The Potential Structure of Sample Paths and Performance Sensitivities of Markov Systems

Xi-Ren Cao

***Abstract*—We study the structure of sample paths of Markov systems by using performance potentials as the fundamental units. With a sample path-based approach, we show that performance sensitivity formulas (performance gradients and performance differences) of Markov systems can be constructed intuitively, by first principles, with performance potentials (or equivalently, perturbation realization factors) as building blocks. In particular, we derive sensitivity formulas for two Markov chains with possibly different state spaces. The proposed approach can be used to obtain flexibly the sensitivity formulas for a wide range of problems, including those with partial information. These formulas are the basis for performance optimization of discrete event dynamic systems, including perturbation analysis, Markov decision processes, and reinforcement learning. The approach thus provides insight on on-line learning and performance optimization and opens up new research directions. Sample path based algorithms can be developed.**

***Index Terms*—Markov decision processes, performance sensitivity, perturbation analysis, perturbation realization, reinforcement learning.**

I. INTRODUCTION

MOTIVATED by many engineering, economic, and social problems in the information technology era, researchers in different scientific disciplines have developed various approaches to the optimization of discrete event dynamic systems (DEDSs); among them are perturbation analysis (PA) in control theory [7], [8], [13], [15], [21], [22], Markov decision processes (MDPs) in operations research [2], [5], [6], [27], [28], and reinforcement learning (RL) in computer science [3], [4], [23]–[25], [29]–[33]. These approaches share a common feature: to improve a system's performance based on the information obtained by analyzing the current system behavior. Markov process is the common model used in these approaches.

Recent research indicates that the above different approaches are closely related [10], [12], and that the performance sensitivities are the basis for optimization. The relation among these areas can be illustrated by Fig. 1 (see [12]). At the center are the two types of performance sensitivity equations (the notations in the figure are for the standard Markov models and will be explained later): When the system parameters are continuous variables, the sensitivity is the performance gradient (derivatives) with respect to the parameter(s); when the system is characterized by discrete quantities (e.g., policies), the sensitivity is the performance difference between the system with two different set of parameters (e.g., under two different policies). The fundamental concept for both sensitivities is the performance potential (denoted as g) of a Markov process (which differs from the "relative cost vector" in [5], or the "bias" in [28], by only a constant). Both PA and MDP can be explained from a performance sensitivity point of view: PA gives the performance gradients (called policy gradient in RL literature) and policy iteration can be derived naturally from the performance difference formula [10], [12], [19]. Both of them can be implemented using a single sample path; RL, Q-learning, $\mathrm{TD}(\lambda)$, neuro-dynamic programming, etc, are sample-path-based efficient ways of estimating the performance potentials and other related quantities (e.g., Q-factors). As shown in the figure, performance optimization can be achieved either by using the performance gradients combined with stochastic approximation methods, see, e.g., [16], [20], [26], and [34] (in each step, performance improves by a small amount), or by applying policy iteration algorithms in MDPs (in each iteration policy jumps to another one with a better performance).

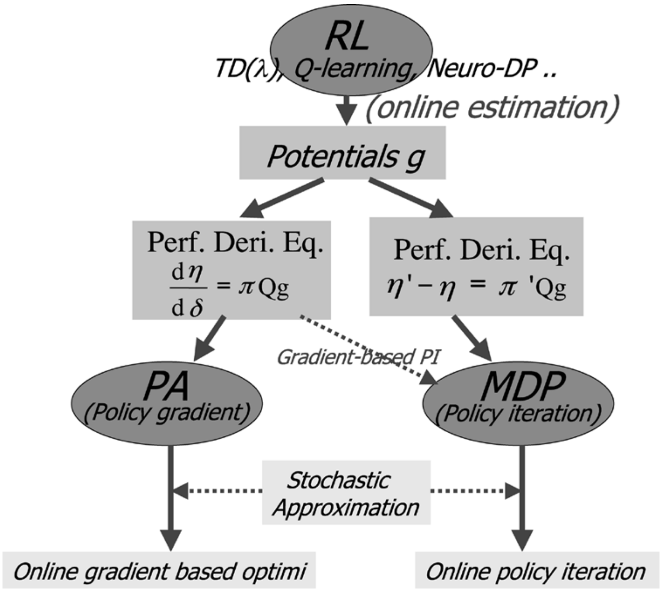

Fig. 1. Map of the optimization world with two sensitivity formulas at the center.

Fig. 1 describes the relations among the different optimization areas and illustrates how the two sensitivity formulas lead to different optimization approaches. However, because these two sensitivity formulas are derived with the standard Markov

Manuscript received April 23, 2003; revised June 22, 2004. Recommended by Associate Editor Y. Wardi. This work was supported in part by a grant from the Hong Kong UGC.

The author is with the Hong Kong University of Science and Technology, Kowloon, Hong Kong (e-mail: eecao@ee.ust.hk).

Digital Object Identifier 10.1109/TAC.2004.838494

model, the results thus obtained do not cover nonstandard problems. It is well known that the standard Markov approach suffers from two major drawbacks: the state–space is usually too large, and the actions have to be chosen independently at different states.

The research reported in this paper is a part of our effort in extending the above sensitivity view and the approaches depicted in Fig. 1 to more general cases. The extension allows approaches similar to those discussed above be applied to problems in which the state space may be different, and/or the actions at different states may be dependent; aggregation can be used to reduce the state space. The central work is to derive the two types of sensitivity formulas for more general problems. As illustrated in Fig. 1, once the sensitivity formulas are obtained, optimization approaches such as policy gradient, policy iteration, and RL, etc, may be derived for these general (nonstandard) problems.

To achieve our goal, we first study the structure of performance sensitivities (i.e., performance gradient and performance difference). We show that both of them can be constructed, by first principles, using performance potentials as building blocks. We show that the construction of the performance sensitivities is based on the sample-path structure determined by performance potentials. More precisely, we can build up a sample path of one system with that of another system and its performance potentials. With the potential structure of sample paths introduced in this paper, we can flexibly construct performance sensitivities for many systems; these sensitivity formulas are the basis for performance optimization. The construction approach is an extension of the previous work [9], [13], in which a simple case was studied: the performance gradient when the parameterized systems are in the same irreducible state space. In this paper, we extend this result to cover performance differences and more general systems. A major breakthrough in this extension is the decomposition in dealing with the coupling effect of two jumps that are close to each other; in PA, this represents a change from infinitesimal perturbations to finite ones. As examples, we apply the approach to both the performance derivatives and the performance differences for two systems that may have possibly two different state spaces with a common subspace.

Following the terminology of PA, we refer to the two systems under comparison as an original system and a perturbed one, respectively, and their sample paths the original path and the perturbed path, respectively. (For performance derivatives with respect to a parameter θ, the original system is the one with θ, and the perturbed one, with $\theta + \Delta\theta$). The main idea is as follows: Any change in system parameters or even in the system structure is reflected by "jumps" on the system's sample path; a jump here refers to the case that from the same sate, the original path transits to state i, while the perturbed one transits to state j. The effect of such a single jump from i to j on the system performance can be measured by a quantity called the *realization factor* $d(i,j)$ which equals $g(j) - g(i)$, where $g(i)$ is the *performance potential* at state i. Both $d(i,j)$ and $g(i)$ can be estimated on sample paths. Finally, the performance sensitivities, which reflect the effect of the change in parameters and/or structure, can be decomposed into the effects of many single jumps on the system's sample path and can be therefore constructed by using realization factors or potentials as building blocks.

Using the aforementioned idea, we can derive performance sensitivity formulas by first principles. Such an approach is common in physics where researchers first formulate and solve problems based on experimental evidence by first principles and then prove their results rigorously. In this aspect, we can view the sample path based reasoning for constructing performance sensitivities as "thought experiments." In Section II, we briefly review the concepts of realization factors and performance potentials. In Section III–A, we apply our idea to construct the performance derivatives for the case where the perturbed system and original one are in the same state–space. This is the simplest case and serves as a template for more complicated cases to follow. In Section III-B, we derive a formula for the performance difference of two Markov chains in the same state space. In Section IV, we further illustrate the idea by applying it to more cases. In Section IV-A, we construct both the performance derivatives and differences for two Markov chains with one state space being a subspace of the other. In Section IV-B, we extend the results to the case where the two Markov chains have different state spaces with a common subspace. In Section V, we show, by two examples, that the same principle can be applied to construct performance sensitivities for systems with partial information that can be obtained on a single sample path. These examples also serve to illustrate the flexibility in applying the "building block" idea to construct performance sensitivities: we only need to focus our attention to the states that are affected by the parameter changes. In addition, these examples show that performance potentials can be aggregated and the number of quantities need to be estimated on a sample path can be reduced.

The contributions of this paper are as follows. We propose an intuitive approach to construct, by first principles, performance derivatives and differences by using potentials as the fundamental building blocks; or equivalently, we show that the difference between two sample paths of two different systems can be decomposed into path segments that can be measured on the average by performance potentials of one of the systems. This clearly illustrate the physical meaning of potentials or realization factors and their crucial role in performance optimization of discrete event dynamic systems. Using this approach, we can flexibly derive formulas for performance sensitivities which are otherwise not easy to conceive. Especially, we obtained sensitivities formulas for Markov chains with different state spaces and for systems with partial information. Compared with the traditional MDP solutions where the potentials (i.e., biases) at all states are treated as a vector and considered as a group altogether, our approach offers a novel view to potentials by treating them separately and flexibly. Next, the sensitivity formulas constructed for general problems play the central role as the two standard sensitivity formulas illustrated in Fig. 1. Therefore, these sensitivity formulas lead to new research topics for future studies. For example, new policy iteration algorithms can be developed, and sample-path-based algorithms can be developed for estimating performance gradients and implementing policy iterations. Furthermore, as illustrated in Section V, the approach applies to problems in which actions at different states are dependent and aggregation techniques can be used to reduce the computation complexity. Further research is needed in these directions.

In this paper, we study the long-run average performance criteria; but the results can be easily extended to discounted performance criteria (see [11]). Also, for simplicity, we study systems with discrete time and a finite state space; thus, the model is a finite-state discrete-time Markov process; which we shall refer to as a "Markov chain."

II. Perturbation Realization and Performance Potentials

We first review some fundamental concepts and their related theory. Consider a Markov chain with transition probability matrix P [17]. P may depend on a parameter θ and therefore is sometimes denoted as $P(\theta)$. Let $\mathcal{S} = \{1, 2, \ldots, M\}$ be the state space, $\mathbf{X} = \{X_0, X_1, \ldots, X_n, \ldots\}$ be a sample path, and $f : \mathcal{S} \rightarrow \mathcal{R}$ be the cost function. In this section, we assume that P is irreducible and aperiodic and hence ergodic. Define the steady-state probability as a row vector $\pi = (\pi(1), \ldots \pi(M))$, then its flow balance equation is

$$\pi P = \pi \quad \pi e = 1 \tag{1}$$

where $e = (1, 1, \ldots, 1)^T$ is an M-dimensional column vector whose all components are 1's, and the superscript "T" denotes transpose. We will use subscript to indicate the dimension of e when it is needed (e.g., $e \equiv e_M$ in (1)). The performance measure is defined as

$$\begin{aligned}\eta = \sum_{i=1}^{M} \pi(i) f(i) = \pi f &= \lim_{L\to\infty} \frac{1}{L} \sum_{l=0}^{L-1} f(X_l) \\ &= \lim_{L\to\infty} \frac{F_L}{L}, \qquad w.p.1 \end{aligned} \tag{2}$$

where $f = (f(1), \ldots, f(M))^T$ (we use f as both a function and a vector) and

$$F_L = \sum_{l=0}^{L-1} f(X_l);$$

the limit in (2) exists with probability one.

