ON CONVEX FINITE-DIMENSIONAL VARIATIONAL METHODS IN IMAGING SCIENCES, AND HAMILTON-JACOBI EQUATIONS

JÉRÔME DARBON

October 16, 2013*

Abstract. We consider standard finite-dimensional variational models used in signal/image processing that consist in minimizing an energy involving a data fidelity term and a regularization term. We propose new remarks from a theoretical perspective which give a precise description on how the solutions of the optimization problem depend on the amount of smoothing effects and the data itself. The dependence of the minimal values of the energy is shown to be ruled by Hamilton-Jacobi equations, while the minimizers $\mathbf{u}(\mathbf{x},t)$ for the observed images $\mathbf{x}$ and smoothing parameters t are given by

$$\mathbf{u}(\mathbf{x},t) = \mathbf{x} - t\nabla H(\nabla_{\mathbf{x}} E(\mathbf{x},t))$$

where $E(\mathbf{x},t)$ is the minimal value of the energy and H is a Hamiltonian related to the data fidelity term. Various vanishing smoothing parameter results are derived illustrating the role played by the prior in such limits.

1. Introduction

Many low-level image processing and computer vision problems can be formulated as an optimization problem. A quite standard approach for performing image denoising consists in optimizing an energy that is a weighted combination between a data fidelity term (that embeds the knowledge we have on the nature of the noise that corrupts the image) and a prior (that contains the knowledge we have on the image to be reconstructed). Among such models, the Rudin-Osher-Fatemi (ROF) model which consists in minimizing the Total Variation with a separable quadratic term has received a lot of interest in the image processing and computer vision communities since the seminal works of [8, 35]. Many other priors other than Total Variation have been introduced in image processing and computer vision to get better quality for image reconstruction (see [1, 8, 37] for instance). In this paper, we shall consider variational imaging problems that consist in minimizing a convex data fidelity term with a given convex prior. In a Bayesian framework, this corresponds to consider Maximum A Posteriori estimators. The goal of this paper is to establish new theoretical relationships between the solutions of the energy minimization problems in image processing and Hamilton-Jacobi (H-J) partial differential equations.

A finite-dimensional framework is considered in this paper. Scalars and vectors will be denoted by letter and bold letters, respectively. It is assumed that images are defined on a lattice $\mathcal{V}$ with cardinality $|\mathcal{V}| = n$. The value of an image $\mathbf{x}$ at a site $i \in \mathcal{V}$ is denoted by $x_i \in \mathbb{R}$. It is more convenient for mathematical purposes to see an image $\mathbf{x}$ as an element of $\mathbb{R}^n$ and to access its i^{th} entry by $x_i \in \mathbb{R}$ for $i = 1,\dots,n$. We shall abuse notation by always writing x_i even though i could live in the sets $\mathcal{V}$ or $\{1,\dots,n\}$. This abuse of notation should never be confusing in this paper.

A standard model for image formation is formally given by

$$\mathbf{x} = A\bar{\mathbf{u}} + \boldsymbol{\eta}, \tag{1.1}$$

where $\mathbf{x} \in \mathbb{R}^n$ is the observed signal or image and we aim at estimating $\bar{\mathbf{u}}$. In other words, the observed image $\mathbf{x}$ has been generated from an unknown ideal (noiseless)

* This work has been supported by ONR-N00014-11-1-0749. UCLA Mathematics Department, and on temporary leave from CNRS.

image $\bar{\mathbf{u}} \in \mathbb{R}^n$ that is seen through a linear application (represented by the matrix A with real entries) that generally corresponds to a blur in image processing. In addition, it has also been corrupted by some additive noise $\boldsymbol{\eta}$. It is also assumed that the model "$A\bar{\mathbf{u}}+\boldsymbol{\eta}$" spans $\mathbb{R}^n$, i.e., for any observed image $\mathbf{x}$ there exists at least one ideal image $\bar{\mathbf{u}}$ and a realization of the noise $\boldsymbol{\eta}$ that yields the observation $\mathbf{x}$. In this paper, we shall assume that the matrix A is an invertible matrix. Two scenarios are considered: (a) ideal images live in a subset of $\mathbb{R}^n$ and thus we assume that the noise perturbation spans the whole space or (b) we assume that the noise is bounded and that the set of possible original images is $\mathbb{R}^n$. Note that the latter assumption on the boundedness of the noise represents no limitation for signal/image processing. Indeed sensors readouts are bounded from obvious physical reasons.

Priors contain some knowledge we have on the signal to reconstruct [37]. Several priors used in image processing are convex functions but they are not necessarily differentiable. We briefly present below some of the most common priors encountered in the literature. Priors based on ℓ_1 have been remarkably popular since they promote sparsity. Among them, Total Variation-based prior is a popular choice in image processing since it allows the reconstructed image to exhibit sharp edges. There are several ways to define Total Variation (TV) on lattices. Perhaps the simplest one consists in considering the weighted absolute value of the pairwise difference of some pixels. This yields to the following finite dimension anisotropic version of TV [8, 11, 16, 29]:

$$J(\mathbf{y}) = \sum_{(i,j)\in\mathcal{V}^2} w_{ij}|y_j - y_i|, \tag{1.2}$$

where the weights w_{ij} are finite and non-negative. Another used formulation of TV considers a more isotropic version [35] than the pairwise interactions formulation given by (1.2). It is contained in the following general form:

$$J(\mathbf{y}) = \sum_{(i,j,k)\in\mathcal{V}^3} w_{ijk}\sqrt{(y_j-y_i)^2+(y_k-y_i)^2},$$

where any weight w_{ijk} is still non-negative and finite. Higher order interaction priors can also be considered. Note that both of these formulations can be seen as a particluar case of Non-Local Total Variation [23] that takes the following form

$$J(\mathbf{y}) = \sum_{i\in\mathcal{V}}\sqrt{\sum_{j\in\mathcal{V}} w_{ij}(\mathbf{y}_j-\mathbf{y}_i)^2}$$

where any weight w_{ij} is finite and non-negative.

Another useful prior on images consists in weighted l_1-norms for encouraging sparsity of the image. It takes the following form

$$J(\mathbf{y}) = \sum_{i\in\mathcal{V}} w_i|y_i|,$$

where, again, the weights w_i are non-negative and finite. This prior has received a lot of interest due to its close connection with compressive sensing reconstruction [10, 19].

One can also use priors on images that do not come from l_p-norm. A general form of priors with pairwise interactions takes the following form

$$J(\mathbf{y}) = \sum_{(i,j)\in\mathcal{V}^2} \phi_{ij}(y_i - y_j),$$

where any ϕ_{ij} is a convex function. For example a popular choice for ϕ_{ij} is the Huber prior [37] that corresponds to the following definition:

$$\phi_{ij}(z) = \begin{cases} z^2 & \text{if} |z| \leq \alpha, \\ 2\alpha|z| - \alpha^2 & \text{otherwise}, \end{cases}$$

where the real valued parameter α is non-negative. Note that this prior is differentiable. We refer the reader to [1, 8, 37] for other possible priors.

The data fidelity term corresponds to the knowledge we have on the process that alter the ideal image $\bar{\mathbf{u}}$. A standard assumption in image processing is that the noise that corrupts ideal images is Gaussian and additive. For the sake of simplicity, we should only consider this case in this section. We shall present it formally while rigorous justification will be given later. This corresponds to consider a separable quadratic data fidelity term. More general data fidelity terms, i.e., non-Gaussian noise, will be covered later in the paper. Given an observed image $\mathbf{x} \in \mathbb{R}^n$, a standard imaging problem consists in minimizing in the $\mathbf{y}$ variable the following energy

$$\mathbf{y} \mapsto \frac{1}{2t}\|\mathbf{y} - \mathbf{x}\|_2^2 + J(\mathbf{y}), \tag{1.3}$$

for any fixed $t > 0$. The Euclidean norm in $\mathbb{R}^n$ is denoted by $\|\cdot\|_2$. The real $t > 0$ gives the amount of filtering we wish to consider. It corresponds to a trade-off between the smoothing effect of the prior and the fidelity to the observed image $\mathbf{x}$.

A lot of effort has been devoted to proposing efficient algorithms for minimizing (1.3) when J is given by one of the prior recalled above and for a given *fixed* observed image $\mathbf{x}$ and a given *fixed* smoothing parameter $t > 0$. This is still an active field of research. In this paper, we study the behavior of the minimal values of the energy (1.3) with respect to both the observed data $\mathbf{x} \in \mathbb{R}$ and the smoothing parameter $t > 0$. In other words we study the function $F : \mathbb{R}^n \times (0, +\infty) \to \mathbb{R}$ formally defined by

$$F(\mathbf{x}, t) = \min_{\mathbf{y} \in \mathbb{R}^n} \left\{ \frac{1}{2t}\|\mathbf{y} - \mathbf{x}\|_2^2 + J(\mathbf{y}) \right\}. \tag{1.4}$$

For a fixed $t > 0$, the mapping $\mathbf{x} \mapsto F(\mathbf{x}, t)$ is called the Moreau envelopes of J [34, Def. 1.22, p. 20] or the Moreau-Yosida regularization of J [28, p. 317]. For imaging purposes, one is generally more interested in the minimizer of the energy F itself (i.e., the vector $\mathbf{y}$ that realizes the minimum of F) rather than its minimal value $F(\mathbf{x}, t)$. This minimizer is called the proximal point of $\mathbf{x}$ relatively to J [32]. In that context, the object of interest is the function $\mathbf{v} : \mathbb{R}^n \times (0, +\infty)$ that maps the observed data $\mathbf{x} \in \mathbb{R}^n$ and the smoothing parameter $t > 0$ to the proximal point (i.e., the minimizer of (1.4))

$$\mathbf{v}(\mathbf{x}, t) = \operatorname*{argmin}_{\mathbf{y} \in \mathbb{R}^n} \left\{ \frac{1}{2t}\|\mathbf{y} - \mathbf{x}\|_2^2 + J(\mathbf{y}) \right\}. \tag{1.5}$$

For any fixed $t > 0$, the mapping $\mathbf{x} \mapsto \mathbf{v}(\mathbf{x}, t)$ is called the proximal mapping of J [34, Def. 1.22,p. 20]. In [12], the connection with imaging problems of the form of (1.3) and proximal methods is highlighted. The authors give some properties of the proximal mapping $\mathbf{x} \mapsto \mathbf{v}(\mathbf{x}, t)$ for any fixed $t > 0$. Nevertheless, their study is not sufficient for our goal since we shall consider non-quadratic data fidelity terms. In

[24, chp. 3] and [25], the author studies the proximal mapping for Total Variation in an infinite-dimensional framework.

In this paper, we shall consider imaging problems where the data fidelity term is not restricted to be quadratic. We study both the mappings that take the observed data $\mathbf{x} \in \mathbb{R}^n$ and the smoothing parameter $t \geq 0$ to the minimal values of the energy $F(\mathbf{x},t)$ (1.4) and the minimizers (1.5) $\mathbf{v}(\mathbf{x},t)$.

Let us illustrate the behavior of the minimal values and of the minimizers on a simple example. We consider an energy of the form of (1.3). We recall that the data fidelity term is a separable quadratic term. We set the prior J to an anisotropic Total Variation with pairwise interactions of the form of (1.2), i.e., $J(\mathbf{y}) = \sum_{(i,j)\in\mathcal{V}^2} w_{ij}|y_j - y_i|$. We set the lattice $\mathcal{V}$ to be a regular 2D grid and we endow it with the 4-nearest neighbors [37]. For any $i \in \mathcal{V}$ we denote by $\mathcal{N}(i)$ the set of the 4-nearest neighbors of i. The weights w_{ij} are defined as follows: for any $i \in \mathcal{V}$ and any $j \in \mathcal{N}(i)$ we set $w_{ij} = \frac{1}{2}$. Thus, the energy we consider in this example corresponds to the anisotropic ROF problem [35] that takes the form

$$\frac{1}{2t}\sum_{i\in\mathcal{V}}(y_i - x_i)^2 + \frac{1}{2}\sum_{i\in\mathcal{V}}\sum_{j\in\mathcal{N}(i)}|y_j - y_i|\,. \tag{1.6}$$

The minimizer and the minimal value of the this energy can be computed up to the machine precision using maximum-flow based algorithms [11, 16, 29]. Figure 1.1 depicts two original images: an aerial image of an area of Montpellier and an image of a baboon that are respectively denoted by $\mathbf{x_M}$ and $\mathbf{x_B}$. Figure 1.2-(a-d) depicts the minimizer of energy (1.6) for the two images with two different values of the smoothing parameter t. It also depicts in figure 1.2-(e) the minimal values of (1.6) as a function of the smoothing parameter t with the observed data $\mathbf{x} = \mathbf{x_M}$ or $\mathbf{x} = \mathbf{x_B}$ being fixed. Let us note that this function is convex. We pursue this example by illustrating the behavior of minimizing (1.6) when it is seen as a function of both the observed data $\mathbf{x}$ and the smoothing parameter t. The following toy example aims at illustrating this behavior. We consider the convex combinations between the two observed images $\mathbf{x_B}$ and $\mathbf{x_M}$, and the two smoothing parameters $t_0 = 15$ and $t_1 = 50$. More precisely we consider

$$[0,1] \ni \alpha \mapsto \min_{\mathbf{y}\in\mathbb{R}^n}\left\{\frac{\|\mathbf{y} - (1-\alpha)t_0\mathbf{x_B} + \alpha t_1\mathbf{x_M}\|_2^2}{2((1-\alpha)t_0 + \alpha t_1)} + \frac{1}{2}\sum_{i\in\mathcal{V}}\sum_{j\in\mathcal{N}(i)}|y_j - y_i|\right\}. \tag{1.7}$$

Figure 1.3-(a,c) depicts the convex combination images $(1-\alpha)t_0\mathbf{x_B} + \alpha t_1\mathbf{x_M}$. The minimizer of (1.7) for $\alpha = 0.4$ and $\alpha = 0.6$ are depicted in figure 1.3-(b,d). The plot of the function (1.7) is depicted in Figure 1.3-(e). We also observe that this function is convex.

