Continuous Feedback Control Using Needle Variations for Nonlinear and Hybrid Systems

Giorgos Mamakoukas Aleksandra Kalinowska Malcolm A. MacIver Todd D. Murphey

***Abstract*— This paper proves that the mode insertion gradient (MIG), traditionally used in hybrid systems for optimal mode scheduling, provides a descent direction over the entire time horizon. This result allows us to construct a first-order Needle Variation Controller (NVC) for general nonlinear systems that synthesizes continuous control responses from the MIG, maintaining the benefits of needle variation methods, such as high computational efficiency and a large region of attraction. The proposed algorithm is a generalization of all possible first-order algorithms and, for specific parameters, it is shown to match the conjugate gradient descent algorithm. As a second result of this paper, the proposed control solution guarantees descent even when subject to arbitrary scaling (that can differ both across different inputs and over time) as long as it maintains the sign of each input. The benefits of NVC are showcased in simulations of a three-link biped that demonstrates successful gait initiation, walking, and disturbance tolerance.**

I. Introduction

Robotic tasks commonly involve nonlinear, hybrid, underactuated, and high-dimensional systems that pose challenges for real-time controllers. Although various approaches have been developed, including nonlinear model-based predictive controllers (NMPC) [1], dynamic programming (DP) [2], backstepping [3], sliding mode control (SMC) [4], as well as intelligent [5] and hybrid [6] control, these methods face limitations. NMPC and DP are computationally expensive; backstepping is generally ineffective in the presence of control limits; SMC methods suffer from chattering, which results in high energy consumption and instability risks by virtue of exciting unmodeled high-frequency dynamics [7]; intelligent control methods are subject to data uncertainties [8], while other approaches are often case-specific and do not hold for the level of generality encountered in robotics.

Single-action controllers constitute a distinct approach to optimal control [9]. Based on needle variation methods, these schemes rely on the mode insertion gradient—the sensitivity of the cost function to infinitesimal application of controls—to synthesize single, constant magnitude actions of infinitesimal duration that reduce the objective [10], [11].

Compared to other methods, single-action needle variation schemes exhibit certain advantages. They are shown to exist globally, demonstrate a larger region of attraction, exhibit robustness to unmodeled dynamics, are less sensitive to local minima, and have a simpler Lie Group representation [9], [12]–[14]. Furthermore, when considering the second-order mode insertion gradient, they have provable properties of descent for controllable systems, a condition that is not provided by other nonlinear controllers [15]. Lastly, control solutions of these methods have a closed-form, analytical expression and readily apply saturation limits on the control solutions. As a consequence, single-action needle variation schemes obtain orders of magnitude improvements in computational speed over controllers such as iLQG [16] and, thus, seem well-suited for online robotic tasks. On the other hand, they perform poorly near equilibrium, they often converge slowly because they apply only a short action in each iteration, and their feedback is discontinuous and, therefore, challenging to implement in hardware.

This paper constructs a control algorithm that bridges the gap between traditional trajectory optimization algorithms and single-action feedback schemes. By proving that the mode insertion gradient provides a descent direction for the entire time horizon, we are able to construct a continuous feedback synthesis scheme, referred to as Needle Variation Controller (NVC), out of the single-action solution. We claim that the proposed algorithm is a generalization of all first-order control solutions and, under certain parameter choices that are specified in Section IV, it is shown to match the solution of conjugate gradient descent. We further prove that the suggested control solution remains a direction under arbitrary scaling that addresses a broader set of cases than what has been shown for the conjugate gradient descent algorithm [17], [18].

The proposed controller readily adapts to hybrid systems, as visible in the simulation results of the bipedal walker in Section V. Without using gait specific parameters, NVC performs gait initialization, establishes a quasi-periodic gait cycle, and is robust to force disturbances. It therefore appears promising for the well-studied field of bipedal walking, where currently available control methods lack versatility—they require offline optimization [19], [20], involve tuning of walking-specific parameters, such as touch-down and take-off angles [21], and are sensitive to outside disturbances.

The paper is structured as follows: Section II presents an overview of single-action, needle variation controllers in optimal control. Section III proves that the mode insertion gradient provides a descent direction for smooth optimal control problems, even in the presence of saturation limits, and proposes a continuous needle variation feedback scheme. Section IV relates NVC to the gradient descent algorithm

The authors are with the Department of Mechanical Engineering (Mamakoukas, Kalinowska, MacIver and Murphey), Biomedical Engineering (MacIver), and Neurobiology (MacIver), Northwestern University, Evanston, IL, 60208 USA. Emails: {giorgosmamakoukas, ola}@u.northwestern.edu, {maciver,t-murphey}@northwestern.edu

TABLE I: Notation

Symbol	Description
v	default control
w	control deviation
$D_s f(\cdot)$	partial derivative $\frac{\partial f(\cdot)}{\partial s}$
$\|\cdot\|_A$	norm with metric A (e.g. $\|x\|_A = x^T A x$)

and proves that the two policies are identical for certain parameter choices. Section V includes simulation results that compare the proposed algorithm to single-action policies, gradient descent, and iLQG, demonstrating its robustness to initial conditions, saturation limits and the formulation of the objective. The approach is also tested on a three-link biped. Conclusive remarks and directions for further work are given in Section VI.

II. Needle Variation Methods in Optimal Control

In this section, we provide an overview of single-action needle variation methods in optimal control. We conclude with a conceptual comparison to alternative optimization schemes that highlights the benefits of needle variation schemes in robotic applications.

In hybrid systems literature, the mode insertion gradient (MIG) is the first-order sensitivity of the objective to an infinitesimal perturbation of the default control. Feedback controllers that use needle variation methods, such as Sequential Action Control [9], construct control responses that correspond to a negative MIG. The resulting solution is itself a function of time and returns the optimal infinitesimal input perturbation at an application time.

We show that the time history can be considered an optimal perturbation to the default control. This is an alternative implementation that updates all the control values over the time horizon, similar to linear quadratic methods, but with less computation.

