Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Robust Adaptive Control with Active Learning for Fed-Batch Process based on Approximate Dynamic Programming

Ha-Eun Byun*, Boeun Kim, Jay H. Lee***

**Department of Chemical and Biomolecular Engineering, Korea Advanced Institute of Science and Technology, 291 Daehak-ro, Yuseong-gu, Daejeon, 34141, Republic of Korea, (e-mail: {bhe2515, jayhlee}@kaist.ac.kr)*
*** Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison, WI 53706, USA (e-mail: bkim329@wisc.edu)}*

Abstract: Batch process is often subject to a high degree of uncertainty in raw material quality and other initial feedstock conditions. One of the key objectives in operating a batch process is achieving consistent performance and constraint satisfaction in the presence of these uncertainties. This study presents a method for optimal control of a fed-batch process, which can actively and robustly cope with system uncertainty. As in dual control, the method aims to achieve an optimal balance between control actions (exploitation) and probing actions (exploration), leading to improved process performance by actively reducing system uncertainty. An optimal solution of the dual control problem can be found by stochastic dynamic programming but it is computationally intractable in most practical cases. In this study, an approximate dynamic programming (ADP) method for solving the dual control problem is tailored to a batch process which involves non-stationary and nonlinear dynamics. Rewards are formulated to maximize a given end objective while satisfying path constraints. Performance of the ADP-based dual controller is tested on a fed-batch bioreactor with two uncertain parameters.

Keywords: Robust adaptive control, Dual control, Stochastic optimal control, Approximate dynamic programming, Fed-batch process, Model uncertainty

1. INTRODUCTION

Control of dynamic systems in the presence of uncertain parameters is of great interest in industrial chemical and biological processes. When measurements are available, the general approach is to estimate the unknown parameters with measurements using some state estimator (e.g., extended Kalman filter) and use the parameter estimates in a deterministic control framework, known as certainty equivalence control. To obtain informative data about the uncertain parameters and thus obtain better estimates, exploratory inputs to excite the system, i.e., probing inputs, may be needed. In standard certainty equivalence based control, however, the effect of control inputs on the future parameter estimates is not taken into account, such that the parameter learning is only accidental and thus passive (Wittenmark, 1995).

To actively learn about the uncertain system, control inputs should have a properly designed probing effect. Probing inputs may decrease the performance in short-term (when it conflicts with the control objective), but the improved knowledge about uncertainty can result in better control performance in the future. Thus, an optimal balancing between learning (by maximization of information content in the collected data) and control (by optimization of process performance) is needed, and this problem is called the 'dual control problem'. The dual controller actively reduces uncertainty to the extent necessary to achieve optimal control performance by incorporating probing naturally into control inputs (Heirung et al., 2019).

An optimal solution of the dual control problem can be found by stochastic dynamic programming (SDP) (Feldbaum, 1960), but it is computationally intractable in most practical cases, particularly for systems with continuous state space. To solve the dual control problem approximately, various approximate solution methods have been suggested, which can be classified as implicit and explicit approaches (Filatov, 2000). Implicit dual control methods directly approximate the original dual control problem to obtain actively adaptive suboptimal control policies that retain the dual properties of the system. On the other hand, explicit dual control methods reformulate the stochastic optimal control problem into a simpler one by introducing heuristic-based probing effects. Explicit dual control methods are easier to implement than implicit dual control methods, but they require proper tuning for balancing the trade-off between the control and probing actions.

A growing body of literature on applying ADP or RL to solve the dual control problem has demonstrated that the data-based learning approach can derive a superior control policy, which actively learns parameter values to give significant performance improvement, compared to the standard adaptive control with passive learning (Lee & Lee, 2009; Thompson & Cluett, 2005; Morinelly & Ydstie, 2016). The investigation of the data-based learning approach, however, has been limited to tracking control problems of continuous

processes. In this study, we present a scheme for robust implicit dual control of nonlinear and nonstationary batch processes, based on ADP.

Batch processes play an important role in chemical and biological industry, especially in manufacturing specialty chemicals, pharmaceuticals, and polymers, given its flexibility to meet various product specifications. An important objective in batch process operations is to achieve quality and quantity specifications of the product while satisfying process constraints. However, a high degree of uncertainty in raw materials and other initial charge conditions make it difficult to achieve such specifications consistently. Therefore, in order to compensate for the poor prior knowledge and thus satisfy the constraints and final specifications within finite batch time, the controller needs to actively learn and cope with the uncertainty.

