PSPACE hardness of approximating the capacity of time-invariant Markov channels with perfect feedback

MUKUL AGARWAL

It is proved that approximating the capacity of a time-invariant Markov channel with perfect feedback is PSPACE hard.

1. Introduction

In this paper, it is proved that approximating the capacity of a Markov channel with perfect feedback is PSPACE-hard. By 'approximating,' we mean, computing the capacity within a certain given additive error e. A class of channels will be constructed for which it will be proved that approximating the capacity to within 0.1 bits of the correct capacity is PSPACE-hard.

The observation that approximating the capacity of a Markov channel with perfect feedback can be formulated as a stochastic dynamic programming problem with partial observations thereby connecting it to the work of Tsitsiklis and Papadimitriou on the complexity of Partially Observed Markov decision processes [1] is what is novel here along with the result.

The authors of [1] demonstrate that the complexity of Markov decision processes with partially observed states is PSPACE-hard. The constructions and arguments in our paper depend significantly on, and are in many cases, the same as the arguments in [1]. The output of the channel in this paper is the partial state information in [1]. By carefully chosing the reset probabilities from the final decision states, we are able to map the complexity result about Markov decision process in [1] to result concerning the hardness of capacity approximation in this paper. The utter simplicity with which a result in control theory can be 'transported' into a result in communication theory can be seen.

For the relevant background on complexity theory, see [1] and references therein. An understanding of Section 4 in [1], in particular, the statement of Theorem 6 and its proof is needed to understand the proof here. Once this

is understood, and the reader has an understanding of information theory and basic probability theory, the proofs presented here can be understood.

2. Literature survey

In [1], the complexity of Markov decision processes and partially observed Markov decision processes has been considered and in [2], the Witsenhausen and team decision problems have been considered. In both these papers, it is proved that there are problems in each category which are hard. See further references therein for problems which have been proved to be hard, especially in a control setting. There are various papers in information theory demostrating hardness of source and channel coding algorithms and code constructions, see, for example [3], [4], and [5]. In [3], it has been proved that the generalized Lloyd-Max algorithm is NP-complete. In [4], it has been proved that general decoding problem for linear codes and the general problem of finding the weights of a linear code are both NP-complete. In [5], the problem of encoding complexity of network coding is considered, and one of the results in this paper is that approximating the minimum number of encoding nodes required for the design of multicast coding networks is NP-hard.

In this paper, it is proved that even just approximately calculating the capacity to within a constant additive positive number is PSPACE-hard for the problem of Markov channel with perfect feedback, and thus, for a general network, the problem of approximating the capacity region is PSPACE-hard.

3. The result

This section contains the channel construction, lemmas and the theorem.

For literature on Markov channels with feedback, the reader is referred to [6] and references therein, though this paper is not particularly needed to understand our paper.

3.1. Channel construction and reliable communication

Starting from any quantified formula $Q = \exists x_1 \forall x_2 \exists x_3 \ldots \forall x_n F(x_1, x_2, \ldots, x_n)$ with n variables and m clauses $C_1, C_2, \ldots, C_m$, where F is in conjunctive normal form, we construct a channel as follows:

Channel states $s_0, A_{ij}, A'_{ij}, T_{ij}, T'_{ij}, F_{ij}, F'_{ij}$; sets $A_j, A'_j, T_j, T'_j, F_j, F'_j$ $1 \leq i \leq m, 1 \leq j \leq n$ are the same as those in [1]. The states $A_{i,n+1}$ are lumped into a single state A_{n+1} and the states $A'_{i,n+1}$ are lumped into a

single state A'_{n+1}. Also, there is no terminating state, and from A_{n+1}, the channel goes back to itself or to s_0, and from A'_{n+1}, the channel goes back to itself or to s_0, in a way described later.

The state transitions, which as stated above, are exactly the same as in [1] (but for the exception stated above). In order to make the paper self-contained, we reproduce the same here.

Based on each clause C_i of Q and variable x_j there are 6 states, A_{ij}, A'_{ij}, T_{ij}, T'_{ij}, F_{ij}, F'_{ij}. There are also 2 additional states A_{n+1} and A'_{n+1}.

