Euler Discretization and Inexact Restoration for Optimal Control*

C. Yalçın Kaya† J. M. Martínez‡

December 5, 2006

Abstract

A computational technique for unconstrained optimal control problems is presented. First an Euler discretization is carried out to obtain a finite-dimensional approximation of the continous-time (infinite-dimensional) problem. Then an inexact restoration (IR) method due to Birgin and Martínez is applied to the discretized problem to find an approximate solution. Convergence of the technique to a solution of the continuous-time problem is facilitated by the convergence of the IR method and the convergence of the discrete (approximate) solution as finer subdivisions are taken. It is shown that a special case of the IR method is equivalent to the projected Newton method for equality constrained quadratic optimization problems. The technique is numerically demonstrated by means of a scalar system and the van der Pol system, and comprehensive comparisons are made with the Newton and projected Newton methods.

Key words: Optimal control, inexact restoration, Euler discretization, discrete approximation, projected Newton method, Lagrange multiplier update, costate update, van der Pol system.

1 Introduction

Continuous-time optimal control problems are optimization problems in infinite-dimensional spaces. To obtain an accurate solution of these problems, typically, shooting techniques are employed (see [1, 2] for an exposition and survey of these techniques). However, shooting techniques in general require a good initial guess, and are prone to ill-conditioning which causes numerical instabilities. As a result, these techniques may fail, or take a long time to converge to a solution. One of the approaches to tackle these difficulties is to use some discretization scheme so as to obtain a finite-dimensional approximation of the problem, and then apply standard optimization techniques to get an approximate solution of the original problem. Discretized versions of the problem may eliminate ill-conditioning, but it is still important that a solution is found quickly within a desired accuracy. Once an approximate

*The authors thank the anonymous reviewers whose comments and suggestions improved the manuscript. Helmut Maurer is gratefully acknowledged for insightful e-mail discussions.

†Visiting Professor, Departamento de Sistemas e Computação, Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil; and School of Mathematics and Statistics, Senior Lecturer, University of South Australia, Mawson Lakes, S.A. 5095 Australia. The author gratefully acknowledges support through a fellowship by CAPES, Ministry of Education, Brazil (Grant No. 0138-11/04), for his visit to UFRJ between March 2004 and February 2005. E-mail: yalcin.kaya@unisa.edu.au

‡Professor, Department of Applied Mathematics, IMECC-UNICAMP, University of Campinas, CP 6065, 13081-970 Campinas SP, Brazil. The work of this author was supported by FAPESP (Grant 2001-04597-4) and CNPq. E-mail: martinez@ime.unicamp.br

solution is obtained, it can either be regarded satisfactory and accepted, or be fed into a shooting method as a better initial guess in order to obtain an accurate solution.

There are three important issues in this kind of discretization approach; namely, (i) selection of the discretization scheme, (ii) convergence of the finite-dimensional optimization technique employed, and (iii) convergence to a solution of the original problem as one takes finer subdivisions in the discretization scheme.

For an optimal control problem, many kinds of discretization schemes, or finite difference approximations, can be used, such as Euler, the midpoint (or box), trapezoidal, and Runge-Kutta schemes [1, 3, 4]. In these approximations both the state and control variables are discretized along a given time horizon. In [5] linear and cubic polynomials are fitted, respectively, for the states and controls, between each two consequtive discretization points. Another popular scheme of discretization, also referred to as control parameterization, takes a different approach: the control variables are approximated in each given subdivision by constants, linear functions or splines, whereas the dynamical system equations are solved accurately for the states where the approximate controls are used [6, 7, 8, 9, 10, 11]. Because the controls are approximated by a finite number of parameters, the optimal control problem becomes one of finding these parameters to achieve a minimum integral cost. Rather than this partial discretization approach, we consider full discretization, i.e. discretization of both the states and controls, in particular under the Euler scheme. Euler discretization is the simplest, in the sense that, under Euler, calculations are more straightforward and explicit optimality conditions can be derived more easily than under other schemes. This addresses the issue (i).

Convergence of the solution of the (fully) discretized problem to a solution of the original problem has been studied extensively in the literature (see [12, 13, 3, 4, 14, 15, 16] and the references therein). Under certain assumptions, Dontchev and Hager [13] give a convergence result for Euler discretization. We will use their result, which addresses the issue (iii).

The Inexact Restoration (IR) method and its several variants (tailored for different situations) were introduced by Martínez and his coworkers in [17, 18, 19] for solving finite dimensional constrained optimization problems. IR methods are modern versions of the classical feasible methods [20, 21, 22, 23, 24, 25, 26, 27, 28] for nonlinear programming. Each iteration of the IR method consists of two phases: in the first phase feasibility of the current iterate is improved, and in the second, the value of the cost is reduced in some tangent plane. Recently Birgin and Martínez [17] carried out a local convergence analysis of the IR method for finite dimensional optimization problems subject to equality constraints. They also provided extensive numerical comparisons with a well-known general-purpose optimization software. They demonstrated that the IR method is, overall, considerably more robust.

When discretized and written down as a mathematical programming problem, an unconstrained optimal control problem is transformed into a constrained optimization problem, where the constraints are given by the dynamical equations of the control system. In this paper we implement the IR method due to Birgin and Martínez [17] for the Euler discretization of the optimal control problem, and so address the issue (ii).

The paper is organized as follows. In Section 2, the infinite-dimensional setting is summarized, and assumptions are stated. In Section 3, we describe the Euler discretization of the optimal control problem and its Lagrangian formulation, and cite a convergence theorem from [13]. In Section 4, application of the IR method to the discretized optimal control

problem is described, further assumptions are stated, and the main convergence results are provided in Theorem 2 and Corollary 2. In Section 5, we show that a special case of IR is equivalent to projected Newton for quadratic problems. Finally, in Section 6, we illustrate and discuss a numerical implementation of the method by means of two test problems (one of which is non-quadratic), and provide comprehensive comparisons with the Newton and projected Newton methods.

2 Optimal Control Problem

We consider the optimal control problem

$$(\mathrm{P}) \begin{cases} \text{minimize} & \displaystyle\int_{t_0}^{t_f} f_0(x(t),u(t))\,dt \\ \text{subject to} & \dot{x}(t) = f(x(t),u(t)), \quad x(t_0) = x^0, \end{cases}$$

where the state variable $x(t) \in \mathbb{R}^n$, $\dot{x} = dx/dt$, the control variable $u(t) \in \mathbb{R}^m$, time $t \in [t_0, t_f]$ with fixed t_0 and t_f, and the functions $f_0 : \mathbb{R}^n \times \mathbb{R}^m \longrightarrow \mathbb{R}$ and $f : \mathbb{R}^n \times \mathbb{R}^m \longrightarrow \mathbb{R}^n$. The initial state is prescribed as x^0.

In the rest of this section we adopt the following terminology and notation from Dontchev and Hager [13]. Let $L^\alpha(t_0, t_f; \mathbb{R}^n)$ denote the Lebesgue space of measurable functions $x : [t_0, t_f] \longrightarrow \mathbb{R}^n$ with $\|x(\cdot)\|^\alpha$ integrable, equipped with the norm

$$\|x\|_{L^\alpha} = \left[\int_{t_0}^{t_f} \|x(t)\|^\alpha \, dt\right]^{1/\alpha},$$

where $\|\cdot\|$ is the Euclidean norm. The case $\alpha = \infty$ corresponds to the space of essentially bounded, measurable functions equipped with the essential supremum norm. By $W^{m,\alpha}(t_0, t_f; \mathbb{R}^n)$ we denote the Sobolev space consisting of functions $x : [t_0, t_f] \longrightarrow \mathbb{R}^n$ whose jth derivative lies in L^α for all $0 \le j \le m$ with the norm

$$\|x\|_{W^{m,\alpha}} = \sum_{j=0}^{m} \left\| \frac{d^j x}{dt^j} \right\|_{L^\alpha}.$$

Furthermore, H^m denotes the space $W^{m,2}$.

The Hamiltonian function associated with Problem (P) is defined by

$$H(x, u, \lambda) = f_0(x, u) + \lambda^T f(x, u),$$

where $\lambda(t) \in \mathbb{R}^n$ is the costate variable and the argument t has been supressed for clarity. We pose the following assumptions.

(A1) Problem (P) has a local solution (x^*, u^*) which lies in $W^{2,\infty} \times W^{1,\infty}$.

(A2) In a neighbourhood of (x^*, u^*), f_0 and f have first- and second-order partial derivatives which are Lipschitz continuous.

