

# Nonlinear Preconditioning in Problems of Optimal Control for Fluid Systems

Bartosz Protas \*

\* Department of Mathematics & Statistics, McMaster University, Hamilton, Ontario, Canada

**Abstract** This note discusses certain aspects of computational solution of optimal control problems for fluid systems. We focus on approaches in which the steepest descent direction of the cost functional is determined using the adjoint equations. In the first part we review the classical formulation by presenting it in the context of Nonlinear Programming. In the second part we show some new results concerning determination of descent directions in general Banach spaces without Hilbert structure. The proposed approach is illustrated with computational examples concerning a state estimation problem for the 1D Kuramoto–Sivashinsky equation.

## 1 Introduction

Problems of optimal control arise in very many areas of science and engineering. Given a (possibly nonlinear) system  $\mathbf{u}(\mathbf{x}, \phi) = 0$ , where  $\mathbf{x}$  is the state of the system and  $\phi$  is an actuation, control problems consist in determining the control  $\phi$ , so that this control and the corresponding state minimize some performance criterion, i.e.,

$$\min_{\mathbf{x} \in \mathcal{X}, \phi \in \mathcal{U}} \tilde{J}(\mathbf{x}, \phi) \quad (1a)$$

$$\text{subject to } \mathbf{u}(\mathbf{x}, \phi) = 0, \quad (1b)$$

where  $\mathcal{U}$  represents the set of admissible controls, whereas  $\mathcal{X}$  is the space of system states. Applications of such problems in Fluid Mechanics are ubiquitous. Here we mention just some of the most important examples, admitting that this list is far from being exhaustive:

- shape optimization with application to aircraft design, e.g., Mohammadi and Pironneau (2001); Martins et al. (2004),
- flow control for drag reduction, e.g., Bewley et al. (2001); Protas and Styczek (2002),
- variational data assimilation in dynamic meteorology, e.g., Kalnay (2003),

- mixing enhancement.

In the above examples the performance criterion  $\tilde{J}$  and the control  $\phi$  may take different forms reflecting the structure of the problem at hand. The equation (1b) governing the state of the system is usually some form of the Navier–Stokes equation. In fact, from the formal point of view, optimal control problems are examples of *inverse problems* [see, e.g., Isakov (1997)].

In practice, problems of the type (1) involving minimization of a cost functional subject to some constraints are solved using optimization methods. Since the constraint is a partial differential equation (PDE), such problems are examples of *PDE-constrained optimization*. One of the first studies to analyze systematically such problems was the seminal work by Lions (1968). In the context of Fluid Mechanics these problems were further investigated by Abergel and Temam (1990) and Gunzburger (2002). When such infinite-dimensional problems are solved in practice, suitable discretization is used to obtain a corresponding finite-dimensional problem which, at least in principle, can be solved using methods of Nonlinear Programming (NLP). There are, however, some aspects of the problem that make this approach quite challenging. First of all, since the discrete systems are obtained from discretizations of PDEs, the dimension of the discrete state vector  $\mathbf{x}$  can be extremely large. Consequently, it is impossible to store the linear operators involved in the solution process as matrices. Consequently, many existing software packages designed to solve finite-dimensional NLP problems cannot be used and “matrix-free” alternatives have to be developed. Secondly, given the size of the discrete system and difficulties involved in calculating second-order derivatives of the cost functional, the Hessian information is usually unavailable and Newton’s method can rarely be used. Consequently, one needs to use first-order (gradient) approaches such as, for instance, the Conjugate Gradient (CG) method. Moreover, the physical systems of interest to us are often characterized by a broad range of interacting length- and time-scales and, as a result, the optimization problem is very poorly conditioned. The purpose of the present paper is to discuss some recent ideas useful for accelerating convergence of iterative solution to such optimization problems. In particular, we will focus on nonlinear preconditioning strategies which, by performing locally a nonlinear change of the metric, attempt to increase the range of validity of the tangent linear approximation which is crucial to the present approach.

