb{R}^{M_\ell}$ for each $\ell \in \{1, \ldots, L\}$ and $M_1 + \cdots + M_L = M$.

Numerous constraints usually involved in the formulation of inverse problems can be modelled by a block-decomposable function. Popular examples are the Kullback-Leibler divergence [16], ℓ_2-norm composed by Anscombe transformation [19], ℓ_1-norm [11], $\ell_{1,\infty}$-norm [43], and total variation [21] or total generalized variation [20] semi-norms. A possible solution to deal with these constraints is to exploit the Lagrangian equivalence between projection and proximity operator, which boils down to the problem of finding the zero of a nonlinear equation [2, Section 6.6.1]. However, this approach turns out to be efficient only when the proximity operator admits a simple form, which is the case of Kullback-Leibler divergence [16] and ℓ_1-norm [11].

The present work aims at designing an efficient method to address Problem 1 when the projection onto the involved constraint does not have a closed-form expression and the standard approach mentioned above is infeasible or inefficient. More specifically:

(i) We propose a splitting technique that replaces the constraint in Problem 1 with a collection of epigraphs and a closed half-space constraint. So doing, we trade the problem of computing the projection onto the original constraint with the problem of computing the projection onto smaller epigraphs.

(ii) We enrich the list of functions for which the projection onto the associated epigraph can be efficiently computed. In this regard, we provide some theoretical results concerning the epigraphical projection of several functions of practical interest, such as the absolute value raised to a power $q \in [1, +\infty[$, the distance to a convex set and the ℓ_p-norm with $p \in \{2, +\infty\}$.

(iii) We illustrate through an image restoration example that regularity constraints based on Total Variation [44] and Non-Local Total Variation [45] can be efficiently handled by the proposed epigraphical splitting, which significantly speeds up the convergence (in terms of execution time) with respect to standard numerical solutions [11, 43].

1.3 Organization

The paper is organized as follows. In section 2, we detail a new splitting approach to deal with a constraint expressed as the lower level set of a decomposable function. Since the proposed splitting introduces some epigraphs in the minimization process, we provide in Section 3 the expression of specific epigraphical projections. To demonstrate the flexibility of our approach, we illustrate some numerical experiments in Sections 4 and 5: the former concerns an image recovery problem, while the latter is related to a pulse shape design for digital communications. Finally, some conclusions are drawn in Section 6.

Notation: Let $\mathscr{H}$ be a real Hilbert space endowed with the norm $\|\cdot\|$ and the scalar product $\langle\cdot\mid\cdot\rangle$. $\Gamma_0(\mathscr{H})$ denotes the set of proper lower-semicontinuous convex functions from $\mathscr{H}$ to $]-\infty,+\infty]$. Recall that a function $\varphi\colon\mathscr{H}\to\;]-\infty,+\infty]$ is proper if $\operatorname{dom}\varphi=\left\{y\in\mathscr{H}\mid\varphi(y)<+\infty\right\}$ is nonempty. The lower level set of φ at height $\zeta\in\mathbb{R}$ is the closed convex subset of $\mathscr{H}$ defined as $\operatorname{lev}_{\leq\zeta}\varphi=\left\{y\in\mathscr{H}\mid\varphi(y)\leq\zeta\right\}$, and the epigraph of $\varphi\in\Gamma_0(\mathscr{H})$ is the closed convex subset of $\mathscr{H}\times\mathbb{R}$ defined as $\operatorname{epi}\varphi=\left\{(y,\zeta)\in\mathscr{H}\times\mathbb{R}\mid\varphi(y)\leq\zeta\right\}$. A subgradient of φ at $y\in\mathscr{H}$ is the subset of $\mathscr{H}$ defined as $\partial\varphi(y)=\left\{t\in\mathscr{H}\mid(\forall u\in\mathscr{H})\;\varphi(u)\geq\varphi(y)+\langle t\mid u-y\rangle\right\}$. When φ is Gâteaux-differentiable at y, $\partial\varphi(y)=\{\nabla\varphi(y)\}$. Let C be a nonempty closed convex subset C of $\mathscr{H}$. The relative interior of C is denoted by $\operatorname{ri}C$. For every $y\in\mathscr{H}$, the proximity operator of φ reads $\operatorname{prox}_\varphi(y)=\operatorname{argmin}_{u\in\mathscr{H}}\frac{1}{2}\|u-y\|^2+\varphi(u)$. The indicator function $\iota_C\in\Gamma_0(\mathscr{H})$ of C is equal to 0 for $y\in C$ and $+\infty$ otherwise. The projection onto C reads $P_C(y)=\operatorname{argmin}_{u\in C}\|u-y\|$. The distance to C is given by $d_C(y)=\|y-P_C(y)\|$.

2 Proposed method

We now turn our attention to the constraint in Problem 1 and we illustrate how to deal with it when the associated projection does not have a closed form. More precisely, we assume that the function h in Problem 1 can be modelled as:

$$(\forall y\in\mathbb{R}^M)\qquad h(y)=\sum_{\ell=1}^L h_\ell(y^{(\ell)}),\tag{1}$$

where the generic vector y is decomposed into blocks of coordinates as follows

$$y=\left[\underbrace{\left(y^{(1)}\right)^\top}_{\operatorname{size}M_1},\ldots,\underbrace{\left(y^{(L)}\right)^\top}_{\operatorname{size}M_L}\right]^\top\in\mathbb{R}^M,\tag{2}$$

with $M_1+\cdots+M_L=M$ and, for every $\ell\in\{1,\ldots,L\}$, $y^{(\ell)}\in\mathbb{R}^{M_\ell}$ and $h_\ell\in\Gamma_0(\mathbb{R}^{M_\ell})$ is such that $\operatorname{ri}(\operatorname{dom}h_\ell)\neq\varnothing$.

2.1 Epigraphical splitting

Our approach consists of introducing an auxiliary vector $\zeta=\left(\zeta^{(\ell)}\right)_{1\leq\ell\leq L}\in\mathbb{R}^L$ in the minimization process, so that the constraint in Problem 1 can be decomposed into a collection of block-wise convex sets and a closed half-space:[1]

$$\sum_{\ell=1}^L h_\ell(y^{(\ell)})\leq\eta\quad\Leftrightarrow\quad\begin{cases}(\forall\ell\in\{1,\ldots,L\})\quad h_\ell(y^{(\ell)})\leq\zeta^{(\ell)},\\[2mm]\displaystyle\sum_{\ell=1}^L\zeta^{(\ell)}\leq\eta.\end{cases}\tag{3}$$

Consequently, Problem 1 can be equivalently formulated in the following form:

$$\operatorname*{minimize}_{(x,\zeta)\in\mathscr{H}\times\mathbb{R}^L}\quad\sum_{r=1}^R g_r(T_r x)\quad\text{s.t.}$$
$$\begin{cases}(\forall\ell\in\{1,\ldots,L\})\quad h_\ell\left([Fx]^{(\ell)}\right)\leq\zeta^{(\ell)},\\[2mm]\displaystyle\sum_{\ell=1}^L\zeta^{(\ell)}\leq\eta.\end{cases}\tag{4}$$

Note that the above minimization problem is defined with respect to x and ζ, so we have increased the dimensionality of our problem and we have replaced the lower level set of h with simpler constraints given by the epigraphs of $h_1,\ldots,h_L$.

2.2 Connections with proximal algorithms

Within the proposed constrained optimization framework, Problem (4) can be rewritten in a more compact form as follows:

Problem 2

$$\operatorname*{minimize}_{(x,\zeta)\in\mathscr{H}\times\mathbb{R}^M}\sum_{r=1}^R g_r(T_r x)\quad\text{s.t.}\quad\begin{cases}(Fx,\zeta)\in E,\\\zeta\in V,\end{cases}$$

where

$$E=\{(y,\zeta)\in\mathbb{R}^M\times\mathbb{R}^L\mid$$
$$\qquad(\forall\ell\in\{1,\ldots,L\})\quad(y^{(\ell)},\zeta^{(\ell)})\in\operatorname{epi}h_\ell\},\tag{5}$$
$$V=\left\{\zeta\in\mathbb{R}^L\mid 1_L^\top\zeta\leq\eta\right\},\tag{6}$$

with $1_L=(1,\ldots,1)^\top\in\mathbb{R}^L$.

The resolution of Problem 2 requires an efficient algorithm for dealing with large-size problems involving possibly nonsmooth functions and linear operators. As already mentioned in the introduction, we resort to proximal algorithms, which allow us to deal individually with the operators

---

[1] Note that the linear inequality over the auxiliary vector ζ can be also replaced by an equality, even though it makes little difference in our approach.

$(T_r)_{1 \le r \le R}$, F, $(\mathrm{prox}_{g_r})_{1 \le r \le R}$, P_E, and P_V. In the present case, we assume that $(\mathrm{prox}_{g_r})_{1 \le r \le R}$ have closed-form expressions. In addition, the projection onto V is well-known [2, Section 6.2.3], whereas the projection onto E is given by

$$(\forall (y, \zeta) \in \mathbb{R}^M \times \mathbb{R}^L) \qquad P_E(y, \zeta) = (p, \theta), \qquad (7)$$

where $\theta = (\theta^{(\ell)})_{1 \le \ell \le L} \in \mathbb{R}^L$, $p = (p^{(\ell)})_{1 \le \ell \le L} \in \mathbb{R}^M$ is block-wise decomposed as in (2), and

$$(\forall \ell \in \{1, \dots, L\}) \qquad (p^{(\ell)}, \theta^{(\ell)}) = P_{\mathrm{epi}\, h_\ell}(y^{(\ell)}, \zeta^{(\ell)}). \qquad (8)$$

Therefore, in order to solve Problem 2, we need to compute the projection onto $\mathrm{epi}\, h_\ell$ for each $\ell \in \{1, \dots, L\}$, which yields two potential benefits with respect to Problem 1. Firstly, the epigraphical projection involves the lower-dimensional problem of determining the projection onto the convex subset $\mathrm{epi}\, h_\ell$ of $\mathbb{R}^{M_\ell} \times \mathbb{R}$. Secondly, these projections can be computed in parallel, since they are defined over disjoint blocks [46]. An example of an algorithm that converges to a solution to Problem 2 (and thus to a solution of Problem 1) will be provided in Section 4.

