Insurance: Mathematics and Economics 96 (2021) 262–275

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

A hybrid deep learning method for optimal insurance strategies: Algorithms and convergence analysis

Zhuo Jin [a,*], Hailiang Yang [b], G. Yin [c]

[a] Centre for Actuarial Studies, Department of Economics, The University of Melbourne, VIC 3010, Australia
[b] Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong
[c] Department of Mathematics, University of Connecticut, Storrs, CT, 06269-1009, United States of America

ARTICLE INFO

Article history:
Received February 2020
Received in revised form November 2020
Accepted 30 November 2020
Available online xxxx

Keywords:
Neural network
Deep learning
Markov chain approximation
Stochastic approximation
Investment
Reinsurance
Dividend management
Convergence

ABSTRACT

This paper develops a hybrid deep learning approach to find optimal reinsurance, investment, and dividend strategies for an insurance company in a complex stochastic system. A jump–diffusion regime-switching model with infinite horizon subject to ruin is formulated for the surplus process. A Markov chain approximation and stochastic approximation-based iterative deep learning algorithm is developed to study this type of infinite-horizon optimal control problems. Approximations of the optimal controls are obtained by using deep neural networks. The framework of Markov chain approximation plays a key role in building iterative algorithms and finding initial values. Stochastic approximation is used to search for the optimal parameters of neural networks in a bounded region determined by the Markov chain approximation method. The convergence of the algorithm is proved and the rate of convergence is provided.

1. Introduction

For insurance companies, due to the nature of insurance products, insurers tend to accumulate relatively large amounts of cash or cash equivalents and invest the surplus in a financial market in order to pay future claims and avoid financial ruin. Meanwhile, redundant surplus will be paid out to policyholders before deficit occurs. Hence, to optimize the cash flow management, the decision makers of insurance companies will manage the risk sharing, investment performance and dividend payment schemes. Thus, how to build the strategies of reinsurance, investment, and dividend payout is crucial to insurance industry.

Reinsurance is a standard risk sharing tool to reduce and eliminate risks borne by primary insurance carriers. The primary insurance carrier pays to reinsurance company a certain part of premiums in return for protections against the adverse claim volatilities. Since the pioneering work of Borch (1960) and Arrow (1963), there has been extensive research on optimal reinsurance. The recent book on reinsurance, Albrecher et al. (2017), provides an impressive list of references on the subject.

The optimal portfolio selection problem is of practical importance. Earlier work in this area can be traced back to Markowitz's mean–variance model, see Markowitz (1952). The asset allocation problem for an insurance portfolio is different from that in finance, since an insurer needs to pay claims. Browne (1995) considers a model in which aggregate claims are modelled by a Brownian motion with drift, and the risky asset is modelled by a geometric Brownian motion. Hipp and Plum (2000) use the Cramér–Lundberg model to formulate the risk process of an insurance company and assume that the surplus of the insurance company can be invested in a risky asset (market index) that follows a geometric Brownian motion.

Dividend payment scheme represents an important signal about a company's financial status and future growth opportunities. Miller and Modigliani (1961) demonstrate the relationship between a company's dividend policy and the valuation of its shares. Instead of considering the safety aspect, optimal dividend strategies for insurance companies are first studied by De Finetti (1957), who proposed a random walk to model the surplus process and obtained that the optimal dividend payment strategy was of barrier type. This research focuses on the economic performance instead of the safety aspect to maximize the discounted total dividend payment until ruin. Gerber (1972) provides solutions for optimal dividend problem under both discrete and continuous models. Højgaard and Taksar (1999) study the reinsurance and dividend strategies in a diffusion model and provide closed-form solutions for optimal strategies.

* Corresponding author.
E-mail addresses: zjin@unimelb.edu.au (Z. Jin), hlyang@hku.hk (H. Yang), gyin@uconn.edu (G. Yin).

https://doi.org/10.1016/j.insmatheco.2020.11.012
0167-6687/

In past decades, extensive research has been devoted to finding optimal insurance strategies using analytic techniques under various discrete-time and continuous-time models. Types of controls such as regular, singular, or impulse controls are investigated under various models such as random walk, compound Poisson process, jump diffusion model, regime-switching model, etc. Due to increasing complexity of stochastic systems such as considering multiple types of controls simultaneously, adopting nonlinear insurance/reinsurance premium principles, and multiple decision makers in a game-theoretical framework etc., closed-form solutions are not available in many cases. Recently, there is emerging research on numerically solving insurance problems using finite difference or similar type of methods; see Jin et al. (2012, 2013a,b), and Van Staden et al. (2018).

On the other hand, the fast developments of machine learning, big data analytics, and artificial intelligence are changing our community and insurance market in almost all aspects. There are emerging efforts to figure out the impacts of data science on insurance industry, and to see how we can apply the novel data science approach to insurance industry such as reducing losses, claim reserve estimation, policy design, and key parameter estimation; see Wüthrich (2018a,b), Hainaut (2018), and Aleandri (2018). A comprehensive summary of machine learning techniques in non-life insurance pricing and data science such as regression trees, neural networks, and unsupervised learning is presented in Wüthrich and Buser (2017).

When managing a portfolio with multiple insurance products, the decision maker generally faces a stochastic control problem. Depending on the structures of insurance products, the control problem is categorized into two types: finite-time horizon and infinite-time horizon. There exists some literature on applying deep learning methods to solve finite-time horizon problems. Han and E (2016) and E et al. (2017) utilize neural networks to approximate the controls. The expectation of the objective function at terminal time is approximated by the average value of Monte-Carlo paths. Hence, finding optimal controls becomes searching the optimal parameters of approximating neural networks under a certain criteria guided by the rewarding function. Bachouch et al. (2018) and Huré et al. (2018) integrate deep learning methods into Monte Carlo backward optimization algorithms. Parametric neural networks are adopted and the optimization is executed backwards at discrete times. The approximating error analysis is provided. In summary, determining optimal controls in such finite-time horizon problems can be viewed as Monte Carlo projections starting from an initial value.

For infinite-time horizon problems, since there is no fixed terminal time, we can hardly use the maximization of a simple expectation of projections to design the reward function. There exists very few literature on applying deep learning methods to find stochastic optimal controls in infinite-time horizon. Cheng et al. (2020) develop a Markov chain approximation-based deep learning algorithm to approximate the optimal insurance strategies using neural networks. The idea of using Markov chain approximation method to find initial guesses is proposed. The reinsurance strategy and dividend strategy, considered as regular and singular controls respectively, are approximated by two neural networks separately. The classical gradient descent algorithm is adopted to find the weights of the two neural networks. A couple of numerical examples are presented to show that the neural-network approximating strategies converge to the analytical solutions obtained in Højgaard and Taksar (1999). In this paper, we further modify the algorithm and replace gradient descent method by stochastic approximation to calibrate the parameters of neural networks. The stochastic approximation theory provides a well-established framework to guarantee the convergence of the iterations in the weak sense. A rigorous convergence proof of the algorithm is provided in this work, while Cheng et al. (2020) are the first work to develop a hybrid Markov chain approximation-based deep learning algorithm and presents several numerical examples.

The hybrid feature of the proposed algorithm lies in an integration of neural network, Markov chain approximation, and stochastic approximation to solve a stochastic optimization problem. Markov chain approximation method (MCAM) and stochastic approximation (SA) are the main building blocks in the approximation procedures. A comprehensive introduction of the development of Markov chain approximation methods and stochastic approximation methods, together with the literature can be found in Kushner and Dupuis (2001) and Kushner and Yin (2003), respectively.

In this work, we apply our method to a complex jump–diffusion system with regime-switching. The controls are approximated by neural networks. To obtain the optimal parameters of the neural networks, we have developed two major steps. (1) Applying the Markov chain approximation method with coarse scale to estimate the initial guess of the neural network; (2) Applying stochastic approximation with fine scale to estimate the accurate parameters in a bounded region. The convergence of the numerical scheme is proved.

Comparing with the existing numerical methods on stochastic control problems, our proposed deep-learning algorithm has two main advantages. First, the introduction of machine learning framework enables us to improve the computation efficiency by using the two-scale numerical method. As it is well known, it is inevitable that one faces the problem of "curse of dimensionality" that the number of computation nodes grows exponentially when dealing with optimization problems with multiple control variables and states. We replace the optimization over the piecewise control grid for every state value by finding optimal parameters of neural networks for all state values. Now the computational complexity mainly comes from the evaluation of gradients for every state value. By using the stochastic approximation to calculate the optimum, the number of computation nodes increases linearly with respect to the number of points in the state lattice. In addition, the coarse-scale Markov chain approximation provides an initial value with small neighbourhood to conduct the stochastic approximation with fine-scale computation. Hence the computation efficiency for optimal controls can be largely improved. Second, the accuracy of numerical results can be improved by the developed algorithm. Traditional approximation methods generally use piecewise constant controls to approximate the optimal control. Then the accuracy of control strategy is subject to the denseness of the grid. The denseness of grids depends on the types and ranges of controls and states. When the ranges of controls and states are not comparable, the computation efficiency and accuracy are largely affected since it is difficult to find suitable stepsize for the lattice. On the contrary, neural networks allow the control strategy to take values in a continuous range and easily conquer the difficulty of effectively choosing a precision in control spaces with significant different scales.

The rest of the paper is organized as follows. A general formulation of surplus, dividend, investment, reinsurance strategies, and related assumptions are presented in Section 2 together with a complex regime-switching jump diffusion model. Section 3 shows the construction of an approximating Markov chain. In Section 4, the main steps of deep learning algorithms are established. The neural networks are constructed accordingly. Convergence of the algorithm is provided in Section 5. Some concluding remarks are provided in Section 6.

2. Formulation

Let us work with a complete filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$, where $\{\mathcal{F}_t\}$ (or simply $\mathcal{F}_t$) is a filtration satisfying the usual condition. That is, $\mathcal{F}_t$ is a family of σ-algebras such that $\mathcal{F}_s \subset \mathcal{F}_t$ for $s \le t$ and that $\mathcal{F}_0$ contains all null sets.

An insurance company adopts reinsurance, investment and dividend strategies to manage the insurance portfolios. The surplus process depends on regimes of the market, which is modelled by a continuous-time finite-state Markov chain. The Markov chain, $\alpha(t)$ takes values in a finite space $\mathcal{M} = \{1, \dots, m\}$. The states of economy are represented by the Markov chain $\alpha(t)$. Let the continuous-time Markov chain $\alpha(t)$ be generated by $Q = (q_{ij}) \in \mathbb{R}^{m\times m}$. That is,

$$\mathbb{P}\{\alpha(t+\delta) = j | \alpha(t) = i, \alpha(s), s \le t\} = \begin{cases} q_{ij}\delta + o(\delta), & \text{if } j \ne i, \\ 1 + q_{ii}\delta + o(\delta), & \text{if } j = i, \end{cases} \tag{2.1}$$

where $q_{ij} \ge 0$ for $i, j = 1, 2, \dots, m$ with $j \ne i$ and $q_{ii} = -\sum_{j\ne i} q_{ij} < 0$ for each $i = 1, 2, \dots, m$.

We consider a Poisson measure in lieu of the widely used Poisson process. Suppose $\mathcal{R} \subset \mathbb{R}_+$ is a compact set.

$$N(t, H) := \text{ number of claims on } [0, t] \text{ with claim size taking values in } H \in \mathcal{R} \tag{2.2}$$

counts the number of claims up to time t, which is a Poisson counting process. The claim size A has a distribution $\Pi(\cdot)$. Due to the regime switching, the Poisson measure in each regime is represented as N_i for all $i \in \mathcal{M}$. Then the Poisson measure $N_i((\cdot), (\cdot))$ has intensity $\lambda(i)dt \times \Pi(d\rho)$, where $\Pi(d\rho) = f(\rho)d\rho$, and $f(\rho)$ is the density function. Let $Y(t)$ be the aggregate claims process

$$Y(t) = \int_0^t \int_{\mathbb{R}} \rho N_{\alpha(t)}(dt, d\rho).$$

$Y(t)$ is a jump process representing claims with arrival rate $\lambda(i)$, for $i \in \mathcal{M}$. Note that claim frequencies depend on the states of economy and financial market.

Furthermore, for each $i \in \mathcal{M}$, we assume that the premium rate $c(i)$ collected by the primary insurance company follows the expectation premium principle:

$$c(i) = (1 + \varphi)\lambda(i)\mathbb{E}[A],$$

where φ is the safety loading for the primary insurer.

Let κ be the fraction of each claim paid by the primary insurance company. Then the aggregate claims amount paid by the primary insurance company is denoted as $Y^\kappa(t)$. The reinsurance premium rate is denoted as $g(\kappa)$. We consider proportional reinsurance strategy in this work. Hence $\kappa \in [0, 1]$. By using the variance premium principle, the reinsurance premium rate at time t is

$$g(\kappa) = (1 - \kappa)\mathbb{E}[A] + \beta(1 - \kappa)^2 \mathrm{Var}[A], \tag{2.3}$$

where $\beta > 0$ is the safety loading for the reinsurer. Note that different premium principles are adopted for insurers and reinsurance companies to make the formulation more general.

Following the work in Yang and Zhang (2005), we assume the surplus process is invested in a financial market with a risk free asset whose price follows

$$dS_0(t) = S_0(t)r_0(\alpha(t))dt$$

and $\mathcal{N}$ risky assets whose prices are governed by

$$dS_\iota(t) = S_\iota(t)(\mu_\iota(\alpha(t))dt + \sum_{j=1}^{\mathcal{N}} \sigma_{\iota j}(\alpha(t))dW_j(t)),$$

where $W(t) = (W_1(t), \dots, W_{\mathcal{N}}(t))'$ is an $\mathcal{N}$-dimensional standard Brownian motion. In the above and thereafter, B' denotes the transpose of B with B being either a vector or a matrix with appropriate dimension, and $|B|$ denotes the Euclidean norm of B. Set

$$B(\alpha(t)) = (\mu_1(\alpha(t)) - r_0(\alpha(t)), \dots, \mu_{\mathcal{N}}(\alpha(t)) - r_0(\alpha(t)))', \quad \text{and} \quad \sigma(\alpha(t)) = (\sigma_{\iota j}(\alpha(t)))_{\mathcal{N}\times\mathcal{N}}. \tag{2.4}$$

We use the proportional portfolio $\phi(t) = (\phi_1(t), \dots, \phi_{\mathcal{N}}(t))'$ to represent an investment strategy, where $\phi_\iota(t)$ is the percentage of the total capital invested in asset ι. To better reflect the reality in certain markets where short selling is not allowed, we further set a borrowing constraint on the investment strategy, which means $\sum_{\iota=1}^{\mathcal{N}} \phi_\iota(t) \le 1$ at any time. Denote

$$[0, 1]^{\mathcal{N}} = [0, 1] \times [0, 1] \times \cdots \times [0, 1],$$

and denote the constraint set of the controls as

$$\Gamma := \left\{ \phi \in [0, 1]^{\mathcal{N}} : \sum_{\iota=1}^{\mathcal{N}} \phi_\iota \le 1 \right\}. \tag{2.5}$$

A dividend strategy $D(\cdot)$ is an $\mathcal{F}_t$-adapted process $\{D(t) : t \ge 0\}$ corresponding to the accumulated amount of dividends paid up to time t such that $D(t)$ is a nonnegative and nondecreasing stochastic process that is right continuous and have left limits with $D(0^-) = 0$.

