http://dx.doi.org/10.5370/KIEE.2012.61.3.451

논 문
61-3-16

연속시간 선형시스템에 대한 탐색화된 정책반복법

Explorized Policy Iteration For Continuous-Time Linear Systems

이 재 영* · 전 태 윤* · 최 윤 호** · 박 진 배†
(Jae Young Lee · Tae Yoon Chun · Yoon Ho Choi · Jin Bae Park)

Abstract - This paper addresses the problem that policy iteration (PI) for continuous-time (CT) systems requires explorations of the state space which is known as persistency of excitation in adaptive control community, and as a result, proposes a PI scheme explorized by an additional probing signal to solve the addressed problem. The proposed PI method efficiently finds in online fashion the related CT linear quadratic (LQ) optimal control without knowing the system matrix A, and guarantees the stability and convergence to the LQ optimal control, which is proven in this paper in the presence of the probing signal. A design method for the probing signal is also presented to balance the exploration of the state space and the control performance. Finally, several simulation results are provided to verify the effectiveness of the proposed explorized PI method.

Key Words : Policy iteration, LQR, Adaptive optimal control, Exploration, Persistency of excitation

1. 서 론

정책반복법 (policy iteration)은 최적 의사결정 및 최적 제어문제의 해를 구하기 위한 반복계산 기법이다 [1],[2]. 이러한 정책반복법은 정책평가 (policy evaluation) 루틴과 정책향상 (policy improvement) 루틴으로 구성되어 있으며, 이 두 개의 루틴을 교대로 반복실행하여 최적화 문제를 해결한다. 여기서, 정책평가 루틴은 현재 에이전트 (agent)의 행동에 대한 가치함수 (value function)을 정확히 추정하여 에이전트의 성능을 평가하고, 정책향상 루틴은 추정된 가치함수를 기반으로 에이전트의 성능을 향상시키는 역할을 한다.

이러한 정책반복법에 대한 연구는 유한 마르코프 의사결정 모델 (Markov decision process: MDP) [1],[2]로부터 시작하여 최근에는 동적 시스템 (dynamical system)을 대상으로 하는 연구로 확장되었다 [3]-[9]. 하지만, 연속시간 동적 시스템에 대한 많은 정책반복법들이 그 안정도와 수렴성에 대한 증명이 이루어지지 않은 채로 사용되었으며, 이는 제어 시스템의 안정도 및 성능에 좋지 않은 영향을 미칠 수 있음을 의미한다. 안정성과 수렴성이 보장된 연속시간 동적시스템의 정책반복법은 Murray와 Cox, Lendaris, Saeks (2002)에 의해 처음 제안되었다 [4]. 하지만, 이는 시스템 내부 모델 정보를 모르는 경우, 상태변수의 미분치를 측정해야만 하는 단점이 존재한다. 이를 발전시킨 적분구간 강화학습 방법 (interval reinforcement learning)이라 명명된 Vrabie와 Pastravanu, Abu-Khalaf, Lewis가 제안한 정책반복법 [5],[8],[9]은 시스템의 내부 모델과 상태변수 미분치를 모르는 상황에서도 적용 가능하며, 제어이론 관점 [10],[11]에서 안정성과 수렴성이 증명된 정책반복법이다. 이와 같은 시스템 정보를 완전히 알지 못하는 상황에서도 적용 가능한, 안정도와 수렴성이 보장된 정책반복법은 제어이론적 관점으로 볼 때 적응최적 제어기법으로 분류된다 [5],[8].

한편, 정책반복법 및 이와 연관된 강화학습 (reinforcement learning)등의 학습이론에서는 이용 (exploitation)과 탐색 (exploration)사이의 균형문제가 존재한다 [1],[2]. 에이전트는 보상을 얻기 위해 이미 알고 있는 것을 "이용"해야 하지만 한편, 역시 미래에 더 좋은 행동 선택을 위해서 환경에 대한 "탐색"도 해야 한다. 동적 시스템을 대상으로 하는 정책반복법에서의 이러한 균형문제는 프로빙 (probing) 잡음을 통한 상태공간 탐색과 상태변수의 수렴성 사이의 균형문제로 나타난다 [5],[6]. 즉, 학습을 위해서는 프로빙 잡음을 통해 상태공간을 충분히 탐색해야 하지만, 이는 상태변수의 수렴성을 저해시키는 요인으로 작용하여, 이 둘 사이의 균형이 필요하다. 하지만, 연속시간 시스템의 내부 모델의 정보를 모를 때에도 적용 가능한 [5],[8] 등에서 소개된 정책반복법의 경우에는, 이러한 균형문제가 고려되지 않았고, 상태변수가 수렴하게 되면, 더 이상 학습이 불가능한 상태에 놓이게 된다. 이러한 프로빙 잡음을 통한 상태변수 탐색은 적응제어의 영속여기 조건 (persistent excitation condition)에 해당한다 [12]. 적응제어에서 영속여기 조건은 시스템 파라미터의 수렴을 위해 필수적이며, 영속여기 조건을 만족하지 않을 경우, 더 이상 파라미터 학습이 불가능하게 된다.

본 논문에서는, Vrabie 등에 의해 제안된 연속시간 선형시스템에 대한 정책반복법 [8]을 기반으로 프로빙 (probing)

* 정 회 원 : 연세대학교 전기전자공학과 박사과정
** 정 회 원 : 경기대학교 전자공학부 교수
† 교신저자, 정회원 : 연세대학교 전기전자공학과 교수
E-mail : jbpark@yonsei.ac.kr
접수일자 : 2011년 11월 29일
최종완료 : 2012년 2월 20일

신호를 도입한 탐색화된 정책반복법을 제안한다. 제안한 알고리즘은 기존의 방법 [8]에 대해 일반화된 기법으로, 시스템의 내부 모델정보를 알지 못하는 상황에서도 상태변수 변화율의 측정 없이 LQ-최적제어의 최적해를 구할 수 있는 적응형 최적제어 기법이다. 또한, 도입된 프로빙 신호를 효과적으로 다룰 수 있으며, 프로빙 신호가 존재하는 상황에서 제안한 알고리즘의 안정도와 수렴성이 보장됨을 본 논문에서 증명하였다. 이와 함께, 프로빙 신호를 이용한 상태공간 탐색을 통해 기존방법에 비해 매 단계 수치적 오차가 작은 가치함수를 얻을 수 있음을 본 논문에서 보이고, 효과적인 탐색 및 이용의 균형을 달성하기 위한 프로빙 신호의 설계 방법을 제안한다. 마지막으로, 모의실험을 통해 기존 정책반복법 [8],[9]에 대한 제안된 알고리즘의 유용성과 정확성을 검증하였다.