The goal of this paper is to show that for a broad class of variational problems (including the ones presented above) the function E which maps the observed signals/images $\mathbf{x} \in \mathbb{R}^n$ and the smoothing parameters $t \geq 0$ to the minimal values is ruled by the solution of a Hamilton-Jacobi partial differential equation. The initial datum of this Hamilton-Jacobi equation is the convex prior J. We provide a closed formula in Proposition 3.1 that gives an explicit representation of the minimizers $\mathbf{u}$ in function of $\mathbf{x}, t > 0, E$ and some Hamiltonians related to the data fidelity term. In addition, the behavior of both mappings for the limiting case $t \to 0$ is also described.

(a) (b)

FIG. 1.1. *Two original images: (a) an aerial image of an area of Montpellier, and (b) an image of a baboon.*

Indeed, theorem 2.6 gives the value of E when $t \to 0$ and proposition 3.2 provides the convergence of the minimizer $\mathbf{u}$ when $t \to 0$. The behavior of the semi-derivative of E at $(\mathbf{x},0)$ and the evolution rule of the minimizer when $t \to 0$ are also characterized; see proposition 3.3 and proposition 3.4. These results when the smoothing parameter $t \to 0$ are important for imaging purposes.

In other words, the study gives the dependency of the minimal value of the energy and of the minimizer of the variational problem with respect to the observed data $\mathbf{x}$ and the value of the smoothing parameter t.

The remainder of this paper is organized as follows: section 2 studies the minimal value of the imaging problems in function of the observed image $\mathbf{x} \in \mathbb{R}^n$ and the smoothing parameter $t > 0$. Not only lemma 2.1 shows that the image processing problem is well-posed: for any $\mathbf{x} \in \mathbb{R}^n$ and $t > 0$ it has a unique solution); but it also proves properties on the Hamiltonian that are essential for Hamilton-Jacobi equations. Indeed, theorem 2.6 shows that the minimal value function is convex and obeys a Hamilton-Jacobi equation with initial datum. Section 3 gives the formulas for the minimizer of imaging problems. Proposition 3.1 gives formulas that connect directly the minimizers to the minimal values of the imaging problem: Proposition 3.2, 3.3 and 3.4 give a precise description of the behavior of the minimizers when the smoothing parameter t ends to 0. We draw some conclusions in section 4.

2. Convex problems and Hamilton-Jacobi Equations

The goal of this section is to describe theoretical connections between the minimal value of convex image processing problems with Hamilton-Jacobi equations with convex initial datum. In this paper we shall make use of the book on convex analysis in finite dimension of Hiriart-Urruty and Lemaréchal [27] and [28]. We also refer the reader to the monographs [6, 18, 20, 33]. We first introduce some useful definitions of convex analysis.

2.1. Preliminaries

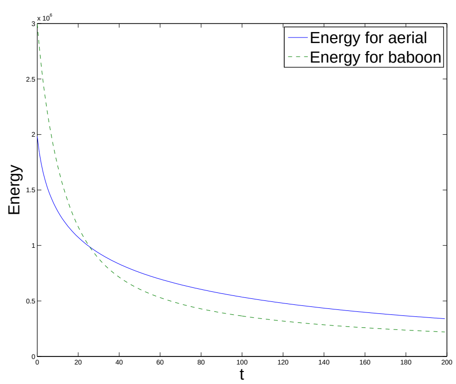
We first recall some standard definitions of convex analysis.

(a)

(b)

(c)

(d)



(e)

FIG. 1.2. *Minimizers of the quadratic + anisotropic TV model for two values of the smoothing parameter t: In (a) and (c) the aerial image respectively filtered with $t=25$ and $t=50$ while in (c) and (d) is the minimizer for the baboon image with the same values of the smoothing parameter t. The figure (e) depicts the plot of the minimal values of the energy with respect to the smoothing parameter t for the two images.*

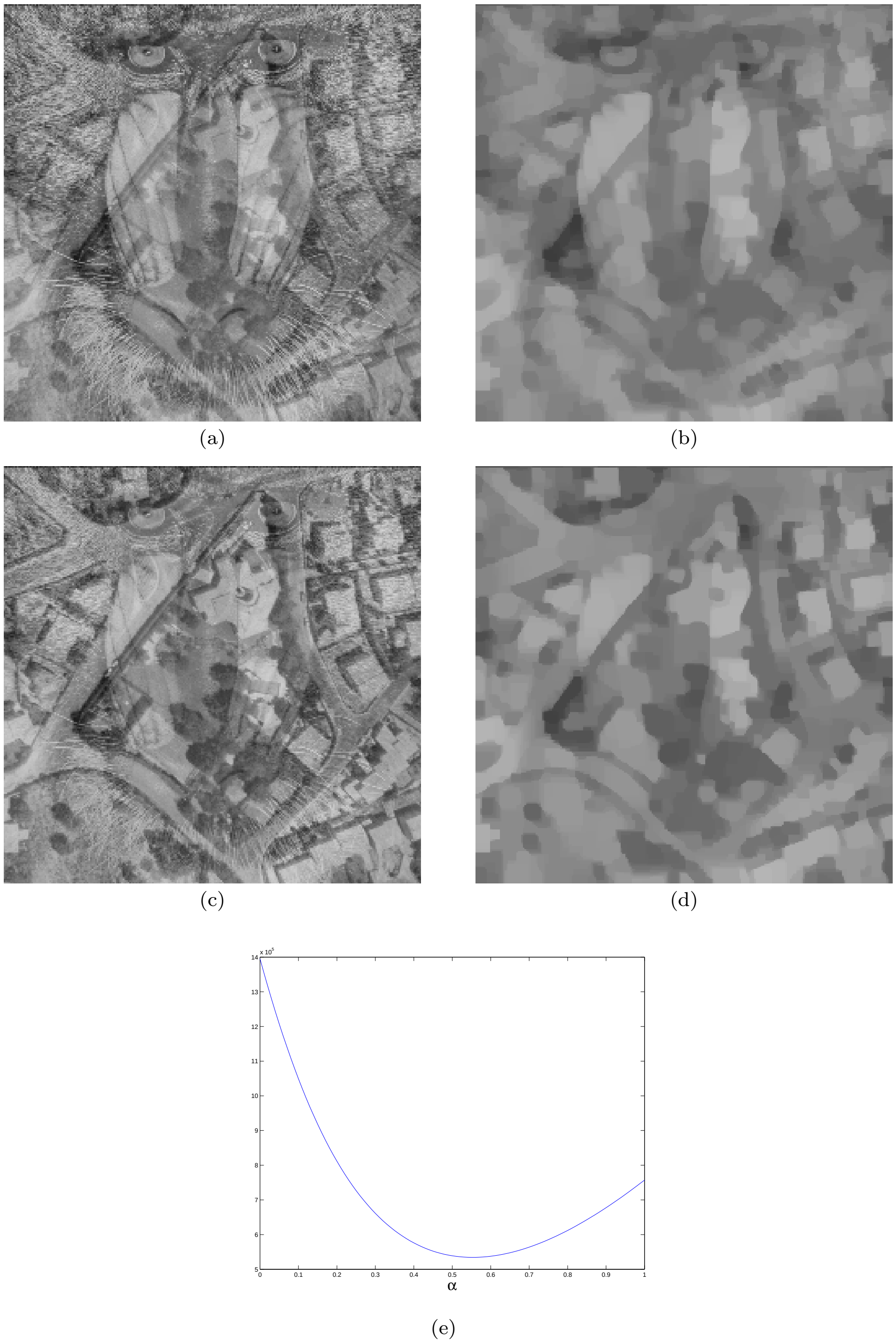


Fig. 1.3. *Figure (a) and (b) depicts the convex combination* $(1-\alpha)t_0\mathbf{x_B}+\alpha t_1$ *for* $\alpha = 0.4$ *and* $\alpha = 0.6$ *respectively. The corresponding minimizers that solves* (1.7) *are respectively depicted in (b) and (d). The plot of the function* (1.7) *that corresponds to the minimal energy against the convex combination coefficient* $\alpha \in [0,1]$ *is shown in (e).*

Definition 2.1 (Convex functions and the set $\Gamma_0(\mathbb{R}^n)$). *A function $f:\mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$, not identically $+\infty$, is said to be convex when, for all $(\mathbf{x},\mathbf{y})\in\mathbb{R}^n\times\mathbb{R}^n$ and for all $\alpha\in(0,1)$, there holds*

$$f(\alpha\mathbf{x}+(1-\alpha)\mathbf{y}) \le \alpha f(\mathbf{x})+(1-\alpha)f(\mathbf{y}), \tag{2.1}$$

considered as an inequality in $\mathbb{R}^n\cup\{+\infty\}$. The class of convex functions that are lower semicontinuous is denoted by $\Gamma_0(\mathbb{R}^n)$. The function f is said to be strictly convex if the inequality is strict in (2.1).

The class of convex functions $\Gamma_0(\mathbb{R}^n)$ is the one of interest in this paper. The domain of $f\in\Gamma_0(\mathbb{R}^n)$ is the nonempty set $\operatorname{dom} f=\{\mathbf{x}\in\mathbb{R}^n, f(\mathbf{x})<+\infty\}$. The standard Euclidean scalar product of $\mathbb{R}^n$ is denoted by $\langle\cdot,\cdot\rangle$ and its associated norm by $\|\cdot\|_2$.

Definition 2.2 (Subdifferential/subgradients [27, p. 241]). *The subdifferential $\partial f(\mathbf{x})$ of $f\in\Gamma_0(\mathbb{R}^n)$ at $\mathbf{x}\in\operatorname{dom} f$ is the set (possibly empty) of vectors $\mathbf{s}\in\mathbb{R}^n$ satisfying*

$$\forall\mathbf{y}\in\mathbb{R}^n,\ f(\mathbf{y})\ge f(\mathbf{x})+\langle\mathbf{s},\mathbf{y}-\mathbf{x}\rangle.$$

Any vector $\mathbf{s}\in\partial f(\mathbf{x})$ is called a subgradient of f at $\mathbf{x}$.

Definition 2.3 (Fenchel transform [28, p. 37]). *Let $f\in\Gamma_0(\mathbb{R}^n)$. The Fenchel-Legendre transform $f^*\in\Gamma_0(\mathbb{R}^n)$ of f is given by*

$$\mathbf{s}\mapsto f^*(\mathbf{s})=\sup_{\mathbf{x}\in\operatorname{dom} f}\{\langle\mathbf{s},\mathbf{x}\rangle-f(\mathbf{x})\}.$$

Definition 2.4 (Infimal-convolution [27, p. 163]). *Let $f_1\in\Gamma_0(\mathbb{R}^n)$ and $f_2\in\Gamma_0(\mathbb{R}^n)$. The infimal-convolution of f_1 and f_2 is the function from $\mathbb{R}^n$ to $\mathbb{R}\cup\{+\infty\}$ defined by*

$$(f_1\,\square\, f_2)(\mathbf{x})=\inf_{\mathbf{x}_1+\mathbf{x}_2=\mathbf{x}}\{f_1(\mathbf{x}_1)+f_2(\mathbf{x}_2)\}.$$

The infimal convolution is said to be exact if the infimum is attained at $(\bar{\mathbf{x}}_1,\bar{\mathbf{x}}_2)\in\operatorname{dom} f_1\times\operatorname{dom} f_2$, and the infimum can be replaced by a minimum.

Definition 2.5 (Differentiability and gradient). *A function $f:\Omega\to\mathbb{R}$ with $\Omega\neq\emptyset$ an open subset of $\mathbb{R}^n$ is said to be differentiable at $\mathbf{x}\in\Omega$ if there exists a linear form l on $\mathbb{R}^n$ such that*

$$f(\mathbf{x}+\mathbf{h})=f(\mathbf{x})+l(h)+o(\|h\|_2).$$

This linear form l is donted by $Df(\mathbf{x})$ and is called the differential of f at $\mathbf{x}$. It can be represented by a unique vector of $\mathbb{R}^n$ that we denote by $\nabla f(\mathbf{x})\in\mathbb{R}^n$. It is defined for all $\mathbf{x}\in\Omega$ by

$$Df(\mathbf{x})(\mathbf{h})=\langle\nabla f(\mathbf{x}),\mathbf{h}\rangle\ \text{ for all } \mathbf{h}\in\mathbb{R}^n.$$

We call this element the gradient of f at $\mathbf{x}$.

2.2. H-J equations with convex initial datum and convex Hamiltonians We now describe the connection between a large class of convex imaging problems and Hamilton-Jacobi equations.

We consider imaging problems that take formally the following form: for any observed data $\mathbf{x} \in \mathbb{R}^n$ and any $t>0$ solve

$$\inf_{\mathbf{y}\in\mathbb{R}^n} \left\{ J(\mathbf{y}) + tH^*\left(\frac{\mathbf{x}-\mathbf{y}}{t}\right)\right\}. \tag{2.2}$$

Here J corresponds to the prior we have on the image to be reconstructed. The other term, namely $tH^*\left(\frac{\mathbf{x}-\cdot}{t}\right)$ corresponds to the data fidelity. For example, the canonical imaging problem (1.4) is obtained by setting $H^* = \frac{1}{2}\|\cdot\|_2^2$. We shall make some assumptions on J and H. Throughout this paper, the following assumptions are made

(H1) $\operatorname{dom} H = \mathbb{R}^n$,
(H2) H is differentiable,
(H3) H strictly convex,
(H4) $J \in \Gamma_0(\mathbb{R}^n)$ (see definition 2.1).