The structure of the paper is as follows: in the next Section we introduce a simple, yet relevant from the Fluid Mechanics perspective, optimization problem based on the Kuramoto–Sivashinsky equation that we will use as our “toy model”, then we present a standard adjoint-based optimization approach typically used to solve such problems; in Section 3 we will introduce the idea of nonlinear preconditioning and show how it can be formulated in terms of gradient extraction in spaces without Hilbert structure; in Section 4 we will present some computational

results indicating the utility of the proposed method; conclusions and discussion of further perspectives are deferred to Section 5. The present report is of a rather exploratory nature, and more complete results concerning this problem are already available in Protas (2008).

## 2 Adjoint–Based Optimization

Here we show how problem (1) can be efficiently solved using methods of Non-linear Programming. In its initial formulation this is a *constrained* optimization problem in which both the state  $\mathbf{x}$  and the control  $\phi$  are variables to be optimized. This is a rather inconvenient situation, since  $\mathbf{x}$  is a solution of a (time–dependent) PDE and its discretization may contain a very large number of degrees of freedom in space and in time. On the other hand, the state  $\mathbf{x}$  may be considered a function of the control, i.e.,  $\mathbf{x} = \mathbf{x}(\phi)$ , which allows us to express problem (1) in the corresponding *unconstrained* form

$$\min_{\phi \in \mathcal{U}} \tilde{\mathcal{J}}(\mathbf{x}(\phi), \phi) = \min_{\phi \in \mathcal{U}} \mathcal{J}(\phi), \quad (2)$$

where  $\mathcal{J} : \mathcal{U} \rightarrow \mathbb{R}$  is called the *reduced cost functional*<sup>1</sup>. An advantage of this formulation over (1) is that now optimization is carried out with respect to one variable only with discretization usually involving much fewer degrees of freedom. Moreover, problem (2) is unconstrained so that optimization methods required to solve it are simpler, however, the price to be paid for this is that the functional dependence of  $\mathcal{J}$  on  $\phi$  is now much more involved.

As mention in Introduction, we are concerned here with situations where calculation of the Hessian of (2) is impossible or impractical. We will therefore focus on first–order gradient–based methods. The necessary condition characterizing the minimum of the cost functional  $\mathcal{J}(\phi)$  is the vanishing of its Gâteaux differential  $\mathcal{J}' : \mathcal{U} \times \mathcal{U} \rightarrow \mathbb{R}$ , i.e.

$$\mathcal{J}'(\phi_{opt}; \phi') = 0, \quad \forall \phi' \in \mathcal{U}, \quad (3)$$

where the Gâteaux differential is defined as  $\mathcal{J}'(\phi; \phi') = \lim_{\varepsilon \rightarrow 0} \frac{\mathcal{J}(\phi + \varepsilon \phi') - \mathcal{J}(\phi)}{\varepsilon}$  and  $\phi_{opt}$  is the minimizer. In most applications, and also in the case considered here, the cost functional  $\tilde{\mathcal{J}}$  is quadratic in both  $\mathbf{x}$  and  $\phi$ , however,  $\mathbf{x} = \mathbf{x}(\phi)$  is often a nonlinear mapping and the optimization problem (2) may be therefore nonconvex. As a result, it may admit nonunique solutions and (3) will characterize only a local minimizer  $\phi_{opt}$ . Given some initial guess  $\phi^{(0)}$ , such a minimizer can be found using gradient–based descent method of the general form

$$\phi^{(k+1)} = \phi^{(k)} + \mathbf{d}^{(k)}, \quad k = 1, 2, \dots, \quad (4)$$

---

<sup>1</sup>Since this is the formulation we will focus on below, hereafter we will skip the adjective “reduced”, unless needed for clarity.

such that  $\lim_{k \rightarrow \infty} \boldsymbol{\phi}^{(k)} = \boldsymbol{\phi}_{opt}$ , where  $k$  is the iteration count. At every iteration  $k$  the descent direction  $\mathbf{d}^{(k)}$  is determined based on the gradient  $\nabla J$  of the cost functional calculated at  $\boldsymbol{\phi}^{(k)}$ . As will be shown below, this gradient can be extracted from  $J'(\boldsymbol{\phi}^{(k)}; \boldsymbol{\phi}')$ . A convenient expression for  $J'(\boldsymbol{\phi}^{(k)}; \boldsymbol{\phi}')$  can be found using methods of Nonlinear Programming [see Lewis (2001) for a discussion of NLP techniques in the context of PDE-constrained optimization]

