Set-valued Hamilton-Jacobi-Bellman Equations

Melih İşeri* Jianfeng Zhang†

March 31, 2024

Abstract

Building upon the dynamic programming principle for set-valued functions arising from many applications, in this paper we propose a new notion of set-valued PDEs. The key component in the theory is a set-valued Itô formula, characterizing the flows on the surface of the dynamic sets. In the contexts of multivariate control problems, we establish the wellposedness of the set-valued HJB equations, which extends the standard HJB equations in the scalar case to the multivariate case. As an application, a moving scalarization for certain time inconsistent problems is constructed by using the classical solution of the set-valued HJB equation.

MSC2020. 49L12, 49J53, 60H30, 49J20, 47H04

Keywords. Set valued functions, dynamic programming principle, set valued Itô formula, set valued HJB equation, multivariate control problems, time inconsistent problems, surface evolution equations

*Department of Mathematics, University of Michigan, United States, iseri@umich.edu.

†Department of Mathematics, University of Southern California, United States, jianfenz@usc.edu. This author is supported in part by NSF grant DMS-2205972.

1 Introduction

In this paper we consider set-valued functions taking the form:

$$\mathbb{V} : [0,T] \times \mathbb{R}^d \to 2^{\mathbb{R}^m}.$$

That is, for each $(t,x) \in [0,T]\times\mathbb{R}^d$, the value $\mathbb{V}(t,x)$ is a subset of $\mathbb{R}^m$ satisfying appropriate properties. Such set-valued functions, or their variants, have appeared in many applications, for example, stochastic viability problems (cf. Aubin-Da Prato [2]), multivariate superhedging problems (cf. Kabanov [15] and Bouchard-Touzi [5]), multivariate dynamic risk measures (cf. Feinstein-Rudloff [9]), time inconsistent optimization problems (cf. Karman-Ma-Zhang [16]), stochastic target problems (cf. Soner-Touzi [24, 26, 27]), and recently, nonzero sum (mean field) games with multiple equilibria (Feinstein-Rudloff-Zhang [11] and İşeri-Zhang [14]). Many of these problems were considered non-standard or even ill-posed in the literature, and overall we lacked convenient mathematical tools to treat them. When viewed as set values, however, their value functions (named set value functions[1] in this paper) enjoy many nice properties as the value function of standard control problems, in particular the crucial *dynamic programming principle*, or say the time consistency. Notice that, for a standard control problem, the combination of the dynamic programming principle and the Itô formula leads to the popular PDE approach. Then a natural question is:

$$\textit{Can we characterize these set value functions via set-valued PDEs?} \tag{1.1}$$

This is exactly the goal of the present paper: to introduce the PDE approach and hence recover the standard language for these challenging problems. To be precise, the main contributions of this paper are as follows:

- Introduce derivatives for set-valued functions and establish the set-valued Itô formula.
- Propose a notion of set-valued PDEs, and establish its wellposedness in the contexts of multivariate stochastic control problems.
- As an important application, construct a so called moving scalarization for a time inconsistent problem by using the classical solution of the corresponding set-valued PDE.

We hope this paper serves as the first step of our long term project on providing a convenient tool and systematic study for multivariate problems, including games.

Our first main result is the *set-valued Itô formula*, which roughly reads:

$$d\mathbb{V}(t,X_t) = \Big[\partial_t\mathbb{V} + \partial_x\mathbb{V}\cdot b + \frac{1}{2}\mathrm{tr}\,(\sigma^\top\partial_{xx}\mathbb{V}\sigma) - \mathcal{K}_{\mathbb{V}}\zeta + \xi\Big]dt + \Big[\partial_x\mathbb{V}\sigma + \zeta\Big]dB_t, \tag{1.2}$$

where $dX_t = b_t dt + \sigma_t dB_t$ is an arbitrary diffusion. We refer to Theorem 3.1 below for the precise meaning of the above formula. In particular, $\partial_t\mathbb{V}, \partial_x\mathbb{V}, \partial_{xx}\mathbb{V}$ are derivatives of $\mathbb{V}$ defined on the graph $\mathbb{G}_{\mathbb{V}}$, which is the graph of the boundary surface $\mathbb{V}_b$ and consists of all points (t,x,y) where y lies on $\mathbb{V}_b(t,x)$. The essence of the Itô formula is to characterize flows

[1] We use this for the value functions from the applications, to distinguish from general set-valued functions.

on the boundary surface. Given the surface's invariance under tangential deformations, a key feature of the set valued Itô formula is the inclusion of arbitrary driving forces ξ and ζ on $\mathbb{G}_{\mathbb{V}}$, which take values in the tangent space. This, along with the appropriate correction term $\mathcal{K}_{\mathbb{V}}\zeta$, ensures that the flows are not pushed away from the boundary surface.

Our set-valued HJB equation takes the following form: for some Hamiltonian function $h_{\mathbb{V}}$ and with appropriate terminal condition,

$$\sup_{a,\zeta} \mathbf{n}_{\mathbb{V}}(t,x,y)\cdot\Big[\partial_t\mathbb{V}(t,x,y)+h_{\mathbb{V}}(t,x,y,\partial_x\mathbb{V},\partial_{xx}\mathbb{V},a,\zeta)\Big]=0,\quad (t,x,y)\in\mathbb{G}_{\mathbb{V}}, \tag{1.3}$$

where a takes values in a control set, ζ takes values in the tangent space, and $\mathbf{n}_{\mathbb{V}}$ is the unit outward normal vector. This is derived by applying the above Itô formula on the dynamic programming principle for the set value function of the multivariate control problem. As we see, the introduction of ζ (and ξ) in (1.2) is crucial. We note that ξ disappears in the equation since $\mathbf{n}_{\mathbb{V}}\cdot\xi=0$. However, $\mathcal{K}_{\mathbb{V}}\zeta$ is nonlinear in ζ and thus $\mathbf{n}_{\mathbb{V}}\cdot\mathcal{K}_{\mathbb{V}}\zeta$ is an important component in the equation. The equation (1.3) can be rewritten equivalently in terms of the signed distance function $\mathbf{r}_{\mathbb{V}}$, see (5.4) below. We emphasize that $\mathbf{n}_{\mathbb{V}}$ is part of the solution here and the equation is satisfied only on the graph $\mathbb{G}_{\mathbb{V}}$, so the wellposedness of (1.3) has a completely different nature than that of standard PDEs.

In the scalar case: $m=1$, we can easily see that $\mathbb{V}(t,x)=[\underline{v}(t,x),\overline{v}(t,x)]$, where $\underline{v}$ and $\overline{v}$ are the value functions of the standard minimization and maximization problems, respectively. In this case, $\mathbf{n}_{\mathbb{V}}=1$ or -1, and the tangent space is degenerate and thus $\zeta=0$. Then (1.3) reduces exactly back to the standard HJB equations for $\underline{v}$ and $\overline{v}$. So our set-valued HJB equation is a natural extension of the standard HJB equation to the multivariate setting. Indeed, (1.3) means the optimization over each direction $\mathbf{n}_{\mathbb{V}}(t,x,y)$. Moreover, note that $\underline{v}$ and $\overline{v}$ are the boundaries of $\mathbb{V}$ in this case, namely $\mathbb{V}_b(t,x)=\{\underline{v}(t,x),\overline{v}(t,x)\}$, which inspires us to focus on the boundary surface $\mathbb{V}_b$ instead of on the whole set $\mathbb{V}$. We would like to point that, again since $\xi=0$, $\zeta=0$, in this case (1.2) also reduces back to the standard Itô formula for $\underline{v}(t,X_t)$ and $\overline{v}(t,X_t)$.

Our main result of the paper is that the dynamic set value function of the multivariate stochastic control problem is the unique classical solution of the set-valued HJB equation (1.3), provided $\mathbb{V}$ has sufficient regularity. We thus obtain the positive answer to our question (1.1) in this setting, which further opens the door to the PDE approach for more general multivariate problems. Such PDE characterization helps to understand better the structure and the nice properties of the dynamic set value function. In particular, it helps to construct (approximate) optimal controls with certain Markovian structure. Indeed, when $\mathbb{V}$ is smooth, as in standard verification theorem we may use the optimal arguments (a^*,ζ^*) of the Hamiltonian in (1.3) to construct an optimal control for a scalarized optimization problem, as we will explain in a later paragraph.

It should be noted that, under natural Lipschitz conditions our set values turn out to be convex and the set value function is uniformly continuous in (t,x), both are basic properties important in theory and in applications. However, we emphasize that our set-valued Itô formula does not require the convexity, and our arguments for the wellposedness of the set-valued HJB equations does not rely on this convexity either.

As an important application of our wellposedness result, we construct a moving scalarization for some time inconsistent problems, proposed by [16][2] and Feinstein-Rudloff [10]. Note that we are in the multivariate setting and in general it is not feasible to optimize the multiple objects simultaneously. In practice quite often one considers a scalarized optimization problem: $\max_{y\in\mathbb{V}(0,x_0)}\varphi(y)$, where $\varphi:\mathbb{R}^m\to\mathbb{R}$. This scalarized problem, unfortunately, is typically time inconsistent. The idea of a moving scalarization is to find a dynamic scalarization function $\Phi(t,X_{[0,t]};y)$, with $\Phi(0,x_0;y)=\varphi(y)$, such that the dynamic problem $\max_{y\in\mathbb{V}(t,X_t)}\Phi(t,X_{[0,t]};y)$ becomes time consistent. In Section 7 below we shall investigate this interesting application. In particular, we shall construct a moving scalarization for the mean variance problem explicitly.

At this point we would like to mention that the present paper considers classical solutions only. In particular, this requires that the set value $\mathbb{V}(t,x)$ is non degenerate and its boundary $\mathbb{V}_b(t,x)$ is a smooth $m-1$ dimensional manifold, namely the co-dimension is 1. It is our strong interest to remove these constraints and study viscosity solutions of more general set-valued PDEs, thereby broadening the applicability of the theory. We shall leave this important question to future research.

Some related literature. There is a large literature on set-valued analysis, we refer to the books Aubin-Frankowska [3], Kisielewicz [17, 18], and the reference therein. However, our approach is completely different from those in the set-valued analysis. First, the input (t,x) of our set-valued functions $\mathbb{V}$ is vector valued, while in the standard literature it is typically set-valued. Next, we focus on the dynamics of the boundary surface, rather than the dynamics of the whole set. Roughly speaking, we focus only on those special selectors whose flow remains on the boundary. These selectors have nice properties and are sufficient to characterize the whole sets. Moreover, the boundary is essentially the frontier which has intrinsic optimality and thus is also important from practical point of view. We shall mention the recent paper Ararat-Ma-Wu [1] on set-valued backward SDEs, which is highly relevant to our paper. Given our set value function $\mathbb{V}(t,x)$ and a state process X (e.g. a Brownian motion), we may introduce a set-valued process $\mathbb{Y}_t:=\mathbb{V}(t,X_t)$. In spirit the process $\mathbb{Y}$ should satisfy a set-valued backward SDE. However, besides that we employ completely different approaches, except in some simple cases the set-valued process $\mathbb{Y}$ does not satisfy the equation in [1]. That is, the objectives of the two works are different. We should mention that the applications mentioned in the beginning of this introduction fall into our framework, although technically our current results do not cover many of them (which we intend to study in our future research).

Our approach is strongly motivated by the studies on surface evolution equations, see e.g. Sethian [22], Evans-Spruck [8], Soner [23], Barles-Soner-Souganidis [4], the monograph Giga [12], and the references therein. These equations arise in various applications such as evolutions of phase boundaries, crystal growths, image processing, and mean-curvature flows, to mention a few. These works consider the dynamics of a set-valued function $\mathbb{V}(t)$, more precisely the boundary $\mathbb{V}_b(t)$, without the state variable x. In our terms, roughly speaking these works study first order set-valued ODEs, while we extend to second order set-valued PDEs. In particular, the set-valued Itô formula is not involved there. Another difference is, due to the nature of different applications, they study forward equations with initial conditions while we study backward problems with terminal conditions. This differ-

[2]In [16] it's called dynamic utility function, instead of moving scalarization.

ence would be crucial when one concerns path dependent setting (not covered in this paper), where one cannot do the time change freely due to the intrinsic adaptedness requirement.

Furthermore, within the surface evolution literature, our work is closely related to Soner-Touzi [27] which studies stochastic target problems by using mean curvature type geometric flows. In our contexts, their approach amounts to studying the following set value function via its signed distance function $\mathbf{r}_{\widehat{\mathbb{V}}}$:

$$\widehat{\mathbb{V}}(t) := \big\{(x,y) : x \in \mathbb{R}^d, y \in \mathbb{V}(t,x)\big\}, \text{ and thus } \widehat{\mathbb{V}}_b(t) := \big\{(x,y) : x \in \mathbb{R}^d, y \in \mathbb{V}_b(t,x)\big\}.$$

Clearly $\widehat{\mathbb{V}}$ and $\mathbb{V}$ are equivalent, with the same graph: $\mathbb{G}_{\widehat{\mathbb{V}}} = \mathbb{G}_{\mathbb{V}}$. The major difference here is that, while $\mathbf{r}_{\mathbb{V}}$ and $\mathbf{r}_{\widehat{\mathbb{V}}}$ agree on the graph (both are 0 by definition), their derivatives are different on the graph, and consequently, the equation derived in [27] is different from our set-valued HJB equation (1.3). In particular, in the scalar case: $m = 1$, as mentioned (1.3) reduces back to the standard HJB equations, but the equation for $\mathbf{r}_{\widehat{\mathbb{V}}}$ appears in a quite different form. Moreover, the normal vector $\mathbf{n}_{\widehat{\mathbb{V}}}$ of $\widehat{\mathbb{V}}$ is also different from $\mathbf{n}_{\mathbb{V}}$, and does not serve as a moving scalarization as we discussed. We shall provide more detailed discussions in Section 8.2 below.

Finally, we remark that there are some very interesting studies on (possibly discontinuous) viscosity solutions along this line, see e.g. [4], Chen-Giga-Goto [7], Soner-Touzi [24, 25, 26]. It will be interesting to explore these ideas in our setting.

The rest of the paper is organized as follows. In Section 2 we introduce the setting and define the intrinsic derivatives of set-valued functions. In Section 3 we prove the crucial set-valued Itô formula. In Section 4 we present the multivariate control problem. In Section 5 we introduce the set-valued HJB equation and show that the value function of the multivariate control problem is a classical solution, and the uniqueness of classical solutions is established in Section 6. Section 7 is devoted to the application of moving scalarization. In particular we solve it explicitly for the mean variance problem. In Section 8 we offer further discussions, including an extension to the case that the terminal condition is non-degenerate, and comparisons with [1] and [27]. Finally we complete some technical proofs in Appendix.

Notation. For the convenience of the readers, we list some frequently used notation here.

- $\mathbb{D}$: subsets of $\mathbb{R}^m$;
- $\mathbb{V} : [0,T] \times \mathbb{R}^d \to 2^{\mathbb{R}^m}$: set-valued functions, where in particular $\mathbb{V}(t,x) \subset \mathbb{R}^m$ is a set;
- $\mathbb{V}_b$ and $\mathbb{V}_o$: the boundary and interior of $\mathbb{V}$, respectively;
- $\mathbb{G}_{\mathbb{V}}$: the graph of $\mathbb{V}_b$, see (2.9) and (2.19);
- $\mathbf{r}$: the signed distance function, see (2.2);
- $\mathbf{n}$: the outward unit normal vector, see (2.4);
- $\mathbb{T}$: the tangent space, see (2.5);
- π: the projection onto the boundary, see (2.3);

- $\mathbf{I}_m$: the $m \times m$ identity matrix; and $\mathbf{I}_m - \mathbf{n}\mathbf{n}^\top$ is the projection operator onto $\mathbb{T}$;
- $\nabla \widehat{f}$: the standard derivatives of a function $\widehat{f}$;
- $\partial_\cdot f$: intrinsic derivatives of $f : \mathbb{G}_{\mathbb{V}} \to \mathbb{R}$,
see (2.7), (2.17), (2.20), (2.21), (2.22), and (2.23);
- X and (Y, Z): solutions to SDEs and BSDEs, respectively, see e.g. (4.1);
- Υ: forward dynamics typically on $\mathbb{V}_b$, see e.g. (3.2);
- (ξ, ζ): vector fields taking values on the tangent space $\mathbb{T}$, see e.g. (3.2);
- $\mathcal{K}^\sigma_{\mathbb{V}}\zeta$: correction term of ζ in the Itô formula, see (3.1);
- $\mathcal{L}^{b,\sigma}$: differential operator in the Itô formula, see (3.1);
- $(h^0_{\mathbb{V}}, h_{\mathbb{V}}, H_{\mathbb{V}})$: the related Hamiltonians, see (5.2);
- $\mathcal{L}$: the differential operator of the HJB equation, see (5.3).

2 Intrinsic derivatives of set-valued functions

Throughout the paper all vectors are viewed as column vectors, $\cdot$ denotes the inner product, and $^\top$,^{c} denote the transpose and complement, respectively. We denote by ∇ the gradient operator, and for a function $f : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}$, we take the convention that the second derivative $\nabla_{xy} f(x, y) := [\partial_{x_1 y} f, \cdots, \partial_{x_d y} f] \in \mathbb{R}^{m\times d}$.

2.1 Some basic materials

In this subsection we present some basic materials in geometry, which will be the starting point of our set-valued functions in this paper.

Let $\mathcal{D}^m_0 \subset 2^{\mathbb{R}^m}$ denote the space of closed sets $\mathbb{D}$ in $\mathbb{R}^m$, and denote by $\mathbb{D}_o$, $\mathbb{D}_b$, $\mathbb{D}^c$, the interior, the boundary[3], and the complement of $\mathbb{D}$, respectively. Define the Hausdorff distance:

$$d(\mathbb{D}, \tilde{\mathbb{D}}) := \big(\sup_{y\in\mathbb{D}} d(y, \tilde{\mathbb{D}})\big) \vee \big(\sup_{\tilde{y}\in\tilde{\mathbb{D}}} d(\tilde{y}, \mathbb{D})\big), \text{ where } d(y, \mathbb{D}) := \inf_{y'\in\mathbb{D}} |y - y'|, \ \forall y \in \mathbb{R}^m. \tag{2.1}$$

The following signed distance function of $\mathbb{D}$ plays a central role in our analysis:

$$\mathbf{r}_{\mathbb{D}}(y) := \begin{cases} d(y, \mathbb{D}_b), & y \in \mathbb{D}^c; \\ -d(y, \mathbb{D}_b), & y \in \mathbb{D}. \end{cases} \tag{2.2}$$

It is obvious that

$$\mathbb{D}_o = \{y \in \mathbb{R}^m : \mathbf{r}_{\mathbb{D}}(y) < 0\}, \ \mathbb{D}_b = \{y \in \mathbb{R}^m : \mathbf{r}_{\mathbb{D}}(y) = 0\}, \ \mathbb{D}^c = \{y \in \mathbb{R}^m : \mathbf{r}_{\mathbb{D}}(y) > 0\}.$$

[3] It may seem more natural to use $\partial\mathbb{D}$ to denote the boundary, however, in this paper we reserve the notation ∂ for the intrinsic (partial) derivatives.

We next let $\mathcal{D}_2^m$ denote the space of $\mathbb{D} \in \mathcal{D}_0^m$ such that $\mathbf{r}_{\mathbb{D}}$ is twice continuously differentiable with bounded derivatives on $O_\varepsilon(\mathbb{D}_b) := \{y \in \mathbb{R}^m : |\mathbf{r}_{\mathbb{D}}(y)| < \varepsilon\}$ for some $\varepsilon > 0$. We remark that the boundary $\mathbb{D}_b$ is a manifold without boundary, as regular as $\mathbf{r}_{\mathbb{D}}$. For each $y \in \mathbb{D}_b$, let $\mathbf{n}_{\mathbb{D}}(y) \in \mathbb{R}^m$ denote the outward unit normal vector at y. Moreover, for any $y \in O_\varepsilon(\mathbb{D}_b)$, let $\pi_{\mathbb{D}}(y)$ denote the unique projection of y on $\mathbb{D}_b$, i.e. $\pi_{\mathbb{D}}(y) \in \mathbb{D}_b$ satisfies:

$$y = \pi_{\mathbb{D}}(y) + \mathbf{r}_{\mathbb{D}}(y)\mathbf{n}_{\mathbb{D}}(\pi_{\mathbb{D}}(y)), \quad y \in O_\varepsilon(\mathbb{D}_b). \tag{2.3}$$

We refer to [19] and [6, Corollary 3.4.5] for discussions on the connections between the regularity of the signed distance function $\mathbf{r}_{\mathbb{D}}$, the surface regularity of $\mathbb{D}_b$, and the projection $\pi_{\mathbb{D}}$. It is clear that:

$$\mathbf{n}_{\mathbb{D}}(\pi_{\mathbb{D}}(y)) = \nabla_y \mathbf{r}_{\mathbb{D}}(y), \quad \text{and} \quad |\nabla_y \mathbf{r}_{\mathbb{D}}(y)| = 1, \quad y \in O_\varepsilon(\mathbb{D}_b). \tag{2.4}$$

For any $y \in \mathbb{D}_b$, let $\mathbb{T}_{\mathbb{D}}(y)$ denote the tangent space:

$$\mathbb{T}_{\mathbb{D}}(y) := \big\{\xi \in \mathbb{R}^m : \xi \cdot \mathbf{n}_{\mathbb{D}}(y) = 0\big\}, \quad y \in \mathbb{D}_b. \tag{2.5}$$

For a differentiable function $f : \mathbb{D}_b \to \mathbb{R}$, we define its intrinsic derivative $\partial_y f(y) \in \mathbb{T}_{\mathbb{D}}(y)$:

$$\lim_{\varepsilon \to 0} \frac{f(\theta(\varepsilon)) - f(y)}{\varepsilon} = \partial_y f(y) \cdot \theta'(0), \text{ for any smooth curve } \theta : \mathbb{R} \to \mathbb{D}_b \text{ with } \theta(0) = y. \tag{2.6}$$

Alternatively, for any smooth extension $\widehat{f} : \mathbb{R}^m \to \mathbb{R}$, i.e. $\widehat{f} = f$ on $\mathbb{D}_b$, we have

$$\partial_y f(y) = \nabla_y \widehat{f}(y) - [\nabla_y \widehat{f}(y) \cdot \mathbf{n}_{\mathbb{D}}(y)]\mathbf{n}_{\mathbb{D}}(y) = (\mathbf{I}_m - \mathbf{n}_{\mathbb{D}}(y)\mathbf{n}_{\mathbb{D}}(y)^\top)\nabla_y \widehat{f}(y). \tag{2.7}$$

where $\mathbf{I}_m \in \mathbb{R}^{m \times m}$ is the identity matrix, and $\mathbf{I}_m - \mathbf{n}_{\mathbb{D}}(y)\mathbf{n}_{\mathbb{D}}(y)^\top$ is the projection operator onto $\mathbb{T}_{\mathbb{D}}(y)$. We emphasize that $\partial_y f(y)$ does not depend on the choice of the extension $\widehat{f}$.[4]

We also recall the shape operator $\partial_y \mathbf{n}_{\mathbb{D}}(y) = [\partial_y \mathbf{n}_{\mathbb{D}}^1(y), \cdots, \partial_y \mathbf{n}_{\mathbb{D}}^m(y)] \in \mathbb{R}^{m \times m}$, which captures the curvatures of $\mathbb{D}_b$ at y.

2.2 Set-valued functions

Consider a function $\mathbb{V} : \mathbb{R} \to \mathcal{D}_2^m$. Denote

$$\begin{gathered} \mathbb{V}_b(x) := (\mathbb{V}(x))_b, \quad \mathbf{r}_{\mathbb{V}}(x,y) := \mathbf{r}_{\mathbb{V}(x)}(y), \quad \mathbf{n}_{\mathbb{V}}(x,y) := \mathbf{n}_{\mathbb{V}(x)}(y), \\ \pi_{\mathbb{V}}(x,y) := \pi_{\mathbb{V}(x)}(y), \quad \mathbb{T}_{\mathbb{V}}(x,y) := \mathbb{T}_{\mathbb{V}(x)}(y), \end{gathered} \tag{2.8}$$

and introduce the graph of $\mathbb{V}_b$:

$$\mathbb{G}_{\mathbb{V}} := \big\{(x,y) : x \in \mathbb{R}, y \in \mathbb{V}_b(x)\big\}. \tag{2.9}$$

When there is no notational confusion, we may drop the subscript $_{\mathbb{V}}$ in $\mathbf{r}_{\mathbb{V}}$, $\mathbf{n}_{\mathbb{V}}$, $\pi_{\mathbb{V}}$ and denote them as $\mathbf{r}, \mathbf{n}, \pi$. We say $\mathbb{V} \in C^2(\mathbb{R}; \mathcal{D}_2^m)$ if $\mathbf{r}_{\mathbb{V}}$ is twice continuously differentiable with bounded derivatives on $O_\varepsilon(\mathbb{G}_{\mathbb{V}})$ for some $\varepsilon > 0$. Note that in this case $\mathbb{V}(x) \in \mathcal{D}_2^m, \forall x \in \mathbb{R}$.

[4] We note that here we require f to be differentiable (as a function on a manifold). Then (2.6) and (2.7) are the usual definition of a derivative of a differentiable function on a manifold, however, we can identify them as vectors since $\mathbb{D}_b$ is a submanifold of $\mathbb{R}^m$. We also note that tangent space is defined as an equivalence class of curves on the manifold, but we similarly identify it by vectors.

Remark 2.1 *(i) We note that our results in the paper will only involve $\mathbf{r}_{\mathbb{V}}$ and its derivatives near $\mathbb{G}_{\mathbb{V}}$. For the convenience of our arguments, throughout the paper, we shall modify $\mathbf{r}_{\mathbb{V}}$ outside of $O_\varepsilon(\mathbb{G}_{\mathbb{V}})$, so that the modified function $\widehat{\mathbf{r}}_{\mathbb{V}}$ satisfies:*

- $\widehat{\mathbf{r}}_{\mathbb{V}} = \mathbf{r}_{\mathbb{V}}$ *on* $O_\varepsilon(\mathbb{G}_{\mathbb{V}})$*;*
- $\widehat{\mathbf{r}}_{\mathbb{V}} \in C^2(\mathbb{R}\times\mathbb{R}^m;\mathbb{R})$ *with bounded derivatives;*
- $|\widehat{\mathbf{r}}_{\mathbb{V}}(x,y)| \geq \frac{\varepsilon}{2}$ *for all* $(x,y)\notin O_\varepsilon(\mathbb{G}_{\mathbb{V}})$.

We emphasize that all our results will not rely on the choice of such a modification. For notational simplicity, we may identify the notation $\widehat{\mathbf{r}}_{\mathbb{V}}$ with $\mathbf{r}_{\mathbb{V}}$.

(ii) Similarly we may extend $\pi_{\mathbb{V}}$ outside of $O_\varepsilon(\mathbb{G}_{\mathbb{V}})$, still denoted as $\pi_{\mathbb{V}}$, such that

- *On $O_\varepsilon(\mathbb{G}_{\mathbb{V}})$, $\pi_{\mathbb{V}}(x,y)$ is the original unique projection of y on $\mathbb{V}_b(x)$ such that the counterpart of (2.3) holds true;*
- *$\pi_{\mathbb{V}}(x,y)$ is jointly measurable and $\pi_{\mathbb{V}}(x,y)\in\mathbb{V}_b(x)$ for all $(x,y)\in\mathbb{R}\times\mathbb{R}^m$;*
- *There exists a constant $C = C_{\mathbb{V}}$ such that*

$$|y-\pi_{\mathbb{V}}(x,y)| \leq C|\mathbf{r}_{\mathbb{V}}(x,y)|. \tag{2.10}$$

We remark that (2.10) follows from (2.3) when $(x,y)\in O_\varepsilon(\mathbb{G}_{\mathbb{V}})$, and the existence of C for arbitrary $(x,y)\notin O_\varepsilon(\mathbb{G}_{\mathbb{V}})$ is due to the facts that $|\mathbf{r}_{\mathbb{V}}(x,y)|\geq\frac{\varepsilon}{2}$ and that, if needed, we may modify the extension of $\mathbf{r}_{\mathbb{V}}$ further when $|y-\pi_{\mathbb{V}}(x,y)|$ is large.

(iii) We can also extend $\mathbf{n}_{\mathbb{V}}$ to the whole space $\mathbb{R}\times\mathbb{R}^m$, still denoted as $\mathbf{n}_{\mathbb{V}}$, such that $\mathbf{n}_{\mathbb{V}}$ is continuously differentiable with bounded derivatives. For convenience, we fix $\mathbf{n}_{\mathbb{V}}(x,y) := \nabla_y\mathbf{r}_{\mathbb{V}}(x,y)$ for all $(x,y)\in\mathbb{R}\times\mathbb{R}^m$. ■

Fix $x_0\in\mathbb{R}$. For each $y_0\in\mathbb{V}_b(x_0)$, in light of the modifications in Remark 2.1, we consider the following ODE which clearly has a unique solution $\Upsilon^{x_0,y_0}:\mathbb{R}\to\mathbb{R}^m$:

$$\Upsilon^{x_0,y_0}(x) = y_0 - \int_{x_0}^x \nabla_x\mathbf{r}_{\mathbb{V}}\ \mathbf{n}_{\mathbb{V}}(\tilde{x},\Upsilon^{x_0,y_0}(\tilde{x}))\,d\tilde{x}. \tag{2.11}$$

The operator Υ is crucial for characterizing the flows on the boundary surface, see Proposition 2.2. We note that the drift $-\nabla_x\mathbf{r}_{\mathbb{V}}\mathbf{n}_{\mathbb{V}}$ here is intrinsic in certain sense, and Υ leads to a local geodesic, see Remark 2.4, Proposition 2.5, and Remark 5.3 below.

Proposition 2.2 *Assume $\mathbb{V}\in C^2(\mathbb{R};\mathcal{D}_2^m)$ and $x_0\in\mathbb{R}$. Then, for any $x\in\mathbb{R}$,*

$$\mathbb{V}_b(x) = \big\{\Upsilon^{x_0,y_0}(x) : y_0\in\mathbb{V}_b(x_0)\big\}. \tag{2.12}$$

Consequently, (2.11) involves $\mathbf{r}_{\mathbb{V}}$ and $\mathbf{n}_{\mathbb{V}}$ only on $\mathbb{G}_{\mathbb{V}}$ and thus does not depend on the modifications of $\mathbf{r}_{\mathbb{V}}$ and $\mathbf{n}_{\mathbb{V}}$ in Remark 2.1.

Proof For notational simplicity, we drop the subscripts and denote $\mathbf{r}, \mathbf{n}, \pi$.

We first show that, for any $y_0 \in \mathbb{V}_b(x_0)$ and $x > x_0$, $\Upsilon(x) := \Upsilon^{x_0,y_0}(x) \in \mathbb{V}_b(x)$. Let $\varepsilon > 0$ be such that the original $\mathbf{r}$ in (2.2) is twice continuously differentiable on $O_\varepsilon(\mathbb{G}_\mathbb{V})$. Note that $(x_0, y_0) \in \mathbb{G}_\mathbb{V} \subset O_\varepsilon(\mathbb{G}_\mathbb{V})$. Denote

$$\tau := \inf\big\{x > x_0 : (x, \Upsilon(x)) \notin O_\varepsilon(\mathbb{G}_\mathbb{V})\big\}.$$

Then, for $x \in [x_0, \tau)$, apply the chain rule we have

$$\begin{aligned}\frac{d}{dx}\mathbf{r}(x,\Upsilon(x)) &= \nabla_x \mathbf{r}(x,\Upsilon(x)) - \nabla_y \mathbf{r}(x,\Upsilon(x)) \cdot [\nabla_x \mathbf{r}\ \mathbf{n}(x,\Upsilon(x))] \\ &= \nabla_x \mathbf{r}(x,\Upsilon(x))\Big[1 - \nabla_y \mathbf{r}(x,\Upsilon(x)) \cdot \mathbf{n}(x,\Upsilon(x))\Big].\end{aligned}$$

Here we used the fact that $\nabla_x \mathbf{r}$ is scalar. Recall the extension $\mathbf{n}(x,\Upsilon(x)) = \nabla_y \mathbf{r}(x,\Upsilon(x))$ in Remark 2.1 (iii) and by the second equality in (2.4) we have

$$\frac{d}{dx}\mathbf{r}(x,\Upsilon(x)) = \nabla_x \mathbf{r}(x,\Upsilon(x))\Big[1 - |\nabla_y \mathbf{r}(x,\Upsilon(x))|^2\Big] = 0, \quad x \in [x_0,\tau).$$

Note that $\mathbf{r}(x_0,\Upsilon(x_0)) = \mathbf{r}(x_0,y_0) = 0$. Then the above equation implies $\mathbf{r}(x,\Upsilon(x)) = 0$ for all $x \in [x_0,\tau)$, which in turn implies $\tau = \infty$. Thus $\mathbf{r}(x,\Upsilon(x)) = 0$ and hence $\Upsilon(x) \in \mathbb{V}_b(x)$ for all $x \geq x_0$. This implies that $\{\Upsilon^{y_0}(x) : y_0 \in \mathbb{V}_b(x_0)\} \subset \mathbb{V}_b(x)$ for all $x \geq x_0$. Similarly we can show that $\{\Upsilon^{y_0}(x) : y_0 \in \mathbb{V}_b(x_0)\} \subset \mathbb{V}_b(x)$ for all $x \leq x_0$, and hence for all $x \in \mathbb{R}$.

On the other hand, for any $y \in \mathbb{V}_b(x)$, set $y_0 := \Upsilon^{x,y}(x_0) \in \mathbb{V}_b(x_0)$. By the flow property of the ODE (2.11), we have $\Upsilon^{x,y}(x') = \Upsilon^{x_0,y_0}(x')$ for all $x' \in \mathbb{R}$, and thus $y = \Upsilon^{x,y}(x) = \Upsilon^{x_0,y_0}(x)$. This proves the opposite inclusion in (2.12). ■

Remark 2.3 *(i) Later on we will define $\partial_x \mathbb{V}(x,y) := -\nabla_x \mathbf{r}_\mathbb{V} \mathbf{n}_\mathbb{V}(x,y)$ for $(x,y) \in \mathbb{G}_\mathbb{V}$, see (2.20) below. Then (2.11) can be rewritten as*

$$\Upsilon^{x_0,y_0}(x) = y_0 + \int_{x_0}^x \partial_x \mathbb{V}(\tilde{x}, \Upsilon^{x_0,y_0}(\tilde{x}))d\tilde{x}. \tag{2.13}$$

Thus (2.12) can be viewed as the fundamental theorem of calculus for set-valued functions:

$$\mathbb{V}_b(x) = \mathbb{V}_b(x_0) + \int_{x_0}^x \partial_x \mathbb{V}(\tilde{x}, \mathbb{V}_b(\tilde{x}))d\tilde{x}. \tag{2.14}$$

(ii) However, (2.14) should be interpreted as (2.13) and (2.12), rather than the meaning in the standard set-valued analysis, which roughly speaking considers

$$\tilde{\Upsilon}(x) := y_0 + \int_{x_0}^x \partial_x \mathbb{V}(\tilde{x}, \tilde{\gamma}(\tilde{x}))d\tilde{x}, \quad \forall y_0 \in \mathbb{V}_b(x_0), \tilde{\gamma}(\tilde{x}) \in \mathbb{V}_b(\tilde{x}).$$

The above $\tilde{\Upsilon}(x)$ is in general not in $\mathbb{V}_b(x)$. See also related discussions in Section 8.3 below.

(iii) The set-valued Itô formula in the next section, which is one of the main results of this paper, can be viewed as the stochastic version of Proposition 2.2. ■

Remark 2.4 *For any $\Upsilon \in C^1(\mathbb{R};\mathbb{R}^m)$ taking values on the boundary surface, namely $\Upsilon(x) \in \mathbb{V}_b(x)$ for all $x \in \mathbb{R}$, by (2.4) we have*

$$\begin{aligned} 0 = \frac{d}{dx}\mathbf{r}_{\mathbb{V}}(x,\Upsilon(x)) &= \nabla_x \mathbf{r}_{\mathbb{V}}(x,\Upsilon(x)) + \nabla_y \mathbf{r}_{\mathbb{V}}(x,\Upsilon(x)) \cdot \nabla_x \Upsilon(x) \\ &= \nabla_x \mathbf{r}_{\mathbb{V}}(x,\Upsilon(x)) + \mathbf{n}_{\mathbb{V}}(x,\Upsilon(x)) \cdot \nabla_x \Upsilon(x). \end{aligned} \tag{2.15}$$

That is, for any such Υ with $\Upsilon(x) = y$, the projection of $\nabla_x\Upsilon(x)$ on $\mathbf{n}_{\mathbb{V}}(x,y)$ is independent of the choice of Υ: $(\nabla_x\Upsilon \cdot \mathbf{n}_{\mathbb{V}})\mathbf{n}_{\mathbb{V}}(x,y) = -\nabla_x \mathbf{r}_{\mathbb{V}}\mathbf{n}_{\mathbb{V}}(x,y) = \partial_x\mathbb{V}(x,y)$, so our definition of $\partial_x\mathbb{V}(x,y)$ is intrinsic in the sense that it is independent of the choice of Υ.

However, if we do not require $\partial_x\mathbb{V}$ to go along the normal direction, then there are alternative ways to define $\partial_x\mathbb{V}$. This relaxation will not affect our main results though, see Remark 5.3 below. ■

The next result, although technically not used in this paper, is interesting in its own right. We postpone its proof to Appendix.

