Discrete Energy Laws for the First-Order System Least-Squares Finite-Element Approach

J. H. Adler[1(✉)], I. Lashuk[1], S. P. MacLachlan[2], and L. T. Zikatanov[3]

[1] Department of Mathematics, Tufts University, Medford, MA 02155, USA
james.adler@tufts.edu, ilya.lashuk@gmail.com
[2] Department of Mathematics and Statistics, Memorial University of Newfoundland, St. John's, NL A1C 5S7, Canada
smaclachlan@mun.ca
[3] Department of Mathematics, Penn State, University Park, PA 16802, USA
ludmil@psu.edu

Abstract. This paper analyzes the discrete energy laws associated with first-order system least-squares (FOSLS) discretizations of time-dependent partial differential equations. Using the heat equation and the time-dependent Stokes' equation as examples, we discuss how accurately a FOSLS finite-element formulation adheres to the underlying energy law associated with the physical system. Using regularity arguments involving the initial condition of the system, we are able to give bounds on the convergence of the discrete energy law to its expected value (zero in the examples presented here). Numerical experiments are performed, showing that the discrete energy laws hold with order $\mathcal{O}\left(h^{2p}\right)$, where h is the mesh spacing and p is the order of the finite-element space. Thus, the energy law conformance is held with a higher order than the expected, $\mathcal{O}\left(h^{p}\right)$, convergence of the finite-element approximation. Finally, we introduce an abstract framework for analyzing the energy laws of general FOSLS discretizations.

1 Introduction

First-order system least squares (FOSLS) is a finite-element methodology that aims to reformulate a set of partial differential equations (PDEs) as a system of first-order equations [8,9]. The problem is posed as a minimization of a functional in which the first-order differential terms appear quadratically, so that the functional norm is equivalent to a norm meaningful for the given problem. In equations of elliptic type, this is usually a product H^1 norm. Some of the compelling features of the FOSLS methodology include: self-adjoint discrete equations stemming from the minimization principle; good operator conditioning stemming from the use of first-order formulations of the PDE; and finite-element and multigrid performance that is optimal and uniform in certain parameters (e.g., Reynolds number for the Navier-Stokes equations), stemming from uniform product-norm equivalence.

Successful FOSLS formulations have been developed for a variety of applications [5,7]. One example of a large-scale physical application is in magneto-hydrodynamics (MHD) [1–3]. These numerical methods have led to substantial improvements in MHD simulation technology; however, several important estimates remain to be analyzed to confirm their quantitive accuracy. One of these is the energy of the system. Using an energetic-variational approach [11–13,15], energy laws of the MHD system can be derived that show that the total energy should decay as a direct result of the dissipation in the system. Initial computations show that the FOSLS method indeed captures this energy law, but it remains to be shown why it should.

In this paper, we describe the *discrete* energy laws associated with FOSLS discretizations of time-dependent PDEs, such as the heat equation or Stokes' equation, and show quantitatively how they are related to the continuous physical law. While we only show results for these "simple" linear systems, the results appear generalizable to more complicated systems, such as MHD. Getting the correct energy law is not only important for numerical stability, but it is crucial for capturing the correct physics, especially if singularities or high contrasts in the solution are present.

The paper is outlined as follows. In Sect. 2, we discuss the energy laws of a given system and describe their discrete analogues. Section 3 analyzes the energy laws associated with the FOSLS discretizations of the heat equation, and the same is done for Stokes' equations in Sect. 4. For both examples, we present numerical simulations in Sect. 5. Finally, we discuss a generalization of the concepts presented here in Sect. 6, and some concluding remarks in Sect. 7.

2   Energy Laws

The energetic-variational approach (EVA) [11–13,15] of hydrodynamic systems in complex fluids is based on the second law of thermodynamics and relies on the fundamental principle that the change in the total energy of a system over time must equal the total dissipation of the system. This energy principle plays a crucial role in understanding the interactions and coupling between different scales or phases in a complex fluid. In general, any set of equations that describe the system can be derived from the underlying energy laws. The energetic variational principle is based on the energy dissipation law for the whole coupled system:

$$\frac{\partial E_{total}}{\partial t} = -\mathcal{D}, \tag{1}$$

where E_{total} is the total energy of the system, and $\mathcal{D}$ is the dissipation.

Simple fluids, where we assume no internal (or elastic) energies, can also be described in this setting and yield the following energy law:

$$\frac{\partial}{\partial t} \left(\frac{1}{2} \int_{\Omega} |\mathbf{u}|^2 \, d\mathbf{x} \right) = - \int_{\Omega} \nu |\nabla \mathbf{u}|^2 \, d\mathbf{x}, \tag{2}$$

where $\mathbf{u}$ represents the fluid velocity and ν is the fluid viscosity, accounting for the dissipation in the system. Applying the so-called least-action principle results in the integral equation,

$$\left\langle \frac{\partial \mathbf{u}}{\partial t} + \nabla p, \mathbf{y} \right\rangle = \langle \nabla \cdot \nu \nabla \mathbf{u}, \mathbf{y} \rangle, \forall \mathbf{y} \in \mathcal{V},$$

where we assume an incompressible fluid, $\nabla \cdot u = 0$, and an appropriate Hilbert space, $\mathcal{V}$. Here, we use $\langle \cdot, \cdot \rangle$ to denote the $L^2(\Omega)$ inner product. In strong form, we obtain the time-dependent Stokes' equations (assuming appropriate boundary conditions):

$$\frac{\partial \mathbf{u}}{\partial t} + \nabla p - \nabla \cdot \nu \nabla \mathbf{u} = 0, \tag{3}$$

$$\nabla \cdot \mathbf{u} = 0. \tag{4}$$

Note that the energy law can also be derived directly from the PDE itself. First, we consider the weak form of (3)–(4), multiplying (3) by $\mathbf{u}$ and (4) by p and integrate over Ω. After integration by parts we obtain the following relations:

$$0 = \left\langle \frac{\partial \mathbf{u}}{\partial t} + \nabla p - \nabla \cdot \nu \nabla \mathbf{u}, \mathbf{u} \right\rangle + \langle \nabla \cdot \mathbf{u}, p \rangle$$

$$= \left\langle \frac{\partial \mathbf{u}}{\partial t}, \mathbf{u} \right\rangle + \langle \nabla p, \mathbf{u} \rangle - \langle \nabla \cdot \nu \nabla \mathbf{u}, \mathbf{u} \rangle + \langle \nabla \cdot \mathbf{u}, p \rangle$$

$$= \frac{1}{2} \frac{\partial}{\partial t} \langle \mathbf{u}, \mathbf{u} \rangle + \langle \nabla p, \mathbf{u} \rangle + \langle \nu \nabla \mathbf{u}, \nabla \mathbf{u} \rangle - \langle \mathbf{u}, \nabla p \rangle.$$

Here, we have assumed that the boundary conditions are such that the boundary terms, resulting from the integration by parts, vanish. Hence, we have

$$\frac{1}{2} \frac{\partial}{\partial t} \langle \mathbf{u}, \mathbf{u} \rangle = - \langle \nu \nabla \mathbf{u}, \nabla \mathbf{u} \rangle.$$

This approach can also be applied to other PDEs, such as the heat equation, to show similar energy dissipation relations. Let ν be the thermal diffusivity of the body Ω, and u its temperature. Then the PDE describing the temperature distribution in Ω is as follows,

$$\frac{\partial u}{\partial t} - \nabla \cdot \nu \nabla u = 0, \quad \text{on } \Omega, \quad u = 0, \quad \text{on } \partial \Omega. \tag{5}$$

As before, we multiply (5) by u and integrate over Ω to obtain that

$$0 = \left\langle \frac{\partial u}{\partial t} - \nabla \cdot \nu \nabla u, u \right\rangle = \left\langle \frac{\partial u}{\partial t}, u \right\rangle - \langle \nabla \cdot \nu \nabla u, u \rangle$$

$$= \frac{1}{2} \frac{\partial}{\partial t} \langle u, u \rangle + \langle \nu \nabla u, \nabla u \rangle$$

Hence,

$$\frac{1}{2} \frac{\partial}{\partial t} \langle u, u \rangle = - \langle \nu \nabla u, \nabla u \rangle,$$

which is the scalar version of (2).