A. Single-Action Needle Variation Controller

Consider systems with state $x : \mathbb{R} \mapsto \mathbb{R}^N$ and control $u : \mathbb{R} \mapsto \mathbb{R}^M$. Single-action policies are needle variation methods developed for control-affine dynamics of the form

$$\dot{x}(t) = f(x(t), u(t), t) = g(x(t)) + h(x(t), t)u(t), \quad (1)$$

where $g(x(t))$ is the drift vector field. These feedback schemes consider objectives of the form

$$J = \int_{t_o}^{t_f} L(x(t), t)\,\mathrm{d}t + m(x(t_f)) \quad (2)$$

that have sometimes been control-independent. Let $v(t)$ be the default control and $w(t)$ be the control perturbation vector, such that

$$u(t) = v(t) + w(t). \quad (3)$$

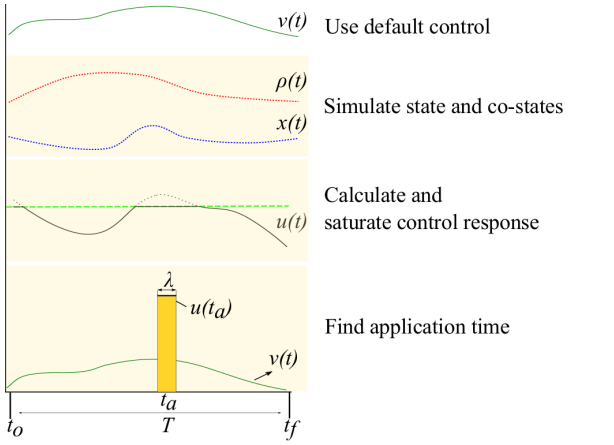


Fig. 1: The algorithmic steps of single-action controllers. Using default control, the state and costate variables are forward simulated in time. The optimal control response is computed from a closed-form analytical expression and saturated. In the last step, the application time of a single inserted action is chosen to correspond to a negative MIG.

Needle variation methods consider a time period $[t_o, t_f]$ and switched control modes described by

$$\dot{x}(t) = \begin{cases} f_1, & t_o \le t < t_a - \frac{\lambda}{2} \\ f_2, & t_a - \frac{\lambda}{2} \le t < t_a + \frac{\lambda}{2} \\ f_1, & t_a + \frac{\lambda}{2} \le t \le t_f, \end{cases} \quad (4)$$

where f_1 and f_2 are dynamics associated with *default* and *inserted* control $v(t)$ and $u(t)$, respectively, and defined as

$$f_1 \triangleq f(x(t), v(t))$$
$$f_2 \triangleq f(x(t), u(t_a)).$$

Parameters t_a and λ describe the switching time between the two modes and the (infinitesimal) duration of the inserted dynamics f_2.

Note that, while the default control $v(t)$ of the switched mode sequence described in (4) may be time-dependent, dynamics f_2 consider a time-*invariant* control $u(t_a)$, which is the control response evaluated at the application time t_a. Dynamics of the form (4) appear in optimal control of hybrid systems to optimize the time scheduling of *a priori* known modes [22], but single-action feedback schemes use them to obtain a new control mode that will optimally perturb the trajectory of any type of system with a needle action.

Single-action policies, as introduced in [9], use the mode insertion gradient to compute a closed-form analytical expression of the control solution, determine the application time t_a, and find the control duration λ with a backtracking line search. The MIG expression has been derived for control-independent costs of the form shown in (2) and is given by

$$\frac{dJ}{d\lambda_+} = \rho^T(t)(f_2 - f_1), \quad (5)$$

where the elements of $\rho : \mathbb{R} \mapsto \mathbb{R}^N$ are the adjoint states (costates) calculated from the default trajectory via

$$\dot{\rho}(t) = -D_x f_1^T \rho(t) - D_x L(x(t), v(t), t)^T \quad (6)$$

subject to $\rho(t_f) = D_x m(x(t_f))^T$. Work in [12] uses α_d as the desired sensitivity and computes controls given by

$$\begin{aligned} u(t) &= \underset{u}{\operatorname{argmin}} \ \frac{1}{2}(\frac{dJ}{d\lambda_+} - \alpha_d)^2 + \frac{1}{2}\|u - v\|_R^2 \\ &= v(t) + \alpha_d(\Lambda + R)^{-1}h^T(t)\rho(t), \end{aligned} \tag{7}$$

where $\Lambda \triangleq h(t)^T\rho(t)\rho(t)^T h(t) \succeq 0$. The first term in (7) drives the MIG to a desired negative sensitivity (α_d), while the second term penalizes large deviations from the nominal control. Given (3), the control update becomes

$$w(t) = \alpha_d(\Lambda + R)^{-1}h^T(t)\rho(t), \tag{8}$$

where $\alpha_d = -\gamma J(v(t))$ and $\gamma \in \mathbb{R}^+$. Although α_d changes across iterations, it remains constant throughout $[t_o, t_f]$ of any individual control update. A visual overview of single-action controllers is presented in Fig. 1. For more details, the reader can refer to [9], [12].

B. Comparison to Traditional Optimal Control Policies

Several methods in optimal control minimize a local approximation of the cost function by iteratively searching in high-dimensional spaces, a computationally expensive process. On the other hand, needle variation methods aim to *reduce*, not minimize, the objective. They avoid the approximation of the value function, and instead exploit the time-evolving sensitivity of the objective to infinitesimal switched dynamics to optimally perturb the default trajectory. As a result, needle variation methods are less computationally expensive than methods such as iLQR and other algorithms that expand the cost function to second order [23], [24].

While first-order needle variation controls compute the evolution of the state and first-order costate equations via two $n \times 1$ differential equations in each iteration, the iLQR algorithm solves the Riccati equations to calculate a descent direction and, together with the simulation of the state, it computes in total three $n \times 1$ and one $n \times n$ differential equations.

Needle variation methods exist globally and exhibit robustness (e.g., to unmodeled fluid flow [14]), demonstrate a large region of attraction [9], [13], and have a less complicated representation on Lie Groups [13]. Further, using information from the second-order mode insertion gradient, they have formal guarantees of descent for certain controllable systems, not provided by other schemes [15].

On the other hand, needle variation methods based on the mode insertion gradient have so far been implemented only as single-action feedback. As a result, control responses are discontinuous without filtering realizable by hardware. Additionally, single-action controllers exhibit poor performance near equilibrium and often switch to iLQR policies to ensure stability [9], [12].

III. Continuous feedback based on needle variations

In this section, we show that the mode insertion gradient, and by extension the solution of single-action controllers, is a descent direction for the control over the entire time horizon. As a result, one can apply the control curve and avoid the sparse implementation of single-action feedback schemes.

A. Descent Direction

We make the following assumptions.

Assumption 1: The vector elements of dynamics f_1 and f_2 are real, bounded, $\mathcal{C}^1$ in x, and $\mathcal{C}^0$ in u and t.

