Journal of Optimization Theory and Applications (2024) 203:985–1026
https://doi.org/10.1007/s10957-024-02541-z

Implicit Peer Triplets in Gradient-Based Solution Algorithms for ODE Constrained Optimal Control

Jens Lang[1] · **Bernhard A. Schmitt**[2]

Received: 16 April 2023 / Accepted: 15 September 2024 / Published online: 1 October 2024

Abstract
It is common practice to apply gradient-based optimization algorithms to numerically solve large-scale ODE constrained optimal control problems. Gradients of the objective function are most efficiently computed by approximate adjoint variables. High accuracy with moderate computing time can be achieved by such time integration methods that satisfy a sufficiently large number of adjoint order conditions and supply gradients with higher orders of consistency. In this paper, we upgrade our former implicit two-step Peer triplets constructed in [Algorithms, 15:310, 2022] to meet those new requirements. Since Peer methods use several stages of the same high stage order, a decisive advantage is their lack of order reduction as for semi-discretized PDE problems with boundary control. Additional order conditions for the control and certain positivity requirements now intensify the demands on the Peer triplet. We discuss the construction of 4-stage methods with order pairs (3, 3) and (4, 3) in detail and provide three Peer triplets of practical interest. We prove convergence of order $s-1$, at least, for s-stage methods if state, adjoint and control satisfy the corresponding order conditions. Numerical tests show the expected order of convergence for the new Peer triplets.

Keywords Implicit peer two-step methods · Nonlinear optimal control · Gradient-based optimization · First-discretize-then-optimize · Discrete adjoints

Mathematics Subject Classification 34H05 · 49J15 · 65L05 · 65L06

✉ Jens Lang
lang@mathematik.tu-darmstadt.de

Bernhard A. Schmitt
schmitt@mathematik.uni-marburg.de

[1] Technical University of Darmstadt, Dolivostraße 15, 64293 Darmstadt, Germany

[2] Philipps-Universität Marburg, Hans-Meerwein-Straße 6, 35043 Marburg, Germany

1 Introduction

The numerical solution of optimal control problems governed by time-dependent differential equations is still a challenging task in designing and analyzing higher-order time integrators. An essential solution strategy is the so called *first-discretize-then-optimize* approach, where the continuous control problem is first discretized into a nonlinear programming problem which is then solved by state-of-the-art gradient-based optimization algorithm. Nowadays this direct approach is the most commonly used method due to its easy applicability and robustness. Consistent gradients of the objective function are derived from control, state and adjoint variables given by first-order necessary conditions of Karush-Kuhn-Tucker type, and are used in an iterative minimization algorithm to calculate the wanted optimal control. In this solution strategy, the unique gain of higher-order time integrators is twofold: increasing the efficiency in computing time by the use of larger and fewer time steps and, even more important for large-scale problems, thus reducing at the same time the memory requirement caused by the necessity to store all variables for the computation of gradients.

There are one-step as well as multistep time integrators in common use to solve ODE constrained optimal control problems. Symplectic Runge-Kutta methods [5, 21, 26] and backward differentiation formulas [1, 4] are prominent classes, but also partitioned and implicit-explicit Runge-Kutta methods [14, 23] and explicit stabilized Runge–Kutta-Chebyshev methods [2] have been proposed. However, fully implicit one-step methods often request the solution of large systems of coupled stages and might suffer from serious order reduction due to their lower stage order. This is especially the case when they are applied to semi-discretized PDEs with general time-dependent boundary conditions arising from boundary control problems [18, 24]. Matlab codes for [18] are available at [19]. In general, further consistency conditions have to be satisfied [11, 20] in order to achieve higher classical order for the discrete adjoint variables. Multistep methods avoid order reduction and have a simple structure, but higher order comes with restricted stability properties and adjoint initialization steps are usually inconsistent approximations which may lead to severe losses of accuracy. Moreover, the appropriate approximation of initial values and its structural consequences for the adjoint variables are further unsolved inherent difficulties that have limited the application of higher-order multistep methods for optimal control problems in a first-discretize-then-optimize solution strategy.

Recently, we have proposed a new class of implicit two-step Peer triplets that aggregate the attractive properties of one- and multistep methods through the use of several stages of one and the same high order and at the same time avoid their deficiencies with the aid of a two-step form [16–18]. The incorporation of different but matching start and end steps increases the flexibility of Peer methods for the solution of optimal control problems, especially also allowing higher-order approximations of the control, adjoint variables and the gradient of the objective function.

The class of s-stage implicit two-step Peer methods was introduced in [27] in linearly implicit form to solve stiff ODEs of the form $y'(t) = f(y(t))$, $y_0 = y(0)$, $y \in \mathbb{R}^m$. Later, the methods were simplified as implicit two-step schemes

$$Y_n = (Q \otimes I_m)Y_{n-1} + h(R \otimes I_m)F(Y_n),\ n = 0, 1, \ldots, \tag{1}$$

where the constant time step $h > 0$ will be considered here. An application of (1) to large scale problems with Krylov solvers was discussed in [3]. The stage solutions $Y_n = (Y_{ni})_{i=1}^s$ are approximations of $(y(t_n + c_i h))_{i=1}^s$ with equal accuracy and stability properties, which motivates the attribute *peer* and is the key for avoiding order reduction. The off-step nodes $c_1, \ldots, c_s$ are associated with the interval $[0, 1]$ but some may lie outside. In (1), $R \in \mathbb{R}^{s\times s}$ is lower triangular and invertible, $Q \in \mathbb{R}^{s\times s}$, $I_m \in \mathbb{R}^{m\times m}$ is the identity matrix, and $F(Y_n) = (f(Y_{ni}))_{i=1}^s$. Note that the stages Y_{ni} can be successively computed for $i = 1, \ldots, s$ due to the triangular structure of R. Several variants of Peer methods have been developed and successfully applied to a broad class of differential equations, e.g. [9, 15, 22, 25, 28, 30].

The application of Peer methods to optimal control problems requires a couple of modifications. A first direct attempt in [31] was unsatisfactory, mainly due to the restricted, first-order approximation of the adjoint variables. In [16], we found that the general redundant formulation of the above standard Peer method,

$$(A \otimes I_m)Y_n = (B \otimes I_m)Y_{n-1} + h(K \otimes I_m)F(Y_n),\ n = 0, 1, \ldots, \tag{2}$$

with invertible lower triangular matrix $A \in \mathbb{R}^{s\times s}$ and diagonal matrix $K \in \mathbb{R}^{s\times s}$ is admittedly equivalent in terms of the state variables to (1), but surprisingly not for the adjoint variables with the same coefficients Q, R. The additional degrees of freedom given by K together with a careful design of start and end methods with different coefficient matrices laid the foundation for improved Peer triplets with higher-order convergence. Triplets with good stability properties could be found [17]. Formulation (2) will also be the starting point in this paper. In contrast to our former approach in [16, 17], where the control $u(t)$ has been eliminated and a boundary value problem has been solved, we now compute the optimal control in an iterative procedure, making use of gradients of the objective function.

An important advantage of the approximations of the state $Y_{ni} \approx y(t_n + c_i h)$ and the adjoint multipliers $P_{ni} \approx p(t_n + c_i h)$ in the discrete time points $t_n + c_i h$ (see the next Chapter for the details of the notation) with equal high accuracy is the opportunity to simply use the discrete control variables

$$U_{ni} = \Phi(Y_{ni}, P_{ni}) \approx u(t_n + c_i h),\ i = 1, \ldots, s,\ n = 0, 1, \ldots, \tag{3}$$

in a gradient-based optimization algorithm. Although, the function Φ is only implicitly given, the higher order of Y_{ni} and P_{ni} is directly transferable to the control vector U_{ni}. Interpolation in time is easily realizable, the data structure keeps simple. However, additional order conditions for the control derived from (3) and positivity requirements for column sums in the matrix triplet (K_0, K, K_N), where K_0 and K_N are the matrices for the start and end method, intensify the demands on the Peer methods in the triplet. The arising bottlenecks in the design caused by stronger entanglement of all matrices must be resolved by a more sophisticated analysis.

The paper is organized as follows. In Sect. 2, we formulate the optimal control problem and define its discretization. The gradient of the cost function is derived in

Sect. 3. Order conditions and their algebraic consequences are discussed in Sect. 4. Two classes of four-stage Peer triplets are studied in Sects. 5 and 6 and three triplets of practical interest are constructed. In Sect. 7, we give a detailed convergence proof for unconstrained controls. Numerical examples collected in Sect. 8 illustrate the theoretical findings. The paper concludes with a summary in Sect. 9.

2 The Optimal Control Problem and Its Discretization

We are interested in the numerical solution of the following ODE-constrained nonlinear optimal control problem:

$$\text{minimize } \mathcal{C}\big(y(T)\big) \tag{4}$$
$$\text{subject to } y'(t) = f\big(y(t), u(t)\big), \quad u(t) \in U_{ad},\ t \in (0, T], \tag{5}$$
$$y(0) = y_0, \tag{6}$$

where the state $y(t) \in \mathbb{R}^m$, the control $u(t) \in \mathbb{R}^d$, $f : \mathbb{R}^m \times \mathbb{R}^d \mapsto \mathbb{R}^m$, the objective function $\mathcal{C} : \mathbb{R}^m \mapsto \mathbb{R}$, and the set of admissible controls $U_{ad} \subset \mathbb{R}^d$ is closed and convex. Introducing for any $u \in U_{ad}$ the normal cone mapping

$$N_U(u) = \{w \in \mathbb{R}^d : w^T(v - u) \leq 0 \text{ for all } v \in U_{ad}\},$$

the first-order optimality conditions read [11, 32]

$$y'(t) = f\big(y(t), u(t)\big), \quad t \in (0, T], \quad y(0) = y_0, \tag{7}$$
$$p'(t) = -\nabla_y f\big(y(t), u(t)\big)^{\mathsf{T}} p(t), \quad t \in [0, T), \quad p(T) = \nabla_y \mathcal{C}\big(y(T)\big)^{\mathsf{T}},$$
$$-\nabla_u f\big(y(t), u(t)\big)^{\mathsf{T}} p(t) \in N_U\big(u(t)\big), \quad t \in [0, T]. \tag{8}$$

Under appropriate regularity conditions, there exists a local solution $(y^\star, u^\star)$ of the optimal control problem (4–6) and a Lagrange multiplier $p^\star$ such that the first-order optimality conditions (7–8) are necessarily satisfied at $(y^\star, p^\star, u^\star)$. If, in addition, the Hamiltonian $H(y, p, u) := p^{\mathsf{T}} f(y, u)$ satisfies a coercivity assumption, then these conditions are also sufficient [11]. The control uniqueness property introduced in [11] yields the existence of a locally unique minimizer $u = u(\hat{y}, \hat{p})$ of the Hamiltonian over all $u \in U_{ad}$, if $(\hat{y}, \hat{p})$ is sufficiently close to $(y^\star, p^\star)$.

Many other optimal control problems can be transformed to the Mayer form $\mathcal{C}(y(T))$ which only uses terminal solutions. For example, terms given in the Lagrange form

$$\mathcal{C}_L(y, u) := \int_0^T l(y(t), u(t))\, dt \tag{9}$$

can be equivalently reduced to the Mayer form by adding a new differential equation $y'_{m+1}(t) = l(y(t), u(t))$ and initial values $y_{m+1}(0) = 0$ to the constraints. Then (9) simply reduces to $y_{m+1}(T)$.

On a time grid $\{t_0, \ldots, t_N\} \subset [0, T]$ with fixed step size length $h = t_{n+1} - t_n$ Peer methods use s stage approximations $Y_{ni} \approx y(t_{ni})$ and $U_{ni} \approx u(t_{ni})$ per time step at points $t_{ni} = t_n + c_i h$, $i = 1, \ldots, s$, associated with fixed nodes $c_1, \ldots, c_s$. All s stages share the same properties like a common high stage order equal to the global order, preventing order reduction. Applying the two-step Peer method for $n \geq 1$ and an exceptional starting step for $n = 0$ to the problem (5–6), we get the discrete constraint nonlinear optimal control problem

$$\text{minimize } \mathcal{C}\big(y_h(T)\big)$$

$$\text{subject to } A_0 Y_0 = a \otimes y_0 + h K_0 F(Y_0, U_0), \tag{10}$$

$$A_n Y_n = B_n Y_{n-1} + h K_n F(Y_n, U_n), \; n = 1, \ldots, N, \tag{11}$$

with long vectors $Y_n = (Y_{ni})_{i=1}^s \in \mathbb{R}^{sm}$, $U_n = (U_{ni})_{i=1}^s \in \mathbb{R}^{sd}$, and $F(Y_n, U_n) = \big(f(Y_{ni}, U_{ni})\big)_{i=1}^s$. Further, $y_h(T) = (w^\top \otimes I) Y_N \approx y(T)$, $a, w \in \mathbb{R}^s$, $A_n, B_n, K_n \in \mathbb{R}^{s\times s}$, and $I \in \mathbb{R}^{m\times m}$ being the identity matrix. As a change to the introduction, we will use the same symbol for a coefficient matrix like A and its Kronecker product $A \otimes I$ as a mapping from the space $\mathbb{R}^{sm}$ to itself. Throughout the paper, e_i denotes the i-th cardinal basis vector and $\mathbb{1} := (1, \ldots, 1)^\top \in \mathbb{R}^s$, sometimes with an additional index indicating a different space dimension.

On each subinterval $[t_n, t_{n+1}]$, Peer methods may be defined by three coefficient matrices A_n, B_n, K_n, where A_n is assumed to be non-singular. For practical reasons, this general version will not be used. We choose a fixed Peer method $(A_n, B_n, K_n) \equiv (A, B, K)$, $n = 1, \ldots, N-1$, in the inner grid points with lower triangular non-singular A, which allows a consecutive computation of the solution vectors Y_{ni}, $i = 1, \ldots, s$, in (10), (11). Exceptional coefficients (A_0, K_0) and (A_N, B_N, K_N) in the first and last forward steps are taken to allow for a better approximation in the initial step and of the end conditions.

The first order optimality conditions now read

$$A_0 Y_0 = a \otimes y_0 + h K_0 F(Y_0, U_0), \tag{12}$$

$$A_n Y_n = B_n Y_{n-1} + h K_n F(Y_n, U_n), \; n = 1, \ldots, N, \tag{13}$$

$$A_N^\top P_N = w \otimes p_h(T) + h \nabla_Y F(Y_N, U_N)^\top K_N^\top P_N, \tag{14}$$

$$A_n^\top P_n = B_{n+1}^\top P_{n+1} + h \nabla_Y F(Y_n, U_n)^\top K_n^\top P_n, \; 0 \leq n \leq N-1, \tag{15}$$

$$-h \nabla_U F(Y_n, U_n)^\top K_n^\top P_n \in N_{U^s}\big(U_n\big), \; 0 \leq n \leq N. \tag{16}$$

Here, $p_h(T) = \nabla_y \mathcal{C}\big(y_h(T)\big)^\top$ and the Jacobians of F are block diagonal matrices of the form $\nabla_Y F(Y_n, U_n) = \text{diag}_i\big(\nabla_{Y_{ni}} f(Y_{ni}, U_{ni})\big)$ and $\nabla_U F(Y_n, U_n) = \text{diag}_i\big(\nabla_{U_{ni}} f(Y_{ni}, U_{ni})\big)$. The generalized normal cone mapping $N_{U^s}\big(U_n\big)$ is defined by

$$N_{U^s}(u) = \left\{ w \in \mathbb{R}^{sd} : w^T(v-u) \le 0 \text{ for all } v \in U^s_{ad} \subset \mathbb{R}^{sd} \right\}.$$

If $K_n := (\kappa^{[n]}_{ij})$ is diagonal and $\kappa^{[n]}_{ii} = 0$, then (16) is automatically satisfied for stage $i \in \{1, \dots, s\}$. Assuming otherwise $k_{ni} := e_i^T K_n^T 1\!\!1 \neq 0$, and defining

$$Q_{ni} := \frac{1}{k_{ni}} \sum_j \kappa^{[n]}_{ji} P_{nj},\ 0 \le n \le N,\ i = 1, \dots, s,$$

(16) can be equivalently reformulated as

$$-k_{ni}\, \nabla_{U_{ni}} f(Y_{ni}, U_{ni})^\top Q_{ni} \in N_U\big(U_{ni}\big),\ 0 \le n \le N,\ i = 1, \dots, s. \tag{17}$$

Note that $Q_{ni} = P_{ni}$, if the matrix K_n is diagonal. A severe new restriction on the Peer triplet comes from the need to preserve the correct sign in (17) requiring that $k_{ni} > 0$. Then, we can divide by it and the control uniqueness property guarantees the existence of a local minimizer U_{ni} of the Hamiltonian $H(Y_{ni}, Q_{ni}, U)$ over all $U \in U_{ad}$ since Q_{ni} can be seen as an approximation to the multiplier $P_{ni} \approx p(t_n + c_i h)$. Such positivity conditions also arise in the context of classical Runge–Kutta methods or W-methods, see e.g. [11, Theorem 2.1] and [20, Chapter 5.2].

The need to sacrifice the triangular resp. diagonal form of the matrix coefficients A_n, K_n in the boundary steps comes from the fact that the starting steps (12), and backwards (14) are single-step methods with s outputs. With a triangular form of A_0, A_N and K_0, K_N their first stages (backward for $n = N$) would represent simple implicit Euler steps with a local order limited to 2, see Section 5 in [16] for a discussion.

3 The Gradient of the Cost Function

We first introduce the vector of control values for the entire interval $[0, T]$

$$U = (U_{01}^\top, \dots, U_{0s}^\top, U_{11}^\top, \dots, U_{Ns}^\top)^\top \in \mathbb{R}^{sd(N+1)}$$

and let $\mathcal{C}(U) := \mathcal{C}(y_h(U))$ be the cost function associated with these controls. The first order system (12–16) provides a convenient way to compute the gradient of $\mathcal{C}(U)$ with respect to U. Following the approach in [12], we find

$$\nabla_{U_{ni}} \mathcal{C}(U) = h \nabla_{U_{ni}} f(Y_{ni}, U_{ni})^\top (e_i^\top K_n^\top \otimes I) P_n, \tag{18}$$

for $0 \le n \le N$ and $i = 1, \dots, s$, where all approximations (Y_n, P_n) are computable by a forward-backward marching scheme. The state variables Y_n are obtained from the discrete state Eqs. (12–13) for $n = 0, \dots, N$, using the given values of the control vector U. Then, using the updated values Y_n, one computes $p_h(T) = \nabla_y \mathcal{C}\big(y_h(T)\big)^\top$ with $y_h(T) = (w^\top \otimes I) Y_N$ before marching the steps (14–15) backwards for $n = N, \dots, 0$, solving the discrete costate equations for all P_n.

