

ON DECOMPOSITION MODELS IN IMAGING SCIENCES AND MULTI-TIME HAMILTON-JACOBI PARTIAL DIFFERENTIAL EQUATIONS

JÉRÔME DARBON AND TINGWEI MENG

ABSTRACT. This paper provides new theoretical connections between multi-time Hamilton-Jacobi partial differential equations and variational image decomposition models in imaging sciences. We show that the minimal values of these optimization problems are governed by multi-time Hamilton-Jacobi partial differential equations. The minimizers of these optimization problems can be represented using the momentum in the corresponding Hamilton-Jacobi partial differential equation. Moreover, variational behaviors of both the minimizers and the momentum are investigated as the regularization parameters approach zero. In addition, we provide a new perspective from convex analysis to prove the uniqueness of convex solutions to Hamilton-Jacobi equations. Finally, we consider image decomposition models that do not have unique minimizers and we propose a regularization approach to perform the analysis using multi-time Hamilton-Jacobi partial differential equations.

1. INTRODUCTION

In the late 20th century, the Hamilton-Jacobi (HJ) equation was widely studied in the field of partial differential equations (PDEs). To be specific, the solution $S(x, t)$ defined for $x \in \mathbb{R}^n$, $t \geq 0$ satisfies the following Cauchy problem

$$\begin{cases} \frac{\partial S(x, t)}{\partial t} + H(x, t, S(x, t), \nabla_x S(x, t)) = 0, & x \in \mathbb{R}^n, t > 0; \\ S(x, 0) = J(x), & x \in \mathbb{R}^n, \end{cases}$$

where H is the Hamiltonian and J is the initial data. When the Hamiltonian only depends on the spatial gradient $\nabla_x S(x, t)$, under some regularity and convexity assumptions, the solution is given by the Hopf formula or Lax formula [18, 68]

$$S(x, t) = \sup_{p \in \mathbb{R}^n} \langle p, x \rangle - J^*(p) - tH(p) \quad (\text{Hopf formula})$$

$$= \inf_{u \in \mathbb{R}^n} J(u) + tH^* \left(\frac{x - u}{t} \right) \quad (\text{Lax formula})$$

where J^* and H^* are the Legendre transform of the functions J and H , respectively. From the physics point of view, HJ PDE describes the movement of a particle in a physics model whose energy function is given by the Hamiltonian H . To be specific, the variables x and t are the current position and time of the particle. The characteristic line of the PDE gives the trajectory of the particle. The momentum is given by the spatial gradient $\nabla_x S(x, t)$ which coincides with the maximizer in the Hopf formula. The velocity is given by $\frac{x-u}{t}$ where u is the minimizer in the Lax formula.

We refer the readers to the review paper [48] for thorough details and [49, 69] for connections between convex analysis and HJ equations. An extension of this PDE is to consider the time variable t in a higher dimensional space \mathbb{R}^N , in which case the PDE system is called the multi-time Hamilton-Jacobi equation, first discussed by Rochet from an economic point of view [80]. Later, Lions and Rochet [71] considered the multi-time HJ equations when the Hamiltonians are convex functions which only depend on the momentum. They proposed the generalized Hopf formula by writing it as the composition of several semigroups of the corresponding single-time HJ operators. Following their work, several existence and uniqueness results [20, 32, 73, 78, 88] were provided in more general cases, for example, when the Hamiltonians have spatial or time dependence.

It is well known that the HJ equation has a deep relationship with optimal control [26] and differential games [57, 84]. Later, Darbon [49] provided a representation formula for the minimizers of a specific kind

of optimization problem, which relates the minimizers to the spatial gradients of the solutions to the HJ equations. As we will see below, many models in imaging sciences can be viewed from a perspective of HJ PDEs. Following that work, we generalize the results to multi-time HJ equations and a larger set of optimization problems, including the decomposition models in image processing.

In the past few decades, many decomposition models have been proposed in image processing. These models are applied to different practical problems, such as inpainting [23, 56], image classification [12], and road detection [62]. Here, we give a brief overview of convex variational models in this area. There are many models that cannot be fully listed here, for which we refer the readers to [44, 61].

The basic idea of image decomposition is to regard an image x as a summation of several components $\{u_j\}$, and solve the following minimization problem:

$$(1) \quad \arg \min_{u_0 + \dots + u_N = x} f_0(u_0) + \sum_{j=1}^N \lambda_j f_j(u_j).$$

Here, each function f_j is designed to characterize the corresponding component u_j . One may tune the parameters $\{\lambda_j\}$ to put emphasis on different components. There are many celebrated decomposition models in the literature of imaging sciences. In the introduction we mention the continuous versions of the models, while later in the main part of this paper we will work with their discrete versions. The first widely used decomposition model is the Rudin-Osher-Fatemi (ROF) model, proposed in [83], which applies the total variation (TV) semi-norm and $\|\cdot\|_{L^2}^2$ to recognize the geometry and noise in an image, respectively. In the continuous setting, for any function $u \in L^1(\Omega)$ and $\Omega \subset \mathbb{R}^2$, the TV semi-norm of u is defined by

$$\|u\|_{TV} := \sup \left\{ \int_{\Omega} u(x) \operatorname{div} \phi(x) dx : \phi \in C_c^1(\Omega, \mathbb{R}^2), \|\phi\|_{L^\infty} \leq 1 \right\}.$$

Here and after in the introduction, the derivatives and divergence are in the distribution sense. The space $BV(\Omega)$ is the space containing all functions of bounded variation, defined by

$$BV(\Omega) = \{u \in L^1(\Omega) : \|u\|_{TV} < +\infty\}.$$

Under these settings, the ROF model solves the following problem

$$\arg \min_{u \in BV(\Omega)} \|u\|_{TV} + \frac{1}{2\lambda} \|x - u\|_{L^2}^2.$$

The mathematical analysis for the ROF model is provided in [1, 2, 3, 4, 7, 28, 29, 33, 34, 35, 36, 38, 40, 41, 46, 47, 51, 63, 64, 76, 79, 89, 91]. Later, Meyer [72] pointed out the disadvantage of $\|\cdot\|_{L^2}^2$ in capturing oscillating patterns. In order to overcome this disadvantage, he suggested using the norm in either of the three spaces E, F, G to replace it, where these three spaces are defined as follows. We use the notations of Meyer to describe these spaces [72]. First, define the space of functions of bounded mean oscillation (BMO) by

$$BMO := \left\{ f \in L_{loc}^1(\mathbb{R}^n) : \sup \left\{ \frac{1}{|Q|} \int_Q |f(x) - f_Q| dx : Q \text{ is any ball in } \mathbb{R}^n \right\} < +\infty \right\}$$

$$\text{where the symbol } f_Q \text{ is defined by } f_Q := \frac{1}{|Q|} \int_Q f(x) dx,$$

and the homogeneous Besov space $\dot{B}_1^{1,1}$ by

$$\dot{B}_1^{1,1} := \left\{ f \in L^{\frac{n}{n-1}}(\mathbb{R}^n) : \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}^n} |c(j, k)| 2^{j(1-n/2)} < +\infty, \right.$$

$$\left. \text{where } \{c(j, k)\} \text{ are the wavelet coefficients of } f \right\}.$$

Let $\dot{B}_\infty^{-1,\infty}$ be the dual space of $\dot{B}_1^{1,1}$. Then, define E, F, G by $E := \dot{B}_\infty^{-1,\infty}$, $F := \operatorname{div}(BMO)$ and $G := \operatorname{div}(L^\infty)$. To be specific, the space G and G -norm are defined as follows

$$G := \{f = \partial_1 g_1 + \partial_2 g_2 : g_1, g_2 \in L^\infty(\mathbb{R}^2)\},$$

$$\|f\|_G := \inf \{ \|(g_1^2 + g_2^2)^{1/2}\|_{L^\infty} : f = \partial_1 g_1 + \partial_2 g_2 \}.$$

The space F is similarly defined by replacing the space L^∞ in the above definition with the BMO space. The corresponding models proposed by Meyer are stated as follows

$$(2) \quad \arg \min_{u \in BV(\Omega)} \|u\|_{TV} + \lambda \|x - u\|_X, \quad \text{where the space } X \text{ can be } E, F \text{ or } G.$$

For mathematical analysis of these models, we refer the readers to [59, 62, 70]. In [59], the space E is also generalized to any homogeneous Besov space $\dot{B}_p^{\alpha-2,q}$, where $p, q \in [1, +\infty]$ and $\alpha \in (0, 2)$. However, Meyer's models are hard to solve numerically. There are mainly two approaches to numerically solve the model with G -norm. The first approach is approximating L^∞ in the definition of G by L^p [90]. Osher et al. [77] proposed an equivalent formulation called OSV when $p = 2$. In a word, OSV uses the square of H^{-1} -norm instead of G -norm. To be specific, the OSV model solves

$$\arg \min_{u \in BV(\Omega)} \|u\|_{TV} + \lambda \int_{\Omega} |\nabla(\Delta)^{-1}(x - u)|^2 dx dy.$$

The other approach called A^2BC model is proposed by Aujol et al. [9, 10], replacing the G -norm with the indicator function of balls in the space G . In other words, it solves the following problem

$$(3) \quad \arg \min_{u \in BV(\Omega)} \|u\|_{TV} + I\{\|x - u\|_G \leq \mu\},$$

where $I\{\cdot\}$ denotes the indicator function whose definition will be given in section 2. It is shown that this A^2BC model gives the solution to Meyer's model eq. (2) with $X = G$ when the parameter μ is appropriately chosen. In practice, they use a Moreau-Yosida type approximation and solve the following problem instead

$$(4) \quad \arg \min_{u \in BV(\Omega), v \in G} \|u\|_{TV} + I\{\|v\|_G \leq \mu\} + \frac{1}{2\lambda} \|x - u - v\|_{L^2}^2,$$

This regularized model converges to eq. (3) as the parameter λ approaches zero. Moreover, it is easy to implement using Chambolle's projection method [37]. Similarly, in [11], the indicator function of the E -ball is used to replace the E -norm, which provides a similar numerical implementation approach to the Meyer's model eq. (2) with $X = E$.

In the above models, an image is decomposed into a geometrical part and an oscillating part. However, for a noisy image, the oscillating part may contain both the texture in the original image and the noise. To split these two parts, a $u + v + w$ model is proposed in [11], which constrains the G -norm of the texture part and the E -norm of the noisy part. Later, Gilles [60] modified the $u + v + w$ model with a coefficient assigned to each pixel to smoothly indicate whether it is in texture or noise. He also modified the A^2BC model by requiring the G -norm of the noise to be much smaller than the G -norm of the texture. In [15, 53, 54], the authors extended some of the abovementioned models, which are originally proposed for gray-scale images, to color images. Besides, there are many other functions used in image decomposition. For example, the L^1 -norm [5, 14, 42, 75] is used to promote sparsity or remove salt and pepper noise. In [13, 14], the quadratic form $\langle \cdot, K \cdot \rangle$, where K is a linear symmetric positive operator, is used for adaptive kernel selection of the texture component. Note that this quadratic form generalizes the L^2 term in ROF and the H^{-1} term in OSV.

The previous work [49] clarifies the relationship between single-time HJ equations and decomposition models with two terms (i.e. $N = 1$ in eq. (1)), such as the ROF model, Meyer's models and some of their variations. However, as mentioned above, there are many other models handling three or more components. Also, in practice, one may modify a model by adding a quadratic term for numerical consideration, such as in eq. (4). This kind of modification is applied to most of the above models. As a result, the objective function in the numerical implementation actually contains three or more terms. On the other hand, new models can be constructed by regarding the functions mentioned above as building blocks and combining them together. For instance, the morphological component analysis [58, 85, 86] combines ROF model and L^1 minimization for the coefficients with respect to two sets of dictionaries chosen for the representation of texture and geometry. Another example is [45], which adds a higher order term $\alpha \|\Delta v\|_{L^2}^2$ to the models introduced above, in order to reduce the staircase effect. Actually, the higher order terms in image processing are widely studied in the literature. Two important models are the TV-TV² infimal convolution model [41] and the Total Generalized Variation (TGV) model [25]. In fact, after discretization, the higher order linear operators are discretized using some matrices. In other words, the results in this paper can be applied to the discrete models with higher order terms by regarding them as matrix multiplication. In conclusion, it is

valuable to generalize the previous work [49] and provide a framework to analyze the models involving more than two components. Also, our proposed framework is suitable for a large class of discrete decomposition models in imaging sciences, even including some models containing higher order terms.

Now, we briefly introduce the intuition and the basic setup for our framework and demonstrate the idea using some experimental results of the discrete A^2BC model. In general, for a discrete decomposition model eq. (1), an image is regarded as a vector $x \in \mathbb{R}^n$, where n is the number of pixels. If we can relate each f_j , $j \geq 1$, to a Hamiltonian and f_0 to an initial function, then the minimal value, regarded as a function of the input data x and the parameters $\{\lambda_j\}$, relates to the solution of the corresponding multi-time HJ equation. Here, the parameters $\{\lambda_j\}$ are regarded as time variables.

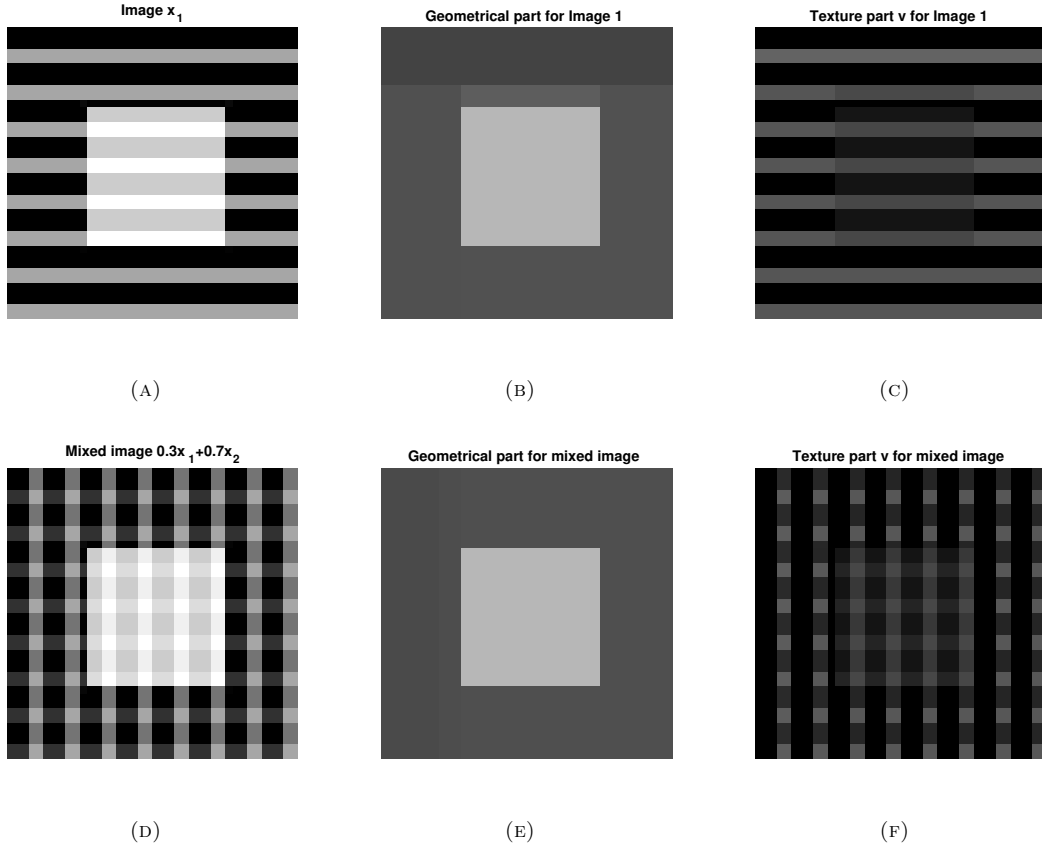


FIGURE 1. The A^2BC model is applied to an artificial image. The original image x_1 and the corresponding minimizers u, v are shown in (a)-(c). The convex combination $0.3x_1 + 0.7x_2$ of x_1 and its rotation x_2 is shown in (d), whose minimizers are shown in (e)-(f).

For example, the discrete A^2BC model solves the following optimization problem:

$$(5) \quad S(x, \mu, \lambda) := \min_{u, v \in \mathbb{R}^n} J(u) + J^* \left(\frac{v}{\mu} \right) + \frac{1}{2\lambda} \|x - u - v\|_2^2.$$

The desired quantities are the minimizers, denoted as $u(x, \mu, \lambda)$ and $v(x, \mu, \lambda)$. Here, the discrete total variation semi-norm $J : \mathbb{R}^{m_1 \times m_2} \rightarrow \mathbb{R}$ is defined as follows

$$(6) \quad J(u) := \sum_{i=1}^{m_1-1} \sum_{j=1}^{m_2-1} |u_{i+1,j} - u_{i,j}| + |u_{i,j+1} - u_{i,j}|.$$

In this paper, we identify the space $\mathbb{R}^{m_1 \times m_2}$ containing all matrices with m_1 rows and m_2 columns with the Euclidean space \mathbb{R}^n where $n = m_1 m_2$. The discrete total variation J defined above is the anisotropic

version, which will be used in this paper. Its Legendre transform J^* is the indicator function of the unit ball in the dual space. To be specific, let $\|\cdot\|_G$ be the dual norm of J , which is given by

$$\|v\|_G = \inf \left\{ \sup_{\substack{1 \leq i \leq m_1 \\ 1 \leq j \leq m_2}} \sqrt{(g_{i,j})^2 + (h_{i,j})^2} : \begin{array}{l} v_{i,j} = g_{i,j} - g_{i-1,j} + h_{i,j} - h_{i,j-1}, \\ g_{0,j} = g_{m_1,j} = h_{i,0} = h_{i,m_2} = 0, \quad g_{i,j}, h_{i,j} \in \mathbb{R} \quad \forall 1 \leq i \leq m_1, 1 \leq j \leq m_2 \end{array} \right\}.$$

Then, we have $J^*(p) = I\{\|p\|_G \leq 1\}$ for any $p \in \mathbb{R}^n$ where $I\{\cdot\}$ denotes the indicator function. Notice that any indicator function is invariant under multiplication with a positive constant, then we have $\mu J^* = J^*$. Hence, the above optimization problem is equivalent to

$$S(x, \mu, \lambda) = \min_{u, v \in \mathbb{R}^n} J(u) + \mu J^*\left(\frac{v}{\mu}\right) + \frac{\lambda}{2} \left\| \frac{x - u - v}{\lambda} \right\|_2^2.$$

We shall see that such a representation for S will allow us to show that S satisfies the following multi-time HJ equation

$$\begin{cases} \frac{\partial S(x, \mu, \lambda)}{\partial \mu} + J(\nabla_x S(x, \mu, \lambda)) = 0, & x \in \mathbb{R}^n, \mu > 0, \lambda > 0; \\ \frac{\partial S(x, \mu, \lambda)}{\partial \lambda} + \frac{1}{2} \|\nabla_x S(x, \mu, \lambda)\|_2^2 = 0, & x \in \mathbb{R}^n, \mu > 0, \lambda > 0; \\ S(x, 0, 0) = J(x), & x \in \mathbb{R}^n. \end{cases}$$

In figs. 1 to 6, the minimizers u, v and the minimal values S for the corresponding input images are shown. To compute the minimizers, we apply a splitting algorithm to convert the optimization problem (5) to two subproblems involving computing the proximal point of λJ and computing the projection to a μ -ball of Meyer's norm. The second subproblem is the dual problem to the first one. As a result, for both subproblems, we can apply the algorithm in [39, 50, 67] to obtain the exact minimizers.

In the first example, the test image x_1 is shown in fig. 1a. We consider the following parameters $\mu_1 = 1, \lambda_1 = 0.01$. The corresponding minimizers u and v are shown in figs. 1b and 1c. When $x = x_1, \lambda = \lambda_1$ are fixed, the minimal values $S(x_1, \mu, \lambda_1)$ can be regarded as a function of μ , whose graph is plotted in fig. 2b. Similarly, the graph of $S(x_1, \mu_1, \lambda)$ is plotted in fig. 2c. To illustrate the variation of S with respect to x , we choose another image x_2 with corresponding suitable parameters μ_2, λ_2 , and plot the function values $f : \alpha \mapsto S(\alpha x_1 + (1 - \alpha)x_2, \alpha \mu_1 + (1 - \alpha)\mu_2, \alpha \lambda_1 + (1 - \alpha)\lambda_2)$ with $\alpha \in [0, 1]$. In this example, x_2 is chosen to be a rotation of x_1 , and the parameters remain the same: $\mu_2 = \mu_1, \lambda_2 = \lambda_1$. The graph of f is plotted in fig. 2a. We also show an example of the mixed image $x = \alpha x_1 + (1 - \alpha)x_2$ for $\alpha = 0.3$ and the corresponding minimizers u, v in figs. 1d to 1f. In addition, the A^2BC model (with parameters $\mu = 0.06, \lambda = 0.01$) is applied to a noisy image shown in fig. 3a, whose minimizers are shown in figs. 3b and 3c.

The test image “Barbara” is used in the second example. The original image and the corresponding minimizers u, v in the A^2BC model with parameters $\mu = 30, \lambda = 8$ are shown in fig. 4. To demonstrate the variations of the minimal values, we choose two parts x_1, x_2 of the image, shown in figs. 5a and 5d, and repeat the experiment in the first example. Setting $\mu_1 = 16, \mu_2 = 24, \lambda_1 = 8$, and $\lambda_2 = 12$, the corresponding minimizers u, v are shown in figs. 5b, 5c, 5e, and 5f. The mixed image ($\alpha = 0.5$) and minimizers are shown in figs. 5g to 5i, and the dependence of S on x, μ, λ is shown in figs. 6a to 6c.