Assumption 2: The incremental cost $l(\cdot)$ and the terminal cost $m(\cdot)$ are real, bounded, and $\mathcal{C}^1$ in x.

Assumption 3: Default and inserted controls v and u are real, bounded, and $\mathcal{C}^0$ in t.

We use Pontryagin's Maximum Principle to determine the local minimizer of a system.

Definition 1: A trajectory that is the local minimizer is given by a pair (x^*, u^*) if and only if the pair satisfies the optimality conditions posed by Pontryagin's Maximum principle:

$$\frac{\partial H}{\partial \rho} = \dot{x}(t) \quad \frac{\partial H}{\partial x} = -\dot{\rho}(t) \quad \frac{\partial H}{\partial u} = 0, \tag{9}$$

where H, the Hamiltonian function, is defined as

$$H \triangleq L(x(t), u(t), t) + \rho^T(t) f(x(t), u(t), t).$$

Proposition 1: Consider systems with state x and control u and an objective given by (2). Then, the control policy given by (8) is a descent direction for all $t \in [t_o, t_f]$. Further, if there exists $t \in [t_o, t_f]$ for which the MIG is negative, then the feedback policy in (8) will decrease the cost. Moreover, this feedback policy converges to the local minimizer trajectory, as stated in Definition 1.

Proof: Using a first-order Taylor expansion, we write

$$J(v(t) + w(t)) \approx J(v(t)) + \frac{\partial J}{\partial u(t)}\Big|_{v(t)} \cdot w(t). \tag{10}$$

For objectives of the form in (2), we use the Gâteaux derivative to calculate the gradient of the cost with respect to the control

$$\frac{d}{d\epsilon} J(v(t) + \epsilon w(t))\Big|_{\epsilon=0} = \int_{t_o}^{t_f} \rho(t)^T h(t) \cdot w(t) \, \mathrm{d}t. \tag{11}$$

From (10) and (11), the first-order change in cost can be approximated with

$$\Delta J \approx \int_{t_o}^{t_f} \rho^T(t) h(t) \cdot w(t) \mathrm{d}t. \tag{12}$$

Equivalently, using (5), we can write

$$\Delta J \approx \int_{t_o}^{t_f} \frac{dJ}{d\lambda_+} \mathrm{d}t,$$

which, with the update of (8) and given that $\alpha_d < 0$, becomes

$$\Delta J \approx \int_{t_o}^{t_f} \alpha_d \|\rho^T(t) h(t)\|^2_{(\Lambda(t)+R)^{-1}} \mathrm{d}t \leq 0.$$

We examine the case of the equality, such that

$$\begin{aligned} \Delta J = 0 &\Leftrightarrow \rho^T(t) h(t) = 0 \ \forall \ t \in [t_o, t_f] \\ &\Leftrightarrow \frac{dJ}{d\lambda_+} = 0 \ \forall \ t \in [t_o, t_f]. \end{aligned}$$

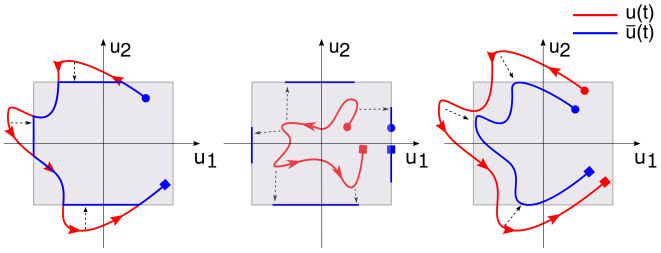


Fig. 2: Cases of control scaling that remain valid descent directions for the proposed needle-variation controller. The left figure shows control clipping, where values are saturated at a specific threshold; the middle figure shows arbitrary stretching to the saturation limits; the right figure shows proportional scaling that maintains the same direction of the applied control. Control curves are a function of time and arbitrarily shown for a 2-input system for easier visualization. Simulation results in this paper use control clipping.

If there exists $t \in [t_o, t_f]$ for which $dJ/d\lambda_+ < 0$, then $\Delta J < 0$. The change in cost, to first order, will be zero if and only if $\rho^T(t)h(t) = 0$ for all $t \in [t_o, t_f]$, which is the condition for a minimum according to Pontryagin's Maximum Principle (for objectives of the form in (2)). ■

Next, we prove that one can scale the descent direction in an arbitrary way, differently across the time-horizon, and still obtain a descent direction, provided that the direction of the update is maintained.

Proposition 2: Consider systems with state x and control u and an objective metric given by (2). Let $\Gamma(t) \succ 0$ be a diagonal matrix. The control update $w(t)$ given by (8) remains a descent direction for the entire trajectory even when scaled to $w_s(t) = \Gamma(t)w(t)$.

Proof: Consider the update policy in (8) with scaled control $\bar{u}(t)$. Let $w_s(t)$ indicate the perturbation after scaling, such that $\bar{u}(t) = v(t) + w_s(t)$, where $w_s(t) = \Gamma(t)w(t)$, where the diagonal elements of $\Gamma(t)$, $\gamma_i(t) \geq 0$, can be different from each other.

From Proposition 1, the cost change, approximated to first-order, is then

$$\Delta J \approx \int_{t_o}^{t_f} \alpha_d \|\rho^T(t)h(t)\|^2_{\Gamma(t)(\Lambda(t)+R)^{-1}} \mathrm{d}t. \tag{13}$$

Given that $\alpha_d < 0$, $\Gamma(t) \succ 0$, and $\Lambda(t) + R \succ 0$, (13) is negative if there exists time t in $[t_o, t_f]$ such that $\rho^T(t)h(t) \neq 0 \iff w_s(t) \neq 0$. Hence,

$$\Delta J < 0$$
$$\iff \exists\ t \in [t_o, t_f] \text{ such that } w_s(t) = \Gamma(t)w(t) \neq 0.$$

Therefore, given the update (8), the cost—approximated to first-order—is guaranteed to decrease provided that there exists $t \in [t_o, t_f]$ such that $\bar{u}(t) \neq v(t)$. ■

We note that Proposition 2 holds even when control inputs are scaled differently in time. Fig. 2 shows valid cases of control distortion that remain a descent direction.

