Semi-parametric Approaches to Learning in Model-Based Hierarchical Control of Complex Systems

Munzir Zafar, Areeb Mehmood, Mouhyemen Khan,
Shimin Zhang, Muhammad Murtaza, Victor Aladele,
Evangelos A. Theodorou, Seth Hutchinson, and Byron Boots

Georgia Institute of Technology, Atlanta GA 30332, USA

Abstract. For systems with complex and unstable dynamics, such as humanoids, the use of model-based control within a hierarchical framework remains the tool of choice. This is due to the challenges associated with applying model-free reinforcement learning on such problems, such as sample inefficiency and limits on exploration of state space in the absence of safety/stability guarantees. However, relying purely on physics-based models comes with its own set of problems. For instance, the necessary limits on expressiveness imposed by committing to fixed basis functions, and consequently, their limited ability to learn from data gathered on-line. This gap between theoretical models and real-world dynamics gives rise to a need to incorporate a learning component at some level within the model-based control framework. In this work, we present a highly redundant wheeled inverted-pendulum humanoid as a testbed for experimental validation of some recent approaches proposed to deal with these fundamental issues in the field of robotics, such as: 1. Semi-parametric Gaussian Process-based approaches to computed-torque control of serial robots [1] 2. Probabilistic Differential Dynamic Programming framework for trajectory planning by high-level controllers [2,3] 3. Barrier Certificate based safe-learning approaches for data collection to learn the dynamics of inherently unstable systems [4]. We discuss how a typical model-based hierarchical control framework can be extended to incorporate approaches for learning at various stages of control design and hierarchy, based on the aforementioned tools.

Keywords: Hierarchical Control, Model-based Control, Wheeled Inverted Pendulum Humanoids, Semi-Parametric Model, Safe learning, Probabilistic Trajectory Optimization

1 Introduction

Wheeled inverted pendulum (WIP) systems offer fast and efficient locomotion along with the ability to deal with very heavy payloads. This ability allows them to compensate large external forces by readily adjusting their center of mass (CoM). **Golem Krang** [6] is a tree-structured serial robot with two serial

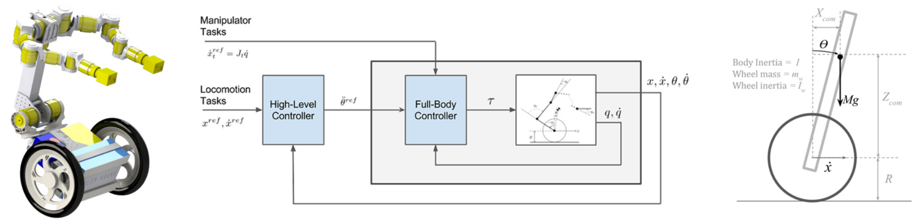

Fig. 1. Krang (left) with hierarchical control (middle) and its 2D model (right)

manipulators (each with 7-degrees-of-freedom (DoF)) mounted on a 3-DoF serial link (base, spine, torso). The system is mounted on a differential drive able to dynamically balance itself in an inverted pendulum configuration (see Figure 1).

A large body of literature exists to deal with the nonlinear, unstable, and under-actuated dynamics of WIP system [13–15]. Based on this, efforts are invested on WIP humanoid by attempting to control WIP independently from the control of upper body manipulators [10]. However, this approach comes with limitations in performing tasks that impose constraints on the kinematics of upper body while satisfying locomotion and stability objectives. This motivates the case for a unified approach in handling both locomotion and manipulation.

We utilize a framework similar to [5] for whole-body control of **Krang** through a hierarchical optimization-based approach and relying on physics-based modeling of the system. The framework has two levels of hierarchy: 1) *low-level controller* that considers the complete dynamics and optimizes only for one time-step, and 2) *high-level controller* that plans for longer horizons on a simplified model of the system for computational efficiency. The approach is grounded in a physics-based model of the system, associated non-holonomic constraints, and stability consideration of the zero dynamics of the system [16].

