J Electr Eng Technol.2018; 13(5): 2074-2079
http://doi.org/10.5370/JEET.2018.13.5.2074

ISSN(Print) 1975-0102
ISSN(Online) 2093-7423

Optimal Controller Design of One Link Inverted Pendulum Using Dynamic Programming and Discrete Cosine Transform

Namryul Kim* and Bumjoo Lee†

Abstract – Global state space's optimal policy is used for offline controller in the form of table by using Dynamic Programming. If an optimal policy table has a large amount of control data, it is difficult to use the system in a low capacity system. To resolve these problem, controller using the compressed optimal policy table is proposed in this paper. A DCT is used for compression method and the cosine function is used as a basis. The size of cosine function decreased as the frequency increased. In other words, an essential information which is used for restoration is concentrated in the low frequency band and a value of small size that belong to a high frequency band could be discarded by quantization because high frequency's information doesn't have a big effect on restoration. Therefore, memory could be largely reduced by removing the information. The compressed output is stored in memory of embedded system in offline and optimal control input which correspond to state of plant is computed by interpolation with Inverse DCT in online. To verify the performance of the proposed controller, computer simulation was accomplished with a one link inverted pendulum.

Keywords: Optimal policy, Dynamic programming, Compress, Discrete Cosine Transform (DCT), Inverse Discrete Cosine Transform (IDCT).

1. Introduction

Dynamic Programming is an algorithm which can find optimal policy based on accurate model. Above all, value iteration is a method of deriving the optimal policy that optimizes the value function through iterative calculation of Bellman Optimality Equation [1-3]. Since the entire control input is compared for each state of the global state space, a large cost may be occurred in terms of time in some cases. Therefore, the initial value function and policy are created with the Linear Quadratic Regulator (LQR) [2], [4]. Then, the discretized control input is randomly selected and applied to the global state space for value iteration. The optimal policy is determined as a more optimal value by comparing the initial policy and the policy obtained by value iteration for each iteration [5-7]. If the error between the previous value function and the current value function satisfies the set value, the iteration is terminated. The optimal policy of the global state space obtained by the above procedure is stored in the form of a table. In the case of a dynamic system with a large non-linearity depending on the problem, the state is discretized in high resolution because the dynamic system changes even in the smallness state variation. Therefore, if the optimal policy table having a large amount of control data is utilized in a low-capacity system such as an embedded system, extra costs may be occurred or difficulties may arise in use. To solve this problem, a controller using a compressed optimal policy table to reduce memory is proposed in this paper. In this paper, the optimal policy table is treated as still image and compressed. A DCT which widely used in the image and voice fields is used for compression method. After performing the DCT on the optimal policy table, an information needed for reconstruction is concentrated in the low frequency band and a data with little effect on the restoration is collected in the high frequency band. The data gathered in the high frequency band is removed due to quantization according to performance needs. The optimal policy table compressed by DCT and quantization is stored offline in the memory of the target system and the control input corresponding to the state of the plant is interpolated in real time by IDCT [8-10]. The performance of the controller according to the compression ratio is verified and analyzed through a computer simulation for a one link inverted pendulum system.

Preliminaries of Dynamic Programming and LQR are described in Section 2. DCT and IDCT for compression are described in Section 3, and computer simulation which is verified performance of proposed controller is treated in Section 4. Finally, conclusions are summarized in Section 5.

† Corresponding Author: Dept. of Electrical Engineering, Myongji University, Korea. (bjlee@mju.ac.kr)
* Dept. of Electrical Engineering, Myongji University, Korea. (kimnamryul@naver.com)
Received: October 23, 2017; Accepted: April 24, 2018

2. Preliminary

2.1 Dynamic Programming (DP)

Dynamic Programming is an algorithm based on a Principle of Optimality [11] and Markov Decision Process (MDP) which is a discrete time decision model composed

of state, action, state transition function, reward function and discount factor. A value function is used as a way of characterizing the policy. The optimal policy is obtained by repeatedly calculating the Bellman Optimal Equation so that all state in the global state space can store the largest or smallest value depending on the optimization problem.

