Optimization of dynamic motions for legged robots

Ioannis Havoutis[†], Jonas Buchli[‡], Darwin G. Caldwell[†] and Claudio Semini[†]
[†]Department of Advanced Robotics, Istituto Italiano di Tecnologia, (IIT), Genova, Italy
{*ioannis.havoutis, darwin.caldwell, claudio.semini*}@*iit.it*
[‡]Agile & Dexterous Robotics Lab (ADRL), ETH Zurich, Switzerland
buchlij@ethz.ch

1 Motivation

Legged robots and animals can exhibit a variety of locomotory skills, trotting and walking being the most widely used modes of locomotion. Such behaviours are often defined with the overall stability of the system in mind and are inherently of periodic nature. Legged robots, possibly equipped with arms as well, in everyday scenarios, e.g. service or forestry/agriculture, will require also skills that are not of periodic nature but rather better described as single-shot manoeuvres.

Such motions can be dynamic in nature and require the coordination of the whole body of the robot in order to accomplish their goals. Examples in a quadrupedal locomotion context can include rearing and balancing on two legs or propping the free legs against a wall to maximize reach, jumping over an obstacle or gap and squat-jumping in place. All such motions can be part of a larger locomotion-vocabulary but are not periodic in nature.

Everyday tasks in complex environments require successful agents to perform such whole body dynamic manoeuvres. The specification of such motions using a traditional motion-design approach is often cumbersome if possible at all. In addition motion authoring often fails to take into account the dynamic properties and the capabilities of the robot in question. Instead we are investigating ways of leveraging benefits from learning and optimization approaches that can be used to optimize whole body manoeuvres, for robots with legs and arms, through a somewhat high-level task specification – often an evaluation (cost) function over the overall outcome of the behaviour.

2 State of the art

Learning and optimization have a long tradition with autonomous agents but until recently were mostly restricted to problems amendable to discrete representation and/or low-dimensionality. Theodorou et. al [1] developed a framework, *Policy Improvement with Path Integrals* (PI^2), based on stochastic optimal control principles that is able to cope with high-dimensional problems, common in multi-degree-of-freedom robotics scenarios. Stulp et. al [2] demonstrated how the PI^2 approach can be used for optimizing trajectories and gain schedules for humanoids in an every-day task context. Coming close to our research focus on quadrupedal robots, Yeuhi et. al [3] used a general stochastic optimization approach, namely *covariance matrix adaptation* (CMA) [5], to synthesize dynamic manipulation motions, that require whole-body manoeuvres, to optimize the performance of the system. Fankhauser et. al [4] used a combination of simulation and hardware-based optimization with PI^2 to optimize the behaviour of a single-leg hopper over a number of scenarios and metrics, e.g. hopping height, hopping length, and periodic hopping.

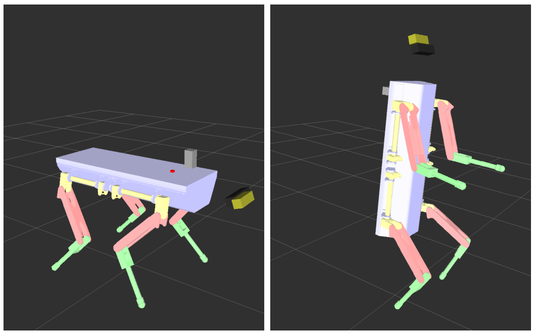

Figure 1: The robot performing a dynamic rearing motion in simulation. *Left*: the initial pose of the rearing behaviour. *Right*: the final pose after performing a dynamic rearing motion optimized with CMA [5].

3 Our approach

We follow a similar direction using the CMA algorithm [5] to optimize/learn a dynamic *rearing* motion on the quadrupedal robot - HyQ (Fig 1). Rearing is a motion common to quadrupedal animals, during which the front legs push the torso in an upright orientation and the support of the body is passed to the hind legs. This can serve to prop the front feet onto another surface or balance in a bipedal posture. Note that often such posture cannot be reached in a kinematic (static) manner.

Formally the rearing problem in the context of robotic motions is to find the appropriate joint motions that realize a rearing manoeuvre. Direct optimization of time parametrized torque profiles results in a search space that is inconveniently large. To reduce the problem dimensionality we exploit the

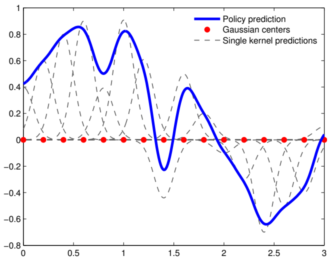

Figure 2: An example policy encoding with Gaussian kernels. The μ's are regularly spaced over time, the σ^2's are set to 0.01, and weights are randomized in the interval [-1,1].

