Multigrid optimization for DNS-based optimal control in turbulent channel flows

C. Nita^{a,b,*}, S. Vandewalle^b, J. Meyers^a

^aDepartment of Mechanical Engineering, KU Leuven, Celestijnenlaan 300A, B3001 Leuven, Belgium

^bDepartment of Computer Science, KU Leuven, Celestijnenlaan 200A, B3001 Leuven, Belgium

Abstract

The use of PDE-constrained optimization techniques in combination with transient three-dimensional turbulent flow simulations such as Direct Numerical Simulation (DNS) or Large-Eddy Simulation (LES) involves large computational cost and memory resources. To date, the minimization of a DNS-based cost functional is typically achieved by applying classical single-grid gradient-based iterative methods of quasi-Newton or non-linear conjugate gradient type. In the current study, a multigrid optimization (MG/OPT) strategy is investigated in order to speed up gradient-based algorithms designed for large scale optimization problems. The method employs a hierarchy of optimization problems defined on different representation levels. It aims to reduce the computational resources associated with the cost functional improvement on the finest level. We apply the MG/OPT method in the context of direct numerical simulations of a fully developed channel flow problem. The performance of the multigrid optimization technique is compared against the single-grid optimization method in terms of equivalent function and gradient evaluations. Also the influence of the optimization problem properties and algorithmic parameters are investigated. It is found that, in some cases, the MG/OPT method accelerates the single-grid damped L-BFGS method by a factor of four.

Keywords: DNS-based optimization, multigrid methods, damped L-BFGS, adjoint method, turbulent flow

1. Introduction

The application of optimal control theory to complex three-dimensional turbulence requires detailed numerical simulations to represent the desired effects of the controls and to capture the flow field response [1–8]. Examples cover different areas such as drag reduction in turbulent boundary layers [1, 9], turbulent mixing [10, 11], noise reduction [6, 12] and wind-farm energy extraction [7, 8, 13]. In this context, the use of direct numerical simulation (DNS) or large-eddy simulation (LES) of turbulent flows implies large computational resources associated with the numerical evaluation of the objective functional that is used to define the optimization problem. Recently, various gradient based optimization algorithms have been tested for DNS-based optimal control in terms of number of forward and adjoint simulations required for cost functional improvement [14]. However, these methods were based on a single-grid optimization approach. In the current study, we examine multigrid optimization [15] in order to accelerate classical gradient-based optimization algorithms [16, 17] for unconstrained turbulent flow minimization problems. As a single-grid reference method, we employ a damped L-BFGS quasi-Newton method, which emerged in Ref. [14] as the most efficient single-grid method for the optimal-control problems that we consider in the current study.

In recent years, there has been an increasing interest in the application of multigrid schemes to the optimization of systems governed by partial differential equations. This interest is motivated by the need for an efficient algorithm in terms of computational cost [15, 18]. In the context of optimization different multigrid strategies for linear and non-linear problems have been proposed [15, 19–21]. A feature to distinguish between these techniques is the step in the computational algorithm where the multigrid method is employed. The most straightforward approach is to use the multigrid scheme as an inner solver in the optimization algorithm, i.e., by solving the forward and adjoint problems using a multilevel method. This technique has been successfully applied for parameter and shape optimization problems [22, 23], for design spaces of size $O(10^2)$. A second approach is the so-called one-shot

*Corresponding author

Email addresses: cornelia.nita@kuleuven.be (C. Nita), stefan.vandewalle@kuleuven.be (S. Vandewalle), johan.meyers@kuleuven.be (J. Meyers)

multigrid method [24–26]. In this case, the optimality system is solved for the state, adjoint and design variables simultaneously by a multigrid method. A collective smoother strategy is performed for the variables, coupling the state equation, the adjoint equation and the optimality conditions [18, 20, 25]. However, flow optimization problems that use DNS or LES are characterized by extremely large state spaces, so that simultaneous approaches may well be infeasible. Instead a reduced space formalism is typically used, in which the state equation is treated implicitly in the optimization problem [18, 27]. To that end, at every step in the optimization process, a DNS is performed in order to keep the state equations satisfied given the evolving design parameters (see, e.g., [28–30]). In the current study, we also follow this approach, and concentrate on the multigrid scheme proposed by Nash and Lewis [15], which is best constructed for such an approach.

The multigrid optimization strategy denoted by MG/OPT in Refs. [15, 31, 32] aims to reduce the computational resources associated with the cost functional by performing many optimization steps on coarser (and, hence, cheaper) discretization levels. The search directions and step lengths computed on lower fidelity grids are injected back into the fine level to determine a final search direction and step size. Numerical experiments indicate that under appropriate conditions MG/OPT accelerates classical one-grid optimization algorithms. It has been applied successfully in combination with various methods, such as the truncated Newton method [15, 33], the steepest descent method [34] and the nonlinear conjugate gradient method [35–37].

In the current work, the optimization algorithm is applied in the context of optimal control of direct numerical simulations of a fully developed channel flow problem. We consider two different objective functionals in the optimal control problem, similar to the earlier work in Ref. [14]. The first objective functional aims at decreasing the turbulent kinetic energy at the end of the optimal control time horizon, which is relevant for achieving drag reduction [1, 9]. The second objective functional aims at increasing the total extracted energy by an internal distributed force in a part of the channel. This problem is related to the problem of maximizing wind-farm performance in a turbulent boundary layer [7, 8, 13]. The performance of the multigrid algorithm is compared with single-grid optimization in terms of equivalent functional evaluations on the finest level.

The paper is organized as follows. Section 2 provides the details of the problem formulation for a turbulent channel flow optimization problem and the discretization methods used in this study. Subsequently, in Section 3 the multigrid optimization strategy is elaborated in the context of DNS-based optimal control. Section 4 presents the set-up of the cases and the results obtained with the MG/OPT algorithm. Finally in Section 5, we present some conclusions.

2. Problem formulation

We focus on two different test problems in a turbulent channel flow, i.e., one related to maximizing energy extraction, and one related to classical drag reduction. Details of the related cost functionals are provided in §2.1. They are both symbolically represented by $\mathcal{J}(\mathbf{f}, \mathbf{u})$, with $\mathbf{f}$ the control parameter, and $\mathbf{u}(\mathbf{x}, t)$ the state variable, i.e., the turbulent velocity field. The optimization problem is cast in its reduced form, i.e.,

$$\min_{\mathbf{f}} \hat{\mathcal{J}}(\mathbf{f}) \equiv \mathcal{J}(\mathbf{f}, \mathbf{u}(\mathbf{f})), \quad (1)$$

where $\mathbf{u}(\mathbf{f})$ represents the state variables given the parameter $\mathbf{f}$, implicitly defined as the solution of the governing model:

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} + \nabla p - \frac{1}{Re} \nabla^2 \mathbf{u} = \mathbf{f} + f_\infty \mathbf{e}_1 \quad \text{in } \Omega \times (0, T], \quad (2)$$

$$\nabla \cdot \mathbf{u} = 0 \quad \text{in } \Omega \times (0, T], \quad (3)$$

$$\mathbf{u}|_{t=0} = \mathbf{u}_0 \quad \text{on } \Omega, \quad (4)$$

$$\mathbf{u} = 0 \quad \text{on } \Gamma_3^+ \cup \Gamma_3^-, \quad (5)$$

$$\mathbf{u}|_{\Gamma_1^-} = \mathbf{u}|_{\Gamma_1^+} \quad \text{and} \quad \mathbf{u}|_{\Gamma_2^-} = \mathbf{u}|_{\Gamma_2^+}. \quad (6)$$

The constraints are the incompressible Navier–Stokes equations, Eqs. (2, 3), completed with a proper set of boundary conditions on the walls Eq. (5, 6). Periodic boundary conditions are imposed in the streamwise (x_1) and the spanwise (x_2) directions and a no-slip boundary condition on the lower Γ_3^- and upper Γ_3^+ walls. The flow is driven by a constant pressure gradient $f_\infty \mathbf{e}_1 \equiv -\partial p_\infty / \partial x$, with $\mathbf{e}_1$ the unit vector in the streamwise direction. The term $\mathbf{f}$ represents a volume force distributed over a volume $\Omega_f \subset \Omega$, with Ω the full channel-flow domain. For sake of convenience, we express the state equations Eq. (2-3) using the short-hand notation $B(\mathbf{u}(\mathbf{f}), \mathbf{f}) = 0$.

Given the PDE optimization problem, the sensitivity of the objective functional (Eq. 1) w.r.t. changes $\delta \mathbf{f}$ in $\mathbf{f}$ is typically estimated using the adjoint-based gradient technique (see, e.g., [28, 29, 38]):

$$\hat{\mathcal{J}}'(\mathbf{f}, \delta \mathbf{f}) \equiv (\nabla \hat{\mathcal{J}}, \delta \mathbf{f}), \quad (7)$$

where the gradient $\nabla \hat{\mathcal{J}}$ is expressed in terms of the adjoint variables. The L^2 inner product $(\cdot, \cdot)$ between the arbitrary changes $\delta \mathbf{f}$ in the control variable $\mathbf{f}$ and the gradient is defined by:

$$(\mathbf{a}, \mathbf{b}) = \int_0^T \int_{\Omega} a_i \cdot b_i \, dx dt. \quad (8)$$

We define the adjoint $[\partial \mathcal{B} / \partial \mathbf{f}]^*$ of the operator $\partial \mathcal{B} / \partial \mathbf{f}$ through the following identity [7, 18]:

$$\left(\mathbf{q}, \frac{\partial \mathcal{B}}{\partial \mathbf{f}} \delta \mathbf{f} \right) \equiv \left(\left[\frac{\partial \mathcal{B}}{\partial \mathbf{f}} \right]^* \mathbf{q}, \delta \mathbf{f} \right). \quad (9)$$

The gradient of the reduced cost functional evaluated at the current value of the control is given by [18, 27]:

$$\nabla \hat{\mathcal{J}} = \frac{\partial \mathcal{J}}{\partial \mathbf{f}} + \left[\frac{\partial \mathcal{B}}{\partial \mathbf{f}} \right]^* \mathbf{q} \quad (10)$$

We introduce the Lagrange multipliers $\mathbf{q} = (\boldsymbol{\xi}, \pi)$ with $\boldsymbol{\xi}$ the multiplier corresponding to Eq. (2) and π the multiplier to Eq. (3). In the current study, the adjoint equations for the gradient evaluation are derived using the continuous adjoint approach, leading to the following backward-in-time system [10, 27, 39]:

$$-\frac{\partial \boldsymbol{\xi}}{\partial t} + (\nabla \mathbf{u})^T \boldsymbol{\xi} - (\mathbf{u} \cdot \nabla) \boldsymbol{\xi} - \nabla \pi - \frac{1}{Re} \nabla^2 \boldsymbol{\xi} = \mathbf{f}^* \quad \text{in } \Omega \times (0, T], \quad (11)$$

$$-\nabla \cdot \boldsymbol{\xi} = 0 \quad \text{in } \Omega \times (0, T], \quad (12)$$

$$\boldsymbol{\xi}|_{t=T} = \frac{\partial \hat{\mathcal{J}}}{\partial \mathbf{u}_T} \quad \text{on } \Omega, \quad (13)$$

$$\boldsymbol{\xi} = 0 \quad \text{on } \Gamma_3^+ \cup \Gamma_3^-, \quad (14)$$

$$\boldsymbol{\xi}|_{\Gamma_1^-} = \boldsymbol{\xi}|_{\Gamma_1^+} \quad \text{and} \quad \boldsymbol{\xi}|_{\Gamma_2^-} = \boldsymbol{\xi}|_{\Gamma_2^+}. \quad (15)$$

The adjoint initial condition, Eq. (13), is referred to as the ‘terminal’ condition, as it is defined at the time horizon T with $\mathbf{u}_T = \mathbf{u}(\mathbf{x}, T)$. The forcing term $\mathbf{f}^*$ from the adjoint momentum equation Eq. (11), is identified from the cost functional variation with respect to the state variables:

$$\mathbf{f}^* = \frac{\partial \hat{\mathcal{J}}}{\partial \mathbf{u}}. \quad (16)$$

The boundary conditions associated with the adjoint problem are similar to those that are imposed on the state system. In the wall-normal direction, no-slip conditions are required for the top and the bottom wall Eq. (14); in streamwise and spanwise direction periodic boundary conditions Eq. (15) are to be imposed.