(A3) There exists a nontrivial costate $\lambda^* \in W^{2,\infty}$ associated with Problem (P) for which the following first-order necessary conditions of optimality (maximum principle) [29] is satisfied at (x^*, u^*, λ^*):

$$\dot{x} = f(x,u) , \qquad x(t_0) = x^0 , \tag{1}$$

$$-\dot{\lambda}^T = \frac{\partial H}{\partial x}(x,u,\lambda) = \frac{\partial f_0}{\partial x}(x,u) + \lambda^T \frac{\partial f}{\partial x}(x,u) , \qquad \lambda(t_f) = 0 , \tag{2}$$

$$0 = \frac{\partial H}{\partial u}(x,u,\lambda) = \frac{\partial f_0}{\partial u}(x,u) + \lambda^T \frac{\partial f}{\partial u}(x,u) . \tag{3}$$

We say that (x^*, u^*, λ^*) is a *critical triplet* of Problem (P) if it satisfies (1)-(3). We assume the Legendre condition $\partial^2 H/\partial u^2(x^*, u^*, \lambda^*) > 0$ holds so that we can solve the control u from (3) as $u = u(x, \lambda)$. Then the differential algebraic equations (1)-(3) with the given end conditions constitute a two-point boundary-value problem (TPBVP).

One more assumption, the so-called *coercivity*, is posed below, as in [13]. Let $w^* = (x^*, u^*, \lambda^*)$ and define

$$A^* = \frac{\partial f}{\partial x}(x^*, u^*) , \qquad B^* = \frac{\partial f}{\partial u}(x^*, u^*) ;$$

$$Q^* = \frac{\partial^2 H}{\partial x^2}(w^*) , \qquad M^* = \frac{\partial^2 H}{\partial x \partial u}(w^*) , \qquad R^* = \frac{\partial^2 H}{\partial u^2}(w^*) .$$

Let $\mathcal{B}$ be the quadratic form defined by

$$\mathcal{B}(x,u) = \frac{1}{2} \int_{t_0}^{t_f} \left[x^T(t) Q^* x(t) + u^T(t) R^* u(t) + 2\, x^T(t) M^* u(t) \right] dt .$$

(A4) There exists a constant $\alpha > 0$ such that

$$\mathcal{B}(x,u) \geq \alpha \, \|u\|_{L^2}^2 \quad \text{for all } (x,u) \in \mathcal{M}$$

where

$$\mathcal{M} = \{(x,u) : x \in H^1, u \in L^2,\ \dot{x} - A^* x - B^* u = 0,\ x(0) = 0\} .$$

As pointed in [13], (A4) is a strong form of a second-order sufficient optimality condition.

3 Discretization of the Optimal Control Problem

We subdivide the time horizon $[t_0, t_f]$ into N pieces, with the subdivision points t_i, $i = 0, 1, \ldots, N$, such that

$$t_0 < t_1 < t_2 < \cdots < t_N = t_f .$$

Define the partition

$$\pi := \{t_0, t_1, \ldots, t_N\} .$$

Without loss of generality, we take the partition points equidistant, namely that $\Delta t = t_{i+1} - t_i$, $i = 0, 1, 2, \ldots, N-1$. Now we consider the following one-step finite difference approximation, the so-called Euler scheme, of the system dynamics.

$$x_{i+1} = x_i + \Delta t \, f(x_i, u_i)$$

where x_i and u_i are the approximations of $x(t_i)$ and $u(t_i)$, respectively, but x_0 is prescribed. We approximate the integral

$$\int_{t_0}^{t_f} f_0(x(t), u(t))\, dt \ ,$$

by the Riemann sum

$$\Delta t \sum_{i=0}^{N-1} f_0(x_i, u_i) \ .$$

Now a finite difference approximation for Problem (P) can be given by

$$\text{(PE)} \begin{cases} \text{minimize} & \Delta t \sum_{i=0}^{N-1} f_0(x_i, u_i) \\ \text{subject to} & \Delta t\, f(x_i, u_i) - x_{i+1} + x_i = 0 \end{cases}$$

$i = 0, 1, 2, \ldots, N-1$, with x_0 prescribed. Problem (PE) is a finite-dimensional optimization problem, to which we will apply the IR method in Section 4 as given in [17].

Define

$$x_\pi := (x_1^T, x_2^T, \ldots, x_N^T)^T \in \mathbb{R}^{nN}, \quad u_\pi := (u_0^T, u_1^T, \ldots, u_{N-1}^T)^T \in \mathbb{R}^{mN}.$$

We will be speaking of x_π as the (*state*) *trajectory* of the discretized system. The Lagrangian for Problem (PE) is given as

$$L(x_\pi, u_\pi, \lambda_\pi) = \Delta t \sum_{i=0}^{N-1} f_0(x_i, u_i) + \sum_{i=0}^{N-1} \lambda_i^T \left[\Delta t\, f(x_i, u_i) - x_{i+1} + x_i\right] \tag{4}$$

where

$$\lambda_\pi := (\lambda_0^T, \lambda_1^T, \ldots, \lambda_{N-1}^T)^T \in \mathbb{R}^{nN}$$

is the Lagrange multiplier vector.

The first-order necessary conditions, namely the Karush-Kuhn-Tucker conditions, for Problem (PE) are given by

$$\nabla L(x_\pi, u_\pi, \lambda_\pi) = 0 \ ,$$

that is,

$$x_{i+1} = x_i + \Delta t\, f(x_i, u_i) \tag{5}$$

$$\lambda_{i-1}^T = \lambda_i^T + \Delta t \left[\frac{\partial f_0}{\partial x}(x_i, u_i) + \lambda_i^T \frac{\partial f}{\partial x}(x_i, u_i)\right] \tag{6}$$

$$0 = \frac{\partial f_0}{\partial u}(x_i, u_i) + \lambda_i^T \frac{\partial f}{\partial u}(x_i, u_i) \tag{7}$$

for $i = 0, 1, \ldots, N-1$, with $x_0 = x^0$ and $\lambda_{N-1} = 0$. The multiplier λ_{-1} can be associated with the trivial constraint $x_0 - x^0 = 0$. Therefore for $i = 0$ Equation (7) should be considered redundant. We say that $(x_\pi^*, u_\pi^*, \lambda_\pi^*)$ is a critical triplet of Problem (PE) if it satisfies (5)-(7). Note that Equation (5) is the Euler approximation of the ODE given in (1). On the other hand, Equation (6) provides the Euler approximation of (2), backwards in time.

Let $x_i' := (x_{i+1} - x_i)/\Delta t$. The discrete analogues of L^2, L^∞ and H^1 norms are defined as follows.

$$\|x_\pi\|_{L^2} = \left[\sum_i \Delta t \|x_i\|^2\right]^{1/2} , \qquad \|x_\pi\|_{L^\infty} = \sup_i \|x_i\| ,$$
$$\|x_\pi\|_{H^1} = \left[\|x_\pi\|_{L^2}^2 + \|x_\pi'\|_{L^2}^2\right]^{1/2} ,$$

where $\|x_i'\|_{L^2}^2 = \sum_i \Delta t \, \|x_i'\|^2$. The index i ranges over 1 to N for x_π and over 0 to $N-1$ for u_π and λ_π. The norm of a variable is taken over the relevant index range. Also define $x_\pi - x^*$ to be discrete such that $(x_\pi - x^*)_i := x_i - x^*(t_i)$. A discrete variable is said to be Lipschitz continuous in (discrete) time with Lipschitz constant M if $\|x_i'\| < M$ for each $i = 0, 1, \ldots, N-1$.

Theorem 1 *(Dontchev and Hager [13]) If Assumptions (A1)-(A4) hold, then for all sufficiently small Δt, there exists a local solution (x_π^*, u_π^*) of Problem (PE) and an associated Lagrange multiplier λ_π^* such that*

$$\|x_\pi^* - x^*\|_{H^1} + \|u_\pi^* - u^*\|_{L^2} + \|\lambda_\pi^* - \lambda^*\|_{H^1} \le c_1 \, \Delta t \ , \tag{8}$$

and

$$\|x_\pi^* - x^*\|_{W^{1,\infty}} + \|u_\pi^* - u^*\|_{L^\infty} + \|\lambda_\pi^* - \lambda^*\|_{W^{1,\infty}} \le c_2 \, (\Delta t)^{2/3} \ , \tag{9}$$

where c_1 and c_2 are constants independent of Δt.

Corollary 1 *If Assumptions (A1)-(A4) hold, then, as $\Delta t \longrightarrow 0$, a local solution of Problem (PE) converges to a local solution of Problem (P).*

4 IR for Optimal Control

In this section, we will formulate the IR method for solving optimal control problems.

Let

$$h_i(x_i, x_{i+1}, u_i) := \Delta t \, f(x_i, u_i) - x_{i+1} + x_i \ , \tag{10}$$

for $i = 0, 1, \ldots, N-1$, and define

$$h(x_\pi, u_\pi) := \left(h_0^T(x_0, x_1, u_0), h_1^T(x_1, x_2, u_1), \ldots, h_{N-1}^T(x_{N-1}, x_N, u_{N-1})\right)^T \ .$$

Problem (PE) can be rewritten as

$$\text{(PIR)} \begin{cases} \text{minimize} & \tilde{f}_0(x_\pi, u_\pi) \\ \text{subject to} & h(x_\pi, u_\pi) = 0 \end{cases}$$

with x_0 prescribed, where $\tilde{f}_0(x_\pi, u_\pi) := \Delta t \sum_{i=0}^{N-1} f_0(x_i, u_i)$.

