DYNAMIC SYSTEMS COUPLED WITH SOLUTIONS OF STOCHASTIC NONSMOOTH CONVEX OPTIMIZATION *

JIANFENG LUO† AND XIAOJUN CHEN‡

Abstract. In this paper, we study ordinary differential equations (ODE) coupled with solutions of a stochastic nonsmooth convex optimization problem (SNCOP). We use the regularization approach, the sample average approximation and the time-stepping method to construct discrete approximation problems. We show the existence of solutions to the original problem and the discrete problems. Moreover, we show that the optimal solution of the SNCOP with a strong convex objective function admits a linear growth condition and the optimal solution of the regularized SNCOP converges to the least-norm solution of the original SNCOP, which are crucial for us to derive the convergence results of the discrete problems. We illustrate the theoretical results and applications for the estimation of the time-varying parameters in ODE by numerical examples.

Key words. Dynamic system, stochastic nonsmooth optimization, regularization method, sample average approximation, convergence analysis.

MSC codes. 90C15, 90C33, 90C39

1. Introduction.

Let ξ be a random variable defined in the probability space $(\Omega, \mathcal{F}, \mathcal{P})$ with support set $\Xi := \xi(\Omega) \subseteq \mathbb{R}^d$. Let $f : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \times \Xi \to \mathbb{R}^n$, $g : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \times \Xi \to \mathbb{R}$, $A : \Xi \to \mathbb{R}^{q\times n}$, $B : \Xi \to \mathbb{R}^{q\times m}$ and $Q : \mathbb{R} \times \Xi \to \mathbb{R}^q$ be given mappings. In this paper, we consider the following dynamic system coupled with solutions of stochastic nonsmooth convex optimization:

$$\dot{x}(t) = \mathbb{E}[f(t, x(t), y(t), \xi)], \;\; x(0) = x_0, \tag{1.1}$$

$$y(t) \in \arg\min_{\mathbf{y}\in\mathbb{R}^m} \mathbb{E}[g(t, x(t), \mathbf{y}, \xi)] \quad \text{s.t.} \;\; \mathbf{y} \in K(t, x(t)), \tag{1.2}$$

where $x_0 \in \mathbb{R}^n$ is an initial vector, and the set-valued function $K : \mathbb{R}_+ \times \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is defined as below

$$K(t, x(t)) \triangleq \{\mathbf{y} \in \mathbb{R}^m : \mathbb{E}[A(\xi)]x(t) + \mathbb{E}[B(\xi)]\mathbf{y} + \mathbb{E}[Q(t, \xi)] \le 0\}.$$

We assume that all expected values in problem (1.1)-(1.2) are well defined.

Let $\|\cdot\|$ denote the Euclidean norm of a vector and a matrix. We suppose that there exists a measurable function $\kappa_f : \Xi \to \mathbb{R}_+$ with $\mathbb{E}[\kappa_f(\xi)] < \infty$ such that for any $t_1,\, t_2 \in \mathbb{R}_+$, $u_1,\, u_2 \in \mathbb{R}^n$, $v_1,\, v_2 \in \mathbb{R}^m$ and almost everywhere (a.e.) $\xi \in \Xi$,

$$\|f(t_1, u_1, v_1, \xi) - f(t_2, u_2, v_2, \xi)\| \le \kappa_f(\xi)(|t_1 - t_2| + \|u_1 - u_2\| + \|v_1 - v_2\|). \tag{1.3}$$

We also assume that $g(t, x(t), \cdot, \xi)$ is convex for any $t \in \mathbb{R}_+$, $x(t) \in \mathbb{R}^n$ and a.e. $\xi \in \Xi$, the functions $g(\cdot, \cdot, \cdot, \xi)$ and $Q(\cdot, \xi)$ are both continuous for a.e. $\xi \in \Xi$, and there exists a measurable function $\kappa_Q : \Xi \to \mathbb{R}_+$ with $\mathbb{E}[\kappa_Q(\xi)] < \infty$ such that $\|Q(t, \xi)\| \le \kappa_Q(\xi)t$

*1 September 2023, revisions 9 March 2024, 14 August 2024, 24 January 2025,

Funding: This work is supported by CAS-Croucher Funding Scheme for the CAS AMSS-PolyU Joint Laboratory in Applied Mathematics and Hong Kong Research Grant Council project PolyU15300022.

†School of Mathematics, North University of China, Taiyuan, China; CAS AMSS-PolyU Joint Laboratory of Applied Mathematics (Shenzhen), The Hong Kong Polytechnic University Shenzhen Research Institute, Shenzhen, China (luojf2024@163.com).

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

for any $t \in \mathbb{R}_+$ and a.e. $\xi \in \Xi$. Assume that $g(\cdot,\cdot,\cdot,\xi)$ is dominated by an integrable function for a.e. $\xi \in \Xi$. Then we know $\mathbb{E}[f(t,x(t),y(t),\xi)]$ is Lipschitz continuous and $\mathbb{E}[g(t,x(t),y(t),\xi)]$ is continuous with respect to (w.r.t.) $(t,x(t),y(t))$.

The optimization problem (1.2) is a stochastic convex program for any fixed $t \in \mathbb{R}_+$ and $x(t) \in \mathbb{R}^n$, since the objective function $\mathbb{E}[g(t,x(t),\cdot,\xi)]$ is convex and the feasible set $K(t,x(t))$ is a convex set. The objective function $\mathbb{E}[g(t,x(t),\cdot,\xi)]$ is not necessarily differentiable and the solution set of (1.2) may have multiple elements. Problem (1.1)-(1.2) can be equivalently written as the following dynamic generalized stochastic variational inequality:

$$\begin{cases} \dot{x}(t) = \mathbb{E}[f(t,x(t),y(t),\xi)], \quad x(0)=x_0, \\ 0 \in \partial_{y(t)}\mathbb{E}[g(t,x(t),y(t),\xi)] + \mathcal{N}_{K(t,x(t))}(y(t)), \end{cases} \tag{1.4}$$

where $\partial_{y(t)}\mathbb{E}[g(t,x(t),y(t),\xi)]$ is the subdifferential of $\mathbb{E}[g(t,x(t),y(t),\xi)]$ at the point $y(t)$ and $\mathcal{N}_{K(t,x(t))}(y(t))$ denotes the normal cone of $K(t,x(t))$ at $y(t)$ [24]. When the function $\mathbb{E}[g(t,x(t),\cdot,\xi)]$ is continuously differentiable, we can derive a differential stochastic variational inequality (DSVI):

$$\begin{cases} \dot{x}(t) = \mathbb{E}[f(t,x(t),y(t),\xi)], \quad x(0)=x_0, \\ 0 \in \nabla_{y(t)}\mathbb{E}[g(t,x(t),y(t),\xi)] + \mathcal{N}_{K(t,x(t))}(y(t)). \end{cases} \tag{1.5}$$

It is easy to see that problem (1.5) is a special case of (1.4) and problem (1.1)-(1.2). The DSVI (1.5) includes the deterministic differential variational inequality (DVI), which has many important applications in engineering, economics and biology. The DVI involves dynamics, variational inequalities and equilibrium conditions, and has been studied in [3, 8, 9, 10, 22, 28, 30, 31].

It should be noted that, when the variational inequality (VI) or optimization problem (1.2) has multiple solutions, a wrong selection of solutions may make the corresponding ODE unsolvable or numerical scheme divergent. The authors in [14] proposed to use the least-norm solution of the VI to ensure the convergence of time-stepping method for a special class of monotone DVI, which inspires our regularization approach for problem (1.1)-(1.2) in this paper.

The dynamic systems coupled with solutions of an optimization problem have a wide applications in many fields such as atmospheric chemistry [4, 16] and dynamic flux balance analysis in biological systems [33]. They have been extended to stochastic case in [23], where the authors investigated a dynamic flux balance analysis model with uncertainty. As mentioned in [16], the deterministic dynamic systems coupled with solutions of a convex optimization problem can be seen as an ODE-constrained optimization problem, which is proposed to estimate parameters for the ODE in [11]. In [16], the authors proposed a numerical method for the differential equations coupled with a smooth nonconvex optimization problem and applied the Karush-Kuhn-Tucker conditions to reformulate the problem as the DVI. However, as we mentioned before, when the objective function is nonsmooth, we cannot transform problem (1.1)-(1.2) as the DVI and apply the existing methods and results. Therefore, we present the existence of solutions, numerical methods, convergence analysis and applications of (1.1)-(1.2) in this paper.

The main contributions of this paper are twofold. (i) We give sufficient conditions for the existence of a solution (x,y) of problem (1.1)-(1.2) on $[0,T]$, where x is absolutely continuous and y is integrable. In addition, if the objective function in (1.2) is strongly convex, then problem (1.1)-(1.2) has a solution (x,y) over $[0,\tilde{T}]$

with x being continuously differentiable, and y being continuous for a positive number $\tilde{T}$ and admitting a linear growth condition. (ii) We propose a regularization method to approximate the objective function in (1.2) by a strongly convex function and show the unique optimal solution of the regularized optimization problem converges to the least-norm optimal solution of (1.2) when the regularization parameter goes to zero. Moreover, we prove the existence of solutions to the discrete regularization problem using the sample average approximation (SAA) and the implicit Euler time-stepping scheme. We show the solution of the approximation problem constructed by the regularization approach, SAA and time-stepping method converges to a solution of (1.1)-(1.2) with probability 1 (w.p.1) by the repeated limits in the order of the regularization parameter goes to zero, the SAA sample size goes to infinity and the time-stepping step size goes to zero.

The paper is organised as follows: Section 2 deals with the existence of solutions of problem (1.1)-(1.2). Section 3 studies the existence of solutions of the regularized problem of (1.1)-(1.2) and the convergence to the original problem as the regularization parameter approaches to zero. In Section 4, we present the existence of solutions of the SAA of (1.1)-(1.2) and the convergence analysis. In Section 5, we study the convergence of the time-stepping scheme and show the convergence properties of the discrete method using the SAA and the implicit Euler time-stepping scheme. Section 6 gives a numerical example to illustrate the theoretical results obtained in this paper. And Section 7 shows the application of the estimation of the time-varying parameters in ODE. Some final conclusion remarks are presented in Section 8.

1.1. Notation. Denote by $\mathcal{B}(v,r)$ the open ball centered by $v \in \mathbb{R}^n$ with the radius of r in the Euclidean norm. For sets $S_1, S_2 \subseteq \mathbb{R}^n$, we denote the distance from $v \in \mathbb{R}^n$ to S_1 and the deviation of the set S_1 from the set S_2 by $\mathrm{dist}(v, S_1) = \inf_{v' \in S_1} \|v - v'\|$, and $\mathbb{D}(S_1, S_2) = \sup_{v \in S_1} \mathrm{dist}(v, S_2)$, respectively. We also define the Hausdorff distance between the set S_1 and the set S_2 by $\mathbb{H}(S_1, S_2) = \max\{\mathbb{D}(S_1, S_2), \mathbb{D}(S_2, S_1)\}$. We define $S_1 + S_2 = \{z_1 + z_2 : z_1 \in S_1, z_2 \in S_2\}$. For a set S, $\mathrm{int}S$ denotes the interior of S and $\tau S = \{\tau z : z \in S\}$ with a scalar τ. Let $C^1([a,b])$ and $C^0([a,b])$ be the spaces of continuously differentiable vector-valued functions and continuous vector-valued functions on $[a,b]$, respectively.

2. Existence of solutions.

In this section, we show the existence of solutions to problem (1.1)-(1.2).

DEFINITION 2.1. *[12, 26]*

(i) (lower semicontinuity). A set-valued mapping $\mathcal{S} : \mathbb{R}^{n_1} \rightrightarrows \mathbb{R}^{m_1}$ is lower semicontinuous at $\bar{z} \in \mathbb{R}^{n_1}$ if for any open set $\mathcal{B}_{\mathcal{S}}$ with $\mathcal{B}_{\mathcal{S}} \cap \mathcal{S}(\bar{z}) \neq \emptyset$, there exists $\sigma > 0$ such that $\mathcal{S}(z) \cap \mathcal{B}_{\mathcal{S}} \neq \emptyset$ for any $z \in \mathcal{B}(\bar{z}, \sigma)$.

(ii) (upper semicontinuity). A set-valued mapping $\mathcal{S} : \mathbb{R}^{n_1} \rightrightarrows \mathbb{R}^{m_1}$ is upper semicontinuous at $\bar{z} \in \mathbb{R}^{n_1}$ if for any open set $\mathcal{B}_{\mathcal{S}}$ with $\mathcal{S}(\bar{z}) \subseteq \mathcal{B}_{\mathcal{S}}$, there exists $\sigma > 0$ such that $\mathcal{S}(z) \subseteq \mathcal{B}_{\mathcal{S}}$ for any $z \in \mathcal{B}(\bar{z}, \sigma)$.

A set-valued mapping $\mathcal{S}$ is said to be continuous if and only if it is both upper and lower semicontinuous. Obviously, upper (lower) semicontinuity is nothing else than continuity if $\mathcal{S}$ is single-valued.

Let $\mathcal{S}(t, x(t))$ denote the optimal solution set of (1.2) for fixed $t \in \mathbb{R}_+$ and $x(t) \in \mathbb{R}^n$. For some $T > 0$, if there exists $(x, y) \in C^1([0,T]) \times C^0([0,T])$ fulfilling problem (1.1)-(1.2), we call (x, y) a classic solution of problem (1.1)-(1.2) on $[0,T]$. We call (x, y) a weak solution of problem (1.1)-(1.2) over $[0,T]$ if x is absolutely continuous and y is integrable over $[0,T]$ with $y(t)\in\mathcal{S}(t,x(t))$ and

$$x(t)=x_0+\int_0^t \mathbb{E}[f(\tau,x(\tau),y(\tau),\xi)]d\tau.$$

We first show the existence of solutions of problem (1.1)-(1.2) under the following assumption.

Assumption 2.2. The set $\mathrm{int}K(0,x_0)$ is not empty, the function $\mathbb{E}[g(0,x_0,\cdot,\xi)]$ is level-bounded over $K(0,x_0)$ (i.e. all sets $\{\mathbf{y}\in K(0,x_0):\mathbb{E}[g(0,x_0,\mathbf{y},\xi)]\le\alpha\}$ for $\alpha\in\mathbb{R}$ are bounded) and the function $f(t,\mathbf{x},\cdot,\xi)$ is affine for any $t\in\mathbb{R}$, $\mathbf{x}\in\mathbb{R}^n$ and a.e. $\xi\in\Xi$.