2.2 Linear Quadratic Regulator (LQR)

The initial policy is created by the LQR which is one example of applying the maximum principle to a linear system before the optimal policy is obtained by using Dynamic Programming. The control gain of the input which optimizes the value function of the quadratic equation form is obtained using an Algebraic Riccati Equation (ARE). A performance measure in the LQR is defined as follows:

$$L = \frac{1}{2}x^T Qx + \frac{1}{2}u^T Ru \tag{1}$$

where Q and R are weight matrix of the state x and action u, respectively. In LQR, the termination cost ψ is not considered. Adopting co-state vector, $\lambda = Px$, control input is calculated as follows:

$$u = -R^{-1}B^T\lambda = (-R^{-1}B^TP)x \tag{2}$$

The optimal control gain K is equal to $-R^{-1}B^TP$ and the unknown quantity P can be obtained using the ARE.

$$PA + A^TP + Q - PBR^{-1}B^TP = 0 \tag{3}$$

2.3 Implementation of optimal controller

Optimal control algorithm is consist of two parts: off-line process and on-line process as shown in Fig. 1. After derivation of the dynamic equation and the cost function, optimal policy is calculated and updated by Optimization Programming method such as DP while off-line. During at control, control data is indexed and interpolated according to the feed-back state. In this paper, data compression scheme was adopted to reduce the data table size that is increased by the resolution of the state and the control.

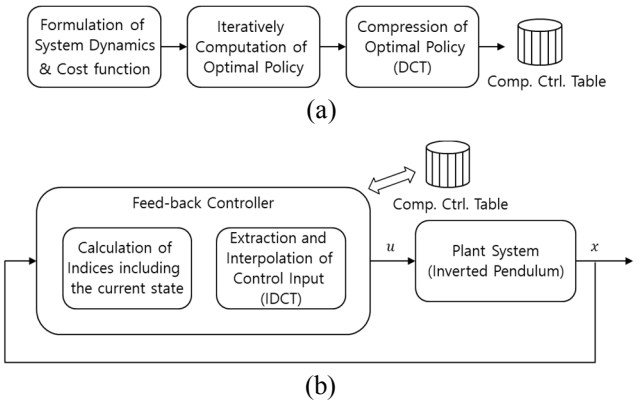


Fig. 1. Optimal control algorithm: (a) Off-line process: creation of control table; (b) On-line process: indexation and interpolation of control table

3. Compression Algorithm for Optimal Policy

3.1 Size Reduction of Control Data

Dynamic properties can also be significantly changed for very small state variation in some non-linear dynamic system. In the case, state discretization may be performed at high resolution. Thus, a large of amount of control data which is equal to optimal policy is generated depending on the number of state and resolution. The optimal policy is stored in the target system in the form of a table for use in the offline controller. If the target system is a low capacity system, for example, an embedded system can be difficult to use. This is because separate costs such as communication or memory expansion may occur for the use of the optimal policy table. There is a various way to solve the above problem. In everyday life, Large files can be compressed for transmission and reception. If the optimal policy table is compressed to reduce the memory, it can be used in a low capacity system. In the case of ideal compression, an original information which is equal to control data can be restored to the target system without distortion. Since the purpose is to use the controller, information's distortion should not occur. For this reason, this paper proposed a method to utilize the compressed optimal policy table for the controller. There are many methods to compress the optimal policy table. Discrete Fourier Transform and Discrete Cosine Transform are similar with each other and are widely used in various application, among them compression method for image and audio fields.

