Trial-and-Error Learning of Mismatch for Humanoid QP-based Whole-Body Control

Jonathan Spitz Karim Bouyarmane Serena Ivaldi Jean-Baptiste Mouret

I. INTRODUCTION

Like humans, humanoid robots need to carefully coordinate dozens of degrees of freedom for even the most basic tasks, like standing upright or walking [1]. This challenge can be tackled in a principled way with Quadratic Programming-based Whole-Body Controllers (QP-based WBC) [2]: at each time-step, an optimizer minimizes a cost function that describes the task(s), under constraints that model the dynamics of the robot interacting with its environment [3]. By formulating the problem as a quadratic program with linear constraints it can be solved many times per second on modern computers.

The most fundamental assumption of QP-based WBC is that the model accurately captures the dynamics of both the robot and the environment. Unfortunately, no model is ever perfect and thus such controllers often fail when a mismatch exists between the model and the real world. Even with a good dynamics model extracted by CAD and refined by dynamics parameters identification [4], [5], there are elasticities, nonlinearities and coupled dynamics effects which are impossible to model and measure accurately on a complex platform like a humanoid, especially in presence of multiple contacts [6]. The robot may also be damaged [7], or some parts may be worn out. Overall, in practice, setting a QP-based WBC for humanoids almost always involves long hand-tuning sessions of the model and the cost function [8].

The fact that a WBC can fail when its model is imperfect is not a problem *per se*: humans often fail when they have imperfect information or when their "internal model" is different (*e.g.*, when they move under perturbations [9] or after an injury). Humans, however, *learn from their mistakes*, *i.e.* they adapt their behavior until they find a way to achieve their objective. By contrast, a QP-based WBC with a fixed model and tasks structure will keep performing the same faulty behaviors.

Ideally, we would like to see humanoid robots that attempt to achieve a task with their WBC, and are able to learn from failures to improve the controller, until they achieve the desired task. We would also like the learning process to succeed after only a few trials (less than a dozen) and a few minutes [10], [11], [12], in particular because of the limited energetic autonomy of robots. The main question here is: "how to incorporate new information from the real world into a QP-based WBC?"

e-mail: firstname.lastname@inria.fr
karim.bouyarmane@loria.fr
All authors have the following affiliations:
- Inria, Villers-lès-Nancy, F-54600, France
- CNRS, Loria, UMR 7503, Vandœvre-lès-Nancy, F-54500, France
- Université de Lorraine, Loria, UMR 7503, Vandœuvre-lès-Nancy, F-54500, France

This work received funding from the Grand Est region (project "LocoLearn"), Inria (project "wbCub"), the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (GA no. 637972, project "ResiBots"), and the European Commission (GA no. 731540, H2020 project "An.Dy").

II. DEALING WITH MODEL MISMATCH

Since a QP-based WBC assumes a perfect QP optimizer, only two elements can be updated in such a trial-and-error process: the cost function (*i.e.*, the tasks) and the model (*i.e.*, the constraints). The most classic approach is to update the model according to the data acquired during each trial, *i.e.* perform a classic *model identification* [13]. Nevertheless, identifying the model of a full humanoid is far from being straightforward, as (I) identification can seldom be performed with only proprioceptive sensors [4], and (II) it might require exciting the system in specific ways which may be unsafe for the robot [14]. More importantly, some effects cannot be captured by tuning the parameters of classic models.

Our main insight is that even if a model makes inaccurate predictions for some behaviors, this is not necessarily the case for all the behaviors [15], [16], [7], [17]. Therefore a learning process could discover where the WBC makes inaccurate predictions and avoid such behaviors. In other words, we can use an imperfect model *if we know its limits*.