THEOREM 2.3. *Suppose that Assumption 2.2 holds. Then there exists $T_0>0$ such that problem (1.1)-(1.2) has at least a weak solution (x^*,y^*) on $[0,T]$ for any $T\le T_0$.*

Proof. Let $K_1(\mathbf{x},\mathbf{q})=\{\mathbf{y}\in\mathbb{R}^m:\mathbb{E}[A(\xi)]\mathbf{x}+\mathbb{E}[B(\xi)]\mathbf{y}+\mathbf{q}\le 0\}$ for $\mathbf{x}\in\mathbb{R}^n$ and $\mathbf{q}\in\mathbb{R}^q$. Denote $\mathbf{q}_0=\mathbb{E}[Q(0,\xi)]$. It is easy to see $K_1(x_0,\mathbf{q}_0)=K(0,x_0)$. Since $\mathrm{int}K_1(x_0,\mathbf{q}_0)\ne\emptyset$, there are $\breve{\sigma}>0$ and $\breve{\delta}>0$ such that $\mathrm{int}K_1(\mathbf{x},\mathbf{q})\ne\emptyset$ for any $(\mathbf{x},\mathbf{q})\in\mathcal{B}(x_0,\breve{\sigma})\times\mathcal{B}(\mathbf{q}_0,\breve{\delta})$. By the continuity of $\mathbb{E}[Q(\cdot,\xi)]$, we conclude that there are $\hat{\sigma}>0$ and $\hat{\delta}>0$ such that $\mathrm{int}K(t,x(t))\ne\emptyset$ for any $(t,x(t))\in[0,\hat{\sigma}]\times\mathcal{B}(x_0,\hat{\delta})$ with $x\in C^0([0,\hat{\sigma}])$.

It is easy to verify that $K_1(\tau\mathbf{x}_1+(1-\tau)\mathbf{x}_2,\tau\mathbf{q}_1+(1-\tau)\mathbf{q}_2)\supset\tau K_1(\mathbf{x}_1,\mathbf{q}_1)+(1-\tau)K_1(\mathbf{x}_2,\mathbf{q}_2)$ for $\tau\in(0,1)$, $\mathbf{x}_1,\,\mathbf{x}_2\in\mathcal{B}(x_0,\breve{\sigma})$ and $\mathbf{q}_1,\,\mathbf{q}_2\in\mathcal{B}(\mathbf{q}_0,\breve{\delta})$, which means that K_1 is graph-convex. Then following from [24, Corollary 9.34], K_1 is strictly continuous at any point of $\mathcal{B}(x_0,\breve{\sigma})\times\mathcal{B}(\mathbf{q}_0,\breve{\delta})$.

Therefore, by the continuity of $\mathbb{E}[g(\cdot,\cdot,\cdot,\xi)]$ and Assumption 2.2, there are two scalars σ and δ with $\hat{\sigma}\ge\sigma>0$ and $\hat{\delta}\ge\delta>0$ such that K is continuous over $[0,\sigma]\times\mathcal{B}(x_0,\delta)$ and $\mathbb{E}[g(t,x(t),\cdot,\xi)]$ is level-bounded over $K(t,x(t))$ for any $(t,x(t))\in[0,\sigma]\times\mathcal{B}(x_0,\delta)$ with $x\in C^0([0,\sigma])$. According to [24, Example 1.11], we know that $\mathcal{S}(t,x(t))$ is nonempty and compact for any $(t,x(t))\in[0,\sigma]\times\mathcal{B}(x_0,\delta)$ with $x\in C^0([0,\sigma])$. It then derives that, by [26, Theorem 3.1], $\mathcal{S}$ is convex-valued and upper semicontinuous over $[0,\sigma]\times\mathcal{B}(x_0,\delta)$, which means that there exists $\rho_s>0$ (independent of $(t,x(t))$) such that $\sup\{\|\mathbf{y}\|:\mathbf{y}\in\mathcal{S}(t,x(t))\}\le\rho_s$ over $[0,\sigma]\times\mathcal{B}(x_0,\delta)$.

Since $\mathcal{S}$ is convex-valued and $\mathbb{E}[f(t,x(t),\cdot,\xi)]$ is affine for $t\in\mathbb{R}$ and $x(t)\in\mathbb{R}^n$, the set-valued mapping

$$f_{\mathcal{S}}(t,x(t))=\{\mathbb{E}[f(t,x(t),\mathbf{y},\xi)]:\mathbf{y}\in\mathcal{S}(t,x(t))\}$$

is convex-valued. Following from the Lipschitz property of $\mathbb{E}[f(\cdot,\cdot,\cdot,\xi)]$, we know that there exists $\breve{\rho}_f>0$ such that $\|\mathbb{E}[f(t,x(t),y(t),\xi)]\|\le\breve{\rho}_f(1+|t|+\|x(t)\|+\|y(t)\|)$. Therefore, there exists $\rho_f>0$ such that the following linear growth condition holds for any $(t,x(t))\in[0,\sigma]\times\mathcal{B}(x_0,\delta)$

$$\sup\{\|\mathbb{E}[f(t,x(t),y(t),\xi)]\|:y(t)\in\mathcal{S}(t,x(t))\}\le\rho_f(1+\|x(t)\|). \tag{2.1}$$

It is obvious that $f_{\mathcal{S}}$ is closed by the compactness of $\mathcal{S}$. We then know that $f_{\mathcal{S}}$ is upper semicontinuous over $(t,x(t))\in[0,\sigma]\times\mathcal{B}(x_0,\delta)$ since it is bounded on compact sets [1, Corollary 1 in Section 1 of Chapter 1].

According to [12, Theorem 5.1] and [22, Lemma 6.1], we know that the following differential inclusion

$$\begin{cases} \dot{x}(t) \in f_{\mathcal{S}}(t, x(t)), \\ x(0) = x_0, \end{cases}$$

has at least one absolutely continuous solution x^*. According to [1, Corollary 1 in Section 14 of Chapter 1] and [22, Lemma 6.3], there exists an integrable function $y^*(t) \in \mathcal{S}(t, x^*(t))$ such that

$$x^*(t) = x_0 + \int_0^t \mathbb{E}[f(\tau, x^*(\tau), y^*(\tau), \xi)]d\tau.$$

Clearly, there exists $\sigma_T > 0$ such that $x^*(t) \in \mathcal{B}(x_0, \delta)$ for any $t \in [0, \sigma_T]$. By choosing $T_0 = \min\{\sigma_T, \sigma\}$, we can conclude the result. ∎

If optimization problem (1.2) has equality constraints with $K(t, x(t)) = \{\mathbf{y} \in \mathbb{R}^m : \mathbb{E}[A(\xi)]x(t) + \mathbb{E}[B(\xi)]\mathbf{y} + \mathbb{E}[Q(t, \xi)] = 0\}$, we can replace "int$K(0, x_0)$ is not empty" by "$K(0, x_0)$ is not empty" in Assumption 2.2 and consider the relaxation set $K(t, x(t), \epsilon) \triangleq \{\mathbf{y} \in \mathbb{R}^m : \|\mathbb{E}[A(\xi)]x(t) + \mathbb{E}[B(\xi)]\mathbf{y} + \mathbb{E}[Q(t, \xi)]\|_\infty \le \epsilon\}$, where $\epsilon \ge 0$ is a scalar. Since $K(0, x_0)$ is not empty, we have int$K(0, x_0, \epsilon)$ with $\epsilon > 0$ is not empty. Let $\tilde{K}(\epsilon) = K(0, x_0, \epsilon)$. It is easy to see that $\tilde{K}(\epsilon)$ is graph-convex and the graph $\tilde{K}(\epsilon)$ is polyhedral, which implies from [24, Example 9.35] that $\tilde{K}(\epsilon)$ is Lipschitz continuous w.r.t. ϵ. Therefore, from the function $\mathbb{E}[g(0, x_0, \cdot, \xi)]$ is level-bounded over $K(0, x_0)$, we have that $\mathbb{E}[g(0, x_0, \cdot, \xi)]$ is level-bounded over $K(0, x_0, \epsilon)$ with any sufficiently small $\epsilon > 0$, which means that Assumption 2.2 holds to the relaxation optimization problem with int$K(0, x_0, \epsilon) \neq \emptyset$ for any sufficiently small $\epsilon > 0$.

It is obvious that $K(t, x(t), \epsilon)$ is also Lipschitz continuous w.r.t. ϵ with any given t and $x(t)$. It means from [24, Definition 9.26, Corollary 4.7] that $K(t, x(t), \epsilon) \downarrow K(t, x(t))$ as $\epsilon \downarrow 0$. Moreover, from [24, Proposition 7.4(f), Exercise 7.8(a)], we know that $\mathbb{E}[g(t, x(t), \cdot, \xi)] + I_{K(t,x(t),\epsilon)} \to^{epi} \mathbb{E}[g(t, x(t), \cdot, \xi)] + I_{K(t,x(t))}$ as $\epsilon \downarrow 0$, where I_K is the indicator function of set K. It then concludes by [24, Theorem 7.33] that $\lim_{\epsilon\downarrow 0} \mathbb{D}(\mathcal{S}^\epsilon(t, x(t)), \mathcal{S}(t, x(t))) = 0$, where $\mathcal{S}(t, x(t))$ and $\mathcal{S}^\epsilon(t, x(t))$ denote the optimal solution sets of optimization problem (1.2) with equality constraints and its relaxation optimization problem with $\epsilon > 0$, respectively. Hence by using this relaxation method, the results of this paper are also applicable without assume that int$K(t, x(t)) \neq \emptyset$.

2.1. Existence in the strong convex case. In this subsection, we consider a special case of (1.2) where the objective function is strongly convex.

Assumption 2.4. There exists a measurable function $\varrho : \Xi \to \mathbb{R}_{++}$ with $0 < \mathbb{E}[\varrho(\xi)] < \infty$ such that for any $\mathbf{y}_1$, $\mathbf{y}_2 \in \mathbb{R}^m$ and $\tau \in (0, 1)$,

$$g(t, \mathbf{x}, (1-\tau)\mathbf{y}_1 + \tau\mathbf{y}_2, \xi) \le (1-\tau)g(t, \mathbf{x}, \mathbf{y}_1, \xi) + \tau g(t, \mathbf{x}, \mathbf{y}_2, \xi) - \frac{1}{2}\varrho(\xi)\tau(1-\tau)\|\mathbf{y}_1 - \mathbf{y}_2\|^2$$

holds for any fixed $t \in \mathbb{R}_+$, $\mathbf{x} \in \mathbb{R}^n$ and a.e. $\xi \in \Xi$.

Assumption 2.4 means that $g(t, \mathbf{x}, \cdot, \xi)$ is strongly convex for any fixed $t \in \mathbb{R}_+$, $\mathbf{x} \in \mathbb{R}^n$ and a.e. $\xi \in \Xi$ and $\mathbb{E}[g(t, \mathbf{x}, \cdot, \xi)]$ is also strongly convex. Under Assumption 2.4 we have the following result about the existence of solutions of problem (1.1)-(1.2).

THEOREM 2.5. *Suppose that Assumption 2.4 holds and int*$K(0, x_0) \neq \emptyset$. *Then there exists* $\tilde{T} > 0$ *such that problem* (1.1)-(1.2) *has a classic solution* $(\tilde{x}, \tilde{y})$ *on* $[0, \tilde{T}]$. *In addition, there exists* $\rho > 0$ *such that*

$$\|\tilde{y}(t)\| \le \rho(1 + |t| + \|\tilde{x}(t)\|) \quad \text{for} \quad t \in [0, \tilde{T}]. \tag{2.2}$$

Proof. Following the proof of Theorem 2.3, there are $\sigma > 0$ and $\delta > 0$ such that $K(t, x(t))$ is convex and nonempty for any $(t, x(t)) \in [0, \sigma] \times \mathcal{B}(x_0, \delta)$ with $x \in C^0([0, \sigma])$. It then derives the existence of a unique optimal solution $\hat{y}(t, x(t))$ for any $(t, x(t)) \in [0, \sigma] \times \mathcal{B}(x_0, \delta)$ by the strong convexity of $\mathbb{E}[g(t, x(t), \cdot, \xi)]$. In addition, we can obtain that the optimal solution set $\mathcal{S}$ of the optimization problem (1.2) is also upper semicontinuous over $[0, \sigma] \times \mathcal{B}(x_0, \delta)$, which means that $\hat{y}(t, x(t))$ is continuous w.r.t. t and $x(t)$ for any $x \in C^0([0, \sigma])$ since $\mathcal{S}$ is single-valued.

Therefore, applying the Peano existence theorem [27], we find that

$$\begin{cases} \dot{x}(t) = \mathbb{E}[f(t, x(t), \hat{y}(t, x(t)), \xi)], \\ x(0) = x_0, \end{cases} \tag{2.3}$$

has a solution $\tilde{x}(t)$, where $\tilde{x} \in C^1([0, \sigma])$. Write $\tilde{y}(t) = \hat{y}(t, \tilde{x}(t))$ and then $\tilde{y} \in C^0([0, \sigma])$. Noting

$$\tilde{x}(t) = x_0 + \int_0^t \mathbb{E}[f(\tau, \tilde{x}(\tau), \tilde{y}(\tau), \xi)] d\tau.$$

Clearly, there exists $\sigma_{\tilde{T}} > 0$ such that $\tilde{x}(t) \in \mathcal{B}(x_0, \delta)$ for $t \in [0, \sigma_{\tilde{T}}]$, which derives that problem (1.1)-(1.2) has a classic solution $(\tilde{x}, \tilde{y})$ on $[0, \tilde{T}]$ with $\tilde{T} = \min\{\sigma, \sigma_{\tilde{T}}\}$.

Now we prove (2.2). Following from [24, Example 9.14] and the continuity and convexity of $\mathbb{E}[g(t, x(t), \cdot, \xi)]$, $\partial_{y(t)} \mathbb{E}[g(t, x(t), y(t), \xi)]$ is nonempty and compact for any $t \in \mathbb{R}_+$, $x(t) \in \mathbb{R}^n$ and $y(t) \in \mathcal{C}$ with a compact subset $\mathcal{C}$ of $\mathbb{R}^m$. Since $\mathbb{E}[g(t, \tilde{x}(t), \cdot, \xi)]$ is strongly convex, the set-valued mapping $\partial_{y(t)} \mathbb{E}[g(t, \tilde{x}(t), \cdot, \xi)]$ is strongly monotone with constant $\mathbb{E}[\varrho(\xi)]$ over $K(t, \tilde{x}(t))$. To derive (2.2), it suffices to show the linear growth condition of $\hat{y}(t, \tilde{x}(t))$ w.r.t. t and $\tilde{x}(t)$. It is known that $\hat{y}(t, \tilde{x}(t))$ is the unique optimal solution of the optimization problem (1.2) if and only if $(\hat{y}(t, \tilde{x}(t)), \hat{z}(t, \tilde{x}(t)))$ with $\hat{z}(t, \tilde{x}(t)) \in \partial_{y(t)} \mathbb{E}[g(t, \tilde{x}(t), \hat{y}(t, \tilde{x}(t)), \xi)]$ is the unique solution of the generalized variational inequality: find $(\hat{\mathbf{y}}, \hat{\mathbf{z}})$ with $\hat{\mathbf{z}} \in \partial_{y(t)} \mathbb{E}[g(t, \tilde{x}(t), \hat{\mathbf{y}}, \xi)]$ such that $(\mathbf{y} - \hat{\mathbf{y}})^\top \hat{\mathbf{z}} \geq 0$ for any $\mathbf{y} \in K(t, \tilde{x}(t))$.

