

# An optimization method for chaotic turbulent flow

Seung Whan Chung, Jonathan B. Freund

*University of Illinois at Urbana–Champaign*

---

## Abstract

Evidence indicates that **quantities-of-interest** in some turbulent flows can be controlled despite the overall chaotic dynamics. It is typically thought that this is via relatively deterministic, larger-scale components of the turbulence. However, finding such controls, if they exist, is challenging because chaos causes sensitivity gradients to explode and the search space to become intractably non-convex. This challenge is analyzed, and a penalty method is introduced to cope with it. In the new approach, the time domain is broken into segments approximately matching the chaos time scales, so that the solution within each segment is both physical and relatively deterministic. The initial condition of each segment is included in an adjoint-based gradient optimization, which temporarily introduces artificial  $\Delta\mathbf{q}$  discontinuities in the overall solution. The optimization then proceeds in stages with increasing penalization of  $\Delta\mathbf{q}$ . The method is developed and illustrated for a logistic map, the Lorenz Equation, and an advection augmented Kuramoto–Sivashinsky Equation. These examples show how the  $\Delta\mathbf{q}$  temporarily increases the search scale prior to the strong  $\Delta\mathbf{q} \rightarrow 0$  penalization that recovers a physical solution. It is then applied to turbulent Kolmogorov flow, for which it also far outperforms a standard adjoint-based gradient search. The utility of such an optimized chaotic solution is discussed.

*Keywords:* turbulence, optimization, control

*2020 MSC:* 76F70, 37D45, 35Q93

*Email addresses:* [schung58@illinois.edu](mailto:schung58@illinois.edu) (Seung Whan Chung),  
[jbfreund@illinois.edu](mailto:jbfreund@illinois.edu) (Jonathan B. Freund)

*Preprint submitted to Journal of Computational Physics*

*January 11, 2022*

---

## 1. Introduction

### 1.1. Optimization in flow computations

It is well-understood that predictive simulations can be augmented with optimization methods to increase scientific understanding or improve engineering design [1–12]. Mathematically, this means adjusting input parameters  $\Theta$  to reduce an objective functional  $\mathcal{J}$ . For high-dimensional  $\Theta$ , adjoint gradient-based optimization is particularly attractive. Both passive and active flow controls can be designed [1, 2], such as for airfoil geometry [13, 14], jet nozzle shape [15], and acoustic dampers [16]. Related applications include error estimation and control of the simulations themselves, including grid adaptation [17, 18] and data-assimilated modeling [3–5, 19, 20]. Similarly, flow instability beyond the linear regime can be analyzed via optimization, such as in thermoacoustics [6], bypass transition [7, 8, 21], Rayleigh–Taylor instability [9], and sound generation [10–12].

The flow optimizations we consider share two main characteristics. First, they involve large numbers of optimization variables. Optimization in high-dimensional systems benefits from using the gradient of the  $\mathcal{J}$  to find a local minimum in the neighborhood of the starting point [14]. However, this may or may not lead to significantly reduced  $\mathcal{J}$ . Second, we consider flows governed by trusted governing equations, which leads to equality-constrained optimization [22–24]. In such cases, adjoint methods for equality-constrained optimization provide the gradient in **the high-dimensional parameter space with a computational cost similar to the** baseline predictive simulation. These methods, however, are limited when the flow is turbulent and therefore chaotic.

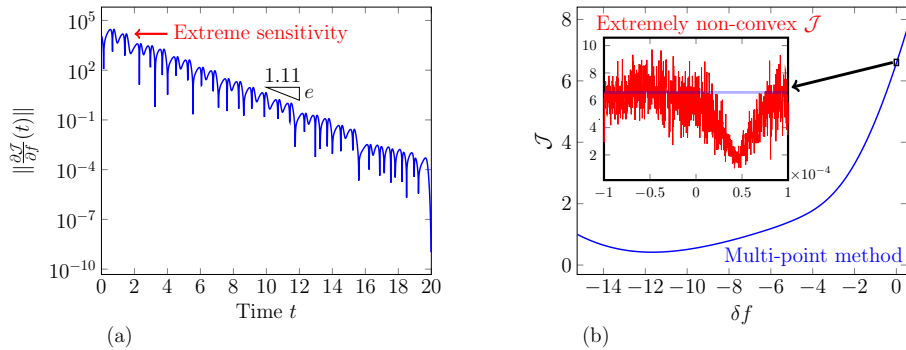


Figure 1: The impact of chaos on the optimal control of the Lorenz system: (a) the gradient of  $\mathcal{J}$  to  $f(t)$  growing exponentially backward in time, and (b) highly non-convex  $\mathcal{J}[f]$  along the direction of gradient  $\frac{\partial \mathcal{J}}{\partial f}(t)$ . For (b), these data are generated by brute-force evaluation of  $\mathcal{J}$  for  $10^4$  values of  $\delta f$ , with  $f(t) = \delta f \frac{\partial \mathcal{J}}{\partial f}(t)$  (see Section 3). The new multi-point method's smooth  $\mathcal{J}_A$  for  $\mu = 10^{-4}$  is included for comparison (see Section 4).

## 25 1.2. Challenge of optimization for turbulent flows

The particular challenge of flow turbulence we address can be illustrated with the Lorenz system [25],

$$\begin{aligned}
 \frac{dx}{dt} &= \sigma(y - x) \\
 \frac{dy}{dt} &= x(\rho - z) - y \\
 \frac{dz}{dt} &= xy - \beta z + f(t),
 \end{aligned} \tag{1}$$

to which we will also appeal for demonstrating the new method, before applying it to turbulence. The Lorenz state  $(x, y, z)$  famously orbits around two unstable fixed points. (Mathematical details of this problem are in Section 3.) A time-varying controller  $f(t)$  is sought to minimize a  $\mathcal{J}$  that favors one of the fixed points. However, chaos hinders gradient-based optimization of  $f(t)$  in two ways, both shared with turbulence. The extreme sensitivity of  $\frac{\partial \mathcal{J}}{\partial f}$  to  $f$ , manifests as the exponential growth of  $\|\frac{\partial \mathcal{J}}{\partial f}\|$  in reverse time seen in Figure 1 (a). This biases control toward  $t = 0$ , which limits its utility. In addition,  $\mathcal{J}$  becomes highly non-convex as shown in Figure 1 (b). The local minima created, each with a tiny neighborhood and a  $\mathcal{J}$  value that is not significantly improved [2], prevent

gradient-based optimization from finding a useful local optimum. These challenges are observed however  $\mathcal{J}$  is defined [9, 26], and they hinder optimization of turbulent flows [2, 9, 12, 27, 28]. For example, Kim *et al.* [12] attempted to reduce the far-field sound of a turbulent jet using an adjoint method. The optimized control was modestly effective, yet concentrated at the early times. It remained unclear whether the optimization was limited by an inherent lack of controllability or the challenge of the chaotic dynamics. A similar challenge was recognized by Pérez *et al.* [29] for a backward-facing step flow.

Unfortunately, there are few, if any, alternatives to gradient-based optimization. Non-convex optimization techniques [30], such as Lipschitz optimization [31, 32], statistical-model-based global optimization [33, 34], and random search methods [35, 36], are attractive, but the non-convexity caused by the chaos is often too challenging [31, 32]. Typical approaches either utilize an assumed statistical model for  $\mathcal{J}$  [33, 34] or resort to extensive searching [35, 36].

Turbulent flows lack sufficient statistical models for typical  $\mathcal{J}$  and are too expensive to compute for extensive searching.

### 1.3. The hope for optimization of turbulent flows

While Figure 1 makes the optimal control of the Lorenz equation seem impossible, this system is well-known to be controllable. Figure 2 shows such a controlled trajectory. This implies that the chaos of the system only obscures gradient-based methods from finding an effective local minimum but does not preclude control. The details of finding this solution and further investigation of its usefulness are in Section 4.

For turbulence, our starting-point conjecture is that many important flows have a relatively chaotic part that prevents gradient-based methods from arriving at effective control via relatively deterministic parts of the flow. It might be expected that this chaotic part is loosely associated with smaller-scale turbulence, while the less chaotic part is associated with relatively large-scale, coherent structures. The same idea motivates efforts to average or filter turbulence to extract reduced models; that approach is, of course, preferable, but

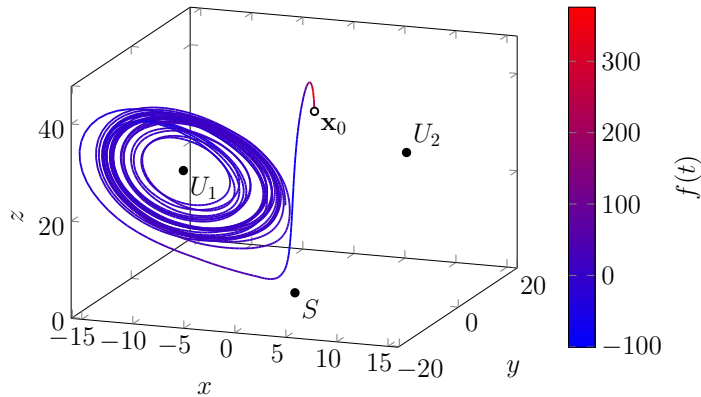


Figure 2: A controlled trajectory of the Lorenz equation, with color indicating the strength of the instantaneous control forcing  $f(t)$ .

it has not been quantitatively successful in all circumstances. In Section 5, we introduce a one-dimensional model based on the Kuramoto–Sivashinsky (K–S) equation to show this property.

Analyses support this perspective even when accurate reduced models have  
70 not been identified. Linear response theory for dissipative chaos [37] proves that averages based on ergodic state distributions are differentiable, which supports the possibility of optimization of some chaotic systems. Algorithms have been proposed based on this theory: ensemble-averaging of adjoint sensitivities [38, 39], using the fluctuation–dissipation theorem [40], a pdf-based approach  
75 [41, 42], least-square shadowing [43–45], cumulant-truncation [46], and space-splitting [47]. However, most of them are not yet applicable for optimization. **These methods** also suffer from poor convergence due to diverging sensitivity [38, 39]; they require prohibitive amounts of computations [41–45]; and they may be inaccurate when the system is not exactly ergodic or uniformly hyperbolic  
80 [40, 42, 45, 46]. Chandramoorthy and Wang [47] recently proposed a promising gradient computation method that requires computation of all covariant Lyapunov vectors; however, this is not yet feasible for large-scale turbulent flows. Chung *et al.* [48] used a continuum limit to regularize the gradient of particle dynamics discretization; however, that method relies on the particular prop-

85 erties of the particle dynamics, which precludes direct extension to turbulent flows.

Support also comes from experience with particular flows. For jet noise, the link from jet turbulence to radiated sound is intricate, but the success of wave-packet descriptions suggests sufficient determinism for controllability [10, 49, 50].  
90 For example, low-frequency, low-wavenumber eigenmodes from stability analysis show reasonable agreement with time-averaged statistics [51, 52]. The successful passive control of low-frequency, aft-angle jet noise with nozzle modification further supports controllability of these large-scale flow structures [53–57]. However, representing the real-time dynamics of such coherent structures with  
95 a reduced model remains challenging, especially for novel configurations. The underlying nonlinearity of the turbulence seems to manifest as intermittency of the large-scale coherent structure, making complete separation of their dynamics difficult [50]. Many ansatzes are suggested for useful bases to extract these structures [50, 58]; however, none of them rival the predictive accuracy of  
100 a well-resolved simulation.

Another example is passive control of thermoacoustic oscillations in rocket engines and gas turbines [6, 16, 59, 60]. Although the turbulence is intense [16, 60], sensitivity and optimization need focus only on the key mechanism of the thermoacoustic instability, with turbulence only seeding the instabilities [6, 16,  
105 59]. However, obtaining further prognostic information to prevent the instability requires analysis of fluctuations in turbulent combustors, which includes their chaotic behavior [60]. Intermittency and multi-dimensional chaos in turbulent combustion dynamics has been extensively studied and measured [61–64], but a design optimization harnessing this understanding in the form of a predictive  
110 model remains a goal.

#### *1.4. Overview*

Given the challenges and potential opportunities, we develop a procedure that circumvents non-convex features during the optimization, leaving the exact governing equation intact. To develop the method, we quantify the impact

115 of chaos on the optimization of both model systems and turbulence. Error  
amplification in the gradient computation has been presented as the main chal-  
lenge [9, 12, 26, 27], for which the Lyapunov exponent is a well-established  
metric [65, 66]. However,  $\mathcal{J}$  reduction is also hindered by non-convexity [2, 28].  
In the spirit of entropy and fractal dimension [67–71], we develop a useful indi-  
120 cator for non-convexity to assess the challenge based on a range of utility in the  
parameter space [48].

The method overall is an extension of standard adjoint gradient-based meth-  
ods for equality-constrained optimization, which is compactly formulated in Sec-  
tion 2. With the turbulence problem in mind, Section 3 uses the Lorenz system  
125 to quantify two ways that chaos impedes gradient-based optimization. The met-  
rics **introduced** are used to illustrate opportunities that can also be exploited in  
turbulence. The new multi-point penalty-based method is proposed in Section 4,  
where it is demonstrated with the Lorenz equation example. It is then shown in  
Section 5 to far outperform the standard approach for an advection augmented  
130 one-dimensional Kuramoto–Sivashinsky Equation (adv+KS) [72, 73], **which is**  
**constructed to have an exact reduced model available**. The same behavior is  
confirmed in Section 6 for three-dimensional turbulent Kolmogorov flow.

## 2. Optimal control formulation

### 2.1. Equality-constrained optimization

A governing equation,

$$\mathcal{N}[\mathbf{q}; \Theta] = \mathbf{0}, \quad \mathcal{N} \in \mathbb{N}, \quad (2)$$

provides an equality constraint as the state  $\mathbf{q} \in \mathbb{Q}$  and control input  $\Theta \in \mathbb{T}$   
are optimized. **Specific  $\mathbf{q}$  and  $\Theta$  will defined for each dynamical system we**  
**consider**. We use  $\mathbb{N}$ ,  $\mathbb{Q}$ , and  $\mathbb{T}$  to denote the spaces on which  $\mathcal{N}$ ,  $\mathbf{q}$ , and  $\Theta$  are  
respectively defined. In general,  $\mathbb{N} \equiv \mathbb{Q}$ , so  $\mathbf{q}$  is determined completely by  $\mathcal{N}$ .  
The control space  $\mathbb{T}$  in general can be defined independently of  $\mathbb{Q}$ . Both  $\mathbb{Q}$  and

$\mathbb{T}$  are Hilbert spaces. Optimization seeks to minimize scalar objective functional  $\mathcal{J} : \mathbb{Q} \times \mathbb{T} \rightarrow \mathbb{R}$ ,

$$\text{minimize } \mathcal{J}[\mathbf{q}, \Theta] \in \mathbb{R} \quad \text{such that } \mathcal{N}[\mathbf{q}; \Theta] = \mathbf{0}. \quad (3)$$

135 Significantly improving  $\mathcal{J}$  is often sufficient, so we forego the greater challenge of seeking a global optimum.

## 2.2. Standard adjoint-based optimization

Extending the method of Lagrange multipliers [22–24] to equality-constrained optimization leads to standard adjoint-based optimization [6, 8–12, 21]. For (2) a differential equation in time, it can be expressed

$$\mathcal{N}[\mathbf{q}; \Theta] \equiv \frac{\partial \mathbf{q}}{\partial t} - \mathcal{R}[\mathbf{q}; \Theta] = 0, \quad (4a)$$

with initial condition

$$\mathbf{q}(t_i) = \mathbf{q}_i. \quad (4b)$$

We define the inner product for  $\mathbb{Q}$  as

$$\langle \mathbf{p}, \mathbf{q} \rangle_{\mathbb{Q}} = \int_{t_i}^{t_f} \langle \mathbf{p}(t), \mathbf{q}(t) \rangle_{\mathbb{Q}^+} dt, \quad (5)$$

with  $t_i, t_f$  the initial and final times, and the inner product  $\langle \mathbf{p}(t), \mathbf{q}(t) \rangle_{\mathbb{Q}^+}$  for the subspace  $\mathbb{Q}^+$  of the instantaneous state  $\mathbf{q}(t)$  at any time  $t$ . For the Lorenz example (1),  $\mathbf{q}(t) = (x, y, z) \in \mathbb{Q}^+ = \mathbb{R}^3$  and  $\mathbb{Q} = U^3$  where  $U = H^0(\mathbb{R}_0^+)$  is the space of  $L^2$ -functions from  $[0, \infty)$  to  $\mathbb{R}$ , and the instantaneous inner product is

$$\langle \mathbf{p}(t), \mathbf{q}(t) \rangle_{\mathbb{Q}^+} = \mathbf{p}^T(t) \mathbf{q}(t). \quad (6)$$

Inner products for  $\mathbb{T}$  are defined similarly,

$$\langle \Theta, \Psi \rangle_{\mathbb{T}} = \int_{t_i}^{t_f} \langle \Theta(t), \Psi(t) \rangle_{\mathbb{T}^+} dt, \quad (7)$$

with the subspace  $\mathbb{T}^+$  for instantaneous control.

