Applications of Numerical Methods for Stochastic Controlled Switching Diffusions with a Hidden Markov Chain: Case Studies on Distributed Power Management and Communication Resource Allocation

Zhixin Yang[1], Le Yi Wang[2], George Yin[3](✉), Qing Zhang[4], and Hongwei Zhang[5]

[1] Department of Mathematics, University of Wisconsin-Eau Claire, Eau Claire, WI 54701, USA
yangzhix@uwec.edu

[2] Department of Electrical and Computer Engineering, Wayne State University, Detroit, MI 48202, USA
lywang@wayne.edu

[3] Department of Mathematics, Wayne State University, Detroit, MI 48202, USA
gyin@math.wayne.edu

[4] Department of Mathematics, The University of Georgia, Athens, GA 30602, USA
qingz@math.uga.edu

[5] Department of Computer Science, Wayne State University, Detroit, MI 48202, USA
hongwei@wayne.edu

1 Introduction

Recently, considerable attention has been drawn to stochastic controlled systems with hidden Markov chains. Much motivation stems from applications in distributed power management and platoon inter-vehicle distance maintenance, among others. The dynamic systems of interest are controlled diffusions with switching, known as switching diffusions [6]. Different from the extensive studies contained in the aforementioned reference, the switching process in this paper is assumed to be a continuous-time Markov chain that is hidden. We can only observe the state of the Markov chain with additive noise. Mean-variance control problems were first considered in the Nobel prize winning paper of Markowitz [2]. It was subsequently considered by a host of researchers. The recent advances in backward stochastic differential equations enable the treatment of the mean-variance controls in continuous time, which is otherwise impossible because of the so-called indefinite control weights; see Zhou and Li [7] for the first paper in this direction and further details. Further work in conjunction with regime-switching models can be found in Zhou and Yin [8], among others.

As a new twist of the mean-variance portfolio selections, our recent work focuses on using the mean-variance formulation to treat networked control systems. That is, we borrow the idea in financial engineering to treat problems

I. Dimov et al. (Eds.): FDM 2014, LNCS 9045, pp. 117–128, 2015.
DOI: 10.1007/978-3-319-20239-6_11

arising in networked control problems. Much of the motivation stems from applications arising in cyber-physical systems. It has been observed in [4] that a large class of problems arising from networked systems and platoon controls can be formulated as such systems, similar to the mean variance control problems that were originally pursued in financial engineering [8]. In [4], we outlined three potential applications in platoon controls based on mean-variance controls. The first problem concerns the longitudinal inter-vehicle distance control. To increase highway utility, it is desirable to reduce the total length of a platoon, resulting in smaller overall inter-vehicle distances. The drawback of this strategy, however, is the increase in the risk of collision due to traffic uncertainties. The task of minimizing the risk with desired inter-vehicle distance fits naturally to a mean-variance optimization framework. The second one is communication resource allocation of bandwidths for vehicle-to-vehicle (V2V) communications. For a given maximum throughput of a platoon communication system, the communication system operator must find a way to assign this resource to different V2V channels, which may also be formulated as a mean-variance control problem. The third one is the platoon fuel consumption. Due to variations in vehicle sizes and speeds, each vehicle's fuel consumption is a controlled random process. Tradeoff between a platoon's team acceleration/maneuver capability and fuel consumption can be summarized in a desired platoon fuel consumption rate. Assigning fuels to different vehicles results in coordination of vehicle operations modeled by subsystem fuel rate dynamics. This problem may also be casted into the framework of mean-variance control. Such problems are highly nonlinear, it is virtually impossible to find closed-form solutions. Our objective is thus devoted to finding feasible algorithms for the desired tasks. Recently, in our work [5], numerical approximation methods have been developed. The convergence of the algorithms is proved. The basic idea is first to convert the partially observable stochastic control problems to completely observed systems by means of the Wonham filtering methodologies. Then we use relaxed controls and Markov chain approximation techniques to build convergent numerical schemes. Based on that work, this paper aims to provide case studies of two typical problems in applications. Our main effort is to demonstrate using numerical methods solving the problems arising in the specific applications.

The rest of the paper is arranged as follows. Section 2 formulates the problem. Section 3 introduces the Markov chain approximation methods and provides the approximation of the optimal controls. Sections 4 and 5 present two case studies to illustrate the wide applications of the scheme developed in our work.

2 Problem Formulation

Consider a given probability space $(\Omega, \mathcal{F}, P)$ in which there is $w_1(t)$, a standard ρ-dimensional Brownian motion with $w_1(t) = (w_1^1(t), w_1^2(t), \dots, w_1^\rho(t))'$, where z' denotes the transpose of z. Let $\alpha(t)$ be a continuous-time finite-state Markov chain, independent of $w_1(t)$, taking values in $\mathcal{M} = \{1, 2, \dots, m\}$ with generator $Q = (q_{ij})_{m\times m}$. We consider a networked system that consists of $\rho + 1$ nodes (subsystems), which is modeled for $t \in [s, T]$ by