This study tailors the ADP method to dual control of fed-batch processes with uncertain parameters. We adopt the idea of ADP approach proposed by Lee and Lee (2009), in which dynamic programming is solved on a restricted space of hyper-state sampled through stochastic closed-loop simulations performed with suboptimal control policies. For the batch process under uncertainty, formulation of the objective function in terms of economic optimization instead of set-point tracking can lead to a superior economic performance (Lucia et al., 2014). Thus, 'profit-to-go' in this study is formulated to maximize a given end objective while satisfying various path and end constraints, to optimally balance learning and control effects such as to maximize the overall economic gains. A case study of fed-batch bioreactor with two uncertain parameters demonstrates the performance of the ADP-based dual optimizing controller.

2. ADP-BASED ROBUST DUAL ADAPTIVE CONTROL

2.1 Problem Formulation

Consider the discrete-time stochastic system

$$x_{k+1} = f\left(x_k, u_k, \theta_k, e_k\right), \quad k \in \mathbb{N}, \tag{1}$$

where x_k is a state vector, u_k is a control input, θ_k is a vector of unknown parameters of the model, and e_k represents exogenous noises. The state x_k is assumed to be fully measured at each k and the model structure f is assumed to be known. The aim of control is to maximize the performance index represented as follow:

$$\max_{u_0,\cdots,u_{t_f-1}} E\left[\sum_{t=0}^{t_f-1} \phi\left(x_t, u_t\right) + \bar{\phi}_T\left(x_{t_f}\right)\right], \tag{2}$$

subject to

$$g_j\left(x_t\right) \le 0, \quad \forall j \in G, \tag{3}$$

$$u^L \le u \le u^U \tag{4}$$

where ϕ and $\bar{\phi}_T$ are stage-wise reward and terminal reward, respectively. The expectation E is taken over the distribution of θ and e . g_i denotes the path constraints which should be satisfied. The inputs are to be decided based on the measured state information, and so the problem is to find the optimal control policy $\pi = \{\pi_k\}_{k=0}^{t_f-1}$, where π_k is a map between state and control input at time k .

In the framework of dynamic programming, the optimal 'profit-to-go' function at time k can be represented as

$$J_k^*\left(\xi_k\right) = \max_{\pi} E\left[\sum_{t=k}^{t_f-1} \phi\left(x_t, u_t\right) + \bar{\phi}_T\left(x_{t_f}\right) \middle| \xi_k\right], \tag{5}$$

where hyper-state ξ_k is an extended state including the information about the uncertain parameters, i.e., the parameter estimates and their covariances, in addition to the x_k . J_k^* , which maps the hyper-state to the profit-to-go value under the optimal control, satisfies the following Bellman's optimality equation:

$$J_k^*\left(\xi_k\right) = \max_{u_k} E\left[\phi\left(x_k, u_k\right) + J_{k+1}^*\left(\xi_{k+1}\right) \middle| \xi_k\right]. \tag{6}$$

Once J_k^* is determined, the optimal control policy can be derived by solving

$$u_k = \pi^*\left(\xi_k\right) = \arg\max_{u_k} E\left[\phi\left(x_k, u_k\right) + J_{k+1}^*\left(\xi_{k+1}\right) \middle| \xi_k\right]. \tag{7}$$

For each evaluation of a candidate u_k , the expectation needs to be calculated, which involves the integration of the successor hyper-state ξ_{k+1} for all its possible value. To solve the Bellman equation numerically, the value iteration or policy iteration is performed after discretization of the hyper-state space. However, this is computationally intractable in most practical cases, especially when the hyper-state space is continuous.

2.2 ADP Algorithm

The ADP based approach proposed by Lee and Lee (2009) circumvents the curse-of-dimensionality of the traditional DP approach by solving the DP only for a restricted space of the hyper-state, sampled from Monte Carlo simulations of the closed-loop system with some suboptimal control policies. The same idea is adopted and tailored to the characteristics of a nonstationary batch process with path constraints. Construction and improvement of the profit-to-go approximation proceed as follows. Note that these steps are performed off-line and only the converged profit-to-go approximator is used on-line.