2.3 Examples of epigraphical constraints

We now illustrate some examples of functions that can be handled with the epigraphical splitting presented above. The mathematical expression of the associated projection will be derived in Section 3.

(i) **ℓ_q-norm**. Let $q \in [1, +\infty[$, $(\tau^{(\ell)})_{1 \le \ell \le M} \in]0, +\infty[^M$, and

$$(\forall y \in \mathbb{R}^M) \qquad h(y) = \sum_{\ell=1}^M \tau^{(\ell)} |y^{(\ell)}|^q. \qquad (9)$$

Then, the above function can be modelled as in (1) with $L = M$ and, for every $\ell \in \{1, \dots, M\}$,

$$(\forall y^{(\ell)} \in \mathbb{R}) \qquad h_\ell(y^{(\ell)}) = \tau^{(\ell)} |y^{(\ell)}|^q. \qquad (10)$$

The corresponding epigraphical projection will be given in Proposition 2 for $q = 1$ and Proposition 3 for $q > 1$. Note that the ℓ_1 and ℓ_2 norms are widely used for the regularization of inverse problems [47, 48].

(ii) **Distance function**. Let $L \in \mathbb{N}^*$, $(\tau^{(\ell)})_{1 \le \ell \le L} \in]0, +\infty[^L$, $(q^{(\ell)})_{1 \le \ell \le L} \in [1, +\infty[^L$ and, for every $\ell \in \{1, \dots, L\}$, let $C^{(\ell)}$ be a nonempty closed convex subset of $\mathbb{R}^{M_\ell}$, and

$$(\forall y \in \mathbb{R}^M) \qquad h(y) = \sum_{\ell=1}^L \tau^{(\ell)} d_{C^{(\ell)}}^{q^{(\ell)}}(y^{(\ell)}). \qquad (11)$$

Then, the above function can be modelled as in (1) with, for every $\ell \in \{1, \dots, L\}$,

$$(\forall y^{(\ell)} \in \mathbb{R}^{M_\ell}) \qquad h_\ell(y^{(\ell)}) = \tau^{(\ell)} d_{C^{(\ell)}}^{q^{(\ell)}}(y^{(\ell)}). \qquad (12)$$

The corresponding epigraphical projection will be given in Proposition 4. Such a function is relevant for relaxing constraints on support or dynamics range [49].

(iii) **$\ell_{1,2}$-norm**. Let $L \in \mathbb{N}^*$, $(\tau^{(\ell)})_{1 \le \ell \le L} \in]0, +\infty[^L$, and

$$(\forall y \in \mathbb{R}^M) \qquad h(y) = \sum_{\ell=1}^L \tau^{(\ell)} \|y^{(\ell)}\|_2. \qquad (13)$$

Then, the above function can be modelled as in (1) with, for every $\ell \in \{1, \dots, L\}$,

$$(\forall y^{(\ell)} \in \mathbb{R}^{M_\ell}) \qquad h_\ell(y^{(\ell)}) = \tau^{(\ell)} \|y^{(\ell)}\|_2. \qquad (14)$$

The associated epigraphical projection will be given in Corollary 1. Note that the $\ell_{1,2}$-norm is useful to define multivariate sparsity constraints [50] or total variation bounds [44], which typically involve a sum of functions like (14) composed with linear operators corresponding to analysis transforms or gradient operators.

(iv) **$\ell_{1,\infty}$-norm**. Let $L \in \mathbb{N}^*$ and, for each $\ell \in \{1, \dots, L\}$, let $(\tau^{(\ell,m)})_{1 \le m \le M_\ell} \in]0, +\infty[^{M_\ell}$. We assume that:

$$(\forall y \in \mathbb{R}^M) \qquad h(y) = \sum_{\ell=1}^L \max_{1 \le m \le M_\ell} \tau^{(\ell,m)} |y^{(\ell,m)}|. \qquad (15)$$

Then, the above function can be modelled as in (1) with, for every $\ell \in \{1, \dots, L\}$,

$$(\forall y^{(\ell)} \in \mathbb{R}^{M_\ell}) \qquad h_\ell(y^{(\ell)}) = \max_{1 \le m \le M_\ell} \tau^{(\ell,m)} |y^{(\ell,m)}|. \qquad (16)$$

The corresponding epigraphical projection will be given in Proposition 5. When $\tau^{(\ell,m)} \equiv 1$, the function in (16) reduces to the standard infinity norm $\| \cdot \|_\infty$, which has recently attracted much interest for regularization purposes [43, 51–53].

2.4 Differences with existing splitting techniques

Several splitting techniques have been recently proposed to efficiently handle functions composed with a linear operator. In this context, the approaches inspired by the Alternating Direction Method of Multipliers are very popular, since they deal with optimization problems of the form

$$\underset{x \in \mathcal{H}}{\text{minimize}} \quad g_1(T_1 x) + g_2(x)$$

by resorting to the following reformulation

$$\underset{(x,v) \in \mathcal{H} \times \mathbb{R}^{N_1}}{\text{minimize}} \quad g_1(v) + g_2(x) \quad \text{s.t.} \quad T_1 x = v.$$

This kind of splitting has been often used in image restoration [54] and, more recently, for distributed optimization problems [55]. A similar form of splitting has been considered in [6], where the constraint $T_1 x = v$ is handled by computing the projection onto the nullspace of $[T_1 \ -\mathrm{Id}]$, which has a closed-form expression for some specific choices of T_1, such as circulant matrices involved in image restoration.

The solution that we propose in this work also introduces auxiliary variables. However, our objective is not to deal with linear transformations, but rather with a projection that does not have a closed-form expression. Consequently, the proposed solution departs from the usual splitting methods, in the sense that our approach leads to a collection of epigraphs, while the usual splitting techniques involve linear constraints.

3 Epigraphical projections

The key point in the proposed splitting is the introduction of some epigraphs in the minimization process, in order to facilitate the computational steps. Therefore, it is of paramount importance that the projection onto the epigraph can be efficiently computed. The problem of determining such an *epigraphical projection* is formalized in the following proposition (all the proofs are given in appendix).[2]

Proposition 1 *Let $\mathscr{H}$ be a real Hilbert space and let $\mathscr{H} \times \mathbb{R}$ be equipped with the standard product space norm. Let φ be a function in $\Gamma_0(\mathscr{H})$ such that $\mathrm{dom}\,\varphi$ is open. For every $(y, \zeta) \in \mathscr{H} \times \mathbb{R}$, the projector $P_{\mathrm{epi}\,\varphi}$ onto $\mathrm{epi}\,\varphi$ is given by:*

$$P_{\mathrm{epi}\,\varphi}(y, \zeta) = (p, \theta), \tag{17}$$

where

$$\begin{cases} p = \mathrm{prox}_{\frac{1}{2}(\max\{\varphi - \zeta, 0\})^2}(y), \\ \theta = \max\{\varphi(p), \zeta\}. \end{cases} \tag{18}$$

The previous result shows that the proximity operator in (18) plays a prominent role in the calculation of the projection onto $\mathrm{epi}\,\varphi$. We now provide several examples of function φ for which this proximity operator admits a simple form.

Proposition 2 *Let $\tau \in]0, +\infty[$. Assume that*

$$(\forall y \in \mathbb{R}) \qquad \varphi(y) = \tau|y|. \tag{19}$$

For every $(y, \zeta) \in \mathbb{R} \times \mathbb{R}$, $P_{\mathrm{epi}\,\varphi}(y, \zeta) = (p, \theta)$ is given by

$$p = \begin{cases} y, & \text{if } \tau|y| \leq \zeta, \\ \dfrac{\mathrm{sign}(y)}{1 + \tau^2} \max\{|y| + \tau\zeta, 0\}, & \text{otherwise.} \end{cases} \tag{20}$$

and $\theta = \max\{\tau|p|, \zeta\}$.

The above result follows by the fact that $\frac{1}{2}(\max\{\tau|y| - \zeta, 0\})^2 = (\tau^2/2)y^2 - \tau\zeta|y| + \zeta^2/2$ if $\tau|y| > \zeta$ and 0 otherwise, for which the proximity operator is known [24, Example 4.6]. This example clearly shows that Proposition 1 allows us to leverage the results on proximity operators already known in the literature. The following proposition considers the case when the absolute value is raised by a positive power.

---

² Alternative characterizations of the epigraphical projection can be found in [40, Prop. 9.17 and 28.28], whereas the result reported in [2, Section 6.6.2] refers to a preliminary version of this paper.

Proposition 3 *Let $q \in]1, +\infty[$ and $\tau \in]0, +\infty[$. Assume that*

$$(\forall y \in \mathbb{R}) \qquad \varphi(y) = \tau|y|^q. \tag{21}$$

For every $(y, \zeta) \in \mathbb{R} \times \mathbb{R}$, $P_{\mathrm{epi}\,\varphi}(y, \zeta) = (p, \theta)$ is given by

$$p = \begin{cases} \mathrm{sign}(y)\chi_0, & \text{if } \zeta \leq 0, \\ y, & \text{if } \zeta > 0 \text{ and } \tau|y|^q \leq \zeta, \\ \mathrm{sign}(y)\chi_\zeta, & \text{otherwise,} \end{cases} \tag{22}$$

and $\theta = \max\{\tau|p|^q, \zeta\}$, where $\chi_\cdot$ is the unique solution on $[(\cdot/\tau)^{1/q}, +\infty[$ of the following equation

$$q\tau^2\chi^{2q-1} - q\tau\zeta\chi^{q-1} + \chi = |y|. \tag{23}$$

Note that, when q is a rational number, (23) is equivalent to a polynomial equation for which either closed form solutions are known or standard numerical solutions exist.

The previous propositions allows us to establish a result concerning the distance function to a convex set.

Proposition 4 *Let C be a nonempty convex subset of $\mathscr{H}$. Let $q \in [1, +\infty[$, $\tau \in]0, +\infty[$, and $\zeta \in \mathbb{R}$. Assume that*

$$(\forall y \in \mathscr{H}) \qquad \varphi(y) = \tau d_C^q(y). \tag{24}$$

For every $(y, \zeta) \in \mathscr{H} \times \mathbb{R}$, $P_{\mathrm{epi}\,\varphi}(y, \zeta) = (p, \theta)$ is given by

$$p = \begin{cases} y, & \text{if } y \in C, \\ \alpha y + (1 - \alpha)P_C(y), & \text{otherwise,} \end{cases} \tag{25}$$

and $\theta = \max\{\tau d_C^q(p), \zeta\}$, where

$$\alpha = \frac{\mathrm{prox}_{\frac{1}{2}(\max\{\tau|\cdot|^q - \zeta, 0\})^2}\big(d_C(y)\big)}{d_C(y)} \tag{26}$$

and the above expression is provided by Propositions 2–3.

