

DIFFERENTIAL STOCHASTIC VARIATIONAL INEQUALITIES WITH PARAMETRIC OPTIMIZATION*

XIAOJUN CHEN[†], JIAN GUO[‡], AND GUAN WANG[§]

Abstract. This paper is focused on studying the differential stochastic variational inequality with parametric convex optimization (DSVI-O), that is, an ordinary differential equation whose right-hand side involves a stochastic variational inequality and solutions of several dynamic and random parametric convex optimization problems. Let the involved random variable possess the time-dependent distribution, and be measurable and integrable. We first show that the DSVI-O has a weak solution coupled with measurable and integrable solutions of the parametric optimization problems. Next we propose a discrete scheme of DSVI-O by using the time-stepping approximation and sample average approximation and then prove its convergence. Finally, we highlight the superiority of our theoretical results on DSVI-O over the existing ones through applications to an embodied intelligence system for the elderly health, using synthetic health care data generated by Multimodal Large Language Models.

Key words. Differential variational inequality, stochastic variational inequality, parametric optimization, weak solution, measurability, sample average approximation, time-stepping scheme.

MSC codes. 90C15, 90C33, 90C39

1. Introduction. Let $X \subset \mathbb{R}^n$ be a nonempty closed convex set. At each time $t \in \mathbb{R}_+$, uncertainty is represented by a random vector $\xi : \Omega \rightarrow \mathbb{R}^\ell$, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P}_t)$, with time-dependent distribution $P_t := \mathbb{P}_t \circ \xi^{-1}$ concentrated on a fixed Borel set $\Xi \subseteq \mathbb{R}^\ell$. Assume that there exists a fixed σ -finite Borel measure μ on Ξ such that $P_t \ll \mu$ for every $t \in \mathbb{R}_+$. We consider the following differential stochastic variational inequality with parametric convex optimization (DSVI-O):

$$(1.1) \quad \begin{cases} \dot{x}(t) = \Pi_X(x(t) - \mathbb{E}_{P_t}[\Phi(t, \xi, x(t), y(t, \xi))] - x(t), & t \geq 0, \\ x(0) = x_0, \end{cases}$$

$$(1.2) \quad \begin{cases} y_i(t, \xi) \in \mathcal{S}_i(t, \xi, x(t)) := \arg \min_{\mathbf{y}_i \in \mathcal{Y}_i(t, \xi)} f_i(t, \xi, x(t), \mathbf{y}_i), & i = 1, \dots, k, \\ y(t, \xi) = (y_1(t, \xi), \dots, y_k(t, \xi)), & \text{a.e. } \xi \in \Xi. \end{cases}$$

In (1.1), $\Phi : \mathbb{R}_+ \times \mathbb{R}^\ell \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$ is a continuous function and has the form

$$(1.3) \quad \Phi(t, \xi, x(t), y(t, \xi)) = \Phi_1(t, \xi, x(t)) + \sum_{i=1}^k B_i(t, \xi, x(t)) y_i(t, \xi),$$

where $\Phi_1 : \mathbb{R}_+ \times \mathbb{R}^\ell \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a continuous function and $B_i : \mathbb{R}_+ \times \mathbb{R}^\ell \times \mathbb{R}^n \rightarrow \mathbb{R}^{n \times m_i}$ is a continuous matrix-valued function for each $i \in [k] := \{1, \dots, k\}$. The operator

*14 February 2026.

Funding: This work is supported by RGC grant JLFS/P-501/24 for the CAS AMSS-PolyU Joint Laboratory in Applied Mathematics and Hong Kong Research Grant Council projects PolyU15300123, PolyU15300124.

[†]Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong (maxjchen@polyu.edu.hk).

[‡]Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong (jiguo@polyu.edu.hk).

[§]Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong (guan1.wang@polyu.edu.hk).

$\Pi_X : \mathbb{R}^n \rightarrow X$ denotes the metric projection onto X . To compute the function value of Φ at each time $t \geq 0$ and event $\xi \in \Xi$, we need to solve k convex optimization problems in (1.2), where for each $i \in [k]$, the objective function $f_i : \mathbb{R}_+ \times \mathbb{R}^\ell \times \mathbb{R}^n \times \mathbb{R}^{m_i} \rightarrow \mathbb{R}$ is continuous, and convex in $\mathbf{y}_i$ for any fixed $(t, \xi, x(t)) \in \mathbb{R}_+ \times \mathbb{R}^\ell \times \mathbb{R}^n$, and the feasible set is a set-valued map $\mathcal{Y}_i : \mathbb{R}_+ \times \mathbb{R}^\ell \rightrightarrows \mathbb{R}^{m_i}$ taking nonempty closed-convex values $\mathcal{Y}_i(t, \xi) \subseteq \bar{\mathcal{Y}}_i \subset \mathbb{R}^{m_i}$ for all $(t, \xi) \in \mathbb{R}_+ \times \mathbb{R}^\ell$. Here $\bar{\mathcal{Y}}_i$ is a compact set for $i \in [k]$ and $m = \sum_{i=1}^k m_i$.

In many applications under dynamic and stochastic environments, the parameters (mean, variance, etc.) of probability distributions change over time. Time-dependent distributions are more proper for modelling non-stationary dynamic systems than static distributions [10, 20]. In Section 5, we use the time-varying P_t to study an embodied intelligence system for elderly care with heterogeneous data streams that exhibit temporal drift.

If f_i is continuously differentiable with respect to $\mathbf{y}$ for each $i \in [k]$, then (1.2) can be rewritten as the following stochastic variational inequality

$$(1.4) \quad 0 \in \Psi(t, \xi, x(t), y(t, \xi)) + \mathcal{N}_{\mathcal{Y}(t, \xi)}(y(t, \xi)),$$

where

$$\Psi(t, \xi, x(t), y(t, \xi)) = \begin{pmatrix} \nabla_{\mathbf{y}_1} f_1(t, \xi, x(t), y_1(t, \xi)) \\ \vdots \\ \nabla_{\mathbf{y}_k} f_k(t, \xi, x(t), y_k(t, \xi)) \end{pmatrix},$$

$\mathcal{Y}(t, \xi) = \mathcal{Y}_1(t, \xi) \times \dots \times \mathcal{Y}_k(t, \xi)$ and $\mathcal{N}_{\mathcal{Y}(t, \xi)}(y(t, \xi)) = \mathcal{N}_{\mathcal{Y}_1(t, \xi)}(y_1(t, \xi)) \times \dots \times \mathcal{N}_{\mathcal{Y}_k(t, \xi)}(y_k(t, \xi))$ is the normal cone of $\mathcal{Y}(t, \xi)$ at $y(t, \xi)$. When the distribution P and feasible set $\mathcal{Y}$ are time-independent, then the DSVI-O (1.1)-(1.2) reduces to the DSVI [14]

$$(DSVI) \quad \begin{cases} \dot{x}(t) = \Pi_X(x(t) - \mathbb{E}[\Phi(t, \xi, x(t), y(t, \xi))]) - x(t), & t \geq 0, \\ x(0) = x_0, \\ 0 \in \Psi(t, \xi, x(t), y(t, \xi)) + \mathcal{N}_{\mathcal{Y}(\xi)}(y(t, \xi)), & \text{a.e. } \xi \in \Xi. \end{cases}$$

The existence of a solution to the (DSVI) in the space of continuously differentiable functions and in the space of measurable functions, as well as the convergence of the sample average approximation (SAA) and the time-stepping approximation of the (DSVI) are proved in [14]. These results are established under the assumption that (1.4) has a unique solution $y_{\mathbf{x}}(t, \xi)$ for any $(t, \mathbf{x}) \in \mathbb{R}_+ \times \mathbb{R}^n$ and for a.e. $\xi \in \Xi$. Moreover, the function $(t, \mathbf{x}) \mapsto \mathbb{E}[\Phi(t, \xi, \mathbf{x}, y_{\mathbf{x}}(t, \xi))]$ is assumed to be locally Lipschitz continuous at any given $(t, \mathbf{x}) \in \mathbb{R}_+ \times \mathbb{R}^n$.

If the (DSVI) does not involve random variables, then it reduces to the deterministic (DVI) [9, 15, 16, 18, 42, 43]

$$(DVI) \quad \begin{cases} \dot{x}(t) = \Pi_X(x(t) - \Phi(t, x(t), y(t))) - x(t), & t \geq 0, \\ x(0) = x_0, \\ 0 \in \Psi(t, x(t), y(t)) + \mathcal{N}_{\mathcal{Y}}(y(t)). \end{cases}$$

If the data in (DSVI) are time-independent and we seek an equilibrium point at which $\dot{x}(t) = 0$, then the system reduces to the static two-stage stochastic variational inequality (SVI) [11, 12, 13, 44]

$$(SVI) \quad \begin{cases} 0 \in \mathbb{E}[\Phi(\xi, x, y(\xi))] + \mathcal{N}_X(x), \\ 0 \in \Psi(\xi, x, y(\xi)) + \mathcal{N}_{\mathcal{Y}(\xi)}(y(\xi)), & \text{a.e. } \xi \in \Xi. \end{cases}$$

66 If $X = \mathbb{R}^n$, and the objective functions and constraints in (1.2) are deterministic,
 67 then (1.1)-(1.2) reduces to the optimization-constrained ordinary differential equation
 68 (OCODE) [37, 39]

$$\begin{aligned}
 69 \quad (\text{OCODE}) \quad & \begin{cases} \dot{x}(t) = -\Phi(t, x(t), y(t)), & t \geq 0, \\ x(0) = x_0, \\ y(t) \in \arg \min_{\mathbf{y} \in \mathcal{Y}(t)} f(t, x(t), \mathbf{y}). \end{cases}
 \end{aligned}$$

70 The models (DSVI), (DVI), (SVI) and (OCODE) have many applications in trans-
 71 portation sciences, atmospheric chemistry, biological systems and economics and have
 72 been studied extensively over the last decades (e.g., [14, 43, 44]). Nowadays, Artificial
 73 Intelligence (AI) applications leverage static, dynamic, and random data in diverse
 74 forms, including images, text, and audio [38] in time-varying environments. How-
 75 ever, the models (DSVI), (DVI), (SVI) and (OCODE) have limitations in handling
 76 random data with time-dependent distributions from different sources in different
 77 forms simultaneously. Motivated by existing research on (DSVI), (DVI), (SVI) and
 78 (OCODE), and new applications in AI, in this paper, we study the DSVI-O (1.1)-
 79 (1.2) where k optimization problems in (1.2) handle different learning problems with
 80 time-dependent distributions simultaneously. The objective functions in these opti-
 81 mization problems may include high-dimensional data and nonsmooth regularizers,
 82 and thus are not necessarily strongly convex and differentiable. Therefore, DSVI-O
 83 (1.1)-(1.2) cannot be written as the (DSVI), (DVI), (SVI) or (OCODE) and existing
 84 theoretical results cannot be applied to obtain the existence of solutions of DSVI-O
 85 (1.1)-(1.2) directly.

86 We call a pair $(x^* : [0, T] \rightarrow \mathbb{R}^n, y^* : [0, T] \times \Xi \rightarrow \mathbb{R}^m)$ a *weak solution* of (1.1)-
 87 (1.2) if x^* is absolutely continuous in $[0, T]$ and

$$88 \quad (1.5) \quad x^*(t) = x_0 + \int_0^t \left(\Pi_X(x^*(\tau) - \mathbb{E}_{P_\tau} [\Phi(\tau, \xi, x^*(\tau), y^*(\tau, \xi))]) - x^*(\tau) \right) d\tau,$$

89 and y^* is jointly measurable in (t, ξ) and satisfies

$$90 \quad \int_0^T \mathbb{E}_{P_t} [\|y^*(t, \xi)\|] dt < \infty,$$

91 with $y_i^*(t, \xi) \in \mathcal{S}_i(t, \xi, x^*(t))$ for $i \in [k]$, for a.e. $t \in [0, T]$ and a.e. ξ .

92 The main contributions of this paper are summarized as follows.

- 93 1. We prove the existence of a weak solution of (1.1)-(1.2) for any $T > 0$ without
 94 assuming the differentiability of the objective functions $f_i(t, \xi, x(t), \cdot)$ and
 95 uniqueness of solutions $y_i(t, \xi)$ for $i \in [k]$. When f_i is strongly convex in $\mathbf{y}$,
 96 we show that (1.1)-(1.2) has a classical solution (x^*, y^*) with $x^* \in C^1[0, T]$
 97 and $y^*(\cdot, \xi)$ is continuous in $[0, T]$ for a.e. $\xi \in \Xi$. Moreover, for any weak
 98 solution $(x^*(t), y^*(t, \xi))$, we show that $x^*(t) \in X$ for any $t \in [0, T]$ if $x_0 \in X$.
- 99 2. We propose a discrete approximation scheme for (1.1)-(1.2), which first uses the
 100 time-stepping method based on the forward Euler scheme over a uniform time
 101 grid to the ordinary differential equation (ODE) in (1.1) and then uses the
 102 SAA to approximate the expectation value at each discrete time point. We
 103 give convergence theorems of the discrete approximation scheme to a weak
 104 solution of (1.1)-(1.2).

3. We generate synthetic health care data by Multimodal Large Language Models, and test (1.1)-(1.2) with the proposed discrete approximation scheme for problems in the embodied intelligence system for the elderly health. Preliminary results show the efficiency and effectiveness of the DSVI-O model.

1.1. Preliminaries. In this subsection, we introduce the notation and basic definitions used throughout the paper.

Notation. Let $\|\cdot\|$ denote the Euclidean norm of a vector or a matrix. Let $\mathbb{R}_+$ denote the set of all nonnegative real numbers. Given a convex set $X \subseteq \mathbb{R}^n$ and $\mathbf{z} \in \mathbb{R}^n$, $\text{dist}(\mathbf{z}, X) := \min_{\mathbf{x} \in X} \|\mathbf{x} - \mathbf{z}\|$ denotes the distance from $\mathbf{z}$ to X . Given two sets $A, C \subseteq \mathbb{R}^n$, $\mathbb{D}(A, C) := \sup_{\mathbf{x} \in A} \text{dist}(\mathbf{x}, C)$ denotes the deviation of the set A from the set C . The Hausdorff distance between A and C is defined by $\mathbb{H}(A, C) := \max\{\mathbb{D}(A, C), \mathbb{D}(C, A)\}$. Given an $\mathbf{x} \in X$, $\mathcal{N}_X(\mathbf{x}) = \{\mathbf{v} \in \mathbb{R}^n : \mathbf{v}^\top(\mathbf{z} - \mathbf{x}) \leq 0, \forall \mathbf{z} \in X\}$ denotes the normal cone of X at $\mathbf{x}$, and $\mathcal{T}_X(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}^n : \liminf_{\lambda \rightarrow 0^+} \lambda^{-1} \text{dist}(\mathbf{x} + \lambda \mathbf{y}, X) = 0\}$ denotes the contingent cone to X at $\mathbf{x}$. Let $C_b(X)$ denote the set of real-valued bounded continuous functions on X , and $\text{conv}(C)$ denote the convex hull of a set C .

DEFINITION 1.1 (Upper/lower semicontinuous [2]). *We say that a set-valued map $G : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is upper semicontinuous (u.s.c.) if for any $\mathbf{x}_0 \in \mathbb{R}^n$, for any open set $\mathcal{D}$ containing $G(\mathbf{x}_0)$, there exists a neighborhood $\mathcal{N}_{\mathbf{x}_0}$ of $\mathbf{x}_0$ such that for any $\mathbf{x} \in \mathcal{N}_{\mathbf{x}_0}$, $G(\mathbf{x}) \subset \mathcal{D}$. Similarly, we say that G is lower semicontinuous (l.s.c.) if for any $\mathbf{x}_0 \in \mathbb{R}^n$, for any $\mathbf{y}_0 \in G(\mathbf{x}_0)$ and any neighborhood $\mathcal{N}_{\mathbf{y}_0}$ of $\mathbf{y}_0$, there exists a neighborhood $\mathcal{N}_{\mathbf{x}_0}$ of $\mathbf{x}_0$ such that for any $\mathbf{x} \in \mathcal{N}_{\mathbf{x}_0}$, $G(\mathbf{x}) \cap \mathcal{N}_{\mathbf{y}_0} \neq \emptyset$. Moreover, G is said to be continuous if and only if it is both u.s.c. and l.s.c.*

DEFINITION 1.2 (Measurability [3]). *Consider a measurable space $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$ and a set-valued map $G : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ with closed images. The map G is called measurable if for every open subset $\mathcal{O} \subset \mathbb{R}^n$, we have $G^{-1}(\mathcal{O}) := \{\omega \in \mathbb{R}^m : G(\omega) \cap \mathcal{O} \neq \emptyset\} \in \mathcal{B}(\mathbb{R}^m)$.*

DEFINITION 1.3 (Random closed set [41]). *For a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a set-valued map $\zeta : \Omega \rightrightarrows \mathbb{R}^n$ with closed values is called a random closed set if, for every compact set $\mathcal{O} \subseteq \mathbb{R}^n$, $\{\omega \in \Omega : \zeta(\omega) \cap \mathcal{O} \neq \emptyset\} \in \mathcal{F}$.*

DEFINITION 1.4 (Integrably bounded [41]). *A random closed set ζ is said to be integrably bounded if $\|\zeta\| := \sup\{\|\varsigma\| : \varsigma \in \zeta\}$ has a finite expectation, i.e., $\mathbb{E}[\|\zeta\|] < \infty$.*

DEFINITION 1.5 (Integrable selections [41]). *Let $\zeta : \Omega \rightrightarrows \mathbb{R}^n$ be a random closed set. The family of integrable selections of ζ is defined as $L^1(\zeta) := L^0(\zeta) \cap L^1(\mathbb{R}^n)$, where $L^0(\zeta)$ denotes the family of all measurable selections of ζ , and $L^1(\mathbb{R}^n)$ is the space of measurable random elements ξ with $\mathbb{E}[\|\xi\|] < \infty$.*

The rest of this paper is organized as follows. In Section 2, we prove the existence of a weak solution of (1.1)-(1.2) and the feasibility of any weak solution of (1.1)-(1.2). In Section 3, we study the discrete scheme for (1.1)-(1.2) and prove its convergence. In Section 4, we illustrate the theoretical results on the existence of solutions and discrete scheme with numerical examples. In Section 5, we study applications of DSVI-O in the embodied intelligence system for the elderly health by synthetic health care data generated by Multimodal Large Language Models. In Section 6, we present concluding remarks.