We will add one of the following two assumptions:
either

(H5) H is 1-coercive, i.e., $\lim_{\|\mathbf{x}\|_2\to+\infty} \frac{H(\mathbf{x})}{\|\mathbf{x}\|_2} = +\infty$,

or either

(H5') H^* is differentiable at point whenever it has a subgradient at this point, i.e., $\forall \mathbf{x} \in \operatorname{dom} H^*, \partial H^*(\mathbf{x}) \neq \emptyset \Rightarrow \partial H^*(\mathbf{x}) = \{\nabla H^*(\mathbf{x})\}$,
(H6') H is bounded from below by a constant,
(H7') $\operatorname{dom} J = \mathbb{R}^n$.

Let us briefly review the impact of these assumptions on the general image processing problem (2.2). The set of assumptions (H1), (H2), (H3) and ($H5$) on H corresponds to consider that $\operatorname{dom} H^* = \mathbb{R}^n$ and that H^* is strictly convex, continuously differentiable and 1-coercive. These properties comes from [28, Corollary 4.1.4, p. 82]. The set of assumptions (H1-H5) are widely used in image processing since the only requirement on the prior J is (H4), i.e., $J \in \Gamma_0(\mathbb{R}^n)$ (see definition 2.1). This allows to consider priors J that have bounded domains. The set of assumptions (H1-H4) and (H5'-H7') allows for the data fidelity term H^* to have a bounded domain. This means that the perturbation due to the noise is bounded. Assumption (H5') is a technical assumption on H^* (an example will be given later in this section). This assumption will be used for making the connection between the general problem (2.2) and Hamilton-Jacobi equations. Assumption (H7') ensures that for any observed data $\mathbf{x} \in \mathbb{R}^n$ and any smoothing parameter $t>0$ the imaging problem has always a solution. The assumption (H6') implies that $H^*(\mathbf{0})$ is finite.

The two sets of assumptions (H1)-(H4) with either (H5) or (H5')-(H7') yield to data fidelity terms H^* and priors J where, for each case, the domain of definition of J or H^* is $\mathbb{R}^n$. This ensures that there is a solution to the imaging problem (2.2) for any observed data $\mathbf{x} \in \mathbb{R}^n$ and any smoothing parameter $t > 0$. Indeed, the next lemma shows that under the above assumptions, the problem (2.2) admits a unique minimizer for any observed data $\mathbf{x}$ and positive smoothing parameter t. Note that if one assumption is removed then it may happen that the existence of a minimizer might be lost, or that the uniqueness does not hold, or that the connection of the imaging problem (2.2) with Hamilton-Jacobi equations does not hold (at least in the classical sense). This means that these two sets of minimal assumptions are, in a sense, minimal for imaging purposes.

LEMMA 2.1. *Suppose the assumptions (H1)-(H4) along with either (H5) or (H5')-(H7') hold. Then, the following properties hold for any* $\mathbf{x} \in \mathbb{R}^n$ *and for any postive smooting parameter* $t > 0$*:*

i) The infimum in (2.2) *exists and is attained at a unique point* $\bar{\mathbf{y}} \in \operatorname{dom} J$.

ii) In addition, $\nabla H^*\left(\frac{\mathbf{x}-\bar{\mathbf{y}}}{t}\right)$ exists.

iii) Futhermore, $\nabla H^*\left(\frac{\mathbf{x}-\bar{\mathbf{y}}}{t}\right) \in \partial J(\bar{\mathbf{y}})$.

The proof is given in Appendix A. This Lemma means that the image restoration problem (2.2) is well-posed: it has a unique minimizer. There is exactly one minimizer and it corresponds to the estimated restored image. Note that this only states that the minimization problem (2.2) is well-behaved. In practice, one needs to find the minimizer of (2.2) by some theoretical or numerical/algorithmical means. The next section provides a constructive proof for finding this minimizer in function of the observed data $\mathbf{x} \in \mathbb{R}^n$ and the smoothing parameter $t \geq 0$.

We are now ready to relate the minimal values of the problem (2.2) to Hamilton-Jacobi equations. To this purpose, we consider the initial-value problem for the Hamilton-Jacobi equation that takes the following form:

$$\text{(H-J)} \begin{cases} \dfrac{\partial E}{\partial t}(\mathbf{x},t) + H(\nabla_{\mathbf{x}} E(\mathbf{x},t)) = 0 & \text{in } \mathbb{R}^n \times (0,+\infty), \\ E(\mathbf{x},0) = J(\mathbf{x}) & \forall \mathbf{x} \in \mathbb{R}^n, \end{cases}$$

where the unknown is the function E that maps $\mathbb{R}^n \times [0,+\infty)$ into $\mathbb{R}$. Here, $\partial E/\partial t$ and $\nabla_{\mathbf{x}} E$ respectively denote the partial derivatives with respect to t and the gradient vector with respect to $\mathbf{x} \in \mathbb{R}^n$ of the function E.

The function H is called the Hamiltonian. In general, Hamilton-Jacobi equations do not have *global* classical solutions in the sense that one cannot find a differentiable function that satisfies (H-J) everywhere in $\mathbb{R}^n \times (0,+\infty)$. The theory of viscosity solutions has been developed in [13, 14] to provide an appropriate notion on weak solutions of (H-J) and has been widely studied since. The formalism proposed in this paper does not require the theory of viscosity solutions; the goal of this paper is to exhibit the connections between (H-J) and convex variational problems in imaging sciences. When either the Hamiltonian H or the initial datum J is convex along with some continuity assumptions, then the solutions can be obtained through the Hopf and Lax formulas [4], [21, chp. 10]. In this paper, we shall use these formulas and make strong assumptions on both the Hamiltonian and the initial datum such that the solutions are classical.
The solution of the Hamilton-Jacobi equation is given by

THEOREM 2.6 (Hopf and Lax formula). *Suppose the assumptions (H1)-(H4) along with either (H5) or (H5')-(H7') hold. Then, the unique differentiable and convex function* $E : \mathbb{R}^n \times [0,+\infty) \to \mathbb{R}$ *that satifies the Hamilton-Jacobi equation (H-J) with initial datum* J *in* $\mathbb{R}^n \times (0,+\infty)$ *is defined by*

$$E(\mathbf{x},t) = (J^* + tH)^*(\mathbf{x}) \qquad (Hopf formula) \tag{2.3}$$

$$= (J \square (tH)^*)(\mathbf{x}) \qquad (Lax formula) \tag{2.4}$$

$$= \inf_{\mathbf{y} \in \mathbb{R}^n} \left\{ J(\mathbf{y}) + tH^*\left(\frac{\mathbf{x}-\mathbf{y}}{t}\right) \right\}. \tag{2.5}$$

Furthermore, for any $\mathbf{x} \in \operatorname{dom} J$ *the pointwise limit of* $\lim_{t\to 0, t>0} E(\mathbf{x},t)$ *exists and is given by*

$$\lim_{\substack{t\to 0\\ t>0}} E(\mathbf{x},t) = J(\mathbf{x}).$$

In addition, the infimum in (2.5) *is attained and is unique.*

The proof relies on convex analysis and is done in [26, exercise 7.28, p. 358]. Note that with our assumptions, both the Hopf and Lax formulas yield the same solution. This comes from the fact that both the Hamiltonian H and the initial datum J live in $\Gamma_0(\mathbb{R}^n)$ as defined in definition 2.1.
The set of assumptions (H1)-(H4) along with either (H5) or (H5')-(H7') are rather minimal. Indeed, if one assumption is relaxed, then a solution to the H-J equation with initial datum may not be defined on $\mathbb{R}^n \times [0,+\infty)$ but only on a subset of it. Also, the solution may not be classical as one cannot expect to find a solution that is differentiable everywhere on the interior of its domain of definition.

Note that for any fixed $t > 0$ the Hopf and Lax formula corresponds exactly to the imaging problem of interest (2.2) which consists in estimating the ideal image $\bar{\mathbf{u}}$ while observing the data $\mathbf{x} = A\bar{\mathbf{u}} + \boldsymbol{\eta}$ given by the image formation model (1.1). The convex behavior of the minimal energies that are observed in Figures 1.2-(e) and 1.3-(e) follows directly from the fact the Hopf-Lax solution is a convex function. Compared to many Hamilton-Jacobi equations used in physics and optimal-control [3], the dimension of the problem is very high since it involves $(n+1)$ variables where n is the number of pixels of the image. In addition, the initial datum of the Hamilton-Jacobi equation is the prior we set on the image to reconstruct. The use of an imaging prior J as the convex initial datum of the Hamilton-Jacobi equation is not common in the partial differentiable equation literature compared to standard Hamilton-Jacobi based problems [3].
Some examples. Let us consider the particular case of the separable quadratic Hamiltonian, i.e., $H = 1/2\|\cdot\|_2^2$ in (H-J). Its Fenchel-Legendre transform is itself, i.e., $H^* = 1/2\|\cdot\|_2^2 = H$. We consider the following Hamilton-Jacobi equation

$$\begin{cases} \dfrac{\partial E}{\partial t}(\mathbf{x},t) + \dfrac{1}{2}\|\nabla_{\mathbf{x}} E(\mathbf{x},t)\|_2^2 = 0 & \text{in } \mathbb{R}^n \times (0,+\infty), \\ E(\mathbf{x},0) = J(\mathbf{x}) & \forall \mathbf{x} \in \mathbb{R}^n. \end{cases}$$

The solution is given by the Lax formula which gives for any $t > 0$ and any $\mathbf{x} \in \mathbb{R}^n$

$$E(\mathbf{x},t) = \min_{\mathbf{y}\in\mathbb{R}^n} \left\{ J(\mathbf{y}) + \frac{1}{2t}\|\mathbf{x}-\mathbf{y}\|_2^2 \right\}.$$

This corresponds exactly to the minimal value of the variational image restoration problem when the regularization term is J and the perturbation is zero-mean, additive and Gaussian. If J is a Total Variation then we obtain the Rudin-Osher-Fatemi model [35].

This approach can also deal with more elaborated observation models as (1.1).

For example, more general quadratic Hamiltonians of the form $H = \frac{1}{2}\langle A^{-1}\cdot,\cdot\rangle$ can be considered in (H-J). Here A is a $n \times n$ symmetric positive-definite invertible matrix with real entries. The Fenchel-Legendre transform of the Hamiltonian is $H^* = \frac{1}{2}\langle A\cdot,\cdot\rangle$. Suppose we observe the image $A^{-1}\mathbf{x}$, then the Lax formula yields

$E(A^{-1}\mathbf{x},t) = \min_{\mathbf{y}\in\mathbb{R}^n}\left\{J(\mathbf{y}) + \frac{1}{2t}\|\mathbf{x} - A\mathbf{y}\|_2^2\right\}$. This corresponds to a variational formulation for a deconvolution problem with additive Gaussian noise. In other words, the approach can deal with observations $\mathbf{x}$ that could be obtained from blurry versions of the ideal image $\bar{\mathbf{u}}$ and then corrupted by some realisation of the noise.

The assumption on the perturbation $\boldsymbol{\eta}$ is fairly weak. For instance, non-Gaussian noise can also be considered. Take $f \in \Gamma_0(\mathbb{R})$ as defined in definition 2.1. Suppose $H = f(\|\cdot\|)$ in (H-J) with $\|\cdot\|$ any norm on $\mathbb{R}^n$ and $f \in \Gamma_0(\mathbb{R})$ is an even function. Following [20, Prop. 4.2, p. 19], we have $H^* = f^*(\|\cdot\|_*)$, where $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$. For instance, if we consider the ℓ_p norm with $1 < p < +\infty$ and $f(x) = \frac{1}{p}x^p$, then $H(\mathbf{x}) = \frac{1}{p}\|\mathbf{x}\|_p^p$ and we get $H^*(\mathbf{x}) = \frac{1}{q}\|\mathbf{x}\|_q^q$ with q such that $\frac{1}{p} + \frac{1}{1} = 1$. By doing so, we can consider generalized Gaussian noise distribution [8, 30] as well.

So far, all the Hamiltonians considered satisfy assumption (H5). Under the (H5) assumption, we can consider convex regularization terms J that include some constraints, i.e., that $\operatorname{dom} J$ can be different from $\mathbb{R}^n$. For instance, suppose it is *a priori* known that the signal to reconstruct is piecewise constant and is nonnegative. A standard regularization term under these assumptions takes the following form

$$J(\mathbf{v}) = \begin{cases} \sum_{(i,j)\in\mathcal{V}^2} w_{ij}|v_j - v_i| & \text{if} v_i \geq 0 \\ +\infty & \text{otherwise.} \end{cases}$$

where each coefficient $w_{ij} \geq 0$.

Now, let us exhibit one that satisfies (H5'). Assume that $H(\mathbf{x}) = \sqrt{1 + \|\mathbf{x}\|_2^2}$. Its Fenchel-Legendre transform is

$$H^*(\mathbf{x}) = \begin{cases} -\sqrt{1 - \|\mathbf{x}\|_2^2} & \text{if } \|\mathbf{x}\|_2 \leq 1, \\ +\infty & \text{otherwise}, \end{cases}$$

and satisfies (H6'). Indeed H^* is differentiable in the open ball $\{\mathbf{x}\in\mathbb{R}^n \,|\, \|\mathbf{x}\|_2 < 1\}$ and has empty subdifferentials for any $\mathbf{x}\in\mathbb{R}^n$ such that $\|\mathbf{x}\|_2 \geq 1$.

More generally, the Lax formula given by Eq. (2.5) says that the regularization term in the image processing problem always become the initial datum of the Hamilton-Jacobi equation. The function H^* aims at taking into account some noise effects and implicitly defines the Hamiltonian through the Fenchel-Legendre transform. The amount of regularization is obtained by adjusting the smoothing parameter $t > 0$ through the perspective scaling $tH^*\left(\frac{\cdot}{t}\right)$.

3. Behavior of the minimizer

So far we have only considered the behavior of the minimal value of the optimization problem. Recall that we are more interested in the minimizer of the optimization problem since it corresponds in estimating the ideal image. We shall now address the issue of the behavior of the minimizers.