III. OUR APPROACH

In the present paper, we introduce this idea in the QP-based WBC framework, leading to a novel learning approach. We define the model mismatch as the difference in the QP cost calculated using the measured state and the estimated state of the robot. When the robot attempts to perform a task, we sample the robot's state and observe the mismatch for that state. We use these samples and observations to train a Gaussian Process (GP) regression model of the mismatch. For the next attempt, we modify the behavior of the QP by using two intermediate waypoints. We optimize the position of the waypoints in simulation by minimizing the tracking cost of the task while minimizing the mismatch error, as captured by the learned GP model. We re-attempt the task in the real world using the new waypoints and gather new data points for the mismatch model. This process is repeated until the QP controller is able to solve the task.

In this way, the robot identifies the regions where the model is incorrect and learns to avoid them. This new episodic, trial-and-error algorithm enables QP-based whole-body controllers to adapt in a few trials to unknown situations, like damage, and to imperfect models of the robot or its environment.

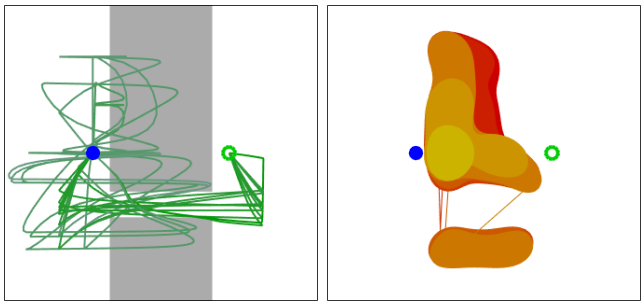

Fig. 1. Different trajectories followed by changing the intermediate waypoints in the QP task (green, left inset) and the progressively learned mismatch model (yellow to red, right inset).

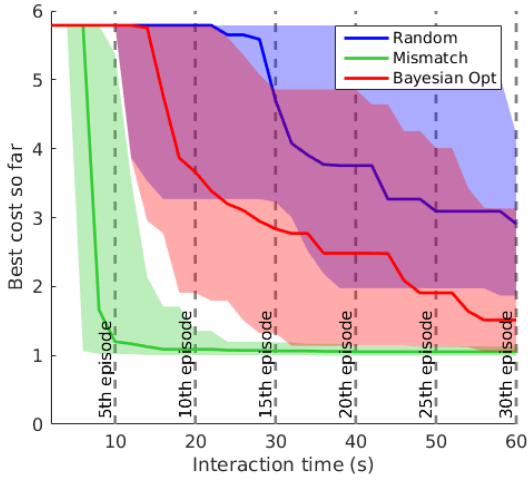

Fig. 2. Performance comparison of our mismatch learning method against random search and bayesian optimization. Mismatch learning solves the task consistently in only 5 trials.

IV. TEST CASE AND RESULTS

We show the effectiveness of our approach on a toy-example (a particle moving in 2D), shown in Fig. 1 (left inset). The task requires moving from the start (blue dot) to the goal (green circle). The model, however, is unaware of the mismatch (gray area) where the system dynamics are different. In this example, a force in the $-x$ direction is applied inside the gray area.

The mismatch model is learned gradually as the QP controller attempts to complete the task, as seen in Fig. 1. Once a good mismatch model is learned, the optimal waypoints guide the particle through the gap between the two mismatch areas, thus improving the performance of the controller.

We compared our approach against two other methods: (I) random search and (II) bayesian optimization. Fig. 2 shows that our approach obtains consistently better results faster than the other two methods. By learning the mismatch we can solve the task in as little as 5 trials.

V. CONCLUSIONS

Preliminary results for the toy-example show that (I) we can learn the mismatch area with a GP model, (II) we can complete the task reliably by staying within the areas where the QP model is correct, and (III) our method is very data efficient, i.e. it solves the task in a few trials. Future work will focus on implementing and testing the mismatch learning algorithm on the iCub robot (simulated and real).

Overall, mismatch learning offers a new view of humanoid robot learning that bridges the gap between modern whole body control and machine learning. We believe it opens many new research avenues to make humanoids robots that can both benefit from sophisticated control methods and adapt to unexpected situations.

REFERENCES

[1] S. Ivaldi, O. Sigaud, B. Berret, and F. Nori, "From humans to humanoids: the optimal control framework," *Paladyn*, vol. 3, no. 2, pp. 75–91, 2012.

