Available online at www.sciencedirect.com
SCIENCE DIRECT®
Computers and Chemical Engineering 27 (2003) 157–166
Computers & Chemical Engineering
www.elsevier.com/locate/compchemeng

An efficient multiple shooting based reduced SQP strategy for large-scale dynamic process optimization. Part 1: theoretical aspects

Daniel B. Leineweber [a], Irene Bauer [b,1], Hans Georg Bock [b,*], Johannes P. Schlöder [b]

[a] *Process Technology, Bayer AG, D-51368 Leverkusen, Germany*
[b] *Interdisciplinary Center for Scientific Computing (IWR), University of Heidelberg, Im Neuenheimer Feld 368, D-69120 Heidelberg, Germany*

Received 31 July 2001; received in revised form 23 May 2002; accepted 23 May 2002

Abstract

Optimal design and operation of complex chemical processes often require the solution of intricate dynamic optimization problems. A tailored simultaneous solution strategy based on multiple shooting and reduced SQP is presented. This *reduced-space boundary value problem* (BVP) *approach* allows an efficient and robust solution of multistage optimal control and design optimization problems for large, sparse DAE process models of index one. The current paper describes the theoretical aspects of the method. Utilizing the natural decomposition of the states into differential and algebraic variables, the structured NLP problem which results from the multiple shooting discretization of the optimization BVP is projected onto the reduced space of differential variables and control parameters. It is shown that this projection can be obtained very efficiently through direct computation of the reduced linearized constraint system via directional sensitivities. Like the original full-space BVP approach, the reduced-space formulation lends itself well to parallel computation. An implementation of the new strategy is provided by the modular optimal control package MUSCOD-II. Software aspects and applications are discussed in a second paper (Part II Software Aspects and Applications, 2002).

Keywords: Large-scale optimal control; Index one DAEs; Multiple shooting; Structured reduced SQP methods; Directional derivatives; Sparse equation systems; Parallel computation

1. Introduction

Today most practical optimal control problems are solved with NLP methods (or *direct* methods) based on some discretization of the original continuous problem. Sequential strategies just parameterize the control functions and employ numerical integration to discretize the dynamic model in a black box manner (e.g., Sargent & Sullivan, 1978; Vassiliadis, Sargent & Pantelides, 1994). Simultaneous strategies, on the other hand, explicitly include a suitable discretization of the dynamic model equations in their NLP formulation (e.g., Biegler, 1984), i.e., simulation and optimization may proceed simultaneously if an infeasible path optimization algorithm is employed. However, the great potential of simultaneous strategies in terms of both efficiency and robustness seems hard to realize in practice due to the very large size of the resulting NLP problems and the difficulty of adaptively controlling the discretization error, especially if *collocation* is applied (Logsdon & Biegler, 1989, 1992).

In this context, a promising alternative way of discretizing the dynamic model is provided by *multiple shooting* (Bock & Plitt, 1984; Bock, Eich, & Schlöder, 1988; Leineweber, 1999). Unlike the more common collocation based strategies, direct use is made of existing advanced, fully adaptive DAE solvers. Therefore, a *fixed* discretization can be employed at the optimization level by restricting adaptivity to the error-controlled evaluation of the NLP functions and gradients. The NLP problems to be solved are typically much smaller than in case of collocation. A further advantage lies in the fact that the integrations on different multiple

* Corresponding author
E-mail address: bock@iwr.uni-heidelberg.de (H.G. Bock).
[1] Present address: Process Systems Enterprise Ltd, London W6 9DA, United Kingdom.

0098-1354/02/$ - see front matter
PII: S0098-1354(02)00158-8

shooting intervals are completely decoupled and can therefore be performed in parallel.

In order to exploit the structure of large, sparse DAE process models with many algebraic variables, a new partially reduced SQP (PRSQP) strategy has been developed (Leineweber, 1999). The structured NLP problem resulting from the multiple shooting DAE discretization is projected onto the reduced space of differential variables plus control parameters, using the fact that the algebraic variables are uniquely determined by the consistency conditions due to the index one assumption. It is demonstrated that the expensive calculation of the full set of differential state sensitivities can be avoided—instead, it suffices to compute a reduced number of *directional* sensitivities if the projection and gradient evaluation steps are suitably intertwined.

A comprehensive implementation of this *reduced-space* BVP approach is provided by the modular optimal control package MUSCOD-II (Leineweber, 1999; Leineweber, Schäfer, Bock, & Schlöder, 2002). The BDF code DAESOL (Bauer, Finocchi, Duschl, Gail, & Schlöder, 1997; Bauer, 2000) is used for the efficient calculation of the required directional derivatives.

2. Problem formulation

2.1. A multistage optimal control problem in DAEs

Many dynamic process optimization problems in chemical engineering can be expressed as multistage optimal control problems in DAEs. Throughout this paper, we will consider the following general class of M-stage optimal control problems, where the time horizon of interest $[t_0,\ t_M]$ is divided into M model stages corresponding to the subintervals $[t_i,\ t_{i+1}]$, $i=0,\ 1,\ldots,\ M-1$:

$$\min_{x_i,z_i,u_i,p,t_i} \sum_{i=0}^{M-1} \phi_i(x_i(t_{i+1}), z_i(t_{i+1}), p, t_{i+1}) \tag{1a}$$

subject to the DAE model stages

$$\begin{aligned} &B_i(\ )\dot{x}_i(t) = f_i(x_i(t), z_i(t), u_i(t), p, t)\\ &0 = g_i(x_i(t), z_i(t), u_i(t), p, t),\\ &t\in[t_i, t_{i+1}], \quad i=0,1,\ldots,M-1, \end{aligned} \tag{1b}$$

the control and path constraints

$$\begin{aligned} &h_i(x_i(t), z_i(t), u_i(t), p, t) \geqslant 0,\\ &t\in[t_i, t_{i+1}], \quad i=0,1,\ldots,M-1, \end{aligned} \tag{1c}$$

the stage transition conditions

$$\begin{aligned} &x_{i+1}(t_{i+1}) = c_i(x_i(t_{i+1}), z_i(t_{i+1}), p, t_{i+1}),\\ &\quad i=0,1,\ldots,M-2, \end{aligned} \tag{1d}$$

and the multipoint boundary conditions

$$\sum_{i=0}^{M-1} (r_i^s(x_i(t_i), z_i(t_i), p, t_i) + r_i^e(x_i(t_{i+1}), z_i(t_{i+1}), p, t_{i+1})) \begin{Bmatrix} = \\ \geq \end{Bmatrix} 0. \tag{1e}$$

The performance index (Eq. (1a)) of generalized Mayer type is minimized with respect to the differential and algebraic state profiles x_i and z_i, respectively, the control profiles u_i, a time-independent global parameter vector p, and the model stage grid points t_i. The vectors $x_i(t)$, $z_i(t)$, $u_i(t)$, and p are of dimensions n_i^x, n_i^z, n_i^u, and n^p (typically, n_i^z is larger than n_i^x, and n_i^x is larger than n_i^u, n^p). Of course, the dynamic model equations as well as the relevant dimensions may change from one model stage to the next. For the quasilinear-implicit nonlinear DAEs (Eq. (1b)) we assume (*differential*) *index one*, i.e., B_i and $\partial g_i/\partial z_i$ must be nonsingular. In most practical cases, B_i as well as the Jacobians of f_i and g_i are quite sparse.

The possibility to directly consider sequences of *dynamic model stages* as shown in (Eqs. (1a), (1b), (1c), (1d) and (1e)) is valuable particularly for the optimization of coupled (or integrated) batch processes, e.g., a batch reactor connected to a batch distillation column. Observe that additional 'recycle loops' can be easily specified for a given linear sequence of coupled batch processes by means of the multipoint boundary conditions (Eq. (1e)) and the stage transition conditions (Eq. (1d)). This enables the optimization of cyclic batch operations, where material and energy can be transferred from one batch to the next, e.g., a batch distillation process with recycling of slop cuts.

