ROBOTICS
SCIENCE AND SYSTEMS
July 12-16, 2014, University of California, Berkeley

Workshop on Dynamic Locomotion

July 13th 2014

Organized by:

Aaron Ames, Texas A & M University

Koushil Sreenath, Carnegie Mellon University

Schedule

Session on Dynamics

Time	Speaker	Title
07:55am - 08:00am	Aaron & Koushil	Welcome and Introduction
08:00am - 08-25am	Andre Seyfarth	Conceptual Models for Real-World Locomotion
08:25am - 08:50am	Hartmut Geyer	Decentralized Control in Natural and Artificial Legged Systems
08:50am - 09:15am	Jonathan Hurst	A Case for Spring-Mass Physics in Legged Robots
09:15am - 09:40am	Katja Mombaur	-
09:40am - 10:05am	Manoj Srinivasan	Energy Optimality in Novel Locomotion Tasks: Experiments, Theory, and Simple Models
10:05am - 10:20am	Coffee Break	Coffee Break
10:20am - 10-40am	Ioannis Poulakakis	Quadrupedal Running with Torso Compliance

Session on Control

Time	Speaker	Title
10:40am - 11:05am	Emo Todorov	Estimation and control as dual trajectory optimization problems
11:05am - 11:30pm	Anton Shiraev	Dynamical Walking with two and more passive degrees of freedom
11:30pm - 11:50pm	Kaveh Hamed (Jessy Grizzle)	Continuous-Time Controllers for Robust Stabilization of 3D Bipedal Walking
11:50pm - 03:00pm	World Cup & Lunch Break	Lunch and FIFA World Cup Final
03:00pm - 03-20pm	Koushil Sreenath	Control Lyapunov Function based Quadratic Programs for Torque Saturated Bipedal Walking
03:20pm - 03:40pm	Aaron Ames	Dynamic Multi-Contact Bipedal Walking
03:40pm - 04:00pm	Robert Gregg	Virtual Constraint Control of a Powered Prosthetic Leg: Experiments with Transfemoral Amputees

Session on Robotics

Time	Speaker	Title
04:00pm - 04:20pm	Aaron Johnson (Dan Koditschek)	Gait Design Using Self–Manipulation
04:20pm - 04:35pm	Sergey Levine (Pieter Abbeel)	Learning Locomotion Controllers via Trajectory Optimization
04:35pm - 05:00pm	Coffee Break	Coffee Break
05:00pm - 05:20pm	Scott Kuindersma (Russ Tedrake)	Whole-body Dynamic Locomotion Planning and Control for a Hydraulic Humanoid Robot
05:20pm - 05:40pm	Siyuan Feng (Chris Atkeson)	Optimization-Based Full Body Control For The DARPA Robotics Challenge
05:40pm - 06:00pm	Cenk Oguz Saglam (Katie Byl)	Biped Locomotion as a Metastable Markov Decision Process
06:00pm - 06:20pm	Pranav Bhousle	Gait planning and control of walking robots based on energy regulation between steps
06:20pm - 06:35pm	Massimo Vespignani (Auke Ijspeert)	Sensorized Foot Design for Robust Locomotion: A Study Using Cheetah-Cub
06:35pm - 06:50pm	John Schulman (Pieter Abbeel)	Learning Locomotion Controllers with a Policy Iteration Algorithms
06-50pm-06:55pm	Aaron & Koushil	Closing Remarks

Conceptual models for real-world locomotion

Andre Seyfarth, Lauflabor Locomotion Laboratory (www.lauflabor.de) , TU Darmstadt, Germany

***Abstract*—During human and animal locomotion, the dynamics of the body can be represented with the help of simplified biomechanical models, such as the spring-mass model (often described as spring-loaded inverted pendulum, SLIP model). In this model, it is assumed that the force generated by each limb resembles that of a linear spring. Furthermore, it is assumed that force generated by the leg mainly acts in leg axis, e.g. the connecting line between contact point (foot) and body (center of mass). Finally, the contact point at the ground is considered fixed during ground contact. Although all of these assumptions are considerable simplifications compared to most biological and artificial legged systems, this gait model provides a useful template for developing more detailed gait models taking leg segmentation (thigh, shank, foot), non-elastic leg properties (e.g. damping, variable leg spring properties), upper body (trunk), neuro-muscular properties (e.g. reflexes) and multiple legs (e.g. bipedal configuration) into account.**

Keywords: Legged locomotion; conceptual model; template; biomechanics; walking robot.

I. Gait Models with Rigid Legs

In the “simplest walking model” [1], bipedal legged locomotion is described with a pair of rigid legs supporting a body represented by a point mass. This model provides a theoretical basis for a series of passive dynamical walking robots [2]. These robots exhibit some mechanical stability and high energy efficiency (partially even better than in human walking). Theoretically, it is even possible to predict walking and running gaits for rigid legs with extremely high leg forces at transitions between contact and air phases [3]. Although mathematically appealing, the dynamics of these gait models and robots are considerably different to human and animal locomotion. This gap can be closed (or at least largely reduced) when taking compliant leg properties into account [4].

II. Models Based on Spring-Like Leg Function

Compliant leg function is a key property in biological limbs, which results from elastic structures (e.g. tendons) recruited by interactions between body mechanics, muscle mechanics and neural control. Assuming a spring-like leg function in a bipedal SLIP model [4] leads to predictions of large regions of stable walking and running for proper tunings of leg stiffness and leg angle of attack to locomotion speed (system energy). In contrast to previous models relying on rigid legs, this model also shows similarities between predicted and experimental ground reaction forces in both gaits. Another advantage of this model is it’s ability to be represented with larger levels of details (Fig. 1) regarding body structure (number of legs [4], leg segmentation, foot [5], trunk [6]), mechanical properties (leg masses, variable leg spring [7]) and neuro-muscular control (e.g. muscle-reflex schemes).

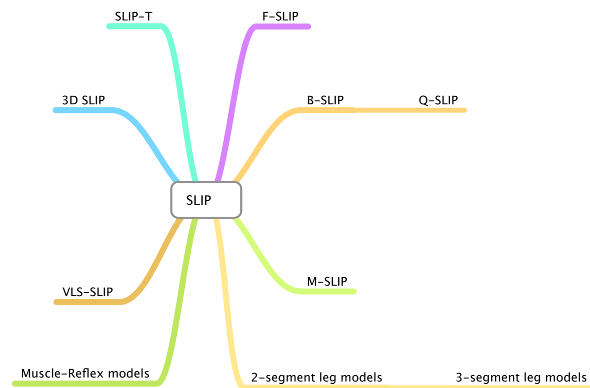


Figure 1: Hierarchy of conceptual gait models based on the SLIP template. (F=foot, B=bipedal, Q=quadrupedal, M=leg masses, T=trunk, VSL=variable leg spring).

III. Outlook

In the future, these model extensions need to be considered more systematically in order to better understand their inter-relations and to design more advanced models for biological or technical locomotor systems. Here, the focus is not only on a more realistic representation of locomotor mechanics but - and equally important - on a model-based design of biologically plausible control schemes for legged systems of different architectures.

References

[1] Garcia, M., Chatterjee, A., Ruina, A., & Coleman, M. (1998). The simplest walking model: stability, complexity, and scaling. Journal of biomechanical engineering, 120(2), 281-288.

[2] Collins, S., Ruina, A., Tedrake, R., & Wisse, M. (2005). Efficient bipedal robots based on passive-dynamic walkers. Science, 307(5712), 1082-1085.

[3] Srinivasan, M., & Ruina, A. (2006). Computer optimization of a minimal biped model discovers walking and running. Nature, 439(7072), 72-75.

[4] Geyer, H., Seyfarth, A., & Blickhan, R. (2006). Compliant leg behaviour explains basic dynamics of walking and running. Proceedings of the Royal Society B: Biological Sciences, 273(1603), 2861-2867.

[5] Maykranz, D., & Seyfarth, A. (2014). Compliant ankle function results in landing-take off asymmetry in legged locomotion. Journal of Theoretical Biology.

[6] Maus, H. M., Lipfert, S. W., Gross, M., Rummel, J., & Seyfarth, A. (2010). Upright human gait did not provide a major mechanical challenge for our ancestors. Nature communications, 1, 70.

[7] Riese, S., & Seyfarth, A. (2012). Stance leg control: variation of leg parameters supports stable hopping. Bioinspiration & biomimetics, 7(1), 16.

Decentralized control in natural and artificial legged systems

Hartmut Geyer, The Robotics Institute, Carnegie Mellon University

***Abstract*—For legged locomotion, animals and humans rely more on decentralized feedback control than on the central control approach common in humanoid robotics. We present evidence that the decentralization does not hamper fidelity and may present an alternative approach to robust locomotion of legged robots from prosthesis to humanoids.**

***Keywords*-locomotion, decentralized control, bioinspiration**

I. Introduction

Humanoids are becoming increasingly dextrous in locomotion tasks. This success is based on central control approaches that use kinematic and dynamic models of the entire robot. However, these approaches are difficult to transfer to powered prosthetic legs and exoskeletons, which form only part of the entire human-robot system. As a result, scientists and engineers in rehabilitation robotics face the challenge of having to identify alternative, decentralized control approaches to legged locomotion. This talk will present evidence that the challenge may rather be an opportunity, generating an alternative approach to robust locomotion in legged robots (Fig. 1).

II. Approach and Results

Three major functional tasks characterize bipedal systems governed by gravity: trunk balance, compliant leg behavior in stance [11, 1], and foot placement in swing [7, 6]. While the first task amounts to a separate inverted pendulum stabilization of a single body, the remaining tasks involve multiple segments whose individual controls are less clear. We show that these remaining tasks can robustly be realized using decentralized feedback control without giving up fidelity [5, 3]. We then interpret these feedbacks within computational models of the human neuromuscular system and find that these models not only demonstrate a range of human locomotion behaviors, but also generate muscle activation patterns similar to those observed in human experiments [9]. The similarity inspires decentralized, reflex-like control algorithms for artificial legs whose functionality may eventually approach the functionality of human limbs [4, 10].

III. Discussion

Collectively, the evidence presented here and obtained in animal experiments [2] suggests that humans and other legged animals decentralize control far more than common in humanoid robotics. Whether the difference is merely due to evolutionary constraints [8] or indicates a necessary shift in control paradigm for achieving high performance legged robots remains an open debate.

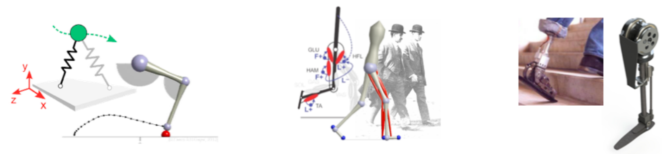


Fig. 1. Realizing major functional tasks of legged locomotion with decentralized control (left) leads to new hypotheses about the structure and function of muscle-reflex pathways (middle) and inspires control algorithms for assistive devices (right).

References

[1] R. Blickhan. The spring-mass model for running and hopping. *J. Biomech.*, 22:1217–1227, 1989.

[2] G. Courtine, Y. Gerasimenko, R. van den Brand, *et al*. Transformation of nonfunctional spinal circuits into functional states after the loss of brain input. *Nat Neurosci*, 12(10):1333–1342, 2009.

[3] R. Desai and H. Geyer. Robust swing leg placement under large disturbances. In *Proc. IEEE Int Conf Robot. Biomimetics*, 265–270, 2012.

[4] M. Eilenberg, H. Geyer, and H. Herr. Control of a Powered Ankle-Foot Prosthesis Based on a Neuromuscular Model. *IEEE Trans Neural Syst Rehabil Eng*, 18(2):164–173, 2010.