The central concept of optimization of DEDSs is the *perturbation realization*. The perturbation *realization factor* $d(i,j)$ measures the effect of a jump (or called a perturbation) from state i to state j on F_L and is defined as follows. Consider two independent Markov chains $\mathbf{X} = \{X_n; n \geq 0\}$ and $\mathbf{X}' = \{X'_n; n \geq 0\}$ with $X_0 = i$ and $X'_0 = j$; both of them have the same transition matrix P. We define [9], [13]

$$\begin{aligned} d(i,j) &= \lim_{L\to\infty} E[F'_L - F_L \,|\, X'_0 = j, X_0 = i] \\ &= \lim_{L\to\infty} E\left[\sum_{l=0}^{L-1} (f(X'_l) - f(X_l)) \,\middle|\, X'_0 = j, X_0 = i \right], \\ &\qquad i,j = 1, \ldots, M. \end{aligned} \tag{3}$$

This is the average of the difference of F_L defined in (2) starting from two different states j and i. If P is irreducible, then with probability one the two sample paths of $\mathbf{X}$ and $\mathbf{X}'$ will merge together. That is, there is a random number L^* such that $X'_{L^*} = X_{L^*}$. Therefore, by the strong Markov property, (3) becomes [9], [13]

$$d(i,j) = E\left[\sum_{l=0}^{L^*-1} (f(X'_l) - f(X_l)) \,\middle|\, X'_0 = j, X_0 = i \right], \qquad i,j = 1, \ldots, M. \tag{4}$$

From the definition (3), we have $d(i,j) = -d(j,i)$, and $d(i,i) = 0$. $d(i,j)$ can also be expressed with a single sample path: let $X_0 = i$ and $L_i(j) = \min |\{n \geq 0, X_n = j | X_0 = i\}$ be the first passage time to state j. Then, it is not difficult to prove [13]

$$d(j,i) = E\left\{ \sum_{l=0}^{L_i(j)-1} [f(X_l) - \eta] \,\middle|\, X_0 = i \right\}. \tag{5}$$

The matrix $D \in \mathcal{R}^{M\times M}$, with $d(i,j)$ as its (i,j)th element, is called a *perturbation realization matrix*. We have $D^T = -D$. From the definition (3) or (4), by simple calculation, we can verify the Lyapunov equation

$$D - PDP^T = F \tag{6}$$

where $F = ef^T - fe^T$. Again, by (3), we have

$$d(i,j) = d(i,k) + d(k,j), \qquad i,j,k \in \mathcal{S}. \tag{7}$$

Thus, D takes the form

$$D = eg^T - ge^T$$

where $g = (g(1), \ldots, g(M))^T$ is called a *performance potential* (or simply *potential*) vector and $g(i)$ the potential at state i. This equation is equivalent to

$$d(i,j) = g(j) - g(i), \qquad i,j = 1, 2, \ldots, M.$$

Since $d(i,j)$ measures the difference of the performance starting from states j and i, $g(i)$ measures the average contribution to F_L of every visit to state i. Furthermore, only the difference between different $g(i)$s are important for performance sensitivities. (7) resembles the conservation law of the potential energy in physics. From (6), we can prove that g satisfies the Poisson equation

$$(I - P)g + \eta e = f. \tag{8}$$

The solution to (8) is only up to an additive constant; i.e., if g satisfies (8), then for any constant c, $g + ce$ also does. Therefore, there must be one particular solution to (8) such that $\pi g = \eta$ (With any solution g_0, let $c = \eta - \pi g_0$ and $g = g_0 + ce$). For this solution, (8) becomes

$$(I - P + e\pi)g = f. \tag{9}$$

For ergodic Markov chains, $(I - P + e\pi)$ is invertible, thus $g = (I - P + e\pi)^{-1} f$, where $(I - P + e\pi)^{-1}$ is called a *fundamental matrix*. From this, we can prove (up to an additive constant)

$$g(i) = \lim_{L\to\infty} E\left\{ \sum_{l=0}^{L-1} [f(X_l) - \eta] \,\middle|\, X_0 = i \right\} \tag{10}$$

which is finite for ergodic chains.

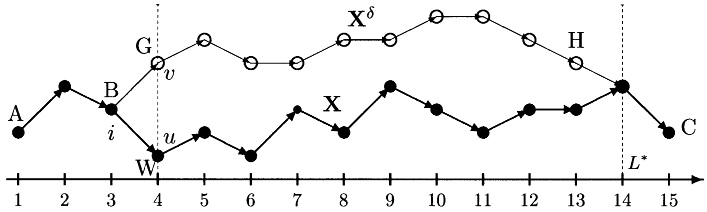

Fig. 2. Perturbation in a sample path of a Markov chain and its effect.

III. Markov Chains on the Same State Space

In this section, we derive performance sensitivity formulas by applying first principles for the case where the Markov chains under comparison are defined on the same state space. A brief intuitive explanation of the performance derivative in Section III-A has appeared in [9], we provide here a more detailed derivation, which motivates the study in subsequent sections and makes the material in this paper complete.

A. Performance Derivatives

We first study the simplest problem, i.e., the performance derivative, to introduce the idea. Given a Markov chain with transition probability matrix P and state space $\mathcal{S} = \{1, 2, \ldots, M\}$, let P' be another irreducible transition matrix on the same state space $\mathcal{S}$, $P'e = 1$, and set $Q = P' - P$. Thus $Qe = 0$. For any $0 \le \delta \le 1$, define $P^\delta = P + Q\delta$. We have $P^\delta e = 1$ and P^δ is also an irreducible transition matrix. The quantities associated with P^δ are denoted as π^δ and η^δ, etc. We view δ as very small when we study derivatives.

Our approach is sample-path based, so we first consider the simulation of a Markov chain with transition probability matrix P. At any time l with $X_l = i, l = 0, 1, \ldots$, we generate a uniformly distributed random variable $\xi_l \in [0, 1)$. If

$$\sum_{k=0}^{u-1} p(i,k) \le \xi_l < \sum_{k=0}^{u} p(i,k), \qquad p(i,0) = 0, \qquad u \in \mathcal{S} \tag{11}$$

then we set $X_{l+1} = u$. For example, consider the case where $p(i,u) = 0.5, p(i,v) = 0.5$, and $p(i,k) = 0$ for all $k \neq u, v$. If $0 \le \xi < 0.5$, then the Markov chain jumps into state u; otherwise, it jumps into state v. Suppose that the transition probabilities change to $p^\delta(i,u) = 0.5 - \delta, p^\delta(i,v) = 0.5 + \delta$, and $p^\delta(i,k) = 0$, if $k \neq u, v$. (i.e., $q(i,u) = -1, q(i,v) = 1$, and $q(i,k) = 0$ if $k \neq u, v$.) We use the same sequence of random variable ξ_l to determine the transition of the Markov chain with P^δ at time $l, l = 0, 1, \ldots$, its sample path is denoted as $\mathbf{X}^\delta$. We observe that if it happens that $\xi_l \in [0.5 - \delta, 0.5)$, then $\mathbf{X}$ transits to state u, but $\mathbf{X}^\delta$ transits to v; however, because δ is very small, most likely we have $\xi_l \notin [0.5 - \delta, 0.5)$, in this case both $\mathbf{X}$ and $\mathbf{X}^\delta$ transit in the same way.

Because δ is very small, P^δ is very close to P. Thus, the previous discussion indicates that starting from the same initial state X_0 and with the same random sequence $\xi_l, l = 0, 1, \ldots$, the two sample paths $\mathbf{X}^\delta = \{X_0^\delta, X_1^\delta, \ldots, \}$ and $\mathbf{X} = \{X_0, X_1, \ldots, \}$ are also very close. Suppose that with the same values of $\xi_n, n = 0, 1, \ldots, l-2$, we have $X_n^\delta = X_n$, for $n = 0, 1, \ldots, l-1$. Denote $X_{l-1} = i$. Furthermore, we assume that with the same value of ξ_{l-1}, applying (11) to P determines that $\mathbf{X}$ transits to state $X_l = u$, but applying (11) to P^δ determines that $\mathbf{X}^\delta$ transits to state $X_l^\delta = v$. We say that the perturbed chain $\mathbf{X}^\delta$ has a *jump* (or perturbation) from u to v at time l. In Fig. 2, $\mathbf{X}$ and $\mathbf{X}^\delta$ are illustrated by the solid dots (A-B-W-C) and hollow circles (A-B-G-C), respectively; the perturbed path $\mathbf{X}^\delta$ has a jump from u to v at $l = 4$. After this time, the two sample paths differ until at L^* ($L^* = 14$ in Fig. 2) they merge together. Because δ is very small, we can assume that such jumps occur rarely; in particular, we can assume that between l ($= 4$ in Fig. 2) and L^* ($= 14$ in Fig. 2) both $\mathbf{X}$ and $\mathbf{X}^\delta$ evolve in the same way, i.e., according to the same transition probability P. In other words, all the transitions on A-B-G-C (including those points between G and H) except the one from X_3^δ to X_4^δ look the same as if they follow the transition matrix P. From Fig. 2, it is clear that $d(u,v)$ defined in (4) measures the average affect of a jump from u to v on F_L in (2). Note the states of $\mathbf{X}$ and $\mathbf{X}^\delta$ between $l = 4$ to $l = 13$ are completely different, we may assume the transitions of $\mathbf{X}$ and $\mathbf{X}^\delta$ in this period are independent.

Now we consider a sample path $\mathbf{X}$ consisting of $L, L \gg 1$, transitions. Among these transitions, on the average there are $L\pi(i)$ transitions at which the system is at state i. Each time when $\mathbf{X}$ visits state i, because of the change from P to P^δ, the perturbed path $\mathbf{X}^\delta$ may have a jump, denoted as from state u to state v, as shown in Fig. 2. For convenience, we allow $u = v$ as a special case. A "real jump" (with $u \neq v$) happens rarely. Denote the probability of a jump from u to v after visiting state i as $b(i,u,v)$. We have

$$b(i,u,v) = p(i,u)p^\delta(i,v \,|\, i,u)$$

where $p^\delta(i,v \,|\, i,u)$ denotes the conditional probability that $\mathbf{X}^\delta$ transits from state i to state v given that $\mathbf{X}$ transits from state i to u. Thus

$$\sum_{v=1}^{M} b(i,u,v) = p(i,u). \tag{12}$$

Similarly

$$\sum_{u=1}^{M} b(i,u,v) = p^\delta(i,v) \tag{13}$$

and $\sum_{u,v=1}^{M} b(i,u,v) = 1$. On the average, in these L transitions there are $L\pi(i)b(i,u,v)$ jumps from u to v following visiting i. Each has on the average an effect of $d(u,v)$ on F_L.

