9

Numerical Methods for Control and Optimization

9.1 Introduction

One of the main techniques presented in Chapters 7 and 8 is to reduce the complexity of singularly perturbed systems by studying the corresponding limit systems that are easier to handle than the original problems. The optimal or nearly optimal controls of the limit problems can be used to construct nearly optimal controls of the original systems. Although the limit systems are substantially simpler than the original pre-limit ones, very often closed-form solutions are still difficult to obtain, except in special cases. For example, in the context of stochastic manufacturing systems, a closed-form solution for optimal production planning is obtained for a system with one-machine and one-part-type by Akella and Kumar [2] for a discounted cost problem, and Zhang and Yin [251] for a finite horizon counterpart. Such closed-form solutions do not seem possible for more general manufacturing systems such as flowshops and jobshops (see Sethi and Zhang [192]). For many applications, one has to resort to a viable alternative – numerical methods.

As a complement to our discussion of singularly perturbed control problems for Markov chains, this chapter focuses on numerical methods for solutions of the control problems. This is a necessary step for many control and optimization problems and alleviates considerably the difficulties encountered. In fact, such a step often plays a crucial role in applications. To take up this issue, we examine the underlying problems from two different

G.G. Yin and Q. Zhang, *Continuous-Time Markov Chains and Applications*, 319
Stochastic Modelling 37, DOI 10.1007/978-1-4614-4346-9_9,

angles, namely, numerical approximation of optimal control problems and stochastic optimization formulation for finding optimal controls under threshold policies.

Treating the optimal control problems, we use the finite difference approximation method developed by Kushner (see Kushner [138], Kushner and Dupuis [141], and the references therein), which has been proven to be very useful for various stochastic systems. Having in our mind a wide variety of applications, we formulate the problem as a nonlinear controlled Markov chain. Our setup is general enough to include, for example, many problems in manufacturing models as special cases. The results obtained are applicable to various dynamical systems and controlled piecewise-deterministic processes.

For various control and optimization problems with long-run average costs, one is often content with a nearly optimal or suboptimal solution. One of the most easily implementable and monitoring strategies in practice is the class of threshold control policies, which provides an enticing alternative. Kimemia and Gershwin brought in the idea of the use of hedging (threshold) policies. Further work along this line may be found in Caramanis and Liberopoulos [24] among others. Under the threshold policy, a control problem can conveniently be transferred to an optimization procedure. The idea is to develop a systematic procedure for finding the optimal threshold values. The essence is to utilize stochastic approximation/optimization methods to resolve the problem. By focusing our attention to the class of threshold controls and considering the expected cost as a function of the threshold levels, we generate a sequence of noisy gradient estimates and update the estimate of the optimal threshold values by use of stochastic recursive algorithms.

The rest of the chapter is arranged as follows. In Section 9.2, we develop a finite difference approximation procedure. Section 9.3 concentrates on the stochastic optimization methods for long-run average cost under threshold policies. Further discussions and citation of related references are in Section 9.4.

9.2 Numerical Methods for Optimal Control

Consider numerical solutions for solving the following control problem:

$$\mathcal{P}: \begin{cases} \text{minimize: } J(x,\alpha,u(\cdot)) = E\displaystyle\int_0^\infty e^{-\rho t} G(x(t),\alpha(t),u(t))dt, \\ \text{subject to: } \dfrac{dx(t)}{dt} = f(x(t),\alpha(t),u(t)), \\ \qquad x(0)=x,\ u(\cdot)\in\mathcal{A},\ \alpha(0)=\alpha, \\ \text{value function: } v(x,\alpha) = \displaystyle\inf_{u(\cdot)\in\mathcal{A}} J(x,\alpha,u(\cdot)), \end{cases}$$

where $\alpha(\cdot)$ is a finite-state Markov chain generated by Q, and $\mathcal{A}$ denotes the set of all admissible controls (i.e., controls that are progressively measurable with respect to $\mathcal{F}(t) = \sigma\{\alpha(s) : \ s \le t\}$ and $u(t) \in \Gamma$, a compact subset of $\mathbb{R}^{n_1}$).

The HJB equation of the control problem $\mathcal{P}$ is

$$\rho v(x,\alpha) = \min_{u\in\Gamma}\Bigg\{ f(x,\alpha,u)\frac{\partial v(x,\alpha)}{\partial x} + G(x,\alpha,u)\Bigg\} + Qv(x,\cdot)(\alpha), \tag{9.1}$$

where as noted in Remark 8.4, $f(x,\alpha,u)(\partial v/\partial x)$ means $\langle f, (\partial v/\partial x)\rangle$, the usual inner product of f and $(\partial v/\partial x)$.

In view of the verification theorem (Theorem A.31), to find an optimal control for the problem, the dynamic programming approach requires a solution to the associated HJB equation. However, more often than not, a closed-form solution of the corresponding HJB equation is not obtainable. Thus, it is necessary to develop numerical algorithms to resolve the problem. In this section, we adopt Kushner's numerical methods for stochastic controls. Our approach consists of using an approximation method for the partial derivatives of the value function $v(x,\alpha)$ within a finite grid of the state vector x and a finite grid for the control vector, which transforms the original optimization problem to an auxiliary discounted Markov decision process. This transformation allows us to apply the well-known techniques, such as a successive approximation or the policy improvement, to solve the HJB equations and then the underlying optimization problems.

Let $\Delta x_i > 0$ denote the length of the finite difference interval of the variables x_i for $i = 1,\dots,n$. Using this finite difference interval, approximate the value function $v(x,\alpha)$ by a sequence of functions $v^{\Delta}(x,\alpha)$ and the partial derivatives $(\partial v(x,\alpha)/\partial x_i)$ by

$$\begin{cases} \dfrac{1}{\Delta x_i}(v^{\Delta}(x(\Delta x_i,+),\alpha) - v^{\Delta}(x,\alpha)), & \text{if } f_i(x,\alpha,u) \ge 0, \\ \dfrac{1}{\Delta x_i}(v^{\Delta}(x,\alpha) - v^{\Delta}(x(\Delta x_i,-),\alpha)), & \text{if } f_i(x,\alpha,u) < 0, \end{cases}$$

where $f(x,\alpha,u) = (f_1(x,\alpha,u),\dots,f_n(x,\alpha,u))'$ and

$$x(\Delta x_i,+) = (x_1,\dots,x_{j-1},x_i+\Delta x_i,x_{j+1},\dots,x_n)',$$

$$x(\Delta x_i,-) = (x_1,\dots,x_{j-1},x_i-\Delta x_i,x_{j+1},\dots,x_n)'.$$

This leads to

$$\begin{aligned} f_i(x,\alpha,u)\frac{\partial}{\partial x_i}v(x,\alpha) \doteq\ & \frac{|f_i(x,\alpha,u)|}{\Delta x_i}v^\Delta(x(\Delta x_i,+),\alpha)I_{\{f_i(x,\alpha,u)\geq 0\}} \\ & +\frac{|f_i(x,\alpha,u)|}{\Delta x_i}v^\Delta(x(\Delta x_i,-),\alpha)I_{\{f_i(x,\alpha,u)<0\}} \\ & -\frac{|f_i(x,\alpha,u)|}{\Delta x_i}v^\Delta(x,\alpha). \end{aligned}$$

