DINCON 2017
CONFERÊNCIA BRASILEIRA DE DINÂMICA, CONTROLE E APLICAÇÕES
30 de outubro a 01 de novembro de 2017 – São José do Rio Preto/SP

Transfer on Count-based Quadratic Control of Markov Jump Linear Systems with Unknown Transition Probabilities

Rafael Lemes Beirigo[1]
Coordenação de Métodos Matemáticos e Computacionais, LNCC, Petrópolis, RJ
Marcos Garcia Todorov[2]
Coordenação de Métodos Matemáticos e Computacionais, LNCC, Petrópolis, RJ
André da Motta Salles Barreto[3]
Coordenação de Métodos Matemáticos e Computacionais, LNCC, Petrópolis, RJ
Google DeepMind, London, UK

Abstract. This paper proposes a transfer strategy to accelerate the model-based optimal quadratic control of discrete-time Markov jump linear systems (MJLS) with unknown Markov chain transition probabilities, but complete information of the jump process. Our approach is based upon a recently developed adaptive control strategy that incrementally builds a transition model via maximum-likelihood estimation, based on online measurements of the Markov chain, and uses it to adjust the current policy in a certainty equivalence fashion. Despite the advantages presented by that strategy, it relies on repetitive executions of a subroutine used to obtain the policy by solving the subjacent Riccati equation, which can be computationally expensive. Here we propose an attempt to mitigate this computational cost by *transferring* intermediate solutions across subsequent calls to the subroutine that solves the Riccati equation during the model improvement steps and policy calculations. The method's performance is illustrated on a numerical example regarding Samuelson's macroeconomic model. The experimental results suggest the method was able to incrementally decrease the computational cost needed to calculate the optimal policy approximations.

Keywords. Markov jump linear systems, adaptive control, learning, transfer.

1 Introduction

When considering the complexity of contemporary control systems, serious constraints emerge that need to be taken into account in the design process, such as the possibility of failures or, more generally, of the susceptibility of these systems to *abrupt changes*.

[1]rafaelb@lncc.br, Coord. de Aperf. de Pessoal de Nível Superior – CAPES, Grant 1528969/2015-5.
[2]todorov@lncc.br, Brazilian National Research Council – CNPq, Grant 458456/2014-4
[3]amsb@lncc.br, Brazilian National Research Council – CNPq, Grant 461739/2014-3

Notwithstanding the vast amount of related literature which came out in recent years (see [7] for a recent account on the subject), several open problems still exist.

This paper is concerned with Markov Jump Linear Systems (MJLS), a particular class of stochastic switching systems that have received increasing attention through the last four decades or so, with applications spanning solar power stations, wireless communications, robotics, flight systems, and economics, for instance. In their simplest form, MJLS are characterized by the switching, orchestrated by a finite-state homogeneous Markov chain, within an ensemble of linear state-space systems (each of which represents an individual operating mode). The increasing interest in this class of systems stems from a well-balanced combination of solid theory and applications in several fields, such as flight systems, networked control, robotics, and economics, for instance. A recent reference on the subject is [6]. Here we consider the quadratic optimal control of Markov jump linear systems, also known as the JLQ problem [4], in a scenario where the transition probabilities of the Markov chain are uncertain (possibly, entirely unknown).

The precise knowledge of the transition probabilities of the Markov chain has long been recognized in the MJLS literature to be a rather stringent assumption, and much effort has been devoted towards relaxing it. To this end, several approaches such as *polytopic uncertainty*, *multi-simplex* setups, the *partially known* case, and the *norm-bounded* setting, among others, have been investigated (a state-of-the-art textbook is [9]). More recently, [1] proposed a method that incrementally builds an online approximation of the Markov chain transition probabilities and uses this approximate model[4] to calculate intermediate better policies by solving a certainty equivalence-based Riccati equation. Despite the advantages presented by this strategy, it needs to carry out value iteration at each control update step, and this can be computationally intensive. To circumvent this problem, in this paper we propose a modification to the method introduced in [1], that saves the intermediate solutions of the Riccati equations and subsequently uses this information in an attempt to accelerate them by means of a better *initialization.* This strategy can be categorized as *parameter transfer in the space of value functions* in the specialized literature. To the best of our knowledge, this is the first time that the subject of transfer is addressed in the MJLS context (see [8] for a survey on reinforcement learning approaches to transfer).