The gradients from (18) can now be employed in gradient-based optimization algorithms which have been developed extensively since the 1950s. Many good algorithms are now available to solve nonlinear optimization problems in an iterative procedure

$$U^{(k+1)} := U^{(k)} + \triangle U^{(k)}, \quad k = 0, 1, \ldots \tag{19}$$

starting from an initial estimate $U^{(0)}$ for the control vector. Evaluating the objective function, its gradient and, in some cases, its Hessian, an efficient update $\triangle U^{(k)}$ of the control can be computed. Based on the principle (19), several good algorithms have been implemented in commercial software packages like MATLAB, MATHEMATICA, and others. We will use the MATLAB routine `fmincon` in our numerical experiments. It offers several optimization algorithms including `interior-point` [7] and `trust-region-reflective` [8] for large-scale sparse problems with continuous objective function and first derivatives.

Since the optimal control $u(t)$ minimizes the Hamiltonian $H(y, p, u) = p^{\mathsf{T}} f(y, u)$, we may compute an improved approximation of the control by the following minimum principle:

$$U_{ni}^{\ddagger} = \underset{U \in U_{ad}}{\arg\min}\, H(Y_{ni}, P_{ni}, U), \quad 0 \le n \le N,\ i = 1, \ldots, s, \tag{20}$$

if Y_{ni} and P_{ni} are approximations of higher-order. We note that the function Φ in (3) provides the solution in (20), when H is replaced by its discrete approximation defined by the Peer triplet.

4 Order Conditions for the Peer Triplet in the Unconstrained Case

We recall the conditions for local order $r \le s$ for the forward schemes and order $q \le s$ for the adjoint schemes, see [16, 17]. These conditions use the Vandermonde matrices $V_q := (\mathbb{1}, \mathbf{c}, \mathbf{c}^2, \ldots, \mathbf{c}^{q-1}) \in \mathbb{R}^{s \times q}$ with the column vector of nodes $\mathbf{c} = (c_i)_{i=1}^s \in \mathbb{R}^s$, and the non-singular Pascal matrix $\mathcal{P}_q = \left(\binom{j-1}{i-1}\right) = \exp(\tilde{E}_q) \in \mathbb{R}^{q \times q}$, where $\tilde{E} = \left(i\delta_{i+1,j}\right) \in \mathbb{R}^{q \times q}$ is nilpotent. There are 5 conditions for the forward scheme and its adjoint method:

$$A_0 V_r = a e_1^{\mathsf{T}} + K_0 V_r \tilde{E}_r,\ n = 0, \tag{21}$$

$$A_n V_r = B_n V_r \mathcal{P}_r^{-1} + K_n V_r \tilde{E}_r,\ 1 \le n \le N, \tag{22}$$

$$w^{\mathsf{T}} V_r = \mathbb{1}^{\mathsf{T}}, \tag{23}$$

$$A_n^{\mathsf{T}} V_q = B_{n+1}^{\mathsf{T}} V_q \mathcal{P}_q - K_n^{\mathsf{T}} V_q \tilde{E}_q,\ 0 \le n \le N-1, \tag{24}$$

$$A_N^{\mathsf{T}} V_q = w \mathbb{1}^{\mathsf{T}} - K_N^{\mathsf{T}} V_q \tilde{E}_q,\ n = N. \tag{25}$$

We remind that the coefficient matrices from interior grid intervals belong to a standard scheme $(A_n, B_n, K_n) \equiv (A, B, K),\ 1 \le n \le N-1$. The whole triplet consists of 8 coefficient matrices $(A_0, K_0),\ (A, B, K),\ (A_N, B_N, K_N)$.

Next, we focus on the new optimality condition (16) in the unconstrained case with $N_U = \{0\}$. It reads stage-wise

$$\nabla_u f(Y_{nj}, U_{nj})^\top \sum_{i=1}^{s} P_{ni}\kappa_{ij}^{[n]} = 0,\ j = 1, \ldots, s,\ 0 \le n \le N. \tag{26}$$

Order conditions are obtained by Taylor expansions, where approximations are replaced by exact solutions $(y^\star(t_n + c_i h), p^\star(t_n + c_i h), u^\star(t_n + c_i h))$ and the (continuous) optimality condition

$$\nabla_u f\big(y^\star(t), u^\star(t)\big)^\top p^\star(t) = 0,\ t \in [0, T] \tag{27}$$

is used. Defining the partial sums $\exp_q(z) := \sum_{j=0,\ldots,q-1} z^j/j!$ with q terms, Taylor's theorem for the expansion of a smooth function $v(t)$, $v \in C^q[0, T]$, at $t_{nj} := t_n + c_j h$ may be written as

$$v(t_n + c_i h) = \exp_q((c_i - c_j)z)v|_{t=t_{nj}} + O(h^q),\ z := h\frac{d}{dt},$$

with some slight abuse of notation. Then, the corresponding expansion of the residuals in (26) for order $q+1$ gives for $j = 1, \ldots, s$,

$$\nabla_u f\big(y^\star(t_{nj}), u^\star(t_{nj})\big)^\top \sum_{i=1}^{s} \kappa_{ij}^{[n]} \exp_q\big((c_i - c_j)z\big) p^\star(t_{nj}) \overset{!}{=} O(z^q). \tag{28}$$

Lemma 4.1 *Let the solution $p^\star$ be smooth, $p^\star \in C^q[0, T]$, $C := diag(c_1, \ldots, c_s)$ the diagonal matrix containing the nodes, and assume that*

$$(\mathbf{c}^{l-1})^\top K_n - 1\!\!1^\top K_n C^{l-1} = 0,\ l = 2, \ldots, q, \tag{29}$$

for all $n = 0, \ldots, N$. Then, for $j = 1, \ldots, s$ it holds with k_{nj} from (17)

$$\sum_{i=1}^{s} \kappa_{ij}^{[n]} p^\star(t_{ni}) - k_{nj} p^\star(t_{nj}) = O(h^q), \tag{30}$$

$$\tau_{nj}^U := \nabla_u f(y^\star(t_{nj}), u^\star(t_{nj}))^\top \sum_{i=1}^{s} p^\star(t_{ni})\kappa_{ij}^{[n]} = O(h^q). \tag{31}$$

Proof The assumption (29) is obtained from (28) by removing mixed powers in the conditions $\sum_{i=1}^{s} \kappa_{ij}^{[n]}(c_i - c_j)^{l-1} = 0$ with the corresponding equations for lower degrees. In fact, these condition prove the stronger version (30) which will be needed below. Of course, (31) is a simple consequence due to (27). □

For the standard method with a diagonal matrix $K_n \equiv K$, $1 \le n < N$, the condition (24) is sufficient for adjoint local order q since (29) is satisfied trivially. However, for the more general matrices K_0, K_N required in the first and last forward steps, it has been shown in [17, Chapter 2.2.4] that additional conditions have to be satisfied due to an unfamiliar form of one-leg-type applied to the linear adjoint equation $p' = -J(t)p$ with $J(t) = \nabla_y f(y(t), u(t))^\mathsf{T}$. These conditions are now covered by (29) for $l = 2$. Since the control is unknown now, the additional constraint Eq. (26) leads to much stronger restrictions on the design of the whole Peer triplet. As discussed in connection with (17), we also require positive column sums in the boundary steps and non-negative ones in the standard scheme,

$$1\!\!1^\mathsf{T} K_0 > 0^\mathsf{T},\ \ 1\!\!1^\mathsf{T} K_N > 0^\mathsf{T},\ \ 1\!\!1^\mathsf{T} K \geq 0^\mathsf{T}. \tag{32}$$

In the special case of a Peer triplet of FSAL type as constructed in Sect. 5.2, we allow $e_1^\mathsf{T} K_0 = 0^\mathsf{T}$ since the corresponding control component U_{01} can be eliminated. Even in the unconstrained case, $N_U = \{0\}$, positivity (32) is required in order to preserve the positive definiteness of the Hesse matrix.

4.1 Combined Conditions for the Order Pair (r, q)

With $r, q \leq s$ the full set of conditions may lead to practical problems for deriving formal solutions with the aid of algebraic manipulation software due to huge algebraic expressions. Possible alternatives like numerical search procedures will suffer from the large dimension of the search space consisting of the entries of 8 coefficient matrices. Fortunately, many of these parameters may be eliminated temporarily by solving certain condensed necessary conditions first. Afterwards, the full set (21–25) may be more easily solved in decoupled form. The combined conditions will be formulated with the aid of the linear operator $X \mapsto \mathcal{L}_{q,r}(X) := \tilde{E}_q^\mathsf{T} X + X\tilde{E}_r,\ \ X \in \mathbb{R}^{q\times r}$.

We note that the map $\mathcal{L}_{q,r}$ is singular since $\tilde{E}_r$ is nilpotent and $\tilde{E}_r e_1 = 0$ for any $r \in \mathbb{N}$. Hence, the first entry of its image always vanishes

$$\left(\mathcal{L}_{q,r}\right)_{11} = 0. \tag{33}$$

The combined conditions are presented in the order in which they would be applied in practice, with the standard scheme (A, B, K) in the first place. In all these conditions the matrix $\mathcal{Q}_{q,r} := V_q^\mathsf{T} B V_r \mathcal{P}_r^{-1}$ plays a central role.

Lemma 4.2 *Let the matrices of the Peer triplet* (A_0, K_0)*,* (A, B, K)*, and* (A_N, B_N, K_N) *satisfy the order conditions* (21–25) *with* $q \leq r \leq s$*. Then, also the following equations hold,*

$$\mathcal{L}_{q,r}(V_q^\mathsf{T} K V_r) = \mathcal{P}_q^\mathsf{T} V_q^\mathsf{T} B V_r - V_q^\mathsf{T} B V_r \mathcal{P}_r^{-1} = \mathcal{P}_q^\mathsf{T} \mathcal{Q}_{q,r} \mathcal{P}_r - \mathcal{Q}_{q,r}, \tag{34}$$

$$\mathcal{L}_{q,r}(V_q^\mathsf{T} K_0 V_r) = \mathcal{P}_q^\mathsf{T} V_q^\mathsf{T} B V_r - V_q^\mathsf{T} a e_1^\mathsf{T} = \mathcal{P}_q^\mathsf{T} \mathcal{Q}_{q,r} \mathcal{P}_r - V_q^\mathsf{T} a e_1^\mathsf{T}, \tag{35}$$

$$\mathcal{L}_{q,r}(V_q^\mathsf{T} K_N V_r) = 1\!\!1_q 1\!\!1_r^\mathsf{T} - V_q^\mathsf{T} B V_r \mathcal{P}_r^{-1} = 1\!\!1_q 1\!\!1_r^\mathsf{T} - \mathcal{Q}_{q,r}. \tag{36}$$

Proof Considering the cases $1 \le n < N$ first, Eq. (22) is multiplied by V_q^T from the left and the transposed condition (24) by V_r from the right. This gives

$$V_q^\mathsf{T} A V_r = V_q^\mathsf{T} B V_r \mathcal{P}_r^{-1} + (V_q^\mathsf{T} K V_r)\tilde{E}_r,$$
$$V_q^\mathsf{T} A V_r = \mathcal{P}_q^\mathsf{T} V_q^\mathsf{T} B V_r - \tilde{E}_q^\mathsf{T}(V_q^\mathsf{T} K V_r).$$

Subtracting both equations eliminates A and yields (34). Equation (35) follows in the same way from (21) and (24) for $n = 0$. The end method has to satisfy three conditions. Multiplying (22) for $n = N$ again from the left by V_q^T and the transposed end condition (25) by V_r from the right and combining both eliminates A_N and gives

$$\mathcal{L}_{q,r}(V_q^\mathsf{T} K_N V_r) = 1\!\!1_q 1\!\!1_r^\mathsf{T} - V_q^\mathsf{T} B_N V_r \mathcal{P}_r^{-1} = 1\!\!1_q 1\!\!1_r^\mathsf{T} - \mathcal{Q}_{q,r} \tag{37}$$

since $V_r^\mathsf{T} w = 1\!\!1$ by (23). The third condition for B_N is (24) with $n = N - 1$. It reduces to $B_N^\mathsf{T} V_q \mathcal{P}_q = A^\mathsf{T} V_q + K^\mathsf{T} V_q \tilde{E}_q = B^\mathsf{T} V_q \mathcal{P}_q$, which means $V_q^\mathsf{T} B_N = V_q^\mathsf{T} B$. Hence, the matrix B_N in (37) may simply be replaced by B. □

Since the operator $\mathcal{L}_{q,r}$ is singular, solutions for (34–36) exist for special right-hand sides only. For instance, in (36), (35) the property (33) requires that

$$1 = e_1^\mathsf{T} \mathcal{Q}_{q,r} e_1 = 1\!\!1^\mathsf{T} B 1\!\!1 \text{ and } 1\!\!1_q^\mathsf{T} a = 1\!\!1^\mathsf{T} B 1\!\!1 = 1. \tag{38}$$

Also the map $X \mapsto \mathcal{P}_q^\mathsf{T} X \mathcal{P}_r - X$ in (34) is singular, but here, (33) imposes no restrictions since we always have $e_1^\mathsf{T}\big(\mathcal{P}_q^\mathsf{T} X \mathcal{P}_r - X\big)e_1 = 0$. However, many further restrictions are due to special structural properties of the matrices $V_q^\mathsf{T} K_n V_r$ which are the arguments of $\mathcal{L}_{q,r}$. The following algebraic background highlights the hidden Hankel structure of certain matrices.

A matrix $X = (x_{ij}) \in \mathbb{R}^{q \times r}$ is said to have Hankel form, if its elements are constant along anti-diagonals, i.e. if $x_{ij} = \xi_{i+j-1}$ for $1 \le i \le q$, $1 \le j \le r$. Some simple, probably well-known, properties are the following.

Lemma 4.3 (a) *If $K \in \mathbb{R}^{s \times s}$ is a diagonal matrix, then $V_q^\mathsf{T} K V_r$ has Hankel form.*
(b) *Congruence transformations with Pascal matrices and the operator $\mathcal{L}_{q,r}$ preserve Hankel form: If $X \in \mathbb{R}^{q \times r}$ has Hankel form, then also $\mathcal{P}_q^\mathsf{T} X \mathcal{P}_r$ and $\mathcal{L}_{q,r}(X)$.*
(c) *The operator $\mathcal{L}_{q,r}$ is homogeneous for multiplication with Pascal matrices:*

$$\mathcal{P}_q^\mathsf{T} \mathcal{L}_{q,r}(X) = \mathcal{L}_{q,r}(\mathcal{P}_q^\mathsf{T} X),\ \mathcal{L}_{q,r}(X)\mathcal{P}_r = \mathcal{L}_{q,r}(X\mathcal{P}_r), \quad X \in \mathbb{R}^{q \times r}.$$

Proof (a) For $K = \operatorname{diag}(\kappa_{ii})$, we get $(\mathbf{c}^{i-1})^\mathsf{T} K \mathbf{c}^{j-1} = \sum_{\ell=1}^{s} \kappa_{\ell\ell} c_\ell^{i+j-2}$. (b) For $X = (\xi_{i+j-1})$ the explicit expression

$$e_i^\mathsf{T} \mathcal{P}_q^\mathsf{T} X \mathcal{P}_r e_j = \sum_{k=1}^{i+j-1} \binom{i+j-2}{k-1} \xi_k$$

should be well-known and is easily shown with the aid of the Vandermonde identity. And from $\tilde{E}_q = (i\delta_{i+1,j})$ it follows that

$$e_i^\mathsf{T}(\tilde{E}_q^\mathsf{T} X + X\tilde{E}_r)e_j = (i+j-2)\xi_{i+j-1},$$

where the factor $i+j-2=0$ leads to (33) for $i=j=1$. Assertion (c) holds, since $\tilde{E}_q$ and $\mathcal{P}_q = \exp(\tilde{E}_q)$ commute. □

4.2 Consequences of the One-Leg-Conditions

For convenience, the trivial case $l=1$ is included in the one-leg conditions (29) for adjoint order q. Recalling the matrix $C := \operatorname{diag}(c_i)$, these conditions may be written as

$$\left.\begin{array}{c}(\mathbf{c}^{l-1})^\mathsf{T} K_n = \mathbb{1}^\mathsf{T} K_n C^{l-1}\\ l=1,\dots,q\end{array}\right\} \iff V_q^\mathsf{T} K_n = \begin{pmatrix}\mathbb{1}^\mathsf{T} K_n\\ \mathbb{1}^\mathsf{T} K_n C\\ \vdots\\ \mathbb{1}^\mathsf{T} K_n C^{q-1}\end{pmatrix} \in \mathbb{R}^{q\times s}. \tag{39}$$

In particular, the matrix $V_q^\mathsf{T} K_n$ depends on the row vector $\mathbb{1}^\mathsf{T} K_n$ of the column sums of K_n only. This carries over to the matrices $V_q^\mathsf{T} K_n V_r$ in Lemma 4.2 and will lead to bottlenecks in the design of Peer triplets for $q+r > s+2$. A first simple restriction is shown now.

Lemma 4.4 *If a matrix $K_n \in \mathbb{R}^{s\times s}$ satisfies (39), then the matrices $V_q^\mathsf{T} K_n V_k$ and $\mathcal{L}_{q,k}(V_q^\mathsf{T} K_n V_k)$ have Hankel form for any $k \in \mathbb{N}$.*

Proof For $l \le q$ and $k \in \mathbb{N}$, Hankel form follows from (39) by

$$(\mathbf{c}^{l-1})^\mathsf{T} K_n \mathbf{c}^{k-1} = \mathbb{1}^\mathsf{T} K_n C^{l-1}\mathbf{c}^{k-1} = \mathbb{1}^\mathsf{T} K_n \mathbf{c}^{l+k-2}.$$

In fact, $V_q^\mathsf{T} K_n V_k = V_q^\mathsf{T} D_K V_k$ with the diagonal matrix D_K containing the column sums of K_n. The Hankel property of $\mathcal{L}_{q,k}(V_q^\mathsf{T} K_n V_k)$ then is a consequence of Lemma 4.3. □

Remark 4.5 By this Lemma, the conditions for the end method lead to severe restrictions on the standard method through Eq. (36). Since its left hand side $\mathcal{L}_{q,r}(V_r^\mathsf{T} K_N V_r)$ is in Hankel form, also its right hand side $\mathbb{1}_q\mathbb{1}_r^\mathsf{T} - \mathcal{Q}_{q,r}$, has to be so restricting the shape of B further. In particular it means that

$$\mathcal{Q}_{q,r} = V_r^\mathsf{T} B V_q \mathcal{P}_q^{-1} \text{ has Hankel form.} \tag{40}$$

Then, by Lemma 4.3, also $\mathcal{P}_q^\mathsf{T}\mathcal{Q}_{q,r}\mathcal{P}_r$ and $\mathcal{L}_{q,r}(V_q^\mathsf{T} K_0 V_r)$ for the starting method have Hankel form, which leaves only $V_q^\mathsf{T} a = e_1$ for the slack variable in (35).