[2] S. Ivaldi, J. Babič, M. Mistry, and R. Murphy, "Special issue on whole-body control of contacts and dynamics for humanoid robots," *Autonomous Robots*, vol. 40, no. 3, pp. 425–428, Mar 2016. [Online]. Available: http://dx.doi.org/10.1007/s10514-016-9545-5

[3] K. Bouyarmane and A. Kheddar, "On the dynamics modeling of free-floating-base articulated mechanisms and applications to humanoid whole-body dynamics and control," in *Humanoid Robots (Humanoids), 2012 12th IEEE-RAS International Conference on*. IEEE, 2012.

[4] S. Traversaro, A. Del Prete, R. Muradore, L. Natale, and F. Nori, "Inertial parameter identification including friction and motor dynamics," in *Humanoid Robots (Humanoids), 2013 13th IEEE-RAS International Conference on*. IEEE, 2013.

[5] S. Traversaro, A. Del Prete, S. Ivaldi, and F. Nori, "Inertial parameters identification and joint torques estimation with proximal force/torque sensing," in *Robotics and Automation (ICRA), 2015 IEEE International Conference on*. IEEE, 2015.

[6] R. Calandra, S. Ivaldi, M. P. Deisenroth, E. Rueckert, and J. Peters, "Learning inverse dynamics models with contacts," in *Robotics and Automation (ICRA), 2015 IEEE International Conference on*. IEEE, 2015.

[7] A. Cully, J. Clune, D. Tarapore, and J.-B. Mouret, "Robots that can adapt like animals," *Nature*, vol. 521, no. 7553, pp. 503–507, 2015.

[8] V. Modugno, U. Chervet, G. Oriolo, and S. Ivaldi, "Learning soft task priorities for safe control of humanoid robots with constrained stochastic optimization," in *Humanoid Robots (Humanoids), 2016 IEEE-RAS 16th International Conference on*. IEEE, 2016.

[9] E. Burdet, R. Osu, D. W. Franklin, T. E. Milner, and M. Kawato, "The central nervous system stabilizes unstable dynamics by learning optimal impedance," *Nature*, vol. 414, no. 6862, pp. 446–449, Nov. 2001.

[10] J.-B. Mouret, "Micro-data learning: The other end of the spectrum," *ERCIM News*, no. 107, p. 2, 2016.

[11] K. Chatzilygeroudis, R. Rama, R. Kaushik, D. Goepp, V. Vassiliades, and J.-B. Mouret, "Black-box data-efficient policy search for robotics," *Proc. of IEEE/RSJ IROS*, 2017.

[12] M. P. Deisenroth, D. Fox, and C. E. Rasmussen, "Gaussian processes for data-efficient learning in robotics and control," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 2, pp. 408–423, 2015.

[13] J. Hollerbach, W. Khalil, and M. Gautier, "Model identification," in *Springer Handbook of Robotics*. Springer, 2016, pp. 113–138.

[14] K. Yamane, "Practical kinematic and dynamic calibration methods for force-controlled humanoid robots," in *Humanoid Robots (Humanoids), 2011 11th IEEE-RAS International Conference on*. IEEE, 2011.

[15] S. Koos, A. Cully, and J.-B. Mouret, "Fast damage recovery in robotics with the t-resilience algorithm," *The International Journal of Robotics Research*, vol. 32, no. 14, pp. 1700–1723, 2013.

[16] S. Koos, J.-B. Mouret, and S. Doncieux, "The transferability approach: Crossing the reality gap in evolutionary robotics," *IEEE Transactions on Evolutionary Computation*, vol. 17, no. 1, pp. 122–145, 2013.

[17] M. Oliveira, S. Doncieux, J.-B. Mouret, and C. P. Santos, "Optimization of humanoid walking controller: Crossing the reality gap," in *Humanoid Robots (Humanoids), 2013 13th IEEE-RAS International Conference on*. IEEE, 2013.