2.2. Problem discretization

The original continuous problem (Eqs. (1a), (1b), (1c), (1d) and (1e)) is reformulated as an NLP problem with a finite number of optimization variables. On each model stage $i=0,\ 1,\ldots,\ M-1$, we employ the time transformation

$$\theta_i(\tau, v) := t_i + \tau d_i, \qquad t_i = t_0 + \sum_{k=0}^{i-1} d_k, \quad \tau\in[0,1] \tag{2}$$

with $v := (t_0,\ d_0,\ d_1,\ldots,\ d_{M-1})$, and we choose a fixed, dimensionless discretization grid

$$0 = \tau_{i,0} < \tau_{i,1} < \cdots < \tau_{i,m_i} = 1. \tag{3}$$

Then we define a piecewise approximation $\hat{u}_i$ of the control vector u_i by

$$\hat{u}_i(\tau) := \varphi_{ij}(\tau, q_{ij}),$$
$$\tau \in I_{ij} = [\tau_{ij}, \tau_{i,j+1}], \quad j = 0, 1, \ldots, m_i - 1, \tag{4}$$

using 'local' control parameters q_{ij}. The functions φ_{ij} are given basis functions, typically vectors of polynomials (e.g., constant, linear, or cubic).

The DAEs (Eq. (1b)) are discretized on the same grid (Eq. (3)) by a specific variant of *multiple shooting* which has been proposed by Bock et al. (1988). We introduce additional optimization parameters s_{ij}^x, s_{ij}^z and solve the following set of relaxed decoupled initial value problems (IVPs),

$$B_i(\ )\mathrm{d}x_i(\tau)/\mathrm{d}\tau = f_i(x_i(\tau), z_i(\tau), \varphi_{ij}(\tau, q_{ij}), p, \theta_i(\tau, v))d_i$$

$$0 = g_i(x_i(\tau), z_i(\tau), \varphi_{ij}(\tau, q_{ij}), p, \theta_i(\tau, v)) - \alpha_{ij}(\tau) g_i(s_{ij}^x, s_{ij}^z, \varphi_{ij}(\tau_{ij}, q_{ij}), p, \theta_i(\tau_{ij}, v)), \quad \tau \in I_{ij}, \tag{5}$$

using $x_i(\tau_{ij}) = s_{ij}^x$ and $z_i(\tau_{ij}) = s_{ij}^z$ as initial values. Note that *relaxed* versions of the original DAEs (Eq. (1b)) appear in Eq. (5) to allow inconsistent values of s_{ij}^x and s_{ij}^z. The damping factor α_{ij} is given by a scalar function which is nonincreasing and nonnegative on I_{ij} and satisfies $\alpha_{ij}(\tau_{ij}) = 1$, e.g.,

$$\alpha_{ij}(\tau) = \exp\left(-\bar{\alpha}\frac{\tau - \tau_{ij}}{\tau_{i,j+1} - \tau_{ij}}\right), \quad \bar{\alpha} \geqslant 0. \tag{6}$$

Let $x_i(\tau_{i,j+1}; s_{ij}^x, s_{ij}^z, q_{ij}, p, v)$ denote the differential state values at $\tau = \tau_{i,j+1}$ obtained by numerical integration of Eq. (5). Then, by including the continuity conditions

$$x_i(\tau_{i,j+1}; s_{ij}^x, s_{ij}^z, q_{ij}, p, v) - s_{i,j+1}^x = 0, \quad j = 0, 1, \ldots, m_i - 1 \tag{7}$$

and the consistency conditions

$$g_i(s_{ij}^x, s_{ij}^z, \varphi_{ij}(\tau_{ij}, q_{ij}), p, \theta_i(\tau_{ij}, v)) = 0, \quad j = 0, 1, \ldots, m_i \tag{8}$$

as equality constraints in the NLP problem, we ensure that the optimal solution will satisfy the original DAEs (Eq. (1b)) on each model stage $i = 0, 1, \ldots, M-1$. However, since the NLP problem is solved by an infeasible path algorithm, the intermediate iterates generated in the solution process are in general *infeasible* with respect to the above DAE discretization (Eqs. (7) and (8)), i.e., optimization and simulation may proceed simultaneously.

Allowing inconsistent values of the discretized state variables at the multiple shooting nodes through the relaxed DAE formulation (Eq. (5)) is an important feature of our BVP approach, because quite frequently, discontinuities are introduced at the discretization grid points τ_{ij} (e.g., if a piece-wise constant control approximation is used). A feasible path approach would normally require a consistent initialization of the DAE model after each discontinuity, and this has been found to become very expensive for larger process models, even in the context of dynamic simulation. Therefore, we deal with the consistency issue at the level of the optimization problem, and our infeasible path strategy requires consistency only for the final optimal solution. It is interesting to note that the use of the damping factor $\alpha_{ij}(\tau)$ within the relaxed IVPs (Eq. (5)) in general amounts to a homotopy-like strategy which forces the solution more and more toward the 'correct' manifold during integration. Our experience indicates that damping is indeed beneficial, i.e., $\bar{\alpha}$ in Eq. (6) should be nonzero, but there fortunately seems to be little dependence on the exact magnitude of $\bar{\alpha}$. For practical nonlinear DAE problems, $\bar{\alpha} = 5.0$ provides an adequate choice (Leineweber, 1999).

Finally, the continuous control and path constraints (Eq. (1c)) are discretized at the points τ_{ij} of the grid (Eq. (3)), leading to a set of constraints of the form

$$h_i(s_{ij}^x, s_{ij}^z, \varphi_{ij}(\tau_{ij}, q_{ij}), p, \theta_i(\tau_{ij}, v)) \geqslant 0, \quad j = 0, 1, \ldots, m_i \tag{9}$$

for each model stage $i = 0, 1, \ldots, M-1$. By choosing a proper parameterization of the controls, one can ensure that the potential violation of constraints is of second or higher order in $\Delta\tau$, the local control mesh width. Alternatively, one might consider applying one of the well-known techniques to enforce strict feasibility, e.g., the transformation of each nonlinear path constraint to an equivalent terminal constraint for an artificial state variable (Sargent & Sullivan, 1978). Such a strategy can be easily realized within the general framework of formulation (Eqs. (1a), (1b), (1c), (1d) and (1e)). However, since techniques of this kind may give rise to numerical difficulties (introduction of non-differentiabilities, zero constraint gradients at the solution), we do not generally recommend their use in the context of our approach—we rather suggest refining the grid (Eq. (3)) if required.

2.3. Structured NLP problem

The reformulation of the continuous M-stage optimal control problem (Eqs. (1a), (1b), (1c), (1d) and (1e)) leads to the following structured NLP problem:

$$\min_{s_{ij}^x, s_{ij}^z, q_{ij}, p, v} \sum_{i=0}^{M-1} \phi_i(s_{i,m_i}^x, s_{i,m_i}^z, p, \theta_i(\tau_{i,m_i}, v)) \tag{10a}$$

subject to the continuity and consistency conditions

$$x_i(\tau_{i,j+1}; s_{ij}^x, s_{ij}^z, q_{ij}, p, v) - s_{i,j+1}^x = 0, \quad j = 0, 1, \ldots, m_i - 1$$

$$g_i(s_{ij}^x, s_{ij}^z, \varphi_{ij}(\tau_{ij}, q_{ij}), p, \theta_i(\tau_{ij}, v)) = 0, \quad j = 0, 1, \ldots, m_i$$

$$i = 0, 1, \ldots, M-1, \tag{10b}$$

the discretized control and path constraints

$$h_i(s^x_{ij}, s^z_{ij}, \varphi_{ij}(\tau_{ij}, q_{ij}), p, \theta_i(\tau_{ij}, v)) \geqslant 0, \quad j = 0, 1, \ldots, m_i, \quad i = 0, 1, \ldots, M-1, \tag{10c}$$

the stage transition conditions

$$c_i(s^x_{i,m_i}, s^z_{i,m_i}, p, \theta_i(\tau_{i,m_i}, v)) - s^x_{i+1,0} = 0, \quad i = 0, 1, \ldots, M-2, \tag{10d}$$

and the linearly coupled multipoint constraints

$$\sum_{i=0}^{M-1} (r^s_i(s^x_{i,0}, s^z_{i,0}, p, \theta_i(\tau_{i,0}, v)) + r^e_i(s^x_{i,m_i}, s^z_{i,m_i}, p, \theta_i(\tau_{i,m_i}, v))) \left\{ \begin{matrix} = \\ \geqslant \end{matrix} \right\} 0. \tag{10e}$$

