Abstract. We consider one-dimensional distributed optimal control problems with the state equation being given by the viscous Burgers equation. We discretize using a space-time discontinuous Galerkin approach. We use upwind flux in time and the symmetric interior penalty approach for discretizing the viscous term. Our focus is on the discretization of the convection terms. We aim for using conservative discretizations for the convection terms in both the state and the adjoint equation, while ensuring that the approaches of discretize-then-optimize and optimize-then-discretize commute. We show that this is possible if the arising source term in the adjoint equation is discretized properly, following the ideas of well-balanced discretizations for balance laws. We support our findings by numerical results.

1. Introduction. We consider the following optimal control problem (OCP) on the space-time domain $Q = \Omega \times (0, T)$ with $\Omega = (x_L, x_R) \subset \mathbb{R}$:

$$
\min_{q, u} J(q, u) = \frac{1}{2} \|u - u_d\|_{L^2(Q)}^2 + \frac{\alpha}{2} \|q\|_{L^2(Q)}^2
$$

subject to

$$
\begin{align*}
    u_t + f(u)_x - \varepsilon u_{xx} &= q + g & \text{in } Q, \\
    u(x_L, \cdot) &= u(x_R, \cdot) = 0 & \text{in } (0, T), \\
    u(\cdot, 0) &= u_0 & \text{in } \Omega.
\end{align*}
$$

Here, $q$ denotes the control and $u$ the state. The parameter $\alpha$ enforces the Tikhonov regularization and we assume $\varepsilon > 0$ to be constant. In this contribution we focus on $f(u) = \frac{1}{2} u^2$. Then, the state equation corresponds to the viscous Burgers equation with distributed control $q$ and a source term $g$. Considering the viscous Burgers equation is a suitable intermediate step towards extending methods to the (compressible) Navier-Stokes equations as we need to deal with a non-linear convection term.

There exists a variety of contributions covering OCPs for viscous Burgers equation with distributed control, see for example [4, 7, 9, 20, 28, 30] and the references cited therein. In this work we solve (1) and (2) by means of the reduced approach using a space-time discontinuous Galerkin (DG) discretization for the state and

---

2020 Mathematics Subject Classification. Primary: 49M41; Secondary: 65M60.

Key words and phrases. Optimal control, discontinuous Galerkin method, optimize-then-discretize, viscous Burgers equation, conservative formulation.

* Corresponding author: Xenia Kerkhoff.
the adjoint equation. We discretize the diffusion term with the symmetric interior penalty (SIPG) discretization \[3\] and use an upwind flux in time. For the non-linear convection term in the state equation we consider two different fluxes, the Lax-Friedrichs (LF) flux and the Engquist-Osher (EO) flux. We focus on finding corresponding matching fluxes for the linearized convection term in the adjoint equation.

Ideally, the discretization of the state and of the adjoint equation should be chosen such that the approaches of \textit{discretize-then-optimize} (DO) and \textit{optimize-then-discretize} (OD) commute. Leykekhman \[22\] examined the commutativity for DG discretizations of a steady advection diffusion equation. He showed the commutative properties of the SIPG discretization. Akman and Karasözen \[1\] extended this to the time-dependent case using DG in time. We will focus on the proper treatment of the (non-)linear convection terms in this contribution.

We start with a standard conservative space-time DG discretization of the state equation (2). In the DO approach, the discretization of the adjoint equation follows from a straightforward but potentially lengthy computation. In the OD approach, one is free to discretize the adjoint equation. In a continuous setting the adjoint equation (dropping initial and boundary conditions) is given by

\[-z_t - f'(u)z_x - \varepsilon z_{xx} = u - u_d \text{ in } Q,\]

with \(z\) denoting the adjoint state. Different to (2), the convection term is given in non-conservative form.

It therefore seems natural to use a non-conservative discretization. This requires to implement a new discretization. It is preferable though to reuse as much as possible from the code for solving the state equation for solving the adjoint equation as well. This way, it is fairly straightforward to extend an existing solver for the state equation (2) to a solver for the OCP (1) and (2) by means of the reduced approach.

This requires to rewrite the convection term in the adjoint equation in a conservative formulation in order to apply the given conservative discretization. On a continuous level, one can easily do this by introducing a source term, which has the form of a reaction term:

\[-z_t + (-f'(u)z)_x - \varepsilon z_{xx} = -f'(u)z + u - u_d \text{ in } Q.\]  

Let us focus on the reduced equation

\[-z_t + (-f'(u)z)_x = -f'(u)z.\]  

This equation belongs to the category of balance laws. Balance laws allow for non-trivial stationary solutions, for which the effects of the convective term and the source term, which might come, e.g., from geometrical or reactive considerations \[24\], cancel each other. A good discretization should (exactly) preserve these stationary solutions. Such schemes are known as well-balanced schemes. There exists a significant amount of literature covering well-balanced methods, often focusing on the shallow water equations, see for example the overview article \[24\] as well as \[5,13,14,17\] and the references cited therein. To stay within the terminology of well-balanced schemes, we will refer in the following to the term \(-f'(u)z\) as a source term instead of calling it a reaction term.

We will see that for our OCP there is a close connection between using a well-balanced DG discretization for the convection term in the adjoint equation and the commutativity property of OD vs. DO: when we use a standard discretization
for (3), we violate commutativity and do not preserve stationary solutions of (4). However, a discretization for (3) that ensures commutativity also preserves the stationary solutions of (4).

The paper is structured as follows: In section 2, we formulate the considered OCP in more detail and describe the discretization of the state equation that we use. In section 3, we briefly discuss the commutative properties of the linear terms in the discretization of the state equation as this is not the focus of our work. Sections 4 and 5 are devoted to the commutative properties of the convection terms: in section 4 we present the resulting discrete adjoint equation when using the DO approach; in section 5 we apply the OD approach and discuss a suitable discretization for the adjoint equation, which is in agreement with the well-balanced idea. In section 6 we present numerical results to support our theoretical findings. We conclude with a short summary in section 7.

2. Problem setup. We consider the OCP given by (1) and (2) on the space-time domain $Q = \Omega \times (0, T)$ with $\Omega = (x_L, x_R) \subset \mathbb{R}$. By $L^2(0, T; H^1_0(\Omega))$ we denote the space of square integrable functions in the sense of Bochner from $(0, T)$ to $H^1_0(\Omega)$. We then define

$$W = W(0, T; H^1_0(\Omega)) = \{ v : v \in L^2(0, T; H^1_0(\Omega)), v_t \in L^2(0, T; H^{-1}(\Omega)) \}.$$ 

We assume $\varepsilon > 0$ to be constant (and to be potentially small) and we will focus on the case of $f(u) = \frac{1}{2} u^2$, i.e., the viscous Burgers equation. Further, $u_d \in L^2(Q)$ denotes the desired state, $g \in L^2(0, T; L^2(\Omega))$ a source term, and $u^0 \in L^2(\Omega)$ given initial data.

In the following, we will assume that there exists a unique solution $(q, u) \in L^2(Q) \times W$ to (1) and (2) as discussed by Volkwein [28].

For better readability of the remaining paper, we already provide the adjoint equation in strong non-conservative form: find $z \in W$ such that it satisfies a weak formulation of

$$-z_t - f'(u)z_x - \varepsilon z_{xx} = u - u_d \quad \text{in } Q,$$

$$z(x_L, \cdot) = z(x_R, \cdot) = 0 \quad \text{in } (0, T),$$

$$z(\cdot, T) = 0 \quad \text{in } \Omega.$$ 

2.1. OD vs. DO. Generally, there are two approaches for discretizing the optimal control problem (1) and (2):

- **Optimize-then-discretize** (OD): one first sets up the optimality system on the continuous level and then discretizes each of the three equations (state equation, adjoint equation, optimality condition), potentially independent of each other.

- **Discretize-then-optimize** (DO): one first discretizes the optimal control problem including the state equation and then sets up the optimality system on a discrete level.

In the DO approach one needs to differentiate through the discretization of the state equation. This approach guarantees an exact discrete gradient but might be very tedious to execute. This process automatically implies a discretization for the adjoint problem, which might not be consistent. Also, error estimates for the resulting discretization of the adjoint equation might not be available, which are typically needed for proving error estimates for the full optimal control problem. In
the OD approach on the other hand, it is easy to ensure a consistent discretization of the adjoint equation but the resulting discrete gradient may be inaccurate.

In the ideal case, both approaches commute:

| Continuous OCP | Optimize | Discretize | Optimize | Discrete optimality system |
|----------------|----------|------------|----------|-----------------|

We note that for a given discretization of the state equation, the DO approach is fully specified. If the resulting discretization of the adjoint is not consistent, we cannot expect commutativity. We will therefore focus on a space-time DG discretization, for which the DO approach results in a consistent discretization of the adjoint equation. Therefore, one can also view this contribution as giving a suitable reinterpretation of the discretization of the adjoint equation in the DO approach.

For commutativity, the OD approach should coincide with this formulation. In this work, we focus on using a conservative discretization of the convection term for both the state and adjoint equation in order to be able to use the same flow solver for the state and the adjoint equation. We will see that one needs to be very careful when discretizing the arising source term $-f'(u)xz$ in the adjoint equation in order to ensure commutativity.

