Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Computational Physics 195 (2004) 49–89
JOURNAL OF COMPUTATIONAL PHYSICS
www.elsevier.com/locate/jcp

A computational framework for the regularization of adjoint analysis in multiscale PDE systems

Bartosz Protas *, Thomas R. Bewley, Greg Hagen

Flow Control Lab, Department of MAE, UC San Diego, La Jolla, CA 92093-0411, USA

Received 5 November 2002; received in revised form 7 August 2003; accepted 28 August 2003

Abstract

This paper examines the regularization opportunities available in the adjoint analysis and optimization of multiscale PDE systems. Regularization may be introduced into such optimization problems by modifying the form of the evolution equation and the forms of the norms and inner products used to frame the adjoint analysis. Typically, L_2 brackets are used in the definition of the cost functional, the adjoint operator, and the cost functional gradient. If instead we adopt the more general Sobolev brackets, the various fields involved in the adjoint analysis may be made smoother and therefore easier to resolve numerically. The present paper identifies several relationships which illustrate how the different regularization options fit together to form a general framework. The regularization strategies proposed are exemplified using a 1D Kuramoto–Sivashinsky forecasting problem, and computational examples are provided which exhibit their utility. A multiscale preconditioning algorithm is also proposed that noticeably accelerates convergence of the optimization procedure. Application of the proposed regularization strategies to more complex optimization problems of physical and engineering relevance is also discussed.

Keywords: Adjoint analysis; Optimization; Regularization; Flow control; 4DVAR

1. Introduction

Adjoint analysis forms a foundation for many applications of model-based control and estimation theory to nonlinear fluid systems, including:
(A) transonic airfoil shape optimization [1],
(B) optimization of open-loop control distributions for transitional and turbulent flow systems [2–5], and
(C) state reconstruction and parameter estimation in numerical weather prediction (known operationally as "4D-VAR") [6].

* Corresponding author. Present address: Department of Mathematics & Statistics, McMaster University, Hamilton, ON, Canada L8S 4K1. Fax: +1-905-522-0935.
E-mail addresses: bprotas@mcmaster.ca (B. Protas), bewley@ucsd.edu (T.R. Bewley), gregh@turbulence.ucsd.edu (G. Hagen).

0021-9991/$ - see front matter
doi:10.1016/j.jcp.2003.08.031

For recent general reviews we refer the reader to, e.g. Gunzburger [7] and Sritharan [8]. In order to apply adjoint analysis, an appropriately defined *cost functional* is first expressed which represents mathematically the physical objective in performing the computational optimization. In problem A, the objective is typically to maximize the lift/drag ratio of the airfoil for a range of different cruise configurations while respecting a variety of practical "feasibility" constraints related to the construction of the airfoil. In problem B, the objective is typically to reduce drag, surface pressure fluctuations, or flow-induced noise or to reduce the kinetic energy of the flow perturbations in order to inhibit transition to turbulence, though in combustion applications the objective is typically the opposite – that is, to excite the flow with minimal control input in order to enhance turbulent mixing. In problem C, the objective is typically to reconcile the numerical weather model with recent weather measurements in order to obtain accurate weather forecasts. All of the above problems in fact represent *inverse problems*, in which one seeks to determine inputs to the system such that the outputs will have desired properties. Once the control objective is expressed mathematically as a cost functional, an appropriately defined *adjoint field* may be used as a tool to determine an appropriately defined *gradient* of the cost functional with respect to the control variables. The adjoint field calculation is thus a central component of high-dimensional gradient-based control optimization strategies. Refs. [5,9] contain brief reviews of our perspective on a few of the relevant issues related to such problems.

Even though the mathematical framework for adjoint-based optimization is fairly mature and has already been used successfully in a broad range of applications in fluid mechanics, many flow systems still present fundamental challenges to this approach. Due to their nonlinear nature, fluid-mechanical systems are often characterized by energetic motions over a broad range of length scales and time scales. Numerical computations of such flows must be performed with care to account properly for this broad range of scales. Adjoint analyses of such flows must also be crafted with care in order to be well behaved over this full range of scales. Inverse problems are often ill-posed in the sense that multiple solutions which locally minimize the relevant cost functional exist, and the solution found by the optimization algorithm does not necessarily have a continuous dependence on the data provided. For instance, in numerical weather prediction, the problem of finding the future state of the nonlinear model based on past measurements is often ill-posed in this regard. In such problems, errors magnify exponentially in time in the linearized ("perturbation") analysis. In the iterative numerical solution of such an inverse problem in the nonlinear setting, a small change in the data provided (the measurements) can sometimes lead to a large change in the forecast. Even in the control of laminar flows, adjoint fields typically grow exponentially when they are calculated (in reverse time), and can thus be exceedingly difficult to resolve in regions where flow perturbations amplify quickly, such as in thin shear layers.

The presence of a broad range of interacting length- and time-scales thus complicates an adjoint-based analysis of a nonlinear multiscale system by causing difficulties of twofold nature: on the one hand, the dependence of the solution on the data in such cases is irregular, resulting in the presence of many local minima of the cost functional; on the other hand, the various fields involved in an adjoint analysis are not smooth and therefore difficult to resolve numerically. These two issues are related, and may be addressed, at least partially, by considering the regularity of the various fields involved in the analysis of such systems. We therefore define the term *regularization* rather broadly in the present work as a reformulation of an adjoint-based algorithm in such a manner that at least some of the fields involved in this analysis are more "smooth", in the sense that the energy spectrum in these fields decays more rapidly with wavenumber at the length scales of concern from the perspective of a numerical implementation (throughout this paper by "energy spectrum" we will mean the "kinetic energy spectrum"). Such regularization will thus render a given optimization problem more amenable to numerical treatment, and may sometimes turn an initially ill-posed problem into a well-posed problem. A more narrow definition of the term regularization is often adopted in the precise mathematical study of ill-posed inverse problems (see [10–12]); however our broader usage of the term is adequate for the present investigation.

Though great attention has been paid to the appropriate treatment/modeling of subgrid-scale (SGS) effects in the numerical simulation literature, much less is known about how to address the multiscale nature of fluid-mechanical systems in adjoint analysis. The central issue is that norms and inner products, collectively to be referred to in this paper as "brackets", are implied, if not explicitly stated, at three distinct steps of the derivation of an adjoint analysis. Each of these brackets implies a relative weighting between the various length scales and time scales in the relationship expressed. "Traditional" approaches have typically selected L_2 brackets at most, if not all, of these steps in the adjoint formulation. However, this choice is by no means unique. Other choices are sometimes more appropriate when the system must be solved on a computer with finite-precision arithmetic using a discrete numerical grid in both space and time. Modifying the definitions of the three brackets used to frame the adjoint analysis facilitates a range of distinct but related regularization opportunities. Capitalizing on these regularization opportunities can result in significantly smoother space–time systems requiring numerical approximation in the adjoint analysis. The purpose of the present paper is to clarify these important regularization opportunities and to illustrate how they may be used in concert to increase the speed, stability, and accuracy of adjoint-based numerical optimization algorithms.

The possibility of achieving a regularizing effect through modification of some of the brackets used in the derivation of an adjoint algorithm was already recognized by Collis et al. [13]. A similar set of ideas, but in the finite-dimensional setting, was earlier suggested by Heinkenschloss and Vicente [14]. Concepts related to the use of Sobolev inner products to extract different cost functional gradients were developed by Neuberger [15], who employed these ideas in the solution of direct problems involving differential equations. Applications to optimization problems were recently considered by Sial et al. [16]. Other approaches explicitly addressing data assimilation in multiscale environments include the works of Liu [17], Brandt and Zaslavsky [18], Grimstad and Mannseth [19], and Brusdal and Mannseth [20]. In particular, [17] and [20] invoke the concept of "basis norm rescaling", which is further elaborated here. Multigrid approaches to optimization problems were studied by Lewis and Nash [21]. Similar ideas also appear in the Numerical Weather Prediction (NWP) literature: e.g., Lorenc [22] discussed performing optimization with respect to a rescaled variable, whereas Thepaut and Moll [23] recognized the possibility of using various inner products to precondition the adjoint algorithm. An adaptive method to enforce smoothness constraints in data assimilation was recently presented by Lin et al. [24]. In the present study, we attempt to synthesize these various regularization opportunities into a more unified framework and characterize the interplay between the various definitions of the brackets upon which adjoint analyses are based.

To make this study concrete, the bulk of the paper illustrates how these regularization techniques may be applied to a Kuramoto–Sivashinsky model system. This 1D model problem illustrates self-sustained chaotic behavior of a multiscale nature, and thus provides an appropriate testbed for the present study. The Kuramoto–Sivashinsky equation was used successfully as a low-dimensional prototype for complicated fluid systems by Holmes et al. [25]. In fact, many advanced flow control strategies were initially investigated using similar 1D models; see, e.g. [26–28].

In Section 2 we identify and discuss in general terms the four fundamental classes of regularization opportunities available in the framing of an adjoint analysis. As a point of reference, the standard optimization framework for the Kuramoto–Sivashinsky model is formulated in Section 3. The different regularization possibilities available in framing the adjoint analysis, and interrelationships between these different formulations, are discussed in Section 4, and some numerical examples are presented in Section 5. Several extensions of this study are also underway, and are briefly introduced in Section 6. Concluding remarks are presented in Section 7.

2. The four fundamental classes of regularization opportunities

In the adjoint-based optimization of unsteady PDE systems in general, there are three spatial domains of interest: the domain on which cost functional is defined, which we denote Ω_1, the domain over which the

state of the system modeled, which we denote Ω_2, and the domain on which the control is applied, which we denote Ω_3. In an unsteady problem, the system model is defined on Ω_2 over some time interval $[0,T]$. The cost functional which measures this model on Ω_1 may average over the interval $[0,T]$, as in "regulation" problems, or may focus the attention on the state at time $t=T$, which is called a "terminal control" problem. The control on Ω_3 can also be defined over $[0,T]$, when an unsteady control distribution is sought, or may be defined at time $t=0$, as done in the forecasting problem (where the "control" is actually the initial condition). In the process of adjoint-based optimization, brackets are used (or implied, if not explicitly stated) on all three of these space–time domains.

In the continuous setting, the form of each of these three brackets may incorporate either derivatives or "anti-derivatives" in both space and time. Mathematically, these brackets are related to the natural measures for functions defined in the Sobolev space $H^p(0,T;H^q(\Omega_i))$, where q is the differentiability order in space, p is the differentiability order in time, and Ω_i denotes the spatial domain. Note that Sobolev brackets with negative differentiability indices can also be considered in this framework by taking p and/or q negative. Such brackets are natural alternatives to the L_2 bracket when considering functions of different degrees of regularity in both space and time. How each of these brackets is defined, in addition to any smoothing that might be applied to the state equation itself, has important consequences on the smoothness of the several variables in the optimization problem, as summarized in Fig. 1. As a shorthand, we use Ψ_1, Ψ_2, and Ψ_3 to identify the brackets selected for the three space–time domains of interest in this problem.

Note that, even though we have borrowed certain concepts from the functional analysis literature in the present discussion, and even though much of the presentation in this paper will be in a compact infinite-dimensional notation to facilitate interpretation, the present paper specifically does *not* concern the problem of functional analysis, which would involve the mathematical characterization of the precise degree of differentiability of each of the quantities involved in the optimization process. Functional analysis is essential for establishing well-posedness (i.e., existence and uniqueness of solutions, and their continuous dependence on data) for many optimization problems governed by PDEs. However, it provides little practical advice regarding the efficient numerical solution of the corresponding optimization problems in the finite-dimensional setting, an issue which is of primary interest to us here. Furthermore, well-posedness has not yet been established (even in the uncontrolled setting) for many systems important from the physical and engineering point of view, such as 3D Navier–Stokes systems with large data. Nevertheless, optimization problems governed by such PDEs are very important from the application perspective, and in this work we develop a generic *computational* framework which does not depend on how much is known a priori about well-posedness of the underlying PDE system. We will assume (without further justification) that the energy content of all of the fields involved (the control, the state, etc.) eventually decays exponentially with wavenumber, and thus these fields belong to the class $C^\infty(0,T;C^\infty(\Omega_i))$, so that they may be

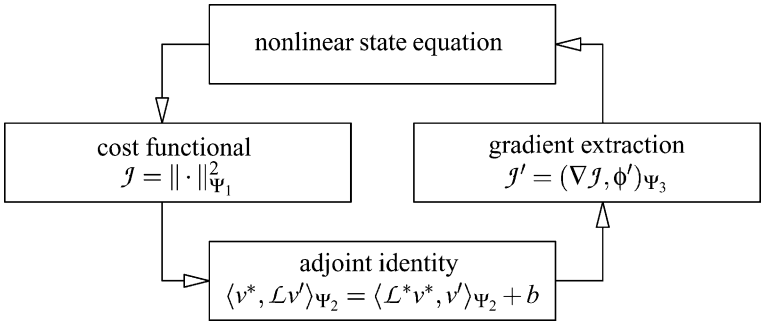

Fig. 1. The four essential components of the adjoint-based optimization process. As outlined in the text, each component of this process is associated with a distinct opportunity for regularization.

differentiated as many times as necessary in the analysis that follows. In fact, in our finite-dimensional calculations, we will typically form some type of spectral truncation so that the energy content of the fields we actually compute is precisely zero at sufficiently small length scales. To summarize, the present paper is not about the mathematical characterization of when a particular optimization problem governed by PDEs is well-posed, but rather it is about the engineering problem of how to get adjoint analysis to work efficiently on multiscale systems obtained by discretizing a given PDE optimization problem, whether or not such a characterization of well-posedness of the corresponding infinite-dimensional problem is tractable.

The first regularization opportunity is given by adding an artificial (but well-motivated) term to the discretized state equation itself. Two common examples are dynamic subgrid-scale models (in turbulence research) and hyperviscosity (in numerical weather prediction). Addition of such a term to the numerical model is useful for tuning the behavior of the numerical model at the unresolvable scales, and may also be needed for numerical stability. In addition to modifying the actual governing equation, we can also consider its different derived forms (e.g., the vorticity form instead of the velocity–pressure form of the Navier–Stokes equation). These different yet equivalent forms may serve to focus on different aspects of the dynamics in numerical simulations and adjoint analyses thereof.

The second regularization opportunity is given by the definition of the cost functional. As mentioned previously, the cost functional can take any of a wide variety of forms depending on the problem under consideration. However, in most such formulations, the cost functional involves the norm of a flow quantity taken over some subdomain of the space–time domain under consideration. Selecting for this purpose a norm Ψ_1 which incorporates either derivatives or anti-derivatives, instead of using the standard L_2 norm, effectively builds in a "filter" into the definition of the cost functional, thereby allowing extra emphasis to be placed on certain scales of interest in the multiscale problem. Note that the cost functional may also incorporate a term penalizing the magnitude of the control on Ω_3, using an appropriate norm in order to limit the magnitude of the control that results from the optimization. [1] Such an approach is known as Tikhonov regularization [29]. Analysis of the effect of this additional term on the stability of the optimization algorithm is deferred to Section 4.5.6.

The third regularization opportunity is given by the form of the inner product which is used to define the adjoint operator and, ultimately, the adjoint field itself. Incorporating derivatives or anti-derivatives into the definition of the inner product Ψ_2, instead of using the standard L_2 inner product, can be useful to obtain well-behaved (that is, numerically tractable) adjoint operators.

Finally, the fourth regularization opportunity is the definition of the inner product used to extract the cost functional gradient. Incorporating derivatives into the inner product Ψ_3, instead of using the standard L_2 inner product, has the effect of scale-dependent filtering, and allows one to extract smoother gradients, thereby preconditioning the optimization process.

The above regularization opportunities fall into two categories: those which affect the descent direction (modifying the cost functional and the gradient extraction procedure) and those which affect how a given gradient is computed (modifying the form of the governing equation and the form of the adjoint identity). In the discrete setting, the options belonging to the first category affect the conditioning of the optimization problem, whereas the options belonging to the second category affect the complexity of the gradient computation. In the present paper we characterize the interplay of the different regularization options in the discrete setting and discuss how they can be used in concert to improve adjoint-based analyses of difficult multiscale problems of both physical and engineering interest, such as high Reynolds number turbulence. In our analysis below we will seek to delineate the different generic opportunities available, but will not

[1] Note that inclusion of such a term is sometimes, but not always, necessary to insure that the optimization problem has a bounded solution. See [5] for further discussion.

attempt to formulate specific recommendations regarding how they should be applied in a given problem-specific context.

3. Forecasting a Kuramoto–Sivashinsky system: the standard framework

In this section, we first describe three different yet equivalent forms of a dynamically rich 1D model system governed by the Kuramoto–Sivashinsky equation, then outline a relatively standard adjoint-based algorithm for the forecasting of this system based on limited noisy measurements. In the section that follows, we illustrate, in turn, the application of the four regularization opportunities summarized above. The Kuramoto–Sivashinsky equation [30,31] has been widely studied due to its chaotic, pattern-forming behavior. Out of the several different normalizations of the parameters of this system which are available in the literature, we have selected the one proposed by Hyman and Nicolaenko [32], in which the system is written

$$\begin{cases} \partial_t u + 4\partial_x^4 u + \kappa\left[\partial_x^2 u + \frac{1}{2}(\partial_x u)^2\right] = 0, & x \in \Omega,\ \ t \in [0,T], \\ \partial_x^i u(0,t) = \partial_x^i u(2\pi,t), & t \in [0,T],\ \ i = 0,\ldots,3, \\ u(x,0) = \psi, & x \in \Omega, \end{cases} \tag{1}$$

where we define $\partial_x^i \triangleq (\partial^i/\partial x^i)$. Integrating this system over the domain $\Omega \triangleq [0,2\pi]$, the evolution of the mean of u is given by $\partial_t \int_0^{2\pi} u\,\mathrm{d}x = -(\kappa/2)\int_0^{2\pi}(\partial_x u)^2\,\mathrm{d}x \neq 0$. For this reason, it is common to transform the system (1) into a different form, which is achieved by first differentiating it with respect to x and then re-expressing it in terms of a new variable $v \triangleq \partial_x u$ such that

$$\begin{cases} \partial_t v + 4\partial_x^4 v + \kappa\left(\partial_x^2 v + v\partial_x v\right) = 0, & x \in \Omega,\ \ t \in [0,T], \\ \partial_x^i v(0,t) = \partial_x^i v(2\pi,t), & t \in [0,T],\ \ i = 0,\ldots,3, \\ v(x,0) = \partial_x \psi \triangleq \phi, & x \in \Omega. \end{cases} \tag{2}$$

This is the form of the Kuramoto–Sivashinsky system that is studied most often. As $v = \partial_x u$ and u is periodic in x, it follows immediately that $\int_0^{2\pi} v\,\mathrm{d}x = 0$ for all t. The variable u can be recovered from v by inverting the differential operator ∂_x and accounting for the mean of u properly. For this purpose we define the "anti-derivative" operator ∂_x^{-1} such that

$$\partial_x^{-1} z(x) \triangleq \int_0^x z(x')\mathrm{d}x' - \frac{1}{2\pi}\int_0^{2\pi}\int_0^s z(x')\,\mathrm{d}x'\,\mathrm{d}s.$$

Note that the constant is selected such that $\int_0^{2\pi} \partial_x^{-1} z(x)\,\mathrm{d}x = 0$. Using this operator, we may recover u from v with

$$u(x,t) = \partial_x^{-1} v(x,t) + \frac{1}{2\pi}\int_0^{2\pi}\psi(x')\,\mathrm{d}x' - \frac{\kappa}{4\pi}\int_0^t\int_0^{2\pi}[v(x',t')]^2\,\mathrm{d}x'\,\mathrm{d}t'.$$

Yet another form of the Kuramoto–Sivashinsky system can be obtained by further differentiating the system (2) and defining $w \triangleq \partial_x v$

$$\begin{cases} \partial_t w + 4\partial_x^4 w + \kappa\left(\partial_x^2 w + w^2 + \partial_x^{-1} w\partial_x w\right) = 0, & x \in \Omega,\ \ t \in [0,T], \\ \partial_x^i w(0,t) = \partial_x^i w(2\pi,t), & t \in [0,T],\ \ i = 0,\ldots,3, \\ w(x,0) = \partial_x \phi \triangleq \varphi, & x \in \Omega. \end{cases} \tag{3}$$

In the derivations to come, we will primarily focus on system (2), which we will dub the *primitive formulation*. By analogy with the equations of fluid dynamics, we will refer to the integral form (1) as the

streamfunction formulation and the derivative form (3) as the *vorticity formulation* of the Kuramoto–Sivashinsky system; the same qualifiers will be used with regard to the corresponding variables. In the above equations the parameter κ has an interpretation similar to Reynolds number in fluid systems. For sufficiently large values of this parameter, the Kuramoto–Sivashinsky system exhibits self-sustained, chaotic dynamics. Issues related to the functional setting of this equation and characterization of its attractor are discussed in Temam [33].