Combining the proportional reinsurance, investment and dividend strategies, the surplus process of the insurance company, denoted by $X(t)$, follows

$$\begin{cases} dX(t) = \big\{X(t)[r_0(\alpha(t)) + \phi'(t)B(\alpha(t))] + c(\alpha(t)) \\ \qquad\qquad - \lambda(\alpha(t))g(\kappa)\big\}\, dt \\ \qquad\qquad + X(t)\phi'(t)\sigma(\alpha(t))dW(t) - dY^\kappa(t) - dD(t), \\ X(0) = x, \end{cases} \tag{2.6}$$

where

$$Y^\kappa(t) = \kappa Y(t) = \kappa \int_0^t \int_{\mathbb{R}} \rho N_{\alpha(t)}(dt, d\rho).$$

In this paper, $\mathcal{F}_t$ is the σ-algebra generated by $\{\alpha(s), W(s), N_{\alpha(s)}(\cdot) : 0 \le s \le t,\ \alpha \in \mathcal{M}\}$.

By choosing the optimal reinsurance, investment and dividend payment strategies, we aim to maximize the present value of cumulative discounted dividend payments until financial ruin. Let γ be the discount factor. A strategy $\pi(\cdot) = \{\pi(t) := (\kappa(t), \phi(t), D(t)), t \ge 0\}$ being progressively measurable with respect to $\mathcal{F}_t$ is called an admissible strategy. For an arbitrary triplet of controls $\pi(\cdot) = (\kappa(\cdot), \phi(\cdot), D(\cdot))$, the objective function is defined as

$$J(x, i, \pi) = \mathbb{E}_{x,i}\left(\int_0^\tau e^{-\gamma t} dD(t)\right), \tag{2.7}$$

where $\tau = \inf\{t \ge 0 : X(t) < 0\}$ represents the time of ruin, and $\mathbb{E}_{x,i}$ denotes the expectation conditioned on $X(0) = x$ and $\alpha(0) = i$. Hence, the value function

$$V(x, i) = \sup_\pi J(x, i, \pi). \tag{2.8}$$

In this paper we consider absolutely continuous dividend strategies, and assume there is an upper bound $\widetilde{M}$ on the dividend rate. We write $D(t)$ as

$$dD(t) = u(t)dt, \quad 0 \le u(t) \le \widetilde{M}, \tag{2.9}$$

where $u(t)$ is an $\mathcal{F}_t$-adapted process and $0 < \widetilde{M} < \infty$.

Remark 2.1. We focus on developing the algorithm and providing convergence analysis in this paper. The case of restricted

dividend payment rate is presented to illustrate the idea and methodology. The case of unrestricted dividend payment rate does not add much difficulty to the algorithm design. A numerical example of the unrestricted dividend payment rate is presented in Cheng et al. (2020).

Then (2.6) can be rewritten as

$$\begin{cases} dX(t) = \big\{X(t)[r_0(\alpha(t)) + \phi'(t)B(\alpha(t))] + c(\alpha(t)) \\ \qquad - \lambda(\alpha(t))g(\kappa) - u(t)\big\}\,dt \\ \qquad +X(t)\phi'(t)\sigma(\alpha(t))dW(t) - dY^{\kappa}(t), \\ X(0) = x. \end{cases} \tag{2.10}$$

In this case, denote the control by $\pi := (\kappa, u, \phi) \in [0,1] \times [0, \widetilde{M}] \times [0,1]^{\mathcal{N}}$. Then the expected discounted dividend until ruin is given by

$$J(x, i, \pi(\cdot)) = \mathbb{E}_{x,i}\left[\int_0^{\tau} e^{-\gamma t} u(t)dt\right]. \tag{2.11}$$

The value function of maximizing expected dividend payoff is defined by the following optimization problem:

$$V(x, i) = \sup_{\pi \in [0,1]\times[0,\widetilde{M}]\times[0,1]^{\mathcal{N}}} J(x, i, \pi(\cdot)). \tag{2.12}$$

For $i \in \mathcal{M}$, and $V(\cdot, i) \in C^2(\mathbb{R})$, define an operator $\mathcal{L}$ by

$$\begin{aligned} \mathcal{L}V(x, i) \ &= V_x(x, i)[x(r_0(i) + \phi' B(i)) + c(i) - \lambda_i g(\kappa) - u] \\ &\quad + \frac{1}{2}x^2 V_{xx}|\phi'\sigma(i)|^2 \\ &\quad + \lambda(i)\int_0^x [V(x - \kappa z, i) - V(x, i)]f(z)dz + QV(x, \cdot)(i), \end{aligned} \tag{2.13}$$

where V_x and V_{xx} denote the first and second derivatives with respect to x, and

$$QV(x, \cdot)(i) = \sum_{i \neq j} q_{ij}(V(x, j) - V(x, i)).$$

The operator $\mathcal{L}$ will be used to design the approximating Markov chain in the following section.

3. Approximating Markov chain

We will construct an approximating Markov chain for the regime-switching jump diffusion model. The discrete-time controlled Markov chain is so defined that it is locally consistent with (2.10). First, we will approximate the terms of discrete claims.

There is an equivalent way to define the process (2.10) by working with the claim times and values. To do this, set $\nu_0 = 0$, and let ν_n, $n \geq 1$, denote the time of the nth claim, and ρ_n be the corresponding claim severity. Let $\{\nu_{n+1} - \nu_n, \rho_n, n < \infty\}$ be mutually independent random variables with $\nu_{n+1} - \nu_n$ being exponentially distributed, and let ρ_n have a distribution $\Pi(\cdot)$. Furthermore, let $\{\nu_{k+1} - \nu_k, \rho_k, k \geq n\}$ be independent of $\{X(s), \alpha(s), s < \nu_n, \nu_{k+1} - \nu_k, \rho_k, k < n\}$, then the nth claim term is ρ_n.

Because $\nu_{n+1} - \nu_n$ is exponentially distributed, we can write

$$\begin{aligned} &\mathbb{P}\{\text{claim occurs on } [t, t+\delta) | X(s), \alpha(s), W(s), N(s, \cdot), s \leq t\} \\ &\quad = \lambda(\alpha(t))\delta + o(\delta). \end{aligned} \tag{3.1}$$

It is implied by the above discussion that $X(\cdot)$ satisfying (2.10) can be viewed as a process that involves regime-switching diffusion with claims according to the claim rate defined by (3.1). To begin, we construct a discrete-time, finite-state, controlled Markov chain to approximate the controlled diffusion process with regime-switching, and the dynamic system is given by

$$\begin{cases} dX(t) = \big\{X(t)[r_0(\alpha(t)) + \phi'(t)B(\alpha(t))] + c(\alpha(t)) \\ \qquad - \lambda(\alpha(t))g(\kappa) - u\big\}\,dt \\ \qquad +X(t)\phi'(t)\sigma(\alpha(t))dW(t) \\ X(0) = x. \end{cases} \tag{3.2}$$

Note that the state of the process has two components x and α. Hence in order to use the methodology in Kushner and Dupuis (2001), our approximating Markov chain must have two components: one component delineates the diffusive behaviour whereas the other keeps track of the regimes. Let $h > 0$ be a discretization parameter representing the stepsize. Define $\widetilde{S}_h = \{x : x = kh, k = 0, \pm 1, \pm 2, \dots\}$ and $S_h = \widetilde{S}_h \cap \widetilde{G}_h$, where $\widetilde{G}_h = (0, \mathcal{B} + h)$ and $\mathcal{B}$ is an upper bound introduced for numerical computation purpose. Moreover, assume without loss of generality that the boundary point $\mathcal{B}$ is an integer multiple of h. Let $\{(\xi_n^h, \alpha_n^h), n < \infty\}$ be a controlled discrete-time Markov chain on $S_h \times \mathcal{M}$ and denote by $p_D^h((x, i), (y, j)|\pi)$ the transition probability from a state (x, i) to another state (y, j) under the control π. We need to define p_D^h so that the chain's evolution well approximates the local behaviour of the controlled regime-switching diffusion (3.2).

π is a control parameter and takes values in the compact set $\mathcal{U}$. We use π_n^h to denote the random variable that is the actual control action for the chain at discrete time n. To approximate the continuous-time Markov chain, we need another approximation sequence. Suppose that there is an $\Delta t^h(x, \alpha, \pi) > 0$ and define the "interpolation interval" as $\Delta t_n^h = \Delta t^h(\xi_n^h, \alpha_n^h, \pi_n^h)$ on $S_h \times \mathcal{M} \times \mathcal{U}$. Define the interpolation time $t_n^h = \sum_{k=0}^{n-1} \Delta t_k^h(\xi_k^h, \alpha_k^h, \pi_k^h)$. The piecewise constant interpolations $(\xi^h(\cdot), \alpha^h(\cdot))$, $\pi^h(\cdot)$ and $\beta^h(t)$ are defined as

$$\xi^h(t) = \xi_n^h,\ \alpha^h(t) = \alpha_n^h,\ \pi^h(t) = \pi_n^h,\ \beta^h(t) = n \ \text{ for } \ t \in [t_n^h, t_{n+1}^h). \tag{3.3}$$

Let $\{p_D^h((x, i), (y, j)|\pi)\}$ for $(x, i), (y, j) \in S^h \times \mathcal{M}$, and $\pi \in \mathcal{U}$ be a collection of well defined transition probabilities for the Markov chain (ξ_n^h, α_n^h), an approximation to $(X(\cdot), \alpha(\cdot))$. Define the difference $\Delta\xi_n^h = \xi_{n+1}^h - \xi_n^h$. Assume $\inf_{x,i,\pi} \Delta t^h(x, i, \pi) > 0$ for each $h > 0$ and $\lim_{h\to\infty} \Delta t^h(x, i, \pi) \to 0$. Let $\mathbb{E}_{x,i,n}^{\pi,h}$, $\mathrm{Var}_{x,i,n}^{\pi,h}$, and $p_{x,i,n}^{\pi,h}$ denote the conditional expectation, variance, and marginal probability given $\{\xi_k^h, \alpha_k^h, u_k^h, k \leq n, \xi_n^h = x, \alpha_n^h = i, \pi_n^h = \pi\}$, respectively. The sequence $\{(\xi_n^h, \alpha_n^h)\}$ is said to be locally consistent with the diffusion and regime switching, if

$$\begin{aligned} &\mathbb{E}_{x,i,n}^{\pi,h}\Delta\xi_n^h = (x[r_0(i) + \phi'(t)B(i)] + c(i) - \lambda(i)g(\kappa) - u) \\ &\Delta t^h(x, i, \pi) + o(\Delta t^h(x, i, \pi)), \\ &\mathrm{Var}_{x,i,n}^{\pi,h}\Delta\xi_n^h = x^2|\phi'\sigma(i)|^2\Delta t^h(x, i, \pi) + o(\Delta t^h(x, i, \pi)), \\ &p_{x,i,n}^{\pi,h}\{\alpha_{n+1}^h = j\} = \Delta t^h(x, i, \pi)q_{ij} + o(\Delta t^h(x, i, \pi)), \text{ for } j \neq i, \\ &p_{x,i,n}^{\pi,h}\{\alpha_{n+1}^h = i\} = \Delta t^h(x, i, \pi)(1 + q_{ii}) + o(\Delta t^h(x, i, \pi)), \\ &\sup_{n,\omega} |\Delta\xi_n^h| \to 0 \text{ as } h \to 0. \end{aligned} \tag{3.4}$$

Once we have a locally consistent approximating Markov chain, we can approximate the value function. Let $\mathcal{U}^h$ denote the collection of controls, which are determined by a sequence of measurable functions $F_n^h(\cdot)$ such that

$$\pi_n^h = F_n^h(\xi_k^h, \alpha_k^h, k \leq n; \pi_k^h, k \leq n). \tag{3.5}$$

Note that $S_h \times \mathcal{M}$ is a finite state space. Let N_h denote the first time that $\{\xi_n^h\}$ leaves S_h. Then the first exit time of $\xi^h(\cdot)$ from S_h is $\tau_h = t_{N_h}^h$. Natural reward functions for the chain is

$$J^h(x, i, \pi^h) = \mathbb{E}_{x,i}\sum_{n=0}^{N_h - 1} e^{-\gamma\,\Delta t_n^h} u_n^h \Delta t_n^h. \tag{3.6}$$

Denote

$$V^h(x, i) = \sup_{\pi^h \in \mathcal{U}^h} J^h(x, i, \pi^h). \tag{3.7}$$

In view of (2.6), the transition probabilities can be constructed as follows

$$
\begin{aligned}
&p_D^h((x,i),(x+h,i)|\pi)\\
&\quad=\frac{(x^2|\phi'\sigma(i)|^2/2)+h(x[r_0(i)+\phi'(t)B(i)]+c(i)-\lambda(i)g(\kappa)-u)^+}{\widetilde{D}-\gamma h^2},\\
&p_D^h((x,i),(x-h,i)|\pi)\\
&\quad=\frac{(x^2|\phi'\sigma(i)|^2/2)+h(x[r_0(i)+\phi'(t)B(i)]+c(i)-\lambda(i)g(\kappa)-u)^-}{\widetilde{D}-\gamma h^2},\\
&p_D^h((x,i),(x,j)|\pi)=\frac{h^2}{\widetilde{D}-\gamma h^2}q_{ij},\quad \text{for } j\neq i,\\
&p_D^h(\cdot)=0,\quad \text{otherwise},\\
&\Delta t^h(x,i,\pi)=\frac{h^2}{\widetilde{D}},
\end{aligned}
\tag{3.8}
$$

with

$$
\begin{aligned}
\widetilde{D}=&\,x^2|\phi'\sigma(i)|^2+h|x[r_0(i)+\phi'(t)B(i)]+c(i)-\lambda(i)g(\kappa)\\
&-u|+h^2(\gamma-q_{ii})
\end{aligned}
$$

being well defined.