Compared to other first-order continuous policies and to the best of the authors' knowledge, the scalability results of the control inputs shown in 2 guarantee descent over a broader set of cases of control distortion. For example, work in [17] considers only particular cases of our scalability results shown in Figure 2, that of control clipping. Their method suggests dividing, based on a guess, the time horizon into regions of saturated and unsaturated control inputs and updating actuation only for the latter regions. Work in [18] provides a control clipping and proofs that the modified control update remains a descent direction, however it also considers only a special case of the results shown here. In particular, we are not aware of a method that considers the arbitrary scaling of the inputs to the saturation limits, which would be a desired traits for applications capable only of discrete actuation.

IV. Connection to Steepest Descent

Steepest (or gradient) descent algorithm [25] is arguably the simplest and most popular first-order trajectory optimization algorithm for general nonlinear optimal control problems. The update policy is given by

$$w(t) = -Rv(t) - h^T(t)\rho(t). \tag{14}$$

Conjugate gradient descent [26], a first-order technique that provably outperforms steepest descent in terms of convergence at comparable computational cost, is only the conjugate variant of steepest descent. The only difference is the conjugate update, which modifies the control update in (14) to ensure successive controls are conjugate with respect to one another. In that sense, all algorithms can have their conjugate variant. For the purposes of comparison, we relate the proposed algorithm to the gradient descent policy (14); if the solutions match, then the conjugate variants of both methods will also be identical. To that extent, by showing that NVC can, under certain parameter choices, match steepest descent, the reader should also assume that NVC can also, under the same choices and by using a conjugate update, also match the conjugate gradient descent algorithm.

In general, the solution of NVC in (8) is not equivalent to the gradient descent.

$$\begin{aligned}\alpha_d(\Lambda + R)^{-1}(h^T\rho) &= -Rv - h^T\rho \\ \alpha_d h^T\rho &= -(\Lambda + R)Rv - (\Lambda + R)h^T\rho.\end{aligned}$$

To show the above relationship is not always true, we can consider the case $h^T\rho = 0$. Then,

$$0 = -RRv,$$

which is only satisfied if $v = 0$. One could, however, refer to a similar form of (8), derived in [9], specifically

$$\begin{aligned}u(t) &= \underset{u}{\operatorname{argmin}}\ \frac{1}{2}\left(\frac{dJ}{d\lambda_+} - \alpha_d\right)^2 + \frac{1}{2}\|u\|^2_R \\ &= v(t) - (\Lambda + R)^{-1}(Rv(t) - \alpha_d h^T\rho(t)),\end{aligned}$$

such that

$$w(t) = -(\Lambda + R)^{-1}(Rv(t) - \alpha_d h^T\rho(t)) \tag{15}$$

The difference with (7) is that (15) does not penalize the norm of the control update $w = u - v$, but rather the magnitude of the control solution u.

We prove that, under a specific choice of the parameter α_d, that is always feasible, and for $R = I$, the update policy in (15) matches the solution of the gradient descent.

Proposition 3: Consider control affine dynamics (1). If $R = I$ and $\alpha_d = -(\rho^T h R u_o + \rho^T h h^T \rho + 1)$, then the update policies of (15) and (14) match.

Proof: Setting the control updates of the proposed algorithm and the Gradient Descent, shown in (15) and (14) respectively, equal to each other,

$$\begin{aligned}(\Lambda + R)^{-1}(Rv - \alpha_d h^T \rho) =& Rv + h^T \rho \\ Rv - \alpha_d h^T \rho =& (\Lambda + R)(Rv + h^T \rho),\end{aligned}$$

where $\Lambda = h^T \rho \rho^T h \succeq 0$ and $\Lambda + R \succ 0$. We consider two cases: i) $h^T \rho = 0$ and ii) $h^T \rho \neq 0$. If $h^T \rho = 0$, then also $\Lambda = 0$ and the above equation becomes

$$\begin{aligned}Rv =& R(Rv) \\ (I - R)Rv =& 0 \\ R =& I.\end{aligned}$$

If $h^T \rho \neq 0$, and using $R = I$,

$$\begin{aligned}v - \alpha_d \underbrace{h^T \rho}_{\boldsymbol{p}} =& (\Lambda + I)(v + h^T \rho) \\ 0 =& \boldsymbol{p}\boldsymbol{p}^T v + v + \boldsymbol{p}\boldsymbol{p}^T \boldsymbol{p} + \boldsymbol{p} - v + \alpha_d \boldsymbol{p} \\ 0 =& \boldsymbol{p}(\boldsymbol{p}^T v + \boldsymbol{p}^T \boldsymbol{p} + 1 + \alpha_d) \\ \alpha_d =& -(\boldsymbol{p}^T v + \boldsymbol{p}^T \boldsymbol{p} + 1) \\ \alpha_d =& -(\rho^T h v + \rho^T h h^T \rho + 1) \end{aligned} \tag{16}$$

■

It is worth noting that α_d, as given by (16), is not necessarily negative. Similarly, it can easily be shown that the control solution in (15) would not always generate a descent direction, that is a negative , as shown in (12).

Next, we prove that for α_d given by (16), the policies shown in (8) and (14) always match. Although this may seem trivial, a solution may not always exist. For example, consider two vectors x and z. Even though a scalar $c = \frac{z^T x}{\|z\|}$ algebraically satisfies the equation $x = cz$, such a solution is infeasible if the vectors are not parallel to each other.

Proposition 4: Consider control affine dynamics (1). If $R = I$ and $\alpha_d = -(\rho^T h R u_o + \rho^T h h^T \rho + 1)$, then the control policies of (8) and (14) match.

Proof: Setting the update policies of the proposed algorithm and Gradient Descent equal to each other,

$$\begin{aligned}v - \alpha_d \boldsymbol{p} =& \Lambda v + v + \Lambda \boldsymbol{p} + \boldsymbol{p} \\ \Lambda(v + \boldsymbol{p}) + \boldsymbol{p} + \alpha_d \boldsymbol{p} =& 0,\end{aligned} \tag{17}$$

where

$$\Lambda = \boldsymbol{p}\boldsymbol{p}^T = \begin{bmatrix} p_1^2 & \cdots & p_1 p_n \\ & \ddots & \\ \vdots & p_i p_j & \vdots \\ & & \ddots \\ p_n p_1 & \cdots & p_n^2 \end{bmatrix},$$

such that

$$\begin{bmatrix} p_1^2 & \cdots & p_1 p_n \\ & \ddots & \\ \vdots & p_i p_j & \vdots \\ & & \ddots \\ p_n p_1 & \cdots & p_n^2 \end{bmatrix} \begin{bmatrix} v_1 + p_1 \\ v_2 + p_2 \\ \vdots \\ v_n + p_n \end{bmatrix} + \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix} + \alpha_d \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix} = 0.$$

