Proceedings of KIIS Spring Conference 2008 Vol. 18, No. 1.

# 강화학습과 메니폴드 제어기법을 이용한 걷는 로봇의 제어

# Control of Walking Robot based on Reinforcement Learning and Manifold Control

**문영준, 박주영**

고려대학교 제어계측공학과
E-mail: {dreamhill, parkj} @ korea.ac.kr

**요 약**

최근 인간을 모방하는 휴머노이드 로봇(Humanoid robot)에 대한 관심이 증가함에 따라, 기계공학, 생체공학, 제어이론 등 여러 분야에서 관련 연구가 활발히 진행되고 있다. 이에 본 논문에서는 액츄에이터(Actuator)가 없이 경사진 지면을 걸을 수 있는 두 발을 가진 패시브 로봇(Passive robot)을 대상으로 강화학습과 메니폴드(Manifold control) 기법을 사용하여 안정적으로 걸을 수 있도록 제어기(Controller)를 설계하는 방안을 고려한다.

**Key Words** : Biped robot, Limit cycle, Reinforcement learning, Manifold control, RLS-NAC

## 1. 서 론

최근 인간을 모방하는 휴머노이드 로봇(Humanoid robot)에 관심이 증가함에 따라 기계공학, 생체공학, 제어이론 등 여러 분야에서 로봇에 대한 연구가 활발히 진행되고 있으며 이를 토대로 현재는 혼다의 아시모(ASIMO)[1] 등의 놀라운 성과를 창출하고 있다. 그리고 걷고, 달리고, 뛰는 동작 등 더욱 인간과 유사한 움직임을 안정적이고, 자연스럽게 작동할 수 있게 하기 위한 연구가 지속적으로 진행되고 있다. 그러나 위와 같은 로봇의 분석적인(Analytical) 방법으로 제어 문제를 풀기 위해서는 많은 어려움이 있다[2]. 예로, 동적으로 움직이는 로봇의 많은 자유도와 그에 따른 액추에이터의 증가로 제어기의 복잡함, 지면의 불확실성, 조인트(Joint)의 비선형적인 특성, 지면과의 충돌로 인한 메카니즘(Mechanism)의 비정확성 등이 있다. 그래서 이를 극복하기 위한 대안으로 학습 알고리즘을 이용한 연구로 관심을 돌리고 있다. 이런 선행 연구에 발맞추어 환경(Environment)의 변화에 따라 자동적으로 제어 솔루션(Solution)을 찾는 강화학습과 메니폴드 컨트롤(Manifold Control)을 걷는 로봇에 적용하여 제어기를 설계한다.

강화학습[3]은 정확한 시스템의 모델없이 환경에 따라 액션(Action)을 선택하여 원하는 목표(Goal)로 가는 학습방법이다. 강화학습의 목표는 환경과 상호작용을 통해 발생하는 보상값(Reward)을 최대화하는 것이다.

메니폴드 제어(Manifold control) 기법은 고차원의 상태공간을 가지는 시스템에서의 계산적인 수고를 덜 수 있는 방법으로 각각의 로컬 정책(Local policy)을 이용해서 글로벌 정책(Global policy)을 얻어 차원의 문제를 극복한다.

위의 두 방법론을 이용하여 본 논문에서는 다양한 기계적 메카니즘을 가진 다리를 가진 로봇 중의 하나인 McGeer에 의해 제안된 단순한 걷는 로봇[4]을 안정적으로 걸을 수 있도록 제어기(Controller)를 설계한다.

본 논문의 구성은 다음과 같다: 우선 2장에서는 대상이 되는 로봇에 대한 설명을 한다. 그리고 3장에서는 로봇에 적용된 알고리즘을 설명하고, 4장에서는 시뮬레이션을 통한 실험결과를 보고, 마지막 5장에서는 결론 및 향후 연구 방향 등을 제시한다.

## 2. 로봇 모델

본 논문에서는 아래의 그림 1과 같이 무릎이 없는 단순한 메카니즘을 가진 로봇을 고려한다[5][6].