Note that $K(t, \tilde{x}(t))$ is a polyhedron for any given t and $\tilde{x}(t)$. Let $\tilde{\mathbf{y}}$ be the least-norm element of $K(t, \tilde{x}(t))$. By Hoffman's error bound for linear systems [13, Lemma 3.2.3], we know that there exists $\alpha > 0$ (independent of t) such that $\|\tilde{\mathbf{y}}\| \leq \alpha(1 + |t| + \|\tilde{x}(t)\|)$ for all t and $\tilde{x}(t)$ with $K(t, \tilde{x}(t)) \neq \emptyset$. Let $\tilde{\mathbf{z}} \in \partial_{y(t)} \mathbb{E}[g(t, \tilde{x}(t), \tilde{\mathbf{y}}, \xi)]$, we have

$$0 \leq (\tilde{\mathbf{y}} - \hat{y}(t, \tilde{x}(t)))^\top \hat{z}(t, \tilde{x}(t)).$$

By the strong monotonicity of $\partial_{y(t)} \mathbb{E}[g(t, \tilde{x}(t), \cdot, \xi)]$, we have

$$\begin{aligned} \mathbb{E}[\varrho(\xi)] \|\tilde{\mathbf{y}} - \hat{y}(t, \tilde{x}(t))\|^2 &\leq (\tilde{\mathbf{y}} - \hat{y}(t, \tilde{x}(t)))^\top (\tilde{\mathbf{z}} - \hat{z}(t, \tilde{x}(t))) \\ &\leq (\tilde{\mathbf{y}} - \hat{y}(t, \tilde{x}(t)))^\top \tilde{\mathbf{z}} \leq \|\tilde{\mathbf{y}} - \hat{y}(t, \tilde{x}(t))\| \|\tilde{\mathbf{z}}\|, \end{aligned}$$

which implies that $\|\tilde{\mathbf{y}} - \hat{y}(t, \tilde{x}(t))\| \leq \mathbb{E}[\varrho(\xi)]^{-1} \|\tilde{\mathbf{z}}\|$. By $\|\tilde{\mathbf{y}}\| \leq \alpha(1 + |t| + \|\tilde{x}(t)\|)$ and the boundeness of $\partial_{y(t)} \mathbb{E}[g(t, \tilde{x}(t), \tilde{\mathbf{y}}, \xi)]$, there exists $\rho > 0$ such that $\|\hat{y}(t, \tilde{x}(t))\| \leq \rho(1 + |t| + \|\tilde{x}(t)\|)$. ∎

Remark 2.6. Following the proofs of Theorems 2.3 and 2.5, The linear growth condition (2.2) in Theorem 2.5 plays an important role on the subsequent convergence analysis. The paper [15] investigated a parameterized convex program with linear constraints and a nonsmooth objective function. By assuming the superquadratic and subquadratic growth conditions for the objective function and the Mangasarian-Fromovitz regularity condition (MFC), the authors showed the upper Lipschitz continuity of the unique optimal solution. We can also derive (2.2) by the upper Lipschitz continuity of the optimal solution of problem (1.2) w.r.t. (t, x) at the point $(0, x_0)$. Our conditions (conditions of Theorem 2.5) are easier to verify and weaker than the conditions in [15].

The authors in [22] also established a linear growth condition for the algebraic variable (the solution of a VI) to ensure the convergence of the implicit Euler method for the DVI. Moreover, in [14], Han et al. derived a linear growth condition for the least-norm solution of a monotone linear complementarity problem and proposed an implicit time-stepping method using the least-norm solutions for differential complementarity systems. Without computing the least-norm solution for a monotone DVI, Chen and Wang [9] proposed a regularized time-stepping method for the DVI and provided the corresponding convergence analysis. These results can be extended to problem (1.1)-(1.2) if g is continuously differentiable and independent of ξ. This papers focus on the case that g is nonsmooth and random.

3. Regularization method.

Since $g(t, \mathbf{x}, \cdot, \xi)$ is convex for any $t \in \mathbb{R}_+$, $\mathbf{x} \in \mathbb{R}^n$ and a.e. $\xi \in \Xi$, $\mathbb{E}[g(t, \mathbf{x}, \cdot, \xi)]$ is convex for any $t \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^n$. Therefore, we add a regularization term $\mu\|\mathbf{y}\|^2$ with $\mu > 0$ to the objective function in (1.2) and get the following regularization optimization problem:

$$y^\mu(t) = \arg\min_{\mathbf{y}\in\mathbb{R}^m} g^\mu(t, x(t), \mathbf{y}) \quad \text{s.t.} \quad \mathbf{y} \in K(t, x(t)), \tag{3.1}$$

where $g^\mu(t, x(t), \mathbf{y}) = \mathbb{E}[g(t, x(t), \mathbf{y}, \xi)] + \mu\|\mathbf{y}\|^2$.

Obviously, under the assumption that $\mathrm{int}K(0, x_0) \neq \emptyset$, there are $\sigma > 0$ and $\delta > 0$ such that the optimization problem (3.1) has a unique optimal solution $\hat{y}^\mu(t, x(t))$ over $K(t, x(t))$ for any $(t, x(t)) \in [0, \sigma] \times \mathcal{B}(x_0, \delta)$ with $x \in C^0([0, \sigma])$.

PROPOSITION 3.1. *Suppose that Assumption 2.2 holds. Let $\hat{y}^\mu(t, x(t))$ be the unique optimal solution of problem (3.1) with some $\mu > 0$, $t \in \mathbb{R}_+$ and $x(t) \in \mathbb{R}^n$. Then it holds that*

$$\|\hat{y}^\mu(t, x(t))\| \leq \min_{\mathbf{y}\in\mathcal{S}(t,x(t))} \|\mathbf{y}\|, \tag{3.2}$$

where $\mathcal{S}(t, x(t))$ is the optimal solution set of problem (1.2) with $t \in \mathbb{R}_+$ and $x(t) \in \mathbb{R}^n$.

Proof. Since $\hat{y}^\mu(t, x(t))$ is the unique optimal solution of the optimization problem (3.1), there exists $\hat{z}^\mu(t, x(t)) \in \partial_{y(t)}\mathbb{E}[g(t, x(t), \hat{y}^\mu(t, x(t)), \xi)]$ such that

$$(\mathbf{y} - \hat{y}^\mu(t, x(t)))^\top(\hat{z}^\mu(t, x(t)) + \mu\hat{y}^\mu(t, x(t))) \geq 0, \quad \forall\, \mathbf{y} \in K(t, x(t)).$$

Let $\bar{z}(t, x(t)) \in \partial_{y(t)}\mathbb{E}[g(t, x(t), \bar{y}(t, x(t)), \xi)]$, where $\bar{y}(t, x(t))$ is the least-norm element of $\mathcal{S}(t, x(t))$. Then we have

$$(\bar{y}(t, x(t)) - \hat{y}^\mu(t, x(t)))^\top(\hat{z}^\mu(t, x(t)) + \mu\hat{y}^\mu(t, x(t))) \geq 0 \tag{3.3}$$

and

$$(\hat{y}^\mu(t, x(t)) - \bar{y}(t, x(t)))^\top\bar{z}(t, x(t)) \geq 0. \tag{3.4}$$

Since $\mathbb{E}[g(t,x(t),\cdot,\xi)]$ is convex, the set-valued mapping $\partial_{y(t)}\mathbb{E}[g(t,x(t),\cdot,\xi)]$ is monotone [24, Theorem 12.17]. Therefore, from (3.4), we can obtain

$$(\hat{y}^\mu(t,x(t)) - \bar{y}(t,x(t)))^\top \hat{z}^\mu(t,x(t)) \geq 0.$$

We then get from (3.3) that $\mu(\bar{y}(t,x(t)) - \hat{y}^\mu(t,x(t)))^\top \hat{y}^\mu(t,x(t)) \geq 0$, which implies that $\|\hat{y}^\mu(t,x(t))\| \leq \|\bar{y}(t,x(t))\|$. ∎

THEOREM 3.2. *Suppose that the set* int$K(0,x_0)$ *is not empty. Then there exists* $\hat{T}(\mu) > 0$ *such that problem* (1.1) *with* (3.1) *has a solution* $(x^\mu, y^\mu) \in C^1([0,\hat{T}(\mu)]) \times C^0([0,\hat{T}(\mu)])$ *for any* $\mu > 0$*. Moreover, there is a positive number* $\hat{T}_0$ *such that* $\hat{T}(\mu) \geq \hat{T}_0$ *for any* $\mu > 0$ *if the optimal solution set of problem* (1.2) *is not empty.*

Proof. Similar with the proof of Theorem 2.5, there are $\sigma > 0$ and $\bar{\sigma}(\mu) > 0$ such that $\hat{T}(\mu) = \min\{\sigma, \bar{\sigma}(\mu)\}$.

Now we illustrate the existence of $\hat{T}_0$. By (3.2), there is $\rho_\alpha > 0$ such that $\|\hat{y}^\mu(t,x(t))\| \leq \rho_\alpha$ for any t, $x(t)$ and μ. Obviously, if $\bar{\sigma}(\mu) \geq \sigma$ for any $\mu > 0$, we have $\hat{T}_0 = \sigma$. If $\bar{\sigma}(\mu) < \sigma$ for some $\mu > 0$, we know that there is $\delta_0 \in (0,\delta]$ such that $\|x^\mu(\bar{\sigma}(\mu)) - x_0\| = \delta_0$. From $\|x^\mu(t) - x_0\| \leq \delta$ for any $t \in [0,\bar{\sigma}(\mu))$, we obtain

$$\delta_0 = \left\| \int_0^{\bar{\sigma}(\mu)} \mathbb{E}[f(\tau, x^\mu(\tau), \hat{y}^\mu(\tau, x^\mu(\tau)), \xi)] d\tau \right\| \leq \int_0^{\bar{\sigma}(\mu)} (\check{\rho}_f \tau + \Theta) d\tau$$

which means that $\bar{\sigma}(\mu) \geq \sqrt{\Theta^2 + 2\check{\rho}_f\delta_0} - \Theta > 0$, where $\Theta = \check{\rho}_f(1 + \|x_0\| + \delta + \rho_\alpha)$. Therefore, we conclude the desired result. ∎

In the convergence analysis of the regularization method as $\mu \downarrow 0$, we use the following notations. Let $\mathcal{X}_T$ denote the space of n-dimensional vector-valued continuous functions over $[0,T]$ equipped with the norm

$$\|u\|_s := \sup_{t\in[0,T]} \|u(t)\|$$

and $\mathcal{Y}_T$ denote the space of m-dimensional vector-valued square integrable functions over $[0,T]$ equipped with the norm

$$\|v\|_{L^2} := \left(\int_0^T \|v(t)\|^2 dt \right)^{\frac{1}{2}}.$$

We define the norm for $(u,v) \in \mathcal{X}_T \times \mathcal{Y}_T$ by

$$\|(u,v)\|_{\mathcal{X}_T\times\mathcal{Y}_T} = \|u\|_s + \|v\|_{L^2}.$$

Let $\mathfrak{X}_T$ and $\mathfrak{Y}_T$ denote the space of real-valued continuous functions and real-valued square integrable functions over $[0,T]$, respectively. When $n = 1$, we have $\mathfrak{X}_T = \mathcal{X}_T$. Let $\mathcal{Z}_T$ denote the space of m-dimensional vector-valued continuous functions over $[0,T]$. Similarly, we define

$$\begin{aligned}
\|(u,v)\|_{\mathcal{X}_T\times\mathcal{Z}_T} &= \|u\|_s + \sup_{t\in[0,T]} \|v(t)\|, \ \forall\ (u,v) \in \mathcal{X}_T \times \mathcal{Z}_T, \\
\|(u,v)\|_{\mathcal{X}_T\times\mathfrak{X}_T} &= \|u\|_s + \sup_{t\in[0,T]} |v(t)|, \ \forall\ (u,v) \in \mathcal{X}_T \times \mathfrak{X}_T, \\
\|(u,v)\|_{\mathcal{X}_T\times\mathfrak{Y}_T} &= \|u\|_s + \left(\int_0^T v^2(\tau) d\tau \right)^{\frac{1}{2}}, \ \forall\ (u,v) \in \mathcal{X}_T \times \mathfrak{Y}_T.
\end{aligned}$$

Denote the optimal value function of optimization problem (1.2) by $g_{min}(t, x(t))$ with $t \in \mathbb{R}_+$ and $x(t) \in \mathbb{R}^n$. According to [26, Theorem 3.1], we know that g_{min} is continuous over $[0, \sigma] \times \mathcal{B}(x_0, \delta)$ under Assumption 2.2 for some σ and δ in the proof of Theorem 2.3. Define

$$\Phi(x, y)(t) = \begin{pmatrix} x(t) - x_0 - \int_0^t \mathbb{E}[f(\tau, x(\tau), y(\tau), \xi)]d\tau \\ \mathbb{E}[g(t, x(t), y(t), \xi)] - g_{min}(t, x(t)) \end{pmatrix}.$$

Let some suitable T with $\sigma \geq T > 0$ be fixed. Obviously, we have $\Phi(x, y) \in \mathcal{X}_T \times \mathfrak{Y}_T$ for any $(x, y) \in \mathcal{X}_T \times \mathcal{Y}_T$, and $\Phi(x, y) \in \mathcal{X}_T \times \mathfrak{X}_T$ for any $(x, y) \in \mathcal{X}_T \times \mathcal{Z}_T$. Moreover, we know that $\|\Phi(x, y)\|_{\mathcal{X}_T \times \mathfrak{Y}_T} = 0$ and $y(t) \in K(t, x(t))$ imply that (x, y) is a weak solution of problem (1.1)-(1.2). And for a continuous function $y \in \mathcal{Z}_T$, $\|\Phi(x, y)\|_{\mathcal{X}_T \times \mathfrak{X}_T} = 0$ and $y(t) \in K(t, x(t))$ imply that (x, y) is a classic solution of problem (1.1)-(1.2). Similarly, let $g^\mu_{min}(t, x(t))$ denote the optimal value function of the optimization problem (3.1) with $t \in \mathbb{R}$, $x(t) \in \mathbb{R}^n$ and $\mu > 0$, and define

$$\Phi^\mu(x, y)(t) = \begin{pmatrix} x(t) - x_0 - \int_0^t \mathbb{E}[f(\tau, x(\tau), y(\tau), \xi)]d\tau \\ \mathbb{E}[g(t, x(t), y(t), \xi)] + \mu\|y(t)\|^2 - g^\mu_{min}(t, x(t)) \end{pmatrix}. \tag{3.5}$$

If $(x^\mu, y^\mu) \in C^1([0, T]) \times C^0([0, T])$ is a solution of problem (1.1) with (3.1), we have $\|\Phi^\mu(x^\mu, y^\mu)\|_{\mathcal{X}_T \times \mathfrak{X}_T} = 0$ and then $\|\Phi^\mu(x^\mu, y^\mu)\|_{\mathcal{X}_T \times \mathfrak{Y}_T} = 0$.