[5] H. Geyer, A. Seyfarth, and R. Blickhan. Positive force feedback in bouncing gaits? *Proc. R. Soc. Lond. B*, 270: 2173–2183, 2003.

[6] S. Kajita, F. Kanehiro, K. Kaneko, *et al.* The 3D linear inverted pendulum mode: a simple modeling for a biped walking pattern generation. In *Proc. 2001 IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, pages 239–246, 2001.

[7] M. H. Raibert. *Legged robots that balance*. MIT press, Cambridge, 1986.

[8] H. A. Simon. The architecture of complexity. *Proc. Am. Philos. Soc.*, 106(6):467–482, 1962.

[9] S. Song and H. Geyer. Robust 3D locomotion models using primarily reflex control. In *Online Proc Dyn. Walk.*, 2013.

[10] N. Thatte and H. Geyer. Towards Local Reflexive Control of a Powered Transfemoral Prosthesis for Robust Amputee Push and Trip Recovery. *IEEE Conf. Intell. Robot. Syst.*, 2014.

[11] M. van Gurp, H. C. Schamhardt, and A. Crowe. The ground reaction force pattern from the hindlimb of the horse simulated by a spring model. *Acta Anat.*, 129: 31–33, 1987.

A Case for Spring-Mass Physics in Legged Robots

Jonathan Hurst, Dynamic Robotics Laboratory (http://mime.oregonstate.edu/research/drl/), Corvallis, Oregon

I. Introduction

We suggest that spring-mass physics is the foundation for legged machines that will achieve the agility, robustness, and energy economy we see from animal locomotion.

Utilizing spring-mass physics necessarily splits the passive dynamics of the mechanical system from the control. This is antithetical to the classical control idea of creating all desired behaviors through control, using actuators simply as an imperfect implementor of the desired behaviors; instead, the passive dynamics and the controller meet in the middle, both playing their own role as partners in generating the behavior.

II. Control Benefits

The bipedal spring-mass model describes walking and running gaits, as shown in Figure1, building on prior work by Geyer et al. [1], Salazar and Carbajal [4], and Rummel et al. [3], among others. The oval point cloud in the figure shows reachable states for walking gaits at *a fixed energy level*, meaning that no energy is gained or lost in this conservative model. Different energy levels caused by actuation or uneven ground would add a third dimension out of the page. The red points in the plot represent equilibrium gaits, where each stride is the same as the last, and red points at $\psi = 0$ represent symmetrical equilibrium gaits, where the first and last half of the stance forces are exactly mirrored. The red points extend to running and hopping, but non-equilibrium reachable states for running or hopping are not shown.

Simple controllers can enable continuous transitions across this state space of gaits, passing through the snapshots shown at the top of Figure 1, from hopping in place at the top left, to running, to grounded running, to walking with a flat force profile, to walking with a classic double-hump force profile, finally transitioning to a double-leg hopping gait at the far right (not shown). Further, this diverse state map of reachable states allows convenient handling of disturbances and unpredictable terrain. A disturbance simply places the system at a different place in the state map, and it is possible to generate control policies that are exponentially stable over a large basin of attraction.

III. Implementation Benefits

The benefits of spring-mass physics are only realized when a physical system generates the appropriate passive dynamics. (1) Appropriately tuned springs are more efficient in absorbing and releasing energy in a cyclic gait than any actuator; (2) they significantly amplify the power output of actuators for cyclic gaits; (3) springs handle unexpected ground impacts elegantly, and (4) eliminate most of the inelastic collision losses; and springs (5) eliminate the massive force spikes that occur during

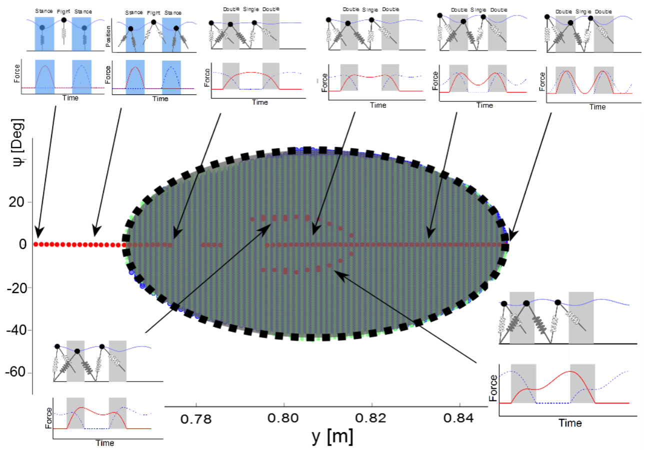


Fig. 1. This very simple model describes a continuum of gaits from hopping in place on alternative feet, through grounded running, walking, and hopping with paired feet.

ground impact with standard actuators. A spring-mass system is (6) not sensitive to impact uncertainties, such as touchdown timing errors, because the transition of force is gradual and the dynamics of the spring-mass system generally dominate whatever ground dynamics may exist. Mass-spring systems are inherently series-elastic, which means that (7) force control during stance is a natural capability[2]. Using force control, variations in animal leg stiffness or damping that have been reported in the biomechanics literature can be explained.

IV. Conclusion

Spring-mass machines can be built. They can be controlled. While there remains work to do, we suggest that design and control approaches based on spring-mass physics will be the foundation of future legged machines.

References

[1] Hartmut Geyer, Andrew Seyfarth, and Reinhard Blickhan. Compliant leg behaviour explains basic dynamics of walking and running. *Proceedings of the Royal Society B-Biological Sciences*, 273(1603):2861–2867, 2006.

[2] D. Koepl and J.W. Hurst. Impulse control for planar spring-mass running. *Journal of Intelligent and Robotic Systems*, 2013.

[3] J. Rummel, Y. Blum, and A. Seyfarth. Robust and efficient walking with spring-like legs. *Journal of Bioinspiration and Biomimetics*, 5, 2010.

[4] Harold Roberto Martinez Salazar and Juan Pablo Carbajal. Exploiting the passive dynamics of a compliant leg to develop gait transitions. *Physical Review E*, 83, 2011.

Energy optimality in novel locomotion tasks: experiments, theory, and simple models

Manoj Srinivasan
The Ohio State University, Columbus OH 43210

***Abstract*—Experiments over the last fifty years have shown that aspects of human locomotion can be predicted using energy optimality, at least approximately. Here, we describe how theoretical predictions from energy optimality can predict outcomes in novel tasks: sideways walking, reaching a destination by a time-deadline using a mixture of walking and running, and walking in non-straight-line paths (circles and other complex paths). Next, we will discuss current work with complex muscle-driven models in explaining human metabolic data and the reliability of these predictions.**

Energy optimality is a potential predictive theory for human locomotion and there is about fifty years of experimental evidence, at least partly in favor of this hypothesis. In this talk, we will briefly review the evidence for energy optimality, describe three human behavioral experiments and corresponding theory for novel locomotion patterns, such as walking sideways, using walk-run mixtures to reach a destination on a deadline, and walking in non-straight-line curves. We will discuss the goodness of match between experiment and theory in these novel tasks, and then discuss energy optimality in more complex muscle-driven models.

I. Optimality of walk-run-rest mixtures

We asked subjects (with no intentional practice) to travel a given distance overground (i.e., not on a treadmill) in a given amount of time. They could use any mixture of walking and running. The subjects mostly walked when given a lot of time and mostly ran when given very little time. Most interestingly they used a mixture of walking and running for intermediate amounts of time. This walk-run mixture is energy optimal, arising from the non-convexity of lower envelope of the walking and running energy cost curves. (Work with Leroy Long [3] and Nicholas Baker.)

II. Sideways walking

We asked subjects to walk sideways at their "comfortable" sideways walking speed. Then, we measured the energy cost of sideways walking at various speeds on a treadmill and determined the optimal speed for each person. Remarkably, we found that the distribution of preferred speeds and the optimal speeds for the 10 subjects had almost the same mean, differing only by 0.03 m/s. However, individuals were further away from the best fit of their optimal speeds – one reason could be the remarkable flatness of the energetic landscape, so that every subject was within 2.5% of their optimal energy cost (and most within 1%), despite greater variation in speed. (Work with Matthew Handford [2].)

III. Not walking in a straight line

We measured the energy cost of humans walking in circles. We found that for a given tangential speed, the cost increases with decreasing radius, as predicted by simple point-mass models. Further, using the empirically derived energy cost as a quasi-steady model of human energy costs, we can predict that humans would prefer to slow down at the higher curvatures when walking along complex curves. The quasi-steady approximation allows us to perform trajectory optimization calculations, which have classically used minimum jerk or other similar cost functions, explaining human paths qualitatively. (Work with Geoff Brown [1].)

IV. Energy optimality in muscle-driven models

We consider simple and planar muscle-driven models of humans walking, such as considered by Mombaur, Todorov, van den Bogert, Miller, Pandy, Anderson, and many other colleagues. We minimize simple models of metabolic cost and compare them to steady state human experiments. We find, perhaps not surprisingly, that while the kinematics are easier to predict, it can be more challenging to predict kinetics, especially given the flatness of the energy landscape with respect to these variables. We will discuss uniqueness of optima and motions within a given range of energies around the putative optimum.

V. Discussion

An open question is whether human coordination is predicted only on average by energy optimality, or if small individual differences in behavior can be explained by energy optimality as having origins in corresponding small differences in the individual's physiology. Answering this question may require better experimental energetics, better muscle models, better individual characterization, and better optimizations.

References

[1] G. Brown and M. Srinivasan. Curvilinear locomotion. In *World Congress of Biomechanics*, 2014.

[2] M. L. Handford and M. Srinivasan. Sideways walking: preferred is slow, slow is optimal, and optimal is expensive. *Biol. Lett.*, 2014.

[3] L. L. Long and M. Srinivasan. Walking, running, and resting under time, distance, and average speed constraints: optimality of walk-run-rest mixtures. *J R Soc Interface*, 10:20120980, 2013.

Quadrupedal Running with Torso Compliance

Ioannis Poulakakis and Qu Cao, University of Delaware, Newark DE

***Abstract*—A variety of conceptual models has been introduced to study legged locomotion. Such models offer unifying descriptions of task-level locomotion behaviors, and inform control design for legged robots. Our work focuses on models for quadrupedal running, with the objective to better understand the effect of torso compliance on gait stability and efficiency.**

Keywords: Reduced-order models, quadrupeds, compliance.

I. Introduction

On a macroscopic level, legged locomotion can be understood through reductive models, the purpose of which is to capture the dominant features of an observed gait without delving into the fine details of a robot's (or animal's) structure and morphology [1]. Restricting attention to running, much of the relevant literature revolves around simple pendulum-based, spring-mass systems, with the Spring Loaded Inverted Pendulum (SLIP) being a representative example [1]. However, the SLIP and its direct extensions cannot capture the leg-torso coordination dynamics that characterize commonly employed quadrupedal running gaits, such as the gallop and the bound. To address this issue, a series of quadrupedal models has been proposed; e.g., [2]. Yet, these models describe systems with rigid, non-deformable torsos. On the contrary, the work presented here focuses on torso flexibility, and on investigating its effects in a template setting.

II. Running Quadrupeds: Modeling and Control

Along the philosophy of the SLIP, our analysis begins with a sagittal-plane quadrupedal model with a segmented flexible torso and compliant massless legs; see Fig. 1. Despite the sensitive dependence of the motion on the torso's bending oscillations, return map studies reveal that a large variety of cyclic bounding motions can be realized passively, through the natural interaction of the model with its environment. Furthermore, for certain combinations of the torso and leg relative stiffness self-stable bounding motions emerge [3].

