Learning Dynamic Arm Motions for Postural Recovery

Scott Kuindersma, Roderic Grupen, Andrew Barto
Department of Computer Science
University of Massachusetts Amherst
Email: {scottk, grupen, barto}@cs.umass.edu

***Abstract*—The biomechanics community has recently made progress toward understanding the role of rapid arm movements in human stability recovery. However, comparatively little work has been done exploring this type of control in humanoid robots. We provide a summary of recent insights into the functional contributions of arm recovery motions in humans and experimentally demonstrate advantages of this behavior on a dynamically stable mobile manipulator. Using Bayesian optimization, the robot efficiently discovers policies that reduce total energy expenditure and recovery footprint, and increase ability to stabilize after large impacts.**

I. Introduction

The successful deployment of mobile humanoid robots in dynamic environments will require solutions to many challenging hardware, perception, and control problems. One particularly challenging control problem is maintaining stability in the face of postural perturbations, e.g., caused by impacts or unpredicted terrain changes. The best solutions to this problem will exhibit a high degree of resourcefulness, exploiting many actuators and innate dynamics to achieve rapid, robust, and efficient stabilization. Indeed, a typical adult human exhibits a remarkable ability to generate whole-body recovery strategies. These strategies frequently involve rapid arm movements that occur simultaneously with activation of lower limb musculature [1], [2]. Biomechanics researchers have recently made significant progress toward understanding the functional contributions of these movements under different experimental conditions [3], [4], [5], [6]. However, relatively little work has focused on developing a computational understanding of this behavior by producing controlled arm responses in artificial systems.

This paper provides an overview of previous research on upper body recovery motions and highlights the practical and scientific value of developing solutions to this problem. Toward the latter goal, we present experimental results involving a dynamically balancing mobile manipulator that learns rapid open-loop arm responses to impact perturbations by performing an efficient policy search using Bayesian optimization [7], [8], [9]. The resulting policies exhibit decreased total energy expenditure, decreased recovery footprint, and an increased ability to stabilize after large impacts.

Section II provides a review of previous investigations into arm recovery motions in both humans and artificial systems. Sections III and IV give a detailed account of two learning experiments performed on a humanoid robot. Finally, in Section V we discuss the possible implications of our results and directions for future work.

II. Background

A. Arm Recovery Motions in Humans

McIlroy and Maki [1] were perhaps the first to specifically consider arm responses to external disturbances. In this study, subjects stood upon a platform that delivered translational perturbations while shoulder and lower leg muscle responses were measured. They observed that the magnitude of the shoulder response was correlated with the magnitude and direction of the perturbation. Furthermore, the authors concluded that these movements are unlikely to be startle responses because no apparent habituation was present over multiple trials. Together, these observations suggested a possible *functional role* of arm movements in the recovery behavior.

Researchers have since begun to uncover more about the functional contributions of the upper extremities during balance recovery. Marigold et al. [2] observed rapid elevation of the arms during slip recovery in young adults. The authors noted a marked change in responses after repeated exposure to the same perturbation, suggesting that whole-body recovery strategies can be short-term adaptive. Troy et al. [5] observed a similar rapid elevation behavior in slipping experiments performed on both young and old adults. Using a simplified sagittal plane model, the authors concluded that arm responses served to reduce trunk rotational velocity immediately following the slip while repositioning the upper body center of mass away from the rear support boundary.

Similar arm response characteristics have been observed for tripping perturbations [4], [6] and hip disturbances [10], [3]. Misiaszek and Krauss [3] observed that recovery responses of leg musculature *increased in magnitude* when arm motions were voluntarily suppressed. Several studies have demonstrated significant differences between the responses of young and old subjects [4], [5], [11]. Generally speaking, younger adults who were capable of faster movements and reduced reaction times tended to produce fast movements that affected the body angular momentum, while older subjects tended to resort to more protective strategies such as grasping and bracing.

Perhaps the most complete functional analysis to date is from Pijnapples et al. [6]. Using a 3D physical model, the

authors analyzed the contribution of arm responses in tripping experiments by calculating what the body angular velocity *would have been* had the arms not been present between the perturbation onset and recovery step. The results of this analysis suggest that, for tripping perturbations during normal walking, arm recovery motions contribute most significantly to controlling rotation in the transverse (yaw) plane which helps position the body to successfully take a recovery step [6]. However, because tripping perturbations induce a rotation in the transverse plane toward the tripped foot which must be counteracted, it is possible that similar analyses for a different perturbation modality would produce different results.