2. Existence of weak solutions of (1.1)–(1.2). The right-hand side of the ODE in (1.1) involves k solutions $y_i(t, \xi), i \in [k]$ of the k optimization problems in

(1.2) for $t > 0$ and a.e. $\xi \in \Xi$. Denote the joint solution set of the optimization problems by

$$\mathcal{S}(t, \xi, x(t)) = \mathcal{S}_1(t, \xi, x(t)) \times \mathcal{S}_2(t, \xi, x(t)) \times \cdots \times \mathcal{S}_k(t, \xi, x(t)).$$

For $T > 0$, define a set-valued map $F : [0, T] \times \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ as

$$(2.1) \quad F(t, x(t)) = \mathbb{E}_{P_t} \{ \Phi(t, \xi, x(t), y(t, \xi)) : y(t, \xi) \in \mathcal{S}(t, \xi, x(t)) \}.$$

Then (1.1) can be rewritten as:

$$(2.2) \quad \dot{x}(t) \in H(t, x(t)) := \left\{ \Pi_X(x(t) - \varphi) - x(t) : \varphi \in F(t, x(t)) \right\}.$$

In this section, we show that the linear combination structure of y_i in (1.3) preserves the convexity of the set-valued map F , which is commonly used in differential variational inequality (DVI) and control works (e.g., [4, 22, 43]). If Φ depends on $\mathbf{y}$ in a general nonlinear structure, F may become nonconvex-valued, and existence results need additional structural conditions (see, e.g., [7]) or conditions on $\mathcal{S}(t, \xi, x(t))$ [14, A.2]. These additional conditions are too strong for some applications, when stochastic optimization problems for data fitting in (1.2) have multiple solutions.

Due to the selection of solutions from stochastic optimization problems in (1.2), we cannot expect that (1.1) has a differentiable solution x for any $T > 0$ without strong reasonable conditions. One may ask if we can select the least-norm solution

$$\bar{y}(t, \xi) = \arg \min_{\mathbf{y} \in \mathcal{S}(t, \xi, x(t))} \|\mathbf{y}\|^2.$$

Since the solution set is convex, there is a unique $\bar{y}(t, \xi)$ for any $t \in [0, T]$, a.e. $\xi \in \Xi$. However, even if $\mathcal{S}(t, \xi, \mathbf{x})$ is Lipschitz continuous as a set-valued map in Hausdorff distance, the least-norm solution $\bar{y}(t, \xi)$ may fail to be Lipschitz continuous in $(t, \mathbf{x}) \in [0, T] \times \mathbb{R}^n$, as illustrated by [2, Counterexample, pp. 70-72]. Moreover, the solution set $\mathcal{S}(t, \xi, \mathbf{x})$ is rarely Lipschitz continuous in general. Under typical assumptions, one can at most expect upper semicontinuity, which may not even admit a continuous selection [2, pp. 39].

In this section, we focus on the existence of a weak solution of the DSVI-O (1.1)–(1.2), which allows us to work with integrable and measurable selections while preserving the meaningful evolution of the system state $x(t)$.

For $x_0 \in X$ and $T > 0$, we make the following assumptions.

- (A1) $\mathbb{P}_{(\cdot)}(\cdot) : [0, T] \times \mathcal{F} \rightarrow [0, 1]$ is a *probability kernel* from $[0, T]$ to Ω , that is, for each $A \in \mathcal{F}$, the mapping $t \mapsto \mathbb{P}_t(A)$ is $\mathcal{B}_{[0, T]}$ measurable, and for each $t \in [0, T]$, $\mathbb{P}_t(\cdot)$ is a probability measure on $(\Omega, \mathcal{F})$ [35].
- (A2) There exist measurable functions $\kappa_{\Phi_1}, \kappa_{B_i} : [0, T] \times \Xi \rightarrow \mathbb{R}_+$, $i \in [k]$, such that their expectations $\bar{\kappa}_{\Phi_1}(t) := \mathbb{E}_{P_t}[\kappa_{\Phi_1}(t, \xi)]$ and $\bar{\kappa}_{B_i}(t) := \mathbb{E}_{P_t}[\kappa_{B_i}(t, \xi)]$ are finite for every $t \in [0, T]$ and integrable over $[0, T]$, and for all $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^n$,

$$\|\Phi_1(t, \xi, \mathbf{x}) - \Phi_1(t, \xi, \mathbf{x}')\| \leq \kappa_{\Phi_1}(t, \xi) \|\mathbf{x} - \mathbf{x}'\|,$$

and

$$\|B_i(t, \xi, \mathbf{x}) - B_i(t, \xi, \mathbf{x}')\| \leq \kappa_{B_i}(t, \xi) \|\mathbf{x} - \mathbf{x}'\|.$$

- (A3) There exists a constant $\bar{\beta} > 0$ such that, for $i \in [k]$,

$$\sup_{t \in [0, T]} \max \{ \mathbb{E}_{P_t} [\|\Phi_1(t, \xi, x_0)\|], \mathbb{E}_{P_t} [\|B_i(t, \xi, x_0)\|] \} \leq \bar{\beta}, \quad i \in [k].$$

(A4) The set-valued map $\mathcal{Y} : [0, T] \times \Xi \rightrightarrows \bar{\mathcal{Y}} := \bar{\mathcal{Y}}_1 \times \cdots \times \bar{\mathcal{Y}}_k \subset \mathbb{R}^m$ is continuous.

Remark 2.1. Assumption (A1) requires only that $\mathbb{P}_t$ is a probability kernel from $[0, T]$ to Ω , without imposing any continuity or smoothness in time. This condition ensures that the random environment is in a measurably well-defined way, and enables us to prove the measurability of the set-valued map F in the right-hand side of (2.2) at $t \in [0, T]$. Assumptions (A2)-(A3) impose Lipschitz continuity conditions on the mappings Φ_1 and B_i in x , with integrable expected Lipschitz moduli, supporting the boundedness and linear-growth of the right-hand side of (2.2). Finally, Assumption (A4) guarantees that the solution set map $\mathcal{S}$ is u.s.c. and takes nonempty and compact values.

Let $M := \max_{i \in [k]} \sup_{\mathbf{y} \in \bar{\mathcal{Y}}_i} \|\mathbf{y}\|$. Before establishing the existence of weak solutions, we first present properties of the set-valued maps $\mathcal{S}$ and F .

LEMMA 2.2. Under Assumptions (A1)-(A4), the following statements hold.

- (i) $\mathcal{S}(t, \xi, \mathbf{x})$ is nonempty, convex and compact for any fixed $t \geq 0$, $\xi \in \Xi$ and $\mathbf{x} \in \mathbb{R}^n$, and $\mathcal{S}$ is u.s.c. ;
- (ii) $F(t, \mathbf{x})$ is nonempty, convex and compact for any fixed $t \geq 0$ and $\mathbf{x} \in \mathbb{R}^n$, $F(t, \cdot)$ is u.s.c. for any $t \geq 0$ and $F(\cdot, \mathbf{x})$ is measurable for any fixed $\mathbf{x} \in \mathbb{R}^n$.

Proof. (i) For fixed $t \in \mathbb{R}_+$, $\xi \in \Xi$, and $\mathbf{x} \in \mathbb{R}^n$, since the feasible set $\mathcal{Y}_i(t, \xi)$ is closed and contained in a compact set $\bar{\mathcal{Y}}_i$, and the objective f_i is continuous in $\mathbf{y}_i$, by the Weierstrass theorem, one knows that each set-valued map $\mathcal{S}_i$ is nonempty-valued and compact-valued, and so is the joint map $\mathcal{S}$. Also, noting Assumption (A4), it then from f_i is continuous on $\mathbb{R}_+ \times \Xi \times \mathbb{R}^n \times \mathbb{R}^{m_i}$, $f_i(t, \xi, \mathbf{x}, \cdot)$ is convex on $\mathbb{R}^{m_i}$ and [50, Theorem 3.1] that the set-valued map $\mathcal{S}_i$ is convex-valued and u.s.c. on $\mathbb{R}_+ \times \Xi \times X$. Hence, it follows from [45, Exercise 2.20] that $\mathcal{S}$ is convex-valued. Noticing the compactness of $\mathcal{S}$, we conclude from [2, Proposition 4 pp.64 and Theorem 2 pp.62] that $\mathcal{S}$ is u.s.c.

(ii) For any $(t, \mathbf{x}) \in \mathbb{R}_+ \times \mathbb{R}^n$ and a given realization $\xi \in \Xi$, define the set-valued map

$$(2.3) \quad \mathcal{G}(t, \xi, \mathbf{x}) := \{ \Phi(t, \xi, \mathbf{x}, \mathbf{y}) : \mathbf{y} \in \mathcal{S}(t, \xi, \mathbf{x}) \}.$$

Since $\mathcal{S}$ is u.s.c. with nonempty closed values, it is measurable by [3, Proposition 8.2.1]. Moreover, since Φ is continuous, it follows from [3, Lemma 8.2.8] that $\mathcal{G}$ is measurable. In addition, for each fixed $(t, \xi, \mathbf{x})$, the set $\mathcal{S}(t, \xi, \mathbf{x})$ is compact and $\Phi(t, \xi, \mathbf{x}, \cdot)$ is continuous, which implies that $\mathcal{G}(t, \xi, \mathbf{x})$ is compact. Thus, it follows from the u.s.c. property of $\mathcal{S}$, the continuity of Φ , and [3, Proposition 1.4.14] that $\mathcal{G}$ is u.s.c.

We now study the properties of F . For fixed $(t, \mathbf{x}) \in \mathbb{R}_+ \times \mathbb{R}^n$, $F(t, \mathbf{x})$ is defined as the Aumann integral of the random set $\mathcal{G}(t, \xi, \mathbf{x})$ with respect to the time-varying probability measure $\mathbb{P}_t$

$$F(t, \mathbf{x}) = \int \mathcal{G}(t, \xi, \mathbf{x}) dP_t(\xi) = \left\{ \mathbb{E}_{P_t}[g(\xi)] : g(\xi) \in L^1(\mathcal{G}(t, \xi, \mathbf{x})) \right\}.$$

By Assumptions (A2) and (A3) and recalling Definition 1.4, for fixed $(t, \mathbf{x})$, it follows

$$\begin{aligned}
& \mathbb{E}_{P_t} \left[\sup_{\mathbf{z} \in \mathcal{G}(t, \xi, \mathbf{x})} \|\mathbf{z}\| \right] = \mathbb{E}_{P_t} \left[\sup_{\mathbf{y} \in \mathcal{S}(t, \xi, \mathbf{x})} \left\| \Phi_1(t, \xi, \mathbf{x}) + \sum_{i=1}^k B_i(t, \xi, \mathbf{x}) \mathbf{y}_i \right\| \right] \\
& \leq \mathbb{E}_{P_t} [\|\Phi_1(t, \xi, x_0)\|] + \|\mathbf{x} - x_0\| \bar{\kappa}_{\Phi_1}(t) \\
& + M \sum_{i=1}^k \left(\mathbb{E}_{P_t} [\|B_i(t, \xi, x_0)\|] + \|\mathbf{x} - x_0\| \bar{\kappa}_{B_i}(t) \right) \\
& \leq (kM + 1)\bar{\beta} + (\bar{\kappa}_{\Phi_1}(t) + M \sum_{i=1}^k \bar{\kappa}_{B_i}(t)) \|\mathbf{x} - x_0\| < \infty,
\end{aligned}$$

which implies $\mathcal{G}(t, \xi, \mathbf{x})$ is integrably bounded. Hence, given its measurability as established, [3, Theorem 8.1.3] ensures that $L^1(\mathcal{G}(t, \xi, \mathbf{x})) \neq \emptyset$ for fixed $(t, \mathbf{x})$, and thus $F(t, \mathbf{x})$ is nonempty. In addition, since $\Phi(t, \xi, \mathbf{x}, \cdot)$ is affine and $\mathcal{S}$ is convex-valued, the mapping $\mathcal{G}$ is convex-valued. Thus, Since $\mathcal{G}$ takes values in the finite-dimensional space $\mathbb{R}^n$, it follows from [41, Theorems 2.1.30 and 2.1.37] that the set-valued map F takes convex and closed values. Hence, $F(t, \cdot)$ is compact-valued.

Moreover, since for each fixed $(t, \mathbf{x}) \in \mathbb{R}_+ \times \mathbb{R}^n$, $\mathcal{G}(t, \cdot, \mathbf{x})$ is measurable, by [3, Theorem 8.1.4], $\mathcal{G}(t, \cdot, \mathbf{x}) : \xi \rightrightarrows \mathbb{R}^n$ has a measurable graph. Therefore, noting that $\mathcal{G}$ is u.s.c., [52, Theorem 3.1 and Remark 3.1] ensures that for fixed $t \in \mathbb{R}_+$, the map $F(t, \cdot)$ is u.s.c. Now, we fix $\mathbf{x} \in \mathbb{R}^n$. Note that $\mathcal{G}(\cdot, \cdot, \mathbf{x})$ is a measurable set-valued map on $[0, T] \times \Xi$ with compact values in $\mathbb{R}^n$, and (2.4) gives the P_t -integrable boundedness of $\mathcal{G}(t, \cdot, \mathbf{x})$ for each $t \in [0, T]$. Hence, by [1, Corollary 4.7], the map $F(\cdot, \mathbf{x})$ is measurable. $\square$

THEOREM 2.3 (Existence of a weak solution). *Under Assumptions (A1)–(A4), if $n = 1$ or $X = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{b}\}$ for some given $A \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$, then for any $T > 0$, the DSVI-O (1.1)–(1.2) has at least one weak solution (x^*, y^*) .*

Proof. We now investigate the differential equation (1.1). Note that (1.1) with $y_i(t, \xi) \in \mathcal{S}_i(t, \xi, x(t))$ for $i \in [k]$, can be rewritten using $F(t, x(t))$ as (2.2). Recall the set-valued map $H(t, \mathbf{x}) := \{\Pi_X(\mathbf{x} - \varphi) - \mathbf{x} : \varphi \in F(t, \mathbf{x})\}$. We next verify the assumptions of [21, Theorem 5.2].

Compact convex-valuedness of H . Since $F(t, \mathbf{x})$ takes nonempty and compact values and, for each $(t, \mathbf{x})$, the map $\varphi \mapsto \Pi_X(\mathbf{x} - \varphi) - \mathbf{x}$ is continuous, it follows that $H(t, \mathbf{x})$ is nonempty and compact. If $X = \{\mathbf{x} : A\mathbf{x} = \mathbf{b}\}$ with $A \in \mathbb{R}^{m \times n}$ (and $X \neq \emptyset$), using the explicit formula $\Pi_X(\mathbf{z}) = \mathbf{z} - A^\dagger(A\mathbf{z} - \mathbf{b})$, where $A^\dagger$ denotes the Moore–Penrose pseudoinverse of A , we get $\Pi_X(\mathbf{x} - \varphi) - \mathbf{x} = -A^\dagger(A\mathbf{x} - \mathbf{b}) - (I - A^\dagger A)\varphi$, which is affine in φ and therefore maps the convex set $F(t, \mathbf{x})$ to a convex set. If $n = 1$, then both X and $F(t, \mathbf{x})$ are intervals, and since the one-dimensional projection Π_X maps intervals to intervals, $H(t, \mathbf{x})$ is an interval and therefore convex.

Measurability of $H(\cdot, \mathbf{x})$ and upper semicontinuity of $H(t, \cdot)$. By Lemma 2.2(ii), we have $F(\cdot, \mathbf{x})$ is measurable for each fixed $\mathbf{x}$. Noting that the $\Pi_X(\cdot)$ is Lipschitz continuous, it follows that $H(\cdot, \mathbf{x})$ is measurable. By [3, Proposition 1.4.14], the upper semicontinuity of $H(t, \mathbf{x})$ in $\mathbf{x}$ for each fixed t follows from the upper semicontinuity of $F(t, \mathbf{x})$ and the continuity of the map $(\mathbf{x}, \varphi) \mapsto \Pi_X(\mathbf{x} - \varphi) - \mathbf{x}$.

Linear growth of H . In addition, for any $T > 0$, by the non-expansiveness of Π_X

and noting (2.4) and $x_0 \in X$, it follows

$$\begin{aligned}
 \sup_{\mathbf{v} \in H(t, \mathbf{x})} \|\mathbf{v}\| &= \sup_{\varphi \in F(t, \mathbf{x})} \|\Pi_X(\mathbf{x} - \varphi) - \mathbf{x}\| \\
 &= \sup_{\varphi \in F(t, \mathbf{x})} \|\Pi_X(\mathbf{x} - \varphi) - \Pi_X(\mathbf{x}) + \Pi_X(\mathbf{x}) - \mathbf{x}\| \\
 (2.5) \quad &\leq \sup_{\varphi \in F(t, \mathbf{x})} \|\varphi\| + \|\mathbf{x} - \Pi_X(\mathbf{x})\| \\
 &\leq (kM + 1)\bar{\beta} + (\bar{\kappa}_{\Phi_1}(t) + M \sum_{i=1}^k \bar{\kappa}_{B_i}(t) + 1) \|\mathbf{x} - x_0\| \\
 &\leq \tilde{L}_1(t) \|\mathbf{x}\| + \tilde{L}_2(t),
 \end{aligned}$$

where $\tilde{L}_1(t) := \bar{\kappa}_{\Phi_1}(t) + M \sum_{i=1}^k \bar{\kappa}_{B_i}(t) + 1$ and $\tilde{L}_2(t) := (kM + 1)\bar{\beta} + \tilde{L}_1(t) \|x_0\|$. Since $\bar{\kappa}_{\Phi_1}$ and $\bar{\kappa}_{B_i}$ are integrable over the interval $[0, T]$, so are $\tilde{L}_1$ and $\tilde{L}_2$. Thus $H(t, \mathbf{x})$ satisfies the linear-growth condition.

Tangency condition. Finally, for fixed $(t, \mathbf{x}) \in [0, T] \times X$, pick $\varphi \in F(t, \mathbf{x})$ and set $\mathbf{v} = \Pi_X(\mathbf{x} - \varphi) - \mathbf{x} \in H(t, \mathbf{x})$. Since $\Pi_X(\mathbf{x} - \varphi) \in X$ and X is convex, the segment $\mathbf{x} + \lambda \mathbf{v} = (1 - \lambda)\mathbf{x} + \lambda \Pi_X(\mathbf{x} - \varphi)$ lies in X for every $\lambda \in (0, 1]$. Hence, $\frac{\mathbf{x} + \lambda \mathbf{v} - \mathbf{x}}{\lambda} = \mathbf{v}$, which means $\mathbf{v}$ is in the tangent cone to X at $\mathbf{x}$, and thus $\mathbf{v} \in H(t, \mathbf{x}) \cap \mathcal{T}_X(\mathbf{x}) \neq \emptyset$ for any $(t, \mathbf{x}) \in [0, T] \times X$.