With these approximations, we can "rewrite" the HJB equation (9.1) in terms of $v^\Delta(x,\alpha)$ as

$$\begin{aligned} v^\Delta(x,\alpha) =& \min_{u\in\Gamma}\left(\rho+|q_{\alpha\alpha}|+\sum_{i=1}^n\frac{|f_i(x,\alpha,u)|}{\Delta x_i}\right)^{-1} \\ & \times\left\{\sum_{i=1}^n\frac{|f_i(x,\alpha,u)|}{\Delta x_i}\left(v^\Delta(x(\Delta x_i,+),\alpha)I_{\{f_i(x,\alpha,u)\geq 0\}}\right.\right. \\ & \left. +v^\Delta(x(\Delta x_i,-),\alpha)I_{\{f_i(x,\alpha,u)<0\}}\right) \\ & \left.+G(x,\alpha,u)+\sum_{\beta\neq\alpha}q_{\alpha\beta}v^\Delta(x,\beta)\right\}. \end{aligned} \tag{9.2}$$

The theorem below shows that $v^\Delta(x,\alpha)$ converges to $v(x,\alpha)$ as the step size Δx_i goes to zero. For simplicity, we only consider the case that

$$\Delta x_1=\Delta x_2=\cdots=\Delta x_n=\Delta>0.$$

Theorem 9.1. *Assume* (A9.1) *and* (A9.2). *Suppose that* $v^\Delta(x,\alpha)$ *is a solution to* (9.2) *and*

$$0\leq v^\Delta(x,\alpha)\leq K(1+|x|^\kappa),$$

for some constants $K>0$ *and* $\kappa>0$. *Then*

$$\lim_{\Delta\to 0}v^\Delta(x,\alpha)=v(x,\alpha). \tag{9.3}$$

Proof: We only give a brief sketch here; for a detailed account, see Kushner and Dupuis [141]. Note that (9.2) can be written as

$$v^\Delta(x,\alpha)=\mathcal{T}v^\Delta(x,\alpha), \tag{9.4}$$

for an operator $\mathcal{T}$. The problem becomes a fixed point iteration procedure. It is not difficult to check that for each $\Delta>0$, the operator $\mathcal{T}$ is a contraction mapping. The contraction mapping principle then implies that (9.2) has a unique solution $v^\Delta(x,\alpha)$.

To proceed, define a sequence $\{v_k^\Delta(x,\alpha)\}$ as

$$v_0^\Delta(x,\alpha) = 0, \text{ and } v_{k+1}^\Delta(x,\alpha) := \mathcal{T} v_k^\Delta(x,\alpha), \ k \geq 0.$$

Using this sequence, we can show that the solution to (9.4) is continuous.

For any fixed positive Δ ($= \Delta x_i$, for $i = 1, \ldots, n$) and $\alpha \in \mathcal{M}$,

$$0 < \rho \leq \rho + |q_{\alpha\alpha}| + \sum_{i=1}^{n} \frac{|f_i(x,\alpha,u)|}{\Delta x_i} \leq K_1(1 + \Delta^{-1}),$$

for some constant K_1. As a result, (9.2) is equivalent to

$$\begin{aligned} 0 \ = \min_{u\in\Gamma}\Bigg\{ & \sum_{i=1}^{n} \frac{|f_i(x,\alpha,u)|}{\Delta x_i}\Bigg([v^\Delta(x(\Delta x_i,+),\alpha) - v^\Delta(x,\alpha)] I_{\{f_i(x,\alpha,u)\geq 0\}} \\ & + [v^\Delta(x(\Delta x_i,-),\alpha) - v^\Delta(x,\alpha)] I_{\{f_i(x,\alpha,u)<0\}} \Bigg) \\ & + G(x,\alpha,u) + \sum_{\beta\neq\alpha} q_{\alpha\beta}[v^\Delta(x,\beta) - v^\Delta(x,\alpha)] - \rho v^\Delta(x,\alpha) \Bigg\}. \end{aligned}$$

For each $x \in \mathbb{R}^n$ and $\alpha \in \mathcal{M}$, let

$$v^*(x,\alpha) := \limsup_{\delta\to 0}\Bigg(\limsup_{\Delta\to 0} \left[\sup\{v^\Delta(\tilde{x},\alpha) : |x - \tilde{x}| \leq \delta\}\right] \Bigg)$$

and

$$v_*(x,\alpha) := \liminf_{\delta\to 0}\Bigg(\liminf_{\Delta\to 0} \left[\inf\{v^\Delta(\tilde{x},\alpha) : |x - \tilde{x}| \leq \delta\}\right] \Bigg).$$

It is clear that $v^*(x,\alpha) \geq v_*(x,\alpha)$. Moreover, it can be shown that $v^*(x,\alpha)$ is upper semicontinuous and $v_*(x,\alpha)$ is lower semicontinuous.

To obtain the convergence result, it remains to derive the reverse inequality, $v^*(x,\alpha) \leq v_*(x,\alpha)$. In fact, we need only show that $v^*(x,\alpha)$ and $v_*(x,\alpha)$ are viscosity subsolution and viscosity supersolution to (9.1), respectively. This can be done as in Kushner and Dupuis [141, Theorem 14.3.1]. Consequently, by virtue of the uniqueness of the viscosity solution to the HJB equation (see Theorem A.24), $v^*(x,\alpha) \leq v_*(x,\alpha)$. Hence,

$$v^*(x,\alpha) = v_*(x,\alpha) = v(x,\alpha)$$

as desired. □

Remark 9.2. To obtain an optimal control via the dynamic programming approach, one needs to use the corresponding value function as in the verification theorem (see Theorem A.31). Usually, the numerical scheme produces only an approximate value function, which can be regarded as a perturbation of the true value function. The rationale is that by using the

approximate value function in the verification theorem, one can construct a feedback control policy that is approximately optimal. In fact, under fairly mild conditions and using a viscosity solution approach, Yan and Zhang [221] have shown that the control policy obtained using an approximate value function is indeed nearly optimal as the perturbations go to 0.

Remark 9.3. In view of the discussion above, there is nothing so special about the problem $\mathcal{P}$. The same approach can equally be applied to the singularly perturbed problem $\mathcal{P}^\varepsilon$ defined in Chapter 8. However, following our previous consideration, for a large and complex system, one would be better off to obtain a "reduced-order" system (limit system) first and to apply the numerical method only to the limit problem. The proof of Theorem 9.1 uses viscosity solution techniques. An alternative approach is to apply the method of weak convergence via Markov chain approximation techniques as in the setup of Kushner [138] or Kushner and Dupuis [141].

9.3 Optimization under Threshold Policy

This section consists of several subsections. First an optimal control problem is reformulated as a stochastic optimization problem. The next subsection gives the convergence proof of the recursive algorithm followed by a couple of examples in production planning with unreliable machines. The last subsection derives the estimation error for the approximation.

