Proceedings of the 42nd IEEE
Conference on Decision and Control
Maui, Hawaii USA, December 2003

Constructing Performance Sensitivities of Markov Systems with Potentials as Building Blocks

Xi-Ren Cao*
Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong
eecao@ee.ust.hk

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 sensitivities of Markov systems can be constructed by using performance potentials (or equivalently, perturbation realization factors) as building blocks. We propose an intuitive approach to derive, by first principles, formulas for performance derivatives and performance differences for two Markov chains. These formulas are the basis for performance optimization of discrete event dynamic systems, including perturbation analysis, Markov decision processes, and reinforcement learning.**

I. INTRODUCTION

Recent research indicates that both perturbation analysis (PA) and Markoc decision processe (MDP) can be explained from a performance sensitivity point of view; the fundamental concept for both PA and MDP is the performance potential of a Markov process; both of them can be implemented using a single sample path; and RL, D(λ), neuro-dynamic programming, etc, are sample-path-based efficient ways of estimating the performance potentials and other related quantities (e.g., Q-factors).

There are two types of sensitivities: 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 two systems (e.g., under two different policies). With the Markov model, different systems correspond to different transition probability matrices determined by different policies. Moreover, the parameters in a transition matrix may depend on continuous parameters (e.g., in case of random policies). Performance optimization can be achieved either by using the performance gradients combined with stochastic approximation methods, or by applying policy iteration algorithms in MDP, which can be easily derived from performance difference formulas.

In previous studies [2], it is shown that the performance derivatives can be constructed by using performance potentials as building blocks. In this paper, we will extend this idea to general cases including the performance derivatives and performance difference for two systems. 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 ideas are as follows: Any change in system parameters or even in 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 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 one sample paths. Finally, the performance sensitivity, which reflects 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 above idea, we can derive performance sensitivity formulas by first principles. This approach is common in physics where researchers first formulate and solve problems based on experimental evidence and then prove their results rigorously. In this vein, 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 the formula for performance difference of two Markov chains in the same state space. The results can be extended to the performance derivatives and differences for two Markov chains with one state space being a subspace of the other, and to the case where the two Markov chains have different state spaces with a common subspace. These examples also serve to illustrate the flexibility in applying the "building block" idea to construct performance sensitivities: we need to focus our attention to only the states that are affected by the parameter changes.

The contributions of the paper are as follows. We proposed an intuitive approach to construct, by first principles,

*Supported in part by a grant from Hong Kong UGC. Tel: (852) 2358-7048 Fax: (852) 2358-1485

performance derivatives and difference by using potentials as the fundamental building blocks. This clearly illustrate the physical meaning of potentials 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.

II. PERTURBATION AND PERFORMANCE POTENTIALS

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

$$\pi P = \pi, \qquad \pi e = 1, \tag{1}$$

where $e = (1, 1, \cdots, 1)^T$ 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 when it is needed (e.g., $e \equiv e_M$ in (1)). The performance measure is defined as

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

where $f = (f(1), \cdots, f(M))^T$ (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 DEDS 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 [2]:

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

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 [2]

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

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 $L_i(j) = min\{n \leq 0, X_n = j\}$ be the first passage time to state j. Then

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

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*. $D^T = -D$. From (4), we can prove the Lyapunov equation

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

where $F = ef^T - fe^T$. It was shown that D takes the form

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

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

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

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. From (6), we can prove that g satisfies the Poisson equation

$$(I - P)g + \eta e = f. \tag{7}$$

The solution to (7) is only up to an additive constant; i.e., if g satisfies (7), then for any constant c, $g + ce$ also does. Therefore, there must be one particular solution to (7) (still denoted as g) such that $\pi g = \eta$. For this solution, (7) becomes

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

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*. We have (up to an additive constant)

$$g(i) = \lim_{L\to\infty} E\{\sum_{l=0}^{L-1} [f(X_l) - \eta] | X_0 = i\}, \tag{9}$$

which is finite for ergodic chains. It has been shown that the potential thus defined is an extension of the α-potential from $0 < \alpha < 1$ to $\alpha = 1$.

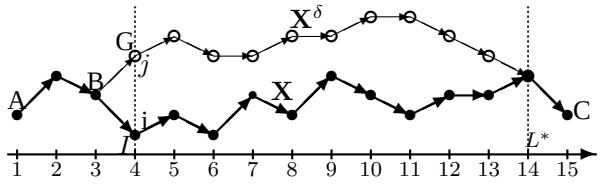

Fig. 1. A Perturbation in a Sample Path of a Markov Chain and Its Effect

III. MARKOV CHAINS ON THE SAME STATE SPACE

In this section, we derive, by applying first principles, performance sensitivities to 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 [2], 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 $\mathcal{S} = \{1, 2, \cdots, M\}$, 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 $0 \leq \delta \leq 1$, 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 since we are concerned about 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 $X_l = k$, $l = 0, 1, \cdots$, we generate a uniformly distributed random variable $\xi_l \in [0, 1)$. If

