

Set Valued Hamilton-Jacobi-Bellman Equations

Melih İseri* Jianfeng Zhang†

November 13, 2023

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 boundary 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:

$$\mathbb{V} : [0, T] \times \mathbb{R}^d \rightarrow 2^{\mathbb{R}^m}. \quad (1.1)$$

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. These include, among many others, stochastic viability problems (Aubin-Da Prato [2]), multivariate super-hedging problems (Kabanov [13] and Bouchard-Touzi [5]), multivariate dynamic risk measures (Feinstein-Rudloff [8]), time inconsistent optimization problems (Karman-Ma-Zhang [14]), stochastic target problems (Soner-Touzi [18, 19]), and recently, nonzero sum games with multiple equilibria (Feinstein-Rudloff-Zhang [10]), and mean field games with multiple mean field equilibria (İşeri-Zhang [12]). One crucial property shared by these set valued functions is the *dynamic programming principle*, or say the time consistency. Then a natural question is:

can we characterize the set valued function via a set valued PDE?

This is exactly the goal of the present paper. We shall propose a notion of set valued PDEs. In the contexts of multivariate stochastic control problems as in [14], which includes the time inconsistent problems such as the mean variance optimization problem, we shall characterize the corresponding dynamic value function via a *set valued HJB equation*.

As in standard stochastic control theory, the key to derive the PDE from the dynamic programming principle is the Itô formula. For this purpose, we first introduce appropriate notions of intrinsic derivatives for set valued functions, through the signed distance function \mathbf{r} . Our approach builds on analyses on the boundary of $\mathbb{V}(t, x)$, denoted as $\mathbb{V}_b(t, x)$, which is a surface and clearly determines the set $\mathbb{V}(t, x)$. So our derivatives, e.g. $\partial_x \mathbb{V}(t, x)$, are considered as functions on $\mathbb{V}_b(t, x)$, or equivalently, we write $\partial_x \mathbb{V}(t, x, y)$ for $(t, x, y) \in \mathbb{G}_{\mathbb{V}}$, the graph of $(t, x) \mapsto \mathbb{V}_b(t, x)$. It turns out that the first order derivatives $\partial_t \mathbb{V}, \partial_x \mathbb{V}$ are parallel to the normal vector \mathbf{n} , but the second order derivative $\partial_{xx} \mathbb{V}$ is not. Interestingly, $\partial_{xx} \mathbb{V}$ is in general not symmetric, but its projection on the normal direction is.

The first main contribution of this paper is the *set valued Itô formula*, which roughly reads: given a diffusion $dX_t = b_t dt + \sigma_t dB_t$,

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

We refer to Theorem 3.1 below for the precise meaning of the above formula. The main feature here is the involvement of arbitrary functions ξ and ζ , which takes values in the

tangent space, and $\mathcal{K}_\mathbb{V}\zeta$ is an appropriate correction term. Note that the main point of the above Itô formula is to characterize the flows on the boundary surface, the driving forces ξ, ζ in the tangent directions will not push the flows away from the surface. Our proof of the Itô formula relies heavily on the signed distance function \mathbf{r} .

Applying the Itô formula on the dynamic programming principle for the set valued function of the multivariate stochastic control problem, we derive our set valued HJB equation with terminal condition in the following form: for some Hamiltonian function $h_\mathbb{V}$,

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

where a takes values in a control set, and ζ takes values in the tangent space. As we see, the introduction of ζ (and ξ) in (1.2) is crucial. We note that ξ disappears in the equation since $\mathbf{n} \cdot \xi = 0$. However, $\mathcal{K}_\mathbb{V}\zeta$ is nonlinear in ζ and thus $\mathbf{n} \cdot \mathcal{K}_\mathbb{V}\zeta$ is a crucial component in the equation. The equation (1.3) can be rewritten equivalently in terms of the signed distance function \mathbf{r} , see (5.3) below. We emphasize that $\mathbf{n} = \mathbf{n}_\mathbb{V}$ is part of the solution here and the equation is satisfied only on the graph $\mathbb{G}_\mathbb{V}$, so wellposedness of (1.3) has a completely different nature than that of standard PDEs.

Our main result of the paper is that the dynamic 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. Without surprise, the proof is based on the dynamic programming principle and the main tool is our new Itô formula. Moreover, 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 the optimal controls, optimal in certain sense.

In the scalar case: $m = 1$, we can easily see that $\mathbb{V}(t, x) = [\underline{v}(t, x), \bar{v}(t, x)]$, where \underline{v}, \bar{v} are the value functions of the standard minimization and maximization problems, respectively. In this case, $\mathbf{n} = 1$ or -1 , and the tangent space is degenerate and thus $\zeta = 0$. Then (1.3) reduces exactly back to the standard HJB equation for \underline{v} and \bar{v} . So our set valued HJB equation is indeed a natural extension of the standard HJB equation to the multivariate setting. Moreover, noting again that \underline{v}, \bar{v} are the boundaries of \mathbb{V} in the scalar case, this partially justifies our focus on the boundary surface \mathbb{V}_b rather than on the whole set \mathbb{V} .

As an important application of our wellposedness result, we construct the moving scalarization for some time inconsistent problems, proposed by [14]¹ and Feinstein-Rudloff [9]. Note that we are in a multivariate setting and in general it is not feasible to optimize the multidimensional objects simultaneously. In practice quite often one considers the scalarized optimization problem: $\max_{y \in \mathbb{V}(0, x_0)} \varphi(y)$, where $\varphi \in \mathbb{R}^m \rightarrow \mathbb{R}$. This scalarized problem,

¹It's called dynamic utility function in [14]

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 7.1 below we solve this problem when φ and hence Φ are linear.

At this point we would like to mention that the present paper considers classical solutions only. In particular, this requires 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. Both the regularity and the nondegeneracy requirements are serious constraints in applications, for example they can hardly be satisfied for the set values of games in [10, 12]. It is our strong interest to remove these constraints and study viscosity solutions of more general set valued PDEs, so that we can apply our theory to more applications. 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 book Aubin-Frankowska [3] and the reference therein. However, our approach is completely different from those in set valued analysis. 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 remain on the boundary. These selectors have nice properties and are sufficient to characterize the whole sets. In particular, we need to mention the recent paper Ararat-Ma-Wu [1] on set valued backward SDE, which is highly relevant to our paper. Given our set valued 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 technically, except in some simple cases our 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.

Our approach is strongly motivated by the studies on surface evolution equations, see e.g. Sethian [16], Evans-Spruck [7], Soner [17], Barles-Soner-Souganidis [4], and the monograph Giga [11] 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 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 difference 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.

Our work is closely related to Soner-Touzi [19] which studies stochastic target problems by using mean curvature type geometric flows. In our contexts, their approach amounts to studying the following set valued function via its signed distance function $\widehat{\mathbf{r}}$:

$$\widehat{\mathbb{V}}(t) := \{(x, y) : x \in \mathbb{R}^d, y \in \mathbb{V}(t)\}, \text{ and thus } \widehat{\mathbb{V}}_b(t) := \{(x, y) : x \in \mathbb{R}^d, y \in \mathbb{V}_b(t)\}. \quad (1.4)$$

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} and $\widehat{\mathbf{r}}$ agree on the graph (both are 0 by definition), their derivatives are different on the graph, and consequently, the equation derived in [19] 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 $\widehat{\mathbf{r}}$ does not seem to connect to the standard HJB equation directly. Moreover, the normal vector $\widehat{\mathbf{n}}$ of $\widehat{\mathbb{V}}$ is also different from \mathbf{n} , and does not serve as a moving scalarization as we discussed. We shall provide more detailed discussions in Section 7.3 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 [6], Soner-Touzi [18]. 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 the classical solution is established in Section 6. In Section 7 we provide some further discussions, including the application on moving scalarization, the comparisons with [19] and [1], as well as an extension to the case that the terminal condition is non-degenerate. Finally we complete some technical proofs in Appendix.

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 $f : \mathbb{R}^d \times \mathbb{R}^m \rightarrow \mathbb{R}$, we take the convention that the second derivative $\nabla_{xy} f(x, y) := [\partial_{x_1 y} f, \dots, \partial_{x_d y} f] \in \mathbb{R}^{m \times d}$.

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 $\mathcal{D}_0^m \subset 2^{\mathbb{R}^m}$ denote the space of closed set \mathbb{D} in \mathbb{R}^m , and denote by \mathbb{D}_o and \mathbb{D}_b the interior and the boundary of \mathbb{D} , respectively. We equip \mathcal{D}_0^m with the metric:

$$\mathbf{d}(\mathbb{D}, \tilde{\mathbb{D}}) := d(\mathbb{D}, \tilde{\mathbb{D}}) \vee d(\mathbb{D}_b, \tilde{\mathbb{D}}_b), \quad (2.1)$$

where d is the standard Hausdorff distance, i.e.

$$d(\mathbb{D}, \tilde{\mathbb{D}}) := \left(\sup_{y \in \mathbb{D}} d(y, \tilde{\mathbb{D}}) \right) \vee \left(\sup_{\tilde{y} \in \tilde{\mathbb{D}}} d(\tilde{y}, \mathbb{D}) \right), \quad d(y, \tilde{\mathbb{D}}) := \inf_{\tilde{y} \in \tilde{\mathbb{D}}} |y - \tilde{y}|.$$

Introduce the signed distance function of \mathbb{D} : denoting by \mathbb{D}^c the complement of \mathbb{D} ,

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

It is obvious that

$$\mathbb{D}_o = \{y \in \mathbb{R}^m : \mathbf{r}_{\mathbb{D}}(y) < 0\}, \quad \mathbb{D}_b = \{y \in \mathbb{R}^m : \mathbf{r}_{\mathbb{D}}(y) = 0\}, \quad (2.3)$$

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 . It is clear that:

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

Moreover, for any $y \in O_\varepsilon(\mathbb{D}_b)$, for possibly a smaller $\varepsilon > 0$, 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:

$$y = \pi_{\mathbb{D}}(y) + \mathbf{r}_{\mathbb{D}}(y) \mathbf{n}_{\mathbb{D}}(y), \quad y \in O_\varepsilon(\mathbb{D}_b). \quad (2.5)$$

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

$$\mathbb{T}_{\mathbb{D}}(y) := \{\xi \in \mathbb{R}^m : \xi \cdot \mathbf{n}_{\mathbb{D}}(y) = 0\}, \quad y \in \mathbb{D}_b. \quad (2.6)$$

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

$$\lim_{\varepsilon \rightarrow 0} \frac{f(\theta(\varepsilon)) - f(y)}{\varepsilon} = \partial_y f(y) \cdot \theta'(0), \text{ for any smooth curve } \theta : \mathbb{R} \rightarrow \mathbb{D}_b \text{ with } \theta(0) = y. \quad (2.7)$$

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

$$\partial_y f(y) = \nabla_y \hat{f}(y) - [\nabla_y \hat{f}(y) \cdot \mathbf{n}_{\mathbb{D}}(y)] \mathbf{n}_{\mathbb{D}}(y) = \nabla_y \hat{f}(y) - \mathbf{n}_{\mathbb{D}}(y) \mathbf{n}_{\mathbb{D}}(y)^\top \nabla_y \hat{f}(y). \quad (2.8)$$

We emphasize that $\partial_y f(y)$ does not depend on the choice of the extension \hat{f} .

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

2.2 Set valued functions

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

$$\begin{aligned} \mathbb{V}_b(x) &:= (\mathbb{V}(x))_b, & \mathbf{r}_{\mathbb{V}}(x, y) &:= \mathbf{r}_{\mathbb{V}(x)}(y), & \mathbf{n}_{\mathbb{V}}(x, y) &:= \mathbf{n}_{\mathbb{V}(x)}(y), \\ \pi_{\mathbb{V}}(x, y) &:= \pi_{\mathbb{V}(x)}(y), & \mathbb{T}_{\mathbb{V}}(x, y) &:= \mathbb{T}_{\mathbb{V}(x)}(y), \end{aligned} \quad (2.9)$$

and introduce the graph of \mathbb{V} :

$$\mathbb{G}_{\mathbb{V}} := \{(x, y) : x \in \mathbb{R}, y \in \mathbb{V}_b(x)\}. \quad (2.10)$$

When there is no confusion, for notational simplicity 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 $\varepsilon > 0$.

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 $\hat{\mathbf{r}}_{\mathbb{V}}$ satisfies:*

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

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

(ii) *We can then extend $\mathbf{n}_{\mathbb{V}}(x, y) := \nabla_y \mathbf{r}_{\mathbb{V}}(x, y)$ 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.*

(iii) *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 (2.5) 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 (modifying the extension of $\mathbf{r}_\mathbb{V}$ if needed)

$$|y - \pi_\mathbb{V}(x, y)| \leq C|\mathbf{r}_\mathbb{V}(x, y)|. \quad (2.11)$$

We remark that (2.11) follows from (2.5) when $(x, y) \in O_\varepsilon(\mathbb{G}_\mathbb{V})$, and the existence of C for arbitrary (x, y) is due to the fact that $|\mathbf{r}_\mathbb{V}(x, y)| \geq \frac{\varepsilon}{2}$ for $(x, y) \notin O_\varepsilon(\mathbb{G}_\mathbb{V})$.

Fix $x_0 \in \mathbb{R}$. For each $y \in \mathbb{V}_b(x_0)$, consider the ODE: in light of Remark 2.1,

$$\Upsilon^y(x) = y - \int_{x_0}^x \nabla_x \mathbf{r}_\mathbb{V} \mathbf{n}_\mathbb{V}(\tilde{x}, \Upsilon^y(\tilde{x})) d\tilde{x}. \quad (2.12)$$

Then clearly the above ODE has a unique solution.

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

$$\mathbb{V}_b(x) = \{\Upsilon^y(x) : y \in \mathbb{V}_b(x_0)\}. \quad (2.13)$$

Consequently, (2.12) involves $\mathbf{r}_\mathbb{V}$ and $\mathbf{n}_\mathbb{V}$ only on $\mathbb{G}_\mathbb{V}$ and thus does not depend on the modification of $\mathbf{r}_\mathbb{V}$.

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 $x > x_0$, $\Upsilon(x) := \Upsilon^{y_0}(x) \in \mathbb{V}_b(x)$. Let $\varepsilon > 0$ be such that the original 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

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

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

$$\frac{d}{dx} \mathbf{r}(x, \Upsilon(x)) = \nabla_x \mathbf{r}(x, \Upsilon(x)) - \nabla_y \mathbf{r}(x, \Upsilon(x)) \cdot [\nabla_x \mathbf{r} \mathbf{n}(x, \Upsilon(x))].$$

Recall (2.5) and denote $\pi(x) := \pi(x, \Upsilon(x)) \in \mathbb{V}_b(x)$. By (2.4) we have

$$\frac{d}{dx} \mathbf{r}(x, \Upsilon(x)) = \nabla_x \mathbf{r}(x, \Upsilon(x)) [\nabla_y \mathbf{r}(x, \pi(x)) \cdot \mathbf{n}(x, \pi(x)) - \nabla_y \mathbf{r}(x, \Upsilon(x)) \cdot \mathbf{n}(x, \Upsilon(x))].$$

Therefore, by (2.5) and the continuous differentiability of $\nabla_y \mathbf{r}$ and \mathbf{n} , we have

$$\frac{d}{dx} \mathbf{r}(x, \Upsilon(x)) = \tilde{b}(x) \mathbf{r}(x, \Upsilon(x)), \quad x \in [x_0, \tau),$$

for some appropriate continuous function $\tilde{b} : \mathbb{R} \rightarrow \mathbb{R}$. Note that $\mathbf{r}(x_0, \Upsilon(x_0)) = \mathbf{r}(x_0, y_0) = 0$. Then the above ODE 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 $\{\Upsilon^y(x) : y \in \mathbb{V}_b(x_0)\} \subset \mathbb{V}_b(x)$ for all $x \geq x_0$. Similarly we can show that $\{\Upsilon^y(x) : y \in \mathbb{V}_b(x_0)\} \subset \mathbb{V}_b(x)$ for all $x \leq x_0$, and hence for all $x \in \mathbb{R}$.

On the other hand, for any $y \in \mathbb{V}_b(x)$, consider (2.12) starting from x :

$$\tilde{\Upsilon}^y(x') = y - \int_x^{x'} \nabla_x \mathbf{r} \mathbf{n}(\tilde{x}, \tilde{\Upsilon}^y(\tilde{x})) d\tilde{x}.$$

Then by the above result we have $y_0 := \tilde{\Upsilon}^y(x_0) \in \mathbb{V}_b(x_0)$. By the wellposedness of the ODE (2.12), one can easily show that $\tilde{\Upsilon}^y(x') = \Upsilon^{y_0}(x')$ for all $x' \in \mathbb{R}$, and thus $y = \tilde{\Upsilon}^y(x) = \Upsilon^{y_0}(x)$. This proves the opposite inclusion in (2.13). \blacksquare

Remark 2.3 (i) Later on we will define $\partial_x \mathbb{V}(x, y) = -\nabla_x \mathbf{r} \mathbf{n}_{\mathbb{V}}(x, y)$ for $(x, y) \in \mathbb{G}_{\mathbb{V}}$, see (2.23) below. Then (2.12) can be rewritten as

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

Thus (2.13) can be viewed as the fundamental theorem of calculus for set valued functions:

$$\mathbb{V}_b(x) = \mathbb{V}_b(x_0) + \int_{x_0}^x \partial_x \mathbb{V}(\tilde{x}, \mathbb{V}_b(\tilde{x})) d\tilde{x}. \quad (2.15)$$

(ii) However, (2.15) should be interpreted as (2.14) and (2.13), rather than the meaning in the standard set valued analysis, which roughly speaking considers

$$\tilde{\Upsilon}(x) := y + \int_{x_0}^x \partial_x \mathbb{V}(\tilde{x}, \tilde{\gamma}(\tilde{x})) d\tilde{x}, \quad \forall y \in \mathbb{V}_b(x_0), \tilde{\gamma}(\tilde{x}) \in \mathbb{V}_b(\tilde{x}). \quad (2.16)$$

The above $\tilde{\Upsilon}(x)$ is in general not in $\mathbb{V}_b(x)$. See also related discussion in Section 7.4 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.

