# Dynamic Programming Based Adaptive Optimal Control for Inverted Pendulum

Dao Phuong Nam[(✉)], Nguyen Van Huong, Ha Duc Minh,
and Nguyen Thanh Long

Hanoi University of Science and Technology, Hanoi, Vietnam
nam.daophuong@hust.edu.vn

**Abstract.** This work presents the problem of adaptive optimal control law for a class of continuous-time systems with input disturbance and unknown parameters. The main objective is to find an adaptive optimal control law based on the adaptive dynamic programming (ADP) method and it is able to stabilize the closed-loop system. Besides, the convergence properties of proposed algorithm is pointed out. The theoretical analysis and simulation results demonstrate the performance of the proposed algorithm for inverted pendulum.

**Keywords:** Adaptive Dynamic Programming (ADP) · Adaptive optimal control law · Inverted pendulum

## 1 Introduction

In order to design optimal control law for uncertain systems, the approximate/adaptive dynamic programming (ADP) approach is a biologically – inspired, non-model-based, computational method that has been used in numerous researches [1], such that appropriate reinforcement learning systems design by Werbos or neuro-dynamic programming by Bertsekas. In [1], the recent development of ADP theory consist of three approaches: operations research, real-time control problems for dynamical systems, applying previous ADP to nonlinear uncertain systems.

The non-model-based approach has been extensively implemented to investigate adaptive optimal stabilization and corresponding tracking problems [5, 6]. In [5], the control law was obtained after transforming the robust control problem into an optimal control problem. The corresponding optimal control depend on discrete-time HJB equation and it was solved by using a neural network. The proposed control law in [5] ensures closed-loop locally asymptotic stability of uncertain nonlinear system with an inequality condition. In [6], Fan et al. proposed the sliding mode controller based on adaptive optimal control theory for partially unknown nonlinear systems with input disturbances. The nearly optimal control design ensures stability of the equivalent sliding-mode dynamics by using policy iteration algorithm [6]. The critic network is utilized to approximate the cost function to overcome the difficulty at the second step of policy iteration algorithm. The proposed controller in [6] ensures closed-loop UUB stability of uncertain nonlinear system depend on the property of bounded signal.

In [7], Jiang and Jiang pointed out the control design based on continuous time systems and the equivalent HJB equation. However, the analytical solution of HJB equation is difficult to be obtained and [7] proposed the PI online technique. The stability analysis of closed-loop system pointed out input state stability (ISS) property and the estimation of attraction region depend on KL functions. In [8], Jiang proposed adaptive optimal control law based on algebraic Ricatti equation for uncertain linear systems without external disturbance. The computational adaptive optimal control algorithm was developed from Kleinman (1968) result. In [2, 9], the discretized model is utilized to propose PI, VI – based output ADP design control techniques. However, the results of these proposed control law much depend on the sample time of discrete systems.

Remarkably, this paper extends [2, 9], which focuses on the continuous time systems with external disturbance, to robust control law of uncertain systems. Additionally, we develop a new adaptive optimal control for continuous-time systems with uncertainties under the framework of the idea of ADP problem and external disturbances.

## 2  Adaptive Optimal Control Design

In this paper, we study a class of continuous-time systems described by:

$$\begin{cases} \dot{z} = g(z; y; v) \\ \dot{x} = Ax + B(u + \Delta(z; y; v)) + Dv \\ \dot{v} = Ev \\ e = Cx + Fv \end{cases} \tag{1}$$

where $x \in \mathbb{R}^n$ is the measured component of the state available for feedback control, $u \in \mathbb{R}^m$ is the input, $y = Cx \in \mathbb{R}^r$ represents the output of plant, $y_d = -Fv \in \mathbb{R}^r$ is the reference signal to be tracked, $e \in \mathbb{R}^r$ is tracking error. $z \in \mathbb{R}^p; v \in \mathbb{R}^q$ are the states of the exosystem. The functions $g : \mathbb{R}^p \times \mathbb{R}^r \times \mathbb{R}^q \to \mathbb{R}^p$ and $\Delta : \mathbb{R}^p \times \mathbb{R}^r \times \mathbb{R}^q \to \mathbb{R}^m$ are two locally Lipschitz functions satisfying, $g(0; 0; 0) = 0$ and $\Delta(0; 0; 0) = 0$. Suppose $A \in \mathbb{R}^{n \times n}; B \in \mathbb{R}^{n \times m}; C \in \mathbb{R}^{r \times n}; D \in \mathbb{R}^{n \times q}; E \in \mathbb{R}^{q \times q}; F \in \mathbb{R}^{r \times q}; g \in \mathbb{R}^p; \Delta \in \mathbb{R}^m$ are unknown and $z; v; y; y_d$ are unmeasurable.

