

Epigraphical Projection and Proximal Tools for Solving Constrained Convex Optimization Problems: Part I

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 Projection and Proximal Tools for Solving Constrained Convex Optimization Problems: Part I. 2012. hal-00744603v2

HAL Id: hal-00744603
https://hal.science/hal-00744603v2

Submitted on 15 Nov 2012

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 Projection and Proximal Tools for Solving Constrained Convex Optimization Problems: Part I

Giovanni Chierchia*, Nelly Pustelnik†,
Jean-Christophe Pesquet‡, and Béatrice Pesquet-Popescu*

November 15, 2012

Abstract

We propose a proximal approach to deal with convex optimization problems involving nonlinear constraints. A large family of such 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, but possibly overlapping, blocks of the signal. For this class of constraints, the associated projection operator generally does not have a closed form. We circumvent this difficulty by splitting the lower level set into as many epigraphs as functions involved in the sum. A closed half-space constraint is also enforced, in order to limit the sum of the introduced epigraphical variables to the upper bound of the original lower level set.

In this paper, we focus on a family of constraints involving linear transforms of $\ell_{1,p}$ balls. Our main theoretical contribution is to provide closed form expressions of the epigraphical projections associated with the Euclidean norm ($p = 2$) and the sup norm ($p = +\infty$). The proposed approach is validated in the context of image restoration with missing samples, by making use of TV-like constraints. Experiments show that our method leads to significant improvements in term of convergence speed over existing algorithms for solving similar constrained problems.

1 Introduction

As an offspring of the wide interest in frame representations and sparsity promoting techniques for data recovery, there has been an emergence of proximal methods to efficiently solve convex optimization problems arising in inverse problems. Examples of applications of these methods can be found in f-MRI reconstruction [1, 2], satellite image restoration [3, 4], microscopy image deconvolution [5, 6], computed tomography [7], Positron Emission Tomography [8, 9], texture-geometry decomposition [10, 11, 12], machine learning [13, 14], stereo vision [15], and audio processing [16, 17]. Proximal algorithms have gained much popularity in solving large-size optimization problems involving non-differentiable (or even non finite) functions. One of the main advantages of these methods is that they are amenable to parallel implementations. For a survey on proximal algorithms and their applications, the reader is referred to [18, 13]. Note also that some of these methods

*G. Chierchia (Corresponding author) and B. Pesquet-Popescu are with Télécom ParisTech/Institut Télécom, LTCI, UMR CNRS 5141, 75014 Paris, France (e-mail: first.last@telecom-paristech.fr).

†N. Pustelnik is with ENS Lyon, Laboratoire de Physique, UMR CNRS 5672, F69007 Lyon, France (e-mail: nelly.pustelnik@ens-lyon.fr).

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

are closely related to augmented Lagrangian approaches [19, 20, 21]. Even if proximal algorithms and the associated convergence properties have been deeply investigated [22, 23, 24, 25], some questions persist in their use for solving inverse problems. A first question is: how can we set the parameters serving to enforce the regularity of the solution in an automatic way? Another question is related to the selection of the most appropriate algorithm within the class of proximal algorithms for a given application. This also raises the question of the computation of the proximity operators associated with the different functions involved in the criterion. Various strategies were proposed in order to address the first question [26, 27, 28, 29, 30], but the computational cost of these methods is often high, especially when several regularization parameters have to be set. Alternatively, it has been recognized for a long time that incorporating constraints directly on the solutions [31, 32, 33, 34, 35] instead of considering regularized functions may often facilitate the choice of the involved parameters. Indeed, in a constrained formulation, the constraint bounds are usually related to some physical properties of the target solution or some knowledge of the degradation process, e.g. the noise statistical properties. Note also that there exist some conceptual Lagrangian equivalences between regularized solutions to inverse problems and constrained ones, although some caution should be taken when the regularization functions are nonsmooth (see [36] where the case of a single regularization parameter is investigated).

In this context, the objective of this paper is to provide an answer to the second question, that is to propose efficient parallel algorithms for solving the following constrained convex optimization problem:

Problem 1.1.

$$\underset{x\in\mathcal{H}}{\text{minimize}} \sum_{r=1}^{R} g_r(T_r x) \quad \text{s. t.} \quad \begin{cases} H_1 x \in C_1, \\ \vdots \\ H_S x \in C_S, \end{cases} \tag{1}$$

where

(i) for every $r \in \{1,\ldots,R\}$, T_r is a bounded linear operator from a real Hilbert space $\mathcal{H}$ to $\mathbb{R}^{N_r}$;

(ii) for every $r \in \{1,\ldots,R\}$, g_r is a proper lower-semicontinuous convex function from $\mathbb{R}^{N_r}$ to $]-\infty,+\infty]$;

(iii) for every $s \in \{1,\ldots,S\}$, H_s is a bounded linear operator from $\mathcal{H}$ to $\mathbb{R}^{M_s}$;

(iv) for every $s \in \{1,\ldots,S\}$, C_s is a nonempty closed convex subset of $\mathbb{R}^{M_s}$.

A large number of inverse problems can be formulated under the form of Problem 1.1. A classical application example is image recovery from blurred and noisy observations. Let z denote the vector of observed data. When the noise is assumed to be zero-mean additive white Gaussian, the restored data can be obtained by solving Problem 1.1 where $R = 2$, $S = 0$, $g_1 = \lambda\|\cdot\|_1$ with $\lambda \in]0,+\infty[$, T_1 is an analysis frame operator [37] allowing us to sparsify signal x, $g_2 = \|\cdot - z\|^2$, and T_2 denotes the matrix associated with the degradation blur [38, 23]. On the other hand, a constrained formulation of the same restoration problem leads to Problem 1.1 with $R = 1$, $S = 1$, $g_1 = \|\cdot\|_1$, T_1 is the same frame operator as previously, and

$$(\forall x \in \mathcal{H}) \qquad H_1 x \in C_1 \quad \Leftrightarrow \quad \|H_1 x - z\|^2 \leq \eta_1.$$

Here, H_1 denotes the matrix associated with the degradation blur, and η_1 is a positive constant which is typically chosen proportional to the noise variance. This specific constraint formulation has been considered in [39].

The present work aims at designing efficient methods in order to deal with Problem 1.1 when the convex constraints are expressed as follows: for every $s \in \{1,\ldots,S\}$,

$$(\forall x \in \mathcal{H}) \qquad H_s x \in C_s \quad \Leftrightarrow \quad h_s(H_s x) \leq \eta_s, \tag{2}$$

where $\eta_s \in \mathbb{R}$ and h_s is a proper lower-semicontinuous function from $\mathbb{R}^{M_s}$ to $]-\infty,+\infty]$.

The paper is organized as follows. In Section 2, we motivate the choice of proximal tools and recall some of their theoretical properties. In addition, closed form expressions for specific proximity operators are derived. Then, in order to deal with a constraint expressed under the general form (2), a splitting approach involving an epigraphical projection is proposed. This epigraphical projection technique is described in detail in Section 3. Experiments in the context of image reconstruction are presented in Section 4. Finally, some conclusions are drawn in Section 5.

Notation: Let $\mathcal{H}$ be a real Hilbert space endowed with the norm $\|\cdot\|$ and the scalar product $\langle\cdot \mid \cdot\rangle$. $\Gamma_0(\mathcal{H})$ denotes the set of proper lower-semicontinuous convex functions from $\mathcal{H}$ to $]-\infty,+\infty]$. Recall that a function $\varphi\colon \mathcal{H} \to]-\infty,+\infty]$ is proper if its domain $\operatorname{dom}\varphi = \big\{y \in \mathcal{H} \;\big|\; \varphi(y) < +\infty\big\}$ is nonempty. The epigraph of $\varphi \in \Gamma_0(\mathcal{H})$ is the nonempty closed convex subset of $\mathcal{H}\times\mathbb{R}$ defined as $\operatorname{epi}\varphi = \big\{(y,\zeta) \in \mathcal{H}\times\mathbb{R} \;\big|\; \varphi(y) \le \zeta\big\}$ and the lower level set of φ at height $\zeta \in \mathbb{R}$ is the nonempty closed convex subset of $\mathcal{H}$ defined as $\operatorname{lev}_{\le\zeta}\varphi = \big\{y \in \mathcal{H} \;\big|\; \varphi(y) \le \zeta\big\}$. A subgradient of φ at $y \in \mathcal{H}$ is an element of its subdifferential defined as $\partial\varphi(y) = \big\{t \in \mathcal{H} \;\big|\; (\forall u \in \mathcal{H})\ \ \varphi(u) \ge \varphi(y) + \langle t \mid u-y\rangle\big\}$.[1] The indicator function $\iota_C \in \Gamma_0(\mathcal{H})$ of a nonempty closed convex subset C of $\mathcal{H}$ is given by

$$(\forall y \in \mathcal{H}) \qquad \iota_C(y) = \begin{cases} 0, & \text{if } y \in C, \\ +\infty, & \text{otherwise.} \end{cases} \tag{3}$$

The relative interior of a subset C of $\mathcal{H}$ is denoted by $\operatorname{ri} C$.

2 Proximal tools

2.1 From gradient descent to proximal algorithms

The first methods for finding a solution to an inverse problem were restricted to the use of a differentiable cost function [40], i.e. Problem 1.1 where $S = 0$ and, for every $r \in \{1,\dots,R\}$, g_r denotes a differentiable function. In this context, gradient-based algorithms, e.g. nonlinear conjugate gradient or quasi-Newton methods, are popular (see [41] and references therein for recent developments concerning these approaches). However, in order to model additional properties, sparsity promoting penalizations or hard constraints ($S \ge 1$) may be introduced and the diffentiability property is not satisfied anymore. One way to circumvent this difficulty is to resort to smart approximations in order to smooth the involved non-differentiable functions [42, 43, 44]. If one wants to address the original nonsmooth problem without introducing approximation errors, one may apply some specific algorithms e.g. Gauss-Seidel or Uzawa methods, the convergence of which is guaranteed under restrictive assumptions [45]. Interior point methods [46] can also be employed for small to medium size optimization problems.

On the other hand, in order to solve convex feasibility problems, i.e. to find a vector belonging to the intersection of convex sets (Problem 1.1 with $R = 0$), iterative projection methods were developed. The projection onto convex sets algorithm (POCS) is one of the most popular approach to solve data recovery problems [47, 48, 31, 49]. A drawback of POCS is that it is not well-suited for parallel implementations. The Parallel Projection Method (PPM) and Method of Parallel Projections (MOPP) are variants of POCS making use of parallel projections. Moreover, these algorithms were designed to efficiently solve inconsistent feasibility problems (when the intersection of the convex set is empty). Thorough comparisons between projection methods have been performed in [50, 51].

[1] When φ is Gâteaux-differentiable at y, $\partial\varphi(y) = \{\nabla\varphi(y)\}$ where $\nabla\varphi(y)$ is the gradient of φ at y.

Computing the projection P_C onto a nonempty closed convex subset C of a real Hilbert space $\mathcal{H}$ requires to solve a constrained quadratic minimization problem:

$$(\forall y \in \mathcal{H}) \qquad P_C(y) = \operatorname*{argmin}_{u \in C} \|u - y\|. \tag{4}$$

The distance to C of every point $y \in \mathcal{H}$ is then given by $d_C(y) = \|y - P_C(y)\|$. However, it turns out that a closed form expression of the solution to (4) is available in a limited number of instances. One such well-known example is the projection onto a closed half-space:

Proposition 2.1. [44] *Let $\eta \in \mathbb{R}$, let $t \in \mathcal{H} \setminus \{0\}$, and let $C = \{u \in \mathcal{H} \,|\, \langle u \,|\, t\rangle \leq \eta\}$. The projection of $y \in \mathcal{H}$ onto C is expressed as*

$$P_C(y) = \begin{cases} y, & \text{if} \quad \langle y \,|\, t\rangle \leq \eta, \\ y + \dfrac{\eta - \langle y \,|\, t\rangle}{\|t\|^2} t, & \text{otherwise.} \end{cases} \tag{5}$$

When an expression of the direct projection is not available, the convex set C can be approximated by a half-space, which leads to the concept of subgradient projection.