A further slight modification of the discretized problem (Eqs. (10a), (10b), (10c), (10d) and (10e)) is favorable in practice: the global model parameters (p, v) are 'localized' by introducing new NLP variables $(p, v)_{ij}$ together with additional linear constraints enforcing

$$(p, v)_{ij} - (p, v)_{0,0} = 0 \tag{11}$$

for all $(i, j) \neq (0, 0)$. Then only *linear coupling* occurs between variables corresponding to different discretization intervals (some of the problem functions introduce no coupling at all), i.e., all problem functions are second-order decoupled.

For typical chemical process control applications, the structured NLP problem (Eqs. (10a), (10b), (10c), (10d) and (10e)) may have several thousand variables due to the state discretization, but only relatively few degrees of freedom—say, up to a few hundred—according to the number of control parameters and global model parameters. Therefore, the performance of solution strategies very much depends on an adequate treatment of the large, structured system of nonlinear equality constraints (Eq. (10b)).

3. A natural decomposition for DAE optimization

3.1. General reduced SQP framework

Large-scale NLP problems like Eqs. (10a), (10b), (10c), (10d) and (10e) having many equality constraints can be efficiently solved using reduced SQP (RSQP) methods (e.g., Locke, Westerberg, & Edahl, 1983; Biegler, Nocedal, & Schmid, 1995; Cervantes, Wächter, Tutuncu, & Biegler, 2000). A generalization of the classical RSQP concept is provided by PRSQP methods, where only a *subset* of the equalities is used in defining the range and null space decomposition, and the reduced linearizations of the inequalities and the remaining equalities are kept in the QP subproblem. A detailed investigation of such methods including convergence properties can be found in Schulz (1996). Thus within a coordinate bases approach, we may consider a generic partitioned NLP problem of the type

$$\min_{w_1, w_2} F(w_1, w_2) \quad \text{subject to} \quad \begin{cases} G_1(w_1, w_2) = 0 \\ G_2(w_1, w_2) = 0, \\ H(w_1, w_2) \geqslant 0 \end{cases} \tag{12}$$

where $w_1 \in \mathbb{R}^{m_1}$ and $w_2 \in \mathbb{R}^{n-m_1}$ denote the basic (or dependent) variables and the nonbasic (or free) variables, respectively. (For simplicity of presentation, we have not included simple bounds here.) Of course, the m_1 basic variables w_1 and the corresponding m_1 constraint functions G_1 are chosen such that $\nabla_{w_1} G_1(w_1, w_2)^T$ is nonsingular. We then obtain the coordinate bases

$$S^{\mathcal{N}}_k = \begin{pmatrix} -(\nabla_{w_1} G^T_{1,k})^{-1} \nabla_{w_2} G^T_{1,k} \\ I \end{pmatrix} \in \mathbb{R}^{n \times (n-m_1)}, \quad S^{\mathcal{R}}_k = \begin{pmatrix} I \\ 0 \end{pmatrix} \in \mathbb{R}^{n \times m_1}. \tag{13}$$

For this particular choice of bases, the range and null space decomposition of the correction step becomes

$$\begin{pmatrix} \Delta w_{1,k} \\ \Delta w_{2,k} \end{pmatrix} = S^{\mathcal{N}}_k y^{\mathcal{N}}_k + S^{\mathcal{R}}_k y^{\mathcal{R}}_k = \begin{pmatrix} D_k y^{\mathcal{N}}_k + y^{\mathcal{R}}_k \\ y^{\mathcal{N}}_k \end{pmatrix}, \tag{14}$$

where we have introduced the abbreviation

$$D_k := -(\nabla_{w_1} G^T_{1,k})^{-1} \nabla_{w_2} G^T_{1,k} \in \mathbb{R}^{m_1 \times (n-m_1)}.$$

By applying the change of variables (Eq. (14)) to the full-space QP subproblem which corresponds to the partitioned NLP problem (Eq. (12)), it directly follows that the range space component $y^{\mathcal{R}}_k$ is given by

$$y^{\mathcal{R}}_k = -(\nabla_{w_1} G^T_{1,k})^{-1} G_{1,k}, \tag{15}$$

and that the null space component $y^{\mathcal{N}}_k = \Delta w_{2,k}$ solves the QP subproblem

$$\min_{y^{\mathcal{N}}} \ (\nabla_{w_2} F_k + D^T_k \nabla_{w_1} F_k)^T y^{\mathcal{N}} + \frac{1}{2} y^{\mathcal{N}T} B^{\mathcal{N}}_k y^{\mathcal{N}}$$

$$\text{subject to} \quad \begin{cases} G_{2,k} + \nabla_{w_1} G^T_{2,k} y^{\mathcal{R}}_k + (\nabla_{w_2} G^T_{2,k} + \nabla_{w_1} G^T_{2,k} D_k) y^{\mathcal{N}} = 0 \\ H_k + \nabla_{w_1} H^T_k y^{\mathcal{R}}_k + (\nabla_{w_1} H^T_k + \nabla_{w_1} H^T_k D_k) y^{\mathcal{N}} \geqslant 0 \end{cases}. \tag{16}$$

In the quadratic objective of Eq. (16), we have neglected a cross term which involves both $y^{\mathcal{R}}_k$ and $y^{\mathcal{N}}_k$, and $B^{\mathcal{N}}_k$ denotes a variable metric approximation to the (partially) reduced Hessian

$$S^{\mathcal{N}T}_k \nabla^2_w L_k S^{\mathcal{N}}_k = D^T_k \nabla^2_{w_1} L_k D_k + 2 \nabla_{w_1} (\nabla_{w_2} L_k) D_k + \nabla^2_{w_2} L_k. \tag{17}$$

RSQP methods of this type have been shown to be two-step q-superlinearly convergent if the reduced Hessian approximations $B^{\mathcal{N}}_k$ are generated by an asymptotically correct update scheme (e.g., Nocedal & Overton, 1985). There also exist one-step superlinearly convergent RSQP variants which either ensure that the omitted cross term actually becomes small (at the cost of

one additional function or gradient evaluation at an intermediate point) or explicitly include some approximation of this term (e.g., Biegler et al., 1995).

Consistent first-order estimates $\tilde{\lambda}_{1,k}$ of the Lagrange multipliers corresponding to the constraints G_1 can be obtained from

$$\tilde{\lambda}_{1,k} = (\nabla_{w_1} G_{1,k})^{-1}(\nabla_{w_1} F_k - \nabla_{w_1} G_{2,k}\tilde{\lambda}_{2,k} - \nabla_{w_1} H_k \tilde{\mu}_k), \quad (18)$$

where $\tilde{\lambda}_{2,k}$ and $\tilde{\mu}_k$ denote the multipliers of the reduced QP subproblem (Eq. (16)). The latter formula is motivated by the fact that $\nabla_w L(w^*, \lambda^*, \mu^*) = 0$.