Suppose that the current state is $\xi_n^h=x$, $\alpha_n^h=i$, and the control is $\pi_n^h=\pi$. To present the claim terms, we determine the next state $(\xi_{n+1}^h,\alpha_{n+1}^h)$ by noting:

1. No claims occur in $[t_n^h,t_{n+1}^h)$ with probability $(1-\lambda(i)\Delta t^h(x,i,\pi)+o(\Delta t^h(x,i,\pi)))$; we determine $(\xi_{n+1}^h,\alpha_{n+1}^h)$ by transition probability $p_D^h(\cdot)$ as in (3.8).
2. There is claim loss q in $[t_n^h,t_{n+1}^h)$ with probability $\lambda(i)\Delta t^h(x,i,\pi)+o(\Delta t^h(x,i,\pi)))$, we determine $(\xi_{n+1}^h,\alpha_{n+1}^h)$ by

   $$\xi_{n+1}^h=\xi_n^h-q^h,\ \alpha_{n+1}^h=\alpha_n^h,$$

   where $q^h\in S_h\subseteq\mathbb{R}_+$ such that q^h is the nearest value of q so that $\xi_{n+1}^h\in S_h$. Then $|q_h-q|\to 0$ as $h\to 0$, uniformly in x.

Let H_n^h denote the event that $(\xi_{n+1}^h,\alpha_{n+1}^h)$ is determined by the first alternative above and use T_n^h to denote the event of the second case. Let $I_{H_n^h}$ and $I_{T_n^h}$ be corresponding indicator functions, respectively. Then $I_{H_n^h}+I_{T_n^h}=1$. Then we need a new definition of the local consistency for Markov chain approximation of jump diffusion process with regime-switching.

Definition 3.1. A controlled Markov chain $\{(\xi_n^h,\alpha_n^h),n<\infty\}$ is said to be locally consistent with (2.6), if there is an interpolation interval $\Delta t^h(x,i,\pi)\to 0$ as $h\to 0$ uniformly in x, i, and π such that

1. there is a transition probability $p_D^h(\cdot)$ that is locally consistent with (3.2) in the sense that (3.4) holds.
2. there is a $\delta^h(x,i,\pi)=o(\Delta t^h(x,i,\pi))$ such that the one-step transition probability $\{p^h((x,i),(y,j))|\pi\}$ is given by

$$
\begin{aligned}
p^h(((x,i),(y,j))|\pi)=&(1-\lambda(i)\Delta t^h(x,i,\pi)+\delta^h(x,i,\pi))\\
&p_D^h((x,i),(y,j))\\
&+(\lambda(i)\Delta t^h(x,i,\pi)+\delta^h(x,i,\pi))\\
&\Pi\{\rho^h=x-y\}.
\end{aligned}
\tag{3.9}
$$

Furthermore, the system of dynamic programming equations in the kth iteration follows

$$
V_{k+1}^h(x,i)=\begin{cases}S(x,i,V_k^h,\pi),&\text{for } x\in S_h,\\ 0,&\text{for } x=0.\end{cases}
\tag{3.10}
$$

where

$$
\begin{aligned}
S(x,i,V_k^h,\pi)=&\max_{\pi\in\mathcal{U}}\Big[(1-\lambda(i)\Delta t^h(x,i,\pi)+\delta^h(x,i,\pi))\\
&\times e^{-\gamma\Delta t^h(x,i,\pi)}\sum_{y,j}(p_D^h((x,i),(y,j))|\pi)V_k^h(y,j)\\
&+(\lambda(i)\Delta t^h(x,i,\pi)+\delta^h(x,i,\pi))e^{-\gamma\Delta t^h(x,i,\pi)}\\
&\times\int_0^x V_k^h(x-\kappa\rho^h,i)\Pi(d\rho)+u\Delta t^h(x,i,\pi)\Big].
\end{aligned}
$$

4. Numerical algorithm

In this section, we give details of the numerical algorithm. In Section 4.1, we present the idea of approximating controls with neural networks and introduce the Markov chain approximation method to find the initial values with coarse scale. In Section 4.2, details of stochastic approximation method are provided to find accurate approximations with fine scale. A comprehensive description of the method is shown in Section 4.3.

4.1. MCAM

According to our approach, the control variables are approximated by neural networks and computed in a lattice. Without loss of generality, given different admissible ranges of different strategies, independent neural networks are adopted for different controls. The computational structure of computation nodes of each neural network follows the pattern in Fig. 4.1.

Remark 4.1. Fig. 4.1 provides a generic computational structure of one neural node with inputs and outputs. Every computation neural node follows the same pattern as in Fig. 4.1. When choosing neural networks practically, the architecture of neural networks depends on the complexity of the problem. Generally neural networks with more layers are equipped with stronger ability to learn more complicated control strategies. But over-complicated neural networks and excess parameters may lead to issues of gradient vanishing.

Remark 4.2. When multiple controls exist, separate and independent neural networks will be designed. Comparable controls can adopt similar architecture of neural networks to improve the computation efficiency. More explanation and figures about designation of neural networks can be found in Cheng et al. (2020). For example, Figure 1 in Cheng et al. (2020) presents an example showing that two controls are approximated by two independent neural networks with similar architecture.

Define θ as the collection of all weights and bias terms in the neural networks, then denote the neural network control strategy by $\mathbb{N}(x,i,\theta)$. Given the policy and value space as designed in Section 3, the stepsize is h while the surplus is approximated as $\{x_\iota\}_{\iota=1}^n$. That is, the range of the surplus is approximated by n spots.

Define an approximation of (3.6) and (3.7) as

$$
\begin{aligned}
&\widetilde{J}^h(x_\iota,i,\mathbb{N}(x_\iota,i,\theta^h))=\mathbb{E}_{x_\iota,i}\sum_{s=0}^{N_h-1}e^{-\gamma\Delta t_s^h}\mathbb{N}(x_\iota,i,\theta^h)\Delta t_s^h,\\
&\widetilde{V}^h(x_\iota,i)=\sup_{\theta^h}\widetilde{J}^h(x_\iota,i,\mathbb{N}(x_\iota,i,\theta^h)),
\end{aligned}
\tag{4.1}
$$

for $\iota=1,\dots,n$, $i=1,\dots,m$.

We are using Markov chain approximation method to find the initial values of the controls. The initial values of the controls are obtained by the value iteration. Assuming we are currently in the kth iteration with the iterative value function $\widetilde{V}_{k-1}^h$ obtained from the previous iteration, the optimal parameters in the current

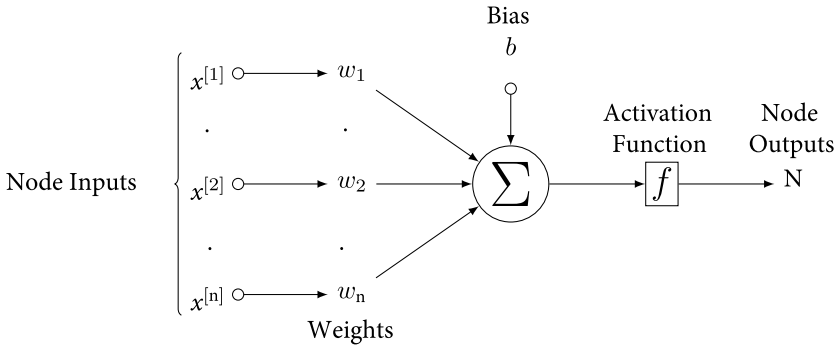

Fig. 4.1. Neural node computation.

iteration is denoted as θ_k^h. We show how to search for θ_k^h by stochastic approximation in Section 4.2. Then the kth iterative control strategy is expressed as $\mathbb{N}(x, i, \theta_k^h)$.

The kth iterative value function follows the dynamic programming equation

$$\widetilde{V}_k^h(x_\iota, i) = S(x_\iota, i, \widetilde{V}_{k-1}^h, \mathbb{N}(x, i, \theta_k^h)),\ \ 1 \le \iota \le n,\ \ 1 \le i \le m.$$

Our objective is

$$\lim_{k\to\infty, h\to 0} \theta_k^h = \theta^*. \tag{4.2}$$

Repeat the above iteration until the termination condition

$$\sum_{\iota=1}^{n}\sum_{i=1}^{m} \left(\widetilde{V}_k^h(x_\iota, i) - \widetilde{V}_{k-1}^h(x_\iota, i)\right)^2 < \epsilon_I$$

is met, where ϵ_I is a predefined small positive number.

4.2. SA algorithm

In the stochastic approximation, the fundamental goal is to use $\mathbb{N}(x_\iota, i, \theta^h)$ and choose θ^h to maximize the global improvement function $G^h(\theta^h)$ as the following

$$G^h(\theta^h) = G^h(\widetilde{J}^h(x_\iota, i, \mathbb{N}(x_\iota, i, \theta^h)) : \iota = 1, \ldots, n, i = 1, \ldots, m). \tag{4.3}$$

The global improvement function G^h reflects how much the approximating cost function will improve globally. The choice of G^h should serve the goal that the value function will be improved on most states rather than on every state of the state lattice. The global improvement is achieved by iteratively adopting the θ_k^h which is optimized in every iteration k. That is, $\theta_k^h = \text{argmax}_{\theta^h} G^h$. Practically, we can choose general global improvement functions such as the weighted average of the value function depending on the problem formulation and performance of the algorithm.

To proceed, we provide a general setting of the stochastic approximation algorithm. The general setting will provide an effective framework for the proof of convergence in the next section. Without loss of generality, we assume that the parameters of neural network θ is an r-dimension vector. Let e_j denote the standard unit vector with the jth component being 1 and all other components being 0, for $j = 1, \ldots, r$. Let θ_l denote the lth estimate of the optimum. Let $\delta_l > 0$ be the stepsize of finite difference intervals and ε_l be the stepsize of iterations, respectively.

The stochastic approximation algorithm proposed above can be described by the following steps in each iteration k for the lattice with stepsize h. Then the lth estimate of the optimum of θ in the kth iteration is denoted as $\theta_{k,l}^h$. Given an initial value $\theta_{k,0}^h$ in each iteration k, we aim to verify

$$\lim_{l\to\infty} \theta_{k,l}^h = \theta_k^h. \tag{4.4}$$

To simplify the notation in the algorithm description, we omit the subscript k and the superscript h in each term, and we write the global improvement function $G^h(\theta^h)$ as $G(\theta)$.

(1) Initialization: Take an initial guess θ_0.
(2) Estimate θ_1:
   - Take noisy observations of $G(\theta)$ at $\theta_0 \pm \delta_0 e_j$ and denote the observations by $\widehat{G}(\theta_0 \pm \delta_0 e_j, \eta_{0,j}^{\pm})$. Here and henceforth, $\eta_{0,j}^{\pm}$ denotes the observation noise associated with $\theta_0 \pm \delta_0 e_j$.
   - Define the gradient estimate $K_{0,j} = \mathbb{D}\widehat{G}(\theta_0, \eta_{0,j}^{\pm}) = \frac{\widehat{G}(\theta_0+\delta_0 e_j, \eta_{0,j}^{+}) - \widehat{G}(\theta_0-\delta_0 e_j, \eta_{0,j}^{-})}{2\delta_0}$, for $j = 1, \ldots, r$.
   - Construct $\theta_1 = \theta_0 + \varepsilon_0 K_0$, where $\varepsilon_0 > 0$ is a step size and $K_0 = (K_{0,1}, K_{0,2}, \ldots, K_{0,r})$.
(3) Iteration step: Repeat Step 2 with $\theta_{l+1} = \theta_l + \varepsilon_l K_l$ in which $K_l = (K_{l,1}, K_{l,2}, \ldots, K_{l,r})$ and $K_{l,i} = \frac{\widehat{G}(\theta_l+\delta_l e_j, \eta_{l,j}^{+}) - \widehat{G}(\theta_l-\delta_l e_j, \eta_{l,j}^{-})}{2\delta_l}$, $\eta_{l,j}^{\pm}$ are the random noises for $l \ge 1$. We further assume that the sequences $\eta_{l,j}^{\pm}$ are stationary processes with $\mathbb{E}\widehat{G}(\theta, \eta_l^{\pm}) = G(\theta)$ for each θ.
(4) A termination criterion: A tolerance level is reached.