It can be seen from figs. 2 and 6 that S is a convex function with respect to the input image x and the parameters. This can be proved with a similar argument as in the proof of proposition 3.1. In this paper, more properties about S and the minimizers u, v are revealed.

Our contribution. The contribution of this paper is the theoretical results connecting the multi-time HJ equation and some optimization models such as decomposition models in imaging sciences. There are three parts in this paper. In the first part, we consider the decomposition models and the corresponding dual problems, and investigate the properties of their optimizers and optimal values. To be specific, for some optimization problems, the minimal value coincides with the solution $S(x, t_1, \dots, t_N)$ to a corresponding multi-time HJ equation. This relationship in the case of single-time HJ equations has been studied in [49]. We generalize the representation formula for the minimizer u_j and the variational analysis results of S and $\nabla_x S$ in [49] to the case of multi-time HJ equations. Moreover, we present a new variational analysis of the scaled minimizer $\frac{u_j}{t_j}$. In the variational analysis, we consider a sequence $\{(x_k, t_{1,k}, \dots, t_{N,k})\}_k$, whose

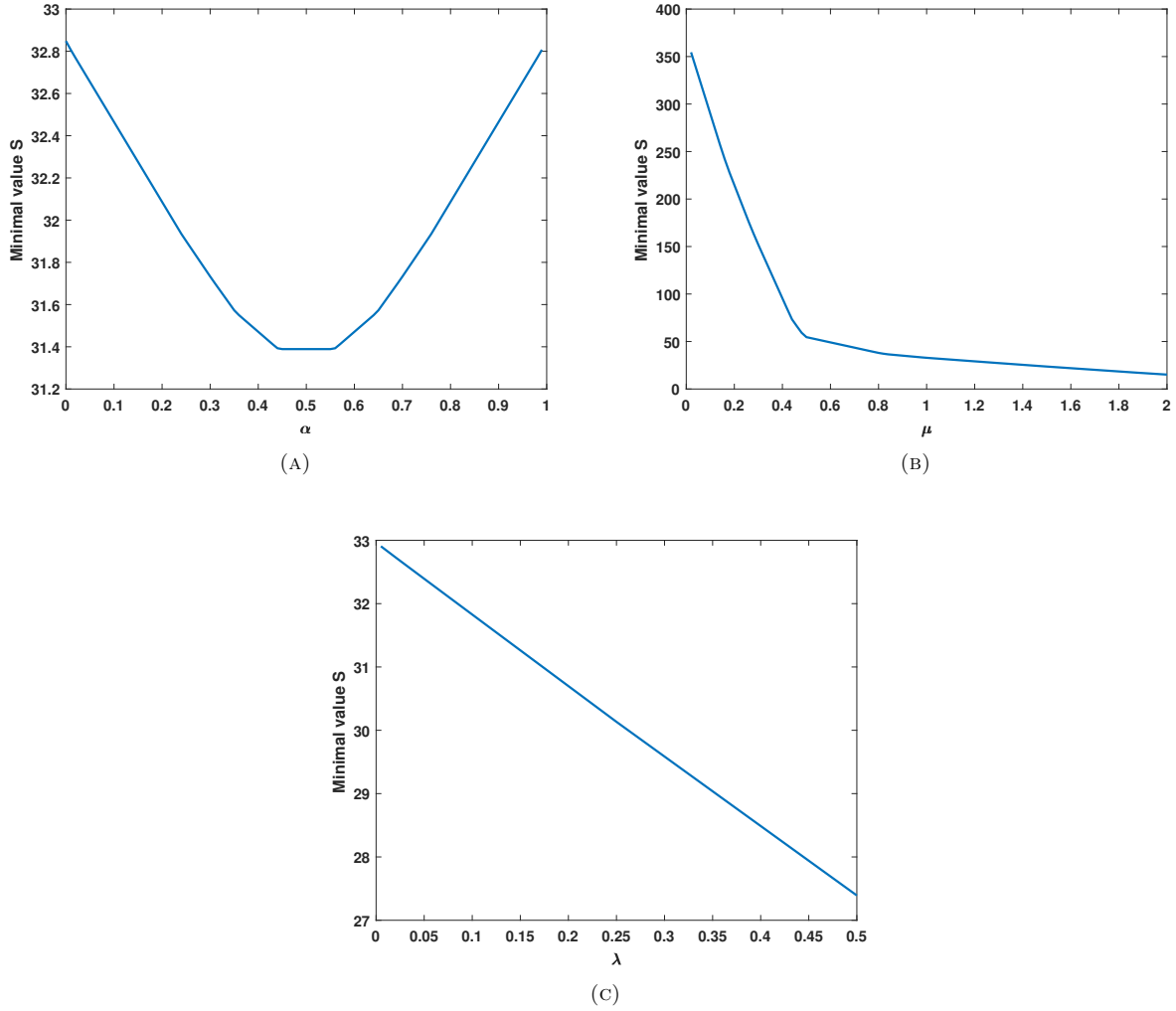


FIGURE 2. The graphs of the minimal values S with respect to the variables α , μ and λ in the first example are shown in (a)-(c), respectively. To be specific, (a) shows the function $\alpha \mapsto S(\alpha x_1 + (1 - \alpha)x_2, \alpha\mu_1 + (1 - \alpha)\mu_2, \alpha\lambda_1 + (1 - \alpha)\lambda_2)$, (b) shows the function $\mu \mapsto S(x_1, \mu, \lambda_1)$, and (c) shows the function $\lambda \mapsto S(x_1, \mu_1, \lambda)$.

elements are perturbed variables near the point $(x, 0, \dots, 0)$ and the perturbation becomes smaller when k is larger. We show that the limits of the corresponding spatial gradients $\nabla_x S$ and the scaled minimizers $\frac{u_j}{t_j}$ solve two optimization problems which are dual to each other. In the second part, we prove the uniqueness of the convex solution to the multi-time HJ equation under some specific assumptions. In the field of PDEs, the uniqueness of the viscosity solution has been widely studied, for which we refer the readers to [48] and the references listed there. Here, our contribution is to provide a new perspective from convex analysis and use the duality technique to prove the uniqueness of the convex solution. At last, we propose a regularization method for the decomposition problems which may have non-unique minimizers or non-differentiable minimal values. The regularization method is used to select a unique minimizer $u_{\lambda, \mu}$ and a unique gradient $p_{\lambda, \mu}$ of the minimal function where λ and μ are some positive parameters. In fact, the gradient $p_{\lambda, \mu}$ coincides with the maximizer in the corresponding dual problem. This regularization method can be regarded as a generalization of the Moreau-Yosida approximation, which is introduced, for example, in [8, 27]. Instead of only considering the primal problem as in the Moreau-Yosida approximation, our contribution here is to consider both the primal problem and the dual problem at the same time. Then, we apply the variational

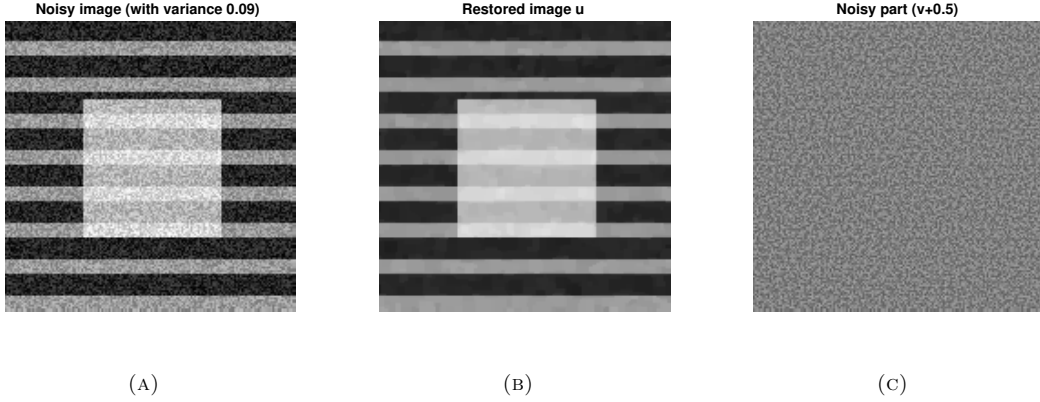


FIGURE 3. The A^2BC model is applied to the noisy test image shown in (a). The corresponding minimizers u and v are shown in (b) and (c), respectively.

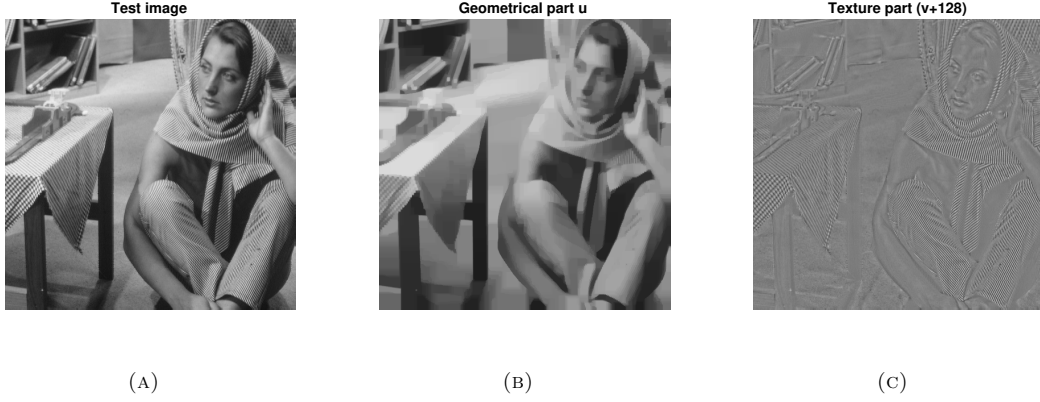


FIGURE 4. The A^2BC model is applied to the test image “Barbara”. The original image is shown in (a). The corresponding minimizers u and v are shown in (b) and (c), respectively.

analysis result in the first part to prove the convergence of $u_{\lambda,\mu}$ and $p_{\lambda,\mu}$. We show that they converge to the l^2 -projection of zero onto the corresponding sets of the original problems, when the regularization parameters λ and μ approach zero in a comparable rate.

Organization of the paper. The paper is organized as follows. Section 2 gives a brief review of the convex optimization theorems which are used in the later proofs. The main results are stated in sections 3 to 5. In section 3, the connection between some decomposition models and the multi-time HJ equation is shown. Proposition 3.2 provides the representation formula for the minimizers u_j of some decomposition models. Also, we investigate the variational behaviors of the minimal value S , the momentum $\nabla_x S$ and the velocities $\frac{u_j}{t_j}$ in proposition 3.4. Section 4 is devoted to the proof of the uniqueness of the convex solution to the multi-time HJ equation. In section 5, we present a regularization method for the degenerate cases which do not satisfy the assumptions in section 3. The method is demonstrated using a specific example but the analysis can be easily applied to other models. Finally, some conclusions are drawn in section 6.

2. MATHEMATICAL BACKGROUND

In this section, several basic definitions and theorems in convex analysis are reviewed. All the results and notations can be found in [65, 66]. We also refer the readers to [22, 24, 81].

First, a set C in \mathbb{R}^n is convex if $\alpha x + (1 - \alpha)y \in C$ whenever $x, y \in C$ and $\alpha \in [0, 1]$. The relative interior of C , denoted as $\text{ri } C$, is the interior of C with respect to the minimal hyperplane containing C in \mathbb{R}^n . For

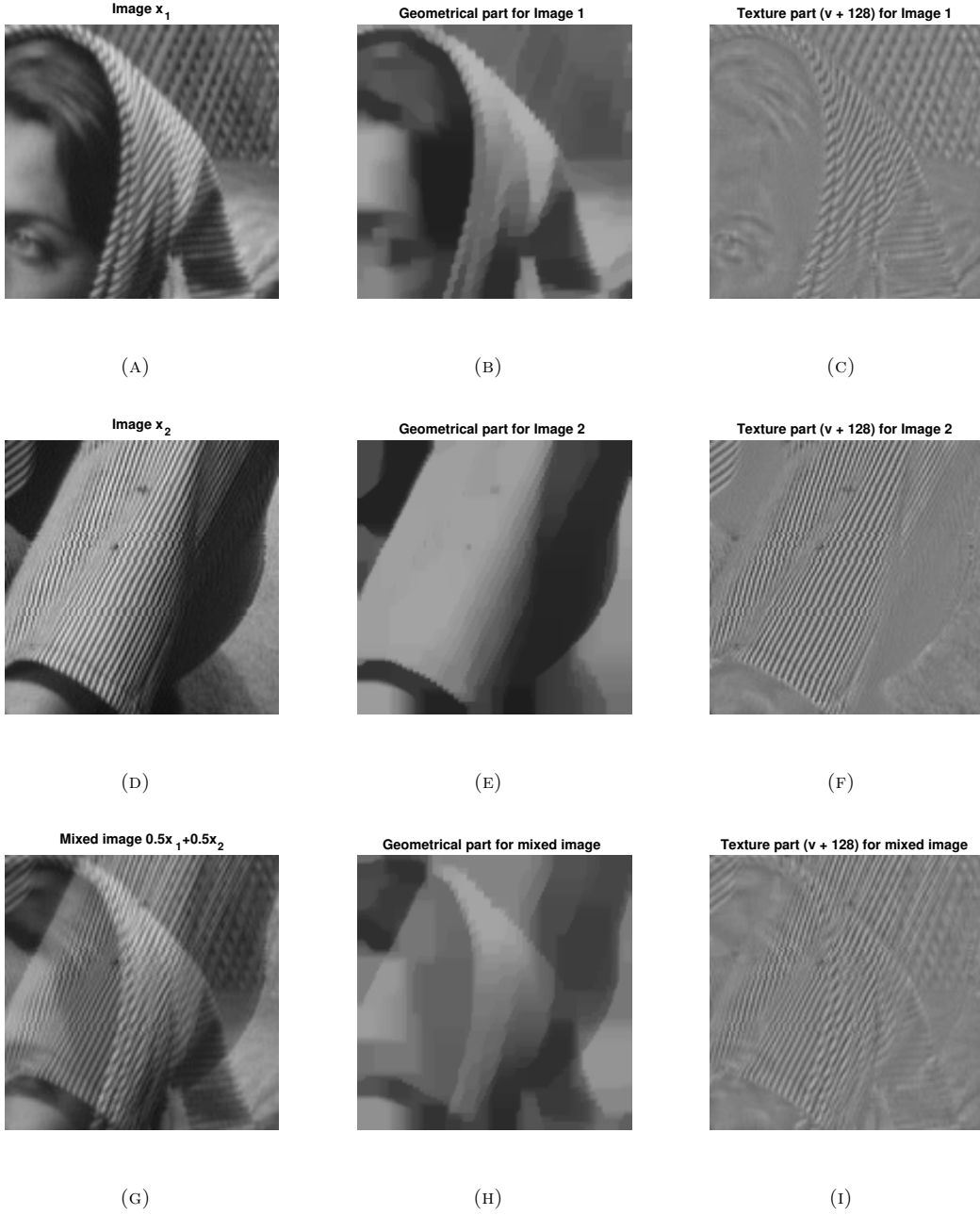


FIGURE 5. The A^2BC model is applied to two parts of the image “Barbara”. The original image x_1 , x_2 and corresponding minimizers u, v are shown in (a)-(f). The convex combination $0.5x_1 + 0.5x_2$ and its minimizers are shown in (g)-(i).

any convex set C , the normal cone of C at $x \in C$, denoted by $N_C(x)$, can be characterized by

$$(7) \quad q \in N_C(x) \text{ if and only if } \langle q, y - x \rangle \leq 0 \text{ for any } y \in C.$$

Here, we use the angle bracket $\langle \cdot, \cdot \rangle$ to denote the inner product operator in any Euclidean space \mathbb{R}^n . For any closed convex set C and any point $x \in C$, one can define the asymptotic cone of C , denoted as $C_\infty(x)$, by

$$(8) \quad C_\infty(x) = \{d \in \mathbb{R}^n : x + td \in C \text{ for all } t > 0\}.$$

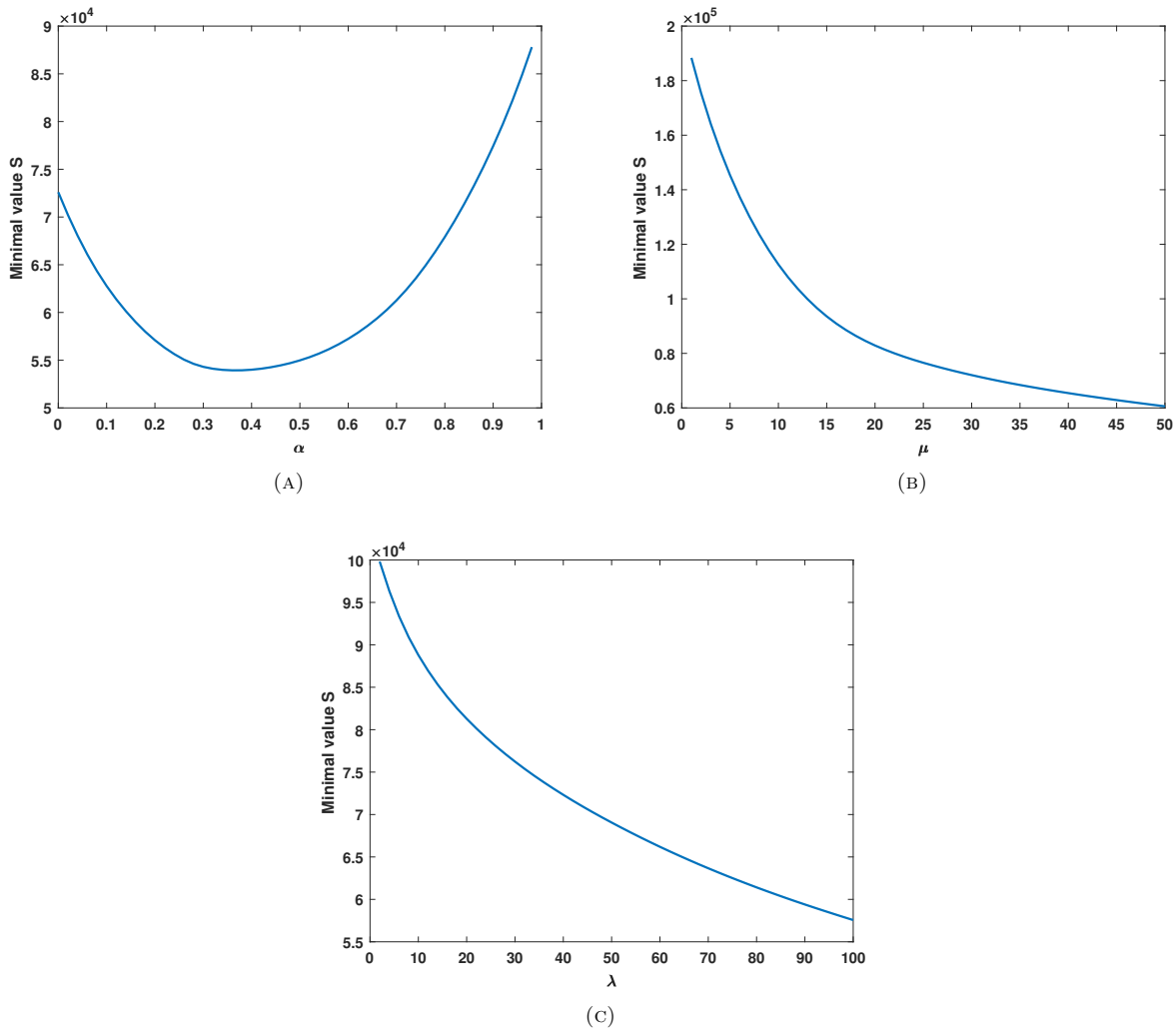


FIGURE 6. The graphs of the minimal values S with respect to the variables α , μ and λ in the second example are plotted in (a)-(c), respectively.

In fact, the asymptotic cone is independent of x , as stated in the following result.

Proposition 2.1. [65, Prop.III.2.2.1] *Let C be a closed convex set and $x, y \in C$. Then $C_\infty(x) = C_\infty(y)$. In other words, for any $d \in C_\infty(x)$, $y + td \in C$ for any $t > 0$.*

A function $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is said to be convex if for any $\alpha \in (0, 1)$ and any $x, y \in \mathbb{R}^n$,

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y).$$

The function f is called proper if it is not identically equal to $+\infty$. The domain of f , denoted by $\text{dom } f$, is defined to be the set where f does not take the value $+\infty$. The epigraph of f , denoted as $\text{epi } f$, is defined by:

$$\text{epi } f := \{(x, t) : x \in \text{dom } f, t \geq f(x)\}.$$

Then, f is convex (proper, or lower semi-continuous, respectively) if and only if $\text{epi } f$ is convex (non-empty, or closed, respectively). We denote $\Gamma_0(\mathbb{R}^n)$ to be the set of proper, convex and lower semi-continuous (l.s.c) functions from \mathbb{R}^n to $\mathbb{R} \cup \{+\infty\}$. In this section, we only consider the functions in $\Gamma_0(\mathbb{R}^n)$. These functions have good continuity properties, which are stated below.

TABLE 1. Notations used in this paper. Here, we use C to denote a set, f to denote a function and x, d to denote vectors in \mathbb{R}^n .