B. Arm Recovery Motions in Artificial Systems

There is a very rich literature devoted to robust humanoid locomotion and recovery from perturbation. However, relatively little work exists which aims to create postural stability controllers that exploit articulated upper bodies, especially in the context of rapid balance recovery. That is not to say this field has not enjoyed much success. Indeed, for the case of bipedal postural stability, the coordination of ankle, hip, and stepping recovery strategies has yielded impressive results on real systems (e.g. [12]). However, given our increasing understanding of human balance recovery, there is reason to suspect that coordination of the arms may offer significant advantages.

Several researchers have studied model systems that have provided valuable insights. Pratt et al. [13] introduced the Linear Inverted Pendulum Plus Flywheel model that abstractly models the angular momentum induced by upper body motions as a flywheel about the body center of mass. Atkeson and Stephens [14] used a multi-link pendulum model to show that different impact recovery strategies can arise from a single quadratic optimization criterion, suggesting that whole-body responses in humans may similarly be the product of a unified control scheme. A recent paper from Nakada et al. [15] described an increase in balance recovery of a simulated biped using a learned arm rotation strategy. Other related work has considered quasi-static contributions of free arm movements in real systems [16], [17].

In the character animation literature, several researchers have produced controllers for generating whole-body recovery responses. Kudoh et al. [18], [19] formulated a quadratic programming problem to produce arm swinging motions that stabilized the system after impacts. Shiratori et al. [20] used human motion capture data during tripping experiments to create controllers that produced human-like responses in characters that were tripped under different initial conditions. Macchietto et al. [21] described a method for directly controlling linear and angular momenta that produced realistic whole-body balance recovery strategies for standing characters. These results are among the most impressive in the literature, but it remains unclear how they will translate to robotic systems with imprecise sensors and models, constrained actuators, and lower bandwidth control.

III. Experiments

We performed two experiments to quantify the advantages of whole-body recovery strategies in a real humanoid robot with limited exposure to a perturbation stimulus. These experiments involved a dynamically balancing mobile manipulator and an apparatus designed to impart controlled impact perturbations to the upper torso of the robot. After describing the experimental hardware in Sections III-A and III-B, we outline our optimal control formulation in Section III-C and our algorithmic approach in Section III-D.

A. The uBot-5

The uBot-5 is an 11-DoF mobile manipulator developed at the University of Massachusetts Amherst. The uBot-5 has two 4-DoF arms, a rotating trunk, and two wheels in a differential drive configuration. The robot stands approximately 60 cm from the ground and has a total mass of 19 kg. The robot's torso is roughly similar to an adult human in terms of geometry and scale, but instead of legs, the uBot has two wheels attached at the hip. The robot balances by controlling its wheels using a linear-quadratic regulator (LQR) with feedback from an onboard inertial measurement unit (IMU) to stabilize around the vertical fixed point. This controller has proved to be robust throughout five years of frequent usage and it remains fixed in all of our experiments.

The robot's wheeled base permits a fast and energy efficient solution to upright stability that is achieved using well understood techniques from optimal control. This makes the uBot a unique and attractive experimental platform for this problem because it allows one to assess the influence of arm motions on the stabilized system without first solving the difficult legged recovery problem.

B. Impact Pendulum

The robot was placed in a balancing configuration with the dorsal side of its torso aligned with a 3.3 kg mass suspended from the ceiling (Figure 1). The mass was pulled away from the robot to a fixed angle and released, producing a controlled impact between the swinging mass and the robot's upper torso. This device is similar that used by Hasson et al. [22] in a human study aimed at developing predictive models for step recovery after impact perturbations. The robot was attached to the ceiling with a loose-fitting safety rig designed to prevent the robot from falling completely, while not affecting the performance of the controlled response.

Impacts were detected using the robot's onboard IMU and arm responses were initiated within approximately 50 ms. The arm initial conditions were the same for each experiment and the shoulder and elbow pitch joints were controlled using a cubic spline. The spline was parameterized by a single waypoint for each joint along with two time parameters: the time-to-waypoint and time-to-home.