Next, to examine the relationship between elastic elements within the torso and the energy requirement for maintaining a gait, the basic model is extended to include non-trivial mass in the legs. By comparing the cost of transport with a rigid-torso model with the same leg mass, it is deduced that torso compliance significantly enhances energy efficiency, but *only* when the Froude number exceeds a particular value. Interestingly, this value corresponds to the Froude number at which transitions from trotting to galloping are observed in animals with drastically different morphological characteristics.

The implications of self-stable bounding orbits to control design are also discussed in the context of a hybrid control law that coordinates the torso bending oscillations with the movements of the legs. When the leg mass is considered

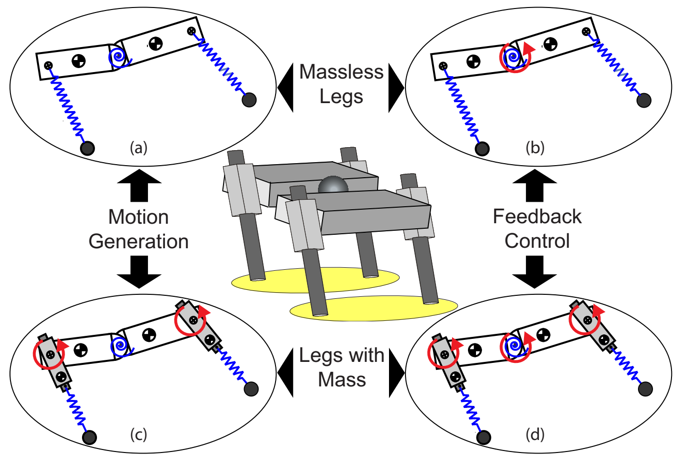


Fig. 1. A hierarchy of models for quadrupedal bounding with a flexible torso. Center: a three-dimensional virtual prototype. (a) An open-loop, energy-conservative template used to generate self-stable motions; (b): The template in (a) with an input toque available at the torso joint for feedback coordination. (c): The template in (a) with non-trivial leg mass and hip actuation to examine the effect of torso compliance on energy efficiency; (d): The template in (c) with a control input at the torso joint to realize stable bounding motions.

negligible, a single actuator at the torso is enough to stabilize the four-degree-of-freedom (4-DOF) system, rejecting significantly large disturbances without excessive effort. It turns out that the same principle of coordinating the torso bending movements with the leg hybrid oscillations is sufficient to stabilize bounding orbits in a model with non-negligible leg mass. In this case, however, a swing-leg retraction controller is needed to enlarge the corresponding domain of attraction.

III. Conclusions

A hierarchy of models with increasingly varying complexity is studied in an effort to probe the effect of compliance in the torso and legs on the generation, stability and energy efficiency of periodic quadrupedal bounding gaits. These results can be used toward a general framework for designing control laws for robotic quadrupeds with torso compliance.

IV. Acknowledgments

This work is supported in part by NSF grant CMMI-1130372 and ARO contract W911NF-12-1-0117.

References

[1] R. Full and D. E. Koditschek, "Templates and anchors: Neuromechanical hypotheses of legged locomotion on land," *Journal of Experimental Biology*, vol. 202, pp. 3325–3332, 1999.

[2] I. Poulakakis, E. G. Papadopoulos, and M. Buehler, "On the stability of the passive dynamics of quadrupedal running with a bounding gait," *The International Journal of Robotics Research*, vol. 25, no. 7, pp. 669–687, 2006.

[3] Q. Cao and I. Poulakakis, "Quadrupedal bounding with a segmented flexible torso: passive stability and feedback control," *Bioinspiration & Biomimetics*, vol. 8, no. 4, p. 046007, 2013.

Estimation and control as dual trajectory optimization problems

Emo Todorov, University of Washington

***Abstract*—We introduce a computational framework for using trajectory optimization to solve both estimation and control problems online. The trajectory is defined over a sliding window centered at the present moment. Costs over future states correspond to task errors and energy, while costs over past states correspond to data likelihoods and priors. Physical consistency priors are evaluated by a general-purpose physics engine.**

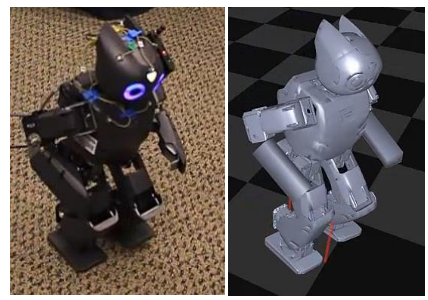
Keywords: estimation; control; optimization

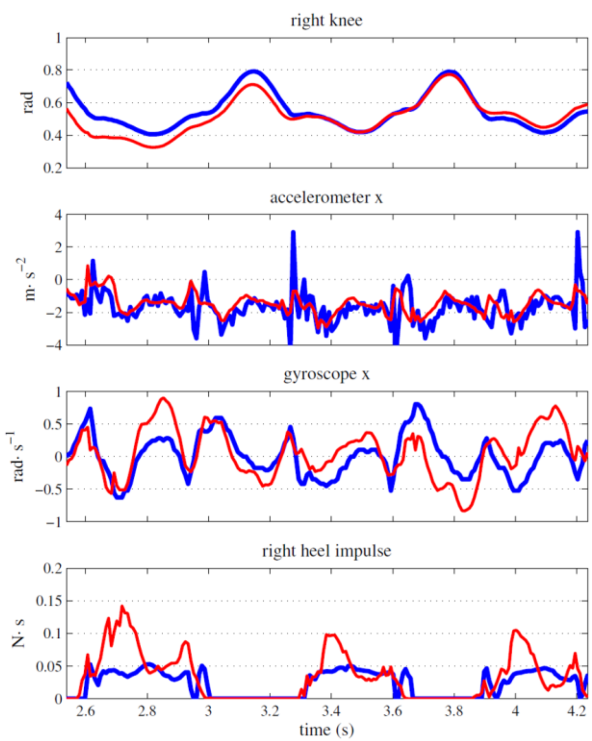
I. Introduction

Bayesian estimation and optimal control are closely related problems. The formal duality between them in the LQG case was known since Kalman, and was more recently extended to certain nonlinear systems [1, 2]. Such formulations however are global and suffer from the curse of dimensionality. In contrast, local trajectory-based methods for control have been able to generate remarkably complex behaviors, especially in contact-rich domains [3, 4]. This talk will summarize our latest results in terms of control, and then develop a related approach to estimation. The estimation approach is based on fixed-lag smoothing: the trajectory over some interval into the past is optimized with respect to the usual estimation costs. What is new here is the use of a general-purpose physics engine (MuJoCo) to define the prior. This results in dynamically-consistent estimation, incorporating not only the sensor data but also the laws of physics. The new prior is made possible by our convex contact model [5] which has an analytical inverse. Given a kinematic representation of the trajectory, we can recover both the control forces and contact forces, and evaluate all relevant cost terms. The computation scales linearly with the number of contacts and is faster than the forward dynamics simulation. This enables real-time estimation with 20-timestep trajectories for humanoid robots.

II. Results

Results will be demonstrated on tracking a walking Darwin robot. The actual walking is done by a pre-existing ZMP-like controller. The focus here is on the estimation. We record sensor data from joint potentiometers, IMU, foot contact sensors, as well as motion capture markers for ground-truth comparison. The figure shows leave-one-out cross validation. The measured output of every sensor can be predicted given the trajectory estimated from the other sensors.





References

[1] S. Mitter and N. Newton. A variational approach to nonlinear estimation. *SIAM J Control Opt*, 42, 2003.

[2] E. Todorov. General duality between optimal control and estimation. CDC 2009.

[3] I. Mordath, E. Todorov and Z. Popovic. Discovery of complex behaviors through contact-invariant optimization. SIGGRAPH 2012.

[4] Y. Tassa, T. Erez and E. Todorov. Synthesis and stabilization of complex behaviors through online trajectory optimization. IROS 2012.

[5] E. Todorov. Analytically-invertible dynamics with contacts and constraints: Theory and implementation in MuJoCo. ICRA 2014.

Dynamical walking with two and more passive degrees of freedom

Anton Shiriaev
Department of Engineering Cybernetics
Norwegian University or Science and Technology
Trondheim, Norway
E-mail: anton.shiriaev@itk.ntnu.no

Department of Automatic Control
Lund University, Lund, Sweden

Planning and controlling a motion of a mechanical system subject to dynamic constraints is the challenging task. There are many conceptual examples in the subject, where none of available analytical methods can be applied and provide solutions. Dynamical locomotion represents one of such conceptual cases where even the task of planning a gait for a system experiencing by scenario a phase with underactuation in several degrees of freedom is difficult and commonly approached by the blind search. Similar open questions are present in planning behaviors of a robot when nominal behavior should be consistent with a certain force profile specified between some of degrees of freedom or the robot and environment.

The talk is aimed at a brief discussion of several examples with original arguments proposed in searching and efficient representing feasible motions of mechanical systems including

- two degrees of freedom passive compass-gait biped;
- three degrees of freedom compass-gait biped with a suspended by springs torso and one actuator, see Fig. 1

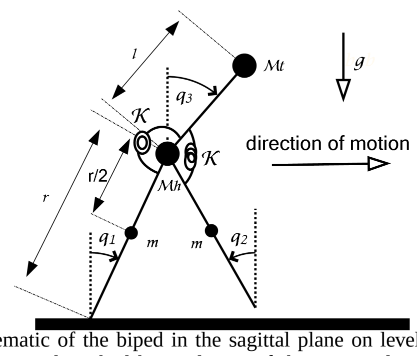


Fig. 1. Schematic of the biped in the sagittal plane on level ground. The walking motion is described by evolution of the support-leg angle q_1, the swing leg angle q_2 and the torso angle q_3. The length of each leg is denoted by r and the torso's by l. The masses of the legs, denoted by m, are lumped at $r/2$. The hip mass is denoted by M_h and the torso's by M_t, and it is lumped at a distance l from the hip. Each of the two torsional springs between the legs and the torso has stiffness coefficient K. The system is equipped with one actuator that can apply external control torque between the legs.

Both examples are important for the analysis and rich for new discoveries. Indeed, even though the compass-gait biped is one of well studied walkers in the subject several fundamental questions are left open:

- How to show that there are only two symmetric gaits for the classical settings? How to find all gaits?
- How to approximate a region of attraction for a stable gait?
- What is the mechanism of asymptotic orbital stability? To what extent the arguments of hybrid zero-dynamics, or Lyapunov function reasoning are applied? How to find such Lyapunov function for a stable gait?
- How to characterize a sensitivity of the stable gait to structural perturbations?

These and other questions become critical for the next example depicted on Fig. 1. It is likely that the system has a plenty of induced gaits and, presumably any gait of such machine should heavily exploit the passive dynamics. But how to find them? How to define a structure of a feedback controller to achieve orbital stabilization similar to observed for the passive gait of the previous example? We will partly answer some of the questions and indicate some of old and new mathematical tools helpful in solving the tasks. The following references can compliment the overview

[1] Shiriaev A.S., L. Freidovich, and M.W. Spong. Controlled invariants and trajectory planning for underactuated mechanical systems, to appear in *IEEE Transactions on Automatic Control*, 10.1109/TAC.2014.2308641

[2] La Hera P.X., A.S. Shiriaev, L.B. Freidovich, U. Mettin and S.V. Gusev. 'Stable walking gaits for three-link planar biped walker with one actuator, *IEEE Transactions on Robotics*, 29(3): 589-601, 2013.

[3] Shiriaev A.S., L.B. Freidovich and S.V. Gusev. 'Transverse Linearization for Controlled Mechanical Systems with Several Passive Degrees of Freedom, *IEEE Transactions on Automatic Control*, 55(4): 893906, 2010.

