

Direct Multiple Shooting for Nonlinear Optimum Experimental Design

Dennis Janka, Stefan Körkel, and Hans Georg Bock

Abstract Optimum experimental design (OED) for parameter identification has become a key technique in the model validation process for dynamical systems. This paper deals with optimum experimental design for systems modelled by differential-algebraic equations. We show how to formulate OED as a nonstandard nonlinear optimal control problem. The direct multiple shooting method is a state of the art method for the solution of standard optimal control problems that leads to structured nonlinear programs. We present two possibilities how to adapt direct multiple shooting to OED by introducing additional variables and constraints. We highlight special structures in the constraint and objective derivatives whose evaluation is usually the bottleneck when solving dynamic optimization problems by multiple shooting. We have implemented a structure exploiting algorithm that takes all these structures into account. Two benchmark examples show the efficiency of the new algorithm.

1 Introduction

Many processes in engineering, chemistry, or physics can be described by dynamical systems given by ordinary differential equations (ODEs) or differential-algebraic equations (DAEs). These equations usually depend on *model parameters*, for example material specific constants that are not directly accessible by measurements. However, as the models are often highly nonlinear, simulation results can vary strongly depending on the values of the model parameters. Thus it is desirable to estimate the model parameters with high precision. This process is called *model validation*. Only a validated model can be used to make meaningful predictions.

The first step in the model validation process is usually a parameter estimation. That means the model is fitted to given measurement data, yielding a first estimate

D. Janka (✉) • S. Körkel • H.G. Bock
Interdisciplinary Center for Scientific Computing, Heidelberg University, Im Neuenheimer Feld
368, 69120 Heidelberg, Germany
e-mail: dennis.janka@iwr.uni-heidelberg.de; stefan.koerkel@iwr.uni-heidelberg.de;
SciCom@iwr.uni-heidelberg.de

for the parameters. Then one can perform a sensitivity analysis to obtain estimates for the covariance matrix of the parameters. The covariance matrix can reveal large uncertainties in the parameter values or correlations between different parameters. In this context, it is also possible to quantify the uncertainty of arbitrary model quantities of interest.

An important observation is that the covariance matrix depends not only on the parameters but also on the experimental conditions. This leads to the task of *optimum experimental design (OED)*: Choose experimental conditions such that a subsequent parameter estimation yields parameters with minimum uncertainty. The uncertainty of the vector of parameters is characterized by a functional on the predicted covariance matrix. OED for general statistical models has been studied for several decades and it is a well-established field of research, see the textbooks [2, 10, 18]. Nonlinear OED for processes modeled by differential equations has been investigated by several authors, see, e.g., [4, 12, 15] for an overview.

In mathematical terms, optimum experimental design can be cast as a special (nonstandard) type of optimal control (OC) problem. As the objective, namely the functional on the covariance matrix, depends on first-order sensitivities of the states, *variational differential equations* or *sensitivity equations* must be explicitly included in the problem formulation, leading to large, but specially structured differential equation systems. We are interested in *direct* methods for OED problems, in particular direct multiple shooting, that transform the infinite dimensional optimal control problem into a finite dimensional nonlinear programming problem (NLP). The direct multiple shooting method for optimal control problems as described by Bock and Plitt [7] makes use of the partial separability of the objective function. This leads to a block-diagonal Hessian of the Lagrangian that can and should be exploited by Newton-type methods. However, a straightforward formulation of the OED objective function lacks the feature of partial separability, so special care must be taken to reformulate the OED problem as a standard optimal control problem. In [14, 16] direct multiple shooting has been applied to OED. In [13], a collocation discretization is applied to a similar problem. The main contribution of this paper consists in detailed descriptions of the structured NLPs that result from a multiple shooting discretization as well as numerical results that demonstrate their benefits and limitations.

The paper is organized as follows: In Sect. 2, we give an introduction to optimum experimental design and formulate it as a nonstandard optimal control problem. In Sect. 3, the direct multiple shooting method is described for standard optimal control problems. Afterwards, in Sect. 4 we propose two ways how to transform the OED problem to a specially structured standard OC problem. In Sect. 5 we show how to efficiently evaluate the constraint Jacobian as well as the gradient and Hessian of the objective. A numerical example from chemical engineering illustrates the efficacy of the approach in Sect. 6. Section 7 concludes.

2 Nonlinear Optimum Experimental Design for Parameter Estimation

Optimum experimental design aims at improving parameter estimation results in a statistical sense. We first introduce the type of parameter estimation problems whose results we seek to improve with OED along with some general notation that we use throughout the paper.

2.1 Parameter Estimation for Dynamical Systems

We consider a dynamical process on a fixed time horizon $[t_0, t_f]$ that is described by the following differential-algebraic equation (DAE) system with interior point and boundary conditions:

$$\dot{y}(t) = f(t, y(t), z(t), p, u(t)), \quad y(t_0) = y_0(p, u) \quad (1a)$$

$$0 = g(t, y(t), z(t), p, u(t)) \quad (1b)$$

$$0 = \sum_{i=1}^{N_r} r_i(t_i, y(t_i), z(t_i), p) \quad (1c)$$

where

$$y(t) \in \mathbb{R}^{n_y} \quad (\text{differential states})$$

$$z(t) \in \mathbb{R}^{n_z} \quad (\text{algebraic states})$$

$$p \in \mathbb{R}^{n_p} \quad (\text{parameters})$$

$$u(t) \in \mathbb{R}^{n_u} \quad (\text{controls})$$

$$r_i(t_i, y(t_i), z(t_i), p) \in \mathbb{R}^{n_r} \quad (\text{boundary conditions}).$$

Throughout the text we will use the notation $x(t) = (y(t)^T, z(t)^T)^T$ to denote both differential and algebraic states.

Assume measurement data $\eta_1, \dots, \eta_{N_M}$ are available at sampling times t_1, \dots, t_{N_M} such that

$$h_i(t_i, y(t_i), z(t_i), p^*) = \eta_i + \varepsilon_i, \quad i = 1, \dots, N_M,$$

where p^* are the true—but inaccessible—parameters, y and z the corresponding states, and ε_i are independently normally distributed with zero mean and standard

deviation σ_i . This assumption states that the model is structurally correct and errors arise only due to inaccuracies in the measurement process. We call

$$h_i(t, y(t), z(t), p), \quad i = 1, \dots, N_M$$

the *model response* or *observable*.

The maximum likelihood parameter estimation problem can be stated as:

$$\min_{y_0, x, p} \sum_{i=1}^{N_m} \left(\frac{h_i(t_i, y(t_i), z(t_i), p) - \eta_i}{\sigma_i} \right)^2 \quad (2a)$$

$$\text{s.t. } \dot{y}(t) = f(t, y(t), z(t), p, u(t)), \quad y(t_0) = y_0(p, u) \quad (2b)$$

$$0 = g(t, y(t), z(t), p, u(t)) \quad (2c)$$

$$0 = \sum_{i=1}^{N_r} r_i(t_i, y(t_i), z(t_i), p) \quad (2d)$$

Parameter estimation problems constrained by differential equations can be solved by different approaches, e.g. by direct multiple shooting in combination with a generalized Gauss-Newton method, see [6].

2.2 Sensitivity Analysis

The solution \hat{p} of the parameter estimation problem (2) is a random variable due to the fact that the measurements η_i are random. The variance-covariance matrix C of \hat{p} is given by:

$$C = (I \ 0) \begin{pmatrix} \mathcal{J}_1^T & \mathcal{J}_1 & \mathcal{J}_2^T \\ \mathcal{J}_2 & & 0 \end{pmatrix}^{-1} \begin{pmatrix} I \\ 0 \end{pmatrix} \quad (3)$$

where $\mathcal{J}_1 \in \mathbb{R}^{N_m \times n_p}$ and $\mathcal{J}_2 \in \mathbb{R}^{n_r \times n_p}$ are the Jacobians of the residual vectors of the parameter estimation problem. We denote by $\mathcal{J}_{1,i}$ the rows of \mathcal{J}_1 and by $\mathcal{J}_{2,i}$ the summands that make up \mathcal{J}_2 :

$$\mathcal{J}_{1,i} = \left(\frac{\sqrt{w_i}}{\sigma_i} \left(\frac{\partial h_i}{\partial x}(t_i, y(t_i), z(t_i), p) \frac{\partial x}{\partial p}(t_i) + \frac{\partial h_i}{\partial p}(t_i, y(t_i), z(t_i), p) \right) \right)_{i=1, \dots, N_m} \quad (4)$$

$$\mathcal{J}_2 = \sum_{i=1}^{N_r} \mathcal{J}_{2,i}, \quad \mathcal{J}_{2,i} = \frac{\partial r_i}{\partial x}(t_i, y(t_i), z(t_i), p) \frac{\partial x}{\partial p}(t_i) + \frac{\partial r_i}{\partial p}(t_i, y(t_i), z(t_i), p). \quad (5)$$

We assume that \mathcal{J}_2 has full rank and that $\mathcal{J}_1^T \mathcal{J}_1$ is positive definite on $\text{Ker } \mathcal{J}_2$ which implies existence of C .