We can translate the idea of the IR method presented by Birgin and Martínez in [17] as follows. Given the current iterate $(x_\pi, u_\pi) \in \mathbb{R}^{nN} \times \mathbb{R}^{mN}$, find, first a "more feasible" point $(y_\pi, u_\pi) \in \mathbb{R}^{nN} \times \mathbb{R}^{mN}$ (the *feasibility phase*), and then a "more optimal" point $(z_\pi, v_\pi) \in$

$\mathbb{R}^{nN} \times \mathbb{R}^{mN}$ in the tangent plane passing through (y_π, u_π) (the *optimality phase*). This tangent plane is formed by (z_π, v_π) which solves

$$h'(y_\pi, u_\pi)\,(z_\pi - y_\pi, v_\pi - u_\pi) = 0\ . \tag{11}$$

The tangent plane can equivalently be expressed, in a more explicit form, by using the linearization of $h_i(z_i, z_{i+1}, v_i)$ at (y_i, y_{i+1}, u_i) from (10):

$$\Delta t\,[A_i\,(z_i - y_i) + B_i\,(v_i - u_i)] - (z_{i+1} - y_{i+1}) + (z_i - y_i) = 0 \tag{12}$$

for $i = 0, 1, \ldots, N-1$, where we use the short-hand notation

$$A_i := \frac{\partial f}{\partial x}(y_i, u_i) \qquad \text{and} \qquad B_i := \frac{\partial f}{\partial u}(y_i, u_i)\ .$$

Note that the matrices A_i and B_i vary with the time index i.

In the optimality phase of the IR method we minimize the Lagrangian $L(z_\pi, v_\pi, \mu_\pi)$ given in (4) subject to the linear constraints described in (12). The Lagrangian associated with this optimization subproblem can be written as

$$\begin{aligned}
\widetilde{L}(z_\pi, v_\pi, \mu_\pi) \;:=\; & L(z_\pi, v_\pi, \lambda_\pi) \\
& + \sum_{i=0}^{N-1} (\mu_i^T - \lambda_i^T)\ \{\Delta t\,[A_i\,(z_i - y_i) + B_i\,(v_i - u_i)] \\
& \qquad - (z_{i+1} - y_{i+1}) + (z_i - y_i)\} \\
=\; & \Delta t \sum_{i=0}^{N-1} f_0(z_i, v_i) + \sum_{i=0}^{N-1} \lambda_i^T\,[\Delta t\, f(z_i, v_i) - z_{i+1} + z_i] \\
& + \sum_{i=0}^{N-1} (\mu_i^T - \lambda_i^T)\ \{\Delta t\,[A_i\,(z_i - y_i) + B_i\,(v_i - u_i)] \\
& \qquad - (z_{i+1} - y_{i+1}) + (z_i - y_i)\}
\end{aligned} \tag{13}$$

where $\mu_\pi = (\mu_0, \ldots, \mu_{N-1}) \in \mathbb{R}^{nN}$, and $(\mu_i - \lambda_i)$ are the Lagrange multipliers corresponding to the linear constraints in (12). The Karush-Kuhn-Tucker conditions in this case are given by

$$\nabla \widetilde{L}(z_\pi, v_\pi, \mu_\pi) = 0\ ,$$

which yields

$$z_{i+1} \;=\; z_i + (y_{i+1} - y_i) + \Delta t\,[A_i\,(z_i - y_i) + B_i\,(v_i - u_i)] \tag{14}$$

$$\mu_{i-1}^T \;=\; \mu_i^T + \Delta t \left[\frac{\partial f_0}{\partial z}(z_i, v_i) + \lambda_i^T \frac{\partial f}{\partial z}(z_i, v_i) + \left(\mu_i^T - \lambda_i^T\right)\ A_i\right] \tag{15}$$

$$0 \;=\; \frac{\partial f_0}{\partial v}(z_i, v_i) + \lambda_i^T \frac{\partial f}{\partial v}(z_i, v_i) + \left(\mu_i^T - \lambda_i^T\right)\ B_i \tag{16}$$

where $i = 0, 1, \ldots, N-1$, with $z_0 = x_0$ and $\mu_{N-1} = 0$. When (z_π, v_π) is reasonably close to (y_π, u_π) affine approximations can provide a good initial guess for an iterative technique such as conjugate gradients or similar, resulting in a solution to (14)-(16) much more easily, compared to (5)-(7).

In what follows we define certain vectors used in the IR method as presented in [17]. Norms of these vectors represent some measure of optimality in the optimality phase. Let

$$G_0(x_0, x_1, u_0, \lambda_0) := \left(\begin{array}{l} x_0^T + \Delta t\, f^T(x_0, u_0) - x_1^T \ , \\ 0^T \ , \\ \dfrac{\partial f_0}{\partial u}(x_0, u_0) + \lambda_0^T \dfrac{\partial f}{\partial u}(x_0, u_0) \end{array} \right)$$

and, for $i = 1, \ldots, N-1$,

$$G_i(x_i, x_{i+1}, u_i, \lambda_{i-1}, \lambda_i) := \left(\begin{array}{l} x_i^T + \Delta t\, f^T(x_i, u_i) - x_{i+1}^T \ , \\ \lambda_{i-1}^T - \Delta t \left[\dfrac{\partial f_0}{\partial x}(x_i, u_i) + \lambda_i^T \dfrac{\partial f}{\partial x}(x_i, u_i) \right] - \lambda_i^T \ , \\ \dfrac{\partial f_0}{\partial u}(x_i, u_i) + \lambda_i^T \dfrac{\partial f}{\partial u}(x_i, u_i) \end{array} \right) . \tag{17}$$

Next we define

$$G(x_\pi, u_\pi, \lambda_\pi) := \left(G_0^T(x_0, x_1, u_0, \lambda_0), \ldots, G_{N-1}^T(x_{N-1}, x_N, u_{N-1}, \lambda_{N-2}, \lambda_{N-1}) \right)^T \ .$$

Let

$$\begin{array}{l} \widetilde{G}_0(z_0, z_1, v_0, \mu_0, y_0, y_1, u_0, \lambda_0) := \\ \quad \left(\begin{array}{l} z_0^T + (y_1^T - y_0^T) + \Delta t \left[(z_0^T - y_0^T)\, A_0^T + (v_0^T - u_0^T)\, B_0^T \right] - z_1^T \\ \quad 0^T \ , \\ \quad \dfrac{\partial f_0}{\partial v}(z_0, v_0) + \lambda_0^T \dfrac{\partial f}{\partial v}(z_0, v_0) + \left(\mu_0^T - \lambda_0^T \right) \dfrac{\partial f}{\partial v}(y_0, u_0) \end{array} \right) \end{array}$$

and, for $i = 1, \ldots, N-1$,

$$\begin{array}{l} \widetilde{G}_i(z_i, z_{i+1}, v_i, \mu_{i-1}, \mu_i, y_i, y_{i+1}, u_i, \lambda_i) := \\ \quad \left(\begin{array}{l} z_i^T + (y_{i+1}^T - y_i^T) + \Delta t \left[(z_i^T - y_i^T)\, A_i^T + (v_i^T - u_i^T)\, B_i^T \right] - z_{i+1}^T \ , \\ \quad \mu_{i-1}^T - \Delta t \left[\dfrac{\partial f_0}{\partial z}(z_i, v_i) + \lambda_i^T \dfrac{\partial f}{\partial z}(z_i, v_i) + \left(\mu_i^T - \lambda_i^T \right) A_i \right] - \mu_i^T \ , \\ \quad \dfrac{\partial f_0}{\partial v}(z_i, v_i) + \lambda_i^T \dfrac{\partial f}{\partial v}(z_i, v_i) + \left(\mu_i^T - \lambda_i^T \right) B_i \end{array} \right) . \end{array} \tag{18}$$

We also define

$$\begin{array}{l} \widetilde{G}(z_\pi, v_\pi, \mu_\pi, y_\pi, u_\pi, \lambda_\pi) := \\ \quad \left(\widetilde{G}_0^T(z_0, z_1, v_0, \mu_0, y_0, y_1, u_0, \lambda_0), \ldots, \right. \\ \qquad \left. \widetilde{G}_{N-1}^T(z_{N-1}, z_N, v_{N-1}, \mu_{N-2}, \mu_{N-1}, y_{N-1}, y_N, u_{N-1}, \lambda_{N-1}) \right)^T . \end{array} \tag{19}$$

Birgin and Martínez give a list of conditions for the two phases of the IR method [17, Conditions (6)-(10)]. When these conditions are satisfied, one says that an IR iteration can

be completed (or is well-defined). We translate these conditions into our setting as follows. We say that an IR iteration starting from $(x_\pi, u_\pi, \lambda_\pi) \in \mathbb{R}^{nN} \times \mathbb{R}^{mN} \times \mathbb{R}^{nN}$ can be completed (or is well-defined) if one can compute $y_\pi \in \mathbb{R}^{nN}$, $(z_\pi, v_\pi, \mu_\pi) \in \mathbb{R}^{nN} \times \mathbb{R}^{mN} \times \mathbb{R}^{nN}$ such that it satisfies the following conditions.

