Optimization for non-periodic dynamic motions of legged systems

Andreea Radulescu[1] Ioannis Havoutis[2] Darwin G. Caldwell[1] Claudio Semini[1]

Abstract— **Autonomous legged robots will be required to handle a wide variety of tasks and environments. While a lot of research has focused on developing their abilities for periodic locomotion tasks, little effort has been invested in devising generalized strategies for dynamic, non-periodic movements. In this work we employ the use of optimization and learning to overcome the use of motion design approaches, frequently by teleoperation, in such scenarios. We employ a realistic simulation of the HyQ2Max quadrupedal system for investigations on two distinctive tasks: rearing and posture recovery. The results obtained show the potential of optimization and learning approaches for motion synthesis in the context of complex tasks.**

Index Terms*— *optimization, parametrized policy, learning, posture recovery, dynamic motions, quadruped, legged system

I. Introduction

Biological legged systems can carry out a variety of whole body movements, in order to manipulate and traverse their environment. In transferring these skills to their robotic counterparts, most research has focused on periodic tasks, frequently designed with respect to a stability criteria, such as trotting and walking.

However, a fully autonomous system would be facing a much more diverse set of tasks, some of which are non-periodic and could be described as single-shot movements. Examples in a quadrupedal locomotion context include rearing, jumping over an obstacle or gap, squat-jumping in place and fall recovery.

Currently, the majority of robots operating in an unsafe and cluttered environment (e.g., search and rescue missions, disaster response) have to rely on teleoperation in order to achieve these objectives. Extending the autonomy of robotic legged systems with such dynamic capabilities would ease the workload of the human operators. Allowing the system to have access to a large motion library would improve the overall performance of the system, especially considering the time sensitive nature of some of the tasks.

Optimization and learning approaches could deliver solutions for such scenarios by using a a high-level task specification, in the form of an evaluation criteria of the overall performance of the emerging behavior.

II. State of the art

The relationship between learning and optimization has a long history [1], [2], but it is only in the recent past that their use has been extended to high dimensional problems, common to modern multi-degree-of-freedom robotics applications. The use of policy based approaches, rather than value function ones, allows the integration of task/domain specific knowledge in the pre-structure of the policy, thus reducing the dimensionality of the search space.

Approaches such as Policy Improvement with Path Integrals (PI^2) [3], based on stochastic optimal control principles, have been successful in generating optimal manipulation solutions for compliant robotic arms [5]. In [6], PI^2 is used on a combination of simulation and hardware based optimizations, to synthesize a periodic hopping behavior for a one-legged system, in a number of scenarios. Using iterative optimal control, the study in [**?**] delivers locally optimal solutions for a set of periodic and non-periodic tasks for a quadrupedal system.

The Covariance Matrix Adaptation (CMA) algorithm [7] has been similarly used to generate whole body movements. The study in [9] employs the CMA Evolution Strategy (CMA-ES) to obtain a optimal fast walking solution for both forward and sideways stepping. A preliminary study on the potential of CMA-ES was presented in [**?**], where the algorithm was used to obtain a rearing solution for the HyQ quadrupedal robotic system. Likewise, in [**?**] the method is employed to achieve a squat-jump movement, as well as various periodic gaits. The CMA method was shown to provide improved performance, when compared with state of the art global search methods [10], thus making it a good method of choice for such investigations.

In spite of the significant efforts in the area of fall avoidance, comparatively little research has focused on developing generalized self-righting techniques. Most work has revolved around devising specific solutions for particular systems, either at hardware design level (low center of mass, invertible robots [11]) or defining embodiment specific strategies [13]. The work in [14] is attempting to develop a generalized method for self-righting strategies, by analyzing and exploiting the given robot structure, but the results are still restricted to small dimensional designs.

III. Our approach

We employ a CMA-ES based approach to address two dynamic non-periodic tasks for a quadrupedal robot: rearing and posture recovery. The method operates by generating and evaluating a set of solutions at each iteration, after which it adapts the covariance matrix of the search distribution. We use a realistic simulation of the 80 kg HyQ2Max [15] quadrupedal robotic system with contacts. The platform was

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

[1] Department of Advanced Robotics, Istituto Italiano di Tecnologia, via Morego, 30, 16163 Genova, Italy. *email*: {`andreea.radulescu`, `darwin.caldwell`, `claudio.semini`}`@iit.it`

[2]Robot Learning and Interaction Group, Idiap Research Institute, Martigny, Switzerland *email*: `ioannis.havoutis@idiap.ch`

designed as a light-weight, robust locomotion vehicle and features 12 hydraulically actuated joints (3 per leg).