In (4) we have also introduced *measurement weights* $w_i \in \{0, 1\}$, $i = 1, \dots, N_m$ for each measurement time. They are fixed in the parameter estimation context but will be design variables in the experimental design where they allow us to select or de-select measurements.

The sensitivities of the states x with respect to the parameters p are subject to the following *variational differential-algebraic equations (VDAE)*, also called *sensitivity equations*:

$$\dot{y}_p(t) = \frac{\partial f}{\partial x}(t, y(t), z(t), p, u(t))x_p(t) + \frac{\partial f}{\partial p}(t, y(t), z(t), p, u(t)) \quad (6)$$

$$0 = \frac{\partial g}{\partial x}(t, y(t), z(t), p, u(t))x_p(t) + \frac{\partial g}{\partial p}(t, y(t), z(t), p, u(t)), \quad (7)$$

where

$$x_p(t) = \frac{\partial x}{\partial p}(t) = \left(\frac{\partial y}{\partial p}(t), \frac{\partial z}{\partial p}(t) \right).$$

Initial values for the VDAE are given by

$$y_p(t_0) = \frac{\partial y_0}{\partial p}$$

for the variational differential states and by (7) for the variational algebraic states. Note that (6) and (7) depend on $y(t)$ and $z(t)$ and therefore have to be solved together with (1a) and (1b).

2.3 The Optimum Experimental Design Problem

Based on the sensitivity analysis, we can predict the variance-covariance matrix for different experimental settings that are characterized by controls $u(t)$ as well as a choice of measurements. An experiment may also be constrained by external process constraints, e.g. safety or cost constraints.

The task of optimum experimental design is to choose experimental settings such that the predicted covariance matrix has the best properties in some statistical sense. The quality of the matrix is measured by a criterion ϕ from statistical experimental design:

- A-criterion: $\phi = \text{tr } C$
- D-criterion: $\phi = \det C$
- E-criterion: $\phi = \max\{\lambda_i, i = 1, \dots, n_p, \lambda_i \text{ eigenvalue of } C\} = \|C\|_2$
- M-criterion: $\phi = \max\{C_{ii}, i = 1, \dots, n_p\}$

The complete optimum experimental design problem is

$$\min_{y_0, x_p, u, w} \phi \left((I \ 0) \begin{pmatrix} \mathcal{J}_1^T & \mathcal{J}_1 & \mathcal{J}_2^T \\ \mathcal{J}_2 & & 0 \end{pmatrix}^{-1} \begin{pmatrix} I \\ 0 \end{pmatrix} \right) \quad (8a)$$

$$\text{s.t. } \dot{y}(t) = f(t, y(t), z(t), p, u(t)), \quad t \in [t_0, t_f] \quad (8b)$$

$$0 = g(t, y(t), z(t), p, u(t)), \quad t \in [t_0, t_f] \quad (8c)$$

$$y(t_0) = y_0(p, u) \quad (8d)$$

$$\dot{y}_p(t) = \frac{\partial f}{\partial x}(t, y(t), z(t), p, u(t))x_p(t) + \frac{\partial f}{\partial p}(t, y(t), z(t), p, u(t)), \quad t \in [t_0, t_f] \quad (8e)$$

$$0 = \frac{\partial g}{\partial x}(t, y(t), z(t), p, u(t))x_p(t) + \frac{\partial g}{\partial p}(t, y(t), z(t), p, u(t)), \quad t \in [t_0, t_f] \quad (8f)$$

$$y_p(t_0) = \frac{\partial y_0}{\partial p} \quad (8g)$$

$$0 = \sum_{i=1}^{N_r} r_i(t_i, y(t_i), z(t_i), p) \quad (8h)$$

$$0 \leq c(t, y(t), z(t), u(t), w), \quad t \in [t_0, t_f] \quad (8i)$$

$$w_i \in \{0, 1\}, \quad i = 1, \dots, N_m \quad (8j)$$

$$\mathcal{J}_{1,i} = \left(\frac{\sqrt{w_i}}{\sigma_i} \left(\frac{\partial h_i}{\partial x}(t_i, y(t_i), z(t_i), p)x_p(t_i) + \frac{\partial h_i}{\partial p}(t_i, y(t_i), z(t_i), p) \right) \right)_{i=1, \dots, N_m} \quad (8k)$$

$$\mathcal{J}_2 = \sum_{i=1}^{N_r} \frac{\partial r_i}{\partial x}(t_i, y(t_i), z(t_i), p)x_p(t_i) + \frac{\partial r_i}{\partial p}(t_i, y(t_i), z(t_i), p) \quad (8l)$$

with the nominal DAE system (8b) and (8c) with initial values (8d), the variational DAE system (8e) and (8f) with initial values (8g), multipoint boundary constraints from the parameter estimation problem (8h), path and control constraints (8i), and integrality constraints for the measurement weights (8j). The Jacobians of the parameter estimation residuals (8k) and (8l) are given to define the covariance matrix on which a functional ϕ is minimized (8a). Note that while initial values for the nominal differential states (8b) may be degrees of freedom in the optimization, initial values for the variational differential states (8e) are explicitly defined by the relation (8g) and for the algebraic states $z(t)$ and $z_p(t)$ they are implicitly defined by the algebraic conditions (8c) and (8f).

3 The Direct Multiple Shooting Method for Optimal Control Problems

The direct multiple shooting method for optimal control problems has been first introduced in [7]. Let us first consider the standard optimal control problem

$$\min_{\tilde{y}_0, \tilde{x}, u} \Phi(\tilde{x}(t_f)) \quad (9a)$$

$$\text{s.t. } \dot{\tilde{y}}(t) = \tilde{f}(t, \tilde{y}(t), \tilde{z}(t), \tilde{u}(t)), \quad \tilde{y}(t_0) = \tilde{y}_0 \quad (9b)$$

$$0 = \tilde{g}(t, \tilde{y}(t), \tilde{z}(t), \tilde{u}(t)) \quad (9c)$$

$$0 \leq \tilde{c}(t, \tilde{y}(t), \tilde{z}(t), \tilde{u}(t)) \quad (9d)$$

$$0 \leq \sum_{i=1}^{N_f} \tilde{r}_i(t_i, \tilde{y}(t_i), \tilde{z}(t_i)). \quad (9e)$$

In direct methods the infinite-dimensional optimal control problem (9) is approximated by a nonlinear programming problem (NLP) which is then solved by suitable numerical methods. The following infinite-dimensional objects of the optimal control problem must be treated adequately when setting up the finite-dimensional NLP:

- control functions \tilde{u}
- differential and algebraic states \tilde{y} and \tilde{z}
- path constraints $0 \leq \tilde{c}(t, \tilde{y}(t), \tilde{z}(t), \tilde{u}(t))$

3.1 Control Functions

We consider a time grid

$$t_0 = \tau_0^c < \tau_1^c < \dots < \tau_{N_c}^c = t_f \quad (10)$$

on which the control function $\tilde{u}(\cdot)$ is parameterized by means of local basis functions:

$$\tilde{u}(t) = \varphi^j(t, q^j), \quad t \in [\tau_j^c, \tau_{j+1}^c],$$

where the $q^j \in \mathbb{R}^{n_u}$ are vectors of finitely many real optimization variables. We define $q := (q^0, \dots, q^{N_c-1})^T$. The local functions φ^j are typically polynomials of low degree, e.g. linear or constant functions.

3.2 States

In shooting methods, initial value problem solvers are employed to obtain representations of the states x for given q and y_0 . In the case of direct single shooting and pure ODEs, the states are regarded as dependent variables, and only q and y_0 are kept as variables in the optimization problem. Thus the tasks of simulation and optimization are kept separate.

The direct multiple shooting method for DAEs is a simultaneous strategy to resolve simulation and optimization in parallel. Again, we consider a discretization of the time horizon

$$t_0 = \tau_0^s < \tau_1^s < \dots < \tau_{N_s}^s = t_f \quad (11)$$

where we assume without loss of generality the grid points to be a subset of the grid points of the control grid (10). On this shooting grid we consider the following set of initial value problems with initial values $s_x^j = (s_y^j, s_z^j)$ that become variables in the optimization problem:

$$\dot{\tilde{y}}(t) = \tilde{f}(t, \tilde{y}(t), \tilde{z}(t), p, \tilde{u}(t)) \quad \tilde{y}(\tau_j^s) = s_y^j \quad (12a)$$

$$0 = \tilde{g}(t, \tilde{y}(t), \tilde{z}(t), p, \tilde{u}(t)) - \theta_j(t) \tilde{g}(\tau_j^s, s_y^j, s_z^j, p, \tilde{u}(t)) \quad \tilde{z}(\tau_j^s) = s_z^j, \quad (12b)$$

where $\theta_j(\cdot)$ is a fast decreasing damping function with $\theta(\tau_j^s) = 1$. This relaxed formulation was proposed in [8] and means that the algebraic condition (12b) is automatically consistent for any initial values s_z^j . That means the DAE solver does not need to solve the (nonlinear) algebraic condition in every iteration of the optimization algorithm to find feasible initial values. Instead, the nonlinear algebraic consistency conditions

$$0 = \tilde{g}(\tau_j^s, s_y^j, s_z^j, \hat{q}^j), \quad j = 0, \dots, N_s$$

are added to the optimization problem which ensures the solution of the original DAE at the solution of the optimization problem.