2.1. Cost functional and controls

The first optimization problem that we consider, focusses on increasing energy extraction by internally distributed forces. The choice of this particular turbulent flow is motivated by its relevance to optimal control of energy extraction in wind farms, a topic of intense interest in wind energy community. In this context, the modelling of aerodynamic forces on the turbine blade and the evaluation of the impact of the wake effects on the wind farm performance is extremely challenging [7, 8, 13]. In this study, we neglect the details of wind turbine modelling. Instead, we use simple volume forces in a subregion of the domain. Thus, our first problem is a canonical case related to the optimal control of a wind farm. The cost functional is given by [14]

$$\mathcal{J} = \int_0^T \int_{\Omega_f} \mathbf{f} \cdot \mathbf{u} \, dx dt, \quad (17)$$

and corresponds to the power extracted from the boundary layer over a time horizon T in the forcing volume Ω_f . In order for the optimal-control problem from Eq. (17) to be mathematically rigorous defined, the maximum amount of energy that can be extracted from the flow should be bounded from above. This is not directly straightforward,

Figure 1: The contours of the instantaneous velocity field $u_1(x, t)/u_\tau$ at $Re_\tau = 395$ evaluated at the initial conditions for optimal control

but can be shown under certain conditions, as further discussed in Appendix A. The control is represented by a streamwise body force distributed over $\Omega_f \subset \Omega$, and is defined as

$$\mathbf{f} = \begin{cases} -f_1(x, y, t) \cdot \mathbf{e}_1, & \Omega_f, \\ 0, & \Omega \setminus \Omega_f. \end{cases} \quad (18)$$

Another important fluid mechanics problem is the control of wall-bounded turbulent flows in order to reduce skin-friction drag. Various control strategies, e.g., the travelling wave-like blowing/suction [40], the streamwise Lorenz force [41], the constant near-wall forcing [42] or the wave-like wall-normal body force [43], have been proposed to suppress the near-wall turbulence, and, as such, achieve the desired goal. Hence, our second optimization problem aims at drag reduction in an incompressible, fully developed channel flow. The problem is formulated as a physically equivalent minimization problem of terminal turbulent kinetic energy (*TKE*) [1, 9, 14]

$$\mathcal{J} = \int_{\Omega} \|\mathbf{u}(\mathbf{x}, T) - \langle \mathbf{u}(\mathbf{x}, T) \rangle\|^2 d\mathbf{x}, \quad (19)$$

where $\langle \mathbf{u}(\mathbf{x}, T) \rangle$ is the horizontally averaged velocity field. In this case, the control force $\mathbf{f}$ is a wave-like wall-normal body force. It is formulated as [14, 43]:

$$\mathbf{f} = \exp\left(-\frac{z^+}{\Delta^+}\right) f_2(x, t) \cdot \mathbf{e}_3, \quad (20)$$

For the *TKE* cases, the design variables are allowed to vary only in streamwise direction and in time. The fixed penetration length Δ^+ has a value of 12.5 measured in wall units. The relevance of this value follows from the parametric study of Mamori & Fukagata [43]. Note that the forcing term for this particular case is only added at the bottom wall.

A common strategy in the cost functional definition is to add a regularization term that explicitly penalizes the control magnitude. This term is, sometimes even necessary to ensure the boundedness of the optimal control. In the current cost functional, however, we omit this term. Based on Ref. [44] and on the numerical experiments from Ref. [1] we note that neglecting this term in optimal control problems governed by Navier–Stokes equations still leads to a bounded solution in a subset of well-defined optimization problems. Moreover, for our specific problem we found that the energy spectrum of f_2 decays rapidly with the wavenumber and frequency, so the distributed control forces are smooth in space and time. Furthermore, the existence, uniqueness and regularity of various optimal control problems governed by Navier–Stokes equations involving both distributed and boundary controls

Figure 2: The mean velocity profile (left) and the root-mean square velocity fluctuations $u_{1,rms}$ (—), $u_{2,rms}$ (---) and $u_{3,rms}$ (-·-·-) (right) for the turbulent channel flow simulations compared to DNS data. $Re_\tau = 110$ simulations ($\diamond$) compared with database results from Ref. [47] ($\bullet$) and $Re_\tau = 395$ simulations ($\triangle$) compared with database results from Ref. [48] (—).

have been extensively investigated in Refs. [2, 28, 45, 46]. However, a rigorous proof for the specific optimal control problems presented above is beyond the scope of the current work as, for the 3D turbulent regime considered here, it relates to the general existence and uniqueness of Navier–Stokes solutions, which is an important unsolved problem in mathematics. Further details of both problem formulations can also be found in Ref. [14].

2.2. Numerical methods

DNS of incompressible turbulent channel flow is performed with the in-house SP-Wind solver developed at KU Leuven [7, 10, 11, 49]. The forward and adjoint equations are discretized by using a parallel mixed pseudo-spectral finite difference method. A fourth-order staggered finite difference scheme [50] is used in the wall normal direction. A pseudo-spectral method with a Fourier discretization is implemented in the wall-parallel directions [51], and dealiasing is performed using the 3/2 rule. Periodic boundary conditions are imposed in the streamwise (x) and the spanwise (y) directions and no-slip conditions are used in the normal direction (z). The grid in the wall-normal direction is stretched using a hyperbolic tangent function. The time integration is performed by a standard fourth order explicit Runge-Kutta scheme. During the forward simulations, the state variables (velocity) are stored at every time step and at every grid point, to be used in the adjoint simulations. The time steps δt are restricted by setting the convective and diffusive Courant-Friedrichs–Lewy (CFL) number to 0.4.

We apply the MG/OPT algorithm for two Reynolds numbers, i.e., $Re_\tau = 110$, and $Re_\tau = 395$, respectively, based on the friction velocity $u_\tau \equiv (\delta/\rho\partial p_\infty/\partial x)^{1/2}$ and the channel half-width $\delta (= 1)$. Furthermore, $\partial p_\infty/\partial x$ is the driving pressure gradient in the channel. We select the computational domain as $L_x \times L_y \times L_z = 2\pi\delta \times \pi\delta \times 2\delta$. The finest grid for the low Reynolds number case is $N = 128 \times 96 \times 144$; a resolution of $N = 256 \times 192 \times 288$ is used for the moderate Reynolds number case. The simulations are initialized using a quadratic streamwise velocity profile to which random perturbations are added. Subsequently, simulations are progressed over 250 through-flow times to spin up realistic turbulence. The turbulent velocity field at the end of the spin up is used as the initial condition for simulations and optimal control. Figure 1 shows the contours of the instantaneous velocity field for the $Re_\tau = 395$ case at the starting point of the optimal control algorithm. We verified that both grids are sufficiently fine to perform adequate DNS of channel flow at these Reynolds numbers. Figure 2 shows profiles of the mean velocity and profiles of the root-mean-square velocity fluctuations respectively, for the two flow cases. The results are compared to the reference database of Refs. [48] and [47]. We observe that the computational mesh and time steps are sufficiently fine for grid-converged DNS, as the horizontally-averaged profiles match the reference data very well. Note that, when the fully turbulent flow regime is generated by employing periodic boundary conditions, longer numerical domains are needed in order to avoid non-physical effects on large scale turbulent flow structures [52]. However, it is shown [53] that these outer-layer flow structures do not significantly contribute to the turbulence in the near wall region at the current Reynolds numbers. Moreover, the optimization for such domain lengths implies very expensive forward simulations and very high memory requirements. To determine how the turbulent kinetic energy is distributed among the wide range of turbulent structures, the energy

Figure 3: One-dimensional spectrum E_{uu} as a function of streamwise direction (k_x) evaluated at wall distance location $z^+ \approx 5$ (solid line), $z^+ \approx 80$ (dot-dashed line) and $z^+ \approx 250$ (dashed line) at $Re_\tau = 395$. The symbols represent the corresponding database results from Ref. [48].

spectrum is analysed. Figure 3 illustrates the variation of the one-dimensional spectrum E_{uu} with the length of the channel domain. These findings are compared with the DNS data at three different distances from the wall. It is shown that the results are consistent with the database for the highest Reynolds number used in our numerical experiments, i.e., $Re_\tau = 395$. Thus, for the present purpose, the simulation domain size is adequate to provide a base flow for testing different optimization algorithms.

Finally, the volume force for the energy extraction optimization case is located between $z^+ \approx [5.5, 15.5]$ at $Re_\tau = 110$ and between $z^+ \approx [19.8, 55.03]$ at $Re_\tau = 395$, in the wall normal direction, and varies only in time and in wall-parallel directions. The symbol z^+ is used to indicate that the distance is given in wall units.

3. The multilevel optimization approach

Nowadays the method of choice to solve the unconstrained minimization problem defined in Eq. (1) is a single-grid gradient-based iterative method for large scale optimization such as quasi-Newton, truncated Newton or non-linear conjugate gradient. In the current study we employ a multigrid optimization approach [15, 33] in which the inner optimization algorithm, the so-called smoother, is a damped limited memory BFGS method. For comparison purposes, the use of a nonlinear conjugate-gradient method as a smoother is also considered. The multigrid optimization strategy aims to reduce the computational demands associated with the cost functional improvement in fine scale optimization problems. The method employs a hierarchy of optimization problems defined on different discretization levels. It computes certain numerical information with less computational cost by considering optimization problems with lower fidelity resolutions. In general, the MG/OPT convergence is conditioned by the implementation of a line search procedure in the coarse-grid correction step [54, 55]. In the optimization context, the effectiveness of applying a multilevel approach can be established by analysing the properties of the reduced Hessian [33]. The challenges to analyse the reduced Hessian are identified in Ref. [31] showing that the strategies used to investigate the classical multigrid are not suitable in optimal control settings. Note that, in our numerical experiments, to guarantee that the MG/OPT or the single grid optimization algorithm converges to a local minimum, we assume that the reduced Hessian is positive definite at the local solution. Sometimes additional constraints on the coarse minimization problem have been implemented in order to improve the performance [33, 56]. Numerical experiments [14] have shown that the damped version of the quasi-Newton method performs well for unconstrained optimal control problems. So, in this paper, we investigate whether this method can be accelerated further by using a multilevel strategy.