In the k^{th} row (where $k \in [1, n]$),

$$\begin{aligned}p_k \sum_{i=1}^{n} p_i(v_i + p_i) + p_k + \alpha_d p_k = 0 \\ \alpha_d p_k = -p_k(\sum_{i=1}^{n} p_i(v_i + p_i) + 1).\end{aligned}$$

If $p_k = 0$, the above equation is always satisfied (for all $k \in [1, n]$), regardless of α_d. If $p_k \neq 0$, we can divide both sides by p_k such that

$$\alpha_d = -\sum_{i=1}^{n} p_i(v_i + p_i) - 1,$$

which is independent of k and, thus, the same equation holds for all $k \in [1, n]$. Thus, n equations collapse to 1, the same as the number of unknowns (α_d). Therefore, (17) always holds and the policies of NVC and Gradient Descent match each other. ■

V. Example Systems

In this section, we use the benchmark cart pendulum system to test NVC and compare it both to trajectory optimization and single-action algorithms. We also use a 3-link biped to demonstrate the performance of NVC on a more challenging system and its robustness against disturbances. The algorithmic steps are shown in Algorithm I and the parameters used in the simulations are reported in Table II.

A. Cart Pendulum

To compare performance across algorithms we use the cart pendulum system, which has been a popular testbed for conventional controllers [6], [27]–[29]. Dynamics are used as in [30] and a state vector $x = [x_c, \dot{x}_c, \theta, \dot{\theta}]$ is utilized.

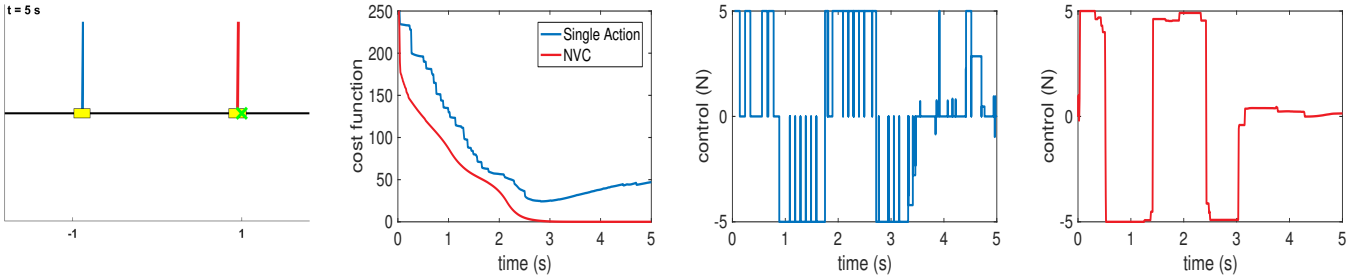
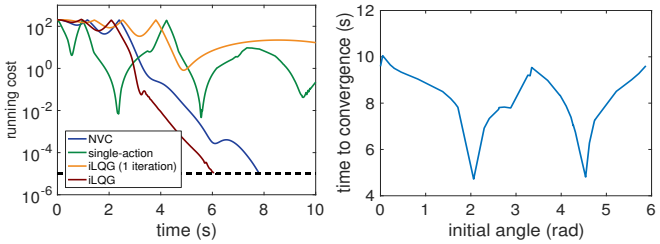


Fig. 3: Performance of a single-action needle variation controllers and NVC on the inversion task of a cart pendulum. NVC successfully inverts the pendulum with saturation limits of 5 N, while single-action controls are reported to require at least 15 N for the four-state system [9]. Given the time, NVC can invert the pendulum with as low as 1N saturation limits. We note that the NVC solution appears similar to a low-pass filtered single-action solution.



(a) (b)

Fig. 4: Fig. 4a compares NVC, single-action feedback, and two implementations of iLQG—one limited to one forward and backward pass per iteration and one with unlimited passes per iteration. The latter provides the optimizer, while the former can be thought of as providing a good enough solution. Fig. 4b shows a Monte Carlo simulation over the initial angle θ_0. NVC successfully leads to inversion in 50 out of 50 trials with convergence times ranging from 4.7 to 10.1 seconds. For the same set of parameters and range of initial conditions, single-action policies do not converge for any of the trials [12].

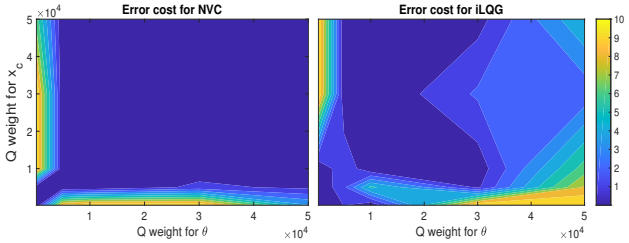


Fig. 5: Comparison of NVC and iLQG performance over a range of cost function formulations. Terminal cost at the end of a 15-second simulation was interpolated over the range of Q sampled. For all simulations Q was defined as a diagonal matrix with weights $= 0$ for all states except θ and x_c, for which weights are shown on the x- and y-axis, respectively. Note the robustness of NVC for a range of task definitions.

Algorithm I Needle Variation Controller

Initialize $k \in (0, 1)$.

Simulate $x(t)$ and $\rho(t)$ in $[t_o, t_o + T]$ from f_1
Compute $J(v(t))$
Compute $w(t)$ given by (8)
Compute $J(u(t))$ and ΔJ
$\beta = 0$
while $\Delta J > \Delta J_{min}$ **do**
    $k = k^{\beta}$
    Update $u(t) = v(t) + kw(t)$
    Saturate $u(t)$
    Compute $J(u(t))$
    $\beta = \beta + 1$
end while

Algorithm 1: $\Delta J = J(v(t)) - J(u(t))$

Fig. 3 illustrates a comparison in performance between single-action and continuous-action needle variation controllers. Saturation limits of 5 N are imposed. These limits are the same as in [9], where they are unable to assure convergence for all four states, and four times lower than those used in [12] to invert the pendulum with single-action control. We further compare NVC to the iLQG controller