[4] Freidovich L.B., U. Mettin, A.S. Shiriaev, and M.W. Spong. 'A Passive 2 DOF Walker: Hunting for Gaits Using Virtual Holonomic Constraints, *IEEE Transactions on Robotics*, vol. 25, no. 5, pp. 12021208, 2009.

[5] Shiriaev A.S. and L.B. Freidovich. 'Transverse Linearization for Impulsive Mechanical Systems with One Passive Link, *IEEE Transactions on Automatic Control*, 54(12): 2882-2888, 2009.

Continuous-Time Controllers for Robust Stabilization of 3D Bipedal Walking

Kaveh Akbari Hamed, Brian G. Buss, and Jessy W. Grizzle, University of Michigan, USA

***Abstract*—This work presents a systematic approach to design continuous-time feedback laws to robustly and exponentially stabilize periodic orbits for legged robots. A parameterized family of continuous-time feedback laws is considered. By investigating the properties of the Poincaré map, a sensitivity analysis is presented to translate the robust stabilization problem into a set of Bilinear Matrix Inequalities (BMIs). A BMI optimization problem is then setup to tune the parameters of the continuous-time controllers.**

Keywords: Legged locomotion; stability; robust walking; bilinear matrix inequalities.

I. Introduction

We address the problem of designing continuous-time controllers to robustly and exponentially stabilize periodic walking orbits of hybrid models of bipedal robots [1]. Previous work on bipedal walking made use of a multi-level feedback control architecture in which parameters of a continuous-time controller were updated in an event-based manner to achieve stable bipedal walking [2, 3]. One drawback of employing event-based controllers to stabilize orbits is the potentially large delay between the occurrence of a disturbance and the event-based control effort.

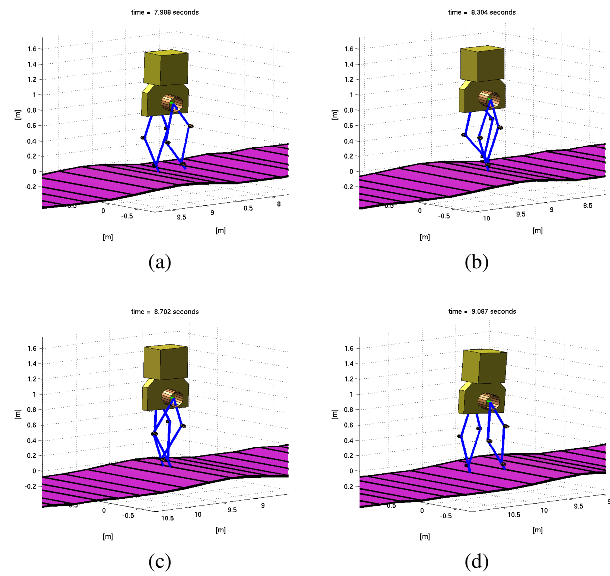


Fig. 1. Simulated 3D walking on uneven ground with point feet.

II. BMI Optimization for Robust Stability

We present a systematic method based on sensitivity analysis and bilinear matrix inequalities (BMI) to design robust and stabilizing continuous-time controllers that provide exponential stability of orbits without relying on event-based controllers. The method assumes that a parameterized family of continuous-time controllers has been designed so that (1) a periodic orbit is induced, and (2) the orbit is invariant under the choice of parameters. These assumptions are satisfied for several classes of feedback controllers, including LQR with feedforward and certain parametrizations of controllers based on virtual constraints [3]. Through a sensitivity analysis of the Jacobian of the Poincaré return map, we show how to translate the problem of selecting parameters of the continuous-time controller into a BMI optimization problem. Such an optimization problem can be solved easily with available software packages, making this a powerful tool for controller design.

We demonstrate the power of this approach in the design of a robust walking controller for an underactuated 3D bipedal robot with 13 degrees of freedom (see Fig. 1).

III. Conclusions

Previous work in the bipedal robot walking literature made use of physical intuition to design virtual constraints [3]. In contrast, the proposed BMI optimization approach provides a *systematic* way to design stabilizing and robust virtual constraints. Furthermore, the method is practical for real legged systems and can be extended to handle additional optimization goals.

IV. Acknowledgments

B. G. Buss was supported by NSF Graduate Student Research Fellowship under Grant No. DGE 1256260 and J. W. Grizzle was supported by NSF Grants ECCS-1343720 and ECCS-1231171.

References

[1] K. Akbari Hamed, B. G. Buss, and J. W. Grizzle, "Continuous-time controllers for stabilizing periodic orbits of hybrid systems: Application to an underactuated 3D bipedal robot," *53rd IEEE CDC*, under review, 2014.

[2] K. Akbari Hamed and J. W. Grizzle, "Event-based stabilization of periodic orbits for underactuated 3D bipedal robots with left-right symmetry," *IEEE TRO*, vol. 30, issue 2, pp. 365–381, 2014.

[3] C. Chevallereau, J. W. Grizzle, and C. L. Shih, "Asymptotically stable walking of a five-link underactuated 3-D bipedal robot," *IEEE TRO*, vol. 25, issue 1, pp. 37–50, 2009.

Control Lyapunov Function based Quadratic Programs for Torque Saturated Bipedal Walking

Kevin Galloway, Koushil Sreenath, Aaron D. Ames and J. W. Grizzle

***Abstract*—This work builds off of recent work on rapidly exponentially stabilizing control Lyapunov functions (RES-CLF) and presents a novel method to address actuator saturation in bipedal walking. The proposed method employs quadratic programming (QP) to implement CLF-based controllers, enabling directly incorporating user-specified input bounds into the controller. We show that even with increasingly stringent levels of actuator saturation, there is only a gradual degradation of performance while still maintaining the task of walking.**

Keywords-dynamic legged locomotion; actuator saturation; control Lyapunov functions.

I. Introduction

The method of Hybrid Zero Dynamics (HZD) [3] has been very successful in dealing with the hybrid and underactuated dynamics of legged locomotion and provides a framework for implementing formally provable controllers for achieving dynamic walking and running. Until recently, this has required either controllers with finite-time convergence or input-output linearization based controllers. Recent work on control Lyapunov function (CLF)-based controllers has enabled implementing hybrid zero dynamics using a variant of CLF, called the rapidly exponentially stabilizing control Lyapunov function (RES-CLF) [1]. This type of CLF incorporates an additional tuning parameter, enabling the control designer to directly control the rate of exponential convergence. This key feature enabled proving local exponential stability of the hybrid periodic orbit corresponding to a walking or running gait. Implementation of this controller on the bipedal robot MABEL was successful, however, user-defined actuator bounds (specified for the sake of safety) were active throughout a large portion of the walking gait. Instead of "blindly" applying such hard torque limits on the computed CLF control, this work explicitly considers the actuator saturation as part of the online control computation.

II. Approach: Control Lyapunov Function based Quadratic Programs

We consider the dynamical model of a bipedal robot and use the method of Hybrid Zero Dynamics (HZD) to design a set of output functions (called virtual constraints), such that when the outputs (and their first derivatives) are driven to zero, the system dynamics is constrained to evolve on a lower dimensional manifold.

Using such a HZD pre-control, we arrive at the closed-loop output dynamics. A control Lyapunov function is then constructed for the output dynamics such that the time-derivative of the Lyapunov function satisfies an rapidly exponential stabilizing property, i.e., $\frac{d}{dt}V \leq -\gamma/\epsilon\ V$, where γ is some

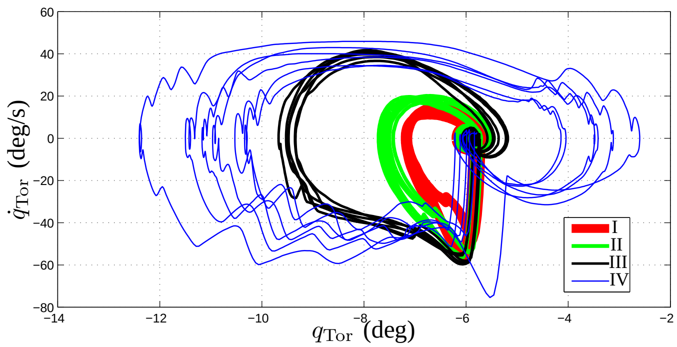


Fig. 1. Phase portrait of the torso angle for four different cases of increasing levels of input bounds. Observe that stricter saturations (increasing from case I to case IV) result in (gradual) deterioration in tracking, as evidenced by deviations of the limit cycle from the nominal orbit. However, the controller still

positive constant, and ϵ is a design parameter to control the rate of exponential convergence. Our novel approach is in implementing a controller as a quadratic program with the above expression as an inequality constraint. These are control Lyapunov function based quadratic programs (CLF-QP), introduced in [2]. Additional constraints, such as torque saturation, can then be easily incorporated into the quadratic program.

III. Results

CLF-QPs have been employed to successfully demonstrate stable walking both in simulations and experiments. Figure 1 considers several different bounds on the actuators, with the bounds getting more stringent from case I to case IV. In each case, simulations on MABEL demonstrate stable walking, although there is a gradual degradation in performance, as can be seen in deviation from the nominal fixed point orbit.

References

[1] A. D. Ames, K. Galloway, K. Sreenath, and J. W. Grizzle, "Rapidly exponentially stabilizing control lyapunov functions and hybrid zero dynamics," *IEEE Transactions on Automatic Control (TAC)*, vol. 59, no. 4, pp. 876–891, April 2014.

[2] K. Galloway, K. Sreenath, A. D. Ames, and J. W. Grizzle, "Torque saturation in bipedal robotic walking through control lyapunov function based quadratic programs," *IEEE Transactions on Robotics (TRO)*, submitted, 2014.

[3] E. R. Westervelt, J. W. Grizzle, C. Chevallereau, J. H. Choi, and B. Morris, *Feedback Control of Dynamic Bipedal Robot Locomotion*. Boca Raton, FL: Taylor & Francis/CRC Press, 2007.

Dynamic Multi-Contact Bipedal Robotic Walking

Huihua Zhao, Hereid Ayonga, Wenlong Ma, Shishir Kolathaya and Aaron D. Ames
AMBER Lab, Texas A&M University, USA

***Abstract*—This work presents a formal framework for achieving multi-contact bipedal robotic walking, and realizes this methodology experimentally on two robotic platforms: AMBER2 and ATRIAS. Inspired by the key feature encoded in human walking—multi-contact behavior—this approach begins with the analysis of human locomotion and uses it to motivate the construction of a hybrid system model representing a multi-contact robotic walking gait. Human-inspired outputs are extracted from human locomotion data to characterize the robot model, and then employed to develop the human-inspired control and optimization problem that yields stable multi-domain walking. Through a trajectory reconstruction strategy motivated by the procedure for formally generating robotic walking gaits, the formal result is successfully translated to the two physical robots.**

Key words: Bipedal walking, multi-contact, hybrid zero dynamics

I. Introduction

Human walking exhibits amazingly robustness properties and, therefore, serves as a prime example in the construction of dynamic robot walking gaits. During the course of a step, humans undergo changes in phase, i.e., changes in contact points with the environment, including a heel-lift and a toe strike. This is potentially one of the predominant features of human walking that results in both robustness and efficiency. Achieving human-like bipedal locomotion has been studied from a variety of viewpoints, yet most of which are constrained to point feet walking (hybrid zero dynamics) and flat foot walking (ZMP walking). Noticeably lacking from existing methods is a formal way to generate multi-contact locomotion in a manner that is both formally correct as well as physically realizable.