2. LQ-최적제어 이론 및 수학적 준비

본 논문에서는, 아래와 같은 선형시스템과 이차 비용함수에 대한 LQ-최적제어 문제를 다룬다.

$$\text{시스템: } \dot{x}(t)=Ax(t)+Bu(t),\ \ x(0)=x_0 \tag{1}$$

$$\text{이차 비용함수: } V_u(x(t),t)=\int_t^{\infty} r(x,u)\,d\tau \tag{2}$$

여기서 $x(t)\in R^n$와 $u(t)\in R^m$는 각각 시스템의 상태변수와 제어입력이고, $r(x,u)$는 준-양확정 (positive semi-definite) 행렬 $Q\geq 0$와 양확정 (positive definite) 행렬 $R>0$에 의해 $r(x,u):=x^TQx+u^TRu$로 정의되는 이차형식의 함수이다. 또한, A와 B는 각각 $n\times n$, $n\times m$의 상수행렬이며, A의 계수 값들은 알려지지 않았다고 가정한다. 비용함수 (2)를 최소화시키는 제어입력 $u^*(t)$와 이에 대한 최적 비용함수 $V^*(x)$는 $u^*=-Kx$와 $V^*(x)=x^TP^*x$와 같이 나타낼 수 있으며, 여기서 K^*는 $K^*=R^{-1}B^TP$로 정의되는 행렬, $P^*\geq 0$는 아래와 같은 대수 리카티 방정식 (algebraic Riccati equation: ARE)을 만족시키는 준-양확정 행렬이다.

$$A^TP^*+P^*A-(K^*)^TRK^*+Q=0$$

이러한 최적해 $P^*\geq 0$의 존재성과 유일성을 위해, 아래의 논의에서는 (A,B)가 가안정 (stabilizable)하고 $(A,Q^{1/2})$가 가검출 (detectable)하다고 가정한다. 본 논문에서 제안하는 적응형 최적 제어방법은 탐색신호가 존재하고 행렬 A의 계수 값들이 알려지지 않은 상황에서, 상태변수 변화율의 측정 없이, (2)에 대한 최적 입력 u^*를 학습한다. 이를 서술하기 위해서는 크로넥커곱 (Kronecker product)과 벡터화 연산이 필요하다. $A\otimes B$를 행렬 A와 B의 크로넥커곱으로 정의하자. 또한, $m\times n$ 행렬 X 에 대하여 $\mathrm{vec}(X)$를 행렬 X의 열벡터들을 일렬로 늘어놓은 $mn\times 1$ 열벡터를 돌려주는 연산자로 정의하자. 이같이 정의된 크로넥터곱 $A\otimes B$와 $\mathrm{vec}(X)$연산 사이에는 다음이 성립한다.

1) $(A\otimes B)^T=A^T\otimes B^T$,
2) $x^TAy=(y\otimes x)^T\mathrm{vec}(A)$,
3) $\mathrm{vec}(AXB)=(B^T\otimes A)\mathrm{vec}(X)$

또한, $n\times n$ 대칭행렬 X에 대한 연산자 $\mathrm{vec}^+(X)$를 대칭행렬 X로부터 X의 대각성분과 상삼각 (upper triangular) 성분에 해당하는 열벡터들을 일렬로 나열한 $n(n+1)/2$차 열벡터에서, 상삼각 성분에 해당하는 성분들에 다시 2를 곱해주는 연산자로 정의하자. 이와 같이 정의된 두 연산자 $\mathrm{vec}(X)$와 $\mathrm{vec}^+(X)$ 는 아래와 같은 성질을 만족시킨다 [4].

4) 임의의 n차 대칭행렬 $X=X^T\in R^{n\times n}$에 대해 $\mathrm{vec}(X)=S\mathrm{vec}^+(X)$를 만족시키는 행렬 S가 존재한다.

본 논문에서는 편의를 위해 벡터 $x\in R^n$와 행렬 $X\in R^{n\times n}$에 대해 각각 $\|x\|$를 x에 대한 유클리드 노름 $\|x\|:=\sqrt{x^Tx}$으로, $\lambda_M(X)$를 X의 최대 고유값, $\lambda_m(X)$를 X의 최소 고유값으로 정의한다. 또한, 행렬 X에 대한 행렬노름 $\|X\|$를 $\|X\|:=\sqrt{\lambda_M(X^TX)}$와 같이 정의한다.

3. 프로빙 신호가 고려된 정책반복법

본 장에서는 A에 대한 정보 없이도 LQ-최적제어 입력 u^*를 학습할 수 있으며, 프로빙 신호을 통해 상태변수를 여기시킬 수 있는 정책반복법을 제안한다. 이를 위해 시스템 (1)에 프로빙 신호 $w(t)$가 추가된 아래와 같은 시스템을 고려하자.

$$\text{시스템 II: } \dot{x}(t)=Ax(t)+B[u(t)+w(t)],\ \ x(0)=x_0 \tag{3}$$

여기서 $w(t)$는 $\|w(t)\|\leq w_M$ 인 유계의 신호로서 상태공간 탐색을 통한 향상된 학습을 위해 인가되는 신호이다. 이후 본 논문에서는 프로빙 신호 $w(t)$를 탐색 $w(t)$라 칭한다. 4, 5장에서 볼 수 있듯이, 잘 설계된 탐색 $w(t)$는 상태변수 $x(t)$를 여기시켜 영속여기 조건을 만족하도록 유지시켜주며, 아래 제안한 정책반복법의 계산 안정도를 향상시킨다. 서술을 위해 임의의 제어입력 $u=-Kx$에 대한 폐루프 시스템 행렬 $A-BK$을 $A_K:=A-BK$로 정의하면, (3)에 대한 폐루프 시스템은 아래와 같이 나타낼 수 있다.

$$\dot{x}=A_Kx+Bw,\ \ x(0)=x_0 \tag{4}$$

이제 탐색 $w(t)$가 고려된 정책반복법의 유도를 위해 먼저 LQ-최적제어 문제 (1)-(2)를 고려하자. 만일 제어입력 $u=-Kx$이 안정하다면, 가치함수 (2)는 유한하며 [10], 이는 $V_u(x)=x^TPx$와 같이 나타낼 수 있다. 여기서 M_K를 $M_K:=K^TRK+Q$로 정의하면, $P>0$는 다음과 같은 리아프노프 (Lyapunov) 방정식의 해를 나타낸다.

$$A_K^TP+PA_K=-M_K \tag{5}$$

[8]에서 제안한 정책반복법에서는 주어진 입력 $u=-Kx$에 대해 아래 Bellman 방정식을 통해 유도된 적분식을 기반으로 (5)를 만족시키는 가치함수 $V_u(x)$의 P를 학습한다 [5],[8],[9].

$$x^T(t)Px(t)=\int_t^{t+T} r(x,u)\,d\tau+x^T(t+T)Px(t+T) \tag{6}$$

하지만, 탐색 $w(t)$가 존재하는 경우, (6)를 통해 구해진 P는

더 이상 (5)를 만족시키지 못한다. 아래 보조정리에서는 탐색신호가 존재하는 경우에 (6)를 대신할 수 있는 수식을 서술한다. 서술을 위해 $\Phi(P,t)$를 아래와 같이 정의한다.