9.3.1 Stochastic Optimization Formulation

As in the previous section, suppose that $\alpha(\cdot)$ is a finite-state Markov chain with stationary transition probability or, equivalently, the generator $Q(\cdot) = Q$, a constant matrix. Let $x(t) \in \mathbb{R}^n$, $u(t) \in \Gamma$, a compact subset of $\mathbb{R}^{n_1}$, $f(\cdot,\cdot,\cdot) : \mathbb{R}^n \times \Gamma \times \mathcal{M} \mapsto \mathbb{R}^n$, and $G(\cdot,\cdot,\cdot) : \mathbb{R}^n \times \Gamma \times \mathcal{M} \mapsto \mathbb{R}$. Consider the following controlled dynamic system

$$\frac{dx(t)}{dt} = f(x(t), \alpha(t), u(t)), \;\; x(0) = x^0, \tag{9.5}$$

with a long-run average cost function

$$J(u) = \lim_{T_1 \to \infty} \frac{1}{T_1} E \int_0^{T_1} G(x(t), \alpha(t), u(t)) dt. \tag{9.6}$$

Instead of seeking optimal controls of the system given above, we reformulate it as a stochastic optimization problem. The main idea lies in concentrating on a class of controls of the threshold type. Under such a setting, our effort is to develop an easily implementable algorithm to approximate the optimal threshold levels. Here and hereafter, the terms threshold values and threshold levels will be used interchangeably.

Definition 9.4. A control policy $u(t) \in \Gamma$ is of threshold type with constant threshold levels if there are sets $A_i \subset \mathbb{R}^n$ and constants $c_i \in \Gamma \subset \mathbb{R}^{n_1}$ for $i = 1, \ldots, n_0$ such that for some integer n_0,

$$u(t) = \sum_{i=1}^{n_0} c_i I_{\{x(t) \in A_i\}}.$$

Typically, the sets A_i depend on some parameter $\theta \in \mathbb{R}^n$. To illustrate, consider the following example. This is an analytically solvable case, and describes the salient features of the threshold type of control policies.

Example 9.5. Consider a failure-prone manufacturing system with production capacity $\alpha(\cdot)$, that is a Markov chain with finite-state space $\mathcal{M}$. For simplicity, assume $\mathcal{M} = \{\alpha_1, \alpha_2\}$, where α_1 means the machine is up and α_2 means that the machine is down. Suppose that the breakdown and repair times are independent and exponentially distributed with parameters λ and μ, respectively. Denote the inventory level and the production rate of the system by $x(t)$, $u(t) \in \mathbb{R}$, respectively. For convenience, let $\alpha_1 = 1$ and $\alpha_2 = 0$. Then, the production constraints are given as

$$0 \le u(t) \le u_{\max}\alpha(t), \ t \ge 0,$$

where $u_{\max}$ is the maximum production rate (since $\alpha(t) = 0$ or 1, $u_{\max}$ is also the maximum capacity) of the machine. Our objective is to find the optimal control $u(\cdot)$ to

$$\begin{cases} \text{minimize: } J(u) = \lim\limits_{T_1 \to \infty} \dfrac{1}{T_1} E \displaystyle\int_0^{T_1} [c^+ x^+(t) + c^- x^-(t)]dt, \\ \text{subject to: } \dfrac{dx(t)}{dt} = u(t) - z, \ \ x(0) = x_0, \end{cases}$$

where z is a constant demand rate, $x^+ = \max\{0, x\}$ and $x^- = \max\{0, -x\}$, and c^+ and c^- are nonnegative constants. By means of dynamical programming equation approach, Bielecki and Kumar [12] derived the optimal control explicitly, and showed that the optimal control is of threshold type given by

$$u^*(t) = \begin{cases} u_{\max} I_{\{\alpha(t)=1\}}, & \text{if } x(t) < \theta^*, \\ z I_{\{\alpha(t)=1\}}, & \text{if } x(t) = \theta^*, \\ 0, & \text{if } x(t) > \theta^*, \end{cases}$$

where θ^* is the optimal threshold value given by

$$\theta^* = \begin{cases} 0, \quad \text{if} \quad \dfrac{u_{\max}\lambda(c^+ + c^-)}{c^+(u_{\max} - z)(\lambda + \mu)} \le 1 \ \text{ and } \ \dfrac{u_{\max} - z}{\lambda} > \dfrac{z}{\mu}, \\ \infty, \quad \text{if} \quad \dfrac{u_{\max} - z}{\lambda} \le \dfrac{z}{\mu}, \\ \dfrac{z(u_{\max} - z)}{\mu(u_{\max} - z) - \lambda z} \log\left(\dfrac{u_{\max}\lambda(c^+ + c^-)}{c^+(u_{\max} - z)(\lambda + \mu)}\right), \quad \text{otherwise.} \end{cases}$$

Moreover, the optimal cost is

$$J(u^*) = \begin{cases} \dfrac{c^- \lambda u_{\max}}{(\lambda+\mu)(\mu u_{\max} - \lambda z - \mu z)}, & \text{if } \theta^* = 0, \\ \dfrac{c^+ z}{\lambda+\mu} + \dfrac{c^+ z(u_{\max} - z)}{\mu(u_{\max} - z) - \lambda z} \log\left(\dfrac{u_{\max}\lambda(c^+ + c^-)}{c^+(u_{\max} - z)(\lambda+\mu)}\right), & \\ & \text{if } \theta^* > 0. \end{cases}$$

Since the optimal control depends on the threshold parameter θ, the expected cost can also be viewed as a function of the threshold. Intuitively, the optimal policy can be described as follows. If the inventory level is below the optimal threshold, one should produce at a full speed; if the inventory level is above the threshold, one should produce nothing; if the inventory level is at the threshold, one should produce exactly the same as the demand.

In view of (9.6), focusing our attention to the class of controls of threshold type, the cost $J(\cdot)$ becomes a function of the threshold levels (i.e., $J = J(\theta)$). Threshold types of control policies have drawn renewed attention lately, since the idea is appealing and the principle is easy to apply. First such policies are fairly simple in form and easily implementable so that they are particularly attractive in applications. Once a threshold value is determined, a controller or an operator can ignore detailed variations and concentrate only on adjusting controls according to the threshold levels. The corresponding control procedure is simpler as compared with the optimal control policies, since only a monitoring device/procedure is needed to keep track of the performance of the underlying system. In lieu of solving the HJB equations, only a few parameters need to be tuned. Moreover, in various situations, one is often content with suboptimality owing to the cost consideration and other limitations. Frequently, a suboptimal control is nearly or virtually as valuable as an optimal control. Furthermore, in many cases, threshold control policies are indeed optimal as in Example 9.5.

Upon transferring the problem to an optimization task, the foremost important task is to locate the optimal threshold values. This dictates the development of stochastic recursive algorithms. Our aim is to develop a systematic approach to approximate the threshold values.

Throughout the rest of the chapter, $\theta \in \mathbb{R}^n$ denotes a column vector. Denote $\xi(t,\theta) = (x(t), \alpha(t))$, and use $\eta > 0$, a small parameter, to represent the step size. The stochastic optimization algorithm takes the form

$$\begin{aligned} \theta_{k+1} &= \theta_k - \eta\,(\text{gradient estimate of } J(\cdot) \text{ at } \theta_k) \\ &= \theta_k - \frac{\eta}{T}\int_{kT}^{(k+1)T} g(\theta_k, \xi(t,\theta_k))dt, \end{aligned} \tag{9.7}$$

for each integer $k \geq 0$, where $g(\cdot)$ is an appropriate function. Example 9.11 and Example 9.12 in what follows, present two examples of such gradient estimates.