Because a real jump happens extremely rarely as $\delta \to 0$, the effects of two real jumps can be decoupled and therefore considered separately. More precisely, consider Fig. 3 which illustrates two jumps at $l = 4$ and $l = 11$. After the first jump, $\mathbf{X}^\delta$ merges with $\mathbf{X}$ at $l = 7$; thus, the effects of the two jumps shown in Fig. 3 can be measured separately. As δ is very small,

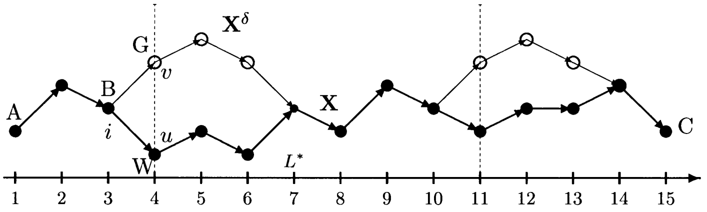

Fig. 3. Effect of two rare perturbations are decoupled.

the probability that another jump occurs before $l = 7$ is of order δ^2; its effect can be ignored for performance derivatives. Thus, on the average the total effect on F_L due to the change in P to $P^\delta = P + \delta Q$ is

$$\begin{aligned} E(F_L^\delta - F_L) &\approx \sum_{i=1}^{M} \left\{ \sum_{u,v=1}^{M} L\pi(i) b(i,u,v) d(u,v) \right\} \\ &= \sum_{i=1}^{M} \left\{ \sum_{u,v=1}^{M} L\pi(i) b(i,u,v)[g(v) - g(u)] \right\} \\ &= \sum_{i=1}^{M} L\pi(i) \left\{ \left\{ \sum_{v=1}^{M} \left[g(v) \sum_{u=1}^{M} b(i,u,v) \right] \right\} \right. \\ &\quad \left. - \left\{ \sum_{u=1}^{M} \left[g(u) \sum_{v=1}^{M} b(i,u,v) \right] \right\} \right\}. \qquad (14) \end{aligned}$$

From (13) and (12), (14) becomes

$$\begin{aligned} E(F_L^\delta - F_L) &\approx \sum_{i=1}^{M} L\pi(i) \left\{ \left\{ \sum_{v=1}^{M} [p^\delta(i,v) g(v)] \right\} \right. \\ &\quad \left. - \left\{ \sum_{u=1}^{M} [p(i,u) g(u)] \right\} \right\} \\ &= \sum_{i=1}^{M} L\pi(i) \left\{ \sum_{j=1}^{M} [p^\delta(i,j) - p(i,j)] g(j) \right\} \\ &= L\pi[P^\delta - P]g = L\pi Q\delta g. \qquad (15) \end{aligned}$$

Thus

$$\eta^\delta - \eta = \lim_{L\to\infty} \frac{1}{L} E(F_L^\prime - F_L) = \pi Q \delta g. \qquad (16)$$

Finally, we get

$$\frac{d\eta}{d\delta} = \pi Q g. \qquad (17)$$

Strictly speaking, the approximation in (14) is not accurate (the difference of both sides may not be small for a large L). It is accurate only after both sides of (14) is divided by L, resulting in (16). Nevertheless, (14) provides a good intuition for thinking.

Note that g, D, and π can be estimated on a single sample path of a Markov chain with transition matrix P; thus, given any P', the performance sensitivity along the direction $Q = P' - P$ can be obtained by (17) using the sample path-based estimates of π and g. Algorithms can be developed for estimating the performance sensitivity based on a single sample path using (17) without estimating each component of g [14], [3], and [4].

B. Performance Differences of Two Markov Chains

In this section, we show how we can use realization factors, or potentials, as building blocks to construct the difference of the performance of two different Markov chains.

Consider the simulation of two Markov chains with transition probability matrices P and P', respectively, on the same state–space $\mathcal{S} = \{1, 2, \ldots, M\}$. As we see in Section III-A, for $P^\delta = P + \delta(P' - P)$ with small δ, if we use the same random sequence for both chains, then the two sample paths $\mathbf{X}^\delta$ and $\mathbf{X}$ are very close, and the jumps happen rarely on $\mathbf{X}^\delta$ and their effects can be treated separately. However, when we consider $P' = P + Q$ ($\delta = 1$), two sample paths $\mathbf{X}'$ and $\mathbf{X}$ are completely different and the effect of the jumps may be coupled (after a jump on $\mathbf{X}'$, another jump may occur before $\mathbf{X}'$ and $\mathbf{X}$ merge together).

We first show how to determine the effect of two "coupled" jumps. In Fig. 4, A-B-W-C is the original sample path $\mathbf{X}$ (with P) and A-B-G-E-H-F is the perturbed path $\mathbf{X}'$ with P'. Suppose the sample path $\mathbf{X}'$ (**not** $\mathbf{X}$!) is generated with a sequence of uniformly distributed $[0, 1)$ random variables $\xi_1, \xi_2, \ldots, \xi_l, \ldots$. We use a similar terminology as for the performance derivative problem: If with ξ_{l-1} from X'_{l-1} (which is most likely different from X_{l-1}) the Markov chain transits to the same state according to both P' and P, we say that the sample path $\mathbf{X}'$ does not have a jump at l. However, if with ξ_{l-1}, X'_{l-1} transits to state u according to P while it transits to state $X'_l = v$ according to P', we say that the perturbed chain $\mathbf{X}'$ has a jump (or a perturbation) from u to v at time l. Fig. 4 illustrates two jumps on $\mathbf{X}'$, one at $l = 4$ from u_1 to v_1, the other at $l = 9$ from u_2 to v_2.

One cannot see u_2 on either $\mathbf{X}$ or $\mathbf{X}'$. In Fig. 4, we have added point R to illustrate the transition at $l = 8$ (according to P) to state u_2. Thus, all the transitions on G-E-R are the same as if they follow the transition matrix P. Now, after R, we add an auxiliary path that follows P until the auxiliary path merges with $\mathbf{X}$ at $l = 14$. Let us denote the path A-B-W-C as path 1, A-B-E-R-C as path 2, and A-B-E-F as path 3. Path 1 follows P (hence $\mathbf{X}$), and Path 3 follows P' (hence $\mathbf{X}'$) on which the segments A-B, G-E, and H-F are the same as if they were generated according to P. With the auxiliary path, segment G-E-R-C also follows P. Now it is clear that the effect of the jump from u_1 to v_1 can be measured by G-E-R-C and W-C, and that of the jump from u_2 to v_2 by H-F and R-C, all these segments follow the transition matrix P.

Let us make the previous observation precise. We use superscripts to indicate the paths associated with. For example, the sequences of states on these three paths are denoted as

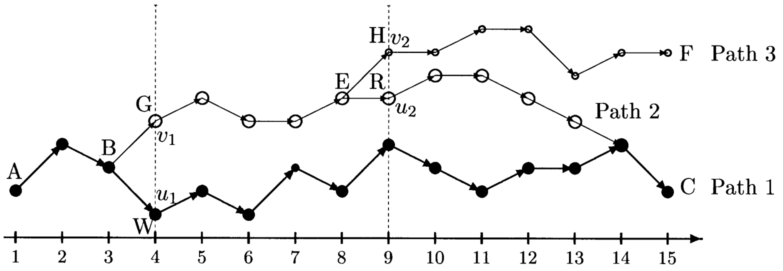

Fig. 4. Effect of two perturbations.

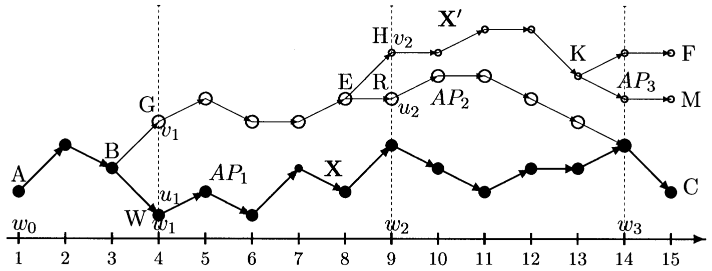

Fig. 5. Potential structure of a sample path.

$X_1^{(1)}, X_2^{(1)}, \ldots; X_1^{(2)}, X_2^{(2)}, \ldots;$ and $X_1^{(3)}, X_2^{(3)}, \ldots;$ respectively. Of course, some states are the same in two or three paths, e.g., $X_1^{(1)} = X_1^{(2)} = X_1^{(3)}$, and $X_8^{(2)} = X_8^{(3)}$. It is clear from Fig. 4 that for any $L > 9$

$$
\begin{aligned}
F_L^{(2)} - F_L^{(1)} &= \sum_{l=1}^{L} f\left(X_l^{(2)}\right) - \sum_{l=1}^{L} f\left(X_l^{(1)}\right) \\
&= \sum_{l=4}^{L} f\left(X_l^{(2)}\right) - \sum_{l=4}^{L} f\left(X_l^{(1)}\right) \qquad (18)
\end{aligned}
$$

$$
F_L^{(3)} - F_L^{(2)} = \sum_{l=9}^{L} f\left(X_l^{(3)}\right) - \sum_{l=9}^{L} f\left(X_l^{(2)}\right). \qquad (19)
$$

Therefore

$$
\begin{aligned}
F_L^{(3)} - F_L^{(1)} = &\left\{ \sum_{l=4}^{L} f\left(X_l^{(2)}\right) - \sum_{l=4}^{L} f\left(X_l^{(1)}\right) \right\} \\
&+ \left\{ \sum_{l=9}^{L} f\left(X_l^{(3)}\right) - \sum_{l=9}^{L} f\left(X_l^{(2)}\right) \right\}. \qquad (20)
\end{aligned}
$$

Because G-E-C follows transition probability P, the expectation of $\sum_{l=4}^{L} f(X_l^{(2)}) - \sum_{l=4}^{L} f(X_l^{(1)})$ as $L \to \infty$ is $d(u_1, v_1)$. Similarly, $H - F$ also follows transition probability P. Path 3 eventually merges with Path 2 (before or after $l = 14$), thus, the expectation of $\sum_{l=9}^{L} f(X_l^{(3)}) - \sum_{l=9}^{L} f(X_l^{(2)})$ as $L \to \infty$ is $d(u_2, v_2)$. Finally, the effect of the two "coupled" jumps at $l = 4$ and $l = 9$ is on the average $d(u_1, v_1) + d(u_2, v_2)$.