Proposition 2.2. [44] *Let $\eta \in \mathbb{R}$ and let $\varphi \in \Gamma_0(\mathcal{H})$. Suppose that $C = \operatorname{lev}_{\leq \eta} \varphi \neq \varnothing$. Let $y \in \mathcal{H}$ and let $t \in \partial\varphi(y)$ be a subgradient of φ at y. Then, C is a subset of the half-space*

$$C_\eta^y = \big\{u \in \mathcal{H} \;\big|\; \langle u - y \,|\, t\rangle \leq \eta - \varphi(y)\big\}. \tag{6}$$

If $t \neq 0$, the projection of y onto C_η^y is a subgradient projection of $y \in \mathcal{H}$ onto C.

The subgradient projection plays a key role in Polyak algorithm that alternates a subgradient projection and an exact projection [52] (see also [53] for alternative projected subgradient approaches). An efficient block iterative surrogate splitting method was proposed in [54] in order to solve Problem 1.1 when, for every $r \in \{1, \ldots, R\}$, $g_r = \|\cdot - z_r\|^2$ where $z_r \in \mathbb{R}^{N_r}$. A main limitation of this method is that the global objective function must be strictly convex. For recent works about subgradient projection methods, the readers may refer to [55, 56].

A way to overcome this difficulty consists of considering proximal approaches. The key tool in these methods is the proximity operator [57] of a function $\varphi \in \Gamma_0(\mathcal{H})$, defined as

$$(\forall y \in \mathcal{H}) \qquad \operatorname{prox}_\varphi(y) = \operatorname*{argmin}_{u \in \mathcal{H}} \frac{1}{2}\|u - y\|^2 + \varphi(u). \tag{7}$$

The proximity operator generalizes the notion of projection onto a nonempty closed convex subset C of $\mathcal{H}$ in the sense that $\operatorname{prox}_{\iota_C} = P_C$. For every $y \in \mathcal{H}$, $p = \operatorname{prox}_\varphi(y)$ is uniquely defined through the inclusion

$$y - p \in \partial\varphi(p). \tag{8}$$

Proximity operators enjoy many additional interesting properties. Some of them are recalled next.

Example 2.3. [23] *Let $\tau \in \,]0, +\infty[$, $\beta \in [1, +\infty[$, and set*

$$\varphi \colon \mathbb{R} \to \,]-\infty, +\infty] : \xi \mapsto \tau|\xi|^\beta. \tag{9}$$

Then, for every $\xi \in \mathbb{R}$, $\operatorname{prox}_\varphi(\xi)$ is given by

$$\begin{cases} \operatorname{sign}(\xi)\max\{|\xi| - \tau, 0\}, & \text{if } \beta = 1, \\ \xi + \frac{4\tau}{3\,.\,2^{1/3}}\big((\epsilon - \xi)^{1/3} - (\epsilon + \xi)^{1/3}\big), & \text{if } \beta = \frac{4}{3}, \\ \xi + \frac{9\tau^2 \operatorname{sign}(\xi)}{8}\Big(1 - \sqrt{1 + \frac{16|\xi|}{9\tau^2}}\Big), & \text{if } \beta = \frac{3}{2}, \\ \frac{\xi}{1+2\tau}, & \text{if } \beta = 2, \\ \operatorname{sign}(\xi)\frac{\sqrt{1+12\tau|\xi|}-1}{6\tau}, & \text{if } \beta = 3, \end{cases} \tag{10}$$

where $\epsilon = \sqrt{\xi^2 + 256\tau^3/729}$ *and* sign *is the signum function.*

It can be noticed that the proximity operator associated with $\beta = 1$ reduces to a soft thresholding [58].

The class of proximal methods includes primal algorithms [59, 23, 24, 60, 61, 62, 63, 18, 21] and primal-dual algorithms [64, 65, 66, 67, 68, 69, 70]. Primal algorithms generally require to compute inverses of some linear operators (typically, $\sum_{r=1}^{R} T_r^* T_r + \sum_{s=1}^{S} H_s^* H_s$), while primal-dual ones only require to compute $(T_r)_{1\leq r\leq R}$ and $(H_s)_{1\leq s\leq S}$, and their adjoints $(T_r^*)_{1\leq r\leq R}$ and $(H_s^*)_{1\leq s\leq S}$. Consequently, primal-dual methods are often easier to implement than primal ones, but their convergence may be slower [71, 72].

2.2 Proximity operators: new closed forms

The projection onto a convex set such as defined in (2) often is non trivial. In Section 3, we will show that this problem can be solved by resorting to a set of epigraphical projections which are easier to compute. The key point is that epigraphical projections are closely related to proximity operators. In the following, we provide some results that will be useful for the calculation of the projection onto the epigraph of a convex function (all the proofs are provided in the appendix).

Proposition 2.4. *Let* $\mathcal{H}$ *be a real Hilbert space and let* $\mathcal{H} \times \mathbb{R}$ *be equipped with the standard product space norm. Let* φ *be a function in* $\Gamma_0(\mathcal{H})$ *such that* $\operatorname{dom}\varphi$ *is open. The projector* $P_{\operatorname{epi}\varphi}$ *onto the epigraph of* φ *is given by:*

$$\big(\forall (y,\zeta) \in \mathcal{H}\times\mathbb{R}\big) \qquad P_{\operatorname{epi}\varphi}(y,\zeta) = (p,\theta) \tag{11}$$

where

$$\begin{cases} p &= \operatorname{prox}_{\frac{1}{2}(\max\{\varphi-\zeta,0\})^2}(y), \\ \theta &= \max\{\varphi(p),\zeta\}. \end{cases} \tag{12}$$

Note that alternative characterizations of the epigraphical projection can be found in [73, Propositions 9.17, 28.28].

Let φ be a function in $\Gamma_0(\mathcal{H})$ such that $\operatorname{dom}\varphi$ is open. From Proposition 2.4, we see that $\operatorname{prox}_{\frac{1}{2}(\max\{\varphi-\zeta,0\})^2}$ with $\zeta \in \mathbb{R}$ plays a prominent role in the calculation of the projection onto $\operatorname{epi}\varphi$. We now provide examples of functions φ for which this proximity operator admits a simple form.

Proposition 2.5. *Let* $\beta \in [1,+\infty[$, $\tau \in]0,+\infty[$. *Assume that*

$$(\forall y \in \mathbb{R}) \qquad \varphi(y) = \tau|y|^\beta. \tag{13}$$

If $\zeta \in]-\infty, 0]$, *then for every* $y \in \mathbb{R}$

$$\operatorname{prox}_{\frac{1}{2}(\max\{\varphi-\zeta,0\})^2}(y) = \begin{cases} \dfrac{\operatorname{sign}(y)}{1+\tau^2}\max\{|y|+\tau\zeta, 0\}, & \text{if } \beta = 1, \\ \operatorname{sign}(y)\chi_0, & \text{if } \beta > 1, \end{cases} \tag{14}$$

where χ_0 *is the unique solution on* $[0,+\infty[$ *of the equation*

$$\beta\tau^2\chi^{2\beta-1} - \beta\tau\zeta\chi^{\beta-1} + \chi = |y|. \tag{15}$$

If $\zeta \in]0,+\infty[$, *then, for every* $y \in \mathbb{R}$,

$$\operatorname{prox}_{\frac{1}{2}(\max\{\varphi-\zeta,0\})^2}(y) = \begin{cases} y, & \text{if } |y| \leq \big(\zeta/\tau\big)^{1/\beta}, \\ \operatorname{sign}(y)\chi_{(\zeta/\tau)^{1/\beta}}, & \text{otherwise,} \end{cases} \tag{16}$$

where $\chi_{(\zeta/\tau)^{1/\beta}}$ *is the unique solution on* $[(\frac{\zeta}{\tau})^{1/\beta}, +\infty[$ *of* (15).

Note that, when β is a rational number, (15) is equivalent to a polynomial equation for which either closed form solutions are known or standard numerical solutions exist.

Proposition 2.6. *Let $\mathcal{H}$ be a real Hilbert space and let C be a nonempty convex subset of $\mathcal{H}$. Let $\beta \in [1, +\infty[$, $\tau \in]0, +\infty[$, and $\zeta \in \mathbb{R}$. Then, for every $y \in \mathcal{H}$,*

$$\operatorname{prox}_{\frac{1}{2}(\max\{\tau d_C^\beta - \zeta, 0\})^2}(y) = \begin{cases} y, & \text{if } y \in C, \\ \alpha y + (1-\alpha) P_C(y), & \text{otherwise,} \end{cases} \tag{17}$$

where $\alpha = \dfrac{\operatorname{prox}_{\frac{1}{2}(\max\{\tau|\cdot|^\beta - \zeta, 0\})^2}\big(d_C(y)\big)}{d_C(y)}$ and the expression of $\operatorname{prox}_{\frac{1}{2}(\max\{\tau|\cdot|^\beta - \zeta, 0\})^2}$ is provided by Proposition 2.5.

The following result is a consequence of the previous one:

Corollary 2.7. *Let $\mathcal{H}$ be a real Hilbert space. Assume that*

$$(\forall y \in \mathcal{H}) \qquad \varphi(y) = \tau \|y - z\| \tag{18}$$

where $z \in \mathcal{H}$, $\zeta \in \mathbb{R}$ and $\tau \in]0, +\infty[$. Then, for every $y \in \mathbb{R}$,

$$\operatorname{prox}_{\frac{1}{2}(\max\{\varphi - \zeta, 0\})^2}(y) = \begin{cases} z, & \text{if } \|y - z\| < -\tau\zeta, \\ y, & \text{if } \|y - z\| < \frac{\zeta}{\tau}, \\ z + \alpha(y - z), & \text{otherwise,} \end{cases} \tag{19}$$

where $\alpha = \frac{1}{1+\tau^2}\Big(1 + \frac{\tau\zeta}{\|y-z\|}\Big)$.

Another result which will be used is the following:

Proposition 2.8. *Let $M \in \mathbb{N}^*$ and let φ be defined as*

$$(\forall y \in \mathbb{R}) \quad \varphi(y) = \frac{1}{2} \sum_{m=1}^{M} \big(\max\{\tau^{(m)}(\nu^{(m)} - y), 0\} \big)^2 \tag{20}$$

where $(\tau^{(1)}, \ldots, \tau^{(M)})^\top \in \mathbb{R}^M$, and $(\nu^{(1)}, \ldots, \nu^{(M)})^\top \in \mathbb{R}^M$ is such that $\nu^{(1)} \leq \ldots \leq \nu^{(M)}$. Set $\nu^{(0)} = -\infty$ and $\nu^{(M+1)} = +\infty$. Then, $\varphi \in \Gamma_0(\mathbb{R})$, and for every $y \in \mathbb{R}$,

$$\operatorname{prox}_\varphi(y) = \frac{y + \sum_{m=1}^{\overline{m}-1} \nu^{(m)} (\tau_-^{(m)})^2 + \sum_{m=\overline{m}}^{M} \nu^{(m)} (\tau_+^{(m)})^2}{1 + \sum_{m=1}^{\overline{m}-1} (\tau_-^{(m)})^2 + \sum_{m=\overline{m}}^{M} (\tau_+^{(m)})^2}, \tag{21}$$

where, for every $m \in \{1, \ldots, M\}$, $\tau_-^{(m)} = \min(\tau^{(m)}, 0)$ and $\tau_+^{(m)} = \max(\tau^{(m)}, 0)$, and $\overline{m}$ is the unique integer in $\{1, \ldots, M+1\}$ such that

$$\nu^{(\overline{m}-1)} < \frac{y + \sum_{m=1}^{\overline{m}-1} \nu^{(m)} (\tau_-^{(m)})^2 + \sum_{m=\overline{m}}^{M} \nu^{(m)} (\tau_+^{(m)})^2}{1 + \sum_{m=1}^{\overline{m}-1} (\tau_-^{(m)})^2 + \sum_{m=\overline{m}}^{M} (\tau_+^{(m)})^2} \leq \nu^{(\overline{m})} \tag{22}$$

(with the convention $\sum_{m=1}^{0} \cdot = \sum_{m=M+1}^{M} \cdot = 0$).

