RESEARCH STATEMENT

HONGJIANG QIAN

My research effort has been devoted to numerical approximation in stochastic systems, large deviations, and moderate deviations. I am interested in developing new numerical schemes, proving their convergence, and ascertaining their convergence rates, which mainly includes numerical approximation for stochastic differential equations (SDEs) and numerics for stochastic control and stochastic filtering problems. For large and moderate deviations theory, I am interested in establishing large and moderate deviation principle for stochastic differential equations, stochastic partial differential equations (SPDEs) under both large time and small diffusion settings. In what follows, I briefly summarize my research work to date and state my future research plans.

1. Research work to date

In this section, my research projects conducted during my Ph.D. program supervised by Professor George Yin are mainly in two directions: numerics in stochastic dynamics, large and moderate deviations theory.

1.1. **Numerics in stochastic dynamic systems.**

Stochastic differential equations. For highly nonlinear stochastic differential equations, it is virtually impossible to obtain their analytic solutions. Thus numerical approximation becomes vitally important. There has been a vast literature devoted to numerical solutions of stochastic differential equations, their convergence, and their convergence rates. The most common practice is to obtain strong and weak convergence. If the drift and diffusion coefficients are Lipschitz, the classical Euler-Maruyama (EM) method converges to the true solution with strong order $1/2$; see [25, Theorem 10.2.2]. Taking advantage of Itô formula to obtain a better approximation, Milstein [30] improves the convergence rate to be of order 1 under more restrictive condition. However, in applications, there are many SDEs having less regular coefficients. It is well recognized that one has to deal with non-Lipschitz and superlinear coefficients. Hutzenthaler et al. [21] introduced the tamed Euler-Maruyama scheme and proved its strong order $1/2$ convergence with one-sided Lipschitz drift coefficient and the linear growth diffusion coefficient; this was further developed in [42]. In addition, the stopped EM method was developed in [27] for nonlinear SDEs as well. Later, the "Drift-truncated Euler scheme" was considered in [20]. The truncated Euler-Maruyama scheme was proposed in [29], and strong convergence was analyzed under local Lipshitz and Khasminskii-type condition. The essence of both tamed EM or truncated EM schemes lies in the idea of using projections and truncation to deal with the highly nonlinearity appeared in coefficients.

The classical Euler-Maruyama schemes might lead to numerical solutions with finite explosion time. Thus some truncation or projection algorithms are used to confine the estimates to be in a certain bounded region. In line with such idea, in [37], we proposed a novel numerical scheme, established the weak convergence using the martingale problem formulation, and proved the weak convergence rate with order $1/2-\varepsilon$ for $\varepsilon > 0$. A distinct feature of our approach is a random truncation region instead of a fixed one is used to bound the sequence of EM numerical approximations of SDEs. This idea stems from stochastic approximation algorithms aiming to solve root-finding or optimization problems under noisy measurements. In this direction of stochastic approximation, see the work of Chen and Zhu [6] by making the sequence of approximation return to a fixed point

Department of Mathematics, University of Connecticut, Storrs, CT 06269, hongjiang.qian@uconn.edu.

and the work of Andradóttir in [1] by projecting the sequence of estimates onto a sequence of increasing sets. Compared to the existing algorithm such as truncated or tamed EM schemes, main advantage of our algorithm is that we need not modify the drift or the diffusion coefficient of the SDE.

Stochastic control. Stochastic Kolmogorov systems have been widely used in biology, ecology, social networking, and statistical physics. Some typical examples are SIS (susceptible-infectious-susceptible) model, stochastic chemostat models, and replicator dynamics.

