ASME Fluids Engineering Conference (July 11, 1996)
Forum on Control of Transitional and Turbulent Flows

A METHOD FOR OPTIMIZING FEEDBACK CONTROL RULES FOR WALL-BOUNDED TURBULENT FLOWS BASED ON CONTROL THEORY

Thomas Bewley and Parviz Moin
Department of Mechanical Engineering
Stanford University
Stanford, California

Roger Temam
The Institute for Scientific Computing and Applied Mathematics
Indiana University
Bloomington, Indiana

ABSTRACT

A new method based on control theory for optimizing feedback control rules with the objective of reducing drag in wall-bounded turbulent flows is presented. Both linear and nonlinear control rules (of the type commonly used in neural networks) are considered. These control rules relate wall measurements of skin friction and pressure to the control, which is applied as a continuous distribution of wall-normal boundary velocity with zero net transpiration. Though the optimization technique itself requires complete information about the flow, and thus can only be performed computationally, it is intended that the resulting optimized rules be scaled appropriately and used in physical boundary layer control implementations.

Using optimal control theory, the sensitivity of some representative cost functional to small modifications in the coefficients of a feedback control rule are found via the solution of an adjoint problem. With this sensitivity field, the coefficients are iteratively updated with a gradient algorithm until the cost functional is minimized. Given that this optimization is performed in a representative situation, the coefficients then may be fixed and the control rule effectively used in other flows with similar configurations, requiring only information about the flow which can be obtained with flush-mounted sensors on the wall.

1. Background

Optimal control theory applied to turbulence provides a rigorous framework to determine the gradient of a cost functional (which represents a physical problem of interest) with respect to small modifications of the control forcing (Abergel and Temam, 1990). With such information, combined with a gradient algorithm to update the control, very effective control distributions may be determined. For example, recent numerical simulations of this approach in a low Reynolds number turbulent channel flow obtained a 50% drag reduction and an order of magnitude turbulent kinetic energy reduction with small levels of boundary velocity control (Moin and Bewley, 1995). Important drawbacks of this approach, however, are 1) it requires complete information about the turbulent fluctuations in the near-wall region, and 2) it is extremely computationally expensive. Thus, it is impossible to apply the optimal control approach directly in an experimental setting.

In order to arrive at a practical scheme, a method was sought to optimize control rules which 1) require only flow information obtainable with wall-mounted sensors, and 2) are computationally inexpensive enough to apply in real time. Possible approaches for this purpose can be divided into two broad categories: *state trajectory approaches*, which attempt to drive some description of the turbulent state (or a portion thereof) in a desired manner, and *direct approaches*, which bypass any description of the turbulent state *per se*, but simply seek a control rule which achieves a desired effect, such as the reduction of drag.

As an example of one state trajectory approach, an adaptive inverse technique has been applied to a low Reynolds number turbulent channel flow, providing approximately 18% drag reduction (Kim, 1996). This approach first develops an approximate "inverse" model between measurable flow quantities (as input) and the control forcing (as output) with an adaptive technique. Each iteration of the adaptation consists of three steps: 1) computing the error of the model output with respect to the desired model output (the actual control forcing used), 2) determining the influence of the weights in the model on this error, then 3) updating all the weights in the model a small amount in a manner that reduces the error. In neural networks, this is commonly referred to as "back-propagation" of the error. Once this approximate inverse model between the flow measurements and the control converges, the inverse model is used to compute a control which will drive the flow measurements to some desired state. In the case of Kim (1996), the desired state is chosen to be a state with reduced spanwise drag fluctuations.

Drawbacks of the adaptive inverse approach are 1) an *ad hoc* desired state must be chosen, 2) a random "dither" signal needs to be applied to the control in order for the inverse model to have "sufficiently exciting" data from which to learn, which reduces the performance of the controller, and 3) it is possible that even at statistical steady state, due to the nonlinear nature of the Navier-Stokes equation, the weights in the inverse model may need to continually adapt in order to represent a temporally evolving relationship between the flow measurements and the control. Thus, if the training of the inverse model does not converge fast enough, it will not have time to keep up with the temporal evolution of the flow (for instance, the movement of the near-wall turbulent coherent structures), and may not develop an accurate model between flow measurements and the control which produces them.

Other state trajectory approaches attempt to control a more complete description of the turbulence using a low-dimensional (10-20 mode) representation of the near-wall coherent structures (Coller *et al.*, 1994). In this approach, the orbit of the near wall structures in this representation is partially stabilized, resulting in a reduced "bursting" frequency and, presumably, reduced drag. Coller *et al.*(1994) showed that the frequency of bursting events could be reduced in their model equations, but did not demonstrate how effective such an approach would be at reducing drag when applied to a fully turbulent flow.

Drawbacks of this low-dimensional representation approach include 1) an accurate estimation of the state in this representation needs to be made from the measurements at the wall, and 2) a desired *ad hoc* state trajectory must be chosen, which can only be selected well if one has a detailed understanding of the cause/effect relationship of the drag-producing phenomena in the near-wall region, which is still under debate.

Direct approaches may be proposed which bypass estimation and control of the state trajectory altogether. In such approaches, one simply represents the control objective mathematically as a cost functional, then attempts to find a control rule which minimizes this functional.

The simplest direct approach is an adaptive "reinforcement learning" approach. In such an approach, the weights of a control rule are initialized randomly and the control rule applied to the flow. Every time a "good" result is seen (for example, the drag is reduced), the weights contributing most to the control at that instant are increased, and every time a "bad" result is seen, the corresponding responsible weights are decreased.

The main drawback of this approach, however, is that this reward/punish training algorithm is not very reliable, especially for complicated nonlinear systems, and thus the scheme may not converge at all.

Thus, we arrive at the motivation for the current work, in which we derive a rigorous algorithm to efficiently optimize a direct control scheme, with the goal in mind simply of reducing some integral measure of the control objective without the prescription of a desired state trajectory. This approach, based on computation of the gradient of a cost functional with respect to modification of the weights in the control rule, will be outlined in the following sections. Numerical simulations that implement the technique described here are currently underway.