We then introduce learning approaches at various stages of the control design and hierarchy to deal with model uncertainty. The aim is to improve control performance given new observations in an online manner. We propose the use of recently developed semi-parametric approaches using Gaussian Processes (GP) to represent unknown non-linearities. We discuss the use of GP-based approaches at three stages of the design: 1) *safe-learning for data collection* on unstable dynamics, 2) *inverse dynamics (ID) using semi-parametric model* for low-level controller's torque computations, and 3) *trajectory planning and optimization* in the high-level controller.

2 Technical Approach

When a WIP humanoid is moving on the ground under nonholonomic constraints of no-slip/skid on the wheels, the robot has $n_{DoF} = 3 + n_{tree}$ DoFs which include: 1) Heading $\dot{x}$ of the robot 2) Spin $\dot{q}_0$ of the robot 3) Pitch $\dot{q}_1$ of the base link and 4) Relative motions of n_{tree} child links in the tree structure mounted on the base link. Choosing the minimum set of n_{DoF} coordinates, equations of motion (EOM) can be derived using Kane's formulation [8] as, unlike Euler-Lagrange, it is able to deal with the quasi-velocity $\dot{x}$. The system is under-actuated since the first 3 DoFs are driven by only 2 actuators *i.e.* the left and right wheel motors.

It is possible to derive the model using Featherstone's approach [9] if we do not insist on the minimum set of coordinates. Generic simulation platforms that provide efficient computation tools for highly articulated structures, like DART, model this system using $n_{full} = 8 + n_{tree}$ coordinates, which include 6 for position/orientation of the floating base, 2 for relative angles of the wheels, and n_{tree} joint motions. The formulation in minimum set of coordinates helps with analysis of zero dynamics of the system, but the use of n_{full} coordinates is indispensable if we are to make use of off-the-shelf algorithms for fast computation of quantities we need for ID control, such as the robot's inertia matrix. For this reason, equations of motion (EOM) using both sets of coordinates are useful. One informs design of the high level controller and analysis of the zero dynamics, while the latter is useful for the low-level controller that needs fast computation of full dynamics of the robot. Given EOM in n_{full} coordinates, we can perform a similarity transformation on them, using a transform Jacobian relating speeds in minimum coordinates to n_{full} speeds, to obtain EOM in minimum coordinates. Similarly, the transform Jacobian can be multiplied with task Jacobians in n_{full} coordinates to obtains Jacobians for minimum set of coordinates.

Figure 1 gives an overview of the hierarchical control approach. It is similar to [5]'s work on a bipedal system that has demonstrated effective performance for walking humanoids. This is the first time that a similar approach is developed for whole-body control of WIP humanoid (see [16] for a detailed account). The low-level controller uses Quadratic Programming (QP) based optimization over joint accelerations to achieve several operational space objectives framed as weighted sums of quadratic costs. The high-level controller uses DDP over the full horizon for trajectory planning of CoM given a locomotion target, and also in a receding horizon fashion online for closed-loop tracking of the optimal trajectory [11,12]. The simplified model is able to capture the essential parts of the zero dynamics of the full system, which explains how the formulation of the high-level controller ensures stability of the zero dynamics.

The above hierarchical approach is model-based. These models, although of immense use, are based on assumptions that may not always hold (like rigid bodies, simplified friction behaviors, and linearity and invariability of identifiable parameters). In Golem Krang, all joints are actuated by harmonic drives with high frictions, and the relationship between current commands and actual torques is not reported to be linear. Thus, it violates key assumptions at play in the hierarchical whole-body control framework discussed above. To address this, we propose the use of semi-parametric representations of our dynamics for predictive control in both levels of the hierarchy.