3.2 Compression by using DCT and IDCT

A Discrete Cosine Transform(DCT) is a sum of cosine functions that vibrates at different frequencies and represents a finite section of data. The result of the transformation is similar with a Discrete Fourier Transform (DFT) which outputs a complex number, but the DCT is relatively simple because it produces only real number. The most significant feature is the energy concentration phenomenon where the energy is concentrated in the low frequency band using the characteristic that the size of each coefficient decreases as the frequency of the cosine function becomes larger. Therefore, DCT is widely used for lossy compression of voice and image fields. The DCT can be divided into four types from 1 to 4. DCT-2 is used for image compression and DCT-3 is used for reconstruction. DCT-2 formula is used to compress the optimal policy table as follows:

$$x(n,m) = \alpha(u)\alpha(k)\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} x(n,m)\cos(\theta)\cos(\psi),$$

where

$$\theta = \frac{(2n+1)u\pi}{2N},\ \psi = \frac{(2m+1)k\pi}{2N},$$
$$\alpha(\sigma) = \begin{cases} \sqrt{1/N}, & \sigma = 0 \\ \sqrt{2/N}, & otherwise \end{cases} \quad (4)$$

$\alpha(u)$ and $\alpha(k)$ are scale factors for making the DCT matrix an orthogonal matrix. If the DCT matrix is made an orthogonal matrix, the transpose and inverse matrix are the same, so it is easy to compute for forward and inverse transform. The DCT-3, IDCT equation, is the same as the DCT-2 except that the scale factor in the series.

$$x(n,m) = \sum_{n=0}^{N-1}\sum_{m=0}^{N-1} \alpha(u)\alpha(k)X(u,k)\cos(\theta)\cos(\psi) \quad (5)$$

3.3 Characteristics of DCT and IDCT

Since information in the high frequency band is removed through quantization, the memory can be greatly reduced. Therefore, it can be used for a low capacity embedded system, so that the extra cost due to the memory can be reduced. The disadvantage is lossy compression with loss of information using quantization, and it is difficult to recover the original signal correctly because it cannot perform computation with infinite time and density and only uses real part without complex domain. Nonetheless, considering the visual phenomenon of people who sensitive to low frequency signal and insensitive to high frequency signals, the difference from the original is not very noticeable.

4. Computer Simulation

4.1 One link inverted pendulum

To verify the effectiveness of proposed algorithm, one link inverted pendulum is examined as shown in Fig. 2. The equation of motion for the system is given as follows:

$$(I + ml^2)\ddot{\theta} = mglsin\theta + \tau \quad (6)$$

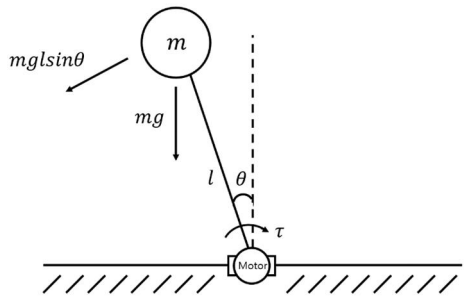


Fig. 2. One link inverted pendulum free body diagram

In this paper, the state is composed of angle ($-2\pi \leq \theta \leq \pi\ [rad]$) and angular velocity ($-20 \leq \dot{\theta} \leq 20\ [rad/sec]$). Also, the range of action is $-1.945 \leq \tau \leq 1.945$ [Nm]. The resolution for state and action is set to 2^{10}-1. Some simulation results were performed with a resolution of $2^9 - 1$. The value function indicates the optimization direction in the form of a concave quadratic function to minimize cost.

$$C_k = x^T Q x + u^T R u \quad (7)$$

where x and u mean state and action, Q and R mean state and action weight matrix, respectively. Since the weight of the state is set to be 10 times smaller than the weight of the action, small variation in the state can also be confirmed.

$$Q = diag\{\alpha, 1\},\ R = \beta \quad (8)$$

α is $0.1 \times$ Ts and β is Ts. In addition, this paper does not consider the loss of angular velocity. The following is a description of the simulation method. The DCT is performed on the optimal policy table obtained by the Dynamic Programming. Since the information of the high frequency band is not discarded through the quantization step, it can be called lossless compression. The quantization is conducted according to memory and performance requirements, and lossy compression results in which high frequency information is discarded are used for the controller. The controller starts the control from the given initial state and calculates the compressed optimal policy table in real time using the IDCT when the input corresponding to the state of the plant is required by trigonometric interpolation with a cosine function. Therefore, this paper confirms the performance of the controller by comparing the method of linear interpolation which be conducted on original policy and the triangular interpolation method using the IDCT for the lossless and lossy compressed policies.

