On weak conditions for optimality inequalities in controlled Markov chains with exponential average cost

Agustin Brau-Rojas
Departamento de Matematicas
Universidad de Sonora.

Emmanuel Fernández-Gaucherand
Dept. of Electrical & Computer Eng. and Computer Science
University of Cincinnati.

1 The Model. In this paper we deal with the standard model for a discrete CMC specified by a four tuple $(\mathbb{X}, \mathbb{A}, P, C)$, where $\mathbb{X}$, the state space, is a countable or finite set; $\mathbb{A}$, the action space, is a finite set; P is a transition probability kernel from $\mathbb{K} := \mathbb{X} \times \mathbb{A}$ to $\mathbb{X}$ and $C : \mathbb{K} \to [0, K]$, $K > 0$, is the cost per stage function, see [1, 8, 10]. Sometimes, we will establish the probability kernel by means of a set of matrices $\{P(a) : a \in \mathbb{A}\}$, so that $P(y \mid x, a) := P_{xy}(a)$.

The controlled Markov chain $\{X_n\}$ is determined in the following way. At each time $t \in \{0, 1, \dots\}$ the state of the system is observed, say $X_t = x \in \mathbb{X}$, and an action $a_0 \in \mathbb{A}$ is chosen. Then a cost $C(x, a)$ is incurred and, regardless of the previous states and actions, the state of the system at time $t + 1$ will be $X_{t+1} = y \in \mathbb{X}$ with probability $P(y \mid x, a)$.

We will restrict attention to stationary deterministic policies, that is, rules for prescribing how to choose actions by means of a decision function $f : \mathbb{X} \to \mathbb{A}$. Such a policy will be denoted by f^∞, meaning that action $f(x)$ is chosen if the system is in state x regardless of time the observation is made. Following standard notation, we will denote by P^f and E^f respectively the probability measure and the expectation operator induced by the policy f^∞ on the canonical product space $(\mathbb{X}^\infty, \mathcal{B}^\infty)$.

The performance index for CMCs discussed here is the so called *exponential average cost* (EAC), which is the (exponential utility) risk-sensitive version of the well known (risk-neutral) average cost (see e.g. [2, 4, 5, 6],). The EAC corresponding to a policy f^∞ is defined as

$$J^f(\gamma, x) := \limsup_{n\to\infty} \frac{1}{n} \frac{1}{\gamma} \log E_x^f \left[\exp(\gamma S_n)\right],$$

where $S_n := \sum_{t=0}^{n-1} C(X_t, A_t)$, and the optimal exponential average cost (OEAC) by

$$J^*(\gamma, x) := \inf_f J^f(\gamma, x),$$

where the infimum above is taken over the class of all stationary deterministic policies.

A standard approach to the problem of finding optimal policies for CMCs with average cost is based on the existence of solutions to an average optimality

equation or an average optimality inequality, see e.g. . Recently, an optimality inequality result (Theorem 4.1, [7]) was presented for the risk-neutral case, purpotedly trying to emulate what have been done previously for the risk-neutral case. The mentioned result proves the existence of solutions to an optimality inequality under (a) a stability condition and (b) a limiting condition on the cost structure. We state below the mentioned result and assumptions respectively as Theorem 1 and Assumption A for ease of reference.

Assumption A (a) There exists a stationary policy f^∞ such that

$$\rho = J^f(\gamma, x) \tag{1}$$

is finite and independent of $x \in \mathbb{X}$.
(b)

$$\liminf_{x\to\infty} \min_{a\in\mathbb{A}} C(x,a) > \rho. \tag{2}$$

Theorem 1. Under Assumption A, there exists a number ρ^* and a (possibly extended) function W on $\mathbb{X}$ such that for all $x \in \mathbb{X}$

$$e^{\gamma(\rho^* + W(x))} \geqslant \inf_{a\in\mathbb{A}} \left\{ e^{\gamma C(x,a)} \sum_y e^{\gamma W(y)} P(y \mid x, a) \right\} \tag{3}$$

and the set $H := \{x \in \mathbb{X} : W(x) \text{ is finite}\}$ is not empty. Moreover, there exists an optimal stationary deterministic policy f^∞ whenever the initial state is in H, and $\rho^* = J^f(\gamma, x)$ for all $x \in H$.