2.2. Discretization of the state equation. We discretize the state equation using a space-time DG approach. Using DG in time is a very natural approach in the context of optimal control as one can easily compute forwards and backwards in time. Further, we avoid strict time stepping conditions due to the presence of the diffusion term, which explicit time stepping schemes require.

We consider a space-time mesh with each space-time element being a tensor product of a spatial cell $K_i = (x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}) \subset \Omega = (x_L, x_R)$ and a time step $I^n = (t^n, t^{n+1}) \subset (0, T)$. For simplicity, we consider $N$ equidistant spatial cells with length $h$, and denote by $N_t$ the number of time steps with $t^0 = 0$ and $t^{N_t} = T$.

The discrete solution is then sought in the space

$$\mathcal{V}^p = \left\{ u^h \in L^2(Q) : u^h|_{K_i \times I^n} = \sum_{l=0}^{p_t} \sum_{m=0}^{p_x} c_{lm} t^l x^m \text{ with } p_t + p_x \leq p \right\}.$$

We will use the following notation for discrete functions $v^h \in \mathcal{V}^p$: indices $+/-$ and $R/L$ denote the following limits (with $\delta > 0$) in time and space

$$v^{n,\pm}(x) = \lim_{\delta \to 0} v^h(x, t^n \pm \delta), \quad v_{i+\frac{1}{2}, R/L}(t) = \lim_{\delta \to 0} v^h(x_{i+\frac{1}{2}} \pm \delta, t).$$

Remark 1. Due to the space-time approach, we always discretize in space and time simultaneously. We typically mark discrete functions by the superscript ‘h’. We choose to use the same polynomial order in time and space to match the expected orders of convergence in time and space. Note also that we do not use a full tensor product approach (Q-space) but consider the reduced P-space. This results in the same orders of convergence but with a reduced number of degrees of freedom.
Generally, one can use different polynomial orders \( p \) and \( q \) in space and time. This is more typical for a full tensor approach, compare, e.g., [8]. All the commutativity results discussed in the following extend to this setting.

**Definition 2.1.** We define the average and the jump for interior edges (w.r.t. space) by

\[
\{\{v^h\}\}_{i+\frac{1}{2}} = \frac{1}{2} \left( v_{i+\frac{1}{2},R} + v_{i+\frac{1}{2},L} \right), \quad [v^h]_{i+\frac{1}{2}} = v_{i+\frac{1}{2},L} - v_{i+\frac{1}{2},R}.
\]

We extend the definition to boundary edges \( x_\frac{1}{2} = x_L \) and \( x_{N+\frac{1}{2}} = x_R \) by

\[
\{\{v^h\}\}_{\frac{1}{2}} = v_{\frac{1}{2},R}, \quad \{\{v^h\}\}_{N+\frac{1}{2}} = v_{N+\frac{1}{2},L}, \quad [v^h]_{\frac{1}{2}} = -v_{\frac{1}{2},R}, \quad [v^h]_{N+\frac{1}{2}} = v_{N+\frac{1}{2},L}.
\]

We further define an additional average by \( \{\{v^h\}\}^* = \{\{v^h\}\} \) for interior edges and

\[
\{\{v^h\}\}^*_{\frac{1}{2}} = \frac{1}{2} v_{\frac{1}{2},R}, \quad \{\{v^h\}\}^*_{N+\frac{1}{2}} = \frac{1}{2} v_{N+\frac{1}{2},L}.
\]

For the discretization of the state equation (2), we use a variant of the space-time DG scheme used in [16, 23]: find \( u^h \in \mathcal{V}^p \) such that for all \( \varphi^h \in \mathcal{V}^p \)

\[
\mathcal{B}_{\text{prim}}(u^h, \varphi^h) + \mathcal{C}_{\text{prim}}(u^h, \varphi^h) + \mathcal{A}(u^h, \varphi^h) = \mathcal{S}^q(q^h, \varphi^h) + \mathcal{S}^g(g, \varphi^h) + l(u^0, \varphi^h). \tag{5}
\]

Here, \( \mathcal{B}_{\text{prim}} + \mathcal{C}_{\text{prim}} \) corresponds to a space-time DG discretization of the scalar conservation law \( u_t + f(u)_x \), which we will describe in more detail below, and \( l \) denotes the implementation of the initial data \( u^0 \). The discretization of the diffusion term is given by \( \mathcal{A} \). Finally, \( \mathcal{S}^q + \mathcal{S}^g \) are discretizations of the source terms on the right hand side and are simply given by

\[
\mathcal{S}^q(q^h, \varphi^h) = \sum_{n,i} \int_{I^n} \int_{K_i} q^h \varphi^h dx \, dt \quad \text{and} \quad \mathcal{S}^g(g, \varphi^h) = \sum_{n,i} \int_{I^n} \int_{K_i} g \varphi^h dx \, dt.
\]

For the discretization of the diffusion term we employ the SIPG discretization [3, 8, 23] given by

\[
\mathcal{A}(u^h, \varphi^h) = \sum_{n,i} \int_{I^n} \int_{K_i} \varepsilon u^h_{x_i} \varphi^h_{x_i} dx \, dt - \sum_{n,i} \int_{I^n} \varepsilon \{\{u^h_{x_i}\}\}_{i+\frac{1}{2}} [\varphi^h]_{i+\frac{1}{2}} \, dt
\]

\[
- \sum_{n,i} \int_{I^n} \varepsilon [u^h]_{i+\frac{1}{2}} \{\{\varphi^h\}\}_{i+\frac{1}{2}} \, dt + \sum_{n,i} \int_{I^n} \sigma [u^h]_{i+\frac{1}{2}} [\varphi^h]_{i+\frac{1}{2}} \, dt.
\]

Here, \( \sigma > 0 \) denotes a stability parameter, which must be chosen sufficiently large to guarantee stability [8]. We note that we use a variant, which uses the diffusion coefficient \( \varepsilon \) in each term. Therefore, the coefficient \( \sigma \) is independent of \( \varepsilon \). It mainly depends on the constant in the inverse trace estimate and roughly scales with \( O(p^2) \).

Finally, \( \mathcal{B}_{\text{prim}} \) and \( \mathcal{C}_{\text{prim}} \) are given by

\[
\mathcal{B}(u^h, \varphi^h) = -\sum_{n,i} \int_{I^n} \int_{K_i} u^h \varphi^h_{x_i} dx \, dt + \sum_{n,i} \int_{K_i} \{U(u^{n+1,-}, u^{n+1,+}) \varphi^{n+1,-} - U(u^{n,-}, u^{n,+}) \varphi^{n,+}\} \, dx \tag{6}
\]
and
\[ C_{\text{prim}}(u^h, \varphi^h) = -\sum_{n,i} \int_{I^n} \int_{K_i} f(u^h) \varphi^h_x \, dx \, dt + \sum_{n,i} \int_{I^n} \left\{ \mathcal{F}(u_{i+\frac{1}{2}}, L, u_{i+\frac{1}{2}}, R) \varphi_{i+\frac{1}{2}, L} - \mathcal{F}(u_{i-\frac{1}{2}}, L, u_{i-\frac{1}{2}}, R) \varphi_{i-\frac{1}{2}, R} \right\} \, dt. \]

Here, $U$ and $\mathcal{F}$ denote the fluxes in time and space. In order to enable proper time marching, we use the upwind flux in time, i.e.,
\[ U(u_{n+1}, u_{n+1/2}) = u_{n+1}, \]
then, $B_{\text{prim}}$ can be derived from $B$ as
\[ B_{\text{prim}}(u^h, \varphi^h) = -\sum_{n,i} \int_{I^n} \int_{K_i} u^h \varphi^h_t \, dx \, dt + \sum_{n,i} \int_{K_i} \left( u_{n+1, -} \varphi_{n+1, -} - u_{n, -} \varphi_{n, -} \right) \, dx \]
with $u^{0, -}$ being defined as 0, the sum over $n$ going from 0 to $N_t - 1$, and the implementation of the initial data being given by
\[ l(u^0, \varphi^h) = \sum_i \int_{K_i} u^0 \varphi^{0, +} \, dx. \]

The flux in space $\mathcal{F}$ will be specified in the next section. At the physical boundaries of the domain we set $u_{\frac{1}{2}, L} = 0$ and $u_{N + \frac{1}{2}, R} = 0$ to respect the homogeneous Dirichlet boundary conditions.

We note that all of $B_{\text{prim}}$, $A$, and $S^q$ are linear in $u^h$ and $q^h$, respectively, whereas $C_{\text{prim}}$ is non-linear in $u^h$. Further, all terms are linear in the test function $\varphi^h$.

