Journal of Computational Physics 517 (2024) 113298

Contents lists available at ScienceDirect

Journal of Computational Physics

journal homepage: www.elsevier.com/locate/jcp

Adjoint-based enforcement of state constraints in PDE optimization problems

Pritpal Matharu [a,b,*], Bartosz Protas [b]

[a] Department of Mathematics, KTH Royal Institute of Technology, Stockholm, Sweden
[b] Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada

ARTICLE INFO

Keywords:
PDE optimization
Adjoint analysis
State constraints
Heat transfer
Turbulence

ABSTRACT

This study demonstrates how the adjoint-based framework traditionally used to compute gradients in PDE optimization problems can be extended to handle general constraints on the state variables. This is accomplished by constructing a projection of the gradient of the objective functional onto a subspace tangent to the manifold defined by the constraint. This projection is realized by solving an adjoint problem defined in terms of the same adjoint operator as used in the system employed to determine the gradient, but with a different forcing. We focus on the "optimize-then-discretize" paradigm in the infinite-dimensional setting where the required regularity of both the gradient and of the projection is ensured. The proposed approach is illustrated with two examples: a simple test problem describing optimization of heat transfer in one direction and a more involved problem where an optimal closure is found for a turbulent flow described by the Navier-Stokes system in two dimensions, both considered subject to different state constraints. The accuracy of the gradients and projections computed by solving suitable adjoint systems is carefully verified and the presented computational results show that the solutions of the optimization problems obtained with the proposed approach satisfy the state constraints with a good accuracy, although not exactly.

1. Introduction

Numerous problems in modern science and engineering are cast in terms of optimization of systems described by partial differential equations (PDEs), often subject to complex constraints imposed on the control (decision) variables. A standard framework for handling such constrained PDE optimization problems is based on Lagrange multipliers [1,2] where the objective functional is considered a function of both the state and the control variable, the two related by suitable constraints. However, this formalism can lead to challenging computational problems since the system of the first-order optimality conditions has the form of a saddle-type problem which is often difficult to solve numerically. As an alternative, one can consider formulations based on the reduced objective functional depending on the control variable only (i.e., where the state variable is *not* a separate independent variable, but is considered a function of the control variable), such that the optimization problem can be solved using some discrete form of the gradient flow in which the constraints are built into the definition of the space in which the gradient flow is constructed. An advantage of this approach is that a saddle-type (min-max) problem is avoided and one can take advantage of computational methods of unconstrained optimization.

* Corresponding author at: Department of Mathematics, KTH Royal Institute of Technology, Stockholm, Sweden.
E-mail address: pritpal@kth.se (P. Matharu).
URL: https://people.kth.se/~pritpal/ (P. Matharu).

https://doi.org/10.1016/j.jcp.2024.113298
Received 3 December 2023; Received in revised form 15 July 2024; Accepted 20 July 2024

Available online 25 July 2024

When dealing with infinite-dimensional PDE optimization problems a key element of this approach is determination of the gradient of the reduced objective functional with respect to the control variable which can be conveniently computed by solving a suitably-defined *adjoint system* [3,4]. These methods are already quite well developed and there exists a large body of literature on this topic, with applications in various areas [5–8]. However, methods based on discrete gradient flows can be challenging to apply in the presence of complicated constraints, especially when they involve both state and control variables as is the case in some applications (one research area where such problems are common is topology optimization [9]). In the present study we demonstrate how such constraints can be approximately enforced using the solution of a suitable adjoint problem, similar but distinct from the adjoint system used to determine the gradient of the reduced objective functional. We focus on the "optimize-then-discretize" approach where all elements of the algorithm are derived in the continuous setting and only then discretized for the purpose of the numerical solution [4]. To the best of our knowledge, this is the first such generalization of adjoint-based techniques in the solution of PDE optimization problems.

To illustrate the ideas described above, we consider the state $u \in \mathcal{X}$ and the control variable $\varphi \in \mathcal{Y}$ with $\mathcal{X}$ and $\mathcal{Y}$ denoting the appropriate functional spaces, to be specified below. The objective functional then is $j : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ and the PDE constraint is given in terms of some function $S : \mathcal{X} \times \mathcal{Y} \to \mathcal{X}^*$ representing the weak form of the PDE with $\mathcal{X}^*$ the dual space of $\mathcal{X}$. In addition, we will also assume that the state and the control variables are subject to $m \in \mathbb{N}^+$ scalar constraints given by the function $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^m$ assumed sufficiently smooth. Our constrained PDE optimization problem can then be stated as

$$\text{(A)} \qquad \begin{aligned} &\min_{(u,\varphi)\in\mathcal{X}\times\mathcal{Y}} j(u,\varphi) \\ &\text{subject to}: \quad \begin{cases} S(u,\varphi) = 0, \\ c(u,\varphi) = 0 \end{cases} \end{aligned}.$$

Introducing the Lagrange multipliers $\lambda \in \mathcal{X}$, $\mu \in \mathbb{R}^m$ and the augmented objective functional, i.e., the Lagrangian, $\mathcal{L}(u,\varphi,\lambda,\mu) := j(u,\varphi) + \langle \lambda, S(u,\varphi)\rangle_{\mathcal{X}\times\mathcal{X}^*} + \langle \mu, c(u,\varphi)\rangle_{\mathbb{R}^m}$, where $\langle\cdot,\cdot\rangle_{\mathcal{X}\times\mathcal{X}^*}$ is the duality pairing between the spaces $\mathcal{X}$ and $\mathcal{X}^*$ [1], $\langle\cdot,\cdot\rangle_{\mathbb{R}^m}$ the inner product in $\mathbb{R}^m$, and ":=" means "equal to by definition", the constrained problem (A) can be recast in the corresponding unconstrained form as [1]

$$\text{(B)} \qquad \min_{(u,\varphi)\in\mathcal{X}\times\mathcal{Y}} \max_{(\lambda,\mu)\in\mathcal{X}^*\times\mathbb{R}^m} \mathcal{L}(u,\varphi,\lambda,\mu).$$

Critical points of this problem are characterized by the appropriate first-order optimality conditions. Besides handling the saddle-type (min-max) nature of these critical points, an extra complication is the need to also approximate the optimal Lagrange multipliers $\breve{\lambda}$ and $\breve{\mu}$ in addition to the optimal state and control variables $\breve{u}$ and $\breve{\phi}$ (hereafter, the symbol $\breve{\cdot}$ will denote the optimal value of a variable).

Defining the *reduced* objective functional $\mathcal{J}(\varphi) := j(u(\varphi),\varphi)$, which assumes that for each φ the map $u = u(\varphi)$ is well defined in terms of the solution of the PDE problem $S(u(\varphi),\varphi) = 0$, Problem (A) can be recast such that minimization is performed with respect to φ only, i.e.,

$$\text{(C)} \qquad \begin{aligned} &\min_{\varphi\in\mathcal{Y}} \mathcal{J}(\varphi) \\ &\text{subject to}: \quad c(u(\varphi),\varphi) = 0 \end{aligned},$$

where the constraint can be interpreted as defining a codimension-m manifold $\mathcal{M} := \{\varphi \in \mathcal{Y} : \quad c(u(\varphi),\varphi) = 0\}$. Thus, since the optimizer $\breve{\varphi}$ must be sought on this constraint manifold, $\breve{\varphi} \in \mathcal{M}$, problem (C) defines a *Riemannian* optimization problem [10].

A local minimizer of Problem (C) can then be approximated with a discrete projected gradient flow as $\breve{\varphi} = \lim_{n\to\infty} \varphi^{(n)}$, where

$$\varphi^{(n+1)} = \varphi^{(n)} - \tau_n P_{\mathcal{T}\mathcal{M}_{\varphi^{(n)}}}\left(\nabla_\varphi \mathcal{J}(\varphi^{(n)})\right), \qquad n = 0,1,\ldots, \tag{1a}$$

$$\varphi^{(0)} = \varphi_0, \tag{1b}$$

in which $P_{\mathcal{T}\mathcal{M}_\varphi} : \mathcal{Y} \to \mathcal{T}\mathcal{M}_\varphi$ is the operator representing the orthogonal projection onto the subspace $\mathcal{T}\mathcal{M}_\varphi \in \mathcal{Y}$ tangent to the constraint manifold $\mathcal{M}$ at $\varphi \in \mathcal{M}$, τ_n is the step size along the projected gradient, $\nabla_\varphi \mathcal{J}$ is the gradient of the reduced functional $\mathcal{J}(\varphi)$ with respect to φ, and φ_0 the initial guess. As illustrated in Fig. 1, in iterations (1) the constraint is satisfied approximately only, with an error $\mathcal{O}(\tau_n^2)$ at each iteration. This error is eliminated if one can introduce the retraction operator $\mathcal{R}_\mathcal{M} : \mathcal{T}\mathcal{M} \to \mathcal{M}$, where $\mathcal{T}\mathcal{M}$ is the tangent bundle, such that the right-hand side (RHS) in (1a) is replaced with $\mathcal{R}_\mathcal{M}\left(\varphi^{(n)} - \tau_n P_{\mathcal{T}\mathcal{M}_{\varphi^{(n)}}}\left(\nabla_\varphi \mathcal{J}(\varphi^{(n)})\right)\right)$ yielding the Riemannian gradient approach [10]. However, for general manifolds $\mathcal{M}$ the retraction operator can be very difficult to implement and as a result one often needs to resort to using (1). While the gradient $\nabla_\varphi \mathcal{J}(\varphi)$ of the (reduced) objective functional can be conveniently computed by solving an adjoint problem [4], the main contribution of the present study is to demonstrate that an analogous computation can in fact be also used to determine the projection operator $P_{\mathcal{T}\mathcal{M}_\varphi}$. This problem is nuanced by the fact that both the gradient $\nabla_\varphi \mathcal{J}(\varphi)$ and the projector $P_{\mathcal{T}\mathcal{M}_\varphi}$ need to be determined with respect to a particular topology, typically induced by the norm in the space $\mathcal{Y}$. This contribution thus expands the scope of applications of adjoint analysis in PDE constrained optimization.

We introduce and illustrate the new approach by way of two examples: the first one is rather academic and concerns a simple control problem involving a one-dimensional (1D) heat equation; the second is more involved and represents an extension of a recently formulated problem concerning finding optimal turbulence closures in two-dimensional (2D) Navier-Stokes flows [11]. While the first example is intended to serve as a simple illustration only of the proposed approach, the second one showcases an application to a

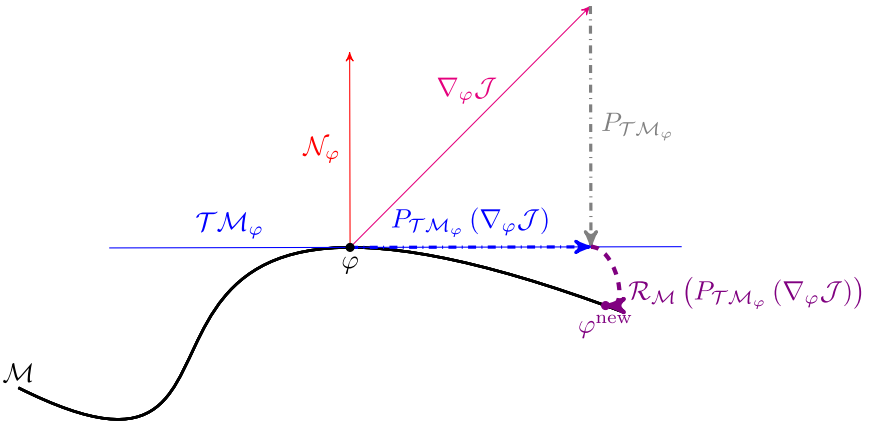

Fig. 1. Schematic representation of the projection of the gradient $\nabla_\varphi \mathcal{J}$ onto the subspace $\mathcal{TM}_\varphi$ tangent to a general manifold $\mathcal{M}$ at φ. The projection operator $P_{\mathcal{TM}_\varphi}$ uses the element $\mathcal{N}_\varphi$ normal to the tangent subspace $\mathcal{TM}_\varphi$, cf. (28). The retraction operator $\mathcal{R}_\mathcal{M}$ then maps elements of the tangent subspace $\mathcal{TM}_\varphi$ back to the manifold $\mathcal{M}$, cf. (29), to a new element φ^{new}. The objects shown in the figure are to be interpreted as elements of a function space, typically a Sobolev space [12], and can be functions of time as in Problems 1 and 2, or state as in Problem 3.

nontrivial problem with a nonstandard structure. The structure of the paper is as follows: in the next section we introduce the two PDE optimization problems whereas the proposed approach is presented in Section 3 with a particular emphasis on how the projection onto the subspace tangent to the constraint manifold can be constructed based on the solution of an adjoint system; computational results are then presented in Section 5 with a summary and conclusions deferred to Section 6.

2. Test problems

In this section we introduce two constrained PDE optimization problems that will serve as our test cases. While the first problem is classical and is used primarily to illustrate our approach, the second one is more challenging due to its non-standard structure. With a slight abuse of notation motivated by the desire to highlight the analogies between them, in the two test problems we reuse the symbols denoting key elements of the formulation of the optimization problems, with the exception of ϕ and φ which represent the control variable in Section 2.1 and Section 2.2, respectively.

2.1. Control of heat conduction in 1D

We consider heat conduction in a finite rod $\Omega := [a,b] \subset \mathbb{R}$, where $-\infty < a < b < \infty$, in which the time-dependent heat flux $\phi(t)$, $t \in [0,T]$ with some $T > 0$, is applied at the left boundary $x = a$ such that the resulting temperature at the right boundary $x = b$ should match some prescribed history $\bar{u}_b(t)$, $t \in [0,T]$, and at the same time the average "energy" of the system should be given by $E_0 > 0$. The problem is thus described by the system

$$\frac{\partial u}{\partial t} - \Delta u = 0, \qquad (t,x) \in (0,T] \times \Omega, \tag{2a}$$

$$\frac{\partial u}{\partial x}\Big|_{x=a} = \phi(t), \qquad t \in (0,T], \tag{2b}$$

$$\frac{\partial u}{\partial x}\Big|_{x=b} = 0, \qquad t \in (0,T], \tag{2c}$$

$$u(t=0) = u_0, \qquad x \in \Omega, \tag{2d}$$

where $u_0 : \Omega \to \mathbb{R}$ is the initial condition and the solution $u = u(t,x;\phi)$ is assumed to depend on the flux ϕ as the control variable. The energy E is defined as

$$E(t;\phi) := \frac{1}{2}\int_\Omega u(t,x;\phi)^2 \, dx, \tag{3}$$

whereas $[f]_T := (1/T)\int_0^T f(t)\,dt$ will denote the time average of some function $f : [0,T] \to \mathbb{R}$. Next, we introduce the Sobolev space $H^p(\Lambda)$ with $p \in \mathbb{N}$ of functions defined on some domain Λ with p square-integrable distributional derivatives and endowed with the inner product

$$\forall\ z_1, z_2 \in H^p(\Lambda) \qquad \langle z_1, z_2 \rangle_{H^p(\Lambda)} := \sum_{q=0}^{p} \ell_q^{2q} \left\langle \frac{d^q z_1}{dx^q}, \frac{d^q z_2}{dx^q} \right\rangle_{L^2(\Lambda)}, \tag{4}$$

where $\langle z_1, z_2 \rangle_{L^2(\Lambda)} := \int_\Lambda z_1 z_2 \, dx$, whereas $\ell_q \in \mathbb{R}^+$, $q = 0, \ldots, p$, are "length-scale" parameters (in the multidimensional case, i.e., when $\Lambda \subseteq \mathbb{R}^n$, $n > 1$, the definition (4) admits a natural generalization [12]). While for different values of $0 < \ell_q < \infty$ the inner

products in (4) are equivalent (in the precise sense of norm equivalence), in Section 5 we will show that the flexibility represented by these parameters significantly improves the convergence of the our iterative minimization algorithm (1).

