ON OPTIMALITY OF MONOTONE CHANNEL-AWARE TRANSMISSION POLICIES: A CONSTRAINED MARKOV DECISION PROCESS APPROACH

Minh Hanh Ngo, Vikram Krishnamurthy

Department of Electrical and Computer Engineering
University of British Columbia
Vancouver, BC, Canada V6T 1Z4
email: minhn@ece.ubc.ca; vikramk@ece.ubc.ca

ABSTRACT

A constrained Markov Decision Process (MDP) approach is deployed to prove the monotone structure of optimal channel-aware transmission policies for packet transmission over a correlated fading wireless channel subject to an average delay constraint. A transmission policy is a function mapping channel state information (CSI), buffer states and numbers of arriving packets to transmit probabilities. The objective is to minimize the average transmission energy cost subject to an average delay constraint. We use the Lagrange multiplier method to convert the constrained MDP to an unconstrained MDP and prove that the unconstrained optimal policy is threshold in the buffer state. It then follows that the constrained optimal transmission policy is a randomized mixture of two pure transmission policies that are threshold in the buffer occupancy.

Index Terms— Markov processes, stochastic optimal control, resource management.

1. INTRODUCTION

Perfect channel state information (CSI) is exploited for optimal point-to-point transmission of data packets, e.g., from a sensor to its cluster head in a multihop sensor network or in a file transferring system from a battery-operated node to a base station, to enhance channel utilization subject to a delay constraint. The wireless communication channel is modelled to evolve as a Finite State Markov Chain [1, 2]. The transmission policies optimization problem is formulated as an average cost, infinite horizon, countable state constrained MDP (CMDP), where the objective is to minimize the average transmission energy cost and the constraint is that the average delay cost must be less than a certain threshold.

The structure of the constrained optimal transmission policy is analysed using the Lagrange multiplier method. First, the CMDP is transformed into an unconstrained average cost MDP by introducing a Lagrange multiplier [3, 4]. Second, it is established based on the concept of supermodularity that the optimal policies of the unconstrained MDP is monotone in the buffer state. Due to a classical result in [3], it follows that the constrained optimal transmission policy is a randomized mixture of two threshold transmission policies.

The main result of this paper is the analytical proof that the constrained optimal transmission policy is monotonically increasing in the buffer state. In particular, the constrained optimal transmission policy is a randomized mixture of two transmission policies that are non-randomized and monotonically increasing (hence threshold) in the buffer state.

Related work: The most comprehensive reference on constrained MDPs is [4]. Important existence results for countable state average cost MDPs are given in [5, 6, 7]. In this paper, monotonicity of the optimal policies is proven by the concept of supermodularity, which is covered in depth in [8]. In [9], the concept of supermodularity is also used to prove optimality of monotone rate control policies. However, therein, the authors assume a finite state space and, additionally, the evolution of the components of the system state are completely decoupled. As a consequence, the analysis is the same for both correlated and i.i.d channels. In comparison, in the system model we consider, the evolution of the channel state affects the evolution of the buffer state as the number of successful transmissions depends on the channel state. Hence, the analysis is much more complicated for the case of correlated fading channels than i.i.d channels.

The organization of the paper is as follows. The system model and the CMDP formulation are given in Section 2. Monotonicity of the optimal transmission policies is established in Section 3. Section 4 concludes the paper.

2. ENERGY AND DELAY CRITICAL PACKET TRANSMISSION

Consider the transmission of packets from a node (e.g., a sensor) to another node (e.g., a local hub node) over a correlated fading wireless link. Assume that time is divided into equal-size slots indexed by $n = 0, 1, 2, \ldots$. During each time slot, exactly one packet can be transmitted. If a packet is transmitted but not successfully received, the packet remains in the buffer for later retransmission. Inherently, it is assumed that

there is an instantaneous error free feedback channel, so that the result (ACK or NACK) of each transmission is perfectly known at the transmitter side.

System state space: At each time slot, the state of the system is defined by the channel state information (CSI), the buffer occupancy, and the number of packets newly admitted to the buffer.

Due to the effect of shadowing, multipath/multiuser interference, and mobility, the communication channel is correlated fading. In a discrete time system, a correlated fading wireless communication channel can be modelled by a Finite State Markov Chain (FSMC) [1, 2]. The procedures for obtaining a FSMC channel model include properly partitioning the channel state domain into a finite number of ranges and computing the corresponding transition matrix.