In the context of such robotic systems, the problem consists of finding the appropriate joint motions that achieve the desired movement for each given task. Direct optimization of the time parametrized joints or torques would require an inconveniently large search space. Hence, we use a parametrized policy to encode these profiles, represented as a weighted average of Gaussian kernels:

$$f(t) = \sum_{i=1}^{M} w_i \phi_i(t) / \sum_{i=1}^{M} \phi_i(t), \qquad (1)$$

$$\phi_i(t) = exp(-\frac{1}{2\sigma^2}(t-\mu_i)), \qquad (2)$$

where w_i are the weights associated with each kernel ϕ_i (defined by mean μ_i and variance σ^2). The CMA algorithm is then used to optimize the weights of all policies according to a task specific cost function (3):

$$J = p(\mathbf{s}_T, \dot{\mathbf{s}}_T) + \int_{t=0}^{T} r(\mathbf{s}_t, \dot{\mathbf{s}}_t, \mathbf{u}_t)\, \mathrm{d}t,\ t \in [0, T], \qquad (3)$$

where $\mathbf{s}_t = [x_{COM},\ y_{COM},\ z_{COM},\ roll,\ pitch,\ yaw]'$ is the trunk state and $\mathbf{u}_t$ is the set of 12 torque actuation commands at time $t \in [0, T]$. The cost function consists of a running term r that seeks to minimize the torques used for producing the motion, including realistic torque limits, and a final cost p that evaluates the success of the motion according to the task goals (desired final state). The policy is initialized to values that maintain an initial pose.

An example of such a resultant policy is depicted in Figure 1 where the $M = 16$ Gaussian kernels' means are equally spaced, the variances σ^2 are all fixed to 0.01 and the weights w_i have been sampled from $[-1, 1]$. In our experiments we use 12 such representations, one for each degree of freedom (DoF) of the quadrupedal system.

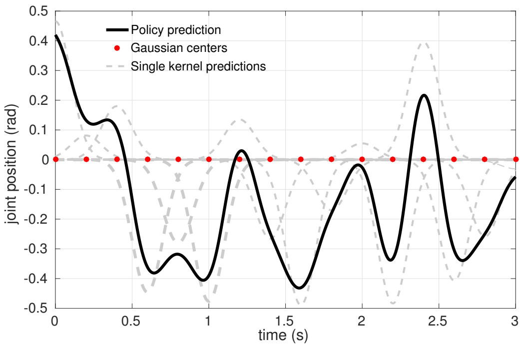

Fig. 1. Example of a policy encoded as a weighted average of Gaussian kernels: the means μ_i are equally spaced and the variances are all fixed to 0.01 and the weights have been sampled from $[-1, 1]$.

IV. Results

We present the results obtained on a realistic simulation of the HyQ2Max robotic platform, as detailed in the previous section. All policies are encoded with a fixed set of $M = 16$ Gaussian kernels evenly spaced in the time interval between 0 to T seconds, while their variance σ^2 is fixed to 0.01.

A. *Rearing*

During rearing the front legs push the torso upwards, while the lower legs are supporting the body. This behavior can serve as a preliminary stage for much more complex maneuvers (e.g., obstacle traversing, transition to bipedal posture). We note that in general such postures cannot be reached in a static manner. To further reduce the complexity of the problem, we exploit the task structure and impose the same policy for the front and hind leg pairs, respectively.

The policy is initialized to values that maintain the default pose of the system, in four legged support (Figure 2, left). The policy converged to a feasible solution within approximately 3000 trials, for an allocated time horizon of $T = 0.3\ s$. Figure 2 (right) shows the final pose reached by the system under the resultant policy, as imposed by the requirements on the position and orientation of the trunk, encoded in the terminal cost p. This also includes terms for minimizing the final angular and linear velocities.