2.3. Choice of numerical flux $\mathcal{F}$. There exists a large literature concerning numerical fluxes, see [18, 21] for a small selection. In the following we will focus on the following two fluxes:

- Global Lax-Friedrichs flux (LF): The flux is given by
  \[ \mathcal{F}(a, b) := \frac{1}{2} (f(a) + f(b)) - \frac{\alpha_g}{2} (b - a) \]
  with a parameter $\alpha_g$ to be discussed below. For $\alpha_g$ independent of $a$ and $b$, the derivatives are given by
  \[ \mathcal{F}'_a(a, b) = \frac{1}{2} f'(a) + \frac{\alpha_g}{2} \quad \text{and} \quad \mathcal{F}'_b(a, b) = \frac{1}{2} f'(b) - \frac{\alpha_g}{2}. \]

- Engquist-Osher flux (EO): For Burgers equation, the EO flux is given by, compare [10, 18],
  \[ \int_0^a \max(f'(u), 0) \, du + \int_0^b \min(f'(u), 0) \, du. \]

The derivatives are given by
\[ \mathcal{F}'_a(a, b) = \max(f'(a), 0) \quad \text{and} \quad \mathcal{F}'_b(a, b) = \min(f'(b), 0). \]

The LF flux is probably the most widespread flux due to its good stability properties and its simple form. The EO flux can be interpreted as an extension of the upwind flux to more general conservation laws. Often upwind fluxes lead to more accurate results than the fairly diffusive LF flux, especially for finite volume or lower order DG schemes.
There are different options for choosing the parameter $\alpha_g$ for the LF flux, for example:

- **LF1:**
  \[ \alpha_g := \max_{x \in \Omega} \{|f'(u^0(x))|\}, \]

- **LF2:**
  \[ \alpha_g := \max_{x \in \Omega} \{|f'(u^h(x,t^n))|\}, \]

- **LF3:**
  \[ \alpha_g := \frac{w \Delta x}{\Delta t}, \quad w = \frac{1}{2p+1}. \]

LF1 and LF2 can be found in standard text books, e.g., LF1 in [6] and LF2 in [15], and LF3 has been suggested in [26]. The difference between LF1 and LF2 is whether to evaluate the constant $\alpha_g$ based on the initial data or whether to reevaluate it in each time step. In the context of finite volume schemes, the LF flux is typically defined using LF3 with $p = 0$.

For explicit time stepping one needs be careful in choosing $\alpha_g$ to guarantee stability. This is especially the case as the source term $g + q$ on the right hand side of (5) can greatly influence the discrete solution by, e.g., creating new extrema. Thanks to the stability properties of using DG in time though, this is not an issue for us. We therefore use the choice that is most convenient in terms of differentiability, and choose LF1 (which we simply call LF from now on) as then $\alpha_g$ only depends on the given data.

**Remark 2.** Other authors have already examined the effect of the choice of the numerical fluxes on optimal control problems. For example, the authors of [2, 19] examine the long-time asymptotic behavior of (a different version of) LF and EO in the context of a finite volume scheme for an OCP with Burgers equation, with the control being given by the initial data, and conclude that LF is too diffusive while EO works well. In our tests we did not observe fundamentally different behaviors for LF and EO. This could be attributed to the observation that often, for higher polynomial degree $p$, the influence of the numerical flux is not as dominant. Our main goal here is to show that our approach can be applied to a variety of fluxes.

This concludes the description of the discretization of the state equation. Next, we discuss the discretization of the adjoint equation. To do so, we first summarize commutativity results from the literature for the time derivative and the diffusion term. This is done in section 3. Afterwards we focus on the discretization of the convection term in the adjoint equation. We will first consider the DO approach in section 4 and afterwards the OD approach in section 5.

**Remark 3.** In [29], the authors examine discretely exact derivatives for DG schemes in the context of linear hyperbolic equations. To avoid quadrature errors, the authors recommend to use a strong DG formulation for the discretization of the adjoint equation if a weak discretization for the state equation is used. We will use weak DG discretizations for both equations and make up by using sufficiently accurate quadrature rules. Our focus is on how to properly discretize a non-conservative equation in a conservative form.

3. **Commutativity of the linear terms of the PDE discretization.** We discuss the commutativity of the linear forms $\mathcal{B}_{\text{prim}}$ and $\mathcal{A}$ in the discretization of the state equation (5). We start with the bilinear form $\mathcal{A}$, the discretization of the diffusion term. As discussed by Leykekhman [22] in the context of a steady advection...
diffusion equation, using the SIPG discretization, results in commutativity with respect to the diffusion term. This is not the case if, for example, the non-symmetric interior penalty (NIPG) discretization was chosen.

Using DG in time with the upwind flux is structurally the same as a DG discretization in space of a constant advection term with upwind flux. Therefore, based on the results by Leykekhman [22] as well as based on the results of Akman and Karasözen [1], we expect the time discretization to commute. For completeness, we briefly recap the computation here. To do so, we introduce the bilinear form $\mathcal{B}_{\text{adj}}$ given by

$$
\mathcal{B}_{\text{adj}}(\varphi^h, z^h) = \sum_{n,i} \int_{I^n} \int_{K_i} z^h \varphi^h_t \, dx \, dt + \sum_{n,i} \int_{K_i} \left( z^{n,+} \varphi^{n,+} - z^{n+1,+} \varphi^{n+1,-} \right) \, dx
$$

with $z^{N_t,+}$ defined to be 0. This form is derived from (6) but respects the fact that in the adjoint equation the term $-z_t$ needs to be discretized and that therefore the upwind flux goes in the opposite direction. Due to the homogeneous initial conditions for $z$, there is no additional term of the form $l$, compare (7).

Using integration by parts in time on each time slat, we get (dropping the sum over the spatial cells for better readability)

$$
\mathcal{B}_{\text{prim}}(u^h, z^h)_i = -\sum_{n=0}^{N_t-1} \int_{I^n} \int_{K_i} u^h z^h \, dx \, dt + \sum_{n=0}^{N_t-1} \int_{K_i} \left( u^{n+1,-} z^{n+1,-} - u^{n,-} z^{n,+} \right) \, dx
$$

$$
= \sum_{n=0}^{N_t-1} \int_{I^n} \int_{K_i} u^h z^h \, dx \, dt - \sum_{n=0}^{N_t-1} \int_{K_i} \left( u^{n+1,-} z^{n+1,-} - u^{n,+} z^{n,+} \right) \, dx
$$

$$
+ \sum_{n=0}^{N_t-1} \int_{K_i} \left( u^{n+1,-} z^{n+1,-} - u^{n,-} z^{n,+} \right) \, dx
$$

$$
= \sum_{n=0}^{N_t-1} \int_{I^n} \int_{K_i} u^h z^h \, dx \, dt + \sum_{n=0}^{N_t-1} \int_{K_i} \left( u^{n,+} z^{n,+} - u^{n,-} z^{n,-} \right) \, dx
$$

$$
= \sum_{n=0}^{N_t-1} \int_{I^n} \int_{K_i} u^h z^h \, dx \, dt + \sum_{n=0}^{N_t-1} \int_{K_i} \left( u^{n,+} z^{n,+} - u^{n+1,-} z^{n+1,+} \right) \, dx,
$$

which coincides with $\mathcal{B}_{\text{adj}}(u^h, z^h)_i$, and therefore confirms the commutativity properties of the time discretization.

Therefore, the discretizations of the linear terms $u_t$ and $-\varepsilon u_{xx}$, resulting in $-z_t$ and $-\varepsilon z_{xx}$ in the adjoint equation, satisfy the requirements for commutativity. So does the linear source term $\mathcal{S}^Q$. Next, we will examine the discretization of $\mathcal{C}_{\text{prim}}$ in more detail, which involves the non-linear term $f(u)_x$. For that purpose, we will examine the DO and the OD approach more formally.
4. **DO.** In this approach we first discretize and then optimize. Therefore, we consider the discrete optimal control problem

\[
\min_{(q^h, u^h) \in \mathbb{V}^p \times \mathbb{V}^p} J(q^h, u^h) = \frac{1}{2} \| u^h - u_d \|^2_{L^2(Q)} + \frac{\alpha}{2} \| q^h \|^2_{L^2(Q)} \tag{10}
\]

s.t. \( B_{\text{prim}}(u^h, \varphi^h) + C_{\text{prim}}(u^h, \varphi^h) + A(u^h, \varphi^h) = S^q(q^h, \varphi^h) + S^g(g, \varphi^h) + l(u^0, \varphi^h) \quad \forall \varphi^h \in \mathbb{V}^p. \)

The formal discrete Lagrangian is then given by

\[
\mathcal{L}_h(q^h, u^h, z^h) = \frac{1}{2} \| u^h - u_d \|^2_{L^2(Q)} + \frac{\alpha}{2} \| q^h \|^2_{L^2(Q)} - B_{\text{prim}}(u^h, z^h) - C_{\text{prim}}(u^h, z^h) + S^q(q^h, z^h) + S^g(g, z^h) + l(u^0, z^h).
\]

This results in the following discrete optimality system: find \((q^h, u^h, z^h) \in \mathbb{V}^p \times \mathbb{V}^p \times \mathbb{V}^p\) such that for \((\varphi^h, \psi^h, \rho^h) \in \mathbb{V}^p \times \mathbb{V}^p \times \mathbb{V}^p\)

\[
\begin{align*}
B_{\text{prim}}(u^h, \varphi^h) + C_{\text{prim}}(u^h, \varphi^h) + A(u^h, \varphi^h) &= S^q(q^h, \varphi^h) + S^g(g, \varphi^h) + l(u^0, \varphi^h) \\
&= \Pi_h u^h, \tag{12}
\end{align*}
\]

\[
B_{\text{adj}}(\psi^h, z^h) + C_{\text{adj}, \text{DO}}(u^h; \psi^h, z^h) + A(\psi^h, z^h) = S^u(u^h - u_d, \psi^h) \\
z_{N_t^+, t^+} = 0, \\
(\alpha q^h + z^h, \rho^h) = 0,
\]

with \(C_{\text{adj}, \text{DO}}\) still to be specified and

\[
S^u(u^h - u_d, \psi^h) = \sum_{n,i} \int_{I_n} \int_{K_i} (u^h - u_d) \psi^h \, dx \, dt, \quad (\psi^h, \rho^h) = \sum_{n,i} \int_{I_n} \int_{K_i} \psi^h \rho^h \, dx \, dt.
\]

Further, \(\Pi_h\) denotes the \(L^2\) projection operator.

