Journal of Process Control 50 (2017) 19–28

Contents lists available at ScienceDirect

Journal of Process Control

journal homepage: www.elsevier.com/locate/jprocont

# Efficient symmetric Hessian propagation for direct optimal control☆

Rien Quirynen[a,c,*], Boris Houska[b], Moritz Diehl[c]

[a] Department ESAT-STADIUS, KU Leuven University, Kasteelpark Arenberg 10, 3001 Leuven, Belgium
[b] School of Information Science and Technology, ShanghaiTech University, 319 Yueyang Road, Shanghai 200031, China
[c] Department IMTEK, University of Freiburg, Georges-Koehler-Allee 102, 79110 Freiburg, Germany

ARTICLE INFO

*Article history:*
Received 15 April 2016
Received in revised form 4 October 2016
Accepted 30 November 2016
Available online 2 January 2017

*Keywords:*
Optimal control
Sensitivity analysis
Algorithms and software
Nonlinear predictive control

ABSTRACT

Direct optimal control algorithms first discretize the continuous-time optimal control problem and then solve the resulting finite dimensional optimization problem. If Newton type optimization algorithms are used for solving the discretized problem, accurate first as well as second order sensitivity information needs to be computed. This article develops a novel approach for computing Hessian matrices which is tailored for optimal control. Algorithmic differentiation based schemes are proposed for both discrete- and continuous-time sensitivity propagation, including explicit as well as implicit systems of equations. The presented method exploits the symmetry of Hessian matrices, which typically results in a computational speedup of about factor 2 over standard differentiation techniques. These symmetric sensitivity equations additionally allow for a three-sweep propagation technique that can significantly reduce the memory requirements, by avoiding the need to store a trajectory of forward sensitivities. The performance of this symmetric sensitivity propagation is demonstrated for the benchmark case study of the economic optimal control of a nonlinear biochemical reactor, based on the open-source software implementation in the ACADO Toolkit.



## 1. Introduction

Optimal control problems (OCP) arise in various applications, including the optimization of an actuation profile, system state and parameter estimation or optimal experiment design problems [1]. Shooting based direct optimal control methods rely on an accurate discretization of the original continuous-time OCP, resulting in a finite dimensional nonlinear program (NLP) [2,3]. In the context of nonlinear model predictive control (NMPC) or moving horizon estimation (MHE) [4], for example, these optimal control problems have to be solved under strict timing constraints [5]. In many practical control applications, the simulation of the nonlinear dynamics as well as the propagation of first and second order derivative information is the main bottleneck in terms of computation time. In particular, Newton type optimization algorithms [6] such as interior point (IP) methods [7] or sequential quadratic programming (SQP) [8] for optimal control require the reliable and efficient evaluation of such derivatives [9]. This is especially crucial in the context of economic objectives, where Gauss–Newton or other Hessian approximation strategies may fail to perform well [10].

Existing direct optimal control algorithms based on shooting methods either employ a *discretize-then-differentiate* [2] or a *differentiate-then-discretize* type of approach. The first technique carries out the differentiation after discretization, e.g., following the technique of internal numerical differentiation (IND) [2] in combination with algorithmic differentiation (AD) [9] to evaluate these derivatives. Other discrete-time sensitivity propagation schemes are tailored for implicit integration methods [11,12]. The *differentiate-then-discretize* approach involves an extension of the dynamic equations with their corresponding sensitivity equations, as implemented for example for first order sensitivities in SUNDIALS [13]. This can be performed for both forward [14,15] and adjoint sensitivity analysis [16,17]. Note that direct transcription methods [18] do not require such a propagation of sensitivities as in shooting based approaches, but they can still benefit from the proposed symmetric evaluation of the Hessian contributions.

Unlike classical forward-over-adjoint (FOA) techniques [9,19] to compute second order derivative information, the symmetric property when evaluating a Hessian matrix can be exploited as

☆ This research was supported by Research Council KUL: PFV/10/002 Optimization in Engineering Center OPTEC, Belgian Federal Science Policy Office: IUAP P7 (DYSCO, Dynamical systems, control and optimization, 2012-2017); EU: FP7-TEMPO (MC ITN-607957), H2020-ITN AWESCO (642682), ERC HIGHWIND (259 166), by the DFG in context of the research unit FOR 2401, National Science Foundation China (NSFC), nr. 61473185, and ShanghaiTech University, grant nr. F-0203-14-012. R. Quirynen holds a PhD fellowship of the Research Foundation – Flanders (FWO).

* Corresponding author.
*E-mail address:* rien.quirynen@esat.kuleuven.be (R. Quirynen).

http://dx.doi.org/10.1016/j.jprocont.2016.11.011
0959-1524/© 2016 Elsevier Ltd. All rights reserved.

discussed in [9,20] for explicit function evaluations. Following this research, a symmetric variant of AD was presented in [10] for explicit integration schemes in the context of exact Hessian based direct optimal control. Later in [21], this symmetric scheme has been extended to the case of an implicit integration method using the implicit function theorem in a discrete-time propagation technique. Similar to [10,21], the present article is concerned with the computation of Hessian matrices as required by Newton type optimization. The work by [22] instead proposes a *differentiate-then-discretize* type approach to compute Hessian vector products directly in the context of truncated Newton (TN) methods. In case of path-constrained optimal control problems, one could additionally use composite adjoints as discussed in [23].

### 1.1. Motivation and contributions

This paper proposes an efficient first and second order algorithmic differentiation scheme for both discrete- and continuous-time optimal control problems. Unlike the initial results from [10,21], this article presents and establishes the correctness of the symmetric Hessian propagation technique for any explicit or implicit integration method. The resulting second order sensitivity analysis typically allows for a computational speedup of about factor 2, by exploiting the symmetry of the Hessian. In addition, we present an extension of these results to continuous-time sensitivity propagation for an implicit system of differential algebraic equations (DAE) of index 1. This discussion in a continuous-time framework allows for a generic sensitivity analysis, before applying a numerical discretization scheme.