$$\Phi(P,t) := \int_t^{t+T} x^T PBw\, d\tau$$

보조정리 1: 상태변수 $x(t)$가 임의의 탐색 $w(t)$와 안정한 제어입력 $u=-Kx$가 인가된 시스템 (4)에 의해 얻어졌다고 가정하면, 리아프노프 방정식 (5)를 만족시키는 $P>0$는 식 (7)에 의해 계산되는 P와 일치한다.

$$x^T(t)Px(t)+2\Phi(P,t)=\int_t^{t+T} r(x,u)\, d\tau + x^T(t+T)Px(t+T) \tag{7}$$

증명: A_K가 안정하다고 가정하면, $M_K \geq 0$에 대해 리아프노프 방정식 (5)을 만족시키는 $P\geq 0$가 항상 존재한다. 시스템 (3)과 제어입력 $u=-Kx$에 대한 리아프노프 함수로 $V(x)=x^TPx$를 고려하면, (4)에 대한 $V(x)$의 시간에 대한 미분은 $\dot{V}(x)=x^T[A_K^TP+PA_K]x+2x^TPBw$와 같고, (5)를 이용하면,

$$\begin{aligned}\int_t^{t+T} r(x,u)\, d\tau &= \int_t^{t+T} x^T M_K x\, d\tau \\ &= -\int_t^{t+T}\left[\dot{V}(x)-2x^TPBw\right] d\tau \\ &= V(x(t))-V(x(t+T))+2\int_t^{t+T} x^TPBw d\tau\end{aligned} \tag{8}$$

와 같은 결과를 얻을 수 있다. 계산된 식 (8)는 (7)과 등가이므로, 임의의 탐색 $w(t)$과 안정한 A_K에 대해 (5)와 (7)의 등가성이 증명되었다. □

알고리즘 1. 탐색화된 정책반복법

1: $P_0=0$으로 설정
2: $i \leftarrow 1$, $u_1=-K_1x$를 임의의 안정한 제어입력으로 설정
3: **do** {
4: 탐색 $w(t)$를 $w(t)\equiv 0$이 아닌 임의의 신호로 설정
5: u_i와 탐색 $w(t)$를 시스템 (3)에 인가
6: **정책평가 루틴:** 아래 식을 만족시키는 P_i계산

$$x^T(t)P_ix(t)+2\Phi(P_i,t)=\int_t^{t+T} r(x,u_i)d\tau + x^T(t+T)P_ix(t+T) \tag{9}$$

7: **정책향상 루틴:** 계산된 P_i를 기반으로 u_{i+1}계산

$$u_{i+1}=-K_{i+1}x,\quad K_{i+1}=R^{-1}B^TP_i \tag{10}$$

8: $i \leftarrow i+1$
9: } **until** $\|K_i-K_{i-1}\|<\delta$

이제 (7)을 기반으로 최적해 u^*와 $V^*(x)$를 구하기 위한 탐색 $w(t)$가 고려된 정책반복법을 알고리즘 1과 같이 유도할 수 있다. 알고리즘 1의 P_i에 대해 $V_i(x)$를 $V_i(x):=x^TP_ix$와 같이 정의하면, 보조정리 1과 그에 관한 논의에서 알 수 있듯이, $V_i(x)$는 i번째 입력 $u_i(t)$에 대한 가치함수가 된다. 이와 관련해 알고리즘 1의 정책평가루틴에서는 측정된 상태변수 $x(t)$, $x(t+T)$와 제어입력 u_i, 탐색 $w(t)$ 등를 이용하여 입력 u_i에 대한 가치함수 $V_i(x)$를 계산한다. 이어서 정책향상루틴에서는 이와 같이 얻어진 u_i에 대한 가치함수 $V_i(x)$를 바탕으로 새로운 제어입력 u_{i+1}을 도출한다. 이러한 정책평가 루틴 (9)과 정책향상 루틴 (10)은 행렬 A의 정보를 이용하지 않음을 식을 통해 알 수 있고, 이들의 반복을 통해 최적의 제어입력 u^*를 도출할 수 있다. 여기서 탐색 $w(t)$는 상태변수를 여기시키기 위해 사용되며, 상태변수의 수렴집합에 관여한다. 이에 대한 수학적 분석을 위해 A_i와 M_i를 각각 $M_i:=K_i^TRK_i+Q$와 $A_i:=A-BK_i$로 정의하면, (3)의 입력 u_i에 대한 폐루프 시스템을 다음과 같이 나타낼 수 있다.

$$\dot{x}=A_ix+Bw \tag{11}$$

여기서 (11)과 (9)는 $K=K_i$일 때의 (4), (5)와 일치하는 수식이며, 따라서 A_i가 안정하면, $P_i>0$는 보조정리 1에 의해

$$(A_i)^TP_i+P_iA_i=-M_i \tag{12}$$

를 만족시킴을 알 수 있다. 이를 이용하면, 시스템 (11)의 안정도에 대한 다음 정리를 얻는다.

정리 1: Q가 양한정이라 가정하자. 만일 초기 제어기 u_1이 안정하고, 모든 $i\in\{1,2,3,...\}$에 대해 P_i와 K_{i+1}이 알고리즘 1의 (9)-(10)에 의해 계산되었다면, 모든 i에 대해 A_i는 항상 안정하며, (11)은 균등궁극유계 (uniform ultimate boundedness)이다. 여기서, 컴팩트 집합 Ω_i과 그 반경 r_i를

$$\Omega_i := \left\{x\in R^n : \|x\| \leq r_i\right\},\ r_i = 2w_M\|RK_i\|/\lambda_m(M_i)$$

와 같이 정의하면, i-번째 시스템 (11)은 상태변수 x를 Ω_i로 유한시간 안에 수렴시킨다.