$$\|h(y_\pi, u_\pi)\| \leq \theta \, \|h(x_\pi, u_\pi)\| \ , \tag{20}$$

$$\|y_\pi - x_\pi\| \leq K_1 \, \|h(x_\pi, u_\pi)\| \ , \tag{21}$$

$$\left\|h'(y_\pi, u_\pi) \, (z_\pi - y_\pi, v_\pi - u_\pi)\right\| \leq K_2 \, \|G(y_\pi, u_\pi, \lambda_\pi)\|^2 \ , \tag{22}$$

$$\left\|\widetilde{G}(z_\pi, v_\pi, \mu_\pi, y_\pi, u_\pi, \lambda_\pi)\right\| \leq \eta \, \|G(y_\pi, u_\pi, \lambda_\pi)\| \ , \tag{23}$$

$$\|z_\pi - y_\pi\| + \|v_\pi - u_\pi\| + \|\mu_\pi - \lambda_\pi\| \leq K_3 \, \|G(y_\pi, u_\pi, \lambda_\pi)\| \ , \tag{24}$$

where $\theta, \eta \in [0, 1)$, $K_1, K_3 > 0$ $K_2 \geq 0$, and $\|\cdot\|$ is any norm in the relevant finite dimensional space.

In the inexact restoration phase we set the initial point for the trajectories x_π and y_π to be the same; namely

(A5) $y_0 = x_0$.

Note that by (A2), there exists $K > 0$ such that for all x, y, u, and t,

$$\|f(y(t), u(t)) - f(x(t), u(t))\| \leq K \|y(t) - x(t)\| \ . \tag{25}$$

Lemma 1 *Suppose Assumptions (A2) and (A5) hold. If Condition* (20) *is satisfied, so is Condition* (21)*.*

Proof. Suppose (20) is satisfied. Then we have

$$\|h(y_\pi, u_\pi)\| \leq \theta \, \|h(x_\pi, u_\pi)\| \leq \|h(x_\pi, u_\pi)\| \ . \tag{26}$$

We can rewrite (10) as

$$x_{i+1} = x_i + \Delta t \, f(x_i, u_i) + h_i(x_i, x_{i+1}, u_i) \ ,$$

and similarly

$$y_{i+1} = y_i + \Delta t \, f(y_i, u_i) + h_i(y_i, y_{i+1}, u_i) \ .$$

for $i = 0, 1, \ldots, N-1$. Now

$$y_{i+1} - x_{i+1} = y_i - x_i + \Delta t \, \left(f(y_i, u_i) - f(x_i, u_i)\right) + h_i(y_i, y_{i+1}, u_i) - h_i(x_i, x_{i+1}, u_i) \ .$$

For notational convenience, define $r_i := h_i(y_i, y_{i+1}, u_i) - h_i(x_i, x_{i+1}, u_i)$, and that $r_\pi = (r_0, \ldots, r_{N-1})$. Note that $r_\pi = h(y_\pi, u_\pi) - h(x_\pi, u_\pi)$. Without loss of generality, in what follows we use the sup-norm, for example, $\|r_\pi\|_{L^\infty} = \sup_i \|r_i\|$, where $\|\cdot\|$ denotes the 1-norm in $\mathbb{R}^{nN}$ or in $\mathbb{R}^n$, appropriately. Now, for $i = 0, 1, \ldots, N-1$,

$$\|y_{i+1} - x_{i+1}\| \leq \|y_i - x_i\| + \Delta t \, \|f(y_i, u_i) - f(x_i, u_i)\| + \|r_i\| \ .$$

Thus, by (25),

$$\|y_{i+1} - x_{i+1}\| \leq (1 + \Delta t \, K) \, \|y_i - x_i\| + \|r_i\| \ . \tag{27}$$

Since $y_0 = x_0$ by Assumption (A5), (27) implies that

$$\|y_1 - x_1\| \leq \|r_0\| \ . \tag{28}$$

Furthermore, the inequalities (27), for $i = 1$, and (28), yield

$$\|y_2 - x_2\| \leq (1 + \Delta t\, K)\, \|r_0\| + \|r_1\| \ .$$

Proceeding inductively, we prove that, for all $i = 0, 1, \ldots, N-1$,

$$\begin{aligned} \|y_i - x_i\| &\leq (1 + \Delta t\, K)^{i-1} \|r_0\| + \ldots + (1 + \Delta t\, K)\, \|r_{i-2}\| + \|r_i\| \\ &\leq (1 + \Delta t\, K)^N \left(\|r_0\| + \ldots + \|r_{N-1}\|\right) \\ &\leq (1 + \Delta t\, K)^N\, N \sup_{0 \leq j \leq N-1} \|r_j\| \\ &= N\, (1 + \Delta t\, K)^N\, \|r_\pi\|_{L^\infty} \end{aligned}$$

This is valid for any i, so

$$\begin{aligned} \|y_\pi - x_\pi\|_{L^\infty} = \sup_{0 \leq j \leq N-1} \|y_i - x_i\| &\leq N\, (1 + \Delta t\, K)^N\, \|r_\pi\|_{L^\infty} \\ &\leq N\, (1 + \Delta t\, K)^N\, \|h(y_\pi, u_\pi) + h(x_\pi, u_\pi)\|_{L^\infty} \ . \end{aligned}$$

By (26), this implies that

$$\|y_\pi - x_\pi\|_{L^\infty} \leq 2N\, (1 + \Delta t\, K)^N \|h(x_\pi, u_\pi)\|_{L^\infty} \tag{29}$$

which is the required conclusion. □

Remark 1 *Inequality* (29) *implies that* K_1 *depends on the Lipschitz constant* K *and the number of subdivisions* N*; namely that we have* $K_1 \geq 2N\, (1 + \Delta t\, K)^N$*. We observe that* $(1+\Delta t\, K)^N = (1+(t_f - t_0)\, K/N)^N$ *is increasing in* N *and that, as* $\Delta t \longrightarrow 0$, $(1+\Delta t\, K)^N \longrightarrow e^{(t_f - t_0)K}$*. So one has* $K_1 \geq 2N\, e^{(t_f - t_0)K}$ *for any* N*.*

Remark 2 *In the optimal control problem we deal with, we can achieve* exact restoration *(i.e.* $h(y_\pi, u_\pi) = 0$*) easily, because we can simply find* y_π *using* $y_{i+1} = y_i + \Delta t\, f(y_i, u_i)$ *recursively with the given* u_π*. This case corresponds to* $\theta = 0$*, for which Lemma 1 still holds.*

Lemma 2 *Suppose Assumption (A2) holds. Then the first-order partial derivative of* ∇h_π *is Lipschitz continuous in* (x_π, u_π)*.*

Proof. It follows from the Lipschitz continuity of the first partial derivative of f. □

In the numerical experiments in Section 6, (14)-(16) will be solved "exactly" in each iteration of the IR method. Then we can set $K_2 = 0$, because we will implement the tangent plane (or the linearization of the constraint) in an exact way. Furthermore because we will achieve optimality exactly in the tangent plane, we have $\widetilde{G}(z_\pi, v_\pi, \mu_\pi, y_\pi, u_\pi, \lambda_\pi) = 0$, satisfying (23), with the choice of $\eta = 0$. Consequently, the pose the following assumption.

(A6) $K_2 = 0$ and $\eta = 0$.

Theorem 2 *Suppose Assumptions (A2), (A5) and (A6) hold, and Conditions* (20) *and* (24) *are satisfied. Then the sequence of IR iterates* $\{x_\pi^{(k)}, u_\pi^{(k)}, \lambda_\pi^{(k)}\}$ *is convergent to the critical triplet* $\{x_\pi^*, u_\pi^*, \lambda_\pi^*\}$ *of Problem (PE). Furthermore, if* $\theta = \eta = 0$*, convergence of the iterates is r-quadratic.*

Proof. By Assumption (A6) and Lemma 1, (21)-(23) are satisfied. Also by Lemma 2, the remaining hypotheses of Theorem 3 in [17] hold, yielding the first conclusion. The second conclusion is then provided by Theorem 5 in [17]. □

Corollary 2 *Suppose Assumptions (A1)-(A6) hold and Conditions* (20) *and* (24) *are satisfied. As* $\Delta t \longrightarrow 0$*, the solution* $\{x_\pi^*, u_\pi^*, \lambda_\pi^*\}$ *found by the IR method tends to the critical triplet* (x^*, u^*, λ^*)*.*

Proof. The statement is furnished by Theorem 2 and Corollary 1. □

5 Quadratic Problems

Consider the problem with quadratic cost and a polynomial dynamics such that

$$\text{(PQ)} \begin{cases} \text{minimize} & \displaystyle\int_{t_0}^{t_f} x^T(t)\, Q(t)\, x(t) + u^T(t)\, R(t)\, u(t)\, dt \\ \text{subject to} & \dot{x}(t) = f(x(t)) + H(x(t))\, u(t), \quad x(t_0) = x^0, \end{cases}$$

where Q is a time-varying $n \times n$ positive semi-definite matrix, R is a time-varying $m \times m$ positive definite matrix. The coordinates of the vector function $f(x)$ are quadratic in x, and $H(x)$ is an $n \times m$ matrix whose entries are linear in x. Dynamical systems of the form in Problem (PQ) commonly arise in mechanics, epidemic and ecological models.