In [48], we developed numerical solutions to the controlled stochastic Kolmogorov systems by virtue of the Markov chain approximation method; we proved convergence of the solution, the cost, and the value function. For more details about Markov chain approximation method, we refer to [24] and references therein. Subsequently, we incorporated Markovian switching and random jumps into consideration. The Markovian switching is used to describe discrete events that coexist with continuous dynamics. Hybrid or switching diffusions are capable of describing complex systems and their inherent uncertainty and randomness in the environment, whereas Poisson jump processes are suitable to model sudden environment changes that cannot be modeled by the usual stochastic differential equations or regime-switching stochastic differential equations. For example, in population dynamics, the population may experience environmental changes such as earthquakes, hurricanes, epidemics, etc. In [37], we constructed numerical schemes using Markov chain approximation method to approximate the controlled switching jump stochastic Kolmogorov systems. Convergence is established, and numerical experiments are exhibited to demonstrate our numerical methods. Take the controlled Lotka-Volterra system in [37] as an example,

$$\begin{cases} dx(t) = x(t)[a(r(t)) - b(r(t))y(t) + u(t)]dt + \alpha(r(t))x(t)dW_1(t) \\ \qquad\qquad + \displaystyle\int_{\mathbb{Y}} \gamma_1(x(t), y(t), r(t), u(t), \eta)N(dt, d\eta) \\ dy(t) = y(t)[-c(r(t)) + d(r(t))x(t) + u(t)]dt + \beta(r(t))y(t)dW_2(t) \\ \qquad\qquad + \displaystyle\int_{\mathbb{Y}} \gamma_2(x(t), y(t), r(t), u(t), \eta)N(dt, d\eta) \\ x(0) = x_0,\ y(0) = y_0,\ r(0) = \iota \end{cases}$$

The objective is to minimize the expected cost

$$J(x_0, y_0, \iota, u(\cdot)) = \mathbb{E}_{x_0,y_0,\iota} \int_0^\tau g(x(s), y(s), r(s), u(s))ds + \widetilde{g}(x(\tau), y(\tau), r(\tau)),$$

where τ is the first exit time of $O = (0, 10) \times (0, 10)$ in our numerical experiment. In addition, for simplicity, we take $\gamma_1(\cdot) = \gamma_2(\cdot) = \eta$, the control $u(t) = \{t/5, 0 \le t \le 10\}$ and let

$$g(\cdot) = 1 + r(t)[x(t) + y(t)](1 + u^2(t)), \quad \widetilde{g}(\cdot) = 1 + r(t)[x(t) + y(t)].$$

The size of jump follows the exponential distribution with density function $c(\eta) = se^{-s\eta}$ where $s = 0.1$, the jump time follows the exponential distribution with mean 10. The Markov chain $r(t) \in \{1, 2\}$ with generator $Q = [-0.5, 0.5; 0.5, -0.5]$. We take the step size $h = 0.01$, tolerance level 10^{-8} and

$$\begin{array}{l} a(1) = 0.6, a(2) = 0.8;\ b(1) = 0.5, b(2) = 0.3 \\ c(1) = 0.45, c(2) = 0.5;\ d(1) = 0.65, d(2) = 0.8. \end{array}$$

Then the approximating value function and optimal control are shown as Figure 1 and Figure 2.

Computational nonlinear filtering. The filtering problem is concerned with state estimation based on partial observation of the system state, given the state of a system being not completely observable. The purpose of nonlinear filtering is to find the "best estimate" for the true state of a system. For a state process x_t and an observation process y_t, it focuses on computing the conditional mean or the conditional distribution given the information of the observation process up to time t. Theoretically, it shows that the conditional mean satisfies the so-called Kushner equation [23]. Later,

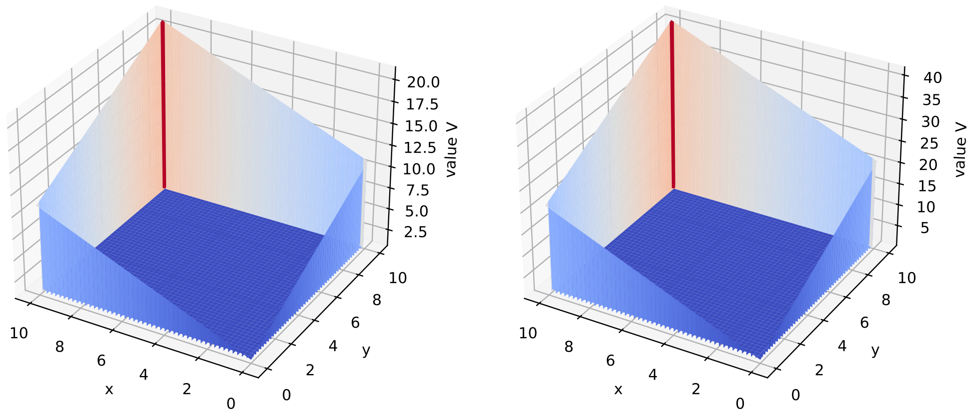


FIGURE 1. The value function in regime 1 (left) and in regime 2 (right).