증명: 본 정리의 증명은 수학적 귀납법에 의해 이루어진다. 먼저, A_{i-1}가 안정하다고 가정하고, (11)에 대한 리아프노프 함수로 $V_{i-1}(x)=x^TP_{i-1}x$를 고려하자. (11)을 따라 $V_{i-1}(x)$를 시간에 대해 미분하면, 다음을 얻는다.

$$\begin{aligned}\dot{V}_{i-1}(x) &= x^T(A_{i-1}^TP_{i-1}+P_{i-1}A_{i-1})x+2u_i^TRw \\ &\quad + x^T\left[P_{i-1}B(K_i-K_{i-1})+(K_i-K_{i-1})^TB^TP_{i-1}\right]x\end{aligned}$$

위 식에 (12)를 대입하고, 완전제곱 형태를 취하면

$$\begin{aligned}\dot{V}_{i-1}(x) &= -x^TM_{i-1}x+2u_i^TRw \\ &\quad + x^T\left[P_{i-1}B(K_i-K_{i-1})+(K_i-K_{i-1})^TB^TP_{i-1}\right]x \\ &\leq -x^TM_ix+2u_i^TRw\end{aligned}$$

를 얻는다. Q가 양한정이라 가정했으므로, M_i 또한 양한정이고, 따라서, 위 식으로부터 다음을 얻는다.

$$\dot{V}_{i-1}(x) \leq -\left[\lambda_m(M_i)\|x\|-2w_M\|RK_i\|\right]\|x\|$$

이 식으로부터, x가 $\|x\|>2w_M\|RK_i\|/\lambda_m(M_i)=r_i$인 경우에 $\dot{V}(x)\leq 0$인 것을 알 수 있다. 따라서, 리아프노프 정리 [12]에 의해, A_{i-1}가 안정한 경우, 폐루프 시스템 $\dot{x}=A_ix+Bw$은 유한시간 안에 Ω_i로 수렴하는 균등궁극유계인 것을 알

수 있다. 또한, $w \equiv 0$ ($w_M = 0$)이면, 항상 $\dot{V}(x) \le 0$이 성립하는 것을 알 수 있으며, 따라서 A_{i-1}이 안정하면, A_i 또한 안정하다. 위와 같은 논의와 수학적 귀납법을 통해, A_1이 안정하면, 모든 $i \in \{1,2,3,...\}$에 대해 A_i 또한 안정하고, 따라서 따라서 i-번째 시스템 (11)은 수렴집합 Ω_i에 대한 균등궁극 유계이다. □

정리 1을 바탕으로, 알고리즘 1의 P^*, K^*로의 수렴성과 u_i의 정책 향상성을 아래와 같이 증명할 수 있다.

유도정리 1: Q가 양한정 행렬이라 가정하자. 만일 초기 제어기 u_1가 안정하면, 알고리즘 1에 의해 계산된 i-번째 가치함수 $V_i(x) = x^T P_i x$는 $V^*(x) \le V_i(x) \le V_{i-1}(x)$를 항상 만족시키며, $i \to \infty$함에 따라 (V_i, u_i)는 (V^*, u^*)로 수렴한다.

증명: 정리 1에 의해 A_i는 항상 안정하며, 따라서 모든 P_i는 (12)를 만족시킨다. (12)는 클래인만 (Kleinman) 뉴턴 방법과 등가이며 [8],[9], 따라서 클레인만 뉴턴방법에 의해 $V^*(x) \le V_i(x) \le V_{i-1}(x)$와 $V^*(x)$로의 수렴성이 증명된다. 또한, V_i의 수렴성과 (10)을 통해 $u_i \to u^*$이 증명된다. □

참조 1: $w(t) \equiv 0$인 경우, 탐색화된 정책반복법은 기존의 정책반복법 [8]과 같게 된다. 즉, 알고리즘 1은 탐색 $w(t)$에 대한 기존방법의 일반화된 알고리즘이다. 여기서 $w(t) \to 0$인 경우, $w_M \to 0$이 되어, 수렴집합 Ω_i는 집합 $\{0\}$으로 수렴하게 된다. 즉, $w(t) \equiv 0$인 경우, 상태변수 x는 평형점 "0"으로 수렴하게 되어, 4장에서 볼 수 있듯이, 더 이상 영속여기조건을 만족시키지 못하게 되며, 이는 알고리즘의 수치적 안정성과 학습능력에 악영향을 미치게 된다.

4. 탐색화된 정책반복법의 구현

본 장에서는 탐색화된 정책반복법 (알고리즘 1)의 최소자승법 기반 구현방법에 대하여 논하고, 이어 상태공간 탐색과 상태변수 수렴성 사이의 균형을 고려한 프로빙 신호 $w(t)$의 설계방법 및 성능에 대해 서술한다.

4.1 최소자승법 기반 알고리즘 구현방법

알고리즘 1의 정책평가 루틴 (9)를 만족시키는 유일한 P_i를 구하기 위해서는 $N_{\min} := n(n+1)/2$개의 방정식이 있어야 하지만, (9)는 1차원의 스칼라 방정식이다. 본 절에서는 이를 극복하기 위한 최소자승법 기반 구현 방법을 소개한다. 먼저 최소자승법을 유도하기 위해 (9)의 각 요소들을 아래와 같은 형식으로 변형한다.

$$x^T P_i x = [(x \otimes x)^T S]\,\mathrm{vec}^+(P_i)$$

$$\Phi(P_i, t) = \left[\int_t^{t+T} (Bw \otimes x)^T S\, d\tau\right] \mathrm{vec}^+(P_i)$$

위 식들은 2장에서 소개한 크로넥커 곱의 4가지 성질을 이용하면 쉽게 유도할 수 있다. 이제, $\bar{x}(t)$를 $\bar{x} := S^T(x \otimes x)$ 와 같이 정의하고, 변형된 위 식들을 이용하면, 식 (9)은 다음과 같이 쓸 수 있다.

$$X^T \mathrm{vec}^+(P_i) = Y \tag{13}$$

여기서 X와 Y는 아래와 같이 정의된다.

$$X := \bar{x}_t - \bar{x}_{t+T} + 2\int_t^{t+T} S^T(Bw \otimes x)\, d\tau$$

$$Y := \int_t^{t+T} x^T Q x + u_i^T R u_i\, d\tau$$

미분방정식 $\dot{V}(t) = x^T Q x + u^T R u$과 $\dot{W} = S^T(Bw \otimes x)$를 고려하면, X, Y는 적분이 없는

$$X = \bar{x}_t - \bar{x}_{t+T} + 2[W(t+T) - W(T)],$$
$$Y = V(t+T) - V(T)$$

와 같은 간단한 형식으로 쓸 수 있다. 이를 통해 T 를 주기로 $N \ge N_{\min}$개의 x를 측정하여 N개의 X, Y 를 계산하였다고 가정하고, 각각을 $X^{(j)}$와 $Y^{(j)}$로 정의하자 ($j = 1,2,3,...,N$). 또한, 이들을 모아놓은 Σ_X와 Σ_Y를 각각 아래와 같이 정의하자.

$$\Sigma_X = [\, X^{(1)},\ X^{(2)},\ \cdots,\ X^{(N)}\,]^T$$
$$\Sigma_Y = [\, Y^{(1)},\ Y^{(2)},\ \cdots,\ Y^{(N)}\,]^T$$

만일 $\mathrm{rank}(\Sigma_X) = N_{\min}$를 만족시킨다면, (9)의 해 P_i는 최소자승법 [8],[14]에 의해 (14)와 같이 유일하게 결정된다.

$$\mathrm{vec}^+(P_i) = (\Sigma_X \Sigma_X^T)^{-1} \Sigma_X \Sigma_Y \tag{14}$$

여기서 (14)의 $(\Sigma_X \Sigma_X^T)^{-1}$에 주목해 보자. 탐색 $w(t)$가 $w(t) \equiv 0$이면, $W(t)$또한 $W(t) \equiv 0$이 되고, 본 최소자승법 기반 구현방법은 [8]에서 제시한 기존의 정책반복법에 대한 구현방법과 동일하게 된다. 이 경우, x가 수렴하거나 그 변화율 $\bar{x}_t - \bar{x}_{t+T}$이 일정하게 되면, Σ_X의 벡터 $X^{(j)}$들은 모두 동일한 값을 가지게 되어 $\mathrm{rank}(\Sigma_X) = N_{\min}$을 더 이상 만족시키지 못해 $(\Sigma_X \Sigma_X^T)^{-1}$를 계산할 수 없게 된다.