We now review the relatively standard framework for adjoint-based data assimilation in PDE systems by focusing on an admittedly contrived, yet dynamically rich, 1D model forecasting problem based on the Kuramoto–Sivashinsky system discussed above. Extensions of these approaches for the regularization of adjoint analyses in more realistic forecasting and control problems in fluid mechanics are discussed in Section 6. In the model problem to be considered, the three spatial domains $\Omega_i,\ i = 1, 2, 3$, mentioned in the previous section happen to coincide. Note that this will not necessarily be the case in general (cf. Section 6.2). More precisely, the "control", which is taken to be the initial condition ϕ in this problem, is defined on $\Omega = [0, 2\pi]$ at time $t = 0$, whereas both the system evolution and the cost functional are defined on $\Omega \times [0, T]$. We begin by first defining the norm

$$\|z\|^2_{L_2(0,T;L_2(\Omega))} \triangleq \int_0^T \int_0^{2\pi} z^2 \,\mathrm{d}x\,\mathrm{d}t, \tag{4}$$

and then attempting to minimize the cost functional

$$\mathscr{J}(\phi) = \frac{1}{2}\|\mathscr{H}v - y\|^2_{L_2(0,T;L_2(\Omega))}, \quad \text{where} \quad y = \mathscr{H}v^{\mathrm{act}} + \eta, \tag{5}$$

v^{act} is the "actual" state (which is unknown to the forecasting algorithm), and v is the "reconstructed" state, which is assumed to be related to the initial state ϕ (the quantity to be determined in the reconstruction problem) via the primitive form (2) of the Kuramoto–Sivashinsky system. Note that $\mathscr{H}$ denotes the "observation" operator, y denotes the corresponding noisy measurements taken of the system, and η denotes the measurement noise. The problem to be solved is to find the initial state ϕ in the reconstruction problem which will minimize $\mathscr{J}$, thereby minimizing the deviation of the measurements from the corresponding quantities in the reconstructed system. The observation operator $\mathscr{H}$ which we have selected for this model problem is based on the cosine decomposition of the flow system. To define $\mathscr{H}$, we will make use of a linear projection operator $\mathscr{P}_r$ defined such that

$$\mathscr{P}_r z \triangleq \left[\frac{1}{\pi}\int_0^{2\pi} \cos(rx')z(x')\,\mathrm{d}x'\right]\cos(rx). \tag{6}$$

Note that the projection operator so defined satisfies $\mathscr{P}_r^2 = \mathscr{P}_r$. We now define the observation operator as

$$\mathscr{H} \triangleq \sum_{r\in\Lambda_r} \mathscr{P}_r, \tag{7}$$

where Λ_r is the set of modes which we choose to observe.

For $\mathscr{J}(\phi)$ to be minimized by ϕ, it is necessary that, in the immediate neighborhood of ϕ, the perturbation $\mathscr{J}'$ of the cost functional $\mathscr{J}$ that arises from perturbations $\varepsilon\phi'$ to the control distribution ϕ vanish for all feasible directions ϕ' as ε is made small. To be precise, the quantity $\mathscr{J}'(\phi;\phi')$ is defined by a limiting process as the differential [2] of the cost functional $\mathscr{J}$ with respect to ϕ such that

[2] In the present work we assume that $\mathscr{J}(\phi)$ is sufficiently smooth that it is differentiable, which is a usual assumption in numerical optimization studies.

$$\mathscr{J}'(\phi;\phi') \triangleq \lim_{\varepsilon\to 0} \frac{\mathscr{J}(\phi+\varepsilon\phi') - \mathscr{J}(\phi)}{\varepsilon}. \tag{8}$$

To summarize, if $\mathscr{J}(\phi)$ is minimized by ϕ, then $\mathscr{J}'(\phi;\phi') = 0$ for all feasible ϕ'; this is referred to as the *first-order necessary optimality condition* for the present problem. Higher-order differentials may also be considered (namely, the second-order sufficient optimality condition), however, we will not make use of such higher-order expressions in this paper.

The differential of the cost functional defined in (5) can now be calculated in the neighborhood of some state $v(\phi)$, which yields

$$\mathscr{J}'(\phi;\phi') = \int_0^T \int_0^{2\pi} (\mathscr{H}v - y)\mathscr{H}v' \,\mathrm{d}x\,\mathrm{d}t, \tag{9}$$

where, by linearization of (2), it follows that $v'(\phi;\phi')$ is the solution of the system

$$\begin{cases} \mathscr{L}v' = 0, & x\in\Omega,\ \ t\in[0,T], \\ \partial_x^i v'(0,t) = \partial_x^i v'(2\pi,t), & t\in[0,T],\ \ i=0,\ldots,3, \\ v'(x,0) = \phi', & x\in\Omega, \end{cases} \tag{10}$$

where the linear operator $\mathscr{L}$ is

$$\mathscr{L}v' \triangleq \partial_t v' + 4\partial_x^4 v' + \kappa\left[\partial_x^2 v' + v\partial_x v' + (\partial_x v)v'\right] \tag{11}$$

and $v(\phi)$ is the solution of (2). Note that, as was the case with v, it follows immediately that $\int_0^{2\pi} v'\,\mathrm{d}x = 0$ for all t.

Numerically, the optimal initial state ϕ and the corresponding evolution of the system v cannot be determined solely from the mathematical statement of the first-order necessary optimality condition, that is, the vanishing of the differential of the cost functional at the optimum distribution of ϕ for all feasible ϕ'. A stable numerical procedure to find such a minimum of $\mathscr{J}(\phi)$ even when no good initial guess of the optimum controls is available (which is quite often the case) is to use an iterative gradient-based optimization procedure: given some initial guess ϕ_0 for the initial conditions ϕ, consecutive refinements $\phi^{(n)}$ are computed using a gradient-based optimization algorithm [3] until convergence to a (local) minimum of $\mathscr{J}$ is obtained. In order to apply such a gradient-based optimization procedure, we need somehow to define a gradient in the space of the control distributions. This is accomplished by identifying the differential (9) as an appropriately defined *inner product* of a quantity, which we will denote $\nabla\mathscr{J}$, with the control perturbation ϕ'. The quantity $\nabla\mathscr{J}$ so defined represents the rate of change in $\mathscr{J}$ when ϕ is updated an infinitesimal amount in the direction ϕ'. We thus identify $\nabla\mathscr{J}$ as a *gradient* in the space where the metric, which effectively defines angles and distances, is given by the inner product selected. Significantly, note that different choices of this inner product will result in different gradients of $\mathscr{J}$ for a particular value of the control distribution ϕ. However, for convex $\mathscr{J}$, all such definitions of the gradient eventually lead to the same minimizer (that is, the optimal value of ϕ), at which $\nabla\mathscr{J} = 0$ regardless of the inner product used to define the gradient.

The most common choice for the inner product used to extract the gradient $\nabla\mathscr{J}$ from the expression for $\mathscr{J}'$ is the L_2 inner product and, for the time being, our derivation is performed using this inner product, that is

[3] In such high-dimensional optimization problems, quasi-Newton methods utilizing Hessians of the cost functional are becoming increasingly popular (see, e.g. [34]). The concepts discussed in this paper appear to extend to such optimization algorithms; such extensions will be considered in future work.

$$\mathscr{J}' \triangleq (\nabla \mathscr{J}, \phi')_{L_2(\Omega)} \triangleq \int_0^{2\pi} (\nabla \mathscr{J}) \phi' \, \mathrm{d}x. \tag{12}$$

In order to identify $\nabla \mathscr{J}$, we first need to transform the expression (9) into a form in which the control perturbation ϕ' is factored out in a convenient manner, as shown above. Note that v' is related to ϕ' through the involved yet linear relationship (10). To accomplish this factorization, we introduce the following inner product:

$$\langle z_1, z_2 \rangle_{L_2(0,T;L_2(\Omega))} \triangleq \int_0^T \int_0^{2\pi} z_1 z_2 \, \mathrm{d}x \, \mathrm{d}t. \tag{13}$$

Based on this bracket, we may derive an adjoint operator $\mathscr{L}^*$ and a corresponding boundary term $b_{\mathscr{L}}$ from the following identity:

$$\langle v^*, \mathscr{L} v' \rangle_{L_2(0,T;L_2(\Omega))} = \langle \mathscr{L}^* v^*, v' \rangle_{L_2(0,T;L_2(\Omega))} + b_{\mathscr{L}}. \tag{14}$$

Using integration by parts and the definition of $\mathscr{L}$ given in (11), it follows that

$$\begin{aligned} \mathscr{L}^* v^* = & -\partial_t v^* + 4\partial_x^4 v^* + \kappa\left(\partial_x^2 v^* - v\partial_x v^*\right), \text{ and } \quad b_{\mathscr{L}} = \left[\int_0^{2\pi} v^* v' \, \mathrm{d}x \right]_{t=0}^{t=T} \\ & + \left\{ \int_0^T 4\left[v^* \partial_x^3 v' - (\partial_x v^*)\partial_x^2 v' + \left(\partial_x^2 v^*\right)\partial_x v' - \left(\partial_x^3 v^*\right) v' \right] + \kappa\left[v^* \partial_x v' - (\partial_x v^*) v' + v^* v v' \right] \mathrm{d}t \right\}_{x=0}^{x=2\pi}. \end{aligned} \tag{15}$$

Making use of the adjoint operator derived above, we may now *define* an adjoint system in the following (as yet, arbitrary) manner:

$$\begin{cases} \mathscr{L}^* v^* = \mathscr{H}^*(\mathscr{H} v - y) \triangleq f, & x \in \Omega, \ \ t \in [0,T], \\ \partial_x^i v^*(0,t) = \partial_x^i v^*(2\pi,t), & t \in [0,T], \ \ i = 0,\ldots,3, \\ v^*(x,T) = 0, & x \in \Omega, \end{cases} \tag{16}$$

where $\mathscr{H}^*$ is defined in a manner analogous to $\mathscr{L}^*$ in (14), and thus it is easy to show that $\mathscr{H}$ is self-adjoint (that is, that $\mathscr{H}^* = \mathscr{H}$). We will refer to (16) as the *primitive* adjoint system and to v^* as the *primitive* adjoint variable. We now combine the state, perturbation, and adjoint systems [(2), (10), and (16)] into the identity given in (14). Note that all the boundary terms in $b_{\mathscr{L}}$ resulting from integrations by parts in space vanish due to periodicity. [4] Due to the choice of the RHS forcing term in the adjoint system (16), we may use (14) to re-express the differential given in (9) in the desired factored form

$$\mathscr{J}'(\phi;\phi') = \int_0^{2\pi} v^*|_{t=0} \phi' \, \mathrm{d}x = \left(v^*\Big|_{t=0}, \phi'. \right)_{L_2(\Omega)}, \tag{17}$$

where v^* denotes the solution of the adjoint problem defined in (16). Finally, note that the mean of the adjoint field defined by (16) is not zero, yet all feasible ϕ' under consideration have zero mean mode. Because of this restriction on the class of ϕ' under consideration, (17) is in fact equivalent to $\mathscr{J}'(\phi;\phi') = (\bar{v}^*|_{t=0}, \phi')_{L_2(\Omega)}$, where the overbar implies that the given variable has the mean mode removed, that is,

[4] Without further mention, we will make use of this fact in many of the transformations to follow.

$$\bar{z} \triangleq z - \frac{1}{2\pi}\int_0^{2\pi} z\,\mathrm{d}x. \tag{18}$$

Note that $\bar{v}^*$ denotes an "orthogonal projection" with respect to the inner product (12) of the adjoint variable v^* onto the space of feasible ϕ'. Thus, the gradient which we seek in the space of feasible ϕ', as indicated in (12), may now be identified as

$$\nabla \mathscr{J} = \bar{v}^*|_{t=0}. \tag{19}$$

The gradient so defined can now be used to find the optimal feasible initial condition using any of a number of standard gradient-based optimization algorithms.

4. Regularizing the Kuramoto–Sivashinsky forecasting problem

In the subsections that follow, we discuss how the regularization opportunities introduced in Section 2 can be applied to fine-tune the adjoint algorithm outlined in Section 3 to better treat multiscale phenomena. In this discussion, we will first investigate adjoint analyses based on the different yet equivalent forms (1) and (3) of the governing equation (2). We will then consider a variety of alternative definitions of the three distinct brackets at the heart of the adjoint formulation, as outlined in Section 2 and listed in the above "standard" formulation as

- the norm $\|\cdot\|_{\Psi_1}$ in (4), which is used to define the cost functional,
- the inner product $\langle\cdot,\cdot\rangle_{\Psi_2}$ in (13), which is used to define the adjoint operator, and
- the inner product $(\cdot,\cdot)_{\Psi_3}$ in (12), which is used to define the gradient.

Note that, in the standard formulation given in Section 3, L_2 brackets over the appropriate space–time domains were used in all three cases. In the subsections to come, we will discuss at length the effects of various choices for Ψ_1, Ψ_2, and Ψ_3. In particular, we will make extensive use of the following brackets:

$$\|z\|^2_{L_2(0,T;H^q(\Omega))} \triangleq \int_0^T \int_0^{2\pi} \left(\partial_x^q z\right)^2 \mathrm{d}x\,\mathrm{d}t, \tag{20a}$$

$$\langle z_1, z_2\rangle_{L_2(0,T;H^q(\Omega))} \triangleq \int_0^T \int_0^{2\pi} \left(\partial_x^q z_1\right)\left(\partial_x^q z_2\right) \mathrm{d}x\,\mathrm{d}t, \tag{20b}$$

$$(z_1, z_2)_{H^q(\Omega)} \triangleq \int_0^{2\pi} \left(\partial_x^q z_1\right)\left(\partial_x^q z_2\right) \mathrm{d}x, \tag{20c}$$

which are related to seminorms on the Sobolev spaces $H^q(\Omega)$. To simplify the nomenclature, we will refer to these brackets as simply H^q inner products or norms (though we will not make use of any of the sophisticated mathematical machinery of functional analysis in Sobolev spaces). For simplicity, we will restrict our attention to the cases with $q = 0$ and ± 1, though higher-order derivatives may also be considered. Note that the special case of $q = 0$ reduces the H^q brackets defined above to the L_2 cases considered previously, as defined in (4), (13), and (12). Also, the present paper will focus on brackets incorporating spatial derivatives only. Formulations generalizing the bracket definitions to include time derivatives as well as space derivatives are also possible, as discussed briefly in Appendix A. Finally, note that it is straightforward to extend these bracket definitions by taking linear combinations of the H^q brackets for various values of q. This fact was recognized previously in [5] for the purpose of extending the definition of the norm used in the cost functional, thereby focusing the cost functional on the particular range of length scales of interest in the system under consideration. In the present work (in Section 4.4), we will develop this extension further by

demonstrating how it may be applied to the definition of the inner product used to extract the gradient, thereby preconditioning the optimization process in a tunable manner. It should be emphasized that, in accordance with our ultimate interest in discrete problems, the subscripts used on the brackets (20a)–(20c) do not imply any differentiability property of the states z_1 and z_2, but rather a different weighing of the Fourier components of their discretizations. In this discussion, the following inner product, defined as a weighted linear combination of the L_2, H^1, and H^{-1} inner products, will be used heavily

$$(z_1,z_2)_{W^{l_1,l_{-1}}} \triangleq \frac{l_{-1}^2}{(1+l_1^2)(1+l_{-1}^2)}\int_0^{2\pi}\left[z_1z_2+\frac{l_1^2l_{-1}^2}{l_1^2+l_{-1}^2}(\partial_x z_1)(\partial_x z_2)+\frac{1}{l_1^2+l_{-1}^2}\left(\partial_x^{-1}z_1\right)\left(\partial_x^{-1}z_2\right)\right]\mathrm{d}x. \tag{21}$$

The justification for the specific choice used above for the coefficients weighing the three terms will become apparent in Section 4.4. Taking the appropriate limits as l_1 and l_{-1} approach zero and infinity, it follows that

$$(z_1,z_2)_{W^{0,\infty}} = (z_1,z_2)_{L_2(\Omega)}, \quad (z_1,z_2)_{W^{\infty,\infty}} = (z_1,z_2)_{H^1(\Omega)}, \quad (z_1,z_2)_{W^{0,0}} = (z_1,z_2)_{H^{-1}(\Omega)}.$$

The form $(z_1,z_2)_{W^{l_1,\infty}}$ is thus a linear combination of the L_2 and H^1 inner products, whereas $(z_1,z_2)_{W^{0,l_{-1}}}$ is a linear combination of the H^{-1} and L_2 inner products. We will use the symbols $W^{l_1,l_{-1}}$, $W^{l_1,\infty}$ and $W^{0,l_{-1}}$ to refer to these different inner products. Symbols representing the spatial components of the different brackets will be used as superscripts to identify the way in which the different objects (that is, the cost functionals, the adjoint operators with the associated adjoint variables, and the cost functional gradients) are constructed. When such symbols are omitted, L_2 brackets are implied (see Section 3).

4.1. The adjoints of alternative forms of the evolution equation

As indicated in (1) and (3), by applying integral or differential operators to the governing equation in the primitive form (2) and suitably redefining the state variable, we obtain a family of systems representing the same conservation law, but emphasizing different aspects (length scales) of the system dynamics. Needless to say, all of these systems are formally equivalent. However, they are characterized by different energy spectra. Thus, we can expect that the adjoint operators derived from these equations might be different as well, with some forms possibly being easier to compute than others. In this subsection, we present two alternative forms of the adjoint operator using the standard L_2 brackets in all three steps of the adjoint derivation, essentially reproducing the "standard" derivation of Section 3, but applying it to the stream-function and vorticity forms of the governing equation presented in (1) and (3). The subsequent three subsections discuss the effects of choosing alternative forms for the three brackets used in the adjoint derivation. A detailed summary of the inter-relationships between these options is presented in Section 4.5.

4.1.1. The standard adjoint derivation based on the vorticity formulation

We now proceed to minimize the original cost functional (5) by modeling the system evolution with the vorticity form (3) of the Kuramoto–Sivashinsky system. Specifically, we consider a cost functional written in the form

$$\mathscr{J}_\varphi(\varphi) = \frac{1}{2}\left\|\mathscr{H}\partial_x^{-1}w - y\right\|^2_{L_2(0,T;L_2(\Omega))}. \tag{22}$$

Note that, as $\partial_x^{-1}w = v$, $\mathscr{J}_\varphi$ is equivalent to $\mathscr{J}$, but depends on the control variable $\varphi = \partial_x\phi$, that is, $\mathscr{J}_\varphi(\partial_x\phi) = \mathscr{J}(\phi)$. The differential of this cost functional can now be expressed as

$$\mathscr{J}'_\varphi(\varphi;\varphi') = -\int_0^T\int_0^{2\pi}\left[\partial_x^{-1}\mathscr{H}^*\left(\mathscr{H}\,\partial_x^{-1}w - y\right)\right]w'\,\mathrm{d}x\,\mathrm{d}t, \tag{23}$$

where integration by parts was used to reveal explicit dependence of $\mathscr{J}'_\varphi$ on the vorticity perturbation $w'(\varphi;\varphi')$. The boundary terms obtained as a result of this transformation vanish due to periodicity of all the variables involved. The field w' satisfies the system obtained by linearizing (3)

$$\begin{cases} \mathscr{M}w' = 0, & x \in \Omega,\ \ t \in [0,T], \\ \partial_x^i w'(0,t) = \partial_x^i w'(2\pi,t), & t \in [0,T],\ \ i = 0,\dots,3, \\ w'(x,0) = \varphi', & x \in \Omega, \end{cases} \tag{24}$$

where the linear operator $\mathscr{M}$ is

$$\mathscr{M}w' \triangleq \partial_t w' + 4\partial_x^4 w' + \kappa\left(\partial_x^2 w' + 2ww' + \partial_x^{-1} w'\partial_x w + \partial_x^{-1} w\partial_x w'\right) \tag{25}$$

and $w(\varphi)$ is the solution of (3). By an identity of the same form as (14), that is,

$$\langle w^*, \mathscr{M}w'\rangle_{L_2(0,T;L_2(\Omega))} = \langle \mathscr{M}^* w^*, w'\rangle_{L_2(0,T;L_2(\Omega))} + b_{\mathscr{M}}, \tag{26}$$

it follows [cf. (15)] that

$$\begin{aligned} &\mathscr{M}^* w^* = -\partial_t w^* + 4\partial_x^4 w^* + \kappa\left[\partial_x^2 w^* + \partial_x^{-1}(w\partial_x w^*) - \partial_x^{-1} w\partial_x w^*\right], \\ &b_{\mathscr{M}} = \left[\int_0^{2\pi} w^* w'\,\mathrm{d}x\right]_{t=0}^{t=T} + \Big[\dots\Big]_{x=0}^{x=2\pi}. \end{aligned} \tag{27}$$

Making use of this adjoint operator, we define the *vorticity* adjoint system [cf. (16)] with

$$\begin{cases} \mathscr{M}^* w^* = -\partial_x^{-1}\mathscr{H}^*\left(\mathscr{H}\partial_x^{-1} w - y\right), & x \in \Omega,\ \ t \in [0,T], \\ \partial_x^i w^*(0,t) = \partial_x^i w^*(2\pi,t), & t \in [0,T],\ \ i = 0,\dots,3, \\ w^*(x,T) = 0, & x \in \Omega. \end{cases} \tag{28}$$

Defining the gradient $\nabla\mathscr{J}_\varphi$ such that

$$\mathscr{J}'_\varphi \triangleq \left(\nabla\mathscr{J}_\varphi, \varphi'\right)_{L_2(\Omega)}, \tag{29}$$

it follows by an analogous derivation that

$$\nabla\mathscr{J}_\varphi = \bar{w}^*|_{t=0}, \tag{30}$$

where the projection operator implied by the overbar is defined as in (18).

