Journal of Computational Physics 371 (2018) 280–314

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

Conservative model reduction for finite-volume models

Kevin Carlberg [*,1], Youngsoo Choi [2], Syuzanna Sargsyan [3]

Sandia National Laboratories, United States of America

ARTICLE INFO

Article history:
Received 1 December 2017
Received in revised form 9 May 2018
Accepted 10 May 2018
Available online xxxx

Keywords:
Nonlinear model reduction
Structure preservation
Finite-volume method
Galerkin projection
Least-squares Petrov–Galerkin projection
Conservative schemes

ABSTRACT

This work proposes a method for model reduction of finite-volume models that guarantees the resulting reduced-order model is conservative, thereby preserving the structure intrinsic to finite-volume discretizations. The proposed reduced-order models associate with optimization problems characterized by a minimum-residual objective function and nonlinear equality constraints that explicitly enforce conservation over subdomains. Conservative Galerkin projection arises from formulating this optimization problem at the time-continuous level, while conservative least-squares Petrov–Galerkin (LSPG) projection associates with a time-discrete formulation. We equip these approaches with hyper-reduction techniques in the case of nonlinear flux and source terms, and also provide approaches for handling infeasibility. In addition, we perform analyses that include deriving conditions under which conservative Galerkin and conservative LSPG are equivalent, as well as deriving *a posteriori* error bounds. Numerical experiments performed on a parameterized quasi-1D Euler equation demonstrate the ability of the proposed method to ensure not only global conservation, but also significantly lower state-space errors than nonconservative reduced-order models such as standard Galerkin and LSPG projection.

© 2018 Published by Elsevier Inc.

1. Introduction

The finite-volume method is commonly employed for discretizing systems of partial differential equations (PDEs) that associate with conservation laws, especially those in fluid dynamics. Rather than operating on the strong form of the PDE, the finite-volume method operates on the integral form of the PDE to numerically enforce conservation over each control volume comprising the computational mesh. Thus, *conservation* is the primary problem structure imposed by finite-volume discretizations; this contrasts with other discretization techniques that aim to preserve other properties, e.g., variational principles in the case of the finite-element discretizations.

Unfortunately, the computational burden imposed by high-fidelity finite-volume models is often prohibitive, as (1) the fine spatiotemporal resolution typically needed to ensure a verified, validated computational model can lead to extremely

[*] Corresponding author.
E-mail addresses: ktcarlb@sandia.gov (K. Carlberg), choi15@llnl.gov (Y. Choi), ssusie@uw.edu (S. Sargsyan).
URLs: http://sandia.gov/~ktcarlb (K. Carlberg), http://people.llnl.gov/choi15 (Y. Choi).

[1] Extreme-scale Data Science and Analytics Department, Sandia National Laboratories, Livermore, CA 94550.

[2] This work was performed while employed in the Extreme-scale Data Science and Analytics Department, Sandia National Laboratories, Livermore, CA 94550. *Current affiliation*: Lawrence Livermore National Laboratories. Lawrence Livermore National Laboratory is operated by Lawrence Livermore National Security, LLC, for the U.S. Department of Energy, National Nuclear Security Administration under Contract DE-AC52-07NA27344.

[3] This work was performed while pursing a Ph.D. at the University of Washington during employment as a summer intern in the Extreme-scale Data Science and Analytics Department, Sandia National Laboratories, Livermore, CA 94550. *Current affiliation*: HERE Technologies.

https://doi.org/10.1016/j.jcp.2018.05.019
0021-9991/© 2018 Published by Elsevier Inc.

large-scale models whose simulations consume months on supercomputers, and (2) many engineering problems are *real time* or *many queries* in nature. Such problems require the (parameterized) computational model to be simulated rapidly either due to a strict time-to-solution constraint in the case of real-time problems (e.g., model predictive control) or due to the need for hundreds or thousands of simulations in the case of many-query problems (e.g., statistical inversion).

Reduced-order models (ROMs) have been developed to mitigate this burden. These techniques first perform an *offline* stage during which they execute computationally costly training tasks (e.g., simulating the high-fidelity model for several parameter instances) to compute a low-dimensional 'trial' basis for the state. Next, these methods execute a computationally inexpensive *online* stage during which they rapidly compute approximate solutions for different points in the parameter space by projection: they compute solutions in the span of the trial basis while enforcing the high-fidelity model residual to be orthogonal to the subspace spanned by a low-dimensional 'test' basis. In the presence of nonlinearities, these techniques also introduce 'hyper-reduction' approximations to ensure the cost of simulating the ROM is independent of the high-fidelity-model dimension.

The most popular model-reduction approach for nonlinear dynamical systems such as those arising from finite-volume discretizations is Galerkin projection [52,21,39], wherein the test basis is set to be equal to the trial basis. The trial basis is often computed via proper orthogonal decomposition (POD) [33], but it can also be computed via the reduced-basis method; see Refs. [31,32,30], which apply the classical reduced-basis method to finite-volume problems. More recently, the least-squares Petrov–Galerkin (LSPG) projection method [16,17,15] was proposed, which has been computationally demonstrated to generate accurate and stable responses for turbulent, compressible flow problems on which Galerkin projection yielded unstable responses. Unfortunately, neither Galerkin nor LSPG projection directly preserves important problem structure related to conservation laws or finite-volume models.

To address this, alternative projection techniques have been developed for improving the performance of reduced-order models when applied to conservation laws, particularly those appearing in fluid dynamics. These include stabilizing inner products applied to finite-difference [46] and finite-element discretizations [9,36]; introducing dissipation via closure models [6,51,12,57,49] or numerical dissipation [34]; performing nonlinear Galerkin projection based on approximate inertial manifolds [41,50,35]; including a pressure-term representation [42,28]; modifying the POD basis by including many modes (such that dissipative modes are captured), changing the norm [34], enabling adaptivity [12,14], or including basis functions that resolve a range of scales [7] or respect the attractor's power balance [8]; modifying the projection by adopting a constrained Galerkin [45,26], constrained Petrov–Galerkin [24], or L^1-norm minimizing projection [1]; developing approaches tailored to the incompressible Navier–Stokes equations by introducing stabilizations based on supremizer-enriched velocity spaces and a pressure Poisson equation [54,53] or by modifying the Galerkin projection [38]; and improving the ROM's ability to capture shocks [43,29,14,55]. Among these contributions, only a subset is applicable to finite-volume discretizations. Further, no model-reduction method to date has been developed to preserve the structure intrinsic to finite-volume models: conservation. In particular, none of the above methods ensures that conservation holds over any subset of the computational domain, which can lead to spurious growth or dissipation of quantities that should be conserved in principle.

To this end, this work proposes a novel projection scheme for finite-volume models that ensures the reduced-order model is conservative over subdomains of the problem. The approach leverages the minimum-residual formulation of both Galerkin and least-squares Petrov–Galerkin projection by equipping their associated optimization problems with (generally nonlinear) equality constraints that explicitly enforce conservation over subdomains. The resulting *conservative* reduced-order models can be expressed as the solution to time-dependent saddle-point problems. The approach does not rely on a particular choice of reduced basis, although the reduced basis can affect feasibility of the associated optimization problems. New contributions in this work include:

1. Conservative Galerkin (Section 4.2) and conservative LSPG (Section 4.3) projection techniques, which ensure that the reduced-order models are conservative over subdomains of the original computational mesh. These methods are equipped with
   (a) techniques for handling infeasible constraints (Section 4.4), and
   (b) hyper-reduction techniques that respect the underlying finite-volume discretization to handle nonlinearities in the flux and source terms (Section 4.5).
2. Analysis, which includes:
   (a) demonstration that conservative Galerkin projection and time discretization are commutative (Theorem 4.3),
   (b) sufficient conditions for feasibility of conservative Galerkin (Proposition 5.1) and conservative LSPG (Proposition 5.2) projection,
   (c) conditions under which conservative Galerkin and conservative LSPG projection are equivalent (Theorem 5.1), and
   (d) *a posteriori* bounds (Section 5.3) for the error in the quantities conserved over subdomains (Theorem 5.3), in the null space (Lemma 5.1) and row space (Lemma 5.2) of the constraints, in the full state (Theorem 5.2), and in the conserved quantities (Lemma 5.3 and Theorem 5.3).
3. Numerical experiments on a parameterized quasi-1D Euler equation associated with modeling inviscid compressible flow in a converging–diverging nozzle (Section 6). These experiments demonstrate the merits of the proposed method and illustrate the importance of ensuring reduced-order models are globally conservative.

We remark that this work was first presented publically at the "Recent Developments in Numerical Methods for Model Reduction" workshop at the Institut Henri Poincaré on November 10, 2016.

Other works have also explored formulating reduced-order models that associate with constrained optimization problems. Zimmermann et al. [59] equip equality 'aerodynamic constraints' to ROMs applied to steady-state external flows, where the constraints associate with matching experimental data or target performance metrics in a design setting. Recently, Reddy et al. [45] propose equipping the time-discrete Galerkin ROM with inequality constraints that enforce solution positivity or a bound on the gas-void fraction. Relatedly, Fick et al. [26] proposed a modified Galerkin optimization problem applicable to the incompressible Navier–Stokes equations, where the inequality constraints associate with bounds on the generalized coordinates; these bounds correspond to the extreme values of the generalized coordinates arising during the training simulations.

The remainder of this paper is organized as follows. Section 2 describes finite-volume discretizations of conservation laws (Section 2.1) discretized in time with a linear multistep scheme (Section 2.2). Section 3 describes the (standard) nonlinear model-reduction methods of Galerkin (Section 3.1) and LSPG (Section 3.2) projection, as well as their hyper-reduced variants (Section 3.3) and interpretations when applied to finite-volume models (Section 3.4). Section 4 describes the proposed methodology, which is based on enforcing conservation over decompositions (Section 4.1) of the computational mesh. Here, Section 4.2 describes the proposed conservative Galerkin projection technique, Section 4.3 describes the proposed conservative LSPG projection method, Section 4.4 describes approaches for handling constraint infeasibility, Section 4.5 describes the application of hyper-reduction to the constraints that respects the underlying finite-volume discretization, and Section 4.6 describes briefly how the quantities required for the proposed ROMs can be constructed from training data. Next, Section 5 performs analysis, including proving sufficient conditions for feasibility (Section 5.1), providing conditions under with the conservative Galerkin and conservative LSPG models are equivalent (Section 5.2), and deriving local *a posteriori* error analysis (Section 5.3). Section 6 demonstrates the benefits off the proposed method on a parameterization of the one-dimensional (compressible) Euler equations applied to a converging–diverging nozzle. Finally, Section 7 concludes the paper.

In this work, matrices are denoted by capitalized bold letters, vectors by lowercase bold letters, and scalars by unbolded letters. The columns of a matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ are denoted by $\boldsymbol{a}_i \in \mathbb{R}^m$, $i \in \mathbb{N}(n)$ with $\mathbb{N}(a) := \{1, \ldots, a\}$ such that $\boldsymbol{A} \equiv [\boldsymbol{a}_1 \ \cdots \ \boldsymbol{a}_n]$. The scalar-valued matrix elements are denoted by $a_{ij} \in \mathbb{R}$ such that $\boldsymbol{a}_j \equiv \left[a_{1j} \ \cdots \ a_{mj}\right]^T$, $j \in \mathbb{N}(n)$. A superscript denotes the value of a variable at that time instance, e.g., $\boldsymbol{x}^n$ is the value of $\boldsymbol{x}$ at time $n\Delta t$, where Δt is the time step.

2. Finite-volume discretization

This work considers parameterized systems of conservation laws. In integral form, the associated governing equations correspond to

$$\frac{d}{dt}\int\limits_{\omega} u_i(\vec{x}, t; \boldsymbol{\mu})\, \mathrm{d}\vec{x} + \int\limits_{\gamma} \boldsymbol{g}_i(\vec{x}, t; \boldsymbol{\mu}) \cdot \boldsymbol{n}(\vec{x})\, \mathrm{d}\vec{s}(\vec{x}) = \int\limits_{\omega} s_i(\vec{x}, t; \boldsymbol{\mu})\, \mathrm{d}\vec{x}, \quad i \in \mathbb{N}(n_u),\ \forall \omega \subseteq \Omega, \tag{2.1}$$

which is solved in time domain $[0, T]$ with final time $T \in \mathbb{R}_+$, and a (parameterized) initial condition denoted by $u_i^0 : \Omega \times \mathcal{D} \to \mathbb{R}$ such that $u_i(\vec{x}, 0; \boldsymbol{\mu}) = u_i^0(\vec{x}; \boldsymbol{\mu})$. Here, ω with $\gamma := \partial\omega$ denotes any subset of the spatial domain of interest $\Omega \subset \mathbb{R}^d$ with $d \leq 3$, whose boundary is $\Gamma := \partial\Omega$, $\mathrm{d}\vec{s}(\vec{x})$ denotes integration with respect to the boundary, $u_i : \Omega \times [0, T] \times \mathcal{D} \to \mathbb{R}$, $i \in \mathbb{N}(n_u)$ denotes the ith conserved variable (per unit volume); $\boldsymbol{g}_i : \Omega \times [0, T] \times \mathcal{D} \to \mathbb{R}^d$, $i \in \mathbb{N}(n_u)$, denotes the flux associated with the ith conserved variable (per unit area per unit time); $\boldsymbol{n} : \gamma \to \mathbb{R}^d$ denotes the outward unit normal to ω; $s_i : \Omega \times [0, T] \times \mathcal{D} \to \mathbb{R}$, $i \in \mathbb{N}(n_u)$ denotes the source associated with the ith conserved variable (per unit volume per unit time); and $\mathcal{D} \subseteq \mathbb{R}^{n_\mu}$ denotes the parameter domain. We assume the domain Ω is independent of the parameters $\boldsymbol{\mu}$ for notational simplicity.

2.1. Spatial discretization

We consider the particular case where the governing equations (2.1) have been discretized in space by a finite-volume method. This implies that the spatial domain has been partitioned into a mesh $\mathcal{M}$ of $N_\Omega \in \mathbb{N}$ non-overlapping (closed, connected) control volumes $\Omega_i \subseteq \Omega$, $i \in \mathbb{N}(N_\Omega)$ such that $\Omega = \cup_{i=1}^{N_\Omega} \Omega_i$, which intersect only on their $(d-1)$-dimensional interface, i.e., $\mathrm{meas}(\Omega_i \cap \Omega_j) = 0$ for $i \neq j$, where $\mathrm{meas}(\omega) := \int_\omega \mathrm{d}\vec{x}$, $\forall \omega \subseteq \Omega$. We define the mesh as $\mathcal{M} := \{\Omega_i\}_{i=1}^{N_\Omega}$, and we denote the boundary of the ith control volume by $\Gamma_i := \partial\Omega_i$. The ith control-volume boundary is partitioned into a set of faces[4] denoted by $\mathcal{E}_i$ such that $\Gamma_i = \{\vec{x} \,|\, \vec{x} \in e,\ \forall e \in \mathcal{E}_i,\ i \in \mathbb{N}(|\mathcal{E}_i|)\}$. Then the full set of N_e faces within the mesh is $\mathcal{E} \equiv \{e_i\}_{i=1}^{N_e} := \cup_{i=1}^{N_\Omega} \mathcal{E}_i$. Applying Eq. (2.1) to each control volume in the mesh yields

[4] We note that this is a set of faces for $d = 3$, faces for $d = 2$, or simply extremities for $d = 1$.

$$\frac{d}{dt}\int_{\Omega_j} u_i(\vec{x},t;\boldsymbol{\mu})\,\mathrm{d}\vec{x} + \int_{\Gamma_j} \boldsymbol{g}_i(\vec{x},t;\boldsymbol{\mu})\cdot \boldsymbol{n}_j(\vec{x})\,\mathrm{d}\vec{s}(\vec{x}) = \int_{\Omega_j} s_i(\vec{x},t;\boldsymbol{\mu})\,\mathrm{d}\vec{x},\quad i\in\mathbb{N}(n_u),\ j\in\mathbb{N}(N_\Omega), \tag{2.2}$$

where $\boldsymbol{u}\equiv\left(u_1,\ldots,u_{n_u}\right)$ and $\boldsymbol{n}_j:\Gamma_j\rightarrow\mathbb{R}^d$ denotes the unit normal to control volume Ω_j. Finite-volume schemes complete the spatial discretization by introducing a state vector $\boldsymbol{x}:[0,T]\times\mathcal{D}\rightarrow\mathbb{R}^N$ with $N=N_\Omega n_u$ whose elements comprise

$$x_{\mathcal{I}(i,j)}(t;\boldsymbol{\mu}) = \frac{1}{|\Omega_j|}\int_{\Omega_j} u_i(\vec{x},t;\boldsymbol{\mu})\,\mathrm{d}\vec{x},\quad i\in\mathbb{N}(n_u),\ j\in\mathbb{N}(N_\Omega), \tag{2.3}$$

where $\mathcal{I}:\mathbb{N}(n_u)\times\mathbb{N}(N_\Omega)\rightarrow\mathbb{N}(N)$ denotes a mapping from conservation-law index and control-volume index to degree of freedom, and a velocity vector $\boldsymbol{f}:(\boldsymbol{\xi},\tau;\boldsymbol{\nu})\mapsto\boldsymbol{f}^g(\boldsymbol{\xi},\tau;\boldsymbol{\nu})+\boldsymbol{f}^s(\boldsymbol{\xi},\tau;\boldsymbol{\nu})$ with $\boldsymbol{f}^g,\boldsymbol{f}^s:\mathbb{R}^N\times[0,T]\times\mathcal{D}\rightarrow\mathbb{R}^N$ whose elements consist of

$$\begin{aligned} f^g_{\mathcal{I}(i,j)}(\boldsymbol{x},t;\boldsymbol{\mu}) &= -\frac{1}{|\Omega_j|}\int_{\Gamma_j}\boldsymbol{g}_i^{\mathrm{FV}}(\boldsymbol{x};\vec{x},t;\boldsymbol{\mu})\cdot\boldsymbol{n}_j(\vec{x})\,\mathrm{d}\vec{s}(\vec{x})\\ f^s_{\mathcal{I}(i,j)}(\boldsymbol{x},t;\boldsymbol{\mu}) &= \frac{1}{|\Omega_j|}\int_{\Omega_j} s_i^{\mathrm{FV}}(\boldsymbol{x};\vec{x},t;\boldsymbol{\mu})\,\mathrm{d}\vec{x}\end{aligned} \tag{2.4}$$

for $i\in\mathbb{N}(n_u),\ \ j\in\mathbb{N}(N_\Omega)$. Here, $\boldsymbol{g}_i^{\mathrm{FV}}:\mathbb{R}^N\times\Omega\times[0,T]\times\mathcal{D}\rightarrow\mathbb{R}^d$, $i\in\mathbb{N}(n_u)$ denotes the approximated (or reconstructed) flux associated with the ith conserved variable (per unit area per unit time); and $s_i^{\mathrm{FV}}:\mathbb{R}^N\times\Omega\times[0,T]\times\mathcal{D}\rightarrow\mathbb{R}$, $i\in\mathbb{N}(n_u)$ denotes the approximated source associated with the ith conserved variable (per unit volume per unit time), which may arise, e.g., from applying a quadrature rule to evaluate the integral. We emphasize that both the approximated flux $\boldsymbol{g}_i^{\mathrm{FV}}$ and approximated source s_i^{FV} will in general depend on the entire state vector $\boldsymbol{x}$, e.g., due to high-order flux reconstructions or reactions, respectively.

Substituting $\int_{\Omega_j} u_i(\vec{x},t;\boldsymbol{\mu})\,\mathrm{d}\vec{x}\leftarrow|\Omega_j|x_{\mathcal{I}(i,j)}(t;\boldsymbol{\mu})$, $\boldsymbol{g}_i\leftarrow\boldsymbol{g}_i^{\mathrm{FV}}$, and $s_i\leftarrow s_i^{\mathrm{FV}}$ in Eq. (2.2) and dividing by $|\Omega_j|$ yields

$$\frac{d\boldsymbol{x}}{dt}=\boldsymbol{f}(\boldsymbol{x},t;\boldsymbol{\mu}),\qquad \boldsymbol{x}(0;\boldsymbol{\mu})=\boldsymbol{x}^0(\boldsymbol{\mu}), \tag{2.5}$$

where $x^0_{\mathcal{I}(i,j)}(\boldsymbol{\mu}):=\frac{1}{|\Omega_i|}\int_{\Omega_j}u_i^0(\vec{x};\boldsymbol{\mu})\,\mathrm{d}\vec{x}$. This is a parameterized system of nonlinear ordinary differential equations (ODEs) characterizing an initial value problem, which we consider to be our full-order model (FOM).

Remark 1 *(Full-order model ODE: finite-volume interpretation).* From the definitions of the state (2.3) and velocity (2.4), the full-order-model ODE residual element $dx_{\mathcal{I}(i,j)}/dt-f_{\mathcal{I}(i,j)}$ can be interpreted as the (normalized) *rate of violation of conservation* in variable u_i in control volume Ω_j at time instance t under one approximation: the flux and source terms are approximated using the finite-volume discretization (i.e., $\boldsymbol{g}_i\leftarrow\boldsymbol{g}_i^{\mathrm{FV}}$, and $s_i\leftarrow s_i^{\mathrm{FV}}$).

Remark 2 *(Flux velocity from face fluxes).* The elements of the flux velocity can be computed from a vector of face fluxes $\boldsymbol{h}:\mathbb{R}^N\times[0,T]\times\mathcal{D}\rightarrow\mathbb{R}^{n_uN_e}$, whose elements are

$$h_{\mathcal{J}(i,j)}(\boldsymbol{x},t;\boldsymbol{\mu})=\int_{e_j}\boldsymbol{g}_i^{\mathrm{FV}}(\boldsymbol{x};\vec{x},t;\boldsymbol{\mu})\cdot\boldsymbol{n}_j^e(\vec{x})\,\mathrm{d}\vec{s}(\vec{x}),\quad i\in\mathbb{N}(n_u),\ j\in\mathbb{N}(N_e), \tag{2.6}$$

where $\boldsymbol{n}_j^e:e_j\rightarrow\mathbb{R}^d$ denotes the unit normal assigned to face e_j (using any convention) and $\mathcal{J}:\mathbb{N}(n_u)\times\mathbb{N}(N_e)\rightarrow\mathbb{N}(n_uN_e)$ denotes a mapping from conservation-law index and face index to degrees of freedom defined on the faces. This mapping is provided by

$$f^g_{\mathcal{I}(i,j)}(\boldsymbol{x},t;\boldsymbol{\mu})=\sum_{k\,|\,e_k\in\Gamma_j} b_{\mathcal{I}(i,j),\mathcal{J}(i,k)}h_{\mathcal{J}(i,k)}(\boldsymbol{x},t;\boldsymbol{\mu}), \tag{2.7}$$

where the elements of $\boldsymbol{B}\in\mathbb{R}^{N\times n_uN_e}$ are

$$b_{\mathcal{I}(i,j),\mathcal{J}(\ell,k)}=\begin{cases}-\delta_{i\ell}/|\Omega_j|, & e_k\in\Gamma_j;\ \boldsymbol{n}_j(\vec{x})=\boldsymbol{n}_k^e(\vec{x}),\ \vec{x}\in e_k\\ \delta_{i\ell}/|\Omega_j|, & e_k\in\Gamma_j;\ \boldsymbol{n}_j(\vec{x})=-\boldsymbol{n}_k^e(\vec{x}),\ \vec{x}\in e_k\\ 0, & \text{otherwise,}\end{cases} \tag{2.8}$$

where δ_{ij} denotes the Kronecker delta. In matrix form, Eq. (2.7) becomes

$$\boldsymbol{f}^g(\boldsymbol{x},t;\boldsymbol{\mu})=\boldsymbol{B}\boldsymbol{h}(\boldsymbol{x},t;\boldsymbol{\mu}). \tag{2.9}$$

This formulation will be exploited in Section 4, where we introduce the proposed method.

The full-order model ODE (2.5) is typically the starting point for developing reduced-order models for nonlinear dynamical systems. In this work, we exploit the particular structure underlying the dynamical system arising from the definitions of the state (2.3) and velocity (2.4).

2.2. Time discretization

A time discretization is required to solve (2.5) numerically. For simplicity, we restrict the focus in this work to linear multistep schemes, although other time integrators could be considered; see, e.g., Ref. [15], which develops LSPG reduced-order models for explicit, fully implicit, and diagonally implicit Runge–Kutta schemes. Applying a linear k-step method to numerically solve Eq. (2.5) at a given parameter instance $\boldsymbol{\mu} \in \mathcal{D}$ can be written as

$$\sum_{j=0}^{k} \alpha_j \boldsymbol{x}^{n-j} = \Delta t \sum_{j=0}^{k} \beta_j \boldsymbol{f}(\boldsymbol{x}^{n-j}, t^{n-j}; \boldsymbol{\mu}), \tag{2.10}$$

where $\Delta t \in \mathbb{R}_+$ denotes the time step, $\boldsymbol{x}^k$ denotes the numerical approximation to $\boldsymbol{x}(t^k)$, i.e.,

$$x^k_{\mathcal{I}(i,j)} = \frac{1}{|\Omega_j|} \int\limits_{\Omega_j} u_i^k(\vec{x})\, \mathrm{d}\vec{x}, \tag{2.11}$$

where $u_i^k(\vec{x})$ denotes the numerical approximation to $u_i(\vec{x}, t^k)$. The coefficients α_j and β_j define a particular linear multistep scheme, $\alpha_0 \neq 0$ and $\sum_{j=0}^{k} \alpha_j = 0$ is necessary for consistency, and the method is implicit if $\beta_0 \neq 0$. For notational simplicity, we employ a uniform time grid $t^k = t^{k-1} + \Delta t$, $k \in \mathbb{N}(N_T)$ with $t^0 = 0$ and $N_T := T/\Delta t$. The fully discrete full-order model, which is sometimes denoted as the FOM OΔE, is characterized by the following system of algebraic equations to be solved at each time instance $n \in \mathbb{N}(N_T)$:

$$\boldsymbol{r}^n(\boldsymbol{x}^n; \boldsymbol{\mu}) = 0, \tag{2.12}$$

where $\boldsymbol{r}^n : \mathbb{R}^N \to \mathbb{R}^N$ denotes the linear multistep residual, which is defined as

$$\boldsymbol{r}^n(\boldsymbol{w}; \boldsymbol{\nu}) := \alpha_0 \boldsymbol{w} - \Delta t \beta_0 \boldsymbol{f}(\boldsymbol{w}, t^n; \boldsymbol{\nu}) + \sum_{j=1}^{k} \alpha_j \boldsymbol{x}^{n-j}(\boldsymbol{\nu}) - \Delta t \sum_{j=1}^{k} \beta_j \boldsymbol{f}(\boldsymbol{x}^{n-j}, t^{n-j}; \boldsymbol{\nu}). \tag{2.13}$$

The unknown vector $\boldsymbol{w} \in \mathbb{R}^N$ can be interpreted as

$$w_{\mathcal{I}(i,j)} = \frac{1}{|\Omega_j|} \int\limits_{\Omega_j} \tilde{u}_i(\vec{x})\, \mathrm{d}\vec{x}, \tag{2.14}$$

and $\tilde{u}_i$ denotes an approximation to the ith conserved variable $u_i(\vec{x}, t^n)$ when evaluating the residual (2.13) at the nth time instance.

Adams methods. Adams methods consider the integrated form of Eq. (2.5)

$$\boldsymbol{x}^n = \boldsymbol{x}^{n-1} + \int\limits_{t^{n-1}}^{t^n} \boldsymbol{f}(\boldsymbol{x}, t; \boldsymbol{\mu}) dt, \quad n = 1, \ldots, N_T, \tag{2.15}$$

and apply a polynomial approximation to the integrand. In particular, the pth-order Adams scheme employs coefficients $\alpha_0 = 1$, $\alpha_1 = -1$, and $\alpha_j = 0$, $j > 1$ and coefficients β_j that associate with a polynomial interpolation of the integrand. In the explicit ($\beta_0 = 0$) case, these are Adams–Bashforth methods with

$$\Delta t \sum_{j=1}^{k} \beta_j \boldsymbol{f}(\boldsymbol{x}^{n-j}, t^{n-j}; \boldsymbol{\mu}) = \int\limits_{t^{n-1}}^{t^n} I_k^{n-1}(\boldsymbol{f}(\boldsymbol{\mu}); t) dt, \tag{2.16}$$

where $\boldsymbol{f}(\boldsymbol{\mu}) := (\boldsymbol{f}(\boldsymbol{x}^0, t^0; \boldsymbol{\mu}), \ldots \boldsymbol{f}(\boldsymbol{x}^{N_T}, t^{N_T}; \boldsymbol{\mu}))$ and the polynomial approximation (in time) of any time-grid-dependent quantity $\boldsymbol{\xi} := (\boldsymbol{\xi}^1, \ldots, \boldsymbol{\xi}^k)$ using data at $(t^n, \ldots, t^{n+1-k})$ (with $k \geq 1$) is

$$I_k^n(\boldsymbol{\xi}; t) := \sum_{i=1}^{k} \boldsymbol{\xi}^{n+1-i} \prod_{j=1, j\neq i}^{k} \frac{t - t^{n+1-j}}{t^{n+1-i} - t^{n+1-j}}. \tag{2.17}$$

In the implicit case (with $\beta_0 \neq 0$), these are Adams–Moulton methods with coefficients β_j satisfying

$$\Delta t \sum_{j=0}^{k} \beta_j \boldsymbol{f}(\boldsymbol{x}^{n-j}, t^{n-j}; \boldsymbol{\mu}) = \int\limits_{t^{n-1}}^{t^n} I^n_{k+1}(\boldsymbol{f}(\boldsymbol{\mu}); t)dt, \tag{2.18}$$

Thus, the time-discrete residual (2.13) becomes

$$\boldsymbol{r}^n(\boldsymbol{x}^n; \boldsymbol{\mu}) = \boldsymbol{x}^n - \boldsymbol{x}^{n-1} - \int\limits_{t^{n-1}}^{t^n} I(\boldsymbol{f}(\boldsymbol{\mu}); t)dt, \tag{2.19}$$

where $I = I^{n-1}_k$ in the explicit case and $I = I^n_{k+1}$ in the implicit case. Substituting the definitions of the time-discrete state (2.11) and velocity (2.4) in (2.19) yields

$$\begin{aligned} r^n_{\mathcal{I}(i,j)}(\boldsymbol{x}^n; \boldsymbol{\mu}) = & \frac{1}{|\Omega_j|} \int\limits_{\Omega_j} u^n_i(\vec{x})\, d\vec{x} - \frac{1}{|\Omega_j|} \int\limits_{\Omega_j} u^{n-1}_i(\vec{x})\, d\vec{x} \\ & + \frac{1}{|\Omega_j|} \int\limits_{t^{n-1}}^{t^n} \int\limits_{\Gamma_j} I(\mathbf{g}^{\mathrm{FV}}_i(\vec{x}; \boldsymbol{\mu}); t) \cdot \boldsymbol{n}_j(\vec{x})\, d\vec{s}(\vec{x})dt - \frac{1}{|\Omega_j|} \int\limits_{t^{n-1}}^{t^n} \int\limits_{\Omega_j} I(s^{\mathrm{FV}}_i(\vec{x}; \boldsymbol{\mu}); t)\, d\vec{x}dt, \end{aligned} \tag{2.20}$$

where

$$\begin{aligned} \mathbf{g}^{\mathrm{FV}}_i(\vec{x}; \boldsymbol{\mu}) &:= (\mathbf{g}^{\mathrm{FV}}_i(\boldsymbol{x}^0; \vec{x}, t^0; \boldsymbol{\mu}), \ldots, \mathbf{g}^{\mathrm{FV}}_i(\boldsymbol{x}^{N_T}; \vec{x}, t^{N_T}; \boldsymbol{\mu})) \\ s^{\mathrm{FV}}_i(\vec{x}; \boldsymbol{\mu}) &:= (s^{\mathrm{FV}}_i(\boldsymbol{x}^0; \vec{x}, t^0; \boldsymbol{\mu}), \ldots, s^{\mathrm{FV}}_i(\boldsymbol{x}^{N_T}; \vec{x}, t^{N_T}; \boldsymbol{\mu})). \end{aligned} \tag{2.21}$$

Remark 3 *(Full-order model OΔE: finite-volume interpretation for Adams methods).* Eq. (2.20) shows that the full-order-model OΔE residual element $r^n_{\mathcal{I}(i,j)}$ in the case of Adams methods can be interpreted as the (normalized) *violation of conservation* in variable u_i in control volume Ω_j over time interval $[t^{n-1}, t^n]$ under two approximations: (1) the flux and source terms are approximated using the finite-volume discretization (i.e., $\mathbf{g}_i \leftarrow \mathbf{g}^{\mathrm{FV}}_i$, and $s_i \leftarrow s^{\mathrm{FV}}_i$), and (2) a polynomial interpolation is used to approximate the integrand for time integration.