We assume a FSMC channel model with K states. Denote the channel state space by $\Gamma = \{\gamma_1, \gamma_2, \ldots, \gamma_K\}$, where γ_i corresponds to a better channel state than γ_j for all $i > j$. Let $(p_{ij})_{i,j=1,2,\ldots,K}$ be the state transition probability matrix, i.e., $\mathbb{P}_\Gamma(x^{n+1} = \gamma_j | x^n = \gamma_i) = p_{ij}$, where x^n denotes the channel state at time slot n.

Assume that each node has an infinite buffer to store packets for transmission. Define the buffer state to be the buffer occupancy, and denote $\mathcal{B} = \{1, 2, \ldots, \}$ as the buffer state space. Denote the buffer state at time n by b^n, where $b^n \in \mathcal{B}$.

Denote the number of new packets arriving at the buffer at time slot n by $y^n \in \mathcal{Y}$, where $\mathcal{Y}$ is the arrival event space. For notational simplicity, assume that at each time slot, one packet arrives at the buffer with probability δ, i.e. assume an i.i.d probability distribution function $p_Y(\cdot) : \mathcal{Y} \to [0, 1]$, where $p_{\mathcal{Y}}(1) = \delta, p_{\mathcal{Y}}(0) = 1 - \delta$.

The system state space can then be denoted by $S = \mathcal{B} \times \Gamma \times \mathcal{Y}$. For every $s = [b, x, y] \in \mathcal{S}$, the action set is $\mathcal{A} = \{0, 1\}$, where 0 stands for the action of non-transmitting and 1 stands for transmitting, except when $b = 0$ then the only allowed action is $a = 0$. Selecting action $a \in \mathcal{A}$ for state $s \in \mathcal{S}$ yields a transmission cost $c(\cdot, \cdot) : S \times \mathcal{A} \to \mathbb{R}_+$ and a delay cost $d(\cdot) : \mathcal{A} \to \mathbb{R}_+$. Assume that $c(\cdot, 0) = 0$ and $d(1) = 0$. Furthermore, $c(\cdot, \cdot)$, which represents the transmission energy cost, should only depend on the channel state and the action. $c(\cdot, \cdot)$ non-increasing in the channel state.

If a transmission is attempted when the system is in state $s = [b, x, y]$, the success probability depends on the channel state x. Denote the success probability function by $f : \Gamma \times \mathcal{A} \to [0, 1]$, where $f(\cdot, 0) = 0$ and $f(\cdot, 1)$ is increasing, i.e. a higher channel state gives to a higher success probability.

The evolution of the system is then given by

$$p(s^{n+1}|s^n, a^n) = \begin{cases} p_\Gamma(x^{n+1}|x^n) p_Y(y^{n+1}) f(x^n, a^n) \\ \qquad \text{if } b^{n+1} = b^n - a^n + y^n \\ p_\Gamma(x^{n+1}|x^n) p_Y(y^{n+1}) \big(1 - f(x^n, a^n)\big) \\ \qquad \text{if } b^{n+1} = b^n + y^n \\ 0 \qquad \text{otherwise.} \end{cases}$$

CMDP formulation: The objective of the CMDP is to find a transmission policy that minimizes the average transmission cost subject to a constraint on the average delay cost. A transmission policy is a function mapping system states to transmit probabilities for every time slot. Whereas, a stationary transmission policy is independent of time and maps system states directly to transmit probabilities: $u(\cdot) : \mathcal{S} \to [0, 1]$. In this paper we focus on stationary policies. A condition on the existence of an optimal stationary policy will be given in the next section.

The average transmission and delay costs associated with a transmission policy u for initial state s_0 are given by

$$C(u|s^0) = \lim_{N\to\infty} \frac{1}{N} \mathbb{E}_u \Big[\sum_{k=1}^{N} c(s^k, a^k)_{|s^0} \Big], \tag{1}$$

$$D(u|s^0) = \lim_{N\to\infty} \frac{1}{N} \mathbb{E}_u \Big[\sum_{k=1}^{N} d(a^k)_{|s^0} \Big]. \tag{2}$$

The problem of optimizing a transmission policy is given by

$$C^*(s^0) = \inf_u C(u|s^0) \text{ s.t. } D(u|s^0) \le \mathbf{D}, \forall s^0 \in \mathcal{S}. \tag{3}$$

3. MONOTONE OPTIMAL CHANNEL-AWARE TRANSMISSION POLICIES

First, we give a sufficient condition for which there exists a stationary constrained optimal transmission policy. Then given the proper selection of a Lagrange multiplier, the unconstrained MDP also has a stationary optimal policy and it can be proved that the unconstrained and constrained optimal transmission policies are monotone in the buffer state.