3.1. The MG/OPT algorithm

In this subsection we elaborate the methodology of the multigrid optimization strategy. In §3.1.1 the general characteristics of the MG/OPT method are detailed. In §3.1.2 and §3.1.3 the downward and backward branches of a V-cycle in the MG/OPT method are discussed. In §3.1.4 the formulation of the algorithm is presented, and finally, in §3.1.5 the MG/OPT correction is discussed.

3.1.1. General characteristics

The MG/OPT approach accelerates single-grid optimization methods by using information from coarser grids to generate a search direction and a step size on finer level problems. To this end, a hierarchy of grids Ω_k is defined indexed as, $k = 0, 1, \dots, K$, with 0 the finest (DNS) level and K th the coarsest. In the current study, computational grids are coarsened in all dimensions (space and time) with a constant coarsening factor n . The number of discretization variables on level k is then a factor of n^{4k} smaller than the corresponding number on the resolved DNS-level ($k = 0$). This clearly indicates that any useful optimization work that can be performed on coarser levels, can lead to significant speed-ups in the optimization.

We introduce the vector $\mathbf{f}_k \in V_k$ to denote the design parameters, and the symbol $\hat{\mathcal{J}}_k$ used to denote the cost function. Both quantities are discretized on grid level k . A scaled classical Euclidean inner product $(\cdot, \cdot)_k$ between vectors in V_k is used, such that $(f, g)_2$ approximates the continuous L_2 inner product. Hence, inner products on different levels lead to identical values for constant functions. MG/OPT also uses inter-grid transfer operators, i.e., a restriction operator:

$$\mathbf{I}_{k-1}^k : V_{k-1} \rightarrow V_k \quad (21)$$

and a prolongation operator:

$$\mathbf{I}_k^{k-1} : V_k \rightarrow V_{k-1}. \quad (22)$$

They are often selected in such a way that:

$$(\mathbf{I}_k^{k-1} \mathbf{a}, \mathbf{b})_{k-1} = (\mathbf{a}, \mathbf{I}_{k-1}^k \mathbf{b})_k \quad (23)$$

for all $\mathbf{a} \in V_k$ and $\mathbf{b} \in V_{k-1}$. Equation (23) is called the adjoint identity for restriction and prolongation. This identity is one of the conditions necessary to guarantee a descent direction for MG/OPT. Details about the effective operators $\mathbf{I}_{k-1}^k$ and $\mathbf{I}_k^{k-1}$ used in the current study, are provided in §3.3 below.

MG/OPT essentially comes down to a full approximation scheme (FAS) applied to the Karush-Kuhn-Tucker (KKT) conditions of the reduced optimization problem $\hat{\mathcal{J}}_0(\mathbf{f}_0)$ [15, 18]. On the fine mesh the optimality condition can be formulated as:

$$\nabla \hat{\mathcal{J}}_0(\mathbf{f}_0) = 0. \quad (24)$$

The method, then, solves problems of the type $\nabla \hat{\mathcal{J}}_k(\mathbf{f}_k) = \boldsymbol{\tau}_k$ with $\boldsymbol{\tau}_k$ a suitably selected correction term, on different grid levels k . It uses the solutions to these problems to improve a guess for the solution $\mathbf{f}_0$ on the finest grid level. For a three level FAS multigrid scheme, the equation on a level $k = 1$ is given by:

$$\nabla \hat{\mathcal{J}}_1(\mathbf{f}_1) = \nabla \hat{\mathcal{J}}_1(\mathbf{I}_0^1 \mathbf{f}_0) - \mathbf{I}_1^0 \nabla \hat{\mathcal{J}}_0(\mathbf{f}_0), \quad (25)$$

where the $\mathbf{f}_0$ in the right hand side denotes the current approximation to the solution of Eq. (24). Note that if $\mathbf{f}_0$ represents the exact solution of Eq. (24), then the second term in the right hand side vanishes. In that case the solution to Eq. (25) becomes $\mathbf{f}_1 = \mathbf{I}_0^1 \mathbf{f}_0$. Hence the right hand side in Eq. (25) can be seen as a kind of correction term. It ensures, that, upon convergence, the solution on a level $k = 1$ becomes identical to that on level $k = 0$, restricted to the coarse grid. Denote the right hand side of Eq. (25) by $\boldsymbol{\tau}_1$, then, the FAS equation on level $k = 2$ becomes:

$$\nabla \hat{\mathcal{J}}_2(\mathbf{f}_2) = \nabla \hat{\mathcal{J}}_2(\mathbf{I}_1^2 \mathbf{f}_1) - \mathbf{I}_2^1 (\nabla \hat{\mathcal{J}}_1(\mathbf{f}_1) - \boldsymbol{\tau}_1). \quad (26)$$

Again, the $\mathbf{f}_1$ in the right hand side denotes the current approximation to the solution on level $k = 1$.

In the MG/OPT method, equations (24), (25) and (26) will be interpreted as the optimality conditions of the optimization problems:

$$\min_{\mathbf{f}_0} \hat{\mathcal{J}}_0(\mathbf{f}_0) \quad (27)$$

$$\min_{\mathbf{f}_1} \hat{\mathcal{J}}_1(\mathbf{f}_1) - (\boldsymbol{\tau}_1, \mathbf{f}_1)_1 \quad (28)$$

$$\min_{\mathbf{f}_2} \hat{\mathcal{J}}_2(\mathbf{f}_2) - (\boldsymbol{\tau}_2, \mathbf{f}_2)_2 \quad (29)$$

Figure 4: Schematic representation of a two-level V-cycle in MG/OPT algorithm with 0 the finest level, 1 the coarsest level. The inter-grid operators, e.g., restriction I_0^1 and prolongation I_1^0 , the optimization problem on the grid level and the downward and upward branches of the V-cycle are sketched.

with τ_2 , denoting the right hand side of Eq. (26). The computation of a single-level update to f_0, f_1, f_2 in the so-called MG/OPT smoothing step could be done by using a classical iterative method applied to Eqs.(24)–(26). In MG/OPT, however, such an update will be done by applying a few steps of a classical, single-level optimization algorithm. The general form of the problem to solve is given by:

$$\nabla \hat{\mathcal{J}}_k(f_k) = \tau_k, \quad (30)$$

with the correction term expressed as:

$$\tau_k = I_{k-1}^k \tau_{k-1} + \nabla \hat{\mathcal{J}}_k(I_{k-1}^k f_{k-1}) - I_{k-1}^k \nabla \hat{\mathcal{J}}_{k-1}(f_{k-1}). \quad (31)$$

In the current manuscript, we will only use V-cycles; extension to other cycles is straightforward. Once a solution f_K on the coarsest level is obtained, consecutive prolongation operations are used to update the solution on finer grid levels, completing the V-cycle (see Upward branch of the V-cycle subsection below). The precise formula for these updates is given further in Eq. (34). Figure 4 depicts a schematic diagram of the MG/OPT method for a two-level V-cycle. The details about the downward and the upward branches and the MG/OPT algorithm are given below.

3.1.2. Downward branch of the V-cycle

During the downward branch of the V-cycle, the grid levels are recursively coarsened. On every level a smoother is used for a few iterations (see below) so that $f_k^0 \rightarrow f_k^1$, with f_k^0 the unsmoothed guess of the solution at level k , and f_k^1 the guess after smoothing. Moreover, we select the initial approximation on level k as $f_k^0 = I_{k-1}^k f_{k-1}^1$. Eq. (31) then becomes:

$$\tau_k = I_{k-1}^k \tau_{k-1} + \nabla \hat{\mathcal{J}}_k(f_k^0) - I_{k-1}^k \nabla \hat{\mathcal{J}}_{k-1}(f_{k-1}^1). \quad (32)$$

As mentioned before, in MG/OPT the smoother is not formulated as a classical multigrid smoother on the system $\nabla \hat{\mathcal{J}}_k(f_k) - \tau_k = 0$, but instead a few iterations of a conventional gradient-based optimizer are used on the equivalent problem

$$\min_{f_k} \hat{\mathcal{J}}_k(f_k) - (\tau_k, f_k)_k. \quad (33)$$

To this end, we will employ the nonlinear conjugate-gradient method, as well as a damped L-BFGS method. Details about the single-grid optimization algorithms are provided in §3.2.

3.1.3. Upward branch of the V-cycle

On the coarsest grid level K a number of smoothing steps are applied to approximately solve the optimization problem at that level. This yields the approximation f_K^1 , and we set $f_K^3 = f_K^1$. Now the upward branch of the V-cycle is started. The prolongation from k to $k-1$ is used to compute an update to the current approximation at level $k-1$, i.e., the value f_{k-1}^2 . The update formula is as follows:

$$f_{k-1}^2 = f_{k-1}^1 + \alpha_{k-1} I_{k-1}^{k-1} (f_k^3 - I_{k-1}^k f_{k-1}^1). \quad (34)$$

Note that the superscripts 0, 1, 2, 3 of the control variable f_k denote the initial, the pre-smoothing, the corrected and the post-smoothing approximations on level k . Details of post-smoothing are provided in §3.1.4. The parameter α_{k-1} is actually a step length parameter. It is selected in such a way that the step corresponds to an effective descent. In the current manuscript, the decrease condition $\hat{\mathcal{J}}_{k-1}(f_{k-1}^2) \leq \hat{\mathcal{J}}_{k-1}(f_{k-1}^1)$ is satisfied by imposing the first Wolfe condition. This condition is first verified for $\alpha_{k-1} = 1$. A Powell line search method [57] is used when $\alpha_{k-1} = 1$ does not satisfy this condition. This line search strategy is employed also in the damped L-BFGS smoother whereas for the non-linear conjugate gradient smoother, the Brent line search method is used. For the MG/OPT method to work efficiently, a value of $\alpha_{k-1} = 1$ needs to work most of the time (in particular on the finest grid levels). Any lengthy line-search on a fine grid level indicates that the prior coarse grid optimization problem was not properly representing the fine-grid problem.

3.1.4. Formulation of the algorithm

One V-cycle of the MG/OPT algorithm can be formulated recursively as follows. The algorithm formulates the V-cycle on level k in terms of a V-cycle on level $k + 1$:

1. Smoothing

- apply γ_k smoothing steps to the optimization problem on level k , starting from f_k^0 , to obtain f_k^1

2. Coarse level problem

- restrict the control variables f_k^1 to level $k + 1$: $f_{k+1}^0 = I_k^{k+1} f_k^1$
- compute the τ_{k+1} using Eq. (32)
- if $k + 1 = K$ then apply γ_K smoothing steps to f_K^0 , to obtain f_K^3
- else apply a V-cycle on level $k + 1$ to get f_{k+1}^3

3. Correction

- transfer the search direction to level k and perform a line search:

$$f_k^2 = f_k^1 + \alpha_k I_{k+1}^k (f_{k+1}^3 - I_k^{k+1} f_k^1)$$

4. Smoothing

- apply γ_k smoothing steps to the optimization problem on level k starting from f_k^2 to obtain f_k^3 .

3.1.5. Multigrid correction approach

An important aspect of the MG/OPT method is the correction τ_k , which needs to ensure that the solution on the coarse grid approximates the solution of the fine grid. In the current work, we restrict ourselves to Eq. (31). Expressing τ_k explicitly in terms of $\mathcal{J}(f, u(f))$ yields

$$\tau_k = I_{k-1}^k \tau_{k-1} + \nabla \mathcal{J}_k(I_{k-1}^k f_{k-1}^1, u_k(I_{k-1}^k f_{k-1}^1)) - I_{k-1}^k \nabla \mathcal{J}_{k-1}(f_{k-1}^1, u_{k-1}(f_{k-1}^1)), \quad (35)$$

where u_k is the Navier–Stokes solution on grid level k , satisfying $\mathcal{B}_k(u_k(f_k), f_k) = 0$. Note that the restriction operator for u_k may differ from that of f_k . Precise definitions used in the current study are provided in §3.3.