$$\sum_{n=0}^{i-1} p(k, n) \leq \xi_l < \sum_{n=0}^{i} p(k, n), \tag{10}$$

with $p(k, 0) = 0$, then we set $X_{l+1} = i$. For example, consider the case where $p(k, i) = 0.5$, $p(k, j) = 0.5$, and $p(k, n) = 0$ for all $n \neq i, j$. If $0 \leq \xi < 0.5$, then the Markov chain jumps into state i; otherwise, it jumps into state j. Suppose that the transition probabilities change to $p'(k, i) = 0.5 - \delta$, $p'(k, j) = 0.5 + \delta$, and $p'(k, n) = 0$, if $n \neq i, j$. (i.e., $q(k, i) = -1$, $q(k, j) = 1$, and $q(k, n) = 0$ if $n \neq i, j$.) We use the same sequence of random variable ξ_l to determine the transition of the Markov chain with $P(\delta)$ at time l, $l = 0, 1, \cdots$, 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 i, but $\mathbf{X}^\delta$ transits to j; 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 above discussion indicates that starting from the same initial state X_0 and with the same random sequence ξ_l, $l = 0, 1, \cdots$, the two sample paths $\mathbf{X}^\delta = \{X_0^\delta, X_1^\delta, \cdots, \}$ and $\mathbf{X} = \{X_0, X_1, \cdots, \}$ are also very close. Suppose that with the same values of ξ_n, $n = 0, 1, \cdots, l-2$, we have $X_n^\delta = X_n$, for $n = 0, 1, \cdots, l-1$. Furthermore, we assume that with the same value of ξ_{l-1}, applying (10) to P determines that $\mathbf{X}$ transits to state $X_l = i$, but applying (10) to P^δ determines that $\mathbf{X}^\delta$ transits to state $X_l^\delta = j$. We say that the perturbed chain $\mathbf{X}^\delta$ has a jump (or perturbation) from i to j at time l. In Figure 1, $\mathbf{X}$ and $\mathbf{X}^\delta$ are illustrated by the solid dots and hollow circles, respectively; the perturbed path $\mathbf{X}^\delta$ has a jump from i to j at $l = 4$. After this time, the two sample paths differ until at L^* ($L^* = 14$ in Figure 1) 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 Figure 1) and L^* ($= 14$ in Figure 1) 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 except the one from X_3^δ to X_4^δ look the same as those following transition matrix P. Apparently, $d(i, j)$ measures the average affect of a jump from i to j on F_L in (2).

Now we consider a sample path $\mathbf{X}$ consisting of L, $L >> 1$, 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 such a jump as from state u to state v. (i.e., after visiting i, $\mathbf{X}$ transits to u and however, $\mathbf{X}^\delta$ transits to v). For convenience, we allow $u = v$ as a special case. We shall refer to $u \neq v$ as a "real jump", which happens rarely. Denote the probability of a jump from u to v after visiting state i as $b(i, u, v)$. We have

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

Thus,

$$\sum_{u=1}^{M} b(i, u, v) = p^\delta(i, v), \tag{11}$$

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

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 Figure 2 which illustrates two jumps, one at $l = 4$ and the other at $l = 11$. After the first jump, $\mathbf{X}^\delta$ merges with $\mathbf{X}$ at $l = 7$; thus, the effects of the two jumps shown in Figure 2 can be measured separately. As δ is very small, the probability that there is another jump occurs before $l = 7$ is of order

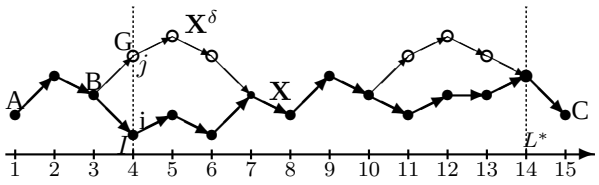

Fig. 2. The Effect of Two Rare Perturbations Are Decoupled

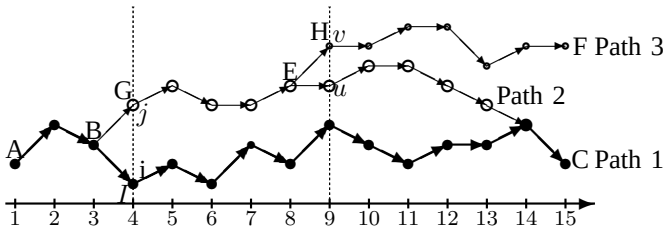

Fig. 3. The Effect of Two Perturbations

δ^2. Thus, on the average the total effect on F_L due to the change in P to $P^\delta = P + \delta Q$ is

$$\begin{aligned} E(F_L^\delta - F_L) &= \sum_{i=1}^{M}\{\sum_{u,v=1}^{M} L\pi(i)b(i,u,v)d(u,v)\} \\ &= \sum_{i=1}^{M}\{\sum_{u,v=1}^{M} L\pi(i)b(i,u,v)[g(v)-g(u)]\} \\ &= \sum_{i=1}^{M}\Bigg\{L\pi(i)\{\sum_{v=1}^{M}[g(v)\sum_{u=1}^{M}b(i,u,v)]\} \\ &\qquad -\{\sum_{u=1}^{M}[g(u)\sum_{v=1}^{M}b(i,u,v)]\}\Bigg\}. \end{aligned} \tag{13}$$