$$\begin{aligned} dx_0(t) &= \mu_0(t, \alpha(t))x_0(t)dt,\ x_0(s) = x_0, \\ dx_l(t) &= x_l(t)\mu_l(t, \alpha(t))dt + x_l(t)\bar{\sigma}_l(t, \alpha(t))dw_1(t),\ x_l(s) = x_l, l = 1, \ldots, \rho, \end{aligned} \tag{1.1}$$

where for each i, $\mu_l(t, i)$ is the drift and $\bar{\sigma}_l(t, i) = (\bar{\sigma}_{l1}(t, i), \ldots, \bar{\sigma}_{l\rho}(t, i))$ is the volatility for the lth node. In our framework, instead of having full information of the Markov chain, we can only observe

$$dy(t) = g(\alpha(t))dt + \beta dw_2(t),\ y(s) = 0, \tag{1.2}$$

where $\beta > 0$ and $w_2(\cdot)$ is a standard scalar Brownian motion, $w_1(\cdot)$, $w_2(\cdot)$, and $\alpha(\cdot)$ are independent. Moreover, the initial data $p(s) = p = (p^1, p^2, \ldots, p^m)$ in which $p^i = p^i(s) = P(\alpha(s) = i)$ is given for $1 \le i \le m$. By distributing the portion $N_l(t)$ of the lth node's flow $x_l(t)$ at time t and denoting the total flows for the whole networked system as $x(t)$, we have $x(t) = \sum_{l=0}^{\rho} N_l(t)x_l(t), t \ge s$. With $x(s) = \sum_{l=0}^{\rho} N_l(s)x_l(s) = x$, the dynamics of $x(t)$ are given as

$$dx(t) = [x(t)\mu_0(t, \alpha(t)) + M(t, \alpha(t))\pi(t)]dt + \pi'(t)\bar{\sigma}(t, \alpha(t))dw_1(t), \tag{1.3}$$

in which $\pi(t) = (\pi_1(t), \ldots, \pi_\rho(t))'$ and $\pi_l(t) = N_l(t)x_l(t)$ for $l = 1, \ldots, \rho$ is the actual flow of the network system for the lth node and $\pi_0(t) = x(t) - \sum_{l=1}^{\rho} \pi_l(t)$ is the actual flow of the networked system for the first node, and $M(t, \alpha(t)) = (\mu_i(t, \alpha(t)) - \mu_0(t, \alpha(t)) : i = 1, \ldots, \rho)$ and $\bar{\sigma}(t, \alpha(t)) = (\bar{\sigma}_{lj}(t, \alpha(t)))_{\rho\times\rho}$. We define $\mathcal{F}_t = \sigma\{w_1(\widetilde{s}), y(\widetilde{s}), x(s) : s \le \widetilde{s} \le t\}$. Our objective is to find an $\mathcal{F}_t$ admissible control $\pi(\cdot)$ in a compact set $\mathbb{\pi}$ under the constraint that the expected terminal flow is $Ex(T) = \kappa$ for some given $\kappa \in \mathbb{R}$, so that the risk measured by the variance of the terminal flow is minimized. Specifically, we have the following goal

$$\begin{aligned} &\min\ J(s, x, p, \pi(\cdot)) := E[x(T) - \kappa]^2 \\ &\text{subject to } Ex(T) = \kappa. \end{aligned} \tag{1.4}$$

We apply the Lagrange multiplier techniques (see, e.g.,[7]) to arrive at the unconstrained optimization problem

$$\begin{aligned} &\min\ J(s, x, p, \pi(\cdot), \lambda) := E[x(T) + \lambda - \kappa]^2 - \lambda^2 \\ &\text{subject to } (x(\cdot), \pi(\cdot)) \text{ admissible}, \end{aligned} \tag{1.5}$$

where λ is the Lagrange multiplier. A pair $(\sqrt{\text{Var }(x(T))}, \kappa) \in \mathbb{R}^2$, corresponding to the optimal control if it exists, is called an *efficient point.*

Next, to treat the partially observed control problem, let $p^i(t) = P(\alpha(t) = i|\mathcal{F}^y(t))$ for $i = 1, 2, \ldots, m$, with $p(t) = (p^1(t), \ldots, p^m(t)) \in \mathbb{R}^{1\times m}$ and $\mathcal{F}^y(t) = \sigma\{y(\widetilde{s}) : s \le \widetilde{s} \le t\}$. It was shown in [3] that this conditional probability satisfies the following system of stochastic differential equations

$$dp^i(t) = \sum_{j=1}^{m} q^{ji}p^j(t)dt + \frac{1}{\beta}p^i(t)(g(i) - \overline{\alpha}(t))d\widehat{w}_2(t),\ p^i(s) = p^i, i = 1, \ldots, m \tag{1.6}$$

where $\overline{\alpha}(t) = \sum_{i=1}^{m} g(i)p^i(t)$ and $\widehat{w}_2(t)$ is the innovation process. Now we have a completely observable system so that $x(s) = x$, $p^i(s) = p^i$, and

$$\begin{aligned} dx(t) &= \mu(x(t), p(t), \pi(t))dt + \sigma(x(t), p(t), \pi(t))dw_1(t) \\ dp^i(t) &= \sum_{j=1}^{m} q^{ji}p^j(t)dt + \frac{1}{\beta}p^i(t)(g(i) - \overline{\alpha}(t))d\widehat{w}_2(t), \text{ for } i \in \{1, \dots, m\} \end{aligned} \tag{1.7}$$