Proposition 2.5 *Assume $\mathbb{V} \in C^2(\mathbb{R};\mathcal{D}_2^m)$ and $(x_0,y_0) \in \mathbb{G}_{\mathbb{V}}$. Then the curve $\Upsilon(x) := \Upsilon^{x_0,y_0}(x)$ determined by (2.11) is a local geodesic of the flow $\mathbb{V}$ in the following sense. For any continuous curve $\theta(x) \in \mathbb{V}_b(x)$ with $\theta(x_0) = y_0$, we have*

$$\overline{\lim_{x\to x_0}} \frac{1}{|x-x_0|}\big[L_\Upsilon(x_0,x) - L_\theta(x_0,x)\big] \le 0,$$

where $L_\Upsilon(x_0,x)$ (resp. $L_\theta(x_0,x)$) denotes the length of Υ (resp. θ) between x_0 and x.

We now turn to functions $f : \mathbb{G}_{\mathbb{V}} \to \mathbb{R}$. For fixed x, the intrinsic derivative $\partial_y f(x,y)$ for $y \in \mathbb{V}_b(x)$ is defined by (2.6) or equivalently by (2.7). We next define the intrinsic derivative of f with respect to x following the local geodesic Υ defined by (2.11):

$$\partial_x f(x_0,y_0) := \lim_{x\to x_0} \frac{f(x,\Upsilon^{x_0,y_0}(x)) - f(x_0,y_0)}{x-x_0}, \quad (x_0,y_0) \in \mathbb{G}_{\mathbb{V}}. \tag{2.16}$$

Equivalently, for any smooth extension $\widehat{f}$ of f, we have

$$\begin{aligned} \partial_x f(x_0,y_0) &= \lim_{x\to x_0} \frac{\widehat{f}(x,\Upsilon^{x_0,y_0}(x)) - \widehat{f}(x_0,y_0)}{x-x_0} \\ &= \nabla_x \widehat{f}(x_0,y_0) - \nabla_x \mathbf{r}_{\mathbb{V}} \nabla_y \widehat{f} \cdot \mathbf{n}_{\mathbb{V}}(x_0,y_0). \end{aligned} \tag{2.17}$$

Again, the right side above does not depend on the choice of the extension $\widehat{f}$.

We say $f \in C^1(\mathbb{G}_{\mathbb{V}};\mathbb{R})$ if f has continuous intrinsic derivatives $\partial_y f$ and $\partial_x f$. By (2.16), it is obvious that $\partial_x f$ is linear on f, and the product rule and the chain rule remain true:

$$\begin{aligned} \partial_x(fg) = g\partial_x f + f\partial_x g, &\text{ for all } f,g \in C^1(\mathbb{G}_{\mathbb{V}};\mathbb{R}); \\ \partial_x[g(f)] = g'(f)\partial_x f, &\text{ for all } f \in C^1(\mathbb{G}_{\mathbb{V}};\mathbb{R}), g \in C^1(\mathbb{R};\mathbb{R}). \end{aligned} \tag{2.18}$$

2.3 Intrinsic derivatives of set-valued functions

We now extend all the above analyses to functions $\mathbb{V} : [0,T] \times \mathbb{R}^d \to \mathcal{D}_2^m$. In this and the next section we may allow infinite time horizon $[0,\infty)$. However, in later sections we require T to be finite, so for simplicity we consider finite T here as well. Introduce $\mathbb{V}_b(t,x)$, $\mathbf{r}_{\mathbb{V}}(t,x,y)$, $\mathbf{n}_{\mathbb{V}}(t,x,y)$, $\pi_{\mathbb{V}}(t,x,y)$, $\mathbb{T}_{\mathbb{V}}(t,x,y)$ in an obvious manner as in (2.8) and denote

$$\mathbb{G}_{\mathbb{V}} := \big\{(t,x,y) : (t,x) \in [0,T] \times \mathbb{R}^d, y \in \mathbb{V}_b(t,x)\big\}. \tag{2.19}$$

As before we may use the simplified notations $\mathbf{r}, \mathbf{n}, \pi$ when there is no confusion, and we will always use their modified version or extension as in Remark 2.1.

Recall (2.16) and (2.17) when $\mathbb{V}$ is defined on $\mathbb{R}$. Now for our more general $\mathbb{V}$ and for any function $f : \mathbb{G}_{\mathbb{V}} \to \mathbb{R}$, we define its intrinsic partial derivatives $\partial_t f \in \mathbb{R}, \partial_x f \in \mathbb{R}^d, \partial_y f \in \mathbb{R}^m$, and the higher order intrinsic derivatives in an obvious manner, for example, the second order derivatives are defined as:

$$\partial_{x_i x_j} f := \partial_{x_i}(\partial_{x_j} f).$$

Moreover, for $f : \mathbb{G}_{\mathbb{V}} \to \mathbb{R}^n$, we define its intrinsic derivatives component wise.

Finally, by considering the special function $f_0(t,x,y) := y$ and its natural extension $\widehat{f}_0(t,x,y) = y$, applying (2.17) component wise we define the intrinsic derivatives of $\mathbb{V}$.

Definition 2.6 *For any $(t,x,y) \in \mathbb{G}_{\mathbb{V}}$ and by denoting $f_0(t,x,y) := y$, define*

$$\begin{aligned} \partial_t \mathbb{V}(t,x,y) &:= \partial_t f_0(t,x,y) = -\nabla_t \mathbf{r}\, \mathbf{n}(t,x,y) \in \mathbb{R}^m; \\ \partial_{x_i} \mathbb{V}(t,x,y) &:= \partial_{x_i} f_0(t,x,y) = -\nabla_{x_i} \mathbf{r}\, \mathbf{n}(t,x,y) \in \mathbb{R}^m, \quad i = 1, \cdots, d. \end{aligned} \tag{2.20}$$

We recall Remark 2.3 and note that (2.17) becomes: for any $f \in C^1(\mathbb{G}_{\mathbb{V}}; \mathbb{R})$,

$$\partial_t f = \nabla_t \widehat{f} + \nabla_y \widehat{f} \cdot \partial_t \mathbb{V}, \quad \partial_{x_i} f = \nabla_{x_i} \widehat{f} + \nabla_y \widehat{f} \cdot \partial_{x_i} \mathbb{V}, \quad \text{on } \mathbb{G}_{\mathbb{V}}. \tag{2.21}$$

Note that $\partial_t \mathbb{V}$ and $\partial_x \mathbb{V}$ are functions on $\mathbb{G}_{\mathbb{V}}$, then we may continue to define higher order derivatives of $\mathbb{V}$ by applying (2.17) or (2.21) repeatedly.

Lemma 2.7 *Assume $\mathbf{r}_{\mathbb{V}} \in C^2([0,T] \times \mathbb{R}^d; \mathcal{D}_2^m)$. Then*

$$\partial_{x_i x_j} \mathbb{V}(t,x,y) = -\nabla_{x_i x_j} \mathbf{r}\ \mathbf{n}(t,x,y) - \nabla_{x_j} \mathbf{r}\ \partial_{x_i} \mathbf{n}(t,x,y); \tag{2.22}$$

$$\partial_x \mathbf{n}^i = \nabla_{x y_i} \mathbf{r}; \qquad \partial_y \mathbf{n}^i = \nabla_{y_i y} \mathbf{r}. \tag{2.23}$$

The proof is quite straightforward, we thus postpone it to Appendix. Throughout the paper, we view $\partial_{xx} \mathbb{V} \in (\mathbb{R}^{d\times d})^m$ as a tensor with components $\partial_{xx} \mathbb{V}^i \in \mathbb{R}^{d\times d}$ for $i = 1, \cdots, m$, and we shall take the notational convention:

$$\begin{aligned} \partial_x \mathbb{V} &:= [\partial_{x_1} \mathbb{V}, \cdots, \partial_{x_d} \mathbb{V}] \in \mathbb{R}^{m\times d}, \quad \partial_{xx} \mathbb{V}^i := [\partial_{x_1 x} \mathbb{V}^i, \cdots, \partial_{x_d x} \mathbb{V}^i] \in \mathbb{R}^{d\times d}; \\ \partial_x \mathbf{n} &= [\partial_{x_1} \mathbf{n}, \cdots, \partial_{x_d} \mathbf{n}] \in \mathbb{R}^{m\times d}, \quad \partial_y \mathbf{n} = [\partial_{y_1} \mathbf{n}, \cdots, \partial_{y_m} \mathbf{n}] \in \mathbb{R}^{m\times m}. \end{aligned} \tag{2.24}$$

Remark 2.8 *(i) At* $(t,x,y)\in\mathbb{G}_{\mathbb{V}}$*, since* $|\mathbf{n}|^2=1$*, by* (2.18) *we have* $\partial_{x_j}\mathbf{n}\cdot\mathbf{n}=0$*. That is,* $\partial_{x_j}\mathbf{n}(t,x,y)\in\mathbb{T}_{\mathbb{V}}(t,x,y)$*. So* (2.22) *provides an orthogonal decomposition of* $\partial_{xx}V$*. In particular, unlike the first order derivatives in* (2.20)*,* $\partial_{x_ix_j}\mathbb{V}$ *is in general not parallel to* $\mathbf{n}$*.*

(ii) It is clear that

$$\partial_{xx}\mathbb{V}\cdot\mathbf{n}:=\big[\partial_{x_ix_j}\mathbb{V}\cdot\mathbf{n}\big]_{1\le i,j\le d}=-\nabla_{xx}\mathbf{r}\in\mathbb{R}^{d\times d}$$

is symmetric. However, in general $\partial_{xx}\mathbb{V}$ *is not symmetric:* $\nabla_{x_j}\mathbf{r}\partial_{x_i}\mathbf{n}\neq\nabla_{x_i}\mathbf{r}\partial_{x_j}\mathbf{n}$*,* $i\neq j$*.*

(iii) $\partial_{\mathbf{y}}\mathbf{n}=\nabla_{yy}\mathbf{r}\in\mathbb{R}^{m\times m}$ *is symmetric. Moreover, since* $\partial_{y_i}\mathbf{n}\cdot\mathbf{n}=0$*, we see that* 0 *is an eigenvalue of* $\partial_{\mathbf{y}}\mathbf{n}$ *with eigenvector* $\mathbf{n}$*.* ■

Example 2.9 *(i) Let* $w:[0,T]\times\mathbb{R}^2\to\mathbb{R}^2$ *and* $u:[0,T]\times\mathbb{R}^2\to(0,\infty)$ *be continuously differentiable. Set, with* $d=m=2$*,*

$$\begin{gathered}\mathbb{V}(t,x):=\big\{y\in\mathbb{R}^2:|y-w(t,x)|\le u(t,x)\big\},\\ \text{and thus}\quad \mathbb{V}_b(t,x)=\big\{y\in\mathbb{R}^2:|y-w(t,x)|=u(t,x)\big\}.\end{gathered}$$

It is clear that

$$\mathbf{r}(t,x,y)=|y-w(t,x)|-u(t,x),\ \ (t,x,y)\in[0,T]\times\mathbb{R}^2\times\mathbb{R}^2.$$

Then, for (t,x,y) *near* $\mathbb{G}_{\mathbb{V}}$ *(so that* $|y-w(t,x)|>0$*), by straightforward calculation,*

$$\begin{aligned}&\nabla_t\mathbf{r}=-\frac{(y-w)\cdot\nabla_t w}{|y-w|}-\nabla_t u;\quad \nabla_{x_i}\mathbf{r}=-\frac{(y-w)\cdot\nabla_{x_i}w}{|y-w|}-\nabla_{x_i}u;\quad \nabla_{y_i}\mathbf{r}=\frac{y_i-w^i}{|y-w|};\\ &\nabla_{x_ix_j}\mathbf{r}=\frac{\nabla_{x_i}w\cdot\nabla_{x_j}w-(y-w)\cdot\nabla_{x_ix_j}w}{|y-w|}-\frac{[(y-w)\cdot\nabla_{x_i}w][(y-w)\cdot\nabla_{x_j}w]}{|y-w|^3}-\nabla_{x_ix_j}u;\\ &\nabla_{x_iy_j}\mathbf{r}=-\frac{\nabla_{x_i}w^j}{|y-w|}+\frac{[(y-w)\cdot\nabla_{x_i}w](y_j-w^j)}{|y-w|^3},\quad \nabla_{y_iy_j}\mathbf{r}=\frac{\mathbf{1}_{\{i=j\}}}{|y-w|}-\frac{[y_i-w^i][y_j-w^j]}{|y-w|^3},\end{aligned}$$

Then, by (2.4)*, Definition 2.6, and Lemma 2.7, at* $(t,x,y)\in\mathbb{G}_{\mathbb{V}}$ *we have*

$$\begin{aligned}&\mathbf{n}=\frac{y-w}{u};\quad \partial_t\mathbb{V}=\big[\nabla_t w\cdot\mathbf{n}+\nabla_t u\big]\mathbf{n};\quad \partial_{x_i}\mathbb{V}=\big[\nabla_{x_i}w\cdot\mathbf{n}+\nabla_{x_i}u\big]\mathbf{n};\\ &\partial_{x_i}\mathbf{n}=\frac{1}{u}\big[-\nabla_{x_i}w+[\mathbf{n}\cdot\nabla_{x_i}w]\mathbf{n}\big],\quad \partial_{y_i}\mathbf{n}^j=\frac{1}{u}\big[\mathbf{1}_{\{i=j\}}-\mathbf{n}^i\mathbf{n}^j\big];\\ &\partial_{x_ix_j}\mathbb{V}=-\Big[\frac{1}{u}\big[\nabla_{x_i}w\cdot\nabla_{x_j}w-(\nabla_{x_i}w\cdot\mathbf{n})(\nabla_{x_j}w\cdot\mathbf{n})\big]-\nabla_{x_ix_j}w\cdot\mathbf{n}-\nabla_{x_ix_j}u\Big]\mathbf{n}\\ &\qquad\qquad -\frac{1}{u}\big[\nabla_{x_j}w\cdot\mathbf{n}+\nabla_{x_j}u\big]\big[\nabla_{x_i}w-(\nabla_{x_i}w\cdot\mathbf{n})\mathbf{n}\big].\end{aligned}\tag{2.25}$$

In particular, we see that in general $\partial_{x_ix_j}\mathbb{V}\neq\partial_{x_jx_i}\mathbb{V}$ *for* $i\neq j$*.*

(ii) Consider a special case that $w=0$ *and* u *satisfies the heat equation:*

$$\nabla_t u+\frac{1}{2}\mathrm{tr}\,(\nabla_{xx}u)=0.$$

Then by (2.25) *we have* $\partial_t\mathbb{V}=\nabla_t u\,\mathbf{n}$*,* $\partial_{x_ix_j}\mathbb{V}=\nabla_{x_ix_j}u\,\mathbf{n}$*, on* $\mathbb{G}_{\mathbb{V}}$*. Thus* $\mathbb{V}$ *satisfies the following equation:* $\partial_t\mathbb{V}+\frac{1}{2}\mathrm{tr}\,(\partial_{xx}\mathbb{V})=0$*, on* $\mathbb{G}_{\mathbb{V}}$*. This clearly implies the following set-valued heat equation:*

$$\mathbf{n}\cdot\big[\partial_t\mathbb{V}+\frac{1}{2}\mathrm{tr}\,(\partial_{xx}\mathbb{V})\big]=0,\ \ \text{on } \mathbb{G}_{\mathbb{V}}.$$ ■

We remark that we assumed $\mathbf{r}$ had bounded derivatives on $\mathbb{G}_{\mathbb{V}}$ in all above analyses. For our applications later, however, $\mathbb{V}(T,x)$ could be degenerate, in the sense that $\mathbb{V}(T,x) = \{g(x)\}$ is a singleton and hence a degenerate manifold in $\mathbb{R}^m$. Note that in (2.25), $\partial_x \mathbf{n}$, $\partial_y \mathbf{n}$, and $\partial_{xx}\mathbb{V}$ explode when $u = 0$. This motivates us to define the following space.

Definition 2.10 *(i) We say $\mathbb{V} \in C^{1,2}([0,T] \times \mathbb{R}^d; \mathcal{D}_2^m)$ if $\mathbf{r}_{\mathbb{V}} \in C^{1,2}(O_\varepsilon(\mathbb{G}_{\mathbb{V}}); \mathbb{R})$ for some $\varepsilon > 0$ such that all the related derivatives are bounded and uniformly Lipschitz continuous in y. Consequently, $\partial_t \mathbb{V}, \partial_x \mathbb{V}, \partial_{xx}\mathbb{V}$, $\partial_x \mathbf{n}$, $\partial_y \mathbf{n}$ are bounded and uniformly Lipschitz continuous in y on $\mathbb{G}_{\mathbb{V}}$.*

(ii) We say $\mathbb{V} \in C^{1,2}([0,T) \times \mathbb{R}^d; \mathcal{D}_2^m)$ if $\mathbb{V} \in C^0([0,T] \times \mathbb{R}^d; \mathcal{D}_0^m)$, and $\mathbb{V} \in C^{1,2}([0,T-\delta] \times \mathbb{R}^d; \mathcal{D}_2^m)$ for all $0 < \delta < T$. Note that we do not require $\mathbb{V}(T,x) \in \mathcal{D}_2^m$ here.

3 The set-valued Itô formula

We start with the probabilistic setting. Let $\Omega := \{\omega \in C([0,T], \mathbb{R}^d) : \omega_0 = 0\}$ be the canonical space, B the canonical process, i.e. $B(\omega) = \omega$, $\mathbb{P}$ the Wiener measure, i.e. B is an $\mathbb{P}$-Brownian motion, and $\mathbb{F} = \mathbb{F}^B$ the augmented filtration generated by B. For a generic Euclidean space E and $p \geq 1$, let $\mathbb{L}^p_{loc}(E)$ denote the space of $\mathbb{F}$-progressively measurable E-valued processes θ with $\int_0^T |\theta_t|^p dt < \infty$, a.s., and $\mathbb{L}^p_{loc}(\mathbb{R}^m; E)$ the space of $\mathbb{F}$-progressively measurable random fields $\xi : (t,\omega,y) \in [0,T] \times \Omega \times \mathbb{R}^m \to E$ such that $\xi(\cdot,\cdot,0) \in \mathbb{L}^p_{loc}(E)$ and ξ is uniformly Lipschitz continuous in y.

Fix $\mathbb{V} \in C^{1,2}([0,T] \times \mathbb{R}^d; \mathcal{D}_2^m)$ as in Definition 2.10 (i) with corresponding $\varepsilon > 0$, and $x_0 \in \mathbb{R}^d$, $b \in \mathbb{L}^1_{loc}(\mathbb{R}^d)$, $\sigma \in \mathbb{L}^2_{loc}(\mathbb{R}^{d\times d})$, $\xi \in \mathbb{L}^1_{loc}(\mathbb{R}^m; \mathbb{R}^m)$, $\zeta \in \mathbb{L}^2_{loc}(\mathbb{R}^m; \mathbb{R}^{m\times d})$. Denote,

$$X_t := x_0 + \int_0^t b_s ds + \int_0^t \sigma_s dB_s,$$

and introduce the (random) differential operators: recalling (2.24),

$$\begin{aligned} \mathcal{L}^{b,\sigma}\mathbb{V}(t,\omega,x,y) &:= \big[\partial_t \mathbb{V} + \partial_x \mathbb{V} b + \tfrac{1}{2}\mathrm{tr}\,(\sigma^\top \partial_{xx}\mathbb{V}\sigma)\big](t,\omega,x,y), \\ \mathcal{K}^\sigma_{\mathbb{V}}\zeta(t,\omega,x,y) &:= \big[\mathrm{tr}\,\big(\zeta^\top \partial_x \mathbf{n}\sigma + \tfrac{1}{2}\zeta^\top \partial_y \mathbf{n}\zeta\big)\,\mathbf{n}\big](t,\omega,x,y), \end{aligned} \tag{3.1}$$

where $\mathrm{tr}\,(\sigma^\top \partial_{xx}\mathbb{V}\sigma) \in \mathbb{R}^m$ with i-th component $\mathrm{tr}\,(\sigma^\top \partial_{xx}\mathbb{V}^i\sigma)$, and recalling Remark 2.1, we may extend the derivatives of $\mathbb{V}$ and $\mathbf{n}$ to $[0,T] \times \mathbb{R}^d \times \mathbb{R}^m$. Here $\partial_x \mathbb{V} b(t,\omega,x,y) = \partial_x \mathbb{V}(t,x,y) b(t,\omega)$, and we take this convention for all combinations of functions involving different variables. Moreover, as usual in the literature, when there is no confusion we omit the variable ω.

We are now ready to establish the set-valued Itô formula.

Theorem 3.1 *Let $\mathbb{V}, x_0, b, \sigma, X, \xi, \zeta$ be as above. Assume, for each i, $\zeta_i(t,\omega,y) \in \mathbb{T}_{\mathbb{V}}(t, X_t(\omega), y)$ holds for all $y \in \mathbb{V}_b(t, X_t(\omega))$, for $dt \times d\mathbb{P}$-a.e. (t,ω). For each $y \in \mathbb{R}^m$, recall Remark 2.1 and let Υ^y denote the unique strong solution of SDE:*

$$\Upsilon^y_t = y + \int_0^t \big[\mathcal{L}^{b,\sigma}\mathbb{V} - \mathcal{K}^\sigma_{\mathbb{V}}\zeta + \xi\big](s, X_s, \Upsilon^y_s) ds + \int_0^t \big[\partial_x \mathbb{V}\sigma + \zeta\big](s, X_s, \Upsilon^y_s) dB_s. \tag{3.2}$$

(i) Assume $\xi_t(y) \in \mathbb{T}_{\mathbb{V}}(t, X_t, y)$, for all $y \in \mathbb{V}_b(t, X_t)$, for $dt \times d\mathbb{P}$-a.e. (t,ω). Then

$$\{\Upsilon_t^y : y \in \mathbb{V}_b(0, x_0)\} \subset \mathbb{V}_b(t, X_t), \text{ for all } 0 \le t \le T, \text{ a.s.}$$

In particular, in this case no extension is needed in (3.2).

Moreover, if $\mathbb{V}_b$ takes values in connected compact sets, then the equality holds:

$$\big\{\Upsilon_t^y : y \in \mathbb{V}_b(0, x_0)\big\} = \mathbb{V}_b(t, X_t), \quad \text{for all } 0 \le t \le T, \text{ a.s.}$$

(ii) Assume $\xi_t(y) \cdot \mathbf{n}_{\mathbb{V}}(t, X_t, y) \le 0$ for all $y \in \mathbb{V}_b(t, X_t(\omega))$, for $dt \times d\mathbb{P}$-a.e. (t,ω). Then

$$\{\Upsilon_t^y : y \in \mathbb{V}_o(0, x_0)\} \subset \mathbb{V}_o(t, X_t), \text{ for all } 0 \le t \le T, \text{ a.s.}$$

(iii) Assume $\xi_t(y) \cdot \mathbf{n}_{\mathbb{V}}(t, X_t, y) \ge 0$ for all $y \in \mathbb{V}_b(t, X_t(\omega))$, for $dt \times d\mathbb{P}$-a.e. (t,ω). Then

$$\{\Upsilon_t^y : y \in \mathbb{V}^c(0, x_0)\} \subset \mathbb{V}^c(t, X_t), \text{ for all } 0 \le t \le T, \text{ a.s.}$$

Remark 3.2 *Note that Theorem 3.1 (i) is mainly to characterize the follows on the boundary surface. Since the forces by ξ, ζ in the tangent space will not push the flow away from the surface, so the Itô formula holds true for arbitrary ξ, ζ, where $\mathcal{K}_{\mathbb{V}}^{\sigma}\zeta$ is the correction term due to the quadratic covariation.*

We remark that, while all the results in this theorem remain true if we choose the trivial one: $\xi \equiv 0$, $\zeta \equiv 0$, this special choice is not sufficient for our main results later. For example, in the dynamic programming principle (4.16) below for the set value function $\mathbb{V}$, besides the standard control α, one may choose another free term $\phi \in \mathbb{V}(\tau, X_\tau^{t,x,\alpha})$ in the right side, which stands for an appropriate selector of $\mathbb{V}$. Roughly speaking, the presence of ϕ will induce appropriate ξ, ζ. In the standard literature where $\mathbb{V}$ is a singleton, ϕ is fixed and correspondingly $\xi \equiv 0$, $\zeta \equiv 0$.

Moreover, since the HJB equation is derived by applying the Itô formula on the dynamic programming equation, our set-valued HJB equation (5.3) below will also involve ζ.[5] ■

Remark 3.3 *In this remark we specify the measurability involved in the theorem. Note that the Hausdorff distance d in (2.1) induces naturally a topology on $\mathcal{D}_0^m$ and we may consider the measurability of $\mathbb{V}$ with respect to the corresponding Borel field. However, since our analysis goes through $\mathbf{r}_{\mathbb{V}}$, such a topology or measurability is not needed. Indeed, the measurability of the events involved in the above theorem and in the rest of the paper can be derived from $\mathbf{r}_{\mathbb{V}}$ directly. For example, in Theorem 3.1 (i), the event*

$$\Big\{\{\Upsilon_t^y : y \in \mathbb{V}_b(0, x_0)\} \subset \mathbb{V}_b(t, X_t)\Big\} = \Big\{\mathbf{r}_{\mathbb{V}}(t, X_t, \Upsilon_t^y) = 0, \ \forall y \in \mathbb{V}_b(0, x_0)\Big\}.$$

Since $\mathbf{r}_{\mathbb{V}}$ is continuous on (t, x, y) and Υ_t^y is continuous in y in $\mathbb{L}^2$-sense, one can easily show that the above event is $\mathcal{F}_t$-measurable (augmenting $\mathcal{F}_t$ as usual). ■

[5] ξ is not involved though as we will explain later.

Proof of Theorem 3.1. (i) Inspired by the arguments for Proposition 2.2, we shall show that

$$\bar{\mathbf{r}}_t := \mathbf{r}(t, X_t, \Upsilon_t) = 0. \tag{3.3}$$

Fix $y_0 \in \mathbb{V}_b(0, x_0)$ and denote $\Upsilon := \Upsilon^{x_0,y_0}$. Introduce

$$\tau := \inf\big\{t \ge 0 : (t, X_t, \Upsilon_t) \notin O_\varepsilon(\mathbb{G}_{\mathbb{V}})\big\} \wedge T. \tag{3.4}$$

Since $(0, x_0, y_0) \in \mathbb{G}_{\mathbb{V}}$, then $\tau > 0$, and $\mathbf{r}$ is smooth on $[0, \tau]$. By the standard Itô formula,

$$\begin{aligned}
&d\bar{\mathbf{r}}_t = d\mathbf{r}(t, X_t, \Upsilon_t) = \Lambda(t, X_t, \Upsilon_t)dt + M(t, X_t, \Upsilon_t)dB_t, \quad \text{where}\\
&\Lambda := \nabla_t \mathbf{r} + \nabla_x \mathbf{r} \cdot b + \nabla_y \mathbf{r} \cdot (\mathcal{L}^{b,\sigma}\mathbb{V} - \mathcal{K}^{\sigma}_{\mathbb{V}}\zeta + \xi)\\
&\quad + \frac{1}{2}\mathrm{tr}\Big(\sigma^\top \nabla_{xx}\mathbf{r}\sigma + (\partial_x \mathbb{V}\sigma + \zeta)^\top \nabla_{yy}\mathbf{r}(\partial_x\mathbb{V}\sigma + \zeta) + 2(\partial_x\mathbb{V}\sigma + \zeta)^\top \nabla_{xy}\mathbf{r}\sigma\Big);\\
&M := \nabla_x \mathbf{r}^\top \sigma + \nabla_y \mathbf{r}^\top (\partial_x \mathbb{V}\sigma + \zeta).
\end{aligned} \tag{3.5}$$

We proceed the rest of the proof in three steps.

Step 1. In this step we show that, for $0 \le t \le \tau$,

$$\Lambda(t, X_t, \Upsilon_t) = \mathbf{n}(t, X_t, \Upsilon_t) \cdot \xi_t(\Upsilon_t), \quad M(t, X_t, \Upsilon_t) = \big[\mathbf{n} \cdot \zeta_1, \cdots, \mathbf{n} \cdot \zeta_d\big](t, X_t, \Upsilon_t). \tag{3.6}$$

Indeed, recall Remark 2.1 (iii) that we have fixed the extension $\mathbf{n}(t, x, y) := \nabla_y \mathbf{r}(t, x, y)$, and similarly we consider this extension in (2.20), (2.22), and (2.23). Then,

$$\begin{aligned}
\mathbf{n} \cdot (\mathcal{L}^{b,\sigma}\mathbb{V} - \mathcal{K}^{\sigma}_{\mathbb{V}}\zeta) &= -\nabla_t \mathbf{r} - \nabla_x \mathbf{r} \cdot b - \tfrac{1}{2}\mathrm{tr}\left(\sigma^\top \nabla_{xx}\mathbf{r}\sigma + 2\zeta^\top \nabla_{xy}\mathbf{r}\sigma + \zeta^\top \nabla_{yy}\mathbf{r}\zeta\right);\\
\big(\nabla_{yy}\mathbf{r}\partial_x\mathbb{V}\big)_{ij} &= \nabla_{y_i y}\mathbf{r} \cdot \partial_{x_j}\mathbb{V} = -\nabla_{y_i y}\mathbf{r} \cdot \nabla_{x_j}\mathbf{r}\ \mathbf{n} = 0, \quad 1 \le i \le m, 1 \le j \le d;\\
\Big((\partial_x\mathbb{V})^\top \nabla_{xy}\mathbf{r}\Big)_{ij} &= \partial_{x_i}\mathbb{V} \cdot \nabla_{x_j y}\mathbf{r} = -\nabla_{x_i}\mathbf{r}\ \mathbf{n} \cdot \nabla_{x_j y}\mathbf{r} = 0, \quad 1 \le i, j \le d.
\end{aligned}$$

Here we used the facts $\partial_{x_i}\mathbf{n} \cdot \mathbf{n} = 0$, $\nabla_{y_i y}\mathbf{r} \cdot \mathbf{n} = 0$, $\nabla_{x_i y}\mathbf{r} \cdot \mathbf{n} = 0$. Plug these into the expression of Λ in (3.5) we obtain $\Lambda = \mathbf{n} \cdot \xi$ straightforwardly. Similarly,

$$M := \nabla_x \mathbf{r}^\top \sigma - \mathbf{n}^\top [\nabla_{x_1}\mathbf{r}\ \mathbf{n}, \cdots, \nabla_{x_d}\mathbf{r}\ \mathbf{n}]\sigma + \mathbf{n}^\top \zeta = \mathbf{n}^\top \zeta.$$

Thus (3.6) holds true.

Step 2. Recall (2.4) that $\mathbf{n}(t, x, y) = \mathbf{n}(t, x, \pi(t, x, y))$ for all $(t, x, y) \in O_\varepsilon(\mathbb{G}_{\mathbb{V}})$. Since $\xi(t, y), \zeta_i(t, y) \in \mathbb{T}_{\mathbb{V}}(t, X_t, y)$ for any $y \in \mathbb{V}_b(t, X_t)$, then, for $1 \le i \le d$,

$$\mathbf{n}(t, X_t, \Upsilon_t) \cdot \xi(t, \pi(t, X_t, \Upsilon_t)) = \mathbf{n}(t, X_t, \Upsilon_t) \cdot \zeta_i(t, \pi(t, X_t, \Upsilon_t)) = 0, \ 0 \le t \le \tau.$$

Note further that, by (2.4) again, $|\Upsilon_t - \pi(t, X_t, \Upsilon_t)| = |\bar{\mathbf{r}}_t|$, $0 \le t \le \tau$. Then

$$\begin{aligned}
\Lambda(t, X_t, \Upsilon_t) &= \mathbf{n}(t, X_t, \Upsilon_t) \cdot \Big(\xi(t, \Upsilon_t) - \xi(t, \pi(t, X_t, \Upsilon_t))\Big) = \tilde{b}_t \bar{\mathbf{r}}_t;\\
M^i(t, X_t, \Upsilon_t) &= \mathbf{n}(t, X_t, \Upsilon_t) \cdot \Big(\zeta_i(t, \Upsilon_t) - \zeta_i(t, \pi(t, X_t, \Upsilon_t))\Big) = \tilde{\sigma}^i_t \bar{\mathbf{r}}_t;
\end{aligned} \tag{3.7}$$

where $\tilde b_t, \tilde\sigma^i_t$ are appropriate processes bounded by the Lipschitz constants of ξ, ζ_i with respect to y, and M^i denotes the i-th column of M. Thus (3.5) becomes:

$$d\bar{\mathbf{r}}_t = \tilde b_t \bar{\mathbf{r}}_t dt + \tilde\sigma_t \bar{\mathbf{r}}_t dB_t, \quad 0 \le t \le \tau. \tag{3.8}$$

Introduce

$$\tilde\Gamma_t := \exp\Big(- \int_0^t \tilde\sigma_s \cdot dB_s - \int_0^t [\tilde b_s + \frac{1}{2}|\tilde\sigma_s|^2] ds\Big). \tag{3.9}$$

Then $\bar{\mathbf{r}}_t = \bar{\mathbf{r}}_0 \tilde\Gamma_t^{-1}$. Since $\bar{\mathbf{r}}_0 = \mathbf{r}(0, x_0, y_0) = 0$, then $\bar{\mathbf{r}}_t = 0$, $0 \le t \le \tau$, and thus, by (3.4) we must have $\tau = T$, a.s. This implies that $\Upsilon_t \in \mathbb{V}_b(t, X_t)$, $0 \le t \le T$, a.s.

Step 3. Moreover, assume further that $\mathbb{V}_b$ takes values in connected, compact sets. Note that $y \mapsto \Upsilon^y_t$ is a homeomorphism almost surely (See Kunita [20]). In particular, it is continuous and locally one-to-one. Since $\mathbb{V}_b(0, x_0)$ is compact, it is mapped to a closed set in $\mathbb{V}_b(0, x_0)$. By invariance of domains for manifolds without boundaries, $y \mapsto \Upsilon^y_t$ is an open mapping in relative topologies of $\mathbb{V}_b(0, x_0)$,$\mathbb{V}_b(t, X_t)$. Therefore, $\mathbb{V}_b(0, x_0)$ maps to an open set in $\mathbb{V}_b(t, X_t)$. This concludes the equality as we assumed connectedness.

(ii) Fix $y_0 \in \mathbb{V}_o(0, x_0)$ and denote $\Upsilon := \Upsilon^{x_0,y_0}$. It suffices to prove

$$\bar{\mathbf{r}}_t := \mathbf{r}(t, X_t, \Upsilon_t) < 0, \quad 0 \le t \le T, \quad \text{a.s.} \tag{3.10}$$

For this purpose, let $\delta < \frac{\varepsilon}{2}$ be small enough so that $\bar{\mathbf{r}}_0 < -\delta$. Introduce recursively a sequence of stopping times: $\tau_0 := 0$, and for $n = 0, 1, \cdots$,

$$\tau_{2n+1} := \inf\big\{t > \tau_{2n} : \bar{\mathbf{r}}_t = -\delta\big\} \wedge T; \qquad \tau_{2n+2} := \inf\big\{t > \tau_{2n+1} : |\bar{\mathbf{r}}_t| = 2\delta\big\} \wedge T.$$

Since $\bar{\mathbf{r}}_0 < -\delta$, by the desired continuity it is clear that

$$\tau_1 > \tau_0, \quad \bar{\mathbf{r}}_t \le -\delta,\ \tau_0 \le t \le \tau_1, \quad \text{and} \quad \bar{\mathbf{r}}_{\tau_1} = -\delta \text{ when } \tau_1 < T.$$

We now assume $\tau_1 < T$. Then $\tau_2 > \tau_1$, due to the desired continuity, and we consider $\tau_1 \le t \le \tau_2$. Since $|\bar{\mathbf{r}}_{\tau_1}| = \delta$, we have $|\bar{\mathbf{r}}_t| \le 2\delta < \varepsilon$, namely $(t, X_t, \Upsilon_t) \in O_\varepsilon(\mathbb{G}_{\mathbb{V}})$. Then (3.5) and (3.6) remain true for $\tau_1 \le t \le \tau_2$, and similarly to (3.7) and (3.8) we have

$$d\bar{\mathbf{r}}_t = \big[\tilde b_t \bar{\mathbf{r}}_t + \eta_t\big] dt + \tilde\sigma_t \bar{\mathbf{r}}_t dB_t,\ \tau_1 \le t \le \tau_2,\ \text{where } \eta_t := \mathbf{n}(t, X_t, \Upsilon_t) \cdot \xi(t, \pi(t, X_t, \Upsilon_t)).$$

for some bounded processes $\tilde b, \tilde\sigma$ constructed as in Step 2. This implies that $\tilde\Gamma_t \bar{\mathbf{r}}_t = \bar{\mathbf{r}}_0 + \int_0^t \tilde\Gamma_s \eta_s ds$ for the $\tilde\Gamma$ in (3.9). Since $\bar{\mathbf{r}}_{\tau_1} = -\delta < 0$, and by our assumption $\eta_t \le 0$, then

$$\bar{\mathbf{r}}_t < 0,\ \tau_1 \le t \le \tau_2, \quad \text{and in particular, } \bar{\mathbf{r}}_{\tau_2} = -2\delta \text{ when } \tau_2 < T.$$

Repeat the arguments, we see that, on $\{\tau_n < T\}$,

$$\tau_{n+1} > \tau_n, \quad \text{and} \quad \bar{\mathbf{r}}_t < 0, \text{ for all } t \le \tau_{n+1}; \tag{3.11}$$

$$\bar{\mathbf{r}}_{\tau_k} = -\delta, \quad \text{for } k \le n \text{ odd}, \quad \text{and} \quad \bar{\mathbf{r}}_{\tau_k} = -2\delta, \quad \text{for } 2 \le k \le n \text{ even}. \tag{3.12}$$

Denote $E := \cap_{n \ge 1}\{\tau_n < T\}$. For each $\omega \in E$, by (3.12) it is clear that $\bar{\mathbf{r}}$ is not left continuous at $\tau_*(\omega) := \lim_{n\to\infty} \tau_n(\omega)$, where the limit exists since τ_n is increasing. However, $\bar{\mathbf{r}}$ is continuous in t, a.s., then we must have $\mathbb{P}(E) = 0$. That is, for a.e. $\omega \in \Omega$, $\tau_n(\omega) = T$ for n large enough. Then by (3.11) we see that $\bar{\mathbf{r}}_t < 0$ for $0 \le t \le T$, a.s.

