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

A PREPRINT

Jérôme Darbon† Tingwei Meng‡

June 25, 2019

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 PDE. 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 momentum $\nabla_x S(x, t)$, the solution is given by the Hopf formula or Lax formula [1][2] under some regularity and convexity assumptions. We refer the readers to the review paper [3] for thorough details and [4][5] 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 [6]. Later, Lions and Rochet generalized the Hopf formula and Lax formula to the multi-time case by proving the commutation property of semigroups [7]. Following their work, several existence and uniqueness results [8][9][10][11] 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 [12] and differential games [13][14]. However, there is a gap between the theory of HJ equation and the field of image processing. To fill this gap, Darbon [4] 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. 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.

†This work was funded by NSF 1820821
‡Department of Applied Mathematics, Brown University, Providence, RI (tingwei_meng@brown.edu).
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 \[15, 16\], image classification \[17\], and road detection \[18\]. Here, we give a brief overview of convex variational models in this area. There are many models cannot be fully listed here, for which we refer the readers to \[19, 20\].

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:

\[
\arg \min_{u_0 + \cdots + u_N = x} f_0(u_0) + \sum_{j=1}^{N} \lambda_j f_j(u_j). \tag{1}
\]

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. The first widely used decomposition model is the Rudin-Osher-Fatemi (ROF) model, proposed in \[21\], which applies the total variation (TV) semi-norm and \(\|\cdot\|_2^2\) to recognize the geometry and noise in an image, respectively. The mathematical analysis for the ROF model is provided in \[22, 23\]. Later, Meyer \[24\] pointed out the disadvantage of \(\|\cdot\|_2^2\) 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: these three spaces are defined as follows:

\[
E := B^{-1, \infty}_1, \quad F := \text{div}(BMO), \quad G := \text{div}(L^\infty)
\]

where \(BMO\) denotes the space of functions of bounded mean oscillation \[24\]. To be specific,

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

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

The space \(F\) is similarly defined by replacing the space \(L^\infty\) in the above definition with \(BMO\). For mathematical analysis of these models, we refer the readers to \[25, 18, 26\]. In \[25\], the space \(E\) is also generalized to any homogeneous Besov space \(B^{p-2, q}_p\), where \(p, q \in [1, \infty]\) and \(\alpha \in (0, 2)\). (Here, we adopt the notation of Besov space in Meyer’s book \[24\].) 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 \(L^\infty\) in the definition of \(G\) by \(L^p\) \[27\]. Osher et al \[28\] proposed an equivalent formulation called OSV when \(p = 2\). In a word, OSV uses the square of \(H^{-1}\)-norm instead of \(G\)-norm. The other approach called \(A^2 BC\) model is proposed by Aujol et al \[29, 30\], replacing the \(G\)-norm with the indicator function of balls in the space \(G\). It is shown that \(A^2 BC\) model gives the solution to Meyer’s model when the radius of the ball in the indicator function is chosen appropriately. This model is easy to solve using Chambolle’s projection method \[31\]. Similarly, in \[32\], the indicator function of \(E\)-ball is used to replace the \(E\)-norm.

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 \[32\], which constrains \(G\)-norm of the texture part and \(E\)-norm of the noisy part. Later, Gilles \[33\] 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 \(A^2 BC\) model by requiring the \(G\)-norm of the noise to be much smaller than the \(G\)-norm of the texture. In \[34, 35, 36\], 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 \[37, 38, 39, 40\] is used to promote sparsity or remove salt and pepper noise. In \[41, 58\], 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 \[4\] 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. 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 \[42, 43, 44\] 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 \(\alpha \| \Delta v \|_2^2\) to the models introduced above, in order to reduce the staircase effect. Therefore, it is valuable to generalize the previous work \[4\] and provide a framework to analyze the models involving more than two components.

Now, we briefly introduce the intuition and the basic setup for our framework and demonstrate the idea using some experimental results of the \(A^2 BC\) 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 $A^2BC$ model solves the following optimization problem:

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

The desired quantities are the minimizers, denoted as $u(x, \mu, \lambda)$ and $v(x, \mu, \lambda)$. Here, $J$ is the total variation semi-norm in $BV(\Omega)$ and its Legendre transform $J^*$ is the indicator function of the unit ball of Meyer’s norm, hence the above optimization problem is equivalent to

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

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

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

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 (2) to two subproblems involving computing the proximal point of $\lambda J$ and computing the projection to a $\mu$–ball of Meyer’s norm. The second subproblem
Figure 2: The graphs of the minimal values $S$ with respect to the variables $\alpha$, $\mu$ and $\lambda$ in the first example are shown in (a)-(c), respectively.

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.
The paper is organized as follows. Section 2 gives a brief review of the convex optimization theorems which are applicable to the multi-time HJ equation. We provide a representation formula for the minimizers \( u_j \) using the spatial gradient of the solution \( S \). Moreover, the variational behaviors of the momentum \( \nabla_x S \) and the velocities \( u_j/t_j \) are investigated, which corresponds to two optimization problems in a duality relation to each other. Also, the optimal values of these two optimization problems give the variational behavior of the solution \( S \). Besides, we also provide a new perspective from convex analysis to prove the uniqueness of the convex solution to multi-time HJ equation. At last, we propose a regularization method for the decomposition problems which may have non-unique minimizers or non-differentiable minimal values. To be specific, the model is modified using some regularization terms so that both the minimizer and the gradient of the minimal values can be regarded as a function of \( \mu \), whose graph is plotted in fig. 2b. Similarly, the graph of \( S(x_1, \mu_1, \lambda) \) is plotted in fig. 2a. To illustrate the variation of \( S \) with respect to \( x \), we choose another image \( x_2 \) with corresponding suitable parameters \( \mu_2, \lambda_2 \), and plot the function values \( f : x \mapsto S(\alpha x_1 + (1-\alpha)x_2, \alpha \mu_1 + (1-\alpha)\mu_2, \alpha \lambda_1 + (1-\alpha)\lambda_2) \) 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. 4d to 4f. 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. 5b and 5c.

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. Set \( \mu_1 = 16, \mu_2 = 24, \lambda_1 = 8, \) 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 \( x, \mu, \lambda \) 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. In this paper, more properties about \( S \) and the minimizers \( u, v \) are revealed.

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. To be specific, for some optimization problems, the minimal value coincides with the solution \( S(x, t_1, \cdots, t_N) \) to a corresponding multi-time HJ equation. We provide a representation formula for the minimizers \( u_j \) using the spatial gradient of the solution \( S \). Moreover, the variational behaviors of the momentum \( \nabla_x S \) and the velocities \( u_j/t_j \) are investigated, which corresponds to two optimization problems in a duality relation to each other. Also, the optimal values of these two optimization problems give the variational behavior of the solution \( S \). Besides, we also provide a new perspective from convex analysis to prove the uniqueness of the convex solution to multi-time HJ equation. At last, we propose a regularization method for the decomposition problems which may have non-unique minimizers or non-differentiable minimal values. To be specific, the model is modified using some regularization terms so that both the minimizer and the gradient of the minimal values are unique and converge to the \( L^2 \)-projection of zero onto the corresponding sets of the original problems, when the regularization parameters approach zero in a comparable rate.

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 \( u_j/t_j \) in proposition 3.4. Section 4 is devoted to the proof of the uniqueness...
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).
Figure 6: The graphs of the minimal values $S$ with respect to the variables $\alpha$, $\mu$ and $\lambda$ in the second example are plotted in (a)-(c), respectively.

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 [49, 50]. We also refer the readers to [51, 52, 53].

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 ri $C$, is the interior of $C$ with respect to the minimal hyperplane containing $C$ in $\mathbb{R}^n$. For any convex set $C$, the normal cone of $C$ at $x \in C$, denoted by $N_C(x)$, can be characterized by

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

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

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

In fact, the asymptotic cone is independent of $x$, as stated in the following result.
Proposition 2.1 \[49\] Proposition 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 \( f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} \) 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.

Proposition 2.2 \[49\] Lemma IV 3.1.1 and Chapter 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 \( y \in \text{ri dom} f \),

\[
f(x) = \lim_{t \to 0^+} f(x + t(y - x)).
\]

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 \( \mathbb{R} \cup \{+\infty\} \). When \( f \) is differentiable at \( x \), \( f'(x, \cdot) = (\nabla f(x), \cdot) \) 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,

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

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

Proposition 2.3 \[50\] Corollary 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. Then one can compute the subdifferential of the indicator function and obtain

\[
\partial I_C(x) = N_C(x).
\]

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

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

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 \( \lim_{\|x\| \to \infty} g(x)/\|x\| = +\infty \).)
Proposition 2.4 [50] Chapter 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. (7).