The aforementioned observation for the two-jump case shed light on the general case (Fig. 5). If P changes to $P' = P + Q$, suppose that there are K jumps on $\mathbf{X}'$ (After the Kth jump, $\mathbf{X}'$ looks the same as if following P). Let $w_1, w_2, \ldots, w_K$, be the instants at which jump occurs, and denote the jump at w_k as from state u_k to state v_k. Let $w_0 = 1$. In Fig. 5, $w_1 = 4, w_2 = 9$, and $w_3 = 14$. By definition, the segments from X'_{w_k} to $X'_{w_{k+1}-1}, k = 0, 1, 2, \ldots$, are the same as if they were generated according to P. (It is possible that $w_{k+1} - 1 = w_k$, in such cases the segment is null.) As in the two-jump case, we add an auxiliary path starting from each $X'_{w_k - 1}$ that follows exactly the transition matrix P (e.g., the paths E-R-C and K-M in Fig. 5). Denote the auxiliary path starting from X'_{w_k-1} as AP_k. AP_1 is B-W-C, which is the same as a part of $\mathbf{X}$, AP_2 is E-R-C, and AP_3 is K-M. Denote the path from $X'_{w_0} (= X'_1)$ to X'_{w_1-1} to AP_1 as path 1 (Path A-B-W-C in Fig. 5), and the path from X'_{w_0} via $X'_{w_1-1}, X'_{w_2-1}, \ldots$, and X'_{w_k-1} and then to AP_k, as path k, etc., with path 1 being $\mathbf{X}$ (a sample path for P). We denote $\mathbf{X}'$ as path $K+1$.

Applying the same reasoning as for the two-jump case illustrated in Fig. 4, we can obtain an equation similar to (20) for the K-jump case (for $L > w_K$)

$$
\begin{aligned}
&F'_L - F_L \\
&\approx \left\{ \sum_{l=w_K}^{L} f\left(X_l^{(K+1)}\right) - \sum_{l=w_K}^{L} f\left(X_l^{(K)}\right) \right\} \\
&\quad + \left\{ \sum_{l=w_{K-1}}^{L} f\left(X_l^{(K)}\right) - \sum_{l=w_{K-1}}^{L} f\left(X_l^{(K-1)}\right) \right\} \\
&\quad + \cdots + \left\{ \sum_{l=w_1}^{L} f\left(X_l^{(2)}\right) - \sum_{l=w_1}^{L} f\left(X_l^{(1)}\right) \right\} \qquad (21)
\end{aligned}
$$

in which the expectation of $\sum_{l=w_k}^{L} f(X_l^{(k+1)}) - \sum_{l=w_k}^{L} f(X_l^{(k)})$ as $L \to \infty$ is $d(u_k, v_k)$.

Fig. 5 illustrates that a sample path, $\mathbf{X}'$, of a Markov system with transition matrix P' can be decomposed into the sum of a sample path, $\mathbf{X}$, with transition matrix P and many segments, G-E-C, H-K-M, and so on, that can be measured on the average by performance potentials of the Markov chain with P. Pictorially, the perturbed sample path $\mathbf{X}'$ in Fig. 5 starts from point A then follows transition matrix P on the original path $\mathbf{X}$ until point B, at which it jumps to point G according to P' then follows P again on another "original path" (with large circles) to

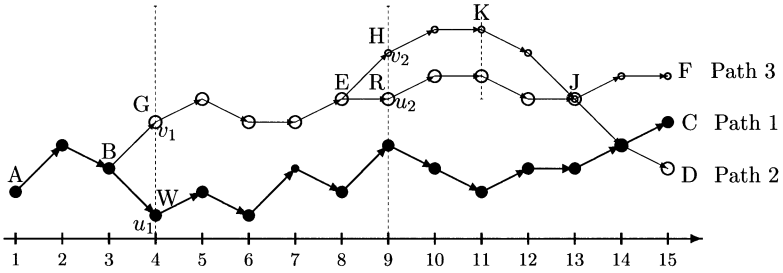

Fig. 6. Perturbation between two different state spaces.

point E, at which it jumps to point H according to P' then follows P again on another "original path" (with small circles) to point K, and so on.

Now, we consider a sample path of the perturbed system with $P' = P + Q$ with L states $\{X'_1, \dots, X'_L\}, L \gg 1$. Among these L states, on the average $L\pi'(i)$ states are at state i. Suppose that after visiting state $i, \mathbf{X}'$ has a jump from u to v (we allow $u = v$). Denote the probability of a jump from u to v after visiting i as $b(i,u,v), \sum_{u,v=1}^{M} b(i,u,v) = 1$. Then on the average, there are $L\pi'(i)b(i,u,v)$ jumps from u to v following visiting i. According to (21), each of such a jump has on the average an effect of $d(u,v)$ on F_L. Thus, on the average the total effect on F_L due to the change in P to P' is

$$\begin{aligned} E(F'_L - F_L) &\approx \sum_{i=1}^{M} \left\{ \sum_{u,v=1}^{M} L\pi'(i)b(i,u,v)d(u,v) \right\} \\ &= \sum_{i=1}^{M} \left\{ \sum_{u,v=1}^{M} L\pi'(i)b(i,u,v)[g(v) - g(u)] \right\}. \end{aligned} \tag{22}$$

Similar to (13) and (12), we have $\sum_{u=1}^{M} b(i,u,v) = p'(i,v)$, and $\sum_{v=1}^{M} b(i,u,v) = p(i,u)$. (22) becomes

$$\begin{aligned} &E(F'_L - F_L) \\ &\quad \approx \sum_{i=1}^{M} \left\{ L\pi'(i) \left\{ \sum_{j=1}^{M} [p'(i,j) - p(i,j)]g(j) \right\} \right\} \\ &\quad = L\pi'[P' - P]g = L\pi' Qg. \end{aligned} \tag{23}$$

Finally, we have

$$\eta' - \eta = \lim_{L \to \infty} \frac{1}{L} E(F'_L - F_L) = \pi' Qg. \tag{24}$$

IV. Markov Chains With Different State Spaces

The construction approach can be applied flexibly to other general problems. In this section, we apply this principle to two special problems to illustrate its flexibility.

A. *One is a Subspace of the Other*

Now, we construct the performance difference between two systems defined on different state spaces, with one as a subspace of the other. Let $\mathcal{S} = \{1, 2, \dots, M\}$ and $\mathcal{S}' = \{1, 2, \dots, M'\}, M < M'$ be such two spaces, with $\mathcal{S} \subset \mathcal{S}'$. An example is the $\mathrm{M/M/1/N}$ and the $\mathrm{M/M/1/N+1}$ queues. Let $P\ (M \times M)$ and $P'(M' \times M')$ be the (irreducible) transition probability matrices for the two Markov chains. We decompose P' into

$$P' = \begin{bmatrix} P_1 & P_{12} \\ P_{21} & P_2 \end{bmatrix} \tag{25}$$

where P_1 is an $M \times M$ matrix corresponding to $\mathcal{S}$. Let $f^T = (f(1), \dots, f(M))$ and $f'^T = (f(1), \dots, f(M), f(M+1), \dots, f(M'))$.

Fig. 6 illustrates the sample paths, in which Path 1 (A-B-W-C) is viewed as $\mathbf{X}$, and Path 3 (A-B-G-E-H-J-F), $\mathbf{X}'$. For any segment in which $\mathbf{X}'$ lies in $\mathcal{S}$, the situation is the same as the case discussed above. For example, at $l = 4$, $\mathbf{X}'$ has a jump from state u_1 to state v_1. If both u_1 and v_1 are in $\mathcal{S}$, then after the jump, $\mathbf{X}'$ may follow the transition matrix P until at $l = 9$ it has another jump from u_2 to v_2. By adding an auxiliary path E-R-J-D that follows P, we have a segment G-E-R-J-D, which follows transition matrix P. Thus, the jump at $l = 4$ from u_1 to $v_1, u_1, v_1 \in \mathcal{S}$, can be treated in the same way as in Section III-B. However, if a jump is from a state in $\mathcal{S}$ to a state outside of $\mathcal{S}$, then after the jump, $\mathbf{X}'$ will follow the sub-matrix $[P_{21}, P_2]$ until it reaches $\mathcal{S}$ again. For example, in Fig. 6 there is a jump from $u_2 \in \mathcal{S}$ to $v_2 \in \mathcal{S}' - \mathcal{S}$ at $l = 9$, and after the jump $\mathbf{X}'$ follows $[P_{21}, P_2]$ until at point K it reaches $\mathcal{S}$ again. (More precisely, $\mathbf{X}'$ follows $[P_2]$ until at $l = 10$ it transits into $\mathcal{S}$ at K following $[P_{21}]$.) Fig. 6 illustrates that there is no jump on $\mathbf{X}'$ until it merges with path 2 at J. If after K there is another jump on $\mathbf{X}'$ before it merges with path 2, we add an auxiliary path and denote it as $K - J$. In both cases, the effect of the jump from u_2 to v_2 can be measured by the difference between the two segments H-J and R-J; R-J follows P, while the first part of H-J, H-K, follows $[P_{21}, P_2]$, and the last part, K-J, follows P. That is, H-J follows the following transition matrix:

$$\tilde{P} = \begin{bmatrix} P & 0 \\ P_{21} & P_2 \end{bmatrix}. \tag{26}$$

Pictorially, $\mathbf{X}'$ in Fig. 6 follows P on the original path $\mathbf{X}$ until point B, at which it jumps to point G according to P' then follows P again on the "large circles" path to point E, at which it jumps to point H according to P' then follows $\tilde{P}$ on the "small circles" to point J, and so on. Since P is a closed submatrix of $\tilde{P}$, following P is the same as following $\tilde{P}$ in a large state space.

From the previous discussion, the effect of a jump from u to $v, u \in \mathcal{S}, v \in \mathcal{S}'$, on the average-cost performance can be measured by the difference of the two segments following the transition matrix $\tilde{P}$. When $u, v \in \mathcal{S}$, the two paths (using auxiliary path if necessary, e.g., W-C and G-E-R-J-D) follow P, and when $u \in \mathcal{S}$ and $v \in \mathcal{S}' - \mathcal{S}$, one path (e.g., R-J) follows P

and the other follows $\tilde{P}$ (e.g., H-K follows $[P_{21}, P_2]$ and K-J follows P). Note that no jump can occur in H-K, which follows $[P_{21}, P_2]$ (no jumps occur from $u \in \mathcal{S}' - \mathcal{S}$).