where

$$\begin{aligned} \mu(x(t), p(t), \pi(t)) &= \sum_{i=1}^{m} \mu_0(t, i)p^i(t)x(t) + \sum_{l=1}^{\rho}\sum_{i=1}^{m}(\mu_l(t, i) - \mu_0(t, i))p^i(t)\pi_l(t) \\ \sigma(x(t), p(t), \pi(t))dw_1(t) &= \sum_{l=1}^{\rho}\sum_{j=1}^{\rho}\sum_{i=1}^{m} \pi_l(t)\bar{\sigma}_{lj}(t, i)p^i(t)dw_1^j(t). \end{aligned}$$

For an arbitrary $\mathbb{\Pi}$ and $\phi(\cdot, \cdot, \cdot) \in C^{1,2,2}(\mathbb{R})$, consider the operator

$$\begin{aligned} \mathcal{L}^r\phi(s, x, p) &= \frac{\partial \phi}{\partial s} + \frac{\partial \phi}{\partial x}\mu(x, p, r) + \frac{1}{2}\frac{\partial^2 \phi}{\partial x^2}[\sigma(x, p, r)\sigma'(x, p, r)] \\ &\quad + \sum_{i=1}^{m} \frac{\partial \phi}{\partial p^i}\sum_{j=1}^{m} q^{ji}p^j + \frac{1}{2}\sum_{i=1}^{m}\frac{\partial^2 \phi}{\partial (p^i)^2}\frac{1}{\beta^2}[p^i(g(i) - \overline{\alpha})]^2. \end{aligned} \tag{1.8}$$

Let $W(s, x, p, \pi)$ be the objective function with $E^{\pi}_{s,x,p}$ denoting the expectation of functionals on $[s, T]$ with $x(s) = x, p(s) = p$, the admissible control $\pi = \pi(\cdot)$, and the value function $V(s, x, p)$

$$V(s, x, p) = \inf_{\pi \in \mathbb{\Pi}} W(s, x, p, \pi) = \inf_{\pi \in \mathbb{\Pi}} E^{\pi}_{s,x,p}(x(T) + \lambda - k)^2 - \lambda^2. \tag{1.9}$$

The value function is a solution of the following equation

$$\inf_{r \in \mathbb{\Pi}} \mathcal{L}^r V(s, x, p) = 0, \tag{1.10}$$

with boundary condition $V(T, x, p) = (x(T) + \lambda - \kappa)^2 - \lambda^2$. Note that (1.10) is known as the Hamilton-Jacobi-Bellman (HJB) equation. To proceed, we use the relaxed control representation. For the σ-algebra $\mathcal{B}(\mathbb{\Pi})$ and $\mathcal{B}(\mathbb{\Pi} \times [s, T])$ of Borel subsets of $\mathbb{\Pi}$ and $\mathbb{\Pi} \times [s, T]$, an admissible relaxed control or simply a relaxed control $m(\cdot)$ is a measure on $\mathcal{B}(\mathbb{\Pi} \times [s, T])$ such that $m(\mathbb{\Pi} \times [s, t]) = t - s$ for all $t \in [s, T]$ for all $t \in [s, T]$. For notational simplicity, for any $B \in \mathcal{B}(\mathbb{\Pi})$, we write $m(B \times [s, T])$ as $m(B, T - s)$. Since $m(\mathbb{\Pi} \times [s, t]) = t - s$ for all $t \in [s, T]$ and $m(B, \cdot)$ is nondecreasing, it is absolutely continuous. Hence the derivative $\dot{m}(B, t) = m_t(B)$ exists almost everywhere for each B. We can further define the relaxed control representation $m(\cdot)$ of $\pi(\cdot)$ by $m_t(B) = I_{\{\pi(t) \in B\}}$ for any $B \in \mathcal{B}(\mathbb{\Pi})$. We say that $M(\cdot)$ is a measure-value $\mathcal{F}_t$ martingale with values $M(B, t)$ if $M(B, \cdot)$ is an $\mathcal{F}_t$ martingale for each $B \in \mathbb{\Pi}$, and for each t, the following holds: $\sup_{B \in \mathbb{\Pi}} EM^2(B, t) < \infty$,

$M(A \cup B, t) = M(A,t) + M(B,t)$ w.p.1. for all disjoint $A, B \in \mathcal{B}(\mathbb{U})$, and $EM^2(B_n, t) \to 0$ if $B_n \to \emptyset$. We say that $M(\cdot)$ is orthogonal if $M(A,\cdot)$ and $M(B,\cdot)$ are $\mathcal{F}_t$ martingales whenever $A \cap B = \emptyset$. If $M(\cdot)$, $\bar{M}(\cdot)$ are $\mathcal{F}_t$ martingale measures and $M(A,\cdot)$, $\bar{M}(B,\cdot)$ are $\mathcal{F}_t$ martingales for any Borel set A, B, then $M(\cdot)$ and $\bar{M}(\cdot)$ are said to be strongly orthogonal. Letting $M(\cdot) = (M_1(\cdot), \ldots, M_\rho(\cdot))'$, a vector valued martingale measure, we impose the following conditions.