4.1.2. The standard adjoint derivation based on the streamfunction formulation

We may also minimize the cost functional (5) by modeling the system evolution with the streamfunction form (1) of the Kuramoto–Sivashinsky system. Specifically, we consider a cost functional written in the form

$$\mathscr{J}_\psi(\psi) = \frac{1}{2}\|\mathscr{H}\partial_x u - y\|^2_{L_2(0,T;L_2(\Omega))}. \tag{31}$$

If we restrict ψ to have zero mean mode, then it follows that $\psi = \partial_x^{-1}\phi$. Noting that $\partial_x u = v$, it is seen that $\mathscr{J}_\psi$ is equivalent to $\mathscr{J}$, that is, $\mathscr{J}_\psi(\partial_x^{-1}\phi) = \mathscr{J}(\phi)$. The differential of $\mathscr{J}_\psi$ is

$$\mathscr{J}'_\psi(\psi;\psi') = -\int_0^T \int_0^{2\pi} [\partial_x\mathscr{H}^*(\mathscr{H}\partial_x u - y)]u'\,\mathrm{d}x\,\mathrm{d}t, \tag{32}$$

where, by linearization of (1), it follows that $u'(\psi;\psi')$ is the solution of the system

$$\begin{cases} \mathscr{K}u' = 0, & x \in \Omega, \ \ t \in [0,T], \\ \partial_x^i u'(0,t) = \partial_x^i u'(2\pi,t), & t \in [0,T], \ \ i = 0,\ldots,3, \\ u'(x,0) = \psi', & x \in \Omega, \end{cases} \tag{33}$$

where the linear operator $\mathscr{K}$ is

$$\mathscr{K}u' \triangleq \partial_t u' + 4\partial_x^4 u' + \kappa\left[\partial_x^2 u' + (\partial_x u)(\partial_x u')\right] \tag{34}$$

and $u(\psi)$ is the solution of (1). By the identity

$$\langle u^*, \mathscr{K}u' \rangle_{L_2(0,T;L_2(\Omega))} = \langle \mathscr{K}^* u^*, u' \rangle_{L_2(0,T;L_2(\Omega))} + b_{\mathscr{K}}, \tag{35}$$

it follows that

$$\mathscr{K}^* u^* = -\partial_t u^* + 4\partial_x^4 u^* + \kappa\left[\partial_x^2 u^* - \partial_x(u^* \partial_x u)\right].$$

Making use of this adjoint operator, we define the *streamfunction* adjoint system with

$$\begin{cases} \mathscr{K}^* u^* = -\partial_x \mathscr{H}^*(\mathscr{H}\partial_x u - y), & x \in \Omega, \ \ t \in [0,T], \\ \partial_x^i u^*(0,t) = \partial_x^i u^*(2\pi,t), & t \in [0,T], \ \ i = 0,\ldots,3, \\ u^*(x,T) = 0, & x \in \Omega. \end{cases} \tag{36}$$

Defining the gradient $\nabla \mathscr{J}_\psi$ such that

$$\mathscr{J}'_\psi \triangleq \left(\nabla \mathscr{J}_\psi, \psi'\right)_{L_2(\Omega)}, \tag{37}$$

it follows that

$$\nabla \mathscr{J}_\psi = u^*|_{t=0}. \tag{38}$$

4.2. Targeting the cost functional

As indicated in (20a), the definition of the cost functional may be generalized by taking the H^q norm (rather than the L_2 norm) of the quantity of interest (in the present case, the measurement misfit [5]). By so doing, we can focus the cost functional on a particular range of length scales of interest. For example, taking the H^1 norm [see (20a)] of the misfit of the measurement y, we define [cf. (5)]

$$\mathscr{J}^{H^1}(\phi) = \|\mathscr{H}v - y\|^2_{L_2(0,T;H^1(\Omega))}. \tag{39}$$

It is straightforward to show that the only modification to the standard formulation of the adjoint analysis which results from this change in the cost functional is the RHS forcing of the evolution equation for the associated adjoint field, which now takes the form [cf. (16)]

$$\mathscr{L}^* v^* = -\partial_x^2 \mathscr{H}^*(\mathscr{H}v - y) = -\partial_x^2 f.$$

[5] We do not consider here the possibility of changing the quantity penalized in the cost functional. An approach in which the cost functional also includes a suitable norm of the control variable is known as Tikhonov regularization [29] and its connection to the present framework is discussed in Section 4.5.6.

Similarly, taking the H^{-1} norm of the misfit of the measurement y, we define

$$\mathscr{J}^{H^{-1}}(\phi) = \|\mathscr{H}v - y\|^2_{L_2(0,T;H^{-1}(\Omega))}. \tag{40}$$

The modification of the RHS forcing of the adjoint field in the standard formulation which results from this change in the cost functional is

$$\mathscr{L}^* v^* = -\partial_x^{-2}\mathscr{H}^*(\mathscr{H}v - y) = -\partial_x^{-2} f.$$

We remark that this regularization option may not be available when the observations $\mathscr{H}v$ are not a well-defined function of space (e.g., are defined pointwise in space). Also, if the measurement error η has the form of white noise, this may render f nondifferentiable and therefore the functional $\mathscr{J}^{H^1}$ may not be applicable.

4.3. Modifying the inner product in the adjoint identity

As indicated in (20b), the definition of the adjoint operator may be generalized by using an H^q inner product rather than the standard L_2 inner product. By so doing, we may determine the same gradient of the same cost functional as found by the standard adjoint framework, but do so via an adjoint system with a different energy spectrum which makes it more or (preferably) less difficult to compute in a numerical simulation.

4.3.1. An adjoint derivation with the H^1 inner product

We again proceed to minimize the original cost functional (5) by modeling the system evolution with the primitive form (2) of the Kuramoto–Sivashinsky system, but now derive the adjoint operator with an H^1 inner product via the identity

$$\langle v^{*,H^1}, \mathscr{L}v' \rangle_{L_2(0,T;H^1(\Omega))} = \langle \mathscr{L}^{*,H^1} v^{*,H^1}, v' \rangle_{L_2(0,T;H^1(\Omega))} + b_1 \tag{41}$$

from which it follows that

$$\begin{aligned} \mathscr{L}^{*,H^1} v^{*,H^1} &= -\partial_t v^{*,H^1} + 4\partial_x^4 v^{*,H^1} + \kappa\left[\partial_x^2 v^{*,H^1} - \partial_x^{-2}(v\partial_x^3 v^{*,H^1})\right], \\ b_1 &= \left[\int_0^{2\pi} \left(\partial_x v^{*,H^1}\right)\left(\partial_x v'\right) \mathrm{d}x\right]_{t=0}^{t=T} + \left[\cdots\right]_{x=0}^{x=2\pi}. \end{aligned} \tag{42}$$

Making use of this adjoint operator, we define an adjoint system with

$$\begin{cases} \mathscr{L}^{*,H^1} v^{*,H^1} = -\partial_x^{-2}\mathscr{H}^*(\mathscr{H}v - y) = -\partial_x^{-2} f, & x \in \Omega,\ \ t \in [0,T], \\ \partial_x^i v^{*,H^1}(0,t) = \partial_x^i v^{*,H^1}(2\pi,t), & t \in [0,T],\ \ i = 0,\ldots,3, \\ v^{*,H^1}(x,T) = 0, & x \in \Omega. \end{cases} \tag{43}$$

Note that the differential of the cost functional (9) may be written in a form consistent with the new inner product

$$\mathscr{J}'(\phi;\phi') = -\int_0^T \int_0^{2\pi} \partial_x\left[\partial_x^{-2}\mathscr{H}^*(\mathscr{H}v - y)\right]\partial_x v' \,\mathrm{d}x\,\mathrm{d}t = \left\langle -\partial_x^{-2}\mathscr{H}^*(\mathscr{H}v - y), v' \right\rangle_{L_2(0,T;H^1(\Omega))}.$$

Combining (10) and (43) with (41) and substituting the above expression, we obtain

$$\mathscr{J}'(\phi;\phi') = -\int_0^{2\pi} \partial_x^2 v^{*,H^1}\Big|_{t=0} \phi' \,\mathrm{d}x = \left(-\partial_x^2 v^{*,H^1}\Big|_{t=0}, \phi'\right)_{L_2(\Omega)}. \tag{44}$$

Using an L_2 inner product to extract the gradient as in (12), we may identify the gradient as

$$\nabla \mathscr{J} = -\partial_x^2 v^{*,H^1}\Big|_{t=0}. \tag{45}$$

Note that the gradient so defined has zero mean mode and thus lies in the space of feasible ϕ'. An analogous derivation using an inner product which incorporates derivatives with respect to the time variable (that is "H^1-in-time") is presented in Appendix A.

4.3.2. An adjoint derivation with the H^{-1} inner product

We again minimize the original cost functional (5) by modeling the system evolution with the primitive form (2) of the Kuramoto–Sivashinsky system, but now derive the adjoint operator with an H^{-1} inner product via the identity

$$\langle v^{*,H^{-1}}, \mathscr{L} v' \rangle_{L_2(0,T;H^{-1}(\Omega))} = \langle \mathscr{L}^{*,H^{-1}} v^{*,H^{-1}}, v' \rangle_{L_2(0,T;H^{-1}(\Omega))} + b_{-1} \tag{46}$$

from which it follows that

$$\begin{aligned} \mathscr{L}^{*,H^{-1}} v^{*,H^{-1}} &= -\partial_t v^{*,H^{-1}} + 4\partial_x^4 v^{*,H^{-1}} + \kappa\left[\partial_x^2 v^{*,H^{-1}} - \partial_x^2 (v \partial_x^{-1} v^{*,H^{-1}})\right], \\ b_{-1} &= \left[\int_0^{2\pi} \left(\partial_x^{-1} v^{*,H^{-1}}\right)\left(\partial_x^{-1} v'\right) \mathrm{d}x\right]_{t=0}^{t=T} + \left[\cdots\right]_{x=0}^{x=2\pi}. \end{aligned} \tag{47}$$

Making use of this adjoint operator, we define an adjoint system with

$$\begin{cases} \mathscr{L}^{*,H^{-1}} v^{*,H^{-1}} = -\partial_x^2 \mathscr{H}^*(\mathscr{H} v - y) = -\partial_x^2 f, & x \in \Omega, \ \ t \in [0,T], \\ \partial_x^i v^{*,H^{-1}}(0,t) = \partial_x^i v^{*,H^{-1}}(2\pi,t), & t \in [0,T], \ \ i = 0,\ldots,3, \\ v^{*,H^{-1}}(x,T) = 0, & x \in \Omega. \end{cases} \tag{48}$$

Note that the differential of the cost functional (9) may be written in a form consistent with the new inner product

$$\mathscr{J}'(\phi;\phi') = -\int_0^T \int_0^{2\pi} \partial_x^{-1}\left[\partial_x^2 \mathscr{H}^*(\mathscr{H} v - y)\right] \partial_x^{-1} v' \,\mathrm{d}x\,\mathrm{d}t = \left\langle -\partial_x^2 \mathscr{H}^*(\mathscr{H} v - y), v' \right\rangle_{L_2(0,T;H^{-1}(\Omega))}.$$

Combining (10) and (48) with (46) and substituting the above expression, we obtain

$$\mathscr{J}'(\phi;\phi') = -\int_0^{2\pi} \partial_x^{-2} v^{*,H^{-1}}\Big|_{t=0} \phi' \,\mathrm{d}x = \left(-\partial_x^{-2} v^{*,H^{-1}}\Big|_{t=0}, \phi'\right)_{L_2(\Omega)}. \tag{49}$$

Using the L_2 inner product (12) to extract the gradient, we may identify the gradient as

$$\nabla \mathscr{J} = -\partial_x^{-2} v^{*,H^{-1}}\Big|_{t=0}. \tag{50}$$

4.4. Preconditioning the gradient

As indicated in (20c), the definition of the gradient may be generalized by taking the H^q inner product (rather than the L_2 inner product) when extracting the gradient from the expression of the cost functional differential. By so doing, we may emphasize the importance of some length scales over others during the iterative gradient-based optimization procedure, a strategy commonly referred to as preconditioning. Note

again that (in a convex problem) the minimizer is not changed by such a procedure, though the gradients are significantly altered. For example, the cost functional differential (12) in the primitive formulation of the adjoint analysis may be rewritten to incorporate either an H^1 inner product or an H^{-1} inner product

$$\mathscr{J}'(\phi;\phi') \triangleq (\nabla^{H^1}\mathscr{J},\phi')_{H^1(\Omega)}, \quad \mathscr{J}'(\phi;\phi') \triangleq (\nabla^{H^{-1}}\mathscr{J},\phi')_{H^{-1}(\Omega)}. \tag{51}$$

By the definition of these inner products and the relation given in (17), it follows that the H^1 gradient, $\nabla^{H^1}\mathscr{J}$, and the H^{-1} gradient, $\nabla^{H^{-1}}\mathscr{J}$, may be identified as

$$\nabla^{H^1}\mathscr{J} = -\partial_x^{-2}\bar{v}^*\Big|_{t=0}, \quad \nabla^{H^{-1}}\mathscr{J} = -\partial_x^2 v^*\Big|_{t=0}. \tag{52}$$

We may thus extract different gradients from a given adjoint field calculation. Note that the H^1 gradient emphasizes the large length scales and the H^{-1} gradient emphasizes the small length scales. Conversely, as shown in (19), (45), and (50), we may also extract a given gradient from different adjoint field calculations.

We now explore the utility of the weighted linear combination of L_2, H^1, and H^{-1} inner products defined in (21) for preconditioning the gradient. For clarity, we first consider for this purpose the inner product $(z_1,z_2)_{W^{l_1,\infty}}$ which, as discussed previously, represents a linear combination of the L_2 and H^1 inner products. Returning to the standard formulation of the adjoint analysis, but extracting the gradient via this inner product, we obtain

$$\begin{aligned}\mathscr{J}' \triangleq (\nabla^{W^{l_1,\infty}}\mathscr{J},\phi')_{W^{l_1,\infty}} &= \frac{1}{1+l_1^2}\int_0^{2\pi}\left[\left(\nabla^{W^{l_1,\infty}}\mathscr{J}\right)\phi' + l_1^2\left(\partial_x\nabla^{W^{l_1,\infty}}\mathscr{J}\right)(\partial_x\phi')\right]\mathrm{d}x \\ &= \left(\frac{1}{1+l_1^2}\left[1-l_1^2\partial_x^2\right]\nabla^{W^{l_1,\infty}}\mathscr{J},\phi'\right)_{L_2(\Omega)} = \left(\bar{v}^*|_{t=0},\phi'\right)_{L_2(\Omega)}.\end{aligned} \tag{53}$$

We may thus identify the desired gradient $\nabla^{W^{l_1,\infty}}\mathscr{J}$ by solving the 1D Helmholtz equation

$$\begin{cases}\frac{1}{1+l_1^2}\left[1-l_1^2\partial_x^2\right]\nabla^{W^{l_1,\infty}}\mathscr{J} = \bar{v}^*\Big|_{t=0}, \\ \nabla^{W^{l_1,\infty}}\mathscr{J}(0) = \nabla^{W^{l_1,\infty}}J(2\pi).\end{cases} \tag{54}$$

The interpretation of the significance of this expression is clear in Fourier space. Using $[\hat{\cdot}]_k$ to denote the corresponding spatial Fourier coefficient at wavenumber k, it follows that

$$\left[\widehat{\nabla^{W^{l_1,\infty}}\mathscr{J}}\right]_k = \frac{k_1^2+1}{k_1^2}\mathscr{F}_{lp}(k)\left[\widehat{\bar{v}^*}|_{t=0}\right]_k, \tag{55}$$

where $\mathscr{F}_{lp}(k) \triangleq k_1^2/(k_1^2+k^2)$ is a low-pass filter (see Fig. 2(a)) with a cut-off wavenumber of $k_1 = 1/l_1$. A gradient defined with such a scale-dependent filter de-emphasizes the spatial wavenumbers greater than k_1 in the gradient-based optimization process. Note that taking $k_1 \to \infty$ recovers the standard L_2 gradient (which weights all wavenumbers equally), whereas taking $k_1 \to 0$ recovers the H^1 gradient. Note also that the inverse Laplacian is commonly used as a "smoother" in this type of problem. The inverse Helmholtz operator used to obtain the solution to (54) is a generalization of the inverse Laplacian, which is used to solve this system in the $l_1 \to \infty$ (that is, $k_1 \to 0$) limit. Both such operations may be used to obtain a "smooth" gradient even when the system is defined on a complicated domain in which Fourier analysis is not possible. Thus, this form of preconditioning has the effect of enforcing smoothness of the control. In Section 4.5.6 we will show that it can be used to stabilize Tikhonov regularization.

In a similar vein, we can extract the cost functional gradient using the inner product $(z_1,z_2)_{W^{0,l_{-1}}}$ which is a linear combination of the L_2 and H^{-1} terms. This yields

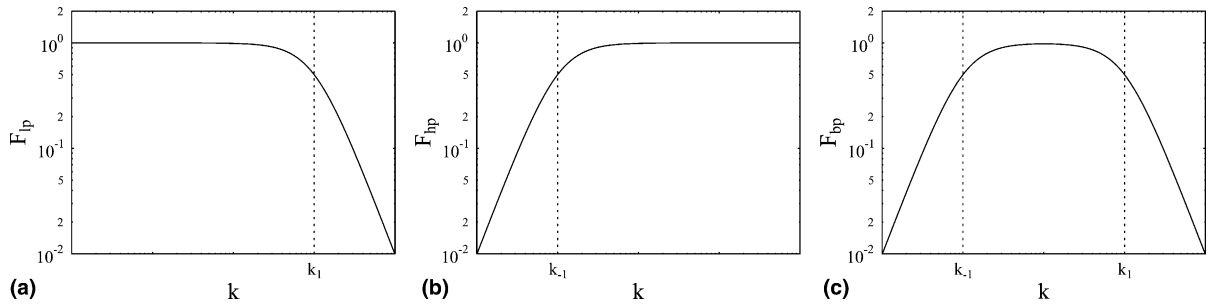

Fig. 2. Interpretation of the systems (54), (57) and (58) in Fourier space as (a) low-pass, (b) high-pass, and (c) band-pass filters which de-emphasize certain ranges of wavenumbers in the extraction of the gradient.

$$\mathscr{J}' \triangleq (\nabla^{W^{0,l_{-1}}}\mathscr{J},\phi')_{W^{0,l_{-1}}} = \frac{l_{-1}^2}{1+l_{-1}^2}\int_0^{2\pi}\left[(\nabla^{W^{0,l_{-1}}}\mathscr{J})\phi' + \frac{1}{l_{-1}^2}\left(\partial_x^{-1}\nabla^{W^{0,l_{-1}}}\mathscr{J}\right)\left(\partial_x^{-1}\phi'\right)\right]\mathrm{d}x$$
$$= \left(\frac{l_{-1}^2}{1+l_{-1}^2}\left[1-\frac{1}{l_{-1}^2}\partial_x^{-2}\right]\nabla^{W^{0,l_{-1}}}\mathscr{J},\phi'\right)_{L_2(\Omega)} = \left(\bar{v}^*|_{t=0},\phi'\right)_{L_2(\Omega)}. \tag{56}$$

We may thus identify the gradient $\nabla^{W^{0,l_{-1}}}\mathscr{J}$ by solving the system

$$\begin{cases} \frac{l_{-1}^2}{1+l_{-1}^2}\left[1-\frac{1}{l_{-1}^2}\partial_x^{-2}\right]\nabla^{W^{0,l_{-1}}}\mathscr{J} = \bar{v}^*\Big|_{t=0}, \\ \nabla^{W^{0,l_{-1}}}\mathscr{J}(0) = \nabla^{W^{0,l_{-1}}}J(2\pi). \end{cases} \tag{57}$$

Solution of this system has a clear interpretation in Fourier space

$$\left[\widehat{\nabla^{W^{0,l_{-1}}}\mathscr{J}}\right]_k = \left(k_{-1}^2+1\right)\mathscr{F}_{\mathrm{hp}}(k)\left[\widehat{\bar{v}^*}|_{t=0}\right]_k,$$

where $\mathscr{F}_{\mathrm{hp}}(k) \triangleq k^2/(k_{-1}^2+k^2)$ is a high-pass filter (see Fig. 2(b)) with a cut-off wavenumber of $k_{-1}=1/l_{-1}$. Thus defined scale-dependent filters de-emphasize spatial wavenumbers smaller than k_{-1}. Furthermore, we note that taking $k_{-1}\to 0$ recovers the standard L_2 gradient, whereas taking $k_{-1}\to\infty$ recovers the H^{-1} gradient.