Herein, we present three examples which highlights (a) the significant differences between the risk-neutral and the risk-sensitive criteria, and (b) the strength and weaknesses of the result in [7].

For the computation of some of the EAC's arising in the examples, we will use the following theorem, see [3].

Theorem 2. If $P = \big(P(x,y)\big)$ is the transition probability matrix of a cost Markov chain X_n then

$$\begin{aligned} J(\gamma, x) :=& \limsup_{n\to\infty} \frac{1}{n} \frac{1}{\gamma} \log E_x \left[\exp(\gamma S_n)\right] \\ =& \frac{1}{\gamma} \max\left\{ \log\, \lambda(\tilde{P}_{\mathcal{C}}) : x \to \mathcal{C} \right\}, \end{aligned}$$

where $S_n := \frac{1}{n}\sum_{j=0}^{n-1} c(X_j)$, c is the cost function, $\mathcal{C}$ denotes a maximal self-communicating class of states,

$$\tilde{P}_{\mathcal{C}} := \big(P(x,y)e^{\gamma c(x)}\big)_{x,y\in\mathcal{C}}$$

is the disutility matriz corresponding to P, and finally, $\lambda(\tilde{P}_{\mathcal{C}})$ denotes the spectral radius of $\tilde{P}_{\mathcal{C}}$.

We begin with an elementary ("multichain") example which focus attention on the stability condition in Assumption A(a). The example is related to the following general, yet simple observation. When the cost function is bounded, say $C(x,a) \leqslant K \ \forall\,(x,a) \in \mathbb{K}$, any couple (ρ^*, W) such that $\rho^* > K$ and $W(\cdot) \equiv K$ is a solution of (3). Indeed,

$$e^{\gamma C(x,a)} \sum_y e^{\gamma W(y)} P(y \mid x,a) \leqslant e^{\gamma K} e^{\gamma K} \leqslant e^{\gamma(\rho^* + W(x))}.$$

For those solutions, $H := \{x \in \mathbb{X} : W(x) \text{ is finite}\} = \mathbb{X}$; however, $\rho^* > J^*(\gamma, x)$ $\forall\, x \in \mathbb{X}$, and consequently the minimizing actions in the right side of (3) have nothing to do with the optimal EAC. Thus, those trivial solutions are generally useless for the main purpose of the optimality inequality, namely, searching for optimal policies. Example 1 below further illustrates this limitation of Theorem 1 , and it does it with a value of ρ^* less than $\sup C(x,a)$.

Example 1. Consider the CMC with state space $S := \{1,2,3\}$, action space $\mathbb{A} = \{a, b\}$, transition probabilities given by the matrices

$$P(a) := \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad P(b) := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix},$$

and cost function C defined by $C(1,a) = C(1,b) = L$, $C(2,a) = C(2,b) := M$, and $C(3,a) = C(3,b) = U$ with $0 < L < U < M$.

Decision functions for this CMC are determined by specifying their value on state 1 because both the transition probabilities and the costs at states 2 and 3 do not depend on the action. Thus, let d and e denote decision functions such that $d(1) = a$ and $e(1) = b$. It is immediate that

$$J^d(\gamma,1) = J^d(\gamma,2) = J^d(\gamma,3) = U \quad \text{and} \quad J^e(\gamma,x) = \begin{cases} U & \text{if } x = 2,3 \\ L & \text{if } x = 1 \end{cases},$$

so that e^∞ is optimal, i.e., $J^* = J^e$. Also, Assumption A(a) is verified since $J^d(\gamma,x)$ is independent of the state x.