(iii) follows from similar arguments as in (ii). ∎

4 A multivariate control problem

Recall the canonical setting introduced in the beginning of Section 3. Given $0 \le t < T$, we shall also consider the shifted Brownian motion $B^t_s := B_s - B_t$, and the shifted filtration $\mathbb{F}^t := \mathbb{F}^{B^t}$ on $[t,T]$. For a generic Euclidean space E, let $\mathbb{L}^2(\mathcal{F}_t, E)$ denote the set of $\mathcal{F}_t$-measurable square integrable E-valued random variables, and $\mathbb{L}^2(\mathbb{F}^t, E)$ the set of $\mathbb{F}^t$-progressively measurable square integrable E-valued processes on $[t,T]$.

Let A be a domain in some Euclidean space. For each $t \in [0,T]$, our set of admissible controls $\mathcal{A}_t$ consists of all $\mathbb{F}^t$-progressively measurable A-valued processes α. We remark that in this paper we consider open loop controls, which is more convenient to study the regularities and to construct desired approximations for our value functions. However, as in standard stochastic control problems, one can easily see that the set values in this section will remain the same if we consider appropriate closed loop controls.

Given $(t,x) \in [0,T] \times \mathbb{R}^d$, consider the following controlled dynamics: for each $\alpha \in \mathcal{A}_t$,

$$\begin{aligned} X^{t,x,\alpha}_s &= x + \int_t^s b(r, X^{t,x,\alpha}_r, \alpha_r)dr + \int_t^s \sigma(r, X^{t,x,\alpha}_r, \alpha_r)dB_r, \\ Y^{t,x,\alpha}_s &= g(X^{t,x,\alpha}_T) + \int_s^T f(r, X^{t,x,\alpha}_r, Y^{t,x,\alpha}_r, Z^{t,x,\alpha}_r, \alpha_r)dr - \int_s^T Z^{t,x,\alpha}_r dB_r. \end{aligned} \tag{4.1}$$

Here X, Y, Z take values in $\mathbb{R}^d, \mathbb{R}^m, \mathbb{R}^{m\times d}$, respectively, and b, σ, f, g are in appropriate dimensions and satisfy certain technical conditions which will be specified later. We emphasize that Y is typically multiple dimensional: $m > 1$. Our set value is defined as:

$$\mathbb{V}(t,x) := \mathrm{cl}\big\{Y^{t,x,\alpha}_t : \alpha \in \mathcal{A}_t\big\} \subset \mathbb{R}^m. \tag{4.2}$$

Here cl denotes the closure. Thus $\mathbb{V}$ is a set-valued mapping $[0,T] \times \mathbb{R}^d \to 2^{\mathbb{R}^m}$. We now motivate this set value function in the following remarks.

Remark 4.1 *In the scalar case: $m = 1$, consider the standard control problems:*

$$\underline{v}(t,x) := \inf_{\alpha \in \mathcal{A}_t} Y^{t,x,\alpha}_t, \quad \overline{v}(t,x) := \sup_{\alpha \in \mathcal{A}_t} Y^{t,x,\alpha}_t.$$

Then it is obvious that

$$\mathbb{V}(t,x) = [\underline{v}(t,x), \overline{v}(t,x)]. \tag{4.3}$$

That is, the standard optimization problems are characterizing the boundary of our set value function. In this paper, we will characterize the boundary of $\mathbb{V}$ through a set-valued HJB equation, and thus we extend the scalar optimization problem to the multivariate setting. ■

Remark 4.2 *The set-valued functions can be used to analyze some time inconsistent optimization problems. Consider the well known mean variance optimization problem:*

$$\begin{aligned} &V_0 := \sup_{\alpha \in \mathcal{A}} \Big[\mathbb{E}[X^{0,x_0,\alpha}_T] - \frac{\lambda}{2} Var(X^{0,x_0,\alpha}_T)\Big], \\ &\text{where } X^{t,x,\alpha}_s = x + \int_t^s \alpha_r dr + \int_t^s \alpha_r dB_r. \end{aligned} \tag{4.4}$$

Here X, B, α are all scalar processes. Note that $Var(X_T) = \mathbb{E}[|X_T|^2] - |\mathbb{E}[X_T]|^2$. Introduce

$$\begin{gathered}\mathbb{V}(t,x) := \mathrm{cl}\big\{Y_t^{t,x,\alpha} : \alpha \in \mathcal{A}_t\big\}, \quad \textit{where}\\ Y_s^{t,x,\alpha,1} = X_T^{t,x,\alpha} - \int_s^T Z_r^{t,x,\alpha,1} dB_r, \quad Y_s^{t,x,\alpha,2} = |X_T^{t,x,\alpha}|^2 - \int_s^T Z_r^{t,x,\alpha,2} dB_r.\end{gathered} \tag{4.5}$$

Then one can easily verify that

$$V_0 := \sup_{y \in \mathbb{V}(0,x_0)} \varphi(y), \quad \textit{where} \quad \varphi(y) := y_1 + \frac{\lambda}{2}|y_1|^2 - \frac{\lambda}{2} y_2. \tag{4.6}$$

Our goal of this paper is to characterize the dynamic set value function $\mathbb{V}$. In fact, in this special case we can solve $\mathbb{V}$ explicitly, following the calculation in Pedersen-Peskir [21, Theorem 3, Part 2] [6]*:*

$$\mathbb{V}(t,x) := \Big\{(y_1,y_2) : y_1 \in \mathbb{R}, y_2 \geq e^{-(T-t)}x^2 + \frac{\big(y_1 - xe^{-(T-t)}\big)^2}{1 - e^{-(T-t)}}\Big\} \subset \mathbb{R}^2. \tag{4.8}$$

Then, given the set $\mathbb{V}(0,x_0)$, it is trivial to solve the deterministic optimization problem (4.6):

$$\begin{gathered}V_0 = x_0 + \frac{1}{2\lambda}[e^T - 1], \quad \textit{with optimal arguments in (4.6):}\\ y_1^\lambda := x_0 + \frac{1}{\lambda}[e^T - 1], \quad y_2^\lambda := |y_1^\lambda|^2 + \frac{1}{\lambda^2}[e^T - 1].\end{gathered} \tag{4.9}$$

We also refer to Section 7 below for the related time inconsistency issue. ■

Remark 4.3 *(i) The problem (4.2) can also be viewed as a stochastic target problem:*

$$\begin{gathered}\mathbb{V}(t,x) = \mathrm{cl}\big\{y \in \mathbb{R}^m : \exists \alpha \in \mathcal{A}_t, Z \in \mathbb{L}^2(\mathbb{F}^t, \mathbb{R}^{m\times d}) \ s.t.\ Y_T^{t,x,y,\alpha,Z} = g(X_T^{t,x,\alpha}),\ a.s.\big\},\\ \textit{where} \quad Y_s^{t,x,y,\alpha,Z} = y - \int_t^s f(r, X_r^{t,x,\alpha}, Y_r^{t,x,y,\alpha,Z}, Z_r, \alpha_r)dr + \int_t^s Z_r dB_r.\end{gathered} \tag{4.10}$$

(ii) Note that at above $\mathbb{V}(T,x) = \{g(x)\}$ is a singleton. In this respect we may easily extend our setting to non-degenerate terminal $\mathbb{G} : \mathbb{R}^d \to \mathcal{D}_0^m$. That is,

$$\mathbb{V}(t,x) := \mathrm{cl}\big\{y \in \mathbb{R}^m : \exists \alpha, Z \ \textit{such that}\ Y_T^{t,x,y,\alpha,Z} \in \mathbb{G}(X_T^{t,x,\alpha}),\ a.s.\big\}. \tag{4.11}$$

All our results in this paper can be extended to this case as well, see Section 8.1 below. ■

---

[6] In the mean variance portfolio selection literature, including [21], typically one uses geometric Brownian motion setting and the controlled dynamics (the wealth process) becomes: with u denoting the control,

$$\tilde X_s^{t,x,u} = x + \int_t^s u_r \tilde X_r^{t,x,u} dr + \int_t^s u_r \tilde X_r^{t,x,u} dB_r. \tag{4.7}$$

Clearly, this is equivalent to our formulation by setting $\alpha = u\tilde X$ and then $\tilde X^{t,x,u} = X^{t,x,\alpha}$. Moreover, as already observed in [21], the optimal u_s^* explodes when $\tilde X_s^{t,x,u} = 0$. Our α^* always exists however, as implied by [21].

Throughout the paper, we shall impose the following technical conditions.

Assumption 4.4 *(i)* $(b, \sigma) : (t, x, a) \in [0, T] \times \mathbb{R}^d \times A \to (\mathbb{R}^d, \mathbb{R}^{d\times d})$ *are bounded, uniformly continuous in* (t, a)*, and uniformly Lipschitz continuous in* x*.*

(ii) $f : (t, x, y, z, a) \in [0, T] \times \mathbb{R}^d \times \mathbb{R}^m \times \mathbb{R}^{m\times d} \times A \to \mathbb{R}^m$ *is uniformly continuous in* (t, x, a) *and* $f(t, x, 0, 0, a)$ *is bounded. Moreover,* f *is continuously differentiable in* (y, z) *with* $\nabla_y f, \nabla_z f$ *bounded and uniformly Lipschitz continuous in* (y, z)*.*

(iii) $g : x \in \mathbb{R}^d \to \mathbb{R}^m$ *is bounded and uniformly continuous in* x*.*

It is clear that (4.1) is wellposed for any $\alpha \in \mathcal{A}_t$, and thus $\mathbb{V}$ is well defined by (4.2).

4.1 Some basic properties

In this subsection we establish two basic properties of $\mathbb{V}$:

- For each (t, x), the set value $\mathbb{V}(t, x)$ is convex;
- The set value function $\mathbb{V}$ is uniformly continuous in (t, x) under the Hausdorff distance d. Consequently, the signed distance function $\mathbf{r}_{\mathbb{V}}$ is also uniformly continuous in (t, x).

These properties are interesting both in theory and in applications. However, we note that, by assuming the smoothness of $\mathbf{r}_{\mathbb{V}}$, our main results in this paper do not rely on these properties. All the proofs in this subsection are postponed to Appendix.

Proposition 4.5 *Under Assumption 4.4, the set value* $\mathbb{V}(t, x)$ *is compact and convex.*

Remark 4.6 *The Lipschitz property of* f *in* (y, z) *is crucial for the convexity. When* f *has quadratic growth, as we see in Example 4.7 below, this convexity is not guaranteed. We shall remark though such quadratic growth violates Assumption 4.4, which is assumed for technical reasons and can be weakened. We refer to Remark 5.2 which provides an intuition to this convexity from the PDE perspective.* ■

Example 4.7 *Consider the example where* f, g *and hence* $\mathbb{V}$ *are independent of* $x \in \mathbb{R}$*. Set* $m = 2$*,* $A = \{a \in \mathbb{R}^2 : |a| \le 1\}$*,* $g = 0$*, and* $f = (f_1, f_2)^\top$ *is specified at below:*

$$f_1(a, y, z) = a_1, \quad f_2(a, y, z) = \frac{2y_1y_2a_1}{1 + y_1^2} + (1 + y_1^2)a_2 - \frac{y_2z_1^2 + 2y_1z_1z_2}{1 + y_1^2} + \frac{4y_1^2y_2z_1^2}{(1 + y_1^2)^2}.$$

Then $\mathbb{V}(t)$ *can be solved explicitly and is nonconvex when* $T - t > \frac{1}{\sqrt{2}}$*:*

$$\begin{gathered}\mathbb{V}_b(t) = \Big\{\mathcal{Y}(t, \theta) : \forall \theta \in [0, 2\pi]\Big\}, \quad \text{where} \\ \mathcal{Y}_1(t, \theta) := (T - t)\cos\theta, \quad \mathcal{Y}_2(t, \theta) := (T - t)\big[1 + (T - t)^2\cos^2\theta\big]\sin\theta.\end{gathered} \tag{4.12}$$

■

Proposition 4.8 *Let Assumption 4.4 hold.*

(i) The set value function $\mathbb{V}$ is uniformly continuous in (t,x) under the Hausdorff distance d in (2.1). That is, for some modulus of continuity function ρ,

$$d\big(\mathbb{V}(t,x),\mathbb{V}(\tilde t,\tilde x)\big) \le \rho(|t-\tilde t|+|x-\tilde x|). \tag{4.13}$$

(ii) The signed distance function $\mathbf{r}_{\mathbb{V}}$ is also uniformly continuous in (t,x): for the above ρ,

$$\sup_{y\in\mathbb{R}^m} |\mathbf{r}_{\mathbb{V}}(t,x,y)-\mathbf{r}_{\mathbb{V}}(\tilde t,\tilde x,y)| \le 3\rho(|t-\tilde t|+|x-\tilde x|). \tag{4.14}$$

We note that $\mathbf{r}_{\mathbb{V}}$ is by definition Lipschitz continuous in y with Lipschitz constant 1.

Remark 4.9 *If we assume further that f,g are uniformly Lipschitz continuous in x and b,σ,f are uniformly Hölder-$\frac12$ continuous in t, then following the same arguments one can easily improve the regularity: both $\mathbb{V}$ and $\mathbf{r}_{\mathbb{V}}$ are uniformly Lipschitz continuous in x and uniformly Hölder-$\frac12$ continuous in t.*

4.2 The dynamic programming principle (DPP)

We now establish the DPP for $\mathbb{V}$. For $0\le t<T$, $x\in\mathbb{R}^d$, $\mathbb{F}^t$-stopping time $\tau\ge t$, $\phi\in\mathbb{L}^2(\mathcal{F}^t_\tau,\mathbb{R}^m)$, and $\alpha\in\mathcal{A}_t$, introduce:

$$Y^{\tau,\phi;t,x,\alpha}_s = \phi + \int_s^\tau f(r,X^{t,x,\alpha}_r,Y^{\tau,\phi;t,x,\alpha}_r,Z^{\tau,\phi;t,x,\alpha}_r,\alpha_r)dr - \int_s^\tau Z^{\tau,\phi;t,x,\alpha}_r dB_r. \tag{4.15}$$

Theorem 4.10 *Let Assumption 4.4 hold and $\mathbb{V}$ be defined by (4.2). For any $0\le t<T$, $x\in\mathbb{R}^d$, and any $\mathbb{F}^t$-stopping time $\tau\ge t$, it holds*

$$\mathbb{V}(t,x) = \mathrm{cl}\big\{Y^{\tau,\phi;t,x,\alpha}_t : \forall\alpha\in\mathcal{A}_t,\ \phi\in\mathbb{L}^2(\mathcal{F}^t_\tau,\mathbb{R}^m)\ s.t.\ \phi\in\mathbb{V}(\tau,X^{t,x,\alpha}_\tau)\ a.s.\big\}. \tag{4.16}$$

Proof Without loss of generality we prove (4.16) only at $(t,x)=(0,x_0)$, and for notational simplicity we omit the superscripts ^{0,x_0}. Denote the right side of (4.16) as $\tilde{\mathbb{V}}(0,x_0)$.

Step 1. In this step we show that $\mathbb{V}(0,x_0)\subset\tilde{\mathbb{V}}(0,x_0)$. Fix arbitrary $y_0\in\mathbb{V}(0,x_0)$ and $\varepsilon>0$. By definition of $\mathbb{V}(0,x_0)$ there exists $\alpha=\alpha^\varepsilon\in\mathcal{A}_0$ such that $|y_0-Y^\alpha_0|\le\varepsilon$. Denote $\phi:=Y^\alpha_\tau$. It is clear that $Y^\alpha_0=Y^{\tau,\phi;\alpha}_0$ and thus $|y_0-Y^{\tau,\phi;\alpha}_0|\le\varepsilon$. We claim that

$$\phi\in\mathbb{V}(\tau,X^\alpha_\tau)\text{ a.s.} \tag{4.17}$$

Then $Y^{\tau,\phi;\alpha}_0\in\tilde{\mathbb{V}}(0,x_0)$, and by the arbitrariness of $\varepsilon>0$ we obtain $y_0\in\tilde{\mathbb{V}}(0,x_0)$.

We first prove (4.17) in the case that $\tau\equiv t$ is deterministic. Consider the shifted canonical space: $\Omega_t:=\{\omega\in C([t,T],\mathbb{R}^d):\omega_t=0\}$. For any $\omega\in\Omega,\tilde\omega\in\Omega_t$, and $\xi\in\mathbb{L}^2(\mathcal{F}_T)$, introduce

$$(\omega\oplus_t\tilde\omega)_s := \omega_s\mathbf{1}_{[0,t)}+(\omega_t+\tilde\omega_s)\mathbf{1}_{[t,T]}(s),\quad \xi^{t,\omega}(\tilde\omega):=\xi(\omega\oplus_t\tilde\omega).$$

Then it is clear that $\omega \oplus_t \tilde{\omega} \in \Omega$, and $\xi^{t,\omega} \in \mathbb{L}^2(\mathcal{F}^t_T)$ for a.e. $\omega \in \Omega$. In particular, for a.e. $\omega \in \Omega$, we have $\alpha^{t,\omega} \in \mathcal{A}_t$, and by (4.1) and denoting $\psi^{\alpha,t,\omega} := (\psi^{0,x_0,\alpha})^{t,\omega}$ for $\psi = X, Y, Z$,

$$X^{\alpha,t,\omega}_s = X^\alpha_t(\omega) + \int_t^s b(r, X^{\alpha,t,\omega}_r, \alpha^{t,\omega}_r)dr + \int_t^s \sigma(r, X^{\alpha,t,\omega}_r, \alpha^{t,\omega}_r)dB^t_r,$$
$$Y^{\alpha,t,\omega}_s = g(X^{\alpha,t,\omega}_T) + \int_s^T f(r, X^{\alpha,t,\omega}_r, Y^{\alpha,t,\omega}_r, Z^{\alpha,t,\omega}_r, \alpha^{t,\omega}_r)dr - \int_s^T Z^{\alpha,t,\omega}_r dB^t_r.$$

This implies (4.17) immediately in the case $\tau \equiv t$.

Next, assume τ is discrete, namely it takes finitely many values $t_1 < \cdots < t_n$. Note that

$$\phi = Y^\alpha_\tau = \sum_{i=1}^n Y^\alpha_{t_i} \mathbf{1}_{\{\tau = t_i\}}, \qquad \mathbb{V}(\tau, X^\alpha_\tau) = \sum_{i=1}^n \mathbb{V}(t_i, X^\alpha_{t_i}) \mathbf{1}_{\{\tau = t_i\}}.$$

Since $Y^\alpha_{t_i} \in \mathbb{V}(t_i, X^\alpha_{t_i})$, a.s., then it is clear that $\phi \in \mathbb{V}(\tau, X^\alpha_\tau)$, a.s.

Finally, for general τ, by standard arguments one can construct discrete τ_n such that $\tau_n \downarrow \tau$. By above $Y^\alpha_{\tau_n} \in \mathbb{V}(\tau_n, X^\alpha_{\tau_n})$, a.s. Note that Y^α, X^α are continuous in t, a.s., and $\mathbb{V}$ is continuous in (t,x), then $Y^\alpha_{\tau_n} \to Y^\alpha_\tau$, $\mathbb{V}(\tau_n, X^\alpha_{\tau_n}) \to \mathbb{V}(\tau, X^\alpha_\tau)$, a.s. This, together with the fact that $\mathbb{V}(\tau, X^\alpha_\tau)$ is closed, implies (4.17) in the general case.

Step 2. We next prove the opposite inclusion: $\tilde{\mathbb{V}}(0,x_0) \subset \mathbb{V}(0,x_0)$. Fix arbitrary $y_0 \in \tilde{\mathbb{V}}(0,x_0)$ and $\varepsilon > 0$. By definition of $\tilde{\mathbb{V}}(0,x_0)$ there exist $\alpha = \alpha^\varepsilon \in \mathcal{A}_0$ and $\phi = \phi^\varepsilon \in \mathbb{L}^2(\mathcal{F}_\tau, \mathbb{R}^m)$ such that $|y_0 - Y^{\tau,\phi;\alpha}_0| \le \varepsilon$ and $\mathbb{P}(E) = 1$, where

$$E := \big\{\omega \in \Omega : \phi(\omega) \in \mathbb{V}(\tau(\omega), X^\alpha_\tau(\omega))\big\}.$$

Our goal is to construct an $\hat{\alpha} \in \mathcal{A}_0$ such that

$$\big|Y^{\hat{\alpha}}_0 - Y^{\tau,\phi;\alpha}_0\big| \le C\varepsilon. \tag{4.18}$$

Then $|y_0 - Y^{\hat{\alpha}}_0| \le \varepsilon + C\varepsilon$. Since $Y^{\hat{\alpha}}_0 \in \mathbb{V}(0,x_0)$ by definition, then $y_0 \in \mathbb{V}(0,x_0)$.

We construct $\hat{\alpha}$ by utilizing the desired regularities as in the standard literature. First let $0 = t_0 < \cdots < t_n = T$ be a partition such that $t_i - t_{i-1} \le \varepsilon^2$, $i = 1, \cdots, n$, and let $\{O^m_j\}_{j\ge1}$ be a partition of $\mathbb{R}^m$ and $\{O^d_k\}_{k\ge1}$ a partition of $\mathbb{R}^d$ such that the diameter of each O^m_j and O^d_k is less than ε. We now denote

$$\begin{aligned}
&E^\tau_i := \{t_{i-1} < \tau \le t_i\}, \quad E^\phi_j := \{\phi \in O^m_j\}, \quad E^\alpha_k := \{X^\alpha_\tau \in O^d_k\},\\
&\tau_\varepsilon := \sum_{i=1}^n t_i \mathbf{1}_{E^\tau_i}, \quad E_\theta := E \cap E^\tau_i \cap E^\phi_j \cap E^\alpha_k, \text{ where } \theta = (i,j,k).
\end{aligned} \tag{4.19}$$

For any $\theta = (i,j,k)$ such that $E_\theta \neq \emptyset$, choose $\omega^\theta \in E_\theta$ such that

$$\mathbb{P}\big(\{\tau > t_\theta\} \cap E_\theta\big) \le \varepsilon^2 \mathbb{P}(E_\theta), \quad \text{where} \quad t_\theta := \tau(\omega^\theta), \quad x_\theta := X^\alpha_{t_\theta}(\omega^\theta). \tag{4.20}$$

Moreover, since $\phi(\omega^\theta) \in \mathbb{V}(t_\theta, x_\theta)$, choose $\alpha^\theta \in \mathcal{A}_{t_\theta}$ such that

$$\big|\phi(\omega^\theta) - Y^{t_\theta,x_\theta,\alpha^\theta}_{t_\theta}\big| \le \varepsilon. \tag{4.21}$$

We then construct $\hat{\alpha} \in \mathcal{A}_0$ by: denoting $\omega^t_s := \omega_s - \omega_t$, $0 \le t \le s \le T$,

$$\hat{\alpha}_t(\omega) := \alpha_t(\omega)\mathbf{1}_{[0,\tau_\varepsilon(\omega))}(t) + \mathbf{1}_{[\tau_\varepsilon(\omega),T]}(t)\Big[\sum_\theta \mathbf{1}_{E_\theta}(\omega)\alpha^\theta_t(\omega^{t_\theta}) + a_0\mathbf{1}_{E^c}\Big],$$

where the summations are over all $\theta = (i,j,k)$ with $i = 1,\cdots,n$ and $j,k \ge 1$, and $a_0 \in A$ is an arbitrary value.

Step 3. We now verify (4.18). First, for any $\theta = (i,j,k)$ such that $E_\theta \neq \emptyset$, a.s. on E_θ we have $\tau_\varepsilon = t_i \ge t_\theta$ and, denoting $(X^\theta, Y^\theta, Z^\theta) := (X^{t_\theta,x_\theta,\alpha^\theta}, Y^{t_\theta,x_\theta,\alpha^\theta}, Z^{t_\theta,x_\theta,\alpha^\theta})$,

$$X^{\hat{\alpha}}_t = X^\alpha_{t_\theta} + \int_{t_\theta}^t b(s, X^{\hat{\alpha}}_s, \alpha_s)ds + \int_{t_\theta}^t \sigma(s, X^{\hat{\alpha}}_s, \alpha_s)dB_s, \quad t \in [t_\theta, t_i],$$

$$X^{\hat{\alpha}}_t = X^{\hat{\alpha}}_{t_i} + \int_{t_i}^t b(s, X^{\hat{\alpha}}_s, \alpha^\theta_s(B^{t_\theta}))ds + \int_{t_i}^t \sigma(s, X^{\hat{\alpha}}_s, \alpha^\theta_s(B^{t_\theta}))dB_s, \quad t \in [t_i, T],$$

$$X^\theta_t = x_\theta + \int_{t_\theta}^t b(s, X^\theta_s, \alpha^\theta_s(B^{t_\theta}))ds + \int_{t_\theta}^t \sigma(s, X^\theta_s, \alpha^\theta_s(B^{t_\theta}))dB_s, \quad t \in [t_\theta, T].$$

Since b, σ are bounded, and $|X^\alpha_{t_\theta} - x_\theta| \le \varepsilon$, $t_i - t_\theta \le \varepsilon^2$, by standard SDE estimates we get

$$\mathbb{E}_{\mathcal{F}_{t_\theta}}\Big[\sup_{t_\theta \le t \le t_i} |X^{\hat{\alpha}}_t - X^\theta_t|^2\Big] \le C\varepsilon^2, \text{ and then } \mathbb{E}_{\mathcal{F}_{t_\theta}}\Big[\sup_{t_i \le t \le T} |X^{\hat{\alpha}}_t - X^\theta_t|^2\Big] \le C\varepsilon^2. \tag{4.22}$$

Similarly, note that

$$Y^{\hat{\alpha}}_t = g(X^{\hat{\alpha}}_T) + \int_t^T f(s, X^{\hat{\alpha}}_s, Y^{\hat{\alpha}}_s, Z^{\hat{\alpha}}_s, \alpha^\theta_s(B^{t_\theta}))ds - \int_t^T Z^{\hat{\alpha}}_s dB_s, \quad t \in [t_i, T],$$

$$Y^{\hat{\alpha}}_t = Y^{\hat{\alpha}}_{t_i} + \int_t^{t_i} f(s, X^{\hat{\alpha}}_s, Y^{\hat{\alpha}}_s, Z^{\hat{\alpha}}_s, \alpha_s)ds - \int_t^{t_i} Z^{\hat{\alpha}}_s dB_s, \quad t \in [t_\theta, t_i],$$

$$Y^\theta_t = g(X^\theta_T) + \int_t^T f(s, X^\theta_s, Y^\theta_s, Z^\theta_s, \alpha^\theta_s(B^{t_\theta}))ds - \int_t^T Z^\theta_s dB_s, \quad t \in [t_\theta, T].$$

Then, by (4.22) and standard BSDE estimates we have

$$\mathbb{E}_{\mathcal{F}_{t_\theta}}\Big[\sup_{t_i \le t \le T} |Y^{\hat{\alpha}}_t - Y^\theta_t|^2\Big] \le C\varepsilon^2, \quad \text{and then} \quad \mathbb{E}_{\mathcal{F}_{t_\theta}}\Big[\sup_{t_\theta \le t \le t_i} |Y^{\hat{\alpha}}_t - Y^\theta_t|^2\Big] \le C\varepsilon^2.$$

In particular, this implies that

$$|Y^{\hat{\alpha}}_{t_\theta} - Y^\theta_{t_\theta}| \le C\varepsilon, \quad \text{a.s. on } E_\theta. \tag{4.23}$$

By Assumption 4.4, one can easily see that Y^θ, $Y^{\hat{\alpha}}$ are bounded. Consider the BSDE:

$$\tilde{Y}^\theta_t = Y^\theta_{t_\theta} + \int_t^{t_\theta} f(s, x_\theta, \tilde{Y}^\theta_s, \tilde{Z}^\theta_s, a_0)ds - \int_t^{t_\theta} \tilde{Z}^\theta_s dB_s, \quad t \in [t_{i-1}, t_\theta]. \tag{4.24}$$

Note that $Y^\theta_{t_\theta}$ is deterministic, then so is $\tilde{Y}^\theta_t$ and thus $\tilde{Z}^\theta_t = 0$. Therefore,

$$\sup_{t_{i-1} \le t \le t_\theta} |\tilde{Y}^\theta_t - Y^\theta_{t_\theta}| \le \int_{t_{i-1}}^{t_\theta} |f(s, x_\theta, \tilde{Y}^\theta_s, 0, a_0)|ds \le C(t_i - t_{i-1}) \le C\varepsilon^2.$$

Moreover, note that $E_\theta \in \mathcal{F}_\tau$ and

$$Y_t^{\hat\alpha} = Y_{t_\theta}^{\hat\alpha} + \int_t^{t_\theta} f(s, X_s^{\hat\alpha}, Y_s^{\hat\alpha}, Z_s^{\hat\alpha}, \alpha_s)ds - \int_t^{t_\theta} Z_s^{\hat\alpha} dB_s, \quad t \in [t_{i-1}, t_\theta].$$

Compare this with (4.24), by (4.23) and standard BSDE estimate we have

$$\mathbb{E}_{\mathcal{F}_\tau}\Big[\sup_{\tau \le t \le t_\theta} |Y_t^{\hat\alpha} - \tilde Y_t^\theta|^2\Big] \le C\varepsilon^2, \quad \text{a.s. on } \{\tau \le t_\theta\} \cap E_\theta.$$

Then

$$|Y_\tau^{\hat\alpha} - Y_{t_\theta}^\theta| \le |Y_\tau^{\hat\alpha} - \tilde Y_\tau^\theta| + |\tilde Y_\tau^\theta - Y_{t_\theta}^\theta| \le C\varepsilon, \quad \text{a.s. on } \{\tau \le t_\theta\} \cap E_\theta.$$

This, together with (4.19) and (4.21), implies that, for a.e. $\omega \in \{\tau \le t_\theta\} \cap E_\theta$,

$$|Y_\tau^{\hat\alpha}(\omega) - \phi(\omega)| \le |Y_\tau^{\hat\alpha}(\omega) - Y_{t_\theta}^\theta| + |Y_{t_\theta}^\theta - \phi(\omega^\theta)| + |\phi(\omega^\theta) - \phi(\omega)| \le C\varepsilon.$$

Note that $\{E_\theta\}$ form a partition of E and $\mathbb{P}(E) = 1$, then by (4.20) we have

$$\mathbb{P}\big(|Y_\tau^{\hat\alpha} - \phi| > C\varepsilon\big) \le \sum_\theta \mathbb{P}\big(\{\tau > t_\theta\} \cap E_\theta\big) \le \sum_\theta \varepsilon^2 \mathbb{P}(E_\theta) = \varepsilon^2.$$

Note again that $Y^{\hat\alpha}$ and ϕ are bounded. Then

$$\mathbb{E}\big[|Y_\tau^{\hat\alpha} - \phi|^2\big] \le C\varepsilon^2 + C\mathbb{P}\big(|Y_\tau^{\hat\alpha} - \phi| > C\varepsilon\big) \le C\varepsilon^2. \tag{4.25}$$

Finally, note that $Y_t^{\hat\alpha} = Y_t^{\tau, Y_\tau^{\hat\alpha}; \alpha}$, $0 \le t \le \tau$. Then, by (4.25) and standard BSDE estimates we have

$$\mathbb{E}\Big[\sup_{0 \le t \le \tau} |Y_t^{\hat\alpha} - Y_t^{\tau,\phi;\alpha}|^2\Big] \le C\varepsilon^2.$$

This clearly implies (4.18) and hence completes the proof. ■

5 Set-valued HJB equations

We now derive the set-valued HJB equation for $\mathbb{V}$, as usual by combining the DPP (4.16) and the Itô formula. Given (t, x, y), α, and tangent fields ξ, ζ, consider the system starting from (t, x, y). First, by Itô's formula we have $\Upsilon_{t+\delta}^{\xi,\zeta} \in \mathbb{V}_b(t+\delta, X_{t+\delta}^\alpha)$; Next, setting $\tau := t+\delta$ and $\phi := \Upsilon_{t+\delta}^{\xi,\zeta}$ in the DPP (4.16), we have $Y_t^{\phi,\alpha} \in \mathbb{V}(t, x)$. Here we are using simplified notations. Then it follows from straightforward calculation that

$$Y_t^{\phi,\alpha} - y = \mathbb{E}\Big[\int_t^T \big[\partial_t \mathbb{V} + h_{\mathbb{V}} + \xi\big] ds\Big],$$

for some appropriate $h_{\mathbb{V}}$ which will be defined later. Unlike standard optimization problems, here $Y_t^{\phi,\alpha} - y$ is a vector, thus one cannot obtain a desired 0 by optimizing over all α (and/or all ϕ). Our remedy relies on the following crucial observation[7]: since $y \in \mathbb{V}_b(t, x)$,

$$\lim_{\delta \to 0} \frac{1}{\delta} \sup_{y' \in \mathbb{V}(t,x), |y'-y| \le \delta} \mathbf{n}_{\mathbb{V}}(t, x, y) \cdot (y' - y) \le 0. \tag{5.1}$$

[7] When $\mathbb{V}(t, x)$ is convex, we have the stronger inequality: $\sup_{y' \in \mathbb{V}(t,x)} (y' - y) \cdot n_{\mathbb{V}}(t, x, y) \le 0$.

Moreover, by optimizing over all α and all ϕ, where the latter effectively means to run over all ξ and ζ, we may obtain 0 at above. Thus formally we have: noting further that $\mathbf{n}_{\mathbb{V}} \cdot \xi = 0$,

$$0 = \sup_{\alpha,\xi,\zeta} \lim_{\delta\to 0} \frac{1}{\delta}\, \mathbf{n}_{\mathbb{V}} \cdot [Y_t^{\phi,\alpha} - y] = \sup_{\alpha,\xi,\zeta} \mathbf{n}_{\mathbb{V}} \cdot \big[\partial_t \mathbb{V} + h_{\mathbb{V}} + \xi\big] = \sup_{\alpha,\zeta} \mathbf{n}_{\mathbb{V}} \cdot \big[\partial_t \mathbb{V} + h_{\mathbb{V}}\big].$$

That is, by focusing on the boundary points, we are actually considering scalarized optimization problems along the direction $\mathbf{n}_{\mathbb{V}}$.

We are now ready to present the set-valued HJB equation precisely. For $(t,x,y) \in \mathbb{G}_{\mathbb{V}}$, $z \in \mathbb{R}^{m\times d}$, $\gamma \in (\mathbb{R}^{d\times d})^m$, $a \in A$, $\zeta \in (\mathbb{T}_{\mathbb{V}}(t,x,y))^d$, inspired by (3.2) in the Itô formula and (4.16) we introduce the Hamiltonian function $H_{\mathbb{V}}$:

$$\begin{aligned}
&\mathcal{K}_{\mathbb{V}}(t,x,y,a,\zeta) := \mathrm{tr}\Big(\zeta^\top \partial_x \mathbf{n}_{\mathbb{V}}(t,x,y)\sigma(t,x,a) + \frac{1}{2}\zeta^\top \partial_y \mathbf{n}_{\mathbb{V}}(t,x,y)\zeta\Big)\mathbf{n}_{\mathbb{V}}(t,x,y);\\
&h^0_{\mathbb{V}}(t,x,y,z,\gamma,a,\zeta) := zb(t,x,a) + \frac{1}{2}\mathrm{tr}\,(\sigma^\top \gamma\sigma(t,x,a)) - \mathcal{K}_{\mathbb{V}}(t,x,y,a,\zeta);\\
&h_{\mathbb{V}}(t,x,y,z,\gamma,a,\zeta) := h^0_{\mathbb{V}}(t,x,y,z,\gamma,a,\zeta) + f\big(t,x,y,z\sigma(t,x,a)+\zeta,a\big);\\
&H_{\mathbb{V}}(t,x,y,z,\gamma) := \sup_{a\in A,\zeta\in(\mathbb{T}_{\mathbb{V}}(t,x,y))^d} \mathbf{n}_{\mathbb{V}}(t,x,y)\cdot h_{\mathbb{V}}(t,x,y,z,\gamma,a,\zeta);
\end{aligned} \tag{5.2}$$

We note that, in the scalar case: $m=1$, then $\mathbf{n} = 1$ or -1, and $H_{\mathbb{V}}$ reduces to standard Hamiltonian functions. See Remark 5.1 (i) below.