3.2. MG/OPT smoother

Two different methods are considered as MG/OPT smoother. The first is the Polak–Ribière conjugate gradient method, which is one of the methods of choice in the DNS/LES–based optimal control community. It is widely used in solving large-scale optimization problems, as its major advantage is limited storage requirements [58]. The algorithm is simple in implementation [1], and often proves to be efficient when only a few iterations are performed [58]. It makes use of gradient information to construct descent directions in minimization of the cost functional and to improve the control variables. This is done by means of a line-search procedure. To that end, derivative-free line-search methods are usually preferred. In the context of DNS-based optimization with the Polak–Ribière method, the Brent line-search algorithm has been often used, see, e.g., [1, 7, 10]. Details of the algorithm can be found in [59].

A second smoother that we consider, is the damped version of the limited-memory BFGS method. This method was originally designed for constrained optimization problems in which the Hessian matrix is too expensive or impractical to be evaluated. It can be used also for unconstrained optimization problems [14]. The approximation

of the inverse Hessian $\mathbf{H}_l$ is constructed by storing only $2m$ vectors ($\mathbf{s}_l$ and $\mathbf{y}_l$) which define the matrix implicitly, using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) relation:

$$\mathbf{H}_l = \mathbf{W}_m^T \mathbf{H}_l^0 \mathbf{W}_m + \sum_{i=2}^m \frac{1}{\mathbf{y}_{l-i}^T \mathbf{s}_{l-i}} \mathbf{W}_{i-1}^T \mathbf{s}_{l-i} \mathbf{s}_{l-i}^T \mathbf{W}_{i-1} + \frac{1}{\mathbf{y}_{l-1}^T \mathbf{s}_{l-1}} \mathbf{s}_{l-1} \mathbf{s}_{l-1}^T, \quad (36)$$

with

$$\mathbf{W}_i = \prod_{j=1}^i \left[\mathbf{I} - 1 / (\mathbf{y}_{l-j}^T \mathbf{s}_{l-j}) \mathbf{y}_{l-j} \mathbf{s}_{l-j}^T \right]. \quad (37)$$

Note that, here, we switched to the classical, unscaled inner product, in order to be consistent with the classical BFGS formulation found in standard textbooks. For the regular rectangular meshes used in our study, the scaling is constant, and would disappear anyway in Eqs. (36-37). The index l denotes the optimization steps in the single grid optimization algorithm. The advantage of the damped version is that the step length in the line search only needs to satisfy the Armijo condition of sufficient decrease of (Eq. 1):

$$\hat{\mathcal{J}}(\mathbf{f}_l + \alpha \mathbf{d}_l) \leq \hat{\mathcal{J}}(\mathbf{f}_l) + c_1 \alpha \nabla \hat{\mathcal{J}}_l^T \mathbf{d}_l, \quad (38)$$

where the step direction is $\mathbf{d}_l = -\mathbf{H}_l \nabla \hat{\mathcal{J}}_l$. Hence, it requires only one additional function evaluation in every line-search step. For the following numerical experiments we set the constant $c_1 = 10^{-4}$ and $m = 5$. For a complete overview of the methods we refer to [14, 58].

Note that we will also use the damped L-BFGS method as a single-grid optimization reference method, as it emerged in an earlier study as the most efficient method for the test cases considered in the current study [14].

3.3. Implementation details

In this subsection some further implementation details are discussed. In §3.3.1 the practical implementation of the restriction and prolongation operators is detailed. In §3.3.2, the implementation of the MG/OPT correction τ is presented. In §3.3.3 some parallelization issues are discussed, and finally, in §3.3.4 convergence criteria are presented.

3.3.1. Restriction and prolongation operators

In the current work, the control variables $\mathbf{f}_k$ vary in time and in the wall-parallel directions. The inter-grid operators are defined as a Kronecker product between time ($\mathbf{T}$) and space ($\mathbf{S}$) operators, i.e., for the restriction $\mathbf{I}_{k-1}^k = \mathbf{T}_{k-1}^k \otimes \mathbf{S}_{k-1}^k$, and for the prolongation $\mathbf{I}_k^{k-1} = \mathbf{T}_k^{k-1} \otimes \mathbf{S}_k^{k-1}$. For higher-dimensional versions, the space components are obtained by taking Kronecker products of the one-dimensional operators. Note that in all numerical experiments we select a coarsening factor of $n = 2$. In the wall-parallel directions, a Fourier spectral discretization is employed (see §2.2). Therefore, the spatial components of the restriction and prolongation operators for the design variables are performed in the Fourier domain. We define the restriction operator by simply truncating the higher wavenumber modes $\mathbf{p} = (\mathbf{p}_x, \mathbf{p}_y)$ that cannot be represented on coarser grid levels. Thus the restriction operator in space corresponds to

$$\mathbf{S}_{k-1}^k \hat{\mathbf{f}}_{k-1}(\mathbf{p}) = \hat{\mathbf{f}}_{k-1}(\mathbf{p}) \quad \forall \mathbf{p} : 0 \leq |\mathbf{p}_x| \leq N_{x,k}/2 \text{ \& } 0 \leq |\mathbf{p}_y| \leq N_{y,k}/2 \quad (39)$$

where $\hat{\mathbf{f}}_{k-1}$ denotes the vector of Fourier coefficients of the control variables at level $k - 1$, and $N_{x,k}$, $N_{y,k}$ the number of grid points in x , and y direction on grid level k . The prolongation operator is defined by zero padding of the additional elements at high wavenumbers on the finer level. Thus,

$$\mathbf{S}_k^{k-1} \hat{\mathbf{f}}_k(\mathbf{p}) = \begin{cases} \hat{\mathbf{f}}_k(\mathbf{p}), & 0 \leq |\mathbf{p}_x| \leq N_{x,k}/2 \text{ \& } 0 \leq |\mathbf{p}_y| \leq N_{y,k}/2, \\ 0, & \text{otherwise.} \end{cases} \quad (40)$$

Remark that the adjoint property $(\mathbf{S}_k^{k-1} \mathbf{a}, \mathbf{b})_{k-1} = (\mathbf{a}, \mathbf{S}_{k-1}^k \mathbf{b})_k$ for the spatial part of the restriction and prolongation operators is satisfied [51, 60].

The temporal components of the inter-grid transfer operators for the control variables are based on linear interpolation and on a local weighted average. More precisely, for a coarsening factor $n = 2$ and for the variable $\mathbf{f}_k$ with time components $(\varphi_k^0, \varphi_k^1, \varphi_k^2, \dots, \varphi_k^{M_k})$, the prolongation operator from M_k to M_{k-1} elements in time is given by

$$\mathbf{T}_k^{k-1} (\varphi_k^0, \varphi_k^1, \varphi_k^2, \dots, \varphi_k^{M_k}) = \left(\varphi_k^0, \frac{\varphi_k^0 + \varphi_k^1}{2}, \varphi_k^1, \frac{\varphi_k^1 + \varphi_k^2}{2}, \dots, \varphi_k^{M_k} \right). \quad (41)$$

The temporal restriction operator is defined such that the adjoint relation holds, i.e., $(\mathbf{T}_k^{k-1} \mathbf{a}, \mathbf{b})_{k-1} = (\mathbf{a}, \mathbf{T}_{k-1}^k \mathbf{b})_k$. In matrix notation, this means $\mathbf{T}_{k-1}^k = \frac{1}{2}(\mathbf{T}_k^{k-1})^T$. This leads to:

$$\mathbf{T}_{k-1}^k \begin{pmatrix} \varphi_{k-1}^0, \varphi_{k-1}^1, \varphi_{k-1}^2, \dots, \varphi_{k-1}^{M_{k-1}} \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \varphi_{k-1}^0 + \frac{1}{2}\varphi_{k-1}^1, \frac{1}{2}\varphi_{k-1}^1 + \varphi_{k-1}^2 + \frac{1}{2}\varphi_{k-1}^3, \dots, \frac{1}{2}\varphi_{k-1}^{M_{k-1}-1} + \varphi_{k-1}^{M_{k-1}} \end{pmatrix}. \quad (42)$$

Remark that using the inter-grid transfer operators defined above the combined space-time restriction and prolongation satisfy Eq. (23). The precise details of the proof are given in Appendix B.

3.3.2. Restriction of the initial condition

To evaluate the term τ_k (Eq. 32, 35), the controls f_{k-1}^1 are restricted to the coarse level. To evaluate $u_k(I_{k-1}^k, f_{k-1}^1)$, also the boundary conditions of B_k need to be restricted from B_{k-1} . For the boundary conditions in space, this is trivial. For the initial condition, the restriction operators for wall-parallel directions are similar as for the control variables. However, in the wall-normal direction, a staggered fourth-order finite difference scheme is applied and the grid is stretched along the z axis. Here, the state variables of the fine grid at initial condition in time are transported to a coarser grid using a 5-point weighted restriction.

3.3.3. Parallelization issues

The simulation cost scales with n^4 , with n the coarsening factor between grids. On a parallel computing platform, this is only true if all grid levels are run on the same number of processors, and if a perfect strong scaling applies. In practice this will not be the case, and the number of grid levels that can be efficiently employed is limited. SP-Wind is parallelized using MPI, based on a pencil decomposition of its computational domain. The simulations are performed on 32 and 64 compute cores, depending on the case, as presented in the next section. However, it is not our intention, here, to study the effects of reduced parallel efficiency on overall MG/OPT performance. Instead, we perform simulations with 2 and 3 grid levels, always using a V-cycle. The number of processors are selected in such a way that parallel efficiency is roughly constant. Still, we will use real wall clock timings when comparing MG/OPT versus single-grid algorithms, so that real speed-up is measured.

3.3.4. Convergence criteria

In our optimization experiments, the MG/OPT algorithm is stopped when the norm of the gradient on the finest level is below a certain threshold, i.e., 5×10^{-5} , or when the reduction in the cost functional on the finest level stagnates. For the latter, our threshold is a reduction of the cost functional (per iteration) smaller than 10^{-4} . We remark that in optimal control of turbulent flows, the optimization algorithm is typically employed only for a limited number of iterations. These optimization problems are often so expensive, that formal convergence is often not achieved before computational resources are exhausted. Also here, convergence with $\|\nabla \hat{J}\| < 5 \times 10^{-5}$ is not always attained, in which case the algorithm is stopped based on the second criterion.

4. Application to DNS-based optimal control problems

In this section we investigate the performance of the MG/OPT algorithm for DNS-based optimal control of a turbulent channel flow. We consider two optimization cost functionals aiming at maximizing energy extraction and decreasing the turbulent kinetic energy, respectively. The same optimal control problems were used for the evaluation of different gradient-based optimization methods in Ref. [14]. In the current work we compare the efficiency of the MG/OPT technique to single-grid optimization methods in terms of equivalent function and gradient evaluations on the finest level. Furthermore, the influence of the optimization problem properties and algorithmic parameters on efficiency are also examined. A full parameter scan to find the optimal setup of the MG/OPT algorithm is, however, impractical due to the large computational cost associated with these problems. Hence, we present the numerical results for a limited number of algorithmic settings. Details of the different cases and computational set-up are provided in §4.1. Subsequently, results are discussed in §4.2–4.7.