The control objective is to find the adaptive optimal control law based on an iterative algorithm ensures that tracking errors converge to zero and convergence properties of this iterative algorithm in presence of uncertain and external disturbance in system. It is known from the adaptive optimal control literature that the following assumptions are given for solving the classical adaptive optimal control problem.

**Assumption 1:** The pair $(A; B)$ is controllable.

**Assumption 2:** The transmission zeros condition holds, i.e.,

$$rank \begin{bmatrix} A - \lambda I & B \\ C & 0 \end{bmatrix} = n + r; \forall \lambda \in \delta(E).$$

**Assumption 3:** The minimal polynomial of $E$ is available, which is:

$$\Gamma_E(s) = \prod_{i=1}^{M} (s - \lambda_i)^{a_j} \prod_{j=1}^{N} \left( s^2 - 2\mu_j s + \mu_j^2 + \omega_j^2 \right)^{b_j}$$

with degree $q_E \leq q$ and $a_i; b_j$ are positive integers and $\lambda_i; \mu_j; \omega_j \in \mathbb{R}$ for $i = \overline{1; M}; j = \overline{1; N}$.

**Assumption 4:** There exists functions $\beta_z; \beta_\Delta$ of class *KL* and a function $\gamma_z; \gamma_\Delta$ of class $K$, both of which are independent of any $v$ satisfies:

$$\|z(t)\| \leq \beta_z(\|z(0)\|, t) + \gamma_z(\|e\|)$$
$$\|\Delta(t)\| \leq \beta_z(\|\Delta(0)\|, t) + \gamma_z(\|e\|)$$

**Assumption 5:** There exist a continuously differentiable, positive definite and radially unbounded function $\Pi : \mathbb{R}^p \rightarrow \mathbb{R}$ and two constant $c_1 > 0; c_2 > 0$, such that:

$$\frac{\partial \Pi}{\partial z} g(z, y, v) \leq - c_1|\Delta(z, y, v)|^2 + c_2|e|^2;$$
$$\forall z \in \mathbb{R}^p; y \in \mathbb{R}^r$$

**Assumption 6:** There exists a constant known number $\xi > 0$ such that matrix $C$ satisfied $\|C\| \leq \xi$.

**Remark 1:** It is different from [2], this paper implement class of systems satisfying $z; v; y; y_d$ are unmeasurable and $x, u, e$ are measurable.

We introduce the classical theorem (Kleinman 1968-Method) [4]: Let $K_0$ be any stabilizing feedback gain matrix, and repeat the following steps for $k = 0; 1; \ldots$

Step 1: Solve for the real symmetric positive definite solution $P_k$ of the Lyapunov equation:

$$A_k^T P_k + P_k A_k + Q + K_k^T K_k = 0 \tag{2}$$

Step 2: Update the feedback gain matrix by:

$$K_{k+1} = B^T P_k \tag{3}$$

Then, the following properties hold:

(1)  $A - BK_k$ is Hurwitz.
(2)  $P^* \leq P_{k+1} \leq P_k$
(3)  $\lim_{k \to \infty} K_k = K^*; \lim_{k \to \infty} P_k = P^*$

We propose the adaptive optimal control law based on the next theorems described as follows:

**Theorem 1:** We denote $\varepsilon = x - Xv$, if a controller is designed as $u = -K^*\varepsilon + Uv$ where $X; U$ solve the following regulator equation:

$$\begin{cases} XE = AX + BU + D \\ 0 = CX + F \end{cases}$$

and $K^* = B^T P^*$ with the symmetric matrix $P^* > 0$ is the unique solution of the well-known algebraic Riccati equation:

$$P^*A + A^T P^* + Q - P^*BB^T P^* = 0 \tag{4}$$

and the weighting matrices $Q$ in (4) satisfying $\lambda_{\min}(Q) > \gamma.\xi^2 > \frac{c_2}{c_1}.\xi^2$ then the closed-loop system achieves disturbance rejection and asymptotic tracking.