For the remainder of the paper, we analyze (2), specifically how closely the FOSLS method can approximate the energy law discretely. We will consider both the scalar (heat equation) and the vector version (Stokes' equation) in the numerical results, as the form of the energy law is identical. First, we discuss how moving to a finite-dimensional space affects the energy law.

3   Heat Equation

First, we consider the heat equation, assuming a constant diffusion coefficient $\nu = 1$ for simplicity, homogeneous Dirichlet boundary conditions, and a given initial condition:

$$\frac{\partial u(\boldsymbol{x}, t)}{\partial t} = \Delta u(\boldsymbol{x}, t) \quad \forall \boldsymbol{x} \in \Omega,\ \forall t > 0 \tag{6}$$

$$u(\boldsymbol{x}, t) = 0 \quad \forall \boldsymbol{x} \in \partial\Omega,\ \forall t \geq 0 \tag{7}$$

$$u(\boldsymbol{x}, 0) = u_0(\boldsymbol{x}) \quad \forall \boldsymbol{x} \in \bar{\Omega}. \tag{8}$$

To discretize the problem in time, we consider a symplectic, or energy-conserving, time-stepping scheme such as Crank-Nicolson. Given a time step size, τ, and time $t_n = \tau n$, we approximate $u_n = u(\mathbf{x}, t_n)$ with the following semi-discrete version of (6),

$$\frac{u_{n+1} - u_n}{\tau} = \frac{\Delta u_{n+1} + \Delta u_n}{2}$$

To simplify the calculations later, we introduce an intermediate approximation, $u_{n+\frac{1}{2}}$, and re-write the semi-discrete problem as

$$\frac{u_{n+\frac{1}{2}} - u_n}{\left(\frac{\tau}{2}\right)} = \Delta u_{n+\frac{1}{2}}$$

$$u_{n+\frac{1}{2}}(\boldsymbol{x}) = 0 \quad \forall \boldsymbol{x} \in \partial\Omega,\ n = 0, 1, 2, \ldots \tag{9}$$

$$u_{n+1} = 2u_{n+\frac{1}{2}} - u_n$$

Remark 1. To obtain the *semi-discrete* energy law for (9), we perform a similar procedure as done in Sect. 2, where we multiply the first equation in (9) by $u_{n+\frac{1}{2}}$ and integrate over the domain. After some simple calculations, we obtain the corresponding energy law, using L^2-norm notation:

$$\frac{||u_{n+1}||^2 - ||u_n||^2}{2\tau} = -||\nabla u_{n+\frac{1}{2}}||^2 \tag{10}$$

To use the FOSLS method, we now put the operator into a first-order system. Since we have reduced the problem to a reaction-diffusion type problem, we introduce a new vector $\boldsymbol{V} = \nabla u$, and use the H^1-elliptic equivalent system [8,9]:

$$L_\tau \begin{pmatrix} u_{n+\frac{1}{2}} \\ \boldsymbol{V}_{n+\frac{1}{2}} \end{pmatrix} = \begin{pmatrix} -\nabla \cdot \boldsymbol{V}_{n+\frac{1}{2}} + \frac{2}{\tau} u_{n+\frac{1}{2}} \\ \boldsymbol{V}_{n+\frac{1}{2}} - \nabla u_{n+\frac{1}{2}} \\ \nabla \times \boldsymbol{V}_{n+\frac{1}{2}} \end{pmatrix} = \begin{pmatrix} \frac{2}{\tau} u_n \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}. \tag{11}$$

Note that Dirichlet boundary conditions on the continuous solution, u, gives rise to tangential boundary conditions on $\mathbf{V}$, $\mathbf{V} \times \mathbf{n} = \mathbf{0}$, where $\mathbf{n}$ is the normal vector to the boundary.

Next, we consider a finite-dimensional subspace of a product H^1 space, $\mathcal{V}^h$, and perform the FOSLS minimization of (11) over $\mathcal{V}^h$:

$$\left(u_{n+\frac{1}{2}}^h, \mathbf{V}_{n+\frac{1}{2}}^h \right) = \operatorname*{arg\,min}_{(u,\mathbf{V}) \in \mathcal{V}^h} \left\| L_\tau \begin{pmatrix} u \\ \mathbf{V} \end{pmatrix} - \begin{pmatrix} \frac{2}{\tau} u_n^h \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix} \right\|,$$

$$u_{n+1}^h = 2 u_{n+\frac{1}{2}}^h - u_n^h.$$

For each n, the above minimization results in the following weak set of equations:

$$\left\langle L_\tau \begin{pmatrix} u_{n+\frac{1}{2}}^h \\ \mathbf{V}_{n+\frac{1}{2}}^h \end{pmatrix} - \begin{pmatrix} \frac{2}{\tau} u_n^h \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, L_\tau \boldsymbol{\phi}^h \right\rangle = 0 \quad \forall \boldsymbol{\phi}^h \in \mathcal{V}^h, \tag{12}$$

where the inner products and norms are all in L^2 (scalar or vector, depending on context), unless otherwise noted.

Note, that with the introduction of $\mathbf{V}$, the discrete form of the FOSLS energy law can now be written,

$$\frac{||u_{n+1}^h||^2 - ||u_n^h||^2}{2\tau} - ||\mathbf{V}_{n+\frac{1}{2}}^h||^2 \to 0, \quad \text{as} \quad h \to 0. \tag{13}$$

The goal of the remainder of this Section is to show how well this energy law is satisfied. To do so, we make use of the following assumption.

Assumption 1. *Assume that the initial condition is smooth enough and the projection onto the finite-element space has the following property,*

$$\left\| u_0 - u_0^h \right\|_{H^1} \leq C h^p \left\| u_0 \right\|_{H^{p+1}},$$

where p is the order of the finite-element space being considered.

Then, using standard regularity estimates we obtain the following Lemma.

Lemma 1. *Let $\{u_i\}_{i=0,1,\dots}$ be a sequence of semi-discrete solutions to (9). Then, for any successive time steps, there exists a constant $C > 0$, such that*

$$\left\| u_{n+1} \right\|_{H^p} \leq C \left\| u_n \right\|_{H^p}$$

A consequence of this regularity estimate is a bound on the error in the approximation.

Lemma 2. *Let $f \in H^p \cap H_0^1$ and let the pair $\left(u^h, \mathbf{V}^h \right) \in \mathcal{V}^h$ solve*

$$\begin{pmatrix} u^h \\ \mathbf{V}^h \end{pmatrix} = \operatorname*{arg\,min}_{(u,\mathbf{V}) \in \mathcal{V}^h} \left\| L_\tau \begin{pmatrix} u \\ \mathbf{V} \end{pmatrix} - \begin{pmatrix} \frac{2}{\tau} f \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix} \right\|^2.$$

Let $\hat{u}$ be the exact solution of the corresponding PDE, i.e.,

$$-\Delta\hat{u} + \frac{2}{\tau}\hat{u} = \frac{2}{\tau}f \quad in \; \partial\Omega,$$

$$\hat{u} = 0 \quad on \; \partial\Omega.$$

Then,

$$\left\|u^h - \hat{u}\right\|_{H^1} \leq \frac{C(\tau)h^p}{\tau}\left\|f\right\|_{H^{p-1}},$$

where the constant $C(\tau)$ may also depend on τ.

Proof. For a fixed τ, the PDE is a reaction-diffusion equation. Therefore, standard results from the FOSLS discretization of reaction-diffusion can be used [8,9]. Note that for a standard FOSLS approach, $C(\tau) = \mathcal{O}\left(\frac{1}{\tau^2}\right)$, but a rescaling of the equations may ameliorate this "worst-case scenario."