It is instructive to consider two formulations: one in which the flux is simply assumed to be a square-integrable function of time together with its first derivative, i.e., $\phi \in \mathcal{S} := H^1(0,T)$, and another one where it is additionally assumed to belong to a quadratic manifold, $\phi \in \mathcal{M} \subset \mathcal{S}$, which is defined as

$$\mathcal{M} := \left\{ \phi \in \mathcal{S} : \quad [E(\cdot;\phi)]_T = \frac{1}{2T} \int_0^T \int_\Omega u(t,x;\phi)^2 \, dx dt = E_0 \right\}. \tag{5}$$

Our (reduced) objective functional $\mathcal{J} : \mathcal{M} \to \mathbb{R}$ is a measure of the error between the actual temperature at the right boundary $u(t,b;\phi)$ and the prescribed profile $\bar{u}_b(t)$

$$\mathcal{J}(\phi) = \frac{1}{2} \int_0^T [u(\phi)|_b - \bar{u}_b]^2 \, dt. \tag{6}$$

We note that the quadratic constraint defining the manifold $\mathcal{M}$ in (5) simplifies when the governing system (2) is subject to the homogeneous initial condition $u_0(x) = 0$, $\forall\ x \in \Omega$ ($u_0 \equiv 0$). In such case the quadratic constraint becomes homogeneous, i.e., $[E(\cdot;\beta\phi)]_T = \beta^2 [E(\cdot;\phi)]_T$ for some $\beta \in \mathbb{R}$, which reduces enforcement of this constraint to a simple rescaling. Otherwise, enforcement of the constraint is more involved and necessitates the use of a retraction operator. For comparison, we thus consider two versions of each of the optimization problems depending on whether or not the initial condition u_0 is homogeneous, i.e.,

Problem 1. *Given system* (2) *with* $u_0 \equiv 0$ *and objective functional* (6)*, find*

$$\breve{\phi} = \arg\min_{\phi \in \mathcal{S}} \mathcal{J}(\phi), \tag{Problem 1.A}$$

$$\breve{\phi} = \arg\min_{\phi \in \mathcal{M}} \mathcal{J}(\phi). \tag{Problem 1.B}$$

Problem 2. *Given system* (2) *with* $u_0 \not\equiv 0$ *and objective functional* (6)*, find*

$$\breve{\phi} = \arg\min_{\phi \in \mathcal{S}} \mathcal{J}(\phi), \tag{Problem 2.A}$$

$$\breve{\phi} = \arg\min_{\phi \in \mathcal{M}} \mathcal{J}(\phi). \tag{Problem 2.B}$$

We add that since in Problem 1.A and Problem 2.A minimization is performed over the entire linear subspace $\mathcal{S}$, we will consider these problems as "unconstrained". In contrast, Problem 1.B and Problem 2.B, where optimization is carried out over the manifold $\mathcal{M} \subset \mathcal{S}$, will be regarded as "constrained".

2.2. Optimal eddy viscosity in closure models for 2D turbulent flows

Turbulent flows governed by the incompressible Navier-Stokes system remain very challenging to compute when the Reynolds number is high due to difficulties resolving motions occurring at small spatial and temporal scales. A solution often adopted in practice relies on solving a filtered version of the governing equations such that the number of resolved degrees of freedom is reduced, an approach referred to as the Large-Eddy Simulation (LES) [13–15]. However, the LES system is not closed as it involves terms explicitly depending on the unresolved degrees of freedom and closing this system in terms of the resolved degrees of freedom leads to the celebrated turbulence closure problem. While a significant body of literature has been devoted to deriving turbulence models adapted to different flows, most of these approaches have been empirical in nature. Recent advances in machine learning have enabled the development of data-based techniques for deducing closure models [16–22]. As an alternative to these approaches, a data-based technique relying on calculus of variations and methods of PDE optimization was recently proposed in [23,11]. It allows one to find an optimal, in a mathematically precise sense, functional form of the eddy viscosity appearing in a commonly used family of turbulence closure models. The problem considered here arises as an extension of that approach.

We consider the flow of viscous incompressible fluids on a two-dimensional (2D) periodic domain $\Omega = [0, 2\pi]^2$ and over a time interval $[0,T]$ for some $T > 0$. Assuming uniform fluid density $\rho = 1$, the motion is governed by the Navier-Stokes system, which written in vorticity-streamfunction form, is

$$\partial_t w + \boldsymbol{\nabla}^\perp \psi \cdot \boldsymbol{\nabla} w = \nu_N \boldsymbol{\Delta} w - \alpha w + f_\omega \quad \text{in} \quad (0,T] \times \Omega, \tag{7a}$$

$$\boldsymbol{\Delta} \psi = -w \quad \text{in} \quad (0,T] \times \Omega, \tag{7b}$$

$$w(t=0) = w_0 \quad \text{in} \quad \Omega, \tag{7c}$$

where $w := -\boldsymbol{\nabla}^{\perp} \cdot \boldsymbol{u}$, with $\boldsymbol{\nabla}^{\perp} := [\partial_{x_2}, -\partial_{x_1}]^T$ and $\boldsymbol{u}$ the velocity field, is the vorticity component perpendicular to the plane of motion, ψ is the streamfunction, ν_N is the coefficient of the kinematic viscosity (for simplicity of notation, we reserve the symbol ν for the eddy viscosity), and w_0 is the initial condition. In (7a) the term proportional to $\alpha > 0$ represents large-scale dissipation (Ekman friction), whereas f_ω is a time-independent band-limited forcing acting on low wavenumbers. The parameters of these two terms are chosen such that the flow is in a statistical equilibrium with a well-developed enstrophy cascade and a rudimentary inverse energy cascade [13].

Using $\widetilde{(\cdot)}$ to denote a suitable low-pass filter, we can write the LES version of (7) as (the reader is referred to [11] for derivation details)

$$\partial_t \widetilde{\omega} + \boldsymbol{\nabla}^{\perp}\widetilde{\widetilde{\psi} \cdot \boldsymbol{\nabla}\widetilde{\omega}} = \boldsymbol{\nabla} \cdot \left(\widetilde{\left[\nu_N + \nu(s)\right] \boldsymbol{\nabla}\widetilde{\omega}} \right) - \alpha\widetilde{\omega} + f_\omega \quad \text{in} \quad (0,T]\times\Omega, \tag{8a}$$

$$\boldsymbol{\Delta}\widetilde{\psi} = -\widetilde{\omega} \quad \text{in} \quad (0,T]\times\Omega, \tag{8b}$$

$$\widetilde{\omega}(t=0) = \widetilde{\omega}_0 := \widetilde{w}_0 \quad \text{in} \quad \Omega, \tag{8c}$$

where $\widetilde{\omega}$ is the LES vorticity, and the initial condition $\widetilde{\omega}_0$ is given as the filtered initial condition (7c) and the forcing term is unaffected by the filter as it acts on the low wavenumbers only. The LES equation (8a) features a Smagorinsky-type closure model with a state-dependent eddy viscosity expressed as

$$\nu(s) = \left[\eta^3 \sqrt{s} + \nu_0\right] \varphi\left(\frac{s}{s_{\max}}\right) \qquad \text{with} \quad s := |\boldsymbol{\nabla}\widetilde{\omega}|^2 \in \mathcal{I} := [0, s_{\max}], \tag{9}$$

where $\nu_0 > 0$, $\eta = 2\pi/k_c$ is the width of the LES filter with k_c the largest resolved wavenumber, $s_{\max} > 0$ is a sufficiently large number to be specified later and $\varphi : [0,1] \to \mathbb{R}$ a non-dimensional function. The form of equation (8a) suggests that $\nu = \nu(s)$, and hence also $\varphi = \varphi(s/s_{\max})$, must be at least piecewise C^1 functions on $\mathcal{I}$ and $[0,1]$, respectively. However, as will become evident in Section 3.3, our solution approach imposes some additional regularity requirements, namely, $\nu = \nu(s)$ needs to be piecewise C^3 on $\mathcal{I}$ with the first and third derivatives vanishing at $s = 0, s_{\max}$. Since gradient-based solution approaches to PDE-constrained optimization problems are preferably formulated in Hilbert spaces [24], we shall look for an optimal function φ parametrizing the eddy viscosity as an element of the following linear space which is a subspace of the Sobolev space $H^2(\mathcal{I})$, cf (4),

$$S := \left\{ \varphi \in C^3([0,1]) \ : \ \frac{d}{d\xi}\varphi(\xi) = \frac{d^3}{d\xi^3}\varphi(\xi) = 0 \text{ at } \xi = 0, 1 \right\}. \tag{10}$$

We will seek an optimal form of the function φ such that solutions of the LES system (8) with eddy viscosity (9) best match the corresponding solutions of the original Navier-Stokes system (7) in a sense to be specified below. We add that when $\nu_0 = 0$ and $\varphi(\xi) = 1$, $\xi \in [0,1]$, then the eddy viscosity (9) reduces to the Leith model which is a counterpart of the Smagorinsky model in 2D flows [25–27].

While in [11] we considered a pointwise match, both in space and in time, between the LES and the Navier-Stokes flows, here we determine an optimal eddy viscosity $\breve{\nu}(s)$ that will allow the LES flow to match the original Navier-Stokes flow in a certain average sense. More specifically, we will formulate the problem in terms of time-averages involving important integral quantities characterizing 2D flows, namely, the enstrophy and palinstrophy which are defined as follows

$$\mathcal{E}(t) := \frac{1}{2}\int_\Omega \widetilde{w}(t,\mathbf{x})^2 \, d\mathbf{x}, \tag{11a}$$

$$\mathcal{P}(t) := \frac{1}{2}\int_\Omega |\boldsymbol{\nabla}\widetilde{w}(t,\mathbf{x})|^2 \, d\mathbf{x}, \tag{11b}$$

where $\widetilde{w}$ is the filtered solution of the Navier-Stokes system (7), whereas $\widetilde{\mathcal{E}}(t;\varphi)$ and $\widetilde{\mathcal{P}}(t;\varphi)$ will denote the corresponding quantities defined in terms of the solution $\widetilde{\omega}$ of the LES system (8) and the notation emphasizes their dependence on the function φ parameterizing the eddy viscosity, cf. ansatz (9).

The optimal eddy viscosity $\breve{\nu}(s)$ will be chosen so as to yield a good match between the palinstrophy in the LES flow (with the given eddy viscosity) and in the original Navier-Stokes flow, so that we shall consider the following error functional

$$\mathcal{J}(\varphi) := \frac{1}{2\,D}\int_0^T \left[\mathcal{P}(t) - \widetilde{\mathcal{P}}(t;\varphi)\right]^2 dt, \tag{12}$$

where $D > 0$ serves as a normalization factor. In addition, we will also require the LES flow with the optimal eddy viscosity to have a time-averaged enstrophy equal to the enstrophy in the original Navier-Stokes flow, i.e.,

$$\left[\widetilde{\mathcal{E}}(\cdot;\varphi)\right]_T = \left[\mathcal{E}(\cdot)\right]_T =: \mathcal{E}_0. \tag{13}$$

This condition can be interpreted as fixing the $L^2([0,T];L^2(\Omega))$ norm of the vorticity field in the LES flow. It thus defines the following nonlinear manifold in the space S, cf. (10), of nondimensional functions parameterizing the eddy viscosity

$$\mathcal{M} := \left\{ \varphi \in S \ : \ \left[\widetilde{\mathcal{E}}(\cdot;\varphi) \right]_T = \mathcal{E}_0, \right\}. \tag{14}$$

The optimal eddy viscosity $\breve{\nu} = \breve{\nu}(s)$ can then be obtained from the solution of the following optimization problems via ansatz (9) with the function φ replaced with $\breve{\varphi}$, where as in the previous section we consider the unconstrained and constrained formulation with and without the nonlinear constraint

Problem 3. *For the system* (8) *and objective functional* (12)*, find*

$$\breve{\varphi} := \underset{\varphi \in S}{\arg\min}\, \mathcal{J}(\varphi), \tag{Problem 3.A}$$

$$\breve{\varphi} := \underset{\varphi \in \mathcal{M}}{\arg\min}\, \mathcal{J}(\varphi). \tag{Problem 3.B}$$

We emphasize that Problem 3 has a fundamentally different structure than Problems 1 and 2, since the control variable $\varphi(s/s_{\max})$ is a function of the *dependent* (state) variable s, cf. (9), rather than the *independent* variables (t or x) as is typical in PDE-constrained optimization problems. In other words, solving Problem 3 can be interpreted as finding an optimal form of the nonlinearity in the closure model. As described in the next section, this aspect will have significant ramifications for how Problem 3 is solved.

3. Solution approach

In this section we first briefly present an adjoint-based approach to compute the gradient $\nabla_\phi \mathcal{J}(\phi)$ of the (reduced) objective functional in Problem 1 and 2, which is quite standard, and then describe how an analogous approach can be used to define the projection operator $P_{\mathcal{T}\mathcal{M}_\phi}$ realizing the orthogonal projection onto the subspace tangent to the constraint manifold $\mathcal{M}$, cf. (1a) and Fig. 1. Finally, we provide some details how these approaches can be adapted to perform analogous tasks in the solution of Problem 3. Throughout these derivations particular attention will be paid to ensuring the required regularity of the obtained solutions which will be done using the framework established in [24].