**Proof:** Using assumption 2 implies the regulator equation is solvable for any matrices $D; F$.

$$\varepsilon = x - Xv \tag{5}$$

$$\Rightarrow \dot{\varepsilon} = \dot{x} - X\dot{v}$$
$$\Rightarrow \dot{\varepsilon} = Ax + B(u + \Delta(z; y; v)) + Dv - XEv \Rightarrow \dot{\varepsilon} = A(\varepsilon + Xv) + B(u + \Delta(z; y; v)) + Dv - XEv$$
$$\Rightarrow \dot{\varepsilon} = A\varepsilon + B(u - Uv + \Delta(z; y; v))$$

$$e = y - y_d = Cx + Fv = C(\varepsilon + Xv) + Fv \qquad \Rightarrow e = C\varepsilon \tag{6}$$

Define $V_1 = \varepsilon^T P^* \varepsilon$, we have:

$$\dot{V}_1 = \varepsilon^T P^*(A\varepsilon + B(-K^*\varepsilon + \Delta)) + (A\varepsilon + B(-K^*\varepsilon + \Delta))^T P^* \varepsilon$$
$$= \varepsilon^T \left(P^*(A - BK^*) + (A - BK^*)^T P^*\right)\varepsilon + \varepsilon^T P^* B\Delta + \Delta^T B^T P^* \varepsilon$$
$$= \varepsilon^T \left(-Q - P^*BB^T P^*\right)\varepsilon + 2\varepsilon^T P^* B\Delta = -\varepsilon^T Q\varepsilon - \left|B^T P^* \varepsilon - \Delta\right|^2 + |\Delta|^2$$

By using assumption 6, we have:

$$Q \geq \lambda_{\min}(Q).I > \gamma.\xi^2 \geq \gamma.\|C\|^2 \geq \gamma C^T C \tag{7}$$

Then

$$\dot{V}_1 \leq -\gamma \varepsilon^T C^T C\varepsilon + |\Delta|^2 = -\gamma|e|^2 + |\Delta|^2$$

Setting $V = V_1 + \frac{1}{c_1}\Pi(z)$, by using assumption 5, we have:

$$\dot{V} \le -\gamma|e|^2 + |\Delta|^2 + \frac{1}{c_1}\left(-c_1|\Delta|^2 + c_2|e|^2\right) \Rightarrow \dot{V} \le -\left(\gamma - \frac{c_2}{c_1}\right)|e|^2$$

It is clear that a direct application of LaSalle's Invariance Principle yields the GAS property of the closed-loop system.

**Theorem 2:** There exists a small constant $\alpha > 0$, such that for all symmetric matrix $P > 0$ satisfying $|P - P^*| < \alpha$, the overall system is GAS with controller is $u = -B^T P\varepsilon + Uv$.

**Proof:** Since    (4),    for    any    symmetric    matrix    $P > 0$    we    have: $A^T P + PA + \widehat{Q} - PBB^T P = 0$.
  where:

$$\widehat{Q} = Q + (P^* - P)A + A^T(P^* - P) + PBB^T P - P^* BB^T P^*$$

Since (7), we have $Q > \gamma C^T C$ then exists a constant $\mu > 0$ such that: $Q - \gamma C^T C > \mu I$ .Then, by continuity, there exists $\alpha > 0$, such that for all symmetric matrix $P > 0$ satisfying $|P - P^*| < \alpha$ we have $\widehat{Q} > Q - \mu I$, which implies $\widehat{Q} > \gamma C^T C$. Therefore, by Theorem 1, the control $u = -B^T P\varepsilon + Uv$ globally asymptotically stabilizes system.