4.1. Description of cases and computational set-up

In the current study, we consider optimal control cases for two different Reynolds numbers, i.e., $Re_\tau = 110$ and $Re_\tau = 395$. Test cases that refer to maximization of extracted energy are referred to with the acronym *ME*, while cases that focus on reduction of turbulent kinetic energy (for drag reduction) are denoted with *TKE*. In the MG/OPT strategy, the simulations are performed using 2 and 3 grid levels in a V-cycle. Details of the simulation set-ups are presented in Table 1 and Table 2 respectively, and are further discussed below. Each entry of the table displays the optimization time horizon, the smoothing steps selected in the V-cycle (γ), together with the control

368 resolution on the levels and the computational cost of one forward evaluation. All simulations for $Re_\tau = 110$
369 are performed on 32 compute cores, on two Ivy Bridge Xeon E5-2680v2 CPU (2.8 GHz, 25 MB cache) nodes.
370 Simulations for $Re_\tau = 395$ are performed on 64 cores, using the same processor type.

Table 1: Summary of simulation cases for $Re_\tau = 110$ with two levels in the V-cycle, where $k = 0$ and $k = 1$ depict the fine and the coarse level. The optimization time horizon, the computational wall clock time for the evaluation of the cost functional ($\mathcal{T}_k$), the control resolutions ($N_{x,k} \times N_{y,k} \times M_k$) and the number of smoothing steps γ_k applied on the $k = 0$ and $k = 1$ levels are also given. H denotes the case when the Hessian initialization strategy (H_{update}) on the fine level is used. Note that the control resolution indicates the number of grid points in streamwise ($N_{x,k}$), spanwise ($N_{y,k}$) and time directions (M_k) at level k .

Case	Time horizon	control resolution		$\mathcal{T}_0$	$\mathcal{T}_1$	$\frac{\mathcal{T}_0}{\mathcal{T}_1}$	γ_0	γ_1	H_{update}
		$k = 0$	$k = 1$						
ME_1	$0.15\delta/u_\tau$	$126 \times 96 \times 500$	$62 \times 48 \times 250$	12'10"	46"	15.9	2	5	–
ME_2	$0.15\delta/u_\tau$	$126 \times 96 \times 500$	$62 \times 48 \times 250$	12'10"	46"	15.9	2	10	–
ME_3	$0.15\delta/u_\tau$	$126 \times 96 \times 500$	$62 \times 48 \times 250$	12'10"	46"	15.9	2	15	–
ME_4	$0.15\delta/u_\tau$	$126 \times 96 \times 500$	$62 \times 48 \times 250$	12'10"	46"	15.9	2	20	–
ME_5	$0.15\delta/u_\tau$	$126 \times 96 \times 500$	$62 \times 48 \times 250$	12'10"	46"	15.9	2	25	–
ME_6	$0.15\delta/u_\tau$	$126 \times 96 \times 500$	$62 \times 48 \times 250$	12'10"	46"	15.9	2	20	H
ME_7	$0.15\delta/u_\tau$	$126 \times 96 \times 500$	$62 \times 48 \times 250$	12'10"	46"	15.9	1	20	H
ME_8	$0.15\delta/u_\tau$	$126 \times 96 \times 500$	$62 \times 48 \times 250$	12'10"	46"	15.9	3	20	H
ME_9	$0.15\delta/u_\tau$	$126 \times 96 \times 500$	$62 \times 48 \times 250$	12'10"	46"	15.9	4	20	H
ME_{10}	$0.03\delta/u_\tau$	$126 \times 96 \times 100$	$62 \times 48 \times 50$	3'48"	14"	15.9	2	20	H
ME_{11}	$0.24\delta/u_\tau$	$126 \times 96 \times 800$	$62 \times 48 \times 400$	17'21"	1'06"	15.8	2	20	H
ME_{12}	$0.15\delta/u_\tau$	$126 \times 96 \times 500$	$62 \times 48 \times 500$	12'10"	1'36"	7.6	2	20	H
ME_{13}	$0.15\delta/u_\tau$	$30 \times 24 \times 500$	$30 \times 24 \times 500$	10'29"	1'24"	7.5	2	20	H
TKE_1	$0.15\delta/u_\tau$	$126 \times 1 \times 500$	$62 \times 1 \times 250$	9'12"	35"	15.8	2	20	H
TKE_2	$0.15\delta/u_\tau$	$30 \times 1 \times 500$	$30 \times 1 \times 250$	7'38"	30"	15.8	2	20	H
TKE_3	$0.24\delta/u_\tau$	$126 \times 1 \times 800$	$62 \times 1 \times 400$	13'24"	51"	15.9	2	20	H

Table 2: Summary of MG/OPT set-up for the case defined at $Re_\tau = 395$. The number of smoothing steps in the V-cycle (γ_k), the computational wall clock time of one functional evaluation ($\mathcal{T}_k$), the ratio $\mathcal{T}_0/\mathcal{T}_k$ and the control resolution ($N_{x,k} \times N_{y,k} \times M_k$) in streamwise ($N_{x,k}$), spanwise ($N_{y,k}$) and time direction ($\times M_k$) for ME_{14} and ME_{15} cases are provided.

Level (k)	γ_k		ME_{14}			ME_{15}		
	2 levels	3 levels	control resolution	$\mathcal{T}_k$	$\frac{\mathcal{T}_0}{\mathcal{T}_k}$	control resolution	$\mathcal{T}_k$	$\frac{\mathcal{T}_0}{\mathcal{T}_k}$
0	2	2	$30 \times 24 \times 100$	7'30"	1	$254 \times 192 \times 300$	19'1"	1
1	20	2	$30 \times 24 \times 100$	59"	7.6	$126 \times 96 \times 150$	1'12"	15.8
2	–	20	$30 \times 24 \times 100$	14"	32.1	$62 \times 48 \times 75$	13"	87.7

371 As already discussed above in §2.2, the fine level grids for $Re_\tau = 110$, and $Re_\tau = 395$ respectively correspond
372 to $128 \times 96 \times 144$, and $256 \times 192 \times 288$. For $Re_\tau = 110$ only one coarser level with resolution $64 \times 48 \times 72$ is used.
373 The through-flow time of the domain corresponds to $0.36\delta/u_\tau$. For $Re_\tau = 395$, two coarse levels are used, i.e.,
374 $128 \times 96 \times 144$ and $64 \times 48 \times 72$ and the through-flow time of the domain is $0.31\delta/u_\tau$. Note that all simulations with
375 the MG/OPT algorithm start at the finest level. Next we provide an overview of the different numerical tests that
376 are performed.
377 In a first set of simulations, we focus on the effect of the smoother on the MG/OPT performance. To this

Figure 5: Evolution of the cost functional ($\Delta J = J - J_{final}$) as a function of the computational *cost* for cases ME_1 - ME_5 . Methods are the single-grid (Δ) damped L-BFGS method (left) and non-linear conjugate gradient (NCG) method (right), and MG/OPT with damped quasi-Newton (left) or NCG (right) as a smoother for ME_1 ($\diamond$), ME_2 ($\times$), ME_3 ($+$), ME_4 ($\bullet$) and ME_5 ($\square$). Symbols are employed to differentiate lines.

end, we compare the non-linear conjugate gradient Polak–Ribière method and the damped L-BFGS method using cases ME_1 - ME_5 . The difference between these simulations consists of the number of smoothing steps chosen on coarser levels γ_1 . Results are presented in §4.2. Subsequently, in §4.3, we concentrate on the effect of the Hessian initialization for the damped L-BFGS smoother. In that section, the impact of the smoothing steps on the finest levels is also discussed based on cases ME_4 , ME_6 - ME_9 . Next, a number of algorithmic variations are investigated. First the effect of the optimal-control time horizon on the algorithm efficiency is studied in §4.4, using ME_{10} and ME_{11} . Subsequently, we investigate the effectiveness of MG/OPT for the reduction of kinetic energy, showing results for cases TKE_1 , TKE_2 and TKE_3 in §4.5. In §4.6 the effect of the coarsening operator is studied using cases ME_{12} and ME_{13} . Finally, the effect of the number of levels and the full multigrid initialization using the ME_{14} and ME_{15} for $Re_\tau = 395$ is presented in §4.7.

In the following sections, all timings are expressed in terms of equivalent function and gradient evaluations on the finest level ($k = 0$). Thus

$$\text{cost} = (F_0 + A_0) + \sum_{k=1}^n (F_k + A_k) \left(\frac{\mathcal{T}_0}{\mathcal{T}_k} \right)^{-1} \quad (43)$$

with F_k and A_k the number of forward and adjoint simulations on level k . Note that a forward and adjoint equation solution takes roughly the same time. The values $\mathcal{T}_k$ represent the average simulation wall clock time for one functional evaluation on level k . Timings are provided in Table 1 and Table 2. The computational results of classical multigrid approach applied to system of equations are usually reported based on the norm of the residual. In an optimization framework, it is often more relevant to report the evolution of the objective functional [32, 33]. In the current study, the decrease in the cost functional is expressed with respect to the constant value J_{final} which is the same for all optimization algorithms.

4.2. NCG and damped L-BFGS smoother

Figure 5 shows results for case ME_1 , optimized with the single-grid damped L-BFGS and non-linear conjugate gradient (NCG) methods. The figure also displays the MG/OPT results for cases ME_1 - ME_5 , in combination with NCG and damped L-BFGS as a smoother which are further discussed below in §4.3. The MG/OPT strategy using the NCG Polak–Ribière method in combination with the Brent line-search approach clearly accelerates the single-grid NCG method. The best case reported here shows an acceleration by a factor of 3.7. Furthermore, we observe that MG/OPT is more effective in accelerating the underlying NCG method than in accelerating the damped L-BFGS method. For the latter only a factor of 1.8 is achieved for the ME_4 case. Nevertheless the MG/OPT-NCG combination is outperformed by both single grid and MG/OPT damped L-BFGS methods. Therefore, we select

Figure 6: Evolution of the cost functional ($\Delta J = J - J_{final}$) as function of equivalent fine grid evaluations (*cost*) for the MG/OPT algorithm (dashed line) in ME_4 (•), ME_6 (*), ME_7 (◊), ME_8 (×), ME_9 (◻) cases and the corresponding single-grid damped L-BFGS method (Δ). Symbols are employed to differentiate lines.

the damped L-BFGS smoother for further testing the effect of different algorithmic parameters in the MG/OPT strategy for DNS-based optimal control applications.

4.3. Selection of number of steps on each level and Hessian update for damped L-BFGS method

To study the influence of the accuracy of the coarse grid solver on the MG/OPT efficiency, cases ME_1 - ME_5 (Table 1) have been defined with different number of coarse grid smoothing steps γ_1 . The results are shown in Figure 5. First of all, we observe that MG/OPT with an increased number of smoother iterations on the coarse level, γ_1 , performs better than MG/OPT with less γ_1 optimization steps. Here, the computational effort is moved to the coarse level which provides less expensive search directions. However, when the optimization steps on coarser level are increased to 25 (case ME_5), we observe that the MG/OPT effectiveness is clearly outperformed by ME_3 and ME_4 ($\gamma_1 = 15$ and 20). This is not the case for the non-linear conjugate gradient smoother. Here we observe the same performance of MG/OPT for the ME_3 - ME_5 cases.