3.1. Evaluation of the cost functional gradients

In order to determine the gradient $\nabla_\phi \mathcal{J}$ of the objective functional (6), we begin by computing its Gâteaux (directional) differential in the direction of some arbitrary perturbation $\phi' \in H^1(0,T)$ of the heat flux

$$\mathcal{J}'(\phi;\phi') = \frac{d}{d\epsilon}\, \mathcal{J}\left(\phi + \epsilon\phi'\right)\Big|_{\epsilon=0} = \int_0^T [u(\phi)|_b - \bar{u}_b]\, u'(x,t;\phi,\phi')\, dt, \tag{15}$$

where $u'(t,x;\phi,\phi')$ satisfies the system obtained as a perturbation of the governing system (2)

$$\mathcal{K}u' := \frac{\partial u'}{\partial t} - \Delta u' = 0 \qquad (t,x) \in (0,T] \times \Omega, \tag{16a}$$

$$\frac{\partial u'}{\partial x}\Big|_{x=a} = \phi'(t) \qquad t \in (0,T], \tag{16b}$$

$$\frac{\partial u'}{\partial x}\Big|_{x=b} = 0 \qquad t \in (0,T], \tag{16c}$$

$$u'(t=0) = 0 \qquad x \in \Omega. \tag{16d}$$

In order to extract the gradient $\nabla_\phi \mathcal{J}$ from the Gâteaux differential (15), we use the fact that the differential is a bounded linear functional on both $L^2(0,T)$ and $H^1(0,T)$ when viewed as a function of ϕ', and invoke the Riesz representation theorem to obtain [28]

$$\mathcal{J}'(\phi;\phi') = \left\langle \nabla_\phi^{L^2} \mathcal{J}, \phi' \right\rangle_{L^2(0,T)} = \left\langle \nabla_\phi \mathcal{J}, \phi' \right\rangle_{H^1(0,T)}, \tag{17}$$

where the gradients $\nabla_\phi^{L^2} \mathcal{J} \in L^2(0,T)$ and $\nabla_\phi \mathcal{J} \in H^1(0,T)$ are the Riesz representers in the two spaces. While in (1) we require the gradient in the space $H^1(0,T)$, it is convenient to first obtain the gradient with respect to the L^2 topology. Since the expression on the RHS of (15) is not consistent with the Riesz form (17) as the perturbation ϕ' does not appear explicitly in it, but is instead hidden in the boundary condition (16b), we introduce the *adjoint field* $u^* : [0,T] \times \Omega \to \mathbb{R}$ and define the following duality-pairing relation

$$\begin{aligned}\left(\mathcal{K}u', u^*\right) &:= \int_0^T \int_\Omega (\mathcal{K}u')\, u^*\, dx\, dt \\ &= \int_0^T \int_\Omega u'\,(\mathcal{K}^* u^*)\, dx\, dt - \overbrace{\int_0^T [u(\phi)|_b - \bar{u}_b]\, u'(b,t;\phi,\phi')\, dt}^{\mathcal{J}'(\phi;\phi')} - \int_0^T u^*\Big|_{x=a} \phi'(t)\, dt = 0,\end{aligned} \tag{18}$$

where integration by parts was performed with respect to both space and time and the field u^* solves *adjoint system*

$$\mathcal{K}^* u^* := -\frac{\partial u^*}{\partial t} - \Delta u^* = 0 \qquad (t,x) \in (0,T] \times \Omega, \tag{19a}$$

$$\frac{\partial u^*}{\partial x}\Big|_{x=a} = 0 \qquad t \in (0,T], \tag{19b}$$

$$\frac{\partial u^*}{\partial x}\Big|_{x=b} = u(\phi)|_b - \bar{u}_b \qquad t \in (0,T], \tag{19c}$$

$$u^*(t=T) = 0 \qquad x \in \Omega. \tag{19d}$$

Collecting (16), (18) and (19), we obtain an expression for the Gâteaux differential consistent with the Riesz form (17), namely, $\mathcal{J}'(\phi;\phi') = -\int_0^T u^*\Big|_{x=a} \phi'(t)\, dt$. The gradient defined with respect to the L^2 topology is then deduced using the first equality in (17)

$$\nabla_\phi^{L^2} \mathcal{J}(t) = -u^*(t,x)\Big|_{x=a}. \tag{20}$$

The required Sobolev gradient $\nabla_\phi \mathcal{J}$ is then obtained using the second equality in (17) and the definition (4) which gives

$$\mathcal{J}'(\phi;\phi') = \int_0^T \nabla_\phi^{L^2} \mathcal{J}\, \phi'\, dt = \int_0^T \nabla_\phi \mathcal{J}\, \phi'\, dt + \ell^2 \int_0^T \frac{d(\nabla_\phi \mathcal{J})}{dt}\, \frac{d\phi'}{dt}\, dt. \tag{21}$$

Performing integration by parts with respect to time t in the second term and assuming that both the perturbation and the gradient vanish at the endpoints of the time interval, $\phi'(0) = \phi'(T) = 0$ and $\nabla_\phi \mathcal{J}\Big|_{t=0} = \nabla_\phi \mathcal{J}\Big|_{t=T} = 0$, we obtain the Sobolev gradient as the solution of the elliptic boundary-value problem [24]

$$\left[\mathrm{Id} - \ell^2 \frac{d^2}{dt^2}\right] \nabla_\phi \mathcal{J}(t) = \nabla_\phi^{L^2} \mathcal{J}(t) \qquad t \in (0,T), \tag{22a}$$

$$\nabla_\phi \mathcal{J}\Big|_{t=0} = \nabla_\phi \mathcal{J}\Big|_{t=T} = 0. \tag{22b}$$

Given the form of the discrete gradient flow (1), this ensures that the behavior of the control variable at the endpoints $t = 0$ and $t = T$ can be prescribed in the initial guess ϕ_0 (in other words, with the Sobolev gradients defined as in (17) and (22), the discrete gradient flow (1) does not affect the boundary values of the control variable, which is desirable in some applications). It can be shown [24] that extraction of Sobolev gradients via the boundary-value problem (22) can be interpreted as application of a low-pass filter to the L^2 gradient with ℓ acting as the cut-off parameter (it is thus a smoothing operation where Fourier components of the gradient with wavelengths shorter than ℓ are damped).

3.2. Projection onto the subspace tangent to the constraint manifold

Projection onto the subspace $\mathcal{TM}_\phi$ tangent to the manifold $\mathcal{M}$ at a given element $\phi \in \mathcal{M}$ is a key step in enforcing the associated constraint, cf. Fig. 1. Since in the present problem the codimension of the manifold is one, cf. (5), the tangent subspace can be characterized in terms of an element $\mathcal{N}_\phi \in H^1(0,T)$ normal to it in a suitable sense (if the codimension of the manifold is higher, $\mathrm{codim}(\mathcal{M}) > 1$, then there would be $\mathrm{codim}(\mathcal{M})$ such elements and the considerations below generalize in a natural way). We note that the projection operator $P_{\mathcal{TM}_\phi}$ in the discrete gradient flow (1) needs to be defined with respect to the H^1 inner product, cf. (4), as this is the topology of the ambient space. However, as we did above, we will first obtain the normal element $\mathcal{N}_\phi^{L^2}$ defined with respect to the L^2 inner product and then deduce its Sobolev counterpart, both as functions of time.

We begin by considering the Gâteaux differential of the constraint $[E(\cdot;\phi)]_T = E_0$, cf. (5), which yields

$$\left[E'(\cdot;\phi,\phi')\right]_T := \frac{d}{d\epsilon}\, \left[E(\cdot;\phi+\epsilon\phi')\right]_T \Big|_{\epsilon=0} = \frac{1}{T} \int_0^T \int_\Omega u(t,x;\phi)\, u'(t,x;\phi,\phi')\, dx\, dt = 0, \tag{23}$$

where $u'(t,x;\phi,\phi')$ is the solution of the perturbation system (16). We then seek to express the directional differential (23), which is a bounded linear functional of ϕ' on $L^2(0,T)$ and $H^1(0,T)$, in terms of Riesz identities in these spaces, i.e.,

$$\left[E'(\cdot;\phi,\phi')\right]_T = \left\langle \mathcal{N}_\phi^{L^2}, \phi' \right\rangle_{L^2(0,T)} = \left\langle \mathcal{N}_\phi, \phi' \right\rangle_{H^1(0,T)} = 0. \tag{24}$$

Noting that the expression on the RHS of (23) does not explicitly depend on the perturbation ϕ' which instead appears in the boundary condition (16b), we introduce a *new* adjoint state $v^* : [0,T]\times\Omega \to \mathbb{R}$ that will satisfy the following adjoint system

$$\mathcal{K}^* v^* := -\frac{\partial v^*}{\partial t} - \Delta v^* = u(t,x;\phi) \qquad (t,x) \in (0,T]\times\Omega, \tag{25a}$$

$$\left.\frac{\partial v^*}{\partial x}\right|_{x=a} = 0 \qquad t \in (0,T], \tag{25b}$$

$$\left.\frac{\partial v^*}{\partial x}\right|_{x=b} = 0 \qquad t \in (0,T], \tag{25c}$$

$$v^*(t=T) = 0 \qquad x \in \Omega. \tag{25d}$$

Then, introducing the duality relation

$$\left(\mathcal{K}u', v^*\right) := \int_0^T \int_\Omega (\mathcal{K}u') v^* \, dx\, dt = \overbrace{\int_0^T \int_\Omega u'(\mathcal{K}^* v^*)\, dx\, dt}^{\left[E'(\cdot;\phi,\phi')\right]_T} + \int_0^T v^*\Big|_{x=a} \phi'(t)\, dt = 0, \tag{26}$$

we conclude that $\left[E'(\cdot;\phi,\phi')\right]_T = -\int_0^T v^*\Big|_{x=a} \phi'(t)\, dt$. Using the first equality in (24) provides an expression for the element normal to the tangent subspace $\mathcal{T}\mathcal{M}_\phi$ with respect to the L^2 inner product, namely,

$$\mathcal{N}_\phi^{L^2}(t) = -v^*(t,a), \qquad t \in [0,T]. \tag{27}$$

Using the second equality in (24) and following the same steps as above, we obtain $\mathcal{N}_\phi \in H^1(0,T)$, the element normal to $\mathcal{T}\mathcal{M}_\phi$ with respect to the H^1 topology, as the solution of the elliptic boundary-value problem (22) with the source term on the RHS in (22a) replaced with expression (27). For a given $\phi \in \mathcal{M}$, the projection operator then takes the form

$$\forall\, \Phi \in H^1(0,T) \qquad P_{\mathcal{T}\mathcal{M}_\phi}\Phi := \Phi - \zeta\, \mathcal{N}_\phi, \quad \text{where } \zeta := \frac{\left\langle \Phi, \mathcal{N}_\phi \right\rangle_{H^1(0,T)}}{\left\langle \mathcal{N}_\phi, \mathcal{N}_\phi \right\rangle_{H^1(0,T)}} \tag{28}$$

can be interpreted as the Lagrange multiplier associated with the condition $\left\langle P_{\mathcal{T}\mathcal{M}_\phi}\Phi, \mathcal{N}_\phi \right\rangle_{H^1(0,T)} = 0$. We emphasize that each iteration of the gradient descent (1) requires the solution of two adjoint systems: the solution of system (19) allows us to determine the gradient of the objective functional, whereas the solution of (25) is needed in order to construct the projection operator (11b). The two adjoint systems are defined in terms of the same operator $\mathcal{K}^*$, but are subject to different boundary conditions and source terms.

As already indicated in Section 1, for general nonlinear manifolds $\mathcal{M}$ the projected discrete gradient flow (1) produces local minimizers which do not satisfy the constraint exactly, but only with an error of order $\mathcal{O}(\tau_n^2)$ at each iteration. If the constraint is homogeneous as is the case in Problem 1, then one can define the retraction operator $\mathcal{R}_\mathcal{M} : \mathcal{T}\mathcal{M} \to \mathcal{M}$ in terms of simple normalization

$$\mathcal{R}_\mathcal{M}(\phi) := \sqrt{\frac{E_0}{[E(\cdot;\phi)]_T}}\, \phi. \tag{29}$$

The resulting Riemannian gradient flow then satisfies the constraint exactly (i.e., up to round-off errors only) [10]. The steps required to solve Problem 1.B and Problem 2.B with the approach described above are summarized as Algorithm 1. As is evident from this algorithm, the additional per-iteration computational cost resulting from approximate enforcement of the constraint $[E(\cdot;\phi)]_T = E_0$ consists in the solution of the adjoint system (25) and the boundary-value problem (22), lines 9–12.

3.3. Solution of Problem 3

Here we provide some details describing how the framework introduced in Sections 3.1 and 3.2 can be used to solve Problem 3. Since the system considered in this problem is more complicated, to balance completeness and brevity, we only state the most important steps and refer the reader to [11] for further details.

First, we compute the gradient of the objective functional (12) and following the same procedure as stated in Section 3.1 begin by determining the Gâteaux differential of (12)

$$\mathcal{J}'(\varphi;\varphi') := \frac{d}{d\epsilon}\, \mathcal{J}\left(\varphi + \epsilon\varphi'\right)\Big|_{\epsilon=0} = \frac{1}{D}\int_0^T \int_\Omega \left(\mathcal{P}(t) - \widetilde{\mathcal{P}}(t;\varphi)\right) \mathbf{\Delta}\widetilde{\omega}(t,\mathbf{x};\varphi)\, \widetilde{\omega}'(t,\mathbf{x};\varphi,\varphi')\, d\mathbf{x}\, dt, \tag{30}$$

Algorithm 1 Solution of Problem 1.B and Problem 2.B, cf. Fig. 1.