Semi-parametric representations can be used for predictive control of dynamics in both hierarchical levels. For low-level controller, the dynamics can be represented similar to [1]. GPs incorporate parametric basis functions as priors while augmenting non-parametric kernel functions to learn unknown non-linearities missed out by their parametric counterparts. For high-level controller,

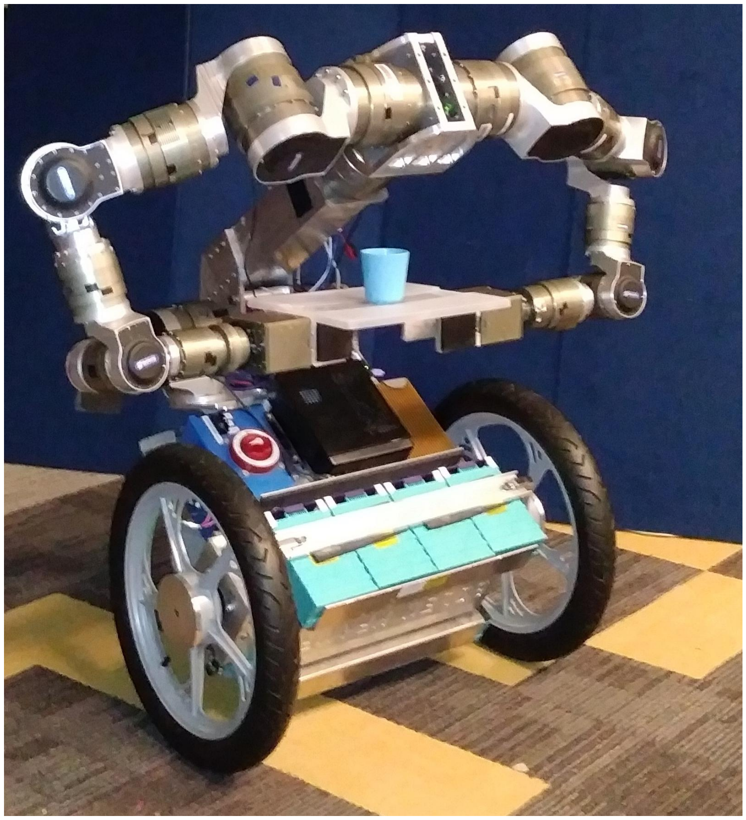

Fig. 2. Krang in its balanced pose carrying a cup on a tray

semi-parametric approach to trajectory optimization proposed in [2] can be utilized. GPs are used to represent uncertainty in the simplified model for Krang and trajectory optimization is performed in belief space. Finally, to address the problem of collecting data on our inherently unsafe and unstable dynamics, the safe-learning framework of [4] can be used. This framework relies on barrier certificates in the presence of GP-based uncertainties to reason about the safety of exploration in unknown regions of the state space for gathering data for learning.

3 Experiments & Results

The platform, Golem Krang (Fig 2), is equipped with an on-board computer, and precise position encoders for all joints. Real-time sensing and control is possible with many dedicated CAN-based physical interfaces, and the real-time operating system installed on the computer. For inter-process communication (IPC) among sensing/control and application processes with real-time guarantees, ACH [18] is used as the underlying framework. The lower body of Krang is very heavy

($\sim$ 100 kg), making it suitable to perform dynamic tasks by manipulating its weight torque intelligently. These characteristics make this platform especially suited for trying out the various ideas for dynamic control of complex systems discussed earlier. This section is divided into three subsections. We first discuss results from the simulation-based investigations of model-based control that are geared specifically to overcome various challenges in implementing model-based hierarchical control on this platform. Secondly, we discuss our result on semi-parametric learning and control of a 7DOF arm of this robot in simulation. Finally, we list the set of follow-up experiments needed to implement the full pipeline discussed in the preceding paragraphs.

3.1 Model-Based Hierarchical Control

A verification of the hierarchical control approach for whole-body control of Krang was done using DART simulation of Krang. Parameters for the model were chosen to reflect properties of the physical system. A simulation video is provided in the supplementary material [1].