For fixed $T > 0$, the gradient estimate of $J(\theta)$ at the kth iterate is of the form

$$\frac{1}{T}\int_{kT}^{(k+1)T} g(\theta_k, \xi(t, \theta_k))dt.$$

In what follows, for notational convenience, we often suppress the θ dependence and write $\xi(t, \theta)$ as $\xi(t)$. Although the gradient estimate of $J(\cdot)$ in (9.7) can be obtained via finite difference approximation in a straightforward way, various alternatives exist. The infinitesimal perturbation analysis (IPA) approach (see Ho and Cao [87] and Glasserman [74]) provides a better alternative, however. While it is more efficient, this approach is application dependent. That is, one needs to figure out the gradient estimate for each application; there are no general forms of the gradient estimates available. We use a constant step size since an iterative algorithm with constant step size has the ability to track slight variation of the parameter and is more robust with respect to the random errors.

Using the IPA approach, for Example 9.5, the gradient estimate takes the form

$$\frac{1}{T}\int_0^T g(\theta_k, \xi(t, \theta_k))dt = \frac{1}{T}\int_0^T \Big(c^+ I_{\{x(t)>0\}} - c^- I_{\{x(t)<0\}}\Big)dt.$$

Example 9.12 gives an illustration for a two-machine system.

Remark 9.6. Equation (9.7) is not a standard stochastic approximation algorithm since averaging is used in the scheme together with continuous time random processes. In Yin, Yan, and Lou [228], with the goal of obtaining an asymptotically unbiased gradient estimator, T is chosen so that $T = T_\eta \to \infty$ as $\eta \to 0$. However, as noted in Kushner and Vázquez-Abad [143], and Kushner and Yin [145], one need not choose T so large. To guarantee the convergence of the algorithm, it is not necessary to use unbiased (or asymptotically unbiased) estimators of the gradient. In fact, large T may result in inefficient performance of the algorithms. A little bias would not and should not concern us.

9.3.2 Convergence

This subsection is devoted to investigating the convergence of the proposed algorithms. To proceed, the following assumptions are needed. For simplicity, assume the initial approximation θ_0 to be nonrandom.

(**A9.1**) For each θ and each k_1,

$$\frac{1}{kT}\sum_{j=k_1}^{k+k_1-1} E^{\mathcal{F}_{k_1}} \int_{jT}^{(j+1)T} g(\theta, \xi(t))dt \to \nabla J(\theta)$$

in probability, as $k \to \infty$, where $E^{\mathcal{F}_{k_1}}$ denotes the conditional expectation on $\mathcal{F}_{k_1 T} = \{\xi(s),\ s \le k_1 T\}$, for an integer $k_1 \ge 0$.

(**A9.2**) For each $T_1 < \infty$, $t \in [0, T_1]$,

$$\lim_{\delta \to 0} E\left(\sup_{|\theta - \tilde{\theta}| < \delta} |g(\theta, \xi(t)) - g(\tilde{\theta}, \xi(t))| \right) = 0.$$

(**A9.3**) For each $N < \infty$, the set of functions

$$\left\{ \sup_{|\theta| \le N} |g(\theta, \xi(t))| \right\} \text{ is uniformly integrable.}$$

Remark 9.7. These assumptions originate from particular applications of manufacturing models. Dealing with specific applications, these conditions can often be verified (see Yan, Yin, and Lou [220]). Condition (A9.1) is an ergodicity condition in the sense of convergence in probability, and is a basic averaging condition. If $\xi(\cdot)$ is a ϕ-mixing process with $E|\xi(t)| < \infty$, then it is a strongly ergodic process and hence (A9.1) holds. In fact, in this case, the convergence is in the sense of with probability one.

Condition (A9.2) indicates that the function $g(\cdot, \xi)$ may not be continuous, but its expectation is continuous such as for the case that $g(\cdot, \xi)$ is an indicator function or a combination of indicator functions.

In various applications, the function $g(\theta, \xi)$ is often bounded. In such a case, (A9.3) is verified. Condition (A9.3) allows us to deal with more complex situations. For example, if

$$|g(\theta, \xi)| \le h_0(\theta)\widetilde{g}_1(\xi) + \widetilde{g}_2(\xi),$$

where $h_0(\theta)$ is a continuous function, and $E|\tilde{g}_i(\xi)|^{1+\gamma} < \infty$, $i = 1, 2$, for some $\gamma > 0$, then condition (A9.3) is also satisfied.

To proceed, we work with continuous time interpolated processes. Let $\theta^\eta(\cdot)$ be defined by $\theta^\eta(0) = \theta_0$ and $\theta^\eta(t) = \theta_k$ for $t \in [k\eta, (k+1)\eta)$. Under the framework of weak convergence (see Kushner [139], and Kushner and Yin [145]), it will be shown that the following limit theorem holds.

Theorem 9.8. *Suppose that* (A9.1)–(A9.3) *are satisfied and the differential equation*

$$\frac{d\theta(t)}{dt} = -\nabla J(\theta) \tag{9.8}$$

has a unique solution for each initial condition θ_0. Assume, for simplicity, that $\theta^\eta(0) = \theta_0$ is independent of η. Then $\{\theta^\eta(t)\}$ is tight in $D([0, \infty); \mathbb{R}^n)$. Every weakly convergent subsequence has the same limit $\theta(\cdot)$ that satisfies the differential equation (9.8).

Remark 9.9. Recall that $D([0,\infty);\mathbb{R}^n)$ denotes the space of $\mathbb{R}^n$-valued functions that are right continuous and have left-hand limits, endowed with the Skorohod topology; see Section A.2 in Appendix A. In lieu of choosing $\theta^\eta(0) = \theta_0$, independent of η, one may use $\theta_0 = \theta_0^\eta$ and hence $\theta^\eta(0) = \theta_0^\eta$ (depending on η). Under an additional condition $\theta_0^\eta \Rightarrow \theta_0$, the result still holds.

Proof of Theorem 9.8: To avoid possible unboundedness, a truncation device will be used (see (A.8) in Appendix for a definition). For each $N < \infty$, let $\theta^{\eta,N}(\cdot)$ be the N-truncation of $\theta^\eta(\cdot)$ such that $\theta^{\eta,N}(t) = \theta^\eta(t)$ up until the first exit from the N-sphere $S_N = \{\theta; |\theta| \le N\}$. A pertinent use of the truncation device requires the use of a truncation function $q_N(\cdot)$, which is a smooth function defined as

$$q_N(\theta) = \begin{cases} 1 & \text{for } |\theta| \le N, \\ 0 & \text{for } |\theta| \ge N+1. \end{cases}$$

One then replace $g(\theta,\xi)$ below by $g_N(\theta,\xi) = g(\theta,\xi)q_N(\theta)$. For notational simplicity, we shall omit the truncation function in what follows, however.