$$J'(\boldsymbol{\phi}; \boldsymbol{\phi}') = \left\langle D_{\boldsymbol{\phi}} J, \boldsymbol{\phi}' \right\rangle_{\mathcal{U} \times \mathcal{U}^*} = \left\langle D_{\boldsymbol{\phi}} \tilde{J}, \boldsymbol{\phi}' \right\rangle_{\mathcal{U} \times \mathcal{U}^*} + \left\langle D_{\mathbf{x}} \tilde{J}, (D_{\boldsymbol{\phi}} \mathbf{x}) \boldsymbol{\phi}' \right\rangle_{\mathcal{X} \times \mathcal{X}^*}, \quad (5)$$

where  $D_a F$  denotes the Fréchet derivative of the mapping  $F = F(a)$  [see Berger (1977)]. In (5)  $\mathcal{U}^*$  is the dual space with respect to  $\mathcal{U}$  and  $\langle \cdot, \cdot \rangle_{\mathcal{U} \times \mathcal{U}^*}$  represents the standard duality pairing between the spaces  $\mathcal{U}$  and  $\mathcal{U}^*$ . Below we will show how the cost functional differential in (5), and in particular the term  $D_{\boldsymbol{\phi}} J$ , can be expressed using an appropriately-defined *adjoint state*. Using the implicit function theorem, the term  $D_{\boldsymbol{\phi}} \mathbf{x}$  can be expressed as

$$D_{\boldsymbol{\phi}} \mathbf{x} = -(D_{\mathbf{x}} u)^{-1} D_{\boldsymbol{\phi}} u, \quad (6)$$

so that the second term on the RHS in (5) can be transformed as follows

$$\begin{aligned} \left\langle D_{\mathbf{x}} \tilde{J}, (D_{\boldsymbol{\phi}} \mathbf{x}) \boldsymbol{\phi}' \right\rangle_{\mathcal{X} \times \mathcal{X}^*} &= - \left\langle D_{\mathbf{x}} \tilde{J}, (D_{\mathbf{x}} u)^{-1} D_{\boldsymbol{\phi}} u \boldsymbol{\phi}' \right\rangle_{\mathcal{X} \times \mathcal{X}^*} \\ &= - \left\langle D_{\boldsymbol{\phi}}^* u (D_{\mathbf{x}} u)^{-*} D_{\mathbf{x}} \tilde{J}, \boldsymbol{\phi}' \right\rangle_{\mathcal{U} \times \mathcal{U}^*} \triangleq \left\langle D_{\boldsymbol{\phi}}^* \mathbf{x} D_{\mathbf{x}} \tilde{J}, \boldsymbol{\phi}' \right\rangle_{\mathcal{U} \times \mathcal{U}^*}, \end{aligned} \quad (7)$$

where an asterisk denotes a Banach space adjoint. Putting together (5) and (7) we see that the adjoint operator  $D_{\boldsymbol{\phi}}^* \mathbf{x} : \mathcal{X}^* \rightarrow \mathcal{U}^*$  can be used to express the differential of the cost functional (5) in a convenient form as

$$J'(\boldsymbol{\phi}; \boldsymbol{\phi}') = \left\langle D_{\boldsymbol{\phi}} \tilde{J} + D_{\boldsymbol{\phi}}^* \mathbf{x} D_{\mathbf{x}} \tilde{J}, \boldsymbol{\phi}' \right\rangle_{\mathcal{U} \times \mathcal{U}^*} = \left\langle D_{\boldsymbol{\phi}} J, \boldsymbol{\phi}' \right\rangle_{\mathcal{U} \times \mathcal{U}^*}. \quad (8)$$