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

Proposition 2.4 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 $\Upsilon(x) := \Upsilon^{y_0}(x)$ determined by (2.12) 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 $\theta(x_0) = y_0$,

$$\overline{\lim}_{x \rightarrow x_0} \frac{1}{|x - x_0|} [L_{\Upsilon}(x_0, x) - L_{\theta}(x_0, x)] \leq 0, \quad (2.17)$$

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

Consider a function $f : \mathbb{G}_\mathbb{V} \rightarrow \mathbb{R}$. For fixed x , the intrinsic derivative $\partial_y f(x, y)$ for $y \in \mathbb{V}_b(x)$ is defined (2.7) or equivalently by (2.8). We next define the intrinsic derivative of f with respect to x following the local geodesic Υ defined by (2.12):

$$\partial_x f(x_0, y_0) := \lim_{x \rightarrow x_0} \frac{f(x, \Upsilon^{y_0}(x)) - f(x_0, y_0)}{x - x_0}, \quad (x_0, y_0) \in \mathbb{G}_\mathbb{V}. \quad (2.18)$$

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

$$\partial_x f(x_0, y_0) = \lim_{x \rightarrow x_0} \frac{\widehat{f}(x, \Upsilon^{y_0}(x)) - \widehat{f}(x_0, y_0)}{x - x_0} = \nabla_x \widehat{f}(x_0, y_0) - \nabla_x \mathbf{r}_\mathbb{V} \nabla_y \widehat{f} \cdot \mathbf{n}_\mathbb{V}(x_0, y_0). \quad (2.19)$$

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

We say $f \in C^1(\mathbb{G}_\mathbb{V})$ if f has continuous intrinsic derivatives $\partial_y f$ and $\partial_x f$. By (2.18), it is obvious that $\partial_x f$ is linear on f , and the product rule and chain rule remain true:

$$\begin{aligned} \partial_x(fg) &= g\partial_x f + f\partial_x g, \text{ for all } f, g \in C^1(\mathbb{G}_\mathbb{V}); \\ \partial_x[f(g)] &= f'(g)\partial_x g, \text{ for all } f \in C^1(\mathbb{R}; \mathbb{R}), g \in C^1(\mathbb{G}_\mathbb{V}). \end{aligned} \quad (2.20)$$

2.3 Intrinsic derivatives of set valued functions

We now extend all the analysis to functions $\mathbb{V} : [0, T] \times \mathbb{R}^d \rightarrow \mathcal{D}_2^m$.² 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 (2.9) and denote

$$\mathbb{G}_\mathbb{V} := \{(t, x, y) : (t, x) \in [0, T] \times \mathbb{R}^d, y \in \mathbb{V}_b(t, x)\}. \quad (2.21)$$

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

Recall (2.18) and (2.19) when \mathbb{V} is defined on \mathbb{R} . Now for our more general \mathbb{V} and for any function $f : \mathbb{G}_\mathbb{V} \rightarrow \mathbb{R}$, 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:

$$\partial_{x_i x_j} f := \partial_{x_i}(\partial_{x_j} f). \quad (2.22)$$

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

Finally, by considering the special function $f_0(x, y) := y$ and its natural extension $\widehat{f}_0(x, y) = y$, applying (2.19) component wise we define the intrinsic derivatives of \mathbb{V} .

²In this and the next section we may allow infinite time horizon $[0, \infty)$. However, for later sections we require T to be finite, so for simplicity we consider finite T here as well.

Definition 2.5 For any $(t, x, y) \in \mathbb{G}_\mathbb{V}$ and by denoting $f_0(x, y) := y$, define

$$\partial_t \mathbb{V}(t, x, y) := \partial_t f_0(t, x, y) = -\nabla_t \mathbf{r}(t, x, y) \mathbf{n}(t, x, y) \in \mathbb{R}^m; \quad (2.23)$$

$$\partial_{x_i} \mathbb{V}(t, x, y) := \partial_{x_i} f_0(t, x, y) = -\nabla_{x_i} \mathbf{r}(t, x, y) \mathbf{n}(t, x, y) \in \mathbb{R}^m, \quad i = 1, \dots, d.$$

We recall Remark 2.3 and note that (2.19) becomes: for any $f \in C^1(\mathbb{G}_\mathbb{V})$,

$$\partial_t f = \nabla_t \widehat{f} + \nabla_y \widehat{f} \cdot \partial_t \mathbb{V}, \quad \partial_{x_i} f = \nabla_{x_i} \widehat{f} + \nabla_y \widehat{f} \cdot \partial_{x_i} \mathbb{V}, \quad \text{on } \mathbb{G}_\mathbb{V}. \quad (2.24)$$

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 (2.19) or (2.24) repeatedly.

Lemma 2.6 Assume $\mathbf{r}_\mathbb{V} \in C^2([0, T] \times \mathbb{R}^d; \mathcal{D}_2^m)$. Then

$$\partial_{x_i x_j} \mathbb{V}(t, x, y) = -\nabla_{x_i x_j} \mathbf{r} \mathbf{n}(t, x, y) - \nabla_{x_j} \mathbf{r} \partial_{x_i} \mathbf{n}(t, x, y); \quad (2.25)$$

$$\partial_x \mathbf{n}^i = \nabla_{xy} \mathbf{r}; \quad \partial_y \mathbf{n}^i = \nabla_{y_i y} \mathbf{r}. \quad (2.26)$$

The proof is quite straightforward, we thus postpone it to Appendix. Throughout the paper, we shall take the notational convention:

$$\partial_x \mathbb{V} := [\partial_{x_1} \mathbb{V}, \dots, \partial_{x_d} \mathbb{V}] \in \mathbb{R}^{m \times d}, \quad \partial_{xx} \mathbb{V}^i := [\partial_{x_1 x} \mathbb{V}^i, \dots, \partial_{x_d x} \mathbb{V}^i] \in \mathbb{R}^{d \times d}; \quad (2.27)$$

$$\partial_x \mathbf{n} = [\partial_{x_1} \mathbf{n}, \dots, \partial_{x_d} \mathbf{n}] \in \mathbb{R}^{m \times d}, \quad \partial_y \mathbf{n} = [\partial_{y_1} \mathbf{n}, \dots, \partial_{y_m} \mathbf{n}] \in \mathbb{R}^{m \times m}.$$

Remark 2.7 (i) At $(t, x, y) \in \mathbb{G}_\mathbb{V}$, since $|\mathbf{n}|^2 = 1$, by (2.20) 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 (2.25) provides an orthogonal decomposition of $\partial_{xx} \mathbb{V}$. In particular, unlike the first order derivatives in (2.23), $\partial_{x_i x_j} \mathbb{V}$ is in general not parallel to \mathbf{n} .

(ii) It is clear that

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

is symmetric. However, in general $\partial_{xx} \mathbb{V}$ is not symmetric:

$$\nabla_{x_j} \mathbf{r} \partial_{x_i} \mathbf{n} \neq \nabla_{x_i} \mathbf{r} \partial_{x_j} \mathbf{n}, \quad i \neq j.$$

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

Example 2.8 (i) Let $w : [0, T] \times \mathbb{R}^2 \rightarrow \mathbb{R}^2$ and $u : [0, T] \times \mathbb{R}^2 \rightarrow (0, \infty)$ be continuously differentiable. Set, with $d = m = 2$,

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

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

It is clear that

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

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

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

Then, by (2.4), Definition 2.5, and Lemma 2.6, at $(t, x, y) \in \mathbb{G}_\mathbb{V}$ we have

$$\begin{aligned} \mathbf{n} &= \frac{y-w}{u}; & \partial_t \mathbb{V} &= [\nabla_t w \cdot \mathbf{n} + \nabla_t u] \mathbf{n}; & \partial_{x_i} \mathbb{V} &= [\nabla_{x_i} w \cdot \mathbf{n} + \nabla_{x_i} u] \mathbf{n}; \\ \partial_{x_i} \mathbf{n} &= \frac{1}{u} [-\nabla_{x_i} w + [\mathbf{n} \cdot \nabla_{x_i} w] \mathbf{n}], & \partial_{y_i} \mathbf{n}^j &= \frac{1}{u} [\mathbf{1}_{\{i=j\}} - \mathbf{n}^i \mathbf{n}^j]; \\ \partial_{x_i x_j} \mathbb{V} &= -\left[\frac{1}{u} [\nabla_{x_i} w \cdot \nabla_{x_j} w - (\nabla_{x_i} w \cdot \mathbf{n})(\nabla_{x_j} w \cdot \mathbf{n})] - \nabla_{x_i x_j} w \cdot \mathbf{n} - \nabla_{x_i x_j} u \right] \mathbf{n} \\ &\quad - \frac{1}{u} [\nabla_{x_j} w \cdot \mathbf{n} + \nabla_{x_j} u] [\nabla_{x_i} w - (\nabla_{x_i} w \cdot \mathbf{n}) \mathbf{n}]. \end{aligned} \tag{2.28}$$

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

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

$$\nabla_t u + \frac{1}{2} \text{tr}(\nabla_{xx} u) = 0. \tag{2.29}$$

Then by (2.28) we have $\partial_t \mathbb{V} = \nabla_t u \mathbf{n}$, $\partial_{x_i x_j} \mathbb{V} = \nabla_{x_i x_j} u \mathbf{n}$, on $\mathbb{G}_\mathbb{V}$. Thus \mathbb{V} satisfies the following equation: $\partial_t \mathbb{V} + \frac{1}{2} \text{tr}(\partial_{xx} \mathbb{V}) = 0$, on $\mathbb{G}_\mathbb{V}$. This clearly implies the following set valued heat equation:

$$\mathbf{n} \cdot [\partial_t \mathbb{V} + \frac{1}{2} \text{tr}(\partial_{xx} \mathbb{V})] = 0, \quad \text{on } \mathbb{G}_\mathbb{V}. \tag{2.30}$$

We remark that we assumed 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 (2.28), $\partial_x \mathbf{n}$, $\partial_y \mathbf{n}$, and $\partial_{xx} \mathbb{V}$ explode when $u = 0$. This motivates us to define the following space.

Definition 2.9 (i) We say $\mathbb{V} \in C^{1,2}([0, T] \times \mathbb{R}^d; \mathcal{D}_2^m)$ if $\mathbf{r}_\mathbb{V} \in C^2(O_\varepsilon(\mathbb{G}_\mathbb{V}); \mathbb{R})$ 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 first introduce 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 natural filtration generated by B . For a generic Euclidian space E and $p \geq 1$, let $\mathbb{L}_{loc}^p(E)$ denote the space of \mathbb{F} -progressively measurable E -valued processes θ such that $\int_0^T |\theta_t|^p dt < \infty$, a.s., and $\mathbb{L}_{loc}^p(\mathbb{R}^m; E)$ the space of \mathbb{F} -progressively measurable random fields $\xi : (t, \omega, y) \in [0, T] \times \Omega \times \mathbb{R}^m \rightarrow E$ such that $\xi(\cdot, \cdot, 0) \in \mathbb{L}_{loc}^p(E)$ and ξ is uniformly Lipschitz continuous in y .

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

$$X_t := x_0 + \int_0^t b_s ds + \int_0^t \sigma_s dB_s, \quad (3.1)$$

and introduce the (random) differential operators: recalling (2.27),

$$\mathcal{L}^{b, \sigma} \mathbb{V}(t, \omega, x, y) := \partial_t \mathbb{V} + \partial_x \mathbb{V} b + \frac{1}{2} \text{tr}(\sigma^\top \partial_{xx} \mathbb{V} \sigma); \quad (3.2)$$

$$\mathcal{K}_{\mathbb{V}}^\sigma \zeta(t, \omega, x, y) := \text{tr}(\zeta^\top \partial_x \mathbf{n} \sigma + \frac{1}{2} \zeta^\top \partial_y \mathbf{n} \zeta) \mathbf{n}. \quad (3.3)$$

where $\text{tr}(\sigma^\top \partial_{xx} \mathbb{V} \sigma) \in \mathbb{R}^m$ with i -th component $\text{tr}(\sigma^\top \partial_{xx} \mathbb{V}^i \sigma)$. Moreover, recalling Remark 2.1, we may extend the derivatives of \mathbb{V} and \mathbf{n} to $[0, T] \times \mathbb{R}^d \times \mathbb{R}^m$.

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, and 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 $y \in \mathbb{R}^m$, let Υ^y denote the unique strong solution of SDE: recalling Remark 2.1,*

$$\Upsilon_t^y = y + \int_0^t [\mathcal{L}^{b, \sigma} \mathbb{V} - \kappa_{\mathbb{V}}^\sigma \zeta + \xi](s, X_s, \Upsilon_s^y) ds + \int_0^t [\partial_x \mathbb{V} \sigma + \zeta](s, X_s, \Upsilon_s^y) dB_s. \quad (3.4)$$

(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*

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

In particular, in this case no extension is needed in (3.4).

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

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

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

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

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

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

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

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

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

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

We claim that, when $y \in \mathbb{V}_b(t, x)$,

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

Indeed, in this case we have $\nabla_y \mathbf{r} = \mathbf{n}$. Then by (2.23), (2.25), and (2.26) we have:

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

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

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

Thus (3.6) holds true.

Since $\xi(t, y), \zeta_i(t, y) \in \mathbb{T}_\mathbb{V}(t, X_t, y)$, we have

$$\Lambda(t, X_t, y) = M(t, X_t, y) = 0, \quad y \in \mathbb{V}_b(t, X_t).$$

Now for any $(t, x, y) \in O_\varepsilon(\mathbb{G}_\mathbb{V})$, since $\pi(t, x, y) \in \mathbb{V}_b(t, x)$, then

$$\Lambda(t, X_t, \pi(t, X_t, \Upsilon_t)) = 0, \quad M(t, X_t, \pi(t, X_t, \Upsilon_t)) = 0, \quad 0 \leq t \leq \tau.$$

Note further that $|\Upsilon_t - \pi(t, X_t, \Upsilon_t)| = |\mathbf{r}(t, X_t, \Upsilon_t)|$. Then, by the desired regularity in Definition 2.9 (i) we have

$$\begin{aligned}\Lambda(t, X_t, \Upsilon_t) &= \Lambda(t, X_t, \Upsilon_t) - \Lambda(t, X_t, \pi(t, X_t, \Upsilon_t)) = \tilde{b}_t \mathbf{r}(t, X_t, \Upsilon_t); \\ M(t, X_t, \Upsilon_t) &= M(t, X_t, \Upsilon_t) - M(t, X_t, \pi(t, X_t, \Upsilon_t)) = \tilde{\sigma}_t \mathbf{r}(t, X_t, \Upsilon_t); \\ \text{where } |\tilde{b}_t| &\leq C[1 + |b_t| + |\sigma_t|^2 + |\zeta_t(0)|^2], \quad |\tilde{\sigma}_t| \leq C[|\sigma_t| + |\zeta_t(0)|].\end{aligned}\tag{3.7}$$

In particular, $\tilde{b} \in \mathbb{L}_{loc}^1(\mathbb{R})$, $\tilde{\sigma} \in \mathbb{L}_{loc}^2(\mathbb{R}^d)$. Note that (3.5) becomes:

$$d\mathbf{r}(t, X_t, \Upsilon_t) = \tilde{b}_t \mathbf{r}(t, X_t, \Upsilon_t) dt + \tilde{\sigma}_t \mathbf{r}(t, X_t, \Upsilon_t) dB_t, \quad 0 \leq t \leq \tau.\tag{3.8}$$

Introduce

$$\tilde{\Gamma}_t := \exp\left(-\int_0^t \tilde{\sigma}_s \cdot dB_s - \int_0^t [\tilde{b}_s + \frac{1}{2}|\tilde{\sigma}_s|^2] ds\right).\tag{3.9}$$

Then, recalling $\mathbf{r}(0, X_0, \Upsilon_0) = \mathbf{r}(0, x_0, y_0) = 0$, we have

$$\mathbf{r}(t, X_t, \Upsilon_t) = \mathbf{r}(0, X_0, \Upsilon_0) \tilde{\Gamma}_t = 0, \quad 0 \leq t \leq \tau.$$

This implies $\tau = T$, a.s. and thus $\Upsilon_t \in \mathbb{V}_b(t, X_t)$, $0 \leq t \leq T$, a.s.