We now consider a weighted linear combination of L_2, H^1 and H^{-1} inner products used to extract the gradient

$$\mathscr{J}' \triangleq (\nabla^{W^{l_1,l_{-1}}}\mathscr{J},\phi')_{W^{l_1,l_{-1}}} = \left(\frac{l_{-1}^2}{(1+l_1^2)(1+l_{-1}^2)}\left[1-\frac{l_1^2 l_{-1}^2}{l_1^2+l_{-1}^2}\partial_x^2-\frac{1}{l_1^2+l_{-1}^2}\partial_x^{-2}\right]\nabla^{W^{l_1,l_{-1}}}\mathscr{J},\phi'\right)_{L_2(\Omega)}$$
$$= \left(\bar{v}^*|_{t=0},\phi'\right)_{L_2(\Omega)}.$$

We may thus identify the desired gradient $\nabla^{W^{l_1,l_{-1}}}\mathscr{J}$ by solving the system

$$\begin{cases} \frac{l_{-1}^2}{(1+l_1^2)(1+l_{-1}^2)}\left[1-\frac{l_1^2 l_{-1}^2}{l_1^2+l_{-1}^2}\partial_x^2-\frac{1}{l_1^2+l_{-1}^2}\partial_x^{-2}\right]\nabla^{W^{l_1,l_{-1}}}\mathscr{J} = \bar{v}^*\Big|_{t=0}, \\ \nabla^{W^{l_1,l_{-1}}}\mathscr{J}(0) = \nabla^{W^{l_1,l_{-1}}}\mathscr{J}(2\pi). \end{cases} \tag{58}$$

Again, the interpretation of this expression is clear in Fourier space. Taking $k_1=1/l_1$ and $k_{-1}=1/l_{-1}$, it follows that

$$\left[\widehat{\nabla^{W^{l_1,l_{-1}}}\mathscr{J}}\right]_k = \frac{\left(k_1^2+k_{-1}^2\right)\left(k_1^2+1\right)\left(k_{-1}^2+1\right)}{k_1^4}\mathscr{F}_{\mathrm{bp}}(k)\left[\widehat{\tilde{v}^*}|_{t=0}\right]_k, \tag{59}$$

where $\mathscr{F}_{\mathrm{bp}}(k) \triangleq \mathscr{F}_{\mathrm{lp}}(k)\cdot\mathscr{F}_{\mathrm{hp}}(k)$ is a band-pass filter (see Fig. 2(c)) formed by the product of the low-pass filter (with a cut-off wavenumber of $k_1 = 1/l_1$) and a high-pass filter (with a cut-off wavenumber of $k_{-1} = 1/l_{-1}$). A band-pass filter of this sort is useful to employ when the optimization process in the multiscale system is designed to target "intermediate-scale" phenomena.

4.5. Relations between different regularization strategies

We now summarize the relations between the various alternatives in the framing of an adjoint analysis, as discussed in detail in the example considered above. We will first (in Section 4.5.1) show how adjoint operators corresponding to alternative forms of the evolution equation (Section 4.1) and alternative inner product used to define the adjoint identity (Section 4.3) are related to the primitive adjoint operator $\mathscr{L}^*$ determined in Section 3. We will then (in Section 4.5.2) discuss how the associated adjoint fields are related, and tabulate how any of three cost functional gradients sought may be determined from any of five alternative forms of the adjoint system. After brief discussions of two interesting special cases (in Sections 4.5.3 and 4.5.4), and an alternative method of deriving an adjoint analysis (in Section 4.5.5), we conclude the section (in Section 4.5.6) with a discussion of the relation between Tikhonov regularization and gradient preconditioning.

4.5.1. Relations between the various adjoint operators

Recall first that $w = \partial_x v$, $w' = \partial_x v'$, and thus, by (11) and (25), that $\partial_x \mathscr{L} z = \mathscr{M}\partial_x z$. By the identity (14), which defines $\mathscr{L}^*$, it thus follows (assuming all variables are periodic in x) that

$$\begin{aligned}
\langle w^*, \mathscr{M}w'\rangle_{L_2(0,T;L_2(\Omega))} &= \langle w^*, \partial_x\mathscr{L}v'\rangle_{L_2(0,T;L_2(\Omega))} = \langle -\partial_x w^*, \mathscr{L}v'\rangle_{L_2(0,T;L_2(\Omega))} \\
&= \left\langle \mathscr{L}^*(-\partial_x w^*), v'\right\rangle_{L_2(0,T;L_2(\Omega))} - \left[\int_0^{2\pi}(\partial_x w^*)v'\,\mathrm{d}x\right]_{t=0}^{t=T} \\
&= \left\langle \partial_x^{-1}\mathscr{L}^*(\partial_x w^*), \partial_x v'\right\rangle_{L_2(0,T;L_2(\Omega))} - \left[\int_0^{2\pi}(\partial_x w^*)v'\,\mathrm{d}x\right]_{t=0}^{t=T}.
\end{aligned}$$

Note that the above derivation computes the adjoint of a composition of operators, $\partial_x\mathscr{L}$, and the result is consistent with a general property of adjoint calculus, namely that $(\mathscr{T}_1\mathscr{T}_2)^* = \mathscr{T}_2^*\mathscr{T}_1^*$, where $\mathscr{T}_1$ and $\mathscr{T}_2$ are two linear operators. Note also that, in general, $\mathscr{T}_2^*\mathscr{T}_1^* \neq \mathscr{T}_1^*\mathscr{T}_2^*$. Noting (26), it follows that

$$\mathscr{M}^* z = \partial_x^{-1}\mathscr{L}^*(\partial_x z). \tag{60}$$

Using a similar approach, it is also straightforward to show that

$$\mathscr{K}^* z = \partial_x\mathscr{L}^*\left(\partial_x^{-1} z\right). \tag{61}$$

Similar relationships may be found for the adjoint operators derived using the H^1 and H^{-1} inner products. For example, it is easily seen (again assuming all variables are periodic in x) that

$$\begin{aligned}
\langle v^{*,H^1}, \mathscr{L}v'\rangle_{L_2(0,T;H^1(\Omega))} &= \left\langle \partial_x v^{*,H^1}, \partial_x\mathscr{L}v'\right\rangle_{L_2(0,T;L_2(\Omega))} = \left\langle -\partial_x^2 v^{*,H^1}, \mathscr{L}v'\right\rangle_{L_2(0,T;L_2(\Omega))} \\
&= \left\langle -\mathscr{L}^*\left(\partial_x^2 v^{*,H^1}\right), v'\right\rangle_{L_2(0,T;L_2(\Omega))} - \left[\int_0^{2\pi}\left(\partial_x^2 v^{*,H^1}\right)v'\,\mathrm{d}x\right]_{t=0}^{t=T} \\
&= \left\langle \partial_x^{-2}\mathscr{L}^*\left(\partial_x^2 v^{*,H^1}\right), v'\right\rangle_{L_2(0,T;H^1(\Omega))} - \left[\int_0^{2\pi}\left(\partial_x^2 v^{*,H^1}\right)v'\,\mathrm{d}x\right]_{t=0}^{t=T}.
\end{aligned}$$

Noting (41), it follows that

$$\mathscr{L}^{*,H^1} z = \partial_x^{-2} \mathscr{L}^* \left(\partial_x^2 z\right). \tag{62}$$

Using a similar approach, it is also straightforward to show that

$$\mathscr{L}^{*,H^{-1}} z = \partial_x^{2} \mathscr{L}^* \left(\partial_x^{-2} z\right). \tag{63}$$

4.5.2. Relations between the various adjoint fields

Substituting (60) into (28), it is seen that $-\partial_x w^*$ satisfies an equivalent set of equations as that defining v^* in (16). We may thus conclude that the primitive adjoint variable v^* and the vorticity adjoint variable w^* are related such that $v^* = -\partial_x w^*$. By (30), it thus follows that

$$\nabla \mathscr{J}_\varphi = -\partial_x^{-1} v^*\big|_{t=0}. \tag{64}$$

The quantity $\nabla \mathscr{J}_\varphi$ is simply the gradient of the cost functional $\mathscr{J}_\varphi(\varphi)$ in the space of φ, where the metric is defined by the L_2 inner product.

We now consider two gradient descent algorithms: one conducted in the space of φ and proceeding at each step in the direction $\nabla \mathscr{J}_\varphi$, and the other conducted in the space of ϕ and proceeding at each step in some direction s_ϕ. We then constrain s_ϕ such that the two descent algorithms are equivalent in the sense that $\varphi^{(n)} = \partial_x \phi^{(n)}$ for all iterations i. It follows that

$$\left.\begin{aligned} \varphi^{(n)} &= \varphi^{(n-1)} - \alpha \nabla \mathscr{J}_\varphi \\ \phi^{(n)} &= \phi^{(n-1)} - \alpha s_\phi \end{aligned}\right\} \quad \Rightarrow \quad s_\phi = \partial_x^{-1}(\nabla \mathscr{J}_\varphi),$$

that is, the corresponding descent direction in the space of ϕ is given by $s_\phi \triangleq \partial_x^{-1}(\nabla \mathscr{J}_\varphi)$. Combining this with (64) and (52), we obtain

$$s_\phi = -\partial_x^{-2} v^*\big|_{t=0} = \nabla^{H^1} \mathscr{J},$$

that is, gradient extraction via the L_2 inner product in the space of $\varphi = \partial_x \phi$ and gradient extraction via the H^1 inner product in the space of ϕ are equivalent. Using a similar approach, it is straightforward to show that gradient extraction via the L_2 inner product in the space of $\psi = \partial_x^{-1}\phi$ (where ψ is again restricted to have zero mean) and gradient extraction via the H^{-1} inner product in the space of ϕ are equivalent. Similar observations regarding gradient computations before and after a transformation of the independent variables in a system (in the finite-dimensional setting) were made by Dennis and Schnabel [35].

Noting the convenient form of the terms b_1 and b_{-1} in (41) and (46), it is seen that the derivation of the H^1 gradient [see (51)] follows naturally from the adjoint field defined with the H^1 inner product, and that the H^{-1} gradient follows naturally from the adjoint field defined with the H^{-1} inner product

$$\nabla^{H^1} \mathscr{J} = v^{*,H^1}\Big|_{t=0}, \quad \text{and} \quad \nabla^{H^{-1}} \mathscr{J} = v^{*,H^{-1}}\Big|_{t=0}.$$

In order to summarize the pattern that emerges from the application of the various regularization strategies to the formulation of adjoint-based analyses, we collect some of the significant relations between the various adjoint operators, the corresponding adjoint fields, and the different gradients in Table 1.

4.5.3. Special case: spatially uniform linearized systems

The relationships between the various alternative forms of adjoint analyses summarized above simplify greatly when the linearization of the governing evolution equation is spatially uniform (that is, it does not

Table 1
Summary of the principal relations resulting from application of various regularization strategies to the formulation of an adjoint-based optimization algorithm

Section introduced	Perturbation system	Inner product in adjoint identity	Adjoint system	$\nabla^{L_2}\mathscr{J}$	$\nabla^{H^1}\mathscr{J}$	$\nabla^{H^{-1}}\mathscr{J}$
Section 3	$\mathscr{L}v' = 0,$ $v'(0) = \phi$	$\langle\cdot,\cdot\rangle_{L_2(0,T;L_2(\Omega))}$	$\mathscr{L}^* v^* = f = \mathscr{H}^*(\mathscr{H}v - y),$ $v^*(T) = 0$	$v^*\big\vert_{t=0}$	$-\partial_x^{-2} v^*\big\vert_{t=0}$	$-\partial_x^2 v^*\big\vert_{t=0}$
Section 4.1.1	$\mathscr{M}w' = \partial_x \mathscr{L}v' = 0,$ $w'(0) = \varphi = \partial_x \phi$	$\langle\cdot,\cdot\rangle_{L_2(0,T;L_2(\Omega))}$	$\mathscr{M}^* w^* = \partial_x^{-1}\mathscr{L}^*(\partial_x w^*) = -\partial_x^{-1} f,$ $w^*(T) = 0$	$-\partial_x w^*\big\vert_{t=0}$	$\partial_x^{-1} w^*\big\vert_{t=0}$	$\partial_x^3 w^*\big\vert_{t=0}$
Section 4.1.2	$\mathscr{K}u' = \partial_x^{-1}\mathscr{L}v' = 0,$ $u'(0) = \psi = \partial_x^{-1}\phi$	$\langle\cdot,\cdot\rangle_{L_2(0,T;L_2(\Omega))}$	$\mathscr{K}^* u^* = \partial_x \mathscr{L}^*(\partial_x^{-1} u^*) = -\partial_x f,$ $u^*(T) = 0$	$-\partial_x^{-1} u^*\big\vert_{t=0}$	$\partial_x^{-3} u^*\big\vert_{t=0}$	$\partial_x u^*\big\vert_{t=0}$
Section 4.3.1	$\mathscr{L}v' = 0,$ $v'(0) = \phi$	$\langle\cdot,\cdot\rangle_{L_2(0,T;H^1(\Omega))}$	$\mathscr{L}^{*,H^1} v^{*,H^1} = \partial_x^{-2}\mathscr{L}^*\ (\partial_x^2 v^{*,H^1}) = \partial_x^{-2} f,$ $v^{*,H^1}(T) = 0$	$-\partial_x^2 v^{*,H^1}\big\vert_{t=0}$	$v^{*,H^1}\big\vert_{t=0}$	$\partial_x^4 v^{*,H^1}\big\vert_{t=0}$
Section 4.3.2	$\mathscr{L}v' = 0,$ $v'(0) = \phi$	$\langle\cdot,\cdot\rangle_{L_2(0,T;H^{-1}(\Omega))}$	$\mathscr{L}^{*,H^{-1}} v^{*,H^{-1}} = \partial_x^2\ \mathscr{L}^*(\partial_x^{-2}\ v^{*,H^{-1}}) = \partial_x^2 f,$ $v^{*,H^{-1}}(T) = 0$	$-\partial_x^{-2} v^{*,H^{-1}}\big\vert_{t=0}$	$\partial_x^{-4} v^{*,H^{-1}}\big\vert_{t=0}$	$v^{*,H^{-1}}\big\vert_{t=0}$

have spatially varying coefficients). This is the case, for instance, when the system (2) is linearized about the state $v = \text{constant}$. The perturbation operator for such a case will be denoted $\mathscr{L}_0$ and its adjoint $\mathscr{L}_0^*$; both may be obtained from (11) and (16) by setting $v = \text{constant}$. The reason for the simplification in this special case is that both $\mathscr{L}_0$ and $\mathscr{L}_0^*$ commute with ∂_x, and thus, by (60)–(62), $\mathscr{L}_0^* = \mathscr{K}_0^* = \mathscr{M}_0^* = \mathscr{L}_0^{*,H^1} = \mathscr{L}_0^{*,H^{-1}}$.

We now consider the system (16), with the operator $\mathscr{L}^*$ replaced by $\mathscr{L}_0^*$. Defining $v_1^* \triangleq -\partial_x^{-2} v^*$ and noting (52), the gradient of the cost functional $\mathscr{J}$ extracted using the H^1 inner product may be written $\nabla^{H^1}\mathscr{J}^{L_2} = v_1^*|_{t=0}$. Thus,

$$\mathscr{L}_0^* v^* = \mathscr{L}_0^*\big(-\partial_x^2 v_1^*\big) = -\partial_x^2 \mathscr{L}_0^* v_1^* = \mathscr{H}^*(\mathscr{H} v - y) \quad \Rightarrow \quad \mathscr{L}_0^* v_1^* = -\partial_x^{-2}\mathscr{H}^*(\mathscr{H} v - y).$$

By the discussion in Section 4.2, it is seen that v_1^* is exactly the adjoint variable used when the cost functional $\mathscr{J}^{H^{-1}}$ is minimized. It follows (in this special case only) that $\nabla^{H^1}\mathscr{J}^{L_2} = \nabla^{L_2}\mathscr{J}^{H^{-1}}$ and, similarly, that $\nabla^{H^{-1}}\mathscr{J}^{L_2} = \nabla^{L_2}\mathscr{J}^{H^1}$.

4.5.4. Special case: preconditioning the system with the nonlinear term removed

Interesting analytical insights regarding preconditioning can be obtained by considering optimization of the system (2) with the nonlinear term removed. To simplify calculations, in the reminder of this subsection we will also assume that the observation operator $\mathscr{H}$ has a particularly simple form of an identity operator, $\mathscr{H} = Id$, and there is no noise in the measurements, i.e., $\eta = 0$. This fairly contrived estimation problem serves simply to illustrate the behavior of a well-preconditioned minimization algorithm. With the above assumptions, we can analytically solve both the perturbation system (10) and the adjoint system (16) which, by (19), allows us to express the L_2 gradient in closed form as

$$(\widehat{\nabla^{L_2}\mathscr{J}})_k = \frac{\hat{\phi}_k^{\text{act}}}{2(4k^4 - \kappa k^2)}\left(1 - \mathrm{e}^{-2(4k^4-\kappa k^2)T}\right) = \hat{\mathscr{A}}_k \hat{\phi}_k^{\text{act}}, \tag{65}$$

where

$$\hat{\mathscr{A}}_k = \frac{1 - \mathrm{e}^{-2(4k^4-\kappa k^2)T}}{2(4k^4 - \kappa k^2)}$$

and ϕ^{act} is the actual initial condition that we seek to reconstruct. We see that, since the operator $\mathscr{A}$ is different from identity, the steepest descent direction given by (65) does not point to the minimizer ϕ^{act}, as shown schematically in Fig. 3(a). However, the form of relation (65) suggests that by extracting the gradient with an $\mathscr{A}^{-1/2}$ inner product defined as $(z_1, z_2)_{\mathscr{A}^{-1/2}} = (\mathscr{A}^{-1/2} z_1, \mathscr{A}^{-1/2} z_2)_{L_2(\Omega)}$ we obtain the gradient

$$\nabla^{\mathscr{A}^{-1/2}}\mathscr{J} = \phi^{\text{act}}, \tag{66}$$

which points directly to the minimizer (see Fig. 3(a)). Alternatively, this procedure can be understood as changing the metric in the space where optimization is performed from the metric induced by the L_2 inner product to the metric induced by the $\mathscr{A}^{-1/2}$ inner product in which the new gradient (66) represents the steepest descent direction (see Fig. 3(b)). In accordance with the discussion in Section 4.5.3, we notice that replacing the cost functional (5) with the cost functional $\mathscr{J}^{\mathscr{A}^{1/2}}(\phi) = \frac{1}{2}\|\mathscr{H} v - y\|^2_{L_2(0,T;\mathscr{A}^{1/2}(\Omega))}$ and extracting the gradient using an L_2 inner product also yields a descent direction which points directly to the minimizer ϕ^{act}. Thus, this example shows how modifying the cost functional and changing the gradient extraction procedure can improve conditioning of an optimization problem. In the present simple case we were able to obtain a perfectly preconditioned problem (i.e., with all Hessian eigenvalues λ_i equal). In more realistic cases when the nonlinear term in (2) is present and only incomplete measurements are available (i.e., $\mathscr{H} \neq Id$) analytical expressions analogous to (65) are not available; in such cases, the approaches described

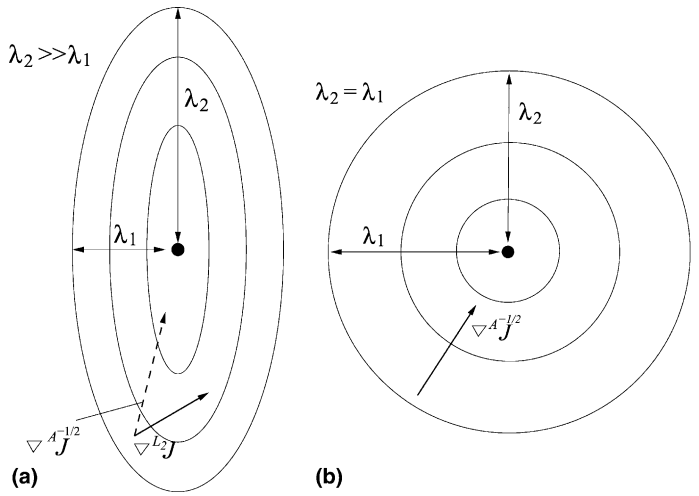

Fig. 3. Schematic showing isolines of the cost functional $\mathscr{J}$ in the metric induced by: (a) the L_2 inner product (the original problem), and (b) the $\mathscr{A}^{-1/2}$ inner product (the rescaled problem). Descent directions corresponding to gradient extraction performed with different inner products are also shown.

in Sections 4.2 and 4.4 appear to be natural strategies for improving the conditioning of the resulting optimization problem, even if they do not make it perfect.