There are various possibilities to approximate the reduced Hessian $S_k^{\mathcal{N}T}\nabla_w^2 L_k S_k^{\mathcal{N}}$. In our context, we will make use of the *reduced Lagrangian gradient*

$$\nabla_{y^{\mathcal{N}}} L(w_k + S_k^{\mathcal{N}} y^{\mathcal{N}}, \lambda, \mu)|_{y^{\mathcal{N}}=0} = S_k^{\mathcal{N}T}\nabla_w L(w_k, \lambda, \mu) \quad (19)$$

to directly calculate a variable metric approximation $B_k^{\mathcal{N}}$ of the reduced Hessian.

Starting from an initial guess $(w_{1,0}, w_{2,0})$, the PRSQP method for solving Eq. (12) calculates the sequence of iterates

$$\begin{pmatrix} w_{1,k+1} \\ w_{2,k+1} \end{pmatrix} = \begin{pmatrix} w_{1,k} \\ w_{2,k} \end{pmatrix} + \alpha_k \begin{pmatrix} \Delta w_{1,k} \\ \Delta w_{2,k} \end{pmatrix}, \quad k = 0, 1, \ldots, \quad (20)$$

where the relaxation factor $\alpha_k \in]0, 1]$ is determined by a line search to force global convergence.

The PRSQP approach allows a very flexible tailoring of the decomposition to large, but highly structured optimization problems like Eqs. (10a), (10b), (10c), (10d) and (10e). While the low-level sparsity structure which may be present in the constraint Jacobian can be directly exploited through the use of coordinate bases, it is possible to conserve certain other structural features of the optimization problem (e.g., the partial separability of the Lagrangian function) in the reduction process. Hence this additional high-level structure can be taken advantage of when formulating and solving the reduced QP subproblem.

3.2. A partitioned coordinate bases approach

The index one assumption for the DAE model (Eq. (1b)) implies that the algebraic variables s_{ij}^z are completely determined through the consistency conditions for given values of the differential variables s_{ij}^x and parameters $\hat{q}_{ij} = (q_{ij}, p_{ij}, v_{ij})$. This observation provides the basis for using s_{ij}^z as dependent variables in a 'natural' decomposition of the discretized dynamic optimization problem. Before outlining our decomposition approach, a few additional points should be made. We first rewrite the continuity and consistency conditions (Eq. (10b)) in the more compact form

$$x_i(\tau_{i,j+1}; s_{ij}^x, s_{ij}^z, \hat{q}_{ij}) - s_{i,j+1}^x = 0 \quad (21a)$$

$$\hat{g}_i(s_{ij}^x, s_{ij}^z, \hat{q}_{ij}, \tau_{ij}) = 0. \quad (21b)$$

These conditions lead to a characteristic 'multiple shooting structure' in the constraint Jacobian of the structured NLP problem (Eqs. (10a), (10b), (10c), (10d) and (10e)),

$$\begin{pmatrix} \ddots & & & & & & & \\ & X_{ij}^x & X_{ij}^z & X_{ij}^{\hat{q}} & -\mathrm{I} & & & \\ & \hat{G}_{ij}^x & \hat{G}_{ij}^z & \hat{G}_{ij}^{\hat{q}} & & & & \\ & & & & X_{i,j+1}^x & X_{i,j+1}^z & X_{i,j+1}^{\hat{q}} & -\mathrm{I} \\ & & & & \hat{G}_{i,j+1}^x & \hat{G}_{i,j+1}^z & \hat{G}_{i,j+1}^{\hat{q}} & \\ & & & & & & & \ddots \end{pmatrix}, \quad (22)$$

shown here only for two successive interior discretization points j and $j+1$ on model stage i. (Stage transition conditions (Eq. (10d)) lead to exactly the same general structure although they do not involve state and sensitivity integrations.) At this point, it is important to realize that the blocks

$$X_{ij}^z := \frac{\partial}{\partial s_{ij}^z} x_i(\tau_{i,j+1}; s_{ij}^x, s_{ij}^z, \hat{q}_{ij})$$

and also X_{ij}^x, $X_{ij}^{\hat{q}}$ are dense, and their calculation is quite expensive since it requires a state and sensitivity integration on subinterval $[\tau_{ij}, \tau_{i,j+1}]$. In contrast, the blocks

$$\hat{G}_{ij}^z := \frac{\partial}{\partial s_{ij}^z} \hat{g}_i(s_{ij}^x, s_{ij}^z, \hat{q}_{ij}, \tau_{ij})$$

as well as $\hat{G}_{ij}^x, \hat{G}_{ij}^{\hat{q}}$ are usually sparse and can be calculated cheaply. Since many DAE process models have a large share of algebraic variables, there is a great incentive to exploit the sparsity of the consistency condition Jacobians $\hat{G}_{ij}^x, \hat{G}_{ij}^z, \hat{G}_{ij}^{\hat{q}}$, and to avoid the expensive calculation of the full set of differential state sensitivities $X_{ij}^x, X_{ij}^z, X_{ij}^{\hat{q}}$. While the former is a standard requirement in large-scale decompositions, the latter is not usually encountered since it relates to very specific features of problem (Eqs. (10a), (10b), (10c), (10d) and (10e)), namely, the occurrence of dense and 'expensive' blocks in the constraint Jacobian (Eq. (22)).

Let us now give a formal statement of the PRSQP decomposition which is employed by MUSCOD-II. Our aim is to eliminate all algebraic variables from the problem by projecting onto the tangent space of the consistency conditions, thus performing the optimization (and the approximation of the Hessian) only within this tangent space. Based on the ordering of variables

$$w = \begin{pmatrix} w_{0,0} \\ w_{0,1} \\ \vdots \\ w_{\hat{M}-1,\hat{m}} \end{pmatrix} \quad \text{with} \quad w_{ij} := \begin{pmatrix} s_{ij}^x \\ s_{ij}^z \\ \hat{q}_{ij} \end{pmatrix},$$

(where $\hat{M}$ denotes the total number of stages, including nontrivial stage transitions) we choose the *partitioned* coordinate bases

$$S^{\mathcal{N}} = \begin{pmatrix} S^{\mathcal{N}}_{0,0} & & & \\ & S^{\mathcal{N}}_{0,1} & & \\ & & \ddots & \\ & & & S^{\mathcal{N}}_{\hat{M}-1,\hat{m}} \end{pmatrix},$$
$$S^{\mathcal{R}} = \begin{pmatrix} S^{\mathcal{R}}_{0,0} & & & \\ & S^{\mathcal{R}}_{0,1} & & \\ & & \ddots & \\ & & & S^{\mathcal{R}}_{\hat{M}-1,\hat{m}} \end{pmatrix} \quad (23a)$$

with diagonal blocks

$$S^{\mathcal{N}}_{ij} = \begin{pmatrix} I & 0 \\ -(\hat{G}^z_{ij})^{-1}\hat{G}^x_{ij} & -(\hat{G}^z_{ij})^{-1}\hat{G}^{\hat{q}}_{ij} \\ 0 & I \end{pmatrix}, \quad S^{\mathcal{R}}_{ij} = \begin{pmatrix} 0 \\ I \\ 0 \end{pmatrix}. \quad (23b)$$