Notation	Meaning	Definition
$\text{dom } f$	domain of f	$\{x \in \mathbb{R}^n : f(x) \in \mathbb{R}\}$
$\text{ri } C$	relative interior of C	the interior of C with respect to the minimal hyperplane containing C in \mathbb{R}^n
$N_C(x)$	normal cone of C at x	$\{q \in \mathbb{R}^n : \langle q, y - x \rangle \leq 0 \text{ for any } y \in C\}$
$C_\infty(x)$	asymptotic cone of C	$\{d \in \mathbb{R}^n : x + td \in C \text{ for all } t > 0\}$
$\text{epi } f$	epigraph of f	$\{(x, t) \in \mathbb{R}^n \times \mathbb{R} : x \in \text{dom } f, t \geq f(x)\}$
$\Gamma_0(\mathbb{R}^n)$	a useful and standard class of convex functions	the set containing all proper, convex, l.s.c. functions from \mathbb{R}^n to $\mathbb{R} \cup \{+\infty\}$
$f'(x, d)$	directional derivative of f at x along the direction d	$\lim_{h \rightarrow 0^+} \frac{1}{h}(f(x + hd) - f(x))$
$\partial f(x)$	subdifferential of f at x	$\{p \in \mathbb{R}^n : f(y) \geq f(x) + \langle p, y - x \rangle \forall y \in \mathbb{R}^n\}$
I_C	the indicator function of C	If $x \in C$, then define $I_C(x) := 0$. Otherwise, define $I_C(x) := +\infty$.
f^*	Legendre transform of f	$f^*(p) := \sup_{x \in \mathbb{R}^n} \langle p, x \rangle - f(x)$
$f \square g$	inf-convolution of f and g	$(f \square g)(x) := \inf_{u \in \mathbb{R}^n} f(u) + g(x - u)$

Proposition 2.2. [65, Lem.IV.3.1.1 and Chap.I.3.1 - 3.2] Let $f \in \Gamma_0(\mathbb{R}^n)$. If $x \in \text{ri dom } f$, then f is continuous at x in $\text{dom } f$. If $x \in \text{dom } f \setminus \text{ri dom } f$, then for any $y \in \text{ri dom } f$,

$$f(x) = \lim_{t \rightarrow 0^+} f(x + t(y - x)).$$

For any $f \in \Gamma_0(\mathbb{R}^n)$ and $x \in \text{dom } f$, the directional derivative at x along any direction d , denoted as $f'(x, d)$, is well-defined in $\mathbb{R} \cup \{\pm\infty\}$. When f is differentiable at x , $f'(x, \cdot) = \langle \nabla f(x), \cdot \rangle$ is a linear function. In general, when f is not differentiable, $f'(x, \cdot)$ is only sublinear, in which case we can consider the linear functions dominated by it. Each normal vector of such linear functions gives a subgradient of f at x , whose formal definition is given below. Also, the rigorous statement about the relation we described above between the directional derivatives and subgradients is given in proposition 2.6.

A vector p is called a subgradient of f at x if it satisfies

$$f(y) \geq f(x) + \langle p, y - x \rangle, \text{ for any } y \in \mathbb{R}^n.$$

The collection of all such subgradients is called the subdifferential of f at x , denoted as $\partial f(x)$. It is easy to check that $0 \in \partial f(x)$ if and only if x is a minimizer of f . As a result, one can check whether x is a minimizer by computing the subdifferential.

As is well known, the subdifferential operator is a (maximal) monotone operator. To be specific,

$$(9) \quad \langle p - q, x - y \rangle \geq 0 \text{ for any } p \in \partial f(x) \text{ and } q \in \partial f(y).$$

Moreover, in most cases, the subdifferential operator commutes with summation.

Proposition 2.3. [66, Cor.XI.3.1.2] Let $f, g \in \Gamma_0(\mathbb{R}^n)$. Assume $\text{ri dom } f \cap \text{ri dom } g \neq \emptyset$. Then $\partial(f+g)(x) = \partial f(x) + \partial g(x)$ for any $x \in \text{dom } f \cap \text{dom } g$.

Here, we give one simple example. For any convex set C , the indicator function I_C is defined by

$$I_C(x) := \begin{cases} 0, & x \in C; \\ +\infty, & x \notin C. \end{cases}$$

In this paper, we also use the notation $I\{\cdot\}$ to denote the indicator function if the set C is given in the form of some constraints. By definition, the indicator function I_C remains the same after multiplying by a positive constant, i.e. we have $\alpha I_C = I_C$ for any $\alpha > 0$. One can compute the subdifferential of the indicator function and obtain

$$(10) \quad \partial I_C(x) = N_C(x).$$

Next, we introduce one important transform in convex analysis called Legendre transform. For any function $f \in \Gamma_0(\mathbb{R}^n)$, the Legendre transform of f , denoted as f^* , is defined by

$$(11) \quad f^*(p) := \sup_{x \in \mathbb{R}^n} \langle p, x \rangle - f(x).$$

Legendre transform gives a duality relationship between f and f^* . In other words, if $f \in \Gamma_0(\mathbb{R}^n)$, then $f^* \in \Gamma_0(\mathbb{R}^n)$ and $f^{**} = f$. Similarly, along with this duality relationship, some properties are dual to others, as stated in the following proposition. (Here and after, a function g is called 1-coercive if $\lim_{\|x\| \rightarrow +\infty} g(x)/\|x\| = +\infty$.)

Proposition 2.4. [66, Chap.X.4.1] *Let $f \in \Gamma_0(\mathbb{R}^n)$. Then f is finite-valued if and only if f^* is 1-coercive. Also, f is differentiable if and only if f^* is strictly convex.*

In particular, the subgradients can be characterized by the maximizers in eq. (11).

Proposition 2.5. [66, Cor.X.1.4.4] *Let $f \in \Gamma_0(\mathbb{R}^n)$ and $p, x \in \mathbb{R}^n$. Then $p \in \partial f(x)$ if and only if $x \in \partial f^*(p)$, if and only if $f(x) + f^*(p) = \langle p, x \rangle$.*

The concepts we introduced above, including directional derivatives, subgradients and Legendre transform, can be linked all together by the following proposition.

Proposition 2.6. [66, Example X.2.4.3] *Let $f \in \Gamma_0(\mathbb{R}^n)$ and $x \in \text{dom } f$ such that $\partial f(x)$ is nonempty, then $(f'(x, \cdot))^* = I_{\partial f(x)}$. Moreover, if $x \in \text{ri dom } f$, then $f'(x, \cdot) \in \Gamma_0(\mathbb{R}^n)$, hence $f'(x, \cdot) = I_{\partial f(x)}^*$.*

Except from Legendre transform, there is another operator to construct convex functions called inf-convolution. Given two functions $f, g \in \Gamma_0(\mathbb{R}^n)$, assume there exists an affine function l such that $f(x) \geq l(x)$ and $g(x) \geq l(x)$ for any $x \in \mathbb{R}^n$. Then, the inf-convolution between f and g , denoted as $f \square g$, is a convex function taking values in $\mathbb{R} \cup \{+\infty\}$. The definition of the inf-convolution $f \square g$ is given by

$$(12) \quad (f \square g)(x) := \inf_{u \in \mathbb{R}^n} f(u) + g(x - u).$$

In the following proposition, the relation between Legendre transform and inf-convolution is stated. Actually, the Hopf formula and Lax formula introduced in the next section are formulated using Legendre transform and inf-convolution operator, respectively. As a result, these two operators play a significant role in our analysis in this paper.

Proposition 2.7. [66, Thm.X.2.3.2 and Thm.XI.3.4.1] *Let $f, g \in \Gamma_0(\mathbb{R}^n)$. Assume the intersection of $\text{ri dom } f^*$ and $\text{ri dom } g^*$ is non-empty. Then $f \square g \in \Gamma_0(\mathbb{R}^n)$ and $f \square g = (f^* + g^*)^*$. Moreover, for any $x \in \text{dom } f \square g$, the optimization problem eq. (12) has at least one minimizer, and $\partial(f \square g)(x) = \partial f(u) \cap \partial g(x - u)$ for any minimizer u .*

3. PROPERTIES OF THE SOLUTIONS TO THE MULTI-TIME HAMILTON-JACOBI EQUATIONS

In this section, we provide a representation formula for the minimizers in the Lax formula and highlight the relation of the minimizers and the momentum in the multi-time HJ equation. Also, we investigate the variational behaviors of both the solution to the multi-time HJ equation and the corresponding momentum when time variables approach zero. Moreover, we also present a new result stating the variational behaviors of the velocities, which has not been developed before, even for the single-time case. Similar to the duality relation of the Hopf and Lax formulas, the cluster points of the minimizers and momentum solve two optimization problems, which are also dual to each other. An illustration is given in the upper part of fig. 7.

We consider the solution $S(x, t_1, \dots, t_N)$ to the following multi-time HJ equation

$$(13) \quad \begin{cases} \frac{\partial S}{\partial t_j} + H_j(\nabla_x S) = 0 \text{ for any } j \in \{1, \dots, N\}, & x \in \mathbb{R}^n, t_1, \dots, t_N > 0; \\ S(x, 0, \dots, 0) = J(x), & x \in \mathbb{R}^n. \end{cases}$$

Here, we only consider the multi-time HJ equations whose Hamiltonians only depend on the momentum $\nabla_x S$. Several conditions are imposed on the Hamiltonians $\{H_j\}$ and the initial data J in this section. To be specific, we assume

- (H1) $H_j : \mathbb{R}^n \rightarrow \mathbb{R}$, is convex and 1-coercive for any $j = 1, \dots, N$. Moreover, at least one of them is strictly convex;
- (H2) $J \in \Gamma_0(\mathbb{R}^n)$.

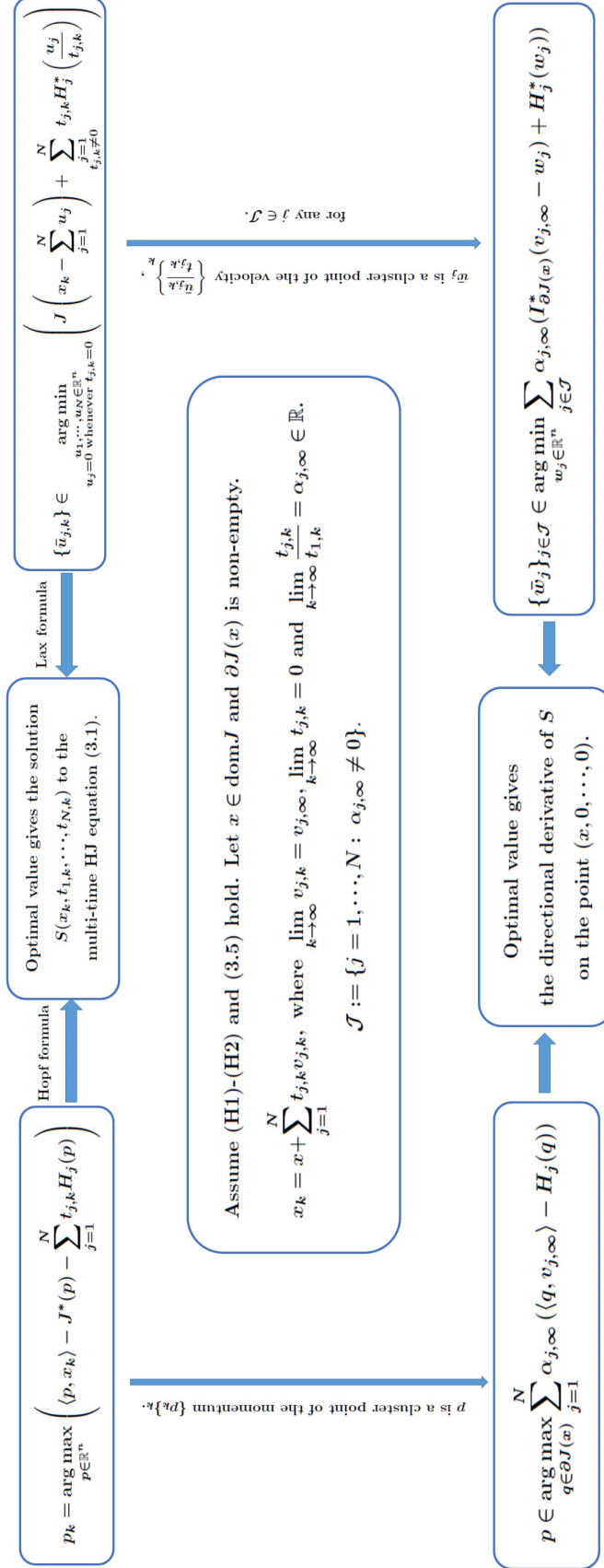


FIGURE 7. This is an illustration for proposition 3.4. It shows the relation of four optimization problems and the multi-time HJ equation. Here, $\bar{u}_{j,k}$ denotes the minimizer $u_j(x_k, t_{1,k}, \dots, t_{N,k})$.

From the assumption (H1), by proposition 2.4, it is known that H_j^* is also finite-valued, convex and 1-coercive for any $j = 1, \dots, N$. Moreover, at least one H_j^* is differentiable.

It is well known that in this case the unique classical solution is given by the Hopf formula [71, 88] stated as follows

$$(14) \quad S_H(x, t_1, \dots, t_N) := \left(J^* + \sum_{j=1}^N t_j H_j \right)^* (x) = \sup_{p \in \mathbb{R}^n} \left(\langle p, x \rangle - J^*(p) - \sum_{j=1}^N t_j H_j(p) \right),$$

and the Lax formula [88] stated as follows

$$(15) \quad \begin{aligned} S_L(x, t_1, \dots, t_N) &:= (J \square (t_1 H_1)^* \square \dots \square (t_N H_N)^*) (x) \\ &= \inf_{\substack{u_1, \dots, u_N \in \mathbb{R}^n \\ u_j = 0 \text{ whenever } t_j = 0}} \left(J \left(x - \sum_{j=1}^N u_j \right) + \sum_{\substack{j=1 \\ t_j \neq 0}}^N t_j H_j^* \left(\frac{u_j}{t_j} \right) \right), \end{aligned}$$

for any $x \in \mathbb{R}^n$ and $t_1, \dots, t_N \geq 0$. We extend S_H and S_L to the whole domain by simply setting the function values to $+\infty$ whenever the function value is not defined. There are some physical interpretations of the HJ PDEs and the optimizers in the above two formulas. Given suitable Hamiltonians $\{H_j\}$ and a suitable initial condition J , the HJ PDE eq. (13) describes the movement of a particle. Roughly speaking, in a time interval with length t_j , a particle moves along the characteristic line of the j -th equation in the PDE system. The velocity in this time interval equals $\frac{u_j}{t_j}$ where (u_1, \dots, u_N) denotes the minimizer in the Lax formula eq. (15). On the other hand, the maximizer in the Hopf formula eq. (14) gives the momentum of the particle, which coincides with the spatial gradient $\nabla_x S(x, t_1, \dots, t_N)$. We refer the reader to [21] for details about HJ PDEs and variational principles in physics.

Under the assumptions (H1) and (H2), $S_H = S_L$, and the value is finite if there exists some $t_j > 0$. In addition, the minimizers in the Lax formula eq. (15) exist whenever the minimal value is finite. This result can be proved using proposition 2.7. Also, by proposition 2.5, it is not hard to check $S_H \in C^1(\mathbb{R}^n \times (0, +\infty)^N)$ and satisfies HJ equation eq. (13). Moreover, the spatial gradient is the unique maximizer in the Hopf formula eq. (14). To conclude, the Hopf and Lax formulas express the classical solution to the multi-time HJ equation as two optimization problems. The Hopf formula provides a physical interpretation and has the momentum $\nabla_x S$ as the maximizer, while its dual problem in the Lax formula is in the same form as some decomposition models in imaging sciences.

The following proposition states that the solution is actually a convex function, hence the techniques in convex analysis can be applied to analyze the solution. The results hold even under weaker assumptions. Actually, a part of the proposition can be further generalized to the case when $J, H_j \in \Gamma_0(\mathbb{R}^n)$ and $\text{dom } J^* \subseteq \text{dom } H_j$ for any j .

Proposition 3.1. *Let $J, H_j \in \Gamma_0(\mathbb{R}^n)$ and $\text{dom } H_j = \mathbb{R}^n$ for any j . Then, $S_H \in \Gamma_0(\mathbb{R}^{n+N})$, whose Legendre transform is given by*

$$S_H^*(p, E^-) = J^*(p) + \sum_{j=1}^N I\{E_j^- + H_j(p) \leq 0\},$$

for any $p \in \mathbb{R}^n$ and $E^- = (E_1^-, \dots, E_N^-) \in \mathbb{R}^N$. Here, $I\{\cdot\}$ denotes the indicator function. Moreover, if the assumptions (H1)-(H2) are satisfied, then $S_H(x, t_1, \dots, t_N)$ is finite for any $x \in \mathbb{R}^n$ and $t_1, \dots, t_N \geq 0$ which are not all zero.

Proof. First, we prove that S_H is the Legendre transform of F , where F is defined by

$$F(p, E^-) := J^*(p) + \sum_{j=1}^N I\{E_j^- + H_j(p) \leq 0\},$$

for any $p \in \mathbb{R}^n$ and any $E^- = (E_1^-, \dots, E_N^-) \in \mathbb{R}^N$. It is easy to check $F \in \Gamma_0(\mathbb{R}^{n+N})$.

By definition, for any $x \in \mathbb{R}^n$ and $t = (t_1, \dots, t_N) \in \mathbb{R}^N$,

$$(16) \quad F^*(x, t) = \sup_{p \in \mathbb{R}^n, E^- \in \mathbb{R}^N} \left(\langle x, p \rangle + \sum_{j=1}^N t_j E_j^- - J^*(p) - \sum_{j=1}^N I\{E_j^- + H_j(p) \leq 0\} \right).$$

First, we consider the case when there exists k such that $t_k < 0$. Take $p \in \text{dom } J^*$. For any $j \neq k$, take $E_j^- = -H_j(p)$, which is a finite value. From the above equation,

$$F^*(x, t) \geq \langle x, p \rangle + \sum_{j \neq k} t_j E_j^- - J^*(p) + \limsup_{\substack{E_k^- \leq -H_k(p) \\ E_k^- \rightarrow -\infty}} t_k E_k^- = +\infty.$$

Hence $F^*(x, t) = +\infty = S_H(x, t)$ if $t_k < 0$ for some k .

Then, consider the case when $t_1, \dots, t_N \geq 0$. Let $x \in \mathbb{R}^n$, from eq. (16), we obtain

$$(17) \quad \begin{aligned} F^*(x, t) &= \sup_{\substack{p \in \mathbb{R}^n \\ E_j^- \leq -H_j(p) \ \forall j}} \left(\langle x, p \rangle + \sum_{j: t_j > 0} t_j E_j^- - J^*(p) \right) \\ &= \sup_{p \in \mathbb{R}^n} \left[\sup_{\substack{E_j^- \leq -H_j(p) \ \forall j \\ E_j^- \geq -H_j(p) \ \forall j: t_j > 0}} \left(\langle x, p \rangle + \sum_{j: t_j > 0} t_j E_j^- - J^*(p) \right) \right] \\ &= \sup_{p \in \mathbb{R}^n} \left(\langle x, p \rangle - \sum_{j: t_j > 0} t_j H_j(p) - J^*(p) \right) = S_H(x, t). \end{aligned}$$

Therefore, $S_H = F^*$, which implies S_H is a convex lower semi-continuous function and $F = S_H^*$. Moreover, if there exists some k such that $t_k > 0$ and $t_j \geq 0$ for any $j \neq k$, then, by assumption (H1), we deduce that $J^* + \sum_j t_j H_j$ is 1-coercive, which, by proposition 2.4, implies its Legendre transform $S_H(\cdot, t_1, \dots, t_N)$ (with respect to x) is finite-valued. \square

By investigating S_H on the boundary of the domain, the solution to a lower time dimensional equation is embedded in the solution to the higher time dimensional equation, in the sense that the restriction of S_H on the subspace $\{(x, t_1, \dots, t_N) : t_j = 0 \ \forall j \in J\}$ for any index set $J \subset \{1, \dots, N\}$ is the solution to the corresponding lower time dimensional HJ equation with Hamiltonians $\{H_j\}_{j \notin J}$.

The following proposition states a representation formula for the minimizers in the Lax formula. In the decomposition model eq. (15), a given image x is decomposed into different components including u_1, \dots, u_N and the residual $x - \sum_{j=1}^N u_j$. However, sometimes the primal minimization problem is difficult to solve, then the following proposition can be applied to compute (u_1, \dots, u_N) using the momentum $\nabla_x S_L(x, t_1, \dots, t_N)$. In fact, the momentum is the maximizer of the dual problem in the Hopf formula eq. (14). In other words, the following proposition gives the relation of the optimizers in the primal decomposition problem and the dual problem.