Our set-valued HJB equation takes the form:

$$\begin{aligned}
&\mathcal{L}\mathbb{V}(t,x,y) = 0,\ \forall (t,x,y)\in \mathbb{G}_{\mathbb{V}}, \quad \text{where}\\
&\mathcal{L}\mathbb{V}(t,x,y) := \partial_t \mathbb{V}(t,x,y)\cdot \mathbf{n}_{\mathbb{V}}(t,x,y) + H_{\mathbb{V}}\big(t,x,y,\partial_x\mathbb{V}(t,x,y),\partial_{xx}\mathbb{V}(t,x,y)\big).
\end{aligned} \tag{5.3}$$

Equivalently, by (2.4), (2.20), (2.22), (2.23), we may rewrite the above equation:

$$\begin{aligned}
\nabla_t \mathbf{r}_{\mathbb{V}} + \inf_{a\in A,\zeta\in(\mathbb{T}_{\mathbb{V}}(t,x,y))^d} \Big[&\nabla_x \mathbf{r}_{\mathbb{V}}\cdot b + \frac{1}{2}\mathrm{tr}\,\big(\sigma^\top \nabla_{xx}\mathbf{r}_{\mathbb{V}}\sigma + 2\zeta^\top \nabla_{xy}\mathbf{r}_{\mathbb{V}}\sigma + \zeta^\top\nabla_{yy}\mathbf{r}_{\mathbb{V}}\zeta\big)\\
&-\nabla_y \mathbf{r}_{\mathbb{V}}\cdot f(t,x,y,-\nabla_y\mathbf{r}_{\mathbb{V}}(\nabla_x\mathbf{r}_{\mathbb{V}})^\top\sigma+\zeta,a)\Big] = 0, \quad (t,x,y)\in\mathbb{G}_{\mathbb{V}}.
\end{aligned} \tag{5.4}$$

Remark 5.1 *(i) For the scalar case as in Remark 4.1, by (4.3) we have*

$$\begin{aligned}
\mathbb{V}_b(t,x) = \{\underline{v}(t,x),\overline{v}(t,x)\},\quad \mathbf{n}(t,x,\underline{v}(t,x)) = -1,\quad \mathbf{n}(t,x,\overline{v}(t,x)) = 1,\\
\mathbb{T}_{\mathbb{V}}(t,x,\underline{v}(t,x)) = \mathbb{T}_{\mathbb{V}}(t,x,\overline{v}(t,x)) = \{0\}.
\end{aligned}$$

In the neighborhood of $y=\underline{v}(t,x)$, we have $\mathbf{r}_{\mathbb{V}}(t,x,y) = \underline{v}(t,x)-y$ and $\mathbf{n}(t,x,\underline{v}(t,x)) = -1$. Then (5.4) reduces to the standard HJB equation for $\underline{v}$:

$$\nabla_t\underline{v} + \inf_{a\in A}\Big[\nabla_x\underline{v}\cdot b + \frac{1}{2}\mathrm{tr}\,\big(\sigma^\top\nabla_{xx}\underline{v}\sigma\big) + f\big(t,x,\underline{v},(\nabla_x\underline{v})^\top\sigma,a\big)\Big] = 0.$$

Similarly, in the neighborhood of $y=\overline{v}(t,x)$, we have $\mathbf{r}_{\mathbb{V}}(t,x,y) = y-\overline{v}(t,x)$ and $\mathbf{n}(t,x,\overline{v}(t,x)) = 1$. Then (5.4) reduces to the standard HJB equation for $\overline{v}$:

$$\nabla_t\overline{v} + \sup_{a\in A}\Big[\nabla_x\overline{v}\cdot b + \frac{1}{2}\mathrm{tr}\,\big(\sigma^\top\nabla_{xx}\overline{v}\sigma\big) + f\big(t,x,\overline{v},(\nabla_x\overline{v})^\top\sigma,a\big)\Big] = 0.$$

(ii) Although $\mathbf{r}_{\mathbb{V}}$ is scalar, we emphasize that (5.4) holds true only on $\mathbb{G}_{\mathbb{V}}$, and the set $\mathbb{G}_{\mathbb{V}}$ is in turn determined by the solution $\mathbf{r}_{\mathbb{V}}$. So (5.4) is actually quite involved, and we can not apply the standard PDE theory on it.

(iii) It is clear that $\mathbb{V}(T,x) = \{g(x)\}$ is degenerate, so we do not require the smoothness of $\mathbb{V}$ at T. See Definition 2.10 and the paragraph above it. ■

Remark 5.2 *Note that $\mathcal{K}_{\mathbb{V}}$ relies on ζ quadratically, and the space of ζ is typically unbounded. This is associated with the fact that the PDE (A.7) in the proof of Proposition 4.5 is quadratic in z, and is related to the convexity of $\mathbb{V}$ in Proposition 4.5.*

(i) In the scalar case: $m = 1$, we have $\mathbb{T}_{\mathbb{V}}(t,x,y) = \{0\}$ for all $(t,x,y) \in \mathbb{G}_{\mathbb{V}}$. Then this issue is trivial. Indeed, in this case the set-valued PDE reduces back to the standard HJB equations, as we saw in Remark 5.1 (i).

(ii) For $m \geq 2$, recall Remark 2.8 (iii) that $\partial_y \mathbf{n}_{\mathbb{V}} = \nabla_{yy}\mathbf{r}_{\mathbb{V}}$ is symmetric and 0 is an eigenvalue with eigenvector $\mathbf{n}$. At any fixed $(t,x,y) \in \mathbb{G}_{\mathbb{V}}$, let $\lambda_1 \leq \cdots \leq \lambda_{m-1}$ be the other eigenvalues. When f has linear growth in z, which is implied by the Lipschitz continuity, and $\lambda_1 > 0$, then clearly $H_{\mathbb{V}} < \infty$. In this case $\mathbb{V}(t,x)$ is strictly convex.

(iii) When f has linear growth in z, one may easily derive from $H_{\mathbb{V}} < \infty$ that $\lambda_1 \geq 0$. Thus, our classical solution $\mathbb{V}$ has to be convex.

(iv) When f has quadratic growth, this convexity is not required, as we saw in Example 4.7. We shall remark though such quadratic growth violates Assumption 4.4, which is assumed for technical reasons and can be weakened. ■

Remark 5.3 *Consider Remarks 2.3 and 2.4 again with $x \in \mathbb{R}$. One can show that, roughly speaking, $\mathbb{V}$ satisfies the fundamental theorem (2.14) if and only if $\partial_x \mathbb{V} := -\nabla_x \mathbf{r}_{\mathbb{V}} \mathbf{n}_{\mathbb{V}} + \zeta_0$ for some appropriate field ζ_0 taking values in the tangent space $\mathbb{T}_{\mathbb{V}}$. So, by restricting $\partial_x \mathbb{V}(x,y)$ to be parallel to $\mathbf{n}_{\mathbb{V}}(t,x,y)$, we see that $\partial_x \mathbb{V} := -\nabla_x \mathbf{r}_{\mathbb{V}} \mathbf{n}_{\mathbb{V}}$ is the unique choice. If we relax this requirement, then we may define $\partial_x \mathbb{V}$ alternatively by choosing certain ζ_0, denoted as $\partial_x^{\zeta_0}\mathbb{V}$. This choice of ζ_0 will not affect our main results though. Indeed, one can show that $\mathbf{n}^\top \partial_{xx}^{\zeta_0}\mathbb{V} = \mathbf{n}^\top \partial_{xx}\mathbb{V} - \zeta_0^\top \partial_x \mathbf{n} - \partial_x \mathbf{n}^\top \zeta_0 - \zeta_0^\top \partial_y \mathbf{n} \zeta_0$. Thus, in the set-valued Itô formula, the presence of ζ_0 will only lead to a different choice of (ξ, ζ) in (3.2). Then, by taking the supremum over ζ, the Hamiltonian $H_{\mathbb{V}}$ in (5.2) will remain the same. Consequently, the set-valued PDE (5.3) also does not depend on ζ_0.*

Finally, due to the local geodesic property in Proposition 2.5, throughout the paper we fix $\zeta_0 \equiv 0$ so that $\partial_x \mathbb{V}$ is align with $\mathbf{n}_{\mathbb{V}}$. ■

We now turn to the wellposedness of (5.3). We first define classical solutions rigorously.

Definition 5.4 *(i) Let $C_0^{1,2}([0,T) \times \mathbb{R}^d; \mathcal{D}_2^m)$ denote the set of $\mathbb{V} \in C^{1,2}([0,T) \times \mathbb{R}^d; \mathcal{D}_2^m)$ such that, for any $T_0 < T$, the eigenvalue λ_1 of $\partial_y \mathbf{n}(t,x,y)$ in Remark 5.2 (ii) has a lower bound $c_{T_0} > 0$ for all $(t,x) \in [0,T_0] \times \mathbb{R}^d$, $y \in \mathbb{V}_b(t,x)$. That is,*

$$\operatorname{tr}\left(\zeta^\top \partial_y \mathbf{n}_{\mathbb{V}}(t,x,y)\zeta\right) \geq c_{T_0}|\zeta|^2 \quad \forall (t,x) \in [0,T_0] \times \mathbb{R}^d, y \in \mathbb{V}_b(t,x), \zeta \in \mathbb{T}_{\mathbb{V}}(t,x,y). \tag{5.5}$$

This implies that $H_{\mathbb{V}}(\cdot, \partial_x \mathbb{V}, \partial_{xx}\mathbb{V})$ is finite and uniformly continuous whenever $t \leq T_0$.

(ii) We say $\mathbb{V} \in C_0^{1,2}([0,T) \times \mathbb{R}^d; \mathcal{D}_2^m)$ is a classical solution to (5.3) if it satisfies (5.3) for all $(t,x) \in [0,T) \times \mathbb{R}^d$ and $y \in \mathbb{V}_b(t,x)$.

We shall provide an example in Example 6.4 below. We next establish a crucial estimate, whose proof is postponed to Appendix.

Lemma 5.5 *Let Assumption 4.4 hold and $\mathbb{V}$ be defined by (4.2). Assume $\mathbb{V} \in C_0^{1,2}([0,T)\times \mathbb{R}^d; \mathcal{D}_2^m)$. Fix $T_0 < T$ and $x_0 \in \mathbb{R}^d$. Let $\varepsilon, \delta > 0$ and $\alpha \in \mathcal{A}_0$ be such that $|\mathbf{r}_{\mathbb{V}}(0, x_0, Y_0^\alpha)| \le \varepsilon$, where $(X^\alpha, Y^\alpha, Z^\alpha) = (X^{0,x_0,\alpha}, Y^{0,x_0,\alpha}, Z^{0,x_0,\alpha})$ are defined by (4.1). Then there exists a constant C_{T_0}, which may depend on T_0 but not on $\varepsilon, \delta, \alpha$, such that*

$$\mathbb{P}\Big(\sup_{0\le t\le T_0} |\mathbf{r}_{\mathbb{V}}(t, X_t^\alpha, Y_t^\alpha)| > \delta\Big) \le C_{T_0}\sqrt{\frac{\varepsilon}{\delta}}. \tag{5.6}$$

In particular, if $Y_0^\alpha \in \mathbb{V}_b(0, x_0)$, then $Y_t^\alpha \in \mathbb{V}_b(t, X_t^\alpha)$, $0 \le t \le T$, a.s.

The main result of this section is the following theorem.

Theorem 5.6 *Let Assumption 4.4 hold and $\mathbb{V}$ be defined by (4.2). Assume $\mathbb{V} \in C_0^{1,2}([0,T)\times \mathbb{R}^d; \mathcal{D}_2^m)$. Then $\mathbb{V}$ is a classical solution of (5.3) with terminal condition $\mathbb{V}(T,x) = \{g(x)\}$.*

Proof It is clear that $\mathbb{V}(T,x) = \{g(x)\}$. Without loss of generality, we shall verify (5.3) only at a fixed $(0, x_0, y_0) \in \mathbb{G}_{\mathbb{V}}$, and for notational simplicity, in this proof we omit the supscripts ^{0,x_0,y_0} and the subscript ${}_{\mathbb{V}}$ in $\mathbf{r}, \mathbf{n}, \pi$. We proceed in two steps, following the ideas in the standard stochastic control literature, but by adopting the scalarization with $\mathbf{n}$ as explained in the beginning of this section.

Step 1. We first show that $\mathcal{L}\mathbb{V}(0, x_0, y_0) \le 0$. For this purpose, we fix an arbitrary $a \in A$ and let $X^a := X^{0,x_0,\alpha}$ be defined by (4.1) for constant control process $\alpha \equiv a$. Moreover, we fix arbitrary $\xi, \zeta_i : [0,T]\times\Omega\times\mathbb{R}^m \to \mathbb{R}^m$, $i = 1, \cdots, d$, which are $\mathbb{F}$-progressively measurable, bounded, continuous in t, uniformly Lipschitz continuous in y, and $\xi(t,\omega,y), \zeta_i(t,\omega,y) \in \mathbb{T}_{\mathbb{V}}(t, X_t^a(\omega), y)$ for all $y \in \mathbb{V}_b(t, X_t^a(\omega))$, for $dt \times d\mathbb{P}$-a.e. (t,ω). Denote $\zeta = (\zeta_1, \cdots, \zeta_d)$ and consider the SDE:

$$\begin{aligned}\Upsilon_t^{a,\xi,\zeta} = y_0 + &\int_0^t \big[\partial_t \mathbb{V} + h^0_{\mathbb{V}}(\cdot, \partial_x\mathbb{V}, \partial_{xx}\mathbb{V}, a, \zeta_s) + \xi\big](s, X_s^a, \Upsilon_s^{a,\xi,\zeta})ds \\ &+ \int_0^t \big[\partial_x\mathbb{V}\sigma(\cdot, a) + \zeta\big](s, X_s^a, \Upsilon_s^{a,\xi,\zeta})dB_s.\end{aligned} \tag{5.7}$$

Applying the Itô formula Theorem 3.1 we have $\Upsilon_t^{a,\xi,\zeta} \in \mathbb{V}_b(t, X_t^a)$, for all $0 \le t < T$.

Now for any $\delta > 0$ small, consider the BSDE (4.15) with terminal condition $(\delta, \Upsilon_\delta^{a,\xi,\zeta})$:

$$Y_t^{\delta,a,\xi,\zeta} = \Upsilon_\delta^{a,\xi,\zeta} + \int_t^\delta f(s, X_s^a, Y_s^{\delta,a,\xi,\zeta}, Z_s^{\delta,a,\xi,\zeta}, a)ds - \int_t^\delta Z_s^{\delta,a,\xi,\zeta} dB_s, 0 \le t \le \delta. \tag{5.8}$$

Since $\Upsilon_\delta^{a,\xi,\zeta} \in \mathbb{V}_b(\delta, X_\delta^a)$, by the DPP (4.16) we see that $Y_0^{\delta,a,\xi,\zeta} \in \mathbb{V}(0, x_0)$. Denote

$$\Delta Y_t^\delta := Y_t^{\delta,a,\xi,\zeta} - \Upsilon_t^{a,\xi,\zeta}, \quad \Delta Z_t^\delta := Z_t^{\delta,a,\xi,\zeta} - \big[\partial_x\mathbb{V}\sigma(\cdot, a) + \zeta\big](t, X_t^a, \Upsilon_t^{a,\xi,\zeta}).$$

Then, by (5.7) and (5.8) we have

$$
\begin{aligned}
\Delta Y_t^\delta &= \int_t^\delta \Big[\partial_t \mathbb{V} + h_{\mathbb{V}}(\cdot, \partial_x \mathbb{V}, \partial_{xx}\mathbb{V}, a, \zeta) + \xi\Big](s, X_s^a, \Upsilon_s^{a,\xi,\zeta})ds - \int_t^\delta \Delta Z_s^\delta dB_s \\
&\quad + \int_t^\delta \Big[f(s, X_s^a, Y_s^{\delta,a,\xi,\zeta}, Z_s^{\delta,a,\xi,\zeta}, a) - f(s, X_s^a, Y_s^{\delta,a,\xi,\zeta} - \Delta Y_s^\delta, Z_s^{\delta,a,\xi,\zeta} - \Delta Z_s^\delta, a)\Big]ds \\
&= \int_t^\delta \Big[\partial_t \mathbb{V} + h_{\mathbb{V}}(\cdot, \partial_x \mathbb{V}, \partial_{xx}\mathbb{V}, a, \zeta) + \xi\Big](s, X_s^a, \Upsilon_s^{a,\xi,\zeta})ds - \int_t^\delta \Delta Z_s^\delta dB_s \\
&\quad + \int_t^\delta \Big[\tilde b_s \Delta Y_s^\delta + \tilde\sigma_s \Delta Z_s^\delta\Big]ds,
\end{aligned}
$$

where $\tilde b, \tilde\sigma$ are appropriate $\mathbb{F}$-progressively measurable bounded processes. Then, for the $\tilde\Gamma$ defined by (3.9), we have

$$
\begin{aligned}
\tilde\Gamma_t \Delta Y_t^\delta &= \int_t^\delta \tilde\Gamma_s \Big[\partial_t \mathbb{V} + h_{\mathbb{V}}(\cdot, \partial_x \mathbb{V}, \partial_{xx}\mathbb{V}, a, \zeta) + \xi\Big](s, X_s^a, \Upsilon_s^{a,\xi,\zeta})ds \\
&\quad - \int_t^\delta \tilde\Gamma_s \big[\Delta Z_s^\delta + \Delta Y_t^\delta \tilde\sigma\big] \cdot dB_s.
\end{aligned}
$$

In particular,

$$
\Delta Y_0^\delta = \mathbb{E}\Big[\int_0^\delta \tilde\Gamma_s \big[\partial_t \mathbb{V} + h_{\mathbb{V}}(\cdot, \partial_x \mathbb{V}, \partial_{xx}\mathbb{V}, a, \zeta) + \xi\big](s, X_s^a, \Upsilon_s^{a,\xi,\zeta})ds\Big]. \tag{5.9}
$$

Given our conditions, it is clear that $|\Delta Y_0^\delta| \le C\delta$, which implies that $\lim_{\delta\to 0} Y_0^{\delta,\alpha,\xi,\zeta} = y_0$. Since $Y_0^{\delta,a,\xi,\zeta} \in \mathbb{V}(0, x_0)$ and $y_0 \in \mathbb{V}_b(0, x_0)$, then it follows from (5.1)[8] that:

$$
\overline{\lim_{\delta\to 0}} \frac{1}{\delta}\big[\mathbf{n}(0, x_0, y_0) \cdot \Delta Y_0^\delta\big] \le 0. \tag{5.10}
$$

Now by (5.9) and the desired continuity of $\tilde\Gamma_s, X_s^a, \Upsilon_s^{a,\xi,\zeta}, \xi_s$ in s as well as the desired regularity of all the involved functions in (x, y), we have

$$
\begin{aligned}
0 &\ge \overline{\lim_{\delta\to 0}} \frac{1}{\delta}\mathbb{E}\Big[\mathbf{n}(0, x_0, y_0) \cdot \int_0^\delta \tilde\Gamma_s \Big[\partial_t \mathbb{V} + h_{\mathbb{V}}(\cdot, \partial_x \mathbb{V}, \partial_{xx}\mathbb{V}, a, \zeta) + \xi\Big](s, X_s^a, \Upsilon_s^{a,\xi,\zeta})ds\Big] \\
&= \overline{\lim_{\delta\to 0}} \frac{1}{\delta}\Big[\mathbf{n}(0, x_0, y_0) \cdot \int_0^\delta \Big[\partial_t \mathbb{V} + h_{\mathbb{V}}(\cdot, \partial_x \mathbb{V}, \partial_{xx}\mathbb{V}, a, \zeta) + \xi\Big](0, x_0, y_0)ds\Big] \\
&= \mathbf{n}(0, x_0, y_0) \cdot \Big[\partial_t \mathbb{V} + h_{\mathbb{V}}(\cdot, \partial_x \mathbb{V}, \partial_{xx}\mathbb{V}, a, \zeta)\Big](0, x_0, y_0),
\end{aligned}
$$

where the last equality is due to the assumption $\xi \in \mathbb{T}_{\mathbb{V}}$. Now by the arbitrariness of a, ζ we obtain $\mathcal{L}\mathbb{V}(0, x_0, y_0) \le 0$.

Step 2. We next show that $\mathcal{L}\mathbb{V}(0, x_0, y_0) \ge 0$. This step is more involved, because one needs to construct approximately "optimal" α, ζ, ξ to make the inequality (5.1) essentially

[8] We emphasize that here we are using (5.1), rather than the stronger inequality in Footnote 7 for convex sets. That is, although our set values turn out to be convex, the arguments here do not rely on this convexity.

reaching 0. For this purpose, fix $T_0 < T$, and throughout this proof, the generic constant C may depend on T_0. Since $\mathbb{V} \in C^{1,2}([0,T_0]\times\mathbb{R}^d;\mathcal{D}_2^m)$, there exists $\varepsilon_0 > 0$ such that $\mathbf{r} \in C^{1,2}(O_{\varepsilon_0}^{T_0}(\mathbb{G}_{\mathbb{V}});\mathbb{R})$, where $O_{\varepsilon_0}^{T_0}(\mathbb{G}_{\mathbb{V}}) := \{(t,x,y) \in [0,T_0]\times\mathbb{R}^d\times\mathbb{R}^m : |\mathbf{r}(t,x,y)| \le \varepsilon_0\}$. Fix a sufficiently small constant $\varepsilon > 0$. Since $y_0 \in \mathbb{V}_b(0,x_0)$, there exists $\alpha = \alpha^\varepsilon \in \mathcal{A}_0$ such that

$$\pi(0,x_0,Y_0^\alpha) = y_0 \quad \text{and} \quad |y_0 - Y_0^\alpha| \le \varepsilon^4,$$

where $(X^\alpha, Y^\alpha, Z^\alpha) = (X^{0,x_0,\alpha}, Y^{0,x_0,\alpha}, Z^{0,x_0,\alpha})$ are defined by (4.1). Define[9]

$$\tau := \tau_{\varepsilon,\alpha} := \inf\{t > 0 : |\mathbf{r}(t,X_t^\alpha,Y_t^\alpha)| > \varepsilon^2\} \wedge T_0.$$

By Lemma 5.5 we have

$$\mathbb{P}(\tau < T_0) \le \mathbb{P}\big(\sup_{0\le t\le T_0}|\mathbf{r}(t,X_t^\alpha,Y_t^\alpha)| > \varepsilon^2\big) \le C\sqrt{\frac{\varepsilon^4}{\varepsilon^2}} = C\varepsilon. \tag{5.11}$$

Step 2.1. Introduce two random fields:

$$\begin{aligned}
&\zeta_t(y) := Z_t^\alpha - \partial_x\mathbb{V}\sigma(t,X_t^\alpha,y,\alpha_t),\\
&\xi_t(y) := -\big[\partial_t\mathbb{V} + h_{\mathbb{V}}(\cdot,\partial_x\mathbb{V},\partial_{xx}\mathbb{V},\alpha_t,\zeta_t)\big](t,X_t^\alpha,y).
\end{aligned} \tag{5.12}$$

Then we may rewrite the BSDE for (Y^α, Z^α) forwardly:

$$\begin{aligned}
Y_t^\alpha \;&=\; Y_0^\alpha + \int_0^t \Big[\partial_t\mathbb{V} + h_{\mathbb{V}}^0(\cdot,\partial_x\mathbb{V},\partial_{xx}\mathbb{V},\alpha_s,\zeta) + \xi\Big](s,X_s^\alpha,Y_s^\alpha)ds\\
&\quad + \int_0^t \big[\partial_x\mathbb{V}\sigma(\cdot,\alpha_s) + \zeta\big](s,X_s^\alpha,Y_s^\alpha)dB_s.
\end{aligned}$$

We note that, if $Y_t^\alpha \in \mathbb{V}_b(t,X_t^\alpha)$ and $\zeta_t(Y_t^\alpha), \xi_t(Y_t^\alpha)$ are in the tangent space $\mathbb{T}_{\mathbb{V}}(t,X_t^\alpha,Y_t^\alpha)$, then by the optimality of $H_{\mathbb{V}}$ we have $\mathcal{L}\mathbb{V}(t,X_t^\alpha,Y_t^\alpha) \ge -\mathbf{n}(t,X_t^\alpha,Y_t^\alpha)\cdot\xi_t(Y_t^\alpha) = 0$, which is the desired inequality. In this and the next substeps, we shall prove these properties in approximate sense.

Denote $\pi_t^\alpha := \pi(t,X_t^\alpha,Y_t^\alpha)$. By (3.5) and (3.6), similarly to (3.7) and (3.8) we have,

$$\begin{aligned}
d\mathbf{r}(t,X_t^\alpha,Y_t^\alpha) &= \Big[\mathbf{r}(t,X_t^\alpha,Y_t^\alpha)\tilde{b}_t - \mathbf{n}\cdot\xi(t,X_t^\alpha,\pi_t^\alpha)\Big]dt\\
&\quad + \Big[\mathbf{r}(t,X_t^\alpha,Y_t^\alpha)\tilde{\sigma}_t - \mathbf{n}^\top\zeta(t,X_t^\alpha,\pi_t^\alpha)\Big]dB_t, \quad 0 \le t \le \tau,
\end{aligned} \tag{5.13}$$

where $\tilde{b}, \tilde{\sigma}$ are $\mathbb{F}$-progressively measurable and satisfy: for some constant $C = C_{T_0}$,

$$|\tilde{\sigma}_t| \le C, \quad |\tilde{b}_t| \le C\big[1 + |Z_t^\alpha|\big]. \tag{5.14}$$

Recall the process $\tilde{\Gamma}$ defined in (3.9), we have

$$\mathbf{r}(0,x_0,Y_0^\alpha) - \tilde{\Gamma}_\tau\mathbf{r}(\tau,X_\tau^\alpha,Y_\tau^\alpha) = \int_0^\tau \tilde{\Gamma}_s\mathbf{n}\cdot\xi(s,X_s^\alpha,\pi_s^\alpha)ds + \int_0^\tau \tilde{\Gamma}_s\mathbf{n}^\top\zeta(s,X_s^\alpha,\pi_s^\alpha)dB_s. \tag{5.15}$$

[9]Since we are using the augmented filtration, the τ here is an $\mathbb{F}$-stopping time.

Moreover, by Assumption 4.4 one can easily see that Y^α is bounded, then $\int_0^{\tau\wedge\cdot} Z_s^\alpha dB_s$ is a BMO martingale, and thus there exist $c_0, C > 0$, such that (cf. [30, Section 7.2])

$$\mathbb{E}\Big[\exp\big(c_0\int_0^\tau |Z_s^\alpha|^2 ds\big)\Big] \le C < \infty.$$

In particular, this implies that, for any $p \ge 1$, there exists a constant $C_p > 0$ such that

$$\mathbb{E}\Big[\sup_{0\le t\le\tau}\big[|\tilde\Gamma_t|^p + |\tilde\Gamma_t|^{-p}\big]\Big] \le C_p. \tag{5.16}$$

Applying the standard Itô formula on $|\tilde\Gamma_t \mathbf{r}(t, X_t^\alpha, Y_t^\alpha)|^2$, by (5.13) we have

$$\begin{aligned}
&\mathbb{E}\Big[\int_0^\tau |\tilde\Gamma_s \mathbf{n}^\top \zeta(s, X_s^\alpha, \pi_s^\alpha)|^2 ds\Big]\\
&= \mathbb{E}\Big[|\tilde\Gamma_\tau \mathbf{r}(\tau, X_\tau^\alpha, Y_\tau^\alpha)|^2 - |\mathbf{r}(0, x_0, Y_0^\alpha)|^2 + 2\int_0^\tau \tilde\Gamma_s^2 \mathbf{r}(s, X_s^\alpha, Y_s^\alpha)\mathbf{n}\cdot\xi(s, X_s^\alpha, \pi_s^\alpha)ds\Big]\\
&\le C\varepsilon^4 + C\varepsilon^2\mathbb{E}\Big[\sup_{0\le t\le\tau}|\tilde\Gamma_t|^2\int_0^\tau [1 + |Z_s^\alpha|^2]ds\Big] \le C\varepsilon^2.
\end{aligned} \tag{5.17}$$

Step 2.2. Introduce two processes

$$\begin{aligned}
\hat\zeta_s &:= (\mathbf{I}_m - \mathbf{n}\mathbf{n}^\top)\zeta_s(\pi_s^\alpha) \in \big(\mathbb{T}_{\mathbb{V}}(s, X_s^\alpha, \pi_s^\alpha)\big)^d;\\
\hat\xi_s &:= -\big[\partial_t\mathbb{V} + h_{\mathbb{V}}(\cdot, \partial_x\mathbb{V}, \partial_{xx}\mathbb{V}, \alpha_s, \hat\zeta_s)\big](s, X_s^\alpha, \pi_s^\alpha).
\end{aligned} \tag{5.18}$$

By (5.3), (5.4), and by Step 1, we have

$$0 \le -\mathcal{L}\mathbb{V}(s, X_s^\alpha, \pi_s^\alpha) \le \mathbf{n}(s, X_s^\alpha, \pi_s^\alpha)\cdot\hat\xi_s. \tag{5.19}$$

Then, by taking expectation on both sides of (5.15) we have

$$\begin{aligned}
&-\mathbb{E}\Big[\int_0^\tau \tilde\Gamma_s\mathcal{L}\mathbb{V}(s, X_s^\alpha, \pi_s^\alpha)ds\Big] \le \mathbb{E}\Big[\int_0^\tau \tilde\Gamma_s\mathbf{n}(s, X_s^\alpha, \pi_s^\alpha)\cdot\hat\xi_s ds\Big]\\
&\le \mathbb{E}\Big[\int_0^\tau \tilde\Gamma_s\mathbf{n}(s, X_s^\alpha, \pi_s^\alpha)\cdot\xi_s ds\Big] + \mathbb{E}\Big[\int_0^\tau \tilde\Gamma_s|\xi_s - \hat\xi_s|ds\Big]\\
&= \mathbb{E}\Big[\mathbf{r}(0, x_0, Y_0^\alpha) - \tilde\Gamma_\tau\mathbf{r}(\tau, X_\tau^\alpha, Y_\tau^\alpha)\Big] + \mathbb{E}\Big[\int_0^\tau \tilde\Gamma_s|\xi_s - \hat\xi_s|ds\Big].
\end{aligned}$$

Recall Remark 2.8 (iii), we see that $\mathrm{tr}\big((\mathbf{n}\mathbf{n}^\top\zeta)^\top\partial_y\mathbf{n}(\mathbf{n}\mathbf{n}^\top\zeta)\big) = 0$. Then, by (5.2),

$$|\xi_s - \hat\xi_s| \le C|\mathbf{n}^\top\zeta(s, X_s^\alpha, \pi_s^\alpha)|,$$

and thus, by (5.17),

$$\begin{aligned}
&-\mathbb{E}\Big[\int_0^\tau \tilde\Gamma_s\mathcal{L}\mathbb{V}(s, X_s^\alpha, \pi_s^\alpha)ds\Big] \le \mathbb{E}\Big[\int_0^\tau \tilde\Gamma_s\mathbf{n}(s, X_s^\alpha, \pi_s^\alpha)\cdot\hat\xi_s ds\Big]\\
&\le \mathbb{E}\Big[\mathbf{r}(0, x_0, Y_0^\alpha) - \tilde\Gamma_\tau\mathbf{r}(\tau, X_\tau^\alpha, Y_\tau^\alpha)\Big] + C\mathbb{E}\Big[\int_0^\tau \tilde\Gamma_s|\mathbf{n}^\top\zeta(s, X_s^\alpha, \pi_s^\alpha)|ds\Big]\\
&\le C\varepsilon^2 + C\Big(\mathbb{E}\Big[\int_0^\tau |\tilde\Gamma_s\mathbf{n}^\top\zeta(s, X_s^\alpha, \pi_s^\alpha)|^2 ds\Big]\Big)^{1/2} \le C\varepsilon.
\end{aligned} \tag{5.20}$$

Step 2.3. Fix another small constant $\delta > 0$. Since $\mathcal{L}\mathbb{V} \le 0$, by (5.11) we have

$$
\begin{aligned}
&-\mathcal{L}\mathbb{V}(0,x_0,y_0) = \mathbb{E}\Big[-\frac{1}{\delta}\int_0^{\tau\wedge\delta} \mathcal{L}\mathbb{V}(0,x_0,y_0)ds - \frac{(\delta-\tau)^+}{\delta}\mathcal{L}\mathbb{V}(0,x_0,y_0)\Big]\\
&= -\frac{1}{\delta}\mathbb{E}\Big[\int_0^{\tau\wedge\delta} \tilde\Gamma_s \mathcal{L}\mathbb{V}(s,X^\alpha_s,\pi^\alpha_s)ds\Big] - \mathcal{L}\mathbb{V}(0,x_0,y_0)\mathbb{E}\Big[(1-\frac{\tau}{\delta})\mathbf{1}_{\{\tau<\delta\}}\Big]\\
&\quad +\frac{1}{\delta}\mathbb{E}\Big[\int_0^{\tau\wedge\delta} \big[\tilde\Gamma_s\mathcal{L}\mathbb{V}(s,X^\alpha_s,\pi^\alpha_s) - \mathcal{L}\mathbb{V}(0,x_0,y_0)\big]ds\Big]\\
&\le \frac{C\varepsilon}{\delta} + \mathbb{E}\Big[\sup_{0\le t\le\tau\wedge\delta}\big|\tilde\Gamma_t\mathcal{L}\mathbb{V}(t,X^\alpha_t,\pi^\alpha_t) - \mathcal{L}\mathbb{V}(0,x_0,y_0)\big|\Big].
\end{aligned}
$$

Since $\mathbb{V} \in C^{1,2}([0,T_0]\times\mathbb{R}^d;\mathcal{D}^m_2)$, $H_{\mathbb{V}}$ is bounded and uniform continuous. Then, for some modulus of continuity function ρ we have

$$
-\mathcal{L}\mathbb{V}(0,x_0,y_0) \le \frac{C\varepsilon}{\delta} + C\mathbb{E}\Big[\sup_{0\le t\le\tau\wedge\delta}\big[|\tilde\Gamma_t - 1| + \rho\big(\delta + |X^\alpha_t - x_0| + |\pi^\alpha_t - y_0|\big)\big]\Big]. \tag{5.21}
$$

Recall (3.9), (5.14), and (5.16), we have

$$
\begin{aligned}
&\mathbb{E}\Big[\sup_{0\le t\le\tau\wedge\delta}|\tilde\Gamma_t - 1|\Big] \le \mathbb{E}\Big[\sup_{0\le t\le\tau\wedge\delta}\big[(\tilde\Gamma_t+1)(|\int_0^t\tilde\sigma_s\cdot dB_s| + \int_0^t[|\tilde b_s| + \frac{1}{2}|\tilde\sigma_s|^2]ds\big)\Big]\\
&\quad \le C\sqrt{\delta} + C\Big(\mathbb{E}\Big[(\int_0^{\tau\wedge\delta}|Z^\alpha_t|dt)^2\Big]\Big)^{\frac{1}{2}} \le C\sqrt{\delta} + C\sqrt{\delta}\Big(\mathbb{E}\Big[\int_0^T|Z^\alpha_t|^2dt\Big]\Big)^{\frac{1}{2}} \le C\sqrt{\delta};\\
&\mathbb{E}\Big[\sup_{0\le t\le\tau\wedge\delta}|X^\alpha_t - x_0|\Big] \le C\sqrt{\delta};\\
&\mathbb{E}\Big[\sup_{0\le t\le\tau\wedge\delta}|\pi^\alpha_t - y_0|\Big] \le C\varepsilon^2 + \mathbb{E}\Big[\sup_{0\le t\le\tau\wedge\delta}|Y^\alpha_t - y_0|\Big]\\
&\quad \le C[\varepsilon^2 + \sqrt{\delta}] + C\mathbb{E}\Big[(\int_0^{\tau\wedge\delta}|Z^\alpha_t|^2dt)^{\frac{1}{2}}\Big].
\end{aligned}
$$

Then (5.21) implies

$$
\begin{aligned}
-\mathcal{L}\mathbb{V}(0,x_0,y_0) &\le C\Big[\frac{\varepsilon}{\delta} + \sqrt{\delta}\Big] + C\rho(\delta + \delta^{\frac{1}{3}}) + \frac{C}{\delta^{\frac{1}{3}}}\mathbb{E}\Big[\sup_{0\le t\le\tau\wedge\delta}[|X^\alpha_t - x_0| + |\pi^\alpha_t - y_0|]\Big]\\
&\le C\Big[\frac{\varepsilon}{\delta} + \sqrt{\delta} + \rho(\delta+\delta^{\frac{1}{3}}) + \frac{\varepsilon^2+\sqrt{\delta}}{\delta^{\frac{1}{3}}}\Big] + \frac{C}{\delta^{\frac{1}{3}}}\mathbb{E}\Big[(\int_0^{\tau\wedge\delta}|Z^\alpha_t|^2dt)^{\frac{1}{2}}\Big]. \quad (5.22)
\end{aligned}
$$

Step 2.4. Recall (5.5) and (5.19). Then by (5.20) we have

$$
C\varepsilon \ge \mathbb{E}\Big[\int_0^{\tau\wedge\delta}\tilde\Gamma_s\mathbf{n}(s,X^\alpha_s,\pi^\alpha_s)\cdot\hat\xi_s ds\Big] \ge \mathbb{E}\Big[\int_0^{\tau\wedge\delta}\tilde\Gamma_s\big[\frac{c_{T_0}}{2}|\hat\zeta_s|^2 - C|\hat\zeta_s| - C\big]ds\Big].
$$

This, together with (5.16), implies that

$$
\mathbb{E}\Big[\int_0^{\tau\wedge\delta}\tilde\Gamma_s|\hat\zeta_s|^2ds\Big] \le C[\varepsilon+\delta]. \tag{5.23}
$$

By (5.12) and (5.18) we have

$$|Z_t^\alpha| \le |\zeta_t(\pi_t^\alpha)| + C \le |\hat{\zeta}_t| + |\mathbf{n}^\top \zeta(t, X_t^\alpha, \pi_t^\alpha)| + C.$$

Then, by (5.17), (5.23), and (5.16), we have

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\int_0^{\tau\wedge\delta} |Z_t^\alpha|^2 dt\Big)^{\frac{1}{2}}\Big] \le C\sqrt{\delta} + C\mathbb{E}\Big[\Big(\int_0^{\tau\wedge\delta} [|\hat{\zeta}_t|^2 + |\mathbf{n}\cdot\zeta(t, X_t^\alpha, \pi_t^\alpha)|^2] dt\Big)^{\frac{1}{2}}\Big] \\
&\le C\sqrt{\delta} + C\mathbb{E}\Big[\Big(\sup_{0\le t\le\tau} \tilde{\Gamma}_t^{-1} \int_0^{\tau\wedge\delta} \tilde{\Gamma}_t |\hat{\zeta}_t|^2 dt\Big)^{\frac{1}{2}} \\
&\qquad\qquad + \Big(\sup_{0\le t\le\tau} \tilde{\Gamma}_t^{-2} \int_0^{\tau} \tilde{\Gamma}_t^2 |\mathbf{n}\cdot\zeta(t, X_t^\alpha, \pi_t^\alpha)|^2] dt\Big)^{\frac{1}{2}}\Big] \\
&\le C\sqrt{\delta} + C\Big(\mathbb{E}\Big[\int_0^{\tau\wedge\delta} \tilde{\Gamma}_t|\hat{\zeta}_t|^2 dt + \int_0^{\tau} \tilde{\Gamma}_t^2 |\mathbf{n}\cdot\zeta(t, X_t^\alpha, \pi_t^\alpha)|^2 dt\Big]\Big)^{\frac{1}{2}} \\
&\le C\sqrt{\delta} + C\big(\varepsilon + \delta + \varepsilon^2\big)^{\frac{1}{2}} \le C[\sqrt{\varepsilon} + \sqrt{\delta}].
\end{aligned}$$

Plug this into (5.22), we get

$$-\mathcal{L}\mathbb{V}(0, x_0, y_0) \le C\Big[\frac{\varepsilon}{\delta} + \sqrt{\delta} + \rho(\delta + \delta^{\frac{1}{3}}) + \frac{\varepsilon^2 + \sqrt{\delta}}{\delta^{\frac{1}{3}}}\Big] + \frac{C}{\delta^{\frac{1}{3}}}[\sqrt{\varepsilon} + \sqrt{\delta}].$$

By first send $\varepsilon \to 0$ and then $\delta \to 0$, we obtain $-\mathcal{L}\mathbb{V}(0, x_0, y_0) \le 0$. ■

6 The uniqueness of the classical solution

We now turn to the uniqueness of the classical solution, including the verification result.