Due to the one-leg conditions (39) the matrix equations (35), (36) collapse to heavily over-determined linear systems for the column sums $\mathbb{1}^\mathsf{T} K_0,\ \mathbb{1}^\mathsf{T} K_N$, implying additional restrictions for their right-hand sides, in particular for the matrix $\mathcal{Q}_{q,r}$. Looking at an equation

$$\mathcal{L}_{q,r}(V_q^\mathsf{T} K_n V_r) = \Theta := (\vartheta_{ij}) \in \mathbb{R}^{q\times r}$$

of the form of (35) or (36), $n \in \{0, N\}$, and considering element (l, k) of it, with (39) we get

$$\begin{aligned} &e_l^\mathsf{T}(\tilde{E}_q^\mathsf{T} V_q^\mathsf{T} K_n V_r + V_q^\mathsf{T} K_n V_r \tilde{E}_r)e_k \\ &= (l-1)(\mathbf{c}^{l-2})^\mathsf{T} K_n \mathbf{c}^{k-1} + (k-1)(\mathbf{c}^{l-1})^\mathsf{T} K_n \mathbf{c}^{k-2} \\ &= (\mathbb{1}^\mathsf{T} K_n)(l+k-2)\mathbf{c}^{l+k-3} \stackrel{!}{=} \vartheta_{lk}, \quad 1 \le l \le q,\ 1 \le k \le r. \end{aligned} \tag{41}$$

These are $q \cdot r$ conditions for the only s degrees of freedom in $\mathbb{1}^\mathsf{T} K_n$, where the index $j := l+k-2$ lies in the range $\{0, \dots, q+r-2\}$. Obviously, this system is only solvable if $\vartheta_{11} = 0$, see (33), and if ϑ_{lk} does only depend on $l+k$, which means that Θ has Hankel form. For $r+s-2 > s$, even more restriction will follow. These additional requirements will be collected farther down for different choices of q and r.

Lemma 4.6 *Let* (39) *hold and assume that the Eqs.* (35) and (36) *have solutions* K_0, K_N. *(a) If* $q+r \ge s+2$, *then the column sums* $\mathbb{1}^\mathsf{T} K_0, \mathbb{1}^\mathsf{T} K_N$ *are uniquely determined by the nodes and the matrix B of the standard scheme alone. These sums are solutions of the non-singular linear systems*

$$(\mathbb{1}^\mathsf{T} K_0) V_s D_s = \big(((\mathbf{c}+\mathbb{1})^{l-1})^\mathsf{T} B \mathbf{c}^{j-l+1}\big)_{j=1}^s, \tag{42}$$

$$(\mathbb{1}^\mathsf{T} K_N) V_s D_s = \mathbb{1}^\mathsf{T} - \big((\mathbf{c}^{l-1})^\mathsf{T} B (\mathbf{c}-\mathbb{1})^{j-l+1}\big)_{j=1}^s, \tag{43}$$

where $D_s = diag(1, 2, \dots, s)$. *In particular, solvability requires that the expressions on the right-hand sides of these equations do not depend on the index* l, $1 \le l \le \min\{q, j+1\}$.
(b) For $r = s$, *the column sums may be given by*

$$\mathbb{1}^\mathsf{T} K_0 = (\mathbf{c}+\mathbb{1})^\mathsf{T} B V_s D_s^{-1} V_s^{-1}, \tag{44}$$

$$\mathbb{1}^\mathsf{T} K_N = (\mathbb{1}^\mathsf{T} - \mathbf{c}^\mathsf{T} B V_s \mathcal{P}_s^{-1}) D_s^{-1} V_s^{-1}. \tag{45}$$

Proof (a) Omitting the trivial first equation $0 = \vartheta_{11}$ in (41), for $q+r-3 \ge s-1$ the next s cases with $1 \le l+k-2 \le s$ may be combined to the system

$$\mathbb{1}^\mathsf{T} K_n\big(1, 2\mathbf{c}, \dots, s\mathbf{c}^{s-1}\big) = \mathbb{1}^\mathsf{T} K_n V_s D_s \stackrel{!}{=} \big(\vartheta_{l,j-l+1}\big)_{j=1}^s,$$

$n \in \{0, N\}$, where the index l selects one of several possible choices. Since the matrix $V_s D_s$ is non-singular, solutions are unique, if they exist. Equations (42) and (43) are obtained with $\Theta = \mathcal{P}_q^\mathsf{T} V_q^\mathsf{T} B V_r - e_1 e_1^\mathsf{T}$ and $\Theta = 1\!\!1_q 1\!\!1_r^\mathsf{T} - \mathcal{Q}_{q,r}$, respectively, where (38) implies $\vartheta_{11} = 0$. (b) For $r=s$, the choice $l=2$ gives a complete row vector of length s in both equations. □

Remark 4.7 There are practical consequences for the design of Peer triplets. In the beginning, one may have hoped that positivity $1\!\!1^\mathsf{T} K_0 > 0^\mathsf{T}$, $1\!\!1^\mathsf{T} K_N > 0^\mathsf{T}$ could be obtained in the final design of the end methods with the aid of the many remaining degrees of freedom in the matrices K_0, K_N. However, the Lemma prohibits that. Instead, for $q + r \geq s + 2$ all row sums are determined by the standard method and the positivity restrictions may now be included in search procedures for the standard method alone having fewer degrees of freedom than the whole triplet.

For methods with forward order $r = s$ and adjoint order $q = s - 1$, it holds $q + r - 2 = 2s - 3 > s$ for $s \geq 4$. In this case there are still more requirements for the solvability of (41) which are discussed for the special case $(r, q) = (4, 3)$ in Sect. 6.2 below.

5 Four-Stage Triplets for the Order Pair $(r, q) = (3, 3)$

The restriction to positive column sums of all matrices K_n reduces the degrees of freedom in the Peer triplets. In order to regain flexibility in their design, we start with the order pair $(r, q) = (3, 3)$ with smaller order $r = 3$ than in [17]. Since Lemma 4.6 represents a bottleneck in the design process, we have to consider the boundary methods first. Now, the large number of parameters for the boundary methods and huge algebraic expressions bring the formal elimination of the 3 original order conditions for the end method to its limits. One may circumvent these difficulties by solving the combined condition (36) formally (with free parameters) for K_N first and then the other conditions in a step-by-step fashion as follows:

$$\text{solve} \begin{cases} \mathcal{L}_{3,3}(V_3^\mathsf{T} K_N V_3) = 1\!\!1_3 1\!\!1_3^\mathsf{T} - \mathcal{Q}_{3,3} & \text{for } K_N \\ A_N^\mathsf{T} V_3 = w 1\!\!1_3^\mathsf{T} - K_N^\mathsf{T} V_3 \tilde{E}_3 & \text{for } A_N, \\ B_N^\mathsf{T} V_3 = B^\mathsf{T} V_3 & \text{for } B_N. \end{cases}$$

The column sums $1\!\!1^\mathsf{T} K_0,\ 1\!\!1^\mathsf{T} K_N$ of the boundary methods are still uniquely determined by the standard method through (41) since $q + r - 2 = 4 = s$. We may write such a system with $n \in \{0, N\}$ in the form

$$\mathbb{R}^{1\times 4} \ni 1\!\!1^\mathsf{T} K_n(1\!\!1, 2\mathbf{c}, 3\mathbf{c}^2, 4\mathbf{c}^4) = \begin{pmatrix} \vartheta_{12} & \vartheta_{13} & * & * \\ \vartheta_{21} & \vartheta_{22} & \vartheta_{23} & * \\ * & \vartheta_{31} & \vartheta_{32} & \vartheta_{33} \end{pmatrix}. \tag{46}$$

This over-determined system with $s = 4$ columns is written with some abuse of notation, where potentially conflicting entries on the right-hand side are stacked in the same column. Asterisks indicate slack variables where no conditions have to be satisfied. Solutions will exist only if entries within each column have the same value.

This corresponds to the Hankel form of $\Theta \in \mathbb{R}^{3\times 3}$, which means Hankel form of $\mathcal{Q}_{3,3}$ for $n = N$. Then, Hankel form of $\mathcal{P}_3^\mathsf{T} V_3^\mathsf{T} B V_3 = \mathcal{P}_3^\mathsf{T} \mathcal{Q}_{3,3}\mathcal{P}_3$ for $n = 0$ follows from Lemma 4.3. Since the matrix $(1\!\!1, 2\mathbf{c}, 3\mathbf{c}^2, 4\mathbf{c}^4) = V_4 D_4$ is non-singular, the row sums $1\!\!1^\mathsf{T} K_n$ are uniquely determined by (46). A possible choice of elements ϑ_{ij} is used in the following equation which enforces the positivity conditions (32) with the representations

$$1\!\!1^\mathsf{T} K_n = (\vartheta_{21}, \vartheta_{22}, \vartheta_{23}, \vartheta_{33}) D_4^{-1} V_4^{-1} \overset{!}{\geq} \kappa_* 1\!\!1^\mathsf{T},\ n \in \{0, N\}, \tag{47}$$

with $\Theta = \mathcal{P}_3^\mathsf{T} \mathcal{Q}_{3,3}\mathcal{P}_4$ for $n = 0$ and $\Theta = 1\!\!1_3 1\!\!1_3^\mathsf{T} - \mathcal{Q}_{3,3}$ for $n = N$.

Considering the good performance of a triplet based on the backward difference formula BDF4 in [17, 18], it is of interest to look for a version with boundary methods satisfying the additional one-leg-conditions (29) or (39) for $q = 3$. According to (40), Hankel form of $\mathcal{Q}_{3,r}$ is required now. However, the matrix $\mathcal{Q}_{4,4}$ for the BDF standard method has the Hankel property in its first 4 anti-diagonals only with $e_1^\mathsf{T} \mathcal{Q}_{4,4} = (1, \frac{1}{8}, \frac{1}{96}, 0)$. Hence, only $\mathcal{Q}_{3,3}$ has Hankel form since it contains only one element from the fifth anti-diagonal. Now, for a method with $q = r = 3$, the column sums of the boundary methods are explicitly determined by the standard method through (46). Here, for BDF the end method satisfies (47) with $\kappa_* = 55/576$, but not the starting method since $1\!\!1^\mathsf{T} K_0 e_2 = -21/64 < 0$. Hence, no positive triplet based on BDF4 exists satisfying the one-leg-condition with $q \geq 3$.

5.1 Method `AP4o33pa`

So far, we have only discussed the normal order conditions for the Peer methods and their adjoints and the resulting restrictions. For methods to be efficient further requirements have to be considered. First, the conditions (22) and (24) for the pair (r, q) relate to the (local) orders of consistency. In order to also establish convergence of (global) order $O(h^r)$ and $O(h^q)$ in [17], the following two conditions for super-convergence of the forward and adjoint scheme have been added,

$$1\!\!1^\mathsf{T}\big(A\mathbf{c}^r - B(\mathbf{c} - 1\!\!1)^r - rK\mathbf{c}^{r-1}\big) = 0, \tag{48}$$

$$1\!\!1^\mathsf{T}\big(A^\mathsf{T}\mathbf{c}^q - B^\mathsf{T}(\mathbf{c} + 1\!\!1)^q + qKc^{q-1}\big) = 0, \tag{49}$$

which cancel the leading term in the global error. In practice, the super-convergence effect may be observed for a sufficiently fast damping of secondary modes of the stability matrix only and requires that

$$|\lambda_2(A^{-1}B)| \leq \gamma < 1, \quad \gamma \cong 0.8, \tag{50}$$

where λ_2 denotes the absolutely second largest eigenvalue of the stability matrix $\bar{B} := A^{-1}B$ of the standard scheme. We note, that the stability matrix $\tilde{B}^\mathsf{T} := (BA^{-1})^\mathsf{T}$ of the adjoint time steps has the same eigenvalues as $\bar{B}$. In several numerical tests in [17], the given value $\gamma = 0.8$ was sufficiently small to produce super-convergence reliably and it does so in our tests at the end. Super-convergence (48), (49) only cancels the leading error terms of the methods. In order to cover other essential error contributions also the norms

$$err_r := \frac{1}{r!}\|\mathbf{c}^r - A^{-1}B(\mathbf{c}-\mathbb{1})^r - rA^{-1}K\mathbf{c}^{r-1}\|_\infty,$$
$$err_q^\dagger := \frac{1}{q!}\|\mathbf{c}^q - A^{-\mathsf{T}}B^{\mathsf{T}}(\mathbf{c}+\mathbb{1})^q + qA^{-\mathsf{T}}Kc^{q-1}\|_\infty,$$

are monitored as the essential error constants, see [17]. Furthermore, the norm $\|A^{-1}B\|_\infty$ of the stability matrix is of interest since it may be a measure for the propagation of rounding errors.

Application of the Peer methods to stiff problems requires good stiff stability properties. $A(\alpha)$-stability is defined here by the requirement that the spectral radius $\varrho\big((A - zK)^{-1}B\big) < 1$ of the stability matrix of the standard scheme is below one for z in the open sector of the complex plain with aperture 2α centered at the negative real axis. The adjoint stability matrix $(A - zK)^{\mathsf{T}}B^{\mathsf{T}}$ and $\big(A - zK\big)^{-1}B$ possess the same eigenvalues. Details on the computation of α can be found in [16, §5.2]. The angles for the different methods are contained in Table 1.

Very mild eigenvalue restrictions for the boundary methods are also taken from [17]. In order to guarantee the solvability of the stage systems for the first and last steps we require

$$\mu_0 := \min_j \Re\lambda_j(K_0^{-1}A_0) > 0, \quad \mu_N := \min_j \Re\lambda_j(K_N^{-1}A_N) > 0. \tag{51}$$

A new requirement is the non-negativity condition (32) imposed by the use of a gradient-based method to update the control vector in (19). Lemma 4.6 has shown that the column sums of the boundary methods are already fully determined by the standard method and their positivity $\mathbb{1}^{\mathsf{T}}K_0 > 0^{\mathsf{T}}$, $\mathbb{1}^{\mathsf{T}}K_N > 0^{\mathsf{T}}$ can be included in the search for it. In practice, the performance of the gradient method (19) may suffer badly if the column sums have differing magnitudes. Hence, the search was narrowed to methods with moderate positive values of the **c**olumn **s**um **q**uotient

$$csq := \max\left\{\frac{\max_i |\mathbb{1}^{\mathsf{T}}K_0e_i|}{\min_i \mathbb{1}^{\mathsf{T}}K_0e_i}, \frac{\max_i |\mathbb{1}^{\mathsf{T}}K_Ne_i|}{\min_i \mathbb{1}^{\mathsf{T}}K_Ne_i}\right\} \overset{!}{>} 0,\ \mathbb{1}^{\mathsf{T}}K \geq 0^{\mathsf{T}}, \tag{52}$$

and (32) where $\mathbb{1}^{\mathsf{T}}K_0$, $\mathbb{1}^{\mathsf{T}}K_N$ are determined by the standard method, see (44), (45). Although there are rather tight restrictions on the column sums of K_0, K_N, there still exists a null-space in the conditions for these matrices and it was necessary to restrict the norms $\|K_0\|$, $\|K_N\|$ in addition to (51) in the final search for the boundary methods.

For easier reference, the full set of conditions is collected in Table 3. We remind that the conditions in line (d) there ensure the existence of the boundary methods and allow for the detached construction of the standard method alone.

Although searches for positive standard methods with $r = 3 = s - 1$ came very close to A-stability, no truly A-stable methods could be found. This might be due to

Table 1 Properties of the 4-stage standard methods of Peer triplets

Triplet	(r, q)	nodes	α	$\|A^{-1}B\|_\infty$	$\|\lambda_2\|$	err_r	$err_q^\dagger$	csq
AP4o33pa	(3, 3)	[0, 1.41]	89.90^o	8.2	0.66	0.050	0.046	33.4
AP4o33pfs	(3, 3)	[0, 1]	77.53^o	16.0	0.46	0.031	0.030	1.72
AP4o43p	(4, 3)	[0.1, 0.9]	59.78^o	8.5	0.58	0.0038	0.024	11.0

Table 2 Properties of the boundary methods of Peer triplets

	Starting method			End method			
Triplet	blksz	μ_0	$\varrho(BA_0^{-1})$	blksz	μ_N	$\varrho(A_N^{-1}B_N)$	$\varrho(B_N A^{-1})$
AP4o33pa	4	2.03	1	1+3	2.21	1	1
AP4o33pfs	1+3	4.92	1	1+3	1.61	1	1
AP4o43p	4	4.13	1	4	4.36	1	1.09

the restriction (50) on the sub-dominant eigenvalue and we suspect that a (formally) A-stable method might have a multiple eigenvalue 1 if it exists. In fact, with a rather unsafe restriction $|\lambda_2(A^{-1}B)| \le 0.9$, a method was found with stability angle 89.976^o extremely close to A-stability. Slightly relaxing the requirement on the angle to $\alpha = 89.90^o$, the following **a**lmost A-stable method `AP4o33pa` was constructed. Its node vector is given by

$$\mathbf{c}^{\mathsf{T}} = \left(\frac{46}{5253}, \frac{29}{51}, \frac{1723}{2193}, \frac{17131}{12189}\right) \doteq (0.00876, 0.5686, 0.7857, 1.4054)$$

with monotonic nodes. It is super-convergent with (48),(49) for $r = q = 3$ with a good damping factor $|\lambda_2(A^{-1}B)| < 0.66$. The error constants are almost equal with $err_3 = 0.298/3! \approx 0.050$ and $err_3^\dagger = 0.279/3! \approx 0.046$. Further data of this method are collected in Table 1. All boundary steps are zero-stable but only for the end method a block structure could be obtained with block sizes `blksz=3` $+ 1$. These data are presented in Table 1. The complete set of coefficients is given in the supplementary material A.1 (Online Resource 1).

5.2 FSAL Method `AP4o33pfs`

The number of stage equations to be solved numerically may be reduced with the aid of the FSAL property (*first stage as last*) frequently used in the design of one-step methods, where the last stage of the previous time step equals the first stage of the new step. For Peer methods, this property has been discussed in [28]. In our formulation (13) it means that

$$c_1 = 0,\ c_s = 1, \quad e_1^{\mathsf{T}} K_n = 0^{\mathsf{T}},\ e_1^{\mathsf{T}} A_n = a_{11}^{[n]} e_1^{\mathsf{T}},\ e_1^{\mathsf{T}} B_n = a_{11}^{[n]} e_s^{\mathsf{T}}, \qquad (53)$$

implying $Y_{n,1} = Y_{n-1,s} \cong y(t_n)$, $n \geq 1$. A convenient benefit of Peer methods is that, due to their high stage order, the interpolation of all s stages provides an accurate polynomial approximation of the solution being also continuous if the FSAL property holds.