Let U_1 and U_2 be the spaces taken either $U_1 = \mathcal{X}_T \times \mathfrak{X}_T$ or $U_1 = \mathcal{X}_T \times \mathfrak{Y}_T$, and $U_2 = \mathcal{X}_T \times \mathcal{Z}_T$ or $U_2 = \mathcal{X}_T \times \mathcal{Y}_T$. A sequence $\{\Psi^k\}_{k=1}^\infty$ is said to be epigraphically convergent to a function Ψ, denoted by $\Psi^k \to^{epi} \Psi$, if

(i) $\liminf_{k\to\infty} \Psi^k(x^k, y^k) \geq \Psi(x, y)$ for any sequence $\{(x^k, y^k)\}_{k=1}^\infty \subseteq U_2$ with $(x^k, y^k) \to (x, y)$ by the norm $\|\cdot\|_{U_2}$;

(ii) $\limsup_{k\to\infty} \Psi^k(x^k, y^k) \leq \Psi(x, y)$ for some sequence $\{(x^k, y^k)\}_{k=1}^\infty \subseteq U_2$ with $(x^k, y^k) \to (x, y)$ by the norm $\|\cdot\|_{U_2}$.

To study the convergence of $\{(x^\mu, y^\mu)\}$ in U_2, we firstly have the following lemma about the mapping $\|\Phi^\mu\|_{U_1}$ is epigraphically convergent to $\|\Phi\|_{U_1}$ as $\mu \downarrow 0$.

LEMMA 3.3. *Suppose that Assumption 2.2 holds. Let $\{\mu_k\}_{k=1}^\infty \downarrow 0$ be given and $\Phi^k = \Phi^{\mu_k}$ be defined in (3.5). Then for any sequence $\{(x^k, y^k)\}_{k=1}^\infty \subset U_2$ with $(x^k, y^k) \to (x, y)$ by the norm $\|\cdot\|_{U_2}$ as $k \to \infty$, we have $\|\Phi^k(x^k, y^k)\|_{U_1} \to \|\Phi(x, y)\|_{U_1}$ and $\|\Phi^k\|_{U_1} \to^{epi} \|\Phi\|_{U_1}$.*

Proof. For any given $\mu > 0$, $t \in \mathbb{R}_+$ and $x(t) \in \mathbb{R}^n$, it is clear that $g_{min}(t, x(t)) \leq g^\mu_{min}(t, x(t))$ as $\hat{y}^\mu(t, x(t)) \in K(t, x(t))$. In addition,

$$\begin{aligned} g^\mu_{min}(t, x(t)) &= \mathbb{E}[g(t, x(t), \hat{y}^\mu(t, x(t)), \xi)] + \mu\|\hat{y}^\mu(t, x(t))\|^2 \\ &= \min_{\mathbf{y}\in K(t,x(t))} \{\mathbb{E}[g(t, x(t), \mathbf{y}, \xi)] + \mu\|\mathbf{y}\|^2\} \\ &\leq \min_{\mathbf{y}\in \mathcal{S}(t,x(t))} \{\mathbb{E}[g(t, x(t), \mathbf{y}, \xi)] + \mu\|\mathbf{y}\|^2\} \\ &\leq g_{min}(t, x(t)) + \mu \min_{\mathbf{y}\in \mathcal{S}(t,x(t))} \|\mathbf{y}\|^2, \end{aligned}$$

which means that $|g^\mu_{min}(t, x(t)) - g_{min}(t, x(t))| \leq \mu \min_{\mathbf{y}\in\mathcal{S}(t,x(t))} \|\mathbf{y}\|^2$ for any given $(t, x(t)) \in [0, \sigma] \times \mathcal{B}(x_0, \delta)$, since $\mathbb{E}[g(t, x(t), \hat{y}^\mu(t, x(t)), \xi)] \geq g_{min}(t, x(t))$ and $\hat{y}^\mu(t, x(t)) \in K(t, x(t))$. By the uniform boundedness of $\mathcal{S}(t, x(t))$ for any $(t, x(t)) \in [0, \sigma] \times \mathcal{B}(x_0, \delta)$ and $g^{\mu_1}_{min}(t, x(t)) \leq g^{\mu_2}_{min}(t, x(t))$ for $\mu_1 \leq \mu_2$, we can obtain that g^μ_{min} converges to g_{min} uniformly as $\mu \downarrow 0$ over $(t, x(t)) \in [0, \sigma] \times \mathcal{B}(x_0, \delta)$.

Let $(x^k, y^k) \to (x, y)$ by the norm $\|\cdot\|_{U_2}$ as $k \to \infty$. Taking $U_1 = \mathcal{X}_T \times \mathfrak{X}_T$ and $U_2 = \mathcal{X}_T \times \mathcal{Z}_T$, we have

$$\begin{aligned}&\|\Phi^k(x^k, y^k) - \Phi(x^k, y^k)\|_{U_1}\\ &\le \mu_k \sup_{t\in[0,T]} \|y^k(t, x^k(t))\|^2 + \sup_{t\in[0,T]} |g^{\mu_k}_{min}(t, x^k(t)) - g_{min}(t, x^k(t))| \to 0 \quad \text{as} \quad \mu_k \downarrow 0.\end{aligned}$$

If we take $U_1 = \mathcal{X}_T \times \mathfrak{Y}_T$ and $U_2 = \mathcal{X}_T \times \mathcal{Y}_T$, we have

$$\begin{aligned}&\|\Phi^k(x^k, y^k) - \Phi(x^k, y^k)\|_{U_1}\\ &\le \mu_k \|y^k(\cdot, x^k)\|^2_{L^2} + \left(\int_0^T (g^{\mu_k}_{min}(t, x^k(t)) - g_{min}(t, x^k(t)))^2 dt\right)^{\frac{1}{2}} \to 0 \quad \text{as} \quad \mu_k \downarrow 0.\end{aligned}$$

Moreover $\|\Phi(x^k, y^k)\|_{U_1} \to \|\Phi(x, y)\|_{U_1}$ as $k \to \infty$ since $\|\Phi\|_{U_1}$ is continuous. We then obtain $\|\Phi^k(x^k, y^k)\|_{U_1} \to \|\Phi(x, y)\|_{U_1}$ by

$$\begin{aligned}&\|\Phi^k(x^k, y^k) - \Phi(x, y)\|_{U_1}\\ &\le \|\Phi^k(x^k, y^k) - \Phi(x^k, y^k)\|_{U_1} + \|\Phi(x^k, y^k) - \Phi(x, y)\|_{U_1}.\end{aligned}$$

It then implies that $\|\Phi^k\|_{U_1} \to^{epi} \|\Phi\|_{U_1}$. ∎

THEOREM 3.4. *Suppose that Assumption 2.2 holds. Let* $(x^\mu, y^\mu) \in C^1([0, \hat{T}_0]) \times C^0([0, \hat{T}_0])$ *be a solution of problem* (1.1) *with* (3.1) *for any* $\mu > 0$. *Then there exists a sequence* $\{\mu_k\}_{k=1}^\infty \downarrow 0$ *such that* $x^{\mu_k} \to x^*$ *as* $k \to \infty$ *uniformly over* $[0, \hat{T}_0]$ *and* $y^{\mu_k} \to y^*$ *as* $k \to \infty$ *weakly in* $\mathcal{Y}_{\hat{T}_0}$. *In addition,*

(i) if $y^{\mu_k} \to y^*$ *w.r.t.* $\|\cdot\|_{L^2}$ *as* $k \to \infty$, *then* (x^*, y^*) *is a weak solution of* (1.1)-(1.2) *over* $[0, \hat{T}_0]$;

(ii) if $y^{\mu_k} \to y^*$ *uniformly as* $k \to \infty$, *then* (x^*, y^*) *is a classic solution of* (1.1)-(1.2) *over* $[0, \hat{T}_0]$; *moreover,* $y^*(t)$ *is the unique least-norm optimal solution of problem* (1.2) *with* t *and* $x^*(t)$.

Proof. Notice that the Lipschitz property (1.3) implies that $\mathbb{E}[f(\cdot, \cdot, \cdot, \xi)]$ has linear growth in $(t, \mathbf{x}, \mathbf{y}) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m$, i.e., there exists $\check{\rho}_f > 0$ such that $\|\mathbb{E}[f(t, \mathbf{x}, \mathbf{y}, \xi)]\| \le \check{\rho}_f(1 + |t| + \|\mathbf{x}\| + \|\mathbf{y}\|)$. Let $(x^\mu, y^\mu) \in C^1([0, \hat{T}_0]) \times C^0([0, \hat{T}_0])$ be a solution of problem (1.1) with (3.1) for any $\mu > 0$. According to (3.2) and the compactness of the optimal solution set of problem (1.2), there is $\rho_\alpha > 0$ (independent of μ, t and $x(t)$) such that $\|y^\mu(t)\| \le \rho_\alpha$. We then have

$$\begin{aligned}\|x^\mu(t)\| &\le \|x_0\| + \int_0^t \|\mathbb{E}[f(\tau, x^\mu(\tau), y^\mu(\tau), \xi)]\| d\tau\\ &\le \|x_0\| + \check{\rho}_f \int_0^t (1 + \rho_\alpha + |\tau| + \|x^\mu(\tau)\|) d\tau,\end{aligned}$$

which implies that for any $t \in [0, \hat{T}_0]$ there exists $\bar{\rho}_f > 0$ such that

$$\|x^\mu(t)\| \le \|x_0\| + \bar{\rho}_f \int_0^t (1 + \|x^\mu(\tau)\|) d\tau.$$

We then have $\|x^\mu\|_s \le (1 + \|x_0\|)\exp(\bar{\rho}_f \hat{T}_0) - 1$, according to [9, Lemma 2.6]. Hence, $\{x^\mu\}$ is uniformly bounded on $[0, \hat{T}_0]$ for any $\mu > 0$ and then so is $\{\dot{x}^\mu\}$, which means that $\{x^\mu\}$ is equicontinuous over $[0,\hat{T}_0]$ for any $\mu>0$. By Arzelá-Ascoli theorem [17], there exists a sequence $\{\mu_k\}_{k=1}^\infty \downarrow 0$ such that $\{x^{\mu_k}\}$ is convergent to a point $x^*\in\mathcal{X}_{\hat{T}_0}$ uniformly over $[0,\hat{T}_0]$.

In addition, we know that $\{y^\mu\}$ is uniformly bounded on $[0,\hat{T}_0]$ for any $\mu>0$ by $\|y^\mu(t)\|\le\rho_\alpha(1+|t|+\|x^\mu(t)\|)$. By Alaglu's theorem [17], there exists a subsequence of $\{y^{\mu_k}\}$, which we may assume without loss of generality to be $\{y^{\mu_k}\}$ itself, has a weak* limit, named y^*, in $\mathcal{Y}_{\hat{T}_0}$. Since $\mathcal{Y}_{\hat{T}_0}$ is a Hilbert space, it is a reflexive Banach space, which implies that weak* convergent sequences are also weakly convergent sequences.

By [24, Example 9.35], we know that $K(t,\cdot)$ is Lipschitz continuous on its domain for any $t\in\mathbb{R}$, which means that $\mathbb{H}(K(t,x^\mu(t)),K(t,x^*(t)))\to 0$ as $x^\mu\to x^*$ uniformly. Then by $y^\mu(t)\in K(t,x^\mu(t))$, we have $y^*(t)\in K(t,x^*(t))$. Then following Lemma 3.3, we know that if $y^{\mu_k}\to y^*$ w.r.t. $\|\cdot\|_{L^2}$, (x^*,y^*) is a weak solution of (1.1)-(1.2) over $[0,\hat{T}_0]$; if $y^{\mu_k}\to y^*$ uniformly as $k\to\infty$, then (x^*,y^*) is a classic solution of (1.1)-(1.2) over $[0,\hat{T}_0]$.

Let $\hat{y}^{\mu_k}(t,x(t))$ denote the unique optimal solution of problem (3.1) with any $\mu_k>0$, $t\in\mathbb{R}_+$ and $x(t)\in\mathbb{R}^n$. Then by $y^{\mu_k}(t)=\hat{y}^{\mu_k}(t,x^{\mu_k}(t))$, $y^{\mu_k}\to y^*$ uniformly as $k\to\infty$, and the continuity of $\hat{y}^{\mu_k}$, we know that $\lim_{k\to\infty}\|\hat{y}^{\mu_k}(t,x^*(t))-y^*(t)\|=0$. Since (x^*,y^*) is a classic solution of (1.1)-(1.2), we obtain that $y^*(t)\in\mathcal{S}(t,x^*(t))$ and then $y^*(t)$ is the unique least-norm element of $\mathcal{S}(t,x^*(t))$ by (3.2). ∎

When there exists a constant $\tilde{\varrho}>0$ such that for any $\mathbf{y}_1$, $\mathbf{y}_2\in\mathbb{R}^m$ and $\tau\in(0,1)$,

$$\begin{aligned}\mathbb{E}[g(t,\mathbf{x},(1-\tau)\mathbf{y}_1+\tau\mathbf{y}_2,\xi)]\le&(1-\tau)\mathbb{E}[g(t,\mathbf{x},\mathbf{y}_1,\xi)]+\tau\mathbb{E}[g(t,\mathbf{x},\mathbf{y}_2,\xi)]\\&-\frac{1}{2}\tilde{\varrho}\tau(1-\tau)\|\mathbf{y}_1-\mathbf{y}_2\|^2\end{aligned}$$

holds for any fixed $t\in\mathbb{R}_+$ and $\mathbf{x}\in\mathbb{R}^n$, the objective function of optimization problem (1.2) is strongly convex w.r.t. $\mathbf{y}$, and (1.2) admits a unique optimal solution y^* which is continuous w.r.t. t for any $x\in C^0([0,\sigma])$ by Theorem 2.5. Fix some $x\in C^0([0,\sigma])$ and let $z^\mu(t)\in\partial_{y(t)}\mathbb{E}[g(t,x(t),y^\mu(t),\xi)]$ and $z^*(t)\in\partial_{y(t)}\mathbb{E}[g(t,x(t),y^*(t),\xi)]$. Then we have

$$(y^*(t)-y^\mu(t))^\top(z^\mu(t)+\mu y^\mu(t))\ge 0 \text{ and } (y^\mu(t)-y^*(t))^\top z^*(t)\ge 0. \tag{3.6}$$

In this case, by the strong monotonicity of $\partial_{y(t)}\mathbb{E}[g(t,x(t),\cdot,\xi)]$ and (3.6) and (3.2), we have

$$\begin{aligned}\tilde{\varrho}\|y^\mu(t)-y^*(t)\|^2&\le(y^\mu(t)-y^*(t))^\top(z^\mu(t)-z^*(t))\le(y^\mu(t)-y^*(t))^\top z^\mu(t)\\&\le\mu(y^*(t)-y^\mu(t))^\top y^\mu(t)\le\mu\|y^*(t)-y^\mu(t)\|\|y^\mu(t)\|\\&\le\mu\|y^*(t)-y^\mu(t)\|\|y^*(t)\|.\end{aligned}$$

We then obtain that the pointwise convergence of y^μ to y^* as $\mu\downarrow 0$, which means that $y^\mu\to y^*$ w.r.t. $\|\cdot\|_{L^2}$ as $\mu\downarrow 0$ and the uniform convergence by the continuity of y^μ and y^*.