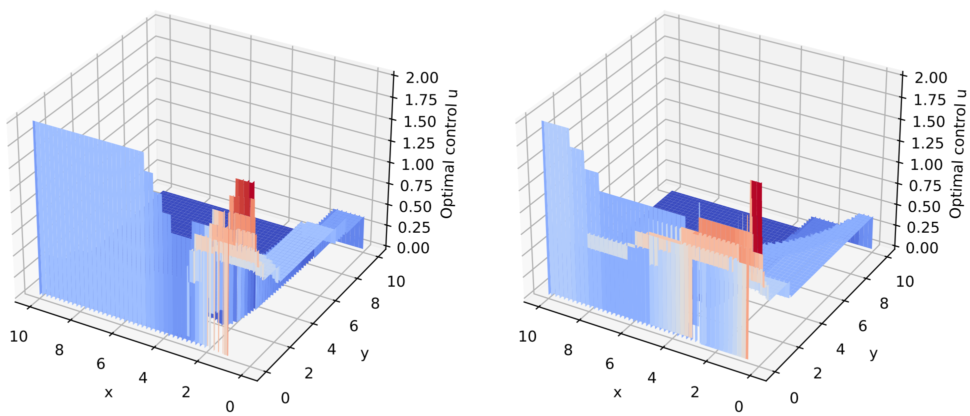


FIGURE 2. The optimal control u in regime 1 (left) and in regime 2 (right).

Duncan, Mortensen, and Zakai independently worked on the utilization of unnormalized conditional distribution (that is, the conditional distribution is not a probability measure). The equation has been referred as the Duncan-Mortensen-Zakai equation. To approximate the infinite-dimensional systems, one has to face the inherent "curse of dimensionality".

The recent advent of machine learning and neural network (NN) provides an alternative for us to solve the filtering problem numerically. By virtue of the NN, we are able to solve it in a finite-dimensional space by finding the optimal weights of NN. We refer the work of Nielsen [33], LeCun et al. [26], and reference therein for an introduction to deep learning and applications. Recently, a new method named deep filter was studied by Wang et al. in [47]. By generating Monte Carlo samples, they used the observation data to train the deep neural network (DNN) and used the state trajectories as the target. It turns out that results of deep filter compare favorably to that of the classical Kalman filter, and this method can also deal with regime-switching systems which cannot be treated by traditional filtering approaches. In [39], we further investigated the deep filtering with adaptive learning rates. The training of DNN is normally executed by back-propagation with constant learning rates (LRs). Thus the choice of learning rate is important to affect training effects. If the learning rate is too small, it may take forever to converge; if too large, it tends to overshoot leading to oscillation divergence. Several strategies are made by machine learning group.

For example, time-based and step-based learning rate schedules were used in [32], adaptive learning rate strategies using explicit functions of the gradient of the loss function like RMSprop and Adam can be found in [41], and adaptive learning rate proportional to the loss function was studied by Park et al. in [35]. We note the potential drawback of the work of Park et al. lies its reliance on the magnitude of the loss function. In [39], our approach focuses on fully recursive stochastic gradient descent algorithms and can be adaptive to various noise level. In contrast to existing literature, we proposed the following algorithm by adaptively updating the learning rates:

$$\begin{aligned} &\theta_{k+1}^{n\ell} = \theta_k^{n\ell} - \rho_n \nabla_\theta J(\theta_k^{n\ell}, \xi_k^{n\ell}),\ \ k = 0, \dots, \ell - 1, \\ &\rho_{n+1} = \rho_n - \varepsilon \widehat{G}_n, \end{aligned}$$