3. Reduced-order models

During the online stage, projection-based reduced-order models compute an approximate solution $\tilde{\boldsymbol{x}} \approx \boldsymbol{x}$ that lies in a low-dimensional affine trial subspace $\tilde{\boldsymbol{x}}(t; \boldsymbol{\mu}) \in \boldsymbol{x}^0(\boldsymbol{\mu}) + \mathrm{Ran}(\boldsymbol{\Phi})$, i.e.,

$$\tilde{\boldsymbol{x}}(t; \boldsymbol{\mu}) = \boldsymbol{x}^0(\boldsymbol{\mu}) + \boldsymbol{\Phi}\hat{\boldsymbol{x}}(t; \boldsymbol{\mu}), \tag{3.1}$$

where $\boldsymbol{\Phi} \in \mathbb{R}^{N \times p}$ is the reduced-basis matrix of dimension $p \le N$, which we assume without loss of generality satisfies $\boldsymbol{\Phi}^T\boldsymbol{\Phi} = \boldsymbol{I}$, $\hat{\boldsymbol{x}} : [0, T] \times \mathcal{D} \to \mathbb{R}^p$ denotes the generalized coordinates, and $\mathrm{Ran}(\boldsymbol{A})$ denotes the range of a matrix $\boldsymbol{A}$. This basis can be computed in a variety of ways during the offline stage, e.g., eigenmode analysis, POD [33], or the reduced-basis method [44,47]. Substituting the approximation $\boldsymbol{x} \leftarrow \tilde{\boldsymbol{x}}$ into governing equations (2.5) yields an overdetermined system of N equations in p unknowns. To compute a unique solution, reduced-order models must enforce the residual to be orthogonal to a p-dimensional test subspace. Galerkin and LSPG projection differ in their choices of this subspace; each choice leads to an approximate solution that exhibits a particular notion of optimality.

3.1. Galerkin projection

Galerkin projection employs a test subspace of $\mathrm{Ran}(\boldsymbol{\Phi})$ and thus enforces the residual to be orthogonal to $\mathrm{Ran}(\boldsymbol{\Phi})$, i.e., the Galerkin ODE is

$$\frac{d\hat{\boldsymbol{x}}}{dt} = \boldsymbol{\Phi}^T \boldsymbol{f}(\boldsymbol{x}^0(\boldsymbol{\mu}) + \boldsymbol{\Phi}\hat{\boldsymbol{x}}, t; \boldsymbol{\mu}), \quad \hat{\boldsymbol{x}}(0) = \boldsymbol{0}. \tag{3.2}$$

Applying a linear multistep scheme to integrate Eq. (3.2) in time yields the Galerkin OΔE

$$\boldsymbol{\Phi}^T \boldsymbol{r}^n(\boldsymbol{x}^0(\boldsymbol{\mu}) + \boldsymbol{\Phi}\hat{\boldsymbol{x}}^n(\boldsymbol{\mu}); \boldsymbol{\mu}) = 0. \tag{3.3}$$

As demonstrated, e.g., in Ref. [15], Galerkin projection exhibits continuous optimality if the reduced basis is orthogonal, i.e., $\boldsymbol{\Phi}^T\boldsymbol{\Phi} = \boldsymbol{I}$, as the Galerkin ROM computes the approximated velocity that minimizes the ℓ^2-norm of the FOM ODE residual (2.5) over $\mathrm{Ran}(\boldsymbol{\Phi})$, i.e.,

$$\frac{d\tilde{\boldsymbol{x}}}{dt}\left(\boldsymbol{x}^0(\boldsymbol{\mu})+\boldsymbol{\Phi}\hat{\boldsymbol{x}},t;\boldsymbol{\mu}\right)=\underset{\boldsymbol{v}\in\mathrm{Ran}(\boldsymbol{\Phi})}{\arg\min}\,\|\boldsymbol{r}(\boldsymbol{v},\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}},t;\boldsymbol{\mu})\|_2 \tag{3.4}$$

or equivalently

$$\frac{d\hat{\boldsymbol{x}}}{dt}\left(\boldsymbol{x}^0(\boldsymbol{\mu})+\boldsymbol{\Phi}\hat{\boldsymbol{x}},t;\boldsymbol{\mu}\right)=\underset{\hat{\boldsymbol{v}}\in\mathbb{R}^p}{\arg\min}\|\boldsymbol{r}(\boldsymbol{\Phi}\hat{\boldsymbol{v}},\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}},t;\boldsymbol{\mu})\|_2, \tag{3.5}$$

where

$$\boldsymbol{r}(\boldsymbol{v},\boldsymbol{\xi},\tau;\boldsymbol{\nu}):=\boldsymbol{v}-\boldsymbol{f}(\boldsymbol{\xi},\tau;\boldsymbol{\nu}) \tag{3.6}$$

denotes the FOM ODE residual.

Remark 4 *(Galerkin ROM ODE: finite-volume interpretation).* From the time-continuous optimality of the Galerkin ROM ODE (3.5) and the finite-volume interpretation of the FOM ODE in Remark 1, the Galerkin ROM ODE (3.2) can be interpreted as minimizing the sum of squared (normalized) *rates of violation of conservation* across all variables u_i, $i\in\mathbb{N}(n_u)$ and control volumes Ω_j, $j\in\mathbb{N}(N_\Omega)$ at time instance t under one approximation: the flux and source terms are approximated using the finite-volume discretization (i.e., $\boldsymbol{g}_i\leftarrow\boldsymbol{g}_i^{\mathrm{FV}}$, and $s_i\leftarrow s_i^{\mathrm{FV}}$).

3.2. LSPG projection

In contrast, LSPG projection associates with a minimum-residual formulation applied to the (time-discrete) O∆E (2.12), i.e.,

$$\tilde{\boldsymbol{x}}^n=\underset{\boldsymbol{z}\in\boldsymbol{x}^0(\boldsymbol{\mu})+\mathrm{Ran}(\boldsymbol{\Phi})}{\arg\min}\|\boldsymbol{r}^n(\boldsymbol{z};\boldsymbol{\mu})\|_2 \tag{3.7}$$

or equivalently

$$\hat{\boldsymbol{x}}^n=\underset{\hat{\boldsymbol{z}}\in\mathbb{R}^p}{\arg\min}\|\boldsymbol{r}^n(\boldsymbol{x}^0(\boldsymbol{\mu})+\boldsymbol{\Phi}\hat{\boldsymbol{z}};\boldsymbol{\mu})\|_2. \tag{3.8}$$

The necessary optimality conditions for problem (3.8) associate with stationarity of the objective function, i.e., the solution $\hat{\boldsymbol{x}}^n$ satisfies

$$\boldsymbol{\Psi}^n(\hat{\boldsymbol{x}}^n;\boldsymbol{\mu})^T\boldsymbol{r}^n(\boldsymbol{x}^0(\boldsymbol{\mu})+\boldsymbol{\Phi}\hat{\boldsymbol{x}}^n;\boldsymbol{\mu})=\boldsymbol{0}, \tag{3.9}$$

where the LSPG test basis $\boldsymbol{\Psi}^n:\mathbb{R}^p\times\mathcal{D}\to\mathbb{R}^{N\times p}$ is

$$\boldsymbol{\Psi}^n(\hat{\boldsymbol{w}};\boldsymbol{\nu}):=\frac{\partial\boldsymbol{r}^n}{\partial\boldsymbol{w}}(\boldsymbol{x}^0(\boldsymbol{\mu})+\boldsymbol{\Phi}\hat{\boldsymbol{w}};\boldsymbol{\mu})\boldsymbol{\Phi}=\left(\alpha_0\boldsymbol{I}+\beta_0\Delta t\frac{\partial\boldsymbol{f}}{\partial\boldsymbol{\xi}}(\boldsymbol{x}^0(\boldsymbol{\mu})+\boldsymbol{\Phi}\hat{\boldsymbol{w}},t^n;\boldsymbol{\mu})\right)\boldsymbol{\Phi}. \tag{3.10}$$

Eq. (3.10) reveals that LSPG projection adds the term $\frac{\beta_0\Delta t}{\alpha_0}\frac{\partial\boldsymbol{f}}{\partial\boldsymbol{\xi}}(\boldsymbol{x}^0(\boldsymbol{\mu})+\boldsymbol{\Phi}\hat{\boldsymbol{w}},t^n;\boldsymbol{\mu})\boldsymbol{\Phi}$ to the test basis employed by Galerkin projection.

Remark 5 *(LSPG ROM O∆E: finite-volume interpretation for Adams methods).* From the time-discrete optimality of the LSPG ROM O∆E (3.8) and the finite-volume interpretation of the FOM O∆E for Adams methods in Remark 3, the LSPG ROM O∆E (3.8) can be interpreted as minimizing the sum of squared (normalized) *violation of conservation* across all variables u_i, $i\in\mathbb{N}(n_u)$ and control volumes Ω_j, $j\in\mathbb{N}(n_u)$ over time interval $[t^{n-1},t^n]$ under two approximations: (1) the flux and source terms are approximated using the finite-volume discretization (i.e., $\boldsymbol{g}_i\leftarrow\boldsymbol{g}_i^{\mathrm{FV}}$, and $s_i\leftarrow s_i^{\mathrm{FV}}$), and (2) a polynomial interpolation is used to approximate the integrand for time integration.

3.3. Hyper-reduction

In the case of nonlinear dynamical systems, projection is insufficient to yield computational savings, as high-dimensional nonlinear quantities $\boldsymbol{r}$ and $\boldsymbol{r}^n$ must be repeatedly computed, projected as $\boldsymbol{\Phi}^T\boldsymbol{r}$ and $(\boldsymbol{\Psi}^n)^T\boldsymbol{r}^n$, and differentiated (in the case of implicit time integrators) for Galerkin and LSPG ROMs, respectively. To reduce this computational bottleneck, several 'hyper-reduction' techniques have been developed that require computing only a sample of the elements of these nonlinear vector-valued functions. These techniques include collocation [5,48,37], gappy POD [23,13,5,16,17], the empirical interpolation method (EIM) [10,19,27,22,4], reduced-order quadrature [3], finite-element subassembly methods [2,25], and reduced-basis-sparsification techniques [18].

In the present context, hyper-reduction can be achieved by replacing the residuals appearing in the objective functions of (3.5) and (3.8) by $\tilde{\boldsymbol{r}}(\approx\boldsymbol{r})$ and $\tilde{\boldsymbol{r}}^n(\approx\boldsymbol{r}^n)$, respectively, such that the hyper-reduced optimization problems for Galerkin and LSPG projection become

$$\frac{d\hat{\boldsymbol{x}}}{dt}\left(\boldsymbol{x}^0(\boldsymbol{\mu})+\boldsymbol{\Phi}\hat{\boldsymbol{x}},t;\boldsymbol{\mu}\right)=\underset{\hat{\boldsymbol{v}}\in\mathbb{R}^p}{\arg\min}\|\tilde{\boldsymbol{r}}(\boldsymbol{\Phi}\hat{\boldsymbol{v}},\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}},t;\boldsymbol{\mu})\|_2, \tag{3.11}$$

and

$$\hat{\boldsymbol{x}}^n(\boldsymbol{\mu})=\underset{\hat{\boldsymbol{z}}\in\mathbb{R}^p}{\arg\min}\|\tilde{\boldsymbol{r}}^n(\boldsymbol{x}^0(\boldsymbol{\mu})+\boldsymbol{\Phi}\hat{\boldsymbol{z}};\boldsymbol{\mu})\|_2, \tag{3.12}$$

respectively. These residual approximations are typically constructed in one of two ways. Later, Section 4.5 proposes a third technique tailored to finite-volume discretizations.

1. **Residual hyper-reduction.** This approach amounts to

$$\tilde{\boldsymbol{r}}=\boldsymbol{\Phi}_r(\boldsymbol{P}_r\boldsymbol{\Phi}_r)^+\boldsymbol{P}_r\boldsymbol{r},\quad \tilde{\boldsymbol{r}}^n=\boldsymbol{\Phi}_r(\boldsymbol{P}_r\boldsymbol{\Phi}_r)^+\boldsymbol{P}_r\boldsymbol{r}^n \tag{3.13}$$

in the case of gappy POD hyper-reduction, or simply

$$\tilde{\boldsymbol{r}}=\boldsymbol{P}_r^T\boldsymbol{P}_r\boldsymbol{r},\quad \tilde{\boldsymbol{r}}^n=\boldsymbol{P}_r^T\boldsymbol{P}_r\boldsymbol{r}^n \tag{3.14}$$

in the case of collocation. Here, $\boldsymbol{P}_r\in\{0,1\}^{n_{p,r}\times N}$ denotes a sampling matrix comprising selected rows of the $N\times N$ identity matrix, while $\boldsymbol{\Phi}_r\in\mathbb{R}_\star^{N\times p_r}$ denotes a $p_r(\leq N)$-dimensional reduced-basis matrix constructed for the residual, a superscript $+$ denotes the Moore–Penrose pseudoinverse, and $\mathbb{R}_\star^{m\times n}$ denotes the set of full-column rank $m\times n$ matrices (the non-compact Stiefel manifold). This approach has the advantage of associating hyper-reduced optimization problems (3.11) and (3.12) with a weighted-norm variant of the original optimization problems (3.5) and (3.8), i.e.,

$$\frac{d\hat{\boldsymbol{x}}}{dt}\left(\boldsymbol{x}^0(\boldsymbol{\mu})+\boldsymbol{\Phi}\hat{\boldsymbol{x}}\right)=\underset{\hat{\boldsymbol{v}}\in\mathbb{R}^p}{\arg\min}\|\boldsymbol{A}\boldsymbol{r}(\boldsymbol{\Phi}\hat{\boldsymbol{v}},\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}},t;\boldsymbol{\mu})\|_2,\quad \hat{\boldsymbol{x}}^n=\underset{\hat{\boldsymbol{z}}\in\mathbb{R}^p}{\arg\min}\|\boldsymbol{A}\boldsymbol{r}^n(\boldsymbol{x}^0(\boldsymbol{\mu})+\boldsymbol{\Phi}\hat{\boldsymbol{z}};\boldsymbol{\mu})\|_2, \tag{3.15}$$

where $\boldsymbol{A}=(\boldsymbol{P}_r\boldsymbol{\Phi}_r)^+\boldsymbol{P}_r$ and $\boldsymbol{A}=\boldsymbol{P}_r$ in the case of gappy POD and collocation, respectively.

2. **Velocity hyper-reduction.** This approach employs an approximated residual constructed from hyper-reduction performed on the velocity vector only, i.e.,

$$\tilde{\boldsymbol{r}}(\boldsymbol{v},\boldsymbol{\xi},\tau;\boldsymbol{\nu})=\boldsymbol{v}-\tilde{\boldsymbol{f}}(\boldsymbol{\xi},\tau;\boldsymbol{\nu}) \tag{3.16}$$

$$\tilde{\boldsymbol{r}}^n(\boldsymbol{w};\boldsymbol{\nu})=\alpha_0\boldsymbol{w}-\Delta t\beta_0\tilde{\boldsymbol{f}}(\boldsymbol{w},t^n;\boldsymbol{\nu})+\sum_{j=1}^{k}\alpha_j\boldsymbol{x}^{n-j}(\boldsymbol{\nu})-\Delta t\sum_{j=1}^{k}\beta_j\tilde{\boldsymbol{f}}(\boldsymbol{x}^{n-j},t^{n-j};\boldsymbol{\nu}), \tag{3.17}$$

where

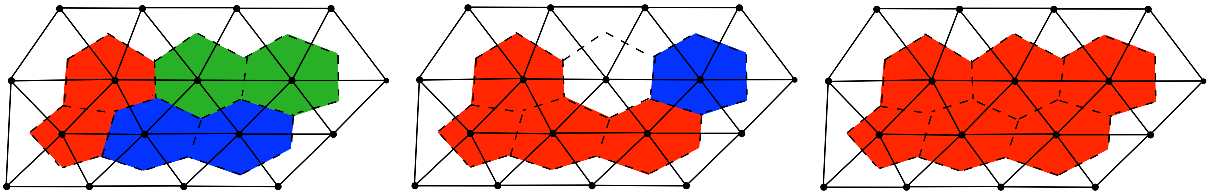
$$\tilde{\boldsymbol{f}}=\boldsymbol{\Phi}_f(\boldsymbol{P}_f\boldsymbol{\Phi}_f)^+\boldsymbol{P}_f\boldsymbol{f}\quad\text{or}\quad\tilde{\boldsymbol{f}}=\boldsymbol{P}_f^T\boldsymbol{P}_f\boldsymbol{f} \tag{3.18}$$

in the case of gappy POD or collocation, respectively. Here, $\boldsymbol{P}_f\in\{0,1\}^{n_{p,f}\times N}$ denotes a sampling matrix comprising selected rows of the identity matrix, while $\boldsymbol{\Phi}_f\in\mathbb{R}_\star^{N\times p_f}$ denotes a $p_f(\leq N)$-dimensional reduced-basis matrix constructed for the velocity. This approach has the advantage of limiting the hyper-reduction approximation to the nonlinear component of the residual.

We note that the gappy POD approximations are equivalent to empirical interpolation when the number of samples is equal to the number of reduced-basis elements (i.e., $n_{p,r}=p_r$, $n_{p,f}=p_f$), as the pseudo-inverse is equal to the inverse and the approximation interpolates the nonlinear function at the sampled elements in this case. Further, the POD–(D)EIM method [19] corresponds to Galerkin projection with gappy POD velocity hyper-reduction and $n_{p,f}=p_f$, in which case the hyper-reduced Galerkin ODE becomes

$$\frac{d\hat{\boldsymbol{x}}}{dt}=\boldsymbol{\Phi}^T\boldsymbol{\Phi}_f(\boldsymbol{P}_f\boldsymbol{\Phi}_f)^{-1}\boldsymbol{P}_f\boldsymbol{f}(\boldsymbol{x}^0(\boldsymbol{\mu})+\boldsymbol{\Phi}\hat{\boldsymbol{x}},t;\boldsymbol{\mu}),\quad \hat{\boldsymbol{x}}(0)=\boldsymbol{0}. \tag{3.19}$$

In addition, the GNAT method [16,17] corresponds to LSPG projection with gappy POD residual hyper-reduction. In principle, the two projection techniques and two hyper-reduction approaches above yield four possible (hyper-reduced) reduced-order models that could be constructed.



(a) Decomposed mesh $\bar{\mathcal{M}}$ with $N_{\bar{\Omega}} = 3$

(b) Decomposed mesh $\bar{\mathcal{M}}$ with $N_{\bar{\Omega}} = 2$

(c) Decomposed mesh $\bar{\mathcal{M}}_{\text{global}}$ with $N_{\bar{\Omega}} = 1$, $\bar{\Omega}_1 = \Omega$, and $\bar{\Gamma}_1 = \Gamma$

Fig. 1. Examples of decomposed meshes $\bar{\mathcal{M}}$ for a vertex-centered finite-volume model. Solid lines denote the primal mesh, and dashed lines the control-volume interfaces Γ_j defining the dual mesh, and colors denote separate subdomains $\bar{\Omega}_i$. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

3.4. Lack of conservation

Remarks 4 and 5 demonstrated that Galerkin and LSPG ROMs minimize the violation of conservation in the case of finite-volume models in particular senses; Galerkin performs this minimization at the time-continuous level, while LSPG does so at the time-discrete level. While this is an attractive property, it does not guarantee that the model is conservative in any sense: because the minimum value of the objective functions in Eqs. (3.4) and (3.7) may be non-zero, conservation is generally violated by each of these approaches. We interpret this as violating the structure intrinsic to finite-volume models. This provides the motivation for this work: we aim to develop reduced-order models that ensure the resulting model is conservative globally and—more generally—over subdomains.

4. Proposed method

This section describes the proposed method, which equips the optimization problems characterizing the online ROM solution with equality constraints that explicitly enforce conservation over subdomains. The approach requires no modification to the offline stage except when hyper-reduction is applied to the nonlinear terms appearing in the constraints. Section 4.1 introduces the concept of conservation over subdomains, Section 4.2 introduces conservative Galerkin projection, Section 4.3 describes conservative LSPG projection, Section 4.4 described approaches for handling infeasibility, and Section 4.5 describes hyper-reduction techniques applicable to objective function and constraint, and Section 4.6 describes snapshot-based (offline) training procedures that may be used for generating the reduced-basis matrices required by the method.

4.1. Domain decomposition

To begin, we decompose the mesh $\mathcal{M}$ into subdomains, each of which comprises the union of control volumes. That is, we define a decomposed mesh $\bar{\mathcal{M}}$ of $N_{\bar{\Omega}}(\leq N_\Omega)$ subdomains $\bar{\Omega}_i = \cup_{j\in\mathcal{K}\subseteq\mathbb{N}(N_\Omega)}\Omega_j$, $i \in \mathbb{N}(N_{\bar{\Omega}})$ with $\bar{\mathcal{M}} := \{\bar{\Omega}_i\}_{i=1}^{N_{\bar{\Omega}}}$. We note that the subdomains need not be non-overlapping, closed, or connected. Denoting the boundary of the ith subdomain by $\bar{\Gamma}_i := \partial\bar{\Omega}_i$, we have $\bar{\Gamma}_i = \{\vec{x}\,|\,\vec{x} \in e,\ \forall e \in \bar{\mathcal{E}}_i,\ i \in \mathbb{N}(|\bar{\mathcal{E}}_i|)\} \subseteq \cup_{j=1}^{N_\Omega}\Gamma_j$, $i \in \mathbb{N}(N_{\bar{\Omega}})$ with $\bar{\mathcal{E}}_i \subseteq \mathcal{E}$ representing the set of faces belonging to the ith subdomain. We denote the full set of faces within the decomposed mesh by $\bar{\mathcal{E}} := \cup_{i=1}^{N_{\bar{\Omega}}}\bar{\mathcal{E}}_i \subseteq \mathcal{E}$. Fig. 1 depicts several decompositions that satisfy the above conditions. We emphasize that the subdomains can overlap, their union need not correspond to the global domain, and the global domain can be considered by employing $\bar{\mathcal{M}} = \bar{\mathcal{M}}_{\text{global}}$, which is characterized by $N_{\bar{\Omega}} = 1$ subdomain that corresponds to the global domain, i.e., $\bar{\Omega}_1 = \Omega$ and $\bar{\Gamma}_1 = \Gamma$, as depicted in Fig. 1c.

Enforcing conservation (2.1) on each subdomain in the decomposed mesh yields

$$\frac{d}{dt}\int\limits_{\bar{\Omega}_j} u_i(\vec{x}, t; \boldsymbol{\mu})\,\mathrm{d}\vec{x} + \int\limits_{\bar{\Gamma}_j} \boldsymbol{g}_i(\vec{x}, t; \boldsymbol{\mu}) \cdot \bar{\boldsymbol{n}}_j(\vec{x})\,\mathrm{d}\vec{s}(\vec{x}) = \int\limits_{\bar{\Omega}_j} s_i(\vec{x}, t; \boldsymbol{\mu})\,\mathrm{d}\vec{x}, \quad i \in \mathbb{N}(n_u),\ j \in \mathbb{N}(N_{\bar{\Omega}}), \tag{4.1}$$

where $\bar{\boldsymbol{n}}_j : \Gamma_j \to \mathbb{R}^d$ denotes the unit normal to subdomain $\bar{\Omega}_j$. We propose applying a finite-volume discretization to Eq. (4.1) that operates on the decomposed mesh $\bar{\mathcal{M}}$. That is, we introduce a 'decomposed' state vector $\bar{\boldsymbol{x}} : \mathbb{R}^N \times [0, T] \times \mathcal{D} \to \mathbb{R}^{\bar{N}}$ with $\bar{N} = N_{\bar{\Omega}} n_u$ and elements

$$\bar{x}_{\bar{\mathcal{I}}(i,j)}(\boldsymbol{x}, t; \boldsymbol{\mu}) = \frac{1}{|\bar{\Omega}_j|}\int\limits_{\bar{\Omega}_j} u_i(\vec{x}, t; \boldsymbol{\mu})\,\mathrm{d}\vec{x}, \quad i \in \mathbb{N}(n_u),\ j \in \mathbb{N}(N_{\bar{\Omega}}), \tag{4.2}$$

where $\bar{\mathcal{I}} : \mathbb{N}(n_u) \times \mathbb{N}(N_{\bar{\Omega}}) \to \mathbb{N}(\bar{N})$ denotes a mapping from conservation-law index and subdomain index to decomposed degree of freedom. The decomposed state vector can be computed from the state vector $\boldsymbol{x}$ as

$$\bar{x}_{\bar{\mathcal{I}}(i,j)}(\boldsymbol{x},t;\boldsymbol{\mu}) = \frac{1}{|\bar{\Omega}_j|}\sum_{k\,|\,\Omega_k\subseteq\bar{\Omega}_j}|\Omega_k|x_{\mathcal{I}(i,k)}(t;\boldsymbol{\mu}) \tag{4.3}$$

or equivalently

$$\bar{\boldsymbol{x}}(\boldsymbol{x}) = \bar{\boldsymbol{C}}\boldsymbol{x}, \tag{4.4}$$

where $\bar{\boldsymbol{C}} \in \mathbb{R}_+^{\bar{N}\times N}$ has elements $\bar{c}_{\bar{\mathcal{I}}(i,j),\mathcal{I}(\ell,k)} = |\Omega_k|/|\bar{\Omega}_j|\delta_{i\ell}I(\Omega_k\subseteq\bar{\Omega}_j)$, where I is the indicator function, which evaluates to one if its argument is true, and zero if its argument is false. We note that this matrix can be decomposed as $\bar{\boldsymbol{C}} = \bar{\boldsymbol{V}}^{-1}\bar{\boldsymbol{E}}\boldsymbol{V}$, where the elements of the volumetic matrices $\boldsymbol{V}\in\mathbb{R}^{N\times N}$, $\bar{\boldsymbol{V}}\in\mathbb{R}^{\bar{N}\times\bar{N}}$ and aggregation matrix $\bar{\boldsymbol{E}}\in\{0,1\}^{\bar{N}\times N}$ comprise

$$v_{\mathcal{I}(i,j),\mathcal{I}(\ell,k)} = \delta_{i\ell}\delta_{jk}\Omega_k,\quad \bar{v}_{\bar{\mathcal{I}}(i,j),\bar{\mathcal{I}}(\ell,k)} = \delta_{i\ell}\delta_{jk}\bar{\Omega}_k,\quad \bar{e}_{\bar{\mathcal{I}}(i,j),\mathcal{I}(\ell,k)} = \delta_{i\ell}I(\Omega_k\subseteq\bar{\Omega}_j). \tag{4.5}$$

Similarly, we write the velocity vector $\bar{\boldsymbol{f}}:(\boldsymbol{\xi},\tau;\boldsymbol{\nu})\mapsto\bar{\boldsymbol{f}}^g(\boldsymbol{\xi},\tau;\boldsymbol{\nu})+\bar{\boldsymbol{f}}^s(\boldsymbol{\xi},\tau;\boldsymbol{\nu})$ with $\bar{\boldsymbol{f}}^g,\bar{\boldsymbol{f}}^s:\mathbb{R}^N\times[0,T]\times\mathcal{D}\to\mathbb{R}^{\bar{N}}$ whose elements consist of

$$\bar{f}^g_{\bar{\mathcal{I}}(i,j)}(\boldsymbol{x},t;\boldsymbol{\mu}) = -\frac{1}{|\bar{\Omega}_j|}\int_{\bar{\Gamma}_j}\boldsymbol{g}_i^{\mathrm{FV}}(\boldsymbol{x};\vec{x},t;\boldsymbol{\mu})\cdot\bar{\boldsymbol{n}}_j(\vec{x})\,\mathrm{d}\vec{s}(\vec{x}) \tag{4.6}$$

$$\bar{f}^s_{\bar{\mathcal{I}}(i,j)}(\boldsymbol{x},t;\boldsymbol{\mu}) = \frac{1}{|\bar{\Omega}_j|}\int_{\bar{\Omega}_j}s_i^{\mathrm{FV}}(\boldsymbol{x};\vec{x},t;\boldsymbol{\mu})\,\mathrm{d}\vec{x}, \tag{4.7}$$

for $i\in\mathbb{N}(n_u)$, $j\in\mathbb{N}(N_{\bar{\Omega}})$, which can be computed from the underlying finite-volume model as

$$\bar{\boldsymbol{f}}^s(\boldsymbol{x},t;\boldsymbol{\mu}) = \bar{\boldsymbol{C}}\boldsymbol{f}^s(\boldsymbol{x},t;\boldsymbol{\mu}),\quad \bar{\boldsymbol{f}}^g(\boldsymbol{x},t;\boldsymbol{\mu}) = \bar{\boldsymbol{B}}\boldsymbol{h}(\boldsymbol{x},t;\boldsymbol{\mu}) \tag{4.8}$$

where the elements of $\bar{\boldsymbol{B}}\in\mathbb{R}^{\bar{N}\times n_uN_e}$ are

$$\bar{b}_{\bar{\mathcal{I}}(i,j),\mathcal{J}(\ell,k)} = \begin{cases} -\delta_{i\ell}/|\bar{\Omega}_j|, & e_k\in\bar{\Gamma}_j;\ \bar{\boldsymbol{n}}_j(\vec{x}) = \boldsymbol{n}_k^e(\vec{x}),\ \vec{x}\in e_k \\ \delta_{i\ell}/|\bar{\Omega}_j|, & e_k\in\bar{\Gamma}_j;\ \bar{\boldsymbol{n}}_j(\vec{x}) = -\boldsymbol{n}_k^e(\vec{x}),\ \vec{x}\in e_k \\ 0, & \text{otherwise.}\end{cases} \tag{4.9}$$

Critically, noting that $\bar{\boldsymbol{B}} = \bar{\boldsymbol{C}}\boldsymbol{B}$ due to the fact that neighboring control volumes have outward unit normals of opposite sign along a shared face, we have

$$\bar{\boldsymbol{f}}^g(\boldsymbol{x},t;\boldsymbol{\mu}) = \bar{\boldsymbol{C}}\boldsymbol{f}^g(\boldsymbol{x},t;\boldsymbol{\mu}) \tag{4.10}$$

such that

$$\bar{\boldsymbol{f}}(\boldsymbol{x},t;\boldsymbol{\mu}) = \bar{\boldsymbol{C}}\boldsymbol{f}(\boldsymbol{x},t;\boldsymbol{\mu}). \tag{4.11}$$

Thus, conservation on the decomposed mesh $\bar{\mathcal{M}}$ given an underlying finite-volume discretization on mesh $\mathcal{M}$ can be expressed as

$$\bar{\boldsymbol{C}}\frac{d\boldsymbol{x}}{dt} = \bar{\boldsymbol{C}}\boldsymbol{f}(\boldsymbol{x},t;\boldsymbol{\mu}) \tag{4.12}$$

or equivalently

$$\bar{\boldsymbol{C}}\boldsymbol{r}(\frac{d\boldsymbol{x}}{dt},\boldsymbol{x},t;\boldsymbol{\mu}) = \boldsymbol{0}. \tag{4.13}$$

Applying a linear multistep scheme to discretize (4.12) in time yields

$$\bar{\boldsymbol{C}}\boldsymbol{r}^n(\boldsymbol{x}^n;\boldsymbol{\mu}) = \boldsymbol{0}. \tag{4.14}$$

Note that the decomposed ODE (4.12) and decomposed OΔE (4.14) are underdetermined, as they comprise $\bar{N}$ equations in $N(\geq\bar{N})$ unknowns.

We now demonstrate that conservation that is enforced over a decomposed mesh automatically leads to conservation over a coarser mesh that embeds the decomposed mesh.

Theorem 4.1 *(Conservation over coarser decompositions). Define a decomposed mesh $\bar{\bar{\mathcal{M}}}$ as a decomposition of a non-overlapping decomposed mesh $\bar{\mathcal{M}}$ satisfying* $\mathrm{meas}(\bar{\Omega}_i \cap \bar{\Omega}_j) = 0$ *for $i \neq j$ such that $\bar{\bar{\Omega}}_i = \cup_{j\in\bar{\mathcal{K}}\subseteq\mathbb{N}(N_{\bar{\Omega}})}\bar{\Omega}_j$, $i \in \mathbb{N}(N_{\bar{\bar{\Omega}}})$, with $\bar{\bar{\mathcal{M}}} := \{\bar{\bar{\Omega}}_i\}_{i=1}^{N_{\bar{\bar{\Omega}}}}$ and $N_{\bar{\bar{\Omega}}} \leq N_{\bar{\Omega}} (\leq N_\Omega)$. Then, satisfaction of time-continuous conservation on $\bar{\mathcal{M}}$ (i.e., Eq. (4.12)) implies satisfaction of time-continuous conservation on $\bar{\bar{\mathcal{M}}}$, i.e.,*

$$\bar{\bar{\boldsymbol{C}}}\frac{d\boldsymbol{x}}{dt} = \bar{\bar{\boldsymbol{C}}}\boldsymbol{f}(\boldsymbol{x}, t; \boldsymbol{\mu}) \tag{4.15}$$

and satisfaction of time-discrete conservation on $\bar{\mathcal{M}}$ (i.e., Eq. (4.14)) implies satisfaction of time-discrete conservation on $\bar{\bar{\mathcal{M}}}$, i.e.,

$$\bar{\bar{\boldsymbol{C}}}\boldsymbol{r}^n(\boldsymbol{x}^n; \boldsymbol{\mu}) = 0, \tag{4.16}$$

where $\bar{\bar{\boldsymbol{C}}} := \bar{\bar{\boldsymbol{V}}}^{-1}\bar{\bar{\boldsymbol{E}}}\boldsymbol{V} \in \mathbb{R}_+^{\bar{\bar{N}}\times N}$, the elements of $\bar{\bar{\boldsymbol{V}}} \in \mathbb{R}^{\bar{\bar{N}}\times\bar{\bar{N}}}$ are $\bar{\bar{v}}_{\bar{\bar{\mathcal{I}}}(i,j),\bar{\bar{\mathcal{I}}}(\ell,k)} = \delta_{i\ell}\delta_{jk}\bar{\bar{\Omega}}_k$, and the elements of $\bar{\bar{\boldsymbol{E}}}$ are $\bar{\bar{e}}_{\bar{\bar{\mathcal{I}}}(i,j),\mathcal{I}(\ell,k)} = \delta_{i\ell}I(\Omega_k \subseteq \bar{\bar{\Omega}}_j)$.