Optimal control seeks the solution  $\mathbf{q}$  and control  $\Theta$  that minimizes

$$\mathcal{J}[\mathbf{q}, \Theta] \equiv \Phi[\mathbf{q}(t_f)] + \int_{t_i}^{t_f} \mathcal{I}[\mathbf{q}(t), \Theta(t)] dt, \quad (8a)$$

preserving

$$\frac{d\mathbf{q}}{dt} - \mathcal{R}[\mathbf{q}; \Theta] = \mathbf{0}, \quad (8b)$$

where  $\Phi$  is an objective functional for the final state  $\mathbf{q}(t_f)$ , and  $\mathcal{I}$  is an instantaneous analog of  $\mathcal{J}$ . **Specific  $\Phi$  and  $\mathcal{I}$  are introduced in context of specific examples in Section 3, 5 and 6.** The corresponding Lagrangian,

$$\begin{aligned} \mathcal{L}[\mathbf{q}, \mathbf{q}^\dagger, \Theta] &= \mathcal{J}[\mathbf{q}, \Theta] - \left\langle \mathbf{q}^\dagger, \frac{d\mathbf{q}}{dt} - \mathcal{R}[\mathbf{q}; \Theta] \right\rangle_{\mathbb{Q}} \\ &= \Phi[\mathbf{q}(t_f)] + \int_{t_i}^{t_f} \mathcal{I}[\mathbf{q}(t), \Theta(t)] dt \\ &\quad - \int_{t_i}^{t_f} \left\langle \mathbf{q}^\dagger, \frac{d\mathbf{q}}{dt} - \mathcal{R}[\mathbf{q}; \Theta] \right\rangle_{\mathbb{Q}^+} dt, \end{aligned} \quad (9)$$

includes the adjoint variable  $\mathbf{q}^\dagger \in \mathbb{Q}$ . A local minimum of (8) is a stationary point of  $\mathcal{L}$ , where its gradients with respect to  $\mathbf{q}$ ,  $\mathbf{q}^\dagger$ , and  $\Theta$  are zero [22, 23, 74].

Recasting the first-order variation of  $\mathcal{L}$  yields

$$\begin{aligned} \delta\mathcal{L} &= \left\langle \frac{\partial\Phi}{\partial\mathbf{q}}, \delta\mathbf{q}(t_f) \right\rangle_{\mathbb{Q}^+} + \left\langle \frac{\partial\mathcal{I}}{\partial\mathbf{q}}, \delta\mathbf{q} \right\rangle_{\mathbb{Q}} + \left\langle \frac{\partial\mathcal{I}}{\partial\Theta}, \delta\Theta \right\rangle_{\mathbb{T}} \\ &\quad - \left\langle \mathbf{q}^\dagger, \frac{d\delta\mathbf{q}}{dt} - \frac{\partial\mathcal{R}}{\partial\mathbf{q}}\delta\mathbf{q} - \frac{\partial\mathcal{R}}{\partial\Theta}\delta\Theta \right\rangle_{\mathbb{Q}} \\ &\quad - \left\langle \delta\mathbf{q}^\dagger, \frac{d\mathbf{q}}{dt} - \mathcal{R}[\mathbf{q}; \Theta] \right\rangle_{\mathbb{Q}}, \end{aligned} \quad (10)$$

where  $\frac{\partial\mathcal{I}}{\partial\mathbf{q}} \in \mathbb{Q}$  and  $\frac{\partial\mathcal{I}}{\partial\Theta} \in \mathbb{T}$ , and  $\frac{\partial\mathcal{R}}{\partial\mathbf{q}} : \mathbb{Q} \rightarrow \mathbb{Q}$  and  $\frac{\partial\mathcal{R}}{\partial\Theta} : \mathbb{T} \rightarrow \mathbb{Q}$  are defined in the sense of a Frechét derivative. Integration by parts yields the adjoint of  $\frac{\partial\mathcal{R}}{\partial\mathbf{q}}$  as  $\frac{\partial\mathcal{R}}{\partial\mathbf{q}}^\dagger$  [75, 76],

$$\begin{aligned} \delta\mathcal{L} &= \\ &\left. \begin{aligned} & - \left\langle \delta\mathbf{q}^\dagger, \frac{d\mathbf{q}}{dt} - \mathcal{R}[\mathbf{q}; \Theta] \right\rangle_{\mathbb{Q}} \end{aligned} \right\} \text{Governing equation} \\ &\left. \begin{aligned} & + \left\langle \frac{\partial\Phi}{\partial\mathbf{q}} - \mathbf{q}^\dagger(t_f), \delta\mathbf{q}(t_f) \right\rangle_{\mathbb{Q}^+} \\ & + \left\langle \frac{d\mathbf{q}^\dagger}{dt} + \frac{\partial\mathcal{R}^\dagger}{\partial\mathbf{q}} \mathbf{q}^\dagger + \frac{\partial\mathcal{I}}{\partial\mathbf{q}}, \delta\mathbf{q} \right\rangle_{\mathbb{Q}} \end{aligned} \right\} \text{Adjoint equation} \end{aligned} \quad (11)$$

$$\left. \begin{aligned} & + \left\langle \frac{\partial \mathcal{R}^\dagger}{\partial \Theta} \mathbf{q}^\dagger + \frac{\partial \mathcal{I}}{\partial \Theta}, \delta \Theta \right\rangle_{\mathbb{T}} \\ & + \langle \mathbf{q}^\dagger(t_i), \delta \mathbf{q}(t_i) \rangle_{\mathbb{Q}^+} \end{aligned} \right\} \text{Gradient}$$

The state  $\mathbf{q}$  and control  $\Theta$  must satisfy the governing equation (8b) so the first term in (11) is zero. The second and third terms are also zero if  $\mathbf{q}^\dagger$  satisfies

$$\frac{d\mathbf{q}^\dagger}{dt} + \frac{\partial \mathcal{R}^\dagger}{\partial \mathbf{q}} \mathbf{q}^\dagger + \frac{\partial \mathcal{I}}{\partial \mathbf{q}} = \mathbf{0} \quad (12a)$$

$$\mathbf{q}^\dagger(t_f) = \frac{\partial \Phi}{\partial \mathbf{q}}, \quad (12b)$$

so  $\delta \mathcal{L}$  (11) depends only on the control variation  $\delta \Theta$  and the initial condition variation  $\delta \mathbf{q}(t_i)$ . This provides a gradient via

$$\text{Gradient} \quad \delta \mathcal{L} = \left\langle \frac{\partial \mathcal{R}^\dagger}{\partial \Theta} \mathbf{q}^\dagger + \frac{\partial \mathcal{I}}{\partial \Theta}, \delta \Theta \right\rangle_{\mathbb{T}} + \langle \mathbf{q}^\dagger(t_i), \delta \mathbf{q}(t_i) \rangle_{\mathbb{Q}^+}, \quad (13a)$$

where

$$\text{Governing equation} \left\{ \begin{aligned} & \frac{d\mathbf{q}}{dt} - \mathcal{R}[\mathbf{q}; \Theta] = \mathbf{0} \quad (13b) \\ & \mathbf{q}(t_i) = \mathbf{q}_i \quad (13c) \end{aligned} \right.$$

$$\text{Adjoint equation} \left\{ \begin{aligned} & \frac{d\mathbf{q}^\dagger}{dt} + \frac{\partial \mathcal{R}^\dagger}{\partial \mathbf{q}} \mathbf{q}^\dagger + \frac{\partial \mathcal{I}}{\partial \mathbf{q}} = \mathbf{0} \quad (13d) \\ & \mathbf{q}^\dagger(t_f) = \frac{\partial \Phi}{\partial \mathbf{q}}. \quad (13e) \end{aligned} \right.$$

Although our methods should be broadly adaptable to gradient-informed searches [22, 77, 78], we focus on conjugate gradient line searches [78]. For example, with the line search direction  $\delta \Theta_k$  set, the iteration  $k$  search step  $\alpha_k$  is sought via a one-dimensional minimization problem,

$$\alpha_k = \underset{\alpha}{\operatorname{argmin}} \mathcal{J}[\mathbf{q}, \Theta_{k-1} + \alpha \delta \Theta_k] \quad \text{for} \quad \mathcal{N}[\mathbf{q}_k; \Theta_{k-1} + \alpha \delta \Theta_k] = 0. \quad (14)$$

A typical inverse parabolic interpolation combined with golden-section search is used to stably estimate  $\alpha_k$  [78].

As formulated, the governing equation  $\mathcal{N}[\mathbf{q}; \Theta] = 0$  constraint exactly preserves  $\mathcal{L} \equiv \mathcal{J}$  throughout the optimization. In Section 3, we consider how this

145 strict constraint, combined with chaotic dynamical systems, limits exploration  
in a way that frustrates the search for a significantly better local minimum. Its  
temporary relaxation is a key component of the new algorithm in Section 4.

### 3. Chaos and gradient-based optimization

Two main aspects of how chaos disrupts gradient-based optimization are  
150 demonstrated for the Lorenz system [already introduced descriptively](#) in Sec-  
tion 1. Standard results from the theory of dynamical systems are used to  
anticipate how these also [present challenges for turbulent flow](#) control and mo-  
tivate our circumvention strategy. Flexible metrics are selected that can also be  
evaluated for turbulence simulations.

#### 155 3.1. Lorenz equation illustration

##### 3.1.1. System definition

The Lorenz equation (1) for the state  $\mathbf{q} = (x, y, z) \in \mathbb{Q} = U^3$  with ac-  
tuation  $\Theta = f(t) \in \mathbb{T} = U$ , is solved numerically with  $\sigma = 10$ ,  $\beta = 8/3$ ,  
and  $\rho = 28$  and  $\mathbf{q}_i = (1.49, 1.49, 37)^T$ . The unstable fixed points are  $U_{1,2} =$   
160  $(\pm 6\sqrt{2}, \pm 6\sqrt{2}, 27)^T$ .

##### 3.1.2. Control

To seek orbits of only  $U_1$ ,  $\mathcal{J}$  is defined

$$\mathcal{J} = \frac{1}{t_f - t_i} \int_{t_i}^{t_f} \mathcal{I}[\mathbf{q}] dt, \quad (15)$$

with

$$\mathcal{I}[\mathbf{q}] = \begin{cases} \frac{1}{2} \left( \frac{2x + y}{5} \right)^2 & 2x + y \geq 0 \\ 0 & \text{otherwise.} \end{cases} \quad (16)$$

The optimization period is  $t_i = 0$  to  $t_f = 20$ .

For the standard gradient-based optimization that was shown in Figure 1, the  
gradient of  $\mathcal{J}$  to  $f(t)$  grows exponentially in reverse time, and  $\mathcal{J}$  has numerous  
165 local extrema along a line in the parameter space. Starting from  $\mathbf{q}_i$  with  $f = 0$ ,

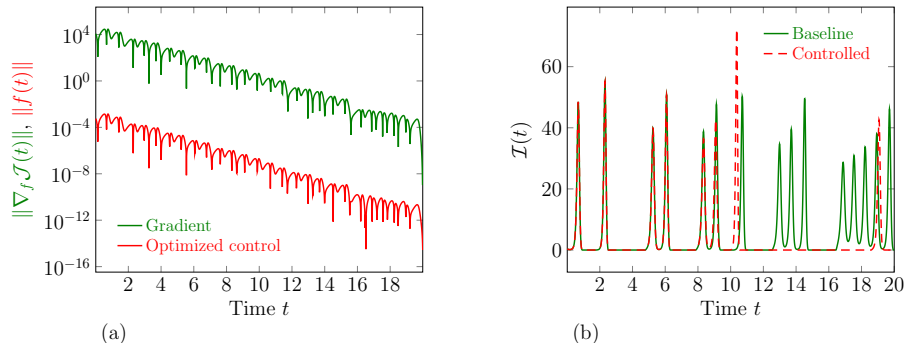


Figure 3: Standard gradient-based optimization for the Lorenz system (1): (a) the magnitude of the control gradient  $\nabla_f \mathcal{J}$  and the optimized control  $f(t)$ , and (b) the instantaneous objective functional  $\mathcal{I}(t)$  (16) for the baseline and controlled trajectories.

the standard gradient-based method finds an  $f$  that reduces  $\mathcal{J}$  by 44.6%, as shown in Figure 3.

### 3.1.3. Biased line search due to gradient growth

The exponentially amplified gradients at early times dominate the overall  
 170 gradient, leading to control being ineffectively concentrated near  $t = t_i$  (Figure 3 a). This masks the moments at which  $f$  can best minimize  $\mathcal{J}$ . Figure 3 (b) show how  $\mathcal{I}(t)$  is reduced only at late times, well removed in time from the strongest  $f(t)$ .

It might be tempting to say that a system showing this behavior is not  
 175 very controllable, but that would be wrong. A well-known nonlinear feedback control [79–81] shows that  $f(t)$  can be activated only when the term  $2x + y$  from (16) nears 0. This control suppressed all  $\mathcal{I}(t)$  peaks, confirming that late-time peaks of  $\mathcal{I}(t)$  do not need to exploit  $f(t)$  at early times. Similarly, the early-time  $\mathcal{I}$  peaks need not remain uncontrolled [81].

### 180 3.1.4. Non-convexity of $\mathcal{J}$

The  $f(t)$  for the standard gradient-based method is also significantly different from the nonlinear feedback control. The nonlinear feedback control has maximum  $|f(t)| = 4623$  and oscillates in  $-713 < f(t) < 378$  after  $t > 0.5$ , whereas

the standard gradient-based method has peak  $|f(t)| = 0.00141$  at  $t = 0.76$ .  
 185 While it is impressive that such a small control perturbation can achieve more  
 than 40% reduction, amplitude is not penalized so there is no reason for such  
 an extraordinarily low-amplitude control. It is the non-convexity of  $\mathcal{J}[\Theta]$  that  
 limits the control magnitude. The small features of  $\mathcal{J}[\Theta]$  in Figure 1 (b) impede  
 the gradient-based method from finding effective solutions far from the starting  
 190 point.

### 3.2. Signature of chaos: horseshoe mapping

The challenges of the previous section reflect well-known properties of chaotic  
 dynamical systems. It is simplest to describe these for discrete mappings from  
 one time to the next [65, 66, 82], which are analogous to solutions of (4) at  
 intervals,

$$\mathbf{q}_{n+1} = \mathbf{q}(\mathbf{q}_n, \Delta T) \equiv \mathbf{q}_n + \int_{t_i+n\Delta T}^{t_i+(n+1)\Delta T} \mathcal{R}[\mathbf{q}] dt. \quad (17)$$

The one-dimensional **logistic** map is a specific example of (17) [65],

$$q_{n+1} = q(q_n) \equiv (2q_n - 1)^2 \quad q_n \in [0, 1], \quad (18)$$

which is called a horseshoe map for the way it stretches and folds a distribution  
 of possible  $q_n$  to yield a  $q_{n+1}$  distribution [66, 82]. These are the two essential  
 aspects of chaotic dynamical systems. The stretching at each step leads to the  
 195 exponential growth of the gradient discussed in Section 3.1. The folding creates  
 new local extrema. Figure 4 (a) shows subsequent states  $q_n(q_0)$ , whose number  
 of local extrema doubles at every mapping.

The increasing number of local extrema blocks routes to any useful (small  
 $\mathcal{J}$ ) local minima that lay outside a small neighborhood of the starting point.  
 The simple objective function

$$\mathcal{J} = q_2 + \frac{1}{2}q_0 \quad (19)$$

illustrates this. The two mappings to  $q_2$  generate the three extrema shown in  
 Figure 4 (b). The local maximum at  $q_0 = 0.5$  blocks the path to  $\mathcal{J}_1$  for  $q_0 > 0.5$ .

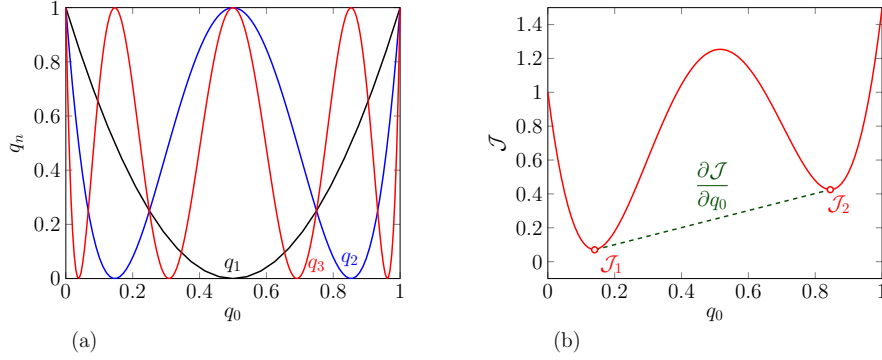


Figure 4: **Logistic** map example (18): (a) states  $q_1$ ,  $q_2$  and  $q_3$  as functions of the initial state  $q_0$ , and (b) the example objective function  $\mathcal{J}(q_0)$  (19).

200 The space of subsequent states becomes so non-convex by horseshoe mapping, that any objective functional dependent on them will also reflect this character.

### 3.3. Quantification of the impact of chaos on optimization

To guide design of the new framework, we quantify both the exponential growth of sensitivity and restriction of the search step size due to the increasing non-convexity of  $\mathcal{J}$ , recognizing that these metrics need to be computable even in a turbulent flow.