4. Sample average approximation. We apply the sample average approximation (SAA) approach to solve problem (1.1)-(1.2). We consider an independent identically distributed (i.i.d) sample of $\xi(\omega)$, which is denoted by $\{\xi_1,\cdots,\xi_\nu\}$, and use the following SAA problem to approximate problem (1.1)-(1.2):

$$\dot{x}(t) = \frac{1}{\nu}\sum_{\ell=1}^{\nu} f(t, x(t), y(t), \xi_\ell), \;\; x(0) = x_0, \tag{4.1}$$

$$\begin{aligned} y(t) \in \arg\min_{\mathbf{y}\in\mathbb{R}^m} \; & \frac{1}{\nu}\sum_{\ell=1}^{\nu} g(t, x(t), \mathbf{y}, \xi_\ell) \\ \text{s.t.} \;\; & \mathbf{y} \in K^\nu(t, x(t)), \end{aligned} \tag{4.2}$$

where

$$K^\nu(t, x(t)) \triangleq \left\{ \mathbf{y} \in \mathbb{R}^m : \frac{1}{\nu}\sum_{\ell=1}^{\nu}(A(\xi_\ell)x(t) + B(\xi_\ell)\mathbf{y} + Q(t, \xi_\ell)) \le 0 \right\}.$$

In this paper, by saying a property holds with probability 1 (w.p.1) for sufficiently large ν, we mean that there exists a set $\Omega_0 \subset \Omega$ of $\mathcal{P}$-measure zero such that for all $\omega \in \Omega\backslash\Omega_0$ there exists a positive integer $\nu^*(\omega)$ such that the property holds for all $\nu \ge \nu^*(\omega)$.

THEOREM 4.1. *Suppose that Assumption 2.4 holds and* int$K(0, x_0) \neq \emptyset$. *Then there exists* $T^* > 0$ *such that problem (4.1)-(4.2) has a solution* $(x^\nu, y^\nu) \in C^1([0, T^*]) \times C^0([0, T^*])$ *w.p.1 for sufficiently large* ν. *Moreover, there exists* $\rho^* > 0$ *such that*

$$\|y^\nu(t)\| \le \rho^*(1 + |t| + \|x^\nu(t)\|), \quad \text{for} \quad t \in [0, T^*] \tag{4.3}$$

w.p.1 for sufficiently large ν.

Proof. By int$K(0, x_0) \neq \emptyset$ and the strong Law of Large Number, we can conclude int$K^\nu(0, x_0) \neq \emptyset$ w.p.1 for sufficiently large ν.

Similar with the proof of Theorem 2.3, we can also conclude that there are $\sigma_1 > 0$ and $\delta_1 > 0$ such that int$K^\nu(t, x(t)) \neq \emptyset$ for any $(t, x(t)) \in [0, \sigma_1] \times \mathcal{B}(x_0, \delta_1)$ w.p.1 for sufficiently large ν.

Assumption 2.4 implies that $\frac{1}{\nu}\sum_{\ell=1}^{\nu} g(t, x(t), \cdot, \xi_\ell)$ is strongly convex w.p.1 for sufficiently large ν. Similar with the proof of Theorem 2.5, we obtain our results. ◻

THEOREM 4.2. *Suppose that Assumption 2.4 holds and* int$K(0, x_0) \neq \emptyset$. *Let* $(x^\nu, y^\nu) \in C^1([0, T^*]) \times C^0([0, T^*])$ *be a solution of problem (4.1)-(4.2). Then there are* $\bar{T}$ *with* $T^* \ge \bar{T} > 0$ *and a sequence* $\{\nu_k\}_{k=1}^\infty$ *with* $\nu_k \to \infty$ *such that* $x^{\nu_k} \to x^*$ *as* $k \to \infty$ *w.p.1 uniformly over* $[0, \bar{T}]$ *and* $y^{\nu_k} \to y^*$ *w.p.1 as* $k \to \infty$ *w.r.t.* $\|\cdot\|_{L^2}$ *in* $\mathcal{Y}_{\bar{T}}$, *where* (x^*, y^*) *is a weak solution of (1.1)-(1.2) over* $[0, \bar{T}]$.

Proof. Since $(x^\nu, y^\nu) \in C^1([0, T^*]) \times C^0([0, T^*])$ is a solution of problem (4.1)-(4.2), by the linear growth condition (4.3), we obtain that there exists $\hat{\rho}_f > 0$ such that

$$\|x^\nu(t)\| \le \|x_0\| + \hat{\rho}_f \int_0^t (1 + \|x^\nu(\tau)\|)d\tau$$

holds w.p.1 for sufficiently large ν, which implies that $\|x^\nu\|_s \le (1+\|x_0\|)\exp(\hat{\rho}_f T^*)-1$ w.p.1 for sufficiently large ν. Hence, we obtain that $\{x^\nu\}$ is uniformly bounded w.p.1 for sufficiently large ν and then so is $\{\dot{x}^\nu\}$, which means that $\{x^\nu\}$ is equicontinuous over $[0, T^*]$ w.p.1 for sufficiently large ν. By Arzelá-Ascoli theorem, there exists a sequence $\{\nu_k\}$ such that $\{x^{\nu_k}\}$ is convergent to a point $x^* \in \mathcal{X}_{T^*}$ as $\nu_k \to \infty$ w.p.1 uniformly over $[0, T^*]$.

Let $\hat{y}(t,x(t))$ and $\hat{y}^\nu(t,x(t))$ denote the optimal solutions of optimization problems (1.2) and (4.2) with t and $x(t)$, respectively. Theorems 2.5 and 4.1 imply that there are $\tilde{\sigma}>0$ and $\tilde{\delta}>0$ such that $\hat{y}$ and $\hat{y}^\nu$ are bounded and $\mathrm{int}K(t,x(t))\neq\emptyset$ over $(t,x(t))\in[0,\tilde{\sigma}]\times\mathcal{B}(x_0,\tilde{\delta})$. It implies that there exists a compact set $\mathcal{C}\subseteq\mathbb{R}^m$ such that $\hat{y}(t,x(t))\in\mathcal{C}$ and $\hat{y}^\nu(t,x(t))\in\mathcal{C}$ w.p.1 for sufficiently large ν. Following from [25, Theorem 7.53], we obtain that for any given $t\in\mathbb{R}_+$ and $x(t)\in\mathbb{R}^n$, $\frac{1}{\nu}\sum_{\ell=1}^{\nu} g(t,x(t),\mathbf{y},\xi_\ell)$ converges to $\mathbb{E}[g(t,x(t),\mathbf{y},\xi)]$ w.p.1 uniformly on $\mathbf{y}\in\mathcal{C}$ as $\nu\to\infty$. In addition, following from the strong Large Law of Number, we can obtain that $\mathbb{D}(K^\nu(t,x(t)),K(t,x(t)))\to 0$ w.p.1 as $\nu\to\infty$ for any $t\in\mathbb{R}_+$ and $x(t)\in\mathbb{R}^n$, which means that for any fixed t and $x(t)$ if $\mathbf{y}^\nu\in K^\nu(t,x(t))$ and $\mathbf{y}^\nu$ converges w.p.1 to a point $\mathbf{y}$, then $\mathbf{y}\in K(t,x(t))$. According to [25, Remark 8], we conclude that there exists a sequence $\{y^\nu(t,x(t))\}$ with $y^\nu(t,x(t))\in K^\nu(t,x(t))$ such that $y^\nu(t,x(t))\to\hat{y}(t,x(t))$ w.p.1 as $\nu\to\infty$ since $\mathrm{int}K(t,x(t))\neq\emptyset$ for any $(t,x(t))\in[0,\tilde{\sigma}]\times\mathcal{B}(x_0,\tilde{\delta})$. Therefore, following from [25, Theorem 5.5], we can obtain that $\hat{y}^\nu(t,x(t))$ converges to $\hat{y}(t,x(t))$ w.p.1 as $\nu\to\infty$ on $(t,x(t))\in[0,\tilde{\sigma}]\times\mathcal{B}(x_0,\tilde{\delta})$. According to Lebesgue Dominated Convergence Theorem and (4.3), we then conclude that there exists $\tilde{T}>0$ such that $\hat{y}^\nu(\cdot,x)$ converges to $\hat{y}(\cdot,x)$ w.p.1 w.r.t. $\|\cdot\|_{L^2}$ as $\nu\to\infty$, where x is continuously differentiable over $t\in[0,\tilde{T}]$.

Let $\bar{T}=\min\{T^*,\tilde{T}\}$. It is clear that $y^\nu(t)=\hat{y}^\nu(t,x^\nu(t))$ and $\hat{y}^{\nu_k}(\cdot,x^{\nu_k})$ converges to $\hat{y}(\cdot,x^*)$ w.p.1 w.r.t. $\|\cdot\|_{L^2}$ as $k\to\infty$, following from the continuity of $\hat{y}^\nu$ and $\hat{y}$. Denote $y^*(t)=\hat{y}(t,x^*(t))$. Then taking $\{\nu_k\}$ with $\nu_k\to\infty$, for any $t\in[0,\bar{T}]$, we have

$$\begin{aligned}&\left\|\int_0^t \frac{1}{\nu_k}\sum_{\ell=1}^{\nu_k} f(\tau,x^{\nu_k}(\tau),y^{\nu_k}(\tau),\xi_\ell)d\tau-\int_0^t \mathbb{E}[f(\tau,x^*(\tau),y^*(\tau),\xi)]d\tau\right\|\\ &\leq \frac{1}{\nu_k}\sum_{\ell=1}^{\nu_k}\kappa_f(\xi_\ell)\left(\bar{T}\|x^{\nu_k}-x^*\|_s+\sqrt{\bar{T}}\|y^{\nu_k}-y^*\|_{L^2}\right)+\mathcal{L}_{\bar{T}},\end{aligned}$$

where

$$\mathcal{L}_{\bar{T}}=\bar{T}\left\|\frac{1}{\nu_k}\sum_{\ell=1}^{\nu_k} f(\cdot,x^*,y^*,\xi_\ell)-\mathbb{E}[f(\cdot,x^*,y^*,\xi)]\right\|_s.$$

Similarly, by [25, Theorem 7.53], we obtain that $\frac{1}{\nu}\sum_{\ell=1}^{\nu} f(t,x^*(t),y^*(t),\xi_\ell)$ converges to $\mathbb{E}[f(t,x^*(t),y^*(t),\xi)]$ w.p.1 uniformly on $t\in[0,\bar{T}]$ as $\nu\to\infty$. Therefore, we can conclude that

$$x^*(t)=x_0+\int_0^t \mathbb{E}[f(\tau,x^*(\tau),y^*(\tau),\xi)]d\tau$$

by $x^{\nu_k}(t)=x_0+\int_0^t\frac{1}{\nu_k}\sum_{\ell=1}^{\nu_k} f(\tau,x^{\nu_k}(\tau),y^{\nu_k}(\tau),\xi)d\tau$. By $x^*\in\mathcal{X}_{\bar{T}}$, we obtain $y^*\in\mathcal{Y}_{\bar{T}}$, which means that (x^*,y^*) is a weak solution of problem (1.1)-(1.2) over $[0,\bar{T}]$. ◻

For the case that $\mathbb{E}[g(t,\mathbf{x},\cdot,\xi)]$ is convex, we can choose a measurable function $\hat{\varrho}:\Xi\to\mathbb{R}_{++}$ with $0<\mathbb{E}[\tilde{\rho}(\xi)]<\infty$ and consider the regularized function

$$\tilde{g}(t,x(t),y(t),\xi)=g(t,x(t),y(t),\xi)+\frac{\mu}{2}\tilde{\rho}(\xi)\|y(t)\|^2.$$

Then Assumption 2.4 holds for $\tilde{g}$ with $\mu\tilde{\rho}$ and $\mu>0$. We apply SAA method to (1.1)-(1.2) with $\tilde{g}$ and obtain

$$y^\nu_\mu(t)=\arg\min_{\mathbf{y}\in K^\nu(t,x(t))}\frac{1}{\nu}\sum_{\ell=1}^{\nu} g(t,x(t),\mathbf{y},\xi_\ell)+\frac{\mu}{2\nu}\sum_{\ell=1}^{\nu}\tilde{\rho}(\xi_\ell)\|\mathbf{y}\|^2. \tag{4.4}$$

According to Theorems 3.4 and 4.2, we can obtain the following result.

THEOREM 4.3. *Suppose that Assumption 2.2 holds. Let* (x_μ^ν, y_μ^ν) *be a solution of problem* (4.1) *with* (4.4) *for some* $\mu > 0$ *and* $\nu > 0$. *Then there are* $\check{T} > 0$, $(x^*, y^*) \in C^1([0,\check{T}]) \times C^0([0,\check{T}])$, *a sequence* $\{\mu_k\}_{k=1}^\infty$ *with* $\mu_k \downarrow 0$ *and a sequence* $\{\nu_k\}_{k=1}^\infty$ *with* $\nu_k \to \infty$ *such that*

$$\lim_{\mu_k \downarrow 0} \lim_{\nu_k \to \infty} \|x_{\mu_k}^{\nu_k} - x^*\|_s = 0, \ \ w.p.1$$

and $y_{\mu_k}^{\nu_k} \to y^*$ *weakly w.p.1 in* $\mathcal{Y}_{\check{T}}$ *by the order of* $\mu_k \downarrow 0$ *and* $\nu_k \to \infty$. *If*

$$\lim_{\mu_k \downarrow 0} \lim_{\nu_k \to \infty} \|y_{\mu_k}^{\nu_k} - y^*\|_{L^2} = 0, \ \ w.p.1$$

then (x^*, y^*) *is a weak solution of* (1.1)-(1.2) *over* $[0, \check{T}]$.