We can check that if we take ρ^*, with $U < \rho^* < M$ and the function W such that $\infty > W(1) > W(2) > W(3)$ and $W(2) - W(3) \geqslant M - \rho^*$, then (ρ^*, W) is a solution of (3) and $H := \{W(x) < \infty\} = S$. However $J^*(\gamma,x) \neq \rho^*$ $\forall\, x$. Moreover, the action that minimizes the right-hand side of (3) for state 1 is a and $e(1) = b \neq a$. Summarizing, the described solution of (3) does not satisfy the conclusion of Theorem 1, i.e., it does not provide the optimal decision function on H.

The next two examples will address another limitation of Theorem 1. They both consist of a CMC with state space $\mathbb{X} = \{1,2,\dots\}$ for which the optimal

EAC has an infinite number of different values. Thus, they illustrate the fact that the optimal average inequality in Theorem 1 may not provide the optimal decision function in most of the state space. Morevoer, the CMC in Example 3 has a strong recurrence structure, namely, it satisfies a simultaneous Doeblin condition.

Example 2. Consider first the CMC with state space $S := \{1, 2\}$, action space $\mathbb{A} = \{a, b\}$, transition probabilities given by the matrices

$$P(a) := \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad P(b) := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and cost function determined by $C(1, a) = C(1, b) := M$ and $C(2, a) = C(2, b) := L$, with $0 < L < M$.

Decision functions for this CMC are determined by specifying their value on state 2 because both the transition probabilities and the costs at state 1 do not depend on the action. Thus, let d and e denote decision functions such that $d(2) = a$ and $e(2) = b$. It is immediate that

$$J^d(\gamma, 1) = J^d(\gamma, 2) = M \quad \text{and} \quad J^e(\gamma, x) = \begin{cases} M & \text{if } x = 1 \\ L & \text{if } x = 2 \end{cases},$$

so that e^∞ is optimal. Also, Assumption A(a) is verified since $J^d(\gamma, x)$ is independent of the state x and Assumption A(b) is trivially satisfied. However, the optimal value function depends on the state space and thus the optimal decision function can not be obtained from (3).

Based on the previous finite structure, now we construct a CMC with state space $\mathbb{X}$, as the model in [7], for which difficulties similar to those in the above finite model appears, for an infinite number of states.

Let us extend the previous finite model by appending the "tail" $\{3, 4, \dots\}$ and defining $P(2 \mid x, a) = P(2 \mid x, b) = p_x$, $P(x \mid x, a) = P(x \mid x, b) = 1 - p_x$ with $0 < p_x < 1$, and $C(x, a) = C(x, b) = R > M \ \forall\, x \geqslant 3$. In that way, Part (b) of Assumption A in Theorem 1 is satisfied:

$$\liminf_{x \to \infty} \min_{a \in \mathbb{A}} C(x, a) = R > M = J^d(\gamma, x).$$

Moreover, the values of p_x can be chosen so that $L < R + \frac{1}{\gamma} \log p_x < M$ because $L < M < R$ and

$$\lim_{\gamma \to +\infty} R + \frac{1}{\gamma} \log p_x = R \text{ and } \lim_{\gamma \to 0} R + \frac{1}{\gamma} \log p_x = -\infty, \tag{4}$$

and, by Theorem 2, we have now

$$J^{\bar{d}}(\gamma, x) = M \ \forall\, x \quad \text{and} \quad J^{\bar{e}}(\gamma, x) = \begin{cases} M & \text{if } x = 1 \\ L & \text{if } x = 2 \\ R + \frac{1}{\gamma} \log p_x & \text{if } x \geqslant 3 \end{cases},$$

where $\bar{d}$ and $\bar{e}$ are the obvious extensions of d and 3 respectively. Consequently, $\bar{e}^\infty$ is optimal. Therefore as in the finite model, (3) provides the optimal decision function only in one out of an infinite number of states. One might conjecture that the observed behaviour is caused by the weak recurrence structure of the CMC in the example . Nevertheless, as Example 3 below will show, even if we add a Doeblin Condition to Assumption A of Theorem 4.1 in [7], the result will still show the observed fragility.