Figure 3 shows the reference trajectory determined by DDP on the simplified model, and the state of the simplified model that results by applying the receding horizon control in a closed-loop manner on the simplified model. The control and state trajectories produced by this high-level controller are used as references for the low-level controller, which is also responsible for controlling the position and orientation of the end-effector. We see that the state follows the reference trajectory very closely with disturbances occurring during fast transitions owing to the disturbances caused by unmodeled full body dynamics (i.e. not modeled by the simplified model used by the high-level controller). Also, the last plot to the right is the plot of six joint torques from the full body. It is clear that the entire body is participating in performing the three tasks.

Snapshots of the full body at 4 instances during execution are shown in Figure 4. Fig 4A shows the perspective and side views of the case when we perform ID based low-level control for the full body. This control assumes perfect knowledge of model parameters. And it shows that the hierarchical controller is able to perform unified control of locomotion and manipulation. This is demonstrated by the cup on the tray carried by the arms. If we turn off whole-body control and perform simple locomotion task on the simplified model, the cup ends up falling. This is due to the changes in pitch of the body affecting the orientation of the end-effector. This however is prevented when the control is unified as shown in the figure.

In order to perform this experiment on the hardware, we needed to verify two more cases in simulation. The waist joint on the robot supports the bulk of weight of the upper body, and should not be unlocked after reaching the desired height. We investigated in simulation if it is necessary to unlock this joint to perform the tray-carrying task. We achieved control of simulated robot with a locked waist joint, by (a) excluding waist joint acceleration in the vector of

[1] Visit https://vimeo.com/user90167025/review/292912849/f662c39b8f

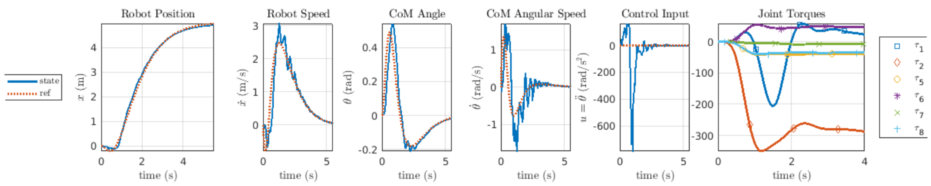

Fig. 3. Simplified model reference and state trajectories (first 5 plots). The last plot shows the resulting joint torques for $t \in (0, 4)$ seconds.

decision variables for QP-based optimization in the low-level controller, and (b) deleting the columns and rows in the various Jacobian-based cost functions [16], and EOMs for ID that correspond to the waist joint. Fig 4B shows the results. We see that when the waist is unable to lift the upper body to satisfy unified objectives, the base joint has to tilt the body a lot more to ensure CoM targets are being followed as the arms are busy in maintaining tray orientation.

A third point of investigation before hardware experiments could begin is the use of IK for control of the arms. This allows us to relax the assumption of perfect knowledge of system parameters, as kinematic references are followed by joint level PID controllers. Fig 4C shows the result. Again the waist joint is locked in this formulation. A key point to note for IK is that the robot joints are all harmonic drives with the effect of frictions much more significant than other dynamic effects. We reflected this fact in choosing our simulation frictions. In our experience, IK is a lot easier to perform in the presence of high viscous frictions, as the frictions are helpful in stabilizing the system around any configuration. This however is not the case for base, waist and torso joints. This is due to the inertial and gravitational effects of the articulated structures mounted on these three joints far outweigh the respective frictions on these joints. Therefore the use of ID-based control for these joints is indispensable. In the absence of perfect system knowledge, learning based approaches discussed above have to be utilized.

3.2 Semi-Parametric Learning On Krang's Arm

To verify the semi-parametric based approach on low-level controller, we first implement it on one of the serial manipulators on Krang. This is a necessary investigation before deploying the framework on the whole system. Non-linear error terms were included in the dynamics with the objective of GP capturing such dynamics. We successfully implemented it and verified it in simulation.

The experimental pipeline in performing the semi-parametric based approach incorporated the following steps: follow reference trajectory in operation space and collect training/testing datasets, compute hyper-parameters from training set, and perform online control.