4.2 Simulation results

The compression ratio of Table 1 is calculated as

Table 1. Capacity and cost according to compression ratio at resolution 2^9-1

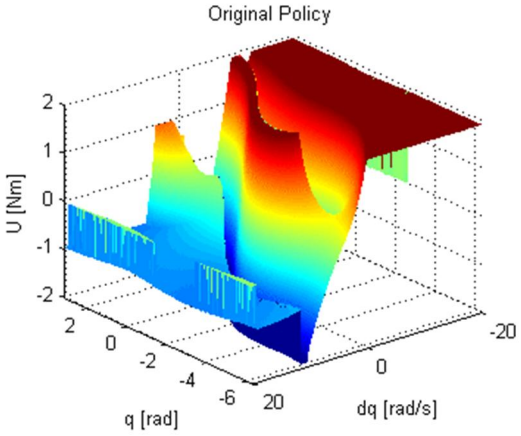
Index	Compression Ration [%]	Capacity [Bytes]	Cost
1	0	1,044,484	3.3333
2	10	940,060	3.3348
3	20	835,180	3.3336
4	30	730,060	3.3320
5	40	627,700	3.3315
6	50	521,220	3.3318
7	60	416,784	3.3326
8	70	312,840	3.3352
9	80	209,304	3.3355
10	90	104,424	3.3371

$\frac{A-B}{A} \times 100$, A is total number of data in original policy and B is Non-zero number of data in compressed policy. In addition, the capa city was calculated based on the 4 bytes data type float. When comparing the memory sizes of the original policy and the 50% compressed policy, a difference of about 2 times occurs. In extreme cases, a difference of about 10 times occurs between the original policy and the 90% compressed policy. The following figure shows the graphs and action trajectory of all policies including original policy and compressed policy.

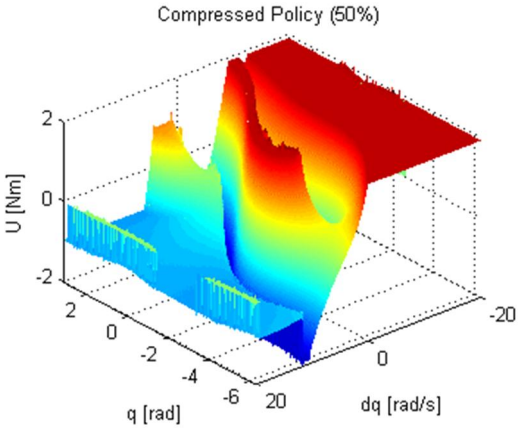
The original policy and the 50% compressed optimal policy picture are shown in Fig. 3. In Fig. 3, (a) and (b) are a little difference visually. Nevertheless, as shown in Table 1, memory is about twice the difference.

The angle and angular velocity trajectory according to the interpolation method shown in Fig. 4 50% compression policy's result which is equal to (c) in Fig. 4 converges to a goal state exactly as compared with the result of the original and lossless compression.

Fig. 5 shows the performance of the controller proposed in this paper. Since the value of the original policy which is equal to green line of (b) in Fig. 5 is divided into a negative number an d a positive number for a small state variation, the shape of the original policy is vibrated due to the characteristic of a linear interpolation method connecting two points. Compared with this, the proposed controller shows a smooth curve shape with no vibration in the transient state and steady state by using the triangular interpolation method using the cosine function.