The evolution of the position and orientation of the robot's trunk for the final trial is depicted in Figure 3, with each individual policy taking a shape similar to that in Figure 1.

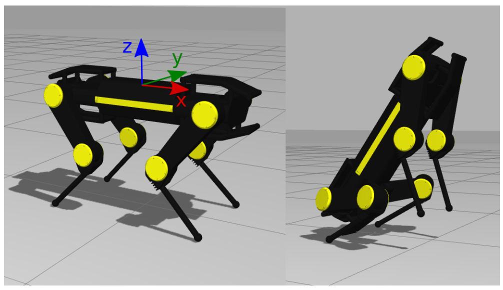

Fig. 2. HyQ2Max performing the rearing task in simulation under the resultant policy. *Left:* initial pose. *Right:* final pose.

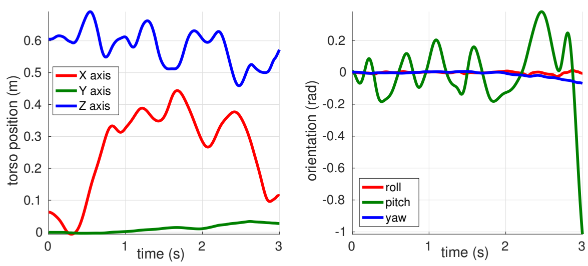

Fig. 3. HyQ2Max performing the rearing task under the resultant policy. *Left:* absolute body position. *Right:* body orientation.

B. *Posture recovery*

In the context of posture recovery we introduce a scenario where the regular locomotion task fails, due to an unexpected obstacle, and the robot finds itself in an unforeseen state (Figure 4, left). The task consists of returning the system to an upright position, that allows the resuming of the locomotion task. Unlike in the rearing scenario, due to the nature of the task, the policies of each leg are independent of each other.

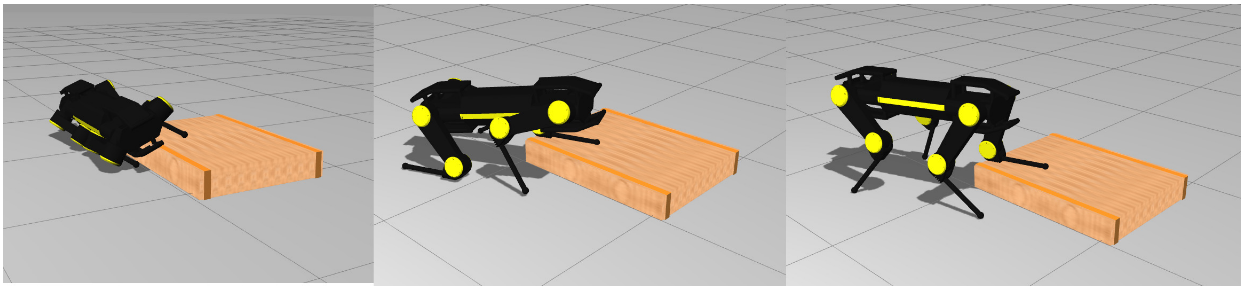

Fig. 4. The HyQ2Max performing the posture recovery task in simulation. *Left:* initial pose (unexpected fall). *Middle:* intermediary pose under the resultant policy. *Right:* final pose (after the policy is executed and the oversight is passed to a balancing controller).

The final cost p penalizes deviations from a set final state of the robot's trunk (in both linear and angular DoF). In this study the desired positions and orientations have been empirically determined, with an additional objective on minimizing final velocities.

The policy converged to a feasible solution within approximately 4000 trials, with a time horizon $T = 0.3\ s$. Figure 4 (middle) shows and intermediary pose reached by the system under the resultant policy. Figure 4 (right) depicts the final pose following the execution of the policy and after passing the oversight to a balancing controller.

The current results are limited to the simulation environment in the context of relatively simple scenarios. The presented results indicate that the behavior can be transferred to the real hardware, which we we aim to achieve in the near future, while expanding the range posture recovery and rearing scenarios addressed.

V. Future directions and Conclusion

The overall results depicted in this work serve as an example of the possibilities that optimization and learning approaches can offer to motion synthesis for complex tasks. For example, the goal of a rearing motion can be to reach the basin of attraction of a balancing controller, keeping the quadruped upright.