(A1) $M(\cdot) = (M_1(\cdot), \ldots, M_\rho(\cdot))'$ is square integrable and continuous; each component is orthogonal; and the pairs are strongly orthogonal.

Under (A1), there are measure-valued random processes $m_i(\cdot)$ such that the quadratic variation processes satisfy, for each t and $A, B \in \mathcal{B}(\mathbb{U})$, $\big\langle M_i(A,\cdot),\, M_j(B,\cdot)\big\rangle(t) = \delta_{ij} m_i(A \cap B, t)$.

(A2) m_i does not depend on i, $m_i(\cdot) = m(\cdot)$, and $m(\mathbb{U}, t) = t$ for all t.

With the help of the martingale measures and relaxed controls, we can represent our control system in the following way:

$$
\begin{aligned}
x(t) &= x + \int_s^t \int_{\mathbb{U}} \mu(x(z), p(z), c) m_z(dc)dz + \int_s^t \int_{\mathbb{U}} \sigma(x(z), p(z), c) M(dc, dz) \\
p^i(t) &= \int_s^t \sum_{j=1}^m q^{ji} p^j(z)dz + \int_s^t \frac{1}{\beta}[p^i(z)(g(i) - \overline{\alpha}(z))]d\widehat{w}_2(z),\ i \in \{1, \ldots, m\}
\end{aligned}
$$

(A3) $\mu(\cdot,\cdot,\cdot)$ and $\sigma(\cdot,\cdot,\cdot)$ are continuous; $\mu(\cdot, p, c)$ and $\sigma(\cdot, p, c)$ are Lipschitz continuous uniformly in p, c and bounded.

(A4) $\sigma(x,p,c) = (\sigma_1(x,p,c), \ldots, \sigma_\rho(x,p,c)) > 0$.

3 Approximation Algorithms

To facilitate subsequent numerical computations, let $v^i(t) = \log p^i(t)$. Itô's rule leads to the dynamics of $v^i(t)$. We can then obtain the following discrete-time approximation of the Wonham filter

$$
\begin{aligned}
v_{n+1}^{h_2,i} &= v_n^{h_2,i} + h_2\Big[\sum_{j=1}^m q^{ji}\frac{p_n^{h_2,j}}{p_n^{h_2,i}} - \frac{1}{2\beta^2}(g(i) - \bar{\alpha}_n^{h_2})^2\Big] + \sqrt{h_2}\frac{1}{\beta}(g(i) - \bar{\alpha}_n^{h_2})\varepsilon_n, \\
v_0^{h_2,i} &= \log(p^i),\ p_{n+1}^{h_2,i} = \exp(v_{n+1}^{h_2,i}),
\end{aligned} \tag{1.11}
$$

where $\bar{\alpha}_n^{h_2} = \sum_{i=1}^m g(i) p_n^{h_2,i}$ and $\{\varepsilon_n\}$ is a sequence of i.i.d. random variables satisfying $E\varepsilon_n = 0$, $E\varepsilon_n^2 = 1$, and $E|\varepsilon_n|^{2+\gamma} < \infty$ for some $\gamma > 0$ with $\varepsilon_n = \frac{\widehat{w}_2((n+1)h_2) - \widehat{w}_2(nh_2)}{\sqrt{h_2}}$. Here $p_n^{h_2,i}$ appeared as a denominator in (1.11) and we have concentrated on the case that $p_n^{h_2,i}$ stays away from 0. Let $h_1 > 0$ be a discretization parameter for state variables, and recall that $h_2 > 0$ is the step size for the time variable. We construct a discrete-time finite-states Markov chain to approximate the controlled diffusion process, $x(t)$. Let $N_{h_2} = (T-s)/h_2$ be an integer and define $S_{h_1} = \{x : x = kh_1, k = 0, \pm 1, \pm 2, \ldots\}$. We use $\pi_n^{h_1,h_2}$ to denote the random variable that is the control action for the chain at discrete time n. Let $\pi^{h_1,h_2} = (\pi_0^{h_1,h_2}, \pi_1^{h_1,h_2}, \ldots)$ denote the sequence of $\mathbb{\pi}$-valued random variables which are the control actions at time $0, 1, \ldots$ and $p^{h_2} = (p_0^{h_2}, p_1^{h_2}, \ldots)$ be the corresponding posterior probabilities in which $p_n^{h_2} = (p_n^{h_2,1}, p_n^{h_2,2}, \ldots, p_n^{h_2,m})$. We define the difference $\Delta\xi_n^{h_1,h_2} = \xi_{n+1}^{h_1,h_2} - \xi_n^{h_1,h_2}$ and let $E_{x,p,n}^{h_1,h_2,r}$, $V_{x,p,n}^{h_1,h_2,r}$ denote the conditional expectation and variance given $\{\xi_k^{h_1,h_2}, \pi_k^{h_1,h_2}, p_k^{h_2}, k \le n, \xi_n^{h_1,h_2} = x, p_n^{h_2} = p, \pi_n^{h_1,h_2} = r\}$. By stating that $\{\xi_n^{h_1,h_2}, n < \infty\}$ is a controlled discrete-time Markov chain on a discrete-time state space S_{h_1} with transition probabilities denoted by $p^{h_1,h_2}((x,y)|r,p)$, we mean that the transition probabilities are functions of a control variable r and posterior probability p. The sequence $\{\xi_n^{h_1,h_2}, n < \infty\}$ is said to be locally consistent with (1.7) if it satisfies