The initial state s_0 should be thought of as a set. For each variable j, the states $A_{ij}, 1 \leq i \leq m$ form a set A_i, and the states $A'_{ij}, 1 \leq j \leq m$ form a set A'_{ij} and similarly, the sets T_i, T'_i, F_i, F'_i are defined (this will be the partial state information, as we shall see below).

The channel can transition from s_0 to states $A'_{i1}, 1 \leq i \leq m$ with equal probability. If x_j is an existential variable, there are two possible state transitions out of the set A_j leading with certainty from A_{ij} to T_{ij} and F_{ij} respectively, and similarly, there are two transitions out of the set A'_j leading with certainty from A'_{ij} to T'_{ij} and F'_{ij} respectively. If x_j is a universal variable, there is one transition out of the set A_j leading with equal probability from A_{ij} to T_{ij} and F_{ij} and similarly, there is one transition out of the set A'_j leading with equal probability from A'_{ij} to T'_{ij} and F'_{ij}. From the sets T_j, F_j, T'_j, F'_j sets, there is only one transition which leads with certainty from T_{ij}, F_{ij}, T'_{ij} and F'_{ij} to (respectively) $A_{i,j+1}, A_{i,j+1}, A'_{i,j+1}, A'_{i,j+1}$ with two exceptions: If x_j appears positively in C_i, the transition from T'_{ij} is to $A_{i,j+1}$ instead of $A'_{i,j+1}$ and if x_j appears negatively, the transition from $F'_{i,j}$ is to $A_{i,j+1}$. When the channel reaches the state $A_{i,n+1}$, as stated above, the states $A_{i,n+1}$ are lumped into a single state A_{n+1}, and the channel stays in A_{n+1} with probability $1 - p$ and transitions back to the state s_0 with probability p for a value of p stated later. Similarly, when the channel reaches state $A'_{i,n+1}$, as stated previously, the states $A'_{i,n+1}$ are lumped into a single state A'_{n+1}, and the channel stays at state A'_{n+1} with probability $1 - q$ and transitions to the state s_0 with probability q, for a value of q stated later.

The state A_{n+1} will be called 'good' state and the rest of the states will be called 'bad' states. The reason for this will become clear below.

The output of the channel is the partial state information, that is, one of $s_0, A_i, A'_i, T_i, T'_i, F_i, F'_i, 1 \leq i \leq n$, and A_{n+1}, A'_{n+1} along with an output bit either 0 or 1 depending on channel input and channel functioning. Thus, the output space of the channel is $\mathbb{O} \triangleq \{s_0, \{T_i\}, \{T'_i\}, \{F_i\}, \{F'_i\}\{A_i\}, \{A'_i\}, 1 \leq i \leq n, A_{n+1}, A'_{n+1}\}\} \times \{0, 1\}$.

The output of the channel is 'fed back' directly to the encoder *without* delay.

The input to the channel is $\{D_1, D_2\} \times \{0, 1\}$. Intuitively, this should be thought of as a bit being transmitted through 0 or 1 and D_1 and D_2 determine the state transition (note that there are at most 2 possible state transitions out of a state). Of course, another policy can be used instead of transmitting the information bit and the state transition information; the above is just an intuitive way of thinking about the channel input.

The input to the encoder is a sequence of bits, each bit taking a value of 1 with probability $\frac{1}{2}$ and 0 with probability $\frac{1}{2}$, along with the output of the channel, which, as stated above, is 'fed back' directly to the encoder without delay. At each time, the bit input to the encoder should be thought of as a single bit, the bit which has not yet been communicated. Thus, the input space of the encoder is $\{0, 1\} \times \mathbb{O} = \{0, 1\} \times \{s_0, \{T_i\}, \{T_i'\}, \{F_i\}, \{F_i'\}\{A_i\}, \{A_i'\}, 1 \le i \le n, A_{n+1}, A_{n+1}'\} \times \{0, 1\}$.