Note that the DAEs (12) are solved independently on the smaller time intervals $[\tau_j^s, \tau_{j+1}^s]$ as the initial values s_x^j are variables of the optimization problem. To ensure equivalence to the original system (1), continuity conditions are added to the optimization problem for every shooting interval. Let us denote by $\tilde{y}(\tau_{j+1}^s; s_y^j, s_z^j, \hat{q}^j)$ a representation of the solution to problem (12) on the intervals $[\tau_j^s, \tau_{j+1}^s]$, where \hat{q}^j denotes the subvector of q that represents \tilde{u} on the interval $[\tau_j^s, \tau_{j+1}^s]$. Then the continuity conditions read as:

$$\tilde{y}(\tau_{j+1}^s; s_y^j, s_z^j, \hat{q}^j) = s_y^{j+1}, \quad j = 0, \dots, N_s - 1.$$

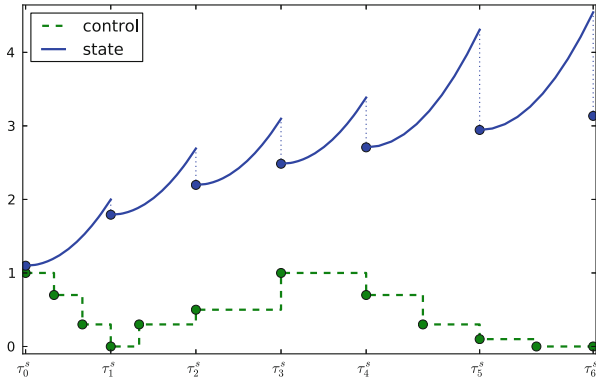


Fig. 1 Concept of direct multiple shooting for one state and one piecewise constant control. The continuity conditions are violated (*vertical dotted lines*). Note how the control is also allowed to switch within shooting intervals

Figure 1 illustrates the concept of direct multiple shooting. Note that we explicitly maintain separate grids for controls and states. A special case is of course to choose the same grid for both. However, in our experience, the decoupling of grids provides greater flexibility and a smaller number of shooting intervals can greatly accelerate convergence for problems where a relatively fine discretization of the controls is desirable.

3.3 Path Constraints

All path constraints such as (9d) that are required to hold at infinitely many points are evaluated on finitely many checkpoints only. Let us assume—without loss of generality—that the checkpoints are the grid points of the multiple shooting grid (10). Then the discretized path constraints read as

$$0 \leq \tilde{c}(\tau_j^s, s_y^j, s_z^j, \hat{q}^j), \quad j = 0, \dots, N_s. \tag{13}$$

Depending on the choice of the time grid, the constraints might be violated in between grid points. There exist strategies how to adaptively add checkpoints, see, e.g., [17], but to keep the notation simple we assume for the scope of this paper that they match the grid points of the shooting grid.

3.4 Structured NLP

We have now addressed all constraints of the optimal control problem and can formulate the structured multiple shooting NLP as follows:

$$\min_{s_y, s_z, q} \phi(s_x^N) \quad (14a)$$

$$\text{s.t. } 0 = \tilde{y}(\tau_0^s) - s_y^0 \quad (14b)$$

$$0 = \tilde{y}(\tau_{j+1}^s; s_y^j, s_z^j, \hat{q}^j) - s_y^{j+1}, \quad j = 0, \dots, N_s - 1 \quad (14c)$$

$$0 = \tilde{g}(\tau_j^s, s_y^j, s_z^j, \hat{q}^j), \quad j = 0, \dots, N_s \quad (14d)$$

$$0 \leq c(\tau_j^c, \tilde{y}(\tau_j^c), \tilde{z}(\tau_j^c), \hat{q}^j), \quad j = 0, \dots, N_s \quad (14e)$$

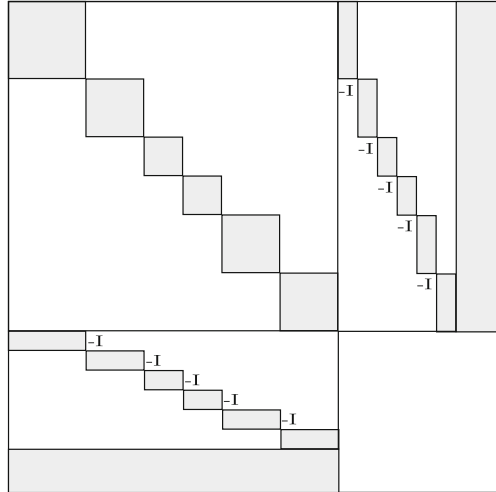
$$0 \leq \sum_{j=0}^{N_s} \sum_{\substack{i, \\ \tau_j \leq t_i < \tau_{j+1}}}^{N_r} \tilde{r}_i(t_i, \tilde{y}(t_i; s_y^j, s_z^j, \hat{q}^j), \tilde{z}(t_i; s_y^j, s_z^j, \hat{q}^j)), \quad (14f)$$

where (14c) and (14d) are the continuity and consistency conditions that guarantee the solution of the original DAE (1) at the solution of the optimization problem.

In Newton-type methods, the Jacobian of the constraints and the Hessian of the Lagrangian are of special importance. It is clear that the evaluation of the continuity and consistency constraints with index j only depend on variables s_x^j and \hat{q}^j in a nonlinear way. This leads to a constraint Jacobian that has a banded structure and a Hessian of the Lagrangian with a block diagonal structure according to the shooting discretization. These structures can be seen in the structure of the KKT matrix as depicted in Fig. 2.

Depending on the shooting discretization, problems of type (14) can be very large, but sparse. Algorithmic techniques such as condensing (see [7]) exploit this sparsity and reduce the additional effort considerably that is caused by the larger matrices when using a fine multiple shooting discretization.

Fig. 2 Sparsity pattern of the KKT matrix of a multiple shooting discretized optimal control problem. The constraint Jacobian comprises continuity constraints that are responsible for the banded structure. Linearly coupled constraints give rise to a dense block. The Hessian of the Lagrangian (*upper left*) has block diagonal structure. Different block sizes may occur if the numbers of control variables on two intervals differ



4 Optimum Experimental Design as Separable NLP

We now want to apply the multiple shooting discretization as described in the previous section to the optimum experimental design problem (8). In particular we need to extend the problem formulation (14) to cope with the special kind of coupled objective that is characteristic for OED. Multiple Shooting for OED problems has first been applied in [16] and has been further investigated in [14].

4.1 Measurements

The grid of possible measurements depends on the process and should be independent of the shooting and control grid. In particular, more than one measurement could be taken at the same time, see [15].

In the original formulation, integrality of the measurement weights is required. In our formulation we employ a continuous relaxation:

$$0 \leq w_i \leq 1, \quad i = 1, \dots, N_m$$

In practice, this often yields satisfactory results, as a bang-bang structure is observed for the measurements and so integrality is satisfied automatically. In fact there is also some theoretical evidence for this, see [20].

The measurement weights, along with the controls, are experimental design variables. All simple bounds and linear constraints on the measurement weights fit into the framework of general path constraints and linearly coupled interior point constraints (9d) and (9e).

4.2 Dynamical System

We combine nominal and variational states into one system:

$$\tilde{y}(t) = \begin{cases} y(t) \\ \downarrow y_p(t) \downarrow \end{cases} \in \mathbb{R}^{n_y+n_y \cdot n_p}$$

for the differential states and

$$\tilde{z}(t) = \begin{cases} z(t) \\ \downarrow z_p(t) \downarrow \end{cases} \in \mathbb{R}^{n_z+n_z \cdot n_p}$$

for the algebraic states, where we denote by $\downarrow \cdot \downarrow$ the map that combines the columns of an $m \times n$ matrix into a single $m \cdot n$ column vector by stacking them one below the other.

That leaves us with the new DAE system

$$\dot{\tilde{y}}(t) = \tilde{f}(t, \tilde{y}, \tilde{z}, p, u) = \begin{cases} f(t, y(t), z(t), p, u(t)) \\ \downarrow \frac{\partial f}{\partial x}(t, y(t), z(t), p, u(t))x_p(t) + \frac{\partial f}{\partial p}(t, y(t), z(t), p, u(t)) \downarrow \end{cases} \quad (15)$$

$$0 = \tilde{g}(t, \tilde{y}, \tilde{z}, p, u) = \begin{cases} g(t, y(t), z(t), p, u(t)) \\ \downarrow \frac{\partial g}{\partial x}(t, y(t), z(t), p, u(t))x_p(t) + \frac{\partial g}{\partial p}(t, y(t), z(t), p, u(t)) \downarrow \end{cases} \quad (16)$$

This system has of course a special structure that can and should be exploited in an efficient implementation. We will give details on this in Sect. 5.