Next, we make the following observation, which follows from the well-posedness of the FOSLS formulation [8,9].

Lemma 3. *Let $(u_1, \boldsymbol{V}_1) \in \mathcal{V}^h$ and $(u_2, \boldsymbol{V}_2) \in \mathcal{V}^h$ be two solutions to the following FOSLS weak forms with different right-hand sides,*

$$\left\langle L_\tau\begin{pmatrix}u_1\\\boldsymbol{V}_1\end{pmatrix} - F_1, L_\tau\boldsymbol{\phi}^h\right\rangle = 0, \quad \left\langle L_\tau\begin{pmatrix}u_2\\\boldsymbol{V}_2\end{pmatrix} - F_2, L_\tau\boldsymbol{\phi}^h\right\rangle = 0 \quad \forall\boldsymbol{\phi}^h \in \mathcal{V}.$$

Then,

$$\left\|u_1 - u_2\right\|_{H^1} + \left\|\boldsymbol{V}_1 - \boldsymbol{V}_2\right\|_{H^1} \leq C(\tau)\left\|F_1 - F_2\right\|.$$

This, then, yields the following result.

Lemma 4. *Given the solution to the semi-discrete Eq. (9), and the fully discrete solution, we can bound the error in the L^2 norm:*

$$\left\|u^h_{n+\frac{1}{2}} - u_{n+\frac{1}{2}}\right\| \leq \frac{C_1(\tau)}{\tau}h^p\left\|u_n\right\|_{H^p} + C_2(\tau)\left\|u^h_n - u_n\right\|. \tag{14}$$

Proof. Let $\tilde{u}^h_{n+\frac{1}{2}}$ be the scalar part of the FOSLS solution $\left(\tilde{u}^h_{n+\frac{1}{2}}, \tilde{\boldsymbol{V}}^h_{n+\frac{1}{2}}\right)$ of

$$-\Delta u + \frac{2}{\tau}u = \frac{2}{\tau}u_n, \text{ in } \Omega, \quad u = 0 \text{ on } \partial\Omega, \tag{15}$$

where the exact semi-discrete solution u_n, at the previous time step, is used in the right-hand side. By the triangle inequality,

$$\left\|u^h_{n+\frac{1}{2}} - u_{n+\frac{1}{2}}\right\| \leq \left\|u^h_{n+\frac{1}{2}} - \tilde{u}^h_{n+\frac{1}{2}}\right\| + \left\|\tilde{u}^h_{n+\frac{1}{2}} - u_{n+\frac{1}{2}}\right\|. \tag{16}$$

By Lemma 3, we have

$$\left\|u^h_{n+\frac{1}{2}} - \tilde{u}^h_{n+\frac{1}{2}}\right\| \leq \left\|u^h_{n+\frac{1}{2}} - \tilde{u}^h_{n+\frac{1}{2}}\right\|_{H^1} + \left\|\boldsymbol{V}^h_{n+\frac{1}{2}} - \tilde{\boldsymbol{V}}^h_{n+\frac{1}{2}}\right\|_{H^1}$$

$$\leq C_2(\tau)\left\|u^h_n - u_n\right\|. \tag{17}$$

The functions $\tilde{u}^h_{n+\frac{1}{2}}$ and $u_{n+\frac{1}{2}}$ are, respectively, FOSLS and exact solutions of the same boundary value problem (15). Hence, from Lemma 2, we have

$$\left\|\tilde{u}^h_{n+\frac{1}{2}} - u_{n+\frac{1}{2}}\right\| \leq \frac{C(\tau)}{\tau} h^p \|u_n\|_{H^{p-1}} \leq \frac{C(\tau)}{\tau} h^p \|u_n\|_{H^p} \tag{18}$$

Combining (16), (17) and (18), we obtain (14).

Finally, we have the following result on the approximation of the exact energy law (13).

Theorem 1. *Let* $\begin{pmatrix} u^h_n \\ \mathbf{V}^h_n \end{pmatrix}$ *be the solution to the FOSLS system,* (12), *at time step* n *(with* $u^h_{n+\frac{1}{2}}$ *and* $V^h_{n+\frac{1}{2}}$ *defined as before). There exists* $C(\tau) > 0$ *such that*

$$\left| \frac{\|u^h_{n+1}\|^2 - \|u^h_n\|^2}{2\tau} + \|\mathbf{V}^h_{n+\frac{1}{2}}\|^2 \right| \leq C(\tau) \frac{2}{\tau} \|u^h_n - u_n\| \min_{\phi^h \in \mathcal{V}^h} \left\| \begin{pmatrix} u^h_{n+\frac{1}{2}} \\ \mathbf{V}^h_{n+\frac{1}{2}} \\ \mathbf{0} \end{pmatrix} - L_\tau \phi^h \right\|.$$

Proof. To simplify the notation, define the energy law we wish to bound as

$$E^h_n := \frac{\|u^h_{n+1}\|^2 - \|u^h_n\|^2}{2\tau} + \|\mathbf{V}^h_{n+\frac{1}{2}}\|^2.$$

Note that

$$\frac{1}{2} \frac{\|u^h_{n+1}\|^2 - \|u^h_n\|^2}{\tau} = \left\langle \frac{u^h_{n+1} - u^h_n}{\tau}, u^h_{n+\frac{1}{2}} \right\rangle = \left\langle \frac{u^h_{n+\frac{1}{2}} - u^h_n}{\frac{\tau}{2}}, u^h_{n+\frac{1}{2}} \right\rangle,$$

and

$$\left\|\mathbf{V}^h_{n+\frac{1}{2}}\right\|^2 = \left\langle -\nabla \cdot \mathbf{V}^h_{n+\frac{1}{2}}, u^h_{n+\frac{1}{2}} \right\rangle + \left\langle \mathbf{V}^h_{n+\frac{1}{2}} - \nabla u^h_{n+\frac{1}{2}}, \mathbf{V}^h_{n+\frac{1}{2}} \right\rangle,$$

where the latter equation is obtained by integration by parts, continuity of the spaces, and appropriate boundary conditions. Thus,

$$E^h_n = \left\langle -\nabla \cdot \mathbf{V}^h_{n+\frac{1}{2}} + \frac{u^h_{n+\frac{1}{2}} - u^h_n}{\frac{\tau}{2}}, u^h_{n+\frac{1}{2}} \right\rangle + \left\langle \mathbf{V}^h_{n+\frac{1}{2}} - \nabla u^h_{n+\frac{1}{2}}, \mathbf{V}^h_{n+\frac{1}{2}} \right\rangle$$

$$= \left\langle L_\tau \begin{pmatrix} u^h_{n+\frac{1}{2}} \\ \mathbf{V}^h_{n+\frac{1}{2}} \end{pmatrix} - \begin{pmatrix} \frac{2}{\tau} u^h_n \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} u^h_{n+\frac{1}{2}} \\ \mathbf{V}^h_{n+\frac{1}{2}} \\ \mathbf{0} \end{pmatrix} \right\rangle.$$

Using (12), for any $\phi^h \in \mathcal{V}^h$,

$$E^h_n = \left\langle L_\tau \begin{pmatrix} u^h_{n+\frac{1}{2}} \\ \mathbf{V}^h_{n+\frac{1}{2}} \end{pmatrix} - \begin{pmatrix} \frac{2}{\tau} u^h_n \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} u^h_{n+\frac{1}{2}} \\ \mathbf{V}^h_{n+\frac{1}{2}} \\ \mathbf{0} \end{pmatrix} - L_\tau \phi^h \right\rangle.$$