Proof. The conditions $\bar{\bar{\Omega}}_i = \cup_{j\in\bar{\mathcal{K}}\subseteq\mathbb{N}(N_{\bar{\Omega}})}\bar{\Omega}_j$, $i \in \mathbb{N}(N_{\bar{\bar{\Omega}}})$; $\bar{\Omega}_i = \cup_{j\in\mathcal{K}\subseteq\mathbb{N}(N_\Omega)}\Omega_j$, $i \in \mathbb{N}(N_{\bar{\Omega}})$; and $\mathrm{meas}(\bar{\Omega}_i \cap \bar{\Omega}_j) = 0$ for $i \neq j$ imply that the aggregation operator characterizing the mesh $\bar{\bar{\mathcal{M}}}$ can be applied in two stages, i.e.,

$$\bar{\bar{e}}_{\bar{\bar{\mathcal{I}}}(i,j),\mathcal{I}(\ell,k)} = \delta_{i\ell}I(\Omega_k \subseteq \bar{\bar{\Omega}}_j) = \delta_{i\ell}\sum_{m=1}^{N_{\bar{\Omega}}} I(\bar{\Omega}_m \subseteq \bar{\bar{\Omega}}_j)I(\Omega_k \subseteq \bar{\Omega}_m) = \delta_{i\ell}\delta_{\ell h}\sum_{m=1}^{N_{\bar{\Omega}}}\sum_{\ell=1}^{n_u}\bar{\bar{e}}'_{\bar{\bar{\mathcal{I}}}(i,j),\bar{\mathcal{I}}(\ell,m)}\bar{e}_{\bar{\mathcal{I}}(\ell,m),\mathcal{I}(h,k)}$$

and thus $\bar{\bar{\boldsymbol{E}}} = \bar{\bar{\boldsymbol{E}}}'\bar{\boldsymbol{E}}$, where the elements of $\bar{\bar{\boldsymbol{E}}}'$ are $\bar{\bar{e}}'_{\bar{\bar{\mathcal{I}}}(i,j),\bar{\mathcal{I}}(\ell,k)} = \delta_{i\ell}I(\bar{\Omega}_k \subseteq \bar{\bar{\Omega}}_j)$. Substituting $\bar{\bar{\boldsymbol{E}}} = \bar{\bar{\boldsymbol{E}}}'\bar{\boldsymbol{E}}$ in the definition of $\bar{\bar{\boldsymbol{C}}}$ yields

$$\bar{\bar{\boldsymbol{C}}} = \bar{\bar{\boldsymbol{V}}}^{-1}\bar{\bar{\boldsymbol{E}}}'\bar{\boldsymbol{E}}\boldsymbol{V} = \bar{\bar{\boldsymbol{V}}}^{-1}\bar{\bar{\boldsymbol{E}}}'\bar{\boldsymbol{V}}\bar{\boldsymbol{C}}. \tag{4.17}$$

Thus, Eqs. (4.15) and (4.16) can be rewritten as

$$\bar{\bar{\boldsymbol{V}}}^{-1}\bar{\bar{\boldsymbol{E}}}'\bar{\boldsymbol{V}}\bar{\boldsymbol{C}}\frac{d\boldsymbol{x}}{dt} = \bar{\bar{\boldsymbol{V}}}^{-1}\bar{\bar{\boldsymbol{E}}}'\bar{\boldsymbol{V}}\bar{\boldsymbol{C}}\boldsymbol{f}(\boldsymbol{x}, t; \boldsymbol{\mu}) \tag{4.18}$$

$$\bar{\bar{\boldsymbol{V}}}^{-1}\bar{\bar{\boldsymbol{E}}}'\bar{\boldsymbol{V}}\bar{\boldsymbol{C}}\boldsymbol{r}^n(\boldsymbol{x}^n; \boldsymbol{\mu}) = 0, \tag{4.19}$$

which are clearly satisfied if Eqs. (4.12) and (4.14) are satisfied, respectively. □

Corollary 4.1 *(Full-order model conservation). The full-order model satisfies time-continuous and time-discrete conservation over any decomposed mesh.*

Proof. This corresponds to a particular case of Theorem 4.1 with $\bar{\mathcal{M}} = \mathcal{M}$, as any decomposed mesh $\bar{\bar{\mathcal{M}}}$ must satisfy $\bar{\bar{\Omega}}_i = \cup_{j\in\mathcal{K}\subseteq\mathbb{N}(n_u)}\Omega_j$, $i \in \mathbb{N}(N_{\bar{\bar{\Omega}}})$ and the original mesh is non-overlapping, i.e., $\mathrm{meas}(\Omega_i \cap \Omega_j) = 0$ for $i \neq j$. □

Corollary 4.2 *(Global conservation). If the decomposed mesh $\bar{\mathcal{M}}$ satisfies $\cup_{i=1}^{N_{\bar{\Omega}}}\bar{\Omega}_i = \Omega$ and is non-overlapping, i.e.,* $\mathrm{meas}(\bar{\Omega}_i \cap \bar{\Omega}_j) = 0$ *for $i \neq j$, then satisfaction of time-continuous conservation on $\bar{\mathcal{M}}$ implies satisfaction of time-continuous (global) conservation on $\bar{\mathcal{M}}_{global} := \{\Omega\}$, and satisfaction of time-discrete conservation on $\bar{\mathcal{M}}$ implies satisfaction of time-discrete (global) conservation on $\bar{\mathcal{M}}_{global}$.*

Proof. This corresponds to a particular case of Theorem 4.1 with $\bar{\bar{\mathcal{M}}} = \bar{\mathcal{M}}_{\mathrm{global}}$, as the required condition $\bar{\bar{\Omega}}_i = \cup_{j\in\bar{\mathcal{K}}\subseteq\mathbb{N}(N_{\bar{\Omega}})}\bar{\Omega}_j$, $i \in \mathbb{N}(N_{\bar{\bar{\Omega}}})$ is satisfied for $N_{\bar{\bar{\Omega}}} = 1$, $\bar{\bar{\Omega}}_1 = \Omega$, and $\bar{\mathcal{K}} = \mathbb{N}(N_{\bar{\Omega}})$ under the stated assumptions. □

We now derive the proposed conservative Galerkin and conservative LSPG projection techniques, which equip their associated optimization problems with equality constraints that enforce conservation over the decomposed mesh $\bar{\mathcal{M}}$.

4.2. Conservative Galerkin projection

To enable a Galerkin-like projection scheme that enforces conservation, we equip the unconstrained optimization problems (3.4)–(3.5)—which are defined at the time-continuous level—with equality constraints corresponding to (time-continuous) conservation (4.13) over the decomposed mesh $\bar{\mathcal{M}}$. The resulting conservative Galerkin solution $\frac{d\tilde{\boldsymbol{x}}}{dt}\left(\boldsymbol{x}^0(\boldsymbol{\mu}) + \boldsymbol{\Phi}\hat{\boldsymbol{x}}, t; \boldsymbol{\mu}\right)$ satisfies

$$\begin{aligned}&\underset{\boldsymbol{v}\in \mathrm{Ran}(\boldsymbol{\Phi})}{\text{minimize}}\ \|\boldsymbol{r}(\boldsymbol{v},\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}},t;\boldsymbol{\mu})\|_2\\&\text{subject to } \bar{\boldsymbol{C}}\boldsymbol{r}(\boldsymbol{v},\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}},t;\boldsymbol{\mu})=\boldsymbol{0}.\end{aligned}\tag{4.20}$$

Equivalently, the conservative Galerkin generalized coordinates $\frac{d\hat{\boldsymbol{x}}}{dt}\left(\boldsymbol{x}^0(\boldsymbol{\mu})+\boldsymbol{\Phi}\hat{\boldsymbol{x}},t;\boldsymbol{\mu}\right)$ satisfy

$$\begin{aligned}&\underset{\hat{\boldsymbol{v}}\in \mathbb{R}^p}{\text{minimize}}\ \|\boldsymbol{r}(\boldsymbol{\Phi}\hat{\boldsymbol{v}},\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}},t;\boldsymbol{\mu})\|_2\\&\text{subject to } \bar{\boldsymbol{C}}\boldsymbol{r}(\boldsymbol{\Phi}\hat{\boldsymbol{v}},\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}},t;\boldsymbol{\mu})=\boldsymbol{0}.\end{aligned}\tag{4.21}$$

We now provide a finite-volume interpretation of the conservative Galerkin model, define the feasible set, and provide an algebraic description of the solution.

Remark 6 *(Conservative Galerkin ROM ODE: interpretation).* From Remark 4, the conservative Galerkin ROM ODE (4.20) can be interpreted as minimizing the sum of squared (normalized) *rates of violation of conservation* across all variables u_i, $i\in\mathbb{N}(n_u)$ and control volumes Ω_j, $j\in\mathbb{N}(N_\Omega)$ at time instance t subject to the enforcement of conservation of all variables u_i, $i\in\mathbb{N}(n_u)$ over subdomains $\bar{\Omega}_j$, $j\in\mathbb{N}(N_{\bar{\Omega}})$ at time instance t under one approximation: the flux and source terms are approximated using the finite-volume discretization (i.e., $\boldsymbol{g}_i\leftarrow\boldsymbol{g}_i^{\mathrm{FV}}$, and $s_i\leftarrow s_i^{\mathrm{FV}}$).

Definition 1 *(Feasibility of conservative Galerkin projection).* Problem (4.21) is feasible if the Galerkin feasible set $\mathcal{F}_{\mathrm{G}}(\boldsymbol{x}^0(\boldsymbol{\mu})+\boldsymbol{\Phi}\hat{\boldsymbol{x}},t;\boldsymbol{\mu})$, defined as

$$\mathcal{F}_{\mathrm{G}}(\boldsymbol{\xi},\tau;\boldsymbol{\nu}):=\{\boldsymbol{w}\in\mathbb{R}^p\,|\,\bar{\boldsymbol{C}}\boldsymbol{r}(\boldsymbol{\Phi}\boldsymbol{w},\boldsymbol{\xi},\tau;\boldsymbol{\nu})=\boldsymbol{0}\},\tag{4.22}$$

is non-empty.

Theorem 4.2. *If Problem (4.21) is feasible, then the solution is unique and satisfies the time-dependent saddle-point problem*

$$\begin{bmatrix}\boldsymbol{I} & \boldsymbol{\Phi}^T\bar{\boldsymbol{C}}^T\\ \bar{\boldsymbol{C}}\boldsymbol{\Phi} & \boldsymbol{0}\end{bmatrix}\begin{bmatrix}\frac{d\hat{\boldsymbol{x}}}{dt}\\ \frac{d\boldsymbol{\lambda}_G}{dt}\end{bmatrix}=\begin{bmatrix}\boldsymbol{\Phi}^T\boldsymbol{f}(\boldsymbol{x}^0(\boldsymbol{\mu})+\boldsymbol{\Phi}\hat{\boldsymbol{x}},t;\boldsymbol{\mu})\\ \bar{\boldsymbol{C}}\boldsymbol{f}(\boldsymbol{x}^0(\boldsymbol{\mu})+\boldsymbol{\Phi}\hat{\boldsymbol{x}},t;\boldsymbol{\mu})\end{bmatrix},\quad \hat{\boldsymbol{x}}(0)=\boldsymbol{0},\tag{4.23}$$

which can be expressed equivalently as

$$\frac{d\hat{\boldsymbol{x}}}{dt}=\boldsymbol{\Phi}^T\boldsymbol{f}(\boldsymbol{x}^0(\boldsymbol{\mu})+\boldsymbol{\Phi}\hat{\boldsymbol{x}},t;\boldsymbol{\mu})+\delta\boldsymbol{f}_{\mathrm{G}}(\boldsymbol{x}^0(\boldsymbol{\mu})+\boldsymbol{\Phi}\hat{\boldsymbol{x}},t;\boldsymbol{\mu}),\quad \hat{\boldsymbol{x}}(0)=\boldsymbol{0}\tag{4.24}$$

with

$$\delta\boldsymbol{f}_{\mathrm{G}}(\boldsymbol{\xi},\tau;\boldsymbol{\nu}):=(\bar{\boldsymbol{C}}\boldsymbol{\Phi})^{+}[\bar{\boldsymbol{C}}\boldsymbol{f}(\boldsymbol{\xi},\tau;\boldsymbol{\nu})-\bar{\boldsymbol{C}}\boldsymbol{\Phi}\boldsymbol{\Phi}^T\boldsymbol{f}(\boldsymbol{\xi},\tau;\boldsymbol{\nu})]\tag{4.25}$$

and Lagrange multipliers

$$\frac{d\boldsymbol{\lambda}_G}{dt}=-(\bar{\boldsymbol{C}}\boldsymbol{\Phi})^{+T}(\bar{\boldsymbol{C}}\boldsymbol{\Phi})^{+}[\bar{\boldsymbol{C}}\boldsymbol{f}(\boldsymbol{\xi},\tau;\boldsymbol{\nu})-\bar{\boldsymbol{C}}\boldsymbol{\Phi}\boldsymbol{\Phi}^T\boldsymbol{f}(\boldsymbol{\xi},\tau;\boldsymbol{\nu})].\tag{4.26}$$

Proof. The Lagrangian associated with problem (4.21) can be written as

$$\mathcal{L}_G(\hat{\boldsymbol{v}},\boldsymbol{\gamma},t;\boldsymbol{\mu}):=\frac{1}{2}\|\boldsymbol{\Phi}\hat{\boldsymbol{v}}-\boldsymbol{f}(\boldsymbol{x}^0(\boldsymbol{\mu})+\boldsymbol{\Phi}\hat{\boldsymbol{x}},t;\boldsymbol{\mu})\|_2^2+\boldsymbol{\gamma}^T\left[\bar{\boldsymbol{C}}\boldsymbol{\Phi}\hat{\boldsymbol{v}}-\bar{\boldsymbol{C}}\boldsymbol{f}(\boldsymbol{x}^0(\boldsymbol{\mu})+\boldsymbol{\Phi}\hat{\boldsymbol{x}},t;\boldsymbol{\mu})\right].\tag{4.27}$$

We note that problem (4.21) corresponds to a convex linear least-squares problem with linear equality constraints; thus, $(\frac{d\hat{\boldsymbol{x}}}{dt},\frac{d\boldsymbol{\lambda}_{\mathrm{G}}}{dt})$ is a unique solution if and only if it satisfies the stationarity conditions

$$\frac{\partial\mathcal{L}_G}{\partial\hat{\boldsymbol{v}}}\left(\frac{d\hat{\boldsymbol{x}}}{dt},\frac{d\boldsymbol{\lambda}_{\mathrm{G}}}{dt},t;\boldsymbol{\mu}\right)=\boldsymbol{0},\quad\frac{\partial\mathcal{L}_G}{\partial\boldsymbol{\gamma}}\left(\frac{d\hat{\boldsymbol{x}}}{dt},\frac{d\boldsymbol{\lambda}_{\mathrm{G}}}{dt},t;\boldsymbol{\mu}\right)=\boldsymbol{0}.$$

Noting that $\boldsymbol{\Phi}^T\boldsymbol{\Phi}=\boldsymbol{I}$, these conditions are equivalent to Eq. (4.23). The proof of Eqs. (4.24)–(4.26) follows the null-space method for solving optimization problems with linear equality constraints. Feasibility implies that the feasible set $\mathcal{F}_{\mathrm{G}}(\boldsymbol{x}^0(\boldsymbol{\mu})+\boldsymbol{\Phi}\hat{\boldsymbol{x}},t;\boldsymbol{\mu})$ is non-empty, which in turn implies that the second block of Eqs. (4.23) is consistent and $\bar{\boldsymbol{C}}\boldsymbol{f}(\boldsymbol{\xi},\tau;\boldsymbol{\nu})\in\mathrm{Ran}\left(\bar{\boldsymbol{C}}\boldsymbol{\Phi}\right)$ or equivalently

$$\bar{\boldsymbol{C}}\boldsymbol{f}(\boldsymbol{\xi},\tau;\boldsymbol{\nu})=\boldsymbol{U}_{\mathrm{G}}\hat{\hat{\boldsymbol{f}}}(\boldsymbol{\xi},\tau;\boldsymbol{\nu}),\tag{4.28}$$

with $\hat{\hat{\boldsymbol{f}}}:\mathbb{R}^N\times[0,T]\times\mathcal{D}\to\mathbb{R}^{\mathrm{rank}(\bar{\boldsymbol{C}}\boldsymbol{\Phi})}$, where

$$\bar{\boldsymbol{C}}\boldsymbol{\Phi}=\boldsymbol{U}_{\mathrm{G}}\boldsymbol{\Sigma}_{\mathrm{G}}\boldsymbol{V}_{\mathrm{G}}^T\tag{4.29}$$

is the singular value decomposition with $\boldsymbol{U}_{\mathrm{G}} \in \mathbb{R}^{\bar{N}\times\mathrm{rank}(\bar{\boldsymbol{C}}\boldsymbol{\Phi})}$ and $\boldsymbol{U}_{\mathrm{G}}^T\boldsymbol{U}_{\mathrm{G}} = \boldsymbol{I}$, $\boldsymbol{\Sigma}_{\mathrm{G}} \equiv \mathrm{diag}\left(\sigma_1, \ldots, \sigma_{\mathrm{rank}(\bar{\boldsymbol{C}}\boldsymbol{\Phi})}\right)$ and $\sigma_1 \geq \cdots \geq \sigma_{\mathrm{rank}(\bar{\boldsymbol{C}}\boldsymbol{\Phi})} > 0$, and $\boldsymbol{V}_{\mathrm{G}} \in \mathbb{R}^{p\times\mathrm{rank}(\bar{\boldsymbol{C}}\boldsymbol{\Phi})}$ with $\boldsymbol{V}_{\mathrm{G}}^T\boldsymbol{V}_{\mathrm{G}} = \boldsymbol{I}$. Because $\mathrm{Ran}\,(\boldsymbol{V}_{\mathrm{G}}) \oplus \mathrm{Ran}\,(\boldsymbol{Z}_{\mathrm{G}}) = \mathbb{R}^p$ with $\boldsymbol{Z}_{\mathrm{G}} \in \mathbb{R}^{p\times(p-\mathrm{rank}(\bar{\boldsymbol{C}}\boldsymbol{\Phi}))}$ an orthogonal basis for the null space of $\bar{\boldsymbol{C}}\boldsymbol{\Phi}$, we can decompose the unknown vector $\hat{\boldsymbol{v}} \in \mathbb{R}^p$ appearing in optimization problem (4.21) as

$$\hat{\boldsymbol{v}} = \boldsymbol{V}_{\mathrm{G}}\hat{\boldsymbol{v}}_1 + \boldsymbol{Z}_{\mathrm{G}}\hat{\boldsymbol{v}}_2 \tag{4.30}$$

with $\hat{\boldsymbol{v}}_1 \in \mathbb{R}^{\mathrm{rank}(\bar{\boldsymbol{C}}\boldsymbol{\Phi})}$ and $\hat{\boldsymbol{v}}_2 \in \mathbb{R}^{p-\mathrm{rank}(\bar{\boldsymbol{C}}\boldsymbol{\Phi})}$. Substituting Eqs. (4.28), (4.29), and (4.30) into the constraints of Problem (4.21) and noting that $\boldsymbol{V}_{\mathrm{G}}^T\boldsymbol{Z}_{\mathrm{G}} = \boldsymbol{0}$ yields

$$\hat{\boldsymbol{v}}_1 = \boldsymbol{\Sigma}_{\mathrm{G}}^{-1}\boldsymbol{U}_{\mathrm{G}}^T\bar{\boldsymbol{C}}\boldsymbol{f}(\boldsymbol{x}^0(\boldsymbol{\mu}) + \boldsymbol{\Phi}\hat{\boldsymbol{x}}, t; \boldsymbol{\mu}). \tag{4.31}$$

Pre-multiplying (4.31) by $\boldsymbol{V}_{\mathrm{G}}$, substituting $(\bar{\boldsymbol{C}}\boldsymbol{\Phi})^+ = \boldsymbol{V}_{\mathrm{G}}\boldsymbol{\Sigma}_{\mathrm{G}}^{-1}\boldsymbol{U}_{\mathrm{G}}^T$, and using (4.28) yields

$$\boldsymbol{V}_{\mathrm{G}}\hat{\boldsymbol{v}}_1 = (\bar{\boldsymbol{C}}\boldsymbol{\Phi})^+\bar{\boldsymbol{C}}\boldsymbol{f}(\boldsymbol{x}^0(\boldsymbol{\mu}) + \boldsymbol{\Phi}\hat{\boldsymbol{x}}, t; \boldsymbol{\mu}). \tag{4.32}$$

Thus, decomposing the solution as

$$\frac{d\hat{\boldsymbol{x}}}{dt} = \boldsymbol{V}_{\mathrm{G}}\left[\frac{d\hat{\boldsymbol{x}}}{dt}\right]_1 + \boldsymbol{Z}_{\mathrm{G}}\left[\frac{d\hat{\boldsymbol{x}}}{dt}\right]_2, \tag{4.33}$$

we have

$$\boldsymbol{V}_{\mathrm{G}}\left[\frac{d\hat{\boldsymbol{x}}}{dt}\right]_1 = (\bar{\boldsymbol{C}}\boldsymbol{\Phi})^+\bar{\boldsymbol{C}}\boldsymbol{f}(\boldsymbol{x}^0(\boldsymbol{\mu}) + \boldsymbol{\Phi}\hat{\boldsymbol{x}}, t; \boldsymbol{\mu}). \tag{4.34}$$

Now, substituting Eqs. (4.30) with $\hat{\boldsymbol{v}}_1$ defined in (4.32) into Problem (4.21) yields an unconstrained optimization problem in $\hat{\boldsymbol{v}}_2$ only, i.e., $\left[\frac{d\hat{\boldsymbol{x}}}{dt}\right]_2$ is the solution to

$$\underset{\hat{\boldsymbol{v}}_2\in\mathbb{R}^{p-\mathrm{rank}(\bar{\boldsymbol{C}}\boldsymbol{\Phi})}}{\mathrm{minimize}} \; \|\boldsymbol{\Phi}[(\bar{\boldsymbol{C}}\boldsymbol{\Phi})^+\bar{\boldsymbol{C}}\boldsymbol{f}(\boldsymbol{x}^0(\boldsymbol{\mu}) + \boldsymbol{\Phi}\hat{\boldsymbol{x}}, t; \boldsymbol{\mu}) + \boldsymbol{Z}_{\mathrm{G}}\hat{\boldsymbol{v}}_2] - \boldsymbol{f}(\boldsymbol{x}^0(\boldsymbol{\mu}) + \boldsymbol{\Phi}\hat{\boldsymbol{x}}, t; \boldsymbol{\mu})\|_2, \tag{4.35}$$

which—using orthogonality of $\boldsymbol{\Phi}\boldsymbol{Z}_{\mathrm{G}}$—is simply

$$\left[\frac{d\hat{\boldsymbol{x}}}{dt}\right]_2 = \boldsymbol{Z}_{\mathrm{G}}^T\boldsymbol{\Phi}^T[\boldsymbol{f}(\boldsymbol{x}^0(\boldsymbol{\mu}) + \boldsymbol{\Phi}\hat{\boldsymbol{x}}, t; \boldsymbol{\mu}) - \boldsymbol{\Phi}(\bar{\boldsymbol{C}}\boldsymbol{\Phi})^+\bar{\boldsymbol{C}}\boldsymbol{f}(\boldsymbol{x}^0(\boldsymbol{\mu}) + \boldsymbol{\Phi}\hat{\boldsymbol{x}}, t; \boldsymbol{\mu})]. \tag{4.36}$$

Applying Eqs. (4.34), and (4.36) to Eq. (4.33) yields

$$\frac{d\hat{\boldsymbol{x}}}{dt} = (\bar{\boldsymbol{C}}\boldsymbol{\Phi})^+\bar{\boldsymbol{C}}\boldsymbol{f}(\boldsymbol{x}^0(\boldsymbol{\mu}) + \boldsymbol{\Phi}\hat{\boldsymbol{x}}, t; \boldsymbol{\mu}) + \boldsymbol{Z}_{\mathrm{G}}\boldsymbol{Z}_{\mathrm{G}}^T[\boldsymbol{\Phi}^T\boldsymbol{f}(\boldsymbol{x}^0(\boldsymbol{\mu}) + \boldsymbol{\Phi}\hat{\boldsymbol{x}}, t; \boldsymbol{\mu}) - (\bar{\boldsymbol{C}}\boldsymbol{\Phi})^+\bar{\boldsymbol{C}}\boldsymbol{f}(\boldsymbol{x}^0(\boldsymbol{\mu}) + \boldsymbol{\Phi}\hat{\boldsymbol{x}}, t; \boldsymbol{\mu})]. \tag{4.37}$$

Applying $\boldsymbol{Z}_{\mathrm{G}}^T(\bar{\boldsymbol{C}}\boldsymbol{\Phi})^+ = \boldsymbol{0}$ to Eq. (4.37) and observing that $\boldsymbol{Z}_{\mathrm{G}}\boldsymbol{Z}_{\mathrm{G}}^T\hat{\boldsymbol{v}} = (\boldsymbol{I} - \boldsymbol{V}_{\mathrm{G}}\boldsymbol{V}_{\mathrm{G}}^T)\hat{\boldsymbol{v}} = (\boldsymbol{I} - (\bar{\boldsymbol{C}}\boldsymbol{\Phi})^+\bar{\boldsymbol{C}}\boldsymbol{\Phi})\hat{\boldsymbol{v}}$ yields Eq. (4.24). Finally, Eq. (4.26) arises from substituting (4.24) into (4.23). □

Comparing Eqs. (3.2) and (4.24) reveals that equipping the Galerkin-ROM optimization problem with equality constraints associated with conservation has the effect of modifying the velocity vector through the addition of the term $\delta\boldsymbol{f}_{\mathrm{G}}$ defined in Eq. (4.25). Note that Eq. (4.24) corresponds to an initial-value problem that can be integrated in time, e.g., using a linear multistep method.

We now show that the conservative Galerkin velocity can be expressed as the orthogonal projection of the standard Galerkin velocity onto the feasible set.

Corollary 4.3. *If Problem* (4.20) *is feasible, then the solution corresponds to the orthogonal projection of the standard Galerkin velocity* (3.2) *onto the feasible set, i.e.,*

$$\frac{d\hat{\boldsymbol{x}}}{dt}\left(\boldsymbol{x}^0(\boldsymbol{\mu}) + \boldsymbol{\Phi}\hat{\boldsymbol{x}}, t; \boldsymbol{\mu}\right) = \underset{\boldsymbol{v}\in\mathcal{F}_{\mathrm{G}}(\boldsymbol{x}^0(\boldsymbol{\mu})+\boldsymbol{\Phi}\hat{\boldsymbol{x}},t;\boldsymbol{\mu})}{\arg\min} \|\boldsymbol{v} - \boldsymbol{\Phi}^T\boldsymbol{f}(\boldsymbol{x}^0(\boldsymbol{\mu}) + \boldsymbol{\Phi}\hat{\boldsymbol{x}}, t; \boldsymbol{\mu})\|_2. \tag{4.38}$$

Proof. We first identify the feasible set from Eqs. (4.33) and (4.34) as

$$\mathcal{F}_{\mathrm{G}}(\boldsymbol{x}^0(\boldsymbol{\mu}) + \boldsymbol{\Phi}\hat{\boldsymbol{x}}, t; \boldsymbol{\mu}) = (\bar{\boldsymbol{C}}\boldsymbol{\Phi})^+\bar{\boldsymbol{C}}\boldsymbol{f}(\boldsymbol{x}^0(\boldsymbol{\mu}) + \boldsymbol{\Phi}\hat{\boldsymbol{x}}, t; \boldsymbol{\mu}) + \mathrm{Ran}\,(\boldsymbol{Z}_{\mathrm{G}}). \tag{4.39}$$

Noting that the orthogonal projection of a vector $\boldsymbol{\xi}$ onto an affine subspace $\bar{\boldsymbol{\xi}} + \mathrm{Ran}\,(\boldsymbol{Q})$ with $\boldsymbol{Q}$ an orthogonal matrix with more rows than columns is simply $\bar{\boldsymbol{\xi}} + \boldsymbol{Q}\,\boldsymbol{Q}^T(\xi - \bar{\xi})$, we identify $\frac{d\hat{\boldsymbol{x}}}{dt}$ as the orthogonal projection of the standard Galerkin velocity $\boldsymbol{\Phi}^T\boldsymbol{f}(\boldsymbol{x}^0(\boldsymbol{\mu}) + \boldsymbol{\Phi}\hat{\boldsymbol{x}}, t; \boldsymbol{\mu})$ onto the affine subspace corresponding to the feasible set $\mathcal{F}_{\mathrm{G}}(\boldsymbol{x}^0(\boldsymbol{\mu}) + \boldsymbol{\Phi}\hat{\boldsymbol{x}}, t; \boldsymbol{\mu})$; Eq. (4.38) derives from this result and the optimality property of orthogonal projectors. □

Of course, numerically solving the conservative Galerkin ROM ODE, requires introducing a time integrator. Applying a linear multistep scheme to solve Eq. (4.23) characterizing the conservative Galerkin ROM ODE yields at time instance n yields the conservative Galerkin ROM OΔE

$$\begin{aligned}\sum_{j=0}^{k}\alpha_j\hat{\boldsymbol{x}}^{n-j}+\sum_{j=0}^{k}\alpha_j\boldsymbol{\Phi}^T\bar{\boldsymbol{C}}^T\boldsymbol{\lambda}_{\mathrm{G}}^{n-j}&=\Delta t\sum_{j=0}^{k}\beta_j\boldsymbol{\Phi}^T\boldsymbol{f}(\boldsymbol{x}^0(\boldsymbol{\mu})+\boldsymbol{\Phi}\hat{\boldsymbol{x}}^{n-j},t;\boldsymbol{\mu})\\ \sum_{j=0}^{k}\alpha_j\bar{\boldsymbol{C}}\boldsymbol{\Phi}\hat{\boldsymbol{x}}^{n-j}&=\Delta t\sum_{j=0}^{k}\beta_j\bar{\boldsymbol{C}}\boldsymbol{f}(\boldsymbol{x}^0(\boldsymbol{\mu})+\boldsymbol{\Phi}\hat{\boldsymbol{x}}^{n-j},t;\boldsymbol{\mu}).\end{aligned}\tag{4.40}$$

We now demonstrate that conservative Galerkin projection and time discretization are commutative.

Theorem 4.3 *(Commutativity of conservative Galerkin projection and time discretization). Conservative Galerkin projection is equivalent to computing an approximate solution $(\tilde{\boldsymbol{x}}(t;\boldsymbol{\mu}),\boldsymbol{\lambda}_G(t;\boldsymbol{\mu}))\in\boldsymbol{x}^0(\boldsymbol{\mu})+\mathrm{Ran}(\boldsymbol{\Phi})\times\mathbb{R}^{\bar{N}}$ via Galerkin projection applied to the system*

$$\begin{bmatrix}\boldsymbol{I} & \bar{\boldsymbol{C}}^T\\ \bar{\boldsymbol{C}} & \boldsymbol{0}\end{bmatrix}\begin{bmatrix}\frac{d\boldsymbol{x}}{dt}\\ \frac{d\boldsymbol{\lambda}_G}{dt}\end{bmatrix}=\begin{bmatrix}\boldsymbol{f}(\boldsymbol{x},t;\boldsymbol{\mu})\\ \bar{\boldsymbol{C}}\boldsymbol{f}(\boldsymbol{x},t;\boldsymbol{\mu})\end{bmatrix}.\tag{4.41}$$

Further, performing conservative Galerkin projection on Eq. (4.41) and subsequently applying time discretization yields the same model as first applying time discretization on Eq. (4.41) and subsequently performing conservative Galerkin projection.