Moreover, assume further that \mathbb{V}_b takes values in connected, compact sets. Note that $y \mapsto \Upsilon_t^y$ is a homeomorphism almost surely (See Kunita [15]). 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, $y \mapsto \Upsilon_t^y$ is an open mapping in relative topologies of $\mathbb{V}_b(0, x_0), \mathbb{V}_b(t, X_t^{0, x_0})$. Therefore, $\mathbb{V}_b(0, x_0)$ maps to an open set in $\mathbb{V}_b(t, X_t^{0, x_0})$. This concludes the equality as we assumed connectedness.

(ii) In this case, by (3.6) we have

$$\Lambda(t, X_t, y) = \mathbf{n}(t, X_t, y) \cdot \xi_t(y) \leq 0, \quad M(t, X_t, y) = 0, \quad \text{for all } y \in \mathbb{V}_b(t, X_t).\tag{3.10}$$

Fix $y_0 \in \mathbb{V}_b(0, x_0)$ and denote $\Upsilon = \Upsilon^{y_0}$. Let $\delta < \frac{\varepsilon}{2}$ be small enough so that $\mathbf{r}(0, x_0, y_0) < -\delta$. Introduce recursively a sequence of stopping times: $\tau_0 := 0$, and for $n = 0, 1, \dots$,

$$\begin{aligned}\tau_{2n+1} &:= \inf \{t > \tau_{2n} : \mathbf{r}(t, X_t, \Upsilon_t) = -\delta\} \wedge T; \\ \tau_{2n+2} &:= \inf \{t > \tau_{2n+1} : |\mathbf{r}(t, X_t, \Upsilon_t)| = 2\delta\} \wedge T.\end{aligned}$$

Since $\mathbf{r}(0, x_0, y_0) < -\delta$, it is clear that $\mathbf{r}(t, X_t, \Upsilon_t) \leq -\delta$, $\tau_0 \leq t \leq \tau_1$. Now for $\tau_1 \leq t \leq \tau_2$, note that $|\mathbf{r}(t, X_t, \Upsilon_t)| \leq 2\delta$. Then by (3.10) and following the same arguments as in (3.8) we derive: for $\tau_1 \leq t \leq \tau_2$ and denoting $\pi_t := \pi(t, X_t, \Upsilon_t)$,

$$d\mathbf{r}(t, X_t, \Upsilon_t) = [\tilde{b}_t \mathbf{r}(t, X_t, \Upsilon_t) + \mathbf{n}(t, X_t, \pi_t) \cdot \xi_t(\pi_t)] dt + \tilde{\sigma}_t \mathbf{r}(t, X_t, \Upsilon_t) dB_t.$$

Since $\mathbf{r}(\tau_1, X_{\tau_1}, \Upsilon_{\tau_1}) < 0$ and $\mathbf{n}(t, X_t, \pi_t) \cdot \xi_t(\pi_t) \leq 0$, we can easily see that $\mathbf{r}(t, X_t, \Upsilon_t) < 0$ for all $\tau_1 \leq t \leq \tau_2$. In particular, $\mathbf{r}(\tau_2, X_{\tau_2}, \Upsilon_{\tau_2}) = -2\delta < -\delta$ on $\{\tau_2 < T\}$. Repeating the arguments we see that $\mathbf{r}(t, X_t, \Upsilon_t) < 0$ for all $0 \leq t \leq \tau_n$ and for all n .

It remains to show that $\tau_n = T$ for all n large, which clearly implies that $\mathbf{r}(t, X_t, \Upsilon_t) < 0$ for all $0 \leq t \leq T$. Assume by contradiction that $\tau_n < T$ for all n . Then

$$\mathbf{r}(\tau_{2n+1}, X_{\tau_{2n+1}}, \Upsilon_{\tau_{2n+1}}) = -\delta, \quad \mathbf{r}(\tau_{2n+2}, X_{\tau_{2n+2}}, \Upsilon_{\tau_{2n+2}}) = -2\delta, \quad \forall n.$$

Denote $\tau_* := \lim_{n \rightarrow \infty} \tau_n$. Sending $n \rightarrow \infty$ at above and by the continuity of X_t and Υ_t , we obtain $\mathbf{r}(\tau_*, X_{\tau_*}, \Upsilon_{\tau_*}) = -\delta$ and $\mathbf{r}(\tau_*, X_{\tau_*}, \Upsilon_{\tau_*}) = -2\delta$, which is a desired contradiction.

(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 $0 \leq t < T$, we shall also consider the shifted Brownian motion $B_s^t := B_s - B_t$, and the shifted filtration $\mathbb{F}^t := \mathbb{F}^{B^t}$ on $[t, T]$. For a generic Euclidian 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 \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 $\alpha \in \mathcal{A}_t$,

$$\begin{aligned} X_s^{t,x,\alpha} &= x + \int_t^s b(r, X_r^{t,x,\alpha}, \alpha_r) dr + \int_t^s \sigma(r, X_r^{t,x,\alpha}, \alpha_r) dB_r, \\ Y_s^{t,x,\alpha} &= g(X_T^{t,x,\alpha}) + \int_s^T f(r, X_r^{t,x,\alpha}, Y_r^{t,x,\alpha}, Z_r^{t,x,\alpha}, \alpha_r) dr - \int_s^T Z_r^{t,x,\alpha} dB_r. \end{aligned} \quad (4.1)$$

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:

$$\mathbb{V}(t, x) := \text{cl}\{Y_t^{t,x,\alpha} : \alpha \in \mathcal{A}_t\} \subset \mathbb{R}^m. \quad (4.2)$$

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

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

$$\underline{v}(t, x) := \inf_{\alpha \in \mathcal{A}_t} Y_t^{t,x,\alpha}, \quad \bar{v}(t, x) := \inf_{\alpha \in \mathcal{A}_t} Y_t^{t,x,\alpha}. \quad (4.3)$$

Then it is obvious that

$$\mathbb{V}(t, x) = [\underline{v}(t, x), \bar{v}(t, x)]. \quad (4.4)$$

That is, the standard optimization problems are characterizing the boundary of our set valued 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:*

$$V_0 := \sup_{\alpha \in \mathcal{A}} \left[\mathbb{E}[X_T^{0,x_0,\alpha}] - \frac{\lambda}{2} \text{Var}(X_T^{0,x_0,\alpha}) \right], \quad \text{where } X_s^{t,x,\alpha} = x + \int_t^s \alpha_r dr + \int_t^s \alpha_r dB_r. \quad (4.5)$$

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

$$\begin{aligned} \mathbb{V}(t, x) &:= \text{cl}\{Y_t^{t,x,\alpha} : \alpha \in \mathcal{A}_t\}, \quad \text{where} \\ Y_s^{t,x,\alpha,1} &= X_T^{t,x,\alpha} - \int_s^T Z_r^{t,x,\alpha,1} dB_r, \quad Y_s^{t,x,\alpha,2} = |X_T^{t,x,\alpha}|^2 - \int_s^T Z_r^{t,x,\alpha,2} dB_r. \end{aligned} \quad (4.6)$$

Then one can easily verify that

$$V_0 := \sup_{y \in \mathbb{V}(0, x_0)} \varphi(y), \quad \text{where } \varphi(y) := y_1 + \frac{\lambda}{2} |y_1|^2 - \frac{\lambda}{2} y_2. \quad (4.7)$$

Our goal of this paper is to characterize the dynamic set valued function \mathbb{V} . Provided we know the set $\mathbb{V}(0, x_0)$, it is trivial to solve the deterministic optimization problem (4.7). We also refer to Section 7.1 below for the related time inconsistency issue.

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

$$\begin{aligned} \mathbb{V}(t, x) &= \text{cl}\{y \in \mathbb{R}^m : \exists \alpha, Z \text{ such that } Y_T^{t,x,y,\alpha,Z} = g(X_T^{t,x,\alpha}), \text{ a.s.}\}, \\ \text{where } Y_s^{t,x,y,\alpha,Z} &= y - \int_t^s f(r, X_r^{t,x,\alpha}, Y_r^{t,x,y,\alpha,Z}, Z_r, \alpha_r) dr + \int_t^s Z_r dB_r. \end{aligned} \quad (4.8)$$

(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 $\mathbb{G} : \mathbb{R}^d \rightarrow \mathcal{D}_0^m$. That is,

$$\mathbb{V}(t, x) := \text{cl}\{y \in \mathbb{R}^m : \exists \alpha, Z \text{ such that } Y_T^{t,x,y,\alpha,Z} \in \mathbb{G}(X_T^{t,x,\alpha}), \text{ a.s.}\}. \quad (4.9)$$

All our results in this paper can be extended to this case as well, see Section 7.2 below.

In the rest of this section, we establish the dynamic programming principle (DPP) for \mathbb{V} . For this purpose, we first specify the technical conditions on the coefficients.

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

(ii) $f : (t, x, y, z, a) \in [0, T] \times \mathbb{R}^d \times \mathbb{R}^m \times \mathbb{R}^{m \times d} \times A \rightarrow \mathbb{R}^m$ 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) $g : x \in \mathbb{R}^d \rightarrow \mathbb{R}^m$ 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). Now for $0 \leq t < T$, $x \in \mathbb{R}^d$, \mathbb{F}^t -stopping time $\tau \geq t$, $\phi \in \mathbb{L}^2(\mathcal{F}_\tau^t, \mathbb{R}^m)$, and $\alpha \in \mathcal{A}_t$, introduce:

$$Y_s^{\tau, \phi; t, x, \alpha} = \phi + \int_s^\tau f(r, X_r^{t, x, \alpha}, Y_r^{\tau, \phi; t, x, \alpha}, Z_r^{\tau, \phi; t, x, \alpha}, \alpha_r) dr - \int_s^\tau Z_r^{\tau, \phi; t, x, \alpha} dB_r. \quad (4.10)$$

We then have the crucial DPP.

Theorem 4.5 Let Assumption 4.4 hold and \mathbb{V} be defined by (4.2). For any $0 \leq t < T$, $x \in \mathbb{R}^d$, and any \mathbb{F}^t -stopping time $\tau \geq t$, it holds

$$\mathbb{V}(t, x) = \text{cl}\{Y_t^{\tau, \phi; t, x, \alpha} : \forall \alpha \in \mathcal{A}_t, \phi \in \mathbb{L}^2(\mathcal{F}_\tau^t, \mathbb{R}^m) \text{ s.t. } \phi \in \mathbb{V}(\tau, X_\tau^{t, x, \alpha}) \text{ a.s.}\}. \quad (4.11)$$

Proof Without loss of generality we prove (4.11) only at $(t, x) = (0, x_0)$, and for notational simplicity we omit the superscripts $0, x_0$. Denote the right side of (4.11) as $\tilde{\mathbb{V}}(0, x_0)$.

Step 1. We first 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 $|y_0 - Y_0^\alpha| \leq \varepsilon$. Denote $\phi := Y_\tau^\alpha$. It is clear that $Y_0^\alpha = Y_0^{\tau, \phi; \alpha}$ and thus $|y_0 - Y_0^{\tau, \phi; \alpha}| \leq \varepsilon$. We claim that

$$\phi \in \mathbb{V}(\tau, X_\tau^\alpha) \text{ a.s.} \quad (4.12)$$

Then $Y_0^{\tau, \phi; \alpha} \in \tilde{\mathbb{V}}(0, x_0)$, and by the arbitrariness of $\varepsilon > 0$ we obtain $y_0 \in \tilde{\mathbb{V}}(0, x_0)$.

To see (4.12), we 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

$$(\omega \oplus_t \tilde{\omega})_s := \omega_s \mathbf{1}_{[0, t)} + (\omega_t + \tilde{\omega}_s) \mathbf{1}_{[t, T]}(s), \quad \xi^{t, \omega}(\tilde{\omega}) := \xi(\omega \oplus_t \tilde{\omega}). \quad (4.13)$$

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 $\psi = X, Y, Z$,

$$\begin{aligned} X_s^{\alpha, \tau, \omega} &= X_\tau^\alpha(\omega) + \int_\tau^s b(r, X_r^{\alpha, \tau, \omega}, \alpha_r^{\tau, \omega}) dr + \int_\tau^s \sigma(r, X_r^{\alpha, \tau, \omega}, \alpha_r^{\tau, \omega}) dB_r^\tau, \\ Y_s^{\alpha, \tau, \omega} &= g(X_T^{\alpha, \tau, \omega}) + \int_s^T f(r, X_r^{\alpha, \tau, \omega}, Y_r^{\alpha, \tau, \omega}, Z_r^{\alpha, \tau, \omega}, \alpha_r^{\tau, \omega}) dr - \int_s^T Z_r^{\alpha, \tau, \omega} dB_r^\tau. \end{aligned} \quad (4.14)$$

This implies (4.12) immediately.

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 $|y_0 - Y_0^{\tau, \phi; \alpha}| \leq \varepsilon$ and $\mathbb{P}(E) = 1$, where

$$E := \{\omega \in \Omega : \phi(\omega) \in \mathbb{V}(\tau(\omega), X_\tau^\alpha(\omega))\}.$$

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

$$|Y_0^{\hat{\alpha}} - Y_0^{\tau, \phi; \alpha}| \leq C\varepsilon. \quad (4.15)$$

Then $|y_0 - Y_0^{\hat{\alpha}}| \leq \varepsilon + C\varepsilon$. Since $Y_0^{\hat{\alpha}} \in \mathbb{V}(0, x_0)$ by definition, then $y \in \mathbb{V}(0, x_0)$.

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

$$\begin{aligned} E_i^\tau &:= \{t_{i-1} < \tau \leq t_i\}, & E_j^\phi &:= \{\phi \in O_j^m\}, & E_k^\alpha &:= \{X_\tau^\alpha \in O_k^d\}, \\ \tau_\varepsilon &:= \sum_{i=1}^n t_i \mathbf{1}_{E_i^\tau}, & E_\theta &:= E \cap E_i^\tau \cap E_j^\phi \cap E_k^\alpha, \text{ where } \theta = i, j, k. \end{aligned} \quad (4.16)$$

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

$$\mathbb{P}(\{\tau > t_\theta\} \cap E_\theta) \leq \varepsilon^2 \mathbb{P}(E_\theta), \quad \text{where } t_\theta := \tau(\omega^\theta), \quad x_\theta := X_{t_\theta}^\alpha(\omega^\theta). \quad (4.17)$$

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

$$|\phi(\omega^\theta) - Y_{t_\theta}^{t_\theta, x_\theta, \alpha^\theta}| \leq \varepsilon. \quad (4.18)$$

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

$$\hat{\alpha}_t(\omega) := \alpha_t(\omega) \mathbf{1}_{[0, \tau_\varepsilon(\omega))}(t) + \mathbf{1}_{[\tau_\varepsilon(\omega), T]}(t) \left[\sum_{\theta} \mathbf{1}_{E_\theta}(\omega) \alpha_t^\theta(\omega^{t_\theta}) + a_0 \mathbf{1}_{E^c} \right], \quad (4.19)$$

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

Step 3. We now verify (4.15). First, for any $\theta = (i, j, k)$ such that $E_\theta \neq \emptyset$, a.s. on E_θ we have $\tau_\varepsilon = t_i \geq t_\theta$ and, denoting $(X^\theta, Y^\theta, Z^\theta) := (X^{t_\theta, x_\theta, \alpha^\theta}, Y^{t_\theta, x_\theta, \alpha^\theta}, Z^{t_\theta, x_\theta, \alpha^\theta})$,

$$\begin{aligned} X_t^{\hat{\alpha}} &= X_{t_\theta}^\alpha + \int_{t_\theta}^t b(s, X_s^{\hat{\alpha}}, \alpha_s) ds + \int_{t_\theta}^t \sigma(s, X_s^{\hat{\alpha}}, \alpha_s) dB_s, \quad t \in [t_\theta, t_i], \\ X_t^{\hat{\alpha}} &= X_{t_i}^{\hat{\alpha}} + \int_{t_i}^t b(s, X_s^{\hat{\alpha}}, \alpha_s^\theta(B^{t_\theta})) ds + \int_{t_i}^t \sigma(s, X_s^{\hat{\alpha}}, \alpha_s^\theta(B^{t_\theta})) dB_s, \quad t \in [t_i, T], \\ X_t^\theta &= x_\theta + \int_{t_\theta}^t b(s, X_s^\theta, \alpha_s^\theta(B^{t_\theta})) ds + \int_{t_\theta}^t \sigma(s, X_s^\theta, \alpha_s^\theta(B^{t_\theta})) dB_s, \quad t \in [t_\theta, T]. \end{aligned}$$

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

$$\mathbb{E}_{\mathcal{F}_{t_\theta}} \left[\sup_{t_\theta \leq t \leq t_i} |X_t^{\hat{\alpha}} - X_t^\theta|^2 \right] \leq C\varepsilon^2, \quad \text{and then} \quad \mathbb{E}_{\mathcal{F}_{t_\theta}} \left[\sup_{t_i \leq t \leq T} |X_t^{\hat{\alpha}} - X_t^\theta|^2 \right] \leq C\varepsilon^2. \quad (4.20)$$