**Remark 2: The proposed optimal control law guarantees the GAS property of the closed-loop system in presence of uncertain parameters and external disturbances.**
  Using assumption 3, we can always find a vector $\bar{v}(t) \in \mathbb{R}^{q_E}$ and a matrix $\bar{E} \in \mathbb{R}^{q_E \times q_E}$ such that:

$$\begin{aligned} \dot{\bar{v}}(t) &= \bar{E}.\bar{v}(t) \\ v(t) &= G.\bar{v}(t) \end{aligned} \tag{8}$$

with $G \in \mathbb{R}^{q \times q_E}$ is an unknown constant matrix.
  From (5); (8) and Theorem 1, the LOORP is solved, if we design a controller $u = K^*(x - XG\bar{v}(t)) + UG\bar{v}(t)$

**PI based output ADP design:**
  Suppose $\Delta$ is available during the learning phase.

$$\begin{aligned} \dot{\varepsilon} &= A\varepsilon + B(u - UG\bar{v} + \Delta(z; y; v)) = (A - BK_k)\varepsilon + B(u + K_k\varepsilon - UG\bar{v} + \Delta) \\ \dot{\varepsilon} &= A_k\varepsilon + B(w + K_k\varepsilon - UG\bar{v}) \end{aligned}$$

Define $\bar{P}_k = C^T P_k C$ (8), from (6), we have:

$$
\begin{aligned}
e(t+\delta t)^T \overline{P}_k e(t+\delta t) - e(t)^T \overline{P}_k e(t) &= \varepsilon(t+\delta t)^T P_k \varepsilon(t+\delta t) - \varepsilon(t)^T P_k \varepsilon(t) \\
&= \int\limits_{t}^{t+\delta t} \left[ \varepsilon^T \left( A_k^T P_k + P_k A_k \right) \varepsilon + 2(w + K_k \varepsilon - UG\bar{v})^T B^T P_k \varepsilon \right] d\tau
\end{aligned}
\tag{9}
$$

$$
= - \int\limits_{t}^{t+\delta t} \varepsilon^T \left( Q + K_k^T K_k \right) \varepsilon d\tau + 2 \int\limits_{t}^{t+\delta t} (w + K_k \varepsilon)^T K_{k+1} \varepsilon d\tau - 2 \int\limits_{t}^{t+\delta t} \bar{v}^T (UG)^T K_{k+1} \varepsilon d\tau
\tag{10}
$$

$$
\text{Define } XG = \bar{X}; \ UG = \bar{U}; \ Q + K_k^T K_k = Q_k.
\tag{11}
$$

Since (5); (8); (11), we have:

$$
\begin{aligned}
\varepsilon^T \left( Q + K_k^T K_k \right) \varepsilon &= \left( x^T - \bar{v}^T \bar{X}^T \right) Q_k (x - \bar{X}\bar{v}) \\
&= x^T Q_k x - \bar{v}^T \bar{X}^T Q_k x - x^T Q_k \bar{X}\bar{v} + \bar{v}^T \bar{X}^T Q_k \bar{X}\bar{v}
\end{aligned}
$$

$$
\begin{aligned}
(w + K_k \varepsilon)^T K_{k+1} \varepsilon &= w^T K_{k+1} \varepsilon + \varepsilon^T K_k^T K_{k+1} \varepsilon = w^T K_{k+1} (x - \bar{X}\bar{v}) + \left( x^T - \bar{v}^T \bar{X}^T \right) K_k^T K_{k+1} (x - \bar{X}\bar{v}) \\
&= w^T K_{k+1} x - w^T K_{k+1} \bar{X}\bar{v} + x^T K_k^T K_{k+1} x - \bar{v}^T \bar{X}^T K_k^T K_{k+1} x - x^T K_k^T K_{k+1} \bar{X}\bar{v} + \bar{v}^T \bar{X}^T K_k^T K_{k+1} \bar{X}\bar{v}
\end{aligned}
$$

$$
\bar{v}^T (UG)^T K_{k+1} \varepsilon = \bar{v}^T \bar{U}^T K_{k+1} (x - \bar{X}\bar{v}) = \bar{v}^T \bar{U}^T K_{k+1} x - \bar{v}^T \bar{U}^T K_{k+1} \bar{X}\bar{v}
$$

