HAL
open science

Localization of flow structures using infinity-norm optimization

D. P. G. Foures, C. P. Caulfield, Peter J. Schmid

To cite this version:

D. P. G. Foures, C. P. Caulfield, Peter J. Schmid. Localization of flow structures using infinity-norm optimization. Journal of Fluid Mechanics, 2013, 729, pp.672-701. 10.1017/jfm.2013.333 . hal-00996436

HAL Id: hal-00996436

https://polytechnique.hal.science/hal-00996436v1

Submitted on 30 Aug 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

J. Fluid Mech. (2013), *vol.* 729, *pp.* 672–701. © Cambridge University Press 2013
doi:10.1017/jfm.2013.333

Localization of flow structures using ∞-norm optimization

D. P. G. Foures[1,†], C. P. Caulfield[2,1] and P. J. Schmid[3]

[1]DAMTP, University of Cambridge, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, UK

[2]BP Institute, University of Cambridge, Madingley Rise, Madingley Road, Cambridge CB3 0EZ, UK

[3]Laboratoire d'Hydrodynamique (LadHyX), CNRS-Ecole Polytechnique, 91128 Palaiseau, France

(Received 25 October 2012; revised 6 February 2013; accepted 23 June 2013; first published online 24 July 2013)

Stability theory based on a variational principle and finite-time direct-adjoint optimization commonly relies on the kinetic perturbation energy density $E_1(t) = (1/V_\Omega)\int_\Omega e(\boldsymbol{x}, t)\,\mathrm{d}\Omega$ (where $e(\boldsymbol{x}, t) = |\boldsymbol{u}|^2/2$) as a measure of disturbance size. This type of optimization typically yields optimal perturbations that are global in the fluid domain Ω of volume V_Ω. This paper explores the use of p-norms in determining optimal perturbations for 'energy' growth over prescribed time intervals of length T. For $p = 1$ the traditional energy-based stability analysis is recovered, while for large $p \gg 1$, localization of the optimal perturbations is observed which identifies confined regions, or 'hotspots', in the domain where significant energy growth can be expected. In addition, the p-norm optimization yields insight into the role and significance of various regions of the flow regarding the overall energy dynamics. As a canonical example, we choose to solve the ∞-norm optimal perturbation problem for the simple case of two-dimensional channel flow. For such a configuration, several solutions branches emerge, each of them identifying a different energy production zone in the flow: either the centre or the walls of the domain. We study several scenarios (involving centre or wall perturbations) leading to localized energy production for different optimization time intervals. Our investigation reveals that even for this simple two-dimensional channel flow, the mechanism for the production of a highly energetic and localized perturbation is not unique in time. We show that wall perturbations are optimal (with respect to the ∞-norm) for relatively short and long times, while the centre perturbations are preferred for very short and intermediate times. The developed p-norm framework is intended to facilitate worst-case analysis of shear flows and to identify localized regions supporting dominant energy growth.

Key words: instability, mathematical foundations, variational methods

1. Introduction

Stability analysis in fluid mechanics may be defined as the study of fluid behaviour in the vicinity of a predefined state (the *base state*). If an initial deviation from the base state decays over time, the flow is considered to be stable; if an initial deviation

† Email address for correspondence: dpgf2@cam.ac.uk

increases over time, the flow is considered unstable. The deviation in this analysis is traditionally measured by the energy of the perturbation field or a variation thereof.

It is becoming increasingly accepted that the non-normal properties of the Navier–Stokes operator need to be considered in such stability analyses. Non-modal stability analysis tools have been developed (Trefethen *et al.* 1993; Schmid & Henningson 2001; Schmid 2007) which accurately capture the finite-time evolution of infinitesimal perturbations and allow the study of transient energy amplification for subcritical Reynolds numbers (Gustavsson 1991; Butler & Farrell 1992; Reddy & Henningson 1993 and others). Initially, these developed tools relied on decomposition methods from linear algebra, most notably the singular value decomposition (SVD), and inherited from them the prevalence of L_2-norms, which translated into energy measures or their variants (such as, e.g., enstrophy) of the physical variables. In particular, the vast majority of early studies are based on a measure of the local kinetic energy density $e(\boldsymbol{x}, t)$ defined as

$$e(\boldsymbol{x}, t) = \frac{1}{2}|\boldsymbol{u}|^2 = \frac{1}{2}\sum_{i=1}^{N} {u_i}^2, \tag{1.1}$$

where $\boldsymbol{u} = (u_i)_{i=1\ldots N}$ represents the pointwise velocity vector at time t, and N is the dimension of the flow. The common choice is then the 1-norm of the energy density field (L_2-norm of the velocity vector, representing the integrated energy over the considered domain Ω) given as

$$E_1(t) = \|e(\boldsymbol{x}, t)\|_1 = \frac{1}{V_\Omega}\int_\Omega e(\boldsymbol{x}, t)\,\mathrm{d}\Omega, \tag{1.2}$$

where V_Ω is the volume of the domain.

As more complex flows were tackled, the linear-algebra-based direct analysis gave way to a variational approach, introducing an iterative scheme based on the direct and adjoint evolution equations for solving an optimization problem constrained by partial differential equations. With this more versatile approach, time-periodic, generally time-dependent (Farrell & Ioannou 1996; Schmid 2007; Arratia, Caulfield & Chomaz 2013), stochastic (Farrell & Ioannou 2003) and even nonlinear stability problems (see, e.g., Cherubini *et al.* 2010; Pringle & Kerswell 2010; Monokrousos *et al.* 2011; Rabin, Caulfield & Kerswell 2012) could be treated, and constraints based on seminorms could be accommodated (Foures, Caulfield & Schmid 2012). However, while the constraining governing equations became increasingly more complex, scant attention has been paid to the manner in which the cost functional has been expressed, and the persistence of the classical, energy-based measure of amplification (or decay) has resulted in a bias towards global optimal structures. A notable exception is Monokrousos *et al.* (2011) who, based on thermodynamics arguments, successfully used the total dissipation integrated over a time interval to identify the nonlinear optimal initial perturbation in plane Couette flow. Nevertheless, in this case, the objective functional, although not a perturbation kinetic energy, was still a quadratic measure of the perturbation velocity vector. Physical mechanisms associated with energy-based optimization, whether optimal, suboptimal or asymptotic, are often based on large-scale structures, harvesting energy from any available source in the flow, since no restrictions on their spatial extent have been imposed explicitly. Global structures are thus widespread in energy-based optimization.

There is considerable interest, in particular driven by industrial applications, to take a more local approach and ask which localized regions of the flow are responsible for

optimal energy growth. This type of question constitutes a first attempt of a worst-case analysis of the flow, as it would pinpoint and identify 'hotspots' that exhibit strong amplification of energy density. To accomplish this analysis, a cost functional must be chosen that promotes localization of the optimal perturbation which is identified. To address this question, we generalize the energy-based cost functional to a functional based on a p-norm. The p-norm of the energy density is

$$E_p(t) = \|e(\boldsymbol{x}, t)\|_p = \left(\frac{1}{V_\Omega}\int_\Omega e(\boldsymbol{x}, t)^p \,\mathrm{d}\Omega\right)^{1/p}, \tag{1.3}$$

which, in the limit of large p, approaches the ∞-norm and yields localized optimal (worst-case) perturbations. We introduce a general mathematical framework to consider such 'stability' problems, including computational implementation issues, and we validate it by considering a representative case.

As a testbed for this validation and a demonstration of novel features revealed by p-norm optimization, we chose two-dimensional channel flow at a Reynolds number of $Re = 4000$. Despite the apparent simplicity of this flow, and the large body of literature on its stability and transition behaviour (see, e.g., Orszag 1971; Farrell 1988; Reddy, Schmid & Henningson 1993), this choice has been made deliberately. Applying the p-norm framework to channel flow and thereby uncovering novel features for this much-studied flow should enable and motivate the application of this type of technique to flow configurations that are less well known or whose stability and transition scenarios are additionally complicated by geometric features or other complexities. For such inhomogeneous flows, for example flow over a backward-facing step (see Blackburn, Barkley & Sherwin 2008; Marquet *et al.* 2008) or flow over a cavity, it has been shown that optimal initial perturbations based on the L_2-norm of the velocity vector are rather localized in space and often coincide with regions of maximum shear, e.g. close to detachment zones. However, the localization of the initial perturbation does not necessarily lead to a localized optimal terminal perturbation. Therefore, even for configurations for which the common global optimization approach results in a spatial localization of the perturbation, the use of a p-norm could conceivably yield even more localized structures at the end of the specified time horizon. Moreover, even for such complex configurations, the identification of several zones of localized energy production makes the optimization of the ∞-norm an interesting and instructive choice. Starting with revisiting the classical Orr mechanism with a local energy density analysis, the p-norm optimization will then expose a system of solution branches that are suboptimal in a traditional analysis but come to the fore as one searches for localized flow features that grow optimally over a prescribed time span. In effect, this article is not an analysis of two-dimensional channel flow, but rather a study of localization effects on perturbations and their associated energy dynamics once we deviate from the traditional energy measure that favours global input and output structures.

The remainder of the paper is organized as follows. In § 2 we define the governing equations and energy measures, while in § 3 we present the variational framework, paying particular attention to the challenges of optimizing with respect to a p-norm. Equipped with these insights, we revisit in § 4 the classic $p = 1$ case for two-dimensional plane Poiseuille channel flow, but analyse this flow by focusing on the local energy mechanisms rather than the previously considered global measures. This analysis provides the reference case for optimizing the p-norm, allowing us to study the effects of localization in § 5 and to explore the ∞-norm limit. In

addition to producing localization effects, the p-norm optimization framework allows the distinction of different mechanisms at play in a given flow, manifested by co-existing solutions branches. In § 6, we solve the ∞-norm optimal perturbation problem for a varying horizon (optimization) time and show that the different mechanisms leading to energy hotspots are in competition for dominance at a given time: different mechanisms to produce localized energy perturbations are identified depending on the specific time horizon. We discuss the implications of p-norm optimization and draw conclusions in the final § 7. In the appendices, we offer more details on a robust, essentially geometric, implementation of the optimization algorithm used throughout this article (appendix A) and we present a series of computations at $Re = 250$ to show an example of the Re-dependence of the solutions (appendix B). Finally, we detail in § C the convergence properties of the various optimal solutions presented in this paper.

2. Flow configuration and governing equations

We wish to demonstrate the mathematical framework for the computation of general p-norm optimal perturbations for a simple two-dimensional plane channel flow driven by a constant pressure gradient in a rectangular domain of streamwise size L_x, which varies depending on the time integration window chosen, and wall-normal extent $L_y = 2$. We non-dimensionalize lengths with half the channel depth h^*, and velocities with the centreline velocity u^*. Despite the simplicity of this flow and a considerable body of literature on its perturbation dynamics, Poiseuille flow represents a fitting testbed for our study of localization effects by p-norm optimization and of the associated energy dynamics; moreover, it is prototypical of more complex shear flows. Within this non-dimensionalization, the base flow is given by the familiar parabolic profile

$$\overline{\boldsymbol{u}} = y(2-y)\boldsymbol{e}_{\boldsymbol{x}}. \tag{2.1}$$

The central governing equations for our study are the (non-dimensionalized) two-dimensional linearized Navier–Stokes equations in a Cartesian coordinate system with periodic boundary conditions in the streamwise direction and standard no-slip boundary conditions at the walls:

$$\partial_t \boldsymbol{u} + \overline{\boldsymbol{u}} \cdot \nabla \boldsymbol{u} + \boldsymbol{u} \cdot \nabla \overline{\boldsymbol{u}} + \nabla p - \frac{1}{Re}\nabla^2 \boldsymbol{u} = 0, \tag{2.2a}$$

$$\nabla \cdot \boldsymbol{u} = 0, \tag{2.2b}$$

where $\overline{\boldsymbol{u}}$ stands for the base flow and $(\boldsymbol{u}, p)$ describe the perturbation velocity and pressure (not to be confused with p, the parameter of the p-norm), and $Re = u^* h^*/\nu^*$, where ν^* is the (dimensional) kinematic viscosity. The initial (in general spatially varying) velocity field is denoted by $\boldsymbol{u}(\boldsymbol{x}, 0) = \boldsymbol{u}_0$.