The Lyapunov exponent is, of course, a common measure of the gradient growth [65, 66]. The leading Lyapunov exponent is straightforward to estimate from numerical calculation, typically by averaging the growth rate of random adjoint states between times  $t_i$  and  $t_f$  during the course of the simulation [83–85]. Samples

$$\tilde{\lambda}_1(t_{i,n}, t_{f,n}) = \frac{1}{t_{f,n} - t_{i,n}} \log \frac{\|\mathbf{q}^\dagger(t_{i,n})\|_{\mathbb{Q}^+}}{\|\mathbf{q}^\dagger(t_{f,n})\|_{\mathbb{Q}^+}} \quad \text{for } n = 1, \dots, N \quad (20)$$

are averaged to provide an estimate of the gradient amplification rate  $\lambda_1$  and corresponding  $e$ -folding time  $\tau_\lambda$  [86],

$$\lambda_1 \equiv \frac{1}{\tau_\lambda} = \frac{1}{N} \sum_{n=1}^N \tilde{\lambda}_1(t_{i,n}, t_{f,n}). \quad (21)$$

The step size restriction due to non-convexity is estimated based upon the range for which the  $\frac{\partial \mathcal{J}}{\partial \Theta}$  gradient provides an accurate estimate of the behavior of  $\mathcal{J}$ . This can be estimated by comparing with a finite-difference approximation of the gradient,

$$\frac{\Delta \mathcal{J}}{\Delta \Theta} = \frac{\mathcal{J}[\mathbf{q}; \Theta_0 + \Delta \Theta \mathbf{e}_\theta] - \mathcal{J}[\mathbf{q}; \Theta_0]}{\Delta \Theta}. \quad (22)$$

In (22),  $\mathbf{e}_\theta = \frac{\partial \mathcal{J}}{\partial \Theta} / \|\frac{\partial \mathcal{J}}{\partial \Theta}\|_{\mathbb{T}}$  is the unit vector in  $\mathbb{T}$  along the gradient direction. A corresponding error

$$\epsilon[\Delta \Theta] = \left| \frac{\frac{\Delta \mathcal{J}}{\Delta \Theta} - \langle \frac{\partial \mathcal{J}}{\partial \Theta}, \mathbf{e}_\theta \rangle_{\mathbb{T}}}{\langle \frac{\partial \mathcal{J}}{\partial \Theta}, \mathbf{e}_\theta \rangle_{\mathbb{T}}} \right| \quad (23)$$

quantifies how accurately the gradient predicts the actual variation  $\Delta \mathcal{J}$  for a finite step  $\Delta \Theta$ . The growth of finite-precision errors sets a floor for numerical accuracy. However, even when the precision threshold is not reached, utility is  
 210 limited by a viable step size, which decreases exponentially in (reverse) time.

A nominal viable step size  $\delta \Theta$  is defined based on when a nominal objective functional  $\Phi = \langle \mathbf{q}^\dagger(t_f), \mathbf{q}(t_f) \rangle_{\mathbb{Q}^+}$  has a relative error  $\epsilon = 0.5$ . Step size restriction—a surrogate for increasing non-convexity—is estimated by the decay of  $\delta \Theta$  over a time interval  $t_f - t_i$

$$\tilde{\phi}(\Phi, \mathbf{q}(t_i), t_i, t_f) = -\frac{\log \delta \Theta[\Phi, \mathbf{q}(t_i), t_i, t_f]}{t_f - t_i}. \quad (24)$$

To sample the entire state distribution, we ensemble-average  $\tilde{\phi}$  (24) to define non-convexity growth time-scale estimate  $\tau_\phi$ ,

$$\frac{1}{\tau_\phi} \equiv \phi = \frac{1}{N} \sum_{n=1}^N \tilde{\phi}(\Phi_n, \mathbf{q}(t_{i,n}), t_{i,n}, t_{f,n}), \quad (25)$$

where  $\Phi_n = \langle \mathbf{q}^\dagger(t_{f,n}), \mathbf{q}(t_{f,n}) \rangle_{\mathbb{Q}^+}$ .

### 3.4. Application to the model systems and turbulence

The  $e$ -folding time (21) of the Lorenz equation (1) is estimated with  $N = 500$  adjoint final states,

$$\mathbf{q}^\dagger(t_f) = (\xi_x, \xi_y, \xi_z), \quad (26)$$

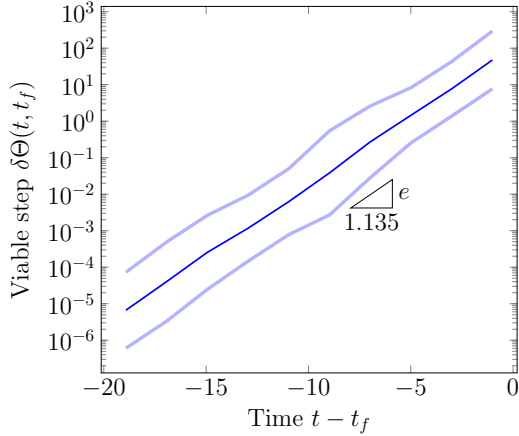


Figure 5: Viable step for the Lorenz example.

where  $\xi_x$ ,  $\xi_y$ , and  $\xi_z$  are pseudo-random numbers with uniform distribution  $U[-10^{-5}, 10^{-5}]$ . The adjoint initial state  $\mathbf{q}^\dagger(t_i)$  is computed from each adjoint final state per (13d) with  $\mathcal{I} = \mathbf{0}$ , for 500 consecutive intervals of length  $t_f - t_i = 20$ . A geometric linear fit of the time evolution of the ensemble average,

$$\overline{\mathbf{q}^\dagger} = \left( \prod_{k=1}^N \|\mathbf{q}_k^\dagger\|_{\mathbb{Q}^+} \right)^{\frac{1}{N}} \quad (27)$$

estimates  $\tau_\lambda \approx 1.11$ , and the predicted gradient increase  $\exp[(t_f - t_i)/\tau_\lambda] \approx 7 \times 10^7$  is indeed consistent with the observation in Figure 3 (a). The sample final  
 215 adjoint states (26) are reused to evaluate  $\Phi = \langle \mathbf{q}^\dagger(t_f), \mathbf{q}(t_f) \rangle_{\mathbb{Q}^+}$  for estimating  $\tau_\phi$ . Per (13) with  $\mathcal{I} = \mathbf{0}$  and  $\Theta = \mathbf{0}$ , the gradient of  $\Phi$  with respect to  $\mathbf{q}(t_i)$  is equivalent to  $\frac{\partial \Phi}{\partial \mathbf{q}(t_i)} = \mathbf{q}^\dagger(t_i)$ . This  $\mathbf{q}^\dagger(t_i)$  is already computed in (13d) for estimating  $\tau_\lambda$ . The viable step  $\delta\Theta(t, t_f)$  at varying initial times  $t_i = t$  are ensemble-averaged with  $N = 100$  samples in Figure 5. The fitted decay time  
 220 scale (25) is  $\tau_\phi = 1.14$ . The actual gradient growth and inferred viable step for  $\mathcal{J}$  (15) with control  $f(t)$  in (1) confirms that these  $\tau_\lambda$  and  $\tau_\phi$  estimated from state perturbations quantify the impact of chaos on the optimization.

With its sole positive Lyapunov exponent, the similarity of  $\tau_\phi = 1.14$  to  $\tau_\lambda = 1.11$  is expected for the Lorenz system, although this is not necessarily  
 225 the case for more complex systems. The  $\tau_\lambda/\tau_\phi$  ratio for several more complex

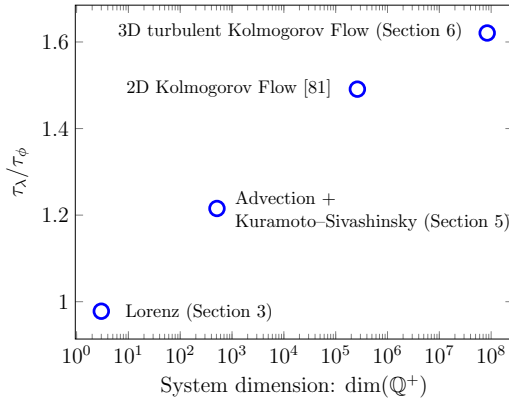


Figure 6: The ratio between inferred  $\epsilon$ -folding time  $\tau_\lambda$  and viable step decay time  $\tau_\phi$ .

examples considered subsequently and elsewhere are shown in Figure 6. These all have  $\tau_\phi < \tau_\lambda$ , suggesting that increasing non-convexity is the greater concern.

#### 4. Multi-step penalty method

The proposed optimization framework temporarily suppresses the non-convexity of  $\mathcal{J}$  to escape blocking local minima. It is a quadratic penalty augmentation of the equality-constrained optimization framework introduced in Section 2. The formulation developed in Sections 4.1 through 4.3 is first applied for illustration to the **logistic** map example of Section 3.2. It is then demonstrated for the Lorenz example in Section 4.4. Sections 5 and 6 will cover its application to chaotic advection and turbulence.

##### 4.1. Penalty

Penalty methods replace equality-constrained optimization (3) for  $\mathcal{J}$ , with a sequence of unconstrained optimizations,

$$\text{minimize } \mathcal{J}_A[\mathbf{q}, \Theta, \mu] \equiv \mathcal{J}[\mathbf{q}, \Theta] + \mathcal{P}[\mathcal{N}[\mathbf{q}, \Theta], \mu] \in \mathbb{R}, \quad (28)$$

where the functional  $\mathcal{P}$  penalizes violation of  $\mathcal{N}[\mathbf{q}, \Theta] = 0$  with strength  $\mu > 0$  [22, 23]. Each subproblem (28) is solved with standard gradient-based meth-

ods. Direct differentiation of the residual  $\mathcal{N}$  provides a formulation nearly the  
 240 same as the adjoint-based gradient.

The quadratic penalty method uses a squared norm based on the equality  
 constraint [22, 23, 77],

$$\mathcal{P}[\mathcal{N}[\mathbf{q}, \Theta], \mu] = \frac{\mu}{2} \|\mathcal{N}[\mathbf{q}, \Theta]\|^2, \quad (29)$$

with  $\|\cdot\|$  defined for  $\mathbb{N} \equiv \mathbb{Q}$ . Sequentially solving (28) with (29) for  $\mu \rightarrow$   
 $\infty$  [22, 23] converges to a local optimum. The quadratic penalty method is suf-  
 ficient for our purposes, though it is straightforward to formulate an augmented  
 Lagrangian method [22, 23, 81].

245 On its own, however, the penalty method is insufficient. Since it relaxes  
 the entire equation  $\mathcal{N}[\mathbf{q}, \Theta] = 0$ ,  $\mathbf{q}$  is fully included in the optimization space,  
 independently of  $\Theta$ , which significantly increases the optimization dimension.  
 A key to our approach is to only modestly increase the dimension. To do this  
 we restrict violation of  $\mathcal{N} = 0$  to particular times approximately separated by  
 250  $\tau_\phi$  (and  $\tau_\lambda$ ). This is the essential component of the proposed method.

#### 4.2. Illustration for the *logistic* map of Section 3

Figure 7 (a) shows (19) with the basin of attraction for the global minimum  
 $\min \mathcal{J}_1$ : gradient-based optimization for any initial  $q_0$  outside this basin  $q_0 \in$   
 $[0, 0.515]$  converges to the other local minimum  $\min \mathcal{J}_2$ . This is emblematic of  
 255 the local minima landscape that the horseshoe mapping creates (Section 3).

To expose a path that reaches  $\min \mathcal{J}_1$  for any  $q_0$ , we introduce an intermedi-  
 ate state  $q_1^+$ , which corresponds to a possible  $q_1$ , though not one that is strictly  
 tied to the current  $q_0$ . Instead of viewing  $q_2 \equiv q(q(q_0))$  as a direct function of  
 $q_0$ , we replace  $q_1 \equiv q(q_0)$  with  $q_1^+$  and consider  $\mathcal{J}$  as it depends on  $q_0$  and  $q_1^+$ ,

$$\mathcal{J}(q_0, q_1^+) = q_2(q_1^+) + \frac{1}{2}q_0 = (2q_1^+ - 1)^2 + \frac{1}{2}q_0, \quad (30a)$$

with an additional equality constraint for the intermediate state

$$q_1^+ = q_1(q_0). \quad (30b)$$

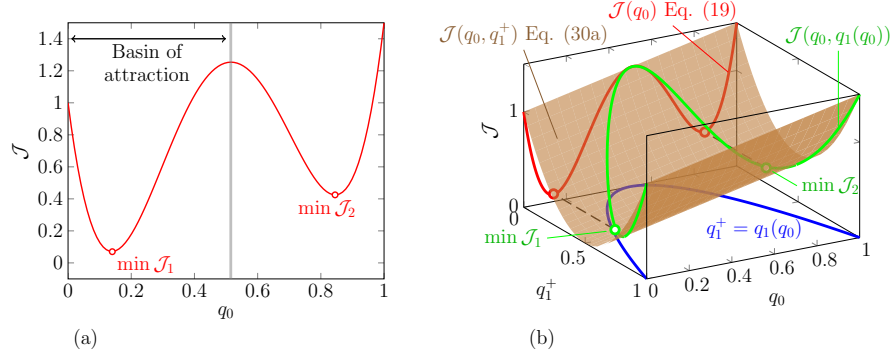


Figure 7: **Logistic** map example: (a) the basin of attraction for the global minimum of  $\mathcal{J}$  (19), and (b) the corresponding objective  $\mathcal{J}$  (30a) in  $(q_0, q_1^+)$  coordinates.

This is an exact rearrangement of the original problem, however without the intermediate constraint (30b), the two-dimensional objective functional  $\mathcal{J}(q_0, q_1^+)$  has the unique global minimum at  $(q_0, q_1^+) = (0, 0.5)$ , as shown in Figure 7 (b). The basin of attraction for the global minimum is the entire  $(q_0, q_1^+)$ -domain. Only the constraint  $q_1^+ = q_1(q_0)$  limits the basin of attraction. With this constraint,  $\mathcal{J}(q_0, q_1(q_0))$  is a curve through the two-dimensional space, shown as the green line in Figure 7 (b). Three extrema (two minima) are created in the green line  $\mathcal{J}(q_0, q_1(q_0))$ , as the blue curve  $q_1^+ = q_1(q_0)$  is projected and folded onto the surface  $\mathcal{J}(q_0, q_1)$ . The result is the non-convex  $\mathcal{J}(q_0)$  of Figure 7 (a): while  $\mathcal{J}$  is defined to be convex on  $\mathbb{Q}(q_0, q_1^+)$ , the set of feasible  $(q_0, q_1^+)$  (constrained by  $\mathcal{N} \equiv q_1^+ - q_1(q_0) = 0$ ) becomes non-convex, thus limiting the neighborhood of a local minimum.

The strategy is therefore to relax this intermediate constraint (30b) to expand the basin of attraction for  $\mathcal{J}(q_0, q_1^+)$  as in Figure 7 (b), avoiding the confinement introduced by the horseshoe mapping. To do this the objective functional is augmented with the penalty,

$$\mathcal{J}_A = (2q_1^+ - 1)^2 + \frac{1}{2}q_0 + \frac{\mu}{2}\{q_1^+ - (2q_0 - 1)^2\}^2, \quad (31)$$

and minimized over  $(\mathbf{q}, \Theta) = (q_1^+, q_0)$ . The intermediate state  $q_1^+$  is optimized together with  $q_0$ , until  $q_1^+ \rightarrow q_1 \equiv (2q_0 - 1)^2$  as  $\mu$  increases.

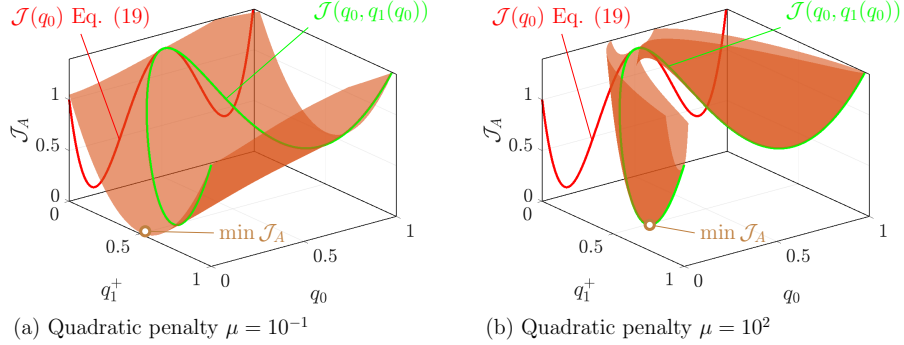


Figure 8: Logistic map demonstration objective functionals,  $\mathcal{J}$  for the initial problem and  $\mathcal{J}_A$  (31) for the quadratic penalty method with (a)  $\mu = 10^{-1}$  and (b)  $\mu = 10^2$ .