Example 3. Similar to what we did in Example 2, the present example will be based on the following basic scheme provided by a CMC with state space $S := \{1, 2\}$, action space $\mathbb{A} := \{a, b\}$ and transition probabilities given by

$$P(a) := \begin{pmatrix} 1 & 0 \\ 1-p_2 & p_2 \end{pmatrix} \quad \text{and} \quad P(b) := \begin{pmatrix} p_1 & 1-p_1 \\ 1-p_2 & p_2 \end{pmatrix}, \tag{5}$$

where $0 < p_1 < 1$, $0 < p_2 < 1$. The cost structure for this CMC will be defined by $C(1, a) = C(1, b) := L$ and $C(2, a) = C(2, b) := M$, where $0 < L < M$.

Again, decision functions for this CMC are determined by specifying their value on state 1 because both the transition probabilities and the costs at state 2 do not depend on the action. Consequently, let us just consider decision functions d and e such that $d(1) = a$ and $e(1) = b$. According to Theorem 2, the EAC's for the corresponding stationary deterministic policies d^∞ and e^∞ are, respectively,

$$J^d(\gamma, 1) = L, \quad J^d(\gamma, 2) = \max\left\{L,\, M + \frac{1}{\gamma}\log\frac{1}{2}\right\} \tag{6}$$

and

$$J^e(\gamma, 1) = J^e(\gamma, 2) = \frac{1}{\gamma}\log\lambda(\tilde{P}_e), \tag{7}$$

where $\lambda(\tilde{P}_e)$ is the spectral radius (see [9] of the disutility matrix

$$\tilde{P}_e := \begin{pmatrix} p_1\, e^{\gamma L} & (1-p_1)\, e^{\gamma L} \\ (1-p_2)e^{\gamma M} & p_2\, e^{\gamma M} \end{pmatrix}. \tag{8}$$

We have then

$$L < \frac{1}{\gamma}\log\lambda(\tilde{P}_e) < M \quad \text{and} \quad M + \frac{1}{\gamma}\log p_2 < \frac{1}{\gamma}\log\lambda(\tilde{P}_e) < M. \tag{9}$$

The inequalities on the right above can be checked directly by carrying out the computation of $\lambda(\tilde{P}_e)$ in terms of the entries of $\tilde{P}_e$ or by appealing to a Perron-Frobenius theory argument as follows. If v is the Perron-Frobenius vector of $\tilde{P}_e$ and we set $B := (B_{ij})$, the 2×2 matrix with $B_{22} = p_2 e^{\gamma M}$ and $B_{ij} = 0$ otherwise, then we have

$$\lambda(\tilde{P}_e)\, v \;=\; \tilde{P}_e\, v \;>\; B\, v,$$

because $v > 0$. The fact that $v > 0$ allows us as well to use a simple result from P-F theory (Corollary 8.1.29, [9]) to conclude that $\lambda(\tilde{P}_e) > \lambda(B) = p_2 e^{\gamma M}$. The inqualities on the left can be obtained with a similar argument.

Let us now "extend the previous CMC the state space $\mathbb{X} = \{1, 2, \dots\}$. In this case, let us define the transition probabilities for $x = 3, 4, \dots$ as

$$P(x \mid x, a) = P(x \mid x, b) = p_x, \quad P(1 \mid x, a) = P(1 \mid x, b) = 1 - p_x,$$

where $0 < p_x < 1$, and extend the cost function by defining $C(x, a) := M$ for those states.

As before, it is sufficient to consider two stationary deterministic policies for this CMC, namely, those respectively determined by the decision functions

$$g(x) = \begin{cases} b & \text{if } x = 1, \\ a & \text{elsewhere} \end{cases} \quad \text{and} \quad h(x) = a \quad \forall\, x.$$

Now, take p_x, $x = 3, 4, \dots$ in the interval $(0, p_2)$ and such that $L < M + \frac{1}{\gamma} \log p_x$. That choice is possible because $\lim_{p \to 0} M + \frac{1}{\gamma} \log p = -\infty$. Then, taking into account that $p_x < p_2$, $x = 3, 4, \dots$ and the second inequality in (9) we have

$$L < M + \frac{1}{\gamma} \log p_x < M + \frac{1}{\gamma} \log p_2 < \frac{1}{\gamma} \log \lambda(\tilde{P}_e) < M, \tag{10}$$

where $\tilde{P}_e$ is matrix (8).