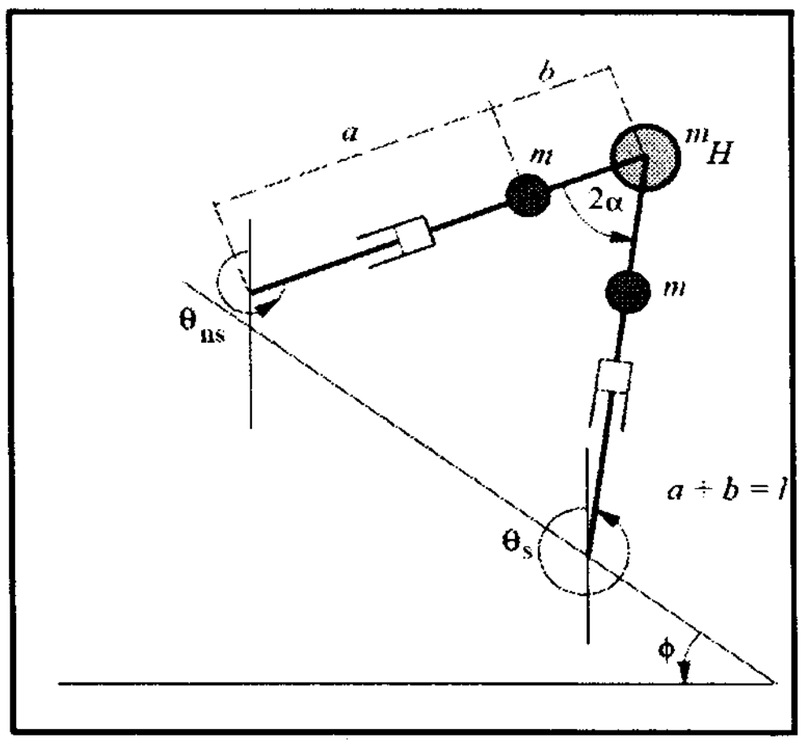


그림 1. Model of a compass gait biped robot on a slope[5].

위의 그림 1처럼 로봇은 $\phi$의 각도로 경사진 면에서 움직인다. $m$과 $m_H$는 각각 다리와 엉덩이의 점 질량이고, $l$은 발 끝에서부터 엉덩이까지의 길이이다. $\theta_{ns}$와 $\theta_s$는 수직축과 다리에 의해 만들어지는 각도이다. 마지막으로 $\alpha$는 두 다리가 경사면에 디디고 있을 때의 다리사이의 각도로 정의한다.

이 모델에 대해서 다음의 가정을 따른다.

(a) 두 다리는 동일하다.

(b) 로봇의 모션은 스윙단계(Swing stage)와 변화단계(Transition stage)로 구성되어 있다.

(c) 다리가 땅에 닿을 때 마찰로 인한 탄성과 미끄러짐이 없다.

상태는 $x=[\theta_{ns}\,\theta_s\,\dot{\theta}_{ns}\,\dot{\theta}_s]$로 표현한다. 그리고 위의 가정에 따라 스윙단계에서의 운동방정식은 다음과 같다.

$$M(\theta)\ddot{\theta}+N(\theta,\dot{\theta})\dot{\theta}+G(\theta)=Su, \quad (1)$$
$$\theta=[\theta_{ns}\,\theta_s]^T$$
$$u=[u_{ns}\,u_H\,u_s]^T$$

약자 $ns, s, H$는 각각 no swing leg, swing leg, hip을 의미한다. 그리고 $M, N, G, S$는 다음과 같이 주어진다.

$$M(\theta)=\begin{bmatrix} mb^2 & -mlbcos(\theta_s-\theta_{ns}) \\ -mlbcos(\theta_s-\theta_{ns}) & (m_H+m)l^2+ma^2 \end{bmatrix}$$
$$N(\theta,\dot{\theta})=\begin{bmatrix} 0 & mlb\dot{\theta}_s sin(\theta_s-\theta_{ns}) \\ -mlb\dot{\theta}_{ns} sin(\theta_s-\theta_{ns}) & 0 \end{bmatrix}$$
$$G(\theta)=\begin{bmatrix} mbsin\theta_{ns} \\ -(m_H l+ma+ml)\sin\theta_s \end{bmatrix} g$$
$$S=\begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

변화단계에서는 움직이는 다리가 지면에 닿고 지지하고 있는 다리는 지면으로부터 떨어지는 동작이 동시에 일어난다. 위의 가정에 의해 로봇의 다리가 지면에 디딜 때, 충격이 없으므로 로봇의 각 운동량이 보존된다[5]. 그러므로 다음과 같은 관계가 성립한다.