270 Figure 8 illustrates how this circumvents the non-convexity  $\mathcal{J}$ . In Figure 8 (a),  $\mathcal{J}_A$  with a weak  $\mu = 10^{-1}$  has a nearly identical shape to  $\mathcal{J}(q_0, q_1^+)$  (30a) in Figure 7 (b). So the entire domain  $(q_0, q_1^+)$  is still the basin of attraction for its global minimum. As  $\mu$  increases,  $\mathcal{J}_A$  converges to the original  $\mathcal{J}(q_0, q_1(q_0))$  as shown in Figure 8 (b). This, of course, has the similarly limited  
275 basin of attraction as the original  $\mathcal{J}(q_0)$ . However, if used correctly, the optimized  $(q_0, q_1^+)$  will have already entered the  $\mathcal{J}_1$  basin before  $\mu$  is increased. For any initial  $(q_0, q_1^+)$ , the local minimizer of  $\mathcal{J}_A$  converges to the global minimizer of  $\mathcal{J}$  in (19) as  $\mu$  increases.

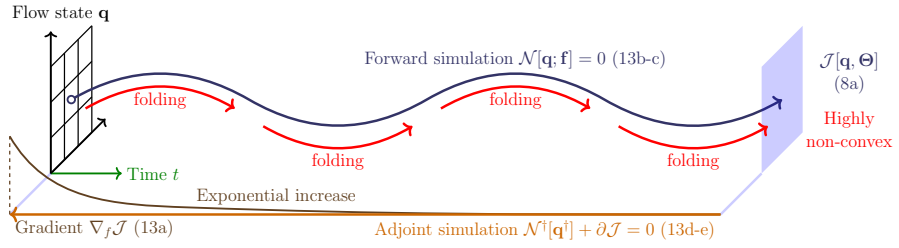
### 4.3. Formulation for time-continuous dynamical systems

#### 280 4.3.1. Added intermediate constraints

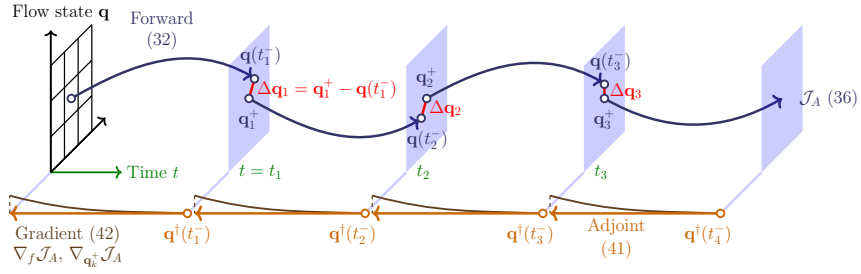
This is extended to time-continuous dynamical systems by adding auxiliary intermediate conditions to the state  $\mathbf{q}$ . Starting from the time-explicit governing equation (4), intermediate conditions  $\mathbf{q}_k^+$  are introduced at times  $t_k = t_i + kT$ ,

$$\frac{d\mathbf{q}}{dt} - \mathcal{R}[\mathbf{q}; \Theta] = \mathbf{0} \quad \text{for } t \in [t_k, t_{k+1}) \quad (32a)$$

$$\mathbf{q}(t_k) = \mathbf{q}_k^+ \quad \text{for } k = 0, \dots, N-1. \quad (32b)$$



(a) Standard gradient-based method



(b) Multi-step penalty-based method

Figure 9: Schematic of (a) standard gradient-based optimization and (b) multi-step penalty-based optimization with three intermediate conditions.

In summary, the intermediate constraints analogous to (30b) are

$$\Delta \mathbf{q}_k \equiv \mathbf{q}_k^+ - \mathbf{q}(t_k^-) = 0 \quad \text{for } k = 1, \dots, N - 1, \quad (32c)$$

where  $\mathbf{q}(t_k^-)$  is the terminal state of the previous interval  $t \in [t_{k-1}, t_k)$ ,

$$\mathbf{q}(t_k^-) = \mathbf{q}_{k-1}^+ + \int_{t_{k-1}}^{t_k^-} \mathcal{R}[\mathbf{q}; \Theta] dt. \quad (33)$$

Figure 9 schematically illustrates the original and modified systems. Only the (32c) constraint is relaxed; the initial condition (32b) and the governing equation  $\mathcal{N} = 0$  (32a) within the intervals are strictly enforced **so solutions within intervals are as physical as the original description**. The resulting state trajectory  $\mathbf{q} \in \mathbb{Q}$  is piecewise-continuous with intermediate discontinuities  $\Delta \mathbf{q}_k$  at each  $t = t_k$ .

For the adjoint formulation, the inner-product for the state space  $\mathbb{Q}$  is modified with sub-inner-products for piecewise-continuous trajectories,

$$\langle \mathbf{p}, \mathbf{q} \rangle_{\mathbb{Q}} = \sum_{k=1}^N \langle \mathbf{p}, \mathbf{q} \rangle_{\mathbb{Q}_k}, \quad (34)$$

where the subspaces  $\mathbb{Q}_k \subset \mathbb{Q}$  cover time intervals  $t \in [t_{k-1}, t_k)$ ,

$$\langle \mathbf{p}, \mathbf{q} \rangle_{\mathbb{Q}_k} = \int_{t_{k-1}}^{t_k^-} \langle \mathbf{p}(t), \mathbf{q}(t) \rangle_{\mathbb{Q}^+} dt. \quad (35)$$

Similar time-splitting is used in multiple shooting methods for various problems [45, 87–89]. A key difference of our framework is the combination with the penalty method, where increasing penalty strength  $\mu$  gradually restricts search steps, presumably after reaching the neighborhood of a useful local minimum. The multiple shooting method has been mostly used to find unstable periodic orbits of chaotic dynamical systems [87–89]. Blonigan *et al.* [45] used a multiple shooting method to compute ergodic sensitivities of chaotic dynamical systems, where optimization is pursued for the linearized system with a filtering parameter. This filtering parameter is similar to the inverse penalty strength  $1/\mu$ , allowing discontinuities at intermediate time points. However, in their procedure this parameter is fixed, determining the level of compromise between better conditioning and accuracy.

#### 4.3.2. Penalty-based optimization

An augmented objective functional  $\mathcal{J}_A$  (28) penalizes the intermediate constraints (32c),

$$\mathcal{J}_A[\mathbf{q}, \Theta; \{\mathbf{q}_k^+\}, \mu] = \mathcal{J}[\mathbf{q}, \Theta] + \mathcal{P}[\{\Delta\mathbf{q}_k\}, \mu], \quad (36)$$

with  $\{\mathbf{q}_k^+\} = (\mathbf{q}_1^+, \mathbf{q}_2^+, \dots, \mathbf{q}_{N-1}^+)$  and  $\{\Delta\mathbf{q}_k\} = (\Delta\mathbf{q}_1, \Delta\mathbf{q}_2, \dots, \Delta\mathbf{q}_{N-1})$ . The subproblem is then formulated as

$$({}_i\{\mathbf{q}_k^+\}, {}_i\Theta) = \underset{\{\mathbf{q}_k^+\}, \Theta}{\operatorname{argmin}} \mathcal{J}_A[\mathbf{q}, \Theta; \{\mathbf{q}_k^+\}, i\mu], \quad \text{with (32a) and (32b)}, \quad (37)$$

300 so the jumps are now included in the optimization. It is solved via a gradient-based method that is formulated in the following subsection.

#### 4.3.3. Adjoint-based gradient for the subproblem (37)

The Lagrangian associated with  $\mathcal{J}_A$  (36) is

$$\begin{aligned} \mathcal{L}_A[\mathbf{q}, \mathbf{q}^\dagger, \Theta] &= \mathcal{J}[\mathbf{q}, \Theta] + \mathcal{P}[\{\Delta\mathbf{q}_k\}, \mu] - \left\langle \mathbf{q}^\dagger, \frac{d\mathbf{q}}{dt} - \mathcal{R}[\mathbf{q}; \Theta] \right\rangle_{\mathbb{Q}} \\ &= \mathcal{J}[\mathbf{q}, \Theta] + \mathcal{P}[\{\Delta\mathbf{q}_k\}, \mu] - \sum_{k=1}^N \left\langle \mathbf{q}^\dagger, \frac{d\mathbf{q}}{dt} - \mathcal{R}[\mathbf{q}; \Theta] \right\rangle_{\mathbb{Q}_k}. \end{aligned} \quad (38)$$

Linearizing it and formulating its adjoint yields

$$\begin{aligned} \delta\mathcal{L}_A[\mathbf{q}, \mathbf{q}^\dagger, \Theta] &= \\ &\left. \begin{aligned} & - \sum_{k=1}^N \left\langle \delta\mathbf{q}^\dagger, \frac{d\mathbf{q}}{dt} - \mathcal{R}[\mathbf{q}; \Theta] \right\rangle_{\mathbb{Q}_k} \\ & - \langle \mathbf{q}^\dagger(t_i), \delta\mathbf{q}(t_i) \rangle_{\mathbb{Q}_+} \end{aligned} \right\} \text{Governing equation} \\ &\left. \begin{aligned} & - \left\langle \mathbf{q}^\dagger(t_f) - \frac{\partial\Phi}{\partial\mathbf{q}(t_f)}, \delta\mathbf{q}(t_f) \right\rangle_{\mathbb{Q}_+} \\ & - \sum_{k=1}^{N-1} \left\langle \mathbf{q}^\dagger(t_k^-) - \frac{\partial\mathcal{P}}{\partial\mathbf{q}(t_k^-)}, \delta\mathbf{q}(t_k^-) \right\rangle_{\mathbb{Q}_+} \\ & + \sum_{k=1}^N \left\langle \frac{d\mathbf{q}^\dagger}{dt} + \frac{\partial\mathcal{R}^\dagger}{\partial\mathbf{q}} \mathbf{q}^\dagger + \frac{\partial\mathcal{I}}{\partial\mathbf{q}}, \delta\mathbf{q} \right\rangle_{\mathbb{Q}_k} \end{aligned} \right\} \text{Adjoint equation} \quad (39) \end{aligned}$$

$$\left. \begin{aligned} & + \left\langle \frac{\partial \mathcal{R}^\dagger}{\partial \Theta} \mathbf{q}^\dagger + \frac{\partial \mathcal{I}}{\partial \Theta}, \delta \Theta \right\rangle_{\mathbb{T}} \\ & + \sum_{k=1}^{N-1} \left\langle \mathbf{q}^\dagger(t_k) + \frac{\partial \mathcal{P}}{\partial \mathbf{q}_k^+}, \delta \mathbf{q}_k^+ \right\rangle_{\mathbb{Q}^+}, \end{aligned} \right\} \text{Gradient}$$

where  $\frac{\partial \mathcal{P}}{\partial \mathbf{q}_k^+}$  and  $\frac{\partial \mathcal{P}}{\partial \mathbf{q}(t_k^-)}$  are weak-form gradients of  $\mathcal{P}$ , so its variation is

$$\delta \mathcal{P} = \sum_{k=1}^{N-1} \left[ \left\langle \frac{\partial \mathcal{P}}{\partial \mathbf{q}_k^+}, \delta \mathbf{q}_k^+ \right\rangle_{\mathbb{Q}^+} + \left\langle \frac{\partial \mathcal{P}}{\partial \mathbf{q}(t_k^-)}, \delta \mathbf{q}(t_k^-) \right\rangle_{\mathbb{Q}^+} \right]. \quad (40)$$

In (39), the first two inner products vanish because the governing equation (32a) is enforced and  $\delta \mathbf{q}(t_0) = \mathbf{0}$  since the initial condition in (32b) is fixed. The next three inner products are related to the adjoint solution. For each time interval  $t \in [t_k, t_{k+1})$ , the adjoint equation

$$-\frac{d\mathbf{q}^\dagger}{dt} = \frac{\partial \mathcal{R}^\dagger}{\partial \mathbf{q}} \mathbf{q}^\dagger + \frac{\partial \mathcal{I}}{\partial \mathbf{q}}, \quad (41a)$$

is solved in reverse time with conditions

$$\mathbf{q}^\dagger(t_f) = \frac{\partial \Phi}{\partial \mathbf{q}(t_f)} \quad \text{for } k = N - 1 \quad (41b)$$

$$\mathbf{q}^\dagger(t_{k+1}^-) = \frac{\partial \mathcal{P}}{\partial \mathbf{q}(t_{k+1}^-)} \quad \text{for } k = 0, 1, \dots, N - 2, \quad (41c)$$

so these corresponding terms in (39) are also zero. Each interval starts from intermediate condition  $\mathbf{q}^\dagger(t_{k+1}^-)$ , then progresses in reverse time to  $\mathbf{q}^\dagger(t_k)$ . Then the adjoint for the previous interval  $t \in [t_{k-1}, t_k)$  is solved starting with its own distinct intermediate condition  $\mathbf{q}^\dagger(t_k^-)$ . This procedure is illustrated schematically in Figure 9 (b). The last two inner products in (39) provide the gradient for updating the control and the intermediate conditions,

$$\frac{\partial \mathcal{L}_A}{\partial \mathbf{q}_k^+} = \mathbf{q}^\dagger(t_k) + \frac{\partial \mathcal{P}}{\partial \mathbf{q}_k^+} \quad (42a)$$

$$\frac{\partial \mathcal{L}_A}{\partial \Theta} = \frac{\partial \mathcal{R}^\dagger}{\partial \Theta} \mathbf{q}^\dagger + \frac{\partial \mathcal{I}}{\partial \Theta}, \quad (42b)$$

which are used for solving the subproblem (37).

For the quadratic penalty method,

$$\mathcal{P}[\{\Delta \mathbf{q}_k\}, \mu] = \frac{\mu}{2} \sum_{k=1}^{N-1} \|\mathbf{q}_k^+ - \mathbf{q}(t_k^-)\|_{\mathbb{Q}^+}^2, \quad (43)$$

with gradients for (41) and (42),

$$\frac{\partial \mathcal{P}}{\partial \mathbf{q}_k^+} = \mu [\mathbf{q}_k^+ - \mathbf{q}(t_k^-)] \quad (44a)$$

$$\frac{\partial \mathcal{P}}{\partial \mathbf{q}(t_k^-)} = -\mu [\mathbf{q}_k^+ - \mathbf{q}(t_k^-)]. \quad (44b)$$

#### 4.4. Demonstration on the Lorenz system

The optimization period  $t_f - t_0 = 20$  is split evenly into  $n = 200$  intervals based on  $e$ -folding time of this system  $t_\lambda \approx 1.11$  estimated in Section 3. The optimal length of time intervals has not been identified. In this and other examples, a conservative approach is used to keep gradient growth below a factor of 1.3. For the given interval length  $T = 0.1$ , the sensitivity is anticipated to be amplified by only about a factor of 1.09 within intervals. The minimization of each subproblem is deemed sufficient when

$$\left\| \frac{\partial \mathcal{L}_A}{\partial \Theta} \right\|_{\mathbb{T}}^2 + \sum_{k=1}^{N-1} \left\| \frac{\partial \mathcal{L}_A}{\partial \mathbf{q}_k^+} \right\|_{\mathbb{Q}^+}^2 < 10^{-8}. \quad (45)$$

305 The penalty strength  $\mu$  is increased by a factor of 10 for each subproblem, starting from  $\mu = 10^{-5}$ .

Figure 10 (a) compares the result to the standard gradient-based method. The multi-point method achieves a 99.99% reduction of  $\mathcal{J}$  versus the 44.6% for the standard gradient-based optimization. The intermediate discontinuities are  
 310 decreased to  $\|\Delta \mathbf{q}_k\|_{\mathbb{Q}^+} < 10^{-13}$ . Figure 10 (b) compares the step size  $\delta \Theta = \|\alpha_k \delta \Theta_k\|_{\mathbb{T}}$  taken in (14) between the standard and the multi-point gradient-based methods. Clearly, the multi-point method searches a much larger region, only decreasing its step size with increasing  $\mu$ . This finds the control shown in Figure 10 (c) and (d). It is no longer concentrated toward early times, and it  
 315 suppresses all the peaks of  $\mathcal{I}(t)$ . The optimized state trajectory was shown in Figure 2.

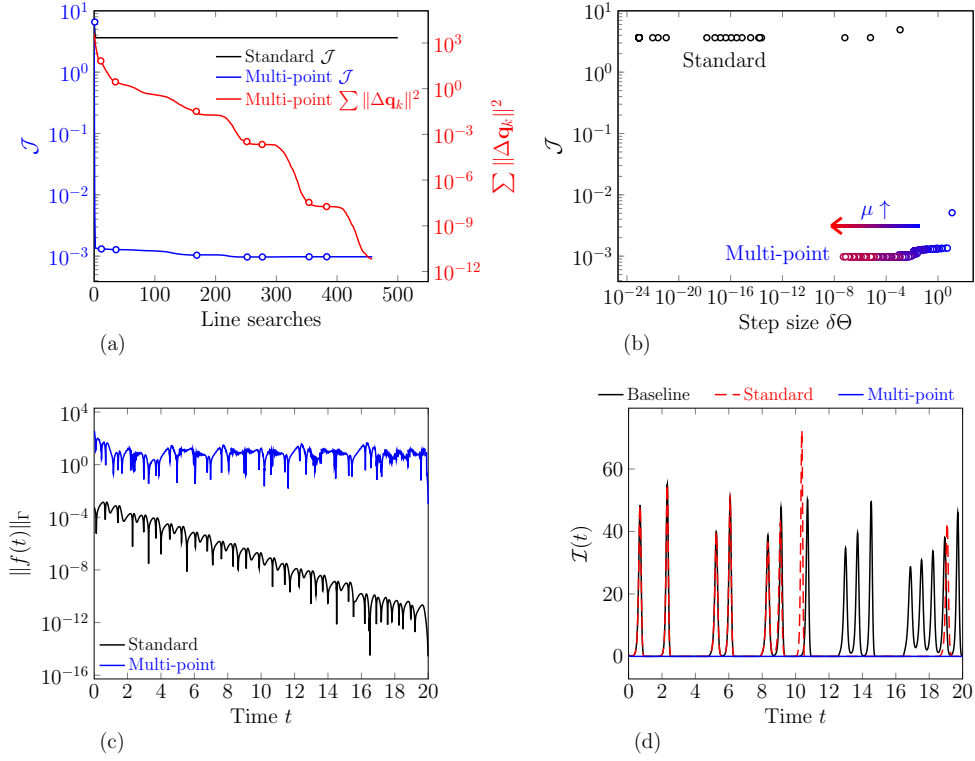


Figure 10: Multi-point penalty-based method applied to the Lorenz example of Section 3.1: (a) reduction of  $\mathcal{J}$  (15) and the intermediate discontinuities, with markers indicating the updates of  $\mu$ ; (b) the step sizes taken in the optimizations, with color for the multi-point methods changing from blue to red indicates the increase of  $\mu$ ; (c) the control strength  $f(t)$  of the optimized controls; and (d) the instantaneous objective functional  $\mathcal{I}(t)$  (16) of the baseline and controlled solutions.

