Journal of Optimization Theory and Applications (2023) 196:1106–1118
https://doi.org/10.1007/s10957-022-02154-4

Exact Discrete Solutions of Boundary Control Problems for the 1D Heat Equation

Jens Lang[1] · Bernhard A. Schmitt[2]

Received: 23 October 2022 / Accepted: 22 December 2022 / Published online: 9 January 2023

Abstract

Method-of-lines discretizations are demanding test problems for stiff integration methods. However, for PDE problems with known analytic solution, the presence of space discretization errors or the need to use codes to compute reference solutions may limit the validity of numerical test results. To overcome these drawbacks, we present in this short note a simple test problem with boundary control, a situation where one-step methods may suffer from order reduction. We derive exact formulas for the solution 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. This analytical setting is used to compare the numerically observed convergence orders for selected implicit Runge–Kutta and Peer two-step methods of classical order four, which are suitable for optimal control problems.

Keywords Optimal control · Implicit Peer two-step methods · Symplectic Runge–Kutta methods · First-discretize-then-optimize · Discrete adjoints

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

Communicated by Lorenz Biegler.

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

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

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

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

1 Introduction

One main area of application for stiff integration methods is semi-discretizations in space of time-dependent partial differential equations in the method-of-lines approach. In order to test new methods in this area, one may rely on PDE test problems with known analytic solution or reference solutions computed by other numerical methods. However, both approaches have its drawbacks. For PDE problems, the accuracy is limited by the level of space discretization errors and in computing reference solutions one has to trust the reliability of the used code. This background was our motivation to develop the current test problems with exact discrete solutions for a finite difference semi-discretization in space with arbitrarily fine grids.

It is known that, in contrast to multi-step-type methods, one-step methods may suffer from order reduction if applied to MOL systems especially with time-dependent boundary conditions, see [7, 8]. Motivated by our recent work [5, 6] on Peer two-step methods in optimal control, the present example is formulated as a problem with boundary control.

The paper is organized as follows. In Sect. 2, we apply a finite difference discretization with a shifted equi-spaced grid for the 1D heat equation with general Robin boundary conditions and derive exact formulas for the solutions of the discrete heat equation and an optimal boundary control problem. These analytical solutions are used in a sparse setting to study the numerically observed convergence orders in Sect. 3 for several one-step and two-step integration methods which are suitable for optimal control. Conclusions are given in Sect. 4.

2 A Discrete Heat Equation with Boundary Control

2.1 Finite Difference Discretization of the 1D Heat Equation

We consider the initial-boundary-value problem for a function $Y(x,t)$ governed by the heat equation

$$\begin{aligned} \partial_t Y(x,t) &= \partial_{xx} Y(x,t),\ \ (x,t)\in[0,1]\times[0,T], && (1)\\ \partial_x Y(0,t) &= 0,\ \ \beta_0 Y(1,t)+\beta_1\partial_x Y(1,t)=u(t), && (2)\\ Y(x,0) &= \Psi(x), \end{aligned}$$

where $\Psi(x)$ and $u(t)$ are given functions. The homogeneous Neumann condition at $x=0$ may be considered as a shortcut for space-symmetric solutions $Y(-x,t)\equiv Y(x,t)$. The coefficients of the general Robin boundary condition are nonnegative, $\beta_0,\beta_1\ge 0$ and nontrivial $(\beta_0,\beta_1)\neq(0,0)$.

Equation (1) is approximated by finite differences with a shifted equi-spaced grid with step size $h=1/m$, $m\in\mathbb{N}$:

$$x_j=\left(j-\frac{1}{2}\right)h,\ \ j=1,\dots,m.$$

For the approximation of the boundary conditions, also the outside points $x_0 = -h/2$ and $x_{m+1} = 1 + h/2$ will be considered temporarily. In the method-of-lines approach with central differences, approximations $y_j(t),\ j = 1, \ldots, m$, are defined by the differential equations

$$y_j' = \frac{1}{h^2}\left(y_{j-1} - 2y_j + y_{j+1}\right),\ j = 2, \ldots, m-1, \tag{3}$$

for the grid points in a distance to the boundary. The symmetric difference approximation $0 \overset{!}{=} hY_x(0,t) \cong (y_1 - y_0)$ leads to the symmetry condition $y_0 \equiv y_1$ and yields the MOL equation

$$y_1' = \frac{-y_1 + y_2}{h^2}. \tag{4}$$

In a similar way, the Robin boundary condition is approximated by the equation

$$\beta_0 \frac{y_m + y_{m+1}}{2} + \beta_1 \frac{y_{m+1} - y_m}{h} = u(t),$$

which may be solved for y_{m+1} by

$$y_{m+1} = \frac{2\beta_1 - \beta_0 h}{2\beta_1 + \beta_0 h} y_m + \frac{2h}{2\beta_1 + \beta_0 h} u(t).$$