Then

$$
e(t+\delta t)^T \overline{P}_k e(t+\delta t) - e(t)^T \overline{P}_k e(t)
$$

$$
= - \int\limits_{t}^{t+\delta t} \left[ x^T Q_k x - \bar{v}^T \bar{X}^T Q_k x - x^T Q_k \bar{X}\bar{v} + \bar{v}^T \bar{X}^T Q_k \bar{X}\bar{v} \right] d\tau + 2 \int\limits_{t}^{t+\delta t} \begin{bmatrix} w^T K_{k+1} x - w^T K_{k+1} \bar{X}\bar{v} \\ + x^T K_k^T K_{k+1} x - \bar{v}^T \bar{X}^T K_k^T K_{k+1} x \\ -x^T K_k^T K_{k+1} \bar{X}\bar{v} + \bar{v}^T \bar{X}^T K_k^T K_{k+1} \bar{X}\bar{v} \end{bmatrix} d\tau
$$

$$
- 2 \int\limits_{t}^{t+\delta t} \left[ \bar{v}^T \bar{U}^T K_{k+1} x - \bar{v}^T \bar{U}^T K_{k+1} \bar{X}\bar{v} \right] d\tau
$$

$$
= 2 \int\limits_{t}^{t+\delta t} (w^T K_{k+1} x) d\tau - 2 \int\limits_{t}^{t+\delta t} (w^T K_{k+1} \bar{X}\bar{v}) d\tau + \int\limits_{t}^{t+\delta t} x^T \left( 2 K_k^T K_{k+1} - Q_k \right) x d\tau + \int\limits_{t}^{t+\delta t} x^T \left( Q_k - 2 K_k^T K_{k+1} \right) \bar{X}\bar{v} d\tau
$$

$$
+ \int\limits_{t}^{t+\delta t} \bar{v}^T \bar{X}^T \left( Q_k - 2 K_k^T K_{k+1} \right) x d\tau + + \int\limits_{t}^{t+\delta t} \bar{v}^T \left[ \bar{X}^T \left( 2 K_k^T K_{k+1} - Q_k \right) + 2 \bar{U}^T K_{k+1} \right] \bar{X}\bar{v} d\tau + - 2 \int\limits_{t}^{t+\delta t} (\bar{v}^T \bar{U}^T K_{k+1} x) d\tau
$$

$$
= 2 \int\limits_{t}^{t+\delta t} (w^T K_{k+1} x) d\tau - 2 \int\limits_{t}^{t+\delta t} (w^T K_{k+1} \bar{X}\bar{v}) d\tau + 2 \int\limits_{t}^{t+\delta t} x^T K_k^T K_{k+1} x d\tau - \int\limits_{t}^{t+\delta t} x^T Q_k x d\tau
$$

$$
+ 2 \int\limits_{t}^{t+\delta t} x^T \left( Q_k - 2 K_k^T K_{k+1} \right) \bar{X}\bar{v} d\tau + \int\limits_{t}^{t+\delta t} \bar{v}^T \left[ \bar{X}^T \left( 2 K_k^T K_{k+1} - Q_k \right) + 2 \bar{U}^T K_{k+1} \right] \bar{X}\bar{v} d\tau - 2 \int\limits_{t}^{t+\delta t} (\bar{v}^T \bar{U}^T K_{k+1} x) d\tau
$$