Another aspect that we investigate is the initialization of the damped quasi-Newton smoother in the MG/OPT algorithm. In the previous results, the damped L-BFGS smoother is initialized with $H_k^0 = I$ in every upward branch of the V-cycle as suggested by the construction of the MG/OPT framework. This can lead to poor scaling of the search direction in the first smoothing step. To study the influence of the smoother initialization, in case ME_6 in Table 1, we store the fine level vector pairs used to update the approximation of the inverse Hessian, i.e., Eq. (36), during the V-cycles. These are used to estimate the initial search direction each time a smoother is used on the fine level. Figure 6 shows the decrease in the fine cost functional as a function of the number of equivalent fine evaluations. In the initial stages of the optimization no further saving in the total cost is observed compared to case ME_4 . However, in later stages, a better approximation of the initial inverse Hessian accelerates the convergence of MG/OPT, with 10% less equivalent fine evaluations.

Next to that we investigate also the effect of fine smoothing steps on the MG/OPT performance. We define the ME_7 - ME_9 cases by varying only the number of γ_0 iterations and using the inverse Hessian update strategy. These results are also shown in Figure 6. First of all, we observe that in terms of equivalent fine evaluation required, the MG/OPT algorithm applied to case ME_6 performs best. The case reported here shows an acceleration of damped L-BFGS by a factor of 2. When comparing to MG/OPT applied to ME_8 (with $\gamma_0 = 3$) and ME_9 (with $\gamma_0 = 4$) a factor of 1.4 is obtained. Finally, MG/OPT applied to ME_7 (with $\gamma_0 = 1$) is less effective in accelerating the underlying algorithm. In summary, we obtained the best results for a fine level smoothing number $\gamma_0 = 2$, a coarse level solve with $\gamma_1 = 20$ iterations, and using the inverse Hessian update strategy.

4.4. Effect of the optimization time horizon

To study the impact of the optimization time horizon on the MG/OPT performance, we select two time intervals, i.e., a relatively short time horizon $T = 0.03\delta/u_\tau$ and a long time horizon $T = 0.24\delta/u_\tau$ corresponding

Figure 7: (left) study of the effect of the horizon length on the MG/OPT convergence for the ME_{10} ($\circ$) and ME_{11} ($\blacksquare$) cases; (right) study of the effect of the cost functional on the MG/OPT convergence ($\circ$) for the TKE_1 (solid line), TKE_2 (dot dashed line) and TKE_3 (dashed line) cases. In both studies the corresponding single-grid damped L-BFGS method results are illustrated by a triangle symbol (Δ). Symbols are employed to differentiate lines.

to cases ME_{10} and ME_{11} (see Table 1). Figure 7 (left) shows the optimization results. The figure demonstrates a significant improvement for MG/OPT on both cases. Moreover, we observe that in terms of goal function improvement, the multilevel approach achieves more reduction than the single-grid method. It is found that for a shorter time interval (ME_{10}), the MG/OPT method requires 53% less fine evaluations than the single-grid damped L-BFGS method. Also, MG/OPT appears to be effective in accelerating the underlying algorithm for a longer optimization time horizon. The results show that MG/OPT for the ME_{11} case achieves the same improvement in the cost functional with 33% fewer evaluations than the single level reference case.

4.5. Effect of the cost functional

In Figure 7 (left) we demonstrate the speed-ups obtained with MG/OPT for the ME cost functional. In Figure 7 (right) we illustrate its use (and failure) for the TKE cost functional. For the latter case, and over the full convergence history, the MG/OPT algorithm appears to require more equivalent fine forward and adjoint evaluations to reach the same decrease in the cost functional as the single-grid damped L-BFGS method. This may be due to a lack of consistency between the coarse and fine PDE discretization. The latter was also identified as one of the causes for MG/OPT not to work well in Ref. [31]. In low Reynolds number turbulent channel flows it is known that most kinetic energy is situated in the buffer region close to the wall, with dominant length scales that are small and that are lost on the coarse grid level. Looking in more detail at the convergence history, we do find that MG/OPT outperforms single-grid optimization in the initial stages (up to 50%). In later stages, this advantage is lost. For the power maximization cost functional we observe that the search direction computed on the coarse level is a descent direction and typically in almost all the V-cycles performed is well scaled, i.e., $\alpha = 1$ is successful. For the TKE cost functional, when we examine the step value α from the coarse grid correction stage of the MG/OPT algorithm, we notice that only in the first three V-cycles $\alpha = 1$. Furthermore, for this case, the search direction from the upward branch of the V-cycle attains almost no decrease in the cost functional on the fine level in later stages. This suggests that the coarse optimization problem is not a good approximation of the fine level problem at later iterations. Thus, the computational effort to find a search direction on the low resolution problem is mainly wasted. We tested different configurations and coarsening strategies, but did not manage to find a case for which MG/OPT works with the TKE functional. Furthermore, in a variant of the optimal-control of terminal TKE , we add a penalty term to the objective functional that corresponds to a cost of the control. We observe that the optimization problem becomes more challenging as the weighting factor of the penalty term goes to zero. With respect to the relative performance of MG/OPT versus the single grid algorithm we do not find any major changes, and the best algorithm remains the same. For the sake of brevity, these results are not further reported here.

Figure 8: Evolution of the cost functional ($\Delta J = J - J_{final}$) as a function of the computational *cost* for the damped L-BFGS method (Δ) and the MG/OPT algorithm applied to ME_6 ($*$) and ME_{10} ($\circ$) (on the left) and ME_{11} ($\square$) (on the right). Symbols are employed to differentiate lines.

4.6. Effect of semi-coarsening

Next, we test a less aggressive coarsening strategy. To this end, we define the cases ME_{10} and ME_{11} from Table 1. Case ME_{10} represents a test in which the only coarsening happens in space. The coarse grid simulations are advanced in time using the same time steps as the fine grid discretization. In case ME_{11} , we use the same settings in time but now we coarsen only the state variables in space. Thus, for this case, the control force is defined on a limited number of modes that can be represented on both grids without need for coarsening.

Figure 8 shows the decrease in the cost functional as a function of the fine grid equivalent computational cost. The results show that MG/OPT for case ME_{10} requires 33% fewer evaluations than single-grid damped L-BFGS. The corresponding number is 50% for MG/OPT when the full coarsening strategy is chosen (see case ME_6). These results indicate that time coarsening is effective. Figure 8 shows also the results when no coarsening is applied at the control level. We observe a computational cost reduction of about 58%. Without control-coarsening, the computational effort to solve the coarse level problems is increased (see ratio $\mathcal{T}_0/\mathcal{T}_1$ from Table 1). However, the fine optimization variables are better approximated by the coarse model, so that the MG/OPT strategy is more effective for this case.

4.7. Effect of the number of levels and the full multigrid initialization

Next, we investigate the effect of the number of levels in MG/OPT. We select the best algorithmic settings found in §4.2-§4.6 and apply the MG/OPT algorithm to the energy extraction maximization case at $Re_\tau = 395$. Table 2 presents the details of the simulation set-ups. Case ME_{14} denotes a fixed size of the control distribution with an optimization time horizon of $0.03\delta/u_\tau$ (100 time steps). The control is defined only in the first 30 and 24 Fourier modes in the x and y directions. The control is represented on all levels in the hierarchy with the same accuracy. The temporal resolution is the same on all grids in the V-cycle. This leads to a fixed number of degrees of freedom in the control space $O_k = 7.2 \times 10^4$ ($k = 0, 1, 2$) after discretization. Case ME_{15} represents a full coarsening study with a longer time horizon, i.e., $0.09\delta/u_\tau$.

Figure 9 illustrates the results for both cases with 2 and 3 levels in the V-cycle. The 3-level MG/OPT for case ME_{14} costs about 50% less than the 2 level MG/OPT. Compared to single-grid optimization, the 3-level MG/OPT method is 4 times less expensive. For a relatively longer time horizon and a full coarsening strategy, i.e., case ME_{15} , the 3-level MG/OPT is 1.3 times cheaper than 2-level MG/OPT. The 3-level case accelerates single-level damped L-BFGS by a factor of 2.2.

Up till now, we have initialized the MG/OPT optimization with the same specific estimate on the fine optimization level as for the single-grid optimization algorithm. In practice, in classical multigrid methods, an alternative approach is to provide a good initial approximation obtained from a coarse level problem, e.g., the full multigrid algorithm [61–63]. Next, we investigate the MG/OPT method initialized using the full-multigrid approach for the

Figure 9: Study of the effect of the number of levels and the full multigrid initialization on MG/OPT performance for the ME_{14} (solid lines) and ME_{15} (dashed line). The single grid damped L-BFGS results are indicated by (Δ) whereas the MG/OPT results with 2 levels in the V-cycle by ($\blacksquare$), with 3 levels by ($\circ$) and with 2 levels and full multigrid initialization by ($\bullet$). Symbols are employed to differentiate lines.

ME_{14} and ME_{15} cases. To estimate an initial guess for the fine problem, the single-grid damped L-BFGS algorithm on coarse level problem is terminated when the norm of the reduced gradient is below to a certain threshold, e.g. 5×10^{-5} . Subsequently, this coarse solution is interpolated to the fine level.

Figure 9 shows the results when a full multigrid initialization technique is used for the 2-level configuration. We find that the MG/OPT algorithm with full initialization technique for case ME_{14} requires 65% fewer equivalent evaluations than the single-grid algorithm. The approach accelerates the 2-level MG/OPT by a factor of 1.4. For case ME_{15} , we find a factor of 1.72. In this case, the same cost functional decrease is achieved with an equivalent of only 44 evaluations on the finest level. When comparing the full multigrid initialization MG/OPT with single-grid damped L-BFGS, a factor of 3 is achieved.

5. Conclusion

This study set out with the aim of assessing the effectiveness of MG/OPT to accelerate conventional single-grid optimization algorithms for DNS-based optimal control. In MG/OPT, the descent directions are predicted by using lower fidelity resolutions of the optimization problem. The definition of the coarse cost functional is altered by a term involving the current gradient of the fine model.

The multilevel strategy was tested for unconstrained optimization of turbulent wall-bounded flows. The objective functionals aimed at a decrease of the turbulent kinetic energy or an increase of the total extracted energy by internal distributed forces. We estimated the cost functional gradient by means of a continuous adjoint approach. The smoother in the MG/OPT hierarchy was the damped limited-memory BFGS method. Moreover, as a point of reference, the non-linear conjugate-gradient method was also investigated as a smoother in the MG/OPT algorithm. We compared the performance of the multilevel algorithm with the single-grid optimization methods in terms of equivalent function and gradient evaluations on the finest level. The results for the energy extraction cost functional indicated an acceleration of the smoother, in some cases, by a factor of 4. It is interesting to note that for the optimization time horizons tested here, the results of all MG/OPT variants showed a saving in the number of equivalent fine evaluations. For the drag reduction problem, in which the cost functional was represented by the turbulent kinetic energy, MG/OPT was outperformed by the single-grid optimization method. A possible explanation for this might be that the coarse model does not adequately resolve the small turbulent structures close to the wall, that have an important effect on the cost functional.

