Epigraphical splitting for solving constrained convex optimization problems with proximal tools
Giovanni Chierchia, Nelly Pustelnik, Jean-Christophe Pesquet, Béatrice Pesquet-Popescu

To cite this version:
Giovanni Chierchia, Nelly Pustelnik, Jean-Christophe Pesquet, Béatrice Pesquet-Popescu. Epigraphical splitting for solving constrained convex optimization problems with proximal tools. Signal, Image and Video Processing, Springer Verlag, 2015, 9 (8), pp.1737-1749. 10.1007/s11760-014-0664-1. hal-01372568

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

Submitted on 27 Sep 2016

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

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Epigraphical splitting for solving constrained convex optimization problems with proximal tools – extended version

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

Received: 03 June 2014 / Accepted: 17 June 2014

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 $\ell^q$-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

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:

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

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 $\lambda$ is a positive regularization parameter. A constrained formulation of the above problem may be also considered [11], yielding

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

where $\eta$ is a positive constraint bound. Both optimization problems are equivalent for some specific values of $\lambda$ and $\eta$, 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 $\eta$ may be easier to set than the parameter $\lambda$, 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 non-smooth 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 $\varphi$ from a real Hilbert space $\mathcal{H}$ to $]-\infty, +\infty]$, defined as

$$(\forall y \in \mathcal{H}) \quad \text{prox}_\varphi(y) = \arg\min_{u \in \mathcal{H}} \frac{1}{2} \|u - y\|^2 + \varphi(u).$$

The proximity operator can be interpreted as a sort of subgradient step for the function $\varphi$, as $p = \text{prox}_\varphi(y)$ 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 $C \subset \mathcal{H}$, in the sense that

$$(\forall y \in \mathcal{H}) \quad \text{prox}_{1_C}(y) = P_C(y) = \arg\min_{u \in C} \frac{1}{2} \|u - y\|^2,$$

where $1_C$ 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. $\text{prox}_{1 - \|\cdot\|_1^2}$ and $P_{\|\cdot\| \leq \eta}$ 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 a hypercube, a closed half-space, a hyperslab, an affine set, a simplex and an $\ell_2$-norm ball [2, 7, 40]. However, more sophisticated convex sets are usually necessary in a large number of inverse problems (e.g. the $\ell_1$-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 $\ell_1$-ball [11, 42] and an $\ell_1$-norm 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**

minimize $\sum_{r=1}^{R} g_r(T_r x)$ s.t. $h(F x) \leq \eta,$

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

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

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

with $y^{(\ell)} \in \mathbb{R}^{M_{\ell}}$ 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], $\ell_2$-norm composed with an Ancombe transformation [19], $\ell_1$-norm [11], $\ell_1$-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 $\ell_1$-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 $\ell_p$-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 $\mathcal{H}$ be a real Hilbert space endowed with the norm $\| \cdot \|$ and the scalar product $\langle \cdot, \cdot \rangle$. $\mathcal{I}_0(\mathcal{H})$ denotes the set of proper lower-semicontinuous convex functions from $\mathcal{H}$ to $]-\infty, +\infty]$. Recall that a function $\varphi : \mathcal{H} \to ]-\infty, +\infty]$ is proper if $\text{dom} \varphi = \{ y \in \mathcal{H} \mid \varphi(y) < +\infty \}$ is nonempty. The lower level set of $\varphi$ at height $\zeta \in \mathbb{R}$ is the closed convex subset of $\mathcal{H}$ defined as $\text{lev}_\zeta \varphi = \{ y \in \mathcal{H} \mid \varphi(y) \leq \zeta \}$. The epigraph of $\varphi$ in $\mathcal{I}_0(\mathcal{H})$ is the closed convex subset of $\mathcal{H} \times \mathbb{R}$ defined as $\text{epi} \varphi = \{ (y, \zeta) \in \mathcal{H} \times \mathbb{R} \mid \varphi(y) \leq \zeta \}$. A subgradient of $\varphi$ at $y \in \mathcal{H}$ is the subset of $\mathcal{H}$ defined as $\partial \varphi(y) = \{ t \in \mathcal{H} \mid \langle u, \varphi(y) \rangle \geq \varphi(y) + \langle t, u - y \rangle \}$. When $\varphi$ is Gâteaux-differentiable at $y$, $\partial \varphi(y) = \{ \varphi(y) \}$. Let $C$ be a nonempty convex closed subset of $\mathcal{H}$. The relative interior of $C$ is denoted by $\text{ri}C$. For every $y \in \mathcal{H}$, the proximity operator of $\varphi$ reads $\text{prox}_{\varphi}(y) = \text{argmin}_{u \in \mathcal{H}} \frac{1}{2} \| u - y \|^2 + \varphi(u)$. The indicator function $\text{1}_C \in \mathcal{I}_0(\mathcal{H})$ of $C$ is equal to 0 for $y \in C$ and $+\infty$ otherwise. The projection onto $C$ reads $P_C(y) = \text{argmin}_{u \in \text{ri}C} \| u - y \|$. 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:

$$\text{argmin}_{y \in \mathbb{R}^M} h(y) = \sum_{l=1}^{L} h_l(y^{(l)}),$$

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

$$y = \begin{bmatrix} (y^{(1)})^\top, \ldots, (y^{(L)})^\top \end{bmatrix}^\top \in \mathbb{R}^M,$$

with $M_1 + \cdots + M_L = M$ and, for every $l \in \{ 1, \ldots, L \}$, $y^{(l)} \in \mathbb{R}^{M_l}$ and $h_l \in \mathcal{I}_0(\mathbb{R}^{M_l})$ is such that $\text{ri}(\text{dom} h_l) \neq \emptyset$.

2.1 Epigraphical splitting

Our approach consists of introducing an auxiliary vector $\zeta = (\zeta^{(l)})_{1 \leq l \leq L} \in \mathbb{R}^L$ 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

$$\sum_{l=1}^{L} h_l(y^{(l)}) \leq \eta \quad \Leftrightarrow \quad \left\{ \begin{array}{l} \sum_{l=1}^{L} \zeta^{(l)} \leq \eta, \\ \langle \forall \ell \in \{ 1, \ldots, L \} \rangle \quad h_l(y^{(\ell)}) \leq \zeta^{(\ell)}, \end{array} \right.$$  

(3)

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

$$\text{minimize}_{(x, \zeta) \in \mathcal{H} \times \mathbb{R}^L} \sum_{r=1}^{R} g_r(T_r x) \quad \text{s.t.} \quad \left\{ \begin{array}{l} \langle \forall \ell \in \{ 1, \ldots, L \} \rangle \quad h_l\left(F x^{(\ell)}\right) \leq \zeta^{(\ell)}, \\ \sum_{l=1}^{L} \zeta^{(l)} \leq \eta. \end{array} \right.$$  

(4)

Note that the above minimization problem is defined with respect to $x$ and $\zeta$, 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

$$\text{minimize}_{(x, \zeta) \in \mathcal{H} \times \mathbb{R}^M} \sum_{r=1}^{R} g_r(T_r x) \quad \text{s.t.} \quad \left\{ \begin{array}{l} (F x, \zeta) \in E, \\ \zeta \in V, \end{array} \right.$$  

where

$$E = \{ (y, \zeta) \in \mathbb{R}^M \times \mathbb{R}^L \mid \langle \forall \ell \in \{ 1, \ldots, L \} \rangle \quad y^{(\ell)}, \zeta^{(\ell)} \in \text{epi} h_l \},$$

$$V = \{ \zeta \in \mathbb{R}^L \mid 1_L^\top \zeta \leq \eta \},$$  

(5)

(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 $\zeta$ can be also replaced by an equality, even though it makes little difference in our approach.
We now illustrate some examples of functions that can be decomposed as in (2), and the projection onto \( E \) is given by
\[
(\forall y \in \mathbb{R}^M) \quad h(y) = \sum_{\ell=1}^{L} \tau^{(\ell)}_y \|y^{(\ell)}\|_2.
\] (13)

Then, the above function can be modelled as in (1) with, for every \( \ell \in \{1, \ldots, L\}, \)
\[
(\forall y \in \mathbb{R}^M) \quad h(y) = \max_{1 \leq m \leq M_{\ell}} \tau^{(\ell, m)}_y |y^{(\ell, m)}|.
\] (15)

The corresponding epigraphical projection will be given in Proposition 5. When \( \tau^{(\ell, m)} = 1 \), 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
\[
\min_{x \in \mathcal{H}} \quad g_1(T_1x) + g_2(x)
\]
by resorting to the following reformulation
\[
\min_{(x,v) \in \mathcal{H} \times \mathbb{R}^q} \quad g_1(y) + g_2(x) \quad \text{s.t.} \quad T_1x = v.
\]