Thus, y_{m+1} may be eliminated from Eq. (3) with $j = m$ yielding

$$y_m' = \frac{1}{h^2}(y_{m-1} - \theta y_m) + \gamma\, u(t) \tag{5}$$

with

$$\theta = \frac{2\beta_1 + 3\beta_0 h}{2\beta_1 + \beta_0 h} = 3 - \frac{4\beta_1}{2\beta_1 + \beta_0 h}, \quad \gamma = \frac{2}{(2\beta_1 + \beta_0 h)h}. \tag{6}$$

Hence, we have $\theta = 3$ for the Dirichlet condition and $\theta = 1$ for the pure Neumann condition. Collecting all Eqs. (3), (4) and (5), the following MOL system for the vector $y(t) = \left(y_j(t)\right)_{j=1,\ldots,m}$ is obtained:

$$y' = My + \gamma e_m u(t), \tag{7}$$

$$M = \frac{1}{h^2}\begin{pmatrix} -1 & 1 & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ & & & 1 & -\theta \end{pmatrix}, \tag{8}$$

where e_m is the m-th unit vector. The initial conditions are simple evaluations of the function Ψ on the grid,

$$y(0) = \psi, \quad \psi = \left(\Psi(x_j)\right)_{j=1}^{m}. \tag{9}$$

The basis of our construction is that the eigenvalues and eigenvectors of the symmetric matrix M are known, which is well known for special values of θ, at least.

Lemma 2.1 *For $m \geq 2$, the eigenvalues of the matrix $M \in \mathbb{R}^{m\times m}$ from (8) are given by*

$$\lambda_k = -4m^2 \sin^2\left(\frac{\omega_k}{2m}\right),\ k = 1, \ldots, m, \tag{10}$$

where $\omega_k,\ k = 1, \ldots, m$, are the m first nonnegative solutions of the equation

$$\tan(\omega)\tan\left(\frac{\omega}{2m}\right) = \frac{\beta_0}{2m\beta_1}, \tag{11}$$

with the convention that $\omega_k = (k-\frac{1}{2})\pi$, $k = 1, \ldots, m$, for $\beta_1 = 0$. The corresponding normalized eigenvectors $v^{[k]}$ have the components

$$v_j^{[k]} = \nu_k \cos\left(\omega_k \frac{2j-1}{2m}\right),\ j = 1, \ldots, m, \tag{12}$$

with constants $\nu_k = 2/\sqrt{2m + \sin(2\omega_k)/\sin(\omega_k/m)}$.

Proof In the main Eq. (3), the ansatz $v = \left(\Re e^{i\omega x_j}\right)_{j=1}^{m}$ gives

$$\frac{1}{h^2}(v_{j-1} - 2v_j + v_{j+1}) = \frac{1}{h^2}\Re e^{i\omega x_j}\left(e^{-i\omega h)} - 2 + e^{i\omega h}\right) = -\frac{4}{h^2}\sin^2\left(\frac{\omega h}{2}\right)v_j.$$

In the first equation, we have

$$\begin{aligned}\frac{1}{h^2}(-v_1 + v_2) &= \frac{1}{h^2}\left(-\cos\frac{\omega h}{2} + \cos\left(3\frac{\omega h}{2}\right)\right) = \frac{4}{h^2}\left(\cos^3\left(\frac{\omega h}{2}\right) - \cos\frac{\omega h}{2}\right)\\ &= -\frac{4}{h^2}\sin^2\left(\frac{\omega h}{2}\right)v_1,\end{aligned}$$

with the same factor $\lambda := -(4/h^2)\sin^2\left(\omega h/2\right)$. In order to satisfy the eigenvalue condition in the last component, we consider the equation $0 = e_m^T(Mv - \lambda v)$, i.e.,

$$\begin{aligned}0 &\overset{!}{=} v_{m-1} - \left(\theta + \lambda h^2\right)v_m = \cos\left(\omega(x_m - h)\right) - \left(\theta + \lambda h^2\right)\cos(\omega x_m)\\ &= \left(\cos(\omega h) + 4\sin^2\left(\frac{\omega h}{2}\right) - \theta\right)\cos(\omega x_m) + \sin(\omega h)\sin(\omega x_m)\end{aligned}$$

$$= \left(1 + 2\sin^2\left(\frac{\omega h}{2}\right) - \theta\right)\cos(\omega x_m) + \sin(\omega h)\sin(\omega x_m),$$