As is evident from the above relationship, the first argument in the duality pairing can be identified with the *gradient* of the reduced cost functional  $J : \mathcal{U} \rightarrow \mathbb{R}$  in the metric induced by the space  $\mathcal{U}$ . It must be emphasized that the gradient in fact belongs to the dual space  $D_{\boldsymbol{\phi}} J \in \mathcal{U}^*$  and, since in most infinite-dimensional cases the dual space  $\mathcal{U}^*$  is *not* contained in the original space  $\mathcal{U}$ , this gradient may not be used as a descent direction in  $\mathcal{U}$ . In the special case when  $\mathcal{U}$  is a Hilbert space we can invoke Riesz' representation theorem [Berger (1977)] which allows us to map  $D_{\boldsymbol{\phi}} J \in \mathcal{U}^*$  to the corresponding element  $\nabla J \in \mathcal{U}$  as

$$J'(\boldsymbol{\phi}; \boldsymbol{\phi}') = \left\langle D_{\boldsymbol{\phi}} J, \boldsymbol{\phi}' \right\rangle_{\mathcal{U} \times \mathcal{U}^*} = \left( \nabla J, \boldsymbol{\phi}' \right)_{\mathcal{U}}, \quad (9)$$

where  $(\cdot, \cdot)_{\mathcal{U}}$  represents the inner product on the Hilbert space  $\mathcal{U}$ , so that now  $\nabla \mathcal{J} \in \mathcal{U}$  can be used to construct a descent direction in  $\mathcal{U}$ . On the other hand, when  $\mathcal{U}$  is not a Hilbert space, Riesz' theorem does not apply and identification (9) is not possible. However, in Section 3 we will present a method for constructing an equivalent of  $\nabla \mathcal{J}$  in the space  $\mathcal{U}$  in such a general case.

Now we illustrate these somewhat abstract considerations by analyzing a concrete example of PDE-constrained optimization. We will focus on a model problem introduced in Protas et al. (2004) which concerns estimation of the initial condition for the 1D Kuramoto–Sivashinsky equation. This particular problem is selected as it models the variational data assimilation, known as 4DVAR, in Dynamic Meteorology [see Kalnay (2003)]. The Kuramoto–Sivashinsky equation is chosen, since it is endowed with chaotic and multiscale behavior and as such is an attractive model for the Navier–Stokes system. We follow here Protas et al. (2004) as regards the set-up of this problem and below highlight only the main points of the derivation, while the Reader is referred to the original source for further details.

For simplicity, we will consider the 1D Kuramoto–Sivashinsky equation on a periodic spatial domain  $\Omega = [0, 2\pi]$  and a time interval  $[0, T]$

$$\begin{cases} \partial_t v + 4\partial_x^4 v + \kappa(\partial_x^2 v + v\partial_x v) = 0, & x \in \Omega, \quad t \in [0, T], \\ \partial_x^i v(0, t) = \partial_x^i v(2\pi, t), & t \in [0, T], \quad i = 0, \dots, 3, \\ v(x, 0) = \phi, & x \in \Omega. \end{cases} \quad (10)$$

Given incomplete and possibly noisy measurements  $y = \mathcal{H}v_{act} + \eta \in \mathcal{Y}$ , where  $v_{act}(\cdot, t) \in \mathcal{X}$  is the actual system trajectory,  $\mathcal{H} : \mathcal{X} \rightarrow \mathcal{Y}$  is an observation operator and  $\eta$  is (Gaussian) noise, our optimization problem consists in finding an initial condition  $\phi$  in (10) such that the corresponding system trajectory best matches the available measurements  $y$ . In other words, we minimize the following cost functional

$$\mathcal{J}(\phi) = \frac{1}{2} \left\| \mathcal{H}v(\phi) - y \right\|_{L_2(0, T; L_2(\Omega))}^2 = \frac{1}{2} \int_0^T (\mathcal{H}v(\phi) - y)^2 d\tau. \quad (11)$$

Consistently with the properties of system (10), we will assume that  $\phi \in \mathcal{U} = L_2(\Omega)$ . Since  $\mathcal{J}$  depends on the control variable  $\phi$  implicitly through the state equation (10), expression (11) represents in fact the reduced cost functional [cf. (2)]. We will assume that the observation operator  $\mathcal{H}$  has the form of projection on a set of cosine modes with the wavenumbers in some set  $\Lambda_r$ , i.e.