This paper is organized as follows. Some notation, basic definitions and relevant results from the literature are put together in Section 2. The problem statement and proposed solution are provided in Section 3. Section 4 features a numerical example that illustrates the proposed approach. Some concluding remarks and directions for future work are the subject of Section 5.

2 Background and notation

Consider a homogeneous Markov chain $\theta = \{\theta(k); k = 0, 1, 2, \ldots\}$ in a complete stochastic basis $(\Omega, \mathfrak{F}, \{\mathfrak{F}_k\}, \mathbb{P})$, with state space $\mathcal{S} \triangleq \{1, \ldots, N\}$, and transition matrix $P = [p_{ij}] \in \mathbb{R}^{N \times N}$, such that $p_{ij} \geq 0$, $\sum_{j=1}^{N} p_{ij} \equiv 1$ for all $i, j \in \mathcal{S}$. The mathematical expectation with respect to $\mathbb{P}$ will be represented as $\mathbb{E}$, and 1_A will stand for the indicator

[4] In this work a "transition model" denotes *an estimate of the true transition matrix P.*

of the event A (a random variable that can be equal to 1 or 0, depending on whether the event occurs or not). For later use, consider also the following operation:

$$\mathcal{E}_i(Y) \triangleq \sum_{j=1}^{N} p_{ij} Y_j, \quad i \in \mathcal{S}, \tag{1}$$

for any N-sequence of matrices of the form $Y = (Y_1, \ldots, Y_N)$. The set of all N-sequences of positive semidefinite matrices, i.e., those such that $Y_i \geq 0$ for all $i \in \mathcal{S}$, will be denoted $\mathbb{H}_N^{n+}$.

The subject of our study is the following discrete-time control system:

$$\begin{cases} x(k+1) = A_{\theta(k)} x(k) + B_{\theta(k)} u(k) \\ \quad\;\; z(k) = C_{\theta(k)} x(k) + D_{\theta(k)} u(k) \\ \quad\;\; x(0) = x_0 \in \mathbb{R}^n, \end{cases} \tag{2}$$

where $x = \{x(k) \in \mathbb{R}^{n_x};\, k = 0, 1, 2, \ldots\}$ is the state, $u = \{u(k) \in \mathbb{R}^{n_u};\, k = 0, 1, 2, \ldots\}$ is the control, and $z = \{z(k) \in \mathbb{R}^{n_z};\, k = 0, 1, 2, \ldots\}$ is a controlled output, whose performance we wish to optimize. For simplicity, the initial state x_0 is assumed deterministic. Throughout the paper, we shall write $A_{\theta(k)} = A_i$, $B_{\theta(k)} = B_i$, $C_{\theta(k)} = C_i$, and $D_{\theta(k)} = D_i$, whenever $\theta(k) = i$.

The quadratic optimal control problem of system (2), following the approach described in details in [5, Section 4.3], consists of finding a control that minimizes the *infinite horizon quadratic cost*, given by

$$\mathfrak{J}(\theta(0), x_0, u) = \sum_{k=0}^{\infty} \mathbb{E}\big(\|z(k)\|^2\big) \tag{3}$$

and that also pertains to the class $\mathcal{U} \triangleq \{u = \{u(k);\, k = 0, 1, \ldots\};\, u(k)$ *is* $\mathfrak{F}_k$*-measurable for each* k*, and such that* $\lim_{k\to\infty} \mathbb{E}\big(\|x(k)\|^2\big) = 0\}$. As the next lemma (borrowed from [5, Chapter 4]) shows, the search for such a control can be restricted without loss of generality to "Markovian" controllers of the form $u(k) = K_{\theta(k)} x(k)$.

Lemma 2.1. *Suppose there is a solution* $X = (X_1, \ldots, X_N) \in \mathbb{H}_N^{n+}$ *to the following set of coupled algebraic Riccati equations (CARE):*

$$X_i = A_i' \mathcal{E}_i(X) A_i + C_i' C_i - A_i' \mathcal{E}_i(X) B_i \big(B_i' \mathcal{E}_i(X) B_i + D_i' D_i\big)^{-1} B_i' \mathcal{E}_i(X) A_i, \tag{4}$$