since $\cos(\omega h) = 1 - 2\sin^2(\omega h/2)$. The last grid point is $x_m = 1 - h/2$ and with the trigonometric formulas for $\cos(\omega - \omega h/2)$, $\sin(\omega - \omega h/2)$ and the identity $\sin(\omega h) = 2\sin(\omega h/2)\cos(\omega h/2)$, we may proceed with

$$\begin{aligned}
0 &= \left(1 + 2\sin^2\left(\frac{\omega h}{2}\right) - \theta\right)\left(\cos(\omega)\cos\left(\frac{\omega h}{2}\right) + \sin(\omega)\sin\left(\frac{\omega h}{2}\right)\right)\\
&\quad + 2\sin\left(\frac{\omega h}{2}\right)\cos\left(\frac{\omega h}{2}\right)\left(\sin(\omega)\cos\left(\frac{\omega h}{2}\right) - \cos(\omega)\sin\left(\frac{\omega h}{2}\right)\right)\\
&= (1-\theta)\cos\left(\frac{\omega h}{2}\right)\cos(\omega)\\
&\quad + \left(1 + 2\sin^2\left(\frac{\omega h}{2}\right) - \theta + 2\cos^2\left(\frac{\omega h}{2}\right)\right)\sin\left(\frac{\omega h}{2}\right)\sin(\omega)\\
&= (1-\theta)\cos\left(\frac{\omega h}{2}\right)\cos(\omega) + (3-\theta)\sin\left(\frac{\omega h}{2}\right)\sin(\omega).
\end{aligned}$$

Hence, the different versions of θ in (6) verify the condition (11) for $m = 1/h$. Rearranging (11) as $\tan(\omega) = \beta_0/(2m\beta_1)\cot(\omega/(2m))$, for $\beta_0 > 0$ it is seen that exactly m solutions exist in $(0, m\pi)$ since the function $\omega \mapsto \cot(\omega/(2m))$ is monotonically decreasing and positive. Finally, the vector norms are computed for $\omega \neq 0$. Abbreviating $\omega/m =: \Omega$ and using $\cos^2(x) = (1 + \cos(2x))/2$, we get

$$\begin{aligned}
\sum_{j=1}^{m}\cos^2\left(\left(j - \frac{1}{2}\right)\frac{\omega}{m}\right) &= \frac{m}{2} + \frac{1}{2}\sum_{j=1}^{m}\cos\left((2j-1)\Omega\right)\\
&= \frac{m}{2} + \frac{1}{2}\Re\sum_{j=1}^{m} e^{i(2j-1)\Omega} = \frac{m}{2} + \frac{1}{2}\Re\frac{e^{i2\Omega m} - 1}{e^{i\Omega} - e^{-i\Omega}}\\
&= \frac{m}{2} + \frac{1}{4}\Im\frac{e^{i2\Omega m} - 1}{\sin(\Omega)} = \frac{m}{2} + \frac{1}{4}\frac{\sin(2\omega)}{\sin(\Omega)},
\end{aligned}$$

which leads to the value of the normalizing factor ν in (12). □

For later use, we introduce the diagonal matrix $\Lambda = \operatorname{diag}(\lambda_k)$ and the unitary matrix $V = (v^{[1]}, \dots, v^{[m]})$ satisfying $M = V\Lambda V^T$.

Remark 2.2 The well-known frequencies for Dirichlet boundary conditions are $\omega_k = (k - \frac{1}{2})\pi$ and $\omega_k = (k-1)\pi$, $k = 1, \dots, m$ for Neumann conditions. For general values $\beta_0, \beta_1 > 0$, Eq. (11) may be rewritten in fixed point form

$$\omega = f_k(\omega) := (k-1)\pi + \arctan\left(\frac{\beta_0}{2m\beta_1}\cot\left(\frac{\omega}{2m}\right)\right), \quad k = 1, \dots, m.$$

The functions f_k are monotonically decreasing in ω, and an iteration with initial value $\omega = \max\{1, (k-1)\pi\}$ converges to the desired solution ω_k at least for $2m > \beta_0/\beta_1 \geq 1$, since $1/|f_k'(\omega)| = (4m^2\beta_1/\beta_0)\sin^2(\omega/(2m)) + (\beta_0/\beta_1)\cos^2(\omega/(2m))$.