In view of the definition of the interpolation (without loss of generality, assume that t/η and $(t+s)/\eta$ are integers) and choosing a sequence of integers $\{k_\eta\}$ such that $k_\eta \to \infty$ as $\eta \to 0$ and $\eta k_\eta = \delta_\eta \to 0$, we have

$$\begin{aligned} \theta^{\eta,N}(t) &= \theta^{\eta,N}(0) - \frac{\eta}{T}\sum_{j=0}^{t/\eta-1}\int_{jT}^{(j+1)T} g(\theta_j^{\eta,N},\xi(\upsilon))d\upsilon \\ &= \theta^{\eta,N}(0) - \sum_{0\le i\delta_\eta\le t}\frac{\delta_\eta}{k_\eta T} \\ &\qquad \times \sum_{ik_\eta\le j\le (i+1)k_\eta-1}\int_{jT}^{(j+1)T} g(\theta_j^{\eta,N},\xi(\upsilon))d\upsilon \\ &= \theta^{\eta,N}(0) - \int_0^t B^\eta(\tau)d\tau, \end{aligned} \tag{9.9}$$

where $B^\eta(\cdot)$ is a piecewise-constant function on $[i\delta_\eta,(i+1)\delta_\eta)$, that is,

$$B^\eta(t) = \frac{1}{k_\eta T}\sum_{ik_\eta\le j\le (i+1)k_\eta-1}\int_{jT}^{(j+1)T} g(\theta_j^{\eta,N},\xi(\upsilon))d\upsilon \tag{9.10}$$

for $t \in [i\delta_\eta,(i+1)\delta_\eta)$. It follows from (9.9) that

$$\frac{d\theta^{\eta,N}(t)}{dt} = -B^\eta(t).$$

Condition (A9.3) implies that

$$\left\{\frac{1}{T}\int_{jT}^{(j+1)T} g(\theta_j^N,\xi(\upsilon))d\upsilon : \ j = 1,2,\dots\right\} \text{ is uniformly integrable.}$$

Then by virtue of Theorem A.15, $\{\theta^{\eta,N}(\cdot), B^{\eta}(\cdot)\}$ is tight and the limit of any weakly convergent subsequence has continuous paths with probability one.

Pick out an arbitrary convergent subsequence and denote the limit by $(\theta^N(\cdot), \overline{B}(\cdot))$. By the Skorohod representation (without changing notation), we may assume that

$$(\theta^{\eta,N}(\cdot), B^{\eta}(\cdot)) \to (\theta^N(\cdot), \overline{B}(\cdot)) \text{ w.p.1}$$

and the convergence is uniform on any finite time interval.

Define

$$M^N(t) = \theta^N(t) - \theta^N(0) + \int_0^t \overline{B}(\theta^N(\upsilon))d\upsilon. \tag{9.11}$$

It will be seen in what follows that $\overline{B}(\cdot)$ is equal to $\nabla J(\cdot)$. If we can show that $M^N(t)$ is a continuous martingale, the limit theorem will hold for the truncated process. Note that $M^N(0) = 0$ and $M^N(t)$ is Lipschitz continuous. If it is a martingale, it must satisfy $M^N(t) \equiv 0$ (see Theorem A.21). Therefore, we need only verify the martingale property.

To verify the martingale property, let $h(\cdot)$ be any bounded and continuous function, κ be any positive integer, and t_{i_1} be such that $t_{i_1} < t < t + s$ for $i_1 \le \kappa$. In view of the weak convergence and the Skorohod representation, we have

$$\begin{aligned} &\lim_{\eta\to 0} Eh(\theta^{\eta,N}(t_{i_1}), i_1 \le \kappa)\left(\theta^{\eta,N}(t+s) - \theta^{\eta,N}(t)\right) \\ &\quad = Eh(\theta^N(t_{i_1}), i_1 \le \kappa)\left(\theta^N(t+s) - \theta^N(t)\right). \end{aligned} \tag{9.12}$$

Recall that $E^{\mathcal{F}_j}$ denotes the conditional expectation with respect to the σ-algebra $\mathcal{F}_{jT} = \sigma\{\xi(t), t \le jT\}$. Using the recursion (9.9), we have

$$\begin{aligned} &\lim_{\eta\to 0} Eh(\theta^{\eta,N}(t_{i_1}), i_1 \le \kappa)\left(\theta^{\eta,N}(t+s) - \theta^{\eta,N}(t)\right) \\ &\quad = \lim_{\eta\to 0} Eh(\theta^{\eta,N}(t_{i_1}), i_1 \le \kappa) \\ &\qquad \times\left(-\sum_{i\delta_\eta = t}^{t+s} \frac{\delta_\eta}{k_\eta T} \sum_{ik_\eta \le j \le (i+1)k_\eta - 1} E^{\mathcal{F}_j} \int_{jT}^{(j+1)T} g(\theta_j^{\eta,N}, \xi(\upsilon))d\upsilon\right) \\ &\quad = \lim_{\eta\to 0} Eh(\theta^{\eta,N}(t_{i_1}), i_1 \le \kappa) \\ &\qquad \times\left(-\sum_{i\delta_\eta = t}^{t+s} \frac{\delta_\eta}{k_\eta T} \sum_{ik_\eta \le j \le (i+1)k_\eta - 1} E^{\mathcal{F}_j} \int_{jT}^{(j+1)T} g(\theta^{\eta,N}(\tau), \xi(\upsilon))d\upsilon\right). \end{aligned} \tag{9.13}$$

The last equality above follows from the weak convergence, the Skorohod representation, (A9.2), and $\eta j \to \tau$ as $\eta \to 0$ for j satisfying $ik_\eta \le j \le (i+1)k_\eta$.

Now for any $\Delta > 0$, there exists a function $\theta^{N,\Delta}(\cdot)$ that takes only finitely many values (say $\overline{\theta}_1, \ldots, \overline{\theta}_{n_0}$) such that

$$|\theta^N(\tau) - \theta^{N,\Delta}(\tau)| < \Delta.$$

Consequently, by applying (A9.2), the limit in (9.13) is the same as that of

$$\lim_{\eta\to 0} Eh(\theta^{\eta,N}(t_{i_1}), i_1 \le \kappa) \times \left(-\sum_{i\delta_\eta = t}^{t+s} \frac{\delta_\eta}{k_\eta T} \sum_{ik_\eta \le j \le (i+1)k_\eta - 1} E^{\mathcal{F}_j} \int_{jT}^{(j+1)T} g(\theta^{N,\Delta}(\tau), \xi(\upsilon)) d\upsilon \right).$$

By virtue of (9.10), the limit of $B^\eta(\tau)$ is the same as that of

$$\begin{aligned} &\frac{1}{k_\eta T} \sum_{ik_\eta \le j \le (i+1)k_\eta - 1} E^{\mathcal{F}_j} \int_{jT}^{(j+1)T} g(\theta^{N,\Delta}(\tau), \xi(\upsilon)) d\upsilon \\ &= \sum_{i_2=1}^{n_0} \frac{1}{k_\eta T} \sum_{ik_\eta \le j \le (i+1)k_\eta - 1} E^{\mathcal{F}_j} \int_{jT}^{(j+1)T} g(\overline{\theta}_{i_2}, \xi(\upsilon)) d\upsilon \, I_{\{\theta^{N,\Delta}(\tau) = \overline{\theta}_{i_2}\}} \\ &\to \sum_{i_2=1}^{n_0} \nabla J(\overline{\theta}_{i_2}) I_{\{\theta^{N,\Delta}(\tau) = \overline{\theta}_{i_2}\}} \quad \text{in probability} \\ &= \nabla J(\theta^{N,\Delta}(\tau)). \end{aligned}$$