We derive \(C_{\text{adj}, \text{DO}}\) from \(C_{\text{prim}}\) formally by considering the directional derivative of \(C_{\text{prim}}(u^h, z^h)\) with respect to \(u^h\) in direction \(\varphi^h\) given by

\[
C_{\text{prim}, u^h}(u^h, z^h)(\varphi^h) = -\sum_{n,i} \int_{I_n} \int_{K_i} f'(u^h) \varphi^h z_i^h \, dx \, dt
\]

\[
+ \sum_{n,i} \int_{I_n} \left\{ F_{u^h}(u_{i+\frac{1}{2}, L}, u_{i+\frac{1}{2}, R})(\varphi^h) z_{i+\frac{1}{2}, L} - F_{u^h}(u_{i-\frac{1}{2}, L}, u_{i-\frac{1}{2}, R})(\varphi^h) z_{i-\frac{1}{2}, R} \right\} \, dt.
\]

We apply integration by parts in space to get

\[
C_{\text{prim}, u^h}(u^h, z^h)(\varphi^h)
= \sum_{n,i} \int_{I_n} \int_{K_i} f'(u^h) \varphi^h z_i^h \, dx \, dt
+ \sum_{n,i} \int_{I_n} \int_{K_i} f'(u^h) \varphi^h z_i^h \, dx \, dt
\]

\[
- \sum_{n,i} \int_{I_n} \left\{ (f'(u^h) \varphi^h z_i^h)_{i+\frac{1}{2}, L} - (f'(u^h) \varphi^h z_i^h)_{i-\frac{1}{2}, R} \right\} \, dt \tag{13}
\]

\[
+ \sum_{n,i} \int_{I_n} \left\{ F_{u^h}(u_{i+\frac{1}{2}, L}, u_{i+\frac{1}{2}, R})(\varphi^h) z_{i+\frac{1}{2}, L} - F_{u^h}(u_{i-\frac{1}{2}, L}, u_{i-\frac{1}{2}, R})(\varphi^h) z_{i-\frac{1}{2}, R} \right\} \, dt.
\]

We will now discuss the derivatives for our two numerical fluxes \(F\) separately.
4.1. LF flux. From (8) we get for the directional derivative of $F(u_L, u_R)$ with respect to $u^h$ in direction $\varphi^h$

$$F_{u^h}(u_L, u_R)(\varphi^h) = \left\langle \{ f'(u^h) \varphi^h \} \right\rangle^* + \frac{\alpha_k}{2} \left\| \varphi^h \right\| .$$

Using this in (13) results in

$$C_{\text{prim}, u^h, z^h}(\varphi^h)$$

$$= \sum_{n,i} \int_{I^n} \int_{K_i} f'(u^h)_{x^h} z^h \, dx \, dt + \sum_{n,i} \int_{I^n} \int_{K_i} f'(u^h) \varphi^h_{x^h} z^h \, dx \, dt$$

$$- \sum_{n,i} \int_{I^n} \left\{ (f'(u^h) \varphi^h)_{i+\frac{1}{2}, L} - (f'(u^h) \varphi^h_{z^h})_{i-\frac{1}{2}, R} \right\} \, dt$$

$$+ \sum_{n,i} \int_{I^n} \left\{ \left\langle \{ f'(u^h) \varphi^h \} \right\rangle^*_{i+\frac{1}{2}} + \frac{\alpha_k}{2} \left\| \varphi^h \right\|_{i+\frac{1}{2}} \right\} z_{i+\frac{1}{2}, L}$$

$$- \left\langle \{ f'(u^h) \varphi^h \} \right\rangle^*_{i-\frac{1}{2}} + \frac{\alpha_k}{2} \left\| \varphi^h \right\|_{i-\frac{1}{2}} \right\} z_{i-\frac{1}{2}, R} \right\} \, dt.$$

We collect all terms that are associated with test functions $\varphi_{i+\frac{1}{2}, L}$ and $\varphi_{i-\frac{1}{2}, L}$ and get for interior edges:

$$\varphi_{i+\frac{1}{2}, L} : - f'(u_{i+\frac{1}{2}, L}) z_{i+\frac{1}{2}, L} + \frac{1}{2} f'(u_{i+\frac{1}{2}, L}) z_{i+\frac{1}{2}, L} + \frac{\alpha_k}{2} z_{i+\frac{1}{2}, L}$$

$$= - \frac{1}{2} f'(u_{i+\frac{1}{2}, L}) z_{i+\frac{1}{2}, L} + \frac{\alpha_k}{2} z_{i+\frac{1}{2}, L},$$

$$\varphi_{i-\frac{1}{2}, L} : - f'(u_{i-\frac{1}{2}, L}) z_{i-\frac{1}{2}, L} - \frac{\alpha_k}{2} z_{i-\frac{1}{2}, L}.$$

Reordering the sum to identify $\varphi_{i-\frac{1}{2}, L}$ with $\varphi_{i+\frac{1}{2}, L}$ while respecting the homogeneous Dirichlet boundary conditions results in

$$- f'(u_{i+\frac{1}{2}, L}) \left\langle \{ z \} \right\rangle^*_{i+\frac{1}{2}} + \frac{\alpha_k}{2} \left\| z \right\|_{i+\frac{1}{2}}, \quad i = 1, \ldots, N.$$

Similarly, we get for terms that are associated with $\varphi_{i+\frac{1}{2}, R}$ and $\varphi_{i-\frac{1}{2}, R}$, respectively, (preferring the index $i - \frac{1}{2}$)

$$f'(u_{i-\frac{1}{2}, R}) \left\langle \{ z \} \right\rangle^*_{i-\frac{1}{2}} - \frac{\alpha_k}{2} \left\| z \right\|_{i-\frac{1}{2}}, \quad i = 1, \ldots, N.$$

This results in the following definition of the convection term in the discretization of the adjoint equation for the DO approach:

$$C_{\text{adj}, DO}(u^h; \varphi^h, z^h)$$

$$= \sum_{n,i} \int_{I^n} \int_{K_i} f'(u^h)_{x^h} z^h \, dx \, dt + \sum_{n,i} \int_{I^n} \int_{K_i} f'(u^h) \varphi^h_{x^h} z^h \, dx \, dt$$

$$+ \sum_{n,i} \int_{I^n} \left\{ \left(- f'(u_{i+\frac{1}{2}, L}) \left\langle \{ z \} \right\rangle^*_{i+\frac{1}{2}} + \frac{\alpha_k}{2} \left\| z \right\|_{i+\frac{1}{2}} \right) \varphi_{i+\frac{1}{2}, L} \right. \right. \right.$$  \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{1cm} \hspace{
4.2. EO flux. Next, we consider the EO flux. Using the derivatives given by (9) in (13) and reordering the terms with respect to \( \varphi^h \) gives

\[
C_{\text{prim},u^h}(u^h, z^h)(\varphi^h) = \sum_{n,i} \int_{I^n} \int_{K_i} f'(u^h)z^h \varphi^h dx \, dt + \sum_{n,i} \int_{I^n} \int_{K_i} f'(u^h)z^h \varphi_x^h dx \, dt
\]

\[
- \sum_{n,i} \int_{I^n} \left\{ (f'(u^h)z^h)_{i+\frac{1}{2},L} - (f'(u^h)z^h)_{i-\frac{1}{2},R} \right\} dt
\]

\[
+ \sum_{n,i} \int_{I^n} \max(f'(u_{i-\frac{1}{2},R}), 0) \left[ z^h \right]_{i-\frac{1}{2}} \varphi_{i-\frac{1}{2},R} \right\} dt.
\]

We further simplify (with \( z_{\frac{1}{2},L} = 0 \) and \( z_{N+\frac{1}{2},R} = 0 \))

\[
\max(f'(u_{i+\frac{1}{2},L}), 0) \left[ z^h \right]_{i+\frac{1}{2}} - f'(u_{i+\frac{1}{2},L})z_{i+\frac{1}{2},L}
\]

\[
= - \min(f'(u_{i-\frac{1}{2},R}), 0)z_{i-\frac{1}{2},L} - \max(f'(u_{i+\frac{1}{2},L}), 0)z_{i+\frac{1}{2},R}
\]

and

\[
\min(f'(u_{i-\frac{1}{2},R}), 0) \left[ z^h \right]_{i-\frac{1}{2}} + f'(u_{i-\frac{1}{2},R})z_{i-\frac{1}{2},R}
\]

\[
= \min(f'(u_{i-\frac{1}{2},R}), 0)z_{i-\frac{1}{2},L} + \max(f'(u_{i-\frac{1}{2},R}), 0)z_{i-\frac{1}{2},R}.
\]

This implies

\[
C_{\text{adj,DO}}(u^h; \varphi^h, z^h)
\]

\[
= \sum_{n,i} \int_{I^n} \int_{K_i} f'(u^h)x^h \varphi^h dx \, dt + \sum_{n,i} \int_{I^n} \int_{K_i} f'(u^h)x^h \varphi_x^h dx \, dt
\]

(15)

\[
+ \sum_{n,i} \int_{I^n} \left\{ - \min(f'(u_{i+\frac{1}{2},L}), 0)z_{i+\frac{1}{2},L} - \max(f'(u_{i+\frac{1}{2},L}), 0)z_{i+\frac{1}{2},R} \right\} \varphi_{i+\frac{1}{2},L}
\]

\[
+ \left\{ \min(f'(u_{i-\frac{1}{2},R}), 0)z_{i-\frac{1}{2},L} + \max(f'(u_{i-\frac{1}{2},R}), 0)z_{i-\frac{1}{2},R} \right\} \varphi_{i-\frac{1}{2},R} \right\} dt.
\]

with \( z_{\frac{1}{2},L} = 0 \) and \( z_{N+\frac{1}{2},R} = 0 \).