Proof. The first part of the theorem can be derived by noticing that substituting Eq. (3.1) in (4.41) and premultiplying by $\begin{bmatrix}\boldsymbol{\Phi} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}\end{bmatrix}$ yields the conservative Galerkin saddle-point system (4.23). Then, applying a linear multistep scheme to solve Eq. (4.23) yields the conservative Galerkin ROM OΔE (4.40) above. Now, applying a linear multistep scheme to integrate (4.41) in time yields

$$\begin{aligned}\sum_{j=0}^{k}\alpha_j\boldsymbol{x}^{n-j}+\sum_{j=0}^{k}\alpha_j\bar{\boldsymbol{C}}^T\boldsymbol{\lambda}_{\mathrm{G}}^{n-j}&=\Delta t\sum_{j=0}^{k}\beta_j\boldsymbol{f}(\boldsymbol{x}^{n-j},t;\boldsymbol{\mu})\\ \sum_{j=0}^{k}\alpha_j\bar{\boldsymbol{C}}\boldsymbol{x}^{n-j}&=\Delta t\sum_{j=0}^{k}\beta_j\bar{\boldsymbol{C}}\boldsymbol{f}(\boldsymbol{x}^{n-j},t;\boldsymbol{\mu}).\end{aligned}\tag{4.42}$$

Because applying conservative Galerkin projection to Eq. (4.42) yields Eq. (4.40), we conclude that conservative Galerkin projection and time discretization are commutative. □

4.3. Conservative LSPG projection

Analogously to the procedure employed to derive the conservative Galerkin ROM, we now equip the unconstrained optimization problem (3.7)–(3.8)—which is defined at the time-discrete level—with equality constraints corresponding to (time-discrete) conservation (4.14) over the decomposed mesh $\bar{\mathcal{M}}$. The resulting conservative LSPG solution $\tilde{\boldsymbol{x}}^n$ satisfies

$$\begin{aligned}&\underset{\boldsymbol{z}\in\boldsymbol{x}^0(\boldsymbol{\mu})+\mathrm{Ran}(\boldsymbol{\Phi})}{\text{minimize}}\quad\|\boldsymbol{r}^n(\boldsymbol{z};\boldsymbol{\mu})\|_2\\ &\quad\text{subject to }\bar{\boldsymbol{C}}\boldsymbol{r}^n(\boldsymbol{z};\boldsymbol{\mu})=0.\end{aligned}\tag{4.43}$$

Equivalently, the conservative LSPG generalized coordinates $\hat{\boldsymbol{x}}^n$ satisfy

$$\begin{aligned}&\underset{\hat{\boldsymbol{z}}\in\mathbb{R}^p}{\text{minimize}}\,\|\boldsymbol{r}^n(\boldsymbol{x}^0(\boldsymbol{\mu})+\boldsymbol{\Phi}\hat{\boldsymbol{z}};\boldsymbol{\mu})\|_2\\ &\text{subject to }\bar{\boldsymbol{C}}\boldsymbol{r}^n(\boldsymbol{x}^0(\boldsymbol{\mu})+\boldsymbol{\Phi}\hat{\boldsymbol{z}};\boldsymbol{\mu})=0.\end{aligned}\tag{4.44}$$

We now provide a finite-volume interpretation of the proposed model, define the feasible set, and provide an algebraic description of the solution.

Remark 7 *(Conservative LSPG ROM OΔE: interpretation for Adams methods).* From Remark 5, the conservative LSPG ROM OΔE (4.43) can be interpreted as minimizing the sum of squared (normalized) *violation of conservation* across all variables u_i, $i\in\mathbb{N}(n_u)$ and control volumes Ω_j, $j\in\mathbb{N}(n_u)$ over time interval $[t^{n-1},t^n]$ subject to the enforcement of conservation of all variables u_i, $i\in\mathbb{N}(n_u)$ over subdomains $\bar{\Omega}_j$, $j\in\mathbb{N}(N_{\bar{\Omega}})$ and time interval $[t^{n-1},t^n]$ under two approximations: (1) the flux and source terms are approximated using the finite-volume discretization (i.e., $\boldsymbol{g}_i\leftarrow\boldsymbol{g}_i^{\mathrm{FV}}$, and $s_i\leftarrow s_i^{\mathrm{FV}}$), and (2) a polynomial interpolation is used to approximate the integrand for time integration.

Definition 2 *(Feasibility of conservative LSPG projection).* Problem (4.44) is feasible if the LSPG feasible set $\mathcal{F}_{\mathrm{P}}^{n}(\boldsymbol{\mu})$, defined as

$$\mathcal{F}_{\mathrm{P}}^{n}(\boldsymbol{\nu}) := \{\boldsymbol{w} \in \mathbb{R}^{p} \mid \bar{\boldsymbol{C}}\boldsymbol{r}^{n}(\boldsymbol{x}^{0}(\boldsymbol{\mu}) + \boldsymbol{\Phi}\boldsymbol{w};\, \boldsymbol{\mu}) = 0\}, \tag{4.45}$$

is non-empty.

Proposition 4.1. *If Problem (4.44) is feasible, then a solution exists and satisfies the nonlinear saddle-point problem*

$$\begin{aligned} \boldsymbol{\Psi}^{n}(\hat{\boldsymbol{x}}^{n};\, \boldsymbol{\mu})^{T}\left[\boldsymbol{r}^{n}(\boldsymbol{x}^{0}(\boldsymbol{\mu}) + \boldsymbol{\Phi}\hat{\boldsymbol{x}}^{n};\, \boldsymbol{\mu}) + \bar{\boldsymbol{C}}^{T}\boldsymbol{\lambda}_{P}^{n}\right] = \boldsymbol{0} \\ \bar{\boldsymbol{C}}\boldsymbol{r}^{n}(\boldsymbol{x}^{0}(\boldsymbol{\mu}) + \boldsymbol{\Phi}\hat{\boldsymbol{x}}^{n};\, \boldsymbol{\mu}) = \boldsymbol{0}, \end{aligned} \tag{4.46}$$

where $\boldsymbol{\lambda}_{\mathrm{P}}^{n} \in \mathbb{R}^{\bar{N}}$ denote Lagrange multipliers.

Proof. Defining the Lagrangian associated with problem (4.43) as

$$\mathcal{L}_{L}^{n}(\hat{\boldsymbol{z}}, \boldsymbol{\gamma};\, \boldsymbol{\mu}) := \frac{1}{2}\|\boldsymbol{r}^{n}(\boldsymbol{x}^{0}(\boldsymbol{\mu}) + \boldsymbol{\Phi}\hat{\boldsymbol{z}};\, \boldsymbol{\mu})\|_{2}^{2} + \boldsymbol{\gamma}^{T}\bar{\boldsymbol{C}}\boldsymbol{r}^{n}(\boldsymbol{x}^{0}(\boldsymbol{\mu}) + \boldsymbol{\Phi}\hat{\boldsymbol{z}};\, \boldsymbol{\mu}), \tag{4.47}$$

the solution $(\hat{\boldsymbol{x}}^{n}, \boldsymbol{\lambda}_{\mathrm{P}}^{n})$ satisfies the first-order necessary optimality conditions associated with problem (4.43), i.e., $\partial\mathcal{L}_{L}^{n}/\partial\hat{\boldsymbol{z}}(\hat{\boldsymbol{x}}^{n}, \boldsymbol{\lambda}_{\mathrm{P}}^{n};\, \boldsymbol{\mu}) = \boldsymbol{0}$ and $\partial\mathcal{L}_{L}^{n}/\partial\boldsymbol{\gamma}(\hat{\boldsymbol{x}}^{n}, \boldsymbol{\lambda}_{\mathrm{P}}^{n};\, \boldsymbol{\mu}) = \boldsymbol{0}$, which—using the definition of the test basis in Eq. (3.10)—are equivalent to Eqs. (4.46). □

Any appropriate optimization algorithm could be applied to solve minimization problem (4.43) characterizing the conservative LSPG ROM at each time instance. In this work, we propose solving problem (4.43) using the sequential quadratic programming (SQP) method with the Gauss–Newton Hessian approximation. This amounts to applying Newton's method (with globalization) to the first-order necessary optimality conditions (4.46) and neglecting the term involving differentiation of the test basis $\boldsymbol{\Psi}^{n}(\hat{\boldsymbol{x}}^{n};\, \boldsymbol{\mu})$. After choosing an initial guess $\hat{\boldsymbol{x}}^{n(0)}$, this approach leads to the following iterations for $k = 0, \ldots, K$

$$\begin{aligned} &\begin{bmatrix} \boldsymbol{\Psi}^{n}(\hat{\boldsymbol{x}}^{n(k)};\, \boldsymbol{\mu})^{T}\boldsymbol{\Psi}^{n}(\hat{\boldsymbol{x}}^{n(k)};\, \boldsymbol{\mu}) & \boldsymbol{\Psi}^{n}(\hat{\boldsymbol{x}}^{n(k)};\, \boldsymbol{\mu})^{T}\bar{\boldsymbol{C}}^{T} \\ \bar{\boldsymbol{C}}\boldsymbol{\Psi}^{n}(\hat{\boldsymbol{x}}^{n(k)};\, \boldsymbol{\mu}) & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \delta\hat{\boldsymbol{x}}^{n(k)} \\ \delta\boldsymbol{\lambda}_{\mathrm{P}}^{n(k)} \end{bmatrix} \\ &\quad = -\begin{bmatrix} \boldsymbol{\Psi}^{n}(\hat{\boldsymbol{x}}^{n(k)};\, \boldsymbol{\mu})^{T}\left(\boldsymbol{r}^{n}(\boldsymbol{x}^{0}(\boldsymbol{\mu}) + \boldsymbol{\Phi}\hat{\boldsymbol{x}}^{n(k)};\, \boldsymbol{\mu}) + \bar{\boldsymbol{C}}^{T}\boldsymbol{\lambda}_{\mathrm{P}}^{n(k)}\right) \\ \bar{\boldsymbol{C}}\boldsymbol{r}^{n}(\boldsymbol{x}^{0}(\boldsymbol{\mu}) + \boldsymbol{\Phi}\hat{\boldsymbol{x}}^{n(k)};\, \boldsymbol{\mu}) \end{bmatrix}. \end{aligned} \tag{4.48}$$

$$\begin{bmatrix} \hat{\boldsymbol{x}}^{n(k+1)} \\ \boldsymbol{\lambda}_{\mathrm{P}}^{n(k+1)} \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{x}}^{n(k)} \\ \boldsymbol{\lambda}_{\mathrm{P}}^{n(k)} \end{bmatrix} + \eta^{n(k)}\begin{bmatrix} \delta\hat{\boldsymbol{x}}^{n(k)} \\ \delta\boldsymbol{\lambda}_{\mathrm{P}}^{n(k)} \end{bmatrix}, \tag{4.49}$$

where $\eta^{n(k)} \in \mathbb{R}$ is the step length that can be chosen, e.g., to satisfy the strong Wolfe conditions to ensure global convergence to a local solution of (4.44).

4.4. Handling infeasibility

Of course, the optimization problems characterizing conservative Galerkin projection (i.e., problems (4.20)–(4.21)) and conservative LSPG projection (i.e., problems (4.43)–(4.44)) may not be feasible for arbitrary decomposed meshes $\bar{\mathcal{M}}$ and reduced basis matrices $\boldsymbol{\Phi}$. For example, if the decomposed mesh corresponds to the original mesh (i.e., $\bar{\mathcal{M}} = \mathcal{M}$) and the reduced basis is low-dimensional (i.e., $p \ll N$), then the constraints in these problems correspond to *exactly* satisfying the full-order-model equations over a low-dimensional subspace; it is likely impossible to do so.

In practice, infeasibility of a given model can be detected by identifying that the feasible set is empty. In the case of conservative Galerkin projection, this occurs at a given time instance t^{n} and parameter instance $\boldsymbol{\mu}$ if $\mathcal{F}_{\mathrm{G}}(\boldsymbol{x}^{0}(\boldsymbol{\mu}) + \boldsymbol{\Phi}\hat{\boldsymbol{x}}^{n}, t^{n};\, \boldsymbol{\mu}) = \emptyset$, which implies that $\bar{\boldsymbol{C}}\boldsymbol{f}(\boldsymbol{x}^{0}(\boldsymbol{\mu}) + \boldsymbol{\Phi}\hat{\boldsymbol{x}}^{n}, t^{n};\, \boldsymbol{\mu}) \notin \mathrm{Ran}\left(\bar{\boldsymbol{C}}\boldsymbol{\Phi}\right)$. Similarly, in the case of conservation-preserving LSPG projection, infeasibility is detected if a given time instance t^{n} and parameter instance $\boldsymbol{\mu}$ yield $\mathcal{F}_{\mathrm{P}}^{n}(\boldsymbol{\mu}) = \emptyset$, which implies that no value of $\boldsymbol{w}$ can set $\bar{\boldsymbol{C}}\boldsymbol{r}^{n}(\boldsymbol{x}^{0}(\boldsymbol{\mu}) + \boldsymbol{\Phi}\boldsymbol{w};\, \boldsymbol{\mu})$ to zero. We now describe two approaches for handling the case where infeasibility is detected.

1. **Coarsen the decomposed mesh.** First, the number of constraints can be reduced by coarsening the decomposed mesh, i.e., replace $\bar{\mathcal{M}}$ by another decomposed mesh characterized by fewer subdomains $N_{\bar{\Omega}}$. As this reduces the number of constraints, the likelihood of feasibility increases, although feasibility remains not guaranteed. This procedure can be repeated until the decomposed mesh leads to a nonempty feasible set or a decomposed mesh characterized by one subdomain ($N_{\bar{\Omega}} = 1$) is infeasible.

If a decomposed mesh leading to feasibility is constructed via coarsening, the conservative reduced-order model can be redefined using the new decomposed mesh and the reduced-order-model simulation can be either (1) reinitialized and restarted from $t = 0$, or (2) resumed from the time instance t^n where infeasibility was detected. The first approach facilitates analysis, as the reduced-order-model trajectory association with a fixed decomposed mesh, while the latter precludes the need to re-simulate any part of the time interval. Further, if the new decomposed mesh is a decomposition of the previous decomposed mesh, and the previous decomposed mesh is non-overlapping, then conservation over the new decomposed mesh holds over the first part of the time interval (see Theorem 4.1). We note that this approach is not guaranteed to ensure feasibility, as it is possible for infeasibility to exist even in the case of $N_{\bar{\Omega}} = 1$.

2. **Penalty formulation.** Alternatively, infeasibility can be addressed by including the constraints in the objective function via penalization. In the case of conservative Galerkin projection, problem (4.21) is reformulated as

$$\underset{\hat{\boldsymbol{v}}\in\mathbb{R}^p}{\text{minimize}}\ \|\boldsymbol{r}(\boldsymbol{\Phi}\hat{\boldsymbol{v}},\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}},t;\boldsymbol{\mu})\|_2^2+\rho\|\bar{\boldsymbol{C}}\boldsymbol{r}(\boldsymbol{\Phi}\hat{\boldsymbol{v}},\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}},t;\boldsymbol{\mu})\|_2^2, \tag{4.50}$$

while the conservative LSPG projection problem (4.44) is reformulated as

$$\underset{\hat{\boldsymbol{z}}\in\mathbb{R}^p}{\text{minimize}}\ \|\boldsymbol{r}^n(\boldsymbol{x}^0(\boldsymbol{\mu})+\boldsymbol{\Phi}\hat{\boldsymbol{z}};\boldsymbol{\mu})\|_2^2+\rho\|\bar{\boldsymbol{C}}\boldsymbol{r}^n(\boldsymbol{x}^0(\boldsymbol{\mu})+\boldsymbol{\Phi}\hat{\boldsymbol{z}};\boldsymbol{\mu})\|_2^2, \tag{4.51}$$

where $\rho \in \mathbb{R}_+$ is a penalty parameter. This approach does not enforce conservation over any subdomain of the problem.

4.5. Hyper-reduction

To enable hyper-reduction for the proposed conservative reduced-order models, in addition to approximating the nonlinear objective functions that appear in optimization problems (4.21) and (4.44) as previously described in Section 3.3, we must also approximate the nonlinear constraints $\bar{\boldsymbol{C}}\boldsymbol{r}(\boldsymbol{\Phi}\hat{\boldsymbol{v}},\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}},t;\boldsymbol{\mu})=\boldsymbol{0}$ and $\bar{\boldsymbol{C}}\boldsymbol{r}^n(\boldsymbol{x}^0(\boldsymbol{\mu})+\boldsymbol{\Phi}\hat{\boldsymbol{z}};\boldsymbol{\mu})=\boldsymbol{0}$. To accomplish this, we propose applying hyper-reduction to the nonlinear residuals that appears in the constraints, i.e., the constraints become

$$\bar{\boldsymbol{C}}\tilde{\tilde{\boldsymbol{r}}}(\boldsymbol{\Phi}\hat{\boldsymbol{v}},\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}},t;\boldsymbol{\mu})=\boldsymbol{0}\quad\text{and}\quad\bar{\boldsymbol{C}}\tilde{\tilde{\boldsymbol{r}}}^n(\boldsymbol{x}^0(\boldsymbol{\mu})+\boldsymbol{\Phi}\hat{\boldsymbol{z}};\boldsymbol{\mu})=\boldsymbol{0} \tag{4.52}$$

for conservative Galerkin and LSPG projection, respectively. Here, approximations $\tilde{\tilde{\boldsymbol{r}}}(\approx\boldsymbol{r})$ and $\tilde{\tilde{\boldsymbol{r}}}^n(\approx\boldsymbol{r}^n)$ can be constructed using any of the approaches described in Section 3.3; we note that, in general, different approximations can be employed for the objective and constraints such that $\tilde{\tilde{\boldsymbol{r}}}\neq\tilde{\boldsymbol{r}}$ and $\tilde{\tilde{\boldsymbol{r}}}^n\neq\tilde{\boldsymbol{r}}^n$.

In addition to the two forms of hyper-reduction introduced in Section 3.3, we also propose a third type that leverages the underlying finite-volume discretization of the governing equations:

3. **Flux and source hyper-reduction.** This approach respects the underlying decomposition of the velocity vector. It adopts the same residual approximation (3.16)–(3.17) as velocity hyper-reduction (approach 2 in Section 3.3), but employs separate approximations for each term comprising the velocity, i.e.,

$$\tilde{\boldsymbol{f}}=\tilde{\boldsymbol{f}}^s+\tilde{\boldsymbol{f}}^g,\quad\tilde{\boldsymbol{f}}^g=\boldsymbol{B}\tilde{\boldsymbol{h}} \tag{4.53}$$

where

$$\tilde{\boldsymbol{f}}^s=\boldsymbol{\Phi}_s(\boldsymbol{P}_s\boldsymbol{\Phi}_s)^+\boldsymbol{P}_s\boldsymbol{f}^s,\quad\tilde{\boldsymbol{h}}=\boldsymbol{\Phi}_h(\boldsymbol{P}_h\boldsymbol{\Phi}_h)^+\boldsymbol{P}_h\boldsymbol{h} \tag{4.54}$$

in the case of gappy POD, or

$$\tilde{\boldsymbol{f}}^s=\boldsymbol{P}_s^T\boldsymbol{P}_s\boldsymbol{f}^s,\quad\tilde{\boldsymbol{h}}=\boldsymbol{P}_h^T\boldsymbol{P}_h\boldsymbol{h} \tag{4.55}$$

in the case of collocation. Here, $\boldsymbol{P}_s\in\{0,1\}^{n_{p,s}\times N}$ and $\boldsymbol{P}_h\in\{0,1\}^{n_{p,h}\times n_uN_e}$ denote sampling matrices comprising selected rows of the identity matrix, while $\boldsymbol{\Phi}_s\in\mathbb{R}_\star^{N\times p_s}$ and $\boldsymbol{\Phi}_h\in\mathbb{R}_\star^{n_uN_e\times p_h}$ denote reduced-basis matrices constructed for the source and flux, respectively.

One can consider a hierarchy of models that employ objective functions and constraints, each of which may or may not employ one of the three proposed hyper-reduction techniques. For this purpose, we define the Tier-1 and Tier-2 Galerkin and LSPG objective functions as

$$f_{\mathrm{G,I}}(\hat{\boldsymbol{v}},t;\boldsymbol{\mu}):=\|\boldsymbol{r}(\boldsymbol{\Phi}\hat{\boldsymbol{v}},\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}},t;\boldsymbol{\mu})\|_2^2,\quad f_{\mathrm{G,II}}(\hat{\boldsymbol{v}},t;\boldsymbol{\mu}):=\|\tilde{\boldsymbol{r}}(\boldsymbol{\Phi}\hat{\boldsymbol{v}},\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}},t;\boldsymbol{\mu})\|_2^2, \tag{4.56}$$

$$f_{\mathrm{P,I}}^n(\hat{\boldsymbol{z}};\boldsymbol{\mu}):=\|\boldsymbol{r}^n(\boldsymbol{x}^0(\boldsymbol{\mu})+\boldsymbol{\Phi}\hat{\boldsymbol{z}};\boldsymbol{\mu})\|_2^2,\quad f_{\mathrm{P,II}}^n(\hat{\boldsymbol{z}};\boldsymbol{\mu}):=\|\tilde{\boldsymbol{r}}^n(\boldsymbol{x}^0(\boldsymbol{\mu})+\boldsymbol{\Phi}\hat{\boldsymbol{z}};\boldsymbol{\mu})\|_2^2, \tag{4.57}$$

and the Tier-0 (unconstrained), Tier-1, and Tier-2 Galerkin and LSPG constraints as

$$\boldsymbol{c}_{\mathrm{G},0}(\hat{\boldsymbol{v}},t;\boldsymbol{\mu}):=\boldsymbol{0},\ \ \boldsymbol{c}_{\mathrm{G,I}}(\hat{\boldsymbol{v}},t;\boldsymbol{\mu}):=\bar{\boldsymbol{C}}\boldsymbol{r}(\boldsymbol{\Phi}\hat{\boldsymbol{v}},\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}},t;\boldsymbol{\mu}),\ \ \boldsymbol{c}_{\mathrm{G,II}}(\hat{\boldsymbol{v}},t;\boldsymbol{\mu}):=\bar{\boldsymbol{C}}\tilde{\tilde{\boldsymbol{r}}}(\boldsymbol{\Phi}\hat{\boldsymbol{v}},\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}},t;\boldsymbol{\mu}) \tag{4.58}$$

$$\boldsymbol{c}_{\mathrm{P},0}^n(\hat{\boldsymbol{z}};\boldsymbol{\mu}):=\boldsymbol{0},\ \ \boldsymbol{c}_{\mathrm{P,I}}^n(\hat{\boldsymbol{z}};\boldsymbol{\mu}):=\bar{\boldsymbol{C}}\boldsymbol{r}^n(\boldsymbol{x}^0(\boldsymbol{\mu})+\boldsymbol{\Phi}\hat{\boldsymbol{z}};\boldsymbol{\mu}),\ \ \boldsymbol{c}_{\mathrm{P,II}}^n(\hat{\boldsymbol{z}};\boldsymbol{\mu}):=\bar{\boldsymbol{C}}\tilde{\tilde{\boldsymbol{r}}}^n(\boldsymbol{x}^0(\boldsymbol{\mu})+\boldsymbol{\Phi}\hat{\boldsymbol{z}};\boldsymbol{\mu}). \tag{4.59}$$

Then, we say the Tier A-B Galerkin ROM solution $\frac{d\hat{\boldsymbol{x}}}{dt}$ is the solution to

$$\underset{\hat{\boldsymbol{v}}\in\mathbb{R}^p}{\text{minimize}}\ f_{\text{G,A}}(\hat{\boldsymbol{v}}, t; \boldsymbol{\mu}) \text{ subject to } \boldsymbol{c}_{\text{G,B}}(\hat{\boldsymbol{v}}, t; \boldsymbol{\mu}) = \boldsymbol{0} \tag{4.60}$$

and the Tier A-B LSPG ROM solution $\hat{\boldsymbol{x}}^n$ is the solution to

$$\underset{\hat{\boldsymbol{z}}\in\mathbb{R}^p}{\text{minimize}}\ f^n_{\text{P,A}}(\hat{\boldsymbol{z}}; \boldsymbol{\mu}) \text{ subject to } \boldsymbol{c}^n_{\text{P,B}}(\hat{\boldsymbol{z}}; \boldsymbol{\mu}) = \boldsymbol{0}. \tag{4.61}$$

Note that Tier i-0 models correspond to the (original) unconstrained models, Tier i-1 models enforce conservation over subdomains, and Tier i-2 models enforce *approximate* conservation over subdomains. The penalty-method variants of the Tier A-B Galerkin and LSPG ROMs are, respectively,

$$\underset{\hat{\boldsymbol{v}}\in\mathbb{R}^p}{\text{minimize}}\ f_{\text{G,A}}(\hat{\boldsymbol{v}}, t; \boldsymbol{\mu}) + \rho\|\boldsymbol{c}_{\text{G,B}}(\hat{\boldsymbol{v}}, t; \boldsymbol{\mu})\|_2^2 \tag{4.62}$$

$$\underset{\hat{\boldsymbol{z}}\in\mathbb{R}^p}{\text{minimize}}\ f^n_{\text{P,A}}(\hat{\boldsymbol{z}}; \boldsymbol{\mu}) + \rho\|\boldsymbol{c}^n_{\text{P,B}}(\hat{\boldsymbol{z}}; \boldsymbol{\mu})\|_2^2. \tag{4.63}$$

Remark 8 *(Computational cost of evaluating the objective function and constraints).* We note that the computational cost incurred by evaluating the constraints is often significantly lower than the cost of evaluating the objective function. For example, for a linear or zero source term, the only nonlinear contribution to the constraints arises from the face flux along the boundary of the subdomains comprising the decomposed mesh. For a small number of subdomains (e.g., global conservation with $N_{\bar{\Omega}} = 1$), this requires computing only a small number of the elements of the face-flux vector $\boldsymbol{h}$, even without hyper-reduction. Thus, applying hyper-reduction to the objective function is generally more important for computational-cost reduction than applying hyper-reduction to the constraints, i.e., Tier 2–1 ROMs may be preferable to Tier 2–2 ROMs, as their cost is often similar and the former strictly enforces conservation.

4.6. Snapshot-based training

Here, we propose to construct the reduced-basis matrices $\boldsymbol{\Phi}$, $\boldsymbol{\Phi}_r$, $\boldsymbol{\Phi}_f$, $\boldsymbol{\Phi}_h$, and $\boldsymbol{\Phi}_s$ during the offline stage using proper orthogonal decomposition (POD). In particular, given a set of training parameter instances $\mathcal{D}_{\text{train}} := \{\boldsymbol{\mu}^1_{\text{train}}, \ldots, \boldsymbol{\mu}^{n_{\text{train}}}_{\text{train}}\} \subset \mathcal{D}$, we execute training simulations from which we compute ‘data tensors’

$$\mathcal{X}_{ijk} := x_i(t^j; \boldsymbol{\mu}^k_{\text{train}}) - x_i^0(\boldsymbol{\mu}^k_{\text{train}}), \quad i \in \mathbb{N}(N),\ j \in \mathbb{N}(N_T),\ k \in \mathbb{N}(n_{\text{train}}) \tag{4.64}$$

$$\mathcal{R}_{ijk\ell}(\boldsymbol{\xi}) := r_i^j(\boldsymbol{\xi}^{j(\ell)}; \boldsymbol{\mu}^k_{\text{train}}), \quad i \in \mathbb{N}(N),\ j \in \mathbb{N}(N_T),\ k \in \mathbb{N}(n_{\text{train}}),\ \ell \in \mathbb{N}(k_{\max}(\boldsymbol{\xi}, t^j; \boldsymbol{\mu}^k_{\text{train}})) \tag{4.65}$$

$$\mathcal{F}_{ijk}(\boldsymbol{\xi}) := f_i(\boldsymbol{\xi}^j, t^j; \boldsymbol{\mu}^k_{\text{train}}), \quad i \in \mathbb{N}(N),\ j \in \mathbb{N}(N_T),\ k \in \mathbb{N}(n_{\text{train}}) \tag{4.66}$$

$$\mathcal{H}_{ijk}(\boldsymbol{\xi}) := h_i(\boldsymbol{\xi}^j, t^j; \boldsymbol{\mu}^k_{\text{train}}), \quad i \in \mathbb{N}(n_u N_e),\ j \in \mathbb{N}(N_T),\ k \in \mathbb{N}(n_{\text{train}}) \tag{4.67}$$

$$\mathcal{S}_{ijk}(\boldsymbol{\xi}) := f_i^g(\boldsymbol{\xi}^j, t^j; \boldsymbol{\mu}^k_{\text{train}}), \quad i \in \mathbb{N}(N),\ j \in \mathbb{N}(N_T),\ k \in \mathbb{N}(n_{\text{train}}). \tag{4.68}$$

Here, a superscript $j(\ell)$ denotes the value of a variable at the ℓth Newton(-like) iteration during the solution of its nonlinear OΔE at time instance t^j and $k_{\max}(\boldsymbol{\xi}, t^j; \boldsymbol{\mu})$ denotes the maximum number of Newton(-like) iterations taken during the simulation of solution $\boldsymbol{\xi}$ at time instance t^j and parameter instance $\boldsymbol{\mu}$.

Note that constructing the state tensor $\mathcal{X}$ requires solving the full-order model (2.5) at training instances $\boldsymbol{\mu} \in \mathcal{D}_{\text{train}}$, while constructing the other tensors requires computing the solution $\boldsymbol{\xi}$ at these parameter instances; $\boldsymbol{\xi}$ can correspond to the full-order model state (i.e., $\boldsymbol{\xi} = \boldsymbol{x}$) or Tier A-B reduced-order model states (i.e., $\boldsymbol{\xi} = \tilde{\boldsymbol{x}}$) for $\text{A} \in \{1, 2\}$ and $\text{B} \in \{0, 1\}$. Clearly, the least computationally expensive approach is to employ either $\boldsymbol{\xi} = \boldsymbol{x}$—as the training full-order-model simulations are already required to construct the state tensor $\mathcal{X}$—or $\boldsymbol{\xi} = \tilde{\boldsymbol{x}}$ corresponding to the Tier 2-B model for $\text{B} \in \{0, 1\}$, as the hyper-reduced objective function reduces the simulation cost significantly.

The reduced-basis matrix associated with each data tensor can be computed as the dominant left singular vectors of its mode-1 unfolding; for example, the state basis $\boldsymbol{\Phi} \equiv [\boldsymbol{\phi}_1\ \cdots\ \boldsymbol{\phi}_p]$ is computed as

$$\boldsymbol{X}_{(1)} := \left[\boldsymbol{X}(\boldsymbol{\mu}^1_{\text{train}}) \quad \ldots \quad \boldsymbol{X}(\boldsymbol{\mu}^{n_{\text{train}}}_{\text{train}})\right] = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T \in \mathbb{R}^{N\times N_T n_{\text{train}}} \tag{4.69}$$

$$\boldsymbol{\phi}_i = \boldsymbol{u}_i, \quad i \in \mathbb{N}(p), \tag{4.70}$$

where $\boldsymbol{X}(\boldsymbol{\mu}) := [\boldsymbol{x}^1(\boldsymbol{\mu})\ \cdots\ \boldsymbol{x}^{N_T}(\boldsymbol{\mu})]$ is often referred to as the ‘snapshot matrix’.

Further, we propose to construct the sampling matrices $\boldsymbol{P}_r$, $\boldsymbol{P}_f$, $\boldsymbol{P}_h$, and $\boldsymbol{P}_s$ using the sample-mesh greedy method presented in Ref. [17], which allows for oversampling to enable least-squares regression via gappy POD and also constructs a ‘sample mesh’ wherein all residual elements associated with a given control volume are sampled. However, rather than constructing each of these sampling matrices independently, we propose to construct $\boldsymbol{P}_r$ according to the greedy method executed with basis $\boldsymbol{\Phi}_r$ and subsequently set $\boldsymbol{P}_f = \boldsymbol{P}_s = \boldsymbol{P}_r$. Further, we construct $\boldsymbol{P}_h$ to select the faces associated with the control volumes sampled by $\boldsymbol{P}_r$; this corresponds to selecting the sampling matrix $\boldsymbol{P}_h$ with the maximum number of rows such that $\boldsymbol{P}_r\boldsymbol{B} = \boldsymbol{P}_r\boldsymbol{B}\boldsymbol{P}_h^T\boldsymbol{P}_h$.

5. Analysis

This section performs analysis of the proposed conservative Galerkin and conservative LSPG techniques. For simplicity, we focus on the models without hyper-reduction; the hyper-reduced variants of the results can be derived in a similar manner by making the obvious substitutions.

5.1. Feasibility conditions

We first derive sufficient conditions under which the optimization problems characterizing conservative Galerkin and conservative LSPG projection are feasible.