Now, we study the potential of $\tilde{P}$. Let $\tilde{D}$ $(M' \times M')$ be its realization matrix, then

$$\tilde{D} - \tilde{P}\tilde{D}\tilde{P}^T = F' \tag{27}$$

where $F' = e_{M'} f'^T - f' e_{M'}^T$. We have $\tilde{D}^T = -\tilde{D}, \tilde{D} = e_{M'}\tilde{g}^T - \tilde{g} e_{M'}^T$, and $\tilde{g}$ is the potential satisfying the Poisson equation

$$(I - \tilde{P})\tilde{g} + \tilde{\eta} e_{M'} = f'. \tag{28}$$

Later, we will see that the solutions to (27) and (28) exist. In the Markov chain with transition matrix $\tilde{P}$, all recurrent states are in $\mathcal{S}$. Thus its steady-state probability is

$$\tilde{\pi} = (\pi(1), \ldots, \pi(M), 0, \ldots, 0) \tag{29}$$

or in a vector form $\tilde{\pi} = (\pi, 0)$, with $\pi = (\pi(1), \ldots, \pi(M))$ being the steady-state probability corresponding to P. We have $\tilde{\pi} e_{M'} = \pi e_M = 1, \tilde{\pi} f' = \pi f = \eta$, and $\tilde{\pi}(I - \tilde{P}) = 0$. Left-multiplying both sides of (28) with $\tilde{\pi}$, we get

$$\tilde{\eta} = \tilde{\pi} f' = \pi f = \eta. \tag{30}$$

Recall that in (28) $\tilde{g}$ is determined only up to an additive constant.We may set

$$\tilde{\pi}\tilde{g} = \tilde{\eta}. \tag{31}$$

Denote

$$\tilde{g}^T = \left(g^T, g_2^T\right) \tag{32}$$

where g is an M-dimensional vector. (31) becomes $\pi g = \eta$. Denote $\tilde{D}$ as

$$\tilde{D} = \begin{bmatrix} D & D_{12} \\ D_{21} & D_2 \end{bmatrix} \tag{33}$$

with D being an $M \times M$ matrix. Putting (26) and (33) into (27), we get (34)–(36)

$$D - PDP^T = F \tag{34}$$

where $F = e_M f^T - f e_M^T$, which shows that the up-left submatrix of $\tilde{D}$ is the same as the realization factor matrix for the Markov chain with transition matrix P; and

$$D_{12} - \left(PDP_{21}^T + PD_{12}P_2^T\right) = F_{12} \tag{35}$$

where $F_{12} = e_M f_2^T - f e_{M'-M}^T$ is an $M \times (M' - M))$ matrix, $f_2^T = (f(M+1), \ldots, f(M'))$ is an $(M' - M)$ dimensional vector, $f'^T = (f^T, f_2^T)$; and

$$D_2 - P_2 D_2 P_2^T = F_2 + (P_{21}D + P_2 D_{21})P_{21}^T + P_{21}D_{12}P_2^T \tag{36}$$

where $F_2 = e_{M'-M} f_2^T - f_2 e_{M'-M}^T$.

From (34), we have $D = e_M g^T - g e_M^T$ and furthermore, from (32) we have

$$D_{12} = e_M g_2^T - g e_{M'-M}^T.$$

Substituting the previous equation into (35) and using $Pe_M = e_M, P_2 e_{M'-M} + P_{21} e_M = e_{M'-M}$, we get

$$\begin{aligned} e_M g_2^T - e_M g_2^T P_2^T = e_M f_2^T - f e_{M'-M}^T + g e_{M'-M}^T \\ + e_M g^T P_{21}^T - P g e_{M'-M}^T. \end{aligned}$$

Left-multiplying this equation with π and using $\pi e_M = 1, \pi P = \pi$, we obtain

$$g_2 - P_2 g_2 = f_2 - \eta e_{M'-M} + P_{21} g \tag{37}$$

or (the inverse $(1 - P_2)^{-1}$ exists for unichains, see [28])

$$g_2 = (I - P_2)^{-1}\{f_2 - \eta e_{M'-M} + P_{21} g\}. \tag{38}$$

Let $D_2 = e_{M'-M} g_2^T - g_2 e_{M'-M}^T$. Substituting this into (36), we can verify that (37) is indeed a solution to (36). We also conclude that solutions to (27) and (28) indeed exist.

After determining the effect of one jump on the performance, $\tilde{d}(i,j)$ and $\tilde{g}(j), i,j = 1, \ldots, M$, the next step is to determine the total effect of all the jumps caused by the changes in the transition probability matrix as well as the state space. Consider a sample path of the Markov chain with transition probability matrix P' for L states $\{X'_1, \ldots, X'_L\}, L \gg 1$. Recall that $b(i,u,v)$ is the probability that after visiting state i the chain jumps from u to v. We can follow the same procedure as described in Section III-B with only one exception: there would be jumps only when the system is in $\mathcal{S}$ (There is no jump in between H and K Fig. 6). Therefore, corresponding to (22), we have

$$E(F'_L - F_L) \approx \sum_{i=1}^{M} \left\{ \sum_{u=1}^{M} \sum_{v=1}^{M'} [L\pi'(i) b(i,u,v) \tilde{d}(u,v)] \right\} \tag{39}$$

where $\tilde{d}(u,v) = \tilde{g}(v) - \tilde{g}(u)$ is the (u,v)th component of $\tilde{D}$. We have $\sum_{u=1}^{M} b(i,u,v) = p'(i,v)$ and $\sum_{v=1}^{M'} b(i,u,v) = p(i,u)$, for $i = 1, \ldots, M$. Thus

$$\begin{aligned} E(F'_L - F_L) \approx \sum_{i=1}^{M} \Bigg\{ L\pi'(i) \Bigg\{ \sum_{v=1}^{M'} [p'(i,v)\tilde{g}(v)] \Bigg\} \\ - \Bigg\{ \sum_{u=1}^{M} [p(i,u)\tilde{g}(u)] \Bigg\} \Bigg\}. \end{aligned}$$

Setting $p(i,u) = 0$ for $i = 1, \ldots, M, u = M+1, \ldots, M'$, we have

$$\begin{aligned} E(F'_L - F_L) \approx \sum_{i=1}^{M} \Bigg\{ L\pi'(i) \Bigg\{ \sum_{v=1}^{M'} [p'(i,v)\tilde{g}(v)] \Bigg\} \\ - \Bigg\{ \sum_{u=1}^{M'} [p(i,u)\tilde{g}(u)] \Bigg\} \Bigg\} \end{aligned}$$

Finally, we have

$$\eta' - \eta = \lim_{L \to \infty} \frac{1}{L} E(F'_L - F_L) = \pi'^{*} Q^{*} \tilde{g} \tag{40}$$

where $\pi'^{*} = (\pi'(1), \ldots, \pi'(M))$ and

$$Q^{*} = [P_1, P_{12}] - [P, 0]$$

where “0” denotes an $M \times (M' - M)$ matrix in which all components are zero.

The intuitively obtained (40) can be easily verify. Left-multiplying (28) by π' and using $\pi' e_{M'} = 1, \tilde{\eta} = \eta, \pi' P' = \pi'$, we have

$$\eta' - \eta = \pi' f' - \tilde{\eta} = \pi'(I - \tilde{P})\tilde{g} = \pi'(P' - \tilde{P})\tilde{g} = \pi'^{*} Q^{*} \tilde{g}.$$

For performance sensitivity, we define

$$\begin{aligned} P^{\delta} &= \tilde{P} + \delta[P' - \tilde{P}] \\ &= \begin{bmatrix} P + \delta(P_1 - P) & \delta P_{12} \\ P_{21} & P_2 \end{bmatrix}, \qquad 0 \le \delta \le 1. \end{aligned}$$

Thus, $P^{\delta}|_{\delta=1} = P'$ and $P^{\delta}|_{\delta=0} = \tilde{P}$, which has the same steady-state performance as P. Superscript δ is added to quantities associated with Markov chain P^{δ}. Applying (40) to P^{δ} and P, we obtain $\eta^{\delta} - \eta = \pi^{*\delta} Q^{*} \delta \tilde{g}$. Letting $\delta \to 0$, we get

$$\frac{d\eta}{d\delta} = \pi Q^{*} \tilde{g}. \qquad (41)$$

In the previous analysis, we view the sample path associated with transition matrix P in the smaller state–space $\mathcal{S}$ as the original one and that of the Markov chain with P' in the larger state–space $\mathcal{S}'$ as the perturbed one. The role of the two sample paths can be reversed, i.e., we may view the sample path with P' as the original one and that with P as the perturbed one. In this way, we follow the perturbed path and observe the jumps from states in $\mathcal{S}$ to states in $\mathcal{S}'$. The realization factor matrix $D' = [d'(i,j)] = e_{M'} g'^{T} - g' e_{M'}^{T}$ is an $(M' \times M')$ matrix satisfying

$$D' - P' D P'^{T} = F'$$

and g' satisfies

$$(I - P')g' + \eta' e = f'. \qquad (42)$$

Again, consider a sample path of P for L transitions. Among them, on the average $L\pi(i)$ are from state $i, i \in \mathcal{S}$. After it, $b(i,u,v)$ will jump from $u \in \mathcal{S}'$ to $v \in \mathcal{S}$, etc. Following the same reasoning as we did before, we eventually obtain

$$\eta - \eta' = \pi Q'^{*} g' \qquad (43)$$

where $Q'^{*} = [P, 0] - [P_1, P_{12}] = -Q^{*}$. (43) can be verified simply by left-multiplying both sides of (42) by $\tilde{\pi} = (\pi(1), \pi(2), \ldots, \pi(M), 0, \ldots, 0)$.

To study performance sensitivity, we use $\tilde{P}$ defined in (26) again. Set

$$\begin{aligned} P^{\delta} &= P' + \delta(\tilde{P} - P') \\ &= \begin{bmatrix} P_1 + \delta(P - P_1) & (1-\delta)P_{12} \\ P_{21} & P_2 \end{bmatrix}, \qquad 0 \le \delta \le 1 \end{aligned}$$

with $P^{\delta}|_{\delta=1} = \tilde{P}$, which has the same steady-state performance η as P, and $P^{\delta}|_{\delta=0} = P'$. From (43), we have $\eta^{\delta} - \eta' = \pi^{\delta} Q'^{*} \delta g'$. Therefore

$$\left.\frac{d\eta}{d\delta}\right|_{at\, \eta'} = \pi'^{*} Q'^{*} g' \qquad (44)$$

where $\pi'^{*} = (\pi'(1), \ldots, \pi'(M))$.

Both g' and π' in (44) can be estimated with a single sample path of the Markov chain P'. Thus, when the original state–space is larger, the performance derivative from a large state space to a small state–space can be determined based on a sample path of the original Markov chain. However, in (41), $\tilde{g}$ is determined by $\tilde{P}$ in (26), which depends $[P_{21}, P_2]$. Therefore, for performance derivatives from a small state space to a large state space, additional information is needed besides a sample path from the original Markov chain.

B. General Case

In this section, we study the case where two state spaces have a common subspace. Consider two Markov chains $\mathbf{X}$ and $\mathbf{X}'$. $\mathbf{X}$ is defined on $\mathcal{S} = \{M, M-1, \ldots, 2, 1\}$; both Markov chains have a common subspace denoted as $\mathcal{S}_0 = \{m, \ldots, 2, 1\}, m < M$; and $\mathbf{X}'$ is defined on $\mathcal{S}' = \{m, \ldots, 1, 0, -1, \ldots, -n+1, -n\}$, which has $M' = n + m + 1$ states. Denote $\mathcal{S}_1 = \{M, M-1, \ldots, m+1\}$ and $\mathcal{S}_{-1} = \{0, -1, \ldots, -n\}$. We have $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_0$ and $\mathcal{S}' = \mathcal{S}_0 \cup \mathcal{S}_{-1}$. Let $\tilde{\mathcal{S}} = \{M, \ldots, 1, 0, -1, \ldots, -n+1, -n\}$, which has $\tilde{M} = M + n + 1$ states.