TABLE II: Parameter values for single-action control, iLQG, and NVC

	Cart Pendulum	Biped
x	$[x_c, \dot{x}_c, \theta, \dot{\theta}]$	$[\theta_s, \theta_m, \theta_t, x_{hip}, y_{hip}, \dot{\theta}_s, \dot{\theta}_m, \dot{\theta}_t, \dot{x}_{hip}, \dot{y}_{hip}]$
Q	$[2, 0, 3, 0] \cdot 10^4$	$[10, 350, 350, 0, 0, 0, 0, 0, 0, 0]$
R	$[0.2]$	$[0.1, 0.1]$
T	2.2 s	1.0 s
t_f	15 s	10 s
t_s	0.005 s	0.01 s
u_{max}	$\pm[20]$	$\pm[30, 30]$

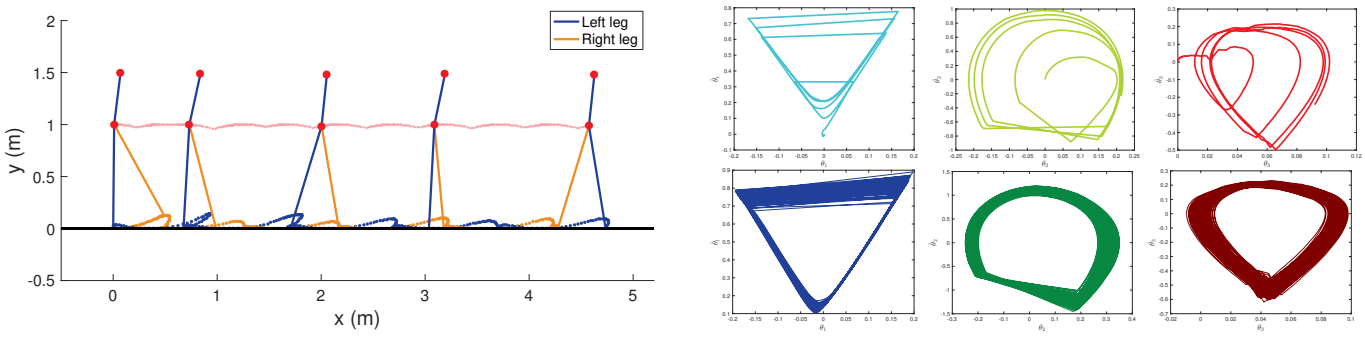
with results shown in Fig. 4a.

We demonstrate the robustness of NVC through a Monte Carlo simulation over the initial pendulum angle θ_0 as well as over the relative state weights in Q and R matrices that define the quadratic objective. Fig. 4b demonstrates that NVC converges in all 50 trials randomly sampled over $[0, 2\pi]$. Fig. 5 shows that NVC is robust to changes in the Q and R matrices within almost two orders of magnitude, while iLQG[1] does not converge as reliably.

B. 3-link Biped

We use a 3-link model that consists of a torso and two identical legs without knees or ankles; all links are rigid. The behavior of the system is described by the swing phase

[1]We implement iLQG using the code that is available at http://www.mathworks.com/matlabcentral/fileexchange/52069-ilqg-ddp-trajectory-optimization and which is described in [31].



(a) (b)

Fig. 6: Forward locomotion of a planar biped using a continuous-action needle variation controller. We observe that the NVC-controlled biped is able to perform gait initialization and establish a quasi-periodic gait pattern. Fig. 6a shows the walking pattern of the biped. Fig. 6b shows phase plots of the legs and torso angles; the upper plots correspond to the gait initiation (first five seconds); the lower plots correspond to the simulation from $t = 10$ s until the end of the five-minute walk. The imperfect gait cycle can allude to a certain level of robustness with respect to deviations from a nominal gait phase.

dynamics, derived using the method of Lagrange [32] and the impact update on the states [33], [34]. To handle impulses that occur at impact, we modify the adjoint variable using reset maps, similar to [9]. Specifically, we simulate the adjoint variables using the same differential equation as (6) and utilize a reset map at time of impact to account for state jumps, as outlined in Equation (40) in [9].

Parameters of the dynamic model can be found in [34]. The biped states are described by $x = [\theta_s, \theta_m, \theta_t, x_{hip}, y_{hip}, \dot{\theta}_s, \dot{\theta}_m, \dot{\theta}_t, \dot{x}_{hip}, \dot{y}_{hip}]$, where θ_s, θ_m, and θ_{hip} are the angles of the support leg, moving leg, and torso from the vertical, respectively; x_{hip} and y_{hip} are the Cartesian coordinates of the hip. As legs alternate between support and swing phase, θ_s and θ_m swap upon impact. We illustrate the performance of NVC for this biped system in Fig. 6. The target states are given by $x_d = [\pi/8, \pi/8, \pi/18, 0, 0, 1, 1, 0, 0, 0]$ and provide the system with a command of driving the swing leg forward and maintaining a slightly forward-leaning posture of the torso that is commonly observed in everyday walking.

As such, NVC enables the generation of a dynamic gait rather than tracking pre-computed reference trajectories [35]–[38]. As seen in Fig. 6, NVC creates a quasi-periodic gait without gait-specific constraints, such as touch-down or take-off angles used by other controllers [21], [39]. What is more, we do not initiate walking from favorable configurations, contrary to other approaches [40], [41] that either ignore gait initiation or treat it separately. Lastly, we use the full dynamic model of the biped [42], while other online controllers take advantage of simplified dynamics or linear model approximations [43]–[45].

Fig. 7 demonstrates the algorithm's robustness against disturbances, which are introduced in the form of discontinuities in the angular velocity of the torso. Fig. 8 introduces forces on the torso for 0.02 s at various points throughout a gait cycle. The force disturbances that the biped can recover from range between -32 N and 18 N depending on the biped's position in the cycle. This range, when scaled by the size of our biped (35 kg compared to 57^+ kg), corresponds to the tolerable range reported in literature [46], [47] for different controllers for kneed bipeds.

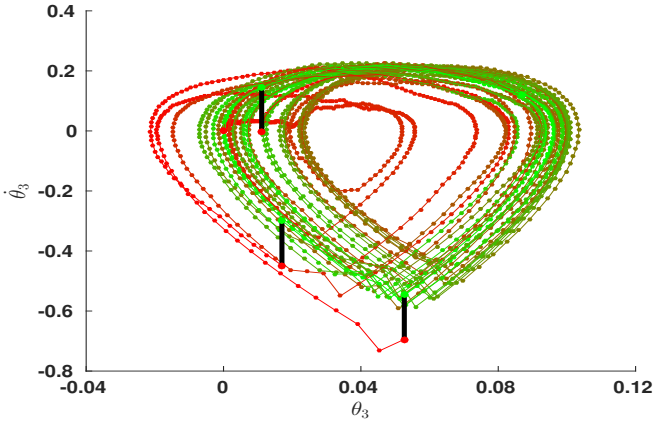


Fig. 7: Gait recovery from disturbances on the torso angular velocity. The figure shows the phase plane related with the torso angle given disturbances of -0.15 rad/s applied every 8 seconds over a 30-second simulation. The introduced disturbance discontinuities are marked with a solid black line. The trajectory in-between disturbances changes linearly with time from red to green color. As a result, the green trajectory is the converging gait cycle and the red indicates deviations present immediately after the disturbances.