3.1. Buffer Stability

Lemma 1 presents a condition for which every policy satisfying the average delay constraint induces a stable buffer.

Lemma 1. *Denote* $\min_{x\in\Gamma}\{f(x, 1)\} = \underline{f}$. *If*

$$\frac{\mathbf{D}}{d(0)} < 1 - \frac{\delta}{\underline{f}} \tag{4}$$

then every policy u *that satisfies the constraint* $d(u|s_0) < \mathbf{D}$ *induces a stable buffer, and hence a recurrent Markov chain.*

Proof. Let u be a transmission policy that satisfies the constraint $D(u|s_0) < \mathbf{D}$, i.e.

$$\mathbf{D} \ge D(u|s_0) = \lim \sup_{N\to\infty} \frac{1}{N} \mathbb{E}_u \Big[\sum_{k=1}^{N} d(x^k, 0) \mathbf{I}(a^k = 0|s^0) \Big]$$
$$\ge \lim \sup_{N\to\infty} \frac{1}{N} \mathbb{E}_u \Big[\sum_{k=1}^{N} d(\gamma_0, 0) \mathbf{I}(a^k = 0|s^0)$$

The average successful transmission rate of u is given by

$$\begin{aligned} r(u|s_0) &= \lim \inf_{N\to\infty} \frac{1}{N}\mathbb{E}_u\Big[\sum_{k=1}^{N} f(x^k, a^k|s^0)\Big] \\ &\geq \lim \inf_{N\to\infty} \frac{1}{N}\mathbb{E}_u\Big[\sum_{k=1}^{N} \underline{\mathrm{f}}\mathbf{I}(a^k = 1|s^0)\Big] \\ &= \underline{\mathrm{f}}\Big(1 - \lim \sup_{N\to\infty} \frac{1}{N}\mathbb{E}_u\Big[\sum_{k=1}^{N} \mathbf{I}(a^k = 0|s^0)\Big]\Big) \\ &\geq \underline{\mathrm{f}}\big(1 - \frac{\mathbf{D}}{d(\gamma_0, 0)}\big) \quad > \delta \quad \text{By (4).} \end{aligned}$$

Therefore the buffer is stable almost surely for policy u, i.e. the Markov Chain is recurrent due to Foster's Theorem [10]. □

3.2. Reformulation of the CMDP using the Lagrange multiplier method

For each Lagrange multiplier λ, define the instantaneous cost of the corresponding unconstrained MDP as follows

$$c(s^k, a^k; \lambda) = c(s^k, a^k) + \lambda d(a^k). \tag{5}$$

The new average cost of a policy is given by

$$C(u|s_0; \lambda) = \lim_{N\to\infty} \frac{1}{N}\mathbb{E}_u\Big[\sum_{k=1}^{N} c(s^k, a^k|s_0; \lambda)\Big]. \tag{6}$$

Then the dynamic programming problem becomes

$$C_{\lambda|s_0} = C(u^*_\lambda|s_0; \lambda) = \inf_u C(u|s_0; \lambda) \tag{7}$$

In the classical use of the Lagrange multiplier technique, the multiplier λ is chosen so that the constraint is met consistently with the original optimization problem [3]. With such selections of λ, due to Lemma 1, the optimal transmission policy of (7) induces a recurrent Markov chain. Hence, it follows that the optimal average cost is independent of the initial states and the unconstrained MDP can be rewritten as

$$\begin{aligned} C^*_\lambda &= C(u^*_\lambda; \lambda) = \inf_u C(u; \lambda) \\ u^*_\lambda &= \arg\inf_u C(u; \lambda). \end{aligned} \tag{8}$$

3.3. Monotonicity of optimal transmission policies

Having formulated the unconstrained MDP with a Lagrange multiplier, we prove here using the concept of supermodularity that the unconstrained optimal policies of (7) is monotonically increasing in the buffer state. The interpretation is that if the buffer occupancy is larger, which indirectly implies a relatively large average delay cost, it is beneficial to transmit more aggressively.

Definition 1. *A function $F(x, y) : X \times Y \to \mathbb{R}$ is supermodular in (x, y) if $F(x_1, y_1) + F(x_2, y_2) \geq F(x_1, y_2) + F(x_2, y_1)$ $\forall x_1, x_2 \in X, y_1, y_2 \in Y, x_1 > x_2, y_1 > y_2$. If the inequality is reversed, $F(\cdot, \cdot)$ is called submodular.*

Supermodularity is a sufficient condition for optimality of monotone policies. Specifically, if $F(x, y)$ defined as in Definition 1 is supermodular (submodular) in (x, y) then $y(x) = \arg\max_y F(x, y)$ is non-decreasing (non-increasing) in x [8].