5. OD. In this approach, we first set up the continuous optimality system. Then we are free to discretize each equation separately.

The state equation is given by: find \( u \in W \) such that it satisfies a weak formulation of

\[
u_t + f(u)_x - \varepsilon u_{xx} = g + q \quad \text{in } Q,
\]

\[
u(x_L, \cdot) = u(x_R, \cdot) = 0 \quad \text{in } (0, T),
\]

\[
u(\cdot, 0) = u^0 \quad \text{in } \Omega.
\]

The discrete state equation is given by (5). The adjoint equation is naturally given in a non-conservative formulation and has the form: find \( z \in W \) such that it satisfies a weak formulation of

\[
-z_t - f'(u)z_x - \varepsilon z_{xx} = u - u_d \quad \text{in } Q,
\]

\[
z(x_L, \cdot) = z(x_R, \cdot) = 0 \quad \text{in } (0, T),
\]

\[
z(\cdot, T) = 0 \quad \text{in } \Omega.
\]
We rewrite it in a conservative formulation so that we can apply the existing space-time DG discretization for conservation laws:
\[- z_t - f'(u)z_x - \varepsilon z_{xx} = - f(u)xz + u - u_d \quad \text{in } Q, \]
\[z(x_L, \cdot) = z(x_R, \cdot) = 0 \quad \text{in } (0, T), \]
\[z(\cdot, T) = 0 \quad \text{in } \Omega. \]

We will discuss suitable discretizations of the adjoint equation in the conservative formulation in the following in detail.

Finally, the continuous optimality condition is given by: find \( q \in L^2(Q) \) such that
\[ \alpha q + z = 0 \quad \text{in } Q. \]

The discretization of this equation is trivial.

5.1. Discretization of the adjoint equation. We discretize (16) using a conservative space-time DG discretization: find \( z^h \in V^p \) such that for all \( \varphi^h \in V^p \)
\[ E_{\text{adj}}(\varphi^h, z^h) + C_{\text{adj,OD}}(u^h; \varphi^h, z^h) + A(\varphi^h, z^h) = S_{\text{OD}}^{\text{cons}}(u^h, \varphi^h) + S(u^h - u_d, \varphi^h), \]
with \( C_{\text{adj,OD}} \) and \( S_{\text{OD}}^{\text{cons}} \) still to be defined. Here, \( S_{\text{OD}}^{\text{cons}} \) refers to a discretization of the new source term \(-f'(u)xz\).

This approach has the advantage that we can define a new flux function \( \tilde{F}(z) = -f'(u)z \) and use a discretization similar to (5), i.e., choose \( \tilde{F} \) to be a matching flux to \( F \). Then,
\[ C_{\text{adj,OD}}(u^h; \varphi^h, z^h) \]
\[ = \sum_{n,i} \int_{I_n} \int_{K_i} f'(u^h)z^h \varphi^h_x \, dx \, dt \]
\[ + \sum_{n,i} \int_{I_n} \left\{ \tilde{F}(u_i^{i+1/2}L, z_i^{i+1/2}L), (u_i^{i+1/2}R, z_i^{i+1/2}R) \right\} \varphi^h_{i^{i+1/2}L} \]
\[ - \tilde{F}(u_i^{i-1/2}L, z_i^{i-1/2}L), (u_i^{i-1/2}R, z_i^{i-1/2}R) \right\} \varphi^h_{i^{i-1/2}R} \right) \, dt. \]

The simplest way to discretize the new source term \(-f'(u)xz\) on the right hand side of (16) is by defining \( S_{\text{OD}}^{\text{cons}} = S_{\text{OD,pre}}^{\text{cons}} \) with
\[ S_{\text{OD,pre}}^{\text{cons}}(u^h; \varphi^h, z^h) = - \sum_{n,i} \int_{I_n} \int_{K_i} f'(u^h)xz^h \varphi^h \, dx \, dt. \]

We will see that this straight-forward approach is not the correct one here.

**Remark 4.** Comparing (18)-(19) with (14) and (15), we observe that for the case of \( \left[ f'(u^h) \right]_{i^{i+1/2}} \neq 0 \) it is not possible to find a flux function \( \tilde{F} \) so that (18)-(19) matches with (14) or (15), as the terms in front of \( \varphi_{i^{i+1/2}L} \) and \( \varphi_{i^{i-1/2}R} \) in (14) and (15), respectively, differ. We will solve this problem by adding suitable edge correction terms to \( S_{\text{OD,pre}}^{\text{cons}} \), resulting in \( S_{\text{OD}}^{\text{cons}} \).

In the context of developing numerical methods for balance laws, it is an open secret that one needs to be very careful in designing the schemes in order to make sure that steady states of the continuous system are preserved by the discretization as good as possible (ideally up to machine precision) and not only up to the accuracy of the underlying scheme. This is, for example, well studied for the shallow water
equations and in particular for the so called lake at rest steady state, see, e.g., [5, 13, 17] and the references cited therein.

We will follow a similar idea here. Let us focus on (4). A stationary solution of this equation (on an infinite domain) is described by

\[(f'(u)z)_z = f'(u)_z z.\]

The constant solution \(z = c = \text{constant}\) solves this equation. We note that it also solves the original non-conservative equation \(-z_t - f'(u)z_x = 0\) on an infinite domain (before we applied integration by parts on a continuous level). We therefore also want our conservative discretization of these terms to imitate this correct behavior, i.e., there should hold

\[C_{\text{adj,OD}}(u^h; \varphi^h, c) - S_{\text{cons,OD}}^\text{pre}(u^h; \varphi^h, c) = 0.\] (20)

We will adjust the discretization of the source term \(S_{\text{cons,OD}}^\text{pre}\) to ensure this behavior locally, with the precise corrections depending on the choice of the numerical fluxes \(F\) and \(\tilde{F}\).

\[\text{Remark 5.}\] Our current focus is on the proper discretization of the convection term, in particular on ensuring (20). We will therefore temporarily consider transmissive boundary conditions for \(z^h\), i.e., set \(z_{\frac{1}{2},L} = z_{\frac{1}{2},R}\) and \(z_{N+\frac{1}{2},R} = z_{N+\frac{1}{2},L}\), instead of the homogeneous Dirichlet boundary conditions, which are not well-posed for a convection equation.

There holds

\[-S_{\text{OD,pre}}^\text{cons}(u^h; \varphi^h, c) + C_{\text{adj,OD}}(u^h; \varphi^h, c) = \sum_{n,i} \int_{I_n} \int_{K_i} f'(u^h)_x c \varphi^h_x \, dx \, dt + \sum_{n,i} \int_{I_n} \int_{K_i} f'(u^h)c \varphi_x^h \, dx \, dt + \sum_{n,i} \int_{I_n} \left\{ \tilde{F}((u_{i+\frac{1}{2},L}, c), (u_{i+\frac{1}{2},R}, c)) \varphi_{i+\frac{1}{2},L} \right. - \left. \tilde{F}((u_{i-\frac{1}{2},L}, c), (u_{i-\frac{1}{2},R}, c)) \varphi_{i-\frac{1}{2},R} \right\} \, dt \right. \]

\[= \sum_{n,i} \int_{I_n} \left\{ (c f'(u^h) \varphi^h)_{i+\frac{1}{2},L} - (c f'(u^h) \varphi^h)_{i-\frac{1}{2},R} \right\} \, dt \right. + \left. \sum_{n,i} \int_{I_n} \left\{ \tilde{F}((u_{i+\frac{1}{2},L}, c), (u_{i+\frac{1}{2},R}, c)) \varphi_{i+\frac{1}{2},L} \right. - \left. \tilde{F}((u_{i-\frac{1}{2},L}, c), (u_{i-\frac{1}{2},R}, c)) \varphi_{i-\frac{1}{2},R} \right\} \, dt.\]

Next, we consider the two fluxes separately.