VI. CONCLUSIONS

This study presents the Needle Variation Controller (NVC), a nonlinear feedback method. The algorithm synthesizes its control response in ways similar to the needle variation controllers shown in [9] and [15], but differs from these single-action policies in that it utilizes the entire time horizon of control values at each iteration. The resulting feedback of the proposed scheme guarantees descent, even in the presence of actuation limits. NVC maintains the desirable traits of existing needle variation methods, namely a large region of attraction, a closed-form expression, and computational efficiency.

Compared to popular feedback schemes, such as iLQG or DDP, the proposed method uses only first-order information and is computationally faster, which makes it suitable for online applications. Further, as the example of the biped model shows, the proposed scheme lends itself to hybrid system applications without additional overhead associated

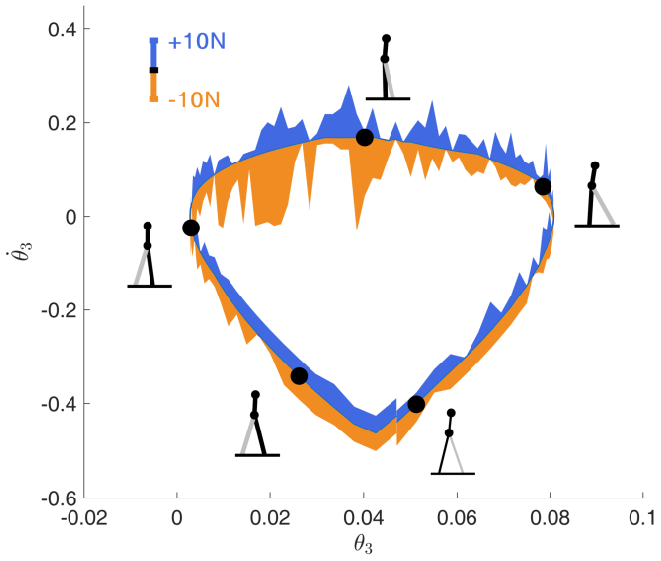


Fig. 8: Gait recovery from external forces applied on the torso for 0.02 s at various times of a full gait cycle. The shaded regions indicate the maximum magnitude of positive (blue) and negative (orange) forces tolerated by the biped without losing balance. The legend bar in the figure indicates the magnitude scale of the forces.

with switched dynamics and impacts. The controller design is not problem-specific, which makes NVC an attractive controller for versatile robotic applications.

Finally, the robustness of the controller can become especially meaningful when dealing with human-in-the-loop approaches, where safety is of utmost importance. It is also critical in applications such as rehabilitation equipment, exoskeletons, and other assistive devices that require low failure rates in the presence of unanticipated system inputs or changes in the environment.

References

[1] F. Allgöwer, R. Findeisen, and Z. K. Nagy, "Nonlinear model predictive control: From theory to application," *Journal of the Chinese Institute of Chemical Engineers*, vol. 35, no. 3, pp. 299–315, 2004.

[2] D. P. Bertsekas, *Nonlinear programming*. Athena scientific Belmont, 1999.

[3] P. V. Kokotovic, "The joy of feedback: nonlinear and adaptive," *IEEE Control systems*, vol. 12, no. 3, pp. 7–17, 1992.

[4] V. I. Utkin, *Sliding Modes in Control and Optimization*. Springer-Verlag, 1992.

[5] C. J. Harris, C. G. Moore, and M. Brown, *Intelligent Control: Aspects of Fuzzy Logic and Neural Nets*. Singapore: World Scientific, 1993, vol. 6.

[6] R. Fierro, F. L. Lewis, and A. Lowe, "Hybrid control for a class of underactuated mechanical systems," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 29, no. 6, pp. 649–654, 1999.

[7] H. K. Khalil, *Nonlinear Systems*, 3rd ed. Upper Saddle River, NJ: Prentice-Hall, 2002.

[8] A. M. El-Nagar, M. El-Bardini, and N. M. EL-Rabaie, "Intelligent control for nonlinear inverted pendulum based on interval type-2 fuzzy PD controller," *Alexandria Engineering Journal*, vol. 53, no. 1, pp. 23–32, 2014.

[9] A. R. Ansari and T. D. Murphey, "Sequential action control: closed-form optimal control for nonlinear and nonsmooth systems," *IEEE Transactions on Robotics*, vol. 32, no. 5, pp. 1196–1214, 2016.

[10] S. Aseev and V. Veliov, "Needle variations in infinite-horizon optimal control," *Variational and Optimal Control Problems on Unbounded Domains*, vol. 619, pp. 1–17, 2014.

[11] M. S. Shaikh and P. E. Caines, "On the hybrid optimal control problem: theory and algorithms," *IEEE Transactions on Automatic Control*, vol. 52, no. 9, pp. 1587–1603, 2007.

[12] E. Tzorakoleftherakis and T. Murphey, "Iterative sequential action control for stable, model-based control of nonlinear systems," *arXiv preprint arXiv:1706.08932*, 2017.

[13] T. Fan and T. Murphey, "Online feedback control for input-saturated robotic systems on Lie groups," in *Robotics: Science and Systems Conference (RSS)*, 2016.

[14] G. Mamakoukas, M. A. MacIver, and T. D. Murphey, "Sequential action control for models of underactuated underwater vehicles in a planar ideal fluid," in *American Control Conference (ACC)*. IEEE, 2016, pp. 4500–4506.

[15] ——, "Feedback synthesis for controllable underactuated systems using sequential second order actions," 2017.

[16] E. Todorov and W. Li, "A generalized iterative LQG method for locally-optimal feedback control of constrained nonlinear stochastic systems," in *Proceedings of the American Control Conference*, 2005, pp. 300–306.

[17] B. Pagurek and C. M. Woodside, "The conjugate gradient method for optimal control problems with bounded control variables," *Automatica (Journal of IFAC)*, vol. 4, no. 5-6, pp. 337–349, 1968.