In what follows we consider a temporal optimization problem whereby an initial condition $\boldsymbol{u}_0$ is sought which maximizes a given functional, such as the energy gain, over a specified time interval. Our analysis will use energy considerations and their generalizations to isolate and quantify physical mechanisms that contribute to the optimal disturbance growth in time, and it is thus appropriate to obtain an evolution equation for both the global and the local pointwise perturbation energy densities, as the latter equation will provide more insight in localization issues and energy-redistribution processes and will thus yield a more detailed picture of the underlying physical mechanisms. We derive a temporal evolution equation for the pointwise energy density field defined in (1.1) (for $N = 2$) by multiplying the linearized

momentum equations (2.2) by the fluctuating velocity field. We obtain

$$\partial_t e = \underbrace{-\frac{1}{2}\overline{\boldsymbol{u}}\cdot\nabla(\boldsymbol{u}\cdot\boldsymbol{u})}_{a}\underbrace{-\boldsymbol{u}\cdot(\boldsymbol{u}\cdot\nabla\overline{\boldsymbol{u}})}_{s}\underbrace{-\boldsymbol{u}\cdot\nabla p}_{i}\underbrace{-\frac{1}{Re}\nabla\boldsymbol{u}:\nabla\boldsymbol{u}}_{d}\underbrace{+\frac{1}{Re}\nabla\cdot(\boldsymbol{u}\cdot\nabla\boldsymbol{u})}_{r}, \quad (2.3)$$

where the terms a, s, i, d and r denote advection, shear production, incompressibility redistribution, dissipation and diffusive redistribution of kinetic energy, respectively. The analysis that follows will concentrate on the various terms of the right-hand side and deduce interactions and energy-exchange processes during the optimal growth of perturbations.

By integrating the above equation over the computational domain Ω, we find the standard Reynolds–Orr equation, as derived by Schmid & Henningson (2001) for example, describing the evolution of the global energy $E(t)$:

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \underbrace{-\frac{1}{V_\Omega}\int_\Omega \boldsymbol{u}\cdot(\boldsymbol{u}\cdot\nabla\overline{\boldsymbol{u}})\,\mathrm{d}\Omega}_{S}\underbrace{-\frac{1}{ReV_\Omega}\int_\Omega \nabla\boldsymbol{u}:\nabla\boldsymbol{u}\,\mathrm{d}\Omega}_{D}, \quad (2.4)$$

where ':' represents the term by term (or Frobenius) product. We note that only the shear production and dissipation term remain; the advection term and incompressibility constraint do not contribute to the global energy budget.

3. A variational framework for p-norm optimization

3.1. *Cost functional and constraints*

We use a variational technique to compute perturbations that optimize a given objective (cost functional) subject to prescribed constraints. These constraints are added to the cost functional via scalar products with Lagrange multipliers or adjoint variables to yield an unconstrained optimization problem. This technique has been employed previously for energy-based optimization (see Schmid (2007) for a review); here, we extend the framework to more general measures of disturbance size. In particular, the cost functional is generalized to include the p-norm of the state vector given by (1.3).

This definition includes the common energy disturbance measure (for $p = 1$), but also less typical measures such as the (worst-case) infinity-norm (for $p \to \infty$). Under the same governing equations and constraints, the choice of norm for the cost functional has a significant influence on the optimal solution. We illustrate this dependence qualitatively in figure 1, in a discrete two-dimensional case.

We solve the p-norm maximization problem below, but the choice of the norm is naturally just as crucial for maximization problems as for minimization problems (as sketched in the corresponding figure). Adopting the p-norm concept introduced in (1.3), the cost functional of interest may be expressed as

$$\mathscr{J}(\boldsymbol{u}) = E_p(T), \quad (3.1)$$

that is, the measure of the p-norm energy density integrated over the domain Ω and evaluated at the end of a specified time-interval $t \in [0, T]$. With this definition, the classical energy gain is recovered as the special case E_1, while (3.1) represents a family of objective functionals parameterized by p. For large values of p, the cost functional still has the dimension of an energy although we believe no physical interpretation should be attributed to it. We only use it for algorithmic convergence to approximate, in a continuous fashion, the spatial maximum of the energy density. For

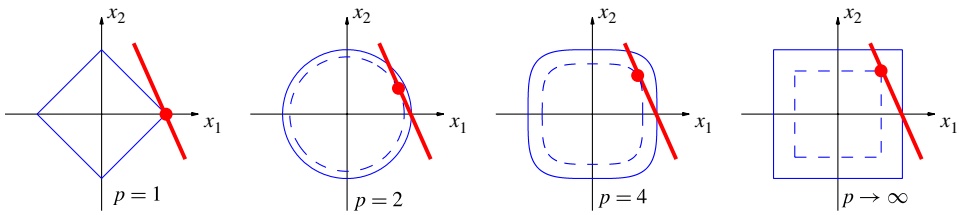

FIGURE 1. (Colour online) Locus of a two-dimensional unit-p-norm vector $\boldsymbol{x}$ for different values of p (thin solid line, shown in blue online). We recall that the p-norm in a two-dimensional discrete case can be expressed as $\|\boldsymbol{x}\|_p = ({x_1}^p + {x_2}^p)^{1/p}$. The solution (dot, shown in red online) of a constrained p-norm minimization problem for different values of p The constraint is given by the thick solid line (shown in red online). The minimum p-norm of the solution is indicated by the thin dashed line (shown in blue online) (which for the case $p = 1$ coincides with the thin solid line, shown in blue online).

$p \to \infty$ the p-norm approaches the infinity-norm which is given by the largest entry in the state vector (or energy density field). We define $E_\infty(T)$ as the infinity-norm of the energy density field at time $t = T$ as

$$E_\infty(T) = \lim_{p\to\infty} E_p(T) = \max_{\boldsymbol{x}\in\Omega} e(\boldsymbol{x}, T). \tag{3.2}$$

The objective functional (3.1) is continuous with respect to $\boldsymbol{u}$, and we can thus naturally pose the problem of infinity-norm optimization of the energy. At this point, it is important to note that the objective functional is positive definite regardless of the value p, which explains our use of the term 'energy' no matter the value of p, which can in general take any real value. We nevertheless focus on the parameter range $p \geqslant 1$.

We incorporate the constraints imposed on our cost functional via scalar products with adjoint variables which act as independent quantities for the overall optimization. We enforce dynamic constraints stemming from the governing equations by adding them to the cost functional via adjoint velocity and pressure fields as these constraints have to be satisfied locally throughout the computational domain Ω. We impose additional constraints corresponding to the boundary and initial conditions in an analogous manner, choosing an appropriate inner product for the associated adjoint variable. We choose to normalize the initial condition $\boldsymbol{u}_0$ such that its energy density 1-norm E_1 is equal to unity:

$$E_1(0) = \frac{1}{V_\Omega}\int_\Omega \frac{1}{2}{\mathbf{u}_0}^2\,\mathrm{d}\Omega = 1. \tag{3.3}$$

By specifically accounting for the amplitude of the initial perturbation, the variational formulation is suitable even for a nonlinear analysis where the dependence on the initial amplitude enters as an additional parameter.

A pure p-norm amplification could be optimized by normalizing with respect to the p-norm instead of the 1-norm of the energy density field. However, we believe it is more appropriate to consider a finite initial integrated value of the energy, i.e. a finite initial value of E_1. Indeed, normalizing in a p-norm, and especially in the ∞-norm, would lead to highly delocalized solutions since only the maximum value of the initial condition would be constrained, but none of the other spatial components.

We now need to find the gradient of the cost functional (3.1) with respect to the initial condition $\boldsymbol{u}_0$ in order to employ an ascent method and to identify the optimal perturbation that will maximize the p-norm of the energy density field at $t = T$.

3.2. *Augmented Lagrangian and gradient information*

We merge the cost functional and constraints into an augmented Lagrangian $\mathscr{L}$ given

$$\mathscr{L}(\boldsymbol{u}, p, \boldsymbol{u}_0, \boldsymbol{u}^\dagger, p^\dagger, \boldsymbol{u}_0^\dagger) = \mathscr{J}(\boldsymbol{u}) - \left\langle \boldsymbol{u}^\dagger, \partial_t \boldsymbol{u} + \overline{\boldsymbol{u}} \cdot \nabla \boldsymbol{u} + \boldsymbol{u} \cdot \nabla \overline{\boldsymbol{u}} + \nabla p - \frac{1}{Re} \nabla^2 \boldsymbol{u} \right\rangle$$
$$- \langle p^\dagger, \nabla \cdot \boldsymbol{u} \rangle - (\boldsymbol{u}_0^\dagger, \boldsymbol{u}_0 - \boldsymbol{u}(0)), \tag{3.4}$$

where $\langle \cdot, \cdot \rangle$ and $(\cdot, \cdot)$, respectively, represent the spatiotemporal and spatial scalar products:

$$\langle a(\boldsymbol{x}, t), b(\boldsymbol{x}, t) \rangle = \frac{1}{V_\Omega} \int_0^T \int_\Omega a(\boldsymbol{x}, t) b(\boldsymbol{x}, t) \, \mathrm{d}\Omega \, \mathrm{d}t, \tag{3.5a}$$

$$(c(\boldsymbol{x}), d(\boldsymbol{x})) = \frac{1}{V_\Omega} \int_\Omega c(\boldsymbol{x}) d(\boldsymbol{x}) \, \mathrm{d}\Omega, \tag{3.5b}$$

with a, b, c and d being arbitrary functions. For simplicity, the time integration is not normalized by the time interval length T.

The Lagrangian $\mathscr{L}$ depends on the direct $(\boldsymbol{u}, p, \boldsymbol{u}_0)$ and adjoint $(\boldsymbol{u}^\dagger, p^\dagger, \boldsymbol{u}_0^\dagger)$ variables, and we find the optimal perturbation by equating to zero the first variation of $\mathscr{L}$ with respect to all of its independent variables.

The first variation with respect to the adjoint variables will reproduce the original constraints, i.e. the linearized Navier–Stokes equations and the initial condition. The first variation with respect to the direct variables produces a set of equations for the adjoint variables given by

$$\partial_t \boldsymbol{u}^\dagger + \overline{\boldsymbol{u}} \cdot \nabla \boldsymbol{u}^\dagger - \boldsymbol{u}^\dagger \cdot (\nabla \overline{\boldsymbol{u}})^\top + \nabla p^\dagger + \frac{1}{Re} \nabla^2 \boldsymbol{u}^\dagger = 0, \tag{3.6a}$$

$$\nabla \cdot \boldsymbol{u}^\dagger = 0, \tag{3.6b}$$

where $^\top$ is the transpose operator. In addition, the adjoint terminal condition at $t = T$ is found by taking the first variation of the objective functional with respect to $\boldsymbol{u}(T)$ and matching the result to the term $\boldsymbol{u}^\dagger(T)$, stemming from the integration by parts of the partial time derivative. We obtain

$$\boldsymbol{u}^\dagger(T) = \left(\frac{1}{V_\Omega} \int_\Omega e(T)^p \, \mathrm{d}\Omega \right)^{1/p-1} e(T)^{p-1} \boldsymbol{u}(T), \tag{3.7}$$

and the gradient of the cost functional with respect to the initial condition $\boldsymbol{u}_0$ is given as

$$\nabla_{\boldsymbol{u}_0} \mathscr{J} = \boldsymbol{u}_0^\dagger, \tag{3.8}$$

with $\boldsymbol{u}_0^\dagger = \boldsymbol{u}^\dagger(\boldsymbol{x}, 0)$. As expected, this gradient does not account for the unit-norm constraint imposed in (3.3). Even if starting from a point on the hyper-sphere defined by this norm constraint and marching in the direction of the gradient vector (3.8), initial conditions can result that violate this unit-norm constraint and we find that a simple gradient-descent algorithm is inadequate. We instead decide to follow Douglas, Amari & Kung (1998) and use a geometric optimization algorithm which explicitly imposes this normalization constraint (see appendix A for details).