$$Q^+(\alpha)\dot{\theta}^+=Q^-(\alpha)\dot{\theta}^-, \quad (2)$$
$$Q^-=\begin{bmatrix} (m_H l^2+2ml^2)\cos(2\alpha)-mab-2mblcos(2\alpha)-mab \\ -mab \qquad\qquad 0 \end{bmatrix}$$
$$Q^+=\begin{bmatrix} mb^2-mblcos(2\alpha) & (ml^2+ma^2+m_H l^2)-mblcos(2\alpha) \\ mb^2 & -mblcos(2\alpha) \end{bmatrix}$$
$$2\alpha=\theta_s-\theta_{ns}$$

그림 2는 4차원의 상태를 2차원으로 투영하여 단지 한쪽 다리만을 고려한 속도에 대한 실험 결과이다. 로봇은 경사가 $\phi=3^\circ$로 기울어져 있으며 제어입력이 $u=0$일 때의 Limit cycle를 나타내었다. 기본적으로 로봇의 안정도는 주기적으로 얻어지는 Limit cycle에서 상태가 벗어남의 정도로 알 수 있다. 참고문헌[5],[6]을 통해서, 주기적으로 일정한 해를 갖는 시스템의 안정도를 궤도 안정도(Orbital stability)로 용어로 정의하며, 일정한 궤도에서 시작하여 다른 궤도를 그릴 때 이전과 유사한 형태의 궤도를 그리면 시스템은 안정하다는 의미를 갖는다.

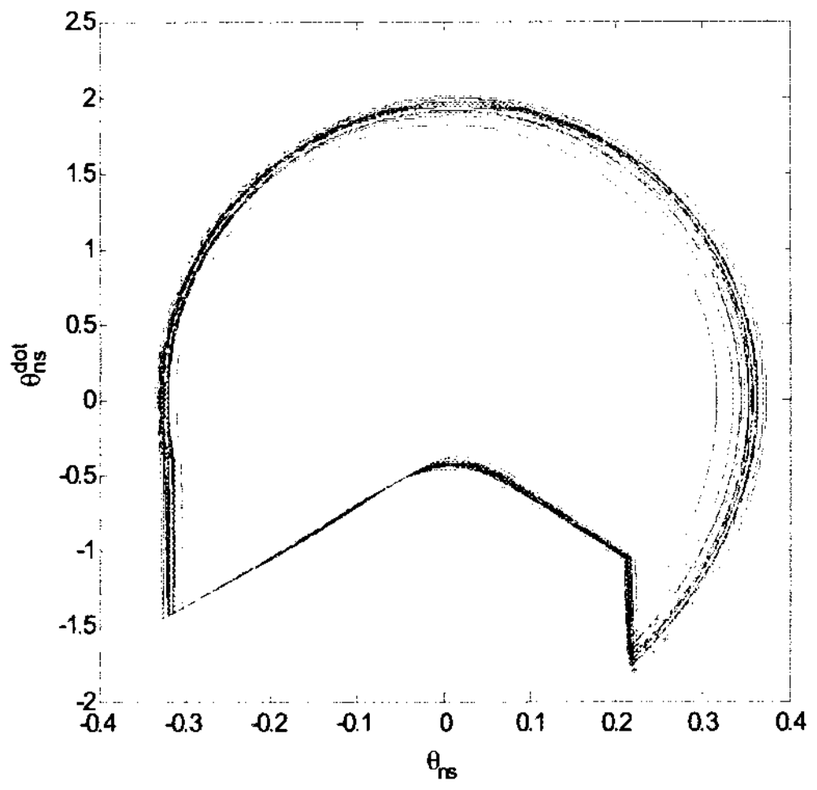


그림 2. Limit cycle for compass gait biped.

경사진 면을 이동하는 이 로봇은 지면에 닿을 때마다 위치에너지는 감소하며 그에 따라 운동에너지는 증가한다. 그래서 전체 에너지는 아래의 식 (3)을 통해 얻을 수 있으며 일정한 값을 유지한다.

$$E=0.5\dot{\theta}^T M\dot{\theta}+PE \quad (3)$$

## 3. 강화학습과 Manifold Control

2절에서 설명한 로봇을 대상으로 강화학습과 Manifold Control을 접목하여 로봇이 더 안정적으로 걷기위한 피드백 정책(Policy)을 찾는다. 동적으로 움직이는 걷는 로봇은 제어기의 복잡성과 환경에 대한 불확실성 그리고 비선형 부분으로 인해 많은 어려움이 있다. 이를 극복하기 위한 방법으로 본 논문에서는 목표를 향해 학습하는 강화학습을 구조를 취한다. 또한 상태와 액션 공간의 문제를 Manifold Control을 통해 해결한다.