Since the order conditions for methods of type `AP4o33*` leave a 10-parameter family of standard methods, the additional restrictions (53) can easily be satisfied for (A, B, K). However, some properties of the boundary methods imply further restrictions on the standard method through (46). Although the matrices K_0, K_N in the boundary steps are not restricted to diagonal form, the one-leg conditions (39) with $q = s - 1$ request that they are rank-1-changes of diagonal matrices only. Then, the condition $e_1^\mathsf{T} K_n = 0^\mathsf{T}$ leaves off-diagonal elements in their first columns only. However, for matrices of such a form, the constraint (26) reads

$$\nabla_u f(Y_{nj}, U_{nj})^\mathsf{T} P_{nj} \kappa_{jj}^{[n]} = 0, \; j = 2, \ldots, s,$$

$$\nabla_u f(Y_{n1}, U_{n1})^\mathsf{T} \sum_{i=2}^{s} P_{ni} \kappa_{i1}^{[n]} = 0.$$

These are s conditions on $P_{n2}, \ldots, P_{ns}$, which may not always be satisfiable since $\nabla_u f^\mathsf{T}$ is evaluated at different places. Hence, the condition $e_1^\mathsf{T} K_n = 0^\mathsf{T}$ requires that K_n is diagonal with zero as the first diagonal element. However, the property $K_n e_1 = 0$ also leads to $k_{n1} = 1\!\!1^\mathsf{T} K_n e_1 = 0$ and introduces via (46) one additional restriction on the matrix $\mathcal{Q}_{3,3}$ from the standard method both for K_0 and K_N. In condition (47), the first component, being zero, can be deleted since also the control U_{n1} is no longer present in (26).

Unfortunately, only in the starting step a diagonal matrix $K_0 \succeq 0$ with $\kappa_{11}^{[0]} = 0$ is possible leading to an exact start $Y_{01} = y_0$. However, no non-negative triplet seems to exist with a final FSAL step. Hence, K_N is chosen lower triangular, having a dense first column and the first row $e_1^\mathsf{T} K_N = \frac{1}{3} e_1^\mathsf{T}$, leading to a small jump $Y_{n1} - Y_{n-1,s} = O(h^3)$ at t_N only.

Computer searches found the method `AP4o33pfs` with stability angle $\alpha = 77.53^o$, having node vector $c^\mathsf{T} = (0, \frac{9}{86}, \frac{321}{602}, 1)$, error constants $err_3 = 0.187/3! \approx 0.031$, $err_3^\dagger = 0.180/3! \approx 0.030$ and a small damping factor $\gamma = 0.46$. More data are given in Tables 1 and 2. Of course, the computation of the quotient csq in (52) and the real part μ_0 in (51) was restricted to the nontrivial lower 3×3 block of K_0. The coefficients of `AP4o33pfs` are given in the supplementary material A.2 (Online Resource 1).

6 Four-Stage Triplets for the Order Pair $(r, q) = (s, q) = (4, 3)$

Obeying the additional order and super-convergence conditions for $r = s = 4$ means that the Peer method will have global order 4 for pure initial value problems without control. Unfortunately, in Sect. 7 we will not be able to prove the same improvement for the state solutions of the control problem if the adjoint equations remain at order $q = 3$. However, the numerical experience in our previous papers [17, 18] indicates that the coupling between both errors is often rather weak and the higher order may show up in numerical tests. The A-stable methods from [17] already used this combination of (r, q) for problems where u may be explicitly eliminated, leading to a boundary value problem for (y, p). However, these methods are not suited within a gradient-based optimization method discussed in Sect. 3 since K possesses negative diagonal elements. Still, two methods `AP4o43bdf` and `AP4o43dif` satisfy condition (32) and may be used in a gradient-based optimization. In our numerical tests, `AP4o43bdf` will be compared to the new methods derived now and which satisfy one set of one-leg conditions more, see (29). These stronger requirements on all methods of the triplet lead to a severe bottleneck in the augmented order conditions: any appropriate standard method (A, B, K), $K = \operatorname{diag}(\kappa_{ii})$, has a blind third stage with $\kappa_{33} = 0$. This observation is a consequence of equation (36) and Lemma 4.4. The full set of conditions will be collected at the end of this section.

6.1 Consequences of the Hankel Form of $\mathcal{Q}_{3,4}$

In this subsection, it is shown that the restriction $\kappa_{33} = 0$ is a consequence of only the forward order conditions (22) with $r = s = 4$ and the $q \times s = 3 \times 4$-Hankel form (40) of the matrix

$$\mathcal{Q}_{q,s} = V_q^{\mathsf{T}} B V_s \mathcal{P}_s^{-1} = V_q^{\mathsf{T}} (A V_s - K V_s \tilde{E}_s). \tag{54}$$

With the shift matrices $S_q = \big(\delta_{i,j-1}\big) \in \mathbb{R}^{q \times q}$ and the projection $\check{I}_q := I_q - e_q e_q^{\mathsf{T}}$, Hankel form of the matrix (54) is equivalent with

$$0 = \check{I}_q (S_q \mathcal{Q}_{q,s} - \mathcal{Q}_{q,s} S_s^{\mathsf{T}}) \check{I}_s. \tag{55}$$

Also, column shifts in the Vandermonde matrix have a simple consequence, $V_q S_q^{\mathsf{T}} = C V_q \check{I}_q$.

Theorem 6.1 *(a) Hankel form of the matrix* (54) *is equivalent with the condition*

$$V_{q-1}^{\mathsf{T}} (AC - CA - K) V_{s-1} = 0. \tag{56}$$

(b) For $q = 3$, $r = s = 4$, equations (54), (55) *imply* $\kappa_{33} = e_3^{\mathsf{T}} K e_3 = 0$.

Proof Since $\tilde{E}_r = D_r S_r = S_r (D_r - I_r)$ with $D_r = \operatorname{diag}_{i=1}^r(i)$, we have $\tilde{E}_r S_r^{\mathsf{T}} = D_r \check{I}_r$ and $C V_r \tilde{E}_r = V_r (D_r - I_r)$. Now, the Hankel condition (55) for the matrix (54) reads

$$\begin{aligned}
0 =& \check{I}_q \big(S_q V_q^{\mathsf{T}} (A V_s - K V_s \tilde{E}_s) - V_q^{\mathsf{T}} (A V_s - K V_s \tilde{E}_s) S_s^{\mathsf{T}}\big) \check{I}_s \\
=& \check{I}_q (V_q^{\mathsf{T}} C (A V_s - K V_s \tilde{E}_s) - V_q^{\mathsf{T}} A C V_s + V_q^{\mathsf{T}} K V_s D) \check{I}_s \\
=& \check{I}_q V_q^{\mathsf{T}} (C A V_s - A C V_s - K (C V_s \tilde{E}_s - V_s D)) \check{I}_s \\
=& \check{I}_q V_q^{\mathsf{T}} (C A - A C + K) V_s \check{I}_s.
\end{aligned}$$

Since the matrices $\check{I}_q$, $\check{I}_s$ simply eliminate the last row or column, this equation is equivalent with the assertion (56). b) For $q = 3$, $s = 4$, condition (56) consists of 6 equations. The commutator $[A, C] = AC - CA$ is strictly lower triangular since the diagonal of A cancels out. Ignoring the diagonal, the map $A \mapsto V_{q-1}^{\mathsf{T}}(AC - CA - K)V_{s-1}$ still has a rank deficiency if it is considered as a function of the 6 subdiagonal elements of A only. In fact, there exists a nontrivial kernel of its adjoint having rank-1 structure. Consider

$$V_3 \begin{pmatrix} c_1c_2 \\ -c_1 - c_2 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ (c_3 - c_1)(c_3 - c_2) \\ (c_4 - c_1)(c_4 - c_2) \end{pmatrix} =: x_R$$

and

$$V_2 \begin{pmatrix} -c_4 \\ 1 \end{pmatrix} = \begin{pmatrix} * \\ * \\ c_4 - c_3 \\ 0 \end{pmatrix} =: x_L.$$

Since $[A, C]$ is strictly lower triangular, the vector $[A, C]x_R$ has 3 leading zeros and its inner product with x_L vanishes. Hence,

$$\begin{aligned}(-c_4, 1)V_2^{\mathsf{T}}(AC - CA - K)V_3 \begin{pmatrix} c_1c_2 \\ -c_1 - c_2 \\ 1 \end{pmatrix} &= -\, x_L^{\mathsf{T}} K x_R \\ &= -\,(c_4 - c_3)(c_3 - c_2)(c_3 - c_1)\kappa_{33}\end{aligned}$$

and (56) implies $\kappa_{33} = 0$ for non-confluent nodes. □

Since K is diagonal, the matrix $V_{q-1}^{\mathsf{T}} K V_{s-1}$ has Hankel form again and (56) is an over-determined system for the column sums $\mathbb{1}^{\mathsf{T}} K = (\kappa_{11}, \ldots, \kappa_{ss})$. Solutions may only exist if elements of $V_{q-1}^{\mathsf{T}}(AC - CA)V_{s-1}$ within each anti-diagonal have the same value. For $q = 3,\ s = 4$, this leads to the following restrictions on A alone:

$$(\mathbf{c}^2)^{\mathsf{T}} A\mathbf{c}^j - 2\mathbf{c}^{\mathsf{T}} A\mathbf{c}^{j+1} + \mathbb{1}^{\mathsf{T}} A\mathbf{c}^{j+2} = 0,\ \ j = 0, 1. \tag{57}$$

Since a 2×3 matrix possesses 4 antidiagonals, under assumption (57) the system (56) reduces to

$$\begin{aligned} & \mathbb{1}^{\mathsf{T}} K V_4 = \mathbb{1}^{\mathsf{T}}(AC - CA)(\mathbb{1}, \mathbf{c}, \mathbf{c}^2, 0) + \mathbf{c}^{\mathsf{T}}(AC - CA)\mathbf{c}^2 \\ \iff\ & \mathbb{1}^{\mathsf{T}} K = \mathbb{1}^{\mathsf{T}}(AC - CA) + \beta e_4^{\mathsf{T}} V_4^{-1}, \\ & \text{where } \beta = \mathbf{c}^{\mathsf{T}}(AC - CA)\mathbf{c}^2 - \mathbb{1}^{\mathsf{T}}(AC - CA)\mathbf{c}^3 = \mathbb{1}^{\mathsf{T}}[C, [A, C]]\mathbf{c}^2. \end{aligned}$$

Remark 6.2 Since $\kappa_{33} = 0$, the third stage of the standard method uses no additional function evaluation of $f(Y_{n3}, U_{n3})$ and it seems that it does not provide any additional information. In fact, this stage can be eliminated but the resulting method will be a 3-stage 3-step Peer method.

Remark 6.3 The blind third stage has consequences both for the analysis and the implementation of the standard Peer method. In the equations (12–16) of the Peer steps, it is seen that there is no coupling between the controls U_n of different time steps. Now, the contribution to the Lagrange function from the third stage of the standard method in time step n with multiplier P_{n3} is given by

$$\ldots + P_{n3}^{\mathsf{T}} \left(\sum_{j=1}^{3} a_{3j} Y_{nj} - \sum_{j=1}^{4} b_{nj} Y_{n-1,j} \right) + \ldots$$

missing the unknown U_{n3}. Since U_{n3} does not appear anywhere else, it is non-existent and the unknown U_{n3} should be discarded as well as the corresponding stage equation $0 \cdot \nabla_u f(Y_{n3}, U_{n3})^{\mathsf{T}} P_{n3} = 0$ from (70). This measure will also be used in the analysis of Sect. 7.

6.2 Further Requirements for the Existence of Peer Triplets

For $q + r - 2 > s$, the restriction of $\mathcal{Q}_{q,r}$ to Hankel form is not the whole picture yet. If $q + r - 2 = s + 1$, which case occurs for $r = s = 4$ and $q = s - 1 = 3$, the system (41) does not possess full rank having more than s columns. The vector $(\psi^{\mathsf{T}}, 1)^{\mathsf{T}}$ with $\psi := -V_s^{-1}\mathbf{c}^s$ spans the kernel of the extended Vandermonde matrix $V_{s+1} = (1\!\!1, \mathbf{c}, \ldots, \mathbf{c}^s)$ since it contains the coefficients of the node polynomial $\hat{\psi}(t) = (t - c_1) \cdots (t - c_s) = t^s + \sum_{j=1}^{s} \psi_j t^{j-1}$. Rewriting this property in the form of (41),

$$1\!\!1^{\mathsf{T}} K_n \big(1, 2\mathbf{c}, \ldots, (s+1)\mathbf{c}^s\big) \cdot D_{s+1}^{-1} \begin{pmatrix} \psi \\ 1 \end{pmatrix} = 0,$$

it follows that solutions only exist if also

$$\big(\vartheta_{l,j-l+1}\big)_{j=1}^{s+1} \cdot D_{s+1}^{-1} \begin{pmatrix} \psi \\ 1 \end{pmatrix} = 0. \tag{58}$$

Since $r = s$, a convenient choice for the indices l here is

$$(\vartheta_{21}, \ldots, \vartheta_{2s}, \vartheta_{3s}) = \big(1\!\!1^{\mathsf{T}} - \mathbf{c}^{\mathsf{T}} B V_s \mathcal{P}_s^{-1}, 1 - (\mathbf{c}^2)^{\mathsf{T}} B (\mathbf{c} - 1\!\!1)^{s-1}\big) \text{ for } K_N, \tag{59}$$

$$(\vartheta_{21}, \ldots, \vartheta_{2s}, \vartheta_{3s}) = \big((\mathbf{c} + 1\!\!1)^{\mathsf{T}} B V_s, ((\mathbf{c} + 1\!\!1)^2)^{\mathsf{T}} B \mathbf{c}^{s-1}\big) \text{ for } K_0. \tag{60}$$

We summarize all conditions in the following lemma.

Lemma 6.4 *Let $r = 4$ and $q = 3$. Then, necessary conditions for the existence of four-stage Peer triplets satisfying* (21–25) and (39) *are Hankel form of $\mathcal{Q}_{3,4} = \mathcal{P}_3^{\mathsf{T}} B V_4 \mathcal{P}_4^{-1}$ as well as*

$$\mathbf{c}^{\mathsf{T}} B V_4 P_4^{-1} D_4^{-1} \psi + \frac{1}{5} (\mathbf{c}^2)^{\mathsf{T}} B (\mathbf{c} - 1\!\!1)^3 = \int_0^1 \hat{\psi}(t) dt \tag{61}$$

$$(\mathbf{c}+1\!\!1)^{\mathsf{T}} B V_4 D_4^{-1}\psi + \frac{1}{5}((\mathbf{c}+1\!\!1)^2)^{\mathsf{T}} B\mathbf{c}^3 = 0. \tag{62}$$

Proof As a first step, we consider the constant term $1\!\!1_{s+1}^{\mathsf{T}}$ in the Eq. (59). In (58) it gives rise to the contribution

$$1\!\!1_s^{\mathsf{T}} D_s^{-1}\psi + \frac{1}{s+1} = \sum_{j=0} \frac{1}{j}\psi_j + \frac{1}{s+1} = \int_0^1 \hat{\psi}(t)dt.$$

Now, the conditions (61), (62) correspond to the vanishing of the inner products (58) with the two vectors from (59), (60). □

Remark 6.5 In practice it was found that the two conditions (61), (62) seem to be equivalent with the two simpler equations

$$\int_0^1 \hat{\psi}(t)dt = 0, \tag{63}$$

$$\mathbf{c}^{\mathsf{T}} B V_4 P_4^{-1} D_4^{-1}\psi + \frac{1}{5}(\mathbf{c}^2)^{\mathsf{T}} B(\mathbf{c}-1\!\!1)^3 = 0, \tag{64}$$

in conjunction with the many other order conditions.

In general, Peer methods are invariant under a common shift of the nodes, which means in practice that this shift may be fixed after the construction of some method, e.g. by choosing $c_s = 1$. However, orthogonality (63) strongly depends on the absolute positions of the nodes. Still, for $\int_0^1 \hat{\psi}(t)dt$ this dependence is only linear for the node differences and (63) may be easily solved for one of those, e.g. for $d_4 = c_4 - c_2$.

6.3 Method `AP4o43p`

Without the non-negativity condition several different regions in the parameter space of Peer triplets did exist for $(r, q) = (4, 3)$ in [17]. Some of the standard methods found there have non-monotonic nodes and negative diagonal elements in K. Now, non-negativity seems to leave only one such region with ordered nodes c_i and we present only one such method with a nearly maximal stability angle. Method `AP4o43p` has a stability angle of $\alpha = 59.78^o$ with node vector

$$\mathbf{c}^{\mathsf{T}} = \Big(\frac{4657}{46172}, \frac{43}{97}, \frac{3991}{6596}, \frac{21111803999}{23798723875}\Big) \doteq \big(0.1009, 0.4432, 0.6050, 0.8871\big).$$

The node c_4 has a rather long representation since it was used to solve condition (63). The damping factor $\gamma = 0.58$ from (50) is well below one, the error constants are $err_4 = 0.092/4! \approx 0.0038$ and $err_3^{\dagger} = 0.144/3! \approx 0.024$ and the quotient (52) is $csq = 11.0$. Further data are collected in Table 1. In order to obtain acceptable properties for the stability and definiteness (51) of the boundary methods, the matrices A_0, A_N have full block size 4, denoted by `blksz=4` in Table 2. The coefficients of `AP4o43p` are given in the supplementary material A.3 (Online Resource 1).

7 The Global Error

Convergence of the Peer triplets for $h \to 0$ will be discussed for the unconstrained case $N_U = \{0\}$ only. Here, the additional constraint (16), (26) complicates the situation compared to [16, 17] and we will extend these proofs here. Node vectors of the exact solution are denoted by bold face, e.g., $\mathbf{y}_n = \big(y^\star(t_{ni})\big)_{i=1}^s$, $\mathbf{y} := \big(\mathbf{y}_n\big)_{n=0}^N$, and the global errors by checks, e.g., $\check{Y}_n = Y_n - \mathbf{y}_n$, $\check{Y} := \big(\check{Y}_n\big)_{n=0}^N$. For ease of writing, we also introduce the combined vectors $\mathbf{z} := (\mathbf{y}^\top, \mathbf{p}^\top, \mathbf{u}^\top)^\top$ and $\check{Z} = (\check{Y}^\top, \check{P}^\top, \check{U}^\top)^\top$, and use abbreviations like $F(Z_n) := F(Y_n, U_n)$ and $\nabla_z f = (\nabla_y f, 0, \nabla_u f)$.