1) Perform Monte-Carlo runs of the closed-loop system with known suboptimal control policies, e.g., MPC. It is recommended to simulate several policies with different characteristics in order to cover a broad range of potential operating space.

2) For each state visited during the simulation runs, calculate the profit-to-go J_k^0 using the simulation data according to

$$J_k^0\left(\xi_k\right)=\sum_{t=k}^{t_f-1}\phi\left(x_k,u_k\right)+\bar{\phi}_T\left(x_{t_f}\right)\left|\xi_k\right. . \tag{8}$$

In this study, the satisfaction of path constraint is considered as the stage-wise reward:

$$\phi\left(x_k,u_k\right)=-\lambda\sum_{j\in G_{cv}}g_j\left(x_{k+1}\right). \tag{9}$$

G_{cv} is a set of indexes of constraints violated. λ is a weighting parameter for the penalty for constraint violation.

3) Construct an initial function approximator $\tilde{J}$ using calculated profit-to-go values for the sampled points to approximate the profit-to-go with respect to the continuous hyper-state. In this work, a local averager, i.e., a modified k-nearest neighbor (kNN), suggested in (Lee et al., 2006) is used as the approximator. Considering the nonstationary, finite-time characteristics of the batch process, the value function approximation is performed for each time step *k* as below:

$$\tilde{J}_k\left(\xi_{k,0}\right)=\sum_{n=1}^{N}w_nJ_k\left(\xi_{k,n}\right) \tag{10}$$

with

$$w_n=\frac{1/d_n}{\sum_N 1/d_n}, \tag{11}$$

where $\xi_{k,0}$ is a query point at time k, and N is the number of neighboring points in the data set. Each neighboring point is weighted inversely proportional to its Euclidean distance d_n. To avoid excessive extrapolation, a quadratic penalty term based on the local density $J_{k,P}$ is subtracted.

$$\tilde{J}_k\left(\xi_{k,0}\right)\Leftarrow\tilde{J}_k\left(\xi_{k,0}\right)-J_{k,P}\left(\xi_{k,0}\right), \tag{12}$$

$$J_{k,P}\left(\xi_{k,0}\right)= A_k\cdot H\left(\frac{1}{f_\Omega\left(\xi_{k,0}\right)}-\rho\right)\cdot\left[\frac{1/f_\Omega\left(\xi_{k,0}\right)-\rho}{\rho}\right]^2, \tag{13}$$

$$f_\Omega\left(\xi_{k,0}\right)=\frac{1}{N\sigma_B^{m_0}}\sum_{n=1}^{N}K\left(\frac{\xi_{k,0}-\xi_{k,n}}{\sigma_B}\right). \tag{14}$$

where $f_\Omega\left(\xi_{k,0}\right)$ is a density estimate around $\xi_{k,0}$, obtained as a sum of kernel functions $K(\cdot)$ placed at each sample. $H(\cdot)$ is a Heaviside step function, A_k is a scaling parameter, ρ is a threshold value, and σ_B is a bandwidth parameter. Detailed description of the penalty term can be found in (Lee et al., 2006). Note that in this study the penalty term is set as $J_{k,\max}-J_{k,\min}$ whenever $J_{k,P}(\xi_{k,0})\geq J_{k,\max}-J_{k,\min}$, to bound the profit-to-go in the iteration steps.

4) Improve the profit-to-go approximation through value iteration

$$J_k^{i+1}\left(\xi_k\right)=\max_{u_k}E\left[\phi\left(x_k,u_k\right)+\tilde{J}_{k+1}^{i}\left(\xi_{k+1}\right)\left|\xi_k\right.\right], \tag{15}$$

where superscript *i* denotes *i*th iteration step and J_k^{i+1} is calculated for all the sampled states ξ_k from simulations. To evaluate the expectation, Monte Carlo simulation is performed and the average of data ensemble is used. The control input space is discretized and the expectation is evaluated for each candidate input. The iteration is repeated until $\left\|J_k^i\left(\xi\right)-J_k^{i-1}\left(\xi\right)\right\|_\infty$ becomes negligibly small for all k.

Once the profit-to-go values converge, it can be used on-line as a control policy by solving

$$u_k=\pi^*\left(\xi_k\right)=\arg\max_{u_k}E\left[\phi\left(x_k,u_k\right)+\tilde{J}_{k+1}^{N_c}\left(\xi_{k+1}\right)\left|\xi_k\right.\right], \tag{16}$$

at each sampling time. This single-stage optimization requires much less on-line computation time than the original multi-stage optimization problem.