### 3.1 강화학습

강화학습의 목표는 환경으로부터 피드백(Feedback)되어 오는 보상값을 최대로 하는 것이 목표이다. 알고리즘은 이전에 소개된 바 있는 RLS-NAC(Recursive least-square - natural actor-critic) 방법론을 채택하였다[7]. 기본적으로 액터-크리틱 방법론(Actor-critic method)의 구조를 가지고 있다. 액터 부분은 natural gradient method를 사용하여 액션공간에서 분포된 정책을 통해 액션을 찾는 역할이며 크리틱 부분은 가치함수(Value function)를 평가하는 부분이다. 이상의 자세한 설명과 수식은 참고문헌 [7]을 참조할 수 있다.

### 3.2 메니폴드 제어(Manifold Control)

최적제어인 강화학습을 이용한 제어방법은 상태와 액션공간에서 차원에 대한 문제에 부딪치게 된다. 걷는 로봇 같은 경우 고차원의 상태공간을 가진다. 그러므로 참고 문헌 [8]에 의해 제안된 Manifold control 방법론을 적용하여 안정하게 걸을 수 있도록 피드백 정책을 찾는다. Manifold control은 상태의 변화에 따른 공간적으로 나누어진 수용범위(Receptive field)[9]의 활성화(Activation)에 따라 최적의 정책을 찾는 것이다. 그래서 특정 범위의 로컬 정책(Local policy)은 아래 식 (3)같이 정의한다.

$$\mu_i(x) = \left[1\ (x-c_i)^T M^T\right] G_i \tag{4}$$

$c_i$는 입력공간의 위치, $M$은 스케일을 결정하는 대각행렬, $G_i$는 $(n+1)\times m$ 행렬로서 선형함수인 로컬 정책의 파라미터이고, 집합 $G=\{G_i\}$는 로컬의 정책을 발생시킨다. 여기에서 $m$은 액션의 차원이고, $n$은 상태공간의 차원이다. 이들 로컬 정책을 결합하여 최종적으로 글로벌 정책(Global policy)을 얻을 수 있으며, 아래의 식 (5)와 같다.

$$u(x) = \frac{\sum_{i=1}^{n} w_i \mu_i}{\sum_{i=1}^{n} w_i} \tag{5}$$

가중치($w_i$)는 아래의 식 (6)과 같이 가우시안 커널(Gaussian kernel) 함수로 얻는다.

$$w_i = \exp\left(-(x-c_i)^T M^T \sigma M (x-c_i)\right) \tag{6}$$

실험에서는 $M$, $\sigma$는 고정된 상수를 사용하였다.

## 4. 실 험

본 장에서는 강화학습과 Manifold Control을 이용하여 로봇에 적용한 실험과 결과에 대해 살펴본다.

본 실험은 로컬(Local) 정책을 통해 글로벌(Global) 정책 얻어 로봇의 엉덩이 부분에만 가해지는 경우를 고려한다. 로봇은 20번의 걸음을 걷는 것을 한번으로 총 5번의 에피소드(Episode)를 구성된다. 보상값은 현재의 에너지와 원하는 에너지의 차이를 이용해 참고문헌 [10]과 유사한 형태의 아래의 식 (7)과 같이 정의한다.

$$Reward = \frac{1}{2}|E - E_d|^2 \tag{7}$$

강화학습의 파라미터는 다음 표 1과 같다.

| | |
|---|---|
| Discount factor | 0.96 |
| Learning rate | 0.001 |
| forgetting factor | 0.99 |
| trace-decay parameter | 0.5 |

표 1. 강화학습에 필요한 파라미터.

Manifold Control에서 4개의 상태공간의 위치와 로컬 정책을 사용하였으며, $M=6$, $\sigma=0.3$으로 두었다.

본 논문에 적용된 알고리즘을 사용한 결과는 그림 3과 같다. 그림 3에서 볼 수 있듯이 궤도의 안정도(Orbital Stability)의 정의에 의해 그림 2보다 더 안정하다고 볼 수 있다. 이 실험을 통해 강화학습과 Manifold control을 사용하여 적절한 제어기를 설계할 수 있었다.