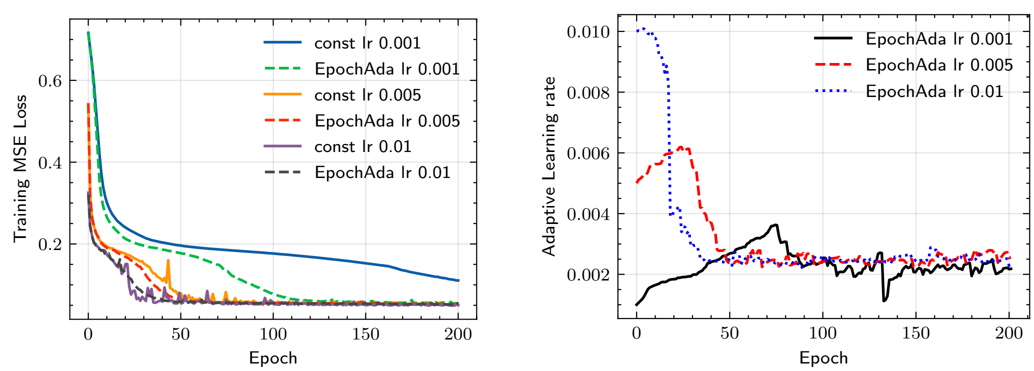
where θ is the neural network parameter, $\nabla_\theta J(\theta_k^{n\ell}, \xi_k^{n\ell})$ denotes the noisy gradients of J w.r.t. θ, $\varepsilon > 0$ is a small parameter serving as a step size, and $\widehat{G}_n$ denotes a sequence of n-dependent estimates of the partial derivative w.r.t. ρ of the loss function $\bar{\chi}(\rho, \theta_e)$. We proved the convergence of the algorithms making use of stochastic averaging and martingale methods. Numerically, we examined the robustness and efficiency of the algorithm using several highly nonlinear dynamics involved with random switching. Take a two dimensional nonlinear model as example, consider

$$\begin{cases} x_{n+1} = x_n + \eta \begin{bmatrix} \sin[(0.3x_n^0 + 0.5x_n^1)\alpha_n] \\ \sin(0.3x_n^1 \alpha_n) \end{bmatrix} + \sqrt{\eta}\sigma w_n, \\ y_{n+1} = y_n + \eta G x_n + \sqrt{\eta}\sigma_1 v_n \end{cases}$$

where $x_n = (x_n^0, x_n^1)'$, the Markov chain $\alpha_n \in \{1, 2\}$, and

$$\sigma = \begin{bmatrix} 1 & -0.3 \\ 0 & 1 \end{bmatrix}, G = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \sigma_1 = \begin{bmatrix} 0.2 & 0.05 \\ 0 & 0.2 \end{bmatrix}.$$

The initial state is chosen as $x_0 = (1, -1)'$. The training loss function with constant LRs (learning rates) $(0.001, 0.005, 0.01)$ as well as adaptive LRs with such initial rates is given in Figure 3(a). The adaptive LRs appears to converges to 0.0025 shown in Figure 3(b). The sample paths of the state x_n and the corresponding out-of-sample deep filtering sample paths with constant LR 0.0025 and adaptive LR with $\rho_0 = 0.0025$ are shown in Figure 4.



(a) (b)

FIGURE 3. The loss function (left) and paths of the adaptive learning rates with initials 0.001, 0.005 and 0.01 (right).

Subsequently, we also considered the filtering problem with degenerate observation noise where the Kalman-Bucy filter fails; see [40]. Traditionally, the Kalman-Bucy filter requires the gain matrix to be invertible, which does not hold when the observation noise is degenerate. Thus we replaced

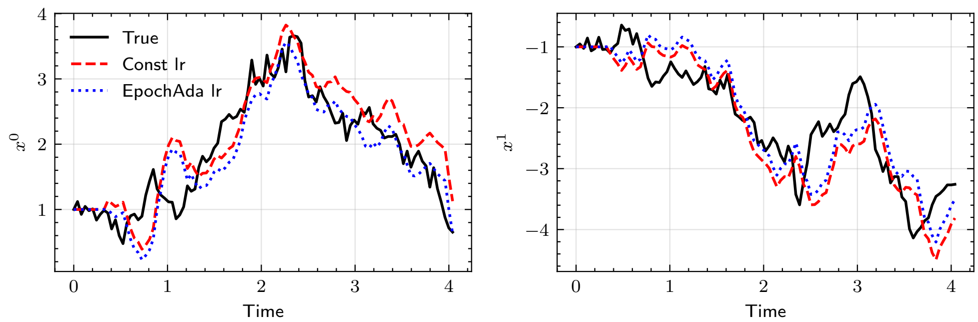


FIGURE 4. Sample paths of out-of-sample state and the sample paths of the deep filters with constant LR $\rho = 0.0025$ and adaptive LR with initial $\rho_0 = 0.0025$.