4. A new look at energy production in a two-dimensional channel

4.1. *Numerical implementation*

We use a numerical code based on a fractional treatment of the various terms in the Navier–Stokes equations. The right-hand side of the direct governing equation (2.2) can be written as a sum of several terms as

$$\partial_t \boldsymbol{u} = \sum_{i=1}^{4} \boldsymbol{T}_i, \tag{4.1}$$

with

$$\boldsymbol{T}_1 = -\overline{\boldsymbol{u}} \cdot \nabla \boldsymbol{u}, \quad \boldsymbol{T}_2 = -\boldsymbol{u} \cdot \nabla \overline{\boldsymbol{u}}, \quad \boldsymbol{T}_3 = \frac{1}{Re} \nabla^2 \boldsymbol{u}, \quad \boldsymbol{T}_4 = -\nabla p. \tag{4.2}$$

The fractional-step algorithm treats each term separately by solving sequentially each of the four equations

$$\partial_t \boldsymbol{u} = \boldsymbol{T}_i. \tag{4.3}$$

The known velocity at the current time t is denoted by $\boldsymbol{u}^n(\boldsymbol{x}) = \boldsymbol{u}(\boldsymbol{x}, t)$ and $\boldsymbol{u}^{n+1}(\boldsymbol{x}) = \boldsymbol{u}(\boldsymbol{x}, t + \mathrm{d}t)$ is the unknown velocity field we wish to compute. First, advection by the base flow over a time interval $\mathrm{d}t$ is computed using a semi-Lagrangian advection scheme with 'back-and-forth error compensation and correction' (BFECC; see Selle *et al.* (2008)). The algorithm is as follows. The governing partial differential equation for advection is (4.3) for $i = 1$ and is solved using a semi-Lagrangian scheme which integrates the advection equation by following its characteristics. Under the assumption that $\boldsymbol{u}$ does not vary significantly over a time interval of length $\mathrm{d}t$, the advected flow at time $t + \mathrm{d}t$ (denoted $\boldsymbol{u}_1$) can be written as

$$\boldsymbol{u}_1 = \boldsymbol{u}(\boldsymbol{x} - \boldsymbol{u}\,\mathrm{d}t, t) = \boldsymbol{u}^n(\boldsymbol{x} - \boldsymbol{u}\,\mathrm{d}t). \tag{4.4}$$

The right-hand side is approximated by bilinear interpolation of the velocity vector $\boldsymbol{u}^n$ at the position $\boldsymbol{x} - \boldsymbol{u}\,\mathrm{d}t$ from the velocity values at the four surrounding grid points. The resulting velocity vector will be denoted by $\boldsymbol{u}_1^*$. In this procedure, two sources of error can be identified: the assumption of a constant advection velocity over the time interval and the interpolation. We denote the total error by $\boldsymbol{e}$. By transporting $\boldsymbol{u}_1^*$ backward in time using the same process (and the same time step), we find $\boldsymbol{u}^{n*}$, which gives us an approximation of the initial velocity field $\boldsymbol{u}^n$. Under the assumption that the accumulated error in either substep is the same, we can write

$$\boldsymbol{u}_1^* = \boldsymbol{u}^n(\boldsymbol{x} - \boldsymbol{u}\,\mathrm{d}t) + \boldsymbol{e}, \boldsymbol{u}^{n*} = \boldsymbol{u}^n(\boldsymbol{x}) + 2\boldsymbol{e}. \tag{4.5}$$

We thus can derive an expression for the error $\boldsymbol{e}$ by combining $\boldsymbol{u}^{n*}$ and $\boldsymbol{u}^n$, which can be exploited to achieve a second-order accurate approximation of $\boldsymbol{u}_1$ according to

$$\boldsymbol{u}_1 \simeq \boldsymbol{u}_1^* - \frac{1}{2}(\boldsymbol{u}^{n*} - \boldsymbol{u}^n). \tag{4.6}$$

The second advection term $\boldsymbol{T}_2$ is treated analytically by solving the partial differential equation (4.3) for $i = 2$ from t to $t + \mathrm{d}t$ which yields

$$\boldsymbol{u}_2 = \exp(-\mathrm{d}t \nabla \overline{\boldsymbol{u}}) \boldsymbol{u}_1 \tag{4.7}$$

where $\boldsymbol{u}_1$, the velocity found at the previous step (4.6), has been chosen as an initial condition. We solve the diffusion equation, (4.3) for $i = 3$, using a standard

finite-difference discretization of the Laplacian operator ∇^2 in space and a first-order implicit Euler scheme in time:

$$\boldsymbol{u}_3 = (\boldsymbol{I} - \mathrm{d}t\nabla^2)^{-1}\boldsymbol{u}_2. \tag{4.8}$$

We can express the incompressibility condition within a continuous formulation using (4.3) for $i = 4$. The time discretization of this partial differential equation is then equivalent to projecting the velocity vector onto the space of divergence-free functions (similar to Chorin's algorithm for simulating incompressible flows, see Chorin (1968)), yielding the velocity vector at step $n + 1$:

$$\boldsymbol{u}^{n+1} = \boldsymbol{u}_3 - \mathrm{d}t\boldsymbol{\nabla}p. \tag{4.9}$$

The pressure p is chosen appropriately to ensure that the velocity vector $\boldsymbol{u}^{n+1}$ is divergence free. Indeed, taking the divergence of (4.9) yields the following Poisson equation for the pressure

$$\nabla^2 p = \frac{1}{\mathrm{d}t}\boldsymbol{\nabla}\cdot\boldsymbol{u}_3, \tag{4.10}$$

which subsequently enforces the incompressibility condition (2.2). After this last stage, the procedure can be repeated to evolve the velocity field forward in time. The corresponding adjoint equations are solved in a completely analogous manner.

The time step $\mathrm{d}t$ is chosen such that $CFL = u_{max}\,\mathrm{d}t/h < 1$ (Courant–Friedrichs–Lewy condition, see Thomas (1995)), where $u_{max} = 1$ is the maximum non-dimensional advection velocity and $h = \min(\mathrm{d}x, \mathrm{d}y)$ stands for the smallest spatial discretization length. In practice, we choose $CFL = 0.5$. The spatial discretization consists of $N_x = 600$ points in the streamwise direction and $N_y = 200$ in the spanwise direction for a domain of dimension $L_x = 2\pi$ and $L_y = 2$. For domains of different size (which we use for relatively few computations associated with variations in horizon time), the resolution is adapted proportionally. Specific computations, which were deemed sensitive to discretization parameters, have also been performed at double the resolution to ensure grid convergence and mesh-independent results.

4.2. *Two-dimensional energy production: the Orr mechanism revisited*

As is apparent from (2.3) and (2.4), linear energy amplification in two-dimensional shear flows can only arise from the production term s since advection as well as the pressure term conserve energy while the viscous term dissipates or redistributes energy. Base flow shear is thus the sole source of energy production. The associated mechanism is known as the Orr mechanism (Orr 1907; Lindzen 1988): a mechanism based on Kelvin's circulation theorem which states that in an inviscid fluid the circulation around a contour moving with the local fluid velocity is conserved. By constructing an initial condition consisting of vortical structures tilted against the mean shear, we observe that the initial contour that encloses this structure shortens as it is distorted by the mean shear. As a consequence, the velocity along the contour has to increase to preserve the initial circulation. The largest velocities are obtained when the vortical structures tip over, after which the contour stretches again and the velocity along it has to decay. This explanation and previous research account for the global energy budget of the perturbation; a local description of the Orr-mechanism identifying the location(s) of maximum energy growth is lacking. While we will first establish a baseline case using the familiar 1-norm optimal perturbation, this analysis will serve as a reference point for more complex p-norm optimizations and localization studies.

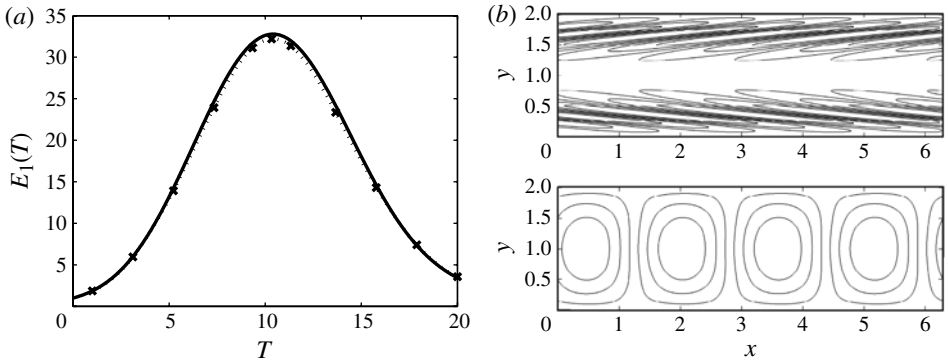

FIGURE 2. (*a*) The 1-norm energy, as defined in (1.2), given by the SVD analysis of the Orr–Sommerfeld operator (solid line) and by the direct-adjoint procedure (black crosses, interpolated by the dotted line). (*b*) Streamlines plots of the associated optimal perturbations (computed using the global variational framework) at $Re = 4000$. Optimal initial condition (top) and perturbation at $t = T_{opt} = 10.40$ (bottom).

The optimization of energy growth, using the technique presented in § 3 for $p = 1$, produces the well-known optimal initial perturbation in the familiar L_2 velocity norm. For this special case, we use a more efficient technique based on the formulation of the Orr–Sommerfeld operator followed by a SVD along with the direct-adjoint optimization algorithm in order to validate the present algorithm. We recall that for harmonic solutions the perturbation can be expressed in the form $\boldsymbol{u} = \mathrm{Re}(\hat{\boldsymbol{u}}(y)\exp(\mathrm{i}\alpha x))$, where $\hat{\boldsymbol{u}}(y)$ is the (complex) amplitude and α is the streamwise wavenumber. We choose a streamwise wavenumber $\alpha = 2$, following the work of Reddy & Henningson (1993). The optimal gain follows from the largest singular value of the evolution operator; for optimization horizons between $T = 0$ and $T = 20$ and for a Reynolds number of $Re = 4000$ we reproduce the results given by Reddy & Henningson (1993). The optimal time and gain at $Re = 4000$ are $T_{opt} = 10.4$ and $E_1(T_{opt}) \simeq 32$. These results are displayed in figure 2. As expected, the optimal initial perturbation consists of an array of vortices aligned against the mean flow shear, whereas the disturbance corresponding to $T = T_{opt}$ is composed of rectangular vortical structures normal to the shear direction (y-axis). The regions of largest velocities are located at the edge of each vortex, either close to the wall or in the centre of the channel. Strikingly, the streamlines (and energy) move from the wall towards the centre of the channel, even though the background flow is strictly parallel. In the next section we focus on the mechanisms driving this migration of energy density from both walls to the centre region.

4.3. *Spatiotemporal organization of the local energy density*

Figure 3 presents a sequence of snapshots displaying the temporal evolution of the 1-norm optimal perturbation. The main feature to note is that energy is initially produced in an intermediate region between the wall and the centre after which it is transported both towards the wall and towards the centre. This drift of energy density leads to a 'transition' from a single to two local maxima of the energy density field as time evolves. This is indicated in figure 3 using black dots which mark the local maxima of the energy density field.

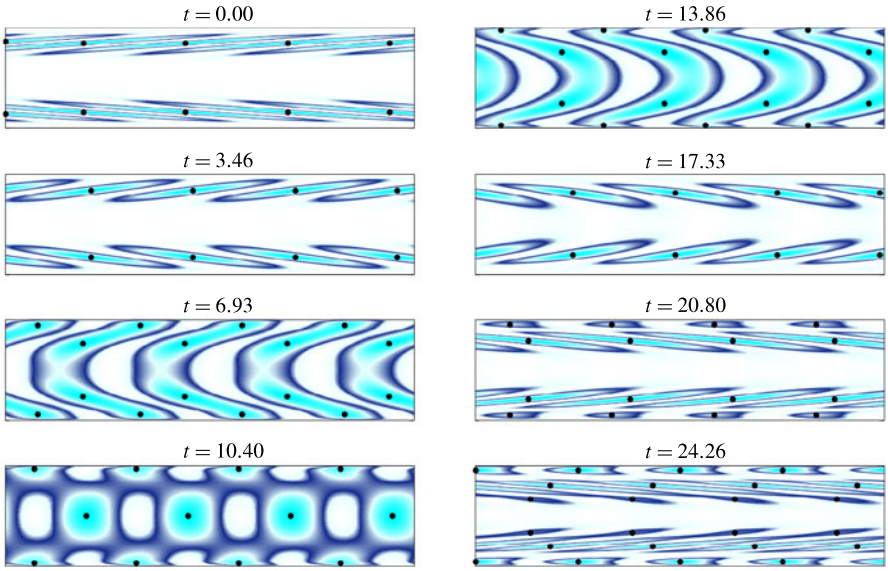

FIGURE 3. Evolution of the energy density e, as defined in (1.1), for the optimal perturbation computed for $T = T_{opt} = 10.40$. This figure illustrates the Orr mechanism: energy is gained by the tilting of vortical structures initially oriented against the mean shear. Black dots represent the positions of the local maxima of the energy density field.

A detailed understanding of the mechanisms involved in local energy production and spatial reorganization of the energy density field cannot, by definition, come from a global analysis. The reason for this lies in the pressure term i of the local energy evolution equation (2.3) that is responsible for a redistribution and localization of energy density structures but does not contribute to the global energy balance. Motivated by the structure formation in figure 3, we introduce an averaged quantity that is local in the normal direction but integrated over the homogeneous streamwise coordinate direction. For any local quantity q (such as, e.g., the terms labelled a, i in (2.3)) we define

$$q_y = \frac{1}{L_x} \int_0^{L_x} q(x, y, t)\, \mathrm{d}x. \tag{4.11}$$

With this definition we can now take a closer look at the different terms responsible for the local energy production in (2.3). We display in figure 4 the contributions of s_y and i_y describing, respectively, Reynolds stress production by mean-shear interaction and pressure-based energy redistribution, both as a function of y and t. The contribution from advection a_y is identically zero, and the dissipation term d_y is, by definition, strictly energy-dissipative with a maximum close to the wall during the large-amplitude phase. We find that the term r_y representing spatial redistribution due to viscosity is one order of magnitude smaller than the pressure term i_y, and we therefore consider it to be negligible in the spatial distribution of the energy.

Focusing on the channel centreline we observe that only the pressure term i_y has a non-zero contribution to the energy density balance: energy is produced by an interaction of the perturbations with the mean shear (by way of the term s_y), after which it is redistributed spatially along the y-axis by the pressure term i_y.