Input:
- $\bar{u}_b(t)$ — prescribed profile
- E_0 — value of the constraint
- $\Delta x, \Delta t$ — numerical discretization parameters
- ℓ — Sobolev length scale
- $\epsilon_{\mathcal{J}}$ — tolerance in the termination criterion
- $\phi_0(t)$ — initial guess

Output:
- $\breve{\phi}(t)$ — optimal flux

1: • set $n = 0$
2: • set $\phi^{(0)} = \phi_0$
3: **repeat**
4: • set $n = n + 1$
5: • solve the governing system (2) with the boundary condition (2b) given by $\phi^{(n)}$
6: • solve the adjoint system (19)
7: • determine the L^2 gradient $\nabla_\phi^{L^2}\mathcal{J}(\phi^{(n)})$, cf. (20)
8: • determine the Sobolev gradient $\nabla_\phi\mathcal{J}(\phi^{(n)})$, cf. (22)
9: • solve the adjoint system (25)
10: • determine the normal element in L^2 $\mathcal{N}_\phi^{L^2}$, cf. (27)
11: • determine the normal element in the Sobolev space H^1 $\mathcal{N}_\phi$, cf. (22)
12: • determine the projection of the Sobolev gradient onto the tangent subspace $\mathcal{TM}_{\phi^{(n)}}$, (28)
13: • determine the optimal step length $\tau^{(n)}$ by solving a line-minimization problem while applying retraction (29) (Problem 1.B only)
14: • update the flux $\phi^{(n)} \leftarrow \mathcal{R}_{\mathcal{M}}\left(\phi^{(n)} - \tau_n P_{\mathcal{TM}_{\phi^{(n)}}}\phi^{(n)}\right)$
15: **until** $\left(\mathcal{J}(\phi^{(n-1)}) - \mathcal{J}(\phi^{(n)})\right)/\mathcal{J}(\phi^{(n-1)}) < \epsilon_{\mathcal{J}}$

where φ' is an arbitrary perturbation of $\varphi \in \mathcal{S}$, and $\widetilde{\omega}'(t,\mathbf{x};\varphi,\varphi')$ satisfies the corresponding perturbation system obtained by linearizing the governing system (8), namely,

$$\mathcal{K}\begin{bmatrix}\widetilde{\omega}'\\ \widetilde{\psi}'\end{bmatrix} := \begin{bmatrix} \partial_t\widetilde{\omega}' + \boldsymbol{\nabla}^\perp\widetilde{\psi}'\cdot\boldsymbol{\nabla}\widetilde{\omega} + \boldsymbol{\nabla}^\perp\widetilde{\psi}\cdot\boldsymbol{\nabla}\widetilde{\omega}' + \alpha\widetilde{\omega}' \\ -\boldsymbol{\nabla}\cdot\left(2(\boldsymbol{\nabla}\widetilde{\omega}\cdot\boldsymbol{\nabla}\widetilde{\omega}')(\frac{dv}{ds}\,\varphi\boldsymbol{\nabla}\widetilde{\omega} + \frac{\eta^3\sqrt{s}+v_0}{s_{\max}}\,\frac{d\varphi}{d\sigma}\,\boldsymbol{\nabla}\widetilde{\omega}) + (v_N+v)\boldsymbol{\nabla}\widetilde{\omega}'\right) \\ \boldsymbol{\Delta}\widetilde{\psi}' + \widetilde{\omega}' \end{bmatrix} = \begin{bmatrix} \boldsymbol{\nabla}\cdot\left((\eta^3\sqrt{s}+v_0)\varphi'\boldsymbol{\nabla}\widetilde{\omega}\right) \\ 0 \end{bmatrix}, \tag{31a}$$

$$\widetilde{\omega}'(t=0,\mathbf{x}) = 0, \tag{31b}$$

where for ease of notation, we have denoted $\sigma := s/s_{\max}$. Defining the adjoint system as

$$\mathcal{K}^*\begin{bmatrix}\widetilde{\omega}^*\\ \widetilde{\psi}^*\end{bmatrix} := \begin{bmatrix} -\partial_t\widetilde{\omega}^* - \boldsymbol{\nabla}^\perp\widetilde{\psi}\cdot\boldsymbol{\nabla}\widetilde{\omega}^* + \alpha\widetilde{\omega}^* + \widetilde{\psi}^* \\ -\boldsymbol{\nabla}\cdot\left(2(\boldsymbol{\nabla}\widetilde{\omega}\cdot\boldsymbol{\nabla}\widetilde{\omega}^*)(\frac{dv}{ds}\,\varphi\boldsymbol{\nabla}\widetilde{\omega} + \frac{\eta^3\sqrt{s}+v_0}{s_{\max}}\,\frac{d\varphi}{d\sigma}\,\boldsymbol{\nabla}\widetilde{\omega}) + (v_N+v)\boldsymbol{\nabla}\widetilde{\omega}^*\right) \\ \boldsymbol{\Delta}\widetilde{\psi}^* - \boldsymbol{\nabla}^\perp\cdot(\widetilde{\omega}^*\boldsymbol{\nabla}\widetilde{\omega}) \end{bmatrix} = \begin{bmatrix} W \\ 0 \end{bmatrix}, \tag{32a}$$

$$\widetilde{\omega}^*(t=T,\mathbf{x}) = 0, \tag{32b}$$

with source term $W(t,\mathbf{x}) := \frac{1}{D}\left(\mathcal{P}(t) - \widetilde{\mathcal{P}}(t;\varphi)\right)\boldsymbol{\Delta}\widetilde{\omega}(t,\mathbf{x};\varphi)$, we obtain the duality-pairing relation

$$\begin{aligned}\left(\mathcal{K}\begin{bmatrix}\widetilde{\omega}'\\ \widetilde{\psi}'\end{bmatrix}, \begin{bmatrix}\widetilde{\omega}^*\\ \psi^*\end{bmatrix}\right) &:= \int_0^T\int_\Omega \mathcal{K}\begin{bmatrix}\widetilde{\omega}'\\ \widetilde{\psi}'\end{bmatrix}\cdot\begin{bmatrix}\widetilde{\omega}^*\\ \widetilde{\psi}^*\end{bmatrix} d\mathbf{x}\,dt \\ &= \int_0^T\int_\Omega \begin{bmatrix}\widetilde{\omega}'\\ \widetilde{\psi}'\end{bmatrix}\cdot\mathcal{K}^*\begin{bmatrix}\widetilde{\omega}^*\\ \widetilde{\psi}^*\end{bmatrix} d\mathbf{x}\,dt = \left(\begin{bmatrix}\widetilde{\omega}'\\ \widetilde{\psi}'\end{bmatrix}, \mathcal{K}^*\begin{bmatrix}\widetilde{\omega}^*\\ \psi^*\end{bmatrix}\right),\end{aligned} \tag{33}$$

where integration by parts was performed with respect to both space and time and all boundary terms vanish due to periodic boundary conditions. This relation together with the adjoint system (32) allows us to re-express the Gâteaux differential (30) as

$$\begin{aligned}\mathcal{J}'(\varphi;\varphi') &= \int_0^T\int_\Omega W(t,\mathbf{x})\,\widetilde{\omega}'(t,\mathbf{x};\varphi,\varphi')\,d\mathbf{x}\,dt, \\ &= -\int_0^T\int_\Omega \left(\eta^3\sqrt{s}+v_0\right)\left(\boldsymbol{\nabla}\widetilde{\omega}\cdot\boldsymbol{\nabla}\widetilde{\omega}^*\right)\varphi'\,d\mathbf{x}\,dt,\end{aligned}$$

$$= \int_0^1 \left[-\int_0^T \int_\Omega \delta\left(\frac{\nabla\widetilde{\omega}\cdot\nabla\widetilde{\omega}}{s_{\max}} - \sigma \right) \left(\eta^3 \sqrt{s} + \nu_0 \right) \nabla\widetilde{\omega}\cdot\nabla\widetilde{\omega}^* \, d\mathbf{x}\, dt \right] \varphi'(\sigma)\, d\sigma, \tag{34}$$

where the substitution $\varphi'(\nabla\widetilde{\omega}\cdot\nabla\widetilde{\omega}) = \int_0^1 \delta\left(\frac{\nabla\widetilde{\omega}\cdot\nabla\widetilde{\omega}}{s_{\max}} - \sigma \right) \varphi'(\sigma)\, d\sigma$ with $\delta(\cdot)$ denoting the Dirac delta distribution was made and Fubini's theorem was used to swap the order of integration. These last two steps are needed in order to change the integrations variables in (34) from $d\mathbf{x}\, dt$ to ds, such that the Gâteaux differential (30) can be expressed in the required Riesz form

$$\mathcal{J}'(\varphi;\varphi') = \left\langle \nabla_\varphi \mathcal{J}, \varphi' \right\rangle_{H^2([0,1])} = \left\langle \nabla_\varphi^{L^2} \mathcal{J}, \varphi' \right\rangle_{L^2([0,1])}. \tag{35}$$

This then allows us to extract the gradient with respect to the L^2 topology as

$$\nabla_\varphi^{L^2} \mathcal{J}(\sigma) = -\int_0^T \int_\Omega \delta\left(\frac{\nabla\widetilde{\omega}\cdot\nabla\widetilde{\omega}}{s_{\max}} - \sigma \right) \left(\eta^3 \sqrt{s} + \nu_0 \right) \nabla\widetilde{\omega}\cdot\nabla\widetilde{\omega}^* \, d\mathbf{x}\, dt, \qquad \sigma \in [0,1]. \tag{36}$$

As was done in Section 3.1, we use the Riesz representation in the Sobolev space $H^2(0,T)$, cf. (24), and the definition of the subspace $\mathcal{S}$ in (10) which after integration by parts with respect to σ, cf. (21), give the Sobolev gradient $\nabla_\varphi \mathcal{J}$ as the solution of the following elliptic boundary-value problem

$$\left[\mathrm{Id} - \ell_1^2 \frac{d^2}{d\sigma^2} + \ell_2^4 \frac{d^4}{d\sigma^4} \right] \nabla_\varphi \mathcal{J}(\sigma) = \nabla_\varphi^{L^2} \mathcal{J}(\sigma), \qquad \sigma \in [0,1], \tag{37a}$$

$$\left. \frac{d^{(1)}(\nabla_\varphi \mathcal{J})}{d\sigma^{(1)}} \right|_{\sigma=0,1} = \left. \frac{d^{(3)}(\nabla_\varphi \mathcal{J})}{d\sigma^{(3)}} \right|_{\sigma=0,1} = 0. \tag{37b}$$

We add that the boundary conditions in (37b), see also (10), were selected in order to ensure the vanishing of the terms at $\sigma = 0, 1$ which result from integration by parts.

Similarly, we determine the normal element $\mathcal{N}_\varphi$ by considering the Gâteaux differential of the constraint (13) with respect to φ and invoking the Riesz representation theorem, i.e.,

$$\begin{aligned} \left[\widetilde{\mathcal{E}}'(\cdot;\varphi,\varphi') \right]_T := \frac{d}{d\epsilon} \left[\widetilde{\mathcal{E}}(\cdot;\varphi+\epsilon\varphi') \right]_T \Big|_{\epsilon=0} &= \frac{1}{T} \int_0^T \int_\Omega \widetilde{\omega}(t,\mathbf{x};\varphi)\, \widetilde{\omega}'(t,\mathbf{x};\varphi,\varphi')\, d\mathbf{x}\, dt, \\ &= \left\langle \mathcal{N}_\varphi, \varphi' \right\rangle_{H^2([0,1])} = \left\langle \mathcal{N}_\varphi^{L^2} \mathcal{J}, \varphi' \right\rangle_{L^2([0,1])} = 0. \end{aligned} \tag{38}$$

Introducing the new *adjoint fields* $\breve{\omega}^*$ and $\breve{\psi}^*$, which satisfy the adjoint system (32), with the source term $W(t,\mathbf{x}) := \frac{1}{T}\widetilde{\omega}(t,\mathbf{x};\varphi)$, we can use the duality pairing (33) to conclude that

$$\begin{aligned} \left[\widetilde{\mathcal{E}}'(\cdot;\varphi,\varphi') \right]_T &= -\int_0^T \int_\Omega \left(\eta^3 \sqrt{s} + \nu_0 \right) \left(\nabla\widetilde{\omega}\cdot\nabla\breve{\omega}^* \right) \varphi'\, d\mathbf{x}\, dt, \\ &= \int_0^1 \left[-\int_0^T \int_\Omega \delta\left(\frac{\nabla\widetilde{\omega}\cdot\nabla\widetilde{\omega}}{s_{\max}} - \sigma \right) \left(\eta^3 \sqrt{s} + \nu_0 \right) \nabla\widetilde{\omega}\cdot\nabla\breve{\omega}^* \, d\mathbf{x}\, dt \right] \varphi'(\sigma)\, d\sigma. \end{aligned}$$

Then, using the second equality in (38), we obtain an expression for the element normal to the subspace $\mathcal{T}\mathcal{M}_\varphi$ with respect to the L^2 topology as

$$\mathcal{N}_\varphi^{L^2}(\sigma) = -\int_0^T \int_\Omega \delta\left(\frac{\nabla\widetilde{\omega}\cdot\nabla\widetilde{\omega}}{s_{\max}} - \sigma \right) \left(\eta^3 \sqrt{s} + \nu_0 \right) \nabla\widetilde{\omega}\cdot\nabla\breve{\omega}^* \, d\mathbf{x}\, dt, \qquad \sigma \in [0,1]. \tag{39}$$

Finally, employing the last equality in (38) and performing analogous steps as described above, we obtain the normal element $\mathcal{N}_\varphi \in H^2(0,1)$ as a solution of the boundary-value problem (37), but with the source term in (37a) replaced with the RHS of (39). This then allows us to construct the projection operator $P_{\mathcal{T}\mathcal{M}_\varphi}$ in the form (28). The steps required to solve Problem 3 are analogous to Algorithm 1, with obvious modifications.

4. Numerical discretization and validation

In this section we briefly describe the numerical discretizations employed to solve Problems 1, 2 and 3 using the discrete gradient flow (1). Then we validate key elements of these computations requiring solution of adjoint systems, namely, evaluation of the cost functional gradients (20) and (36) and the normal elements (27) and (39). The proposed approach has been implemented in MATLAB and the code is available in [29] (it can be used to generate the figures shown in Sections 4.1 and 5.1).

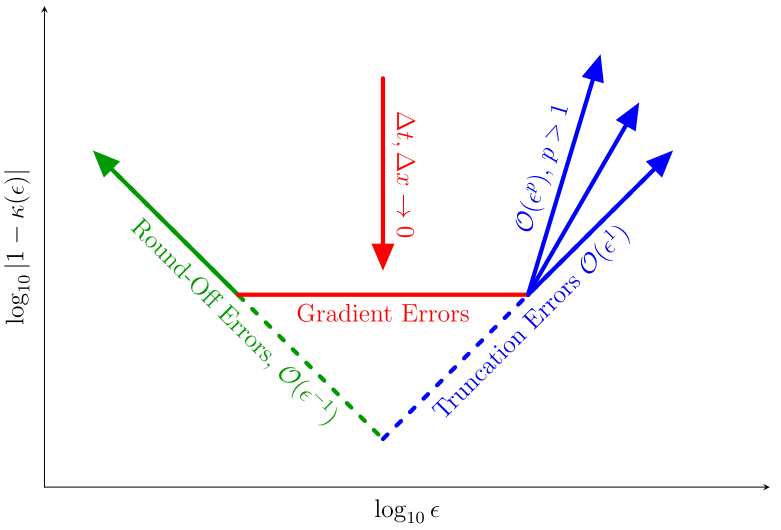

Fig. 2. Schematic of the interplay between the three types of errors in the validation of accuracy of the gradients and of the normal elements based on formulas (40) and (43), cf. Section 4.1. The slope of the right branch (blue) depends on the order of the finite-difference formula used to approximate the Gâteaux differential in the denominator ($p=1$ in (40) and (43)). (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

4.1. Problems 1 and 2

The governing system (2) and the adjoint systems (19) and (25) are discretized in space with a standard second-order finite-difference scheme and integrated in time using a second-order Crank-Nicolson scheme.