$$\begin{aligned} &E_{x,p,n}^{h_1,h_2,r}\Delta\xi_n^{h_1,h_2} = \mu(x,p,r)h_2 + o(h_2), \\ &V_{x,p,n}^{h_1,h_2,r}\Delta\xi_n^{h_1,h_2} = \sigma(x,p,r)\sigma'(x,p,r)h_2 + o(h_2), \\ &\sup_n |\Delta\xi_n^{h_1,h_2}| \to 0, \text{ as } h_1, h_2 \to 0. \end{aligned} \tag{1.12}$$

With the approximating Markov chain given above, we can approximate the cost function $W^{h_1,h_2}(s,x,p,\pi^{h_1,h_2})$ in which $x(T)$ is replaced by $\xi_{N_{h_2}}^{h_1,h_2}$ and can find approximation of $V(s,x,p)$. Now we will proceed to find a reasonable Markov chain that is locally consistent. We first suppose that the control space has a unique admissible control $\pi^{h_1,h_2} \in \mathbb{\pi}^{h_1,h_2}$, so that we can drop inf in (1.10). We discretize (1.8) by a finite difference method using step-size $h_1 > 0$ for the state variable and $h_2 > 0$ for the time variable as mentioned above. For simplicity, we omit the details. We can show that the approximating Markov chain constructed above satisfies local consistency. Note that we have used local transitions here so that we can avoid the problem of "numerical noise"or "numerical viscosity", which appears in non-local transitions cases, and is even more serious in higher dimension scenarios, see [1] for more details. We omit most of the details and please refer to [5] for further demonstration.

It can be shown that the Markov chain $\{\xi_n^{h_1,h_2}, n < \infty\}$ with transition probabilities $p^{h_1,h_2}(\cdot)$ properly defined is locally consistent with (1.7). Next, we give the discrete-time approximation algorithm for the controlled Markov chain. Based on the local consistency, we can represent $\xi_{n+1}^{h_1,h_2}$ as

$$\xi_{n+1}^{h_1,h_2} = \xi_n^{h_1,h_2} + \mu(\xi_n^{h_1,h_2}, p_n^{h_2}, \pi_n^{h_1,h_2})h_2 + \sigma(\xi_n^{h_1,h_2}, p_n^{h_2}, \pi_n^{h_1,h_2})\Delta w_n^{h_1,h_2} + o(1), \tag{1.13}$$

where $o(1)$ can be written as $\varepsilon_n^{h_1,h_2}$ in which $\varepsilon_n^{h_1,h_2} \to 0$ as $h_1, h_2 \to 0$. To approximate the continuous-time process $(x(t), p(t), m(t), M(t))$, we use continuous-time interpolation. For $t \in [nh_2, (n+1)h_2)$, we define the piecewise constant interpolations by

$$\xi^{h_1,h_2}(t) = \xi_n^{h_1,h_2},\ p^{h_2}(t) = p_n^{h_2},\ \bar{\alpha}^{h_1,h_2}(t) = \sum_{i=1}^{m} g(i)p_n^{h_2},\ \pi^{h_1,h_2}(t) = \pi_n^{h_1,h_2},$$
$$z^{h_2}(t) = n,\ w_l^{h_1,h_2}(t) = \sum_{k=0}^{z^{h_2}(t)-1} \Delta w_{l,k}^{h_1,h_2},\ \varepsilon^{h_1,h_2}(t) = \varepsilon_n^{h_1,h_2}. \tag{1.14}$$

With most of the technical details omitted, which can be found in [5], we present the main approximation theorem below.

Theorem 1. *Assuming* (A1)-(A4)*, let* $\{\xi_n^{h_1,h_2}, n < \infty\}$*, the approximating chain be constructed with transition probabilities properly defined. Let* $\{\pi_n^{h_1,h_2}, n < \infty\}$ *be a sequence of admissible controls,* $\xi^{h_1,h_2}(\cdot)$ *and* $p^{h_2}(\cdot)$ *be the continuous time interpolation defined in* (1.14)*,* $m^{h_1,h_2}(\cdot)$ *be the relaxed control representation of* $\pi^{h_1,h_2}(\cdot)$ *(continuous time interpolation of* $\pi_n^{h_1,h_2}$*). Then*

$$(\xi^{h_1,h_2}(\cdot), p^{h_2}(\cdot), m^{h_1,h_2}(\cdot), M^{h_1,h_2}(\cdot))\ \textit{istight},$$