it with a matrix-valued parameter and used the stochastic approximation method to find the best gain matrix parameter. Namely, we consider the filtering scheme

$$d\widehat{X}_t = f(\widehat{X}_t)dt + R(dY_t - h(\widehat{X}_t)dt), \quad \widehat{X}_0 = \mathbb{E}X_0$$

where $\widehat{X}_t$ denote an estimate of X_t, and R is the constant gain matrix parameter. Thus, we obtained a scheme to solve the filtering problem with degenerate noise numerically. This advances the literature in that in the past only nonsingular diffusion matrices can be treated.

1.2. **Large and moderate deviations theory.** Since the pioneer work of Freidlin [17], much of effort are put to the study of large deviation principle (LDP) for stochastic dynamical systems with small diffusion. In [5], Cerrai and Freidlin studied the large deviation principle for the Langevin equation with strong damping by using the integration by parts formula. Later, Nguyen and Yin [34] extended the above results by considering the LDP of the time-inhomogeneous Langevin equations with strong damping in general random environment. They showed that the solution of the second order Langevin dynamics and that of a first order equation possess the same LDP assuming the first order equation satisfies the local LDP. In [38], we studied the moderate deviation principle for a class of Langevin dynamic systems with strong damping and Markovian switching. Specifically, we consider

$$\begin{cases} \varepsilon^2\ddot{q}_\varepsilon(t) = b(q_\varepsilon(t), r_\varepsilon(t)) - \alpha_\varepsilon(q_\varepsilon(t))\dot{q}_\varepsilon(t) + \sqrt{\varepsilon}\sigma(q_\varepsilon(t), r_\varepsilon(t))\dot{w}(t) \\ q_\varepsilon(0) = q \in \mathbb{R}^d, \quad \dot{q}_\varepsilon(0) = \frac{p}{\varepsilon} \in \mathbb{R}^d. \end{cases}$$

where $r_\varepsilon(t)$ is a fast-varying continuous-time Markov chain with a finite state space $\mathcal{M}$.The main contribution is that we demonstrated not only do the solution of the second order Langevin dynamics and that of the corresponding first order equation verify the same LDP, but also they satisfy the same moderate deviation principle.

2. Future Research Plans

In the future, I am interested in pursuing such areas related to stochastic (partial) differential equations, stochastic control, stochastic approximation, large and moderate deviations. I am eager to learn new materials to expand my research horizon with a wide range of applications. In what follows, I outline some potential research projects.

2.1. **Large and moderate deviations.** Large deviations (LDs) theory is very important in many fields, including statistics, finance, and statistical mechanics. It is concerned with the asymptotic behavior, the exponential decay in specific, of probabilities of rare events. Take the i.i.d. real-valued random variables $X_1, \ldots, X_n$ for instance, let $S_n = X_1 + \cdots + X_n$ be the partial sums, the Central Limit Theorem (CLT) quantifies the probability that S_n differs from μn by amount of order $\sqrt{n}$

where $\mu = \mathbb{E}X_1$. Deviations of this $\sqrt{n}$ size are called "normal". However, large deviations deal with events where S_n differs from μn by an amount of order n which goes beyond the scope of CLT. A typical example is the event $\{S_n \geq (\mu + a)n, a > 0\}$, and it shows that under certain conditions, the decay is exponential in n:

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq (\mu + a)n) = -I(a) < 0, \quad a > 0.$$

where I is called the rate function satisfying certain properties. To fill the gap between the normal scale and the large deviation scale, moderate deviation takes the stage. The mathematical theory of large deviations was initiated by Cramer [7] in 1930s and further developed by Donsker and Varadhan [10, 11, 12, 13]. Freidlin and Wentzell [17] established the LDs theory for diffusion processes in 1970s. For more information about large deviations theory, we refer the work [14, 19].