$$\mathcal{H} = \sum_{r \in \Lambda_r} \mathcal{P}_r, \quad \text{where} \quad \mathcal{P}_r z = \left[ \frac{1}{\pi} \int_0^{2\pi} \cos(rx') z(x') dx' \right] \cos(rx). \quad (12)$$

The Gâteaux differential of (11) is given by [cf. (5)]

$$\mathcal{J}'(\phi; \phi') = \int_0^T \int_0^{2\pi} (\mathcal{H}v - y) \mathcal{H}v' dx dt, \quad (13)$$

where the perturbation  $v'(\phi; \phi')$  is obtained by solving the Kuramoto–Sivashinsky equation linearized around the state  $\phi$ , i.e.

$$\begin{cases} \mathcal{L}v' = \partial_t v' + 4\partial_x^4 v' + \kappa [\partial_x^2 v' + v \partial_x v' + (\partial_x v) v'] = 0, & x \in \Omega, t \in [0, T], \\ \partial_x^i v'(0, t) = \partial_x^i v'(2\pi, t), & t \in [0, T], i = 0, \dots, 3, \\ v'(x, 0) = \phi', & x \in \Omega, \end{cases} \quad (14)$$

with the operator  $\mathcal{L} : \mathcal{X} \rightarrow \mathcal{X}^*$  understood in the weak sense. Relation (13) can now be transformed to a form consistent with (8) by introducing an adjoint operator  $\mathcal{L}^* : \mathcal{X} \rightarrow \mathcal{X}^*$  and the corresponding adjoint state  $v^* \in \mathcal{X}^*$  via the following identity

$$\langle v^*, \mathcal{L}v' \rangle_{\mathcal{X} \times \mathcal{X}^*} = \langle \mathcal{L}^* v^*, v' \rangle_{\mathcal{X} \times \mathcal{X}^*} + b_{\mathcal{L}}. \quad (15)$$

Using integration by parts and the definition of  $\mathcal{L}$  in (14), we obtain

$$\begin{aligned} \mathcal{L}^* v^* &= -\partial_t v^* + 4\partial_x^4 v^* + \kappa (\partial_x^2 v^* - v \partial_x v^*), \quad \text{and} \\ b_{\mathcal{L}} &= \left[ \int_0^{2\pi} v^* v' dx \right]_{t=0}^{t=T}. \end{aligned} \quad (16)$$

We remark that  $b_{\mathcal{L}}$  does not contain any boundary terms (resulting from integration by parts), since all of them vanish due to periodicity. Defining an adjoint system as

$$\begin{cases} \mathcal{L}^* v^* = \mathcal{H}^*(\mathcal{H}v - y), & x \in \Omega, t \in [0, T], \\ \partial_x^i v^*(0, t) = \partial_x^i v^*(2\pi, t), & t \in [0, T], i = 0, \dots, 3, \\ v^*(x, T) = 0, & x \in \Omega, \end{cases} \quad (17)$$

and using (14), (15) and (16) we can now express the Gâteaux differential (13) in the desired form (5)

$$j'(\phi; \phi') = \int_0^{2\pi} v^* \Big|_{t=0} \phi' dx. \quad (18)$$

Thus, this differential (i.e., the sensitivity of the cost functional  $j$  with respect to perturbations of the initial condition) can be expressed using the solution of the adjoint system (17).

Relationship (18) can now be employed to extract the gradient required in a descent optimization algorithm. Since  $\mathcal{U} = L_2(0, 2\pi)$ , we immediately obtain

$$j'(\phi; \phi') = \int_0^{2\pi} v^* \Big|_{t=0} \phi' dx = \left( \nabla^{L_2} j, \phi' \right)_{L_2(\Omega)} \implies \nabla^{L_2} j = v^* \Big|_{t=0}. \quad (19)$$