It is not difficult to see that the optimality phase equations (14)-(16) for Problem (PQ) are linear in z_i, μ_i and v_i. Furthermore, (16) can be solved for v_i in z_i and μ_i uniquely. Substitution of the expression obtained for v_i into (14)-(15) yields a linear system of state-costate equations which can in general be solved much more easily than the original nonlinear equations (5)-(7). We provide a simple example to Problem (PQ) in Section 6.1.

5.1 IR and Projected Newton Methods

In this Subsection we show that a special case of the IR method is equivalent to the projected Newton method. For brevity, rewrite Problem (PIR), with $x = (x_\pi, u_\pi)$ only in this subsection, as

$$\text{(PIRa):} \qquad \text{minimize } \widetilde{f}_0(x), \quad \text{subject to } \; h(x) = 0\ .$$

Recall the classical Lagrangian function

$$L(x, \lambda) = \widetilde{f}_0(x) + \lambda^T h(x)$$

Denote the current iterate by (x, λ) and the next iterate by (z, μ). Newton's method finds (z, μ) by solving

$$\begin{bmatrix} \nabla_{xx} L(x,\lambda) & \nabla h(x)^T \\ \nabla h(x) & 0 \end{bmatrix} \begin{bmatrix} z-x \\ \mu-\lambda \end{bmatrix} = - \begin{bmatrix} \nabla_x L(x,\lambda)^T \\ h(x) \end{bmatrix} \tag{30}$$

Let the kth iterate be denoted by $(x^{(k)}, \lambda^{(k)})$. Then Newton iterations are carried out using (30) with $(x^{(k)}, \lambda^{(k)}) = (x, \lambda)$ and $(x^{(k+1)}, \lambda^{(k+1)}) = (z, \mu)$, $k = 0, 1, 2, \ldots$. If one sets $\lambda^{(k+1)} = \mu$, but chooses $x^{(k+1)}$ to satisfy $h(x^{(k+1)}) = 0$, then the method is referred to as a *projected Newton method* [30].

Proposition 1 *Suppose that Problem (PIRa) is quadratic, namely that $\widetilde{f}_0(x)$ and $h(x)$ are quadratic in x. If Assumption (A6) holds and $\theta = 0$, the IR method is a projected Newton method.*

Proof. Rewrite the "modified Lagrangian function" used in the optimality phase of IR for Problem (PIRa) as

$$\mathcal{L}(z,\mu) = L(z,\lambda) + (\mu-\lambda)^T \nabla h(y)\,(z-y)$$

Suppose $\widetilde{f}_0(x)$ and $h(x)$ are quadratic, namely, without loss of generality,

$$\widetilde{f}_0(x) = x^T Q x\ , \quad \text{and} \quad h_i(x) = x^T H_i x,$$

where Q and H_i, $i = 1, \ldots, m$, are $n \times n$ symmetric matrices. Then the necessary conditions of optimality dictates that

$$\begin{aligned} \nabla \mathcal{L}(z,\mu)^T &= \begin{bmatrix} \nabla \widetilde{f}_0(z)^T + \nabla h(z)^T \lambda + \nabla h(y)^T (\mu-\lambda) \\ \nabla h(y)^T (z-y) \end{bmatrix} \\ &= \begin{bmatrix} Qz + \sum_{i=1}^m \lambda_i H_i z + \nabla h(y)^T (\mu-\lambda) \\ \nabla h(y)^T (z-y) \end{bmatrix} \\ &= \begin{bmatrix} \left(Q + \sum_{i=1}^m \lambda_i H_i\right)(z-y) + \nabla h(y)^T (\mu-\lambda) + \left(Q + \sum_{i=1}^m \lambda_i H_i\right) y \\ \nabla h(y)^T (z-y) \end{bmatrix} = 0 \end{aligned} \tag{31}$$

After rearranging, one gets

$$\begin{bmatrix} Q + \sum_{i=1}^m \lambda_i H_i & \nabla h(y)^T \\ \nabla h(y) & 0 \end{bmatrix} \begin{bmatrix} z-y \\ \mu-\lambda \end{bmatrix} = - \begin{bmatrix} \left(Q + \sum_{i=1}^m \lambda_i H_i\right) y \\ 0 \end{bmatrix}$$

that is

$$\begin{bmatrix} \nabla_{yy} L(y,\lambda) & \nabla h(y)^T \\ \nabla h(y) & 0 \end{bmatrix} \begin{bmatrix} z-y \\ \mu-\lambda \end{bmatrix} = - \begin{bmatrix} \nabla_y L(y,\lambda)^T \\ 0 \end{bmatrix} \tag{32}$$

With $\theta = 0$, the pair (y, λ) is the current iterate such that $h(y) = 0$, i.e. y is a feasible point. Let the kth iterate be denoted by $(y^{(k)}, \lambda^{(k)})$, $k = 0, 1, 2, \ldots$. The next iterate $(y^{(k+1)}, \lambda^{(k+1)})$ is found as follows: The system (32) is solved for (z, μ) with $(y, \lambda) = (y^{(k)}, \lambda^{(k)})$. One sets $\lambda^{(k+1)} = \mu$, and $y^{(k+1)}$ is chosen to satisfy $h(y^{(k+1)}) = 0$.

In the case of projected Newton method, $h(x) = 0$ in (30), and so (32) is identical to (30). This furnishes the proposition. □

Remark 3 *Tapia and Whitley [30] establish super-quadratic convergence of order* $1+\sqrt{2}$ *for the projected Newton method applied to the eigenvalue problem of symmetric matrices, which is an equality constrained quadratic problem. Their discovery does not hold, however, for the eigenvalue problem of non-symmetric matrices. With Proposition 1, the convergence of IR sets conditions for the convergence of the projected Newton method for general equality constrained quadratic problems.*

If $\widetilde{f}_0(x)$ *or* $h(x)$ *is not quadratic in* x*, or* $\theta \neq 0$*, then IR and projected Newton are clearly different methods.*

6 Numerical Implementation

In this section we illustrate an implementation of the IR method through two test problems, one with a quadratic scalar system and the other with the van der Pol system. Recall that the IR method as described in [17] is a local technique. For comparison purposes we use the Newton and the projected Newton methods as applied to Equations (14)-(16). In the first example, which involves a quadratic problem, the IR method is the same as the projected Newton method, by virtue of Proposition 1. In the case of van der Pol system, the problem is not quadratic, so we provide comparisons with both the Newton and projected Newton methods.

6.1 A scalar system

Consider the problem

$$\text{(P1)} \begin{cases} \text{minimize} & \dfrac{1}{2}\displaystyle\int_0^1 (u^2(t)+x^2(t))\,dt \\ \text{subject to} & \dot{x}(t) = 1 - x(t) + x(t)u(t), \quad x(0) = 1\ , \end{cases}$$

where the state $x(t)$ is scalar. Necessary conditions (1)-(3) yield the following TPBVP.

$$\dot{x}(t) = 1 - x(t) - \lambda(t)\,x^2(t), \qquad x(t_0) = 1\ , \tag{33}$$

$$\dot{\lambda}(t) = (\lambda^2(t) - 1)\,x(t) + \lambda(t) \qquad \lambda(1) = 0\ , \tag{34}$$

where the optimal control $u(t) = -\lambda(t)\,x(t)$ has been substituted. The TPBVP (33)-(34) can be solved accurately by using shooting techniques if a good initial guess for $\lambda(0)$ is provided. The solution is found to be the initial value problem (IVP) with the initial conditions

$$x(0) = 1 \quad \text{and} \quad \lambda(0) = 0.492349\ ,$$

and the cost incurred is obtained as 0.439560, up to the given accuracy.