$i \in \mathcal{S}$*, with* $\mathcal{E}_i(X) \triangleq \sum_{j=1}^{N} p_{ij} X_j$ *as in* (1)*. Suppose also that the control* $\bar{u} = \{\mathcal{F}_{\theta(k)}(X) x(k);$ $k = 0, 1, 2, \ldots\}$*, with:*

$$\mathcal{F}_i(X) \equiv -\big(B_i' \mathcal{E}_i(X) B_i + D_i' D_i\big)^{-1} B_i' \mathcal{E}_i(X) A_i, \tag{5}$$

stabilizes system (2) *in the mean square sense. Then we have* $\mathfrak{J}(\theta(0), x_0, \bar{u}) \equiv x_0' X_{\theta(0)} x_0 \leq \mathfrak{J}(\theta(0), x_0, u)$ *for all* $u \in \mathcal{U}$. $\triangledown$

In scenarios where P is uncertain, the obvious shortcoming of the preceding lemma is that, in order to obtain X, one must know the true transition probability matrix P beforehand. In order to overcome this, [1] replaces P by a transition model of the form $\widehat{P}^{(k)} = (1-\alpha_k)\widehat{P}^{(k-1)} + \alpha_k Q^{(k)}$, where $Q^{(k)}$ is a maximum-likelihood estimate of P on the basis of the available history of the process θ up until the instant k. This leads us to the Riccati equations

$$X_i^t = A_i'\mathcal{E}_i^t(X^t)A_i + C_i'C_i - A_i'\mathcal{E}_i^t(X^t)B_i\big(B_i'\mathcal{E}_i^t(X^t)B_i + D_i'D_i\big)^{-1}B_i'\mathcal{E}_i^t(X^t)A_i, \quad i \in \mathcal{S}, \tag{6}$$

where $\mathcal{E}_i^t(X^t) \triangleq \sum_{j=1}^N \widehat{p}_{ij}^{(k_t)} X_j^t$, which, in an online setting, must be solved at certain time steps $\{k_t;\, t = 0, 1, \ldots, \tau\}$. Whenever existing, the solution of (6) for a given t can be obtained iteratively by applying dynamic programming [3], viz. by iterating as follows until numerical convergence:

$$X_i^t \leftarrow A_i'\mathcal{E}_i^t(X^t)A_i + C_i'C_i - A_i'\mathcal{E}_i^t(X^t)B_i\big(B_i'\mathcal{E}_i^t(X^t)B_i + D_i'D_i\big)^{-1}B_i'\mathcal{E}_i^t(X^t)A_i, \quad i \in \mathcal{S}. \tag{7}$$

For this process to be consistent, it is necessary to initialize X^t at the first iterate, which is commonly made by setting all its components to zero: $X_i^t = 0$, $i \in \mathcal{S}$. As it will be detailed in the next section, here we are interested on repeating this entire process for subsequent improvements on approximations of P, i.e., such iteration needs to be carried out for a succession of values of t. In this case, we propose initializing the tth iterate with the preceding Riccati solution, i.e., using $X^t \leftarrow X^{t-1}$ instead of $X^t \leftarrow 0$, for each $t > 1$.

3 Problem statement and proposed algorithm

In this section we consider a variant of the JLQ control problem posed in Section 2, in which the true transition matrix P is irreducible and, despite there being no perfect knowledge of it, a sample path of Markov states $\{\theta(0) = i_0, \theta(1) = i_1, \ldots, \theta(k) = i_k\}$ obeying P is available at the current time k (for each $k = 0, 1, \ldots$). We also assume that there exists a stabilizing solution to the CARE (4) we are trying to solve and all the remaining system parameters are perfectly known.

In this work we propose Algorithm 1, which adds to the one presented in [1] the capability of reusing previous solutions of the Riccati equation calculated in the step (5). The objective of this modification is to mitigate the computational cost incurred in step (5), by initializing the corresponding iterative process with a previous calculated approximation of X. The expected net result is an increased convergence speed in this step, thus accelerating the global process comprised on Algorithm 1.

Algorithm 1. *Given a strictly increasing sequence of time steps* $0 = k_0 < k_1 < k_2 < \cdots < k_\tau$ *and some bootstrapping horizon* $k_s \in \{k_0, \ldots, k_{\tau-1}\}$, *along with initial control gains* $K_i^0 \in \mathbb{R}^{n_u \times n}$, $i \in \mathcal{S}$, *optionally an initial estimate* $\widehat{P}$ *of* P, *and* $X_i^0 \equiv 0$ *for all* $i \in \mathcal{S}$, *let* $t \leftarrow 0$, $k \leftarrow k_t$, *and proceed as follows:*

(1) *Apply the control* $u(k) = K_{\theta(k)}^t x(k)$, *and observe the resulting transition* $(x(k), \theta(k)) \rightarrow (x(k+1), \theta(k+1))$.