Two learning experiments were performed using different impact magnitudes. We aimed to evaluate the hypotheses that the robot could learn to exploit dynamic interactions between its arms and the LQR stabilized base to:

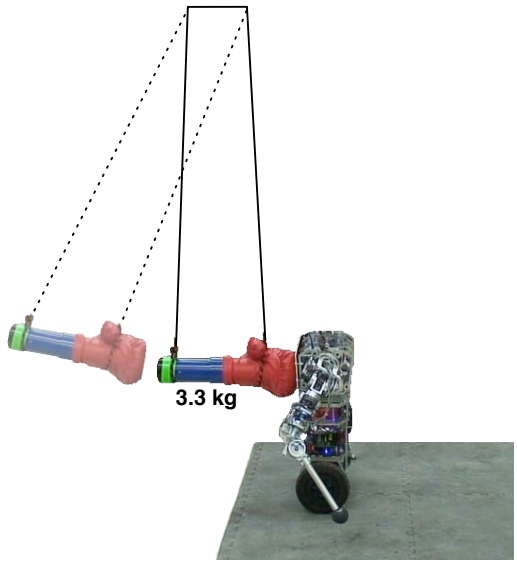

Fig. 1: The uBot-5 situated in the impact pendulum apparatus.

1) reduce the spatial footprint of the recovery,
2) reduce the total energy expenditure, and
3) increase robustness to large perturbations.

In the first experiment, the robot was situated at the base of the impact pendulum, and the release angle was chosen such that the robot could reliably recover balance using only the wheel LQR controller. The momentum of the pendulum mass prior to impact was estimated to be 5.6 N·s with a measurement error of ± 0.8 N·s by analyzing video footage of the experiment. The impact duration could not be accurately inferred from the video, but it appeared to be between 1 and 2 video frames, or $1/25$ to $2/25$ seconds. In the second experiment, the impact magnitude was increased so that a fixed arm policy would fail to stabilize the system a significant fraction of the time. The perturbation in this case was approximately 6.7 ± 1.0 N·s.

C. Optimal Control Formulation

To encourage spatially and energetically efficient solutions, we define a simple cost function:

$$J(\bar{\boldsymbol{x}}) = \int_0^T \left[x_{wheel}^2(t) + \dot{x}_{wheel}^2(t) + g(\boldsymbol{x}(t)) I(t) V \right] dt, \quad (1)$$

where $x_{wheel}(t)$ and $\dot{x}_{wheel}(t)$ are the wheel position and velocity at time t, respectively, $I(t)$ is the total absolute current being drawn by all motors, and $V = 13.1$ volts is the system voltage. The notation $\bar{\boldsymbol{x}}$ denotes a state trajectory $\{\boldsymbol{x}(0), \boldsymbol{x}(dt), \ldots, \boldsymbol{x}(T)\}$, where the state vector $\boldsymbol{x}(t)$ contains the IMU readings, a failure bit, and positions, velocities, and motor currents for all joints at time t. The function $g(\boldsymbol{x}(t))$ captures the additional energetic cost associated with a failure to recover. If $\boldsymbol{x}(t) \in FailureStates$, then $g(\boldsymbol{x}(t)) = 0.005$. Otherwise, $g(\boldsymbol{x}(t)) = 0.001$. A state $\boldsymbol{x}(t) \in FailureStates$ if and only if the state $\boldsymbol{x}(t)$ is detected as a failure or $\exists t' < t$ such that $\boldsymbol{x}(t') \in FailureStates$. Failure states were detected reliably as larges spikes in the IMU data. In all experiments, $T = 3.5$ seconds and the sampling frequency was 100 Hz.

Arm motions were constrained to be symmetric in the sagittal plane, so a single cubic spline parameterization describes the motions for both arms. The spline parameters are: $\boldsymbol{\theta} = [\theta_{\text{shoulder}}, \theta_{\text{elbow}}, t_{\text{wp}}, t_{\text{f}}]$, where θ_{shoulder} and θ_{elbow} are the shoulder and elbow waypoint positions, respectively. The remaining two time parameters describe the desired time to reach the waypoint positions and the time to return to the starting configuration. Using prior knowledge about what policies are feasible, these parameters are conservatively constrained:

$$1.5 \text{ rad} \geq \theta_{\text{shoulder}} \geq -1.5 \text{ rad} \quad (2)$$
$$1.0 \text{ rad} \geq \theta_{\text{elbow}} \geq -1.0 \text{ rad} \quad (3)$$
$$1.0 \geq t_{\text{wp}} \geq d(\theta_{\text{shoulder}}, \theta_{\text{elbow}}) \quad (4)$$
$$1.5 \geq t_{\text{f}} \geq d(\theta_{\text{shoulder}}, \theta_{\text{elbow}}) + t_{\text{wp}}, \quad (5)$$

where the function $d(\theta_{\text{shoulder}}, \theta_{\text{elbow}})$ returns the minimum time required to move to the waypoint positions given the maximum joint velocity, $5\pi/4$ rad/s.