Proposition 5.1 *(Sufficient conditions for feasibility of conservative Galerkin projection). Problem (4.21) is feasible if* $\operatorname{rank}\left(\bar{\boldsymbol{C}}\boldsymbol{\Phi}\right) = \bar{N}$, *which in turn requires $p \geq \bar{N}$, i.e., the number of reduced basis vectors exceeds the number of constraints.*

Proof. If $\operatorname{rank}\left(\bar{\boldsymbol{C}}\boldsymbol{\Phi}\right) = \bar{N}$, then $\operatorname{Ran}\left(\bar{\boldsymbol{C}}\boldsymbol{\Phi}\right) = \mathbb{R}^{\bar{N}}$ and thus $\bar{\boldsymbol{C}}\boldsymbol{f}(\boldsymbol{\xi}, \tau; \boldsymbol{\nu}) \in \operatorname{Ran}\left(\bar{\boldsymbol{C}}\boldsymbol{\Phi}\right)$ regardless of its arguments. □

Proposition 5.2 *(Sufficient conditions for feasibility of conservative LSPG projection). Problem (4.44) is feasible if (1) an explicit scheme is employed and* $\operatorname{rank}\left(\bar{\boldsymbol{C}}\boldsymbol{\Phi}\right) = \bar{N}$, *(2) the limit $\Delta t \to 0$ is taken, or (3) the velocity $\boldsymbol{f}$ is linear in its first argument and* $\operatorname{rank}\left(\bar{\boldsymbol{C}}[\alpha_0\boldsymbol{I} - \Delta t\beta_0 \partial \boldsymbol{f}/\partial\boldsymbol{\xi}(\cdot, t^n; \boldsymbol{\mu})]\boldsymbol{\Phi}\right) = \bar{N}$.

Proof. *Case 1.* If an explicit scheme is employed, then $\beta_0 = 0$ and the feasible set becomes

$$\mathcal{F}_{\mathrm{P}}^n(\boldsymbol{\nu}) = \{\boldsymbol{w} \in \mathbb{R}^p \,|\, \alpha_0\bar{\boldsymbol{C}}\boldsymbol{\Phi}\boldsymbol{w} = -\sum_{j=1}^{k}\alpha_j\bar{\boldsymbol{C}}\boldsymbol{\Phi}\hat{\boldsymbol{x}}^{n-j}(\boldsymbol{\mu}) + \Delta t\sum_{j=1}^{k}\beta_j\bar{\boldsymbol{C}}\boldsymbol{f}(\boldsymbol{x}^0(\boldsymbol{\mu}) + \boldsymbol{\Phi}\hat{\boldsymbol{x}}^{n-j}, t^{n-j}; \boldsymbol{\mu})\}. \tag{5.1}$$

If $\operatorname{rank}\left(\bar{\boldsymbol{C}}\boldsymbol{\Phi}\right) = \bar{N}$, then $\operatorname{Ran}\left(\bar{\boldsymbol{C}}\boldsymbol{\Phi}\right) = \mathbb{R}^{\bar{N}}$ and right-hand-side of the constraints in (5.1) must lie in $\operatorname{Ran}\left(\bar{\boldsymbol{C}}\boldsymbol{\Phi}\right)$.

Case 2. If the limit $\Delta t \to 0$ is taken, then the feasible set becomes

$$\mathcal{F}_{\mathrm{P}}^n(\boldsymbol{\nu}) = \{\boldsymbol{w} \in \mathbb{R}^p \,|\, \alpha_0\bar{\boldsymbol{C}}\boldsymbol{\Phi}\boldsymbol{w} = -\sum_{j=1}^{k}\alpha_j\bar{\boldsymbol{C}}\boldsymbol{\Phi}\hat{\boldsymbol{x}}^{n-j}(\boldsymbol{\mu})\} \tag{5.2}$$

and the right-hand-side of the constraints in (5.2) will lie in $\operatorname{Ran}\left(\bar{\boldsymbol{C}}\boldsymbol{\Phi}\right)$ regardless of its rank.

Case 3. If the velocity is linear in the state, the feasible set becomes

$$\begin{aligned}\mathcal{F}_{\mathrm{P}}^n(\boldsymbol{\nu}) = \{\boldsymbol{w} \in \mathbb{R}^p \,|\, &\bar{\boldsymbol{C}}[\alpha_0\boldsymbol{\Phi} - \Delta t\beta_0\partial\boldsymbol{f}/\partial\boldsymbol{\xi}(\cdot, t^n; \boldsymbol{\mu})\boldsymbol{\Phi}]\boldsymbol{w} = \\ &-\sum_{j=1}^{k}\alpha_j\bar{\boldsymbol{C}}\boldsymbol{\Phi}\hat{\boldsymbol{x}}^{n-j}(\boldsymbol{\mu}) + \Delta t\sum_{j=1}^{k}\beta_j\bar{\boldsymbol{C}}\boldsymbol{f}(\boldsymbol{x}^0(\boldsymbol{\mu}) + \boldsymbol{\Phi}\hat{\boldsymbol{x}}^{n-j}, t^{n-j}; \boldsymbol{\mu})\}\end{aligned} \tag{5.3}$$

and (as above) $\operatorname{rank}\left(\bar{\boldsymbol{C}}[\alpha_0\boldsymbol{I} - \Delta t\beta_0\partial\boldsymbol{f}/\partial\boldsymbol{\xi}(\cdot, t^n; \boldsymbol{\mu})]\boldsymbol{\Phi}\right) = \bar{N}$ ensures the right-hand-side of the constraints in (5.1) will lie in $\operatorname{Ran}\left(\bar{\boldsymbol{C}}\boldsymbol{\Phi}\right)$. □

5.2. Equivalence conditions

We now derive conditions under which conservative Galerkin and conservative LSPG projection are equivalent.

Theorem 5.1 *(Equivalence). The discrete-time conservative Galerkin ROM solution is equivalent to the conservative LSPG solution if either (1) an explicit scheme is employed or (2) the limit $\Delta t \to 0$ is taken. Further, under these conditions, the Lagrange multipliers are related as*

$$\boldsymbol{\lambda}_P^n = \sum_{j=0}^{k}\alpha_j\boldsymbol{\lambda}_G^{n-j}. \tag{5.4}$$

Proof. We first note that the discrete-time conservative Galerkin ROM solution $\hat{\boldsymbol{x}}_{\mathrm{G}}^n$ satisfies Eqs. (4.40), which can be re-written as

$$\begin{aligned}\boldsymbol{\Phi}^T\left[\boldsymbol{r}^n(\boldsymbol{x}^0(\boldsymbol{\mu})+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{G}}^n;\boldsymbol{\mu})+\sum_{j=0}^{k}\alpha_j\bar{\boldsymbol{C}}^T\boldsymbol{\lambda}_{\mathrm{G}}^{n-j}\right]&=\mathbf{0}\\ \bar{\boldsymbol{C}}\boldsymbol{r}^n(\boldsymbol{x}^0(\boldsymbol{\mu})+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{G}}^n;\boldsymbol{\mu})&=\mathbf{0}.\end{aligned}\tag{5.5}$$

Comparing Eqs. (4.46) and (5.5) reveals that the discrete-time conservative LSPG and conservative Galerkin solutions are equivalent if (1) $\boldsymbol{\Psi}^n(\hat{\boldsymbol{x}}^n;\boldsymbol{\mu})=a\boldsymbol{\Phi}$ and (2) $\boldsymbol{\Psi}^n(\hat{\boldsymbol{x}}^n;\boldsymbol{\mu})^T\bar{\boldsymbol{C}}^T\boldsymbol{\lambda}_{\mathrm{P}}^n=a\sum_{j=0}^k\alpha_j\boldsymbol{\Phi}^T\bar{\boldsymbol{C}}^T\boldsymbol{\lambda}_{\mathrm{G}}^{n-j}$ for any constant $a\in\mathbb{R}$. As was shown in Ref. [15], the first condition holds for $a=\alpha_0$ if either the scheme is explicit (i.e., $\beta_0=0$) or the limit $\Delta t\to 0$ is taken. The same conditions apply to the second condition above. To see this, note that

$$\bar{\boldsymbol{C}}\boldsymbol{\Psi}^n(\hat{\boldsymbol{x}}^n;\boldsymbol{\mu})=\alpha_0\bar{\boldsymbol{C}}\boldsymbol{\Phi}-\Delta t\beta_0\bar{\boldsymbol{C}}\frac{\partial\boldsymbol{f}}{\partial\boldsymbol{\xi}}(\boldsymbol{w},t^n;\boldsymbol{\mu})\boldsymbol{\Phi}.\tag{5.6}$$

If either $\beta_0=0$ or the limit $\Delta t\to 0$ is taken, the second term vanishes such that we have

$$\boldsymbol{\Psi}^n(\hat{\boldsymbol{x}}^n;\boldsymbol{\mu})^T\bar{\boldsymbol{C}}^T\boldsymbol{\lambda}_{\mathrm{P}}^n=\alpha_0\boldsymbol{\Phi}^T\bar{\boldsymbol{C}}^T\boldsymbol{\lambda}_{\mathrm{P}}^n.\tag{5.7}$$

This expression is equivalent to $a\sum_{j=0}^k\alpha_j\boldsymbol{\Phi}^T\bar{\boldsymbol{C}}^T\boldsymbol{\lambda}_{\mathrm{G}}^{n-j}$ with $a=\alpha_0$ if Eq. (5.4) holds. □

5.3. Error analysis

We now derive several *a posteriori* error bounds for (components of) the solution computed by the proposed conservative model-reduction methods. We employ some of the same techniques used for error analysis in Ref. [15]. For notational simplicity, we drop dependence of the operators on the parameters $\boldsymbol{\mu}$.

We begin by writing the discrete equations characterizing the full-order, Galerkin, and LSPG models as

$$\alpha_0^n\boldsymbol{x}_\star^n=\beta_0^n\Delta t\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{x}_\star^n,t^n)+\bar{\boldsymbol{r}}^n[\boldsymbol{x}_\star^{n-k},\ldots,\boldsymbol{x}_\star^{n-1}]\tag{5.8}$$

$$\alpha_0^n\hat{\boldsymbol{x}}_{\mathrm{G}}^n=\beta_0^n\Delta t\boldsymbol{\Phi}^T\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{G}}^n,t^n)+\boldsymbol{\Phi}^T\bar{\boldsymbol{r}}^n[\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{G}}^{n-k},\ldots,\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{G}}^{n-1}]-\sum_{j=0}^{k}\alpha_j^n\boldsymbol{\Phi}^T\bar{\boldsymbol{C}}^T\boldsymbol{\lambda}_{\mathrm{G}}^{n-j},\tag{5.9}$$

$$\begin{aligned}\alpha_0^n\hat{\boldsymbol{x}}_{\mathrm{P}}^n=&\beta_0^n\Delta t((\boldsymbol{\Psi}^n)^T\boldsymbol{\Phi})^{-1}(\boldsymbol{\Psi}^n)^T\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^n,t^n)+((\boldsymbol{\Psi}^n)^T\boldsymbol{\Phi})^{-1}(\boldsymbol{\Psi}^n)^T\bar{\boldsymbol{r}}^n[\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-k},\ldots,\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-1}]\\&-((\boldsymbol{\Psi}^n)^T\boldsymbol{\Phi})^{-1}(\boldsymbol{\Psi}^n)^T\bar{\boldsymbol{C}}^T\boldsymbol{\lambda}_{\mathrm{P}}^n,\end{aligned}\tag{5.10}$$

respectively, where $\boldsymbol{\Psi}^n:=\boldsymbol{\Psi}^n(\hat{\boldsymbol{x}}_{\mathrm{P}}^n)$, as well as

$$\alpha_0^n\bar{\boldsymbol{C}}\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{G}}^n=\beta_0^n\Delta t\bar{\boldsymbol{C}}\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{G}}^n,t^n)+\bar{\boldsymbol{C}}\bar{\boldsymbol{r}}^n[\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{G}}^{n-k},\ldots,\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{G}}^{n-1}]\tag{5.11}$$

$$\alpha_0^n\bar{\boldsymbol{C}}\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^n=\beta_0^n\Delta t\bar{\boldsymbol{C}}\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^n,t^n)+\bar{\boldsymbol{C}}\bar{\boldsymbol{r}}^n[\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-k},\ldots,\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-1}]\tag{5.12}$$

with $\boldsymbol{x}_\star^0=\mathbf{0}$ and $\hat{\boldsymbol{x}}_{\mathrm{G}}^0=\hat{\boldsymbol{x}}_{\mathrm{P}}^0=\mathbf{0}$. Here, we have defined

$$\bar{\boldsymbol{r}}^n[\boldsymbol{x}^{n-k},\ldots,\boldsymbol{x}^{n-1}]:=\sum_{\ell=1}^{k}\left(\beta_\ell^n\Delta t\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{x}^{n-\ell},t^{n-\ell})-\alpha_\ell^n\boldsymbol{x}^{n-\ell}\right).\tag{5.13}$$

We also assume Lipschitz continuity of $\boldsymbol{f}$ in its first argument:

A₁ There exists a constant $\kappa>0$ such that for $\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^N$

$$\|\boldsymbol{f}(\boldsymbol{x},t)-\boldsymbol{f}(\boldsymbol{y},t)\|_2\le\kappa\|\boldsymbol{x}-\boldsymbol{y}\|_2,\quad\forall t\in[0,T].$$

To simplify notation, we define the Galerkin and LSPG operators as

$$\mathbb{V}:=\boldsymbol{\Phi}\boldsymbol{\Phi}^T,\quad\mathbb{P}^n:=\boldsymbol{\Phi}((\boldsymbol{\Psi}^n)^T\boldsymbol{\Phi})^{-1}(\boldsymbol{\Psi}^n)^T,\tag{5.14}$$

respectively, and the Galerkin and LSPG state-space errors at time instance n as

$$\delta\hat{\boldsymbol{x}}_{\mathrm{G}}^n:=\boldsymbol{x}_\star^n-\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{G}}^n,\quad\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^n:=\boldsymbol{x}_\star^n-\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^n,\tag{5.15}$$

respectively. Because the time instance of the first and second arguments of $\boldsymbol{f}$ always match for linear multistep schemes, we omit the second argument (time) from $\boldsymbol{f}$ in the remainder of this section. All norms in this section correspond to the Euclidean norm, i.e., $\|\cdot\|=\|\cdot\|_2$.

We proceed by deriving *a posteriori* error bounds for the proposed conservative techniques. We remark that derivation of state-space error bounds for the proposed constrained ROMs is complicated by the presence of Lagrange multipliers in the discrete equations (5.9)–(5.10). Thus, we derive bounds that relate to the null-space and row-space of the associated constraint matrices, which enables elimination of these Lagrange multipliers from the analysis. To accomplish this, we make use of three decompositions of $\mathbb{R}^N$. The first is $\mathbb{R}^N = \mathrm{Ran}\left(\bar{\boldsymbol{V}}_{\mathrm{G}}\right) \oplus \mathrm{Ran}\left(\bar{\boldsymbol{Z}}_{\mathrm{G}}\right)$, where $\bar{\boldsymbol{V}}_{\mathrm{G}} \in \mathbb{R}^{N\times \mathrm{rank}(\bar{\boldsymbol{C}}\boldsymbol{\Phi})}$ and $\bar{\boldsymbol{Z}}_{\mathrm{G}} \in \mathbb{R}^{N\times N-\mathrm{rank}(\bar{\boldsymbol{C}}\boldsymbol{\Phi})}$ are orthogonal matrices satisfying

$$\bar{\boldsymbol{C}}\boldsymbol{\Phi} = \boldsymbol{U}_{\mathrm{G}}\boldsymbol{\Sigma}_{\mathrm{G}}\boldsymbol{V}_{\mathrm{G}}^T, \quad \bar{\boldsymbol{V}}_{\mathrm{G}} = \boldsymbol{\Phi}\boldsymbol{V}_{\mathrm{G}}, \quad \bar{\boldsymbol{Z}}_{\mathrm{G}}^T\bar{\boldsymbol{V}}_{\mathrm{G}} = \boldsymbol{0}. \tag{5.16}$$

The second is $\mathbb{R}^N = \mathrm{Ran}\left(\bar{\boldsymbol{V}}_{\mathrm{P}}^n\right) \oplus \mathrm{Ran}\left(\bar{\boldsymbol{Z}}_{\mathrm{P}}^n\right)$, where $\bar{\boldsymbol{V}}_{\mathrm{P}}^n \in \mathbb{R}^{N\times \mathrm{rank}(\bar{\boldsymbol{C}}\boldsymbol{\Psi}^n)}$ and $\bar{\boldsymbol{Z}}_{\mathrm{P}}^n \in \mathbb{R}^{N\times N-\mathrm{rank}(\bar{\boldsymbol{C}}\boldsymbol{\Psi}^n)}$ are orthogonal matrices satisfying

$$\bar{\boldsymbol{C}}\boldsymbol{\Psi}^n(\boldsymbol{\Phi}^T\boldsymbol{\Psi}^n)^{-1} = \boldsymbol{U}_{\mathrm{P}}^n\boldsymbol{\Sigma}_{\mathrm{P}}^n[\boldsymbol{V}_{\mathrm{P}}^n]^T, \quad \bar{\boldsymbol{V}}_{\mathrm{P}}^n = \boldsymbol{\Phi}\boldsymbol{V}_{\mathrm{P}}^n, \quad [\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\bar{\boldsymbol{V}}_{\mathrm{P}}^n = \boldsymbol{0}. \tag{5.17}$$

Finally, we consider $\mathbb{R}^N = \mathrm{Ran}\left(\boldsymbol{V}_{\bar{\boldsymbol{C}}}\right) \oplus \mathrm{Ran}\left(\boldsymbol{Z}_{\bar{\boldsymbol{C}}}\right)$, where $\boldsymbol{V}_{\bar{\boldsymbol{C}}} \in \mathbb{R}^{N\times \mathrm{rank}(\bar{\boldsymbol{C}})}$ and $\boldsymbol{Z}_{\bar{\boldsymbol{C}}} \in \mathbb{R}^{N\times N-\mathrm{rank}(\bar{\boldsymbol{C}})}$ are orthogonal matrices satisfying

$$\bar{\boldsymbol{C}} = \boldsymbol{U}_{\bar{\boldsymbol{C}}}\boldsymbol{\Sigma}_{\bar{\boldsymbol{C}}}\boldsymbol{V}_{\bar{\boldsymbol{C}}}^T, \quad \boldsymbol{Z}_{\bar{\boldsymbol{C}}}^T\boldsymbol{V}_{\bar{\boldsymbol{C}}} = \boldsymbol{0}. \tag{5.18}$$

Note that $\bar{\boldsymbol{C}}^+\bar{\boldsymbol{C}} = \boldsymbol{V}_{\bar{\boldsymbol{C}}}\boldsymbol{V}_{\bar{\boldsymbol{C}}}^T$.

Lemma 5.1 *(Local a posteriori error bounds: null-space error). If* $\mathbf{A_1}$ *holds and* $\Delta t < |\alpha_0^n|/(|\beta_0^n|\kappa)$*, then*

$$\|\bar{\boldsymbol{Z}}_{\mathrm{G}}^T\delta\hat{\boldsymbol{x}}_{\mathrm{G}}^n\| \leq \sum_{\ell=0}^{k}\varepsilon_\ell^n\left(\kappa\|\bar{\boldsymbol{V}}_{\mathrm{G}}^T\delta\hat{\boldsymbol{x}}_{\mathrm{G}}^{n-\ell}\| + \|(\boldsymbol{I}-\mathbb{V})\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{G}}^{n-\ell})\|\right) + \sum_{\ell=1}^{k}\gamma_\ell^n\|\bar{\boldsymbol{Z}}_{\mathrm{G}}^T\delta\hat{\boldsymbol{x}}_{\mathrm{G}}^{n-\ell}\| \tag{5.19}$$

$$\|[\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^n\| \leq \sum_{\ell=0}^{k}\varepsilon_\ell^n\left(\kappa\|[\bar{\boldsymbol{V}}_{\mathrm{P}}^n]^T\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell}\| + \|(\boldsymbol{I}-\mathbb{P}^n)\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell})\|\right) + \sum_{\ell=1}^{k}\gamma_\ell^n\|[\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell}\|, \tag{5.20}$$

where $\varepsilon_\ell^m := |\beta_\ell^m|\Delta t/h^m$, $\gamma_\ell^m := (|\alpha_\ell^m| + |\beta_\ell^m|\kappa\Delta t)/h^m$, *and* $h^m := |\alpha_0^m| - |\beta_0^m|\kappa\Delta t$.

Proof. Subtracting the premultiplication of Eq. (5.10) by $[\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\boldsymbol{\Phi}$ from the premultiplication of Eq. (5.8) by $[\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T$ yields

$$\alpha_0^n[\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^n = \beta_0^n\Delta t[\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\left(\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{x}_\star^n) - \mathbb{P}^n\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^n)\right) + [\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\delta\boldsymbol{r}_{\mathrm{P}}^{n-1}, \tag{5.21}$$

where $\delta\boldsymbol{r}_{\mathrm{P}}^{n-1} := \bar{\boldsymbol{r}}^n[\boldsymbol{x}_\star^{n-k},\ldots,\boldsymbol{x}_\star^{n-1}] - \mathbb{P}^n\bar{\boldsymbol{r}}^n[\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-k},\ldots,\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-1}]$. Adding and subtracting $\beta_0^n\Delta t[\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^n)$ from Eq. (5.21) yields

$$\alpha_0^n[\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^n = \beta_0^n\Delta t[\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T[\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{x}_\star^n) - \boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^n) + \boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^n) - \mathbb{P}^n\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^n)] + [\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\delta\boldsymbol{r}_{\mathrm{P}}^{n-1}. \tag{5.22}$$

Applying the triangle inequality yields

$$|\alpha_0^n|\|[\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^n\| \leq |\beta_0^n|\Delta t\left(\|\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{x}_\star^n) - \boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^n)\| + \|(\boldsymbol{I}-\mathbb{P}^n)\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^n)\|\right) + \|[\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\delta\boldsymbol{r}_{\mathrm{P}}^{n-1}\|. \tag{5.23}$$

Now, using Lipschitz continuity and $\boldsymbol{y} = \bar{\boldsymbol{V}}_{\mathrm{P}}^n[\bar{\boldsymbol{V}}_{\mathrm{P}}^n]^T\boldsymbol{y} + \bar{\boldsymbol{Z}}_{\mathrm{P}}^n[\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\boldsymbol{y}$ for all $\boldsymbol{y} \in \mathbb{R}^N$, we have

$$|\alpha_0^n|\|[\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^n\| \leq |\beta_0^n|\Delta t\left(\kappa\|[\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^n\| + \kappa\|[\bar{\boldsymbol{V}}_{\mathrm{P}}^n]^T\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^n\| + \|(\boldsymbol{I}-\mathbb{P}^n)\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^n)\|\right) + \|[\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\delta\boldsymbol{r}_{\mathrm{P}}^{n-1}\|. \tag{5.24}$$

Using $\Delta t < |\alpha_0^n|/(|\beta_0^n|\kappa)$, we have

$$\|[\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^n\| \leq \frac{|\beta_0^n|\Delta t}{h^n}\left(\kappa\|[\bar{\boldsymbol{V}}_{\mathrm{P}}^n]^T\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^n\| + \|(\boldsymbol{I}-\mathbb{P}^n)\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^n)\|\right) + \frac{1}{h^n}\|[\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\delta\boldsymbol{r}_{\mathrm{P}}^{n-1}\|. \tag{5.25}$$

Next, we estimate $\|[\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\delta\boldsymbol{r}_{\mathrm{P}}^{n-1}\|$. First, we have

$$[\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\delta\boldsymbol{r}_{\mathrm{P}}^{n-1} = \sum_{\ell=1}^{k}\beta_\ell^n\Delta t[\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\left(\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{x}_\star^{n-\ell}) - \mathbb{P}^n\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell})\right) - \sum_{\ell=1}^{k}\alpha_\ell^n[\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\left(\boldsymbol{x}_\star^{n-\ell} - \boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell}\right). \tag{5.26}$$

Following similar steps to those above, adding and subtracting $\sum_{\ell=1}^k\beta_\ell^n\Delta t[\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell})$ and applying the triangle inequality yields

$$\|[\bar{\boldsymbol{Z}}_{\mathrm{P}}^{n}]^{T}\delta\boldsymbol{r}_{\mathrm{P}}^{n-1}\| \le \sum_{\ell=1}^{k}|\beta_{\ell}^{n}|\Delta t\left(\kappa\|[\bar{\boldsymbol{V}}_{\mathrm{P}}^{n}]^{T}\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell}\| + \|(\boldsymbol{I}-\mathbb{P}^{n})\boldsymbol{f}(\boldsymbol{x}^{0}+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell})\|\right) + \sum_{\ell=1}^{k}(|\beta_{\ell}^{n}|\kappa\Delta t + |\alpha_{\ell}^{n}|)\|[\bar{\boldsymbol{Z}}_{\mathrm{P}}^{n}]^{T}\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell}\|. \tag{5.27}$$

Combining inequalities (5.25) and (5.27) yields the final result (5.20). The Galerkin counterpart (5.19) can be derived by following the same steps with the Galerkin operators. □

Lemma 5.1 shows that the component of the error in the null space of the constraints behaves very similarly to the full-space error in the case of standard, unconstrained ROMs as reported in [15, Theorem 6.1]; the only difference is the addition of the terms arising from the row-space errors, which is $\sum_{\ell=0}^{k}\varepsilon_{\ell}^{n}\kappa\|\bar{\boldsymbol{V}}_{\mathrm{G}}^{T}\delta\hat{\boldsymbol{x}}_{\mathrm{G}}^{n-\ell}\|$ and $\sum_{\ell=0}^{k}\varepsilon_{\ell}^{n}\kappa\|[\bar{\boldsymbol{V}}_{\mathrm{P}}^{n}]^{T}\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell}\|$ for conservative Galerkin and conservative LSPG projection, respectively.

Lemma 5.2 *(Local a posteriori error bounds: row-space error). If $\mathbf{A_1}$ holds and $\Delta t < |\alpha_{0}^{n}|/(|\beta_{0}^{n}|\kappa)$, then*

$$\|\bar{\boldsymbol{V}}_{\mathrm{G}}^{T}\delta\hat{\boldsymbol{x}}_{\mathrm{G}}^{n}\| \le \sum_{\ell=0}^{k}\varepsilon_{\ell}^{n}\left(\kappa\|\bar{\boldsymbol{Z}}_{\mathrm{G}}^{T}\delta\hat{\boldsymbol{x}}_{\mathrm{G}}^{n-\ell}\| + \zeta_{\mathrm{G}}\|(\boldsymbol{I}-\mathbb{V})\boldsymbol{f}(\boldsymbol{x}^{0}+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{G}}^{n-\ell})\|\right) + \sum_{\ell=1}^{k}\gamma_{\ell}^{n}\|\bar{\boldsymbol{V}}_{\mathrm{G}}^{T}\delta\hat{\boldsymbol{x}}_{\mathrm{G}}^{n-\ell}\| \tag{5.28}$$

$$\begin{aligned}\|[\bar{\boldsymbol{V}}_{\mathrm{P}}^{n}]^{T}\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^{n}\| \le{}& \sum_{\ell=0}^{k}\varepsilon_{\ell}^{n}\left(\kappa\|[\bar{\boldsymbol{Z}}_{\mathrm{P}}^{n}]^{T}\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell}\| + \zeta_{\mathrm{P}}^{n}\|(\boldsymbol{I}-[\mathbb{P}^{n}]^{T})\boldsymbol{f}(\boldsymbol{x}^{0}+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell})\|\right) + \sum_{\ell=1}^{k}\gamma_{\ell}^{n}\|[\bar{\boldsymbol{V}}_{\mathrm{P}}^{n}]^{T}\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell}\| \\ &+ \frac{\zeta_{\mathrm{P}}^{n}}{h^{n}}\|\Delta^{n}\|\sum_{\ell=0}^{k}|\alpha_{\ell}^{n}|\|\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell}\|,\end{aligned} \tag{5.29}$$

where $\zeta_{\mathrm{G}} := \|\boldsymbol{\Sigma}_{\mathrm{G}}^{-1}\boldsymbol{U}_{\mathrm{G}}^{T}\bar{\boldsymbol{C}}\|$, $\zeta_{\mathrm{P}}^{n} := \|[\boldsymbol{\Sigma}_{\mathrm{P}}^{n}]^{-1}[\boldsymbol{U}_{\mathrm{P}}^{n}]^{T}\bar{\boldsymbol{C}}\|$, and $\Delta^{n} := \boldsymbol{\Psi}^{n}(\boldsymbol{\Phi}^{T}\boldsymbol{\Psi}^{n})^{-1} - \boldsymbol{\Phi}$.