In the case when $q = 1$ and $C = \{z\}$ for some $z \in \mathscr{H}$, the previous result can be specialized with $d_C(y) = \|y - z\|$ and $P_C(y) = z$. So doing, we obtain a corollary about the Euclidean norm, for which the corresponding epigraph is called the *Lorentz cone* and the associated epigraphical projection is known in the literature (see e.g. [2, Section 6.3.2]).

Corollary 1 *Let $\tau \in]0, +\infty[$, $\zeta \in \mathbb{R}$ and $z \in \mathscr{H}$. Assume that*

$$(\forall y \in \mathscr{H}) \qquad \varphi(y) = \tau\|y - z\|. \tag{27}$$

For every $(y, \zeta) \in \mathscr{H} \times \mathbb{R}$, $P_{\mathrm{epi}\,\varphi}(y, \zeta) = (p, \theta)$ is given by

$$p = \begin{cases} z, & \text{if } y = z, \\ y, & \text{if } \tau\|y - z\| \leq \zeta, \\ \alpha y + (1 - \alpha)z, & \text{otherwise,} \end{cases} \tag{28}$$

and $\theta = \max\{\tau\|p - z\|, \zeta\}$, where

$$\alpha = \frac{1}{1 + \tau^2} \max\left\{1 + \frac{\tau\zeta}{\|y - z\|}, 0\right\}. \tag{29}$$

We conclude the section with a result about the weighted maximum of absolute values. When the weights are all equal, this function reduces to the standard infinity norm $\|\cdot\|_\infty$, for which the expression of the epigraphical projection has been recently given in [56]. The following proposition provides a slightly more general result.

Proposition 5 *Let $(\tau^{(m)})_{1\leq m\leq M} \in \,]0,+\infty[^M$. Assume that*

$$(\forall y \in \mathbb{R}^M) \qquad \varphi(y) = \max_{1\leq m\leq M} \tau^{(m)}|y^{(m)}|, \tag{30}$$

where the values $(\nu^{(m)} = \tau^{(m)}|y^{(m)}|)_{1\leq m\leq M_\ell}$ are in ascending order. Then, for every $(y,\zeta) \in \mathbb{R}^M \times \mathbb{R}$, $P_{\mathrm{epi}\,\varphi}(y,\zeta) = (p,\theta)$ is given by

$$p = \left(\mathrm{sign}(y^{(m)}) \min\left\{ |y^{(m)}|, \theta/\tau^{(m)} \right\} \right)_{1\leq m\leq M} \tag{31}$$

and

$$\theta = \max\left\{ \frac{\zeta + \sum\limits_{m=\overline{m}}^{M} \nu^{(m)}(\tau^{(m)})^{-2}}{1 + \sum\limits_{m=\overline{m}}^{M} (\tau^{(m)})^{-2}}, 0 \right\}, \tag{32}$$

where $\overline{m}$ is the unique integer in $\{1,\dots,M+1\}$ such that

$$\nu^{(\overline{m}-1)} < \frac{\zeta + \sum\limits_{m=\overline{m}}^{M} \nu^{(m)}(\tau^{(m)})^{-2}}{1 + \sum\limits_{m=\overline{m}}^{M} (\tau^{(m)})^{-2}} \leq \nu^{(\overline{m})} \tag{33}$$

(with the conventions $\sum_{m=1}^{0} \cdot = \sum_{m=M+1}^{M} \cdot = 0$, $\nu^{(0)} = -\infty$ and $\nu^{(M+1)} = +\infty$).

4 Image recovery problems

We propose to evaluate the performance of the proposed epigraphical solution in the context of image restoration, where $\mathscr{H} = \mathbb{R}^N$, $\bar{x} \in \mathbb{R}^N$ denotes the image to be recovered, and $z \in \mathbb{R}^K$ is an observation vector such that

$$z = DA\bar{x} + b. \tag{34}$$

We assume that $A \in \mathbb{R}^{N\times N}$ is a blurring operator, $D \in \mathbb{R}^{K\times N}$ is a decimation operator,[3] and $b \in \mathbb{R}^K$ is a realization of zero-mean white Gaussian noise.

---

[3] D thus corresponds to $K \leq N$ lines of the identity $N \times N$ matrix.

4.1 Constrained formulation

A usual approach to recover $\bar{x}$ from the degraded observations is to follow a variational approach that aims at solving a convex optimization problem defined as:

$$\underset{x\in[0,255]^N}{\text{minimize}} \ \|DAx - z\|_2^2 \quad \text{s.t.} \quad h(Fx) \leq \eta, \tag{35}$$

where $\eta \geq 0$ and $F \in \mathbb{R}^{M\times N}$ is the linear operator associated with an analysis transform. Hereabove, the quadratic term aims at insuring that the solution is close to the observations, while the constraint imposes some regularity on the solution. Regularization is essential for solving such an ill-posed problem, as it allows one to select, among competing solutions, the one that presents some form of *parsimony* or *smoothness*.

Natural images usually exhibit a smooth spatial behaviour, except around some locations (e.g. object edges), where discontinuities arise. Therefore, the quality of the results obtained through the aforementioned variational approach strongly depends on the ability of the operator F and the function h to model such a specific kind of regularity. Among the sophisticated regularization terms that have been recently developed in the field of image restoration, the most commonly-used ones can be generalized as

$$(\forall y \in \mathbb{R}^M) \qquad h(y) = \sum_{\ell=1}^{L} \|\mathsf{y}^{(\ell)}\|_p^q, \tag{36}$$

where $q \geq 1$, $p \geq 1$ and y is block-decomposed as in (2).

The above function can model the ℓ_q-norm in (9) when $p = q$ and $L = M$. This norm has been largely used in combination with a sparsifying transform F, such as frames [24] or Laplacian operators. Indeed, the case $q = 2$ leads to Tikhonov regularization, whereas the case $q = 1$ yields a sparsity-inducing regularization [57,58].

The above function also matches the $\ell_{1,2}$-norm in (13) when $q = 1$ and $p = 2$. Such a mixed-norm is used in the classical *total variation* (TV) regularization to penalize the image gradient [44]. Although TV has emerged as a simple and successful convex optimization tool, it often fails to preserve textures, details and fine structures, because they are hardly distinguishable from noise. To improve this behaviour, the TV model has been extended by using some generalizations based on higher-order spatial differences [59] or a non-locality principle [45,60,61]. The latter yields the *non-local-TV* operator, defined as

$$Fx = \begin{bmatrix} F_1 x \\ \dots \\ F_N x \end{bmatrix} = \begin{bmatrix} \left(\omega_{1,n}(x^{(1)} - x^{(n)}) \right)_{n\in\mathscr{N}_1} \\ \dots \\ \left(\omega_{N,n}(x^{(N)} - x^{(n)}) \right)_{n\in\mathscr{N}_N} \end{bmatrix}, \tag{37}$$

where, for every $\ell \in \{1,\dots,N\}$, $\mathscr{N}_\ell \subset \{1,\dots,N\} \setminus \{\ell\}$ denotes the neighborhood of ℓ and $(\omega_{\ell,n})_{n\in\mathscr{N}_\ell}$ are positive

weights. Note that (37) reduces to the classical gradient operator when $\mathcal{N}_\ell$ only contains the horizontal/vertical nearest neighbours of ℓ and $\omega_{\ell,n} \equiv 1$.[4]

Several types of $\boldsymbol{\ell}_{q,p}$-norms can be used to penalize the nonlocal gradient. The case $q = p = 2$ (i.e. the $\boldsymbol{\ell}_2$-norm) may be seen as a variational extension of *NonLocal Means* [60], while the case $q = p = 1$ (i.e. the $\boldsymbol{\ell}_1$-norm) leads to *NonLocal Medians* [63]. The case $q = 1$ and $p \in \{2, +\infty\}$ (i.e. the $\boldsymbol{\ell}_{1,p}$-norm) is instead preferred in color imagery [64, 65]. In our experiments, we will present the results of $\boldsymbol{\ell}_{q,p}$-non-local-TV for several values of q and p, in order to show that the proposed epigraphical splitting may be more efficient that the existing solutions for handling this class of constraints.

4.2 Optimization method

As mentioned in the introduction, proximal algorithms can efficiently solve Problem (35) as long as the proximity operator of the involved functions can be efficiently computed. Unfortunately, the projection onto the lower level set of (36) cannot be computed in closed form. A possible approach to circumvent this issue is given by the epigraphical splitting presented in Section 2. Indeed, Problem (35) can be rewritten as Problem 2 with $R = 2$, $T_1 = A$, $T_2 = I$, $g_1 = \|D \cdot -z\|_2^2$, $g_2 = \iota_{[0,255]^N}$ and $h_\ell = |\cdot|^q$ (for $p = q \geq 1$) or $h_\ell = \|\cdot\|_p$ (for $q = 1$ and $p \in \{2, +\infty\}$). In order to solve this specific instance of Problem 2, we will consider two proximal methods: the primal-dual algorithm called *Monotone+Lipschitz Forward Backward Forward* (M+LFBF) [34], and the primal algorithm called *Simultaneous-Direction Method of Multipliers* (SDMM) [1]. According to the general results in [34, Theorem 4.2] and [1], the sequence $\left(x^{[i]}\right)_{i \in \mathbb{N}}$ generated by M+LFBF or SDMM is guaranteed to converge to a (global) minimizer of Problem 2.

4.3 Numerical results

The aim of our experiments is twofold. Firstly, we will display an example of image reconstructed with the different types of non-local-TV regularization, in order to evaluate the visual impact of each regularization. Secondly, in order to show the efficiency of the proposed epigraphical splitting, we will compare the execution time required by M+LFBF and SDMM to solve Problem (35) using: 1) the epigraphical approach, and 2) the standard approach that employs specific numerical methods to deal with a $\boldsymbol{\ell}_1$- or $\boldsymbol{\ell}_{1,2}$-ball [11] and a $\boldsymbol{\ell}_{1,\infty}$-ball [43].

