

Nonlinear optimals and their role in sustaining turbulence in channel flow

Dario Simon Klingenberg^{1,2}  and Rich R. Kerswell¹ 

¹DAMTP, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, UK

²ZARM, University of Bremen, Am Fallturm 2, 28359 Bremen, Germany

Corresponding author: Dario Simon Klingenberg, dsk34@cam.ac.uk

(Received 14 February 2025; revised 17 May 2025; accepted 21 June 2025)

We investigate the energy transfer from the mean profile to velocity fluctuations in channel flow by calculating nonlinear optimal disturbances, i.e. the initial condition of a given finite energy that achieves the highest possible energy growth during a given fixed time horizon. It is found that for a large range of time horizons and initial disturbance energies, the nonlinear optimal exhibits streak spacing and amplitude consistent with direct numerical simulation (DNS) at least at $Re_\tau = 180$, which suggests that they isolate the relevant physical mechanisms that sustain turbulence. Moreover, the time horizon necessary for a nonlinear disturbance to outperform a linear optimal is consistent with previous DNS-based estimates using eddy turnover time, which offers a new perspective on how some turbulent time scales are determined.

Key words: nonlinear instability, computational methods, channel flow

1. Introduction

When trying to understand how turbulence is sustained in shear flow, an essential part of the picture is the transfer of energy from the mean profile to the fluctuations. To shed light on this, a popular approach has been to linearise the Navier–Stokes equations around the mean profile and to study the growth of linear optimal disturbances on top of this mean. Some aspects of the dynamics observed in fully developed turbulence can be explained in this way, for example, streak formation along with the correct streak spacing (Butler & Farrell 1993), as well as streak breakdown (Schoppa & Hussain 2002; Hoepffner, Brandt & Henningson 2005; Cassinelli *et al.* 2017).

This energy transfer is usually understood as a self-sustaining process (Kim, Kline & Reynolds 1971; Jiménez & Moin 1991; Hamilton, Kim & Waleffe 1995; Waleffe

1997; Jiménez & Pinelli 1999; Farrell & Ioannou 2012). The current prevailing picture is that of a two-stage linear process (Schoppa & Hussain 2002; Lozano-Durán *et al.* 2021; Markeviciute & Kerswell 2024). In a primary linear process, streaks form and grow due to transient growth. If one subsequently changes the decomposition so as to include these streaks in the base profile, a secondary linear process can describe how secondary disturbances grow on top of this streaky field (again fuelled by transient growth), ultimately causing the streaks to break down and restarting the cycle (Hamilton *et al.* 1995; Schoppa & Hussain 2002; Hoepffner *et al.* 2005; Lozano-Durán *et al.* 2021).

Even though breaking this process down into linear subprocesses offers great practical advantages associated with the simplicity of linear theories in general, it also introduces a number of questions that are difficult to answer in a self-contained manner. For example, it remains unclear how to choose the correct streak amplitude for the secondary linear process without appealing to direct numerical simulation (DNS) results (Markeviciute & Kerswell 2024).

A natural generalisation, which also helps address these challenges, is to consider the full nonlinear optimisation problem. Thanks to the presence of the nonlinear term, the primary and secondary linear processes can be coupled together, thus eliminating the need for formulating a two-stage process. Of course, it is unclear *a priori* if the nonlinear optimal actually involves such a two-stage process, or if completely different dynamics is more effective. More generally, the central question to be addressed in this work is which aspects of the nonlinear optimal resemble the dynamics of real turbulence.

The motivation for studying this problem is twofold. First, a better understanding of how turbulence is sustained could inform modelling and flow control tasks. Second, Malkus (1956) formulated the seminal idea that the mean profile actually observed in practice is in some way optimal in its ability to sustain turbulence (in concrete terms, he asserted that the mean profile should be marginally stable in a statistical sense). Using the methods available at the time, he focused on long-term (linear) stability rather than short-term transient energy growth as the main mechanism responsible for maintaining turbulence and ultimately failed (e.g. Reynolds & Tiederman 1967) to formulate a suitable definition of optimality. However, the general notion is still appealing today and we hope that the present study could contribute to a discussion about reviving this idea.

Attempts to understand turbulence by linearising around a suitable base flow have been around for a long time (Malkus 1956; Reynolds & Tiederman 1967). However, it was not until Farrell (1988) that transient growth was popularised as an important linear mechanism for describing energy transfer from the mean profile to the fluctuations. In a seminal work, Schoppa & Hussain (2002) furthermore showed that when linearising around a streaky base flow, it is again transient growth that can best describe energy transfer from the streaky mean field to the disturbances, thus giving rise to the two-step picture described previously. Whereas the primary linear process is fairly well-understood, much debate has surrounded the question of which linear process can best describe the secondary process (Waleffe 1997; Reddy *et al.* 1998; Schoppa & Hussain 2002; Hoepffner *et al.* 2005; Farrell & Ioannou 2012). Recently, however, in an extensive cause-and-effect study, Lozano-Durán *et al.* (2021) were able to confirm Schoppa & Hussain's (2002) assertion that transient growth is more important than other proposed mechanisms such as linear instability or parametric instability, at least at the relatively low Reynolds number of $Re_\tau = 180$.

By comparison, the nonlinear optimisation problem has received much less attention. For one, nonlinear optimisation relying on solving the full three-dimensional Navier–Stokes equations only became computationally feasible during the 2010s (Pringle & Kerswell 2010; Cherubini *et al.* 2010, 2011; Monokrousos *et al.* 2011; Pringle, Willis &

Kerswell 2012; Farano *et al.* 2016). However, even today, the computational effort remains high, leading to relatively few works applying nonlinear optimisation to the turbulent setting (Cherubini & De Palma 2013; Farano *et al.* 2015, 2017). Notably, however, these works employ a different formulation than used here, rendering these works unsuitable to answer the questions with which we are presently concerned. This is discussed in greater detail in § 2.

Including the nonlinear term in the optimisation problem can lead to fundamentally different optimals. Whereas linear optimals are global in structure and only include a single streamwise and spanwise wave number mode, nonlinear optimals combine multiple wave numbers and, under some circumstances, can be highly localised in space. Moreover, they can achieve much higher growth than linear optimals, often by coupling together multiple linear mechanisms (Pringle *et al.* 2012; Kerswell 2018). An important point to note is that unlike in the transition problem, the disturbances relevant to the turbulent problem are not necessarily small, which makes it more difficult to justify dropping the nonlinear term. Even though, as discussed by Lozano-Durán *et al.* (2021), there are still good reasons to consider linear theories, it does imply that the evolution of the linear optimal is in practice influenced by the nonlinear term, which, at high energies, may actually limit the gains it can produce. As a result, nonlinear optimals are likely more practically relevant, while at the same time also offering deeper insights into the relevant dynamics of turbulence.

This paper is organised as follows: in § 2, we discuss the problem formulation on a theoretical level and § 3 describes the numerical implementation. Results are shown and discussed in § 4, and § 5 concludes.

2. Problem formulation

In the present work, we consider the same problem as Lozano-Durán *et al.* (2021), i.e. incompressible Newtonian channel flow at friction Reynolds number $Re_\tau = u_\tau h/\nu = 180$, where u_τ is the wall friction velocity, h the channel half-height and ν the kinematic viscosity. The problem is governed by the Navier–Stokes equations

$$\mathcal{N} := \frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} + \nabla p + \mathbf{f} - \frac{1}{Re_\tau} \nabla^2 \mathbf{u} = 0, \quad (2.1)$$

$$\mathcal{M} := \nabla \cdot \mathbf{u} = 0, \quad (2.2)$$

where $\mathbf{u}$ and p are the dimensionless flow velocity and pressure, respectively, and $\mathbf{f}$ is an external volume forcing term, the reasons for including which will become apparent later. The coordinate system is oriented such that x points in the streamwise direction, y in the wall-normal direction and z in the spanwise direction. We decompose velocity and pressure into a base and a disturbance part,

$$\mathbf{u} = \mathbf{U} + \tilde{\mathbf{u}}, \quad p = P + \tilde{p}, \quad (2.3)$$

where capital letters denote the base and tilde denotes the disturbance fields. In this study, the base is obtained by temporal averaging an *a priori* DNS run. Since it is calculated in advance, $\mathbf{U}$ can be thought of as a fixed parameter entering the problem. In the following, we set $-\nabla P = \hat{\mathbf{x}}$, with $\hat{\mathbf{x}}$ denoting the unit vector in the x -direction. Note that the streamwise-spanwise (but not temporal) average of $\tilde{\mathbf{u}}$, which we denote $\langle \tilde{\mathbf{u}} \rangle_{xz}$, is allowed to be non-zero. Here, angular brackets denote averaging across the directions in the index, so, for example,

$$\langle \tilde{\mathbf{u}} \rangle_{txz} = \frac{1}{TL_x L_z} \int_0^T \int_0^{L_x} \int_0^{L_z} \tilde{\mathbf{u}} \, dz \, dx \, dt. \quad (2.4)$$

This ensures consistency because $\tilde{\mathbf{u}}$ is able to change the streamwise-spanwise averaged velocity $\langle \mathbf{u} \rangle_{xz}$ as the disturbance grows on the short time scales in which we are presently interested, even though $\mathbf{U}$ remains fixed. However, in the limit of very long time horizons, the temporal average $\langle \tilde{\mathbf{u}} \rangle_t$ vanishes. Inserting (2.3) into (2.1) and (2.2) yields

$$\begin{aligned} \mathcal{N} := & \frac{\partial \mathbf{U}}{\partial t} + \mathbf{U} \cdot \nabla \mathbf{U} - \hat{\mathbf{x}} + \mathbf{f} - \frac{1}{Re_\tau} \nabla^2 \mathbf{U} \\ & + \frac{\partial \tilde{\mathbf{u}}}{\partial t} + \mathbf{U} \cdot \nabla \tilde{\mathbf{u}} + \tilde{\mathbf{u}} \cdot \nabla \mathbf{U} + \tilde{\mathbf{u}} \cdot \nabla \tilde{\mathbf{u}} + \nabla \tilde{p} - \frac{1}{Re_\tau} \nabla^2 \tilde{\mathbf{u}} = 0, \end{aligned} \quad (2.5)$$

$$\mathcal{M} := \nabla \cdot \mathbf{U} + \nabla \cdot \tilde{\mathbf{u}} = 0. \quad (2.6)$$

We now set

$$\mathbf{f} = \hat{\mathbf{x}} + \frac{1}{Re_\tau} \nabla^2 \mathbf{U}. \quad (2.7)$$

Physically, this forcing term represents the background turbulence fluctuations maintaining $\mathbf{U}$. It is important to note that these background fluctuations are separate from the disturbances $\tilde{\mathbf{u}}$ in our present model. This point is further discussed later. Together with $\partial \mathbf{U} / \partial t = 0$ due to statistical stationarity and $\mathbf{U} \cdot \nabla \mathbf{U} = \nabla \cdot \mathbf{U} = 0$ in channel flow, this yields an equation system for the disturbances which reads

$$\tilde{\mathcal{N}} := \frac{\partial \tilde{\mathbf{u}}}{\partial t} + \mathbf{U} \cdot \nabla \tilde{\mathbf{u}} + \tilde{\mathbf{u}} \cdot \nabla \mathbf{U} + \tilde{\mathbf{u}} \cdot \nabla \tilde{\mathbf{u}} + \nabla \tilde{p} - \frac{1}{Re_\tau} \nabla^2 \tilde{\mathbf{u}} = 0, \quad (2.8)$$

$$\tilde{\mathcal{M}} := \nabla \cdot \tilde{\mathbf{u}} = 0. \quad (2.9)$$

For a given base profile $\mathbf{U}$, time horizon T and initial condition $\tilde{\mathbf{u}}(\mathbf{x}, 0)$, (2.9) and (2.8) can be integrated forward in time to yield $\tilde{\mathbf{u}}(\mathbf{x}, T)$ and, in particular, the final disturbance energy $e_T := 1/(2V) \int_V |\tilde{\mathbf{u}}(\mathbf{x}, T)|^2 dV$. Normalising e_T by the initial energy e_0 yields

the energy gain $G := e_T/e_0$. By further fixing the initial energy e_0 , one can ask which initial condition $\tilde{\mathbf{u}}(\mathbf{x}, 0)$ experiences the highest energy gain G by the end of the time horizon T . This question naturally leads to an optimisation problem which has been studied extensively in previous works (Pringle & Kerswell 2010; Cherubini *et al.* 2010, 2011; Monokrousos *et al.* 2011; Cherubini & De Palma 2013; Farano *et al.* 2017), and which can be solved using adjoint looping. Details of the solution algorithm are given in § 3. In this study, only (2.8) and (2.9) are actually solved numerically. The base velocity is imposed to be the time-averaged mean profile obtained from a previous DNS (Markeviciute & Kerswell 2024).