A model of the system is not available, and the robot is only able to acquire noisy samples of (1), $\hat{J}(\boldsymbol{\theta}) \sim J(\boldsymbol{\theta}) + \epsilon$, where $\epsilon \sim \mathcal{N}(0, \sigma_n^2)$. We write $J(\boldsymbol{\theta})$ as a function of the policy parameters since we assume the initial conditions remain fixed across trials.

D. Bayesian Optimization

We employed a Bayesian optimization algorithm to optimize the policy parameters in both experiments. Our selection of this general class of algorithms is motivated by the high experimental cost associated with obtaining samples in the impact pendulum. These algorithms involve two major steps: 1) computing a posterior distribution over cost functions given all observations, and 2) selecting the next best policy parameterization to try by optimizing an acquisition criterion computed on the posterior. For an excellent tutorial on Bayesian optimization, see Brochu et al. [9].

1) Prior Representation: We represented the prior distribution over cost functions as a *Gaussian process* (GP). A GP is defined as a (possibly infinite) set of random variables, any finite subset of which is jointly Gaussian distributed [23]. To fully specify the GP, one must define a mean function and a covariance function:

$$\begin{aligned} m(\boldsymbol{\theta}) &= \mathbb{E}[J(\boldsymbol{\theta})] \\ k(\boldsymbol{\theta}_p, \boldsymbol{\theta}_q) &= \mathbb{E}[(J(\boldsymbol{\theta}_p) - m(\boldsymbol{\theta}_p))(J(\boldsymbol{\theta}_q) - m(\boldsymbol{\theta}_q))]. \end{aligned}$$

Given these functions and a set of observations, $\mathcal{D} = \{(\boldsymbol{\theta}_1, \hat{J}(\boldsymbol{\theta}_1)), \ldots, (\boldsymbol{\theta}_N, \hat{J}(\boldsymbol{\theta}_N))\}$, both the log likelihood of the data given the model and the Gaussian posterior, $P(\hat{J}(\boldsymbol{\theta}')|\boldsymbol{\theta}', \mathcal{D})$, for a point $\boldsymbol{\theta}'$ can be computed straightforwardly [23].

We use a squared exponential covariance function,

$$k(\boldsymbol{\theta}_p, \boldsymbol{\theta}_q) = \sigma_f^2 \exp(-\frac{1}{2}(\boldsymbol{\theta}_p - \boldsymbol{\theta}_q)^\top M (\boldsymbol{\theta}_p - \boldsymbol{\theta}_q)) + \sigma_n^2 \delta_{pq}, \quad (6)$$

where σ_f^2 and σ_n^2 are the signal and noise variance, respectively, $M = diag(\boldsymbol{l}^{-2})$ is a diagonal matrix of length-scales,

$\boldsymbol{l} = [l_1, l_2, l_3, l_4]$, and δ_{pq} is the Kronecker delta function. Thus, our covariance function has six hyperparameters. To avoid a potentially laborious tuning process and allow for greater flexibility, the hyperparameters were automatically optimized after each trial with respect to the maximum a posteriori (MAP) criterion. To achieve cost scale invariance, the maximum likelihood mean was computed analytically after each trial and used in the log likelihood computation [8].

A prior was placed over the logarithm of the length-scale hyperparameters: $\log(\boldsymbol{l}) \sim \mathcal{N}(\mathbf{0}, 3^2 I)$. Intuitively, the length-scales describe how much each policy parameter must be changed before a significant difference in cost is likely to be observed. Although this prior is quite broad for our application[1], it provides a flexible way to constrain the optimization process in the early stages of learning [8].

The gradients of the log likelihood and log prior terms were computed analytically and the optimization of hyperparameters was performed using the NLOPT [24] implementation of the Method of Moving Asymptotes [25]. After each trial, the hyperparameters were optimized starting from the MAP estimate from the previous trial, and 30 random restarts were performed to decrease the chance of arriving at a poor local optimum.