Theorem 6.1 *Let Assumption 4.4 hold and $\mathbb{V}$ be defined by (4.2).*

(i) Assume $\mathbb{U} \in C_0^{1,2}([0,T) \times \mathbb{R}^d; \mathcal{D}_2^m)$ is a classical solution of (5.3) with terminal condition $\mathbb{V}(T, x) = \{g(x)\}$. Then $\mathbb{U} = \mathbb{V}$, and consequently (5.3) has a unique classical solution with terminal condition $\{g(x)\}$.

(ii) Assume further that the Hamiltonian $H_{\mathbb{U}}(\cdot, \partial_x\mathbb{U}, \partial_{xx}\mathbb{U})$ has an optimal argument:

$$a^* = I_1^{\mathbb{U}}(t, x, y) \in A, \quad \zeta^* = I_2^{\mathbb{U}}(t, x, y) \in (\mathbb{T}_{\mathbb{U}}(t, x, y))^d.$$

Moreover, recall Remark 2.1 and denote

$$\begin{aligned}
\tilde{I}_3^{\mathbb{U}}(t, x, y) &:= -\Big[\partial_t\mathbb{U} + h_{\mathbb{U}}(\cdot, \partial_x\mathbb{U}, \partial_{xx}\mathbb{U}, I_1^{\mathbb{U}}, I_2^{\mathbb{U}})\Big](t, x, y); \\
I_3^{\mathbb{U}}(t, x, y) &:= (\mathbf{I}_m - \mathbf{n}_{\mathbb{U}}\mathbf{n}_{\mathbb{U}}^\top)\ \tilde{I}_3^{\mathbb{U}}(t, x, y);
\end{aligned} \tag{6.1}$$

and assume, for given $(0, x_0, y_0) \in \mathbb{G}_{\mathbb{U}}$, the following SDE has a strong solution:

$$\begin{aligned}
X_t^* &= x_0 + \int_0^t b(\cdot, I_1^{\mathbb{U}})(s, X_s^*, \Upsilon_s^*) ds + \int_0^t \sigma(\cdot, I_1^{\mathbb{U}})(s, X_s^*, \Upsilon_s^*) dB_s; \\
\Upsilon_t^* &= y_0 + \int_0^t \Big[\partial_t\mathbb{U} + h_{\mathbb{U}}^0(\cdot, \partial_x\mathbb{U}, \partial_{xx}\mathbb{U}, I_1^{\mathbb{U}}, I_2^{\mathbb{U}}) + I_3^{\mathbb{U}}\Big](s, X_s^*, \Upsilon_s^*) ds \\
&\quad + \int_0^t [\partial_x\mathbb{U}\sigma(\cdot, I_1^{\mathbb{U}}) + I_2^{\mathbb{U}}](s, X_s^*, \Upsilon_s^*) dB_s.
\end{aligned} \tag{6.2}$$

Then, for $\alpha^*_t := I_1^{\mathbb{U}}(t, X^*_t, \Upsilon^*_t)$, we have $Y_t^{0,x_0,\alpha^*} = \Upsilon^*_t \in \mathbb{V}_b(t, X^*_t)$, $0 \le t \le T$, a.s. In particular, $Y_0^{0,x_0,\alpha^*} = y_0$.

Remark 6.2 *From Step 2 in the proof, especially (6.6) below, we see that (6.2) actually becomes the following simpler and more natural SDE:*

$$
\begin{aligned}
X^*_t &= x_0 + \int_0^t b(\cdot, I_1^{\mathbb{U}})(s, X^*_s, \Upsilon^*_s)ds + \int_0^t \sigma(\cdot, I_1^{\mathbb{U}})(s, X^*_s, \Upsilon^*_s)dB_s;\\
\Upsilon^*_t &= y_0 - \int_0^t f(\cdot, \partial_x \mathbb{U}\sigma(\cdot, I_1^{\mathbb{U}}) + I_2^{\mathbb{U}}, I_1^{\mathbb{U}})(s, X^*_s, \Upsilon^*_s)ds \qquad (6.3)\\
&\quad + \int_0^t [\partial_x \mathbb{U}\sigma(\cdot, I_1^{\mathbb{U}}) + I_2^{\mathbb{U}}](s, X^*_s, \Upsilon^*_s)dB_s. \qquad \blacksquare
\end{aligned}
$$

Remark 6.3 *(i) Under the setting of above (ii), the α^* is an optimal argument (at least locally) for the scalarized optimization problem: $\sup_{\alpha\in\mathcal{A}_0} \mathbf{n}(0, x_0, y_0) \cdot Y_0^{0,x_0,\alpha}$. We refer to Subsection 7 below for more detailed analysis along this line.*

(ii) When σ is nondegenerate, by (6.3) we have

$$
\begin{aligned}
\Upsilon^*_t = y_0 &+ \int_0^t \Big[\big[\partial_x \mathbb{U}\sigma(\cdot, I_1^{\mathbb{U}}) + I_2^{\mathbb{U}}\big]\sigma^{-1}(\cdot, I_1^{\mathbb{U}})\Big](s, X^*_s, \Upsilon^*_s)dX^*_s\\
&- \int_0^t \Big[f(\cdot, \partial_x \mathbb{U}\sigma(\cdot, I_1^{\mathbb{U}}) + I_2^{\mathbb{U}}, I_1^{\mathbb{U}}) + \big[\partial_x \mathbb{U}\sigma(\cdot, I_1^{\mathbb{U}}) + I_2^{\mathbb{U}}\big]\sigma^{-1}b(\cdot, I_1^{\mathbb{U}})\Big](s, X^*_s, \Upsilon^*_s)ds
\end{aligned}
$$

Then we may write Υ^*_t as a function of $X^*_{[0,t]}$, thus as a closed loop control $\alpha^*_t = \alpha^*(t, X^*_{[0,t]})$ is path dependent. Such path dependence appears often in multivariate setting. However, we note that (X^*, Υ^*) is jointly Markovian, so by adding the state Υ^*, the optimal control α^* becomes Markovian, or more precisely state dependent. Therefore, the above verification theorem does help to construct Markovian optimal controls in this sense. $\blacksquare$

Proof of Theorem 6.1. We proceed in three steps. Denote $T_\delta := T - \delta$ for $\delta > 0$ small.

Step 1. We first show that $\mathbb{V}(0, x_0) \subset \mathbb{U}(0, x_0)$. By the same arguments, we can also show that $\mathbb{V}(t, x) \subset \mathbb{U}(t, x)$ for all $(t, x) \in [0, T) \times \mathbb{R}^d$.

Fix $\delta > 0$ small and $\alpha \in \mathcal{A}_0$. Denote $(X^\alpha, Y^\alpha, Z^\alpha) := (X^{0,x_0,\alpha}, Y^{0,x_0,\alpha}, Z^{0,x_0,\alpha})$. Since $\mathbb{V}(T, x) = \{g(x)\} = \mathbb{U}(T, x)$, by Assumption 4.4 and the continuity of $\mathbb{U}$, there exists $\phi_\delta \in \mathbb{L}^2(\mathcal{F}_{T_\delta})$ such that $\phi_\delta \in \mathbb{U}_b(T_\delta, X^\alpha_{T_\delta})$, a.s. and

$$
\mathbb{E}\big[|Y^\alpha_{T_\delta} - \phi_\delta|^2\big] \le C\mathbb{E}\big[|Y^\alpha_{T_\delta} - g(X^\alpha_T)|^2 + |g(X^\alpha_T) - g(X^\alpha_{T_\delta})|^2 + |\phi_\delta - g(X^\alpha_{T_\delta})|^2\big] \to 0, \qquad (6.4)
$$

as $\delta \to 0$. Recall (4.15) and set $(Y^{\alpha,\delta}, Z^{\alpha,\delta}) := (Y^{T_\delta,\phi_\delta;0,x_0,\alpha}, Z^{T_\delta,\phi_\delta;0,x_0,\alpha})$. Then by the standard BSDE estimates we have

$$
\lim_{\delta\to 0} |Y_0^{\alpha,\delta} - Y_0^\alpha| = 0. \qquad (6.5)
$$

As in (3.5), by standard Itô's formula, we have

$$
\begin{aligned}
&d\mathbf{r}_{\mathbb{U}}(t, X^\alpha_t, Y^{\alpha,\delta}_t) = \Lambda(t, X^\alpha_t, Y^{\alpha,\delta}_t, Z^{\alpha,\delta}_t, \alpha_t)dt + \tilde{Z}^{\alpha,\delta}_t dB_t, \quad \text{where}\\
&\Lambda := \nabla_t \mathbf{r}_{\mathbb{U}} + \nabla_x \mathbf{r}_{\mathbb{U}} \cdot b - \nabla_y \mathbf{r}_{\mathbb{U}} \cdot f + \frac{1}{2}\mathrm{tr}\left(\sigma^\top \nabla_{xx}\mathbf{r}_{\mathbb{U}}\sigma + 2z^\top \nabla_{xy}\mathbf{r}_{\mathbb{U}}\sigma + z^\top \nabla_{yy}\mathbf{r}_{\mathbb{U}} z\right);\\
&\tilde{Z}^{\alpha,\delta} := \nabla_x \mathbf{r}_{\mathbb{U}}\sigma + (\nabla_y \mathbf{r}_{\mathbb{U}})^\top Z^{\alpha,\delta}.
\end{aligned}
$$

Denote

$$\pi_t^{\alpha,\delta} := \pi_{\mathbb{U}}(t, X_t^\alpha, Y_t^{\alpha,\delta}), \quad \zeta_t^{\alpha,\delta} := Z_t^{\alpha,\delta} - \mathbf{n}_{\mathbb{U}}\mathbf{n}_{\mathbb{U}}^\top Z_t^{\alpha,\delta}.$$

Then, by (3.6) and since $\mathbb{U}$ is a classical solution of (5.3), we have

$$\Lambda(t, X_t^\alpha, \pi_t^{\alpha,\delta}, \partial_x \mathbb{U}\sigma + \zeta_t^{\alpha,\delta}, \alpha_t) = -\mathbf{n}_{\mathbb{U}} \cdot \big[\partial_t \mathbb{U} + h_{\mathbb{U}}\big(t, X_t^\alpha, \pi_t^{\alpha,\delta}, \partial_x \mathbb{U}, \partial_{xx}\mathbb{U}, \alpha_t, \zeta_t^{\alpha,\delta}\big)\big] \geq 0.$$

Note that, for appropriate processes $\tilde{b}, \tilde{\sigma}$,

$$\Lambda(t, X_t^\alpha, Y_t^{\alpha,\delta}, Z_t^{\alpha,\delta}, \alpha_t) - \Lambda(t, X_t^\alpha, \pi_t^{\alpha,\delta}, \partial_x \mathbb{U}\sigma + \zeta_t^{\alpha,\delta}, \alpha_t) = -\big[\tilde{b}_t \mathbf{r}_{\mathbb{U}} + \tilde{\sigma}_t \tilde{Z}_t^{\alpha,\delta}\big].$$

Here, due to the regularity of $\mathbb{U} \in C^{1,2}([0, T_\delta] \times \mathbb{R}^d; \mathcal{D}_2^m)$, as in (3.7) there exists a constant $C_\delta > 0$, which may depend on δ, such that for $0 \leq t \leq T_\delta$,

$$|\tilde{b}_t| \leq C_\delta[1 + |Z_t^{\alpha,\delta}|^2], \quad |\tilde{\sigma}_t| \leq C_\delta[1 + |Z_t^{\alpha,\delta}|].$$

Then, for the $\tilde{\Gamma}$ in (3.9) we have

$$d\Big(\tilde{\Gamma}_t \mathbf{r}_{\mathbb{U}}(t, X_t^\alpha, Y_t^{\alpha,\delta})\Big) = \tilde{\Gamma}_t \Lambda(t, X_t^\alpha, \pi_t^{\alpha,\delta}, \partial_x \mathbb{U}\sigma + \zeta_t^{\alpha,\delta}, \alpha_t)dt + \tilde{\Gamma}_t(\tilde{Z}_t^{\alpha,\delta} - \mathbf{r}_{\mathbb{U}}\tilde{\sigma}_t)dB_t.$$

Since $\mathbf{r}(T_\delta, X_{T_\delta}^\alpha, Y_{T_\delta}^{\alpha,\delta}) = 0$, a.s. then,

$$\mathbf{r}_{\mathbb{U}}(0, x_0, Y_0^{\alpha,\delta}) = -\mathbb{E}\Big[\int_0^{T_\delta} \tilde{\Gamma}_t \Lambda(t, X_t^\alpha, \pi_t^{\alpha,\delta}, \partial_x \mathbb{U}\sigma + \zeta_t^{\alpha,\delta}, \alpha_t)\Big] \leq 0.$$

That is, $Y_0^{\alpha,\delta} \in \mathbb{U}(0, x_0)$. Send $\delta \to 0$, by (6.5) and the closedness of $\mathbb{U}(0, x_0)$, we have $Y_0^\alpha \in \mathbb{U}(0, x_0)$. Moreover, since $\alpha \in \mathcal{A}_0$ is arbitrary, we obtain $\mathbb{V}(0, x_0) \subset \mathbb{U}(0, x_0)$.

Step 2. We next prove (ii) and show that in this case $\mathbb{U}(0, x_0) \subset \mathbb{V}(0, x_0)$. Indeed, consider an arbitrary $y_0 \in \mathbb{U}_b(0, x_0)$. First by the Itô formula Theorem 3.1 we see that $\Upsilon_t^* \in \mathbb{V}_b(t, X_t^*)$, $0 \leq t \leq T$, a.s. In particular, this implies $\Upsilon_T^* = g(X_T^*)$. Note that, by the optimality of $I_1^{\mathbb{U}}, I_2^{\mathbb{U}}$,

$$h_{\mathbb{U}}(\cdot, \partial_x \mathbb{U}, \partial_{xx}\mathbb{U}, I_1^{\mathbb{U}}, I_2^{\mathbb{U}})(s, X_s^*, \Upsilon_s^*) = H_{\mathbb{U}}(\cdot, \partial_x \mathbb{U}, \partial_{xx}\mathbb{U}, I_1^{\mathbb{U}}, I_2^{\mathbb{U}})(s, X_s^*, \Upsilon_s^*).$$

Since $\mathbb{U}$ satisfies the PDE (5.3) and by (5.2), at $(s, X_s^*, \Upsilon_s^*) \in \mathbb{G}_{\mathbb{U}}$ we have

$$\mathbf{n}_{\mathbb{U}} \cdot \tilde{I}_3^{\mathbb{U}} = 0; \quad \partial_t \mathbb{U} + h_{\mathbb{U}}^0(\cdot, \partial_x \mathbb{U}, \partial_{xx}\mathbb{U}, I_1^{\mathbb{U}}, I_2^{\mathbb{U}}) + I_3^{\mathbb{U}} = -f(\cdot, \partial_x \mathbb{U}\sigma(\cdot, I_1^{\mathbb{U}}) + I_2^{\mathbb{U}}, I_1^{\mathbb{U}}). \quad (6.6)$$

This implies that $Y_t^{0,x_0,\alpha^*} = \Upsilon_t^*$. In particular, $y_0 = \Upsilon_0^* = Y_0^{0,x_0,\alpha^*} \in \mathbb{V}(0, x_0)$. Thus $\mathbb{U}_b(0, x_0) \subset \mathbb{V}(0, x_0)$, which implies that $\mathbb{U}(0, x_0) \subset \mathbb{V}(0, x_0)$.

Step 3. We now prove $\mathbb{U}(0, x_0) \subset \mathbb{V}(0, x_0)$ in the general case, without assuming the additional conditions in (ii). As in Step 2 of Theorem 5.6, this step is more involved than Step 2 above, because it requires to construct approximately "optimal controls". Fix $(0, x_0, y_0) \in \mathbb{G}_{\mathbb{U}}$ and $\delta > 0$. Since $\mathbb{U} \in C^{1,2}([0, T_\delta] \times \mathbb{R}^d; \mathcal{D}_2^m)$, we assume $\mathbf{r}_{\mathbb{U}}$ is smooth in $O_{\varepsilon_0}^{T_\delta}(\mathbb{G}_{\mathbb{U}})$ for some $\varepsilon_0 > 0$. In the rest of this proof, let C_δ be a generic constant which may depend on δ, more precisely on the c_{T_δ} in (5.5) and the regularity of $\mathbb{U}$ on $[0, T_\delta] \times \mathbb{R}^d$.

Since $\mathbb{U}$ satisfies (5.3), by (5.5) there exist $\bar{a}_0 \in A$ and $\bar{\zeta}^0 \in (\mathbb{T}_{\mathbb{U}}(0, x_0, y_0))^d$ such that

$$|\bar{\zeta}^0| \le C_\delta \quad \text{and} \quad 0 \le -\mathbf{n}_{\mathbb{U}} \cdot \big[\partial_t \mathbb{U} + h_{\mathbb{U}}(\cdot, \partial_x \mathbb{U}, \partial_{xx}\mathbb{U}, \bar{a}_0, \bar{\zeta}^0)\big](0, x_0, y_0) < \delta. \tag{6.7}$$

Set $\tau_0 := 0$, $\alpha^1_t \equiv \bar{a}_0$, $0 \le t \le T_\delta$, and define

$$X^1_t = x_0 + \int_0^t b(s, X^1_s, \alpha^1_s) ds + \int_0^t \sigma(s, X^1_s, \alpha^1_s) dB_s, \quad 0 \le t \le T_\delta.$$

Recall Remark 2.1 and introduce random fields $(\xi^1, \zeta^1) : [0, T_\delta] \times \Omega \times \mathbb{R}^m \to (\mathbb{R}^m, \mathbb{R}^{m\times d})$:

$$\zeta^1_t(y) := \bar{\zeta}^0 - \mathbf{n}_{\mathbb{U}} \mathbf{n}_{\mathbb{U}}^\top \bar{\zeta}^0(t, X^1_t, y), \quad \xi^1_t(y) := \tilde{\xi}^1_t(y) - [\mathbf{n}_{\mathbb{U}} \cdot \tilde{\xi}^1_t] \mathbf{n}_{\mathbb{U}}(t, X^1_t, y),$$
$$\text{where} \quad \tilde{\xi}^1_t(y) := -\big[\partial_t \mathbb{U} + h_{\mathbb{U}}(\cdot, \partial_x \mathbb{U}, \partial_{xx}\mathbb{U}, \alpha^1, \zeta^1)\big](t, X^1_t, y).$$

Then $\xi^1_t(y) \in \mathbb{T}_{\mathbb{U}}(t, X^1_t, y)$, $\zeta^1_t(y) \in (\mathbb{T}_{\mathbb{U}}(t, X^1_t, y))^d$, $\forall y \in \mathbb{U}_b(t, X^1_t)$, and ξ^1, ζ^1 are uniformly Lipschitz continuous in y, with a Lipschitz constant depending on δ. Consider the SDE:

$$\begin{aligned}\Upsilon^1_t = y_0 &+ \int_0^t \Big[\partial_t \mathbb{U} + h^0_{\mathbb{U}}(\cdot, \partial_x \mathbb{U}, \partial_{xx}\mathbb{U}, \alpha^1_s, \zeta^1_s) + \xi^1\Big](s, X^1_s, \Upsilon^1_s) ds \\ &+ \int_0^t \Big[\partial_x \mathbb{U}(s, X^1_s, \Upsilon^1_s)\sigma(s, X^1_s, \alpha^1_s) + \zeta^1_s(\Upsilon^1_s)\Big] dB_s.\end{aligned} \tag{6.8}$$

By the Itô formula Theorem 3.1 we have $\Upsilon^1_t \in \mathbb{U}_b(t, X^1_t)$, $0 \le t \le T_\delta$. Note that (6.7) implies $\mathbf{n}_{\mathbb{U}}(0, X^1_0, \Upsilon^1_0) \cdot \tilde{\xi}^1_0(\Upsilon^1_0) \le \delta$, and by our construction, α^1, ζ^1 and hence $\tilde{\xi}^1$ are continuous in t. We then set

$$\tau_1 := \inf\Big\{t > \tau_0 : \mathbf{n}_{\mathbb{U}}(t, X^1_t, \Upsilon^1_t) \cdot \tilde{\xi}^1_t(\Upsilon^1_t) \ge 2\delta\Big\} \wedge T_\delta.$$

Next, on $\{\tau_1 < T_\delta\}$, by measurable selection theorem, there exist $\mathcal{F}_{\tau_1}$-measurable random variables $\bar{\alpha}^1_{\tau_1} \in A$ and $\bar{\zeta}^1_{\tau_1} \in (\mathbb{T}_{\mathbb{V}}(\tau_1, X^1_{\tau_1}, \Upsilon^1_{\tau_1}))^d$ such that

$$|\bar{\zeta}^1_{\tau_1}| \le C_\delta \quad \text{and} \quad 0 \le -\mathbf{n}_{\mathbb{U}} \cdot \big[\partial_t \mathbb{U} + h_{\mathbb{U}}(\cdot, \partial_x \mathbb{U}, \partial_{xx}\mathbb{U}, \bar{\alpha}^1_{\tau_1}, \bar{\zeta}^1_{\tau_1})\big](\tau_1, X^1_{\tau_1}, \Upsilon^1_{\tau_1}) < \delta.$$

Set $\alpha^2_t \equiv \bar{\alpha}^1_{\tau_1}$, $\tau_1 \le t \le T_\delta$, and define

$$X^2_t = X^1_{\tau_1} + \int_{\tau_1}^t b(s, X^2_s, \alpha^2_s) ds + \int_{\tau_1}^t \sigma(s, X^2_s, \alpha^2_s) dB_s, \quad \tau_1 \le t \le T_\delta.$$

Similarly introduce, for $\tau_1 \le t \le T_\delta$,

$$\zeta^2_t(y) := (\mathbf{I}_m - \mathbf{n}_{\mathbb{U}} \mathbf{n}_{\mathbb{U}}^\top) \bar{\zeta}^1_{\tau_1}(t, X^2_t, y), \quad \xi^2_t(y) := (\mathbf{I}_m - \mathbf{n}_{\mathbb{U}} \mathbf{n}_{\mathbb{U}}^\top) \tilde{\xi}^2_t(y) \mathbf{n}_{\mathbb{U}}(t, X^2_t, y),$$
$$\text{where} \quad \tilde{\xi}^2_t(y) := -\big[\partial_t \mathbb{U} + h_{\mathbb{U}}(\cdot, \partial_x \mathbb{U}, \partial_{xx}\mathbb{U}, \alpha^2, \zeta^2)\big](t, X^2_t, y),$$

and consider the SDE:

$$\begin{aligned}\Upsilon^2_t = \Upsilon^1_{\tau_1} &+ \int_{\tau_1}^t \Big[\partial_t \mathbb{U} + h^0_{\mathbb{U}}(\cdot, \partial_x \mathbb{U}, \partial_{xx}\mathbb{U}, \alpha^2_s, \zeta^2_s) + \xi^2\Big](s, X^2_s, \Upsilon^2_s) ds \\ &+ \int_{\tau_1}^t \Big[\partial_x \mathbb{U}(s, X^2_s, \Upsilon^2_s)\sigma(s, X^2_s, \alpha^2_s) + \zeta^2_s(\Upsilon^2_s)\Big] dB_s.\end{aligned}$$

Then $\Upsilon^2_t \in \mathbb{U}_b(t, X^2_t)$, $\tau_1 \le t \le T_\delta$, and we may set

$$\tau_2 := \inf\Big\{t > \tau_1 : \mathbf{n}_{\mathbb{U}}(t, X^2_t, \Upsilon^2_t) \cdot \tilde{\xi}^2_t(\Upsilon^2_t) \ge 2\delta\Big\} \wedge T_\delta.$$

Repeat the arguments, we obtain a sequence $(\tau_n, \alpha^n, \zeta^n, \tilde{\xi}^n, \xi^n, X^n, \Upsilon^n)$, $n \ge 0$, satisfying the desired properties. We first show that $\tau_n = T_\delta$ for n large enough, a.s. Indeed, on $E_\delta := \cap_{n\ge 1}\{\tau_n < T_\delta\}$, we have,

$$\mathbf{n}_{\mathbb{U}}(\tau_n, X_{\tau_n}, \Upsilon_{\tau_n}) \cdot \tilde{\xi}^n_{\tau_n}(\Upsilon_{\tau_n}) \le \delta, \quad \mathbf{n}_{\mathbb{U}}(\tau_{n+1}, X_{\tau_{n+1}}, \Upsilon_{\tau_{n+1}}) \cdot \tilde{\xi}^n_{\tau_{n+1}}(\Upsilon_{\tau_{n+1}}) = 2\delta, \quad \forall n.$$

Then, for any n,

$$\delta \mathbb{P}(E_\delta) \le \mathbb{E}\Big[\Big|\mathbf{n}_{\mathbb{U}}(\tau_{n+1}, X_{\tau_{n+1}}, \Upsilon_{\tau_{n+1}}) \cdot \tilde{\xi}^n_{\tau_{n+1}}(\Upsilon_{\tau_{n+1}}) - \mathbf{n}_{\mathbb{U}}(\tau_n, X_{\tau_n}, \Upsilon_{\tau_n}) \cdot \tilde{\xi}^n_{\tau_n}(\Upsilon_{\tau_n})\Big|\Big].$$

Send $n \to \infty$, by the desired regularity and in particular $|\zeta| \le C_\delta$, we obtain $\mathbb{P}(E_\delta) = 0$.

We now define

$$(\alpha_t, \zeta_t, X_t, \Upsilon_t, \xi_t) := (\alpha^n_t, \zeta^n_t, X^n_t, \Upsilon^n_t, \xi^n_t), \quad t \in [\tau_n, \tau_{n+1}), n = 0, 1, \cdots.$$

Note that $X_{T_\delta} := \lim_{t\uparrow T_\delta} X_t$ and $\Upsilon_{T_\delta} := \lim_{t\uparrow T_\delta} \Upsilon_t$ are well defined. Define

$$Z_t := \partial_x \mathbb{U}(t, X_t, \Upsilon_t)\sigma(t, X_t, \alpha_t) + \zeta_t(\Upsilon_t), \quad \eta_t := [\mathbf{n}_{\mathbb{U}} \cdot \tilde{\xi}_t]\mathbf{n}_{\mathbb{U}}(t, X_t, \Upsilon_t), \quad 0 \le t < T_\delta.$$

Then, $|\eta| \le 2\delta$, and by (5.2) and (6.8) we have

$$\Upsilon_t = y_0 - \int_0^t \Big[f(s, X_s, \Upsilon_s, Z_s, \alpha_s) + \eta_s\Big]ds + \int_0^t Z_s dB_s, \quad 0 \le t \le T_\delta.$$

Equivalently, we may rewrite it backwardly:

$$\Upsilon_t = \Upsilon_{T_\delta} + \int_t^{T_\delta} \Big[f(s, X_s, \Upsilon_s, Z_s, \alpha_s) + \eta_s\Big]ds - \int_t^{T_\delta} Z_s dB_s, \quad 0 \le t \le T_\delta.$$

Compare this with (4.15), by standard BSDE estimates we have

$$\big|y_0 - Y_0^{T_\delta, \Upsilon_{T_\delta}; 0, x_0, \alpha}\big|^2 = \big|\Upsilon_0 - Y_0^{T_\delta, \Upsilon_{T_\delta}; 0, x_0, \alpha}\big|^2 \le C\mathbb{E}\Big[\int_0^{T_\delta} |\eta_s|^2 ds\Big] \le C\delta^2. \tag{6.9}$$

Finally, fix an arbitrary $a_* \in A$, and extend α with $\alpha_t \equiv a_*$, $t \in [T_\delta, T]$. Since $\mathbb{V}(T, x) = \{g(x)\} = \mathbb{U}(T, x)$, by Assumption 4.4 and the continuity of $\mathbb{U}$, similarly to (6.4) we have

$$\begin{aligned}&\mathbb{E}\big[|Y^\alpha_{T_\delta} - \Upsilon_{T_\delta}|^2\big]\\ &\le C\mathbb{E}\Big[|Y^\alpha_{T_\delta} - g(X^\alpha_T)|^2 + |g(X^\alpha_T) - g(X^\alpha_{T_\delta})|^2 + |\Upsilon_{T_\delta} - g(X^\alpha_{T_\delta})|^2\Big] \le \rho(\delta),\end{aligned}$$

for some modulus of continuity function ρ, independent of α. Then, by standard BSDE estimates again,

$$\big|Y_0^{T_\delta, \Upsilon_{T_\delta}; 0, x_0, \alpha} - Y^\alpha_0\big|^2 = \big|Y_0^{T_\delta, \Upsilon_{T_\delta}; 0, x_0, \alpha} - Y_0^{T_\delta, Y^\alpha_{T_\delta}; 0, x_0, \alpha}\big|^2 \le \mathbb{E}\big[|Y^\alpha_{T_\delta} - \Upsilon_{T_\delta}|^2\big] \le \rho(\delta).$$

Combine this with (6.9), we have

$$|y_0 - Y^\alpha_0| \le C\delta + \sqrt{\rho(\delta)}.$$

Since $Y^\alpha_0 \in \mathbb{V}(0, x_0)$ and $\delta > 0$ is arbitrary, we obtain $y_0 \in \mathbb{V}(0, x_0)$. ■

We conclude this section with a simple example where $\mathbb{V}$ is indeed a classical solution.

Example 6.4 *Set $d = 1$, $m = 2$, $A = \{a \in \mathbb{R}^2 : |a| \le 1\}$, and*

$$b = 0, \quad \sigma = 1, \quad f = f^0(t,x) + a,$$

where f^0 and g are smooth and bounded. Then it is straightforward to check that

$$\mathbb{V}(t,x) = \Big\{ y \in \mathbb{R}^2 : |y - w(t,x)| \le T - t \Big\},$$

where $w = (w_1, w_2)^\top$ is the classical solution to the following heat equations:

$$\nabla_t w_i + \frac{1}{2}\nabla_{xx} w_i + f_i^0 = 0, \quad w_i(T,x) = g_i(x), \quad i = 1, 2.$$

We shall prove in Appendix that $\mathbb{V} \in C_0^{1,2}([0,T) \times \mathbb{R}; \mathcal{D}_2^2)$, and the conditions in Theorem 6.1 (ii) hold true. Then it follows from Theorems 5.6 and 6.1 that $\mathbb{V}$ is the unique classical solution of the HJB equation (5.4). ■

7 An application: the moving scalarization

Recall Remark 4.2, in particular (4.4) and (4.6) for the mean variance optimization problem. This problem is time inconsistent in the following sense. Consider the general setting (4.1) and (4.2). Given $(0, x_0)$ and $\varphi \in C(\mathbb{R}^m; \mathbb{R})$, let $\alpha^*_{[0,T]}$ be an optimal control for the problem

$$V_0 := \sup_{\alpha \in \mathcal{A}_0} \varphi(Y_0^{0,x_0,\alpha}) = \sup_{y \in \mathbb{V}(0,x_0)} \varphi(y). \tag{7.1}$$

If we follow α^* on $[0,t]$ and denote $X_t^* := X_t^{0,x_0,\alpha^*}$. Then $\alpha^*_{[t,T]}$ is not optimal for the optimization problem at t[10]:

$$\sup_{\alpha_{[t,T]}} \varphi\big(Y_t^{t,X_t^*,\alpha_{[t,T]}}\big) = \sup_{y \in \mathbb{V}(t,X_t^*)} \varphi(y).$$

It was proposed in [16] to find a so called dynamic utility function $\Phi(t, \mathbf{x}_{[0,t]}, y)$ such that $\Phi(0, x_0, y) = \varphi(y)$ and $\alpha^*_{[t,T]}$ remains optimal for the alternative optimization problem

$$\sup_{\alpha_{[t,T]}} \Phi\Big(t, X^*_{[0,t]}, Y_t^{t,X_t^*,\alpha_{[t,T]}}\Big) = \sup_{y \in \mathbb{V}(t,X_t^*)} \Phi\big(t, X^*_{[0,t]}, y\big). \tag{7.2}$$

In Subsection 7.2 below we will find such an Φ for the mean variance problem explicitly. In the next subsection we first consider the case that φ is linear.

[10] Here we are using the notations heuristically. Rigorously we shall either consider ess sup in the left side or consider X^* and $\alpha_{[t,T]}$ in a pathwise manner.

7.1 The linear scalarization

When φ is linear: $\varphi(y) = \lambda_0 \cdot y$ for some $\lambda_0 \in \mathbb{R}^m$, we require Φ to be linear as well: $\Phi(t, \mathbf{x}_{[0,t]}, y) = \Lambda(t, \mathbf{x}_{[0,t]}) \cdot y$. This Λ is exactly the moving scalarization proposed in [10]. That is, we want to find Λ such that $\Lambda(0, x_0) = \lambda_0$ and $\alpha^*_{[t,T]}$ is optimal for the problem:

$$\sup_{\alpha_{[t,T]}} \Lambda(t, X^*_{[0,t]}) \cdot Y_t^{t, X^*_t, \alpha_{[t,T]}} = \sup_{y \in \mathbb{V}(t, X^*_t)} \Lambda(t, X^*_{[0,t]}) \cdot y. \tag{7.3}$$

Our set-valued HJB equation provides a solution to this interesting problem, provided that (5.3) is wellposed in the sense of Theorem 6.1 (ii) and $\mathbb{V}(t,x)$ is strictly convex. Consider a slightly more general setting by letting $\lambda : \mathbb{R}^d \to \mathbb{R}^m$ be such that $\lambda(x_0) = \lambda_0$. Assume without loss of generality that $|\lambda(x)| = 1$ for all $x \in \mathbb{R}^d$. Since $\mathbb{V}(0,x)$ is compact and strict convex, we may find a unique optimal argument $y_\lambda(x) \in \mathbb{V}_b(0,x)$ for the problem: $V(0,x) := \sup_{y \in \mathbb{V}(0,x)} \lambda(x) \cdot y$. Recalling $\mathbb{U} = \mathbb{V}$, we construct $X^*, \Upsilon^*, \alpha^*$ as in Theorem 6.1 (ii) with initial data $(0, x, y_\lambda(x)) \in \mathbb{G}_{\mathbb{V}}$. Assume further that $\sigma \in \mathbb{R}^d$ is nondegenerate, then as in Remark 6.3 (ii) Υ^* is $\mathbb{F}^{X^*}$-progressively measurable and hence there exists Λ such that

$$\Lambda(t, X^*_{[0,t]}) = \mathbf{n}_{\mathbb{V}}(t, X^*_t, \Upsilon^*_t). \tag{7.4}$$

We argue that this Λ is a desired moving scalarization.

First, since $\lambda(x) \cdot y_\lambda(x) = \sup_{y \in \mathbb{V}(0,x)} \lambda(x) \cdot y$ and $|\lambda(x)| = 1$, we see that

$$\lambda(x) = \mathbf{n}_{\mathbb{V}}(0, x, y_\lambda(x)) = \Lambda(0, x).$$

Next, from the construction in Theorem 6.1 (ii), it is clear that $\Upsilon^*_t = Y_t^{t, X^*_t, \alpha^*_{[t,T]}}$. Then, since $\mathbb{V}(t, X^*_t)$ is convex, by (7.4) we see that

$$\begin{aligned}
&\Lambda(t, X^*_{[0,t]}) \cdot Y_t^{t, X^*_t, \alpha^*_{[t,T]}} = \Lambda(t, X^*_{[0,t]}) \cdot \Upsilon^*_t \\
&= \sup_{y \in \mathbb{V}(t, X^*_t)} \Lambda(t, X^*_{[0,t]}) \cdot y = \sup_{\alpha_{[t,T]}} \Lambda(t, X^*_{[0,t]}) \cdot Y_t^{t, X^*_t, \alpha_{[t,T]}}.
\end{aligned}$$

This exactly means $\alpha^*_{[t,T]}$ is an optimal control for the dynamic optimization problem (7.3).