Observe that $S^{\mathcal{N}}_{ij}$ and $S^{\mathcal{R}}_{ij}$ may be interpreted as *local* bases which act only on the subset of variables belonging to discretization point (i, j). (These bases always exist since $\hat{G}^z_{ij}$ is nonsingular due to the index one assumption.) Upon substitution of

$$\Delta w = S^{\mathcal{N}} y^{\mathcal{N}} + S^{\mathcal{R}} y^{\mathcal{R}} \quad (24)$$

into the full-space QP subproblem which comes from the structured NLP problem (Eqs. (10a), (10b), (10c), (10d) and (10e)), we find that the range space component $y^{\mathcal{R}}$ is directly given by

$$y^{\mathcal{R}} = \begin{pmatrix} y^{\mathcal{R}}_{0,0} \\ y^{\mathcal{R}}_{0,1} \\ \vdots \\ y^{\mathcal{R}}_{\hat{M}-1,\hat{m}} \end{pmatrix} \quad (25)$$

with $y^{\mathcal{R}}_{ij} = -(\hat{G}^z_{ij})^{-1}\hat{g}_i(s^x_{ij}, s^z_{ij}, \hat{q}_{ij}, \tau_{ij})$,

while the null space component $y^{\mathcal{N}}$ takes the form

$$y^{\mathcal{N}} = \begin{pmatrix} y^{\mathcal{N}}_{0,0} \\ y^{\mathcal{N}}_{0,1} \\ \vdots \\ y^{\mathcal{N}}_{\hat{M}-1,\hat{m}} \end{pmatrix}, \quad \text{where } y^{\mathcal{N}}_{ij} = \begin{pmatrix} \Delta s^x_{ij} \\ \Delta \hat{q}_{ij} \end{pmatrix}, \quad (26)$$

and is obtained as the solution of a reduced QP subproblem. This reduced QP problem—which still contains reduced linearizations of the inequalities and of all equalities except the consistency conditions—has the characteristic structure known from the ODE case, i.e., a *blockdiagonal Hessian* and a *block sparse constraint Jacobian*. Thus it is possible to employ the high-rank block updates of Bock & Plitt (1984) in a variable metric approximation of the partially reduced Hessian, and existing highly efficient block QP solvers can be used for determining $y^{\mathcal{N}}$ (Leineweber, 1999).

We briefly summarize the advantages of our specific choice of partitioned coordinate bases (Eqs. (23a) and (23b)) for problem (Eqs. (10a), (10b), (10c), (10d) and (10e)). These bases

- are guaranteed to exist for index one problems,
- provide a very useful reduction if there are many algebraic variables,
- can be cheaply calculated on the basis of sparse factorizations of $\hat{G}^z_{ij}$,
- allow to conserve the separability structure of the original problem, and
- are constructed from and act on *local* information only.

The final point implies that the whole process of generating the reduced QP subproblem (including function and gradient evaluation, basis construction, and projection) can be directly parallelized, (Leineweber et al., 2002).

4. Generation of the reduced QP subproblem

4.1. Intertwining of constraint linearization and projection

We have argued that it should be avoided to calculate all differential state sensitivities $X^x_{ij}, X^z_{ij}, X^{\hat{q}}_{ij}$ when generating the reduced QP subproblem. In order to see how this can be accomplished, let us once again have a closer look at the linearizations of conditions (Eqs. (21a) and (21b)),

$$X^x_{ij}\Delta s^x_{ij} + X^z_{ij}\Delta s^z_{ij} + X^{\hat{q}}_{ij}\Delta\hat{q}_{ij} - \Delta s^x_{i,j+1} = -(x_{ij} - s^x_{i,j+1}) \quad (27a)$$
$$\hat{G}^x_{ij}\Delta s^x_{ij} + \hat{G}^z_{ij}\Delta s^z_{ij} + \hat{G}^{\hat{q}}_{ij}\Delta\hat{q}_{ij} = -\hat{g}_{ij}, \quad (27b)$$

where we have used the shorthands $x_{ij} = x_i(\tau_{i,j+1}; s^x_{ij}, s^z_{ij}, \hat{q}_{ij})$ and $\hat{g}_{ij} = \hat{g}_i(s^x_{ij}; s^z_{ij}, \hat{q}_{ij}, \tau_{ij})$. Since $\hat{G}^z_{ij}$ is nonsingular, the linearized consistency conditions (Eq. (27b)) can be solved for Δs^z_{ij} in terms of the remaining solution components Δs^x_{ij} and $\Delta\hat{q}_{ij}$, yielding

$$\Delta s^z_{ij} = -(\hat{G}^z_{ij})^{-1}(\hat{G}^x_{ij}\Delta s^x_{ij} + \hat{G}^{\hat{q}}_{ij}\Delta\hat{q}_{ij} + \hat{g}_{ij})$$
$$=: D^x_{ij}\Delta s^x_{ij} + D^{\hat{q}}_{ij}\Delta\hat{q}_{ij} + d^g_{ij}. \quad (28)$$

Of course, D^x_{ij}, $D^{\hat{q}}_{ij}$ and d^g_{ij} are determined by solving the linear equation system

$$\hat{G}^z_{ij}(D^x_{ij}|D^{\hat{q}}_{ij}|d^g_{ij}) = -(\hat{G}^x_{ij}|\hat{G}^{\hat{q}}_{ij}|\hat{g}_{ij}) \quad (29)$$

based on a sparse factorization of $\hat{G}^z_{ij}$, i.e., without calculating $(\hat{G}^z_{ij})^{-1}$. In MUSCOD-II, we employ Harwell subroutine MA48, an advanced direct sparse solver (Reid & Duff, 1993), for the solution of system (Eq. (29)).

Substituting Δs^z_{ij} from Eq. (28) into Eq. (27a), we arrive at the *reduced* linearized continuity conditions

$$(X^x_{ij} + X^z_{ij}D^x_{ij})\Delta s^x_{ij} + (X^{\hat{q}}_{ij} + X^z_{ij}D^{\hat{q}}_{ij})\Delta\hat{q}_{ij} - \Delta s^x_{i,j+1}$$
$$= -(x_{ij} - s^x_{i,j+1}) - X^z_{ij}d^g_{ij}, \quad (30)$$

to be included as equality constraints in the reduced QP subproblem. For an efficient computation of these

reduced linearizations it is now crucial to observe that

$$(X^x_{ij} + X^z_{ij} D^x_{ij}) =: \mathscr{V}^x_{ij}, \quad (X^{\hat{q}}_{ij} + X^z_{ij} D^{\hat{q}}_{ij}) =: \mathscr{V}^{\hat{q}}_{ij} \quad \text{and} \quad X^z_{ij} d^g_{ij} =: \mathscr{V}^g_{ij} \tag{31}$$

can be interpreted as *directional derivatives* of $x_i(\tau_{i,j+1}; s^x_{ij}, s^z_{ij}, \hat{q}_{ij})$. The corresponding directions in $(s^x_{ij}, s^z_{ij}, \hat{q}_{ij})$ space are given by the columns of

$$\mathscr{D}^x_{ij} := \begin{pmatrix} I \\ D^x_{ij} \\ 0 \end{pmatrix}, \quad \mathscr{D}^{\hat{q}}_{ij} := \begin{pmatrix} 0 \\ D^{\hat{q}}_{ij} \\ I \end{pmatrix}, \quad \text{and} \quad \mathscr{D}^g_{ij} := \begin{pmatrix} 0 \\ d^g_{ij} \\ 0 \end{pmatrix}, \tag{32}$$

respectively. While it is certainly possible to calculate all $n^x_i + n^z_i + n^{\hat{q}}_i$ differential state sensitivities $X^x_{ij}, X^z_{ij}, X^{\hat{q}}_{ij}$ and combine them according to (Eq. (31)), a direct computation of the required $n^x_i + n^{\hat{q}}_i + 1$ directional sensitivities $\mathscr{V}^x_{ij}, \mathscr{V}^{\hat{q}}_{ij}, \mathscr{V}^g_{ij}$ is much more efficient if n^z_i is large. Observe that the latter strategy necessitates *intertwining* the constraint linearization and projection steps to some degree (i.e., performing for each discretization point (i, j) first the linearization of consistency conditions, then the solution of linear system (Eq. (29)), and finally the directional linearization of continuity conditions). However, the directional sensitivity calculation can still be parallelized with respect to the discretization points (i, j) since $\mathscr{V}^x_{ij}, \mathscr{V}^{\hat{q}}_{ij}, \mathscr{V}^g_{ij}$ depend on strictly local information only.