II. Approach

With the goal of exploring a method to produce multi-contact robotic bipedal locomotion, our approach begins by noting that the multi-contact behavior of human locomotion can be represented as a hybrid system. Therefore, a hybrid system with multiple domains is constructed to describe the multi-contact robotic locomotion in a general form. Further motivated by the human locomotion data, the *extended canonical walking function* (ECWF) is utilized to serve as a low dimensional representation of the human locomotion system. Human-inspired control is utilized to drive outputs of the robot to this representation at an exponential rate. Finally, a novel multi-domain optimization problem is proposed to obtain controller parameters that yield invariant tracking even through impacts. More importantly, this optimization problem is also subject to specific physical constraints, such as torque bounds and scuffing preventions; therefore, the obtained parameters can be implemented on the physical robot directly.

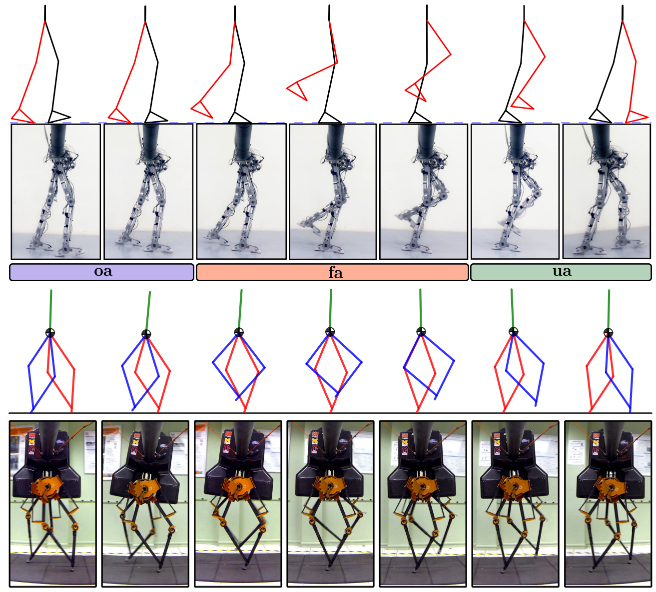


Fig. 1: Comparisons of walking tiles of simulated and experimental walking of AMBER2 (top) and ATRIAS (bottom).

III. Results

The formal results are then implemented on two different robot platforms: AMBER2 (**A**&**M** **B**ipedal **E**xperimental **R**obot 2) with the goal to achieve human-like multi-contact locomotion and ATRIAS (**A**ssume **T**he **R**obot **I**s **A** **S**phere) with the goal of emulating the Spring Loaded Inverted Pendulum (SLIP) multi-contact locomotion. Utilizing a procedure motivated by the formal constructions, we are able to successfully achieve robust human-like multi-contact walking on AMBER2 [1, 2] and SLIP-like multi-contact walking on ATRIAS [3, 4].

Acknowledgment

We would like to thank to Professor Johnathan Hurst and his team for a fruitful collaboration on ATRIAS.

References

[1] Walking of AMBER2: http://youtu.be/VvkIdCKlL54.
[2] Huihua Zhao, Wenlong Ma, and Michael B. Zeagler. Human-inspired multi-contact locomotion with amber2. In *5th ACM Intl. Conf. on ICCPS*, Berlin, 2014.
[3] Walking of ATRIAS 2.1: http://youtu.be/yiEbWwC-sR0.
[4] A. Hereid, S. Kolathaya, M. Jones, J. Van Why, J. Hurst, and A. Ames. Dynamic multi-domain bipedal walking with atrias through slip based human-inspired control. In *17th ACM Intl. Conf. on HSCC*, Berlin, April 2014.

Virtual Constraint Control of a Powered Prosthetic Leg: Experiments with Transfemoral Amputees

Robert D. Gregg, University of Texas at Dallas, rgregg@ieee.org

***Abstract*—Powered prosthetic legs could significantly improve mobility for lower-limb amputees, but the performance and clinical viability of these devices are currently limited by complex control systems that independently control different joints and time periods of the gait cycle. Recent efforts have been made to address these challenges with a unifying control model used by recent bipedal robots, in which virtual constraints define joint patterns as functions of a monotonic variable that continuously represents the gait cycle phase. This talk reports initial results of virtual constraint control of a powered prosthetic leg with three transfemoral amputee subjects walking overground and at variable cadences on a treadmill.**

Keywords: powered prosthetics; control; virtual constraints

I. Introduction

With the addition of sensors and motors, powered prosthetic legs must continuously make control decisions throughout the gait cycle [4], thus increasing the complexity of these devices. This complexity is currently handled by discretizing the gait cycle into multiple discrete "phases," each having its own separate control model—sometimes with more than a dozen control parameters per joint. Multiple tasks (e.g., walking, standing, and stair climbing) add up to hundreds of parameters for a multi-joint prosthetic leg, presenting a critical challenge to the clinical viability of these high-tech devices. In an initial effort to address these challenges, this talk presents recent experimental results with a novel nonlinear control method that continuously parameterizes the gait cycle with a mechanical phase variable, by which the prosthesis matches the human body's progression through the cycle [1, 2].

II. Methods and Results

This autonomous control method was inspired by recent breakthroughs in walking robots, which can walk, run, and climb stairs by "virtually" enforcing kinematic constraints that define desired joint patterns as functions of a mechanical variable [5]. We modeled two virtual constraints using a popular concept in the prosthetics field known as *effective shape*—the trajectory of the center of pressure (COP) mapped into a leg-based reference frame [3]. These virtual constraints depend on the COP as a phase variable, which moves monotonically from heel to toe during steady gait. We designed and implemented a control system to enforce these virtual constraints on the powered above-knee Vanderbilt leg (designed in [4]).

Recent experiments with three above-knee amputee subjects at the Rehabilitation Institute of Chicago will be presented. All three subjects achieved stable walking overground (Fig. II) and on a treadmill at variable cadences, demonstrating the clinical viability of this novel control approach.

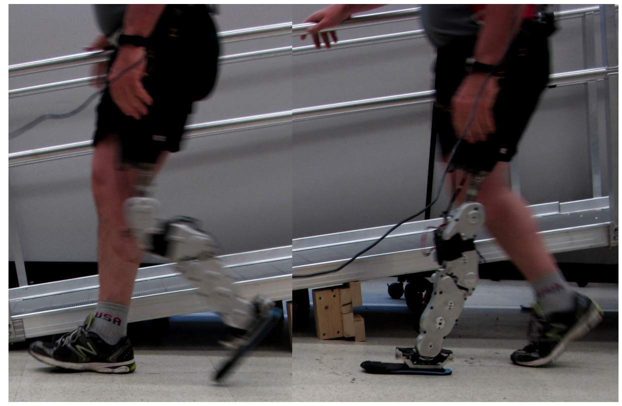


Fig. 1. Images of transfemoral amputee subject walking on the Vanderbilt prosthetic leg (designed in [4]) using virtual constraints.

Acknowledgments

Tommaso Lenzi, Levi Hargrove, and Jonathon Sensinger were co-investigators in this research. This work was supported by USAMRAA grant W81XWH-09-2-0020. Robert D. Gregg, Ph.D. holds a Career Award at the Scientific Interface from the Burroughs Wellcome Fund. This research was also supported by the National Institutes of Health under Award Number DP2HD080349. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References

[1] R. D. Gregg and J. W. Sensinger. Towards biomimetic virtual constraint control of a powered prosthetic leg. *IEEE Trans. Control Sys. Tech.*, 22(1):246–254, 2014.

[2] R. D. Gregg, T. Lenzi, N. P. Fey, L. J. Hargrove, and J. W. Sensinger. Experimental effective shape control of a powered transfemoral prosthesis. In *IEEE Int. Conf. Rehab. Robotics*, Seattle, WA, 2013.

[3] A.H. Hansen and D.S. Childress. Investigations of roll-over shape: Implications for design, alignment, and evaluation of ankle-foot prostheses and orthoses. *Disability & Rehab.*, 32(26):2201–2209, 2010.

[4] F. Sup, H.A. Varol, and M. Goldfarb. Upslope walking with a powered knee and ankle prosthesis: Initial results with an amputee subject. *IEEE Trans. Neural Sys. & Rehab. Eng.*, 19(1):71–78, 2011.

[5] E. Westervelt, J. Grizzle, C. Chevallereau, J. Choi, and B. Morris. *Feedback Control of Dynamic Bipedal Robot Locomotion*. CRC Press, New York, NY, 2007.

Gait Design Using Self–Manipulation

Aaron M. Johnson, University of Pennsylvania, Philadelphia, PA, USA

***Abstract*—Design insight into steady–state locomotion requires careful analysis of the dynamics within and transitions among the combinatorially many possible contact conditions that a legged robot will experience. To that end the self–manipulation formalism, originally intended to aid in the analysis of dynamical transitions, leans on the well established manipulation literature to systematically populate these continuous and discrete dynamics and apply a consistent structure that has proven to be useful in the analysis of grasping tasks. These ideas are instantiated on RHex with a pronking gait in which all legs are used together. Here a new behavior is presented that is both significantly faster than prior pronking gaits and also inherently stable allowing for open-loop operation. This new, stably robust pronk will enable dynamic transitions such leaping from running in any stride without requiring a decision two strides ahead of the leap.**

Keywords–Legged locomotion; robotics; motion analysis

I. Introduction

This talk applies legged self–manipulation analysis [4] (which leverages the tools of manipulation [7] for the similar, but not identical, problem of locomotion) to the design of a new pronking gait for the hexapedal robot, RHex [8]. Pronking, or stotting, has been hypothesized in animals to have many different functions from predator detection to play [3]. For legged robots it can be an efficient running gait [6] as well as a convenient way to prepare [2] leaping or other behaviors that use all of the limbs in concert [5]. On RHex there have been several studies of pronking both on the robot and in simulation [6, 1], with some success at moderate speeds using a variety of controllers to stabilize the pitch of the robot.

II. Gait Design

Self–manipulation analysis [4] provides three levels of insight into the design of this new pronking gait. First, at the behavioral level, this analysis reveals the importance of a nominally positive pitch and the inertial effects leg recirculation. Second, at the level of controller design, the self–manipulation framework underscores the great affordance given by stubbing the toes. Finally, at the level of platform design, self–manipulation analysis demonstrates that rolling contact of the legs enables a higher forward speed bound, as compared to a point toe. These ideas were previously used to design better leaping behaviors [5], but here they further provide the necessary insight to stabilize steady-state pronking.

III. Results

Using these insights, an open–loop stable pronk is shown in Fig. 1, where the legs of the robot are slightly splayed on average resulting in a net upward pitch. It appears that there is no "level" pitched limit–cycle (stable or unstable) for this system, and requiring a level pitch would overconstrain the task. The robot ran at up to 1.68m/s on outdoors on bricks (about 73% faster than [6]), with 26% duty factor (i.e. the robot is in the air 74% of the time). Indoors on smooth tiles, as shown in Fig. 1, the resulting behavior was a little slower at about 1.25 m/s. In these trials it appears that the rear leg begins to slow down before it reaches the end of stance, suggesting that some slight toe stubbing is aiding the pitch stability in an open–loop manor. These results are preliminary and additional tuning and testing is required to fully characterize the efficiency, stability, and robustness of the behavior.

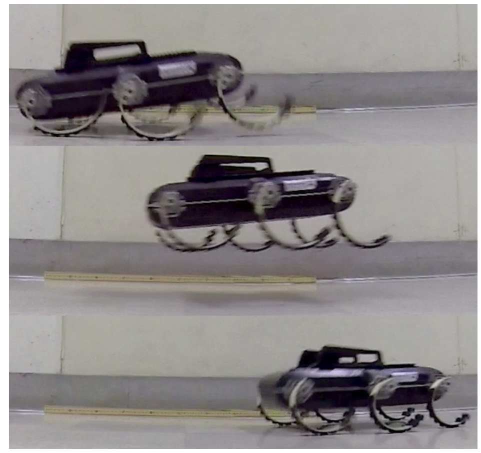


Fig. 1: RHex pronking on smooth tiles. Frames taken .2s apart.