Proposition 3.2. *Suppose the assumptions (H1)-(H2) hold. Let $x \in \mathbb{R}^n, t_1, \dots, t_N \geq 0$ and assume the time variables $\{t_j\}$ are not all zero. Denote (u_1, \dots, u_N) to be any minimizer of the minimization problem in eq. (15) with parameters x and t_1, \dots, t_N . Here, each u_j can be regarded as a function of (x, t_1, \dots, t_N) . Then, for any j ,*

$$(18) \quad u_j(x, t_1, \dots, t_N) \in t_j \partial H_j(\nabla_x S_L(x, t_1, \dots, t_N)).$$

Specifically, if a stronger assumption is imposed, say, all the Hamiltonians are differentiable, then the minimizer (u_1, \dots, u_N) is unique and satisfies

$$u_j(x, t_1, \dots, t_N) = t_j \nabla H_j(\nabla_x S_L(x, t_1, \dots, t_N)) \text{ for any } j.$$

Proof. Since $\text{dom } H_j = \mathbb{R}^n$ for each j , by proposition 2.7 and induction, the minimizers u_j exist if $S_L(x, t_1, \dots, t_N) < +\infty$, and

$$(19) \quad \begin{aligned} \partial_x S_L(x, t_1, \dots, t_N) &= \partial J \left(x - \sum_{j=1}^N u_j \right) \cap \left(\bigcap_{j=1}^N \partial \left(t_j H_j^* \left(\frac{\cdot}{t_j} \right) \right) (u_j) \right) \\ &= \partial J \left(x - \sum_{j=1}^N u_j \right) \cap \left(\bigcap_{j=1}^N \partial H_j^* \left(\frac{u_j}{t_j} \right) \right). \end{aligned}$$

From the assumption (H1), there exists some j such that H_j^* is differentiable, hence the intersection above contains at most one element. On the other hand, $\partial_x S_L$ is non-empty in the interior of the domain of $S_L(\cdot, t_1, \dots, t_N)$, which is the whole space \mathbb{R}^n because $S_L = S_H$ is finite-valued when the time variables are not all zero. Therefore, the above intersection contains exactly one element. In other words, S_L is differentiable with respect to x for any $t_1, \dots, t_N \geq 0$ which are not all zero and $x \in \mathbb{R}^n$. Moreover, by eq. (19), $\nabla_x S_L \in \partial H_j^*(u_j/t_j)$, which implies $u_j/t_j \in \partial H_j(\nabla_x S_L(x, t_1, \dots, t_N))$ for any j . \square

In the remaining part of this section, we investigate the multi-time HJ equation eq. (13) and the minimization problem eq. (15) in a variational point of view. To be specific, let $v_{j,k} \in \mathbb{R}^n$ and $t_{j,k} > 0$ for any $j \in \{1, \dots, N\}$ and $k \in \mathbb{N}$ such that they satisfy $\lim_{k \rightarrow +\infty} t_{j,k} = 0$ and $\lim_{k \rightarrow +\infty} v_{j,k} = v_{j,\infty}$ for any j . Let $x \in \mathbb{R}^n$ and $x_k = x + \sum_{j=1}^N t_{j,k} v_{j,k}$ for any k . We are interested in the convergence behavior of the momentum $\nabla_x S_H$ and the minimizers u_j evaluated at $(x_k, t_{1,k}, \dots, t_{N,k})$. We will demonstrate one application in section 5.

Among all the sequences $\{t_{j,k}\}_k, j = 1, \dots, N$, by taking subsequences, we can assume there is a sequence with the lowest convergence rate. According to the symmetry of the time variables, without loss of generality, we can assume $\{t_{1,k}\}_k$ is the slowest sequence converging to zero compared to $\{t_{j,k}\}_k$ for any $j > 1$, i.e., we assume that $\left\{ \frac{t_{j,k}}{t_{1,k}} \right\}_k$ has a finite limit denoted as $\alpha_{j,\infty} \in \mathbb{R}$ for any j . In summary, the following notations and assumptions are adopted:

$$(20) \quad \begin{cases} x_k = x + \sum_{j=1}^N t_{j,k} v_{j,k}, \text{ where } t_{j,k} > 0, x, v_{j,k} \in \mathbb{R}^n \text{ for any } j \in \{1, \dots, N\} \text{ and } k \in \mathbb{N}; \\ \lim_{k \rightarrow +\infty} t_{j,k} = 0 \text{ and } \lim_{k \rightarrow +\infty} v_{j,k} = v_{j,\infty}; \\ \lim_{k \rightarrow +\infty} \frac{t_{j,k}}{t_{1,k}} = \alpha_{j,\infty} \in \mathbb{R}. \end{cases}$$

In the decomposition models, $\{x_k\}$ is given by a sequence of observed images. In each x_k there is a constant component denoted by x and several other components denoted by $t_{j,k} v_{j,k}$ for $j = 1, \dots, N$. In the remaining part of this section, we investigate the behavior of the minimizers of the decomposition model in eq. (15) when the components $t_{j,k} v_{j,k}$ converge to zero and the parameters $t_{j,k}$ in the model vanish.

First, we show the convergence of u_j to zero, which is stated in (i) in the following proposition. In other words, the decomposition model recovers the constant component x when the other components $t_{j,k} v_{j,k}$ and the parameters $t_{j,k}$ in the model converge to zero. Then, (ii) and (iii) in the following proposition are technical results about the convergence rate, which will be used in later proofs.

Proposition 3.3. *Assume (H1)-(H2) and eq. (20) hold. Let (u_1, \dots, u_N) be any minimizer of the minimization problem in eq. (15). Let $x \in \text{dom } J$. Then,*

(i) *For any $j = 1, \dots, N$,*

$$(21) \quad \lim_{k \rightarrow +\infty} u_j(x_k, t_{1,k}, \dots, t_{N,k}) = 0.$$

(ii) *If $\partial J(x) \neq \emptyset$ and $\alpha_{j,\infty} = 0$, then*

$$\lim_{k \rightarrow +\infty} \frac{1}{t_{1,k}} u_j(x_k, t_{1,k}, \dots, t_{N,k}) = 0.$$

(iii) *If $\partial J(x) \neq \emptyset$ and $\alpha_{j,\infty} \neq 0$, then the sequence $\left\{ \frac{1}{t_{j,k}} u_j(x_k, t_{1,k}, \dots, t_{N,k}) \right\}_k$ is bounded.*

Proof. Denote $\bar{u}_{j,k} := u_j(x_k, t_{1,k}, \dots, t_{N,k})$ for any $j = 1, \dots, N$, and $\bar{u}_{0,k} := x_k - \sum_{j=1}^N \bar{u}_{j,k}$. Define $I := \{j : \{\|\bar{u}_{j,k}\|/t_{j,k}\}_k \text{ is not bounded}\}$. Recall that for each $j = 1, \dots, N$, $\{v_{j,k}\}_k \subset \mathbb{R}^n$ and $\{t_{j,k}\}_k \subset (0, +\infty)$ are two sequences satisfying $\lim_{k \rightarrow +\infty} v_{j,k} = v_{j,\infty}$ and $\lim_{k \rightarrow +\infty} t_{j,k} = 0$, respectively. And the k -th spatial variable x_k is defined to be $x + \sum_{j=1}^N t_{j,k} v_{j,k}$.

Proof of (i): By Lax formula eq. (15),

$$(22) \quad \begin{aligned} J(\bar{u}_{0,k}) + \sum_{j=1}^N t_{j,k} H_j^* \left(\frac{\bar{u}_{j,k}}{t_{j,k}} \right) &\leq J \left(x_k - \sum_{j=1}^N t_{j,k} v_{j,k} \right) + \sum_{j=1}^N t_{j,k} H_j^*(v_{j,k}) \\ &= J(x) + \sum_{j=1}^N t_{j,k} H_j^*(v_{j,k}). \end{aligned}$$

Since J is a convex function, there exists $z \in \text{dom } J$ such that $\partial J(z) \neq \emptyset$. Let $q \in \partial J(z)$. Then, using the convexity of J and Cauchy-Schwarz inequality, we get

$$(23) \quad J(\bar{u}_{0,k}) \geq J(z) + \langle q, \bar{u}_{0,k} - z \rangle \geq J(z) - \|q\| \sum_{j=1}^N \|\bar{u}_{j,k}\| - \|q\| \|x_k - z\|.$$

Combining eq. (22) and eq. (23), we get

$$(24) \quad \sum_{j=1}^N t_{j,k} H_j^* \left(\frac{\bar{u}_{j,k}}{t_{j,k}} \right) \leq J(x) - J(z) + \sum_{j=1}^N t_{j,k} H_j^*(v_{j,k}) + \|q\| \sum_{j=1}^N \|\bar{u}_{j,k}\| + \|q\| \|x_k - z\|.$$

For any $j \in I$, since $\|\bar{u}_{j,k}\|/t_{j,k}$ is not bounded, without loss of generality, by taking subsequences, we can assume $\|\bar{u}_{j,k}\|/t_{j,k}$ increases to infinity. Since H_j^* is 1-coercive, for any $M > 0$, there exists K such that for any $k > K$, $H_j^*(\bar{u}_{j,k}/t_{j,k}) \geq M \|\bar{u}_{j,k}\|/t_{j,k}$. Together with eq. (24), we get

$$(25) \quad \begin{aligned} \sum_{j \in I} (M - \|q\|) \|\bar{u}_{j,k}\| &\leq \sum_{j \in I} \left(t_{j,k} H_j^* \left(\frac{\bar{u}_{j,k}}{t_{j,k}} \right) - \|q\| \|\bar{u}_{j,k}\| \right) \\ &\leq J(x) - J(z) + \|q\| \|x_k - z\| + \sum_{j=1}^N t_{j,k} H_j^*(v_{j,k}) + \sum_{j \notin I} \left(\|q\| \|\bar{u}_{j,k}\| - t_{j,k} H_j^* \left(\frac{\bar{u}_{j,k}}{t_{j,k}} \right) \right). \end{aligned}$$

Since $\{t_{j,k}\}_k$ and $\{v_{j,k}\}_k$ are bounded, and H_j^* is continuous in \mathbb{R}^n for any j , then the right hand side is bounded. However, M can be arbitrarily large, then the boundedness of left hand side (deduced by the boundedness of the right hand side) implies $\|\bar{u}_{j,k}\| \rightarrow 0$ for any $j \in I$. If $j \notin I$, then $\|\bar{u}_{j,k}\|/t_{j,k}$ is bounded by the definition of I , hence $\bar{u}_{j,k}$ also converges to zero.

Proof of (ii): We can apply the same argument as above and set $z = x$, because $\partial J(x) \neq \emptyset$. From eq. (25), using the definition of x_k in eq. (20) and triangle inequality, we have

$$\begin{aligned} \sum_{j \in I} (M - \|q\|) \|\bar{u}_{j,k}\| &\leq \|q\| \|x_k - x\| + \sum_{j=1}^N t_{j,k} H_j^*(v_{j,k}) + \sum_{j \notin I} \left(\|q\| \|\bar{u}_{j,k}\| - t_{j,k} H_j^* \left(\frac{\bar{u}_{j,k}}{t_{j,k}} \right) \right) \\ &\leq \sum_{j=1}^N t_{j,k} (H_j^*(v_{j,k}) + \|q\| \|v_{j,k}\|) + \sum_{j \notin I} t_{j,k} \left(\|q\| \left\| \frac{\bar{u}_{j,k}}{t_{j,k}} \right\| - H_j^* \left(\frac{\bar{u}_{j,k}}{t_{j,k}} \right) \right). \end{aligned}$$

Dividing both sides by $t_{1,k}$, we can obtain

$$\begin{aligned} (M - \|q\|) \sum_{j \in I} \frac{\|\bar{u}_{j,k}\|}{t_{1,k}} &\leq \sum_{j=1}^N \frac{t_{j,k}}{t_{1,k}} (H_j^*(v_{j,k}) + \|q\| \|v_{j,k}\|) \\ &\quad + \sum_{j \notin I} \frac{t_{j,k}}{t_{1,k}} \left(\|q\| \left\| \frac{\bar{u}_{j,k}}{t_{j,k}} \right\| - H_j^* \left(\frac{\bar{u}_{j,k}}{t_{j,k}} \right) \right). \end{aligned}$$

With the same argument as in the proof of (i), we deduce that the right hand side is bounded, while M can be arbitrarily large. Therefore, $\|\bar{u}_{j,k}\|/t_{1,k}$ converges to zero for any $j \in I$. If $j \notin I$ and $\alpha_{j,\infty} = 0$, then

$\|\bar{u}_{j,k}\|/t_{j,k}$ is bounded by the definition of I and $t_{j,k}/t_{1,k}$ converges to zero by the definition of $\alpha_{j,\infty}$, hence $\|\bar{u}_{j,k}\|/t_{1,k}$ also converges to zero.

Proof of (iii): It suffices to prove the contrapositive statement. To be specific, let $j \in I$, i.e. $\|\bar{u}_{j,k}\|/t_{j,k}$ is unbounded, it suffices to prove $\alpha_{j,\infty} = 0$. In the proof of (ii), we know that $\|\bar{u}_{j,k}\|/t_{1,k}$ converges to zero if $j \in I$. Then, the unboundedness of $\{\bar{u}_{j,k}/t_{j,k}\}_k$ implies that $t_{j,k}/t_{1,k}$ converges to 0, hence $\alpha_{j,\infty} = 0$ and (iii) is proved. \square

Similarly, we also consider the maximizers $\nabla_x S_H$ in the dual problem eq. (14) with the observed data x_k and the parameters $\{t_{j,k}\}_{j=1}^N$. The following lemma states the boundedness of the maximizers $\{\nabla_x S_H(x_k, t_{1,k}, \dots, t_{N,k})\}_k$ which will be used in the later proofs.

Lemma 3.1. *Under the assumptions (H1)-(H2) and eq. (20), for any $x \in \text{dom } J$ such that $\partial J(x) \neq \emptyset$, the sequence $\{\nabla_x S_H(x_k, t_{1,k}, \dots, t_{N,k})\}_k$ is bounded and any cluster point p is in $\partial J(x)$.*

Proof. Recall that for each $j \in \{1, \dots, N\}$, $\{v_{j,k}\}_k \subset \mathbb{R}^n$ and $\{t_{j,k}\}_k \subset (0, +\infty)$ are two sequences satisfying the assumptions in eq. (20). Denote $p_k := \nabla_x S_H(x_k, t_{1,k}, \dots, t_{N,k})$. Then, p_k is a maximizer of the maximization problem in eq. (14). Hence, for any q in $\partial J(x)$,

$$\langle x_k, p_k \rangle - J^*(p_k) - \sum_{j=1}^N t_{j,k} H_j(p_k) \geq \langle x_k, q \rangle - J^*(q) - \sum_{j=1}^N t_{j,k} H_j(q).$$

Since $q \in \partial J(x)$, we have $x \in \partial J^*(q)$, hence $J^*(p_k) \geq J^*(q) + \langle x, p_k - q \rangle$. Combining this inequality and the above one we can obtain

$$\sum_{j=1}^N t_{j,k} H_j(p_k) - \sum_{j=1}^N t_{j,k} H_j(q) \leq \langle x_k - x, p_k - q \rangle \leq \sum_{j=1}^N t_{j,k} \|v_{j,k}\| (\|p_k\| + \|q\|).$$

Here, for the second inequality above, we used the definition of x_k in eq. (20) and Cauchy-Schwarz inequality. Then, rearranging the terms and dividing by $t_{1,k}$, we get

$$(26) \quad \sum_{j=1}^N \frac{t_{j,k}}{t_{1,k}} (H_j(p_k) - \|v_{j,k}\| \|p_k\|) \leq \sum_{j=1}^N \frac{t_{j,k}}{t_{1,k}} (H_j(q) + \|v_{j,k}\| \|q\|).$$

If $\{p_k\}_k$ is not bounded, without loss of generality, we can assume $\|p_k\|$ increases to infinity. Since H_j is 1-coercive for all j , then for any $M > 0$, there exists K such that $H_j(p_k) \geq M\|p_k\|$ for any $k > K$ and any $j = 1, \dots, N$. Then, from eq. (26), for any $k > K$, we obtain

$$\sum_{j=1}^N \frac{t_{j,k}}{t_{1,k}} (M - \|v_{j,k}\|) \|p_k\| \leq \sum_{j=1}^N \frac{t_{j,k}}{t_{1,k}} (H_j(q) + \|v_{j,k}\| \|q\|).$$

The right hand side is bounded. However, since $\|p_k\|$ goes to infinity, the term for $j = 1$ on the left hand side is unbounded, while the terms for $j > 1$ is non-negative. As a result, the left hand side can be arbitrarily large, which leads to a contradiction. Therefore, we can conclude that $\{p_k\}_k$ is bounded.

For the remaining part, let p be a cluster point, then there exists a subsequence converging to p , still denoted as p_k . Since S_H solves the multi-time HJ equation eq. (13) and H_j is continuous for any j , then we have

$$\begin{aligned} \lim_{k \rightarrow +\infty} \nabla S_H(x_k, t_{1,k}, \dots, t_{N,k}) &= \lim_{k \rightarrow +\infty} (p_k, -H_1(p_k), \dots, -H_N(p_k)) \\ &= (p, -H_1(p), \dots, -H_N(p)). \end{aligned}$$

By the continuity property [66, Prop.XI.4.1.1] of the subdifferential operator ∂S_H of the convex lower semi-continuous function S_H , we can conclude that

$$(p, -H_1(p), \dots, -H_N(p)) \in \partial S_H(x, 0, \dots, 0),$$

which implies $p \in \partial J(x)$. \square

The variational behaviors of the momentum $\nabla_x S$ and the velocities u_j/t_j are presented in the following proposition. To be specific, the cluster points of the momenta and the velocities solve two optimization problems, respectively, and the two problems are dual to each other. An illustration of this result is given in fig. 7.

Proposition 3.4. *Assume (H1)-(H2) and eq. (20) hold. Let $x \in \text{dom } J$ and $\partial J(x) \neq \emptyset$. Then,*

(i) *the directional derivative of S_H corresponds to a maximization problem:*

$$(27) \quad \lim_{k \rightarrow +\infty} \frac{S_H(x_k, t_{1,k}, \dots, t_{N,k}) - S_H(x, 0, \dots, 0)}{t_{1,k}} = \max_{q \in \partial J(x)} \sum_{j=1}^N \alpha_{j,\infty} (\langle q, v_{j,\infty} \rangle - H_j(q)).$$

Moreover, let p be any cluster point of $\{\nabla_x S_H(x_k, t_{1,k}, \dots, t_{N,k})\}_k$, then,

$$(28) \quad p \in \arg \max_{q \in \partial J(x)} \sum_{j=1}^N \alpha_{j,\infty} (\langle q, v_{j,\infty} \rangle - H_j(q)).$$

(ii) *the directional derivative of S_L corresponds to the dual minimization problem:*

$$(29) \quad \begin{aligned} & \lim_{k \rightarrow +\infty} \frac{S_L(x_k, t_{1,k}, \dots, t_{N,k}) - S_L(x, 0, \dots, 0)}{t_{1,k}} \\ &= \min_{w_j \in \mathbb{R}^n} \sum_{j=1}^N \alpha_{j,\infty} (I_{\partial J(x)}^*(v_{j,\infty} - w_j) + H_j^*(w_j)). \end{aligned}$$

Moreover, if \bar{w}_j is a cluster point of $\{u_j(x_k, t_{1,k}, \dots, t_{N,k})/t_{j,k}\}_k$ for any j satisfying $\alpha_{j,\infty} \neq 0$, then

$$(30) \quad \bar{w}_j \in \arg \min_{w_j \in \mathbb{R}^n} (I_{\partial J(x)}^*(v_{j,\infty} - w_j) + H_j^*(w_j)).$$

Specially, if H_j is strictly convex and $\alpha_{j,\infty} \neq 0$ for some j , then the maximizer in eq. (28) is unique, which implies the convergence of $\nabla_x S_H(x_k, t_{1,k}, \dots, t_{N,k})$ to the unique maximizer. Similarly, for any j such that H_j is differentiable and $\alpha_{j,\infty} \neq 0$, we can conclude that $u_j(x_k, t_{1,k}, \dots, t_{N,k})/t_{j,k}$ converges to the unique minimizer in eq. (30).

Remark 3.1. *It is straightforward to obtain $\lim_{k \rightarrow +\infty} \frac{S_H(x_k, t_{1,k}, \dots, t_{N,k}) - S_H(x, 0, \dots, 0)}{\|(t_{1,k}, \dots, t_{N,k})\|_2}$ using the following computation*

$$\begin{aligned} & \lim_{k \rightarrow +\infty} \frac{S_H(x_k, t_{1,k}, \dots, t_{N,k}) - S_H(x, 0, \dots, 0)}{\|(t_{1,k}, \dots, t_{N,k})\|_2} \\ &= \lim_{k \rightarrow +\infty} \frac{S_H(x_k, t_{1,k}, \dots, t_{N,k}) - S_H(x, 0, \dots, 0)}{t_{1,k}} \cdot \frac{t_{1,k}}{\|(t_{1,k}, \dots, t_{N,k})\|_2} \\ &= \lim_{k \rightarrow +\infty} \frac{S_H(x_k, t_{1,k}, \dots, t_{N,k}) - S_H(x, 0, \dots, 0)}{t_{1,k}} \cdot \frac{1}{\|(\alpha_{1,\infty}, \dots, \alpha_{N,\infty})\|_2}, \end{aligned}$$

where the last equality follows from the assumption that $\alpha_{j,\infty} = \lim_{k \rightarrow +\infty} t_{j,k}/t_{1,k}$ for any $j = 1, \dots, N$.

Proof. Recall that the k -th spatial variable x_k is defined to be $x + \sum_{j=1}^N t_{j,k} v_{j,k}$, where $\{v_{j,k}\}_k \subset \mathbb{R}^n$ and $\{t_{j,k}\}_k \subset (0, +\infty)$ are two sequences satisfying the assumptions in eq. (20). Denote $\Delta S_k := S_H(x_k, t_{1,k}, \dots, t_{N,k}) - S_H(x, 0, \dots, 0)$.