Since the the minimizer in the Hopf-Lax formula is unique for any observed data $\mathbf{x} \in \mathbb{R}^n$ and any smoothing parameter $t > 0$, we can introduce the function that maps the observed data and the regularization parameter to the minimizer; that is the function $\mathbf{u}: \mathbb{R}^n \times [0, +\infty) \to \mathbb{R}^n$ defined by

$$\mathbf{u}(\mathbf{x},t) = \begin{cases} \arg\min_{\mathbf{y}\in\mathbb{R}^n}\left\{J(\mathbf{y}) + tH^*\left(\dfrac{\mathbf{x}-\mathbf{y}}{t}\right)\right\} & \text{if } t > 0, \\ \mathbf{x} & \text{if } t = 0. \end{cases} \tag{3.1}$$

The justification for setting $\mathbf{u}(\mathbf{x},0)=\mathbf{x}$ will be justified in proposition 3.2. We first study a simple example to highlight the behavior of (3.1) before considering the general case.

3.1. An example with l_1 prior

Let us first consider an example in order to understand the kind of formula we wish to establish. We set an ℓ_1 prior on images (as given by Eq. (1.3)) $J=\|\cdot\|_1$ and consider the quadratic Hamiltonian $H=1/2\|\cdot\|_2^2$. The corresponding Hamilton-Jacobi equation is

$$\begin{cases} \frac{\partial E}{\partial t}(\mathbf{x},t)+\frac{1}{2}\|\nabla_{\mathbf{x}}E(\mathbf{x},t)\|_2^2=0 & \text{in } \mathbb{R}^n\times(0,+\infty), \\ E(\mathbf{x},0)=\|\mathbf{x}\|_1 & \forall \mathbf{x}\in\mathbb{R}^n. \end{cases}$$

The Hopf-Lax formula yields for any $t>0$ and any $\mathbf{x}\in\mathbb{R}^n$

$$E(\mathbf{x},t)=\min_{u\in\mathbb{R}^n}\left\{\frac{1}{2t}\|\mathbf{u}-\mathbf{x}\|_2^2+\|\mathbf{u}\|_1\right\}. \tag{3.2}$$

It is well known that the optimal solution corresponds to a soft thresholding/shrink applied component by component [31, 22, 17]. This operation is widely used in image processing, computer vision, machine learning and compressive sensing as it promotes sparsity and can be computed efficiently. The soft thresholding operator is defined for any real value a and any positive real number α as

$$S(a,\alpha)=\begin{cases} a-\alpha & \text{if } a\in(\alpha,+\infty), \\ 0 & \text{if } a\in[-\alpha,\alpha], \\ a+\alpha & \text{otherwise.} \end{cases} \tag{3.3}$$

The minimizer of (3.2) is then given component-wise by

$$u_i(\mathbf{x},t)=S(x_i,t)$$

for any $i=1,\ldots,n$. Another formulation of the solution consists in noting that the soft thresholding can be expressed as an Euclidean projection onto a closed convex set. More precisely, let π_{C_1} be the Euclidean projection onto the closed convex set $C_1=[-1,1]^n$. The soft thresholding reads as follows

$$\mathbf{u}(\mathbf{x},t)=\mathbf{x}-\pi_{tC_1}(\mathbf{x})=\mathbf{x}-t\pi_{C_1}\left(\frac{\mathbf{x}}{t}\right).$$

Let us note here that this problem is a specific Riemann problem and that the solutions and inequalities involving generalizations of this problem have originally been obtained by Bardi and Osher in [5].

This approach is not specific to the ℓ_1-norm and can be generalized. Indeed, instead of choosing $J=\|\cdot\|_1$, we consider $J\in\Gamma_0(\mathbb{R}^n)$ that is positively 1-homogeneous, i.e., $J(\lambda\mathbf{y})=\lambda J(\mathbf{y})$ for any $\mathbf{y}\in\mathbb{R}^n$ and $\lambda\geq 0$. The set convex closed set C_1 is replaced by the convex closed set C defined by

$$C=\partial J(\mathbf{0}),$$

and the minimizer then satisfies

$$\mathbf{u}(\mathbf{x},t)=\mathbf{x}-\pi_{tC}(\mathbf{x})=\mathbf{x}-t\pi_{C}\left(\frac{\mathbf{x}}{t}\right).$$

Note that the behavior of the minimizer could be studied through the stability of the Euclidean projector onto a closed convex set [7, 36]. However, in this paper we do not wish to resctrict the study to 1-homogeneous prior J as many priors used in signal/image processing do not enjoy this property.

3.2. Main results

We now describe the general case that will include the above examples. Our goal is to establish proposition 3.1 that gives an explicit formula of the minimizers in function of $\mathbf{x} \in \mathbb{R}^n, t>0, E$ and H. The behavior of the minimizers when the smoothing parameter $t \to 0$ is given by Prop. 3.2. The behaviors of the minimal values E and $\nabla_{\mathbf{x}} E$ at $(\mathbf{x},0)$ are also characterized by proposition 3.3 and proposition 3.4.

Lemma 2.1 states that the infimal-convolution (2.5) of J with $tH^*(\frac{\cdot}{t})$, which corresponds to the Hopf-Lax formula, is exact for any $t>0$. Thus, we can invoke [28, Prop. 3.4.1, p. 119] and deduce that the subdifferential of this infimal-convolution with respect to the variable $\mathbf{x}$ is given for any $t>0$ and any $\mathbf{x} \in \mathbb{R}^n$ by

$$\partial(\mathbf{y} \mapsto E(\mathbf{y},t))(\mathbf{x}) = \partial J(\mathbf{u}(\mathbf{x},t)) \cap \partial H^* \left(\frac{\mathbf{x} - \mathbf{u}(\mathbf{x},t)}{t} \right). \tag{3.4}$$

By Lemma 2.1-ii) we have that $\partial H^* \left(\frac{\mathbf{x}-\mathbf{u}(\mathbf{x},t)}{t} \right) = \left\{ \nabla H^* \left(\frac{\mathbf{x}-\mathbf{u}(\mathbf{x},t)}{t} \right) \right\}$, and by Lemma 2.1-iii), we obtain that for any $t>0$ and any $\mathbf{x} \in \mathbb{R}^n$ that $\nabla H^*(\frac{\mathbf{x}-\mathbf{u}(\mathbf{x},t)}{t}) \in \partial J(\mathbf{u}(\mathbf{x},t))$. This implies that the subdifferential of (3.4) is a singleton. In other words it is differentiable and its partial derivative with respect to $\mathbf{x}$ is given by

$$\nabla_{\mathbf{x}} E(\mathbf{x},t) = \nabla H^* \left(\frac{\mathbf{x} - \mathbf{u}(\mathbf{x},t)}{t} \right) \in \partial J(\mathbf{u}(\mathbf{x},t)). \tag{3.5}$$

In addition, under the assumptions (H1)-(H4) with (H5) (the proofs can be adapted to cope with the other set of assumptions, i.e., (H1)-(H4) with (H5')-(H7')) we can invoke [28, Corollary 4.1.4, p. 82] or [18, Prop. 23.2, p. 289] to obtain for any $\mathbf{y} \in \operatorname{dom} H^*$ such that $\partial H^*(\mathbf{y}) \neq \emptyset$ the following formula about the differentials

$$DH^*(\mathbf{y}) = (DH)^{-1}(\mathbf{y}). \tag{3.6}$$

Using (3.6) and Definition 2.5 into (3.5), we get

$$\forall \mathbf{x} \in \mathbb{R}^n \ \forall t>0 \quad \mathbf{u}(\mathbf{x},t) = \mathbf{x} - t\nabla H \left(\nabla_{\mathbf{x}} E(\mathbf{x},t) \right).$$

Thus, we have proved the following result.

Proposition 3.1. *Suppose assumptions (H1)-(H4) along with either (H5) or (H5')-(H7') hold. Then, for any $\mathbf{x} \in \mathbb{R}^n$ and any $t > 0$ we have*

$$\mathbf{u}(\mathbf{x},t) = \mathbf{x} - t\nabla H \left(\nabla_{\mathbf{x}} E(\mathbf{x},t) \right). \tag{3.7}$$

The above formula states that the behavior of the minimizers are completely dictated by the spatial derivative $\nabla_{\mathbf{x}} E$ of the solution of (H-J) and the gradient of the Hamiltonian. To the best of our knowledge, this relationship that links the restored image $\mathbf{u}(\mathbf{x},t)$ with the minimal value of the imaging problem (2.2) was unknown.

However, proposition 3.1 is valid only for smoothing parameters $t>0$. The next propositions study the behavior of the minimizers when $t\to 0$ with $t>0$. For this purpose we shall consider any sequence of positive smoothing parameter $t_k>0$ converging to 0. We also consider the sequence of observed data $(\mathbf{x}+t_k\mathbf{d}_k)_{k\in\mathbb{N}}$ where $\mathbf{x}\in\mathbb{R}^n$ and $\mathbf{d}_k\in\mathbb{R}^n$ for any $k\in\mathbb{N}$ such that $\lim_{k\to+\infty}\mathbf{d}_k=\mathbf{d}$ with $\mathbf{d}\in\mathbb{R}^n$. In other words, we consider sequences of the form $(\mathbf{x}+t_k\mathbf{d}_k,t_k)_{k\in\mathbb{N}}$ that are converging from the direction $(\mathbf{d},1)$ to $(\mathbf{x},0)$ [34, p. 197].

Our goal is to prove that the sequences $(\mathbf{u}(\mathbf{x}+t_k\mathbf{d}_k,t_k))_{k\in\mathbb{N}}$, $(E(\mathbf{x}+t_k\mathbf{d}_k,t_k))_{k\in\mathbb{N}}$ and $(\nabla_{\mathbf{x}}E(\mathbf{x}+t_k\mathbf{d}_k,t_k)))_{k\in\mathbb{N}}$ are convergent and to provide explicit formulas of these limits.

The next proposition describes the behavior of the minimizer when the smoothing parameter vanishes. It shows that the minimizer of the imaging problem converges to the final observed data $\mathbf{x}$. This means that vanishing smoothing parameters corresponds vanishing noise.

Proposition 3.2. *Suppose assumptions (H1)-(H4) along with either (H5) or (H5')-(H7') hold. Let $(t_k)_{k\in\mathbb{N}}$ be a sequence of positive real numbers converging to 0 and let $(\mathbf{d}_k)_{k\in\mathbb{N}}$ be a sequence of elements of $\mathbb{R}^n$ converging to $\mathbf{d}\in\mathbb{R}^n$. Then, the following properties hold:*

i) For any $\mathbf{x}\in\operatorname{dom}J$ we have for $\mathbf{u}$ given by (3.7)

$$\lim_{k\to+\infty}\mathbf{u}(\mathbf{x}+t_k\mathbf{d}_k,t_k)=\mathbf{x}.$$

ii) For any $\mathbf{x}\in\operatorname{dom}J$ such that $\partial J(\mathbf{x})\neq\emptyset$ then

$$\text{the sequence}\left(\frac{\mathbf{x}+t_k\mathbf{d}_k-\mathbf{u}(\mathbf{x}+t_k\mathbf{d}_k,t_k)}{t_k}\right)_{k\in\mathbb{N}}\text{ is bounded.}$$

Part-ii) of the above proposition is a technical result that is useful to prove the next proposition. It shows that the sequence $(\nabla E_{\mathbf{x}}(\mathbf{x}+t_k\mathbf{d}_k,t_k))_{k\in\mathbb{N}}$ is bounded. It also provides the set where any accumulation point of $(\nabla E_{\mathbf{x}}(\mathbf{x}+t_k\mathbf{d}_k,t_k))_{k\in\mathbb{N}}$ lives.

Proposition 3.3. *Suppose assumptions (H1)-(H4) with either (H5) or (H5')-(H7') hold. Let $(t_k)_{k\in\mathbb{N}}$ be a sequence of positive real numbers converging to 0 and $(\mathbf{d}_k)_{k\in\mathbb{N}}$ be a sequence of elements of $\mathbb{R}^n$ converging to $\mathbf{d}\in\mathbb{R}^n$. Then, the sequence $(\nabla E_{\mathbf{x}}(\mathbf{x}+t_k\mathbf{d}_k,t_k))_{k\in\mathbb{N}}$ is bounded and any accumulation point $\mathbf{q}$ satisfies*

$$\mathbf{q}\in\partial J(\mathbf{x}).$$

We can refine the above proposition by looking at the semiderivates (see [34, Def. 7.20, p. 256]) of E for $t=0$. The next proposition describes the limits of the variations of E and $\nabla_{\mathbf{x}}E$ in function of the data and the smoothing parameter as the latter vanishes and the data converges to $\mathbf{x}$ along the direction $\mathbf{d}$. It provides explicit formulas of the limits of these variations in function of H and J.

Proposition 3.4. *Suppose assumptions (H1)-(H4) along with either (H5) or (H5')-(H7') hold. Let $\mathbf{x}\in\operatorname{dom}J$ such that $\partial J(\mathbf{x})\neq\emptyset$. Let $(t_k)_{k\in\mathbb{N}}$ be a sequence of positive* *real numbers converging to 0 and let* $(\mathbf{d}_k)_{k\in\mathbb{N}}$ *be a sequence of elements of* $\mathbb{R}^n$ *converging to* $\mathbf{d}\in\mathbb{R}^n$*. Then, the semiderivative of* E *at* $(\mathbf{x},0)$ *along the direction* $(\mathbf{d},1)$ *is given by*

$$\lim_{k\to+\infty}\frac{E(\mathbf{x}+t_k\mathbf{d}_k,t_k)-E(\mathbf{x},0)}{t_k}=\max_{\mathbf{y}\in\partial J(\mathbf{x})}\{\langle\mathbf{d},\mathbf{y}\rangle\ -H(\mathbf{y})\}.$$

Furthermore,

$$\lim_{k\to+\infty}\nabla_{\mathbf{x}}E(\mathbf{x}+t_k\mathbf{d}_k,t_k)=\arg\max_{\mathbf{y}\in\partial J(\mathbf{x})}\{\langle\mathbf{d},\mathbf{y}\rangle\ -H(\mathbf{y})\}.$$

The above proposition shows that $\nabla E_{\mathbf{x}}$ and the variations of E correspond to the maximum deviation between the linear form $\langle\mathbf{d},\cdot\rangle$ and H over the closed convex set $\partial J(\mathbf{x})$ as the smoothing parameter t vanishes. It gives thus explicit formulas for the role played by the image prior J on the minimal value E and its variations as $t\to 0$.