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 \( T_v = v \) is handled by computing the projection onto the nullspace of \( [T_1 - \text{Id}] \), 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 \( \mathcal{H} \) be a real Hilbert space and let \( \mathcal{H} \times \mathbb{R} \) be equipped with the standard product space norm. Let \( \varphi \) be a function in \( \Gamma_0(\mathcal{H}) \) such that \( \text{dom} \varphi \) is open. For every \( (y, \zeta) \in \mathcal{H} \times \mathbb{R} \), the projector \( P_{\text{epi} \varphi} \) onto \( \text{epi} \varphi \) is given by:

\[
P_{\text{epi} \varphi}(y, \zeta) = (p, \theta),
\]

where

\[
\begin{align*}
    p &= \text{prox}_{\frac{1}{2} \max \{ \varphi(y) - \zeta \}^2}(y), \\
    \theta &= \max \{ \varphi(p), \zeta \}.
\end{align*}
\]

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

**Proposition 2** Let \( \tau \in \mathbb{R} \). Assume that

\[
(\forall y \in \mathcal{H}) \quad \varphi(y) = \tau |y|.
\]

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

\[
p = \begin{cases} y, & \text{if } |\tau y| \leq \zeta, \\ \frac{\text{sign}(y)}{1 + \tau^2} \max \{ |y| + \tau \zeta, 0 \}, & \text{otherwise}. \end{cases}
\]

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.

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

\[
(\forall y \in \mathcal{H}) \quad \varphi(y) = \tau |y|^q. 
\]

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

\[
p = \begin{cases} \text{sign}(y)\zeta, & \text{if } \zeta \leq 0, \\ y, & \text{if } \zeta > 0 \text{ and } |\tau y|^q \leq \zeta, \\ \text{sign}(y)\zeta, & \text{otherwise}, \end{cases}
\]

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

\[
q \chi^{2q-1} - q \tau \chi q^{q-1} + \chi = |\tau y|^q. 
\]

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 \( \mathcal{H} \). Let \( q \in [1, +\infty[ \), \( \tau \in [0, +\infty[ \), and \( \zeta \in \mathbb{R} \). Assume that

\[
(\forall y \in \mathcal{H}) \quad \varphi(y) = \tau d_C^q(y). 
\]

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

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

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

\[
\alpha = \frac{\text{prox}_{\frac{1}{2} \max \{ |\tau y - \zeta|, 0 \}^2 \cdot d_C^q(y)}{d_C(y)}
\]

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

In the case when \( q = 1 \) and \( C = \{ z \} \) for some \( z \in \mathcal{H} \), 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 \( z \in \mathcal{H} \). Assume that

\[
(\forall y \in \mathcal{H}) \quad \varphi(y) = \tau \| y - z \|. 
\]

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

\[
p = \begin{cases} z, & \text{if } y = z, \\ \alpha y + (1 - \alpha)z, & \text{otherwise}, \end{cases}
\]

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

\[
\alpha = \frac{1}{1 + \tau^2} \max \{ 1 + \frac{\tau \zeta}{\| y - z \|}, 0 \}.
\]

\(^2\) 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.
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  
\[
(\forall y \in \mathbb{R}^M) \quad \varphi(y) = \max_{1 \leq m \leq M} \tau^{(m)} |y^{(m)}|,
\]
where the values \( (\nu^{(m)} = \tau^{(m)} |y^{(m)}|)_{1 \leq m \leq M} \) are in ascending order. Then, for every \( (y, \zeta) \in \mathbb{R}^M \times \mathbb{R} \), \( P_{\varphi, \theta}(y, \zeta) = (p, \theta) \) is given by  
\[
p = \left( \text{sign}(y^{(m)}) \min \left\{ \frac{y^{(m)} \theta / \tau^{(m)}}{1 + \sum_{m=1}^M \tau^{(m)^{-2}}} \right\}, \right)_{1 \leq m \leq M}
\]
and  
\[
\theta = \max \left\{ \frac{\zeta + \sum_{m=\mathbf{m}}^M \nu^{(m)}(\tau^{(m)^{-2}})}{1 + \sum_{m=\mathbf{m}}^M \tau^{(m)^{-2}}} \right\},
\]
where \( \mathbf{m} \) is the unique integer in \( \{1, \ldots, M + 1\} \) such that  
\[
\nu^{(\mathbf{m}+1)} / \nu^{(\mathbf{m})} \leq \left( \frac{1}{\tau^{(m)} - 2} \right)^{1/2} \times \nu^{(\mathbf{m})}
\]
(33)