References

[1] M.M. Ankaralı and U. Saranlı. Control of underactuated planar pronking through an embedded spring-mass hopper template. *Autonomous Robots*, 30(2):217–231, 2011.

[2] R. Burridge, A. Rizzi, and D. Koditschek. Sequential composition of dynamically dexterous robot behaviors. *Intl. J. of Robotics Research*, 18(6):534–555, 1999.

[3] T. M. Caro. The functions of stotting: a review of the hypotheses. *Animal Behaviour*, 34(3):649–662, 1986.

[4] A. M. Johnson and D. E. Koditschek. Legged self-manipulation. *IEEE Access*, 1:310–334, May 2013.

[5] A. M. Johnson and D. E. Koditschek. Toward a vocabulary of legged leaping. In *IEEE Intl. Conference on Robotics and Automation*, pages 2553–2560, May 2013.

[6] D. McMordie. Towards pronking with a hexapod robot. Master's thesis, McGill Univ., 2002.

[7] R. M. Murray, Z. Li, and S. S. Sastry. *A Mathematical Introduction to Robotic Manipulation*. CRC Press, Boca Raton, FL, 1994.

[8] U. Saranlı, M. Buehler, and D. E. Koditschek. RHex: A Simple and Highly Mobile Hexapod Robot. *The Intl. Journal of Robotics Research*, 20(7):616–631, 2001.

Learning Locomotion Controllers via Trajectory Optimization

Sergey Levine and Pieter Abbeel
Department of Electrical Engineering and Computer Sciences, UC Berkeley
E-mail: svlevine@eecs.berkeley.edu, pabbeel@cs.berkeley.edu

I. Introduction

Policy search methods can in principle learn controllers for a wide range of locomotion tasks automatically [8, 3, 7, 9, 1]. However, these algorithms typically require a carefully engineered policy class for each locomotion task. A policy class designed for one task, such as fast running, may not be effective for learning another task, such as rough terrain traversal. Recently developed guided policy search methods can learn general-purpose policies represented by neural networks, without task-specific engineering, by using trajectory optimization to find successful task executions [4, 5, 6]. We summarize our recent results on learning locomotion controllers with guided policy search, and present a novel trajectory optimization algorithm that can be used with guided policy search to learn policies even under unknown system dynamics.

II. Guided Policy Search

Guided policy search (GPS) methods optimize the parameters θ of a policy $\pi_\theta(\mathbf{u}_t|\mathbf{x}_t)$ with respect to a cost $\ell(\mathbf{x}_t, \mathbf{u}_t)$ by using trajectory optimization to guide the policy toward good solutions. A sketch of the guided policy search method is provided in Algorithm 1. The key component of GPS is the use of samples around optimized trajectories to improve the policy. These samples serve to illustrate successful task executions, and allow the difficult temporal aspects of the control problem to be handled with trajectory optimization. A second key component is the iterative reoptimization of the trajectories with an objective that encourages low cost and agreement with the current policy π_θ. This adaptation procedure gradually forces trajectory optimization to converge to a solution that is realizable under the policy.

In Figure 1, we show some simulated locomotion controllers trained with GPS under known dynamics. These results include 3D humanoid running on uneven terrain and recovery from strong lateral pushes. The push recovery controller is trained on four different pushes to capture a variety of recovery strategies, and the learned policy can recover from pushes of 250 to 500 Newtons delivered over 100 ms at different points in the gait. All experiments used a single example demonstration to initialize the trajectory, and a simulator of the dynamics was used during trajectory optimization. Videos are available on the websites associated with these papers.[1] [2]

[1]http://graphics.stanford.edu/projects/cgpspaper/index.htm
[2]http://graphics.stanford.edu/projects/gpspaper/index.htm

Algorithm 1 Guided policy search sketch

Initialize the trajectories $\{\tau_1, \dots, \tau_M\}$ with trajectory optimization and/or examples
for iteration $k = 1$ to K **do**
  Generate samples $\mathcal{S}$ around $\{\tau_1, \dots, \tau_M\}$
  Use samples $\mathcal{S}$ to optimize θ and improve the policy
  Reoptimize $\{\tau_1, \dots, \tau_M\}$ to agree with π_θ
end for

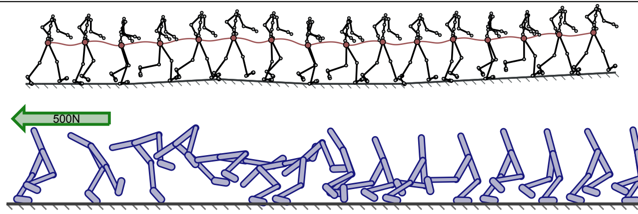


Fig. 1. Controllers trained with guided policy search for running on uneven terrain and push recovery. Adapted from [4, 6].

III. Learning with Unknown Dynamics

Using simulated dynamics models presents serious challenges, since even an accurately modeled robotic platform may respond differently from the simulation in the presence of contacts. We are currently developing a trajectory optimization method for guided policy search that does not rely on known dynamics. Similarly to differential dynamic programming [2], we use linearized dynamics and an LQR backward pass to obtain time-varying linear feedbacks. Rollouts using these feedbacks can then define a new trajectory. To work with unknown dynamics, we construct a stochastic linear Gaussian controller, which induces a distribution over trajectories. We then sample trajectories from this distribution using rollouts of the stochastic controller, use them to refit the time-varying linear dynamics model, and repeat the process.

One challenge with such local dynamics models is that the LQR update can drastically change the trajectory, leading to instability and divergence when new samples are generated from the new trajectory distribution. Inspired by work in model-free policy search [7], we bound the KL-divergence between the new and old trajectory distributions, allowing the method to make consistent, stable progress. The constrained problem can still be solved by an LQR-like method, and can be used with GPS to learn neural network policies under unknown dynamics. Preliminary results show that this method can learn simple walking controllers with 30 minutes of experience, and we are working to further improve sample efficiency.

References

[1] M. Deisenroth and C. Rasmussen. PILCO: a model-based and data-efficient approach to policy search. In *International Conference on Machine Learning (ICML)*, 2011.

[2] D. Jacobson and D. Mayne. *Differential Dynamic Programming*. Elsevier, 1970.

[3] Nate Kohl and Peter Stone. Policy gradient reinforcement learning for fast quadrupedal locomotion. In *Proceedings of the IEEE International Conference on Robotics and Automation*, 2004.

[4] S. Levine and V. Koltun. Guided policy search. In *International Conference on Machine Learning (ICML)*, 2013.

[5] S. Levine and V. Koltun. Variational policy search via trajectory optimization. In *Advances in Neural Information Processing Systems (NIPS)*, 2013.

[6] S. Levine and V. Koltun. Learning complex neural network policies with trajectory optimization. In *International Conference on Machine Learning (ICML)*, 2014.

[7] J. Peters and S. Schaal. Reinforcement learning of motor skills with policy gradients. *Neural Networks*, 21(4):682–697, 2008.

[8] R. Tedrake, T. Zhang, and H. Seung. Stochastic policy gradient reinforcement learning on a simple 3d biped. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2004.

[9] E. Theodorou, J. Buchli, and S. Schaal. Reinforcement learning of motor skills in high dimensions: a path integral approach. In *International Conference on Robotics and Automation (ICRA)*, 2010.

Whole-body Dynamic Locomotion Planning and Control for a Hydraulic Humanoid Robot

Scott Kuindersma, Hongkai Dai, Robin Deits, Andrés Valenzuela, and Russ Tedrake
Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, MA 02139
Email: {scottk,daih,rdeits,avalenzu,russt}@csail.mit.edu

***Abstract*—We summarize recent efforts of our research group toward developing algorithms for generating whole-body dynamic walking gaits from observations of (possibly complex) terrain and stabilizing their execution on physical robots. Our approach is to break the planning problem into three subproblems that we solve in sequence, each stage exploiting solutions from the previous. We are actively evaluating these algorithms using a physical Atlas humanoid robot at MIT.**

Keywords: locomotion; optimization; model-based control; legged robots

I. Introduction

Achieving dynamically-stable locomotion over irregular terrain is a central goal of legged robotics. The problem of generating dynamic walking motions can be extraordinarily complex, involving combinatoric selection of footsteps and optimization of a kinodynamically-feasible trajectory for a high-dimensional, hybrid dynamical system. We tackle this problem by subdividing it into three stages: segmentation of the terrain into a sequence of convex safe regions, optimizing footsteps and centroidal dynamics trajectories through those regions, and finally optimizing the whole-body trajectory. Stable plan execution on the robot is achieved by efficiently solving convex quadratic programs (QP) in real time.

II. Summary of Approach

A. Terrain Segmentation and Assignment

To find candidate safe regions for footsteps, we expand sampled terrain points into convex regions that are safe for foot placement using the IRIS algorithm [1]. Terrain safety criteria include avoiding foot placements on steep or uneven terrain, as well as avoiding step locations which would bring the robot's upper body into collision with the environment.

Once safe regions have been generated, footsteps locations can be planned within those regions. We search over the number of steps to take, the leading foot to choose, and the assignment of each step to one of the safe regions. This can be formulated as a mixed-integer QP on a simplified model of the robot's kinematically-reachable steps or more generally as a graph search over the safe regions combined with a nonlinear optimization to find the footstep positions and orientations, both of which are efficiently solvable.

B. Centroidal Dynamics and Footstep Planning

Given the safe polygonal terrain regions and their nominal footstep assignments, we next simultaneously searching for contact (foot) positions within these regions, ground reaction forces, and the center of mass trajectory while ensuring dynamic feasibility. The centroidal angular momentum of the robot is minimized throughout the walking motion (although incorporating more general angular momentum behaviors would be straightforward). The resulting optimization takes the form of a nonlinear program (NLP) with linear constraints. By applying sequential quadratic programming to this problem, we can guarantee that the solution to each QP in the sequence is feasible, thus creating the opportunity for early termination if hard time constraints are present.

C. Whole-body Planning

The final step is to solve a NLP to optimize a whole-body plan that is consistent with the centroidal dynamics trajectory while incorporating kinematic constraints, such as end-effector positioning and collision avoidance.

D. Stabilization

The stabilization problem can be formulated as a convex QP by exploiting the fact that the instantaneous dynamics, contact, and input constraints can be expressed linearly. A key observation about this approach is that the set of active inequality constraints typically changes infrequently between consecutive control steps, permitting the use of efficient active-set solvers for real time control [2].

III. Results

We will describe our results in achieving stable execution of walking trajectories with a 34-DoF Atlas humanoid robot.

References

[1] Robin L.H. Deits and Russ Tedrake. Computing large convex regions of obstacle-free space through semidefinite programming. *Under Review*, August 2014.

[2] Scott Kuindersma, Frank Permenter, and Russ Tedrake. An efficiently solvable quadratic program for stabilizing dynamic locomotion. In *Proceedings of the International Conference on Robotics and Automation*, May 2014.