Applying Kronecker product representation gives:

$$e^T \bar{P}_k e = \left(e^T \otimes e^T\right)vec\left(\bar{P}_k\right); w^T K_{k+1} x = \left(x^T \otimes w^T\right)vec(K_{k+1}); w^T K_{k+1} \bar{X} \bar{v}$$
$$= \left(\bar{v}^T \otimes w^T\right)vec(K_{k+1}\bar{X});$$

$$x^T K_k^T K_{k+1} x = \left(x^T \otimes (K_k x)^T\right)vec(K_{k+1}); x^T \left(Q_k - 2K_k^T K_{k+1}\right)\bar{X}\bar{v}$$
$$= \left(\bar{v}^T \otimes x^T\right)vec\left(\left(Q_k - 2K_k^T K_{k+1}\right)\bar{X}\right);$$

$$\bar{v}^T\left[\bar{X}^T\left(2K_k^T K_{k+1} - Q_k\right) + 2\bar{U}^T K_{k+1}\right]\bar{X}\bar{v} = \left(\bar{v}^T \otimes \bar{v}^T\right)vec\left[\left[\bar{X}^T\left(2K_k^T K_{k+1} - Q_k\right) + 2\bar{U}^T K_{k+1}\right]\bar{X}\right]$$

$$\bar{v}^T \bar{U}^T K_{k+1} x = \left(x^T \otimes \bar{v}^T\right)vec\left(\bar{U}^T K_{k+1}\right);$$

Define: $\left(Q_k - 2K_k^T K_{k+1}\right)\bar{X} = G_{1;k};$

$$\left[\bar{X}^T\left(2K_k^T K_{k+1} - Q_k\right) + 2\bar{U}^T K_{k+1}\right]\bar{X} = G_{2;k}; \bar{U}^T K_{k+1} = G_{3;k}$$

$$\Phi_k = \begin{bmatrix} \Upsilon\left(t_0^{(k)}\right) & \vartheta\left(t_0^{(k)}\right) & v\left(t_0^{(k)}\right) & \chi\left(t_0^{(k)}\right) & \pi\left(t_0^{(k)}\right) & \sigma\left(t_0^{(k)}\right) \\ \Upsilon\left(t_1^{(k)}\right) & \vartheta\left(t_1^{(k)}\right) & v\left(t_1^{(k)}\right) & \chi\left(t_1^{(k)}\right) & \pi\left(t_1^{(k)}\right) & \sigma\left(t_1^{(k)}\right) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \Upsilon\left(t_{l-1}^{(k)}\right) & \vartheta\left(t_{l-1}^{(k)}\right) & v\left(t_{l-1}^{(k)}\right) & \chi\left(t_{l-1}^{(k)}\right) & \pi\left(t_{l-1}^{(k)}\right) & \sigma\left(t_{l-1}^{(k)}\right) \end{bmatrix}; \Psi_k = \begin{bmatrix} \rho\left(t_0^{(k)}\right) \\ \rho\left(t_1^{(k)}\right) \\ \cdots \\ \rho\left(t_{l-1}^{(k)}\right) \end{bmatrix}$$

$$\Upsilon(t) = e^T(t) \otimes e^T(t) - e^T(t+\delta t) \otimes e^T(t+\delta t); \vartheta(t) = 2\int_t^{t+\delta t} \left(x^T \otimes (w + K_k x)^T\right)d\tau$$

$$v(t) = -2\int_t^{t+\delta t}\left(\bar{v}^T \otimes w^T\right)d\tau; \chi(t) = 2\int_t^{t+\delta t}\left(\bar{v}^T \otimes x^T\right)d\tau; \pi(t) = \int_t^{t+\delta t}\left(\bar{v}^T \otimes \bar{v}^T\right)d\tau$$

$$\delta(t) = -2\int_t^{t+\delta t}\left(x^T \otimes \bar{v}^T\right)d\tau; \rho(t) = \int_t^{t+\delta t} x^T Q_k x d\tau$$

Consequently, we have:

$$\Phi_k \cdot \begin{bmatrix} vec(\bar{P}_k) \\ vec(K_{k+1}) \\ vec(K_{k+1}\bar{X}) \\ vec(G_{1;k}) \\ vec(G_{2;k}) \\ vec(G_{3;k}) \end{bmatrix} = \Psi_k \qquad (12)$$