4.5.5. Optimization derivations based on Lagrange multipliers

It is important to note that the four distinct regularization opportunities considered in this paper are also available when the evolution equation of the system is incorporated into the optimization problem with a Lagrange multiplier approach. In such derivations, the cost functional selected is first augmented with a selected form of the inner product (cf. Section 4.3) of a Lagrange multiplier with a selected form of the governing equation (cf. Section 4.1). This augmented cost functional is then minimized with respect to both the chosen control variable and the Lagrange multiplier, often using a gradient-based strategy using a selected form of an inner product to define the gradient. This setting effectively renders the optimization problem as "unconstrained", and the Lagrange multiplier itself turns out to be equivalent to the adjoint field used in the present derivation. In derivations based on such Lagrange multiplier techniques, all four of the regularization opportunities outlined in this paper are still available and may be selected to achieve the desired regularizing effect.

4.5.6. Tikhonov regularization in the gradient optimization setting

One common technique to regularize an optimization problem is to modify the cost functional (see Section 4.2) by adding a term which explicitly penalizes the irregularity of the control ϕ. This approach is commonly referred to as Tikhonov regularization (see, e.g. [10–12,29]). When applied to the data assimilation problem formulated in Section 3, this results in a new cost functional

$$\mathscr{J}_r(\phi) = \mathscr{J}(\phi) + \ell_r^2 \|\phi\|^2_{H^r(\Omega)}, \tag{67}$$

where ℓ_r and $r > 0$ are constants characterizing the degree and form of the regularization. The differential of this functional is given by

$$\mathscr{J}'_r(\phi;\phi') = \mathscr{J}'(\phi;\phi') + \ell_r^2 \int_0^{2\pi} \left(\partial_x^r \phi\right)\left(\partial_x^r \phi'\right) \mathrm{d}x,$$

from which we may extract the L_2 gradient of the functional as

$$\widehat{\nabla \mathscr{J}}_r = \widehat{\nabla \mathscr{J}} + \ell_r^2 k^{2r} \hat{\phi}, \tag{68}$$

where $\nabla \mathscr{J}$ may be determined as in Section 3 and for convenience we adopt the Fourier-space representation.

The optimization problem which we are attempting to solve may be written as $\widehat{\nabla \mathscr{J}}_r(\phi) = 0$. Further, the gradient-based optimization strategy which we have employed to solve this problem may be interpreted as a method to find the stationary solution of the following system, which evolves in the artificial "pseudo-time" coordinate τ

$$\begin{cases} \frac{\mathrm{d}\hat{\phi}}{\mathrm{d}\tau} = -\widehat{\nabla \mathscr{J}}_r = -\widehat{\nabla \mathscr{J}} - \ell_r^2 k^{2r} \hat{\phi} \quad \text{on } \tau \in (0, \infty), \\ \hat{\phi} = \hat{\phi}_0 \quad \text{at } \tau = 0. \end{cases} \tag{69}$$

Effectively, we are attempting to march this artificial system as rapidly as possible to the steady state characterized by $(\mathrm{d}\hat{\phi}/\mathrm{d}\tau) \approx 0$. Time accuracy during this artificial march is not required. This interpretation facilitates solution of the optimization problem at hand by adopting a variety of different time-discretization strategies applied to (69). Due to complexity involved in its evaluation (employing both forward and adjoint simulations), the term $\widehat{\nabla \mathscr{J}}$ must be calculated explicitly. However, the term $\ell_r^2 k^{2r} \hat{\phi}$ is easily handled with a variety of either implicit or explicit treatments. This leads to many possible forms of the optimization algorithm, including:

1. explicit (Euler) treatment of the penalty term

$$\hat{\phi}^{(n+1)} = \hat{\phi}^{(n)} - \Delta\tau (\widehat{\nabla \mathscr{J}}_r)^{(n)} = -\Delta\tau (\widehat{\nabla \mathscr{J}})^{(n)} + \left[1 - \Delta\tau \ell_r^2 k^{2r}\right] \hat{\phi}^{(n)},$$

2. implicit (Cranck–Nicholson) treatment of the penalty term

$$\hat{\phi}^{(n+1)} = \hat{\phi}^{(n)} - \Delta\tau \left[(\widehat{\nabla \mathscr{J}})^{(n)} + \ell_r^2 k^{2r} \left(\hat{\phi}^{(n)} + \hat{\phi}^{(n+1)} \right) \Big/ 2 \right]$$

$$= -\frac{\Delta\tau}{1 + \frac{1}{2}\Delta\tau \ell_r^2 k^{2r}} (\widehat{\nabla \mathscr{J}})^{(n)} + \left[\frac{1 - \frac{1}{2}\Delta\tau \ell_r^2 k^{2r}}{1 + \frac{1}{2}\Delta\tau \ell_r^2 k^{2r}} \right] \hat{\phi}^{(n)} = \hat{\phi}^{(n)} - \frac{\Delta\tau}{1 + \frac{1}{2}\Delta\tau \ell_r^2 k^{2r}} (\widehat{\nabla \mathscr{J}}_r)^{(n)},$$

where $\Delta\tau$ is some discrete stepsize in the pseudo-time coordinate τ. Taking $\Delta\tau =$ constant results in what is sometimes referred to as Landweber iteration (see, e.g. [10]), and is often the approach most amenable to numerical analysis. In practice, however, it is usually more efficient to adjust the stepsize $\Delta\tau$ at every iteration in order to minimize $\mathscr{J}_r$. Note that the explicit method #1 suffers from a stability constraint $\Delta\tau \leqslant 2\ell_r^{-2} k_{\max}^{-2r}$ which is reminiscent of a CFL condition and, if violated, will result in an unstable explicit march and amplification of the small scales of the field $\hat{\phi}$. In practice, method #1 is therefore generally not preferred. On the other hand, no such restriction applies to method #2. Furthermore, we observe that the semi-implicit method #2 may in fact be regarded as an explicit approach utilizing the cost functional $\mathscr{J}_r(\phi)$ and a smoothed gradient extracted with the inner product $(z_1, z_2)_{L_2(\Omega)} + (\ell_r^2 \Delta\tau/2)(z_1, z_2)_{H^r(\Omega)}$ (see Section 4.4). We thus see that adoption of a suitable gradient extraction strategy in the context of Tikhonov regularization may help bypass restrictive limitations on the length of the step $\Delta\tau$. As described above, the semi-implicit variant of Tikhonov regularization may in fact be viewed as a special case of the regularization framework proposed in the present study, incorporating appropriate forms of the cost functional and the inner product used to extract the gradient in an explicit optimization procedure.

5. Computational examples

In this section, we provide a few computational examples to illustrate how the different brackets used in the framing of an adjoint analysis may be used to affect the computational accuracy and the rate of convergence of a numerical optimization algorithm. Based on this analysis, we then propose a family of promising multiscale preconditioning approaches that improve the convergence of the state reconstruction problem highlighted in the previous two sections. Note that the present computational study is by no means meant to be exhaustive, but simply to indicate the utility of pursuing the various regularization opportunities outlined previously. Many interesting related questions are left to be characterized numerically in future work.

In the PDE setting, the descent direction determined via an adjoint analysis depends solely on the choice of the norm in the cost function (see Section 4.2) and the inner product used to extract the gradient (see Section 4.4). The choice of the form of the evolution equation (see Section 4.1) and the inner product used to define the adjoint identity (see Section 4.3) affect only *how* the desired gradient of the cost functional selected is determined.

The specific problem considered in the data assimilation results reported here is obtained by setting $\kappa = 4000$ in (2). This rather high value for κ insures the system under consideration exhibits chaotic multiscale dynamics. The peak of the energy spectrum of the system is generally between $k = 20$ and $k = 25$ and, for higher wavenumbers, the spectrum rolls off rapidly after that. Around 22–23 peaks may usually be counted in the domain Ω at any given time. A typical numerical simulation of this system is shown in Fig. 4. The initial condition (selected on the chaotic attractor of the system) which we will seek to reconstruct, based solely on measurements of the system on $[0, T]$, is that shown in Fig. 4(a). The length of the optimization horizon T used, which corresponds to about 300 time steps, is sufficient to demonstrate significant

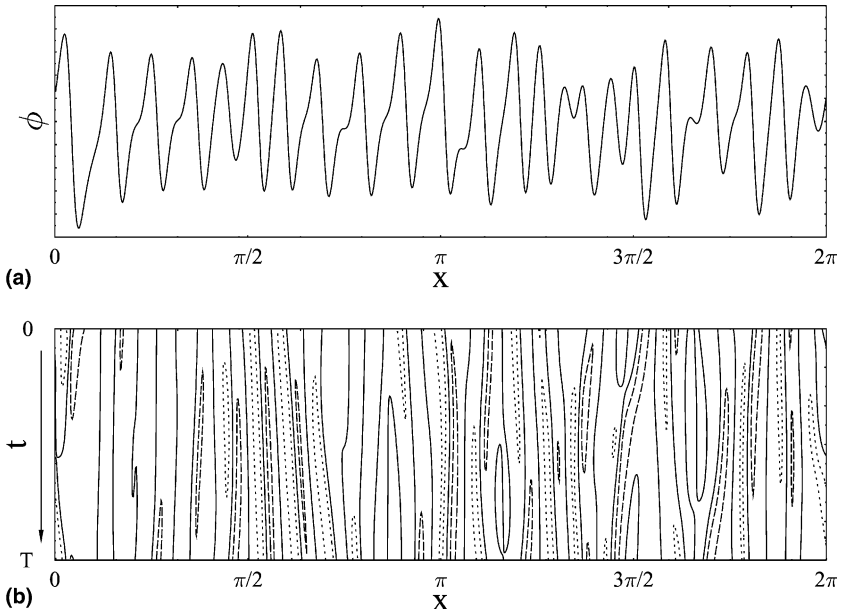

Fig. 4. Dynamics of the Kuramoto–Sivashinsky system for $\kappa = 4000$: (a) initial condition $v(0) = \phi$ (chosen on the chaotic attractor of the system), and (b) spatio-temporal evolution of the system (visualized are the zero (solid), several positive (dotted) and negative (dashed) isocontours of v in the space–time plane).

dynamics of the chaotic system, as illustrated in Fig. 4(b). The observation operator $\mathscr{H}$ selected represents observation of the real part of the first 50 Fourier coefficients (i.e., the first 50 coefficients of the cosine decomposition) of the primitive variable v [that is, $\Lambda_r = \{1, \ldots, 50\}$ in (7)]. We will assume for the purpose of this discussion that our measurements are not corrupted by noise [that is, $\eta = 0$ in (5)]. The initial guess for the initial conditions, $\phi^{(0)}$, will be taken to be zero in all optimizations attempted. These several choices make the (admittedly contrived) state reconstruction problem studied here difficult, yet still solvable in a reasonable number of iterations. This problem thus provides a tractable 1D multiscale chaotic testbed which is useful in quantifying the effectiveness of the various regularization strategies proposed. Extension of these strategies from the present 1D model problem to 2D and 3D systems of engineering relevance are straightforward – a few such extensions of particular interest are discussed briefly in Section 6.

In the present work, the state and adjoint systems are both solved in the well-resolved setting (on 1024 grid points) using a dealiased pseudo-spectral method. Time advancement is performed using an RK3 scheme on the nonlinear term and a generalized trapezoidal method with $\theta = 5/8$ (see [36]) on the linear terms. Gradient iterations are carried out using the Polak–Ribiere version of the Conjugate Gradient (CG) method (see, e.g. [34]). The "momentum" term in the CG method is calculated using a standard L_2 inner product, and is reset to zero every 20 iterations. Minimization in the descent direction is performed using Brendt's method [37] at each iteration. A gradient method has been selected for the optimization rather than a quasi-Newton method (which is an attractive alternative) in order to provide a simple environment for comparison of the different adjoint formulations.

5.1. Numerical characterization of the adjoint systems obtained with alternative formulations

In Section 4.5.2 we concluded that any given cost functional gradient can be extracted from any adjoint field calculation. In the present section we characterize certain numerical properties of the different adjoint operators derived in Sections 4.1 and 4.3. In Fig. 5 we compare the energy spectra of the adjoint variable at $t = 0$ (i.e., after the backward-time adjoint march) at the 100th iteration obtained with adjoint systems corresponding to different forms of the evolution equation and different forms of the inner product employed in the adjoint identity. The energy spectrum is defined as $E(k) = \hat{z}_k \cdot \mathrm{conj}(\hat{z}_k)$, where $\hat{z}_k$ is the Fourier transform of an appropriate adjoint variable and $\mathrm{conj}(\hat{z}_k)$ denotes the complex conjugate of $\hat{z}_k$. For

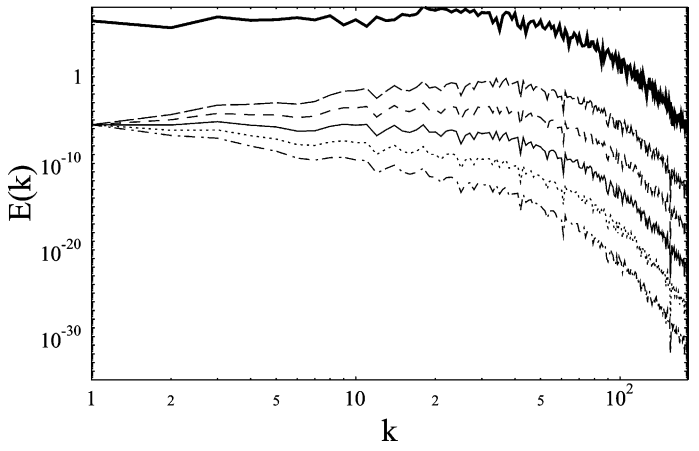

Fig. 5. Energy spectra of (thick solid) a typical solution of the Kuramoto–Sivashinsky equation (2) and of five different definitions of the corresponding adjoint system: (long dashed) $v^{*,H^{-1}}$, as defined in (48), (short dashed) u^*, as defined in (36), (thin solid) v^*, as defined in (16), (dotted) w^*, as defined in (28), and (dot-dashed) v^{*,H^1}, as defined in (43).

comparison, we also show the energy spectrum of a typical solution $v(t)$ of the Kuramoto–Sivashinsky system (2) on the chaotic attractor. Note that the decay rates of the spectra are consistent with the relations shown in columns 5–7 of Table 1. Specifically, we see that there is a difference of the factor k^2 in the roll-off of the spectra between every two consecutive variables $v^{*,H^{-1}}$, u^*, v^*, w^*, and v^{*,H^1}. Despite this fact, however, close inspection of the systems (48), (36), (16), (28), and (43) shows that discretization of the corresponding adjoint variables with a given cut-off wavenumber $k_{\max}$ retains in the resolved modes precisely the same information in each case. This can be visualized by drawing a vertical line at $k = k_{\max}$ in Fig. 5 and retaining the modes to the left of this line only. The resolved modes carry the same information, but have different weights in the Fourier-space representation. Thus, extraction of a given cost functional gradient from different adjoint variables may be regarded as readjusting these different weights so that they are consistent with the definition of the gradient sought.

Regardless of the specific approximation scheme employed, a spatial discretization $\mathbb{L}^*$ of the primitive adjoint operator $\mathscr{L}^*$ can be represented as

$$\mathbb{L}^*\mathbf{v}^* = -\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{v}^* + (\mathbb{B} + \mathbb{D})\mathbf{v}^*, \tag{70}$$

where $\mathbf{v}^*$ is the corresponding discretization of the adjoint variable v^*, and $\mathbb{B}$ and $\mathbb{D}$ are matrices representing discretizations of the operators $-\kappa v\partial_x$ and $4\partial_x^4 + \kappa\partial_x^2$, respectively. Defining now a transformation matrix $\mathbb{S}$ such that

$$[\mathbb{S}]_{ij} = \begin{cases} i, & i = j \\ 0, & i \neq j \end{cases}, \quad i, j = 1, \ldots, k_{\max},$$

we can express the spatial discretizations of the operators $\mathscr{L}^{*,H^{-1}}$, $\mathscr{K}^*$, $\mathscr{M}^*$, and $\mathscr{L}^{*,H^1}$ as follows:

$$\mathbb{L}^{*,H^{-1}}\mathbf{v}^{*,H^{-1}} = -\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{v}^{*,H^{-1}} + (\mathbb{S}^2\mathbb{B}\mathbb{S}^{-2} + \mathbb{D})\mathbf{v}^{*,H^{-1}}, \tag{71a}$$

$$\mathbb{K}^*\mathbf{u}^* = -\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{u}^* + (\mathbb{S}\mathbb{B}\mathbb{S}^{-1} + \mathbb{D})\mathbf{u}^*, \tag{71b}$$

$$\mathbb{M}^*\mathbf{w}^* = -\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{w}^* + (\mathbb{S}^{-1}\mathbb{B}\mathbb{S} + \mathbb{D})\mathbf{w}^*, \tag{71c}$$

$$\mathbb{L}^{*,H^1}\mathbf{v}^{*,H^1} = -\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{v}^{*,H^1} + (\mathbb{S}^{-2}\mathbb{B}\mathbb{S}^2 + \mathbb{D})\mathbf{v}^{*,H^1}. \tag{71d}$$

We note that, for any arbitrary square matrix $\mathbb{T}$, $\mathbb{T}$ and $\mathbb{S}^{-n}\mathbb{T}\mathbb{S}^n$, $n \in \ldots, -2, -1, 0, 1, 2, \ldots$, have exactly the same eigenvalues. Relating this fact to explicit time-integration of systems (70) and (71a)–(71d), we observe that the value of the CFL number is the same for all of these systems. This allows us to conclude that truncation errors related to spatial discretization and stability restrictions concerning explicit time-integration affect in the same way the adjoint systems obtained based on different forms of the evolution equation and different inner products used to define the adjoint identity.

On the other hand, each of the adjoint systems (48), (36), (16), (28), and (43) will be to a different degree prone to numerical precision (round-off) errors. This fact can be appreciated by drawing in Fig. 5 two *horizontal* lines: one for a large value, and the other for a small value of the ordinate, and retaining Fourier modes only with amplitudes between the values corresponding to the two lines. We observe that, depending on the choice of the lower and upper limit, the resulting loss of information will be different for the different adjoint variables as a result of the finite-precision arithmetic of the numerical calculation.

5.2. Optimization

As defined in (40), (5), (39), the three cost functionals $\mathscr{J}^{H^{-1}}$, $\mathscr{J}$, and $\mathscr{J}^{H^1}$ effectively measure the misfit of the model with the observed measurements with particular focus, respectively, on the large length scales, on all length scales, and on the small length scales. By large and small length-scales we mean the ranges of length-scales approaching the largest and the smallest length-scales resolved in the discrete representation of the system. In the present study the largest length-scale, equal to the size of the computational domain, is 2π, whereas the smallest resolved length-scale is approximately 10^{-3}. In the discussion of the computational results, by large scales we will mean length-scales from the interval $[10^0, 2\pi]$, by intermediate length-scales from the interval $[10^{-2}, 10^0]$, and by small length-scales from the interval $[10^{-3}, 10^{-2}]$. In this section, we will consider optimizations based on the minimization of all three of these cost functionals. To perform the optimizations, we will consider gradients extracted using the $W^{l_1,\infty}$ and $W^{0,l_{-1}}$ inner products, as defined in (53) and (56), for various values of l_1 and l_{-1}. Recall that the $W^{l_1,\infty}$ inner product reduces to the L_2 inner product in the $l_1 \to 0$ limit, and to the H^1 inner product in the $l_1 \to \infty$ limit. To simplify the notation, the different cases considered in this section will be referred to using a shorthand notation $\{z_1, z_2\}$ to characterize the spatial component of the brackets Ψ_1 and Ψ_3. For example, $\{L_2, L_2\}$ will be used denote the standard (L_2-based) formulation discussed in Section 3, whereas $\{H^1, W^{l_1,\infty}\}$ will be used to denote the formulation derived from the H^1 cost functional (39) together with the $W^{l_1,\infty}$ inner product used to define the gradient.