3. CASE STUDY

3.1 Fed-Batch Bioreactor

We illustrate the ADP-based dual control of the batch process with an example of fed-batch ethanol fermentation. The system can be described by

$$x_k=f_k\left(x_{k-1},u_{k-1},\theta_k\right)+e_k,\qquad e_k\sim N\left(0,R_e\right), \tag{17}$$

$$y_k=h_k\left(x_k\right)+v_{k,}\qquad v_k\sim N\left(0,R_v\right), \tag{18}$$

$$\theta_k=\theta_{k-1}+w_k,\qquad w_k\sim N\left(0,R_w\right), \tag{19}$$

where e_k is exogenous noises and θ_k is a set of uncertain parameters. The system model f_k is a discretized form of the following nonlinear differential equations (Chen & Hwang, 1990) using explicit Euler method.

$$\dot{X}=\mu X-\frac{u}{V}X, \tag{20}$$

$$\dot{S}=-\frac{\mu X}{Y_x}+\frac{u}{V}\left(S_{in}-S\right), \tag{21}$$

$$\dot{P}=\eta X-\frac{u}{V}P, \tag{22}$$

$$\dot{V}=u, \tag{23}$$

with

$$\mu = \frac{\mu_{max} S}{\left(1 + P / K_P\right)\left(K_S + S\right)}, \tag{24}$$

$$\eta = \frac{\eta_{max} S}{\left(1 + P / K_P'\right)\left(K_S' + S\right)}, \tag{25}$$

The model includes mass balances for biomass X, substrate S, ethanol P, and liquid volume V. The control input u is the feeding rate of substrate. Table 1 lists the nominal values of the model parameters and initial conditions of the states. There are various sources of uncertainty in bioprocess including variability in feedstock (Kim et al., 2019), product yield (Heinzle et al., 2007), growth rate, and so on, but for the sake of simplicity, we consider two uncertain parameters in this study, the inlet sugar concentration $S_{in} \sim N(150, 20^2)$ and the maximum specific productivity $\eta_{max} \sim N(1, 0.2^2)$, and assume them to be normally distributed. In addition, we assume the uncertain parameters remain constant during a batch. A perfect measurement of the physical states is assumed ($R_v = 0$) and thus $y_k = x_k = [X_k, S_k, P_k, V_k]^T$ and $\theta_k = [\eta_{\max,k}, S_{in,k}]^T$. In reality, some state variables in bioprocess cannot be easily measured on-line, and even they can be measured, there exist measurement noises. In such a case, we might need to perform estimation of states as well as parameters, and the dual controller might try to actively learn about uncertainties in state as well as parameters. In this study, we focus on investigating the active learning feature of the proposed dual controller with respect to uncertainties in parameters.

Table 1. Nominal parameters of the fed-batch bioreactor (Chen & Hwang, 1990)

Parameter	Nominal	Unit
$\mu_{\max}$	0.408	h^{-1}
K_P	16	g/L
K_S	0.22	g/L
$\eta_{\max}$	1	h^{-1}
K_P'	71.5	g/L
K_S'	0.44	g/L
Y_X	0.1	g/g
S_{in}	150	g/L
X_0	1	g/L
S_0	150	g/L
P_0	0	g/L
V_0	10	L

In this study, the extended Kalman filter is employed to estimate θ_k from the measurements. We set the initial covariance matrix Σ_0 as $diag\{0.2^2, 20^2\}$, R_e as $I_{4\times4}$, and R_w as $diag\{0.01^2, 0.01^2\}$. The hyper-state of the process is defined as a 9-dimensional vector consisting of four physical states and five information states:

$$\xi_k = \left[X_k, S_k, P_k, V_k, \hat{\eta}_{max,k+1|k}, \hat{S}_{in,k+1|k}, \Sigma^{11}_{k+1|k}, \Sigma^{22}_{k+1|k}, \Sigma^{12}_{k+1|k} \right]^T \tag{26}$$

where $\hat{\theta}_{k+1|k}$ and $\Sigma^{ij}_{k+1|k}$ represent the parameter estimates and the ij[th] element of covariance matrix based on the measurements up to time k.