Proposition 2.5 [50] Corollary 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 [50] 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

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

(8)

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 [50] Theorem X 2.3.2 and Theorem XI 3.4.1] Let \( f, g \in \Gamma_0(\mathbb{R}^n) \). Assume \( \text{ri dom} f^* \cap \text{ri dom} g^* \neq \emptyset \). 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. (8) has at least one minimizer, and \( \partial(f \square g)(x) = \partial f(u) \cap \partial g(y - u) \) for any minimizer \( u \).

### 3 Properties of the Solutions to the Multi-time Hamilton-Jacobi (HJ) 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 fig. (7)

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

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

(9)

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

(\text{H1}) \quad H_j : \mathbb{R}^n \rightarrow \mathbb{R}, \text{ is convex and } 1\text{-coercive for any } j = 1, \cdots, N. \text{ Moreover, at least one of them is strictly convex;}

(\text{H2}) \quad J \in \Gamma_0(\mathbb{R}^n).

From the assumption (H1), by proposition 2.3 it is known that \( H_j^* \) is also finite-valued, convex and 1-coercive for any \( j = 1, \cdots, N \). 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:

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

(10)
and the Lax formula:
\[ S_L(x, t_1, \cdots, t_N) := \left( J \square (t_1 H_1)^* \cdots \square (t_N H_N)^* \right) (x) \]
\[ = \inf_{u_1, \cdots, u_N \in \mathbb{R}^n \atop u_j = 0 \text{ whenever } t_j = 0} \left( J \left( x - \sum_{j=1}^N u_j \right) + \sum_{j=1}^N t_j H_j^* \left( \frac{u_j}{t_j} \right) \right), \tag{11} \]

for any \( x \in \mathbb{R}^n \) and \( t_1, \cdots, t_N \geq 0 \). 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.

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. (11) 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. (9). Moreover, the spatial gradient is the unique maximizer in the Hopf formula eq. (10). 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, \( S_H \in \Gamma_0(\mathbb{R}^{n+N}) \) with \( \text{dom } S_H = (\mathbb{R}^n \times [0, +\infty)^N) \setminus ((\text{dom } J)^c \times \{0\}^N) \). Moreover, for any \( p \in \mathbb{R}^n \) and \( E^- = (E_1^-, \cdots, E_N^-) \in \mathbb{R}^N \),
\[ S_H^*(p, E^-) = J^*(p) + \sum_{j=1}^N I\{E_j^- + H_j(p) \leq 0\}. \]

Here, \( I\{\cdot\} \) denotes the indicator function.

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 \( \{ (x, t_2, \cdots, t_N) : t_j = 0 \ \forall j \in J \} \) for any index set \( J \subset \{1, \cdots, N\} \) 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, which suggests obtaining the minimizers \( u_j \) by solving its dual optimization problem in the Hopf formula when it is difficult to directly solve the original problem in the Lax formula.

**Proposition 3.2** Suppose the assumptions (H1)-(H2) hold. Let \( x \in \mathbb{R}^n, t_1, \cdots, t_N \geq 0 \) and assume the time variables \( \{t_j\} \) are not all zero. Denote \( (u_1, \cdots, u_N) \) to be any minimizer of the minimization problem in eq. (11) with parameters \( x \) and \( t_1, \cdots, t_N \). Here, each \( u_j \) can be regarded as a function of \( (x, t_1, \cdots, t_N) \). Then, for any \( j \),
\[ u_j(x, t_1, \cdots, t_N) = t_j \nabla H_j (\nabla_x S_L(x, t_1, \cdots, t_N)). \tag{12} \]

Specifically, if a stronger assumption is imposed, say, all the Hamiltonians are differentiable, then the minimizer \( u_1, \cdots, u_N \) is unique and satisfies
\[ u_j(x, t_1, \cdots, t_N) = t_j \nabla H_j (\nabla_x S_L(x, t_1, \cdots, t_N)) \]
for any \( j \).

In the remaining part of this section, we investigate the multi-time HJ equation eq. (9) and the minimization problem eq. (11) in a variational point of view. To be specific, let \( x_k = x + \sum_{j=1}^N t_{j,k} v_{j,k} \) such that \( \lim_{k \to \infty} t_{j,k} = 0 \) and \( \lim_{k \to \infty} v_{j,k} = v_{j,\infty} \) for any \( j \). We are interested in the convergence behavior of the momentum \( \nabla_x S_H \) and the minimizers \( u_j \) evaluated at \( (x_k, t_{1,k}, \cdots, t_{N,k}) \). This result can be used in the convergence analysis for the penalty method and other optimization problems. We demonstrate one application in section 5.
\( \{t_{1,k}\}_k \) converges the slowest. Then, by taking a subsequence again, we assume that the sequence of the time ratios \( \left\{ \frac{t_{j,k}}{t_{1,k}} \right\}_k \) has a limit for any \( j \), denoted as \( \alpha_{j,\infty} \in \mathbb{R} \). In summary, the following notations and assumptions are adopted:

\[
\begin{align*}
\begin{cases}
x_k = x + \sum_{j=1}^{N} t_{j,k}v_{j,k}, \text{ where } t_{j,k}, v_{j,k} \in \mathbb{R}^{n} \text{ for any } j \in \{1, \cdots, N\} \text{ and } k \in \mathbb{N}; \\
\lim_{k \to \infty} t_{j,k} = 0 \quad \text{and} \quad \lim_{k \to \infty} v_{j,k} = v_{j,\infty}; \\
\lim_{k \to \infty} \frac{t_{j,k}}{t_{1,k}} = \alpha_{j,\infty} \in \mathbb{R}.
\end{cases}
\end{align*}
\]

(13)

First, we show the convergence of \( u_j \) to zero. Also, the boundedness of the velocities \( u_j/t_j \) and the momentum \( \nabla_x S_H \) is shown in the following two results.

**Proposition 3.3** Assume (H1)-(H2) and eq. [13] hold. Let \( (u_1, \cdots, u_N) \) be any minimizer of the minimization problem in eq. (11). Let \( x \in \text{dom } J \). Then,

(i). For any \( j = 1, \cdots, N \),

\[
\lim_{k \to \infty} u_j(x_k, t_{1,k}, \cdots, t_{N,k}) = 0.
\]

(14)

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

\[
\lim_{k \to \infty} \frac{1}{t_{1,k}} u_j(x_k, t_{1,k}, \cdots, t_{N,k}) = 0.
\]

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

**Lemma 3.1** Under the assumptions (H1)-(H2) and eq. [13], for any \( x \in \text{dom } J \) such that \( \partial J(x) \neq \emptyset \), the sequence \( \{ \nabla_x S_H(x_k, t_{1,k}, \cdots, t_{N,k}) \}_k \) is bounded and any cluster point \( p \) is in \( \partial J(x) \).

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. [13] 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:

\[
\lim_{k \to \infty} \frac{S_H(x_k, t_{1,k}, \cdots, t_{N,k}) - S_H(x, 0, \cdots, 0)}{t_{1,k}} = \max_{q \in \partial J(x)} \sum_{j=1}^{N} \alpha_{j,\infty} \left( \langle q, v_{j,\infty} \rangle - H_j(q) \right).
\]

(15)

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

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

(16)

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

\[
\lim_{k \to \infty} \frac{S_L(x_k, t_{1,k}, \cdots, t_{N,k}) - S_L(x, 0, \cdots, 0)}{t_{1,k}} = \min_{w_j \in \mathbb{R}^{n}} \sum_{j=1}^{N} \alpha_{j,\infty} (I_{\partial J(x)}^*(v_{j,\infty} - w_j) + H_j^*(w_j)).
\]

(17)

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

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

(18)

Specially, if \( H_j \) is strictly convex and \( \alpha_{j,\infty} \neq 0 \) for some \( j \), then the maximizer in eq. (16) is unique, which implies the convergence of \( \nabla_x S_H(x_k, t_{1,k}, \cdots, t_{N,k}) \) to the unique maximizer. Similarly, for any \( j \) such that \( H_j \) is differentiable and \( \alpha_{j,\infty} \neq 0 \), we can conclude that \( u_j/(t_{j,k}) \) converges to the unique minimizer in eq. (18).
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 \( u_j(x_k, t_{1,k}, \cdots, t_{N,k}) \).
3.1 Proof of proposition 3.1

In this subsection, we prove proposition 3.1 by showing $S_H$ is the Legendre transform of $F$, where $F$ is defined as

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

for any $p \in \mathbb{R}^n$ and any $E^- = (E^-_1, \ldots, E^-_N) \in \mathbb{R}^N$. It is easy to check $F \in \Gamma_0(\mathbb{R}^{n+N})$.