For the algorithm to converge, we need to choose the stepsize so that the following conditions satisfy

$$\varepsilon_l \to 0, \quad \varepsilon_l/\delta_l \to 0, \quad \textstyle\sum_l \varepsilon_l = \infty, \quad \sum_l \varepsilon_l^2/\delta_l^2 < \infty. \tag{4.5}$$

By using the initialization of the MCAM, the search region of the optimal control is confined to a bounded neighbourhood centred at the MCAM's piecewise optimal control. In addition, to ensure that the iterations remain in a bounded region, we consider the case that θ is bounded. Therefore, the iterations of the neural network's parameters should be confined to a bounded region. For simplicity, take the projection region to be

$$M = \{\theta : \mathbb{N}(x, i, \theta) \in [\mathbb{N}(x, i, \theta_0) - \delta_l, \mathbb{N}(x, i, \theta_0) + \delta_l], |\theta_{l,j}| \le \widetilde{B}, \widetilde{B} \in \mathbb{R}, i = 1, \ldots, m, j = 1, \ldots, r\}, \tag{4.6}$$

where $\widetilde{B}$ is an arbitrarily large positive number, $\theta_{l,j}$ is the jth element of vector θ_l. Now we propose a projection procedure

$$\theta_{l+1} = \mathcal{P}_M[\theta_l + \varepsilon_l K_l], \tag{4.7}$$

where $\mathcal{P}_M$ denotes the projection operator onto the constraint set M and $\mathcal{P}_M(\theta)$ is the closest point in M to θ. Thus if the iteration is within the region we keep their values. If they ever exit from this

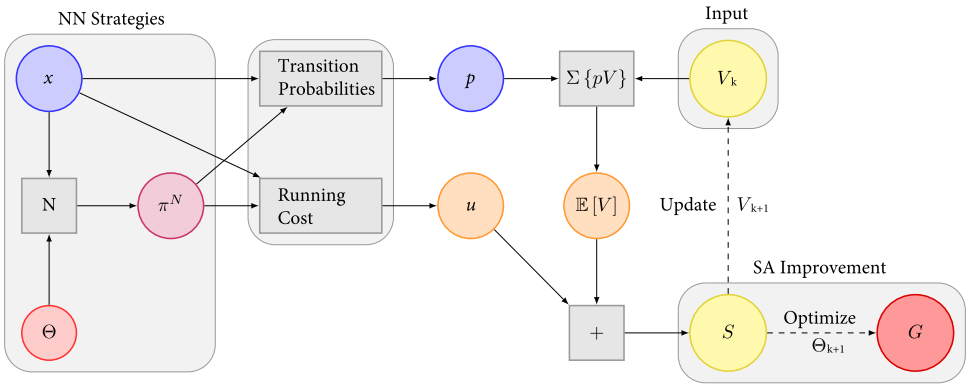

Fig. 4.2. Iterative learning cycle.

interval, we push them back to the boundaries. For more details of the projection procedure, we refer readers to Yin et al. (2002).

4.3. Algorithm summary

To summarize all above constructions, a complete algorithm will be given in the following. The algorithm starts from the following initialization steps. To simplify the notations, we omit the index h for each term.

Initialization 1: Construct the state lattice for deep learning algorithm denoted as $\{x_\iota\}_{\iota=1}^{n}$, and the state lattice for obtaining initial value of θ denoted as $\{y_\ell\}_{\ell=1}^{\tilde{n}}$. These two state lattices satisfy following conditions:

$$x_0 = y_0,\ x_n = y_{\tilde{n}},\ \tilde{n} \le n.$$

Initialization 2: Choose the sets of computation precision ϵ and maximal number of learning times. They are used to obtain initial value of θ_0, to determine iterative control strategy $\mathbb{N}(x, i, \theta_k)$, and to stop the MCAM iteration respectively.

Initialization 3: Pick up an appropriate function $f(\cdot)$ to compute initial value for iteration. The choice is subject to properties of the problem. Compute U_0 as:

$$U_0(y_\ell, i) = f(y_\ell, i),\ \ell = 1, \dots, \tilde{n}, i = 1, \dots, m.$$

Initialization 4: Use the same function $f(\cdot)$ as in *Initialization 3* to compute V_0 as:

$$\widetilde{V}_0(x_\iota, i) = f(x_\iota, i),\ \iota = 1, \dots, n, i = 1, \dots, m.$$

After initialization, the algorithm will repeat below iterative steps. The repetition will stop until the algorithm achieves the desired precision, which is set up in *Initialization 2*.

Step 1: For $i = 1, \dots, m$, denote by $\tilde{\pi}_k(y_\ell, i)$ the optimal control obtained from standard MCAM. The input values U_{k-1} are from *Initialization 3* or *Step 4* in the last round.

Step 2: Fit against $\tilde{\pi}_k(y_\ell, i)$ to obtain parameter starting values $\theta_{k,0}$:

$$\theta_{k,0} = \text{argmin}_\theta \sum_{\ell=1}^{\tilde{n}} \sum_{i=1}^{m} (\tilde{\pi}_k(y_\ell, i) - \mathbb{N}(y_\ell, i, \theta))^2.$$

The fitting process will stop if the desired precision is achieved or if the maximal number of fitting iteration is reached, whichever comes first.

Step 3: Maximize $G(\theta_k)$ by the stochastic approximation algorithm to obtain the iterative control strategy. The learning process will stop if the desired precision is achieved or if the maximal number of learning iteration is reached, whichever comes first. Now, we have θ_k, which yields $\mathbb{N}(x, i, \theta_k)$.

Step 4: For $\ell = 1, \dots, \tilde{n}$, iterate to $U_k(y_\ell, i))$ in the following way:

$$U_k(y_\ell, i) = S(y_\ell, i, U_{k-1}, \mathbb{N}(y_\ell, i, \theta_k)).$$

Step 5: For $\iota = 1, \dots, n$, iterate to $V_k(x_\iota, i)$ in the following way:

$$\widetilde{V}_k(x_\iota, i) = S(x_\iota, i, \widetilde{V}_{k-1}, \mathbb{N}(x_\iota, i, \theta_k)).$$

Step 6: Compute $\sum_{\iota=1}^{n} \sum_{i=1}^{m} \left(\widetilde{V}_k(x_\iota, i) - \widetilde{V}_{k-1}(x_\iota, i)\right)^2$, and then check the termination condition:

- If $\sum_{\iota=1}^{n} \sum_{i=1}^{m} \left(\widetilde{V}_k(x_\iota, i) - \widetilde{V}_{k-1}(x_\iota, i)\right)^2 < \epsilon$, stop.
- If $\sum_{\iota=1}^{n} \sum_{i=1}^{m} \left(\widetilde{V}_k(x_\iota, i) - \widetilde{V}_{k-1}(x_\iota, i)\right)^2 > \epsilon$,
  - if the maximal number of iterations is reached, stop;
  - otherwise, go to *Step 1*.

One should bear in mind that we are using a general deep learning algorithm to solve for the proposed optimization problems. In specific cases, the structures of neural networks can be different. For example, if the ranges of controls are comparable, we use one neural network to output different controls. If the ranges of controls are not comparable, we use independent neural networks to output different controls. Then more parameters and more precise grid required. To guarantee the feasibility and efficiency of the algorithm, we need build neural networks case by case. The neural networks will be calibrated and trained by stochastic approximation method. A brief computation graph is provided in Fig. 4.2 to illustrate the algorithm.

In the following section, we focus on the convergence proof of the algorithm. Several numerical examples are presented in our previous work (Cheng et al., 2020). The algorithms are coded by Python with TensorFlow package and run on x64 platform of Intel Xeon E-2186 2.90 GHz CPU with 64 GB RAM and NVIDIA Quadro P5200 GPU with 16 GB RAM. Detailed settings of neural networks and performance of the algorithm can be found in Section 6 of Cheng et al. (2020).

5. Convergence

In this section, we prove the convergence of the algorithm. That is, by starting with an initial guess $\theta_{k,0}^h$, the iteration will lead to the optimal set of parameters θ^*. Particularly, we will prove that (4.2) and (4.4) hold.

5.1. Convergence of Markov chain approximation

This section deals with the convergence proof of (4.2). Note that θ_k^h is the parameters of neural networks that optimally fits

the piecewise constant control obtained by MCAM in the kth iteration with stepsize h. The convergence of (4.2) can be guaranteed by the convergence of piecewise constant control in MCAM. Hence, we need only prove the convergence of π^h.

5.1.1. Local consistency

To proceed, we first present the local consistency for our approximating Markov chain. Basically, it says that the approximation we constructed is consistent with the given dynamic system.

Lemma 5.1. *The Markov chain $\{\xi_n^h, \alpha_n^h\}$ with transition probabilities $(p_D^h(\cdot))$ defined in (3.8) is locally consistent with the stochastic differential equation in (3.2).*

Proof. Using (3.8), it is readily seen that

$$\begin{aligned}\mathbb{E}^{\pi,h}_{x,i,n}\Delta\xi_n^h &= hp_D^h((x,i),(x+h,i)|\pi) - hp_D^h((x,i),(x-h,i)|\pi)\\ &= (x[r_0(i)+\phi'(t)B(i)] + c(i) - \lambda(i)g(\kappa) - u)\Delta t^h(x,i,\pi)\\ &\quad + o(\Delta t^h(x,i,\pi)),\end{aligned}$$

Likewise, we obtain

$$\begin{aligned}\mathbb{E}^{\pi,h}_{x,i,n}(\Delta\xi_n^h)^2 &= h^2p_D^h((x,i),(x+h,i)|\pi) - h^2p_D^h((x,i),(x-h,i)|\pi)\\ &= x^2|\phi'\sigma(i)|^2\Delta t^h(x,i,\pi) + \Delta t^h(x,i,\pi)O(h).\end{aligned}$$

As a result,

$$\begin{aligned}\mathrm{Var}^{\pi,h}_{x,i,n}\Delta\xi_n^h &= x^2|\phi'\sigma(i)|^2\Delta t^h(x,i,\pi) + \Delta t^h(x,i,\pi)O(h)\\ &\quad - (x[r_0(i)+\phi'(t)B(i)] + c(i) - \lambda(i)g(\kappa) - u)\\ &\quad \times \Delta t^h(x,i,\pi) + o(\Delta t^h(x,i,\pi))^2\\ &= x^2|\phi'\sigma(i)|^2\Delta t^h(x,i,\pi) + o(\Delta t^h(x,i,\pi)).\end{aligned}$$

Thus both equations in (3.4) are verified. The desired local consistency follows with the use of local properties of claims specified. □

5.1.2. Interpolations of approximation sequences

Based on the Markov chain approximation constructed in the last section, piecewise constant interpolation is obtained here with appropriately chosen interpolation intervals. Using (ξ_n^h, α_n^h) to approximate the continuous-time process $(X(\cdot), \alpha(\cdot))$, we defined the continuous-time interpolation $(\xi^h(\cdot), \alpha^h(\cdot))$, $\pi^h(\cdot)$ and $\beta^h(t)$ as in (3.3). Recall N_h is defined in the paragraph above (3.6), we define the first exit time of $\xi^h(\cdot)$ from S_h by

$$\tau_h = t^h_{N_h}. \tag{5.1}$$

Let the discrete times at which claims occur be denoted by $\nu_j^h, j = 1, 2, \ldots$ Define $\mathcal{D}_n^h$ as the smallest σ-algebra of $\{\xi_k^h, \alpha_k^h, \pi_k^h, H_k^h, k \le n; \nu_k^h, \rho_k^h : \nu_k^h \le t_n\}$. Then τ_h is a $\mathcal{D}_n^h$-stopping time. Using the interpolation process, we can rewrite (3.6) as

$$J^h(x,i,\pi^h) = \mathbb{E}_{x,i}\int_0^{\tau_h} e^{-\gamma s}u^h(s)ds. \tag{5.2}$$

Let $\xi_0^h = x$, $\alpha_0^h = \alpha$, $\mathbb{E}_n^h$ denote the expectation conditioned on the information up to time n, that is, conditioned on $\mathcal{D}_n^h$. In addition, $\mathcal{U}^h$ defined by (3.5) is equivalent to the collection of all piecewise constant admissible controls with respect to $\mathcal{D}_n^h$.

Then we can write

$$\begin{aligned}\xi_n &= x + \sum_{k=0}^{n-1}[\Delta\xi_k^h I_{H_k^h} + (\Delta\xi_k^h(1 - I_{H_k^h}))]\\ &= x + \sum_{k=0}^{n-1}\mathbb{E}_k^h\Delta\xi_k^h I_{H_k^h} + \sum_{k=0}^{n-1}(\Delta\xi_k^h - \mathbb{E}_k^h\Delta\xi_k^h)I_{H_k^h}\\ &\quad + \sum_{k=0}^{n-1}(\Delta\xi_k^h(1 - I_{H_k^h})).\end{aligned} \tag{5.3}$$

The local consistency leads to

$$\begin{aligned}&\sum_{k=0}^{n-1}\mathbb{E}_k^h\Delta\xi_k^h I_{H_k^h}\\ &= \sum_{k=0}^{n-1}((\xi_k^h[r_0(\alpha_k^h) + \phi'(t)B(\alpha_k^h)] + c(\alpha_k^h) - \lambda(\alpha_k^h)g(\kappa_k^h) - u_k^h)\Delta t_k^h\\ &\quad + o(\Delta t_k^h))I_{H_k^h}\\ &= \sum_{k=0}^{n-1}((\xi_k^h[r_0(\alpha_k^h) + \phi'(t)B(\alpha_k^h)] + c(\alpha_k^h) - \lambda(\alpha_k^h)g(\kappa_k^h) - u_k^h)\Delta t_k^h\\ &\quad + o(\Delta t_k^h)) - (\max_{k'\le n}\Delta t_{k'}^h)O(\sum_{k=0}^{n-1}I_{T_k^h})\end{aligned} \tag{5.4}$$

Denote

$$\begin{aligned}M_n^h &= \sum_{k=0}^{n-1}(\Delta\xi_k^h - \mathbb{E}_k^h\Delta\xi_k^h)I_{H_k^h},\\ R_n^h &= -\sum_{k=0}^{n-1}(\Delta\xi_k^h(1 - I_{H_k^h})) = \sum_{k:\nu_k<n}\rho_k^h,\end{aligned} \tag{5.5}$$

where M_n^h is a martingale with respect to $\mathcal{D}_n^h$. Note that

$$\mathbb{E}\sum_{k=0}^{n-1}I_{T_k^h} = \mathbb{E}[\text{number of } n : \nu_n^h \le t] \to \lambda t \quad \text{as} \quad h \to 0.$$

This implies

$$(\max_{k'\le n}\Delta t_{k'}^h)O(\sum_{k=0}^{n-1}I_{T_k^h}) \to 0 \text{ in probability as } h \to 0.$$

Hence we can drop the term involving $I_{H_k^h}$ without affecting the limit in (5.4). We attempt to represent $M^h(t)$ similar to the diffusion term in (3.2). Define $W^h(\cdot)$ as

$$\begin{aligned}W^h(t) &= \sum_{k=0}^{n-1}(\Delta\xi_k^h - \mathbb{E}_k^h\Delta\xi_k^h)/|\xi_k^h(\phi_k^h)'\sigma(\alpha_k^h)|,\\ &= \int_0^t \frac{1}{|x(s)(\phi^h)'\sigma(\alpha^h(s))|}dM^h(s).\end{aligned} \tag{5.6}$$

Combining (5.4)–(5.6), we rewrite (5.3) by

$$\begin{aligned}\xi^h(t) &= x + \int_0^t(\xi^h[r_0(\alpha^h(s)) + (\phi^h)'(s)B(\alpha^h(s))] + c(\alpha^h(s))\\ &\quad -\lambda(\alpha^h(s))g(\kappa^h(s)) - u^h(s))ds\\ &\quad + \int_0^t \xi^h|(\phi^h)'\sigma(\alpha^h(s))|dW^h(s) - R^h(t) + \varepsilon^h(t)\\ R^h(t) &= \sum_{\nu_n^h\le t}\rho_n^h\kappa_{j,n}^h(\nu_n),\end{aligned} \tag{5.7}$$

where $\varepsilon^h(t)$ is a negligible error satisfying

$$\lim_{h\to 0}\sup_{0\le t\le T}\mathbb{E}|\varepsilon^h(t)| \to 0 \text{ for any } 0 < T < \infty. \tag{5.8}$$

We can also rewrite (5.7) as

$$\begin{aligned}X(t) &= x + \int_0^t\{X(s)[r_0(\alpha(s)) + \phi'(s)B(\alpha(s))] + c(\alpha(s))\\ &\quad -\lambda(\alpha(s))g(\kappa) - u(s)\}ds\\ &\quad + \int_0^t X(s)\phi'(s)\sigma(\alpha(s))dW(s) - Y^\kappa(t),\end{aligned} \tag{5.9}$$

where

$$Y^{\kappa}(t)=\sum_{\nu_n\le t}\rho_n\kappa_n=\kappa(t)\int_0^t\int_{\mathbb{R}_+}\rho N(dsd\rho).$$

Now we give the definition of existence and uniqueness of weak solution.