4.3 Objective Function

An important difference between the problem of optimum experimental design (8) and the standard optimal control problem (9) is the nonlinear coupling in time in the objective that is due to the inversion when computing the covariance matrix, as it has been noted in [16]. In particular this violates the property of partial separation of the Lagrange function that is responsible for its sparse, block-diagonal Hessian. We present two approaches to resolve that nonlinear coupling.

4.3.1 Linearly Coupled Constraint for Information Matrix

Recall the definition of the covariance matrix:

$$C = (I \ 0) \begin{pmatrix} \mathcal{J}_1^T \mathcal{J}_1 & \mathcal{J}_2^T \\ \mathcal{J}_2 & 0 \end{pmatrix}^{-1} \begin{pmatrix} I \\ 0 \end{pmatrix}$$

with the rows of \mathcal{J}_1 and the summands that constitute \mathcal{J}_2 as defined by (4) and (5):

$$\mathcal{J}_{1,i} = \left(\frac{\sqrt{w_i}}{\sigma_i} \left(\frac{\partial h_i}{\partial x} x_p(t_i) + \frac{\partial h_i}{\partial p} \right) \right)_{i=1, \dots, N_m}$$

$$\mathcal{J}_{2,i} = \frac{\partial r_i}{\partial x}(t_i, y(t_i), z(t_i), p) x_p(t_i) + \frac{\partial r_i}{\partial p}(t_i, y(t_i), z(t_i), p).$$

In [16] it has been pointed out that

$$\begin{pmatrix} \mathcal{J}_1^T \mathcal{J}_1 & \mathcal{J}_2^T \\ \mathcal{J}_2 & 0 \end{pmatrix} = \begin{pmatrix} \sum_{i=0}^{N_m} \mathcal{J}_{1,i}^T \mathcal{J}_{1,i} & \sum_{i=0}^{N_r} \mathcal{J}_{2,i}^T \\ \sum_{i=0}^{N_r} \mathcal{J}_{2,i} & 0 \end{pmatrix}. \quad (17)$$

Note that in particular, $\mathcal{J}_{1,i}$ and $\mathcal{J}_{2,i}$ depend on evaluations of the nominal and variational states x and x_p at individual points t_i . Thus, (17) implies that the matrices $\mathcal{J}_1^T \mathcal{J}_1$ and \mathcal{J}_2 only exhibit a linear coupling in time. In a multiple shooting context, we assign the points t_i to the proper shooting intervals and plug in the representation of the solution $x(\tau_{j+1}^s; s_y^j, s_z^j, \hat{q}^j)$ and $x_p(\tau_{j+1}^s; s_y^j, s_z^j, \hat{q}^j)$, respectively. We write this as

$$\mathcal{J}_1^T \mathcal{J}_1 = \sum_{j=0}^{N_s-1} \sum_{\tau_j < t_i \leq \tau_{j+1}}^{N_m} \mathcal{J}_{1,i}(s_y^j, s_z^j, \hat{q}^j, w_i)^T \mathcal{J}_{1,i}(s_y^j, s_z^j, \hat{q}^j, w_i)$$

$$\mathcal{J}_2 = \sum_{j=0}^{N_s-1} \sum_{\tau_j < t_i \leq \tau_{j+1}}^{N_r} \mathcal{J}_{2,i}(s_y^j, s_z^j, \hat{q}^j).$$

We introduce additional variables H and J and linearly coupled constraints that fit into the framework of (14). The objective then only depends on the newly introduced variables H and J and we obtain the following structured NLP:

$$\min_{s_y, s_z, q, w, H, J} \phi \left((I \ 0) \begin{pmatrix} H & J^T \\ J & 0 \end{pmatrix}^{-1} \begin{pmatrix} I \\ 0 \end{pmatrix} \right) \quad (18a)$$

$$\text{s.t. } 0 = H - \sum_{j=0}^{N_s-1} \sum_{\tau_j < t_i \leq \tau_{j+1}}^{N_m} \mathcal{J}_{1,i}(s_y^j, s_z^j, \hat{q}^j, w_i)^T \mathcal{J}_{1,i}(s_y^j, s_z^j, \hat{q}^j, w_i) \quad (18b)$$

$$0 = J - \sum_{j=0}^{N_s-1} \sum_{\tau_j < t_i \leq \tau_{j+1}}^{N_r} \mathcal{J}_{2,i}(s_y^j, s_z^j, \hat{q}^j) \quad (18c)$$

$$0 = \tilde{y}(\tau_0^s; \hat{q}^0) - s_y^0 \quad (18d)$$

$$0 = \tilde{y}(\tau_{j+1}^s; s_y^j, s_z^j, \hat{q}^j) - s_y^{j+1}, \quad j = 0, \dots, N_s - 1 \quad (18e)$$

$$0 = g(\tau_j^s, s_y^j, s_z^j, \hat{q}^j), \quad j = 0, \dots, N_s \quad (18f)$$

$$0 \leq c(\tau_j^s, s_y^j, s_z^j, \hat{q}^j, w^j), \quad j = 0, \dots, N_s \quad (18g)$$

$$0 \leq w_i \leq 1, \quad i = 1, \dots, N_m \quad (18h)$$

$$0 \leq \sum_{j=0}^{N_s} \sum_{\tau_j \leq t_i < \tau_{j+1}}^{N_r} r_i(t_i, y(t_i; s_y^j, s_z^j, \hat{q}^j), z(t_i; s_y^j, s_z^j, \hat{q}^j)). \quad (18i)$$

4.3.2 Pseudo States for Covariance Matrix

Another possibility to resolve the coupling in the objective is to move the computation of the covariance to the constraints by deriving a recursion formula.

Using this formula, we introduce matrix-valued variables

$$C^j = \begin{pmatrix} C_1^j & C_2^{jT} \\ C_2^j & C_3^j \end{pmatrix} \in \mathbb{R}^{(n_p+n_r) \times (n_p+n_r)}, j = 1, \dots, N_s \quad (19)$$

and constraints for the recursion at the multiple shooting nodes and add them as additional variables to the NLP. This resembles the treatment of dynamical states in a multiple shooting method, hence we refer to C^j as pseudo states for the covariance matrix.

Let us first derive the formula in the unconstrained case: Let H^j denote the information matrix including all terms up to time τ_j . Then H^j is given as the sum of H^{j-1} and the information gain by measurements and constraint evaluations in the interval $(\tau_{j-1}, \tau_j]$:

$$H^{j+1} = H^j + \sum_{i: \tau_j < t_i \leq \tau_{j+1}} \mathcal{J}_{1,i}(s_y^j, s_z^j, \hat{q}^j, w_i)^T \mathcal{J}_{1,i}(s_y^j, s_z^j, \hat{q}^j, w_i), \quad (20)$$

where we start with some initial information given as a positive definite matrix H^0 , e.g., from previous experiments or from literature.

At every grid point τ_j the covariance matrix taking into account all measurements up to time τ_j is the inverse of H^j . From (20) we obtain as recursion formula for the covariance matrix:

$$C^0 = (H^0)^{-1} \quad (21)$$

$$C^{j+1} = \left((C^j)^{-1} + \sum_{i:\tau_j < t_i \leq \tau_{j+1}} \mathcal{J}_{1,i}(s_y^j, s_z^j, \hat{q}^j, w_i)^T \mathcal{J}_{1,i}(s_y^j, s_z^j, \hat{q}^j, w_i) \right)^{-1}. \quad (22)$$

Equation (22) can be simplified to

$$C^{j+1} = C^j \left(I + \sum_{i:\tau_j < t_i \leq \tau_{j+1}} \mathcal{J}_{1,i}(s_y^j, s_z^j, \hat{q}^j, w_i)^T \mathcal{J}_{1,i}(s_y^j, s_z^j, \hat{q}^j, w_i) C^j \right)^{-1}.$$

This can be easily generalized to the case of constrained parameter estimation problems and we obtain the complete NLP by replacing the coupled constraints (18b) and (18c) by the pseudo continuity constraints

$$0 = \left(\left(\begin{array}{cc} C_1^j & C_2^{jT} \\ C_2^j & C_3^j \end{array} \right)^{-1} + \sum_{i:\tau_j < t_i \leq \tau_{j+1}} \left(\begin{array}{cc} \mathcal{J}_{1i}^T \mathcal{J}_{1i} & \mathcal{J}_{2i}^T \\ \mathcal{J}_{2i} & 0 \end{array} \right)^{-1} - \left(\begin{array}{cc} C_1^{j+1} & C_2^{j+1T} \\ C_2^{j+1} & C_3^{j+1} \end{array} \right) \right), \quad (23)$$

$$j = 0, \dots, N-1.$$

As initial values for the pseudo states we take some a priori uncertainty given by a positive definite matrix C_1^0 and a full rank matrix C_2^0 such that $C_2^0 C_2^{0T}$ is positive definite to ensure invertibility of C^0 . The a priori uncertainty should be chosen several orders of magnitude larger than the expected uncertainty after the OED to make sure that the choice of C_1^0 and C_2^0 does not interfere significantly with the solution. The objective for this formulation simplifies to

$$\phi \left((I \ 0) \left(\begin{array}{cc} C_1^{N_s} & C_2^{N_s T} \\ C_2^{N_s} & C_3^{N_s} \end{array} \right) \begin{pmatrix} I \\ 0 \end{pmatrix} \right). \quad (24)$$

While this formulation seems to be very much in the spirit of multiple shooting—the covariance of the system is modelled as a kind of state on each interval which is coupled via continuity-type constraints, the nonlinearity of the objective is distributed over the shooting intervals—it is computationally much less appealing: The objective (24) is simpler because the matrix inversion is now contained in the constraints, however, the pseudo continuity constraints (23) are numerically delicate, each containing two matrix inversions (in fact, we can reduce this to one by appropriate reformulation) of potentially ill-conditioned matrices.

