Copyright © 2002 IFAC
15th Triennial World Congress, Barcelona, Spain

SEMI-MARKOV DECISION PROBLEMS AND PERFORMANCE SENSITIVITY ANALYSIS

Xi-Ren Cao [*,1]

[*] *Department of Electrical and Electronic Engineering*
The Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong

Abstract: We extend the results about performance potentials, perturbation realization matrices, policy iteration of Markov decision processes, etc., to semi-Markov processes (SMPs). Starting with the concept of perturbation realization, we define a realization matrix and prove that it satisfies the Lyapunov equation. From the realization matrix we define a performance potential and prove that it satisfies the Poisson equation. Sensitivity formulas and policy iteration algorithms of Semi-Markov decision process (SMDPs) can be derived. The performance sensitivities can be obtained and policy iteration of SMDPs can be implemeted on a single sample path of the SMPs.

Keywords: Potentials, Lyapunov equations, Poisson equations, perturbation analysis, policy iteration

1. INTRODUCTION

Recent research shows that Markov decision processes (MDPs) can be viewed from a sensitivity point of view, and both MDPs and perturbation analysis (PA) of Markov processes are based on an important concept, called performance potential, which is strongly related to perturbation realization in PA [5] [4]. In this paper, we extend these results to semi-Markov processes (SMPs). Starting with the concept of perturbation realization, we define a realization matrix and prove that it satisfies the Lyapunov equation. From the realization matrix we define a performance potential and prove that it satisfies the Poisson equation. Sensitivity formulas and policy iteration algorithms of semi-Markov decision processes (SMDP) can be derived then. It is also shown that the potenials can be estimated on a single sample path and hence online algorithms can be derived for performance sensitivities and policy iteration of SMDPs.

2. FUNDAMENTALS FOR SEMI-MARKOV PROCESSES

We study a semi-Markov process defined on a countable state space $\mathcal{E} = \{1, 2, \cdots\}$. Let $T_0, T_1, \cdots, T_n, \cdots$ be the transition epoches. The process is right continuous so the state at each transition epoch is the state after the transition. Let $X_n = X_{T_n}$, $n = 0, 1, 2, \cdots$.

Define the semi-Markov kernel [6] as

$$Q(i, j, t) = P\{X_{n+1} = j, T_{n+1} - T_n \le t | X_n = i\}.$$

Set

$$\begin{aligned} Q(i,t) &= \sum_{j \in \mathcal{E}} Q(i,j,t) \\ &= P\{T_{n+1} - T_n \le t | X_n = i\}, \\ H(i,t) &= 1 - Q(i,t), \end{aligned}$$

[1] Supported in part by a grant from Hong Kong UGC. Tel: (852)2358-7048 Fax: (852)2358-1485 Email: eecao@ust.hk

$$\begin{aligned} Q(i,j) &= \lim_{t\to\infty} Q(i,j,t) \\ &= P\{X_{n+1} = j | X_n = i\}, \end{aligned}$$

and

$$\begin{aligned} G(i,j,t) &= \frac{Q(i,j,t)}{Q(i,j)} \\ &= P\{T_{n+1} - T_n \leq t | X_n = i, X_{n+1} = j\}. \end{aligned}$$

Define the hazard rates

$$q(i,t) = \frac{Q'(i,t)}{H(i,t)},$$

where the prime denote the deivative with respect to t, and

$$q(i,j,t) = \frac{Q'(i,j,t)}{H(i,t)},$$

the latter is the rate that the process jumps from i to j in $[t, t+\Delta t)$ given that the process does not jump out from state i in $[0,t)$.

Let $P_t(i,j) = P\{X_t = j | X_0 = i\}$. Then we have

$$\begin{aligned} P_{t+\Delta t}(i,j) &= \sum_{k\in\mathcal{E}} P_t(i,k) \int_0^\infty p_t(s|k) \\ &\times\{I(k,j)[1 - q(k,s)\Delta t] + q(k,j,s)\Delta t\} ds \quad (1) \end{aligned}$$

where $p_t(s|k)ds$ is the probability that given the state at time t is k the process has been in state k for a period of s to $s + ds$. $I(j,k) = 1$ if $j = k$, $I(j,k) = 0$ if $j \neq k$. Letting $\Delta t \to 0$, we get

$$\begin{aligned} \frac{dP_t(i,j)}{dt} &= -\sum_{k\in\mathcal{E}} P_t(i,k) \int_0^\infty \{p_t(s|k) \\ &\cdot [I(k,j)q(k,s) - q(k,j,s)]\} ds. \quad (2) \end{aligned}$$