Next, consider adding and subtracting the solutions to the semi-discrete, (11), and fully discrete, (12), FOSLS system from the previous time step,

$$
E_n^h = \left\langle L_\tau \begin{pmatrix} u_{n+\frac{1}{2}}^h \\ \boldsymbol{V}_{n+\frac{1}{2}}^h \end{pmatrix} - \begin{pmatrix} \frac{2}{\tau} u_n^h \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix} + \frac{2}{\tau} \begin{pmatrix} u_n^h - u_n \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} u_{n+\frac{1}{2}}^h \\ \boldsymbol{V}_{n+\frac{1}{2}}^h \\ \boldsymbol{0} \end{pmatrix} - L_\tau \phi^h \right\rangle
$$

$$
- \frac{2}{\tau} \left\langle \begin{pmatrix} u_n^h - u_n \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} u_{n+\frac{1}{2}}^h \\ \boldsymbol{V}_{n+\frac{1}{2}}^h \\ \boldsymbol{0} \end{pmatrix} - L_\tau \phi^h \right\rangle
$$

$$
= \left\langle L_\tau \begin{pmatrix} u_{n+\frac{1}{2}}^h \\ \boldsymbol{V}_{n+\frac{1}{2}}^h \end{pmatrix} - \frac{2}{\tau} \begin{pmatrix} u_n \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} u_{n+\frac{1}{2}}^h \\ \boldsymbol{V}_{n+\frac{1}{2}}^h \\ \boldsymbol{0} \end{pmatrix} - L_\tau \phi^h \right\rangle
$$

$$
- \frac{2}{\tau} \left\langle \begin{pmatrix} u_n^h - u_n \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} u_{n+\frac{1}{2}}^h \\ \boldsymbol{V}_{n+\frac{1}{2}}^h \\ \boldsymbol{0} \end{pmatrix} - L_\tau \phi^h \right\rangle
$$

$$
\leq \left\| L_\tau \begin{pmatrix} u_{n+\frac{1}{2}}^h \\ \boldsymbol{V}_{n+\frac{1}{2}}^h \end{pmatrix} - \frac{2}{\tau} \begin{pmatrix} u_n \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix} \right\| M_n^h + \frac{2}{\tau} M_n^h \left\| u_n^h - u_n \right\|,
$$

where we have defined $M_n^h := \min\limits_{\phi^h \in \mathcal{V}^h} \left\| \begin{pmatrix} u_{n+\frac{1}{2}}^h \\ \boldsymbol{V}_{n+\frac{1}{2}}^h \\ \boldsymbol{0} \end{pmatrix} - L_\tau \phi^h \right\|$. Then, adding and

subtracting $L_\tau \begin{pmatrix} u_{n+\frac{1}{2}} \\ \boldsymbol{V}_{n+\frac{1}{2}} \end{pmatrix}$ yields

$$
E_n^h \leq \left\| L_\tau \begin{pmatrix} u_{n+\frac{1}{2}}^h - u_{n+\frac{1}{2}} \\ \boldsymbol{V}_{n+\frac{1}{2}}^h - \boldsymbol{V}_{n+\frac{1}{2}} \end{pmatrix} + L_\tau \begin{pmatrix} u_{n+\frac{1}{2}} \\ \boldsymbol{V}_{n+\frac{1}{2}} \end{pmatrix} - \frac{2}{\tau} \begin{pmatrix} u_n \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix} \right\| M_n^h + \frac{2}{\tau} M_n^h \left\| u_n^h - u_n \right\|.
$$

Using the continuity of L_τ, followed by Lemma 3, gives

$$
|E_n^h| \leq C(\tau) M_n^h \left\| \begin{pmatrix} u_{n+\frac{1}{2}}^h - u_{n+\frac{1}{2}} \\ \boldsymbol{V}_{n+\frac{1}{2}}^h - \boldsymbol{V}_{n+\frac{1}{2}} \end{pmatrix} \right\|_{H^1} + \frac{2}{\tau} M_n^h \left\| u_n^h - u_n \right\|
$$

$$
\leq C(\tau) \frac{2}{\tau} M_n^h \left\| u_n^h - u_n \right\| + \frac{2}{\tau} M_n^h \left\| u_n^h - u_n \right\|.
$$

Combining the two terms completes the proof.

To provide a better bound for the FOSLS energy law (13), we introduce a measure for the truncation error defined as

$$
\delta_n = \max_{v \in H^{p+1}(\Omega)} \frac{1}{\|v\|_{H^{p+1}}} \min_{\phi^h \in \mathcal{V}^h} \left\| \begin{pmatrix} u_{n+\frac{1}{2}}^h(v) \\ \boldsymbol{V}_{n+\frac{1}{2}}^h(v) \\ \boldsymbol{0} \end{pmatrix} - \mathcal{L}\phi^h \right\|, \tag{19}
$$

where $u_{n+\frac{1}{2}}^h(v)$ and $\mathbf{V}_{n+\frac{1}{2}}^h(v)$ are the corresponding solutions to the fully discrete problem with $u_0 = v$ as the initial condition.

Corollary 1. *Using the same assumptions as Theorem 1 and Assumption 1,*

$$\left| \frac{\|u_{n+1}^h\|^2 - \|u_n^h\|^2}{2\tau} + \|\mathbf{V}_{n+\frac{1}{2}}^h\|^2 \right| \leq \frac{C(\tau)\delta}{\tau} h^p \|u_0\|_{H^{p+1}}^2, \quad \delta = \max_n \delta_n. \quad (20)$$

Proof. Using the definitions of $u_{n+\frac{1}{2}}^h$ and $u_{n+\frac{1}{2}}$, the triangle inequality, and Lemma 4,

$$\|u_{n+1}^h - u_{n+1}\| \leq 2\|u_{n+\frac{1}{2}}^h - u_{n+\frac{1}{2}}\| + \|u_n^h - u_n\|$$
$$\leq \frac{2}{\tau} C_1(\tau) h^p \|u_n\|_{H^p} + (C_2(\tau) + 1) \|u_n^h - u_n\|.$$

An induction argument then gives

$$\|u_n^h - u_n\| \leq \frac{2}{\tau} C_1(\tau) h^p \sum_{j=1}^n (C_2(\tau) + 1)^{j-1} \|u_{n-j}\|_{H^p} + (C_2(\tau) + 1)^n \|u_0^h - u_0\|.$$

With Assumption 1,

$$\|u_n^h - u_n\| \leq \frac{2}{\tau} C_1(\tau) h^p \sum_{j=1}^n (C_2(\tau) + 1)^{j-1} \|u_{n-j}\|_{H^p} + (C_2(\tau) + 1)^n \|u_0\|_{H^{p+1}}.$$

Using some regularity arguments for each u_i, we get,

$$\|u_n^h - u_n\| \leq C(n) h^p \|u_0\|_{H^{p+1}}.$$

Then, with the definition of δ and the result from Theorem 1, the proof is complete.

We note that the bound in Corollary 1 is a rather pessimistic one. At a fixed time, t, we expect the quality of both the fully discrete and semi-discrete approximations to the true solution to improve as $\tau \to 0$ and more time-steps are used to reach time t; thus, $\|u_n^h - u_n\|$ should decrease as $\tau \to 0$ for $n = t/\tau$. Furthermore, for the unforced heat equation, we expect both u_n^h and u_n to decrease in magnitude with n, but this is not accounted for in the bound in Corollary 1. The bound above worsens with smaller τ and bigger n, showing the limitations of bounding $\|u_n^h - u_n\|$ by terms depending only on u_0 and the finite-element space.

Remark 2. As shown in the numerical experiments, Sect. 5, the constant δ defined in (19) is of order h^p for a smooth solution. This indicates that the energy law (13) holds with order $\mathcal{O}\left(h^{2p}\right)$. While the theoretical justification of such statement may be plausible, it is nontrivial as the discrete quantities involved in the definition of δ do not possess enough regularity (they are just finite-element functions, only in H^1).