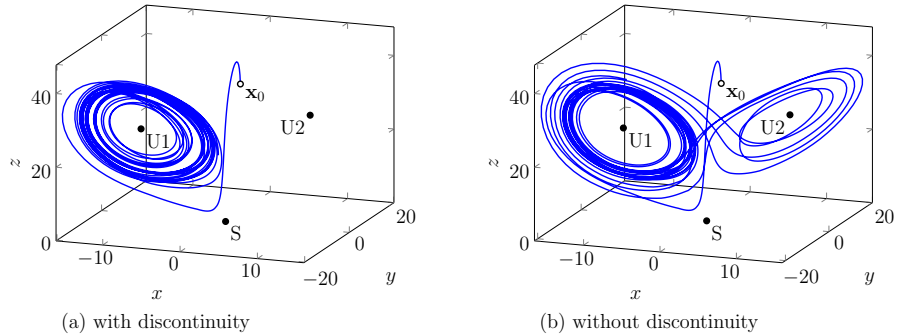


Figure 11: (a) The piecewise continuous optimized state trajectory of the Lorenz system (1), and (b) the state trajectory with the optimized control forcing and with the intermediate constraints strictly enforced.

The control is successful, but it must be recognized that it retains a pathological sensitivity to initial condition, so the utility of this optimized solution is questionable: Does it approximate a realistic trajectory? Can it be used for the actual control? These questions are discussed in the following two subsections.

#### 4.5. Approximation in a shadowing sense

As expected, due to the exponential sensitivity, even the small residual  $\|\Delta \mathbf{q}_k\|_{\mathbb{Q}^+} < 10^{-13}$  affects the optimized solution. Although the optimized state trajectory in Figure 11 (a) is essentially smooth, omitting the  $\Delta \mathbf{q}_k$  leads the trajectory to return to orbit  $U_2$ , as shown in Figure 11 (b). This same issue underlies any finite-precision numerical solution of a chaotic dynamical system [43, 86, 90, 91]. The sense in which such a solution approximates a true solution has been studied extensively [92–95]. In many cases, there is thought to exist a related *shadowing* trajectory with slightly different initial conditions that tracks the computed solution [94], lending credence to it. In the low-cost Lorenz system, it is possible, by advancing the steps sequentially, for us to calculate an exact numerical analog of a shadowing solution, also shown in Figure 12. (This is done by strictly enforcing  $\Delta \mathbf{q}_k = 0$  in stages through the time series, and re-optimizing the remainder with the multi-point method; full details of this exercise are reported elsewhere [81], where it is also done for the K–S

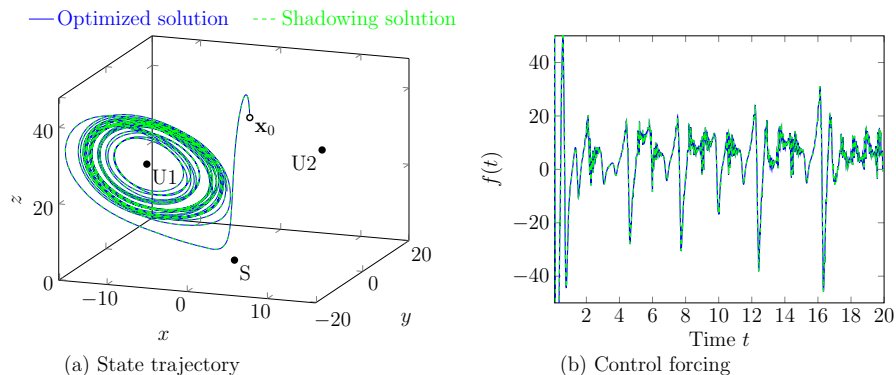


Figure 12: A shadowing solution of the Lorenz system: (a) its state trajectory, and (b) control  $f(t)$ .

equation.) For the Lorenz system, this further supports the existence of a true shadowing solution and supports the consistency of the numerically optimized solution. However, the cost of this full procedure approximately scales with the square of the number of time intervals times that of a machine precision converged single multi-step optimization, so it is beyond what can be managed for significantly more costly systems. Still, there does not seem to be any fundamental restriction to doing this for more complex cases. Chandramoorthy and Wang [96] raise a concern about shadow solutions that do not closely reproduce their original statistics, but such pathological shadowing solutions are not found in our study. This is probably because the shadowing distance is also controlled by the penalty, which was demonstrated for both the Lorenz and K-S equations [81]. Practically, subsequent analysis by the user will likely be decisive regarding whether or not the optimized solution is indeed useful.

#### 4.6. Utility of the optimized solution

Of course, even the existence of a true shadowing solution does not necessarily mean that the optimized trajectory is useful. If it requires such extremely fine resolution to have the expected performance, is the control usefully robust? Or, is the system actually controllable in practice?

The perspectives of closed-loop control exposes promising utility. As opti-

355 mized, the combined  $\mathbf{q}(t)$  and  $\Theta(t)$  solution is an open-loop control [97, 98]. Many open-loop controls are well-understood to be highly sensitive to disturbances and errors, and only effective when the underlying dynamical system is stable [97]. From this perspective, the failure in Figure 11 is only a consequence of passively applying the control in an open-loop fashion.

A path forward is to use  $\Theta(t) = \Theta[\mathbf{q}(t)]$  as one instance of a successful feedback control law to inform a control law  $\Theta[\mathbf{q}]$ . For the Lorenz example this can be demonstrated by simply fitting  $\mathbf{q}(t)$  and  $\Theta(t) = f(t)$  to find a robust closed-loop control. The optimized solution  $\mathbf{q}$  and  $\Theta$  in  $t \in [0, 20]$  is used to provide 2000 training samples of  $(\mathbf{q}(t), f(t))$ . A standard regression tree, as implemented in the MATLAB Toolbox [99], is used as a model function with minimum leaf size 4, leading to a tree of 685 nodes. The trained regression tree has

$$R^2 = 1 - \frac{\sum_{k=1}^{2000} [f_k - f(\mathbf{q}_k)]^2}{\sum_{k=1}^{2000} (f_k - \bar{f})^2} \approx 0.91. \quad (46)$$

360 Figure 13 shows the application of the inferred control law for four different initial conditions. Although only the initial condition shown in Figure 13 (a) is used for training, all of them are well controlled (Figure 13 c-d).

In this example, the optimization procedure is useful, because it provides valuable training data. Similarly, it is often unknown whether a flow is control-  
 365 lable. In such situations, the control found by the optimization also confirms the controllability of the flow system, in addition to providing a pathway to analyze its mechanism. Human-learning is also possible, as in the analysis of Wei and Freund [11] for the optimized control of two-dimensional shear layer sound.

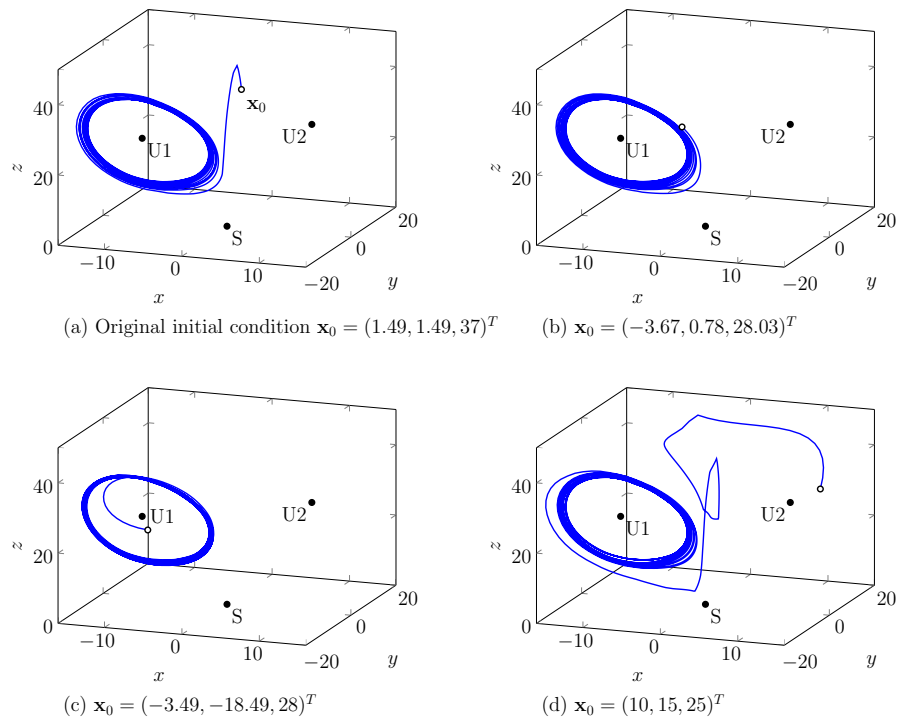


Figure 13: Solutions controlled by the learned control law starting from 4 different initial conditions: (a) in-sample and (b) out-of-sample cases.

370 **5. Demonstration: Advection plus Kuramoto–Sivashinsky (adv+KS)**

5.1. *Governing equation*

The generalized Kuramoto–Sivashinsky (K–S) equation [100],

$$\frac{\partial v}{\partial t} + \frac{1}{2} \frac{\partial v^2}{\partial x} + \alpha_1 \frac{\partial^2 v}{\partial x^2} + \alpha_2 \frac{\partial^4 v}{\partial x^4} = 0, \quad (47)$$

is considered on a periodic domain  $x \in \mathbb{D} = [0, 2\pi]$ , so the state  $v(x, t)$  is defined in  $\mathbb{Q} = H^0(\mathbb{D}) \times H^0(\mathbb{R}_0^+)$ . The inner product (5) is defined

$$\langle u(t), v(t) \rangle_{\mathbb{Q}^+} = \int_0^{2\pi} u(x, t)v(x, t) dx, \quad (48)$$

where  $\mathbb{Q}^+ = H^0(\mathbb{D})$ . Parameters  $\alpha_1 = 1.0$  and  $\alpha_2 = 0.02991$  are chosen so that the system (47) exhibits chaos [87].

To illustrate how a relatively chaotic portion of a turbulent flow might impede optimization of a relatively deterministic (presumably larger-scale) component, this K–S equation is augmented with deterministic advection. The combined solution  $\mathbf{q} = u(x, t) \in \mathbb{Q}$  is taken to be

$$u(x, t) = U(x, t) + \epsilon v(x, t), \quad (49)$$

where  $\epsilon = 0.01$ , and the deterministic  $U(x, t) = U(x - U_0 t, 0)$  part simply advects at speed  $U_0 = 2$ :

$$\frac{\partial U}{\partial t} + U_0 \frac{\partial U}{\partial x} = 0. \quad (50)$$

The initial conditions are

$$U(x, 0) = \sin 2x \quad \text{and} \quad v(x, 0) = \sin 5x. \quad (51)$$

375 Figure 14 (a) shows how  $U$  and  $v$  constitute the solution  $u$ . The independence of the dynamics is taken to be unknown, as it is for turbulence, though in this case the advection component provides an exact reduced model for the deterministic dynamics.

For numerical solution,  $\mathbb{D}$  is discretized with  $N_g = 512$  mesh points, with second-order central finite-difference first derivatives. Higher-order derivatives

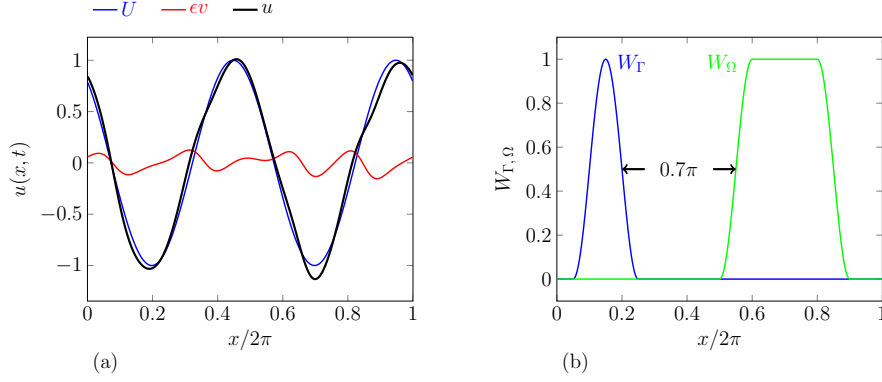


Figure 14: The adv+KS system: (a) evolution of the solution  $u = U + \epsilon v$ , and (b) mollifying support for the target region  $\Omega$  (53) and the control region  $\Gamma$  (56).

380 are computed with repeated first derivatives. It is integrated in time with a  
semi-explicit four-step Runge–Kutta scheme with  $\Delta t = 0.01$  [81], wherein only  
the nonlinear term from (47) is explicit. The discrete-exact adjoint is used  
to compute  $\nabla_{\Theta} \mathcal{J}$ . Using the same methods as for the Lorenz system, we estimate  
 $\tau_{\lambda} = 0.412$  and  $\tau_{\phi} = 0.339$ . Time step independence is confirmed for a short  
385 time  $t_f - t_i = 0.5$  similar to  $\tau_{\lambda}$ . Halving the time step changed  $u(x, t_f)$  by only  
0.7 percent.

## 5.2. Optimization

The control objective is to locally suppress the solution per

$$\mathcal{J} = \int_{t_0}^{t_f} \int_0^L |u(x, t)|^2 W_{\Omega}(x, t) dx dt, \quad (52)$$

where

$$W_{\Omega}(x) = \begin{cases} 0.5 + 0.5 \sin\left(\frac{x-1.1\pi}{0.2}\right) & \frac{x}{2\pi} \in [0.5, 0.6) \\ 1 & \frac{x}{2\pi} \in [0.6, 0.8) \\ 0.5 - 0.5 \sin\left(\frac{x-1.7\pi}{0.2}\right) & \frac{x}{2\pi} \in [0.8, 0.9) \\ 0 & \text{otherwise,} \end{cases} \quad (53)$$

is  $\mathcal{C}^3$ -continuous and exactly compact (Figure 14 b). For  $\epsilon = 0.01$ ,  $U$  contributes most of  $\mathcal{J}$ .

The entire  $u$  is forced by  $\Theta = f(x, t) \in \mathbb{T} = \mathbb{Q}$  in the  $\Gamma$  control region as

$$\frac{\partial u}{\partial t} + \mathcal{R}[u] = (1 + \epsilon)W_\Gamma(x)f(x, t), \quad (54)$$

which is distributed to  $U$  and  $v$  accordingly,

$$\frac{\partial U}{\partial t} + U_0 \frac{\partial U}{\partial x} = W_\Gamma(x)f(x, t) \quad (55a)$$

$$\frac{\partial v}{\partial t} + \frac{1}{2} \frac{\partial}{\partial x} (v^2) + \alpha_1 \frac{\partial^2 v}{\partial x^2} + \alpha_2 \frac{\partial^4 v}{\partial x^4} = W_\Gamma(x)f(x, t). \quad (55b)$$

The control mollifying support  $W_\Gamma(x)$  is shown in Figure 14 (b),

$$W_\Gamma(x) = \begin{cases} 0.5 + 0.5 \sin\left(\frac{x-0.2\pi}{0.2}\right) & \frac{x}{2\pi} \in [0.05, 0.25] \\ 0 & \text{otherwise.} \end{cases} \quad (56)$$

An estimated time for control effects to reach the target region is

$$\tau_A = \frac{\min \|\mathbf{x}_\Omega - \mathbf{x}_\Gamma\|}{U_c}, \quad (57)$$

390 where  $\min \|\mathbf{x}_\Omega - \mathbf{x}_\Gamma\|$  is the minimum distance between the target region  $\Omega$  and the control region  $\Gamma$ , and  $U_c$  is the characteristic advection speed. For  $U$ ,  $\tau_A = 1.1$ , and for  $v$  the advection speed has zero mean, so notionally  $\tau_A \rightarrow \infty$ , meaning there is no direct path by which control effort in  $v$  can propagate to the target region. Hence, the control effort is expected to be primarily impeded  
395 by the chaotic dynamics of  $v$ .