---

[4] Methods for building $\mathcal{N}_\ell$ and setting the associated weights are described in [60–62]. In our experiments, we use the $\boldsymbol{\ell}_{q,p}$-TV regularization to obtain an estimate of x, which we subsequently use to compute the weights through the *self-similarity* measure proposed in [62].

Fig. 1 illustrates the restoration of a RGB color image ($N = 320 \times 480 \times 3$) degraded by a 3×3 uniform blur (applied component-by-component), a decimation that randomly removes 60% of the pixels ($K = 0.4 \times N$) and an additive white Gaussian noise with variance 10^2. The constraint bounds η for the several $\boldsymbol{\ell}_{q,p}$-balls were hand-tuned in order to achieve the best SNR values. These results show the interest of considering non-local operators and $\boldsymbol{\ell}_{1,\infty}$-norms for modelling the regularity present in color images.

Fig. 2 compares the speed of the considered algorithms (w.r.t. the example in Fig. 1), showing the distance $\|x^{[i]} - x^{[\infty]}\|$ as a function of the execution time for the epigraphical approach and the approach based on the direct computation of projections. The stopping criterion is set to $\|x^{[i+1]} - x^{[i]}\| \leq 10^{-5}\|x^{[i]}\|$, and $x^{[\infty]}$ denotes the solution obtained when such a criterion is reached. Note that $x^{[\infty]}$ may not be unique, hence it was computed for each algorithm independently. These plots indicate that the proposed epigraphical approach yields a faster convergence.

Since the constraint bound η may not be known precisely, it is important to evaluate the impact of its choice on our method performance (it is out of the scope of this paper to devise an optimal strategy to set this bound). In Tables 1-4, we compare the execution times of epigraphical and direct approaches for different choices of regularization constraints and values of η. However, in order to reduce the computation burden of the direct projection onto the $\boldsymbol{\ell}_{1,\infty}$-ball, we present the results obtained with the grayscale image *boat* cropped at 256×256.[5] The stopping criterion is set to $\|x^{[i+1]} - x^{[i]}\| \leq 10^{-4}\|x^{[i]}\|$. For more readability, the values of η are expressed as a multiplicative factor of the $\boldsymbol{\ell}_{q,p}$-non-local-TV semi-norm evaluated on the original image.

- Tables 1 and 2 report the comparison for $\boldsymbol{\ell}_{1,2}$-TV and $\boldsymbol{\ell}_{1,\infty}$-TV, respectively. The convergence times indicate that the epigraphical approach yields a faster convergence although it requires more iterations in order to converge. This can be explained by the computational cost of the subiterations required by the direct projections onto the $\boldsymbol{\ell}_{1,p}$-ball. Moreover, the numerical results show that errors within $\pm 20\%$ from the optimal value for η lead to SNR variations within 2%.

- Tables 3 and 4 collects the results of $\boldsymbol{\ell}_{1,2}$-non-local-TV and $\boldsymbol{\ell}_{1,\infty}$-non-local-TV for different sizes of the neighbourhood $\mathcal{N}_\ell$. The convergence times show that the epigraphical approach is faster than the direct one for both considered algorithms. Moreover, it can be noticed that errors within $\pm 20\%$ from the optimal bound value lead to SNR variations within 1%.

Note that our codes were developed in MATLAB R2011b (the operators F and $F^\top$ are implemented in C) and executed on an Intel Xeon CPU at 2.80 GHz and 8 GB of RAM.

---

[5] Note that the $\boldsymbol{\ell}_{1,2}$-norm performs better on grayscale images.

(a) Original image.

(b) Degraded image.

(c) $\ell_{2,2}$-non-local-TV: 18.12–0.761.

(d) $\ell_{1,1}$-TV: 17.78–0.787.

(e) $\ell_{1,2}$-TV: 18.36–0.821.

(f) $\ell_{1,\infty}$-TV: 18.91–0.824.

(g) $\ell_{1,1}$-non-local-TV: 18.93–0.828.

(h) $\ell_{1,2}$-non-local-TV: 19.47–0.839.

(i) $\ell_{1,\infty}$-non-local-TV: **20.17–0.847**.
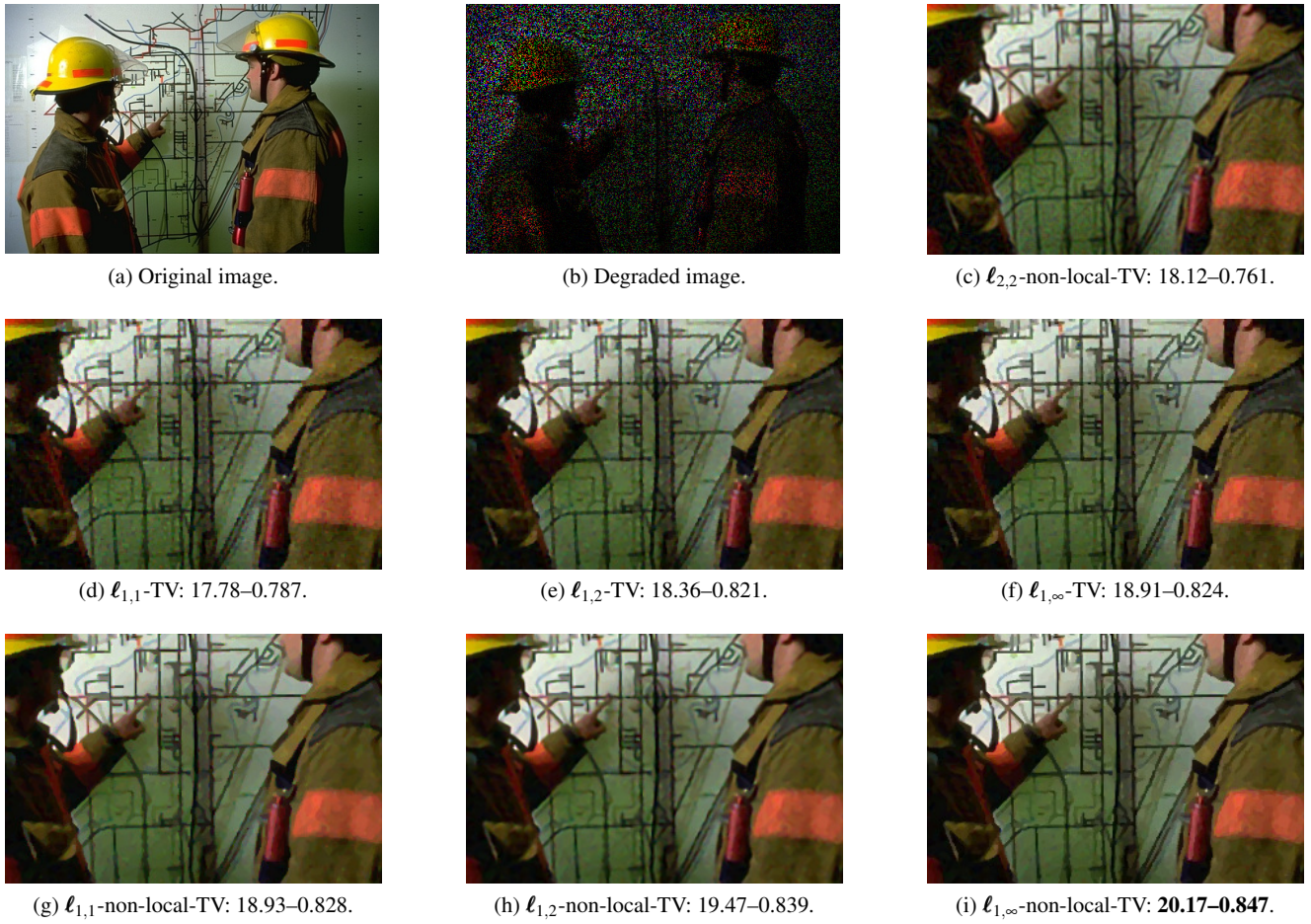
Fig. 1 Comparison of SNR (dB) – SSIM [66] indices for an example of restoration.

(a) $\ell_{1,2}$-TV.

(b) $\ell_{1,\infty}$-TV.

(c) $\ell_{1,2}$-non-local-TV.

(d) $\ell_{1,\infty}$-non-local-TV.

Fig. 2 Plots (distance to $x^{[\infty]}$ vs execution time) comparing the epigraphical and direct approaches implemented with M+LFBF and SDMM.

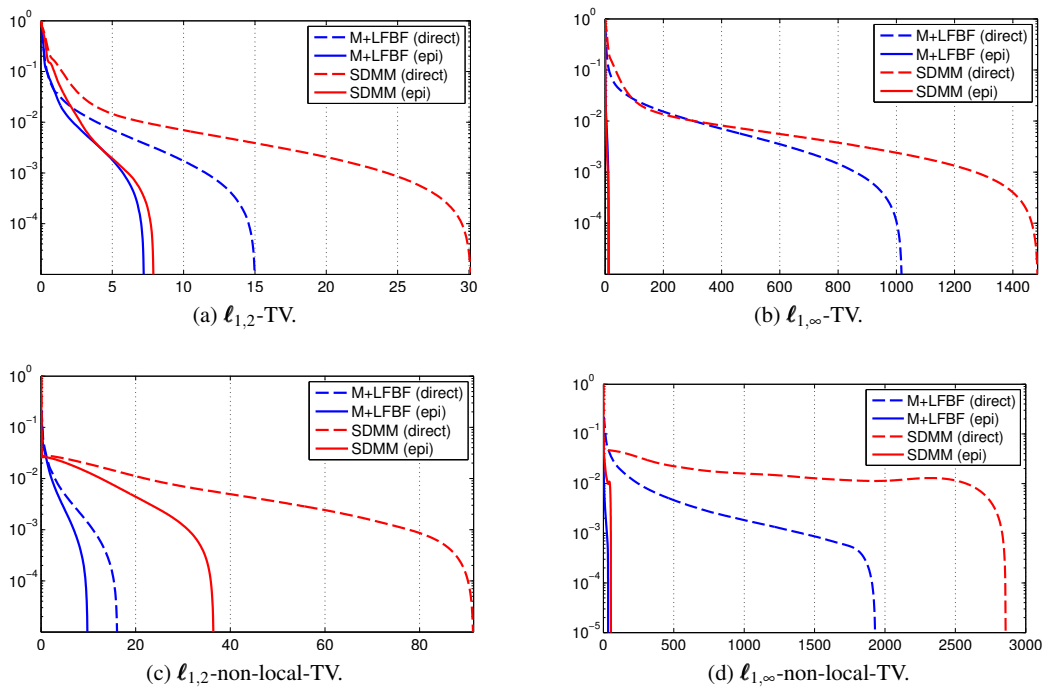**Table 1** Results for the $\ell_{1,2}$-TV constraint and different values of η (w.r.t. the grayscale image *boat* cropped at 256×256)