Recently, in [44, 45], Veretennikov obtained sufficient conditions under which the rate function for the Euler approximation scheme for a solution of an SDE is close to that for an exact solution. For a solution of an SDE X_t, the log-asymptotics of probabilities

$$\mathbb{P}\Big(\frac{1}{t}\int_0^t f(X_s)ds \in A\Big), \quad A \subset \mathbb{R}$$

is described by the rate function L that is a Frendchel-Legendre transformation of the function H where

$$L(\alpha) = \sup_\beta\{\alpha\beta - H(\beta)\}, \quad H(\beta) := \lim_{t\to\infty} \frac{1}{t} \log \mathbb{E} \exp\left(\beta \int_0^t f(X_s)ds\right).$$

Denote by X_t^h the Euler approximation of the solution of such SDE with step size h, the corresponding function H^h, if well-defined, approximates to H in the following way,

$$|H^h(\beta) - H(\beta)| \leq Ch^{1/2}\Delta_0^{-1}$$

Thus we conclude that the rate function $L^h(\alpha)$ will be close to $L(\alpha)$.

The natural question is whether we can have similar results when the diffusion is modulated by an additional Markov switching process. That is, we are considering regime switching diffusion instead. One of the future work will devote to answer this question.

2.2. **Stochastic control.** Stochastic control or stochastic optimal control theory is aimed to find the best strategies or decisions to reach goals termed cost functional, such as minimizing risk or maximizing profit in financial markets, where the system is governed by the ordinary differential equations (ODEs) or stochastic dynamics. An extremely well-studied one is the linear system with quadratic cost functional known as LQ problem, which dates back to [2] for ODEs and the work of Kushner for diffusions; see also [46]. For the theory of stochastic control in finite dimensional space, we refer the work of Yong and Zhou [49]. In addition, there are a lot of work concerning the optimal control problems in infinite dimensional space. We refer the work of Fabbri et al. [15].

Stochastic reaction diffusion equations. Reaction diffusion systems are common to describe the change in space and time of the concentration of one or more chemical reactants. Stochastic perturbations are used to model the random influences of the environment. In [4], Cerrai investigated the optimal control problems for the stochastic reaction-diffusion equations with non-Lipschitz reaction coefficients in $\mathbb{R}^d, d \leq 3$. A verification theorem was proved saying the value function satisfies a Hamilton-Jacobi-Bellman (HJB) equation. However, the existence of such optimal control still remains open, except the case $d = 1$ with additional assumptions; see [4, pp. 1780]. Such existence was further studied by Brzezniak and Serrano [3] in 2013 using relaxed controls. They proved the existence of the optimal relaxed control for a class of semilinear stochastic PDEs on general Banach space. For more details about control for stochastic partial differential equations, we refer the recent book [28].

However, the driving noise of above stochastic reaction-diffusion systems or SPDEs are normally modeled by cylindrical Wiener process in a Hilbert space, i.e., $W(t) = \sum_{i=1}^{\infty} e_i w_i(t)$, where $\{e_i\}$ is a complete orthonormal basis of a Hilbert space, and $\{w_i(t)\}$ is a family of independent, real-valued, standard Brownian motions. Nevertheless, empirical data implies Brownian motion is inappropriate to explain many physical phenomenons, like turbulence and flicker noise. More widely used in physics is the fractional Brownian motion (fBM), which is first defined by Kolmogorov in 1940s. Motivated by this, it is interesting to study the optimal control (or relaxed control) problems for stochastic equations driven by cylindrical fractional Brownian motions in infinite dimensional space. For the LQ problem in a Hilbert space, Duncan et al. [9] proved the existence of the optimal control in a feedback form for the stochastic evolution equations driven by cylindrical fBM. However, when the nonlinear reaction term is considered in the drift, the existence of the optimal control (or relaxed control) for such stochastic equations driven by cylindrical fractional Brownian motions still remains unknown. Thus one of future work will aim to answer this question.