A comparison of Eqs. (23a) and (23b) and Eq. (25) with Eq. (32) shows that the local bases $S^{\mathscr{N}}_{ij}$ and range space moves $S^{\mathscr{R}}_{ij} y^{\mathscr{R}}_{ij}$ of our PRSQP decomposition can simply be interpreted as

$$S^{\mathscr{N}}_{ij} = (\mathscr{D}^x_{ij} \quad \mathscr{D}^{\bar{q}}_{ij}), \quad S^{\mathscr{R}}_{ij} y^{\mathscr{R}}_{ij} = \mathscr{D}^g_{ij} \tag{33}$$

which provides the obvious connection with the general PRSQP framework presented before. A certain price must be paid, however, for *not* calculating the complete set of constraint gradients. In particular, we can no longer calculate first-order estimates of the Lagrange multipliers for the eliminated consistency conditions from a Lagrangian stationarity condition, compare Eq. (18), because this would require explicit knowledge of the differential state sensitivities X^z_{ij}. Consequently, it is not possible to employ update schemes which are based on the original Lagrangian gradient, and a partially multiplier-free globalization strategy must be used.

4.2. Efficient calculation of directional sensitivities

We now provide some background on the efficient calculation of the directional sensitivities $\mathscr{V}^x_{ij}, \mathscr{V}^{\hat{q}}_{ij}, \mathscr{V}^g_{ij}$ as realized by the BDF integrator DAESOL (Bauer et al., 1997). For the sake of lucidity, we restrict ourselves to the simplified DAE system

$$\mathrm{d}x_i/\mathrm{d}\tau = \hat{f}_i(x_i, z_i, \hat{q}_{ij}, \tau)$$

$$0 = \hat{g}_i(x_i, z_i, \hat{q}_{ij}, \tau) - \alpha_{ij}(\tau)\hat{g}_i(s^x_{ij}, s^z_{ij}, \hat{q}_{ij}, \tau_{ij}) \quad ,$$

$$\tau \in [\tau_{ij}, \tau_{i,j+1}] \tag{34}$$

with initial values $x_i(\tau_{ij}) = s^x_{ij}$ and $z_i(\tau_{ij}) = s^z_{ij}$. (For discussion of the general case (Eq. (5)) which includes the left-hand side matrix $B_i(\cdot)$, see Bauer, 2000).

The differential state sensitivities $X^x_{ij}, X^z_{ij}, X^{\hat{q}}_{ij}$ can be obtained by integrating the standard variational DAE system

$$\frac{\mathrm{d}}{\mathrm{d}\tau}(X^x_{ij}|X^z_{ij}|X^{\hat{q}}_{ij}) = \frac{\partial \hat{f}_i}{\partial x_i}(X^x_{ij}|X^z_{ij}|X^{\hat{q}}_{ij}) + \frac{\partial \hat{f}_i}{\partial z_i}(Z^x_{ij}|Z^z_{ij}|Z^{\hat{q}}_{ij}) + \frac{\partial \hat{f}_i}{\partial \hat{q}_{ij}}(0|0|I)$$

$$0 = \frac{\partial \hat{g}_i}{\partial x_i}(X^x_{ij}|X^z_{ij}|X^{\hat{q}}_{ij}) + \frac{\partial \hat{g}_i}{\partial z_i}(Z^x_{ij}|Z^z_{ij}|Z^{\hat{q}}_{ij}) + \frac{\partial \hat{g}_i}{\partial \hat{q}_{ij}}(0|0|I) - \alpha_{ij}(\tau)(\hat{G}^x_{ij}|\hat{G}^z_{ij}|\hat{G}^{\hat{q}}_{ij}) \tag{35a}$$

from the initial values

$$(X^x_{ij}|X^z_{ij}|X^{\hat{q}}_{ij})|_{\tau=\tau_{ij}} = (I|0|0),$$

$$(Z^x_{ij}|Z^z_{ij}|Z^{\hat{q}}_{ij})|_{\tau=\tau_{ij}} = (0|I|0) \tag{35b}$$

along with the original DAE (Eq. (34)). The algebraic state sensitivities $Z^x_{ij}, Z^z_{ij}, Z^{\bar{q}}_{ij}$ are not required on the optimization level, but must nevertheless be computed. Since the variational DAE (Eq. (35a)) is *linear* in the sensitivities, we can formally multiply from the right by the matrix

$$\mathscr{D}_{ij} := (\mathscr{D}^x_{ij} \quad \mathscr{D}^{\hat{q}}_{ij} \quad \mathscr{D}^g_{ij}) = \begin{pmatrix} I & 0 & 0 \\ D^x_{ij} & D^{\hat{q}}_{ij} & d^g_{ij} \\ 0 & I & 0 \end{pmatrix}. \tag{36}$$

Defining in addition the directional algebraic sensitivities

$$(Z^x_{ij} + Z^z_{ij} D^x_{ij}) =: \mathscr{W}^x_{ij}, \quad (Z^{\hat{q}}_{ij} + Z^z_{ij} D^{\hat{q}}_{ij}) =: \mathscr{W}^{\hat{q}}_{ij}, \quad \text{and} \quad Z^z_{ij} d^g_{ij} =: \mathscr{W}^g_{ij} \tag{37}$$

we obtain the 'directional' variational DAE

$$\frac{\mathrm{d}}{\mathrm{d}\tau}(\mathscr{V}^x_{ij}|\mathscr{V}^{\hat{q}}_{ij}|\mathscr{V}^g_{ij}) = \frac{\partial \hat{f}_i}{\partial x_i}(\mathscr{V}^x_{ij}|\mathscr{V}^{\hat{q}}_{ij}|\mathscr{V}^g_{ij}) + \frac{\partial \hat{f}_i}{\partial z_i}(\mathscr{W}^x_{ij}|\mathscr{W}^{\hat{q}}_{ij}|\mathscr{W}^g_{ij}) + \frac{\partial \hat{f}_i}{\partial \hat{q}_{ij}}(0|I|0)$$

$$0 = \frac{\partial \hat{g}_i}{\partial x_i}(\mathscr{V}^x_{ij}|\mathscr{V}^{\hat{q}}_{ij}|\mathscr{V}^g_{ij}) + \frac{\partial \hat{g}_i}{\partial z_i}(\mathscr{W}^x_{ij}|\mathscr{W}^{\hat{q}}_{ij}|\mathscr{W}^g_{ij}) + \frac{\partial \hat{g}_i}{\partial \hat{q}_{ij}}(0|I|0) - \alpha_{ij}(\tau)(0|0|-\hat{g}_{ij}) \tag{38a}$$

with initial values

$$(\mathscr{V}_{ij}^{x}|\mathscr{V}_{ij}^{\hat{q}}|\mathscr{V}_{ij}^{g})|_{\tau=\tau_{ij}} = (I|0|0),$$

$$(\mathscr{W}_{ij}^{x}|\mathscr{W}_{ij}^{\hat{q}}|\mathscr{W}_{ij}^{g})|_{\tau=\tau_{ij}} = (D_{ij}^{x}|D_{ij}^{\hat{q}}|d_{ij}^{g}). \quad (38b)$$

This transformed variational DAE system for the direct calculation of $\mathscr{V}_{ij}^{x}, \mathscr{V}_{ij}^{\hat{q}}, \mathscr{V}_{ij}^{g}$ is the one which is actually solved together with the original DAE system (Eq. (34)). DAESOL uses the principle of *internal numerical differentiation* (see e.g., Bock & Plitt (1984) or Bauer (2000)), which makes sure that the sensitivities are the exact derivatives of the *discretized* solution of the original DAE system. To our knowledge, DAESOL is currently the only BDF code capable of automatically generating and solving the directional system (Eqs. (38a) and (38b)) given a matrix of directions (Eq. (36)). Due to the use of sparse linear solvers like MA48 (Reid & Duff, 1993), an efficient order and step size control, and a special monitoring strategy which saves evaluations and factorizations of the iteration matrix, DAESOL allows an efficient integration of large nonlinear DAE systems (Bauer et al., 1997).