Proof of (i): For any $q \in \partial J(x)$, by Hopf formula eq. (14), we obtain

$$\Delta S_k = \left(J^* + \sum_{j=1}^N t_{j,k} H_j \right)^* (x_k) - J(x) \geq \langle q, x_k \rangle - J^*(q) - \sum_{j=1}^N t_{j,k} H_j(q) - J(x).$$

Since $q \in \partial J(x)$, we have $J^*(q) + J(x) = \langle q, x \rangle$. Hence, together with the definition of x_k in eq. (20), we get

$$\Delta S_k \geq \langle q, x_k - x \rangle - \sum_{j=1}^N t_{j,k} H_j(q) = \sum_{j=1}^N t_{j,k} (\langle q, v_{j,k} \rangle - H_j(q)).$$

Therefore, we have

$$\liminf_{k \rightarrow +\infty} \frac{\Delta S_k}{t_{1,k}} \geq \liminf_{k \rightarrow +\infty} \sum_{j=1}^N \frac{t_{j,k}}{t_{1,k}} (\langle q, v_{j,k} \rangle - H_j(q)) = \sum_{j=1}^N \alpha_{j,\infty} (\langle q, v_{j,\infty} \rangle - H_j(q)),$$

where we recall that $\lim_{k \rightarrow +\infty} v_{j,k} = v_{j,\infty}$ and $\lim_{k \rightarrow +\infty} t_{j,k}/t_{1,k} = \alpha_{j,\infty}$ by eq. (20). Here, q is an arbitrary element in $\partial J(x)$, hence we obtain

$$(31) \quad \liminf_{k \rightarrow +\infty} \frac{\Delta S_k}{t_{1,k}} \geq \sup_{q \in \partial J(x)} \sum_{j=1}^N \alpha_{j,\infty} (\langle q, v_{j,\infty} \rangle - H_j(q)).$$

On the other hand, for any k , consider the function $\phi_k : [0, +\infty) \rightarrow \mathbb{R}$ defined by $\phi_k(t) := S_H \left(x + \sum_{j=1}^N t \alpha_{j,k} v_{j,k}, \alpha_{1,k} t, \dots \right)$ where $\alpha_{j,k} := t_{j,k}/t_{1,k}$. Since S_H is a convex function and ϕ_k is its restriction on a line, then $\phi_k \in \Gamma_0(\mathbb{R})$ with $\text{dom } \phi_k = [0, +\infty)$. Also, ϕ_k is differentiable in $(0, +\infty)$ since S_H is differentiable. The derivative of ϕ_k at $t_{1,k}$ is given by the chain rule:

$$\phi'_k(t_{1,k}) = \sum_{j=1}^N \alpha_{j,k} \left(\langle \nabla_x S_H(x_k, t_{1,k}, \dots, t_{N,k}), v_{j,k} \rangle + \frac{\partial S_H}{\partial t_j}(x_k, t_{1,k}, \dots, t_{N,k}) \right).$$

Since S_H satisfies the multi-time HJ equation eq. (13), we obtain

$$\phi'_k(t_{1,k}) = \sum_{j=1}^N \alpha_{j,k} (\langle \nabla_x S_H(x_k, t_{1,k}, \dots, t_{N,k}), v_{j,k} \rangle - H_j(\nabla_x S_H(x_k, t_{1,k}, \dots, t_{N,k}))).$$

From straightforward computation and the convexity of ϕ_k , we get

$$(32) \quad \frac{\Delta S_k}{t_{1,k}} = \frac{\phi_k(t_{1,k}) - \phi_k(0)}{t_{1,k}} \leq \phi'_k(t_{1,k}) = \sum_{j=1}^N \alpha_{j,k} (\langle p_k, v_{j,k} \rangle - H_j(p_k)),$$

where $p_k := \nabla_x S_H(x_k, t_{1,k}, \dots, t_{N,k})$.

Let p be a cluster point of $\{p_k\}$. Take a subsequence converging to p and still denote it as $\{p_k\}$. Since $p \in \partial J(x)$ by lemma 3.1 and H_j is continuous for any j , we have

$$\limsup_{k \rightarrow +\infty} \frac{\Delta S_k}{t_{1,k}} \leq \sum_{j=1}^N \alpha_{j,\infty} (\langle p, v_{j,\infty} \rangle - H_j(p)) \leq \sup_{q \in \partial J(x)} \sum_{j=1}^N \alpha_{j,\infty} (\langle q, v_{j,\infty} \rangle - H_j(q)).$$

Together with eq. (31), the equation eq. (27) is proved. Moreover, any cluster point p is a maximizer.

Proof of (ii): Here, we adopt the notations $\bar{u}_{j,k}$ and $\bar{u}_{0,k}$ defined in the proof of proposition 3.3 to represent the minimizers in the Lax formula. According to the Lax formula eq. (15) evaluated at the point $(x_k, t_{1,k}, \dots, t_{N,k})$ and by the convexity of J we deduce that

$$S_L = J(\bar{u}_{0,k}) + \sum_{j=1}^N t_{j,k} H_j^* \left(\frac{\bar{u}_{j,k}}{t_{j,k}} \right) \geq J(x) + \langle q, \bar{u}_{0,k} - x \rangle + \sum_{j=1}^N t_{j,k} H_j^* \left(\frac{\bar{u}_{j,k}}{t_{j,k}} \right),$$

for any $q \in \partial J(x)$. Since $S_L = S_H$, we have $S_L(x_k, t_{1,k}, \dots, t_{N,k}) - S_L(x, 0, \dots, 0) = \Delta S_k$. By the definition of x_k and $\bar{u}_{0,k}$, we can compute $\bar{u}_{0,k} - x = x_k - x - \sum_j \bar{u}_{j,k} = \sum_j (t_{j,k} v_{j,k} - \bar{u}_{j,k})$, hence we have

$$\frac{\Delta S_k}{t_{1,k}} \geq \sum_{j=1}^N \left(\alpha_{j,k} \langle q, v_{j,k} \rangle - \left\langle q, \frac{\bar{u}_{j,k}}{t_{j,k}} \right\rangle + \alpha_{j,k} H_j^* \left(\frac{\bar{u}_{j,k}}{t_{j,k}} \right) \right),$$

where $\alpha_{j,k} := t_{j,k}/t_{1,k}$. According to proposition 3.2 we have $\bar{u}_{j,k}/t_{j,k} \in \partial H_j(p_k)$. Therefore we get

$$\alpha_{j,k} H_j^* \left(\frac{\bar{u}_{j,k}}{t_{j,k}} \right) = \alpha_{j,k} \left(\left\langle \frac{\bar{u}_{j,k}}{t_{j,k}}, p_k \right\rangle - H_j(p_k) \right) = \left\langle \frac{\bar{u}_{j,k}}{t_{1,k}}, p_k \right\rangle - \alpha_{j,k} H_j(p_k).$$

Combining the above two equations we obtain

$$(33) \quad \begin{aligned} \frac{\Delta S_k}{t_{1,k}} &\geq \sum_{\substack{j=1 \\ \alpha_{j,\infty}=0}}^N \left(\alpha_{j,k} \langle q, v_{j,k} \rangle + \left\langle p_k - q, \frac{\bar{u}_{j,k}}{t_{1,k}} \right\rangle - \alpha_{j,k} H_j(p_k) \right) \\ &\quad + \sum_{\substack{j=1 \\ \alpha_{j,\infty} \neq 0}}^N \left(\alpha_{j,k} \left\langle q, v_{j,k} - \frac{\bar{u}_{j,k}}{t_{j,k}} \right\rangle + \alpha_{j,k} H_j^* \left(\frac{\bar{u}_{j,k}}{t_{j,k}} \right) \right). \end{aligned}$$

From proposition 3.3 (ii), $\|\bar{u}_{j,k}\|/t_{1,k}$ converges to zero if $\alpha_{j,\infty} = 0$. Also, p_k are bounded by lemma 3.1, hence the first sum in the right hand side of eq. (33) converges to zero as k approaches infinity. On the other hand, for j such that $\alpha_{j,\infty} \neq 0$, $\bar{u}_{j,k}/t_{j,k}$ is bounded by proposition 3.3 (iii). Taking a subsequence, we can assume that $\bar{u}_{j,k}/t_{j,k}$ converges to some vector, denoted as \bar{w}_j . In conclusion, as k approaches infinity in eq. (33), we have

$$(34) \quad \lim_{k \rightarrow +\infty} \frac{\Delta S_k}{t_{1,k}} \geq \sum_{\substack{j=1 \\ \alpha_{j,\infty} \neq 0}}^N \alpha_{j,\infty} (\langle q, v_{j,\infty} - \bar{w}_j \rangle + H_j^*(\bar{w}_j)) \geq \sum_{j=1}^N \alpha_{j,\infty} (\langle q, v_{j,\infty} \rangle - H_j(q)),$$

where the second inequality holds by the definition of Legendre transform eq. (11). From eq. (27), for any maximizer p in eq. (28),

$$(35) \quad \lim_{k \rightarrow +\infty} \frac{\Delta S_k}{t_{1,k}} = \sum_{j=1}^N \alpha_{j,\infty} (\langle p, v_{j,\infty} \rangle - H_j(p)).$$

Taking $q = p$ in eq. (34) and comparing it with eq. (35), we can conclude that the inequalities in eq. (34) become equalities when $q = p$. As a result, when $\alpha_{j,\infty} \neq 0$ we have $\langle p, \bar{w}_j \rangle = H_j^*(\bar{w}_j) + H_j(p)$, which implies that $p \in \partial H_j^*(\bar{w}_j)$. Then, we deduce that

$$(36) \quad \lim_{k \rightarrow +\infty} \frac{\Delta S_k}{t_{1,k}} = \sum_{\substack{j=1 \\ \alpha_{j,\infty} \neq 0}}^N \alpha_{j,\infty} (\langle p, v_{j,\infty} - \bar{w}_j \rangle + H_j^*(\bar{w}_j)).$$

On the other hand, for an arbitrary $q \in \partial J(x)$, by eq. (34) and eq. (36), we have

$$\begin{aligned} \sum_{\substack{j=1 \\ \alpha_{j,\infty} \neq 0}}^N \alpha_{j,\infty} (\langle p, v_{j,\infty} - \bar{w}_j \rangle + H_j^*(\bar{w}_j)) &= \lim_{k \rightarrow +\infty} \frac{\Delta S_k}{t_{1,k}} \\ &\geq \sum_{\substack{j=1 \\ \alpha_{j,\infty} \neq 0}}^N \alpha_{j,\infty} (\langle q, v_{j,\infty} - \bar{w}_j \rangle + H_j^*(\bar{w}_j)), \end{aligned}$$

which implies that $\langle p - q, v_{j,\infty} - \bar{w}_j \rangle \geq 0$ for any $q \in \partial J(x)$, when $\alpha_{j,\infty} \neq 0$. By eq. (7) and eq. (10), we can deduce that $v_{j,\infty} - \bar{w}_j \in N_{\partial J(x)}(p) = \partial I_{\partial J(x)}(p)$. proposition 2.5 gives the equality $\langle p, v_{j,\infty} - \bar{w}_j \rangle = I_{\partial J(x)}^*(v_{j,\infty} - \bar{w}_j)$. Then, eq. (29) follows from this equality and eq. (36).

It remains to prove eq. (30). Consider any j such that $\alpha_{j,\infty} \neq 0$. Define $f : \mathbb{R}^n \rightarrow \mathbb{R}$ by $f(w) := I_{\partial J(x)}^*(v_{j,\infty} - w) + H_j^*(w)$. Then it suffices to prove $0 \in \partial f(\bar{w}_j)$. So far, we have proved $p \in \partial H_j^*(\bar{w}_j)$ and $v_{j,\infty} - \bar{w}_j \in \partial I_{\partial J(x)}(p)$, which implies $p \in \partial I_{\partial J(x)}^*(v_{j,\infty} - \bar{w}_j)$. By straightforward computation and proposition 2.3,

$$\partial f(\bar{w}_j) = -\partial I_{\partial J(x)}^*(v_{j,\infty} - \bar{w}_j) + H_j^*(\bar{w}_j) \ni -p + p = 0.$$

Therefore, \bar{w}_j is a minimizer of f , which concludes the proof. \square

The above proposition provides the explicit formulas for the variations of S , $\nabla_x S$ and $\frac{u_j}{t_j}$ where u_j denotes the j -th component of the minimizer of the decomposition model in the form of eq. (15). Specifically, the limits of these quantities are related to the two optimization problems given by eqs. (28) and (30). From the perspective of image processing, given an observed image x_k which is a summation of a constant component x and other components $t_{j,k}v_{j,k}$, the decomposition model eq. (15) gives $N + 1$ components. In these $N + 1$

components, one component converges to the constant component x and the other components u_j vanish as the parameters $t_{j,k}$ approach zero, by proposition 3.3. Then, proposition 3.4(ii) states that the component u_j converges to 0 from a direction \bar{w}_j [82, p. 197]. On the other hand, proposition 3.4(i) provides a representation formula for the cluster point of the maximizers of the dual problem in the form of eq. (14).

4. UNIQUENESS OF THE CONVEX SOLUTIONS TO THE MULTI-TIME HAMILTON-JACOBI EQUATIONS

In the previous section, we have discussed the relation of the optimization problems in the Hopf formula and Lax formula with the classical solution of the multi-time HJ equation. In fact, some results can be generalized to weaker assumptions in which case the solution provided by Hopf and Lax formulas is not classical. In this section, we prove that the only convex solution is given by the two formulas.

In the field of PDEs, a type of solution called viscosity solution is considered for solving the HJ equation when no classical solution exists. The uniqueness of the viscosity solution has been widely studied under different assumptions [17, 19]. However, the functions in convex analysis and optimization may take the value $+\infty$, which is an unusual condition in the field of PDEs. Therefore, to maintain the connection of the HJ equations and convex optimization problems, we consider the convex solution which may be infinity in some area and prove the uniqueness using the techniques in convex analysis.

We start with the proof for the classical convex solution, in order to demonstrate the idea of utilizing the convexity assumptions. After that, we state the uniqueness of nonsmooth convex solution under more general assumptions in corollary 4.1. When proving the uniqueness of the classical convex solution, we assume the properties (H1) and (H2) hold. Moreover, the solution S satisfies:

- (S1) $S \in \Gamma_0(\mathbb{R}^n \times [0, +\infty)^N) \cap C^1(\mathbb{R}^n \times (0, +\infty)^N)$;
- (S2) S solves the multi-time Hamilton-Jacobi equation eq. (13).

As it is discussed in section 3, S_H defined in the Hopf formula eq. (14) is a solution satisfying the assumptions (S1) and (S2). Hence, we just need to prove $S = S_H$ for any S satisfying (S1)-(S2). First, we consider the single-time case when the time dimension $N = 1$, and formulate its Legendre transform $S^*(p, E^-)$ for $p \in \mathbb{R}^n$ and $E^- \in \mathbb{R}$ in the following lemma.

Lemma 4.1. *Assume (H1)-(H2) hold and S satisfies (S1)-(S2). Let $N = 1$. Then there exists a convex function $\tilde{H} : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$, such that $S^*(p, E^-) = J^*(p) + I_V(p, E^-)$, where $V := \{(p, E^-) : E^- \leq -\tilde{H}(p)\}$.*

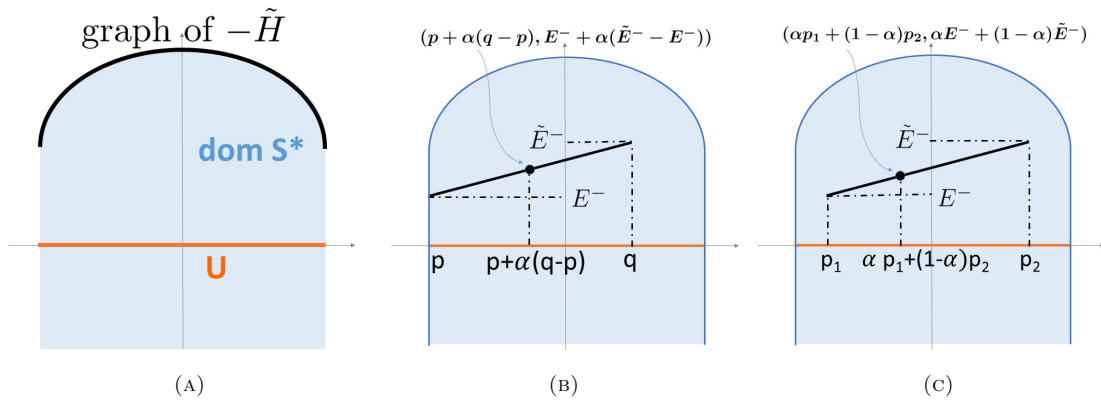


FIGURE 8. Illustrations for different steps in the proof of lemma 4.1.

Proof. In this proof, we only consider the single-time HJ equation. For the single-time case, H is used to denote the Hamiltonian, instead of H_1 , for simplicity. First, consider the domain of S^* . For each $p \in \mathbb{R}^n$, define

$$(37) \quad \tilde{H}(p) := \inf\{-E^- : (p, E^-) \in \text{dom } S^*\} \in \bar{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}.$$

For the illustration of this definition, see fig. 8a. The function \tilde{H} defined here is an extended-valued function taking values in $\bar{\mathbb{R}}$. In the last step of this proof, we will show the convexity and specify the range of this function. From this definition, it is obvious that $\text{dom } S^* \subseteq V$, where $V = \{(p, E^-) : E^- \leq -\tilde{H}(p)\}$, as defined in the statement of this lemma. Moreover, denote $V_1 = \{(p, E^-) : E^- < -\tilde{H}(p)\}$, then we prove $V_1 \subseteq \text{dom } S^*$ by using the monotonicity of $S^*(p, \cdot)$. To be specific, let $p \in \mathbb{R}^n$ and $-\infty < \tilde{E}^- \leq E^- < +\infty$, then, we have

$$(38) \quad S^*(p, \tilde{E}^-) = \sup_{x \in \mathbb{R}^n, t \geq 0} \langle p, x \rangle + t\tilde{E}^- - S(x, t) \leq \sup_{x \in \mathbb{R}^n, t \geq 0} \langle p, x \rangle + tE^- - S(x, t) = S^*(p, E^-).$$

Hence, $S^*(p, E^-)$ is non-decreasing with respect to E^- . As a result, $(p, E^-) \in \text{dom } S^*$ implies $\{p\} \times (-\infty, E^-] \subseteq \text{dom } S^*$. Therefore we obtain $V_1 \subseteq \text{dom } S^* \subseteq V$.

In the next step, we prove $\text{dom } S^* = V$.

Denote $U := \{p \in \mathbb{R}^n : \tilde{H}(p) < +\infty\}$ (see fig. 8a). Here and after in this section, we use the bold character $\mathbf{0}$ to denote the zero vector in \mathbb{R}^n . Since U is the projection of $\text{dom } S^*$ along the direction $(\mathbf{0}, 1)$, U is a convex set. Let $p \in \text{ri } U$. Take $E^- < -\tilde{H}(p)$, then $\partial S^*(p, E^-) \neq \emptyset$ because $(p, E^-) \in \text{ri dom } S^*$. Let $(x, t) \in \partial S^*(p, E^-)$, which implies $(p, E^-) \in \partial S(x, t)$. If $t > 0$, then $E^- = \frac{\partial S}{\partial t}(x, t)$ and $p = \nabla_x S(x, t)$. Since S satisfies the HJ equation eq. (13), $E^- + H(p) = 0$. In other words, if $(x, t) \in \partial S^*(p, E^-)$ with $E^- \neq -H(p)$, then we can conclude that $t = 0$. Therefore, for any $E^- < -\tilde{H}(p)$ and $E^- \neq -H(p)$, by proposition 2.6, the directional derivative of S^* in the direction $(\mathbf{0}, 1)$ is:

$$(S^*)'((p, E^-), (\mathbf{0}, 1)) = \sup_{(x, t) \in \partial S^*(p, E^-)} \langle (x, t), (\mathbf{0}, 1) \rangle = \sup_{(x, t) \in \partial S^*(p, E^-)} \langle (x, 0), (\mathbf{0}, 1) \rangle = 0.$$

As a result, $S^*(p, \cdot)$ is a constant function in its domain. Denote this value as $f(p)$. By the continuity of S^* when restricting to the straight line $\{p\} \times \mathbb{R}$, the value $S^*(p, -\tilde{H}(p))$ is also $f(p)$ if $\tilde{H}(p)$ is finite. Hence, $S^*(p, E^-) = f(p)$ for any $p \in \text{ri } U$ and $E^- \leq -\tilde{H}(p)$.

Now, we consider the case when $p \in U \setminus \text{ri } U$. For the illustration, see fig. 8b. Let $E^- < -\tilde{H}(p)$. Take $q \in \text{ri } U$ and $\tilde{E}^- < -\tilde{H}(q)$, then by proposition 2.2,

$$(39) \quad S^*(p, E^-) = \lim_{\alpha \rightarrow 0^+} S^*(p + \alpha(q - p), E^- + \alpha(\tilde{E}^- - E^-)) = \lim_{\alpha \rightarrow 0^+} f(p + \alpha(q - p)).$$

Hence, the value of $S^*(p, E^-)$ does not depend on E^- if $E^- < -\tilde{H}(p)$. Denote this value as $f(p)$. By continuity, $S^*(p, -\tilde{H}(p)) = f(p)$ if $\tilde{H}(p)$ is finite. Therefore, we have proved that the domain of S^* coincides with the set V and $S^*(p, E^-) = f(p)$ in the domain of S^* .