5. Time-stepping method. We now adopt the time-stepping method for solving problem (4.1)-(4.2) with a fixed sample $\{\xi_1, \dots, \xi_\nu\}$, which uses a finite-difference formula to approximate the time derivative $\dot{x}$. For a fixed $\bar{T}$ in Theorem 4.2, it begins with the division of the time interval $[0, \bar{T}]$ into N subintervals for a fixed step size $h = \bar{T}/N = t_{i+1} - t_i$ where $i = 0, \cdots, N-1$. Inspired by the DVI-specific time-stepping approach in [22], we propose to solve the optimization problem (4.2) independently of the first equation (4.1). This method is different with the time-stepping method which is usually adopted in [7, 20, 21]. Therefore, starting from $\mathbf{x}_0^\nu = x_0$, we compute two finite sets of vectors $\{\mathbf{x}_1^\nu, \mathbf{x}_2^\nu, \cdots, \mathbf{x}_N^\nu\} \subset \mathbb{R}^n$ and $\{\mathbf{y}_1^\nu, \mathbf{y}_2^\nu, \cdots, \mathbf{y}_N^\nu\} \subset \mathbb{R}^m$ in the following manner for $i = 0, \cdots, N-1$:

$$\mathbf{x}_{i+1}^\nu = \mathbf{x}_i^\nu + \frac{h}{\nu} \sum_{\ell=1}^{\nu} f(t_{i+1}, \mathbf{x}_{i+1}^\nu, \mathbf{y}_{i+1}^\nu, \xi_\ell), \tag{5.1}$$

$$\begin{aligned} \mathbf{y}_{i+1}^\nu = \arg \min_{\mathbf{y} \in \mathbb{R}^m} & \frac{1}{\nu} \sum_{\ell=1}^{\nu} g(t_{i+1}, \mathbf{x}_i^\nu, \mathbf{y}, \xi_\ell) \\ \text{s.t.} \ \ & \mathbf{y} \in K^\nu(t_{i+1}, \mathbf{x}_i^\nu). \end{aligned} \tag{5.2}$$

THEOREM 5.1. *Suppose that Assumption 2.4 holds and* int$K(0, x_0) \neq \emptyset$. *Then problem* (5.1)-(5.2) *has a unique solution* $\{\mathbf{x}_i^\nu, \mathbf{y}_i^\nu\}_{i=1}^N$ *w.p.1 for sufficiently large* ν *and sufficiently small* h. *Moreover, there exists* $\hat{\rho} > 0$ *such that for any* $i \in \{0, \cdots, N-1\}$,

$$\|\mathbf{y}_{i+1}^\nu\| \leq \hat{\rho}(1 + \|\mathbf{x}_i^\nu\|)$$

holds w.p.1 for sufficiently large ν *and* N.

Proof. For any $i \in \{0, \cdots, N-1\}$, $\mathbf{y}_{i+1}^\nu$ is a unique optimal solution of problem (5.2) w.p.1 for sufficiently large ν. Similar with the proof of Theorem 2.5 and for a fixed t_i, there exists $\hat{\rho}_i > 0$ such that

$$\|\mathbf{y}_{i+1}^\nu\| \leq \hat{\rho}_i(1 + \|\mathbf{x}_i^\nu\|). \tag{5.3}$$

Following from the Lipschitz property of $f(\cdot, \cdot, \cdot, \xi)$ in (1.3), we obtain that for any $\tilde{\mathbf{x}}$ and $\bar{\mathbf{x}} \in \mathbb{R}^n$,

$$\begin{aligned} & \left\| \frac{h}{\nu} \sum_{\ell=1}^{\nu} f(t_{i+1}, \tilde{\mathbf{x}}, \mathbf{y}_{i+1}^\nu, \xi_\ell) - \frac{h}{\nu} \sum_{\ell=1}^{\nu} f(t_{i+1}, \bar{\mathbf{x}}, \mathbf{y}_{i+1}^\nu, \xi_\ell) \right\| \\ \leq & \frac{h}{\nu} \sum_{\ell=1}^{\nu} \| f(t_{i+1}, \tilde{\mathbf{x}}, \mathbf{y}_{i+1}^\nu, \xi_\ell) - f(t_{i+1}, \bar{\mathbf{x}}, \mathbf{y}_{i+1}^\nu, \xi_\ell) \| \leq \kappa h \|\tilde{\mathbf{x}} - \bar{\mathbf{x}}\|, \end{aligned}$$

where $\kappa \geq \mathbb{E}[\kappa_f(\xi)] \geq \frac{1}{\nu}\sum_{\ell=1}^{\nu}\kappa_f(\xi_\ell)$ w.p.1 for sufficiently large ν. Therefore, if $h < \frac{1}{\kappa}$, we know that $\frac{h}{\nu}\sum_{\ell=1}^{\nu} f(t_{i+1}, \cdot, \mathbf{y}_{i+1}^{\nu}, \xi_\ell)$ is a contractive mapping. Moreover, there exists $\tilde{\rho}_f > 0$ such that for any $i = 0, \cdots, N-1$

$$\|\mathbf{x}_{i+1}^{\nu}\| \leq \|\mathbf{x}_i^{\nu}\| + \frac{h}{\nu}\sum_{\ell=1}^{\nu}\|f(t_{i+1}, \mathbf{x}_{i+1}^{\nu}, \mathbf{y}_{i+1}^{\nu}, \xi_\ell)\| \leq \|\mathbf{x}_i^{\nu}\| + h\tilde{\rho}_f(1 + \|\mathbf{x}_{i+1}^{\nu}\|).$$

It implies that there exists $0 < h_0 < \frac{1}{\tilde{\rho}_f}$ such that $\|\mathbf{x}_{i+1}^{\nu}\| \leq \exp(\frac{\tilde{\rho}_f \bar{T}}{1-h_0\tilde{\rho}_f})(1+\|x_0\|)+1$ for $h \in (0, h_0]$. The contraction mapping theorem implies that there exists unique $\mathbf{x}_{i+1}^{\nu}$ such that (5.1) holds with $i = 0, \cdots, N-1$. We then conclude that problem (5.1)-(5.2) has a unique solution $\{\mathbf{x}_i^{\nu}, \mathbf{y}_i^{\nu}\}_{i=1}^{N}$ w.p.1 for sufficiently large ν and sufficiently small h and the linear growth condition (5.3) holds by $\hat{\rho} = \max_{i\in\{1,\cdots,N\}}\{\hat{\rho}_i\}$. ∎

Let $\{\mathbf{x}_i^{\nu}, \mathbf{y}_i^{\nu}\}_{i=1}^{N}$ be a solution of (5.1)-(5.2). We define a piecewise linear function x_h^{ν} and a piecewise constant function y_h^{ν} on $[0, \bar{T}]$ as below:

$$x_h^{\nu}(t) = \mathbf{x}_i^{\nu} + \frac{t - t_i}{h}(\mathbf{x}_{i+1}^{\nu} - \mathbf{x}_i^{\nu}), \quad y_h^{\nu}(t) = \mathbf{y}_{i+1}^{\nu}, \quad \forall\, t \in (t_i, t_{i+1}]. \tag{5.4}$$

THEOREM 5.2. *Suppose that Assumption 2.4 holds and* $\mathrm{int}K(0, x_0) \neq \emptyset$. *Let* (x_h^{ν}, y_h^{ν}) *be defined in* (5.4) *associated with a solution* $\{\mathbf{x}_i^{\nu}, \mathbf{y}_i^{\nu}\}_{i=1}^{N}$ *of* (5.1)-(5.2). *Then there are sequences* $\{\nu_k\}$ *and* $\{h_k\}$ *with* $\nu_k \to \infty$ *and* $h_k \downarrow 0$ *as* $k \to \infty$, *such that*

$$\lim_{\nu_k\to\infty}\lim_{h_k\downarrow 0}\|x_{h_k}^{\nu_k} - x^*\|_s = 0, \ \ w.p.1$$

and

$$\lim_{\nu_k\to\infty}\lim_{h_k\downarrow 0}\|y_{h_k}^{\nu_k} - y^*\|_{L^2} = 0, \ \ w.p.1,$$

where (x^*, y^*) *is a weak solution of* (1.1)-(1.2) *over* $[0, \bar{T}]$.

Proof. According to Theorems 5.1, we get the family of functions $\{x_h^{\nu}(t)\}$ is uniformly bounded on $[0, \bar{T}]$ w.p.1 for sufficiently large ν and sufficiently small h. Moreover, for any $\nu > 0$,

$$\|\mathbf{x}_{i+1}^{\nu} - \mathbf{x}_i^{\nu}\| \leq h\tilde{\rho}_f(1 + \|\mathbf{x}_{i+1}^{\nu}\|) \leq h\tilde{\rho}_f\left(2 + \exp(\frac{\tilde{\rho}_f\bar{T}}{1-h_0\tilde{\rho}_f})(1+\|x_0\|)\right) \triangleq h\hat{\alpha}.$$

Then for any $t \in [t_i, t_{i+1}]$, $\tau \in [t_{i+p}, t_{i+p+1}]$, $i \in \{0, \cdots, N-1\}$ and $p \in \{-i, 1-i, \cdots, N-i-1\}$, we have

$$\begin{aligned}\|x_h^{\nu}(\tau) - x_h^{\nu}(t)\| &= \left\|(x_h^{\nu}(\tau) - \mathbf{x}_{i+p}^{\nu}) + \sum_{j=1}^{p-1}(\mathbf{x}_{i+j+1}^{\nu} - \mathbf{x}_{i+j}^{\nu}) + (\mathbf{x}_{i+1}^{\nu} - x_h^{\nu}(t))\right\| \\ &\leq (\tau - t_{i+p} + (p-1)h + t_{i+1} - t)\hat{\alpha} = |\tau - t|\hat{\alpha}.\end{aligned}$$

It implies that the piecewise interpolant x_h^{ν} is Lipschitz continuous on $[0, \bar{T}]$ and the Lipschitz constant is independent of h and ν. Hence we obtain that $\{x_h^{\nu}(t)\}$ is equicontinuous. Then according to the Arzelá-Ascoli theorem, there are sequences $\{h_k\}$ and $\{\nu_k\}$ with $h_k \downarrow 0$ and $\nu_k \to \infty$ as $k \to \infty$ and an $x^* \in \mathcal{X}_{\bar{T}}$ such that $\lim_{\nu_k\to\infty}\lim_{h_k\downarrow 0}\|x_{h_k}^{\nu_k} - x^*\|_s = 0$ w.p.1.

Let $\hat{y}(t, x(t))$ and $\hat{y}^{\nu}(t, x(t))$ denote the optimal solutions of optimization problems (1.2) and (4.2) with t and $x(t)$, respectively. For any $t \in (t_i, t_{i+1}]$, it is clear that

$\hat{y}^\nu(t_{i+1}, \mathbf{x}_i^\nu) = \mathbf{y}_{i+1}^\nu = y_h^\nu(t)$. Denote $y^*(t) = \hat{y}(t, x^*(t))$. Then for any $t \in (t_i, t_{i+1}]$ with $i \in \{0, \cdots, N-1\}$, we have

$$\begin{aligned}\|y_h^\nu(t) - y^*(t)\| \leq & \|\hat{y}^\nu(t_{i+1}, \mathbf{x}_i^\nu) - \hat{y}(t_{i+1}, \mathbf{x}_i^\nu)\| + \|\hat{y}(t_{i+1}, \mathbf{x}_i^\nu) - \hat{y}(t, x_h^\nu(t))\| \\ & + \|\hat{y}(t, x_h^\nu(t)) - \hat{y}(t, x^*(t))\|.\end{aligned}$$

From the proof of Theorem 4.2, we obtain that $\hat{y}^\nu(t, x(t))$ converges to $\hat{y}(t, x(t))$ w.p.1 as $\nu \to \infty$ for any $t \in \mathbb{R}_+$ and $x(t) \in \mathbb{R}^n$. When $h \downarrow 0$, $t \in (t_i, t_{i+1}]$ and $i \in \{0, \cdots, N-1\}$, it is easy to obtain $\|\mathbf{x}_{i+1}^\nu - x_h^\nu(t)\| \to 0$ w.p.1 for sufficiently large ν from (5.4). Since $\hat{y}$ is continuous, with $h_k \downarrow 0$ and $\nu_k \to \infty$ as $k \to \infty$ such that $\lim_{\nu_k \to \infty} \lim_{h_k \downarrow 0} \|x_{h_k}^{\nu_k} - x^*\|_s = 0$ w.p.1, we obtain

$$\lim_{\nu_k \to \infty} \lim_{h_k \downarrow 0} \|y_{h_k}^{\nu_k} - y^*\|_{L^2} = 0, \text{ w.p.1.}$$

Now we show that (x^*, y^*) is a weak solution of problem (1.1)-(1.2) over $[0, \bar{T}]$. For $x_h^\nu(0) = x_0$ and any $t \in (0, \bar{T}]$ (without loss of generality, we assume $t \in (t_i, t_{i+1}]$ with some $i \in \{0, \cdots, N-1\}$), we have

$$\begin{aligned}&\|\mathcal{W}_h^\nu(t)\| \triangleq \left\| x_h^\nu(t) - x_h^\nu(0) - \int_0^t \frac{1}{\nu} \sum_{\ell=1}^{\nu} f(\tau, x_h^\nu(\tau), y_h^\nu(\tau), \xi_\ell) d\tau \right\| \\ &\leq \frac{1}{\nu} \sum_{\ell=1}^{\nu} \kappa_f(\xi_\ell) \left(\sum_{j=0}^{i-1} \int_{t_j}^{t_{j+1}} \|\mathbf{x}_{j+1}^\nu - x_h^\nu(\tau)\| d\tau + \int_{t_i}^{t} \|\mathbf{x}_{i+1}^\nu - x_h^\nu(\tau)\| d\tau \right. \\ &\left. + \frac{(i+1)h^2}{2} - \frac{(t_{i+1} - t)^2}{2} \right) \leq \frac{h}{\nu} \sum_{\ell=1}^{\nu} \kappa_f(\xi_\ell) \left(\frac{1}{2} \sum_{j=0}^{i-1} \|\mathbf{x}_{j+1}^\nu - \mathbf{x}_j^\nu\| + \|\mathbf{x}_{i+1}^\nu - \mathbf{x}_i^\nu\| + \frac{\bar{T}}{2} \right) \\ &\leq \frac{h(1+\hat{\alpha})\bar{T}}{\nu} \sum_{\ell=1}^{\nu} \kappa_f(\xi_\ell),\end{aligned}$$

where $\kappa_f(\xi)$ is the Lipschitz constant of $f(\cdot, \cdot, \xi)$. Therefore, we conclude that for any $t \in (0, \bar{T}]$ and two sequences $\{h_k\}$ and $\{\nu_k\}$ with $h_k \downarrow 0$ and $\nu_k \to \infty$ as $k \to \infty$,

$$\lim_{\nu_k \to \infty} \lim_{h_k \downarrow 0} \|\mathcal{W}_{h_k}^{\nu_k}\|_s = 0, \text{ w.p.1.}$$