Remark 2.3 The eigenvalues λ_k in (10) are $O(h^2)$-approximations of the exact eigenvalues $\hat{\lambda}_k = -\varphi_k^2$, where $y_k(x) = \cos(\varphi_k x)$, $k \in \mathbb{N}$, are the eigenfunctions of the boundary value problem with frequencies φ_k satisfying $\varphi\tan(\varphi) = \beta_0/\beta_1$. For Dirichlet ($\beta_1 = 0$) and Neumann ($\beta_0 = 0$) conditions, the discrete frequencies are exact, $\omega_k = \varphi_k$, $k = 1, \ldots, m$. Here, Taylor expansion in (10) shows that $\lambda_k = -4h^{-2}\sin^2(h\omega_k/2) \cong -\omega_k^2(1 + h^2\omega_k^2/12) = \hat{\lambda}_k + O(h^2\omega_k^3)$. This shows convergence of second order for fixed k. However, for $k \to m$ the estimate becomes meaningless since then $h\omega_k = O(1)$. For general Robin conditions, $\beta_0, \beta_1 > 0$, an additional error is added since (11) corresponds to $\beta_0/\beta_1 = 2h^{-1}\tan(h\omega/2)\tan(\omega) \cong \omega\tan(\omega)(1+h^2\omega^2/12)$, which is an $O(h^2)$-perturbation of the condition for the exact frequencies φ_k.

2.2 Exact Solution of the Discrete Heat Equation

Knowing the eigenvectors and eigenvalues of the linear problem (7), the computation of its solution is straightforward. The representation $y(t) = \sum_{k=1,\ldots,m} \eta_k(t)v^{[k]}$ leads to

$$\sum_{k=1}^{m} \eta_k'(t)v^{[k]} = \sum_{k=1}^{m} \lambda_k \eta_k(t)v^{[k]} + \gamma e_m u(t).$$

Since the matrix M is symmetric, the inner product with $v^{[j]}$ yields the decoupled equations $\eta_j'(t) = \lambda_j\eta_j(t) + \gamma v_m^{[j]}u(t)$, which can be solved easily leading to the following result.

Lemma 2.4 *With the data from Lemma 2.1, the solution of the initial value problem (7), (9) is given by*

$$y(t) = \sum_{k=1}^{m} \left(e^{\lambda_k t} v^{[k]^T} \psi + \gamma v_m^{[k]} \int_0^t e^{\lambda_k(t-\tau)} u(\tau)\,\mathrm{d}\tau \right) v^{[k]}. \tag{13}$$

Remark 2.5 The presence of the terms $v_m^{[k]}$ indicates that simple sparse solutions with only a few terms in (13) may not exist due to the inhomogeneous boundary condition (2).

2.3 Exact Solution of an Optimal Control Problem

The inhomogeneity $u(t)$ inherited from the boundary condition (2) may be considered as a control to approach a given target profile $\hat{y} \in \mathbb{R}^m$ at some given time $T > 0$. In an optimal control context, controls are searched for minimizing an objective function like

$$C = \frac{1}{2}\|y(T) - \hat{y}\|_2^2 + \frac{\alpha}{2}\int_0^T u(t)^2 dt,$$

with the Euclidean vector norm $\|\cdot\|_2$ in $\mathbb{R}^m$ and $\alpha > 0$. The unique optimal solution may be computed by using some multiplier function $p(t)$ for the ODE restriction (7) and considering the Lagrangian

$$\begin{aligned} L :=& C + \int_0^T p^T\left(y' - My - \gamma e_m u\right)dt + p^T(0)(y(0) - \psi) \\ =& C - \int_0^T \left(\left(p'\right)^T y + p^T\left(My + \gamma e_m u\right)\right)dt + p^T(T)y(T) - p^T(0)\psi. \end{aligned}$$

The partial derivatives of the Lagrangian L with respect to $p(t)$ and $p(T)$ recover (7), (9), and the other ones are

$$\begin{aligned} \partial_{y(t)}L =& -p' - M^T p = -p' - Mp, \qquad (14)\\ \partial_{y(T)}L =& y(T) - \hat{y} + p(T), \\ \partial_{u(t)}L =& \alpha u - \gamma e_m^T p. \qquad (15) \end{aligned}$$

Hence, the Karush–Kuhn–Tucker conditions, $\partial_{(\cdot)}L = 0$ in (14)–(15), show that the control $u(t)$ may be eliminated by

$$u(t) = \frac{\gamma}{\alpha}e_m^T p(t) = \frac{\gamma}{\alpha}p_m(t), \qquad (16)$$

and a necessary condition for the optimal solution is that it solves the following boundary value problem:

$$\begin{aligned} y' =& My + \frac{\gamma}{\alpha}e_m e_m^T p, \quad y(0) = \psi, \qquad (17)\\ p' =& -Mp, \quad p(T) = \hat{y} - y(T). \qquad (18) \end{aligned}$$