2) Acquisition Criterion: Given the optimized posterior distribution, we use the maximum expected improvement criterion [26] to identify the next best policy parameterization to attempt. The expected improvement (EI) given the GP model is a function of the policy parameters:

$$\mathrm{EI}(\boldsymbol{\theta}) \quad = \quad (\mu^- - \mu(\boldsymbol{\theta}) - \xi)\Phi(Z) + \sigma(\boldsymbol{\theta})\phi(Z), \qquad (7)$$

where

$$Z = \frac{\mu^- - \mu(\boldsymbol{\theta}) - \xi}{\sigma(\boldsymbol{\theta})}, \qquad \mu^- = \min_{i=1:N} \mu(\boldsymbol{\theta}_i),$$

$\mu(\boldsymbol{\theta})$ and $\sigma(\boldsymbol{\theta})$ are the mean and standard deviation of the posterior distribution over $\hat{J}(\boldsymbol{\theta})$, and $\Phi(\cdot)$ and $\phi(\cdot)$ are the CDF and PDF of the normal distribution, respectively. If $\sigma(\boldsymbol{\theta}) = 0$, the expected improvement is defined to be 0. Intuitively, (7) defines the expected reduction in cost over the best policy tried so far. The parameter ξ balances exploration and exploitation, where $\xi = 0$ leads to exploitative behavior that can leave points with high variance unexplored. In our experiments, $\xi = 0.1 \cdot \sigma_f$.

We again used NLOPT to perform the maximization of (7) while satisfying the inequality constraints on the policy parameters, (2)–(5). Forty random restarts were performed and the maximum among these was used to select the next data point.

IV. Results

We applied the Bayesian optimization algorithm in both high impact and low impact pendulum configurations. A total of 35 trials were performed in the high impact case and 30 in the low impact case. After the learning trials, a greedy policy was selected by maximizing the probability of improvement [27] with respect to the posterior distribution:

$$P(\hat{J}(\boldsymbol{\theta}) \leq \mu^-) = \Phi\left(\frac{\mu^- - \mu(\boldsymbol{\theta})}{\sigma(\boldsymbol{\theta})}\right). \qquad (8)$$

The greedy policies were $\boldsymbol{\theta}^*_{\text{low}} = [-0.681, 0.681, 0.174, 1.5]$ and $\boldsymbol{\theta}^*_{\text{high}} = [-0.562, -0.562, 0.143, 1.478]$ for the low and high impact cases, respectively. The symmetry in the shoulder and elbow displacements appears to be a consequence of the constraints (4,5) and the desire to maximize joint displacements over a short initial response time. This symmetry was not strictly observed during the learning process. Interestingly, the rotations of the shoulder and elbow joints are opposite in the low impact policy. This produces a contracted backward arm motion as opposed to the extended backward arm motion in the high impact policy. We also observed a 25% higher peak shoulder torque 0.1 seconds post-impact for the high impact policy.

To evaluate our three hypotheses regarding spatial footprint, total energy, and robustness, we performed 10 trials using the learned greedy policy and a control (fixed arm) policy for each impact magnitude. The learned policies exhibited a 17.1% reduction in average cost (1554.59 to 1288.34) in the low impact case and a 61.6% reduction in average cost (4507.36 to 1728.64) in the high impact case. The fixed arm policy failed to stabilize in 5 out of the 10 high impact trials. If we exclude the failure trials from the high impact control experiment the reduction in cost is 29.7% (2458.98 to 1728.64).

A. Efficiency Gains

We observed a statistically significant decrease in the recovery footprint using the learned arm motions for both impact magnitudes. The wheel trajectories in Figure 2 illustrate this difference. Perhaps more surprisingly, we also observed a statistically significant reduction in *total energy expenditure* when using the learned arm recovery motions. The total energy was calculated as $E = \int_0^T I(t)V dt$, where $I(t)$ is the total absolute current through all motors at time t, and $V = 13.1$ volts. Table I summarizes the reduction in average energy expenditure. Since we could not quantify the true energetic requirements of recovering from a failure, we only included the successful fixed arm trials in these statistics. Thus, the energy savings reported for the high impact case should be viewed as very conservative. These data suggest that the reduction in wheel motor energy consumption more than compensates for the additional energy consumed by the shoulder and elbow motors in the learned policies.