Therefore, according to [21, Theorem 5.2], for any given initial state $x(0) = x_0 \in X$, there exists at least one absolutely continuous trajectory $x^*(t)$ to (2.2) on $[0, T]$ such that $x^*(0) = x_0$.

At last, we construct a jointly measurable selection y^* for the optimization problems in (1.2) that is compatible with the trajectory x^* . Let v^* be a measurable version of $\dot{x}^*$, and, for a.e. $t \in [0, T]$, define $\mathcal{Q}(t) := \{\varphi \in F(t, x^*(t)) : v^*(t) = \Pi_X(x^*(t) - \varphi) - x^*(t)\}$. Since x^* satisfies the differential inclusion, $\mathcal{Q}(t)$ is nonempty for a.e. $t \in [0, T]$. The set-valued map $(t, \xi) \mapsto \mathcal{G}(t, \xi, x^*(t))$ is jointly measurable. Applying the parametrized-integral measurability result [1] to this set-valued map shows that $t \mapsto F(t, x^*(t))$ is measurable. Consequently, $\mathcal{Q}$ has a measurable graph and compact values. Thus, by [3, Theorem 8.1.3], there exists a measurable selection $\varphi^*(t) \in \mathcal{Q}(t)$. Since $\varphi^*(t) \in F(t, x^*(t)) = \int_{\Xi} \mathcal{G}(t, \xi, x^*(t)) P_t(d\xi)$, it follows from [1, Corollary 4.7] that there exists a jointly measurable selection $g^*(t, \xi) \in \mathcal{G}(t, \xi, x^*(t))$ such that $\varphi^*(t) = \int_{\Xi} g^*(t, \xi) P_t(d\xi)$ for a.e. $t \in [0, T]$. Define $\mathcal{R}(t, \xi) := \{\mathbf{y} \in \mathcal{S}(t, \xi, x^*(t)) : \Phi(t, \xi, x^*(t), \mathbf{y}) = g^*(t, \xi)\}$. By the definition of $\mathcal{G}$, $\mathcal{R}$ has nonempty compact values and a measurable graph. Hence, by [3, Theorem 8.1.3], there exists a jointly measurable selection $y^*(t, \xi) \in \mathcal{R}(t, \xi)$. Therefore, $\varphi^*(t) = \mathbb{E}_{P_t}[\Phi(t, \xi, x^*(t), y^*(t, \xi))]$ for a.e. $t \in [0, T]$. Moreover, $\|y^*(t, \xi)\| \leq \sqrt{k}M$, and hence $\int_0^T \mathbb{E}_{P_t}[\|y^*(t, \xi)\|] dt \leq \sqrt{k}MT < \infty$. Together with the state trajectory x^* , the selection y^* satisfies, for every $t \in [0, T]$,

$$x^*(t) = x_0 + \int_0^t (\Pi_X(x^*(\tau) - \mathbb{E}_{P_\tau}[\Phi(\tau, \xi, x^*(\tau), y^*(\tau, \xi))]) - x^*(\tau)) d\tau. \quad \square$$

We complete the proof.

Remark 2.4. The additional condition on X in Theorem 2.3 is used to keep the induced set-valued map H in (2.2) convex-valued. Without this condition, the projection $\Pi_X(\cdot)$ cannot preserve convexity for a convex set-valued map, and [21, Theorem 5.2] is invalid. For nonconvex differential inclusions, existence results require

additional structural conditions (see, e.g., [7]), which are hard to verify here since the induced map is governed by the parametric second-stage problem (1.2).

To relax this condition to a general convex set X , following [14], we can use an alternative assumption that for any given $(t, \mathbf{x})$ there exists a measurable selection $y_{\mathbf{x}}(t, \xi) \in \mathcal{S}(t, \xi, \mathbf{x})$ such that the induced right-hand-side term $G(t, \mathbf{x}) := \mathbb{E}_{P_t} [\Phi(t, \xi, \mathbf{x}, y_{\mathbf{x}}(t, \xi))]$ is (locally) Lipschitz continuous at $(t, \mathbf{x})$. If $f_i(t, \xi, \mathbf{x}, \cdot)$ is uniformly strongly convex and $\nabla_{y_i} f_i(t, \xi, \mathbf{x}, y_i)$ is uniformly Lipschitz continuous with respect to $(t, \mathbf{x})$, then there exists such measurable selection.

Now we consider a special case of the DSVI-O (1.1)-(1.2) where the objective function f_i is uniformly strongly convex in $\mathbf{y}_i$ for $i \in [k]$ in the sense that there is $\eta_i > 0$ such that for any $\lambda \in (0, 1)$,

$$(2.6) \quad f_i(t, \xi, \mathbf{x}, \lambda \mathbf{y} + (1 - \lambda) \mathbf{z}) \leq \lambda f_i(t, \xi, \mathbf{x}, \mathbf{y}) + (1 - \lambda) f_i(t, \xi, \mathbf{x}, \mathbf{z}) - \eta_i \lambda (1 - \lambda) \|\mathbf{y} - \mathbf{z}\|^2$$

holds for every $(t, \xi, \mathbf{x}) \in [0, T] \times \Xi \times \mathbb{R}^n$ and all $\mathbf{y}, \mathbf{z} \in \mathbb{R}^{m_i}$.

COROLLARY 2.5 (Existence of a classical solution). *Suppose that f_i satisfy (2.6) for $i \in [k]$, and that $\Phi_1(t, \xi, x_0)$, $B_i(t, \xi, x_0)$, $\kappa_{\Phi_1}(t, \xi)$, and $\kappa_{B_i}(t, \xi)$ have uniformly bounded p -th moments under P_t on $[0, T]$ for some $p > 1$ and $i \in [k]$. Assume also that P_t is a Feller kernel [6], i.e., the function $t \mapsto \mathbb{E}_{P_t}[\psi(\xi)]$ is continuous and bounded in $[0, T]$ whenever $\psi \in C_b(\Xi)$. Then under Assumptions (A1)–(A4), the DSVI-O (1.1)–(1.2) has at least one classical solution (x^*, y^*) for any $T > 0$ in the sense that (x^*, y^*) is a weak solution of (1.1)–(1.2) with x^* being continuously differentiable and $y^*(\cdot, \xi)$ being continuous for every $\xi \in \Xi$.*

Proof. For each $i \in [k]$, since f_i , $i \in [k]$ is strongly convex (2.6), there exists a unique solution $\hat{y}_i(t, \xi, \mathbf{x})$ of (1.2) for any $(t, \xi, \mathbf{x}) \in \mathbb{R}_+ \times \Xi \times X$. Moreover, by following the proof of Lemma 2.2, we obtain that the joint solution map $\mathcal{S}$ is u.s.c. Since it is single-valued, we conclude that $\hat{y}(t, \xi, \mathbf{x}) := (\hat{y}_1(t, \xi, \mathbf{x}), \dots, \hat{y}_k(t, \xi, \mathbf{x}))$ is continuous on $\mathbb{R}_+ \times \Xi \times X$. From Lemma 2.2, the function $F(t, \mathbf{x}) = \mathbb{E}_{P_t} [\Phi(t, \xi, \mathbf{x}, \hat{y}(t, \xi, \mathbf{x}))]$ is well-defined. Now we prove the continuity of F . Let $(t_\nu, \mathbf{x}_\nu) \rightarrow (t, \mathbf{x})$ be any convergent sequence in $[0, T] \times X$. For each $\nu \in \mathbb{N}$, define $\psi_\nu(\xi) := \Phi(t_\nu, \xi, \mathbf{x}_\nu, \hat{y}(t_\nu, \xi, \mathbf{x}_\nu))$ and $\psi(\xi) := \Phi(t, \xi, \mathbf{x}, \hat{y}(t, \xi, \mathbf{x}))$.

Since P_t is a Feller kernel, we have $P_{t_\nu} \Rightarrow P_t$ as $\nu \rightarrow \infty$. Moreover, the joint continuity of Φ and $\hat{y}$ implies that $\lim_{\nu \rightarrow \infty, \xi' \rightarrow \xi} \psi_\nu(\xi') = \psi(\xi)$, $\xi \in \Xi$. By (2.4), the boundedness of $\{\mathbf{x}_\nu\}_{\nu \geq 1} \cup \{\mathbf{x}\}$, and the uniformly bounded p -th moments, we have

$$\sup_{\nu \geq 1} \mathbb{E}_{P_{t_\nu}} [\|\psi_\nu(\xi)\|^p] + \mathbb{E}_{P_t} [\|\psi(\xi)\|^p] < \infty.$$

Since $p > 1$, the sequence $\{\psi_\nu\}$ is uniformly integrable with respect to $\{P_{t_\nu}\}$. Therefore, by the Lebesgue's dominated convergence theorem for weakly converging measures, applied componentwise,

$$F(t_\nu, \mathbf{x}_\nu) = \mathbb{E}_{P_{t_\nu}} [\psi_\nu] \longrightarrow \mathbb{E}_{P_t} [\psi] = F(t, \mathbf{x}).$$

Thus, F is continuous on $[0, T] \times X$.

Moreover, since $\Pi_X(\cdot)$ is Lipschitz continuous with Lipschitz constant one, the right-hand side of the ODE

$$(2.7) \quad \dot{x}(t) = \Pi_X(x(t) - F(t, x(t))) - x(t)$$

is continuous in $[0, T] \times X$. Thus, the linear growth condition (2.5) and the classical Peano and extension theorem (e.g. [29, Theorems 1.1 and 2.1]) imply that the ODE

(2.7) has at least a solution $x^* \in C^1([0, T])$. Set $y^*(t, \xi) := \hat{y}(t, \xi, x^*(t))$ for any $t \in [0, T]$ and $\xi \in \Xi$. Since $\hat{y}$ is continuous and $x^* \in C^1([0, T])$, it follows that the mapping $t \mapsto y^*(t, \xi)$ is continuous for every $\xi \in \Xi$. Moreover, as $\hat{y}(t, \xi, \mathbf{x})$ is bounded, we have $\mathbb{E}_{P_t}[\|y^*(t, \xi)\|] < \infty$ for all $t \in [0, T]$. Therefore, the pair (x^*, y^*) forms a classical solution on $[0, T]$, which completes the proof. $\square$

Remark 2.6. Corollary 2.5 establishes the existence of a classical solution to the DSVI-O (1.1)–(1.2). However, uniqueness is not guaranteed in general, as it typically requires Lipschitz continuity of the right-hand side of the ODE in (2.7). Below, we provide a special example in which uniqueness holds. Suppose that for each $i \in [k]$, the objective function is given by

$$(2.8) \quad f_i(t, \xi, \mathbf{x}, \mathbf{y}_i) = g_i(t, \xi, \mathbf{x}, \mathbf{y}_i) + h_i(\mathbf{y}_i),$$

where $g_i: [0, T] \times \Xi \times X \times \mathbb{R}^{m_i} \rightarrow \mathbb{R}$ is twice continuously differentiable with respect to $\mathbf{y}$, and $h_i: \mathbb{R}^{m_i} \rightarrow \mathbb{R} \cup \{+\infty\}$ is proper, l.s.c., and convex (possibly nonsmooth). The feasible set is unconstrained, i.e., $\mathcal{Y}_i(t, \xi) = \mathbb{R}^{m_i}$ for all $(t, \xi) \in [0, T] \times \Xi$ and $i \in [k]$. Moreover, assume that $g_i(t, \xi, \mathbf{x}, \cdot)$ is strongly convex for every $(t, \xi, \mathbf{x})$, i.e., the Hessian $\nabla_{\mathbf{y}_i}^2 g_i(t, \xi, \mathbf{x}, \mathbf{y}_i)$ is positive definite.

Under these conditions, the unique solution $\hat{y}_i(t, \xi, \mathbf{x})$ of (1.2) exists, and the solution mapping $p := (t, \xi, \mathbf{x}) \mapsto \hat{y}_i(p)$ is Lipschitz continuous by [25, Corollary 2B.10]. Indeed, for any fixed $\bar{\mathbf{y}}_i \in \mathcal{S}_i(\bar{p})$, define the set-valued mapping

$$\Psi(\mathbf{z}) := \nabla_{\mathbf{y}_i} g_i(\bar{p}, \bar{\mathbf{y}}) + \nabla_{\mathbf{y}_i}^2 g_i(\bar{p}, \bar{\mathbf{y}})(\mathbf{z} - \bar{\mathbf{y}}) + \partial h_i(\mathbf{z}),$$

which is the subgradient of the strongly convex function

$$\psi(\mathbf{z}) := g_i(\bar{p}, \bar{\mathbf{y}}) + (\mathbf{z} - \bar{\mathbf{y}})^\top \nabla_{\mathbf{y}_i} g_i(\bar{p}, \bar{\mathbf{y}}) + \frac{1}{2}(\mathbf{z} - \bar{\mathbf{y}})^\top \nabla_{\mathbf{y}_i}^2 g_i(\bar{p}, \bar{\mathbf{y}})(\mathbf{z} - \bar{\mathbf{y}}) + h_i(\mathbf{z}).$$

It is easy to see that $0 \in \partial\psi(\bar{\mathbf{y}}_i) = \Psi(\bar{\mathbf{y}}_i)$. The strong convexity of ψ implies that its Fenchel conjugate ψ^* is differentiable with Lipschitz continuous gradient (cf. [45, Proposition 12.60]). In particular, $\Psi^{-1} = \nabla\psi^*$ is single-valued and Lipschitz continuous around 0. Therefore, the solution map $\mathcal{S}_i$ is Lipschitz continuous by [25, Corollary 2B.10].

Recall that $H(t, \mathbf{x}) = \Pi_X(\mathbf{x} - \mathbb{E}_{P_t}[\Phi(t, \xi, \mathbf{x}, \hat{y}(t, \xi, \mathbf{x}))]) - \mathbf{x}$. Since $\Phi(t, \xi, \mathbf{x}, \mathbf{y})$ is Lipschitz continuous in $\mathbf{x}$ and $\mathbf{y}$, the projection Π_X is Lipschitz continuous, and the solution map $\hat{y}(t, \xi, \mathbf{x}) = \mathcal{S}_i(t, \xi, \mathbf{x})$ is Lipschitz continuous in $\mathbf{x}$, H is Lipschitz continuous in $\mathbf{x}$. Moreover, if P_t is a Feller kernel, H is also continuous in t . In particular, the Lipschitz property further implies a linear growth property. Consequently, by the Picard–Lindelöf theorem [29, Theorem 3.1], the associated differential equation admits a unique C^1 solution $x^*(\cdot)$ on the entire interval $[0, T]$. This result partially extends [14, Lemma 2.1] to the unconstrained setting.

We next establish the feasibility of the solution to the system (1.1)–(1.2). Specifically, we show that if the initial condition belongs to the closed convex set X , then the entire trajectory remains within X over a finite time horizon.

LEMMA 2.7. *For any $T > 0$, if $x_0 \in X$, then any weak solution (x^*, y^*) of (1.1)–(1.2) satisfies $x^*(t) \in X$ for all $t \in [0, T]$.*

Proof. Define the Lyapunov map $V(x) = \frac{1}{2} \text{dist}^2(x, X) = \frac{1}{2} \|x - \Pi_X(x)\|^2$. Since X is nonempty, convex and closed, $x \in X$ iff $V(x) = 0$ and V is continuously differentiable with $\nabla V(x) = x - \Pi_X(x)$ [27, Theorem 3.2]. Since (x^*, y^*) is a weak solution

of (1.1)-(1.2), $x^*(t)$ satisfies

$$\dot{x}^*(t) \in H(t, x^*(t)) := \{\Pi_X(x^*(t) - \varphi) - x^*(t) : \varphi \in F(t, x^*(t))\}, \text{ a.e. } t \in [0, T],$$

where $F(t, x^*(t))$ is defined in (2.1). By the chain rule [49, Theorem 2.2], for almost every $t \in [0, T]$

$$(2.9) \quad \frac{d}{dt} V(x^*(t)) \in \left\{ \nabla V(x^*(t))^\top v : v \in H(t, x^*(t)) \right\}.$$

Now fix such a time t and pick any $v \in H(t, x^*(t))$. Then there exists $\varphi \in F(t, x^*(t))$ with $v = \Pi_X(x^*(t) - \varphi) - x^*(t)$. Since X is convex, by the first-order optimality condition of $\min_{x \in X} \frac{1}{2} \|x^*(t) - x\|^2$, we have $\langle x^*(t) - \Pi_X(x^*(t)), \Pi_X(x^*(t) - \varphi) - \Pi_X(x^*(t)) \rangle \leq 0$. Consequently,

$$\begin{aligned} \nabla V(x^*(t))^\top v &= \langle x^*(t) - \Pi_X(x^*(t)), \Pi_X(x^*(t) - \varphi) - x^*(t) \rangle \\ &= \langle x^*(t) - \Pi_X(x^*(t)), \Pi_X(x^*(t) - \varphi) - \Pi_X(x^*(t)) \rangle - \|x^*(t) - \Pi_X(x^*(t))\|^2 \\ &\leq -\|x^*(t) - \Pi_X(x^*(t))\|^2 \leq 0. \end{aligned}$$

Because the right-hand side of (2.9) consists entirely of nonpositive values, we conclude that $\frac{d}{dt} V(x^*(t)) \leq 0$ for almost every $t \in [0, T]$. Since $V(x_0) = 0$ (as $x_0 \in X$) and V is non-increasing along the trajectory x^* , it follows that $V(x^*(t)) \equiv 0$ for all $t \in [0, T]$. Hence, $x^*(t) \in X$ for all $t \in [0, T]$. $\square$

3. Discrete approximation and convergence analysis. In this section, we propose a discrete approximation scheme to compute numerical solutions of DSVI-O (1.1)-(1.2). We first apply a time-stepping method based on the forward Euler scheme over a uniform time grid to the ODE in (1.1). Next we apply the SAA to compute approximate values of the expectation at each discrete time. Finally, we define piecewise linear continuous-time functions by linear interpolation using these discrete points and prove the convergence to the x part of a weak solution of DSVI-O (1.1)-(1.2) as the step size in the time-stepping method goes to zero and the sample size of the SAA goes to infinity. Throughout Section 3, we assume that the conditions of Theorem 2.3 hold, so that a weak solution exists.

Throughout Section 3, we assume that the conditions of Theorem 2.3 hold. In addition, we assume that H is jointly upper semicontinuous on $[0, T] \times X$, and that $\bar{\kappa}_{\Phi_1}$ and $\bar{\kappa}_{B_i}$, $i \in [k]$, are uniformly bounded on $[0, T]$.