The optimality phase equations (14)-(16) for this example become

$$z_{i+1} \;=\; z_i + (y_{i+1} - y_i) + \Delta t\,[(u_i - 1)\,(z_i - y_i) + y_i\,(v_i - u_i)]\ , \tag{35}$$

$$\mu_{i-1} \;=\; \mu_i + \Delta t\,[z_i + \lambda_i\,(v_i - 1) + (\mu_i - \lambda_i)\,(u_i - 1)]\ , \tag{36}$$

$$v_i \;=\; -\lambda_i\,z_i - (\mu_i - \lambda_i)\,y_i\ . \tag{37}$$

for $i = 0, \ldots, N-1$, with $x_0 = 1$ and $\mu_{N-1} = 0$. These form a linear system of equations with the unknowns z_π, v_π and μ_π, where the corresponding $(3N \times 3N)$ coefficient matrix

is sparse. Note that (35)-(37) can be put into a computationally more convenient form, by substituting v_i in (37) into (35) and (36). After doing this substitution and rearranging (36) one gets the recursive expressions

$$v_i = -\lambda_i z_i - (\mu_i - \lambda_i) y_i \ , \tag{38}$$

$$z_{i+1} = z_i + (y_{i+1} - y_i) + \Delta t \left[(u_i - 1)(z_i - y_i) + y_i (v_i - u_i) \right] \ , \tag{39}$$

$$\mu_{i+1} = \frac{\mu_i - \Delta t \left[(1 - \lambda_{i+1}^2) z_{i+1} + \lambda_{i+1}^2 y_{i+1} - \lambda_{i+1} u_{i+1} \right]}{1 - \Delta t (1 + \lambda_{i+1} y_{i+1} - u_{i+1})} \ , \tag{40}$$

with $x_0 = 1$ and $\mu_{N-1} = 0$. Note that (40) is well-defined for sufficiently small Δt. With the substitution of v_i, Equations (39)-(40) constitute a standard discrete-time TPBVP in the variables z_π and μ_π. Because this TPBVP is linear (in z_i and μ_i), unless it is ill-conditioned, a solution to it can be found in just one iteration. This is more efficient than solving the sparse linear system of equations in (35)-(37).

We carry out IR iterations with $\theta = \eta = K_2 = 0$. In other words, we solve the subproblems in the feasibility and optimality phases *exactly*, where the linearization of the constraint is also exact. We adopt this exact approach, because it is computationally less demanding than the inexact version. We tabulate in Tables 1 and 2 the IR and Newton iterations with $N = 8$ (or 2^3) and $N = 32768$ (or 2^{15}), respectively. In each table, we list the values

$$d^{(k)} = \|z_\pi - x_\pi^*\|_{L^\infty} + \|v_\pi - u_\pi^*\|_{L^\infty} + \|\mu_\pi - \lambda_\pi^*\|_{L^\infty} \ , \tag{41}$$

$$K_1^{(k)} = \frac{\|y_\pi - x_\pi\|_{L^\infty}}{\|h(x_\pi, u_\pi)\|_{L^\infty}} \ , \tag{42}$$

$$K_3^{(k)} = \frac{\|z_\pi - y_\pi\|_{L^\infty} + \|v_\pi - u_\pi\|_{L^\infty} + \|\mu_\pi - \lambda_\pi\|_{L^\infty}}{\|G(y_\pi, u_\pi, \lambda_\pi)\|_{L^\infty}} \ , \tag{43}$$

where the superscript k has been omitted from the terms on the right-hand side for clarity. We also tabulate the cost, i.e. the value of $\widetilde{f}_0$, at each iteration. Note that the maxima of the sequences $K_1^{(k)}$ and $K_3^{(k)}$ throughout all of the iterations roughly constitute lower bounds for the parameters K_1 and K_3, respectively. At the beginning, we choose

$$\lambda_i^{(0)} = \lambda_0^{(0)} + \frac{i}{N-1} \left(\lambda_{N-1}^{(0)} - \lambda_0^{(0)} \right) \ , \tag{44}$$

which are points equally spaced along a straight line between the guess for the initial costate $\lambda_0^{(0)}$ and the terminal costate $\lambda_{N-1}^{(0)} = 0$. We also use at the start $u_i^{(0)} = -\lambda_i^{(0)} y_i^{(0)}$ for $i = 0, 1, \ldots, N$, recursively. In each iteration ($k = 0, 1, \ldots$) of IR,

$$y_{i+1}^{(k)} = y_i^{(k)} + \Delta t \left(1 - y_i^{(k)} + y_i^{(k)} u_i^{(k)} \right), \quad y_0 = x_0,$$

is used to achieve feasibility.

In both of the cases when $N = 2^3$ and $N = 2^{15}$, the IR conditions (21) and (23) can be satisfied by setting K_1 and K_3 to appropriate values. Recall that the rest of the conditions, where we choose $\theta = \eta = K_2 = 0$, are also satisfied. This assures that IR iterations can be completed in each iteration and that convergence of the vectors x_π, u_π and λ_π is achieved.

Numerical experiments with the IR and Newton methods show interesting similarities as well as differences (see Tables 1 and 2). Based on the wider numerical evidence, including the other experiments we did with various starting guesses, IR seems to be at least as fast

	IR					Newton		
k	$K_1^{(k)}$	$K_3^{(k)}$	$d^{(k)}$	$\frac{d^{(k)}}{(d^{(k-1)})^2}$	*cost*	$d^{(k)}$	$\frac{d^{(k)}}{(d^{(k-1)})^2}$	*cost*
0	—	16.2	4.2×10^{0}		1.441334	4.2×10^{0}		1.441334
1	2.7	1.5	1.6×10^{0}	0.091	0.622911	2.3×10^{0}	0.130	0.500589
2	3.0	1.8	1.2×10^{-1}	0.046	0.442613	3.2×10^{-1}	0.061	0.396110
3	2.2	2.4	8.5×10^{-4}	0.060	0.441798	7.3×10^{-3}	0.071	0.442573
4	2.7	1.5	5.0×10^{-8}	0.070	0.441798	1.9×10^{-6}	0.035	0.441798

Table 1: IR and Newton iterations with $N = 8$ and $\lambda_0^{(0)} = -1.2$. Solution is obtained with $\lambda_0^* = 0.479518$.

	IR					Newton		
k	$K_1^{(k)}$	$K_3^{(k)}$	$d^{(k)}$	$\frac{d^{(k)}}{(d^{(k-1)})^2}$	*cost*	$d^{(k)}$	$\frac{d^{(k)}}{(d^{(k-1)})^2}$	*cost*
0	—	80995.7	4.3×10^{0}		1.563850	4.3×10^{0}		1.563850
1	10488	2.0	2.3×10^{0}	0.128	0.800903	3.4×10^{0}	0.186	0.689536
2	12221	2.7	2.1×10^{-1}	0.039	0.442607	7.1×10^{-1}	0.062	0.325750
3	9159	3.8	2.9×10^{-3}	0.064	0.439561	4.8×10^{-2}	0.097	0.446216
4	9836	2.7	7.9×10^{-7}	0.095	0.439561	1.4×10^{-4}	0.061	0.439552
5						4.4×10^{-10}	0.021	0.439561

Table 2: IR and Newton iterations with $N = 2^{15}$ (or 32768) and $\lambda_0^{(0)} = -1.2$. Solution is obtained with $\lambda_0^* = 0.492346$.

as Newton for this particular problem. In fact, with the given initial guess, IR reaches the solution one step earlier than Newton, pointing to robustness in the early iterations. Local convergence of IR for this problem seems to be at least quadratic, rather than just r-quadratic. Moreover, in most of the experiments we have done, we observe that while the cost is decreasing monotonously with IR, this is not the case with Newton.

It is well-known that Newton's method obeys the so-called mesh independence principle [31, 32]; namely its convergence behaviour does not change with the mesh size, Δt, or with N. We note in this example application that convergence behaviour of IR also seems to be independent from the mesh size. An inspection of the convergence proof in [17] reveals that K_1 is one of the parameters on which the convergence rate depends. However, by Remark 1, K_1 depends on N, which is also evident from Tables 1 and 2. Therefore convergence rate should be expected to depend on N. Nevertheless, the evidence observed on mesh independence of IR in Tables 1 and 2 warrants further investigation.

6.2 The van der Pol system

The van der Pol system has been used as a test problem in various optimal control studies [10, 33]. We consider the van der Pol system with unbounded control,

$$\dot{x}_1(t) = x_2(t) \ , \tag{45}$$

$$\dot{x}_2(t) = -x_1(t) - (x_1^2(t) - 1)\, x_2(t) + u(t) \ , \tag{46}$$

where $x(0) = (1,\ 1)$, with the aim of minimizing the quadratic cost

$$\frac{1}{2}\int_0^1 (x_1^2(t) + x_2^2(t) + u^2(t))dt\ .$$

The necessary conditions (1)-(3) yield the following TPBVP, where the explicit expression for the optimal control, $u(t) = -\lambda_2(t)$, has been substituted.

$$\dot{x}_1 = x_2\ , \tag{47}$$
$$\dot{x}_2 = -x_1 - (x_1^2 - 1)\, x_2 - \lambda_2\ , \tag{48}$$
$$\dot{\lambda}_1 = -x_1 + (1 + 2x_1x_2)\, \lambda_2\ , \tag{49}$$
$$\dot{\lambda}_2 = -x_2 - \lambda_1 + (x_1^2 - 1)\, \lambda_2\ , \tag{50}$$
$$\text{with} \quad x(0) = (-2,\ 4) \quad \text{and} \quad \lambda(1) = (0, 0)\ .$$

An accurate solution to these equations can be obtained with $\lambda^*(0) = (7.107556,\ 2.943488)$, where the (critical) cost is found as 5.544564.