2. Problem statement

Our goal is to determine a control rule which takes as input the measurable flow quantities on the wall (localized measurements of streamwise drag, spanwise drag, and pressure) and produces as output a control Φ (the normal component of velocity on the wall) which effectively controls the flow system. The flow system we consider is fully developed turbulent channel flow with periodic boundary conditions in the streamwise and spanwise directions; however, the control obtained should apply well to turbulent boundary layers as well due to the similar near-wall behaviors of these flows.

The flow is governed by the incompressible Navier-Stokes equations, which may be written in symbolic form as:

$$\mathcal{N}(U) = 0, \tag{1a}$$

(interior operators such as $\mathcal{N}(U)$ are written out in full in the Appendix) with control Φ on the wall-normal component of velocity at the walls

$$u_i = n_i \Phi \qquad \text{on walls,} \tag{1b}$$

where we will take

$$\Phi = \sum_{\lambda=1}^{\Lambda} W_\lambda \, g(\xi_\lambda),$$

$$\xi_\lambda = w_{1\lambda} * \mu \frac{\partial u_1}{\partial n} + w_{2\lambda} * p + w_{3\lambda} * \mu \frac{\partial u_3}{\partial n} + B_\lambda,$$

and with fully developed turbulent channel flow initial conditions

$$u_i = u_i(0) \qquad \text{at } t = 0, \tag{1c}$$

where n_i is a unit normal on each wall facing *into* the flow, Λ is the number of "nodes" in the "hidden layer" of the network, and $g(\xi_\lambda)$ is an "activation function" (linear or nonlinear) which will be prescribed. Control rules of the form shown above, used commonly in neural networks, have seen a broad range of application and are capable of representing very general nonlinear relationships (Hertz *et al.*, 1991). The task at hand is simply to optimize the weighting functions $w_{\kappa\lambda}$ and the discrete weights B_λ and W_λ such that Φ effectively controls the flow system.

Note that the $w_{\kappa\lambda}$ are convolution functions, where the convolution is defined such that, for example,

$$w_{1\lambda} * \mu \frac{\partial u_1}{\partial n} = \int_{\partial\Omega} w_{1\lambda}(\bar{x})\ \mu \frac{\partial u_1}{\partial n}(x - \bar{x})\ d\bar{S}.$$

By optimizing the convolution functions $w_{\kappa\lambda}(\bar{x})$, we take into account "nearby" flow measurements (in the direction $\bar{x}$) from a specific actuator location (x). In fact, the extent to which these convolution functions are nonzero when converged will indicate how far in each direction from a specific actuator flow measurements are relevant when computing an effective control.

Note that the weighting functions $w_{\kappa\lambda}$ and the discrete weights B_λ and W_λ are prescribed at the outset to be *invariant in time.* Though the method used requires that the weights be optimized by considering finite time horizon $[0, T]$, we seek to approximate the steady-state weights at the "infinite time horizon" in which turbulent fluctuations near the wall are countered by a fixed control rule at the wall in an efficient drag-reducing manner.

As $g(\xi_\lambda)$ may be nonlinear, Einstein notation may not be used for the index λ, as an extra term involving ξ_λ arises in the differentiation of $g(\xi_\lambda)$. Thus, for clarity, Greek subscripts will be summed only if explicitly stated, and products involving the same Greek subscript more than twice, such as equation (3), are not necessarily typos.

3. Control networks

We now consider two specific control networks. The first is a simple one-layer linear network, as shown in figure 1a. In terms of equation (1b), we will restrict the network to a single node $\Lambda = 1$, the weight W_1 will remain fixed at $W_1 = 1$, and the activation function will be a simple linear relationship $g(\xi_1) = \xi_1$. The weights $w_{\kappa 1}$ and B_1 will be optimized with a systematic procedure.

The second network to be considered, which is similar to that commonly used in neural networks, is a two-layer nonlinear form, as shown in figure 1b. In terms of equation (1b), the number of nodes in the hidden layer Λ will remain (for now) unspecified, the activation function at the first node $\lambda = 1$ will again be $g(\xi_1) = \xi_1$, and the activation functions at the other hidden nodes $\lambda > 1$ will be sigmoid saturation functions given by $g(\xi_\lambda) = \tanh(\xi_\lambda)$, as illustrated in figure 2. The weights $w_{\kappa\lambda}$, B_λ, and W_λ will be optimized with a systematic procedure.

In the case of the linear network of figure 1a, the weights may all be initialized to zero and updated iteratively with a gradient algorithm. In the case of the nonlinear network of figure 2b, the weights $w_{\kappa\lambda}$, B_λ, and W_λ must *not* all be initialized with the same values, for if they are, the relationship "learned" by each node will be identical. In this case, then, all weights are randomly initialized such that the initial inputs to the hidden nodes are $O(1)$ quantities. With such a scheme, the different nodes will converge to various different important features of a nonlinear relationship between the flow measurements and an effective control strategy. Assuming convergence of this network for a sufficiently large number of nodes Λ, performance should be superior to that of the linear network.

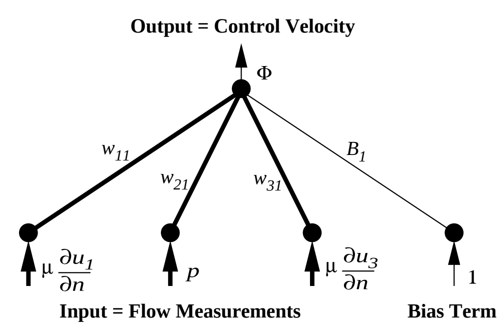

Figure 1a. Single layer linear network. The flow measurements which we take as inputs are localized measurements on the wall of the streamwise drag, the pressure, and the spanwise drag. A

bias term is also included as an input clamped to unity. The flow measurements are convolved with the weighting functions $w_{\kappa 1}$, summed, and added to the bias weight B_1 to determine the control Φ. The input flow measurements are field variables and are indicated with heavy arrows—the corresponding weights are convolution functions (in the continuous case) or two dimensional arrays containing a stencil of weights (in the discrete case).