Since $\Delta > 0$ is arbitrary,

$$\frac{1}{k_\eta T} \sum_{ik_\eta \le j \le (i+1)k_\eta - 1} E^{\mathcal{F}_j} \int_{jT}^{(j+1)T} g(\theta^{N,\Delta}(\tau), \xi(\upsilon)) d\upsilon \to \nabla J(\theta^N(\tau))$$

in probability as $\eta \to 0$. Incorporating this with (9.12) and (9.13) yields

$$\begin{aligned} &\lim_{\eta\to 0} Eh(\theta^{\eta,N}(t_{i_1}), i_1 \le \kappa) \left(\theta^{\eta,N}(t+s) - \theta^{\eta,N}(t)\right) \\ &\quad = Eh(\theta^N(t_{i_1}), i_1 \le \kappa) \left(\theta^N(t+s) - \theta^N(t) + \int_t^{t+s} \nabla J(\theta^N(\tau)) d\tau \right). \end{aligned} \tag{9.14}$$

Combining (9.12) to (9.14), we arrive at

$$Eh(\theta^N(t_i), i \le \kappa) \left(\theta^N(t+s) - \theta^N(t) + \int_t^{t+s} \nabla J(\theta^N(\tau)) d\tau \right) = 0.$$

Hence $M^N(t)$ is a martingale.

Finally, use the idea of Kushner [139, Theorem 2.2 and the Corollary], to finish the proof. The main idea is outlined below. Let $P_{\theta(0)}(\cdot)$ (the subscript $\theta(0)$ signifies the dependence on the initial data) and $P^N(\cdot)$ be the measures induced by $\theta(\cdot)$ and $\theta^N(\cdot)$, respectively, on $\mathcal{B}$, the σ-algebra of Borel subsets of $D([0,\infty);\mathbb{R}^n)$. $P_{\theta(0)}(\cdot)$ is unique since there is a unique solution to the ordinary differential equation for the initial value $\theta(0)$. Thus, for each $T_1<\infty$,

$$P_{\theta(0)}(\theta(\cdot)\in A) = P^N(\theta^N(\cdot)\in A)$$

for each $A\in\mathcal{B}$ such that $\theta(t)$ takes values in S_N (the N-sphere). As a result,

$$\lim_{N\to\infty} P_{\theta(0)}\left(\sup_{t\le T_1}|\theta(t)|\le N\right)=1.$$

This, together with the weak convergence of $\theta^{\eta,N}(\cdot)$, implies that $\theta^\eta(\cdot)\Rightarrow\theta(\cdot)$. Since the limit is unique, it does not depend on the chosen subsequence. The proof of the theorem is completed. □

Theorem 9.8 is similar to the law of large numbers. It gives information on the location and/or distribution of $\theta^\eta(\cdot)$ for small η and for large but bounded t. There is a natural connection between the recursive procedure and the corresponding ordinary differential equation. The optimal threshold values sought are stable points of the differential equation (9.8).

Theorem 9.10. *Assume that the conditions of Theorem 9.8 hold. Suppose the differential equation in (9.8) has a unique asymptotically stable point θ^* (in the sense of Liapunov stability) and the set*

$$\{\theta_k;\ k<\infty,\eta>0\} \tag{9.15}$$

is bounded in probability in that for each $\Delta>0$, there is a $\kappa_\Delta>0$ such that for all $\eta>0$, and all k,

$$P(|\theta_k|\ge\kappa_\Delta)\le\Delta.$$

Let $t_\eta\to\infty$ as $\eta\to0$. Then $\theta^\eta(t_\eta+\cdot)$ is tight in $D([0,\infty;\mathbb{R}^n)$ and any weak limit is equal to θ^*.

Equation (9.15) can be established by using a perturbed Liapunov function method (see Kushner [139], and Kushner and Yin [145]). Theorem 9.10 can be deduced analogously as in the aforementioned reference (see also Kushner and Yin [144, Theorem 5.1]). We give the main idea below. Let $T_1>0$, and consider the pair $\{\theta^\eta(t_\eta+\cdot),\theta^\eta(t_\eta-T_1+\cdot)\}$, which is tight. Choose a weakly convergent subsequence (still indexed by η) with limit denoted by $(\theta(\cdot),\theta_{T_1}(\cdot))$. Then $\theta(0)=\theta_{T_1}(T_1)$. The "initial value" $\theta_{T_1}(0)$ may not be known, but all possible values of $\theta_{T_1}(0)$ belong to a set that is bounded in probability for all T_1 and all convergent subsequences.

The asymptotic stability then implies that for each $\Delta > 0$, there is a $T_\Delta > 0$ such that for all $T_1 > T_\Delta$,

$$P(|\theta_{T_1}(T_1) - \theta^*| > \Delta) < \Delta.$$

Hence the theorem follows.

9.3.3 Examples

To illustrate the idea of approximation of threshold control policies, we consider two manufacturing models in this subsection. The reasons include: (a) Demonstrate that for the example treated below, threshold control policies are indeed optimal. (b) Illustrate the use of stochastic optimization procedure for the long-run average cost criteria.

In the first example, our approximation results compare well with those of Bielecki and Kumar [12]. The second example deals with a two-machine system, in which no closed-form solution (analytic solution or explicit formula) has been found up to date. To reformulate the problem using optimization formulation, we develop stochastic algorithms to estimate the optimal threshold values.

Example 9.11. Return to Example 9.5. Choose $\lambda = 0.1$, $\mu = 0.125$, $z = 1.0$, $c^+ = 2.0$, $c^- = 9.0$, and $u_{\max} = 2.0$. Applying the result of [12], the optimal threshold level and the optimal cost are $\theta^* = 66.96$ and $J(\theta^*) = 142.89$, respectively. Using our algorithm with step size $\eta = 0.5$ and initial value $\theta_0 = 20$, and taking averages of 100 replications, the approximation method gives $\tilde{\theta}^* = 67.23$ (with a 95% confidence interval $[66.64, 67.80]$), and $J(\tilde{\theta}^*) = 139.43$. Different initial conditions yield equally good approximation results. The already existing analytical result allows us to compare the performance of the approximation algorithm with the closed-form solution. We would like to note that even if the explicit solution is available, various parameters (λ, μ, etc.) may not be known; these parameters are not required in our approach. Thus the stochastic optimization approach provides a viable alternative and effective procedure.

Example 9.12. The example to be presented was considered by Yan, Yin, and Lou [220], in which a combination of infinitesimal perturbation analysis initiated by Ho (see Ho and Cao [87]) and stochastic approximation was suggested. Kushner and Vázquez-Abad [143] further examined this model and relaxed the conditions for convergence. For $i = 1, 2$, use $x_i(t)$ to denote the inventory levels of machine i, and $u_i(t)$ the production rate of machine i. Since we are not solving the dynamic programming equations, the demand processes can be quite general. They do not have to be constants although a constant demand rate is used here for simplicity. In what follows, we formulate the surplus control model, and construct the approximation procedure.