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

$$F^*(x, t) = \sup_{p \in \mathbb{R}^n, E^- \in \mathbb{R}^N} \left( \langle x, p \rangle + \sum_{j=1}^N t_j E^-_j - J^*(p) - \sum_{j=1}^N I\{E^-_j + H_j(p) \leq 0\} \right). \tag{19}$$

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,

$$F^*(x, t) \geq \langle x, p \rangle + \sum_{j \neq k} t_j E^-_j - J^*(p) + \limsup_{E^-_k \to -\infty} t_k E^-_k = +\infty.$$ 

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

Then, consider the case when $t_1, \ldots, t_N \geq 0$. Let $x \in \mathbb{R}^n$, from eq. (19), we obtain

$$F^*(x, t) = \sup_{p \in \mathbb{R}^n} \left( \langle x, p \rangle + \sum_{j: t_j > 0} t_j E^-_j - J^*(p) \right)$$

$$= \sup_{p \in \mathbb{R}^n} \left[ \sup_{E^-_j \leq -H_j(p) \forall j} \left( \langle x, p \rangle + \sum_{j: t_j > 0} t_j E^-_j - J^*(p) \right) \right] \tag{20}$$

$$= \sup_{p \in \mathbb{R}^n} \left( \langle x, p \rangle - \sum_{j: t_j > 0} t_j H_j(p) - J^*(p) \right) = S_H(x, t).$$

Therefore, $S_H = F^*$, which implies $S_H$ is a convex lower semi-continuous function and $F = S_H^*$. Moreover, by proposition 2.4 if there exists some $k$ such that $t_k > 0$ and $t_j \geq 0$ for any $j \neq k$, then $J^* + \sum t_j H_j$ is 1-coercive, which implies its Legendre transform (with respect to $x$), $S_H(\cdot, t)$, is finite-valued. Hence dom $S_H = (\mathbb{R}^n \times [0, +\infty)^N) \setminus ((\text{dom} J)^c \times \{0\}^N)$.

3.2 Proof of proposition 3.2

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

$$\partial_x S_L(x, t_1, \ldots, t_N) = \partial J\left( x - \sum_{j=1}^N u_j \right) \bigcap \left( \bigcap_{j=1}^N \partial \left( t_j H^*_j \left( \frac{u_j}{t_j} \right) \right) \right)(u_j)$$

$$= \partial J\left( x - \sum_{j=1}^N u_j \right) \bigcap \left( \bigcap_{j=1}^N \partial H^*_j \left( \frac{u_j}{t_j} \right) \right). \tag{21}$$

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 $S_L(\cdot, t_1, \ldots, t_N)$, 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 $t_1, \ldots, t_N \geq 0$ which are not all zero and $x \in \mathbb{R}^n$. Moreover, by eq. (21), $\nabla_x S_L \in \partial H^*_j(u_j/t_j)$, which implies $u_j/t_j \in \partial H_j(\nabla_x S_L(x, t_1, \ldots, t_N))$ for any $j$. 

13
3.3 Proof of proposition 3.3

Denote \( \bar{u}_{j,k} := u_j(x_k, t_{1,k}, \ldots, t_{N,k}) \) for any \( j = 1, \ldots, N \), and \( \bar{u}_{0,k} := x_k - \sum_{j=1}^{N} \bar{u}_{j,k} \). Define \( I := \{ j : \| \bar{u}_{j,k} \|/t_{j,k} \text{ is not bounded} \} \).

Proof of (i): By Lax formula eq. (11),

\[
J(\bar{u}_{0,k}) + \sum_{j=1}^{N} t_{j,k} H_j^* \left( \frac{\bar{u}_{j,k}}{t_{j,k}} \right) \leq J \left( x_k - \sum_{j=1}^{N} t_{j,k} v_{j,k} \right) + \sum_{j=1}^{N} t_{j,k} H_j^* (v_{j,k})
\]

\[
= J(x) + \sum_{j=1}^{N} t_{j,k} H_j^* (v_{j,k}).
\]

(22)

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

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

(23)

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

\[
\sum_{j=1}^{N} t_{j,k} H_j^* \left( \frac{\bar{u}_{j,k}}{t_{j,k}} \right) \leq J(x) - J(z) + \|q\| \|x_k - z\| + \sum_{j=1}^{N} t_{j,k} H_j^* (v_{j,k}) + \sum_{j \not\in I} \|q\| \|\bar{u}_{j,k}\| + \|q\| \|x_k - z\|.
\]

(24)

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

\[
\sum_{j \in I} (M - \|q\|) \|\bar{u}_{j,k}\| \leq \sum_{j \in I} \left( t_{j,k} H_j^* \left( \frac{\bar{u}_{j,k}}{t_{j,k}} \right) - \|q\| \|\bar{u}_{j,k}\| \right)
\]

\[
\leq J(x) - J(z) + \|q\| \|x_k - z\| + \sum_{j=1}^{N} t_{j,k} H_j^* (v_{j,k}) + \sum_{j \not\in I} \|q\| \|\bar{u}_{j,k}\| - t_{j,k} H_j^* \left( \frac{\bar{u}_{j,k}}{t_{j,k}} \right).
\]

(25)

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 boundeness of left hand side (deduced by the boundeness of the right hand side) implies \( \|\bar{u}_{j,k}\| \to 0 \) for any \( j \in I \). If \( j \not\in I \), 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. (13) and triangle inequality, we have

\[
\sum_{j \in I} (M - \|q\|) \|\bar{u}_{j,k}\| \leq \|q\| \|x_k - x\| + \sum_{j=1}^{N} t_{j,k} H_j^* (v_{j,k}) + \sum_{j \not\in I} \|q\| \|\bar{u}_{j,k}\| - t_{j,k} H_j^* \left( \frac{\bar{u}_{j,k}}{t_{j,k}} \right)
\]

\[
\leq \sum_{j=1}^{N} t_{j,k} (H_j^*(v_{j,k}) + \|q\| \|v_{j,k}\|) + \sum_{j \not\in I} t_{j,k} \left( \|q\| \|\bar{u}_{j,k}\| - H_j^* \left( \frac{\bar{u}_{j,k}}{t_{j,k}} \right) \right).
\]

Dividing both sides by \( t_{1,k} \), we can obtain

\[
(M - \|q\|) \sum_{j \in I} \|\bar{u}_{j,k}\|/t_{1,k} \leq \sum_{j=1}^{N} t_{j,k} \left( H_j^*(v_{j,k}) + \|q\| \|v_{j,k}\| \right)
\]

\[
+ \sum_{j \not\in I} t_{j,k} \left( \|q\| \|\bar{u}_{j,k}\| - H_j^* \left( \frac{\bar{u}_{j,k}}{t_{j,k}} \right) \right).
\]

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 \( j \not\in I \) 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 \( \{ \bar{u}_{j,k}/t_{j,k} \} \) implies that \( t_{j,k}/t_{1,k} \) converges to 0, hence \( \alpha_{j,\infty} = 0 \) and (iii) is proved.