In order to validate the computation of the gradient (20) and of the normal element (27) we define the following quantities

$$\kappa_1(\epsilon) := \frac{\epsilon^{-1}\left[\mathcal{J}(\phi+\epsilon\phi') - \mathcal{J}(\phi)\right]}{\left\langle \nabla_\phi^{L^2}\mathcal{J}(\phi), \phi' \right\rangle_{L^2(0,T)}}, \qquad \epsilon > 0 \tag{40a}$$

$$\kappa_2(\epsilon) := \frac{\epsilon^{-1}\left[E(\cdot;\phi+\epsilon\phi') - E(\cdot;\phi)\right]_T}{\left\langle \mathcal{N}_\phi^{L^2}, \phi' \right\rangle_{L^2(0,T)}}, \tag{40b}$$

in which the numerators represent first-order finite-difference approximations of the Gâteaux differentials (15) and (23), whereas the denominators are the corresponding Riesz forms (17) and (24). Clearly, we expect that $\kappa_i(\epsilon) \approx 1$, $i = 1, 2$, and any deviations of these quantities from unity results from a combination of errors of the following three distinct types:

- errors in the numerical solution of the PDE systems (2), (19) and (25), and in the evaluation of the integrals in (5) and (6); these errors result from the discretization of these equations and expressions in space and in time; and are therefore controlled by the corresponding discretization parameters (Δx and Δt, but in general there may also be some other numerical parameters) and are independent of ϵ; hereafter, we refer to these errors as "gradient errors",
- truncation errors in the finite-difference formula in (40a)–(40b) which are proportional to ϵ (if a finite-difference formula of order $p > 1$ were used in (40a)–(40b), then the truncation errors would be proportional to ϵ^p), and
- subtractive cancellation (round-off) errors proportional to ϵ^{-1}.

An interplay of these types of errors in formulas such as (40a)–(40b) is shown schematically in Fig. 2. Evidently, both the truncation and subtractive cancellation (round-off) errors are artefacts of the structure of these formulas and accuracy of the gradient $\nabla_\phi^{L^2}\mathcal{J}$ and of the normal element $\mathcal{N}_\phi^{L^2}$ is controlled solely by the gradient errors. Therefore, in order to validate their computation, one must show that these errors vanish as the numerical parameters are refined, i.e., as $\Delta x, \Delta t \to 0$, which is manifested by the lowering of the plateau labelled "Gradient Errors" in Fig. 2 until it ultimately disappears. We add that the order of the finite-difference formula used to approximate the Gâteaux differentials in the denominators of (40a)–(40b) does not affect the accuracy with which the gradient $\nabla_\phi^{L^2}\mathcal{J}$ and the normal element $\mathcal{N}_\phi^{L^2}$ are computed (it only affects the slope of the branch corresponding to large ϵ marked in blue in Fig. 2).

To fix attention, we focus here on the more general Problem 2 and add that very similar results were also obtained for Problem 1. In our tests we set $a = 0$, $b = 1$, $T = 1$ and

$$u_0(x) = 10\cos(\pi x), \quad x \in [0,1], \qquad \bar{\phi}(t) = e^{-4t}\left[18 - 1000\cos\left(\frac{15\pi}{2}t\right)\right], \quad t \in [0,T], \tag{41}$$

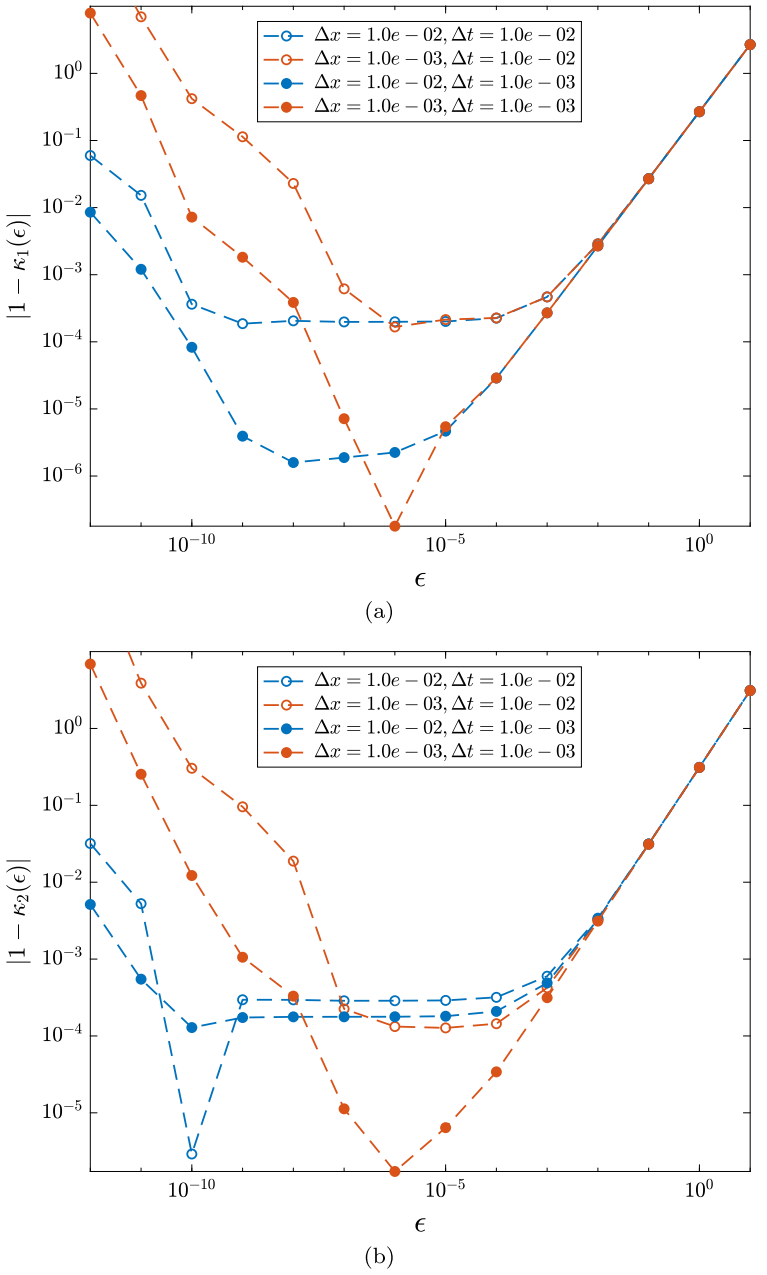

Fig. 3. [Problem 2] Dependence of (a) $|1-\kappa_1(\epsilon)|$ and (b) $|1-\kappa_2(\epsilon)|$ on ϵ for different spatial and temporal discretizations Δx and Δt.

where $\bar{\phi}(t)$ is the "true" flux defining, via the solution of system (2), the target profile $\bar{u}_b(t)$ appearing in the error functional (6). In the diagnostic quantities (40a)–(40b) we use

$$\phi_0 := \phi(t) = 18\,\sin\left(\frac{\pi}{2}\,t\right) e^{-4t}, \qquad \phi'(t) = 4t\,e^{(-4+\pi)t}, \qquad t \in [0,T], \tag{42}$$

representing, respectively, the "point" (in the function space $H^1(0,T)$) where the Gâteaux differentials are computed and the direction.

In Figs. 3a and 3b we show the dependence of the quantities $|1-\kappa_i(\epsilon)|$, $i=1,2$, on ϵ for different indicated values of the space and time discretization parameters Δx and Δt, which reveals the expected behavior, cf. Fig. 2. In particular, we observe that $\kappa_1(\epsilon)$ and $\kappa_2(\epsilon)$ deviate from the unity for very small and very large values of ϵ which is due to, respectively, the subtractive cancellation (round-off) errors and the truncation errors in the finite-difference formula in (40a)–(40b). However, we also observe that for intermediate values of ϵ spanning several orders of magnitude both $\kappa_1(\epsilon)$ and $\kappa_2(\epsilon)$ exhibit plateaus corresponding to gradient errors of order $\mathcal{O}\left(C_{\Delta x}(\Delta x)^2 + C_{\Delta t}(\Delta t)^2\right)$ for some $C_{\Delta x}, C_{\Delta t} > 0$ (the structure of this error reflects the order of accuracy of the discretization techniques used to approximate relation (2), (5), (6), (19) and (25)). Inspection of the results in Fig. 3a shows that

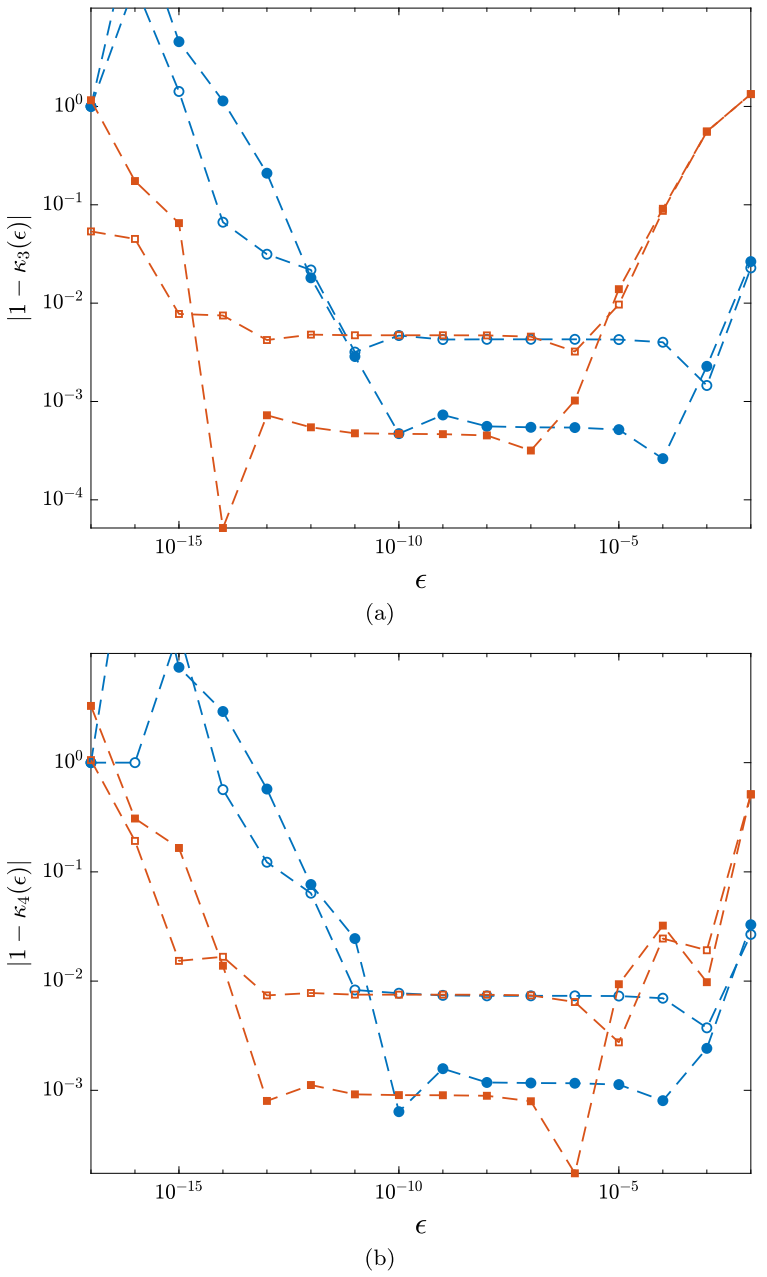

Fig. 4. [Problem 3:] Dependence of (a) $|1-\kappa_3(\epsilon)|$ and (b) $|1-\kappa_4(\epsilon)|$ on ϵ for different temporal discretizations with $\Delta t = 10^{-2}$ (empty symbols) and $\Delta t = 10^{-3}$ (filled symbols). The results are shown for the perturbations φ' corresponding to ν' given in (45a) (blue circles) and (45b) (red squares).

the gradient errors are reduced resulting in the lowering of the plateaus when Δt is refined, indicating that the accuracy of gradient evaluations is controlled by the time discretization (i.e., the gradient error is dominated by the term proportional to $(\Delta t)^2$ and errors resulting from the discretization in space are relatively unimportant). On the other hand, the results in Fig. 3b show that the accuracy with which the normal element is evaluated is more or less equally controlled by both Δx and Δt. In both cases, the plateaus ultimately disappear when the resolution is sufficiently refined indicating that the gradient errors become smaller than the subtractive cancellation (round-off) and truncation errors. This demonstrates the convergence of the approximations of the cost functional gradient and the normal element. Since the Sobolev gradients and normal directions are obtained by solving the elliptic boundary value problem (22) and can be therefore viewed as low-pass filtered versions of the corresponding elements found in L^2 [24], their correctness follows from the validation provided above together with relations (17) and (24). In the computations presented in Section 5.1 the Sobolev gradients and normal directions are computed using the inner product (4) with $\ell_1 = 0.01$.

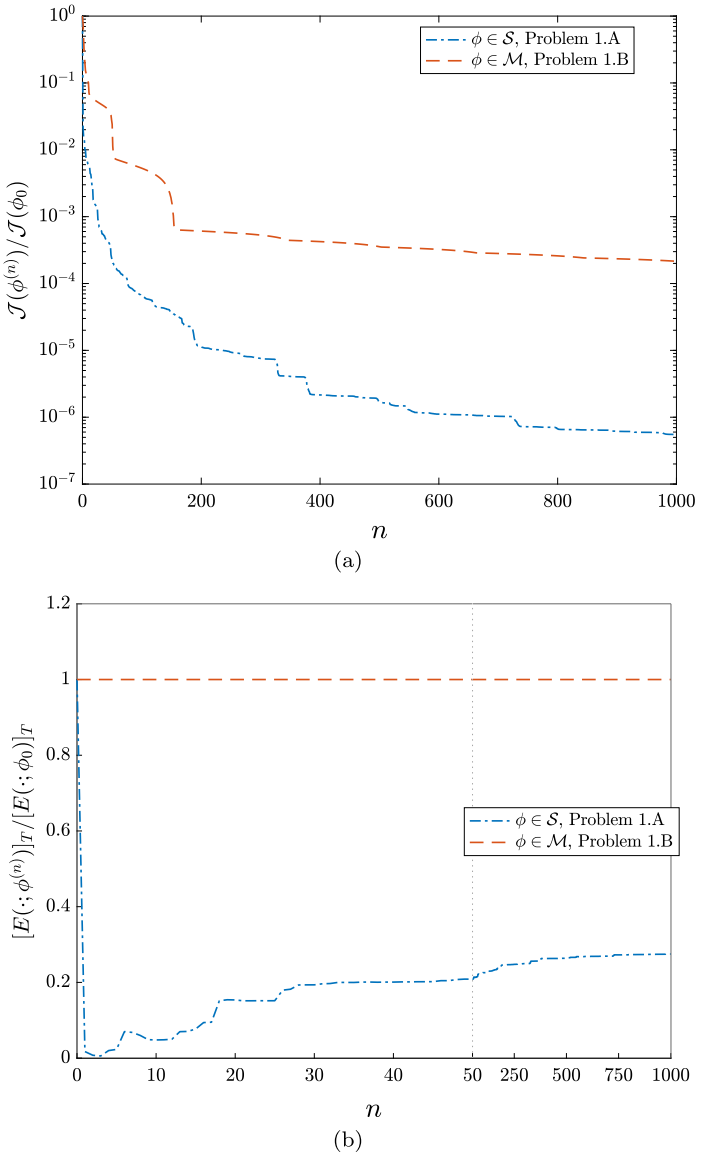

Fig. 5. [Problem 1:] Dependence of (a) the normalized error functional $\mathcal{J}(\phi^{(n)})/\mathcal{J}(\phi_0)$ and (b) normalized constraint $\left[E\left(\cdot;\phi^{(n)}\right)\right]_T / \left[E\left(\cdot;\phi_0\right)\right]_T$ on the iteration n in the solutions of the unconstrained and constrained optimization problem. For clarity, in (b), the horizontal axis is split into two regions with different linear scaling separated by the vertical dotted line.