For reference, we first take advantage of the explicit separability of  $u = U + \epsilon v$ , as if there were a fully reliable reduced model. In this case, the optimization can target

$$\mathcal{J}_{\text{reduced}} = \int_{t_0}^{t_f} \int_0^L |U(x, t)|^2 W_\Omega(x, t) dx dt. \quad (58)$$

where the nominal reduced model (55a),

$$\frac{\partial U}{\partial t} + U_0 \frac{\partial U}{\partial x} = W_\Gamma(x)f(x, t), \quad (59)$$

governs  $U$  dynamics. With  $\mathcal{J}_{\text{reduced}}$  and (59), standard gradient-based optimization quickly identifies an effective control, which reduces  $\mathcal{J}_{\text{reduced}}$  from 0.4699

to 0.1266. When applied to the full dynamics, this same control based on the reduced models is only slightly less effective, yielding  $\mathcal{J} = 0.1314$ .

400 However, when exposed to the full dynamics (55), the corresponding optimization with a standard gradient method fails to significantly reduce  $\mathcal{J}$  (Figure 15 a). Figure 15 (b) suggests that this optimization stalls in narrow features of the  $\mathcal{J}$  optimization space.

For the multi-point method, the full simulation time  $t_f - t_0 = 5$  is evenly split  
 405 into  $n = 125$  intervals. Based on the  $e$ -folding time of this system  $t_\lambda = 0.412$ , the sensitivity should amplify a factor of 1.1 during each interval. The initial penalty strength  ${}_1\mu = 10^{-4}$  increases by a factor of 10 each step up to  ${}_3\mu = 10^{-2}$ , and by a factor of 4 subsequently. In Figure 15 (a), the early iterations with small  $\mu$  reduce  $\mathcal{J}$  beyond the ideal model, nearly to zero. This comes from the  
 410  $t - t_0 < \tau_A \approx 1.1$  period, which is physically uncontrollable due to the finite advection speed. Weak penalization (small  $\mu$ ) allows the jumps between time segments to affect this period. This is then removed by stronger penalization. Eventually, with significantly decreasing  $\sum \|\Delta \mathbf{q}_k\|^2$ , the multi-point method converges to a solution that is as effective as the ideal reduced model. The step  
 415 sizes in Figure 15 (b) shows that it too seems to avoid poor local minima. In addition, Figure 15 (c) shows that the optimized control nearly matches the ideal reduced model, though without explicit representation of it.

## 6. Application to turbulence

Our ultimate test is three-dimensional turbulent Kolmogorov flow, for which  
 420 the larger quasi-two-dimensional structures can be anticipated to be relatively deterministic amidst the overall chaos of the turbulence. (The multi-point method is also demonstrated elsewhere [81] for two-dimensional Kolmogorov flow [101, 102].)

### 6.1. Compressible flow with Kolmogorov forcing

The flow state, control, and the governing equations for compressible flow dynamics follow the formulation of Vishnampet [103]. For a triply periodic

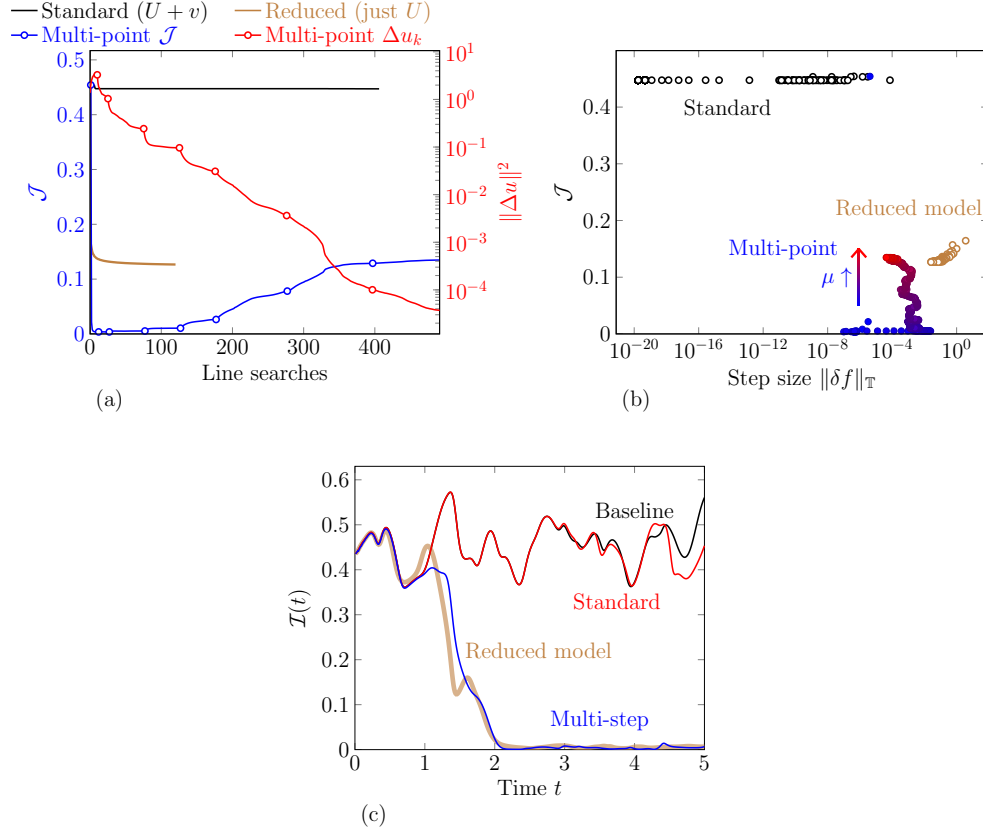


Figure 15: The optimization result for the adv+KS model. (a) Reduction of  $\mathcal{J}$  (52) and the discontinuity  $\|\Delta u\|^2$ . Standard gradient-based results for both the full dynamics and the reduced model are included for reference. Circle markers indicate the penalty strength updates. (b) The evolving  $\mathcal{J}$  and corresponding line search step sizes of  $f(x, t)$ . (c) The instant objective  $\mathcal{I}(t) = \int_0^L |u(x, t)|^2 W_{\Omega}(x, t) dx$  of the baseline and controlled solutions.

domain  $(x_1, x_2, x_3) \in \mathbb{D} = [0, L_x] \times [0, L_t] \times [0, L_t]$ , we denote  $V = H^0(\mathbb{D}) \times H^0(\mathbb{R}_0^+)$  as the space of  $L^2$ -functions from  $\mathbb{D} \times [0, \infty)$  to  $\mathbb{R}$ . The flow state  $\mathbf{q} \in \mathbb{Q} = V^5$  is

$$\mathbf{q} = \begin{pmatrix} \rho & \rho u_1 & \rho u_2 & \rho u_3 & \rho E \end{pmatrix}^T, \quad (60)$$

with  $\rho$  the density,  $\mathbf{u} = (u_1, u_2, u_3) \in V^3$  the velocity, and  $\rho E = \rho C_v T + \frac{1}{2} \rho u_i u_i$  the volume-specific total energy. The time  $t$  inner product (5) for flow states  $\mathbf{p}(t), \mathbf{q}(t) \in \mathbb{Q}^+ \equiv H^0(\mathbb{D})^5$  is

$$\langle \mathbf{p}(t), \mathbf{q}(t) \rangle_{\mathbb{Q}^+} = \int_{\mathbb{D}} \mathbf{p}^T(\mathbf{x}, t) \mathbf{q}(\mathbf{x}, t) d^3 \mathbf{x}. \quad (61)$$

A sinusoidal  $x_1$ -direction excitation force is added to the usual compressible flow equation to maintain turbulence,

$$\frac{\partial \mathbf{q}}{\partial t} + \frac{\partial}{\partial x_i} (\mathbf{F}_i^I - \mathbf{F}_i^V) = \begin{pmatrix} 0 & \chi \sin\left(2\pi n \frac{x_2}{L_t}\right) & 0 & 0 & 0 \end{pmatrix}^T, \quad (62)$$

425 where  $\mathbf{F}_i^I \in \mathbb{Q}$  and  $\mathbf{F}_i^V \in \mathbb{Q}$  are the usual  $x_i$ -direction advective and diffusive fluxes [103]. The gas is taken to be ideal, and the fluid is Newtonian with viscosity  $\eta$  and bulk viscosity  $\eta_B = 0.6\eta$  [104]. For simplicity, we assume a thermal energy sink that exactly cancels the work by the body force.

We include a generic thermal actuator  $\Theta = f(\mathbf{x}, t) \in \mathbb{T} = \mathbb{Q}$  in the governing equation (62),

$$\frac{\partial \mathbf{q}}{\partial t} + \frac{\partial}{\partial x_i} (\mathbf{F}_i^I - \mathbf{F}_i^V) = \begin{pmatrix} 0 & \chi \sin\left(2\pi n \frac{x_2}{L_t}\right) & 0 & 0 & W_\Gamma(\mathbf{x})f(\mathbf{x}, t) \end{pmatrix}^T, \quad (63)$$

where  $W_\Gamma(\mathbf{x}) \in H^0(\mathbb{D})$  is a mollifying compact support that defines the actuator and the control region  $\Gamma \subseteq \mathbb{D}$ . In the form of (4), the right-hand side is

$$\mathcal{R}[\mathbf{q}, \Theta] = -\frac{\partial}{\partial x_i} (\mathbf{F}_i^I - \mathbf{F}_i^V) + \begin{pmatrix} 0 & \chi \sin\left(2\pi n \frac{x_2}{L_t}\right) & 0 & 0 & W_\Gamma(\mathbf{x})f(\mathbf{x}, t) \end{pmatrix}^T. \quad (64)$$

## 6.2. Numerical Discretization

430 Spatial derivatives in (63) are discretized with sixth-order centered finite differences [103]. Second and mixed derivatives are discretized using repeated first

derivatives. This non-dissipative scheme necessitates the use of mild low-pass-filter-like artificial dissipation [103, 105]. Integration in time is by a standard fourth-order Runge–Kutta. The discrete-exact, dual-consistent adjoint solver developed by Vishnampet [103] is used, which, in the framework of (13), provides the gradient (13a) by solving the adjoint equation (13d-e) corresponding to (64).

### 6.3. Configuration

The domain and the external forcing parameters in (63) are chosen to exhibit strong chaos in a two-dimensional counterpart [101, 102, 106]. The Reynolds number based on the forcing strength  $\chi$  is

$$\text{Re}_\chi \equiv \frac{\sqrt{\chi}}{\eta_\infty} \left( \frac{L_t}{2\pi} \right)^{\frac{3}{2}} \approx 284, \quad (65)$$

with dynamic viscosity  $\eta_\infty$ . This is higher than the reported threshold ( $\text{Re}_\chi = 200$  at maximum) for which a chaotic flow state becomes the global attractor [101, 102, 106]. The aspect ratio  $\alpha \equiv \frac{L_t}{L_x} \approx 1.11$  is close to the  $\alpha = 1.1$  value at which intensely chaotic oscillations of vortices are observed [102]. Lastly, the widely-used  $n = 4$  body-force wavenumber used in (63) [101, 102, 106]. Time and velocity scales are,

$$\tau_c = \sqrt{\frac{\rho_\infty L_t}{\chi}} \quad u_c = \sqrt{\frac{L_t \chi}{\rho_\infty}}. \quad (66)$$

The reference state, with zero velocity and no excitation force, is chosen to have speed of sound  $a_\infty = \sqrt{20}u_c > u_c$ , so compressibility effect by the external forcing is weak. The initial condition has  $u_1(\mathbf{x}) = \sqrt{5}u_c$ , with added perturbations. A statistically stationary state is achieved before any control is sought. Turbulence statistics are collected at 200 times during  $t/\tau_c \in [13.42, 17.89]$ . The domain is discretized with  $256^3$  uniform mesh points and numerical time step  $\Delta t = 2.236 \times 10^{-4}\tau_c$ . **Turbulence statistics** are confirmed to be independent of resolution. Energy spectra in Figure 16 show that the turbulence is broadband over wavenumber  $k = \sqrt{k_1^2 + k_2^2 + k_3^2}$ . Following the same approach as for the Lorenz equation yields  $\tau_\lambda \approx 0.49\tau_c$  and  $\tau_\phi \approx 0.30\tau_c$ .

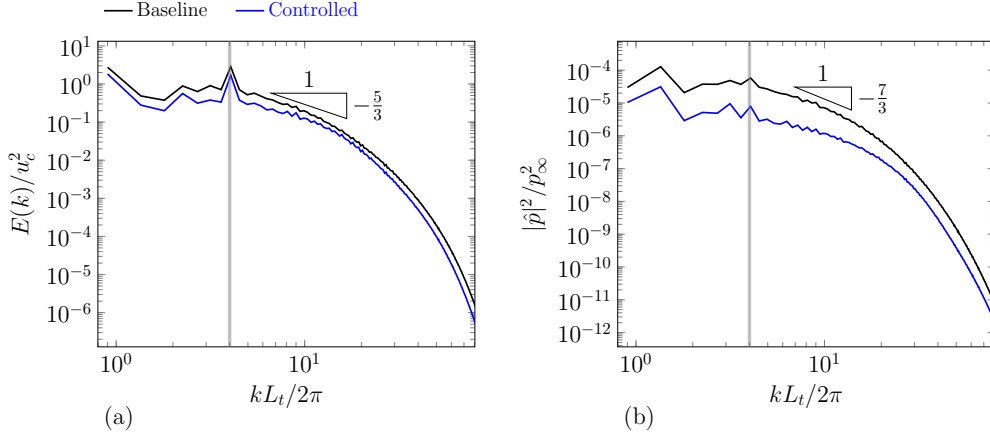


Figure 16: Kolmogorov flow spectra: (a) turbulence kinetic energy, and (b) pressure fluctuations. Vertical lines indicate the excitation wavenumber.

#### 6.4. Control formulation

The objective functional is

$$\mathcal{J} = \int_{t_i}^{t_f} \mathcal{I}(t) dt, \quad (67a)$$

where  $t_i = 13.42\tau_c$ ,  $t_f = 17.89\tau_c$ , and

$$\mathcal{I}(t) = \int_{\Omega} (p - p_\infty)^2 W_\Omega(\mathbf{x}) d^2\mathbf{x}, \quad (67b)$$

with constant reference pressure  $p_\infty = \frac{\rho_\infty a_\infty^2}{\gamma} = 14.286L_t\chi$ , and

$$W_\Omega = 0.5 \{ \tanh[4(x_1^\Omega - 0.075)] - \tanh[4(x_1^\Omega - 0.925)] \}, \quad (68a)$$

where

$$x_1^\Omega = \begin{cases} \frac{1}{0.4} \left( \frac{x_1}{L_x} - 0.6 \right) & \left| \frac{x_1}{L_x} - 0.8 \right| \leq 0.2 \\ 0 & \text{otherwise.} \end{cases}$$

For the control region,

$$W_\Gamma \propto B_{0,3}(4x_1^\Gamma - 2), \quad (68b)$$

where  $B_{0,3}(x)$  is the cubic B-spline basis function, and

$$x_1^\Gamma = \frac{1}{0.1L_x}(x_1 - 0.25L_x).$$

450 The support is normalized so that  $\max W_\Gamma = 1$ . The visualization in Figure 17 (c) also includes  $W_\Omega$  and  $W_\Gamma$ .

As for the previous example, the configuration is set up using the control propagation time  $\tau_A = 0.125\tau_c$  (57), based on the minimum distance  $\Delta x_{\Gamma-\Omega} \lesssim 0.279L_t$  between half maxima of  $W_\Omega$  and  $W_\Gamma$  and advection speed  $\sqrt{5}u_c$ . This  
 455 is also shorter than  $\tau_\lambda \approx 0.49\tau_c$  **to be consistent with our supposition that some component of the turbulence will be controllable. The control has the potential to reduce  $\mathcal{J}$  before being rendered ineffective by the chaotic dynamics.** However, chaos will still present a challenge since the simulation time  $t_f - t_i = 4.47\tau_c$  is much longer than  $\tau_\lambda \approx 0.49\tau_c$ . Hence, it is anticipated that the optimization will  
 460 be impacted significantly by chaos, despite the potential for an effective control. The gradient of  $\mathcal{J}$  (67) **with respect to control** and its viable step match these time scales for  $t - t_f \lesssim -2.5\tau_c$ .