3 Projection computation

We now turn our attention to convex sets for which the associated projection does not have a closed form, and we show that under some appropriate assumptions, it is possible to circumvent this difficulty. Let C denote such a nonempty closed convex subset of $\mathbb{R}^M$ and assume that, for every $y \in \mathbb{R}^M$,

$$y \in C \quad \Leftrightarrow \quad h(y) = \sum_{\ell=1}^{L} h^{(\ell)}(\mathsf{y}^{(\ell)}) \leq \eta \tag{23}$$

where $\eta \in \mathbb{R}$. Hereabove, the generic vector y has been decomposed into blocks of coordinates as follows

$$y = [\underbrace{(\mathsf{y}^{(1)})^\top}_{\text{size } M^{(1)}}, \ldots, \underbrace{(\mathsf{y}^{(L)})^\top}_{\text{size } M^{(L)}}]^\top \tag{24}$$

and, for every $\ell \in \{1, \ldots, L\}$, $\mathsf{y}^{(\ell)} \in \mathbb{R}^{M^{(\ell)}}$ and $h^{(\ell)}$ is a function in $\Gamma_0(\mathbb{R}^{M^{(\ell)}})$ such that $\mathrm{ri}(\mathrm{dom}\, h^{(\ell)}) \neq \varnothing$. Actually, any closed convex subset C can be expressed in this way by setting $\eta = 0$, $L = 1$ and $h = d_C$.

By introducing now the auxiliary vector $\zeta = \left(\zeta^{(\ell)}\right)_{1 \leq \ell \leq L} \in \mathbb{R}^L$, Condition (23) can be equivalently rewritten as[2]

$$\begin{cases} \displaystyle\sum_{\ell=1}^{L} \zeta^{(\ell)} \leq \eta, & (25) \\ (\forall \ell \in \{1, \ldots, L\}) \qquad h^{(\ell)}(\mathsf{y}^{(\ell)}) \leq \zeta^{(\ell)}. & (26) \end{cases}$$

Let us now introduce the closed half-space of $\mathbb{R}^L$ defined as

$$V = \left\{\zeta \in \mathbb{R}^L \;\middle|\; 1_L^\top \zeta \leq \eta\right\}, \tag{27}$$

with $1_L = (1, \ldots, 1)^\top \in \mathbb{R}^L$, and the closed convex set

$$E = \left\{(y, \zeta) \in \mathbb{R}^M \times \mathbb{R}^L \;\middle|\; (\forall \ell \in \{1, \ldots, L\})\ (\mathsf{y}^{(\ell)}, \zeta^{(\ell)}) \in \mathrm{epi}\, h^{(\ell)}\right\}. \tag{28}$$

Then, Constraint (25) means that $\zeta \in V$, whereas Constraint (26) is equivalent to $(y, \zeta) \in E$. In other words, the constraint set C can be split into the two constraint sets V and E provided that L additional scalar variables $(\zeta^{(\ell)})_{1 \leq \ell \leq L}$ are introduced in the original optimization problem. Dealing with additional constraints in the original problem is not a problem for proximal splitting algorithms as far as the projections onto the associated constraint sets can be computed. An example of use of such an algorithm will be described in detail in Section 4 .

In the present case, the projection onto V is given by Proposition 2.1, whereas the projection onto E is given by

$$(\forall (y, \zeta) \in \mathbb{R}^M \times \mathbb{R}^L) \qquad P_E(y, \zeta) = (p, \theta) \tag{29}$$

where $\theta = (\theta^{(\ell)})_{1 \leq \ell \leq L}$, vector $p \in \mathbb{R}^M$ is blockwise decomposed as $p = [(\mathsf{p}^{(1)})^\top, \ldots, (\mathsf{p}^{(L)})^\top]^\top$ like in (24), and

$$(\forall \ell \in \{1, \ldots, L\}) \quad (\mathsf{p}^{(\ell)}, \theta^{(\ell)}) = P_{\mathrm{epi}\, h^{(\ell)}}(\mathsf{y}^{(\ell)}, \zeta^{(\ell)}). \tag{30}$$

So, the problem reduces to the lower-dimensional problem of the determination of the projection onto the convex subset $\mathrm{epi}\, h^{(\ell)}$ of $\mathbb{R}^{M^{(\ell)}} \times \mathbb{R}$ for each $\ell \in \{1, \ldots, L\}$. By using the results in Section 2.2, these projections can be shown to have a closed form expression in a number of cases of particular interest. Two examples are now given:

[2]Note that the inequality in (25) can be also replaced by an equality although it makes little difference in our approach.

- Euclidean norm functions of the form

$$\big(\forall \mathsf{y}^{(\ell)} \in \mathbb{R}^{M^{(\ell)}}\big) \qquad h^{(\ell)}(\mathsf{y}^{(\ell)}) = \tau^{(\ell)} \|\mathsf{y}^{(\ell)}\| \tag{31}$$

where $\ell \in \{1, \ldots, L\}$ and $\tau^{(\ell)} \in]0, +\infty[$.

Proposition 3.1. *Assume that $h^{(\ell)}$ is given by* (31).
Then, for every $(\mathsf{y}^{(\ell)}, \zeta^{(\ell)}) \in \mathbb{R}^{M^{(\ell)}} \times \mathbb{R}$,

$$P_{\mathrm{epi}\, h^{(\ell)}}(\mathsf{y}^{(\ell)}, \zeta^{(\ell)}) = \begin{cases} (0,0), & \text{if } \|\mathsf{y}^{(\ell)}\| < -\tau^{(\ell)}\zeta^{(\ell)}, \\ (\mathsf{y}^{(\ell)}, \zeta^{(\ell)}), & \text{if } \|\mathsf{y}^{(\ell)}\| < \frac{\zeta^{(\ell)}}{\tau^{(\ell)}}, \\ \alpha^{(\ell)} \left(\mathsf{y}^{(\ell)}, \tau^{(\ell)} \|\mathsf{y}^{(\ell)}\|\right), & \text{otherwise,} \end{cases} \tag{32}$$

where $\alpha^{(\ell)} = \dfrac{1}{1 + (\tau^{(\ell)})^2} \Big(1 + \dfrac{\tau^{(\ell)} \zeta^{(\ell)}}{\|\mathsf{y}^{(\ell)}\|}\Big)$.

The result of this proposition is a direct application of Proposition 2.4 and Corollary 2.7. As it will be shown in Section 4, this result is useful to deal with multivariate sparsity constraints [74] or total variation bounds [75, 76], since such constraints typically involve a sum of functions like (31) composed with linear operators corresponding to analysis transforms or gradient operators.

- Infinity norms defined as: for every $\ell \in \{1, \ldots, L\}$ and $\mathsf{y}^{(\ell)} = (\mathsf{y}^{(\ell,m)})_{1 \leq m \leq M^{(\ell)}} \in \mathbb{R}^{M^{(\ell)}}$,

$$h^{(\ell)}(\mathsf{y}^{(\ell)}) = \max \left\{ \frac{|\mathsf{y}^{(\ell,m)}|}{\tau^{(\ell,m)}} \mid 1 \leq m \leq M^{(\ell)} \right\}. \tag{33}$$

where $(\tau^{(\ell,m)})_{1 \leq m \leq M^{(\ell)}} \in]0, +\infty[^{M^{(\ell)}}$.

Proposition 3.2. *Assume that $h^{(\ell)}$ is given by* (33). *Let $(\nu^{(\ell,m)})_{1 \leq m \leq M^{(\ell)}}$ be a sequence of reals obtained by sorting $(|\mathsf{y}^{(\ell,m)}|/\tau^{(\ell,m)})_{1 \leq m \leq M^{(\ell)}}$ in ascending order, and set $\nu^{(\ell,0)} = -\infty$ and $\nu^{(\ell,M^{(\ell)}+1)} = +\infty$. Then, for every $\zeta^{(\ell)} \in \mathbb{R}$, $(\mathsf{p}^{(\ell)}, \theta^{(\ell)}) = P_{\mathrm{epi}\, h^{(\ell)}}(\mathsf{y}^{(\ell)}, \zeta^{(\ell)})$ is such that $\mathsf{p}^{(\ell)} = (\mathsf{p}^{(\ell,m)})_{1 \leq m \leq M^{(\ell)}}$, with*

$$\mathsf{p}^{(\ell,m)} = \begin{cases} \mathsf{y}^{(\ell,m)}, & \text{if } |\mathsf{y}^{(\ell,m)}| \leq \tau^{(\ell,m)}\theta^{(\ell)}, \\ \tau^{(\ell,m)}\theta^{(\ell)}, & \text{if } \mathsf{y}^{(\ell,m)} > \tau^{(\ell,m)}\theta^{(\ell)}, \\ -\tau^{(\ell,m)}\theta^{(\ell)}, & \text{if } \mathsf{y}^{(\ell,m)} < -\tau^{(\ell,m)}\theta^{(\ell)}, \end{cases} \tag{34}$$

$$\theta^{(\ell)} = \frac{\max\Big(\zeta^{(\ell)} + \sum_{m=\overline{m}}^{M^{(\ell)}} \nu^{(\ell,m)} (\tau^{(\ell,m)})^2, 0\Big)}{1 + \sum_{m=\overline{m}}^{M^{(\ell)}} (\tau^{(\ell,m)})^2}, \tag{35}$$

and $\overline{m}$ is the unique integer in $\{1, \ldots, M^{(\ell)} + 1\}$ such that

$$\nu^{(\ell,\overline{m}-1)} < \frac{\zeta^{(\ell)} + \sum_{m=\overline{m}}^{M^{(\ell)}} \nu^{(\ell,m)} (\tau^{(\ell,m)})^2}{1 + \sum_{m=\overline{m}}^{M^{(\ell)}} (\tau^{(\ell,m)})^2} \leq \nu^{(\ell,\overline{m})}. \tag{36}$$

The proof of this proposition is given in Appendix F.

Note that when $\tau^{(\ell,m)} \equiv 1$, function $h^{(\ell)}$ in (33) reduces to the standard infinity norm $\|\cdot\|_\infty$. This proposition can be thus employed to efficient deal with $\boldsymbol{\ell}_{1,\infty}$ regularization which has attracted much interest recently [77, 78, 79].

4 Experimental Results

In this section, we provide numerical examples to illustrate the usefulness of the proposed epigraphical projection method. The presented experiments focus on applications involving projections onto $\boldsymbol{\ell}_{1,p}$-balls where $p \in \{2, +\infty\}$.

As already mentioned in Section 2.1, various algorithms can be used to solve non-smooth convex optimization problems and would potentially benefit from the proposed epigraphical projection technique. In this work, we will employ a primal-dual algorithm, namely the Monotone+Lipschitz Forward Backward Forward (M+LFBF) algorithm, which was recently proposed in [68]. This proximal algorithm is able to address a wide class of convex optimization problems without requiring any matrix inversion and it offers a good performance and robustness to numerical errors. Its convergence is guaranteed (under weak conditions) and its structure makes it suitable for implementation on highly parallel architectures.