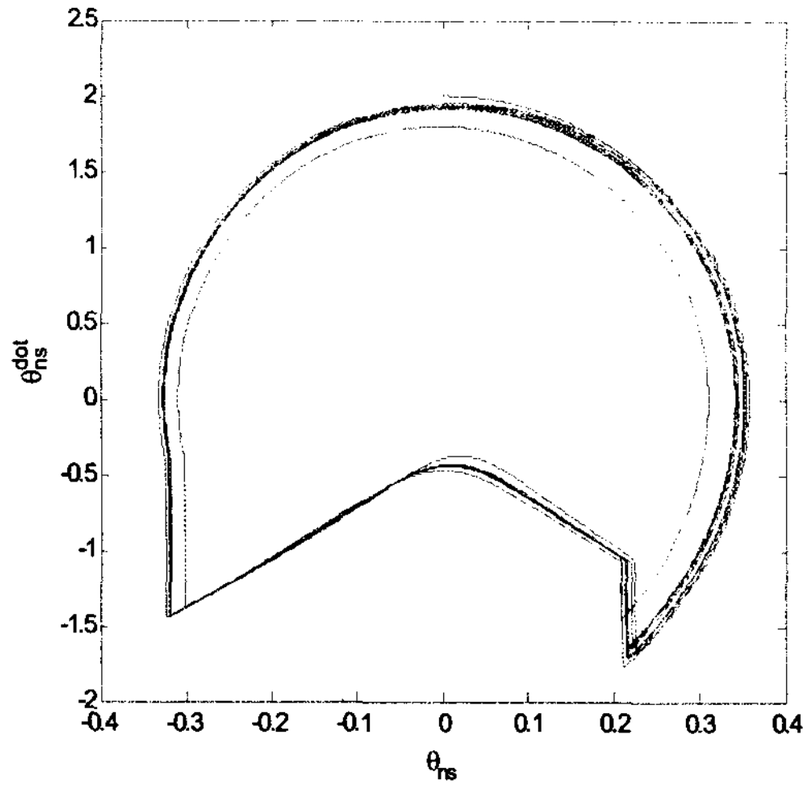


그림 3. Limit cycle for robot applied RL and Manifold control.

## 5. 결론 및 향후 연구방향

본 논문에서는 엑츄에이터(Actuator)가 없는 패시브 로봇(Passive robot)을 대상으로 로봇이 걸을 때의 안정성을 향상시키기 위해 강화학습과 Manifold control을 적용하였다. 본 논문에서는 로봇의 엉덩이 부분에만 액츄에이트되어지는 경우의 로봇을 대상으로 강화학습과 Manifold control을 적용하여 제어기를 설계하였다. 실험의 결과를 통해 엑츄에이트되지 않는 로봇보다 걷는 로봇의 안정성이 향상됨을 볼 수 있었다.

본 논문에서 다루고 있는 로봇은 실제 시스템으로 적용하기 다소 무리가 있으므로 향후 연구에 있어서 더욱 인간과 유사한 모델을 대상을 가지고 연구를 진행시킬 것이며, 실제 걷는 로봇의 부딪칠 수 있는 다양한 조건과 환경에서 강화학습의 유용성을 판단할 수 있는 선행연구를 이어가겠다.

## 참 고 문 헌

[1] Honda Motor Co., http://world.honda.com/asimo/, 2004

[2] R. L. Tedrake, "Applied Optimal Control for Dynamically Stable Legged Locomotion", PhD Thesis, MIT, 2004.

[3] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, MIT Press, 1998.

[4] T. McGeer, "Passive Dynamic Walking", International Journal of Robotics Research, vol. 9, No.2, pp. 62-82, 1990.

[5] A. Goswami, B. Espiau and A. Keramane, "Limit Cycles in a Passive Compass Gait Biped and Passivity-Mimicking Control Laws", Autonomous Robots, vol. 4, No. 3, Decemember, 1997.

[6] A. Goswami, "Compass-like biped robot Part I : Stability and bifurcation of passive gaits", Technical report, INRIA, No. 2996, 1996.

[7] J. Park, J. Kim, and D. Kang, "An RLS-based natural actor-critic algorithm for locomotion of a two-linked robot arm", Lecture Notes in Artificial Intelligence, vol. 3801, pp. 65-72, December, 2005.

[8] T. Erez and W. D. Smart, "Bipedal walking on rough terrain using manifold control", In IEEE/RSJ International Conference on Robots and Systems(IROS 2007), pp. 1539-1544, October, 2007.

[9] S. Schaal and C. G. Atkeson, "Constructive Incremental Learning From Only Local Information", vol. 10, No. 8, pp. 2047-2084, November, 1998.

[10] 주백석, Reinforcement Learning Based Controller Design for Unanalyzable Systems, 고려대학교 기계공학과 박사학위논문, 2006.