5 Evaluation of Problem Functions

After discretization we need to solve the large but structured NLP (18). In derivative based methods such as sequential quadratic programming often the number of variables and constraints corresponds to the overall runtime of the method. In shooting methods for optimal control and especially optimum experimental design, however, the runtime is often dominated by the evaluation of states and their derivatives within the continuity constraints because they comprise possibly expensive calls to external numerical integrators. Thus it is worthwhile to have a closer look at the structures of constraints and objective and derive efficient and accurate evaluation schemes. We concentrate on the formulation (18) of the problem—linearly coupled constraints for the information matrix—as it is numerically more promising.

5.1 Constraint Derivatives

The OED problem (18) has a special structure in the constraints due to the fact that the dynamic system consists of closely related states, namely nominal and corresponding variational states. Ideally, they are evaluated together using the principle of *internal numerical differentiation (IND)*, see [1, 5].

Let us now have a closer look at the derivatives of continuity and consistency constraints (18e) and (18f) for the system that consists of nominal and variational states.

We denote by

$$s_{\tilde{x}}^j = \left(s_x^j, s_{x,p_1}^j, \dots, s_{x,p_{N_p}}^j \right)^T$$

the shooting variables for the nominal and variational states at one shooting node.

Observation 1. Variational states for different parameters are independent. This means

$$\frac{\partial x_{p_i}(\tau_{j+1})}{\partial s_{x,p_k}^j} = 0, \quad i \neq k.$$

Observation 2. By differentiating (6) and (7), we see that the derivative of a variational state with respect to its initial value satisfies the following $n_x \times n_x$ variational DAEs:

$$\left(\frac{\partial y_{p_i}(t)}{\partial s_{x,p_i}^j} \right) = \frac{\partial}{\partial x_{p_i}} \left(\frac{\partial f}{\partial x} x_{p_i}(t) + \frac{\partial f}{\partial p} \right) \cdot \frac{\partial x_{p_i}(t)}{\partial s_{x,p_i}^j} = \frac{\partial f}{\partial x} \cdot \frac{\partial x_{p_i}(t)}{\partial s_{x,p_i}^j}$$

$$0 = \frac{\partial}{\partial x_{p_i}} \left(\frac{\partial g}{\partial x} x_{p_i}(t) + \frac{\partial g}{\partial p} \right) \cdot \frac{\partial x_{p_i}(t)}{\partial s_{x,p_i}^j} = \frac{\partial g}{\partial x} \cdot \frac{\partial x_{p_i}(t)}{\partial s_{x,p_i}^j}.$$

These are the same equations that describes the sensitivity of the nominal states x with respect to their initial values and hence

$$\frac{\partial x_{p_i}(\tau_{j+1})}{\partial s_{x,p_i}^j} = \frac{\partial x(\tau_{j+1})}{\partial s_x^j}.$$

In particular that means that we do not need to evaluate any additional state sensitivities when we explicitly discretize the variational equations by multiple shooting. Instead $\frac{\partial x(\tau)}{\partial s_x}$ needs to be computed only once for each shooting interval and then can be used multiple times in the constraint Jacobian.

The part of the constraint Jacobian that corresponds to the derivative of the continuity constraints (18e) coupling the states starting at τ_{j-1} and at τ_j has the following structure:

$$\begin{pmatrix} \frac{\partial y(\tau_j)}{\partial s_y^{j-1}} & & \frac{\partial y(\tau_j)}{\partial s_z^{j-1}} & & \frac{\partial y(\tau_j)}{\partial q^{j-1}} & 0 \\ \frac{\partial y_{p_1}(\tau_j)}{\partial s_y^{j-1}} & \frac{\partial y(\tau_j)}{\partial s_y^{j-1}} & \frac{\partial y_{p_1}(\tau_j)}{\partial s_z^{j-1}} & \frac{\partial y(\tau_j)}{\partial s_z^{j-1}} & \frac{\partial y_{p_1}(\tau_j)}{\partial q^{j-1}} & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ \frac{\partial y_{p_{N_p}}(\tau_j)}{\partial s_y^{j-1}} & & \frac{\partial y(\tau_j)}{\partial s_y^{j-1}} & \frac{\partial y_{p_{N_p}}(\tau_j)}{\partial s_z^{j-1}} & \frac{\partial y(\tau_j)}{\partial s_z^{j-1}} & \frac{\partial y_{p_{N_p}}(\tau_j)}{\partial q^{j-1}} & 0 \end{pmatrix},$$

where the variable vector is ordered as follows:

$$\left(s_y^{j-1}, s_{y,p_1}^{j-1}, \dots, s_{y,p_{N_p}}^{j-1}, s_z^{j-1}, s_{z,p_1}^{j-1}, \dots, s_{z,p_{N_p}}^{j-1}, q^{j-1}, w^{j-1} \right)^T.$$

A similar structure can be observed for the consistency conditions (18f).

Note that the derivatives $\frac{\partial y}{\partial(\cdot)}$ are first-order and $\frac{\partial y_p}{\partial(\cdot)}$ are second-order sensitivities of the states and must be supplied by the integrator.

5.2 Objective Derivatives

The objective (18a) also deserves special attention as it is a nontrivial—but more or less fixed for the whole problem class—function that in particular comprises the inversion of a symmetric matrix.

First of all we note that (18a) only depends on the newly introduced variables H and J . This allows us to derive explicit formulas for the first and second derivative of the objective. Furthermore H and J enter the constraints only *linearly* through the coupled constraints (18b) and (18c), so the second derivative of the objective contains the entire curvature of the problem with respect to H

and J . This can be used to cheaply compute part of the Hessian approximations in SQP methods without the need to evaluate state sensitivities of higher order.

We only discuss the case of unconstrained parameter estimation problems (i.e. no matrix J) and $\phi(\cdot) = \text{tr}(\cdot)$ as optimization criterion.

5.2.1 Objective Gradient

We have an objective ϕ that maps a symmetric matrix to a scalar, however, we only include the lower triangular matrix as degrees of freedom in the nonlinear program. So whenever we need the gradient of the objective in a derivative based method we need derivatives of ϕ with respect to every entry H_{uv} , $1 \leq v \leq u \leq n_p$.

In the unconstrained case we have $H^{-1} = C$ and for every entry H_{uv}

$$\frac{\partial \phi}{\partial H_{uv}} = \frac{\partial \phi}{\partial C} \cdot \frac{\partial C}{\partial H_{uv}}.$$

Using the general formula

$$\frac{\partial H_{ij}^{-1}}{\partial H_{kl}} = -(H^{-1})_{ik} \cdot (H^{-1})_{lj} \quad (25)$$

we obtain for the derivative with respect to a fixed entry H_{uv} :

$$\frac{\partial \phi}{\partial H_{uv}} = \sum_{\substack{1 \leq i, j \leq n_p \\ 1 \leq k, l \leq n_p}} \frac{\partial \phi}{\partial C_{ij}} \cdot \frac{\partial C_{ij}}{\partial H_{kl}} \cdot \frac{\partial H_{kl}}{\partial H_{uv}} = - \sum_{\substack{1 \leq i, j \leq n_p \\ 1 \leq k, l \leq n_p}} \frac{\partial \phi}{\partial C_{ij}} \cdot C_{ik} \cdot C_{lj} \cdot \frac{\partial H_{kl}}{\partial H_{uv}}.$$

Taking symmetry into account, we have

$$\frac{\partial H_{kl}}{\partial H_{uv}} = \begin{cases} 1, & \text{if } (k, l) = (u, v) \text{ or } (l, k) = (u, v) \\ 0, & \text{else,} \end{cases}$$

and thus

$$\frac{\partial \phi}{\partial H_{uu}} = - \sum_{1 \leq i, j \leq n_p} \frac{\partial \phi}{\partial C_{ij}} \cdot C_{iu} \cdot C_{uj} \quad (26)$$

$$\frac{\partial \phi}{\partial H_{uv}} = -2 \sum_{1 \leq i, j \leq n_p} \frac{\partial \phi}{\partial C_{ij}} \cdot C_{iu} \cdot C_{vj}, \quad u > v. \quad (27)$$

For $\phi = \text{tr}$ this simplifies to

$$\frac{\partial \phi}{\partial H_{uu}} = - \sum_{1 \leq i \leq n_p} C_{iu}^2 \quad (28)$$

$$\frac{\partial \phi}{\partial H_{uv}} = -2 \sum_{1 \leq i \leq n_p} C_{iu} \cdot C_{vi}, \quad u > v. \quad (29)$$

This can be calculated immediately once the covariance matrix C is computed.