When $t \to \infty$, we have $\frac{dP_t(i,j)}{dt} \to 0$ and $P_t(i,j) \to p(j)$, the steady-state probability of j. We further observe that as $t \to \infty$, $p_t(s|k)$ is the probability that the interval $[s, s+ds)$ appears in the entire sojourn time of state k. Thus, we have

$$\lim_{t\to\infty} p_t(s|k) ds = \frac{dsH(k,s)}{\int_0^\infty sQ(k,ds)}.$$

Therefore,

$$\lim_{t\to\infty} p_t(s|k) = \frac{H(k,s)}{m_k},$$

where

$$m_k = \int_0^\infty sQ(k,ds)$$

is the mean of the sojourn time at state k. Letting $t \to \infty$ in both sides of (2), we get

$$\begin{aligned} 0 &= -\sum_{k\in\mathcal{E}} p(k) \int_0^\infty \frac{1}{m_k} [I(k,j)Q'(k,s) - Q'(k,j,s)] ds \\ &= -\sum_{k\in\mathcal{E}} p(k) \{\frac{1}{m_k} [I(k,j) - Q(k,j)]\} \\ &= -\sum_{k\in\mathcal{E}} p(k) \{\lambda_k [I(k,j) - Q(k,j)]\}, \end{aligned}$$

where

$$\lambda_k = \frac{1}{m_k}.$$

Finally, we have

$$\sum_{k\in\mathcal{E}} p(k)A(k,j) = 0 \quad \text{for all } j \in \mathcal{E}, \qquad (3)$$

where

$$A(k,j) = -\lambda_k [I(k,j) - Q(k,j)]\}.$$

In a matrix form, we can write

$$p^T A = 0, \qquad (4)$$

where $p^T = (p(1), p(2), \cdots,)$ is the steady state probability vector, the superscript "T" denotes transpose, and A is a matrix whose kth row and jth column is $A(k,j)$. In addition, we have

$$Ae = 0,$$

where $e = (1, 1, \cdots)^T$ is a column vector whose components are all ones.

Equation (4) is exactly the same as the Markov process with A as the infinitesimal generator. This means that the steady-state probability is insensitive to the high order statistics for the sojourn times at all states. Also, the steady-state probabiliy does not depend on whether the sojourn time at state i depends on j, the state it jumps into from i.

3. REALIZATION MATRICES AND PERFORMANCE POTENTIALS

Consider a semi-Markov process starting from a transition epoch $X_0 = j$. Denote the instant at which the process jumps into state i for the first time as

$$S^j(i) = inf\{t \geq 0, X_t = i | X_0 = j\}.$$

We consider the general case where the performance measurement in $[T_n, T_n + 1)$ can depend on both X_n and X_{n+1}. Denote $f : \mathcal{E} \times \mathcal{E} \to R$ be the performance function. At any time $t \in [T_n, T_n + 1)$,denote $Y_t = X_{T_n+1}$.Thus, the performance measure at any time t is $f(X_t, Y_t)$.

Now we define the perturbation realization factors as

$$D(i,j) = E\{\int_0^{S^j(i)} [f(X_t,Y_t)-\eta]dt|X_0=j\}.(5)$$

where

$$\eta = \lim_{T\to\infty}\frac{1}{T}\int_0^T f(X_t,Y_t)dt$$

is the steady-state performance. Let $p(i,j)$ be the steady-state probability of $X_t=i$ and $Y_t=j$ and $p(j|i)$ be the conditional steady-state probability of $Y_t=j$ given that $X_t=i$, e.g., $\lim_{t\to\infty}P(Y_t=j|X_t=i)$ (not to be confused with $\lim_{n\to\infty}P(X_{n+1}=j|X_n=i)$). We have

$$p(j|i) = \frac{\int_0^\infty sQ(i,j,ds)}{\int_0^\infty sQ(i,ds)} = \frac{Q(i,j)m_{i,j}}{m_i},$$

where

$$m_{i,j} = \int_0^\infty sG(i,j,ds), \qquad (6)$$

and

$$m_i = \sum_{j\in\mathcal{E}} Q(i,j)m_{i,j} = \int_0^\infty sQ(i,ds). \qquad (7)$$