5.1.1. LF flux. A natural choice for \(\tilde{F}\) in case of the LF flux is given by

\[\tilde{F}((u_L, z_L), (u_R, z_R)) = \langle -f'(u) \rangle^\star \{z\} + \frac{\alpha_g}{2} \|[z]\|.\]

Then,

\[\tilde{F}((u_L, c), (u_R, c)) = \langle -f'(u) \rangle^\star \{c\} + \frac{\alpha_g}{2} \|[c]\] = -c \langle f'(u) \rangle^\star.\]
This gives
\[ -S_{OD,pre}^{\text{cons}}(u^h; \varphi^h, c) + C_{adj,OD}(u^h; \varphi^h, c) \]
\[ = \sum_{n,i} \int_{I_n} \left\{ \left( c f'(u^h) \varphi^h \right)_{i+\frac{1}{2},L} - \left( c f'(u^h) \varphi^h \right)_{i-\frac{1}{2},R} \right\} \, dt \]
\[ + \sum_{n,i} \int_{I_n} \left\{ -c \left\{ f'(u^h) \right\}^*_{i+\frac{1}{2}} \varphi_{i+\frac{1}{2},L} + c \left\{ f'(u^h) \right\}^*_{i-\frac{1}{2}} \varphi_{i-\frac{1}{2},R} \right\} \, dt \]
\[ = \sum_{n,i} \int_{I_n} \left\{ \frac{1}{2} c \left\{ f'(u^h) \right\}^*_{i+\frac{1}{2}} \varphi_{i+\frac{1}{2},L} + \frac{1}{2} c \left\{ f'(u^h) \right\}^*_{i-\frac{1}{2}} \varphi_{i-\frac{1}{2},R} \right\} \, dt. \]

Therefore, at each spatial edge \( i + \frac{1}{2} \), we are left with a term of the form
\[ c \left\{ f'(u^h) \right\}^*_{i+\frac{1}{2}}. \]

In other words: our current discretization does not preserve a standing wave if \( \left\{ f'(u^h) \right\}^*_{i+\frac{1}{2}} \neq 0 \).

Reincorporating the homogeneous Dirichlet boundary conditions for the adjoint state \( z \), we therefore define for the LF flux
\[ S_{OD}^{\text{cons}}(u^h; \varphi^h, z^h) = S_{OD,pre}^{\text{cons}}(u^h; \varphi^h, z^h) + \sum_{n,i} \int_{I_n} \left\{ f'(u^h) \right\}^*_{i+\frac{1}{2}} \left\{ \varphi^h \right\}^*_{i+\frac{1}{2}} \varphi_{i+\frac{1}{2},L} + \frac{1}{2} c \left\{ f'(u^h) \right\}^*_{i-\frac{1}{2}} \varphi_{i-\frac{1}{2},R} \right\} \, dt. \]

(21)

**Remark 6.** Note that for this definition \( C_{adj,OD} - S_{OD}^{\text{cons}} \) coincides with \( C_{adj,DO} \), compare (14), and therefore the OD approach, which uses the source term \( S_{OD}^{\text{cons}} \), commutes with the DO approach.

5.1.2. **EO flux.** A natural flux \( \tilde{F} \) in case of EO, which respects the upwind character of the flux, is given by
\[ \tilde{F}((u_L, z_L), (u_R, z_R)) = \int_0^{z_L} \max(-f'(u_L), 0) \, ds + \int_0^{z_R} \min(-f'(u_R), 0) \, ds \]
\[ = -\min(f'(u_L), 0)z_L - \max(f'(u_R), 0)z_R. \]

We then get
\[ -S_{OD,pre}^{\text{cons}}(u^h; \varphi^h, c) + C_{adj,OD}(u^h; \varphi^h, c) \]
\[ = \sum_{n,i} \int_{I_n} \left\{ \left( c f'(u^h) \varphi^h \right)_{i+\frac{1}{2},L} - \left( c f'(u^h) \varphi^h \right)_{i-\frac{1}{2},R} \right\} \, dt \]
\[ + \sum_{n,i} \int_{I_n} \left\{ -\min(f'(u_{i+\frac{1}{2},L}), 0)c + \max(f'(u_{i+\frac{1}{2},R}), 0)c \right\} \varphi_{i+\frac{1}{2},L} \]
\[ + \left( \min(f'(u_{i-\frac{1}{2},L}), 0)c + \max(f'(u_{i-\frac{1}{2},R}), 0)c \right) \varphi_{i-\frac{1}{2},R} \right\} \, dt. \]

After some manipulations, this gives
\[ -S_{OD,pre}^{\text{cons}}(u^h; \varphi^h, c) + C_{adj,OD}(u^h; \varphi^h, c) \]
\[ = \sum_{n,i} \int_{I_n} \left\{ \left( \max(f'(u_{i+\frac{1}{2},L}), 0)c - \max(f'(u_{i+\frac{1}{2},R}), 0)c \right) \varphi_{i+\frac{1}{2},L} \right. \]
\[ + \left( \min(f'(u_{i-\frac{1}{2},L}), 0)c - \min(f'(u_{i-\frac{1}{2},R}), 0)c \right) \varphi_{i-\frac{1}{2},R} \right\} \, dt. \]
Therefore, again, our discretization does not satisfy the property of preserving a standing wave locally. To ensure this property and to guarantee commutativity with $\mathcal{L}_{\text{adj,DO}}$ given by (15) we define for the EO flux, setting $z_{N+\frac{1}{2},L} = 0$ to respect the homogeneous Dirichlet boundary conditions for $z$,

$$
\mathcal{S}^{\text{cons}}_{\text{OD}}(u^h; \phi^h, z^h) = \mathcal{S}^{\text{cons}}_{\text{OD,pre}}(u^h; \phi^h, z^h) + \sum_{n,i} \left\{ \left( \max(f'(u_{i+\frac{1}{2}, L}), 0) z_{i+\frac{1}{2}, R} - \max(f'(u_{i+\frac{1}{2}, L}), 0) z_{i+\frac{1}{2}, R} \right) \varphi_{i+\frac{1}{2}, L} \right\} dt.
$$

**Remark 7.** Note that for $\|f'(u^h)\| = 0$ the additional edge stabilizations in (22) and (21) drop. This is for example the case for the linear advection equation.

6. **Numerical results.** In this section we present numerical results for solving the OCP (1) and (2). We use the reduced approach and rewrite $J(q^h, u^h)$ as $j_h(q^h) = J(q^h, S_h(q^h))$ with $S_h$ being the discrete solution operator given by (5). Then, for each evaluation of $\nabla j_h(q^h)$ we once need to solve the discrete state equation and the discrete adjoint equation, which is given by (12) for the DO approach and by (17) for the OD approach. We solve the resulting optimization problem in $j_h$ using the L-BFGS algorithm with Armijo line search, see, e.g., Failor [11] for more information.

We focus on examining the effect of adding the additional correction terms in the discretization of the source term $-f'(u)z$. Therefore, we only use the discretization of the adjoint equation that is given by (17) and compare the following two versions:

- **OD-with (OD-w):** In this approach we add the correction terms in $\mathcal{S}^{\text{cons}}_{\text{OD}}$ (given by (21) for the LF flux and by (22) for the EO flux). The resulting discretization then coincides with the discretization of the adjoint equation in the DO approach.
- **OD-without (OD-wo):** In this approach we use $\mathcal{S}^{\text{cons}}_{\text{OD}} = \mathcal{S}^{\text{cons}}_{\text{OD,pre}}$, which results in a discretization, which violates commutativity and does not preserve steady solutions of the convective terms in the adjoint equation.

For the penalty parameter $\sigma$ in the SIPG formulation we choose $\sigma = 3p(p + 1)$. As basis functions $\phi^h \in \mathcal{V}^p$ we use a scaled monomial basis of the form

$$
\varphi_i^n |_{K_i \times I_n} = \left( \frac{t - t^n+1}{\Delta t} \right)^{p_t} \left( \frac{x - x_i}{h} \right)^{p_x}
$$

with $x_i$ denoting the cell centroid of $K_i$.

Before presenting the results, we briefly comment on the expected convergence orders. Numerically, we typically observe convergence orders of $p + 1$ for solving just the state equation (2) with the space-time discretization used in this contribution (if the solution is sufficiently smooth). The results presented in the following will show that these orders extend to the full OCP for the OD-w approach.

Proving these rates is non-trivial. For an error estimate for the state equation, one can rely on [12]. Next, one needs to extend this to an optimal control setting. For the case of the linear advection diffusion equation this is done in [22] for the steady problem and in [1] for the unsteady problem. As we consider a non-linear problem here, it will not be possible to directly follow the argumentation used there. Instead, second-order sufficient optimality conditions need to be derived and exploited appropriately. This will be addressed in future work. For background
information about how to derive error estimates for non-linear OCPs, we refer to [25].

6.1. Test 1: Manufactured solution. We start with a manufactured solution to test convergence properties. On the domain \((x_L, x_R) = (-1, 1)\), we define the solution

\[
\begin{align*}
  u(x, t) &= \sin \left( \frac{\pi}{2} (x + 1) \right), \\
  z(x, t) &= -\sin \left( \frac{\pi}{2} (x + 1) \right) \exp(-\varepsilon(T - t))(T - t), \\
  q(x, t) &= -\alpha^{-1} z,
\end{align*}
\]

which results in the data \(u^0 = u(\cdot, 0)\),

\[
\begin{align*}
  u_d(x, t) &= z(x, t) \left( \varepsilon \left( \frac{\pi^2}{4} - 1 \right) - \frac{\pi}{2} \cos \left( \frac{\pi}{2} (x + 1) \right) \right) + \sin \left( \frac{\pi}{2} (x + 1) \right) \\
  &\quad + \sin \left( \frac{\pi}{2} (x + 1) \right) \exp(-\varepsilon(T - t)), \\
  g(x, t) &= u(x, t) \left( \frac{\pi}{2} \cos \left( \frac{\pi}{2} (x + 1) \right) + \varepsilon \frac{\pi^2}{4} - \alpha^{-1} \exp(-\varepsilon(T - t))(T - t) \right).
\end{align*}
\]

The specific choices of \(\alpha, \varepsilon, \) and \(T\) will be made precise below.