Thus, noting that $\{1, 2\}$ is a closed class under both policies g^∞ and h^∞, it follows from Theorem 2 that $J^g(\gamma, x) = \frac{1}{\gamma} \log \lambda(\tilde{P}_e)\ \forall\, x$ and

$$J^h(\gamma, x) = \begin{cases} L & \text{if } x = 1 \\ M + \frac{1}{\gamma} \log p_x & \text{elsewhere.} \end{cases} \tag{11}$$

Policy h^∞ is then optimal and the optimal EAC depends on initial state. Moreover, for fixed γ, the p_x's can be chosen to be different for different x's. Consequently, also for this example the optimality inequality gives the optimal EAC only in one out of an infinite number of states.

It must be noted that, as announced earlier, the CMC in this example satisfies a simultaneous Doeblin Condition. Indeed, if we set

$$\tau_1 := \inf \{n > 0 : X_n = 1\},$$

then

$$E_x^f[\tau_1] = E_x^g[\tau_1] = (1 - p_x) \sum_{n=1}^{\infty} n p_x^{n-1} < \sum_{n=1}^{\infty} n p_2^{n-1} < \infty,$$

for $x \geqslant 2$, $E_1^h[\tau_1] = 1$, and

$$E_1^g[\tau_1] = p_1 + (1 - p_1)(1 - p_2) \sum_{n=1}^{\infty} n p_2^{n-1} < \infty.$$

As announced, the previous example illustrates that even adding a simultaneous Doeblin condition, the stability condition used in [7] the CMC exhibits (perhaps undesirable) pathologies not found in the better understood risk-neutral model.

References

[1] A. Arapostathis, V. S. Borkar, E. Fernández-Gaucherand, M. K. Ghosh, and S. I. Marcus. Discrete-time controlled Markov processes with average cost criterion: a survey. *SIAM J. Control and Optimization*, 31(2):282–344, March 1993.

[2] A. Brau and E. Fernandéz-Gaucherand. Controlled Markov chains with risk-sensitive exponential average cost criterion. In *Proceedings of the 36th IEEE Conference on Decision and Control*, pages 2260–2264, San Diego, CA, 1997.

[3] A. Brau-Rojas. *Controlled Markov Chains with risk-sensitive average cost criterion.* PhD thesis, Department of Mathematics, University of Arizona, 1999.

[4] R. Cavazos-Cadena and E. Fernandéz-Gaucherand. Controlled Markov chains with risk-sensitive criteria: Average cost, optimality equations, and optimal solutions. *Mathematical Methods of Operations Research*, 49:299–324, 1999.

[5] R. Cavazos-Cadena and E. Fernandéz-Gaucherand. Risk sensitive optimal control in communicating average markov decision chains. In M. Dror, P. L'Ecuyer, and D. F. Szidarovszky, editors, *Modeling Uncertainty. An Examination of Stochastic Theory, Methods and Applications.* Kluwer Academic Publishers, 2002.

[6] W. H. Fleming and D. Hernández-Hernández. Risk sensitive control of finite state machines on an infinite horizon I. *SIAM Journal on Control and Optimization*, 35(5):1970–1810, September 1997.

[7] D. Hernández-Hernández and S. Marcus. Existence of risk sensitive optimal stationary policies for controlled Markov processes. *Applied Mathematics and Optimization*, 40(3):273–285, November 1999.

[8] O. Hernández-Lerma. *Adaptive Markov Control Processes.* Springer-Verlag, New York, 1989.

[9] R. A. Horn and C. R. Johnson. *Matrix Analysis.* Cambridge University Press, Cambridge, 1982.

[10] M. L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming.* John Wiley & Sons, Inc., New York, 1994.