### 3.4 Proof of lemma 3.1

Denote \( p_k = \nabla_x S_H(x_k, t_{1,k}, \cdots, t_{N,k}) \). Then, \( p_k \) is a maximizer of the maximization problem in eq. (10). 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

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

Here, for the second inequality above, we used the definition of \( x_k \) in eq. (13) and Cauchy-Schwarz inequality. Then, rearranging the terms and dividing by \( t_{1,k} \), we get

\[
\sum_{j=1}^N \frac{t_{j,k}}{t_{1,k}} (H_j(p_k) - \| \nu_{j,k} \| \| p_k \|) \leq \sum_{j=1}^N \frac{t_{j,k}}{t_{1,k}} (H_j(q) + \| \nu_{j,k} \| \| q \|).
\]

(26)

If \( \{ p_k \} \) 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 \( j = 1, \cdots, N \). Then, from eq. (26), for any \( k > K \), we obtain

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

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 \( \{ p_k \} \) 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. (9) and \( H_j \) is continuous in the interior of its domain for any \( j \), then we have

\[
\lim_{k \to \infty} \nabla S_H(x_k, t_{1,k}, \cdots, t_{N,k}) = \lim_{k \to \infty} (p_k, -H_1(p_k), \cdots, -H_N(p_k)) = (p, -H_1(p), \cdots, -H_N(p)).
\]

By the continuity property [25] Proposition XI 4.1.1] of the subdifferential operator \( \partial S_H \) of the convex lower semi-continuous function \( S_H \), we can conclude that

\[
(p, -H_1(p), \cdots, -H_N(p)) \in \partial S_H(x, 0, \cdots, 0),
\]

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

### 3.5 Proof of proposition 3.4

Denote \( \Delta S_k := S_H(x_k, t_{1,k}, \cdots, t_{N,k}) - S_H(x, 0, \cdots, 0) \). Proof of (i): For any \( q \in \partial J(x) \), by Hopf formula eq. (10), 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. (13), we get

\[
\Delta S_k \geq \langle q, x_k - x \rangle - \sum_{j=1}^N t_{j,k} H_j(q) = \sum_{j=1}^N t_{j,k} (\langle q, \nu_{j,k} \rangle - H_j(q)).
\]
Therefore, we have
\[
\liminf_{k \to \infty} \frac{\Delta S_k}{t_{1,k}} \geq \liminf_{k \to \infty} \sum_{j=1}^{N} \frac{t_{j,k}}{t_{1,k}} (\langle q, v_{j,k} \rangle - H_j(q)) = \sum_{j=1}^{N} \alpha_{j,\infty} (\langle q, v_{j,\infty} \rangle - H_j(q)),
\]
where we recall that \(\lim_{k \to \infty} v_{j,k} = v_{j,\infty}\) and \(\lim_{k \to \infty} t_{j,k}/t_{1,k} = \alpha_{j,\infty}\) by eq. (13). Here, \(q\) is an arbitrary element in \(\partial J(x)\), hence we obtain
\[
\liminf_{k \to \infty} \frac{\Delta S_k}{t_{1,k}} \geq \sup_{q \in \partial J(x)} \sum_{j=1}^{N} \alpha_{j,\infty} (\langle q, v_{j,\infty} \rangle - H_j(q)).
\tag{27}
\]

On the other hand, for any \(k\), consider the function \(\phi_k : [0, \infty) \to \mathbb{R}\) defined by \(\phi_k(t) := S_H\left(x + \sum_{j=1}^{N} t \alpha_{j,k} v_{j,k}, \alpha_{1,k}, \ldots, \alpha_{N,k}\right)\), where \(\alpha_{j,k} := t_{j,k}/t_{1,k}\). Since \(S_H\) is a convex function and \(\phi_k\) is its restriction on a line, then \(\phi_k \in \Gamma_0(\mathbb{R})\) with \(\text{dom} \phi_k = [0, \infty)\). Also, \(\phi_k\) is differentiable in \((0, \infty)\) since \(S_H\) is differentiable. The derivative of \(\phi_k\) at \(t_{1,k}\) is given by the chain rule:
\[
\phi'_k(t_{1,k}) = \sum_{j=1}^{N} \alpha_{j,k} \left(\langle \nabla x S_H(x_k, t_{1,k}, \ldots, t_{N,k}), v_{j,k} \rangle + \frac{\partial S_H}{\partial t}(x_k, t_{1,k}, \ldots, t_{N,k})\right).
\]

Since \(S_H\) satisfies the multi-time HJ equation eq. (9), we obtain
\[
\phi'_k(t_{1,k}) = \sum_{j=1}^{N} \alpha_{j,k} (\langle \nabla x S_H(x_k, t_{1,k}, \ldots, t_{N,k}), v_{j,k} \rangle - H_j(\nabla x S_H(x_k, t_{1,k}, \ldots, t_{N,k}))).
\]

From straightforward computation and the convexity of \(\phi_k\), we get
\[
\frac{\Delta S_k}{t_{1,k}} = \frac{\phi_k(t_{1,k}) - \phi_k(0)}{t_{1,k}} \leq \phi'_k(t_{1,k}) = \sum_{j=1}^{N} \alpha_{j,k} (\langle p_k, v_{j,k} \rangle - H_j(p_k)),
\tag{28}
\]
where \(p_k := \nabla x S_H(x_k, t_{1,k}, \ldots, t_{N,k})\).

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

Together with eq. (27), the equation eq. (13) 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. (11) evaluated at the point \((x_k, t_{1,k}, \ldots, t_{N,k})\) and by the convexity of \(J\) we deduce that
\[
S_L = J(\bar{u}_{0,k}) + \sum_{j=1}^{N} t_{j,k} H^*_j (\bar{u}_{j,k}) \geq J(x) + \langle q, \bar{u}_{0,k} - x \rangle + \sum_{j=1}^{N} t_{j,k} H^*_j (\bar{u}_{j,k}),
\]
for any \(q \in \partial J(x)\). Since \(S_L = S_H\), we have \(S_L(x_k, t_{1,k}, \ldots, t_{N,k}) = S_L(x, 0, \ldots, 0) = \Delta S_k\). 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
\[
\frac{\Delta S_k}{t_{1,k}} \geq \sum_{j=1}^{N} \alpha_{j,k} \langle q, v_{j,k} \rangle - \langle q, \bar{u}_{j,k} \rangle \geq \alpha_{j,k} H^*_j (\bar{u}_{j,k}),
\]
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
\[
\alpha_{j,k} H^*_j (\bar{u}_{j,k}) = \alpha_{j,k} \left(\langle \bar{u}_{j,k}, p_k \rangle - H_j(p_k)\right) = \langle \bar{u}_{j,k}, p_k \rangle - \alpha_{j,k} H_j(p_k).
\]

16
Combing the above two equations we obtain
\[
\frac{\Delta S_k}{t_{1,k}} \geq \sum_{j=1}^{N} \left( \alpha_{j,k}(q_j, v_{j,k}) + \left\langle p_k - q_j, \bar{u}_{j,k} \right\rangle - \alpha_{j,k} H^*_j(p_k) \right) \\
+ \sum_{j=1}^{N} \left( \alpha_{j,k} \left\langle q_j, v_{j,k} - \bar{u}_{j,k} \right\rangle + \alpha_{j,k} H^*_j \left( \frac{\bar{u}_{j,k}}{t_{j,k}} \right) \right).
\] (29)

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. (29) 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. (29), we have
\[
\lim_{k \to \infty} \frac{\Delta S_k}{t_{1,k}} \geq \sum_{j=1}^{N} \alpha_{j,\infty} \left( \langle q_j, v_{j,\infty} - \bar{w}_j \rangle + H^*_j (\bar{w}_j) \right) \geq \sum_{j=1}^{N} \alpha_{j,\infty} (\langle q_j, v_{j,\infty} \rangle - H_j(q)),
\] (30)

where the second inequality holds by the definition of Legendre transform eq. (7). From eq. (15), for any maximizer \( p \) in eq. (16),
\[
\lim_{k \to \infty} \frac{\Delta S_k}{t_{1,k}} = \sum_{j=1}^{N} \alpha_{j,\infty} \left( \langle p, v_{j,\infty} \rangle - H_j(p) \right).
\] (31)

Taking \( q = p \) in eq. (30) and comparing it with eq. (31), we can conclude that the inequalities in eq. (30) 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
\[
\lim_{k \to \infty} \frac{\Delta S_k}{t_{1,k}} = \sum_{j=1}^{N} \alpha_{j,\infty} \left( \langle p, v_{j,\infty} - \bar{w}_j \rangle + H^*_j (\bar{w}_j) \right).
\] (32)

On the other hand, for an arbitrary \( q \in \partial J(x) \), by eq. (30) and eq. (32), we have
\[
\sum_{j=1}^{N} \alpha_{j,\infty} \left( \langle p - q, v_{j,\infty} - \bar{w}_j \rangle + H^*_j (\bar{w}_j) \right) = \lim_{k \to \infty} \frac{\Delta S_k}{t_{1,k}} \geq \sum_{j=1}^{N} \alpha_{j,\infty} \left( \langle q, v_{j,\infty} - \bar{w}_j \rangle + H^*_j (\bar{w}_j) \right),
\]
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. (3) and eq. (6), we can deduce that \( v_{j,\infty} - \bar{w}_j \in N_{\partial J(x)}(p) = \partial J(x)(p) \). 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. (17) follows from this equality and eq. (32).