The point cloud of end-effector positions are shown in Figure 5. Hyper-parameters were determined based on minimizing the log-marginal likelihood

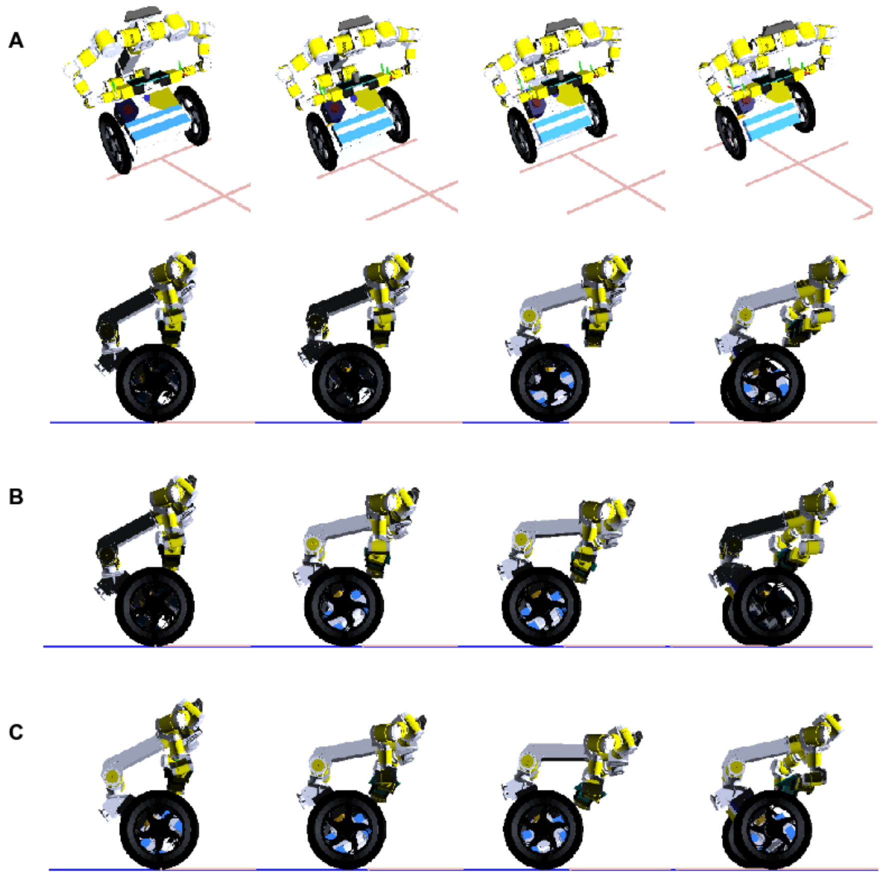

Fig. 4. Whole-body control of Krang unifying locomotion with end-effector orientation control to carry a tray with a cup. MPC based high-level controller in all cases. A) ID based low-level controller B) Same as A with locked waist joint. C) Low-level controller with IK based arm control and ID based lower body control (locked waist joint).

and online control is performed on test trajectory. The computation time for performing online control was of the order $5 \sim 12$ milliseconds per query point. The performance of GP-based torque computation outperformed purely parametric based rigid-body dynamics torque computations as shown in Figure 6. Joint 7 is not included because we are not interested in performing learning on this joint due to its relative ease of control through feedback controller. In the joint-space, the normalized mean-square error is computed and the ratio of improvement is tabulated in Table 1.

3.3 Follow-up Experiments

After having verified the results in simulations, we have the following set of hardware experiments lined up: (a) perform MPC based control of WIP on the

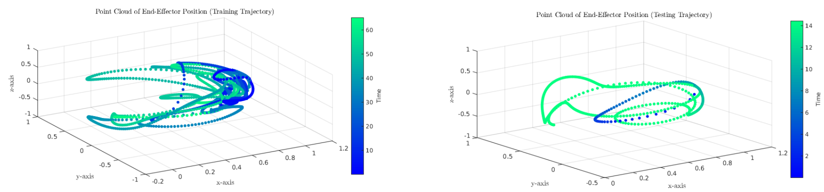

Fig. 5. Point cloud data of end-effector positions for data collection of training and testing sets.