As mentioned previously, $\langle \tilde{\mathbf{u}} \rangle_{xz}$ can be non-zero and, as a consequence, $\langle \mathbf{u} \rangle_{xz} = \mathbf{U} + \langle \tilde{\mathbf{u}} \rangle_{xz} \neq \mathbf{U}$. Thus, it might appear like the initial disturbance $\tilde{\mathbf{u}}(\mathbf{x}, 0)$ is free to adjust $\langle \mathbf{u}(\mathbf{x}, 0) \rangle_{xz}$ to any arbitrary profile, in particular one that might be more suitable for generating disturbance energy growth than $\mathbf{U}$. However, the initial energies considered here are $\lesssim 0.02\%$ of the base energy (except for one run at 0.04% in § 4.3) and so this is not an issue.

The formulation given by (2.8) and (2.9) is intended to be the simplest possible nonlinear model that could be used to investigate disturbance growth in turbulent flows. We note that in the literature, an alternative formulation exist (e.g. Farano *et al.* 2017), which differs from our work in that no background turbulence fluctuations are presumed. This, in turn, implies a feedback mechanism. Since in that model, the disturbances are responsible for maintaining the base flow $\mathbf{U}$, they are constrained to be of an order of magnitude that would be necessary for maintaining the mean flow in a DNS. For the questions to be

investigated in the present work, this constrained behaviour of the disturbances would not make sense, as we are interested in how the disturbances grow on top of the base profile, drawing energy only from it.

Some authors considering the linear optimisation problem (Reynolds & Hussain 1972; Pujals *et al.* 2009; Cossu & Hwang 2017) have advocated for including an eddy viscosity in the momentum equation for the disturbances, i.e. (2.8). This introduces an *ad hoc* assumption that the eddy viscosity felt by the mean profile is also the same as would be felt by the fluctuations. To keep the model as simple as possible, we decide not to include an eddy viscosity in this study, though its effect could be an interesting topic for future work.

To formulate the optimisation algorithm, we follow Kerswell (2018) and consider the Lagrangian

$$\begin{aligned} \mathcal{L} = & \iiint_V \frac{1}{2} |\tilde{\mathbf{u}}(\mathbf{x}, T)|^2 dV + \lambda \left[\iiint_V \frac{1}{2} |\tilde{\mathbf{u}}(\mathbf{x}, 0)|^2 dV - e_0 \right] \\ & + \int_0^T \iiint_V \tilde{\mathbf{v}}(\mathbf{x}, t) \cdot \tilde{\mathbf{N}} dV dt + \int_0^T \iiint_V \tilde{q}(\mathbf{x}, t) \tilde{\mathcal{M}} dV dt, \end{aligned} \quad (2.10)$$

where λ , $\tilde{\mathbf{v}}$ and $\tilde{q}$ are Lagrange multipliers. Setting the first variation of $\mathcal{L}$ to zero leads, after some algebra, to the further conditions for the Lagrange multiplier fields that

$$\tilde{\mathbf{u}}(\mathbf{x}, T) + \tilde{\mathbf{v}}(\mathbf{x}, T) = 0, \quad (2.11)$$

$$\frac{\partial \tilde{\mathbf{v}}}{\partial t} + (\mathbf{U} + \tilde{\mathbf{u}}) \cdot \nabla \tilde{\mathbf{v}} - \tilde{\mathbf{v}} \cdot (\nabla(\mathbf{U} + \tilde{\mathbf{u}}))^T + \nabla \tilde{q} + \frac{1}{Re_\tau} \nabla^2 \tilde{\mathbf{v}} = 0, \quad (2.12)$$

$$\nabla \cdot \tilde{\mathbf{v}} = 0, \quad (2.13)$$

provided that the boundary conditions for $\tilde{\mathbf{v}}$ are the same as for $\tilde{\mathbf{u}}$. Furthermore,

$$\frac{\delta \mathcal{L}}{\delta \tilde{\mathbf{u}}(\mathbf{x}, 0)} = \lambda \tilde{\mathbf{u}}(\mathbf{x}, 0) - \tilde{\mathbf{v}}(\mathbf{x}, 0) = 0. \quad (2.14)$$

Note that (2.12) and (2.13) form the adjoint Navier–Stokes equations, and the Lagrange multipliers $\tilde{\mathbf{v}}$ and $\tilde{q}$ can be interpreted as adjoint velocity and pressure (divided by the density), respectively. This set of equations naturally gives rise to a looping algorithm to obtain the optimal initial condition $\tilde{\mathbf{u}}_{opt}(\mathbf{x}, 0)$ that generates the highest energy gain G at the end of the time horizon T . First, make an initial guess for $\tilde{\mathbf{u}}_{opt}(\mathbf{x}, 0)$, which we call $\tilde{\mathbf{u}}(\mathbf{x}, 0)$, and evolve it forward in time using the Navier–Stokes equations (2.9) and (2.8) until time $t = T$. Second, obtain the initial condition for the adjoint system $\tilde{\mathbf{v}}(\mathbf{x}, T)$ using (2.11), and integrate (2.12) and (2.13) backward in time from $t = T$ to $t = 0$. Third, now that the final value of the adjoint variable $\tilde{\mathbf{v}}(\mathbf{x}, 0)$ is known, calculate $\delta \mathcal{L} / \delta \tilde{\mathbf{u}}(\mathbf{x}, 0)$ using (2.14) and use this gradient to improve the current guess $\tilde{\mathbf{u}}(\mathbf{x}, 0)$. Repeat this loop until some convergence criterion is reached. More technical details of this algorithm are discussed in § 3.

3. Computational method

A special-purpose solver has been implemented in the programming language Python to integrate (2.9) and (2.8) forward in time, and to integrate the adjoint system (2.13) and (2.12) backward in time. This solver uses the library JAX (Bradbury *et al.*

2018), which makes the code fully differentiable (although this is not used much in the present work) and provides GPU acceleration, thus delivering a crucial speed-up and making the investigation feasible. The source code of the solver is available at <https://github.com/dakling/jax-spectral-dns>.

Following Kim, Moin & Moser (1987) and Lluesma-Rodríguez *et al.* (2021), a velocity-vorticity formulation is employed. Spatial directions are discretised using a pseudospectral method, with Fourier modes in periodic directions and Chebyshev modes in the wall-normal direction. A third-order Runge–Kutta method derived by Spalart, Moser & Rogers (1991) is used for time discretisation. De-aliasing using the 3/2-rule (Orszag & Patterson 1972) is done in the Fourier directions (but not in Chebyshev directions for performance reasons), though this is probably unnecessary at the employed resolution. The nonlinear terms are formulated in skew-symmetric form for increased numerical stability (Gibson 2014).

As is further discussed in Appendix A.2, we choose a computational domain size of $\pi \times 2 \times \pi$, which is bigger than the minimal channel unit ($0.6\pi \times 2 \times 0.3\pi$ at the considered Reynolds number) but still small enough to keep the numerical effort reasonably low (Jiménez & Moin 1991). This domain size should be enough to observe some localisation of the nonlinear optimal. Following Kim *et al.* (1987), the domain is discretised using a resolution of at least $48 \times 129 \times 80$, which corresponds to $\Delta x^+ \approx 11.8$, $\Delta y_{\max}^+ \approx 4.2$ and $\Delta z^+ \approx 7.1$, though some calculations were done using a higher resolution to ensure that it does not make a difference. To make full use of JAX's GPU acceleration, the time step size should be set *a priori*, as JAX is better able to optimise loops of fixed size compared with loops of variable size. Taking into account the base velocity and the expected magnitude of the disturbances (which is either roughly known from the previous looping iteration or can be estimated depending on the nature of the initial guess), the time step is chosen such that the Courant–Friedrichs–Lewy number (CFL) is kept below 0.6 with a safety factor of 0.9. Since the time stepping algorithm has been reported to be stable up to $\text{CFL} = 0.7$ (Lluesma-Rodríguez *et al.* 2021), this choice is conservative, and indeed no issues with time stepping instabilities occurred with this setting. Thanks to the GPU acceleration provided by JAX, the duration of a single DNS forward run is of the order of a few minutes (compared with a few hours when running on a single CPU), rendering the looping approach computationally feasible.

More important than runtime, a major limiting factor for the present looping algorithm is memory. Note that (2.12) contains $\tilde{\mathbf{u}}$, i.e. the full velocity space–time history of the forward run. Depending on the details of the set-up, in our present investigation, the size of this variable is of the order of magnitude between 10 and 100 GB, which, given that this is just a single variable, would cause out-of-memory issues on GPUs for all but the smallest cases considered here if it needed to be saved entirely. However, memory can be traded for runtime efficiency by employing a strategy known as checkpointing (Berggren 1998; Griewank & Walther 2000; Hinze, Walther & Sternberg 2005). The idea is that rather than storing the entire history of the forward run, only some snapshots, or checkpoints, of the forward calculation are saved, serving as initial conditions for recomputing intervals as needed during the adjoint calculation. The strategy employed in this work is fairly simple: snapshots are saved at equally spaced time intervals. The number of time intervals is chosen such that the number of snapshots is roughly equal to the number of time steps between two snapshots. This reduces the memory requirements from N to $2\sqrt{N}$, where N characterises the size of the space–time history of the velocity field, while increasing the runtime of the adjoint calculation by a factor of two.

The effectiveness and efficiency of the looping approach crucially depends on the employed gradient-based optimisation algorithm. In the present work, we follow Cherubini *et al.* (2011) and use a conjugate gradient method, although we determine the step size adaptively using a two-way backtracking line search (Truong & Nguyen 2020). This line search is similar to a regular backtracking line search, but the step size from the last iteration is used as an initial guess for the new step size, thus reducing the number of necessary function evaluations (which are quite expensive as they involve the forward run of the DNS). If the step size is found to be sufficiently small, it is also periodically tested if the step size can be increased, which is not necessary in classical backtracking line search. Note that the formula for the gradient (2.14) contains the Lagrange multiplier λ , which ensures that the initial energy constraint is fulfilled. Its appropriate value depends on the step size and direction taken, so the gradient itself depends on the parameters of the conjugate descent and must be recomputed appropriately. For a given set of descent parameters, λ is then obtained using a standard Newton method. To ensure that numerical issues with determining λ can never lead to a violation of the initial energy constraint, the resulting initial condition is also appropriately rescaled after each optimisation iteration. The algorithm is documented in detail in [Appendix A.3](#).