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 \( \mathcal{H} = \mathbb{R}^N \), \( \mathbf{x} \in \mathbb{R}^N \) denotes the image to be recovered, and \( z \in \mathbb{R}^K \) is an observation vector such that  
\[
z = D A \mathbf{x} + b.
\]

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 \( \mathbf{x} \) from the degraded observations is to follow a variational approach that aims at solving a convex optimization problem defined as:

\[
\begin{align*}
\text{minimize} \quad & \|D A \mathbf{x} - z\|_2^2 \\
\text{s.t.} \quad & h(F \mathbf{x}) \leq \eta,
\end{align*}
\]

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

\[
(\forall y \in \mathbb{R}^M) \quad h(y) = \sum_{\ell=1}^L \|y^{(\ell)}\|_p^q,
\]

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

The above function can model the \( \ell_p \)-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

\[
F \mathbf{x} = \begin{bmatrix}
F_1 \mathbf{x} \\
\vdots \\
F_N \mathbf{x}
\end{bmatrix} = \begin{bmatrix}
(\omega_{1,n}(x^{(1)} - x^{(n)}))_{n \in A_{\ell}} \\
\vdots \\
(\omega_{N,n}(x^{(N)} - x^{(n)}))_{n \in A_{\ell}}
\end{bmatrix},
\]

where, for every \( \ell \in \{1, \ldots, N\} \), \( A_{\ell} \subset \{1, \ldots, N\} \setminus \{\ell\} \) denotes the neighborhood of \( \ell \) and \( (\omega_{l,n})_{n \in A_{\ell}} \) are positive...
weights. Note that (37) reduces to the classical gradient operator when \( \mathcal{N} \) only contains the horizontal/vertical nearest neighbours of \( \ell \) and \( \partial_{\ell,p} \equiv 1.\)

Several types of \( \ell_{q,p} \)-norms can be used to penalize the nonlocal gradient. The case \( q = p = 2 \) (i.e. the \( \ell_2 \)-norm) may be seen as a variational extension of NonLocal Means [60], while the case \( q = p = 1 \) (i.e. the \( \ell_1 \)-norm) leads to NonLocal Medians [63]. The case \( q = 1 \) and \( p \in \{2, +\infty\} \) (i.e. the \( \ell_{1,p} \)-norm) is instead preferred in color imagery [64, 65]. In our experiments, we will present the results of \( \ell_{q,p} \)-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 \( R = 2 \), \( T_1 = A \), \( T_2 = I \), \( g_1 = \|D \cdot z\|^2 \), \( g_2 = t_{0.253}^\| \) and \( h_1 = |z|^q \) (for \( q = 1 \) or \( q = 2 \)) and \( h_2 = \| \|_p \) (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 \( \ell_1 \)- or \( \ell_{1,2} \)-ball [11] and a \( \ell_{1,\infty} \)-ball [43].

Fig. 1 illustrates the restoration of a RGB color image \( (N = 320 \times 480 \times 3) \) degraded by a 3 \( \times \) 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 bound \( \eta \) for the several \( \ell_{q,p} \)-balls were hand-tuned in order to achieve the best SNR values. These results show the interest of considering non-local operators and \( \ell_{1,\infty} \)-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 \( \|\hat{x}^{[t]} - x^{[m]}\| \) 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 \( x^{[m]} \) denotes the solution obtained when such a criterion is reached. Note that \( x^{[m]} \) 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 \( \eta \) 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 \( \eta \). However, in order to reduce the computation burden of the direct projection onto the \( \ell_{1,\infty} \)-ball, we present the results obtained with the grayscale image \( \text{boat} \) cropped at \( 256 \times 256 \). The stopping criterion is set to \( \|x^{[i+1]} - x^{[i]}\| \leq 10^{-4}\|x^{[i]}\| \). For more readability, the values of \( \eta \) are expressed as a multiplicative factor of the \( \ell_{q,p} \)-non-local-TV semi-norm evaluated on the original image.