3.2 Control Problem

The control goal is to maximize the production of ethanol PV at the fixed final time while satisfying an upper bound constraint on volume, i.e., $V \le V_{max} = 200\,L$. The batch time is $t_f = 50\,h$ and the sampling time of the controller is $t_s = 5\,h$. The control input u_k is bounded in $[0, 12]$. In this example, the stage-wise reward and terminal reward are defined as below:

$$\phi(x_i, u_i) = \begin{cases} \lambda(V_{\max} - V_{i+1}) & if\ V_{i+1} > V_{\max} \\ 0 & otherwise \end{cases}, \tag{27}$$

$$\bar{\phi}_T\left(x_{t_f}\right) = PV\big|_{t_f}, \tag{28}$$

where the penalty factor λ is set as 10^6 in this case study. The penalty factor must be chosen to sufficiently large value for ensuring the given constraint.

For the data generation, stochastic closed-loop simulations were performed with the suboptimal policies, i.e., a shrinking-horizon adaptive economic nonlinear model predictive control (eNMPC) with and without dithered inputs. NMPC is a well-known advanced batch control strategy and an adaptive NMPC is an MPC method in which the uncertain parameters of the prediction model are adapted in a certainty equivalent manner. The dither signals were introduced to generate more broad range of potential operating data, and they were uniformly sampled from $[-1, 1]$. Four sets of dithered inputs from different realizations were used and 200 runs of simulations with each suboptimal policy were conducted. The total number of data points for each time step obtained from the simulations was 1000. With the sampled data, the value function approximators were constructed using 4-nearest neighbor averager, and the value iteration was performed to improve the approximation, by solving (14). The expectation operator was evaluated explicitly by simulating 30 realizations of uncertainties θ_k and e_k, randomly sampled from $N(\theta_{k+1|k}, P_{k+1|k})$ and $N(0, R_e)$, respectively, for each candidate control input. The candidate input set was composed of the input values corresponding to the four nearest neighbors sampled during simulation step as well as discretized values over the input space. The value iteration converged after 7 runs using the following convergence criterion

$$e_{rel} = \left\| \frac{J_k^i(\xi_k) - J_k^{i-1}(\xi_k)}{J_k^{i-1}(\xi_k)} \right\|_\infty < 0.05. \quad (29)$$

The coefficients for penalty term of the averager were set as $\rho = 966.77$, $\sigma_B = 0.9$, and $A = \{A_k\}_{k=0}^{t_f} =$ [160225, 160225, 160225, 160225, 160225, 160225, 158744, 140894, 107543, 60346, 12].

4. RESULTS AND DISCUSSION

The performances of shrinking-horizon adaptive eNMPC and ADP-based robust dual optimizing control are compared for one hundred new uncertainty realizations. The average amount of ethanol produced and the number of constraint violations for each control policy are reported in Table 2. It can be seen that the adaptive eNMPC cannot satisfy the constraints for more than half of the uncertainty realizations, indicating that adapting uncertain parameters alone is not enough to handle the uncertainty, especially in the case of control with economic goals which forces the process to operate at one of its constraints. On the other hand, the ADP-based controller, in which the penalty for constraint violation is incorporated as a stage-wise reward, shows substantial robustness against stochastic system uncertainties.

The average amount of ethanol produced during a batch is compared for the feasible cases where both controllers do not violate the constraint (the fourth row in Table 2). The ADP-based robust dual control shows similar (even slightly better) ethanol production compared to the adaptive eNMPC even though it is significantly more robust with less constraint violation against one hundred uncertainty realizations. In general, there exists a trade-off between the robustness and performance, however, the active learning feature of the dual control allows for better estimation of uncertain parameters, leading to improved performance in a less conservative way.