Definition 5.2. By a weak solution of (5.9), we mean that there exist a probability space $(\Omega,\mathcal{F},\mathbb{F},P)$, a filtration $\mathcal{F}_t$, and process $(X(\cdot),\alpha(\cdot),\pi(\cdot),W(\cdot),N(\cdot))$ such that $W(\cdot)$ is a standard $\mathcal{F}_t$-Wiener process, $N(\cdot)$ is an $\mathcal{F}_t$-Poisson measure with claim rate λ and claim size distribution $\Pi(\cdot)$, $\alpha(\cdot)$ is a Markov chain with generator Q and state space $\mathcal{M}$, $\pi(\cdot)$ is admissible with respect to $(\alpha(\cdot),W(\cdot),N(\cdot))$, $X(\cdot)$ is $\mathcal{F}_t$-adapted, and (5.9) is satisfied. For an initial condition (x,i), by the weak sense uniqueness, we mean that the probability law of the admissible process $(\alpha(\cdot),\pi(\cdot),W(\cdot),N(\cdot))$ determines the probability law of solution $(X(\cdot),\alpha(\cdot),\pi(\cdot),W(\cdot),N(\cdot))$ to (5.9), irrespective of probability space.

We need one more assumption.

(A1) Let $\widehat{\tau}(\varphi)=\infty$ and $\widetilde{G}^o$ be an interior of an compact set, if $\varphi(t)\in\widetilde{G}^o$, for all $t<\infty$, otherwise, define $\widehat{\tau}(\varphi)=\inf\{t:\varphi\notin\widetilde{G}^o\}$. The function $\widehat{\tau}(\cdot)$ is continuous (as a map from $D[0,\infty)$, the space of functions that are right continuous and have left limits endowed with the Skorohod topology to the interval $[0,\infty]$ (the extended and compactified positive real numbers)) with probability one relative to the measure induced by any solution to (5.9) with initial condition (x,i).

5.1.3. Convergence of surplus processes

This section deals with convergence of surplus processes.

Lemma 5.3. *Using the transition probabilities $\{p^h(\cdot)\}$ defined in (3.4) and (3.9), the interpolated process of the constructed Markov chain $\{\alpha^h(\cdot)\}$ converges weakly to $\alpha(\cdot)$, the Markov chain with generator $Q=(q_{ij})$.*

Proof. The proof can be obtained similar to Theorem 3.1 in Yin et al. (2003). □

Theorem 5.4. *Let the approximating chain $\{\xi_n^h,\alpha_n^h,n<\infty\}$ constructed with transition probabilities defined in (3.8) be locally consistent with (2.10), $\{\pi_n^h,n<\infty\}$ be a sequence of admissible controls, and $(\xi^h(\cdot),\alpha^h(\cdot))$ be the continuous-time interpolation defined in (3.3). Let $\{\widetilde{\tau}_h\}$ be a sequence of $\mathcal{F}_t^h$-stopping times. Then $\{\xi^h(\cdot),\alpha^h(\cdot),\pi^h(\cdot),W^h(\cdot),N^h(\cdot),\widetilde{\tau}_h\}$ is tight.*

Proof. Using one point compactification, $\widetilde{\tau}\in[0,\infty]$. In view of Lemma 5.3, $\{\alpha^h(\cdot)\}$ is tight. The sequence $\{\pi^h(\cdot),\widetilde{\tau}_h\}$ is always tight since the corresponding range space is compact. Let $T<\infty$, and let ν_h be an $\mathcal{F}_t$-stopping time which is no bigger than T. Then for $\delta>0$,

$$\mathbb{E}^{u^h}_{\widetilde{\nu}_h}(W^h(\nu_h+\delta)-W^h(\nu_h))^2=\delta+\widetilde{\varepsilon}_h,\tag{5.10}$$

where $\widetilde{\varepsilon}_h\to 0$ uniformly in ν_h. Taking $\limsup_{h\to0}$ followed by $\lim_{\delta\to0}$ yield the tightness of $\{W^h(\cdot)\}$. In view of Theorem 9.2.1in Kushner and Dupuis (2001), the sequence $\{N^h(\cdot)\}$ is tight because the mean number of claims on any bounded interval $[t,t+s]$ is bounded by $\lambda(\alpha(t))s+\delta_1^h(s)$, where $\delta_1^h(s)$ goes to zero as $h\to0$, and

$$\liminf_{\delta\to0}\inf_{h,n}\mathbb{P}\{\nu_{n+1}^h-\nu_n^h>\delta|\text{data up to }\nu_n^h\}=1.$$

This also implies the tightness of $\{R^h(\cdot)\}$. These results and the boundedness of $c(\cdot)$, $g(\cdot)$ and $u(\cdot)$ implies the tightness of $\{\xi^h(\cdot)\}$. Thus, $\{\xi^h(\cdot),\alpha^h(\cdot),\pi^h(\cdot),W^h(\cdot),N^h(\cdot),\widetilde{\tau}_h\}$ is tight. □

Theorem 5.5. *Let $(\xi(\cdot),\alpha(\cdot),\pi(\cdot),W(\cdot),N(\cdot),\widetilde{\tau})$ be the limit of a weakly convergent subsequence and $\mathcal{F}_t$ the σ-algebra generated by $\{X(s),\alpha(s),\pi(s),W(s),N(s),s\le t,\widetilde{\tau}I_{\{\widetilde{\tau}<t\}}\}$. Then $W(\cdot)$ and $N(\cdot)$ are a standard $\mathcal{F}_t$-Wiener process and Poisson measure, respectively, and $\widetilde{\tau}$ is an $\mathcal{F}_t$-stopping time and $\pi(\cdot)$ is an admissible control. Let the claim times and claim sizes of $N(\cdot)$ be denoted by ν_n, ρ_n. Then, (5.9) is satisfied.*

Proof. Since $\{\xi^h(\cdot),\alpha^h(\cdot),\pi^h(\cdot),W^h(\cdot),N^h(\cdot),\widetilde{\tau}_h\}$ is tight, we can extract a weakly convergent subsequence by Prohorov's theorem. Denote the limit by $(\xi(\cdot),\alpha(\cdot),\pi(\cdot),W(\cdot),N(\cdot),\widetilde{\tau})$. To characterize $W(\cdot)$, let $t>0$, $\delta>0$, p, $\tilde{\kappa}$, $\{t_k:k\le p\}$ be given such that $t_k\le t\le t+\widetilde{t}$ for all $k\le p$, $\mathrm{P}(\widetilde{\tau}_h=t_k)$ is zero. Let $\{\Gamma_j^{\tilde{\kappa}},j\le\tilde{\kappa}\}$ be a sequence of nondecreasing partition of Γ such that $\Pi(\partial\Gamma_j^{\tilde{\kappa}})=0$ for all j and all $\tilde{\kappa}$, where $\partial\Gamma_j^{\tilde{\kappa}}$ is the boundary of the set $\Gamma_j^{\tilde{\kappa}}$. As $\tilde{\kappa}\to\infty$, let the diameter of the sets $\Gamma_j^{\tilde{\kappa}}$ go to zero. By (5.6), $W^h(\cdot)$ is an $\mathcal{F}_t$-martingale. Thus we have for any bounded and continuous function $\mathcal{H}(\cdot)$

$$\mathbb{E}\mathcal{H}(\xi^h(t_k),\alpha^h(t_k),W^h(t_k),\pi^h(t_k),N^h(t_k,\Gamma_j^{\tilde{\kappa}}),j\le\tilde{\kappa},\\ k\le p,\widetilde{\tau}_hI_{\{\widetilde{\tau}_h\le t\}})\times[W^h(t+\widetilde{t})-W^h(t)]=0.\tag{5.11}$$

By using the Skorohod representation and the dominant convergence theorem, letting $h\to0$, we obtain

$$\mathbb{E}\mathcal{H}(X(t_k),\alpha(t_k),W(t_k),\pi(t_k),N(t_k,\Gamma_j^{\tilde{\kappa}}),j\le\tilde{\kappa},\\ k\le p,\widetilde{\tau}I_{\{\widetilde{\tau}\le t\}})[W(t+\widetilde{t})-W(t)]=0.\tag{5.12}$$

Since $W(\cdot)$ has continuous sample paths, (5.12) implies that $W(\cdot)$ is a continuous $\mathcal{F}_t$-martingale. On the other hand, since $\mathbb{E}[(W^h(t+\delta))^2-(W^h(t))^2]=\mathbb{E}[(W^h(t+\delta)-W^h(t))^2]$, by using the Skorohod representation and the dominant convergence theorem together with (5.10), we have

$$\mathbb{E}\mathcal{H}(X(t_k),\alpha(t_k),W(t_k),\pi(t_k),N(t_k,\Gamma_j^{\tilde{\kappa}}),j\le\tilde{\kappa},\\ k\le p,\widetilde{\tau}I_{\{\widetilde{\tau}\le t\}})[W^2(t+\delta)-W^2(t)-\delta]=0.\tag{5.13}$$

The quadratic variation of the martingale $W(t)$ is t. Then $W(\cdot)$ is an $\mathcal{F}_t$-Wiener process.

Now we need to show that $N(\cdot)$ is an $\mathcal{F}_t$-Poisson measure. Let $\varphi(\cdot)$ be a continuous function on $\mathbb{R}_+$, and define the process

$$\varphi_N(t)=\int_0^t\int_{\mathbb{R}_+}\varphi(\rho)N(dsd\rho).$$

By an argument which is similar to the Wiener process above, if $f(\cdot)$ is a continuous function with compact support, then

$$\mathbb{E}\mathcal{H}(X(t_k),\alpha(t_k),W(t_k),\pi(t_k),N(t_k,\Gamma_j^{\tilde{\kappa}}),j\le\tilde{\kappa},k\le p,\widetilde{\tau}I_{\{\widetilde{\tau}\le t\}})\\ \times\Big[f(\varphi_N(t+\widetilde{t}))-f(\varphi_N(t))-\lambda\int_t^{t+\widetilde{t}}\int_{\mathbb{R}_+}[f(\varphi_N(s)+\varphi(\rho))\\ -f(\varphi_N(s))]\Pi(dsd\rho)\Big]=0.\tag{5.14}$$

Eq. (5.14) and the arbitrariness of $\mathcal{H}(\cdot)$, p, $\tilde{\kappa}$, t_k, $\Gamma_j^{\tilde{\kappa}}$, $f(\cdot)$ and $\varphi(\cdot)$ imply that $N(\cdot)$ is an $\mathcal{F}_t$-Poisson measure.

For $\delta>0$, define the process $\tilde{\phi}(\cdot)$ by $\tilde{\phi}^{h,\delta}(t)=\tilde{\phi}^h(n\delta),t\in[n\delta,(n+1)\delta)$. Then, by the tightness of $\{\xi^h(\cdot),\alpha^h(\cdot)\}$, (5.7) can be

rewritten as

$$\begin{aligned}\xi^h(t) = x &+ \int_0^t (\xi^h[r_0(\alpha^h(s)) + (\tilde{\phi}^h(s))'(t)B(\alpha^h(s))] + c(\alpha^h(s)) \\ &-\lambda(\alpha^h(s))g(\kappa^h(s)) - u^h(s))ds \\ &+ \int_0^t \xi^h|(\tilde{\phi}^h)'\sigma(\alpha^{h,\delta}(s))|dW^h(s) - R^h(t) + \varepsilon^{h,\delta}(t),\end{aligned} \tag{5.15}$$

where

$$\lim_{\delta\to 0}\limsup_{h\to 0}\mathbb{E}|\varepsilon^{h,\delta}(t)| = 0. \tag{5.16}$$

Letting $h \to 0$, by using the Skorohod representation, we obtain

$$\begin{aligned}\mathbb{E}|&\int_0^t (\xi^h[r_0(\alpha^h(s)) + (\tilde{\phi}^h(s))'(t)B(\alpha^h(s))] + c(\alpha^h(s)) \\ &-\lambda(\alpha^h(s))g(\kappa^h(s)) - u^h(s))ds \\ &- \int_0^t (\xi[r_0(\alpha(s)) + (\tilde{\phi}(s))'(t)B(\alpha(s))] + c(\alpha(s)) \\ &-\lambda(\alpha(s))g(\kappa(s)) - u(s))ds| = 0\end{aligned} \tag{5.17}$$

uniformly in t with probability one. Furthermore, the Skorohod representation implies that as $h \to 0$,

$$\begin{aligned}&\int_0^t (\xi^h[r_0(\alpha^h(s)) + (\tilde{\phi}^h(s))'(t)B(\alpha^h(s))] + c(\alpha^h(s)) \\ &\quad -\lambda(\alpha^h(s))g(\kappa^h(s)) - u^h(s))ds \\ &\to \int_0^t (\xi[r_0(\alpha(s)) + (\tilde{\phi}(s))'(t)B(\alpha(s))] + c(\alpha(s)) \\ &\quad -\lambda(\alpha(s))g(\kappa(s)) - u(s))ds\end{aligned} \tag{5.18}$$

uniformly in t with probability one on any bounded interval.