The findings further support the idea that the efficiency of MG/OPT is influenced by a number of factors [56], e.g., the nature of the objective functional, the selection of the underlying optimization algorithm (the smoother), the algorithmic parameters and the model equations. We observed that MG/OPT for the energy extraction cost functional was more effective in accelerating the non-linear conjugate gradient method than the damped L-BFGS

method. This is related to the properties of the underlying optimization algorithm, as the non-linear conjugate-gradient method is already quite efficient when only a few iterations are affordable. Thus, MG/OPT is more effective than the single-grid NCG method. However, when comparing to the damped L-BFGS method, the combination MG/OPT-NCG is clearly outperformed by MG/OPT-damped L-BFGS.

We also showed that the MG/OPT performance with a damped L-BFGS smoother is substantially improved by increasing the accuracy of the coarse level solve and by increasing the number of levels in the V-cycle. However, in a continued coarsening strategy, caution must be applied, as the DNS with coarse resolutions represents a poor approximation to the solution, affecting the efficiency of the algorithm. With respect to the results for different variants of the intergrid transfer operators, we observe that MG/OPT with no coarsening at control level outperforms the other cases.

In the current work, we have showed that the MG/OPT algorithm is a viable tool to reduce the cost associated with the evaluations required for cost functional improvement in high-fidelity simulations. However, more research is required to assess the impact of unresolved scales from coarse levels on the effectiveness of multi-grid optimization. In the drag reduction optimization problem, the MG/OPT algorithm accelerates the single-grid smoother only in the first few iterations. As the optimization process evolves, we observe that the coarse grid model does not provide a good approximation of the fine level solution any more. This leads to an inconsistency between the models in the MG/OPT hierarchy. A reasonable approach to tackle this issue could be to employ a two-stage strategy where MG/OPT is used in the initial stages, followed by a switch to single-grid optimization when the algorithm approaches the optimum.

Acknowledgements

The authors acknowledge support from the Flemish Science Foundation (FWO, grant number G0C3115N), and from OPTEC (OPTimization in Engineering Center of Excellence, KU Leuven), which is funded by the KU Leuven Research Council under grant no PFV/10/002. Simulations were performed on the computing infrastructure of the VSC Flemish Supercomputer Center, funded by the FWO and the Flemish Government.

Appendix A. Energy extraction cost functional

We briefly analyse the formulation of power extraction from a fully developed turbulent channel flow by a volume force. In order for the optimization problem related to the cost function in Eq. (17) to be well posed, it is important that $\hat{\mathcal{J}}(\mathbf{f})$ is bounded from below, or equivalently the maximum amount of energy that can be extracted from the flow is bounded from above. In case of Eq. (17), this is not directly straightforward, since $\mathcal{J}(\mathbf{f}, \mathbf{u})$ itself is not bounded, so that we have to invoke extra arguments related to the solution $\mathbf{u}(\mathbf{f})$. To this end, we can multiply the Navier–Stokes equations with $\mathbf{u}$ to obtain an evolution equation for the kinetic energy (see, e.g., Ref [64], for classical derivations of the energy equation). Integrated over the computational domain and over the time horizon, using integration by parts, and taking into account the boundary conditions, this yields following energy balance

$$\int_0^T \int_{\Omega} \frac{de}{dt} d\mathbf{x} dt + \frac{2}{Re} \int_0^T \int_{\Omega} \mathbf{S} : \mathbf{S} d\mathbf{x} dt - f_{\infty} |\Omega| \int_0^T U(t) dt = \int_0^T \int_{\Omega} \mathbf{f} \cdot \mathbf{u} d\mathbf{x} dt, \quad (44)$$

with $e = \mathbf{u} \cdot \mathbf{u}/2$, and $\mathbf{S} = (\nabla \mathbf{u} + (\nabla \mathbf{u})^T)/2$, and where the mean streamwise velocity $U = (1/|\Omega|) \int_{\Omega} \mathbf{u} \cdot \mathbf{e}_1 d\mathbf{x}$. Since $\mathbf{S} : \mathbf{S}$, and $e(t)$ are strictly positive, we find

$$\hat{\mathcal{J}}(\mathbf{f}) = \int_0^T \int_{\Omega} \mathbf{f} \cdot \mathbf{u} d\mathbf{x} dt \geq - \int_{\Omega} e(\mathbf{x}, 0) d\mathbf{x} - f_{\infty} |\Omega| \int_0^T U(t) dt, \quad (45)$$

with $e(\mathbf{x}, 0)$ the initial kinetic energy in the domain. Thus $\hat{\mathcal{J}}(\mathbf{f})$ is bounded from below provided $U(t)$ is bounded from above.

In case $f_1 \geq 0$, we can straightforwardly find an upper bound on the mean streamwise velocity U in the domain. To this end, we integrate the streamwise momentum equation over the domain. Using the boundary conditions, and integration by parts, this leads to

$$|\Omega| \frac{dU}{dt} = (f_{\infty} - f_1) |\Omega| - \tau_w^+ |\Gamma_3^+| - \tau_w^- |\Gamma_3^-|, \quad (46)$$

where τ_w^+ and τ_w^- are the averaged wall stresses at the upper and lower wall, i.e.

$$\tau_w^- \equiv - \frac{1}{Re |\Gamma_3^-|} \int_{\Gamma_3^-} (\nabla \mathbf{u}|_{\Gamma_3^-} \cdot \mathbf{n}) \cdot \mathbf{e}_1 d\mathbf{x} = \frac{1}{Re |\Gamma_3^-|} \int_{\Gamma_3^-} \frac{\partial u_1}{\partial x_3} \Big|_{\Gamma_3^-} d\mathbf{x}, \quad (47)$$

573 and a similar definition holds for τ_w^+ .

574 Since we are looking for an upper bound on $U(t)$, we can safely limit ourselves to situations for which $\mathbf{u} \cdot \mathbf{e}_1 \geq 0$.
 575 Together with the no-slip boundary conditions, this always yields $\tau_w^+ \geq 0$, and $\tau_w^- \geq 0$. Thus,

$$U(t) = U(0) + \int_0^t \frac{dU}{dt'} dt' \leq U(0) + f_\infty |\Omega| t, \quad (48)$$

576 with $U(0)$ the initial mean flow. Note that the above derivation assumes uniqueness and existence of solutions
 577 of the Navier–Stokes equations, which is often assumed to hold given well-posed boundary conditions [65, 66].
 578 However, to date this is still an important unsolved problem in mathematics.

579 The condition $f_1 \geq 0$ may not be necessary to find a lower bound for $\hat{\mathcal{J}}(f)$, but a rigorous proof is much
 580 less obvious. When f_1 is allowed to go to $-\infty$ at some places and instances in time, Eq. (48) is not a valid upper
 581 bound anymore for $U(t)$. In that case, an approach can be to use a scaling argument for $\mathbf{S} : \mathbf{S}$ in (44). From a
 582 physical point of view, we expect that addition of energy to the flow (by negative f) cannot be recuperated later or
 583 at other locations (by positive f) because of the effect of dissipation. Moreover, we expect dissipation to increase
 584 more than linearly with velocity (given the no-slip boundary conditions at the wall). For instance, in laminar flow,
 585 typically $(2/Re) \int_\Omega \mathbf{S} : \mathbf{S} d\mathbf{x} \sim |U|^2$, while in turbulent flow $\sim |U|^3$ is usually found. A further derivation is not
 586 in the scope of the current manuscript. Finally, we note that, although we do not formally impose $f_1 \geq 0$ in our
 587 optimization problem, we never encountered situations with $f_1 \leq 0$ during our numerical experiments, suggesting
 588 that the upper bound for the case $f_1 \geq 0$ may in fact also be valid for the more general case.

589 Appendix B. The space-time inter-grid transfer operators

590 We show that the combined space-time transfer operators defined in §3.3.1 satisfy the adjoint identity, Eq. (23).
 591 To that end, we consider the TKE minimization problem, where the control force varies with time and with one
 592 direction in space. For a coarse grid k , we denote with N the number of elements in space and with M the time
 593 elements. For a coarsening factor 2, the prolongation operator in Fourier space (Eq. 40) from N to $2N$ elements is
 594 written as the following $2N \times N$ matrix:

$$S_k^{k-1} = \begin{bmatrix} 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & 0 \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 \end{bmatrix} \quad (49)$$

595 Furthermore, the restriction operator in space (Eq. 39) has size $N \times 2N$ and is given by:

$$S_{k-1}^k = \begin{bmatrix} 0 & \dots & 0 & 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & 1 & \dots & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & 0 & 0 & \dots & 1 & 0 & \dots & 0 \end{bmatrix} \quad (50)$$

596 The time prolongation and restriction operators are defined in Eqs. (41-42) and correspond to the matrices:

$$T_k^{k-1} = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \dots & \dots & 0 & 1 \end{bmatrix}; \quad T_{k-1}^k = \frac{1}{2} \begin{bmatrix} 1 & \frac{1}{2} & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{2} & 1 & \frac{1}{2} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \dots & \frac{1}{2} & 0 \\ 0 & \dots & \dots & \dots & \frac{1}{2} & 1 \end{bmatrix} \quad (51)$$

597 with sizes $2M \times M$ and $M \times 2M$.

$$I_k^{k-1} = T_k^{k-1} \otimes S_k^{k-1} = \begin{bmatrix} 1 \cdot S_k^{k-1} & 0 & \dots & 0 & 0 \\ \frac{1}{2} \cdot S_k^{k-1} & \frac{1}{2} \cdot S_k^{k-1} & \dots & 0 & 0 \\ 0 & 1 \cdot S_k^{k-1} & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & \frac{1}{2} \cdot S_k^{k-1} & \frac{1}{2} \cdot S_k^{k-1} \\ 0 & \dots & \dots & 0 & 1 \cdot S_k^{k-1} \end{bmatrix}. \quad (52)$$

The space-time restriction operator is defined as:

$$I_{k-1}^k = T_{k-1}^k \otimes S_{k-1}^k = \frac{1}{2} \begin{bmatrix} 1 \cdot S_{k-1}^k & \frac{1}{2} \cdot S_{k-1}^k & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{2} \cdot S_{k-1}^k & 1 \cdot S_{k-1}^k & \frac{1}{2} \cdot S_{k-1}^k & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \dots & \frac{1}{2} \cdot S_{k-1}^k & 0 \\ 0 & \dots & \dots & \dots & \frac{1}{2} \cdot S_{k-1}^k & 1 \cdot S_{k-1}^k \end{bmatrix} \quad (53)$$

Remark that the adjoint identity holds as the restriction, I_{k-1}^k , corresponds to the transpose of the prolongation matrix, I_k^{k-1} , up to a constant factor:

$$(I_k^{k-1})^T = (T_k^{k-1} \otimes S_k^{k-1})^T = (T_k^{k-1})^T \otimes (S_k^{k-1})^T = 2T_{k-1}^k \otimes S_{k-1}^k = 2I_{k-1}^k. \quad (54)$$