Obviously,

$$\begin{aligned}&\sup_{t \in [0, \bar{T}]} \left\| x^*(t) - x_0 - \int_0^t \mathbb{E}[f(\tau, x^*(\tau), y^*(\tau), \xi)] d\tau \right\| \\ &\leq \lim_{k \to \infty} \left(\sup_{t \in [0, \bar{T}]} \left\| x^*(t) - x_0 - \int_0^t \mathbb{E}[f(\tau, x^*(\tau), y^*(\tau), \xi)] d\tau - \mathcal{W}_{h_k}^{\nu_k}(t) \right\| + \|\mathcal{W}_{h_k}^{\nu_k}\|_s \right) \\ &\leq \lim_{k \to \infty} \left(\left(1 + \frac{\bar{T}}{\nu_k} \sum_{\ell=1}^{\nu_k} \kappa_f(\xi_\ell) \right) \|x^* - x_{h_k}^{\nu_k}\|_s + \left\| \mathbb{E}[f(\cdot, x^*, y^*, \xi)] - \frac{1}{\nu_k} \sum_{\ell=1}^{\nu_k} f(\cdot, x^*, y^*, \xi_\ell) \right\|_s \right. \\ &\left. + \frac{\sqrt{\bar{T}}}{\nu_k} \sum_{\ell=1}^{\nu_k} \kappa_f(\xi_\ell) \|y_{h_k}^{\nu_k} - y^*\|_{L^2} + \|\mathcal{W}_{h_k}^{\nu_k}\|_s \right) = 0,\end{aligned}$$

which implies that (x^*, y^*) is a weak solution of (1.1)-(1.2) over $[0, \bar{T}]$. ◻

For any fixed $i \in \{1, \ldots, N\}$, solving problem (5.1)-(5.2) should address two issues: the nonsmooth fixed point problem and nonsmooth convex optimization problem. For the nonsmooth convex optimization problem (5.2), we can adopt the well-known existing algorithms such as proximal schemes. To solve the nonsmooth fixed point problem (5.1), we can adopt the EDIIS algorithm [6, 7] which is a modified Anderson acceleration. The Anderson acceleration is designed to solve the fixed point problem when computing the Jacobian of the function in the problem is impossible or too costly [2]. We have known that $\frac{h}{\nu}\sum_{\ell=1}^{\nu} f(t_{i+1}, \cdot, \mathbf{y}_{i+1}^{\nu}, \xi_\ell)$ is a contractive mapping w.p.1 for sufficiently large ν and sufficiently small h. Then following from [6, Theorem 2.1], we can obtain that the sequence $\{\mathbf{x}_{i+1}^{(\nu,k)}\}$ generated by the EDIIS algorithm converges to the unique solution $\mathbf{x}_{i+1}^{\nu}$ of (5.1) as the iteration step $k \to \infty$.

6. Numerical experiment. In this section, we verify our theoretical results by a numerical example, which is performed in MATLAB 2017b on a Lenovo laptop (2.60GHz, 32.0GB RAM).

Example 6.1. We consider the following problem:

$$\begin{aligned} &\dot{x}(t) = \mathbb{E}\left[\begin{pmatrix} \xi_1 & 2 \\ \xi_1^2 & \xi_2 \end{pmatrix} x(t) + \begin{pmatrix} 2x_1(t) & x_2(t) & \xi_2 \\ 2t & 0 & \xi_1 x_1(t) \end{pmatrix} y(t)\right], \\ &y(t) \in \arg\min_{\mathbf{y}\in\mathbb{R}^3} \mathbb{E}[\|M(\xi)\mathbf{y} - b(x(t),\xi)\|^2 + \|\mathbf{y}\|_1] \\ &\qquad\qquad \text{s.t. } \mathbf{y} \in K(t, x(t)) = \{\mathbf{y} : \mathbb{E}[A(\xi)]x(t) + \mathbb{E}[B(\xi)]\mathbf{y} + \mathbb{E}[Q(t,\xi)] \le 0\}, \end{aligned} \tag{6.1}$$

where $x(t) = (x_1(t), x_2(t))^\top$, $x(0) = x_0 = (-1, -2)^\top$,

$$\begin{aligned} &M(\xi) = \begin{pmatrix} 2+\xi_1 & 0 & -\xi_2 \\ 0 & \xi_1+\xi_2 & -1 \end{pmatrix}, \qquad b(x(t),\xi) = \begin{pmatrix} x_1(t) + \xi_2 \\ \xi_1 x_2(t) \end{pmatrix}, \\ &A(\xi) = \begin{pmatrix} -2-\xi_1 & 1 \\ -1 & \xi_2 \end{pmatrix},\ B(\xi) = \begin{pmatrix} 1 & \xi_1^2 & \xi_2 \\ \xi_2 & 0 & 2 \end{pmatrix},\ Q(t,\xi) = \begin{pmatrix} t-\xi_1 \\ \xi_2 \end{pmatrix}. \end{aligned}$$

We set the terminal time $T = 1$, $\xi_1 \sim \mathcal{N}(1, 0.01)$ and $\xi_2 \sim \mathcal{U}(-1, 1)$. It can be verified easily that all functions in this example fulfill our settings in the beginning of this paper. It is obvious that the objective function $\mathbb{E}[g(t, x(t), \cdot)]$ in (6.1) is convex and $\mathbb{E}[g(0, x_0, \mathbf{y})] \ge \|\mathbf{y}\|_1$, which means that $\mathbb{E}[g(0, x_0, \cdot)]$ is level-coercive and then is level-bounded, following from [24, Corollary 3.27]. We can also obtain that $K(0, x_0) = \{(\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3) : \mathbf{y}_1 + 1.01\mathbf{y}_2 \le 0, 2\mathbf{y}_3 + 1 \le 0\}$ and then $\mathrm{int}K(0, x_0) \ne \emptyset$. Hence, we know that Assumption 2.2 holds for this example.

Now we illustrate that this example exists a solution on $[0, 1]$. Following from the proof of Theorem 2.3, the solution existing interval mainly depends on the range of $(t, \mathbf{x})$ which is such that $\mathrm{int}K(t, \mathbf{x}) \ne \emptyset$ and $\mathbb{E}[g(t, \mathbf{x}, \cdot)]$ is level-bounded. By $K(t, \mathbf{x}) = \{(\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3) : \mathbf{y}_1 + 1.01\mathbf{y}_2 \le 3\mathbf{x}_1 - \mathbf{x}_2 - t + 1, 2\mathbf{y}_3 \le -\mathbf{x}_1\}$, we know that for any $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$ there always holds that $\mathrm{int}K(t, \mathbf{x}) \ne \emptyset$ because two linear constraints are independent of each other. For any $(t, \mathbf{x})$, we can also have $\mathbb{E}[g(t, \mathbf{x}, \mathbf{y})] \ge \|\mathbf{y}\|_1$, which means that $\mathbb{E}[g(t, \mathbf{x}, \cdot)]$ is also level-coercive and then is level-bounded. Hence, we know that the optimal set $\mathcal{S}(t, \mathbf{x})$ of the optimization problem in (6.1) is nonempty and bounded for any $(t, \mathbf{x})$, and is upper semicontinuous. It then derives that (6.1) has at least a weak solution on $[0, 1]$ from Theorem 2.3.

We add a regularization term $\mu\|\mathbf{y}\|^2$ to the objective function in (6.1). For the regularization numerical form of (6.1), we use the EDIIS(1) method to solve the fixed point problem and the Matlab toolbox CVX to obtain the optimal solution of the convex optimization problem. The EDIIS(1) method is used in the numerical example of [7] since the minimization problem in EDIIS has a closed-form solution in this case. The stop criterion of EDIIS(1) for each i is $\|\mathbf{x}^{(\nu,k+1)} - \mathbf{x}^{(\nu,k)}\| \leq 10^{-6}$.

In the numerical experiments, let $\hat{x} = (\hat{x}_1, \hat{x}_2)$ be a numerical solution of the ODE in (6.1) with regularization parameter $\mu = 10^{-5}$, sample size $\nu = 5000$ and step size $h = 10^{-4}$. For the fixed step size $h = 10^{-4}$, we carry out tests with the regularization parameter $\mu = 10^{-4}$, 0.001, 0.01 and 0.1, and the sample size $\nu = 3000$, 2000, 1000 and 500. We compute the numerical solution $x^{\mu,\nu} = (x_1^{\mu,\nu}, x_2^{\mu,\nu})$ and

$$R_1 = \frac{1}{10000} \sum_{i=1}^{10000} |\hat{x}_1(ih) - x_1^{\mu,\nu}(ih)|, \;\; R_2 = \frac{1}{10000} \sum_{i=1}^{10000} |\hat{x}_2(ih) - x_2^{\mu,\nu}(ih)|$$

50 times and averages them. The decreasing tendencies of R_1 and R_2 as ν increases and μ decreases are shown in FIG. 1.

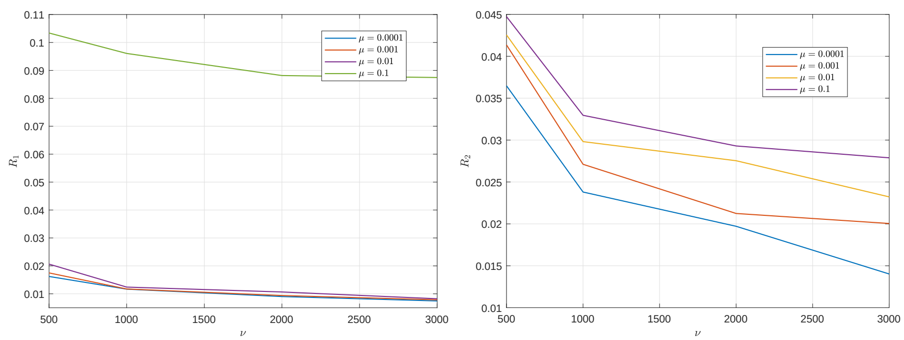

FIG. 1. *The decreasing tendencies of R_1 and R_2 as ν increases and μ decreases.*

7. Application in time-varying parameter estimation for an ODE. In this section, we apply (1.1)-(1.2) to estimate the time-varying parameter for an ODE. Several strategies can be employed to estimate the time-varying parameters for an ODE based on noisy data, such as the local polynomial method [5], the nonlinear least squares method [19] and the spline-based method [18]. As a Bayesian approach, Gaussian process is also widely used to infer dynamics of ODE (see [32] and the references therein). A Gaussian process can be viewed as a distribution over functions, while its inference takes place directly in the function space. It is a collection of random variables, any finite number of which have a joint Gaussian distribution. It is also a non-parametric probabilistic model for function estimation that is widely used in tasks such as regression and classification. Therefore, we use Gaussian process to infer the dynamics of an ODE based on noisy data in order to estimate its time-varying parameter.

A system of ODE with initial value $x(0) = x_0$ takes the form

$$\dot{x}(t) = f_1(x(t))y(t) + f_2(x(t)), \;\; t \in [0, T], \tag{7.1}$$

where $f_1 : \mathbb{R}^n \to \mathbb{R}^{n \times m}$, $f_2 : \mathbb{R}^n \to \mathbb{R}^n$ are given functions and $y(t) \in \mathbb{R}^m$ is unknown. It is well known that a Gaussian process is completely characterized by a mean function and a covariance or a kernel function.

We assume that we can observe the values of states and their derivatives at the given time points $\{t_i\}_{i=1}^N$. If the observation data of derivatives is not available, we can estimate the derivatives by the first-order method, that is $\dot{x}_j(t_i) \approx \frac{x(t_{i+1})-x(t_i)}{t_{i+1}-t_i}$. Let $Y_i = [Y_i^1, \cdots, Y_i^n]^\top$ be the measurement of true value of state variable x at time t_i, that is $Y_i^j = x_j(t_i) + \epsilon_j$ for $j = 1, \cdots, n$, where ϵ_j denotes the measurement error. We also let $Z_i^j = \dot{x}_j(t_i) + \epsilon_j^d$ for given derivatives observation or $Z_i^j = \frac{Y_{i+1}^j - Y_i^j}{t_{i+1}-t_i}$, where ϵ_j^d denotes the measurement error. The errors ϵ_j and ϵ_j^d are assumed to follow a Gaussian distribution with zero mean and variance σ_j^2 and $\hat{\sigma}_j^2$, respectively.

We employ the Gaussian processes to obtain the distributions of the state variables and their derivatives, denoted as $\hat{x}(t,\xi)$ and $\dot{\hat{x}}(t,\xi)$, where ξ denotes a random variable. It should be noticed that $\hat{x}(t,\xi)$ and $\dot{\hat{x}}(t,\xi)$ have no closed forms but can obtain their values at any given t. By [29], the n-dimensional variable of $\hat{x}(t,\xi)$ and $\dot{\hat{x}}(t,\xi)$ can be obtained by stacking n independent Gaussian processes to model each state and the derivative independently.

Therefore, we can estimate the time-varying coefficients $y(t)$ by solving the following optimization problem

$$(7.2)\qquad y(t) \in \arg\min_{\mathbf{y}\in\mathbb{K}} \mathbb{E}\left[\|\dot{\hat{x}}(t,\xi) - f_1(x(t))\mathbf{y} - f_2(x(t))\|_1 + \|\hat{x}(t,\xi) - x(t)\|_1\right],$$

where $x(t)$ fulfills (7.1), and $\mathbb{K}$ is a nonempty closed convex set which can be some inaccurate information for the coefficients such as upper or lower bounds. Obviously, the objective function in (7.2) is nonsmooth and convex in $\mathbf{y}$. Note that the objective function is not strongly convex in $\mathbf{y}$, then we introduce the regularization method into it. By using the regularization method with parameter μ, SAA with sample size ν and time-stepping method with step size h, we obtain the following discrete form of (7.1)-(7.2):

$$(7.3)\qquad \mathbf{x}_{i+1} = \mathbf{x}_i + h(f_1(\mathbf{x}_{i+1})\mathbf{y}_{i+1} + f_2(\mathbf{x}_{i+1})),$$

$$(7.4)\qquad \mathbf{y}_{i+1} = \arg\min_{\mathbf{y}\in\mathbb{K}} \frac{1}{\nu}\sum_{\ell=1}^{\nu} \|\dot{\hat{x}}(t_{i+1},\xi_\ell) - f_1(\mathbf{x}_i)\mathbf{y} - f_2(\mathbf{x}_i)\|_1 + \mu\|\mathbf{y}\|^2.$$

It should be noted that $\dot{\hat{x}}(t_{i+1},\xi_\ell)$ does not need any information of $\mathbf{x}_i$ and $\mathbf{x}_{i+1}$ in (7.4), as it is obtained by the Gaussian process based on the observation data independently. As we mentioned before, we adopt EDIIS algorithm to solve the fixed point problem (7.3). For the nonsmooth convex optimization problem (7.4), we use the CVX tool box to solve it. At last, we can obtain the approximation solution of (7.1)-(7.2) and the estimation of time-varying coefficients.