We remark that the mapping $\Lambda : [0,T] \times C([0,T]; \mathbb{R}^d) \to \mathbb{R}^m$, which is path dependent in an adapted way, is time consistent in the following sense. Consider the problem at time 0 with initial condition (x, λ). Let X^* and Λ be as above, but denoted as $X^{0,x,\lambda,*}$ and $\Lambda^{0,\lambda}$ to indicate their dependence on the initial conditions. Now fix $0 < t < T$, consider the problem on $[t,T]$ with initial condition $X^{0,x,\lambda,*}_{[0,t]}$ and $\Lambda^{0,\lambda}(t,\cdot)$, we can easily see that the moving scalarization we find following the same procedure coincides with the original Λ found at time 0:

$$X_s^{t, X^{0,x,\lambda,*}_{[0,t]}, \Lambda^{0,\lambda}(t,\cdot)} = X_s^{0,x,\lambda,*}, \quad \Lambda^{t, \Lambda^{0,\lambda}(t,\cdot)}(s, \cdot) = \Lambda^{0,\lambda}(s,\cdot), \quad t \le s \le T.$$

Remark 7.1 *When $\mathbb{V}(t, X^*_t)$ is nonconvex, as in Example 4.7, the Λ in (7.4) can be viewed as a local asymptotic moving scalarization in the following sense:*

$$\Lambda(t, X^*_{[0,t]}) \cdot \Upsilon^*_t \ge \Lambda(t, X^*_{[0,t]}) \cdot y - o(|y - \Upsilon^*_t|), \quad \forall y \in \mathbb{V}(t, X^*_t); \quad \textit{or equivalently,}$$

$$\Lambda(t, X^*_{[0,t]}) \cdot Y_t^{t, X^*_t, \alpha^*_{[t,T]}} \ge \Lambda(t, X^*_{[0,t]}) \cdot Y_t^{t, X^*_t, \alpha_{[t,T]}} - o\big(\big|Y_t^{t, X^*_t, \alpha_{[t,T]}} - Y_t^{t, X^*_t, \alpha^*_{[t,T]}}\big|\big), \forall \alpha_{[t,T]}.$$

■

Remark 7.2 *(i) When the φ in (7.1) is nonlinear, since $\mathbb{V}(0,x_0)$ is compact, one may still find an optimal argument $y_0 \in \mathbb{V}(0,x_0)$ for the problem in the right side of (7.1). We emphasize that it is possible that $y_0 \in \mathbb{V}_o(0,x_0)$ and such y_0 may not be unique. Fix an arbitrary $\alpha^0 \in \mathcal{A}_0$ and Z^0, for example $\alpha^0 \equiv a_0 \in A$ and $Z^0 \equiv 0$. Denote $X^0 := X^{0,x_0,\alpha^0}$ and*

$$Y^0_t = y_0 - \int_0^t f(s, X^0_s, Y^0_s, Z^0_s, \alpha^0_s) ds + \int_0^t Z^0_s dB_s,$$
$$\tau_0 := \inf\big\{t \geq 0 : (t, X^0_t, Y^0_t) \in \mathbb{G}_{\mathbb{V}}\big\}.$$

It is clear that $\tau_0 \leq T$ and $(\tau, X^0_\tau, Y^0_\tau) \in \mathbb{G}_{\mathbb{V}}$. Applying Theorem 6.1 (ii) on $(\tau, X^0_\tau, Y^0_\tau)$ (assuming all the conditions are satisfied) and following the measurable selection theorem we may construct α^* on $[\tau_0, T]$ with initial condition $(\tau, X^0_\tau, Y^0_\tau)$. Then one can easily see that $\alpha^0 \oplus_{\tau_0} \alpha^*$ is an optimal argument for the left side of (7.1). That is, Theorem 6.1 (ii) can help us to construct an optimal control for (7.1) even when φ is nonlinear. However, in this case it is not clear how to construct naturally a (nonlinear) moving scalarization Φ as in (7.2). In particular, when φ has certain structure, for example the linear quadratic structure for the mean variance problem in Remark 4.2, we may naturally expect Φ to have the same structure, which will add the difficulty for constructing a desired Φ.

(ii) For some nonlinear φ, it is possible to linearize it through certain transformation. Indeed, let ψ be a diffeomorphism[11] *on $\mathbb{R}^m$ and set $\tilde{\mathbb{V}}(t,x) := \{\psi(y) : y \in \mathbb{V}(t,x)\}$. Then*

$$\sup_{y \in \mathbb{V}(t,x)} \varphi(y) = \sup_{\tilde{y} \in \tilde{\mathbb{V}}(t,x)} \tilde{\varphi}(\tilde{y}), \quad \text{where} \quad \tilde{\varphi}(\tilde{y}) := \varphi(\psi^{-1}(\tilde{y})).$$

If one can choose ψ such that $\tilde{\varphi}$ is linear, then one can apply the analysis in this subsection to find a linear moving scalarization $\tilde{\Phi}$ for $\tilde{\mathbb{V}}$, which leads to a desired nonlinear moving scalarization for the original $\mathbb{V}$: $\Phi(t, X^*_{[0,t]}, y) := \tilde{\Phi}(t, X^*_{[0,t]}, \psi(y))$. We remark that X^* stands for X^{α^*} for some optimal control α^*, so it remains the same after the transformation. However, in this case $\mathbf{n}_{\mathbb{V}}(t, X^*_t, \Upsilon^*_t)$ does not lead to a desired moving scalarization. ■

7.2 The mean variance problem

In this subsection we find a desired moving scalarization for the mean variance problem in Remark 4.2, by employing the idea in Remark 7.2 (ii). We first remark that in this case $\mathbb{V}$ is not bounded. However, since $\mathbb{V}$ is explicit as in (4.8), we may still apply the results in Theorem 6.1 (ii).

Theorem 7.3 *Consider the optimization problem (4.4) and introduce:*

$$\Lambda(t, \mathbf{x}_{[0,t]}) := \frac{\lambda e^{T-t}}{e^T - \lambda(\mathbf{x}_t - \mathbf{x}_0)} \quad \forall \mathbf{x} \in C([0,T], \mathbb{R}) \ \textit{s.t.} \ \sup_{0 \leq t \leq T} [\mathbf{x}_t - \mathbf{x}_0] < \frac{1}{\lambda} e^T. \tag{7.5}$$

Then the following dynamic mean variance problem is time consistent:

$$V_t := \operatorname*{ess\,sup}_{\alpha} \Big\{ \mathbb{E}[X^\alpha_T | \mathcal{F}_t] - \tfrac{1}{2}\Lambda(t, X^*_{[0,t]}) \mathrm{Var}(X^\alpha_T | \mathcal{F}_t) \Big\}. \tag{7.6}$$

[11] We refer to [13, Theorem A] for a characterization of diffeomorphisms.

Here X^* is the optimal trajectory for (4.4) and it satisfies $\sup_{0\le t\le T}[X^*_t - x_0] < \frac{1}{\lambda}e^T$, *a.s. Moreover, the optimal control and the optimal value are*[12]*:*

$$\begin{aligned} \alpha^*_t &= -X^*_t + x_0 + \tfrac{1}{\lambda}e^T; \\ V_t &= \tfrac{1}{2}(1+e^{t-T})X^*_t + \tfrac{1}{2}(1-e^{t-T})x_0 + \tfrac{e^T}{2\lambda}(1-e^{t-T}). \end{aligned} \tag{7.7}$$

Proof In light of Remark 7.2 (ii), we introduce an obvious diffeomorphism

$$\psi(y_1,y_2) := (y_1,\tilde y_2) := (y_1, y_2 - |y_1|^2). \tag{7.8}$$

Then, by (4.8) we have

$$\begin{aligned} \tilde{\mathbb{V}}(t,x) :=& \Big\{\psi(y) : y\in \mathbb{V}(t,x)\Big\} = \Big\{(y_1,\tilde y_2) : y_1\in\mathbb{R}, \tilde y_2 \ge \phi_1(t)(y_1-x)^2\Big\}, \\ \tilde{\mathbb{V}}_b(t,x) =& \Big\{(y_1,\tilde y_2) : y_1\in\mathbb{R}, \tilde y_2 = \phi_1(t)(y_1-x)^2\Big\}, \text{ where } \phi_1(t) := \frac{1}{e^{T-t}-1}. \end{aligned} \tag{7.9}$$

Note that $\tilde{\mathbb{V}}$ is convex, so the concern in Remark 7.1 is irrelevant and we are finding a true moving scalarization. We shall denote $\tilde y = (y_1,\tilde y_2)$, and for $\tilde y \in \tilde{\mathbb{V}}_b(t,x)$, clearly it suffices to specify y_1. Moreover, recall (4.5) and denote $\tilde Y := \psi(Y) = (Y^1,\tilde Y^2)$, by the standard Itô formula we have

$$Y^{t,x,\alpha,1}_s = X^{t,x,\alpha}_T - \int_s^T Z^{t,x,\alpha,1}_r dB_r,\ \tilde Y^{t,x,\alpha,2}_s = \int_s^T |Z^{t,x,\alpha,1}_r|^2 dr - \int_s^T \tilde Z^{t,x,\alpha,2}_r dB_r.$$

That is, in light of (4.1) and denoting $\tilde z = (z_1,\tilde z_2)$,

$$\tilde f(t,x,\tilde y,\tilde z,a) = (0,|z_1|^2)^\top. \tag{7.10}$$

Given (7.9), one can easily compute that

$$\begin{aligned} \tilde{\mathbf{n}}(t,x,\tilde y) := \mathbf{n}_{\tilde{\mathbb{V}}}(t,x,\tilde y) = \frac{1}{\phi_3}\begin{bmatrix}\phi_2\\ -1\end{bmatrix}(t,x,\tilde y), \quad (t,x,\tilde y)\in \mathbb{G}_{\tilde{\mathbb{V}}}, \\ \text{where}\quad \phi_2(t,x,\tilde y) := 2\phi_1(t)(y_1-x), \quad \phi_3 := \sqrt{1+|\phi_2|^2}. \end{aligned} \tag{7.11}$$

Next, fix $(t,x,\tilde y)\in\mathbb{G}_{\tilde{\mathbb{V}}}$ and set $\Upsilon(x') := \big(y_1, \phi_1(t)(y_1-x')^2\big)^\top$, $x'\in\mathbb{R}$. Clearly $\Upsilon(x')\in \tilde{\mathbb{V}}_b(x')$ for all x' and $\frac{d}{dx'}\Upsilon(x')\Big|_{x'=x} = (0,-\phi_2(t,x,\tilde y))^\top$. Then by (2.20) and (2.15) we have

$$\partial_x\tilde{\mathbb{V}}(t,x,\tilde y) = \Big((0,-\phi_2)^\top\cdot\tilde{\mathbf{n}}\Big)\tilde{\mathbf{n}}(t,x,\tilde y) = \frac{\phi_2}{\phi_3^2}\begin{bmatrix}\phi_2\\ -1\end{bmatrix}(t,x,\tilde y).$$

The right side of above and (7.11) provide natural extensions of $\tilde{\mathbf{n}}$ and $\partial_x\tilde{\mathbb{V}}$ on $[0,T]\times\mathbb{R}\times\mathbb{R}^2$. Then by (2.7) and (2.17) we may compute straightforwardly that, at $(t,x,\tilde y)\in\mathbb{G}_{\tilde{\mathbb{V}}}$,

$$\partial_{xx}\tilde{\mathbb{V}} = \frac{2\phi_1}{\phi_3^6}\begin{bmatrix}-2\phi_2\\ 1-|\phi_2|^2\end{bmatrix}, \quad \partial_x\tilde{\mathbf{n}} = -\frac{2\phi_1}{\phi_3^5}\begin{bmatrix}1\\ \phi_2\end{bmatrix}, \quad \partial_{\tilde y}\tilde{\mathbf{n}} = \frac{2\phi_1}{\phi_3^5}\begin{bmatrix}1\\ \\ \phi_2\end{bmatrix}\begin{bmatrix}1 & \phi_2\end{bmatrix}.$$

[12]The optimal control α^* is the same as the static optimal control in [21, Theorem 3.3 (A)], with the correspondence $\alpha^*_t = X^*_t u^s_*(t,X^*_t)$. However, our V_t is neither equal to the static optimal value nor to the dynamic optimal value in [21, Theorem 3.3], except that at $t=0$ it is equal to the static optimal value there.

Moreover, as the tangent space is one dimensional, it is clear that

$$\zeta \in \mathbb{T}_{\tilde{\mathbb{V}}}(t,x,\tilde{y}) \iff \exists \zeta_0 \in \mathbb{R} \text{ such that } \zeta = \zeta_0 \begin{bmatrix} 1 \\ \phi_2 \end{bmatrix}.$$

Then, recalling (5.2) and (7.10), we may compute straightforwardly that

$$\begin{aligned}
&\tilde{\mathbf{n}} \cdot h_{\tilde{\mathbb{V}}}(t,x,\tilde{y},\partial_x\tilde{\mathbb{V}},\partial_{xx}\tilde{\mathbb{V}},a,\zeta) \\
&= \tilde{\mathbf{n}} \cdot \Big[a\partial_x\tilde{\mathbb{V}} + \frac{a^2}{2}\partial_{xx}\tilde{\mathbb{V}} + (0, |a(\partial_x\tilde{\mathbb{V}})_1 + \zeta_1|^2)^\top\Big] - \Big[a\zeta^\top\partial_x\tilde{\mathbf{n}} + \frac{1}{2}\zeta^\top\partial_{\tilde{y}}\tilde{\mathbf{n}}\zeta\Big] \\
&= \frac{\phi_2}{\phi_3^2}a - \frac{\phi_1}{\phi_3^5}a^2 - \frac{1}{\phi_3}(\frac{\phi_2^2}{\phi_3^2}a + \zeta_0)^2 + \frac{2\phi_1}{\phi_3^3}a\zeta_0 - \frac{\phi_1}{\phi_3}\zeta_0^2 \\
&= -\frac{1}{\phi_3}\Big[(1+\phi_1)\zeta_0^2 + \frac{\phi_1+\phi_2^4}{\phi_3^4}a^2 - \frac{2(\phi_1-\phi_2^2)}{\phi_3^2}a\zeta_0 - \phi_2 a\Big].
\end{aligned}$$

This is quadratic in (a,ζ_0), and one may obtain immediately the optimal arguments:

$$\begin{aligned}
I_1^{\tilde{\mathbb{V}}} = a^* &= \frac{(1+\phi_1)\phi_2}{2\phi_1} = (1+\phi_1(t))(y_1 - x), \\
\zeta_0^* = \frac{\phi_2(\phi_1-\phi_2^2)}{2\phi_1\phi_3^2}, &\quad \text{and thus} \quad I_2^{\tilde{\mathbb{V}}} = \zeta^* = \frac{\phi_2(\phi_1-\phi_2^2)}{2\phi_1\phi_3^2}\begin{bmatrix} 1 \\ \phi_2 \end{bmatrix}.
\end{aligned} \tag{7.12}$$

We next derive (6.2) for $\tilde{\mathbb{V}}$, with the solution denoted as $(X^*, \tilde{\Upsilon}^*) = (X^*, \Upsilon^{*,1}, \tilde{\Upsilon}^{*,2})$. Since by Theorem 6.1 we have $\tilde{\Upsilon}_t^* \in \tilde{\mathbb{V}}_b(t, X_t^*)$, it suffices to specify the equations for $(X^*, \Upsilon^{*,1})$. Note that, with $(\cdot)_1$ denoting the first component,

$$\Big(\partial_x\tilde{\mathbb{V}}\sigma(\cdot, I_1^{\tilde{\mathbb{V}}}) + I_2^{\tilde{\mathbb{V}}}\Big)_1(t,x,\tilde{y}) = \frac{\phi_2^2}{\phi_3^2} \times \frac{(1+\phi_1)\phi_2}{2\phi_1} + \frac{\phi_2(\phi_1-\phi_2^2)}{2\phi_1\phi_3^2} = \phi_1(t)(y_1 - x).$$

Note further that $\tilde{f}_1 = 0$. Then, by recalling (6.3) in Remark 6.2 and (4.9), we have

$$\begin{aligned}
X_t^* &= x_0 + \int_0^t (1+\phi_1(s))(\Upsilon_s^{*,1} - X_s^*)ds + \int_0^t (1+\phi_1(s))(\Upsilon_s^{*,1} - X_s^*)dB_s; \\
\Upsilon_t^{*,1} &= x_0 + \frac{1}{\lambda}[e^T - 1] + \int_0^t \phi_1(s)(\Upsilon_s^{*,1} - X_s^*)dB_s.
\end{aligned} \tag{7.13}$$

Thus:

$$\Upsilon_t^{*,1} - X_t^* = \frac{1}{\lambda}[e^T - 1] - \int_0^t (1+\phi_1(s))(\Upsilon_s^{*,1} - X_s^*)ds - \int_0^t (\Upsilon_s^{*,1} - X_s^*)dB_s. \tag{7.14}$$

Clearly we can solve this and hence (7.13) explicitly. More relevantly for the moving scalarization, as in Remark 6.3 (ii) we may rewrite (7.14) as

$$\Upsilon_t^{*,1} - X_t^* = \frac{1}{\lambda}[e^T - 1] - \int_0^t \frac{1}{1+\phi_1(s)}dX_s^* - \int_0^t \phi_1(s)(\Upsilon_s^{*,1} - X_s^*)ds.$$

Then, denoting $\Gamma_t := e^{\int_0^t \phi_1(s)ds} = \frac{e^T-1}{e^T-e^t}$,

$$\begin{aligned}\Gamma_t(\Upsilon^{*,1}_t - X^*_t) &= \frac{1}{\lambda}[e^T-1] - \int_0^t \frac{\Gamma_s}{1+\phi_1(s)} dX^*_s = \frac{1}{\lambda}[e^T-1] - \int_0^t (1-e^{-T}) dX^*_s \\ &= \frac{1}{\lambda}[e^T-1] - (1-e^{-T})X^*_t + (1-e^{-T})x_0.\end{aligned}$$

Thus

$$\Upsilon^{*,1}_t - X^*_t = \frac{1}{\lambda}(e^T - e^t) - (1-e^{t-T})(X^*_t - x_0). \tag{7.15}$$

By abusing the notation Λ with the previous subsection, our goal in this subsection is to find a moving scalarization $\Lambda_t := \Lambda(t, X^*_{[0,t]})$ such that the following dynamic problem

$$\sup_{\tilde y \in \tilde{\mathbb{V}}(t, X^*_t)} \Big(y_1 - \frac{\Lambda_t}{2}\tilde y_2\Big) \text{ is time consistent.} \tag{7.16}$$

From the analysis in the previous subsection, this implies that $(1, -\frac{\Lambda_t}{2})^\top$ is parallel to $\tilde{\mathbf{n}}(t, X^*_t, \Upsilon^*_t)$. By (7.11), this implies that $-\frac{\Lambda_t}{2} = \frac{-1}{\phi_2(t,X^*_t,\Upsilon^*_t)}$. Thus, by (7.15),

$$\Lambda(t, X^*_{[0,t]}) = \Lambda_t = \frac{2}{\phi_2(t, X^*_t, \Upsilon^*_t)} = \frac{e^{T-t}-1}{\Upsilon^{*,1}_t - X^*_t} = \frac{\lambda e^{T-t}}{e^T - \lambda(X^*_t - X^*_0)}.$$

This proves (7.5). We remark that, by (7.14) clearly $\Upsilon^{*,1}_t - X^*_t > 0$, then it follows from (7.15) that $\sup_{0\le t\le T}[X^*_t - x_0] < \frac{1}{\lambda}e^T$, a.s.

For the original $\mathbb{V}$, by (7.16) the following dynamic problem is time consistent:

$$\sup_{y\in\mathbb{V}(t,X^*_t)} \Phi(t, X^*_{[0,t]}, y), \quad \text{where} \quad \Phi(t, X^*_{[0,t]}, y) := y_1 + \frac{\Lambda(t, X^*_{[0,t]})}{2}|y_1|^2 - \frac{\Lambda(t, X^*_{[0,t]})}{2} y_2.$$

This is clearly equivalent to the time consistency of the dynamic problem (7.6).

Finally, plugging (7.15) into the first line of (7.12), we obtain the expression of α^* in (7.7) immediately. Moreover, by (7.9), (7.5), and (7.15) we have

$$\begin{aligned}V_t &= \Upsilon^{*,1}_t - \tfrac{\Lambda_t}{2}\tilde\Upsilon^{*,2}_t = \Upsilon^{*,1}_t - \frac{e^{T-t}-1}{2(\Upsilon^{*,1}_t - X^*_t)} \times \phi_1(t)(\Upsilon^{*,1}_t - X^*_t)^2 \\ &= \tfrac{1}{2}(\Upsilon^{*,1}_t - X^*_t) + X^*_t = \tfrac{1}{2}(1+e^{t-T})X^*_t + \tfrac{1}{2}(1-e^{t-T})x_0 + \tfrac{e^T}{2\lambda}(1-e^{t-T}).\end{aligned}$$

This proves (7.7), and completes the proof of the theorem. ■

8 Further discussions

8.1 The case with nondegenerate terminal

As pointed out in Remark 4.3 (ii), given a general $\mathbb{G} : \mathbb{R}^d \to \mathcal{D}^m_0$, we may define $\mathbb{V}$ by (4.11). This is equivalent to

$$\mathbb{V}(t,x) := \mathrm{cl}\big\{Y^{T,\phi;t,x,\alpha}_t : \alpha \in \mathcal{A}_t, \phi \in \mathbb{L}^2(\mathcal{F}^t_T) \text{ s.t. } \phi \in \mathbb{G}(X^{t,x,\alpha}_T), \text{ a.s.}\big\}. \tag{8.1}$$

Then we have

Theorem 8.1 *Let Assumption 4.4 (i), (ii) hold, and $\mathbb{G}$ is bounded and continuous. Assume the $\mathbb{V}$ defined by (4.11) or (8.1) is in $C_0^{1,2}([0,T)\times\mathbb{R}^d;\mathcal{D}_2^m)$. Then $\mathbb{V}$ is the unique classical solution of the HJB equation (5.3) with terminal condition $\mathbb{V}(T,x)=\mathbb{G}(x)$.*

The proof is essentially the same as in the previous sections, we thus omit it. In particular, when $\mathbb{G}(x)\in\mathcal{D}_2^m$ and $\mathbb{V}\in C^{1,2}([0,T]\times\mathbb{R}^d;\mathcal{D}_2^m)$, the proof is actually slightly easier.

8.2 Comparison with Soner-Touzi [27]

In the contexts of stochastic target problem, [27] derived a geometric equation to characterize the reachable set of the problem. This work is very closely related to our problem. In this subsection we provide some detailed analyses on the connection and the differences between the two works. We shall introduce their approach, but in our contexts and using our notations, and all the discussions are heuristic.

We first note that, the stochastic target problem (4.10) (or the more general one (4.11)) can be rewritten equivalently as:

$$\widehat{\mathbb{V}}(t):=\Big\{(x,y)\in\mathbb{R}^{d+m}:\exists(\alpha,Z)\text{ such that }Y_T^{t,x,y,\alpha,Z}=g(X_T^{t,x,\alpha}),\text{a.s.}\Big\}.$$

Here $(X^{t,x,\alpha},Y^{t,x,y,\alpha,Z})$ becomes a $d+m$-dimensional controlled state process with control (α,Z). It is clear that $\widehat{\mathbb{V}}$ and our $\mathbb{V}$ are equivalent in the following sense:

$$\begin{aligned}&\widehat{\mathbb{V}}(t)=\big\{(x,y):x\in\mathbb{R}^d,y\in\mathbb{V}(t,x)\big\},\quad \widehat{\mathbb{V}}_b(t)=\big\{(x,y):x\in\mathbb{R}^d,y\in\mathbb{V}_b(t,x)\big\};\\&\qquad\text{and}\quad \mathbb{V}(t,x)=\big\{y:(x,y)\in\widehat{\mathbb{V}}(t)\big\},\quad \mathbb{V}_b(t,x)=\big\{y:(x,y)\in\widehat{\mathbb{V}}_b(t)\big\}.\end{aligned}\tag{8.2}$$

Then $\mathbb{G}_{\widehat{\mathbb{V}}}=\mathbb{G}_{\mathbb{V}}$. Naturally we may define, for some $\varepsilon>0$,

$$\begin{aligned}&\mathbf{n}_{\widehat{\mathbb{V}}}(t,x,y):=\mathbf{n}_{\widehat{\mathbb{V}}(t)}(x,y)\in\mathbb{R}^{d+m},\quad (t,x,y)\in\mathbb{G}_{\widehat{\mathbb{V}}};\\&\mathbf{r}_{\widehat{\mathbb{V}}}(t,x,y):=\mathbf{r}_{\widehat{\mathbb{V}}(t)}(x,y)\in\mathbb{R},\quad (t,x,y)\in O_\varepsilon(\mathbb{G}_{\widehat{\mathbb{V}}}).\end{aligned}$$

The work [27] characterized the square of the distance function[13]: $\eta(t,x,y):=\frac12|\mathbf{r}_{\widehat{\mathbb{V}}}(t,x,y)|^2$ by the following PDE: denoting $\hat y:=(x,y)$ and noting the time change in [27],

$$\begin{aligned}&\nabla_{\hat y}\nabla_t\eta(t,\hat y)+\nabla_{\hat y}\Big[F(t,\hat y,\nabla_{\hat y}\eta(t,\hat y),\nabla_{\hat y\hat y}\eta(t,\hat y))\Big]=0,\quad (t,\hat y)\in\mathbb{G}_{\widehat{\mathbb{V}}},\quad\text{where}\\&F(t,\hat y,\nabla_{\hat y}\eta,\nabla_{\hat y\hat y}\eta):=\inf_{(a,z)\in\mathcal{N}(t,\hat y,\nabla_{\hat y}\eta)}\Big[b(t,x,a)\cdot\nabla_x\eta(t,\hat y)-f(t,\hat y,z,a)\cdot\nabla_y\eta(t,\hat y)\\&\quad+\frac12\mathrm{tr}\left(\sigma^\top(t,x,a)\nabla_{xx}\eta(t,\hat y)\sigma(t,x,a)+2z^\top\nabla_{xy}\eta(t,\hat y)\sigma(t,x,z)+z^\top\nabla_{yy}\eta(t,\hat y)z\right)\Big],\\&\mathcal{N}(t,\hat y,\nabla_{\hat y}\eta):=\Big\{(a,z):[\sigma^\top(t,x,a),z^\top]^\top[\sigma^\top(t,x,a),z^\top]\nabla_{\hat y}\eta(t,\hat y)=0\Big\}.\end{aligned}\tag{8.3}$$

We remark that (8.3) holds only on $\mathbb{G}_{\widehat{\mathbb{V}}}$, and thus it is also not a standard PDE.

We first note that $\hat{\mathbb{V}}$ is a function of t only, so it does not invoke the set-valued Itô formula. While this may seem to be technically easier, the set-valued Itô formula has

[13] The reason to consider the squared function is that, in the degenerate case, $\mathbf{r}_{\widehat{\mathbb{V}}}$ is typically not smooth while $|\mathbf{r}_{\widehat{\mathbb{V}}}|^2$ is. In the nondegenerate case as in this paper, actually one may study $\mathbf{r}_{\widehat{\mathbb{V}}}$ directly.

independent interest and is one of the main contributions of this paper. For example, it provides microstructure of the flow on the boundary surface, as we see in Theorem 6.1 (ii). More importantly, the roles of x and y are symmetric in $\hat{\mathbb{V}}(t)$. In many applications, however, the input x and the output y play different roles: x is the observed state, while the value y (more precisely the set of y) is our objective, so it is more natural to study $\mathbb{V}(t,x)$.

Technically, the major difference is that, as we see in Example 8.2 below,

$$\mathbf{r}_{\widehat{\mathbb{V}}}(t,x,y) \neq \mathbf{r}_{\mathbb{V}}(t,x,y).$$

In general, recalling (8.2) we have

$$\begin{aligned}|\mathbf{r}_{\widehat{\mathbb{V}}}(t,x,y)|^2 &= \inf_{(\tilde{x},\tilde{y})\in\widehat{\mathbb{V}}_b(t)} \Big[|x-\tilde{x}|^2 + |y-\tilde{y}|^2\Big] = \inf_{\tilde{x}\in\mathbb{R}^d}\inf_{\tilde{y}\in\mathbb{V}_b(t,\tilde{x})} \Big[|x-\tilde{x}|^2 + |y-\tilde{y}|^2\Big]\\ &= \inf_{\tilde{x}\in\mathbb{R}^d} \Big[|x-\tilde{x}|^2 + |\mathbf{r}_{\mathbb{V}}(t,\tilde{x},y)|^2\Big] \le |\mathbf{r}_{\mathbb{V}}(t,x,y)|^2\end{aligned}$$

Example 8.2 *Set $d=m=1$ and consider time invariant set values:*

$$\mathbb{V}(x) = [x-1, x+1] \subset \mathbb{R}, \quad \widehat{\mathbb{V}} := \{(x,y) : x\in\mathbb{R}, y\in[x-1,x+1]\} \subset \mathbb{R}^2.$$

Clearly $\mathbb{V}_b(x) = \{x-1, x+1\}$. One can easily verify that,

$$\begin{aligned}&\mathbf{r}_{\mathbb{V}}(x,y) = y-(x+1), \quad \textit{for } y\approx x+1; \quad \textit{and} \quad \mathbf{r}_{\mathbb{V}}(x,y) = (x-1)-y, \quad \textit{for } y\approx x-1;\\ &\mathbf{r}_{\widehat{\mathbb{V}}}(x,y) = \frac{y-(x+1)}{\sqrt{2}}, \quad \textit{for } y\approx x+1; \quad \textit{and} \quad \mathbf{r}_{\widehat{\mathbb{V}}}(x,y) = \frac{(x-1)-y}{\sqrt{2}}, \quad \textit{for } y\approx x-1.\end{aligned}$$

We also observe directly from above that, although $\mathbf{r}_{\mathbb{V}} = \mathbf{r}_{\widehat{\mathbb{V}}} = 0$ on $\mathbb{G}_{\mathbb{V}} = \mathbb{G}_{\widehat{\mathbb{V}}}$, their derivatives are in general not equal. ■

Consequently, although both (5.4) and (8.3) characterize the same set $\mathbb{G}_{\mathbb{V}} = \mathbb{G}_{\widehat{\mathbb{V}}}$, the two equations are different. This is partially explained by the above observation that $\mathbf{r}_{\mathbb{V}}$ and $\mathbf{r}_{\widehat{\mathbb{V}}}$ have different derivatives on $\mathbb{G}_{\mathbb{V}}$. At below we provide more detailed calculation for the set-valued heat equation in Example 2.9 (ii), but with $d=m=1$.

Example 8.3 *Set $d=m=1$, $b=0$, $\sigma=1$, $f=0$, and the terminal $\mathbb{G}(x) = [-\psi(x), \psi(x)]$, where $\psi:\mathbb{R}\to(0,\infty)$ is smooth. Then, similar to Example 2.9 (ii), we have*

$$\mathbb{V}(t,x) = [-u(t,x), u(t,x)], \quad \mathbb{V}_b(t,x) = \{-u(t,x), u(t,x)\},$$

where u is the unique classical solution of the heat equation

$$\partial_t u + \frac{1}{2}\partial_{xx} u = 0, \quad u(T,x) = \psi(x). \tag{8.4}$$

We shall prove in Appendix that $\widehat{\mathbf{r}} := \mathbf{r}_{\widehat{\mathbb{V}}}$ satisfies the following equation:

$$\nabla_t\widehat{\mathbf{r}} + \frac{1}{2}\Big[\nabla_{xx}\widehat{\mathbf{r}} - 2\nabla_{xy}\widehat{\mathbf{r}}\frac{\nabla_x\widehat{\mathbf{r}}}{\nabla_y\widehat{\mathbf{r}}} + \nabla_{yy}\widehat{\mathbf{r}}\Big|\frac{\nabla_x\widehat{\mathbf{r}}}{\nabla_y\widehat{\mathbf{r}}}\Big|^2\Big] = 0, \quad \textit{on } \mathbb{G}_{\widehat{\mathbb{V}}}. \tag{8.5}$$

■

In this scalar case, by Remark 5.1 (i) we see that the set-valued HJB equation (5.4) reduces back to the standard PDE for $\overline{v}(t,x) = u(t,x)$ and $\underline{v}(t,x) = -u(t,x)$, both of which identify with the heat equation (8.4). So (5.4) is indeed a natural extension of the HJB equation to the multivariate case. The equation (8.5), however, is quite different from (8.4). So in this sense, it is more natural to study (5.4) than to study (8.3).

Another advantage of (5.4) is that, as we saw in Section 7, the normal vector $\mathbf{n}_{\mathbb{V}}(t, X_t^*, \Upsilon_t^*)$ provides naturally a moving scalarization for the time inconsistent multivariate optimization problem. The vector $\mathbf{n}_{\widehat{\mathbb{V}}}$ (at certain optimal paths) does not serve for this purpose. In fact, $\mathbf{n}_{\widehat{\mathbb{V}}} \in \mathbb{R}^{d+m}$, while a moving scalarization Λ is by nature m-dimensional.

Finally, we remark that [27, Theorem 2.1] showed that $\widehat{\mathbb{V}}$ is the unique classical solution of (8.3), provided that optimal controls exist, in the same spirit of Theorem 6.1 (ii). We instead proved the existence and uniqueness under weak conditions in Theorems 5.6 and 6.1 (i).

8.3 Comparison with Ararat-Ma-Wu [1]

Mainly motivated by dynamic set-valued risk measures for multi-asset financial models, [1] studied the following set-valued BSDE:

$$\mathbb{Y}_t = \mathbb{E}\Big[\mathbb{G}(B_T) + \int_t^T \mathbb{F}(s, B_s, \mathbb{Y}_s)ds \big| \mathcal{F}_t\Big]. \tag{8.6}$$

Here, denoting by $\mathcal{D}_{cc}^m$ the space of convex compact sets $\mathbb{D} \in \mathcal{D}_0^m$, the terminal $\mathbb{G} : \mathbb{R}^d \to \mathcal{D}_{cc}^m$, and the driver $\mathbb{F} : [0,T] \times \mathbb{R}^d \times \mathcal{D}_{cc}^m \to \mathcal{D}_{cc}^m$ (abusing the notation $\mathbb{F}$ here). We note that [1] actually allows $\mathbb{G}$ and $\mathbb{F}$ to depend on the paths of B. By relying on the sophisticated set-valued stochastic analysis, especially the Hukuhara difference, [1] established the well-posedness of the above set-valued BSDE. The general case that $\mathbb{F}$ depends on $\mathbb{Z}$, and the martingale representation with the term $\mathbb{Z}_t dB_t$ seem to be a quite remote goal.

Formally, the set-valued BSDE (8.6) is associated with our set-valued HJB equation (5.3) in the case $x_0 = 0$, $b = 0$, $\sigma = 1$, $f = f(t,x,y,a)$. Then $X = B$, and we may naturally define

$$\mathbb{Y}_t := \mathbb{V}(t, B_t), \quad \mathbb{F}(t,x,\mathbb{D}) := \big\{f(t,x,y,a) : y \in \mathbb{D}, a \in A\big\}. \tag{8.7}$$

In the linear case: $f = f(t,x,a)$ and thus $\mathbb{F}(t,x) = \big\{f(t,x,a) : a \in A\big\}$ is independent of $\mathbb{D}$, the random set-valued process $\mathbb{Y}_t := \mathbb{V}(t, B_t)$ indeed satisfies (8.6) in the sense of [1].

However, when f depends on y, the $\mathbb{Y}, \mathbb{F}$ in (8.7) do not satisfy (8.6). That is, (8.6) is not the stochastic counterpart of (5.3). The reason is the same as in Remark 2.3 (ii). In (4.1), the Y in the left side and that in the right side of the BSDE are required to be the same process. In (8.6), however, one allows to consider different selectors for the $\mathbb{Y}$ in the left side and that in the right side of the equation. Consequently, the solutions to (5.3) and to (8.6) are typically not equal. We shall remark that, the applications mentioned in Introduction typically fall into our framework, although technically many of them are not covered by the current form of our HJB equation (5.3).

A Some technical proofs

Proof of Proposition 2.5. Again we denote $\mathbf{r}, \mathbf{n}$ for notational simplicity. We prove it only for $x > x_0$. Fix $x_1 > x_0$. Without loss of generality, we assume θ is absolutely continuous in $x \in [x_0, x_1]$ with appropriate derivative function θ', as otherwise the length of θ would be ∞. Thus we have

$$\theta(x) = y_0 - \int_{x_0}^{x} \theta'(\tilde{x}) d\tilde{x},$$

Note that $\mathbf{r}(x, \theta(x)) = 0$. Then, for Lebesgue-a.e. x,

$$\begin{aligned} 0 &= \frac{d}{dx}\mathbf{r}(x, \theta(x)) = \nabla_x \mathbf{r}(x, \theta(x)) - \nabla_y \mathbf{r}(x, \theta(x)) \cdot \theta'(x) \\ &= \nabla_x \mathbf{r}(x, \theta(x)) - \mathbf{n}(x, \theta(x)) \cdot \theta'(x), \end{aligned}$$

and thus

$$\zeta(x) := \theta'(x) - \nabla_x \mathbf{r}\ \mathbf{n}(x, \theta(x)) \in \mathbb{T}_{\mathbb{V}}(x, \theta(x)).$$

Therefore,

$$\begin{aligned} L_\theta(x_0, x_1) &= \int_{x_0}^{x_1} \sqrt{1 + |\theta'(x)|^2}\ dx = \int_{x_0}^{x_1} \sqrt{1 + \big|\nabla_x \mathbf{r}\ \mathbf{n}(x, \theta(x)) + \zeta(x)\big|^2}\ dx \\ &= \int_{x_0}^{x_1} \sqrt{1 + \big|\nabla_x \mathbf{r}(x, \theta(x))\big|^2 + |\zeta(x)|^2}\ dx \geq \int_{x_0}^{x_1} \sqrt{1 + \big|\nabla_x \mathbf{r}(x, \theta(x))\big|^2}\ dx \end{aligned}$$

This implies that

$$\begin{aligned} &\overline{\lim_{x_1 \downarrow x_0}} \frac{1}{x_1 - x_0} \Big[L_\Upsilon(x_0, x_1) - L_\theta(x_0, x_1)\Big] \\ &\leq \overline{\lim_{x_1 \downarrow x_0}} \frac{1}{x_1 - x_0} \Big[\int_{x_0}^{x_1} \sqrt{1 + \big|\nabla_x \mathbf{r}(x, \Upsilon(x))\big|^2}\ dx - \int_{x_0}^{x_1} \sqrt{1 + \big|\nabla_x \mathbf{r}(x, \theta(x))\big|^2}\ dx\Big] \\ &\leq \overline{\lim_{x_1 \downarrow x_0}} \frac{1}{x_1 - x_0} \Big[\int_{x_0}^{x_1} \Big|\sqrt{1 + \big|\nabla_x \mathbf{r}(x, \Upsilon(x))\big|^2} - \sqrt{1 + \big|\nabla_x \mathbf{r}(x_0, y_0)\big|^2}\Big|\ dx \\ &\qquad + \int_{x_0}^{x_1} \Big|\sqrt{1 + \big|\nabla_x \mathbf{r}(x, \theta(x))\big|^2} - \sqrt{1 + \big|\nabla_x \mathbf{r}(x_0, y_0)\big|^2}\Big|\ dx\Big] = 0. \end{aligned}$$

■

Proof of Lemma 2.7. Recall (2.4) and consider the natural extension $\widehat{\mathbf{n}} = \nabla_y \mathbf{r}$. By (2.20), (2.18), and (2.17) we have, for $i, j = 1, \cdots, d$, and $(t, x, y) \in \mathbb{G}_{\mathbb{V}}$,

$$\begin{aligned} \partial_{x_i x_j} \mathbb{V}(t, x, y) &= -\partial_{x_i}\big(\nabla_{x_j} \mathbf{r}\mathbf{n}\big)(t, x, y) = -\Big[\partial_{x_i}(\nabla_{x_j} \mathbf{r})\mathbf{n} + \nabla_{x_j} \mathbf{r}\partial_{x_i} \mathbf{n}\Big](t, x, y) \\ &= -\Big[\nabla_{x_i x_j} \mathbf{r} + \nabla_{x_i} r \nabla_{x_j y} \mathbf{r} \cdot \mathbf{n}\Big]\mathbf{n}(t, x, y) - \nabla_{x_j} \mathbf{r}\partial_{x_i} \mathbf{n}(t, x, y). \end{aligned} \tag{A.1}$$

Recall (2.4), at $(t, x, y) \in O_\varepsilon(\mathbb{G}_{\mathbb{V}})$ we have

$$\nabla_{x_j y} \mathbf{r} \cdot \nabla_y \mathbf{r} = \frac{1}{2}\nabla_{x_j}\big(|\nabla_y \mathbf{r}|^2\big) = 0.$$

In particular, $\nabla_{x_j y}\mathbf{r}\cdot\mathbf{n}(t,x,y) = 0$ for $(t,x,y)\in\mathbb{G}_{\mathbb{V}}$. Plugging this into (A.1) we obtain (2.22) immediately.