Proof. Noting that $[\bar{\boldsymbol{V}}_{\mathrm{P}}^{n}]^{T}\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n} = [\boldsymbol{V}_{\mathrm{P}}^{n}]^{T}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n}$, we have from adding and subtracting $\alpha_{0}^{n}\bar{\boldsymbol{C}}[\boldsymbol{\Psi}^{n}(\boldsymbol{\Phi}^{T}\boldsymbol{\Psi}^{n})^{-1}]\hat{\boldsymbol{x}}_{\mathrm{P}}^{n}$ to Eq. (5.12) and premultiplying by $[\boldsymbol{\Sigma}_{\mathrm{P}}^{n}]^{-1}[\boldsymbol{U}_{\mathrm{P}}^{n}]^{T}$

$$\alpha_{0}^{n}[\bar{\boldsymbol{V}}_{\mathrm{P}}^{n}]^{T}\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n} = \beta_{0}^{n}\Delta t[\boldsymbol{\Sigma}_{\mathrm{P}}^{n}]^{-1}[\boldsymbol{U}_{\mathrm{P}}^{n}]^{T}\bar{\boldsymbol{C}}\boldsymbol{f}(\boldsymbol{x}^{0}+\hat{\boldsymbol{x}}_{\mathrm{P}}^{n}) + \alpha_{0}^{n}[\boldsymbol{\Sigma}_{\mathrm{P}}^{n}]^{-1}[\boldsymbol{U}_{\mathrm{P}}^{n}]^{T}\bar{\boldsymbol{C}}\Delta^{n}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n} + [\boldsymbol{\Sigma}_{\mathrm{P}}^{n}]^{-1}[\boldsymbol{U}_{\mathrm{P}}^{n}]^{T}\bar{\boldsymbol{C}}\bar{\boldsymbol{r}}^{n}[\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-k},\ldots,\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-1}]. \tag{5.30}$$

Premultiplying Eq. (5.8) by $[\bar{\boldsymbol{V}}_{\mathrm{P}}^{n}]^{T} = [\boldsymbol{\Sigma}_{\mathrm{P}}^{n}]^{-1}[\boldsymbol{U}_{\mathrm{P}}^{n}]^{T}\bar{\boldsymbol{C}}[\mathbb{P}^{n}]^{T}$ yields

$$\alpha_{0}^{n}[\bar{\boldsymbol{V}}_{\mathrm{P}}^{n}]^{T}\boldsymbol{x}_{\star}^{n} = \beta_{0}^{n}\Delta t[\boldsymbol{\Sigma}_{\mathrm{P}}^{n}]^{-1}[\boldsymbol{U}_{\mathrm{P}}^{n}]^{T}\bar{\boldsymbol{C}}[\mathbb{P}^{n}]^{T}\boldsymbol{f}(\boldsymbol{x}^{0}+\boldsymbol{x}_{\star}^{n}) + [\boldsymbol{\Sigma}_{\mathrm{P}}^{n}]^{-1}[\boldsymbol{U}_{\mathrm{P}}^{n}]^{T}\bar{\boldsymbol{C}}[\mathbb{P}^{n}]^{T}\bar{\boldsymbol{r}}^{n}[\boldsymbol{\Phi}\boldsymbol{x}_{\star}^{n-k},\ldots,\boldsymbol{\Phi}\boldsymbol{x}_{\star}^{n-1}]. \tag{5.31}$$

Subtracting (5.30) from (5.31) yields

$$\begin{aligned}\alpha_{0}^{n}[\bar{\boldsymbol{V}}_{\mathrm{P}}^{n}]^{T}\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^{n} ={}& \beta_{0}^{n}\Delta t[\boldsymbol{\Sigma}_{\mathrm{P}}^{n}]^{-1}[\boldsymbol{U}_{\mathrm{P}}^{n}]^{T}\bar{\boldsymbol{C}}\left[[\mathbb{P}^{n}]^{T}\boldsymbol{f}(\boldsymbol{x}^{0}+\boldsymbol{x}_{\star}^{n}) - \boldsymbol{f}(\boldsymbol{x}^{0}+\hat{\boldsymbol{x}}_{\mathrm{P}}^{n})\right] - \alpha_{0}^{n}[\boldsymbol{\Sigma}_{\mathrm{P}}^{n}]^{-1}[\boldsymbol{U}_{\mathrm{P}}^{n}]^{T}\bar{\boldsymbol{C}}\Delta^{n}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n} \\ &+ [\boldsymbol{\Sigma}_{\mathrm{P}}^{n}]^{-1}[\boldsymbol{U}_{\mathrm{P}}^{n}]^{T}\bar{\boldsymbol{C}}\delta\boldsymbol{r}_{\mathrm{P},\star}^{n-1},\end{aligned} \tag{5.32}$$

where $\delta\boldsymbol{r}_{\mathrm{P},\star}^{n-1} := [\mathbb{P}^{n}]^{T}\bar{\boldsymbol{r}}^{n}[\boldsymbol{x}_{\star}^{n-k},\ldots,\boldsymbol{x}_{\star}^{n-1}] - \bar{\boldsymbol{r}}^{n}[\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-k},\ldots,\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-1}]$. Adding and subtracting $[\mathbb{P}^{n}]^{T}\boldsymbol{f}(\boldsymbol{x}^{0}+\hat{\boldsymbol{x}}_{\mathrm{P}}^{n})$ to the bracketed quantity, applying the triangle inequality, and using Lipschitz continuity yields

$$\begin{aligned}|\alpha_{0}^{n}|\|[\bar{\boldsymbol{V}}_{\mathrm{P}}^{n}]^{T}\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^{n}\| \le{}& |\beta_{0}^{n}|\Delta t\kappa\|[\boldsymbol{\Sigma}_{\mathrm{P}}^{n}]^{-1}[\boldsymbol{U}_{\mathrm{P}}^{n}]^{T}\bar{\boldsymbol{C}}[\mathbb{P}^{n}]^{T}\|(\|[\bar{\boldsymbol{V}}_{\mathrm{P}}^{n}]^{T}\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^{n}\| + \|[\bar{\boldsymbol{Z}}_{\mathrm{P}}^{n}]^{T}\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^{n}\|) \\ &+ |\beta_{0}^{n}|\Delta t\|[\boldsymbol{\Sigma}_{\mathrm{P}}^{n}]^{-1}[\boldsymbol{U}_{\mathrm{P}}^{n}]^{T}\bar{\boldsymbol{C}}\|\|(\boldsymbol{I}-\mathbb{P}^{n})\boldsymbol{f}(\boldsymbol{x}^{0}+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n})\| \\ &+ |\alpha_{0}^{n}|\|[\boldsymbol{\Sigma}_{\mathrm{P}}^{n}]^{-1}[\boldsymbol{U}_{\mathrm{P}}^{n}]^{T}\bar{\boldsymbol{C}}\Delta^{n}\|\|\hat{\boldsymbol{x}}_{\mathrm{P}}^{n}\| + \|[\boldsymbol{\Sigma}_{\mathrm{P}}^{n}]^{-1}[\boldsymbol{U}_{\mathrm{P}}^{n}]^{T}\bar{\boldsymbol{C}}\delta\boldsymbol{r}_{\mathrm{P},\star}^{n-1}\|.\end{aligned} \tag{5.33}$$

Noting that $\|[\boldsymbol{\Sigma}_{\mathrm{P}}^{n}]^{-1}[\boldsymbol{U}_{\mathrm{P}}^{n}]^{T}\bar{\boldsymbol{C}}[\mathbb{P}^{n}]^{T}\| = \|\boldsymbol{\Phi}\boldsymbol{V}_{\mathrm{P}}^{n}\| = 1$ and $\Delta t < |\alpha_{0}^{n}|/(|\beta_{0}^{n}|\kappa)$ yields

$$\begin{aligned}\|[\bar{\boldsymbol{V}}_{\mathrm{P}}^{n}]^{T}\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^{n}\| \le{}& \varepsilon_{0}^{n}\kappa\|[\bar{\boldsymbol{Z}}_{\mathrm{P}}^{n}]^{T}\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^{n}\| + \varepsilon_{\mathrm{P},0}^{n}\|(\boldsymbol{I}-\mathbb{P}^{n})\boldsymbol{f}(\boldsymbol{x}^{0}+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n})\| \\ &+ \frac{|\alpha_{0}^{n}|\|[\boldsymbol{\Sigma}_{\mathrm{P}}^{n}]^{-1}[\boldsymbol{U}_{\mathrm{P}}^{n}]^{T}\bar{\boldsymbol{C}}\Delta^{n}\|}{h^{n}}\|\hat{\boldsymbol{x}}_{\mathrm{P}}^{n}\| + \frac{1}{h^{n}}\|[\boldsymbol{\Sigma}_{\mathrm{P}}^{n}]^{-1}[\boldsymbol{U}_{\mathrm{P}}^{n}]^{T}\bar{\boldsymbol{C}}\delta\boldsymbol{r}_{\mathrm{P},\star}^{n-1}\|.\end{aligned} \tag{5.34}$$

Next, we estimate $\|[\boldsymbol{\Sigma}_{\mathrm{P}}^n]^{-1}[\boldsymbol{U}_{\mathrm{P}}^n]^T\bar{\boldsymbol{C}}\delta\boldsymbol{r}_{\mathrm{P},\star}^{n-1}\|$. First, we have from adding and subtracting $\sum_{\ell=1}^k\boldsymbol{\Psi}^n(\boldsymbol{\Phi}^T\boldsymbol{\Psi}^n)^{-1}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell}$ that

$$\begin{aligned}[\boldsymbol{\Sigma}_{\mathrm{P}}^n]^{-1}[\boldsymbol{U}_{\mathrm{P}}^n]^T\bar{\boldsymbol{C}}\delta\boldsymbol{r}_{\mathrm{P},\star}^{n-1} &= \sum_{\ell=1}^{k}\beta_\ell^n\Delta t[\boldsymbol{\Sigma}_{\mathrm{P}}^n]^{-1}[\boldsymbol{U}_{\mathrm{P}}^n]^T\bar{\boldsymbol{C}}\left([\mathbb{P}^n]^T\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{x}_\star^{n-\ell})-\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell})\right)\\ &\quad-\sum_{\ell=1}^{k}\alpha_\ell^n[\boldsymbol{\Sigma}_{\mathrm{P}}^n]^{-1}[\boldsymbol{U}_{\mathrm{P}}^n]^T\bar{\boldsymbol{C}}\left([\mathbb{P}^n]^T\boldsymbol{x}_\star^{n-\ell}-\boldsymbol{\Psi}^n(\boldsymbol{\Phi}^T\boldsymbol{\Psi}^n)^{-1}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell}+\Delta^n\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell}\right).\end{aligned}\tag{5.35}$$

Following similar steps to those above, adding and subtracting $\sum_{\ell=1}^k\beta_\ell^n\Delta t[\boldsymbol{\Sigma}_{\mathrm{P}}^n]^{-1}[\boldsymbol{U}_{\mathrm{P}}^n]^T\bar{\boldsymbol{C}}[\mathbb{P}^n]^T\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell})$ and applying the triangle inequality yields

$$\begin{aligned}\|[\boldsymbol{\Sigma}_{\mathrm{P}}^n]^{-1}[\boldsymbol{U}_{\mathrm{P}}^n]^T\bar{\boldsymbol{C}}\delta\boldsymbol{r}_{\mathrm{P},\star}^{n-1}\| &\le \sum_{\ell=1}^{k}|\beta_\ell^n|\Delta t\kappa(\|[\bar{\boldsymbol{V}}_{\mathrm{P}}^n]^T\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell}\|+\|[\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell}\|)\\ &\quad+\sum_{\ell=1}^{k}|\beta_\ell^n|\Delta t\|[\boldsymbol{\Sigma}_{\mathrm{P}}^n]^{-1}[\boldsymbol{U}_{\mathrm{P}}^n]^T\bar{\boldsymbol{C}}\|\|(\boldsymbol{I}-\mathbb{P}^n)\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell})\|\\ &\quad+\sum_{\ell=1}^{k}|\alpha_\ell^n|\|[\bar{\boldsymbol{V}}_{\mathrm{P}}^n]^T\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell}\|+\sum_{\ell=1}^{k}|\alpha_\ell^n|\|[\boldsymbol{\Sigma}_{\mathrm{P}}^n]^{-1}[\boldsymbol{U}_{\mathrm{P}}^n]^T\bar{\boldsymbol{C}}\Delta^n\|\|\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell}\|.\end{aligned}\tag{5.36}$$

Combining inequalities (5.34) and (5.36) yields the final result (5.29). The Galerkin counterpart (5.28) can be derived by following the same steps with the Galerkin operators; the main modification is that adding and subtracting $\sum_{\ell=0}^k\beta_\ell^n\Delta t\boldsymbol{\Sigma}_{\mathrm{G}}^{-1}\boldsymbol{U}_{\mathrm{G}}^T\bar{\boldsymbol{C}}\mathbb{V}\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{G}}^{n-\ell})$ is not needed in the Galerkin case. □

Lemma 5.2 shows that—as was the case with null-space error bounds in Lemma 5.1—the row-space error bounds are affected by the error incurred in the null space through the terms $\sum_{\ell=0}^k\varepsilon_\ell^n\kappa\|\bar{\boldsymbol{Z}}_{\mathrm{G}}^T\delta\hat{\boldsymbol{x}}_{\mathrm{G}}^{n-\ell}\|$ and $\sum_{\ell=0}^k\varepsilon_\ell^n\kappa\|[\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell}\|$ for conservative Galerkin and conservative LSPG projection, respectively. Further, these bounds are quite similar to the null-space error bounds with two exceptions. First, the projection-error term is multiplied by a constant, which is ζ_{G} in the case of conservative Galerkin projection and is ζ_{P}^n in the case of conservative LSPG projection; this constant arises from the fact that these bounds are derived from the discrete equations associated with subdomain conservation (5.11)–(5.12). We also note that the conservative LSPG row-space error bound employs the transpose of the typical LSPG projector, i.e., $[\mathbb{P}^n]^T$, which is the oblique projection onto $\mathrm{Ran}\left(\boldsymbol{\Psi}^n\right)$ orthogonal to $\mathrm{Ran}\left(\boldsymbol{\Phi}\right)$; this arises from appearance of the *transpose* of the constraints in the discrete equations (5.10). This also leads to the appearance of the term proportional to $\|\Delta^n\|$ in the conservative LSPG error bound.

Theorem 5.2 *(Local a posteriori error bounds). If* $\mathbf{A_1}$ *holds and* $\Delta t<|\alpha_0^n|/(|\beta_0^n|\kappa)$*, then*

$$\|\delta\hat{\boldsymbol{x}}_{\mathrm{G}}^n\|\le\sum_{\ell=0}^{k}(1+\zeta_{\mathrm{G}})\varepsilon_\ell^n\|(\boldsymbol{I}-\mathbb{V})\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{G}}^{n-\ell})\|+\sum_{\ell=1}^{k}\gamma_\ell^n\|\delta\hat{\boldsymbol{x}}_{\mathrm{G}}^{n-\ell}\|\tag{5.37}$$

$$\begin{aligned}\|\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^n\|&\le\sum_{\ell=0}^{k}\varepsilon_\ell^n\|(\boldsymbol{I}-\mathbb{P}^n)\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell})\|+\zeta_{\mathrm{P}}^n\sum_{\ell=0}^{k}\varepsilon_\ell^n\|(\boldsymbol{I}-[\mathbb{P}^n]^T\mathbb{P}^n)\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell})\|\\ &\quad+\frac{\zeta_{\mathrm{P}}^n}{h^n}\|\Delta^n\|\sum_{\ell=0}^{k}|\alpha_\ell^n|\|\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell}\|+\sum_{\ell=1}^{k}\gamma_\ell^n\|\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell}\|.\end{aligned}\tag{5.38}$$

Proof. Adding the premultiplication of Eq. (5.21) by $\bar{\boldsymbol{Z}}_{\mathrm{G}}$ to the premultiplication of Eq. (5.32) by $\bar{\boldsymbol{V}}_{\mathrm{P}}^n$ and noting that $[\bar{\boldsymbol{V}}_{\mathrm{P}}^n]^T=[\boldsymbol{\Sigma}_{\mathrm{P}}^n]^{-1}[\boldsymbol{U}_{\mathrm{P}}^n]^T\bar{\boldsymbol{C}}[\mathbb{P}^n]^T$ yields

$$\begin{aligned}\alpha_0^n\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^n&=\beta_0^n\Delta t\left[\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{x}_\star^n)-(\bar{\boldsymbol{Z}}_{\mathrm{P}}^n[\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\mathbb{P}^n+\bar{\boldsymbol{V}}_{\mathrm{P}}^n[\boldsymbol{\Sigma}_{\mathrm{P}}^n]^{-1}[\boldsymbol{U}_{\mathrm{P}}^n]^T\bar{\boldsymbol{C}})\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^n)\right]-\alpha_0^n\bar{\boldsymbol{V}}_{\mathrm{P}}^n[\boldsymbol{\Sigma}_{\mathrm{P}}^n]^{-1}[\boldsymbol{U}_{\mathrm{P}}^n]^T\bar{\boldsymbol{C}}\Delta^n\hat{\boldsymbol{x}}_{\mathrm{P}}^n\\ &\quad+\bar{\boldsymbol{Z}}_{\mathrm{P}}^n[\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\delta\boldsymbol{r}_{\mathrm{P}}^{n-1}+\bar{\boldsymbol{V}}_{\mathrm{P}}^n[\boldsymbol{\Sigma}_{\mathrm{P}}^n]^{-1}[\boldsymbol{U}_{\mathrm{P}}^n]^T\bar{\boldsymbol{C}}\delta\boldsymbol{r}_{\mathrm{P},\star}^{n-1}.\end{aligned}\tag{5.39}$$

Adding and subtracting $\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^n)+\bar{\boldsymbol{V}}_{\mathrm{P}}^n[\bar{\boldsymbol{V}}_{\mathrm{P}}^n]^T\mathbb{P}^n\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^n)$ from the bracketed quantity, using $[\bar{\boldsymbol{V}}_{\mathrm{P}}^n]^T[\bar{\boldsymbol{V}}_{\mathrm{P}}^n]^T\mathbb{P}^n=\bar{\boldsymbol{V}}_{\mathrm{P}}^n[\boldsymbol{\Sigma}_{\mathrm{P}}^n]^{-1}[\boldsymbol{U}_{\mathrm{P}}^n]^T\bar{\boldsymbol{C}}[\mathbb{P}^n]^T\mathbb{P}^n$, applying the triangle inequality, and using Lipschitz continuity yields

$$
\begin{aligned}
|\alpha_0^n|\|\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^n\| \leq{}& |\beta_0^n|\Delta t\left[\kappa\|\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^n\| + \|(\boldsymbol{I}-\mathbb{P}^n)\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^n)\| + \|[\boldsymbol{\Sigma}_{\mathrm{P}}^n]^{-1}[\boldsymbol{U}_{\mathrm{P}}^n]^T\bar{\boldsymbol{C}}\|\|(\boldsymbol{I}-[\mathbb{P}^n]^T\mathbb{P}^n)\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^n)\|\right] \\
&+ |\alpha_0^n|\|[\boldsymbol{\Sigma}_{\mathrm{P}}^n]^{-1}[\boldsymbol{U}_{\mathrm{P}}^n]^T\bar{\boldsymbol{C}}\Delta^n\|\|\hat{\boldsymbol{x}}_{\mathrm{P}}^n\| + \|\bar{\boldsymbol{Z}}_{\mathrm{P}}^n[\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\delta\boldsymbol{r}_{\mathrm{P}}^{n-1} + \bar{\boldsymbol{V}}_{\mathrm{P}}^n[\boldsymbol{\Sigma}_{\mathrm{P}}^n]^{-1}[\boldsymbol{U}_{\mathrm{P}}^n]^T\bar{\boldsymbol{C}}\delta\boldsymbol{r}_{\mathrm{P},\star}^{n-1}\|.
\end{aligned}
\tag{5.40}
$$

Now, using $\Delta t < |\alpha_0^n|/(|\beta_0^n|\kappa)$ yields

$$
\begin{aligned}
\|\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^n\| \leq{}& \varepsilon_0^n\|(\boldsymbol{I}-\mathbb{P}^n)\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^n)\| + \varepsilon_0^n\zeta_{\mathrm{P}}^n\|(\boldsymbol{I}-[\mathbb{P}^n]^T\mathbb{P}^n)\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^n)\| \\
&+ \frac{\zeta_{\mathrm{P}}^n}{h^n}\|\Delta^n\||\alpha_0^n|\|\hat{\boldsymbol{x}}_{\mathrm{P}}^n\| + \frac{1}{h^n}\|\bar{\boldsymbol{Z}}_{\mathrm{P}}^n[\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\delta\boldsymbol{r}_{\mathrm{P}}^{n-1} + \bar{\boldsymbol{V}}_{\mathrm{P}}^n[\boldsymbol{\Sigma}_{\mathrm{P}}^n]^{-1}[\boldsymbol{U}_{\mathrm{P}}^n]^T\bar{\boldsymbol{C}}\delta\boldsymbol{r}_{\mathrm{P},\star}^{n-1}\|.
\end{aligned}
\tag{5.41}
$$

Next, we estimate $\|\bar{\boldsymbol{Z}}_{\mathrm{P}}^n[\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\delta\boldsymbol{r}_{\mathrm{P}}^{n-1} + \bar{\boldsymbol{V}}_{\mathrm{P}}^n[\boldsymbol{\Sigma}_{\mathrm{P}}^n]^{-1}[\boldsymbol{U}_{\mathrm{P}}^n]^T\bar{\boldsymbol{C}}\delta\boldsymbol{r}_{\mathrm{P},\star}^{n-1}\|$.

$$
\bar{\boldsymbol{Z}}_{\mathrm{P}}^n[\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\delta\boldsymbol{r}_{\mathrm{P}}^{n-1} = \bar{\boldsymbol{Z}}_{\mathrm{P}}^n[\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\bar{\boldsymbol{r}}^n[\boldsymbol{x}_\star^{n-k},\ldots,\boldsymbol{x}_\star^{n-1}] - \bar{\boldsymbol{Z}}_{\mathrm{P}}^n[\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\mathbb{P}^n\bar{\boldsymbol{r}}^n[\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-k},\ldots,\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-1}]
\tag{5.42}
$$

$$
\begin{aligned}
\bar{\boldsymbol{V}}_{\mathrm{P}}^n[\boldsymbol{\Sigma}_{\mathrm{P}}^n]^{-1}[\boldsymbol{U}_{\mathrm{P}}^n]^T\bar{\boldsymbol{C}}\delta\boldsymbol{r}_{\mathrm{P},\star}^{n-1} &= \bar{\boldsymbol{V}}_{\mathrm{P}}^n[\boldsymbol{\Sigma}_{\mathrm{P}}^n]^{-1}[\boldsymbol{U}_{\mathrm{P}}^n]^T\bar{\boldsymbol{C}}[\mathbb{P}^n]^T\bar{\boldsymbol{r}}^n[\boldsymbol{x}_\star^{n-k},\ldots,\boldsymbol{x}_\star^{n-1}] - \bar{\boldsymbol{V}}_{\mathrm{P}}^n[\boldsymbol{\Sigma}_{\mathrm{P}}^n]^{-1}[\boldsymbol{U}_{\mathrm{P}}^n]^T\bar{\boldsymbol{C}}\bar{\boldsymbol{r}}^n[\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-k},\ldots,\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-1}] \\
&= \bar{\boldsymbol{V}}_{\mathrm{P}}^n[\bar{\boldsymbol{V}}_{\mathrm{P}}^n]^T\bar{\boldsymbol{r}}^n[\boldsymbol{x}_\star^{n-k},\ldots,\boldsymbol{x}_\star^{n-1}] - \bar{\boldsymbol{V}}_{\mathrm{P}}^n[\boldsymbol{\Sigma}_{\mathrm{P}}^n]^{-1}[\boldsymbol{U}_{\mathrm{P}}^n]^T\bar{\boldsymbol{C}}\bar{\boldsymbol{r}}^n[\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-k},\ldots,\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-1}].
\end{aligned}
\tag{5.43}
$$

Thus,

$$
\begin{aligned}
&\bar{\boldsymbol{Z}}_{\mathrm{P}}^n[\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\delta\boldsymbol{r}_{\mathrm{P}}^{n-1} + \bar{\boldsymbol{V}}_{\mathrm{P}}^n[\boldsymbol{\Sigma}_{\mathrm{P}}^n]^{-1}[\boldsymbol{U}_{\mathrm{P}}^n]^T\bar{\boldsymbol{C}}\delta\boldsymbol{r}_{\mathrm{P},\star}^{n-1} = \\
&\qquad\sum_{\ell=1}^{k}\beta_\ell^n\Delta t\left[\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{x}_\star^{n-\ell}) - (\bar{\boldsymbol{Z}}_{\mathrm{P}}^n[\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\mathbb{P}^n + \bar{\boldsymbol{V}}_{\mathrm{P}}^n[\boldsymbol{\Sigma}_{\mathrm{P}}^n]^{-1}[\boldsymbol{U}_{\mathrm{P}}^n]^T\bar{\boldsymbol{C}})\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell})\right] \\
&\qquad-\sum_{\ell=1}^{k}\alpha_\ell^n\left[\boldsymbol{x}_\star^{n-\ell} - (\bar{\boldsymbol{Z}}_{\mathrm{P}}^n[\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\mathbb{P}^n + \bar{\boldsymbol{V}}_{\mathrm{P}}^n[\boldsymbol{\Sigma}_{\mathrm{P}}^n]^{-1}[\boldsymbol{U}_{\mathrm{P}}^n]^T\bar{\boldsymbol{C}})\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell}\right].
\end{aligned}
\tag{5.44}
$$

Adding and subtracting $\sum_{\ell=1}^{k}\beta_\ell^n\Delta t(\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell}) + \bar{\boldsymbol{V}}_{\mathrm{P}}^n[\bar{\boldsymbol{V}}_{\mathrm{P}}^n]^T\mathbb{P}^n\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell})) + \sum_{\ell=1}^{k}\alpha_\ell^n\bar{\boldsymbol{V}}_{\mathrm{P}}^n\bar{\boldsymbol{V}}_{\mathrm{P}}^n\mathbb{P}^n\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell}$, using $\mathbb{P}^n\boldsymbol{\Phi} = \boldsymbol{\Phi}$, and applying the triangle inequality yields

$$
\begin{aligned}
&\|\bar{\boldsymbol{Z}}_{\mathrm{P}}^n[\bar{\boldsymbol{Z}}_{\mathrm{P}}^n]^T\delta\boldsymbol{r}_{\mathrm{P}}^{n-1} + \bar{\boldsymbol{V}}_{\mathrm{P}}^n[\boldsymbol{\Sigma}_{\mathrm{P}}^n]^{-1}[\boldsymbol{U}_{\mathrm{P}}^n]^T\bar{\boldsymbol{C}}\delta\boldsymbol{r}_{\mathrm{P},\star}^{n-1}\| \leq \\
&\qquad\sum_{\ell=1}^{k}|\beta_\ell^n|\Delta t\left[\kappa\|\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell}\| + \|(\boldsymbol{I}-\mathbb{P}^n)\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell})\| + \zeta_{\mathrm{P}}^n\|(\boldsymbol{I}-[\mathbb{P}^n]^T\mathbb{P}^n)\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell})\|\right] \\
&\qquad+\sum_{\ell=1}^{k}|\alpha_\ell^n|\left[\|\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell}\| + \zeta_{\mathrm{P}}^n\|\Delta^n\|\|\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell}\|\right].
\end{aligned}
\tag{5.45}
$$

Combining inequalities (5.41) and (5.45) yields the final result. The Galerkin result is derived similarly, except we note that $\mathbb{V}^T\mathbb{V} = \boldsymbol{I}$ and the Δ^n term associated with Galerkin projection is zero. □

Comparing Theorem 5.2 with [15, Theorem 6.1] shows that the state-space error bounds for the conservative Galerkin and conservative LSPG models are in general larger than the bounds for their unconstrained counterparts; the conservative Galerkin bound has the addition of the constant ζ_{G}, while the second and third terms in the conservative LSPG bound are added. This is to be expected, as the proposed models do not strictly minimize their associated residuals; they do so only subject to the satisfaction of nonlinear equality constraints. Thus, general state-space error bounds that are related to the full-space residual alone will lead to larger bounds. Instead, if we consider the components of the error associated with the constraints themselves, we can derive more favorable bounds.

Lemma 5.3 *(Local a posteriori error bounds in conserved quantities). The error in the conserved quantities can be bounded as*

$$
\|\bar{\boldsymbol{C}}\delta\hat{\boldsymbol{x}}_{\mathrm{G}}^n\| \leq \sum_{\ell=0}^{k}\frac{|\beta_\ell^n|\Delta t}{|\alpha_0^n|}\|\bar{\boldsymbol{C}}\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{x}_\star^{n-\ell}) - \bar{\boldsymbol{C}}\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{G}}^{n-\ell})\| + \sum_{\ell=1}^{k}\frac{|\alpha_\ell^n|}{|\alpha_0^n|}\|\bar{\boldsymbol{C}}\delta\hat{\boldsymbol{x}}_{\mathrm{G}}^{n-\ell}\|
\tag{5.46}
$$

$$
\|\bar{\boldsymbol{C}}\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^n\| \leq \sum_{\ell=0}^{k}\frac{|\beta_\ell^n|\Delta t}{|\alpha_0^n|}\|\bar{\boldsymbol{C}}\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{x}_\star^{n-\ell}) - \bar{\boldsymbol{C}}\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell})\| + \sum_{\ell=1}^{k}\frac{|\alpha_\ell^n|}{|\alpha_0^n|}\|\bar{\boldsymbol{C}}\delta\hat{\boldsymbol{x}}_{\mathrm{P}}^{n-\ell}\|.
\tag{5.47}
$$

Proof. The result can be obtained trivially by subtracting Eq. (5.11) from the premultiplication of Eq. (5.8) by $\bar{\boldsymbol{C}}$ and applying the triangle inequality. □

Lemma 5.3—while very simple—highlights an important attribute of the proposed methods. In particular, the new contribution to the error at time instance t^n is due to a term comprising a scalar multiple of $\|\bar{\boldsymbol{C}}\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{x}_\star^n)-\bar{\boldsymbol{C}}\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_\mathrm{G}^n)\|$ and $\|\bar{\boldsymbol{C}}\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{x}_\star^n)-\bar{\boldsymbol{C}}\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_\mathrm{P}^n)\|$ for conservative Galerkin and conservative LSPG, respectively. In the absence of source terms, this error associates with the *error in the flux along the faces $\bar{\mathcal{E}}$ of the decomposed mesh $\bar{\mathcal{M}}$*. In many cases, this error will be quite small. For example, in the case of global conservation characterized by $\bar{\mathcal{M}}=\bar{\mathcal{M}}_\mathrm{global}$, $N_{\bar{\Omega}}=1$, $\bar{\Omega}_1=\Omega$ and $\bar{\Gamma}_1=\Gamma$, the error in the globally conserved quantities arises entirely from the error in the flux computed along the boundary of the domain.

Theorem 5.3 *(Local a posteriori error bounds in conserved quantities). If $\mathbf{A_1}$ holds and $\Delta t<|\alpha_0^n|/(|\beta_0^n|\kappa\,\mathrm{cond}\left(\bar{\boldsymbol{C}}\right))$, then*

$$\|\bar{\boldsymbol{C}}\delta\hat{\boldsymbol{x}}_\mathrm{G}^n\|\leq\sum_{\ell=0}^{k}\bar{\varepsilon}_\ell^n\kappa\|\boldsymbol{Z}_{\bar{\boldsymbol{C}}}^T\delta\hat{\boldsymbol{x}}_\mathrm{G}^{n-\ell}\|+\sum_{\ell=1}^{k}\bar{\gamma}_\ell^n\|\bar{\boldsymbol{C}}\delta\hat{\boldsymbol{x}}_\mathrm{G}^{n-\ell}\| \tag{5.48}$$

$$\|\bar{\boldsymbol{C}}\delta\hat{\boldsymbol{x}}_\mathrm{P}^n\|\leq\sum_{\ell=0}^{k}\bar{\varepsilon}_\ell^n\kappa\|\boldsymbol{Z}_{\bar{\boldsymbol{C}}}^T\delta\hat{\boldsymbol{x}}_\mathrm{P}^{n-\ell}\|+\sum_{\ell=1}^{k}\bar{\gamma}_\ell^n\|\bar{\boldsymbol{C}}\delta\hat{\boldsymbol{x}}_\mathrm{P}^{n-\ell}\|, \tag{5.49}$$

where $\bar{\varepsilon}_\ell^m:=|\beta_\ell^m|\|\bar{\boldsymbol{C}}\|\Delta t/\bar{h}^m$, $\bar{\gamma}_\ell^m:=(|\alpha_\ell^m|+|\beta_\ell^m|\kappa\,\mathrm{cond}\left(\bar{\boldsymbol{C}}\right)\Delta t)/\bar{h}^m$, $\bar{h}^m:=|\alpha_0^m|-|\beta_0^m|\kappa\,\mathrm{cond}\left(\bar{\boldsymbol{C}}\right)\Delta t$.

Proof. Subtracting Eq. (5.11) from the premultiplication of Eq. (5.8) by $\bar{\boldsymbol{C}}$ yields

$$\alpha_0^n\bar{\boldsymbol{C}}\delta\hat{\boldsymbol{x}}_\mathrm{G}^n=\beta_0^n\Delta t\bar{\boldsymbol{C}}\left(\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{x}_\star^n)-\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_\mathrm{G}^n)\right)+\bar{\boldsymbol{C}}\delta\bar{\boldsymbol{r}}_\mathrm{G}^{n-1}, \tag{5.50}$$

where $\delta\bar{\boldsymbol{r}}_\mathrm{G}^{n-1}:=\bar{\boldsymbol{r}}^n[\boldsymbol{x}_\star^{n-k},\ldots,\boldsymbol{x}_\star^{n-1}]-\bar{\boldsymbol{r}}^n[\boldsymbol{\Phi}\hat{\boldsymbol{x}}_\mathrm{G}^{n-k},\ldots,\boldsymbol{\Phi}\hat{\boldsymbol{x}}_\mathrm{G}^{n-1}]$. Applying the triangle inequality yields

$$|\alpha_0^n|\|\bar{\boldsymbol{C}}\delta\hat{\boldsymbol{x}}_\mathrm{G}^n\|\leq|\beta_0^n|\Delta t\|\bar{\boldsymbol{C}}\|\|\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{x}_\star^n)-\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_\mathrm{G}^n)\|+\|\bar{\boldsymbol{C}}\delta\bar{\boldsymbol{r}}_\mathrm{G}^{n-1}\|. \tag{5.51}$$

Now, using Lipschitz continuity and $\boldsymbol{y}=\boldsymbol{V}_{\bar{\boldsymbol{C}}}\boldsymbol{V}_{\bar{\boldsymbol{C}}}^T\boldsymbol{y}+\boldsymbol{Z}_{\bar{\boldsymbol{C}}}\boldsymbol{Z}_{\bar{\boldsymbol{C}}}^T\boldsymbol{y}$ for all $\boldsymbol{y}\in\mathbb{R}^N$ with $\bar{\boldsymbol{C}}^+\bar{\boldsymbol{C}}=\boldsymbol{V}_{\bar{\boldsymbol{C}}}\boldsymbol{V}_{\bar{\boldsymbol{C}}}^T$, we have

$$|\alpha_0^n|\|\bar{\boldsymbol{C}}\delta\hat{\boldsymbol{x}}_\mathrm{G}^n\|\leq|\beta_0^n|\Delta t\|\bar{\boldsymbol{C}}\|\kappa\left(\|\bar{\boldsymbol{C}}^+\|\|\bar{\boldsymbol{C}}\delta\hat{\boldsymbol{x}}_\mathrm{G}^n\|+\|\boldsymbol{Z}_{\bar{\boldsymbol{C}}}^T\delta\hat{\boldsymbol{x}}_\mathrm{G}^n\|\right)+\|\bar{\boldsymbol{C}}\delta\bar{\boldsymbol{r}}_\mathrm{G}^{n-1}\|. \tag{5.52}$$

Now, using $\Delta t<|\alpha_0^n|/(|\beta_0^n|\kappa\,\mathrm{cond}\left(\bar{\boldsymbol{C}}\right))$ and $\mathrm{cond}\left(\bar{\boldsymbol{C}}\right)=\sigma_\mathrm{max}(\bar{\boldsymbol{C}})/\sigma_\mathrm{min}(\bar{\boldsymbol{C}})=\|\bar{\boldsymbol{C}}\|\|\bar{\boldsymbol{C}}^+\|$, we have

$$\|\bar{\boldsymbol{C}}\delta\hat{\boldsymbol{x}}_\mathrm{G}^n\|\leq\frac{|\beta_0^n|\Delta t\|\bar{\boldsymbol{C}}\|\kappa}{\bar{h}^n}\|\boldsymbol{Z}_{\bar{\boldsymbol{C}}}^T\delta\hat{\boldsymbol{x}}_\mathrm{G}^n\|+\frac{1}{\bar{h}^n}\|\bar{\boldsymbol{C}}\delta\bar{\boldsymbol{r}}_\mathrm{G}^{n-1}\|. \tag{5.53}$$

Next, we estimate $\|\bar{\boldsymbol{C}}\delta\bar{\boldsymbol{r}}_\mathrm{G}^{n-1}\|$. First, we have

$$\bar{\boldsymbol{C}}\delta\bar{\boldsymbol{r}}_\mathrm{G}^{n-1}=\sum_{\ell=1}^{k}\beta_\ell^n\Delta t\bar{\boldsymbol{C}}\left(\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{x}_\star^{n-\ell})-\boldsymbol{f}(\boldsymbol{x}^0+\boldsymbol{\Phi}\hat{\boldsymbol{x}}_\mathrm{G}^{n-\ell})\right)-\sum_{\ell=1}^{k}\alpha_\ell^n\bar{\boldsymbol{C}}\left(\boldsymbol{x}_\star^{n-\ell}-\boldsymbol{\Phi}\hat{\boldsymbol{x}}_\mathrm{G}^{n-\ell}\right). \tag{5.54}$$

Applying the triangle inequality and following the above steps yields

$$\|\bar{\boldsymbol{C}}\delta\bar{\boldsymbol{r}}_\mathrm{G}^{n-1}\|\leq\sum_{\ell=1}^{k}|\beta_\ell^n|\kappa\Delta t\left(\mathrm{cond}\left(\bar{\boldsymbol{C}}\right)\|\bar{\boldsymbol{C}}\delta\hat{\boldsymbol{x}}_\mathrm{G}^{n-\ell}\|+\|\bar{\boldsymbol{C}}\|\|\boldsymbol{Z}_{\bar{\boldsymbol{C}}}^T\delta\hat{\boldsymbol{x}}_\mathrm{G}^{n-\ell}\|\right)+\sum_{\ell=1}^{k}\alpha_\ell^n\|\bar{\boldsymbol{C}}\delta\hat{\boldsymbol{x}}_\mathrm{G}^{n-\ell}\|. \tag{5.55}$$

Combining inequalities (5.53) and (5.55) produces the final result. □

Theorem 5.3 shows that—at a given time instance t^n—the only new contribution to the error bound arises from the term $\bar{\varepsilon}_0^n\kappa\|\boldsymbol{Z}_{\bar{\boldsymbol{C}}}^T\delta\hat{\boldsymbol{x}}_\mathrm{G}^n\|$, which associates with error incurred in the null space to the constraint matrix $\bar{\boldsymbol{C}}$. Thus, even though the methods explicitly enforce conservation over subdomains, the actual values of those conserved variables may deviate from their full-order-model counterparts. This can be interpreted as a closure problem: the errors in the state component not restricted by the constraints can lead to errors in the state component restricted by the constraints, i.e., the conserved variables.