Example 6.2. For the following ODE with time-varying coefficients,

$$(7.5)\qquad \begin{aligned} \dot{x}_1(t) &= x_1(t) + \sin(t)x_1(t), \\ \dot{x}_2(t) &= x_1(t) - 2tx_2(t), \quad t \in [0,5], \end{aligned}$$

where $x_1(0) = 1$ and $x_2(0) = 0$. Let $t_i = 0.04i$, $i = 0, \cdots, 125$. Obviously, for any given t_i, we can obtain the values $x(t_i)$ and then $\dot{x}(t_i)$. We estimate the parameters $(\sin(t), -2t)$ of (7.5) under two cases: (i) both the noisy data of $Y_i = x(t_i) + \epsilon$ and $Z_i = \dot{x}(t_i) + \epsilon$ are given, where $\epsilon \sim \mathcal{N}(0, 0.4)$; (ii) only the noisy data Y_i is given.

We estimate the time-varying parameters of (7.5) by solving the problem (7.5) with an optimization problem (7.2), where we estimate the parameters with the set $\mathbb{K} = \{(\mathbf{y}_1, \mathbf{y}_2) : -1 \le \mathbf{y}_1 \le 1, -10 \le \mathbf{y}_2 \le 0\}$. When we adopt the regularization approach, SAA method and the time-stepping method, we set the regularization parameter $\mu = 10^{-4}$, the sample size $\nu = 1000$ and step size $h = 0.001$. For problem (7.3), we also adopt EDIIS(1), where the stop criterion for each i is $\|\mathbf{x}_i^{(k+1)} - \mathbf{x}_i^{(k)}\| \leq 10^{-6}$. We obtain the estimation of parameters by averaging 50 independent repetitions. The visualization of estimates of parameters $(\sin(t), -2t)$ in (7.5) for the two cases are shown in FIGs. 2 and 3. Let $y(t)$ and $\tilde{y}(t)$ denote the true functions and their estimations, respectively. In the figures, the dash lines denote the 95% simultaneous l_∞ credible bands, where the radius is estimated by the 95% quantile of $\|y - \tilde{y}\|_s \triangleq \max_i |y(t_i) - \tilde{y}(t_i)|$. The shaded area denotes the estimation area between the 25% and 75% quantiles of 50 independent repetitions.

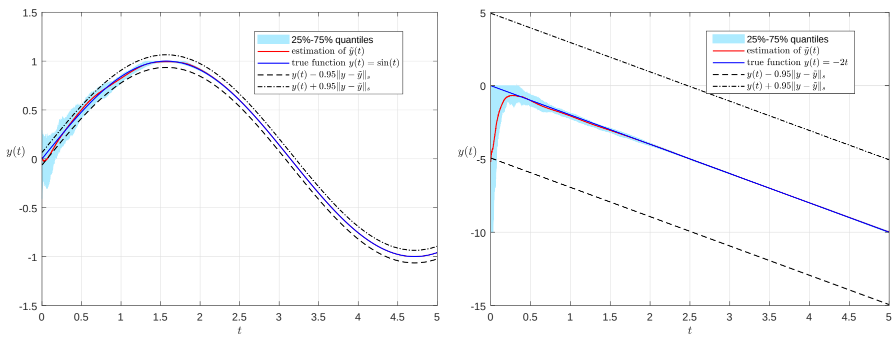

Fig. 2. *Visualization of estimates of parameters* $(\sin(t), -2t)$ *in* (7.5) *under case (i).*

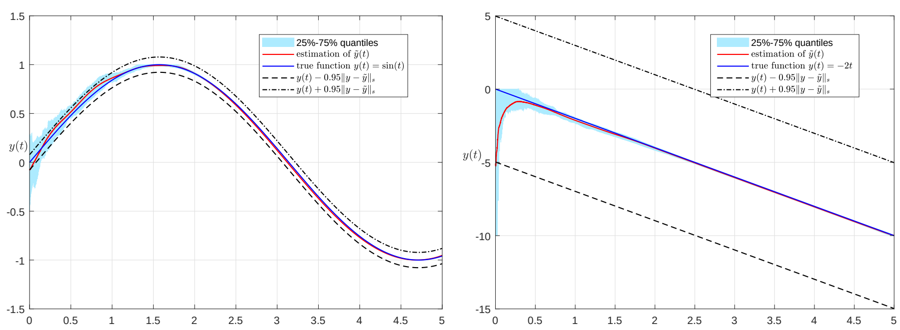

Fig. 3. *Visualization of estimates of parameters* $(\sin(t), -2t)$ *in* (7.5) *under case (ii).*

Under the case (i), for the fixed step size $h = 0.001$, we also carry out tests with the regularization parameter $\mu = 0.001$, 0.01, 0.1 and 1, and the sample size $\nu = 1000$, 500, 100 and 50. We compute the numerical solution $x^{\mu,\nu} = (x_1^{\mu,\nu}, x_2^{\mu,\nu})$ and

$$R_3 = \frac{1}{5000}\sqrt{\sum_{i=1}^{5000}(x_1^*(ih) - x_1^{\mu,\nu}(ih))^2},\ R_4 = \frac{1}{5000}\sqrt{\sum_{i=1}^{5000}(x_2^*(ih) - x_2^{\mu,\nu}(ih))^2}$$

50 times and averages them, where $(x_1^*(t), x_2^*(t))$ is the true solution of problem (7.5)

$$x_1^*(t) = e^{t-\cos(t)+1}, \quad x_2^*(t) = e^{1-t^2}\int_0^t e^{-\tau^2-\cos(\tau)+\tau}d\tau.$$

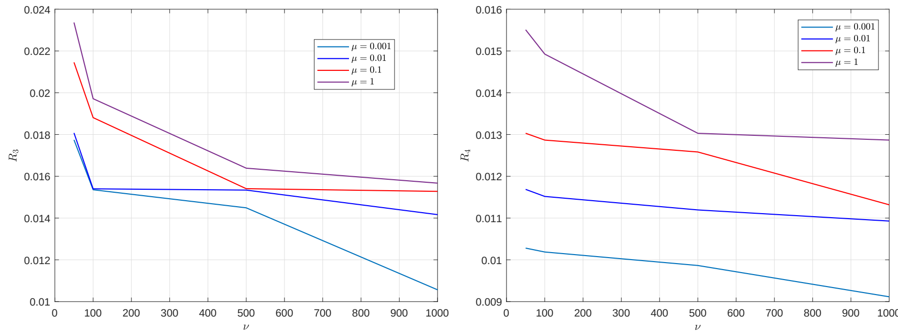

FIG. 4. *The decreasing tendencies of R_3 and R_4 as ν increases and μ decreases for (7.5) under case (i).*

The decreasing tendencies of R_3 and R_4 as ν increases and μ decreases are shown in FIG. 4.

From FIGs 2 and 3, we can observe that our model (7.1)-(7.2) can be applied to approximate the time-varying parameters in an ODE system (7.5), which means the potential application of (1.1)-(1.2) in estimating the time-varying parameters in ODE system. FIG 4 also verifies the theoretical results for our numerical methods proposed by this paper.

8. Conclusions. In this paper, we show the existence of weak solutions of the dynamic system coupled with solutions of stochastic nonsmooth convex optimization problem (1.1)-(1.2). By adding a regularization term $\mu\|\mathbf{y}\|^2$ to the convex objective function in (1.2), the convex optimization problem becomes a strongly convex problem, which has a unique continuous optimal solution. We show that the unique optimal solution of nonsmooth optimization with strong convexity admits a linear growth condition and the regularized dynamic system has a classic solution. Moreover, we prove that the solutions of regularized problem converge to the solutions of original problem as the regularization parameter goes to zero. Moreover, we show that the unique optimal solution of the regularized optimization problem (3.1) converges to the least-norm optimal solution of the original problem (1.2). We adopt the sample average approximation scheme and implicit Euler method to discretize the dynamic system coupled with solutions of stochastic nonsmooth strongly convex optimization problem and present the corresponding convergence analysis. We give a numerical example to demonstrate our theoretical results. Finally, the effectiveness of our model is verified by an example of the estimation of the time-varying parameters in ODE.

Acknowledgement We would like to thank the Associate Editor and two referees for their very helpful comments.

REFERENCES

[1] J. Aubin and A. Cellina, *Differential Inclusions*, Springer-Verlag, New Nork, 1984.
[2] W. Bian and X. Chen, *Anderson acceleration for nonsmooth fixed point problems*, SIAM J. Numer. Anal., 60 (2022), pp. 2568–2591.
[3] B. Brogliato and A. Tanwani, *Dynamical systems coupled with monotone set-valued operators: Formalisms, applications, well-posedness, and stability*, SIAM Rev., 62 (2020),

pp. 3–129.
[4] A. Caboussat, C. Landry, and J. Rappaz, *Optimization problem coupled with differential equations: A numerical algorithm mixing an interior-point method and event detection*, J. Optim. Theory Appl., 147 (2010), pp. 141–156.
[5] J. Chen and H. Wu, *Efficient local estimation for time-varying coefficients in deterministic dynamic models with applications to HIV-1 dynamics*, J. Am. Statist. Assoc., 103 (2008), pp. 369–384.
[6] X. Chen and T. Kelley, *Convergence of the EDIIS algorithm for nonlinear equations*, SIAM J. Sci. Comput., 41 (2019), pp. 365–379.
[7] X. Chen and J. Shen, *Dynamic stochastic variational inequalities and convergence of discrete approximation*, SIAM J. Optim., 32 (2022), pp. 2909–2937.
[8] X. Chen and Z. Wang, *Computational error bounds for differential linear variational inequality*, IMA J. Numer. Anal., 32 (2012), pp. 957–982.
[9] X. Chen and Z. Wang, *Convergence of regularized time-stepping methods for differential variational inequalities*, SIAM J. Optim., 23 (2013), pp. 1647–1671.
[10] X. Chen and Z. Wang, *Differential variational inequality approaches to dynamic games with shared constraints*, Math. Program., 146 (2014), pp. 379–408.
[11] M. Chung, J. Krueger, and M. Pop, *Identification of microbiota dynamics using robust parameter estimation methods*, Math. Biosci., 294 (2017), pp. 71–84.
[12] K. Deimling, *Multivalued Differential Equations*, Walter de Gruyter, Berlin, 1992.
[13] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer-Verlag, New York, 2003.
[14] L. Han, A. Tiwari, M. K. Camlibel, and J.-S. Pang, *Convergence of time-stepping schemes for passive and extended linear complementarity systems*, SIAM J. Numer. Anal., 47 (2009), pp. 3768–3796.
[15] R. Janin and J. Gauvin, *Lipschitz-type stability in nonsmooth convex programs*, SIAM J. Control Optim., 38 (1999), pp. 124–137.
[16] C. Landry, A. Caboussat, and E. Hairer, *Solving optimization-constrained differential equations with discontinuity points, with application to atmosphereic chemistry*, SIAM J. Sci. Comput., 31 (2009), pp. 3806–3826.
[17] S. Lang, *Real and Fuctional Analysis, 3rd edn*, Springer, Berlin, 1993.
[18] H. Liang and H. Wu, *Parameter estimation for differential equation models using a framework of measurement error in regression models*, J. Am. Statist. Assoc., 103 (2008), pp. 1570–1583.
[19] D. Lin and Z. Ying, *Semiparametric and nonparametric regression analysis of longitudinal data*, J. Am. Statist. Assoc., 96 (2001), pp. 103–126.
[20] J. Luo and X. Chen, *An optimal control problem with terminal stochastic linear complementarity constraints*, SIAM J. Control Optim., 61 (2023), pp. 3370–3389.
[21] J. Luo, X. Wang, and Y. Zhao, *Convergence of discrete approximation for differential linear stochastic complementarity systems*, Numer. Algorithms, 87 (2021), pp. 223–262.
[22] J.-S. Pang and D. E. Stewart, *Differential variational inequalities*, Math. Program., 113 (2008), pp. 345–424.
[23] J. A. Paulson, M. Martin-Casas, and A. Mesbah, *Fast uncertainty quantification for dynamic flux balance analysis using non-smooth polynomial chaos expansions*, PLoS Comput. Biol., 15 (2019), p. e1007308.
[24] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis, 3rd*, Springer-Verlag, Heidelberg, 2009.
[25] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory, 2nd*, SIAM, Philadelphia, 2014.
[26] Y. Terazono and A. Matani, *Continuity of optimal solution functions and their conditions on objective functions*, SIAM J. Optim., 25 (2015), pp. 2050–2060.
[27] W. Walter, *Ordinary Differential Equations*, Springer-Verlag, New Nork, 1998.
[28] Z. Wang and X. Chen, *An exponential integrator based discontinuous Galerkin method for linear complementarity systems*, IMA J. Numer. Anal., 38 (2018), pp. 2145–2165.
[29] P. Wenk, G. Abbati, M. A. Osborne, B. Schölkopf, A. Krause, and S. Bauer, *ODIN: ODE-informed regression for parameter and state inference in time-continuous dynamicals systems*, AAAI 2020, pp. 6364–6371.
[30] S. Wu and X. Chen, *A parallel iterative algorithm for differential linear complementarity problems*, SIAM J. Sci. Comput., 39 (2017), pp. A3040–A3066.
[31] S. Wu, T. Zhou, and X. Chen, *A Gauss-Seidel type method for dynamic nonlinear complementarity problems*, SIAM J. Control Optim., 58 (2020), pp. 3389–3412.
[32] S. Yang, S. W. K. Wong, and S. C. Kou, *Inference of dynamic systems from noisy and*

sparse data via manifold-constrained Gaussian processes, Proc. Natl. Acad. Sci. USA, 118 (2021), p. e2020397118.

[33] X. Zhao, S. Noack, W. Wiechert, and E. von Lieres, *Dynamic flux balance analysis with nonlinear objective function*, J. Math. Biol., 75 (2017), pp. 1487–1515.