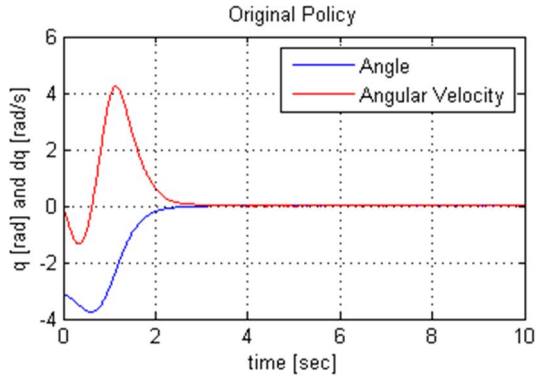


(a)

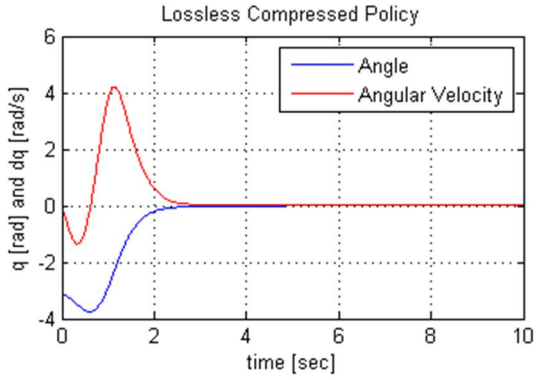


(b)

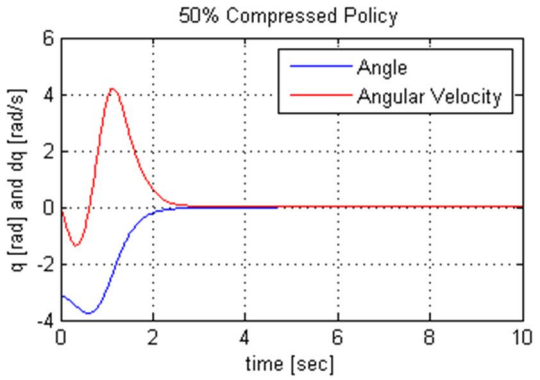
Fig. 3. Comparing the original optimal policy and the compressed optimal policy at resolution 2^{10}-1: (a) Original policy; (b) 50% compressed policy



(a)

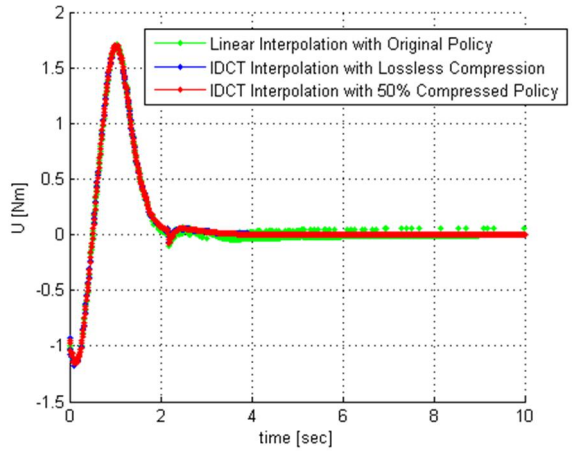


(b)

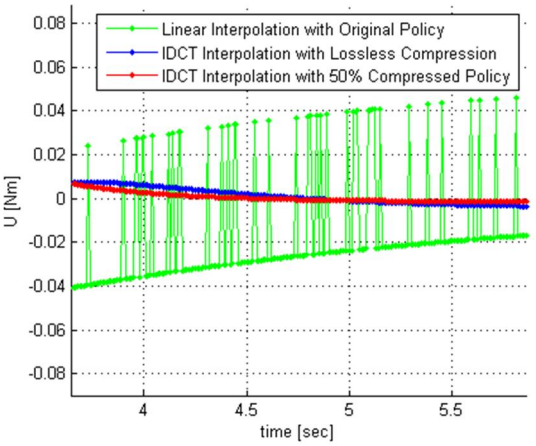


(c)

Fig. 4. Angle and angular velocity trajectory according to interpolation method: (a) Original policy using linear interpolation; (b) Lossless compressed policy using the proposed method; (c) 50% compressed policy using the proposed method



(a)



(b)

Fig. 5. Action trajectory according to the interpolation method: (a) All action trajectory; (b) Expansion of about 3.7 seconds to 5.8 seconds

5. Conclusion

In this paper, optimal policy is obtained by using Dynamic Programming based on Principle of Optimality and MDP, and it is stored in table form and used in offline controller. In the optimal policy table, the dynamic system with large non-linearity has a rapid variation of dynamic system even with the small state variation. Therefore, the optimal policy table has a huge amount of control data. Thus, a low capacity system is hard to use optimal policy table. To solve this problem, controller using a compressed optimal policy table is proposed in this paper. As a compression method, DCT which is widely used for lossy compression of video and audio fields is used. It is shown that energy concentration phenomenon, which is a feature of DCT, can reduce memory by eliminating data of high frequency band with little information needed for reconstruction through quantization.