5.2.2 Objective Hessian

We now use formulae (28) and (29) to derive an explicit formula for the Hessian of $\text{tr}(H^{-1})$ with respect to the entries of H .

When computing $\frac{\partial^2 \text{tr}(H^{-1})}{\partial H_{uv} \partial H_{rs}}$ we distinguish between three cases:

- H_{uv} and H_{rs} diagonal elements
- H_{uv} diagonal, H_{rs} off-diagonal element
- H_{uv} and H_{rs} off-diagonal elements.

For two diagonal elements we obtain:

$$\begin{aligned} \frac{\partial^2 \phi(C)}{\partial H_{uu} \partial H_{rr}} &= \frac{\partial}{\partial H_{uu}} \left(\frac{\partial \phi}{\partial H_{rr}} \right) = \frac{\partial}{\partial H_{uu}} \left(- \sum_{1 \leq i \leq n_v} C_{ir}^2 \right) \\ &= -2 \sum_{1 \leq i \leq n_v} \frac{\partial C_{ir}}{\partial H_{uu}} \cdot C_{ir} = 2 \sum_{1 \leq i \leq n_v} C_{iu} \cdot C_{ur} \cdot C_{ir}. \end{aligned}$$

If we have one diagonal and one off-diagonal element we compute:

$$\begin{aligned} \frac{\partial^2 \phi(C)}{\partial H_{uu} \partial H_{rs}} &= \frac{\partial}{\partial H_{uu}} \left(\frac{\partial \phi}{\partial H_{rs}} \right) = \frac{\partial}{\partial H_{uu}} \left(-2 \sum_{1 \leq i \leq n_v} C_{ir} \cdot C_{si} \right) \\ &= -2 \sum_{1 \leq i \leq n_v} \frac{\partial C_{ir}}{\partial H_{uu}} \cdot C_{si} + C_{ir} \cdot \frac{\partial C_{si}}{\partial H_{uu}} \\ &= 2 \sum_{1 \leq i \leq n_v} C_{iu} \cdot C_{ur} \cdot C_{si} + C_{ir} \cdot C_{sp} \cdot C_{ui}. \end{aligned}$$

Taking the second derivative with respect to two off-diagonal elements yields:

$$\begin{aligned}
\frac{\partial^2 \phi(C)}{\partial H_{pq} \partial H_{rs}} &= \frac{\partial}{\partial H_{uv}} \left(\frac{\partial \phi}{\partial H_{rs}} \right) + \frac{\partial}{\partial H_{vu}} \left(\frac{\partial \phi}{\partial H_{rs}} \right) \\
&= \frac{\partial}{\partial H_{uv}} \left(-2 \sum_{1 \leq i \leq n_v} C_{ir} \cdot C_{si} \right) + \frac{\partial}{\partial H_{vu}} \left(-2 \sum_{1 \leq i \leq n_v} C_{ir} \cdot C_{si} \right) \\
&= -2 \sum_{1 \leq i \leq n_v} \frac{\partial C_{ir}}{\partial H_{uv}} \cdot C_{si} + C_{ir} \cdot \frac{\partial C_{si}}{\partial H_{uv}} + \frac{\partial C_{ir}}{\partial H_{vu}} \cdot C_{si} + C_{ir} \cdot \frac{\partial C_{si}}{\partial H_{vu}} \\
&= 2 \sum_{1 \leq i \leq n_v} C_{iu} \cdot C_{vr} \cdot C_{si} + C_{ir} \cdot C_{su} \cdot C_{vi} + C_{iv} \cdot C_{ur} \cdot C_{si} + C_{ir} \cdot C_{sv} \cdot C_{ui}.
\end{aligned}$$

As the variables H enter the constraints only linearly we note that the objective Hessian with respect to H is in fact the same as the Hessian of the Lagrangian with respect to H .

6 Numerical Results

We test our methods on two examples: The first one is a predator-prey model adapted to OED [19] that serves as proof of concept for both formulations. The second one is a more involved example from chemical engineering: the urethane reaction [15]. We report SQP iterations which basically amount to derivative evaluations as well as CPU time and compare the results to a single shooting implementation.

6.1 Implementation

We implemented direct multiple shooting for OED within our software package VPLAN [15] that allows to formulate, simulate and optimize DAE models. From a user specified formulation of the nominal DAE system, parameters, controls, and process constraints it generates structured NLPs for OED as described in Sect. 4. A special focus is on the efficient sparse evaluation of the constraints and their derivatives as outlined in Sect. 5 and parallelization of state integration and derivative evaluation on multi experiment and shooting node level.

As integrator we use DAESOL [3], a variable order and stepsize BDF method that can efficiently compute sensitivities of first and second order. The absolute and relative tolerance for all computations was set to 10^{-9} .

For the solution of the structured nonlinear programs, we implemented a filter based line search SQP method as described in [21]. The block structured quadratic

subproblems are solved by a modified version of the parametric active set solver qpOASES [11] that uses direct sparse linear algebra. The Hessian of the Lagrangian is approximated blockwise depending on the number of shooting intervals. For each block a positive definite damped BFGS update is employed which is scaled by the centered Oren-Luenberger sizing factor as described in [9]. Both full space and limited memory updates are available. In a multiple shooting context, the exact objective Hessian is computed cheaply as discussed in Sect. 5.2 and used for the lowermost diagonal block instead of a BFGS approximation. The optimality and nonlinear feasibility tolerance are set to 10^{-5} . Details of the SQP implementation will be discussed in an upcoming publication.

All results were obtained on a workstation with two Intel Xeon hexacore CPUs (2.4 GHz) allowing 24 parallel threads in total and 32 GB RAM running Ubuntu 12.04.

6.2 Example 1: Predator-Prey Model

We consider a predator-prey dynamics taken from [19] on the time horizon $[0, 12]$ with fixed initial values and an additional fishing term $0 \leq u(t) \leq 1$ as control:

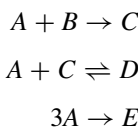
$$\begin{aligned}\dot{y}_1(t) &= y_1(t) - p_1 \cdot y_1(t) \cdot y_2(t) - 0.4 \cdot u(t) \cdot y_1(t), & y_1(0) &= 0.5 \\ \dot{y}_2(t) &= -y_2(t) + p_2 \cdot y_1(t) \cdot y_2(t) - 0.2 \cdot u(t) \cdot y_2(t), & y_2(0) &= 0.7.\end{aligned}$$

We assume $p_1 = p_2 = 1$ and that both states can be observed at most four times during the experiment with constant variances for the measurement errors, i.e., $\sigma_i = 1$. Both the control $u(t)$ and the grid of possible measurements are discretized on 50 equidistant intervals. The objective is to minimize the average variance of p_1 and p_2 , i.e. $\frac{1}{2} \text{tr}(C)$.

The initial guess for the controls is $u(t) \equiv 0.3$ and all 50 measurements selected, i.e. $w_i^1 = w_i^2 = 1$, $i = 1, \dots, 50$. It yields an objective function value (average variance) of $\frac{1}{2} \text{tr}(C) = 0.00683$ but note that this design is infeasible: all 50 possible measurements for both states are selected but only four per state are allowed.

6.3 Example 2: Urethane Reaction

The Urethane reaction is a well-known example from chemical reaction kinetics, see [15]. The reaction scheme is the following:



Educts are phenylisocyanate A and butanol B in the solvent dimethylsulfoxide L . During the reaction the product urethane C , the byproduct allophanate D and the byproduct isocyanate E are formed.