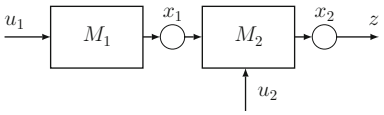

FIGURE 9.1. A Two-Machine System

The two machines are in a cascade form and the inventory levels are given by

$$\begin{aligned} \frac{dx_1(t)}{dt} &= u_1(t) - u_2(t), \\ \frac{dx_2(t)}{dt} &= u_2(t) - z, \\ x_1(t) &\geq 0,\ t \geq 0. \end{aligned} \tag{9.16}$$

For each $i = 1, 2$, let the machine capacity be $\alpha_i(t)$ with

$$\alpha_i(t) = \begin{cases} 1, & \text{machine } i \text{ is working;} \\ 0, & \text{otherwise.} \end{cases}$$

We then have

$$0 \leq u_i(t) \leq u_{i\max}\alpha_i(t),\ i = 1, 2,$$

where $u_{i\max}$ is the maximum capacity of machine i for $i = 1, 2$. Assume that

$$u_{1\max} > u_{2\max} > z.$$

This scenario is depicted in Figure 9.1.

Surplus at machine i is defined as the difference between accumulative production and accumulated demand, i.e., it is the inventory level (or work in progress) at machine i plus the inventory level of all downstream machines. Let $s_i(t)$ be the surplus for machine i, for $i = 1, 2$:

$$s_1(t) = x_1(t) + x_2(t) \text{ and } s_2(t) = x_2(t).$$

Note that the surplus can be positive or negative. A negative surplus means that there is a backlog. With these definitions, the system dynamics can also be written as

$$\begin{aligned} \frac{ds_1(t)}{dt} &= u_1(t) - z, \\ \frac{ds_2(t)}{dt} &= u_2(t) - z, \\ s_1(t) &\geq s_2(t),\ t \geq 0. \end{aligned} \tag{9.17}$$

Comparing (9.17) with the one-machine model in the work of Akella and Kumar [2], and Bielecki and Kumar [12], the surplus control policy is more or less like having two machines operating independently.

Let θ_i denote the surplus threshold levels of machine i. The control policy is given by

$$u_1(t) = \begin{cases} u_{1\max} I_{\{\alpha_1(t)=1\}}, & \text{if } s_1(t) < \theta_1, \\ z I_{\{\alpha_1(t)=1\}}, & \text{if } s_1(t) = \theta_1, \\ 0, & \text{if } s_1(t) > \theta_1; \end{cases}$$

$$u_2(t) = \begin{cases} u_{2\max} I_{\{\alpha_2(t)=1\}}, & \text{if } s_2(t) < \theta_2,\ s_1(t) - s_2(t) > 0, \\ z I_{\{\alpha_2(t)=1\}}, & \text{if } s_2(t) = \theta_2,\ s_1(t) - s_2(t) > 0, \\ 0, & \text{if } s_2(t) > \theta_2,\ s_1(t) - s_2(t) > 0. \end{cases}$$

The interpretation of the control policies is similar to that of the one machine case. The problem to be investigated is to find the optimal threshold value $\theta^* = (\theta_1^*, \theta_2^*)$ such that the cost functional

$$J(\theta) = \lim_{T\to\infty} \frac{1}{T} E \int_0^T (c_1 x_1(t) + c_2^+ x_2^+(t) + c_2^- x_2^-(t))dt \tag{9.18}$$

is minimized.

Define

$$\tau_0^1(\theta) = 0,\ \tau_0^2(\theta) = \inf\{t > 0;\ s_1(t,\theta) = \theta_1\},$$

and $\tau_k^i(\theta)$ for $i = 1, 2$ and $k > 0$, recursively by

$$\tau_k^1(\theta) = \min\{t \geq \tau_{k-1}^2(\theta); s_1(t,\theta) = s_2(t,\theta)\} \quad \text{and}$$

$$\tau_k^2(\theta) = \min\{t \geq \tau_k^1(\theta); s_2(t,\theta) = \theta_2\}.$$

Moreover, define $\gamma_0^2(\theta) = 0$ and

$$\gamma_k^1(\theta) = \min\{t \geq \gamma_{k-1}^2(\theta);\ s_2(t,\theta) = \theta_2\},$$

$$\gamma_k^2(\theta) = \min\{t \geq \gamma_k^1(\theta);\ s_1(t,\theta) = s_2(t,\theta)\}.$$

Furthermore, let

$$w_1(t,\theta) = \sum_{k=1}^{\infty} I_{\{\tau_k^1(\theta) \leq t \leq \tau_k^2(\theta)\}} \quad \text{and}$$

$$w_2(t,\theta) = \sum_{k=1}^{\infty} I_{\{\gamma_k^1(\theta) \leq t \leq \gamma_k^2(\theta)\}}.$$

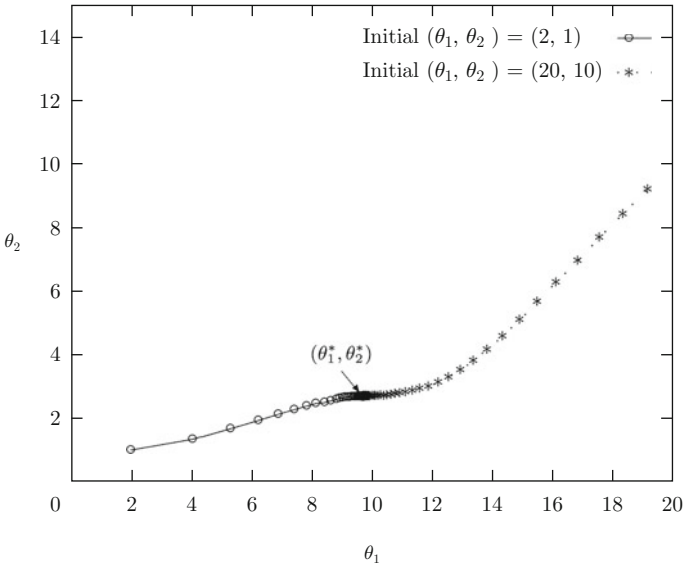

FIGURE 9.2. Convergence of the Iterates

Note that at most one of the indicator functions in the sums above can be positive at a given time t. The summation is thus well defined. Then the integrand of the gradient estimates can be written as

$$g_1(\theta, \xi(t)) = c_1 I_{\{t \geq \tau_0^2(\theta)\}} + c_2^+ w_1(t, \theta) I_{\{s_2(t,\theta) \geq 0\}} - c_2^- w_1(t, \theta) I_{\{s_2(t,\theta) < 0\}},$$

$$g_2(\theta, \xi(t)) = c_2^+ w_2(t, \theta) I_{\{s_2(t,\theta) \geq 0\}} - c_2^- w_2(t, \theta) I_{\{s_2(t,\theta) < 0\}}$$

via perturbation analysis (see Ho and Cao [87] and Glasserman [74]). The notation of the stopping times, suggested in Kushner and Vázquez-Abad [143], allows us to write the gradient estimates in a compact form. In the original paper of Yan, Yin, and Lou [220], some auxiliary processes were used in lieu of the stopping times. The ideas are the same, however.

Figure 9.2 demonstrates the performance of the algorithm for two machine case. One may generate contour curves via simulation for each set of threshold values, the approximation obtained in our algorithm can be seen to belong to the region of optimality. Our numerical results demonstrate that the initial conditions do not affect the algorithm much and the algorithm is robust with respect to the initial data.

9.3.4 Error Bounds

This subsection continues our investigation of Algorithm (9.7). We derive an error bound on the approximation sequence. The consideration of this subsection falls into the category of rates of convergence.