Let P and P' be the transition probability matrices of the two Markov chains. Let $f^{T} = (f(M), \ldots, f(1)))$ and $f'^{T} = (f(m), \ldots, f(1), f(0), f(-1), \ldots, f(-n))$ be the performance vectors, and π, η, and π', η' be the steady-state probabilities and average-cost performance of the two chains. Assume both are irreducible. We decompose P and P' into

$$P = \begin{bmatrix} P_1 & P_{10} \\ P_{01} & P_0 \end{bmatrix} \qquad (45)$$

and

$$P' = \begin{bmatrix} P'_0 & P'_{0-1} \\ P'_{-10} & P'_{-1} \end{bmatrix} \qquad (46)$$

where $[P_1, P_{10}]$ in P corresponds to $\mathcal{S}_1$, $[P_{01}, P_0]$ in P and $[P'_0, P'_{0-1}]$ in P' correspond to subspace $\mathcal{S}_0$, and $[P'_{-10}, P'_{-1}]$ to $\mathcal{S}_{-1}$.

Without loss of generality, we assume that both $\mathbf{X}$ and $\mathbf{X}'$ start from the same state in $\mathcal{S}_0$. Following the same argument as in Section IV-A, we use Fig. 6 to construct the performance difference, with Path 1 (A-B-W-C) as $\mathbf{X}$ and Path 3 (A-B-G-E-H-J-F) as $\mathbf{X}'$. We can see that using auxiliary paths if necessary, the paths that determine the realization factors follow the transition matrix on $\tilde{\mathcal{S}}$ [cf. (26)]

$$\tilde{P} = \begin{bmatrix} P_1 & P_{10} & 0 \\ P_{01} & P_0 & 0 \\ 0 & P'_{-10} & P'_{-1} \end{bmatrix}. \qquad (47)$$

Similar to (27), we have

$$\tilde{D} - \tilde{P} \tilde{D} \tilde{P}^{T} = \tilde{F} \qquad (48)$$

where $\tilde{D}$ is an $\tilde{M} \times \tilde{M}$ realization matrix, $\tilde{F} = e_{\tilde{M}} \tilde{f}^{T} - \tilde{f} e_{\tilde{M}}^{T}$, and

$$\tilde{f} = (f(M), \ldots, f(1), f(0), \ldots, f(-n))^{T} = (f, f_{-1})$$

with $f_{-1} = (f(0), f(-1), \ldots, f(-n))$.

We have $\tilde{D}^{T} = -\tilde{D}, \tilde{D} = e_{\tilde{M}} \tilde{g}^{T} - \tilde{g} e_{\tilde{M}}^{T}$, and $\tilde{g}$ is the potential satisfying

$$(I - \tilde{P})\tilde{g} + \tilde{\eta} e_{\tilde{M}} = \tilde{f}. \qquad (49)$$

Now, we consider a sample path of the Markov chain with transition probability matrix P' for L states $\{X'_1, \dots, X'_L\}, L \gg 1$. Similar to (39), we have

$$E(F'_L - F_L) = \sum_{i \in \mathcal{S}_0} \left\{ \sum_{u \in \mathcal{S}} \sum_{v \in \mathcal{S}'} [L\pi'(i) b(i,u,v) \tilde{d}(u,v)] \right\}.$$

We have $\sum_{u \in \mathcal{S}} b(i,u,v) = p'(i,v)$ and $\sum_{v \in \mathcal{S}'} b(i,u,v) = p(i,u)$. Thus

$$E(F'_L - F_L) \approx \sum_{i \in \mathcal{S}_0} \left\{ L\pi'(i) \left\{ \sum_{v \in \mathcal{S}'} [p'(i,v)\tilde{g}(v)] \right\} - \left\{ \sum_{u \in \mathcal{S}} [p(i,u)\tilde{g}(u)] \right\} \right\}.$$

Setting $p(i,u) = 0$ for $i \in \mathcal{S}_0, u \in \mathcal{S}_{-1}$, and $p'(i,v) = 0$ for $i \in \mathcal{S}_0, v \in \mathcal{S}_1$, we have

$$E(F'_L - F_L) \approx \sum_{i \in \mathcal{S}_0} \left\{ L\pi'(i) \left\{ \sum_{v \in \tilde{\mathcal{S}}} [p'(i,v)\tilde{g}(v)] \right\} - \left\{ \sum_{u \in \tilde{\mathcal{S}}} [p(i,u)\tilde{g}(u)] \right\} \right\}.$$

Finally, we have

$$\eta' - \eta = \lim_{L \to \infty} \frac{1}{L} E(F'_L - F_L) = \pi'^* Q^* \tilde{g} \tag{50}$$

where $\pi'^* = (\pi'(m), \dots, \pi'(1))$ and

$$Q^* = [0, P'_0, P'_{0-1}] - [P_{01}, P_0, 0].$$

Now, we consider the performance sensitivity. To this end, we define

$$\tilde{P}^\delta = \begin{bmatrix} P_1 & P_{10} & 0 \\ (1-\delta)P_{01} & P_0 + \delta(P'_0 - P_0) & \delta P_{0-1} \\ 0 & P'_{-10} & P'_{-1} \end{bmatrix}.$$

Thus, $\tilde{P}^\delta|_{\delta=0} = \tilde{P}$, which has the same steady-state performance as P, and $\tilde{P}^\delta|_{\delta=1}$ has the same steady-state performance as P'. We have $\eta^\delta - \eta = \pi^{\delta *} Q^* \delta \tilde{g}$. Letting $\delta \to 0$, we get

$$\frac{d\eta}{d\delta} = \pi^* Q^* \tilde{g}$$

with $\pi^* = (\pi(m), \dots, \pi(1))$.

We have constructed performance sensitivities with intuitions. The results need to be rigorously proved. First, we define an $M + n + 1$-dimensional row vector

$$\tilde{\pi} = (\pi(M), \dots, \pi(1), 0, \dots, 0).$$

The nonzero part is π with $\pi P = \pi$. Thus, $\tilde{\pi}\tilde{P} = \tilde{\pi}$. Left-multiplying both sides of (49) with $\tilde{\pi}$ and noting $\tilde{\pi} e_{\tilde{M}} = 1$, we get

$$\tilde{\eta} = \tilde{\pi}\tilde{f} = \pi f = \eta.$$

Let $\tilde{\pi}' = (0, \pi')$ be an $M + n + 1$ dimensional row vector with "0" denoting an $M - m$-dimensional row vector with all components being zeros. We have $\tilde{\pi}'\tilde{f} = \pi' f' = \eta'$. Left-multiplying both sides of (49) with $\tilde{\pi}'$, we get

$$\eta' - \eta = \tilde{\pi}'(I - \tilde{P})\tilde{g}$$

in which

$$\tilde{\pi}'(I - \tilde{P}) = (0, \pi') \left\{ \begin{bmatrix} 0 & 0 & 0 \\ 0 & P'_0 & P'_{0-1} \\ 0 & P'_{-10} & P'_{-1} \end{bmatrix} - \begin{bmatrix} P_1 & P_{10} & 0 \\ P_{01} & P_0 & 0 \\ 0 & P'_{-10} & P'_{-1} \end{bmatrix} \right\}.$$

It is then clear that $\tilde{\pi}'(I - \tilde{P}) = \pi'^* Q^*$, and (50) is proved.

Equations (29)–(38) hold with some minor modifications. For clarity, we repeat these equations with only notation changes to fit the general case. We rewrite (47) as

$$\tilde{P} = \begin{bmatrix} P & 0 \\ P'_{-1*} & P'_{-1} \end{bmatrix} \tag{51}$$

where $P'_{-1*} = [0, P'_{-10}]$. Next, we denote

$$\tilde{g}^T = (g^T, g^T_{-1}) = (g^T_1, g^T_0, g^T_{-1}) \tag{52}$$

where g is defined on $\mathcal{S}, g_1, g_0$, and g_{-1} on $\mathcal{S}_1, \mathcal{S}_0$, and $\mathcal{S}_{-1}$, respectively. Denote $\tilde{D}$ as

$$\tilde{D} = \begin{bmatrix} D & D_{*-1} \\ D_{-1*} & D_{-1} \end{bmatrix} \tag{53}$$

with D being an $M \times M$ matrix corresponding to $\mathcal{S}$. Putting (51) and (53) into (48), we get

$$D - PDP^T = F$$

where $F = e_M f^T - f e^T_M$. This is the same for the Markov chain with transition matrix P; thus g in (52) satisfies the Poisson equation

$$(I - P)g + \eta e_M = f.$$

Finally, we can obtain [cf. (38)]

$$\begin{aligned} g_{-1} &= (I - P'_{-1})^{-1}\{f_{-1} - \eta e_{n+1} + P'_{-1*} g\} \\ &= (I - P'_{-1})^{-1}\{f_{-1} - \eta e_{n+1} + P'_{-10} g_0\}. \end{aligned}$$

V. Parameterized and Partially Observable Systems

It is well known that the standard Markov model and policy iteration approach may encounter some difficulties: the dimension of the problem is usually too large, the transition matrix P may not be explicitly known, and the actions at different states may not be independent. In this section, we show that our construction approach can be applied in a more flexible way to some special problems to obtain sensitivity equations; based on these equations the above mentioned difficulties may be overcome or alleviated. In particular, in parameterized systems, the construction approach can be applied to a part of the system that is affected by the changes of the values of the parameters; no information about the other part of the system is needed. More precisely, if a parameter change only affects the transition probabilities $p(k,i)$ and $p(k,j)$, then only the potentials $g(i)$ and $g(j)$ have to be estimated. In the same spirit, the approach can be applied to systems in which only a part of the states are observable (partially observable); in this case, the potentials are aggregated. In both cases, the actions at different states may be co-related and we may obtain performance sensitivities without estimating potentials for all the states and without knowing the

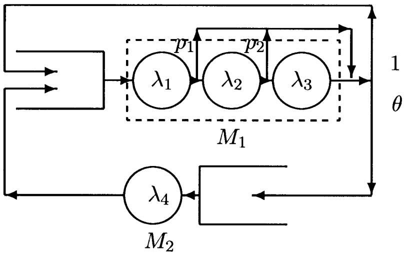

Fig. 7. Manufacturing system.

transition probability matrix. In addition, sample-path-based algorithms can be developed for performance sensitivities. The ideas are illustrated by two examples.

A. Parameterized Systems: An Example

We consider a manufacturing system consisting of two machines and N parts circulating between the two machines, as shown in Fig. 7. Machine 1 (M_1) can perform three operations (1–3); their service times are exponentially distributed with rates λ_1, λ_2, and λ_3, respectively. Some parts only require operation 1, some require operations 1 and 2, and others need to go through all three operations in the sequence of 1–3. The probabilities that a part belongs to these three types are $p_1, (1-p_1)p_2, (1-p_1)(1-p_2)$, respectively, as shown in Fig. 7. Machine 1 works on only one part at a time. Machine 2 (M_2) has only one operation; its service time is exponential with rate λ_4. Machine 1 can also be viewed as having a coxian distributed service time [18].

The system can be modeled as a Markov process with states denoted as (n, i), where n is the number of parts at M_1 and $i = 1, 2$, or 3 denotes the operation that M_1 is performing. The state–space is $\mathcal{S} = \{(n, i)\} \cup \{0\}, n = 1, \ldots, N, i = 1, 2, 3$. A part after completion of service at M_1 goes to M_2 with probability $\theta(n) \in [0, 1]$ (assumed to be independent of i), or immediately returns to M_1 with probability $1 - \theta(n)$. Let f be the performance function.