For the simulations presented in this section, to bypass further consideration of the numerical resolution issues discussed in the previous section, we will continue to use fine resolution, discretizing the system on 1024 grid points (that is, 684 degrees of freedom after dealiasing). Thus, as summarized in Table 1, we may determine the gradient sought via appropriate use of any of these definitions of the adjoint operator. For simplicity, all calculations discussed in the present section are performed using just the primitive adjoint Eq. (15), which is based on the primitive form of the Kuramoto–Sivashinsky system and the standard L_2 inner product for adjoint definition. This allows us to focus our attention in this section on the effects of modifying the brackets Ψ_1 and Ψ_3.

5.2.1. Analysis after one iteration

We first analyze the effect of the choice of Ψ_1 and Ψ_3 after just one iteration. The reason for focusing on the first iteration is that we intend to compare different gradients, and such a comparison is meaningful only when the state at which they are evaluated is the same for all gradients. The progress made towards the minimum on the large length scales, over all length scales, and on the small length scales will be assessed based on the reduction of $\mathscr{J}^{H^{-1}}$, $\mathscr{J}$, and $\mathscr{J}^{H^1}$, respectively, regardless of which cost functional is actually minimized in the case considered.

We begin by comparing the shapes of the gradients themselves in the case $\{L_2, W^{l_1,\infty}\}$ for different values of l_1 (Fig. 6). Note that the various choices for l_1 which have been used result in substantially different gradients, and that, as l_1 increases, the gradients $\nabla^{W^{l_1,\infty}}\mathscr{J}$ become significantly smoother (that is, as l_1 increases, the energy in the gradient field rolls off more rapidly with wavenumber). However, it is difficult to determine visually which of the gradients best captures the actual initial condition.

In Fig. 7, we present values of the functionals $\mathscr{J}^{H^{-1}}$, $\mathscr{J}$, and $\mathscr{J}^{H^1}$ obtained after the first iteration in all the cases considered as a function of the lengths l_1 and l_{-1} which parameterize the inner products used in gradient extraction. Note in all the cases that the three functionals vary smoothly with l_1 and l_{-1} revealing similar trends. The top plot in Fig. 7(a) illustrates the effect on $\mathscr{J}^{H^{-1}}$ when a control strategy is used which targets $\mathscr{J}^{H^{-1}}$. Similarly, the bottom plot in Fig. 7(c) illustrates the effect on $\mathscr{J}^{H^1}$ when a control strategy is used which targets $\mathscr{J}^{H^1}$. Note in both cases that, after a single iteration, the functional targeted by the control algorithm is reduced substantially. In Fig. 7(a) we see that optimization strategies targeting large scales may also work well at intermediate and small scales, provided the $W^{0,l_{-1}}$ inner product with a suitable

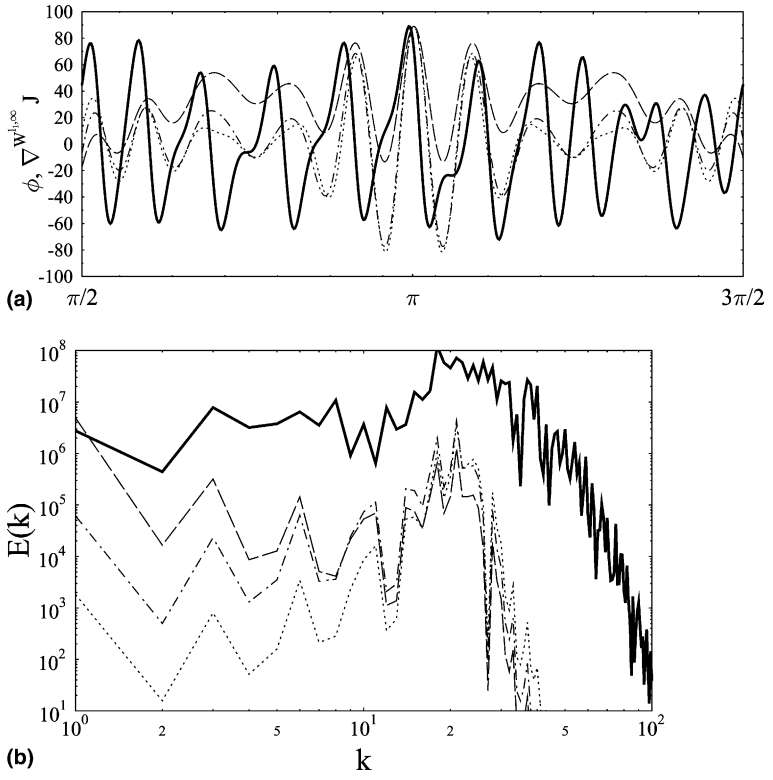

Fig. 6. (a) The shape of the gradients $\nabla^{W^{l_1,\infty}}\mathcal{J}$ in physical space, normalized by their peak value on the subinterval shown. (b) The energy contained in the gradient field as a function of spatial wavenumber. The cases considered are: (dashed) $l_1 = 0.5$, (dot-dashed) $l_1 = 0.1$, and (dotted) $l_1 = 0$. For comparison, the solid lines depict the actual initial condition sought.

value of l_{-1} is selected for gradient extraction. Likewise, strategies targeting all scales (Fig. 7(b)) work best in conjunction with the inner product $W^{l_1,\infty}$ selected for gradient extraction. On the contrary, strategies targeting small scales (Fig. 7(c)) reveal a rather weak dependence on the form of the inner product used for gradient extraction and, moreover, their performance is degraded on the larger scales. We note that when either $\mathcal{J}$ and $\mathcal{J}^{H^{-1}}$ is minimized, a clever choice of an inner product for gradient extraction can render an optimization strategy effective also on length scales other than those explicitly targeted by the cost functional. We thus see that, when posing an optimization problem of this sort, it is useful to select appropriate definitions of both the cost functional (by appropriate selection of Ψ_1) and its gradient (by appropriate selection of Ψ_3) in order to tune the performance on the length scales of interest.

5.2.2. Analysis after many iterations

We now analyze the effect of the choice of Ψ_1 and Ψ_3 after 100 iterations are performed. Several cases were run; for brevity, the following cases of particular interest are reported here:

(1) $\{H^{-1}, L_2\}$,
(2) $\{H^{-1}, W^{0,l_{-1}}\}$ with $l_{-1} = 0.2$,
(3) $\{H^{-1}, W^{0,l_{-1}}\}$ with $l_{-1}^{(n)} = l_{-1}^{(0)}\zeta^n$, where $l_{-1}^{(0)} = 0.2$ and $\zeta = 0.95$,
(4) $\{H^{-1}, W^{0,l_{-1}}\}$ with $l_{-1}^{(n)} = l_{-1}^{(0)}\zeta^n$, where $l_{-1}^{(0)} = 0.2$ and $\zeta = 1.05$,
(5) $\{L_2, L_2\}$,
(6) $\{L_2, W^{l_1,\infty}\}$ with $l_1 = 0.1$,

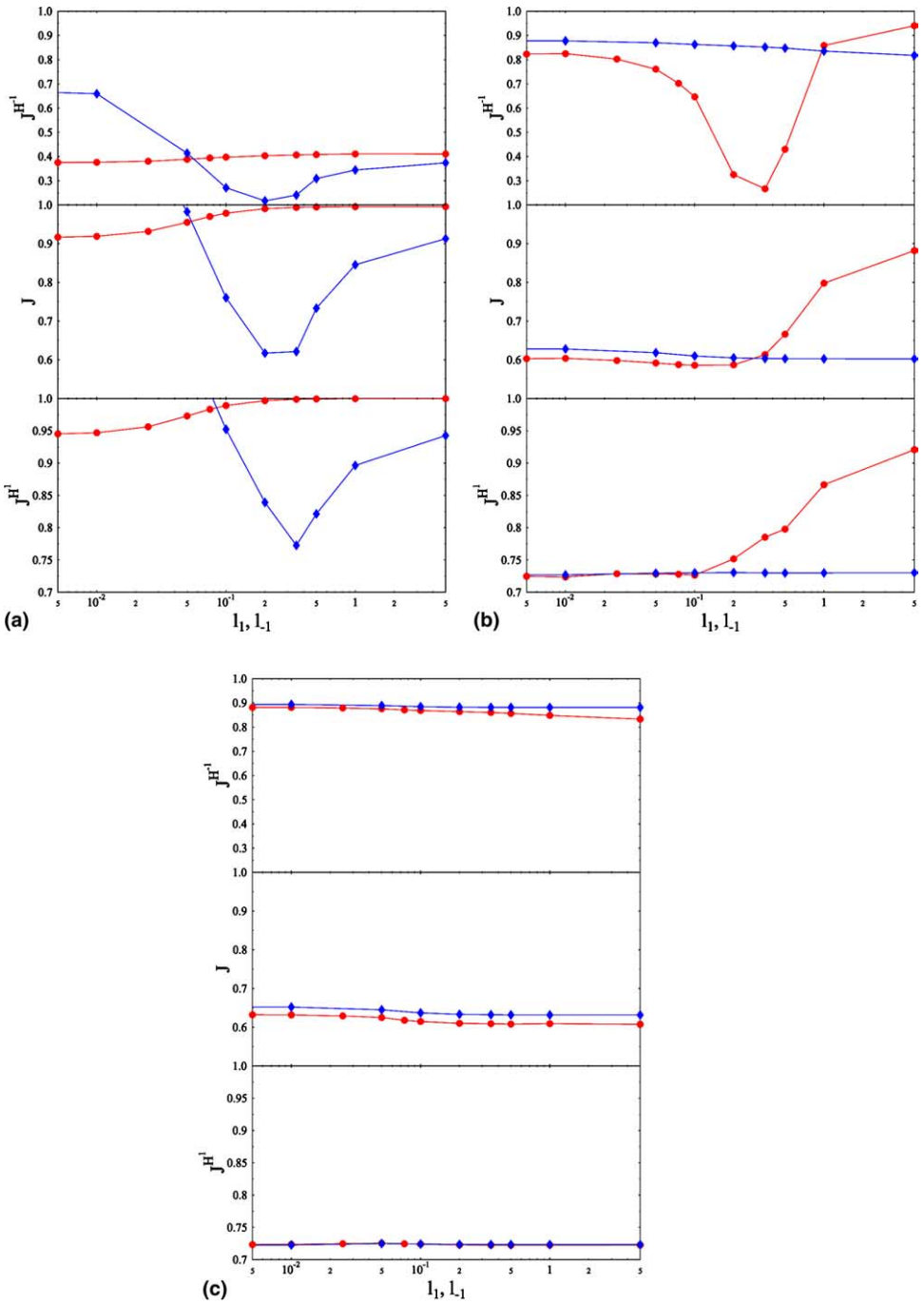

Fig. 7. Dependence of the functionals $\mathcal{J}^{H^{-1}}$, $\mathcal{J}$, and $\mathcal{J}^{H^1}$ on the lengths l_1 and l_{-1} parameterizing the inner products $W^{l_1,\infty}$ (circles) and $W^{0,l_{-1}}$ (diamonds) used to extract the gradients during the first iteration. In all figures, the vertical axis is normalized by the value of the corresponding cost functional for the initial guess $\phi^{(0)}$ (that is, before the first iteration). Thus, a cost functional value of 0.6 in the above figures implies a 40% reduction of the corresponding cost functional after one iteration. (a) Minimizing $\mathcal{J}^{H^{-1}}$; (b) minimizing $\mathcal{J}$; (c) minimizing $\mathcal{J}^{H^1}$.

(7) $\{L_2, W^{l_1,\infty}\}$ with $l_1^{(n)} = l_1^{(0)}\zeta^n$, where $l_1^{(0)} = 0.1$ and $\zeta = 2/3$,
(8) $\{L_2, W^{l_1,\infty}\}$ with $l_1^{(n)} = l_1^{(0)}\zeta^n$, where $l_1^{(0)} = 0.1$ and $\zeta = 3/2$.

The cases #3,4,7,8 will be referred to as *multiscale preconditioning* approaches. In these strategies gradients are extracted with the inner products $W^{l_1,\infty}$ and $W^{0,l_{-1}}$ in which l_1 and l_{-1} vary monotonically with the iteration number; cases of l_1 and l_{-1} both decreasing and increasing are considered. Gentle variation of $l_1^{(n)}$ and $l_{-1}^{(n)}$ thus provides a convenient "knob" controlling the cut-off length scales as a function of the iteration number n. This approach may be regarded as a *multiscale* version of the *variable preconditioning* method discussed in the context of finite-dimensional linear systems by Axelsson [38]. Appropriate values of $l_1^{(0)}$, $l_{-1}^{(0)}$ and ζ for the present system were found by trial and error.

Note that, when computing the descent direction at every iteration of the present conjugate gradient descent algorithm, we need to evaluate a "momentum" term formed by a ratio of inner products of the recently calculated gradients. Though there is some discussion of this issue in the literature, there appears to be no commonly accepted strategy for selecting the inner product to use to calculate the momentum term when a variable preconditioning strategy is employed. We have used simple L_2 inner products to evaluate the momentum term in the present work. Other strategies were also tried, including the use of inner products in this calculation that varied from one iteration to the next. Unfortunately, none of these strategies were found to significantly accelerate convergence.

In Fig. 8 we show the reduction of the three metrics $\mathscr{J}^{H^{-1}}$, $\mathscr{J}$, and $\mathscr{J}^{H^1}$ in the eight cases mentioned above. We note that when the cost functional $\mathscr{J}$ is minimized, the multiscale approach starting with a "good" value of l_1 and then progressively decreasing it to zero gives better results than standard optimi-

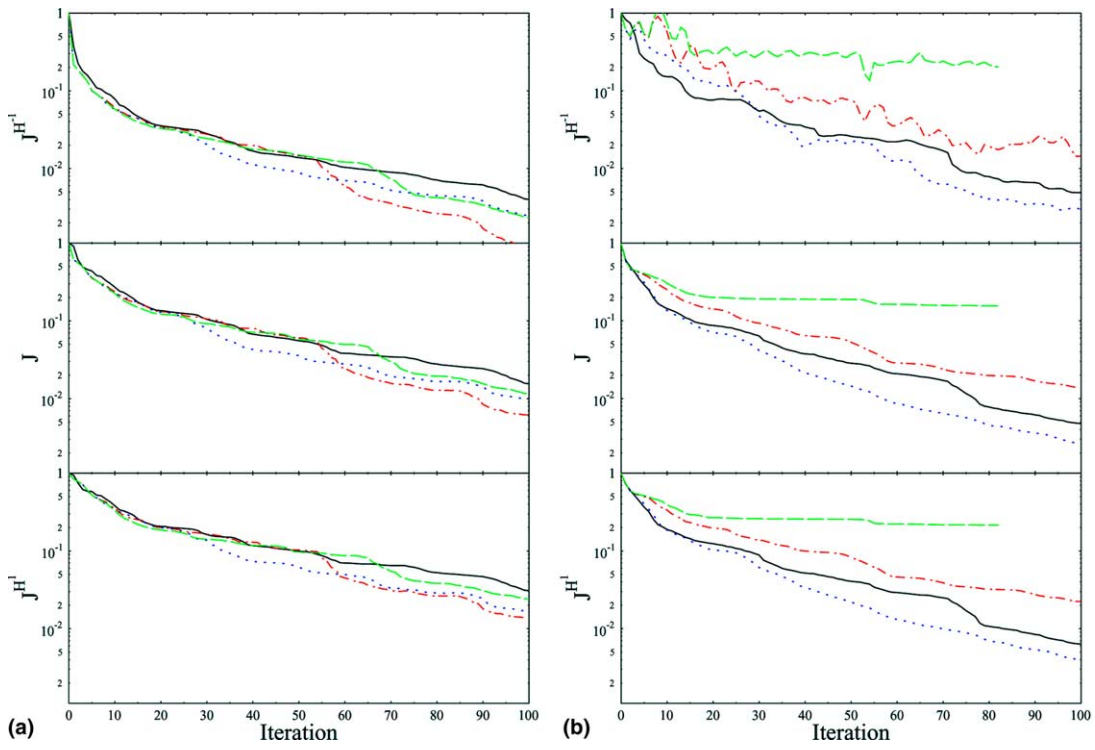

Fig. 8. Variation of the functionals $\mathscr{J}^{H^{-1}}$, $\mathscr{J}$, and $\mathscr{J}^{H^1}$ as a function of the iteration count for reconstructions based on minimizing $\mathscr{J}^{H^{-1}}$ (a) and $\mathscr{J}$ (b), and using the following inner products: (a) (solid) L_2, (dash-dotted) $W^{0,l_{-1}}$ with l_{-1} fixed, (dotted) $W^{0,l_{-1}}$ with l_{-1} decreasing to zero, and (dashed) $W^{0,l_{-1}}$ with l_{-1} increasing; (b) (solid) L_2, (dash-dotted) $W^{l_1,\infty}$ with l_1 fixed, (dotted) $W^{l_1,\infty}$ with l_1 decreasing to zero, and (dashed) $W^{l_1,\infty}$ with l_1 increasing.

zation employing L_2 gradients. When the cost functional $\mathscr{J}^{H^{-1}}$ is minimized, all approaches based on $W^{0,l_{-1}}$ gradients perform better than the standard strategy using L_2 gradients. As expected, minimizing $\mathscr{J}^{H^{-1}}$ instead of $\mathscr{J}$ results in a better convergence on large scales, while compromising slightly convergence on intermediate scales.

Some theoretical insights regarding the impact of the norm selected in the cost functional and the inner product used to extract the gradient on convergence of a gradient minimization algorithm applied to an optimization problem governed by a linear operator such as $\mathscr{L}$ can be gleaned by examining construction of the corresponding Krylov spaces. We consider a general case $\{H^m, H^n\}$, where m and n are the differentiation orders in the definition of the cost functional and in the inner product used for gradient extraction. After p iterations (starting with zero initial guess) the cost functional is minimized over the following Krylov space:

$$K_{m,n}(0,p) = \operatorname{span}\left\{ \partial_x^{-2n}(\mathscr{L}^*)^{-1}\partial_x^{2m}\mathscr{H}^*\mathscr{H}y, \partial_x^{-2n}(\mathscr{L}^*)^{-1}\partial_x^{2m}(\mathscr{L})^{-1}\partial_x^{-2n}(\mathscr{L}^*)^{-1}\partial_x^{2m}\mathscr{H}^*\mathscr{H}y, \ldots, \prod_{k=1}^{p}\left[\partial_x^{-2n}(\mathscr{L}^*)^{-1}\partial_x^{2m}(\mathscr{L})^{-1}\right]^k \partial_x^{-2n}(\mathscr{L}^*)^{-1}\partial_x^{2m}\mathscr{H}^*\mathscr{H}y \right\}.$$

We see that the structure of the Krylov spaces $K_{m,n}(0,p)$, in which the pth approximation of the minimizer is constructed, depends in an intricate way on an interplay of the operators ∂_x^{-2n} and ∂_x^{2m} with the evolution operator $\mathscr{L}$ and its adjoint $\mathscr{L}^*$. In particular, we note that in the special case of spatially uniform system considered in Section 4.5.3, the differential operators commute with both $(\mathscr{L})^{-1}$ and $(\mathscr{L}^*)^{-1}$ and therefore the case $\{H^m, H^n\}$ is equivalent to $\{L_2, H^{n-m}\}$.

6. Extensions

In Section 4 we presented a comprehensive picture of different regularization strategies applied to a simple model system. Since the ultimate goal is to apply these methods to real systems of physical and engineering interest, such as the Navier–Stokes equation in a bounded domain, below we show that such generalization is in fact straightforward. One significant difference is the more complicated structure of the governing equation and its adjoint when working in higher spatial dimensions. Another significant difference is related to the fact that various terms obtained from the integration by parts do not vanish on the solid boundaries. In order to highlight some of the issues, below we will outline how applying selected regularization options modifies the formulation of an adjoint-based optimization of the Navier–Stokes system in a periodic domain (Section 6.1) and the Kuramoto–Sivashinsky equation in a bounded domain (Section 6.2). Due to space limitations, we restrict ourselves to presenting only the numerical framework and do not show any computational results. For the same reason, the case of the Navier–Stokes system in a bounded domain is also deferred to a forthcoming paper.