3.1. Discrete approximation. We describe the time-stepping method for the ODE in (1.1). For a fixed $T > 0$, we define the time step size $h := T/N$ with a fixed $N \in \mathbb{N}$. Consider the uniform partition of the interval $[0, T]$ with nodes $t_\nu := \nu h$ for $\nu = 0, \dots, N$. Given the initial condition $\mathbf{x}_0^N = x(0) = x_0 \in X$, we recursively construct a discrete trajectory $\{\mathbf{x}_\nu^N\}_{\nu=0}^N \subset \mathbb{R}^n$ along with measurable selections $\{y_{\nu,i}^N(\xi)\}_{\nu=0}^N \subset \mathbb{R}^{m_i}$ for $i \in [k]$ as follows.

Forward Euler scheme

$$(3.1) \quad \begin{cases} \mathbf{x}_{\nu+1}^N = \mathbf{x}_\nu^N + h (\Pi_X(\mathbf{x}_\nu^N - \mathbb{E}_{P_{t_\nu}}[\Phi(t_\nu, \xi, \mathbf{x}_\nu^N, y_\nu^N(\xi))]) - \mathbf{x}_\nu^N), \\ y_{\nu,i}^N(\xi) \in \arg \min_{\mathbf{y}_i \in \mathcal{Y}_i(t_\nu, \xi)} f_i(t_\nu, \xi, \mathbf{x}_\nu^N, \mathbf{y}_i), \quad i \in [k], \end{cases}$$

where $y_\nu^N(\xi) := (y_{\nu,1}^N(\xi), \dots, y_{\nu,k}^N(\xi))$.

The forward Euler scheme (3.1) involves an intractable expectation with respect to the distribution P_{t_ν} at each time t_ν . To approximate this expectation, we apply the SAA method.

At each node t_ν , we draw a batch of i.i.d. samples $\{\xi_\nu^j\}_{j=1}^J$ from P_{t_ν} , where the batch size $J \in \mathbb{N}$ is independent of ν . Replacing the expectation by the empirical mean leads to the following SAA–Euler scheme:

SAA-Euler scheme

$$(3.2) \quad \begin{cases} \mathbf{x}_{\nu+1}^{N,J} = \mathbf{x}_\nu^{N,J} + h \left(\Pi_X(\mathbf{x}_\nu^{N,J}) - \frac{1}{J} \sum_{j=1}^J \Phi(t_\nu, \xi_\nu^j, \mathbf{x}_\nu^{N,J}, y_\nu^{N,J}(\xi_\nu^j)) \right) - \mathbf{x}_\nu^{N,J}, \\ y_{\nu,i}^{N,J}(\xi_\nu^j) \in \arg \min_{\mathbf{y}_i \in \mathcal{Y}_i(t_\nu, \xi_\nu^j)} f_i(t_\nu, \xi_\nu^j, \mathbf{x}_\nu^{N,J}, \mathbf{y}_i), \quad i \in [k], j \in [J], \end{cases}$$

where $y_\nu^{N,J}(\xi_\nu^j) := (y_{\nu,1}^{N,J}(\xi_\nu^j), \dots, y_{\nu,k}^{N,J}(\xi_\nu^j))$.

We define the piecewise linear interpolations $x^N, x^{N,J} : [0, T] \rightarrow \mathbb{R}^n$, respectively, as follows

$$(3.3) \quad x^N(t) = \frac{t_{\nu+1} - t}{h} \mathbf{x}_\nu^N + \frac{t - t_\nu}{h} \mathbf{x}_{\nu+1}^N, \quad t \in [t_\nu, t_{\nu+1}], \quad \nu = 0, \dots, N-1.$$

$$(3.4) \quad x^{N,J}(t) = \frac{t_{\nu+1} - t}{h} \mathbf{x}_\nu^{N,J} + \frac{t - t_\nu}{h} \mathbf{x}_{\nu+1}^{N,J}, \quad t \in [t_\nu, t_{\nu+1}], \quad \nu = 0, \dots, N-1.$$

3.2. Convergence analysis. By the proof of Theorem 2.3, we know that the set-valued map $H(t, \mathbf{x}) = \{\Pi_X(\mathbf{x} - \varphi) - \mathbf{x} : \varphi \in F(t, \mathbf{x})\}$ with $F(t, \mathbf{x})$ defined in (2.1) satisfies that $H(\cdot, \mathbf{x})$ is measurable and $H(t, \cdot)$ is u.s.c. However, H may fail to be jointly u.s.c. in $(t, \mathbf{x})$, which prevents a direct application of classical convergence results for difference inclusions [23]. In some special cases, for instance when P_t is Feller together with suitable uniform integrability conditions, the map $H(t, \mathbf{x})$ can be jointly u.s.c. in $(t, \mathbf{x})$. To address the lack of joint u.s.c. only under Assumptions (A1)–(A4), we construct an auxiliary set-valued map H_ϵ that is jointly u.s.c. and satisfies the same linear growth condition. Moreover, $H_\epsilon(t, \mathbf{x}) \subset H(t, \mathbf{x})$, and both inclusions admit the same set of solutions on a large subset of $[0, T]$. The key properties of this approximating inclusion are summarized in the following lemma.

3.3. Convergence analysis. By the proof of Theorem 2.3, we know that the set-valued map $H(t, \mathbf{x}) = \{\Pi_X(\mathbf{x} - \varphi) - \mathbf{x} : \varphi \in F(t, \mathbf{x})\}$ with $F(t, \mathbf{x})$ defined in (2.1) satisfies that $H(\cdot, \mathbf{x})$ is measurable and $H(t, \cdot)$ is u.s.c. However, H may fail to be jointly u.s.c. in $(t, \mathbf{x})$, which prevents a direct application of classical convergence results for difference inclusions [23]. In some special cases, for instance when P_t is Feller together with suitable uniform integrability conditions, the map $H(t, \mathbf{x})$ can be jointly u.s.c. in $(t, \mathbf{x})$. To address the lack of joint u.s.c., we construct an auxiliary set-valued map H_ϵ that is jointly u.s.c. and satisfies the same linear growth condition. Moreover, $H(t, \mathbf{x}) \subseteq H_\epsilon(t, \mathbf{x})$, and the two maps coincide outside a finite set of times. Consequently, they define the same differential inclusion a.e. on $[0, T]$. The key properties of this approximating inclusion are summarized in the following lemma.

LEMMA 3.1. *Suppose that there exists a finite set $E \subset [0, T]$ such that H is jointly u.s.c. on $([0, T] \setminus E) \times X$, and that $\bar{\kappa}_{\Phi_1}$ and $\bar{\kappa}_{B_i}$, $i \in [k]$, are uniformly bounded on $[0, T]$. For any $\epsilon > 0$, there exists a closed set $T_\epsilon \subset [0, T]$ satisfying $\lambda([0, T] \setminus T_\epsilon) \leq \epsilon$ with $\lambda(\cdot)$ being the Lebesgue measure on $[0, T]$, and a jointly u.s.c. set-valued map*

$H_\epsilon : [0, T] \times X \rightrightarrows \mathbb{R}^n$ with closed convex values such that H_ϵ satisfies the linear growth condition and any sequence $\{x^N\}_{N \in \mathbb{N}}$ generated by the time-stepping scheme (3.1) and (3.3) satisfies, for all sufficiently large N , $\dot{x}^N(t) \in H_\epsilon(t_\nu, \bar{x}^N(t))$ for a.e. $t \in [0, T]$ with $\bar{x}^N(t) := \mathbf{x}_\nu^N$ for $t \in (t_\nu, t_{\nu+1})$, $\nu = 0, \dots, N-1$.

Proof. By (2.5) and the uniform boundedness assumptions, there exist constants $a, b > 0$ such that $\sup_{\mathbf{v} \in H(t, \mathbf{x})} \|\mathbf{v}\| \leq a\|\mathbf{x}\| + b$ on $[0, T] \times X$. For every $\epsilon > 0$, choose a closed set $T_\epsilon \subset [0, T] \setminus E$ satisfying $\lambda([0, T] \setminus T_\epsilon) \leq \epsilon$. Define $H_\epsilon(t, \mathbf{x}) := H(t, \mathbf{x})$ if $t \notin E$, and $H_\epsilon(t, \mathbf{x}) := \mathbb{B}(0, a\|\mathbf{x}\| + b)$ if $t \in E$. Then H_ϵ has nonempty compact convex values, contains H , and satisfies the same linear-growth bound. Moreover, $H_\epsilon = H$ on $([0, T] \setminus E) \times X$, and hence on $T_\epsilon \times X$. Since E is finite, H_ϵ is jointly u.s.c. at every $(t, \mathbf{x})$ with $t \notin E$. At a point $(t, \mathbf{x})$ with $t \in E$, upper semicontinuity follows from $H_\epsilon(s, \mathbf{z}) \subseteq \mathbb{B}(0, a\|\mathbf{z}\| + b)$ and the continuity of $\mathbf{z} \mapsto a\|\mathbf{z}\| + b$. Thus H_ϵ is jointly u.s.c. on $[0, T] \times X$. When $0 < h \leq 1$, the convexity of X and (3.1) imply that $\mathbf{x}_\nu^N \in X$. Therefore,

$$\dot{x}^N(t) \in H(t_\nu, \bar{x}^N(t)) \subseteq H_\epsilon(t_\nu, \bar{x}^N(t))$$

for a.e. $t \in (t_\nu, t_{\nu+1})$. This completes the proof. $\square$

The following theorem shows the uniform convergence of $\{x^N\}_{N \in \mathbb{N}}$.

THEOREM 3.2. *Let $\{x^N\}_{N \in \mathbb{N}}$ be the sequence generated by (3.3). Then, under Assumptions (A1)–(A4) and the assumption of Lemma 3.1, every sequence $\{x^N\}_{N \in \mathbb{N}}$ has a uniformly convergent subsequence $\{x^{N_\tau}\}_{\tau \in \mathbb{N}}$ and a limit x such that*

$$\sup_{t \in [0, T]} \|x^{N_\tau}(t) - x(t)\| \rightarrow 0 \quad \text{as } \tau \rightarrow \infty,$$

and the limit x is the x -part of a weak solution of (1.1)–(1.2).

Proof. Let $\epsilon_i \downarrow 0$ as $i \rightarrow \infty$. For each i , by Lemma 3.1 and [23, Theorem 2.2], we have a closed set $T_{\epsilon_i} \subset [0, T]$, a jointly u.s.c. set-valued map H_{ϵ_i} , and a subsequence $\{x^{N_i}(\cdot)\}_{N_i \in \mathbb{N}_{\epsilon_i} \subset \mathbb{N}}$ of $\{x^N(\cdot)\}_{N \in \mathbb{N}}$ that converges uniformly on $[0, T]$ to an absolutely continuous limit $x_i : [0, T] \rightarrow \mathbb{R}^n$ satisfies that $\dot{x}_i(t) \in H_{\epsilon_i}(t, x_i(t))$ for a.e. $t \in [0, T]$ and, since $H_{\epsilon_i} = H$ outside the finite set E , $\dot{x}_i(t) \in H(t, x_i(t))$ for a.e. $t \in [0, T]$. Moreover, $x_i(0) = x_0$ and $x_i(t) \in X$ for all $t \in [0, T]$.

In addition, by [21, Theorem. 7.1], we can extract a uniformly convergent subsequence $\{x_{i_\tau}(\cdot)\}_{\tau \in \mathbb{N}}$ converging to an absolutely continuous limit $x : [0, T] \rightarrow \mathbb{R}^n$ satisfying $\dot{x}(t) \in H(t, x(t))$ a.e. on $[0, T]$. For each τ , choose $K(\tau) \in \mathbb{N}_{\epsilon_{i_\tau}}$ large enough to satisfy $\sup_{t \in [0, T]} \|x^{N_\tau}(t) - x_{i_\tau}(t)\| < \frac{1}{\tau}$ for any $N \in \mathbb{N}_{\epsilon_{i_\tau}} \cap [K(\tau), \infty)$. In particular, choose $K(\tau)$ to be strictly increasing in $\tau \in \mathbb{N}$. Define $N_\tau := K(\tau)$. The subsequence $\{x^{N_\tau}\}_{\tau \in \mathbb{N}}$ satisfies

$$\sup_{t \in [0, T]} \|x^{N_\tau}(t) - x(t)\| \leq \frac{1}{\tau} + \sup_{t \in [0, T]} \|x_{i_\tau}(t) - x(t)\| \rightarrow 0 \quad \text{as } \tau \rightarrow \infty.$$

Finally, by the measurable selection argument in the proof of Theorem 2.3, x is the x -part of a weak solution of (1.1)–(1.2). This completes the proof. $\square$

Remark 3.3. A notable property of the coupled time-stepping scheme (3.1) is that the sequence of first-stage trajectories $\{x^N\}_{N \in \mathbb{N}}$ admits a uniformly convergent subsequence for any selections $\{y^N\}_{N \in \mathbb{N}}$. This indicates that the first-stage dynamics can remain stable with respect to fluctuations in the second-stage variables. To illustrate this, we give the following example. Let $\xi \sim \text{U}(0, 1)$, P denote the corresponding

probability distribution on $\Xi = [0, 1]$, $X = \mathbb{R}$, $k = 1$, $\mathcal{Y}(t, \xi) = \{\mathbf{y} \in \mathbb{R}_+^2 : \mathbf{y}_1 + \mathbf{y}_2 = 1\}$
and

$$f(t, \xi, \mathbf{x}, \mathbf{y}) = (1 + \mathbf{x}^2 + \xi)(\mathbf{y}_1 + \mathbf{y}_2), \quad \Phi(t, \xi, \mathbf{x}, \mathbf{y}) = \mathbf{x} + \sin(2\pi\xi) + (\mathbf{y}_1 - \mathbf{y}_2).$$

Then the optimal set satisfies $\mathcal{S}(t, \xi, \mathbf{x}) = \mathcal{Y}(t, \xi)$ for all $(t, \xi, \mathbf{x})$.

Define a function $r_\nu(\xi) := (-1)^{\lfloor 2^{\nu+1}\xi \rfloor}$ and let $y_\nu^N(\xi) = (\frac{1+r_\nu(\xi)}{2}, \frac{1-r_\nu(\xi)}{2})$ for each $\nu \in \mathbb{N}$. Then $y_\nu^N(\xi) \in \mathcal{S}(t_\nu, \xi, \mathbf{x}_\nu^N)$, $\mathbb{E}_P[y_{\nu,1}^N(\xi) - y_{\nu,2}^N(\xi)] = \mathbb{E}_P[r_\nu] = 0$ and $\mathbb{E}_P[\sin(2\pi\xi)] = 0$. Thus the Euler update reduces to

$$\mathbf{x}_{\nu+1}^N = \mathbf{x}_\nu^N + h(\mathbf{x}_\nu^N - \mathbb{E}_P[\Phi] - \mathbf{x}_\nu^N) = (1 - h)\mathbf{x}_\nu^N,$$

which implies $\mathbf{x}_\nu^N = (1 - h)^\nu \mathbf{x}_0$. After interpolation, the first-stage trajectories converge uniformly to $x(t) = x_0 e^{-t}$. However, define $y^N(t, \xi) := y_\nu^N(\xi)$, $t \in [t_\nu, t_{\nu+1})$. For $M \geq 2N$, we have $\|y^N - y^M\|_{L^2([0, T] \times [0, 1])}^2 = T - \frac{T}{M} \geq \frac{T}{2}$. Thus no subsequence is strongly convergent in this space. This example illustrates that the sequence $\{x^N\}$ may still admit a uniformly convergent subsequence, even when the second-stage selections $\{y^N\}$ do not admit any convergent subsequence.

Moreover, this example also shows that the selection rule plays a critical role in the joint convergence of (x^N, y^N) . For instance, employing the constant selector $y(t, \xi) \equiv (1, 0)$ yields convergent subsequences for both stages. Note that the function f is continuously differentiable and convex with respect to $\mathbf{y}$, and thus the minimization problem $\min_{\mathbf{y} \in \mathcal{Y}(t, \xi)} f(t, \xi, \mathbf{x}, \mathbf{y})$ is equivalent to a variational inequality problem. For continuous selection of solutions to parametric variational inequalities, see [17, 31].

Theorem 3.2 establishes the subsequential uniform convergence of the continuous-time interpolants $\{x^N\}$ generated by the Euler scheme, and shows that any uniform limit is the x -part of a weak solution to (1.1)–(1.2). We next strengthen this result by deriving an error bound analysis on $[0, T]$.

Throughout the analysis, we additionally assume the Riemann integrability of $\bar{\kappa}_{\Phi_1}(t)$ and $\bar{\kappa}_{B_i}(t)$, $i \in [k]$. Set $M_i := \sup_{t \in [0, T]} \tilde{L}_i(t)$, $i = 1, 2$, $R := e^{TM_1}(\|x_0\| + TM_2)$, and $D := X \cap \bar{\mathbb{B}}(0, R)$, where $\tilde{L}_1(\cdot)$ and $\tilde{L}_2(\cdot)$ are given by (2.5). Then any solution $x(\cdot)$ of (2.2) on $[0, T]$ satisfies $\|x(t)\| \leq R$ for all $t \in [0, T]$ by Grönwall's inequality. Moreover, we assume the support set Ξ is compact. We then give the following assumption.

Assumption 3.4. (a) For each $i \in [k]$ and for $\xi \in \Xi$, there exists a measurable function $\alpha_i : \Xi \rightarrow (0, \infty)$ and a constant $\alpha_0 > 0$ such that $\alpha_i(\xi) \geq \alpha_0$ for all $\xi \in \Xi$, and for all $t \in [0, T]$, $\mathbf{x} \in X$, and $\mathbf{y} \in \bar{\mathcal{Y}}_i$,

$$f_i(t, \xi, \mathbf{x}, \mathbf{y}) - v_i(t, \xi, \mathbf{x}) \geq \alpha_i(\xi) \text{dist}(\mathbf{y}, \mathcal{S}_i(t, \xi, \mathbf{x})),$$

with $v_i(t, \xi, \mathbf{x}) := \min_{\mathbf{y} \in \mathcal{Y}_i(t, \xi)} f_i(t, \xi, \mathbf{x}, \mathbf{y})$.