B. Stability Gains

During the evaluation of the learned high impact policy, the robot successfully recovered in 10 out of 10 trials. In contrast, the robot only recovered in 5 out of 10 trials with the control (fixed arm) policy. Figure 3 compares the learned response to a failure control trial. It is interesting that a single policy and impact magnitude can produce different stabilization results. Careful analysis of the experiment video

[1]Our maximum parameter range is 3 units while the prior states that there is about a 95% chance that the length-scales are between 403 and 0.002.

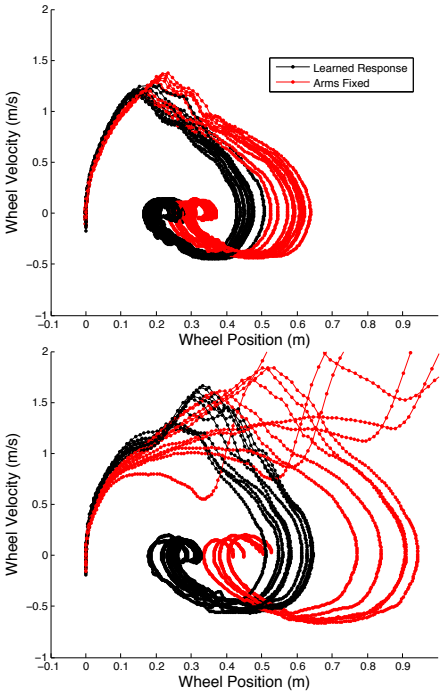

Fig. 2: Wheel position and velocity trajectories for the learned and fixed arm policies in both the low impact (above) and high impact (below) cases.

TABLE I: Comparison of mean energy expenditure averaged over 10 trials. The 5 fixed arm failure trials were excluded from the high impact data. Thus, we expect the true energetic gain in this case to be much larger than reported.

	Fixed Arms	Learned Response	Behrens-Fisher
Low impact	194.03 joules	176.37 joules	$p < 0.0001$
High impact	242.16 joules	215.67 joules	$p = 0.0046$

showed that the pendulum motion varied very little between trials. However, the state of the robot's slight back-and-forth balancing motion at the time of impact seemed to be loosely correlated with the trial outcome—though not perfectly so. Thus, the system appears to exhibit some degree of sensitivity to initial conditions. We have not yet ascertained whether the robot can distinguish between these cases. This suggests that it may be necessary to characterize the variance of each policy separately, allowing the robot to select predictable recovery strategies when the stakes are high.

The learned policies were successfully deployed during unconstrained operation of the robot. A simple filter on IMU data allowed the robot to accurately classify impacts that were similar to those seen during learning. The robot successfully responded to various uncontrolled impact perturbations: small bumps caused by a person walking into it (no arm response), pushing the robot (low impact arm response), kicking the robot (high impact response), and throwing a large exercise ball at the robot (high impact response).[2]

[2]A video is available at http://www.cs.umass.edu/~scottk/videos.

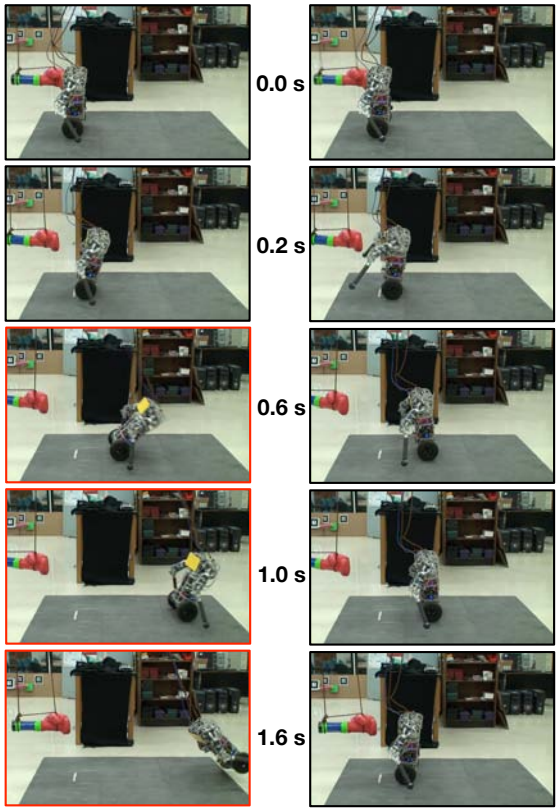

Fig. 3: Comparison of the recovery behavior without (left) and with (right) arm recovery motions after a large impact perturbation. The bottom three panels on the left outlined in red indicate the point of failure when the safety rig was engaged.