Theorem 1 and Corollary 1 are on the monotonicity of unconstrained and constrained optimal transmission policy respectively.

Theorem 1. *$u^*_\lambda(\cdot)$ is deterministic and monotonically increasing, and hence threshold, in the buffer state component b of the system state $s = [b, x, y]$, i.e., $u^*_\lambda(\cdot)$ is of the form*

$$u^*_\lambda([b, x, y]) = \begin{cases} 1 & \text{if } b \geq b^*_{xy;\lambda} \\ 0 & \text{otherwise,} \end{cases} \tag{9}$$

where $b^*_{xy;\lambda}$ is the threshold buffer state for the channel state x and the arrival event y for the given Lagrange multiplier λ.

Proof. See the appendix. □

Corollary 1. *The constrained optimal transmission policy u^* of (3) is a randomized mixture of two pure threshold transmission policies, i.e.*

$$u^*([b, x, y]) = q u^*_{\lambda_1}([b, x, y]) + (1 - q) u^*_{\lambda_2}([b, x, y]), \tag{10}$$

where $u^*_{\lambda_1}([b, x, y]), u^*_{\lambda_2}([b, x, y])$ are of the form (9).

Proof. By Theorem 4.3 of [3] and Theorem 1. □

4. CONCLUSION

It is proved using the concept of supermodularity that the optimal constrained transmission policy is a randomized mixture of two pure transmission policies that are threshold in the buffer state. This structural result can be exploited to derive efficient algorithm for estimating the optimal policy. The analysis of the paper can be easily generalized for the case of a finite number of actions and much more general packet arrival processes.

Appendix - Proof of Theorem 1

For a proper selection of the Lagrange multiplier λ the optimal policy of (7) induces a stable buffer. It is then possible to verify that Assumptions 1, 2, and 3 in [5] holds. Therefore, due to [5], there exist a stationary policy $u(\cdot)$, which is a limit point of a sequence of discounted cost optimal policies for a sequence of discount factors that increase to 1, a scalar C^*_λ and a vector $h(s)$ satisfying

$$C^*_\lambda + h(s) \geq \min_{a\in\mathcal{A}}\Big[c(s, u(s); \lambda) + \sum_{s'\in\mathcal{S}} \mathbb{P}(s'|s, u(s))h(s')\Big]$$

for $s \in \mathcal{S}$. Since $u(\cdot)$ is a limit point of a sequence of discounted cost optimal policies, $u(\cdot)$ inherits the structure of discounted cost optimal policies [11], [5]. For any discount factor α, the optimal discounted cost satisfies

$$V_\alpha(s) = \min_{a \in \mathcal{A}_s} \Big\{ c(s,a;\lambda) + \alpha \sum_{s'} \mathbb{P}(s'|s,a) V_\alpha(s') \Big\}$$

and can be computed by the recursion

$$V_\alpha^{t+1}(s) = \min_{a \in \mathcal{A}_s} Q_\alpha^{t+1}(s,a)$$
$$Q_\alpha^{t+1}(s,a) = c(s,a;\lambda) + \alpha \sum_{s' \in \mathcal{S}} \mathbb{P}(s^{'}|s,a) V_\alpha^t(s^{'},a). \quad (11)$$

Furthermore, the optimal discounted cost policy is given by $u_\alpha^*(s) = \arg\min_{a \in \mathcal{A}_s} Q_\alpha(s,a)$. We now show that $u_\alpha^*(s)$, where $s = [b,x,y]$, is monotonically increasing in the buffer state b by proving using mathematical induction that $Q_\alpha([b,x,y],a)$ is submodular in the buffer state and the action, i.e.

$$Q_\alpha([b,x,y],1) - Q_\alpha([b,x,y],0) = c(s,1;\lambda) - c(s,0;\lambda)$$
$$+ \sum_{x' \in \Gamma, y' \in Y} \mathbb{P}_\Gamma(x^{'}|x) p_Y(y^{'}) f(x,1) \Big[V_\alpha(b+y-1, x^{'}, y^{'})$$
$$- V_\alpha(b+y, x^{'}, y^{'}) \Big]$$

is monotonically decreasing in b for $b \geq 1$, which holds if $V_\alpha([b,x,y])$ has increasing differences in b for all $b \geq 0$ (as $c(\cdot,\cdot;\lambda)$ is constant in b).