6.1.1. Gradient test. We start with a gradient test to assess the accuracy of the discrete gradient for the two methods OD-w and OD-wo. More precisely, we compare the results for computing the directional derivative \(j_h'(q^h)(\delta q^h)\) in a direction \(\delta q^h\) by means of the adjoint equation with the result of using central difference quotients for different step sizes \(\rho\), i.e., we evaluate

\[
\left| \frac{j_h(q^h + \rho \delta q^h) - j_h(q^h - \rho \delta q^h)}{2\rho} - j_h'(q^h)(\delta q^h) \right|. 
\]

We set \(\varepsilon = 0.01, \alpha = 0.01, T = 1,\) and choose \(q^h\) to be the \(L^2\) projection of the minimum and \(\delta q^h\) to be the \(L^2\) projection of \(10(x + 1)(1 - x)(0.2 - t)\). We choose a coarse discretization with \(h = 0.1\) and time step length \(\Delta t = 0.1\) and vary \(\rho\) between \(10^{-1}\) and \(10^{-9}\).

![Figure 1](image_url)

**Figure 1.** Test 1: Results for gradient test. The \(x\)-axis denotes \(\frac{1}{\rho}\) with \(\rho\) being the step length in the difference quotient, the \(y\)-axis denotes the error of the gradient as given by (24).
The results for LF and EO flux are shown in figure 1. As expected, we obtain an accurate discrete gradient for OD-w. For OD-wo, we observe

- an inaccurate gradient discretization,
- but with increasing accuracy for increasing polynomial degree $p$.

### 6.1.2. Convergence of the OCP

Next, we run convergence tests for the solution given by (23). We choose $T = 0.2$, $\alpha = 0.1$, and $\Delta t = 0.5 \Delta x$. The stopping tolerance in the L-BFGS algorithm is set to be $10^{-8}$. A gradient size bigger than that indicates that the line search broke (due to an inconsistent discrete gradient).

In table 1, we show the results for using the LF flux for $\varepsilon = 10^{-3}$. Besides errors and orders of convergence we present the number of iterations in the L-BFGS algorithm and the gradient norm when the algorithm exited. Roughly speaking we observe convergence behavior of $O(h^{p+1})$ for the state $u$, the adjoint $z$, and the control $q$, except for 2 cases: If the error size is too close to the residual of the gradient (for example for $N = 80$ and $p = 3$) our convergence breaks. This is to be expected. The more interesting case is the case of $p = 1$ for OD-wo. Here, we observe a deterioration in the convergence order for the adjoint $z$, and marginally for the control $q$ as well. This also reflects in larger errors compared to using OD-w.

The same effect is more obvious when we decrease the diffusion parameter. In table 2, we present the results for the LF flux for $\varepsilon = 10^{-5}$. Here, we observe a decay of the convergence order in the adjoint $z$ to first order for $p = 1$ for OD-wo,
Table 2. Test 1: Errors and orders of convergence for $\varepsilon = 10^{-5}$.

| $p$ | $N$ | error $u$ | order | error $q$ | order | error $z$ | order | iter. | gradient |
|-----|-----|----------|-------|----------|-------|----------|-------|-------|----------|
| OD-w |       |          |       |          |       |          |       |       |          |
| 1   | 40  | 2.79e-04 |       | 1.97e-03 |       | 2.47e-04 |       | 24    | 5.33e-05 |
| 1   | 80  | 7.51e-05 | 1.89  | 5.51e-04 | 1.84  | 9.20e-05 | 1.42  | 27    | 2.81e-05 |
| 1   | 160 | 2.28e-05 | 1.72  | 1.93e-04 | 1.51  | 4.08e-05 | 1.18  | 29    | 1.26e-05 |
| 1   | 320 | 8.15e-06 | 1.49  | 7.90e-05 | 1.29  | 1.99e-05 | 1.04  | 33    | 7.13e-06 |
| 2   | 20  | 2.52e-05 |       | 1.72e-04 |       | 1.72e-05 |       | 44    | 2.25e-07 |
| 2   | 40  | 3.98e-06 | 2.06  | 2.18e-05 | 2.98  | 2.18e-06 | 2.97  | 51    | 4.63e-08 |
| 2   | 80  | 6.28e-07 | 2.67  | 3.09e-06 | 2.82  | 3.10e-07 | 2.82  | 54    | 9.51e-09 |
| 2   | 160 | 1.01e-07 | 2.64  | 5.60e-07 | 2.46  | 5.42e-08 | 2.51  | 54    | 9.51e-09 |
| OD-w |       |          |       |          |       |          |       |       |          |
| 1   | 40  | 2.74e-04 |       | 1.89e-03 |       | 1.89e-04 |       | 54    | 9.51e-09 |
| 1   | 80  | 6.99e-05 | 1.97  | 4.68e-04 | 2.01  | 4.68e-05 | 2.01  | 54    | 9.51e-09 |
| 1   | 160 | 1.76e-05 | 2.00  | 1.17e-04 | 2.01  | 1.17e-05 | 2.01  | 54    | 9.51e-09 |
| 1   | 320 | 4.39e-06 | 2.00  | 2.91e-05 | 2.00  | 2.91e-06 | 2.00  | 54    | 9.51e-09 |
| 2   | 20  | 2.46e-05 |       | 1.72e-04 |       | 1.72e-05 |       | 54    | 9.51e-09 |
| 2   | 40  | 3.84e-06 | 2.68  | 2.12e-05 | 3.02  | 2.12e-06 | 3.02  | 54    | 9.51e-09 |
| 2   | 80  | 5.86e-07 | 2.71  | 3.19e-06 | 3.00  | 3.19e-07 | 3.00  | 54    | 9.51e-09 |
| 2   | 160 | 8.82e-08 | 2.73  | 3.89e-07 | 2.77  | 3.89e-08 | 2.77  | 54    | 9.51e-09 |
| 3   | 10  | 4.63e-06 |       | 4.87e-05 |       | 4.87e-06 |       | 54    | 9.51e-09 |
| 3   | 20  | 2.76e-07 | 4.07  | 3.12e-06 | 3.96  | 3.08e-07 | 3.98  | 54    | 9.51e-09 |
| 3   | 40  | 3.58e-08 | 2.94  | 3.19e-07 | 3.29  | 3.19e-08 | 4.04  | 54    | 9.51e-09 |
| 3   | 80  | 2.68e-08 | 0.42  | 2.12e-07 | 0.50  | 2.94e-09 | 2.67  | 54    | 9.51e-09 |

i.e., for not using the correction terms. For OD-w we observe perfect second order convergence. For $p = 2$, the difference between the two discretizations is already very small. This is consistent with the results of the gradient test, which imply that the gradient becomes increasingly better for higher polynomial degree.

The results for the EO flux are qualitatively very similar, and are not shown here for brevity.

6.2. Test 2. Next, we consider a problem from the literature [9, 27], which we modified slightly to see more activity for later time $t$. We set $(x_L, x_R) = (0, 1)$, $T = 1$, $\alpha = 0.5$, and $\varepsilon = 10^{-3}$. Initial conditions $u^0$ are given by

$$u^0(x) = \sin(4\pi x).$$

We further choose $g = 0$ and $u_d = \sin(4\pi x)$. The choice of $u_d$ is the main difference to the test in [9, 27] where $u_d = 0$ is used, which causes the discrete solution $u_h$ to decay to zero very rapidly. We choose the time step size $\Delta t = 0.25\Delta x$, and we initialize the L-BFGS algorithm with $q_h^0 = 0$ and set the stopping tolerance to be $10^{-7}$. We use the EO flux for the results presented in the following. The usage of the LF flux gives very similar results.

When solving the inviscid Burgers equation $u_t + f(u)_x = 0$ (outside the context of an OCP) for the given initial data $u^0$, two shocks form at $x = \frac{1}{4}$ and $x = \frac{3}{4}$ and rarefaction waves form at $x = 0$, $x = \frac{1}{2}$, and $x = 1$. To keep the discrete state
solution $u^h$ close to the desired state $u_d$ (within the OCP context), the control $q^h$ needs to counteract this behavior. We show the solution for the control $q^h$ and the state $u^h$ on a mesh with $p = 2$ and $N = 160$ in figure 2. We observe steep gradients in the control around $x = 0$, $x = \frac{1}{2}$, and $x = 1$, counterbalancing the arising rarefaction waves in $u^h$.

We note that the solution shown in figure 2 is pretty continuous. This is partially due to the fact that we solve the viscous Burgers equation here (with $\varepsilon = 10^{-3}$) and enforce weak continuity through the penalty term in $A$, and combine that with a fairly high resolution. For a strong zoom (not shown here), one can identify small jumps between the polynomial solutions of different cells, especially around the peaks of the discrete state solution $u^h$ for $t = 0.25$.

In table 3 we present results for the functional values $J$ given by (1) for $p = 1$ for both approaches OD-w and OD-wo. Despite the fact that the computation for OD-wo breaks very early due to the inconsistent gradient, the functional values for $J$ are very similar to each other.