V. Discussion and Future Work

Our results suggest that the integration of arm motions in balance recovery can reduce the recovery footprint and total energy expenditure, and increase the robot's ability to stabilize after large perturbations. Although the uBot's wheeled base is very different from that of a bipedal humanoid, there is considerable practical value in being able to experimentally determine the dynamic effects of upper body responses using this simpler system. In addition to having direct practical implications for wheeled mobile manipulators [28], [29], [30], we expect the observed benefits to translate across morphologies. Indeed, our results agree with previous observations that the magnitude of human lower body recovery responses increased when arm motions were suppressed [3].

This general problem also has several attributes that make it interesting from a machine learning perspective: expensive evaluations, nonlinearity, stochasticity, and high-dimensionality. In our experiments, a low-dimensional policy space was identified, allowing the robot to apply a Bayesian optimization algorithm to discover effective policies in a small number of trials. Another benefit of this approach is that we are able to *interpret* the robot's state of knowledge. For example, by examining the MAP length-scale hyperparameters, we can

learn something about the relative sensitivity of the cost with respect to the policy parameters. The length-scales after learning in the high impact experiment suggest that the cost is most sensitive to changes in initial response time and shoulder angle, with total movement time and elbow angle having considerably lower sensitivity.

In future work, we will consider a larger range of initial conditions. For example, we expect that the arm configurations of a mobile manipulator will vary considerably under normal operation and that solving this problem will require not only the ability to generate rapid collision-free arm motions, but also apply knowledge about successful policies to avoid vulnerable configurations. We will also address the problem of sensitivity to initial conditions observed in higher energy impact situations. In cases where there is a significant chance to destabilize, efficiency concerns should fall behind the ability to predictably recover, and the robot's ability to make such judgements will depend on its ability capture the shape of the cost distribution at different points in policy space.

VI. Conclusion

Humanoid postural stability is an important and difficult control problem that has received much attention. Despite significant successes, the benefits of exploiting dynamic arm motions in balance recovery have not been fully understood. Our experiments with a humanoid robot demonstrate the ability to learn recovery policies in a small number of trials, and that coordinated arm motions can increase the efficiency and robustness of responses to impact perturbations.

Acknowledgments

The authors would like to thank Brian Umberger for several helpful discussions. Scott Kuindersma is supported by a NASA GSRP Fellowship. Roderic Grupen was supported by the ONR MURI award N00014-07-1-0749. Andrew Barto was supported by the AFOSR under grant FA9550-08-1-0418.

References

[1] W. E. McIlroy and B. E. Maki, "Early activation of arm muscles follows external perturbation of upright stance," *Neuroscience Letters*, vol. 184, no. 3, pp. 177–180, 1995.

[2] D. S. Marigold, A. J. Bethune, and A. E. Patla, "Role of the unperturbed limb and arms in the reactive recovery response to an unexpected slip during locomotion," *J Neurophysiol*, vol. 89, pp. 1727–1737, 2003.

[3] J. E. Misiaszek and E. M. Krauss, "Restricting arm use enhances compensatory reactions of leg muscles during walking," *Experimental Brain Research*, vol. 161, no. 4, pp. 474–485, 2005.

[4] P. E. Roos, M. P. McGuigan, D. G. Kerwin, and G. Trewartha, "The role of arm movement in early trip recovery in younger and older adults," *Gait & Posture*, vol. 27, pp. 352–356, 2008.

[5] K. L. Troy, S. J. Donovan, and M. D. Grabiner, "Theoretical contribution of the upper extremities to reducing trunk extension following a laboratory-induced slip," *Journal of Biomechanics*, vol. 42, pp. 1339–1344, 2009.

[6] M. Pijnappels, I. Kingma, D. Wezenberg, G. Reurink, and J. H. van Dieën, "Armed against falls: the contribution of arm movements to balance recovery after tripping," *Experimental Brain Research*, vol. 201, 2010.

[7] M. Frean and P. Boyle, "Using gaussian processes to optimize expensive functions," in *AI 2008: Advances in Artificial Intelligence*, 2008, pp. 258–267.

[8] D. J. Lizotte, "Practical bayesian optimization," Ph.D. dissertation, University of Alberta, Edmonton, Alberta, 2008.

[9] E. Brochu, V. Cora, and N. de Freitas, "A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning," University of British Columbia, Department of Computer Science, Tech. Rep. TR-2009-023, 2009.