In the error discussion, a notational difficulty arises since the right-hand sides of the adjoint equations already contain a first derivative $(\nabla_y f)^\top p$. In order to avoid ambiguities with second derivatives, we introduce an additional notation $\langle ., . \rangle$ for the standard inner product in $\mathbb{R}^m$ which is exclusively dedicated to the product of the Lagrange multiplier p and the components of the function $f = (f_i)_{i=1}^m$ or its derivatives as in

$$\langle p, f\rangle = \sum_{i=1}^m (e_i^\top p) f_i, \quad \langle p, \nabla_z f\rangle = \sum_{i=1}^m (e_i^\top p)\nabla_z f_i.$$

The notation is particularly used for second derivatives, where the matrix $\nabla_{uu}\langle p, f\rangle = \sum_{i=1}^m (e_i^\top p)\nabla_{uu} f_i$ is symmetric and a linear combination of Hessian matrices of the components of f. The notation carries over to compound expressions of a whole time step, e.g.

$$\Big\langle K_n^\top P_n, \nabla_{UU} F(Y_n, U_n)\Big\rangle = \operatorname{diag}_{j=1,\dots,s}\Bigg(\Bigg\langle \sum_{i=1}^s P_{ni}\kappa_{ij}^{[n]}, \nabla_{uu} f(Y_{nj}, U_{nj})\Bigg\rangle\Bigg).$$

With these notations, we may introduce the local errors for the two-step equations

$$\begin{aligned}
\tau_0^Y &:= \mathbf{y}_0 - \mathbb{1}\otimes y_0 - hA_0^{-1}K_0 F(\mathbf{z}_0),\\
\tau_n^Y &:= \mathbf{y}_n - A_n^{-1}\big(B_n \mathbf{y}_{n-1} + hK_n F(\mathbf{z}_n)\big), \quad 1 \le n \le N, \qquad (65)\\
\tau_n^P &:= \mathbf{p}_n - A_n^{-\top}\big(B_{n+1}^\top \mathbf{p}_{n+1} + h\nabla_Y F(\mathbf{z}_n)^\top K_n^\top \mathbf{p}_n\big),\ 0 \le n < N,\\
\tau_N^P &:= \mathbf{p}_N - \mathbb{1}\otimes p(T) - hA_N^{-\top}\nabla_Y F(\mathbf{z}_N)^\top K_N^\top \mathbf{p}_N. \qquad (66)
\end{aligned}$$

The order conditions from Table 3 have the following consequences for these errors (see [17] and Lemma 4.1),

$$\tau_n^Y = O(h^r \|(y^\star)^{(r)}\|_\infty), \quad \tau_n^P = O(h^q \|(p^\star)^{(q)}\|_\infty),\ 0 \le n \le N.$$

Table 3 Combined order conditions for the Peer triplets AP4o43p and AP4o33pa

	Steps	forward: $r \le s = 4$	adjoint: $q = s - 1 = 3$
(a)	Start, $n = 0$	(21)	(24), $n = 0$, (29), (32)
(b)	Standard, $1 \le n < N$	(22)	(24)
(c)	Super-convergence	(48)	(49)
(d)	Compatibility	(38), (40)	
	… for $r = s = 4$	(63), (64)	
(e)	Last step	(22), $n = N$	(24), $n = N - 1$
(f)	End point	(23)	(25), (29), (32)

Furthermore, line (c) of Table 3 gives for the interior time steps with $1 \le n \le N-1$ that

$$1\!\!1^{\mathsf{T}} A\tau_n^Y = O(h^{r+1}\|(y^\star)^{(r+1)}\|_\infty), \quad 1\!\!1^{\mathsf{T}} A^{\mathsf{T}}\tau_n^P = O(h^{q+1}\|(p^\star)^{(q+1)}\|_\infty).$$

The global error in these time steps may be analyzed by multiplying (12), (13) by inverses A_n^{-1} and (14), (15) by $A_n^{-\mathsf{T}}$ and subtracting the corresponding Eqs. (65) or (66), respectively. This yields the equations

$$\begin{aligned} \check{Y}_n - \bar{B}_n\check{Y}_{n-1} &= R_n^Y(\check{Z}) - \tau_n^Y,\ n = 0,\dots,N, \\ R_n^Y(\check{Z}) &:= h\bar{K}_n\big(F(\mathbf{z}_n + \check{Z}_n) - F(\mathbf{z}_n)\big), \end{aligned} \tag{67}$$

$$\begin{aligned} \check{P}_n - \tilde{B}_{n+1}^{\mathsf{T}}\check{P}_{n+1} &= R_n^P(\check{Z}_n) - \tau_n^P,\ n = 0,\dots,N-1, \\ R_n^P(\check{Z}) &:= hA_n^{-\mathsf{T}}\big(\langle K_n^{\mathsf{T}}(\mathbf{p}_n + \check{P}_n), \nabla_Y F(\mathbf{z}_n + \check{Z}_n)\rangle \\ &\qquad - \langle \tilde{K}_n^{\mathsf{T}}\mathbf{p}_n, \nabla_Y F(\mathbf{z}_n)\rangle\big), \end{aligned} \tag{68}$$

with $\bar{K}_n = A_n^{-1}K_n$. Recall that $\bar{B}_n = A_n^{-1}B_n$ and $\tilde{B}_n^{\mathsf{T}} = A_n^{-\mathsf{T}}B_n^{\mathsf{T}}$. The starting step fits in (67) by setting $\bar{B}_0 = 0$. Restricting the objective function $\mathcal{C}$ to polynomials of degree two at most the equation for the error in step (14) becomes

$$\begin{aligned} \check{P}_N - \big((1\!\!1 w^{\mathsf{T}}) \otimes \nabla_{yy}\mathcal{C}\big)\check{Y}_N &= R_N^P(\check{Z}) - \tau_N^P, \\ R_N^P(\check{Z}) &= hA_N^{-1}\big(\langle K_N^{\mathsf{T}}(\mathbf{p}_N + \check{P}_N), \nabla_Y F(\mathbf{z}_N + \check{Z}_N)\rangle \\ &\qquad - \langle \tilde{K}_n^{\mathsf{T}}\mathbf{p}_N, \nabla_Y F(\mathbf{z}_N)\rangle\big). \end{aligned} \tag{69}$$

The convergence proof will employ a fixed-point argument and its basic principle is to collect all terms of size $O(1)$ (i.e. independent of h) implicitly as part of a fixed linear system and terms of size h as Lipschitz functions. Accordingly, in the Eqs. (67–69) the implicit two-step terms were written on the left-hand sides and $O(h)$-terms on the right. Unfortunately, this separation is not so obvious for the new constraint (26) which may be written as

$$\Big\langle K_n^\top P_n, \nabla_U F(Y_n, U_n)\Big\rangle^\top = 0 \tag{70}$$

in our new notation. We remind that by Remark 6.3 the unknowns U_{n3} and the corresponding equations for blind stages ($\kappa_{33}^{[n]} = 0$, $1 \le n < N$) should be discarded.

For (70) we resort to Taylor expansion at the exact solution $\mathbf{z}$ and use the Jacobian at this point in the implicit part of the error equation. Considering the definition (31) of the local error τ_n^U and that the exact solution $(y^\star(t), p^\star(t), u^\star(t))$ satisfies $\nabla_u f(y^\star, u^\star)^\top p^\star = \langle p^\star, \nabla_y f(y^\star, u^\star)\rangle^\top \equiv 0$, see (27), we have

$$\begin{aligned}
-(\tau_n^U)^\top &= \langle K_n^\top P_n, \nabla_U F(Z_n)\rangle - \langle K_n^\top \mathbf{p}_n, \nabla_U F(\mathbf{z}_n)\rangle \\
&= \langle K_n^\top \check{P}_n, \nabla_U F(\mathbf{z}_n)\rangle + \langle K_n^\top \mathbf{p}_n, \nabla_U F(Z_n) - \nabla_U F(\mathbf{z}_n)\rangle \\
&\quad + \langle K_n^\top \check{P}_n, \nabla_U F(Z_n) - \nabla_U F(\mathbf{z}_n)\rangle \\
&= \langle K_n^\top \check{P}_n, \nabla_U F(\mathbf{z}_n)\rangle \\
&\quad + \langle K_n^\top \mathbf{p}_n, \nabla_{UY} F(\mathbf{z}_n)\check{Y}_n\rangle + \langle K_n^\top \mathbf{p}_n, \nabla_{UU} F(\mathbf{z}_n)\check{U}_n\rangle \\
&\quad + \langle K_n^\top \mathbf{p}_n, \nabla_U F(Z_n) - \nabla_U F(\mathbf{z}_n)\rangle - \langle K_n^\top \mathbf{p}_n, \nabla_{UY} F(\mathbf{z}_n)\check{Y}_n\rangle \\
&\quad - \langle K_n^\top \mathbf{p}_n, \nabla_{UU} F(\mathbf{z}_n)\check{U}_n\rangle + \langle K_n^\top \check{P}_n, \nabla_U F(Z_n) - \nabla_U F(\mathbf{z}_n)\rangle.
\end{aligned}$$

Hence, in each time step there is an additional equation

$$\begin{aligned}
&\Big(\langle K_n^\top \mathbf{p}_n, \nabla_{UY} F(\mathbf{z}_n)\rangle^\top, \nabla_U F(\mathbf{z}_n)^\top K_n, \langle K_n^\top \mathbf{p}_n, \nabla_{UU} F(\mathbf{z}_n)\rangle^\top\Big)\check{Z}_n \\
&= -\tau_n^U + R_n^U(\check{Z}_n).
\end{aligned} \tag{71}$$

The function on its right-hand side is given by

$$\begin{aligned}
R_n^U(\check{Z}) = &- \langle K_n^\top \check{P}_n, \nabla_U F(\mathbf{z}_n + \check{Z}_n) - \nabla_U F(\mathbf{z}_n)\rangle^\top \\
&- \langle K_n^\top \mathbf{p}_n, \nabla_U F(\mathbf{z}_n + \check{Z}_n) - \nabla_U F(\mathbf{z}_n)\rangle^\top. \\
&+ \langle K_n^\top \mathbf{p}_n, \nabla_{UY} F(\mathbf{z}_n)\check{Y}_n + \nabla_{UU} F(\mathbf{z}_n)\check{U}_n\rangle^\top.
\end{aligned} \tag{72}$$

The important point in this equation is that the matrix of the left-hand side of (71) is independent of $\check{Z}$ and that the right-hand side (72) has a Lipschitz constant $O(\epsilon)$ in an ϵ-neighborhood of the origin as will be shown below.

Combining now all equations (67), (68), (69) and (71), the complete system for the error $\check{Z}$ has the form

$$\begin{aligned}
\mathbb{M}_0 \check{Z} &= \begin{pmatrix} -\tau^Y + R^Y(\check{Z}) \\ -\tau^P + R^P(\check{Z}) \\ -\tau^U + R^U(\check{Z}) \end{pmatrix}, \\
\mathbb{M}_0 &:= \begin{pmatrix} M_{11} \otimes I_m & 0 & 0 \\ M_{21} \otimes \nabla_{yy}\mathcal{C}_N & M_{22} \otimes I_m & 0 \\ \langle \mathbf{K}^\top \mathbf{p}, \nabla_{UY} F(\mathbf{z})\rangle^\top & \nabla_U F(\mathbf{z})^\top \mathbf{K} & \Omega \end{pmatrix},
\end{aligned} \tag{73}$$

where the last block in $\mathbb{M}_0$, being a symmetric block diagonal matrix, was abbreviated as

$$\Omega := \langle \mathbf{K}^\top \mathbf{p}, \nabla_{UU} F(\mathbf{z})\rangle = \operatorname{diag}_n(\Omega_n),\ \ \Omega_n := \langle K_n^\top \mathbf{p}_n, \nabla_{UU} F(\mathbf{z}_n)\rangle, \tag{74}$$

since it requires detailed investigation. In order to avoid confusion with the coefficient K from the standard scheme, we denoted the block diagonal matrix of all K_n by $\mathbf{K} := \operatorname{diag}(K_0, K, \dots, K, K_N)$. We note that $R^Y(0) = R^P(0) = 0$, $R^U(0) = 0$.

For ease of reading, we recall a few details for $\mathbb{M}_0$ from [16]. The index range $0 \dots N$ of the grid is also used for this matrix and its different blocks. From (67) and (68), we see that the submatrices M_{11} and M_{22} are block bi-diagonal matrices with identities I_s in the main diagonal. M_{11} is lower bi-diagonal with subdiagonal blocks $(M_{11})_{n,n-1} = -\bar{B}_n := -A_n^{-1} B_n$, $1 \le n \le N$, and M_{22} is upper bi-diagonal with super-diagonals $(M_{22})_{n,n+1} = -\tilde{B}_{n+1}^\top := -(BA^{-1})^\top$, $0 \le n < N$. Due to requirement (50) there exist norms such that $\|\bar{B}_n\| = \|\tilde{B}_{n+1}^\top\| = 1$ hold. Hence, all non-trivial blocks of the inverses M_{11}^{-1} and M_{22}^{-1} have norm one (with the possible exception of one single block at each boundary). The third block M_{21} is very sparse with a rank-one entry $\mathbb{1} w^\top$ in its last diagonal $s \times s$-block only due to (14).

Temporarily assuming non-singularity of Ω (which will be considered later on) the system (73) may be transformed to fixed-point form

$$\check{Z} = \Phi(\check{Z}) = \begin{pmatrix} \Phi^Y \\ \Phi^P \\ \Phi^U \end{pmatrix} := -\mathbb{M}_0^{-1}\tau + \mathbb{M}_0^{-1} R(\check{Z}) \tag{75}$$

with the vectors

$$\tau := \begin{pmatrix} \tau^Y \\ \tau^P \\ \tau^U \end{pmatrix}, \quad R(\check{Z}) := \begin{pmatrix} R^Y(\check{Z}) \\ R^P(\check{Z}) \\ R^U(\check{Z}) \end{pmatrix}.$$

Now, the inverse of the fixed matrix $\mathbb{M}_0$ may be given in factored form as

$$\mathbb{M}_0^{-1} = \begin{pmatrix} I & & \\ 0 & I & \\ -\Omega^{-1}\langle \mathbf{K}^\top \mathbf{p}, \nabla_{UY} F(\mathbf{z})\rangle^\top & -\Omega^{-1}\nabla_U F(\mathbf{z})^\top \mathbf{K}\ \Omega^{-1} & \end{pmatrix} \tag{76}$$

$$\cdot \begin{pmatrix} M_{11}^{-1} \otimes I_m & & \\ -M_{22}^{-1} M_{21} M_{11}^{-1} \otimes \nabla_{yy}\mathcal{C}_N & M_{22}^{-1} \otimes I_m & \\ 0 & 0 & I \end{pmatrix}. \tag{77}$$

Since there is no coupling between the stages of different time steps in $R(\check{Z})$, all Jacobians $\nabla_Y R$, $\nabla_P R$, $\nabla_U R$ for each part R^Y, R^P, R^U are block diagonal matrices with blocks of size $(sm) \times (sm)$ or $(sd) \times (sd)$, since the two-step structure of the Peer methods is completely represented in the block bi-diagonal matrices M_{11}, M_{22}. Since every single nontrivial $s \times s$-block of the block triangular inverses M_{11}^{-1}, M_{22}^{-1}

has norm one according to the discussion following Eq. (74), the inverse (77) exists with a norm of size $O(N) = O(h^{-1})$. Together with smoothness requirements on the function f Lipschitz estimates for the $\check{Y}$, $\check{P}$ equations in (75) will follow below by standard arguments. However, the last part with R^U, lacking the factor h, which is essentially covered by the left factor (76) of $\mathbb{M}_0^{-1}$, needs additional inspection.

Obviously, boundedness of this factor (76) requires that the block diagonal matrix Ω from (74) has a uniformly bounded inverse. Since $p^\star(t)$ is assumed to be smooth, Lemma 4.1 shows that $K_n^\mathsf{T}\mathbf{p}_n = D_{K_n}\mathbf{p}_n + O(h^q)$, where $D_{K_n} \succ 0$ is the diagonal matrix with the row sums of K_n^T. Hence, $D_{K_n}^{-1}\Omega_n$ is a small perturbation of a block diagonal matrix with blocks being the Hesse matrices $\nabla_{uu}H(\mathbf{y}_{ni}, \mathbf{p}_{ni}, \mathbf{u}_{ni}) = \langle \mathbf{p}_{ni}, \nabla_{uu} f(\mathbf{y}_{ni}, \mathbf{u}_{ni})\rangle$, $1 \le i \le s$, of the Hamiltonian $H(y, p, u) = p^\mathsf{T} f(y, u)$ at the solution. Now, the *control-uniqueness property* from [11] assumes that for any $t \in [0, T]$ the Hamiltonian $H\big(y(t), p(t), \tilde{u}\big)$ has a unique minimum with respect to $\tilde{u}$ in small neighborhoods of $u(t)$. An appropriate condition for this property is the definiteness of the Hessian

$$\nabla_{uu}H\big(y(t), p(t), u(t)\big) \succ \eta I_d,\ t \in [0, T],\ \eta > 0, \tag{78}$$

implying bounded invertibility of this Hessian which is not essentially affected by small perturbations of $p(t)$. In the main theorem below, we will use a slightly weaker assumption (82), but we will show now that (78) is satisfied for an interesting class of control problems.

Example 7.1 A common type of optimal control problems of tracking type has right-hand sides $f(y, u)$ which depend linearly on $u \in \mathbb{R}^d$. Only the objective function is quadratic in the form (9),

$$\frac{1}{2}\int_0^T (y^\mathsf{T}\Upsilon y + u^\mathsf{T} W u)dt \tag{79}$$

with positive definite matrices $\Upsilon, W \succ 0$. The transformation to standard form (4) uses the additional differential equation

$$y'_{m+1} = \frac{1}{2}\big(y^\mathsf{T}\Upsilon y + u^\mathsf{T} W u\big),\ y_{m+1}(0) = 0, \tag{80}$$

and (79) becomes $\mathcal{C}(\bar{y}(T)) = y_{m+1}(T)$ with extended $\bar{y}^\mathsf{T} = (y^\mathsf{T}, y_{m+1})$. Also, $\bar{f}$, $\bar{p}$ are extended versions. Since the right-hand side of (80) does not depend on y_{m+1}, the adjoint equation for the last Lagrange multiplier simply reads $p'_{m+1} = 0$ with end condition $p_{m+1}(T) = 1$ yielding $p^\star_{m+1}(t) \equiv 1$. Hence, $\langle p^\star, \nabla_{uu}H(z^\star)\rangle = \langle \bar{p}^\star, \nabla_{uu}\bar{f}(y^\star, u^\star)\rangle = W \succ 0$ is definite.

We note that here even the $m+1$-th component of the discrete solution $\bar{P}_n$ is exact. Denoting the stage vector of its $m+1$-th component by $\phi_n \in \mathbb{R}^s$, one sees that the first column of (25) simply states $A_N^\mathsf{T} V_s e_1 = A_N^\mathsf{T}\mathbb{1} = w$ for $r \ge s-1$ and that the end condition (14) reads $A_N^\mathsf{T}\phi_N = w$. Hence $\phi_N = \mathbb{1}$ is the unique solution due to the non-singularity of A_N. Then, the adjoint recursion (15) reduces to $A_n^\mathsf{T}\phi_n = B_{n+1}^\mathsf{T}\phi_{n+1}$, which leaves the vector $\mathbb{1}$ unchanged by (24). Hence, $\phi_n \equiv \mathbb{1}$ for $n = 0, \dots, N$. Since the original right-hand side f is linear in u, condition (16) depends on u only in the $m+1$-th component of $\bar{F}$ and its derivative with respect to U is $D_{K_n} \otimes W \succ 0$ by (32). We remind that for `AP4o43p` equations from the third stage with $\kappa_{33} = 0$ have been removed from the system, see Remark 6.3. Hence, the corresponding diagonal block in Newton's method is non-singular.