η	SNR (dB) – SSIM	SDMM					M+LFBF				
		direct		epigraphical		speed up	direct		epigraphical		speed up
		# iter.	sec.	# iter.	sec.		# iter.	sec.	# iter.	sec.	
0.45	19.90 – 0.733	107	6.07	174	2.03	2.99	113	6.15	182	3.49	1.76
0.50	20.18 – 0.745	117	6.95	159	1.95	3.57	116	6.97	168	3.44	2.03
0.56	**20.23 – 0.745**	**129**	**8.36**	**153**	**1.90**	**4.41**	**124**	**8.17**	**159**	**3.01**	**2.72**
0.62	20.16 – 0.737	141	9.44	155	1.83	5.16	131	8.62	159	3.26	2.65
0.67	20.00 – 0.724	154	10.20	162	2.17	4.71	140	10.00	164	2.84	3.52

Table 2 Results for the $\ell_{1,\infty}$-TV constraint and different values of η (w.r.t. the grayscale image *boat* cropped at 256×256)

η	SNR (dB) – SSIM	SDMM					M+LFBF				
		direct		epigraphical		speed up	direct		epigraphical		speed up
		# iter.	sec.	# iter.	sec.		# iter.	sec.	# iter.	sec.	
0.45	19.52 – 0.726	160	312.55	231	3.89	80.43	183	347.10	252	6.43	53.96
0.50	19.71 – 0.728	168	342.01	215	3.75	91.31	185	368.24	236	5.83	63.17
0.56	**19.71 – 0.734**	**180**	**373.60**	**211**	**3.49**	**106.93**	**189**	**386.29**	**229**	**5.53**	**69.91**
0.62	19.59 – 0.715	196	412.68	216	3.67	112.50	198	411.04	229	5.86	70.15
0.67	19.39 – 0.698	211	448.77	223	3.76	119.27	207	437.66	234	5.76	75.96

Table 3 Results for the $\ell_{1,2}$-non-local-TV constraint and different values of η (w.r.t. the grayscale image *boat* cropped at 256×256)

η	SNR (dB) – SSIM	SDMM					M+LFBF				
		direct		epigraphical		speed up	direct		epigraphical		speed up
		# iter.	sec.	# iter.	sec.		# iter.	sec.	# iter.	sec.	
		Neighbourhood size: 3×3									
0.43	20.82 – 0.757	208	20.67	211	10.93	1.89	82	6.95	93	3.76	1.85
0.49	20.97 – 0.765	167	16.84	177	9.01	1.87	75	6.61	83	3.47	1.91
0.54	21.02 – 0.767	147	15.31	157	7.93	1.93	71	6.45	77	3.15	2.04
0.59	20.98 – 0.764	134	14.44	148	7.67	1.88	72	6.58	77	3.24	2.03
0.65	20.88 – 0.757	133	14.82	136	7.11	2.08	76	7.53	80	3.27	2.30
		Neighbourhood size: 5×5									
0.43	21.00 – 0.766	301	56.03	343	45.18	1.24	82	8.51	90	5.43	1.57
0.49	21.15 – 0.773	260	49.03	302	39.64	1.24	75	7.90	81	4.90	1.61
0.54	**21.20 – 0.775**	**242**	**46.31**	**283**	**37.72**	**1.23**	**71**	**8.26**	**75**	**4.47**	**1.85**
0.59	21.17 – 0.773	231	46.20	268	36.56	1.26	70	7.94	74	4.49	1.77
0.65	21.08 – 0.767	220	44.64	252	34.46	1.30	73	8.40	76	4.59	1.83

Table 4 Results for the $\ell_{1,\infty}$-non-local-TV constraint and different values of η (w.r.t. the grayscale image *boat* cropped at 256×256)

η	SNR (dB) – SSIM	SDMM					M+LFBF				
		direct		epigraphical		speed up	direct		epigraphical		speed up
		# iter.	sec.	# iter.	sec.		# iter.	sec.	# iter.	sec.	
		Neighbourhood size: 3×3									
0.43	20.78 – 0.762	434	1470.46	449	25.03	58.76	225	730.26	244	12.35	59.15
0.49	20.86 – 0.764	395	1319.64	413	22.86	57.72	221	692.25	237	11.92	58.08
0.54	20.83 – 0.760	363	1193.61	382	21.46	55.62	217	667.50	233	11.46	58.22
0.59	20.73 – 0.752	340	1093.26	354	19.77	55.30	216	653.79	230	11.67	56.01
0.65	20.58 – 0.740	322	1007.55	336	18.64	54.06	216	643.00	229	11.45	56.18
		Neighbourhood size: 5×5									
0.43	20.91 – 0.769	384	2069.62	452	64.42	32.13	233	863.01	252	18.47	46.73
0.49	20.97 – 0.767	326	1700.34	412	58.66	28.99	231	822.06	247	18.36	44.77
0.54	**20.98 – 0.771**	**290**	**1476.98**	**389**	**55.35**	**26.69**	**229**	**787.61**	**245**	**17.90**	**43.99**
0.59	20.88 – 0.759	276	1336.16	374	52.64	25.38	230	772.42	245	17.57	43.96
0.65	20.75 – 0.749	268	1220.14	362	51.45	23.72	231	760.86	245	17.81	42.72

5 Pulse shape design

We revisit the problem of a pulse shape design for digital communications presented in [49], where it was addressed in terms of constrained convex optimization. Denote by $x = (x^{(\ell)})_{0 \le \ell \le N-1} \in \mathbb{R}^N$ the pulse, by $\chi = (\chi^{(\ell)})_{0 \le \ell \le N-1}$ its discrete Fourier transform, and assume that the underlying sampling rate is 2560 Hz. The following constraints arise from engineering specifications [49].

(i) The modulus of the Fourier transform should not exceed the bound $\gamma = 10^{-3/2}$ beyond 300 Hz. This leads to

$$(\forall \ell \in \mathbb{D}_1) \qquad C_1^{(\ell)} = \{x \in \mathbb{R}^N \mid |\chi^{(\ell)}| \le \gamma\}, \qquad (38)$$

where $\mathbb{D}_1$ represents frequencies beyond 300 Hz.

(ii) Vanishing frequencies of χ at the zero frequency and at integer multiples of 50 Hz:

$$C_2 = \{x \in \mathbb{R}^N \mid (\forall \ell \in \mathbb{D}_2) \quad \chi^{(\ell)} = 0\} \qquad (39)$$

where $\mathbb{D}_2$ denotes the frequencies where the Fourier transform χ vanishes.

(iii) Symmetry of the pulse and its mid-point value should be equal to 1:

$$C_3 = \{x \in \mathbb{R}^N \mid x^{(N/2)} = 1 \quad \text{and}$$
$$(\forall \ell \in \{0, \dots, N/2\}) \quad x^{(\ell)} = x^{(N-1-\ell)})\}. \quad (40)$$

(iv) Pulse duration should be 50 ms and it should have periodic zero crossings every 3.125 ms:

$$C_4 = \{x \in \mathbb{R}^N \mid (\forall \ell \in \mathbb{D}_4) \quad x^{(\ell)} = 0\}, \qquad (41)$$

where $\mathbb{D}_4$ is the set of time indices in the zero areas.

(v) Pulse energy should be as small as possible, in order to avoid interference with other systems.

5.1 Optimization method

Some of the aforementioned constraints are incompatible (e.g. the second and the fourth ones). Hence, in order to make the problem feasible, we propose to replace the convex sets $(C_1^{(\ell)})_{1 \le \ell \le N}$ with the following constraint:

$$C_1 = \{x \in \mathbb{R}^N \mid \sum_{\ell \in \mathbb{D}_1} d_{C_1^{(\ell)}}^q(\chi^{(\ell)}) \le \eta\}, \qquad (42)$$

where $\eta > 0$ and $q \in [1, +\infty[$. Therefore, the resulting optimization problem reads

$$\underset{x \in \mathbb{R}^N}{\text{minimize}} \quad \|x\|_2^2 + \sum_{s=1}^{4} \iota_{C_s}(x). \qquad (43)$$

As mentioned in the introduction, proximal algorithms can efficiently solve the above problem as long as the projection associated to each constraint $(C_s)_{1 \le s \le 5}$ can be efficiently

computed. Unfortunately, the projection onto C_1 cannot be computed in closed form. We thus propose to circumvent this issue by using the epigraphical splitting presented in Section 2. Indeed, (43) can be rewritten as Problem 2 with $R = 4$, $T_1 = \cdots = T_4 = I$, $g_1 = \|\cdot\|_2^2$, $g_s = \iota_{C_s}$ for each $s \in \{2, \dots, 4\}$, $L = N$ and

$$E = \{(y, \zeta) \in \mathbb{R}^N \times \mathbb{R}^N \mid (\forall \ell \in \mathbb{D}_1) \quad (y^{(\ell)}, \zeta^{(\ell)}) \in \operatorname{epi} d_{C_1^{(\ell)}}^q\}, \qquad (44)$$

$$V = \{\zeta \in \mathbb{R}^N \mid 1_{\mathbb{D}_1}^\top \zeta \le \eta\}. \qquad (45)$$

5.2 Numerical results

Fig. 3 presents state-of-the-art results (from [49]), while Fig. 4 presents the results obtained with the proposed solution for $q = 2$ and different values of η. Note that for large values of η, the solutions converge to a solution of the unconstrained (without imposing C_1) problem (cf. Fig 3-right). It is also interesting to experimentally observe that the estimated pulse for the smallest value of η leading an admissible solution (cf. Fig 4-left) is similar to the solution proposed in [49] (cf. Fig. 3 - middle). As illustrated by these experiments, the proposed approach allows us to gain more design flexibility at the expense of a small additional computational cost.