The homogeneous differential equation (18) for p has the simple solution

$$p(t) = e^{(T-t)M}p(T) = \sum_{\ell=1}^m e^{\lambda_\ell(T-t)}v^{[\ell]}v^{[\ell]^T}(\hat{y} - y(T))$$

with the matrix exponential

$$e^{sM} = V\ \mathrm{diag}\left(e^{\lambda_k s}\right)V^T, \quad 0 \le s \le T.$$

With u given by (16), this solution may be used in (13) to yield the solution for (17) with the coefficient functions

$$\begin{aligned}\eta_k(t) &= e^{\lambda_k t}\eta_k(0) + \frac{\gamma^2}{\alpha} v_m^{[k]} \int_0^t e^{\lambda_k(t-\tau)} p_m(\tau) \mathrm{d}\tau \\ &= e^{\lambda_k t}\eta_k(0) + \frac{\gamma^2}{\alpha} v_m^{[k]} \sum_{\ell=1}^m v_m^{[\ell]} \int_0^t e^{\lambda_k t + \lambda_\ell T - (\lambda_k+\lambda_\ell)\tau} \mathrm{d}\tau \cdot v^{[\ell]^T}(\hat{y} - y(T)).\end{aligned} \tag{19}$$

Considering the function $\varphi_1(z) = \int_0^1 e^{zt} dt$ satisfying $\varphi_1(z) = (e^z - 1)/z$ for $z \neq 0$ and $\varphi_1(0) = 1$, the integral may be written as

$$\int_0^t e^{\lambda_k t + \lambda_\ell T - (\lambda_k+\lambda_\ell)\tau} \mathrm{d}\tau = t e^{\lambda_\ell (T-t)} \varphi_1\left((\lambda_k + \lambda_\ell)t\right).$$

The result (19) may be used in different ways. The first application is computing the solution for a given target profile $\hat{y}$.

Lemma 2.6 *Let $\hat{y} \in \mathbb{R}^m$ be given. Then, the coefficient vector $\eta(T)$ of the solution $y(t) = V\eta(t)$ of the boundary value problem* (17), (18) *is given by the unique solution of the linear system*

$$(I + Q)\eta(T) = e^{T\Lambda}\eta(0) + QV^T\hat{y}, \tag{20}$$

with the positive semi-definite matrix $Q = (q_{k\ell})_{k,\ell=1}^m$ having the elements

$$q_{k\ell} = \frac{\gamma^2 T}{\alpha} v_m^{[k]} \varphi_1\left((\lambda_k + \lambda_\ell)T\right) v_m^{[\ell]}, \; k, \ell = 1, \ldots, m. \tag{21}$$

Proof At the end point T, the formula (19) simplifies to

$$\eta_k(T) = e^{\lambda_k T}\eta_k(0) + \frac{\gamma^2 T}{\alpha} v_m^{[k]} \sum_{\ell=1}^m v_m^{[\ell]} \varphi_1\left((\lambda_k + \lambda_\ell)T\right) \cdot \left(v^{[\ell]^T}\hat{y} - \eta_\ell(T)\right).$$

This equation may be reordered to the form given in (20) with the matrix elements (21). Finally, we consider the quadratic form of the matrix Q with some vector $w = (w_j)$, obtaining

$$\begin{aligned} w^T Q w &= \frac{\gamma^2 T}{\alpha} \sum_{k,\ell=1}^m v_m^{[k]} w_k \int_0^1 e^{(\lambda_k+\lambda_\ell)T\tau} \mathrm{d}\tau \cdot v_m^{[\ell]} w_\ell \\ &= \frac{\gamma^2 T}{\alpha} \int_0^1 \sum_{k,\ell=1}^m (e^{\lambda_k T\tau} v_m^{[k]} w_k)(e^{\lambda_\ell T\tau} v_m^{[\ell]} w_\ell) \mathrm{d}\tau \\ &= \frac{\gamma^2 T}{\alpha} \int_0^1 \Big(\sum_{k=1}^m e^{\lambda_k T\tau} v_m^{[k]} w_k\Big)^2 \mathrm{d}\tau \geq 0. \end{aligned}$$

This means that Q is semi-definite and $I + Q$ definite and the system (20) always has a unique solution. □