(b) For each $i \in [k]$ and for $\xi \in \Xi$, the map $f_i(\cdot, \xi, \cdot, \cdot)$ is Lipschitz continuous in $(t, \mathbf{x}, \mathbf{y})$ on $[0, T] \times X \times \bar{\mathcal{Y}}_i$, $\mathcal{Y}_i(\cdot, \xi)$ is Hausdorff-Lipschitz continuous on $[0, T]$, and $\Phi_1(\cdot, \xi, \cdot)$ and $B_i(\cdot, \xi, \cdot)$ are Lipschitz continuous in $(t, \mathbf{x})$ on $[0, T] \times X$, with measurable and bounded moduli $\kappa_{f_i}(\xi)$, $\kappa_{\mathcal{Y}_i}(\xi)$, $\bar{\kappa}_{\Phi_1}(\xi)$, and $\bar{\kappa}_{B_i}(\xi)$, respectively.

(c) There exists $L_P > 0$ such that for all $t, t' \in [0, T]$,

$$\|P_t - P_{t'}\|_{\text{TV}} \leq L_P |t - t'|,$$

where $\|\cdot\|_{\text{TV}}$ denotes the total variation distance; see [51, Definition 2.4].

Remark 3.5. Under the standing assumptions of this section, the existence of a weak solution follows from Theorem 2.3. Moreover, Assumption 3.4(a) is the weak sharp minima property in the sense of Burke–Ferris [8]. It allows $\mathcal{S}_i(t, \xi, \mathbf{x})$ to be multi-valued, and together with the remaining parts of Assumption 3.4, it yields the Hausdorff-Lipschitz continuity of the map $(t, \mathbf{x}) \mapsto H(t, \mathbf{x})$.

We further illustrate that Assumption 3.4(c) holds for several standard classes of probability measures by three examples.

(i) Let $\Xi = \{z_1, \dots, z_m\}$ and $P_t = \sum_{j=1}^m p_j(t) \delta_{z_j}$. If each p_j is a Lipschitz continuous function on $[0, T]$ with a constant $L_j > 0$, then

$$\|P_t - P_{t'}\|_{\text{TV}} = \frac{1}{2} \sum_{j=1}^m |p_j(t) - p_j(t')| \leq \frac{1}{2} \sum_{j=1}^m L_j |t - t'|.$$

Hence Assumption 3.4(c) holds with $L_P = \frac{1}{2} \sum_{j=1}^m L_j$.

(ii) Let $P_t = \text{Unif}([a(t), b(t)]^\ell)$ on $\mathbb{R}^\ell$ with $a(t) = a_0 - tc_0$, $b(t) = b_0 + tc_0$, where $a_0 < b_0$ and $c_0 > 0$. Set $I_t = [a(t), b(t)]^\ell$ and $r(t) = b(t) - a(t) = b_0 - a_0 + 2tc_0$. For $t \geq t'$, one has $I_{t'} \subset I_t$. Since the density of P_t is $r(t)^{-\ell} \mathbf{1}_{I_t}$, by [51, Lemma 2.1], we obtain

$$\begin{aligned} \|P_t - P_{t'}\|_{\text{TV}} &= \frac{1}{2} \int_{\mathbb{R}^\ell} \left| \frac{1}{r(t)^\ell} \mathbf{1}_{I_t}(z) - \frac{1}{r(t')^\ell} \mathbf{1}_{I_{t'}}(z) \right| dz \\ &\leq \frac{2\ell c_0(t - t')}{b_0 - a_0 + 2tc_0} \leq \frac{2\ell c_0}{b_0 - a_0} |t - t'|. \end{aligned}$$

The case $t < t'$ is analogous. Thus Assumption 3.4(c) holds with $L_P = 2\ell c_0/(b_0 - a_0)$.

(iii) Let $P_t = \mathcal{N}(m(t), \Sigma)$ on $\mathbb{R}^\ell$, where Σ is a symmetric positive definite matrix and m is a Lipschitz continuous function with a constant $L_m > 0$. By Pinsker's inequality [51, Lemma 2.5] and the Kullback-Leibler divergence between Gaussian distributions with the same covariance matrix [19, Chapter 9, Section 5],

$$\begin{aligned} \|P_t - P_{t'}\|_{\text{TV}} &\leq \frac{1}{2} \left((m(t) - m(t'))^\top \Sigma^{-1} (m(t) - m(t')) \right)^{1/2} \\ &\leq \frac{L_m}{2\sqrt{\lambda_{\min}(\Sigma)}} |t - t'|. \end{aligned}$$

Therefore Assumption 3.4(c) holds with $L_P = L_m/(2\sqrt{\lambda_{\min}(\Sigma)})$.

LEMMA 3.6 (Hausdorff-Lipschitz continuity of H). *Under Assumption 3.4, there exists a constant $L_H > 0$ such that for all $t, t' \in [0, T]$ and all $\mathbf{x}, \mathbf{x}' \in D$,*

$$\mathbb{H}(H(t, \mathbf{x}), H(t', \mathbf{x}')) \leq L_H (|t - t'| + \|\mathbf{x} - \mathbf{x}'\|).$$

Proof. Fix $t, t' \in [0, T]$ and $\mathbf{x}, \mathbf{x}' \in D$, and set $\Delta := |t - t'| + \|\mathbf{x} - \mathbf{x}'\|$. Fix $i \in [k]$ and $\xi \in \Xi$. Take any $\mathbf{y}' \in \mathcal{S}_i(t', \xi, \mathbf{x}') \subseteq \mathcal{Y}_i(t', \xi)$. By Assumption 3.4(b),

$$f_i(t, \xi, \mathbf{x}, \mathbf{y}') \leq f_i(t', \xi, \mathbf{x}', \mathbf{y}') + \kappa_{f_i}(\xi) \Delta = v_i(t', \xi, \mathbf{x}') + \kappa_{f_i}(\xi) \Delta.$$

Let $\mathbf{y} \in \mathcal{S}_i(t, \xi, \mathbf{x})$. Since $\mathcal{Y}_i(\cdot, \xi)$ is Hausdorff-Lipschitz with modulus $\kappa_{\mathcal{Y}_i}(\xi)$, there exists $\bar{\mathbf{y}} \in \mathcal{Y}_i(t', \xi)$ such that $\|\bar{\mathbf{y}} - \mathbf{y}\| \leq \kappa_{\mathcal{Y}_i}(\xi) |t - t'|$. Using Assumption 3.4(b) again,

$$\begin{aligned} v_i(t', \xi, \mathbf{x}') &\leq f_i(t', \xi, \mathbf{x}', \bar{\mathbf{y}}) \leq f_i(t, \xi, \mathbf{x}, \mathbf{y}) + \kappa_{f_i}(\xi) (|t - t'| + \|\mathbf{x} - \mathbf{x}'\| + \|\bar{\mathbf{y}} - \mathbf{y}\|) \\ &\leq v_i(t, \xi, \mathbf{x}) + (1 + \kappa_{\mathcal{Y}_i}(\xi)) \kappa_{f_i}(\xi) \Delta. \end{aligned}$$

Combining the above inequalities yields $f_i(t, \xi, \mathbf{x}, \mathbf{y}') \leq v_i(t, \xi, \mathbf{x}) + (2 + \kappa_{\mathcal{Y}_i}(\xi)) \kappa_{f_i}(\xi) \Delta$,
 and Assumption 3.4(a) implies $\text{dist}(\mathbf{y}', \mathcal{S}_i(t, \xi, \mathbf{x})) \leq \frac{(2 + \kappa_{\mathcal{Y}_i}(\xi)) \kappa_{f_i}(\xi)}{\alpha_0} \Delta$. Swapping
 $(t, \mathbf{x})$ and $(t', \mathbf{x}')$ gives

$$(3.5) \quad \mathbb{H}(\mathcal{S}_i(t, \xi, \mathbf{x}), \mathcal{S}_i(t', \xi, \mathbf{x}')) \leq \frac{(2 + \kappa_{\mathcal{Y}_i}(\xi)) \kappa_{f_i}(\xi)}{\alpha_0} \Delta.$$

Recalling $\mathcal{G}$ in (2.3), using translation invariance and subadditivity of $\mathbb{H}$, we have

$$(3.6) \quad \begin{aligned} \mathbb{H}(\mathcal{G}(t, \xi, \mathbf{x}), \mathcal{G}(t', \xi, \mathbf{x}')) &\leq \|\Phi_1(t, \xi, \mathbf{x}) - \Phi_1(t', \xi, \mathbf{x}')\| \\ &\quad + \sum_{i=1}^k \mathbb{H}(B_i(t, \xi, \mathbf{x}) \mathcal{S}_i(t, \xi, \mathbf{x}), B_i(t', \xi, \mathbf{x}') \mathcal{S}_i(t', \xi, \mathbf{x}')). \end{aligned}$$

Fix $i \in [k]$. By Assumption 3.4(b) and the compactness of Ξ , it follows that for any
 $t \in [0, T]$, $\mathbf{x} \in D$, and $\xi \in \Xi$

$$\|B_i(t, \xi, \mathbf{x})\| \leq \|B_i(0, \xi, x_0)\| + \tilde{\kappa}_{B_i}(\xi)(T + 2R) < \infty.$$

Moreover, $\mathcal{S}_i(t, \xi, \mathbf{x}) \subseteq \bar{\mathcal{Y}}_i$ implies the boundedness of $\mathcal{S}_i(t, \xi, \mathbf{x})$. Combining (3.5)
 with the Lipschitz continuity of $\Phi_1(\cdot, \xi, \cdot)$ and $B_i(\cdot, \xi, \cdot)$ in Assumption 3.4(b) and
 using (3.6), we obtain

$$\mathbb{H}(\mathcal{G}(t, \xi, \mathbf{x}), \mathcal{G}(t', \xi, \mathbf{x}')) \leq C_{\mathcal{G}}(\xi) \Delta,$$

where $C_{\mathcal{G}}(\xi) := \tilde{\kappa}_{\Phi_1}(\xi) + \sum_{i=1}^k \left((\|B_i(0, \xi, x_0)\| + \tilde{\kappa}_{B_i}(\xi)(T + 2R)) \frac{(2 + \kappa_{\mathcal{Y}_i}(\xi)) \kappa_{f_i}(\xi)}{\alpha_0} + \right.$
 $\left. \tilde{\kappa}_{B_i}(\xi) M \right)$ with $M := \max_{i \in [k]} \sup_{\mathbf{y} \in \bar{\mathcal{Y}}_i} \|\mathbf{y}\|$. Since the moduli in Assumption 3.4(b)
 are bounded on Ξ , there exists a constant $L_{\mathcal{G}} > 0$ such that $C_{\mathcal{G}}(\xi) \leq L_{\mathcal{G}}$ for all $\xi \in \Xi$,
 and hence

$$\mathbb{H}(\mathcal{G}(t, \xi, \mathbf{x}), \mathcal{G}(t', \xi, \mathbf{x}')) \leq L_{\mathcal{G}} \Delta.$$

Next, by the definition of the Aumann integral under the fixed measure P_t , we
 have

$$(3.7) \quad \mathbb{H}(\mathbb{E}_{P_t}[\mathcal{G}(t, \xi, \mathbf{x})], \mathbb{E}_{P_{t'}}[\mathcal{G}(t', \xi, \mathbf{x}')]) \leq \mathbb{E}_{P_t}[\mathbb{H}(\mathcal{G}(t, \xi, \mathbf{x}), \mathcal{G}(t', \xi, \mathbf{x}'))] \leq L_{\mathcal{G}} \Delta.$$

For the fixed $\mathbf{x}' \in D$ and $t' \in [0, T]$ and for any $\|u\| \leq 1$, define $h_{\mathcal{G}(t', \xi, \mathbf{x}')} (u) :=$
 $\sup_{g \in \mathcal{G}(t', \xi, \mathbf{x}')} u^\top g$. By (2.3) and Assumption 3.4(b) and the compactness of Ξ , there
 exists a constant $\bar{\Gamma}_R > 0$ such that $\|h_{\mathcal{G}(t', \cdot, \mathbf{x}')} (u)\|_\infty \leq \bar{\Gamma}_R$ for all $\|u\| \leq 1$. Using
 $h_{\mathbb{E}_P[\mathcal{G}]}(u) = \mathbb{E}_P[h_{\mathcal{G}}(u)]$ and [26, Proposition D.2.4], we obtain

$$(3.8) \quad \begin{aligned} &\mathbb{H}(\mathbb{E}_{P_t}[\mathcal{G}(t', \xi, \mathbf{x}')] , \mathbb{E}_{P_{t'}}[\mathcal{G}(t', \xi, \mathbf{x}')]) \\ &= \sup_{\|u\| \leq 1} \left| \mathbb{E}_{P_t}[h_{\mathcal{G}(t', \xi, \mathbf{x}')} (u)] - \mathbb{E}_{P_{t'}}[h_{\mathcal{G}(t', \xi, \mathbf{x}')} (u)] \right| \\ &\leq \sup_{\|u\| \leq 1} \|h_{\mathcal{G}(t', \cdot, \mathbf{x}')} (u)\|_\infty \|P_t - P_{t'}\|_{\text{TV}} \leq \bar{\Gamma}_R \|P_t - P_{t'}\|_{\text{TV}} \\ &\leq \bar{\Gamma}_R L_P |t - t'| \leq \bar{\Gamma}_R L_P \Delta. \end{aligned}$$

Recalling $F(t, \mathbf{x}) = \mathbb{E}_{P_t}[\mathcal{G}(t, \xi, \mathbf{x})]$, by (3.7) and (3.8), we have

$$\mathbb{H}(F(t, \mathbf{x}), F(t', \mathbf{x}')) \leq (L_{\mathcal{G}} + \bar{\Gamma}_R L_P) \Delta.$$

Finally, by nonexpansiveness of Π_X and the definition of H , it follows

$$\mathbb{H}(H(t, \mathbf{x}), H(t', \mathbf{x}')) \leq 2\|\mathbf{x} - \mathbf{x}'\| + \mathbb{H}(F(t, \mathbf{x}), F(t', \mathbf{x}')) \leq (2 + L_G + \bar{\Gamma}_R L_P)\Delta.$$

Thus, the proof is completed with $L_H := 2 + L_G + \bar{\Gamma}_R L_P$. $\square$

THEOREM 3.7 (Euler discretization error). *Under Assumption 3.4, for any $x(\cdot)$ of a weak solution to (1.1)-(1.2) on $[0, T]$, there exists an Euler sequence $\{\mathbf{x}_\nu^N\}_{\nu=0}^N$ such that*

$$\max_{0 \leq \nu \leq N} \|x(t_\nu) - \mathbf{x}_\nu^N\| = O(h).$$

Proof. By Lemma 3.6 and (2.5), the set-valued map $(t, \mathbf{x}) \mapsto H(t, \mathbf{x})$ is Hausdorff-Lipschitz continuous and bounded on $[0, T] \times D$. Hence, applying [24, Proposition, p. 350] completes the proof. $\square$

Using Theorem 3.7, the next corollary shows that, along an Euler sequence $\{\mathbf{x}_\nu^N\}$ generated by (3.1), there exists a selection of second-stage solutions $\{y_{\nu,i}^N\}$ that inherits the convergence of $\{\mathbf{x}_\nu^N\}$.

COROLLARY 3.8. *Fix $i \in [k]$ and $\xi \in \Xi$. Define, for each $t \in [0, T]$, $y_i(t, \xi) := \Pi_{\mathcal{S}_i(t, \xi, x(t))}(0)$, and for each t_ν , $y_{i,\nu}^N(\xi) := \Pi_{\mathcal{S}_i(t_\nu, \xi, \mathbf{x}_\nu^N)}(y_i(t_\nu, \xi))$, $\nu = 0, \dots, N$. Then, under Assumption 3.4, we have*

$$(3.9) \quad \max_{0 \leq \nu \leq N} \|y_{i,\nu}^N(\xi) - y_i(t_\nu, \xi)\| = O(h).$$

Proof. For $\nu = 0, \dots, N$, by the definition of $y_{i,\nu}^N(\xi)$ and $y_i(t_\nu, \xi) \in \mathcal{S}_i(t_\nu, \xi, x(t_\nu))$, we have

$$\begin{aligned} \|y_{i,\nu}^N(\xi) - y_i(t_\nu, \xi)\| &= \text{dist}(y_{i,\nu}^N(\xi), \mathcal{S}_i(t_\nu, \xi, \mathbf{x}_\nu^N)) \\ &\leq \mathbb{H}(\mathcal{S}_i(t_\nu, \xi, x(t_\nu)), \mathcal{S}_i(t_\nu, \xi, \mathbf{x}_\nu^N)). \end{aligned}$$

Therefore, using (3.5) and noting the compactness of Ξ , the result (3.9) follows from Theorem 3.7. $\square$

The SAA-Euler scheme (3.2) replaces the true expectations in the forward Euler method with empirical averages and provides a fully implementable version. We now give a convergence result for the SAA-Euler scheme (3.2).

THEOREM 3.9. *Under Assumptions (A1)-(A4), every sequence $\{x^{N,J}\}_{N,J \in \mathbb{N}}$ defined in (3.2) has a uniformly convergent subsequence $\{x^{N_j, J_j}\}_{j \in \mathbb{N}}$ in $[0, T]$ with $N_j \rightarrow \infty$ and $J_j \rightarrow \infty$ as $j \rightarrow \infty$ to an absolutely continuous function $\bar{x}$ such that for any $r > 0$,*

$$(3.10) \quad \dot{\bar{x}}(t) \in \text{conv}(H^r(t, \bar{x}(t))), \text{ a.e. } t \in [0, T],$$

where $H^r(t, \bar{x}(t)) := \bigcup_{\mathbf{x}' \in \mathbb{B}(\bar{x}(t), r)} H(t, \mathbf{x}')$, $\mathbb{B}(\bar{x}(t), r)$ denotes the open ball of radius r centered at $\bar{x}(t)$ in $\mathbb{R}^n$.

Proof. Recall that (1.1)-(1.2) can be rewritten as the differential inclusion (2.2) and the forward Euler method (3.1) can be written as $\mathbf{x}_{\nu+1}^N \in \mathbf{x}_\nu^N + hH(t_\nu, \mathbf{x}_\nu^N)$, for $\nu = 0, 1, \dots, N-1$ with $\mathbf{x}_0^N = x_0$. Similarly, the SAA-Euler scheme (3.2) can be expressed as $\mathbf{x}_{\nu+1}^{N,J} \in \mathbf{x}_\nu^{N,J} + hH_J(t_\nu, \mathbf{x}_\nu^{N,J})$, where $H_J(t, \mathbf{x}) := \{\Pi_X(\mathbf{x} - \varphi - \mathbf{x} : \varphi \in \frac{1}{J} \sum_{j=1}^J \mathcal{G}(t, \xi^j, \mathbf{x}))\}$ with $\{\xi^j\}_{j=1}^J$ being i.i.d. samples drawn from the distribution P_{t_ν} at each time node t_ν . In the following, we consider the case where F is jointly u.s.c. in (t, x) . When $F(\cdot, \mathbf{x})$ is measurable and $F(t, \cdot)$ is u.s.c., similar results follow from arguments as in Lemma 3.1.