4.0.1 Degradation model

Set $\mathcal{H} = \mathbb{R}^{\overline{N}}$. Denote by $\overline{x} = (\overline{x}^{(n)})_{1 \leq n \leq \overline{N}} \in \mathbb{R}^{\overline{N}}$ the signal of interest, and by $z \in \mathbb{R}^N$ an observation vector such that $z = A\overline{x} + b$. It is assumed that $A \in \mathbb{R}^{N \times \overline{N}}$ is a linear operator, and $b \in \mathbb{R}^N$ is a realization of a zero-mean white Gaussian noise. The recovery of $\overline{x}$ from the degraded observations is performed by following a variational approach which aims at solving the following problem

$$\underset{x \in [\underline{\mu}, \overline{\mu}]^{\overline{N}}}{\text{minimize}} \ \|Ax - z\|^2 \quad \text{s. t.} \quad \sum_{\ell=1}^{L} \|\Omega_\ell \, B_\ell \, F \, x\|_p \leq \eta, \tag{37}$$

where $(\underline{\mu}, \overline{\mu}) \in \mathbb{R}^2$ with $\underline{\mu} \leq \overline{\mu}$, η is a real positive constant, and $F \in \mathbb{R}^{K \times \overline{N}}$ is the linear operator associated with an analysis transform. Furthermore, for every $\ell \in \{1, \ldots, L\}$, $B_\ell \in \mathbb{R}^{M^{(\ell)} \times K}$ is a *block-selection linear operator* which selects a block of $M^{(\ell)}$ data from its input vector.[3] For every $\ell \in \{1, \ldots, L\}$, Ω_ℓ denotes an $M^{(\ell)} \times M^{(\ell)}$ diagonal matrix of real positive weights.

The term $\|Ax - z\|^2$ is the *data fidelity* corresponding to the minus log-likelihood of x. The bounds $\underline{\mu}$ and $\overline{\mu}$ allow us to take into account the *value range* of each component of $\overline{x}$.

The second constraint involved in Problem (37) conveys a *smoothness* condition. It exploits the fact that natural signals usually exhibit a smooth spatial behaviour, except around some locations (e.g. object edges in images) where discontinuities arise. The proposed constraint is based on a weighted $\boldsymbol{\ell}_{1,p}$-norm, where $p \in \{2, +\infty\}$. The associated block-sparsity measure extends many of the smoothness terms used in the literature. It reduces to the weighted $\boldsymbol{\ell}_1$-norm criterion found in [80] when each block reduces to a singleton (i.e. $L = K$, and, for every $\ell \in \{1, \ldots, L\}$, $M^{(\ell)} = 1$ and $B_\ell \, y = y^{(\ell)}$).[4] It captures the $\boldsymbol{\ell}_{1,2}$ criteria present in [75, 81, 82, 83] when $p = 2$. It matches the $\boldsymbol{\ell}_{1,\infty}$ criterion proposed in [78] when $p = +\infty$.

Let us define $\Lambda = [B_1^\top \Omega_1, \ \ldots, B_L^\top \Omega_L]^\top$ and

$$C_2 = \big\{ y \in \mathbb{R}^M \ \big| \ \sum_{\ell=1}^{L} \|\mathrm{y}^{(\ell)}\|_p \leq \eta \big\} \tag{38}$$

[3] This means that there exist distinct indices $m_1, \ldots, m_{M^{(\ell)}}$ in $\{1, \ldots, K\}$ such that, for every $y = (y^{(k)})_{1 \leq k \leq K} \in \mathbb{R}^K$, $B_\ell y = (y^{(m_j)})_{1 \leq j \leq M^{(\ell)}}$.

[4] In this case, the constraint in (37) reduces to a sum of absolute values.

where $M = M^{(1)} + \cdots + M^{(L)}$ and the same decomposition as in (24) is performed. Then, it can be observed that Problem (37) is a particular case of Problem 1.1 where $S = 2$, $R = 1$, $g_1 = \|\cdot - z\|^2$, $T_1 = A$, $H_1 = I$, $C_1 = [\underline{\mu}, \overline{\mu}]^{\overline{N}}$, $H_2 = \Lambda F$, and C_2 is the above $\ell_{1,p}$-ball.

4.0.2 Algorithmic solution

The main difficulty in solving Problem (37) stems from the second constraint. The point is that most of the applicable algorithms require to compute the projection onto C_2. Specific numerical methods [84, 85, 78, 86] have been developed for this purpose. The aim of this section is to propose an alternative method based on the splitting principle presented in Section 3. So doing, the resulting problem can be efficiently addressed by proximal algorithms. The two possible approaches are now detailed.

- *Epigraphical method* – The principle of this method is to decompose C_2 into the closed half-space defined by (27) and the closed convex set defined by (28) with $h^{(\ell)} = \|\cdot\|_p$. The advantage of this decomposition is that the projections onto V and E have closed forms. Indeed, P_V is given by Proposition 2.1, while P_E is given by Proposition 3.1 for $p = 2$ and Proposition 3.2 for $p = +\infty$. We are then able to solve Problem (37) by means of the M+LFBF algorithm. The associated iterations are given in Algorithm 1.

Algorithm 1 Epigraphical method: M+LFBF for solving Problem (37)

Initialization
$(v^{[0]}, \nu^{[0]}) \in \mathbb{R}^M \times \mathbb{R}^L$
$(x^{[0]}, \zeta^{[0]}) \in \mathbb{R}^{\overline{N}} \times \mathbb{R}^L$
$\theta = 2\|A\|^2 + \max\{\|\Lambda F\|, 1\}$
$\epsilon \in]0, \frac{1}{\theta+1}[$
For $i = 0, 1, \ldots$
$\gamma_i \in [\epsilon, \frac{1-\epsilon}{\theta}]$
$(\widehat{x}^{[i]}, \widehat{\zeta}^{[i]}) = (x^{[i]}, \zeta^{[i]}) - \gamma_i \big(2A^\top (Ax^{[i]} - z) + F^\top \Lambda^\top v^{[i]}\ ,\ \nu^{[i]}\big)$
$(p^{[i]}, \rho^{[i]}) = \big(P_{[\underline{\mu}, \overline{\mu}]^{\overline{N}}}(\widehat{x}^{[i]}), P_V(\widehat{\zeta}^{[i]})\big)$
$(\widehat{v}^{[i]}, \widehat{\nu}^{[i]}) = (v^{[i]}, \nu^{[i]}) + \gamma_i (\Lambda F\, x^{[i]}, \zeta^{[i]})$
$(a^{[i]}, \alpha^{[i]}) = (\widehat{v}^{[i]}, \widehat{\nu}^{[i]}) - \gamma_i\, P_E\big(\gamma_i^{-1}\, \widehat{v}^{[i]},\ \gamma_i^{-1}\, \widehat{\nu}^{[i]}\big)$
$(\widetilde{v}^{[i]}, \widetilde{\nu}^{[i]}) = (a^{[i]}, \alpha^{[i]}) + \gamma_i (\Lambda F\, p^{[i]}, \rho^{[i]})$
$(v^{[i+1]}, \nu^{[i+1]}) = (v^{[i]} - \widehat{v}^{[i]} + \widetilde{v}^{[i]}, \nu^{[i]} - \widehat{\nu}^{[i]} + \widetilde{\nu}^{[i]})$
$(\widetilde{x}^{[i]}, \widetilde{\zeta}^{[i]}) = (p^{[i]}, \rho^{[i]}) - \gamma_i \big(2A^\top (Ap^{[i]} - z) + F^\top \Lambda^\top a^{[i]}, \alpha^{[t]}\big)$
$(x^{[i+1]}, \zeta^{[i+1]}) = (x^{[i]} - \widehat{x}^{[i]} + \widetilde{x}^{[i]}, \zeta^{[i]} - \widehat{\zeta}^{[i]} + \widetilde{\zeta}^{[i]})$

- *Direct method* – For completeness, we also consider the projection onto C_2 with the algorithm in [84] when $p = 2$,[5] or the iterative algorithm in [78] when $p = +\infty$.[6] Then, M+LFBF can still be used to solve Problem (37).

Note that, according to the general result in [68, Theorem 4.2], the sequence $\big(x^{[i]}\big)_{i\in\mathbb{N}}$ generated by Algorithm 1 is guaranteed to converge to a (global) minimizer of Problem (37).

[5] Code available at `www.cs.ubc.ca/~mpf/spgl1`
[6] Code available at `www.lsi.upc.edu/~aquattoni`

4.0.3 Smoothness constraint

We consider two smoothness constraints based on recent Non-Local TV measures [81, 87]. They constitute particular cases of the one considered in (37) when $L = \overline{N}$. They are described for 2D data in the following.

- $\boldsymbol{\ell_2}$*-Non-Local TV* – This constraint has the form

$$\sum_{\ell=1}^{\overline{N}} \Big(\sum_{n \in \mathcal{N}_\ell \subset \mathcal{W}_\ell} \omega_{\ell,n} (x^{(\ell)} - x^{(n)})^2 \Big)^{1/2} \leq \eta, \tag{39}$$

  where $\mathcal{N}_\ell$ is the *neighbourhood support* of ℓ, and $\mathcal{W}_\ell$ is the set of positions $n \in \{1, \ldots, \overline{N}\} \setminus \{\ell\}$ located into a $Q \times Q$ window centered at ℓ, where $Q \in \mathbb{N}$ is odd. This constraint is a particular case of the one considered in (37) where $K = (Q^2 - 1)\overline{N}$ and F is a concatenation of discrete difference operators F_{q_1,q_2} with $(q_1, q_2) \in \{-(Q-1)/2, \ldots, (Q-1)/2\}^2 \setminus \{(0,0)\}$. More precisely, for every (q_1, q_2), F_{q_1,q_2} is a 2D filter with impulse response: for every $(n_1, n_2) \in \mathbb{Z}^2$,

$$f_{q_1,q_2}^{(n_1,n_2)} = \begin{cases} 1, & \text{if } n_1 = n_2 = 0, \\ -1, & \text{if } n_1 = q_1 \text{ and } n_2 = q_2, \\ 0, & \text{otherwise.} \end{cases} \tag{40}$$

  In addition, for every $\ell \in \{1, \ldots, \overline{N}\}$, $M^{(\ell)} \leq Q^2 - 1$, B_ℓ selects the components of Fx corresponding to differences $(x^{(\ell)} - x^{(n)})_{n \in \mathcal{N}_\ell}$, and the positive weights $(\omega_{\ell,n})_{n \in \mathcal{N}_\ell}$ are gathered in the diagonal matrix Ω_ℓ.

- $\boldsymbol{\ell_\infty}$*-Non-Local TV*– We consider the following constraint

$$\sum_{\ell=1}^{\overline{N}} \max_{n \in \mathcal{N}_\ell} \left\{ \omega_{\ell,n} \, |x^{(\ell)} - x^{(n)}| \right\} \leq \eta. \tag{41}$$

  We proceed similarly to the previous constraint, except that the $\boldsymbol{\ell_\infty}$-norm is now substituted for the $\boldsymbol{\ell_2}$-norm.

Note that the classical isotropic TV constraint (designated by $\boldsymbol{\ell_2}$-TV in the following) constitutes a particular case of the $\boldsymbol{\ell_2}$-NLTV one, where each neighbourhood $\mathcal{N}_\ell$ only contains the horizontal/vertical neighbouring pixels ($M^{(\ell)} = 2$) and the weights are $\omega_{\ell,n} \equiv 1$. Similarly, the $\boldsymbol{\ell_\infty}$-TV constraint is a special case of the $\boldsymbol{\ell_\infty}$-NLTV one.