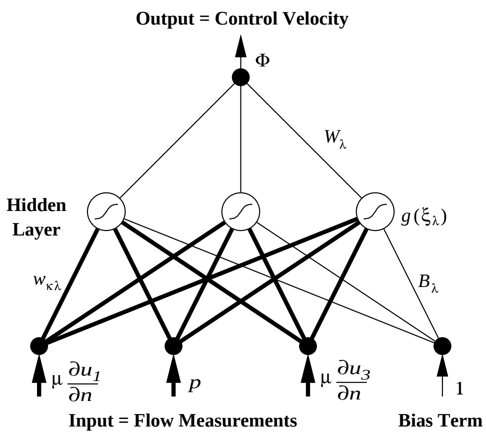

Figure 1b. Two layer nonlinear network. The output of several simple networks λ similar to the one depicted in figure 1a are used as the arguments ξ_λ to activation functions g at the hidden nodes. The output of all of the hidden nodes $g(\xi_\lambda)$ are then weighted with the W_λ and summed to produce the control Φ.

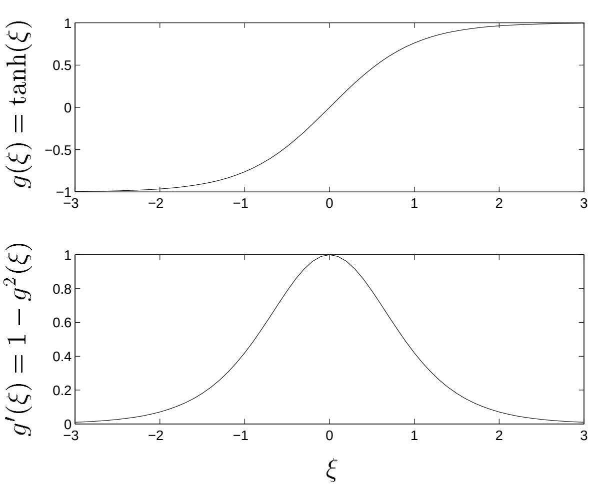

Figure 2. The sigmoid saturation function $g(\xi)$ and its derivative.

4. Cost functional

The objective of applying the control in this problem is to reduce the drag without using excessive amounts of control forcing. Additionally, we will constrain the control rule so that it applies approximately zero net transpiration, as motivated both by physical flow control devices and the current simulations which require a control with zero net mass flux.

Mathematically, a cost functional for this problem may thus be expressed as

$$\begin{aligned} \mathscr{J} = &\frac{1}{A\,T}\int_{\partial\Omega}\int_0^T \left(\mu\,\frac{\partial u_1}{\partial n} + \frac{\ell}{2}\,\Phi^2\right)\,dt\,dS \\ &+ \frac{1}{T}\int_0^T \frac{m}{2}\,\bar{\Phi}^2\,dt, \end{aligned} \tag{2}$$

where

$$\bar{\Phi} = \frac{1}{A}\int_{\partial\Omega} \Phi\,dS.$$

The term involving $\mu\,\partial u_1/\partial n$ is the average drag. The term involving Φ^2 is an expression of the magnitude of the control. These two terms are weighted together with a factor ℓ, which represents the price of the control. This quantity is small if the control is "cheap", and large if applying the control is "expensive". The term involving $\bar{\Phi}$ is the mean value of the control—a large value for m will be prescribed to result in small amounts of net transpiration. (Note that, in practice, the net transpiration is set exactly to zero at each time step).

Thus, minimization of $\mathscr{J}$ corresponds to reducing drag while maintaining a small amount of control forcing and a (very) small net transpiration.

5. Gradient of cost functional

We now develop a technique to compute the gradient of the cost functional $\mathscr{J}$ with respect to the weighting functions w. Similar techniques can be used to express the gradients of $\mathscr{J}$ with respect to the discrete weights B and W, as discussed in Bewley (1996).

Differential field

Consider first the Fréchet differential of the flow U with respect to w, which is defined such that

$$\begin{aligned} \dot{U}^w &\equiv \frac{1}{A}\lim_{\epsilon\to 0}\frac{U(w+\epsilon\dot{w},B,W) - U(w,B,W)}{\epsilon} \\ &= \frac{1}{A}\int_{\partial\Omega}\sum_{\kappa=1}^{3}\sum_{\lambda=1}^{\Lambda}\frac{\mathscr{D}U(w,B,W)}{\mathscr{D}w_{\kappa\lambda}}\dot{w}_{\kappa\lambda}\,dS \end{aligned}$$

where $\dot{w}$ is an arbitrary "update direction" to the weighting function w. This update direction will remain undetermined and will later be isolated and removed from the equation for the differential of the cost functional.

The differential field $\dot{U}^w$ is governed by the Fréchet differential of (1) with respect to w, which may be written:

$$\mathcal{A}\dot{U}^w = 0, \tag{3a}$$

with boundary conditions

$$\dot{u}_i^w = n_i \dot{\Phi}^w \qquad \text{on walls}, \tag{3b}$$

where

$$\dot{\Phi}^w = \sum_{\lambda=1}^{\Lambda} W_\lambda \, g'(\xi_\lambda)\, \dot{\xi}_\lambda^w$$

$$\dot{\xi}_\lambda^w = w_{1\lambda} * \mu \frac{\partial \dot{u}_1^w}{\partial n} + w_{2\lambda} * \dot{p}^w + w_{3\lambda} * \mu \frac{\partial \dot{u}_3^w}{\partial n} + \dot{w}_{1\lambda} * \mu \frac{\partial u_1}{\partial n} + \dot{w}_{2\lambda} * p + \dot{w}_{3\lambda} * \mu \frac{\partial u_3}{\partial n},$$