To increase the autonomy of the system under the suggested approach, real-time sensory information from the environment and a methodology for determining the ideal final pose of the policy will have to be integrated. In the long term we aim to develop a general tool for generating optimal dynamic whole-body motions that are not necessarily periodic in nature.

The speed of computing such solutions might not always allow for on-line optimization using conventional approaches. Machine learning methods could be employed to speed up the solution delivery time. In [16] the solutions of an optimization task are used to guide the learning of neural network controllers, for a variety of locomotion tasks on generic robotic models in simulation. Such dynamic movements will serve in complementing and extending the capabilities of robots with arms and legs.

References

[1] K. P. Bennett and E. Parrado-Hernández, "The interplay of optimization and machine learning research," *Journal of Machine Learning Research*, vol. 7, no. Jul, pp. 1265–1281, 2006.

[2] J. Kober, J. A. Bagnell, and J. Peters, "Reinforcement learning in robotics: A survey," *The International Journal of Robotics Research*, p. 0278364913495721, 2013.

[3] E. Theodorou, J. Buchli, and S. Schaal, "A generalized path integral control approach to reinforcement learning," *Journal of Machine Learning Research*, vol. 11, no. Nov, pp. 3137–3181, 2010.

[4] F. Stulp, J. Buchli, E. Theodorou, and S. Schaal, "Reinforcement learning of full-body humanoid motor skills," in *2010 10th IEEE-RAS International Conference on Humanoid Robots*, pp. 405–410, IEEE, 2010.

[5] M. Kalakrishnan, L. Righetti, P. Pastor, and S. Schaal, "Learning force control policies for compliant manipulation," in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 4639–4644, IEEE, 2011.

[6] P. Fankhauser, M. Hutter, C. Gehring, M. Bloesch, M. A. Hoepflinger, and R. Siegwart, "Reinforcement learning of single legged locomotion," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 188–193, IEEE, 2013.

[7] N. Hansen and A. Ostermeier, "Completely derandomized self-adaptation in evolution strategies," *Evolutionary computation*, vol. 9, no. 2, pp. 159–195, 2001.

[8] Y. Abe, B. Stephens, M. P. Murphy, and A. A. Rizzi, "Dynamic whole-body robotic manipulation," in *SPIE Defense, Security, and Sensing*, pp. 87410V–87410V, International Society for Optics and Photonics, 2013.

[9] N. Shafii, N. Lau, and L. P. Reis, "Learning to walk fast: Optimized hip height movement for simulated and real humanoid robots," *Journal of Intelligent & Robotic Systems*, vol. 80, no. 3-4, pp. 555–571, 2015.

[10] N. Hansen and S. Kern, "Evaluating the cma evolution strategy on multimodal test functions," in *International Conference on Parallel Problem Solving from Nature*, pp. 282–291, Springer, 2004.

[11] P. Ben-Tzvi, A. A. Goldenberg, and J. W. Zu, "Design, simulations and optimization of a tracked mobile robot manipulator with hybrid locomotion and manipulation capabilities," in *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, pp. 2307–2312, IEEE, 2008.

[12] B. Brown and G. Zeglin, "The bow leg hopping robot," in *Robotics and Automation, 1998. Proceedings. 1998 IEEE International Conference on*, vol. 1, pp. 781–786, IEEE, 1998.

[13] U. Saranli, A. A. Rizzi, and D. E. Koditschek, "Model-based dynamic self-righting maneuvers for a hexapedal robot," *The International Journal of Robotics Research*, vol. 23, no. 9, pp. 903–918, 2004.

[14] C. C. Kessens, C. T. Lennon, and J. Collins, "A metric for self-rightability and understanding its relationship to simple morphologies," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3699–3704, IEEE, 2014.

[15] C. Semini, J. Goldsmith, B. U. Rehman, M. Frigerio, V. Barasuol, M. Focchi, and D. G. Caldwell, "Design overview of the hydraulic quadruped robots hyq2max and hyq2centaur," in *The Fourteenth Scandinavian International Conference on Fluid Power (SICFP)*, 2015.

[16] I. Mordatch and E. Todorov, "Combining the benefits of function approximation and trajectory optimization," in *Robotics: Science and Systems (RSS)*, 2014.