This constraint $\partial J(\mathbf{x})$ shows how that image processing prior J acts on the minimal value and its variation of the image processing problem (2.2).

Let us make some remarks for the particular case of the quadratic Hamiltonian $H=\frac{1}{2}\|\cdot\|_2^2$ in (H-J). In this case, for any fixed $\mathbf{x}\in\mathbb{R}^n$ the mapping $(0,+\infty)\ni t\mapsto \nabla_{\mathbf{x}}E(\mathbf{x},t)$ is called the "Yosida approximation" of the subdifferential ∂J [2, Chp. 3, p. 144]. We apply Proposition 3.4 with the direction $(\mathbf{0},1)$ and we get

$$\lim_{k\to+\infty}\nabla_{\mathbf{x}}E(\mathbf{x},t_k)=\pi_{\partial J(\mathbf{x})}(\mathbf{0}),\tag{3.8}$$

where $\pi_{\partial J(\mathbf{x})}(\mathbf{0})$ corresponds to the element of the subdifferential of J at the point $\mathbf{x}$ that has the minimal Euclidean norm (i.e., the Euclidean projection of $\mathbf{0}$ onto the convex closed set $\partial J(\mathbf{x})$). The result given by (3.8) on Yosida approximation is widely known and can be found for instance in [9, Thm. 3.1, p. 54] and [2, Thm. 2, p 144].

‘

3.3. Proof of Proposition 3.2

We set $\mathbf{x}_k=\mathbf{x}+t_k\mathbf{d}_k$.

Case with (H5).

Proof of i). Since for all $k\in\mathbb{N}$, $\mathbf{u}(\mathbf{x}_k,t_k)$ is the minimizer of (2.5), we have

$$\forall\mathbf{y}\in\mathbb{R}^n\ J(\mathbf{u}(\mathbf{x}_k,t_k))+t_kH^*\left(\frac{\mathbf{x}_k-\mathbf{u}(\mathbf{x}_k,t_k)}{t_k}\right)\le J(\mathbf{y})+t_kH^*\left(\frac{\mathbf{x}_k-\mathbf{y}}{t_k}\right).$$

In particular, for $\mathbf{y}=\mathbf{x}$ we get for any $k\in\mathbb{N}$

$$J(\mathbf{u}(\mathbf{x}_k,t_k))+t_kH^*\left(\frac{\mathbf{x}_k-\mathbf{u}(\mathbf{x}_k,t_k)}{t_k}\right)\le J(\mathbf{x})+t_kH^*\left(\frac{\mathbf{x}_k-\mathbf{x}}{t_k}\right).\tag{3.9}$$

By (H5), H is 1-coercive and by invoking [28, Prop. 1.3.8, p. 46] we get that $\operatorname{dom}H^*=\mathbb{R}^n$. Since H^* is also convex it is continuous on $\mathbb{R}^n$. In addition, since $\mathbf{x}_k=\mathbf{x}+t_k\mathbf{d}_k$ we deduce that $\frac{\mathbf{x}_k-\mathbf{x}}{t_k}\to\mathbf{d}$ as $k\to+\infty$. We then obtain by the continuity of H^* that $\lim_{k\to+\infty}H^*(\frac{\mathbf{x}_k-\mathbf{x}}{t_k})=H^*(\mathbf{d})$. Thus, there exists $c_0\in\mathbb{R}$ such that for any $k\in\mathbb{N}$ $H^*(\frac{\mathbf{x}_k-\mathbf{x}}{t_k})\le c_0$. From (3.9) we get that for any $k\in\mathbb{N}$

$$J(\mathbf{u}(\mathbf{x}_k,t_k))+t_kH^*\left(\frac{\mathbf{x}_k-\mathbf{u}(\mathbf{x}_k,t_k)}{t_k}\right)\le J(\mathbf{x})+t_kc_0.\tag{3.10}$$

Since $\operatorname{dom} J \neq \emptyset$ and J is convex, there exists $\mathbf{y}_0 \in \operatorname{dom} J$ such that $\partial J(\mathbf{y}_0) \neq \emptyset$. Let $s_{\mathbf{y}_0} \in \partial J(\mathbf{y}_0)$. Using the convex inequality and the Cauchy-Schwarz inequality we get

$$J(\mathbf{u}(\mathbf{x}_k,t_k)) \geq J(\mathbf{y}_0) + \langle \mathbf{s}_{\mathbf{y}_0}, \mathbf{u}(\mathbf{x}_k,t_k) - \mathbf{y}_0 \rangle \geq J(\mathbf{y}_0) - \|\mathbf{s}_{\mathbf{y}_0}\|_2 \, \|\mathbf{u}(\mathbf{x}_k,t_k) - \mathbf{y}_0\|_2 \,, \tag{3.11}$$

which yields when combined with (3.10) to

$$t_k H^* \left(\frac{\mathbf{x}_k - \mathbf{u}(\mathbf{x}_k,t_k)}{t_k} \right) \leq J(\mathbf{x}) + t_k c_0 - J(\mathbf{y}_0) + \|\mathbf{s}_{\mathbf{y}_0}\|_2 \, \|\mathbf{u}(\mathbf{x}_k,t_k) - \mathbf{y}_0\|_2 \,. \tag{3.12}$$

By the triangle inequality, we have that for any $k \in \mathbb{N}, \|\mathbf{u}(\mathbf{x}_k,t_k) - \mathbf{y}_0\|_2 \leq \|\mathbf{y}_0 - \mathbf{x}_k\|_2 + \|\mathbf{x}_k - \mathbf{u}(\mathbf{x}_k,t_k)\|_2$. Since the sequence $(\mathbf{x}_k)_{k \in \mathbb{N}}$ converges to $\mathbf{x}$, it is bounded and so is $(\|\mathbf{y}_0 - \mathbf{x}_k\|_2)_{k \in \mathbb{N}}$. Also, the sequence $(t_k c_0)_{k \in \mathbb{N}}$ is bounded since it converges to 0. Thus, there exists a constant $c_1 \in \mathbb{R}$ such that for any $k \in \mathbb{N}$ it holds $J(\mathbf{x}) + t_k c_0 - J(\mathbf{y}_0) + \|\mathbf{s}_{\mathbf{y}_0}\|_2 \|\mathbf{y}_0 - \mathbf{x}_k\|_2 \leq c_1$. Thus, we obtain that for any $k \in \mathbb{N}$

$$t_k H^* \left(\frac{\mathbf{x}_k - \mathbf{u}(\mathbf{x}_k,t_k)}{t_k} \right) \leq c_1 + \|\mathbf{s}_{\mathbf{y}_0}\|_2 \, \|\mathbf{x}_k - \mathbf{u}(\mathbf{x}_k,t_k)\|_2 \,. \tag{3.13}$$

Suppose by contradiction that $(\|\mathbf{x}_k - \mathbf{u}(\mathbf{x}_k,t_k)\|_2)_{k \in \mathbb{N}}$ does not converge to 0, i.e.,

$$\exists \varepsilon > 0, \forall k_0 \in \mathbb{N}, \exists l \in \mathbb{N}, \, l \geq k_0, \, \|\mathbf{x}_l - \mathbf{u}(\mathbf{x}_l,t_l)\|_2 > \varepsilon \,.$$

Dividing by $\|\mathbf{x}_l - \mathbf{u}(\mathbf{x}_l,t_l)\|_2 \geq \varepsilon > 0$ in (3.13) we get

$$\frac{t_l}{\|\mathbf{x}_l - \mathbf{u}(\mathbf{x}_l,t_l)\|_2} H^* \left(\frac{\mathbf{x}_l - \mathbf{u}(\mathbf{x}_l,t_l)}{t_l} \right) \leq \frac{c_1}{\|\mathbf{x}_l - \mathbf{u}(\mathbf{x}_l,t_l)\|_2} + \|\mathbf{s}_{\mathbf{y}_0}\|_2 \leq \frac{c_1}{\epsilon} + \|\mathbf{s}_{\mathbf{y}_0}\|_2 \,. \tag{3.14}$$

Since $(t_k)_{k \in \mathbb{N}}$ is converging to 0 and $\|\mathbf{x}_l - \mathbf{u}(\mathbf{x}_l,t_l)\|_2 \geq \varepsilon$ we can make the quantity $\frac{\|\mathbf{x}_l - \mathbf{u}(\mathbf{x}_l,t_l)\|_2}{t_l}$ as large as desired by choosing k_0 sufficiently large. By (H1), i.e., $\operatorname{dom} H = \mathbb{R}^n$ and [28, Prop. 1.3.9, p. 46] we have that H^* is 1-coercive, i.e., there exists a constant $c_2 \in \mathbb{R}$ such that for any $\mathbf{z} \in \mathbb{R}^n$ with $\|\mathbf{z}\|_2 \geq c_2$ it holds $H^*(\mathbf{z}) \geq (\frac{c_1}{\varepsilon} + \|\mathbf{s}_{\mathbf{y}_0}\|_2 + 1)\|\mathbf{z}\|_2$. Letting $\mathbf{z} = \frac{\mathbf{x}_l - \mathbf{u}(\mathbf{x}_l,t_l)}{t_l}$ with k_0 large enough such that $\|\mathbf{z}\|_2 \geq c_2$ we get

$$\frac{t_l}{\|\mathbf{x}_l - \mathbf{u}(\mathbf{x}_l,t_l)\|_2} H^* \left(\frac{\mathbf{x}_l - \mathbf{u}(\mathbf{x}_l,t_l)}{t_l} \right) \geq \frac{c_1}{\varepsilon} + \|\mathbf{s}_{\mathbf{y}_0}\|_2 + 1$$

which contradicts (3.14). Thus, $(\|\mathbf{x}_k - \mathbf{u}(\mathbf{x}_k,t_k)\|_2)$ converges to 0. Since $\|\mathbf{x} - \mathbf{u}(\mathbf{x}_k,t_k)\|_2 \leq \|\mathbf{x}_k - \mathbf{u}(\mathbf{x}_k,t_k)\|_2 + t_k \|\mathbf{d}_k\|_2$, we get the desired result by taking the limit on $k \to +\infty$, i.e., $\lim_{k \to +\infty} \mathbf{u}(x_k,t_k) = \mathbf{x}$. The proof of i) is complete.

Proof of ii). By assumption we have $\mathbf{x} \in \operatorname{dom} J$ and $\partial J(\mathbf{x}) \neq \emptyset$. Thus we can choose $\mathbf{y}_0 = \mathbf{x}$ and $\mathbf{s}_0 \in \partial J(\mathbf{x})$ in (3.11). Thus, inequality (3.12) becomes

$$t_k H^* \left(\frac{\mathbf{x}_k - \mathbf{u}(\mathbf{x}_k,t_k)}{t_k} \right) \leq J(\mathbf{x}) + t_k c_0 - J(\mathbf{x}) + \|\mathbf{s}_0\|_2 \, \|\mathbf{u}(\mathbf{x}_k,t_k) - \mathbf{x}\|_2 \,.$$

We apply the triangle inequality $\|\mathbf{u}(\mathbf{x}_k,t_k) - \mathbf{x}\|_2 \leq \|\mathbf{u}(\mathbf{x}_k,t_k) - \mathbf{x}_k\|_2 + \|\mathbf{x}_k - \mathbf{x}\|_2$ to get

$$t_k H^* \left(\frac{\mathbf{x}_k - \mathbf{u}(\mathbf{x}_k,t_k)}{t_k} \right) \leq t_k c_0 + \|\mathbf{s}_0\|_2 \, \|\mathbf{u}(\mathbf{x}_k,t_k) - \mathbf{x}_k\|_2 + \|\mathbf{s}_0\|_2 \, \|\mathbf{x} - \mathbf{x}_k\|_2 \,.$$

We divide by $t_k > 0$ and since $\mathbf{x}_k = \mathbf{x} + t_k \mathbf{d}_k$, we obtain

$$H^*\left(\frac{\mathbf{x}_k - \mathbf{u}(\mathbf{x}_k, t_k)}{t_k}\right) \leq \|\mathbf{s}_0\|_2 \frac{\|\mathbf{u}(\mathbf{x}_k, t_k) - \mathbf{x}_k\|_2}{t_k} + c_0 + \|\mathbf{s}_0\|_2 \|\mathbf{d}_k\|_2.$$

We can bound the sequence $(c_0 + \|\mathbf{s}_0\|_2 \|\mathbf{d}_k\|_2)_{k\in\mathbb{N}}$ from above by $c_3 \in \mathbb{R}$ (since $\mathbf{d}_k \to \mathbf{d}$ and $\|\mathbf{d}_k\|_2 \to \|\mathbf{d}\|_2$ by continuity of the $l_2(\mathbb{R}^n)$ norm, when $k \to +\infty$) to obtain

$$H^*\left(\frac{\mathbf{x}_k - \mathbf{u}(\mathbf{x}_k, t_k)}{t_k}\right) \leq \|\mathbf{s}_0\|_2 \frac{\|\mathbf{u}(\mathbf{x}_k, t_k) - \mathbf{x}_k\|_2}{t_k} + c_3 . \tag{3.15}$$

Suppose by contradiction that the sequence $\left(\frac{\mathbf{x}_k - \mathbf{u}(\mathbf{x}_k, t_k)}{t_k}\right)_{k\in\mathbb{N}}$ is not bounded, i.e.,

$$\forall \varepsilon > 0, \forall k_0 \in \mathbb{N}, \exists l \in \mathbb{N}, l \geq k_0, \left\|\frac{\mathbf{x}_k - \mathbf{u}(\mathbf{x}_k, t_k)}{t_k}\right\|_2 \geq \epsilon. \tag{3.16}$$

Since H^* is 1-coercive, there exists a constant $c_4 \in \mathbb{R}$ such that for any $\mathbf{z} \in \mathbb{R}^n$ with $\|\mathbf{z}\|_2 \geq c_4$ it holds $H^*(\mathbf{z}) \geq (1 + \|\mathbf{s}_0\|_2)\|\mathbf{z}\|_2$. Choosing $\varepsilon \geq \max\{c_4, c_3\}$ and letting $\mathbf{z} = \frac{\mathbf{x}_l - \mathbf{u}(\mathbf{x}_l, t_l)}{t_l}$ we get

$$(\|\mathbf{s}_0\|_2 + 1)\left\|\frac{\mathbf{x}_l - \mathbf{u}(\mathbf{x}_l, t_l)}{t_k}\right\|_2 \leq H^*\left(\frac{\mathbf{x}_l - \mathbf{u}(\mathbf{x}_l, t_l)}{t_k}\right).$$

Together with (3.15) this gives

$$1 + c_3 \leq \varepsilon \leq \left\|\frac{\mathbf{x}_l - \mathbf{u}(\mathbf{x}_l, t_l)}{t_k}\right\|_2 \leq c_3 ,$$

which is a contradiction. Thus $\left(\frac{\mathbf{x}_k - \mathbf{u}(\mathbf{x}_k, t_k)}{t_k}\right)_{k\in\mathbb{N}}$ is bounded.