and with initial conditions

$$\dot{u}_i^w = 0 \qquad \text{at } \; t = 0. \tag{3c}$$

The Fréchet differential of the cost functional $\mathcal{J}$ with respect to w is:

$$\begin{aligned}
\dot{\mathcal{J}}^w &\equiv \frac{1}{A} \lim_{\epsilon \to 0} \frac{\mathcal{J}(w + \epsilon \dot{w}, B, W) - \mathcal{J}(w, B, W)}{\epsilon} \\
&= \frac{1}{A} \int_{\partial\Omega} \sum_{\kappa=1}^{3} \sum_{\lambda=1}^{\Lambda} \frac{\mathscr{D}\mathcal{J}(w, B, W)}{\mathscr{D} w_{\kappa\lambda}} \, \dot{w}_{\kappa\lambda} \, dS \qquad (4) \\
&= \frac{1}{A\,T} \int_{\partial\Omega} \int_0^T \left[\mu \frac{\partial \dot{u}_1^w}{\partial n} + \left(\ell\,\Phi + m\,\bar{\Phi}\right) \dot{\Phi}^w \right] dt\, dS
\end{aligned}$$

It is seen that the differential of the cost functional $\dot{\mathcal{J}}^w$ is a function of the differential of the flow $\dot{U}^w$. The linear dependence of $\dot{U}^w$ on $\dot{w}$ may, in theory, be found directly from (3). However, in practice, this is not a tractable approach due to the excessively large dimension of the problem under consideration. Thus, we seek a simpler way to express the above equation in a way that we may determine the gradient $\mathscr{D}\mathcal{J}(w, B, W)/\mathscr{D} w_{\kappa\lambda}$. It is for this reason that we now propose the definition of an adjoint field.

Adjoint field

As shown in the Appendix, an adjoint operator $\mathcal{A}^*$ may be defined by the identity

$$< \mathcal{A}\dot{U}^w, \tilde{U} > = < \dot{U}^w, \mathcal{A}^* \tilde{U} > + b^w. \tag{5}$$

In order to express the differential of the cost functional (4) in a usable form, we now define an adjoint state by the system of equations

$$\mathcal{A}^* \tilde{U} = 0, \tag{6a}$$

with boundary conditions on the walls

$$\begin{aligned}
\tilde{u}_1 &= 1 + \sum_{\lambda=1}^{\Lambda} \left(\tilde{h}\, W_\lambda \, g'(\xi_\lambda)\right) * \check{w}_{1\lambda} \\
\tilde{u}_2 &= -n_2 \sum_{\lambda=1}^{\Lambda} \left(\tilde{h}\, W_\lambda \, g'(\xi_\lambda)\right) * \check{w}_{2\lambda} \qquad (6b) \\
\tilde{u}_3 &= \sum_{\lambda=1}^{\Lambda} \left(\tilde{h}\, W_\lambda \, g'(\xi_\lambda)\right) * \check{w}_{3\lambda},
\end{aligned}$$

where

$$\begin{aligned}
\tilde{f} &\equiv \tilde{p} - 2\,\rho\,\tilde{u}_2\, u_2 - \mu \frac{\partial \tilde{u}_2}{\partial x_2} \\
\tilde{h} &\equiv -\tilde{f} + \ell\,\Phi + m\,\bar{\Phi} \\
\check{b}(x) &\equiv b(-x),
\end{aligned}$$

and with initial conditions

$$\tilde{u}_i = 0 \qquad \text{at } \; t = T. \tag{6c}$$

Gradient

Using the identity (5) and the definition of the adjoint in (6), we can algebraically manipulate equation (4) to the form

$$\begin{aligned}
&\frac{1}{A} \int_{\partial\Omega} \sum_{\kappa=1}^{3} \sum_{\lambda=1}^{\Lambda} \frac{\mathscr{D}\mathcal{J}(w, B, W)}{\mathscr{D} w_{\kappa\lambda}} \, \dot{w}_{\kappa\lambda} \, dS \\
&\qquad = \frac{1}{A} \int_{\partial\Omega} \sum_{\kappa=1}^{3} \sum_{\lambda=1}^{\Lambda} \tilde{G}_{w_{\kappa\lambda}} \, \dot{w}_{\kappa\lambda} \, dS,
\end{aligned}$$

where $\tilde{G}_{w_{\kappa\lambda}}$ is some function of the solution to the adjoint problem (6). As $\dot{w}$ is arbitrary, we may then identify the expression for $\tilde{G}_{w_{\kappa\lambda}}$ as $\mathscr{D}\mathcal{J}(w, B, W)/\mathscr{D} w_{\kappa\lambda}$. It is straightforward to show (Bewley, 1996) that the resulting expression for the gradient is

$$\begin{aligned}
\frac{\mathscr{D}\mathcal{J}(w, B, W)}{\mathscr{D} w_{1\lambda}} &= \frac{1}{T} \int_0^T \left(\tilde{h}\, W_\lambda \, g'(\xi_\lambda)\right) * \mu \frac{\partial \check{u}_1}{\partial n} \, dt \\
\frac{\mathscr{D}\mathcal{J}(w, B, W)}{\mathscr{D} w_{2\lambda}} &= \frac{1}{T} \int_0^T \left(\tilde{h}\, W_\lambda \, g'(\xi_\lambda)\right) * \check{p} \, dt \\
\frac{\mathscr{D}\mathcal{J}(w, B, W)}{\mathscr{D} w_{3\lambda}} &= \frac{1}{T} \int_0^T \left(\tilde{h}\, W_\lambda \, g'(\xi_\lambda)\right) * \mu \frac{\partial \check{u}_3}{\partial n} \, dt \\
\frac{\partial \mathcal{J}(w, B, W)}{\partial B_\lambda} &= \frac{1}{A\,T} \int_{\partial\Omega} \int_0^T \tilde{h}\, W_\lambda \, g'(\xi_\lambda) \, dt\, dS \\
\frac{\partial \mathcal{J}(w, B, W)}{\partial W_\lambda} &= \frac{1}{A\,T} \int_{\partial\Omega} \int_0^T \tilde{h}\, g(\xi_\lambda) \, dt\, dS.
\end{aligned}$$