- Tables 1 and 2 report the comparison for \( \ell_{1,2} \)-TV and \( \ell_{1,\infty} \)-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 \( \ell_{1,p} \)-ball. Moreover, the numerical results show that errors within \( \pm 20\% \) from the optimal value for \( \eta \) lead to SNR variations within 2\%.

- Tables 3 and 4 collects the results of \( \ell_{1,2} \)-non-local-TV and \( \ell_{1,\infty} \)-non-local-TV for different sizes of the neighbourhood \( \mathcal{N} \). 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 \( F^T \) are implemented in C) and executed on an Intel Xeon CPU at 2.80 GHz and 8 GB of RAM.

---

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

5 Note that the \( \ell_{1,2} \)-norm performs better on grayscale images.
Fig. 1 Comparison of SNR (dB) – SSIM [66] indices for an example of restoration.

(a) Original image.
(b) Degraded image.
(c) $\ell_2$-non-local-TV: 18.12–0.761.
(d) $\ell_1$-$\ell_2$: 17.78–0.787.
(e) $\ell_1$-$\ell_2$: 18.36–0.821.
(f) $\ell_1$-$\ell_\infty$: 18.91–0.824.
(g) $\ell_1$-non-local-TV: 18.93–0.828.
(h) $\ell_1$-non-local-TV: 19.47–0.839.
(i) $\ell_1$-non-local-TV: 20.17–0.847.

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 $\eta$ (w.r.t. the grayscale image boat cropped at 256 $\times$ 256)