4.0.4 Weight estimation and neighbourhood choice

To set the weights, we got inspired from the Non-Local Means approach originally described in [88]. Here, for every $\ell \in \{1, \ldots, \overline{N}\}$ and $n \in \mathcal{N}_\ell$, the weight $\omega_{\ell,n}$ depends on the similarity between patches built around the pixels ℓ and n of the image. Since our degradation process involves some missing data, a two-step approach has been adopted. In the first step, the $\boldsymbol{\ell_2}$-TV approach is used in order to obtain an estimate $\widetilde{x}$ of the target image. This estimate is subsequently used in the second step to compute the weights through a *self-similarity* measure, yielding

$$\omega_{\ell,n} = \widetilde{\omega}_\ell \exp\left(-\delta^{-2} \, \|\widetilde{B}_\ell \widetilde{F}_\ell \widetilde{x} - \widetilde{B}_n \widetilde{F}_n \widetilde{x}\|^2 \right), \tag{42}$$

where $\delta \in \mathbb{R} \setminus \{0\}$, $\widetilde{\omega}_\ell \in]0, +\infty[$, $\widetilde{B}_\ell$ (resp. $\widetilde{B}_n$) selects a $\widetilde{Q} \times \widetilde{Q}$ patch centered at position ℓ (resp. n) and $\widetilde{F}_\ell$ (resp. $\widetilde{F}_n$) is a linear processing of the image depending on the position ℓ (resp. n). The constant $\widetilde{\omega}_\ell$ is set so as to normalize the weights (i.e. $\sum_{n \in \mathcal{N}_\ell} \omega_{\ell,n} = 1$).

The measure in (42) generalizes the one proposed in [88], which corresponds to the case when $\widetilde{F}_\ell$ (resp. $\widetilde{F}_n$) reduces to a Gaussian function with mean ℓ (resp. n). In the present work, we consider the *foveated self-similarity* measure recently introduced in [89], due to its better performance in denoising. This approach can be derived from (42) by setting $\widetilde{F}_\ell$ (resp. $\widetilde{F}_n$) to a set of low-pass Gaussian filters whose variances increase as the spatial distance from the patch center ℓ (resp. n) grows.

For every $\ell \in \{1, \ldots, \overline{N}\}$, the neighbourhood $\mathcal{N}_\ell$ is built according to the procedure described in [90]. In practice, we limit the size of the neighbourhood, so that $M^{(\ell)} \leq \overline{M}$.

4.0.5 Simulation results

The purpose of this section is twofold. First, the quality of images reconstructed with our variational approach is evaluated for different choices of regularization constraints and comparisons are made with a state-of-the-art method. Secondly, the convergence speed of the proposed epigraphical technique is evaluated w.r.t. the direct methods.

In the following experiments, if not specified otherwise, the degradation matrix A is a decimated convolution which consists of a 3×3 uniform blur followed by a decimation which randomly deletes 60% of the pixels ($N = 0.4 \times \overline{N}$). The standard deviation of the additive white Gaussian noise is equal to $\sigma = 10$. Since we deal with natural images, the data range bounds are $\underline{\mu} = 0$ and $\overline{\mu} = 255$. For the smoothness constraint, we set $Q = 11$, $\widetilde{Q} = 5$, $\delta = 35$ and $\overline{M} = 14$.

Extensive tests have been carried out on several standard images of different sizes. The SNR and SSIM [91] results obtained by using the various previously introduced TV-like constraints are collected in Table 1. In addition, a comparison is performed between our method and the Gradient Projection for Sparse Reconstruction (GPSR) method [60], which also relies on a variational approach. The constraint bound for both methods was hand-tuned in order to achieve the best SNR values. The best results are highlighted in bold. A visual comparison is made in Figure 1, where two representative images are displayed. These results demonstrate the interest of considering non-local smoothness measures. Non-Local TV with $\boldsymbol{\ell}_{1,2}$-norm indeed proves to be the most effective constraint with gains in SNR and SSIM (up to 1.82 dB and 0.042) with respect to $\boldsymbol{\ell}_2$-TV, which in turn outperforms GPSR. The better performance of NLTV seems to be related to its ability to better preserve edges and thin structures present in images. In terms of computational time, GPSR is about twice faster than $\boldsymbol{\ell}_2$-NLTV. Our codes were developed in MATLAB[7], the operators F and $F^\top$ being implemented in C using mex files.

In order to complete the analysis, we report in Figure 2 SNR/SSIM comparisons between $\boldsymbol{\ell}_2$-NLTV and $\boldsymbol{\ell}_2$-TV for different blur and noise configurations. These plots show that $\boldsymbol{\ell}_2$-NLTV provides better results regardless of the degradation conditions.

In the second part of the section, we focus on the convergence speed of epigraph and direct methods. All the results refer to the *culicoidae* image cropped at 256×256 ($\overline{N} = 256^2$), since a similar behaviour was observed for other images. The stopping criterion is set to: $\|x^{[i+1]} - x^{[i]}\| \leq 10^{-4} \|x^{[i]}\|$. For the $\boldsymbol{\ell}_{1,p}$-ball projectors needed by the direct method, we used the software publicly available on-line [84, 78].

[7] R2011b version on an Intel Xeon CPU at 2.80 GHz and 8 GB of RAM.

Table 1: $\mathrm{SNR}_{\mathrm{dB}}$ and SSIM results of our method and GPSR (noise parameters: blur = 3×3, $\sigma = 10$, decimation = 60%)

SNR (dB) – SSIM	$\overline{N}$	ℓ_2-TV	ℓ_∞-TV	ℓ_2-NLTV	ℓ_∞-NLTV	GPSR
Culicoidae	256^2	20.80 – 0.855	20.25 – 0.853	**22.62 – 0.897**	22.38 – 0.897	17.03 – 0.738
Lena	256^2	23.18 – 0.783	22.77 – 0.769	**24.18 – 0.812**	24.14 – 0.812	20.26 – 0.678
Cameraman	256^2	20.06 – 0.774	19.68 – 0.755	**20.71 – 0.801**	20.17 – 0.743	17.92 – 0.673
Boat	256^2	20.25 – 0.739	19.74 – 0.718	**21.13 – 0.770**	20.77 – 0.741	18.06 – 0.649
House	256^2	25.47 – 0.823	24.70 – 0.808	**26.31 – 0.836**	25.87 – 0.823	22.14 – 0.734
Man	256^2	19.24 – 0.725	18.96 – 0.714	**19.66 – 0.741**	19.51 – 0.736	17.11 – 0.629
Peppers	512^2	23.69 – 0.801	23.25 – 0.786	**24.80 – 0.829**	24.45 – 0.813	21.94 – 0.709
Barbara	512^2	16.74 – 0.653	16.64 – 0.642	**17.02 – 0.673**	16.99 – 0.652	15.97 – 0.562
Hill	512^2	22.18 – 0.723	21.89 – 0.715	**22.55 – 0.735**	22.43 – 0.733	20.21 – 0.637
Culicoidae	1024^2	20.84 – 0.855	20.49 – 0.812	**23.25 – 0.885**	22.57 – 0.810	17.19 – 0.725

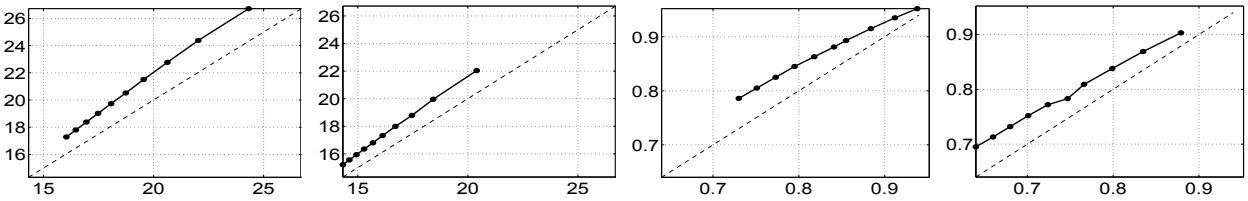


(a) Culicoidae. (b) Degraded. (c) Zoom. (d) GPSR, SNR: 17.03 dB. (e) Lena. (f) Degraded. (g) Zoom. (h) GPSR, SNR: 20.26 dB.



(i) ℓ_2-TV, SNR: 20.80 dB. (j) ℓ_∞-TV, SNR: 20.25 dB. (k) ℓ_2-NLTV, SNR: **22.62 dB**. (l) ℓ_∞-NLTV, SNR: 22.38 dB. (m) ℓ_2-TV, SNR: 23.18 dB. (n) ℓ_∞-TV, SNR: 22.77 dB. (o) ℓ_2-NLTV, SNR: **24.18 dB**. (p) ℓ_∞-NLTV, SNR: 24.14 dB.

Figure 1: Image restoration examples (noise parameters: blur = 3×3, $\sigma = 10$, decimation = 60%)



(a) SNR comparison (3×3 blur). (b) SNR comparison (5×5 blur). (c) SSIM comparison (3×3 blur). (d) SSIM comparison (5×5 blur).

Figure 2: SNR and SSIM values for ℓ_2-NLTV (vertical axes) and ℓ_2-TV (horizontal axes), for the *culicoidae* image. The plots show the results obtained for $\sigma \in \{5, 10, \ldots, 50\}$, where lower SNR or SSIM values correspond to higher σ values. No decimation is applied in this experiment.

- *Total Variation* – Tables 2 and 3 report a comparison between the direct and epigraphical methods for different values of η and for $\boldsymbol{\ell}_2$-TV and $\boldsymbol{\ell}_\infty$-TV, respectively. For more readability, these are expressed as a multiplicative factor of the $\boldsymbol{\ell}_p$-TV-semi-norm of the original image. It can be noticed that the parameter η influences both the quality of the results and the convergence speed. For the $\boldsymbol{\ell}_{1,2}$-norm, the epigraphical method is 3.5 faster than the direct one. For the $\boldsymbol{\ell}_{1,\infty}$-norm, it is 65 times faster.

  Figs. 3-a and 3-b show the relative error $\|x^{[i]} - x^{[\infty]}\|/\|x^{[\infty]}\|$ as a function of the computational time, where $x^{[\infty]}$ denotes the solution computed after a large number of iterations (typically, 5000 iterations). The dashed line presents the results for the direct method while the solid line refers to the epigraphical one. These plots show that the epigraphical approach is faster despite 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.

Table 2: Results for the $\boldsymbol{\ell}_2$-TV constraint and different values of η

η	SNR (dB) – SSIM	DIRECT		EPIGRAPH.		SPEED UP
		# iter.	sec.	# iter.	sec.	
0.6	19.70 – 0.839	116	5.74	146	2.23	2.58
0.7	20.62 – 0.862	132	7.14	151	2.29	3.11
0.8	**20.80 – 0.855**	**160**	**9.17**	**171**	**2.59**	**3.54**
0.9	20.45 – 0.826	195	11.95	196	2.98	4.01

Table 3: Results for the $\boldsymbol{\ell}_\infty$-TV constraint and different values of η

η	SNR (dB) – SSIM	DIRECT		EPIGRAPH.		SPEED UP
		# iter.	sec.	# iter.	sec.	
0.6	19.58 – 0.839	195	373.84	230	6.21	60.22
0.7	**20.25 – 0.853**	**206**	**417.71**	**231**	**6.39**	**65.34**
0.8	20.24 – 0.835	233	486.17	245	6.86	70.82
0.9	19.80 – 0.800	263	551.03	267	7.44	74.06

- $\boldsymbol{\ell}_2$*-Non-Local Total Variation* – Table 4 reports the results for $\boldsymbol{\ell}_2$-NLTV. Different combinations of neighbourhood size Q and bound value η are considered. To set the weights, the first TV estimate is computed with $\eta = 0.8$. The best SNR improvement (1.82 dB over $\boldsymbol{\ell}_2$-TV) is observed when a relatively small neighbourhood is used ($Q = 11$). This behaviour was already observed in [90]. It may be related to the bias-variance trade-off commonly encountered in Non-Local filters [92, 93].