Based on the symmetric sensitivity equations, a resulting three-sweep Hessian propagation (TSP) scheme has been proposed [10]. This technique is studied here both in discrete- and continuous-time, and shown to considerably reduce the memory requirements, in addition to the reduced computational burden, over classical forward-over-adjoint (FOA) approaches. The proposed TSP scheme avoids the need to store a trajectory of forward sensitivities based on the symmetric sensitivity equations, which generally results in a much smaller memory footprint to compute the Hessian. An implementation of these symmetric Hessian propagation techniques in the open-source ACADO Toolkit software is presented and its performance is illustrated on the numerical case study of a nonlinear biochemical reactor.

### 1.2. Notation and preliminaries

This paper denotes first order total and partial derivatives, respectively using the compact notation $\mathrm{D}_a F(a,b) = \frac{\mathrm{d}F(a,b)}{\mathrm{d}a}$ and $\partial_a F(a,b) = \frac{\partial F(a,b)}{\partial a}$. In addition, let us write the second order directional derivatives:

$$\langle c, \mathrm{D}^2_{a,b} F(a,b)\rangle = \sum_{k=1}^{n} c_k \frac{\mathrm{d}^2 F_k(a,b)}{\mathrm{d}a\,\mathrm{d}b}, \tag{1}$$

where $c \in \mathbb{R}^n$ is a constant vector and $F : \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}^n$ a twice differentiable function. Notice that the map $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^{n\times l\times m} \to \mathbb{R}^{l\times m}$ does not denote a standard scalar product, since the first argument is a vector while the second argument is a tensor. We occasionally use the shorthand notation $\langle c, \mathrm{D}^2_{a,b} F\rangle$, if it is clear from the context that $F$ depends on $a$ and $b$. We write $\langle c, \mathrm{D}^2_a F(a)\rangle$ rather than $\langle c, \mathrm{D}^2_{a,a} F(a)\rangle$, in case $F$ has only one argument. If the second order derivatives of $F$ are continuous, the matrix $\langle c, \mathrm{D}^2_a F(a)\rangle$ is symmetric. When using partial instead of total derivatives, a similar compact notation for the directional second order derivatives is adopted:

$$\langle c, \partial^2_{a,b} F(a,b)\rangle = \sum_{k=1}^{n} c_k \frac{\partial^2 F_k(a,b)}{\partial a\,\partial b}. \tag{2}$$

The article is organized as follows. Section 2 introduces direct optimal control in a simplified setting in order to illustrate the need for efficient sensitivity analysis. Section 3 discusses discrete-time propagation techniques for a generic implicit integration method. We first present the classical first and second order techniques, then we propose and motivate our alternative symmetric propagation scheme. Section 4 then presents the continuous-time extension of these sensitivity equations, considering an implicit DAE system of index 1. A three-sweep Hessian propagation technique is introduced and discussed in Section 5, including implementation details. The open-source ACADO code generation software using these novel sensitivity propagation techniques is briefly discussed in Section 6. Numerical results for an illustrative case study are finally presented and discussed further in Section 7.

## 2. Problem statement

Let us briefly introduce the problem formulation in which we are interested, including the system of differential algebraic equations (DAE). We then introduce the framework of direct optimal control and Newton type algorithms to solve the resulting optimization problem, using first and second order sensitivity analysis.

### 2.1. Differential algebraic equations

Let us consider the following semi-explicit DAE system:

$$\dot{x}(t) = f(x(t), z(t)), \quad x(0) = x_0(p), \quad 0 = g(x(t), z(t)), \tag{3}$$

in which $x(t) \in \mathbb{R}^{n_x}$ denotes the differential states, $z(t) \in \mathbb{R}^{n_z}$ the algebraic variables and $f : \mathbb{R}^{n_x} \times \mathbb{R}^{n_z} \to \mathbb{R}^{n_x}$, $g : \mathbb{R}^{n_x} \times \mathbb{R}^{n_z} \to \mathbb{R}^{n_z}$. The parameters $p \in \mathbb{R}^{n_p}$ are additional variables which define the initial value function $x_0 : \mathbb{R}^{n_p} \to \mathbb{R}^{n_x}$. The latter DAE system is of index 1 [17] if the Jacobian $\partial_z g(\cdot)$ is non-singular. In case there are no algebraic variables, the system instead denotes a set of ordinary differential equations (ODE):

$$\dot{x}(t) = f_{\mathrm{ODE}}(x(t)), \quad x(0) = x_0(p). \tag{4}$$

We introduce the following two important assumptions.

**Assumption 1.** The functions $f(x(t), z(t))$, $g(x(t), z(t))$ and $x_0(p)$ are twice continuously differentiable in all arguments.

**Assumption 2.** The semi-explicit DAE system in Eq. (3) has differential index 0 or index 1, which means that either $n_z = 0$, or the Jacobian matrix $\partial_z g(\cdot)$ must be invertible.

Finally, let us refer to the following definition for consistent initial conditions such that the initial value problem in Eq. (3) has a unique solution $x(t, p)$ and $z(t, p)$ $\forall t \in [0, T]$, $p$ given the previous two assumptions [14,24].

**Definition 3.** The initial values $(x(0, p), z(0, p))$ are called consistent when the following conditions hold:

$$\begin{aligned} x(0,p) &= x_0(p) \\ 0 &= g(x(0,p), z(0,p)). \end{aligned} \tag{5}$$

This is a well defined nonlinear system in $(x(0, p), z(0, p))$ given the parameter values $p$ and an index 1 DAE.