Theorem 9.13. *Assume that the conditions of Theorem 9.8 are satisfied and there is a twice continuously differentiable Liapunov function $V(\cdot)$ such that*

$$V(\theta) \geq 0,\ V(\theta) \to \infty \text{ as } |\theta| \to \infty,$$

$$(V_\theta(\theta))' \nabla J(\theta) \geq \kappa_0 V(\theta),$$

for some $\kappa_0 > 0$, and $V_{\theta\theta}(\cdot)$ is bounded, where $V_\theta(\cdot)$ and $V_{\theta\theta}(\cdot)$ denote the first and the second derivatives of $V(\cdot)$, respectively, and ζ' denotes the transpose of a vector $\zeta \in \mathbb{R}^{n\times 1}$. Suppose that for each θ,

$$\begin{aligned} &\left| \sum_{j=k}^{\infty} E^{\mathcal{F}_k} \frac{1}{T} \int_{jT}^{(j+1)T} (g(\theta, \xi(t)) - \nabla J(\theta))\, dt \right| \leq K \text{ and} \\ &\left| \sum_{j=k}^{\infty} E^{\mathcal{F}_k} \frac{1}{T} \int_{jT}^{(j+1)T} (g(\theta, \xi(t)) - \nabla J(\theta))_\theta\, dt \right| \leq K \end{aligned} \tag{9.19}$$

for some $K > 0$, where $E^{\mathcal{F}_k}$ denotes the conditional expectation with respect to $\mathcal{F}_{kT} = \sigma\{\xi(s), s \leq kT\}$. Assume that

$$|g(\theta, \xi)|^2 + |\nabla J(\theta)|^2 \leq K(1 + V(\theta)).$$

Then

$$\limsup_{k\to\infty} V(\theta_k) = O(\eta). \tag{9.20}$$

Remark 9.14. An alternative form of the first inequality in (9.19) is

$$\left| \int_{kT}^{\infty} E^{\mathcal{F}_k} (g(\theta, \xi(t)) - \nabla J(\theta))\, dt \right| \leq K, \tag{9.21}$$

and similar analogue holds for the second inequality in (9.19). It is readily seen that if $\xi(\cdot)$ is a ϕ-mixing process with mixing rate $\rho(\cdot)$ such that $\int_0^\infty \rho(t) < \infty$, the mixing inequality (see Kushner [139, p. 82]) implies that

$$\left| \int_{kT}^{\infty} E^{\mathcal{F}_k} (g(\theta, \xi(t)) - \nabla J(\theta))\, dt \right| \leq 2 \int_{kT}^{\infty} \rho(t - kT) dt \leq K,$$

with similar estimates regarding the second inequality in (9.19).

Outline of Proof of Theorem 9.13: We use a technique known as perturbed Liapunov function method (see Kushner and Yin [145] and the references therein). Since the proof is similar to that of [145, Chapter 10], only an outline of the idea is given.

By virtue of a Taylor expansion, direct calculation leads to

$$\begin{aligned} &E^{\mathcal{F}_k} V(\theta_{k+1}) - V(\theta_k) \\ &\quad = -\eta (V_\theta(\theta_k))' \nabla J(\theta_k) + O(\eta^2)(1 + V(\theta_k)) \\ &\qquad - \frac{\eta (V_\theta(\theta_k))'}{T} E^{\mathcal{F}_k} \int_{kT}^{(k+1)T} (g(\theta_k, \xi(t)) - \nabla J(\theta_k))\, dt. \end{aligned}$$

Define

$$V_1^\eta(k) = -\frac{\eta E^{\mathcal{F}_k} (V_\theta(\theta_k))'}{T} \sum_{j=k}^{\infty} \int_{jT}^{(j+1)T} (g(\theta_k, \xi(t)) - \nabla J(\theta_k))\, dt,$$

$$V^\eta(k) = V(\theta_k) + V_1^\eta(k).$$

It is easily seen that

$$|V_1^\eta(k)| \le \eta K (1 + V(\theta_k)). \tag{9.22}$$

Detailed calculation leads to

$$E^{\mathcal{F}_k} V^\eta(k+1) - V^\eta(k) \le -\eta \kappa_0 V(\theta_k) + O(\eta^2)(1 + V(\theta_k)).$$

Equation (9.22) then yields that

$$E^{\mathcal{F}_k} V^\eta(k+1) \le V^\eta(k) - \eta \kappa_0 V^\eta(k) + O(\eta^2)(1 + V^\eta(k)).$$

By choosing η small enough, we obtain

$$E^{\mathcal{F}_k} V^\eta(k+1) \le \left(1 - \frac{\eta \kappa_0}{2}\right) V^\eta(k) + O(\eta^2).$$

Iterating on the above inequality, taking expectation and lim sup as $k \to \infty$, and using (9.22), the desired result follows. □

Remark 9.15. If the Liapunov function is locally (near θ^*) quadratic, it can be shown that there is an N_η such that

$$\left\{ U_k = \frac{\theta_k - \theta^*}{\sqrt{\eta}} : \ k \ge N_\eta \right\}$$

is tight. Define

$$U^\eta(t) = U_k \text{ for } t \in [(k - N_\eta)\eta, (k - N_\eta + 1)\eta).$$

Under further conditions, one can obtain a local results in connection with a stochastic differential equation.

9.4 Notes

This chapter has been devoted to the numerical solutions of the control problems. It consists of numerical methods for solving the HJB equations and approximation for threshold control policies. The main techniques used are the finite difference approximation methods (see Kushner [138], Kushner and Dupuis [141]) and the stochastic optimization methods (see Kushner and Yin [145]).

The computational methods for the optimal control problem presented here are equivalent to methods of computing the optimal controls for discrete Markov chain models. For a general background and discussion on the method and many references for controlled diffusion and jump diffusion processes, we refer to Kushner [138] and Kushner and Dupuis [141]. For applications of such methods in manufacturing models, we refer the reader to Yan and Zhang [221] among others. In the implementation of the numerical method, one may use either "value iteration" that is essentially a fixed point iteration, or "policy iteration," and the variation and/or modification of the aforementioned procedures, such as Jacobi iteration, Gauss-Seidel method, and accelerated Jacobi and Gauss-Seidel methods (see [141] for a detailed discussion on this and related matters). In practice, one often wishes to use an accelerated procedure to speed up the computation. The recent advances on multigrid and domain decomposition methods give new hope for solving large-dimensional systems. As its deterministic counterpart, by and large, this is still a current research topic for stochastic systems.

Converting optimal control problems into optimization problems under threshold type of control policies is in Yin, Yan, and Lou [228], and Yan, Yin, and Lou [220]. Early work on developing hedging policies is in Kimemia and Gershwin [121]; related work along this direction is in Caramanis and Liberopoulos [24]. For systems arising in production planning see [220] where a combined approach of the infinitesimal perturbation analysis with stochastic optimization methods is utilized. The computation presented here was done by Houmin Yan. We are very grateful for his help. For further approaches on using stochastic approximation based algorithms for optimization with long-run average costs, see Kushner and Yin [145, Chapter 9] and the references therein. In applications, one may use a projection or truncation algorithm. Treatments of such algorithms and a comprehensive study on stochastic approximation algorithms can be found in [145].