  In Figure 3-c, a plot similar to those in Figs. 3-a and 3-b show the convergence profile. The epigraphical method requires about the same number of iterations as the direct one in order to converge. This results in a time reduction (1.8 times), as a single iteration of the epigraphical method is faster than one iteration of the direct method.

- $\boldsymbol{\ell}_\infty$*-Non-Local Total Variation* – Table 5 and Figure 3-d show the results obtained with the $\boldsymbol{\ell}_\infty$-NLTV constraint. Note that $\boldsymbol{\ell}_\infty$-NLTV requires more iterations than $\boldsymbol{\ell}_2$-NLTV in order to converge. In what concerns the convergence speed, the epigraphical method is 19 times faster.

Table 4: Results for the $\boldsymbol{\ell}_2$-NLTV constraint and some values of η and Q

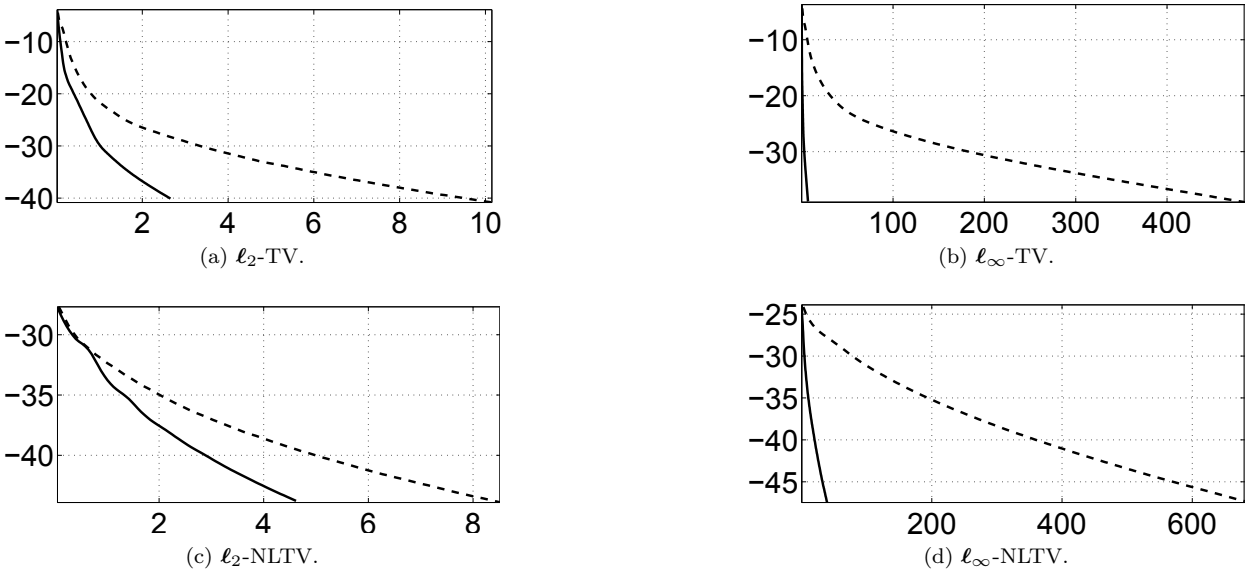
η	SNR (dB) – SSIM	DIRECT		EPIGRAPH.		SPEED UP
		# iter.	sec.	# iter.	sec.	
Neighbourhood size: $Q = 3$						
0.7	22.26 – 0.895	70	5.82	77	3.07	1.90
0.8	22.41 – 0.893	72	6.39	75	3.00	2.13
0.9	22.06 – 0.875	88	7.97	89	3.58	2.23
Neighbourhood size: $Q = 5$						
0.7	22.44 – 0.900	70	7.13	74	4.37	1.63
0.8	22.58 – 0.898	72	7.85	75	4.44	1.77
0.9	22.25 – 0.880	87	9.71	88	5.24	1.85
Neighbourhood size: $Q = 11$						
0.7	22.50 – 0.901	72	7.52	76	4.51	1.67
0.8	**22.61 – 0.897**	**75**	**8.21**	**78**	**4.64**	**1.77**
0.9	22.27 – 0.879	89	9.83	91	5.41	1.82

Table 5: Results for the $\boldsymbol{\ell}_\infty$-NLTV constraint and some values of η and Q

η	SNR (dB) – SSIM	DIRECT		EPIGRAPH.		SPEED UP
		# iter.	sec.	# iter.	sec.	
Neighbourhood size: $Q = 5$						
0.6	21.78 – 0.889	229	766.21	251	25.61	29.92
0.7	22.28 – 0.894	213	653.31	226	22.87	28.56
0.8	22.10 – 0.876	207	596.90	216	21.89	27.26
Neighbourhood size: $Q = 5$						
0.6	21.92 – 0.894	231	914.03	252	43.32	21.10
0.7	**22.38 – 0.897**	**219**	**749.32**	**232**	**39.73**	**18.86**
0.8	22.21 – 0.880	212	673.60	224	38.05	17.70
Neighbourhood size: $Q = 11$						
0.6	21.95 – 0.894	236	940.38	256	43.75	21.49
0.7	22.38 – 0.897	222	760.49	234	39.84	19.09
0.8	22.20 – 0.879	216	689.07	227	38.80	17.76

5 Conclusions

We have proposed a new epigraphical technique to solve constrained convex optimization problems with the help of proximal algorithms. In this paper (Part I), our attention has been turned to block-sparsity constraints based on weighted $\boldsymbol{\ell}_{1,p}$-norms with $p \in \{2, +\infty\}$. The obtained results demonstrate the better performance of non-local measures in terms of image quality. Our results also show that the $\boldsymbol{\ell}_{1,2}$-norm has to be preferred over the $\boldsymbol{\ell}_{1,\infty}$-norm for image recovery problems. However, it would be interesting to consider alternative applications of $\boldsymbol{\ell}_{1,\infty}$-norms such as regression problems [94, 95]. 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. Full implementations in C of the proposed algorithms and parallelization of our codes should even allow us to accelerate them [96]. Note that, although the considered application involves two constraint sets, the proposed approach can handle an arbitrary number of convex constraints. In a companion paper (Part II), we will show that the epigraphical approach can also be used to develop approximation methods for dealing with more general convex constraints.



(a) ℓ_2-TV. (b) ℓ_∞-TV. (c) ℓ_2-NLTV. (d) ℓ_∞-NLTV.

Figure 3: Comparison between epigraphical method (*solid line*) and direct method (*dashed line*): $\frac{\|x^{[i]}-x^{[\infty]}\|}{\|x^{[\infty]}\|}$ in dB vs time

A Proof of the Proposition 2.4

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

$$\begin{aligned}(\forall u \in \mathcal{H}) \quad 0 &= \frac{1}{2}\|p-y\|^2 + \frac{1}{2}\big(\max\{\varphi(p)-\zeta,0\}\big)^2 \\ &\leq \frac{1}{2}\|u-y\|^2 + \frac{1}{2}\big(\max\{\varphi(u)-\zeta,0\}\big)^2, \end{aligned} \tag{43}$$

which shows that (12) holds. Let us now consider the case when $\varphi(y) > \zeta$. From the definition of the projection, we get

$$(p,\theta) = \operatorname*{argmin}_{(u,\xi)\in\operatorname{epi}\varphi} \|u-y\|^2 + (\xi-\zeta)^2. \tag{44}$$

From the Karush-Kuhn-Tucker theorem [97, Theorem 5.2],[8] there exists $\alpha \in [0,+\infty[$ such that

$$(p,\theta) = \operatorname*{argmin}_{(u,\xi)\in\mathcal{H}\times\mathbb{R}} \frac{1}{2}\|u-y\|^2 + \frac{1}{2}(\xi-\zeta)^2 + \alpha(\varphi(u)-\xi) \tag{45}$$

where the Lagrange multiplier α is such that

$$\alpha(\varphi(p)-\theta) = 0. \tag{46}$$

Since the value $\alpha = 0$ is not allowable (since 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 (45) w.r.t. ξ yields

$$\varphi(p) = \theta = \zeta + \alpha \geq \zeta. \tag{47}$$

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

Hence, (p, θ) given by (44) is such that

$$p = \underset{\substack{u\in\mathcal{H}\\ \varphi(u)\geq\zeta}}{\operatorname{argmin}} \|u-y\|^2 + (\varphi(u)-\zeta)^2 \tag{48}$$

$$\theta = \varphi(p) = \max\{\varphi(p), \zeta\}. \tag{49}$$

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

$$\inf_{\substack{u\in\mathcal{H}\\ \varphi(u)\leq\zeta}} \|u-y\|^2 = \|P_{\operatorname{lev}_{\leq\zeta}\varphi}(y) - y\|^2 = \inf_{\substack{u\in\mathcal{H}\\ \varphi(u)=\zeta}} \|u-y\|^2 \tag{50}$$

where we have used the fact that $P_{\operatorname{lev}_{\leq\zeta}\varphi}(y)$ belongs to the boundary of $\operatorname{lev}_{\leq\zeta}\varphi$ which is equal to $\big\{u\in\mathcal{H} \;\big|\; \varphi(u)=\zeta\big\}$ since φ is lower-semicontinuous and $\operatorname{dom}\varphi$ is open [73, Corollary 8.38]. We have then

$$\begin{aligned}\inf_{\substack{u\in\mathcal{H}\\ \varphi(u)\leq\zeta}} \|u-y\|^2 &= \inf_{\substack{u\in\mathcal{H}\\ \varphi(u)=\zeta}} \|u-y\|^2 \\ &\geq \inf_{\substack{u\in\mathcal{H}\\ \varphi(u)\geq\zeta}} \|u-y\|^2 + (\varphi(u)-\zeta)^2. \end{aligned}\tag{51}$$

Altogether, (48) and (51) lead to

$$p = \underset{u\in\mathcal{H}}{\operatorname{argmin}} \frac{1}{2}\|u-y\|^2 + \frac{1}{2}\big(\max\{\varphi(u)-\zeta, 0\}\big)^2 \tag{52}$$

which is equivalent to (12) since $\frac{1}{2}\big(\max\{\varphi-\zeta,0\}\big)^2 \in \Gamma_0(\mathcal{H})$.