Case with (H5')-(H7'). Since for all $k \in \mathbb{N}$, $\mathbf{u}(\mathbf{x}_k, t_k)$ is the minimer of (2.5), we have for any $\mathbf{y} \in \mathbb{R}^n$

$$J(\mathbf{u}(\mathbf{x}_k, t_k)) + t_k H^*\left(\frac{\mathbf{x}_k - \mathbf{u}(\mathbf{x}_k, t_k)}{t_k}\right) \leq J(\mathbf{y}) + t_k H^*\left(\frac{\mathbf{x}_k - \mathbf{y}}{t_k}\right).$$

In particular, for $\mathbf{y} = \mathbf{x}_k$ we get that for any $k \in \mathbb{N}$

$$J(\mathbf{u}(\mathbf{x}_k, t_k)) + t_k H^*\left(\frac{\mathbf{x}_k - \mathbf{u}(\mathbf{x}_k, t_k)}{t_k}\right) \leq J(\mathbf{x}_k) + t_k H^*(\mathbf{0}) . \tag{3.17}$$

The quantity $H^*(\mathbf{0})$ is finite since by (H6') H is assumed to be bounded from below by a constant. Since J is a real-valued convex function on $\mathbb{R}^n$ we have that $\partial J(\mathbf{y}) \neq \emptyset$ for any $\mathbf{y} \in \mathbb{R}^n$. Let $\mathbf{s}_k \in \partial J(\mathbf{x}_k)$. Using the convex inequality and the Cauchy-Schwarz inequality we get that

$$J(\mathbf{u}(\mathbf{x}_k, t_k)) \geq J(\mathbf{x}_k) + \langle \mathbf{s}_k, \mathbf{u}(\mathbf{x}_k, t_k) - \mathbf{x}_k \rangle \geq J(\mathbf{x}_k) - \|\mathbf{s}_k\|_2 \|\mathbf{u}(\mathbf{x}_k, t_k) - \mathbf{x}_k\|_2 , \tag{3.18}$$

which yields when combined with (3.17)

$$t_k H^*\left(\frac{\mathbf{x}_k - \mathbf{u}(\mathbf{x}_k, t_k)}{t_k}\right) \leq t_k H^*(\mathbf{0}) + \|\mathbf{s}_k\|_2 \|\mathbf{u}(\mathbf{x}_k, t_k) - \mathbf{x}_k\|_2 . \tag{3.19}$$

By assumption, $\mathbf{x}_k \to \mathbf{x}$ when $k \to +\infty$ which implies that the sequence $(\mathbf{x}_k)_{k\mathbb{N}}$ is bounded. Thus, there exists a convex compact set C of $\mathbb{R}^n$ such that $\mathbf{x}_k \in C$ for any $k \in \mathbb{N}$. By (H7'), we have that $\operatorname{dom} J = \mathbb{R}^n$. Then [27, Thm. 3.1.2, p. 174] implies that the function J restricted on the convex compact set C, i.e., $J|_C : C \to \mathbb{R}$, is Lipschitz with some constant L. This yields that $\|\mathbf{s}_k\|_2 \leq L$ for any $k \in \mathbb{N}$. We get

$$t_k H^* \left(\frac{\mathbf{x}_k - \mathbf{u}(\mathbf{x}_k, t_k)}{t_k} \right) \leq t_k H^*(\mathbf{0}) + L \, \|\mathbf{u}(\mathbf{x}_k, t_k) - \mathbf{x}_k\|_2 . \tag{3.20}$$

From there we proceed as in the previous case. We assume that the results do not hold and we obtain again a contradicton by using the 1-coercivity of H^*.

3.4. Proof of Proposition 3.3

We note $\mathbf{x}_k = \mathbf{x} + t_k \mathbf{d}_k$.

Case with (H5). By Proposition 3.2, we have that the sequence $\left(\frac{\mathbf{x}_k - \mathbf{u}(\mathbf{x}_k, t_k)}{t_k} \right)_{k\mathbb{N}}$ is bounded. Thus, we can find a compact convex set $C \subset \mathbb{R}^n$, e.g., a closed ball in R^n, such that for any $k \in \mathbb{N}$ we have $\frac{\mathbf{x}_k - \mathbf{u}(\mathbf{x}_k, t_k)}{t_k} \in C$. By (H1), i.e., $\operatorname{dom} H = \mathbb{R}^n$, and [28, Prop. 1.3.8, p.46] we have that $\operatorname{dom} H^* = \mathbb{R}^n$. This yields that the restriction of H^* to C, i.e., $H^*|_C : C \to \mathbb{R}$, is convex and Lipschitz of constant L by invoking [27, Thm. 3.1.2, p. 174]. For any $\mathbf{y} \in C$, we thus have $\|\nabla H^*(\mathbf{y})\|_2 \leq L$. Recall that equation (3.5) states that $\nabla_{\mathbf{x}} E(\mathbf{x}_k, t_k) = \nabla H^* \left(\frac{\mathbf{x}_k - \mathbf{u}(\mathbf{x}_k, t_k)}{t_k} \right)$. Thus, we have that for any $k \in \mathbb{N}$ $\|\nabla_{\mathbf{x}} E(\mathbf{x}_k, t_k)\|_2 = \left\| \nabla H^* \left(\frac{\mathbf{x}_k - \mathbf{u}(\mathbf{x}_k, t_k)}{t_k} \right) \right\|_2 \leq L$, which shows that $(\nabla_{\mathbf{x}} E(\mathbf{x}_k, t_k))_{k\in\mathbb{N}}$ is bounded.

Since the sequence $(\nabla_{\mathbf{x}} E(\mathbf{x}_k, t_k))_{k\in\mathbb{N}}$ is bounded in $\mathbb{R}^n$, we can extract a subsequence that converges to $\mathbf{q} \in \mathbb{R}^n$ by the Bolzano-Weierstrass theorem. This subsequence is denoted by $(\nabla_{\mathbf{x}} E(\mathbf{x}_{\psi(k)}, t_{\psi(k)}))_{k\in\mathbb{N}}$ with $\psi : \mathbb{N} \to \mathbb{N}$ increasing. By Lemma 2.1-iii) we have that $\partial J(\mathbf{u}(\mathbf{x}, t))$ is non-empty. According to the Fenchel-Legendre inequality for J and since $\nabla_{\mathbf{x}} E(\mathbf{y}, t) \in \partial J(\mathbf{u}(\mathbf{y}, t))$ for any $\mathbf{y} \in \mathbb{R}^n$, we thus have

$$\forall \mathbf{y} \in \mathbb{R}^n, J(\mathbf{y}) \geq J(\mathbf{u}(\mathbf{x}_{\psi(k)}, t_{\psi(k)})) + \langle \nabla_{\mathbf{x}} E(\mathbf{x}_{\psi(k)}, t_{\psi(k)}), \mathbf{y} - \mathbf{u}(\mathbf{x}_{\psi(k)}, t_{\psi(k)}) \rangle \ .$$

According to Propoposition 3.2 $\lim_{k\to+\infty} \mathbf{u}(\mathbf{x}_k, t_k) = \mathbf{u}(\mathbf{x}_{\psi(k)}, t_{\psi(k)}) = \mathbf{x}$ and since $(\nabla E_{\mathbf{x}}(\mathbf{x}_{\psi(k)}, t_{\psi(k)}))_{k\in\mathbb{N}} \to \mathbf{q}$ as $k \to +\infty$ we have

$$\forall \mathbf{y} \in \mathbb{R}^n, \lim_{k\to+\infty} \langle \nabla_x E(\mathbf{x}_{\psi(k)}, t_{\psi(k)}), \mathbf{y} - \mathbf{u}(\mathbf{x}_{\psi(k)}, t_{\psi(k)}) \rangle = \langle \mathbf{q}, \mathbf{y} - \mathbf{x} \rangle \ ,$$

by the continuity of the Euclidean scalar product. In addition, since J is lower semicontinuous, we get

$$\forall \mathbf{y} \in \mathbb{R}^n, J(\mathbf{y}) \geq \liminf_{k\to+\infty} \left\{ J(\mathbf{u}(\mathbf{x}_{\psi(k)}, t_{\psi(k)})) + \langle \nabla_x E(\mathbf{x}_{\psi(k)}, t_{\psi(k)}), \mathbf{y} - \mathbf{u}(\mathbf{x}_{\psi(k)}, t_{\psi(k)}) \rangle \right\},$$

hence

$$\forall \mathbf{y} \in \mathbb{R}^n, J(\mathbf{y}) \geq J(\mathbf{x}) + \langle \mathbf{q}, \mathbf{y} - \mathbf{x} \rangle,$$

that is $\mathbf{q} \in \partial J(\mathbf{x})$.

Case with (H5')-(H7'). We only show that the sequence $(\nabla_{\mathbf{x}} E(\mathbf{x} + t_k \mathbf{d}_k, t_k))_{k\in\mathbb{N}}$ is bounded since the rest of the proof is the same as in the previous case.

By Proposition 3.2, $\mathbf{u}(\mathbf{x}_k, t_k) \to \mathbf{x}$ as $k \to +\infty$. Thus we can find a compact $C \subset \mathbb{R}^n$, e.g., a closed ball of $\mathbb{R}^n$, such that $\mathbf{u}(\mathbf{x}_k, t_k) \in C$ for any $k \in \mathbb{N}$. Recall that (H7') states that $\operatorname{dom} J = \mathbb{R}^n$. Thus, by [27, Thm. 3.1.2, p. 174] the restriction of $J: \mathbb{R}^n \to \mathbb{R}$ to C, i.e., $J|_C : C \to \mathbb{R}$ is Lipschitz with some constant L. Thus, for any $\mathbf{s} \in \partial J(\mathbf{u}(\mathbf{x}_k, t_k))$ we have that $\|\mathbf{s}\|_2 \leq L$. Since for any $t > 0$, $\nabla_{\mathbf{x}} E(\mathbf{x}, t) \in \partial J(\mathbf{u}(\mathbf{x}, t))$, we get that $\|\nabla_{\mathbf{x}} E(\mathbf{x} + t_k \mathbf{d}_k, t_k)\|_2 \leq L$.

3.5. Proof of Proposition 3.4

We proceed in three steps: the first two steps respectively consist in bounding from below $\liminf_{k\to+\infty} \frac{E(\mathbf{x}+t_k\mathbf{d}_k,t_k)-E(\mathbf{x},0)}{t_k}$ and bounding from above $\limsup_{k\to+\infty} \frac{E(\mathbf{x}+t_k\mathbf{d}_k,t_k)-E(\mathbf{x},0)}{t_k}$. The third step shows that these two bounds are actually equal. Thus the quantity $\frac{E(\mathbf{x}+t_k\mathbf{d}_k,t_k)-E(\mathbf{x},0)}{t_k}$ converges as $k \to +\infty$. In addition, this last step provides the limit of $\nabla_{\mathbf{x}} E(\mathbf{x} + t_k \mathbf{d}_k, t_k)$.