Since $\xi^{h,\delta}(\cdot)$ and $\alpha^{h,\delta}(\cdot)$ are piecewise constant functions, we obtain

$$\int_0^t X(s)\tilde{\phi}'(s)\sigma(\alpha^{h,\delta}(s))dW^h(s) \to \int_0^t X(s)\tilde{\phi}'(s)\sigma(\alpha^{\delta}(s))dW(s) \text{ as } h \to 0 \tag{5.19}$$

with probability one. Combining (5.11)–(5.19), we have

$$\begin{aligned}X(t) = x &+ \int_0^t (\xi[r_0(\alpha(s)) + (\tilde{\phi}(s))'(t)B(\alpha(s))] + c(\alpha(s)) \\ &-\lambda(\alpha(s))g(\kappa(s)) - u(s))ds \\ &+ \int_0^t X(s)\tilde{\phi}'(s)\sigma(\alpha^{\delta}(s))dW(s) - Y^{\kappa}(t) + \varepsilon^{\delta}(t),\end{aligned} \tag{5.20}$$

where $\lim_{\delta\to 0}\mathbb{E}|\varepsilon^{\delta}(t)| = 0$. Finally, taking limits in the above equation as $\delta \to 0$, (5.9) is obtained. □

5.1.4. Convergence of value functions

This section deals with the convergence of the reward and value functions. Note that the reward $J^h(x, i, \pi^h)$ is given by (5.2). By virtue of Theorem 5.4, with the use of τ_h in (5.1), each sequence $\{\xi^h(\cdot), \alpha^h(\cdot), \pi^h(\cdot), W^h(\cdot), N^h(\cdot), \tau_h\}$ has a weakly convergent subsequence with the limit satisfying (5.9). Slightly abusing the notation, still index the convergent subsequence by h with the limit denoted by $(X(\cdot), \alpha(\cdot), \pi(\cdot), W(\cdot), N(\cdot), \widetilde{\tau})$. By assumption (A1), $\{\tau_h\}$ is uniformly integrable. Using the Skorohod representation and the weak convergence, as $h \to 0$,

$$\mathbb{E}_{x,i}\int_0^{\tau_h} e^{-\gamma s}u^h(s)ds \to \mathbb{E}_{x,i}\int_0^{\widetilde{\tau}} e^{-\gamma s}u(s)ds. \tag{5.21}$$

Assumption (A1) guarantees that the exit time of $X(\cdot)$ from $\widetilde{G}^o$ is $\widetilde{\tau} = \tau$. This leads to

$$J^h(x, i, \pi^h) \to J(x, i, \pi) \text{ as } h \to 0. \tag{5.22}$$

Theorem 5.6. *Assume* (A1). *$V^h(x, i)$ and $V(x, i)$ are value functions defined in (3.7) and (2.8), respectively. Then $V^h(x, i) \to V(x, i)$ as $h \to 0$.*

Proof. Since $V(x, i)$ is the maximizing reward function, for any admissible control $\pi(\cdot)$,

$$J(x, i, m) \le V(x, i).$$

Let $\widetilde{\pi}^h(\cdot)$ be an optimal control for $\{\xi^h(\cdot)\}$. That is,

$$V^h(x, i) = J^h(x, i, \widetilde{\pi}^h) = \sup_{\pi^h} J^h(x, i, \pi^h).$$

Choose a subsequence $\{\widetilde{h}\}$ of $\{h\}$ such that

$$\limsup_{h\to 0} V^h(x, i) = \lim_{\widetilde{h}\to 0} V^{\widetilde{h}}(x, i) = \lim_{\widetilde{h}\to 0} J^{\widetilde{h}}(x, i, \widetilde{\pi}^{\widetilde{h}}).$$

Without loss of generality (passing to an additional subsequence if needed), we may assume that $(\xi^{\widetilde{h}}(\cdot), \alpha^{\widetilde{h}}(\cdot), \pi^{\widetilde{h}}(\cdot), W^{\widetilde{h}}(\cdot), N^{\widetilde{h}}(\cdot), \tau^{\widetilde{h}})$ converges weakly to $(X(\cdot), \alpha(\cdot), \pi(\cdot), W(\cdot), N(\cdot), \tau)$, where $\pi(\cdot)$ is an admissible related control. Then the weak convergence and the Skorohod representation yield that

$$\limsup_h V^h(x, i) = J(x, i, \pi) \le V(x, i). \tag{5.23}$$

We proceed to prove the reverse inequality.

We claim that

$$\liminf_h V^h(x, i) \ge V(x, i). \tag{5.24}$$

Suppose that $\overline{u}$ is an optimal control with respect to $(\alpha(\cdot), W(\cdot), N(\cdot))$ such that $\overline{x}(\cdot)$ and $\overline{\tau}$ are the associated trajectory and the stopping time, and $J(x, i, \overline{u}) = V(x, i)$. Given any $h > 0$, there are an $\varepsilon > 0$ and an ordinary control $\overline{\pi}^h(\cdot)$ that takes only finite many values, that $\overline{\pi}^h(\cdot)$ is a constant on $[k\varepsilon, k\varepsilon + \varepsilon)$, that $\overline{\pi}^h(\cdot)$ is its corresponding optimal control representation, and let $\overline{X}^h(\cdot)$ and $\overline{\tau}^h$ be the associated solution and stopping time. Then if $(\overline{\pi}^h(\cdot), \alpha(\cdot), W(\cdot), N(\cdot))$ converges weakly to $(\overline{\pi}(\cdot), \alpha(\cdot), W(\cdot), N(\cdot))$, we also have $(\overline{X}^h(\cdot), \overline{\pi}^h(\cdot), \alpha(\cdot), W(\cdot), N(\cdot), \overline{\tau}^h)$ converges weakly to $(X(\cdot), \overline{\pi}(\cdot), \alpha(\cdot), W(\cdot), N(\cdot), \overline{\tau})$, where (5.9) holds for the limit and $\overline{\tau}$ is the associate stopping time by Theorem 5.4. With assumption (A1), $J^h(x, i, \overline{\pi}^h) \to J(x, i, \overline{\pi})$, and that $J^h(x, i, \overline{\pi}^h) \ge V(x, i) - h$. Thus,

$$\liminf_h V^h(x, i) \ge J^h(x, i, \overline{\pi}^h) \ge V(x, i) - h.$$

The arbitrariness of h then implies that $\liminf_h V^h(x, i) \ge V(x, i)$.

Using (5.23) and (5.24) together with the weak convergence and the Skorohod representation, we obtain the desired result. The proof of the theorem is concluded. □

5.2. Convergence of stochastic approximation

In this section, we will work on the convergence of SA in each iteration k. That is, we will prove the convergence of (4.4). Similar to Section 4.2, since all terms are computed under the MCAM with stepsize h in all iterations, for notation simplicity, we omit the h in the superscript and k in the subscript for all terms in all iterations. We first rewrite (4.7) as

$$\theta_{l+1} = \theta_l + \varepsilon_l K_l + \varepsilon_l z_l, \tag{5.25}$$

where $\varepsilon_l z_l$ is the vector having the shortest Euclidean length necessary to bring $\theta_l + \varepsilon_l K_l$ back to M if it escapes from M. Then we have $\varepsilon_l z_{n,j} = \theta_{l+1,j} - \theta_{l,j} - \varepsilon_l \frac{\widehat{G}(\theta_l+\delta_l e_j, \eta^+_{n,j}) - \widehat{G}(\theta_l-\delta_l e_j, \eta^-_{n,j})}{2\delta_l}$. To establish convergence of the algorithm, we present some sufficient conditions first.

(A2) For each η, the observed or simulated solution $\widehat{G}(\cdot,\eta)$ is three-times continuously differentiable.

(A3) $\widehat{G}(\theta,\eta) = G_0(\theta,\widetilde{\eta}) + \widehat{\eta}$, such that $G_0(\cdot,\widetilde{\eta})$ is three times continuously differentiable for each $\widetilde{\eta}$, that $\{\widetilde{\eta}_l^{\pm}\}$ are sequences of bounded and stationary ϕ-mixing processes with mixing measure $\phi(k)$ satisfying $\sum_k \phi^{1/2}(k) < \infty$ and $\mathbb{E}_l G_0(\theta,\widetilde{\eta}_l^{\pm}) = \bar{G}(\theta)$ for each θ, where $\mathbb{E}_l$ is the conditional expectation with respect to the $\sigma-$field generated by $\{\widetilde{\eta}_l^{\pm}\}$; that $\{\widehat{\eta}_l^{\pm}\}$ are stationary martingale difference sequences satisfying $\mathbb{E}|\widehat{\eta}_l^{\pm}|^2 < \infty$, and that $\{\widetilde{\eta}_l^{\pm}\}$ and $\{\widehat{\eta}_l^{\pm}\}$ are independent.

Remark 5.7. Note that in (A2), we assumed that $\widehat{G}$ is a smooth function. The assumption is common in the treatment of stochastic optimization and is satisfied for the applications we are interested in. Non-smooth functions can be dealt with. A new development is in Nguyen and Yin (2020), which allows the nonsmoothness appear in the algorithms and the limit dynamics. The key depends on the use of differential inclusions and newly developed stochastic differential inclusions. However, we decide to not to get involved in the technical details.

Assumption (A3) covers a broad range of random noise processes. It includes additive noise such as the case $\widehat{G}(\theta,\xi) = \bar{G}(\theta)$ + noise as well as non-additive noise in a rather general form (a nonlinear function of θ and the noise). In (A3), $\widetilde{\eta}$ is the nonadditive noise and $\widehat{\eta}$ is the additive noise.

Define

$$\zeta_{l,j} = [\bar{G}(\theta_l+\delta_l e_j) - G_0(\theta_l+\delta_l e_j, \widetilde{\eta}_{l,j}^{+})] - [\bar{G}(\theta_l-\delta_l e_j) - G_0(\theta_l-\delta_l e_j, \widetilde{\eta}_{l,j}^{-})],$$

$$\psi_{l,j} = \widehat{\eta}_{l,j}^{+} - \widehat{\eta}_{l,j}^{-},$$

$$\varpi_{l,j} = \bar{G}_{l,j,\theta_{l,j}}(\theta_l) - \frac{\bar{G}(\theta_l+\delta_l e_j) - \bar{G}(\theta_l-\delta_l e_j)}{2\delta_l}.$$

Using the notation

$$\zeta_l = (\zeta_{l,1},\ldots,\zeta_{l,r})', \quad \psi_l = (\psi_{l,1},\psi_{l,2},\ldots,\psi_{l,r})',$$
$$\varpi_l = (\varpi_{l,1},\ldots,\varpi_{l,r})',$$

$$\bar{G}_l(\cdot) = \bar{G}_{l,\theta}(\theta_l) = (\bar{G}_{l,1,\theta_{l,1}}(\cdot), \bar{G}_{l,2,\theta_{l,2}}(\cdot),\ldots,\bar{G}_{l,r,\theta_{l,r}}(\cdot))',$$

the algorithm (5.25) can be written as

$$\theta_{l+1} = \theta_l + \varepsilon_l \bar{G}_{l,\theta}(\theta_l) + \varepsilon_l \frac{\zeta_l}{2\delta_l} + \varepsilon_l \frac{\psi_l}{2\delta_l} + \varepsilon_l \varpi_l + \varepsilon_l z_l. \tag{5.26}$$

In the above, $\{\varpi_l\}$ is known as the bias in the finite difference estimate of $\bar{J}_{l,\theta}(\theta_l)$. We separate the noise into two parts, uncorrelated noise $\{\zeta_l\}$ and correlated noise $\{\psi_l\}$. The algorithm is of the KW (Kiefer and Wolfowitz) type. Actually, in lieu of a sequence of decreasing $\{\delta_l\}$, we may use a fixed finite-difference stepsize $\delta > 0$. The main requirement is that the stepsize goes to 0 much faster than the finite difference intervals do. Further discussion on the choice of stepsizes can be found in Kushner and Yin (2003). For notational simplicity, $\widehat{G}_0(\theta,\widetilde{\eta}_l)$ and $\widehat{\eta}_l$ are often used to represent $\widehat{G}_0(\theta,\widetilde{\eta}_l^{\pm})$ and $\widehat{\eta}_l^{\pm}$ in what follows.

To proceed, let

$$t_0 = 0 \quad \text{and} \quad t_l = \sum_{i=0}^{n-1} \varepsilon_j,$$

and

$$m(t) = \begin{cases} n \text{ satisfying } t_l \le t < t_{l+1}, & \text{for } t \ge 0, \\ 0, & \text{for } t < 0. \end{cases}$$

Therefore, $m(t)$ is the unique value of l such that $t_l \le t < t_{l+1}$. We define the continuous-time interpolation $\theta_0(\cdot)$ on $(-\infty,\infty)$ by $\theta_0(t) = \theta_0$ for $t \le 0$, and for $t \ge 0$, $\theta_0(t) = \theta_l$, for $t_l \le t < t_{l+1}$. We then define the sequence of shifted process $\theta^l(\cdot)$ by

$$\theta_l(t) = \theta_0(t_l + t), \qquad -\infty < t < \infty.$$

As for the reflection term, we let $z_i = 0$ for $i < 0$. Define

$$Z_0(t) = 0, \text{ for } t \le 0,$$
$$Z_0(t) = \sum_{k=0}^{m(t)-1} \varepsilon_k z_k, \quad t \ge 0.$$

Define the shifted process as below:

$$Z_l(t) = Z_0(t_l+t) - Z_0(t_l) = \sum_{k=l}^{m(t_l+t)-1} \varepsilon_k z_k, \quad t \ge 0,$$

$$Z_l(t) = -\sum_{k=m(t_l+t)}^{l-1} \varepsilon_k z_k, \quad t < 0.$$

Then we can rewrite the interpolated process $\theta_l(\cdot)$ as follows

$$\begin{aligned}\theta_l(t) &= \theta_l + \sum_{k=l}^{m(t_l+t)-1} \varepsilon_k \bar{G}_{k,\theta}(\theta_k) + \sum_{k=l}^{m(t_l+t)-1} \varepsilon_k \frac{\zeta_k}{2\delta_k} + \sum_{k=n}^{m(t_l+t)-1} \varepsilon_k \frac{\psi_k}{2\delta_k} \\ &\quad + \sum_{k=l}^{m(t_l+t)-1} \varepsilon_k \varpi_k + \sum_{k=l}^{m(t_l+t)-1} \varepsilon_k z_k \\ &= \theta_l + g_l(t) + \zeta_l(t) + \psi_l(t) + \varpi_l(t) + Z_l(t),\end{aligned}$$

where for $t \ge 0$,

$$g_l(t) = \sum_{k=n}^{m(t_l+t)-1} \varepsilon_k \bar{G}_{k,\theta}(\theta_k), \qquad \zeta_l(t) = \sum_{k=n}^{m(t_l+t)-1} \varepsilon_k \frac{\zeta_k}{2\delta_k},$$
$$\psi_l(t) = \sum_{k=n}^{m(t_l+t)-1} \varepsilon_k \frac{\psi_k}{2\delta_k}, \qquad \varpi_l(t) = \sum_{k=n}^{m(t_l+t)-1} \varepsilon_k \varpi_k.$$

Recall the notion of the projected ODE (ordinary differential equation). According to the setup in Chapter 4 of Kushner and Yin (2003), for $\theta \in M$, define $C(\theta)$ as follows. For $\theta \in M^0$, the interior of M, $C(\theta)$ contains the zero element only; and for $\theta \in \partial M$, the boundary of M, let $C(\theta)$ be the infinite convex cone generated by the outer normals at θ of the faces on which θ lies. The projected ODE is defined by

$$\begin{aligned}\dot{\theta}(t) &= -\bar{G}_\theta(\theta(t)) + z(t), \qquad z \in C(\theta), \\ Z(t) &= \int_0^t z(u)du,\end{aligned} \tag{5.27}$$

where $z(\cdot)$ is the projection or the constraint term that is the minimum force needed to keep $\theta(\cdot)$ in M.