- [1] T. R. Bewley, P. Moin, R. Temam, DNS-based predictive control of turbulence: an optimal benchmark for feedback algorithms, *Journal of Fluid Mechanics* 447 (2001) 179–225.
- [2] M. D. Gunzburger, *Perspectives in flow control and optimization*, Vol. 5, SIAM, Philadelphia, 2003.
- [3] M. Wei, J. B. Freund, A noise-controlled free shear flow, *Journal of Fluid Mechanics* 546 (2006) 123–152.
- [4] D. Thevenin, G. Janiga, *Optimization and Computational Fluid Dynamics*, Springer Berlin Heidelberg, 2008.
- [5] D. Marinc, H. Foyi, Investigation of a continuous adjoint-based optimization procedure for aeroacoustic control of plane jets, *International Journal of Heat and Fluid Flow* 38 (2012) 200–212.
- [6] J. Kim, D. J. Bodony, J. B. Freund, Adjoint-based control of loud events in a turbulent jet, *Journal of Fluid Mechanics* 741 (2014) 28–59.
- [7] J. P. Goit, J. Meyers, Optimal control of energy extraction in wind-farm boundary layers, *Journal of Fluid Mechanics* 768 (2015) 5–50.
- [8] J. P. Goit, W. Munters, J. Meyers, Optimal coordinated control of power extraction in LES of a wind farm with entrance effects, *Energies* 9 (1) (2016) 29.
- [9] S. S. Collis, Y. Chang, S. Kellogg, R. Prabhu, Large eddy simulation and turbulence control, *AIAA paper* 2564 (2000) 2000.
- [10] S. Delpont, M. Baelmans, J. Meyers, Maximizing dissipation in a turbulent shear flow by optimal control of its initial state, *Physics of Fluids* 23 (4) (2011) 045105.
- [11] H. Badreddine, S. Vandewalle, J. Meyers, Sequential Quadratic Programming (SQP) for optimal control in direct numerical simulation of turbulent flow, *Journal of Computational Physics* 256 (2014) 1–16.
- [12] J. B. Freund, Adjoint-based optimization for understanding and suppressing jet noise, *Journal of Sound and Vibration* 330 (17) (2011) 4114–4122.
- [13] W. Munters, J. Meyers, An optimal control framework for dynamic induction control of wind farms and their interaction with the atmospheric boundary layer, *Phil. Trans. R. Soc. A* 375 (2017) 20160100.
- [14] C. Nita, S. Vandewalle, J. Meyers, On the efficiency of gradient based optimization algorithms for DNS-based optimal control in a turbulent channel flow, *Computers & Fluids* 125 (2016) 11–24.
- [15] S. G. Nash, A multigrid approach to discretized optimization problems, *Optimization Methods and Software* 14 (1-2) (2000) 99–116.
- [16] D. C. Liu, J. Nocedal, On the limited memory BFGS method for large scale optimization, *Mathematical programming* 45 (1-3) (1989) 503–528.
- [17] S. G. Nash, J. Nocedal, A numerical study of the limited memory BFGS method and the truncated-Newton method for large scale optimization, *SIAM Journal on Optimization* 1 (3) (1991) 358–372.
- [18] A. Borzi, V. Schulz, *Computational optimization of systems governed by partial differential equations*, SIAM, Philadelphia, 2012.
- [19] M. Hinze, M. Köster, S. Turek, Space-time Newton-multigrid strategies for nonstationary distributed and boundary flow control problems, in: *Trends in PDE Constrained Optimization*, Springer, 2014, pp. 383–401.
- [20] A. Borzi, V. Schulz, Multigrid methods for PDE optimization, *SIAM review* 51 (2) (2009) 361–395.
- [21] R. M. Lewis, S. G. Nash, Using inexact gradients in a multilevel optimization algorithm, *Computational Optimization and Applications* 56 (1) (2013) 39–61.
- [22] V. H. Schulz, Solving discretized optimization problems by partially reduced SQP methods, *Computing and visualization in science* 1 (2) (1998) 83–96.
- [23] T. Dreyer, B. Maar, V. Schulz, Multigrid optimization in applications, *Journal of Computational and Applied Mathematics* 120 (1) (2000) 67–84.
- [24] S. Ta’asan, Pseudo-time methods for constrained optimization problems governed by PDE, *ICASE Report* 94-52.
- [25] M. Vallejos, A. Borzi, Multigrid optimization methods for linear and bilinear elliptic optimal control problems, *Computing* 82 (1) (2008) 31–52.
- [26] S. B. Hazra, Multigrid one-shot method for PDE-constrained optimization problems, *Applied Mathematics* 3 (2012) 1565–1571.
- [27] F. Tröltzsch, *Optimal control of partial differential equations: theory, methods, and applications*, Grad. Stud. Math. 112, American Mathematical Society, Providence, RI, 2010.
- [28] F. Abergel, R. Temam, On some control problems in fluid mechanics, *Theoretical and Computational Fluid Dynamics* 1 (6) (1990) 303–325.

- [29] A. Jameson, L. Martinelli, N. Pierce, Optimum aerodynamic design using the Navier–Stokes equations, *Theoretical and computational fluid dynamics* 10 (1-4) (1998) 213–237.
- [30] H. Choi, M. Hinze, K. Kunisch, Instantaneous control of backward-facing step flows, *Applied Numerical Mathematics* 31 (2) (1999) 133–158.
- [31] S. G. Nash, R. M. Lewis, Assessing the performance of an optimization-based multilevel method, *Optimization Methods and Software* 26 (4-5) (2011) 693–717.
- [32] S. G. Nash, Properties of a class of multilevel optimization algorithms for equality-constrained problems, *Optimization Methods and Software* 29 (1) (2014) 137–159.
- [33] R. M. Lewis, S. G. Nash, Model problems for the multigrid optimization of systems governed by differential equations, *SIAM Journal on Scientific Computing* 26 (6) (2005) 1811–1837.
- [34] M. P. Rumpfkeil, D. J. Mavriplis, Optimization-based multigrid applied to aerodynamic shape design, Tech. Rep., Department of Mechanical Engineering, University of Wyoming, Laramie, WY.
- [35] M. Vallejos, MGOPT with gradient projection method for solving bilinear elliptic optimal control problems, *Computing* 87 (1-2) (2010) 21–33.
- [36] A. Borzi, U. Hohenester, Multigrid optimization schemes for solving Bose-Einstein condensate control problems, *SIAM Journal on Scientific Computing* 30 (1) (2008) 441–462.
- [37] A. Borzi, E.-J. Park, M. V. Lass, Multigrid optimization methods for the optimal control of convection–diffusion problems with bilinear control, *Journal of Optimization Theory and Applications* 168 (2) (2016) 510–533.
- [38] M. Hinze, R. Pinnau, M. Ulbrich, S. Ulbrich, *Optimization with PDE constraints*, Vol. 23, Springer Science & Business Media, 2008.
- [39] M. Gunzburger, Adjoint equation-based methods for control problems in incompressible, viscous flows, *Flow, Turbulence and Combustion* 65 (3-4) (2000) 249–272.
- [40] T. Min, S. M. Kang, J. L. Speyer, J. Kim, Sustained sub-laminar drag in a fully developed channel flow, *Journal of Fluid Mechanics* 558 (2006) 309–318.
- [41] H. Mamori, K. Fukagata, Drag reduction by streamwise traveling wave-like Lorenz Force in channel flow, in: *Journal of Physics, Conference Series*, Vol. 318, IOP Publishing, 2011, p. 022030.
- [42] J. Xu, S. Dong, M. R. Maxey, G. E. Karniadakis, Turbulent drag reduction by constant near-wall forcing, *Journal of Fluid Mechanics* 582 (2007) 79–101.
- [43] H. Mamori, K. Fukagata, Drag reduction effect by a wave-like wall-normal body force in a turbulent channel flow, *Physics of Fluids* (1994-present) 26 (11) (2014) 115104.
- [44] B. Protas, T. R. Bewley, G. Hagen, A computational framework for the regularization of adjoint analysis in multiscale PDE systems, *Journal of Computational Physics* 195 (1) (2004) 49–89.
- [45] M. Desai, K. Ito, Optimal controls of Navier–Stokes equations, *SIAM Journal on Control and Optimization* 32 (5) (1994) 1428–1446.
- [46] A. V. Fursikov, M. D. Gunzburger, L. Hou, Boundary value problems and optimal boundary control for the Navier–Stokes system: the two-dimensional case, *SIAM Journal on Control and Optimization* 36 (3) (1998) 852–894.
- [47] K. Iwamoto, Y. Suzuki, N. Kasagi, Reynolds number effect on wall turbulence: toward effective feedback control, *International journal of heat and fluid flow* 23 (5) (2002) 678–689.
- [48] R. D. Moser, J. Kim, N. N. Mansour, Direct numerical simulation of turbulent channel flow up to $Re = 590$, *Phys. Fluids* 11 (4) (1999) 943–945.
- [49] S. Delport, M. Baelmans, J. Meyers, Constrained optimization of turbulent mixing-layer evolution, *Journal of Turbulence* 10 (18) (2009) 1–26.
- [50] R. Verstappen, A. Veldman, Symmetry-preserving discretization of turbulent flow, *Journal of Computational Physics* 187 (1) (2003) 343–368.
- [51] C. Canuto, M. Y. Hussaini, A. Quarteroni, T. A. Zang, *Spectral Methods, Fundamentals in Single Domains*, Springer Berlin Heidelberg, 2006.
- [52] J. C. Del Alamo, J. Jiménez, P. Zandonade, R. D. Moser, Scaling of the energy spectra of turbulent channels, *Journal of Fluid Mechanics* 500 (2004) 135–144.
- [53] J. Jimenez, J. C. Del Alamo, O. Flores, The large-scale dynamics of near-wall turbulence, *Journal of Fluid Mechanics* 505 (2004) 179–199.
- [54] A. Borzi, On the convergence of the MG/OPT method, *PAMM* 5 (1) (2005) 735–736.
- [55] S. G. Nash, Convergence and descent properties for a class of multilevel optimization algorithms (2010).
- [56] R. M. Lewis, S. G. Nash, Factors affecting the performance of optimization-based multigrid methods, in: *Multiscale Optimization Methods and Applications*, Springer, 2006, pp. 151–172.
- [57] M. J. Powell, A fast algorithm for nonlinearly constrained optimization calculations, in: *Numerical analysis*, Springer Berlin Heidelberg, 1978, pp. 144–157.
- [58] J. Nocedal, S. J. Wright, *Numerical optimization*, Springer New York, 2006.
- [59] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical recipes in Fortran 77 and Fortran 90*, Cambridge University Press Cambridge, 1996.
- [60] T. A. Zang, Y. S. Wong, M. Y. Hussaini, Spectral multigrid methods for elliptic equations II, *Journal of Computational Physics* 54 (3) (1984) 489–507.
- [61] U. Trottenberg, C. W. Oosterlee, A. Schuller, *Multigrid*, Academic Press, 2000.
- [62] I. Yavneh, Why multigrid methods are so efficient, *Computing in science & engineering* 8 (6) (2006) 12–22.
- [63] C. Rodrigo, F. J. Gaspar, C. W. Oosterlee, I. Yavneh, Accuracy measures and Fourier analysis for the full multigrid algorithm, *SIAM Journal on Scientific Computing* 32 (5) (2010) 3108–3129.
- [64] S. B. Pope, *Turbulent flows*, IOP Publishing, 2001.
- [65] H. Koch, D. Tataru, Well-posedness for the Navier–Stokes equations, *Advances in Mathematics* 157 (1) (2001) 22–35.
- [66] J. Nordström, M. Svärd, Well-posed boundary conditions for the Navier–Stokes equations, *SIAM Journal on Numerical Analysis* 43 (3) (2005) 1231–1255.