The optimality phase equations for the van der Pol system can be rearranged in the same fashion as those given in (38)-(40), giving rise to a standard discrete-time TPBVP. We are not writing out these equations here, because of their complex appearance. At the beginning, we choose $\lambda_i^{(0)}$ in the same way it was chosen in (44) with some given $\lambda_0^{(0)}$ and with $\lambda_{N-1}^{(0)} = 0$. We also use at the start $u_i^{(0)} = -\lambda_i^{(0)}$ for $i = 0, 1, \ldots, N$, recursively. In each iteration ($k = 0, 1, \ldots$), we set

$$y_{i+1}^{(k)} = y_i^{(k)} + \Delta t\, f(y_i^{(k)}, u_i^{(k)}), \quad y_0 = x^0\ .$$

In this way, we obtain exact feasibility.

The IR iterations, as well as the Newton and Projected Newton iterations, for $N = 2^3$ (or 8) and $N = 2^{15}$ (or 32768) are listed in Tables 3 and 4, respectively. The values $d^{(k)}$, $K_1^{(k)}$, $K_3^{(k)}$ are defined as in (41)-(43).

Finite values of K_1 and K_3, along with other requirements, assure the convergence of IR. As in Subsection 6.1, IR seems to be at least as fast as Newton, exhibiting convergence at a quadratic rate. Because the problem is not quadratic, IR is different from the projected Newton, this time. Therefore we also include the projected Newton in the comparisons. As in Subsection 6.1 we observe that cost does not necessarily decrease monotonously with Newton, as opposed to both IR and the projected Newton. Tables 3 and 4 verify this behaviour.

The fact that the behaviour of the IR iterations with $N = 2^3$ and $N = 2^{15}$ are similar prompts again the issue of mesh independence. However, as discussed in the scalar case before, we should leave this as part of a separate investigation.

	IR					Newton			Projected Newton		
k	$K_1^{(k)}$	$K_3^{(k)}$	$d^{(k)}$	$\frac{d^{(k)}}{(d^{(k-1)})^2}$	*cost*	$d^{(k)}$	$\frac{d^{(k)}}{(d^{(k-1)})^2}$	*cost*	$d^{(k)}$	$\frac{d^{(k)}}{(d^{(k-1)})^2}$	*cost*
0	—	3.3	2.9×10^{1}		20.485664	2.9×10^{1}		20.485664	2.9×10^{1}		20.485664
1	4.4	7.6	1.6×10^{1}	0.0184	5.411196	2.2×10^{1}	0.0248	5.993229	1.9×10^{1}	0.0224	6.166872
2	3.8	2.6	5.2×10^{-1}	0.0021	4.402118	1.7×10^{0}	0.0036	4.255352	1.2×10^{0}	0.0033	4.403729
3	2.0	2.0	8.9×10^{-5}	0.0003	4.401793	1.3×10^{-3}	0.0005	4.401682	6.8×10^{-4}	0.0004	4.401793
4	1.4	4.1	1.3×10^{-10}	0.0171	4.401793	8.1×10^{-9}	0.0047	4.401793	8.9×10^{-10}	0.0020	4.401793

Table 3: IR, Newton and projected Newton iterations with $N = 8$ and $\lambda_0^{(0)} = (10,\ 10)$. Solution is obtained with $\lambda_0^* = (2.223707,\ 1.912639)$.

	IR					Newton			Projected Newton		
k	$K_1^{(k)}$	$K_3^{(k)}$	$d^{(k)}$	$\frac{d^{(k)}}{(d^{(k-1)})^2}$	*cost*	$d^{(k)}$	$\frac{d^{(k)}}{(d^{(k-1)})^2}$	*cost*	$d^{(k)}$	$\frac{d^{(k)}}{(d^{(k-1)})^2}$	*cost*
0	—	3651	2.4×10^{1}		18.640039	2.4×10^{1}		18.640039	2.4×10^{1}		18.640039
1	25660	48962	1.6×10^{1}	0.0273	6.548757	2.2×10^{1}	0.0377	7.188487	2.0×10^{1}	0.0349	7.427308
2	2854	26756	1.1×10^{0}	0.0044	5.546653	4.5×10^{0}	0.0093	5.291185	3.3×10^{0}	0.0080	5.574305
3	4764	20580	3.0×10^{-3}	0.0025	5.544268	2.3×10^{-1}	0.0116	5.546968	8.8×10^{-2}	0.0080	5.544293
4	4652	2	1.9×10^{-8}	0.0021	5.544268	5.3×10^{-4}	0.0099	5.544279	5.7×10^{-5}	0.0074	5.544268
5						2.6×10^{-9}	0.0092	5.544268	2.2×10^{-11}	0.0067	5.544268

Table 4: IR, Newton and projected Newton iterations with $N = 2^{15}$ (or 32768) and $\lambda_0^{(0)} = (10,\ 10)$. Solution is obtained with $\lambda_0^* = (7.106053,\ 2.943200)$.

We define the piecewise-linear function approximation $x^{(k)}(t)$ in the kth IR iteration by

$$x^{(k)}(t) = x_i^{(k)} + (Nt - i)\,(x_{i+1}^{(k)} - x_i^{(k)}), \qquad \text{for } t \in \left[\frac{i}{N}, \frac{i+1}{N}\right) ,$$

where $i = 0, 1, \ldots, N$. The piecewise-linear approximations $u^{(k)}(t)$ and $\lambda^{(k)}(t)$ are defined similarly. In Figure 1, we depict the IR iterations with $N = 2^{15}$. In the first column, solid curves illustrate a current iterate $x^{(k)}(t)$, $k = 1, \ldots, 4$, and the dashed curves a previous iterate, $x^{(k-1)}(t)$, $k = 1, \ldots, 4$. The second and third columns similarly depict $u^{(k)}(t)$ and $\lambda^{(k)}(t)$, and their respective previous iterates, $u^{(k-1)}(t)$ and $\lambda^{(k-1)}(t)$, $k = 1, \ldots, 4$.

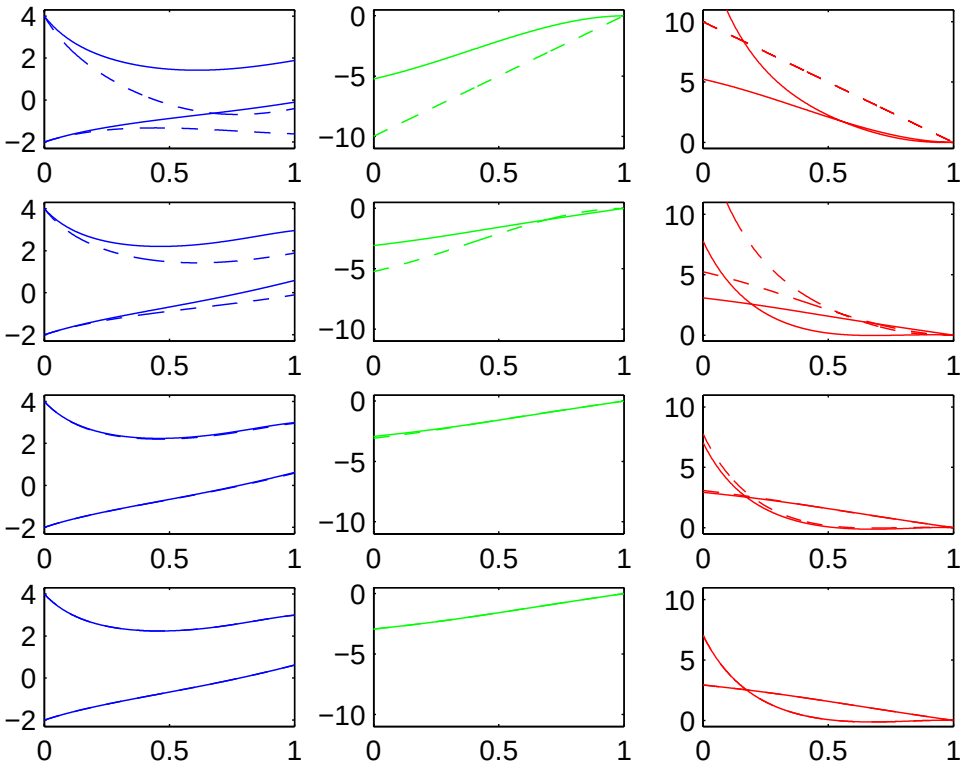

Figure 1: A graphical depiction of the IR iterations for the van der Pol system with $N = 2^{15}$. Solid curves in each respective column represent $x^{(k)}(t)$, $u^{(k)}(t)$ and $\lambda^{(k)}(t)$, $k = 1, \ldots, 4$; dashed curves in each respective column represent $x^{(k-1)}(t)$, $u^{(k-1)}(t)$ and $\lambda^{(k-1)}(t)$, $k = 1, \ldots, 4$.