Then, we prove $f = J^*$ when restricting to $\text{dom } f$. By setting $f(p) = +\infty$ if $p \notin U$, we can regard f as a function from \mathbb{R}^n to $\mathbb{R} \cup \{+\infty\}$. It is not hard to check the convexity of f . To be specific, for any $p_1, p_2 \in \text{dom } f$ and $\alpha \in (0, 1)$, choose $E^- < -\tilde{H}(p_1)$ and $\tilde{E}^- < -\tilde{H}(p_2)$ (see fig. 8c), then we have

$$\begin{aligned} f(\alpha p_1 + (1 - \alpha)p_2) &= S^*(\alpha p_1 + (1 - \alpha)p_2, \alpha E^- + (1 - \alpha)\tilde{E}^-) \\ &\leq \alpha S^*(p_1, E^-) + (1 - \alpha)S^*(p_2, \tilde{E}^-) \\ &= \alpha f(p_1) + (1 - \alpha)f(p_2). \end{aligned}$$

Hence f is a convex function taking values in $\mathbb{R} \cup \{+\infty\}$. Also, for each $x \in \mathbb{R}^n$, we have

$$\begin{aligned} J(x) &= S(x, 0) = \sup_{(p, E^-) \in V} \langle x, p \rangle - S^*(p, E^-) \\ &= \sup_{\substack{p \in \mathbb{R}^n \\ \exists E^- \mid (p, E^-) \in V}} \langle x, p \rangle - f(p) = \sup_{p \in \text{dom } f} \langle x, p \rangle - f(p) = f^*(x). \end{aligned}$$

Therefore, $f^{**} = J^*$, which implies $\text{ri dom } f = \text{ri dom } J^*$ and $f(p) = J^*(p)$ if $p \in \text{ri dom } f$. Moreover, according to proposition 2.2 and eq. (39), we deduce that

$$f(p) = \lim_{\alpha \rightarrow 0^+} f(p + \alpha(q - p)) = \lim_{\alpha \rightarrow 0^+} J^*(p + \alpha(q - p)) = J^*(p),$$

for any $p \in \text{dom } f \setminus \text{ri dom } f$ and $q \in \text{ri dom } f$. As a result we have $f = J^*$ in the domain of definition. In conclusion, we get the following formula for S^*

$$(40) \quad S^*(p, E^-) = J^*(p) + I_V(p, E^-).$$

The final part is to prove that \tilde{H} is a convex function taking values in $\mathbb{R} \cup \{+\infty\}$.

First, we prove that \tilde{H} cannot take the value $-\infty$ by contradiction. Suppose there exists $p \in \mathbb{R}^n$ such that $\tilde{H}(p)$ equals $-\infty$. Then, by definition of \tilde{H} we have $\{p\} \times \mathbb{R} \subseteq \text{dom } S^*$. Together with the formula of S^* in eq. (40), we derive

$$\{p\} \times \mathbb{R} \times \{J^*(p)\} \subseteq \text{epi } S^*.$$

Therefore, $(\mathbf{0}, 1, 0)$ and $(\mathbf{0}, -1, 0)$ are in the asymptotic cone of $\text{epi } S^*$ by definition eq. (8). Then, by proposition 2.1, for any $q \in U$, we obtain

$$\{q\} \times \mathbb{R} \times \{J^*(q)\} \subseteq \text{epi } S^*,$$

which implies $\{q\} \times \mathbb{R} \subseteq \text{dom } S^*$. Since q is an arbitrary vector in U , we deduce that $\text{dom } S^* = U \times \mathbb{R}$. Moreover, according to eq. (40), the function S^* is a constant on the line $\{q\} \times \mathbb{R}$ for any $q \in U$, which implies that the directional derivative of S^* in the direction $(\mathbf{0}, 1)$ is zero. In other words, we have

$$(41) \quad (S^*)'((p, E^-), (\mathbf{0}, 1)) = 0 \text{ for any } p \in U \text{ and } E^- \in \mathbb{R}.$$

On the other hand, consider any $y \in \mathbb{R}^n$ and $s > 0$ such that $\partial S(y, s)$ is nonempty. Let $(p, E^-) \in \partial S(y, s)$. This implies $(y, s) \in \partial S^*(p, E^-)$. Hence, according to proposition 2.6, we get

$$(S^*)'((p, E^-), (\mathbf{0}, 1)) \geq \sup_{(x, t) \in \partial S^*(p, E^-)} \langle (x, t), (\mathbf{0}, 1) \rangle \geq \langle (y, s), (\mathbf{0}, 1) \rangle = s > 0,$$

which contradicts eq. (41). Therefore, \tilde{H} cannot take the value $-\infty$.

At last, the convexity of \tilde{H} follows from the convexity of $\text{dom } S^*$. In fact, $\text{epi } \tilde{H} = \{(p, -E^-) : (p, E^-) \in \text{dom } S^*\}$, which is a reflection of the convex set $\text{dom } S^*$, hence it is also convex. Therefore, \tilde{H} is a convex function from \mathbb{R}^n to $\mathbb{R} \cup \{+\infty\}$. \square

Based on this lemma, the following proposition states the uniqueness result. It can be easily seen in the above lemma that the Legendre transform of S has a similar form as S_H^* . Actually, the following proposition is proved by equating the two functions S^* and S_H^* .

Proposition 4.1. *The solution to the multi-time Hamilton-Jacobi equation is unique. Specifically, under the assumptions (H1) and (H2), if S satisfies (S1)-(S2), then $S = S_H$.*

Proof. In the proof of this proposition, we first consider the case of single-time. Let $N = 1$, and H be the Hamiltonian.

From lemma 4.1, it is proved that $S^*(p, E^-) = J^*(p) + I_V(p, E^-)$, where $V = \{(p, E^-) : E^- \leq -\tilde{H}(p)\}$ and \tilde{H} is a convex function whose domain is the projection of $\text{dom } S^*$ along $(\mathbf{0}, 1)$. Moreover, $\text{ri dom } J^* = \text{ri dom } \tilde{H}$ (note that the domains of \tilde{H} and f are the same).

First, we prove that $\tilde{H}(p) = H(p)$ for any $p \in \text{ri dom } \tilde{H}$ by contradiction. Assume there exists $p \in \text{ri dom } \tilde{H}$ such that $\tilde{H}(p) \neq H(p)$. Let $E^- = -\tilde{H}(p)$. Then, by proposition 2.3 and eq. (10), we deduce that

$$(42) \quad \partial S^*(p, E^-) = \partial J^*(p) \times \{0\} + N_V(p, E^-) = \partial J^*(p) \times \{0\} + \{t(v, 1) : v \in \partial \tilde{H}(p), t \geq 0\},$$

where the last equality holds because V is the reflection of $\text{epi } \tilde{H}$. Here $N_V(p, E^-)$ denotes the normal cone of the set V at (p, E^-) . Let $x_0 \in \partial J^*(p)$, $t > 0$ and $v \in \partial \tilde{H}(p)$. Denote $x = x_0 + tv$. Then, by eq. (42) we have $(x, t) \in \partial S^*(p, E^-)$, which implies $(p, E^-) \in \partial S(x, t)$. However, $E^- + H(p) = -\tilde{H}(p) + H(p) \neq 0$, hence the HJ equation eq. (13) does not hold at (x, t) , which is a contradiction. Therefore, $\tilde{H} = H$ when restricting to the relative interior of the domain of \tilde{H} , which implies

$$S^*(p, E^-) = J^*(p) + I\{E^- \leq -\tilde{H}(p)\} = J^*(p) + I\{E^- \leq -H(p)\} = S_H^*(p, E^-),$$

for any $p \in \text{ri dom } \tilde{H}$.

Actually, the values of any convex lower semi-continuous function on the relative boundary of its domain is fully determined by the values in the relative interior. It is not hard to check that

$$\text{ri dom } S^* = \text{ri dom } S_H^* = \{(p, E^-) : p \in \text{ri dom } J^*, E^- < -H(p)\}.$$

Hence, we have proved that S^* and S_H^* agree in the relative interior of the domain. Therefore, $S^* = S_H^*$ in the whole domain, which implies $S = S_H$ and gives the uniqueness of the convex solution to the single-time HJ equation.

Then, we can consider the case of multi-time. Now, we assume $N > 1$. It suffices to prove S and S_H coincide for any $x \in \mathbb{R}^n$ and any $t_1, \dots, t_N > 0$. Let $\alpha_1, \dots, \alpha_N$ be arbitrary positive real numbers and denote $\alpha := (\alpha_1, \dots, \alpha_N)$. Define $T(x, s) := S(x, s\alpha_1, \dots, s\alpha_N)$ for any $x \in \mathbb{R}^n$ and $s \geq 0$. Then $T \in \Gamma_0(\mathbb{R}^{n+1})$. We can compute the gradient of T with respect to s for any $x \in \mathbb{R}^n$ and $s > 0$ using chain rule and the assumption that S satisfies the multi-time HJ equation eq. (13) to obtain

$$\frac{\partial T(x, s)}{\partial s} = \sum_{j=1}^N \alpha_j \frac{\partial S(x, s\alpha)}{\partial t_j} = - \sum_{j=1}^N \alpha_j H_j(\nabla_x S(x, s\alpha)) = - \sum_{j=1}^N \alpha_j H_j(\nabla_x T(x, s)).$$

It is easy to check that T satisfies the initial condition given by J , i.e. $T(x, 0) = J(x)$ for any $x \in \mathbb{R}^n$. Hence, T is a solution to the single-time HJ equation with Hamiltonian $H = \sum_{j=1}^N \alpha_j H_j$, which is finite-valued, 1-coercive and strictly convex. Therefore, for the single-time HJ equation, the conditions (H1)-(H2) and (S1)-(S2) are satisfied. Then, the solution T is unique and equal to the Hopf formula with respect to the Hamiltonian H . Hence, for any $x \in \mathbb{R}^n, s > 0$ and any $\alpha_1, \dots, \alpha_N > 0$, we have

$$S(x, s\alpha_1, \dots, s\alpha_N) = (J^* + sH)^*(x) = \left(J^* + \sum_{j=1}^N s\alpha_j H_j \right)^*(x) = S_H(x, s\alpha_1, \dots, s\alpha_N).$$

Therefore, $S = S_H$ in the relative interior of the domain, which implies $S = S_H$ in the whole space, because of the lower semi-continuity of S and S_H . The uniqueness of the solution to the multi-time HJ equation follows. \square

One can actually apply the above arguments to weaker assumptions and obtain a generalized result, which is stated in the following corollary. In this generalized result, it is possible that the solution S is not a classical solution, hence the subgradients of S , instead of the gradients, are assumed to satisfy the HJ equation, which is a natural generalization of the classical solution when we want to consider the solution which is convex and lower semi-continuous.

Corollary 4.1. *Let $J \in \Gamma_0(\mathbb{R}^n)$, and H_1, H_2, \dots, H_N be arbitrary extended-valued functions defined on \mathbb{R}^n . Assume there exists a function $S \in \Gamma_0(\mathbb{R}^n \times [0, +\infty)^N)$ satisfying:*

- (i) *If $p \in \mathbb{R}^n$ and $E_1^-, \dots, E_N^- \in \mathbb{R}$ satisfy $(p, E_1^-, \dots, E_N^-) \in \partial S(x, t_1, \dots, t_N)$ for some $x \in \mathbb{R}^n$ and $t_1, \dots, t_N > 0$, then $E_j^- + H_j(p) = 0$ for any $j = 1, \dots, N$.*
- (ii) *$S(x, 0, \dots, 0) = J(x)$ for any $x \in \mathbb{R}^n$.*

Then, the following statements hold:

1. *For the case of single time, i.e. $N = 1$, denote $H = H_1$ to be the Hamiltonian. If there exists $x \in \mathbb{R}^n$, $t > 0$ such that $S(x, t) \neq +\infty$, then S is unique and $S = F^*$, where F is defined by*

$$(43) \quad F(p, E^-) := J^*(p) + I\{E^- \leq -H(p)\} + I\{p \in \text{ri dom } J^*\},$$

for any $p \in \mathbb{R}^n$ and $E^- \in \mathbb{R}$. Moreover, the restriction of H on $\text{ri dom } J^$ is finite-valued and convex.*

2. *For the multi-time case, i.e. $N > 1$, if \tilde{S} is another function satisfying the assumptions (i)-(ii) with $\text{ri dom } \tilde{S} = \text{ri dom } S$, then $\tilde{S} = S$. In other words, the solution is unique when the relative interior of the domain is given.*

Proof. The proof of this corollary is similar to the proof of proposition 4.1, so we just give a brief sketch here. First, we adjust the proof of lemma 4.1 by changing the gradients of S to the subgradients of S . The argument still holds because we assume in (i) that the subgradients of S satisfy the HJ equation. Then, we draw the same conclusion as in lemma 4.1. In other words, with the function \tilde{H} defined in eq. (37), we have

$$(44) \quad S^*(p, E^-) = J^*(p) + I\{E^- \leq -\tilde{H}(p)\}.$$

Also, the part of $N = 1$ in the proof of proposition 4.1 still holds. So we derive that the two functions \tilde{H} and H coincide in the relative interior of $\text{dom } J^*$. Together with eq. (44), we derive eq. (43), and hence the first statement in this corollary follows.

For the case when $N > 1$, it suffices to prove that S and \tilde{S} coincide in the relative interior of the domain. Let (y, t_1, \dots, t_N) be an arbitrary point in $\text{ri dom } S$. It remains to prove that S and \tilde{S} are equal at the point (y, t_1, \dots, t_N) . Notice that we have $t_i > 0$ for any $i = 1, \dots, N$, then we can choose the positive number α_i

in the proof of proposition 4.1 to be t_i for any i . As in the proof of proposition 4.1, we define the functions T and \tilde{T} by

$$T(x, s) := S(x, s\alpha_1, \dots, s\alpha_N), \quad \text{and} \quad \tilde{T}(x, s) := \tilde{S}(x, s\alpha_1, \dots, s\alpha_N),$$

for any $x \in \mathbb{R}^n$ and $s \geq 0$. Since there exists a point (y, t_1, \dots, t_N) in the relative interior of $\text{dom } S$, one can easily check that the assumptions in [66, Thm.XI.3.2.1] hold. Then, by [66, Thm.XI.3.2.1], the chain rule for the subgradients of S holds. Similarly, the chain rule also holds for the subgradients of \tilde{S} . Therefore, the argument in the proof of proposition 4.1 in the multi-time case remains valid by changing the gradients to the subgradients. As a result, we conclude that both T and \tilde{T} solve the single-time HJ equation with the Hamiltonian $\sum_{j=1}^N \alpha_j H_j$. Then, by the first statement in this corollary, we have $T \equiv \tilde{T}$, which implies that S and \tilde{S} coincide at the point (y, t_1, \dots, t_N) , and the proof is complete. \square

5. A REGULARIZATION METHOD FOR THE DEGENERATE CASES

In the previous two sections, we discussed the relation between some optimization problems and the multi-time HJ equations under the assumptions (H1) and (H2). In general, if those assumptions are not satisfied, some results may collapse. For example, if there is no strictly convex Hamiltonian, then the solution may be non-differentiable, which leads to the non-uniqueness of the maximizer p (called momentum) in the Hopf formula eq. (14). Also, the minimizer u in the Lax formula eq. (15) may be non-unique if the Hamiltonians are not differentiable. However, these are two common situations for optimization problems such as the decomposition models. In fact, any norm or indicator function is neither strictly convex nor differentiable. As a result, it is an important problem to select a meaningful momentum p or minimizer u in the solution set when it contains more than one element.

In this section, we propose a regularization method to select a unique momentum p and a unique minimizer u simultaneously, and provide the representation formulas for both selected quantities by using the results stated in the previous sections. Intuitively, to select a minimizer u , we modify the degenerate term by adding λH to it where λ is a positive parameter and H is a differentiable function satisfying (H1). When λ approaches zero, the minimizer of the modified problem will converge to the unique minimizer \bar{u} in the solution set of the original problem which minimizes the function H . The procedure to select p is the same except performing the inf-convolution with $\lambda H^*(\cdot/\lambda)$ to the degenerate term instead of the addition of λH .

In the literature, the special case selecting the momentum p using inf-convolution with $\|\cdot\|^2/(2\lambda)$ is well-known as Moreau-Yosida approximation, which is introduced, for instance, in [8, Thm.2, p.144] and [27, Thm.3.1, p.54]. Generally, a Moreau-Yosida based regularization method usually selects a unique minimizer u only or a momentum p only, but not both. Our contribution here is that we consider the primal problem and the dual problem simultaneously. In other words, one can select the momentum p and the minimizer u at the same time using our method. This analysis can be adapted easily to other decomposition models with more degenerate terms. Moreover, one can also use the same procedure with other function H or even use two different functions in the two added terms. One alternative choice is $\|\cdot\|_\alpha^\alpha/\alpha$ for any $\alpha > 1$, for example. In fact, if H is chosen to be any non-negative, finite-valued, 1-coercive, differentiable and strictly convex function, the statements in this section still hold. To be specific, the proofs of lemma 5.1, lemma 5.3 and proposition 5.1 hold after subtle adjustment, and one can use subdifferential calculus to prove lemma 5.2. In this paper, for simplicity, we mainly focus on the quadratic regularization terms, which are usually preferred in practice because of the simplicity and efficiency of numerical implementation.

Now, we focus on a specific decomposition model, and the regularization function H is chosen to be $\|\cdot\|_2^2/2$. Some other models can be analyzed using similar arguments. Let $\|\cdot\|$ and $\|\!\|\!\cdot\!\|$ be two arbitrary norms whose dual norms are denoted as $\|\cdot\|_*$ and $\|\!\|\!\cdot\!\|_*$. In fact, all the results remain valid if $\|\cdot\|$ and $\|\!\|\!\cdot\!\|$ are two semi-norms, in which case the corresponding dual norms $\|\cdot\|_*$ and $\|\!\|\!\cdot\!\|_*$ are finite in some subspaces and equal to $+\infty$ otherwise. The set of minimizers is defined as follows

$$U(x, t) := \arg \min_{u \in \mathbb{R}^n} \|u\| + I\{\|\!\|x - u\!\|_* \leq t\}.$$

We can regard the minimal value as a solution to the HJ equation given by the Lax formula with spatial variable $x \in \mathbb{R}^n$ and time variable $t > 0$ and define

$$S(x, t) := \min_{u \in \mathbb{R}^n} \|u\| + I\{\|x - u\|_* \leq t\}.$$

Note that in the corresponding HJ equation, the initial function is $\|\cdot\|$ and the Hamiltonian is $\|\cdot\|_*$, hence the assumption (H1) is not satisfied. As a result, we need to apply the regularization method in this example. For simplicity we also use F_1, F_2 to denote these two norms, then $F_2^*(y) = I\{\|y\|_* \leq t\}$. We assume $t = 1$ and drop the variable t in the remainder of this section because the variation of t is not considered in this problem. Then, we can rewrite the problem as the following

$$(45) \quad \begin{aligned} U(x) &= \arg \min_{u \in \mathbb{R}^n} F_1(u) + F_2^*(x - u), \\ S(x) &= \min_{u \in \mathbb{R}^n} F_1(u) + F_2^*(x - u). \end{aligned}$$

In fact, there are in practice some useful models in the literature which can fit in this form. Now, we give two examples. In what follows, we use $\|\cdot\|_{TV}$, $\|\cdot\|_E$ and $\|\cdot\|_G$ to denote the discrete total variation semi-norm, the discrete E -norm and the discrete G -norm, respectively. First, in [9, 10], it is shown that the Meyer's model in the following form

$$\arg \min_{u \in \mathbb{R}^n} \|u\|_{TV} + \alpha \|x - u\|_G$$

is equivalent to

$$\arg \min_{u \in \mathbb{R}^n} \|u\|_{TV} + I\{\|x - u\|_G \leq \beta\},$$

for some suitable positive parameter β . In this example, both F_1 and F_2 are the discrete total variation because the discrete G -norm is the dual norm of $\|\cdot\|_{TV}$. Similarly, another Meyer's model stated as follows

$$\arg \min_{u \in \mathbb{R}^n} \|u\|_{TV} + \alpha \|x - u\|_E$$

is equivalent to

$$\arg \min_{u \in \mathbb{R}^n} \|u\|_{TV} + I\{\|x - u\|_E \leq \beta\},$$

for some suitable positive parameter β [11]. In this example, the functions F_1 and F_2 are the discrete total variation and the dual norm of the discrete E -norm, respectively.

As mentioned above, we apply two operators to the function F_1 and obtain its approximation

$$(46) \quad F_{1,\lambda,\mu} := \left(F_1 + \frac{\lambda}{2} \|\cdot\|_2^2 \right) \square \frac{1}{2\mu} \|\cdot\|_2^2,$$

where $\lambda, \mu > 0$ are small regularization parameters. Here, we choose to modify the function F_1 , but one may instead apply the operators to the function F_2 and the analysis is similar. Then, the problem reads

$$(47) \quad \begin{aligned} u_{\lambda,\mu}(x) &= \arg \min_{u \in \mathbb{R}^n} F_{1,\lambda,\mu}(u) + F_2^*(x - u), \\ S_{\lambda,\mu}(x) &= \min_{u \in \mathbb{R}^n} F_{1,\lambda,\mu}(u) + F_2^*(x - u). \end{aligned}$$

We expand the inf-convolution to get

$$(48) \quad \begin{aligned} u_{\lambda,\mu} &= x - w_{\lambda,\mu}, \\ (v_{\lambda,\mu}, w_{\lambda,\mu}) &:= \arg \min_{v, w \in \mathbb{R}^n} F_1(v) + \frac{\lambda}{2} \|v\|_2^2 + F_2^*(w) + \frac{1}{2\mu} \|x - v - w\|_2^2, \\ S_{\lambda,\mu} &:= \min_{v, w \in \mathbb{R}^n} F_1(v) + \frac{\lambda}{2} \|v\|_2^2 + F_2^*(w) + \frac{1}{2\mu} \|x - v - w\|_2^2. \end{aligned}$$

Here and later in this section, we omit the variable x when there is no ambiguity.