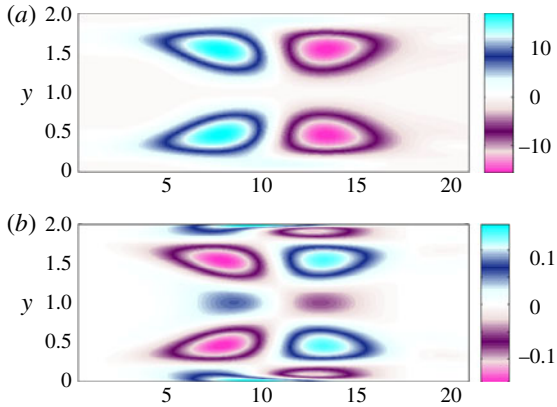

FIGURE 4. Spatiotemporal evolution of the streamwise-averaged energy density contributions (defined using (2.3) and (4.11)) from (*a*) mean-shear interaction s_y and (*b*) pressure-based redistribution i_y during the Orr-mechanism scenario corresponding to the 1-norm optimal perturbation for $t = T_{opt} = 10.40$.

Two locations for this redistribution are favoured: the centreline and the walls, corresponding to the production zones (blue regions) in the pressure contribution plot figure 4(*b*) around $t \simeq 7$.

The first qualitative observation based on the snapshot sequence in figure 3 is further corroborated by tracking in time the positions of local maxima of the energy density which we plot in figure 5. We observe that for $t \simeq 7$ a new local (in space) maximum appears close to the wall at $y = 0.15$ (and its symmetric counterpart with respect to the centreline). This position is clearly closer than the initial energy density maximum for the initial optimal perturbation which was located at $y = 0.31$. This appearance coincides with a marked displacement of the initial local maximum toward the centreline of the channel where they finally merge at $t \simeq 8.8$ into a single maximum. This structure eventually vanishes in the same fashion as it appeared: the central maximum breaks up into two local maxima (at $t \simeq 13.1$) which subsequently move toward the region that originally contained the maximum of the initial condition. We plot the energy associated with each of these local maxima in figure 5(*b*). The maximum value of the energy density is reached at $t = T_1 \simeq 10.6$ and corresponds to the centreline maximum. The energy density of the wall-bounded local maximum reaches its largest value for a time slightly shorter than T_1 (i.e. $t = T_2 \simeq 10.2$). Unsurprisingly these two local optimal times are very close to $T_{opt} = 10.4$. However, it is an interesting feature that the wall local maximum reaches its largest value earlier than the centre one and that the mean of the local optimal times corresponds exactly (although perhaps coincidentally) to the global optimal time $T_{opt} = (T_1 + T_2)/2$.

The above observations suggest that there co-exist two distinct sources of energy growth: one located in the wall regions, the other acting in the centre of the channel. The wall-bounded energy production mechanism is needed to maximize the integrated value of the energy (in the $p = 1$ case) but is apparently not the most dangerous. Rather, the centreline energy density peak has the largest value and should thus be considered the most pertinent source of transient growth. However, the norm used so far (the 1-norm of the energy density) is not able to distinguish between these two mechanisms since they both contribute to its maximization. The aim of the following

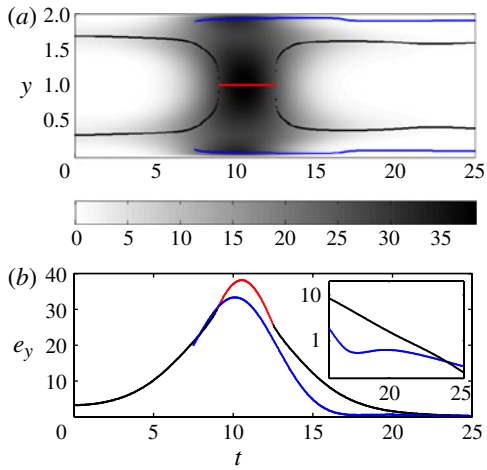

FIGURE 5. (*a*) Cross-stream location y of the local maxima of the energy density (coloured dots: black for the local maxima associated with the centre local maximum, red for the centre local maximum itself and blue for the wall maximum) and energy density integrated in the x-direction (black and white shading) as a function of time t. (*b*) Value of the energy density field at these locations (using the same colours as in (*a*)) as a function of t. These plots correspond to the time evolution of the 1-norm optimal solution for $T = T_{opt} = 10.40$. The inset shows the dominance of the wall local maximum for the late times.

section is to use a higher-order norm (p-norm for $p > 1$) in order to detect and isolate the mechanisms of energy density production in two-dimensional shear flows in general and in plane Poiseuille flow in particular.

5. p-norm optimal perturbations

Having described the mechanism leading to the energy growth for the 1-norm optimal perturbation, in this section we present the results of a p-norm optimization, for $p > 1$. It is important to distinguish between the two usages of the expressions local and global. Depending on context, we use local and global to describe the spatial character of maxima of the energy density of the perturbation, and alternatively to describe the nature in solution space of the identified maxima of the cost functional $\mathscr{J}$. We will find in particular that the p-norm cost functional is multimodal, in the sense that it allows a rich variety (local or global) of maxima and is thus able to separate the two spatial components of energy growth (centre and wall) identified in the previous section.

5.1. *General observations*

In the derivation of the p-norm final energy density optimization, we can identify the step in the algorithm (see the adjoint initial condition (3.7)) where the p-norm directly localizes the energy field around its maximum leading theoretically to both maximum final energy-value optimization and a strong localization effect. If we consider the limit of large p, the adjoint initial condition approaches a delta function centred on the maximum value of the direct energy density field. When advected backwards in time, this will therefore lead to spatially localized gradient information making the optimization procedure ultimately converge toward a localized initial perturbation.

From this simple fact, we can infer the regularity of the problem with respect to the optimization time and discuss the role of the initial guess in the optimization algorithm, which is commonly chosen to be random noise. Let us first consider optimization over a short time interval. At the end of such a short time integration, there is not likely to be any preferential region within the domain for the (spatial) global maximum to occur. Indeed, in some sense, any particular physical process is unlikely to have had enough time to act on the random noise to extract a global behaviour. Therefore, during the localization step (3.7), energy is focused in a largely arbitrary fashion, and the optimization algorithm will converge toward a local optimum (of the cost functional) arbitrarily localized in space since from the very first localization step, any possibilities of exploring other areas of the solution space of initial conditions would have been removed. Indeed, it is important to appreciate that for short times the gradient (3.8) is not smooth and that several local maxima of the cost functional may well exist. The limitation of any gradient method can be felt since it will become trapped in a local optimum instead of exploring other potential optimal candidates in the solution space. However, over a large number of Monte–Carlo realizations, we would expect to find the global optimal perturbation, particularly for small time horizons.

For larger times, physical mechanisms have time to organize the (initially random) field coherently, and some of the features of the 1-norm optimal solution can be identified. Therefore, the same global maximum is more likely to emerge from time integration, leading through the same process of localization, backward adjoint integration and gradient descent to a maximum of the cost functional which has some physical meaning, corresponding in some way to the global behaviour of the perturbation.

However, it is also possible that other local maxima of the cost functional exist and do not coincide with the global behaviour. Indeed, we can imagine a complex solution space for which a random noise time integration gives a localized energy density in the zone that produces the largest amount of perturbation energy growth. A second zone of energy production can however exist and be identified only if the initial guess is not random noise but already localized around this zone. This observation raises the question of the most appropriate form of the initial guess for such optimization problems, a question we consider in more detail below, when considering the actual solutions we identify. In effect, any random field with no spatial weighting will most likely lead to the identification of an optimal perturbation which is the closest to the global behaviour.

As the length of the optimization time interval increases, the many an various cost functional local optima that exist for short times are smoothed out by the coherent mechanisms acting on the flow. In our simple case of a two-dimensional incompressible flow, the only relevant time scale is given by the diffusion coefficient, quantified by the Reynolds number *Re*. The larger the Reynolds number, the longer the time to experience the effect of coherent mechanisms will be.

5.2. *Multiple solutions*

Based on the observations in the previous section, we choose the length of the optimization time interval to be the global 1-norm optimal time denoted by T_{opt}; this choice will avoid (for the moment) the problem of local maxima for small optimization times. The length of the domain is chosen to be $L_x = 2\pi$. This choice is based on the requirement that localized solutions found for this optimization time should never interact with themselves due to the periodicity of the domain; a localized

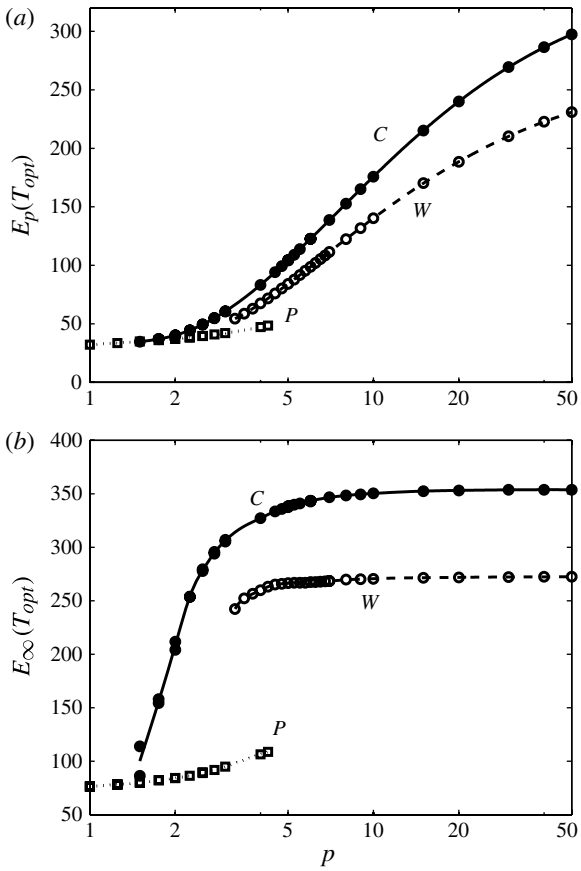

FIGURE 6. Energies $E_p(T_{opt})$ (*a*) and $E_\infty(T_{opt})$ (*b*) as a function of p for $Re = 4000$. The dotted, solid and dashed lines represent the periodic (*P*), centre (*C*) and wall (*W*) solutions branches, respectively. Each symbol represents a completed optimization (squares, filled and open circles denote, respectively, periodic, centre and wall solutions). The plateau that both wall and centre solutions reach in the $E_\infty(T_{opt})$ diagram shows that the p-norm optimization converges to the ∞-norm optimization for a finite and relatively small value of p ($p \simeq 5$ for the wall solutions branch and $p \simeq 10$ for the centre solutions branch).

(periodic) solution thus never overlaps with its adjacent structure. A longer time horizon would require the use of a larger domain to avoid such overlapping. In § 6, where we vary the optimization time horizon T, we adapt the size of the domain appropriately to avoid any self-interaction due to overlapping. We then compute the p-norm optimal perturbation for several values of p, starting from the familiar case $p = 1$ to larger values of p (up to $p = 50$). We present results of these computations in figure 6 where we plot the value of $E_p(T)$ at the end of the time integration, the quantity to be optimized (figure 6*a*), as well as the corresponding maximum of the energy $E_\infty(T)$ (figure 6*b*) .

As alluded to in the previous section, we find several solutions to the optimization problem for increasing values of the parameter p, which we can classify into three branches. The first periodic '*P*' solutions branch (square symbols), which exists for relatively small values of p, results from a continuation of the optimal solution for

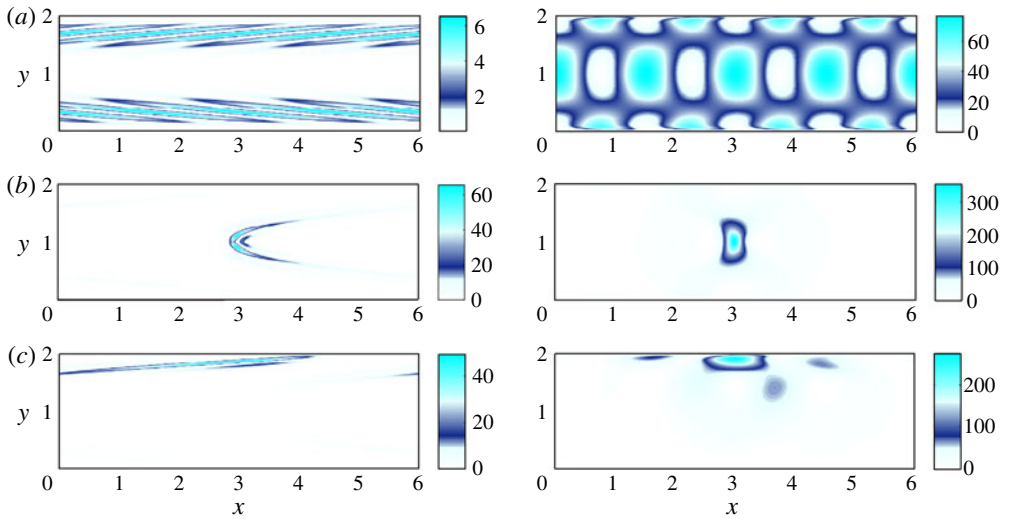

FIGURE 7. Energy density plots for three solutions of the p-norm optimization problem: (a) periodic P solution ($p = 1$); (b) centre C solution ($p = 50$); (c) wall W solution ($p = 50$). The left-hand side represents the initial perturbation at $t = 0$ while the associated right-hand side displays the solution at the final time $t = T_{opt}$. For clarity the localized solutions (at both $t = 0$ and $t = T$) have been placed in the centre of the domain. For this reason, the x-axis cannot be used to deduce the advection length.