We can use uniformization to convert this model to a discrete-time Markov chain so that we can apply the results in Section II. (A parallel theory can be developed for continuous-time Markov processes; see [13].) The transition probability matrix of this Markov chain can be easily derived by using $\lambda_i, i = 1, 2, 3, 4$, and θ. We, however, will not do so because its explicit form is not needed in our approach.

Following the same procedure as in Section III, we consider a sample path with $L \gg 1$ transitions. Let $p(n)$ be the probability that a transition is due to a service completion of M_1 and there are n customers in it. Now, suppose that $\theta(n)$ changes to $\theta(n) + \delta_n, \delta_n > 0$, for a particular n. This change may cause a state jump from $(n, 1)$ to $(n-1, 1)$. The probability of such a jump is δ_n and its average effect is measured by the realization factor $d[(n, 1), (n-1, 1)]$. Similar to (15), the average number of transitions corresponding to M_1's service completion time is $Lp(n)$, the average number of jumps after such service completions is $Lp(n)\delta_n$. Thus, we have

$$E\left(F_L^\delta - F_L\right) \approx Lp(n)\delta_n d[(n, 1), (n-1, 1)].$$

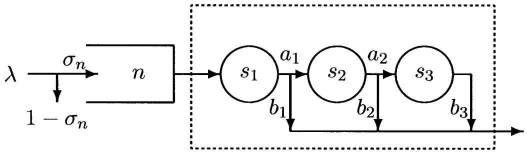

Fig. 8. M/G/1/N queue.

Therefore

$$\begin{aligned}\frac{d\eta}{d\theta(n)} &= p(n)d[(n, 1), (n-1, 1)] \\ &= p(n)[g(n-1, 1) - g(n, 1)].\end{aligned}$$

Both $p(n)$ and $d[(n, 1), (n-1, 1)]$ can be directly estimated on a sample path without knowing P and π. From (5), $d[(n, 1), (n-1, 1)]$ can be estimated by averaging the sum of $f(X_l)$ over the periods from state $(n, 1)$ to state $(n-1, 1)$. Thus, to obtain $(d\eta/d\theta(n))$ for a particular n, we need to estimate only $d[(n, 1), (n-1, 1)]$. We can obtain $(d\eta/d\theta(n))$ for all n if we estimate $g(n, 1)$ for all $n = 0, 1, \ldots, N$. These derivatives can be used in various optimization schemes (even with constrains). For example, if $\theta(n)$ changes to $\theta(n) + \alpha_n\delta$ for all n with a set of fixed α_n, then the performance derivative is

$$\begin{aligned}\frac{d\eta}{d\delta} &= \sum_{n=1}^{N} \alpha_n p(n) d[(n, 1), (n-1, 1)] \\ &= \sum_{n=1}^{N} \alpha_n p(n)[g(n-1, 1) - g(n, 1)].\end{aligned}$$

Now, suppose we have another system working under parameters $\theta'(n)$ $n = 1, 2 \ldots, N$. Following the same procedure as in Section III-B, we obtain

$$\begin{aligned}\eta' - \eta &= \sum_{n=1}^{N} p'(n)[\theta'(n) - \theta(n)]d[(n, 1), (n-1, 1)] \\ &= \sum_{n=1}^{N} p'(n)[\theta'(n) - \theta(n)][g(n-1, 1) - g(n, 1)].\end{aligned} \tag{54}$$

This example shows that by our performance sensitivity construction method, we may obtain the sensitivities by analyzing a sample path without knowing the transition probability P and even without estimating $g(i)$ for all the states. In this problem, the same action $\theta(n)$ applies to different state $(n, 1), (n, 2)$, and $(n, 3)$; the standard MDP formulation does not apply.

B. Partially Observable Systems: The M/G/1/N Queue

In this section, we show that in systems where only a part of the states is observable, we can aggregate the potentials on a set of states that share the same observable part of states and obtain the performance gradients. The aggregated potential can be estimated on a single sample path.

Consider an M/G/1/N queue, in which the service time distribution is a Coxian distribution consisting of K stages, each of them is exponentially distributed with mean $s_k, k = 1, \ldots, K$ ($K = 3$ in Fig. 8). For simplicity and without loss of generality, we assume $s_k \equiv s$ for all $k = 1, 2, \ldots, K$. Let $\mu =$

$(1/s)$. After receiving service at stage k, a customer enters stage $k+1$ with probability a_k, and leaves the station with probability $b_k = 1 - a_k, k = 1, 2, \ldots, K$, and $b_K = 1$. Let n be the number of customers in the system, and N be the buffer size: when an arriving customer finds $n = N$, the customer is lost. The arrival process is Poisson with rate λ. An arriving customer enters the queue with probability σ_n (depending on n), and is rejected by the system with probability $1 - \sigma_n$ (the admission control problem). This is equivalent to a load-dependent arrival rate $\lambda_n = \lambda\sigma_n$. The state space of the system is $\{0, (n,k), n = 1, 2, \ldots, N; k = 1, \ldots, K\}$, with k denoting the stage of the customer being served.

We assume that σ_n can be taken from a finite set, and let the cost function be $f(n, \sigma_n)$ (independent of the stage k). Then, σ_n can be viewed as the action taken when the state is (n,k). However, the stage k in a real system is not observable, i.e., the controller only knows the number of customers, n, in the system, but does not know the stage k. This problem becomes an MDP with partially observable states (POMDP). The actions can only depend on n, or its history.

It is well-known that POMDP is hard to solve. A sensitivity-based approach may be feasible for performance optimization. Thus, we shall construct the performance sensitivity with respect to σ_n using realization factors. To illustrate the idea, we set $\lambda = \mu$ (thus, $\mu > \lambda\sigma_n$ for all n). The system parameters are $\sigma_n, n = 0, 1, \ldots, N-1, (\sigma_N = 0)$, and λ. Let the discrete-time Markov chain be denoted as $\mathbf{X} = \{X_0, X_1, \ldots, X_l, \ldots\}$.

We observe the system for $L \gg 1$ transitions and obtain a sample path $\{X_0, X_1, \ldots, X_L\}$. Similar to the example in Section V-A, let $p(n)$ be the probability that a transition in the Markov chain is due to a customer arrival when there are n customers in the system

$$p(n) = \frac{L_n}{L}$$

where L_n is the number of transitions in the observation period that belong to the arrival process and the number of customers is n before the arrival point. Thus, $L_n = Lp(n)$, and the number of customers accepted by the system in the observation period when there are n customers in the system is $L_n^+ = Lp(n)\sigma_n$. Now, suppose $\sigma_n, n < N$, changes to $\sigma_n + \Delta\sigma_n$. Then, $Lp(n)\Delta\sigma_n$ is the number of additional arrivals that are admitted to the system when there are n customers in it due to the change $\Delta\sigma_n$. We have

$$p(n) = \sum_{k=1}^{K} p(n,k)$$

where $p(n,k)$ is the probability that a transition in $\mathbf{X}$ is a customer arrival when the state is (n,k). Therefore, $Lp(n,k)\Delta\sigma_n$ is the number of additional arrivals that are admitted to the system when the state is (n,k) due to the change $\Delta\sigma_n$. The average effect of one such additional arrival is $d[(n,k);(n+1,k)]$. Thus, the total effect of all these additional admitted customers on the performance is

$$\Delta F_L \approx \sum_{k=1}^{K} Lp(n,k) d[(n,k);(n+1,k)]\Delta\sigma_n$$

or letting $L \to \infty$

$$\Delta\eta = \frac{\Delta F_L}{L} = \sum_{k=1}^{K} p(n,k) d[(n,k);(n+1,k)]\Delta\sigma_n.$$

Letting $\Delta\sigma_n \to 0$, we get

$$\frac{d\eta}{d\sigma_n} = \sum_{k=1}^{K} p(n,k) d[(n,k);(n+1,k)].$$

Define the conditional probability $p(k\,|\,n) = (p(n,k)/p(n)), \quad k = 1, \ldots, K$. Then

$$\frac{d\eta}{d\sigma_n} = p(n)\sum_{k=1}^{K} p(k\,|\,n) d[(n,k);(n+1,k)]$$
$$= p(n)\bar{d}(n, n+1)$$

where $\bar{d}(n,n+1) = \sum_{k=1}^{K} p(k\,|\,n) d[(n,k);(n+1,k)]$ is the mean of the realization factor in the subset $\mathcal{S}_n = \{(n,k) : \text{all } k\}$ when an arrival customer is accepted. Note

$$d[(n,k);(n+1,k)] = g(n+1,k) - g(n,k).$$

Finally, we have

$$\frac{d\eta}{d\sigma_n} = p(n)\left[\sum_{k=1}^{K} p(k\,|\,n) g(n+1,k) - \sum_{k=1}^{K} p(k\,|\,n) g(n,k)\right]$$
$$= p(n)[g^+(n+1) - g^-(n)]$$

where $g^+(n+1) = \sum_{k=1}^{K} p(k\,|\,n) g(n+1,k)$ is the mean of the potentials in set $\mathcal{S}_{n+1} = \{(n+1,k) : k = 1, \ldots, K\}$ given that n jumps to $n+1$, and $g^-(n) = \sum_{k=1}^{K} p(k\,|\,n) g(n,k)$ is the mean of the potentials in set $\mathcal{S}_n$ given that n remains the same.

Finally, $p(n), g^+(n+1)$, and $g^-(n)$ can be estimated on a single sample path. For example, $g^+(n)$ can be estimated in the following way. In the observation period $\{X_0, X_1, \ldots, X_L\}$, denote the sequence of arrival points that find n customers in the system as $t_1, \ldots, t_{L_n}$. Among these points, denote those instants at which the system jumps from n to $n+1$ (i.e., the customers are accepted at these points) as $t_{v_1}, t_{v_2}, \ldots, t_{v_{L_n^+}}$. Then at $t_{v_i}, i = 1, 2, \ldots, L_n^+$, there are $n+1$ customers in the system. Choose a large integer D. Set

$$h_{v_i} = \sum_{l=t_{v_i}}^{t_{v_i}+D} [f(X_l) - \eta].$$

Now, we partition the set $\mathcal{T} = \{t_{v_i}, i = 1, 2, \ldots, L_n^+\}$ into K subsets $\mathcal{T} = \cup_{k=1}^{K}\mathcal{T}_k$, such that if $t_{v_i} \in \mathcal{T}_k$ then $X_{t_{v_i}} = (n+1,k)$. Let $L_{n,k}$ be the number of instants in $\mathcal{T}_k$. Then we have

$$\frac{1}{L_n^+}\sum_{i=1}^{L_n^+} h_{v_i} = \frac{1}{L_n^+}\sum_{k=1}^{K}\sum_{t_{v_i}\in\mathcal{T}_k} h_{v_i}$$
$$= \sum_{k=1}^{K}\left\{\frac{L_{n,k}}{L_n^+}\frac{1}{L_{n,k}}\sum_{t_{v_i}\in\mathcal{T}_k} h_{v_i}\right\}.$$

Because

$$\lim_{D\to\infty}\left\{\lim_{L_{n,k}\to\infty}\frac{1}{L_{n,k}}\sum_{t_{v_i}\in\mathcal{T}_k}h_{v_i}\right\}=g(n+1,k)$$

and

$$\lim_{L_n^+\to\infty}\frac{L_{n,k}}{L_n^+}=p(k\,|\,n)$$

we have

$$\lim_{D\to\infty}\left\{\lim_{L_n\to\infty}\frac{1}{L_n^+}\sum_{i=1}^{L_n^+}h_{v_i}\right\}=\sum_{k=1}^{K}p(k\,|\,n)g(n+1,k)=g^+(n+1). \quad (55)$$

Equation (55) shows that $g^+(n+1)$ can be estimated on a sample path in a similar way as estimating $g(n+1)$ [cf. (10) and [13]].