Despite its simplicity, in many cases this is not an optimal choice, as it may result in poor conditioning of the corresponding discrete optimization problem. In Protas

et al. (2004) a set of regularization options was identified which can, at least partially, alleviate some of such difficulties. In relation to gradient extraction it was shown that it can be beneficial to extract the cost functional gradient in a more general Hilbert space, Sobolev spaces being natural candidates [see also Neuberger (1997)]. In particular, gradient extraction was considered in the Sobolev space  $H^1(\Omega)$  characterized by the inner product

$$\left( z_1, z_2 \right)_{H^1(\Omega)} = \frac{1}{(1+l_2^2)} \int_0^{2\pi} \left[ z_1 z_2 + l_2^2 (\partial_x z_1) (\partial_x z_2) \right] dx, \quad (20)$$

where  $l_2$  is an adjustable length-scale. Identification  $J'(\phi; \phi') = \left( \nabla^{H^1} J, \phi' \right)_{H^1(\Omega)}$  yields, after integration by parts, the gradient  $\nabla^{H^1} J$  defined via solutions of the following Helmholtz boundary value problem

$$\begin{cases} \frac{1}{1+l_2^2} [1 - l_2^2 \partial_x^2] \nabla^{H^1} J = \bar{v}^* \Big|_{t=0}, \\ \nabla^{H^1} J(0) = \nabla^{H^1} J(2\pi). \end{cases} \quad (21)$$

Thus, the Sobolev space gradient  $\nabla^{H^1} J$  is obtained by applying the inverse Helmholtz operator to the classical  $L_2$  gradient. Interestingly, when regarded in Fourier space, the inverse Helmholtz operator is equivalent to a low-pass filter with the cut-off given by the inverse of the length-scale  $l_2$  parametrizing the inner product (20). Consequently, extracting gradients in Sobolev spaces with inner products given by (20) has the effect of de-emphasizing components with characteristic length-scales smaller than  $l_2$ . As was shown in Protas et al. (2004), adjusting this length-scale during solution of an optimization problem can accelerate convergence of iterations. In particular, starting with  $l_2$  large and then progressively decreasing it to zero results in a multiscale procedure targeting first the large-scale structures and then homing in on smaller scale components of the solution  $\phi_{opt}$ .

### 3 Nonlinear Preconditioning using Descent Directions in General Banach Spaces

In this Section we address the issue of gradient extraction in general Banach spaces and the potential advantage this technique may offer as a method of nonlinear preconditioning. Similar ideas were already discussed by Lewis (2001) and elaborated in greater detail by Neuberger (1997), however, they were not concerned with preconditioning nonlinear optimization problems. The present approach relies on the assumption that the Banach space  $\mathcal{U}$ , where the descent direction is to be identified, be *reflexive*, i.e., that  $\mathcal{U}^{**} = \mathcal{U}$ . As already mentioned in Section 2 in relation

to formula (9), the gradient is a linear functional on the space  $\mathcal{U}$  and therefore belongs to the dual space  $\mathcal{U}^*$ . For example, if  $\mathcal{U}$  is the Sobolev space  $W_0^{1,p}$ ,  $p \neq 2$ , defined as

$$W_0^{1,p}(\Omega) = \{u : \Omega \rightarrow \mathbb{R}, \|u\|_{W^{1,p}} < \infty, v|_{\partial\Omega} = 0\}, \quad (22)$$

$$\text{where } \|u\|_{W^{1,p}} = \left[ \int_{\Omega} (|v|^p + l_p^p |\partial_x v|^p) d\Omega \right]^{1/p},$$

where  $l_p \in \mathbb{R}^+$  is a weight, then the dual space  $\mathcal{U}^* = W^{-1,q}$ , where  $\frac{1}{p} + \frac{1}{q} = 1$  [see Adams and Fournier (2005)]. Since a dual space is usually “larger”, its elements do not necessarily belong to the original space  $\mathcal{U}$  and therefore cannot be used to represent descent directions in that space. Consequently, it is necessary to propose a different approach which allows one to extract a descent direction  $\tilde{\mathbf{g}}$  from  $\mathcal{J}'(\phi; \phi')$  such that  $\tilde{\mathbf{g}} \in \mathcal{U}$ . As shown by Lewis (2001) and Neuberger (1997), this can be done defining  $\tilde{\mathbf{g}}$  as a unit-norm element of  $\mathcal{U}$  which minimizes expression (9). In other words, we postulate to find  $\tilde{\mathbf{g}}$  as a solution of the following constrained minimization problem