[18] V. Quintana and E. Davison, "Clipping-off gradient algorithms to compute optimal controls with constrained magnitude," *International Journal of Control*, vol. 20, no. 2, pp. 243–255, 1974.

[19] T. Erez and E. Todorov, "Trajectory optimization for domains with contacts using inverse dynamics," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2012, pp. 4914–4919.

[20] C. Liu, C. G. Atkeson, and J. Su, "Biped walking control using a trajectory library," *Robotica*, vol. 31, no. 2, pp. 311–322, 2013.

[21] M. Rutschmann, B. Satzinger, M. Byl, and K. Byl, "Nonlinear model predictive control for rough-terrain robot hopping," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2012, pp. 1859–1864.

[22] M. Egerstedt, Y. Wardi, and H. Axelsson, "Transition-time optimization for switched-mode dynamical systems," *IEEE Transactions on Automatic Control*, vol. 51, no. 1, pp. 110–115, 2006.

[23] E. Theodorou and F. J. Valero-Cuevas, "Optimality in neuromuscular systems," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2010, pp. 4510–4516.

[24] J. Hauser, "A projection operator approach to the optimization of trajectory functionals," *IFAC Proceedings Volumes*, vol. 35, no. 1, pp. 377–382, 2002.

[25] H. B. Curry, "The method of steepest descent for non-linear minimization problems," *Quarterly of Applied Mathematics*, vol. 2, no. 3, pp. 258–261, 1944.

[26] L. Lasdon, S. Mitter, and A. Waren, "The conjugate gradient method for optimal control problems," *IEEE Transactions on Automatic Control*, vol. 12, no. 2, pp. 132–138, 1967.

[27] W. Sun, E. A. Theodorou, and P. Tsiotras, "Game theoretic continuous time differential dynamic programming," in *American Control Conference (ACC), 2015*. IEEE, 2015, pp. 5593–5598.

[28] ——, "Continuous-time differential dynamic programming with terminal constraints," in *Adaptive Dynamic Programming and Reinforcement Learning (ADPRL)*. IEEE, 2014, pp. 1–6.

[29] A. N. K. Nasir, M. A. Ahmad, and M. F. Rahmat, "Performance comparison between LQR and PID controllers for an inverted pendulum system," in *AIP Conference Proceedings*, vol. 1052, no. 1. AIP, 2008, pp. 124–128.

[30] E. Tzorakoleftherakis and T. D. Murphey, "Controllers as filters: Noise-driven swing-up control based on maxwell's demon," in *54th Annual Conference on Decision and Control (CDC)*, 2015, pp. 4368–4374.

[31] Y. Tassa, N. Mansard, and E. Todorov, "Control-limited differential dynamic programming," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, May 2014.

[32] M. W. Spong, S. Hutchinson, and M. Vidyasagar, *Robot modeling and control*. Wiley New York, 2006, vol. 3.

[33] F. Plestan, J. W. Grizzle, E. R. Westervelt, and G. Abba, "Stable walking of a 7-dof biped robot," *IEEE Transactions on Robotics and Automation*, vol. 19, no. 4, pp. 653–668, 2003.

[34] E. R. Westervelt, J. W. Grizzle, C. Chevallereau, J. H. Choi, and B. Morris, *Feedback control of dynamic bipedal robot locomotion*. CRC press, 2007, vol. 28.

[35] J. H. Park and K. D. Kim, "Biped robot walking using gravity-compensated inverted pendulum mode and computed torque control," in *1998 IEEE International Conference on Robotics and Automation, 1998. Proceedings.*, vol. 4. IEEE, 1998, pp. 3528–3533.

[36] S. Kajita, F. Kanehiro, K. Kaneko, K. Fujiwara, K. Harada, K. Yokoi, and H. Hirukawa, "Biped walking pattern generation by using preview control of zero-moment point," in *IEEE International Conference on Robotics and Automation, 2003. Proceedings.*, vol. 2. IEEE, 2003, pp. 1620–1626.

[37] Q. Huang, K. Kaneko, K. Yokoi, S. Kajita, T. Kotoku, N. Koyachi, H. Arai, N. Imamura, K. Komoriya, and K. Tanie, "Balance control of a biped robot combining off-line pattern with real-time modification," in *IEEE International Conference on Robotics and Automation, 2000. Proceedings.*, vol. 4. IEEE, 2000, pp. 3346–3352.

[38] P.-B. Wieber and C. Chevallereau, "Online adaptation of reference trajectories for the control of walking systems," *Robotics and Autonomous Systems*, vol. 54, no. 7, pp. 559–566, 2006.

[39] P. M. Wensing and D. E. Orin, "High-speed humanoid running through control with a 3d-slip model," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2013, pp. 5134–5140.

[40] Y. Jian, D. A. Winter, M. G. Ishac, and L. Gilchrist, "Trajectory of the body cog and cop during initiation and termination of gait," *Gait & Posture*, vol. 1, no. 1, pp. 9–22, 1993.

[41] B. G. Buss, K. A. Hamed, B. A. Griffin, and J. W. Grizzle, "Experimental results for 3d bipedal robot walking based on systematic optimization of virtual constraints," in *American Control Conference (ACC)*, 2016, pp. 4785–4792.

[42] E. Tzorakoleftherakis, "Stable control synthesis for human-in-the-loop systems," Ph.D. dissertation, Northwestern University, 2017, unpublished thesis.

[43] E. R. Westervelt, J. W. Grizzle, and D. E. Koditschek, "Hybrid zero dynamics of planar biped walkers," *IEEE Transactions on Automatic Control*, vol. 48, no. 1, pp. 42–56, 2003.

[44] D. Dimitrov, P.-B. Wieber, O. Stasse, H. J. Ferreau, and H. Diedam, "An optimized linear model predictive control solver for online walking motion generation," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2009, pp. 1171–1176.

[45] A. Herdt, H. Diedam, P.-B. Wieber, D. Dimitrov, K. Mombaur, and M. Diehl, "Online walking motion generation with automatic footstep placement," *Advanced Robotics*, vol. 24, no. 5-6, pp. 719–737, 2010.

[46] B. J. Stephens and C. G. Atkeson, "Dynamic balance force control for compliant humanoid robots," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2010, pp. 1248–1255.

[47] J. M. Wang, D. J. Fleet, and A. Hertzmann, "Optimizing walking controllers," *ACM Transactions on Graphics (TOG)*, vol. 28, no. 5, p. 168, 2009.