6 Conclusions

We have proposed a new epigraphical technique to deal with constrained convex optimization problems with the help of proximal algorithms. In particular, we have turned our attention to constraints based on distance functions and weighted $\ell_{q,p}$-norms. In the context of 1D signals, we have shown that constraints based on distance functions are useful for pulse shape design. In the context of images, we have used $\ell_{1,p}$-balls to promote block-sparsity of analysis representations. The obtained results demonstrate the good performance (in terms of image quality) of non-local measures and $\ell_{1,\infty}$-norm in color imagery. Nevertheless, it would be also interesting to consider alternative applications of $\ell_{1,\infty}$-norms, such as regression problems [67].

Furthermore, the experimental part indicates that the epigraphical method converges faster than the approach based on the direct computation of the projections via standard iterative solutions. Parallelization of our codes should even allow us to accelerate them [46]. Note that, although the considered application involves two constraint sets, the proposed approach can handle an arbitrary number of convex constraints. In addition, the epigraphical approach could also be used to develop approximation methods for addressing more general convex constraints.
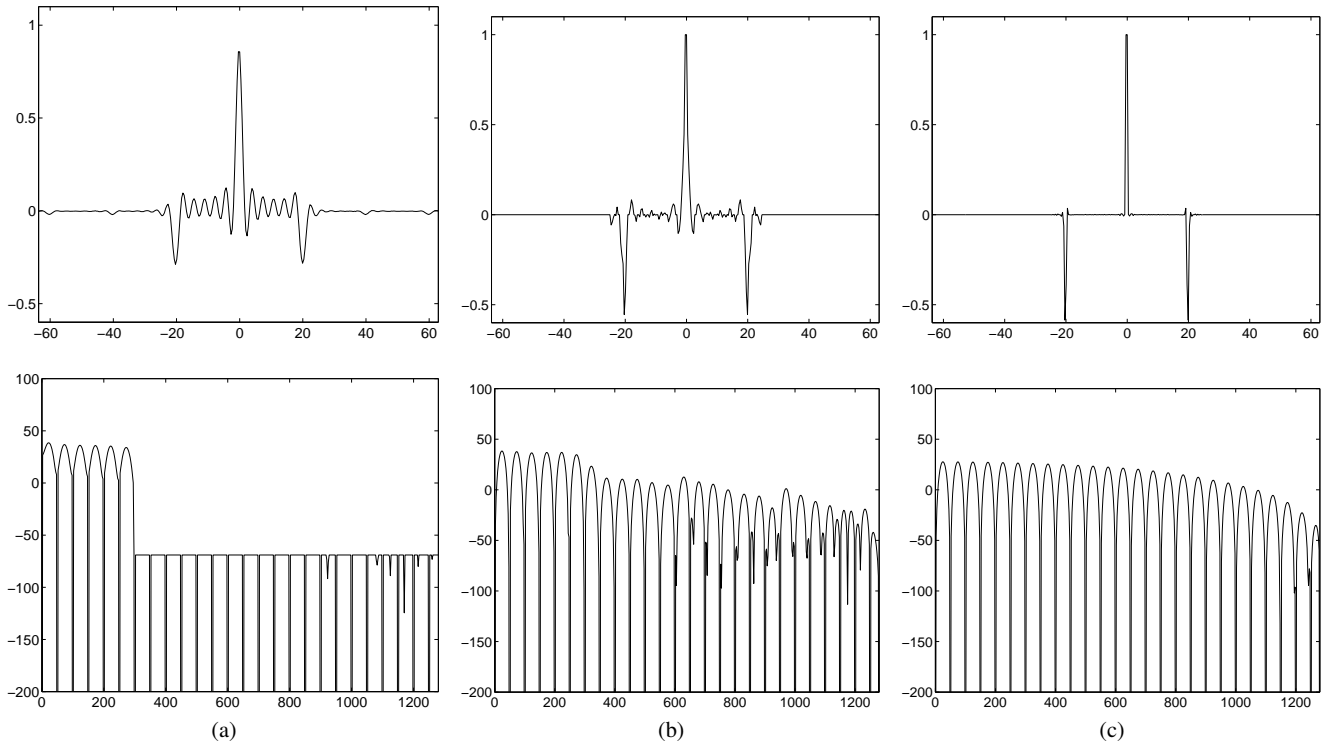
Fig. 3 State-of-the-art results: pulse x (top) and its Fourier transform χ (bottom). (a) Minimizer of $\|x\|_2^2 + d_{C_3}(x) + d_{C_4}(x)$ subject to $x \in \left(\bigcap_{k \in \mathbb{D}_1} C_1^{(k)} \right) \cap C_2$. (b) Minimizer of $\|x\|_2^2 + \sum_{k \in \mathbb{D}_1} d_{C_1^k}^2(x)$ subject to $x \in C_2 \cap C_3 \cap C_4$. (c) Minimizer of $\|x\|_2^2$ subject to $x \in C_2 \cap C_3 \cap C_4$.
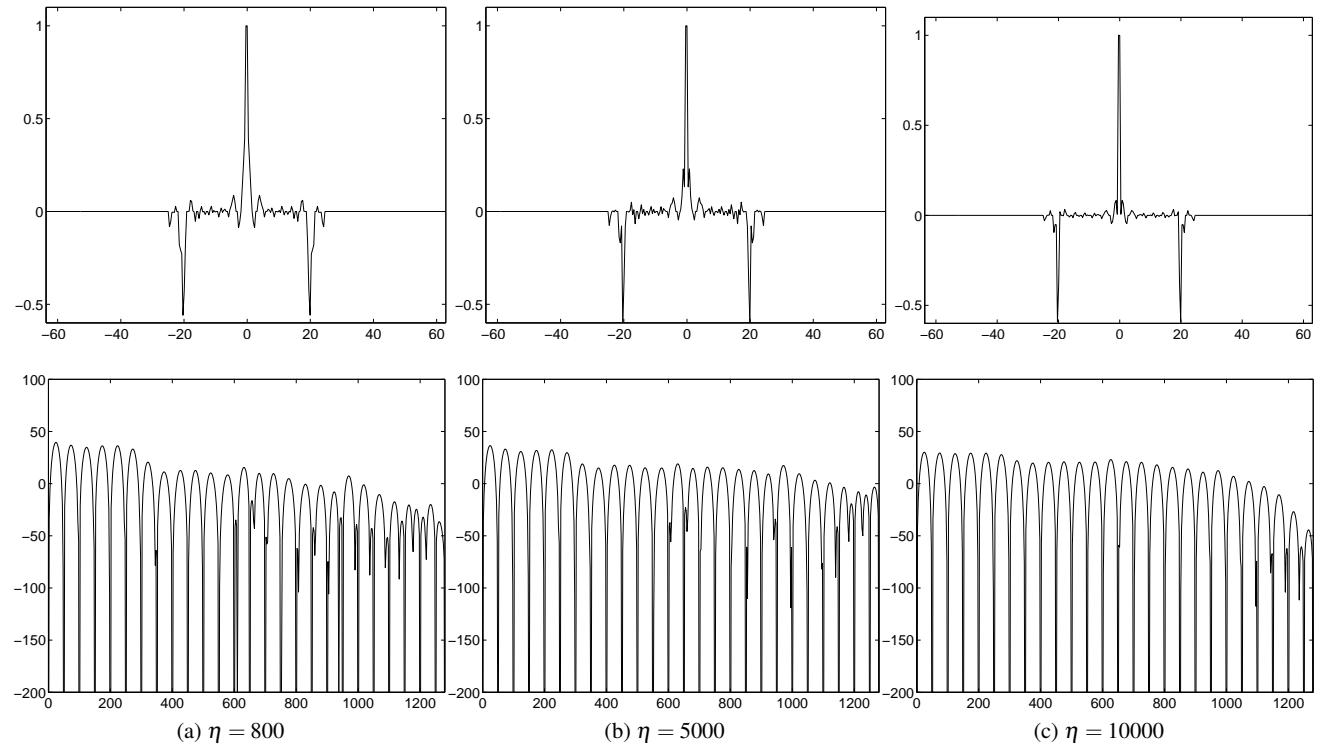
Fig. 4 Results obtained from (43) with $q = 2$: pulse x (top) and its Fourier transform χ (bottom).

7 Appendix

7.1 Proof of Proposition 1

For every $(y, \zeta) \in \mathscr{H} \times \mathbb{R}$, let $(p, \theta) = P_{\text{epi}\,\varphi}(y, \zeta)$. If $\varphi(y) \leq \zeta$, then $p = y$ and $\theta = \zeta = \max\{\varphi(p), \zeta\}$. In addition,

$$(\forall u \in \mathscr{H}) \quad 0 = \frac{1}{2}\|p - y\|^2 + \frac{1}{2}\big(\max\{\varphi(p) - \zeta, 0\}\big)^2$$
$$\leq \frac{1}{2}\|u - y\|^2 + \frac{1}{2}\big(\max\{\varphi(u) - \zeta, 0\}\big)^2, \quad (46)$$

which shows that (18) holds. Let us now consider the case when $\varphi(y) > \zeta$. From the definition of the projection, we get

$$(p, \theta) = \underset{(u, \xi) \in \text{epi}\,\varphi}{\text{argmin}} \|u - y\|^2 + (\xi - \zeta)^2. \quad (47)$$

From the Karush-Kuhn-Tucker theorem [68, Theorem 5.2],[6] there exists $\alpha \in [0, +\infty[$ such that

$$(p, \theta) = \underset{(u, \xi) \in \mathscr{H} \times \mathbb{R}}{\text{argmin}} \frac{1}{2}\|u - y\|^2 + \frac{1}{2}(\xi - \zeta)^2 + \alpha(\varphi(u) - \xi) \quad (48)$$

where the Lagrange multiplier α is such that $\alpha(\varphi(p) - \theta) = 0$. Since the value $\alpha = 0$ is not allowable (otherwise it would lead to $p = y$ and $\theta = \zeta$), it can be deduced from the above equality that $\varphi(p) = \theta$. In addition, differentiating the Lagrange functional in (48) w.r.t. ξ yields $\varphi(p) = \theta = \zeta + \alpha \geq \zeta$. Hence, (p, θ) in (47) is such that $\theta = \varphi(p) = \max\{\varphi(p), \zeta\}$ and

$$p = \underset{\substack{u \in \mathscr{H} \\ \varphi(u) \geq \zeta}}{\text{argmin}} \|u - y\|^2 + (\varphi(u) - \zeta)^2. \quad (49)$$