$$\min_{\|\mathbf{g}\|_{\mathcal{U}}=1} \left\langle D_{\phi} \mathcal{J}, \mathbf{g} \right\rangle_{\mathcal{U}^* \times \mathcal{U}} \quad (23)$$

which can be converted to the more convenient unconstrained form

$$\min_{\mathbf{g} \in \mathcal{U}} \left[ \left\langle D_{\phi} \mathcal{J}, \mathbf{g} \right\rangle_{\mathcal{U}^* \times \mathcal{U}} + \frac{\mu}{p} \|\mathbf{g}\|_{\mathcal{U}}^p \right] = \min_{\mathbf{g} \in \mathcal{U}} \mathcal{G}(\mathbf{g}), \quad (24)$$

where  $p$  is an integer,  $\mu$  is the Lagrange multiplier and  $\mathcal{G} : \mathcal{U} \rightarrow \mathbb{R}$ . This problem can be solved with a method analogous to the approach described earlier in Section 2. Thus, the descent direction  $\tilde{\mathbf{g}}$  is characterized by the vanishing of the Gâteaux differential of (24), i.e.

$$\forall \mathbf{g}' \in \mathcal{U} \quad \mathcal{G}'(\tilde{\mathbf{g}}; \mathbf{g}') = \left\langle D_{\mathbf{g}} \mathcal{G}(\tilde{\mathbf{g}}), \mathbf{g}' \right\rangle_{\mathcal{U}^* \times \mathcal{U}} = 0, \quad (25)$$

where  $D_{\mathbf{g}} \mathcal{G} : \mathcal{U} \rightarrow \mathcal{U}^*$ . Thus, we obtain

$$D_{\mathbf{g}} \mathcal{G}(\tilde{\mathbf{g}}) = 0 \quad \text{in } \mathcal{U}^* \quad (26)$$

as an equation determining the direction  $\tilde{\mathbf{g}} \in \mathcal{U}$ . Below we will show how this direction can be determined when  $\mathcal{U}$  is one of the Banach spaces commonly arising in the analysis on nonlinear PDEs. This analysis will be carried out in the setting of the optimization problem for the Kuramoto–Sivashinsky equation introduced in Section 2. We begin with the Lebesgue spaces  $L_p(\Omega)$  with norms defined as [Adams and Fournier (2005)]

$$\|u\|_{L_p(\Omega)} = \begin{cases} \left( \int_{\Omega} |u|^p d\Omega \right)^{1/p} & 1 \leq p < \infty, \\ \text{ess sup}_{x \in \Omega} |u| & p = \infty. \end{cases} \quad (27)$$



Considering for the moment the case with  $1 \leq p < \infty$ , the unconstrained cost functional (24) and its Gâteaux differential (25) take the form

$$\mathcal{G}(g) = \int_{\Omega} (v^*|_{t=0}g + \frac{\mu}{p}|g|^p) d\Omega, \quad (28a)$$

$$\forall_{g' \in L_p(\Omega)} \mathcal{G}'(g; g') = \int_{\Omega} (v^*|_{t=0}g + \mu g|g|^{(p-2)})g' d\Omega, \quad (28b)$$

so that the descent direction  $\tilde{g}_{L_p}$  is characterized by the algebraic relation

$$\tilde{g}|\tilde{g}|^{(p-2)} = -\frac{1}{\mu}v^*|_{t=0}. \quad (29)$$

The solution of (29) is

$$\tilde{g}_{L_p} = \begin{cases} p^{-1}\sqrt[p-1]{-\frac{1}{\mu}v^*|_{t=0}}, & p \text{ — even,} \\ -\operatorname{sgn}(v^*|_{t=0}) p^{-1}\sqrt[p-1]{\frac{1}{\mu}|v^*|_{t=0}|}, & p \text{ — odd.} \end{cases} \quad (30)$$