### 6.5. Optimization

For the multi-point method, the simulation time  $t_f - t_0 = 4.47\tau_c$  is split into  
 465  $N = 40$  intervals of  $T = 0.11\tau_c$ , which based on  $\tau_\phi$  should amplify sensitivity by a factor of about 1.25 within each interval. Penalty  $\mu$  increases as  $10^{-5}$ ,  $10^{-4}$ ,  $10^{-3}$ ,  $2 \times 10^{-3}$ ,  $4 \times 10^{-3}$ ,  $8 \times 10^{-3}$ ,  $\dots$

Figure 18 (a) shows that the standard approach achieves a factor of 2 reduction of  $\mathcal{J}$ . Though not insignificant, most of the reduction is in one single,  
 470 potentially fortuitous, step of the third line search, and a flexible controller such as considered might be expected to perform better. The multi-point framework achieves a factor of 8 reduction, seemingly by broader exploration of  $\mathbb{T}$  space. In Figure 18 (b), the first three steps of standard approach, which achieved most reduction, are large  $\delta\Theta \sim 10^{-3}$ . However, most steps of the standard approach  
 475 have  $\delta\Theta \sim 10^{-7}$ . The control pursued by the standard approach is biased to early times, as shown in Figure 18 (c). Figure 18 (d) shows that most steps

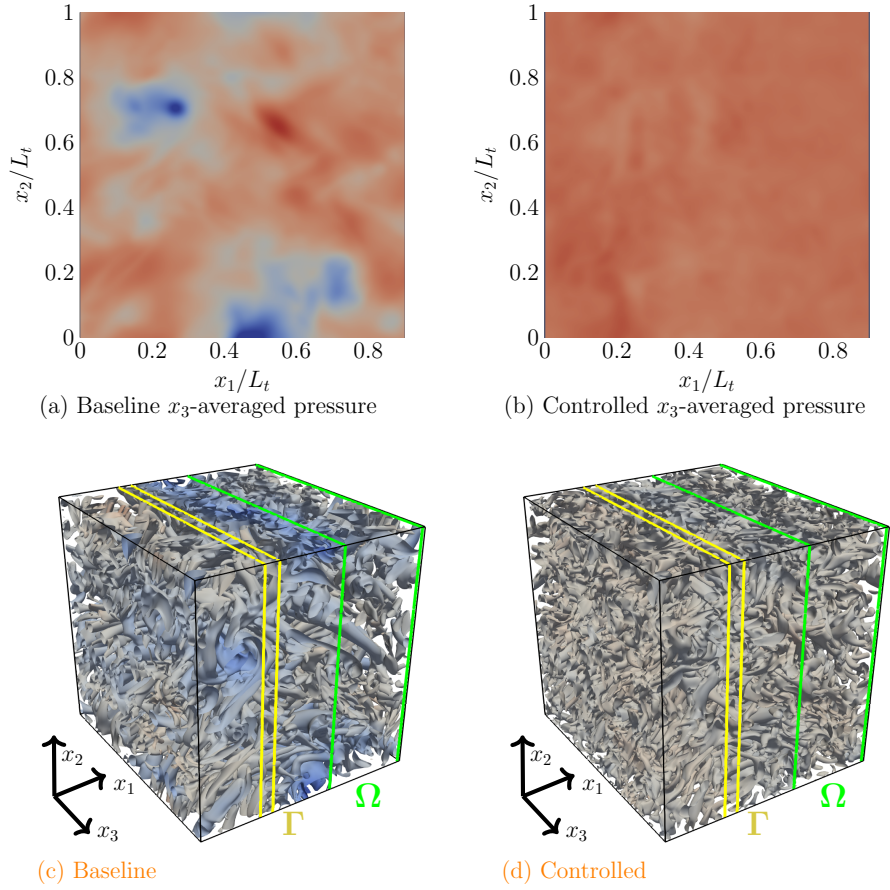


Figure 17: Kolmogorov turbulence at  $t_0 = 14.51\tau_c$ : pressure  $p/p_\infty \in [0.973, 1.008]$  averaged along  $x_3$  direction of (a) the baseline solution and (b) the controlled solution; and, isosurfaces of Q-criterion ( $Q = 20\tau_c^{-2}$ ) of (c) the baseline solution and (d) the controlled solution, colored by the pressure  $p/p_\infty \in [0.95, 1.05]$ . The supports for the control ( $\Gamma$ ) and the target ( $\Omega$ ) are visualized as yellow and green region, respectively.

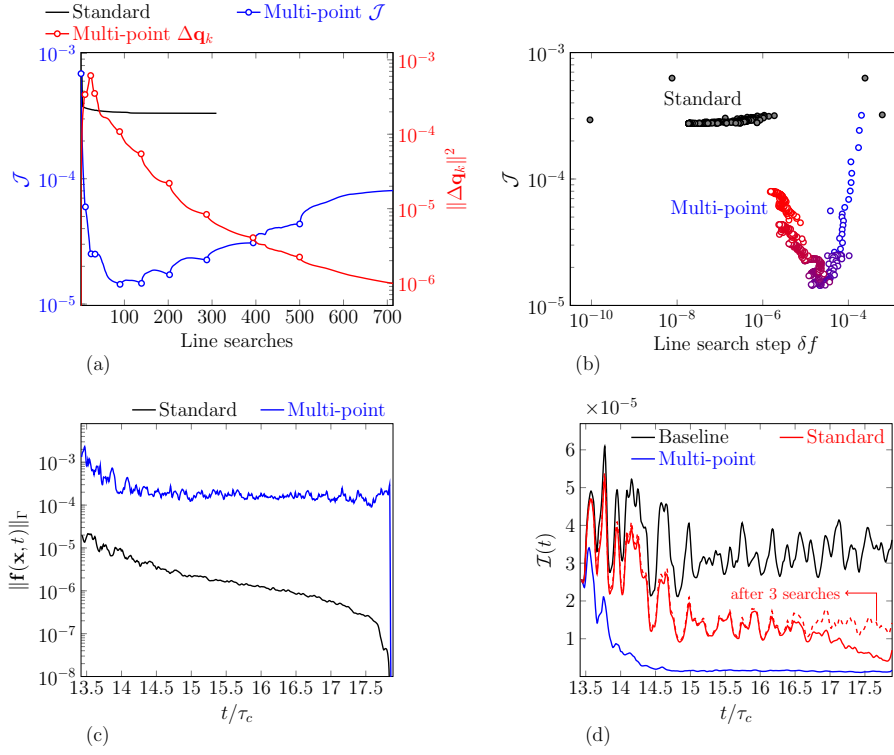


Figure 18: Three-dimensional Kolmogorov flow: (a) reduction of  $\mathcal{J}$  (67) and average  $\|\Delta \mathbf{q}_k\|^2$ , with markers denoting  $\mu$  updates; (b) the search steps in optimization procedure; (c) the control strength of the optimized controls; and (d) the instantaneous objective  $\mathcal{I}(t)$  (67b) of the baseline and controlled solutions.

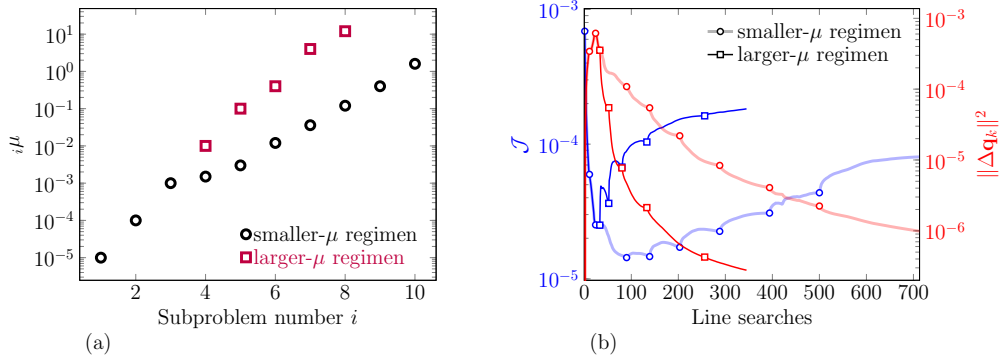


Figure 19: Penalty  $\mu$  schedules for three-dimensional Kolmogorov flow: (a)  $\mu$  used for each subproblem, including previous  $\mu$  schedule, and (b) the corresponding reduction of  $\mathcal{J}$  (67) and  $\|\Delta \mathbf{q}_k\|^2$ . Markers denote  $\mu$  updates.

in the standard approach only affect  $\mathcal{J}$  for late times. This is strikingly different from the multi-point method (Figure 18 b), for which the control found is distributed throughout the simulation time, effectively suppressing  $\mathcal{J}$  after a short initial transient period,  $t \lesssim 14.5\tau_c$ . This transient time  $t - t_i = 1.12\tau_c$  is longer than  $\tau_A = 0.125\tau_c$ , presumably because the control  $\Theta$  is set to directly manipulate only thermal energy. Regardless, it is effective much earlier than the standard approach.

Figure 17 (a) and (b) show that the targeted spanwise-averaged pressure fluctuations are suppressed, suggesting that the multi-point framework indeed targets the relatively deterministic large-scale structures. However, this is not achieved via laminarization; smaller-scale turbulence remains energetic and broadband (Figure 16), with only a slight decrease in small-scale turbulence kinetic energy. Hence the solution seems free of and of the pathological behavior that can hypothetically exist in shadowing trajectories [96].

### 6.6. Alternate $\mu$ penalty scheduling

Figure 19 shows another Kolmogorov flow optimization for the Kolmogorov flow with a faster increase in  $\mu$ . As expected, the  $\Delta \mathbf{q}_k$  discontinuity decreases

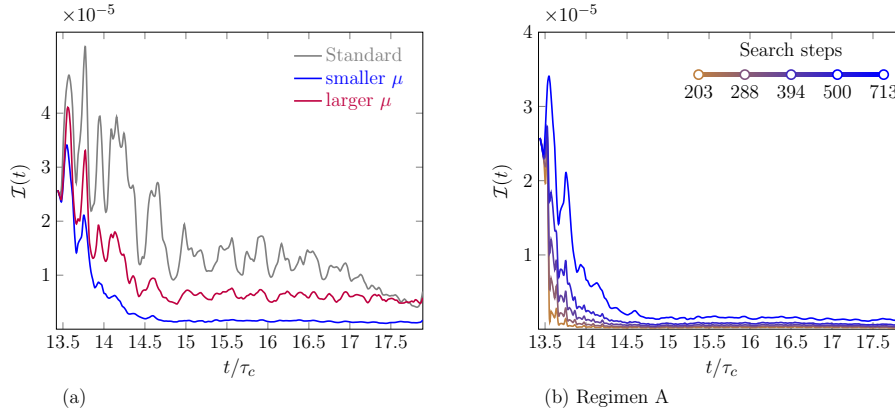


Figure 20: Optimization results from smaller- and larger- $\mu$  regimens in Figure 19: (a) instantaneous objective functional  $\mathcal{I}(t)$  (67b) at their final iterations, with standard gradient-based optimization included for a reference; and (b)  $\mathcal{I}(t)$  optimization of the smaller- $\mu$  regimen after the 203-th line search.

faster, but this compromises the reduction of  $\mathcal{J}$ . For each  $\mu$  update, abrupt  
495 jumps appear in  $\mathcal{J}$ . As the line searches continue,  $\mathcal{J}$  decreases again, but when  
 $\mu$  increases after only a few line searches, these upward jumps are not fully  
countered. Thus, in Figure 19, optimization with faster increase of  $\mu$  leads to  
a long-term inferior  $\mathcal{J}$ . At the same discontinuity  $\|\Delta\mathbf{q}_k\|^2 = 10^{-6}$ , the slower  
case in Figure 19 (b) achieves more  $\mathcal{J}$  reduction.

500 While this indicates a need for caution, it also provides flexibility. The opti-  
mized  $\mathcal{I}(t)$  for the two regimens in Figure 19 show the behavior. Figure 20 (a)  
shows that the larger- $\mu$  regimen reaches an inferior stationary state. However,  
this solution is obtained with about 4 times fewer line searches, as shown in  
Figure 19 (b), and at this point also has smaller  $\|\Delta\mathbf{q}\|$ . This reflects a trade-off  
505 between convergence rate and  $\mathcal{J}$  achieved. Figure 20 (b) shows the evolution of  
 $\mathcal{I}(t)$  of the smaller- $\mu$  regimen. During this part of the optimization, it changes  
mainly at small  $t - t_i$ , remaining small and lower than that of the larger- $\mu$   
regimen for later times. Both solutions, however, have similar transient time  
 $t - t_i \approx 1.12\tau_c$  reflecting the early increase of controllability from zero at  $t = t_i$ ,

510 shorter than that of standard gradient-based optimization.

## 7. Conclusions

By breaking the time domain into intervals with penalized jump discontinuities between them, the new algorithm allows a broad gradient search before being restricted by the non-convexity of short-time-scale chaos. Model problems  
515 demonstrate this in detail, and it is confirmed to be able to target relatively deterministic large scales amongst the chaos of Kolmogorov turbulence. It far outperforms standard gradient search methods, without the need for explicit formulation of a reduced model.

Of course, this approach does not ultimately skirt the chaos of the turbulence. Although arbitrarily close (with iterations) to one without discontinuities,  
520 the resulting flow solution remains subject to the extreme sensitivity of chaos, so its utility might be in question. Still, it is valuable as an instance of an effective control law. It can also establish that a flow is controllable and provide knowledge about how that might be accomplished, including data for training  
525 stable closed-loop controls.

Iterations through the stages of increasing penalty strength make the approach more expensive than a simple gradient, but its escape from nearby local minima seems to justify this added expense. It is also not fundamentally more expensive: its expense does not share the explosive increase in cost that would be  
530 associated with a broad random sampling of a high-dimensional control space. It is expected to be effective so long as there exists a sufficiently deterministic component that serves the objective, which is a common case in turbulent flow. It is also potentially applicable to chaotic systems beyond flow turbulence that include both relatively chaotic and relatively deterministic components.

## 535 Acknowledgement

This material is based in part upon work supported by the Department of Energy, National Nuclear Security Administration, under Award Number DE-

NA0002374 and DE-NA0003963.

## References

- 540 [1] T. R. Bewley, P. Moin, R. Temam, DNS-based predictive control of turbulence: an optimal benchmark for feedback algorithms, *Journal of Fluid Mechanics* 447 (2001) 179–225. doi:10.1017/S0022112001005821.
- [2] J. Kim, T. R. Bewley, A linear systems approach to flow control, *Annual Review of Fluid Mechanics* 39 (2007) 383–417. doi:10.1146/annurev.fluid.39.050905.110153.
- 545 [3] A. P. Singh, K. Duraisamy, Using field inversion to quantify functional errors in turbulence closures, *Physics of Fluids* 28 (4) (2016) 045110.
- [4] A. P. Singh, S. Medida, K. Duraisamy, Machine-learning-augmented predictive modeling of turbulent separated flows over airfoils, *AIAA Journal* 55 (7) (2017) 2215–2227. doi:10.2514/1.J055595.
- 550 [5] C. He, Y. Liu, L. Gan, A data assimilation model for turbulent flows using continuous adjoint formulation, *Physics of Fluids* 30 (10). doi:10.1063/1.5048727.
- [6] M. P. Juniper, Triggering in the horizontal Rijke tube: Non-normality, transient growth and bypass transition, *Journal of Fluid Mechanics* 667 (2011) 272–308. doi:10.1017/S0022112010004453.
- 555 [7] R. Kerswell, Nonlinear nonmodal stability theory, *Annual Review of Fluid Mechanics* 50 (1) (2018) 319–345. doi:10.1146/annurev-fluid-122316-045042.
- 560 [8] T. S. Eaves, C. P. Caulfield, Disruption of SSP=VWI states by a stable stratification, *Journal of Fluid Mechanics* 784 (2015) 548–564. doi:10.1017/jfm.2015.596.

- [9] A. Kord, J. Capecelatro, Optimal perturbations for controlling the growth of a Rayleigh–Taylor instability, *Journal of Fluid Mechanics* 876 (2019) 150–185. doi:10.1017/jfm.2019.532.
- [10] J. B. Freund, Adjoint-based optimization for understanding and suppressing jet noise, *Journal of Sound and Vibration* 330 (17) (2011) 4114–4122.
- [11] M. Wei, J. B. Freund, A noise-controlled free shear flow, *Journal of Fluid Mechanics* 546 (2006) 123–152.
- [12] J. Kim, D. J. Bodony, J. B. Freund, Adjoint-based control of loud events in a turbulent jet, *Journal of Fluid Mechanics* 741 (2014) 28–59.
- [13] A. Jameson, L. Martinelli, N. A. Pierce, Optimum aerodynamic design using the navier–stokes equations, *Theoretical and Computational Fluid Dynamics* 10 (1998) 213–237.
- [14] J. R. Martins, Perspectives on aerodynamic design optimization, *AIAA Scitech 2020 Forum* (January) (2020) 1–21. doi:10.2514/6.2020-0043.
- [15] N. Sikarwar, P. Morris, The use of an adjoint method for optimization of blowing in a convergent-divergent nozzle, *International Journal of Aeroacoustics* 14 (1-2) (2015) 327–351. doi:10.1260/1475-472X.14.1-2.327.
- [16] G. A. Mensah, J. P. Moeck, Acoustic Damper Placement and Tuning for Annular Combustors: An Adjoint-Based Optimization Study, *Journal of Engineering for Gas Turbines and Power* 139 (6). doi:10.1115/1.4035201.
- [17] K. J. Fidkowski, D. L. Darmofal, Review of output-based error estimation and mesh adaptation in computational fluid dynamics, *AIAA Journal* 49 (4) (2011) 673–694. doi:10.2514/1.J050073.
- [18] L. Shi, Z. J. Wang, Adjoint-based error estimation and mesh adaptation for the correction procedure via reconstruction method, *Journal of Computational Physics* 295 (2015) 261–284. doi:10.1016/j.jcp.2015.04.011.