By introducing the quadratic terms, the uniqueness of $(v_{\lambda,\mu}, w_{\lambda,\mu})$ and the differentiability of $S_{\lambda,\mu}$ are guaranteed. When the parameters λ and μ converge to zero in a comparable rate, the reasonable minimizer u and momentum p are selected. In fact, they are the elements with the minimal l^2 norms in the target sets $U(x)$ and $\partial S(x)$. The detailed statements are listed as follows.

Lemma 5.1. *For any $\lambda, \mu > 0$, there is a unique minimizer $(v_{\lambda,\mu}, w_{\lambda,\mu})$ to the problem eq. (48). Moreover, for any positive constant K , the sets $\{v_{\lambda,\mu} : \lambda, \mu \in (0, K)\}$ and $\{w_{\lambda,\mu} : \lambda, \mu \in (0, K)\}$ are bounded.*

Proof. It is easy to check that the objective function in eq. (48) is 1-coercive and strictly convex, because of the 1-coercivity and strict convexity of the quadratic terms. Therefore, there exists a unique minimizer $(v_{\lambda,\mu}, w_{\lambda,\mu})$.

Setting $w = x - v$ and $v \in U(x)$ in eq. (48) and comparing it with eq. (45), we obtain

$$S_{\lambda,\mu}(x) \leq \min_{v \in U(x)} F_1(v) + \frac{\lambda}{2} \|v\|_2^2 + F_2^*(x - v) = S(x) + \lambda \min_{v \in U(x)} \frac{1}{2} \|v\|_2^2.$$

Denote $C := S(x) + \min_{v \in U(x)} \frac{K}{2} \|v\|_2^2$, where K is an arbitrary positive number as defined in the statement. Then C is independent of λ and μ , and $S_{\lambda,\mu}(x) \leq C$ when $0 < \lambda < K$. From this inequality and the definition of $S_{\lambda,\mu}(x)$ in eq. (48), we can derive a bound for $x - v_{\lambda,\mu} - w_{\lambda,\mu}$ that reads

$$(49) \quad \|x - v_{\lambda,\mu} - w_{\lambda,\mu}\|_2^2 \leq 2\mu S_{\lambda,\mu}(x) \leq 2C\mu \leq 2CK \text{ whenever } \mu < K.$$

Therefore, $v_{\lambda,\mu} + w_{\lambda,\mu}$ is bounded by the constant $\|x\|_2 + \sqrt{2CK}$ when we assume $\lambda, \mu \in (0, K)$.

Then, from the constraint given by the indicator function F_2^* in the minimization problem eq. (48), we have $\|w_{\lambda,\mu}\|_* \leq 1$, which implies the boundedness of $w_{\lambda,\mu}$ because all the norms are equivalent in the finite-dimensional space \mathbb{R}^n . As a result, $v_{\lambda,\mu}$ is also bounded whenever $\lambda, \mu \in (0, K)$. Then the conclusion follows. \square

Lemma 5.2. *Let $v_{\lambda,\mu}$ and $w_{\lambda,\mu}$ be defined by eq. (48). Then, we have $\lim_{\lambda,\mu \rightarrow 0+} v_{\lambda,\mu} + w_{\lambda,\mu} = x$. Any cluster point of $v_{\lambda,\mu}$ is also a cluster point of $u_{\lambda,\mu}$ and vice versa. Moreover, any cluster point of $u_{\lambda,\mu}$ and $v_{\lambda,\mu}$ is in $U(x)$.*

Proof. The convergence of $v_{\lambda,\mu} + w_{\lambda,\mu}$ to x follows from eq. (49). Since $u_{\lambda,\mu} = x - w_{\lambda,\mu}$, any cluster point of $u_{\lambda,\mu}$ is also a cluster point of $v_{\lambda,\mu}$ and vice versa. It remains to show that any cluster point of $v_{\lambda,\mu}$ is in $U(x)$.

By the definition of $(v_{\lambda,\mu}, w_{\lambda,\mu})$, we have

$$(50) \quad \begin{aligned} (v_{\lambda,\mu}, w_{\lambda,\mu}) &= \arg \min_{v,w \in \mathbb{R}^n} F_1(v) + \frac{\lambda}{2} \|v\|_2^2 + I\{\|w\|_* \leq 1\} + \frac{1}{2\mu} \|x - v - w\|_2^2 \\ &= \arg \min_{v,w \in \mathbb{R}^n} \mu F_1(v) + \frac{\lambda\mu}{2} \|v\|_2^2 + I\{\|w\|_* \leq 1\} + \frac{1}{2} \|x - v - w\|_2^2 \\ &= \arg \max_{v,w \in \mathbb{R}^n} \langle x, v \rangle + \langle x, w \rangle - \left(\frac{1}{2} \|v + w\|_2^2 + F_2^*(w) + \mu F_1(v) + \frac{\lambda\mu}{2} \|v\|_2^2 \right), \end{aligned}$$

where we first multiply the objective function by μ and then expand the quadratic term. Recall that any indicator function is invariant under multiplication with a positive constant, hence we obtain $I\{\|w\|_* \leq 1\} = \mu I\{\|w\|_* \leq 1\}$ and the second equality in eq. (50) follows. The last maximization problem in eq. (50) is in the form of Hopf formula. The corresponding multi-time HJ equation with time variables μ and $\nu = \lambda\mu$ is given by

$$(51) \quad \begin{cases} \frac{\partial}{\partial \mu} \tilde{S}(y, z, \mu, \nu) + F_1(\nabla_y \tilde{S}(y, z, \mu, \nu)) = 0, & y, z \in \mathbb{R}^n; \mu, \nu > 0; \\ \frac{\partial}{\partial \nu} \tilde{S}(y, z, \mu, \nu) + \frac{1}{2} \|\nabla_y \tilde{S}(y, z, \mu, \nu)\|_2^2 = 0, & y, z \in \mathbb{R}^n; \mu, \nu > 0; \\ \tilde{S}(y, z, 0, 0) = J(y, z), & y, z \in \mathbb{R}^n. \end{cases}$$

Here, J is the l.s.c. convex function such that $J^*(v, w) = \frac{1}{2} \|v + w\|_2^2 + F_2^*(w)$. Although the assumption (H1) is not satisfied, by eq. (50) and lemma 5.1, we know that the Hopf formula is well-defined in $\mathbb{R}^n \times \mathbb{R}^n \times [0, +\infty) \times [0, +\infty)$. Moreover, the solution \tilde{S} is the classical solution to the multi-time HJ equation eq. (51) and its spatial gradient equals $(v_{\lambda,\mu}, w_{\lambda,\mu})$. To be specific, we have

$$(52) \quad (v_{\lambda,\mu}, w_{\lambda,\mu}) = \nabla_{y,z} \tilde{S}(x, x, \mu, \lambda\mu).$$

Then, we want to apply the results in proposition 3.4 (i) to prove that any cluster point of $v_{\lambda,\mu}$ is in $U(x)$. In fact, under the basic assumptions that $H_j, J \in \Gamma_0(\mathbb{R}^n)$ and the Hopf formula is well-defined, the proof of proposition 3.4 (i) only requires the following statements:

- (a) $\partial J(x, x)$ is non-empty;
- (b) the Hamiltonians are finite-valued;
- (c) \tilde{S} is differentiable;

(d) the spatial gradient $\nabla_{y,z}\tilde{S}(x, x, \mu, \lambda\mu)$ is bounded with all limit points in $\partial J(x, x)$.

The statements (b) and (c) are obvious satisfied. It is straightforward to check $\partial J(x, x) \neq \emptyset$. Specifically, $(v, w) \in \partial J(x, x)$ iff $(x, x) \in \partial J^*(v, w)$. By simple computation, $\partial J^*(v, w) = (v + w, v + w + \partial F_2^*(w))$. Then we obtain

$$(53) \quad (v, w) \in \partial J(x, x) \text{ iff } v + w = x \text{ and } \|w\|_* \leq 1.$$

Such v and w always exist, hence $\partial J(x, x) \neq \emptyset$. As for the statement (d), the boundedness of $\nabla_{y,z}\tilde{S}(x, x, \mu, \lambda\mu)$ follows from eq. (52) and lemma 5.1. By eq. (49), $v_{\lambda,\mu} + w_{\lambda,\mu}$ converges to x . Also, $\|w_{\lambda,\mu}\|_* \leq 1$ is given by the constraint imposed by F_2^* in the minimization problem eq. (48). Together with eq. (52), we can conclude that any limit point of $\nabla_{y,z}\tilde{S}(x, x, \mu, \lambda\mu)$, denoted as (v, w) , satisfies $v + w = x$ and $\|w\|_* \leq 1$. Hence, $(v, w) \in \partial J(x, x)$ by eq. (53) and the statement (d) is proved.

Therefore, the conclusion of proposition 3.4 (i) still holds although the assumption (H1) is not satisfied. As a result, for any cluster point (\bar{v}, \bar{w}) of $(v_{\lambda,\mu}, w_{\lambda,\mu})$,

$$(\bar{v}, \bar{w}) \in \arg \max_{(v,w) \in \partial J(x,x)} -F_1(v) = \arg \min_{v+w=x, \|w\|_* \leq 1} F_1(v) = \{(v, w) : v \in U(x), w = x - v\},$$

where the last two equalities follow from eq. (53) and the definition of $U(x)$ in eq. (45). In conclusion, any cluster point \bar{v} of $v_{\lambda,\mu}$ is in $U(x)$. \square

Lemma 5.3. *For any $\lambda, \mu > 0$, the function $S_{\lambda,\mu}$ defined in eq. (47) is differentiable. Let $x \in \mathbb{R}^n$ and define $p_{\lambda,\mu} := \nabla S_{\lambda,\mu}(x)$. Then for any positive constant K , the set of gradients $\{p_{\lambda,\mu} : \lambda, \mu \in (0, K)\}$ is bounded. Moreover, as λ and μ approach zero, any cluster point of $p_{\lambda,\mu}$ is in $\partial S(x)$.*

Proof. Rewriting the formula of $S_{\lambda,\mu}$ in eq. (48), we get

$$S_{\lambda,\mu} = \left(F_1 + \frac{\lambda}{2} \|\cdot\|_2^2 \right) \square F_2^* \square \left(\frac{1}{2\mu} \|\cdot\|_2^2 \right).$$

From straightforward computation, by proposition 2.7 and the definition of $(v_{\lambda,\mu}, w_{\lambda,\mu})$ in eq. (48), we obtain

$$(54) \quad \begin{aligned} \partial S_{\lambda,\mu}(x) &= \partial \left(F_1 + \frac{\lambda}{2} \|\cdot\|_2^2 \right) (v_{\lambda,\mu}) \cap \partial F_2^*(w_{\lambda,\mu}) \cap \left\{ \frac{1}{\mu} (x - v_{\lambda,\mu} - w_{\lambda,\mu}) \right\} \\ &= (\partial F_1(v_{\lambda,\mu}) + \lambda v_{\lambda,\mu}) \cap \partial F_2^*(w_{\lambda,\mu}) \cap \left\{ \frac{1}{\mu} (x - v_{\lambda,\mu} - w_{\lambda,\mu}) \right\}. \end{aligned}$$

As a result, $\partial S_{\lambda,\mu}(x)$ contains at most one element. On the other hand, $S_{\lambda,\mu}$ is convex and finite-valued, which implies the subdifferential of $S_{\lambda,\mu}$ is non-empty. Hence, $S_{\lambda,\mu}$ is differentiable and its gradient is given by

$$(55) \quad p_{\lambda,\mu} := \nabla S_{\lambda,\mu}(x) = \frac{1}{\mu} (x - v_{\lambda,\mu} - w_{\lambda,\mu}).$$

Let K be an arbitrary positive number. Now, we prove that there exists a constant C such that $\|p_{\lambda,\mu}\|_2 \leq C$ whenever $\lambda, \mu \in (0, K)$. By eqs. (54) and (55), $p_{\lambda,\mu}$ is in the set $\partial F_1(v_{\lambda,\mu}) + \lambda v_{\lambda,\mu}$. On the one hand, the subdifferential of the norm F_1 is always bounded. In other words, there exists a constant C_1 such that $\|s\|_2 \leq C_1$ whenever $s \in \partial F_1(z)$ for some $z \in \mathbb{R}^n$. Then, we deduce that the set $\partial F_1(v_{\lambda,\mu})$ is bounded by C_1 . On the other hand, according to lemma 5.1, there exists a constant C_2 such that $\|v_{\lambda,\mu}\|_2 \leq C_2$ whenever $\lambda, \mu \in (0, K)$. Therefore, $\{p_{\lambda,\mu} : \lambda, \mu \in (0, K)\}$ is bounded by $C_1 + C_2 K$.

Let p be a cluster point of $\{p_{\lambda,\mu}\}$. By taking a subsequence we can assume λ_k and μ_k converge to zero and $p_k := p_{\lambda_k, \mu_k}$ converges to p . By lemma 5.1, $v_k := v_{\lambda_k, \mu_k}$ is bounded, hence we can assume v_k converges to a point u by taking a subsequence. Then, $w_k := w_{\lambda_k, \mu_k}$ converges to $x - u$ by lemma 5.2. From eq. (54), we have

$$p_k \in (\partial F_1(v_k) + \lambda_k v_k) \cap \partial F_2^*(w_k).$$

Since the subdifferential operators ∂F_1 and ∂F_2^* are continuous [66, Prop.XI.4.1.1], when k goes to infinity, the above inclusion becomes

$$(56) \quad p \in (\partial F_1(u) + 0 \cdot u) \cap \partial F_2^*(x - u) = \partial F_1(u) \cap \partial F_2^*(x - u).$$

On the other hand, by proposition 2.7 and the definition of $S(x)$ and $U(x)$ in eq. (45), we have

$$(57) \quad \partial S(x) = \partial F_1(\tilde{u}) \cap \partial F_2^*(x - \tilde{u}),$$

for any $\tilde{u} \in U(x)$. Moreover, by lemma 5.2, since u is a cluster point of v_k , we can conclude that $u \in U(x)$. As a result, we can choose $\tilde{u} = u$ in eq. (57) and compare it with eq. (56) to conclude that $p \in \partial S(x)$. \square

Proposition 5.1. *Assume $\{\lambda_k\} \subset (0, +\infty)$ and $\{\mu_k\} \subset (0, +\infty)$ converge to zero and $\lim_{k \rightarrow +\infty} \frac{\lambda_k}{\mu_k} = c \in (0, +\infty)$. Then, the minimizer $u_k := u_{\lambda_k, \mu_k}$ and the gradient $p_k := \nabla S_{\lambda_k, \mu_k}(x)$ converge to the l^2 projections of zero onto the sets $U(x)$ and $\partial S(x)$, respectively. To be specific,*

$$\lim_{k \rightarrow +\infty} u_k = \arg \min_{u \in U(x)} \|u\|_2, \quad \text{and} \quad \lim_{k \rightarrow +\infty} p_k = \arg \min_{p \in \partial S(x)} \|p\|_2.$$

Proof. Define $H(\cdot) := \|\cdot\|_2^2/2$. We will use the general symbol H to replace the quadratic function because this proof holds for a general finite-valued, 1-coercive, differentiable and strictly convex function H . Note that the limit of u_k is the same as the limit of v_k , hence we just need to prove the result for v_k and p_k . Denote

$$(58) \quad \bar{u} := \arg \min_{u \in U(x)} H(u), \quad \text{and} \quad \bar{p} := \arg \min_{p \in \partial S(x)} H(p).$$

Since v_k and p_k are bounded, we can assume that v_k converges to u and p_k converges to p by taking a subsequence. Then it suffices to prove $u = \bar{u}$, $p = \bar{p}$.

By eq. (54) and eq. (55), we have

$$(59) \quad p_k \in (\partial F_1(v_k) + \lambda_k \nabla H(v_k)) \cap \partial F_2^*(w_k) \cap \left\{ \nabla H^* \left(\frac{x - v_k - w_k}{\mu_k} \right) \right\}.$$

By proposition 2.5, we deduce that $w_k \in \partial F_2(p_k)$ and $x - v_k - w_k = \mu_k \nabla H(p_k)$. Together with eq. (59), we obtain

$$(60) \quad \begin{aligned} p_k - \lambda_k \nabla H(v_k) &\in \partial F_1(v_k); \\ x - \mu_k \nabla H(p_k) - v_k &= w_k \in \partial F_2(p_k). \end{aligned}$$

On the other hand, since \bar{u} and \bar{p} are the minimizer and momentum of the original problem eq. (45), we have

$$(61) \quad \bar{p} \in \partial F_1(\bar{u}) \cap \partial F_2^*(x - \bar{u}).$$

Combining eq. (60) and eq. (61), we obtain

$$\begin{cases} p_k - \lambda_k \nabla H(v_k) \in \partial F_1(v_k); & \text{and} \quad \bar{p} \in \partial F_1(\bar{u}). \\ x - \mu_k \nabla H(p_k) - v_k \in \partial F_2(p_k); & \text{and} \quad x - \bar{u} \in \partial F_2(\bar{p}). \end{cases}$$

Since the subdifferential operators ∂F_1 and ∂F_2 are monotone, by eq. (9), we obtain

$$\begin{cases} \langle p_k - \lambda_k \nabla H(v_k) - \bar{p}, v_k - \bar{u} \rangle \geq 0; \\ \langle x - \mu_k \nabla H(p_k) - v_k - (x - \bar{u}), p_k - \bar{p} \rangle \geq 0. \end{cases}$$

We sum up the two inequalities to get

$$\begin{aligned} 0 &\geq -\langle p_k - \lambda_k \nabla H(v_k) - \bar{p}, v_k - \bar{u} \rangle - \langle x - \mu_k \nabla H(p_k) - v_k - (x - \bar{u}), p_k - \bar{p} \rangle \\ &= \lambda_k \langle \nabla H(v_k), v_k - \bar{u} \rangle + \mu_k \langle \nabla H(p_k), p_k - \bar{p} \rangle. \end{aligned}$$

We divide the above inequality by μ_k and take the limit $k \rightarrow +\infty$ to obtain

$$(62) \quad 0 \geq c \langle \nabla H(u), u - \bar{u} \rangle + \langle \nabla H(p), p - \bar{p} \rangle,$$

where the positive constant c is defined in the statement of this proposition to be $c := \lim_{k \rightarrow +\infty} \lambda_k / \mu_k$. From lemma 5.2 and lemma 5.3, we know that $u \in U(x)$ and $p \in \partial S(x)$, hence we have $H(u) \geq H(\bar{u})$ and $H(p) \geq H(\bar{p})$ by eq. (58). Taken together with eq. (62), we obtain

$$(63) \quad 0 \geq c(H(\bar{u}) - H(u)) + H(\bar{p}) - H(p) \geq c \langle \nabla H(u), \bar{u} - u \rangle + \langle \nabla H(p), \bar{p} - p \rangle \geq 0.$$

As a result, the inequalities in eq. (63) become equalities, which implies $H(u) = H(\bar{u})$ and $H(p) = H(\bar{p})$ because c is positive by assumption. Therefore, we conclude that $u = \bar{u}$ and $p = \bar{p}$, since the minimizers in eq. (62) are unique. \square

In practice, if a model has non-unique minimizers, then some existing optimization algorithms may fail to converge, in which case one may consider this modification procedure and perform the optimization algorithm to the modified problem to obtain a sequence converging to the selected minimizer. Here, for simplicity, we only demonstrate the method on a specific optimization problem whose objective function contains two parts including one norm and one constraint. In fact, this method works for more general cases, such as some other decomposition models with more degenerate parts. Now, we give a numerical illustration for this proposed regularization method on the celebrated TVL1 model [5, 6, 14, 42, 43, 52, 74, 75].

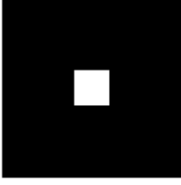
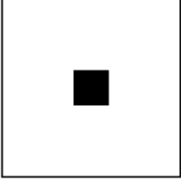
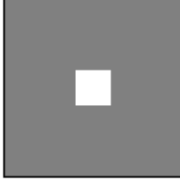
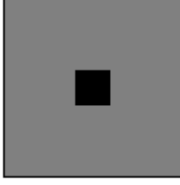

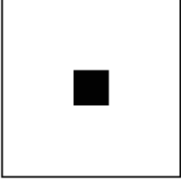

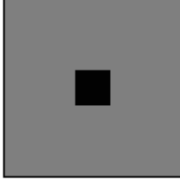
	Example 1	Example 2	Example 3	Example 4
Original Image				
v Component				

TABLE 2. Numerical results of the TVL1 model with the proposed regularization method.