It remains to prove eq. (18). Consider any \( j \) such that \( \alpha_{j,\infty} \neq 0 \). Define \( f : \mathbb{R}^n \to \mathbb{R} \) 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 2.5,
\[
\partial f(\bar{w}_j) = -\partial I_{\partial J(x)}(v_{j,\infty} - \bar{w}_j) + H^*_j(\bar{w}_j) \ni -p + p = 0.
\]
Therefore, \( \bar{w}_j \) is a minimizer of \( f \), which finishes the proof.

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 consider the other direction and prove that the only convex solution is given by the two formulas.

In the field of PDE, 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.
However, the functions in convex analysis and optimization may take the value $+\infty$, which is an unusual condition in the PDE field. 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. (9).

As it is proved in the last section, $S_H$ defined in the Hopf formula eq. (10) 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 $H : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, such that $S^*(p, E^-) = J^*(p) + I_V(p, E^-)$, where $V := \{(p, E^-) : E^- \leq -\tilde{H}(p)\}$.

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 $S = S_H$.

One can actually apply the above arguments to weaker assumptions and obtain a generalized result, which is stated in the following corollary. The proof for this corollary is almost the same, so we omit it here. 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 $H_1, H_2, \ldots, H_N$ 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 $E_1^-, \cdots, E_N^- \in \mathbb{R}$ satisfy $(p, E_1^-, \cdots, E_N^-) \in \partial S(x,t_1,\cdots,t_N)$ for some $x \in \mathbb{R}^n$ and $t_1, \cdots, t_N > 0$, then $E_j^- + H_j(p) = 0$ for any $j = 1, \cdots, N$.

(ii). $S(x,0,\cdots,0) = J(x)$ 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

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

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 $\hat{S}$ is another function satisfying the assumptions (i)-(ii) with $\text{ri dom } \hat{S} = \text{ri dom } S$, then $\hat{S} = S$. In other words, the solution is unique when the relative interior of the domain is given.

**4.1 Proof of lemma 4.1**

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

$$\tilde{H}(p) := \inf\{-E^- : (p, E^-) \in \text{dom } S^*\} \in \mathbb{R} \cup \{\pm \infty\}.$$
As a result, we can conclude that $f$ satisfies the HJ equation eq. (9), is a convex set. Let $U$ be the domain of $S^*$. Hence, $S^* (p, \cdot)$ is non-decreasing with respect to $E^-$. Therefore, for any $p, E^- \in dom S^*$ implies $\{p\} \times (-\infty, E^-] \subseteq dom S^*$. Therefore we obtain $V_1 \subseteq dom S^* \subseteq V$.

In the next step, we prove $S^* = V$.

Denote $U = \{p \in \mathbb{R}^n : \tilde{H}(p) < +\infty\}$ (see fig. 8a). Since $U$ is the projection of $dom S^*$ along the direction $(0, 1)$, $U$ is a convex set. Let $p \in ri U$. Take $E^- < -\tilde{H}(p)$, then $\partial S^*(p, E^-) \neq \emptyset$ because $(p, E^-) \in ri dom S^*$. 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. (9), $E^- + H(p) = 0$. 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 $E^- < -H(p)$ and $E^- \neq -H(p)$, by proposition 6.6, the directional derivative of $S^*$ in the direction $(0, 1)$ is:

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

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 $p \in ri U$ and $E^- \leq -\tilde{H}(p)$.

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

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

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 dom $f$. By setting $f(p) = +\infty$ if $p \not\in U$, 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 $p_1, p_2 \in dom f$ and $\alpha \in (0, 1)$.
choose $E^- < -\tilde{H}(p_1)$ and $\tilde{E}^- < -\tilde{H}(p_2)$ (see fig. 8e), then we have

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

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

$$J(x) = S(x, 0) = \sup_{p, E^-} \langle x, p \rangle - S^*(p, E^-) = \sup_{(p, E^-) \in V} \langle x, p \rangle - f(p) = \sup_{p \in \text{dom } f} \langle x, p \rangle - f(p) = f^*(x).$$

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. (35), we deduce that

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

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 $S^*$

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

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

First, we prove $\tilde{H}$ cannot take the value $-\infty$ by contradiction. If there exists $p \in \mathbb{R}^n$ such that $\tilde{H}(p) = -\infty$. Then \{p\} $\times \mathbb{R} \subseteq \text{dom } S^*$, hence \{p\} $\times \mathbb{R} \times \{J^*(p)\} \subseteq \text{epi } S^*$. Therefore, \{(0, 1, 0)\} and \{(0, -1, 0)\} are in the asymptotic cone of $\text{epi } S^*$ by definition eq. 4. Then, by proposition 2.1 for any $q \in U$, \{q\} $\times \mathbb{R} \times \{J^*(q)\} \subseteq \text{epi } S^*$, which implies \{q\} $\times \mathbb{R} \subseteq \text{dom } S^*$. As a result, $\text{dom } S^* = U \times \mathbb{R}$. Hence, for any $p \in U$ and $E^- \in \mathbb{R}$, the directional derivative of $S^*$ in the direction \{(0, 1)\} is zero, because $S^*$ is a constant on the line \{p\} $\times \mathbb{R}$. In other words,

$$(S^*)'((p, E^-), (0, 1)) = 0 \text{ for any } p \in U \text{ and } E^- \in \mathbb{R}. \quad (35)$$

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

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

which contradicts eq. (35). 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 $\mathbb{R} \cup \{+\infty\}$.  

4.2 Proof of proposition 4.1

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 $V = \{(p, E^-) : E^- \leq -\tilde{H}(p)\}$. Then $\tilde{H}$ is a convex function whose domain is the projection of $\text{dom } S^*$ along \{(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. (6), we deduce that

$$\partial S^*(p, E^-) = \partial J^*(p) \times \{0\} + N_V(p, E^-) = \partial J^*(p) \times \{0\} + \{(tv, 1) : v \in \partial H(p), t \geq 0\}, \quad (36)$$

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 $v \in \partial H(p)$. Denote $x = x_0 + tv$. Then, by eq. (36) 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. (5) 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

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

in the domain of $\tilde{H}$. 


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

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 $t_1, \ldots, t_N > 0$. Let $\alpha_1, \ldots, \alpha_N$ be arbitrary positive real numbers and denote $\alpha = (\alpha_1, \ldots, \alpha_N)$. Define $T(x, s) = S(x, s\alpha_1, \ldots, s\alpha_N)$ 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. (9) 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 $\alpha_1, \ldots, \alpha_N > 0$, we have

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

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.

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. (10). Also, the minimizer $u$ in the Lax formula eq. (11) may be non-unique if the Hamiltonians are not differentiable. Moreover, 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 $\lambda H$ to it where $\lambda$ is a positive parameter and $H$ is a differentiable function satisfying (H1). When $\lambda$ 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 $\lambda H$.

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 [54, Theorem 2, p. 144] and [55, Theorem 3.1, p. 54]. 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. Also, the analysis can be adapted easily to other decomposition models with more degenerate terms. Moreover, one can also use the same procedure with other finite-valued, 1-coercive, differentiable and strictly convex function $H$ and obtain some similar results.
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 the similar arguments. Let $\| \cdot \|$ and $\| \cdot \|_*$ be two arbitrary norms whose dual norms are denoted as $\| \cdot \|_*$ and $\| \cdot \|_*$. 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 $F_2^*(y) = I\{\|y\|_* \leq 1\}$. 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

$$U(x) = \arg \min_{u \in \mathbb{R}^n} F_1(u) + F_2^*(x - u),$$  

$$S(x) = \min_{u \in \mathbb{R}^n} F_1(u) + F_2^*(x - u).$$  

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

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

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

$$u_{\lambda,\mu}(x) = \arg \min_{u \in \mathbb{R}^n} F_{1,\lambda,\mu}(u) + F_2^*(x - u),$$  

$$S_{\lambda,\mu}(x) = \min_{u \in \mathbb{R}^n} F_{1,\lambda,\mu}(u) + F_2^*(x - u).$$  

We expand the inf-convolution to get

$$u_{\lambda,\mu} = x - w_{\lambda,\mu},$$  

$$(v_{\lambda,\mu}, w_{\lambda,\mu}) := \arg \min_{v, w \in \mathbb{R}^n} F_1(v) + \frac{\lambda}{2} \|v\|_2^2 + F_2^*(w) + \frac{1}{2\mu} \|x - v - w\|_2^2,$$  

$$S_{\lambda,\mu} := \min_{v, w \in \mathbb{R}^n} F_1(v) + \frac{\lambda}{2} \|v\|_2^2 + F_2^*(w) + \frac{1}{2\mu} \|x - v - w\|_2^2.$$  

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 $\lambda$ and $\mu$ 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** There is a unique minimizer $(v_{\lambda,\mu}, w_{\lambda,\mu})$ to the problem eq. (40). Moreover, when $\lambda$ and $\mu$ are small, $\{v_{\lambda,\mu}\}$ and $\{w_{\lambda,\mu}\}$ are bounded.