4.2. Problem 3

The Navier-Stokes system (7), the corresponding LES system (8) and the adjoint systems (32) are approximated using a standard Fourier pseudo-spectral method in space and using a third-order IMEX scheme introduced in [30] in time. Evaluation of the closure terms involving the state-dependent eddy viscosity (9) requires interpolation from the physical space Ω to the state space $\mathcal{I}$ and differentiation with respect to the state variable s, steps which are performed using methods based on Chebyshev polynomials [31]. We refer the reader to [11] for further numerical details.

The "target" Navier-Stokes flow w is defined as the solution of system (7) obtained with the parameters $\nu_N = 4\times10^{-4}$, $\alpha = 5\times10^{-3}$, $F = 2$, $k_a = k_b = 4$ over the time window $T = 50 \approx 27.8\, t_e$, where $t_e := \left[\int_0^T \mathcal{E}(t)\, dt/(8\pi^2 T)\right]^{-1/2}$ is the eddy turnover time [32]. This time window is chosen to make Problem 3 physically interesting, but at the same time is not too long as to make the computation of gradients problematic due to exponential divergence of nearby trajectories (it is worth mentioning here that reliable computation of sensitivities of long-time averages of quantities defined for chaotic systems is facilitated by the "shadowing approach" [33,34]). The direct numerical simulation (DNS) of (7) and the LES computation solving (8) use, respectively, $N_x = 256$ and $N_x = 64$ equispaced grid points in each spatial direction. They also use $N_s = 128$ Chebyshev points to discretize the state space $\mathcal{I}$. The LES system (8) is

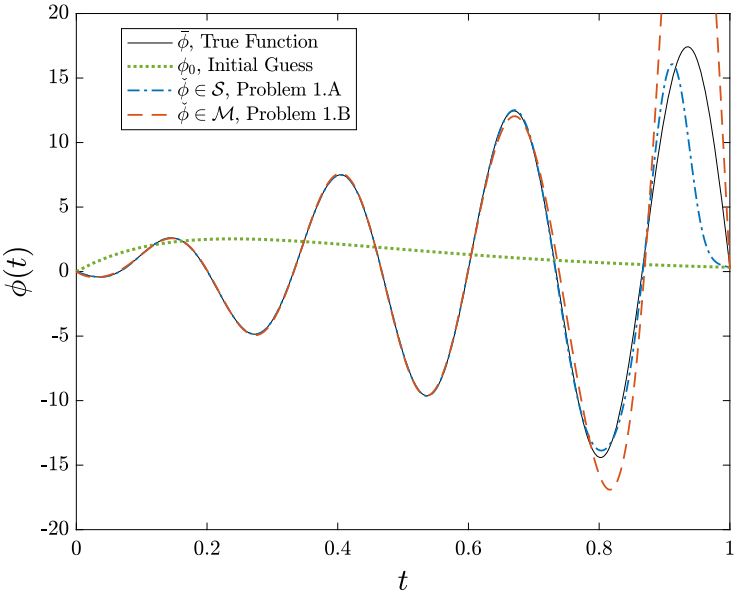

Fig. 6. [Problem 1:] (black, solid line) the true flux $\bar{\phi}(t)$, cf. (41), (green, dotted line) the initial guess $\phi_0(t)$ in (1) and (42), (blue, dot-dashed line) the optimal flux $\breve{\phi}(t)$ in the unconstrained problem, and (red, dot-solid line) the optimal flux $\breve{\phi}(t)$ in the constrained problem, all as functions of time $t \in [0, T]$.

obtained using a "box" filter with the cutoff $k_c = 8$ [11]. The use of this rather "aggressive" filter is dictated by the desire to define a problem where the presence of the constraint (13) has a significant effect on the solutions of Problem 3, which will in turn allow us to more easily elucidate the properties of the proposed approach.

In order to validate the computation of the cost functional gradient and the normal element, we define the following diagnostic quantities analogous to (40a)–(40b)

$$\kappa_3(\epsilon) := \frac{\epsilon^{-1}\left[\mathcal{J}(\varphi + \epsilon\varphi') - \mathcal{J}(\varphi)\right]}{\left\langle \nabla^{L^2}_{\varphi}\mathcal{J}(\varphi), \varphi' \right\rangle_{L^2(0,1)}}, \qquad \epsilon > 0 \tag{43a}$$

$$\kappa_4(\epsilon) := \frac{\epsilon^{-1}\left[\widetilde{\mathcal{E}}(\cdot;\varphi + \epsilon\varphi') - \widetilde{\mathcal{E}}(\cdot;\varphi)\right]_T}{\left\langle \mathcal{N}^{L^2}_{\varphi}, \varphi' \right\rangle_{L^2(0,1)}}, \tag{43b}$$

where $\varphi = \varphi(s)$ and $\varphi' = \varphi'(s)$ are the nondimensional functions corresponding, via ansatz (9), to the eddy viscosity in the Leith model [25–27]

$$\nu_L(s) = (C_L\, k_c)^3 \sqrt{s} \tag{44}$$

with the constant $C_L = 4.702 \times 10^{-3}$ chosen to ensure that $\varphi \in \mathcal{M}$, and to the perturbations

$$\nu'(s) = \nu_L(s), \qquad \text{with} \quad C_L = 4.2 \times 10^{-3}, \tag{45a}$$

$$\nu'(s) = \exp\left(\frac{-2s}{30}\right). \tag{45b}$$

The quantities $|1 - \kappa_3(\epsilon)|$ and $|1 - \kappa_4(\epsilon)|$ are shown as functions of ϵ for different perturbations φ' and two different time steps Δt in Figs. 4a and 4b, respectively. Since the spatial discretization (with respect to $\mathbf{x}$) and the discretization of the state space s we use are both spectrally accurate (i.e., they have exponentially small errors), we focus here on assessing the effect of the time discretization, which has an algebraic only rate of convergence, on the accuracy of the gradient and the normal element. In Figs. 4a and 4b we see the expected behavior of the quantities $\kappa_3(\epsilon)$ and $\kappa_4(\epsilon)$, cf. Fig. 2. In particular, they approach unity for intermediate values of ϵ as the time step Δt is refined indicating the vanishing of the gradient error which is dominated by the time discretization. However, in contrast to the results of analogous tests for Problem 2, cf. Fig. 3a,b, now the convergence is slower with the gradient error of order $\mathcal{O}(\Delta t)$, which is a consequence of the difficulties in evaluating the integrals involving the Dirac distributions in (36) and (39). We can nevertheless conclude that the approximations of the cost functional gradient and of the normal element are convergent. In the computations presented in Section 5.2 the Sobolev gradients and normal directions are computed using the inner product (4) with $\ell_1 = 1000$ and $\ell_2 = 100$.

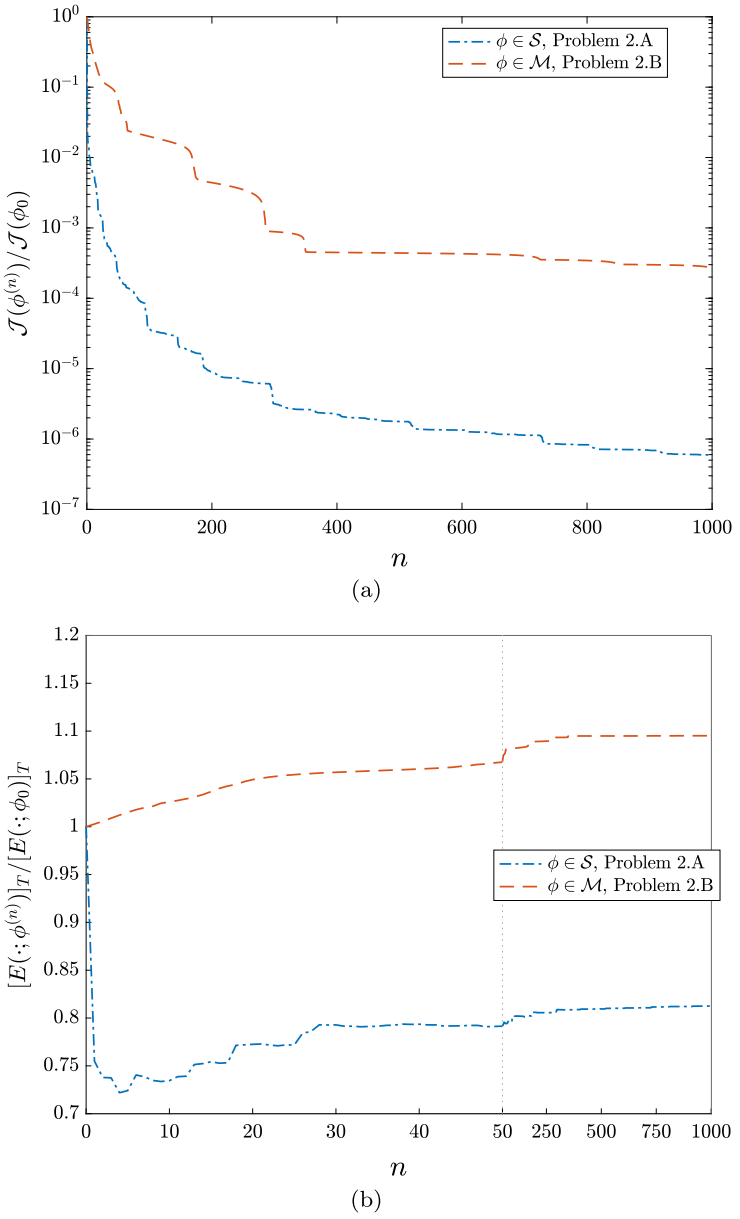

Fig. 7. [Problem 2:] Dependence of (a) the normalized error functional $\mathcal{J}(\phi^{(n)})/\mathcal{J}(\phi_0)$ and (b) normalized constraint $\left[E\left(\cdot;\phi^{(n)}\right)\right]_T / \left[E\left(\cdot;\phi_0\right)\right]_T$ on the iteration n in the solutions of the unconstrained and constrained optimization problem. For clarity, in (b), the horizontal axis is split into two regions with different linear scaling separated by the vertical dotted line.

5. Results

In this section we study solutions to Problems 1, 2 and 3 focusing our attention on how well and how efficiently the constraints involved are satisfied. In iterations (1) the step size τ_n is determined using Brent's algorithm to solve the line-minimization problem [35]. In the unconstrained versions of the problems we use the Polak-Ribière version of the conjugate gradient method [36] to accelerate convergence of iterations. In the constrained versions of the problems when no retraction operator is available, we limit the magnitude of the step size τ_n to reduce drift away from the constraint manifold $\mathcal{M}$.

5.1. Problems 1 and 2

In Problems 1 and 2 we attempt to reconstruct the flux $\phi(t)$ at the left boundary of the domain by minimizing the error functional (6) such that the temperature at the right boundary matches the prescribed target temperature $\bar{u}_b$ over the time window $[0,T]$; the latter temperature is obtained by solving problem (2) with a certain "true" flux $\bar{\phi}(t)$, given in (41), applied at the left boundary.

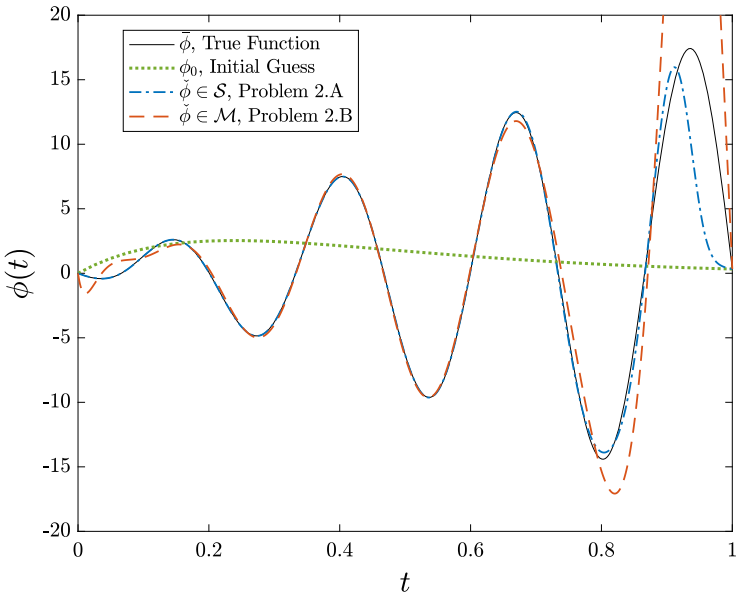

Fig. 8. [Problem 2:] (black, solid line) the true flux $\bar{\phi}(t)$, cf. (41), (green, dotted line) the initial guess $\phi_0(t)$ in (1) and (42), (blue, dot-dashed line) the optimal flux $\check{\phi}(t)$ in the unconstrained problem, and (red, dot-solid line) the optimal flux $\check{\phi}(t)$ in the constrained problem, all as functions of time $t \in [0,T]$.

The true flux $\bar{\phi}(t)$ thus represents the "exact" solution of Problems 1.A and 2.A in the sense that $\mathcal{J}(\bar{\phi}) = 0$ in these cases. However, since it does not satisfy the nonlinear constraint, $\bar{\phi} \notin \mathcal{M}$, the true flux cannot be a solution of Problems 1.B and 2.B. In the solution of Problem 1.B the nonlinear constraint, cf. (5), is enforced exactly using the retraction operator $\mathcal{R}_{\mathcal{M}}$ whereas in the solution of Problem 2.B this has to be done approximately using the orthogonal projection on the tangent subspace (28).