4 Stokes' Equations

Next, we return to the time-dependent Stokes' equations, (3)–(4). For simplicity, we again assume $\nu = 1$, and rewrite the equations using Dirichlet boundary conditions for the normal components of the velocity field, and zero-mean average for the pressure field,

$$\frac{\partial \boldsymbol{u}(\boldsymbol{x},t)}{\partial t} - \Delta \boldsymbol{u}(\boldsymbol{x},t) + \nabla p(\boldsymbol{x},t) = \boldsymbol{0} \quad \forall \boldsymbol{x} \in \Omega, \ \forall t > 0 \tag{21}$$

$$\nabla \cdot \boldsymbol{u}(\boldsymbol{x},t) = 0 \quad \forall \boldsymbol{x} \in \Omega, \ \forall t > 0 \tag{22}$$

$$\mathbf{n} \cdot \boldsymbol{u}(\boldsymbol{x},t) = 0 \quad \forall \boldsymbol{x} \in \partial\Omega, \ \forall t \geq 0 \tag{23}$$

$$\boldsymbol{u}(\boldsymbol{x},0) = \mathbf{g}(\boldsymbol{x}) \quad \forall \boldsymbol{x} \in \bar{\Omega}, \tag{24}$$

$$\int_\Omega p(\boldsymbol{x},t)\,\mathrm{d}V = 0 \quad \forall t \geq 0. \tag{25}$$

Using a similar semi-discretization in time with Crank-Nicolson that was done in (9) yields

$$\frac{\boldsymbol{u}_{n+\frac{1}{2}} - \boldsymbol{u}_n}{\left(\frac{\tau}{2}\right)} - \Delta \boldsymbol{u}_{n+\frac{1}{2}} + \nabla p_{n+\frac{1}{2}} = \boldsymbol{0},$$

$$\nabla \cdot \boldsymbol{u}_{n+\frac{1}{2}} = 0,$$

$$\mathbf{n} \cdot \boldsymbol{u}_{n+\frac{1}{2}}(\boldsymbol{x}) = 0 \quad \forall \boldsymbol{x} \in \partial\Omega, \tag{26}$$

$$\int_\Omega p_{n+\frac{1}{2}}\,\mathrm{d}V = 0 \quad \forall n \geq 0,$$

$$\boldsymbol{u}_{n+1} = 2\boldsymbol{u}_{n+\frac{1}{2}} - \boldsymbol{u}_n,$$

$$p_{n+\frac{1}{2}} = 2p_{n+\frac{1}{2}} - p_n.$$

To use the FOSLS method, we put the operator into a first-order system in a similar fashion to the heat equation. Least-squares formulations are well-studied for Stokes' system and we consider a simple, velocity-gradient-pressure formulation, where a new gradient tensor, $\mathbf{V} = \nabla \boldsymbol{u}$, is used to obtain an H^1-elliptic equivalent system [4,6,14]:

$$L_\tau \begin{pmatrix} \boldsymbol{u}_{n+\frac{1}{2}} \\ \mathbf{V}_{n+\frac{1}{2}} \\ p_{n+\frac{1}{2}} \end{pmatrix} = \begin{pmatrix} -\nabla \cdot \mathbf{V}_{n+\frac{1}{2}} + \nabla p_{n+\frac{1}{2}} + \frac{2}{\tau}\boldsymbol{u}_{n+\frac{1}{2}} \\ \nabla \cdot \boldsymbol{u}_{n+\frac{1}{2}} \\ \mathbf{V}_{n+\frac{1}{2}} - \nabla \boldsymbol{u}_{n+\frac{1}{2}} \\ \nabla \times \mathbf{V}_{n+\frac{1}{2}} \\ \nabla \mathrm{tr} \mathbf{V}_{n+\frac{1}{2}} \end{pmatrix} = \begin{pmatrix} \frac{2}{\tau}\boldsymbol{u}_n \\ 0 \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}. \tag{27}$$

Appropriate boundary conditions on the continuous solution, such as $\mathbf{n} \cdot \boldsymbol{u} = 0$, gives rise to tangential boundary conditions on $\mathbf{V}$, $\mathbf{V} \times \mathbf{n} = \boldsymbol{0}$, where $\mathbf{n}$ is the normal vector to the boundary. Ultimately, the corresponding semi-discrete energy law is

$$\frac{||\boldsymbol{u}_{n+1}||^2 - ||\boldsymbol{u}_n||^2}{2\tau} = -||\mathbf{V}_{n+\frac{1}{2}}||^2. \tag{28}$$

Finally, we minimize the residual of (27) over a finite-dimensional subspace of the product H^1 Sobolev space in the L^2 norm obtaining the weak equations,

$$\left\langle L_\tau \begin{pmatrix} \boldsymbol{u}^h_{n+\frac{1}{2}} \\ \boldsymbol{V}^h_{n+\frac{1}{2}} \\ p^h_{n+\frac{1}{2}} \end{pmatrix} - \begin{pmatrix} \frac{2}{\tau}\boldsymbol{u}^h_n \\ 0 \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, L_\tau \boldsymbol{\phi}^h \right\rangle = 0 \quad \forall \boldsymbol{\phi}^h \in \mathcal{V}^h. \tag{29}$$

Note that the weak system is similar to (12) and the energy law is identical to (13) in vector form. Thus, all the above theory still holds subject to enough regularity of the solution to the time-dependent Stokes' equations [21,22] and a suitable generalization of the definition of δ.

5 Numerical Experiments

For the numerical results presented here, we use a C++ implementation of the FOSLS algorithm, using the modular finite-element library MFEM [20] for managing the discretization, mesh, and timestepping. The linear systems are solved by direct method using the UMFPACK package [10].

5.1 Heat Equation

First, we consider the heat Eq. (6), and its discrete FOSLS formulation, (12), on a triangulation of $\Omega = (0,1) \times (0,1)$. The data is chosen so that the true solution is $u(x,y,t) = \sin(\pi x)\sin(\pi y)e^{-2\pi^2 t}$. Note that this solution satisfies the boundary conditions and other assumptions discussed above.
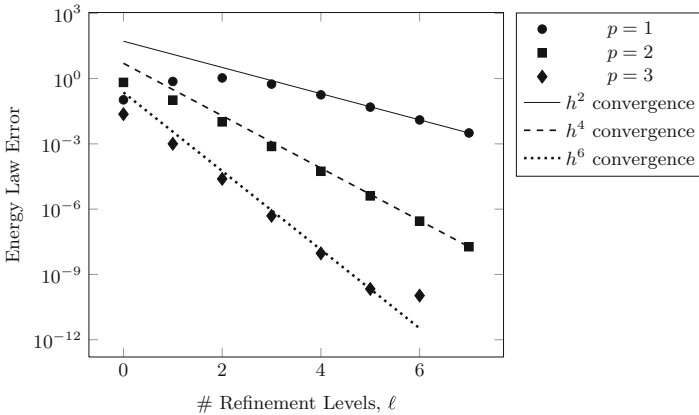

Fig. 1. Energy law error, (13), vs. number of mesh refinements, ℓ ($h = \frac{1}{2^\ell}$), for the FOSLS discretization of the heat equation, (12), using various orders of the finite-element space ($p = 1$ - linear; $p = 2$ - quadratic; and $p = 3$ - cubic). One time step is performed with $\tau = 0.005$.