In general, solutions computed with (20) will not be sparse, i.e., they will have m nontrivial basis coefficients in the state y and the Lagrange multiplier p. Due to the special inhomogeneity in (17), sparse solutions for the state y probably do not exist. However, by adjusting the target profile $\hat{y}$, one may simply start with a sparse multiplier $p(t)$ with, for instance, two terms only,

$$p(t) = \delta_1 e^{\lambda_1(T-t)} v^{[1]} + \delta_2 e^{\lambda_2(T-t)} v^{[2]}, \tag{22}$$

with coefficients δ_1, δ_2 belonging to some *reasonable* form of the control u. Then, by the boundary condition in (18), the corresponding target profile has the form

$$\hat{y} = y(T) + \delta_1 v^{[1]} + \delta_2 v^{[2]}, \tag{23}$$

where, by (19), the coefficients of $y(T)$ are given by

$$\eta_k(T) = e^{\lambda_k T} \eta_k(0) + \frac{\gamma^2 T}{\alpha} v_m^{[k]} \sum_{\ell=1}^{2} \delta_\ell v_m^{[\ell]} \varphi_1\left((\lambda_k + \lambda_\ell)T\right). \tag{24}$$

We will use this construction in our numerical example.

3 Test Case: Dirichlet Boundary Control Problem

To illustrate an application of the derived expressions for the exact discrete solutions of the linear heat equation equipped with different boundary conditions, we consider the following ODE-constrained optimal control problem with an incorporated boundary control of Dirichlet type:

$$\begin{aligned} \min_{(y,u)} C \; &:= \frac{1}{2}\|y(T) - \hat{y}\|_2^2 + \frac{\alpha}{2}\int_0^T u(t)^2\, dt \\ \text{subject to } y'(t) \; &= My(t) + \gamma e_m u(t), \quad t \in (0, T], \\ y(0) \; &= \mathbb{1}, \end{aligned}$$

with $T = 1$, $\gamma = 2/h^2$, $\alpha = 1$, $\mathbb{1} = (1, \ldots, 1)^T \in \mathbb{R}^m$, state vector $y(t) \in \mathbb{R}^m$, and M as defined in (8) with $\theta = 3$. We set $\delta_1 = \delta_2 = -1/75$ in (22) and compute the target profile $\hat{y} \in \mathbb{R}^m$ from (23) with coefficients for $y(T)$ defined in (24).

We will compare numerical results for four time integrators of classical order four: the symmetric 2-stage Gauss method (Appendix: Table 1), the symmetric 3-stage partitioned Runge–Kutta pair Lobatto IIIA-IIIB (Appendix: Table 2) and our recently developed two-step Peer methods AP4o43bdf and AP4o43dif [6]. The two one-step methods are symplectic and therefore well suited for optimal control [4, 9].

Two test scenarios are considered. First, the accuracy of the numerical approximations for $y(T)$ and $p(0)$ is studied, where the exact control $u(t) = \gamma p_m(t)/\alpha$ is used. The initial value for the multiplier is set to $p(T) = \hat{y} - y_\tau(T)$ with $y_\tau(T)$ being the approximation of $y(T)$ with time step τ. In this case, the Karush–Kuhn–Tucker system decouples and only two systems of linear ODEs have to be solved. In the second scenario, the optimal control problem is solved for all unknowns (y, p, u) by a gradient-based interior point algorithm as implemented in the MATLAB routine *fmincon*, see, e.g., [1, 2] for more details, and the errors for the control are discussed.

We first reduce the objective function $C(y, u)$ to the so-called Mayer form, which uses terminal solution values only. Introducing an additional differential equation $y'_{m+1}(t) = u(t)^2$ with initial values $y_{m+1}(0) = 0$ and an extended state vector $\tilde{y} = (y^T, y_{m+1})^T$, the new objective function reads $\tilde{C} = (\|y(T) - \hat{y}\|_2^2 + \alpha y_{m+1}(T))/2$. Let now $U \in \mathbb{R}^{sN}$ denote the vector of approximate control values at the nodes $t_{ni} = t_n + c_i\tau, i = 1, \ldots, s$, used by an s-stage time integrator on a time grid $\{t_0, \ldots, t_N\}$ with step size τ [3, 6], and let $\tilde{C}(U) := \tilde{C}(\tilde{y}(T))$ be the value of the cost functional associated with these discrete controls. Then, the Karush–Kuhn–Tucker system provides a convenient way to compute the gradient $\nabla_U \tilde{C}(U)$ for a given $U^{(k)}$ in an iterative optimization algorithm, solving first the forward Eq. (7) for y with given intermediate values for the control to compute approximations y_τ and then the backward Eq. (18) for p using $y_\tau(T)$ as approximation for $y(T)$. For Runge–Kutta methods, it holds $\nabla_{u_\tau(t_{ni})}\tilde{C}(U) = -\tau b_i p_\tau(t_{ni})\nabla_u f(y_\tau(t_{ni}), u_\tau(t_{ni}))$ [3, Formula (27)], where $y_\tau(t_{ni})$, $p_\tau(t_{ni})$, and $u_\tau(t_{ni})$ are the approximations of $(y(t), p(t), u(t))$ at $t = t_{ni}$, respectively, and b_i are the weights of the Runge–Kutta method, see Appendix. Similar formulas are computable for other time integrators. Eventually, we set $U^{(0)} = 0$ as initial guess and call *fmincon*, where we provide gradients of the objective function for each approximation $U^{(k)}$.