$(\xi^{h_1,h_2}(\cdot), p^{h_2}(\cdot), m^{h_1,h_2}(\cdot), M^{h_1,h_2}(\cdot))$ *converges weakly to* $(x(\cdot), p(\cdot), m(\cdot), M(\cdot))$*, and* $W(s, x, p, m^{h_1,h_2}) \to W(s, x, p, m)$*. Denoting the limit of a weakly convergent subsequence by* $(x(\cdot), p(\cdot), m(\cdot), M(\cdot))$*, the martingale measure* $M(\cdot)$ *has quadratic variation process given by* $m(\cdot)$ *and the desired limit dynamics hold. Moreover,* $V^{h_1,h_2}(s, x, p) \to V(s, x, p)$ *as* $h_1 \to 0$ *and* $h_2 \to 0$*.*

4 Case Study I: Distributed Power Management

Consider a distribution network of three renewable energy generators and energy storage devices. Typically, the distributed generators can be photovoltaic (PV) systems, wind turbines, bio-engines, fuel cells, etc. Energy storage devices can be batteries, super-capacitors, etc. To be concrete, let $x_i(t)$, $i = 1, 2, 3$ be the maximum power generating capacity of the ith generator at time t. In addition, $x_0(t)$ is the available maximum capacity that is allowed to be purchased from the main grid at t.

Let $N_i(t)$ be the portion of the power generated by the ith generator that is used to satisfy total power demand, Then, the total locally generated power at time t is $\sum_{i=1}^3 N_i(t)x_i(t)$. Implicitly, the remaining power will be purchased from the main grid, i.e., $\pi_0(t) = N_0(t)x_0(t) = x(t) - \sum_{i=1}^3 N_i(t)x_i(t)$. A renewable generator's maximum capacity is a stochastic process. For example, a wind turbine's maximum power is determined by the wind speed and direction. Similarly, a PV system's output is determined by how much solar radiation is available at a given time, weather condition, and the angle that the sunlight is shining on the solar panels. Here, $\{x_i(t) :\ i = 0, 1, \ldots, 3\}$ is given by (1.1) with $\alpha(t)$ being a 3-state switching process which takes values in $\{1, 2, 3\}$ with generator

$Q = \begin{pmatrix} -0.5 & 0,2 & 0.3 \\ 0.3 & -0.6 & 0.3 \\ 0.4 & 0.4 & -0.8 \end{pmatrix}$, $\mu_1(\alpha) = 2\alpha$, $\mu_2(\alpha) = \alpha + 1$, $\mu_3(\alpha) = \alpha + 2$,

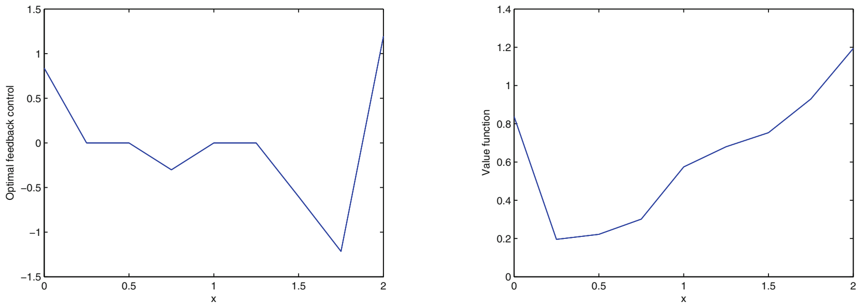
$\sigma_1(\alpha) = (\alpha, 0, 0)$, $\sigma_2(\alpha) = (0, \frac{\alpha}{2}, 0)$ and $\sigma_3(\alpha) = (0, 0, \frac{\alpha}{3})$, for $\alpha = 1, 2, 3$, and $w_1(t) \in \mathbb{R}^3$. Here, the drift term represents average solar radiation values throughout a day; and diffusion term represents solar radiation fluctuations which are caused by many factors such as clouds, weather conditions, etc. The dynamics of the process depend on an event variable α which reflects system structural changes. This is exemplified by scheduled or emergency maintenance of solar modules, failure of a battery cell, addition of super-capacitor banks, tap changes in transformer actions, etc.

It is noted that sometimes such switching actions α cannot be observed directly, such as solar or battery cell failures. However, such switching actions will affect certain measured variables. For example, battery cell failures will cause a jump in terminal voltages. In this study, instead of direct access to α, we assume (1.2) is observable where $g(1) = 1, g(2) = 2$ and $g(3) = 3$, $\beta = 1 > 0$ is a constant, and $w_2(t)$ is a Brownian motion, independent of $w_1(t)$. $y(t)$ is a measured quantity. Distributed power management aims to decide dispatching parameters $N_i(t)$, $i = 1, \ldots, 3$. This can be formulated as a mean-variance control problem. To meet the total power consumption demand $z = 1\,\text{MW}$ (mega watts), it is required that we have the constraint $Ex(T) = z$. On the other hand, to maintain grid stability, smooth operations, and reduced waste, it is desirable that generation-consumption disparity in transient be as small as possible. It is well understood in traditional power flow analysis that transient power fluctuations cause energy loss on lines, affect voltage and frequency stability. In view of (1.4), the Lagrange multiplier technique leads to (1.5). The value function and corresponding control are in Fig. 1 in which x axis is the possible consumption demand of all the generators in the system at $T = 2$ and y axis represent the feedback control π_1 for the first generator and value function V, respectively. The efficient frontier is demonstrated in Fig. 2 in which the x axis is the standard deviation of total generation-consumption of the system and y axis is the expected power consumption. We use the simplex method to find out the optimal λ.