**Assumption 7:** For each $k = 1; 2; \ldots$ there exists an integer $N$ such that, when $k \geq N$, the following rank condition holds:

$$rank(\Phi_k) = \frac{n(n+1)}{2} + (m+q)n$$

By assumption 7, then
$\left[vec(\bar{P}_k), vec(K_{k+1}), vec(K_{k+1}\bar{X}), vec(G_{1;k}), vec(G_{2;k}), vec(G_{3;k})\right]^T$     can     be uniquely determined by:

$$\left[vec(\bar{P}_k), vec(K_{k+1}), vec(K_{k+1}\bar{X}), vec(G_{1;k}), vec(G_{2;k}), vec(G_{3;k})\right]^T$$
$$= \left(\Phi_k^T \Phi_k\right)^{-1} \Phi_k^T \Phi_k \tag{13}$$

Assumption 2 implies that $B$ is in full column rank, so $K_{k+1} = B^T P_k$ is in full row rank, then:

$$\bar{U}^T = G_{3;k} \cdot K_{k+1}^T \cdot \left(K_{k+1} \cdot K_{k+1}^T\right)^{-1} \tag{14}$$

**Remark 3:** It is different from [2], we obtain the adaptive optimal control law for continuous time systems affected by external disturbances.

Now, we are ready to propose the following adaptive optimal control algorithm for practical online implementation.

### Algorithm 1:

```
Select a stabilizing K₀ and a threshold v>0 . j←0 .
Apply a bounded control policy μₖ on [0;k₀;₀).
repeat
Apply uₖ =-Kⱼεₖ +eₖ on [kⱼ;₀;kⱼ;ₛ] with eₖ an exploration
noise. Solve P̄ⱼ;Kⱼ₊₁ from (12).
until |P̄ⱼ -P̄ⱼ₋₁|<v
j* ←j. Get the approximated optimal control policy:
```

$$u_k = -K_{j^*} \cdot \varepsilon + \left( G_{3;j^*} \cdot K_{j^*+1}^T \cdot \left( K_{j^*+1} \cdot K_{j^*+1}^T \right)^{-1} \right)^T \bar{v}$$

**Theorem 3:** Let $K_0$ be any stailizing feedback gain matrix, and let $(P_k; K_{k+1}; \bar{U})$ be obtained from Algorithm 1. Then, under assumption 7, the following properties hold:

(1) $A - BK_k$ is Hurwitz.
(2) $P^* \leq P_{k+1} \leq P_k$
(3) $\lim\limits_{k\to\infty} K_k = K^*$; $\lim\limits_{k\to\infty} P_k = P^*$

**Proof:** From (9); (10) one sees that the $(P_k; K_{k+1})$ obtained from (2); (3) must satisfy the condition (12). In addition, by assumption 7, it is unique. Therefore, the solution in theorem Kleinman 1968 is the same as the solution in (13) for any $k \geq N$.

## 3   Simulation Results

$$
\begin{bmatrix} \dot{x} \\ \ddot{x} \\ \dot{\phi} \\ \ddot{\phi} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & \frac{-(I+ml^2)b}{I(M+m)+Mml^2} & \frac{m^2gl^2}{I(M+m)+Mml^2} & 0 \\ 0 & 0 & 0 & 1 \\ 0 & \frac{-mlb}{I(M+m)+Mml^2} & \frac{mgl(M+m)}{I(M+m)+Mml^2} & 0 \end{bmatrix} \begin{bmatrix} x \\ \dot{x} \\ \phi \\ \dot{\phi} \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{I+ml^2}{I(M+m)+Mml^2} \\ 0 \\ \frac{ml}{I(M+m)+Mml^2} \end{bmatrix} u
$$
$$
y = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ \dot{x} \\ \phi \\ \dot{\phi} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} u
$$

(15)