$p = 1$ (see figure 6a) to higher values of p. These solutions are periodic in x, with a dominant wavenumber of $\alpha = 2$. We plot the corresponding spatial shape in figure 7(a). For larger values of p, P-branch solutions seem to cease to exist ($p \simeq 4.25$). Strictly speaking, there is no reason for this branch to disappear fully. It is more likely though that its basin of attraction in solution space shrinks drastically as p increases. Indeed, for larger values of p any numerical errors (due to discretization, convergence or even machine precision) are amplified. Therefore, even when starting from a previous periodic solution (for slightly smaller p), step (3.7) will converge toward the global maximum as long as a discrepancy, no matter how small, between several spatial local maxima of the energy density field exists at the end of the time integration. This explains why we no longer can track this branch for larger values of p. Choosing an analytical first guess (at least in the x-direction) may remedy this problem. However, we did not pursue such an approach since we are focused on localized solutions in this paper, thus rendering this periodic branch less interesting; moreover, for larger values of p, this branch is evidently suboptimal.

However, two other localized initial perturbations, maximizing the p-norm at time $T = T_{opt}$, arise: the centre 'C' and wall 'W' solutions branches, labelled according to their respective region of localization. The second C-branch (filled circles) corresponds to centre solutions which could be traced to values of p as low as $p \simeq 1.5$ where they gradually merge with the periodic solutions. We find this branch straightforwardly for all values of p and this branch yields the optimal energy density p-norm and thus the global maximum for $t = T_{opt}$. Along this solutions branch, the spatial organization of the energy density field shows a progressive concentration of the initial perturbation in the centre of the channel as p increases. The third suboptimal W-branch exists for $p \gtrsim 3.25$ (open circles) and corresponds to localized wall structures. The 1-norm

energy gain is of course optimal for $p = 1$; therefore, any solution corresponding to $p > 1$ has diminished energy amplification with respect to the 1-norm gain. For large values of p, two localized solutions, the centre and wall solutions, remain, whose spatial energy density distributions we plot in figure 7(*b*,*c*) respectively. In either case, they take on the form of a unique localized spot (or 'hotspot') of energy density at the end of the optimization interval, either in the centre or near one wall of the channel. The uniqueness of this spatial maximum at the end of the time integration explains why the wall solution W cannot be symmetric. If symmetrized, the time evolution of such a solution would produce two hotspots of energy located at each wall and, consequently, the maximum energy would be decreased by a factor very close to two, and would therefore be clearly suboptimal.

Increasing the order p thus leads to a more complex cost functional, characterized by an appearance of several local maxima, in contrast to the $p = 1$ case, for which the maximum is unique. In order to find these different branches for increasing values of p (and below in § 6, for varying time horizons T), an appropriate initial guess for the optimization process is needed. To this end, we perform several optimizations starting from random noise. In most cases, we find that the random noise lies within the basin of attraction of the global optimal C-branch. For relatively few cases, the optimization converged toward a solution located on the W-branch, while we never manage to find the P-branch for $p > 2$. It is important to appreciate that random noise is not likely to explore the entire solution space, particularly for high-dimensional problems. In order to focus on the W- or C-branch, we find it advantageous to start the optimization routine from localized 'random noise', focused either in the centre of the domain (for the C-branch) or at the edge of the channel (for the W-branch). Moreover, whenever possible we use a continuation procedure (as explained previously for the P-branch) to track the different solutions branches.

The main result of this section is reflected in the observed saturation in the E_∞ plot (see figure 6*b*): both centre and wall branches converge to a finite value of E_∞, long before the limit $p \to \infty$ is reached. For finite, and actually rather small, values of p, we are thus able to obtain optimal solutions in the ∞-norm. This phenomenon can be explained physically with an argument based on two competing length scales. The only step in the optimization algorithm where the choice of norm has a direct impact on the energy density field is the so-called localization step (3.7). The value of p controls the strength of localization effects (via the adjoint terminal condition) on the direct energy density maximum. Two length scales thus play a role: the localization length scale l_p and the diffusion length scales l_{Re}. The latter is a fixed parameter of the problem, only dependent on the Reynolds number and quantify the minimum characteristic length over which fine-scale structures are not smoothed out by diffusion, while l_p is the characteristic scale of localization associated with a given p-norm optimization problem. Since localization is stronger for larger values of p, the length scale l_p decreases as p increases. When $l_p \simeq l_{Re}$, a further increase of p will not affect the results since diffusion will act to smooth any further localization of the adjoint initial condition. For this reason, ∞-norm solutions can be computed for finite values of p (when E_∞ approaches a constant value). Apparently and quite reasonably, this limit value of p (denoted by p_{lim}) seems to be an increasing function of the Reynolds number Re, as we demonstrate in appendix B where we consider a flow at $Re = 250$.

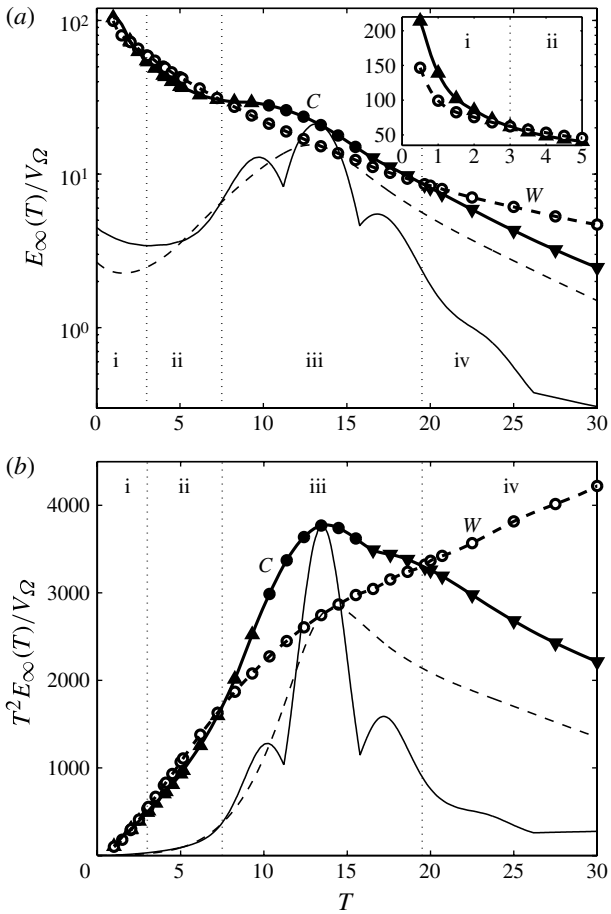

FIGURE 8. (*a*) Normalized (by the computational domain volume) optimal $E_\infty(T)/V_\Omega$ curves for the wall (W) and centre (C) solutions. Open circles (interpolated by the thick dashed line) represent the wall solutions while filled circles and upward and downward triangles (interpolated by the thick plain line) represent symmetric and asymmetric centre solutions of type I and II, respectively. For very short times ($T \lesssim 3$, period i), the centre solution is optimal, after which it becomes the second local optimum for $3 \lesssim T \lesssim 7$ (period ii). Then, for $7 \lesssim T \lesssim 20$ (period iii), the most efficient way to produce localized energy is based on the centre solution. Finally, for times larger than $T = 20$, the wall solutions prevail again as the optimal solution, yielding the least energy decay. The thin, plain and dashed lines correspond to the time evolution of the maximum energy E_∞ as a function of time for the wall and centre optimal perturbation computed at $T = 13.5$, where they are tangent to their corresponding optimal envelope. The inset figure represents the computations for short times which were carried out with a smaller ($L_x = \pi$) domain for accuracy issues. (*b*) Optimal $T^2E_\infty(T)$ curves for the wall (W) and centre (C) solutions. Under this scaling, a purely diffusive process would appear as a constant function. The symbols are the same as in (*a*).

6. Variation of the optimization time horizon T

Recalling our motivating arguments from the introduction, the problem of optimizing the maximum value of the energy density field is of great interest in many applications where information about the details of potential energy hotspot formation is crucial. For a fixed optimization time horizon ($t = T_{opt} = 10.4$), we identify two

new localized solutions branches: the optimal centre solutions C and the suboptimal wall solutions W. Following this analysis, it is legitimate to question the dominance of the centre solution for times other than the global optimal time T_{opt} and to investigate time horizons for which wall solutions might dominate the optimal ∞-norm energy density mechanism. To this end, we determine ∞-norm optimal perturbations (as the limit of large p) for varying time horizons, ranging from $T = 0.5$ to $T = 30$, and plot our results in figure 8. We perform these computations in a channel of length $L_x = 4\pi$, except for very short times (see the inset in figure 8) for which we used a smaller box of length $L_x = \pi$.

Figure 8(a) represents the ∞-norm of the energy density field at time $t = T$ for several optimization time horizons T. A general decaying behaviour for the optimal maximum value of the energy density E_∞ as a function of the time T is observed, which can be attributed to diffusion. For short optimization time intervals, the best way to achieve a large final value for the maximum local value of energy density is to start from an initial condition that closely resembles a delta function, which explains the very large values of $E_\infty(T)$ when T is small. For larger values of T, diffusion has sufficient time to affect the flow, resulting in a significant reduction of the maximal energy density. However, other energy production mechanisms come into play (such as, the Orr mechanism, pressure effects, etc.) which explains the departure from a simple diffusive decay and the richness and complexity of the solutions branches. Under a purely diffusive process, we would expect an amplitude decay proportional to $1/t^2$. We therefore replot in figure 8(b) the results presented in figure 8(a) but multiply the maximum energy by T^2 to compensate for simple diffusion. A purely diffusive decay would thus correspond to a constant function as $T \to \infty$ within this scaling.

A noticeably broader family of solutions is observed compared with the fixed-horizon time optimization. The previously identified W-branch is present at all horizon times considered and we plot it with open circles. It is characterized by a persistent spatial shape of the perturbations near the channel wall over the range of considered values for T. The centre solutions, on the other hand, do not follow such a simple behaviour as the horizon time is varied. In fact, several centre solutions are observed and are plotted with full symbols. For short times, the centre solution is not symmetric. It is nonetheless denoted as a centre solution since it represents the short-time continuation of the symmetric centre solution observed for $t = T_{opt}$ in the previous section. This solution will be referred to as the asymmetric centre solution of type I. The expected symmetric centre solution is found for horizon times ranging from $T \simeq 10$ up to $T \simeq 16$, at which point a new branch appears (corresponding to the new optimal centre solution for larger times); we refer to the solutions on this branch as asymmetric centre solutions of type II. More precisely, for $T \simeq 10$ the asymmetric centre branch merges with the centre branch (which was suboptimal for shorter times). For $T \simeq 16$ the opposite phenomenon happens: the symmetric centre branch becomes suboptimal and a new asymmetric centre branch takes over for longer times. For clarity, we choose not to plot the significantly suboptimal (symmetric) centre solutions for short and long times.

We plot the energy density distribution of these two distinct solutions in figure 9.