This example shows that potentials can be aggregated together over the set of states with the same observable part of the states; thus, our approach can be used to construct the performance gradients with only the observable information. Again, the exact values of the transition probability matrix is not needed (but the system structure as shown in Fig. 8 is known). In this example, the action σ_n depends on n, not its history. This corresponds to the reactive policy in POMDP literature. Further research is going on to explore the applicability of this approach to more general POMDPs in which the policy depends on the observation history.

VI. Conclusion and Discussion

The novelty of the research in this paper can be summarized in twofold. First, we propose an intuitive approach to construct the sensitivity formulas for Markov systems, including those that do not fit the standard MDP formulation. Second, these sensitivity formulas form the basis for performance optimization; system structure can be utilized in the sensitivity construction and, thus, computation and/or memory spaces may be saved and techniques such as aggregation may be implemented.

Specifically, we show that a sample path of a Markov chain with transition probability matrix P' can be built upon a sample path of a Markov chain with transition probability matrix P together with the segments that can be measured on the average by the performance potentials (see Fig. 5). We refer to this as the potential structure of the sample path. We show that this structure allows us to construct performance sensitivities, both performance derivatives and performances differences, by first principles with sample path-based arguments. Performance potentials, or realization factors, are used as building blocks in the construction. When the two systems under comparison have the same state space, or the original system has a larger state space, the potentials used in the sensitivity formulas can be estimated on a single sample path of the original system. This is in the same spirit as perturbation analysis: one can obtain the performance sensitivity by analyzing only the original system. When the two systems have different state–spaces or the perturbed system has a larger state–space, the potentials can be estimated on sample paths with an enlarged transition probability matrix [see (26) and (47)]. For these systems, efficient methods in estimating potentials based on reinforcement learning should be developed.

The approach is flexible in the sense that it can be applied to many systems including those with partial information, and it only requires to estimate the potentials that are directly related to the changes in parameters. The sensitivity formulas obtained have clear meanings and are not so easy to conceive otherwise.

Since the sensitivity formulas are the basis for performance optimization, the construction approach introduced in this paper opens up a new direction for further research in optimization. Two examples are given to illustrate the idea. Both problems do not fit the standard MDP formulation because the same action has to be applied to a group of different states. In the first example, the number of potentials to be estimated is reduced because of the system structure. The second example is on the reactive policy of a POMDP problem. From the sensitivity formulas, the potentials can be aggregated and the number of aggregated potentials to be estimated is also reduced. Finally, it is worth noting that to utilize the system structure, one has to know the system structure in advance; i.e., one has to know some property about the underlying transition probability matrix.

The policy gradient of POMDP proposed in [1] is based on the standard Markov model with an enlarged state space consisting of two components, the natural state and the internal state. Therefore, the state space usually becomes larger and no system structure is utilized.

The performance derivatives can be used together with stochastic approximation algorithms in performance optimization. When the Markov systems are in the same state space, the policy iteration algorithm in MDPs can be easily derived from our performance difference formulas. It has been shown that policy iteration in fact chooses the policy that has the steepest gradient after randomization [10], [11]. Thus, both the performance gradient and performance difference formulas are the basis for performance optimization. The performance sensitivity formulas obtained in this paper open up some new research directions: can we derive approaches similar to policy iteration for systems with different state spaces or with partial information? If so, how?

Finally, our approach applies to discounted performance criteria as well (cf. [11]). Thus, our approach provides a uniform framework for optimization with both average and discounted performance criteria.

References

[1] D. Aberdeen and J. Baxter, "Scaling internal-state policy-gradient methods for partially observable Markov decision processes," in *Proc. 19th Int. Conf. on Machine Learning*, Sydney, Australia, 2002, pp. 3–10.

[2] E. Altman, *Constrained Markov Decision Processes*. Boca Raton, FL: CRC, 1999.

[3] J. Baxter and P. L. Bartlett, "Infinite-horizon policy-gradient estimation," *J. Art. Intell. Res.*, vol. 15, pp. 319–350, 2001.

[4] J. Baxter, P. L. Bartlett, and L. Weaver, "Experiments with infinite-horizon policy-gradient estimation," *J. Art. Intell. Res.*, vol. 15, pp. 351–381, 2001.

[5] D. P. Bertsekas, *Dynamic Programming and Optimal Control*. Belmont, MA: Athena Scientific, 1995, vol. I, II.

[6] D. P. Bertsekas and T. N. Tsitsiklis, *Neuro-Dynamic Programming*. Belmont, MA: Athena Scientific, 1996.

[7] X. R. Cao, "Convergence of parameter sensitivity estimates in a stochastic experiment," *IEEE Trans. Automat. Contr.*, vol. AC-30, pp. 834–843, July 1985.

[8] ——, *Realization Probabilities: The Dynamics of Queueing Systems*. New York: Springer-Verlag, 1994.

[9] X. R. Cao, X. M. Yuan, and L. Qiu, "A single sample path-based performance sensitivity formula for Markov chains," *IEEE Trans. Automat. Contr.*, vol. 41, pp. 1814–1817, Oct. 1996.

[10] X. R. Cao, "The relation among potentials, perturbation analysis, Markov decision processes, and other topics," *Discrete Event Dyna. Syst.: Theory Applicat.*, vol. 8, pp. 71–87, 1998.

[11] ——, "A unified approach to Markov decision problems and performance sensitivity analysis," *Automatica*, vol. 36, pp. 771–774, 2000.

[12] ——, "From perturbation analysis to Markov decision processes and reinforcement learning," *Discrete Event Dyna. Syst.: Theory Applicat.*, vol. 13, pp. 9–39, 2003.

[13] X. R. Cao and H. F. Chen, "Potentials perturbation realization, and sensitivity analysis of Markov processes," *IEEE Trans. Automat. Contr.*, vol. 42, pp. 1382–1393, Oct. 1997.

[14] X. R. Cao and Y. W. Wan, "Algorithms for sensitivity analysis of Markov systems through potentials and perturbation realization," *IEEE Trans. Control Syst. Technol.*, vol. 6, pp. 482–494, Mar. 1998.

[15] C. G. Cassandras and S. Lafortune, *Introduction to Discrete Event Systems*. Norwell, MA: Kluwer, 1999.

[16] E. K. P. Chong and P. J. Ramadge, "Convergence of recursive optimization algorithms using infinitesimal perturbation analysis estimates," *Discrete Event Dyna. Syst.: Theory Applicat.*, vol. 1, pp. 339–372, 1992.

[17] E. Çinlar, *Introduction to Stochastic Processes*. Upper Saddle River, NJ: Prentice-Hall, 1975.

[18] D. R. Cox, "A use of complex probabilities in the theory of stochastic processes," in *Proc. Cambridge Philosophical Soc.*, vol. 51, 1955, pp. 313–319.

[19] H. T. Fang and X. R. Cao, "Single sample path-based recursive algorithms for Markov decision processes," *IEEE Trans. Automat. Contr.*, 2004, to be published.

[20] M. C. Fu, "Convergence of a stochastic approximation algorithm for the GI/G/1 queue using infinitesimal perturbation analysis," *J. Optim. Theory Applicat.*, vol. 65, pp. 149–160, 1990.

[21] Y. C. Ho and X. R. Cao, *Perturbation Analysis of Discrete-Event Dynamic Systems*. Boston, MA: Kluwer, 1991.

[22] ——, "Perturbation analysis and optimization of queueing networks," *J. Optim. Theory Applicat.*, vol. 40, no. 4, pp. 559–582, 1983.

[23] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, "Planning and acting in partially observable stochastic domains," *Art. Intell.*, vol. 101, no. 1–2, pp. 99–134, May 1998.

[24] V. R. Konda and V. S. Borkar, "Actor-critic like learning algorithms for Markov decision processes," in *SIAM J. Control Optim.*, vol. 38, 1999, pp. 94–123.

[25] V. R. Konda and J. N. Tsitsiklis, "Actor-critic algorithms," *SIAM J. Control Optim.*, vol. 42, no. 4, pp. 1143–1166, 2003.

[26] P. Marbach and T. N. Tsitsiklis, "Simulation-based optimization of Markov reward processes," *IEEE Trans. Automat. Contr.*, vol. 46, pp. 191–209, Feb. 2001.

[27] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*. London, U.K.: Springer-Verlag, 1993.

[28] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. New York: Wiley, 1994.

[29] S. P. Singh, "Reinforcement learning algorithms for average-payoff Markovain decision processes," in *Proc. 12th National Conf. Artificial Intelligence*, 1994, pp. 202–207.

[30] W. D. Smart and L. P. Kaelbling, "Practical reinforcement learning in continuous spaces," in *Proc. 17th Int. Conf. Machine Learning*, vol. 17, 2000, pp. 903–910.

[31] R. S. Sutton, "Learning to predict by the methods of temporal differences," *Mach. Learn.*, vol. 3, pp. 835–846, 1988.

[32] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press, 1998.

[33] J. N. Tsitsiklis and B. Van Roy, "Average cost temporal-difference learning," *Automatica*, vol. 35, pp. 1799–1808, 1999.

[34] F. J. Vazquez-Abad, C. G. Cassandras, and V. Julka, "Centralized and decentralized asynchronous optimization of stochastic discrete event systems," *IEEE Trans. Automat. Contr.*, vol. 43, pp. 631–655, May 1998.

[35] X. R. Cao, Ed., "Special issue on learning, optimization, and decision making in DEDS," in *Discrete Event Dynamic Systems: Theory and Applications*, 2003, vol. 13.

Xi-Ren Cao received the M.S. and Ph.D. degrees from Harvard University, Cambridge, MA, in 1981 and 1984, respectively.

He was a Research Fellow at Harvard University from 1984 to 1986. He then worked as a Principal and Consultant Engineer/Engineering Manager at Digital Equipment Corporation, Maynard, MA, until October 1993. Since then, he has been a Professor at the Hong Kong University of Science and Technology (HKUST), Hong Kong, China. His current research areas include discrete-event dynamic systems, optimization theory, performance analysis of communication systems, and signal processing.

Dr. Cao owns three patents in data- and telecommunications and has published two books in the area of discrete-event dynamic systems. He received the Outstanding Transactions Paper Award from the IEEE Control Systems Society in 1987 and the Outstanding Publication Award from the Institution of Management Science in 1990. He is an Associate Editor-at-Large of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL, and he is/was on Board of Governors of the IEEE Control Systems Society, Associate Editor of a number of international journals, and Chairman of a few technical committees of international professional societies.