[10] J. Misiaszek, "Early activation of arm and leg muscles following pulls to the waist during walking," *Experimental Brain Research*, vol. 151, no. 3, pp. 318–329, 2003.

[11] J. H. J. Allum, M. G. Carpenter, F. Honegger, A. L. Adkin, and B. R. Bloem, "Age-dependent variations in the directional sensitivity of balance corrections and compensatory arm movements in man," *The Journal of Physiology*, vol. 542, no. 2, pp. 643–663, 2002.

[12] B. Stephens and C. Atkeson, "Push recovery by stepping for humanoid robots with force controlled joints," in *Proceedings of the International Conference on Humanoid Robots*, Nashville, TN, 2010.

[13] J. Pratt, J. Carff, S. Drakunov, and A. Goswami, "Capture point: A step toward humanoid push recovery," in *Proceedings of the IEEE-RAS International Conference on Humanoid Robots*, 2006, pp. 200–207.

[14] C. G. Atkeson and B. Stephens, "Multiple balance strategies from one optimization criterion," in *Proceedings of the IEEE-RAS International Conference on Humanoid Robots*, December 2007, pp. 57–64.

[15] M. Nakada, B. F. Allen, S. Morishima, and D. Terzopoulos, "Learning arm motion strategies to recover balance in bipedal robots," in *Intl Symposium on Learning and Adaptive Behavior in Robotic Systems*, Canterbury, UK, September 2010.

[16] E. Yoshida and J.-P. Laumond, "Motion planning for humanoid robots: Highlights with HRP-2," *Journées Nationales de la Recherche en Robotique*, pp. 9–12, October 2007.

[17] S. Kuindersma, "Control model learning for whole-body mobile manipulation," in *Proceedings of the Twenty-Fourth Conference on Artificial Intelligence (AAAI-10)*, Atlanta, GA, July 2010, pp. 1939–1940.

[18] S. Kudoh, T. Komura, and K. Ikeuchi, "The dynamic postural adjustment with the quadratic programming method," in *International Conference on Intelligent Robots and Systems (IROS)*, October 2002, pp. 2563–2568.

[19] ——, "Stepping motion for a human-like character to maintain balance against large perturbations," in *Proceedings of the International Conference on Robotics and Automation (ICRA)*, Orlando, FL, May 2006.

[20] T. Shiratori, B. Coley, R. Cham, and J. K. Hodgins, "Simulating balance recovery responses to trips based on biomechanical principles," in *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, Aug. 2009.

[21] A. Macchietto, V. Zordan, and C. R. Shelton, "Momentum control for balance," in *Transactions on Graphics/ACM SIGGRAPH*, 2009.

[22] C. J. Hasson, R. E. V. Emmerik, and G. E. Caldwell, "Predicting dynamic postural instability using center of mass time-to-contact information," *Journal of Biomechanics*, vol. 41, no. 10, pp. 2121–2129, 2008.

[23] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[24] S. G. Johnson, "The NLopt nonlinear-optimization package," http://ab-initio.mit.edu/nlopt.

[25] K. Svanberg, "A class of globally convergent optimization methods based on conservative convex separable approximations," *SIAM J. Optim*, vol. 12, no. 2, pp. 555–573, 2002.

[26] J. Močkus, V. Tiesis, and A. Žilinskas, "The application of bayesian methods for seeking the extremum," in *Toward Global Optimization*. Elsevier, 1978, vol. 2, pp. 117–128.

[27] H. J. Kushner, "A new method of locating the maximum of an arbitrary multipeak curve in the presence of noise," *J. Basic Engineering*, vol. 86, pp. 97–106, 1964.

[28] R. O. Ambrose, R. T. Savely, S. M. Goza, P. Strawser, M. A. Diftler, I. Spain, and N. Radford, "Mobile manipulation using NASA's robonaut," in *Proceedings of the International Conference on Robotics and Automation (ICRA)*, 2004, pp. 2104–2109.

[29] M. Stilman, J. Olson, and W. Gloss, "Golem krang: Dynamically stable humanoid robot for mobile manipulation," in *Proceedings of the International Conference on Robotics and Automation (ICRA)*, 2010.

[30] A. J. McClung, III, Y. Zheng, and J. B. Morrell, "Contact feature extraction on a balancing manipulation platform," in *Proceedings of the International Conference on Robotics and Automation (ICRA)*, Anchorage, Alaska, May 2010.