Theorem 5.8. *Assume that* (A2) *and* (A3) *are satisfied and that*

$$\sup_{l \le k \le m(t_l+T)} (\varepsilon_k/\delta_k^2)/(\varepsilon_l/\delta_l^2) \le c_1(T), \qquad \text{for some} \quad c_1(T) < \infty. \tag{5.28}$$

Then, there is a null set N such that for all $\omega \notin N$, $\{\theta_l(\cdot), Z_l(\cdot)\}$ is equicontinuous in the extended sense. Let $(\theta(\cdot), Z(\cdot))$ denote the limit of a convergent subsequence. Then, it satisfies the projected ordinary differential equation (5.27). If θ^* is an asymptotically stable point of (5.27) and θ_l is in some compact set that is a subset of the domain of attraction of θ^* w.p.1, then $\theta_l \to \theta^*$ w.p.1.

Proof. To prove this theorem, we apply Theorem 6.5.1 of Kushner and Yin (2003). We only need to verify the conditions in that theorem are fulfilled. To begin, for each θ in a bounded set, the

boundedness of $\{\eta_l^{\pm}\}$ implies that $\{\zeta_l\}$ is bounded; thus,

$$\sup_l |\zeta_l| < \infty.$$

The moment bounds on $\{\widehat{\eta}_l^{\pm}\}$ imply that $E|\psi_l|^2 < \infty$. The function $\bar{G}(\cdot)$, (A2), (A3) lead to the conclusion that $G(\theta, \eta)$ is continuous in θ for each η. For the additive noise, since $\{\psi_l\}$ is a martingale difference sequence, by virtue of (5.28), we have

$$\mathbb{E}\left|\sum_{k=m(jT)}^{m(jT+t)-1}(\varepsilon_k/\delta_k)\psi_k\right|^2 \to 0, \qquad \text{as } l \to \infty.$$

As a result, for any $\mu > 0$ and some $T > 0$,

$$\lim_l \mathbb{P}\left(\sup_{j\ge n}\max_{0\le t\le T}\left|\sum_{k=m(jT)}^{m(jT+t)-1}(\varepsilon_k/2\delta_k)\psi_k\right| \ge \mu\right) = 0. \tag{5.29}$$

Recall that the correlated noise is defined as

$$\begin{aligned}\zeta_{k,j} &= [\bar{G}(\theta_k + \delta_k e_j) - G_0(\theta_k + \delta_k e_j, \widetilde{\eta}_{k,j}^{+})]\\ &\quad -[\bar{G}(\theta_k - \delta_k e_j) - G_0(\theta_k - \delta_k e_j, \widetilde{\eta}_{k,j}^{-})]\end{aligned} \tag{5.30}$$

and

$$\zeta_k = (\psi_{k,1}, \psi_{k,2}, \ldots, \psi_{k,r}).$$

Since $\{\eta_l\}$ is stationary, by (A3),

$$\mathbb{E}\zeta_k = 0, \qquad \text{for each } \theta_k.$$

To verify the rate-of-growth condition, we need consider only, for each θ, the sum of terms involving ζ_k. In fact, by virtue of (A3) and the mixing inequality established by Kushner and Yin (2003) as $l \to 0$,

$$\begin{aligned}\mathbb{E}\left|\sum_{k=m(jT)}^{m(jT+t)-1}(\varepsilon_k/2\delta_k)\zeta_k\right|^2 &= \sum_{k=m(jT)}^{m(jT+t)-1}\sum_{l\ge k}^{m(jT+t)-1}(\varepsilon_k/4\delta_k)(\varepsilon_l/\delta_l)E\zeta_k'\zeta_l\\ &\le \mathbb{K}\sum_{k=m(jT)}^{m(jT+t)-1}(\varepsilon_k^2/\delta_k^2)\sum_{l\ge k}^{m(jT+t)-1}\phi(l-k) \to 0,\end{aligned}$$

where $\mathbb{K}$ is a positive constant. Thus, we also have

$$\lim_l \mathbb{P}\left(\sup_{j\ge l}\max_{0\le t\le T}\left|\sum_{k=m(jT)}^{m(jT+t)-1}(\varepsilon_k/2\delta_k)\zeta_k\right| \ge \mu\right) = 0. \tag{5.31}$$

Owing to (5.29) and (5.31), the rate-of-growth conditions for the noise processes are verified.

Finally, concerning the bias term, in view of the central finite difference, using the smoothness of $G_0(\cdot, \eta)$ [hence, that of $\bar{G}(\cdot)$] and taking the Taylor expansion about θ_l, $\alpha_l = O(\delta_l^2)$. Therefore,

$$\left|\sum_{k=m(jT)}^{m(jT+t)-1}\varepsilon_k\alpha_k \le \mathbb{K}\sum_{k=m(jT)}^{m(jT+t)-1}\varepsilon_k O(\delta_k^2) \le \mathbb{K}O(\delta_{m(jT)}^2)\right|,$$

and hence

$$\lim_l \mathbb{P}\left(\sup_{j\ge l}\max_{0\le t\le T}\left|\sum_{k=m(jT)}^{m(jT+t)-1}\varepsilon_k\alpha_k\right| \ge \mu\right) = 0.$$

Thus, all conditions of Theorem 6.5.1 in Kushner and Yin (2003) are verified. The desired result follows. □

5.3. Rate of convergence

In Section 5.2, we have obtained that $\theta_l \to \theta^*$ w.p.1 under suitable conditions. In this section, we will examine the rate of convergence. As in the classical theory of convergence rate, we will focus on the quantity $y_l = l^{\alpha}(\theta_l - \theta^*)$. Roughly speaking, studying the rate of convergence of the stochastic approximation algorithm is to find a choice of α that leads to a nontrivial limit of y_l in distribution. In the following analysis of rate convergence we further assume θ_i^* is in the interior of $\widetilde{B}$ for each i. That is, without loss of generality we may drop the reflection term in the recursion.

(A4) $\theta_l \to \theta^*$ w.p.1 such that $\theta^* \in M_0$, the interior of M, and θ^* is a globally asymptotically stable point of the ODE (5.27). The set $\{l^{\alpha}(\theta_l - \theta^*)\}$ is tight.

We can take the Taylor expansion of $\bar{G}(\theta_l \pm \delta_l e_j) - G_0(\theta_l \pm \delta_l e_j, \eta_l^{\pm})$ about the point $\theta^* \pm \delta_l e_j$ as follows. Denote

$$\begin{aligned}\zeta_{l,j}^* &= [\bar{G}(\theta^* + \delta_l e_j) - G_0(\theta^* + \delta_l e_j, \eta_l^{+})]\\ &\quad - [\bar{G}(\theta^* - \delta_l e_j) - G_0(\theta^* - \delta_l e_j, \eta_l^{-})],\end{aligned}$$

and denote

$$\zeta_l^* = (\zeta_{l,1}^*, \ldots, \zeta_{l,r}^*)'.$$

In view of the condition in (A3), $\{\zeta_l^*\}$ is a stationary ϕ-mixing sequence. Define

$$w^l(t) = \sum_{k=l}^{m(t_l+t)-1}\frac{\zeta_k^* + \psi_k}{2\sqrt{k}}, \quad t \in [0, \infty).$$

Lemma 5.9. *Under conditions (A2)-(A4), $w^l(\cdot)$ converges weakly to a Brownian motion $w(\cdot)$, whose covariance matrix is given by*

$$\Sigma t = \frac{1}{4}(\Sigma_1 + \Sigma_2)t,$$

where

$$\Sigma_1 = E\zeta_1^*(\zeta_1^*)' + \sum_{k=2}^{\infty}E\zeta_1^*(\zeta_k^*)' + \sum_{k=2}^{\infty}E\zeta_k^*(\zeta_1^*)', \tag{5.32}$$

$$\Sigma_2 = E\psi_k\psi_k'. \tag{5.33}$$

Remark 5.10. By the independence of $\{\zeta_l^*\}$ and $\{\psi_l\}$, $\sum_{k=l}^{m(t_l+t)-1}\zeta_k^*/\sqrt{k}$ and $\sum_{k=l}^{m(t_l+t)-1}\psi_k/\sqrt{k}$ can be treated separately. Each of them converges to a Brownian motion with covariance $\Sigma_1 t$ and $\Sigma_2 t$ respectively by virtue of page 235–241 of Kushner and Yin (2003), the desired result then follows.

To proceed, we rewrite $\varpi_{l,j}$ as

$$\begin{aligned}\varpi_{l,j} &= \bar{G}_{l,j,\theta_{l,j}}(\theta_l) - \frac{\bar{G}(\theta_l + \delta_l e_j) - \bar{G}(\theta_l - \delta_l e_j)}{2\delta_l}\\ &= \bar{G}_{l,j,\theta_{l,j}}(\theta_l) - (\bar{G}_{l,j,\theta_{l,j}}(\theta_l) + C_{l,j}(\theta_l)\delta_l^2) + o(\delta_n^2)\\ &= -C_{l,j}(\theta_l)\delta_l^2 + o(\delta_l^2).\end{aligned}$$

Thus we can rewrite (5.26) as follows:

$$\begin{aligned}\theta_{l+1} &= \theta_l + \varepsilon_n\frac{\zeta_l^* + \psi_l}{2\delta_l} - \varepsilon_l C_l(\theta_n)\delta_l^2\\ &\quad -\varepsilon_n[\bar{G}_l(\theta^*) + \bar{G}_{l,\theta}(\theta^*)(\theta_l - \theta^*)] + o(\delta_l^2) + o(\theta_l - \theta^*)\\ &= \theta_l + \varepsilon_l[-\bar{G}_{l,\theta}(\theta^*)(\theta_l - \theta^*) - C_l(\theta_l)\delta_l^2]\\ &\quad +\varepsilon_l[-\bar{G}_l(\theta^*) + o(\theta_l - \theta^*) + o(\delta_l^2) + \frac{\zeta_l^* + \psi_l}{2\delta_l}].\end{aligned} \tag{5.34}$$

where

$$C_l(\cdot) = (C_{l,1}(\cdot), C_{l,2}(\cdot), \ldots, C_{l,r}(\cdot)) \in \mathbb{R}^r,$$

with

$$C_{l,j}(\cdot)=\frac{1}{3!}\bar{G}_{l,\theta_{l,j}\theta_{l,j}\theta_{l,j}}(\cdot).$$

We further define $y^l(\cdot)$ to be the piecewise constant interpolation of y_l, i.e.,

$$y^l(t)=y_l,\quad \text{for } t\in[t_{l+k}-t_l,t_{l+k+1}-t_l).$$

Then, $y^l(\cdot)\in D^r[0,+\infty)$, the space of $\mathbb{R}^r$-valued functions that are right continuous, have left limits, endowed with the Skorohod topology.

For simplicity, we let $\varepsilon_l=O(1/l)$ and $\delta_l=1/l^\beta$, then

$$((l+1)/l)^\alpha=1+\frac{\alpha}{l}+O(\varepsilon_l).$$

Using the scaling factor l^α and (5.34), we expand

$$\begin{aligned}y_{l+1}=&\left(\frac{l+1}{l}\right)^\alpha y_l+\varepsilon_l\left(\frac{l+1}{l}\right)^\alpha\left(-\bar{G}_{l,\theta}(\theta^*)y_l-C_l(\theta_l)\delta_l^2l^\alpha\right)\\&+\varepsilon_l\left(\frac{l+1}{l}\right)^\alpha\frac{1}{2\delta_l}l^\alpha(\zeta_l^*+\psi_l)+\left(\frac{l+1}{l}\right)^\alpha\varepsilon_l(-\bar{G}_l(\theta^*)l^\alpha)\\&+\varepsilon_l\rho_l\\=&\left(1+\frac{\alpha}{l}+o(\varepsilon_l)\right)y_l-\varepsilon_l\left(\frac{l+1}{l}\right)^\alpha\bar{G}_{l,\theta}(\theta^*)y_l\\&-\frac{(l+1)^\alpha}{l^{2\beta}}\varepsilon_lC_l(\theta^*)\\&+\varepsilon_l(l+1)^\alpha\bar{G}_l(\theta^*)+\frac{1}{2}\varepsilon_l(l+1)^\alpha l^\beta(\zeta_l^*+\psi_l)+\varepsilon_l\rho_l,\end{aligned}$$

where

$$\rho_l=o(\theta_l-\theta^*)(l+1)^\alpha+o(\delta_l^2)(l+1)^\alpha.$$

It is clear that we need require $\alpha-2\beta\le0$ and $\alpha+\beta-1/2\le0$ for the weak convergence hold. By choosing the optimal choice of α and β, we further have $\alpha=1/3$, and $\beta=1/6$. Then

$$\begin{aligned}y_{l+1}=&y_l+\frac{1}{l}\left(\frac{I}{3}-\bar{G}_{l,\theta}(\theta^*)\right)y_l-\frac{1}{l}C_l(\theta^*)\\&+\frac{1}{2\sqrt{l}}(\zeta_l^*+\psi_l)+\frac{1}{l}\rho_l.\quad\square\end{aligned}$$

Theorem 5.11. *Recall that $y_l=l^{1/3}(\theta_l-\theta^*)$, and $y^l(\cdot)$ is its continuous-time interpolation. Suppose that* (A2)-(A4) *are satisfied and $y^l(0)\to y_0$. All eigenvalues of $I/3-\bar{G}_{\theta\theta}(\theta^*)$ have negative real parts, then $y^l(\cdot)$ converges weakly to $y(\cdot)$, which is a solution of the stochastic differential equation*

$$\begin{aligned}&dy(t)=\left(\left(\frac{I}{3}-\bar{G}_{\theta\theta}(\theta^*)\right)y(t)-C(\theta^*)\right)dt+dw,\\&y(0)=y_0,\end{aligned}\tag{5.35}$$

where $C(\theta^*)$ is the limit of $C_l(\theta^*)$ and $w(\cdot)$ is the Brownian motion given in Lemma 5.9.