Another critical question concerns constructing an appropriate initial guess. Inconveniently, for many of the parameters considered here, it appears that the linear optimal remains a local optimum, so that it (or slightly perturbed versions of it) are unsuitable as an initial guess. Instead, in this work, we rely on introducing randomness by running the optimisation loop with a deliberately large step size, effectively forcing the algorithm to sample the space of flow fields in a chaotic manner. Another strategy is to use a random snapshot from a DNS as an initial condition. Generally, it should be noted that even though nonlinear optimals outperform linear ones even for relatively short time horizons and low initial energies, to find them initially, longer time horizons and higher initial energies are usually necessary. Once an initial condition that gives rise to a nonlinear optimal is found, this resulting nonlinear optimal can in turn also be used as an initial condition for calculations with different parameters. Most of the parameter combinations – including all on which we focus the following discussion – were also rerun from different initial conditions to ensure that the same optimal was reached, although, since this is a nonlinear optimisation problem, there is no rigorous way to ensure that a global optimum has indeed been found. Nevertheless, the conclusions made in the following are quite robust to changes in the computational set-up, including the domain size, as is further discussed in [Appendix A.2](#).

Since the code written for this study supports automatic differentiation, it would also be possible to obtain the necessary gradient information without solving the adjoint system. However, it has been found that using the classical adjoint-based approach performs better than automatic differentiation, both in terms of memory and runtime. In particular, the ability to more easily control the trade-off between memory usage and runtime through checkpointing makes the adjoint approach preferable since we are currently already close to the limits of available GPU-memory on an NVIDIA A100 card with 40 GB memory. Moreover, the gradients obtained from the adjoint method are fairly smooth, whereas in gradients obtained from automatic differentiation, the discretisation error manifests itself in the form of noise, so that some processing of these gradients would likely be necessary. Automatic differentiation does have the advantage of requiring less development time and thus being more readily applied to new problems, so that this feature will likely be valuable in future studies.

4. Results

As discussed in § 2, the main parameters are the time horizon T and the initial disturbance energy e_0 . Unlike in the transition problem, it is not useful to look at very large time horizons, as the time scale on which disturbances can grow is constrained by large-scale disruptive turbulent events. This argument has been used by Butler & Farrell (1993) to determine the time scale for linear optimal growth, and the general idea also applies to the nonlinear case. From previous investigations (Lozano-Durán *et al.* 2021; Markeviciute & Kerswell 2024), we can estimate that the primary and the secondary linear processes both operate on a time scale of approximately $t_p = t_s = 0.35h/u_\tau$ and that bursting events are roughly $t_{burst} = 4h/u_\tau$ apart. This motivates an initial focus on the time horizon $T = 0.7h/u_\tau = 2t_p$, as this would be enough to see both the primary and the secondary linear process. This is addressed in § 4.1. To get a deeper understanding of the role of this parameter, we additionally study times $t_p, 1.5t_p, 3t_p, 4t_p, 6t_p$ and $8t_p$ in § 4.2.

For any T , as $e_0 \rightarrow 0$, the linear optimisation problem is recovered. Consequently, for small e_0 , a quasilinear optimal, which closely resembles the linear optimal, is obtained. The linear optimal's distinguishing property is that it consists of only a single wave number mode. This is because interactions between different wave number modes would require nonlinear effects, in the absence of which the optimisation algorithm simply picks the wave number mode with the highest energy gain for the given time horizon. The quasilinear optimal is the weakly nonlinear extension of the linear optimal. Since it is finite amplitude, nonlinear coupling between modes exists but is too weak to make much of a difference compared with the linear optimal. As e_0 increases, there comes a point where a structurally completely different nonlinear optimal arises. For large e_0 , the final state of the disturbances may be well in the turbulent regime, which, in the current formulation, usually causes the optimisation algorithm to fail, as the final state dependency on the initial condition is too sensitive and chaotic for the optimisation problem to remain well conditioned. The precise determination of this boundary is difficult and not very important to the present study, but values up to $e_0/E_0 = 10^{-4}$ were found to converge without issues even for the longest time horizons considered here, where E_0 denotes the energy of the base flow. Naturally, we are most interested in nonlinear optimals whose initial energy is below this value.

4.1. Optimal disturbances at $T = 0.7h/u_\tau$

In this subsection, we consider the time horizon $T = 0.7h/u_\tau$. The optimal gain that can be achieved as a function of initial energy e_0 is plotted in figure 1.

It is apparent that at $e_0/E_0 = 2 \times 10^{-5}$, a first nonlinear optimal emerges. As the initial energy increases further, this optimal becomes increasingly more efficient, until the highest gain among all e_0 is found at $e_0/E_0 = 7.2 \times 10^{-5}$. As e_0/E_0 is increased beyond this value, saturation seems to set in and the achieved gains slowly diminish until around $e_0/E_0 = 2 \times 10^{-4}$, where the optimisation problem starts becoming ill-conditioned. Interestingly, the difference between the highest nonlinear optimal gain (34.82) and the linear gain (28.48) is not very high at this T . However, the long-term evolution (obtained by solving (2.8) and (2.9) with an optimal disturbance as the initial condition, but for $t_{final} \gg T$) reveals a major difference: whereas the gain $G(t) = e(t)/e_0 = 1/(2V) \iint_V |\tilde{\mathbf{u}}(\mathbf{x}, t)|^2 dV/e_0$ of the linear optimal tends to reach its peak gain at or close to

$t = T$, the nonlinear optimals keep growing significantly, often almost by another order of magnitude (see figure 2). This continued growth seems to be a consequence of localisation, i.e. it is mainly due to the disturbance further spreading out in space rather than it growing

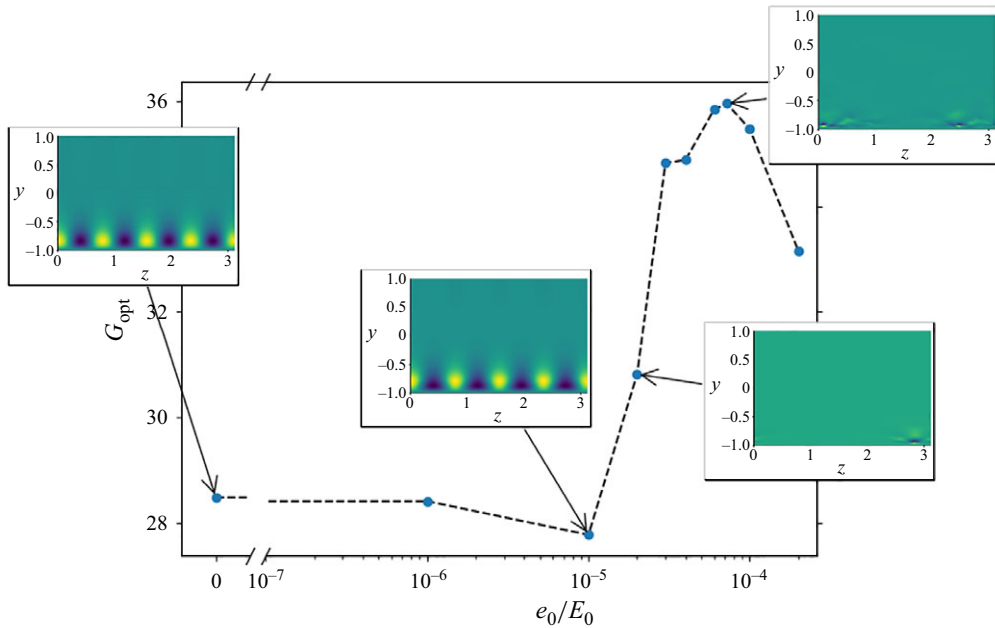


Figure 1. Optimal gain over initial energy normalised by the energy of the base flow E_0 . Note the truncated vertical axis. Plots of the initial condition of the streamwise disturbance velocity for some of the optimals are also included. For more details on these and their time evolution, please refer to [figures 3](#) and [5](#).

in amplitude. The fact that a decay is observed for long times ensures that the long-term temporal average $\langle \tilde{u} \rangle_t$ vanishes.

Next, we study the structure of the observed optimals to better assess their relevance in fully developed turbulence, starting with the linear optimal. Its time evolution is shown in [figure 3](#). As is expected for a linear optimal, only a single wave number mode – in this case, $[k_x, k_z] = [0, 8]$ – is present, with the energy of all other modes combined only making up approximately 10^{-10} times the total energy. This wave number mode grows due to the well-documented transient growth mechanism based on non-normal interaction of modes. Here, $k_x = 2\pi/\lambda_x$, where λ_x stands for the x -wavelength, and k_z is defined analogously. Note that the streak spacing does not match what one would observe in real turbulence. This is expected, however, because, following the argument of [Butler & Farrell \(1993\)](#), the considered time horizon is unrealistically long (by a factor of two) for the primary linear process alone.

Moving to finite but small values for e_0/E_0 , at 10^{-6} and up to 10^{-5} , a quasilinear optimal is found. Its velocity time evolution (not shown) closely resembles the linear optimal ([figure 3](#)), though the effects of the nonlinear term are visible in the combined energy of the non-dominant wave modes, which grows from approximately 3×10^{-4} to 2.6×10^{-3} times the total energy. This becomes apparent in the spectral decomposition of the streamwise velocity shown in [figure 4](#).

Optimal gain is achieved by the nonlinear optimal with initial energy $e_0/E_0 = 7.2 \times 10^{-5}$, and its time evolution is shown in [figures 5](#) and [6](#). It is apparent that the shape of the initial condition is completely different from the linear optimal. Most notably, no streaks are present, and the disturbance is concentrated, or localised, in specific parts of the domain. This localisation is most clear in the spanwise direction, as there is no full localisation in the streamwise direction (see [figure 6](#)). Instead, in the streamwise direction, there seems to be a length scale between the two distinct peaks of the disturbance. Indeed,

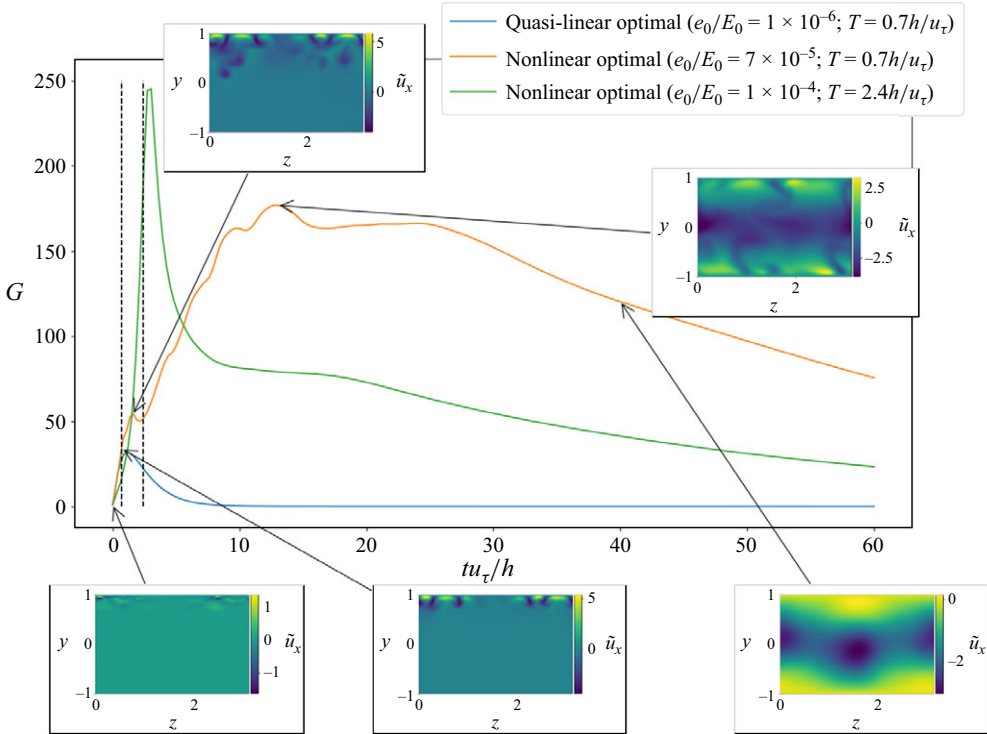


Figure 2. Long-term evolution ($t_{final}u_\tau/h = 60$) of the energy gain G of the quasilinear optimal ($e_0/E_0 = 10^{-6}$, $Tu_\tau/h = 0.7$, blue line), the nonlinear optimal ($e_0/E_0 = 7.2 \times 10^{-5}$, $Tu_\tau/h = 0.7$, orange line), as well as the nonlinear optimal ($e_0/E_0 = 1 \times 10^{-4}$, $Tu_\tau/h = 2.8$, green line). Vertical dashed lines indicate $tu_\tau/h = 0.7$ and $tu_\tau/h = 2.8$, respectively. Snapshots of the streamwise disturbance velocity are shown for the $(e_0/E_0 = 7.2 \times 10^{-5}, Tu_\tau/h = 0.7)$ -optimal (orange line).