<table>
<thead>
<tr>
<th>$\eta$</th>
<th>SNR (dB) – SSIM</th>
<th>SDMM</th>
<th>M+LFBF</th>
</tr>
</thead>
<tbody>
<tr>
<td># iter.</td>
<td>sec.</td>
<td># iter.</td>
<td>sec.</td>
</tr>
<tr>
<td>0.45</td>
<td>19.90 – 0.733</td>
<td>107 6.07</td>
<td>174 2.03</td>
</tr>
<tr>
<td>0.50</td>
<td>20.18 – 0.745</td>
<td>117 6.95</td>
<td>159 1.95</td>
</tr>
<tr>
<td>0.56</td>
<td>20.23 – 0.745</td>
<td>129 8.36</td>
<td>153 1.90</td>
</tr>
<tr>
<td>0.62</td>
<td>20.16 – 0.737</td>
<td>141 9.44</td>
<td>155 1.83</td>
</tr>
<tr>
<td>0.67</td>
<td>20.00 – 0.724</td>
<td>154 10.20</td>
<td>162 2.17</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>$\eta$</th>
<th>SNR (dB) – SSIM</th>
<th>SDMM</th>
<th>M+LFBF</th>
</tr>
</thead>
<tbody>
<tr>
<td># iter.</td>
<td>sec.</td>
<td># iter.</td>
<td>sec.</td>
</tr>
<tr>
<td>0.45</td>
<td>19.52 – 0.726</td>
<td>160 312.55</td>
<td>231 3.89</td>
</tr>
<tr>
<td>0.50</td>
<td>19.71 – 0.728</td>
<td>168 342.01</td>
<td>215 3.75</td>
</tr>
<tr>
<td>0.56</td>
<td>19.71 – 0.734</td>
<td>180 373.60</td>
<td>211 3.49</td>
</tr>
<tr>
<td>0.62</td>
<td>19.59 – 0.715</td>
<td>196 412.68</td>
<td>216 3.67</td>
</tr>
<tr>
<td>0.67</td>
<td>19.39 – 0.698</td>
<td>211 448.77</td>
<td>223 3.76</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>$\eta$</th>
<th>SNR (dB) – SSIM</th>
<th>SDMM</th>
<th>M+LFBF</th>
</tr>
</thead>
<tbody>
<tr>
<td># iter.</td>
<td>sec.</td>
<td># iter.</td>
<td>sec.</td>
</tr>
<tr>
<td>Neighbourhood size: $3 \times 3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.43</td>
<td>20.82 – 0.757</td>
<td>208 20.67</td>
<td>211 10.93</td>
</tr>
<tr>
<td>0.49</td>
<td>20.97 – 0.765</td>
<td>167 16.84</td>
<td>177 9.01</td>
</tr>
<tr>
<td>0.54</td>
<td>21.02 – 0.767</td>
<td>147 15.31</td>
<td>157 7.93</td>
</tr>
<tr>
<td>0.59</td>
<td>20.98 – 0.764</td>
<td>134 14.44</td>
<td>148 7.67</td>
</tr>
<tr>
<td>0.65</td>
<td>20.88 – 0.757</td>
<td>133 14.82</td>
<td>136 7.11</td>
</tr>
<tr>
<td>Neighbourhood size: $5 \times 5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.43</td>
<td>21.00 – 0.766</td>
<td>301 56.03</td>
<td>333 45.18</td>
</tr>
<tr>
<td>0.49</td>
<td>21.15 – 0.773</td>
<td>260 49.03</td>
<td>302 39.64</td>
</tr>
<tr>
<td>0.54</td>
<td>21.20 – 0.775</td>
<td>242 46.31</td>
<td>283 37.72</td>
</tr>
<tr>
<td>0.59</td>
<td>21.17 – 0.773</td>
<td>231 44.20</td>
<td>268 36.56</td>
</tr>
<tr>
<td>0.65</td>
<td>21.08 – 0.767</td>
<td>220 44.64</td>
<td>252 34.46</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>$\eta$</th>
<th>SNR (dB) – SSIM</th>
<th>SDMM</th>
<th>M+LFBF</th>
</tr>
</thead>
<tbody>
<tr>
<td># iter.</td>
<td>sec.</td>
<td># iter.</td>
<td>sec.</td>
</tr>
<tr>
<td>Neighbourhood size: $3 \times 3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.43</td>
<td>20.78 – 0.762</td>
<td>434 1470.46</td>
<td>449 25.03</td>
</tr>
<tr>
<td>0.49</td>
<td>20.86 – 0.764</td>
<td>395 1319.64</td>
<td>413 22.86</td>
</tr>
<tr>
<td>0.54</td>
<td>20.83 – 0.760</td>
<td>363 1191.61</td>
<td>382 21.46</td>
</tr>
<tr>
<td>0.59</td>
<td>20.73 – 0.752</td>
<td>340 1093.26</td>
<td>354 19.77</td>
</tr>
<tr>
<td>0.65</td>
<td>20.58 – 0.740</td>
<td>322 1007.55</td>
<td>336 18.64</td>
</tr>
<tr>
<td>Neighbourhood size: $5 \times 5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.43</td>
<td>20.91 – 0.769</td>
<td>384 2069.62</td>
<td>452 64.42</td>
</tr>
<tr>
<td>0.49</td>
<td>20.97 – 0.767</td>
<td>326 1700.34</td>
<td>412 58.66</td>
</tr>
<tr>
<td>0.54</td>
<td>20.98 – 0.771</td>
<td>290 1476.98</td>
<td>389 55.35</td>
</tr>
<tr>
<td>0.59</td>
<td>20.88 – 0.759</td>
<td>276 1336.16</td>
<td>374 52.64</td>
</tr>
<tr>
<td>0.65</td>
<td>20.75 – 0.749</td>
<td>268 1220.14</td>
<td>362 51.45</td>
</tr>
</tbody>
</table>
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 $x = \{x^{(i)}\}_{0 \leq i \leq N-1} \in \mathbb{R}^N$ the pulse, by $\mathcal{X} = \{x^{(i)}\}_{0 \leq i \leq N-1}$ 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

$$\left(\forall \ell \in \mathbb{D}_1\right) \quad C_1^{(i)} = \{x \in \mathbb{R}^N \mid |x^{(i)}| \leq \gamma\},$$

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

(ii) Vanishing frequencies of $\mathcal{X}$ at the zero frequency and at integer multiples of 50 Hz:

$$C_2 = \{x \in \mathbb{R}^N \mid \langle \ell \in \mathbb{D}_2 \rangle \quad x^{(i)} = 0\}$$

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

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

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

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

$$C_4 = \{x \in \mathbb{R}^N \mid \langle \ell \in \mathbb{D}_4 \rangle \quad x^{(i)} = 0\},$$ (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 $(C_1^{(i)})_{1 \leq i \leq N}$ with the following constraint:

$$C_1 = \{x \in \mathbb{R}^N \mid \sum_{\ell \in \mathbb{D}_1} d^{(\ell)}_{C_1^{(i)}}(x^{(i)}) \leq \eta\},$$ (42)

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

$$\text{minimize} \quad \|x\|^2 + \sum_{s=1}^4 c_s(x).$$ (43)