4.3. Approximation of the reduced Hessian

Since the decomposition based on the partitioned coordinate bases (Eqs. (23a) and (23b)) preserves the separability structure of the full-space problem—as mentioned, the partially reduced Hessian $S^{\mathscr{N}T}\nabla_w^2 L S^{\mathscr{N}}$ is still block diagonal—we may use a suitable generalization of the partitioned variable metric update strategy of Bock and Plitt (1984) for the direct approximation of $S^{\mathscr{N}T}\nabla_w^2 L S^{\mathscr{N}}$. Specifically, we apply the blockwise high-rank updates

$$(B_{ij}^{\mathscr{N}})_{k+1} = (B_{ij}^{\mathscr{N}})_k + U((B_{ij}^{\mathscr{N}})_k, (\delta_{ij}^{\mathscr{N}})_k, (\gamma_{ij}^{\mathscr{N}})_k), \quad (39)$$

where $(B_{ij}^{\mathscr{N}})_k$ is an approximation for $(S_{ij}^{\mathscr{N}T}\nabla_{w_{ij}}^2 L_{ij} S_{ij}^{\mathscr{N}})_k$. The update U is given by the heuristically modified BFGS formula of Powell (1978). Note that the exact partially reduced Hessian will often be indefinite, hence the heuristic modification of the update scheme is required to ensure positive definite Hessian approximations. In Eq. (39), we employ the step

$$(\delta_{ij}^{\mathscr{N}})_k := \alpha_k (y_{ij}^{\mathscr{N}})_k, \quad (40)$$

and the corresponding difference of reduced Lagrangian gradients is defined by

$$(\gamma_{ij}^{\mathscr{N}})_k := (S_{ij}^{\mathscr{N}})_{k+1}^T \nabla_{w_{ij}} L_{ij}((w_{ij})_{k+1}, \lambda_{k+1}, \mu_{k+1}) - (S_{ij}^{\mathscr{N}})_k^T \nabla_{w_{ij}} L_{ij}((w_{ij})_k, \lambda_{k+1}, \mu_{k+1}). \quad (41)$$

It should be noted that the reduced Lagrangian gradients which appear in Eq. (41) do *not* explicitly depend on the multipliers of the consistency conditions. Hence no difficulty arises in our method, where these multipliers are unknown. For the latter reason, most other possible definitions of $(\gamma_{ij}^{\mathscr{N}})_k$ (e.g., Nocedal & Overton, 1985) cannot be used.

5. A partially multiplier-free globalization strategy

Finally we address the changes required to adapt the globalization strategy to our partially multiplier-free PRSQP approach. To keep it simple, we use the generic partitioned NLP problem (Eq. (12)) as the basis for our presentation. Since we do not calculate the multiplier estimates $\tilde{\lambda}_{1,k}$ corresponding to the constraints G_1, we can no longer employ the l_1 exact penalty function P_P of Powell (1978) with heuristic penalty weights derived from the multiplier estimates. Instead, we define the slightly modified penalty function

$$\tilde{P}_P(w_k, \rho, \sigma, \tau) = F_k + \rho \sum_{i=1}^{m_1} |G_{1,k,i}| + \sum_{i=1}^{m-m_1} \sigma_i |G_{2,k,i}| + \sum_{j=1}^{l} \tau_j |\min(0, H_{k,j})| \quad (42)$$

which contains a single scalar penalty factor ρ for all constraint residuals $G_{1,k,i}$. It is easy to show that the directional derivative $D_{\Delta w_k}\tilde{P}_{P,k}$ with $\Delta w_k = S_k^{\mathscr{N}} y_k^{\mathscr{N}} + S_k^{\mathscr{R}} y_k^{\mathscr{R}}$ can be expressed as

$$D_{\Delta w_k}\tilde{P}_P(w_k, \rho, \sigma, \tau) = -y_k^{\mathscr{N}T} B_k^{\mathscr{N}} y_k^{\mathscr{N}} - \rho \|G_{1,k}\|_1 - \tilde{\lambda}_{1,k}^T G_{1,k} - \sum_{i \in \varepsilon_k^{\vee}} (\sigma_i - \tilde{\lambda}_{2,k,i})(-G_{2,k,i}) - \sum_{i \in \hat{\varepsilon}_k} (\sigma_i + \tilde{\lambda}_{2,k,i}) G_{2,k,i} - \sum_{j \in \mathscr{I}_k^{\vee}} (\tau_i + \tilde{\mu}_{k,j})(-H_{k,j}) - \sum_{j \in \hat{\mathscr{I}}_k} \tilde{\mu}_{k,j} H_{k,j}. \quad (43)$$

Clearly, for positive definite Hessian approximations $B_k^{\mathscr{N}}$, compatibility of the search direction Δw_k with the merit function $\tilde{P}_P$ is obtained if

$$\rho \geq \|\tilde{\lambda}_{1,k}\|_\infty, \quad \sigma_i \geq |\tilde{\lambda}_{2,k,i}|, \quad \tau_j \geq \tilde{\mu}_{k,j} \quad (44)$$

for all k. Furthermore, if the method converges to a KKT point (w^*, λ^*, μ^*) which satisfies the Jacobian uniqueness condition, w^* must be a local minimizer of $\tilde{P}_P$ as well, since (Eq. (44)) then automatically implies that

$$\rho \geq \|\tilde{\lambda}_1^*\|_\infty, \quad \sigma_i \geq |\tilde{\lambda}_{2,i}^*|, \quad \tau_j \geq \tilde{\mu}_j^*. \quad (45)$$

However, a weight selection scheme on the basis of Eq. (44) is not applicable in the present case due to the unavailability $\|\tilde{\lambda}_{1,k}\|_\infty$.

At this point it is important to note that the term $\tilde{\lambda}_{1,k}^T G_{1,k}$ in Eq. (43) can be calculated very cheaply in the form

$$\tilde{\lambda}_{1,k}^T G_{1,k} = -\nabla_{w_1} F_k^T y_k^{\mathscr{R}} + \tilde{\lambda}_{2,k}^T \nabla_{w_1} G_{2,k}^T y_k^{\mathscr{R}} + \tilde{\mu}_k^T \nabla_{w_1} H_k^T y_k^{\mathscr{R}}, \quad (46)$$

that is, based on directional derivatives only. This

immediately follows from the Lagrange multiplier formula (Eq. (18)). The gradient terms $\nabla_{w_1} G_{2,k}^T y_k^{\mathscr{R}}$ and $\nabla_{w_1} H_k^T y_k^{\mathscr{R}}$ must be computed anyway in order to formulate the reduced QP subproblem (Eq. (16)). If we now replace the condition $\rho \geqslant \|\tilde{\lambda}_{1,k}\|_\infty$ by the weaker requirement

$$\rho \|G_{1,k}\|_1 \geqslant |\tilde{\lambda}_{1,k}^T G_{1,k}|, \tag{47}$$

we still obtain compatibility of Δw_k and $\tilde{P}_P$, but unless additional precautions are taken, the resulting globalization strategy may fail to force convergence of $G_1 = 0$. In practice, however, the heuristic weight selection scheme

$$\rho_k = \begin{cases} \rho_{k-1} & \text{if } \rho_{k-1}\|G_{1,k}\|_1 \geqslant |\tilde{\lambda}_{1,k}^T G_{1,k}| + \hat{\rho}\|G_{1,k}\|_1 \\ \dfrac{|\tilde{\lambda}_{1,k}^T G_{1,k}|}{\|G_{1,k}\|_1} + 2\hat{\rho} & \text{otherwise} \end{cases} \tag{48}$$

(where $\hat{\rho}$ is some positive constant) can be safely used if in addition $\|G_{1,k}\|_1$ is monitored over a certain number t of subsequent iterations, and ρ_k is increased whenever

$$\|G_{1,k}\|_1 \geqslant \max(\|G_{1,k-1}\|_1, \|G_{1,k-2}\|_1, \ldots, \|G_{1,k-t}\|_1).$$

Observe that ρ_k is never decreased and eventually should become fixed. The other weights $\sigma_{k,i}$ and $\tau_{k,j}$ can be chosen according to a standard rule like the one of Powell (1978) since the corresponding multipliers $\tilde{\lambda}_{2,k,i}$ and $\tilde{\mu}_{k,j}$ are known. A similar multiplier-free scheme has been described by Biegler, Schmid, and Ternet (1997) for the classical RSQP case in order to completely avoid the calculation of multipliers.