Table 2. Performance comparison between adaptive NMPC and robust dual adaptive control based on ADP

	Adaptive NMPC	ADP-based control
Constraint violation (CV)	73/100	2/100
Aver. EtOH production [g]	17287.85	16372.96
Aver. EtOH production [g] (w/o CV for both cases)	17394.38	17543.33

Fig. 1 shows the profiles of control input, ethanol concentration and volume when $\eta_{\max}$ and S_{in} are 0.85 h^{-1} and 132 g/L, respectively. For this realization, both control methods satisfy the volume capacity constraint, but the ethanol production obtained with the proposed ADP-based robust dual control is larger than one obtained with the adaptive eNMPC. It can be seen that the control input sequence generated by the robust dual controller includes more excitation (see Fig. 1), resulting in more accurate estimates of parameters (see Fig. 2). The mean squared error of $\hat{\eta}_{\max}$ and $\hat{S}_{in}$ by ADP-based robust dual control are 0.031 and 53.4, and those by adaptive eNMPC are 0.044 and 114.3, respectively. In addition, the trace of covariance matrix at the end of the batch by ADP-based robust dual control is 9% smaller than that by adaptive eNMPC. The ADP-based robust dual control provides a probing feature to the control inputs for active learning, which is well-balanced with the control activity.

In addition, we analyzed the effects of penalty weight λ in (26) and the number of sampled data points on the performance of ADP-based control. The performances are compared for the same one hundred uncertainty realizations. Table 3 shows its performances with different weight values of the penalty. It can be identified that the conservativeness of the controller increases as the penalty weight increases. Therefore, the penalty weight λ needs to be chosen properly depending on the desired conservativeness, or robustness, of the controller.

Table 4 shows the performance of ADP-based control constructed based on different number of sampled data points. As a result, if the number of samples obtained from stochastic simulation step is too small, i.e., 250 data points, performance of the ADP-based control could be unsatisfactory, i.e., showing more constraint violations with less production, due to a poor approximation of the value function. As a local approximator like the k-nearest neighbor, is strongly affected by the number and the quality of training samples, sufficient sampling is required for satisfactory performance of the ADP-based robust dual control.

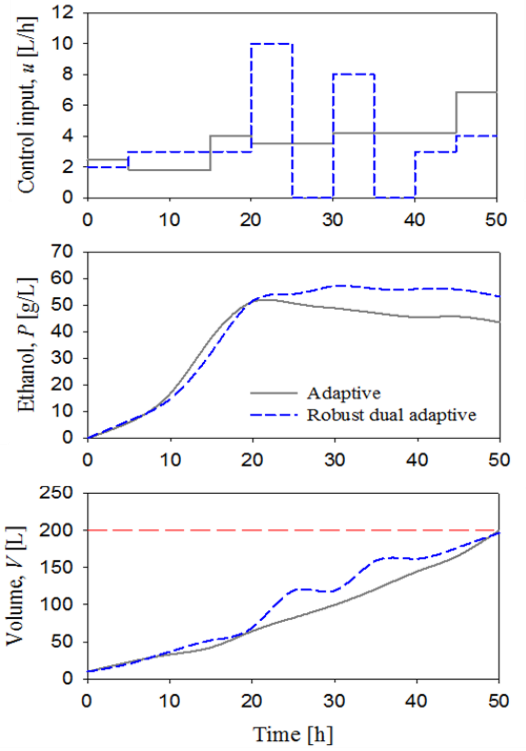

Fig. 1. Profiles of ethanol concentration, volume, and control input when $\eta_{\max} = 0.85 h^{-1}$ and $S_{in} = 132 g/L$.

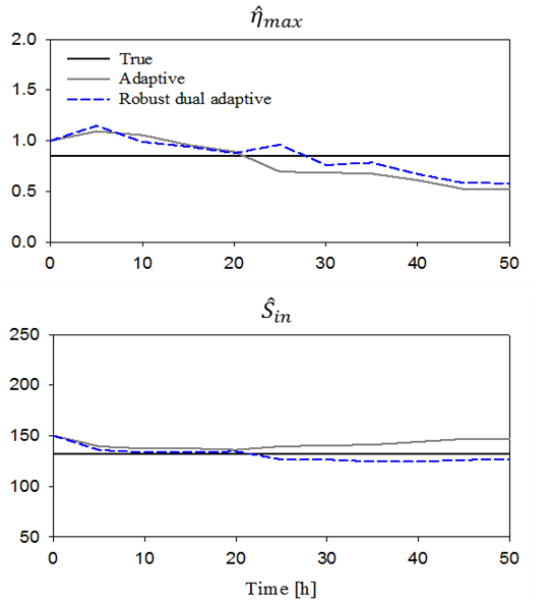

Fig. 2. Parameter estimates when employing the adaptive eNMPC and the ADP-based robust dual adaptive control when $\eta_{\max} = 0.85 h^{-1}$ and $S_{in} = 132\, g/L$.