Now we show that the SAA field H_J inherits the linear-growth property of H . Fix $r > 0$, by [48, Theorem 2], it follows

$$(3.11) \quad \sup_{(t, \mathbf{x}) \in [0, T] \times X} \mathbb{D}(H_J(t, \mathbf{x}), H^r(t, \mathbf{x})) \xrightarrow[J \rightarrow \infty]{\text{w.p. } 1} 0.$$

Hence, almost surely, for any $\varepsilon > 0$ there exists a random index J_0 such that for all $J \geq J_0$, all $\mathbf{x} \in X$ and any $\mathbf{v}_J \in H_J(t, \mathbf{x})$, one can find $\mathbf{x}' \in \mathbb{B}(\mathbf{x}, r)$ and $\mathbf{v} \in H(t, \mathbf{x}')$ with $\|\mathbf{v}_J - \mathbf{v}\| \leq \varepsilon$. Then,

$$\|\mathbf{v}_J\| \leq \|\mathbf{v}\| + \|\mathbf{v}_J - \mathbf{v}\| \leq \sup_{\mathbf{u} \in H(t, \mathbf{x}')} \|\mathbf{u}\| + \varepsilon \leq \tilde{L}_1(t) \|\mathbf{x}'\| + \tilde{L}_2(t) + \varepsilon.$$

Because $\|\mathbf{x}'\| \leq \|\mathbf{x}\| + r$, we obtain $\|\mathbf{v}_J\| \leq \tilde{L}_1(t) (\|\mathbf{x}\| + r) + \tilde{L}_2(t) + \varepsilon$, and taking the supremum over $\mathbf{v}_J \in H_J(t, \mathbf{x})$ yields for any $(t, \mathbf{x}) \in [0, T] \times X$, $J \geq J_0$,

$$(3.12) \quad \sup_{\mathbf{v} \in H_J(t, \mathbf{x})} \|\mathbf{v}\| \leq \tilde{L}_1(t) \|\mathbf{x}\| + \tilde{L}_1(t)r + \tilde{L}_2(t) + \varepsilon.$$

Hence, it follows $\|\mathbf{x}_{\nu+1}^{N,J}\| \leq (1 + ah)\|\mathbf{x}_\nu^{N,J}\| + hb$ with $a := \sup_{t \in [0, T]} \tilde{L}_1(t)$ and $b := \sup_{t \in [0, T]} (\tilde{L}_1(t)r + \tilde{L}_2(t) + \varepsilon)$. Then, after iteration, we have,

$$(3.13) \quad \|\mathbf{x}_\nu^{N,J}\| \leq e^{aT} \|\mathbf{x}_0\| + \frac{b}{a}(e^{aT} - 1), \quad \nu = 0, \dots, N-1.$$

Thus, we conclude

$$(3.14) \quad \limsup_{N \rightarrow \infty} \limsup_{J \rightarrow \infty} \sup_{t \in [0, T]} \|x^{N,J}(t)\| < \infty.$$

Moreover, by the recursive inclusion, the derivative $\dot{x}^{N,J}$ exists a.e. in $[0, T]$, and satisfies $\dot{x}^{N,J}(t) \in H_J(t, \mathbf{x}_\nu^{N,J})$, for $t \in (t_\nu, t_{\nu+1})$, $\nu = 0, \dots, N-1$. Combining (3.14) and (3.12) implies:

$$(3.15) \quad \limsup_{N \rightarrow \infty} \limsup_{J \rightarrow \infty} \sup_{t \in [0, T]} \|\dot{x}^{N,J}(t)\| < \infty, \quad \forall t \in (t_\nu, t_{\nu+1}), \quad \nu = 0, \dots, N-1.$$

Hence the family $\{x^{N,J}\}_{N,J \in \mathbb{N}}$ is equi-bounded and equi-continuous, so for any $r > 0$, by the Arzelà-Ascoli theorem, there are subsequences $\{N_j\}$ and $\{J_j\}$ with $J_j \geq \max\{j, J_0\}$ and $N_j \rightarrow \infty$ and $J_j \rightarrow \infty$ as $j \rightarrow \infty$ such that $\{x^{N_j, J_j}\}_{j \in \mathbb{N}}$ converges uniformly to a function $\bar{x}$. In addition, $\dot{x}(\cdot)$ converges weakly in $L_\infty([0, T])^n$ to a limit $\bar{v}$ for some further subsequence (not relabeled). Then by (3.11) and noting $J_j > J_0$, it follows $\sup_{t \in [0, T]} \mathbb{D}(H_J(t, x^{N_j, J_j}(t)), H^r(t, x^{N_j, J_j}(t))) \leq \varepsilon$. Moreover, from (3.15), we have

$$\sup_{\nu \in [N_j]-1} \max_{t_\nu \leq t \leq t_{\nu+1}} \|x^{N_j, J_j}(t) - x^{N_j, J_j}(t_\nu)\| \rightarrow 0, \quad \text{as } j \rightarrow \infty,$$

and therefore, there exists a j_1 such that for all $j > j_1$

$$\dot{x}^{N_j, J_j}(t) \in H_J(t, x^{N_j, J_j}(t)) + \mathbb{B}(\mathbf{0}, \varepsilon) \text{ for a.e. } t \in [0, T].$$

Then it follows from the uniform convergence $x^{N_j, J_j} \rightarrow \bar{x}$, weak convergence $\dot{x}^{N_j, J_j} \rightharpoonup \dot{\bar{x}}$ and the dominated convergence theorem that, for every measurable $\Delta \subset [0, T]$ and $\mathbf{q} \in \mathbb{R}^n$ with $\|\mathbf{q}\| = 1$,

$$\int_\Delta \mathbf{q}^\top \dot{\bar{x}}(\tau) d\tau \leq \int_\Delta \left[\sup_{\mathbf{a} \in H^r(\tau, \bar{x}(\tau))} \mathbf{q}^\top \mathbf{a} + 2\varepsilon \right] d\tau.$$

Since Δ and ε are arbitrary, we have for any $r > 0$, $\dot{\bar{x}}(t) \in \text{conv}(H^r(t, \bar{x}(t)))$ a.e. $t \in [0, T]$. $\square$

4. Numerical experiment. In this section, we verify our theoretical results by a numerical example. The numerical test is conducted in MATLAB 2025a on a Lenovo desktop (2.60GHz, 32.0GB RAM).

Example. Consider problem (1.1)-(1.2) with the following setting. Choose two independent random variables $\xi_1 \sim \mathcal{N}(1, \sigma)$ and $\xi_2 \sim \mathcal{U}(-1 - t, 1 + t)$ and the set $X = \mathbb{R}_+^2$.

The parametric optimization problem in the DSVI-O is defined as (4.1)

$$y_i(t, \xi) \in \arg \min_{\mathbf{y} \in \mathcal{Y}_i(t, \xi)} f_i(t, \xi, x(t), \mathbf{y}) = \frac{1}{2} \|H_i(t, \xi) \mathbf{y} - c_i(x(t), \xi)\|^2 + \mu \|\mathbf{y}\|_1,$$

for $i = 1, 2$, where $\mu > 0$ is a fixed regularization parameter,

$$H_1(t, \xi) = \begin{pmatrix} t + \xi_1 + 0.5 & t + \xi_1 + 1 & \cdots & t + \xi_1 + 5 \\ t + \xi_2 + 0.5 & t + \xi_2 + 1 & \cdots & t + \xi_2 + 5 \\ \xi_1 + \xi_2 + 0.5 & \xi_1 + \xi_2 + 1 & \cdots & \xi_1 + \xi_2 + 5 \end{pmatrix} \in \mathbb{R}^{3 \times 10},$$

$$H_2(t, \xi) = \begin{pmatrix} 1 & t & \xi_1 & \xi_2 \\ t & t + \xi_1 & t + \xi_2 & \xi_1 + \xi_2 \end{pmatrix} \in \mathbb{R}^{2 \times 4},$$

$$c_1(x(t), \xi) = \begin{pmatrix} x_1(t) + \xi_2 \\ x_2(t) + \xi_1 \end{pmatrix} \in \mathbb{R}^3, \quad c_2(x(t), \xi) = \begin{pmatrix} x_1(t) - x_2(t) \\ (\xi_1 - \xi_2) x_2(t) \end{pmatrix} \in \mathbb{R}^2,$$

and the feasible sets are $\mathcal{Y}_1(t, \xi) = (1 + t + \exp(-|\xi_1|) + |\xi_2|) [-\mathbf{e}_1, \mathbf{e}_1]$ with $\mathbf{e}_1 = (1, \dots, 1)^\top \in \mathbb{R}^{10}$ and $\mathcal{Y}_2(t, \xi) = (2 - t + |\xi_2|) [-\mathbf{e}_2, \mathbf{e}_2]$ with $\mathbf{e}_2 = (1, \dots, 1)^\top \in \mathbb{R}^4$. Let $\Phi(t, \xi, x(t), y(t, \xi)) = A(t, \xi)x(t) + \sum_{i=1}^2 B_i(t, \xi, x(t)) y_i(t, \xi)$, where

$$A(t, \xi) = \begin{pmatrix} 1 + \xi_2 + t & \xi_1 - 1 \\ \xi_2 & 3 + \xi_1 \end{pmatrix}, \quad B_1(t, \xi, x(t)) = \begin{pmatrix} (2x_1(t)\xi_1 + t + \xi_2) \mathbf{e}_1^\top \\ (x_1(t)\xi_2 + t + 2\xi_1) \mathbf{e}_1^\top \end{pmatrix},$$

$$B_2(t, \xi, x(t)) = \begin{pmatrix} (2t + \xi_2) \mathbf{e}_2^\top \\ (2t + 2\xi_1) \mathbf{e}_2^\top \end{pmatrix}.$$

Now we illustrate that this example has a weak solution $(x^* : [0, 1] \rightarrow \mathbb{R}^2, y_1^* : [0, 1] \times \Xi \rightarrow \mathbb{R}^{10}, y_2^* : [0, 1] \times \Xi \rightarrow \mathbb{R}^4)$. We briefly verify that assumptions **(A1)**–**(A4)** hold. The random vector (ξ_1, ξ_2) is drawn from the product law $P_t = \mathcal{N}(1, \sigma^2) \otimes \mathcal{U}(-1 - t, 1 + t)$, hence **(A1)** is fulfilled. Denote $\Phi_1(t, \xi, x(t)) = A(t, \xi)x(t)$ and note $B_i(t, \xi, x(t))$ are affine in $x(t)$. Hence, for all $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^2$,

$$\|\Phi_1(t, \xi, \mathbf{x}) - \Phi_1(t, \xi, \mathbf{x}')\| \leq \|A(t, \xi)\| \|\mathbf{x} - \mathbf{x}'\|,$$

$$\|B_1(t, \xi, \mathbf{x}) - B_1(t, \xi, \mathbf{x}')\| \leq \sqrt{4\xi_1^2 + \xi_2^2} \|\mathbf{e}_1\| \|\mathbf{x} - \mathbf{x}'\|,$$

$$\|B_2(t, \xi, \mathbf{x}) - B_2(t, \xi, \mathbf{x}')\| = 0.$$

Since ξ_2 is bounded and $\mathbb{E}[\xi_1] < \infty$, both $\|A(t, \xi)\|$ and $\sqrt{4\xi_1^2 + \xi_2^2} \|\mathbf{e}_1\|$ have finite expectations uniformly in $t \in [0, T]$. Thus **(A2)** is satisfied. For fixed x_0 , the quantities $\mathbb{E}_{P_t}[\|A(t, \xi)x_0\|]$ and $\mathbb{E}_{P_t}[\|B_i(t, \xi, x_0)\|]$ are uniformly bounded on $[0, 1]$, since ξ_2 is bounded and $\mathbb{E}[\xi_1] < \infty$. Thus the corrected Assumption **(A3)** holds. Finally, the

feasible set $\mathcal{Y}_i(t, \xi)$ is continuous in (t, ξ) and contained in $5[-\mathbf{e}_i, \mathbf{e}_i]$, $i = 1, 2$. Therefore, (A4) is satisfied. Consequently, by Theorem 2.3 the system admits at least a weak solution (x^*, y^*) .

To solve the inner ℓ_1 -regularized least squares problem (4.1), we employ the Fast Iterative Shrinkage–Thresholding Algorithm (FISTA) [5]. In the numerical experiments, choose $\mu = 5 \times 10^{-3}$ and let $\hat{x} = (\hat{x}_1, \hat{x}_2)$ denote the reference numerical solution computed with a large sample size $J = 1000$ and a fixed number of time steps $N = 10^4$. For this fixed N , we investigate the effect of varying the standard deviation $\sigma \in \{0.1, 0.5, 1, 1.5\}$ and the sample size $J \in \{30, 100, 200, 500\}$. For each pair (σ, J) , we compute the numerical solution $\tilde{x}_\sigma^{N,J} = (\tilde{x}_{\sigma,1}^{N,J}, \tilde{x}_{\sigma,2}^{N,J})$ and the accuracy

$$R_1 = \frac{1}{10000} \sum_{i=1}^{10000} |\hat{x}_1(ih) - \tilde{x}_{\sigma,1}^{N,J}(ih)|, \quad R_2 = \frac{1}{10000} \sum_{i=1}^{10000} |\hat{x}_2(ih) - \tilde{x}_{\sigma,2}^{N,J}(ih)|$$

50 times and average them. The decreasing tendencies of R_1 and R_2 as σ decreases and J increases are shown in Figure 1

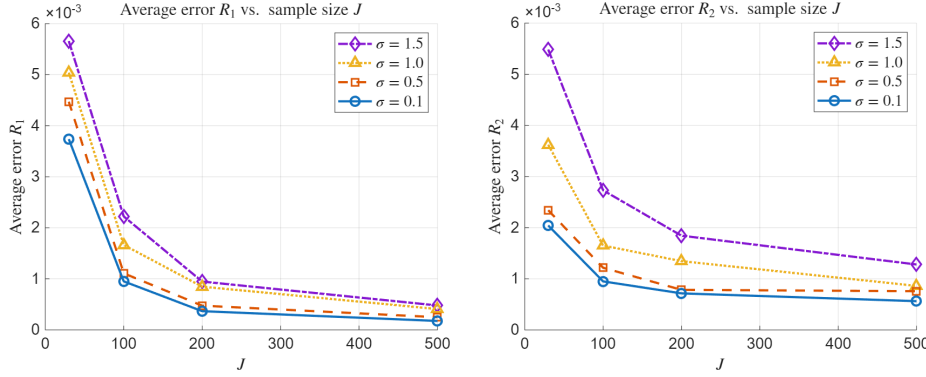


FIG. 1. Decay of R_1 and R_2 with increasing J and decreasing σ .

5. Applications of Embodied Intelligence System. In this section, we show the effectiveness of the DSVI-O with the time-stepping and SAA discrete scheme for delivering personalized health recommendations for elderly individuals aged 60 and above. The DSVI-O framework is specifically designed to provide real-time, individualized health insights through the analysis of multi-source health data, including wearable devices and medical records. Due to privacy constraints and the limited availability of continuous real-world health monitoring data, especially among elderly populations, this study utilizes synthetic data generation methodologies to rigorously evaluate the performance of the DSVI-O with the discrete scheme.

Synthetic data generation has gained significant recognition for its ability to replicate complex, real-world scenarios while preserving data privacy and enabling comprehensive algorithm testing [28, 33, 40]. By creating diverse, physiologically plausible datasets that simulate realistic variations in elderly health conditions, the synthetic approach ensures robust, scalable validation of the proposed framework. Figure 2 illustrates the application structure of the DSVI-O, highlighting the integration and processing of multi-source data streams to generate real-time personalized health recommendations tailored for elderly individuals.

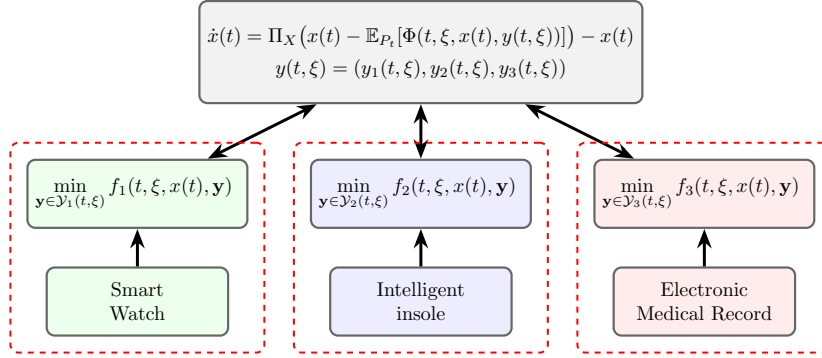


FIG. 2. DSVI-O workflow for personalized elderly health recommendations

5.1. Data Generation. The generated data encompass physiological parameters, clinical records, and wearable device outputs, such as insole sensor readings. This comprehensive yet controlled synthetic approach facilitates rigorous and scalable evaluation of the DSVI-O framework while effectively addressing data privacy and scarcity constraints.

5.1.1. Hybrid Generation of Synthetic Health Data. Traditional rule-based methods lack flexibility and realism, while deep learning approaches typically require large datasets and yield less interpretable results [47]. To address these limitations, our hybrid approach combines rule-based constraints, knowledge distillation from Large Language Models (LLMs), and stochastic mathematical modeling to ensure both realism and interpretability. Specifically, we employ:

1. Physiological rules (e.g., BMI sampled from $\mathcal{N}(22, 2.5^2)$, clipped to $[18.5, 30]$),
2. LLM (Qwen2.5-7B-Instruct) for flexible, guideline-based rule generation,
3. Stochastic mathematical modeling, using Gaussian distributions centered on clinical medians [30].

Individual Attributes. Age groups are 60–69 (50%), 70–79 (30%), 80+ (20%). Gender ratio is 1:1. Height follows age-adjusted distributions: $\mathcal{N}(170 - 0.1(\text{age} - 60), 5^2)$ for men, $\mathcal{N}(160 - 0.1(\text{age} - 60), 5^2)$ for women. Values are clipped to the range $[145 \text{ cm}, 190 \text{ cm}]$. Weight is calculated using the formula $\text{BMI} \times (\text{height}/100)^2$.

Clinical Indicators. Static EMRs (e.g., WBC, RBC, HGB, ...) are sampled from normal distributions around clinical medians, adjusted for age, gender, and BMI, within clinical limits.