7 Conclusions

Many (but not all) Nonlinear Programming problems have the property that, given an arbitrary nonfeasible point x, a (perhaps approximately) feasible point y (related with x) is easy to obtain. The Inexact Restoration theory [19, 17] sets sufficient conditions that the restored point y must satisfy in order that the Inexact Restoration method enjoys global and local convergence properties. Essentially, these conditions are that y should be sufficiently

more feasible than x and that the distance between y and x should be, at most, of the order of the infeasibility measure. In this paper we addressed the discretized version of the control problem (P) and we illustrated that the conditions that make the Inexact Restoration method applicable are fulfilled. From the practical point of view, we explained why IR is computationally attractive in this case and we described the application to classical control problems. The numerical results were encouraging in the sense that practical convergence in a few iterations was verified in all the test examples and the general performance seemed to be at least as good as the Newton and projected Newton methods.

The IR method with relatively small N can obviously solve the discretized problem much more efficiently than shooting techniques can solve the continuous-time problem. So, instead of the computationally demanding first few iterations of shooting techniques, one can use the IR method to obtain a reasonable approximation of the solution quickly, and then use this approximation as a good initial guess in shooting methods so as to sharpen the solution and reach a desired accuracy. Because the IR iterates are approximations of functions over the given time horizon, the approximate solution would constitute a good initial guess particularly for multiple shooting techniques.

Many computational schemes for optimal control involve finding some solution of a local subproblem efficiently, which may not be required to be very accurate - an example to such a scheme is the leapfrog algorithm [34]. The IR method seems to be well-suited for solving such subproblems, too.

Future research should include the application and theoretical analysis of the IR technique to more general control problems, for example, with the addition of constraints on the control and/or the state.

References

[1] ASCHER, U. M., MATTHEIJ, R. M. M., and RUSSELL, R. D., *Numerical Solution of Boundary Value Problems for Ordinary Differential Equations*, SIAM Publications, Philadelphia, 1995.

[2] STOER, J., and BULIRSCH, R., *Introduction to Numerical Analysis*, Second Edition, New York: Springer Verlag, 1993.

[3] HAGER, W. W., *Rates of Convergence for Discrete Approximations to Unconstrained Control Problems*, SIAM Journal on Numerical Analysis, Vol. 13, No. 4, pp. 449–472, 1976.

[4] HAGER, W. W., *Runge-Kutta Methods in Optimal Control and the Transformed Adjoint System*, Numerische Mathematik, Vol. 87, pp. 247–282, 2000.

[5] VON STRYK, O., *Numerical Solution of Optimal Control Problems by Direct Collocation*, in: R. Bulirsch, A. Miele, J. Stoer, K.-H. Well (eds.): Optimal Control - Calculus of Variations, Optimal Control Theory and Numerical Methods, International Series of Numerical Mathematics 111, Basel: Birkhuser, pp. 129-143, 1993.

[6] SIRISENA, H. R., and CHOU, F. S., *Convergence of the Control Parameterization Ritz Method for Nonlinear Optimal Control Problems*, Journal of Optimization Theory and Applications, Vol. 29. No. 3, 1979.

[7] TEO, K. L., GOH, C. J., and WONG, K. H., *A Unified Computational Approach to Optimal Control Problems*, Longman Scientific and Technical, New York, 1991.

[8] BÜSKENS, C., *Optimierungsmethoden und Sensitivitätsanalyse für optimale Steuerprozesse mit Steuer- und Zustands-Beschränkungen*, Dissertation, Institut für Numerische Mathematik, Universität Münster, 1998.

[9] LUUS, R., *Iterative Dynamic Programming*, Chapman & Hall/CRC, 2000.

[10] KAYA, C. Y., and NOAKES, J. L., *Computational Method for Time-Optimal Switching Control*, Journal of Optimization Theory and Applications, Vol. 117, No. 1, pp. 69–92, 2003.

[11] KAYA, C. Y., LUCAS, S. K., and SIMAKOV, S. T., *Computations for Bang–Bang Constrained Optimal Control Using a Mathematical Programming Formulation*, Optimal Control Applications and Methods, Vol. 25, No. 6, pp. 295–308, 2004.

[12] DONTCHEV, A. L., *An a Priori Estimate for Discrete Approximations in Nonlinear Optimal Control*, SIAM Journal on Control and Optimization, Vol. 34, No. 4, pp. 1315–1328, 2000.

[13] DONTCHEV, A. L., and HAGER, W. W., *The Euler Approximation in State Constrained Optimal Control*, Mathematics of Computation, Vol. 70, pp. 173–203, 2000.

[14] MALANOWSKI, K., BÜSKENS, C., and MAURER, H., *Convergence of Approximations to Nonlinear Optimal Control Problems*. In: *Mathematical Programming with Data Perturbations V*, ed. A. V. Fiacco, Lecture Notes in Pure and Applied Mathematics, Vol. 195 (Dekker, 1997), pp. 253–284, 1997.

[15] MORDUKHOVICH, B., *On Difference Approximations of Optimal Control Systems*, Journal of Applied Mathematics and Mechanics, Vol. 42, pp. 452–461, 1978.

[16] VELIOV, V., *On the Time-Discretization of Control Systems*, SIAM Journal on Control and Optimization, Vol. 35, No. 5, pp. 1470–1486, 1997.

[17] BIRGIN, E. G., and MARTÍNEZ, J. M., *Local Convergence of an Inexact-Restoration Method and Numerical Experiments*, Journal of Optimization Theory and Applications, Vol. 127, No. 2, pp. 229-247, 2005.

[18] MARTÍNEZ, J. M., and PILOTTA, E. A., *Inexact Restoration Algorithm for Constrained Optimization*, Journal of Optimization Theory and Applications, Vol. 104, No. 1, pp. 135–163, 2000.

[19] MARTÍNEZ, J. M., *Inexact Restoration Method with Lagrangian Tangent Decrease and New Merit Function for Nonlinear Programming*. Journal of Optimization Theory and Applications, Vol. 111, pp. 39–58, 2001.

[20] ABADIE, J., and CARPENTIER, J., *Generalization of the Wolfe Reduced-Gradient Method to the Case of Nonlinear Constraints*, Optimization, Edited by R. Fletcher, Academic Press, New York, NY, pp. 37–47, 1968.

[21] DRUD, A., *CONOPT: A GRG Code for Large Sparse Dynamic Nonlinear Optimization Problems*, Mathematical Programming, Vol. 31, pp. 153–191, 1985.

[22] LASDON, L.S., *Reduced Gradient Methods*, Nonlinear Optimization 1981, Edited by M.J.D. Powell, Academic Press, New York, NY, pp. 235–242, 1982.

[23] MIELE, A., HUANG, H. Y., and HEIDEMAN, J. C., *Sequential Gradient-Restoration Algorithm for the Minimization of Constrained Functions: Ordinary and Conjugate Gradient Version*, Journal of Optimization Theory and Applications, Vol. 4, pp. 213–246, 1969.

[24] MIELE, A., LEVY, A. V., and CRAGG, E. E., *Modifications and Extensions of the Conjugate-Gradient Restoration Algorithm for Mathematical Programming Problems*, Journal of Optimization Theory and Applications, Vol. 7, pp. 450–472, 1971.

[25] MIELE, A., SIMS, E. M., and BASAPUR, V. K., *Sequential Gradient-Restoration Algorithm for Mathematical Programming Problems with Inequality Constraints, Part 1, Theory*, Rice University, Aero-Astronautics Report No. 168, 1983.

[26] ROSEN, J. B., *The Gradient Projection Method for Nonlinear Programming, Part 1, Linear Constraints*, SIAM Journal on Applied Mathematics, Vol. 8, pp. 181–217, 1960.

[27] ROSEN, J. B., *The Gradient Projection Method for Nonlinear Programming, Part 2, Nonlinear Constraints*, SIAM Journal on Applied Mathematics, Vol. 9, pp. 514–532, 1961.

[28] ROSEN, J. B., and KREUSER, J., *A Gradient Projection Algorithm for Nonlinear Constraints*, Numerical Methods for Nonlinear Optimization, Edited by F.A. Lootsma, Academic Press, London, UK, pp, 297–300, 1972.

[29] HESTENES, M. R., *Calculus of Variations and Optimal Control Theory*, New York: John Wiley & Sons, 1966.

[30] TAPIA, R. A., and WHITLEY, D. L., *The Projected Newton Method Has Order* $1+\sqrt{2}$ *for the Symmetric Eigenvalue Problem*, SIAM Journal on Numerical Analysis, **25**(6), 1376-1382, 1988.

[31] DONTCHEV, A. L., HAGER, W. W., and VELIOV, V. M., *Uniform Convergence and Mesh Independence of Newton's Method for Discretized Variational Problems*, SIAM Journal on Control and Optimization, Vol. 39, No. 3, pp. 961-980.

[32] ALT, W., *Mesh-Independence of the Lagrange-Newton Method for Nonlinear Optimal Control Problems and Their Discretizations*, Annals of Operations Research, Vol. 101, pp. 101–117, 2001.

[33] MAURER H., and OSMOLOVSKII, N. P., *Second Order Sufficient Conditions for Time-Optimal Bang–Bang Control*, SIAM Journal on Control and Optimization, Vol. 42, pp. 2239-2263, 2004.

[34] KAYA, C. Y., and NOAKES, J. L., *Leapfrog for Optimal Control*, submitted.