The products C , D , and E are modelled as differential states while A , B , and L can be computed from C , D , and E using molar number balance. In total, six parameters have to be identified, namely the frequency factors and activation energies for the Arrhenius kinetics. The objective is to minimize $\frac{1}{6} \text{tr}(C)$. To achieve this, one experiment is to be designed for the time horizon $[t_0, t_{end}] = [0h, 80h]$ and one out of three possible measurements can be taken at each of 11 equidistant points in time. The reactor is run in a stirrer tank with two feeds: Feed 1 contains phenylisocyanate and the solvent, feed 2 contains dimethylsulfoxide and the solvent. Both can be fed into the reactor during the process. Furthermore, the temperature can be controlled. For our numerical experiments, we parameterize the derivative of the control functions $\dot{T}(t)$, $\dot{feed}_1(t)$, and $\dot{feed}_2(t)$ by piecewise constant functions on ten equidistant intervals. The actual process controls $T(t)$, $feed_1(t)$, and $feed_2(t)$ are set up as additional differential states, so we end up with a total of six state variables. The constraints on the controls are formulated as path constraints. The full model including process constraints and measurement methods is summarized in Fig. 3. The variances of the measurement errors are assumed constant with $\sigma_i = 1$.

The initial guess for the controls is depicted in Fig. 4. It yields an objective function value (average variance) of $\frac{1}{6} \text{tr}(C) = 1936.3$.

6.4 Results Predator-Prey

We solved the problem with both direct multiple shooting formulations introduced in Sect. 4. We keep the control discretization fixed but vary the number of shooting intervals giving rise to different, yet equivalent, NLPs. We discovered a number of different local minima that differ mainly with respect to the placing of the measurements. Table 1 shows the results for the first multiple shooting formulation—a coupled constraint for the information matrix—as well as the single shooting formulation while Table 2 shows results for the formulation where the covariance is distributed over the shooting intervals. Both algorithms terminated successfully when restarted in the minima found by the other one indicating the structural correctness of our approach.

Table 1 shows that the first formulation converges for all discretizations. However, we note how the number of SQP iterations increases when we increase the number of shooting intervals. Preliminary experiments show that this is probably due to the BFGS update that is unable to reflect negative curvature of the underlying Lagrangian. A block SQP method that can handle indefinite approximations which can reduce this effect is currently under development. When we look at the CPU time, however, we see that the benefits of the efficient evaluation scheme for the multiple shooting formulation outweighs the smaller number of SQP iterations for single shooting, making direct multiple shooting with a moderate number of

<i>States</i>		<i>Educts</i>
$\dot{n}_C = V \cdot (r_1 - r_2 + r_3),$	$n_C(0) = 0$	$n_{A,e} = n_{A,e1,0} \cdot feed_1$
$\dot{n}_D = V \cdot (r_2 - r_3),$	$n_D(0) = 0$	$n_{B,e} = n_{B,e2,0} \cdot feed_2$
$\dot{n}_E = V \cdot r_4,$	$n_E(0) = 0$	$n_{L,e} = n_{L,e1,0} \cdot feed_1 + n_{L,e2,0} \cdot feed_2$
$feed_1 = u_1,$	$feed_1(0) = 0$	$n_A = n_{A,0} + n_{A,e} - n_C - 2 \cdot n_D - 3 \cdot n_E$
$feed_2 = u_2,$	$feed_2(0) = 0$	$n_B = n_{B,0} + n_{B,e} - n_C - n_D$
$\dot{T} = u_3,$	$T(0) = 293.15$	$n_L = n_{L,0} + n_{L,e}$

Reaction Rates

$$r_1 = k_{ref1} \cdot \exp\left(-\frac{E_{a,1}}{R} \cdot \left(\frac{1}{T} - \frac{1}{363.16}\right)\right) \cdot \frac{n_A}{V} \cdot \frac{n_B}{V}$$

$$r_2 = k_{ref2} \cdot \exp\left(-\frac{E_{a,2}}{R} \cdot \left(\frac{1}{T} - \frac{1}{363.16}\right)\right) \cdot \frac{n_A}{V} \cdot \frac{n_C}{V}$$

$$r_3 = k_{ref2} \cdot \exp\left(-\frac{E_{a,2}}{R} \cdot \left(\frac{1}{T} - \frac{1}{363.16}\right)\right) \cdot \left(K_{C2} \cdot e^{-\frac{\Delta H_2}{R} \cdot \left(\frac{1}{T} - \frac{1}{T_{C2}}\right)}\right)^{-1} \cdot \frac{n_D}{V}$$

$$r_4 = k_{ref4} \cdot \exp\left(-\frac{E_{a,4}}{R} \cdot \left(\frac{1}{T} - \frac{1}{363.16}\right)\right) \cdot \left(\frac{n_A}{V}\right)^2$$

$$V = \frac{n_A \cdot M_A}{\rho_A} + \frac{n_B \cdot M_B}{\rho_B} + \frac{n_C \cdot M_C}{\rho_C} + \frac{n_D \cdot M_D}{\rho_D} + \frac{n_E \cdot M_E}{\rho_E} + \frac{n_L \cdot M_L}{\rho_L}$$

Time Dependent Controls

$$0 \leq feed_1, feed_2 \leq 1$$

$$0 \leq u_1, u_2 \leq 0.0125$$

$$273.15 \leq T \leq 473.15$$

$$-40 \leq u_3 \leq 40$$

Time Independent Controls

$$0.1 \leq n_{A,0} \leq 1, \quad 0 \leq n_{B,0}, n_{L,0} \leq 1$$

$$\frac{n_{A,0}M_A + n_{B,0}M_B}{n_{A,0}M_A + n_{B,0}M_B + n_{L,0}M_L} \leq 0.8$$

$$n_{A,0}M_A\rho_A^{-1} + n_{B,0}M_B\rho_B^{-1} + n_{L,0}M_L\rho_L^{-1} \leq 7.5 \cdot 10^{-4}$$

Measurements

$$h_1(t) = 100 \cdot \frac{n_A \cdot M_A}{n_A \cdot M_A + n_B \cdot M_B + n_C \cdot M_C + n_D \cdot M_D + n_E \cdot M_E + n_L \cdot M_L}$$

$$h_2(t) = 100 \cdot \frac{n_C \cdot M_C}{n_A \cdot M_A + n_B \cdot M_B + n_C \cdot M_C + n_D \cdot M_D + n_E \cdot M_E + n_L \cdot M_L}$$

$$h_3(t) = 100 \cdot \frac{n_D \cdot M_D}{n_A \cdot M_A + n_B \cdot M_B + n_C \cdot M_C + n_D \cdot M_D + n_E \cdot M_E + n_L \cdot M_L}$$

$$h_4(t) = 100 \cdot \frac{n_E \cdot M_E}{n_A \cdot M_A + n_B \cdot M_B + n_C \cdot M_C + n_D \cdot M_D + n_E \cdot M_E + n_L \cdot M_L}$$

$$w_i^2 = w_i^3, \quad w_i^1 + w_i^2 + w_i^4 \leq 1 \quad i = 0, \dots, 16$$

Parameters

$$k_{ref1} = 5.0 \cdot 10^{-4}$$

$$E_{a,1} = 35240$$

$$k_{ref2} = 8.0 \cdot 10^{-8}$$

$$E_{a,2} = 85000$$

$$k_{ref4} = 1.0 \cdot 10^{-8}$$

$$E_{a,4} = 35000$$

Constants

$$M_A = 0.11911, \quad \rho_A = 1095$$

$$M_B = 0.07412, \quad \rho_B = 809$$

$$M_C = 0.19323, \quad \rho_C = 1415$$

$$M_D = 0.31234, \quad \rho_D = 1528$$

$$M_E = 0.35733, \quad \rho_E = 1451$$

$$M_L = 0.07806, \quad \rho_L = 1101$$

$$\Delta H_2 = -17031.0, \quad K_{C2} = 0.17$$

Fig. 3 Urethane reaction model

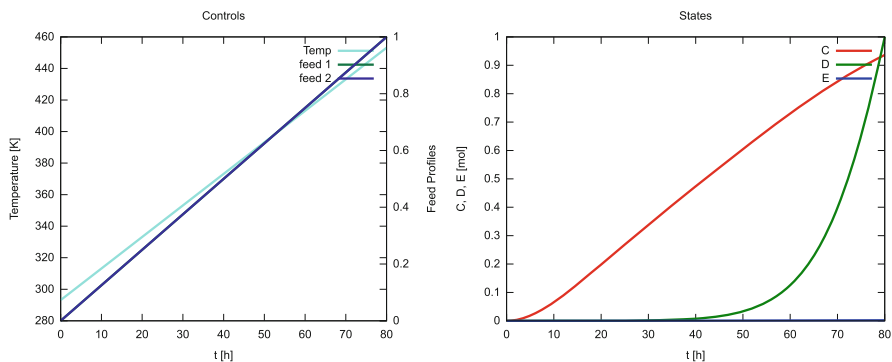


Fig. 4 Initial values for controls and corresponding states for the urethane example. The objective (average variance) is $\frac{1}{6} \text{tr}(C) = 1936.3$

Table 1 Performance of multiple shooting for information matrix for predator-prey

Shooting intervals	Var.	Cons.	SQP iterations	Obj.	Time (s)
Single shooting	150	2	16	7.614527e-03	11.03
2	159	11	22	7.614526e-03	3.47
4	173	25	37	7.908877e-03	2.30
6	187	39	52	7.908865e-03	2.18
12	229	81	82	7.908856e-03	2.27
25	297	149	91	7.908878e-03	2.77
50	447	299	126	7.634122e-03	5.94

Table 2 Performance of multiple shooting for covariance matrix for predator-prey

Shooting intervals	Var.	Cons.	SQP iterations	Obj.	Time (s)
2	162	14	41	7.794461e-03	6.55
4	182	34	122	7.646316e-03	9.73
6	202	54	–	–	–
12	262	114	90	8.270562e-03	6.29
25	369	221	–	–	–
50	594	446	304	7.952309e-03	34.01

shooting intervals the best overall choice. The following aspects are responsible for this:

- derivatives with respect to controls are required only locally that means less directional derivatives of second order are needed
- derivative evaluation is easily parallelized on a multicore machine.