Step 1. By the Hopf formula we have that for any $k \in \mathbb{N}$ and any $\mathbf{x} \in \operatorname{dom} J$ such that $\partial J(\mathbf{x}) \neq \emptyset$

$$E(\mathbf{x}+t_k\mathbf{d}_k,t_k) = \sup_{\mathbf{y}\in\mathbb{R}^n} \left\{ \langle \mathbf{x}+t_k\mathbf{d}_k, \mathbf{y}\rangle - J^*(\mathbf{y}) - t_k H(\mathbf{y}) \right\}.$$

Thus, for any $\mathbf{y} \in \mathbb{R}^n$ and any $k \in \mathbb{N}$ we have

$$E(\mathbf{x}+t_k\mathbf{d}_k,t_k) \geq \langle \mathbf{x}+t\mathbf{d}_k, \mathbf{y}\rangle - J^*(\mathbf{y}) - t_k H(\mathbf{y}).$$

Since $E(\cdot,0) = J(\cdot)$, for any $\mathbf{y} \in \mathbb{R}^n$ and any $k \in \mathbb{N}$ we deduce

$$E(\mathbf{x}+t_k\mathbf{d}_k,t_k) - E(\mathbf{x},0) \geq \langle \mathbf{x},\mathbf{y}\rangle - J^*(\mathbf{y}) - J(\mathbf{x}) + t_k \langle \mathbf{d}_k, \mathbf{y}\rangle - t_k H(\mathbf{y}). \tag{3.21}$$

By assumption, we have $\partial J(\mathbf{x}) \neq \emptyset$. We choose $\mathbf{y} \in \partial J(\mathbf{x})$. We invoke [28, Thm. 1.4.1, p.47] to get

$$J^*(\mathbf{y}) + J(\mathbf{x}) - \langle \mathbf{y}, \mathbf{x}\rangle = 0.$$

Combining this equality with inequality (3.21) we get for any $\mathbf{y} \in \partial J(\mathbf{x})$ and any $k \in \mathbb{N}$

$$\frac{E(\mathbf{x}+t_k\,\mathbf{d}_k,t_k) - E(\mathbf{x},0)}{t_k} \geq \langle \mathbf{d}_k, \mathbf{y}\rangle - H(\mathbf{y}).$$

We take the limit inferior as $k \to +\infty$. The right-hand side has actually a limit since we have that $\mathbf{d}_k \to \mathbf{d}$ and by continuity of the Euclidean scalar product we obtain that $\langle \mathbf{d}_k, \mathbf{y}\rangle \to \langle \mathbf{d}, \mathbf{y}\rangle$ as $k \to +\infty$. Thus, we get for any $\mathbf{y} \in \partial J(\mathbf{x})$

$$\liminf_{k\to+\infty} \frac{E(\mathbf{x}+t_k\,\mathbf{d}_k,t_k) - E(\mathbf{x},0)}{t_k} \geq \liminf_{k\to+\infty} \left(\langle \mathbf{d}_k, \mathbf{y}\rangle - H(\mathbf{y}) \right) = \langle \mathbf{d}, \mathbf{y}\rangle - H(\mathbf{y}).$$

We thus obtain

$$\liminf_{k\to+\infty} \frac{E(\mathbf{x}+t_k\,\mathbf{d}_k,t_k) - E(\mathbf{x},0)}{t_k} \geq \sup_{\mathbf{y}\in\partial J(\mathbf{x})} \left\{ \langle \mathbf{d}, \mathbf{y}\rangle - H(\mathbf{y}) \right\}.$$

Since $H : \mathbb{R}^n \to \mathbb{R}$ is assumed to be strictly convex, coercive, i.e., $\lim_{\|\mathbf{x}\|_2\to+\infty} H(\mathbf{x}) \to +\infty$, we have that the supremum in the right hand side is attained at some $\hat{\mathbf{y}} \in \partial J(\mathbf{x})$ and is unique, i.e.,

$$\hat{\mathbf{y}} = \arg\max_{\mathbf{y}\in\partial J(\mathbf{x})} \left\{ \langle \mathbf{d}, \mathbf{y}\rangle - H(\mathbf{y}) \right\} \tag{3.22}$$

and

$$\liminf_{k\to+\infty}\frac{E(\mathbf{x}+t_k\,\mathbf{d}_k,t_k)-E(\mathbf{x},0)}{t_k}\geq\max_{\mathbf{y}\in\partial J(\mathbf{x})}\{\langle\mathbf{d},\mathbf{y}\rangle-H(\mathbf{y})\}=\langle\mathbf{d},\hat{\mathbf{y}}\rangle-H(\hat{\mathbf{y}})\,. \tag{3.23}$$

Step 2. Set any $\mathbf{e}\in\mathbb{R}^n$, any $\mathbf{x}\in\operatorname{dom}J$ such that $\partial J(\mathbf{x})\neq\emptyset$. Let us introduce the function $\phi:[0,+\infty)\to\mathbb{R}$ defined by

$$\phi:t\mapsto E(\mathbf{x}+t\,\mathbf{e},t).$$

The function ϕ is convex since it is the composition of E convex with an affine mapping ([27, Prop. 2.1.5, p. 159]). It is also differentiable in $(0,+\infty)$. For any $t>0$ the derivative ϕ' of ϕ is given by

$$\phi'(t)=\langle\nabla_x E(\mathbf{x}+t\,\mathbf{e},t),\mathbf{e}\rangle+\frac{\partial E}{\partial t}(\mathbf{x}+t\mathbf{e},t)\,.$$

By the Hamilton-Jacobi equation, we have that for any $t>0$ $\frac{\partial E}{\partial t}(\mathbf{x}+t\,\mathbf{e},t)=-H(\nabla_{\mathbf{x}}E(\mathbf{x}+t\mathbf{e},t))$ and thus

$$\phi'(t)=\langle\nabla_x E(\mathbf{x}+t\mathbf{e},t),\mathbf{d}\rangle-H(\nabla_{\mathbf{x}}E(\mathbf{x}+t\mathbf{d},t))\,.$$

Since ϕ is convex we have that for any $t>0$

$$\frac{\phi(t)-\phi(0)}{t}\leq\phi'(t)\,.$$

This yields that for any $k\in\mathbb{N}$

$$\frac{E(\mathbf{x}+t_k\,\mathbf{d}_k,t_k)-E(\mathbf{x},0)}{t_k}\leq\langle\nabla_x E(\mathbf{x}+t_k\mathbf{d}_k,t_k),\mathbf{d}_k\rangle-H(\nabla_{\mathbf{x}}E(\mathbf{x}+t_k\,\mathbf{d}_k,t_k))\,. \tag{3.24}$$

By proposition 3.3 we have that the sequence $(\nabla_x E(\mathbf{x}+t_k\,\mathbf{d}_k,t_k))_{k\in\mathbb{N}}$ of $\mathbb{R}^n$ is bounded. Since we are working in $\mathbb{R}^n$ with the usual topology induced by the Euclidean metric, Bolzano-Weierstrass theorem yields that the sequence $(\nabla_x E(\mathbf{x}+t_k\,\mathbf{d}_k,t_k))_{k\in\mathbb{N}}$ has a convergent subsequence. Let $\mathbf{q}\in\mathbb{R}^n$ be the limit of a convergent sub-sequence of $(\nabla_{\mathbf{x}}E(\mathbf{x},t_k))_{k\in\mathbb{N}}$ (that we also denote by $(\nabla_{\mathbf{x}}E(\mathbf{x},t_k))_{k\in\mathbb{N}}$). By proposition 3.3 we have that $\mathbf{q}\in\partial J(\mathbf{x})$. We take to the limit superior on $k\to+\infty$ in (3.24). We have that $\nabla_x E(\mathbf{x}+t_k\,\mathbf{d}_k,t_k)\to\mathbf{q}$ (since we have extracted the subsequence with limit $\mathbf{q}$). Futhermore, by assumption we have that $\mathbf{d}_k\to\mathbf{d}$, which gives $\langle\nabla_x E(\mathbf{x}+t_k\mathbf{d}_k,t_k),\mathbf{d}_k\rangle\to\langle\mathbf{q},\mathbf{d}\rangle$ by continuity of the Euclidean scalar product. By continuity of H we get that $H(\nabla_{\mathbf{x}}E(\mathbf{x}+t_k\,\mathbf{d}_k,t_k))\to H(\mathbf{q})$. Thus we obtain

$$\limsup_{k\to+\infty}\frac{E(\mathbf{x}+t_k\,\mathbf{d}_k,t_k)-E(\mathbf{x},0)}{t_k}\leq\langle\mathbf{q},\mathbf{d}\rangle-H(\mathbf{q})\,. \tag{3.25}$$

Step 3. From inequalities (3.23) and (3.25) we get

$$\begin{aligned}\langle\mathbf{d},\hat{\mathbf{y}}\rangle-H(\hat{\mathbf{y}})\leq\liminf_{k\to+\infty}\frac{E(\mathbf{x}+t_k\,\mathbf{d}_k,t_k)-E(\mathbf{x},0)}{t_k}\leq\\ \limsup_{k\to+\infty}\frac{E(\mathbf{x}+t_k\,\mathbf{d}_k,t_k)-E(\mathbf{x},0)}{t_k}\leq\langle\mathbf{q},\mathbf{d}\rangle-H(\mathbf{q}).\end{aligned} \tag{3.26}$$

However, recall that $\langle \mathbf{d},\hat{\mathbf{y}}\rangle - H(\hat{\mathbf{y}}) = \max_{\mathbf{y}\in\partial J(\mathbf{x})}\{\langle \mathbf{y},\mathbf{d}\rangle - H(\mathbf{y})\}$. This yields

$$\max_{\mathbf{y}\in\partial J(\mathbf{x})}\{\langle \mathbf{d},\mathbf{y}\rangle - H(\mathbf{y})\} = \langle \mathbf{d},\hat{\mathbf{y}}\rangle - H(\hat{\mathbf{y}}) \leq \langle \mathbf{q},\mathbf{d}\rangle - H(\mathbf{q})\,, \tag{3.27}$$

where $\hat{\mathbf{y}}\in\partial J(\mathbf{x})$ is given by (3.22) and is the unique element that realizes the maximum in the left-hand side quantity of (3.27). Since $\mathbf{q}\in\partial J(\mathbf{x})$, we necessarily have that $\mathbf{q} = \hat{\mathbf{y}}$ where $\hat{\mathbf{y}}$ is given by (3.22). Indeed, suppose by contradiction that $\mathbf{q} \neq \hat{\mathbf{y}}$. This would yield that $\hat{\mathbf{y}}$ is not the unique maximizer of (3.27). This would be in contradiction with step 1. Thus, we have shown that

$$\max_{\mathbf{y}\in\partial J(\mathbf{x})}\{\langle \mathbf{d},\hat{\mathbf{y}}\rangle - H(\hat{\mathbf{y}}\} = \langle \mathbf{q},\mathbf{d}\rangle - H(\mathbf{q})\,. \tag{3.28}$$

Combining (3.26) and (3.28) yields

$$\lim_{k\to+\infty}\frac{E(\mathbf{x}+t_k\,\mathbf{d}_k,t_k) - E(\mathbf{x},0)}{t_k} = \langle \mathbf{d},\hat{\mathbf{y}}\rangle - H(\hat{\mathbf{y}}).$$

The same argument also shows that for any convergent subsequence of the bounded sequence $(\nabla_x E(\mathbf{x}+t_k\,\mathbf{d}_k,t_k))_{k\in\mathbb{N}}$ the limit is $\hat{\mathbf{y}}$. Recall that the sequence $(\nabla_x E(\mathbf{x}+t_k\,\mathbf{d}_k,t_k))_{k\in\mathbb{N}}$ of $\mathbb{R}^n$ is bounded and we conclude that this sequence converges to $\hat{\mathbf{y}}$, that is

$$\lim_{k\to+\infty}\nabla_x E(\mathbf{x}+t_k\,\mathbf{d}_k,t_k) = \arg\max_{\mathbf{y}\in\partial J(\mathbf{x})}\{\langle \mathbf{d},\mathbf{y}\rangle - H(\mathbf{y})\}\,.$$

4. Conclusion

This work has described some original connections between convex optimization problems in image processing and Hamilton-Jacobi equations. The striking new fact is that the minimal values of these problems are solutions of Hamilton-Jacobi equations. The initial datum corresponds to the prior while the Hamiltonian is related to the data fidelity term. Explicit formulas give the dependence of the minimizers with respect to the observed images $\mathbf{x}\in\mathbb{R}^n$ and the smoothing parameter t. Not only the analysis provides explicit formulas for $t>0$ but also when $t\to 0$.

Obviously, the proposed formalism and formulas apply to vector valued signals/images (e.g., colors, hyperspectral) but also cope with more sophisticated convex models (e.g, nuclear norm, non-local and/or vector valued convex priors).

As a byproduct, this study shows that the solution of certain initial-valued Hamilton-Jacobi problems in many space dimensions can be computed at some points using optimization solvers already developed for imaging purposes.

This paper has only considered the case of invertible matrices A in the image formation model given by (1.1). In [15], non-invertible matrices A are considered and the approach described here is extended to compressive sensing and ℓ_1 related optimization problems.

Appendix A. Proof of Lemma 2.1.

Proof of i). We proceed in two steps: first we show the existence of a minimizer and then its uniqueness.

For any $\mathbf{x}\in\mathbb{R}^n$ and any $t>0$, define the function $G:\mathbb{R}^n\to\mathbb{R}$ by

$$G:\mathbf{y}\mapsto J(\mathbf{y}) + tH^*\left(\frac{\mathbf{x}-\mathbf{y}}{t}\right) \tag{1.1}$$

attains its minimum at a unique point $\bar{\mathbf{y}} \in \mathbb{R}^n$.

Step 1, existence of a minimizer. By assumption (H1), $\operatorname{dom} H = \mathbb{R}^n$. Thus, by invoking [28, Prop. 1.3.9 (ii), p. 46], H^* is 1-coercive. Consequently, for any $a \in \mathbb{R}$ there exists $c \in \mathbb{R}$ such that for any $\mathbf{y} \in \mathbb{R}^n$ with $\|\mathbf{y}\|_2 \geq c$ we have $tH^*\left(\frac{\mathbf{x}-\mathbf{y}}{t}\right) \geq a\|\mathbf{x}-\mathbf{y}\|_2$. By (H4), there exists a point $\mathbf{x}_0 \in \mathbb{R}^n$ such that $\partial J(\mathbf{x}_0) \neq \emptyset$. Let $\mathbf{s} \in \partial J(\mathbf{x}_0)$. Using the convex inequality for J, the Cauchy-Schwarz inequatlity and the triangle inequality, we get that for any $\mathbf{y} \in \mathbb{R}^n$ $J(\mathbf{y}) \geq J(\mathbf{x}_0) + \langle \mathbf{s}, \mathbf{y}-\mathbf{x}_0\rangle \geq J(\mathbf{x}_0) - \|\mathbf{s}\|_2\|\mathbf{y}-\mathbf{x}_0\|_2 \geq J(\mathbf{x}_0) - \|\mathbf{s}\|_2(\|\mathbf{y}-\mathbf{x}\|_2 + \|\mathbf{x}-\mathbf{x}_0\|_2)$. This yields that for any $\mathbf{y} \in \mathbb{R}^n$ with $\|\mathbf{y}\|_2 \geq c$ we have

$$J(\mathbf{x}_0) - \|\mathbf{s}\|_2\|\mathbf{x}-\mathbf{x}_0\|_2 + (a - \|s\|_2)\|\mathbf{x}-\mathbf{y}\|_2 \leq G(\mathbf{y}).$$

By choosing a large enough such $a > \|\mathbf{s}\|_2$, we get the $\lim_{\|\mathbf{y}\|_2 \to +\infty}(a - \|\mathbf{s}\|_2)\|\mathbf{x}-\mathbf{y}\|_2 = +\infty$. Thus, we have that $\lim_{\|\mathbf{y}\|_2 \to +\infty} G(\mathbf{y}) = +\infty$. Futhermore, $G \in \Gamma_0(\mathbb{R}^n)$ since it is the sum of two lower semicontinuous functions. Thus, since G is lower semicontinuous and coercive, it has a minimizer $\bar{\mathbf{y}}$.