5 Case Study II: Communication Resource Allocation

The second case study is concerned with communication resource allocation of bandwidths for vehicle-to-vehicle (V2V) communications. For a given maximum throughput of a platoon communication system, the communication system must find a way to assign this resource to different V2V channels. If the total bandwidth used is lower than the assigned bandwidth, there will be a waste of resource. Conversely, usage of bandwidths over the budget may incur high costs or interfere with other platoons' operations. In this case, each channel's bandwidth usage is the state of the subsystem. Their summation is a random process and is desired to approach the maximum throughput (the desired mean at the terminal time) with variations as small as possible. Consequently, it becomes a mean-variance control problem.

Consider a platoon of five vehicles. Let $B_i(t)$, $i = 0, 1, \ldots, 4$ be the maximum transmission data rate of vehicle i at time t. In practice, the maximum



(a) Optimal feedback control $\pi_1(t) = N_1(t)x_1(t)$ for the first generator by using the step size $h_1 = 0.25$ for the state variable and step size $h_2 = 0.001$ for the time variable with the fixed expectation $z = 1$ MW

(b) Approximate value function V by using the step size $h_1 = 0.25$ for the state variable and step size $h_2 = 0.001$ for the time variable with fixed expectation $z = 1$ MW

Fig. 1. Optimal control for the first generator and value function V for the power management system

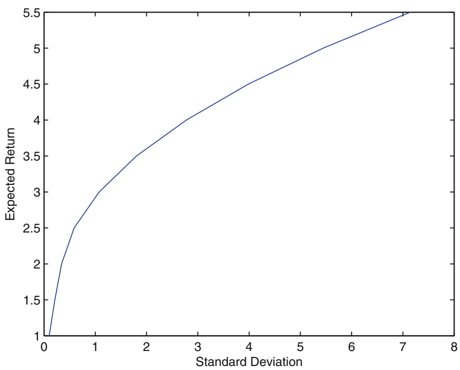
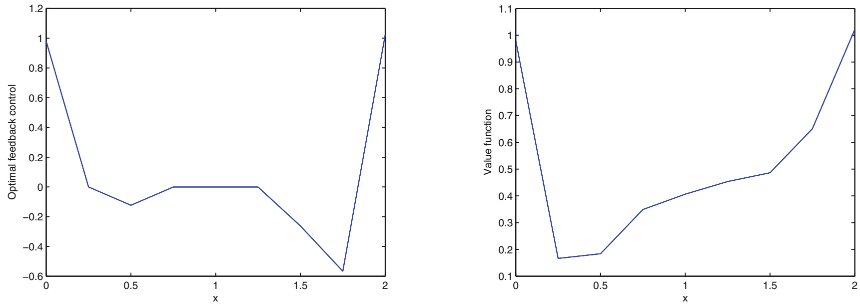


Fig. 2. Mean variance efficient frontier for power management system in which step sizes for the state variable and time variable are $h_1 = 0.25$ and $h_2 = 0.001$, respectively.

data rate is determined by the processing capability limits, the resources used by other tasks of the vehicle's communication system, and the bandwidth allocation scheme between vehicles (e.g., through wireless transmission scheduling). If the platoon is assigned with the total data rate $B(t)$ Mbps (mega bits per second), which must be shared by all the vehicles within the platoon. Let $N_i(t)$ be the portion of $B_i(t)$ that is used in the actual transmission by vehicle i. Then, $N_i(t)B_i(t)$ is the data rate of vehicle i and the total data rate of the entire platoon is desired to be $B(t) = \sum_{i=0}^{4} N_i(t)B_i(t)$. Due to dynamics of many tasks, $B_i(t)$ is a stochastic process. In addition, since vehicles move along roads, we have a communication network whose topology switches. Assume that $\{B_i(t) : \ i = 0, 1, \ldots, 4\}$ obeys the stochastic system (1.1) with the Markov chain $\alpha(t)$ having m states, representing m possible network topologies. To be concrete, suppose that $m = 1, 2, 3, 4$ and the switching process has the generator $Q = \begin{pmatrix} -0.7 & 0.5 & 0.1 & 0.1 \\ 0.4 & -0.8 & 0.2 & 0.2 \\ 0.2 & 0.1 & -0.5 & 0.2 \\ 0.1 & 0.2 & 0.3 & -0.6 \end{pmatrix}$ and $\mu_0(\alpha(t), t) = 0.5\alpha$, $\mu_1(\alpha(t), t) = \alpha + t$, $\mu_2(\alpha(t), t) = 2\alpha + 1.5t$, $\mu_3(\alpha, t) = \alpha - t$, $\sigma_1(\alpha(t), t) = (\alpha, 0, 0, 0)$, $\sigma_2(\alpha(t), t) = (0, \frac{\alpha}{2}, 0, 0)$, $\sigma_3(\alpha(t), t) = (0, 0, \frac{\alpha}{3}, 0)$ for $\alpha = 1, 2, 3, 4$, and $w_1(t) \in \mathbb{R}^4$. Here, the drift term represents average maximum data rates during an operating time interval of the communication system and $B_i(t)\sigma_i(\alpha(t), t)dw_1(t)$ represents fluctuations on B_i, which are determined by other communication tasks such as coding, data compression, packet formation, etc. The dynamics of the process depend on the event variable α which reflects communication network topology changes. Communication link changes typically contain both observable and unobservable elements. It is noted that a communication link can be terminated by the associated vehicles, which is an observable event. However, packet loss can cause a link to be broken which is not observable directly until the data transmission is completed and data were lost. In this sense, this unobservable event can be partially observed from data flows and receipt acknowledgement. Consequently, the event α can be modeled by (1.2) where $g(1) = 2$, $g(2) = 1.5$, $g(3) = 3$ and $g(4) = -1$, and $\beta = 1 > 0$ is a constant. Here $y(t)$ is a measured variable for the event.