Health State Groups. Users are categorized into healthy, weak or ill groups (e.g., heart rate, SpO_2 , sleep, steps, and calories), with specific indicator shifts (e.g., heart rate group: CRP +2, WBC +1.5), clipped within medical ranges.

Physiological Data. The physiological data were generated at 5-second intervals, incorporating circadian rhythms, activity, and health status. The baseline heart rate was defined as $\text{HR} = 70 - 0.2(\text{age} - 60) + 2\mathbb{I}_{\{\text{male}\}} + 0.05(\text{SBP} - 120) + 0.01(\text{HGB} - 140)$, and modulated by ± 8 bpm based on sleep or activity state.

Insole Data. The insole data were generated at 5-second intervals and include 17 indicators such as plantar pressures on four channels (left forefoot, left heel, right forefoot, right heel), step frequency, stride length, left–right symmetry, force estimation, and balance status. These indicators are influenced by factors such as weight, age, gait pattern, and abnormal conditions, with added Gaussian noise and values constrained within physiologically realistic ranges.

5.1.2. Overview of the Generated Dataset. The test dataset consists of 10 individuals, each with 10 days of randomly generated testing data. For each individual, the experiments also use 90 days of historical data. The possible health states for each individual include: healthy (0), weak (1), and ill (2), as well as transitional states between these categories, such as 0.1 or 1.6. Each individual’s daily health status evolves according to one of four temporal patterns: jump (sudden transitions, e.g., acute events), linear (gradual change, e.g., fatigue accumulation), exponential (rapid onset with saturation, e.g., drug response), and logistic (slow-fast-slow transitions, e.g., chronic recovery). These patterns describe transitions between health states over each 24-hour period. Labels are recorded every five seconds, yielding fine-grained time-series trajectories. Such structured modeling of heterogeneous transitions aligns with biomedical insights on personalized health trajectories [34].

Each person is represented by a series of features organized into multiple categories, each stored in a separate CSV file. All generated data are publicly available on GitHub at: <https://github.com/DSVI2025/DSVI-O>.

- **user_profiles.csv**: demographic attributes (user ID, age, gender, height, weight, BMI).
- **user_groups.csv**: health trajectory types and group labels.
- **medical_records.csv**: blood counts, metabolic panels, blood pressure, etc.
- **health_data.csv**: daily metrics including heart rate, SpO₂, sleep quality, steps, calories, skin temperature, hrv, stress index, sbp, dbp, activity intensity, active minutes, floors climbed activity, and status.
- **insole_data.csv**: foot pressure, motion types, and gait information from wearable insole devices.

5.2. Data-Driven Health State Modeling via DSVI-O. We apply DSVI-O (1.1)–(1.2) to dynamically model the evolution of an individual’s health status over 24 hours. The first-stage state variable $x(t) \in \mathbb{R}^n$ denotes the personalized health state of an elderly individual at time t , evolving continuously under the influence of real-time monitoring data. The second stage incorporates inputs from three heterogeneous sources: smart watch signals, intelligent insole data, and EMR. Historical data from each source is used to continuously calculate the time-varying parameters in the ODE, enabling real-time tracking and personalized updates of the health trajectory.

To capture the multi-factor evolution of an elderly individual’s health quality, we model the latent state by an ODE assembled from three physiologically motivated components.

1. *Self-dynamics model.* Following the frailty index framework in [46], the self-evolution of an individual’s health status through an ODE is modeled as

$$\dot{x}(t) = -A(t)x(t) - p(t),$$

where $A(t)$ and $p(t)$ are damage-repair parameters.

2. *Circadian rhythm (CR) model.* Wearable sensors now provide high-resolution, non-invasive activity traces, enabling large-scale modelling of individual CR patterns. We embed this circadian component in the ODE (1.1) to reflect the 24-hours signal by a truncated Fourier series [36]:

$$q(t, \xi) = \sum_{r=1}^M A_r(t) \sin(2\pi\zeta_r t + \phi_r(t)) + \varepsilon_1(t, \xi),$$

where $\phi_r(t)$ and $A_r(t)$ are the phase and amplitude of the r -th harmonic, $\varepsilon_1(t, \xi)$ aggregates noise and ζ_r sets the circadian frequencies.

3. *Latent exogenous drive.* The health state can be perturbed by external factors such as physical activity, gait stability, or clinical interventions. We represent these effects through a *latent exogenous signal* $\ell(t, \xi) = \ell_0 + \Delta\ell_1(t) m(t, \xi)$, where ℓ_0 is the long-run baseline and $\Delta\ell_1(t)$ captures time-varying departures. The composite score $m(t, \xi)$ is built from the latest wearable-sensor read-outs via a linear fusion model

$$m(t, \xi) = \sum_{i=1}^3 \alpha_i(t) B_i(t, \xi) y_i(t, \xi) + \varepsilon_2(t, \xi),$$

with $B_i(t, \xi)$ the i -th physiological channel, $y_i(t, \xi)$ the physiological parameter, and $\alpha_i(t)$ an adaptive weight.

Building upon the three components above, the latent health state $x(t) \in X$ is modeled as the following

$$(5.1) \quad \dot{x}(t) = \Pi_X \left(x(t) - A(t) x(t) - p(t) + \mathbb{E}_{P_t} [q(t, \xi) + \ell(t, \xi)] \right) - x(t).$$

To evaluate $m(t, \xi)$ in $\ell(t, \xi)$, we solve the following constrained quadratic program:

$$(5.2) \quad y_i(t, \xi) \in \arg \min_{\mathbf{y} \in [a_i(\xi), b_i(\xi)]} f_i(t, \xi, x(t), \mathbf{y}), \quad i = 1, 2, 3,$$

where $f_i(t, \xi, \mathbf{x}, \mathbf{y}) = \frac{1}{2} \|H_i(t, \xi) \mathbf{y} + N_i(t, \xi) \mathbf{x} - c_i(t, \xi)\|^2 + \frac{\rho_i}{2} \|B_i(t, \xi) \mathbf{y} - \mathbf{x}\|^2$. Here, the first term ensures that the model follows clinical priors, while the second term keeps the estimated state close to physiological features. These parameter estimates are then used to compute $m(t, \xi)$ and later $\ell(t, \xi)$, forming a feedback loop in the projected ODE. In addition, the matrices $H_i(t, \xi) \in \mathbb{R}^{n_i \times m_i}$ and $c_i(t, \xi) \in \mathbb{R}^{n_i}$ are built from the individual's historical data, where H_i collects past wearable and physiological signals, and c_i gives the related clinical labels. The correction matrix $N_i(t, \xi) \in \mathbb{R}^{n_i \times n}$ helps reduce the effect of past data that are not similar to the current state.

The following conditions and Assumption **(A1)** ensure the existence of a weak solution to (5.1)–(5.2).

The following conditions, together with Assumption **(A1)**, ensure the existence of a weak solution to (5.1)–(5.2).

(B1) The functions $A(\cdot)$, $p(\cdot)$, $\Delta\ell_1(\cdot)$, $\alpha_i(\cdot)$, $i = 1, 2, 3$, $A_r(\cdot)$, and $\phi_r(\cdot)$, $r = 1, \dots, M$, are continuous on $[0, T]$.

(B2) The mappings ε_j , $j = 1, 2$, and B_i , $i = 1, 2, 3$, are continuous, and $\mathbb{E}_{P_t} [\|\varepsilon_j(t, \xi)\|]$ and $\mathbb{E}_{P_t} [\|B_i(t, \xi)\|]$ are uniformly bounded in $t \in [0, T]$, and hence integrable on $[0, T]$.

(B3) The functions $a_i, b_i : \Xi \rightarrow \mathbb{R}^{m_i}$, $i = 1, 2, 3$, are continuous and uniformly bounded, and $a_i(\xi) \leq b_i(\xi)$ componentwise for all $\xi \in \Xi$.

It is easy to verify that **(A2)**–**(A4)** hold in the present example. Indeed, by the definitions of q , ℓ , and m , **(A2)** holds with $\tilde{\kappa}_{\Phi_1}(t, \xi) = \|A(t)\|$, $\tilde{\kappa}_{B_i}(t, \xi) = 0$, $i = 1, 2, 3$. Moreover, **(A3)** holds with $\bar{\beta} := \sup_{t \in [0, T]} \{ \|A(t)x_0\| + \|p(t)\| + \|\ell_0\| + \mathbb{E}_{P_t} [\|q(t, \xi)\|] + |\Delta\ell_1(t)| \mathbb{E}_{P_t} [\|\varepsilon_2(t, \xi)\|] + |\Delta\ell_1(t)| \max_{i \in [3]} |\alpha_i(t)| \mathbb{E}_{P_t} [\|B_i(t, \xi)\|] \} < \infty$. Finally, **(B3)** implies **(A4)**. Thus, by Theorem 2.3, (5.1)–(5.2) has at least one weak solution on $[0, T]$ for any $T > 0$.

5.3. Numerical Experiment Setup. We use the DSVI-O model (5.1)–(5.2) to simulate the evolution of the health status of 10 persons over 10 days by using their health data in the past 90 days. Within every 24-hour cycle, measurements were sampled at 5-second intervals, yielding 17,280 time points per day. We consider

three types of multimodal data: smart watch signals (14 features), intelligent insole measurements (17 features), and electronic medical records (30 features) as illustrated in Figure 2.

For each person, the following three matrices and three vectors contain his/her health features and corresponding labels for $\nu = 1, \dots, 17280$:

- **smart watch** $Z_1(t_\nu) \in \mathbb{R}^{100 \times 14}$, $\eta_1(t_\nu) \in \mathbb{R}^{100}$
- **intelligent insole** $Z_2(t_\nu) \in \mathbb{R}^{100 \times 17}$, $\eta_2(t_\nu) \in \mathbb{R}^{100}$
- **medical records** $Z_3(t_\nu) \in \mathbb{R}^{100 \times 30}$, $\eta_3(t_\nu) \in \mathbb{R}^{100}$.

At each time t_ν , from the first 90 rows of $Z_i(t_\nu)$, we randomly and independently chose 50 rows 20 times. This yields sampled matrices $H_i(t_\nu, \xi^\iota) \in \mathbb{R}^{50 \times m_i}$ and corresponding labels $c_i(t_\nu, \xi^\iota) \in \mathbb{R}^{50}$, $\iota = 1, \dots, 20$ and $i = 1, 2, 3$ with $m_1 = 14$, $m_2 = 17$ and $m_3 = 30$.

We simulate the health dynamics of a single person over one day. For each day $j \in [10]$, we extract the j -th day's data for the person $B_i^j(t_\nu) := (Z_i(t_\nu))_{[90+j, :]}$, $i \in [3]$, $\nu \in [17280]$. This process is repeated across $j = 1, \dots, 10$ to obtain 10 simulated trajectories for the same person.

To reduce the influence of dissimilar historical data, we define a correction vector $N_i^j(t_\nu, \xi^\iota) \in \mathbb{R}^{50}$ based on softmax-normalized Gaussian similarities. Specifically, given the test feature vector $B_i^j(t_\nu)$ and the historical data matrix $H_i(t_\nu, \xi^\iota)$, we compute

$$(5.3) \quad N_i^j(t_\nu, \xi^\iota) = \lambda_i \left(\mathbf{1}_{50} - \text{softmax} \left(-\frac{1}{h_i^2} \cdot \text{pdist2} \left(B_i^j(t_\nu), H_i(t_\nu, \xi^\iota) \right) \right) \right),$$

with h_i being defined as the median of all pairwise distances between rows of $H_i(t_\nu, \xi^\iota)$:

$$h_i = \text{median} \left\{ \|(H_i(t_\nu, \xi^\iota))_{[r, :]} - (H_i(t_\nu, \xi^\iota))_{[s, :]}\| \mid 1 \leq r < s \leq 50 \right\}.$$

The **pdist2** computes a vector in $\mathbb{R}^{50}$ whose the r th component is the squared Euclidean distance between $\mathbf{B} \in \mathbb{R}^{1 \times m_i}$ and the r th row of $\mathbf{H} \in \mathbb{R}^{50 \times m_i}$: $(\text{pdist2}(\mathbf{B}, \mathbf{H}))_r = \sum_{k=1}^{m_i} (\mathbf{B}_{[1, k]} - \mathbf{H}_{[r, k]})^2$, $r = 1, \dots, 50$. For a vector $\mathbf{v} \in \mathbb{R}^{50}$, the **softmax**($\mathbf{v}$) $\in \mathbb{R}^{50}$ is defined by $(\text{softmax}(\mathbf{v}))_r = \frac{e^{v_r}}{\sum_{k=1}^{50} e^{v_k}}$, $r = 1, \dots, 50$.

The correction term introduces a similarity-dependent adjustment to the historical regression targets.

All simulations use the following parameters: $A(t) \equiv \exp(-3)$, $p(t) \equiv 0.01$, $q(t, \xi) = q(t) = 0.15 \sin(2\pi \cdot 0.025t) + 0.17 \sin(2\pi \cdot 0.021t)$, $\varepsilon_2(t, \xi) \equiv 0$, $\ell_0 = 0.02$, $\Delta \ell_1(t) \equiv 1.5$, $\alpha_1(t) \equiv 0.4$, $\alpha_2(t) \equiv 0.4$, $\alpha_3(t) \equiv 0.2$, $\lambda_i = 0.50$, and $\rho_i = 5$ for $i = 1, 2, 3$. The initial state is defined as $\mathbf{x}_0^j = \eta_1(t_0)_{[90+j]}$ for each $j \in [10]$. Overall, for each day $j \in [10]$ and time step $\nu = 0, \dots, 17279$, the latent health state $\mathbf{x}_\nu^j \in [0, 2]$ evolves as

$$(5.4) \quad \mathbf{x}_{\nu+1}^j = \Pi_{[0, 2]} \left(\mathbf{x}_\nu^j - A(t_\nu) \mathbf{x}_\nu^j - p(t_\nu) + q(t_\nu) + \frac{1}{20} \sum_{\iota=1}^{20} \ell^j(t_\nu, \xi^\iota) \right),$$

where $\Pi_{[0, 2]}$ denotes projection onto the interval $[0, 2]$, and

$$\ell^j(t_\nu, \xi^\iota) = \ell_0 + \Delta \ell_1(t_\nu) \left(\sum_{i=1}^3 \alpha_i(t_\nu) B_i^j(t_\nu) \mathbf{y}_i^j(t_\nu, \xi^\iota) + \varepsilon_2(t_\nu, \xi^\iota) \right).$$

At each t_ν , the parameter $\mathbf{y}_i^j(t_\nu, \xi^\iota)$ for $i = 1, 2, 3$ solves

$$\min_{\mathbf{y}_i \in [-10, 10]^{m_i}} \frac{\rho_i}{2} \left\| B_i^j(t_\nu) \mathbf{y}_i - \mathbf{x}_\nu^j \right\|^2 + \frac{1}{2} \left\| H_i(t_\nu, \xi^\iota) \mathbf{y}_i + N_i^j(t_\nu, \xi^\iota) \mathbf{x}_\nu^j - c_i(t_\nu, \xi^\iota) \right\|^2.$$

895 Then, the predicted health state is discretized as

$$896 \quad (5.5) \quad \text{Pred}_\nu^j = \begin{cases} 0, & \text{if } \mathbf{x}_\nu^j \leq \frac{2}{3}, \\ 1, & \text{if } \frac{2}{3} < \mathbf{x}_\nu^j < \frac{4}{3}, \\ 2, & \text{if } \mathbf{x}_\nu^j \geq \frac{4}{3}. \end{cases}$$

897 In addition, the real health state trajectories are set as:
(5.6)

$$898 \quad \tilde{\mathbf{x}}_{\nu+1}^j = \Pi_{[0,2]} \left(\tilde{\mathbf{x}}_\nu^j - A(t_\nu) \tilde{\mathbf{x}}_\nu^j - p(t_\nu) + q(t_\nu) + \ell_0 + \Delta \ell_1(t_\nu) \sum_{i=1}^3 \alpha_i(t_\nu) \eta_i(t_\nu)_{[90+j]} \right).$$

899 Finally, to assess generalizability, the entire procedure above is replicated for 10
900 different persons.

901 **5.4. Numerical Experiment Result.** Following the experimental setup de-
902 scribed above, we now present the evaluation outcomes. The numerical tests are
903 implemented in Python 3.12 on an Apple M3 (4.05 GHz, 16 GB RAM). We adopt an
904 individual-wise testing strategy to evaluate the last 10 days of data for each of the
905 10 individuals in the test set separately. Performance metrics are computed at the
906 individual level and then averaged to obtain the overall performance. This dual per-
907 spective allows us to assess both personalized and aggregate behavior of the DSVI-O.

908 **5.4.1. Performance Metrics.** To quantitatively assess the performance of our
909 classification model, we employ a comprehensive set of evaluation metrics: **Accuracy**,
910 **Precision**, **Recall**, **Specificity**, and **F1-score** [32]. These metrics are computed
911 based on the following quantities for each class i ($i = 0, 1, 2$):

- 912 • **True Positives** (TP_i): Number of samples correctly predicted as class i .
- 913 • **True Negatives** (TN_i): Number of samples correctly predicted as not be-
914 longing to class i .
- 915 • **False Positives** (FP_i): Number of samples incorrectly predicted as class i .
- 916 • **False Negatives** (FN_i): Number of samples incorrectly predicted as not
917 class i .

918 Based on these definitions, we compute the following *per-class metrics*, each quan-
919 tifying a distinct aspect of model performance:

- 920 • **Accuracy**: Measures the overall proportion of correctly classified instances
921 among all predictions

$$922 \quad \text{Accuracy}(i) = \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}.$$

- 923 • **Precision**: Indicates the proportion of positive predictions that are actually
924 correct

$$925 \quad \text{Precision}(i) = \frac{TP_i}{TP_i + FP_i}.$$

- 926 • **Recall**: Represents the proportion of actual positives correctly identified

$$927 \quad \text{Recall}(i) = \frac{TP_i}{TP_i + FN_i}.$$

- 928 • **Specificity**: Quantifies the proportion of actual negatives correctly identified

$$929 \quad \text{Specificity}(i) = \frac{TN_i}{TN_i + FP_i}.$$

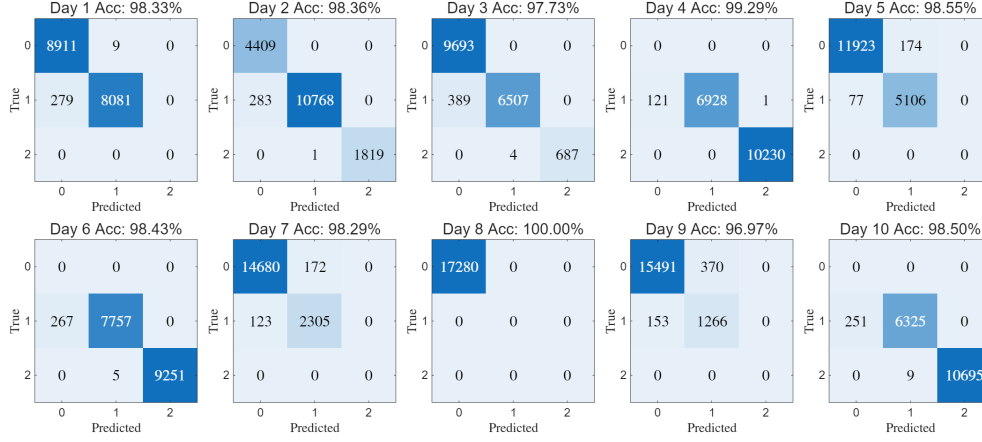


FIG. 3. Confusion matrix comparing true and predicted labels for one person in 10 days.