Figure 8 clearly shows competition between the different, previously identified solutions. The mechanism of production of the optimal hotspot of energy is not unique in time. For very short optimization times, the asymmetric centre solutions are marginally better choices (compared with the wall solutions) at creating such localized zones of high kinetic energy density. This period, identified by the roman numeral i, starts at $T = 0.5$ (no computations were performed for shorter time intervals) and

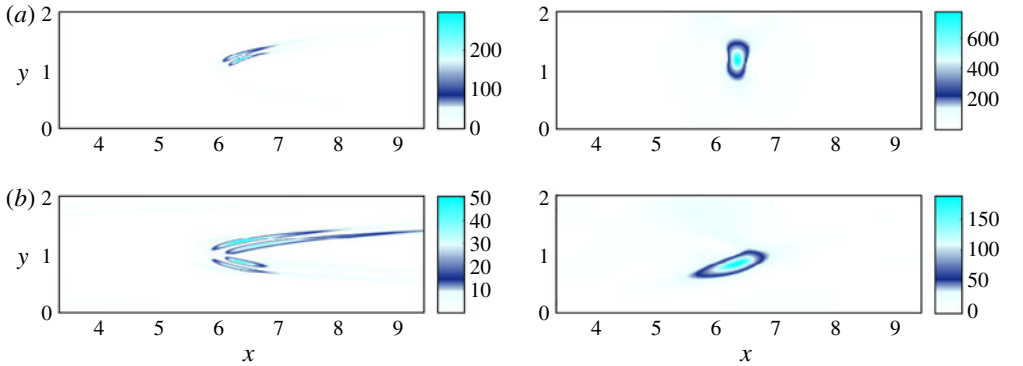

FIGURE 9. (*a*) Energy density plots of the asymmetric centre solution of type I (for $T = 8$) and (*b*) asymmetric centre solution of type II (for $T = 20$). The left figures represent the initial perturbation at $t = 0$ and the right figures display the solution at the final time $t = T$. For clarity the localized solutions (at both $t = 0$ and $t = T$) have been placed in the centre of the domain. For this reason, the x-axis cannot be used to deduce the advection length.

lasts until $T \simeq 3$. A second period ii for which wall solutions are the optimal solutions to the ∞-norm optimization follows and corresponds to optimization horizons in the interval $3 \lesssim T \lesssim 7$. After this time, a new change between wall and centre solutions occurs and marks the start of period iii, which ends at $T \simeq 20$. This period is more complex than periods i and ii since, as briefly explained in the previous paragraph, the centre solutions are not always of the same type within this optimization time interval. Indeed, the solution consists first of an asymmetric centre solution of type I which then becomes symmetric for $T \simeq 10$. This symmetry is then broken for $T \simeq 16$ where qualitatively different asymmetric centre solutions of type II dominate. The final period iv starts at $T \simeq 20$ and is associated with wall solutions which decay at the (slowest) optimal rate.

In figure 8, we also plot the temporal evolution of the global maximum of the energy density for two specific optimal centre and wall solutions, computed for $T = 13.5$. In accordance with the definition of our optimization problem, the two curves lie below their respective optimal envelopes and are tangent to it at the optimization time ($T = 13.5$). The wall solution shows a smooth temporal evolution (in the sense that the location of the energy maximum is continuously evolving in space), while kinks appear in the centre solution curve, demonstrating that a discontinuity in the maximum of the energy density field location occurs around $t \simeq 12$ and $t \simeq 16$. The centre solution displays a marked peak of energy at the optimal time, which subsequently decays rather quickly past the optimal time. After the optimization time, both solutions experience decay that is stronger than a pure diffusion process would produce; this is indicated by the decreasing slope in the $T^2 E_\infty(T)$ plot given in figure 8(*b*). However, the decay rate of the centre solution is substantially larger than the decay rate of the wall solution, which explains why, for longer optimization times, the wall solution is favoured.

Finally, as discussed in more detail in appendix C, although the cost functional appeared to converge strongly to an extremal value, it is important to appreciate that more rigourous gradient-based convergence tests suggest that some of our identified solutions may have imperfectly converged. Therefore, the precise numerical values of

$E_\infty(T)$ and the times separating the various periods should be thought of as subject to slight possible change.

7. Summary and conclusions

Iterative optimization techniques based on direct and adjoint equations have become commonplace in fluid mechanics and are routinely applied to compute optimal initial conditions, optimal control strategies or optimal geometric modifications to a flow configuration. The objective in these optimization studies is often based on (or related to) the kinetic perturbation energy of the flow. The present study has followed this general technique but has abandoned the traditional energy measure (L_2-norm of the velocity field) in favour of a p-norm of the energy as defined in (1.3) (for $p > 1$) which promotes a strong spatial localization effect on the quantity to be optimized. With this technique, a far more detailed analysis of the perturbation dynamics is feasible by isolating localized regions of the flow and assessing their contribution to the local energy density evolution as a function of time. A p-norm analysis (in particular, for large values of p) thus produces a fuller and richer picture of the physical mechanisms of momentum and energy transport.

To demonstrate this analysis, we consider the simple case of two-dimensional channel flow. The p-norm type of analysis provides insight into the mechanisms acting during the transient amplification of global energy. The local character of the energy density e as defined in (1.1), however, yields more detailed information about the reorganization of fluid structures than the more common global energy analysis; in particular, processes with a zero *global* net contribution are retained *locally* and furnish valuable information about the redistribution of energy density over time. Adopting this local point of view, it is possible to reanalyse the periodic 1-norm optimal perturbation in terms of the local dynamics of local maxima of the energy density field, and the associated local energy production mechanisms. The well-known Orr mechanism naturally occurs, where energy is first amplified close to the wall via classical shear tilting before it migrates toward the centre of the domain due to pressure effects (ensuring the incompressibility of the flow). Around the optimal time, the energy density is mainly localized in the centre and close to the wall, suggesting two energy production mechanisms. For larger times, the single centre maximum splits and drifts away from the centreline, as shown in figures 3 and 5.

In order to explain this simple time evolution analysis, we perform a p-norm optimal perturbation study, with the intention to magnify and interpret the local aspects of the energy density dynamics, to isolate a hierarchy of processes associated with localized structures and to study their dominance (or subdominance) in time. Starting with a constant time horizon and a smooth increase in the norm parameter p, two main mechanisms leading to such localized states can be identified in plane channel flow which we have chosen to refer to as the centre C and wall W solutions branches. The underlying mechanisms leading to centre and wall solutions are competing in time to produce the most energetic spot of localized energy. The wall solutions are optimal (with respect to the energy ∞-norm) for relatively short and long times, while the centre solutions are optimal for very short and intermediate times only.

Our results illustrate that, even for as simple a configuration as two-dimensional channel flow, the search for and analysis of highly localized and energetic perturbations is by no means trivial. The departure from classical energy measures towards localization-promoting p-norms yields a more complex but richer view of energy transport processes. As such, p-norm optimization holds great promise in

detecting and separating several local (in space) energy production processes, in gaining more insight into intricate stability and transition scenarios and in quantifying worst-case measures about a particular flow configuration. We believe that more complex flows with multiple, competing instability mechanisms (e.g. in a jet in cross-flow) would benefit from a p-norm analysis, and indeed, generalizations of the introduced framework to p-norm optimization of (localized) frequency responses, control strategies, wall modifications or shape optimizations are conceivable and would reward being the focus of future efforts.

Acknowledgements

The research activities of D.P.G.F. and C.P.C. are supported by EPSRC Research Grant No. EP/H050310/1 'AIM (Advanced Instability Methods) for Industry'. They would also like to acknowledge the generous hospitality of the Hydrodynamics Laboratory (LadHyX) École Polytechnique/CNRS during the production of this manuscript.

Appendix A. Implementation issues: unit hyper-sphere optimization

In this appendix we give details of the numerical algorithm we use to determine p-norm optimal solutions. We start by presenting the limitations of standard techniques to enforce a unit-norm constraint such as (3.3), after which we introduce an efficient alternative based on rotations rather than on gradient descent.

The unit-norm constraint determines the initial energy level of the perturbation and is thus very significant in the optimization of a state vector satisfying the full nonlinear Navier–Stokes equation. In previous studies, the normalization constraint (3.3) has been embedded in the variational formulation via a new (scalar) Lagrange multiplier (usually denoted λ) whose value is determined by requiring the new initial condition to satisfy the constraint (3.3). The objective of the next section is to explain our reasons for departing from this widely used method.

A.1. *Lagrangian formulation for unit-sphere constraints*

For simplicity, we will consider the optimization problem in its discrete form. The state vector is denoted by $\boldsymbol{X} \in \mathbb{R}^N$, with N being the total number of degrees of freedom. The variable J represents the cost functional, a mapping of $\mathbb{R}^N \to \mathbb{R}$. We denote the adjoint vector and the cost functional gradient by $\tilde{\boldsymbol{X}}$ and $\boldsymbol{\nabla}_{\boldsymbol{X}} J$, respectively. The normalization constraint (3.3) can thus be rewritten as

$$\|\boldsymbol{X}\|_2 = (\boldsymbol{X}^\top \boldsymbol{X})^{1/2} = \left(\sum_{i=1}^{N} X_i^2\right)^{1/2} = C, \tag{A 1}$$

where C is a constant. After the normalization constraint is enforced via a Lagrange multiplier λ, the gradient depends explicitly on the direct variable $\boldsymbol{X}$:

$$\boldsymbol{\nabla}_{\boldsymbol{X}} J = \tilde{\boldsymbol{X}} - \lambda \boldsymbol{X}. \tag{A 2}$$

The simplest gradient update at each iteration n is thus

$$\boldsymbol{X}_{n+1} = \boldsymbol{X}_n + \epsilon \boldsymbol{\nabla}_{\boldsymbol{X}_n} J = (1 - \lambda\epsilon)\boldsymbol{X}_n + \epsilon \tilde{\boldsymbol{X}}_n, \tag{A 3}$$

with $\boldsymbol{X}_{n+1}$ denoting the new initial condition, which is as yet undetermined since λ is still unknown, and ϵ is the (assumed 'small') step size. Applying the normalization

constraint (A 1) to the vector $\boldsymbol{X}_{n+1}$ produces the following quadratic equation in λ

$$\lambda^2 - 2\lambda\left(\frac{1}{\epsilon} + \frac{A}{C}\cos\theta\right) + \left(\frac{A}{C}\right)^2 + \frac{2A}{C\epsilon}\cos\theta = 0, \tag{A 4}$$

with $A = \|\tilde{\boldsymbol{X}}\|_2$ and θ representing the angle between $\boldsymbol{X}$ and $\tilde{\boldsymbol{X}}$. This equation admits real solutions when the determinant is positive or zero, for which ϵ has to satisfy the following inequality

$$\epsilon_{min} \equiv -\frac{C}{A\sin\theta} \leqslant \epsilon \leqslant \frac{C}{A\sin\theta} \equiv \epsilon_{max}. \tag{A 5}$$

We will only consider positive values of ϵ since the goal of the optimization is to find a maximum value for J. Under the assumption of real-valued solutions (i.e. with ϵ falling within the appropriate interval), (A 4) yields two values, λ_- and λ_+, given by

$$\lambda_\pm = \frac{1}{\epsilon} + \frac{A}{C}\cos\theta \pm \left(\frac{1}{\epsilon^2} - \left(\frac{A}{C}\right)^2 \sin^2\theta\right)^{1/2}. \tag{A 6}$$

The root λ_- corresponds to the gradient update, while λ_+ represents a spurious solution, as shown in figure 10. It is noteworthy that the solutions $\lambda_\pm$, and consequently the gradient itself, depend on the step size ϵ chosen for the update, which makes a line search a tedious process as the above equation has to be solved at each step. Moreover, such a line search for an optimal step size is confined to the circular arc of valid updates (with $0 < \epsilon \leqslant \epsilon_{max}$). Strictly speaking, the entire circle is reachable from the vector $\boldsymbol{X}$, but it would require the consideration of both roots (λ_- and λ_+) of (A 4), as well as the case $\epsilon < 0$. Let us denote by α the angle between $\boldsymbol{X}_n$ and the updated vector $\boldsymbol{X}_{n+1}$. We see that the Lagrangian update (if we consider valid updates only) is restricted to updates corresponding to $0 < \alpha < \pi/2$ only. The double root case ($\epsilon = \epsilon_{max}$) corresponds to $\alpha = \pi/2$.

The particular case $\epsilon = C/A$ is remarkable as it corresponds to the 'power iteration' update for which any information about the current state vector $\boldsymbol{X}$ is omitted; for this value of ϵ we obtain $1 - \epsilon\lambda_- = 0$, but a non-zero value when substituting the spurious root λ_+. In this case, the update is purely in the gradient direction $\tilde{\boldsymbol{X}}$ and simply given by

$$\boldsymbol{X}_{n+1} = \frac{C}{A}\tilde{\boldsymbol{X}}_n. \tag{A 7}$$

For a conjugate gradient method, the analogous equation for λ becomes an even more complicated, quartic equation (because the update is second-order in the gradient); conditions for the existence of real solutions and the choice of legitimate values for λ from among the four roots are both non-trivial. Moreover, if one were to constrain a general p-norm (for $p > 1$), the corresponding λ equation is at least of order $2p$ for a simple gradient update, and iterative techniques have to be used to determine its solution.

In summary, the Lagrangian treatment of the unit-sphere constraint is complicated (due to the high-order λ equation) leading to an overly involved gradient evaluation and to a limited range of 'valid' state-vector updates.