For the error estimates below norms are required on three different levels. On the highest, the grid level, the maximum norm is used for convenience. On the step level it is essential to use appropriate weighted norms for $\check{Y}_n, \check{P}_n$ such that $\|\bar{B}\| = \|\tilde{B}^\mathsf{T}\| = 1$ holds. On the lowest, the problem level, any norm may be appropriate. If, for instance, (78) is given, the Euclidean norm may be considered for $\check{U}_{ni} \in \mathbb{R}^d$. However, in the following theorem, we will use the slightly more general assumption (82) in an appropriate norm. In order to prove contraction in the equations for $\check{Y}, \check{P}$ in (75), some degree of smoothness is required for the right-hand sides of both differential equations (7), (8). We assume that there exist constants $\Lambda_j,\ j = 1, 2, 3$, such that

$$\|\nabla_z f(z)\| \le \Lambda_1, \quad (1 + \|z\|)\|\nabla_{zz} f(z)\| \le \Lambda_2, \quad \|\langle p, \nabla_{uzz} f(z)\rangle\| \le \Lambda_3, \tag{81}$$

in some tubular neighborhood of the solution $z^\star = ((y^\star)^\mathsf{T}, (p^\star)^\mathsf{T}, (u^\star)^\mathsf{T})^\mathsf{T}$. The constant $\Gamma := \|\nabla_{yy}\mathcal{C}\|$ vanishes for linear objective functions. One further detail concerns the triplet `AP4o33pa` having a node larger than one, which means that the last off-step node t_{Ns} in the grid exceeds T. Hence, the smoothness assumptions on the solution are required in a slightly larger interval $[0, T^*] \supseteq [0, T]$.

Theorem 7.2 *Considering the unconstrained case with $N_U = \{0\}$, let the objective function $\mathcal{C}$ in (4) be a polynomial of degree less or equal two and assume*

$$\|\langle \tilde{p}, \nabla_{uu} f\big(y^\star(t), u^\star(t)\big)\rangle^{-1}\| \le \omega \tag{82}$$

for all $\tilde{p}$ with $\|\tilde{p} - p(t)\| \le \epsilon < 1$, $t \in [0, T^*]$, $T^* > T$. Assume also that a unique solution $\big(y^\star(t), p^\star(t), u^\star(t)\big)$ of (7–8) exists with $y^\star, p^\star, u^\star \in C^q[0, T^*]$. Also let the right-hand side f of (5) satisfy (81) such that

$$(\zeta_1^P \Lambda_1 + \zeta_2^P \Lambda_2 + \zeta_c^P \Gamma \Lambda_1) T \le \frac{2}{3}, \tag{83}$$

$$(\zeta_1^U \Lambda_1 + \zeta_2^U \Lambda_2)\Lambda_1 \omega \le \frac{1}{3}, \tag{84}$$

where the constants ζ are determined by the actual triplet only and do not depend on the grid size.

Let the Peer triplet satisfy the order conditions from lines (a,b,d,e,f) of Table 3 with $2 \le q \le r \le s$, and the eigenvalue conditions (50), (51).

Then, for $h \le h_0$ the fixed point problem (75) has a unique solution $\check{Z}$ in a sufficiently small tubular neighborhood of the exact solution $\big(y^\star(t), p^\star(t), u^\star(t)\big)$ of (7–8). The solution $\check{Z}$ satisfies

$$\|\check{Z}\| = \max\{\|Y - \mathbf{y}\|, \|P - \mathbf{p}\|, \|U - \mathbf{u}\|\} = O(h^{q-1}) \tag{85}$$

where (Y, P, U) *is the solution of the discrete boundary value problem* (12–16).

Proof a) Considering $\tilde{Z}, \hat{Z}$ in a neighborhood $\mathcal{N}_{\mathcal{Z}} := \{\mathcal{Z} : \|\mathcal{Z}\| \le \varepsilon\}$ of the origin, a Lipschitz condition for R_n^Y in (67) follows from

$$\begin{aligned}\|R_n^Y(\tilde{Z}) - R_n^Y(\hat{Z})\| &= h\|\bar{K}_n\big(F(\mathbf{z}_n + \tilde{Z}_n) - F(\mathbf{z}_n + \hat{Z}_n)\big)\| \\ &\le h\|\bar{K}_n\|\Lambda_1\|\tilde{Z}_n - \hat{Z}_n\|. \end{aligned} \tag{86}$$

Now, R^Y is multiplied by a lower block triangular matrix possessing the blocks $(M_{11}^{-1})_{nk} = \bar{B}_n \cdots \bar{B}_{k+1}$, $k < n$. With the possible exception of $\bar{B}_N$, all other factors $\bar{B}$ have norm one in a suitable norm which exists according to (50). So a Lipschitz condition in the equation for $\check{Y}$ holds as

$$\|M_{11}^{-1}\big(R^Y(\tilde{Y}) - R^Y(\hat{Y})\big)\| \le L_Y\|\tilde{Z} - \hat{Z}\|,\ L_Y = \zeta^Y\Lambda_1 T, \tag{87}$$

since $\|M_{11}^{-1}\| \le N\|B_N\| + 1 \le \|B_N\|Th^{-1}$. In (87), $\zeta^Y = \|\bar{B}_N\| \max_n \|\bar{K}_n\|$ is a constant depending on the actual triplet only, while $\Lambda_1 T$ is problem-dependent.

Using the formal identity $2(ab - \alpha\beta) = (a - \alpha)(b + \beta) + (a + \alpha)(b - \beta)$, the Lipschitz difference for R^P may be rewritten as

$$\begin{aligned}&\|R_n^P(\tilde{Z}) - R_n^P(\hat{Z})\| \\ &\quad = h\|A_n^{-\mathsf{T}}\big(\langle K_n^{\mathsf{T}}(\mathbf{p}_n + \tilde{P}_n), \nabla_Y F(\mathbf{z}_n + \tilde{Z}_n)\rangle \\ &\quad - \langle K_n^{\mathsf{T}}(\mathbf{p}_n + \hat{P}_n), \nabla_Y F(\mathbf{z}_n + \hat{Z}_n)\rangle\big)\| \\ &\quad = \frac{h}{2}\|A_n^{-\mathsf{T}}\big(\langle K_n^{\mathsf{T}}(\tilde{P}_n - \hat{P}_n), \nabla_Y F(\mathbf{z}_n + \tilde{Z}_n) + \nabla_Y F(\mathbf{z}_n + \hat{Z}_n)\rangle \\ &\quad + \langle K_n^{\mathsf{T}}(2\mathbf{p}_n + \tilde{P}_n + \hat{P}_n), \nabla_Y F(\mathbf{z}_n + \tilde{Z}_n) - \nabla_Y F(\mathbf{z}_n + \hat{Z}_n)\rangle\big)\| \\ &\le h(\tilde{\zeta}_1^P\Lambda_1 + \tilde{\zeta}_2^P\Lambda_2)\|\tilde{Z}_n - \hat{Z}_n\|,\end{aligned}$$

with constants $\tilde{\zeta}_j^P$. Looking again at (75), (77), the matrix M_{22}^{-1} is block upper triangular with off-diagonal blocks essentially being powers of $\tilde{B}^{\mathsf{T}}$ where $\|\tilde{B}^{\mathsf{T}}\| = 1$. Again, only the norms in the boundary steps may be different and we get $\|M_{22}^{-1}\| \le \|\tilde{B}_1^{\mathsf{T}}\|(N + \|\tilde{B}_N^{\mathsf{T}}\|) \le 2\|\tilde{B}_1^{\mathsf{T}}\|Th^{-1}$ for small h. The matrix $M_{22}^{-1}M_{21}M_{11}^{-1}$ in the subdiagonal of (77) is a simple rank-one matrix [16]. Since the two factors of the only nontrivial last block of $(M_{21})_{NN} = \mathbb{1}w^{\mathsf{T}} = \mathbb{1}\mathbb{1}^{\mathsf{T}}A_N$ are right resp. left eigenvectors of the blocks in M_{22}^{-1} resp. M_{11}^{-1}, each $s \times s$-Block of $M_{22}^{-1}M_{21}M_{11}^{-1}$ is equal to $\mathbb{1}w^{\mathsf{T}}$ and the norm of this matrix is $O(N)$ again. Hence, in the Lipschitz difference of Φ^P, we also inherit some contribution $O(\Gamma L_Y\|\tilde{Z} - \hat{Z}\|)$ from (86) yielding

$$\begin{aligned}\|\Phi^P(\tilde{Z}) - \Phi^P(\hat{Z})\| &\le L_P\|\tilde{Z}_n - \hat{Z}_n\| \\ L_P &= (\zeta_1^P\Lambda_1 + \zeta_2^P\Lambda_2 + \zeta_c^P\Gamma\Lambda_1)T.\end{aligned}$$

In order to simplify assumptions slightly, we will assume $\zeta_1^P \geq \zeta^Y$ yielding $L_P \geq L_Y$.
b) The map R_n^U given explicitly in (72) consists of two parts and we will consider Lipschitz differences $R_n^U(\tilde{Z}) - R_n^U(\hat{Z})$ for both parts separately by using Taylor's Theorem with integral remainder. For the contribution in the first line of (72), we get with $(\nabla_Y, 0, \nabla_U) = \nabla_Z$ that

$$\begin{aligned}
&- \langle K_n^\top(\tilde{P}_n - \hat{P}_n), \nabla_U F(\mathbf{z}_n)\rangle^\top \\
&+ \langle K_n^\top \tilde{P}_n, \nabla_U F(\mathbf{z}_n + \tilde{Z}_n)\rangle^\top - \langle K_n^\top \hat{P}_n, \nabla_U F(\mathbf{z}_n + \hat{Z}_n)\rangle^\top \\
&= \langle K_n^\top(\tilde{P}_n - \hat{P}_n), \nabla_U F(\mathbf{z}_n + \tilde{Z}_n) - \nabla_U F(\mathbf{z}_n)\rangle^\top \\
&+ \langle K_n^\top \hat{P}_n, \int_0^1 \nabla_{UZ} F(\mathbf{z}_n + \xi \tilde{Z}_n + (1-\xi)\hat{Z}_n)\, d\xi\, (\tilde{Z}_n - \hat{Z}_n)\rangle^\top \\
&= \langle K_n^\top(\tilde{P}_n - \hat{P}_n), \int_0^1 \nabla_{UZ} F(\mathbf{z}_n + \xi \tilde{Z}_n)\, d\xi\, \tilde{Z}_n\rangle^\top \\
&+ \langle K_n^\top \hat{P}_n, \int_0^1 \nabla_{UZ} F(\mathbf{z}_n + \xi \tilde{Z}_n + (1-\xi)\hat{Z}_n) d\xi (\tilde{Z}_n - \hat{Z}_n)\rangle^\top.
\end{aligned}$$

This part is bounded by

$$\begin{aligned}
&\|K_n^\top\| \Lambda_2 \|\tilde{P}_n - \hat{P}_n\|(\|\tilde{Y}\| + \|\tilde{U}\|) + L_2 \|\hat{P}_n\| \|\tilde{Z}_n - \hat{Z}_n\| \\
&\leq \zeta^U \Lambda_2(\|\tilde{Z}_n\| + \|\hat{Z}_n\|)\|\tilde{Z}_n - \hat{Z}_n\|.
\end{aligned}$$

In the difference for the remaining term from (72), the part $\langle K_n^\top \mathbf{p}_n, \nabla_U F(\mathbf{z}_n)\rangle^\top$ cancels out and the others contribute

$$\begin{aligned}
&- \langle K_n^\top \mathbf{p}_n, \nabla_U F(\mathbf{z}_n + \tilde{Z}_n) - \nabla_U F(\mathbf{z}_n + \hat{Z}_n)\rangle^\top \\
&+ \langle K_n^\top \mathbf{p}_n, \nabla_{UZ} F(\mathbf{z}_n)(\tilde{Z}_n - \hat{Z}_n)\rangle^\top \\
= &- \langle K_n^\top \mathbf{p}_n, \int_0^1 \nabla_{UZ} F(\mathbf{z}_n + \xi \tilde{Z}_n + (1-\xi)\hat{Z}_n)\, d\xi\, (\tilde{Z}_n - \hat{Z}_n)\rangle^\top \\
&+ \langle K_n^\top \mathbf{p}_n, \nabla_{UZ} F(\mathbf{z}_n)(\tilde{Z}_n - \hat{Z}_n)\rangle^\top \\
= &- \langle K_n^\top \mathbf{p}_n, \int_0^1 \Big(\nabla_{UZ} F(\mathbf{z}_n + \xi \tilde{Z}_n + (1-\xi)\hat{Z}_n))^\top - \nabla_{UZ} F(\mathbf{z}_n)\Big)\, d\xi\, (\tilde{Z}_n - \hat{Z}_n)\rangle^\top.
\end{aligned}$$

Obviously, Lipschitz continuity of the second derivatives would suffice to obtain the desired result. Using the bound from assumption (81), it is seen that this contribution to the Lipschitz difference of R_n^U is bounded by $\tilde{\zeta}_3^U \Lambda_3(\|\tilde{Z}_n\| + \|\hat{Z}_n\|)\|\tilde{Z}_n - \hat{Z}_n\|$. Hence, we have shown that

$$\|R^U(\tilde{Z}) - R^U(\hat{Z})\| \le (\tilde{\zeta}_2^U \Lambda_2 + \tilde{\zeta}_3^U \Lambda_3)(\|\tilde{Z}\| + \|\hat{Z}\|)\|\tilde{Z} - \hat{Z}\|. \tag{88}$$

c) We conclude the proof that (75) is a contractive fixed point problem by showing that the first factor (76) of $\mathbb{M}_0^{-1}$ is bounded uniformly in $h \le h_0$. The matrix Ω in its last block is again a block diagonal matrix with blocks $\Omega_n = \langle K_n^\mathsf{T} \mathbf{p}_n, \nabla_{UU} F(\mathbf{z}_n)\rangle$. Since $\mathbf{p}_n$ contains the node values of the smooth solution $p^\star(t)$, the estimate (30) in Lemma 4.1 shows that

$$\|K_n^\mathsf{T} \mathbf{p}_n - D_{K_n} \mathbf{p}_n\| = O(h^q) < \epsilon$$

for $h \le h_0$ where $D_{K_n} := \operatorname{diag}_i(1\!\!1^\mathsf{T} K_n e_i)$ differs from K_n^T in the boundary steps only, i.e., for $n = 0, N$. Hence, assumption (82) shows that $\max_n \|\Omega_n^{-1}\| \le \omega$ and that the first factor (76) is uniformly bounded. Its subdiagonal blocks add contributions with constants $O(\Lambda_2 L_Y)$, $O(\Lambda_1 L_P)$ to the Lipschitz difference of Φ^U. Ordering terms appropriately, we obtain

$$\|\Phi^U(\tilde{Z}) - \Phi^U(\hat{Z})\| \le (L_0^U + L_1^U \varepsilon)\|\tilde{Z} - \hat{Z}\|,$$
$$L_0^U = (\zeta_1^U \Lambda_1 + \zeta_2^U \Lambda_2)\Lambda_1 \omega,\; L_1^U = \zeta_3^U(\Lambda_2 + \Lambda_3)\omega.$$

Here, $L^U = L_0^U + L_1^U \varepsilon$ depends on the radius ε of the neighborhood $\mathcal{N}_\mathcal{Z}$.

The Lipschitz constant for the whole map Φ is $L = \max\{L^Y, L^P, L_0^U + L_1^U \varepsilon\}$ and we are verifying now that the assumptions are sufficient to show

$$\|\Phi(\tilde{Z}) - \Phi(\hat{Z})\| \le \frac{2}{3}\|\tilde{Z} - \hat{Z}\| \tag{89}$$

for ε sufficiently small. First, (83) corresponds to $L^P \le 2/3$ and also implies $L^Y \le 2/3$ with an appropriate constant $\zeta_1^P \ge \zeta^Y$. Then, by (84) we have $L_0^U \le 1/3$ and $L^U = L_0^U + L_1^U \varepsilon \le 2/3$ for $\varepsilon \le 1/(3L_1^U)$.

Finally, we choose $\tilde{Z} \in \mathcal{N}_\mathcal{Z}$ and $\hat{Z} = 0$ where $\|\Phi(0)\| = \|\mathbb{M}_0^{-1}\tau\| = O(h^{q-1})$. Since $q \ge 2$ by assumption, we may restrict $h_0 \le h_1$ such that $\|\Phi(0)\| \le \varepsilon/3$ for $h \le h_0$, and from (89) follows

$$\|\Phi(\tilde{Z})\| \le \|\Phi(0)\| + \|\Phi(\tilde{Z}) - \Phi(0)\| \le \frac{1}{3}\varepsilon + \frac{2}{3}\varepsilon = \varepsilon.$$

Hence, Φ maps $\mathcal{N}_\mathcal{Z}$ onto itself and is a contraction proving the existence of a unique fixed point $\check{Z} = \Phi(\check{Z}) \in \mathcal{N}_\mathcal{Z}$ and the solution $Z = \mathbf{z} + \check{Z}$ of the discrete boundary value problem. Again from (89) follows that

$$\begin{aligned}\|\check{Z}\| =& 3\|\Phi(\check{Z})\| - 2\|\check{Z}\| \le 3\|\Phi(\check{Z}) - \Phi(0)\| + 3\|\mathbb{M}_0^{-1}\tau\| - 2\|\check{Z}\| \\ \le& (3\frac{2}{3} - 2)\|\check{Z}\| + 3\|\mathbb{M}_0^{-1}\tau\| = 3\|\mathbb{M}_0^{-1}\tau\|.\end{aligned} \tag{90}$$

The assertion now follows from $\|\mathbb{M}_0^{-1}\tau\| = O(h^{q-1})$, see Lemma 4.1 in [16].

The global error estimate (85) is rather pessimistic and may be improved by considering the super-convergence conditions (48), (49). However, since the norm in the estimate (90) computes the maximum of all errors $\check{Z}$, $\check{P}$, $\check{U}$, only the lower order h^q may be verified rigorously.