반면, '0'이 아닌 탐색 $w(t)$를 통해 Σ_X를 구성하고 있는 상태변수 x와 $W(t)$ 벡터를 충분히 여기시키게 되면, Σ_X의 열벡터들의 상호 의존성이 완화되고, 이를 통해 (14)의 $(\Sigma_X \Sigma_X^T)^{-1}$의 계산이 가능하게 되고, P_i에 대한 수치적 오차가 줄어들 수 있다.

이를 알아보기 위해 (14)의 행렬 $\Sigma_X \Sigma_X^T$와 열벡터 $\Sigma_X \Sigma_Y$가 각각 $\triangle\Sigma_X \Sigma_X^T$와 $\triangle\Sigma_X \Sigma_Y$만큼의 오차가 발생하였을 때 (14)에 의해 발생된 각각의 전달오차를 $\triangle \mathrm{vec}^+(P_i)$라 하면, 각각에 대한 상대오차는 아래와 같이 나타낼 수 있다 [14].

$$\frac{\|\triangle \mathrm{vec}^+(P_i)\|}{\|\mathrm{vec}^+(P_i)\|} \le c\, \frac{\|\triangle\Sigma_X \Sigma_Y\|}{\|\Sigma_X \Sigma_Y\|} \tag{15}$$

$$\frac{\|\triangle \mathrm{vec}^+(P_i)\|}{\|\mathrm{vec}^+(P_i) + \triangle \mathrm{vec}^+(P_i)\|} \le c\, \frac{\|\triangle\Sigma_X \Sigma_X^T\|}{\|\Sigma_X \Sigma_X^T\|} \tag{16}$$

여기서 c는 $\Sigma_X \Sigma_X^T$ 에 대한 조건수 (condition number)로 $c :=$

$\lambda_M(\Sigma_X\Sigma_X^T)/\lambda_m(\Sigma_X\Sigma_X^T)$와 같이 정의되는 양이다. 조건수는 행렬의 고유특성을 나타내는 양으로, 그 행렬의 특이성 (singularity)이 커질수록 이에 비례하여 증가하는 양이다. 단적인 예로, 행렬 $\Sigma_X\Sigma_X^T$가 비정칙 행렬 (singular matrix) 인 경우, c는 $c=\infty$의 값을 갖는다 [14]. 위 식 (15)-(16)를 살펴보면, 각각의 상대오차가 공통적으로 조건수 c에 의존함을 알 수 있다. 따라서 탐색 $w(t)$를 통해 Σ_X의 열벡터들의 상호 의존성을 완화시켜 $\Sigma_X\Sigma_X^T$의 조건수를 감소시키게 되면, 그에 따라 위와 같은 계산오차 (15)-(16)도 또한 줄어들게 된다. 이에 대한 수치적 결과를 기존 정책반복법 [8]과 비교하여 본 논문의 5장에서 제공한다.

4.2 탐색신호의 설계

LQ-최적제어의 목적은 비용함수 (2)를 최소화 시키면서 상태변수 x를 평형점(equilibrium point) '0'으로 유지시키는 것이다. 프로빙 신호 $w(t)$를 이용하면, 위에서 논의한 바와 같이 (6)에 대한 수치/계산적 특성이 향상될 수 있다. 하지만, $w(t)$의 도입으로 인해 정리 1에서 알 수 있듯이 $w(t)\equiv 0$가 아닌 경우는 상태변수 x가 평형점으로 수렴하지 못하고, 평형점 '0'를 포함하는 유계집합 Ω_i 안에 머물게 되어, 상태변수 탐색과 수렴성 사이의 균형이 필요하다. 여기서 유계집합 Ω_i의 범위가 탬색 $w(t)$의 상한 w_M에 의존한다는 것에 주목하면, $w(t)$의 크기를 조절함으로써 상태변수 x의 수렴집합을 평형점 '0'에 충분히 가깝게 할 수 있다.

이를 살펴보기 위해 i-번째 단계에서의 $\|x\|$의 정상상태에서의 최대 허용오차를 x_M로 정의하자. 즉, 정상상태에서 상태변수 x는 $\|x\|\le x_M$를 만족시켜야 한다. 이는 i-번째 유계집합 Ω_i의 반경 r_i가 x_M보다 작을 때 즉, $2w_M\|RK_i\|\le x_M\lambda_m(M_i)$ 일 때 항상 성립한다. 따라서 i-번째 시스템에 대한 정상상태에서 $\|x\|\le x_M$을 만족시키는 w_M값은 다음과 같이 구할 수 있다.

$$w_M=\frac{x_M\lambda_m(M_i)}{2\|RK_i\|} \tag{17}$$

따라서 각 i번째 단계에 대해 (17)을 만족시키도록 탐색 $w(t)$를 구현하면, 정상상태에서의 상태변수 $x(t)$가 항상 $\|x\|\le x_M$을 만족하도록 설계할 수 있다. 여기서, w_M이 (17)의 오른쪽 항보다 더 작도록 $w(t)$를 설계하여도 $\|x\|\le x_M$은 만족된다. 하지만, 다음 절에서 볼 수 있듯이, w_M 이 '0'에 가깝게 되면, $w(t)$ 또한 '0'에 가까운 신호가 되어, 상태변수를 충분히 여기시키지 못하게 되고, 이에 따라 조건수 c가 증가하는 결과를 초래한다. 이는 오차 방정식 (15)-(16)에 의해 (14)의 해 $\mathrm{vec}^+(P_i)$에 영향을 미쳐 수치적 오차를 증가시키는 요인으로 작용하게 되며, 이에 대한 수치적 결과가 다음 장에 제시되어 있다.