Time inconsistent control. Time consistency in classical optimal control theory means that if an optimal control is found for a given initial time and initial state, the optimal control remains its optimality on any later time along the optimal trajectory. To keep this nice property needs some restrictions on the cost functional, discounting fact, for example. It then leads to the study of optimal control problems with time inconsistent cost functional. In order to study that, one resort to the game theoretic approach and aim to find the locally optimal, time-consistent equilibrium strategies. There are many results concerning time inconsistent controls for SDEs in finite dimensional spaces. For the infinite dimensional setting, we mention that recently Dou and Lü [8] studied the linear-quadratic optimal control with time-inconsistent cost functional for stochastic evolution equations in a Hilbert space. They established the existence of the equilibrium strategy via multi-person differential games. We note that their model is still driven by cylindrical Wiener processes. Thus another future work will devote to the study of time inconsistent control problems for the stochastic evolution equation driven by cylindrical fBM.

References

[1] S. Andradóttir. A stochastic approximation algorithm with varying bounds. *Operations Res.*, **43** (1995) 1037–1048.

[2] R. Bellman, I. Glicksberg and O. Gross. Some apsects of the Mathematical Theory of Control Processes, RAND Corporation, Santa Monica, CA, 1958.

[3] Z. Brzezzniak and R. Serrano. Optimal relaxed control of dissipative stochastic partial differential equations in Banach spaces. *SIAM J. Control Optim.*, **51** (2013): 2664-2703.

[4] S. Cerrai. Optimal control problems for stochastic reaction-diffusion systems with non-Lipschitz coefficients. *SIAM J. Control Optim.* **39** (2001): 1779-1816.

[5] S. Cerrai and M. Freidlin. Large deviations for the Langevin equation with strong damping. *J. Stat. Phys.* **161** (2015): 859-875.

[6] H.F. Chen and Y.M. Zhu. Stochastic approximation procedures with randomly varying truncations. *Science in China Ser. A* **29** (1986), 914–926.

[7] H. Cramer. Sur un nouveau theoreme-limite de la theorie des probabilites. *Actual. Sci. Ind.* **736** (1938): 5-23.

[8] F. Dou and Q. Lü. Time-inconsistent linear quadratic optimal control problems for stochastic evolution equations. *SIAM J. Control Optim.* **58**.1 (2020): 485-509.

[9] T.E. Duncan, B. Maslowski, and B. Pasik-Duncan. Linear stochastic differential equations driven by Gauss-Volterra processes and related linear-quadratic control problems. *Appl. Math. Optim.* **80** (2019): 369-389.

[10] M. D. Donsker and S. R. S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time. I. *Comm. Pure Appl. Math.* **28** (1975):1–47.

[11] M. D. Donsker and S. R. S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time. II. *Comm. Pure Appl. Math.* **28** (1975):279–301.

[12] M. D. Donsker and S. R. S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time. III. *Comm. Pure Appl. Math.* **29** (1976):389–461.

[13] M. D. Donsker and S. R. S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time. IV. *Comm. Pure Appl. Math.* **36** (1983):183–212.

[14] A. Dembo and O. Zeitouni. *Large deviations techniques and applications.* Springer Science & Business Media, 2009.

[15] G. Fabbri, F. Gozzi and A. Swiech. *Stochastic optimal control in infinite dimension*. Probability and Stochastic Modelling, Springer, 2017.

[16] R. Frey, T. Schmidt and L. Xu. On Galerkin approximations for the Zakai equation with diffusive and point process observations. https://arxiv.org/pdf/1303.0975.pdf (2018).

[17] M. Freidlin and A. D. Wentzell. *Random Perturbations of Dynamical Systems*. Springer, New York, 1998.

[18] J. Gao and H. Tembine. Distributed mean-field type filters for big data assimilation. *Proceedings of the 18th IEEE International Conference on High Performance Computing and Commmunications* (2016): 1446-1453.

[19] F. Hollander. *Large deviations*. American Mathematical Soc., 2000.

[20] M. Hutzenthaler and A. Jentzen. *Numerical approximations of stochastic differential equations with non-globally Lipschitz continuous coefficients*. American Mathematical Soc., 2015.

[21] M. Hutzenthaler, A. Jentzen and P.E. Kloeden. Strong convergence of an explicit numerical method for SDEs with nonglobally Lipschitz continuous coefficients. *Ann. Appl. Probab.* **22** (2012), 1611–1641.