A.2. *Updating through rotation*

To avoid the complications outlined in the previous section, we propose to retain the gradient information (3.8), but alternatively use geometric arguments (following

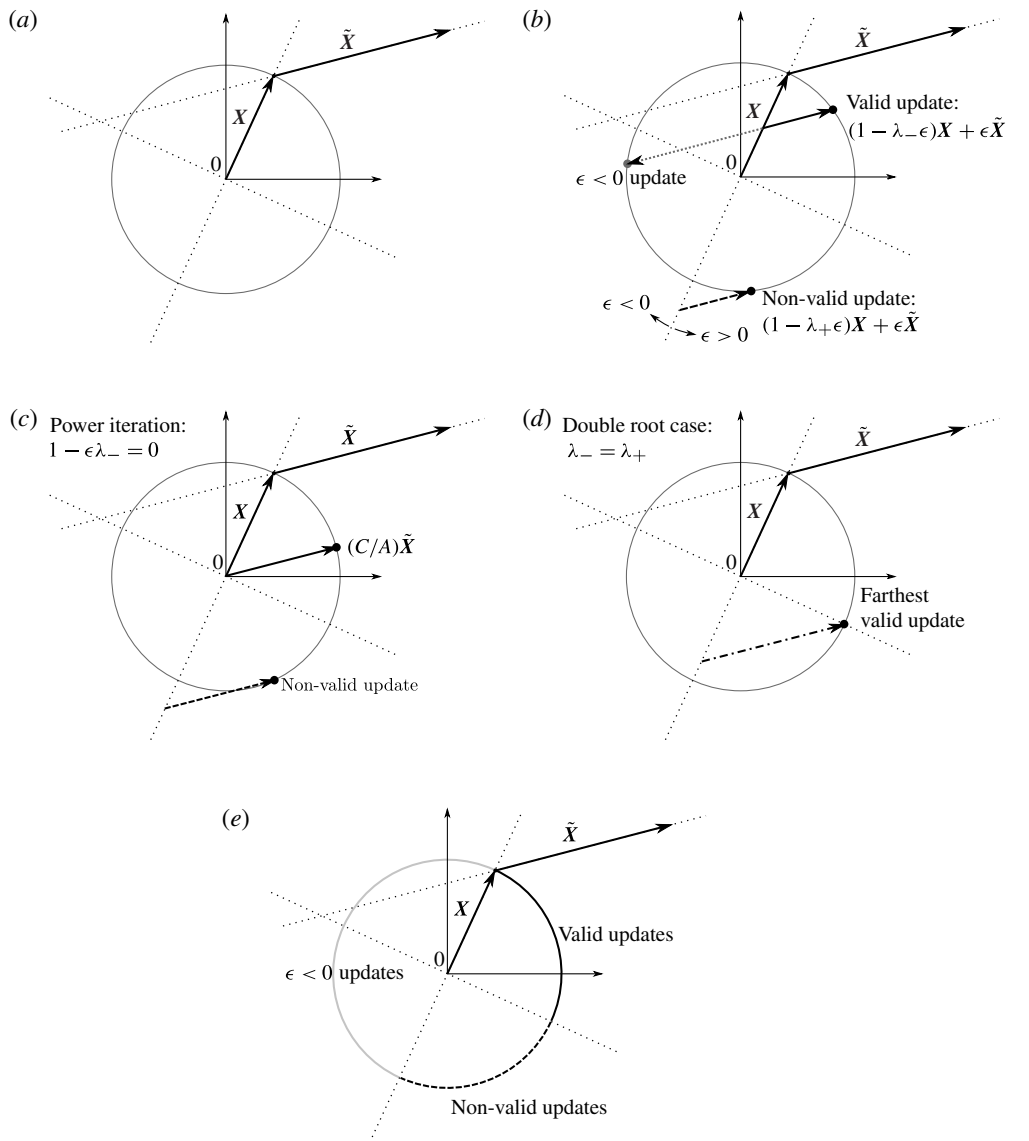

FIGURE 10. Constrained gradient ascent on a unit circle using standard Lagrangian techniques. (*a*) State vector $\boldsymbol{X}$ and gradient representation $\tilde{\boldsymbol{X}}$. (*b*) The circle is divided into two zones determined by the sign of the step size ϵ, as defined implicitly in (A 3). Two solutions for a $\epsilon > 0$ update are possible, corresponding to the two roots (λ_-, λ_+) determined as defined in (A 4), where λ_- yields a valid update (close to the initial state vector) while λ_+ produces a spurious update. (*c*) Power iteration case: the update only follows the gradient direction $\tilde{\boldsymbol{X}}$ (see (A 7)). (*d*) Furthest possible update for $\epsilon = \epsilon_{max}$; this case corresponds to the double root case $\lambda_- = \lambda_+$. (*e*) Different regions of the circle of update. The locus of valid updates is limited to the first quarter of circle. (All panels display the $(\boldsymbol{X}, \tilde{\boldsymbol{X}})$ intersection plane and are thus valid for any number of degrees of freedom N.)

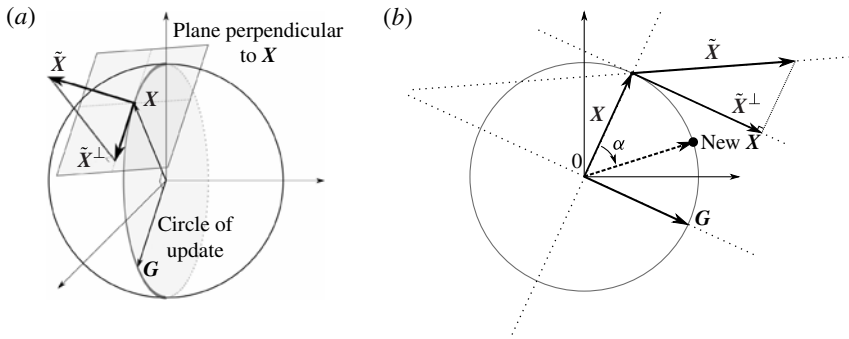

FIGURE 11. Sketch representing the procedure used to update the vector position from $\boldsymbol{X}_n$ to $\boldsymbol{X}_{n+1}$ for a problem of dimension $d = 3$. (*a*) Full sphere constraint: we project the gradient $\nabla_{\boldsymbol{X}} J = \tilde{\boldsymbol{X}}$ onto the plane perpendicular to $\boldsymbol{X}$. The corresponding scaled vector is denoted by $\boldsymbol{G}$ and is used to perform the update $\boldsymbol{X}_n \rightarrow \boldsymbol{X}_{n+1}$ via a one-dimensional rotation, as shown in (*b*).

Douglas *et al.* 1998) to satisfy the constraint on the initial state-vector norm. The gradient (3.8) can be expressed in discrete form as

$$\nabla_{\boldsymbol{X}} J = \tilde{\boldsymbol{X}}. \tag{A 8}$$

To enforce the constraint that the initial state vector has to fall on the hyper-sphere surface consistent with the normalization constraint (A 1), we project the true gradient (A 8) onto the hyperplane normal to the hyper-sphere surface. This is accomplished by simply removing the component of the gradient that is perpendicular to the normal of the hyperplane, i.e. the component of $\tilde{\boldsymbol{X}}$ along the vector $\boldsymbol{X}$. Mathematically, this procedure can be expressed as

$$\tilde{\boldsymbol{X}}^{\perp} = \tilde{\boldsymbol{X}} - \frac{\boldsymbol{X}^{\top}\tilde{\boldsymbol{X}}}{\boldsymbol{X}^{\top}\boldsymbol{X}}\boldsymbol{X}, \tag{A 9}$$

where the superscript $^{\perp}$ indicates that the adjoint variable (and gradient) $\tilde{\boldsymbol{X}}$ has been projected onto the hyper-plane perpendicular to $\boldsymbol{X}$. Next, we need to satisfy constraint (3.3). We therefore normalize the gradient and scale it by the radius of the hyper-sphere (i.e. the L_2-norm of $\boldsymbol{X}$, which is equal to C in our example)

$$\boldsymbol{G} = C\frac{\tilde{\boldsymbol{X}}}{\|\tilde{\boldsymbol{X}}\|_2}, \tag{A 10}$$

which is followed by a rotation of $\boldsymbol{X}$ on the circle defined by the intersection of the plane $(\boldsymbol{X}, \boldsymbol{G})$ with the hyper-sphere.

At step n of the algorithm, we update the state vector $\boldsymbol{X}_n$ to form $\boldsymbol{X}_{n+1}$ through

$$\boldsymbol{X}_{n+1} = \boldsymbol{X}_n \cos\alpha + \boldsymbol{G}_n \sin\alpha. \tag{A 11}$$

This procedure is illustrated in figure 11. It is important to observe that the value of α is no longer restricted, but can span the whole circle, implying a more well-posed and robust procedure to update the state vector $\boldsymbol{X}$ on a hyper-sphere surface.

It is desirable to combine this rotation technique with a more advanced gradient-descent algorithm such as the conjugate gradient method which utilizes a line search as well as a sequence of previously chosen directions in the optimization process to

converge to an optimum. Starting with the first descent direction (denoted $\boldsymbol{L}$) given by the true gradient of the cost functional

$$\boldsymbol{L}_0 = \boldsymbol{G}_0, \tag{A 12}$$

we obtain for all subsequent steps

$$\boldsymbol{L}_n = \boldsymbol{G}_n + \beta_n \boldsymbol{L}_{n-1}^{\perp}, \tag{A 13}$$

where $\boldsymbol{L}_{n-1}^{\perp}$ is the descent direction from the previous iteration, projected onto the hyperplane tangent to the hyper-sphere at the position $\boldsymbol{X}_n$ (the current location). The coefficients β_n are given by the Polak–Ribière formula (Polak 1971):

$$\beta_n = \frac{\boldsymbol{G}_n^{\top}(\boldsymbol{G}_n - \boldsymbol{G}_{n-1})}{\boldsymbol{G}_{n-1}^{\top}\boldsymbol{G}_{n-1}}. \tag{A 14}$$

The conjugate gradient update is then simply given by

$$\boldsymbol{X}_{n+1} = \boldsymbol{X}_n \cos\alpha + \boldsymbol{L}_n \sin\alpha. \tag{A 15}$$

We use this procedure, which yields a stable and robust optimization algorithm for all optimizations performed in this article. As mentioned previously, it also lends itself to optimization problems where the governing equations are nonlinear and, consequently, where the initial amplitude enters the procedure as an additional, user-defined parameter. This procedure can also be extended to normalization with respect to a p-norm by modifying the projection operator appropriately.

Appendix B. p-norm optimal perturbations at $Re = 250$

We conjecture in § 5 that the limit value of p for which the ∞-norm reaches saturation should be a function of the Reynolds number. This conjecture is based on the behaviour of two competing length scales: diffusion and localization. To validate this conjecture, we perform a series of computation at $Re = 250$. We fix the horizon time to the energy 1-norm optimal time which is $T_{opt} = 4.80$ at $Re = 250$, and then vary the order of the norm p. We focus the maximum search on the C solutions branch only for this demonstration. However, the W also appears at $Re = 250$. We present our results in figure 12.

We note that the p-norm converges towards the value of the ∞-norm, but that the latter value saturates on a plateau, confirming the observation at $Re = 4000$ that the ∞-norm can be optimized for relatively small values of p. This limit value of p for which the ∞-norm becomes optimal is $p_{lim} \simeq 5$. This observed value is smaller than in the case $Re = 4000$, which supports our conjectured dependence of p_{lim} on Reynolds number Re. Indeed, it seems that the value of p for which the ∞-norm can be optimized is an increasing function of the Reynolds number.

We plot in figure 13 the energy density distribution of the ∞-norm optimal perturbation for both the initial perturbation at time $t = 0$ and the final perturbation at the optimal time $t = T_{opt}$. The perturbation consists of a localized symmetric structure, very similar to the $Re = 4000$ solution plotted in figure 7(*c*). However, due to the relatively low Reynolds number, the perturbation is less localized than in the higher Reynolds number case. Although the localization is weaker due to viscous effects, the perturbation can still be described as a 'hotspot' of energy at the end of the integration time $t = T_{opt}$.

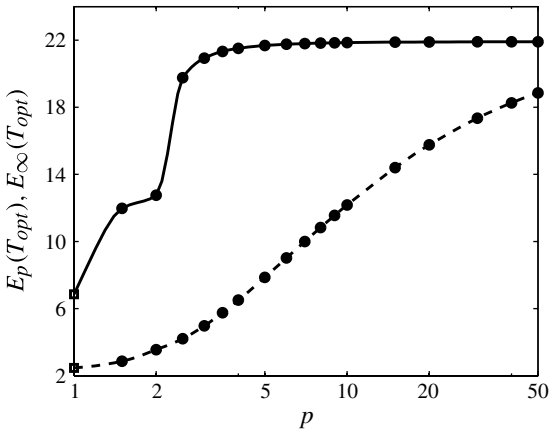

FIGURE 12. Energies $E_p(T_{opt})$ (dashed line) and $E_\infty(T_{opt})$ (solid line) as a function of p for $Re = 250$. All of the maxima presented are centre solutions, except for $p = 1$ where the maximum is the usual periodic perturbation. $E_\infty(T_{opt})$ reaches a plateau for increasing values of p. The limit value of p for which this saturated value is reached is roughly $p_{lim} \simeq 5$.