Since (11) converges for any initial condition, we can select $V_\alpha^0([b,x,y])$ with increasing differences in b. Assume that $V_\alpha^n([b,x,y])$ has increasing differences in b, we will show that $V_\alpha^{n+1}([b,x,y])$ also has increasing differences in b, i.e.

$$V_\alpha^{n+1}([b+1,x,y]) - V_\alpha^{n+1}([b,x,y])$$
$$- \big(V_\alpha^{n+1}([b,x,y]) - V_\alpha^{n+1}([b-1,x,y]) \big) \geq 0. \quad (12)$$

Now assume $V_\alpha^{n+1}([b+1,x,y]) = Q_\alpha^{n+1}([b+1,x,y],a_2)$, $V_\alpha^{n+1}([b,x,y]) = Q_\alpha^{n+1}([b,x,y],a_1)$, $V_\alpha^{n+1}([b-1,x,y]) = Q_\alpha^{n+1}([b-1,x,y],a_0)$ for some $a_2, a_1, a_0 \in \mathcal{A}$. Then (12) becomes

$$Q_\alpha^{n+1}([b+1,x,y],a_2) - Q_\alpha^{n+1}([b,x,y],a_1)$$
$$- Q_\alpha^{n+1}([b,x,y],a_1) + Q_\alpha^{n+1}([b-1,x,y],a_0) \geq 0$$
$$\Leftrightarrow \underbrace{Q_\alpha^{n+1}([b+1,x,y],a_2) - Q_\alpha^{n+1}([b,x,y],a_2)}_{A}$$
$$+ \underbrace{\Big(Q_\alpha^{n+1}([b,x,y],a_2) - Q_\alpha^{n+1}([b,x,y],a_1) \Big)}_{\geq 0 \text{ By optimality}}$$
$$\underbrace{-Q_\alpha^{n+1}([b,x,y],a_1) + Q_\alpha^{n+1}([b,x,y],a_0)}_{\geq 0 \text{ By optimality}}$$
$$- \underbrace{\Big(Q_\alpha^{n+1}([b,x,y],a_0) - Q_\alpha^{n+1}([b-1,x,y],a_0) \Big)}_{B}$$

By (11) and induction hypothesis, it is easy to see that

$$A \geq \sum_{x',y'} P_X(x^{'}|x) P_Y(y^{'}|y) \Big[V_\alpha^n([b+y, x^{'}, y^{'}])$$
$$- V_\alpha^n([b+y-1, x^{'}, y^{'}]) \Big] \geq B$$

Hence $V_\alpha^{n+1}([b,x,y])$ has increasing differences in the buffer state b.

5. REFERENCES

[1] H. S. Wang and N. Moayeri, "Finite-state Markov channel - A useful model for radio communications channels," *IEEE Trans. Vehicular Technology*, vol. 44, no. 1, pp. 163–171, 1995.

[2] A. Zhang and S. A. Kassam, "Finite-state Markov model for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, pp. 1688–1692, November 1999.

[3] F. J. Beutler and K. W. Ross, "Optimal Policies for Controlled Markov Chains with a Constraint," *Journal of Mathematical Analysis and Applications*, vol. 112, pp. 236–252, 1985.

[4] E. Altman, *Constrained Markov Decision Processes*, Chapman and Hall, London, 1999.

[5] L. I. Sennott, "Average Cost Optimal Stationary Policies in Infinite State Markov Decision Processes with Unbounded Costs," *Operations Research*, vol. 37, no. 4, pp. 626–633, July-August 1989.

[6] L. I. Sennott, "Value iteration in countable state average cost Markov Decision Processes with unbounded costs," *Annals of Operations Research*, vol. 28, pp. 261–272, 1991.

[7] R. Cavazos-Cadena, "Undiscounted Value Iteration in Stable Markov Decision Chains with Bounded Rewards," *Journal of Mathematical Systems, Estimation, and Control*, vol. 6, no. 2, pp. 1–34, 1996.

[8] D. M. Topkis, *Supermodularity and Complementarity*, Princeton University Press, 1998.

[9] D.V. Djonin and V. Krishnamurthy, " MIMO Transmission Control in Fading Channels - A Constrained Markov Decision Process Formulation with Monotone Policies," *IEEE Trans. Signal Proc.*, 2006, To appear.

[10] P. Bremaud, *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*, Springer-Verlag, 1999.

[11] S. Ross, *Introduction to Stochastic Dynamic Programming*, Academic Press, San Diego, California., 1983.