Similarly, note that

$$\begin{aligned} Y_t^{\hat{\alpha}} &= g(X_T^{\hat{\alpha}}) + \int_t^T f(s, X_s^{\hat{\alpha}}, Y_s^{\hat{\alpha}}, Z_s^{\hat{\alpha}}, \alpha_s^\theta(B^{t_\theta})) ds - \int_t^T Z_s^{\hat{\alpha}} dB_s, \quad t \in [t_i, T], \\ Y_t^{\hat{\alpha}} &= Y_{t_i}^{\hat{\alpha}} + \int_t^{t_i} f(s, X_s^{\hat{\alpha}}, Y_s^{\hat{\alpha}}, Z_s^{\hat{\alpha}}, \alpha_s) ds - \int_t^{t_i} Z_s^{\hat{\alpha}} dB_s, \quad t \in [t_\theta, t_i], \\ Y_t^\theta &= g(X_T^\theta) + \int_t^T f(s, X_s^\theta, Y_s^\theta, Z_s^\theta, \alpha_s^\theta(B^{t_\theta})) ds - \int_t^T Z_s^\theta dB_s, \quad t \in [t_\theta, T]. \end{aligned}$$

Then, by (4.20) and standard BSDE estimates we have

$$\mathbb{E}_{\mathcal{F}_{t_\theta}} \left[\sup_{t_i \leq t \leq T} |Y_t^{\hat{\alpha}} - Y_t^\theta|^2 \right] \leq C\varepsilon^2, \quad \text{and then} \quad \mathbb{E}_{\mathcal{F}_{t_\theta}} \left[\sup_{t_\theta \leq t \leq t_i} |Y_t^{\hat{\alpha}} - Y_t^\theta|^2 \right] \leq C\varepsilon^2.$$

In particular, this implies that

$$|Y_{t_\theta}^{\hat{\alpha}} - Y_{t_\theta}^\theta| \leq C\varepsilon, \quad \text{a.s. on } E_\theta. \quad (4.21)$$

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

$$\tilde{Y}_t^\theta = Y_{t_\theta}^\theta + \int_t^{t_\theta} f(s, x_\theta, \tilde{Y}_s^\theta, \tilde{Z}_s^\theta, a_0) ds - \int_t^{t_\theta} \tilde{Z}_s^\theta dB_s, \quad t \in [t_{i-1}, t_\theta]. \quad (4.22)$$

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

$$\sup_{t_{i-1} \leq t \leq t_\theta} |\tilde{Y}_t^\theta - Y_{t_\theta}^\theta| \leq \int_{t_{i-1}}^{t_\theta} |f(s, x_\theta, \tilde{Y}_s^\theta, 0, a_0)| ds \leq C(t_i - t_{i-1}) \leq C\varepsilon^2.$$

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 (4.22), by (4.21) and standard BSDE estimate we have

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

Then

$$|Y_\tau^{\hat{\alpha}} - Y_{t_\theta}^\theta| \leq |Y_\tau^{\hat{\alpha}} - \tilde{Y}_\tau^\theta| + |\tilde{Y}_\tau^\theta - Y_{t_\theta}^\theta| \leq C\varepsilon, \quad \text{a.s. on } \{\tau \leq t_\theta\} \cap E_\theta. \quad (4.23)$$

This, together with (4.16) and (4.18), implies that, for a.e. $\omega \in \{\tau \leq t_\theta\} \cap E_\theta$,

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

Note that $\{E_\theta\}$ form a partition of E and $\mathbb{P}(E) = 1$, then by (4.17) we have

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

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

$$\mathbb{E}[|Y_\tau^{\hat{\alpha}} - \phi|^2] \leq C\varepsilon^2 + C\mathbb{P}(|Y_\tau^{\hat{\alpha}} - \phi| > C\varepsilon) \leq C\varepsilon^2. \quad (4.24)$$

Finally, note that $Y_t^{\hat{\alpha}} = Y_t^{\tau, Y_\tau^{\hat{\alpha}}; \alpha}$, $0 \leq t \leq \tau$. Then, by (4.24) and standard BSDE estimates we have

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

This clearly implies (4.15) and hence completes the proof. \blacksquare

5 Set valued HJB equations

We now derive the set valued HJB equation for \mathbb{V} from the DPP (4.11). Introduce the Hamiltonian: for $(t, x, y) \in \mathbb{G}_\mathbb{V}$, $z \in \mathbb{R}^{m \times d}$, $\gamma \in (\mathbb{R}^{d \times d})^m$, $a \in A$, $\zeta \in (\mathbb{T}_\mathbb{V}(t, x, y))^d$,

$$\begin{aligned} \mathcal{K}_\mathbb{V}(t, x, y, a, \zeta) &:= \text{tr}\left(\zeta^\top \partial_x \mathbf{n}_\mathbb{V}(t, x, y) \sigma(t, x, a) + \frac{1}{2} \zeta^\top \partial_y \mathbf{n}_\mathbb{V}(t, x, y) \zeta\right) \mathbf{n}_\mathbb{V}(t, x, y); \\ h_\mathbb{V}^0(t, x, y, z, \gamma, a, \zeta) &:= zb(t, x, a) + \frac{1}{2} \text{tr}(\sigma^\top \gamma \sigma(t, x, a)) - \mathcal{K}_\mathbb{V}(t, x, y, a, \zeta); \\ h_\mathbb{V}(t, x, y, z, \gamma, a, \zeta) &:= h_\mathbb{V}^0(t, x, y, z, \gamma, a, \zeta) + f(t, x, y, z \sigma(t, x, a) + \zeta, a); \\ H_\mathbb{V}(t, x, y, z, \gamma) &:= \sup_{a \in A, \zeta \in (\mathbb{T}_\mathbb{V}(t, x, y))^d} \mathbf{n}_\mathbb{V}(t, x, y) \cdot h_\mathbb{V}(t, x, y, z, \gamma, a, \zeta); \end{aligned} \quad (5.1)$$

where $\text{tr}(\sigma^\top \gamma \sigma) \in \mathbb{R}^m$ with i -th component $\text{tr}(\sigma^\top \gamma^i \sigma)$. Then our set valued HJB equation takes the form:

$$\mathcal{L}\mathbb{V}(t, x, y) = 0, \quad \forall (t, x, y) \in \mathbb{G}_\mathbb{V}, \quad \text{where} \quad (5.2)$$

$$\mathcal{L}\mathbb{V}(t, x, y) := \partial_t \mathbb{V}(t, x, y) \cdot \mathbf{n}_\mathbb{V}(t, x, y) + H_\mathbb{V}(t, x, y, \partial_x \mathbb{V}(t, x, y), \partial_{xx} \mathbb{V}(t, x, y)).$$

Equivalently, by (2.4), (2.23), (2.25), (2.26), we may rewrite the above equation:

$$\begin{aligned} \nabla_t \mathbf{r}_\mathbb{V} + \inf_{a \in A, \zeta \in (\mathbb{T}_\mathbb{V}(t, x, y))^d} \left[\nabla_x \mathbf{r}_\mathbb{V} \cdot b + \frac{1}{2} \text{tr}(\sigma^\top \nabla_{xx} \mathbf{r}_\mathbb{V} \sigma + 2\zeta^\top \nabla_{xy} \mathbf{r}_\mathbb{V} \sigma + \zeta^\top \nabla_{yy} \mathbf{r}_\mathbb{V} \zeta) \right. \\ \left. - \nabla_y \mathbf{r}_\mathbb{V} \cdot f(t, x, y, -\nabla_y \mathbf{r}_\mathbb{V} (\nabla_x \mathbf{r}_\mathbb{V})^\top \sigma + \zeta, a) \right] = 0, \quad (t, x, y) \in \mathbb{G}_\mathbb{V}. \end{aligned} \quad (5.3)$$

Remark 5.1 (i) For the scalar case as in Remark 4.1, by (4.4) we have

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

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 (5.3) reduces to the standard HJB equation for \underline{v} :

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

Similarly, in the neighborhood of $y = \bar{v}(t, x)$, we have $\mathbf{r}_\mathbb{V}(t, x, y) = y - \bar{v}(t, x)$ and $\mathbf{n}(t, x, \bar{v}(t, x)) = 1$. Then (5.3) reduces to the standard HJB equation for \bar{v} :

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

(ii) Although $\mathbf{r}_\mathbb{V}$ is scalar, we emphasize that (5.3) 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 (5.3) 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 2.9 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, so in general $H_\mathbb{V}$ could blow up and then the set valued PDE is not well defined.

(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 2.7 (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 $\lambda_1 \leq \dots \leq \lambda_{m-1}$ 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 $\lambda_1 \geq 0$. So, unfortunately, our classical solution \mathbb{V} has to be convex. Thus one should explore appropriate notions of weak solutions in the nonconvex case, which we will leave to future research.

(iv) When f has quadratic growth, this convexity is not required, as we will see in Example 5.3 below. We shall remark though such quadratic growth violates Assumption 4.4, which is assumed for technical reasons and can be weakened.

Example 5.3 Consider a deterministic example where f, g and hence \mathbb{V} are independent of x . Set $m = 2$, $A = \{a \in \mathbb{R}^2 : |a| \leq 1\}$, $g = 0$, and $f = (f_1, f_2)$ are specified at below:

$$f_1(a, y) = a_1, \quad f_2(a, y) = \frac{2y_1 y_2 a_1}{1 + y_1^2} + (1 + y_1^2) a_2.$$

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

$$\begin{aligned} \mathbb{V}_b(t) &= \left\{ \mathcal{Y}(t, \theta) : \forall \theta \in [0, 2\pi] \right\}, \quad \text{where} \\ \mathcal{Y}_1(t, \theta) &:= (T - t) \cos \theta, \quad \mathcal{Y}_2(t, \theta) := (T - t) [1 + (T - t)^2 \cos^2 \theta] \sin \theta. \end{aligned} \quad (5.4)$$

We postpone its proof to Appendix.

We now turn to the wellposedness of (5.2). 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,

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

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 (5.2) if it satisfies (5.2) 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 6.2 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 $|\mathbf{r}_{\mathbb{V}}(0, x_0, Y_0^\alpha)| \leq \varepsilon$, 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 is independent of $\varepsilon, \delta, \alpha$, such that

$$\mathbb{P} \left(\sup_{0 \leq t \leq T_0} |\mathbf{r}_{\mathbb{V}}(t, X_t^\alpha, Y_t^\alpha)| \geq \delta \right) \leq C_{T_0} \sqrt{\frac{\varepsilon}{\delta}}. \quad (5.6)$$

In particular, if $Y_0^\alpha \in \mathbb{V}_b(0, x_0)$, then $Y_t^\alpha \in \mathbb{V}_b(t, X_t^\alpha)$, $0 \leq t \leq T$, 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 (5.2) 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 (5.2) only at a fixed $(0, x_0, y_0) \in \mathbb{G}_\mathbb{V}$, and for notational simplicity, in this proof we omit the superscripts $^{0, x_0, y_0}$ and the subscript \mathbb{V} in $\mathbf{r}, \mathbf{n}, \pi$. We proceed in two steps.

Step 1. We first show that $\mathcal{L}\mathbb{V}(0, x_0, y_0) \leq 0$. 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 $\xi, \zeta_i : [0, T] \times \Omega \times \mathbb{R}^m \rightarrow \mathbb{R}^m, i = 1, \dots, d$, 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 $\zeta = (\zeta_1, \dots, \zeta_d)$ and consider the SDE:

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

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

Now for any $\delta > 0$ small, consider the BSDE (4.10) with terminal condition $(\delta, \Upsilon_\delta^{a, \xi, \zeta})$:

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

Since $\Upsilon_\delta^{a, \xi, \zeta} \in \mathbb{V}_b(\delta, X_\delta^a)$, by the DPP (4.11) we see that $Y_0^{\delta, a, \xi, \zeta} \in \mathbb{V}(0, x_0)$. Denote

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

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

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

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

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

In particular,

$$\Delta Y_0^\delta = \mathbb{E} \left[\int_0^\delta \tilde{\Gamma}_s [\partial_t \mathbb{V} + h_{\mathbb{V}}(\cdot, \partial_x \mathbb{V}, \partial_{xx} \mathbb{V}, a, \zeta) + \xi] (s, X_s^a, \Upsilon_s^{a, \xi, \zeta}) ds \right]. \quad (5.9)$$

Given our conditions, it is clear that $|\Delta Y_0^\delta| \leq C\delta$, which implies that $\lim_{\delta \rightarrow 0} Y_0^{\delta, \alpha, \xi, \zeta} = y_0$. Since $Y_0^{\delta, \alpha, \xi, \zeta} \in \mathbb{V}_0(0, x_0)$ and $y_0 \in \mathbb{V}_b(0, x_0)$, then

$$\overline{\lim}_{\delta \rightarrow 0} \frac{1}{\delta} [\mathbf{n}(0, x_0, y_0) \cdot \Delta Y_0^\delta] \leq 0. \quad (5.10)$$

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

$$\begin{aligned} 0 &\geq \overline{\lim}_{\delta \rightarrow 0} \frac{1}{\delta} \mathbb{E} \left[\mathbf{n}(0, x_0, y_0) \cdot \int_0^\delta \tilde{\Gamma}_s [\partial_t \mathbb{V} + h_{\mathbb{V}}(\cdot, \partial_x \mathbb{V}, \partial_{xx} \mathbb{V}, a, \zeta) + \xi] (s, X_s^a, \Upsilon_s^{a, \xi, \zeta}) ds \right] \\ &= \overline{\lim}_{\delta \rightarrow 0} \frac{1}{\delta} \left[\mathbf{n}(0, x_0, y_0) \cdot \int_0^\delta [\partial_t \mathbb{V} + h_{\mathbb{V}}(\cdot, \partial_x \mathbb{V}, \partial_{xx} \mathbb{V}, a, \zeta) + \xi] (0, x_0, y_0) ds \right] \\ &= \mathbf{n}(0, x_0, y_0) \cdot [\partial_t \mathbb{V} + h_{\mathbb{V}}(\cdot, \partial_x \mathbb{V}, \partial_{xx} \mathbb{V}, a, \zeta)] (0, x_0, y_0), \end{aligned}$$

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

Step 2. We next show that $\mathcal{L}\mathbb{V}(0, x_0, y_0) \geq 0$. 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 $\mathbf{r} \in C^2(O_{\varepsilon_0}^{T_0}(\mathbb{G}_{\mathbb{V}}))$, where $O_{\varepsilon_0}^{T_0}(\mathbb{G}_{\mathbb{V}}) := \{(t, x, y) \in [0, T_0] \times \mathbb{R}^d \times \mathbb{R}^m : |\mathbf{r}(t, x, y)| \leq \varepsilon_0\}$. 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

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

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

$$\tau := \tau_{\varepsilon, \alpha} := \inf\{t > 0 : |\mathbf{r}(t, X_t^\alpha, Y_t^\alpha)| \geq \varepsilon^2\} \wedge T_0. \quad (5.11)$$

By Lemma 5.5 we have

$$\mathbb{P}(\tau < T_0) \leq \mathbb{P}\left(\sup_{0 \leq t \leq T_0} |\mathbf{r}(t, X_t^\alpha, Y_t^\alpha)| \geq \varepsilon^2\right) \leq C \sqrt{\frac{\varepsilon^4}{\varepsilon^2}} = C\varepsilon. \quad (5.12)$$

Step 2.1. Introduce two random fields:

$$\zeta_t(y) := Z_t^\alpha - \partial_x \mathbb{V} \sigma(t, X_t^\alpha, y, \alpha_t), \quad \xi_t(y) := -[\partial_t \mathbb{V} + h_{\mathbb{V}}(\cdot, \partial_x \mathbb{V}, \partial_{xx} \mathbb{V}, \alpha_t, \zeta_t)](t, X_t^\alpha, y). \quad (5.13)$$

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

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

We remark 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 $\mathcal{L}\mathbb{V}(t, X_t^\alpha, Y_t^\alpha) \geq -\mathbf{n}(t, X_t^\alpha, Y_t^\alpha) \cdot \xi_t(Y_t^\alpha) = 0$, which is the desired inequality. In this and the next substep, 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,

$$\begin{aligned} d\mathbf{r}(t, X_t^\alpha, Y_t^\alpha) &= \left[\mathbf{r}(t, X_t^\alpha, Y_t^\alpha) \tilde{b}_t - \mathbf{n} \cdot \xi(t, X_t^\alpha, \pi_t^\alpha) \right] dt \\ &\quad + \left[\mathbf{r}(t, X_t^\alpha, Y_t^\alpha) \tilde{\sigma}_t - \mathbf{n}^\top \zeta(t, X_t^\alpha, \pi_t^\alpha) \right] dB_t, \quad 0 \leq t \leq \tau, \end{aligned} \tag{5.14}$$

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

$$|\tilde{\sigma}_t| \leq C, \quad |\tilde{b}_t| \leq C[1 + |Z_t^\alpha|]. \tag{5.15}$$

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

$$\mathbf{r}(0, x_0, Y_0^\alpha) - \tilde{\Gamma}_\tau \mathbf{r}(\tau, X_\tau^\alpha, Y_\tau^\alpha) = \int_0^\tau \tilde{\Gamma}_s \mathbf{n} \cdot \xi(s, X_s^\alpha, \pi_s^\alpha) ds + \int_0^\tau \tilde{\Gamma}_s \mathbf{n}^\top \zeta(s, X_s^\alpha, \pi_s^\alpha) dB_s. \tag{5.16}$$