Figure 1 displays the convergence of the energy law to zero as the mesh is refined for a fixed time step. The convergence is $\mathcal{O}\left(h^{2p}\right)$, where p is the order of the finite-element space being considered, confirming Theorem 1. It also suggests that the constant δ is $\mathcal{O}\left(h^{p}\right)$, as is remarked above.
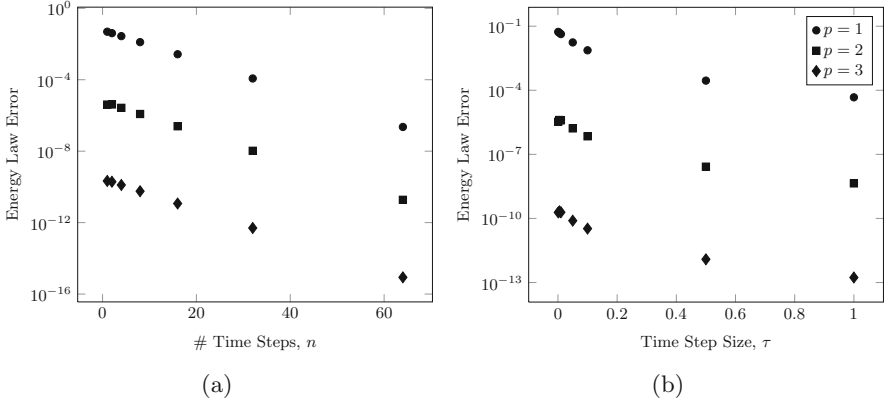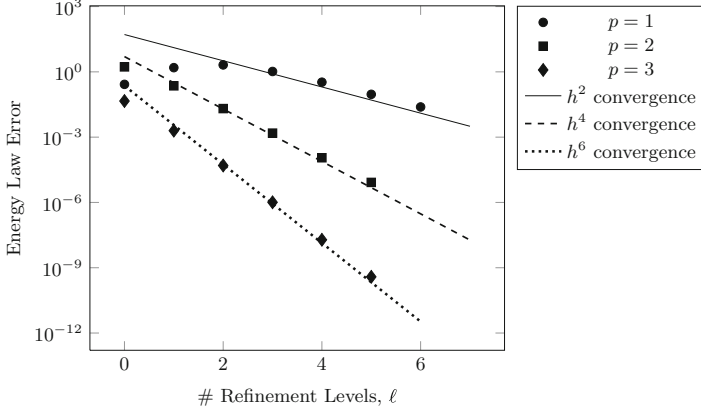

Fig. 2. Energy law error, (13), vs. (a) number of time steps, n (with fixed $\tau = 0.005$), and (b) time step size, τ, for the FOSLS discretization of the heat equation, (12), using various orders of the finite-element space ($p = 1$ - linear; $p = 2$ - quadratic; and $p = 3$ - cubic). Mesh spacing is $h = \frac{1}{32}$.

Figure 2 indicates how the timestepping affects the convergence of the energy law. As discussed above, taking more time steps decreases the error in the energy law, showing that we can improve the results on the bound, $\|u_n - u_n^h\|$. On the other hand, if only one time step is taken, the convergence slightly worsens for small τ, which is consistent with the constants found in Theorem 1 and Corollary 1.

5.2 Stokes' Equations

Next, we consider Stokes' Equations, (21), and the FOSLS discretization described above, (29). The same domain, $\Omega = (0, 1) \times (0, 1)$, is used, and we assume data that yields the exact solution,

$$\mathbf{u}(\mathbf{x}, t) = \begin{pmatrix} \sin(\pi x)\cos(\pi y) \\ -\cos(\pi x)\sin(\pi y) \end{pmatrix} e^{-2\pi^2 t},$$

$$p(\mathbf{x}, t) = 0.$$

This produces a C^∞ solution that satisfies the appropriate boundary conditions and regularity arguments needed for the bounds on the energy law described above.

Similarly to the heat equation, Fig. 3 compares the convergence of the energy law to zero as the mesh is refined for a fixed time step. Again, we see that the convergence is $\mathcal{O}\left(h^{2p}\right)$, where p is the order of the finite-element space being considered, confirming that Theorem 1 can also be applied to the time-dependent Stokes' equations. Thus, the FOSLS discretization can adhere to the energy law for fluid-type systems, and has the potential for capturing the relevant physics of other complex fluids.

Fig. 3. Energy law error, (28), vs. number of mesh refinements, ℓ ($h = \frac{1}{2^\ell}$), for the FOSLS discretization of the Stokes' equation, (27), using various orders of the finite-element space ($p = 1$ - linear; $p = 2$ - quadratic; and $p = 3$ - cubic). One time step is performed with $\tau = 0.005$.
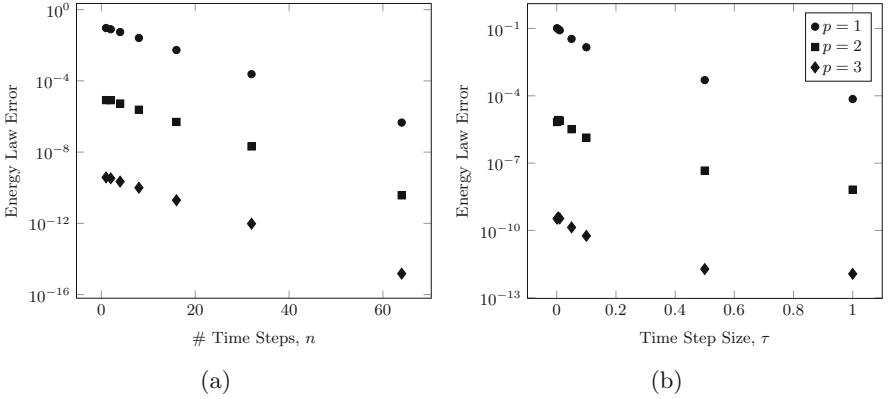

$$(a) \qquad\qquad\qquad (b)$$

Fig. 4. Energy law error, (28), vs. (a) number of time steps, n (with fixed $\tau = 0.005$), and (b) time step size, τ, for the FOSLS discretization of the Stokes' equation, (27), using various orders of the finite-element space ($p = 1$ - linear; $p = 2$ - quadratic; and $p = 3$ - cubic). Mesh spacing is $h = \frac{1}{32}$.

Figure 4 again confirms how we expect the timestepping to affect the convergence of the energy law. Taking more time steps decreases the error in the energy law, while the convergence slightly worsens for small τ. These results also highlight the similarities between the energy laws of the heat equation and the time-dependent Stokes equations. Since both have underlying energy laws that are similar, the FOSLS discretization is capable of capturing both with high accuracy.

6   Discussion: General Discrete Energy Laws

The above results show that FOSLS discretizations of two specific PDEs yield higher-order approximation of their underlying energy laws. In this section, we give a more general result, which suggests ideas for extending this theory for other discrete energy laws using FOSLS discretizations.

6.1   FOSLS Discrete Energy Laws

As encountered earlier, an energy law is an integral relation of the form:

$$\langle \mathcal{L}u, u \rangle = 0, \quad \text{for} \quad u(x,0) = u_0(x), \tag{30}$$

where $\mathcal{L} : \widetilde{\mathcal{V}} \to \widetilde{\mathcal{V}}$ is a linear operator (that involves boundary conditions), $\widetilde{\mathcal{V}}$ is a function space, and $u \in \widetilde{\mathcal{V}}$ is the solution to

$$\mathcal{L}u = 0, \quad u(x,0) = u_0, \quad \text{for example:} \quad \mathcal{L} = \partial_t - \Delta. \tag{31}$$

To match the time-dependent problems considered in earlier sections, $\widetilde{\mathcal{V}}$ corresponds to a computational domain that involves both space and time, or as is often dubbed, a "space-time" domain: $\widetilde{\Omega} = \Omega \times [0,T]$. Further, we define a finite-dimensional space, $\widetilde{V}_h$ on $\widetilde{\Omega}$ corresponding to a triangulation of this space-time domain, as well as a "stationary" finite-dimensional space, V_h, for $t = 0$. Regarding such space-time discrete spaces and the related constructions, we refer the reader to the classical works by Johnson et al. [16,17], to [19] for space-time least squares formulations, and to [18] for space-time iso-geometric analysis and a comprehensive literature review.