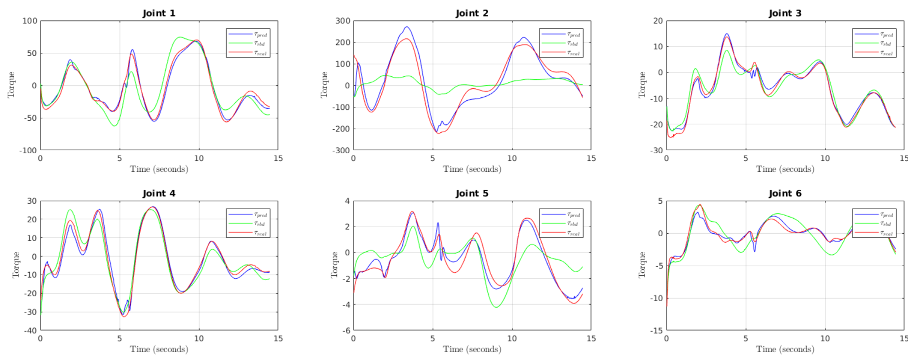

Fig. 6. Red - measured torque, Blue - GP-predicted torque, Green - RBD-based torque. The GP-predicted torque matches the measured torque fares much better than RBD-based torque.

Joint 1	Joint 2	Joint 3	Joint 4	Joint 5	Joint 6
17.4	33.5	5.8	5.1	12.3	2.01

Table 1. Table with normalized mean squared error for each joint. The error is computed between actual torque and predicted torque.

hardware with all body joints locked except the wheels, (b) attempt whole-body control using IK on arms and ID on the lower body (this attempt will have degraded or unstable performance due to erroneous dynamic model estimates), (c) safe learning based data collection to train GPs for WIP dynamics, (d) probabilistic differential dynamic programming [2] for the trajectory generation in the high-level controller using the GPs trained in the previous step, and (e) semi-parametric computed torque control [1] for ID-based low-level control of the full body.

4 Experimental Insights

Some of the key experimental insights from this work are:

1. The hierarchical whole-body control approach performs significantly better (in simulation) than the traditional implementation for WIP humanoid control. For example, in our previous implementation [7], the arms are being manipulated using joint-level PID controllers on trajectories produced by IK. A separate process running an LQR-based balancing algorithm balances the robot by rapidly updating the CoM as the robot changes its pose. In our experience with this implementation, the behavior of the robot is not suited to provide guarantees to higher-level AI algorithms. As the controllers are not unified/coordinated, it is impossible to predict the effects of one on the other, and thus makes it very difficult to plan trajectories for performing complex tasks. As an example, when moving forward, change in pitch of the robot forces the end-effector's position to change if the upper body control is not coordinated with CoM motion. Similarly, during large-force interactions, planning body poses that prevent large shifts in equilibrium positions is only possible with a unified approach. In our simulation-based experiments, we have verified that the hierarchical controller performs better in executing coordinated manipulation and locomotion task.
2. For high-level controllers, the simplification of the robot's body to a single rigid link with an equivalent CoM is sufficient. In the bipedal robots, this simplification is quite common, but this is the first time a similar simplification is being done for whole-body control of a WIP humanoid. Given that WIP dynamics are more unstable, it was important to establish that a similar simplified model succeeds to capture the essential parts of the dynamics.

Furthermore, based on the ongoing experiments, we expect to inquire the following set questions:

1. Does the safe learning approach proposed by [4] for collecting data on unstable systems and demonstrated on a simulated quadrotor example, generalize to other unstable systems like WIP humanoids, and how well does it work in practice? Are they effective in improving the control policy from the policy at work while collecting data?
2. Can fast approximate GP prediction [3] and QP-based ID on semi-parametric models of a 25 DOF system be made fast enough to allow real-time control of the system?
3. Does performing DDP in the belief space [2] significantly outperform the deterministic implementation of DDP for the high-level controller, in terms of robustness and stability?