In both test cases, we use $m = 250$ and $m = 500$ to also study the influence of the system size. The number of time steps are $N = 2^k$ with $k = 4, \ldots, 11$.

In Fig. 1, results for the first test scenario are shown. Not surprisingly, the serious order reduction for the symplectic one-step Runge–Kutta methods is clearly seen. This phenomenon is well understood and occurs particularly drastically for time-dependent Dirichlet boundary conditions [8]. This drawback is shared by all one-step methods due to their insufficient stage order. Note that the number of affected time steps increases when the system size is doubled. In contrast, the newly designed two-step Peer methods for optimal control problems work quite close to their theoretical order four for the state y and the adjoint p. The order reduction for the one-step methods is also visible for the more challenging fully coupled problem. The results plotted in Fig. 2 show a reduction to first order for the approximation of the control, whereas the two-step methods perform with order two for this problem. We refer to [3] for a discussion of the convergence order for general ODE constrained optimal control problems. Once again, the range of the affected time steps depends on the problem size. It increases for finer spatial discretizations.

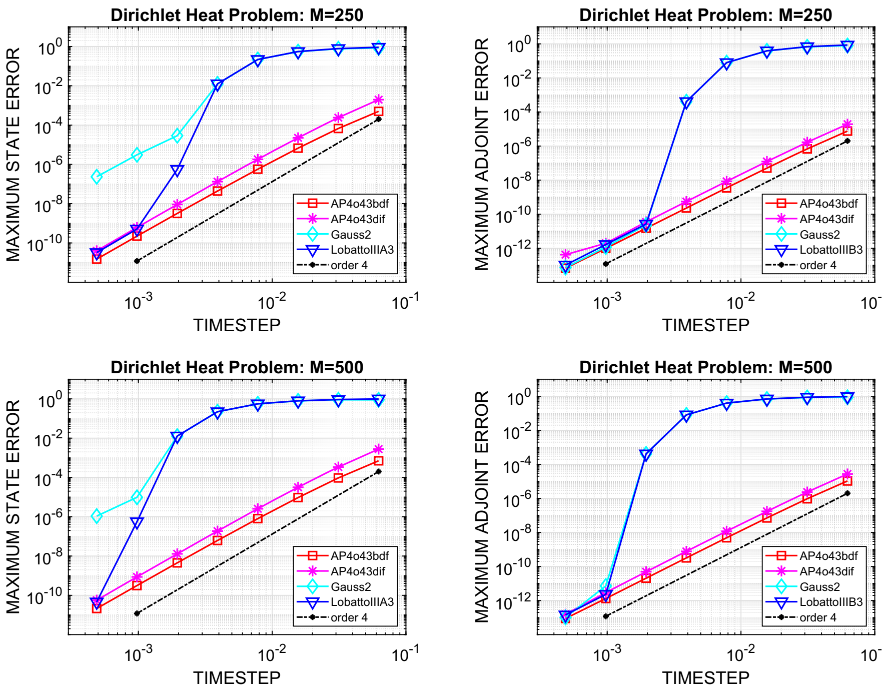

Fig. 1 Dirichlet heat problem with $m = 250, 500$ spatial points and given exact control $u(t)$: $\|y(T) - y_\tau(T)\|_\infty$ (left) and $\|p(0) - p_\tau(0)\|_\infty$ (right)

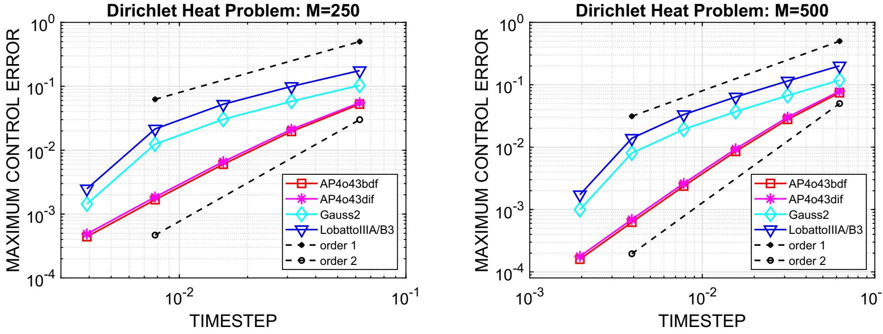

Fig. 2 Dirichlet heat problem with $m = 250, 500$ spatial points, solved by MATLAB's gradient-based *fmincon* with interior point algorithm for (y, p, u): $\max_{n,i} |u(t_{ni}) - u_\tau(t_{ni})|$