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 (c.f. [20, Section 7.2])

$$\mathbb{E} \left[\exp \left(c_0 \int_0^\tau |Z_s^\alpha|^2 ds \right) \right] \leq C < \infty. \tag{5.17}$$

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

$$\mathbb{E} \left[\sup_{0 \leq t \leq \tau} [|\tilde{\Gamma}_t|^p + |\tilde{\Gamma}_t|^{-p}] \right] \leq C_p. \tag{5.18}$$

Applying the standard Itô formula on $|\tilde{\Gamma}_t \mathbf{r}(t, X_t^\alpha, Y_t^\alpha)|^2$, by (5.14) we have

$$\begin{aligned} &\mathbb{E} \left[\int_0^\tau |\tilde{\Gamma}_s \mathbf{n}^\top \zeta(s, X_s^\alpha, \pi_s^\alpha)|^2 ds \right] \\ &= \mathbb{E} \left[|\tilde{\Gamma}_\tau \mathbf{r}(\tau, X_\tau^\alpha, Y_\tau^\alpha)|^2 - |\mathbf{r}(0, x_0, Y_0^\alpha)|^2 + 2 \int_0^\tau \tilde{\Gamma}_s^2 \mathbf{r}(s, X_s^\alpha, Y_s^\alpha) \mathbf{n} \cdot \xi(s, X_s^\alpha, \pi_s^\alpha) ds \right] \\ &\leq C\varepsilon^4 + C\varepsilon^2 \mathbb{E} \left[\sup_{0 \leq t \leq \tau} |\tilde{\Gamma}_t|^2 \int_0^\tau [1 + |Z_s^\alpha|^2] ds \right] \leq C\varepsilon^2. \end{aligned} \tag{5.19}$$

Step 2.2. Denote

$$\begin{aligned}\hat{\zeta}_s &:= \zeta_s(\pi_s^\alpha) - \mathbf{n}\mathbf{n}^\top \zeta_s(\pi_s^\alpha) \in (\mathbb{T}\mathbb{V}(s, X_s^\alpha, \pi_s^\alpha))^d; \\ \hat{\xi}_s &:= -[\partial_t \mathbb{V} + h_{\mathbb{V}}(\cdot, \partial_x \mathbb{V}, \partial_{xx} \mathbb{V}, \alpha_s, \hat{\zeta}_s)](s, X_s^\alpha, \pi_s^\alpha).\end{aligned}\tag{5.20}$$

By (5.2), (5.3), and by Step 1, we have

$$0 \leq -\mathcal{L}\mathbb{V}(s, X_s^\alpha, \pi_s^\alpha) \leq \mathbf{n}(s, X_s^\alpha, \pi_s^\alpha) \cdot \hat{\xi}_s.\tag{5.21}$$

Then, by taking expectation on both sides of (5.16) we have

$$\begin{aligned}-\mathbb{E}\left[\int_0^\tau \tilde{\Gamma}_s \mathcal{L}\mathbb{V}(s, X_s^\alpha, \pi_s^\alpha) ds\right] &\leq \mathbb{E}\left[\int_0^\tau \tilde{\Gamma}_s \mathbf{n}(s, X_s^\alpha, \pi_s^\alpha) \cdot \hat{\xi}_s ds\right] \\ &\leq \mathbb{E}\left[\int_0^\tau \tilde{\Gamma}_s \mathbf{n}(s, X_s^\alpha, \pi_s^\alpha) \cdot \xi_s ds\right] + \mathbb{E}\left[\int_0^\tau \tilde{\Gamma}_s |\xi_s - \hat{\xi}_s| ds\right] \\ &= \mathbb{E}\left[\mathbf{r}(0, x_0, Y_0^\alpha) - \tilde{\Gamma}_\tau \mathbf{r}(\tau, X_\tau^\alpha, Y_\tau^\alpha)\right] + \mathbb{E}\left[\int_0^\tau \tilde{\Gamma}_s |\xi_s - \hat{\xi}_s| ds\right].\end{aligned}$$

Recall Remark 2.7 (iii), we see that $\text{tr}((\mathbf{n}\mathbf{n}^\top \zeta)^\top \partial_y \mathbf{n}(\mathbf{n}\mathbf{n}^\top \zeta)) = 0$. Then, by (5.2),

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

and thus, by (5.19),

$$\begin{aligned}-\mathbb{E}\left[\int_0^\tau \tilde{\Gamma}_s \mathcal{L}\mathbb{V}(s, X_s^\alpha, \pi_s^\alpha) ds\right] &\leq \mathbb{E}\left[\int_0^\tau \tilde{\Gamma}_s \mathbf{n}(s, X_s^\alpha, \pi_s^\alpha) \cdot \hat{\xi}_s ds\right] \\ &\leq \mathbb{E}\left[\mathbf{r}(0, x_0, Y_0^\alpha) - \tilde{\Gamma}_\tau \mathbf{r}(\tau, X_\tau^\alpha, Y_\tau^\alpha)\right] + C \mathbb{E}\left[\int_0^\tau \tilde{\Gamma}_s |\mathbf{n}^\top \zeta(s, X_s^\alpha, \pi_s^\alpha)| ds\right] \\ &\leq C\varepsilon^2 + C \left(\mathbb{E}\left[\int_0^\tau |\tilde{\Gamma}_s \mathbf{n}^\top \zeta(s, X_s^\alpha, \pi_s^\alpha)|^2 ds\right]\right)^{1/2} \leq C\varepsilon.\end{aligned}\tag{5.22}$$

Step 2.3. Fix another small constant $\delta > 0$. Since $\mathcal{L}\mathbb{V} \leq 0$, by (5.12) we have

$$\begin{aligned}-\mathcal{L}\mathbb{V}(0, x_0, y_0) &= \mathbb{E}\left[-\frac{1}{\delta} \int_0^{\tau \wedge \delta} \mathcal{L}\mathbb{V}(0, x_0, y_0) ds - \frac{(\delta - \tau)^+}{\delta} \mathcal{L}\mathbb{V}(0, x_0, y_0)\right] \\ &= -\frac{1}{\delta} \mathbb{E}\left[\int_0^{\tau \wedge \delta} \tilde{\Gamma}_s \mathcal{L}\mathbb{V}(s, X_s^\alpha, \pi_s^\alpha) ds\right] - \mathcal{L}\mathbb{V}(0, x_0, y_0) \mathbb{E}\left[\left(1 - \frac{\tau}{\delta}\right) \mathbf{1}_{\{\tau < \delta\}}\right] \\ &\quad + \frac{1}{\delta} \mathbb{E}\left[\int_0^{\tau \wedge \delta} [\tilde{\Gamma}_s \mathcal{L}\mathbb{V}(s, X_s^\alpha, \pi_s^\alpha) - \mathcal{L}\mathbb{V}(0, x_0, y_0)] ds\right] \\ &\leq \frac{C\varepsilon}{\delta} + \mathbb{E}\left[\sup_{0 \leq t \leq \tau \wedge \delta} |\tilde{\Gamma}_t \mathcal{L}\mathbb{V}(t, X_t^\alpha, \pi_t^\alpha) - \mathcal{L}\mathbb{V}(0, x_0, y_0)|\right].\end{aligned}$$

Since $\mathbb{V} \in C^{1,2}([0, T_0] \times \mathbb{R}^d; \mathcal{D}_2^m)$, $H_{\mathbb{V}}$ is bounded and uniform continuous. Then, for some modulus of continuity function ρ we have

$$-\mathcal{L}\mathbb{V}(0, x_0, y_0) \leq \frac{C\varepsilon}{\delta} + C \mathbb{E}\left[\sup_{0 \leq t \leq \tau \wedge \delta} [|\tilde{\Gamma}_t - 1| + \rho(\delta + |X_t^\alpha - x_0| + |\pi_t^\alpha - y_0|)]\right].\tag{5.23}$$

Recall (3.9), (5.15), and (5.18), we have

$$\begin{aligned}
\mathbb{E}\left[\sup_{0 \leq t \leq \tau \wedge \delta} |\tilde{\Gamma}_t - 1|\right] &\leq \mathbb{E}\left[\sup_{0 \leq t \leq \tau \wedge \delta} [(\tilde{\Gamma}_t + 1)(\left|\int_0^t \tilde{\sigma}_s \cdot dB_s\right| + \int_0^t [|\tilde{b}_s| + \frac{1}{2}|\tilde{\sigma}_s|^2]ds)]\right] \\
&\leq C\sqrt{\delta} + C\left(\mathbb{E}\left[\left(\int_0^{\tau \wedge \delta} |Z_t^\alpha|^2 dt\right)^2\right]\right)^{\frac{1}{2}} \leq C\sqrt{\delta} + C\sqrt{\delta}\left(\mathbb{E}\left[\int_0^T |Z_t^\alpha|^2 dt\right]\right)^{\frac{1}{2}} \leq C\sqrt{\delta}; \\
\mathbb{E}\left[\sup_{0 \leq t \leq \tau \wedge \delta} |X_t^\alpha - x_0|\right] &\leq C\sqrt{\delta}; \\
\mathbb{E}\left[\sup_{0 \leq t \leq \tau \wedge \delta} |\pi_t^\alpha - y_0|\right] &\leq C\varepsilon^2 + \mathbb{E}\left[\sup_{0 \leq t \leq \tau \wedge \delta} |Y_t^\alpha - y_0|\right] \\
&\leq C[\varepsilon^2 + \sqrt{\delta}] + C\mathbb{E}\left[\left(\int_0^{\tau \wedge \delta} |Z_t^\alpha|^2 dt\right)^{\frac{1}{2}}\right].
\end{aligned}$$

Then (5.23) implies

$$\begin{aligned}
-\mathcal{L}\mathbb{V}(0, x_0, y_0) &\leq C\left[\frac{\varepsilon}{\delta} + \sqrt{\delta}\right] + C\rho(\delta + \delta^{\frac{1}{3}}) + \frac{C}{\delta^{\frac{1}{3}}}\mathbb{E}\left[\sup_{0 \leq t \leq \tau \wedge \delta} [|X_t^\alpha - x_0| + |\pi_t^\alpha - y_0|]\right] \\
&\leq C\left[\frac{\varepsilon}{\delta} + \sqrt{\delta} + \rho(\delta + \delta^{\frac{1}{3}}) + \frac{\varepsilon^2 + \sqrt{\delta}}{\delta^{\frac{1}{3}}}\right] + \frac{C}{\delta^{\frac{1}{3}}}\mathbb{E}\left[\left(\int_0^{\tau \wedge \delta} |Z_t^\alpha|^2 dt\right)^{\frac{1}{2}}\right]. \tag{5.24}
\end{aligned}$$

Step 2.4. Recall (5.5) and (5.21). Then by (5.22) we have

$$C\varepsilon \geq \mathbb{E}\left[\int_0^{\tau \wedge \delta} \tilde{\Gamma}_s \mathbf{n}(s, X_s^\alpha, \pi_s^\alpha) \cdot \hat{\xi}_s ds\right] \geq \mathbb{E}\left[\int_0^{\tau \wedge \delta} \tilde{\Gamma}_s \left[\frac{cT_0}{2}|\hat{\zeta}_s|^2 - C|\hat{\zeta}_s| - C\right] ds\right].$$

This, together with (5.18), implies that

$$\mathbb{E}\left[\int_0^{\tau \wedge \delta} \tilde{\Gamma}_s |\hat{\zeta}_s|^2 ds\right] \leq C[\varepsilon + \delta]. \tag{5.25}$$

By (5.13) and (5.20) we have

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

Then, by (5.19), (5.25), and (5.18), we have

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

Plug this into (5.24), we get

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

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

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 (5.2) with terminal condition $\mathbb{V}(T, x) = \{g(x)\}$. Then $\mathbb{U} = \mathbb{V}$, and consequently (5.2) 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:*

$$a^* = I_1^{\mathbb{U}}(t, x, y) \in A, \quad \zeta^* = I_2^{\mathbb{U}}(t, x, y) \in (\mathbb{T}_{\mathbb{U}}(t, x, y))^d. \quad (6.1)$$

Moreover, recall Remark 2.1 and denote

$$\begin{aligned} \tilde{I}_3^{\mathbb{U}}(t, x, y) &:= - \left[\partial_t \mathbb{U} + h_{\mathbb{U}}(\cdot, \partial_x \mathbb{U}, \partial_{xx} \mathbb{U}, I_1^{\mathbb{U}}, I_2^{\mathbb{U}}) \right](t, x, y); \\ I_3^{\mathbb{U}}(t, x, y) &:= \tilde{I}_3^{\mathbb{U}}(t, x, y) - [\mathbf{n}_{\mathbb{U}} \cdot \tilde{I}_3^{\mathbb{U}}] \mathbf{n}_{\mathbb{U}}(t, x, y); \end{aligned} \quad (6.2)$$

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

$$\begin{aligned} X_t^* &= x_0 + \int_0^t b(\cdot, I_1^{\mathbb{U}})(s, X_s^*, \Upsilon_s^*) ds + \int_0^t \sigma(\cdot, I_1^{\mathbb{U}})(s, X_s^*, \Upsilon_s^*) dB_s; \\ \Upsilon_t^* &= y_0 + \int_0^t \left[\partial_t \mathbb{U} + h_{\mathbb{U}}^0(\cdot, \partial_x \mathbb{U}, \partial_{xx} \mathbb{U}, I_1^{\mathbb{U}}, I_2^{\mathbb{U}}) + I_3^{\mathbb{U}} \right](s, X_s^*, \Upsilon_s^*) ds \\ &\quad + \int_0^t [\partial_x \mathbb{V} \sigma(\cdot, I_1^{\mathbb{U}}) + I_2^{\mathbb{U}}](s, X_s^*, \Upsilon_s^*) dB_s. \end{aligned} \quad (6.3)$$

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

Proof 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 $\phi_\delta \in \mathbb{U}_b(T_\delta, X_{T_\delta}^\alpha)$, a.s. and

$$\mathbb{E}[|Y_{T_\delta}^\alpha - \phi_\delta|^2] \leq C \mathbb{E}[|Y_{T_\delta}^\alpha - g(X_{T_\delta}^\alpha)|^2 + |g(X_{T_\delta}^\alpha) - g(X_T^\alpha)|^2 + |\phi_\delta - g(X_{T_\delta}^\alpha)|^2] \rightarrow 0, \quad (6.4)$$

as $\delta \rightarrow 0$. Recall (4.10) 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

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

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

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

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 (5.2), we have

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

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

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

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

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

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

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

That is, $Y_0^{\alpha,\delta} \in \mathbb{U}(0, x_0)$. Send $\delta \rightarrow 0$, 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 $I_1^\mathbb{U}, I_2^\mathbb{U}$, we have

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

Since \mathbb{U} satisfies the PDE 5.2, we have

$$\mathbf{n}_{\mathbb{U}} \cdot \tilde{I}_3^{\mathbb{U}}(s, X_s^*, \Upsilon_s^*) = 0;$$

$$\left[\partial_t \mathbb{U} + h_{\mathbb{U}}^0(\cdot, \partial_x \mathbb{U}, \partial_{xx} \mathbb{U}, I_1^{\mathbb{U}}, I_2^{\mathbb{U}}) + I_3^{\mathbb{U}} \right](s, X_s^*, \Upsilon_s^*) = -f(\cdot, \partial_x \mathbb{V} \sigma(\cdot, I_1^{\mathbb{U}}) + I_2^{\mathbb{U}}, I_1^{\mathbb{U}})(s, X_s^*, \Upsilon_s^*).$$

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). Fix $(0, x_0, y_0) \in \mathbb{G}_{\mathbb{U}}$ and $\delta > 0$. Since $\mathbb{U} \in C^{1,2}([0, T_\delta], \mathbb{R}^d; \mathcal{D}_2^m)$, we assume $\mathbf{r}_{\mathbb{U}}$ is smooth in $O_{\varepsilon_0}^{T_0}(\mathbb{G}_{\mathbb{U}})$ 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 (5.2), 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

$$|\bar{\zeta}^0| \leq C_\delta \quad \text{and} \quad 0 \leq -\mathbf{n}_{\mathbb{U}} \cdot [\partial_t \mathbb{U} + h_{\mathbb{U}}(\cdot, \partial_x \mathbb{U}, \partial_{xx} \mathbb{U}, \bar{a}_0, \bar{\zeta}^0)](0, x_0, y_0) < \delta. \quad (6.6)$$

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

$$X_t^1 = x_0 + \int_0^t b(s, X_s^1, \alpha_s^1) ds + \int_0^t \sigma(s, X_s^1, \alpha_s^1) dB_s, \quad 0 \leq t \leq T_\delta.$$

Introduce the random fields $(\xi^1, \zeta^1) : [0, T_\delta] \times \Omega \times \mathbb{R}^m \rightarrow (\mathbb{R}^m, \mathbb{R}^{m \times d})$: recalling Remark 2.1,

$$\zeta_t^1(y) := \bar{\zeta}^0 - \mathbf{n}_{\mathbb{U}} \mathbf{n}_{\mathbb{U}}^\top \bar{\zeta}^0(t, X_t^1, y), \quad \xi_t^1(y) := \bar{\xi}_t^1(y) - [\mathbf{n}_{\mathbb{U}} \cdot \bar{\xi}_t^1] \mathbf{n}_{\mathbb{U}}(t, X_t^1, y), \quad (6.7)$$

$$\text{where} \quad \bar{\xi}_t^1(y) := -[\partial_t \mathbb{U} + h_{\mathbb{U}}(\cdot, \partial_x \mathbb{U}, \partial_{xx} \mathbb{U}, \alpha^1, \zeta^1)](t, X_t^1, y).$$