Thus

$$p(i,j) = p(j|i)p(i) = p(i)\frac{Q(i,j)m_{i,j}}{m_i}$$

and

$$\eta = \sum_{i,j\in\mathcal{E}} p(i,j)f(i,j) = \sum_{i\in\mathcal{E}} p(i)f(i),$$

where $f(i)$ is defined as

$$f(i) = \frac{\sum_{j\in\mathcal{E}} Q(i,j)f(i,j)m_{i,j}}{m_i}. \qquad (8)$$

From definition, we have

$$\begin{aligned}
D(i,j) &= E\{\int_0^{T_1}[f(X_t,Y_t)-\eta]dt|X_0=j\} \\
&\quad + E\{\int_{T_1}^{S^j(i)}[f(X_t,Y_t)-\eta]dt|X_0=j\} \\
&= \sum_{k\in\mathcal{E}} Q(j,k)\Big\{E\{\int_0^{T_1}[f(X_0,Y_0)-\eta]dt| \\
&\qquad X_0=j,X_1=k\} \\
&\quad + E\{\int_{T_1}^{S^j(i)}[f(X_t,Y_t)-\eta]dt|X_0=j,X_1=k\}\Big\} \\
&= \sum_{k\in\mathcal{E}} Q(j,k)\Big\{[f(j,k)-\eta]E\{T_1|X_0=j,X_1=k\} \\
&\quad + E\{\int_{T_1}^{S^k(i)}[f(X_t,Y_t)-\eta]dt|X_1=k\}\Big\} \\
&= \sum_{k\in\mathcal{E}} Q(j,k)\Big\{[f(j,k)-\eta]m_{j,k} \\
&\quad + E\{\int_{T_1}^{S^k(i)}[f(X_t,Y_t)-\eta]dt|X_1=k\}\Big\},
\end{aligned}$$

From (7) and (8), the above equation leads to

$$D(i,j) = m_j[f(j)-\eta] + \sum_{k\in\mathcal{E}} Q(j,k)D(i,k),$$

or equivalently,

$$\begin{aligned}
-[f(j)-\eta] &= \sum_{k\in\mathcal{E}}\{-\lambda_j[I(j,k)-Q(j,k)]D(i,k)\} \\
&= \sum_{k\in\mathcal{E}}\{A(j,k)D(i,k)\}. \qquad (9)
\end{aligned}$$

In a Matrix form, this is

$$DA^T = -[ef^T - \eta ee^T], \qquad (10)$$

where D is a matrix whose components are $D(i,j)$, and $f^T=(f(1),f(2),\cdots,)$.

Next, on the process X_t, with $T_0=0$ being a transition epoch and $X_0=j$, for any state $i\in\mathcal{E}$ we define a sequence $u_0,u_1,\cdots$, as follows.

$$u_0 = T_0 = 0,$$

$$v_n = inf\{t\geq u_n, X_t=i\}.$$

and

$$u_{n+1} = \inf\{t\geq v_n, X_t=j\},$$

e.g., v_n is the first time when the process reaches i after u_n, and u_{n+1} is the first time when the process reaches j after v_n. Apparently, $u_0,u_1,\cdots$ are stopping times and hence X_t is a regenerative process with $\{u_n, n=0,1,\cdots\}$ as its associated renewal process. By the regenerative theory, we have

$$\begin{aligned}
\eta &= \frac{E\{\int_{u_0}^{u_1} f(X_t,Y_t)dt\}}{E(u_1-u_0)} \\
&= \frac{E\{\int_0^{v_0} f(X_t,Y_t)dt\} + E\{\int_{v_0}^{u_1} f(X_t,Y_t)dt\}}{E[v_0]+E[u_1-v_0]}.
\end{aligned}$$

Thus,

$$E\{\int_0^{v_0}[f(X_t,Y_t)-\eta]dt\}$$
$$+E\{\int_{v_0}^{u_1}[f(X_t,Y_t)-\eta]dt\}=0.$$

By the definition of u_0, v_0 and u_1, we know that the above equation is

$$D(i,j)+D(j,i)=0,$$

or the matrix D is skew-symmetric

$$D^T=-D. \tag{11}$$

Taking transpose of (10), we get

$$-AD=-[fe^T-\eta ee^T].$$

From the above equation and (10), D satisfies the following Lyapunov equation

$$AD+DA^T=-F, \tag{12}$$

where $F=ef^T-fe^T$.

Since D is skew-symmetric, we can write it as

$$D=eg^T-ge^T, \tag{13}$$

where $g^T=(g(1),g(2),\cdots)$ is a column vector. Note that if g fits (13), so does $g+ce$ for any constant c. g is called a performance potential vector, and $g(i)$ the performance potential at state i. As we explained, g may have different versions each of them differs by only a constant.