5. 모의실험

탐색화된 정책반복법 (알고리즘 1)의 탐색 $w(t)$에 따른 성능을 검증하기 위해 다음과 같은 F-16 비행체의 단주기 동역학 (short period dynamics) 에 대한 선형모델을 고려하였다 [15].

$$\frac{d}{dt}\begin{bmatrix}\alpha\\ q\end{bmatrix}=\begin{bmatrix}-1.01887 & 0.90506\\ 0.82225 & -1.07741\end{bmatrix}\begin{bmatrix}\alpha\\ q\end{bmatrix}-\begin{bmatrix}0.00215\\ 0.1755\end{bmatrix}\delta_e \tag{18}$$

여기서 α는 공격각도 (angle of attack), q는 피치율 (pitch rate), δ_e는 승강편향도 (elevator deflection)을 나타내며, 선형화는 502 ft/s의 비행체 속도와 300 psf의 동적압력 등의 평형조건에서 실시되었다 ([15]의 예제 6.4.2 참조). 본 모의실험에서는 δ_e의 액츄에이터에 대한 추가적인 동역학식으로서 [9]에서와 같이 다음과 같은 저역통과필터를 도입하였다.

$$\delta_e(s)=\frac{20.2}{s+20.2}u(s) \tag{19}$$

여기서 s는 라플라스 변수 (Laplace variable)이며, $u(s)$는 s-영역에서 표현된 승강편향도 δ_e에 대한 제어입력이다. (18)과 (19)를 결합하면, F-16 비행체의 단주기 동역학에 대한 다음과 같은 3차 선형모델을 얻는다.

$$A=\begin{bmatrix}-1.01887 & 0.90506 & -0.00215\\ 0.82225 & -1.07741 & -0.1755\\ 0 & 0 & -20.2\end{bmatrix},\quad B=\begin{bmatrix}0\\ 0\\ 20.2\end{bmatrix} \tag{20}$$

여기서 상태변수 x는 $x:=\begin{bmatrix}\alpha & q & \delta_e\end{bmatrix}^T\in R^3$이며, 제어입력 u는 (19)에 의해 정의된 변수이다. 편의를 위해 상태변수 x의 각 성분을 $x_1=\alpha$, $x_2=q$, $x_3=\delta_e$로 표시하기로 한다. 본 절에서는 이와 같은 F-16 비행체의 선형모델 (20)과 $Q=I_3$, $R=1$을 갖는 비용함수 (2)에 대한 LQ-최적제어 문제를 고려하였으며, 이에 대한 최적해 P^*는 [9]에서의 모의실험 환경과 동일하게 다음과 같이 주어진다.

$$P^*=\begin{bmatrix}1.4117 & 1.1540 & -0.0072\\ 1.1540 & 1.4191 & -0.0087\\ -0.0072 & -0.0087 & 0.0206\end{bmatrix} \tag{21}$$

본 모의실험에서는 위 최적해 (21)을 학습하기 위한 4.1절에서 소개된 알고리즘 1의 최소자승법 기반 구현방법을 이용하였으며, 이는 시스템 행렬 A의 정보 없이도, 최적해 (21)에 대한 학습을 가능하게 한다. 모든 모의실험에서의 초기치는 $x_0=[0.2\ \ 0.1\ \ 0.1]^T$ 로 설정하였으며, 알고리즘 1의 데이터 획득 주기 $T>0$는 $T=0.5$ [s]로 설정하여 매 시행마다 $N=8$개의 데이터 쌍 $(X^{(j)},Y^{(j)})$, $\left(j=1,2,3,\cdots,N\right)$을 획득하여 (14)에 의해 $\mathrm{vec}^+(P_i)$를 계산하였다. 여기서 $\mathrm{vec}^+(P_i)$의 계산을 위해서는 $X^{(j)}$에 의해 구성된 행렬 Σ_X가 $\mathrm{rank}(\Sigma_X)=N_{\min}=6$을 만족시켜야 한다.

탐색 $w(t)$의 크기 w_M는 $x_M=0$경우에는 "$w_M=0$"으로 설정하였으며, $x_M\neq 0$인 경우에는 $i=0$에서 "$w_M=0.01$"로, $i=1,2,3,\ldots$에서는 정해진 x_M에 대해 (17)을 만족시키도록 결정하였다. 이러한 w_M에 대해 $w(t)$는 매구간 $[t,\ t+T]$마다 균등확률분포 $[-w_M,\ w_M]$에 의해 얻어진 샘플값 w_s에 대하여 $w(t)\equiv w_s$로 결정하였다. 이렇게 결정된 $w(t)$는 $|w(t)|\le w_M$를

만족하여, 정상상태에서 항상 $\{x \in R^n : \| x \| \le x_M\}$로 x를 구속시킴과 동시에, x에 대한 영속여기조건의 차수를 증가시키는 장점이 있다 [13].

모의실험은 여러 x_M값에 대해 수행하였고, 표 1은 각 x_M에 대한 조건수 c의 변화를 보여준다. 표 1에서의 w_M 값은 P_i가 수렴한 후에 (17)에 의해 계산된 w_M값을 도시하였다. w_M이 (17)을 만족시키도록 정해진다면, K_i $(=R^{-1}B^TP_i)$가 수렴함에 따라 w_M도 일정한 값으로 수렴함을 쉽게 알 수 있다. 표 1을 보면, x_M을 감소시켜 상태변수의 수렴성을 높이면, 이에 따라 w_M이 작아지게 되어 탐색 $w(t)$를 통한 상태변수 탐색 범위가 줄어들게 된다. 이러한 탐색범위의 감소는곧 조건수 c의 증가를 초래하여 수치적 오차를 증가시키는 요인이 되며, 탐색이 고려되지 않은 극단적인 경우 $(x_M=0)$ [8], $\mathrm{rank}(\Sigma_X)<6$이 되어 표 1의 $i \ge 2$의 경우와 같이 더 이상 P_i를 학습하지 못하게 되는 상황이 발생한다.

또한, 탐색범위의 감소에 따른 조건수 c의 증가는 그림 1에서 볼 수 있듯이 큰 계산오차를 초래할 수 있다. 그림

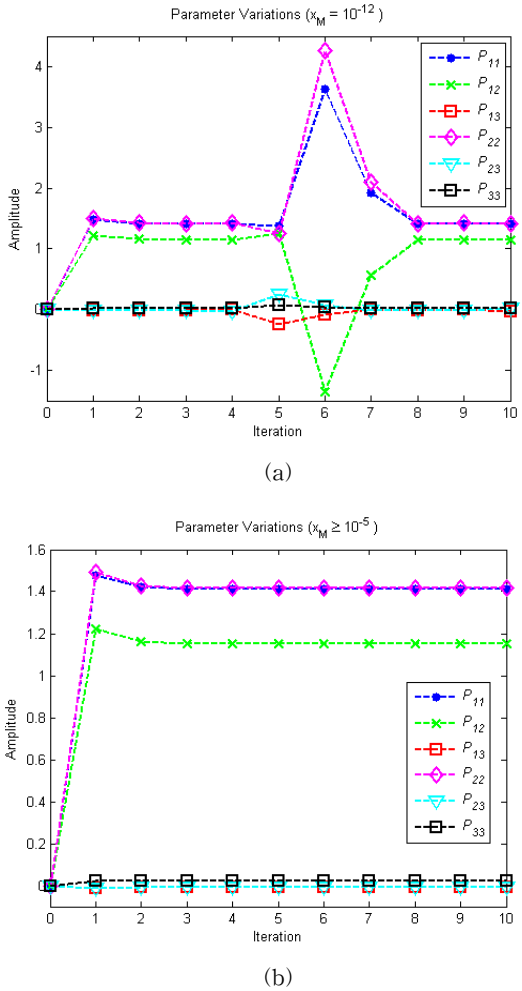


그림 1 (a) $x_M=10^{-2}$, (b) $x_M \ge 10^{-5}$일 때의 탐색화된 정책반복법의 반복시행에 따른 P_i의 변화

Fig. 1 Evolutions of elements of P_i recursively evaluated by the explorized PI−(a) $x_M=10^{-12}$, (b) $x_M \ge 10^{-5}$