As mentioned in the introduction, proximal algorithms can efficiently solve the above problem as long as the projection associated to each constraint $(C_s)_{1 \leq s \leq 4}$ 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 $R = 4$, $T_1 = \cdots = T_2 = I$, $g_1 = \|\cdot\|^2$, $g_s = \ell c_s$, for each $s \in \{2, \ldots, 4\}$, $L = \|N\|$ and

$$E = \{(y, \zeta) \in \mathbb{R}^N \times \mathbb{R}^N \mid \langle y, \zeta \rangle < \eta\},$$ (44)

$$V = \{\zeta \in \mathbb{R}^N \mid 1_d \zeta \leq \eta\}.$$ (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 $\eta$. Note that for large values of $\eta$, 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 $\eta$ 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 $\chi$ (bottom). (a) Minimizer of $\|x\|^2_2 + d_{C_1}(x) + d_{C_2}(x)$ subject to $x \in \left( \bigcap_{k \in D_1} C_k^2 \right) \cap C_2$. (b) Minimizer of $\|x\|^2_2 + \sum_{k \in D_1} d_{C_k^2}(x)$ 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 $\chi$ (bottom).
7 Appendix

7.1 Proof of Proposition 1

For every \((y, \zeta) \in \mathcal{H} \times \mathbb{R}\), let \((p, \theta) = P_{\partial \psi}(y, \zeta)\). If \(\psi(y) \leq \zeta\), then \(p = y\) and \(\theta = \zeta = \max \{\psi(p), \zeta\}\). In addition,\(^{6}\)
\[
(\forall u \in \mathcal{H}) \quad 0 = \frac{1}{2} \|p - y\|^2 + \frac{1}{2} \left( \max \{\psi(p) - \zeta, 0\} \right)^2
\]
\[
\leq \frac{1}{2} \|y - \zeta\|^2 + \frac{1}{2} \left( \max \{\psi(u) - \zeta, 0\} \right)^2 ,
\]
which shows that (18) holds. Let us now consider the case when \(\psi(y) > \zeta\). From the definition of the projection, we get
\[
(p, \theta) = \arg\min_{u \in \mathcal{H}} \|u - y\|^2 + (\zeta - \theta)^2 .
\]

From the Karush-Kuhn-Tucker theorem [68, Theorem 5.2], there exists \(\alpha \in [0, +\infty]\) such that
\[
(p, \theta) = \arg\min_{u \in \mathcal{H}, \zeta \in \mathbb{R}} \frac{1}{2} \|u - y\|^2 + \frac{1}{2} (\zeta - \theta)^2 + \alpha (\psi(u) - \zeta)
\]
where the Lagrange multiplier \(\alpha\) is such that \(\alpha (\psi(y) - \theta) = 0\). 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 \(\psi(p) = \theta\). In addition, differentiating the Lagrange functional in (48) w.r.t. \(\xi\) yields \(\psi(p) = \theta = \zeta + \alpha \geq \zeta\). Hence, \((p, \theta)\) in (47) is such that \(\theta = \psi(p) = \max \{\psi(p), \zeta\}\) and
\[
p = \arg\min_{u \in \mathcal{H}} \|u - y\|^2 + (\psi(u) - \zeta)^2 .
\]

Furthermore, as \(\psi(y) > \zeta\), we obtain
\[
\inf_{u \in \mathcal{H}} \|u - y\|^2 = \|P_{\mathcal{H} \times \mathcal{F}}(y) - y\|^2 = \inf_{u \in \mathcal{H}} \|u - y\|^2 ,
\]
where we have used the fact that \(P_{\mathcal{H} \times \mathcal{F}}(\psi)\) belongs to the boundary of \(\mathcal{H} \times \mathcal{F}\) which is equal to \(\{ u \in \mathcal{H} | \psi(u) = \zeta \}\) since \(\psi\) is lower-semicontinuous and \(\text{dom} \, \psi\) is open [40, Corollary 8.38]. We have then
\[
\inf_{u \in \mathcal{H}} \|u - y\|^2 = \inf_{u \in \mathcal{H}} \|u - y\|^2 + \inf_{u \in \mathcal{H}} (\psi(u) - \zeta)^2 .
\]

Altogether, (49) and (51) lead to
\[
p = \arg\min_{u \in \mathcal{H}} \frac{1}{2} \|u - y\|^2 + \frac{1}{2} (\psi(u) - \zeta)^2
\]
which is equivalent to (18) since \(\frac{1}{2} \max \{\psi - \zeta, 0\}^2 \in \Gamma_0(\mathcal{H})\).