Optimization-Based Full Body Control For The DARPA Robotics Challenge

Siyuan Feng, Eric Whitman, X Xinjilefu and Christopher G. Atkeson
Robotics Institute
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA, 15213

***Abstract*—We describe our full body humanoid control approach developed for the simulation phase of the DARPA Robotics Challenge (DRC), and the modifications made for the DRC Trials. We worked with the Boston Dynamics Atlas robot, which has 28 hydraulic actuators. Our approach was initially targeted at walking, and consisted of two levels of optimization, a high level trajectory optimizer that reasons about center of mass and swing foot trajectories, and a low level controller that tracks those trajectories by solving floating base full body inverse dynamics using quadratic programming. This controller is capable of walking on rough terrain, and also achieves long foot steps, fast walking speeds, and heel-strike and toe-off in simulation. During development of these and other whole body tasks on the physical robot, we introduced an additional optimization component in the low level controller, an inverse kinematics controller. Modeling and torque measurement errors and hardware features of the Atlas robot led us to this three part approach, which was applied to three tasks in the DRC Trials.**

I. Introduction

Originally targeted at rough terrain bipedal walking, we developed a walking control approach that can achieve a sequence of footstep targets, as well as walk fast on level ground [1]. The controller consists of two levels. The high level controller performs online trajectory optimization with a simplified model that only reasons about the center of mass (CoM) of the robot. The low level controller was originally designed to use inverse dynamics (ID) alone, and we added an inverse kinematics (IK) component to cope with modeling error when controlling the physical robot. For the DRC Trials we redesigned the high level controller to also handle ladder climbing and full body manipulation. Figure 1 shows a diagram of our approach.

II. High level controller

The high level controller is application specific. For simulated walking, we optimize the CoM trajectory given a sequence of desired foot steps. The swing foot trajectory is parametrized with quintic splines. For the DRC, we implement static walking instead due to development time constraints. For ladder climbing, the CoM motion and each limb's repositioning is encoded within a state machine. For full body manipulation, we track desired end effector location while maintaining balance.

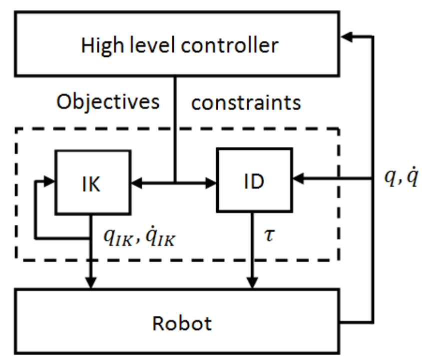


Fig. 1. The high level controller generates desired motions and constraints. Inputs to the high level controller's range from a foot step sequence to a pre-grasp pose or operator commands depending on the specific application. The low level full body controller, which is enclosed by the dashed rectangle, takes the high level objectives and robot states $(q, \dot{q})$ as inputs and outputs desired position q_{IK}, velocity $\dot{q}_{IK}$ and torque τ for each joint. Note that IK uses its own internal states rather than the measured robot states.

III. Low level controller

The low level controller generates full body motion according to the high level controller's plan while obeying various kinematic and dynamic constraints. Since the desired Cartesian motion and various constraints are all linear with respect to the unknowns, we can formulate both ID and IK as quadratic programming problems. For ID, we solve for joint and the floating base acceleration, joint torques and contact forces simultaneously. For IK, we optimize for joint and floating base velocity. Relative importance among different objectives are specified using weights. In the current implementation, IK and ID are solved independently.

IV. Results

We successfully applied the proposed controller on the Atlas robot for the terrain, ladder climbing and full body manipulation tasks during the DRC.

References

[1] S. Feng, X. Xinjilefu, W. Huang, and C. Atkeson. 3d walking based on online optimization. In *Humanoid Robots, 2013. Humanoids 2013. 8th IEEE-RAS International Conference on*, 2013.

Bipedal Locomotion as a Metastable Markov Decision Process

Cenk Oguz Saglam, Katie Byl
University of California, Santa Barbara
{saglam,katiebyl}@ece.ucsb.edu

***Abstract*—We present and explain tools we find useful for modeling and optimizing rough terrain walking. Our aim is to give an intuitive overview, toward encouraging use by others.**

I. Introduction

It is hard to predict long-term dynamics for hybrid, stochastic, nonlinear dynamic systems - such as walking robots. Given realistic noise and actuator limits, guarantees of global stability often cannot be given. Rather than abandoning all hope, we have proposed modeling very good walking as a metastable process, with reliability quantified by the mean first-passage time (MFPT) between falls [1]. Here, we briefly overview our current methods and rationale for MFPT estimation; i.e., we model step-to-step dynamics as a Markov Decision Process, and the resulting dynamics for metastable walking are typically dominated by a single (slow) discrete-time pole.

II. Meshing, Switching, and Metastability

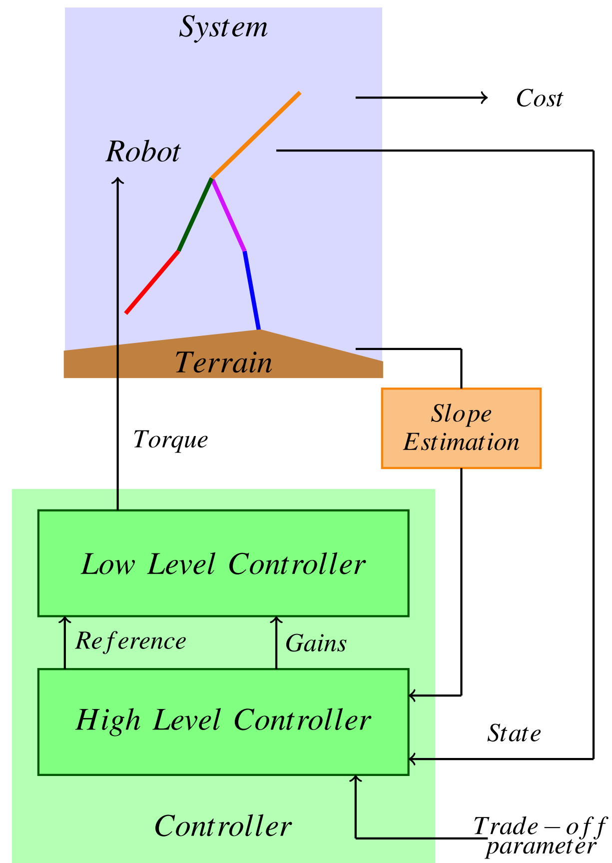
Monte Carlo simulations are not practical to estimate MFPT for metastable systems, motivating our Markov chain modeling. For our approach, this requires the existence of low-level controllers. For piecewise-sloped ground, the next state of the robot, $x[n+1]$, is a function of the current state $x[n]$, the slope ahead $\gamma[n]$, and the controller used $\zeta[n]$, i.e.,

$$x[n+1] = h(x[n], \gamma[n], \zeta[n]). \quad (1)$$

To mesh, we define a Poincaré section immediately before each foot impact. A limit cycle in the full dynamics is a fixed point in this section. Starting with a fixed point and an absorbing failure state (i.e., two points) as our initial mesh, we simulate for each controller/slope combination and add resulting states if a distance metric to all existing nodes is large enough, repeating until no new point is added, i.e., until the mesh covers the reachable part of the state space. Using the mesh, any given control policy that switches on a step-to-step basis between our controllers can be represented as a stochastic transition matrix, and for metastable systems

$$MFPT = \frac{1}{1-\lambda_2}, \quad (2)$$

where λ_2 is the second-largest eigenvalue of this matrix. Given the pre-calculated dynamics on the mesh, we can find optimal policies (based on robot state and/or partial terrain info) and can quantify effects of (for example) bad sensing, bad terrain, and changes to the available low level controllers [2, 3, 4].



Acknowledgments

This work is supported by the Institute for Collaborative Biotechnologies (ICB) through grant W911NF-09-0001 from the U.S. Army Research Office. Contents do not necessarily reflect the position or policy of the Government.

References

[1] K. Byl and R. Tedrake. Metastable walking machines. *Int J Robotics Research*, 28:1040–1064, Aug 2009.
[2] C. O. Saglam and K. Byl. Stability and gait transition of the five-link biped on stochastically rough terrain using a discrete set of sliding mode controllers. In *ICRA*, 2013.
[3] C. O. Saglam and K. Byl. Switching policies for metastable walking. In *CDC*, 2013.
[4] C. O. Saglam and K. Byl. Robust policies via meshing for metastable rough terrain walking. In *RSS*, 2014.

Gait planning and control of walking robots based on energy regulation between steps

Pranav A. Bhounsule
Disney Research Pittsburgh, 4720 Forbes Avenue,
Lower Level, Suite 110, Pittsburgh, PA 15213
pab47@disneyresearch.com

***Abstract*—We show how a simple control scheme that manages energy between key points in the gait cycle using foot placement and ankle push-off can be used to generate walking motions over a wide combinations of walking speeds and step lengths (gait planning) and can be made robust to external disturbances (control).**

I. INTRODUCTION

Our quest is to develop to develop energy-efficient, versatile and robust bipedal robot locomotion using simple models and simple control schemes. We present some progress in this direction.

II. METHODS

Our model is similar to the model used by Kuo [3] and Srinivasan and Ruina [4]. The model consists of mass-less legs and a point mass at the hip. A hip actuator can place the leg instantaneously fast at the desired location. A linear actuator on the legs can generate an impulsive push-off along the leg.

A typical step of the model and is shown in Fig. 1. The walker starts in the upright or mid-stance position in (I). Next, just before heel-strike in (II), the stance leg applies an impulsive push-off P_i and the hip actuator positions the swing leg at an angle $2\theta_i$. Next, after heel-strike in (III), the swing leg becomes the new stance leg. Finally, the walker ends up in the upright position or mid-stance position on the next step in (IV). Note that model moves passively from (I) to (II) and from (III) to (IV) (the cost of moving the mass-less swing leg is zero). The push-off impulse at (II) and the heel-strike impulse at (III) (heel-strike impulse is not shown) serves to re-direct the point mass from one circular arc to the next one. Note that the angular velocity is always perpendicular to the stance leg at all instances of time.

The mid-stance velocity at step $i+1$ ($\dot{\theta}^m_{i+1}$) can be expressed as a function of the mid-stance velocity at step i ($\dot{\theta}^m_i$) and the two controls P_i and θ_i,

$$\dot{\theta}^m_{i+1} = f(\dot{\theta}^m_i, \theta_i, P_i) \tag{1}$$

The key idea behind gait planning and control is to *measure* the mid-stance velocity at the current step, $\dot{\theta}^m_i$, and use a combination of the two *controls*, namely foot placement θ_i and push-off impulse P_i to *regulate* the mid-stance velocity at the next step, $\dot{\theta}^m_{i+1}$ (see Eqn. 1).

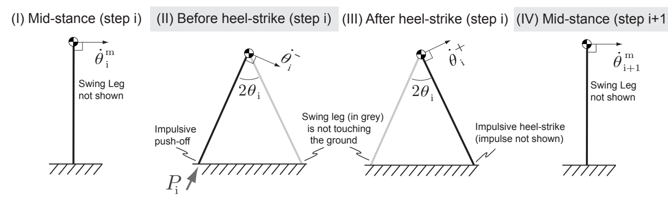


Fig. 1. **A typical step of our point mass model.**

III. RESULTS

We show that by regulating the robot kinetic energy between steps using Eqn. 1, we are able to: 1) generate a wide range of walking speeds and stride lengths, including average human walking; 2) cancel the effect of external disturbance fully in a single step (dead-beat control); and, 3) switch from one periodic gait to another in a single step. More details are in reference [2].

We have also used a similar control scheme to control our legged robot, Cornell Ranger, leading to a 40 mile non-stop walking record [1].

REFERENCES

[1] PA Bhounsule, J Cortell, A Grewal, B Hendriksen, JGD Karrsen, C Paul, and A Ruina. Low-bandwidth reflex-based control for lower power walking: 65 km on a single battery charge. *Intl. Journal of Robotics Research*, 2014 (in press).