6.1. Controlling a 3D Navier–Stokes system

We consider here a Navier–Stokes system

$$\begin{cases} \frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v}\cdot\nabla)\mathbf{v} + \nabla p - \mu\Delta\mathbf{v} = \boldsymbol{\phi}, & \text{in } \Omega\times(0,T), \\ \nabla\cdot\mathbf{v} = 0, & \text{in } \Omega\times(0,T), \\ \mathbf{v} = \mathbf{v}_0, & \text{at } t = 0, \\ \mathbf{v} \text{ periodic in } x_1, x_2, x_3, \end{cases} \tag{72}$$

where $\mathbf{v}$ is the velocity, p is the pressure, μ is the viscosity, and $\boldsymbol{\phi}$ is an externally applied body force representing the control. The system is supplied with the initial condition $\mathbf{v}_0$ and its evolution takes place in a domain $\Omega = [0, 2\pi]^3$ periodic in all three spatial dimensions. The domains Ω_1, Ω_2 and Ω_3 (see Section 2) all coincide with Ω, while the system evolution, the control, and the cost function are all defined on $\Omega \times [0, T]$. Since we now deal with vector quantities depending on three spatial variables, and the control now also depends on time, the brackets (20a)–(20c) used to frame the adjoint analysis need to be suitably redefined

$$\|\mathbf{z}\|^2_{L_2(0,T;H^q(\Omega))} \triangleq \int_0^T \int_\Omega \frac{\partial^q z_i}{\partial x_j^q} \frac{\partial^q z_i}{\partial x_j^q} \,\mathrm{d}\Omega \,\mathrm{d}t, \tag{73a}$$

$$\langle \mathbf{y}, \mathbf{z} \rangle_{L_2(0,T;H^q(\Omega))} \triangleq \int_0^T \int_\Omega \frac{\partial^q y_i}{\partial x_j^q} \frac{\partial^q z_i}{\partial x_j^q} \,\mathrm{d}\Omega \,\mathrm{d}t, \tag{73b}$$

$$(\mathbf{y}, \mathbf{z})_{L_2(0,T;H^q(\Omega))} \triangleq \int_0^T \int_\Omega \frac{\partial^q y_i}{\partial x_j^q} \frac{\partial^q z_i}{\partial x_j^q} \,\mathrm{d}\Omega \,\mathrm{d}t, \tag{73c}$$

where repeated indices imply summation and we will restrict our attention to the cases with $q \geqslant 0$. By analogy with (21), we can define inner products as a weighted linear combination of terms of the form (73c) with different values of q, e.g.

$$(\mathbf{y}, \mathbf{z})_{L_2(0,T;W^{l_1,\infty})} \triangleq \frac{1}{1 + l_1^2} \left[(\mathbf{y}, \mathbf{z})_{L_2(0,T;L_2(\Omega))} + l_1^2 (\mathbf{y}, \mathbf{z})_{L_2(0,T;H^1(\Omega))} \right]. \tag{74}$$

In order to emphasize the differences with respect to the standard approach, we analyze here the problem studied initially in the seminal paper of Abergel and Temam [39], i.e., enstrophy minimization with control in the form of the body force applied to the 3D Navier–Stokes system in a periodic domain. Consequently, we attempt to minimize the functional [6]

$$\mathscr{J}_{\mathrm{ns}}(\boldsymbol{\phi}) \triangleq \frac{1}{2} \|\nabla \times \mathbf{v}\|^2_{L_2(0,T;L_2(\Omega))}. \tag{75}$$

The classical formulation is obtained by following the methodology of Section 3 (see also [39]). The differential of the cost functional is

$$\mathscr{J}'_{\mathrm{ns}}(\boldsymbol{\phi}, \boldsymbol{\phi}') = \int_0^T \int_\Omega (\nabla \times \mathbf{v}) \cdot (\nabla \times \mathbf{v}') \,\mathrm{d}\Omega \,\mathrm{d}t = -\int_0^T \int_\Omega \Delta \mathbf{v} \cdot \mathbf{v}' \,\mathrm{d}\Omega \,\mathrm{d}t, \tag{76}$$

where $\boldsymbol{\phi}'$ is a perturbation of the control and $\mathbf{v}'(\boldsymbol{\phi}, \boldsymbol{\phi}')$ solves the system

$$\begin{cases} \mathscr{N} \begin{bmatrix} \mathbf{v}' \\ p' \end{bmatrix} = \begin{bmatrix} \boldsymbol{\phi}' \\ 0 \end{bmatrix} & \text{in } \Omega \times (0, T), \\ \mathbf{v}' = 0 & \text{at } t = 0, \\ \mathbf{v}' \text{ periodic in } x_1, x_2, x_3, \end{cases} \tag{77}$$

with the linear operator

[6] For the sake of simplicity we skip here the penalty on the control $\boldsymbol{\phi}$. As noted in the numerical experiments of [5], the removal of this control penalty in nonlinear Navier–Stokes control problems apparently leads to bounded control feedback at least in a subset of well-defined problems.

$$\mathscr{N}\begin{bmatrix}\mathbf{v}'\\ p'\end{bmatrix}=\begin{bmatrix}\frac{\partial\mathbf{v}'}{\partial t}+(\mathbf{v}\cdot\nabla)\mathbf{v}'+(\mathbf{v}'\cdot\nabla)\mathbf{v}-\mu\Delta\mathbf{v}'+\nabla p'\\ -\nabla\cdot\mathbf{v}'\end{bmatrix}. \tag{78}$$

We now define an adjoint operator with the identity

$$\left\langle\mathscr{N}\begin{bmatrix}\mathbf{v}'\\ p'\end{bmatrix},\begin{bmatrix}\mathbf{v}^*\\ p^*\end{bmatrix}\right\rangle_{L_2(0,T;L_2(\Omega))}=\left\langle\begin{bmatrix}\mathbf{v}'\\ p'\end{bmatrix},\mathscr{N}^*\begin{bmatrix}\mathbf{v}^*\\ p^*\end{bmatrix}\right\rangle_{L_2(0,T;L_2(\Omega))}+b_{\mathrm{ns}}, \tag{79}$$

where the inner product of vectors is defined in (73b). Consequently, the adjoint operator has the form

$$\mathscr{N}^*\begin{bmatrix}\mathbf{v}^*\\ p^*\end{bmatrix}=\begin{bmatrix}-\frac{\partial\mathbf{v}^*}{\partial t}-\mathbf{v}\cdot[\nabla\mathbf{v}^*+(\nabla\mathbf{v}^*)^{\mathrm{T}}]-\mu\Delta\mathbf{v}^*+\nabla p^*\\ -\nabla\cdot\mathbf{v}^*\end{bmatrix}, \tag{80}$$

and the adjoint system may be defined as

$$\begin{cases}\mathscr{N}^*\begin{bmatrix}\mathbf{v}^*\\ p^*\end{bmatrix}=\begin{bmatrix}-\Delta\mathbf{v}\\ 0\end{bmatrix} & \text{in } \Omega\times(0,T),\\ \mathbf{v}^*=0 & \text{at } t=T,\\ \mathbf{v}^* \text{ periodic in } x_1,x_2,x_3.\end{cases} \tag{81}$$

In such a case we obtain $b_{\mathrm{ns}}=0$ and the relations (79), (77) and (81) can be used to re-express $\mathscr{J}'_{\mathrm{ns}}(\boldsymbol{\phi},\boldsymbol{\phi}')$ as

$$\mathscr{J}'_{\mathrm{ns}}(\boldsymbol{\phi},\boldsymbol{\phi}')=\int_0^T\int_\Omega\boldsymbol{\phi}'\cdot\mathbf{v}^*\,\mathrm{d}\Omega\,\mathrm{d}t=\left(\nabla^{L_2(0,T;L_2(\Omega))}\mathscr{J}_{\mathrm{ns}},\boldsymbol{\phi}'\right)_{L_2(0,T;L_2(\Omega))},$$

which yields the L_2 functional gradient

$$\nabla^{L_2(0,T;L_2(\Omega))}\mathscr{J}_{\mathrm{ns}}=\mathbf{v}^*. \tag{82}$$

In the vein of Sections 4.3 and 4.4, below we examine how this derivation is modified when different forms of the inner product defining the adjoint identity and the gradient extraction is selected.

6.1.1. Adjoint derivation with the H^1 inner product

We now derive the adjoint operator using the identity

$$\left\langle\mathscr{N}\begin{bmatrix}\mathbf{v}'\\ p'\end{bmatrix},\begin{bmatrix}\mathbf{v}^{*,H^1}\\ p^{*,H^1}\end{bmatrix}\right\rangle_{L_2(0,T;H^1(\Omega))}=\left\langle\begin{bmatrix}\mathbf{v}'\\ p'\end{bmatrix},\mathscr{N}^{*,H^1}\begin{bmatrix}\mathbf{v}^{*,H^1}\\ p^{*,H^1}\end{bmatrix}\right\rangle_{L_2(0,T;H^1(\Omega))}+b_{\mathrm{ns},1}, \tag{83}$$

which yields the new adjoint operator

$$\mathscr{N}^{*,H^1}\begin{bmatrix}\mathbf{v}^{*,H^1}\\ p^{*,H^1}\end{bmatrix}=\begin{bmatrix}-\frac{\partial\mathbf{v}^{*,H^1}}{\partial t}-\Delta_0^{-1}\left\{\mathbf{v}\cdot\Delta\left[\nabla\mathbf{v}^{*,H^1}+(\nabla\mathbf{v}^{*,H^1})^{\mathrm{T}}\right]\right\}-\mu\Delta\mathbf{v}^{*,H^1}+\nabla p^{*,H^1}\\ -\nabla\cdot\mathbf{v}^{*,H^1}\end{bmatrix}, \tag{84}$$

where Δ_0^{-1} is the inverse Laplace operator associated with homogeneous Dirichlet boundary conditions. We define the new adjoint system as

$$\begin{cases}\mathscr{N}^{*,H^1}\begin{bmatrix}\mathbf{v}^{*,H^1}\\ p^{*,H^1}\end{bmatrix}=\begin{bmatrix}\Delta_0^{-1}\Delta\mathbf{v}\\ 0\end{bmatrix} & \text{in } \Omega\times(0,T),\\ \mathbf{v}^{*,H^1}=0 & \text{at } t=T,\\ \mathbf{v}^{*,H^1} \text{ periodic in } x_1,x_2,x_3,\end{cases} \tag{85}$$

from which we obtain $b_{\mathrm{ns},1} = 0$ and

$$\mathscr{J}'_{\mathrm{ns}}(\boldsymbol{\phi}, \boldsymbol{\phi}') = \int_0^T \int_\Omega \frac{\partial v_i^{*,H^1}}{\partial x_j} \frac{\partial \phi_i'}{\partial x_j} \,\mathrm{d}\Omega \,\mathrm{d}t.$$

Identifying this expression with either $(\boldsymbol{\nabla}^{L_2(0,T;L_2(\Omega))} \mathscr{J}_{\mathrm{ns}}, \boldsymbol{\phi}')_{L_2(0,T;L_2(\Omega))}$ or $(\boldsymbol{\nabla}^{L_2(0,T;H^1(\Omega))} \mathscr{J}_{\mathrm{ns}}, \boldsymbol{\phi}')_{L_2(0,T;H^1(\Omega))}$ allows us to extract the corresponding gradients as

$$\boldsymbol{\nabla}^{L_2(0,T;L_2(\Omega))} \mathscr{J} = -\Delta \mathbf{v}^{*,H^1}, \quad \boldsymbol{\nabla}^{L_2(0,T;H^1(\Omega))} \mathscr{J} = \mathbf{v}^{*,H^1}.$$

6.1.2. Preconditioning the gradient

We can extract preconditioned gradients by identifying the differential of the cost functional $\mathscr{J}'_{\mathrm{ns}}(\boldsymbol{\phi}, \boldsymbol{\phi}')$ with an alternative form of the inner product, such as that defined in (74). As in Section 4.4, the new gradient can be obtained for all $t \in [0, T]$ as a solution of the Helmholtz equation

$$\begin{cases} \frac{1}{1+l_1^2}[1 - \Delta] \boldsymbol{\nabla}^{L_2(0,T;W^{l_1,\infty})} \mathscr{J}_{\mathrm{ns}} = \mathbf{v}^*, \\ \boldsymbol{\nabla}^{L_2(0,T;W^{l_1,\infty})} \mathscr{J}_{\mathrm{ns}}, \ \frac{\partial}{\partial n} \boldsymbol{\nabla}^{L_2(0,T;W^{l_1,\infty})} \mathscr{J}_{\mathrm{ns}} \text{ periodic in } x_1, x_2, x_3. \end{cases} \tag{86}$$

We remark that we obtain by this approach the same properties with respect to scale-dependent filtering as discussed in Section 4.4. Since in the present case the control $\boldsymbol{\phi}$ is also a function of time, the definition of the inner product used to extract the gradient can also be generalized by incorporating derivatives with respect to time. Using such an inner product will result in smoothing the gradient in the time domain. This approach is discussed further in Section 6.2.4.

6.2. Controlling a Kuramoto–Sivashinsky system in a bounded domain

We now proceed to investigate how the presence of solid boundaries affects the regularization strategies developed in Section 4. We first briefly review the standard formulation and then show how it is modified when each of the four regularization options is applied in turn. For this purpose we consider the system (2) in a bounded domain $\Omega = [0, 2\pi]$ (see, e.g. [40])

$$\begin{cases} \partial_t v + 4\partial_x^4 v + \kappa\left(\partial_x^2 v + v\partial_x v\right) = 0, & x \in \Omega, \ \ t \in [0, T], \\ v(0, t) = v(2\pi, t) = 0, & t \in [0, T], \\ \partial_x v(0, t) = \phi, \partial_x v(2\pi, t) = 0, & t \in [0, T], \\ v(x, 0) = v_0, & x \in \Omega, \end{cases} \tag{87}$$

where a time-dependent control ϕ is applied on one boundary to regulate a quantity defined on the opposite boundary (note that for consistency with the initial data we must have $\partial_x v_0|_{x=0} = \phi|_{t=0}$). Consequently, the norm and the inner product needed to formulate the adjoint analysis are redefined as follows:

$$\|z\|^2_{H^p(0,T)} \triangleq \int_0^T \left(\partial_t^p z\right)^2 \mathrm{d}t, \tag{88a}$$

$$(z_1, z_2)_{H^p(0,T)} \triangleq \int_0^T \left(\partial_t^p z_1\right)\left(\partial_t^p z_2\right) \mathrm{d}x. \tag{88b}$$

(Note that discussion of incorporating time derivatives into the inner product defining the adjoint identity is deferred to Appendix A.) We now select the cost functional as

$$\mathscr{J}_b(\phi) = \frac{1}{2}\left\| \partial_x^2 v\big|_{x=2\pi} \right\|^2_{L_2(0,T)}. \tag{89}$$

In the present problem we have therefore the following relations between the spatial domains of interest: $\Omega_1 = \{2\pi\}$, $\Omega_2 = \Omega$ and $\Omega_3 = \{0\}$. Both the cost functional and the control are defined on $[0,T]$, whereas the system evolution again takes place over $\Omega \times [0,T]$. The differential of the cost functional is

$$\mathscr{J}'(\phi;\phi') = \int_0^T \left[\left(\partial_x^2 v\right)\left(\partial_x^2 v'\right) \right]_{x=2\pi} \mathrm{d}t, \tag{90}$$

where v' is the solution of the system

$$\begin{cases} \mathscr{L} v' = 0, & x \in \Omega, \ \ t \in [0,T], \\ v'(0,t) = v'(2\pi,t) = 0, & t \in [0,T], \\ \partial_x v'(0,t) = \phi', \partial_x v'(2\pi,t) = 0, & t \in [0,T], \\ v'(x,0) = 0, & x \in \Omega, \end{cases} \tag{91}$$

with the operator $\mathscr{L}$ defined as in (11). Note that consistency between the initial and boundary conditions requires that $\phi'(0) = 0$. The adjoint operator $\mathscr{L}^*$ is introduced using the identity (14) and is given by (15). Defining the adjoint state v^* such that

$$\begin{cases} \mathscr{L}^* v^* = 0, & x \in \Omega, \ \ t \in [0,T], \\ v^*(0,t) = v^*(2\pi,t) = 0, & t \in [0,T], \\ \partial_x v^*(0,t) = 0, \partial_x v^*(2\pi,t) = \partial_x^2 v(2\pi,t), & t \in [0,T], \\ v^*(x,T) = 0, & x \in \Omega, \end{cases} \tag{92}$$

we can use (14), (91), and (92) to re-express the differential of the cost functional as

$$\mathscr{J}'_b(\phi;\phi') = -\int_0^T \left(\partial_x^2 v^*\big|_{x=0} \right) \phi' \, \mathrm{d}t = \left(\nabla^{L_2(0,T)} \mathscr{J}_b, \phi' \right)_{L_2(0,T)}$$

from which we obtain the L_2 gradient

$$\nabla^{L_2(0,T)} \mathscr{J}_b = -\partial_x^2 v^*(0,t). \tag{93}$$

6.2.1. Adjoint derivation based on the vorticity formulation

We now proceed to minimize the original cost functional (89) by modeling the system evolution with the vorticity form of the Kuramoto–Sivashinsky system [7]

$$\begin{cases} \partial_t w + 4\partial_x^4 w + \kappa\left(\partial_x^2 w + w^2 + \partial_x^{-1} w \partial_x w\right) = 0, & x \in \Omega, \ \ t \in [0,T], \\ v(0,t) = v(2\pi,t) = 0, & t \in [0,T], \\ w(0,t) = \phi, w(2\pi,t) = 0, & t \in [0,T], \\ w(x,0) = w_0, & x \in \Omega. \end{cases} \tag{94}$$

The cost functional can now be rewritten in the form

$$\mathscr{J}_b(\phi) = \frac{1}{2} \left\| \partial_x w\big|_{x=2\pi} \right\|^2_{L_2(0,T)} \tag{95}$$

[7] Note that, similarly to the vorticity form of the Navier–Stokes system in a bounded domain, the boundary conditions of the vorticity system (94) also involve the primitive variable v.

and its differential can be expressed as

$$\mathscr{J}'(\phi;\phi') = \int_0^T \left[(\partial_x w)(\partial_x w')\right]_{x=2\pi} \mathrm{d}t, \tag{96}$$

where w' is the solution of the system

$$\begin{cases} \mathscr{M}w' = 0, & x \in \Omega, \ \ t \in [0,T], \\ v'(0,t) = v'(2\pi,t) = 0, & t \in [0,T], \\ w'(0,t) = \phi', w'(2\pi,t) = 0, & t \in [0,T], \\ w'(x,0) = 0, & x \in \Omega, \end{cases} \tag{97}$$

where the operator $\mathscr{M}$ is defined as in (25). The adjoint operator is introduced with the bracket (26) and is given by (27). Defining the adjoint state w^* such that

$$\begin{cases} \mathscr{M}^* w^* = 0, & x \in \Omega, \ \ t \in [0,T], \\ \partial_x w^*(0,t) = \partial_x w^*(2\pi,t) = 0, & t \in [0,T], \\ \partial_x^2 w^*(0,t) = 0, \partial_x^2 w^*(2\pi,t) = \partial_x w(2\pi,t), & t \in [0,T], \\ w^*(x,T) = 0, & x \in \Omega, \end{cases} \tag{98}$$

we can now use (26), (97), and (98) to re-express the cost functional differential as

$$\mathscr{J}'_b(\phi;\phi') = -\int_0^T \left(\partial_x^3 w^*\big|_{x=0}\right)\phi' \,\mathrm{d}t = \left(\nabla^{L_2(0,T)}\mathscr{J}_b, \phi'\right)_{L_2(0,T)}$$

from which we obtain the L_2 gradient

$$\nabla^{L_2(0,T)}\mathscr{J}_b = -\partial_x^3 w^*(0,t).$$

6.2.2. Targeting the cost functional

Since the regulated quantity is now a function of time only, an alternative, targeted, cost functional may be selected as, for instance,

$$\mathscr{J}_b^{H^1}(\phi) = \frac{1}{2}\left\|\partial_x^2 v\big|_{x=2\pi}\right\|^2_{H^1(0,T)}, \tag{99}$$

in which case the differential becomes

$$\begin{aligned} \mathscr{J}_b'^{H^1}(\phi;\phi') &= \int_0^T \left[\left(\partial_t\partial_x^2 v\right)\left(\partial_t\partial_x^2 v'\right)\right]_{x=2\pi} \mathrm{d}t \\ &= -\int_0^T \left[\left(\partial_t^2\partial_x^2 v\right)\left(\partial_x^2 v'\right)\right]_{x=2\pi} \mathrm{d}t + \left\{\left[\left(\partial_t\partial_x^2 v\right)\left(\partial_x^2 v'\right)\right]_{x=2\pi}\right\}_{t=0}^{t=T}, \end{aligned} \tag{100}$$

so that the corresponding adjoint system is now

$$\begin{cases} \mathscr{L}^* v^* = 0, & x \in \Omega, \ \ t \in [0,T], \\ v^*(0,t) = v^*(2\pi,t) = 0, & t \in [0,T], \\ \partial_x v^*(0,t) = 0, \partial_x v^*(2\pi,t) = -\partial_t^2\partial_x^2 v(2\pi,t) + \delta(t-T)\left(\partial_t\partial_x^2 v\right)\big|_{x=2\pi}, & t \in [0,T], \\ v^*(x,0) = 0, & x \in \Omega. \end{cases}$$

This system must be interpreted in the sense of a distribution, as one of the boundary conditions involves a "delta function" in time, effectively forcing the adjoint system from the "corner" of the space–time domain.