B Proof of the Proposition 2.5

Since $(\max\{\varphi-\zeta,0\})^2$ is an even function, $\operatorname{prox}_{\frac{1}{2}(\max\{\varphi-\zeta,0\})^2}$ is an odd function [23, Remark 4.1(ii)]. In the following, we thus focus on the case when $y \in]0,+\infty[$.
If $\zeta \in]-\infty, 0]$, then $(\max\{\varphi-\zeta,0\})^2 = (\varphi-\zeta)^2$. When $\beta = 1$, $(\max\{\varphi-\zeta,0\})^2 = \tau^2(\cdot)^2 - 2\tau\zeta|\cdot| + \zeta^2$ and, from Example 2.3, it can be deduced that

$$\operatorname{prox}_{\frac{1}{2}(\max\{\varphi-\zeta,0\})^2}(y) = \operatorname{prox}_{-\frac{\tau\zeta}{1+\tau^2}|\cdot|}\Big(\frac{y}{1+\tau^2}\Big) = \frac{1}{1+\tau^2}\max\{y+\tau\zeta, 0\}. \tag{53}$$

When $\beta > 1$, $(\varphi-\zeta)^2$ is differentiable and, according to (8), $p = \operatorname{prox}_{\frac{1}{2}(\max\{\varphi-\zeta,0\})^2}(y)$ is uniquely defined as

$$p - y + \beta\tau p^{\beta-1}(\tau p^\beta - \zeta) = 0 \tag{54}$$

where, according to [98, Corollary 2.5], $p \geq 0$. This allows us to deduce that $p = \chi_0$.
Let us now focus on the case when $\zeta \in]0,+\infty[$. If $y \in]0, (\zeta/\tau)^{1/\beta}[$, it can be deduced from [98, Corollary 2.5], that $p = \operatorname{prox}_{\frac{1}{2}(\max\{\varphi-\zeta,0\})^2}(y) \in [0, (\zeta/\tau)^{1/\beta}[$. Since $(\forall v \in [0, (\zeta/\tau)^{1/\beta}[)\ \max\{\varphi(v)-\zeta, 0\} = 0$, (8) yields $p = y$. On the other hand if $y > (\zeta/\tau)^{1/\beta}$, as the proximity operator of a function from $\mathbb{R}$ to $\mathbb{R}$ is continuous and increasing [98, Proposition 2.4], $p = \operatorname{prox}_{\frac{1}{2}(\max\{\varphi-\zeta,0\})^2}(y) \geq \operatorname{prox}_{\frac{1}{2}(\max\{\varphi-\zeta,0\})^2}\big((\zeta/\tau)^{1/\beta}\big) = (\zeta/\tau)^{1/\beta}$. Since $(\max\{\varphi-\zeta,0\})^2$ is differentiable in this case, and $(\forall v \geq (\zeta/\tau)^{1/\beta})\ (\max\{\varphi(v)-\zeta,0\})^2 = (\tau v^\beta - \zeta)^2$, (8) allows us to deduce that p is the unique value in $[(\zeta/\tau)^{1/\beta}, +\infty[$ satisfying (54). It can be concluded that, when $\zeta \in]0,+\infty[$, (16) holds.

C Proof of the Proposition 2.6

Let us notice that $\frac{1}{2}(\max\{\tau d_C^\beta - \zeta, 0\})^2 = \psi \circ d_C$ where $\psi = \frac{1}{2}(\max\{\tau|\cdot|^\beta - \zeta, 0\})^2$. According to [25, Proposition 2.7], for every $y \in \mathcal{H}$,

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

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

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

and, according to Proposition 2.5, when $\zeta < 0$, $\beta = 1$ and $d_C(y) \leq -\tau\zeta$, $\operatorname{prox}_\psi\left(d_C(y)\right) = 0$. These show that (55) reduces to (17).

D Proof of the Corollary 2.7

As we have $\varphi = \tau d_C$ where $C = \{z\}$, the result follows from Proposition 2.6 and the expression of $\operatorname{prox}_{\frac{1}{2}(\max\{\tau|\cdot|-\zeta,0\})^2}$ in Proposition 2.5.

E Proof of the Proposition 2.8

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

$$\nu^{(k-1)} < v \leq \nu^{(k)} \qquad \Rightarrow \qquad \varphi(v) = \frac{1}{2}\sum_{m=1}^{k-1}(\tau_-^{(m)})^2\left(v - \nu^{(m)}\right)^2 + \frac{1}{2}\sum_{m=k}^{M}(\tau_+^{(m)})^2\left(v - \nu^{(m)}\right)^2. \tag{57}$$

For every $y \in \mathbb{R}$, as $p = \operatorname{prox}_\varphi(y)$ is characterized by (8), there exists $\overline{m} \in \{1, \ldots, M+1\}$ such that $\nu^{(\overline{m}-1)} < p \leq \nu^{(\overline{m})}$ and

$$y - p = \sum_{m=1}^{\overline{m}-1}(\tau_-^{(m)})^2(p - \nu^{(m)}) + \sum_{m=\overline{m}}^{M}(\tau_+^{(m)})^2(p - \nu^{(m)}). \tag{58}$$

This yields (21), and we have : (22) $\Leftrightarrow \nu^{(\overline{m}-1)} < p \leq \nu^{(\overline{m})}$. The uniqueness of $\overline{m} \in \{1, \ldots, M+1\}$ satisfying this inequality follows from the uniqueness of $\operatorname{prox}_\varphi(y)$.

F Proof of the Proposition 3.2

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

$$\min_{\theta^{(\ell)} \in [0,+\infty[} \Big((\theta^{(\ell)} - \zeta^{(\ell)})^2 + \min_{\substack{|\mathsf{p}^{(\ell,1)}| \leq \tau^{(\ell,1)}\, \theta^{(\ell)} \\ \vdots \\ |\mathsf{p}^{(\ell,M^{(\ell)})}| \leq \tau^{(\ell,M^{(\ell)})}\, \theta^{(\ell)}}} \|\mathsf{p}^{(\ell)} - \mathsf{y}^{(\ell)}\|^2 \Big). \tag{59}$$

For every $\theta^{(\ell)} \in [0, +\infty[$, the inner minimization is achieved when, for every $j \in \{1, \ldots, M^{(\ell)}\}$, $\mathsf{p}^{(\ell,m)}$ is the projection of $\mathsf{y}^{(\ell,m)}$ onto $[-\tau^{(\ell,m)}\, \theta^{(\ell)}, \tau^{(\ell,m)}\, \theta^{(\ell)}]$, which is given by (34). Then, the problem reduces to

$$\underset{\theta^{(\ell)} \in \mathbb{R}}{\text{minimize}} \Big((\theta^{(\ell)} - \zeta^{(\ell)})^2 + \sum_{m=1}^{M^{(\ell)}} (\max\{|\mathsf{y}^{(\ell,m)}| - \tau^{(\ell,m)} \theta^{(\ell)}, 0\})^2 + \iota_{[0,+\infty[}(\theta^{(\ell)}) \Big) \tag{60}$$

which is also equivalent to calculate $\operatorname{prox}_{\varphi + \iota_{[0,+\infty[}}(\zeta^{(\ell)})$, where φ is such that

$$(\forall v \in \mathbb{R}) \qquad \varphi(v) = \frac{1}{2} \sum_{m=1}^{M^{(\ell)}} (\max\{\tau^{(\ell,m)}(\nu^{(\ell,m)} - v), 0\})^2. \tag{61}$$

By using now the fact that $\operatorname{prox}_{\varphi + \iota_{[0,+\infty[}} = P_{[0,+\infty[} \circ \operatorname{prox}_{\varphi}$ ([24, Proposition 12]) and by invoking Proposition 2.8, the expression of the optimal solution in (35) follows.

References

[1] L. Chaari, J.-C. Pesquet, A. Benazza-Benyahia, and Ph. Ciuciu, "A wavelet-based regularized reconstruction algorithm for SENSE parallel MRI with applications to neuroimaging," *Medical Image Analysis*, vol. 15, no. 2, pp. 185–201, Apr. 2011.

[2] 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.*, vol. 30, no. 9, pp. 1649–1660, Sep. 2011.

[3] G. Facciolo, A. Almansa, J.-F. Aujol, and V. Caselles, "Irregular to regular sampling, denoising, and deconvolution," *Multiscale Model. and Simul.*, vol. 7, no. 4, pp. 1574–1608, Avr. 2009.

[4] N. Hajlaoui, C. Chaux, G. Perrin, F. Falzon, and A. Benazza-Benyahia, "Satellite image restoration in the context of a spatially varying point spread function," *J. Opt. Soc. Am.*, vol. 27, no. 6, pp. 1473–1481, 2010.

[5] 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.*, vol. 18, no. 2, pp. 310–321, Feb. 2009.

[6] A. Jezierska, E. Chouzenoux, J.-C. Pesquet, and H. Talbot, "A primal-dual proximal splitting approach for restoring data corrupted with Poisson-Gaussian noise," in *Proc. Int. Conf. Acoust., Speech Signal Process.*, Kyoto, Japan, Mar., 25-30 2012, 4p.

[7] B. Vandeghinste, B. Goossens, J. De Beenhouwer, A. Pizurica, W. Philips, S. Vandenberghe, and S. Staelens, "Split-Bregman-based sparse view CT reconstruction," in *International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, Potsdam, Germany, Jul. 11-15 2011.

[8] N. Pustelnik, C. Chaux, J.-C. Pesquet, and C. Comtat, "Parallel algorithm and hybrid regularization for dynamic PET reconstruction," in *IEEE Med. Imag. Conf.*, Knoxville, Tennessee, Oct. 30 - Nov. 6 2010.

[9] S. Anthoine, J.-F. Aujol, Y. Boursier, and C. Mélot, "Some proximal methods for CBCT and PET tomography," in *Mathematical imaging in interaction with biomedicine*, Edinburgh, Scotland, Sept., 05-09 2011.

[10] J.-F. Aujol, G. Aubert, L. Blanc-Féraud, and A. Chambolle, "Image decomposition into a bounded variation component and an oscillating component," *J. Math. Imag. Vis.*, vol. 22, pp. 71–88, Jan. 2005.

[11] J.-F. Aujol, G. Gilboa, T. Chan, and S. Osher, "Structure-texture image decomposition - modeling, algorithms, and parameter selection," *Int. J. Comp. Vis.*, vol. 67, no. 1, pp. 111–136, Apr. 2006.

[12] L. M. Briceño-Arias, P. L. Combettes, J.-C. Pesquet, and N. Pustelnik, "Proximal algorithms for multicomponent image recovery problems," *J. Math. Imag. Vis.*, vol. 41, no. 1, pp. 3–22, Sep. 2011.

[13] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, "Optimization with sparsity-inducing penalties," *Foundations and Trends in Machine Learning*, vol. 4, no. 1, pp. 1–106, 2012.

[14] S. Theodoridis, K. Slavakis, and I. Yamada, "Adaptive learning in a world of projections," *IEEE Signal Process. Mag.*, vol. 28, no. 1, pp. 97–123, Jan. 2011.

[15] 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.*, 2012, To appear.

[16] M. Kowalski, E. Vincent, and R. Gribonval, "Beyond the narrowband approximation: Wideband convex methods for under-determined reverberant audio source separation," *IEEE Trans. Audio, Speech Language Process.*, vol. 18, no. 7, pp. 1818–1829, Sept. 2010.

[17] O. D. Akyildiz and I. Bayram, "An analysis prior based decomposition method for audio signals," in *Proc. Eur. Sig. and Image Proc. Conference*, Bucharest, Romania, Aug. 27-31 2012.

[18] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds., pp. 185–212. Springer-Verlag, New York, 2011.

[19] S. Setzer, G. Steidl, and T. Teuber, "Deblurring Poissonian images by split Bregman techniques," *J. Visual Communication and Image Representation*, vol. 21, no. 3, pp. 193–199, Apr. 2010.

[20] M. A. T. Figueiredo and J. M. Bioucas-Dias, "Restoration of Poissonian images using alternating direction optimization," *IEEE Trans. Image Process.*, vol. 19, no. 12, pp. 3133–3145, Dec. 2010.

[21] J.-C. Pesquet and N. Pustelnik, "A parallel inertial proximal optimization method," *Pac. J. Optim.*, vol. 8, no. 2, pp. 273–305, Apr. 2012.

[22] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Model. and Simul.*, vol. 4, no. 4, pp. 1168–1200, Nov. 2005.

[23] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs, "A variational formulation for frame-based inverse problems," *Inverse Problems*, vol. 23, no. 4, pp. 1495–1518, Jun. 2007.

[24] P. L. Combettes and J.-C. Pesquet, "A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery," *IEEE J. Selected Topics Signal Process.*, vol. 1, no. 4, pp. 564–574, Dec. 2007.

[25] P. L. Combettes and J.-C. Pesquet, "A proximal decomposition method for solving convex variational inverse problems," *Inverse Problems*, vol. 24, no. 6, Dec. 2008.