**Lemma 5.2** When $\lambda$ and $\mu$ converge to zero, $v_{\lambda,\mu} + w_{\lambda,\mu}$ converges to $x$. Any cluster point of $v_{\lambda,\mu}$ is also a cluster point of $w_{\lambda,\mu}$ and vice versa. Moreover, any cluster point of $u_{\lambda,\mu}$ and $v_{\lambda,\mu}$ is in $U(x)$.

**Lemma 5.3** The function $S_{\lambda,\mu}(x)$ defined in eq. (39) is differentiable. For any fixed $x \in \mathbb{R}^n$, the gradient $p_{\lambda,\mu} := \nabla S_{\lambda,\mu}(x)$ is bounded when $\lambda$ and $\mu$ are small. Moreover, as $\lambda$ and $\mu$ approach zero, any cluster point of $p_{\lambda,\mu}$ is in $\partial S(x)$.

**Proposition 5.1** Assume $\lambda_k$ and $\mu_k$ converge to zero and $\lim_{k \to \infty} \frac{\lambda_k}{\mu_k} = c \in (0, +\infty)$. 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,

$$\lim_{k \to \infty} u_k = \arg \min_{u \in U(x)} \|u\|_2,$$  

$$\lim_{k \to \infty} p_k = \arg \min_{p \in \partial S(x)} \|p\|_2.$$
Here, for simplicity, we only demonstrate the method for a specific optimization problem whose objective function contains only two parts. In fact, this method works for more general cases, such as the decomposition models with more degenerate parts. 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.

5.1 Proof of lemma 5.1

It is easy to check that the objective function in eq. (40) 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. (40) and comparing it with eq. (37), 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 \(C := S(x) + \min_{v \in U(x)} \frac{1}{2} \|v\|_2^2\), then \(C\) is independent of \(\lambda\) and \(\mu\), and \(S_{\lambda,\mu}(x) \leq C\) when \(\lambda\) is small. From this inequality and the definition of \(S_{\lambda,\mu}(x)\) in eq. (40), we can derive a bound for \(x - v_{\lambda,\mu} - w_{\lambda,\mu}\) that reads

\[
\|x - v_{\lambda,\mu} - w_{\lambda,\mu}\|_2^2 \leq 2\mu S_{\lambda,\mu}(x) \leq 2C\mu.
\]

Therefore, \(v_{\lambda,\mu} + w_{\lambda,\mu}\) is bounded when \(\lambda\) and \(\mu\) are small.

Then, from the constraint given by the indicator function \(F_2^*\) in the minimization problem eq. (40), 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. Then the conclusion follows.

5.2 Proof of lemma 5.2

The convergence of \(v_{\lambda,\mu} + w_{\lambda,\mu}\) to \(x\) follows from eq. (41). 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

\[
(v_{\lambda,\mu}, w_{\lambda,\mu}) = \arg \min_{v, w \in \mathbb{R}^n} F_1(v) + \frac{\lambda}{2} \|v\|_2^2 + I\{\|w\|_* \leq 1\} + \frac{1}{2\mu} \|x - v - w\|_2^2
\]

\[
= \arg \min_{v, w \in \mathbb{R}^n} \frac{\lambda \mu}{2} \|v\|_2^2 + I\{\|w\|_* \leq 1\} + \frac{1}{2} \|x - v - w\|_2^2
\]

\[
= \arg \max_{v, w \in \mathbb{R}^n} \langle x, v \rangle + \langle x, w \rangle - \left(\frac{1}{2} \|v + w\|_2^2 + F_2^*(w) + \mu F_1(v) + \frac{\lambda \mu}{2} \|v\|_2^2\right),
\]

where we first multiply the objective function by \(\mu\) and then expand the quadratic term. The last maximization problem in eq. (42) is in the form of Hopf formula. The corresponding multi-time HJ equation with time variables \(\mu\) and \(\nu = \lambda \mu\) is given by

\[
\begin{aligned}
\frac{\partial}{\partial \mu} \hat{S}(y, z, \mu, \nu) &+ F_1(\nabla_y \hat{S}(y, z, \mu, \nu)) = 0, & y, z \in \mathbb{R}^n; \mu, \nu > 0; \\
\frac{\partial}{\partial \nu} \hat{S}(y, z, \mu, \nu) &+ \frac{\lambda \mu}{2} \|\nabla_y \hat{S}(y, z, \mu, \nu)\|_2^2 = 0, & y, z \in \mathbb{R}^n; \mu, \nu > 0; \\
\hat{S}(y, z, 0, 0) & = J(y, z), & y, z \in \mathbb{R}^n.
\end{aligned}
\]

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, it can be checked using the same argument in the previous two sections that the Hopf formula and Lax formula are both well-defined in \(\mathbb{R}^{2n} \times [0, \infty)^2\). Moreover, the solution \(\hat{S}\) is the classical solution to the multi-time HJ equation eq. (43) and its spatial gradient equals \((v_{\lambda,\mu}, w_{\lambda,\mu})\). To be specific, we have

\[
(v_{\lambda,\mu}, w_{\lambda,\mu}) = \nabla_{y, z} \hat{S}(x, x, \mu, \lambda \mu).
\]

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
\begin{equation}
(v, w) \in \partial J(x, x) \text{ iff } v + w = x \text{ and } \|w\|_* \leq 1.
\end{equation}

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. (44) and lemma 5.1. By eq. (41), $v_{\lambda, \mu} + w_{\lambda, \mu}$ converges to $x$. Also, $\|w_{\lambda, \mu}\|_* \leq 1$ is given by the constraint imposed by $F^*_2$ in eq. (40). Together with eq. (44), 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. (45) 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 $(\tilde{v}, \tilde{w})$ of $(v_{\lambda, \mu}, w_{\lambda, \mu})$,
\begin{equation}
(\tilde{v}, \tilde{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\},
\end{equation}
where the last two equalities follow from eq. (45) and the definition of $U(x)$ in eq. (37). In conclusion, any cluster point $\tilde{v}$ of $v_{\lambda, \mu}$ is in $U(x)$.

### 5.3 Proof of lemma 5.3

Rewriting the formula of $S_{\lambda, \mu}$ in eq. (40), we get
\begin{equation}
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).
\end{equation}

From straightforward computation, by proposition 2.7 and the definition of $(v_{\lambda, \mu}, w_{\lambda, \mu})$ in eq. (40), we obtain
\begin{equation}
\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\}
\end{equation}
\begin{equation}
= (\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{equation}

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
\begin{equation}
p_{\lambda, \mu} := \nabla S_{\lambda, \mu}(x) = \frac{1}{\mu} (x - v_{\lambda, \mu} - w_{\lambda, \mu}).
\end{equation}

Moreover, $p_{\lambda, \mu}$ is in the set $\partial F_1(v_{\lambda, \mu}) + \lambda v_{\lambda, \mu}$, which is bounded since the subdifferential of the norm $F_1$ is always bounded and $v_{\lambda, \mu}$ is also bounded by lemma 5.1. Hence, $\{p_{\lambda, \mu}\}$ is bounded when $\lambda$ and $\mu$ are small.