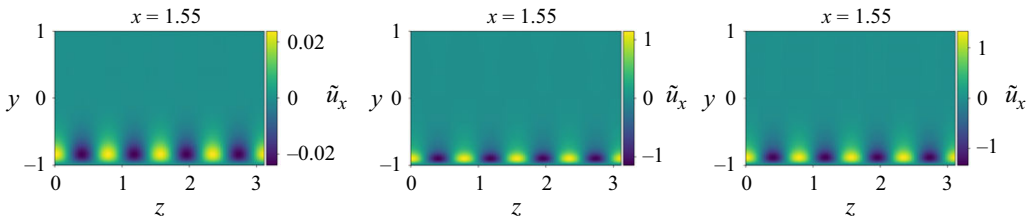


Figure 3. Streamwise velocity evolution ($tu_\tau/h = 0, 0.35, 0.7$) of the linear optimal for $Tu_\tau/h = 0.7$.

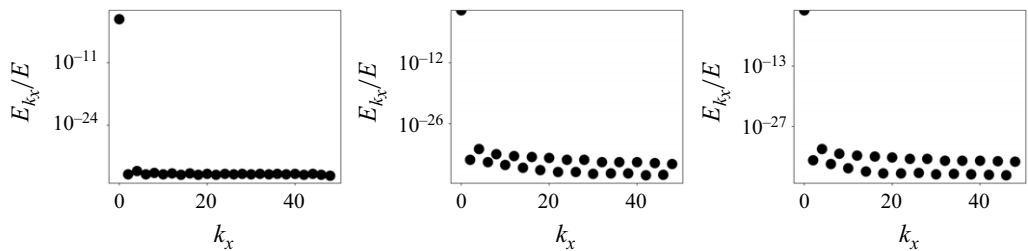


Figure 4. Streamwise spectrum of the streamwise velocity at times $tu_\tau/h = 0, 0.35, 0.7$ of the quasilinear optimal ($e_0/E_0 = 10^{-6}$) for $Tu_\tau/h = 0.7$. Essentially all of the energy is contained in the $k_x = 0$ mode.

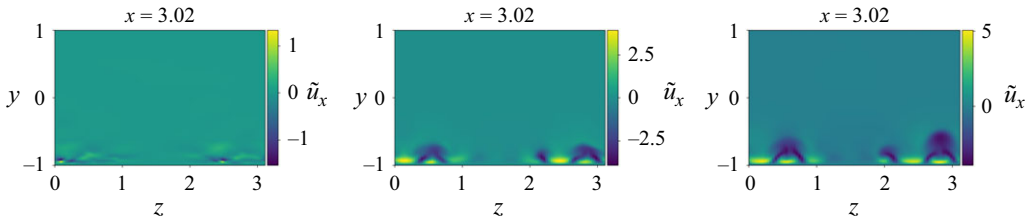


Figure 5. Streamwise velocity evolution ($tu_\tau/h = 0, 0.35, 0.7$) of the nonlinear optimal ($e_0/E_0 = 7.2 \times 10^{-5}$) for $Tu_\tau/h = 0.7$ plotted in the y - z plane.

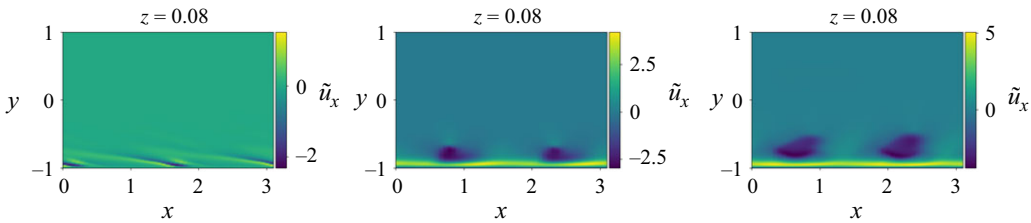


Figure 6. Streamwise velocity evolution ($tu_\tau/h = 0, 0.35, 0.7$) of the nonlinear optimal ($e_0/E_0 = 7.2 \times 10^{-5}$) for $Tu_\tau/h = 0.7$ plotted in the x - y plane.

when moving to a box twice as large in the streamwise direction (i.e. $2\pi \times 2 \times \pi$), four such peaks can be identified (see [Appendix A.2](#)). It is unclear how this length scale is determined, but we may speculate that it is the most efficient spacing to quickly generate continuous streaks. This may also be related to the size of the minimal channel unit, which is roughly half of the current domain in the streamwise direction. Therefore, it is possible that the current structure could be thought of as two minimal channel optimals stitched together in the streamwise direction. Localisation is also visible in the wall-normal direction, with the initial disturbance concentrated close to the wall where the base shear is high. The fact that the entire disturbance is concentrated on one side and not symmetrical can also be explained with localisation. This is in contrast to the linear optimal, for which both the symmetrical and the one-sided versions yield the same gain, though we only show the one-sided version here. Localisation is a well-known phenomenon for nonlinear optimals (Kerswell 2018) and can be rationalised by the optimisation algorithm trying to accommodate the initial energy constraint: instead of distributing the disturbance energy evenly, it may make more sense to have high-amplitude disturbances strategically placed in some small areas and then rely on nonlinear mechanisms to spread out these disturbances over time.

The early part of the evolution is characterised by an absence of streaks (see [figure 9a](#)), and instead we observe tilted flow structures oriented such that they are amplified by an Orr-type mechanism (this is most clearly observed in [figure 8](#)). Following this initial phase, these disturbances quickly organise into streaks (see [figures 6b, 7b and 9b](#)). Note that the streaks that have formed at this point do exhibit correct streak spacing and amplitude (see [figure 5b](#) and the discussion in § 4.2). Unlike the streaks of the linear optimal, the low-speed streaks do not remain straight as they evolve, but meander and wrap around the high-speed streaks, as is revealed by [figure 7](#). Subsequently, these streaks grow due to the lift-up mechanism, and although no clear streak breakdown can be observed, it is clear from [figure 10](#) that second-order disturbances grow on top of the streaks towards the end of the considered time window (from time *(iii)* onwards). Thus, it can be concluded that the

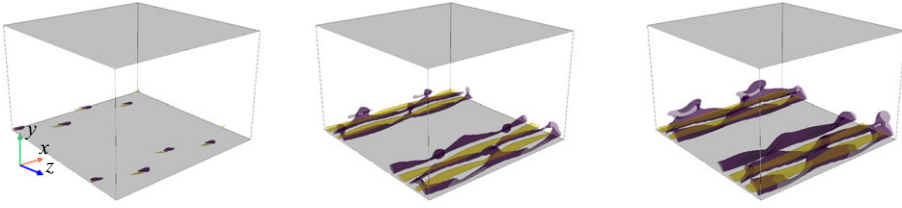


Figure 7. Streamwise velocity evolution ($tu_\tau/h = 0, 0.35, 0.7$) of the nonlinear optimal ($e_0/E_0 = 7.2 \times 10^{-5}$) for $Tu_\tau/h = 0.7$. Yellow isocontours indicate 60 % of the maximum value and blue isocontours 60 % of the minimum value.

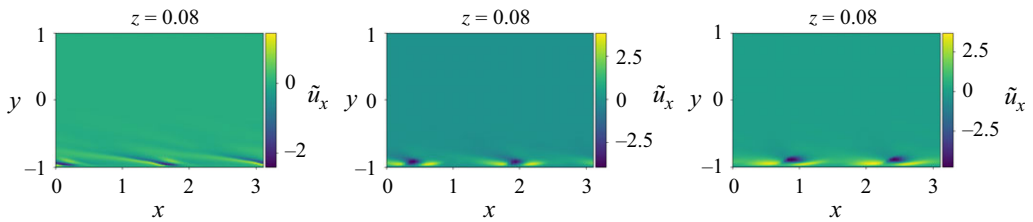


Figure 8. Early part of the streamwise velocity evolution ($tu_\tau/h = 0, 0.35, 0.7$) of the nonlinear optimal ($e_0/E_0 = 7.2 \times 10^{-5}$) for $Tu_\tau/h = 0.7$ plotted in the x - y plane.

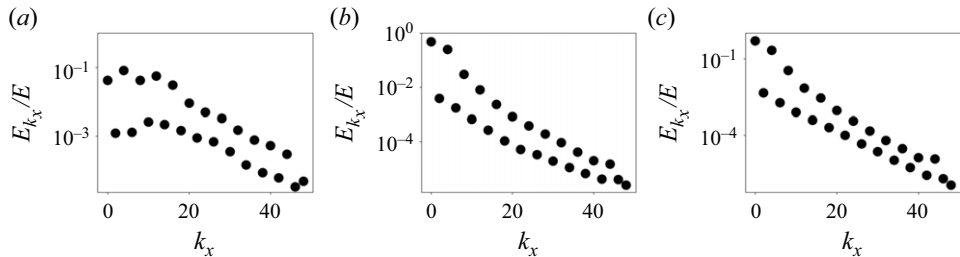


Figure 9. Streamwise spectrum of the streamwise velocity at times $tu_\tau/h = 0, 0.35, 0.7$ of the nonlinear optimal ($e_0/E_0 = 7 \times 10^{-5}$) for $Tu_\tau/h = 0.7$. Compared with [figure 4](#), the spectral dynamics are much more complicated, with streaks ($k_x = 0$) not being dominant initially, but then containing most of the energy later.

primary and, although not as pronounced, the secondary linear processes are both visible in the nonlinear optimum, though the way that they are coupled together by the nonlinear term means that there is significant temporal overlap. This serves as a reminder that characterising them as neatly separated subprocesses constitutes a major simplification. Moreover, the streak growth itself is clearly not exclusively fuelled by transient growth (although this likely plays a significant role), as can be most clearly observed by the temporal evolution of the high-frequency modes ($k_x \geq 6$) between the times (i) and (ii) shown in [figure 10](#). After an initial growth up to the time marked by vertical line (i), these modes start decaying, indicating that they transfer energy to the streaks.