Substituting (13) into (10), we get

$$Ag=-f+\eta e. \tag{14}$$

Since $Ae=0$, A is not invertable. Now suppose g is any solution to (14). Set $c=\eta-p^Tg$ and choose $g'=g+ce$. Then $p^Tg'=\eta$. Thus, there always exists a solution to (14) such that $p^Tg=\eta$. Putting this into (14), we get

$$Ag=-f+(p^Tg)e=-f+e(p^Tg).$$

Thus, we have the Poisson equation for g:

$$(-A+ep^T)g=f. \tag{15}$$

This is the same as the Poisson equation for Markov processes. In particular, for ergodic semi-Markov processes, $(-A+ep^T)$ is invertable. (15) only defines a particular version of the performance potentials. Multiplying both sides of (15) by p^T on the left side, we get

$$p^Tg=\eta.$$

4. SENSITIVITY AND SEMI-MARKOV DECISION PROCESSES

We have shown that by properly defining g and A, semi-Markov processes have the same Poisson equation for potentials and Lyapunov equation for perturbation realization matrices as thoes for Markov processes. Thus, performance sensitivity formulas can be derived in a similar manner and are briefly stated here.

First, for two semi-Markov processes with A', η' f' and A, η, f, multiplying both sides of (15) by p', we get

$$\eta'-\eta=p'^T[(A'-A)g+(f'-f)]. \tag{16}$$

This serves as a foundation for semi-Markov decision processes. Policy iteration for semi-Markov processes can be derived from (15) by noting $p'>0$ componentwise.

Next, suppose A changes to $A(\delta)=A+B\delta$, f changes to $f(\delta)=f+h\delta$, with δ being a small real number and $Be=0$. Then η changes to $\eta(\delta)=\eta+\Delta\eta$ and p changes to $p(\delta)$. From (16), we have

$$\begin{aligned}&\eta(\delta)-\eta(0)\\=&p^T(\delta)[(A(\delta)-A)g+(f(\delta)-f)].\end{aligned}$$

Letting $\delta\to 0$, we get

$$\frac{\partial\eta}{\partial B}=p^T(Bg+h). \tag{17}$$

We can also obtain performance sensitivity using D. We have

$$Dp=(eg^T-ge^T)p=\eta e-g.$$

Replacing g with D in the sensitivity equation, we get

$$\eta'-\eta=p'^T[(A'-A)D^Tp+(f'-f)],$$

and

$$\frac{\partial\eta}{\partial B}=p^T(BD^Tp+h).$$

Equation (5) provides a way to estimate the realization matrix or the potentials on sample paths. From (5), we can also obtain

$$D(i,j)=\lim_{T\to\infty}\{E\{\int_0^T[f(X_t)-\eta]dt|X_0=j\}$$
$$-E\{\int_0^T[f(X_t')-\eta]dt|X_0=i\}\}.$$

Therefore,

$$g(j) = \lim_{T\to\infty} E\{\int_0^T [f(X_t) - \eta]dt | X_0 = j\}$$

is performance potential at j. This is the same as for the Markov process case, except that the integration starts with a transition epoch. Single sample path based algorithms can be developed for potentials, and therefore the performance derivative (17) can be obtained and policy iteration can be implemented with a single sample path.

5. CONCLUSIONS

We have shown that with properly defined A, g and D, the results for potentials, perturbation realization, PA, and MDPs, etc., can be extended to SMPs naturally. This provides a powerful tool in optimization of SMP type of systems. Especially, the potentials, which play a crucial role in both sensitivity analysis and policy iteration, can be measured by the long term performance integration, which has the same physical meaning as for Markov processes. Future research topics include extensions to more general processes such as generalized semi-Markov processes (GSMPs) and applications to queueing networks with general serive time distributions.

6. REFERENCES

[1] D. P. Bertsekas, *Dynamic Programming and Optimal Control,* Vols. I, II, Athena Scientific, Belmont, Massachusetts, 1995.

[2] A. Berman and R. J. Plemmons, "Nonnegative Matrices in the Mathematical Sciences," *SIAM*, Philadelphia, 1994.

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

[4] Xi-Ren Cao, "The Relation Among Potentials, Perturbation Analysis, Markov Decision Processes, and Other Topics," *Journal of Discrete Event Dynamic Systems,* Vol. 8, 71-87, 1998.

[5] 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.

[6] E. Çinlar, *Introduction to Stochastic Processes,* Prentice Hall, Englewood cliffs, NJ, 1975.

[7] P. W. Glynn and S. P. Meyn, "A Lyapunov Bound for Solutions of Poisson's Equation," *Ann. Probab.,* 916-931, Vol. 24, 1996.

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

[9] J. G. Kemeny and J. L. Snell, *Finite Markov Chains,* Van Nostrand, New York, 1960.

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

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