[26] N. P. Galatsanos and A. K. Katsaggelos, "Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation," *IEEE Trans. Image Process.*, vol. 1, no. 3, pp. 322–336, Jul. 1992.

[27] P. C. Hansen and D. P. O'Leary, "The use of the L-curve in the regularization of discrete ill-posed problems," *SIAM J. Sci. Comput.*, vol. 14, no. 6, pp. 1487–1503, 1993.

[28] A. Pizurica and W. Philips, "Estimating the probability of the presence of a signal of interest in multiresolution single- and multiband image denoising," *IEEE Trans. Image Process.*, vol. 15, no. 3, pp. 654–665, Mar. 2006.

[29] S. Ramani, T. Blu, and M. Unser, "Monte-Carlo SURE: A black-box optimization of regularization parameters for general denoising algorithms," *IEEE Trans. Image Process.*, vol. 17, no. 9, pp. 1540–1554, Sep. 2008.

[30] L. Chaari, J.-C. Pesquet, J.-Y. Tourneret, P. Ciuciu, and A. Benazza-Benyahia, "A hierarchical bayesian model for frame representation," *IEEE Trans. Signal Process.*, vol. 58, no. 11, pp. 5560–5571, Nov. 2010.

[31] D. C. Youla and H. Webb, "Image restoration by the method of convex projections. Part I - theory," *IEEE Trans. Med. Imag.*, vol. 1, no. 2, pp. 81–94, Oct. 1982.

[32] H. J. Trussell and M. R. Civanlar, "The feasible solution in signal restoration," *IEEE Trans. Acous., Speech Signal Process.*, vol. 32, no. 2, pp. 201–212, Apr. 1984.

[33] P. L. Combettes, "Inconsistent signal feasibility problems : least-squares solutions in a product space," *IEEE Trans. Signal Process.*, vol. 42, no. 11, pp. 2955–2966, Nov. 1994.

[34] K. Kose, V. Cevher, and A. E. Cetin, "Filtered variation method for denoising and sparse signal processing," in *Proc. Int. Conf. Acoust., Speech Signal Process.*, Kyoto, Japan, March, 27-30 2012.

[35] T. Teuber, G. Steidl, and R. H. Chan, "Minimization and parameter estimation for seminorm regularization models with I-divergence constraints," Tech. Rep., Technische Universität Kaiserslautern, 2012.

[36] R. Ciak, B. Shafei, and G. Steidl, "Homogeneous penalizers and constraints in convex image restoration," *J. Math. Imag. Vis.*, 2012.

[37] L. Jacques, L. Duval, C. Chaux, and G. Peyré, "A panorama on multiscale geometric representations, intertwining spatial, directional and frequency selectivity," *Signal Process.*, vol. 91, no. 12, pp. 2699–2730, Dec. 2011.

[38] M. A. T. Figueiredo and R. D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. Image Process.*, vol. 12, no. 8, pp. 906–916, Aug. 2003.

[39] 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.*, vol. 20, no. 3, pp. 681–695, Mar. 2011.

[40] A. Tikhonov, "Tikhonov regularization of incorrectly posed problems," *Soviet Mathematics Doklady*, vol. 4, pp. 1624–1627, 1963.

[41] E. Chouzenoux, J. Idier, and S. Moussaoui, "A Majorize-Minimize strategy for subspace optimization applied to image restoration," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1517–1528, Jun. 2011.

[42] G. Aubert and R. Tahraoui, "Sur la minimisation d'une fonctionnelle non convexe, non différentiable en dimension 1," *Bolletino UMI*, vol. 5, no. 17B, 1980.

[43] A. Ben-Tal and M. Teboulle, "A smoothing technique for nondifferentiable optimization problems," *Lecture Notes in Mathematics*, vol. 1405, pp. 1–11, 1989.

[44] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex analysis and minimization algorithms, Part I : Fundamentals*, vol. 305 of *Grundlehren der mathematischen Wissenschaften*, Springer-Verlag, Berlin, Heidelberg, N.Y., 2nd edition, 1996.

[45] P. Tseng, "Convergence of a block coordinate descent method for nondifferentiable minimization," *Journal of Opt. Theory and Applications*, vol. 109, no. 3, pp. 475–494, Jun. 2001.

[46] S. J. Wright, *Primal-dual interior-point methods*, SIAM, Philadelphia, PA, 1997.

[47] L. M. Bregman, "The method of successive projection for finding a common point of convex sets," *Soviet Math. Dokl.*, vol. 6, pp. 688–692, 1965.

[48] L. G. Gurin, B. T. Polyak, and E. V. Raik, "Projection methods for finding a common point of convex sets," *Zh. Vychisl. Mat. Mat. Fiz.*, vol. 7, no. 6, pp. 1211–1228, 1967.

[49] P. L. Combettes, "The foundations of set theoretic estimation," *Proceedings of the IEEE*, vol. 81, no. 2, pp. 182–208, Feb. 1993.

[50] P. L. Combettes, "Convex set theoretic image recovery by extrapolated iterations of parallel subgradient projections," *IEEE Trans. Image Process.*, vol. 6, no. 4, pp. 492–506, Apr. 1997.

[51] Y. Censor, W. Chen, P. L. Combettes, R. Davidi, and G. T. Herman, "On the effectiveness of projection methods for convex feasibility problems with linear inequality constraints," *Comput. Optim. Appl.*, vol. 51, no. 3, pp. 1065–1088, 2012.

[52] B. T. Polyak, "Minimization of unsmooth functionals," *USSR Computational Mathematics and Mathematical Physics*, vol. 9, no. 3, pp. 14–29, 1969.

[53] N. Z. Shor, *Minimization Methods for Non-differentiable Functions*, Springer-Verlag, New York, 1985.

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

[55] K. Slavakis, I. Yamada, and N. Ogura, "The adaptive projected subgradient method over the fixed point set of strongly attracting nonexpansive mappings," *Numerical Functional Analysis and Optimization*, vol. 27, no. 7-8, pp. 905–930, Nov. 2006.

[56] P. Bouboulis, K. Slavakis, and S. Theodoridis, "Adaptive learning in complex reproducing kernel Hilbert spaces employing Wirtinger's subgradients," *IEEE Trans Neural Networks*, vol. 23, no. 3, pp. 425–438, Mar. 2012.

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

[58] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inform. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.

[59] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Comm. Pure Applied Math.*, vol. 57, no. 11, pp. 1413–1457, Nov. 2004.

[60] 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*, vol. 5981, no. 4, pp. 586–598, Dec. 2007.

[61] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[62] M. Fornasier and C.-B. Schönlieb, "Subspace correction methods for total variation and ℓ_1-minimization," *IMA J. Numer. Anal.*, vol. 47, no. 8, pp. 3397–3428, 2009.

[63] G. Steidl and T. Teuber, "Removing multiplicative noise by Douglas-Rachford splitting methods," *J. Math. Imag. Vis.*, vol. 36, no. 3, pp. 168–184, 2010.

[64] G. Chen and M. Teboulle, "A proximal-based decomposition method for convex minimization problems," *Math. Program.*, vol. 64, pp. 81–101, 1994.

[65] 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.*, vol. 3, no. 4, pp. 1015–1046, 2010.

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

[67] L. M. Briceño-Arias and P. L. Combettes, "A monotone + skew splitting model for composite monotone inclusions in duality," *SIAM J. Opt.*, vol. 21, no. 4, pp. 1230–1250, 2011.

[68] 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.*, 2011.

[69] B. C. Vũ, "A splitting algorithm for dual monotone inclusions involving cocoercive operators," *Adv. Comput. Math.*, 2011, Accepted, http://www.springerlink.com/content/m177247u22644173.

[70] L. Condat, "A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms," 2012, Submitted, http://hal.archives-ouvertes.fr/hal-00609728/fr/.

[71] N. Pustelnik, J.-C. Pesquet, and C. Chaux, "Relaxing tight frame condition in parallel proximal methods for signal restoration," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 968–973, Feb. 2012.

[72] C. Couprie, L. Grady, L. Najman, J.-C. Pesquet, and H. Talbot, "Constrained TV-based regularization on graphs," *Submitted*, 2012.

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

[74] 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.*, vol. 20, no. 12, pp. 3483–3494, Dec. 2011.

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

[76] J.-F. Aujol, "Some first-order algorithms for total variation based image restoration," *J. Math. Imag. Vis.*, vol. 34, no. 3, pp. 307–327, Jul. 2009.

[77] B. A. Turlach, W. N. Venables, and S. J. Wright, "Simultaneous variable selection," *Technometrics*, vol. 47, no. 3, pp. 349–363, Aug. 2005.

[78] 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.

[79] Y. Chen and A. O. Hero, "Recursive $\ell_{1,\infty}$ lasso," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 3978–3987, Aug. 2012.

[80] E. J. Candés, M. B. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 877–905, Dec. 2008.

[81] G. Gilboa and S. Osher, "Nonlocal Operators with Applications to Image Processing," *Multiscale Model. and Simul.*, vol. 7, no. 3, pp. 1005, 2009.

[82] G. Peyré and J. Fadili, "Group sparsity with overlapping partition functions," in *Proc. Eur. Sig. and Image Proc. Conference*, Barcelona, Spain, Aug. 29 – Sept. 2, 2011, pp. x+5.

[83] I. Bayram and M. Kamasak, "Directional total variation," *IEEE Signal Processing Letters*, 2012.

[84] E. Van Den Berg and M. P. Friedlander, "Probing the Pareto frontier for basis pursuit solutions," *SIAM J. Sci. Comput.*, vol. 31, no. 2, pp. 890–912, Nov. 2008.

[85] P. Weiss, L. Blanc-Féraud, and G. Aubert, "Efficient schemes for total variation minimization under constraints in image processing," *SIAM J. Sci. Comput.*, vol. 31, no. 3, pp. 2047–2080, Apr. 2009.

[86] J. M. Fadili and G. Peyré, "Total variation projection with first order schemes," *Trans. Img. Proc.*, vol. 20, no. 3, pp. 657–669, Mar. 2011.

[87] X. Zhang, M. Burger, X. Bresson, and S. Osher, "Bregmanized nonlocal regularization for deconvolution and sparse reconstruction," *SIAM J. Imaging Sci.*, vol. 3, no. 3, pp. 253–276, 2010.

[88] A. Buades, B. Coll, and J. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Model. and Simul.*, vol. 4, no. 2, pp. 490–530, 2005.

[89] A. Foi and G. Boracchi, "Foveated self-similarity in nonlocal image filtering," in *Proc. SPIE Electronic Imaging 2012, Human Vision and Electronic Imaging XVII*, Burlingame (CA), USA, Jan. 2012, vol. 8291.

[90] G. Gilboa and S. Osher, "Nonlocal linear image regularization and supervised segmentation," *Multiscale Model. and Simul.*, vol. 6, no. 2, pp. 595–630, 2007.

[91] Z. Wang and A. C. Bovik, "Mean squared error: love it or leave it?," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98–117, Jan. 2009.

[92] M. Lebrun, M. Colom, A. Buades, and J. M. Morel, "Secrets of image denoising cuisine," *Acta Numerica*, vol. 21, pp. 475–576, 2012.

[93] C. Kervrann and J. Boulanger, "Local adaptivity to variable smoothness for exemplar-based image regularization and representation," *Int. J. Comp. Vis.*, vol. 79, no. 1, pp. 45–69, 2008.

[94] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. R. Statist. Soc. B*, vol. 58, pp. 267–288, 1996.

[95] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. R. Statist. Soc. B*, vol. 68, pp. 49–67, 2006.

[96] 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.

[97] I. Ekeland and R. Témam, *Convex analysis and variational problems*, SIAM, Philadelphia, 1999.

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