[2] Pranav A. Bhounsule. Control of a compass gait walker based on energy regulation using ankle push-off and foot placement. *Robotica*, FirstView:1–11, May 2014. ISSN 1469-8668. doi: 10.1017/S0263574714000745. URL http://journals.cambridge.org/article_S0263574714000745.

[3] Arthur D Kuo. Energetics of actively powered locomotion using the simplest walking model. *Transactions-American Society of Mechanical Engineers Journal of Biomechanical Engineering*, 124(1):113–120, 2002.

[4] Manoj Srinivasan and Andy Ruina. Computer optimization of a minimal biped model discovers walking and running. *Nature*, 439(7072):72–75, 2006.

Sensorized Foot Design for Robust Locomotion: A Study Using Cheetah-Cub

Massimo Vespignani, Mostafa Ajallooeian, Peter Eckert, Alexandre Tuleu, and Auke J. Ijspeert
Biorobotics Laboratory, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland
{first.last}@epfl.ch

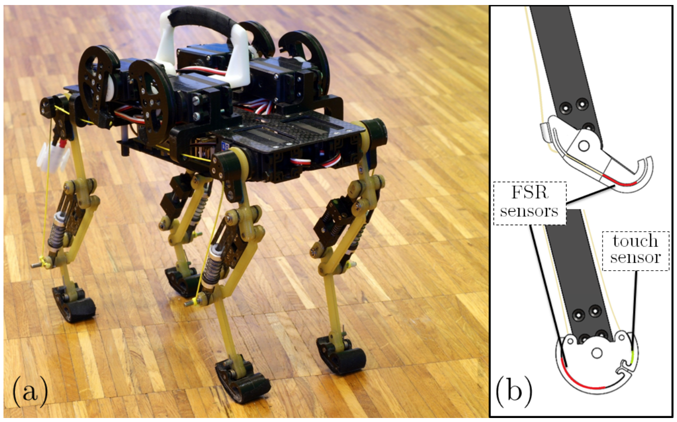


Fig. 1. a) Cheetah-Cub robot. b) Two examples of foot designs: *top*, spring-loaded foot with compliant sole and force sensor; *bottom*, ball-foot equipped with ground force sensor and stumble detector.

***Abstract*—Lightweight quadruped robots such as Cheetah-Cub need design methods and sensor equipment that do not greatly affect the overall weight and dynamics of the robot. 3d printing technology combined with off-the-shelf pressure sensors allows to quickly prototype and test different lightweight feet. We explore the application of touch sensing in modification of Central Pattern Generators with stumbling correction reflex. As a case study, we initiate gaits by height drops of about 50% of the leg-length and explore how sensory feedback can improve robustness.**

Keywords-Legged locomotion; robotics; lightweight design.

I. Design

Sensory feedback such as ground contact detection can provide the necessary information to robustly tackle unexpected external perturbation during locomotion. The addition of the necessary sensors should not limit the dynamics of the system or significantly increase the total weight. This is particularly true for lightweight robots. Here we investigate the use of 3d printing to manufacture cheap and lightweight monolithic feet with the possibility of touch sensor integration.

We developed different types of stiff and compliant feet with embedded touch sensors (Fig. 1b). Contact with the ground can be detected with force-sensing resistors, which provide a rough estimate of the ground reaction force. Stumbling detection is restricted to binary information.

II. Experiments

As experimental testing platform we use the lightweight quadruped robot Cheetah-Cub (Fig. 1a, [1]). One advantage in using this robot is that is fairly simple to obtain a stable open-loop trot gait. The modular design of the robot allows to quickly change parts of the construction to test different hypotheses.

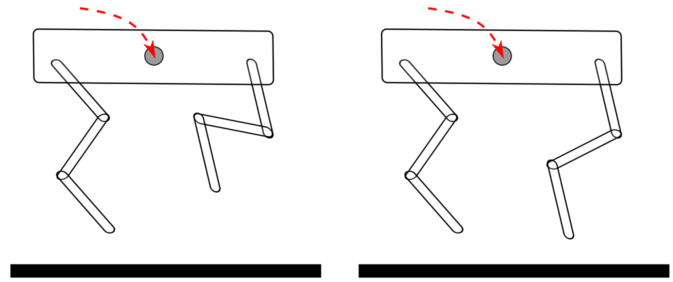


Fig. 2. Landing a height drop. *Left*, landing during CPG control. The leg pairs are asymmetrically flexed/extended. This results in an unstable landing. *Right*, with SCR control the legs will **only** flex if they have contact in the swing phase. This will prevent the aforementioned landing posture.

We compare two methods of generating locomotion profiles. The first method implements open-loop Central Pattern Generators (CPG) [2], as described in [1]. The second method is a modification of the first method where the knee CPG is replaced with a reflex-driven control. Instead of having a joint angle profile for the knee, a stumbling correction reflex (SCR) is implemented [3]. SCR is a simple reflex which flexes the leg if there is contact in the swing phase.

The second method can be useful when the robot lands after a height drop (Fig. 2). In a normal CPG control, if a leg is in stance phase and extended, the ipsilateral and contralateral legs are flexed (trot gait). This can result in an unstable landing. The case is different for the second method where all the four legs are extended as long as there is no contact. SCR initiates the continuation of locomotion after landing. An example of landing using the second method is provided in http://biorob2.epfl.ch/utils/movieplayer.php?id=276.

References

[1] A. Spröwitz et al. *Towards dynamic trot gait locomotion: Design, control, and experiments with Cheetah-cub, a compliant quadruped robot.* IJRR, 2013

[2] A.J. Ijspeert *Central pattern generators for locomotion control in animals and robots: a review.* Neural Networks, 2008

[3] M. Ajallooeian et al. *Modular control of limit cycle locomotion over unperceived rough terrain.* IROS, 2013

Learning Locomotion Controllers with a Policy Iteration Algorithm

John Schulman, Sergey Levine, Pieter Abbeel
Department of Electrical Engineering and Computer Sciences
University of California, Berkeley
Berkeley, California
Email: {joschu,svlevine,pabbeel}@eecs.berkeley.edu

***Abstract*—Reinforcement learning and policy search methods can in principle solve a wide range of control tasks automatically. However, practical robotic applications of policy search typically require a carefully designed policy representation that is specific to each task [2]. For high dimensional robotics tasks where value function estimation is impractical, policy gradient methods usually achieve the best results [6]. However, these methods assume that the policy return is a smooth function of the parameters. Since locomotion is inherently a hybrid task that combines both smooth and discontinuous dynamics [3], such methods are difficult to apply to locomotion without a carefully engineered policy parameterization that subsumes the nonsmooth aspects of the problem.**

We present a policy optimization method that can effectively handle discontinuous dynamics and learn locomotion controllers represented by general-purpose function approximators such as neural networks. Our method combines policy iteration with a trust region method that limits the change in the policy at each iteration. The algorithm can be shown to monotonically improve the policy and converge to a local optimum without assumptions about the smoothness of the objective landscape. Preliminary empirical results suggest that it can learn effective planar swimming, hopping, and walking gaits in simulation. Since the algorithm produces a neural network that directly maps system state to joint torques, it is well suited for real-time control and does not require task-specific policy classes or features.

I. Policy Iteration for Locomotion

General-purpose controller optimization has previously been addressed with direct policy search methods [2] and approximate dynamic programming techniques, such as policy iteration [1]. In direct policy search, samples from the current controller are used to estimate the gradient of the return with respect to the controller parameters. While such methods have achieved impressive results on real robotic systems, they assume that the return is a smooth function of the parameters, and typically require a compact, low-dimensional policy representation [2]. Approximate dynamic programming does not assume smoothness, but such methods are difficult to apply to high-dimensional, continuous dynamical systems.

A common limitation of approximate dynamic programming methods is the need to maintain an estimate of the value function. This is problematic, because the value function is often nonsmooth and has a large dynamic range, requiring a complex function approximator. Furthermore, a value function that has low Bellman error is not necessarily accurate [5], and a good RMS fit to the value function does not necessarily result in a good controller [9].

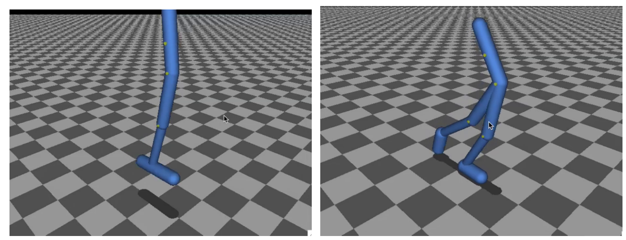


Fig. 1. Neural network controllers for hopping and bipedal walking (in 2D) learned using our method. Simulated using MuJoCo [8].

We avoid fitting a function approximator to the value function by using simulation rollouts to estimate the value of a few randomly chosen actions at each visited state. We then fit a parametric representation of the policy to these value function estimates. For many high-dimensional problems, representing a policy is much easier than representing the value function.

Another critical component of our approach is an explicit bound on the change in the policy at each iteration, to ensure that the new policy does not visit drastically different states where we did not collect samples. This is similar to several recent policy search methods [6, 7]. The difference between these methods and ours is that the policy is fitted to samples of the value function collected for a number of different actions at each visited state, which effectively mitigates many of the problems caused by discontinuous objective landscapes, since the additional action samples serve to "preview" the outcome of taking a different action in each state.

A brief summary of our method is as follows. We use a stochastic policy, which maps each state to a probability distribution over the action space. At each iteration, our method samples states from the current policy by performing rollouts. At each state along these primary rollouts, we sample additional actions and perform branch rollouts to estimate the action's value. These branch rollouts allow us to form a loss function which locally approximates the expected returns of the policy. This loss function is then minimized subject to a constraint that the change in the action distribution compared to the previous policy is small. It can be shown that for a sufficiently small step, the performance of the policy improves with high probability, using an argument similar to the one presented by Kakade and Langford [4], without assumptions about the smoothness of the objective landscape.

Images of preliminary locomotion policies for hopping and bipedal walking for simulated 2D robots are shown in Figure 1.

REFERENCES

[1] Dimitri P Bertsekas. *Dynamic programming and optimal control*, volume 2. Athena Scientific, 2012.

[2] M. Deisenroth, G. Neumann, and J. Peters. A survey on policy search for robotics. *Foundations and Trends in Robotics*, 2(1-2): 1–142, 2013.

[3] J. W. Grizzle, C. Chevallereau, A. D. Ames, and R. W. Sinnet. 3d bipedal robotic walking: Models, feedback control, and open problems. In *Proceedings of the IFAC Symposium on Nonlinear Control Systems*, 2010.

[4] Sham Kakade and John Langford. Approximately optimal approximate reinforcement learning. In *ICML*, volume 2, pages 267–274, 2002.

[5] Remi Munos, Leemon C Baird, and Andrew W Moore. Gradient Descent Approaches to Neural-Net- Based Solutions of the Hamilton-Jacobi-Bellman Equation Gradient Descent Approaches to Neural-Net-Based Solutions of the Hamilton-Jacobi-Bellman Equation. 1999.

[6] J. Peters and S. Schaal. Reinforcement learning of motor skills with policy gradients. *Neural Networks*, 21(4):682–697, 2008.

[7] J. Peters, K. Mülling, and Y. Altün. Relative entropy policy search. In *AAAI Conference on Artificial Intelligence*, 2010.

[8] Emanuel Todorov, Tom Erez, and Yuval Tassa. Mujoco: A physics engine for model-based control. In *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, pages 5026–5033. IEEE, 2012.

[9] Mingyuan Zhong, Mikala Johnson, Yuval Tassa, Tom Erez, and Emanuel Todorov. Value function approximation and model predictive control. In *Adaptive Dynamic Programming And Reinforcement Learning (ADPRL), 2013 IEEE Symposium on*, pages 100–107. IEEE, 2013.