To be specific, the TVL1 model solves the following optimization problem

$$(64) \quad U^{TVL1}(x) := \arg \min_{u \in \mathbb{R}^n} \alpha \|u\|_{TV} + \|x - u\|_1,$$

where $\|\cdot\|_{TV}$ denotes the discrete total variation semi-norm defined in eq. (6). However, it is well-known that this minimization problem may have non-unique minimizers [42, 50]. For instance, let Ω be the domain of an image and Ω_1 be any small rectangle in Ω such that $2|\Omega_1| < |\Omega|$. Let I be the set of indices whose corresponding pixels are in Ω_1 . Let m_1, m_2 be the numbers of pixels on the two adjacent sides of the small rectangle Ω_1 . In other words, there are $m_1 m_2$ pixels in Ω_1 and $2(m_1 + m_2)$ pixels on the boundary of Ω_1 . Let a and b be two different real numbers in $[0, 1]$ and set the discretized image x as follows

$$x_{i,j} := \begin{cases} a, & \text{if } (i,j) \in I; \\ b, & \text{if } (i,j) \notin I. \end{cases}$$

Then, the minimizers of the TVL1 model eq. (64) with $\alpha = (m_1 m_2)/(2m_1 + 2m_2)$ are not unique. Moreover, we have

$$U^{TVL1}(x) = \{\beta u_1 + (1 - \beta)u_2 : \beta \in [0, 1]\},$$

where u_1 and u_2 are defined by

$$(u_1)_{i,j} := \begin{cases} \min\{a, b\}, & \text{if } (i,j) \in I; \\ b, & \text{if } (i,j) \notin I, \end{cases} \quad \text{and} \quad (u_2)_{i,j} := \begin{cases} \max\{a, b\}, & \text{if } (i,j) \in I; \\ b, & \text{if } (i,j) \notin I. \end{cases}$$

By applying the proposed regularization method, a unique minimizer is selected in this set of minimizers. To be specific, we solve the following problem

$$(65) \quad (v_{\lambda,\mu}^{TVL1}(x), w_{\lambda,\mu}^{TVL1}(x)) := \arg \min_{v,w \in \mathbb{R}^n} \alpha \|v\|_{TV} + \|w\|_1 + \frac{\lambda}{2} \|v\|_2^2 + \frac{1}{2\mu} \|x - v - w\|_2^2.$$

Note that the above model is related to models incorporating infimal convolution of L^1 and L^2 fidelity terms, which are used for mixed Gaussian and Salt & Pepper noise image restoration, as proposed in [30, 31] for instance. Although this model is different from the example we give in eq. (45), one can adjust the arguments

to prove the same statements for this model. In other words, when the two parameters λ and μ converge to zero in a comparable rate, the v -component $v_{\lambda,\mu}^{TVL^1}(x)$ converges to the element $\bar{u}^{TVL^1}(x)$ defined by

$$\bar{u}^{TVL^1}(x) := \arg \min_{u \in U^{TVL^1}(x)} \|u\|_2 = u_1,$$

and the w -component converges to the residual $x - u_1$. Numerically, we use a splitting method and the algorithm in [39, 50, 67] to solve the minimizer in eq. (65) when $\lambda = \mu = 0.01$. We test the regularization method on the four images shown in the first row in table 2, and the corresponding v -components are shown in the second row.

6. CONCLUSION

In this paper, we provide connections between multi-time Hamilton-Jacobi equations and some optimization problems such as the decomposition models in image processing. To be specific, we show a representation formula for the minimizers u_j and clarify the connection between the minimizers u_j and the spatial gradient p of the minimal values. Moreover, we also study the variational behaviors of the momentum p and the velocities $\frac{u_j}{t_j}$. It turns out that their limits solve two optimization problems which are dual to each other. In addition, we provide a new perspective from convex analysis to prove the uniqueness of the convex solution to the multi-time Hamilton-Jacobi equation, taking advantage of the convexity assumptions to overcome the difficulty that the functions can take the value $+\infty$. At last, we demonstrate a regularization method to modify the decomposition models which have non-unique minimizers.

In this work, we consider the optimization problems which can be written in the form of Lax formula eq. (15). Hence, we assume the observed data x is the summation of different components $\{u_j\}$. We do not consider non-additive perturbation models such as [16, 55, 87]. However, our analysis actually covers a wide range of decomposition models with additive noise and the results can be easily extended to vector-valued images such as color images.

REFERENCES

- [1] R. ACAR AND C. R. VOGEL, *Analysis of bounded variation penalty methods for ill-posed problems*, Inverse Problems, 10 (1994), pp. 1217–1229.
- [2] W. ALLARD, *Total variation regularization for image denoising, I. geometric theory*, SIAM Journal on Mathematical Analysis, 39 (2008), pp. 1150–1190.
- [3] —, *Total variation regularization for image denoising, II. examples*, SIAM Journal on Imaging Sciences, 1 (2008), pp. 400–417.
- [4] —, *Total variation regularization for image denoising, III. examples.*, SIAM Journal on Imaging Sciences, 2 (2009), pp. 532–568.
- [5] S. ALLINEY, *A property of the minimum vectors of a regularizing functional defined by means of the absolute norm*, IEEE Transactions on Signal Processing, 45 (1997), pp. 913–917.
- [6] S. ALLINEY AND S. A. RUZINSKY, *An algorithm for the minimization of mixed l_1 and l_2 norms with application to bayesian estimation*, IEEE Transactions on Signal Processing, 42 (1994), pp. 618–627.
- [7] G. AUBERT AND P. KORNPROBST, *Mathematical Problems in Image Processing*, Springer-Verlag, 2002.
- [8] J. P. AUBIN AND A. CELLINA, *Differential Inclusions: Set-Valued Maps and Viability Theory*, Springer-Verlag, Berlin, Heidelberg, 1984.
- [9] J.-F. AUJOL, G. AUBERT, L. BLANC-FÉRAUD, AND A. CHAMBOLLE, *Image decomposition application to SAR images*, in Scale Space Methods in Computer Vision, L. D. Griffin and M. Lillholm, eds., Berlin, Heidelberg, 2003, Springer Berlin Heidelberg, pp. 297–312.
- [10] —, *Image decomposition into a bounded variation component and an oscillating component*, Journal of Mathematical Imaging and Vision, 22 (2005), pp. 71–88.
- [11] J.-F. AUJOL AND A. CHAMBOLLE, *Dual norms and image decomposition models*, International Journal of Computer Vision, 63 (2005), pp. 85–104.
- [12] J.-F. AUJOL AND T. F. CHAN, *Combining geometrical and textured information to perform image classification*, Journal of Visual Communication and Image Representation, 17 (2006), pp. 1004 – 1023.
- [13] J.-F. AUJOL AND G. GILBOA, *Constrained and SNR-based solutions for TV-Hilbert space image denoising*, Journal of Mathematical Imaging and Vision, 26 (2006), pp. 217–237.

- [14] J.-F. AUJOL, G. GILBOA, T. CHAN, AND S. OSHER, *Structure-texture image decomposition—modeling, algorithms, and parameter selection*, International Journal of Computer Vision, 67 (2006), pp. 111–136.
- [15] J.-F. AUJOL AND S. H. KANG, *Color image decomposition and restoration*, Journal of Visual Communication and Image Representation, 17 (2006), pp. 916 – 928.
- [16] A. AUSLENDER AND M. TEBoulLE, *Interior gradient and proximal methods for convex and conic optimization*, SIAM Journal on Optimization, 16 (2006), pp. 697–725.
- [17] M. BARDI AND I. CAPUZZO-DOLCETTA, *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*, Birkhäuser Basel, 1997.
- [18] M. BARDI AND L. EVANS, *On Hopf’s formulas for solutions of Hamilton-Jacobi equations*, Nonlinear Analysis: Theory, Methods & Applications, 8 (1984), pp. 1373 – 1381.
- [19] G. BARLES, *Solutions de viscosité des équations de Hamilton-Jacobi*, Mathématiques et Applications, Springer-Verlag Berlin Heidelberg, 1994.
- [20] G. BARLES AND A. TOURIN, *Commutation properties of semigroups for first-order Hamilton-Jacobi equations and application to multi-time equations*, Indiana University Mathematics Journal, 50 (2001), pp. 1523–1544.
- [21] J.-L. BASDEVANT, *Variational principles in physics*, Springer, New York, 2007.
- [22] H. H. BAUSCHKE AND P. L. COMBETTES, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer Publishing Company, Incorporated, 2011.
- [23] M. BERTALMIO, L. VESE, G. SAPIRO, AND S. OSHER, *Simultaneous structure and texture image inpainting*, IEEE Transactions on Image Processing, 12 (2003), pp. 882–889.
- [24] J. BORWEIN AND A. S. LEWIS, *Convex Analysis and Nonlinear Optimization: Theory and Examples*, Springer-Verlag New York, 2006.
- [25] K. BREDIES, K. KUNISCH, AND T. POCK, *Total generalized variation*, SIAM Journal on Imaging Sciences, 3 (2010), pp. 492–526.
- [26] A. BRESSAN AND B. PICCOLI, *Introduction to the mathematical theory of control*, vol. 2, American Institute of Mathematical Sciences (AIMS), Springfield, MO, 2007.
- [27] H. BREZIS, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland Mathematics Studies, Notas de Matemática, North-Holland Publishing Company, 1973.
- [28] H. BREZIS, *Remarks on some minimization problems associated with bv norms*, Discrete & Continuous Dynamical Systems - A, 0 (2019), p. 1.
- [29] M. BURGER AND S. OSHER, *Convergence rates of convex variational regularization*, Inverse Problems, 20 (2004), pp. 1411–1421.
- [30] L. CALATRONI, J. C. DE LOS REYES, AND C.-B. SCHNLIEB, *Infimal convolution of data discrepancies for mixed noise removal*, SIAM Journal on Imaging Sciences, 10 (2017), pp. 1196–1233.
- [31] L. CALATRONI AND K. PAPAFITSOROS, *Analysis and automatic parameter selection of a variational model for mixed Gaussian and salt-and-pepper noise removal*, Inverse Problems, 35 (2019), p. 114001.
- [32] F. CARDIN AND C. VITERBO, *Commuting Hamiltonians and Hamilton-Jacobi multi-time equations*, Duke Math. J., 144 (2008), pp. 235–284.
- [33] E. CASAS, K. KUNISCH, AND C. POLA, *Regularization by functions of bounded variation and applications to image enhancement*, Applied Mathematics and Optimization, 40 (1999), pp. 229–257.
- [34] V. CASELLES, A. CHAMBOLLE, AND M. NOVAGA, *The discontinuity set of solutions of the TV denoising problem and some extensions*, Multiscale Modeling & Simulation, 6 (2007), pp. 879–894.
- [35] V. CASELLES, A. CHAMBOLLE, AND M. NOVAGA, *Regularity for solutions of the total variation denoising problem*, Rev. Mat. Iberoam., 27 (2011), pp. 233–252.
- [36] V. CASELLES, A. CHAMBOLLE, AND M. NOVAGA, *Total variation in imaging*, in Handbook of mathematical methods in imaging. Vol. 1, 2, 3, Springer, New York, 2015, pp. 1455–1499.
- [37] A. CHAMBOLLE, *An algorithm for total variation minimization and applications*, Journal of Mathematical Imaging and Vision, 20 (2004), pp. 89–97.
- [38] A. CHAMBOLLE, V. CASELLES, D. CREMERS, M. NOVAGA, AND T. POCK, *An introduction to total variation for image analysis*, in Theoretical foundations and numerical methods for sparse recovery, vol. 9 of Radon Ser. Comput. Appl. Math., Walter de Gruyter, Berlin, 2010, pp. 263–340.
- [39] A. CHAMBOLLE AND J. DARBON, *On total variation minimization and surface evolution using parametric maximum flows*, International Journal of Computer Vision, 84 (2009), pp. 288–307.

- [40] A. CHAMBOLLE, V. DUVAL, G. PEYRÉ, AND C. POON, *Total variation denoising and support localization of the gradient*, Journal of Physics: Conference Series, 756 (2016), p. 012007.
- [41] A. CHAMBOLLE AND P.-L. LIONS, *Image recovery via total variation minimization and related problems*, Numerische Mathematik, 76 (1997), pp. 167–188.
- [42] T. CHAN AND S. ESEDOGLU, *Aspects of total variation regularized L1 function approximation*, SIAM Journal on Applied Mathematics, 65 (2005), pp. 1817–1837.
- [43] T. CHAN, S. ESEDOGLU, AND M. NIKOLOVA, *Algorithms for finding global minimizers of image segmentation and denoising models*, SIAM Journal on Applied Mathematics, 66 (2006), pp. 1632–1648.
- [44] T. CHAN AND J. SHEN, *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods*, Other Titles in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104), 2005.
- [45] T. F. CHAN, S. ESEDOGLU, AND F. E. PARK, *Image decomposition combining staircase reduction and texture extraction*, Journal of Visual Communication and Image Representation, 18 (2007), pp. 464 – 486.
- [46] G. CHAVENT AND K. KUNISCH, *Regularization of linear least squares problems by total bounded variation*, ESAIM: Control, Optimisation and Calculus of Variations, 2 (1997), pp. 359–376.
- [47] R. CHOKSI, I. FONSECA, AND B. ZWICKNAGL, *A few remarks on variational models for denoising*, Communications in Mathematical Sciences, 12 (2014), pp. 843–857.
- [48] M. G. CRANDALL, H. ISHII, AND P.-L. LIONS, *User’s guide to viscosity solutions of second order partial differential equations*, Bulletin of the American mathematical society, 27 (1992), pp. 1–67.
- [49] J. DARBON, *On convex finite-dimensional variational methods in imaging sciences and Hamilton–Jacobi equations*, SIAM Journal on Imaging Sciences, 8 (2015), pp. 2268–2293.
- [50] 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 (2006), pp. 261–276.
- [51] D. DOBSON AND O. SCHERZER, *Analysis of regularized total variation penalty methods for denoising*, Inverse Problems, 12 (1996), pp. 601–617.
- [52] V. DUVAL, J. AUJOL, AND Y. GOUSSEAU, *The TVL1 model: A geometric point of view*, Multiscale Modeling & Simulation, 8 (2009), pp. 154–189.
- [53] V. DUVAL, J.-F. AUJOL, AND L. A. VESE, *Projected gradient based color image decomposition*, in Scale Space and Variational Methods in Computer Vision, Berlin, Heidelberg, 2009, Springer Berlin Heidelberg, pp. 295–306.
- [54] —, *Mathematical modeling of textures: Application to color image decomposition with a projected gradient algorithm*, Journal of Mathematical Imaging and Vision, 37 (2010), pp. 232–248.
- [55] J. ECKSTEIN, *Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming*, Mathematics of Operations Research, 18 (1993), pp. 202–226.
- [56] M. ELAD, J.-L. STARCK, P. QUERRE, AND D. DONOHO, *Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)*, Applied and Computational Harmonic Analysis, 19 (2005), pp. 340 – 358. Computational Harmonic Analysis - Part 1.
- [57] L. C. EVANS AND P. E. SOUGANIDIS, *Differential games and representation formulas for solutions of Hamilton-Jacobi-Isaacs equations*, tech. rep., Mathematics Research Center, University of Wisconsin-Madison, 1983.
- [58] M. J. FADILI, J. STARCK, J. BOBIN, AND Y. MOUDDEN, *Image decomposition and separation using sparse representations: An overview*, Proceedings of the IEEE, 98 (2010), pp. 983–994.
- [59] J. B. GARNETT, T. M. LE, Y. MEYER, AND L. A. VESE, *Image decompositions using bounded variation and generalized homogeneous Besov spaces*, Applied and Computational Harmonic Analysis, 23 (2007), pp. 25 – 56. Special Issue on Mathematical Imaging.
- [60] J. GILLES, *Noisy image decomposition: A new structure, texture and noise model based on local adaptivity*, Journal of Mathematical Imaging and Vision, 28 (2007), pp. 285–295.
- [61] —, *Image decomposition: theory, numerical schemes, and performance evaluation*, Advances in Imaging and Electron Physics, 158 (2009), pp. 89–137.
- [62] J. GILLES AND Y. MEYER, *Properties of BV-G structures+textures decomposition models. application to road detection in satellite images*, IEEE Transactions on Image Processing, 19 (2010), pp. 2793–2800.
- [63] M. HINTERMÜLLER, M. HOLLER, AND K. PAPAITSOROS, *A function space framework for structural total variation regularization with applications in inverse problems*, Inverse Problems, 34 (2018),

- p. 064002.
- [64] M. HINTERMÜLLER, K. PAPAFITSOROS, AND C. N. RAUTENBERG, *Analytical aspects of spatially adapted total variation regularisation*, Journal of Mathematical Analysis and Applications, 454 (2017), pp. 891 – 935.
 - [65] J.-B. HIRIART-URRUTY AND C. LEMARECHAL, *Convex Analysis and Minimization Algorithms I: Fundamentals*, vol. 305, Springer-Verlag Berlin Heidelberg, 1993.
 - [66] ———, *Convex Analysis and Minimization Algorithms II: Advanced Theory and Bundle Methods*, vol. 306, Springer-Verlag Berlin Heidelberg, 1993.
 - [67] D. S. HOCHBAUM, *An efficient algorithm for image segmentation, Markov random fields and related problems*, Journal of the ACM, 48 (2001), pp. 686–701.
 - [68] E. HOPF, *Generalized solutions of non-linear equations of first order*, Journal of Mathematics and Mechanics, 14 (1965), pp. 951–973.
 - [69] C. IMBERT, *Convex analysis techniques for Hopf-Lax formulae in Hamilton-Jacobi equations*, Journal of Nonlinear and Convex Analysis. An International Journal, 2 (2001), pp. 333–343.
 - [70] T. LE AND L. VESE, *Image decomposition using total variation and $\text{div}(BMO)$* , Multiscale Modeling & Simulation, 4 (2005), pp. 390–423.
 - [71] P. L. LIONS AND J.-C. ROCHET, *Hopf formula and multitime Hamilton-Jacobi equations*, Proceedings of the American Mathematical Society, 96 (1986), pp. 79–84.
 - [72] Y. MEYER, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations: The Fifteenth Dean Jacqueline B. Lewis Memorial Lectures*, American Mathematical Society, Boston, MA, USA, 2001.
 - [73] M. MOTTA AND F. RAMPAZZO, *Nonsmooth multi-time Hamilton-Jacobi systems*, Indiana University Mathematics Journal, 55 (2006), pp. 1573–1614.
 - [74] M. NIKOLOVA, *Minimizers of cost-functions involving nonsmooth data-fidelity terms. application to the processing of outliers*, SIAM Journal on Numerical Analysis, 40 (2002), pp. 965–994.
 - [75] M. NIKOLOVA, *A variational approach to remove outliers and impulse noise*, Journal of Mathematical Imaging and Vision, 20 (2004), pp. 99–120.
 - [76] ———, *Weakly constrained minimization: Application to the estimation of images and signals involving constant regions*, Journal of Mathematical Imaging and Vision, 21 (2004), pp. 155–175.
 - [77] S. OSHER, A. SOL, AND L. VESE, *Image decomposition and restoration using total variation minimization and the H^{-1} norm*, Multiscale Modeling & Simulation, 1 (2003), pp. 349–370.
 - [78] S. PLASKACZ AND M. QUINCAMPOIX, *Oleinik-Lax formulas and multitime Hamilton-Jacobi systems*, Nonlinear Analysis: Theory, Methods & Applications, 51 (2002), pp. 957–967.
 - [79] W. RING, *Structural properties of solutions to total variation regularization problems*, ESAIM: Mathematical Modelling and Numerical Analysis, 34 (2000), pp. 799–810.
 - [80] J. ROCHET, *The taxation principle and multi-time Hamilton-Jacobi equations*, Journal of Mathematical Economics, 14 (1985), pp. 113 – 128.
 - [81] R. T. ROCKAFELLAR, *Convex Analysis*, Princeton University Press, 1970.
 - [82] R. T. ROCKAFELLAR AND R. J.-B. WETS, *Variational Analysis*, vol. 317 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, 1998.
 - [83] L. I. RUDIN, S. OSHER, AND E. FATEMI, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.
 - [84] P. E. SOUGANIDIS, *Max-min representations and product formulas for the viscosity solutions of Hamilton-Jacobi equations with applications to differential games*, Nonlinear Analysis: Theory, Methods & Applications, 9 (1985), pp. 217 – 257.
 - [85] J.-L. STARCK, M. ELAD, AND D. DONOHO, *Redundant multiscale transforms and their application for morphological component separation*, Advances in Imaging and Electron Physics - ADV IMAG ELECTRON PHYS, 132 (2004), pp. 287–348.
 - [86] J. L. STARCK, M. ELAD, AND D. L. DONOHO, *Image decomposition via the combination of sparse representations and a variational approach*, IEEE Transactions on Image Processing, 14 (2005), pp. 1570–1582.
 - [87] M. TEBoulLE, *Convergence of proximal-like algorithms*, SIAM Journal on Optimization, 7 (1997), pp. 1069–1083.

- [88] N. THO, *Hopf-Lax-Oleinik type formula for multi-time Hamilton-Jacobi equations*, Acta Math. Vietnamica, 30 (2005), pp. 275–287.
- [89] T. VALKONEN, *The jump set under geometric regularization. part 1: Basic technique and first-order denoising*, SIAM Journal on Mathematical Analysis, 47 (2015), pp. 2587–2629.
- [90] L. A. VESE AND S. J. OSHER, *Image denoising and decomposition with total variation minimization and oscillatory functions*, Journal of Mathematical Imaging and Vision, 20 (2004), pp. 7–18.
- [91] C. VOGEL, *Computational Methods for Inverse Problems*, Society for Industrial and Applied Mathematics, 2002.

DEPARTMENT OF APPLIED MATHEMATICS, BROWN UNIVERSITY, PROVIDENCE, RI
E-mail address: jerome_darbon@brown.edu

DEPARTMENT OF APPLIED MATHEMATICS, BROWN UNIVERSITY, PROVIDENCE, RI
E-mail address: tingwei_meng@brown.edu