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

$$\begin{aligned} \Upsilon_t^1 &= y_0 + \int_0^t \left[\partial_t \mathbb{U} + h_{\mathbb{U}}^0(\cdot, \partial_x \mathbb{U}, \partial_{xx} \mathbb{U}, \alpha_s^1, \zeta_s^1) + \xi^1 \right](s, X_s^1, \Upsilon_s^1) ds \\ &\quad + \int_0^t \left[\partial_x \mathbb{U}(s, X_s^1, \Upsilon_s^1) \sigma(s, X_s^1, \alpha_s^1) + \zeta_s^1(\Upsilon_s^1) \right] dB_s. \end{aligned} \quad (6.8)$$

By the Itô formula Theorem 3.1, $\Upsilon_t^1 \in \mathbb{U}_b(t, X_t^1)$, $0 \leq t \leq T_\delta$. Note that (6.6) implies $\mathbf{n}_{\mathbb{U}}(0, X_0^1, \Upsilon_0^1) \cdot \bar{\xi}_0^1(\Upsilon_0^1) \leq \varepsilon$, and by our construction, α^1, ζ^1 and hence $\bar{\xi}^1$ are continuous in t . We then set

$$\tau_1 := \inf \left\{ t > \tau_0 : \mathbf{n}_{\mathbb{U}}(t, X_t^1, \Upsilon_t^1) \cdot \bar{\xi}_t^1(\Upsilon_t^1) \geq 2\delta \right\} \wedge T_\delta.$$

Next, on $\{\tau_1 < T_\delta\}$, by measurable selection theorem, there exist \mathcal{F}_{τ_1} -measurable random variables $\bar{\alpha}_{\tau_1}^1 \in A$ and $\bar{\zeta}_{\tau_1}^1 \in (\mathbb{T}_{\mathbb{V}}(\tau_1, X_{\tau_1}^1, \Upsilon_{\tau_1}^1))^d$ such that

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

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

$$X_t^2 = X_{\tau_1}^1 + \int_{\tau_1}^t b(s, X_s^2, \alpha_s^2) ds + \int_{\tau_1}^t \sigma(s, X_s^2, \alpha_s^2) dB_s, \quad \tau_1 \leq t \leq T_\delta.$$

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

$$\begin{aligned} \zeta_t^2(y) &:= \bar{\zeta}_{\tau_1}^1 - \mathbf{n}_U \mathbf{n}_U^\top \bar{\zeta}_{\tau_1}^1(t, X_t^2, y), \quad \xi_t^2(y) := \tilde{\xi}_t^2(y) - [\mathbf{n}_U \cdot \tilde{\xi}_t^2] \mathbf{n}_U(t, X_t^2, y), \\ \text{where } \tilde{\xi}_t^2(y) &:= -[\partial_t \mathbb{U} + h_U(\cdot, \partial_x \mathbb{U}, \partial_{xx} \mathbb{U}, \alpha^2, \zeta^2)](t, X_t^2, y), \end{aligned} \quad (6.9)$$

and consider the SDE:

$$\begin{aligned} \Upsilon_t^2 &= \Upsilon_{\tau_1}^1 + \int_{\tau_1}^t \left[\partial_t \mathbb{U} + h_U^0(\cdot, \partial_x \mathbb{U}, \partial_{xx} \mathbb{U}, \alpha_s^2, \zeta_s^2) + \xi_s^2 \right](s, X_s^2, \Upsilon_s^2) ds \\ &\quad + \int_{\tau_1}^t \left[\partial_x \mathbb{U}(s, X_s^2, \Upsilon_s^2) \sigma(s, X_s^2, \alpha_s^2) + \zeta_s^2(\Upsilon_s^2) \right] dB_s. \end{aligned} \quad (6.10)$$

Then $\Upsilon_t^2 \in \mathbb{U}_b(t, X_t^2)$, $\tau_1 \leq t \leq T_\delta$, and we may set

$$\tau_2 := \inf \{ t > \tau_1 : \mathbf{n}_U(t, X_t^2, \Upsilon_t^2) \cdot \tilde{\xi}_t^2(\Upsilon_t^2) \geq 2\delta \} \wedge T_\delta.$$

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

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

Then, for any n ,

$$\delta \mathbb{P}(E_\delta) \leq \mathbb{E} \left[\left| \mathbf{n}_U(\tau_{n+1}, X_{\tau_{n+1}}, \Upsilon_{\tau_{n+1}}) \cdot \tilde{\xi}_{\tau_{n+1}}^n(\Upsilon_{\tau_{n+1}}) - \mathbf{n}_U(\tau_n, X_{\tau_n}, \Upsilon_{\tau_n}) \cdot \tilde{\xi}_{\tau_n}^n(\Upsilon_{\tau_n}) \right| \right].$$

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

We now define

$$(\alpha_t, \zeta_t, \tilde{\xi}_t, \xi_t, X_t, \Upsilon_t) := (\alpha_t^n, \zeta_t^n, X_t^n, \Upsilon_t^n, \xi_t^n), \quad t \in [\tau_n, \tau_{n+1}), n = 0, 1, \dots$$

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

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

Then, $|\eta| \leq 2\delta$, and by (5.1) and (6.8) we have

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

Equivalently, we may rewrite it backwardly:

$$\Upsilon_t = \Upsilon_{T_\delta} + \int_t^{T_\delta} \left[f(s, X_s, \Upsilon_s, Z_s, \alpha_s) + \eta_s \right] ds - \int_t^{T_\delta} Z_s dB_s, \quad 0 \leq t \leq T_\delta. \quad (6.11)$$

Compare this with (4.10), by standard BSDE estimates we have

$$|y_0 - Y_0^{T_\delta, \Upsilon_{T_\delta}; 0, x_0, \alpha}|^2 = |\Upsilon_0 - Y_0^{T_\delta, \Upsilon_{T_\delta}; 0, x_0, \alpha}|^2 \leq C \mathbb{E} \left[\int_0^{T_\delta} |\eta_s|^2 ds \right] \leq C \delta^2. \quad (6.12)$$

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

$$\mathbb{E}[|Y_{T_\delta}^\alpha - \Upsilon_{T_\delta}|^2] \leq C \mathbb{E} \left[|Y_{T_\delta}^\alpha - g(X_{T_\delta}^\alpha)|^2 + |g(X_{T_\delta}^\alpha) - g(X_{T_\delta}^\alpha)|^2 + |\Upsilon_{T_\delta} - g(X_{T_\delta}^\alpha)|^2 \right] \leq \rho(\delta),$$

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

$$|Y_0^{T_\delta, \Upsilon_{T_\delta}; 0, x_0, \alpha} - Y_0^\alpha|^2 = |Y_0^{T_\delta, \Upsilon_{T_\delta}; 0, x_0, \alpha} - Y_0^{T_\delta, Y_{T_\delta}^\alpha; 0, x_0, \alpha}|^2 \leq \mathbb{E}[|Y_{T_\delta}^\alpha - \Upsilon_{T_\delta}|^2] \leq \rho(\delta).$$

Combine this with (6.12), we have

$$|y_0 - Y_0^\alpha| \leq C \delta + \sqrt{\rho(\delta)}.$$

Since $Y_0^\alpha \in \mathbb{V}(0, x_0)$ 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 6.2 Set $d = 1$, $m = 2$, $A = \{a \in \mathbb{R}^2 : |a| \leq 1\}$, 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

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

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 (5.3).

7 Further discussions

7.1 An application: the moving scalarization

Recall Remark 4.2, in particular (4.5) and (4.7) 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

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

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 : by using the notations heuristically,

$$\sup_{\alpha_{[t,T]}^*} \varphi(Y_t^{t, X_t^*, \alpha_{[t,T]}^*}) = \sup_{y \in \mathbb{V}(t, X_t^*)} \varphi(y). \quad (7.2)$$

It was proposed in [14] 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

$$\sup_{\alpha_{[t,T]}^*} \Phi(t, X_{[0,t]}^*, Y_t^{t, X_t^*, \alpha_{[t,T]}^*}) = \sup_{y \in \mathbb{V}(t, X_t^*)} \Phi(t, X_{[0,t]}^*, y). \quad (7.3)$$

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 [9]. That is, we want to find Λ such that $\Lambda(0, x_0) = \lambda_0$ and $\alpha_{[t,T]}^*$ is optimal for the problem:

$$\sup_{\alpha_{[t,T]}^*} \Lambda(t, X_{[0,t]}^*) \cdot Y_t^{t, X_t^*, \alpha_{[t,T]}^*} = \sup_{y \in \mathbb{V}(t, X_t^*)} \Lambda(t, X_{[0,t]}^*) \cdot y. \quad (7.4)$$

Our set valued HJB equation provides a solution to this interesting problem, provided that (5.2) 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 $\lambda : \mathbb{R}^d \rightarrow \mathbb{R}^m$ 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 $\mathbb{F}^{X^*} = \mathbb{F}^B$. Thus Υ^* is \mathbb{F}^{X^*} -progressively measurable and hence there exists Λ such that

$$\Lambda(t, X_{[0,t]}^*) = \mathbf{n}_\mathbb{V}(t, X_t^*, \Upsilon_t^*). \quad (7.5)$$

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 (7.5) 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 (7.4).

We remark that the mapping $\Lambda : [0, T] \times C([0, T]; \mathbb{R}^d) \rightarrow \mathbb{R}^m$, 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 $X_{[0,t]}^{0,x,\lambda,*}$ 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:

$$X_s^{t, X_{[0,t]}^{0,x,\lambda,*}, \Lambda^{0,\lambda}(t, \cdot)} = X_s^{0,x,\lambda,*}, \quad \Lambda^{t, \Lambda^{0,\lambda}(t, \cdot)}(s, \cdot) = \Lambda^{0,\lambda}(s, \cdot), \quad t \leq s \leq T. \quad (7.6)$$

Remark 7.1 When $\mathbb{V}(t, X_t^*)$ is nonconvex, as in Example 5.3, the Λ in (7.5) can be viewed as a local asymptotic moving scalarization in the following sense:

$$\begin{aligned} \Lambda(t, X_{[0,t]}^*) \cdot \Upsilon_t^* &\geq \Lambda(t, X_{[0,t]}^*) \cdot y - o(|y - \Upsilon_t^*|), \quad \forall y \in \mathbb{V}(t, X_t^*); \quad \text{or equivalently,} \\ \Lambda(t, X_{[0,t]}^*) \cdot Y_t^{t, X_t^*, \alpha_{[t,T]}^*} &\geq \Lambda(t, X_{[0,t]}^*) \cdot Y_t^{t, X_t^*, \alpha_{[t,T]}} - o(|Y_t^{t, X_t^*, \alpha_{[t,T]}} - Y_t^{t, X_t^*, \alpha_{[t,T]}^*}|), \quad \forall \alpha_{[t,T]}. \end{aligned} \quad (7.7)$$

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

$$\begin{aligned} Y_t^0 &= y_\lambda(x_0) - \int_0^t f(s, X_s^0, Y_s^0, Z_s^0, \alpha_s^0) ds + \int_0^t Z_s^0 dB_s, \\ \tau_0 &:= \inf \{ t \geq 0 : (t, X_t^0, Y_t^0) \in \mathbb{G}_\mathbb{V} \}. \end{aligned} \quad (7.8)$$

It is clear that $\tau_0 \leq T$ and $(\tau_0, X_{\tau_0}^0, Y_{\tau_0}^0) \in \mathbb{G}_\mathbb{V}$. Applying Theorem 6.1 (ii) on $(\tau_0, X_{\tau_0}^0, Y_{\tau_0}^0)$ (assuming all the conditions are satisfied) and following the measurable selection theorem

we may construct α^* on $[\tau_0, T]$ with initial condition $(\tau, X_\tau^0, Y_\tau^0)$. Then one can easily see that $\alpha^0 \otimes_{\tau_0} \alpha^*$ 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 (7.3). In particular, when φ has certain structure, for example the linear quadratic structure for the mean variance problem in Remark 4.2, we may expect Φ to have the same structure as well, which will add the difficulty for constructing a desired Φ .

7.2 The case with nondegenerate terminal

As pointed out in Remark 4.3 (ii), given a general $\mathbb{G} : \mathbb{R}^d \rightarrow \mathcal{D}_0^m$, we may define \mathbb{V} by (4.9). This is equivalent to

$$\mathbb{V}(t, x) = \text{cl}\{Y_t^{T, \phi; t, x, \alpha} : \alpha \in \mathcal{A}_t, \phi \in \mathbb{L}^2(\mathcal{F}_T^t) \text{ s.t. } \phi \in \mathbb{G}(X_T^{t, x, \alpha}), \text{ a.s.}\}. \quad (7.9)$$

Then we have

Theorem 7.3 *Let Assumption 4.4 (i), (ii) hold, and \mathbb{G} is bounded and continuous. Assume the \mathbb{V} defined by (4.9) or (7.9) 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 (5.2) 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.

7.3 Comparison with Soner-Touzi [19]

In the contexts of stochastic target problem, [19] 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 (4.8) (or the more general one (4.9)) can be rewritten equivalently as:

$$\widehat{\mathbb{V}}(t) := \left\{ (x, y) \in \mathbb{R}^{d+m} : \exists (\alpha, Z) \text{ such that } Y_T^{t, x, y, \alpha, Z} = g(X_T^{t, x, \alpha}), \text{ a.s.} \right\}. \quad (7.10)$$

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:

$$\begin{aligned} \widehat{\mathbb{V}}(t) &= \{(x, y) : x \in \mathbb{R}^d, y \in \mathbb{V}(t, x)\}, & \widehat{\mathbb{V}}_b(t) &= \{(x, y) : x \in \mathbb{R}^d, y \in \mathbb{V}_b(t, x)\}; \\ \text{and } \mathbb{V}(t, x) &= \{y : (x, y) \in \widehat{\mathbb{V}}(t)\}, & \mathbb{V}_b(t) &= \{y : (x, y) \in \widehat{\mathbb{V}}_b(t)\}. \end{aligned} \quad (7.11)$$

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

$$\begin{aligned}\mathbf{n}_{\hat{\mathbb{V}}}(t, x, y) &:= \mathbf{n}_{\hat{\mathbb{V}}(t)}(x, y) \in \mathbb{R}^{d+m}, \quad (t, x, y) \in \mathbb{G}_{\hat{\mathbb{V}}}; \\ \mathbf{r}_{\hat{\mathbb{V}}}(t, x, y) &:= \mathbf{r}_{\hat{\mathbb{V}}(t)}(x, y) \in \mathbb{R}, \quad (t, x, y) \in O_\varepsilon(\mathbb{G}_{\hat{\mathbb{V}}}).\end{aligned}\tag{7.12}$$

The work [19] characterized the square of the distance function³ $\eta(t, x, y) := \frac{1}{2}|\mathbf{r}_{\hat{\mathbb{V}}}(t, x, y)|^2$ by the following PDE: denoting $\hat{y} := (x, y)$ and noting the time change in [19],

$$\begin{aligned}\nabla_{\hat{y}}\nabla_t\eta(t, \hat{y}) + \nabla_{\hat{y}}\left[F(t, \hat{y}, \nabla_{\hat{y}}\eta(t, \hat{y}), \nabla_{\hat{y}\hat{y}}\eta(t, \hat{y}))\right] &= 0, \quad (t, \hat{y}) \in \mathbb{G}_{\hat{\mathbb{V}}}, \quad \text{where} \\ F(t, \hat{y}, \nabla_{\hat{y}}\eta, \nabla_{\hat{y}\hat{y}}\eta) &:= \inf_{(a, z) \in \mathcal{N}(t, \hat{y}, \nabla_{\hat{y}}\eta)} \left[b(t, x, a) \cdot \nabla_x\eta(t, \hat{y}) - f(t, \hat{y}, z, a) \cdot \nabla_y\eta(t, \hat{y}) \right. \\ &\quad \left. + \frac{1}{2}\text{tr} \left(\sigma^\top(t, x, a)\nabla_{xx}\eta(t, \hat{y})\sigma(t, x, a) + 2z^\top\nabla_{xy}\eta(t, \hat{y})\sigma(t, x, z) + z^\top\nabla_{yy}\eta(t, \hat{y})z \right) \right], \\ \mathcal{N}(t, \hat{y}, \nabla_{\hat{y}}\eta) &:= \{(a, z) : [\sigma^\top(t, x, a), z^\top]^\top [\sigma^\top(t, x, a), z^\top] \nabla_{\hat{y}}\eta(t, \hat{y}) = 0\}.\end{aligned}\tag{7.13}$$

We first note that $\hat{\mathbb{V}}$ is a function of t only, and thus it does not lead to the set valued Itô formula as $\mathbb{V}(t, x)$ does. While this may seem to be technically easier, the set valued Itô formula has 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). We note that (7.13) holds only on $\mathbb{G}_{\hat{\mathbb{V}}}$, so it is also not a standard PDE.