While the first formulation converges for all discretizations within a reasonable number of SQP iterations, the second formulation does not converge for every multiple shooting discretization. Furthermore, many of the SQP steps were reduced

steps, which means many additional constraint and objective evaluations were necessary. Overall, the behaviour was competitive to the first formulation and single shooting only for certain problem instances.

6.5 Results Urethane

We solved the problem with the more promising variant of direct multiple shooting meaning we transform it to an NLP of the form (18) where we introduce a coupled constraint for the information matrix. Again, we use different numbers of shooting intervals with the same control discretization. The problem exhibits several structurally different local minima, however, all of them yield significantly better objective values than the initial guess. The states and controls for one of them are depicted in Fig. 5.

The SQP method was able to find a local minimum for every multiple shooting discretization. The results comprising the number of major iterations, the final objective value and the CPU time in seconds are summarized in Table 3.

We see that the method performs comparably well in terms of SQP iterations for single and multiple shooting. For CPU time, again the results shift strongly in favor of multiple shooting because the derivative evaluation can be done much cheaper.

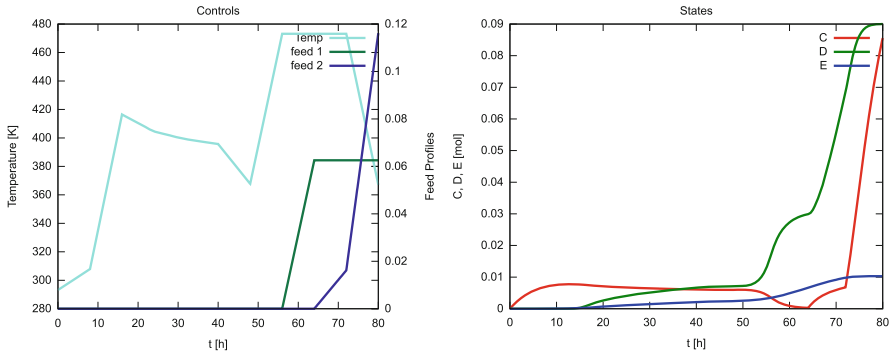


Fig. 5 Optimum experimental design for the urethane example. The objective (average variance) is $\frac{1}{6} \text{tr}(C) = 0.0665$

Table 3 Performance of single and multiple shooting for urethane

Shooting intervals	Var.	Cons.	SQP iterations	Obj.	Time (s)
Single shooting	68	45	68	1.241664e-01	85.87
2	138	112	62	3.441588e-01	30.61
4	242	261	105	6.651601e-02	32.64
10	530	480	117	8.889849e-02	43.07

7 Conclusions

In this paper, we reviewed a nonstandard optimal control problem formulation of OED. For this formulation, we showed how to extend the classical direct multiple shooting method for optimal control problems for OED problems in two ways, leading to highly structured nonlinear programs. Special structures in the constraints and objective derivatives are highlighted that must be taken into account in an efficient implementation. The algorithms presented are implemented within the software package VPLAN. We presented two application examples, one of them a challenging example from chemical engineering that could be solved successfully with one of the new formulations. Our implementation outperforms an existing single shooting implementation in terms of CPU time.

We expect direct multiple shooting for OED to have more benefits for more challenging, large-scale real-life problems. Especially when nontrivial path constraints are present, as it is often the case for real-life systems, direct single shooting can run into problems finding feasible points. The direct multiple shooting method as introduced in this paper allows to choose fine shooting discretizations in critical regions and offers more flexibility for initialization. Another point is that OED, even for small nominal systems, basically requires the solution of an additional $n_x \times n_p$ variational system and even though this can be done efficiently using the principles of IND, its solution becomes very time consuming for large-scale systems. Here the lower sensitivity load as well as the excellent potential for parallelization provide great benefits for multiple shooting.

References

1. Albersmeyer, J.: Adjoint based algorithms and numerical methods for sensitivity generation and optimization of large scale dynamic systems. Ph.D. thesis, Ruprecht-Karls-Universität Heidelberg (2010)
2. Atkinson, A.C., Donev, A.: Optimum Experimental Designs. Oxford Statistical Sciences Series, vol. 8. Oxford University Press, Oxford (1992)
3. Bauer, I., Bock, H.G., Schlöder J.P.: DAESOL – a BDF-code for the numerical solution of differential algebraic equations. Internal Report, IWR, SFB 359, Universität Heidelberg (1999)
4. Bauer, I., Bock, H.G., Körkel, S., Schlöder, J.P.: Numerical methods for optimum experimental design in DAE systems. *J. Comput. Appl. Math.* **120**(1–2), 1–15 (2000)
5. Bock, H.G.: Numerical treatment of inverse problems in chemical reaction kinetics. In: Ebert, K.H., Deuffhard, P., Jäger, W. (eds.) *Modelling of Chemical Reaction Systems*. Springer Series in Chemical Physics, vol. 18, pp. 102–125. Springer, Heidelberg (1981)
6. Bock, H.G.: Randwertproblemmethoden zur parameteridentifizierung in systemen nichtlinearer differentialgleichungen. *Bonner Mathematische Schriften*, vol. 183. Universität Bonn, Bonn (1987)
7. Bock, H.G., Plitt, K.J.: A Multiple Shooting algorithm for direct solution of optimal control problems. In: *Proceedings of the 9th IFAC World Congress*, pp. 242–247. Pergamon Press, Budapest (1984). Available at <http://www.iwr.uni-heidelberg.de/groups/agbock/FILES/Bock1984.pdf>

8. Bock, H.G. Eich, E., Schlöder, J.P.: Numerical solution of constrained least squares boundary value problems in differential-algebraic equations. In: Strehmel, K. (ed.) *Numerical Treatment of Differential Equations. Proceedings of the NUMDIFF-4 Conference, Halle-Wittenberg, 1987. Texte zur Mathematik*, vol. 104, pp. 269–280. Teubner, Leipzig (1988)
9. Contreras, M., Tapia, R.A.: Sizing the BFGS and DFP updates: numerical study. *J. Optim. Theory Appl.* **78**(1), 93–108 (1993)
10. Fedorov, V.V.: *Theory of Optimal Experiments*. Elsevier, Amsterdam (1972)
11. Ferreau, H.J., Kirches, C., Potschka, A., Bock, H.G., Diehl, M.: qpOASES: A parametric active-set algorithm for quadratic programming. *Math. Program. Comput.* **6**(4), 327–363 (2014)
12. Franceschini, G., Macchietto, S.: Model-based design of experiments for parameter precision: state of the art. *Chem. Eng. Sci.* **63**, 4846–4872 (2008)
13. Hoang, M.D., Barz, T., Merchan, V.A., Biegler, L.T., Arellano-Garcia, H.: Simultaneous solution approach to model-based experimental design. *AIChE J.* **59**(11), 4169–4183 (2013)
14. Janka, D.: *Optimum experimental design and multiple shooting*. Master’s thesis, Universität Heidelberg, Heidelberg (2010)
15. Körkel, S.: *Numerische methoden für optimale versuchsplanungsprobleme bei nichtlinearen DAE-modellen*. Ph.D. thesis, Universität Heidelberg, Heidelberg (2002)
16. Körkel, S., Potschka, A., Bock, H.G., Sager, S.: A multiple shooting formulation for optimum experimental design. *Math. Program.* (2012, submitted revisions)
17. Potschka, A., Bock, H.G., Schlöder, J.P.: A minima tracking variant of semi-infinite programming for the treatment of path constraints within direct solution of optimal control problems. *Optim. Methods Softw.* **24**(2), 237–252 (2009)
18. Pukelsheim, F.: *Optimal design of experiments*. In: *Classics in Applied Mathematics*, vol. 50. SIAM, Philadelphia (2006). ISBN:978-0-898716-04-7.
19. Sager, S.: MIOCP benchmark site. (2014) <http://mintoc.de>
20. Sager, S.: Sampling decisions in optimum experimental design in the light of Pontryagin’s maximum principle. *SIAM J. Control. Optim.* **51**(4), 3181–3207 (2013)
21. Wächter, A., Biegler, L.T.: Line search filter methods for nonlinear programming: motivation and global convergence. *SIAM J. Optim.* **16**(1), 1–31 (2005)