From (11) and (12), (13) becomes

$$\begin{aligned} & E(F_L^\delta - F_L) \\ \approx & \sum_{i=1}^{M}\left\{L\pi(i)\{\sum_{v=1}^{M}[p^\delta(i,v)g(v)]\} - \{\sum_{u=1}^{M}[p(i,u)g(u)]\}\right\} \\ = & \sum_{i=1}^{M}\left\{L\pi(i)\{\sum_{j=1}^{M}[p^\delta(i,j)-p(i,j)]g(j)]\right\} \\ = & L\pi[P^\delta - P]g = L\pi Q\delta g. \end{aligned} \tag{14}$$

Thus,

$$\eta^\delta - \eta = \lim_{L\to\infty}\frac{1}{L}E(F_L' - F_L) = \pi Q\delta g. \tag{15}$$

Finally, we get

$$\frac{d\eta}{d\delta} = \pi Q g. \tag{16}$$

Given P and P', g, D, and π can be estimated on a single sample path; thus, the performance sensitivity along any direction $Q = P' - P$ can be obtained by estimating these quantities on a single sample path. Algorithms can be developed for estimating the performance sensitivity based on a single sample path using (16) without estimating each component of g.

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 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 $\mathcal{S} = \{1, 2, \cdots, M\}$. 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 $P' = P + Q$, ($\delta = 1$ and is not small), two sample paths $\mathbf{X}'$ and $\mathbf{X}$ are completely different and the effect of jumps may be coupled (after a jump on $\mathbf{X}'$, another jump may occur before $\mathbf{X}'$ and $\mathbf{X}$ merge together.)

To see how we can evaluate the effects of the two "coupled" jumps on $\mathbf{X}'$, we follow a sample path of $\mathbf{X}'$. As illustrated in Figure 3, we start the simulation at point A. Again, we use the same random sequence to generate both $\mathbf{X}'$ and $\mathbf{X}$ and the same terminology as for the performance derivative problem: Suppose that with ξ_{l-1} from X'_{l-1} the Markov chain transits to the same state $X'_l = X_l$ 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 $X_l = i$ according to P while it transits to state $X'_l = j$ according to P', we say that the perturbed chain $\mathbf{X}'$ has a jump (or a perturbation) from i to j at time l. Figure 3 shows such a jump at $l = 4$.

Different from the performance derivative case discussed in Section III-A, if a jump happens on $\mathbf{X}'$ at l, then because other jumps may happen before $\mathbf{X}'$ and $\mathbf{X}$ merge together, these two paths are usually completely different afterwards. In Figure 3, $A - B - I - C$ illustrates a sample path for P, and $A - B - G - E - F$, a path for P'. Starting from $l = 4$, these two paths can be generated independently. Now, let us follow the path $G - E - H - F$ in the same way as what we did for $A - B - G - \cdots$: We use the same random sequence to determine whether jump occurs at each step (i.e., whether the Markov chains transits to the same sate according to both P and P'.) Figure 3 indicates that there is no jump at $l = 5, 6, 7, 8$, and there is another jump at $l = 9$ from state u to state v. After the jump, $\mathbf{X}'$ follows the path $H - F$.

To explore the idea, we assume that there is no further jump on $\mathbf{X}'$ after $l = 9$. At $l = 8$, the Markov chain transits to state v according to P' and to state u according to P. After u, we add an auxiliary path that follows the transition matrix P until the auxiliary path merges with $\mathbf{X}$ at $l = 14$. Let us denote the path $A-B-I-C$ as path 1, $A-B-E-C$

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 $G-E-C$ also follows P.

With the help of Figure 3, by using the similar argument as for performance derivatives, we can obtain

$$\eta' - \eta = \lim_{L\to\infty} \frac{1}{L} E(F'_L - F_L) = \pi' Q g. \qquad (17)$$

IV. CONCLUSION AND DISCUSSION

The concept of potential is the same as the "bias" or "differential" in the literature of MDPs. The novelty here is summarized as follows. We show that the difference of the potentials, the perturbation realization factor, measures the effect of a single jump on the performance (A basic idea from PA). We further propose an approach that allows us to construct performance sensitivities, both performance derivatives and performances differences, by first principles with "thought experiments". The thought experiments are based on sample paths. Performance potentials, or realization factors, are used as building blocks in the construction. 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.

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 MDP 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. 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?

V. REFERENCES

[1] X. R. Cao, *Realization Probabilities: The Dynamics of Queueing Systems*, Springer-Verlag, New York, 1994.

[2] X. R. Cao and H. F. Chen, "Potentials, Perturbation Realization, and Sensitivity Analysis of Markov Processes," *IEEE Transactions on AC*, Vol. 42, 1382-1393, 1997.

[3] C. G. Cassandras and S. Lafortune, *Introduction to Discrete Event Systems*, Kluwer Academic Publishers, 1999.

[4] Y. C. Ho and X. R. Cao, *Perturbation Analysis of Discrete-Event Dynamic Systems*, Kluwer Academic Publisher, Boston, 1991.

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