- [19] R. A. Nasralla, A. M. Daoud, K. A. Fattah, M. H. Sayyoun, Fast and efficient sensitivity calculation using adjoint method for three-phase field-scale history matching, *Journal of Petroleum Science and Engineering* 77 (3-4) (2011) 338–350. doi:10.1016/j.petrol.2011.04.009.
- 595 [20] E. J. Parish, K. Duraisamy, A paradigm for data-driven predictive modeling using field inversion and machine learning, *Journal of Computational Physics* 305 (2016) 758–774. doi:10.1016/j.jcp.2015.11.012.
- [21] S. M. Rabin, C. P. Caulfield, R. R. Kerswell, Triggering turbulence efficiently in plane Couette flow, *Journal of Fluid Mechanics* 712 (2012) 244–272. doi:10.1017/jfm.2012.417.
- 600 [22] J. Nocedal, S. Wright, *Numerical optimization*, Springer Science & Business Media, 2006.
- [23] D. P. Bertsekas, *Constrained optimization and Lagrange multiplier methods*, Academic press, 2014.
- 605 [24] D. Leonard, N. V. Long, *Optimal control theory and static optimization in economics*, Cambridge University Press, 1992.
- [25] E. N. Lorenz, Deterministic nonperiodic flow, *Journal of the Atmospheric Sciences* 20 (2) (1963) 130–141.
- [26] Q. Wang, J. H. Gao, The drag-adjoint field of a circular cylinder wake at Reynolds numbers 20, 100 and 500, *Journal of Fluid Mechanics* 730 (2013) 145–161. doi:10.1017/jfm.2013.323.
- 610 [27] R. Vishnampet, D. J. Bodony, J. B. Freund, A practical discrete-adjoint method for high-fidelity compressible turbulence simulations, *Journal of Computational Physics* 285 (2015) 173–192.
- 615 [28] P. J. Blonigan, M. Farazmand, T. P. Sapsis, Are extreme dissipation events predictable in turbulent fluid flows?, *Physical Review Fluids* 4 (4) (2019) 1–21. arXiv:1807.10263, doi:10.1103/PhysRevFluids.4.044606.

- [29] J. Otero, R. D. Sandberg, A. Sharma, Direct numerical simulations for adjoint-based optimal flow and noise control of a backward-facing step, in: 22nd AIAA/CEAS Aeroacoustics Conference, 2016, p. 2889. doi: 10.2514/6.2016-2889.
- [30] P. M. Pardalos, A. Zilinskas, J. Zilinskas, Non-Convex Multi-Objective Optimization, Springer, 2017.
- [31] R. Horst, H. Tuy, Global optimization: Deterministic approaches, Springer, 1996. arXiv:arXiv:1011.1669v3.
- [32] R. Horst, Introduction to global optimization, Nonconvex optimization and its applications v. 3, Kluwer Academic Publishers, Dordrecht, 1995.
- [33] J. Mockus, Bayesian approach to global optimization: theory and practice, Mathematics and its applications. Soviet series, Kluwer Academic, Dordrecht, 1989.
- [34] R. Strongin, Y. Sergeyev, Global optimization with non-convex constraints: Sequential and parallel algorithms, Vol. 3, Springer Science & Business Media, 2013.
- [35] R. Schaefer, Foundation of Global Genetic Optimization, Springer, Berlin, 2007.
- [36] A. Zhigljavsky, A. Zilinskas, Stochastic Global Optimization, Springer, Dordrecht, 2008.
- [37] D. Ruelle, A review of linear response theory for general differentiable dynamical systems, Nonlinearity 22 (4) (2009) 855.
- [38] D. J. Lea, T. W. N. Haine, M. R. Allen, J. A. Hansen, Sensitivity analysis of the climate of a chaotic ocean circulation model, Quarterly Journal of the Royal Meteorological Society 128 (586) (2002) 2587-2605.
- [39] G. L. Eyink, T. W. N. Haine, D. J. Lea, Ruelle's linear response formula, ensemble adjoint schemes and Lévy flights, Nonlinearity 17.5.

- 645 [40] R. V. Abramov, A. J. Majda, Blended response algorithms for linear  
fluctuation-dissipation for complex nonlinear dynamical systems, *Nonlin-*  
*earity* 20 (12) (2007) 2793.
- [41] J. Thuburn, Climate sensitivities via a Fokker–Planck adjoint approach,  
*Quarterly Journal of the Royal Meteorological Society* 131 (605) (2005)  
650 73–92.
- [42] P. J. Blonigan, Q. Wang, Probability density adjoint for sensitivity anal-  
ysis of the mean of chaos, *Journal of Computational Physics* 270 (2014)  
660 660–686.
- [43] Q. Wang, Forward and adjoint sensitivity computation of chaotic dynam-  
ical systems, *Journal of Computational Physics* 235 (2013) 1–13.  
655
- [44] P. J. Blonigan, Adjoint sensitivity analysis of chaotic dynamical systems  
with non-intrusive least squares shadowing, *Journal of Computational*  
*Physics* 348 (2017) 803–826. doi:10.1016/j.jcp.2017.08.002.
- [45] P. J. Blonigan, Q. Wang, Multiple shooting shadowing for sensitivity anal-  
ysis of chaotic dynamical systems, *Journal of Computational Physics* 354  
660 (2018) 447–475.
- [46] J. Craske, Adjoint sensitivity analysis of chaotic systems using cumulant  
truncation, *Chaos, Solitons and Fractals: the interdisciplinary journal*  
*of Nonlinear Science, and Nonequilibrium and Complex Phenomena* 119  
665 (2019) 243–254. doi:10.1016/j.chaos.2018.12.024.
- [47] N. Chandramoorthy, Q. Wang, A computable realization of Ruelle’s  
formula for linear response of statistics in chaotic systems (feb 2020).  
arXiv:2002.04117.
- [48] S. W. Chung, S. D. Bond, E. C. Cyr, J. B. Freund, Regular sensitivity  
670 computation avoiding chaotic effects in particle-in-cell plasma methods,  
*Journal of Computational Physics* 400.

- [49] D. G. Crighton, M. Gaster, Stability of slowly diverging jet flow, *Journal of Fluid Mechanics* 77 (2) (1976) 397–413. doi:10.1017/S0022112076002176.
- 675 [50] P. Jordan, T. Colonius, Wave Packets and Turbulent Jet Noise, *Annual Review of Fluid Mechanics* 45 (1) (2013) 173–195. doi:10.1146/annurev-fluid-011212-140756.
- [51] K. Gudmundsson, T. Colonius, Instability wave models for the near-field fluctuations of turbulent jets, *Journal of Fluid Mechanics* 689 (2011) 97–  
680 128. doi:10.1017/jfm.2011.401.
- [52] A. V. Cavalieri, D. Rodríguez, P. Jordan, T. Colonius, Y. Gervais, Wavepackets in the velocity field of turbulent jets, *Journal of Fluid Mechanics* 730 (2013) 559–592. doi:10.1017/jfm.2013.346.
- [53] B. Callender, E. Gutmark, S. Martens, Far-field acoustic investigation  
685 into chevron nozzle mechanisms and trends, *AIAA Journal* 43 (1) (2005) 87–95. doi:10.2514/1.6150.
- [54] B. Callender, E. Gutmark, S. Martens, Near-field investigation of chevron nozzle mechanisms, *AIAA Journal* 46 (1) (2008) 36–45. doi:10.2514/1.17720.
- 690 [55] K. Gudmundsson, T. Colonius, Spatial stability analysis of chevron jet profiles, 13th AIAA/CEAS Aeroacoustics Conference (28th AIAA Aeroacoustics Conference) (2007) 1–14doi:10.2514/6.2007-3599.
- [56] K. Gudmundsson, Instability wave models of turbulent jets from round and serrated nozzles, Ph.D. thesis (2010).
- 695 [57] M. Koenig, Réduction de bruit de jet par injection fluidique en corps central tournant, Ph.D. thesis, Poitiers (2011).
- [58] C. W. Rowley, S. T. M. Dawson, Model reduction for flow analysis and control, *Annual Review of Fluid Mechanics* 49 (2017) 387–417. doi:10.1146/annurev-fluid-010816-060042.

- 700 [59] L. Magri, M. P. Juniper, Sensitivity analysis of a time-delayed thermoacoustic system via an adjoint-based approach, *Journal of Fluid Mechanics* 719 (2013) 183–202. [arXiv:1303.4267](#), [doi:10.1017/jfm.2012.639](#).
- [60] M. P. Juniper, R. I. Sujith, Sensitivity and nonlinearity of thermoacoustic oscillations, *Annual Review of Fluid Mechanics* 50 (2018) 661–689. [doi:10.1146/annurev-fluid-122316-045125](#).
- 705 [61] V. Nair, G. Thampi, S. Karuppusamy, S. Gopalan, R. I. Sujith, Loss of chaos in combustion noise as a precursor of impending combustion instability, *International Journal of Spray and Combustion Dynamics* 5 (4) (2013) 273–290. [doi:10.1260/1756-8277.5.4.273](#).
- [62] V. Nair, G. Thampi, R. I. Sujith, Intermittency route to thermoacoustic instability in turbulent combustors, *Journal of Fluid Mechanics* 756 (2014) 470–487. [doi:10.1017/jfm.2014.468](#).
- [63] J. Tony, E. A. Gopalakrishnan, E. Sreelekha, R. I. Sujith, Detecting deterministic nature of pressure measurements from a turbulent combustor, *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 92 (6) 715 (2015) 1–11. [doi:10.1103/PhysRevE.92.062902](#).
- [64] V. R. Unni, R. I. Sujith, Multifractal characteristics of combustor dynamics close to lean blowout, *Journal of Fluid Mechanics* 784 (2015) 30–50. [doi:10.1017/jfm.2015.567](#).
- 720 [65] E. Ott, *Chaos in Dynamical Systems*, 2002. [doi:10.1017/CB09781107415324.004](#).
- [66] P. Cvitanović, R. Artuso, R. Mainieri, G. Tanner, G. Vattay, N. Whelan, A. Wirzba, *Chaos: Classical and Quantum*, ChaosBook.org (Niels Bohr Institute, Copenhagen 2005), 2005.
- 725 [67] A. N. Kolmogorov, Entropy per unit time as a metric invariant of automorphisms, in: *Dokl. Akad. Nauk SSSR*, Vol. 124, 1959, pp. 754–755.

- [68] Y. G. Sinai, On the notion of entropy of dynamical systems, in: *Doklady Akademii Nauk*, Vol. 124, 1959, pp. 768–771.
- [69] P. Grassberger, Generalized dimensions of strange attractors, *Physics Letters A* 97 (6) (1983) 227–230. doi:10.1016/0375-9601(83)90753-3.
- [70] M. Katětov, On the Rényi dimension, *Commentationes Mathematicae Universitatis Carolinae* 27 (4) (1986) 741–753.
- [71] Y. Chen, Equivalent relation between normalized spatial entropy and fractal dimension, *Physica A: Statistical Mechanics and its Applications* 553 (2020) 124627. arXiv:1608.02054, doi:10.1016/j.physa.2020.124627.
- [72] Y. Kuramoto, T. Tsuzuki, Persistent propagation of concentration waves in dissipative media far from thermal equilibrium, *Progress of Theoretical Physics* 55 (2) (1976) 356–369.
- [73] G. I. Sivashinsky, On flame propagation under conditions of stoichiometry, *SIAM Journal on Applied Mathematics* 39 (1) (1980) 67–82.
- [74] L. S. Pontryagin, *Mathematical Theory of Optimal Processes*, in: L. S. Pontryagin Selected Works, 1st Edition, Vol. 4, Taylor & Francis, London, 2017. arXiv:arXiv:1011.1669v3.
- [75] C. Lanczos, *Linear differential operators*, *Classics in applied mathematics* 18, Society for Industrial and Applied Mathematics SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104, 1996.
- [76] J. Hicken, D. Zingg, The role of dual consistency in functional accuracy: error estimation and superconvergence, in: *20th AIAA Computational Fluid Dynamics Conference*, 2011, p. 3855.
- [77] A. R. Conn, N. I. M. Gould, P. L. Toint, *Trust-Region Methods*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2000.

- [78] W. T. Vetterling, S. A. Teukolsky, B. P. Flannery, W. H. Press, Numerical Recipes in C: The Art of Scientific Computing, Cambridge University Press, 2002.
- 755
- [79] A. Isidori, Nonlinear Control Systems, 3rd Edition, Vol. 53, Springer, 2013.
- [80] L. Chen, Y. Liu, Control of the Lorenz chaos by the exact linearization, Applied Mathematics and Mechanics 19 (1) (1998) 67.
- [81] S. W. Chung, Regular sensitivity calculation and gradient-based optimization of chaotic dynamical systems, Ph.D. thesis, University of Illinois at Urbana-Champaign (2021).
- 760
- [82] J. M. Ottino, The kinematics of mixing: stretching, chaos, and transport, Cambridge University Press, 1989.
- [83] B. Eckhardt, D. Yao, Local Lyapunov exponents in chaotic systems, Physica D 65 (1993) 100–108.
- 765
- [84] P. V. Kuptsov, U. Parlitz, Theory and computation of covariant Lyapunov vectors, Journal of Nonlinear Science 22 (5) (2012) 727–762. doi:10.1007/s00332-012-9126-5.
- [85] S. L. Brunton, C. W. Rowley, Fast computation of finite-time Lyapunov exponent fields for unsteady flows, Chaos 20 (1). doi:10.1063/1.3270044.
- 770
- [86] W. Hayes, Computer simulations, exact trajectories, and the gravitational  $N$ -body problem, American Journal of Physics 72 (9) (2004) 1251–1257.
- [87] F. Christiansen, P. Cvitanović, V. Putkaradze, Spatiotemporal chaos in terms of unstable recurrent patterns, Nonlinearity 10 (1) (1997) 55–70.
- 775
- [88] F. Christiansen, Multipoint shooting method, in: Chaos: Classical and Quantum, ChaosBook.org (Niels Bohr Institute, Copenhagen 2005), 2005, Ch. 16.2.

- 780 [89] Y. Lan, P. Cvitanović, Variational method for finding periodic orbits in a  
general flow, *Physical Review E* 69 (1).
- [90] J. Goodman, D. C. Heggie, P. Hut, On the exponential instability of  $N$ -  
body systems, *The Astrophysical Journal* 415 (1993) 715.
- [91] M. Hemsendorf, D. Merritt, Instability of the gravitational  $N$ -body prob-  
785 lem in the large- $N$  limit, *The Astrophysical Journal* 580 (1) (2002) 606.
- [92] D. V. Anosov, Geodesic flows on closed Riemann manifolds with negative  
curvature, *Proceedings of the Steklov Institute of Mathematics*, no. 90,  
1967, American Mathematical Society, Providence, 1969.
- [93] R. Bowen,  $\omega$ -Limit sets for Axiom A diffeomorphisms, *Journal of Differ-  
790 ential Equations* 18 (2) (1975) 333–339. doi:10.1016/0022-0396(75)  
90065-0.
- [94] C. Grebogi, S. M. Hammel, J. A. Yorke, T. Sauer, Shadowing of physical  
trajectories in chaotic dynamics: Containment and refinement, *Physical  
Review Letters* 65 (13) (1990) 1527–1530. doi:10.1103/PhysRevLett.  
795 65.1527.
- [95] W. B. Hayes, K. R. Jackson, Rigorous shadowing of numerical solutions  
of ordinary differential equations by containment, *SIAM Journal of Nu-  
merical Analysis* 41 (5) (2003) 1948–1973.
- [96] N. Chandramoorthy, Q. Wang, On the probability of finding nonphysi-  
800 cal solutions through shadowing, *Journal of Computational Physics* 440  
(2021) 110389.
- [97] W. L. Hallauer, *Introduction to Linear, Time-Invariant, Dynamic Systems  
for Students of Engineering*, Virginia Polytechnic Institute and State Uni-  
versity, 2021.
- 805 [98] M. F. Golnaraghi, *Automatic control systems*, tenth edition. Edition,  
McGraw-Hill Education, New York, 2017.

- [99] L. Breiman, J. Friedman, C. J. Stone, R. A. Olshen, Classification and regression trees, CRC press, 1984.
- [100] N. A. Kudryashov, Solitary and periodic solutions of the generalized Kuramoto–Sivashinsky equation, Regular and Chaotic Dynamics 13 (3) (2008) 234–238. 810
- [101] G. J. Chandler, R. R. Kerswell, Invariant recurrent solutions embedded in a turbulent two-dimensional Kolmogorov flow, Journal of Fluid Mechanics 722.
- [102] D. Lucas, R. Kerswell, Spatiotemporal dynamics in two-dimensional Kolmogorov flow over large domains, Journal of Fluid Mechanics 750 (2014) 518–554. 815
- [103] R. Vishnampet, An exact and consistent adjoint method for high-fidelity discretization of the compressible flow equations, Ph.D. thesis, University of Illinois at Urbana-Champaign (2015). 820
- [104] P. A. Thompson, Compressible-Fluid Dynamics, Advanced engineering series, McGraw-Hill, 1971.
- [105] K. Mattsson, M. Svård, J. Nordström, Stable and Accurate Artificial Dissipation, Journal of Scientific Computing 21 (1).
- [106] N. Platt, L. Sirovich, N. Fitzmaurice, An investigation of chaotic Kolmogorov flows, Physics of Fluids A 3 (4) (1991) 681–696. doi:10.1063/1.858074. 825