1(a)는 $x_M=10^{-12}$인 경우 알고리즘 1에 의한 P_i의 변화곡선을 나타낸다. 알고리즘 1은 식 (12)에 의해 어떤 탐색 $w(t)$에 대해서도 항상 클레인만의 뉴턴방법과 등가이므로, 그림 1(b)와 같이 반복횟수로 비교했을 때에는 어떤 $x_M>0$ 값에 대해서도 항상 기존의 정책반복법 [8]과 동일한 변화곡선을 유지하여야 한다. 하지만, x_M이 매우 작아지게 되면, 조건수 c가 매우 커지게 되고, 이는 그림 1(a)와 같이 P_i의 계산에 대해 매우 큰 오차를 수반하게 된다.

참조 2: 표 1의 "$x_M=0$"인 경우는 $w(t)\equiv 0$이므로, [8],[9]에서 제안한 기존의 정책반복법과 동일함을 알 수 있다. [9]에서는, 동일한 LQ-최적제어 문제에 대해 $T=0.01[s]$와 $N=6$인 경우 최적해로의 수렴성을 모의실험을 통해 입증하였지만, 본 실험에서와 같이 갱신주기 TN이 증가하여 일정시간 이후 영속여기조건을 더 이상 만족시키지 못하는 경우 $(i \ge 2)$, 표 1과 같이 기존 방법으로는 더 이상 학습이 불가능하거나, 혹은 탐색의 부재로 인한 큰 계산오차를 초래할 수 있다 [16]. 본 논문에서 제안한 탐색화된 정책반복법 (알고리즘 1)은 위에서 보인 바와 같이 "충분히 큰 w_M을 갖는 잘 설계된 탐색 $w(t)$"을 통해 이러한 문제를 해결한다.

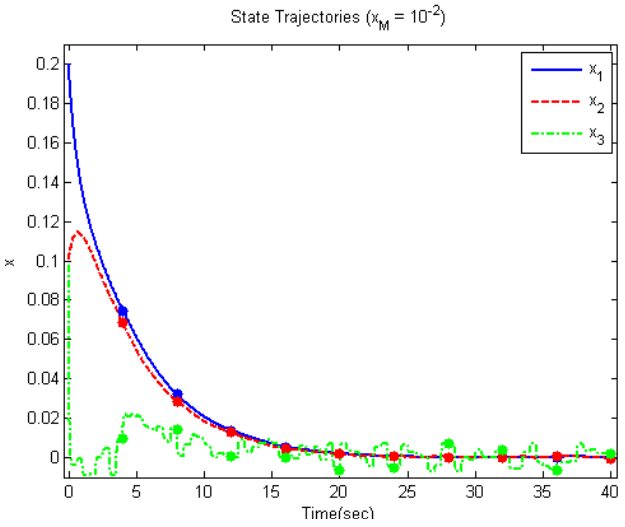
표 1 여러 x_M 대한 각 시행별 조건수 c변화

Table 1 The variations of c of each iteration for various x_M

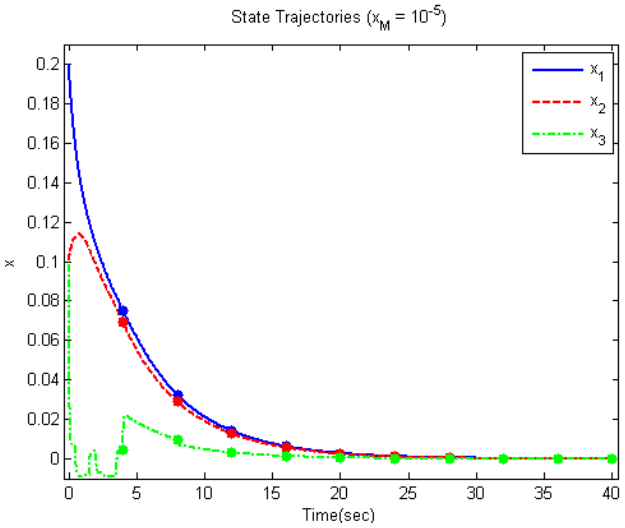
i	조건수 c					
1	7.8e+02	1.1e+03	9.0e+02	6.7e+02	5.0e+02	3.9e+11
2	6.9e+04	4.2e+05	9.1e+05	9.0e+08	9.1e+11	-
3	2.1e+05	3.6e+05	1.9e+07	9.1e+09	1.1e+12	-
4	1.8e+05	4.4e+04	5.1e+07	2.3e+10	3.0e+13	-
5	1.0e+05	4.0e+05	4.8e+08	3.2e+13	1.0e+14	-
6	7.5e+04	2.1e+05	4.0e+07	5.0e+12	7.1e+13	-
7	2.5e+05	5.7e+04	2.1e+07	1.9e+12	3.2e+13	-
8	1.3e+05	1.6e+05	2.5e+07	1.4e+12	1.1e+12	-
9	6.0e+04	1.0e+06	1.6e+06	2.9e+11	3.1e+12	-
10	8.0e+05	1.7e+05	5.8e+05	1.8e+11	6.5e+12	-
w_M	1.05	1.0e-02	1.0e-05	9.1e-09	5.7e-13	0
x_M	1	10^{-2}	10^{-5}	10^{-8}	10^{-12}	0

한편, 그림 2는 $x_M=10^{-2}$와 $x_M=10^{-5}$에 대해 제안한 정책반복법을 적용하였을 때 상태변수의 궤적이다. 예상한 결과와 마찬가지로, x_M이 더 작아지게 되면, 상태변수가 더 '0'에 가까운 상태로 수렴함을 볼 수 있다. 하지만, 충분한 탐색을 통해 조건수 c를 감소시켜 수치적 안정성을 높이기 위하여는 상태변수의 수렴성을 그만큼 희생하여야만 한다. 다른 극단적인 예로 표 1의 $x_M=1$인 경우를 살펴보면, 탐색 $w(t)$를 통해 매우 넓은 상태공간을 탐색하지만, 수치적 특성(조건수)은 크게 좋아지지 않고 $x_M=10^{-2}$경우와 매우 비슷한 수준을 유지하는 것을 볼 수 있다. 따라서, 본 모의실험의 경우, $x_M=10^{-2}$일 때 수렴성과 충분한 탐색간에 적절한 균형을 이룬다고 할 수 있으며, 더 큰 x_M값을 사용하는 것은 무의미하다. 이렇듯 충분한 탐색을 통해 충분히 작은 조건수 c를 가지는 경우, 제안한 정책반복법을 통해 학습한 행

럴 P_i는 (21)의 P^*로 수렴함을 그림 1(b)를 통해 확인할 수 있고, 따라서 모의실험을 통해 제안된 탐색화된 알고리즘의 LQ-최적제어 문제의 최적해 P^*로의 수렴성 및 그 성능을 검증하였다.