The major difference is that, as we see in Example 7.4 below,

$$\mathbf{r}_{\hat{\mathbb{V}}}(t, x, y) \neq \mathbf{r}_{\mathbb{V}}(t, x, y).\tag{7.14}$$

In general, recalling (7.11) we have

$$\begin{aligned}|\mathbf{r}_{\hat{\mathbb{V}}}(t, x, y)|^2 &= \inf_{(\tilde{x}, \tilde{y}) \in \hat{\mathbb{V}}_b(t)} [|x - \tilde{x}|^2 + |y - \tilde{y}|^2] = \inf_{\tilde{x} \in \mathbb{R}^d} \inf_{\tilde{y} \in \mathbb{V}_b(t, \tilde{x})} [|x - \tilde{x}|^2 + |y - \tilde{y}|^2] \\ &= \inf_{\tilde{x} \in \mathbb{R}^d} [|x - \tilde{x}|^2 + |\mathbf{r}_{\mathbb{V}}(t, \tilde{x}, y)|^2] \leq |\mathbf{r}_{\mathbb{V}}(t, x, y)|^2\end{aligned}\tag{7.15}$$

Example 7.4 *Set $d = m = 1$ and consider time invariant set values:*

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

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

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

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

³The main reason to consider the squared function is that, in the degenerate case $\mathbf{r}_{\hat{\mathbb{V}}}$ is typically not smooth while $|\mathbf{r}_{\hat{\mathbb{V}}}|^2$ is. In the nondegenerate case as in this paper, actually one may study $\mathbf{r}_{\hat{\mathbb{V}}}$ directly.

Consequently, although both (5.3) and (7.13) characterize the same set $\mathbb{G}_\mathbb{V} = \mathbb{G}_{\widehat{\mathbb{V}}}$, the two equations are fundamentally 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 2.8 (ii), but with $d = m = 1$.

Example 7.5 *Set $d = m = 1$, $b = 0$, $\sigma = 1$, $f = 0$, and the terminal $\mathbb{G}(x) = [-\psi(x), \psi(x)]$, where $\psi : \mathbb{R} \rightarrow (0, \infty)$ is smooth. Then, similar to Example 2.8 (ii), we see that*

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

where u is the unique classical solution of the heat equation

$$\partial_t u + \frac{1}{2} \partial_{xx} u = 0, \quad u(T, x) = \psi(x). \quad (7.16)$$

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

$$\nabla_t \widehat{\mathbf{r}} + \frac{1}{2} \left[\nabla_{xx} \widehat{\mathbf{r}} - 2 \nabla_{xy} \widehat{\mathbf{r}} \frac{\nabla_x \widehat{\mathbf{r}}}{\nabla_y \widehat{\mathbf{r}}} + \nabla_{yy} \widehat{\mathbf{r}} \left| \frac{\nabla_x \widehat{\mathbf{r}}}{\nabla_y \widehat{\mathbf{r}}} \right|^2 \right] = 0, \quad \text{on } \mathbb{G}_{\widehat{\mathbb{V}}}. \quad (7.17)$$

In this scalar case, by Remark 5.1 (i) we see that the set valued HJB equation (5.3) 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 (7.16). So (5.3) is indeed a natural extension of the HJB equation to the multivariate case. The equation (7.17), however, is quite different from (7.16). So in this sense, it is more natural to study (5.3) than to study (7.13).

Another advantage of (5.3) is that, as we saw in Section 7.1, 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 [19, Theorem 2.1] showed that $\widehat{\mathbb{V}}$ is the unique classical solution of (7.13) under the existence of optimal controls, in the same spirit of our Theorem 6.1 (ii). We instead proved the existence and uniqueness under weak conditions in Theorems 5.6 and 6.1 (i).

7.4 Comparison with Ararat-Ma-Wu [1]

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

$$\mathbb{Y}_t = \mathbb{E} \left[\mathbb{G}(B_T) + \int_t^T \mathbb{F}(s, B_s, \mathbb{Y}_s) ds \mid \mathcal{F}_t \right]. \quad (7.18)$$

Here, denoting by \mathcal{D}_{cc}^m the space of convex compact sets $\mathbb{D} \in \mathcal{D}_0^m$, the terminal $\mathbb{G} : \mathbb{R}^d \rightarrow \mathcal{D}_{cc}^m$, and the driver $\mathbb{F} : [0, T] \times \mathbb{R}^d \times \mathcal{D}_{cc}^m \rightarrow \mathcal{D}_{cc}^m$ (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 $\mathbb{Z}_t dB_t$ seems to be a quite remote goal.

Formally, the set valued BSDE (7.18) is associated with our set valued HJB equation (5.2) in the case $x_0 = 0$, $b = 0$, $\sigma = 0$, $f = f(t, x, y, a)$. Then $X_t = B_t$, and we may naturally define

$$\mathbb{Y}_t := \mathbb{V}(t, B_t), \quad \mathbb{F}(t, x, \mathbb{D}) := \{f(t, x, y, a) : y \in \mathbb{D}, a \in A\}. \quad (7.19)$$

In the linear case: $f = f(t, x, a)$ and thus $\mathbb{F}(t, x) = \{f(t, x, a) : a \in A\}$, the random set valued process $\mathbb{Y}_t := \mathbb{V}(t, B_t)$ indeed satisfies (7.18) in the sense of [1].

However, when f depends on y , the \mathbb{Y}, \mathbb{F} in (7.19) do not satisfy (7.18). That is, (7.18) is not the stochastic counterpart of (5.2). 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 (7.18), however, one allows to consider different selectors for the \mathbb{Y} in the left side and that in the right side of the equation.

8 Appendix

Proof of Proposition 2.4. Again we denote 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 ,

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

and thus

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

Therefore,

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

This implies that

$$\begin{aligned} &\overline{\lim}_{x_1 \downarrow x_0} \frac{1}{x_1 - x_0} [L_\Upsilon(x_0, x_1) - L_\theta(x_0, x_1)] \\ &\leq \overline{\lim}_{x_1 \downarrow x_0} \frac{1}{x_1 - x_0} \left[\int_{x_0}^{x_1} \sqrt{1 + |\nabla_x \mathbf{r}(x, \Upsilon(x))|^2} dx - \int_{x_0}^{x_1} \sqrt{1 + |\nabla_x \mathbf{r}(x, \theta(x))|^2} dx \right] \\ &\leq \overline{\lim}_{x_1 \downarrow x_0} \frac{1}{x_1 - x_0} \left[\int_{x_0}^{x_1} \left| \sqrt{1 + |\nabla_x \mathbf{r}(x, \Upsilon(x))|^2} - \sqrt{1 + |\nabla_x \mathbf{r}(x_0, y_0)|^2} \right| dx \right. \\ &\quad \left. + \int_{x_0}^{x_1} \left| \sqrt{1 + |\nabla_x \mathbf{r}(x, \theta(x))|^2} - \sqrt{1 + |\nabla_x \mathbf{r}(x_0, y_0)|^2} \right| dx \right] = 0. \quad \blacksquare \end{aligned}$$

Proof of Lemma 2.6. Recall (2.4) and consider the natural extension $\hat{\mathbf{n}} = \nabla_y \mathbf{r}$. By (2.23), (2.20), and (2.19) we have, for $i, j = 1, \dots, d$, and $(t, x, y) \in \mathbb{G}_\mathbb{V}$,

$$\begin{aligned} \partial_{x_i x_j} \mathbb{V}(t, x, y) &= -\partial_{x_i} (\nabla_{x_j} \mathbf{r}\mathbf{n})(t, x, y) = -\left[\partial_{x_i} (\nabla_{x_j} \mathbf{r})\mathbf{n} + \nabla_{x_j} \mathbf{r} \partial_{x_i} \mathbf{n} \right](t, x, y) \\ &= -\left[\nabla_{x_i x_j} \mathbf{r} + \nabla_{x_i} r \nabla_{x_j y} \mathbf{r} \cdot \mathbf{n} \right] \mathbf{n}(t, x, y) - \nabla_{x_j} \mathbf{r} \partial_{x_i} \mathbf{n}(t, x, y). \quad (8.1) \end{aligned}$$

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

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

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 (8.1) we obtain (2.25) immediately.

Moreover, again considering the extension $\hat{\mathbf{n}}^i = \nabla_{y_i} \mathbf{r}$, by (2.19) and (2.8) we have

$$\partial_x n^i = \nabla_{x y_i} \mathbf{r} - \nabla_x \mathbf{r} (\nabla_{y_i y} r \cdot \mathbf{n}), \quad \partial_y n^i = \nabla_{y_i y} r - (\nabla_{y_i y} r \cdot \mathbf{n})\mathbf{n}.$$

Similarly, by (2.4) we have $\nabla_{y_i y} r \cdot \mathbf{n} = 0$, which implies (2.26) immediately. \blacksquare

Proof of Example 5.3. We first prove (5.4). Denote

$$\tilde{f}_1(a, y) := a_1, \quad \tilde{f}_2(a, y) := a_2, \quad \tilde{Y}_t^\alpha = \int_t^T \tilde{f}(\alpha_s, \tilde{Y}_s^\alpha) ds.$$

Then one can easily check that

$$\begin{aligned} \tilde{V}(t) &:= \{\tilde{Y}_t^\alpha : \alpha \in \mathcal{A}_t\} = \{\tilde{y} \in \mathbb{R}^2 : |\tilde{y}| \leq T - t\}, \\ \tilde{\mathbb{V}}_b(t) &:= \{\tilde{y} \in \mathbb{R}^2 : |\tilde{y}| = T - t\} = \{(T - t)(\cos \theta, \sin \theta)^\top : \theta \in [0, 2\pi)\}. \end{aligned} \quad (8.2)$$

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

$$Y_t^\alpha := \psi(\tilde{Y}_t^\alpha), \quad \text{where} \quad \psi_1(\tilde{y}) := \tilde{y}_1, \quad \psi_2(\tilde{y}) := [1 + |\tilde{y}_1|^2]\tilde{y}_2.$$

Then we have

$$\begin{aligned} dY_t^{\alpha,1} &= d\tilde{Y}_t^{\alpha,1} = -\alpha_t^1 dt = -f_1(\alpha_t, Y_t^\alpha) dt; \\ dY_t^{\alpha,2} &= 2\tilde{Y}_t^{\alpha,1}\tilde{Y}_t^{\alpha,2}d\tilde{Y}_t^{\alpha,1} + [1 + |\tilde{Y}_t^{\alpha,1}|^2]d\tilde{Y}_t^{\alpha,2} = -\left[2\tilde{Y}_t^{\alpha,1}\tilde{Y}_t^{\alpha,2}\alpha_t^1 + [1 + |\tilde{Y}_t^{\alpha,1}|^2]\alpha_t^2\right] dt \\ &= -\left[\frac{2Y_t^{\alpha,1}Y_t^{\alpha,2}}{1 + |Y_t^{\alpha,1}|^2}\alpha_t^1 + [1 + |Y_t^{\alpha,1}|^2]\alpha_t^2\right] dt = -f_2(\alpha_t, Y_t^\alpha) dt. \end{aligned}$$

That is, Y^α satisfies (4.1) for given f . Therefore, $\mathbb{V}(t) = \{\psi(\tilde{y}) : \tilde{y} \in \tilde{\mathbb{V}}(t)\}$. 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

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

Plug (8.2) into it, we obtain (5.4) immediately.

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

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

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

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

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. \blacksquare

Proof of Lemma 5.5. Denote

$$\tau_\delta := \inf\{t \geq 0 : |\mathbf{r}(t, X_t^\alpha, Y_t^\alpha)| \geq \delta\} \wedge T,$$

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

$$\kappa_t = \mathbf{1}_{\{\tau_\delta \leq T_0\}} - \int_t^{\tau_\delta} \beta_s dB_s, \quad 0 \leq t \leq \tau, \quad (8.3)$$

where $\kappa \in \mathbb{R}$, $\beta \in \mathbb{R}^{1 \times d}$. It is clear that $|\kappa| \leq 1$ and $\int_0^\cdot \beta_s dB_s$ is an BMO martingale. Thus,

$$\mathbb{E} \left[\exp \left(c_0 \int_0^T |\beta_t|^2 dt \right) \right] \leq C_0 < \infty, \quad \text{for some } c_0, C_0 > 0. \quad (8.4)$$

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

$$\kappa_t^n := \kappa_0 + \int_0^t \beta_s^n dB_s.$$

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

$$\mathbb{E} \left[\sup_{0 \leq t \leq \tau_\delta} |\kappa_t^n|^p \right] \leq C_p < \infty; \quad \text{and} \quad c_p^n := \left(\mathbb{E} \left[\sup_{0 \leq t \leq T_0} |\kappa_t^n - \kappa_t|^p \right] \right)^{\frac{1}{p}} \rightarrow 0, \quad \text{as } n \rightarrow \infty. \quad (8.5)$$

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

$$\begin{aligned} \nu_{ij}^n(t, y) &:= - \sum_{k=1}^m \int_0^1 \nabla_{z_{kj}} f^i(t, X_t^\alpha, Y_t^\alpha, Z_t^\alpha + \theta y \beta_t^n, \alpha_t) y_k d\theta; \\ \rho_i^n(t, y) &:= - \int_0^1 \left[\nabla_y f^i \cdot y + \text{tr} \left((\partial_z f^i)^\top \nu^n(t, y) \right) \right] \\ &\quad (t, X_t^\alpha, Y_t^\alpha + \theta \kappa_t^n y, Z_t^\alpha + y \beta_t^n + \theta \kappa_t^n \nu^n(t, y), \alpha_t) d\theta; \end{aligned}$$

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

$$\begin{aligned} \nu^n(t, y) \beta_t^n &:= f(t, X_t^\alpha, Y_t^\alpha, Z_t^\alpha, \alpha_t) - f(t, X_t^\alpha, Y_t^\alpha, Z_t^\alpha + y \beta_t^n, \alpha_t); \\ \kappa_t^n \rho^n(t, y) &:= f(t, X_t^\alpha, Y_t^\alpha, Z_t^\alpha + y \beta_t^n, \alpha_t) - f(t, X_t^\alpha, Y_t^\alpha + \kappa_t^n y, Z_t^\alpha + y \beta_t^n + \kappa_t^n \nu(t, y), \alpha_t). \end{aligned}$$

Moreover, by Assumption 4.4 (ii) we have

$$\begin{aligned} |\nu^n(t, y)| &\leq C|y|, \quad |\rho^n(t, y)| \leq C|y|; \\ |\nu^n(t, y) - \nu^n(t, \tilde{y})| &\leq C_n [1 + |y|] |y - \tilde{y}|, \\ |\rho^n(t, y) - \rho^n(t, \tilde{y})| &\leq C_n [1 + |y| + |\kappa_t^n| + |\kappa_t^n| |y|^2] |y - \tilde{y}|. \end{aligned} \quad (8.6)$$

Next, consider the following SDE:

$$\eta_t^n := \lambda \mathbf{n}_\nabla(0, x_0, \pi(0, x_0, Y_0^\alpha)) + \int_0^t \rho^n(s, \eta_s^n) ds + \int_0^t \nu^n(s, \eta_s^n) dB_s, \quad (8.7)$$

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 (8.6), the above SDE is wellposed, and for any $p \geq 1$,

$$\mathbb{E} \left[\sup_{0 \leq t \leq T_0} |\eta_t^n|^p \right] \leq C_p |\lambda|^p, \quad (8.8)$$

where C_p does not depend on n .