Interestingly, the main features remain remarkably stable throughout the nonlinear regime. Comparing the optimal at $e_0/E_0 = 3 \times 10^{-5}$ with that at $e_0/E_0 = 10^{-4}$ ([figure 11](#)) reveals that only the localisation in the spanwise direction diminishes for higher energies (as one would expect), but apart from this, the dynamics, including time scales and streak spacing, is comparable. In particular, the minimum and maximum values of the streamwise velocity disturbances are very similar, so it can be concluded that the additional initial

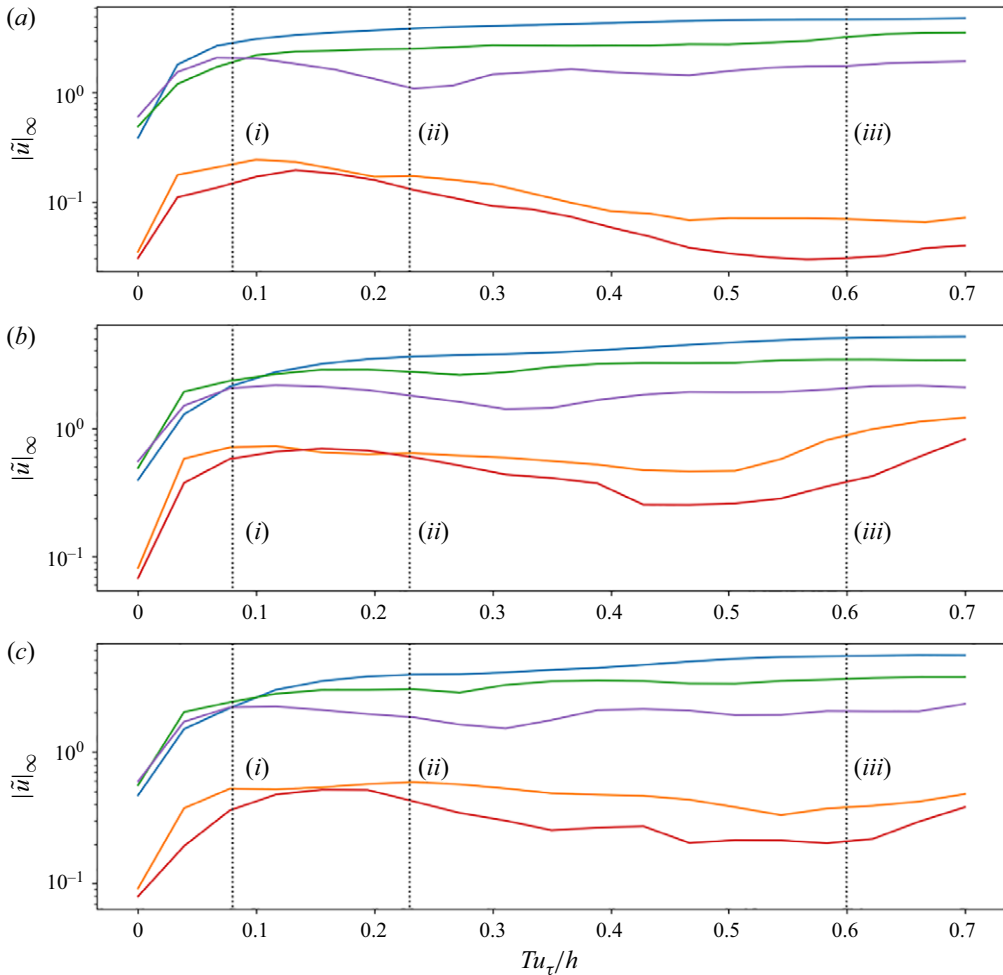


Figure 10. Streamwise velocity amplitudes of x -wave number modes over time for different localised nonlinear optimals, (a) $e_0/E_0 = 3 \times 10^{-5}$, (b) $e_0/E_0 = 7.2 \times 10^{-5}$ and (c) $e_0/E_0 = 10^{-4}$, all for $Tu_\tau/h = 0.7$. Shown are wave numbers $k_x = 0$, i.e. the streaks (blue); $k_x = 2$ (orange); $k_x = 4$ (green); $k_x = 6$ (red) and $k_x = 8$ (purple). Despite the different initial disturbance energies, all exhibit a rapid early growth up to time (I), then a decay of the $k_x \geq 6$ -modes until around time (II) and a final growth of these higher-order modes after time (III).

energy available for the $e_0/E_0 = 10^{-4}$ case only serves to reduce localisation, rather than amplify the magnitude of the disturbance. The similarity of the underlying dynamics is also confirmed by the wave number plot shown in figure 10 and further discussed in § 4.2.

4.2. Exploring time-energy parameter space

It is clear that not any arbitrary parameter combination of T and e_0 will give rise to optimals that distil the relevant dynamics of fully developed turbulence. Although rough estimates for the time and energy scales can be made based on scaling arguments as discussed in § 4.1, it is important to understand how the shape and dynamics of the optimals change in (T, e_0) -space. In particular, we would like to investigate how sensitive the dynamics is to changes in T and e_0 . Figure 12 shows the different regimes that can be identified between $T = 0.35h/u_\tau$ and $T = 2.8h/u_\tau$, with the initial energy chosen small enough to result in an end state that is not too chaotic for the optimisation to converge.

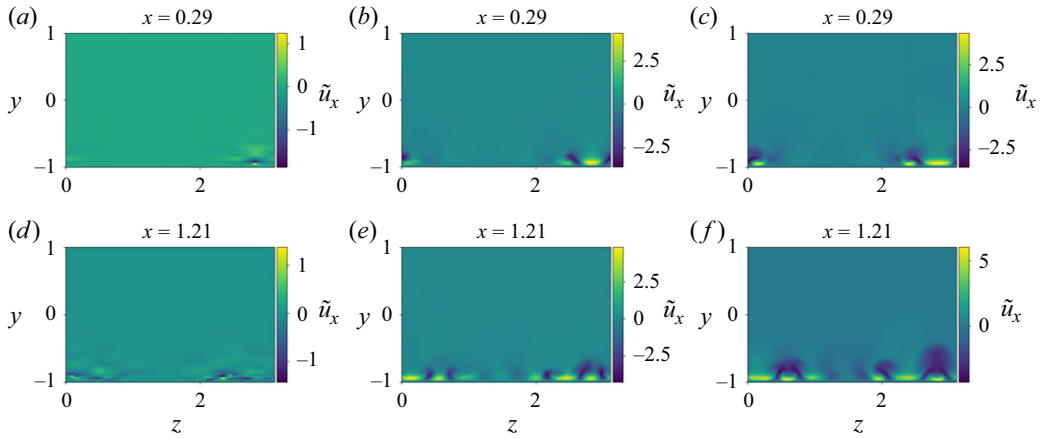


Figure 11. Comparison of nonlinear optimals and their evolution at different initial disturbance energies. (a–c) $e_0/E_0 = 3 \times 10^{-5}$, (d–f) $e_0/E_0 = 10^{-4}$. Panels (a,d), (b,e) and (c,f) correspond to initial ($tu_\tau/h = 0$), intermediate ($tu_\tau/h = 0.35$) and final ($tu_\tau/h = 0.7$) streamwise velocity field, respectively.

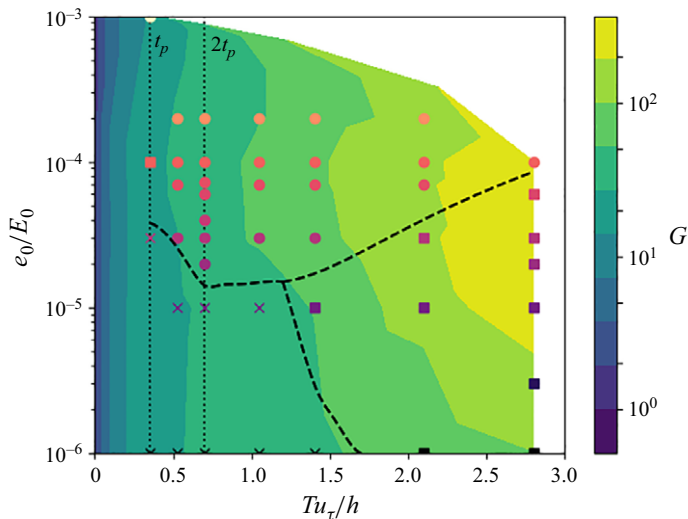


Figure 12. T – e_0 parameter space. Crosses indicate the quasilinear regime, squares the nonlinear non-localised regime and circles the nonlinear localised regime. The dashed lines were added manually to aid visual distinction between the regimes. Furthermore, each dot or cross is also colour-coded according to its e_0 , which is redundant in this plot, but allows establishing connections to [figures 16](#), [17\(a\)](#) and [17\(b\)](#).

The gain achieved by each optimal is also shown, with values between the points obtained by interpolation. For small e_0 , the quasilinear regime is found, which we here identify by the most energetic wave number mode of the initial optimal disturbance containing most (more than 90 %, though the exact value does not matter much as long as it is high) of the total energy. The boundary to the nonlinear regime has not been precisely resolved for all T (unlike in the transition problem, this area is not of primary interest here). It is interesting to note, however, that at $T \leq t_p$, the quasilinear optimal found at low e_0 produces higher gains than optimals found at higher e_0 . This trend is reversed for larger T , which implies

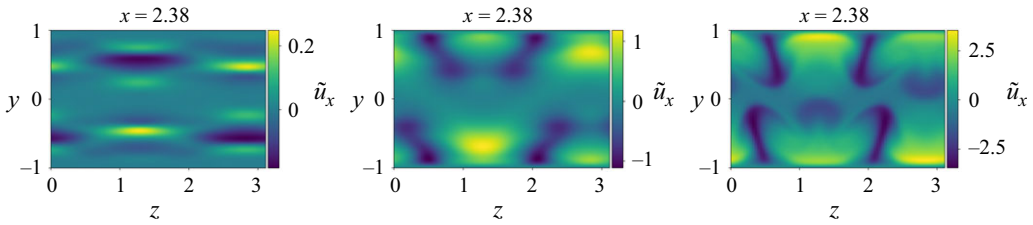


Figure 13. Streamwise velocity evolution ($tu_\tau/h = 0, 1.4, 2.8$) of the optimal for $Tu_\tau/h = 2.8$ and $e_0/E_0 = 3 \times 10^{-5}$.

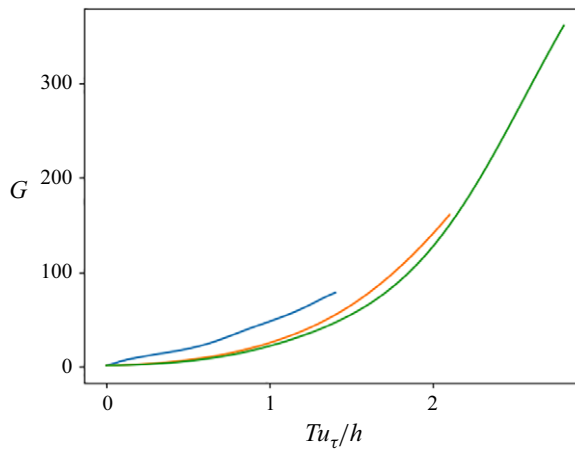


Figure 14. Energy over time of the optimal for the nonlinear localised optimal at $Tu_\tau/h = 1.4$ (blue) compared with the nonlinear non-localised optimals at $Tu_\tau/h = 2.1$ (orange) and $Tu_\tau/h = 2.8$ (green), all at $e_0/E_0 = 3 \times 10^{-5}$. It is apparent that much of the energy growth of the non-localised optimals happens late, which explains why they are not observed for short time horizons.

that the primary growth time scale t_p can also be thought of as the largest time horizon for which the (quasi-) linear optimal outperforms the nonlinear one.

At large time horizons but small initial energies, a completely new nonlinear optimal structure emerges (figure 13). Interestingly, this structure is not localised but global, and, as can be seen in figure 14, much of the energy growth happens in the final part of the calculation, which explains why this structure is not optimal when considering shorter time scales. We distinguish this regime from the nonlinear localised regime by checking that the lowest quarter of wave number modes of the initial optimal disturbance contain almost all ($>99.999\%$) of its energy. However, as is apparent visually and will be discussed in more detail later, the dynamics of how these optimals evolve bears no resemblance to fully developed turbulence, which confirms the expectation discussed at the beginning of § 4 that in a real flow, optimals have to evolve on much shorter time scales. Thus, we can conclude that time scales between approximately $T = 0.7h/u_\tau$ and $T = 2.1h/u_\tau$, and initial energy roughly between $e_0/E_0 = 3 \times 10^{-5}$ and $e_0/E_0 = 10^{-4}$ give rise to nonlinear optimals whose dynamics shares distinctive features with fully developed turbulence.

