

# Model-Based Actor-Critic and Operational Space Control in the context of Motor Learning

Olivier Sigaud<sup>1</sup>, Vincent Padois<sup>1</sup>, Camille Salaün<sup>1</sup> and Anthony Truchet<sup>2</sup>

<sup>1</sup> Université Pierre et Marie Curie - Paris6  
Institut des Systèmes Intelligents et de Robotique (ISIR), CNRS FRE 2507,  
4 place Jussieu, F-75005 Paris, France

`Firstname.Lastname@isir.fr`

<sup>2</sup> CEA, LIST

18 route du Panorama, BP6,  
FONTENAY AUX ROSES, F-92265 France  
`anthony.truchet@cea.fr`

## 1 Introduction

Motor learning can be seen as the combination of two processes, namely motor adaptation and new skill acquisition [Shadmehr & Wise, 2005]. A widely accepted view of motor learning is based on the idea that human motor control is model-based, combines feedforward and feedback control and calls upon optimization processes [Wolpert & Ghahramani, 2000, Wolpert & Kawato, 1998]. Furthermore, Shadmehr promotes the idea that a movement to reach a target is specified in a task space related to the visual point of fixation before being pre-computed as a trajectory in that space and finally realized at the dynamical level in the articular or muscular space. We present elements of a model taking all these features into account and discuss both their biological relevance and their computational efficiency.

## 2 Optimization process: the actor-critic architecture

One of the most convincing computational models of motor control is the Stochastic Optimal Feedback Control (SOFC) framework [Todorov, 2004]. Indeed, this model advocates that muscles inputs are corrupted with noise proportional to their magnitude and, a constraint on reaching movements being to minimize the end-point variance, one must minimize the motor input, giving rise to the minimum intervention principle [Todorov & Jordan, 2003]. However the computational cost of this model makes it unsuitable to solve problems larger than simple planar arm movements such as the synthesis of whole body motion for humanoid robots. Recently, [Todorov & Li, 2005] presented iLQG as a computationally more efficient approach to SOFC, but this approach is still based on optimization and cannot address the control of a system with more than about ten degrees of freedom.

Instead of solving an optimization problem, one can use a reinforcement learning approach that should converge towards an optimal feedback controller

from experience. One difficulty with this perspective is that motor control has to deal with continuous actions. Most reinforcement learning strategies require to find a maximum over the action space, which is an optimization problem. To avoid this problem, alternative learning approaches to control are policy gradient methods which may converge to local maxima. The natural actor-critic algorithm is among the most prominent of such approaches in the context of robotics control problems [Peters *et al.*, 2003] and is a policy gradient method related to the actor-critic architecture. As a matter of fact, from a biological perspective, there is an important body of works suggesting that the action selection process in the brain is based on an actor-critic architecture [Joel *et al.*, 2002]. But standard actor-critic approaches are model free, thus they cannot give account of model adaptation phenomena.

### 3 Model-based control: learning the dynamics

Adaptive control [Slotine & Li, 1987] is a control scheme combining an optimal control process and a learned model of the dynamics of the system. In reinforcement learning research, model-based reinforcement learning [Sutton, 1990] is a very similar framework which combines learning a model of the transitions of the interaction process simultaneously with the incremental optimization of the control policy by applying dynamic programming back-up operations on the model. In this context, motor adaptation consists of the modification of our model of our interactions with the world when the dynamical circumstances of these interactions change.

A standard approach to learning the dynamics of the system consists in using LWPR [Vijayakumar & Schaal, 2000] with instances  $(\mathbf{x}_k, \mathbf{u}_k, \mathbf{x}_{k+1})$  giving the next as a function of the current state and the current control. For instance, the authors of [Mitrovic *et al.*, 2008] learn such a model and reproduce motor adaptation experiments described in [Shadmehr & Mussa-Ivaldi, 1994] by using the motor control approach of [Todorov & Li, 2005]. An open question is whether using LWPR instead of a neural population code as [Donchin *et al.*, 2003] does can reproduce the motor generalization abilities of human subjects studied in [Shadmehr & Moussavi, 2000].

### 4 Operational space control: learning the jacobian

Operational Space Control (OSC) [Khatib, 1987] is a framework to perform control computation in a space relative to the task, which is usually smaller than the articular space. It can thus be applied to large robotic systems and gives rise to a mathematically straightforward way of decoupling a set of tasks ranked by priority [Sentis & Khatib, 2005]. From the dynamical model, one can compute the control input of the system as a function of the joint space error with respect to a target. However, the target is specified in operational space as a desired velocity. Thus it is necessary to transform the desired velocity  $\dot{\xi}$  into a desired joints velocity  $\dot{\mathbf{q}}$  through the jacobian matrix  $J(\mathbf{q})$  associated to the considered

task.  $J(\mathbf{q})$  characterizes the kinematics of the system and depends on the state in a non-linear way. In its matrix form, this relation can be written  $\dot{\mathbf{q}} = J(\mathbf{q})^\# \dot{\xi}$ , where  $J(\mathbf{q})^\#$  is a weighted pseudo-inverse of  $J(\mathbf{q})$ , which is learnt with LWPR too.