To present the FOSLS discretization in an abstract setting, we define an extension of u_0 to the whole of $\widetilde{\Omega}$. Without loss of generality, we assume that the initial condition is a piecewise polynomial and, more precisely, $u_0 \in V_h$. Hence, we define the extension $w_h \in \widetilde{V}_h$ of u_0 so that $w_h(x,0) = u_0(x)$. This gives a non-homogenous problem with zero initial guess, which is equivalent to (31). Its weak form is: Find $u \in \widetilde{\mathcal{V}}$ such that for all $v \in \widetilde{\mathcal{V}}_o$ there holds

$$u = \varphi + w_h, \quad \text{where} \quad \langle \mathcal{L}\varphi, v \rangle = - \langle \mathcal{L}w_h, v \rangle, \tag{32}$$

Here, the space, $\widetilde{\mathcal{V}}_o$, is the subspace of $\widetilde{\mathcal{V}}$ of functions with vanishing trace at $t = 0$ (zero initial condition). In a typical FOSLS setting, for the heat equation, u is a vector-valued function and the extension w_h needs to be modified accordingly. We then have the following space-time FOSLS discrete problem: Find $u_h \in \widetilde{\mathcal{V}}_h$ such that for all $v_h \in \widetilde{\mathcal{V}}_{h,o}$ there holds

$$u_h = w_h + \varphi_h, \quad \text{where,} \quad \langle \mathcal{L}\varphi_h, \mathcal{L}v_h \rangle = -\langle \mathcal{L}w_h, \mathcal{L}v_h \rangle. \tag{33}$$

Restricting $\widetilde{\mathcal{V}}$ to a finite-element space-time space, $\widetilde{\mathcal{V}}_h \subset \widetilde{\mathcal{V}}$, results in a restriction of $\mathcal{L}$ on $\widetilde{\mathcal{V}}_h$, which is often called the "discrete operator".

In the following, we keep $\langle \mathcal{L}u, u \rangle$ in all estimates allowing for a non-homogenous right-hand side in (30). We now estimate the error in the energy law, namely the difference $\langle \mathcal{L}u, u \rangle - \langle \mathcal{L}u_h, u_h \rangle$.

Theorem 2. *If $u_h \in \widetilde{\mathcal{V}}_h$ is the FOSLS solution of* (33). *Then, the following estimate holds:*

$$|\langle \mathcal{L}u, u \rangle - \langle \mathcal{L}u_h, u_h \rangle| \le Ch^p \|u\|_{H^{p+1}}. \tag{34}$$

Proof. For the left side of (34) we have

$$\langle \mathcal{L}u, u \rangle - \langle \mathcal{L}u_h, u_h \rangle = \langle \mathcal{L}u, u \rangle - \langle \mathcal{L}u, u_h \rangle + \langle \mathcal{L}(u - u_h), u_h \rangle$$
$$= \langle \mathcal{L}u, u - u_h \rangle + \langle \mathcal{L}(u - u_h), u_h \rangle.$$

Using the continuity of $\mathcal{L}$ and the standard error estimates for the FOSLS discretization,

$$|\langle \mathcal{L}u, u \rangle - \langle \mathcal{L}u_h, u_h \rangle| \le |\langle \mathcal{L}u, u - u_h \rangle| + |\langle \mathcal{L}(u - u_h), u_h \rangle|$$
$$\le C\|u - u_h\|_{H^1} (\|u\|_{H^1} + \|u_h\|_{H^1})$$
$$\le Ch^p \|u\|_{H^{p+1}}.$$

This concludes the proof.

6.2   Exact Discrete Energy Law

Next, we provide a necessary and sufficient condition for the FOSLS discretization to *exactly* satisfy an energy law, namely conditions under which we have $\langle \mathcal{L}u_h, u_h \rangle = \langle \mathcal{L}u, u \rangle$. Recall the assumption that $u_0 \in \mathcal{V}_h$. Consider two standard projections on the finite-element space, $\widetilde{\mathcal{V}}_{h,o}$: (1) the Galerkin projection $\Pi_h : \widetilde{\mathcal{V}} \mapsto \widetilde{\mathcal{V}}_{h,o}$; and (2) the $L^2(\widetilde{\Omega})$-orthogonal projection, $Q_h : L^2(\widetilde{\Omega}) \mapsto \widetilde{\mathcal{V}}_{h,o}$. These operators are defined in a standard fashion:

$$\langle \mathcal{L}\Pi_h u, v_h \rangle := \langle \mathcal{L}u, v_h \rangle, \quad \text{for all } v_h \in \widetilde{\mathcal{V}}_{h,o} \text{ and } u \in \widetilde{\mathcal{V}},$$
$$\langle Q_h u, v_h \rangle := \langle u, v_h \rangle, \quad \text{for all } v_h \in \widetilde{\mathcal{V}}_{h,o} \text{ and } u \in L^2(\widetilde{\Omega}).$$

Consider a well-known identity (see for example [23] for the case of symmetric $\mathcal{L}$) relating Π_h and Q_h, which is used in the later proof of Theorem 3.

Lemma 5. *The projections Q_h and Π_h satisfy the relation*

$$\mathcal{L}_h \Pi_h = Q_h \mathcal{L}, \tag{35}$$

where $\mathcal{L}_h : \widetilde{\mathcal{V}}_h \mapsto \widetilde{\mathcal{V}}_h$ is the restriction of $\mathcal{L}$ on $\widetilde{\mathcal{V}}_h$, namely,

$$\langle \mathcal{L}_h v_h, w_h \rangle = \langle \mathcal{L} v_h, w_h \rangle, \quad \text{for all} \quad v_h,\ w_h \in \widetilde{\mathcal{V}}_h.$$

Proof. The result easily follows from the definitions of Q_h, Π_h, $\mathcal{L}_h$, and the fact that $\mathcal{L}_h \Pi_h v \in \widetilde{\mathcal{V}}_h$. For $v \in \widetilde{\mathcal{V}}$, and $w \in \widetilde{\mathcal{V}}$ we have

$$\langle \mathcal{L}_h \Pi_h v, w \rangle = \langle \mathcal{L}_h \Pi_h v, Q_h w \rangle = \langle \mathcal{L} \Pi_h v, Q_h w \rangle$$
$$= \langle \mathcal{L} v, Q_h w \rangle = \langle Q_h \mathcal{L} v, Q_h w \rangle = \langle Q_h \mathcal{L} v, w \rangle.$$

This completes the proof.

Note that we use $Q_h \chi_h = \Pi_h \chi_h = \chi_h$ for all $\chi_h \in \widetilde{\mathcal{V}}_{h,o}$. In general, such an identity is not true for $\chi_h \in \widetilde{\mathcal{V}}_h$. However, we can relate the solution to (33) to a discrete analogue of the energy law (30) using Lemma 5. Further, notice that the FOSLS solution, u_h, satisfies $\langle \mathcal{L} u_h, \mathcal{L} \chi_h \rangle = 0$ only for $\chi_h \in \widetilde{\mathcal{V}}_{h,o}$ corresponding to a zero initial guess. Thus, it is not obvious how to estimate $\langle \mathcal{L} u_h, u_h \rangle - \langle \mathcal{L} u, u \rangle$.

Theorem 3. *The solution u_h of (33) satisfies the discrete energy law $\langle \mathcal{L} u_h, u_h \rangle = \langle \mathcal{L} u, u \rangle$ if and only if there exists a $w_h \in \widetilde{\mathcal{V}}_h$ satisfying the initial condition $w_h(x,0) = u_0(x)$ and if $\langle \mathcal{L} u_h, w_h \rangle = \langle \mathcal{L} u, u \rangle$.*

Proof. Let $w_h \in \widetilde{\mathcal{V}}_h$ be any extension of $u_0 \in V_h$ in $\widetilde{\Omega}$, that is, w_h satisfies the initial condition. The following relations follow directly from the definitions given earlier, Eq. (33), and Lemma 5.

$$\langle \mathcal{L} u_h, u_h \rangle = \left\langle \mathcal{L} u_h, \underbrace{(u_h - w_h)}_{\in \widetilde{\mathcal{V}}_{h,o}} \right\rangle + \langle \mathcal{L} u_h, w_h \rangle$$
$$= \langle \mathcal{L} u_h, Q_h(u_h - w_h) \rangle + \langle \mathcal{L} u_h, w_h \rangle$$
$$= \langle \mathcal{L} u_h, Q_h \mathcal{L} \mathcal{L}^{-1}(u_h - w_h) \rangle + \langle \mathcal{L} u_h, w_h \rangle$$
$$= \left\langle \mathcal{L} u_h, \mathcal{L} \underbrace{\Pi_h \mathcal{L}^{-1}(u_h - w_h)}_{v_h \in \widetilde{\mathcal{V}}_{h,o}} \right\rangle + \langle \mathcal{L} u_h, w_h \rangle = \langle \mathcal{L} u_h, w_h \rangle.$$

In the last identity, we use the fact that $v_h = \Pi_h \mathcal{L}^{-1}(u_h - w_h)$ is an element of $\widetilde{\mathcal{V}}_{h,o}$ and the first term on the right side vanishes (by Eq. (33)). As a result, we have

$$\langle \mathcal{L} u, u \rangle - \langle \mathcal{L} u_h, u_h \rangle = \langle \mathcal{L} u, u \rangle - \langle \mathcal{L} u_h, w_h \rangle.$$

which gives the desired necessary and sufficient condition.