6. Numerical experiments

This section compares the performance of several reduced-order models on a parameterization of the quasi-1D Euler equations applied to supersonic flow in a converging–diverging nozzle.

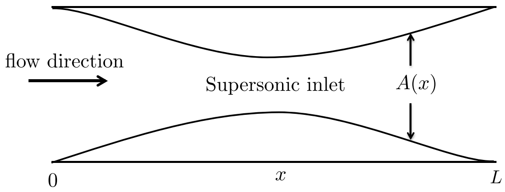


Fig. 2. *Quasi-1D Euler.* Problem geometry for the converging–diverging nozzle.

6.1. Problem description: quasi-1D Euler equation

We consider a parameterized quasi-1D Euler equation associated with modeling inviscid compressible flow in a one-dimensional converging–diverging nozzle with a continuously varying cross-sectional area [40, Chapter 13]; Fig. 2 depicts the problem geometry. In integral form, the governing equations are:

$$\begin{aligned}
&\frac{d}{dt}\int_{\omega} A(x)\rho(x,t;\boldsymbol{\mu})\,\mathrm{d}x && + \int_{\gamma} A(x)\rho(x,t;\boldsymbol{\mu})u(x,t;\boldsymbol{\mu})\mathrm{sign}(n(x))\,\mathrm{d}s(x) = 0\\
&\frac{d}{dt}\int_{\omega} A(x)\rho(x,t;\boldsymbol{\mu})u(x,t;\boldsymbol{\mu})\,\mathrm{d}x && + \int_{\gamma} A(x)\left(\rho(x,t;\boldsymbol{\mu})u(x,t;\boldsymbol{\mu})^2 + p(x,t;\boldsymbol{\mu})\right)\mathrm{sign}(n(x))\,\mathrm{d}s(x)\\
& && = \int_{\omega} p(x,t;\boldsymbol{\mu})\frac{\partial A}{\partial x}(x,t;\boldsymbol{\mu})\,\mathrm{d}x\\
&\frac{d}{dt}\int_{\omega} A(x)e(x,t;\boldsymbol{\mu})\,\mathrm{d}x && + \int_{\gamma} A(x)(e(x,t;\boldsymbol{\mu}) + p(x,t;\boldsymbol{\mu}))u(x,t;\boldsymbol{\mu})\mathrm{sign}(n(x))\,\mathrm{d}s(x) = 0,
\end{aligned}$$

$\forall\omega \subseteq \Omega = [0, L]$. Thus, the governing system of nonlinear partial differential equations is consistent with the conservation-law formulation in Eq. (2.1) with $d = 1$ spatial dimension, $n_u = 3$ conserved variables corresponding to density $u_1 = A\rho$, momentum $u_2 = A\rho u$, and energy density $u_3 = Ae$. The flux corresponds to $g_1 = A\rho u$, $g_2 = A(\rho u^2 + p)$, and $g_3 = A(e + p)u$, and the source corresponds to $s_1 = s_3 = 0$ and $s_2 = p\frac{\partial A}{\partial x}$. In addition, we have $p = (\gamma - 1)\rho\epsilon$, $\epsilon = \frac{e}{\rho} - \frac{u^2}{2}$, and assume a perfect gas (i.e., $p = \rho RT$). Here, ρ denotes density, u denotes velocity, p denotes pressure, ϵ denotes potential energy per unit mass, e denotes total energy density, γ denotes the specific heat ratio, and A denotes the converging–diverging nozzle cross-sectional area. We employ a specific heat ratio of $\gamma = 1.3$ and a specific gas constant of $R = 355.4$ m^2/s^2/K. The spatial domain is $\Omega = [0, L]$ with $L = 0.25$ m. The cross-sectional area $A(x)$ is determined by a cubic spline interpolation over the points

$$\begin{aligned}
(x, A(x)) \in \{&(0, 0.035), (0.0208, 0.0275), (0.0417, 0.0206), (0.0625, 0.0145), (0.0833, 0.0097),\\
&(0.104, 0.0066), (0.125, 0.0055), (0.146, 0.0067), (0.1667, 0.0107), (0.188, 0.0178),\\
&(0.208, 0.0283), (0.229, 0.0427), (0.25, 0.0612)\}.
\end{aligned} \tag{6.1}$$

The final time is $T = 0.29$ s, and we employ the backward Euler scheme with a uniform time step of $\Delta t = 0.01$ s for time discretization. We declare convergence of the Newton(-like) solver at each time instance when the ℓ^2-norm of the residual reaches 1×10^{-5} of its value with the initial guess, which is provided by the solution at the previous time instance.

The initial flow field is created in several steps. First, the following isentropic relations are used to generate a zero pressure-gradient flow field at the inlet ($x = 0$ m) and the outlet ($x = 0.25$ m):

$$M(x) = \frac{M_m A_m}{A(x)}\left(\frac{1 + \frac{\gamma-1}{2}M(x)^2}{1 + \frac{\gamma-1}{2}M_m^2}\right)^{\frac{\gamma+1}{2(\gamma-1)}}, \quad x \in \{0, 0.25\}\ \mathrm{m}, \tag{6.2}$$

where a subscript m indicates the flow quantity at $x = 0.125$ m, and M denotes the Mach number. The initial Mach number at the middle of the domain is employed as the problem parameter (i.e., $\mu = M_m$ with $n_\mu = 1$), from which the initial distribution of the Mach number is defined according to a cubic-spline interpolation with points $\{M(0), \mu, M(0.25)\}$. Then, we use the following relations to obtain the rest of the initial flow field for $x \in \Omega$:

$$p(x) = p_t\left(1 + \frac{\gamma - 1}{2}M(x)^2\right)^{\frac{-\gamma}{\gamma-1}}, \quad T(x) = T_t\left(1 + \frac{\gamma - 1}{2}M(x)^2\right)^{-1}, \tag{6.3}$$

$$\rho(x) = \frac{p(x)}{RT(x)}, \quad c(x) = \sqrt{\gamma \frac{p(x)}{\rho(x)}}, \quad u(x) = M(x)c(x), \tag{6.4}$$

where c denotes the speed of sound, the total temperature is $T_t = 2800$ K, and the total pressure is $p_t = 2.068 \times 10^6$ N/m^2.

6.2. Compared methods

These experiments compare the following methods, which employ the Tier A-B notation established in Section 4.5:

- *FOM*. This model corresponds to the full-order model, i.e., the solution satisfying Eq. (2.5).
- *Galerkin*. This model corresponds to the Tier 1-0 Galerkin ROM.
- *LSPG*. This model corresponds to the Tier 1-0 LSPG ROM.
- *LSPG-FV*. This model corresponds to the Tier 1-1 LSPG ROM, which is conservative.
- *GNAT*. This model corresponds to the Tier 2-0 LSPG ROM, where the residual approximation $\tilde{\boldsymbol{r}}$ is constructed using hyper-reduction method 1 with gappy POD as described in Section 3.3. The snapshots used to construct $\boldsymbol{\Phi}_r$ are constructed during simulation of the FOM method at training instances, i.e., the residual tensor $\mathcal{R}(\boldsymbol{x})$ is employed as described in Section 4.6. This corresponds to the GNAT method [16,17].
- *GNAT-FV*. This model corresponds to the Tier 2-1 LSPG ROM, which is conservative. While objective function is approximated in the same way as in the GNAT method above, no hyper-reduction is applied to the constraints.
- *GNAT-FV(X)*. This model corresponds to the Tier 2-2 LSPG ROM, which is *approximately* conservative. The objective function is approximated in the same way as in the GNAT method above. The residual approximation $\tilde{\tilde{\boldsymbol{r}}}$, which appears in the constraints, is approximated using hyper-reduction method 3 with gappy POD as described in Section 4.5. The snapshots used to construct the required reduced-basis matrices $\boldsymbol{\Phi}_s$ and $\boldsymbol{\Phi}_h$ are constructed from data tensors $\mathcal{S}(\boldsymbol{\xi})$ and $\mathcal{H}(\boldsymbol{\xi})$, where $\boldsymbol{\xi}$ corresponds to the Method 'X' state and X varies during the experiments. The sample matrices satisfy $\boldsymbol{P}_s = \boldsymbol{P}_r$ and $\boldsymbol{P}_r\boldsymbol{B} = \boldsymbol{P}_r\boldsymbol{B}\boldsymbol{P}_h^T\boldsymbol{P}_h$ as described in Section 4.6 such that a single sample mesh can be employed for all approximations.

In all cases that employ constraints (i.e., Tier A-B ROMs with B $\in \{1, 2\}$), the subdomains defining the decomposed mesh $\bar{\mathcal{M}}$ are equally spaced, their union is equal to the global domain (i.e., $\cup_{i=1}^{N_{\bar\Omega}} \bar\Omega_i = \Omega$), and are non-overlapping (i.e., $\mathrm{meas}(\bar\Omega_i \cap \bar\Omega_j) = 0$ for $i \neq j$) such that feasibility implies that all conservative ROMs are globally conservative (Corollary 4.2). If infeasibility is detected with $\bar N \leq p$, then infeasibility-handling approach 1 in Section 4.4 is employed at that time instance; this amounts to coarsening the decomposed mesh $\bar{\mathcal{M}}$ by reducing the number of subdomains by one and updating the operator $\bar{\boldsymbol{C}}$ accordingly; in all cases, global conservation was feasible. If instead $\bar N > p$, in which case feasibility cannot be guaranteed in general (see Proposition 5.1), then infeasibility-handling approach 2 in Section 4.4 is employed; this amounts to applying a penalty formulation with a specified penalty parameter $\rho \in \mathbb{R}_+$. In all cases, a (Newton) step length of $\eta^{n(k)} = 1$ was employed and led to convergence of the solution to the system of nonlinear equations arising at each time instance. All ROMs employ a training set of $\mathcal{D}_{\mathrm{train}} = \{1.7 + 0.1j\}_{j=0}^{3}$ such that $n_{\mathrm{train}} = 4$. The online parameter instance at which the ROMs are simulated is set to $\mu_\star = 1.75$.

We assess the accuracy of any ROM solution $\tilde{\boldsymbol{x}}$ using two metrics: the mean-squared and time-instantaneous state-space error, i.e.,

$$\mathcal{E}_{\boldsymbol{x}} := \sqrt{\sum_{n=1}^{N_T} \|\boldsymbol{x}^n(\boldsymbol{\mu}) - \tilde{\boldsymbol{x}}^n(\boldsymbol{\mu})\|_2^2} \Big/ \sqrt{\sum_{n=1}^{N_T} \|\boldsymbol{x}^n(\boldsymbol{\mu})\|_2^2} \tag{6.5}$$

$$\varepsilon_{\boldsymbol{x}}^n := \|\boldsymbol{x}^n(\boldsymbol{\mu}) - \tilde{\boldsymbol{x}}^n(\boldsymbol{\mu})\|_2 / \|\boldsymbol{x}^n(\boldsymbol{\mu})\|_2, \quad n = 1, \ldots, N_T, \tag{6.6}$$

and the mean-squared and time-instantaneous error in the globally conserved variables

$$\mathcal{E}_{\boldsymbol{x},\mathrm{global}} := \sqrt{\sum_{n=1}^{N_T} \|\bar{\boldsymbol{C}}_{\mathrm{global}}\boldsymbol{x}^n(\boldsymbol{\mu}) - \bar{\boldsymbol{C}}_{\mathrm{global}}\tilde{\boldsymbol{x}}^n(\boldsymbol{\mu})\|_2^2} \Big/ \sqrt{\sum_{n=1}^{N_T} \|\bar{\boldsymbol{C}}_{\mathrm{global}}\boldsymbol{x}^n(\boldsymbol{\mu})\|_2^2} \tag{6.7}$$

$$\varepsilon_{\boldsymbol{x},\mathrm{global}}^n := \|\bar{\boldsymbol{C}}_{\mathrm{global}}\boldsymbol{x}^n(\boldsymbol{\mu}) - \bar{\boldsymbol{C}}_{\mathrm{global}}\tilde{\boldsymbol{x}}^n(\boldsymbol{\mu})\|_2 / \|\bar{\boldsymbol{C}}_{\mathrm{global}}\boldsymbol{x}^n(\boldsymbol{\mu})\|_2, \quad n = 1, \ldots, N_T, \tag{6.8}$$

where $\bar{\boldsymbol{C}}_{\mathrm{global}} \in \mathbb{R}_+^{n_u \times N}$ is the operator $\bar{\boldsymbol{C}}$ associated with the global decomposition $\bar{\mathcal{M}} = \bar{\mathcal{M}}_{\mathrm{global}} := \{\Omega\}$. We also assess the mean-squared and time-instantaneous violation in global conservation as

$$\mathcal{E}_{\boldsymbol{r},\mathrm{global}} = \sqrt{\sum_{n=1}^{N_T} \|\bar{\boldsymbol{C}}_{\mathrm{global}}\boldsymbol{r}^n(\tilde{\boldsymbol{x}}^n(\boldsymbol{\mu}); \boldsymbol{\mu})\|_2} \tag{6.9}$$

$$\varepsilon_{\boldsymbol{r},\mathrm{global}}^n = \|\bar{\boldsymbol{C}}_{\mathrm{global}}\boldsymbol{r}^n(\tilde{\boldsymbol{x}}^n(\boldsymbol{\mu}); \boldsymbol{\mu})\|_2, \quad n = 1, \ldots, N_T. \tag{6.10}$$

All timings are obtained by performing calculations on an Intel(R) Xeon(R) CPU E5-2670 @ 2.60 GHz, 31.4 GB RAM using the `MORTestbed` [58] in Matlab.

6.3. GNAT-FV(X) snapshot study

This section assesses the effect of snapshot-collection method on the performance of the GNAT-FV(X) method; all subsequent experiments employ the snapshot-collection method yielding the best performance.

We set the number of control volumes to $N_\Omega = 100$ such that $N = N_\Omega n_u = 300$, the reduced-basis dimensions to $p = 5$ (which corresponds to a relative statistical energy of 99.78%) and $p_r = p_h$, and employ a sample mesh of 20 control volumes, which corresponds to $n_{p,r} = n_{p,f} = n_{p,h} = 60$. We employ a penalty parameter of $\rho = 10^3$, which is used by infeasibility-handling approach 2 when $\bar{N} > p$. In this setting we vary the number of constraints $\bar{N}$ and flux-basis dimension p_h and report the relative mean-squared violation in global conservation over the time interval, i.e., the value of $\mathcal{E}_{\boldsymbol{r},\text{global}}$ for the given reduced-order model divided by the value of $\mathcal{E}_{\boldsymbol{r},\text{global}}$ for the (unconstrained) GNAT model; note that this value is zero if the constraint-approximation error is zero and a feasible solution is computed at each time instance.

Fig. 3 reports the results for this experiment and elucidates several trends. First, Fig. 3a shows that the GNAT-FV model—for which the constraints are enforced exactly—yields near-exact satisfaction of the conservation laws for $\bar{N} < p$; this implies that a feasible solution was computed at every time instance of that simulation.

Second, we note that for $1 < \bar{N}/p < 2$, the GNAT-FV model yields approximate but accurate satisfaction of the conservation laws, as the relative value of $\mathcal{E}_{\boldsymbol{r},\text{global}}$ is less than 10^{-2} in these cases.

Third, Figs. 3b–3f show that the best results for the GNAT-FV(X) method are obtained for X = LSPG-FV (Fig. 3e) and X = GNAT-FV (Fig. 3f); these techniques yield relative values of $\mathcal{E}_{\boldsymbol{r},\text{global}}$ for the GNAT-FV(X) model less than 10^{-2} for $\bar{N}/p < 2$ in almost all cases. This result is sensible, as the training simulations corresponding to (constrained) LSPG-FV and GNAT-FV are 'closer' to the (constrained) GNAT-FV(X) simulation relative to the (unconstrained) FOM, LSPG, and GNAT simulations. However, in these cases, the relative value of $\mathcal{E}_{\boldsymbol{r},\text{global}}$ for $\bar{N} < p$ is small, but not close to machine zero as in the GNAT-FV case because the constraints are approximated. Thus, these methods—while having a cost independent of N due to the introduction of hyper-reduction—are only *approximately* conservative.

Fourth, we note that the GNAT-FV(LSPG-FV) and GNAT-FV(LSPG-FV) results are insensitive to the flux-basis dimension p_h for p_h sufficiently large ($p_h > 12$).

Finally, we note that while the GNAT-FV(LSPG-FV) and GNAT-FV(GNAT-FV) models yield similar accuracy, the latter method incurs a lower training cost, as the former incurs training simulations with the (Tier 1-1) LSPG-FV model, while the latter incurs training simulations with the (Tier 2-1) GNAT-FV model. Thus, the only GNAT-FV(X) method we consider in subsequent experiments is the GNAT-FV(GNAT-FV) approach.

6.4. Penalty-parameter study

This section assesses the effect of the penalty parameter ρ employed by infeasibility-handling approach 2 when $\bar{N} > p$ on the performance of the (constrained) ROMs LSPG-FV, GNAT-FV, and GNAT-FV(GNAT-FV). All subsequent experiments employ the penalty parameter yielding the best performance.

We again set the number of control volumes to $N_\Omega = 100$ such that $N = N_\Omega n_u = 300$, the reduced-basis dimensions to $p = 5$ and $p_r = p_h = p_s = 20$ and again employ a sample mesh of 20 control volumes, which corresponds to $n_{p,r} = n_{p,f} = n_{p,h} = 60$. We vary the number of constraints $\bar{N}$ and penalty parameter ρ and report the mean-squared state-space error $\mathcal{E}_{\boldsymbol{x}}$ and the (absolute) mean-squared violation in global conservation over the time interval $\mathcal{E}_{\boldsymbol{r},\text{global}}$. We note that a penalty value of $\rho = \infty$ corresponds to minimizing the norm of the constraints only (i.e., the objective function is ignored).

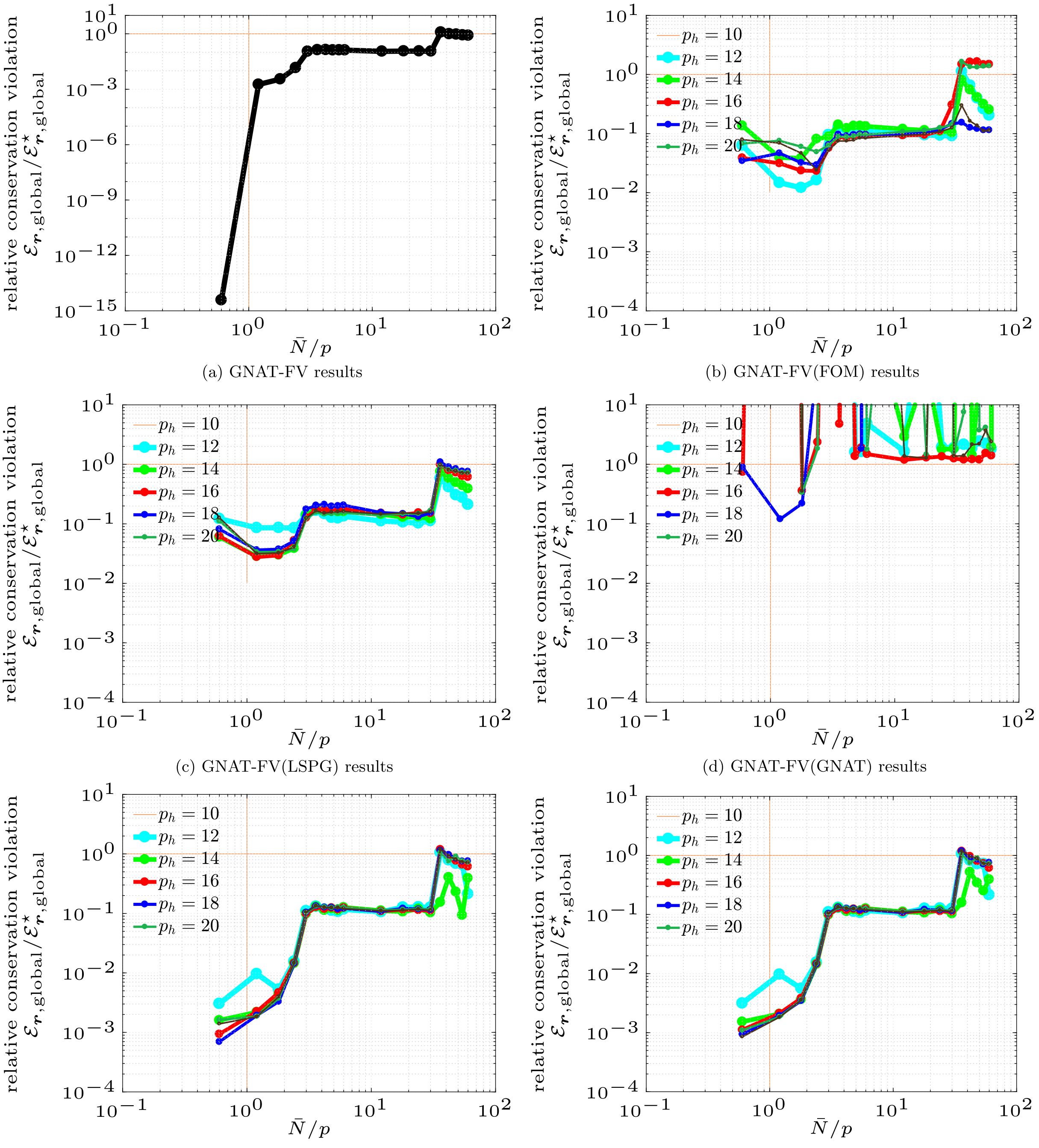
Fig. 4 reports the results for this experiment. First, we note that values $\rho \in \{10, 10^2, 10^3\}$ yield similar performance, which outperforms the other tested values. In particular, values of $\rho \in \{1, \infty\}$ often yield unstable responses, while $\rho = 10$ almost always yields larger errors than employing $\rho \in \{10, 10^2, 10^3\}$.

Second, we note that nearly all cases outperform the unconstrained model, characterized by $\rho = 0$; this implies that employing the proposed constraints can improve accuracy, even if the constraints are employed in a penalty formulation rather than as strictly enforced constraints.

Third, we observe that the two reported metrics are often correlated: larger values of mean-squared violation in global conservation $\mathcal{E}_{\boldsymbol{r},\text{global}}$ typically implies larger values of the relative mean-squared state-space error $\mathcal{E}_{\boldsymbol{x}}$. This lends credibility to the proposed technique, which aims to reduce the violation in global conservation, as it suggests that enforcing this constraint (or employing it as a penalty in the objective function) can lead to more accurate ROMs.

Fourth, the plots indicate that accuracy typically degrades as constraints are added to the problem, i.e., as the decomposed mesh becomes finer. In particular, the case $\bar{N} = N$ is equivalent to the unconstrained case for LSPG-FV for any value of the penalty parameter ρ, as $\bar{\boldsymbol{C}} = \boldsymbol{I}$ in this case and thus the objective function in Problem (4.63) is equal to a scalar multiple of the objective in the (unconstrained) LSPG Problem (4.44). This is not true for the GNAT-FV and GNAT-FV(GNAT-FV) models, as different hyper-reduction approaches are employed for the residual in the objective and constraints such that $\tilde{\tilde{\boldsymbol{r}}} \neq \tilde{\boldsymbol{r}}$.

In subsequent experiments, we employ a penalty-parameter value of $\rho = 10^3$.



(a) GNAT-FV results

(b) GNAT-FV(FOM) results

(c) GNAT-FV(LSPG) results

(d) GNAT-FV(GNAT) results

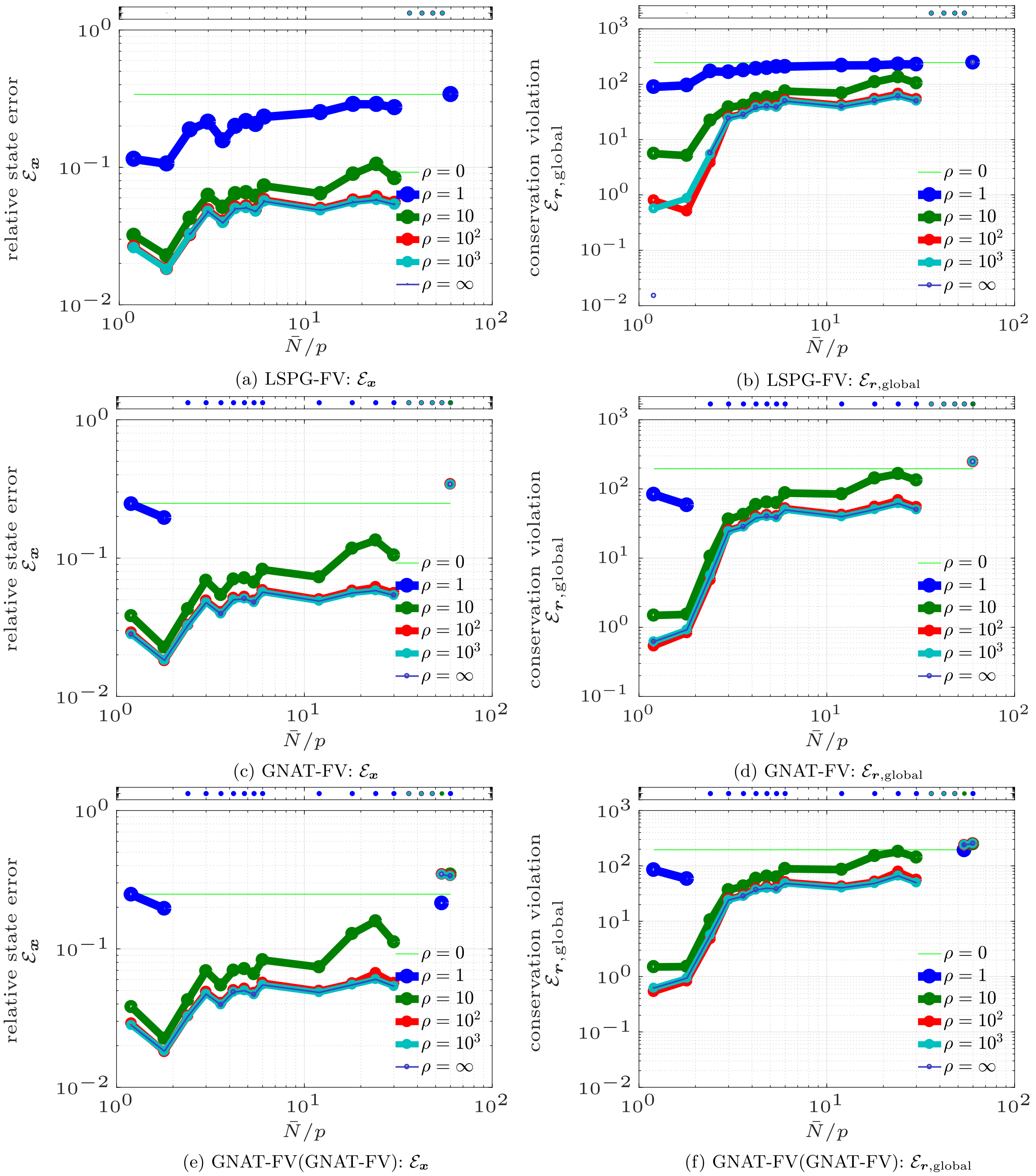
(e) GNAT-FV(LSPG-FV) results

(f) GNAT-FV(GNAT-FV) results

Fig. 3. *One-dimensional Euler equation.* GNAT-FV snapshot study described in Section 6.3. Fig. 3a corresponds to the (conservative) GNAT-FV model; other subfigures correspond to different (approximately conservative) GNAT-FV(X) methods, which employ Method 'X' snapshots to construct the required reduced-basis matrices $\boldsymbol{\Phi}_s$ and $\boldsymbol{\Phi}_h$. Within each subfigure, two parameters vary: the number of constraints $\bar{N}$ (the reduced-basis dimension is fixed to $p = 5$), and the dimension of the reduced-basis matrices $\boldsymbol{\Phi}_s$ and $\boldsymbol{\Phi}_h$, which are enforced to have the same dimension such that $p_r = p_h$. Here, $\mathcal{E}^{\star}_{\mathbf{r},\mathrm{global}}$ denotes the value of $\mathcal{E}_{\mathbf{r},\mathrm{global}}$ obtained for the unconstrained GNAT model.