Qualitative visual inspection hints at a remarkable uniformity of all cases in the nonlinear localised regime. For example, comparing the nonlinear optimal at $T = 0.7h/u_\tau$ and $e_0/E_0 = 7.2 \times 10^{-5}$ (figure 5) with that found at $T = 2.1h/u_\tau$ and $e_0/E_0 = 10^{-4}$ (figure 15), even though the localisation is a bit diminished in the latter one due to the

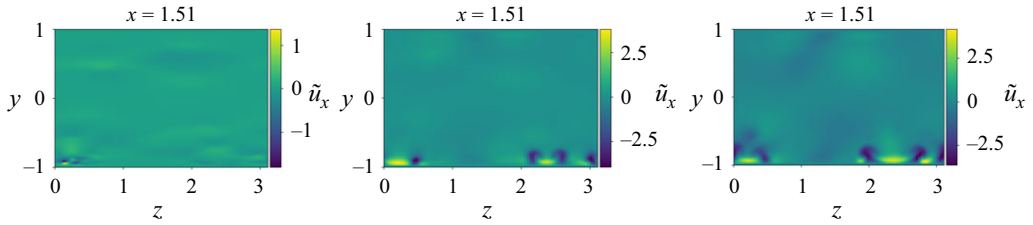


Figure 15. Early part of the streamwise velocity evolution ($tu_\tau/h = 0, 0.35, 0.7$ for comparison with figure 5) of the nonlinear optimal ($e_0/E_0 = 7.2 \times 10^{-5}$) for $Tu_\tau/h = 2.1$ plotted in the y - z plane.

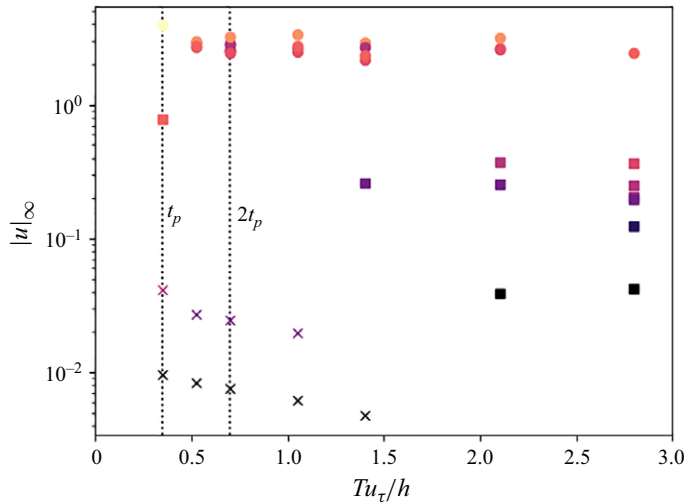


Figure 16. Infinity norm for each case in figure 12 with colours indicating e_0 (see figure 12). Crosses indicate the quasilinear regime, squares the nonlinear non-localised regime and circles the nonlinear localised regime. Despite different initial disturbance energies, all optimals of the nonlinear localised regime exhibit infinity norms in a very narrow range, which is not true for the optimals of the other regimes.

less restrictive initial energy, the general dynamics as well as length and time scales are very similar.

In the following, we pursue a more quantitative analysis to confirm the two observations made previously, i.e. that (i) the optimals in the nonlinear localised regime show very similar dynamics to those seen in real turbulence, and that (ii) this is true for all optimals in the nonlinear localised regime, even across the wide range of T and e_0 spanned by this regime. Plotting the infinity norm of the initial disturbance velocity $|\tilde{\mathbf{u}}(\mathbf{x}, 0)|_\infty$ (figure 16) shows a clear clustering of the different regimes. One might expect $|\tilde{\mathbf{u}}(\mathbf{x}, 0)|_\infty$ to mainly depend on the initial energy e_0 , because it determines the general magnitude of the velocity disturbance. However, especially the cases that fall into the nonlinear localised regime exhibit a remarkable uniformity in their respective $|\tilde{\mathbf{u}}(\mathbf{x}, 0)|_\infty$. This also confirms the qualitative observation in figure 11 that in the nonlinear localised regime, additional initial energy does not lead to higher initial disturbance amplitudes, but mainly to reduced localisation. Furthermore, it is reassuring that while the e_0/E_0 -values are much lower than typical disturbance energies in real turbulent flow, their $|\tilde{\mathbf{u}}(\mathbf{x}, 0)|_\infty$ -values are of the order of 10–20 % of the mean centreline velocity, which is a much more plausible value.

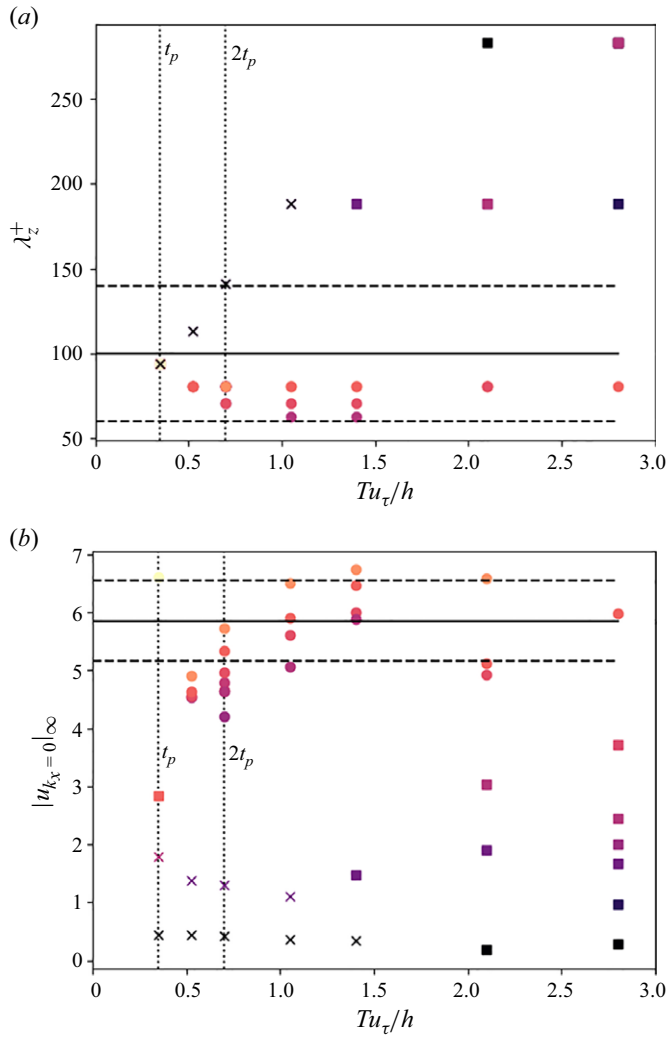


Figure 17. Comparison of streak length scales with DNS. Only the cases of the nonlinear localised regime agree well with DNS, both in terms of (a) streak spacing and (b) streak amplitude.

This uniformity found in the nonlinear localised regime also manifests itself in the streak spacing λ_z^+ , which is shown in figure 17(a). Here, the streak spacing is obtained automatically by computing the dominant ($k_x = 0, k_z$)-mode of the x -velocity at $t = 0.3u_\tau/h$ and then inferring the streak spacing from the k_z -value. This streak spacing detection works very robustly for the quasilinear optimals, but for the nonlinear optimals, the initial state is localised and does not contain any discernable streaks of which to speak. Towards the end of some of the runs, the streaks break down or diminish in importance. As a result, the automatic streak spacing detection is unreliable for early and late times of these runs. However, in a fairly long intermediate phase from approximately $0.1h/u_\tau$ to $0.6h/u_\tau$, when streaks stand out distinctly, the method works well, which motivates the choice of $t = 0.3h/u_\tau$ as the reference time. The result is that all runs that belong to the group of nonlinear localised optimals show streak spacing values inside the expected range, which is only true for some but not all the runs in the other

groups. In particular, it is apparent that linear optimals require a much more precise choice of the time horizon T if they are to produce the correct streak spacing, and that the optimals belonging to the nonlinear non-localised regime exhibit unrealistically high streak spacing. Further evidence for the uniformity and realism of the nonlinear localised optimals is provided in [figure 17\(b\)](#), which shows the infinity norm of the streaks, i.e. the $k_x = 0$ part of the x -velocity fields at $t = 0.8T$ for each case. This relatively late time is chosen here to allow the streaks enough time to not only form but also grow. Note that we consider the infinity norm rather than, say, the energy norm here because the former is less sensitive to localisation, thus allowing for a better comparison. Unlike streak spacing, this metric is not widely discussed in the literature, so that values for the mean and standard deviation of the infinity norm of streaks were obtained from 2000 DNS snapshots. Again, only the optimals belonging to the nonlinear localised regime produce streaks with a realistic infinity norm, though it is apparent that for the cases with relatively short time horizons, the streaks are still a bit small as they have not had the time to fully establish themselves yet. One might object that the comparison is somewhat unfair, because the cases in the quasilinear regime simply lack the initial energy to generate realistic streak amplitudes; however, in the nonlinear non-localised regime, even when there is sufficient initial energy, the streaks that form are not as pronounced and, thus, do not seem as crucial to the overall evolution dynamics.

Another interesting point is that the general shape of these optimals closely resembles that of optimal perturbations found in the transition problem. In the parameter study by Farano *et al.* (2016), both the linear regime and the nonlinear localised regime, which they refer to as highly nonlinear regime, seem to have counterparts that can be observed in the transition problem as well, although the time horizons considered in that study are much longer and the initial energies smaller.

These results are reassuring and, at the same time, highlight the usefulness of considering nonlinear optimals: whereas linear optimals require a very specific time horizon T for their predicted steak spacing to agree with fully developed turbulence, nonlinear optimals are much more robust in this regard, as every optimal in the nonlinear localised regime, which spans a considerable range of T and e_0 , yields dynamics consistent with fully developed turbulence, despite the relatively simple equation system employed here.

4.3. Short-time high-energy optimals

Having expanded the parameter space to very long times, we now also address the question of what happens for very short time horizons. Here, we choose $T = 0.1u_\tau/h$. To compensate for this short time horizon, we select a relatively large initial energy of $e_0/E_0 = 4 \times 10^{-4}$. The evolution of this optimal, which grows by approximately an order of magnitude in the given time horizon, is shown in [figure 18](#).

Interestingly, there is no localisation in x or z , and although the disturbance is concentrated in the y -direction to where the base shear is high, it appears on both walls, again indicating that localisation is not important for these parameters. This is partly expected due to the relatively high initial energy, but the more important point is probably the short time horizon. As can be inferred from the tilting of the disturbances visible in [figure 18\(d–f\)](#), only the Orr mechanism is important here, as there is no time for streaks to develop and make use of the lift-up mechanism. This is expected and consistent with the time scales observed in the temporal evolution of the nonlinear optimals discussed in §§ 4.1 and 4.2. The fact that there is only a single linear mechanism at play due to the short time horizon thus helps explain why localisation is not observed.