Communication system management decides data rate allocation strategies by assigning $N_i(t)$ proportion of data rate to vehicle i, $i = 1, \ldots, 4$. This can be formulated as a mean-variance control problem. To use efficiently the total available data rate $z = 2$ Mbps, we require that at the end of the resource assignment period T, $EB(T) = z$. To ensure that the platoon does not overuse resources (causing interruptions to other platoons, incurring penalty, etc.) or waste resources, it is desirable that the platoon's actual total data rate is as close to 2 Mbps as possible. This is consistent to (1.4), or equivalently (1.5).

The value function and corresponding control are in Fig. 3 in which x axis is the possible value for the resource assignment at $T = 2$ in the platoon communication system and y axis represents π_1- the feedback control or in other words, the data rate of the first vehicle and value function V, respectively. The efficient



(a) Optimal feedback control (data rate) $\pi_1(t)$ for the first vehicle by using the step size $h_1 = 0.25$ for state variable and step size $h_2 = 0.001$ for the time variable with fixed expectation $B = 2$

(b) Approximate value function V by using the step size $h_1 = 0.25$ for state variable and step size $h_2 = 0.001$ for the time variable with fixed expectation $B = 2$

Fig. 3. Optimal control for the first vehicle and value function V for the entire platoon system

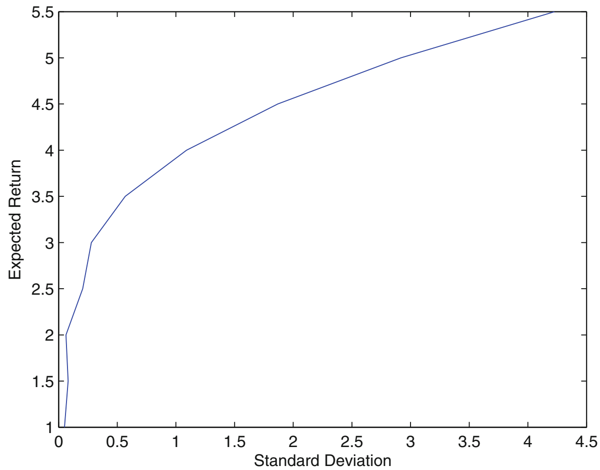


Fig. 4. Mean variance efficient frontier for communication system in which step sizes for the state variable and time variable are $h_1 = 0.25$ and $h_2 = 0.001$, respectively.

frontier is demonstrated in Fig. 4 in which the x axis is the standard deviation of the total data rate of the entire platoon and y axis is the standard deviation of the total data rate allocation for the V2V communications at the end of the resource assignment period.

6 Concluding Remarks

This paper presented case studies on two applications. The main characteristics of the problems are regime-switching diffusions with a hidden Markov chain. Our effort was devoted to the numerical solutions of the problems. After converting the problems into completely observed systems, based on Markov chain approximation techniques, controlled discrete-time Markov chains were constructed for the intended task. Although only two examples have been presented, the techniques used and the methods of approximation can be applied to a wide range of applications.

Acknowledgements. This research was supported in part by the National Science Foundation under CNS-1136007.

References

1. Kushner, H.J.: Consistency issues for numerical methods for variance control with applications to optimization in finance. IEEE. Trans. Automat. Control **44**, 2283–2296 (2000)
2. Markowitz, H.: Portfolio selection. J. Finance **7**, 77–91 (1952)
3. Wonham, W.M.: Some applications of stochastic differential equations to optimal nonlinear filtering. SIAM J. Control **2**, 347–369 (1965)
4. Yang, Z., Yin, G., Wang, L.Y., Zhang, H.: Near-optimal mean-variance controls under two-time-scale formulations and applications. Stochastics **85**, 723–741 (2013)
5. Yang, Z., Yin, G., Zhang, Q.: Mean-variance type controls involving a hidden Markov chain: Models and numerical approximation. IMA J. Control Inf. (to appear)
6. Yin, G., Zhu, C.: Hybrid Switching Diffusions: Properties and Applications. Springer, New York (2010)
7. Zhou, X.Y., Li, D.: Continuous time mean variance portfolio selection: a stochastic LQ framework. Appl. Math. Optim. **42**, 19–33 (2000)
8. Zhou, X.Y., Yin, G.: Markowitz mean-variance portfolio selection with regime switching: a continuous time model. SIAM J. Control Optim. **42**, 1466–1482 (2003)