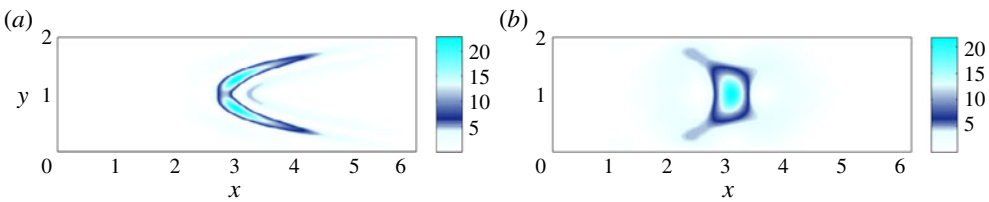

FIGURE 13. (*a*) Energy density plots of the centre solution at $Re = 250$ for $T = 4.80$. The left-hand side represent the initial perturbation at $t = 0$ and the right-hand side display the solution at the final time $t = T$. For clarity the localized solutions (at both $t = 0$ and $t = T$) have been placed in the centre of the domain. For this reason, the x-axis cannot be used to deduce the advection length.

Appendix C. Convergence properties of the solutions

Most of the results presented in this paper are the output of a gradient-based optimization procedure. When an optimization is performed, it is natural to check the value of the norm of the gradient at the corresponding optimal location in order to guarantee the robustness of the maximizer (or minimizer) found. In the case of a constrained optimization (detailed in appendix A), the projection of the gradient on the constraining surface must vanish. Indeed, theoretically, a solution $\boldsymbol{X}_{opt}$ is said to be optimal if we have

$$\|\nabla_{\boldsymbol{X}} J^{\perp}\|_2 = 0. \tag{C 1}$$

In practice, reaching exactly zero is not feasible, but the value of the gradient norm is a good indicator of the quality of convergence of the associated solution. We normalize this value by the norm of the true gradient, and thus define the convergence

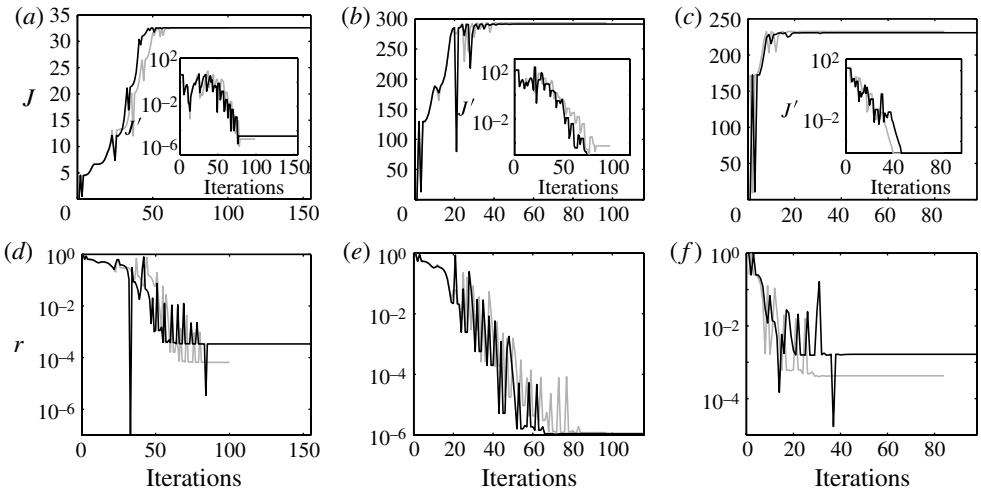

FIGURE 14. Convergence curves for the wall and centre solutions ($p = 50$), and for the periodic solutions ($p = 1$) at $Re = 4000$ for $T = T_{opt} = 10.4$. The first line of this plots array represents the cost functional J as a function of the iterations, respectively for the periodic (*a*), centre (*b*) and wall (*c*) solutions. The inset represents the cost functional residual J' as a function of the iterations. The second line represents the gradient residual r as a function of the iterations, respectively for the periodic (*d*), centre (*e*) and wall (*f*) solutions. One iteration corresponds to one function evaluation. The black curves are obtained for a resolution of $n_x = 600$ and $n_y = 200$ (corresponding to the results presented in the paper), while the grey curves corresponds to a larger resolution of $n_x = 800$ and $n_y = 300$.

residual (similarly to Rabin *et al.* (2012)) r as

$$r = \frac{\left\|\nabla_X J^{\perp}\right\|_2^{\,2}}{\left\|\nabla_X J\right\|_2^{\,2}} = 0. \tag{C 2}$$

Another significant quantity is J' defined at each iteration n as

$$J'_n = \frac{J_n - J_{n-1}}{J_n}. \tag{C 3}$$

This number measures by how much the cost functional is changing at each iteration. A low value of J' indicates algorithmic convergence, but not necessarily numerical convergence which is only really indicated by a low and reducing (with iteration) value of the gradient residual r.

We present in this appendix convergence curves for a few typical solutions found in this paper, and discuss their robustness in terms of actually being extrema of the cost functional J. We consider the value of the cost functional J, its residual J' and the gradient residual r, the latter being the most rigourous test of convergence.

We focus our attention on three typical results obtained for $T = T_{opt}$ at $Re = 4000$ for the periodic solution ($p = 1$ case), and both the centre and wall solutions ($p = 50$ case). The various convergence curves at different numerical resolutions are presented in figure 14.

As is clear on these plots, although the cost functional J appears to converge strongly towards a constant value with associated small values of the quantity J', none of the solutions found is approaching close to machine precision in terms of gradient

residual r. We believe there are at least two reasons for this imperfect convergence. First, this imperfect convergence can be at least partially attributed to the temporal and spatial discretization of the continuous incompressible Navier–Stokes equations (2.2). In addition, for large p values, the cost functional is less and less smooth because of the very strong localization of the adjoint fields (and, thus, of the gradient) which explains the difficulties we encounter in terminating the optimization process.

On the same plots (figure 14), the grey curves represent the convergence curves of the same solutions, for a higher spatial resolution. Interestingly, both the periodic and wall solutions have an improved final value of the gradient residual r (almost one order of magnitude lower), while the centre solution convergence has not improved at all. The 'optimal' cost functional value J varies slightly in each case by 0.1 % for the periodic solution, 0.7 % for the centre solution and 0.8 % for the wall solutions when the resolution is increased. A key distinguishing characteristic is that the periodic and wall solutions unlike the centre solution have strong values of the velocity field close to the wall during their time evolution. Therefore, we believe for our calculations that the interior of the flow is appropriately solved in the sense that the gradient residual we obtain ($r \sim 10^{-6}$) is the best we can expect with such a finite-difference discretization. Moreover, the solver used in this paper produces a spurious divergence boundary layer very close to the wall whose extent is of the order of the spatial discretization. By increasing the resolution, these spurious bands (top and bottom walls) of divergent velocity field are decreasing in size, and the flow at the walls is therefore solved more accurately, which explains the improvement of convergence with the increase of resolution for solutions with significant structure in the immediate vicinity of the walls.

Therefore, the solutions described in this paper cannot be considered unequivocally as being fully converged, although the condition $r \to 0$ is quite a strong condition for such problems. However, the lack of accuracy of our solver only becomes significant when very high accuracy is actually needed, namely when the gradient of the cost functional becomes small.

In order to check the validity of the results we present in this paper, we also performed the various optimizations corresponding to the data points of figure 8 with an independent rewritten two-dimensional incompressible Navier–Stokes solver which solves the incompressibility condition down to machine precision everywhere in the domain, and especially at the walls. The results varied by a small amount because of the increased accuracy (leading to finer convergence of the optimization), but the results of this completely independent set of calculations confirmed that the previously identified optimal solutions are (local or global) extrema of the cost functional. The independent set of calculations also confirmed the existence of the different periods of dominance of either the wall or the centre solutions as shown in figure 8, although the precise locations of the boundaries between the periods shifted slightly. Therefore, we are confident that the basins of attraction of each of the solutions presented in this article do indeed exist.

Hence, the frontiers between regions of dominance of each solution depicted in figure 8 have to be considered strictly as imperfectly converged, and therefore as indicative only. In reality, the values of the switching time (separating the various periods i, ii, iii and iv) as well as the precise values of the maximum of the energy displayed in these plots may well vary by a small amount, but the whole mechanistic picture remains unchanged.

REFERENCES

ARRATIA, C., CAULFIELD, C. P. & CHOMAZ, J.-M. 2013 Transient perturbation growth in time-dependent mixing layers. *J. Fluid Mech.* **717**, 90–133.

BLACKBURN, H. M., BARKLEY, D. & SHERWIN, S. J. 2008 Convective instability and transient growth in flow over a backward-facing step. *J. Fluid Mech.* **603**, 271–304.

BUTLER, K. M. & FARRELL, B. F. 1992 Three-dimensional optimal perturbations in viscous shear flow. *Phys. Fluids* A **4**, 1637–1650.

CHERUBINI, S., DE PALMA, P., ROBINET, J.-CH. & BOTTARO, A. 2010 Rapid path to transition via nonlinear localized optimal perturbations in a boundary-layer flow. *Phys. Rev.* E **82**, 066302.

CHORIN, A. J. 1968 Numerical solution of the Navier–Stokes equations. *Maths Comput.* **22** (104), 745–762.

DOUGLAS, S. C., AMARI, S. & KUNG, S.-Y. 1998 Gradient adaptation under unit-norm constraints. In *Proceedings of the Ninth IEEE SP Workshop on Statistical Signal and Array Processing*, pp. 144–147.

FARRELL, B. F. 1988 Optimal excitation of perturbations in viscous shear flow. *Phys. Fluids* **31**, 2093–2102.

FARRELL, B. F. & IOANNOU, P. J. 1996 Generalized stability theory. Part II. Nonautonomous operators. *J. Atmos. Sci.* **53**, 2041–2053.

FARRELL, B. F. & IOANNOU, P. J. 2003 Structural stability of turbulent jets. *J. Atmos. Sci.* **60**, 2101–2118.

FOURES, D. P. G., CAULFIELD, C. P. & SCHMID, P. J. 2012 Variational framework for flow optimization using seminorm constraints. *Phys. Rev.* E **86**, 026306.

GUSTAVSSON, L. H. 1991 Energy growth of three-dimensional disturbances in plane Poiseuille flow. *J. Fluid Mech.* **224**, 241–260.

LINDZEN, R. S. 1988 Instability of plane parallel shear flow (toward a mechanistic picture of how it works). *Pure Appl. Geophys.* **126** (1), 103–121.

MARQUET, O., SIPP, D., CHOMAZ, J.-M. & JACQUIN, L. 2008 Amplifier and resonator dynamics of a low-Reynolds-number recirculation bubble in a global framework. *J. Fluid Mech.* **605**, 429–443.

MONOKROUSOS, A., BOTTARO, A., BRANDT, L., DI VITA, A. & HENNINGSON, D. S. 2011 Nonequilibrium thermodynamics and the optimal path to turbulence in shear flows. *Phys. Rev. Lett.* **106**, 134502.

ORR, W. M'F. 1907 The stability or instability of the steady motions of a perfect liquid and of a viscous liquid. Part I. A perfect liquid. Part II. A viscous liquid. *Proc. R. Irish Acad. Section* A: Math. Phys. Sci. **27**, 9–138.

ORSZAG, S. A. 1971 Accurate solution of the Orr–Sommerfeld stability equation. *J. Fluid Mech.* **50**, 689–703.

POLAK, E. 1971 *Computational Methods in Optimization: A Unified Approach*, vol. 77. Academic.

PRINGLE, C. C. T. & KERSWELL, R. R. 2010 Using nonlinear transient growth to construct the minimal seed for shear flow turbulence. *Phys. Rev. Lett.* **105**, 154502.

RABIN, S. M. E., CAULFIELD, C. P. & KERSWELL, R. R. 2012 Triggering turbulence efficiently in plane Couette flow. *J. Fluid Mech.* **712**, 244–272.

REDDY, S. C. & HENNINGSON, D. S. 1993 Energy growth in viscous channel flows. *J. Fluid Mech.* **252**, 209–238.

REDDY, S. C., SCHMID, P. J. & HENNINGSON, D. S. 1993 Pseudospectra of the Orr–Sommerfeld operator. *SIAM J. Appl. Maths.* **53** (1), 15–47.

SCHMID, P. J. 2007 Nonmodal stability theory. *Annu. Rev. Fluid Mech.* **39**, 129–162.

SCHMID, P. J. & HENNINGSON, D. S. 2001 *Stability and Transition in Shear Flows*. Springer.

SELLE, A., FEDKIW, R., KIM, B., LIU, Y. & ROSSIGNAC, J. 2008 An unconditionally stable MacCormack method. *J. Sci. Comput.* **35** (2), 350–371.

THOMAS, J. W. 1995 *Numerical Partial Differential Equations: Finite Difference Methods*, vol. 1. Springer.

TREFETHEN, L. N., TREFETHEN, A. E., REDDY, S. C. & DRISCOLL, T. A. 1993 Hydrodynamic stability without eigenvalues. *Science* **261**, 578–584.