Furthermore, as $\varphi(y) > \zeta$, we obtain

$$\inf_{\substack{u \in \mathscr{H} \\ \varphi(u) \leq \zeta}} \|u - y\|^2 = \|P_{\text{lev}_{\leq \zeta}\,\varphi}(y) - y\|^2 = \inf_{\substack{u \in \mathscr{H} \\ \varphi(u) = \zeta}} \|u - y\|^2, \quad (50)$$

where we have used the fact that $P_{\text{lev}_{\leq \zeta}\,\varphi}(y)$ belongs to the boundary of $\text{lev}_{\leq \zeta}\,\varphi$ which is equal to $\{u \in \mathscr{H} \mid \varphi(u) = \zeta\}$ since φ is lower-semicontinuous and $\text{dom}\,\varphi$ is open [40, Corollary 8.38]. We have then

$$\inf_{\substack{u \in \mathscr{H} \\ \varphi(u) \leq \zeta}} \|u - y\|^2 = \inf_{\substack{u \in \mathscr{H} \\ \varphi(u) = \zeta}} \|u - y\|^2 \geq \inf_{\substack{u \in \mathscr{H} \\ \varphi(u) \geq \zeta}} \|u - y\|^2 + (\varphi(u) - \zeta)^2. \quad (51)$$

Altogether, (49) and (51) lead to

$$p = \underset{u \in \mathscr{H}}{\text{argmin}} \frac{1}{2}\|u - y\|^2 + \frac{1}{2}\big(\varphi(u) - \zeta\big)^2 \quad (52)$$

which is equivalent to (18) since $\frac{1}{2}(\max\{\varphi - \zeta, 0\})^2 \in \Gamma_0(\mathscr{H})$.

---

[6] By considering $u_0 \in \text{dom}\,\varphi$ and $\xi_0 > \varphi(u_0)$, the required qualification condition is obviously satisfied.

7.2 Proof of Proposition 3

Since $(\max\{\varphi - \zeta, 0\})^2$ is an even function, $\text{prox}_{\frac{1}{2}(\max\{\varphi - \zeta, 0\})^2}$ is an odd function [24, Remark 4.1(ii)]. In the following, we thus focus on the case when $y \in]0, +\infty[$. If $\zeta \in]-\infty, 0]$, then $(\max\{\varphi - \zeta, 0\})^2 = (\varphi - \zeta)^2$ is differentiable and, according to the fact that $p = \text{prox}_f(y)$ is uniquely defined as $y - p \in \partial f(p)$, we deduce that $p = \text{prox}_{\frac{1}{2}(\varphi - \zeta)^2}(y)$ is such that

$$p - y + q\tau p^{q-1}(\tau p^q - \zeta) = 0, \quad (53)$$

where $p \geq 0$ by virtue of [69, Corollary 2.5]. This allows us to deduce that $p = \chi_0$.

Let us now focus on the case when $\zeta \in]0, +\infty[$. If $y \in]0, (\zeta/\tau)^{1/q}[$, it can be deduced from [69, Corollary 2.5], that $p = \text{prox}_{\frac{1}{2}(\max\{\varphi - \zeta, 0\})^2}(y) \in [0, (\zeta/\tau)^{1/q}[$. Since $(\forall v \in [0, (\zeta/\tau)^{1/q}[) \max\{\varphi(v) - \zeta, 0\} = 0$, we have $p = y$. On the other hand if $y > (\zeta/\tau)^{1/q}$, as the proximity operator of a function from $\mathbb{R}$ to $\mathbb{R}$ is continuous and increasing [69, Proposition 2.4], we obtain

$$p = \text{prox}_{\frac{1}{2}(\max\{\varphi - \zeta, 0\})^2}(y) \geq$$
$$\text{prox}_{\frac{1}{2}(\max\{\varphi - \zeta, 0\})^2}\big((\zeta/\tau)^{1/q}\big) = (\zeta/\tau)^{1/q}. \quad (54)$$

Since $(\max\{\varphi - \zeta, 0\})^2$ is differentiable and, for each $v \geq (\zeta/\tau)^{1/q}$, $(\max\{\varphi(v) - \zeta, 0\})^2 = (\tau v^q - \zeta)^2$, we deduce that p is the unique value in $[(\zeta/\tau)^{1/q}, +\infty[$ satisfying (53).

7.3 Proof of Proposition 4

Let us notice that $\frac{1}{2}(\max\{\tau d_C^q - \zeta, 0\})^2 = \psi \circ d_C$ where $\psi = \frac{1}{2}(\max\{\tau|\cdot|^q - \zeta, 0\})^2$. According to [49, Proposition 2.7], for every $y \in \mathscr{H}$,

$$\text{prox}_{\psi \circ d_C}(y) = \begin{cases} y, & \text{if } y \in C, \\ P_C(y), & \text{if } d_C(y) \leq \max \partial \psi(0), \\ \alpha y + (1 - \alpha)P_C(y), & \text{if } d_C(y) > \max \partial \psi(0) \end{cases} \quad (55)$$

where $\alpha = \frac{\text{prox}_\psi(d_C(y))}{d_C(y)}$. In addition, we have

$$\partial \psi(0) = \begin{cases} [\tau\zeta, -\tau\zeta], & \text{if } \zeta < 0 \text{ and } q = 1, \\ \{0\}, & \text{otherwise}, \end{cases} \quad (56)$$

and, according to Proposition 2, when $\zeta < 0$ and $d_C(y) \leq -\tau\zeta$, $\text{prox}_\psi(d_C(y)) = 0$. This shows that (55) reduces to (25).

7.4 Proof of Proposition 5

For every $(y, \zeta) \in \mathbb{R}^M \times \mathbb{R}$, in order to determine $P_{\text{epi}\,\varphi}(y, \zeta)$, we have to find

$$\min_{\theta \in [0,+\infty[} \left((\theta - \zeta)^2 + \min_{\substack{|p^{(1)}| \leq \theta/\tau^{(1)} \\ \cdots \\ |p^{(M)}| \leq \theta/\tau^{(M)}}} \|p - y\|^2 \right). \quad (57)$$

For all $\theta \in [0, +\infty[$, the inner minimization is achieved when, $\forall m \in \{1, \ldots, M\}$, $p^{(m)}$ is the projection of $y^{(m)}$ onto the range $[-\theta/\tau^{(m)}, \theta/\tau^{(m)}]$, as given by (31). Then, (57) reduces to

$$\min_{\theta \in [0,+\infty[} \left((\theta - \zeta)^2 + \sum_{m=1}^{M} (\max\{|y^{(m)}| - \theta/\tau^{(m)}, 0\})^2 \right) \quad (58)$$

which is equivalent to calculate $\theta = \text{prox}_{\phi + \iota_{[0,+\infty[}}(\zeta)$ where

$$\phi(v) = \frac{1}{2} \sum_{m=1}^{M} (\max\{(\tau^{(m)})^{-1}(v^{(m)} - v), 0\})^2. \quad (59)$$

The function ϕ belongs to $\Gamma_0(\mathbb{R})$ since, for each $m \in \{1, \ldots, M\}$, $\max\{(\tau^{(m)})^{-1}(v^{(m)} - \cdot), 0\}$ is finite convex and $(\cdot)^2$ is finite convex and increasing on $[0, +\infty[$. In addition, ϕ is differentiable and such that, for every $v \in \mathbb{R}$ and $k \in \{1, \ldots, M+1\}$,

$$v^{(k-1)} < v \leq v^{(k)} \Rightarrow \phi(v) = \frac{1}{2} \sum_{m=k}^{M} (\tau^{(m)})^{-2} (v - v^{(m)})^2. \quad (60)$$

By using [25, Prop. 12], $\theta = P_{[0,+\infty[}(\chi)$ with $\chi = \text{prox}_\phi(\zeta)$. Therefore, there exists an $\overline{m} \in \{1, \ldots, M+1\}$ such that $v^{(\overline{m}-1)} < \chi \leq v^{(\overline{m})}$ and $\zeta - \chi = \phi'(\chi)$, so leading to

$$\zeta - \chi = \sum_{m=\overline{m}}^{M} (\tau^{(m)})^{-2}(\chi - v^{(m)}) \quad \Leftrightarrow \quad (32) \quad (61)$$

and $v^{(\overline{m}-1)} < \chi \leq v^{(\overline{m})} \Leftrightarrow (33)$. The uniqueness of $\overline{m} \in \{1, \ldots, M+1\}$ follows from that of $\text{prox}_\phi(\zeta)$.