6. Gradient update to the weights

With the gradients computed using the adjoint field, a control rule may be optimized using a gradient algorithm, such as the simple gradient algorithm

$$w_{\kappa\lambda} = w_{\kappa\lambda} - \alpha \frac{\mathscr{D}\mathcal{J}(w,B,W)}{\mathscr{D}w_{\kappa\lambda}}$$

$$B_\lambda = B_\lambda - \alpha \frac{\partial\mathcal{J}(w,B,W)}{\partial B_\lambda}$$

$$W_\lambda = W_\lambda - \alpha \frac{\partial\mathcal{J}(w,B,W)}{\partial W_\lambda},$$

or a conjugate gradient algorithm, which shows better convergence properties for many problems of this type (Bewley, 1996). The descent parameter α may be adjusted at each iteration to be that value which minimizes $\mathcal{J}$ in the direction of the gradient. After several iterations, this scheme should converge to some local minimum of $\mathcal{J}$ with respect to the weights—note that global convergence can not be assured due to the possibility of multiple minima points of $\mathcal{J}(w,B,W)$ in this nonlinear problem.

7. Discussion

A new technique for optimizing feedback control rules for turbulent flows has been proposed. This technique is based solely on the equations governing the flow and a mathematical statement of the control objective, thus bypassing the *ad hoc* identification of a desired state trajectory often used to determine feedback control rules. Also, the training is based on an adjoint ("sensitivity") field, which determines the gradient of the cost with respect to small modifications of the weights in a rigorous manner. Thus, convergence can be expected to be much better than for an reinforcement learning approach with an adaptive algorithm.

A straightforward extension of the present approach is to take into account *past* measurements in the control rule. Past measurements, which may easily be stored in an experimental implementation, may give additional information about the convection velocity of flow structures which cannot be determined from instantaneous measurements alone. It is also possible that such information can be determined by recurrent networks, in which the inputs of the control network include the outputs of the network from the previous time step.

Drawbacks of the present method include 1) an accurate mathematical model of the flow equations and boundary conditions are needed for the training, and 2) the training algorithm is quite complex, requiring simulation on a supercomputer. However, this method should provide insight into effective new control rules which one could not think of otherwise, and which can be further modified to fit practical problems. In addition, they may be used to guide the development of experimental configurations, revealing the necessary locations of sensors with respect to the actuators in order to obtain information relevant to effective control strategies.

ACKNOWLEDGEMENTS

The authors are grateful for helpful discussions with Prof. Ron Blackwelder, Prof. William C. Reynolds, and Miss Sharon Liu. The financial support of AFOSR Grant No. F49620-93-1-0078 is also gratefully acknowledged.

REFERENCES

Abergel, F. & Temam, R. 1990 On some control problems in fluid mechanics. *Theor. and Comp. Fluid Dynamics.* **1,** 303.

Bewley, T. 1996 Optimal feedback control of turbulence. Ph.D. Thesis (in preparation), Department of Mechanical Engineering, Stanford University.

Coller, B., Holmes, P., & Lumley, J. 1994 Interaction of adjacent bursts in the wall region. *Phys. Fluids.* **6,** 954-961.

Hertz, J., Krogh, A., & Palmer, R. 1991 *Introduction to the theory of neural computation,* Addison Wesley, New York.

Kim, J. 1996 private communication.

Moin, P., & Bewley, T. 1995 Application of control theory to turbulence. *Twelfth Australasian Fluid Mechanics Conference,* December 10-15, Sydney, Australia, p. 109.

APPENDIX A. Differential operators

The fields referred to in this work are the flow field U, the differential field $\dot{U}^w$, and the adjoint field $\tilde{U}$, each of which is composed of three velocity components and a pressure component

$$U = \begin{pmatrix} u_i(x_1,x_2,x_3,t) \\ p(x_1,x_2,x_3,t) \end{pmatrix}, \quad \dot{U}^w = \begin{pmatrix} \dot{u}_i^w(x_1,x_2,x_3,t) \\ \dot{p}^w(x_1,x_2,x_3,t) \end{pmatrix},$$

$$\tilde{U} = \begin{pmatrix} \tilde{u}_i^w(x_1,x_2,x_3,t) \\ \tilde{p}^w(x_1,x_2,x_3,t) \end{pmatrix}.$$

The Navier-Stokes operator is given by

$$\mathcal{N}(U) = \begin{pmatrix} \dfrac{\partial u_i}{\partial t} + u_j \dfrac{\partial u_i}{\partial x_j} - \nu \dfrac{\partial^2 u_i}{\partial x_j^2} + \dfrac{1}{\rho}\dfrac{\partial p}{\partial x_i} + \dfrac{1}{\rho}\delta_{1i} P_x \\ \dfrac{\partial u_j}{\partial x_j} \end{pmatrix}.$$

The Fréchet differential of the (non-linear) Navier-Stokes operator is given by

$$\mathcal{A}\dot{U}^w = \begin{pmatrix} \dfrac{\partial \dot{u}_i^w}{\partial t} + u_j \dfrac{\partial \dot{u}_i^w}{\partial x_j} + \dot{u}_j^w \dfrac{\partial u_i}{\partial x_j} - \nu \dfrac{\partial^2 \dot{u}_i^w}{\partial x_j^2} + \dfrac{1}{\rho}\dfrac{\partial \dot{p}^w}{\partial x_i} \\ -\dfrac{1}{\rho}\dfrac{\partial \dot{u}_j^w}{\partial x_j} \end{pmatrix},$$

which is linear in the differential field $\dot{U}^w$, but is a function of the solution U of the Navier-Stokes problem, so that $\mathcal{A} = \mathcal{A}(U)$. Define an inner product over the domain in space-time under consideration such that

$$< \dot{U}^w, \tilde{U} > = \int_\Omega \int_0^T \dot{U}^w \cdot \tilde{U} \, dt \, dV,$$

and consider the identity

$$< \mathcal{A}\dot{U}^w, \tilde{U} > = < \dot{U}^w, \mathcal{A}^* \tilde{U} > + b.$$