(a)



(b)

그림 2 (a) $x_M = 10^{-12}$, (b) $x_M = 10^{-5}$일 때의 탐색화된 정책반복법의 반복시행에 따른 상태변수 궤적

Fig. 2 System state trajectories when explorized PI is applied with – (a) $x_M = 10^{-2}$, (b) $x_M = 10^{-5}$

6. 결 론

본 논문에서는 상태공간의 탐색을 통해 효율적인 학습을 가능하게 하는 탐색화된 정책반복법을 제안하였다. 제안한 알고리즘의 안정성과 LQ-최적제어 문제의 최적해 P^*, K^*로의 수렴성을 증명하였고, 탐색과 이용의 균형을 위한 탐색신호의 설계방법을 소개하였다. 제안한 방법에 대한 모의실험 결과, "충분히 큰, 잘 설계된 탐색신호"를 통해 알고리즘의 수치적 안정성과 학습능력이 기존 정책반복법 [8],[9]에 비해 향상됨을 확인할 수 있었다. 향후에는, 본 연구를 바탕으로, 탐색과 이용의 균형 및 영속여기 조건과 관련된 더 발전된 탐색신호 설계방법 및 분석, 잡음처리 및 시스템 확장 등의 방향으로 연구를 진행하는 것이 필요하다.

감사의 글

본 연구는 2011년도 두뇌한국 21사업과 지식경제부의 재원으로 한국에너지 기술평가원(KETEP)의 지원을 받아 수행한 연구 과제입니다. (No. 20114010100590)

참 고 문 헌

[1] R. A. Howard, *Dynamic Programming and Markov Processes*, *Cambridge, MA: MIT Press*, 1960.

[2] R. S. Sutton and A. G.. Barto, *Reinforcement Learning: an introduction*, MIT Press, Cambridge, Massachussetts, 1998.

[3] F. Y. Wang, H. Zhang, and D. Liu, "Adaptive dynamic programming: an introduction," *IEEE Computational Intelligent Magazine*, vol. 4, no. 2, pp. 39 - 47, 2009.

[4] J. J. Murray, C. J. Cox, G. G. Lendaris, and R. Saeks, "Adaptive dynamic programming," *IEEE Trans. Systems, Mans and Cybernetics*, vol. 32, no. 2, pp. 140 - 153, 2002.

[5] F. L. Lewis and D. Vrabie, "Reinforcement learning and adaptive dynamic programming for feedback control," *IEEE Circuits and Systems Magazine*, vol. 9, no. 3, pp. 32-50, 2009.

[6] S. J. Bradke and B. E. Ydstie, "Adaptive linear quadratic control using policy iteration," *Proc. American Control Conference*, pp. 3475-3479, 1994.

[7] K. J. Zhang, Y. K. Xu, X. Chen, and X. R. Cao, "Policy iteration based feedback control," *Automatica*, vol. 44, no. 4, pp. 1055-1061, 2008.

[8] D. Vrabie, O. Pastravanu, M. Abu-Khalaf, and F. L. Lewis, "Adaptive optimal control for continuous-time linear systems based on policy iteration," *Automatica*, vol. 45, no. 2, pp. 477-484, 2009.

[9] D. Vrabie, O. Pastravanu, and F. L. Lewis, "Policy iteration for continuous-time systems with unkown internal dynamics," *In Proc. Mediterranean Conf. Control and Automation*, Athens, Greece, 2007.

[10] L. Kleinman, "On an iterative technique for Riccati equation computations," *IEEE Trans. Automatic Control*, vol. AC-13, no. 1, pp. 114-115, 1968.

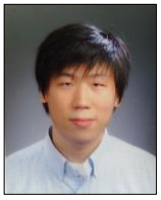
[11] R. Beard, G.. Saridis, and J. Wen, "Approximate solutions to the time-invariant Hamilton-Jacobi-

Bellman equation," *Journal of Optimization Theory and Applications*, vol. 96, no. 3, pp. 589-626, 1998.
[12] H. K. Khalil, *Nonlinear Systems*, Prentice Hall, 2002.
[13] J. C. Willems, P. Rapisarda, I. Markovsky, and B. L. M. Moor, "A note on persistency of excitation," *Systems & Control Letters*, vol. 54, no. 4, pp. 325~329, 2005.
[14] G. Strang, *Linear Algebra and Its Applications*, California: Thomson Higher Edition, 2006.
[15] B. L. Stevens and F. L. Lewis, *Aircraft Control and Simulations*, Willey, 2nd Edition, 2003.
[16] J. Y. Lee, J. B. Park, and Y. H. Choi, 'Policy-iteration-based adaptive optimal control for un-certain continuous-time linear systems with exci-tation signals, *Int'l Conf. on Control, Automation, and Systems (ICCAS)*, Ilsan, South Korea, Oct. 2010.

저 자 소 개

이 재 영 (李 在 英)

2006년 광운대학교 정보제어공학과(공학사), 2007년~현재 연세대학교 전기전자공학과 박사과정. 관심분야는 적응형 최적제어, 강화학습, 근사 동적 프로그래밍, 비선형 제어, 파워 시스템, 멀티-에이전트 시스템.



전 태 윤 (田 泰 潤)

2010년 연세대학교 전기전자공학과(공학사), 2012년 연세대학교 전기전자공학과(공학석사), 현재 동대학원 박사과정. 관심분야는 강화학습, 근사 동적 프로그래밍, 비선형 제어, 파워 시스템.

박 진 배 (朴 珍 培)

1977년 연세대학교 전기공학과(공학사), 1985년 Kansas 주립대학교 전기공학과(공학석사), 1990년 Kansas 주립대학교 전기공학과(공학박사). 1992년~현재 연세대학교 전기전자공학과 교수. 관심분야는 강인제어, 필터, 비선형 제어, 로보틱스, 퍼지 이론, 신경망 회로 이론.

최 윤 호 (崔 允 浩)

1980년 연세대학교 전기공학과(공학사), 1982년 연세대학교 전기공학과(공학석사), 1991년 연세대학교 전기공학과(공학박사). 1993년~현재 경기대학교 전자공학과 교수. 관심분야는 비선형 적응 제어, 지능 제어, 군집 제어, 로보틱스, 웨이블릿 변환 및 응용.