References

1. P. L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212. Springer, 2011.
2. N. Parikh and S. Boyd. Proximal algorithms. *Found. Trends Optim.*, 1(3):123–231, 2013.
3. M. Guerquin-Kern, M. Häberlin, K.P. Pruessmann, and M. Unser. A fast wavelet-based reconstruction method for magnetic resonance imaging. *IEEE Trans. Med. Imag.*, 30(9):1649–1660, Sep. 2011.
4. F.-X. Dupé, M. J. Fadili, and J.-L. Starck. A proximal iteration for deconvolving Poisson noisy images using sparse representations. *IEEE Trans. Image Process.*, 18(2):310–321, Feb. 2009.
5. J.-F. Aujol, G. Gilboa, T. Chan, and S. Osher. Structure-texture image decomposition - modeling, algorithms, and parameter selection. *Int. J. Comp. Vis.*, 67(1):111–136, Apr. 2006.
6. L. M. Briceño-Arias, P. L. Combettes, J.-C. Pesquet, and N. Pustelnik. Proximal algorithms for multicomponent image recovery problems. *J. Math. Imag. Vis.*, 41(1):3–22, Sep. 2011.
7. S. Theodoridis, K. Slavakis, and I. Yamada. Adaptive learning in a world of projections. *IEEE Signal Process. Mag.*, 28(1):97–123, Jan. 2011.
8. C. Chaux, M. El Gheche, J. Farah, J.-C. Pesquet, and B. Pesquet-Popescu. A parallel proximal splitting method for disparity estimation from multicomponent images under illumination variation. *J. Math. Imag. Vis.*, 47(3):167–178, November 2013.
9. S. Mallat. *A wavelet tour of signal processing*. Academic Press, San Diego, USA, 1997.
10. L. Jacques, L. Duval, C. Chaux, and G. Peyré. A panorama on multiscale geometric representations, intertwining spatial, directional and frequency selectivity. *Signal Process.*, 91(12):2699–2730, Dec. 2011.
11. E. Van Den Berg and M. P. Friedlander. Probing the Pareto frontier for basis pursuit solutions. *SIAM J. Sci. Comput.*, 31(2):890–912, Nov. 2008.
12. R. Ciak, B. Shafei, and G. Steidl. Homogeneous penalizers and constraints in convex image restoration. *J. Math. Imag. Vis.*, 2012.
13. D. C. Youla and H. Webb. Image restoration by the method of convex projections. Part I - theory. *IEEE Trans. Med. Imag.*, 1(2):81–94, Oct. 1982.
14. H. J. Trussell and M. R. Civanlar. The feasible solution in signal restoration. *IEEE Trans. Acous., Speech Signal Process.*, 32(2):201–212, Apr. 1984.
15. P. L. Combettes. Inconsistent signal feasibility problems : least-squares solutions in a product space. *IEEE Trans. Signal Process.*, 42(11):2955–2966, Nov. 1994.
16. T. Teuber, G. Steidl, and R. H. Chan. Minimization and parameter estimation for seminorm regularization models with I-divergence constraints. *Inverse Problems*, 29(3):035007, 2013.
17. R. Stück, M. Burger, and T. Hohage. The iteratively regularized Gauss-Newton method with convex constraints and applications in 4Pi microscopy. *Inverse Problems*, 28(1), Jan. 2012.
18. S. Ono and I. Yamada. Poisson image restoration with likelihood constraint via hybrid steepest descent method. In *Proc. Int. Conf. Acoust., Speech Signal Process.*, pages 5929 – 5933, Vancouver, Canada, May 2013.
19. S. Harizanov, J.-C. Pesquet, and G. Steidl. Epigraphical projection for solving least squares anscombe transformed constrained optimization problems. In *Scale Space and Variational Methods in Computer Vision*, volume 7893, pages 125–136. 2013.
20. S. Ono and I. Yamada. Second-order total generalized variation constraint. In *Proc. Int. Conf. Acoust., Speech Signal Process.*, Florence, Italy, May 4-9 2014.
21. M Tofighi, K. Kose, and A. E. Cetin. Signal reconstruction framework based on Projections onto Epigraph Set of a Convex cost function (PESC). 2014. preprint.
22. G. Chen and M. Teboulle. A proximal-based decomposition method for convex minimization problems. *Math. Program.*, 64:81–101, 1994.
23. I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Applied Math.*, 57(11):1413–1457, Nov. 2004.
24. C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs. A variational formulation for frame-based inverse problems. *Inverse Problems*, 23(4):1495–1518, Jun. 2007.
25. P. L. Combettes and J.-C. Pesquet. A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery. *IEEE J. Selected Topics Signal Process.*, 1(4):564–574, Dec. 2007.
26. M. Figueiredo, R. Nowak, and S. Wright. Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems. *IEEE J. Selected Topics Signal Process.: Special Issue on Convex Optimization Methods for Signal Processing*, 5981(4):586–598, Dec. 2007.

27. A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.*, 2(1):183–202, 2009.

28. M. Fornasier and C.-B. Schönlieb. Subspace correction methods for total variation and ℓ_1-minimization. *IMA J. Numer. Anal.*, 47(8):3397–3428, 2009.

29. G. Steidl and T. Teuber. Removing multiplicative noise by Douglas-Rachford splitting methods. *J. Math. Imag. Vis.*, 36(3):168–184, 2010.

30. J.-C. Pesquet and N. Pustelnik. A parallel inertial proximal optimization method. *Pac. J. Optim.*, 8(2):273–305, Apr. 2012.

31. E. Esser, X. Zhang, and T. Chan. A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. *SIAM J. Imaging Sci.*, 3(4):1015–1046, 2010.

32. A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imag. Vis.*, 40(1):120–145, 2011.

33. L. M. Briceño-Arias and P. L. Combettes. A monotone + skew splitting model for composite monotone inclusions in duality. *SIAM J. Opt.*, 21(4):1230–1250, 2011.

34. 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 Var. Anal.*, 20(2):307–330, June 2012.

35. B. C. Vũ. A splitting algorithm for dual monotone inclusions involving cocoercive operators. *Adv. Comput. Math.*, 38(3):667–681, Apr. 2013.

36. L. Condat. A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. *J. Optim. Theory Appl.*, 158(2):460–479, Aug. 2012.

37. P. Chen, J. Huang, and X. Zhang. A primal-dual fixed point algorithm for convex separable minimization with applications to image restoration. *Inverse Problems*, 29(2), 2013.

38. M. A. Alghamdi, A. Alotaibi, P. L. Combettes, and N. Shahzad. A primal-dual method of partial inverses for composite inclusions. *Optim. Letters*, Mar. 2014.

39. J. J. Moreau. Proximité et dualité dans un espace hilbertien. *B. Soc. Math. Fr.*, 93:273–299, 1965.

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

41. P. L. Combettes. A block-iterative surrogate constraint splitting method for quadratic signal recovery. *IEEE Trans. Signal Process.*, 51(7):1771–1782, Jul. 2003.

42. Y. Kopsinis, K. Slavakis, and S. Theodoridis. Online sparse system identification and signal reconstruction using projections onto weighted ℓ_1 balls. 59(3):936–952, Mar. 2011.

43. A. Quattoni, X. Carreras, M. Collins, and T. Darrell. An efficient projection for $\ell_{1,\infty}$ regularization. In *International Conference on Machine Learning*, Montreal, Quebec, Jun., 14-18 2009.

44. L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60(1-4):259–268, Nov. 1992.

45. G. Gilboa and S. Osher. Nonlocal Operators with Applications to Image Processing. *Multiscale Model. and Simul.*, 7(3):1005, 2009.

46. R. Gaetano, G. Chierchia, and B. Pesquet-Popescu. Parallel implementations of a disparity estimation algorithm based on a proximal splitting method. In *Visual Communication and Image Processing*, San Diego, USA, 2012.

47. E. J. Candès, M. B. Wakin, and S. Boyd. Enhancing sparsity by reweighted ℓ_1 minimization. *J. Fourier Anal. Appl.*, 14(5):877–905, Dec. 2008.

48. A. Tikhonov. Regularization of incorrectly posed problems. *Sov. Math. Dokl.*, 4:1624–1627, 1963.

49. P. L. Combettes and J.-C. Pesquet. A proximal decomposition method for solving convex variational inverse problems. *Inverse Problems*, 24(6), Dec. 2008.

50. J. Wu, F. Liu, L. C. Jiao, X. Wang, and B. Hou. Multivariate compressive sensing for image reconstruction in the wavelet domain. *IEEE Trans. Image Process.*, 20(12):3483–3494, Dec. 2011.

51. B. A. Turlach, W. N. Venables, and S. J. Wright. Simultaneous variable selection. *Technometrics*, 47(3):349–363, Aug. 2005.

52. Y. Chen and A. O. Hero. Recursive $\ell_{1,\infty}$ lasso. *IEEE Trans. Signal Process.*, 60(8):3978–3987, Aug. 2012.

53. C. Studer, T. Goldstein, W. Yin, and R. G. Baraniuk. Democratic representations. January 2014. http://arxiv.org/abs/1401.3420.

54. M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo. An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems. *IEEE Trans. Image Process.*, 20(3):681–695, Mar. 2011.

55. S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Machine Learn.*, 8(1):1–122, 2011.

56. C. Ding, D. F. Sun, and K.-C. Toh. An introduction to a class of matrix cone programming. *Math. Programm.*, pages 1–39, Dec. 2012.

57. D. L. Donoho. De-noising by soft-thresholding. *IEEE Trans. Inform. Theory*, 41(3):613–627, May 1995.

58. F. Bach, R. Jenatton, J. Mairal, and G. Obozinski. Optimization with sparsity-inducing penalties. *Foundations and Trends in Machine Learning*, 4(1):1–106, 2012.

59. K. Bredies, K. Kunisch, and T. Pock. Total generalized variation. 3(3):492–526, Sep. 2010.

60. A. Buades, B. Coll, and J. Morel. A review of image denoising algorithms, with a new one. *Multiscale Model. and Simul.*, 4(2):490–530, 2005.

61. G. Gilboa and S. Osher. Nonlocal linear image regularization and supervised segmentation. *Multiscale Model. and Simul.*, 6(2):595–630, 2007.

62. A. Foi and G. Boracchi. Foveated self-similarity in nonlocal image filtering. In *Proc. SPIE Electronic Imaging*, volume 8291, Burlingame, USA, Jan. 2012.

63. Y. Li and S. Osher. A new median formula with applications to pde based denoising. *Commun. Math. Sci.*, 7(3):741–753, Sep. 2009.

64. X. Bresson and T. F. Chan. Fast dual minimization of the vectorial total variation norm and applications to color image processing. *Inverse Problems and Imaging*, 2(4):455–484, Nov. 2008.

65. T. Miyata and Y. Sakai. Vectorized total variation defined by weighted L-infinity norm for utilizing inter channel dependency. In *Proc. Int. Conf. Image Process.*, pages 3057–3060, Sep. 2012.

66. Z. Wang and A. C. Bovik. Mean squared error: love it or leave it? *IEEE Signal Process. Mag.*, 26(1):98–117, Jan. 2009.

67. M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *J. R. Statist. Soc. B*, 68:49–67, 2006.

68. I. Ekeland and R. Témam. *Convex analysis and variational problems*. SIAM, Philadelphia, 1999.

69. P. L. Combettes and J.-C Pesquet. Proximal thresholding algorithm for minimization over orthonormal bases. *SIAM J. Opt.*, 18(4):1351–1376, Nov. 2007.