The dependence of the normalized error functional $\mathcal{J}(\phi^{(n)})/\mathcal{J}(\phi_0)$ and of the normalized constraint $\left[E(\cdot;\phi^{(n)})\right]_T / E_0$ on the iteration count n in the solution of Problems 1.A and 1.B is shown, respectively, in Figs. 5(a) and 5(b). The corresponding optimal fluxes $\check{\phi}(t)$ are shown as functions of time $t \in [0,T]$ in Fig. 6. Fig. 5(a) shows that, as expected, in the solution of Problem 1.A the error functional (6) is reduced to a very low level and this error is significantly higher in the solution of Problem 1.B where an exact solution is not expected to exist. On the other hand, the nonlinear constraint in (5) is satisfied exactly in Problem 1.B, cf. Fig. 5(b). In contrast, in the unconstrained problem we see that the difference $[E(\cdot;\phi)]_T - E_0$ becomes large, on the order of E_0, after a few iterations, which is not be surprising as this constraint is not imposed in Problem 1.A. As is evident from Fig. 6, the true flux is reconstructed rather well in Problem 1.A, except for the final times where a large oscillation is present. The oscillatory character of the optimal flux $\check{\phi}(t)$ in this case indicates that in its original form Problem 1 is ill-posed. Methods for dealing with this issue are well developed and are usually based on Tikhonov regularization [37] or formulation of the problem on spaces of sufficiently regular functions [24]. Since this issue is only tangential to the main topic of the present study, we do not consider it further here.

Moving on to discuss Problem 2, the dependence of the normalized error functional $\mathcal{J}(\phi^{(n)})/\mathcal{J}(\phi_0)$ and of the normalized constraint $\left[E(\cdot;\phi^{(n)})\right]_T / E_0$ on the iteration count n in shown in Figs. 7(a) and 7(b). We emphasize that now the retraction operator $\mathcal{R}_{\mathcal{M}}$ is no longer available and the nonlinear constraint in Problem 2.B can only be enforced approximately via projection on the tangent subspace $\mathcal{TM}_{\phi}$, cf. Fig. 1. We can see that, as a result, the constraint is no longer satisfied exactly, but the drift away from the constraint manifold $\mathcal{M}$ is slow, much less rapid than when the constraint is not enforced at all as in Problem 2.A. In this problem as well we observe a significant reduction of the error functional (6), cf. Fig. 7(a), leading to a good reconstruction of the true flux $\bar{\phi}$ in Problem 2.A, cf. Fig. 8.

5.2. Problem 3

Problem 3 is similar to Problem 2 in that the retraction operator is not available. Solutions to Problems 3.A and 3.B are obtained with iterations (1) where the eddy viscosity corresponding to the Leith model (44) is used as the initial guess. The dependence of the normalized error functional $\mathcal{J}(\varphi^{(n)})/\mathcal{J}(\varphi_0)$ and of the normalized constraint $\left[\widetilde{\mathcal{E}}(\cdot;\varphi^{(n)})\right]_T / \mathcal{E}_0$ on the iteration count n is shown, respectively, in Figs. 9a and 9b. In contrast to Problems 1 and 2, Problem 3 does not admit an "exact" solution and as a result the values of the objective functional (12) attained at the minima are not very small, especially in the constrained Problem 3.B. On the other hand, in analogy with the results for Problem 2, the minimizers found by solving the constrained Problem 3.B reveal a much slower drift away from the constraint manifold $\mathcal{M}$ than is evident in the solution of the unconstrained Problem 3.A. Comparison of the results obtained for Problems 3.A and 3.B shown in Figs. 9a and 9b indicates that in the particularly challenging (due to the constraint) latter problem a significant reduction of the objective functional can be achieved only if the iterates eventually deviate from the constraint manifold $\mathcal{M}$.

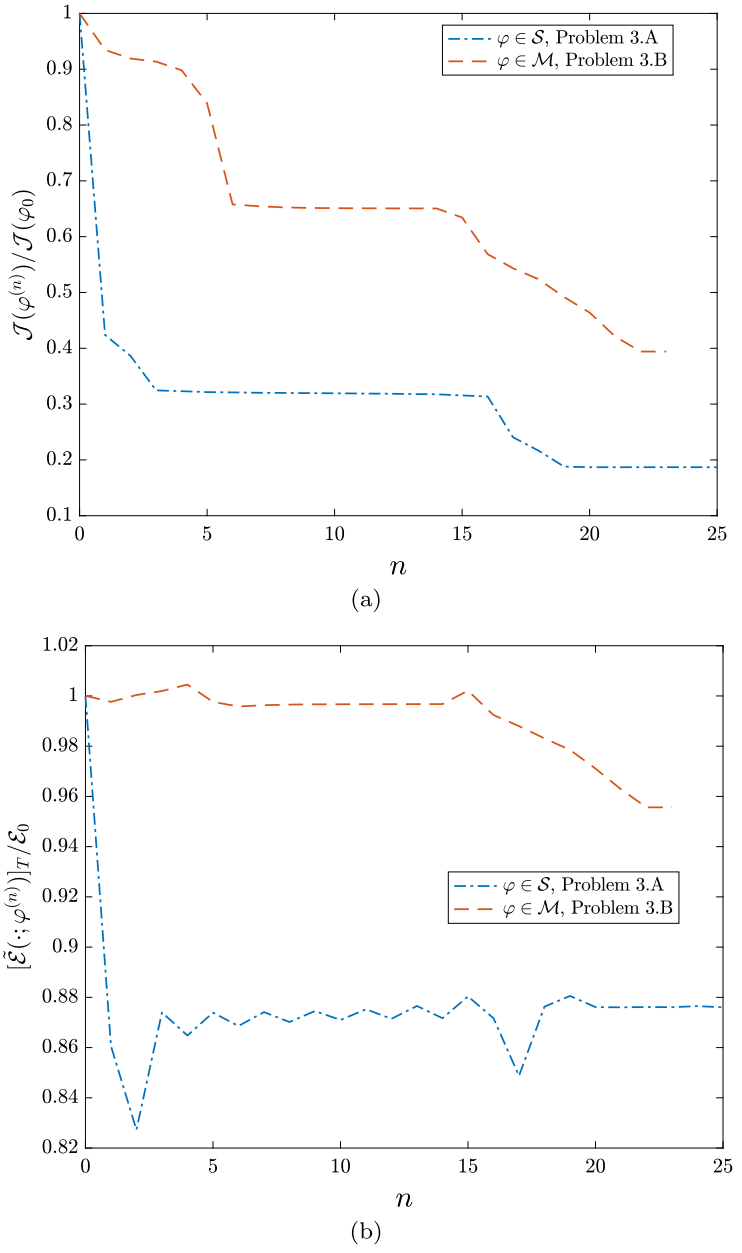

Fig. 9. [Problem 3:] Dependence of (a) the normalized error functional $\mathcal{J}(\varphi^{(n)})/\mathcal{J}(\varphi_0)$ and (b) the normalized constraint $\left[\widetilde{\mathcal{E}}(\cdot;\varphi^{(n)})\right]_T/\mathcal{E}_0$ on the iteration n for the solutions of the unconstrained and constrained optimization problem.

The optimal eddy viscosities corresponding, via ansatz (9), to the solutions of Problems 3.A and 3.B are shown as functions of $s \in \mathcal{I}$ in Fig. 10. Although the eddy viscosities obtained in the two cases have qualitatively similar profiles and possess common features such as a rapid increase for intermediate values of $\sqrt{s}$, they also reveal some fundamentally different physical properties. Most notably, the optimal eddy viscosity obtained in Problem 3.B is strictly dissipative. On the other hand, this quantity obtained in the problem in which the constraint (13) is not imposed is negative for small values of the state variable s, meaning that the closure model in fact injects energy in some regions of the flow. Such behavior has already been observed in closure models [38] and was discussed in detail in [11], where optimal closure models with such properties were obtained using different formulations. In Figs. 11a and 11b we show, respectively, the vorticity field $w(T,\mathbf{x})$ obtained by solving the Navier-Stokes system (7) and the corresponding filtered field $\widetilde{w}(T,\mathbf{x})$ which serves as the "target", cf. (12). As expected, the filtered field reveals fewer small-scale features. The solutions $\widetilde{\omega}(T,\mathbf{x})$ of the LES system (8) with the optimal eddy viscosities $\breve{\nu}(s)$ obtained by solving Problem 3.A and Problem 3.B are shown, respectively, in Figs. 11c and 11d. All fields are shown at the time instant $t = T$ corresponding to the end of the optimization window. Since in the formulation of Problem 3.A and Problem 3.B matching involves quantities defined globally,

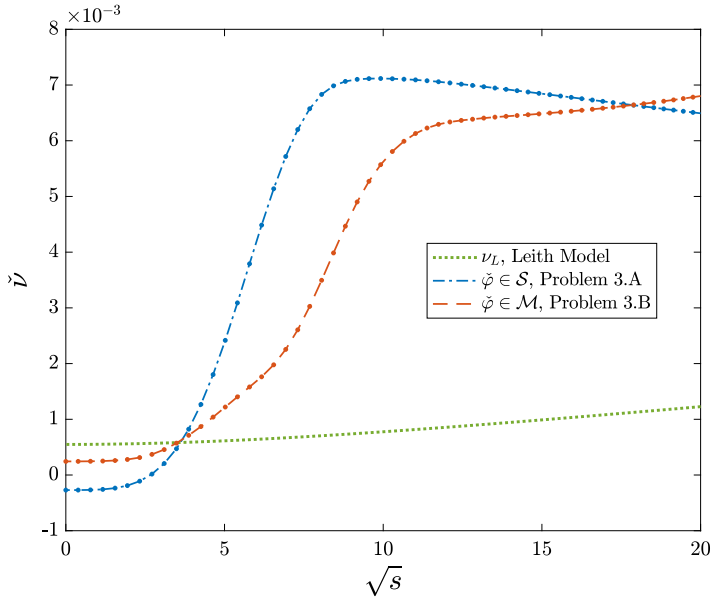

Fig. 10. [Problem 3:] Dependence of the optimal eddy viscosity $\check{\nu}$ on $\sqrt{s}$ for (blue, dashed-dot line) $\varphi \in S$ found by solving Problem 3.A and (red, dashed line) $\varphi \in \mathcal{M}$ found by solving Problem 3.B. The initial guess in (1) is given by the Leith model $\nu_L(s)$, cf. (44), and is shown as green dotted line.

rather than in a pointwise sense (as was done in [11]), the LES fields in Figs. 11c and 11d do not appear well correlated with the filtered target field in Fig. 11b. However, they are constructed to have similar statistical properties in terms of their enstrophy and palinstrophy, cf. (11a)–(11b).

To close this section, we show the enstrophy $\widetilde{\mathcal{E}}(t)$ and the palinstrophy $\widetilde{\mathcal{P}}(t)$ in the filtered DNS flow and in the LES flows with the optimal eddy viscosities obtained by solving Problems 3.A and 3.B, respectively, in Figs. 12a and 12b. These figures illustrate how the matching errors in these two quantities are distributed in time and are presented on a time window twice longer than the "training" window T considered in the error functional (12) and the constraint (13). We see that for $t \in [0, T]$ the behavior of the two quantities is consistent with what was observed in Figs. 9a and 9b. That is, the better reduction of the functional we see in Fig. 9a results in an on average better match of the time history of palinstrophy evident in Fig. 12b and, conversely, a worse match of the average enstrophy noted in Fig. 12a is consistent with deviation from the constraint manifold visible in Fig. 9b. However, for $t \in [T, 2T]$ both enstrophy and palinstrophy in the LES flows deviate more significantly from their values in the filtered DNS flow.

6. Discussion & conclusions

Adjoint-based methods have become a workhorse in the solution of unconstrained PDE optimization problems [4]. They make it possible to conveniently determine the gradient (sensitivity) of the objective functional with respect to a distributed control variable, which can then be used in various gradient descent algorithms [36]. Unlike most constraints imposed on the control variable, constraints on the state variables are generally harder to satisfy since they define, via solutions of the governing system, complicated manifolds in the space of control variables. In the present study we demonstrate how the adjoint-based framework can be extended to handle such constraints approximately, which is done by constructing a projection of the gradient of the objective functional onto a subspace tangent to the constraint manifold at the given iteration. This projection is realized by solving another adjoint problem which is defined in terms of the same adjoint operator as the adjoint system employed to determine the gradient, but with different forcing. Thus, for a state constraint defining a codimension-m manifold, a single iteration of the gradient algorithm (1) requires a total of $(m + 1)$ adjoint solves to evaluate the gradient and the projection on the tangent subspace. We focus on the "optimize-then-discretize" paradigm [4] in the infinite-dimensional setting, where it is more difficult to use "black-box" optimization software that could otherwise facilitate imposition of such constraints. Regularity of both the gradient and the projection is carefully considered, which is done by leveraging the Riesz theorem.

Our approach is illustrated with two test problems, both involving a nonlinear scalar constraint: a simple problem describing heat conduction in 1D which is used to introduce the method in a clean setting and a more complicated problem concerning the optimal design of turbulence closures in 2D incompressible flows governed by the Navier-Stokes system. This latter problem has a nonstandard structure [11] and is characterized by a much higher level of technical complexity typical of problems arising in turbulence research. For each of the optimization problems we consider the constrained and unconstrained version with and without the nonlinear constraint. In addition, for comparison, we also consider a version of the first problem where the constraint is homogeneous and hence can be enforced exactly with a retraction operator.