Based on all past inputs (a bit stream and all past feedback from the channel output), the encoder makes an encoding and 'feeds it' into the channel.

It will be assumed that the channel starts in state s_0. Note that if the channel starts in state s_0, in $2n + 1$ units of time, it reaches either state A_{n+1} or state A_{n+1}'.

For the purpose of understanding, it is best to think of the problem as a sequential problem where a set of bits 'enter' the encoder at a certain rate R which causes an encoding and the channel produces the outputs from which a decoding needs to happen. A rate R is achievable if, for every δ, $\exists t_\delta$ such that for $t > t_\delta$, Rt bits can be communicated and the average error, that is, $\Pr(\hat{B}_{tR} \ne B_{tR})$ is less than δ, where B_{tR} denotes the bit input upto time t and $\hat{B}_{tR}$ is the corresponding decoding.

Let $p = 2^{-(mn)^{100}}$ and $q = 2^{-(mn)^{200}}$.

3.2. Lemmas, theorem and proofs

As has been stated previously, assume, in what follows, that the channel starts in state s_0.

Lemma 1. *Given $\epsilon > 0$. Then, $\exists m_0, n_0$, depending only on ϵ such that for $m > m_0, n > n_0$, if Q is true, capacity of the channel corresponding to Q is larger than $1 - \epsilon$.*

Proof. By [1], Q is true implies that we can choose the channel input (decisions in [1]) so that we always end up in A_{n+1}, not A_{n+1}'. The transitions from s_0 to A_{n+1} takes $2n + 1$ units of time where at worst, no bits can be

communicated, and the channel stays in state A_{n+1} for an average of order of magnitude $2^{(mn)^{100}}$ number of transitions, where 1 bit is transmitted noiselessly per channel use. By these considerations, it follows that given any $\kappa > 0$, $\exists\ m_1, n_1$ such that for $m > m_1$, $n > n_1$, the stationary distribution of the state A_{n+1} of the Markov chain with the above chosen channel inputs is $> 1 - \kappa$. By use of the ergodic theory for Markov chains, the lemma follows. $\square$

Lemma 2. *Given $\alpha > 0$. Then, $\exists m_0, n_0$ sufficiently large, depending only on α such that for $m > m_0, n > n_0$, the capacity of the channel corresponding to the formula Q is larger than α implies that Q is true.*

Proof. If there was some way for the channel to enter the state A'_{n+1} irrespective of the decisions, this would happen with probability at least $\frac{2^{-n}}{m}$ (see [1]), and then, the channel stays in this state for an average of an order of magnitude of $2^{(mn)^{200}}$ amount of time (1 unit of time refers to one state transition). Note that there is no transmission of information possible in state A'_{n+1}. Even if the channel ended in A_{n+1} with the rest of the probability $1 - \frac{2^{-n}}{m}$, the average order of magnitude amount of time the channel stays in state A_{n+1} is $2^{(mn)^{100}}$ which is 'much less' than $2^{(mn)^{200}}\frac{2^{-n}}{m}$. Also, there is the $2n + 1$ units of time when the channel transitions from s_0 to A_{n+1} or A'_{n+1}, which is 'negligible'. It follows that given any $\lambda > 0$, $\exists m_2, n_2$ such that for $m > m_2$, $n > n_2$, the fraction of time the channel spends in states A'_{n+1} is $> (1 - \lambda)$ with high probability. Finally, note that the amount of transmission of information during the $(2n + 1)$ units of time when the channel transitions from s_0 to A_{n+1} or A'_{n+1} is at most $\log\lceil 6mn + 3\rceil$. This is because the output space to the channel has cardinality $6mn + 3$. By taking into account the above numbers, it follows, then, that if the channel could enter A'_{n+1}, there exist m_0, n_0 sufficiently large such that the capacity of the channel will be less than α which will contradict the assumption on the channel. It follows, then, that there is a set of decisions (channel inputs) for which the channel never enters the state A'_{n+1} which implies, by [1], that the formula is true. $\square$