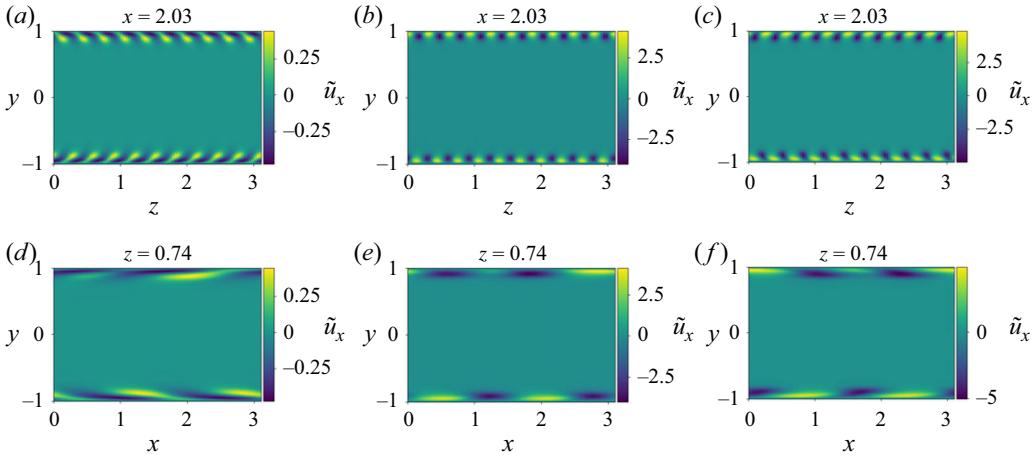


Figure 18. Streamwise velocity evolution ($Tu_\tau/h = 0, 0.05, 0.1$) of the nonlinear optimal ($e_0/E_0 = 4 \times 10^{-4}$) for $Tu_\tau/h = 0.1$ plotted (a–c) in the x – z plane and (d–f) in the x – y plane.

5. Conclusion

In this work, we have investigated the mechanisms by which energy is transferred from the mean to fluctuations in turbulent flow by computing the nonlinear optimals for various initial energies and time horizons. The results show that for a large range of intermediate time horizons, given initial energies large enough to trigger nonlinear effects, localised nonlinear optimals with very similar dynamics emerge. Important aspects of these dynamics, such as the streak spacing and amplitudes, are in good agreement with fully developed turbulence. Interestingly, the lower end of this range of time horizons is consistent with previous estimates of the time scales on which primary and secondary linear processes operate.

These results confirm that the evolution of nonlinear optimals reflects crucial aspects of turbulent dynamics, despite the simplicity of our model. At the same time, the optimisation serves to remove any non-essential noise not contributing to disturbance growth, which would be difficult to filter out in other methods such as DNS. This makes nonlinear optimals an attractive concept for isolating the important dynamics, allowing one to study turbulence in a ‘clean’ environment. At the same time, the present analysis provides upper bounds for the level of energy transfer to the turbulent fluctuations, which could serve as a foundation for future work aimed at better understanding under which circumstances turbulence is sustained. When compared with computationally less expensive linear methods, two major advantages of this approach are (i) the conceptual simplicity – in particular, there is no need to prescribe a streak amplitude as is the case for methods coupling linear processes ‘manually’ (e.g. Ciola *et al.* 2024; Markeviciute & Kerswell 2024) – and (ii) the robustness with which correct values for important characteristics of the dynamics, namely streak spacing and amplitude, are found with respect to parameter variations.

Future work could be aimed at combining the method of nonlinear optimisation with numerical experiments, in which equations are deliberately altered to test the effect. In the past, such numerical experiments have been done using DNS; however, this has the limitation that one relies on DNS snapshots of realistic turbulence as initial conditions to make predictions about an artificial setting. This may make conclusions less reliable compared with using nonlinear optimisation, which are fully self-contained.

Furthermore, the role of the base profile should be investigated. Understanding how different base profile affect disturbance growth could help answer the question of why a particular base profile (i.e. the mean) is selected in reality instead of some different profile. This line of reasoning could ultimately even provide an avenue towards predicting the mean profile in shear flows at moderate Reynolds number if a corrected version of Malkus's (1956) theory could be formulated.

Acknowledgements. We thank Sergio Hoyas and Patrick Keuchel for fruitful discussions about the solver implementation.

Funding. The work leading to this publication was supported by the PRIME programme of the German Academic Exchange Service (DAAD) with funds from the German Ministry of Education and Research (BMBF).

Declaration of interests. The authors report no conflict of interest.

Data availability statement. The data that support the findings of this study are openly available in Pangaea at <https://doi.pangaea.de/10.1594/PANGAEA.983358>.

Appendix A.

A.1. Solver validation

The solver presented in § 3 was validated by reproducing the growth rates of single modes (see figure 19) and by testing the transient growth of linear optimals (see figure 20). Both the single modes and the linear optimals were computed by the solver, which contains an implementation of the relevant linear stability problem in the spirit of Reddy & Henningson (1993) for this specific purpose. As can be seen in the two figures, the agreement is excellent.

A.2. Influence of the domain size

To justify the choice of the fairly small domain used in this study, we compare it with the more commonly used domain size of $2\pi \times 2 \times \pi$, for which a resolution of $96 \times 129 \times 80$ is used. Figure 21 shows the evolution of the optimal at $Tu_\tau/h = 0.7$ and $e_0/E_0 = 3 \times 10^{-5}$ on this larger domain. As can be seen by comparison with figure 11(a,d), the dynamics is essentially indistinguishable. The gain of 33.63 is also very close to the small

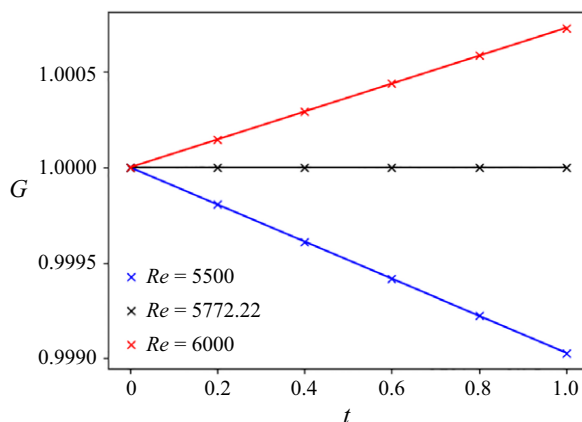


Figure 19. Growth rates of the most unstable mode at $Re = 5500, 5772.22, 6000$, with $\alpha = 1.02056$; $\beta = 0$: analytic values (lines) compared with numerical results (crosses).

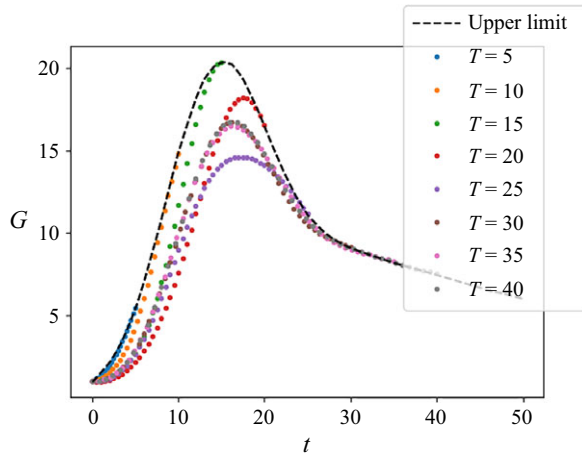


Figure 20. Growth rates of the linear optimal at $Re = 3000$ with $\alpha = 1$, $\beta = 0$ for different values of the time horizon T , compared with the optimal growth plotted over time (dashed line; from Reddy & Henningson 1993).

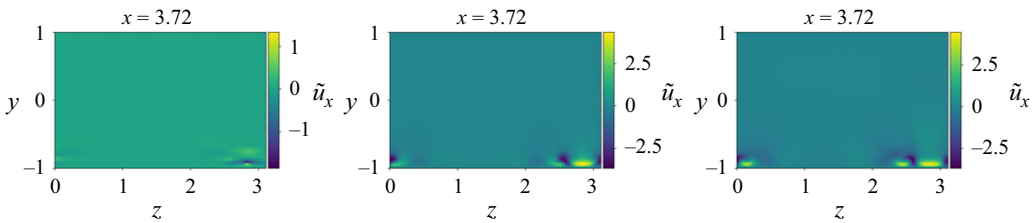


Figure 21. Streamwise velocity evolution ($tu_\tau/h = 0, 0.35, 0.7$) of the optimal for $Tu_\tau/h = 0.7$ and $e_0/E_0 = 3 \times 10^{-5}$ in a larger channel.

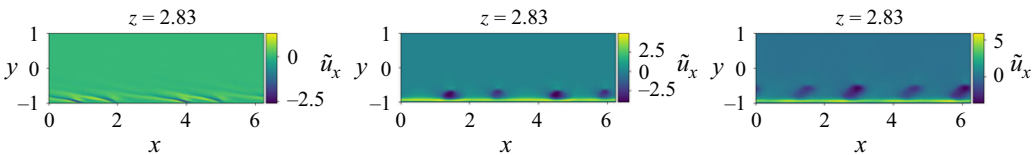


Figure 22. Streamwise velocity evolution ($tu_\tau/h = 0, 0.35, 0.7$) of the optimal for $Tu_\tau/h = 0.7$ and $e_0/E_0 = 3 \times 10^{-5}$ in a larger channel plotted in the x - y -plane.

channel value of 34.82. Interestingly, as is revealed by comparing figure 22 with figure 6 (the slightly higher e_0 has no significant effect on the dynamics), four instead of two clusters are present in the initial condition, and the streamwise length scale of the low-speed streaks is also the same as four instead of two low-speed regions can be seen in this plane.

A similar comparison was also done for the very different time horizon of $T = 2.8$, again at $e_0/E_0 = 3 \times 10^{-5}$, placing the corresponding optimal into the nonlinear non-localised regime. Again, the quantitative agreement of the gain, which was calculated to be 360.85, is in excellent agreement with the small channel value of 361.20. Furthermore, comparison of figure 23 with figure 13 reveals good qualitative agreement of the time evolution of the optimals for each channel size.

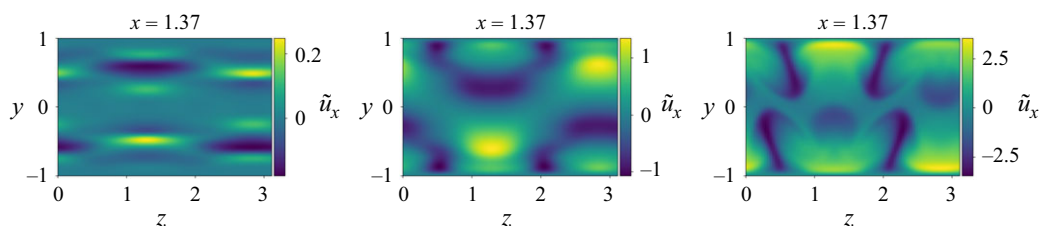


Figure 23. Streamwise velocity evolution ($t u_\tau / h = 0, 1.4, 2.8$) of the optimal for $T u_\tau / h = 2.8$ and $e_0 / E_0 = 3 \times 10^{-5}$ in a larger channel.

We can thus conclude that the smaller channel of $\pi \times 2 \times \pi$ is sufficient for the purposes of this investigation.