Let $p$ be a cluster point of $\{p_{\lambda, \mu}\}$. By taking a subsequence we can assume $\lambda_k$ and $\mu_k$ converge to zero and $p_k := p_{\lambda_k, \mu_k}$ converges to $p$. By lemma 5.1, $v := v_{\lambda_k, \mu_k}$ 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. (46), we have
\begin{equation}
p_k \in (\partial F_1(v_k) + \lambda_k v_k) \cap \partial F^*_2(w_k).
\end{equation}

Since the subdifferential operators $\partial F_1$ and $\partial F^*_2$ are continuous [50 Proposition XI 4.1.1], when $k$ goes to infinity, the above inclusion becomes
\begin{equation}
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).
\end{equation}

On the other hand, by proposition 2.7 and the definition of $S(x)$ and $U(x)$ in eq. (37), we have
\begin{equation}
\partial S(x) = \partial F_1(\tilde{u}) \cap \partial F^*_2(x - \tilde{u}),
\end{equation}
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. (49) and compare it with eq. (48) to conclude that $p \in \partial S(x)$. 

24
5.4 Proof of proposition 5.1

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
\[
\bar{u} := \text{arg min}_{u \in U(x)} \|u\|_2, \quad \bar{p} := \text{arg min}_{p \in \partial S(x)} ||p||_2.
\]
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 \( u = \bar{u}, p = \bar{p} \).

We still consider the solution \( \bar{S} \) to the multi-time HJ equation eq. (43). Denote the two Hamiltonians as \( H_1, H_2 \). Here, we consider the Lax formula. Denote \((y_k, \bar{y}_k, z_k, \bar{z}_k)\) to be the minimizer in the Lax formula, whose existence and uniqueness are easy to check and to hold here. To be specific, we have
\[
(y_k, \bar{y}_k, z_k, \bar{z}_k) = \text{arg min}_{y, \bar{y}, z, \bar{z} \in \mathbb{R}^n} J(x - y - z, x - \bar{y} - \bar{z}) + \mu_k H_1^*(\frac{y}{\mu_k}, \bar{y}; \frac{z}{\lambda_k \mu_k}, \bar{z}).
\] (50)

By straightforward computation, \( J(y, \bar{y}) = \|y\|^2_2 + F_2(\bar{y} - y), H_1^*(y, \bar{y}) = F_1^*(y) + I\{\bar{y} = 0\} \) and \( H_2^*(y, \bar{y}) = \|\bar{y}\|^2_2 + I\{\bar{y} = 0\} \). Hence we deduce \( y_k = \bar{y}_k = 0 \). Notice that eq. (50) is given by the inf-convolution of \( J \) and \( \mu_k H_1^*(\cdot/\mu_k) \) and \( \lambda_k \mu_k H_2^*(\cdot/(\lambda_k \mu_k)) \) and its minimal value is given by \( S(x, x, \mu_k, \lambda_k \mu_k) \). Hence, we can invoke proposition 2.7 to compute the spatial gradient of \( \bar{S} \) that reads
\[
\nabla_{y, z} \bar{S}(x, x, \mu_k, \lambda_k \mu_k) = (v_k, w_k).
\]

As a result, \((v_k, w_k)\) is in the set on the right hand side of eq. (51). Therefore, we get
\[
\begin{align*}
& v_k = \frac{z_k}{\lambda_k \mu_k}; \\
& \frac{y_k}{\mu_k} \in \partial F_1(v_k); \\
& \frac{q_k}{\mu_k} = w_k = x - y_k - z_k - v_k \in \partial F_2(y_k + z_k).
\end{align*}
\] (52)

Notice that \( p_k = (x - v_k - w_k)/\mu_k = (y_k + z_k)/\mu_k \) from eq. (47) and eq. (52). Together with eq. (52), we get
\[
\begin{align*}
p_k - \lambda_k v_k &= p_k - \frac{z_k}{\mu_k} = \frac{y_k}{\mu_k} \in \partial F_1(v_k); \\
x - \mu_k p_k - v_k &= x - y_k - z_k - v_k = q_k \in \partial F_2(y_k + z_k) = \partial F_2\left(\frac{y_k + z_k}{\mu_k}\right) = \partial F_2(p_k).
\end{align*}
\] (53)

On the other hand, since \( \bar{u} \) and \( \bar{p} \) are the minimizer and momentum of the original problem eq. (37), we have
\[
\bar{p} \in \partial F_1(\bar{u}) \cap \partial F_2^*(x - \bar{u}).
\] (54)

Combining eq. (53) and eq. (54), we obtain
\[
\begin{align*}
& \{p_k - \lambda_k v_k \in \partial F_1(v_k); \quad \text{and} \quad \bar{p} \in \partial F_1(\bar{u}). \\
& x - \mu_k p_k - v_k \in \partial F_2(p_k); \quad \text{and} \quad x - \bar{u} \in \partial F_2(\bar{p}).
\end{align*}
\]

Since the subdifferential operators \( \partial F_1 \) and \( \partial F_2 \) are monotone, by eq. (5), we obtain
\[
\begin{align*}
& \langle p_k - \lambda_k v_k - \bar{p}, v_k - \bar{u} \rangle \geq 0; \\
& \langle x - \mu_k p_k - v_k - (x - \bar{u}), p_k - \bar{p} \rangle \geq 0.
\end{align*}
\]

We sum up the two inequalities to get
\[
0 \geq -(p_k - \lambda_k v_k - \bar{p}, v_k - \bar{u}) - (x - \mu_k p_k - v_k - (x - \bar{u}), p_k - \bar{p}) = \lambda_k \langle v_k, v_k - \bar{u} \rangle + \mu_k \langle p_k, p_k - \bar{p} \rangle.
\]

Therefore we have
\[
\frac{\lambda_k}{2} \left( \|v_k - \bar{u}\|^2_2 + \|v_k\|^2_2 - \|\bar{u}\|^2_2 \right) + \frac{\mu_k}{2} \left( \|p_k - \bar{p}\|^2_2 + \|p_k\|^2_2 - \|\bar{p}\|^2_2 \right) \leq 0.
\]
We multiply the above inequality by $2/\mu_k$ and take the limit $k \to \infty$ to obtain
\[ c \left( \| u - \bar{u} \|^2_2 + \| u \|^2_2 - \| u \|^2_2 \right) + \left( \| p - \bar{p} \|^2_2 + \| p \|^2_2 - \| p \|^2_2 \right) \leq 0. \]
From lemma 5.2 and lemma 5.3, we know that $u \in U(x)$ and $p \in \partial S(x)$, hence $\| u \|^2_2 \geq \| \bar{u} \|^2_2$ and $\| p \|^2_2 \geq \| \bar{p} \|^2_2$, which implies
\[ 0 \leq c \left( \| u - \bar{u} \|^2_2 + \| p - \bar{p} \|^2_2 \right) \leq c \left( \| u - \bar{u} \|^2_2 + \| u \|^2_2 - \| u \|^2_2 \right) + \left( \| p - \bar{p} \|^2_2 + \| p \|^2_2 - \| p \|^2_2 \right) \leq 0. \] (55)
Therefore, the inequalities in eq. (55) become equalities and we can conclude that $u = \bar{u}$ and $p = \bar{p}$ because $c > 0$.