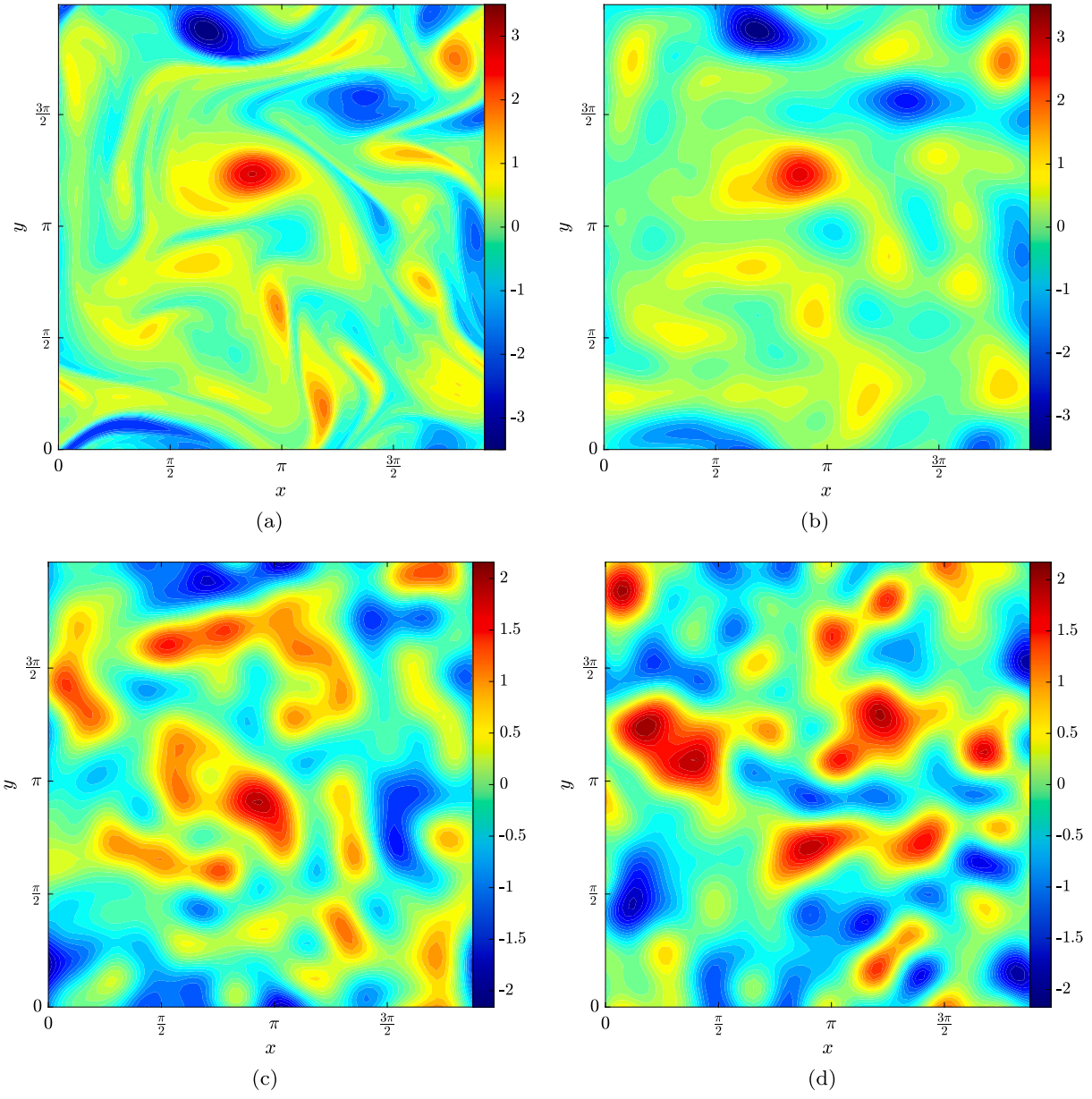

Fig. 11. (a) The vorticity field $w(T,\mathbf{x})$ obtained by solving the Navier-Stokes system (7) and (b) the corresponding filtered field $\widetilde{w}(T,\mathbf{x})$. (c,d) The solutions $\widetilde{\omega}(T,\mathbf{x})$ of the LES system (8) with the optimal eddy viscosities $\check{\nu}(s)$ obtained by solving Problem 3.A and Problem 3.B, respectively. All fields are shown at the time instant $t=T$ corresponding to the end of the optimization window, cf. (12).

Our computation of the gradients and of the normal elements defining the tangent subspaces is carefully validated in Section 4. The computational results presented in Section 5 show that even though in the solutions of Problems 2.B and 3.B the constraints are not satisfied exactly, the obtained optimizers lie much closer to the constraint manifold $\mathcal{M}$ than in Problems 2.A and 3.A where the constraint is not imposed. However, as expected, the optimizers in the constrained problems correspond to larger values of the objective functionals. In this context, we note that it is not a priori known whether Problems 2.B and 3.B admit minimizers belonging to the constraint manifold for which the objective functional would vanish, or nearly vanish (it is, in fact, rather unlikely that this could happen). Therefore, a modest reduction of the objective functional achieved in these problems, cf. Figs. 7a and 9a, should be viewed as a reflection of the difficulty of these problems rather than of limitations of the proposed approach. In practical applications motivating this study even a modest reduction of the objective functional is beneficial if the optimal solution satisfies the required constraints. Importantly, enforcement of the constraint via projection onto the subspace tangent to the constraint manifold is performed in just as straightforward manner as the computation of the gradient. Therefore, the proposed approach is likely to benefit practical applications involving PDE optimization problems.

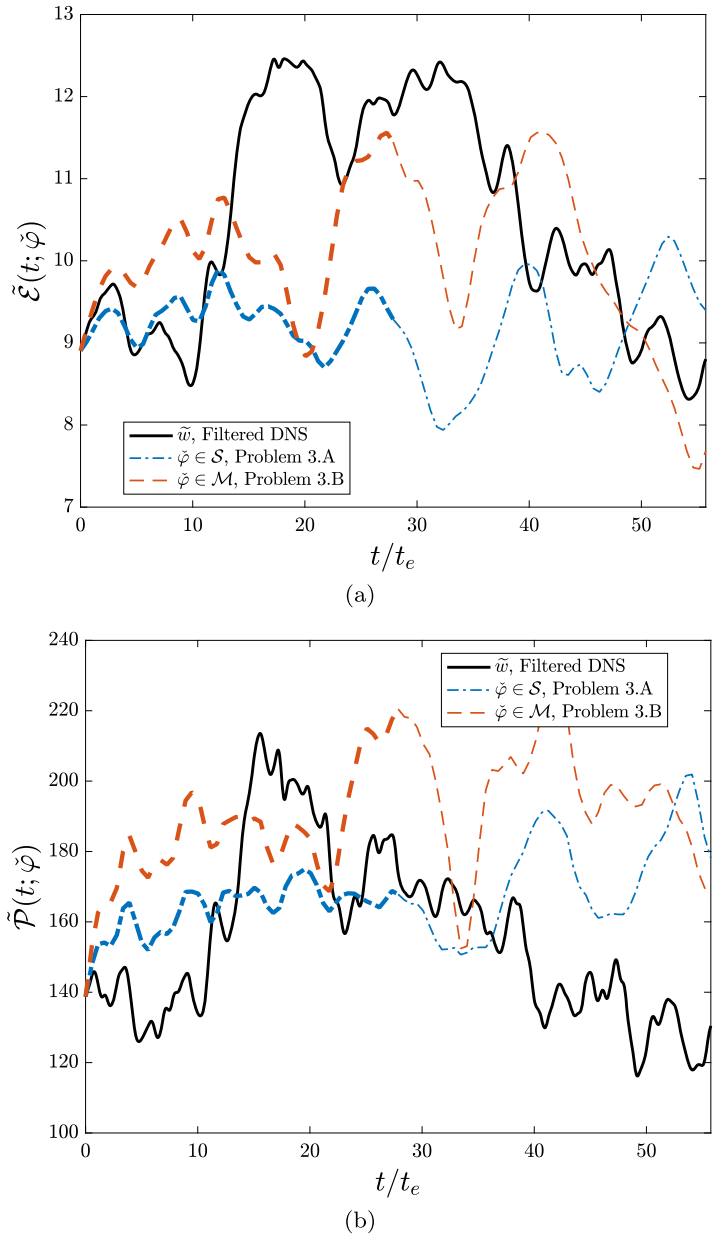

Fig. 12. [Problem 3:] Dependence of (a) the enstrophy (11a) and (b) the palinstrophy (11b) on time for (black, solid line) the filtered DNS flow and the LES flows with the optimal eddy viscosities corresponding to (blue, dashed-dot line) $\breve{\varphi} \in \mathcal{S}$ found by solving Problem 3.A and (red, dashed line) $\breve{\varphi} \in \mathcal{M}$ found by solving Problem 3.B. Thick and thin lines represent to, respectively, time in the "training window" ($t \in [0, T]$) and in the extended window ($t \in (T, 2T]$).

CRediT authorship contribution statement

Pritpal Matharu: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Bartosz Protas:** Supervision, Writing – review & editing, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

Partial support for this research was provided through an NSERC (Canada) Discovery Grant: RGPIN-2020-05710. The first author is supported by Vergstiftelsen.

References

[1] D. Luenberger, Optimization by Vector Space Methods, John Wiley and Sons, 1969.
[2] F. Tröltzsch, Optimal Control of Partial Differential Equations: Theory, Methods and Applications, Graduate Studies in Mathematics, vol. 112, American Mathematical Society, 2010.
[3] J. Lions, Contrôle Optimal de Systèmes Gouvernés par des Équations aux Dérivées Partielles, Dunod, Paris, 1968. English translation: Springer-Verlag, New-York, 1971.
[4] M.D. Gunzburger, Perspectives in Flow Control and Optimization, SIAM, 2003.
[5] V. Bukshtynov, B. Protas, Optimal reconstruction of material properties in complex multiphysics phenomena, J. Comput. Phys. 242 (2013) 889–914, https://doi.org/10.1016/j.jcp.2013.02.034.
[6] E. Vergnault, P. Sagaut, An adjoint-based lattice Boltzmann method for noise control problems, J. Comput. Phys. 276 (2014) 39–61, https://doi.org/10.1016/j.jcp.2014.07.027.
[7] S. Costanzo, T. Sayadi, M. Fosas de Pando, P. Schmid, P. Frey, Parallel-in-time adjoint-based optimization – application to unsteady incompressible flows, J. Comput. Phys. 471 (2022) 111664, https://doi.org/10.1016/j.jcp.2022.111664.
[8] J. Sirignano, J. MacArt, K. Spiliopoulos, PDE-constrained models with neural network terms: optimization and global convergence, J. Comput. Phys. 481 (2023) 112016, https://doi.org/10.1016/j.jcp.2023.112016.
[9] M. Bendsoe, O. Sigmund, Topology Optimization: Theory, Methods, and Applications, Engineering Online Library, Springer Berlin Heidelberg, 2003.
[10] P.-A. Absil, R. Mahony, R. Sepulchre, Optimization Algorithms on Matrix Manifolds, Princeton University Press, 2008.
[11] P. Matharu, B. Protas, Optimal eddy viscosity in closure models for two-dimensional turbulent flows, Phys. Rev. Fluids 7 (2022) 044605, https://doi.org/10.1103/PhysRevFluids.7.044605.
[12] R.A. Adams, J.F. Fournier, Sobolev Spaces, 2nd edition, Elsevier/Academic Press, Amsterdam, 2003.
[13] M. Lesieur, Turbulence in Fluids, 2nd edition, Kluwer Academic Publishers, Dordrecht, Boston, London, 1993.
[14] S.B. Pope, Turbulent Flows, Cambridge University Press, Cambridge, 2000.
[15] P.A. Davidson, Turbulence: An Introduction for Scientists and Engineers, 2nd edition, Oxford University Press, Oxford, 2015.
[16] J.N. Kutz, Deep learning in fluid dynamics, J. Fluid Mech. 814 (2017) 1–4, https://doi.org/10.1017/jfm.2016.803.
[17] M. Gamahara, Y. Hattori, Searching for turbulence models by artificial neural network, Phys. Rev. Fluids 2 (2017) 054604, https://doi.org/10.1103/PhysRevFluids.2.054604.
[18] J. Jimenez, Machine-aided turbulence theory, J. Fluid Mech. 854 (2018) R1, https://doi.org/10.1017/jfm.2018.660.
[19] K. Duraisamy, G. Iaccarino, H. Xiao, Turbulence modeling in the age of data, Annu. Rev. Fluid Mech. 51 (1) (2019) 357–377, https://doi.org/10.1146/annurev-fluid-010518-040547.
[20] K. Duraisamy, Perspectives on machine learning-augmented Reynolds-averaged and large eddy simulation models of turbulence, Phys. Rev. Fluids 6 (2021) 050504, https://doi.org/10.1103/PhysRevFluids.6.050504.
[21] S. Pawar, O. San, Data assimilation empowered neural network parametrizations for subgrid processes in geophysical flows, Phys. Rev. Fluids 6 (2021) 050501, https://doi.org/10.1103/PhysRevFluids.6.050501.
[22] F. Waschkowski, Y. Zhao, R. Sandberg, J. Klewicki, Multi-objective CFD-driven development of coupled turbulence closure models, J. Comput. Phys. 452 (2022) 110922, https://doi.org/10.1016/j.jcp.2021.110922.
[23] P. Matharu, B. Protas, Optimal closures in a simple model for turbulent flows, SIAM J. Sci. Comput. 42 (1) (2020) B250–B272, https://doi.org/10.1137/19M1251941.
[24] B. Protas, T.R. Bewley, G. Hagen, A computational framework for the regularization of adjoint analysis in multiscale PDE systems, J. Comput. Phys. 195 (1) (2004) 49–89, https://doi.org/10.1016/j.jcp.2003.08.031.
[25] C.E. Leith, Diffusion approximation for two-dimensional turbulence, Phys. Fluids 11 (3) (1968) 671–672, https://doi.org/10.1063/1.1691968.
[26] C.E. Leith, Atmospheric predictability and two-dimensional turbulence, J. Atmos. Sci. 28 (2) (1971) 145–161, https://doi.org/10.1175/1520-0469(1971)028<0145:APATDT>2.0.CO;2.
[27] C. Leith, Stochastic models of chaotic systems, Phys. D: Nonlinear Phenom. 98 (2) (1996) 481–491, https://doi.org/10.1016/0167-2789(96)00107-8, Nonlinear Phenomena in Ocean Dynamics.
[28] M.S. Berger, Nonlinearity and Functional Analysis, Academic Press, 1977.
[29] P. Matharu, Constraint optimization - heat equation (HeatOptConstr), https://github.com/pipmath/HeatOptConstr.git, 2023.
[30] R. Alimo, D. Cavaglieri, P. Beyhaghi, T.R. Bewley, Design of IMEXRK time integration schemes via Delaunay-based derivative-free optimization with nonconvex constraints and grid-based acceleration, J. Glob. Optim. 79 (3) (2021) 567–591, https://doi.org/10.1007/s10898-019-00855-1.
[31] L.N. Trefethen, Approximation Theory and Approximation Practice, SIAM, Philadelphia, 2013.
[32] A. Bracco, J.C. McWilliams, Reynolds-number dependency in homogeneous, stationary two-dimensional turbulence, J. Fluid Mech. 646 (2010) 517–526, https://doi.org/10.1017/S0022112009993661.
[33] Q. Wang, R. Hu, P. Blonigan, Least squares shadowing sensitivity analysis of chaotic limit cycle oscillations, J. Comput. Phys. 267 (2014) 210–224, https://doi.org/10.1016/j.jcp.2014.03.002.
[34] P.J. Blonigan, Q. Wang, Multiple shooting shadowing for sensitivity analysis of chaotic dynamical systems, J. Comput. Phys. 354 (2018) 447–475, https://doi.org/10.1016/j.jcp.2017.10.032.
[35] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, Numerical recipes, in: The Art of Scientific Computations, 3rd edition, Cambridge University Press, 2007.
[36] J. Nocedal, S. Wright, Numerical Optimization, 2nd edition, Springer Series in Operations Research and Financial Engineering, Springer, 2006.
[37] H. Engl, M. Hanke, A. Neubauer, Regularization of Inverse Problems, Kluwer, Dordrecht, 1996.
[38] R. Maulik, O. San, J.D. Jacob, Spatiotemporally dynamic implicit large eddy simulation using machine learning classifiers, Physica D 406 (2020) 132409, https://doi.org/10.1016/j.physd.2020.132409.