[22] D. Higham, X. Mao and A. Stuart. Strong convergence of Euler-type methods for nonlinear stochastic differential equations. *SIAM J. Num. Analy.* **40** (2002), 1041-1063

[23] H.J. Kushner. On the differential equations satisfied by conditional probability densities of Markov processes, with application. *J. SIAM Control Ser. A*, **2** (1964), 106-119.

[24] H.J. Kushner and Paul Dupuis. *Numerical methods for stochastic control problems in continuous time*. Springer Science & Business Media, 2001.

[25] P.E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*, Springer, 2013.

[26] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning, *Nature*, Vol. 521, pp. 436-444, (2015).

[27] W. Liu and X. Mao. Strong convergence of the stopped Euler–Maruyama method for nonlinear stochastic differential equations. *Appl. Math. Comput.* **223** (2013): 389-400.

[28] Q. Lü and X. Zhang. *Mathematical control theory for stochastic partial differential equations*. Berlin: Springer, 2021.

[29] Xuerong Mao. The truncated Euler–Maruyama method for stochastic differential equations. *J. Comput. Appl. Math.* **290** (2015): 370-384.

[30] G.N. Milstein. Approximate integration of stochastic differential equations. *Theory Probab. Appl.* **19** (1975): 557-562.

[31] G.N. Milstein. A method of second order accuracy intergration of stochastic differential equations. *Theory Probab. Appl.*. **23** (1978) 396–401.

[32] Y. Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*. Kluwer Academic Publications, Norwell, MA, 2004.

[33] M. Nielsen. Neural networks and deep learning, (online).

[34] N.N. Nguyen and G. Yin. Large deviation principles for Langevin equations in random environment and applications. *J. Math. Phys.* **62** (2021) 088301.

[35] J. Park, D. Yi and S. Ji. A novel learning rate schedule in optimzation for neural networks and it's convergence. *Symmetry*, **660** (2020).

[36] H. Qian, F. Wu and G. Yin. Numerical algorithms for stochastic differential equations with randomly varying trunations. Preprint.

[37] H. Qian, Z. Wen and G. Yin. Numerical Solutions for Optimal Control of Stochastic Kolmogorov Systems with Regime-Switching and Random Jumps. *Stat. Inference Stoch. Process.* (2022): 1-21.

[38] H. Qian and G. Yin. Moderate Deviations for the Langevin Euqations: Strong Damping and Fast Markovian Switching, Preprint.

[39] H. Qian, G. Yin and Q. Zhang. Deep Filtering with Adaptive Learning Rates. *IEEE. Trans. Automat. Contr.* (2022).

[40] H. Qian, Q. Zhang and G. Yin. Filtering with degenerate observation noise: A stochastic approximation approach. *Automatica.* **142** (2022): 110376.

[41] S. Ruder. An overview of gradient descent optmization algorithms, arXiv, (2017).

[42] Sotirios Sabanis. A note on tamed Euler approximations. *Electron. Commun. Probab.* **18** (2013): 1-10.

[43] S.R.S. Varadhan. Asymptotic probability and differential equations. *Comm. Pure Appl. Math.* **19** (1966),261-286.

[44] A.Y. Veretennikov. On large deviations for approximations of SDEs. *Probab. Theory Relat. Fields* **125** (2003): 135-152.

[45] A.Y. Veretennikov. On approximate large deviations for 1D diffusion. (2003): 381-399.

[46] W.M. Wonham. On a matrix Riccatic equation of stochastic control. *SIAM J. Control Optim.* **6** (1968) 312-326.

[47] L.Y. Wang, G. Yin, and Q, Zhang. Deep filtering. *Communications in Information and Systems*, **21**: 651-667.

[48] G. Yin, Z. Wen, H. Qian and H. Nguyen. Numerical solutions for optimal control of stochastic Kolmgorov systems. *J. Syst. Sci. Complex.* **34** (2021): 1703-1722.

[49] Jiongmin Yong and Xun Yu Zhou. *Stochastic controls: Hamiltonian systems and HJB equations.* Springer Science & Business Media, 1999.

[50] M. Zakai. On the optimal filtering of diffusion processes. *Z. Wahrsch Verw. Gebiete.* **11** (1969), 230-243.