Moreover, again considering the extension $\widehat{\mathbf{n}}^i = \nabla_{y_i}\mathbf{r}$, by (2.17) and (2.7) we have

$$\partial_x n^i = \nabla_{xy_i}\mathbf{r} - \nabla_x\mathbf{r}(\nabla_{y_i y}\mathbf{r}\cdot\mathbf{n}), \qquad \partial_y \mathbf{n}^i = (\mathbf{I}_m - \mathbf{n}\mathbf{n}^\top)\nabla_{y_i y}\mathbf{r}.$$

Similarly, by (2.4) we have $\nabla_{y_i y}\mathbf{r}\cdot\mathbf{n} = 0$, which implies (2.23) immediately. ∎

Proof of Proposition 4.5. Under Assumption 4.4, clearly $\mathbb{V}(t,x)$ is bounded. Then the compactness follows from its closeness. To show the convexity, we recall (4.10) and introduce

$$J(t,x,y,\alpha,Z) := \mathbb{E}\Big[\big|Y_T^{t,x,y,\alpha,Z} - g(X_T^{t,x,\alpha})\big|^2\Big],\ W(t,x,y) := \inf_{\alpha,Z} J(t,x,y,\alpha,Z). \tag{A.2}$$

It is clear that $W\ge 0$. Following [16] we have the duality:

$$\mathbb{V}(t,x) = \{y : W(t,x,y) = 0\}. \tag{A.3}$$

Then it suffices to show that W is convex in y.[14] Indeed, for any $\underline{y},\overline{y}\in\mathbb{V}(t,x)$ and $y := \theta\underline{y} + (1-\theta)\overline{y}$, $\theta\in(0,1)$, by (A.3) we have $W(t,x,\underline{y}) = W(t,x,\overline{y}) = 0$, and then by the convexity of W we have

$$0 \le W(t,x,y) \le \theta W(t,x,\underline{y}) + (1-\theta)W(t,x,\overline{y}) = 0.$$

Thus by (A.3) again we have $y\in\mathbb{V}(t,x)$. That is, $\mathbb{V}(t,x)$ is convex.

We now prove the convexity of W in two steps. Throughout this proof, ρ denotes a generic modulus of continuity function which may vary from line to line.

Step 1. In this step we establish the DPP and the desired regularity of W. First, note that $J(t,\cdot,\alpha,Z)$ is uniformly continuous in (x,y), and thus W is uniformly continuous in (x,y). By this regularity, it follows from standard arguments to have the DPP for deterministic time:

$$W(t,x,y) = \inf_{\alpha,Z}\mathbb{E}\Big[W(t+\delta, X_{t+\delta}^{t,x,\alpha}, Y_{t+\delta}^{t,x,y,\alpha,Z})\Big]. \tag{A.4}$$

By setting $\alpha\equiv a_0\in A$ and $Z\equiv 0$ in (A.4), we have,

$$W(t,x,y) \le \mathbb{E}\Big[W(t+\delta, X_{t+\delta}^{t,x,a_0}, Y_{t+\delta}^{t,x,y,a_0,0})\Big]. \tag{A.5}$$

This implies immediately that $W(t,x,y)\le \underline{\lim}_{\delta\downarrow 0} W(t+\delta,x,y)$. Moreover, for any α, Z, by the tower properties of SDEs and conditional expectations, it is clear that

$$J(t,x,y,\alpha,Z) \ge \mathbb{E}\Big[W(t+\delta, X_{t+\delta}^{t,x,\alpha}, Y_{t+\delta}^{t,x,y,\alpha,Z})\Big].$$

Send $\delta\to 0$ we have $J(t,x,y,\alpha,Z)\ge \overline{\lim}_{\delta\downarrow 0} W(t+\delta,x,y)$. Then $W(t,x,y)\ge \overline{\lim}_{\delta\downarrow 0} W(t+\delta,x,y)$. Thus W is right continuous in t:

$$\lim_{\delta\downarrow 0} W(t+\delta,x,y) = W(t,x,y). \tag{A.6}$$

[14]The convexity of W is a sufficient but not a necessary condition of the convexity of $\mathbb{V}$, since the latter involves only the zero level set of W.

We note that, by (A.4), similarly to (A.5) we can show that $\overline{\lim}_{\delta\downarrow 0} W(t-\delta,x,y) \le W(t,x,y)$. So W is upper semicontinuous in t. However, in general it is not clear whether or not W is left continuous in t.

Step 2. We now prove the convexity of W and we use two distinct approaches: the first one based on the viscosity solution theory and the second one on probabilistic arguments. Both approaches have independent interest.

Method 1. By DPP (A.4) and the upper semicontinuity of W in Step 1, it is clear that W is a viscosity subsolution to the following standard HJB equation with unbounded control z:

$$\begin{aligned}\partial_t W(t,x,y) + \inf_{a,z}\Big[&\mathrm{tr}\left(\tfrac{1}{2}\partial_{xx}W\sigma\sigma^\top(t,x,a) + \tfrac{1}{2}\partial_{yy}W zz^\top + \partial_{xy}W\sigma(t,x,a)z^\top\right)\\ &+\partial_x W\cdot b(t,x,a) - \partial_y W\cdot f(t,x,y,z,a)\Big] \ge 0.\end{aligned} \tag{A.7}$$

Here, by abusing notations we use ∂_t, ∂_x etc to denote standard partial derivatives. For any (t,x,y) and any smooth test function φ at (t,x,y), set $a := a_0$ for some fixed $a_0\in A$ and $z := n[e_1,\cdots,e_1]$, where $e_1 := (1,0,\cdots,0)^\top$. Then, by sending $n\to\infty$, it follows from the Lipschitz continuity of f in z that

$$\partial_{y_1y_1}\varphi(t,x,y) \ge 0.$$

Now fix (t,x,y) and introduce the function $U(\tilde y_1) := W(t,x,(\tilde y_1,y_2,\cdots,y_m)^\top)$, $\tilde y_1\in\mathbb{R}$. Then U is uniformly continuous, and it follows from [28, Proposition 5.10] that U is a viscosity subsolution of $U''(\tilde y_1)\ge 0$. Applying [28, Lemma 5.25], we obtain that U is convex in $\tilde y_1$, which implies that W is convex in y_1.

Moreover, for any orthogonal matrix $Q\in\mathbb{R}^{m\times m}$, introduce $\tilde W(t,x,y) := W(t,x,Qy)$. It is clear that $\tilde W$ satisfies a PDE similar to (A.7), with a quadratic term $\frac{1}{2}\mathrm{tr}\left(\partial_{yy}\tilde W(Qz)(Qz)^\top\right)$. Then, by choosing $z = nQ^{-1}[e_1,\cdots,e_1]$, one can show as above that $\tilde W$ is convex in y_1. Now, for any $\underline{y},\overline{y}\in\mathbb{R}^m$ and $y=\theta\underline{y}+(1-\theta)\overline{y}$, $\theta\in[0,1]$, set Q such that $Q^{-1}(\overline{y}-\underline{y}) = |\overline{y}-\underline{y}|e_1$. Then, $Q^{-1}y = Q^{-1}\overline{y} - \theta Q^{-1}(\overline{y}-\underline{y}) = Q^{-1}\overline{y} - \theta|\overline{y}-\underline{y}|e_1$, namely θ affects only the first component. Therefore, by the claimed convexity of $\tilde W$, we have

$$\begin{aligned} W(t,x,y) &= \tilde W(t,x,Q^{-1}y) \le \theta W(t,x,Q^{-1}\underline{y}) + (1-\theta)W(t,x,Q^{-1}\overline{y})\\ &= \theta W(t,x,\underline{y}) + (1-\theta)W(t,x,\overline{y})\end{aligned}$$

This is the desired convexity.

Method 2. Fix $\underline{y}\neq\overline{y}\in\mathbb{R}^m$ and $y=\theta\underline{y}+(1-\theta)\overline{y}$ for some $\theta\in(0,1)$. For any $\delta>0$, α, and n, set $\eta := \frac{\overline{y}-\underline{y}}{|\overline{y}-\underline{y}|}$, $X := X^{0,x_0,\alpha}$, $Y^n := Y^{0,x_0,y,\alpha,n[\eta,\cdots,\eta]}$, and

$$\begin{aligned}\tau^n_\delta &:= \inf\left\{t>0 : \eta\cdot Y^n_t = \eta\cdot\underline{y} \text{ or } \eta\cdot\overline{y}\right\}\wedge\delta,\\ Z^{n,\delta}_t &:= n[\eta,\cdots,\eta]\mathbf{1}_{[0,\tau^n_\delta]}(t), \quad Y^{n,\delta} := Y^{0,x_0,y,\alpha,Z^{n,\delta}}.\end{aligned} \tag{A.8}$$

By DPP (A.4) we have, for some appropriate modulus of continuity function ρ,

$$\begin{aligned} W(0,x_0,y) &\le \mathbb{E}\Big[W(\delta,X_\delta,Y^{n,\delta}_\delta)\Big] \le \mathbb{E}\Big[W(\delta,x_0,Y^{n,\delta}_\delta)\Big] + \rho(\delta)\\ &= \mathbb{E}\Big[W(\delta,x_0,Y^{n,\delta}_\delta)\big[\mathbf{1}_{E^{n,\delta}_1} + \mathbf{1}_{E^{n,\delta}_2} + \mathbf{1}_{\{\tau^n_\delta=\delta\}}\big]\Big] + \rho(\delta),\end{aligned} \tag{A.9}$$

$$\text{where}\quad E^{n,\delta}_1 := \{\tau^n_\delta<\delta, \eta\cdot Y^n_{\tau^n_\delta} = \eta\cdot\underline{y}\},\quad E^{n,\delta}_2 := \{\tau^n_\delta<\delta, \eta\cdot Y^n_{\tau^n_\delta} = \eta\cdot\overline{y}\}.$$

Here the second inequality used the regularity of W in x and the following obvious estimate:

$$\mathbb{E}[|X_\delta - x|^2] \le C\delta.$$

We next prove a crucial estimate:

$$|Y_t^{n,\delta}| \le C(1 + n\tau_\delta^n), \quad \text{and} \quad |f(t, X_t, Y_t^{n,\delta}, Z_t^{n,\delta}, \alpha_s)| \le Cn, \quad t \le \tau_\delta^n. \tag{A.10}$$

Indeed, by (A.8) $\eta \cdot Y_t^{n,\delta}$ is bounded. Denote $\tilde{Y}^{n,\delta} := (\eta \cdot Y^{n,\delta})\eta = \eta\eta^\top Y^{n,\delta}$ and $\Delta Y^{n,\delta} := Y^{n,\delta} - \tilde{Y}^{n,\delta}$. Note that $\eta\eta^\top Z^{n,\delta} = Z^{n,\delta}$. Then, for $0 \le t \le \tau_\delta^n$,

$$\begin{aligned}
\tilde{Y}_t^{n,\delta} &= \eta\eta^\top y + \int_0^t \eta\eta^\top f ds + \int_0^t Z_s^{n,\delta} dB_s; \\
\Delta Y_t^{n,\delta} &= (\mathbf{I}_m - \eta\eta^\top) y + \int_0^t (\mathbf{I}_m - \eta\eta^\top) f(s, X_s, \tilde{Y}_s^{n,\delta} + \Delta Y_s^{n,\delta}, Z_s^{n,\delta}, \alpha_s) ds.
\end{aligned}$$

Since $|\tilde{Y}^{n,\delta}| \le C$ and $|Z^{n,\delta}| \le n$, it follows from the Lipschitz property of f that $|\Delta Y_t^{n,\delta}| \le C(1+nt)$. This, together with the boundedness of $\tilde{Y}^{n,\delta}$ again, implies $|Y_t^{n,\delta}| \le C(1+nt) \le C(1+n\tau_\delta^n)$. Moreover, by Assumption 4.4 and $|Z^{n,\delta}| \le n$, we obtain (A.10).

We now estimate the terms in (A.9). We first consider $E_1^{n,\delta}$. In this case, note that $\eta \cdot (Y_{\tau_\delta^n}^n - \underline{y}) = 0$. Moreover, $(\mathbf{I}_m - \eta\eta^\top)(Y_0^n - \underline{y}) = (\mathbf{I}_m - \eta\eta^\top)(y - \underline{y}) = (1-\theta)|\overline{y} - \underline{y}|(\mathbf{I}_m - \eta\eta^\top)\eta = 0$ and $(\mathbf{I}_m - \eta\eta^\top) Z_t^{n,\delta} = 0, 0 \le t \le \tau_\delta^n$. Then, by (A.10) we have

$$|Y_{\tau_\delta^n}^n - \underline{y}| = \big|(\mathbf{I}_m - \eta\eta^\top)(Y_{\tau_\delta^n}^n - \underline{y})\big| = \Big|\int_0^{\tau_\delta^n} (\mathbf{I}_m - \eta\eta^\top) \cdot f ds\Big| \le Cn\tau_\delta^n. \tag{A.11}$$

Note further that $Z_t^{n,\delta} = 0$, $\tau_\delta^n \le t < \delta$, then by Assumption 4.4, (4.10) and (A.10) we have

$$\begin{aligned}
|Y_t^{n,\delta}| &\le C(|Y_{\tau_\delta^n}^{n,\delta}| + t - \tau_\delta^n) \le C(1 + \delta + n\tau_\delta^n), \quad \tau_\delta^n \le t \le \delta; \\
\big|Y_\delta^{n,\delta} - Y_{\tau_\delta^n}^{n,\delta}\big| &\le \int_{\tau_\delta^n}^\delta |f(s, X_s, Y_s^{n,\delta}, 0, \alpha_s)| ds \le C\delta(1 + n\tau_\delta^n).
\end{aligned}$$

This, together with (A.11), implies

$$\big|Y_\delta^{n,\delta} - \underline{y}\big| \le \big|Y_\delta^{n,\delta} - Y_{\tau_\delta^n}^{n,\delta}\big| + |Y_{\tau_\delta^n}^n - \underline{y}| \le Cn\tau_\delta^n + C\delta, \quad \text{a.s. on } E_1^{n,\delta}. \tag{A.12}$$

Similarly, we have

$$\big|Y_\delta^{n,\delta} - \overline{y}\big| \le Cn\tau_\delta^n + C\delta, \quad \text{a.s. on } E_2^{n,\delta}. \tag{A.13}$$

Plug (A.12), (A.13), and the first estimate of (A.10) into (A.9), we obtain

$$\begin{aligned}
W(0, x_0, y_0) \quad \le \quad & W(\delta, x_0, \underline{y}) \mathbb{P}(E_1^{n,\delta}) + W(\delta, x_0, \overline{y}) \mathbb{P}(E_2^{n,\delta}) \\
& + C\mathbb{E}[n\tau_\delta^n] + C\mathbb{P}(n\tau_\delta^n > \delta) + C(1 + n\delta)\mathbb{P}(\tau_\delta^n = \delta) + \rho(\delta).
\end{aligned} \tag{A.14}$$

Here we used the fact $\rho(n\tau_\delta^n) \le \rho(\delta) + C(\mathbf{1}_{\{n\tau_\delta^n > \delta\}} + n\tau_\delta^n)$, by assuming its linear growth.

We next estimate τ^n_δ. Note that

$$\int_0^t Z^n_s dB_s = n\sqrt{d}\bar{B}_t\eta, \quad \text{where} \quad \bar{B}_t := \frac{1}{\sqrt{d}}\sum_{i=1}^d B^i_t;$$
$$\eta \cdot Y^n_t = \eta \cdot y + \int_0^t \eta \cdot f ds + n\sqrt{d}\bar{B}_t = \eta \cdot y + n\sqrt{d}\bar{B}^n_t, \tag{A.15}$$
$$\text{where} \quad \bar{B}^n_t := \bar{B}_t - \int_0^t \theta^n_s ds, \quad \theta^n_t := -\frac{1}{n\sqrt{d}}\eta \cdot f(t, X_t, Y^n_t, Z^n_t, \alpha_s).$$

Note that $Y^n_t = Y^{n,\delta}_t$ for $t \le \tau^n_\delta$, by (A.10) θ^n_t is bounded for $t \le \tau^n_\delta$. Note further that $\bar{B}$ is a Brownian motion, introduce

$$d\mathbb{P}^n := M^n_t d\mathbb{P}, \quad M^n_t := e^{\int_0^t \theta^n_s d\bar{B}_s - \frac{1}{2}\int_0^t |\theta^n_s|^2 ds}.$$

Then $\bar{B}^n$ is a $\mathbb{P}^n$-Brownian motion. By the middle line of (A.15) and by rescaling the time for the Brownian motion it is straightforward to check that

$$\mathbb{E}^{\mathbb{P}^n}[\tau^n_\delta] \le \frac{C}{n^2}. \tag{A.16}$$

Notice again that θ^n is bounded for $t \le \tau^n_\delta$, then by standard estimates we have $\mathbb{E}^{\mathbb{P}^n}\big[(M^n_{\tau^n_\delta})^{-4}\big] \le C$. This, together with (A.16), implies

$$\mathbb{E}[\tau^n_\delta] \le \mathbb{E}[(\tau^n_\delta)^{\frac{3}{4}}] = \mathbb{E}^{\mathbb{P}^n}\big[(M^n_{\tau^n_\delta})^{-1}(\tau^n_\delta)^{\frac{3}{4}}\big] \le \Big(\mathbb{E}^{\mathbb{P}^n}\big[(M^n_{\tau^n_\delta})^{-4}\big]\Big)^{\frac{1}{4}}\Big(\mathbb{E}^{\mathbb{P}^n}[\tau^n_\delta]\Big)^{\frac{3}{4}} \le \frac{C}{n^{\frac{3}{2}}}. \tag{A.17}$$

In particular, this implies

$$n\mathbb{E}[\tau^n_\delta] \le \frac{C}{\sqrt{n}}, \quad \mathbb{P}(\tau^n_\delta = \delta) \le \frac{1}{\delta}\mathbb{E}[\tau^n_\delta] \le \frac{C}{\delta n^{\frac{3}{2}}}, \quad \text{and} \quad \mathbb{P}(n\tau^n_\delta > \delta) \le \frac{C}{\delta n^{\frac{1}{2}}}. \tag{A.18}$$

Then, by fixing δ and sending $n \to \infty$ in (A.14), we have

$$W(0, x_0, y_0) \le W(\delta, x_0, \underline{y}) \varlimsup_{n\to\infty} \mathbb{P}(E^{n,\delta}_1) + W(\delta, x_0, \overline{y}) \varlimsup_{n\to\infty} \mathbb{P}(E^{n,\delta}_2) + \rho(\delta). \tag{A.19}$$

Moreover, note that

$$\theta\eta \cdot \underline{y} + (1-\theta)\eta \cdot \overline{y} = \eta \cdot y = \eta \cdot \mathbb{E}\Big[Y^n_{\tau^n_\delta} + \int_0^{\tau^n_\delta} f ds\Big]$$
$$= \eta \cdot \underline{y}\mathbb{P}(E^{n,\delta}_1) + \eta \cdot \overline{y}\mathbb{P}(E^{n,\delta}_2) + \eta \cdot \mathbb{E}\Big[Y^n_{\tau^n_\delta}\mathbf{1}_{\{\tau^n_\delta = \delta\}} + \int_0^{\tau^n_\delta} f ds\Big].$$

By the same analysis we have

$$\lim_{n\to\infty} \eta \cdot \mathbb{E}\Big[Y^n_{\tau^n_\delta}\mathbf{1}_{\{\tau^n_\delta=\delta\}} + \int_0^{\tau^n_\delta} f ds\Big] = 0, \quad \lim_{n\to\infty}\Big[\mathbb{P}(E^{n,\delta}_1) + \mathbb{P}(E^{n,\delta}_2)\Big] = 1.$$

Then we must we have

$$\lim_{n\to\infty} |\mathbb{P}(E^{n,\delta}_1) - \theta| = \lim_{n\to\infty} |\mathbb{P}(E^{n,\delta}_2) - (1-\theta)| = 0.$$

Plug this into (A.19), we obtain

$$W(0,x_0,y_0) \le \theta W(\delta,x_0,\underline{y}) + (1-\theta)W(\delta,x_0,\overline{y}) + C\delta + \rho(\delta)$$

Now send $\delta \to 0$, by the right continuity W in t we obtain the desired convexity. ■

Proof of Example 4.7. We first prove (4.12). Denote

$$\tilde f_1(a,y) := a_1, \quad \tilde f_2(a,y) := a_2, \quad \tilde Y_t^\alpha = \int_t^T \tilde f(\alpha_s, \tilde Y_s^\alpha)ds - \int_t^T \tilde Z_s^\alpha dB_s.$$

Then one can easily check that

$$\begin{aligned}\tilde V(t) &:= \big\{\tilde Y_t^\alpha : \alpha \in \mathcal{A}_t\big\} = \big\{\tilde y \in \mathbb{R}^2 : |\tilde y| \le T-t\big\},\\ \tilde{\mathbb{V}}_b(t) &:= \big\{\tilde y \in \mathbb{R}^2 : |\tilde y| = T-t\big\} = \big\{(T-t)\big(\cos\theta, \sin\theta\big)^\top : \theta \in [0,2\pi)\big\}.\end{aligned} \tag{A.20}$$

Consider a function $\psi : \mathbb{R}^2 \to \mathbb{R}^2$ and set

$$\begin{aligned}&Y_t^\alpha := \psi(\tilde Y_t^\alpha), \quad Z_t^\alpha := \big[\nabla\psi_1(\tilde Y_t^\alpha), \nabla\psi_2(\tilde Y_t^\alpha)\big]^\top \tilde Z_t^\alpha,\\ &\text{where} \quad \psi_1(\tilde y) := \tilde y_1, \quad \psi_2(\tilde y) := [1+|\tilde y_1|^2]\tilde y_2.\end{aligned}$$

Then one may check straightforwardly that

$$\begin{aligned}dY_t^{\alpha,1} &= d\tilde Y_t^{\alpha,1} = -\alpha_t^1 dt + \tilde Z_s^{\alpha,1} dB_t = -f_1(\alpha_t, Y_t^\alpha, Z_t^\alpha)dt + Z_s^{\alpha,1}dB_t;\\ dY_t^{\alpha,2} &= 2\tilde Y_t^{\alpha,1}\tilde Y_t^{\alpha,2} d\tilde Y_t^{\alpha,1} + [1+|\tilde Y_t^{\alpha,1}|^2]d\tilde Y_t^{\alpha,2} + \big[\tilde Y_t^{\alpha,2}|\tilde Z_t^{\alpha,1}|^2 + 2\tilde Y_t^{\alpha,1}\tilde Z_t^{\alpha,1}\tilde Z_t^{\alpha,2}\big]dt\\ &= -\Big[2\tilde Y_t^{\alpha,1}\tilde Y_t^{\alpha,2}\alpha_t^1 + [1+|\tilde Y_t^{\alpha,1}|^2]\alpha_t^2 - \tilde Y_t^{\alpha,2}|\tilde Z_t^{\alpha,1}|^2 - 2\tilde Y_t^{\alpha,1}\tilde Z_t^{\alpha,1}\tilde Z_t^{\alpha,2}\Big]dt\\ &\quad + \Big[2\tilde Y_t^{\alpha,1}\tilde Y_t^{\alpha,2}\tilde Z_t^{\alpha,1} + [1+|\tilde Y_t^{\alpha,1}|^2]\tilde Z_t^{\alpha,2}\Big]dB_t\\ &= -\Big[\frac{2Y_t^{\alpha,1}Y_t^{\alpha,2}}{1+|Y_t^{\alpha,1}|^2}\alpha_t^1 + [1+|Y_t^{\alpha,1}|^2]\alpha_t^2\\ &\qquad - \frac{Y_t^{\alpha,2}|Z_t^{\alpha,1}|^2 + 2Y_t^{\alpha,1}Z_t^{\alpha,1}Z_t^{\alpha,2}}{1+|Y_t^{\alpha,1}|^2} + \frac{4|Y_t^{\alpha,1}|^2 Y_t^{\alpha,2}|Z_t^{\alpha,1}|^2}{(1+|Y_t^{\alpha,1}|^2)^2}\Big]dt + Z_t^{\alpha,2}dB_t\\ &= -f_2(\alpha_t, Y_t^\alpha, Z_t^\alpha)dt + Z_t^{\alpha,2}dB_t\end{aligned}$$

That is, Y^α satisfies (4.1) for given f. Therefore, $\mathbb{V}(t) = \big\{\psi(\tilde y) : \tilde y \in \tilde{\mathbb{V}}(t)\big\}$. Note further that $\psi_1(\tilde y_1) = \tilde y_1$, and ψ_2 is strictly increasing in $\tilde y_2$. It is clear that

$$\mathbb{V}_b(t) = \big\{\psi(\tilde y) : \tilde y \in \tilde{\mathbb{V}}_b(t)\big\}.$$

Plug (A.20) into it, we obtain (4.12) immediately.

We next analyze the convexity of $\mathbb{V}(t)$. Assume for simplicity that $t=0$. Note that

$$\begin{aligned}\tilde{\mathbb{V}}_b(0) &= \Big\{\big(y_1, \sqrt{T^2-|y_1|^2}\big),\ \big(y_1, -\sqrt{T^2-|y_1|^2}\big) : |y_1| \le T\Big\};\\ \mathbb{V}_b(0) &= \Big\{\big(y_1, \varphi(y_1)\big), \big(y_1, -\varphi(y_1)\big) : |y_1| \le T\Big\}, \text{ where } \varphi(y_1) := [1+|y_1|^2]\sqrt{T^2-|y_1|^2}.\end{aligned}$$

One may compute straightforwardly that:

$$\varphi''(y_1) = \frac{6|y_1|^4 - 9T^2|y_1|^2 + 2T^4 - T^2}{(T^2 - |y_1|^2)^{\frac{3}{2}}}, \quad |y_1| < T.$$

Note that

$$\sup_{|y_1|<T} \left[6|y_1|^4 - 9T^2|y_1|^2\right] = 0.$$

So when $T \leq \frac{1}{\sqrt{2}}$ and thus $2T^4 - T^2 \leq 0$, we have $\varphi''(y_1) \leq 0$ for $|y_1| < T$, and in this case $\mathbb{V}(0)$ is indeed convex. However, when $T > \frac{1}{\sqrt{2}}$, we find that $\varphi''(y_1) < 0$ for $|y_1| \approx T$, but $\varphi"(0) = \frac{2T^4-T^2}{T^3} > 0$, then $\mathbb{V}(0)$ is nonconvex. ■

Proof of Proposition 4.8. Given the convexity of $\mathbb{V}(t,x)$ in Proposition 4.5, Part (ii) is a direct consequence of Part (i) and Lemma A.1 below. So we shall only prove (i) here. We remark that this regularity does not rely on the convexity of $\mathbb{V}(t,x)$. Throughout the proof, ρ denotes a generic modulus of continuity function, which may vary from line to line. We proceed in four steps.

Step 1. We first prove the regularity in x. For any $y = Y_t^{t,x,\alpha} \in \mathbb{V}(t,x)$, denote $\tilde{y} := Y_t^{t,\tilde{x},\alpha} \in \mathbb{V}(t,\tilde{x})$. By standard SDE and BSDE estimates we have

$$d(y, \mathbb{V}(t,\tilde{x})) \leq |y - \tilde{y}| = |Y_t^{t,x,\alpha} - Y_t^{t,\tilde{x},\alpha}| \leq \rho(|x - \tilde{x}|).$$

Since d is Lipschitz continuous in y with Lipschitz constant y, this clearly implies $\sup_{y\in\mathbb{V}(t,x)} d(y, \mathbb{V}(t,\tilde{x})) \leq \rho(|x-\tilde{x}|)$. Similarly we have $\sup_{\tilde{y}\in\mathbb{V}(t,\tilde{x})} d(\tilde{y}, \mathbb{V}(t,x)) \leq \rho(|x-\tilde{x}|)$. Then we obtain $d(\mathbb{V}(t,x), \mathbb{V}(t,\tilde{x})) \leq \rho(|x-\tilde{x}|)$.

Step 2. In this step we estimate $\sup_{\tilde{y}\in\mathbb{V}(t+\delta,x)} d(\tilde{y}, \mathbb{V}(t,x))$. For $\tilde{y} = Y_{t+\delta}^{t+\delta,x,\tilde{\alpha}} \in \mathbb{V}(t+\delta, x)$ with $\tilde{\alpha} \in \mathcal{A}_{t+\delta}$, fix some $a_0 \in A$ and consider $\alpha \in \mathcal{A}_t$ defined by: $\alpha_s(\omega) := a_0 \mathbf{1}_{[t,t+\delta)}(s) + \tilde{\alpha}_s(\omega^{t+\delta})\mathbf{1}_{[t+\delta,T]}(s)$ where $\omega_s^{t+\delta} := \omega_s - \omega_{t+\delta} \in \Omega_{t+\delta}$. Then $y := Y_t^{t,x,\alpha} \in \mathbb{V}(t,x)$. Note that

$$\left|Y_{t+\delta}^{t,x,\alpha} - \tilde{y}\right| = \left|Y_{t+\delta}^{t+\delta, X_{t+\delta}^{t,x,a_0}, \tilde{\alpha}} - Y_{t+\delta}^{t+\delta,x,\tilde{\alpha}}\right| \leq \rho\big(X_{t+\delta}^{t,x,a_0} - x\big).$$

Recall (4.1) for $Y^{t,x,\alpha}$ and introduce another BSDE on $[t, t+\delta]$:

$$\begin{aligned}
Y_s^{t,x,\alpha} &= Y_{t+\delta}^{t,x,\alpha} + \int_s^{t+\delta} f(r, X_r^{t,x,a_0}, Y_r^{t,x,\alpha}, Z_r^{t,x,\alpha}, a_0)dr - \int_s^{t+\delta} Z_r^{t,x,\alpha} dB_r;\\
\tilde{Y}_s &= \tilde{y} + \int_s^{t+\delta} f(r, X_r^{t,x,a_0}, \tilde{Y}_r, \tilde{Z}_r, a_0)dr - \int_s^{t+\delta} \tilde{Z}_r dB_r.
\end{aligned}$$

By standard BSDE estimates we get, for possibly different modulus of continuity function ρ,

$$|y - \tilde{Y}_t|^2 = |Y_t^{t,x,\alpha} - \tilde{Y}_t|^2 \leq \mathbb{E}\Big[\big|Y_{t+\delta}^{t,x,\alpha} - \tilde{y}\big|^2\Big] \leq \mathbb{E}\Big[\rho^2\big(X_{t+\delta}^{t,x,a_0} - x\big)\Big] \leq \rho(\delta),$$

Moreover, since the FBSDE system for $(X^{t,x,a_0}, \tilde{Y}, \tilde{Z})$ is Markovian and $\tilde{y} \in \mathbb{V}(t+\delta, x)$ is bounded, under Assumption 4.4 one can easily show that $(\tilde{Y}, \tilde{Z})$ is bounded. This implies that $\left|\tilde{Y}_t - \tilde{y}\right| \le C\sqrt{\delta}$. Then $|y - \tilde{y}| \le C\sqrt{\delta} + \rho(\delta) \le \rho(\delta)$, and thus we can easily get:

$$\sup_{\tilde{y} \in \mathbb{V}(t+\delta, x)} d(\tilde{y}, \mathbb{V}(t,x)) \le \rho(\delta). \tag{A.21}$$

Step 3. To see the opposite direction, we first show that $d(g(x), \mathbb{V}(T-\delta, x)) \le \rho(\delta)$. Given arbitrary $y = Y^{T-\delta,x,\alpha}_{T-\delta} \in \mathbb{V}(T-\delta, x)$ with $\alpha \in \mathcal{A}_{T-\delta}$. Denote $\varphi^\alpha := \varphi^{T-\delta,x,\alpha}$ for $\varphi := X, Y, Z$. Under Assumption 4.4, by standard SDE and BSDE estimates we have

$$\mathbb{E}\big[|X^\alpha_T - x|^2\big] \le C\delta, \quad \mathbb{E}\Big[\sup_{T-\delta \le s \le T} |Y^\alpha_s|^2 + \int_{T-\delta}^T |Z^\alpha_s|^2 ds\Big] \le C.$$

Then, for a generic modulus of continuity function ρ,

$$\begin{aligned}
&|y - g(x)| = \Big|\mathbb{E}\big[g(X^\alpha_T) + \int_{T-\delta}^T f(t, X^\alpha_s, Y^\alpha_s, Z^\alpha_s, \alpha_s) ds - g(x)\big]\Big| \\
&\le \mathbb{E}\Big[\rho(|X^\alpha_T - x|) + C\int_{T-\delta}^T [1 + |Y^\alpha_s| + |Z^\alpha_s|] ds\Big] \\
&\le \rho(\delta) + C\Big(\mathbb{E}\Big[\delta\int_{T-\delta}^T [1 + |Y^\alpha_s|^2 + |Z^\alpha_s|^2] ds\Big]\Big)^{\frac{1}{2}} \le \rho(\delta) + C\sqrt{\delta} \le \rho(\delta).
\end{aligned}$$

This implies that $d(g(x), \mathbb{V}(T-\delta, x)) \le \rho(\delta)$.

Step 4. We now show that $\sup_{y \in \mathbb{V}(t,x)} d(y, \mathbb{V}(t+\delta, x)) \le \rho(\delta)$. This, together with the estimate in Step 2, implies the desired regularity in t. Fix $y = Y^{t,x,\alpha}_t \in \mathbb{V}(t,x)$ for some $\alpha \in \mathcal{A}_t$. As in Step 3 we denote $\varphi^\alpha := \varphi^{t,x,\alpha}$ for $\varphi = X, Y, Z$. Abusing the notation $(\tilde{Y}, \tilde{Z})$ with Step 2, we consider the following BSDE on $[t, T-\delta]$:

$$\tilde{Y}_s = g(X^\alpha_{T-\delta}) + \int_t^{T-\delta} f(r, X^\alpha_r, \tilde{Y}_r, \tilde{Z}_r, \alpha_r) dr - \int_t^{T-\delta} \tilde{Z}_r dB_r. \tag{A.22}$$

Note that (Y^α, Z^α) satisfies the BSDE on $[t, T-\delta]$ with terminal condition $Y^\alpha_{T-\delta}$. One can easily see that, as in (4.17) with deterministic τ, $Y^\alpha_{T-\delta} \in \mathbb{V}(T-\delta, X^\alpha_{T-\delta})$. Then, by Step 3, $\left|Y^\alpha_{T-\delta} - g(X^\alpha_{T-\delta})\right| \le \rho(\delta)$. Then, by comparing the two BSDEs, we obtain immediately that

$$|\tilde{Y}_t - y| = |\tilde{Y}_t - Y^\alpha_t| \le \rho(\delta).$$

We next introduce another shifted FBSDE, still on $[t, T-\delta]$:

$$\begin{aligned}
&\bar{X}_s := x + \int_t^s b(r+\delta, \bar{X}_r, \alpha_r) dr + \int_t^s \sigma(r+\delta, \bar{X}_r, \alpha_r) dB_r; \\
&\bar{Y}_s = g(X^\alpha_{T-\delta}) + \int_s^{T-\delta} f(r+\delta, X^\alpha_r, \bar{Y}_r, \bar{Z}_r, \alpha_r) dr - \int_r^{T-\delta} \bar{Z}_r dB_r.
\end{aligned} \tag{A.23}$$

By the continuity of b, σ, f in t, it is clear that $|\bar{Y}_t - \tilde{Y}_t| \le \rho(\delta)$.