Step 2, uniqueness of the minimizer. By (H2), H is differentiable and by invoking [28, Thm. 4.1.3, p. 81], H^* is strictly convex, and so is $tH^*\left(\frac{\mathbf{x}-\cdot}{t}\right)$. J is convex by (H4). Thus G is strictly convex as the sum of a convex function and a strictly convex function. The concludes the proof of i).

Proof of ii). We need to show that $\nabla H^*\left(\frac{\mathbf{x}-\bar{\mathbf{y}}}{t}\right)$ exists.

Case with (H5). By (H5), H is 1-coercive, and thus $\operatorname{dom} H^* = \mathbb{R}^n$ by invoking [28, Thm. 4.1.3, p. 81]. By the strict convexity of H given by (H3) and by [28, Thm. 4.1.1, p. 81] we get that H^* is differentiable on $\mathbb{R}^n$. Thus $\nabla H^*\left(\frac{\mathbf{x}-\bar{\mathbf{y}}}{t}\right)$ exists.

Case with (H5')-(H7'). Recall that $(H5')$ states that H^* is differentiable at point whenever it has a subgradient, i.e., a non-empty subdifferential. It thus enough to prove that $\partial H^*\left(\frac{\mathbf{x}-\bar{\mathbf{y}}}{t}\right) \neq \emptyset$.

Suppose by contradiction that $\partial H^*\left(\frac{\mathbf{x}-\bar{\mathbf{y}}}{t}\right) = \emptyset$. This means that for any $\mathbf{s} \in \mathbb{R}^n$, there exists $\mathbf{y}_0 \in \mathbb{R}^n$ such that

$$H^*\left(\frac{\mathbf{x}-\mathbf{y}_0}{t}\right) < H^*\left(\frac{\mathbf{x}-\bar{\mathbf{y}}}{t}\right) + \langle \mathbf{s}, \frac{\mathbf{x}-\mathbf{y}_0}{t} - \frac{\mathbf{x}-\bar{\mathbf{y}}}{t}\rangle. \tag{1.2}$$

Since $\bar{\mathbf{y}}$ is the unique minimizer of G we have

$$J(\bar{\mathbf{y}}) + tH^*\left(\frac{\mathbf{x}-\bar{\mathbf{y}}}{t}\right) \leq J(\mathbf{y}_0) + tH^*\left(\frac{\mathbf{x}-\mathbf{y}_0}{t}\right). \tag{1.3}$$

Combining (1.2) and (1.3) we obtain

$$J(\bar{\mathbf{y}}) + tH^*\left(\frac{\mathbf{x}-\bar{\mathbf{y}}}{t}\right) < J(\mathbf{y}_0) + tH^*\left(\frac{\mathbf{x}-\bar{\mathbf{y}}}{t}\right) + \langle \mathbf{s}, \bar{\mathbf{y}} - \mathbf{y}_0\rangle.$$

Hence, for any $\mathbf{s} \in \mathbb{R}^n$ there exists $\mathbf{y}_0 \in \mathbb{R}^n$ such that

$$J(\bar{\mathbf{y}}) < J(\mathbf{y}_0) + \langle \mathbf{s}, \bar{\mathbf{y}} - \mathbf{y}_0\rangle. \tag{1.4}$$

By assumption (H7') we have $\operatorname{dom} J = \mathbb{R}^n$, and thus $J(\mathbf{y}_0) < +\infty$. We choose $\mathbf{s} \in \partial J(\mathbf{y}_0)$ in (1.4) and this contradicts the convexity of J. This concludes the proof of ii).

Proof of iii). We need to show that $\nabla H^*(\frac{\mathbf{x}-\bar{\mathbf{y}}}{t}) \in \partial J(\bar{\mathbf{y}})$.

Since H^* is differentiable at $\frac{\mathbf{x}-\bar{\mathbf{y}}}{t}$, there exists an open ball $B\left(\frac{\mathbf{x}-\bar{\mathbf{y}}}{t},\epsilon\right) = \{\mathbf{z} \in \mathbb{R}^n \,|\, \|\mathbf{z} - \frac{\mathbf{x}-\bar{\mathbf{y}}}{t}\|_2 < \epsilon\}$ of radius $\epsilon > 0$ centered at $\frac{\mathbf{x}-\bar{\mathbf{y}}}{t}$, such that $B\left(\frac{\mathbf{x}-\bar{\mathbf{y}}}{t},\epsilon\right) \subset \operatorname{dom} H^*$.

Let $\mathbf{y} \in \operatorname{dom} J$ and $\alpha \in (0,1)$ such that $\alpha\frac{\bar{\mathbf{y}}-\mathbf{y}}{t} \in B\left(\frac{\mathbf{x}-\bar{\mathbf{y}}}{t},\epsilon\right)$. Since $\bar{\mathbf{y}}$ is the unique minimizer of G, we have

$$J(\bar{\mathbf{y}}) + tH^*\left(\frac{\mathbf{x}-\bar{\mathbf{y}}}{t}\right) \leq J(\alpha\mathbf{y} + (1-\alpha)\bar{\mathbf{y}}) + tH^*\left(\frac{\mathbf{x}-(\alpha\mathbf{y}+(1-\alpha)\bar{\mathbf{y}})}{t}\right).$$

Since J is convex, we have $J(\alpha\mathbf{y}+(1-\alpha)\bar{\mathbf{y}}) \leq \alpha J(\mathbf{y}) + (1-\alpha)J(\bar{\mathbf{y}})$ and thus

$$J(\bar{\mathbf{y}}) + tH^*\left(\frac{\mathbf{x}-\bar{\mathbf{y}}}{t}\right) \leq \alpha J(\mathbf{y}) + (1-\alpha)J(\bar{\mathbf{y}}) + tH^*\left(\frac{\mathbf{x}-(\alpha\mathbf{y}+(1-\alpha)\bar{\mathbf{y}})}{t}\right).$$

We divide by $\alpha > 0$ to obtain

$$0 \leq J(\mathbf{y}) - J(\bar{\mathbf{y}}) + \frac{tH^*\left(\frac{\mathbf{x}-\bar{\mathbf{y}}-\alpha(\mathbf{y}-\bar{\mathbf{y}})}{t}\right) - tH^*\left(\frac{\mathbf{x}-\bar{\mathbf{y}}}{t}\right)}{\alpha}.$$

Recall that H^* is differentiable at the point $\frac{\mathbf{x}-\bar{\mathbf{y}}}{t}$. Thus when we take the limit as $\alpha \to 0$, we get

$$0 \leq J(\mathbf{y}) - J(\bar{\mathbf{y}}) + \langle t\nabla H^*\left(\frac{\mathbf{x}-\bar{\mathbf{y}}}{t}\right), -\frac{\mathbf{y}-\bar{\mathbf{y}}}{t}\rangle.$$

We deduce that for any $\mathbf{y} \in \operatorname{dom} J$

$$J(\bar{\mathbf{y}}) + \langle \nabla H^*\left(\frac{\mathbf{x}-\bar{\mathbf{y}}}{t}\right), \mathbf{y}-\bar{\mathbf{y}}\rangle \leq J(\mathbf{y}).$$

This means exactly that $\nabla H^*\left(\frac{\mathbf{x}-\bar{\mathbf{y}}}{t}\right) \in \partial J(\bar{\mathbf{y}})$.

Acknowledgements. This work has been supported by ONR-N00014-11-1-0749. The author deeply thanks Stanley Osher (UCLA Math. Dpt) for his support, fruitful discussions and his patience over the years. Without him this paper would have remained as “the novel of the character Joseph Grand in The Plague” written by Albert Camus. The author deeply thanks Mickael Chekroun (UCLA ATMOS Dpt) and Igor Ciril (IPSA, Paris), Antonio Marquina (Univ. de Valencia), Susana Serna (Univ. Autonoma de Barcelona), Marc Sigelle (Télécom ParisTech), Sylvain Peyronnet (GREYC, Univ. Caen) and Yohann Tendero (UCLA Math. Dpt) for fruitful discussions and careful readings of drafts. He also thanks Arjuna Flenner and Gary Hewer (Naval Air Weapon Center, China Lake) for fruitful discussions. Part of this work has been presented at the Second Monterey Workshop on Computational Issues in Nonlinear Control, Monterey, CA, November 2011.

REFERENCES

[1] G. Aubert and P. Kornprobst. *Mathematical Problems in Image Processing.* Springer-Verlag, 2002.

[2] J.-P. Aubin and A. Cellina. *Differential Inclusions.* Springer-Verlag, 1984.

[3] M. Bardi and I. Dolcetta. *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*. Birkhuser, 2010.

[4] M. Bardi and L.C. Evans. On Hopf's formulas for solutions of Hamilton-Jacobi equations. *Nonlinear Analysis: Theory, Methods & Applications*, 8(11):1373 – 1381, 1984.

[5] M. Bardi and S. Osher. The nonconvex multi-dimensional Riemann problem for Hamilton-Jacobi equations. *SIAM Journal on Mathematical Analysis*, 22(2):344–351, March 1991.

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

[7] F. Bonnans and A. Shapiro. *Perturbation analysis of optimization problems*. Springer, 200.

[8] C. Bouman and K. Sauer. A generalized gaussian image model for edge-preserving map estimation. *IEEE Transactions on Transactions on Signal Processing*, 2(3):296–310, July 1993.

[9] H. Brezis. *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*. North-Holland, 1973.

[10] E. Candes., J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on*, 52(2):489–509, 2006.

[11] A. Chambolle and J. Darbon. On total variation minimization and surface evolution using parametric maximum flows. *International Journal of Computer Vision*, 84(3):288–307, 2009.

[12] P.L. Combettes and V.R. Wajs. Theoretical analysis of some regularized image denoising methods. In *Proceedings of the IEEE International Conference on Image Processing*, volume 1, pages 969–972, 2004.

[13] M.G. Crandall, L. Evans, and P.-L. Lions. Some properties of viscosity solutions of Hamilton-Jacobi equations. *Transactions of the American Mathematical Society*, 282(2), 1984.

[14] M.G. Crandall and P.-L. Lions. Viscosity solutions of Hamilton-Jacobi equations. *Transactions of the American Mathematical Society*, 277(1):1–42, May 1983.

[15] J. Darbon and S. Osher. Initial value problems for Hamilton-Jacobi equations and sparsity for linear systems via ℓ_1 related optimization. Technical report, UCLA Cam report, 2013.

[16] J. Darbon and M. Sigelle. Image restoration with discrete constrained Total Variation part I: Fast and exact optimization. *Journal of Mathematical Imaging and Vision*, 26(3):261–276, December 2006.

[17] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics*, 57(11):1413–1457, 2004.

[18] K. Deimling. *Nonlinear Functional Analysis*. Dover, 2010.

[19] D.L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

[20] I. Ekeland and R. Temam. *Convex Analysis and Variational Problems*. North-Holland, 1976.

[21] L.C. Evans. *Partial Differential Equations (Graduate Studies in Mathematics, V. 19) GSM/19*. American Mathematical Society, June 1998.

[22] M.A.T. Figueiredo and R.D. Nowak. Bayesian wavelet-based signal estimation using non-informative priors. In *Signals, Systems amp; Computers, 1998. Conference Record of the Thirty-Second Asilomar Conference on*, volume 2, pages 1368–1373, 1998.

[23] G. Gilboa and S. Osher. Nonlocal linear image regularization and supervised segmentation. *SIAM jounal on Multiscale Modeling and Simulation*, 6(2):595–630, 2007.

[24] A. Haddad. *Méthodes variationnelles en traitement des images*. PhD thesis, Ecole Normale Supérieure de Cachan, 2005.

[25] A. Haddad. Stability in a class of variational methods. *Applied and Computational Harmonic Analysis*, 23(1):57 – 73, 2007.

[26] J.-B. Hiriart-Urruty. *Optimisation et Analyse Convexe*. Presse Universitaire de France, 1998.

[27] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms Part I*. Springer Verlag, Heidelberg, 1996.

[28] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms Part II*. Springer Verlag, Heidelberg, 1996.

[29] D.S. Hochbaum. An efficient algorithm for image segmentation, Markov Random Fields and related problems. *Journal of the ACM*, 48(2):686–701, July 2001.

[30] S. Kassam and J. Thomas. *Signal detection in non-Gaussian noise*. Springer-Verlag, 1988.

[31] P.-L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16(6):964–979, 1979.

[32] J.J. Moreau. Proximité et dualité dans un espace hilbertien. *Bulletin de la S.M.F.*, 93:273–299, 1965.

[33] R.T. Rockafellar. *Convex Analysys*. Princeton University Press, 1996.
[34] R.T. Rockafellar and R. Wets. *Variational Analysis*. Springer-Verlag, 1998.
[35] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D.*, 60:259–268, 1992.
[36] A. Shapiro. On differentiability of metric projections in $\mathbb{R}^n$. I. boundary case. *Proceedings of the American Mathematical Society*, 99:123–128, 1987.
[37] G. Winkler. *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*. Applications of mathematics. Springer-Verlag, 2003.