- **F1-score:** Provides a harmonic mean of precision and recall

$$F1(i) = \frac{2 \cdot \text{Precision}_i \cdot \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i}.$$

5.4.2. Individual-wise Performance. Figure 3 presents the confusion matrices for one person across ten testing days, offering a detailed view of the DSVI-O framework’s performance at the subject level. Accuracy remains consistently above 96% throughout. Most errors occur when the true label is *weak* (1) but predicted as *healthy* (0), suggesting that the model is more conservative in identifying early or mild degradations. On Day 9, the main misclassification pattern shifts, with *healthy* (0) more often predicted as *weak* (1), possibly due to temporal variations in physiological signals. While the model reliably captures major health trends, day-specific dynamics may influence its sensitivity to borderline states, especially during transitions.

Figure 4 illustrates the health state trajectories $\tilde{\mathbf{x}}_t^j$ in (5.6) and the real-time predictions $\mathbf{x}_t^j$ in (5.4) for one individual over 10 representative days in the test set. For each day, only a local segment containing a state transition is shown. The horizontal axis denotes time, allowing for fine-grained comparison between the true and predicted health states. From these trajectories, we observe that the model responds well to changes in physiological signals and accurately captures transitions among *healthy*, *weak*, and *ill* states. In general, the model tracks health trends effectively, with most predicted trajectories aligning well with ground truth. Sudden state changes are typically detected in real time. In contrast, for gradual transitions, the model often fails to react immediately when the true state begins to shift but successfully identifies the change once the deterioration accumulates. This behavior reflects the model’s reliance on temporal evidence aggregation when detecting slow-onset degradations.

5.4.3. Overall Performance. Figure 5 presents the aggregated confusion matrix for all 10 individuals in the test set, offering a comprehensive summary of the model’s classification performance across the three health states—*healthy* (0), *weak* (1), and *ill* (2). The diagonal entries indicate correctly classified instances, while off-diagonal elements represent misclassifications. Over a total of 1,728,000 time points (10 individuals over 100 days), the model achieves an overall accuracy of 98.18%.

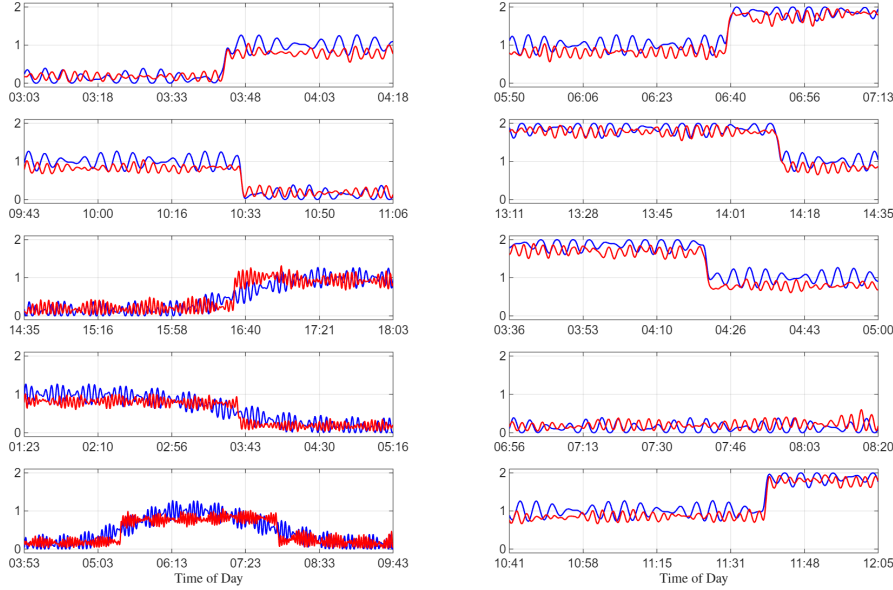


FIG. 4. Tracked health trajectories for one person: estimated (red) $\hat{\mathbf{x}}_\nu^j$ versus true (blue) $\tilde{\mathbf{x}}_\nu^j$.

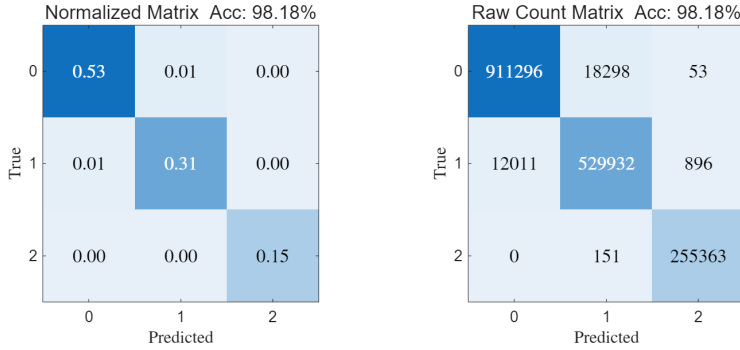


FIG. 5. Confusion matrix of true and predicted labels for ten persons.

Among all predictions, only 53 instances of class *healthy* (0) are incorrectly predicted as *ill* (2), and no instances of class *ill* (2) are misclassified as *healthy* (0). Since these two classes lie at opposite ends of the health spectrum with *weak* (1) in between, this result highlights the model's robustness in distinguishing distant health conditions and avoiding severe misclassifications.

Table 1 reports the class-wise evaluation metrics for the three health states: *healthy* (0), *weak* (1), and *ill* (2). The results demonstrate that the model achieves strong performance across all classes, with overall accuracy above 98% for each. In particular, the model performs exceptionally well in identifying the *ill* state, achieving a Precision of 0.9963, Recall of 0.9994, and an F1-score of 0.9979. This indicates both high sensitivity and specificity in detecting critical health deterioration. The *healthy* class also shows robust metrics, with an F1-score of 0.9836. In contrast, the *weak* class exhibits slightly lower Precision (0.9664) and Recall (0.9762), suggesting that transitional or mild health changes may be more difficult to detect. These observa-

tions highlight a potential direction for model refinement, particularly in improving sensitivity to intermediate health states.

TABLE 1
Class-wise evaluation results for ten persons

Type (i)	Accuracy	Precision	Recall	Specificity	F1-Score
healthy (0)	0.9824	0.9870	0.9803	0.9850	0.9836
weak (1)	0.9819	0.9664	0.9762	0.9844	0.9713
ill (2)	0.9994	0.9963	0.9994	0.9994	0.9979

6. Conclusions. This paper systematically studies fundamental solution properties of the differential stochastic variational inequality with parametric convex optimization (DSVI-O) (1.1)-(1.2). The DSVI-O extends (DSVI), (DVI), (SVI) and (OCODE) by possessing the time-dependent distributions and handling data from different sources in different forms simultaneously. We prove that the DSVI-O has at least one weak solution under standard assumptions. Moreover, we prove that the DSVI-O has at least one classical solution when the objective functions in the parametric convex optimization problems (1.2) are strongly convex. The uniqueness of the classical solution of the DSVI-O is shown in Remark 2.6 for a special class of strongly convex unconstrained parametric optimization problems in (1.2). To find a numerical solution, we propose a discrete scheme of DSVI-O by using a time-stepping approximation and the SAA. We prove the convergence of the discrete scheme, and establish an explicit Euler discretization error bound with $O(h)$ order on the time grid. We apply the DSVI-O with the discrete scheme to an embodied intelligence system for elderly health and generate synthetic health care data by Multimodal Large Language Models to test the performance. Preliminary numerical results demonstrate the efficiency and effectiveness of the DSVI-O. We will study (1.1)-(1.2) with non-convex objective functions f_i from neural network training in the further research. Such extension needs new theory and algorithms for handling parametric nonconvex optimization problems in a dynamic and stochastic system.

Acknowledgments. We are grateful to Prof. Terry Rockafellar for his helpful discussions on solution properties of parametric convex optimization problems and for his comments on the Lipschitz continuity of strongly convex unconstrained optimization problems in Remark 2.6. We also sincerely thank the Associate Editor and the referees for their careful reading and constructive comments, which helped improve the quality of this paper.

REFERENCES

- [1] Z. ARTSTEIN, *Parametrized integration of multifunctions with applications to control and optimization*, SIAM J. Control Optim., 27 (1989), pp. 1369–1380, <https://doi.org/10.1137/0327070>.
- [2] J.-P. AUBIN AND A. CELLINA, *Differential Inclusions*, Springer, 1984.
- [3] J.-P. AUBIN AND H. FRANKOWSKA, *Set-valued Analysis*, Springer, 1999.
- [4] A. BACCIOTTI AND L. ROSIER, *Liapunov Functions and Stability in Control Theory*, Springer, Berlin, Heidelberg, 2nd ed., 2005.
- [5] A. BECK AND M. TEOULLE, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), pp. 183–202, <https://doi.org/10.1137/080716542>.
- [6] R. BENEDUCI, *Positive operator valued measures and Feller Markov kernels*, J. Math. Anal. Appl., 442 (2016), pp. 50–71.

- [7] M. BOUNKHEL, *Existence results of nonconvex differential inclusions*, Port. Math., 59 (2002), pp. 283–310.
- [8] J. V. BURKE AND M. C. FERRIS, *Weak sharp minima in mathematical programming*, SIAM J. Control Optim., 31 (1993), pp. 1340–1359, <https://doi.org/10.1137/0331063>.
- [9] M. K. CAMLIBEL, J.-S. PANG, AND J. SHEN, *Lyapunov stability of complementarity and extended systems*, SIAM J. Optim., 17 (2007), pp. 1056–1101.
- [10] X. CAO, J. ZHANG, AND H. V. POOR, *Online stochastic optimization with time-varying distributions*, IEEE Trans. Automat. Control, 66 (2021), pp. 1840–1847.
- [11] L. CHEN, Y. LIU, X. YANG, AND J. ZHANG, *Stochastic approximation methods for the two-stage stochastic linear complementarity problem*, SIAM J. Optim., 32 (2022), pp. 2129–2155.
- [12] X. CHEN, T. PONG, AND R.-B. WETS, *Two-stage stochastic variational inequalities: an ERM-solution procedure*, Math. Program., 165 (2017), pp. 71–112.
- [13] X. CHEN, A. SHAPIRO, AND H. SUN, *Convergence analysis of sample average approximation of two-stage stochastic generalized equations*, SIAM J. Optim., 29 (2019), pp. 135–161.
- [14] X. CHEN AND J. SHEN, *Dynamic stochastic variational inequalities and convergence of discrete approximation*, SIAM J. Optim., 32 (2022), pp. 2909–2937.
- [15] X. CHEN AND Z. WANG, *Convergence of regularized time-stepping methods for differential variational inequalities*, SIAM J. Optim., 23 (2013), pp. 1647–1671.
- [16] X. CHEN AND Z. WANG, *Differential inequality approaches to dynamic games with shared constraints*, Math. Program., 146 (2014), pp. 379–408.
- [17] X. CHEN AND S. XIANG, *Implicit solution function of P_0 and Z matrix linear complementarity constraints*, Math. Program., 128 (2011), pp. 1–18.
- [18] X. CHEN AND S. XIANG, *Newton iterations in implicit time-stepping scheme for differential linear complementarity systems*, Math. Program., 138 (2013), pp. 579–606.
- [19] T. M. COVER AND J. A. THOMAS, *Elements of Information Theory*, Wiley, Hoboken, NJ, 2nd ed., 2006.
- [20] J. CUTLER, D. DRUSVYATSKIY, AND Z. HARCHAOU, *Stochastic optimization under distributional drift*, J. Mach. Learn. Res., 24 (2023), pp. 1–56.
- [21] K. DEIMLING, *Multivalued Differential Equations*, vol. 1, Walter de Gruyter, 2011.
- [22] T. DONCHEV AND E. FARKHI, *Stability and Euler approximation of one-sided lipschitz differential inclusions*, SIAM J. Control Optim., 36 (1998), pp. 780–796.
- [23] A. DONTCHEV AND F. LEMPIO, *Difference methods for differential inclusions: A survey*, SIAM Rev., 34 (1992), pp. 263–294.
- [24] A. L. DONTCHEV AND E. M. FARKHI, *Error estimates for discretized differential inclusions*, Computing, 41 (1989), pp. 349–358, <https://doi.org/10.1007/BF02241223>.
- [25] A. L. DONTCHEV AND R. T. ROCKAFELLAR, *Implicit Functions and Solution Mappings*, vol. 543, Springer, 2009.
- [26] R. DOUC, E. MOULINES, P. PRIOURET, AND P. SOULIER, *Markov Chains*, Springer, 1 ed., 2018.
- [27] M. FUKUSHIMA, *Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems*, Math. Program., 53 (1992), pp. 99–110.
- [28] M. GIUFFRÈ AND D. L. SHUNG, *Harnessing the power of synthetic data in healthcare: innovation, application, and privacy*, NPJ Digit. Med., 6 (2023), p. 186.
- [29] J. K. HALE, *Ordinary Differential Equations*, Courier Corporation, 1980.
- [30] C. HALMICH, L. HÖSCHLER, C. SCHRANZ, AND C. BORGELT, *Data augmentation of time-series data in human movement biomechanics: A scoping review*, PloS One, 20 (2025), p. e0327038.
- [31] S. HAN AND J.-S. PANG, *Continuous selections of solutions to parametric variational inequalities*, SIAM J. Optim., 34 (2024), pp. 870–892.
- [32] M. HOSSIN AND M. N. SULAIMAN, *A review on evaluation metrics for data classification evaluations*, Int. J. Data Min. Knowl. Manag. Process, 5 (2015), p. 1.
- [33] J. JORDON, L. SZPRUCH, F. HOUSIAU, M. BOTTARELLI, G. CHERUBIN, C. MAPLE, S. N. COHEN, AND A. WELLER, *Synthetic data—what, why and how?*, arXiv preprint arXiv:2205.03257, (2022).
- [34] I. F. JØRGENSEN, A. D. HAUE, D. PLACIDO, J. X. HJALTELIN, AND S. BRUNAK, *Disease trajectories from healthcare data: methodologies, key results, and future perspectives*, Annu. Rev. Biomed. Data Sci., 7 (2024), pp. 251–276.
- [35] O. KALLENBERG, *Foundations of Modern Probability*, vol. 2, Springer, 1997.
- [36] D. W. KIM, C. MAYER, M. P. LEE, S. W. CHOI, M. TEWARI, AND D. B. FORGER, *Efficient assessment of real-world dynamics of circadian rhythms in heart rate and body temperature from wearable data*, J. R. Soc. Interface, 20 (2023), p. 20230030.
- [37] C. LANDRY, A. CABOUSSAT, AND E. HAIRER, *Solving optimization-constrained differential equations with discontinuity points, with application to atmospheric chemistry*, SIAM J. Sci.

- Comput., 31 (2009), pp. 3806–3826.
- [38] H. LIU, D. GUO, AND A. CANGELOSI, *Embodied intelligence: A synergy of morphology, action, perception and learning*, ACM Comput. Surv., 57 (2025), pp. 186:1–186:36.
- [39] J. LUO AND X. CHEN, *Dynamic systems coupled with solutions of stochastic nonsmooth convex optimization*, SIAM J. Control Optim., 63 (2025), pp. 1686–1708.
- [40] Z. R. MCCAW, J. H. GAO, X. H. LIN, AND J. GRONSBELL, *Synthetic surrogates improve power for genome-wide association studies of partially missing phenotypes in population biobanks*, Nat. Genet., 56 (2024), pp. 1527–1536.
- [41] I. MOLCHANOV, *Theory of Random Sets*, vol. 19, Springer, 2017.
- [42] J.-S. PANG, *Three modeling paradigms in mathematical programming*, Math. Program., 125 (2010), pp. 297–323.
- [43] J.-S. PANG AND D. E. STEWART, *Differential variational inequalities*, Math. Program., 113 (2008), pp. 345–424.
- [44] R. T. ROCKAFELLAR AND R.-B. WETS, *Stochastic variational inequalities: single-stage to multistage*, Math. Program., 165 (2017), pp. 331–360.
- [45] R. T. ROCKAFELLAR AND R. J.-B. WETS, *Variational Analysis*, vol. 317, Springer Science & Business Media, 2009.
- [46] K. ROCKWOOD AND A. MITNITSKI, *Frailty in relation to the accumulation of deficits*, J. Gerontol. A Biol. Sci. Med. Sci., 62 (2007), pp. 722–727.
- [47] M. RUJAS, R. M. G. D. M. HERRANZ, G. FICO, AND B. MERINO-BARBANCHO, *Synthetic data generation in healthcare: A scoping review of reviews on domains, motivations, and future applications*, Int. J. Med. Inform., 195 (2025), p. 105763, <https://doi.org/10.1016/j.ijmedinf.2024.105763>.
- [48] A. SHAPIRO AND H. XU, *Uniform laws of large numbers for set-valued mappings and subdifferentials of random functions*, J. Math. Anal. Appl., 325 (2007), pp. 1390–1399.
- [49] D. SHEVITZ AND B. PADEN, *Lyapunov stability theory of nonsmooth systems*, IEEE Trans. Autom. Control, 39 (1994), pp. 1910–1914.
- [50] Y. TERAZONO AND A. MATANI, *Continuity of optimal solution functions and their conditions on objective functions*, SIAM J. Optim., 25 (2015), pp. 2050–2060.
- [51] A. B. TSYBAKOV, *Introduction to Nonparametric Estimation*, Springer, New York, 2009.
- [52] N. C. YANNELIS, *On the upper and lower semicontinuity of the Aumann integral*, J. Math. Econ., 19 (1990), pp. 373–389.