7.2 Proof of Proposition 3

Since \(\max \{\psi - \zeta, 0\}^2\) is an even function, \(\text{prox}_{\frac{1}{2} \max \{\psi - \zeta, 0\}^2}\) is an odd function.\(^{6}\) 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 \(p = \text{proj}_\gamma(y)\) is uniquely defined as \(y - p \in \mathcal{H}\), we deduce that \(p = \text{prox}_{\frac{1}{2} \max \{\psi - \zeta, 0\}^2}(y)\) is such that
\[
p - y + \frac{q \tau p^{q - 1}}{r} (\tau p^q - \zeta) = 0 ,
\]
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 \(y \in [0, (\zeta \gamma \tau)^{1/q}]\), it can be deduced from [69, Corollary 2.5], that \(p = \text{prox}_{\frac{1}{2} \max \{\psi - \zeta, 0\}^2}(y) \in [0, (\zeta \gamma \tau)^{1/q}]\). Since \(\forall v \in [0, (\zeta \gamma \tau)^{1/q}] \max \{\psi(v) - \zeta, 0\} = 0\), we have \(p = y\). On the other hand if \(y > (\zeta \gamma \tau)^{1/q}\), as the proximity operator of a function from \(\mathcal{H}\) to \(\mathcal{H}\) is continuous and increasing [69, Proposition 2.4], we obtain
\[
p = \text{prox}_{\frac{1}{2} \max \{\psi - \zeta, 0\}^2}(y) \geq \text{prox}_{\frac{1}{2} \max \{\psi - \zeta, 0\}^2}\left((\zeta \gamma \tau)^{1/q}\right) = (\zeta \gamma \tau)^{1/q} .
\]

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

7.3 Proof of Proposition 4

Let us notice that \(\frac{1}{2} \max \{\tau d_\psi^2 - \zeta, 0\}^2 = \psi \circ d_C\) where \(\psi = \frac{1}{2} \max \{\tau d_C, 0\}^2\). According to [49, Proposition 2.7], for every \(y \in \mathcal{H}\),
\[
\text{prox}_{\psi \circ d_C}(y) = \begin{cases}
 y , & \text{if } y \in C ,
 P_C(y) , & \text{if } d_C(y) \leq \max \partial \psi(0) ,
 \alpha y + (1 - \alpha) P_C(y) , & \text{if } d_C(y) > \max \partial \psi(0)
\end{cases}
\]
where \(\alpha = \frac{\text{prox}_{\psi} (d_C(y))}{d_C(y)}\). In addition, we have
\[
\partial \psi(0) = \begin{cases}
 [\zeta \gamma, -\zeta \gamma] , & \text{if } \zeta < 0 \text{ and } q = 1 ,
 \{0\} , & \text{otherwise},
\end{cases}
\]
and, according to Proposition 2, when \(\zeta \leq 0 \text{ and } d_C(y) \leq -\tau \zeta\), \(\text{prox}_\psi (d_C(y)) = 0\). 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 \(P_{\psi \theta}(y, \zeta)\), we have to find

\[
\min_{\theta \in [0, +\infty)} \left( (\theta - \zeta)^2 + \min_{|p| \leq \theta/(\psi(1))} \|p - y\|^2 \right). 
\] (57)

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

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

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

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

The function \(\phi\) belongs to \(I_0^2(\mathbb{R})\) since, for each \(m \in \{1, \ldots, M\}\), \(\max\{|\tau(m)^{-1} (v(m) - \zeta)|\} \leq \|v\|\) is finite convex and \((\cdot)^2\) is finite convex and increasing on \([0, +\infty)\). In addition, \(\phi\) is differentiable and such that, for every \(v \in \mathbb{R}\) and \(k \in \{1, \ldots, M + 1\}\),

\[
v^{(k-1)} < v \leq v^{(k)} \Rightarrow \phi(v) = \frac{1}{2} \sum_{m=k}^{M} (\tau(m)^{-1} (v - v(m)) \geq 0. \] (60)

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

\[
\zeta - \chi = \sum_{m=1}^{M} (\tau(m)^{-1})^2 (v - v(m))\quad \Leftrightarrow \quad (32) \] (61)

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

References