## 5 Conclusion

In the talk, we will describe a computational model which integrates all the properties listed above. A first draft of this model has already been presented in [Salaün *et al.*, 2008].

## References

- [Donchin *et al.*, 2003] DONCHIN O., FRANCIS J. T. & SHADMEHR R. (2003). Quantifying generalization from trial-by-trial behavior of adaptive systems that learn with basis functions: Theory and experiments in human motor control. *Journal of Neuroscience*, **23**, 9032–9045.
- [Joel *et al.*, 2002] JOEL D., NIV Y. & RUPPIN E. (2002). Actor-critic models of the basal ganglia: new anatomical and computational perspectives. *Neural Networks*, **15**(4-6), 535–547.
- [Khatib, 1987] KHATIB O. (1987). A unified approach for motion and force control of robot manipulators: The operational space formulation. *IEEE Journal of Robotics and Automation*, **3**(1), 43–53.
- [Mitrovic *et al.*, 2008] MITROVIC D., KLANKE S. & VIJAYAKUMAR S. (2008). Adaptive optimal control for redundantly actuated arms. In *Proceedings of the Tenth International Conference on Simulation of Adaptive Behavior*.
- [Peters *et al.*, 2003] PETERS J., NAKANISHI J. & IJSPEERT A. J. (2003). Learning movement primitives. In *International Symposium on Robotics Research (ISRR)*.
- [Salaün *et al.*, 2008] SALAÜN C., PADOIS V. & SIGAUD O. (2008). A two-level model of anticipation-based motor learning for whole body motion. In *Proceedings of the Fourth Workshop on Anticipatory Behavior in Adaptive Learning Systems*, Munich.
- [Sentis & Khatib, 2005] SENTIS L. & KHATIB O. (2005). Control of free-floating humanoid robots through task prioritization. In *ICRA*.
- [Shadmehr & Moussavi, 2000] SHADMEHR R. & MOUSSAVI Z. M. K. (2000). Spatial generalization from learning dynamics of reaching movements. *Journal of Neuroscience*, **20**, 7807–7815.
- [Shadmehr & Mussa-Ivaldi, 1994] SHADMEHR R. & MUSSA-IVALDI F. A. (1994). Adaptive representation of the dynamics during learning of a motor task. *Journal of Neuroscience*, **14**, 3208–3324.
- [Shadmehr & Wise, 2005] SHADMEHR R. & WISE S. (2005). *The Computational Neurobiology of Reaching and Pointing*. MIT Press.
- [Slotine & Li, 1987] SLOTINE J.-J. & LI W. (1987). On the adaptive control of robot manipulators. *International Journal of Robotics Research*, **6**, 49–59.
- [Sutton, 1990] SUTTON R. S. (1990). Integrated architectures for learning, planning, and reacting based on approximating dynamic programming. In *Proceedings of the Seventh International Conference on Machine Learning*, p. 216–224, San Mateo, CA.: Morgan Kaufmann.

- [Todorov, 2004] TODOROV E. (2004). Optimality principles in sensorimotor control. *Nature Neurosciences*, **7**(9), 907–915.
- [Todorov & Jordan, 2003] TODOROV E. & JORDAN M. (2003). A minimal intervention principle for coordinated movement. In *NIPS*, p. 27–34.
- [Todorov & Li, 2005] TODOROV E. & LI W. (2005). A generalized iterative lqg method for locally-optimal feedback control of constrained nonlinear stochastic systems. In *Proceedings of the American Control Conference*, p. 300–306.
- [Vijayakumar & Schaal, 2000] VIJAYAKUMAR S. & SCHAAL S. (2000). Locally Weighted Projection Regression: An  $O(n)$  algorithm for incremental real time learning in high dimensional space. In *International Conference In Machine Learning (ICML)*, p. 1079–1086.
- [Wolpert & Ghahramani, 2000] WOLPERT D. M. & GHAHRAMANI Z. (2000). Computational principles of movement neuroscience. *Nature Neuroscience*, **3**, 1212–1217.
- [Wolpert & Kawato, 1998] WOLPERT D. M. & KAWATO M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks*, **11**(7-8), 1317–1329.