Integration by parts may be used to move all differential operations from $\dot{U}^w$ on the left hand side of the equation to $\tilde{U}$ on the right hand side, resulting in an expression for the adjoint operator

$$\mathcal{A}^*\tilde{U} = \begin{pmatrix} -\dfrac{\partial \tilde{u}_i}{\partial t} - u_j\Big(\dfrac{\partial \tilde{u}_i}{\partial x_j} + \dfrac{\partial \tilde{u}_j}{\partial x_i}\Big) - \nu \dfrac{\partial^2 \tilde{u}_i}{\partial x_j^2} + \dfrac{1}{\rho}\dfrac{\partial \tilde{p}}{\partial x_i} \\ -\dfrac{1}{\rho}\dfrac{\partial \tilde{u}_j}{\partial x_j} \end{pmatrix},$$

where $\mathcal{A}^* = \mathcal{A}^*(U)$, and an expression for b, which contains all the boundary terms:

$$\begin{aligned} b = \int_{\partial\Omega} \int_0^T -n_j \Big[\tilde{u}_i \Big(u_j \, \dot{u}_i^w + u_i \, \dot{u}_j^w \Big) \\ - \nu \Big(\frac{\partial \dot{u}_i^w}{\partial x_j} \tilde{u}_i - \dot{u}_i^w \frac{\partial \tilde{u}_i}{\partial x_j} \Big) + \frac{1}{\rho} \Big(\dot{p}^w \, \tilde{u}_j - \dot{u}_j^w \, \tilde{p} \Big) \Big] \, dt \, dS \\ + \int_\Omega \dot{u}_i^w \, \tilde{u}_i \, dV \Big|_{t=T} - \int_\Omega \dot{u}_i^w \, \tilde{u}_i \, dV \Big|_{t=0}. \end{aligned}$$

Simplification of the identity (5) by interior equations, boundary conditions, and initial conditions on U, $\dot{U}^w$, and $\tilde{U}$ can provide an expression which recasts $\dot{\mathcal{J}}^w$ from a function of $\dot{U}^w$ to a more manageable function of the solution to an adjoint problem for $\tilde{U}$, as discussed in the text.

APPENDIX B. Useful convolution identities

The convolution $a * b$ is defined:

$$a * b = \int a(\bar{x}) \, b(x - \bar{x}) \, d\bar{S}$$

Note that, for infinite limits or periodic a and b, $a * b = b * a$. Also note that convolution is distributive over addition, so $(a+b)*c = a*c + b*c$.

With this definition, we may derive two useful identities:

$$\begin{aligned} \int a \, (b * c) \, dS &= \int a(x) \int b(\bar{x}) \, c(x - \bar{x}) \, d\bar{S} \, dS \\ &= \int a(x) \int b(x - \bar{x}) \, c(\bar{x}) \, d\bar{S} \, dS \\ &= \int c(x) \int a(\bar{x}) \, b(\bar{x} - x) \, d\bar{S} \, dS \\ &= \int c \, (a * \check{b}) \, dS, \end{aligned} \tag{7}$$

where $\check{b}(x) \equiv b(-x)$, and

$$\begin{aligned} \int a \, (b * c) \, dS &= \int a(x) \int b(\bar{x}) \, c(x - \bar{x}) \, d\bar{S} \, dS \\ &= \int b(x) \int a(\bar{x}) \, c(\bar{x} - x) \, d\bar{S} \, dS \\ &= \int b \, (a * \check{c}) \, dS, \end{aligned} \tag{8}$$

where $\check{c}(x) \equiv c(-x)$.

Further, with $\hat{a}$ denoting the Fourier transform of a, note that convolution in real space corresponds to simple multiplication in Fourier space

$$\widehat{a * b} = \hat{a} \, \hat{b}$$

and that the Fourier transform of $\check{c}$ is the complex conjugate of the Fourier transform of c

$$\hat{\check{c}} = \hat{c}^*.$$

APPENDIX C. Algebra to determine $\mathscr{D}\mathcal{J}(w,B,W)/\mathscr{D}w_{\kappa\lambda}$

Notice that, using (3b), we may write equation (4) as

$$\frac{1}{A}\int_{\partial\Omega}\sum_{\kappa=1}^{3}\sum_{\lambda=1}^{\Lambda}\frac{\mathscr{D}\mathcal{J}(w,B,W)}{\mathscr{D}w_{\kappa\lambda}}\,\dot{w}_{\kappa\lambda}\,dS=\frac{1}{AT}\int_{\partial\Omega}\int_0^T\Bigg[\mu\frac{\partial\dot{u}_1^w}{\partial n}+(\ell\,\Phi+m\,\bar{\Phi})\sum_{\lambda=1}^{\Lambda}W_\lambda\,g'(\xi_\lambda)\cdot \Big(w_{1\lambda}*\mu\frac{\partial\dot{u}_1^w}{\partial n}+w_{2\lambda}*\dot{p}^w+w_{3\lambda}*\mu\frac{\partial\dot{u}_3^w}{\partial n}+\dot{w}_{1\lambda}*\mu\frac{\partial u_1}{\partial n}+\dot{w}_{2\lambda}*p+\dot{w}_{3\lambda}*\mu\frac{\partial u_3}{\partial n}\Big)\Bigg]\,dt\,dS. \tag{9}$$

Using the adjoint field described by (6), we will algebraically manipulate the RHS of (9) to the form

$$\frac{1}{A}\int_{\partial\Omega}\sum_{\kappa=1}^{3}\sum_{\lambda=1}^{\Lambda}\frac{\mathscr{D}\mathcal{J}(w,B,W)}{\mathscr{D}w_{\kappa\lambda}}\,\dot{w}_{\kappa\lambda}\,dS=\frac{1}{A}\int_{\partial\Omega}\sum_{\kappa=1}^{3}\sum_{\lambda=1}^{\Lambda}\tilde{G}_{w_{\kappa\lambda}}\,\dot{w}_{\kappa\lambda}\,dS, \tag{10}$$

where $\tilde{G}_{w_{\kappa\lambda}}$ is some function to the solution of the adjoint problem. As $\dot{w}$ is arbitrary, we may then identify the expression for $\tilde{G}_{w_{\kappa\lambda}}$ as $\mathscr{D}\mathcal{J}(w,B,W)/\mathscr{D}w_{\kappa\lambda}$.