Lemma 7.3 *Let all assumptions of Theorem 7.2 hold with the full set of order conditions from Table 3 for $2 \le q \le r \le s$, $q < s$, and let $y^\star, p^\star \in C^{q+1}[0, T^*]$. Then, the solution (Y, U, P) of (12–16) with $N_{U^s} = \{0\}$ also satisfies*

$$\|\check{Z}\| = \max\{\|Y - \mathbf{y}\|, \|P - \mathbf{p}\|, \|U - \mathbf{u}\|\} = O(h^q).$$

Proof The improved error estimates for Y and P follow from the super-convergence effect where the leading error term of τ is canceled in the product $\mathbb{M}_0^{-1}\tau$ by the conditions (48), (49) and a sufficiently fast damping of the remaining modes through assumption (50) showing that $\|\mathbb{M}_0^{-1}\tau\| = O(h^q)$, see (49) in [17]. A proof by a different technique can be found in [29]. Hence, (90) now yields $\|\check{Y}\|, \|\check{P}\| = O(h^q)$. The errors $\check{U}_n$ of the control variable in different time intervals are independent and (71) may be solved for $\check{U}_n$. There, $\|\tau_n^U\| = O(h^q)$ by (31). Taking norms we get

$$\|\check{U}_n\| \le L\omega\big(\|\tau_n^U\| + \|\check{Y}_n\| + \|\check{P}_n\| + \|R_n^U(\check{Z})\|\big) \tag{91}$$

with some constant L. Since $R^U(0) = 0$ it follows from (88) and (85) that $\|R_n^U(\check{Z})\| \le \tilde{L}\|\check{Z}\|^2 = O(h^{2q-2})$. Finally, (91) yields $\|\check{U}_n\| = O(h^q + h^{2q-2}) = O(h^q)$ for $q \ge 2$. □

Remark 7.4 Since the lowest order in $\|M_0^{-1}\tau\|$ dominates the error estimate (90), only order h^q could be proven rigorously. However, for methods satisfying the conditions (21), (22) and (48) with $r > q$, the standard super-convergence argument leads to $\|M_{11}^{-1}\tau^Y\| = O(h^r)$. Hence, such methods also have global order h^r for pure initial value problems for the state $y(t)$ only. And in many problems coupling between the unknowns y, p, u seems to be so weak that the improved order $r = q + 1$ can also be observed in our numerical experiments following now.

8 Numerical Results

We present numerical results for the Peer triplets `AP4o43pa`, `AP4o33pfs` and `AP4o43p` and compare them with those obtained for our recently developed triplet `AP4o43bdf` from [17, 18]. We note that Matlab codes for [18] are available at [19]. As a well-known standard method also the symmetric fourth-order two-stage Gauss method [13, Table II.1.1] is used. The latter one is implemented along the principles in [11] using intermediate time points $t_n + c_i h$ for the control variables. The standard method `AP4o43bdf` is based on BDF4 and its well-known stability angle is $\alpha = 73.35^o$. It satisfies the positivity requirements (32) and the additional consistency conditions (29) for $q = 2$ which is one order less than for the new Peer triplets. Implicit

Runge–Kutta methods of Gauss type are symplectic making them suitable for optimal control [13, 26]. However, as all one-step methods they may suffer from order reduction due to their lower stage order $s+1$ compared to the classical order $p=2s$.

To illustrate the order of convergence, we first consider two unconstrained problems with known analytic solutions. The first one is a quadratic problem with a mixed term taken from [10, 11] and the second one comes from a method-of-lines discretization of a boundary control problem for the 1D heat equation [18]. Finally, we apply our novel Peer methods to an optimal control problem for an 1D semilinear reaction-diffusion model of Schlögl type with cubic nonlinearity, which was intensively studied in [6]. We pick the problem of stopping a nucleation process to show the potential of higher-order methods.

All calculations have been done with Matlab-Version R2019a on a Latitude 7280 with an i5-7300U Intel processor at 2.7 GHz. We use `fmincon` with stop tolerance 10^{-14}. If not otherwise stated, we apply the `interior-point` algorithm as default choice in `fmincon` and provide the zero control vector as initial guess.

8.1 A Quadratic Problem with a Mixed Term

The first problem is taken from [11]. It was originally proposed in [10, (P2)] and includes a mixed term $y_1(t)u(t)$. We consider

$$\begin{aligned}\text{minimize}\;\; & \frac{1}{2}\int_0^1 \left(1.25\, y_1(t)^2 + y_1(t)u(t) + u(t)^2\right) dt \\ \text{subject to}\;\; & y_1'(t) = 0.5\, y_1(t) + u(t), \quad t \in (0,1], \\ & y_1(0) = 1,\end{aligned}$$

with the optimal solution

$$y_1^\star(t) = \frac{\cosh(1-t)}{\cosh(1)}, \quad u^\star(t) = -\frac{(\tanh(1-t)+0.5)\cosh(1-t)}{\cosh(1)}.$$

The optimal costate can be computed from $p_1^\star(t) = -0.5(y_1^\star(t) + 2u^\star(t))$. Introducing a second component $y_2(t)$ and setting $y_2'(t) = 1.25\, y_1(t)^2 + y_1(t)u(t) + u(t)^2$ with the initial value $y_2(0) = 0$, the objective function can be transformed to the Mayer form $\mathcal{C}(y(1)) = 0.5\, y_2(1)$ with the new state vector $y = (y_1, y_2)^\mathsf{T}$.

Numerical results for $N+1=5, 10, 20, 40$ are shown in Fig. 1. All Peer methods show their theoretical order three for the first component of the adjoint variables, $P_{ni,1}$. The fourth-order Gauss-2 method drops down to order three due to its lower stage order three. Order three is also observed for the state variables $Y_{ni,1}$, except for `AP4o43p` which achieves its super-convergence order four for the first three runs. For the Peer methods, the errors of the control vector U as well as the improved control $U^\ddagger$ obtained from the post-processing in (20) decrease with order three as expected. Since `AP4o43bdf` satisfies the consistency conditions (29) for u with $q=2$ only, its order in U is two, which is nicely seen. However, the third-order approximations in the first

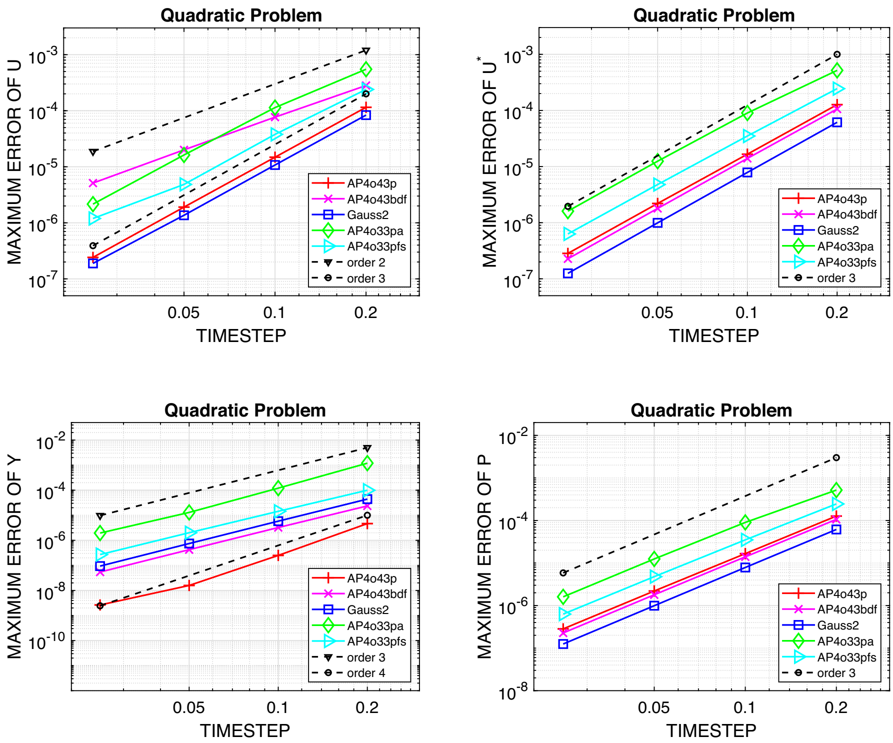

Fig. 1 Quadratic Problem. Convergence of the maximal control errors $\|U_{ni} - u(t_{ni})\|_\infty$ (top left), improved control errors $\|U^{\ddagger}_{ni} - u(t_{ni})\|_\infty$ (top right), state errors $\|Y_{ni,1} - y_1(t_{ni})\|_\infty$ (bottom left), and adjoint errors $\|P_{ni,1} - p_1(t_{ni})\|_\infty$, (bottom right), $n = 0, \ldots, N,\ i = 1, \ldots, s$

components of (Y, P) yields order three for $U^{\ddagger}$ again. Both methods, `AP4o33pa` and `AP4o33pfs`, fall behind the other ones in terms of accuracy. This is not surprising since their better stability properties and the dense output feature of the latter one come with larger error constants.

8.2 Boundary Control of an 1D Discrete Heat Equation

The second problem is taken from [18]. It was especially designed to provide exact formulas for analytic solutions of an optimal boundary control problem governed by a one-dimensional discrete heat equation and an objective function that measures the distance of the final state from the target and the control costs. Since no spatial discretization errors are present, numerical orders of time integrators can be observed with high accuracy without computing reference solutions.

The optimal control problem reads as follows:

$$\text{minimize } \frac{1}{2}\|y(1) - \hat{y}\|_2^2 + \frac{1}{2}\int_0^1 u(t)^2\, dt$$

$$\text{subject to } \; y'(t) = Ay(t) + \gamma e_m u(t), \quad t \in (0, 1],$$
$$y(0) = 1\!\!1,$$

with

$$A = \frac{1}{(\triangle x)^2} \begin{pmatrix} -1 & 1 & & & & \\ 1 & -2 & 1 & & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & -2 & 1 \\ & & & & 1 & -3 \end{pmatrix},$$

state vector $y(t) \in \mathbb{R}^m$, $\triangle x = 1/m$, and $\gamma = 2/(\triangle x)^2$. The components $y_i(t)$ approximate the solution of the continuous 1D heat equation $Y(x, t)$ over the spatial domain $[0, 1]$ at the discrete points $x_i = (i - 0.5)\triangle x, i = 1, \dots, m$. The corresponding boundary conditions are $\partial_x Y(0, t) = 0$ and $Y(1, t) = u(t)$. The matrix $A \in \mathbb{R}^{m \times m}$ results from standard central finite differences. Its eigenvalues λ_k and corresponding normalized orthogonal eigenvectors $v^{[k]}$ are given by

$$\lambda_k = -4m^2 \sin^2\left(\frac{\omega_k}{2m}\right), \; \omega_k = \left(k - \frac{1}{2}\right)\pi,$$
$$v_i^{[k]} = \nu_k \cos\left(\omega_k \frac{2i - 1}{2m}\right), \; \nu_k = \frac{2}{\sqrt{2m + \sin(2\omega_k)/\sin(\omega_k/m)}}, \quad i, k = 1, \dots, m.$$

We follow the test case in [18] and prescribe the sparse control

$$u^\star(t) = -\gamma p_m(t), \quad p^\star(t) = \delta\left(e^{\lambda_1(T-t)}v^{[1]} + e^{\lambda_2(T-t)}v^{[2]}\right),$$

with $\delta = 1/75$, which defines the target vector $\hat{y}$ through

$$\hat{y}(t) = y^\star(T) - \delta\left(v^{[1]} + v^{[2]}\right).$$

The coefficients $\eta_k(T)$ of $y^\star(T) = \sum_{i=1,\dots,m} \eta_k(T)v^{[k]}$ are given by

$$\eta_k(T) = e^{\lambda_k T}\eta_k(0) - \gamma^2 \delta T v_m^{[k]} \sum_{l=1}^{2} v_m^{[l]} \varphi_1((\lambda_k + \lambda_l)T)$$

where $\eta_k(0) = y^\star(0)^\mathsf{T} v^{[k]}$ and $\varphi_1(z) := (e^z - 1)/z$. We will compare the numerical errors for $y(T)$, $p(0)$ and $u(t)$ An approximation $p_h(0)$ for the Peer method is obtained from $p_h(0) = (v^\mathsf{T} \otimes I)P_0$ with $v = V_s^{-\mathsf{T}} e_s$, $e_s = (0, \dots, 0, 1)^\mathsf{T} \in \mathbb{R}^s$. Note that, compared to [18], we have changed the sign of the adjoint variables, i.e., $p \mapsto -p$, to fit into our setting. Introducing an additional component $y_{m+1}(t)$ and adding the

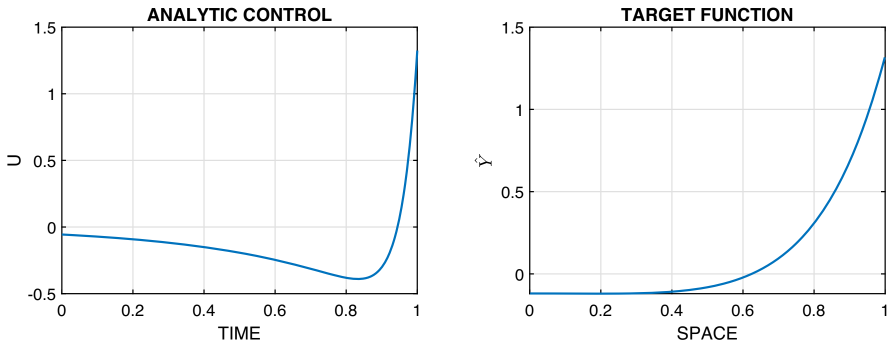

Fig. 2 Dirichlet heat problem with $m = 500$ spatial points. Analytic control $u(t)$ (left) and target function $\hat{y}$ (right)

equations $y'_{m+1}(t) = u(t)^2$, $y_{m+1}(0) = 0$, the objective function can be transformed to the Mayer form

$$\mathcal{C}(y(1)) := \frac{1}{2}\left(\sum_{i=1}^{m}(y_i(1) - \hat{y}_i)^2 + y_{m+1}(1)\right)$$

with the extended vector $\bar{y}(1) = (y_1(1), \ldots, y_m(1), y_{m+1}(1))^{\mathsf{T}}$. We set $m = 500$. In Fig. 2, the analytic control $u(t)$ and the target function $\hat{y} \in \mathbb{R}^m$ are shown.

We will now discuss the numerical errors obtained from applying $N + 1 = 2^k$, $k = 4, \ldots, 9$, time steps. The results are visualized in Fig. 3. As already observed in [18], the one-step Gauss method of order four suffers from a serious order reduction to first order in all variables (y, p, u). This phenomenon is well understood and occurs particularly drastically for time-dependent Dirichlet boundary conditions [24]. This drawback is shared by all one-step methods due to their insufficient stage order. The BDF-based `AP4o43bdf` shows second order convergence in the control, which is in accordance with the fact that it satisfies (29) with $q = 2$ only. This also limits the accuracy of the state and the adjoint at their endpoints to order three and two, respectively. Thus, the improvement in the post-processed control variables $U^{\ddagger}_{ni}$ from (20) is only marginal and does not increase the order. All new Peer methods satisfy (29) for $q = 3$ and deliver approximations of the control with order three, except for certain irregularities in the smallest step. The order of convergence for $y_h(T)$ is three for `AP4o33pa` and `AP4o33pfs`, whereas `AP4o43p` reaches fourth-order superconvergence for nearly all time steps. For $p_h(0)$, `AP4o33pfs` shows an ideal order three. The other two Peer methods vary between order three and five and stagnate at the end when errors are already quite small. The supposed improvement in $U^{\ddagger}_{ni}$ is not achieved for the Peer methods. Quite to the contrary, `AP4o33pfs` loses two orders of magnitude in accuracy, `AP4o43p` loses one order. We infer that for Peer methods, which perform close to their theoretical order, the approximation quality of U is nearly optimal and post-processing is not advisable in general. To summarize, the newly constructed Peer methods significantly improve the approximation of the control

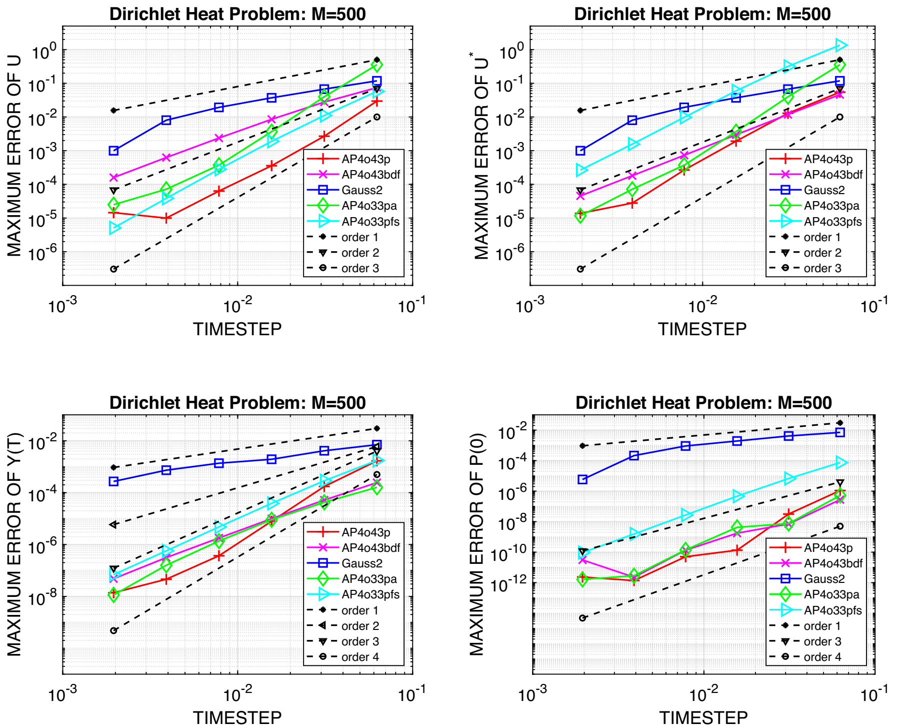

Fig. 3 Dirichlet heat problem with $m = 500$ spatial points. Convergence of the maximal control errors $\|U_{ni} - u(t_{ni})\|_\infty$ (top left), improved control errors $\|U^{\ddagger}_{ni} - u(t_{ni})\|_\infty$ (top right), state errors $\|y(T) - y_h(T)\|_\infty$ (bottom left), and adjoint errors $\|p(0) - p_h(0)\|_\infty$ (bottom right)

with increased order three. AP4o43p gains from its super-convergence property and performs remarkably well for this discrete heat problem.

8.3 Stopping of a Nucleation Process with Distributed Control

In our third study, we consider a PDE-constrained optimal control problem from [6, Chapter 5.4] – stopping of a nucleation process modeled by a nonlinear reaction-diffusion equation of Schlögl-type. It reads

$$\text{minimize } J := \frac{1}{2}\int_Q (Y(x,t) - Y_Q(x,t))^2\, dxdt + \frac{\alpha}{2}\int_Q U(x,t)^2\, dxdt$$

$$\begin{aligned} \text{subject to } \partial_t Y - \partial_{xx} Y &= Y - kY^3 + U(x,t), \quad (x,t) \in Q := (0,L) \times (0,T], \\ Y(x,0) &= Y_0(x), \quad x \in (0,L), \\ \partial_x Y(0,t) &= \partial_x Y(L,t) = 0, \end{aligned}$$

with parameters $\alpha = 10^{-6}$, $L = 20$, $T = 5$, and $k = 1/3$. The initial condition is

$$Y_0(x) = \begin{cases} 1.2\sqrt{3}, & x \in [9, 11], \\ 0, & \text{else}. \end{cases}$$

The solution $Y_{nat}(x, t)$ for $U(x, t) \equiv 0$ describes a natural nucleation process with wavelike dispersion to the left and right. The distributed control $U(x, t)$ should now be chosen in such a way that the dispersion is stopped at $t = 2.5$, forcing a stationary profile $Y_{nat}(x, 2.5)$ for the remaining time interval. Therefore, we set

$$Y_Q(x, t) = \begin{cases} Y_{nat}(x, t), & t \in [0, 2.5], \\ Y_{nat}(x, 2.5), & t \in (2.5, T], \end{cases}$$

in the objective function J. There indeed exists an analytic solution for such a control,

$$U_{stop}(x, t) = \begin{cases} 0, & t \le 2.5, \\ kY_{nat}^3(x, 2.5) - Y_{nat}(x, 2.5) - \partial_{xx} Y_{nat}(x, 2.5), & t > 2.5, \end{cases}$$

since $\partial_t Y(x, t)$ must vanish for $t \ge 2.5$. The authors of [6] proved that the second derivative $\partial_{xx} Y_{nat}(x, 2.5)$ is well defined. The functions Y_{nat}, Y_Q, and U_{stop} are plotted in Fig. 4.