Theorem 1. *Computing the capacity of this set of Markov channels (the set of channels formed by taking a channel corresponding to each formula Q where m, n and the particular formula can be any positive integers) when perfect feedback is available, to within an accuracy of 0.1 bits per channel use, is PSPACE hard.*

Proof. If the capacity of this set of Markov channels with feedback could be computed to within an accuracy of 0.1, it would be known whether the capacity of the channel is less than 0.2 or larger than 0.8. This would imply, from the previous lemmas, that we would know, for sufficiently large m, n, whether Q is true or not, and this problem is PSPACE hard. $\square$

4. Comments

Comment 1. *It has been assumed that the channel starts in state s_0. This is only for simplicity of presentation. Minor modifications can be made in order to make the channel start in any state.*

Comment 2. *In addition, a transition from state $A_{i,n+1}$ to $A'_{i,n+1}$ with a probability $2^{-(mn)^{500}}$ could be added, and thus, from state $A_{i,n+1}$ to state s_0 with probability $1 - 2^{-(mn)^{500}} - 2^{-(mn)^{100}}$ to make the picture a little more realistic. Also, the bound in Lemma 2 concerning the information transmission from input to output, is $\log\lceil 6nm + 3\rceil$; however, a bound should also be possible on the cardinality of the input space of the channel.*

Comment 3. *There is nothing special about the number 0.1 in Theorem 1. Also, p and q need not be doubly exponential. Doubly exponentials work, and for that reason, they have been chosen this way.*

Comment 4. PSPACE *hardness implies* NP *hardness and thus, the problem dealt with in this paper is also* NP *hard.*

Comment 5. *A block-coding model can be considered instead of a sequential model. For the purpose of intuition, it is best to think of a sequential model.*

5. Importance of the result

In information theory, one important research direction is to find single-letter characterizations, appropriately defined, for capacity regions of networks. However, it is not concretely known whether single letter characterizations exist in general. An example is the two-way channel [7]. In order to prove that in general, there is no single letter characterization, all one needs is an example for which it is not possible to get one.

   This paper takes a different view-point and firmly establishes a limit from the view-point of complexity theory by providing an example for which

approximating the capacity of a general network is indeed hard in the sense of PSPACE-hardness.

6. Recap and research directions

A class of Markov channels with perfect feedback was constructed for which it was proved that approximating the capacity to within 0.1 bits is PSPACE hard.

It would be worthwhile exploring the application of this proof idea to other channels with potentially noisy feedback, channels with perfect state information, and networks in general.

It would be helpful to see whether this result puts restrictions on the kind of single-letter characterizations there may exist for capacity regions of networks; for example, this paper will rule out certain characterizations which can be approximated in a way that is not PSPACE-hard.

The simplicity with which a result in which a result in control theory has been used to prove a result in information theory may be noted and further possibilities of the same may be explored.

7. Acknowledgments

References

[1] C. H. Papadimitriou and J. N. Tsitsiklis, *The complexity of Markov Decision Processes*, Mathematics of Operations Research **12** (1987), no. 3.

[2] C. H. Papadimitriou and J. N. Tsitsiklis, *Intractable problems in control theory*, SIAM Journal of Control and Optimization **24** (1986), no. 4.

[3] M. R. Garey, D. S. Johnson, and H. S. Witsenhausen, *The complexity of the generalized Lloyd-Max problem*, IEEE Transactions on Information Theory **IT-28** (1982), no. 2.

[4] E. Berlekamp, R. McEliece, and H. van Tilborg, *On the inherent intractability of certain coding problems*, IEEE Transactions on Information Theory **24** (1978), no. 3.

[5] M. Langberg, A. Sprintson, and J. Bruck, *The encoding complexity of network coding*, IEEE Transactions on Information Theory **52** (2006), no. 6.

[6] S. Tatikonda and S. K. Mitter, *Capacity of channels with feedback*, IEEE Transactions on Information Theory **55** (2009), no. 11.

[7] C. E. Shannon, *Two-way communication channels*, Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics, University of California Press, 1961, pp. 611–644.

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING
BOSTON UNIVERSITY, BOSTON, MA 02215, USA
E-mail address: magar@alum.mit.edu