4 Conclusions

We have derived exact formulas for the solution of an optimal boundary control problem constrained by a one-dimensional discrete heat equation, including Dirichlet and general Robin boundary conditions. These solutions have been used to compare symplectic Runge–Kutta methods and recently developed Peer two-step methods of order four. The numerically observed convergence orders illustrate a serious order reduction for Runge–Kutta methods, which is much less severe for our Peer two-step methods.

Acknowledgements The first author is supported by the Deutsche Forschungsgemeinschaft (DFG, 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).

Funding Open Access funding enabled and organized by Projekt DEAL.

Data Availability Data will be made available on reasonable request.

Appendix: Symplectic Runge–Kutta methods

An implicit s-stage Runge–Kutta method to numerically solve $y'(t) = f(t, y), y(0) = y_0$, with constant step size $\tau > 0$ on a uniform partition $0 < t_1 < \ldots < t_N, t_n = n\tau$, is given by

$$\begin{aligned} k_i &= f\left(t_n + c_i\tau,\, y_n + \tau\sum_{j=1}^{s} a_{ij}k_j\right),\ \ i = 1, \ldots, s, \\ y_{n+1} &= y_n + \tau\sum_{j=1}^{s} b_i k_i, \end{aligned}$$

where b_i, a_{ij} are real numbers, $c_i = \sum_{j=1}^{s} a_{ij}$, and y_n are approximations of $y(t_n)$. The coefficients are usually displayed in a Butcher tableau

$$\begin{array}{c|ccc} c_1 & a_{11} & \ldots & a_{1s} \\ \vdots & \vdots & & \vdots \\ c_s & a_{s1} & \ldots & a_{ss} \\ \hline & b_1 & \ldots & b_s \end{array}\ .$$

We give the coefficients of the symplectic 2-stage Gauss method in Table 1 and the symplectic 3-stage Lobatto IIIA-IIIB pair in Table 2. Further symplectic methods and useful information can be found in [4].

Table 1 Coefficients of the 2-stage Gauss method of order 4

$1/2-\sqrt{3}/6$	$1/4$	$1/4-\sqrt{3}/6$
$1/2+\sqrt{3}/6$	$1/4+\sqrt{3}/6$	$1/4$
	$1/2$	$1/2$

Table 2 Coefficients of the 3-stage Lobatto IIIA-IIIB pair of order 4

0	0	0	0	0	1/6	-1/6	0
1/2	5/24	1/3	-1/24	1/2	1/6	1/3	0
1	1/6	2/3	1/6	1	1/6	5/6	0
	1/6	2/3	1/6		1/6	2/3	1/6

References

1. Betts, J.T.: Practical Methods for Optimal Control and Estimation Using Nonlinear Programming, 2nd edn. Society for Industrial and Applied Mathematics (2010)
2. Byrd, R.H., Hribar, M., Nocedal, J.: An interior point algorithm for large-scale nonlinear programming. SIAM J. Optim. **9**, 877–900 (1999)
3. Hager, W.W.: Runge–Kutta methods in optimal control and the transformed adjoint system. Numer. Math. **87**, 247–282 (2000)
4. Hairer, E., Lubich, C., Wanner, G.: Geometric Numerical Integration, Structure-Preserving Algorithms for Ordinary Differential Equations, Springer Series in Computational Mathematics, vol. 31. Springer, Heidelberg, Berlin (2006)
5. Lang, J., Schmitt, B.A.: Discrete adjoint implicit peer methods in optimal control. J. Comput. Appl. Math. **416**, 114596 (2022)
6. Lang, J., Schmitt, B.A.: Implicit A-stable peer triplets for ODE constrained optimal control problems. Algorithms **15**, 310 (2022)
7. Lubich, C., Ostermann, A.: Runge–Kutta approximation of quasi-linear parabolic equations. Math. Comput. **64**, 601–627 (1995)
8. Ostermann, A., Roche, M.: Runge–Kutta methods for partial differential equations and fractional orders of convergence. Math. Comput. **59**, 403–420 (1992)
9. Sanz-Serna, J.M.: Symplectic Runge–Kutta schemes for adjoint equations, automatic differentiation, optimal control, and more. SIAM Rev. **58**, 3–33 (2016)