In this study, the ADP approach based on value function approximation was used, thus the control inputs were searched over the discretized set. Recently, reinforcement learning algorithms that approximate and optimize directly over the policy space, which is called policy gradient algorithm, have been suggested and successfully applied to a variety of control problems with continuous action space (Mnih et al. 2016; Lillicrap et al., 2015). The application of policy gradient methods could facilitate deriving a better control policy for systems with higher dimension of input space. Our future work will investigate an application of the policy gradient methods for the robust dual optimizing control problem of batch processes.

Table 3. Performance of ADP-based robust dual adaptive control with different penalty weight λ (the number of sampled data points = 1000)

Penalty weight λ	Constraint violation	Average EtOH production [g]
10^4	5	16421.93
10^6	2	16372.96
10^8	1	16296.15

Table 4. Performance of ADP-based robust dual adaptive control with different number of sampled data points ($\lambda = 10^6$)

The number of sampled data points	Constraint violation	Average EtOH production [g]
250	2	15964.61
500	1	16079.37
1000	2	16372.96

5. CONCLUSIONS

This paper demonstrates the potential of an approximate dynamic programming (ADP) for robust dual optimizing control of batch processes. The ADP approach has been applied to solve the dual control problem in a tractable way and the rewards are designed to maximize the performance index of a batch while satisfying process constraints. The performance of ADP-based robust dual adaptive control was tested on a fed-batch ethanol fermentation process with two uncertain parameters. The results show that the proposed robust dual optimizing control can robustly and actively deal with stochastic system uncertainty by introducing a proper probing feature to its control input.

REFERENCES

CHEN, C.T. and HWANG, C., 1990. Optimal control computation for differential-algebraic process systems with general constraints. Chemical Engineering Communications, 97(1), pp.9-26.

Feldbaum, A.A., 1960. Dual control theory. I. Avtomatika i Telemekhanika, 21(9), pp.1240-1249.

Filatov, N.M. and Unbehauen, H., 2000. Survey of adaptive dual control methods. IEE Proceedings-Control Theory and Applications, 147(1), pp.118-128.

Heinzle, E., Biwer, A.P. and Cooney, C.L., 2007. Development of sustainable bioprocesses: modeling and assessment. John Wiley & Sons.

Heirung, T.A.N., Santos, T.L. and Mesbah, A., 2019. Model predictive control with active learning for stochastic systems with structural model uncertainty: online model discrimination. Computers & Chemical Engineering.

Kim, B., Huusom, J.K. and Lee, J.H., 2019. Robust Batch-to-Batch Optimization with Scenario Adaptation. Industrial & Engineering Chemistry Research, 58(30), pp.13664-13674.

Lee, J.M., Kaisare, N.S. and Lee, J.H., 2006. Choice of approximator and design of penalty function for an approximate dynamic programming based control approach. Journal of process control, 16(2), pp.135-156.

Lee, J.M. and Lee, J.H., 2009. An approximate dynamic programming based approach to dual adaptive control. Journal of process control, 19(5), pp.859-864.

Lillicrap, T.P., Hunt, J.J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D. and Wierstra, D., 2015. Continuous control with deep reinforcement learning. arXiv preprint arXiv:1509.02971.

Lucia, S., Andersson, J.A., Brandt, H., Bouaswaig, A., Diehl, M. and Engell, S., 2014. Efficient robust economic nonlinear model predictive control of an industrial batch reactor. IFAC Proceedings Volumes, 47(3), pp.11093-11098.

Mnih, V., Badia, A.P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., Silver, D. and Kavukcuoglu, K., 2016, June. Asynchronous methods for deep reinforcement learning. In International conference on machine learning (pp. 1928-1937).

Morinelly, J.E. and Ydstie, B.E., 2016. Dual mpc with reinforcement learning. IFAC-PapersOnLine, 49(7), pp.266-271.

Thompson, A.M. and Cluett, W.R., 2005. Stochastic iterative dynamic programming: a Monte Carlo approach to dual control. Automatica, 41(5), pp.767-778.

Wittenmark, B., 1995. Adaptive dual control methods: An overview. In Adaptive Systems in Control and Signal Processing 1995 (pp. 67-72). Pergamon.