Finally, note that $\alpha \in \mathcal{A}_t$ means that $\alpha_s = \varphi(s, B^t_{[t,s]})$, $s \in [t,T]$, for some appropriate (deterministic) mapping φ. Introduce $\bar\alpha_s := \varphi(s-\delta, B^{t+\delta}_{[t+\delta,s]})$, $s \in [t+\delta, T]$. It is clear that $(\bar\alpha_s, B^{t+\delta}_s)_{s\in[t+\delta,T]}$ has the same distribution as $(\alpha_s, B^t_s)_{s\in[t,T-\delta]}$. Then, compare (A.23) with the FBSDE for $(X^{\bar\alpha}, Y^{\bar\alpha}, Z^{\bar\alpha})$, we see that $(X^{\bar\alpha}_s, Y^{\bar\alpha}_s, Z^{\bar\alpha}_s)_{s\in[t+\delta,T]}$ has the same distribution as $(\bar X_s, \bar Y_s, \bar Z_s)_{s\in[t,T-\delta]}$. In particular, this means that $Y^{\bar\alpha}_{t+\delta} = \bar Y_t$. Therefore,

$$\big|y - Y^{\bar\alpha}_{t+\delta}\big| = \big|y - \bar Y_t\big| \le \big|y - \tilde Y_t\big| + \big|\tilde Y_t - \bar Y_t\big| \le \rho(\delta).$$

Since $\bar\alpha \in \mathcal{A}_{t+\delta}$, this implies that $d(y, \mathbb{V}(t+\delta, x)) \le \rho(\delta)$, and the proof is complete. ∎

Lemma A.1 *Assume $\mathbb{D}$ and $\tilde{\mathbb{D}}$ are compact and convex. Then*

$$d(\mathbb{D}_b, \tilde{\mathbb{D}}_b) = d(\mathbb{D}, \tilde{\mathbb{D}}), \quad \textit{and} \quad \sup_{y\in\mathbb{R}^m} \big|\mathbf{r}_{\mathbb{D}}(y) - \mathbf{r}_{\tilde{\mathbb{D}}}(y)\big| \le 3d(\mathbb{D}, \tilde{\mathbb{D}}). \tag{A.24}$$

Proof (i) The result $d(\mathbb{D}_b, \tilde{\mathbb{D}}_b) = d(\mathbb{D}, \tilde{\mathbb{D}})$ is not new, see e.g. [29]. However, for the convenience of the readers, we present here an elementary proof in two steps.

Step 1. In this step we show that $\sup_{y\in\mathbb{D}} d(y, \tilde{\mathbb{D}}) \le d(\mathbb{D}_b, \tilde{\mathbb{D}}_b)$. Then similarly we have $\sup_{\tilde y\in\mathbb{D}} d(\tilde y, \mathbb{D}) \le d(\mathbb{D}_b, \tilde{\mathbb{D}}_b)$ and thus $d(\mathbb{D}, \tilde{\mathbb{D}}) \le d(\mathbb{D}_b, \tilde{\mathbb{D}}_b)$. Note that the inequality is trivial when $\mathbb{D} \subset \tilde{\mathbb{D}}$, thus at below we assume $\mathbb{D}\backslash\tilde{\mathbb{D}} \neq \emptyset$, and then,

$$\sup_{y\in\mathbb{D}} d(y, \tilde{\mathbb{D}}) = \sup_{y\in\mathbb{D}\backslash\tilde{\mathbb{D}}} d(y, \tilde{\mathbb{D}}) = \sup_{y\in\mathbb{D}\backslash\tilde{\mathbb{D}}} d(y, \tilde{\mathbb{D}}_b).$$

For each $y \in \mathbb{D}\backslash\tilde{\mathbb{D}}$, by compactness and convexity of $\tilde{\mathbb{D}}$, there exists unique $\tilde y \in \tilde{\mathbb{D}}_b$ such that $d(y, \tilde{\mathbb{D}}_b) = |y - \tilde y|$. Note that, by convexity of $\tilde{\mathbb{D}}$, in this case hyperplane tangent to $y - \tilde y$ at $\tilde y$ is a supporting hyperplane for $\tilde{\mathbb{D}}$. Since $y \in \mathbb{D}$, there exists $\bar y \in \mathbb{D}_b$ such that $\bar y - \tilde y = c(y - \tilde y)$ for some $c \ge 1$. By the convexity of $\tilde{\mathbb{D}}$, we see that $|\bar y - \tilde y| = d(\bar y, \tilde{\mathbb{D}}_b)$. Then, for any $y \in \mathbb{D}\backslash\tilde{\mathbb{D}}$,

$$d(y, \tilde{\mathbb{D}}) = |y - \tilde y| \le |\bar y - \tilde y| = d(\bar y, \tilde{\mathbb{D}}_b) \le d(\mathbb{D}_b, \tilde{\mathbb{D}}_b).$$

This implies the desired estimate: $\sup_{y\in\mathbb{D}} d(y, \tilde{\mathbb{D}}) \le d(\mathbb{D}_b, \tilde{\mathbb{D}}_b)$.

Step 2. In this step we prove the opposite direction: $d(y, \tilde{\mathbb{D}}_b) \le d(\mathbb{D}, \tilde{\mathbb{D}})$ for all $y \in \mathbb{D}_b$. If $y \notin \tilde{\mathbb{D}}$, then clearly $d(y, \tilde{\mathbb{D}}_b) = d(y, \tilde{\mathbb{D}}) \le d(\mathbb{D}, \tilde{\mathbb{D}})$. So we assume $y \in \tilde{\mathbb{D}}$. Since $y \in \mathbb{D}_b$, let η denote a unit outward vector corresponding to a supporting hyperplane of $\mathbb{D}$ at y. Since $y \in \tilde{\mathbb{D}}$, there exists $\tilde y \in \tilde{\mathbb{D}}_b$ such that $\tilde y - y = c\eta$ for some $c > 0$. By the definition of η, this implies that $d(\tilde y, \mathbb{D}) = |\tilde y - y|$. On the other hand, since $\tilde y \in \tilde{\mathbb{D}}_b$, by definition $d(y, \tilde{\mathbb{D}}_b) \le |y - \tilde y|$. So

$$d(y, \tilde{\mathbb{D}}_b) \le |\tilde y - y| = d(\tilde y, \mathbb{D}) \le d(\mathbb{D}, \tilde{\mathbb{D}}), \quad \forall y \in \mathbb{D}_b.$$

Similarly we have $d(\tilde y, \mathbb{D}_b) \le d(\mathbb{D}, \tilde{\mathbb{D}})$ for all $\tilde y \in \tilde{\mathbb{D}}_b$. Thus $d(\mathbb{D}_b, \tilde{\mathbb{D}}_b) \le d(\mathbb{D}, \tilde{\mathbb{D}})$.

(ii) We now estimate $\mathbf{r}_{\mathbb{D}} - \mathbf{r}_{\tilde{\mathbb{D}}}$. Since $\mathbb{D}$ and $\tilde{\mathbb{D}}$ are symmetric, it suffices to prove the following estimate and we proceed in four cases:

$$\mathbf{r}_{\tilde{\mathbb{D}}}(y) - \mathbf{r}_{\mathbb{D}}(y) \le 3d(\mathbb{D}, \tilde{\mathbb{D}}). \tag{A.25}$$

Case 1. $y \in \tilde{\mathbb{D}}\backslash\mathbb{D}$. Then $\mathbf{r}_{\tilde{\mathbb{D}}}(y) \le 0 < \mathbf{r}_{\mathbb{D}}(y)$, and (4.14) is obvious.

Case 2. $y \notin \mathbb{D} \cup \tilde{\mathbb{D}}$. Then $\mathbf{r}_{\tilde{\mathbb{D}}}(y) - \mathbf{r}_{\mathbb{D}}(y) = d(y, \tilde{\mathbb{D}}_b) - d(y, \mathbb{D}_b)$. By the compactness there exists $y' \in \mathbb{D}_b$ such that $\mathbf{r}_{\mathbb{D}}(y) = d(y, \mathbb{D}_b) = |y - y'|$. Then,

$$\mathbf{r}_{\tilde{\mathbb{D}}}(y) = d(y, \tilde{\mathbb{V}}_b) \le |y - y'| + d(y', \tilde{\mathbb{D}}_b) \le \mathbf{r}_{\mathbb{D}}(y) + d(\mathbb{D}_b, \tilde{\mathbb{D}}_b) = \mathbf{r}_{\mathbb{D}}(y) + d(\mathbb{D}, \tilde{\mathbb{D}}).$$

Case 3. $y \in \mathbb{D} \cap \tilde{\mathbb{D}}$. Then $\mathbf{r}_{\tilde{\mathbb{D}}}(y) - \mathbf{r}_{\mathbb{D}}(y) = d(y, \mathbb{D}_b) - d(y, \tilde{\mathbb{D}}_b)$, and (4.14) can be estimated similarly as in Case 2.

Case 4. $y \in \mathbb{D}\backslash\tilde{\mathbb{D}}$. Then $\mathbf{r}_{\tilde{\mathbb{D}}}(y) - \mathbf{r}_{\mathbb{D}}(y) = d(y, \mathbb{D}_b) + d(y, \tilde{\mathbb{D}}_b)$. First,

$$d(y, \tilde{\mathbb{D}}_b) = d(y, \tilde{\mathbb{D}}) \le d(\mathbb{D}, \tilde{\mathbb{D}}),$$

where the first equality is due to the fact that $y \notin \tilde{\mathbb{D}}$ and the inequality is due to the fact that $y \in \mathbb{D}$. Next, it is obvious that

$$d(y, \mathbb{D}_b) \le d(y, \tilde{\mathbb{D}}_b) + d(\tilde{\mathbb{D}}_b, \mathbb{D}_b) \le 2d(\mathbb{D}, \tilde{\mathbb{D}}).$$

Put together with obtain (4.14). ■

We note that the first equality in (A.24) fails in general when $\mathbb{D}$ or $\tilde{\mathbb{D}}$ is not convex. At below we present a simple example where $\lim_{\varepsilon\to 0} d(\mathbb{D}, \tilde{\mathbb{D}}^\varepsilon) = 0 < 1 = \lim_{\varepsilon\to 0} d(\mathbb{D}_b, \tilde{\mathbb{D}}^\varepsilon_b)$.

Example A.2 *For $\varepsilon > 0$, set $\mathbb{D} := \{y \in \mathbb{R}^2 : |y| \le 1\}$ and*

$$\tilde{\mathbb{D}}^\varepsilon := \{r(\cos\theta, \sin\theta) \in \mathbb{R}^2 \;:\; 0 \le r \le 1, \quad \varepsilon \le \theta \le 2\pi - \varepsilon\} \subset \mathbb{D}.$$

One can easily see that $d(\mathbb{D}, \tilde{\mathbb{D}}^\varepsilon) = d((1,0), \tilde{\mathbb{D}}^\varepsilon_b) \to 0$, and $d(\mathbb{D}_b, \tilde{\mathbb{D}}^\varepsilon_b) = d((0,0), \mathbb{D}_b) = 1$. ■

Proof of Lemma 5.5. Denote

$$\tau_\delta := \inf\{t \ge 0 : |\mathbf{r}(t, X^\alpha_t, Y^\alpha_t)| > \delta\} \wedge T_0,$$

and consider the linear BSDE with solution pair (κ, β):

$$\kappa_t = \mathbf{1}_{\{\tau_\delta < T_0\}} - \int_t^{T_0} \beta_s dB_s, \quad 0 \le t \le T_0,$$

where $\kappa \in \mathbb{R}$, $\beta \in \mathbb{R}^{1\times d}$. It is clear that $|\kappa| \le 1$, $\beta_t = 0$ for $\tau_\delta \le t \le T_0$, and $\int_0^\cdot \beta_s dB_s$ is an BMO martingale. Thus,

$$\mathbb{E}\Big[\exp\big(c_0 \int_0^{T_0} |\beta_t|^2 dt\big)\Big] \le C_0 < \infty, \quad \text{for some } c_0, C_0 > 0.$$

For $n \ge 1$, we truncate β by n and denote it as β^n. Define

$$\kappa^n_t := \kappa_0 + \int_0^t \beta^n_s dB_s.$$

Then it is obvious that, for any $p \ge 1$,

$$\mathbb{E}\big[\sup_{0\le t\le T_0} |\kappa^n_t|^p\big] \le C_p < \infty; \text{ and } c^n_p := \Big(\mathbb{E}\big[\sup_{0\le t\le T_0} |\kappa^n_t - \kappa_t|^p\big]\Big)^{\frac{1}{p}} \to 0, \text{ as } n \to \infty. \quad \text{(A.26)}$$

Introduce two random fields $\nu^n(t,\omega,y)$ and $\rho^n(t,\omega,y)$:

$$\begin{aligned}
\nu^n_{ij}(t,y) &:= -\sum_{k=1}^m \int_0^1 \nabla_{z_{kj}} f^i(t,X^\alpha_t,Y^\alpha_t,Z^\alpha_t+\theta y\beta^n_t,\alpha_t) y_k d\theta;\\
\rho^n_i(t,y) &:= -\int_0^1 \Big[\nabla_y f^i\cdot y + \mathrm{tr}\left((\partial_z f^i)^\top \nu^n(t,y)\right)\Big]\\
&\qquad \big(t,X^\alpha_t,Y^\alpha_t+\theta\kappa^n_t y, Z^\alpha_t + y\beta^n_t+\theta\kappa^n_t\nu^n(t,y),\alpha_t\big)d\theta;
\end{aligned}$$

where $\nu^n=[\nu^n_{ij}]_{1\le i\le m,1\le j\le d}\in\mathbb{R}^{m\times d}$, $\rho^n=[\rho^n_i]_{1\le i\le m}\in\mathbb{R}^m$. One can easily verify that

$$\begin{aligned}
&\nu^n(t,y)\beta^n_t := f(t,X^\alpha_t,Y^\alpha_t,Z^\alpha_t,\alpha_t) - f(t,X^\alpha_t,Y^\alpha_t,Z^\alpha_t+y\beta^n_t,\alpha_t);\\
&\kappa^n_t\rho^n(t,y) := f(t,X^\alpha_t,Y^\alpha_t,Z^\alpha_t+y\beta^n_t,\alpha_t)\\
&\qquad\qquad -f(t,X^\alpha_t,Y^\alpha_t+\kappa^n_t y,Z^\alpha_t+y\beta^n_t+\kappa^n_t\nu(t,y),\alpha_t).
\end{aligned}$$

Moreover, by Assumption 4.4 (ii) we have

$$\begin{gathered}
|\nu^n(t,y)|\le C|y|,\quad |\rho^n(t,y)|\le C|y|;\\
|\nu^n(t,y)-\nu^n(t,\tilde y)|\le C_n\big[1+|y|\big]|y-\tilde y|,\\
|\rho^n(t,y)-\rho^n(t,\tilde y)|\le C_n\big[1+|y|+|\kappa^n_t|+|\kappa^n_t||y|^2\big]|y-\tilde y|.
\end{gathered}\tag{A.27}$$

Next, consider the following SDE:

$$\eta^n_t := \lambda\mathbf{n}_{\mathbb{V}}(0,x_0,\pi(0,x_0,Y^\alpha_0)) + \int_0^t \rho^n(s,\eta^n_s)ds + \int_0^t \nu^n(s,\eta^n_s)dB_s,$$

where $\lambda>0$ is a small number which will be determined later. By the standard stopping arguments for stochastic Lipschitz continuous coefficients, and by the uniform linear growth in the first line of (A.27), the above SDE is wellposed, and for any $p\ge 1$,

$$\mathbb{E}\Big[\sup_{0\le t\le T_0}|\eta^n_t|^p\Big]\le C_p|\lambda|^p,\tag{A.28}$$

where C_p does not depend on n.

Denote

$$\tilde Y^n_t := Y^\alpha_t + \kappa^n_t\eta^n_t,\quad \tilde Z^n_t := Z_t + \kappa^n_t\nu^n_t + \eta^n_t\beta^n_t.$$

Then by the standard Itô formula we have

$$\tilde Y^n_t = Y^\alpha_{\tau_\delta} + \kappa^n_{\tau_\delta}\eta^n_{\tau_\delta} + \int_t^{\tau_\delta} f(s,X^\alpha_s,\tilde Y^n_s,\tilde Z^n_s,\alpha_s)ds - \int_t^{\tau_\delta}\tilde Z^n_s dB_s.$$

Moreover, introduce the BSDE

$$\hat Y^n_t = Y^\alpha_{\tau_\delta} + \kappa^n_{\tau_\delta}\eta^n_{\tau_\delta}\mathbf{1}_{\left\{|\kappa^n_{\tau_\delta}|\le 2,|\eta^n_{\tau_\delta}|<\frac{\delta}{2}\right\}} + \int_t^{\tau_\delta} f(s,X^\alpha_s,\hat Y^n_s,\hat Z^n_s,\alpha_s)ds - \int_t^{\tau_\delta}\hat Z^n_s dB_s.$$

By (A.26), (A.28), and noting that $|\kappa_{\tau_\delta}| \le 1$, it follows from standard BSDE estimates that

$$
\begin{aligned}
|\hat{Y}_0^n - \tilde{Y}_0^n|^2 &\le C\mathbb{E}\Big[\big|\kappa^n_{\tau_\delta}\eta^n_{\tau_\delta}\big|^2\big[\mathbf{1}_{\{|\kappa^n_{\tau_\delta}-\kappa_{\tau_\delta}|>1\}} + \mathbf{1}_{\{|\eta^n_{\tau_\delta}|\ge\frac{\delta}{2}\}}\big]\Big] \\
&\le C\mathbb{E}\Big[\big|\kappa^n_{\tau_\delta}\eta^n_{\tau_\delta}\big|^2|\kappa^n_{\tau_\delta} - \kappa_{\tau_\delta}| + \frac{1}{\delta^2}|\kappa^n_{\tau_\delta}|^2|\eta^n_{\tau_\delta}|^4\Big] \le Cc_2^n + C\frac{|\lambda|^4}{\delta^2}.
\end{aligned}
$$

Thus

$$
|\hat{Y}_0^n - \tilde{Y}_0^n| \le C\big[\sqrt{c_2^n} + \frac{|\lambda|^2}{\delta}\big]. \tag{A.29}
$$

Note that

$$
\hat{Y}^n_{\tau_\delta} = Y^\alpha_{\tau_\delta} + \kappa^n_{\tau_\delta}\eta^n_{\tau_\delta}\mathbf{1}_{\{|\kappa^n_{\tau_\delta}|\le 2, |\eta^n_{\tau_\delta}|<\frac{\delta}{2}\}}.
$$

On $\{|\kappa^n_{\tau_\delta}| \le 2, |\eta^n_{\tau_\delta}| < \frac{\delta}{2}\}^c$, we have $\hat{Y}^n_{\tau_\delta} = Y^\alpha_{\tau_\delta} \in \mathbb{V}(\tau_\delta, X^\alpha_{\tau_\delta})$. On $\{|\kappa^n_{\tau_\delta}| \le 2, |\eta^n_{\tau_\delta}| < \frac{\delta}{2}\}$, noting again that $Y^\alpha_{\tau_\delta} \in \mathbb{V}(\tau_\delta, X^\alpha_{\tau_\delta})$, we have $\mathbf{r}_{\mathbb{V}}(\tau_\delta, X^\alpha_{\tau_\delta}, Y^\alpha_{\tau_\delta}) = -\delta$ and $|\kappa^n_{\tau_\delta}\eta^n_{\tau_\delta}| \le \delta$, then $\hat{Y}^n_{\tau_\delta} = Y^\alpha_{\tau_\delta} + \kappa^n_{\tau_\delta}\eta^n_{\tau_\delta} \in \mathbb{V}(\tau_\delta, X^\alpha_{\tau_\delta})$. So in both cases $\hat{Y}^n_{\tau_\delta} \in \mathbb{V}(\tau_\delta, X^\alpha_{\tau_\delta})$. Then by DPP (4.16) we have $\hat{Y}^n_0 \in \mathbb{V}(0, x_0)$. Thus, by (A.29),

$$
\mathbf{r}_{\mathbb{V}}(0, x_0, \tilde{Y}^n_0) \le |\tilde{Y}^n_0 - \hat{Y}^n_0| \le C\big[\sqrt{c_2^n} + \frac{|\lambda|^2}{\delta}\big].
$$

On the other hand, note that $\tilde{Y}^n_0 = Y^\alpha_0 + \kappa_0\lambda\mathbf{n}_{\mathbb{V}}(0, x_0, \pi(0, x_0, Y^\alpha_0))$, for $\kappa_0\lambda$ small we have

$$
\mathbf{r}_{\mathbb{V}}(0, x_0, \tilde{Y}^n_0) = \mathbf{r}_{\mathbb{V}}(0, x_0, Y^\alpha_0) + \kappa_0\lambda.
$$

Thus

$$
\kappa_0\lambda = \mathbf{r}_{\mathbb{V}}(0, x_0, \tilde{Y}^n_0) - \mathbf{r}_{\mathbb{V}}(0, x_0, Y^\alpha_0) \le C\big[\sqrt{c_2^n} + \frac{|\lambda|^2}{\delta}\big] + \varepsilon.
$$

Send $n \to \infty$ and set $\lambda := \sqrt{\varepsilon\delta}$, we obtain (5.6):

$$
\mathbb{P}(\tau_\delta < T_0) = \kappa_0 \le C\frac{\lambda}{\delta} + \frac{\varepsilon}{\lambda} = C\sqrt{\frac{\varepsilon}{\delta}}.
$$

Finally, if $Y^\alpha_0 \in \mathbb{V}_b(0, x_0)$, then $\varepsilon = 0$. We see that $\mathbb{P}(\tau_\delta < T_0) = 0$ for all $\delta > 0$ and all $T_0 < T$. This implies immediately that $Y^\alpha_t \in \mathbb{V}_b(t, X^\alpha_t)$, $0 \le t < T$, a.s. Moreover, note that $Y^\alpha_T = g(X^\alpha_T)$ and $\mathbb{V}(T, x) = \{g(x)\}$, we have $Y^\alpha_T \in \mathbb{V}_b(T, X^\alpha_T)$ as well. ■

Proof of Example 6.4. As usual we drop the subscript $_{\mathbb{V}}$ in $\mathbf{r}$ and $\mathbf{n}$.

(i) We first show that $\mathbb{V} \in C^{1,2}_0([0,T) \times \mathbb{R}; \mathcal{D}^2_2)$. Fix $\delta > 0$ and denote $T_\delta := T - \delta$. By Example 2.9 we have, with $u(t,x) = T - t \ge \delta$ there for $t \in [0, T_\delta]$,

$$
\mathbf{r}(t, x, y) = |y - w(t, x)| - (T - t).
$$

Then it is clear that $\mathbb{V}\in C^{1,2}([0,T_\delta]\times\mathbb{R};\mathcal{D}_2^2)$. By (2.25), for $|y-w(t,x)|=T-t$, we have

$$\mathbf{n}=\frac{y-w}{T-t};\quad \partial_t\mathbb{V}=\big[\nabla_t w\cdot\mathbf{n}-1\big]\mathbf{n};\quad \partial_x\mathbb{V}=\big[\nabla_x w\cdot\mathbf{n}\big]\mathbf{n};$$

$$\partial_x\mathbf{n}=\frac{1}{T-t}\big[-\nabla_x w+[\mathbf{n}\cdot\nabla_x w]\mathbf{n}\big],\quad \partial_y\mathbf{n}=\frac{1}{T-t}\big[I_{2\times 2}-\mathbf{n}\mathbf{n}^\top\big];$$

$$\partial_{xx}\mathbb{V}=-\frac{1}{T-t}\Big[\big[|\nabla_x w|^2-\nabla_{xx}w\cdot\mathbf{n}+|\nabla_x w\cdot\mathbf{n}|^2\big]\mathbf{n}+[\nabla_x w\cdot\mathbf{n}]\big[\nabla_x w-(\nabla_x w\cdot\mathbf{n})\mathbf{n}\big]\Big].$$

In particular, $c_{T_0}=\frac{1}{T}$ in (5.5), and thus $\mathbb{V}\in C_0^{1,2}([0,T)\times\mathbb{R};\mathcal{D}_2^2)$.

(ii) We next verify the conditions in Theorem 6.1 (ii). For any $a\in A$ and $\zeta\in\mathbb{T}_{\mathbb{V}}(t,x,y)$, by (5.2) we have: at $(t,x,y)\in\mathbb{G}_{\mathbb{V}}$,

$$\begin{aligned}
&h^0_{\mathbb{V}}(t,x,y,\partial_x\mathbb{V},\partial_{xx}\mathbb{V},a,\zeta)=\frac12\partial_{xx}\mathbb{V}-\Big[\zeta\cdot\partial_x\mathbf{n}+\frac12\zeta^\top\partial_y\mathbf{n}\zeta\Big]\mathbf{n}\\
&=-\frac{1}{2(T-t)}\Big[\big[|\nabla_x w|^2-\nabla_{xx}w\cdot\mathbf{n}+|\nabla_x w\cdot\mathbf{n}|^2\big]\mathbf{n}+[\nabla_x w\cdot\mathbf{n}]\big[\nabla_x w-(\nabla_x w\cdot\mathbf{n})\mathbf{n}\big]\Big]\\
&\quad-\frac{1}{T-t}\Big[-\zeta\cdot\nabla_x w+\frac12|\zeta|^2\Big]\mathbf{n}.
\end{aligned}$$

Thus

$$\begin{aligned}
&\mathbf{n}\cdot h_{\mathbb{V}}(t,x,y,\partial_x\mathbb{V},\partial_{xx}\mathbb{V},a,\zeta)=\mathbf{n}\cdot h^0_{\mathbb{V}}(t,x,y,\partial_x\mathbb{V},\partial_{xx}\mathbb{V},a,\zeta)+\mathbf{n}\cdot[f^0(t,x)+a]\\
&=-\frac{1}{2(T-t)}\Big[|\nabla_x w|^2-\nabla_{xx}w\cdot\mathbf{n}+|\nabla_x w\cdot\mathbf{n}|^2-2\zeta\cdot\nabla_x w+|\zeta|^2\Big]+\mathbf{n}\cdot\big[f^0+a\big]\\
&=-\frac{1}{2(T-t)}\Big[|\zeta-\nabla_x w|^2-\nabla_{xx}w\cdot\mathbf{n}+|\nabla_x w\cdot\mathbf{n}|^2\Big]+\mathbf{n}\cdot\big[f^0+a\big].
\end{aligned}$$

Recall $|a|\le 1$, then clearly the optimal arguments are:

$$a^*=I_1^{\mathbb{V}}(t,x,y):=\mathbf{n}(t,x,y),\quad \zeta^*=I_2^{\mathbb{V}}(t,x,y):=\nabla_x w-[\nabla_x w\cdot\mathbf{n}]\mathbf{n}.$$

Together with (6.1), this implies further that

$$\begin{aligned}
&\tilde I_3^{\mathbb{V}}:=-\Big[\partial_t\mathbb{V}+h^0_{\mathbb{V}}(\cdot,\partial_x\mathbb{V},\partial_{xx}\mathbb{V},I_1^{\mathbb{V}},I_2^{\mathbb{V}})+f^0+I_1^{\mathbb{V}}\Big]\\
&=-\big[\nabla_t w\cdot\mathbf{n}-1\big]\mathbf{n}+\frac{1}{T-t}\big[-\zeta\cdot\nabla_x w+\frac12|\zeta|^2\big]\mathbf{n}-[f^0+\mathbf{n}]\\
&\quad+\frac{1}{2(T-t)}\Big[\big[|\nabla_x w|^2-\nabla_{xx}w\cdot\mathbf{n}+|\nabla_x w\cdot\mathbf{n}|^2\big]\mathbf{n}+[\nabla_x w\cdot\mathbf{n}]\big[\nabla_x w-(\nabla_x w\cdot\mathbf{n})\mathbf{n}\big]\Big];\\
&I_3^{\mathbb{V}}:=-f^0+[\mathbf{n}\cdot f^0]\mathbf{n}+\frac{1}{2(T-t)}[\nabla_x w\cdot\mathbf{n}]\big[\nabla_x w-(\nabla_x w\cdot\mathbf{n})\mathbf{n}\big].
\end{aligned}$$

Plug $I_1^{\mathbb{V}},I_2^{\mathbb{V}},I_3^{\mathbb{V}}$ into (6.2), clearly the resulted SDE is wellposed. ■

Proof of Example 8.3. We now compute the equation (8.3) in this case. First,

$$\begin{aligned}
\mathcal{N}(t,x,y,\nabla_x\eta,\nabla_y\eta)&=\big\{(a,z):[1,z]^\top[\nabla_x\eta+z\nabla_y\eta]=0\big\}\\
&=\big\{(a,z):z=-\frac{\nabla_x\eta}{\nabla_y\eta}\big\}=\big\{\big(a,-\frac{\nabla_x\widehat{\mathbf{r}}(t,x,y)}{\nabla_y\widehat{\mathbf{r}}(t,x,y)}\big)\big\}.
\end{aligned}$$

Then, recalling $\eta = \frac{1}{2}|\widehat{\mathbf{r}}|^2$,

$$\begin{aligned}
F &= \frac{1}{2}\Big[\nabla_{xx}\eta - 2\nabla_{xy}\eta\frac{\nabla_x\widehat{\mathbf{r}}}{\nabla_y\widehat{\mathbf{r}}} + \nabla_{yy}\eta\Big|\frac{\nabla_x\widehat{\mathbf{r}}}{\nabla_y\widehat{\mathbf{r}}}\Big|^2\Big](t,x,y)\\
&= \frac{1}{2}\Big[\widehat{\mathbf{r}}\nabla_{xx}\widehat{\mathbf{r}} + |\nabla_x\widehat{\mathbf{r}}|^2 - 2[\widehat{\mathbf{r}}\nabla_{xy}\widehat{\mathbf{r}} + \nabla_x\widehat{\mathbf{r}}\nabla_y\widehat{\mathbf{r}}]\frac{\nabla_x\widehat{\mathbf{r}}}{\nabla_y\widehat{\mathbf{r}}} + [\widehat{\mathbf{r}}\nabla_{yy}\widehat{\mathbf{r}} + |\nabla_y\widehat{\mathbf{r}}|^2]\Big|\frac{\nabla_x\widehat{\mathbf{r}}}{\nabla_y\widehat{\mathbf{r}}}\Big|^2\Big](t,x,y)\\
&= \frac{1}{2}\widehat{\mathbf{r}}\Big[\nabla_{xx}\widehat{\mathbf{r}} - 2\nabla_{xy}\widehat{\mathbf{r}}\frac{\nabla_x\widehat{\mathbf{r}}}{\nabla_y\widehat{\mathbf{r}}} + \nabla_{yy}\widehat{\mathbf{r}}\Big|\frac{\nabla_x\widehat{\mathbf{r}}}{\nabla_y\widehat{\mathbf{r}}}\Big|^2\Big](t,x,y).
\end{aligned}$$

Note further that $\widehat{\mathbf{r}} = 0$ on $\mathbb{G}_{\mathbb{V}}$. Then, for $(t,x,y) \in \mathbb{G}_{\mathbb{V}}$, we have

$$\begin{aligned}
\nabla_x F &= \frac{1}{2}\nabla_x\widehat{\mathbf{r}}\Big[\nabla_{xx}\widehat{\mathbf{r}} - 2\nabla_{xy}\widehat{\mathbf{r}}\frac{\nabla_x\widehat{\mathbf{r}}}{\nabla_y\widehat{\mathbf{r}}} + \nabla_{yy}\widehat{\mathbf{r}}\Big|\frac{\nabla_x\widehat{\mathbf{r}}}{\nabla_y\widehat{\mathbf{r}}}\Big|^2\Big](t,x,y);\\
\nabla_y F &= \frac{1}{2}\nabla_y\widehat{\mathbf{r}}\Big[\nabla_{xx}\widehat{\mathbf{r}} - 2\nabla_{xy}\widehat{\mathbf{r}}\frac{\nabla_x\widehat{\mathbf{r}}}{\nabla_y\widehat{\mathbf{r}}} + \nabla_{yy}\widehat{\mathbf{r}}\Big|\frac{\nabla_x\widehat{\mathbf{r}}}{\nabla_y\widehat{\mathbf{r}}}\Big|^2\Big](t,x,y).
\end{aligned}$$

On the other hand, note that $\nabla_t\eta = \widehat{\mathbf{r}}\nabla_t\widehat{\mathbf{r}}$. Then, again at $(t,x,y) \in \mathbb{G}_{\mathbb{V}}$,

$$\nabla_x\nabla_t\eta = \nabla_x\widehat{\mathbf{r}}\nabla_t\widehat{\mathbf{r}}, \quad \nabla_y\nabla_t\eta = \nabla_y\widehat{\mathbf{r}}\nabla_t\widehat{\mathbf{r}}.$$

Plug these into (8.3), we obtain (8.5) immediately. ■

References

[1] Çağın Ararat, Jin Ma, and Wenyuan Wu. Set-valued backward stochastic differential equations. *The Annals of Applied Probability*, 33(5):3418–3448, 2023.

[2] Jean-Pierre Aubin and Giuseppe Da Prato. The viability theorem for stochastic differential inclusions. *Stochastic Analysis and Applications*, 16(1):1–15, 1998.

[3] Jean-Pierre Aubin and Hélène Frankowska. *Set-Valued Analysis*. Modern Birkhäuser Classics. Birkhäuser Boston, Boston, MA, 2009.

[4] Guy Barles, Halil Mete Soner, and Panagiotis E. Souganidis. Front propagation and phase field theory. *SIAM Journal on Control and Optimization*, 31(2):439–469, 1993.

[5] Bruno Bouchard and Nizar Touzi. Explicit solution of the multivariate super-replication problem under transaction costs. *The Annals of Applied Probability*, 10(3):685–708, 2000.

[6] Piermarco Cannarsa and Carlo Sinestrari. *Semiconcave Functions, Hamilton-Jacobi Equations, and Optimal Control*, volume 58 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser Boston, Boston, MA, 2004.

[7] Yun-Gang Chen, Yoshikazu Giga, and Shun'ichi Goto. Uniqueness and existence of viscosity solutions of generalized mean curvature flow equations. *Journal of Differential Geometry*, 33(3):749–786, 1991.

[8] Lawrence C. Evans and Joel Spruck. Motion of level sets by mean curvature. I. *Journal of Differential Geometry*, 33(3):635–681, 1991.

[9] Zachary Feinstein and Birgit Rudloff. Multi-portfolio time consistency for set-valued convex and coherent risk measures. *Finance and Stochastics*, 19(1):67–107, 2015.

[10] Zachary Feinstein and Birgit Rudloff. Time consistency for scalar multivariate risk measures. *Statistics & Risk Modeling*, 38(3-4):71–90, 2021.

[11] Zachary Feinstein, Birgit Rudloff, and Jianfeng Zhang. Dynamic set values for nonzero-sum games with multiple equilibria. *Mathematics of Operations Research*, 47(1):616–642, 2022.

[12] Yoshikazu Giga. *Surface Evolution Equations: A Level Set Approach*, volume 99 of *Monographs in Mathematics.* Birkhäuser Basel, Basel, 2006.

[13] William B. Gordon. On the diffeomorphisms of euclidean space. *The American Mathematical Monthly*, 79(7):755–759, 1972.

[14] Melih İşeri and Jianfeng Zhang. Set values for mean field games. *Transactions of the American Mathematical Society*, 377(10):7117–7174, 2024.

[15] Yuri Kabanov. Hedging and liquidation under transaction costs in currency markets. *Finance and Stochastics*, 3(2):237–248, 1999.

[16] Chandrasekhar Karnam, Jin Ma, and Jianfeng Zhang. Dynamic approaches for some time inconsistent optimization problems. *The Annals of Applied Probability*, 27(6):3435–3477, 2017.

[17] Michal Kisielewicz. *Stochastic differential inclusions and applications*, volume 80 of *Springer Optim. Appl.* Springer, New York, 2013.

[18] Michal Kisielewicz. *Set-valued stochastic integrals and applications*, volume 157 of *Springer Optim. Appl.* Springer, Cham, 2020.

[19] Steven G. Krantz and Harold R. Parks. Distance to C^k hypersurfaces. *Journal of Differential Equations*, 40(1):116–120, 1981.

[20] Hiroshi Kunita. Stochastic differential equations and stochastic flows of diffeomorphisms. In *École d'Été de Probabilités de Saint-Flour XII — 1982*, volume 1097 of *Lecture Notes in Mathematics*, pages 143–303. Springer, Berlin, Heidelberg, 1984.

[21] J. L. Pedersen and G. Peskir. Optimal mean-variance portfolio selection. *Mathematics and Financial Economics*, 11(2):137–160, 2017.

[22] James A. Sethian. Curvature and the evolution of fronts. *Communications in Mathematical Physics*, 101(4):487–499, 1985.

[23] Halil Mete Soner. Motion of a set by the curvature of its boundary. *Journal of Differential Equations*, 101(2):313–372, 1993.

[24] Halil Mete Soner and Nizar Touzi. Dynamic programming for stochastic target problems and geometric flows. *Journal of the European Mathematical Society*, 4(3):201–236, 2002.

[25] Halil Mete Soner and Nizar Touzi. A stochastic representation for the level set equations. *Communications in Partial Differential Equations*, 27(9-10):2031–2053, 2002.

[26] Halil Mete Soner and Nizar Touzi. Stochastic target problems, dynamic programming, and viscosity solutions. *SIAM Journal on Control and Optimization*, 41(2):404–424, 2002.

[27] Halil Mete Soner and Nizar Touzi. A stochastic representation for mean curvature type geometric flows. *The Annals of Probability*, 31(3):1145–1165, 2003.

[28] Nizar Touzi. *Stochastic Control, and Applications to Finance*. Lecture notes at Ecole Polytechnique. 2018.

[29] M. D. Wills. Hausdorff Distance and Convex Sets. *Journal of Convex Analysis*, 14(1):1–6, 2007.

[30] Jianfeng Zhang. *Backward Stochastic Differential Equations: From Linear to Fully Nonlinear Theory*, volume 86 of *Probability Theory and Stochastic Modelling*. Springer, New York, 2017.