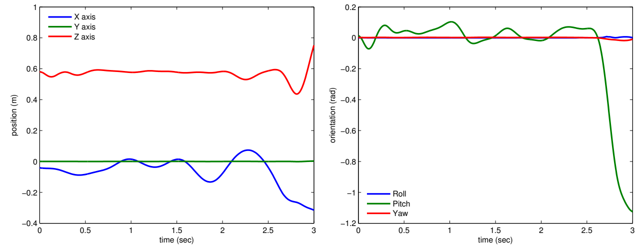

Figure 3: The absolute position (*left*) and orientation (*right*) of the robot's trunk during a rearing motion optimized with CMA.

task structure and impose that the front and hind legs form pairs that act symmetrically. In addition, we use a time-parametrized policy to encode torque profiles, represented as a weighted average of Gaussian kernels. This has the form;

$$f(t) = \sum_{i=1}^{m} w_i \phi_i(t) / \sum_{i=1}^{m} \phi(t), \ \phi_i(t) = exp\left(-\frac{1}{2\sigma_i^2}(t-\mu_i)\right),$$

where the ws are weights associated with each kernel and ϕs are Gaussian kernels described by their *means*, μ's, and *variances*, σ^2's. An example policy is depicted in Fig 2 where the μ's are equally spaced, the σ^2's are all equally set to 0.01, and the weights are randomly sampled from the interval $w \in [-1,1]$. In the case of rearing with the quadruped we use 12 such representations, one for each degree of freedom (DoF) of the robot.

4 Current result

We obtained preliminary results from a realistic simulation of the rigid body dynamics model of the HyQ with contacts. Learning in this setting consists of optimizing the set of 12 time parametrized encodings, one per DoF, as described earlier. The policy is initialized to values that keeps the quadruped standing still on all four legs (Fig. 1, left). All policies are encoded with a fixed set of 16 Gaussian kernels evenly spaced in the time interval between 0 to 3 seconds, while their variance is fixed to 0.01. The CMA algorithm is then used to optimize the weights of all policies according to a cost function. The cost function consists of a *running cost* that seeks to minimize the torques used for producing the motion, and a *final cost* that evaluates the success of the motion according to the task goals. The cost per trial has the form:

$$C(\mathbf{w}) = c_f + \int_{t_s}^{t_f} c_r(t)dt,$$

where $t \in [0,3]sec$, **b** is the vector of weights for the policy in question. The integral part penalizes the magnitude of joint torques, while the final cost term, c_f, penalizes the distance from an upright pose and the linear and angular velocities of the robot's trunk in the final pose of each trial (*rollout*).

In our experimental trials this far the policy converges after evaluating approximately 3000 trials. For the rearing task and a time horizon of 3 seconds, CMA has converged to a policy that moves the robot forward and backward two times to build momentum and then pushes of with the front legs while crouching with the hind legs, subsequently followed by an extension of the hind legs. Fig 3 shows the position and the orientation of the robot's trunk as a rearing trial with the learnt policy is performed. Overall the simulation results suggest that the behaviour can be transferred to the real hardware, i.e. considering required torques and current torque capabilities. This preliminary study is up to now limited to simulation but we aim to transfer and optimize the manoeuvre on the real hardware in the near future.

5 Best possible outcome

The rearing motion serves just as an example of the possibilities that learning/optimization approaches can offer to motion synthesis. For example, the goal of a rearing motion can be to reach the *basin of attraction* of a balancing controller, keeping the quadruped upright. Control is then switched to the balancing controller, the details of which are beyond the scope of this abstract. In the long term we would like to develop a general tool to synthesise and optimize dynamic whole-body motions that are not necessary of periodic nature. Such dynamic manoeuvres will serve to compliment and extend the capabilities of robots with arms and legs.

Acknowledgement

This research is funded by the Fondazione Istituto Italiano di Tecnologia.

References

[1] E. Theodorou, J. Buchli, S. Schaal, "A Generalized Path Integral Control Approach to Reinforcement Learning", Journal of Machine Learning Research, 11, pp.3137-3181, 2010.

[2] F. Stulp, J. Buchli, E. Theodorou, S. Schaal, "Reinforcement learning of full-body humanoid motor skills", 10th IEEE-RAS International Conference on Humanoid Robots (Humanoids), pp.405-410, 2010.

[3] Y. Abe, B. Stephens, M. P. Murphy, A. A. Rizzi, "Dynamic whole-body robotic manipulation", Proc. SPIE 8741, Unmanned Systems Technology XV, 87410V, May 17, 2013.

[4] P. Fankhauser, M. Hutter, C. Gehring, M. Bloesch, M. Hoepflinger, R. Siegwart, Reinforcement Learning of Single Legged Locomotion, Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2013.

[5] N. Hansen, S. Kern, "Evaluating the CMA Evolution Strategy on Multimodal Test Functions", Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2004.