Denote

$$\tilde{Y}_t^n := Y_t^\alpha + \kappa_t^n \eta_t^n, \quad \tilde{Z}_t^n := Z_t + \kappa_t^n \nu_t^n + \eta_t^n \beta_t^n.$$

Then by the standard Itô formula we have

$$\tilde{Y}_t^n = Y_{\tau_\delta}^\alpha + \kappa_{\tau_\delta}^n \eta_{\tau_\delta}^n + \int_t^{\tau_\delta} f(s, X_s^\alpha, \tilde{Y}_s^n, \tilde{Z}_s^n, \alpha_s) ds - \int_t^{\tau_\delta} \tilde{Z}_s^n dB_s. \quad (8.9)$$

Moreover, introduce the BSDE

$$\hat{Y}_t^n = Y_{\tau_\delta}^\alpha + \kappa_{\tau_\delta}^n \eta_{\tau_\delta}^n \mathbf{1}_{\{|\kappa_{\tau_\delta}^n| \leq 2, |\eta_{\tau_\delta}^n| < \frac{\delta}{2}\}} + \int_t^{\tau_\delta} f(s, X_s^\alpha, \hat{Y}_s^n, \hat{Z}_s^n, \alpha_s) ds - \int_t^{\tau_\delta} \hat{Z}_s^n dB_s.$$

By (8.5), (8.8), and noting that $|\kappa_{\tau_\delta}| \leq 1$, it follows from standard BSDE estimates that

$$\begin{aligned} |\hat{Y}_0^n - \tilde{Y}_0^n|^2 &\leq C \mathbb{E} \left[|\kappa_{\tau_\delta}^n \eta_{\tau_\delta}^n|^2 \left[\mathbf{1}_{\{|\kappa_{\tau_\delta}^n - \kappa_{\tau_\delta}| > 1\}} + \mathbf{1}_{\{|\eta_{\tau_\delta}^n| \geq \frac{\delta}{2}\}} \right] \right] \\ &\leq C \mathbb{E} \left[|\kappa_{\tau_\delta}^n \eta_{\tau_\delta}^n|^2 |\kappa_{\tau_\delta}^n - \kappa_{\tau_\delta}| + \frac{1}{\delta^2} |\kappa_{\tau_\delta}^n|^2 |\eta_{\tau_\delta}^n|^4 \right] \leq C c_2^n + C \frac{|\lambda|^4}{\delta^2}. \end{aligned}$$

Thus

$$|\hat{Y}_0^n - \tilde{Y}_0^n| \leq C \left[\sqrt{c_2^n} + \frac{|\lambda|^2}{\delta} \right]. \quad (8.10)$$

Note that

$$\hat{Y}_{\tau_\delta}^n = Y_{\tau_\delta}^\alpha + \kappa_{\tau_\delta}^n \eta_{\tau_\delta}^n \mathbf{1}_{\{|\kappa_{\tau_\delta}^n| \leq 2, |\eta_{\tau_\delta}^n| < \frac{\delta}{2}\}}.$$

On $\{|\kappa_{\tau_\delta}^n| \leq 2, |\eta_{\tau_\delta}^n| < \frac{\delta}{2}\}^c$, we have $\hat{Y}_{\tau_\delta}^n = Y_{\tau_\delta}^\alpha \in \mathbb{V}(\tau_\delta, X_{\tau_\delta}^\alpha)$. On $\{|\kappa_{\tau_\delta}^n| \leq 2, |\eta_{\tau_\delta}^n| < \frac{\delta}{2}\}$, noting again that $Y_{\tau_\delta}^\alpha \in \mathbb{V}(\tau_\delta, X_{\tau_\delta}^\alpha)$, we have $\mathbf{r}_{\mathbb{V}}(\tau_\delta, X_{\tau_\delta}^\alpha, Y_{\tau_\delta}^\alpha) = -\delta$ and $|\kappa_{\tau_\delta}^n \eta_{\tau_\delta}^n| \leq \delta$, then $\hat{Y}_{\tau_\delta}^n = Y_{\tau_\delta}^\alpha + \kappa_{\tau_\delta}^n \eta_{\tau_\delta}^n \in \mathbb{V}(\tau_\delta, X_{\tau_\delta}^\alpha)$. So in both cases $\hat{Y}_{\tau_\delta}^n \in \mathbb{V}(\tau_\delta, X_{\tau_\delta}^\alpha)$. Then by DPP (4.11) we have $\hat{Y}_0^n \in \mathbb{V}(0, x_0)$. Thus, by (8.10),

$$\mathbf{r}_{\mathbb{V}}(0, x_0, \tilde{Y}_0^n) \leq |\tilde{Y}_0^n - \hat{Y}_0^n| \leq C \left[\sqrt{c_2^n} + \frac{|\lambda|^2}{\delta} \right].$$

On the other hand, note that $\tilde{Y}_0^n = Y_0^\alpha + \kappa_0 \lambda \mathbf{r}_{\mathbb{V}}(0, x_0, \pi(0, x_0, Y_0^\alpha))$, for $\kappa_0 \lambda$ small we have

$$\mathbf{r}_{\mathbb{V}}(0, x_0, \tilde{Y}_0^n) = \mathbf{r}_{\mathbb{V}}(0, x_0, Y_0^\alpha) + \kappa_0 \lambda.$$

Thus

$$\kappa_0 \lambda = \mathbf{r}_{\mathbb{V}}(0, x_0, \tilde{Y}_0^n) - \mathbf{r}_{\mathbb{V}}(0, x_0, Y_0^\alpha) \leq C \left[\sqrt{c_2^n} + \frac{|\lambda|^2}{\delta} \right] + \varepsilon.$$

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

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

Finally, if $Y_0^\alpha \in \mathbb{V}_b(0, x_0)$, then $\varepsilon = 0$. We see that $\mathbb{P}(\tau_\delta \leq T_0) = 0$ for all $\delta > 0$ and all $T_0 < T$. This implies immediately that $Y_t^\alpha \in \mathbb{V}_b(t, X_t^\alpha)$, $0 \leq t < T$, a.s. Moreover, note that $Y_T^\alpha = g(X_T^\alpha)$ and $\mathbb{V}(T, x) = \{g(x)\}$, we have $Y_T^\alpha \in \mathbb{V}_b(T, X_T^\alpha)$ as well. \blacksquare

Proof of Example 6.2. As usual we drop the subscript \mathbb{V} in \mathbf{r} and \mathbf{n} .

(i) We first show that $\mathbb{V} \in C_0^{1,2}([0, T] \times \mathbb{R}; \mathcal{D}_2^2)$. Fix $\delta > 0$ and denote $T_\delta := T - \delta$. By Example 2.8 we have, with $u(t, x) = T - t \geq \delta$ 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 (2.28), for $|y - w(t, x)| = T - t$, we have

$$\begin{aligned} \mathbf{n} &= \frac{y - w}{T - t}; \quad \partial_t \mathbb{V} = [\nabla_t w \cdot \mathbf{n} - 1] \mathbf{n}; \quad \partial_x \mathbb{V} = [\nabla_x w \cdot \mathbf{n}] \mathbf{n}; \\ \partial_x \mathbf{n} &= \frac{1}{T - t} [-\nabla_x w + [\mathbf{n} \cdot \nabla_x w] \mathbf{n}], \quad \partial_y \mathbf{n} = \frac{1}{T - t} [I_{2 \times 2} - \mathbf{n} \mathbf{n}^\top]; \\ \partial_{xx} \mathbb{V} &= -\frac{1}{T - t} \left[[|\nabla_x w|^2 - \nabla_{xx} w \cdot \mathbf{n} + |\nabla_x w \cdot \mathbf{n}|^2] \mathbf{n} + [\nabla_x w \cdot \mathbf{n}] [\nabla_x w - (\nabla_x w \cdot \mathbf{n}) \mathbf{n}] \right]. \end{aligned}$$

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 (5.1) we have: at $(t, x, y) \in \mathbb{G}_\mathbb{V}$,

$$\begin{aligned} h_\mathbb{V}^0(t, x, y, \partial_x \mathbb{V}, \partial_{xx} \mathbb{V}, a, \zeta) &= \frac{1}{2} \partial_{xx} \mathbb{V} - \left[\zeta \cdot \partial_x \mathbf{n} + \frac{1}{2} \zeta^\top \partial_y \mathbf{n} \zeta \right] \mathbf{n} \\ &= -\frac{1}{2(T - t)} \left[[|\nabla_x w|^2 - \nabla_{xx} w \cdot \mathbf{n} + |\nabla_x w \cdot \mathbf{n}|^2] \mathbf{n} + [\nabla_x w \cdot \mathbf{n}] [\nabla_x w - (\nabla_x w \cdot \mathbf{n}) \mathbf{n}] \right] \\ &\quad - \frac{1}{T - t} \left[-\zeta \cdot \nabla_x w + \frac{1}{2} |\zeta|^2 \right] \mathbf{n}. \end{aligned}$$

Thus

$$\begin{aligned} \mathbf{n} \cdot h_\mathbb{V}(t, x, y, \partial_x \mathbb{V}, \partial_{xx} \mathbb{V}, a, \zeta) &= \mathbf{n} \cdot h_\mathbb{V}^0(t, x, y, \partial_x \mathbb{V}, \partial_{xx} \mathbb{V}, a, \zeta) + \mathbf{n} \cdot [f^0(t, x) + a] \\ &= -\frac{1}{2(T - t)} \left[[|\nabla_x w|^2 - \nabla_{xx} w \cdot \mathbf{n} + |\nabla_x w \cdot \mathbf{n}|^2 - 2\zeta \cdot \nabla_x w + |\zeta|^2] \right] + \mathbf{n} \cdot [f^0 + a] \\ &= -\frac{1}{2(T - t)} \left[|\zeta - \nabla_x w|^2 - \nabla_{xx} w \cdot \mathbf{n} + |\nabla_x w \cdot \mathbf{n}|^2 \right] + \mathbf{n} \cdot [f^0 + a]. \end{aligned}$$

Recall $|a| \leq 1$, 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 (6.2), this implies further that

$$\begin{aligned}
\tilde{I}_3^\mathbb{V} &:= -\left[\partial_t \mathbb{V} + h_{\mathbb{V}}^0(\cdot, \partial_x \mathbb{V}, \partial_{xx} \mathbb{V}, I_1^\mathbb{V}, I_2^\mathbb{V}) + f^0 + I_1^\mathbb{V}\right] \\
&= -[\nabla_t w \cdot \mathbf{n} - 1] \mathbf{n} + \frac{1}{T-t} \left[-\zeta \cdot \nabla_x w + \frac{1}{2} |\zeta|^2\right] \mathbf{n} - [f^0 + \mathbf{n}] \\
&\quad + \frac{1}{2(T-t)} \left[|\nabla_x w|^2 - \nabla_{xx} w \cdot \mathbf{n} + |\nabla_x w \cdot \mathbf{n}|^2\right] \mathbf{n} + [\nabla_x w \cdot \mathbf{n}] [\nabla_x w - (\nabla_x w \cdot \mathbf{n}) \mathbf{n}]; \\
I_3^\mathbb{V} &:= -f^0 + [\mathbf{n} \cdot f^0] \mathbf{n} + \frac{1}{2(T-t)} [\nabla_x w \cdot \mathbf{n}] [\nabla_x w - (\nabla_x w \cdot \mathbf{n}) \mathbf{n}].
\end{aligned}$$

Plug $I_1^\mathbb{V}, I_2^\mathbb{V}, I_3^\mathbb{V}$ into (6.3), clearly the resulted SDE is wellposed. \blacksquare

Proof of Example 7.5. We now compute the equation (7.13) in this case. First,

$$\begin{aligned}
\mathcal{N}(t, x, y, \nabla_x \eta, \nabla_y \eta) &= \{(a, z) : [1, z]^\top [\nabla_x \eta + z \nabla_y \eta] = 0\} \\
&= \{(a, z) : z = -\frac{\nabla_x \eta}{\nabla_y \eta}\} = \left\{(a, -\frac{\nabla_x \hat{\mathbf{r}}(t, x, y)}{\nabla_y \hat{\mathbf{r}}(t, x, y)}\right\}.
\end{aligned}$$

Then, recalling $\eta = \frac{1}{2} |\hat{\mathbf{r}}|^2$,

$$\begin{aligned}
F &= \frac{1}{2} \left[\nabla_{xx} \eta - 2 \nabla_{xy} \eta \frac{\nabla_x \hat{\mathbf{r}}}{\nabla_y \hat{\mathbf{r}}} + \nabla_{yy} \eta \left| \frac{\nabla_x \hat{\mathbf{r}}}{\nabla_y \hat{\mathbf{r}}} \right|^2 \right] (t, x, y) \\
&= \frac{1}{2} \left[\hat{\mathbf{r}} \nabla_{xx} \hat{\mathbf{r}} + |\nabla_x \hat{\mathbf{r}}|^2 - 2 [\hat{\mathbf{r}} \nabla_{xy} \hat{\mathbf{r}} + \nabla_x \hat{\mathbf{r}} \nabla_y \hat{\mathbf{r}}] \frac{\nabla_x \hat{\mathbf{r}}}{\nabla_y \hat{\mathbf{r}}} + [\hat{\mathbf{r}} \nabla_{yy} \hat{\mathbf{r}} + |\nabla_y \hat{\mathbf{r}}|^2] \left| \frac{\nabla_x \hat{\mathbf{r}}}{\nabla_y \hat{\mathbf{r}}} \right|^2 \right] (t, x, y) \\
&= \frac{1}{2} \hat{\mathbf{r}} \left[\nabla_{xx} \hat{\mathbf{r}} - 2 \nabla_{xy} \hat{\mathbf{r}} \frac{\nabla_x \hat{\mathbf{r}}}{\nabla_y \hat{\mathbf{r}}} + \nabla_{yy} \hat{\mathbf{r}} \left| \frac{\nabla_x \hat{\mathbf{r}}}{\nabla_y \hat{\mathbf{r}}} \right|^2 \right] (t, x, y).
\end{aligned}$$

Note further that $\hat{\mathbf{r}} = 0$ on $\mathbb{G}_{\mathbb{V}}$. Then, for $(t, x, y) \in \mathbb{G}_{\mathbb{V}}$, we have

$$\begin{aligned}
\nabla_x F &= \frac{1}{2} \nabla_x \hat{\mathbf{r}} \left[\nabla_{xx} \hat{\mathbf{r}} - 2 \nabla_{xy} \hat{\mathbf{r}} \frac{\nabla_x \hat{\mathbf{r}}}{\nabla_y \hat{\mathbf{r}}} + \nabla_{yy} \hat{\mathbf{r}} \left| \frac{\nabla_x \hat{\mathbf{r}}}{\nabla_y \hat{\mathbf{r}}} \right|^2 \right] (t, x, y); \\
\nabla_y F &= \frac{1}{2} \nabla_y \hat{\mathbf{r}} \left[\nabla_{xx} \hat{\mathbf{r}} - 2 \nabla_{xy} \hat{\mathbf{r}} \frac{\nabla_x \hat{\mathbf{r}}}{\nabla_y \hat{\mathbf{r}}} + \nabla_{yy} \hat{\mathbf{r}} \left| \frac{\nabla_x \hat{\mathbf{r}}}{\nabla_y \hat{\mathbf{r}}} \right|^2 \right] (t, x, y).
\end{aligned}$$

On the other hand, note that $\nabla_t \eta = \hat{\mathbf{r}} \nabla_t \hat{\mathbf{r}}$. Then, again at $(t, x, y) \in \mathbb{G}_{\mathbb{V}}$,

$$\nabla_x \nabla_t \eta = \nabla_x \hat{\mathbf{r}} \nabla_t \hat{\mathbf{r}}, \quad \nabla_y \nabla_t \eta = \nabla_y \hat{\mathbf{r}} \nabla_t \hat{\mathbf{r}}.$$

Plug these into (7.13), we obtain (7.17) immediately. \blacksquare

References

- [1] Ararat, C.; Ma, J.; Wu, W., *Set-valued backward stochastic differential equations*, *Annals of Applied Probability*, 33(5), 3418-3448, (2023).

- [2] Aubin J.P. and Da Prato, G., *The viability theorem for stochastic differential inclusions*, Stochastic Analysis and Applications, 16, 1-15, (1998).
- [3] Aubin, J.-P. and Frankowska, H., *Set-Valued Analysis*. Birkhauser, Boston, MA, (2009).
- [4] Barles, G.; Soner H. M.; Souganidis, P. E., *Front propagation and phase field theory*, SIAM Journal on Control and Optimization, 31(2), 439-469, (1993).
- [5] Bouchard, B. and Touzi, N., *Explicit solution of the multivariate super-replication problem under transaction costs.*, Annals of Applied Probability, 10, 685-708, (2000).
- [6] Chen Y.-G.; Giga Y.; Goto S., *Uniqueness and existence of viscosity solutions of generalized mean curvature flow equations*, Journal of Differential Geometry, 33, 749-786, (1991).
- [7] Evans L. C. and Spruck J., *Motion of level sets by mean curvature. I*, Journal of Differential Geometry, 33, 635-681, (1991).
- [8] Feinstein, Z. and Rudloff, B., *Multi-portfolio time consistency for set-valued convex and coherent risk measures*, Finance and Stochastic, 19, 67-107, (2015).
- [9] Feinstein, Z. and Rudloff, B., *Time consistency for scalar multivariate risk measures*, Statistics & Risk Modeling, 38, 71-90, (2022).
- [10] Feinstein, Z.; Rudloff, B.; and Zhang, J. *Dynamic set values for nonzero sum games with multiple equilibria*, Mathematics of Operations Research, 47, 616-642, (2022).
- [11] Giga Y., *Surface evolution equations: A level set approach*, Monographs in Mathematics, Birkhäuser Basel, (2006).
- [12] İşeri M. and Zhang J., *Set values for mean field games*, preprint, arXiv:2107.01661.
- [13] Kabanov, Y., *Hedging and liquidation under transaction costs in currency markets*. Finance and Stochastics, 3, 237-248, (1999).
- [14] Karnam, C.; Ma, J.; and Zhang, J. *Dynamic Approaches for Some Time Inconsistent Optimization Problems*, Annals of Applied Probability, 27, 3435-3477, (2017).
- [15] Kunita, H., *Stochastic differential equations and stochastic flows of diffeomorphisms*, Springer Berlin Heidelberg (1984), 143–303.

- [16] Sethian J. A., *Curvature and the evolution of fronts*, Communications in Mathematical Physics, 101, 487-499, (1985).
- [17] Soner H. M., *Motion of a set by the curvature of its boundary*, Journal of Differential Equations, 101, 313-372, (1993).
- [18] Soner H. M. and Touzi N., *Dynamic programming for stochastic target problems and geometric flows*, Journal of the European Mathematical Society, 4, 201-236 (2002).
- [19] Soner H. M. and Touzi N., *A stochastic representation for mean curvature type geometric flows*, The Annals of Probability, Vol. 31 No. 3, (2003).
- [20] Zhang, J., *Backward Stochastic Differential Equations – from linear to fully nonlinear theory*, Probability Theory and Stochastic Modeling 86, Springer, New York, 2017.