6.5. State-basis-dimension study

This section assesses the effect of basis dimension p on the proposed methods. We again employ $N_\Omega = 100$ control volumes in the finite-volume discretization, set reduced-basis dimensions to $p_r = p_h = p_s = 20$, employ a sample mesh with 20 control volumes, and set the penalty parameter to $\rho = 10^3$. We vary both the state-basis dimension p and the



(a) LSPG-FV: $\mathcal{E}_{\boldsymbol{x}}$ (b) LSPG-FV: $\mathcal{E}_{\boldsymbol{r},\text{global}}$

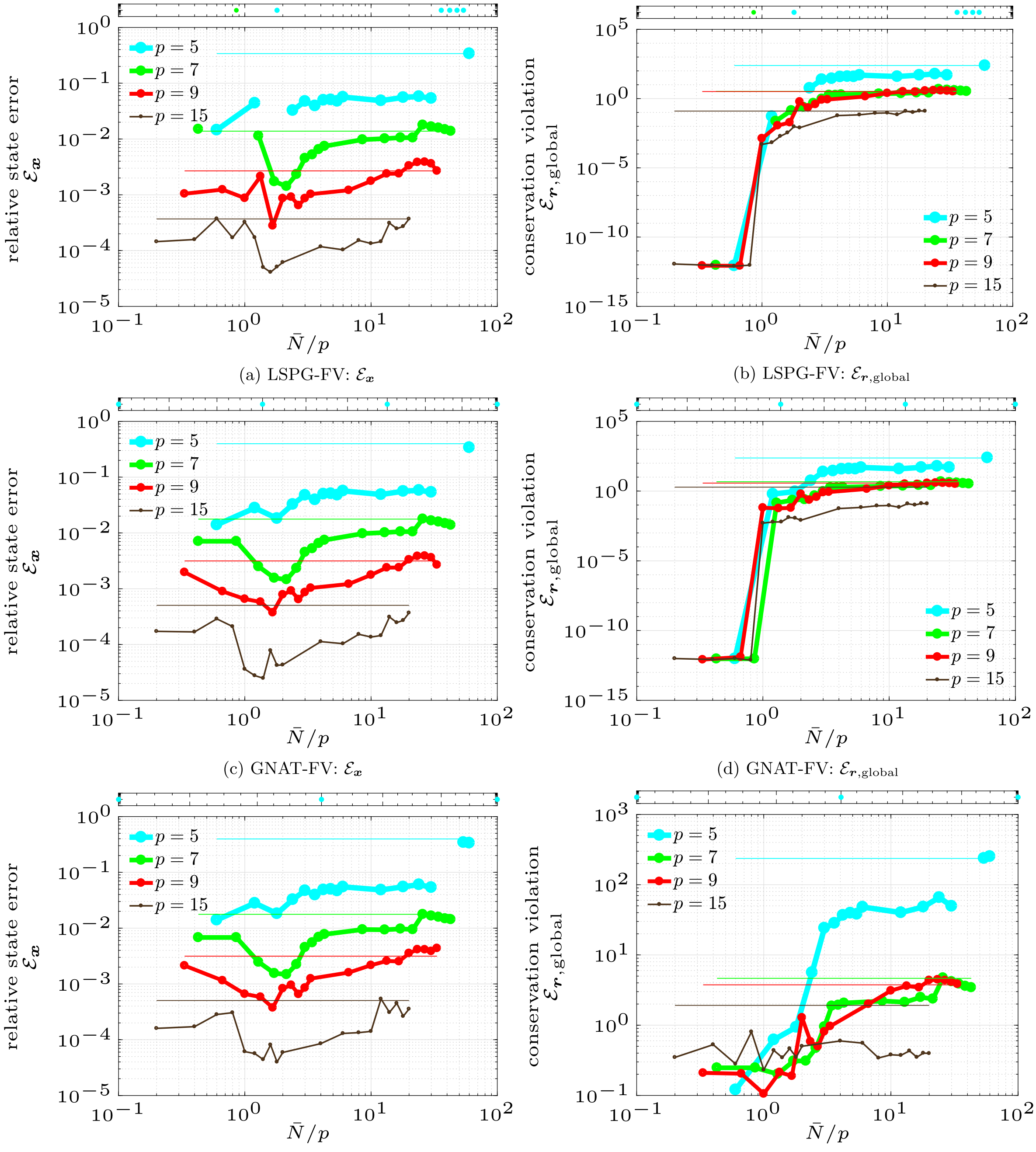
(c) GNAT-FV: $\mathcal{E}_{\boldsymbol{x}}$ (d) GNAT-FV: $\mathcal{E}_{\boldsymbol{r},\text{global}}$

(e) GNAT-FV(GNAT-FV): $\mathcal{E}_{\boldsymbol{x}}$ (f) GNAT-FV(GNAT-FV): $\mathcal{E}_{\boldsymbol{r},\text{global}}$

Fig. 4. *One-dimensional Euler equation.* GNAT-FV penalty-parameter study described in Section 6.4. The top bar reports cases where the reduced-order-model simulation was unstable. Each row of subfigures corresponds to a different ROM method; each column reports a different error measure. Within each subfigure, two parameters vary: the number of constraints $\bar{N}$ (the reduced-basis dimension is fixed to $p = 5$), and the penalty parameter ρ. Note that we only consider $\bar{N}/p > 1$, as infeasibility does not occur for $\bar{N}/p \leq 1$.

number of constraints $\bar{N}$ the relative mean-squared state-space error $\mathcal{E}_{\boldsymbol{x}}$ and the (absolute) mean-squared violation in global conservation over the time interval $\mathcal{E}_{\boldsymbol{r},\text{global}}$.

Fig. 5 reports the results. First, and most importantly, we note that Figs. 5a, 5c, and 5e show that the introduction of constraints yields the most significant improvements for the smallest basis dimension $p = 5$. In these cases, the relative mean-squared state-space error $\mathcal{E}_{\boldsymbol{x}}$ is reduced by over an order of magnitude for all ROMs, as the unconstrained ROMs yield errors exceeding 30%, while their constrained counterparts employing $\bar{N} = 3$ (i.e., global conservation with $\bar{\mathcal{M}} = \mathcal{M}_{\text{global}}$) all yield errors less than 2%. In contrast, for $p \geq 7$, the unconstrained ROMs are already quite accurate, with errors already



(a) LSPG-FV: $\mathcal{E}_{\boldsymbol{x}}$

(b) LSPG-FV: $\mathcal{E}_{\boldsymbol{r},\text{global}}$

(c) GNAT-FV: $\mathcal{E}_{\boldsymbol{x}}$

(d) GNAT-FV: $\mathcal{E}_{\boldsymbol{r},\text{global}}$

(e) GNAT-FV(GNAT-FV): $\mathcal{E}_{\boldsymbol{x}}$

(f) GNAT-FV(GNAT-FV): $\mathcal{E}_{\boldsymbol{r},\text{global}}$

Fig. 5. *One-dimensional Euler equation.* GNAT-FV state-basis-dimension study described in Section 6.5. The top bar reports cases where the reduced-order-model simulation was unstable. The colored horizontal lines correspond to the associated unconstrained ROM. Each row of subfigures corresponds to a different ROM method; each column reports a different error measure. Within each subfigure, two parameters vary: the number of constraints $\bar{N}$ and the reduced-basis dimension $p = 5$).

less than 2%; incorporating constraints in these cases does yield accuracy improvements in most cases, although these improvements are less dramatic. Because the most significant improvements were obtained by enforcing global conservation with $\bar{N} = 3$, subsequent experiments employ ROMs that enforce global conservation by using a decomposed mesh of $\bar{\mathcal{M}} = \bar{\mathcal{M}}_{\text{global}}$.

Second, Figs. 5b and 5d show that the LSPG-FV and GNAT-FV models produce near-exact satisfaction of the conservation laws for $\bar{N} < p$; this implies that a feasible solution was computed at every time instance of the corresponding simulations. In contrast, Fig. 5f shows that the GNAT-FV(GNAT-FV) ROM is only approximately conservative. Nonetheless, this approximate conservation does not adversely impact the actual errors produced by the ROM, as the errors reported in Figs. 5c and 5e are nearly identical in all cases. So, while applying hyper-reduction to the constraints results in a loss of numerically exact satisfaction of global conservation, the results are extremely similar to the case where the constraints are applied exactly.

6.6. Comparison across all methods

This section assesses the relative performance of the methods over time; all ROMs that employ constraints enforce global conservation, i.e., $\bar{N} = 3$ and $\bar{\mathcal{M}} = \mathcal{M}_{\text{global}}$.

We consider two discretizations corresponding to $N_\Omega = 500$ and $N_\Omega = 1000$ control volumes in the finite-volume discretization. We set reduced-basis dimensions to $p = 5$ and $p_r = p_h = p_s = 20$ and employ a sample mesh with 20 control volumes. We report the time-instantaneous state-space errors $\varepsilon^n_{\boldsymbol{x}}$, $n \in \mathbb{N}(N_T)$, errors in the globally conserved variables $\varepsilon^n_{\boldsymbol{x},\text{global}}$, $n \in \mathbb{N}(N_T)$, and violation in global conservation $\varepsilon^n_{\boldsymbol{r},\text{global}}$, $n \in \mathbb{N}(N_T)$.

Fig. 6 reports the results. First, we note that the errors $\varepsilon^n_{\boldsymbol{x}}$ and $\varepsilon^n_{\boldsymbol{x},\text{global}}$ exhibit the same trends in all cases; this suggests that enforcing global conservation—which leads to lower errors in the globally conserved quantities by construction—is an effective approach for also reducing the error in the state itself. This also supports previous observations that enforcing global conservation rather than employing a penalty approach leads to smaller errors in most cases.

Second, we observe that the FOM, LSPG-FV, and GNAT-FV models all lead to global-conservation violations $\varepsilon^n_{\boldsymbol{x},\text{global}}$ near zero as expected. In contrast, the GNAT-FV(GNAT-FV) approach only approximately satisfies global conservation due the introduction of hyper-reduction to the constraints; however, this has no noticeable effect on its response, as the errors reported for GNAT-FV and GNAT-FV(GNAT-FV) are nearly identical in Figs. 6a–6d.

Third, we notice that the conservative methods LSPG-FV and GNAT-FV, as well as the approximately conservative method GNAT-FV(GNAT-FV), all yield significantly lower errors than the unconstrained methods Galerkin, LSPG, and GNAT. Further, these unconstrained methods yield significant violation in global conservation.

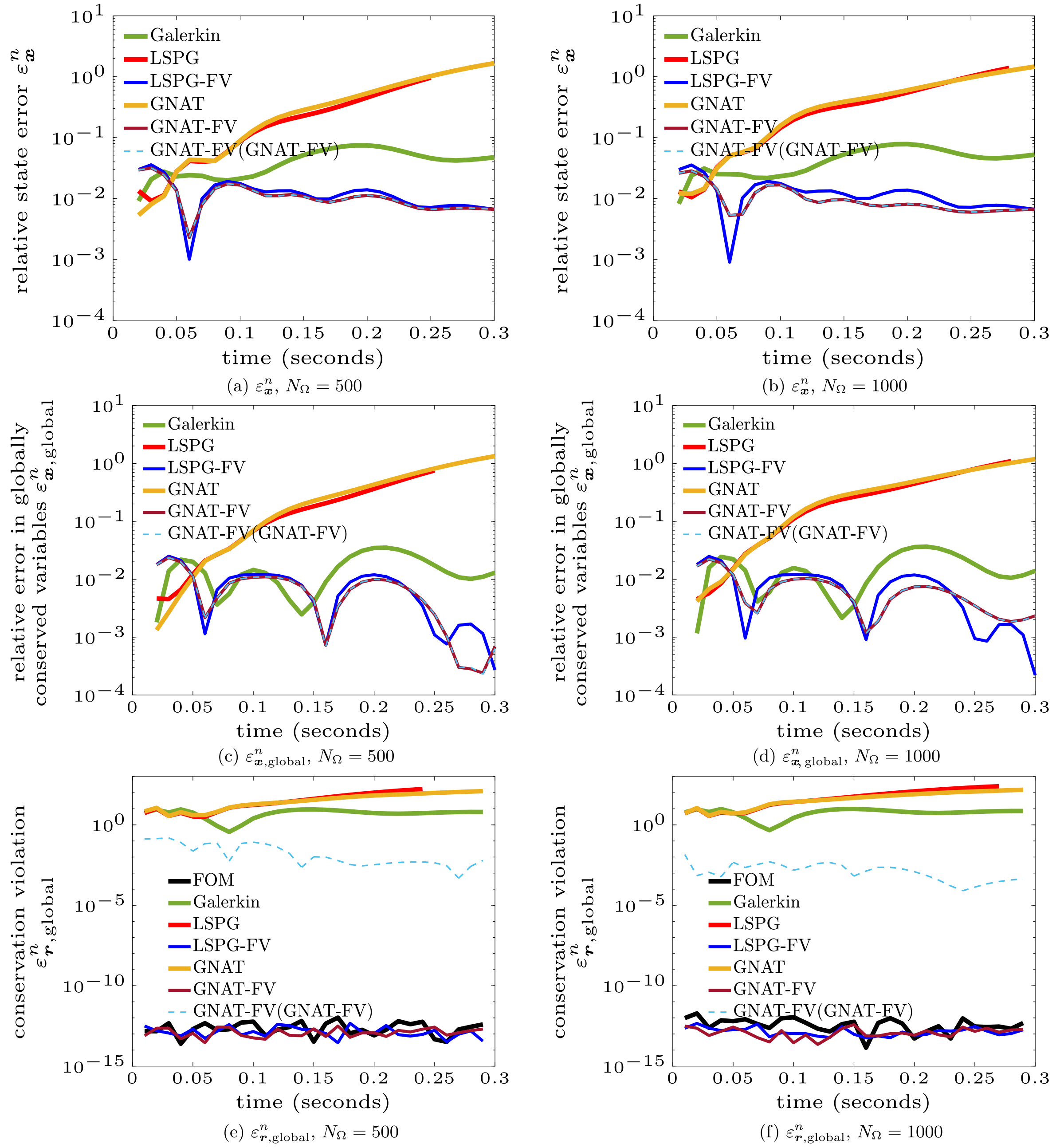
Table 1 reports the timings for these methods. We first note that the LSPG ROM does not have a valid timing for either problem, as the associated simulations yield negative pressures and thus do not successfully run for the entire time interval (see premature termination in Fig. 6). Second, while all other ROMs produce a speedup relative to the FOM, methods that employ hyper-reduction for the objective function (GNAT, GNAT-FV) produce more significant speedups; further applying hyper-reduction to the constraint (GNAT-FV(GNAT-FV)) improves the speedup further.

To enable an objective comparison of the ROM methods, we compare their performance across a wide variation of all method parameters. We subject each model to a parameter study wherein each model parameter is varied between the limits specified in Table 2. From these results, we then construct a Pareto front for each method, which is characterized by the method parameters that minimize the competing objectives of error and wall time.

Fig. 7 reports these Pareto fronts, where both the mean-squared state-space error $\mathcal{E}_{\boldsymbol{x}}$ and mean-squared violation in global conservation $\mathcal{E}_{\boldsymbol{r},\text{global}}$ are considered as error measures, as well as an 'overall' Pareto front that selects the Pareto-optimal methods across all parameter variations. Note that this figure reports the relative wall time with respect to that of the FOM simulation; relative wall times less than one imply the ROM yields a speedup. Here, Fig. 7a shows that the GNAT-FV(GNAT-FV) method is always Pareto dominant for error measure $\mathcal{E}_{\boldsymbol{x}}$, as no other method is both less expensive and more accurate for any tested parameter combination. The method that performs second best is the proposed GNAT-FV method, which exactly enforces constraints; note that it is only slightly more expensive than the GNAT-FV(GNAT-FV) method, as the benefit of performing hyper-reduction on the residual appearing in the constraints with $\bar{N}$ small is much less significant than the benefit of performing hyper-reduction on the residual appearing in the objective function when $\bar{N}$ is small. In particular, note that the Pareto-optimal parameter combinations for the LSPG-FV method yield similar accuracy to the Pareto-optimal GNAT-FV(GNAT-FV) points, but incur significantly larger wall times. Fig. 7b shows that GNAT-FV(GNAT-FV) is Pareto optimal for error measure $\mathcal{E}_{\boldsymbol{r},\text{global}}$ for relative wall times less than 0.28, but GNAT-FV, which enforces constraints exactly, is Pareto optimal for larger relative wall times, yielding near-zero violations in global conservation. We emphasize that both conservative variants of the GNAT method (i.e., GNAT-FV and GNAT-FV(GNAT-FV)) outperform the original GNAT approach, and the conservative variant of the LSPG method (i.e., LSPG-FV) outperforms the original LSPG method; this demonstrates the benefit of the proposed method and the performance improvement gained by enforcing conservation. In particular, note that the introduction of constraints does not adversely affect ROM wall-time performance; in fact, LSPG-FV has *better* wall-time performance relative to the LSPG method. This occurs because global conservation corresponds to only $\bar{N} = 3$ constraints in this case, and because these constraints lead to improved accuracy and thus promote convergence, the associated simulations require fewer iterations to solve the optimization problem at each time instance. We also note that hyper-reduction is needed to realize significant speedups: Pareto-optimal parameter combinations for ROMs employing hyper-reduction lead to relative wall times less than 0.36, while Pareto-optimal parameter combinations for ROMs without hyper-reduction yield relative wall times exceeding 0.6.

7. Conclusions

This work proposed two model-reduction methods for finite-volume models that enforce conservation over subdomains: conservative Galerkin and conservative LSPG projection. These methods associate with optimization problems characterized



(a) $\varepsilon_{\boldsymbol{x}}^n$, $N_\Omega = 500$

(b) $\varepsilon_{\boldsymbol{x}}^n$, $N_\Omega = 1000$

(c) $\varepsilon_{\boldsymbol{x},\text{global}}^n$, $N_\Omega = 500$

(d) $\varepsilon_{\boldsymbol{x},\text{global}}^n$, $N_\Omega = 1000$

(e) $\varepsilon_{\boldsymbol{r},\text{global}}^n$, $N_\Omega = 500$

(f) $\varepsilon_{\boldsymbol{r},\text{global}}^n$, $N_\Omega = 1000$

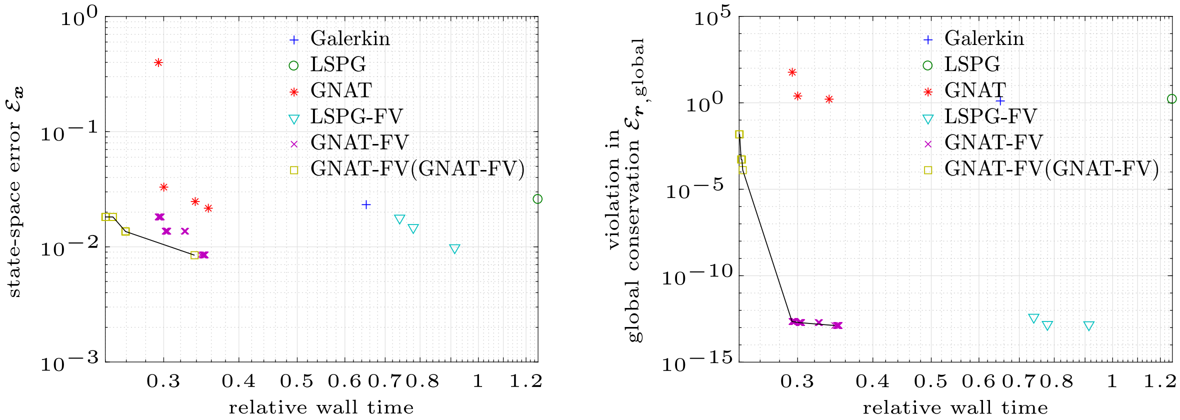
Fig. 6. *One-dimensional Euler equation*. Method comparison over time as described in Section 6.6, with conservative methods enforcing global conservation. Each curve depicts the time evolution of a given error measure for a given ROM. Each row of subfigures corresponds to a different error measure; the left and right subfigure columns correspond to cases $N_\Omega = 500$ and $N_\Omega = 1000$, respectively. Note that the missing data for the LSPG method corresponds to time instances after a negative pressure was generated, thus causing the simulation to end.

Table 1
One-dimensional Euler equation. Timings for the ROM methods assessed in Section 6.6. Here, all ROMs that employ constraints enforce global conservation, i.e., $\bar{N} = 3$ and $\bar{\mathcal{M}} = \bar{\mathcal{M}}_{\text{global}}$. We set reduced-basis dimensions to $p = 5$ and $p_r = p_h = p_s = 20$. ROM methods that use hyper-reduction employ a sample mesh with 20 control volumes.

Method	FOM	Galerkin	LSPG	LSPG-FV	GNAT	GNAT-FV	GNAT-FV(GNAT-FV)
wall time (seconds) for $N_\Omega = 500$	40.9	29.1	N/A	31.9	13.1	11.1	8.1
wall time (seconds) for $N_\Omega = 1000$	81.2	52.2	N/A	58.6	19.8	18.4	14.4

Table 2
One-dimensional Euler equation. Parameters varied for different ROMs to generate the Pareto fronts reported in Fig. 7 as described in Section 6.6.

Method	LSPG, LSPG-FV	GNAT, GNAT-FV, GNAT-FV(GNAT-FV)
p	$\{4, 5, 6\}$	$\{4, 5, 6\}$
$n_{p,r} = n_{p,s} = n_{p,h}$		$\{60, 90\}$
p_r		$\{10, 20, 30\}$
p_h		$\{10 + 5j\}_{j=0}^{4}$
p_s		$\{10 + 5j\}_{j=0}^{4}$



(a) Pareto front with error measure $\mathcal{E}_{\boldsymbol{x}}$

(b) Pareto front with error measure $\mathcal{E}_{\boldsymbol{r},\text{global}}$

Fig. 7. *One-dimensional Euler equation*. Pareto-optimal performance of various methods after varying model parameters reported in Table 2 for $N_\Omega = 500$ as described in Section 6.6. Wall times are reported relative to that of the FOM simulation. Note that the Pareto-optimal ROM methods in terms of minimizing error and wall time are the proposed GNAT-FV and GNAT-FV(GNAT-FV) methods.

by a minimum-residual objective function and nonlinear equality constraints formulated at the time-continuous and time-discrete levels, respectively. We equipped these methods with techniques for handling infeasible constraints, and we also developed hyper-reduction methods to ensure low-cost ROM simulations in the presence of nonlinear flux or source terms.

We performed analysis that demonstrated commutativity of conservative Galerkin projection and time discretization, developed sufficient conditions for feasibility, demonstrated conditions under which conservative Galerkin and conservative LSPG models are equivalent, and derived *a posteriori* error bounds. Numerical experiments on a model problem highlighted the benefit of conservative projection, and also demonstrated that enforcing global conservation led to the most accurate results.

Future work involves implementing the proposed techniques in a production-level computational fluid-dynamics code, demonstrating the methods on truly large-scale finite-volume models, and investigating combining the methodology with space–time projection approaches [56,11,20], as these techniques have demonstrated error bounds that grow slowly in time.

Acknowledgements

We thank Matthew Barone and Irina Tezaur for insightful conversations related to structure-preserving model reduction in fluid dynamics. We also thank J. Nathan Kutz for his help in forging the collaboration. This work was funded by Sandia's Laboratory Directed Research and Development program under Project #190968. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.

References

[1] R. Abgrall, D. Amsallem, R. Crisonovan, Robust model reduction by L^1-norm minimization and approximation via dictionaries: application to nonlinear hyperbolic problems, Adv. Model. Simul. Eng. Sci. 3 (2016) 1–16.

[2] S. An, T. Kim, D. James, Optimizing cubature for efficient integration of subspace deformations, ACM Trans. Graph. (TOG) 27 (2008) 165.

[3] H. Antil, S. Field, F. Herrmann, R. Nochetto, M. Tiglio, Two-step greedy algorithm for reduced order quadratures, J. Sci. Comput. 57 (2013) 604–637.

[4] H. Antil, M. Heinkenschloss, D.C. Sorensen, Application of the discrete empirical interpolation method to reduced order modeling of nonlinear and parametric systems, in: A. Quarteroni, G. Rozza (Eds.), Reduced Order Methods for Modeling and Computational Reduction, in: Springer MS&A, vol. 8, Springer-Verlag Italia, Milano, 2013.

[5] P. Astrid, S. Weiland, K. Willcox, T. Backx, Missing point estimation in models described by proper orthogonal decomposition, IEEE Trans. Autom. Control 53 (2008) 2237–2251.

[6] N. Aubry, P. Holmes, J.L. Lumley, E. Stone, The dynamics of coherent structures in the wall region of a turbulent boundary layer, J. Fluid Mech. 192 (1988) 115–173.
[7] M. Balajewicz, E. Dowell, Stabilization of projection-based reduced order models of the Navier–Stokes equations, Nonlinear Dyn. 70 (2012) 1619–1632.
[8] M. Balajewicz, E. Dowell, B. Noack, Low-dimensional modelling of high-Reynolds-number shear flows incorporating constraints from the Navier–Stokes equation, J. Fluid Mech. 729 (2013) 285–308.
[9] M.F. Barone, I. Kalashnikova, D.J. Segalman, H.K. Thornquist, Stable Galerkin reduced order models for linearized compressible flow, J. Comput. Phys. 228 (2009) 1932–1946.
[10] M. Barrault, Y. Maday, N.C. Nguyen, A.T. Patera, An 'empirical interpolation' method: application to efficient reduced-basis discretization of partial differential equations, C. R. Math. Acad. Sci. 339 (2004) 667–672.
[11] M. Baumann, P. Benner, J. Heiland, Space–time Galerkin POD with application in optimal control of semi-linear parabolic partial differential equations, arXiv preprint arXiv:1611.04050, 2016.
[12] M. Bergmann, C.-H. Bruneau, A. Iollo, Enablers for robust POD models, J. Comput. Phys. 228 (2009) 516–538.
[13] R. Bos, X. Bombois, P. Van den Hof, Accelerating large-scale non-linear models for monitoring and control using spatial and temporal correlations, in: Proceedings of the American Control Conference, vol. 4, 2004, pp. 3705–3710.
[14] K. Carlberg, Adaptive h-refinement for reduced-order models, Int. J. Numer. Methods Eng. 102 (2015) 1192–1210.
[15] K. Carlberg, M. Barone, H. Antil, Galerkin v. least-squares Petrov–Galerkin projection in nonlinear model reduction, J. Comput. Phys. 330 (2017) 693–734.
[16] K. Carlberg, C. Bou-Mosleh, C. Farhat, Efficient non-linear model reduction via a least-squares Petrov–Galerkin projection and compressive tensor approximations, Int. J. Numer. Methods Eng. 86 (2011) 155–181.
[17] K. Carlberg, C. Farhat, J. Cortial, D. Amsallem, The GNAT method for nonlinear model reduction: effective implementation and application to computational fluid dynamics and turbulent flows, J. Comput. Phys. 242 (2013) 623–647.
[18] K. Carlberg, R. Tuminaro, P. Boggs, Preserving Lagrangian structure in nonlinear model reduction with application to structural dynamics, SIAM J. Sci. Comput. 37 (2015) B153–B184.
[19] S. Chaturantabut, D.C. Sorensen, Nonlinear model reduction via discrete empirical interpolation, SIAM J. Sci. Comput. 32 (2010) 2737–2764.
[20] Y. Choi, K. Carlberg, Space–time least-squares Petrov–Galerkin projection for nonlinear model reduction, arXiv preprint arXiv:1703.04560, 2017.
[21] A. Deane, I. Kevrekidis, G.E. Karniadakis, S. Orszag, Low-dimensional models for complex geometry flows: application to grooved channels and circular cylinders, Phys. Fluids A, Fluid Dyn. 3 (1991) 2337–2354.
[22] M. Drohmann, B. Haasdonk, M. Ohlberger, Reduced basis approximation for nonlinear parametrized evolution equations based on empirical operator interpolation, SIAM J. Sci. Comput. 34 (2012) A937–A969.
[23] R. Everson, L. Sirovich, Karhunen–Loève procedure for gappy data, J. Opt. Soc. Am. A 12 (1995) 1657–1664.
[24] F. Fang, C. Pain, I. Navon, A. Elsheikh, J. Du, D. Xiao, Non-linear Petrov–Galerkin methods for reduced order hyperbolic equations and discontinuous finite element methods, J. Comput. Phys. 234 (2013) 540–559.
[25] C. Farhat, P. Avery, T. Chapman, J. Cortial, Dimensional reduction of nonlinear finite element dynamic models with finite rotations and energy-based mesh sampling and weighting for computational efficiency, Int. J. Numer. Methods Eng. 98 (2014) 625–662.
[26] L. Fick, Y. Maday, A.T. Patera, T. Taddei, A reduced basis technique for long-time unsteady turbulent flows, arXiv preprint arXiv:1710.03569, 2017.
[27] D. Galbally, K. Fidkowski, K. Willcox, O. Ghattas, Non-linear model reduction for uncertainty quantification in large-scale inverse problems, Int. J. Numer. Methods Eng. 81 (2009) 1581–1608.
[28] B. Galletti, C. Bruneau, L. Zannetti, A. Iollo, Low-order modelling of laminar flow regimes past a confined square cylinder, J. Fluid Mech. 503 (2004) 161–170.
[29] J.-F. Gerbeau, D. Lombardi, Approximated Lax pairs for the reduced order integration of nonlinear evolution equations, J. Comput. Phys. 265 (2014) 246–269.
[30] B. Haasdonk, M. Ohlberger, Reduced basis method for explicit finite volume approximations of nonlinear conservation laws, in: Proc. 12th International Conference on Hyperbolic Problems: Theory, Numerics, Application, 2008.
[31] B. Haasdonk, M. Ohlberger, Reduced basis method for finite volume approximations of parametrized linear evolution equations, ESAIM: Math. Model. Numer. Anal. 42 (2008) 277–302.
[32] B. Haasdonk, M. Ohlberger, G. Rozza, A reduced basis method for evolution schemes with parameter-dependent explicit operators, Electron. Trans. Numer. Anal. 32 (2008) 145–161.
[33] P. Holmes, J. Lumley, G. Berkooz, Turbulence, Coherent Structures, Dynamical Systems and Symmetry, Cambridge University Press, 1996.
[34] A. Iollo, S. Lanteri, J.A. Desideri, Stability properties of POD-Galerkin approximations for the compressible Navier–Stokes equations, Theor. Comput. Fluid Dyn. 13 (2000) 377–396.
[35] M. Jolly, I. Kevrekidis, E. Titi, Preserving dissipation in approximate inertial forms for the Kuramoto–Sivashinsky equation, J. Dyn. Differ. Equ. 3 (1991) 179–197.
[36] I. Kalashnikova, M. Barone, On the stability and convergence of a Galerkin reduced order model (rom) of compressible flow with solid wall and far-field boundary treatment, Int. J. Numer. Methods Eng. 83 (2010) 1345–1375.
[37] P.A. LeGresley, Application of Proper Orthogonal Decomposition (POD) to Design Decomposition Methods, PhD thesis, Stanford University, 2006.
[38] S. Lorenzi, A. Cammi, L. Luzzi, G. Rozza, POD-Galerkin method for finite volume approximation of Navier–Stokes and RANS equations, Comput. Methods Appl. Mech. Eng. 311 (2016) 151–179.
[39] X. Ma, G.E. Karniadakis, A low-dimensional model for simulating three-dimensional cylinder flow, J. Fluid Mech. 458 (2002) 181–190.
[40] R. MacCormack, Numerical Computation of Compressible Viscous Flow, Tech. rep., Lecture notes for AA214b and AA214c, Stanford University, 2007.
[41] M. Marion, R. Temam, Nonlinear Galerkin methods, SIAM J. Numer. Anal. 26 (1989) 1139–1157.
[42] B. Noack, P. Papas, P. Monkewitz, The need for a pressure-term representation in empirical Galerkin models of incompressible shear flows, J. Fluid Mech. 523 (2005) 339–365.
[43] M. Ohlberger, S. Rave, Nonlinear reduced basis approximation of parameterized evolution equations via the method of freezing, C. R. Math. 351 (2013) 901–906.
[44] C. Prud'homme, D. Rovas, K. Veroy, L. Machiels, Y. Maday, A. Patera, G. Turinici, Reliable real-time solution of parameterized partial differential equations: reduced-basis output bound methods, J. Fluids Eng. 124 (2002) 70–80.
[45] S.R. Reddy, B.A. Freno, P.G. Cizmas, S. Gokaltun, D. McDaniel, G.S. Dulikravich, Constrained reduced-order models based on proper orthogonal decomposition, Comput. Methods Appl. Mech. Eng. 321 (2017) 18–34.
[46] C. Rowley, T. Colonius, R. Murray, Model reduction for compressible flows using POD and Galerkin projection, Phys. D: Nonlinear Phenom. 189 (2004) 115–129.
[47] G. Rozza, D.B.P. Huynh, A.T. Patera, Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations, Arch. Comput. Methods Eng. 15 (2008) 229–275.
[48] D. Ryckelynck, A priori hyperreduction method: an adaptive approach, J. Comput. Phys. 202 (2005) 346–366.
[49] O. San, T. Iliescu, Proper orthogonal decomposition closure models for fluid flows: Burgers equation, arXiv preprint arXiv:1308.3276, 2013.

[50] J. Shen, Long time stability and convergence for fully discrete nonlinear Galerkin methods, Appl. Anal. 38 (1990) 201–229.

[51] S. Sirisup, G. Karniadakis, A spectral viscosity method for correcting the long-term behavior of pod models, J. Comput. Phys. 194 (2004) 92–116.

[52] L. Sirovich, Turbulence and the dynamics of coherent structures. III: dynamics and scaling, Q. Appl. Math. 45 (1987) 583–590.

[53] G. Stabile, S. Hijazi, A. Mola, S. Lorenzi, G. Rozza, Advances in reduced order modelling for CFD: vortex shedding around a circular cylinder using a POD-Galerkin method, arXiv preprint arXiv:1701.03424, 2017.

[54] G. Stabile, G. Rozza, Finite volume POD-Galerkin stabilised reduced order methods for the parametrised incompressible Navier–Stokes equations, arXiv preprint arXiv:1710.11580, 2017.

[55] T. Taddei, S. Perotto, A. Quarteroni, Reduced basis techniques for nonlinear conservation laws, ESAIM: Math. Model. Numer. Anal. 49 (2015) 787–814.

[56] S. Volkwein, S. Weiland, An algorithm for Galerkin projections in both time and spatial coordinates, in: Proc. 17th MTNS, 2006.

[57] Z. Wang, I. Akhtar, J. Borggaard, T. Iliescu, Proper orthogonal decomposition closure models for turbulent flows: a numerical comparison, Comput. Methods Appl. Mech. Eng. 237 (2012) 10–26.

[58] M.J. Zahr, K. Carlberg, D. Amsallem, C. Farhat, Comparison of Model Reduction Techniques on High-Fidelity Linear and Nonlinear Electrical, Mechanical, and Biological Systems, University of California, Berkeley, 2010.

[59] R. Zimmermann, A. Vendl, S. Görtz, Reduced-order modeling of steady flows subject to aerodynamic constraints, AIAA J. 52 (2014).