The reduced-space variant of MUSCOD-II uses a watchdog line search globalization which is based on the above modified penalty function and penalty weights. Of course, the separability structure of problem (Eqs. (10a), (10b), (10c), (10d) and (10e)) allows to calculate a separate penalty weight ρ_{ij} for each discretization point (i, j). Performance on practical problems is further improved by relaxing the weight selection scheme (Eq. (48)) such that ρ_k is allowed to decrease when both $|\tilde{\lambda}_{1,k}^T G_{1,k}|/\|G_{1,k}\|_1$ and $\|G_{1,k}\|_1$ decrease.

6. Summary and conclusions

We have developed an efficient new decompositions approach for dynamic multistage optimization problems which involve large nonlinear DAE process models of index one. Based on multiple shooting and tailored PRSQP, our strategy provides a natural generalization of the classical full-space BVP approach for ODEs (Bock & Plitt, 1984), considerably extending the original range of applicability. The new reduced-space BVP approach has the following central features:

- use of fully adaptive DAE solvers (allows to decouple the integration accuracy from the discretization used on the optimization level)
- relaxed DAE formulation with damping factor (allows inconsistent values of state variables at multiple shooting nodes, improves robustness by forcing solution toward consistency during integration)
- use of sparse linear solvers at the optimization level as well as within integrator (allows to exploit sparsity of large DAE process models)
- direct computation of the reduced linearized constraint set via directional derivatives (allows to significantly decrease the computational work for the linearization of large problems having many algebraic variables)
- complete decoupling of state and sensitivity integration on different multiple shooting intervals (allows parallel function and gradient evaluations)
- combination of the advantages of simultaneous and sequential strategies (explicit discretization of the dynamic model, but still relatively small NLP problem)
- solution of general multistage problems with control and path constraints and multipoint boundary conditions
- large degree of compatibility with existing dynamic process simulators (only requires a DAE solver which is capable of efficiently generating directional sensitivities)

Our approach is well suited for large-scale chemical process optimization as long as the underlying DAE models do not contain too many differential states. Since both the directional sensitivity matrices and the diagonal Hessian blocks are dense, our current PRSQP strategy is limited to DAE models having no more than a few hundred differential states. There is, of course, no such limitation regarding the number of algebraic states.

While many industrial DAE process models have only a moderate number of differentials states, larger DAE models may result, e.g., from a spatial discretization of coupled partial differential and algebraic equations (PDAEs) by the method of lines. In these cases, a further reduction could be obtained by projecting away also the differential variables. As has been shown by Schlöder (1988), this is possible when the initial values are fixed. A corresponding completely reduced variant of our strategy is currently under development and will be described in a forthcoming publication.

References

Bauer, I., Finocchi, F., Duschl, W. J., Gail, H.-P., & Schlöder, J. P. (1997). Simulation of chemical reactions and dust destruction in protoplanetary accretion discs. *Astronomy and Astrophysics 317*, 273–289.

Bauer, I. (2000). Numerische Verfahren zur Lösung von Anfangswertaufgaben und zur Generierung von ersten and zweiten Ableitungen mit Anwendungen bei Optimierungsverfahren in Chemie and Verfahrenstechnik, *Ph.D. thesis*, University of Heidelberg.

Biegler, L. T. (1984). Solution of dynamic optimization problems by successive quadratic programming and orthogonal collocation. *Computers and Chemical Engineering 8*, 243–248.

Biegler, L. T., Nocedal, J., & Schmid, C. (1995). A reduced Hessian method for large-scale constrained optimization. *SIAM Journal of Optimization 5*, 314–347.

Biegler, L. T., Schmid, C., & Ternet, D. (1997). A multiplier-free, reduced Hessian method for process optimization. In L. T. Biegler, T. F. Coleman, A. R. Conn & F. N. Santosa (Eds.), *Large-scale optimization with applications—part II: optimal design and control*. Springer.

Bock, H. G. and Plitt, K. J. (1984). A multiple shooting algorithm for direct solution of optimal control problems. *Proceedings of the ninth IFAC world congress, Budapest*. Pergamon Press.

Bock, H. G., Eich, E., & Schlöder, J. P. (1988). Numerical solution of constrained least squares boundary value problems in differential-algebraic equations. In K. Strehmel (Ed.), *Numerical treatment of differential equations*. Leipzig: Teubner.

Cervantes, A. M., Wächter, A., Tutuncu, R., & Biegler, L. T. (2000). A reduced space interior point strategy for optimization of differential algebraic systems. *Computers and Chemical Engineering 24*, 39–51.

Leineweber, D. B. (1999). Efficient reduced SQP methods for the optimization of chemical processes described by large sparse DAE models. *Fortschritt-Berichte VDI*, Reihe 3, Nr. 613 (ISBN 3-18-361303-4). Düsseldorf: VDI Verlag GmbH.

Leineweber, D. B., Schäfer, A., Bock, H. G., and Schlöder, J. P. (2002). An efficient multiple shooting based reduced SQP strategy for large-scale dynamic process optimization—part II: software aspects and applications.

Locke, M. H., Westerberg, A. W., & Edahl, R. H. (1983). Improved successive quadratic programming optimization algorithm for engineering design problems. *American Institute of Chemical Engineers Journal 29*, 871–874.

Logsdon, J. S., & Biegler, L. T. (1989). Accurate solution of differential-algebraic optimization problems. *Industrial and Engineering Chemistry Research 28*, 1628–1639.

Logsdon, J. S., & Biegler, L. T. (1992). Decomposition strategies for large-scale dynamic optimization problems. *Chemical Engineering Science 47*, 851–864.

Nocedal, J., & Overton, M. L. (1985). Projected Hessian updating algorithms for nonlinearly constrained optimization. *SIAM Journal of Numerical Analysis 22*, 821–850.

Powell, M. J. D. (1978). A fast algorithm for nonlinearly constrained optimization calculations. In G. A. Watson (Ed.), *Numerical analysis, Dundee 1977* (Lecture notes in mathematics 630). Berlin: Springer.

Reid, J. and Duff, I. S. (1993). MA48, a Fortran code for direct solution of sparse unsymmetric linear systems of equations. *Report RAL-93 072*, Rutherford Appleton Laboratory.

Sargent, R. W. H. and Sullivan, G. R. (1977). The development of an efficient optimal control package. In J. Stoer (Ed.), *Proceedings of the eighth IFIP conference on optimization techniques (1977), part 2*. Heidelberg: Springer.

Schlöder, J. P. (1988). Numerische Methoden zur Behandlung hochdimensionaler Aufgaben der Parameteridentifizierung. *Bonner Mathematische Schriften 187*, University of Bonn.

Schulz, V. H. (1996). Reduced SQP methods for large-scale optimal control problems in DAE with application to path planning problems for satellite mounted robots. *Ph.D. thesis*, University of Heidelberg.

Vassiliadis, V. S., Sargent, R. W. H., & Pantelides, C. C. (1994). Solution of a class of multistage dynamic optimization problems. 1. Problems without path constraints. *Industrial and Engineering Chemistry Research 33*, 2111–2122.