We begin by noting that equation (5) may be simplified using (1), (3), and (6). Multiplying by ρ and applying the definition for $\tilde{f}$, equation (5) becomes

$$0=\int_{\partial\Omega}\int_0^T\Big(-\mu\frac{\partial\dot{u}_i^w}{\partial n}\,\tilde{u}_i+\dot{p}^w\,\tilde{u}_2\,n_2-\tilde{f}\,\dot{u}_2^w\,n_2\Big)\,dt\,dS. \tag{11a}$$

Inserting (3b), this becomes

$$\int_{\partial\Omega}\int_0^T\Bigg[-\mu\frac{\partial\dot{u}_i^w}{\partial n}\,\tilde{u}_i+\dot{p}^w\,\tilde{u}_2\,n_2-\tilde{f}\sum_{\lambda=1}^{\Lambda}W_\lambda\,g'(\xi_\lambda)\Big(w_{1\lambda}*\mu\frac{\partial\dot{u}_1^w}{\partial n}+w_{2\lambda}*\dot{p}^w+w_{3\lambda}*\mu\frac{\partial\dot{u}_3^w}{\partial n}\Big)\Bigg]dt\,dS$$
$$=\int_{\partial\Omega}\int_0^T\tilde{f}\sum_{\lambda=1}^{\Lambda}W_\lambda\,g'(\xi_\lambda)\Big(\dot{w}_{1\lambda}*\mu\frac{\partial u_1}{\partial n}+\dot{w}_{2\lambda}*p+\dot{w}_{3\lambda}*\mu\frac{\partial u_3}{\partial n}\Big)dt\,dS. \tag{11b}$$

Applying the identity (7) to the LHS and the identity (8) to the RHS yields:

$$\int_{\partial\Omega}\int_0^T\Bigg[-\mu\frac{\partial\dot{u}_i^w}{\partial n}\,\tilde{u}_i+\dot{p}^w\,\tilde{u}_2\,n_2-\mu\frac{\partial\dot{u}_1^w}{\partial n}\sum_{\lambda=1}^{\Lambda}\Big(\tilde{f}\,W_\lambda\,g'(\xi_\lambda)\Big)*\check{w}_{1\lambda} -\dot{p}^w\sum_{\lambda=1}^{\Lambda}\Big(\tilde{f}\,W_\lambda\,g'(\xi_\lambda)\Big)*\check{w}_{2\lambda}-\mu\frac{\partial\dot{u}_3^w}{\partial n}\sum_{\lambda=1}^{\Lambda}\Big(\tilde{f}\,W_\lambda\,g'(\xi_\lambda)\Big)*\check{w}_{3\lambda}\Bigg]\,dt\,dS \tag{11c}$$
$$=\int_{\partial\Omega}\int_0^T\sum_{\lambda=1}^{\Lambda}\Bigg[\dot{w}_{1\lambda}\Big(\tilde{f}\,W_\lambda\,g'(\xi_\lambda)\Big)*\mu\frac{\partial\check{u}_1}{\partial n}+\dot{w}_{2\lambda}\Big(\tilde{f}\,W_\lambda\,g'(\xi_\lambda)\Big)*\check{p}+\dot{w}_{3\lambda}\Big(\tilde{f}\,W_\lambda\,g'(\xi_\lambda)\Big)*\mu\frac{\partial\check{u}_3}{\partial n}\Bigg]\,dt\,dS.$$

Rearranging and noting that $\partial\dot{u}_2^w/\partial x_2=0$,

$$\int_{\partial\Omega}\int_0^T\Bigg[\mu\frac{\partial\dot{u}_1^w}{\partial n}\Big[\tilde{u}_1+\sum_{\lambda=1}^{\Lambda}\Big(\tilde{f}\,W_\lambda\,g'(\xi_\lambda)\Big)*\check{w}_{1\lambda}\Big]+\dot{p}^w\Big[-\tilde{u}_2\,n_2+\sum_{\lambda=1}^{\Lambda}\Big(\tilde{f}\,W_\lambda\,g'(\xi_\lambda)\Big)*\check{w}_{2\lambda}\Big] +\mu\frac{\partial\dot{u}_3^w}{\partial n}\Big[\tilde{u}_3+\sum_{\lambda=1}^{\Lambda}\Big(\tilde{f}\,W_\lambda\,g'(\xi_\lambda)\Big)*\check{w}_{3\lambda}\Big]\Bigg]\,dt\,dS \tag{11d}$$
$$=-\int_{\partial\Omega}\int_0^T\sum_{\lambda=1}^{\Lambda}\Bigg[\dot{w}_{1\lambda}\Big(\tilde{f}\,W_\lambda\,g'(\xi_\lambda)\Big)*\mu\frac{\partial\check{u}_1}{\partial n}+\dot{w}_{2\lambda}\Big(\tilde{f}\,W_\lambda\,g'(\xi_\lambda)\Big)*\check{p}+\dot{w}_{3\lambda}\Big(\tilde{f}\,W_\lambda\,g'(\xi_\lambda)\Big)*\mu\frac{\partial\check{u}_3}{\partial n}\Bigg]\,dt\,dS.$$