Proof. The proof is divided into three steps. Using a truncation device, we work with an N-truncation in lieu of the original process, obtain its tightness, and derive its weak convergence. Finally, we let $N\to\infty$ to conclude the proof.

For a fixed but otherwise arbitrary $N>0$, write the truncated version of the recursive formula for y_n as

$$\begin{aligned}y_{l+1}^N=&y_l^N+\frac{1}{l}\left(\left(\frac{I}{3}-\bar{G}_{\theta\theta}(\theta^*)\right)y_l^N-C_l(\theta^*)+\rho_l\right)q_N(y_l^N)\\&+\frac{1}{2\sqrt{l}}(\zeta_l^*+\psi_l),\end{aligned}\tag{5.36}$$

where

$$q_N(y)=1,\quad \text{for } y\in S_N,$$

$$q_N(y)=0,\quad \text{for } y\in\mathbb{R}^r-S_{N+1},$$

where $S_N=\{y:|y|\le N\}$ is the sphere with radius N. Let $y^{l,N}(\cdot)$ be the piecewise constant interpolation of y_l^N. That is,

$$y^{l,N}(t)=y_{l+k}^N,\text{ on } t\in[t_{l+k}-t_l,t_{l+k+1}-t_l).$$

Then, $y^{l,N}(t)=y^l(t)$ up until the first exit from S_N, so it is an N-truncation of $y^l(\cdot)$;

Our first step is to derive the tightness of $\{y^{l,N}(\cdot)\}$. For any $\Delta>0$ and $0<s\le\Delta$, we use $\mathbb{E}_t$ to denote the conditional expectation on $\mathcal{F}_t$, the σ-algebra generated by $\{y_0,\xi_j,\psi_j,j<m(t_l+t)\}$. Then, we have that, by using the ϕ-mixing of $\{\eta_l\}$ and the martingale difference property of $\{\widehat{\eta}_l\}$,

$$\mathbb{E}_t\left|\sum_{k=m(t_l+t)}^{m(t_l+t+s)-1}\frac{1}{2\sqrt{k}}(\zeta_k^*+\psi_k)\right|^2\le K\sum_{k=m(t_l+t)}^{m(t_l+t+s)-1}\frac{1}{k}\le Ks=O(\Delta).$$

Using the boundedness of y_l^N, we have

$$\mathbb{E}_t\left|\sum_{k=m(t_l+t)}^{m(t_l+t+s)-1}\frac{1}{k}\left(\frac{I}{3}-\bar{G}_{\theta\theta}(\theta^*)\right)y_k^Nq_N(y_k^N)\right|^2\le Ks^2=O(\Delta^2),$$

$$\mathbb{E}_t\left|\sum_{k=m(t_l+t)}^{m(t_l+t+s)-1}\frac{1}{k}\left(C_k(\theta^*)+\rho_k\right)q_N(y_k^N)\right|^2\le O(\Delta^2).$$

Combining these yields

$$\lim_{\Delta\to0}\limsup_{l\to\infty}E|y^{l,N}(t+s)-y^{l,N}(t)|^2=0.$$

Hence, the tightness of $\{y^{l,N}(\cdot)\}$ follows by virtue of the criterion in page 47 of Kushner (1984).

By the Prohorov theorem, we can extract a convergent subsequence and still use l as its index for convenience. Next, we figure out the limit process. Choose a sequence $\Delta_l\to0$ satisfying

$$\sup_{j\ge l}\frac{1}{l\Delta_n}\to0,\quad \text{as } n\to0.$$

Divide $[m(t_l+t),m(t_l+t+s)-1]$ into subintervals such that $m(t_l+t)=m_1<m_2<\cdots$ and such that $\sum_{k=m_l}^{m_{l+1}-1}(k^{-1}/\Delta_l)\to1$. Then, it can be shown that

$$\begin{aligned}&\sum_{k=m(t_l+t)}^{m(t_l+t+s)-1}\frac{1}{k}\left(\frac{I}{3}-\bar{G}_{\theta\theta}(\theta^*)\right)y_k^N\\&\quad\to\int_t^{t+s}\left(\frac{I}{3}-\bar{G}_{\theta\theta}(\theta^*)\right)y^N(u)q_N(y^N(u))du.\end{aligned}$$

It can also been shown that

$$\sum_{k=m(t_l+t)}^{m(t_l+t+s)-1}\frac{1}{k}C_k(\theta^*)q_N(y_k^N)\to C(\theta^*)\int_t^{t+s}q_N(y^N(u))du,$$

$$\sum_{k=m(t_l+t)}^{m(t_l+t+s)-1}\frac{1}{k}\rho_kq_N(y_k^N)\to0.$$

Thus, $y^{l,N}(\cdot)$ converges weakly to $y^N(\cdot)$, which is a solution of (5.35) with the coefficients involving $y(\cdot)$ truncated, i.e.,

$$\begin{aligned}&dy^N(t)=\left(\left(\frac{I}{3}-\bar{G}_{\theta\theta}(\theta^*)\right)y^N(t)-C(\theta^*)\right)q_N(y^N(t))dt+dw(t),\\&y(0)=y_0.\end{aligned}$$

The final step is to consider the desired result in an unbounded sphere with $N \to \infty$. Using an argument similar to that of page 283–284 of Kushner and Yin (2003), we conclude that $y^l(\cdot)$ also converges to $y(\cdot)$, which is the solution to the SDE (5.35) as desired. □

6. Concluding remarks

This paper develops a hybrid Markov chain approximation and stochastic approximation-based deep learning method to find the optimal investment, reinsurance, and dividend strategies in a complex stochastic system. An infinite-horizon subject to random ruin time optimization problem is formulated. The value function and controls are approximated by deep neural networks. The Markov chain approximation method locates the initial guesses with coarse scale. A stochastic approximation algorithm is developed to find the optimal parameters of the neural networks with fine scale. The approximating neural networks are proved to converge weekly to the optimal controls. The analysis of convergence rate is presented.

The method is different from most existing numerical methods dealing with optimization problems. Such methods mainly focus on solving for the corresponding HJB equations or quasi-variational inequalities and suffer "curse of dimensionality" due to the exponentially increasing computation nodes with finer discretization. Comparing with the classical finite difference method, the computation efficiency of the proposed two-scale method is significantly improved in a high-dimension case. Further, the accuracy of approximating piecewise controls in finite difference depends on the grid density. Especially when the scales of controls and states are significantly incomparable, finding suitable stepsizes for finite differences are quite difficult. The deep learning method implements the stochastic approximation method to find optimal controls. Hence we can achieve more accurate controls. Moreover, the adoption of MCAM with coarse scale provides a feasible way to determine an initial computation node and relatively learning range in the neighbourhood of optimal values to improve the computation efficiency. This paper provides convergence analysis of the algorithm.

In future studies, we will develop a deep learning algorithm to solve for optimization problem with finite horizon. Since finite-horizon problem has one more dimension of time, we need approximate controls by neural networks at each discrete time by integrating Monte Carlo simulation into current algorithm. When approximating complicated expectations, Monte Carlo simulation is of more time efficient than lattice-based iterative methods. Then the amounts of computation nodes only increase linearly with respect to the number of sample paths. Hence, the computation efficiency is largely improved.

Further, we can develop deep learning algorithms to analyse time-inconsistent dividend optimization problems, where HJB equations are generally not available due to non-exponential discounting. The deep learning algorithm will directly approximate the controls and value functions and will provide us some new insights about the forms of optimal dividend strategies comparing with traditional barrier strategies.

Acknowledgements

We are grateful to the editors and anonymous referees for their insightful comments and suggestions. These comments/suggestions greatly improved the quality and readability of the paper. Z. Jin and H. Yang thank the support of the Research Grants Council of the Hong Kong Special Administrative Region (project no. 17330816), Z. Jin's research was also supported by a Faculty Research Grant from The University of Melbourne, Australia. G. Yin's research was supported in part by the National Science Foundation, United States under grant DMS-1710827.

References

Albrecher, H., Beirlant, J., Teugels, J.L., 2017. Reinsurance: Actuarial and Statistical Aspects. Wiley, West Sussex.
Aleandri, M., 2018. Modeling Dynamic Policyholder Behaviour Through Machine Learning Techniques. Working Paper.
Arrow, K., 1963. Uncertainty and the welfare economics of medical care. Amer. Econ. Rev. 53, 941–973.
Bachouch, A., Huré, C., Langrené, N., Pham, H., 2018. Deep neural networks algorithms for stochastic control problems on finite horizon, part 2: numerical applications. arXiv preprint arXiv:1812.05916.
Borch, K., 1960. Reciprocal reinsurance treaties. Astin Bull. 1 (4), 170–191.
Browne, S., 1995. Optimal investment policies for a firm with a random risk process: Exponential utility and minimizing the probability of ruin. Math. Oper. Res. 20 (4), 937–958.
Cheng, X., Jin, Z., Yang, H., 2020. Optimal insurance strategies: A hybrid deep learning Markov chain approximation approach. ASTIN Bull. J. IAA 50 (2), 449–477.
De Finetti, B., 1957. Su un'impostazione alternativa della teoria collettiva del rischio. In: Transactions of the XVth International Congress of Actuaries, Vol. 2. pp. 433–443.
E, W., Han, J., Jentzen, A., 2017. Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations. Commun. Math. Statist. 5 (4), 349–380.
Gerber, H.U., 1972. Games of economic survival with discrete and continuous income processes. Oper. Res. 20 (1), 37–45.
Hainaut, D., 2018. A neural-network analyzer for mortality forecast. ASTIN Bull. J. IAA 48 (2), 481–508.
Han, J., E, W., 2016. Deep learning approximation for stochastic control problems. arXiv preprint arXiv:1611.07422.
Hipp, C., Plum, M., 2000. Optimal investment for insurers. Insurance Math. Econom. 27, 215–228.
Højgaard, B.H., Taksar, M., 1999. Controlling risk exposure and dividends payout schemes: insurance company example. Math. Finance 9 (2), 153–182.
Huré, C., Pham, H., Bachouch, A., Langrené, N., 2018. Deep neural networks algorithms for stochastic control problems on finite horizon, part I: convergence analysis. arXiv preprint arXiv:1812.04300.
Jin, Z., Yang, H., Yin, G., 2013a. Numerical methods for optimal dividend payment and investment strategies of regime-switching jump diffusion models with capital injections. Automatica 49 (8), 2317–2329.
Jin, Z., Yin, G., Wu, F., 2013b. Optimal reinsurance strategies in regime-switching jump diffusion models: stochastic differential game formulation and numerical methods. Insurance: Math. Econom. 53 (2013), 733–746.
Jin, Z., Yin, G., Zhu, C., 2012. Numerical solutions of optimal risk control and dividend optimization policies under a generalized singular control formulation. Automatica 48 (8), 1489–1501.
Kushner, H., 1984. Approximation and Weak Convergence Methods for Random Processes, with Applications to Stochastic Systems Theory. MIT Press, Cambridge, Massachusetts.
Kushner, H., Dupuis, P., 2001. Numerical Methods for Stochstic Control Problems in Continuous Time, second ed. In: Stochastic Modelling and Applied Probability, vol. 24, Springer, New York.
Kushner, H., Yin, G., 2003. Stochastic Approximation and Recursive Algorithms and Applications, second ed. In: Stochastic Modelling and Applied Probability, vol. 35, Springer, New York.
Markowitz, H., 1952. Portfolio selection. J. Finance 7, 77–91.
Miller, M., Modigliani, F., 1961. Dividend policy, growth, and the valuation of shares. J. Bus. 34 (4), 411–433.
Nguyen, N., Yin, G., 2020. Stochastic approximation with discontinuous dynamics, differential inclusions, and applications. submitted.
Van Staden, P.M., Dang, D.M., Forsyth, P.A., 2018. Time-consistent mean–variance portfolio optimization: A numerical impulse control approach. Insurance Math. Econom. 83, 9–28.
Wüthrich, M.V., 2018a. Machine learning in individual claims reserving. Scand. Actuar. J. 6, 465–480.
Wüthrich, M.V., 2018b. Neural networks applied to chain-ladder reserving. Eur. Actuar. J. 8, 407–436.
Wüthrich, M.V., Buser, C., 2017. Data Analytics for Non-Life Insurance Pricing. Swiss Finance Institute Research Paper No. 16–68.
Yang, H., Zhang, L., 2005. Optimal investment for insurer with jump-diffusion risk process. Insurance Math. Econom. 37 (3), 615–634.
Yin, G., Liu, R., Zhang, Q., 2002. Recursive algorithms for stock liquidation: a stochastic optimization approach. SIAM J. Optim. 13 (1), 240–263.
Yin, G., Zhang, Q., Badowski, G., 2003. Discrete-time singularly perturbed Markov chains: Aggregation, occupation measures, and switching diffusion limit. Adv. Appl. Probab. 35, 449–476.