We again use standard finite differences on the shifted spatial mesh $x_i = (i - 0.5)\triangle x$, $i = 1, \ldots, m$, with $\triangle x = L/m$ to discretize the PDE in space. The objective function is approximated by a linear spline. This yields, after transforming to the Mayer form, the discrete control problem

$$\begin{aligned} &\text{minimize } \mathcal{C}\big(y(T)\big) = y_{m+1}(T) \\ &\quad \text{subject to } y'(t) = Ay(t) + G(y(t), u(t)) =: f(y(t), u(t)), \quad t \in (0, T], \\ &\qquad\qquad y(0) = \begin{pmatrix} (Y_0(x_i))_{i=1}^m \\ 0 \end{pmatrix}, \end{aligned}$$

with

$$(G(y, u))_i = \begin{cases} -ky_i^3 + y_i + u_i, & i = 1, \ldots, m, \\ \frac{1}{2}((y - y_Q)_{i=1}^m)^\mathsf{T} M (y - y_Q)_{i=1}^m + \frac{\alpha}{2} u^T M u, & i = m + 1, \end{cases}$$

and the matrices

$$A = \frac{1}{(\triangle x)^2} \begin{pmatrix} -1 & 1 & & & & \\ 1 & -2 & 1 & & & \\ & & \ddots & & & \\ & & 1 & -2 & 1 & \\ & & & 1 & -1 & \\ & & & & & 0 \end{pmatrix}, \quad M = \frac{\triangle x}{12} \begin{pmatrix} 10 & 2 & & & \\ 2 & 8 & 2 & & \\ & & \ddots & & \\ & & 2 & 8 & 2 \\ & & & 2 & 10 \end{pmatrix}.$$

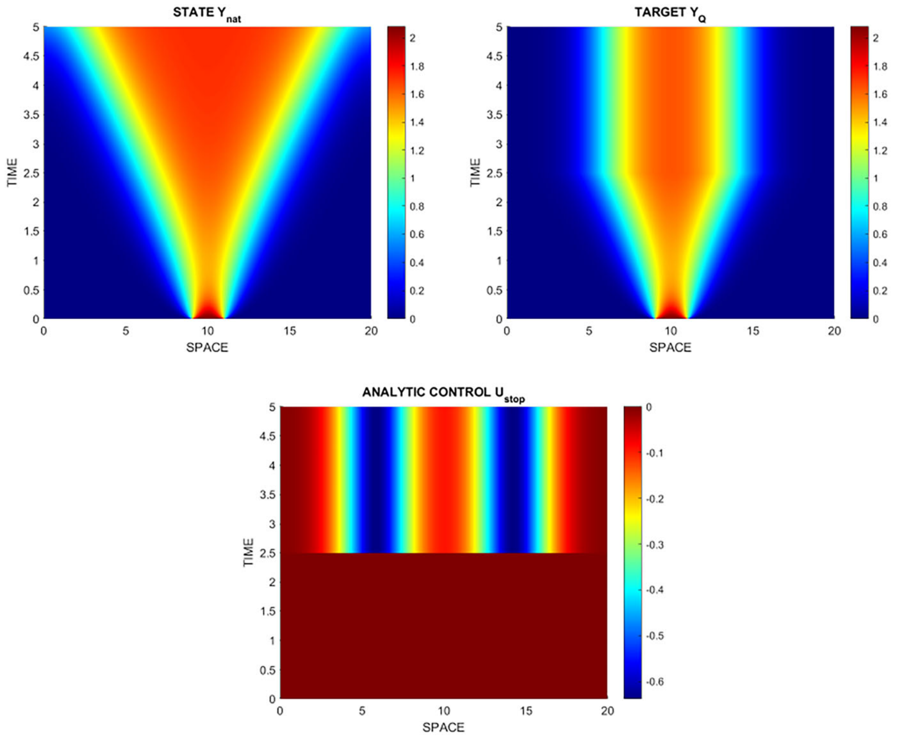

Fig. 4 Nucleation process. Y_{nat} for $U = 0$ (top, left), target function Y_Q (top, right) and analytic control U_{stop} (bottom)

Here, $A \in \mathbb{R}^{m+1,m+1}$, $M \in \mathbb{R}^{m,m}$, $u(t) = (U(x_i, t))_{i=1}^m$, $y_Q(t) = ((Y_Q(x_i, t)_{i=1}^m)$ and $y(t) \approx (Y(x_i, t))_{i=1}^{m+1}$. The total dimension of the discrete control vectors (U_{nj}), $n = 0, \ldots, N$, $j = 1, \ldots, s$, is $ms(N + 1)$. We set $m = 300$ as in [6] and note that $s = 4$ for our Peer methods. The optimal stopping control U_{stop} is discretized by

$$u_{stop}(t) = \begin{cases} 0, & t \le 2.5, \\ ky_Q^3(2.5) - y_Q(2.5) - \hat{A}y_Q(2.5), & t > 2.5, \end{cases}$$

where $\hat{A} = (A)_{i,j=1}^m$. With grid sizes $N \in [24, 399]$ in time the excessive demand of memory for the full Hessian of the objective function prohibits its use in Matlab's `fmincon` subroutine. However, a closer inspection reveals that

$$\nabla_{U_{nj}U_{nj}} \mathcal{C} = h\alpha \sum_{i=1}^{s} \kappa_{ij}^{[n]} (P_{ni})_{m+1} M \in \mathbb{R}^{m,m}, \quad n = 0, \ldots, N, \;\; j = 1, \ldots, s,$$

are the only entries yielding a sparse tridiagonal Hessian, see Example 7.1. Furthermore, controls U_{nj} with $\kappa_{ij}^{[n]} = 0$ for $i = 1, \ldots, s$, are discarded as noted in Sect. 7. Hence, we pass the sparse Hessian to `fmincon` and switch to the

Table 4 Values of the objective function $\mathcal{C}$ and computing times for $U^{(0)} = u_{stop}$ and $N+1 = 400, 200, 100, 50$ uniform time steps. The reference value is $\mathcal{C} = 3.1651 \cdot 10^{-6}$

N+1	400	200	100	50
AP4o43p	2.99e−6	3.24e−6	3.91e−6	6.02e−6
CPU time [s]	176	51	17	7
AP4o33pa	3.76e−6	6.49e−6	8.12e−6	1.53e−5
CPU time [s]	163	56	27	16
AP4o33pfs	4.17e−6	5.82e−6	1.10e−5	2.62e−5
CPU time [s]	126	72	30	17

`trust-region-reflective` algorithm, which allows a simple way for its allocation.

Let us now present the results for the stopping problem and compare them to those documented in [6]. There, the implicit Euler scheme with $h = 1/80$, i.e., 400 uniform time steps, has been applied, together with a nonlinear cg method and different step size rules. Using the optimal control $u_{stop}(t)$ given above in a forward simulation of the ODE, they found $\mathcal{C} = 3.4814 \cdot 10^{-6}$ as reference value. This nicely compares to our value $\mathcal{C} = 3.1651 \cdot 10^{-6}$ for `AP4o43p` applied with the same time steps. In principle, the optimizer should find a solution close to it when started with $U^{(0)} = u_{stop}$ evaluated at the time points $t_n + c_i h$. Computation times and values of the objective function are collected in Table 4. Remarkably, already for $N + 1 = 50$ all Peer methods deliver excellent approximations in very short time compared to 70–100 s reported in [6] for similar calculations. This is a clear advantage of higher-order methods.

Choosing $U^{(0)} = \beta\, u_{stop}$ with $\beta = 0.99$, the authors of [6] already discovered slow convergence and tiny deviations from u_{stop} in the shape of the computed optimal control, which were clearly visible in their plots. In contrast, all controls computed by the Peer methods stay close to the overall picture shown in Fig. 4 even for $\beta = 0.95, 0.50$, and for 50 times steps. The maximal pointwise control errors range around $6 \cdot 10^{-3}$ and $5 \cdot 10^{-2}$, respectively.

As a last (speculative) test, we impose box constraints of the form

$$u(t) \in U_{ad} := \{u(t) \in [L^{\infty}(0,T)]^m : -0.5 \le u_i(t) \le 0,\ \ i = 1,\ldots,m, t \in (0,T)\}.$$

Now the explicitly given optimal control U_{stop} violates the prescribed bounds with its minimum value −0.638. We apply `AP4o43p` with 50 uniform time steps and set $U^{(0)} = u_{stop}$. `fmincon` first restricts the control vector to the admissible set U_{ad} and after 66 s and 142 iteration steps it delivers a solution with $\mathcal{C} = 0.0323$. The stopping process is still quite satisfactory. Details are plotted in Fig. 5. We get nearly the same solution for 400 uniform time steps. Interestingly, the restricted analytic optimal control $\hat{u}_{stop} \in U_{ad}$ only yields $\mathcal{C} = 0.0850$, which is larger than that of the Peer solution by a factor of 2.6.

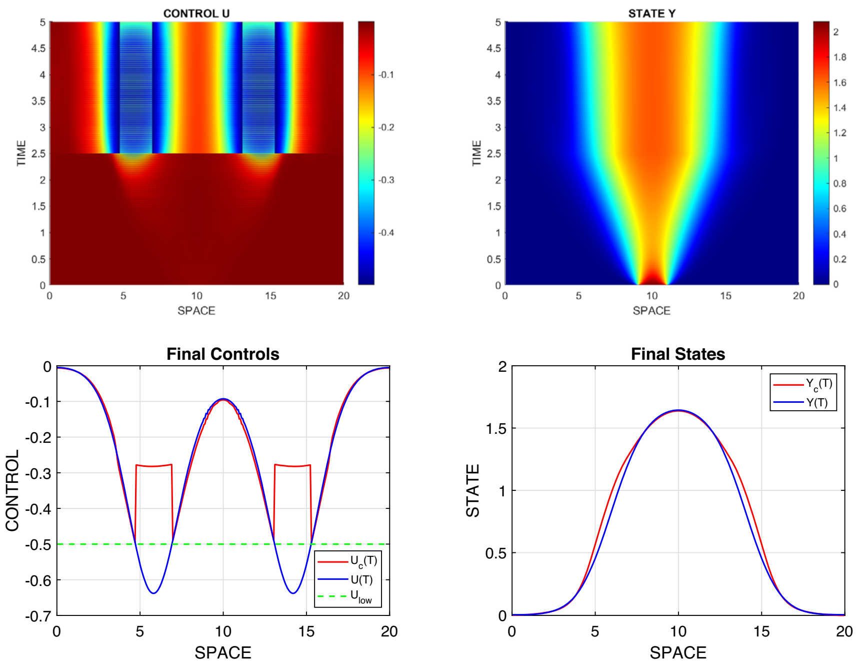

Fig. 5 Nucleation process with box constraints for the control. Computed constrained control $U_c(t)$ (top left), state $Y_h(t)$ (top right), constraint and unconstraint controls at final time $t = T$ (bottom left), and corresponding final states (bottom right) approximated by `AP4o43p` with 50 uniform time steps

9 Summary

We have upgraded our four-stage implicit Peer triplets constructed in [17] to meet the additional order conditions and positivity requirements for an efficient use in a gradient-based iterative solution algorithm for ODE constrained optimal control problems. Using super-convergence for both the state and adjoint variables, an almost A-stable method `AP4o33pa` of the order pair (3, 3) with stability angle $\alpha = 89.90^o$ could be found. We also considered the class of FSAL methods, where the last stage of the previous time step equals the first stage of the new step, and came up with the $A(77.53^o)$-stable method `AP4o33pfs`. A notable theoretical result is that there is no BDF4-based triplet that improves the second-order control approximation of our recently developed `AP4o43bdf` [17] to the present setting. Increasing the order of the forward scheme, an $A(59.78^o)$-stable method `AP4o43p` of the order pair (4, 3) was constructed. All methods show their theoretical orders in the numerical experiments and clearly outperform the fourth-order symplectic Runge–Kutta-Gauss method in the boundary control problem for an 1D discrete heat equation proposed in [18] to study order reduction phenomena. The new Peer triplets also perform remarkably well for a PDE-constrained optimal control problem which models the stopping of a nucleation process driven by a reaction-diffusion equation of Schlögl-type. In future work, we will equip our Peer triplets with variable step sizes to further improve their efficiency.

Funding Open Access funding enabled and organized by Projekt DEAL. The first author is supported by the Deutsche Forschungsgemeinschaft (German Research Foundation) within the collaborative research center TRR154 *"Mathematical modeling, simulation and optimisation using the example of gas networks"* (Project-ID 239904186, TRR154/3-2022, TP B01).

Data Availability Data will be made available on reasonable request.

References

1. Albi, G., Herty, M., Pareschi, L.: Linear multistep methods for optimal control problems and applications to hyperbolic relaxation systems. Appl. Math. Comput. **354**, 460–477 (2019)
2. Almuslimani, I., Vilmart, G.: Explicit stabilized integrators for stiff optimal control problems. SIAM J. Sci. Comput. **43**, A721–A743 (2021)
3. Beck, S., Weiner, R., Podhaisky, H., Schmitt, B.A.: Implicit peer methods for large stiff ODE systems. J. Appl. Math. Comput. **38**, 389–406 (2012)
4. Beigel, D., Mommer, M.S., Wirsching, L., Bock, H.G.: Approximation of weak adjoints by reverse automatic differentiation of BDF methods. Numer. Math. **126**, 383–412 (2014)
5. Bonnans, F.J., Laurent-Varin, J.: Computation of order conditions for symplectic partitioned Runge-Kutta schemes with application to optimal control. Numer. Math. **103**, 1–10 (2006)
6. Buchholz, R., Engel, H., Kammann, E., Tröltzsch, F.: On the optimal control of the Schlögl-model. Comput. Optim. Appl. **56**, 153–185 (2013)
7. Byrd, R.H., Gilbert, J.C., Nocedal, J.: A trust region method based on interior point techniques for nonlinear programming. Math. Program. **89**, 149–185 (2000)
8. Coleman, T.F., Li, Y.: On the convergence of reflective Newton methods for large-scale nonlinear minimization subject to bounds. Math. Program. **67**, 189–224 (1994)
9. Gerisch, A., Lang, J., Podhaisky, H., Weiner, R.: High-order linearly implicit two-step peer - finite element methods for time-dependent PDEs. Appl. Numer. Math. **59**, 624–638 (2009)
10. Hager, W.W.: Rate of convergence for discrete approximations to unconstrained control problems. SIAM J. Numer. Anal. **13**, 449–471 (1976)
11. Hager, W.W.: Runge-Kutta methods in optimal control and the transformed adjoint system. Numer. Math. **87**, 247–282 (2000)
12. Hager, W.W., Rostamian, R.: Optimal coatings, bang-bang controls, and gradient techniques. Optim. Control Appl. Meth. **8**, 1–20 (1987)
13. Hairer, E., Wanner, G., Lubich, Ch.: Geometric Numerical Integration, Structure-preserving Algorithms for Ordinary Differential Equations. Springer series in computational mathematic, Springer, Heidelberg (1970)
14. Herty, M., Pareschi, L., Steffensen, S.: Implicit-explicit Runge-Kutta schemes for numerical discretization of optimal control problems. SIAM J. Numer. Anal. **51**, 1875–1899 (2013)
15. Jebens, S., Knoth, O., Weiner, R.: Explicit two-step peer methods for the compressible Euler equations. Mon. Wea. Rev. **137**, 2380–2392 (2009)
16. Lang, J., Schmitt, B.A.: Discrete adjoint implicit peer methods in optimal control. J. Comput. Appl. Math. **416**, 114596 (2022)

17. Lang, J., Schmitt, B.A.: Implicit A-stable peer triplets for ODE constrained optimal control problems. Algorithms **15**, 310 (2022)
18. Lang, J., Schmitt, B.A.: Exact discrete solutions of boundary control problems for the 1D heat equation. J. Optim. Theory Appl. **196**, 1106–1118 (2023)
19. Lang, J., Schmitt, B.A.: Matlab code LangSchmittOptCtr1DHeat. https://git-ce.rwth-aachen.de/jens.lang/LangSchmittOptCtr1DHeat.git, (2023)
20. Lang, J., Verwer, J.G.: W-methods in optimal control. Numer. Math. **124**, 337–360 (2013)
21. Liu, X., Frank, J.: Symplectic Runge-Kutta discretization of a regularized forward-backward sweep iteration for optimal control problems. J. Comput. Appl. Math. **383**, 113133 (2021)
22. Massa, F.C., Noventa, G., Lorini, M., Bassi, F., Ghidoni, A.: High-order linearly implicit two-step peer methods for the discontinuous Galerkin solution of the incompressible Navier-Stokes equations. Comput. Fluids **162**, 55–71 (2018)
23. Matsuda, T., Miyatake, Y.: Generalization of partitioned Runge-Kutta methods for adjoint systems. J. Comput. Appl. Math. **388**, 113308 (2021)
24. Ostermann, A., Roche, M.: Runge-Kutta methods for partial differential equations and fractional orders of convergence. Math. Comp. **59**, 403–420 (1992)
25. Podhaisky, H., Weiner, R., Schmitt, B.A.: Rosenbrock-type "Peer" two-step methods. Appl. Numer. Math. **53**, 409–420 (2005)
26. Sanz-Serna, J.M.: Symplectic Runge-Kutta schemes for adjoint equations, automatic differentiation, optimal control, and more. SIAM Rev. **58**, 3–33 (2016)
27. Schmitt, B.A., Weiner, R.: Parallel two-step W-methods with peer variables. SIAM J. Numer. Anal. **42**(1), 265–282 (2004)
28. Schmitt, B.A., Weiner, R., Beck, S.: Two-step peer methods with continuous output. BIT **53**, 717–739 (2013)
29. Schneider, D., Lang, J., Hundsdorfer, W.: Extrapolation-based superconvergent implicit-explicit peer methods with A-stable implicit part. J. Comput. Phys. **367**, 121–133 (2018)
30. Schneider, M., Lang, J., Weiner, R.: Super-convergent implicit-explicit Peer methods with variable step sizes. J. Comput. Appl. Math. **387**, 112501 (2021)
31. Schröder, D., Lang, J., Weiner, R.: Stability and consistency of discrete adjoint implicit peer methods. J. Comput. Appl. Math. **262**, 73–86 (2014)
32. Troutman, J.L.: Variational Calculus and Optimal Control. Springer, New York (1996)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.