For future work, adaptive interaction of controllers operating at different timescales is a promising area of investigation. The data collected on manual fine-tuning of the performance based on adjustment of various tuneable parameters can help develop a framework for adaptive online change of the weights/gains of the controllers according to requirements of the tasks being performed. For example, when taking a sharp turn, we may like to enforce stiffness in high-level controller objectives to make the low-level controller strictly follow its targets, but we may want to relax this stiffness when, say, moving in a straight line.

References

1. Nguyen-Tuong, Duy, and Jan Peters. "Using model knowledge for learning inverse dynamics." Robotics and Automation (ICRA), 2010 IEEE International Conference on. IEEE, 2010.
2. Pan, Yunpeng, and Evangelos Theodorou. "Probabilistic differential dynamic programming." Advances in Neural Information Processing Systems. 2014.
3. Pan, Yunpeng, et al. "Prediction under uncertainty in sparse spectrum Gaussian processes with applications to filtering and control." International Conference on Machine Learning. 2017.
4. Wang, Li, Evangelos A. Theodorou, and Magnus Egerstedt. "Safe Learning of Quadrotor Dynamics Using Barrier Certificates." arXiv preprint arXiv:1710.05472 (2017).
5. Feng, Siyuan. "Online hierarchical optimization for humanoid control." (2016).
6. Stilman, Mike, Jon Olson, and William Gloss. "Golem krang: Dynamically stable humanoid robot for mobile manipulation." Robotics and Automation (ICRA), 2010 IEEE International Conference on. IEEE, 2010.
7. Canete, Luis, and Takayuki Takahashi. "Disturbance compensation in pushing, pulling, and lifting for load transporting control of a wheeled inverted pendulum type assistant robot using the extended state observer." Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on. IEEE, 2012.
8. Kane, To R. "Dynamics of nonholonomic systems." Journal of Applied Mechanics 28.4 (1961): 574-578.
9. Featherstone, Roy. "Robot dynamics algorithms." (1984).
10. Stilman, Mike, et al. "Robots using environment objects as tools the MacGyver-paradigm for mobile manipulation." Robotics and Automation (ICRA), 2014 IEEE International Conference on. IEEE, 2014.
11. Jacobson, David H., and David Q. Mayne. "Differential dynamic programming." (1970).
12. Tassa, Yuval, Tom Erez, and Emanuel Todorov. "Synthesis and stabilization of complex behaviors through online trajectory optimization." Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on. IEEE, 2012.
13. Chan, Ronald Ping Man, Karl A. Stol, and C. Roger Halkyard. "Review of modelling and control of two-wheeled robots." Annual Reviews in Control 37.1 (2013): 89-103.
14. Nagarajan, Umashankar, George Kantor, and Ralph L. Hollis. "Trajectory planning and control of an underactuated dynamically stable single spherical wheeled mobile robot." Robotics and Automation, 2009. ICRA'09. IEEE International Conference on. IEEE, 2009.
15. Nagarajan, Umashankar. "Dynamic constraint-based optimal shape trajectory planner for shape-accelerated underactuated balancing systems." (2010): 27-31.
16. Zafar, Munzir, Seth Hutchinson, and Evangelos A. Theodorou. "Hierarchical Optimization for Whole-Body Control of Wheeled Inverted Pendulum Humanoids." arXiv preprint arXiv:1810.03074 (2018).
17. Zafar, Munzir, Akash Patel, Vlahov Bogdan, Nathaniel Glaser, Sergio Aguillera, and Seth Hutchisnon. "Online Center of Mass Estimation for a Humanoid Wheeled Inverted Pendulum Robot." arXiv preprint arXiv:1810.03076 (2018).
18. Dantam, Neil, and Mike Stilman. "Robust and efficient communication for real-time multi-process robot software." Humanoid Robots (Humanoids), 2012 12th IEEE-RAS International Conference on. IEEE, 2012.