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 $u_j/t_j$. 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. (11). 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 $[56, 57, 58]$. 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] M. Bardi and L.C. Evans. On Hopf’s formulas for solutions of Hamilton-Jacobi equations. *Nonlinear Analysis: Theory, Methods & Applications*, 8(11):1373–1381, 1984.
[2] Eberhard Hopf. Generalized solutions of non-linear equations of first order. *Journal of Mathematics and Mechanics*, 14(6):951–973, 1965.
[3] Michael G Crandall, Hitoshi Ishii, and Pierre-Louis Lions. User’s guide to viscosity solutions of second order partial differential equations. *Bulletin of the American Mathematical Society*, 27(1):1–67, 1992.
[4] J. Darbon. On convex finite-dimensional variational methods in imaging sciences and Hamilton-Jacobi equations. *SIAM Journal on Imaging Sciences*, 8(4):2268–2293, 2015.
[5] C. Imbert. Convex analysis techniques for Hopf-Lax formulae in Hamilton-Jacobi equations. *Journal of Nonlinear and Convex Analysis. An International Journal*, 2(3):333–343, 2001.
[6] J.C. Rochet. The taxation principle and multi-time Hamilton-Jacobi equations. *Journal of Mathematical Economics*, 14(2):113 – 128, 1985.
[7] P. L. Lions and J-C. Rochet. Hopf formula and multitime Hamilton-Jacobi equations. *Proceedings of the American Mathematical Society*, 96(1):79–84, 1986.
[8] Guy Barles and Agnès Tourin. Commutation properties of semigroups for first-order Hamilton-Jacobi equations and application to multi-time equations. *Indiana University Mathematics Journal*, 50(4):1523–1544, 2001.
[9] Franco Cardin and Claude Viterbo. Commuting Hamiltonians and Hamilton-Jacobi multi-time equations. *Duke Math. J.*, 144(2):235–284, 08 2008.
[10] Monica Motta and Franco Rampazzo. Nonsmooth multi-time Hamilton-Jacobi systems. *Indiana University Mathematics Journal*, 55(5):1573–1614, 2006.
[11] Slawomir Plaskacz and Marc Quincampoix. Oleinik-Lax formulas and multitime Hamilton-Jacobi systems. *Nonlinear Analysis: Theory, Methods & Applications*, 51(6):957–967, December 2002.
[12] Alberto Bressan and Benedetto Piccoli. *Introduction to the mathematical theory of control*, volume 2. American institute of mathematical sciences Springfield, 2007.
[13] Lawrence Craig Evans and Panagiotis E Souganidis. Differential games and representation formulas for solutions of Hamilton-Jacobi-Isaacs equations. Technical report, Mathematics Research Center, University of Wisconsin-Madison, 1983.

26
[14] Panagiotis 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(3):217 – 257, 1985.

[15] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher. Simultaneous structure and texture image inpainting. *IEEE Transactions on Image Processing*, 12(8):882–889, Aug 2003.

[16] M. Elad, J.-L. Starck, P. Querre, and D.L. Donoho. Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA). *Applied and Computational Harmonic Analysis*, 19(3):340 – 358, 2005. Computational Harmonic Analysis - Part 1.

[17] Jean-François Aujol and Tony F. Chan. Combining geometrical and textured information to perform image classification. *Journal of Visual Communication and Image Representation*, 17(5):1004 – 1023, 2006.

[18] 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(11):2793–2800, Nov 2010.

[19] T.F. 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.

[20] Jérôme Gilles. Image decomposition: theory, numerical schemes, and performance evaluation. *Advances in Imaging and Electron Physics*, 158:89–137, 2009.

[21] Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60(1-4):259–268, November 1992.

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

[23] Antonin Chambolle and Pierre-Louis Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76(2):167–188, Apr 1997.

[24] Yves 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.

[25] John B. Garnett, Triet M. Le, Yves Meyer, and Luminita A. Vese. Image decompositions using bounded variation and generalized homogeneous Besov spaces. *Applied and Computational Harmonic Analysis*, 23(1):25 – 56, 2007. Special Issue on Mathematical Imaging.

[26] T. Le and L. Vese. Image decomposition using total variation and div(BMO). *Multiscale Modeling & Simulation*, 4(2):390–423, 2005.

[27] Luminita A. Vese and Stanley J. Osher. Image denoising and decomposition with total variation minimization and oscillatory functions. *Journal of Mathematical Imaging and Vision*, 20(1):7–18, Jan 2004.

[28] S. Osher, A. Solé, and L. Vese. Image decomposition and restoration using total variation minimization and the $H^{-1}$ norm. *Multiscale Modeling & Simulation*, 1(3):349–370, 2003.

[29] Jean-François Aujol, Gilles Aubert, Laure Blanc-Féraud, and Antonin Chambolle. Image decomposition application to SAR images. In Lewis D. Griffin and Martin Lillholm, editors, *Scale Space Methods in Computer Vision*, pages 297–312, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg.

[30] Jean-François Aujol, Gilles Aubert, Laure Blanc-Féraud, and Antonin Chambolle. Image decomposition into a bounded variation component and an oscillating component. *Journal of Mathematical Imaging and Vision*, 22(1):71–88, Jan 2005.

[31] Antonin Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1):89–97, Jan 2004.

[32] Jean-François Aujol and Antonin Chambolle. Dual norms and image decomposition models. *International Journal of Computer Vision*, 63(1):85–104, Jun 2005.

[33] Jérôme Gilles. Noisy image decomposition: A new structure, texture and noise model based on local adaptivity. *Journal of Mathematical Imaging and Vision*, 28(3):285–295, Jul 2007.

[34] Jean-François Aujol and Sung Ha Kang. Color image decomposition and restoration. *Journal of Visual Communication and Image Representation*, 17(4):916 – 928, 2006.

[35] Vincent Duval, Jean-François Aujol, and Luminita A. Vese. Projected gradient based color image decomposition. In *Scale Space and Variational Methods in Computer Vision*, pages 295–306, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.
[36] Vincent Duval, Jean-François Aujol, and Luminita A. Vese. Mathematical modeling of textures: Application to color image decomposition with a projected gradient algorithm. *Journal of Mathematical Imaging and Vision*, 37(3):232–248, Jul 2010.

[37] 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(4):913–917, April 1997.

[38] Jean-François Aujol, Guy Gilboa, Tony Chan, and Stanley Osher. Structure-texture image decomposition—modeling, algorithms, and parameter selection. *International Journal of Computer Vision*, 67(1):111–136, Apr 2006.

[39] T. Chan and S. Esedoglu. Aspects of total variation regularized L1 function approximation. *SIAM Journal on Applied Mathematics*, 65(5):1817–1837, 2005.

[40] Mila Nikolova. A variational approach to remove outliers and impulse noise. *Journal of Mathematical Imaging and Vision*, 20(1):99–120, Jan 2004.

[41] Jean-François Aujol and Guy Gilboa. Constrained and SNR-based solutions for TV-Hilbert space image denoising. *Journal of Mathematical Imaging and Vision*, 26(1):217–237, Nov 2006.

[42] M. J. Fadili, J. Starck, J. Bobin, and Y. Moudden. Image decomposition and separation using sparse representations: An overview. *Proceedings of the IEEE*, 98(6):983–994, June 2010.

[43] J.-L Starck, Michael Elad, and David Donoho. Redundant multiscale transforms and their application for morphological component separation. *Advances in Imaging and Electron Physics - ADV IMAG ELECTRON PHYS*, 132:287–348, 12 2004.

[44] 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(10):1570–1582, Oct 2005.

[45] Tony F. Chan, Selim Esedoglu, and Frederick E. Park. Image decomposition combining staircase reduction and texture extraction. *Journal of Visual Communication and Image Representation*, 18(6):464 – 486, 2007.

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

[47] Jérôme Darbon and Marc Sigelle. Image restoration with discrete constrained total variation part I: Fast and exact optimization. *Journal of Mathematical Imaging and Vision*, 26(3):261–276, Dec 2006.

[48] D. S. Hochbaum. An efficient algorithm for image segmentation, markov random fields and related problems. *Journal of the ACM*, 48(2):686–701, 2001.

[49] Jean-Baptiste Hiriart-Urruty and Claude Lemarechal. *Convex Analysis and Minimization Algorithms I: Fundamentals*, volume 305. Springer-Verlag Berlin Heidelberg, 1993.

[50] Jean-Baptiste Hiriart-Urruty and Claude Lemarechal. *Convex Analysis and Minimization Algorithms II: Advanced Theory and Bundle Methods*, volume 306. Springer-Verlag Berlin Heidelberg, 1993.

[51] Heinz H. Bauschke and Patrick L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer Publishing Company, Incorporated, 2011.

[52] Jonathan Borwein and Adrian S. Lewis. *Convex Analysis and Nonlinear Optimization: Theory and Examples*. Springer-Verlag New York, 2006.

[53] R. Tyrrell Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[54] Jean Pierre Aubin and A. Cellina. *Differential Inclusions: Set-Valued Maps and Viability Theory*. Springer-Verlag, Berlin, Heidelberg, 1984.

[55] H. Brézis. *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.

[56] Alfred Auslender and Marc Teboulle. Interior gradient and proximal methods for convex and conic optimization. *SIAM J. on Optimization*, 16(3):697–725, March 2006.

[57] Jonathan Eckstein. Nonlinear proximal point algorithms using bregman functions, with applications to convex programming. *Math. Oper. Res.*, 18(1):202–226, February 1993.

[58] Marc Teboulle. Convergence of proximal-like algorithms. *SIAM J. on Optimization*, 7(4):1069–1083, April 1997.