The compressed optimal policy table is stored offline in the memory of the target system. The optimal control action corresponding to the state of the plant is provided by interpolation in real time through the IDCT. The result of the linear interpolation for the original optimal policy, the lossless compressed optimal policy table that has not undergone quantization, and the optimal policy table that has been compressed by 50% after the DCT are subjected to the triangular interpolation using the IDCT and the trajectory and torque trajectory results. In the torque trajectory, the result of original policy which applied linear interpolation method is shown to vibrate in transient and steady state. It is assumed that the original policy has extreme values both positive and negative for small state variation. This problem seems to be solved by increasing the resolution. In contrast, the results of the proposed controller show that the smooth curve shape with no vibration in the transient state and the steady state is confirmed by computer simulation since both the lossless compression and the 50% compressed policy use triangular interpolation using the cosine function. In actual experiments, the performance difference between the conventional method and the proposed controller will be confirmed.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2015R1C1A1A02037641).

References

[1] Busoniu, Lucian, et al., "Reinforcement learning and dynamic programming using function approximators," *CRC Press*, vol. 39, 2010, pp. 14-30.

[2] Jonathan P. How, "3. Dynamic programming : principle of optimality, dynamic programming, discrete LQR," *Principle of Optimal Control*, MIT OCW, pp. 1-27, 2008.

[3] Andrew Ng, "13. Reinforcement Learning and Control," *CS 229 : Machine Learning*, Stanford Univ., pp. 1-15, 2017.

[4] Michael Triantafyllou, "19. Linear Quadratic Regulator," *Maneuvering and Control of Surface and Underwater Vehicles*, MIT OCW, pp. 92-98, 2004.

[5] Atkeson, Christopher G, "Randomly sampling actions in dynamic programming, In *Approximate Dynamic Programming and Reinforcement Learning, 2007. ADPRL 2007. IEEE International Symposium on*, pp. 185-192, 2007.

[6] Atkeson, Chris and Benjamin Stephens, "Randomly sampling of states in dynamic programming," *Advances in neural information processing systems,* pp. 33-40, 2008.

[7] Atkeson, Christopher G. and Chenggang Liu, "Trajectory-based dynamic programming," *Modeling, Simulation and Optimization of Bipedal Walking Cognitive Systems Monographs*, pp. 1-15, 2013.
[8] Cabeen, K. and Gent, P., "Image Compression and the Discrete Cosine Transform," College of the Redwoods, Math 45, pp. 1-11, 1998.
[9] Andrew B. Watson, "Image Compression Using the Discrete Cosine Transform," *Mathmatica Journal*, vol. 4, pp. 81-88, 1994.
[10] Strang, G., "The discrete cosine transform," *SIAM review*, pp. 135-147, 1999.
[11] Bellman, Richard, "The theory of dynamic programming," RAND CORP SANTA MONICA CA, 1954, p. 1-8.

Namryul Kim received the B.S. and M.S. degrees in electrical engineering from Myongji University, Yongin, Korea, in 2016 and 2018, respectively. His research interests include the areas of control systems, especially in motor control algorithm

Bumjoo Lee received the B.S. degree in electrical engineering from Yonsei University, Seoul, Korea, in 2002, and the M.S. and Ph.D. degrees in electrical engineering from Korea Advanced Institute of Science and Technology (KAIST), Korea, in 2004 and 2008, respectively. Since 2012, he has been with the Department of Electrical Engineering, Myongji University, Korea, where he is currently an Assistant Professor. His research interests include the areas of Humanoid Robotics, especially in motion planning and control algorithm.