In this section, we apply the proposed adaptive optimal control law to an inverted pendulum on a cart described as (15) and Table 1. The simulation results in Fig. 1 show the convergence of matrix P and K of proposed algorithm and the tracking errors converge to zero.
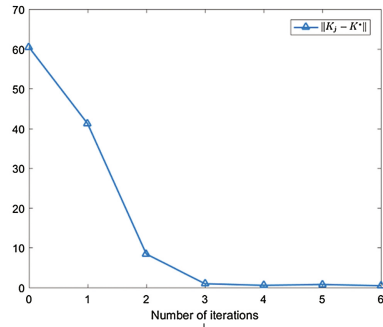
**Table 1.** The parameters of inverted pendullum

| | | | |
|---|---|---|---|
| M | Weight of car | 8.378 | kg |
| m | Weight of link | 0.051 | kg |
| B | Friction coefficient | 12.98 | Ns/m |
| L | ½ length of link | 0.325 | m |
| I | Inertial moment of link | $1.796 \times 10^{-03}$ | kg.m$^2$ |
| G | Gravity | 9.81 | m/s$^2$ |
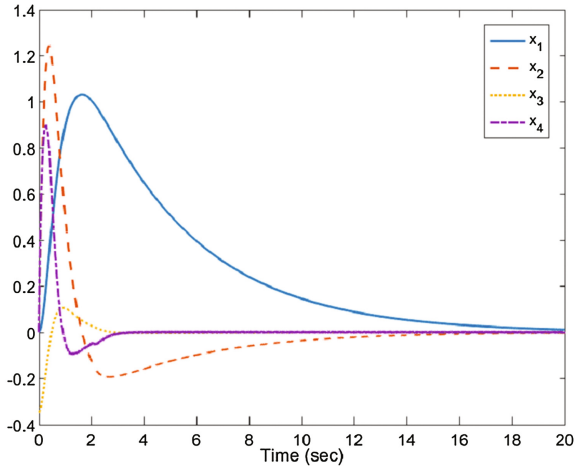
(a)
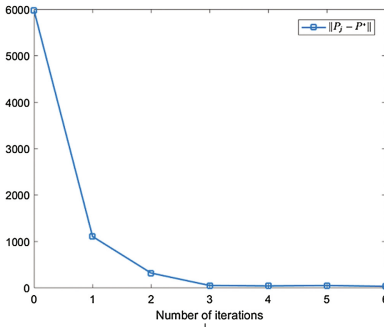


(b)



(c)

**Fig. 1.** Convergence of matrix P, K and tracking errors

## 4 Conclusion

This paper presents an adaptive optimal control algorithm for practical online imple-
mentation of continuous-time systems with unknown system dynamics and external
disturbance. The proposed algorithm pointed out the global asymptotical stability
property and convergence properties. The theory and simulation results illustrate the
effectiveness of proposed algorithm.

# References

1. Jiang, Z.P., Jiang, Y.: Robust adaptive dynamic programming for linear and nonlinear systems: an overview. Eur. J. Control **19**(5), 417–425 (2013)
2. Gao, W., Jiang, Z.P.: Adaptive optimal output regulation via output-feedback: an adaptive dynamic programing approach. In: IEEE 55th Conference on Decision and Control (CDC), Las Vegas, USA, pp. 5845–5850 (2016)
3. Hewer, G.: An iterative technique for the computation of the steady state gains for the discrete optimal regulator. IEEE Trans. Autom. Control **16**(4), 382–384 (1971)
4. Lancaster, P., Rodman, L.: Algebraic Riccati Equations. Oxford University Press Inc., New York (1995)
5. Wang, D., Liu, D., Li, H., Luo, B., Ma, H.: An approximate optimal control approach for robust stabilization of a class of discrete-time nonlinear systems with uncertainties. IEEE Trans. Syst. Man Cybern.: Syst. **46**(5), 713–717 (2016)
6. Fan, Q.Y., Yang, G.H.: Adaptive actor-critic design-based integral sliding-mode control for partially unknown nonlinear systems with input disturbances. IEEE Trans. Neural Netw. Learn. Syst. **27**(1), 165–177 (2016)
7. Jiang, Y., Jiang, Z.P.: Robust adaptive dynamic programming and feedback stabilization of nonlinear systems. IEEE Trans. Neural Netw. Learn. Syst. **25**(5), 882–893 (2014)
8. Jiang, Y., Jiang, Z.P.: Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics. Automatica **48**(10), 2699–2704 (2012)
9. Gao, W., Jiang, Y., Jiang, Z.P., Chai, T.: Output-feedback adaptive optimal control of interconnected systems based on robust adaptive dynamic programming. Automatica **72**, 37–45 (2016)