A.3. Gradient-based optimisation algorithm

We now turn to a more detailed description of the optimisation approach roughly introduced in § 3. Generally speaking, any gradient-based optimisation algorithm has two components: First, selecting a step direction and, second, selecting a step size determining how far to move along this direction. Note that in the literature, optimisation problems are often phrased as minimisation problems. Here, since we are trying to maximise the energy gain, we use language to reflect that we are solving a maximisation problem. The mathematics is obviously completely equivalent up to flipping signs. Notably, we refer to gradient ascent rather than gradient descent. A natural choice for the ascent direction is the steepest direction (given by the gradient); however, in the present work, we follow Cherubini *et al.* (2011) and use the conjugate gradient algorithm with the Polak–Ribière formula (Polak & Ribiere 1969). In concrete terms, the ascent direction $\mathbf{p}^{n+1}$ is determined using

$$\mathbf{p}^{n+1} = \frac{\delta \mathcal{L}}{\delta \tilde{\mathbf{u}}^n(\mathbf{x}, 0)} + \beta^n \mathbf{p}^n = \lambda^n \tilde{\mathbf{u}}^n(\mathbf{x}, 0) - \tilde{\mathbf{v}}^n(\mathbf{x}, 0) + \beta^n \mathbf{p}^n. \quad (\text{A1})$$

At the first step, β^0 is initialised with zero. Afterwards, it is calculated by

$$\beta^{n+1} = \frac{(\mathbf{g}^{n+1})^\top (\mathbf{g}^{n+1} - \mathbf{g}^n)}{(\mathbf{g}^n)^\top \mathbf{g}^n}, \quad (\text{A2})$$

where $\mathbf{g}$ is the steepest ascent direction. Note that some additional complexity arises from the fact that (2.14) depends on the Lagrange multiplier λ , which ensures the initial energy constraint. At step n , for a given step size α^n and conjugate parameter β^n , λ^n is determined implicitly from the formula

$$\frac{1}{2V} \iiint_V |\tilde{\mathbf{u}}^{n+1}|^2 dV = e_0, \quad (\text{A3})$$

where the new guess $\tilde{\mathbf{u}}^{n+1}$ is given by

$$\tilde{\mathbf{u}}^{n+1} = \tilde{\mathbf{u}}^n + \alpha \mathbf{p}^n, \quad (\text{A4})$$

and the ascent direction $\mathbf{p}^n$ is obtained from (A1). Here, we solve (A3) for λ using a standard Newton method, where the necessary gradients are obtained by automatic differentiation.

Unlike Cherubini *et al.* (2011), who adjust the step size depending on whether the last iteration was successful in increasing the objective function, our solver includes the option of using a line search. This can save time under some circumstances, but may also be counterproductive due to its inability to ‘escape’ local optima, so while the line search was used most of the time in obtaining the results presented here, some calculations fully or partially relied on the simpler adaptive step method described by Cherubini *et al.* (2011). The crucial advantage of line search algorithms is that instead of simply accepting an optimisation step, it is first checked if taking this step actually leads to a sufficient increase in the objective function, and only then is the step accepted. Thus, line search ensures that the objective function increases in each iteration. Moreover, even though more function evaluations are necessary in each time step, the step size is adaptively increased if possible, leading to fewer iterations until convergence. Whether this trade-off makes sense depends on the problem, but usually and also in the present case, it does.

The general idea behind a line search algorithm is to maximise the objective function along the chosen ascent direction. However, in practice, it is not a good use of computational resources to find an exact optimum, but it is better to instead settle on a step size that yields a reasonable improvement and is easy to find. Depending on the problem, many approaches exist for this. In the present work, once an ascent direction is determined, a step using the current step size α_i is taken, and the function is evaluated at the new guess. If the increase is big enough, i.e. fulfils the Armijo condition (Armijo 1966)

$$f(x + \alpha_i p) - f(x) \geq \alpha_i t, \quad (\text{A5})$$

where f is the objective function and t is a parameter between 0 and 1, the new guess is either accepted directly or α_i is increased until (A5) is no longer satisfied. Note that trying to increase the step size requires additional function evaluations, so in practice, it should usually not be done in every iteration, but only, say, every third iteration. However, not trying to increase the step size once in a while will likely lead to slow convergence due to a possibly too small step size. If (A5) is not satisfied, α_i is decreased until it is. The step size for the new iteration α_{i+1} is then chosen to be the final step size of the old iteration. Note that upper and lower limits for the step size α_i can be imposed to ensure reasonable behaviour of the solver. This was done in the present work, though the borders were chosen quite liberally as a restriction of the steps size should, in principle, not be required. The convergence criterion is that the relative change of the objective function between two successive iterations lies below a critical value (10^{-8} is chosen in the present work).

REFERENCES

- ARMIGO, L. 1966 Minimization of functions having Lipschitz continuous first partial derivatives. *Pac. J. Maths* **16** (1), 1–3.
- BERGGREN, M. 1998 Numerical solution of a flow-control problem: vorticity reduction by dynamic boundary action. *SIAM J. Sci. Comput.* **19** (3), 829–860.
- BRADBURY, J., FROSTIG, R., HAWKINS, P., JOHNSON, M.J., LEARY, C., MACLAURIN, D., NECULA, G., PASZKE, A., VANDERPLAS, J., WANDERMAN-MILNE, S., *et al.* 2018 JAX: composable transformations of Python+NumPy programs. version 0.3.13. Available at: <http://github.com/jax-ml/jax>.
- BUTLER K.M. & FARRELL B.F. 1993 Optimal perturbations and streak spacing in wall-bounded turbulent shear flow. *Phys. Fluids A: Fluid Dyn.* **5** (3), 774–777.
- CASSINELLI, A., DE GIOVANETTI, M. & HWANG, Y. 2017 Streak instability in near-wall turbulence revisited. *J. Turbul.* **18** (5), 443–464.
- CHERUBINI, S. & DE PALMA, P. 2013 Nonlinear optimal perturbations in a Couette flow: bursting and transition. *J. Fluid Mech.* **716**, 251–279.
- CHERUBINI, S., DE PALMA, P., ROBINET, J.-C. & BOTTARO, A. 2010 Rapid path to transition via nonlinear localized optimal perturbations in a boundary-layer flow. *Phys. Rev. E* **82** (6), 066302.

- CHERUBINI, S., DE PALMA, P., ROBINET, J.-C. & BOTTARO, A. 2011 The minimal seed of turbulent transition in the boundary layer. *J. Fluid Mech.* **689**, 221–253.
- CIOLA, N., DE PALMA, P., ROBINET, J.-C. & CHERUBINI, S. 2024 Large-scale coherent structures in turbulent channel flow: a detuned instability of wall streaks. *J. Fluid Mech.* **997**, A18.
- COSSU, C. & HWANG, Y. 2017 Self-sustaining processes at all scales in wall-bounded turbulent shear flows. *Phil. Trans. R. Soc. A: Math. Phys. Engng Sci.* **375** (2089), 20160088.
- FARANO, M., CHERUBINI, S., ROBINET, J.-C. & DE PALMA, P. 2017 Optimal bursts in turbulent channel flow. *J. Fluid Mech.* **817**, 35–60.
- FARANO, M., CHERUBINI, S., ROBINET, J.-C. & DE PALMA, P. 2015 Hairpin-like optimal perturbations in plane Poiseuille flow. *J. Fluid Mech.* **775**, R2.
- FARANO, M., CHERUBINI, S., ROBINET, J.-C. & PALMA, P.D. 2016 Subcritical transition scenarios via linear and nonlinear localized optimal perturbations in plane Poiseuille flow. *Fluid Dyn. Res.* **48** (6), 061409.
- FARRELL B.F. 1988 Optimal excitation of perturbations in viscous shear flow. *Phys. Fluids* **31** (8), 2093–2102.
- FARRELL, B.F. & IOANNOU, P.J. 2012 Dynamics of streamwise rolls and streaks in turbulent wall-bounded shear flow. *J. Fluid Mech.* **708**, 149–196.
- GIBSON, J.F. 2014 Channelflow: a spectral Navier–Stokes simulator in C++. *Tech. Rep.* U. Channelflow.org.
- GRIEWANK, A. & WALTHER, A. 2000 Algorithm 799: revolve: an implementation of checkpointing for the reverse or adjoint mode of computational differentiation. *ACM Trans. Math. Softw.* **26** (1), 19–45.
- HAMILTON, J.M., KIM, J. & WALEFFE, F. 1995 Regeneration mechanisms of near-wall turbulence structures. *J. Fluid Mech.* **287**, 317–348.
- HINZE, M., WALTHER, A. & STERNBERG, J. 2005 An optimal memory-reduced procedure for calculating adjoints of the instationary Navier–Stokes equations. *Optim. Control Applics. Meth.* **27** (1), 19–40.
- HOEPFFNER, J., BRANDT, L. & HENNINGSON, D.S. 2005 Transient growth on boundary layer streaks. *J. Fluid Mech.* **537** (1), 91–100.
- JIMÉNEZ, J. & MOIN, P. 1991 The minimal flow unit in near-wall turbulence. *J. Fluid Mech.* **225**, 213–240.
- JIMÉNEZ, J. & PINELLI, A. 1999 The autonomous cycle of near-wall turbulence. *J. Fluid Mech.* **389**, 335–359.
- KERSWELL, R.R. 2018 Nonlinear nonmodal stability theory. *Annu. Rev. Fluid Mech.* **50** (1), 319–345.
- KIM, H.T., KLINE, S.J. & REYNOLDS, W.C. 1971 The production of turbulence near a smooth wall in a turbulent boundary layer. *J. Fluid Mech.* **50** (1), 133–160.
- KIM, J., MOIN, P. & MOSER, R. 1987 Turbulence statistics in fully developed channel flow at low Reynolds number. *J. Fluid Mech.* **177**, 133–166.
- LLUESMA-RODRÍGUEZ F., ÁLCANTARA-ÁVILA F., PÉREZ-QUILES M.J. & HOYAS S. 2021 A code for simulating heat transfer in turbulent channel flow. *Mathematics* **9** (7), 756.
- LOZANO-DURÁN, A., CONSTANTINOU, N.C., NIKOLAIDIS, M.-A. & KARP, M. 2021 Cause-and-effect of linear mechanisms sustaining wall turbulence. *J. Fluid Mech.* **914**, A8.
- MALKUS, W.V.R. 1956 Outline of a theory of turbulent shear flow. *J. Fluid Mech.* **1** (5), 521–539.
- MARKEVICIUTE, V.K. & KERSWELL, R.R. 2024 Threshold transient growth as a criterion for turbulent mean profiles. *J. Fluid Mech.* **996**, A32.
- 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** (13), 134502.
- ORSZAG, S.A. & PATTERSON, G.S. 1972 Numerical simulation of three-dimensional homogeneous isotropic turbulence. *Phys. Rev. Lett.* **28** (2), 76–79.
- POLAK, E. & RIBIERE, G. 1969 Note sur la convergence de méthodes de directions conjuguées. *Modél. Math. Anal. Numér.* **3** (1), 35–43.
- PRINGLE, C.C. T., WILLIS, A.P. & KERSWELL, R.R. 2012 Minimal seeds for shear flow turbulence: using nonlinear transient growth to touch the edge of chaos. *J. Fluid Mech.* **702**, 415–443.
- 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** (15), 154502.
- PUJALS G., GARCÍA-VILLALBA M., COSSU C. & DEPARDON S. 2009 A note on optimal transient growth in turbulent channel flows. *Phys. Fluids* **21** (1), 015109.
- 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., BAGGETT, J.S. & HENNINGSON, D.S. 1998 On stability of streamwise streaks and transition thresholds in plane channel flows. *J. Fluid Mech.* **365**, 269–303.
- REYNOLDS, W.C. & HUSSAIN, A.K.M.F. 1972 The mechanics of an organized wave in turbulent shear flow. Part 3. Theoretical models and comparisons with experiments. *J. Fluid Mech.* **54** (2), 263–288.
- REYNOLDS, W.C. & TIEDERMAN, W.G. 1967 Stability of turbulent channel flow, with application to Malkus’s theory. *J. Fluid Mech.* **27** (2), 253–272.

- SCHOPPA, W. & HUSSAIN, F. 2002 Coherent structure generation in near-wall turbulence. *J. Fluid Mech.* **453**, 57–108.
- SPALART, P.R., MOSER, R.D. & ROGERS, M.M. 1991 Spectral methods for the Navier–Stokes equations with one infinite and two periodic directions. *J. Comput. Phys.* **96** (2), 297–324.
- TRUONG, T.T. & NGUYEN, H.-T. 2020 Backtracking gradient descent method and some applications in large scale optimisation. Part 2: Algorithms and experiments. *Appl. Maths Optim.* **84** (3), 2557–2586.
- WALEFFE F. 1997 On a self-sustaining process in shear flows. *Phys. Fluids* **9** (4), 883–900.