(2) *Compute $Q^{(k+1)}$, the updated approximation of P through maximum-likelihood estimation.*

(3) *Increment $k \leftarrow k+1$; if $k < k_{t+1}$, go back to step (1).*

(4) *Let $\alpha_k = \min\{k/k_s, 1\}$, and compute the transition model $\widehat{P}^{(k)} = (1-\alpha_k)\widehat{P}^{(k-1)} + \alpha_k Q^{(k)}$. Also, increment $t \leftarrow t+1$.*

(5) *Initialize the Riccati decision variable as $X^t \leftarrow X^{t-1}$, and iterate (7) until convergence. In case a solution to this process can be found, let $K_i^t \equiv -\big(B_i'\mathcal{E}_i^t(X^t)B_i + D_i'D_i\big)^{-1}B_i'\mathcal{E}_i^t(X^t)A_i$ for each $i \in \mathcal{S}$. If such X^t can't be found, let $K_i^t \equiv K_i^{t-1}$ and $X_i^t \leftarrow X_i^{t-1}$.*

(6) *If $t = \tau$, end the algorithm. Otherwise, go back to step (1).*

Upon exit, return the control gains $K_1^\tau, \ldots, K_N^\tau$ from the last step. ∇

4 Example: Samuelson's macroeconomic model

Samuelson's macroeconomic model [2] analyses the business cycle with the assumptions that the consumption and investment intentions depend on the level and pace of growth of the economic activity, respectively. Another common name to this model is "multiplier-accelerator model", because it is based on the *Keynesian multiplier* to model consumption intentions and the *accelerator theory of investment* to model the investment intentions.

The classical state-space version of Samuelson's model has the following realization:

$$x(k+1) = \begin{bmatrix} 0 & 1 \\ -\alpha & 1-s-\alpha \end{bmatrix} x(k) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(k),$$

where s^{-1} and α correspond to the *multiplier* and *accelerator* coefficients, respectively. Utilizing historical data from the United States Department of Commerce, in [2] the parameters s and α were estimated and grouped into three classes, which define the three *modi operandi* of the system: $\theta(k) = 1$, when both s and α are in mid-range; $\theta(k) = 2$, when s is in low range or α is in high range; and $\theta(k) = 3$, when s is in high range or α is in low range. The transition probability matrix associated with the switching between these three operating modes, along with the system parameters used in this example, were extracted from [2].

To evaluate the proposed method, Algorithm 1 was applied to solve this problem with the following settings: $x_0 = [1\ 1]^T$, $0 = k_0 < k_1 = 10 < k_2 = 20 < \cdots < k_{\tau=100} = 1000$, $k_s = 0$ (so that P is fully unknown), and suboptimal control gains $K_i^0 \in \mathbb{R}^{n_u \times n}$. To account for the inherent randomness, this experiment was repeated 10^3 times with the same settings, but different realizations of the Markov chain. To evaluate the algorithm's performance, it was compared to its "baseline" counterpart: controlling the system using only the initial suboptimal gains K_i^0 and also with the method it is based upon, proposed in [1].

Figure 1 shows the mean number of Riccati iterations needed to calculate intermediate controls using the approximations of P, along with the corresponding standard deviations. As can be seen from this graph, by appling Algorithm 1 ("With transfer"), the number of iterations decreased with the time steps, while the application of the algorithm proposed in [1] ("Without transfer") spended approximately constant 30 Riccati iterations. This result shows that reusing previous Riccati solutions can indeed accelerate the convergence of the process. Figure 1D shows the first component of the state vector x over the 10^3 repetitions *versus* the time steps. As it can be seen from the graph, both controls were able to stabilize the system, presenting a slightly better performance if compared to the "static" counterpart, by converging faster, markedly at the initial time steps. Figures 1B and 1C show the infinity-norm matrix differences $E_P(k_{\hat{\imath}}) \triangleq \|P - \widehat{P}^{(k_{\hat{\imath}})}\|_\infty$ and $E_F(k_{\hat{\imath}}) \triangleq \sum_{i=1}^{N} \|\mathcal{F}_i(X) - K_i^{(k_{\hat{\imath}})}\|_\infty$, both *versus* the time steps $k_{\hat{\imath}}$ (with $\mathcal{F}_i(X)$ as in (5)). As E_P and E_F correspond to the errors on the approximations of P and $\mathcal{F}$, respectively, one can see from these graphs that Algorithm 1 was able to improve the approximations of both P and $\mathcal{F}$ as the time steps $k_{\hat{\imath}}$ increased. These results show that Algorithm 1 was able to accelerate the process of adaptively improving the control gains by using the sampling process to refine the transition model, and then using this model to update the gains.