The fact that here, unlike in all the previous cases, we are strictly able to identify only a weak form of the adjoint system is not considered an insurmountable problem, as there are well-established methods for the numerical approximation of such systems.

6.2.3. Adjoint derivation with the H^1 inner product

We now derive the adjoint using the H^1 inner product (41), and write out the complete form of the term b_1 appearing in this relation as

$$\begin{aligned} b_1 = & \int_0^{2\pi} \left[(\partial_x v') \left(\partial_x v^{*,H^1} \right) \right]_{t=0}^{t=T} \mathrm{d}x + \int_0^T \Big\{ \kappa \Big[\left(\partial_x v^{*,H^1} \right) \left(\partial_x^2 v' \right) - \left(\partial_x^2 v^{*,H^1} \right) \left(\partial_x v' \right) \Big] \\ & + \kappa \Big[\left(\partial_x v^{*,H^1} \right) \partial_x (v v') - \left(\partial_x^2 v^{*,H^1} \right) v v' + \partial_x^{-1} \left(v \partial_x^3 v^{*,H^1} \right) v' \Big] + 4 \Big[\left(\partial_x v^{*,H^1} \right) \left(\partial_x^4 v' \right) - \left(\partial_x^2 v^{*,H^1} \right) \left(\partial_x^3 v' \right) \\ & + \left(\partial_x^3 v^{*,H^1} \right) \left(\partial_x^2 v' \right) - \left(\partial_x^4 v^{*,H^1} \right) \left(\partial_x v' \right) \Big] \Big\}_{x=0}^{x=2\pi} \mathrm{d}t. \end{aligned}$$

Making use of the expression for $\mathscr{L}^{*,H^1}$ in (42), we now define the new adjoint system as

$$\begin{cases} \mathscr{L}^{*,H^1} v^{*,H^1} = 0, & x \in \Omega, \ \ t \in [0,T], \\ \partial_x^2 v^{*,H^1}(0,t) = \partial_x^2 v^{*,H^1}(2\pi,t) = 0, & t \in [0,T], \\ \partial_x^3 v^{*,H^1}(0,t) = 0, \ \partial_x^3 v^{*,H^1}(2\pi,t) = \partial_x^2 v(2\pi,t), & t \in [0,T], \\ v^{*,H^1}(x,T) = 0, & x \in \Omega, \end{cases} \tag{101}$$

which allows us to re-express the differential of the cost functional as

$$\mathscr{J}_b'(\phi;\phi') = -\int_0^T \left(\partial_x^4 v^{*,H^1} \Big|_{x=0} \right) \phi' \, \mathrm{d}t = \left(\nabla^{L_2(0,T)} \mathscr{J}_b, \phi' \right)_{L_2(0,T)}. \tag{102}$$

As a result, the L_2 gradient can be extracted as

$$\nabla^{L_2(0,T)} \mathscr{J}_b = -\partial_x^4 v^{*,H^1}(0,t).$$

By comparing this to the standard formulation culminating with (93), we note that the same L_2 gradient of the cost functional is now obtained by applying a higher-order differential operator to the adjoint field obtained in the new formulation, which is consistent with the relationship between the corresponding expressions for the L_2 gradient in the periodic case as tabulated in Table 1.

6.2.4. Preconditioning the gradient

The control ϕ in the present problem is a function of time only, and new gradients of the cost function (89) can be obtained by identifying its differential with an inner product (88b) incorporating either derivatives ($p = 1$), or anti-derivatives ($p = -1$) with respect to the time variable. In the former case we obtain

$$\mathscr{J}_b'(\phi;\phi') = -\int_0^T \left(\partial_x^2 v^* \Big|_{x=0} \right) \phi' \, \mathrm{d}t = \left(\nabla^{H^1(0,T)} \mathscr{J}_b, \phi' \right)_{H^1(0,T)}, \tag{103}$$

so that the following holds

$$\begin{cases} \partial_t^2 \nabla^{H^1(0,T)} \mathscr{J}_b = \partial_x^2 v^* \big|_{x=0}, \\ \partial_t \nabla^{H^1(0,T)} \mathscr{J}_b(0,0) = \partial_t \nabla^{H^1(0,T)} \mathscr{J}_b(0,T) = 0. \end{cases} \tag{104}$$

We see that the new gradient is obtained by solving this elliptic-in-time boundary-value problem, and therefore will be smoother in the time domain. In the spirit of Section 4.4, this approach can be generalized

by considering an inner product which is a combination of the L_2 and H^1 terms, as this would allow us to focus the optimization on a specific range of time scales that are of interest in a given optimization problem.

7. Discussion and conclusions

In this paper we have identified and related the four opportunities for generalizing the formulation of an adjoint-based gradient optimization algorithm. The first opportunity concerns the choice of the specific form of the equation assumed to govern the system evolution. The remaining three opportunities are related to the choice of the norm and the inner products (collectively referred to in the paper as "brackets") on the three space–time domains that are of interest in a generic optimization problem applied to an unsteady PDE system. Most studies to date have used L_2 brackets on all three of these space–time domains. In the present study we have explored formulations based on the more general Sobolev brackets, which include the L_2 brackets as special cases. Choosing an alternative form of the evolution equation together with the adoption of different Sobolev brackets has the effect of emphasizing or de-emphasizing different length- and time-scales in the definition of the cost functional, the adjoint operator and the associated adjoint field, and the gradient of the cost functional. These opportunities allow one to fine-tune the optimization algorithm to the specific length- and time-scales of interest. By so doing, one may make the original PDE optimization problem more "regular", and thereby easier to discretize and solve numerically.

The four regularization opportunities discussed in Section 4 fall into two categories: those that affect the descent direction (i.e., targeting the cost functional in Section 4.2 and preconditioning the gradient in Section 4.4), and those that affect the manner in which a given descent direction is computed (i.e., changing the form of the evolution equation in Section 4.1 and the inner product in Section 4.3). In the discrete, finite-precision setting, all four of these opportunities are significant, and the different opportunities may be used to amend the different elements of the algorithm. For example, gradient extraction performed using an inner product which combines the H^{-1}, L_2, and H^1 brackets was shown to be equivalent to applying a suitable scale-dependent filter to the adjoint field. A low-pass filter of this sort is useful to employ when the high-frequency components of the system are somehow considered "less significant" or "corrupted by noise" during the optimization process in the multiscale system. In a data assimilation problem this could be the case, for instance, when one is attempting to obtain a long-term forecast, in which the smallest-scale variations of the initial conditions are thought to play a relatively unimportant role. On the other hand, a band-pass filter could be useful to employ when one is attempting to obtain a short-term "meso-scale" forecast, in which the small-scale variations of the initial conditions are again thought to play a relatively unimportant role and the large-scale variations of the initial conditions are determined by a separate (global-scale) optimization code. Optimization based on an alternative form of the evolution equation becomes useful when a numerical implementation of the primitive form of the evolution equation is not readily available, but an implementation of a derived form exists, and the adjoint code is to be generated using automatic techniques (e.g., such as described in [42]). If, as is often the case, one is ultimately interested in obtaining sensitivities with respect to the primitive variables, then the present framework provides guidelines on how such sensitivities can be obtained from optimizations based on derived forms of the evolution equation.

As indicated in the literature survey in Section 1, approaches related to some of the regularization options presented here had already been mentioned in earlier studies. The present paper examines in detail all of the different opportunities and attempts to unify them into a coherent framework by highlighting the relations between the different possibilities. It should be remarked that the same set of regularization opportunities also applies in a straightforward fashion to the "robustified" framework for noncooperative ("worst case") optimization developed by Bewley et al. [41], thus allowing for tunable incorporation of model and measurement errors into a single framework.

The presented framework opens up the possibility for adoption of a wide range of regularization strategies. In order to illustrate these opportunities in a clear and exhaustive fashion, we chose to analyze them in this work based on a simple Kuramoto–Sivashinsky model forecasting problem. Moreover, studying a spatially periodic system allowed us to recast parts of the analysis in Fourier space, which facilitated drawing conclusions regarding spatial regularity of the various fields involved. We also addressed some of the issues arising when the framework presented is extended to systems governed by more complicated evolutions equations (e.g., the 3D Navier–Stokes system), and systems evolving in bounded domains. In such systems, analysis is more difficult, but the fundamental concepts remain the same. A forthcoming paper will discuss the application of some of the regularization opportunities presented here to complex optimization problems involving the Navier–Stokes system in 3D bounded domains.

The computational examples presented in this paper, while far short of exhaustively examining all of the various regularization opportunities, highlighted a few of the computational advantages inherent in the proposed framework. Based on a modified inner-product definition used to extract the gradient, a physically motivated multiscale preconditioning strategy was proposed which noticeably accelerates convergence of an optimization procedure applied to a nonlinear multiscale system. Adoption of similar approaches to the optimization of more complex systems of physical and engineering interest is currently underway.

Acknowledgements

The authors thank Prof. Scott Collis for his inspiring and thoughtful feedback, and gratefully acknowledge the generous funding of the AFOSR programs directed by Dr. John Schmisseur and Prof. Belinda King.

Appendix A. Adjoint derivation with the "H^1-in-time" inner product

We present here yet another way of deriving the adjoint operator, namely using the bracket of the form [cf. (20b)]

$$\langle z_1, z_2 \rangle_{H^p(0,T;L_2(\Omega))} \triangleq \int_0^T \int_0^{2\pi} \left(\partial_t^p z_1\right)\left(\partial_t^p z_2\right) \mathrm{d}x\, \mathrm{d}t. \tag{A.1}$$

We will focus here on the case with $p = 1$, and define the "anti-derivative" operator ∂_t^{-1} as

$$\partial_t^{-1} z(t) \triangleq \int_0^t z(t')\, \mathrm{d}t' - \int_0^T z(t')\, \mathrm{d}t' = \int_T^t z(t')\, \mathrm{d}t', \tag{A.2}$$

so that $\partial_t^{-1} z(T) = 0$ for any $z(t)$. In order not to further complicate the notation, we will use the symbols $\mathscr{L}^*$ and v^* to also denote the new adjoint operator and the new adjoint variable. The adjoint identity has now the following form:

$$\langle v^*, \mathscr{L} v' \rangle_{H^1(0,T;L_2(\Omega))} = \langle \mathscr{L}^* v^*, v' \rangle_{H^1(0,T;L_2(\Omega))} + b_{1t}, \tag{A.3}$$

where

$$\begin{aligned} &\mathscr{L}^* v^* = -\partial_t v^* + 4\partial_x^4 v^* + \kappa\left[\partial_x^2 v^* - \partial_t^{-2}\left(v\partial_x \partial_t^2 v^*\right)\right], \text{ and} \\ &b_{1t} = \left\{ \int_0^{2\pi} \left[(\partial_t v^*)(\partial_t v') + (\partial_t v^*)\partial_x (vv') + v'\partial_t^{-1}(v\partial_x \partial_t^2 v^*) \right] \mathrm{d}x \right\}_{t=0}^{t=T} + \Big[\cdots \Big]_{x=0}^{x=2\pi}. \end{aligned} \tag{A.4}$$

We now define the new adjoint system as

$$\begin{cases} \mathscr{L}^* v^* = -\partial_t^{-2} \mathscr{H}^*(\mathscr{H} v - y) = -\partial_t^{-2} f, & x \in \Omega, \ \ t \in [0,T], \\ \partial_x^i v^*(0,t) = \partial_x^i v^*(2\pi,t), & t \in [0,T], \ \ i = 0,\ldots,3, \\ v^*(x,T) = 0, & x \in \Omega, \end{cases} \tag{A.5}$$

which, when considering the spatially periodic problem defined in Section 3 and combining with (10), (A.3), and (A.2), allows us to re-express the differential (9) as

$$\mathscr{J}'(\phi;\phi') = \int_0^{2\pi} \partial_t^2 v^*\big|_{t=0} \phi' \,\mathrm{d}x = (\nabla^{L_2}\mathscr{J}, \phi')_{L_2(\Omega)}.$$

From this we identify the L_2 gradient in terms of the new adjoint variable as

$$\nabla^{L_2}\mathscr{J} = \partial_t^2 v^*\big|_{t=0}.$$

We note that the new adjoint operator (A.4) and the RHS forcing term used in (A.5) have terms involving ∂_t^{-1} and are therefore nonlocal in time. However, as is evident from (A.2), at a given time instant t the operator ∂_t^{-1} depends on its argument in the interval $[t,T]$ only. Consequently, the system (A.5) can be marched backward in time (i.e., from T to 0) using conventional numerical time-marching methods. We also observe that, as compared to the primitive adjoint operator (15), the new adjoint operator (A.4) has a different "advection" term in which additional time derivatives and anti-derivatives are present. In this sense (A.4) is similar to (42), where the "advection" term includes additional space derivatives and anti-derivatives. Consequently, we can expect system (A.5) to produce adjoint fields which are more regular in the time domain (cf. Section 5.1).

References

[1] J. Reuther, A. Jameson, J. Farmer, L. Martinelli, D. Saunders, Aerodynamic shape optimization of complex aircraft configurations via an adjoint formulation, AIAA Paper 96-0094, 1996.
[2] J.O. Pralits, C. Airiau, A. Hanifi, D.S. Henningson, Sensitivity analysis using adjoint parabolized stability equations for compressible flows, Flow Turbul. Combust. 65 (2000) 321–346.
[3] P. Cathalifaud, P. Luchini, Algebraic growth in boundary layers: optimal control by blowing and suction at the wall, Eur. J. Mech. B 19 (2000) 469–490.
[4] S. Walter, C. Airiau, A. Bottaro, Optimal control of Tollmien–Schlichting waves in a developing boundary layer, Phys. Fluids 13 (2001) 2087–2096.
[5] T.R. Bewley, P. Moin, R. Temam, DNS-based predictive control of turbulence: an optimal benchmark for feedback algorithms, J. Fluid Mech. 447 (2001) 179–225.
[6] F.-X. Le Dimet, O. Talagrand, Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects, Tellus 38A (1986) 97–100.
[7] M.D. Gunzburger, Perspectives in Flow Control and Optimization, SIAM, Philadelphia, 2003.
[8] S.S. Sritharan, Optimal Control of Viscous Flows, SIAM, Philadelphia, 1998.
[9] T.R. Bewley, Flow control: new challenges for a new renaissance, Prog. Aerospace Sci. 37 (2001) 21–58.
[10] H. Engl, M. Hanke, A. Neubauer, Regularization of Inverse Problems, Kluwer, Dordrecht, 1996.
[11] P.Ch. Hansen, Rank-deficient and Discrete Ill-posed Problems, SIAM, Philadelphia, 1998.
[12] C.R. Vogel, Computational Methods for Inverse Problems, SIAM, Philadelphia, 2002.
[13] S.S. Collis, K. Ghayour, M. Heinkenschloss, M. Ulbrich, S. Ulbrich, Optimal control of unsteady compressible viscous flows, Int. J. Numer. Meth. Fluids 40 (2002) 1401–1429.
[14] M. Heinkenschloss, L.N. Vicente, An interface between optimization and application for the numerical simulation of optimal control problems, ACM Trans. Math. Soft. 25 (2) (1999) 157–190.
[15] J.W. Neuberger, Sobolev Gradients and Differential Equations, Springer, Berlin, 1997.
[16] S. Sial, J. Neuberger, T. Lookman, A. Saxena, Energy minimization using Sobolev gradients: application to phase separation and ordering, J. Comput. Phys. 189 (2003) 88–97.

[17] J. Liu, A multiresolution method for distributed parameter estimation, SIAM J. Sci. Comput. 14 (2) (1993) 389–405.
[18] A. Brandt, L.Y. Zaslavsky, Multiscale algorithm for atmospheric data assimilation, SIAM J. Sci. Comput. 18 (3) (1997) 949–956.
[19] A.-A. Grimstad, T. Mannseth, Nonlinearity, scale and sensitivity for parameter estimation problems, SIAM J. Sci. Comput. 21 (6) (2000) 2096–2113.
[20] K. Brusdal, T. Mannseth, Basis norm rescaling for nonlinear parameter estimation, SIAM J. Sci. Comput. 21 (6) (2000) 2114–2125.
[21] R.M. Lewis, S.G. Nash, A multigrid approach to optimization of systems governed by differential equations, AIAA paper 2000-4890, 2000.
[22] A.C. Lorenc, Optimal nonlinear objective analysis, Q. J. R. Meteorol. Soc. 114 (1988) 205–240.
[23] J.-N. Thepaut, P. Moll, Variational inversion of simulated tovs radiances using the adjoint technique, Q. J. R. Meteorol. Soc. 116 (1990) 1425–1448.
[24] C.L. Lin, T. Chai, J. Sun, On the smoothness constraints for four-dimensional data assimilation, J. Comput. Phys. 181 (2002) 430–453.
[25] P. Holmes, J.L. Lumley, G. Berkooz, Turbulence, Coherent Structures, Dynamical Systems and Symmetry, Cambridge University Press, Cambridge, 1996.
[26] H. Choi, R. Temam, P. Moin, J. Kim, Feedback control for unsteady flow and its application to the stochastic Burgers equation, J. Fluid Mech. 253 (1993) 509–543.
[27] K. Kunisch, S. Volkwein, Control of the Burgers equation by a reduced-order approach using proper orthogonal decomposition, J. Optimiz. Theory App. 102 (2) (1999) 345–371.
[28] J.A. Atwell, J.T. Borggaard, B.B. King, Reduced order controllers for Burgers' equation with a nonlinear observer, Appl. Math. Comput. Sci. 11 (6) (2001) 1311–1330.
[29] A.N. Tikhonov, V.Y. Arsenin, Solution of Ill-posed Problems, Wiley, New York, 1977.
[30] Y. Kuramoto, Diffusion induced chaos in reaction systems, Prog. Theoret. Phys. Suppl. 64 (1978) 346–367.
[31] G. Sivashinsky, Nonlinear analysis of hydrodynamic instability in laminar flames, Acta Astronaut. 4 (1977) 1177–1206.
[32] J.M. Hyman, B. Nicolaenko, The Kuramoto–Sivashinsky equation: a bridge between PDE's and dynamical systems, Physica D 18 (1986) 113–126.
[33] R. Temam, Infinite-dimensional Dynamical Systems in Mechanics and Physics. Applied Mathematical Sciences 68, second ed., Springer, Berlin, 1997.
[34] J. Nocedal, S.J. Wright, Numerical Optimization, Springer, Berlin, 1999.
[35] J.E. Dennis, R.B. Schnabel, Numerical Methods for Unconstrained Optimization and Nonlinear Equations, Prentice-Hall, Englewood Cliffs, NJ, 1983.
[36] Ch. Hirsch, Numerical Computation of Internal and External Flows, Wiley, New York, 1989.
[37] W.H. Press, B.P. Flanner, S.A. Teukolsky, W.T. Vetterling, Numerical Recipes: The Art of Scientific Computations, Cambridge University Press, Cambridge, 1986.
[38] O. Axelsson, Iterative Solution Methods, Cambridge University Press, Cambridge, 1994.
[39] F. Abergel, R. Temam, On some control problems in fluid mechanics, Theoret. Comput. Fluid Dyn. 1 (1990) 303–325.
[40] V.M. Eguíluz et al., Average patterns of spatiotemporal chaos: a boundary effect, Phys. Rev. E 59 (3) (1999) 2822–2825.
[41] T.R. Bewley, R. Temam, M. Ziane, A general framework for robust control in fluid mechanics, Physica D 138 (2000) 360–392.
[42] Ch. Faure (Ed.), Automatic Differentiation for Adjoint Code Generation, INRIA Technical Report RR-3555 (available from http://www.inria.fr/rrrt/rr-3555.html), 1998.