With the adjoint boundary conditions defined as in (6b) and applying the identity (7), this may be written

$$\int_{\partial\Omega}\int_0^T\Bigg[\mu\frac{\partial\dot{u}_1^w}{\partial n}+(\ell\,\Phi+m\,\bar{\Phi})\sum_{\lambda=1}^{\Lambda}W_\lambda\,g'(\xi_\lambda)\Big(w_{1\lambda}*\mu\frac{\partial\dot{u}_1^w}{\partial n}+w_{2\lambda}*\dot{p}^w+w_{3\lambda}*\mu\frac{\partial\dot{u}_3^w}{\partial n}\Big)\Bigg]\,dt\,dS \tag{11e}$$
$$=-\int_{\partial\Omega}\int_0^T\sum_{\lambda=1}^{\Lambda}\Bigg[\dot{w}_{1\lambda}\Big(\tilde{f}\,W_\lambda\,g'(\xi_\lambda)\Big)*\mu\frac{\partial\check{u}_1}{\partial n}+\dot{w}_{2\lambda}\Big(\tilde{f}\,W_\lambda\,g'(\xi_\lambda)\Big)*\check{p}+\dot{w}_{3\lambda}\Big(\tilde{f}\,W_\lambda\,g'(\xi_\lambda)\Big)*\mu\frac{\partial\check{u}_3}{\partial n}\Bigg]\,dt\,dS.$$

The identity given in (11e) may now be used to rewrite the differential of the cost functional (9) as a function of the solution to an adjoint problem

$$
\begin{aligned}
&\frac{1}{A}\int_{\partial\Omega}\sum_{\kappa=1}^{3}\sum_{\lambda=1}^{\Lambda}\frac{\mathscr{D}\mathcal{J}(w,B,W)}{\mathscr{D}w_{\kappa\lambda}}\dot{w}_{\kappa\lambda}\,dS=\frac{1}{A\,T}\int_{\partial\Omega}\int_0^T\Bigg[\\
&\quad\sum_{\lambda=1}^{\Lambda}\Big[-\dot{w}_{1\lambda}\Big(\tilde{f}\,W_\lambda\,g'(\xi_\lambda)\Big)*\mu\frac{\partial\check{u}_1}{\partial n}-\dot{w}_{2\lambda}\Big(\tilde{f}\,W_\lambda\,g'(\xi_\lambda)\Big)*\check{p}-\dot{w}_{3\lambda}\Big(\tilde{f}\,W_\lambda\,g'(\xi_\lambda)\Big)*\mu\frac{\partial\check{u}_3}{\partial n}\Big] \qquad (12a)\\
&\quad+\big(\ell\,\Phi+m\,\bar{\Phi}\big)\sum_{\lambda=1}^{\Lambda}W_\lambda\,g'(\xi_\lambda)\Big(\dot{w}_{1\lambda}*\mu\frac{\partial u_1}{\partial n}+\dot{w}_{2\lambda}*p+\dot{w}_{3\lambda}*\mu\frac{\partial u_3}{\partial n}\Big)\Bigg]\,dt\,dS.
\end{aligned}
$$

Applying the identity (8), this may be written

$$
\begin{aligned}
&\frac{1}{A}\int_{\partial\Omega}\sum_{\kappa=1}^{3}\sum_{\lambda=1}^{\Lambda}\frac{\mathscr{D}\mathcal{J}(w,B,W)}{\mathscr{D}w_{\kappa\lambda}}\dot{w}_{\kappa\lambda}\,dS=\frac{1}{A\,T}\int_{\partial\Omega}\int_0^T\sum_{\lambda=1}^{\Lambda}\Bigg[\\
&\quad\dot{w}_{1\lambda}\Big(-\tilde{f}\,W_\lambda\,g'(\xi_\lambda)\Big)*\mu\frac{\partial\check{u}_1}{\partial n}+\dot{w}_{2\lambda}\Big(-\tilde{f}\,W_\lambda\,g'(\xi_\lambda)\Big)*\check{p}+\dot{w}_{3\lambda}\Big(-\tilde{f}\,W_\lambda\,g'(\xi_\lambda)\Big)*\mu\frac{\partial\check{u}_3}{\partial n}+\\
&\quad\dot{w}_{1\lambda}\Big(\big(\ell\,\Phi+m\,\bar{\Phi}\big)W_\lambda\,g'(\xi_\lambda)\Big)*\mu\frac{\partial\check{u}_1}{\partial n}+\dot{w}_{2\lambda}\Big(\big(\ell\,\Phi+m\,\bar{\Phi}\big)W_\lambda\,g'(\xi_\lambda)\Big)*\check{p}+ \qquad (12b)\\
&\quad\dot{w}_{3\lambda}\Big(\big(\ell\,\Phi+m\,\bar{\Phi}\big)W_\lambda\,g'(\xi_\lambda)\Big)*\mu\frac{\partial\check{u}_3}{\partial n}\Bigg]\,dt\,dS.
\end{aligned}
$$

Finally, noting the definition of $\tilde{h}$ and defining

$$
\tilde{G}_{w_{\kappa\lambda}}\equiv\begin{cases}
\dfrac{1}{T}\displaystyle\int_0^T\Big(\tilde{h}\,W_\lambda\,g'(\xi_\lambda)\Big)*\mu\frac{\partial\check{u}_1}{\partial n}\,dt & \kappa=1\\[2ex]
\dfrac{1}{T}\displaystyle\int_0^T\Big(\tilde{h}\,W_\lambda\,g'(\xi_\lambda)\Big)*\check{p}\,dt & \kappa=2\\[2ex]
\dfrac{1}{T}\displaystyle\int_0^T\Big(\tilde{h}\,W_\lambda\,g'(\xi_\lambda)\Big)*\mu\frac{\partial\check{u}_3}{\partial n}\,dt & \kappa=3,
\end{cases} \qquad (13)
$$

we observe that equation (12b) takes the form of equation (10) and thus, since $\dot{w}$ is arbitrary,

$$
\frac{\mathscr{D}\mathcal{J}(w,B,W)}{\mathscr{D}w_{\kappa\lambda}}=\tilde{G}_{w_{\kappa\lambda}}. \qquad (14)
$$