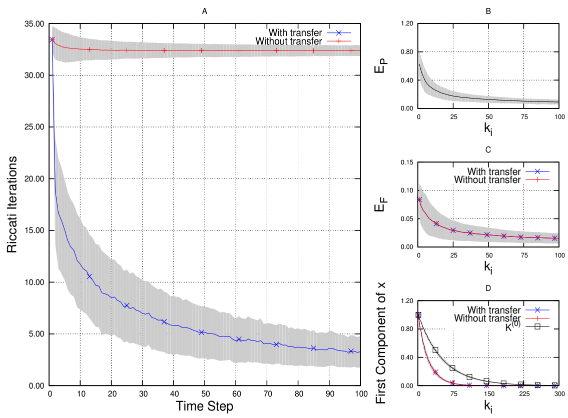

Figure 1: Application of Algorithm 1 to solve the MJLS version of Samuelson's macroeconomic model. Figure A shows the number of Riccati iterations needed to calculate the intermediate policies, based on the transition model approximations. Figure D shows the mean of the first component of the state vector x in three systems, one controlled by the sub-optimal control gains K^0 and the other two by the control gains sequence generated by Algorithm 1 and the one it is based upon [1]. Figures B and C correspond to the errors on the approximation of P and $\mathcal{F}(X)$ as in (5), respectively (along with the corresponding variances), both *versus* the time steps.

5 Concluding remarks

In this paper, we proposed a transfer strategy to accelerate a model-based method for solving the optimal quadratic control of discrete-time Markov jump linear systems (MJLS) when the transition probabilities of the Markov chain are uncertain, yet the jump process can be perfectly measured. The proposed method consisted on a modification of the iterative process used to obtain the solution of the Riccati equation for each control update step. As the transition model is improved and new policies may be calculated by successively solving Riccati equations, the preceding solutions are used to *initialize* this iterative process at the current step. We presented a numerical example, regarding Samuelson's macroeconomic model, where the proposed approach was shown to learn the optimal control with less computational cost.

The idea of reusing previous solutions on current and different control problems with positive transfer results may sound appealing, mostly if one considers the decrease on the computational demands. One possible extension to this work would be to build a *library* of previous Riccati solutions to different control problems. By facing a new problem, the control system could then *search* this library for a suitable initialization data. To improve this search process, it would be interesting to define a problem *metric*, which classifies problems according to structure similarity, thus allowing for better initialization choices.

References

[1] R. L. Beirigo, M. G. Todorov, and A. M. S. Barreto. Count-based quadratic control of markov jump linear systems with unknown transition probabilities. In Proc. of the 56th IEEE Conference on Decision & Control, Melbourne, Australia, (2017) submitted.

[2] W. P. Blair and D. D. Sworder. Feedback control of a class of linear discrete systems with jump parameters and quadratic cost criteria. International Journal of Control, 21:833-844, (1975).

[3] G. C. Chow. Analysis and Control of Dynamic Economic Systems. John Wiley Sons, Inc, New York, (1975).

[4] O. L. V. Costa and M. D. Fragoso. Discrete-time LQ-optimal control problems for infinite Markov jump parameter systems. IEEE Trans. Automat. Control, 40(12):2076-2088, (1995).

[5] O. L. V. Costa, M. D. Fragoso, and R. P. Marques. Discrete-Time Markov Jump Linear Systems. Probability and Its Applications. Springer-Verlag, New York, (2005).

[6] O. L. V. Costa, M. D. Fragoso, and M. G. Todorov. Continuous-Time Markov Jump Linear Systems. Probability and Its Applications. Springer-Verlag, Heidelberg, (2013).

[7] R. Goebel, R. G. Sanfelice, and A. R. Teel. Hybrid Dynamical Systems: Modeling, Stability, and Robustness. Princeton University Press, Princeton, NJ, (2012).

[8] A. Lazaric. Transfer in reinforcement learning: a framework and a survey. In Reinforcement Learning, State of the Art, pages 143-174. Springer-Verlag, (2012).

[9] L. Zhang, T. Yang, P. Shi, and Y. Zhu. Analysis and Design of Markov Jump Systems with Complex Transition Probabilities. Springer, Switzerland, (2016).