From the proof, we immediately obtain the following relation,

$$| \langle \mathcal{L}u, u \rangle - \langle \mathcal{L}u_h, u_h \rangle | = \inf_{w_h} \{ | \langle \mathcal{L}u, u \rangle - \langle \mathcal{L}u_h, w_h \rangle |, \ w_h(\cdot, 0) = u_0 \} . \qquad (36)$$

In addition to the estimate in Theorem 2, it is plausible that one can use the right side of (36) to obtain a sharper result. While this is beyond the scope of this paper, some comments are in order. The difficulties associated with each particular case in hand (heat equation, Stokes' equation, etc.) amount to estimating the quantity on the right side of (36) and such estimates depend on the spaces chosen for discretization and how well the timestepping approximates the space-time formulation. Sharper estimates on the error in discrete energy law, which uses (36), can lead to sharper bounds on the constant defined in (19).

7   Conclusions

In this work, we have shown numerically that convergence of the discrete energy law is of order higher than the finite-element approximation order for two typical transient problems. Thus, while it is known that the FOSLS method may have issues with adherence to some conservation laws (i.e., mass conservation), energy conservation is not such an issue, and can be satisfied with high accuracy. The rigorous theoretical justification of such claims are topics of current and future research.

References

1. Adler, J.H., Manteuffel, T.A., McCormick, S.F., Nolting, J.W., Ruge, J.W., Tang, L.: Efficiency based adaptive local refinement for first-order system least-squares formulations. SIAM J. Sci. Comput. **33**(1), 1–24 (2011). http://dx.doi.org/10.1137/100786897
2. Adler, J.H., Manteuffel, T.A., McCormick, S.F., Ruge, J.W.: First-order system least squares for incompressible resistive magnetohydrodynamics. SIAM J. Sci. Comput. **32**(1), 229–248 (2010). http://dx.doi.org/10.1137/080727282
3. Adler, J.H., Manteuffel, T.A., McCormick, S.F., Ruge, J.W., Sanders, G.D.: Nested iteration and first-order system least squares for incompressible, resistive magnetohydrodynamics. SIAM J. Sci. Comput. **32**(3), 1506–1526 (2010). http://dx.doi.org/10.1137/090766905
4. Bochev, P., Cai, Z., Manteuffel, T.A., McCormick, S.F.: Analysis of velocity-flux first-order system least-squares principles for the Navier-Stokes equations. I. SIAM J. Numer. Anal. **35**(3), 990–1009 (1998). http://dx.doi.org/10.1137/S0036142996313592
5. Bochev, P., Gunzburger, M.: Analysis of least-squares finite element mehtods for the Stokes equations. Math. Comput. **63**(208), 479–506 (1994)

6. Bochev, P., Manteuffel, T.A., McCormick, S.F.: Analysis of velocity-flux least-squares principles for the Navier-Stokes equations. II. SIAM J. Numer. Anal. **36**(4), 1125–1144 (1999). (Electronic). http://dx.doi.org/10.1137/S0036142997324976

7. Bramble, J.H., Kolev, T.V., Pasciak, J.: A least-squares approximation method for the time-harmonic Maxwell equations. J. Numer. Math. **13**, 237–263 (2005)

8. Cai, Z., Lazarov, R., Manteuffel, T.A., McCormick, S.F.: First-order system least squares for second-order partial differential equations. I. SIAM J. Numer. Anal. **31**(6), 1785–1799 (1994). http://dx.doi.org/10.1137/0731091

9. Cai, Z., Manteuffel, T.A., McCormick, S.F.: First-order system least squares for second-order partial differential equations. II. SIAM J. Numer. Anal. **34**(2), 425–454 (1997). http://dx.doi.org/10.1137/S0036142994266066

10. Davis, T.A.: Algorithm 832: Umfpack v4.3–an unsymmetric-pattern multifrontal method. ACM Trans. Math. Softw. **30**(2), 196–199 (2004). http://doi.acm.org/10.1145/992200.992206

11. Feng, J., Liu, C., Shen, J., Yue, P.: A energetic variational formulation with phase field methods for interfacial dynamics of complex fluids: advantages and challenges. In: Calderer, M.C.T., Terentjev, E.M. (eds.) Modeling of Soft Matter. The IMA Volumes in Mathematics and its Applications, vol. 141, pp. 1–26. Springer, New York (2005). https://doi.org/10.1007/0-387-32153-5_1

12. Gelfand, I.M., Fomin, S.V.: Calculus of Variations. Prentice-Hall Inc., Englewood Cliffs (1963). Revised English edition translated and edited by R.A. Silverman

13. Girault, V., Raviart, P.A.: Finite Element Approximation of the Navier-Stokes Equations. LNM, vol. 749. Springer, Berlin (1979). https://doi.org/10.1007/BFb0063447

14. Heys, J.J., Lee, E., Manteuffel, T.A., McCormick, S.F.: An alternative least-squares formulation of the Navier-Stokes equations with improved mass conservation. J. Comput. Phys. **226**(1), 994–1006 (2007). http://dx.doi.org/10.1016/j.jcp.2007.05.005

15. Hyon, Y., Kwak, D.Y., Liu, C.: Energetic variational approach in complex fluids: maximum dissipation principle. Discret. Contin. Dyn. Syst. **26**(4), 1291–1304 (2010). http://dx.doi.org/10.3934/dcds.2010.26.1291

16. Johnson, C.: Numerical Solution of Partial Differential Equations by the Finite Element Method. Dover Publications Inc., Mineola (2009). Reprint of the 1987 edition

17. Johnson, C., Nävert, U., Pitkäranta, J.: Finite element methods for linear hyperbolic problems. Comput. Methods Appl. Mech. Eng. **45**(1–3), 285–312 (1984). http://dx.doi.org/10.1016/0045-7825(84)90158-0

18. Langer, U., Moore, S.E., Neumüller, M.: Space-time isogeometric analysis of parabolic evolution problems. Comput. Methods Appl. Mech. Eng. **306**, 342–363 (2016). http://dx.doi.org/10.1016/j.cma.2016.03.042

19. Masud, A., Hughes, T.J.R.: A space-time Galerkin/least-squares finite element formulation of the Navier-Stokes equations for moving domain problems. Comput. Methods Appl. Mech. Eng. **146**(1–2), 91–126 (1997). http://dx.doi.org/10.1016/S0045-7825(96)01222-4

20. MFEM: Modular finite element methods library (2016). http://mfem.org

21. Solonnikov, V.A.: Estimates for solutions of a non-stationary linearized system of Navier-Stokes equations. Trudy Mat. Inst. Steklov. **70**, 213–317 (1964)

22. Solonnikov, V.A.: On boundary value problems for linear parabolic systems of differential equations of general form. Trudy Mat. Inst. Steklov. **83**, 3–163 (1965)

23. Xu, J.: Iterative methods by space decomposition and subspace correction. SIAM Rev. **34**(4), 581–613 (1992). http://dx.doi.org/10.1137/1034116