To get a better idea of the difference between the schemes, we therefore also compute errors by comparing with a reference solution, which has been produced using $p = 3$ and $N = 640$. In table 4 we show the results for $p = 1$ for the control $q$, both for OD-w and OD-wo. We observe lower convergence orders for OD-wo than OD-w, which results in larger errors. In particular, the error for $N = 160$ is already a factor of 2.4 smaller for OD-w than for OD-wo. This is not a huge factor.
| N  | value J | order | iter. | gradient | value J | order | iter. | gradient |
|----|---------|-------|-------|----------|---------|-------|-------|----------|
| 20 | 0.10640082692 | - | 44 | 8.58e-08 | 0.10641958284 | - | 11 | 2.10e-03 |
| 40 | 0.10818515778 | - | 44 | 7.94e-08 | 0.10819225160 | - | 14 | 1.02e-03 |
| 80 | 0.10855578246 | 2.27 | 43 | 8.63e-08 | 0.10855750188 | 2.28 | 18 | 3.44e-04 |
| 160 | 0.10862035158 | 2.52 | 43 | 9.00e-08 | 0.10862054051 | 2.53 | 22 | 1.09e-04 |

Table 3. Test 2: Values and orders of convergence (computed using values from 3 subsequent meshes) for functional $J$ for $p = 1$.

| N  | error q | order | iter. | gradient | error q | order | iter. | gradient |
|----|---------|-------|-------|----------|---------|-------|-------|----------|
| 20 | 1.67e-02 | - | 44 | 8.58e-08 | 1.77e-02 | - | 11 | 2.10e-03 |
| 40 | 4.56e-03 | 1.87 | 44 | 7.94e-08 | 6.10e-03 | 1.54 | 14 | 1.02e-03 |
| 80 | 1.47e-03 | 1.64 | 43 | 8.63e-08 | 2.83e-03 | 1.11 | 18 | 3.44e-04 |
| 160 | 4.20e-04 | 1.80 | 43 | 9.00e-08 | 1.00e-03 | 1.50 | 22 | 1.09e-04 |

Table 4. Test 2: Errors and orders of convergence for the control $q$ for $p = 1$.

but significant given that the algorithms are identical except for the addition of the correction terms.

7. Conclusion and outlook. We have presented space-time DG discretizations for an optimal control problem for the viscous Burgers equation, for which the approaches of DO and OD commute. We employ conservative discretizations for the convection terms for both the state and the adjoint equation. Following the ideas of well-balanced discretizations of balance laws, we introduce additional edge terms in the discretization of the new source term $-f'(u)z$ in the adjoint equation. This ensures that steady state solutions of the convection terms in the adjoint equation are preserved and that the approaches of DO and OD coincide. Our numerical experiments confirmed reduced convergence rates and reduced gradient accuracy if these additional edge terms are not included for different settings.

In the next step, we proceed to considering the compressible Navier-Stokes equations based on using the space-time DG discretizations suggested in [16, 23]. Due to the symmetry properties introduced by the usage of entropy variables as degrees of freedom (instead of conserved variables), we expect the discretizations of the time derivative term and of the diffusion term to possess the commutative properties. Therefore, similar to the work done here, the challenge will consist in finding appropriate discretizations for the convection terms, for which we will build on insights developed here.

REFERENCES

[1] T. Akman and B. Karasözen, Variational time discretization methods for optimal control problems governed by diffusion-convection-reaction equations, Comput. Appl. Math., 272 (2014), 41–56.
[2] N. Allahverdi, A. Pozo and E. Zuazua, Numerical aspects of large-time optimal control of Burgers equation, ESAIM Math. Model. Numer. Anal., 50 (2016), 1371–1401.
[3] D. N. Arnold, An interior penalty finite element method with discontinuous elements, SIAM J. Numer. Anal., 19 (1982), 742–760.
[4] G. Chen and S. Collis, Optimal control for Burgers flow using the discontinuous Galerkin method, *AIAA Region IV Student Paper Conference*, 2003.

[5] A. Chertock, A. Kurganov and T. Wu, Operator splitting based central-upwind schemes for the shallow water equations with moving bottom topography, *Commun. Math. Sci.*, 18 (2020), 2149–2168.

[6] B. Cockburn, *An introduction to the discontinuous Galerkin Method for convection-dominated problems*, Advanced Numerical Approximation of Nonlinear Hyperbolic Equations (Cetraro, 1997), Lecture Notes in Math., 1697 (1998), 151–268.

[7] J. C. de los Reyes and K. Kunisch, A comparison of algorithms for control constrained optimal control of the Burgers equation, *Calcolo*, 41 (2004), 203–225.

[8] V. Dolejší and M. Feistauer, *Discontinuous Galerkin Methods*, Analysis and Applications to Compressible Flow. Springer Series in Computational Mathematics, 48. Springer, Cham, 2015.

[9] K. T. Elgindy and B. Karasözen, Distributed optimal control of viscous Burgers’ equation via a high-order, linearization, integral, nodal discontinuous Gegenbauer-Galerkin method, *Optimal Control Appl. Methods*, 41 (2020), 253–277.

[10] B. Engquist and S. Osher, *One-sided difference approximations for nonlinear conservation laws*, *Math. Comp.*, 36 (1981), 321–351.

[11] L. Failer, *Optimal Control of Time-Dependent Nonlinear Fluid-Structure Interaction*, PhD thesis, Technical University of Munich, Munich, 2017.

[12] M. Feistauer, V. Kučera, K. Najzar and J. Prokopová, Analysis of space-time discontinuous Galerkin method for nonlinear convection-diffusion problems, *Num. Math.*, 117 (2011), 251–288.

[13] G. Gassner, A. Winters and D. Kopriva, A well balanced and entropy conservative discontinuous Galerkin spectral element method for the shallow water equations, *Appl. Math. Comput.*, 272 (2016), 291–308.

[14] J. Greenberg and A. Leroux, A well-balanced scheme for the numerical processing of source terms in hyperbolic equations, *SIAM J. Numer. Anal.*, 33 (1996), 1–16.

[15] J. Hesthaven and T. Warburton, *Nodal Discontinuous Galerkin Methods: Algorithms, Analysis, and Applications*, Texts in Applied Mathematics, 54. Springer, New York, 2008.

[16] A. Hiltebrand and S. May, Entropy stable spacetime discontinuous Galerkin methods for the two-dimensional compressible Navier-Stokes equations, *Commun. Math. Sci.*, 16 (2018), 2095–2124.

[17] A. Hiltebrand and S. Mishra, Entropy stability and well-balancedness of space-time DG for the shallow water equations with bottom topography, *Netw. Heterog. Media*, 11 (2016), 145–162.

[18] H. Holden and N. H. Risebro, *Front Tracking for Hyperbolic Conservation Laws*, 2nd edition, Applied Mathematical Sciences, 152. Springer, Heidelberg, 2015.

[19] L. I. Ignat, A. Pozo and E. Zuazua, Large-time asymptotics, vanishing viscosity and numerics for 1-D scalar conservation laws, *Math. Comp.*, 84 (2015), 1633–1662.

[20] K. Kunisch and S. Volkwein, Control of the Burgers equation by a reduced-order approach using proper orthogonal decomposition, *J. Optim. Theory Appl.*, 102 (1999), 345–371.

[21] R. J. LeVeque, *Finite Volume Methods for Hyperbolic Problems*, Cambridge University Press, 2002.

[22] D. Leykekhman, Investigation of commutative properties of discontinuous Galerkin methods in PDE constrained optimal control problems, *J. Sci. Comput.*, 53 (2012), 483–511.

[23] S. May, Spacetime discontinuous Galerkin methods for solving convection-diffusion systems, *ESAIM Math. Model. Numer. Anal.*, 51 (2017), 1755–1781.

[24] S. Noelle, Y. Xing and C.-W. Shu, High-order well-balanced schemes, *Numerical methods for Balance Laws, Quad. Mat., Dept. Math., Seconda Univ. Napoli, Caserta*, 24 (2009), 1–66.

[25] J. Pfefferer, *Numerical Analysis for Elliptic Neumann Boundary Control Problems on Polygonal Domains*, PhD thesis, Bundeswehr University Munich, Munich, 2014.

[26] W. J. Rider and R. B. Lowrie, The use of classical Lax-Friedrichs Riemann solvers with discontinuous Galerkin methods, *Internat. J. Numer. Methods Fluids*, 40 (2002), 479–486.

[27] S. Volkwein, *Mesh-Independence of an Augmented Lagrangian-SQP Method in Hilbert Spaces and Control Problems for the Burgers Equation*, PhD thesis, Technical University of Berlin, Berlin, 1997.

[28] S. Volkwein, Distributed control problems for the Burgers equation, *Comput. Optim. Appl.*, 18 (2001), 115–140.
[29] L. Wilcox, G. Stadler, T. Bui-Thanh and O. Ghattas, Discretely exact derivatives for hyperbolic PDE-constrained optimization problems discretized by the discontinuous Galerkin method, *J. Sci. Comput.*, 63 (2015), 138–162.

[30] F. Yılmaz and B. Karasözen, Solving optimal control problems for the unsteady Burgers equation in COMSOL Multiphysics, *J. Comput. Appl. Math.*, 235 (2011), 4839–4850.

Received March 2021; revised July 2021; early access October 2021.

E-mail address: xenia.kerkhoff@tu-dortmund.de
E-mail address: sandra.may@tu-dortmund.de