We thus see that when  $p \neq 2$ , the descent direction in  $L_p(\Omega)$  is obtained by applying a *nonlinear* transformation to the original gradient  $\nabla^{L_2} \mathcal{J} = v^*|_{t=0}$ . In the special case  $p = 2$  we immediately obtain

$$\tilde{g}_{L_2} = -\frac{1}{\mu}v^*|_{t=0} \quad (31)$$

which is the classical expression obtained in Section 2. As regards the constant  $\mu$ , which serves as a Lagrange multiplier in the unconstrained formulation (24), it is chosen to normalize  $\tilde{g}$  to unit norm  $\|\tilde{g}\|_{\mathcal{U}} = 1$ . In the second special case  $p = \infty$ , it can be shown that

$$\tilde{g}_{L_\infty} = -\operatorname{sgn}(v^*|_{t=0}) \quad (32)$$

consistently with taking the limit  $p \rightarrow \infty$  in expressions (30). We also remark that in the case  $p = 1$  the descent direction  $\tilde{g}_{L_1}$  cannot be defined, since the space  $L_1(\Omega)$  is not reflexive.

We now proceed to discuss the problem of determining the descent direction when  $\mathcal{U} = W^{1,p}$ , where  $W^{1,p}$  is the Sobolev space defined in (22). Considering the case  $1 \leq p < \infty$ , the unconstrained cost functional and its Gâteaux differential take the form

$$\mathcal{G}(g) = \int_{\Omega} \left[ v^*|_{t=0}g + \frac{\mu}{p}(|g|^p + I_p^p |\partial_x g|^p) \right] d\Omega, \quad (33a)$$

$$\mathcal{G}'(g; g') = \int_{\Omega} \left\{ v^*|_{t=0}g' + \frac{\mu}{p} \left[ g|g|^{(p-2)}g' - I_p^p \partial_x (\partial_x g |\partial_x g|^{(p-2)})g' \right] \right\} d\Omega. \quad (33b)$$

As before, the boundary terms due to integration by parts vanish because of periodicity of  $\tilde{g}$  and  $g'$ . Since the descent direction  $\tilde{g}$  is characterized by  $\mathcal{G}'(\tilde{g}; g') = 0$ ,  $\forall_{g' \in \mathcal{U}}$ , it can be obtained as a solution of the following problem in  $\mathcal{U}^*$  (due to nondifferentiability of the absolute value function  $|\cdot|$ , this equations is formulated in the weak sense)

$$\begin{cases} \tilde{g}|\tilde{g}|^{(p-2)} - l_p^p \partial_x \left[ (\partial_x \tilde{g}) |\partial_x \tilde{g}|^{(p-2)} \right] = -\frac{1}{\mu} v^*|_{t=0}, \\ \tilde{g}(0) = \tilde{g}(2\pi). \end{cases} \quad (34)$$

The second term on the LHS in the first equation in (34) is usually referred to as the *p-Laplacian* [Neuberger (1997)], as it represents a nonlinear generalization of the familiar Laplace operator. Evidently, in the case  $p = 2$ ,  $W^{1,2}(\Omega) = H^1(\Omega)$  is a Hilbert space and the *p-Laplacian* reduces to the classical Laplace operator. As a result, (34) simplifies to (21) and we recover the Hilbert space framework discussed in Section 2. The Lagrange multiplier  $\mu$  can be adjusted in order to normalize the descent direction to the unit norm. We can conclude that identification of descent directions in Banach spaces (such as  $L_p(\Omega)$ , or  $W^{1,p}(\Omega)$ , when  $p \neq 2$ ) results in *nonlinear* transformations of the adjoint field  $v^*|_{t=0}$ . As regards equations with the *p-Laplacian* operator such as (34), a variety of their interesting properties is discussed by Ishii and Loreti (2005) (see also references contained therein).

We now comment briefly on the utility of extraction descent directions in general Banach spaces as a nonlinear preconditioning technique. The purpose of preconditioning is to modify the metric in which a given iterative process takes place, so as to accelerate convergence. For linear problems with quadratic functionals this can also be regarded as decreasing the condition number of the Hessian of the reduced cost functional. In such cases linear preconditioning techniques are efficient enough (in fact, in many situations there exist specific guidelines regarding the choice of an optimal preconditioner). However, for nonlinear problems linear preconditioning may not be sufficient and a nonlinear change of the metric may lead to better results. In the framework proposed here, choosing a preconditioner is in fact equivalent to choosing a Banach space  $\mathcal{U}$  in which the descent direction is identified. The question of how to choose this space is important. Unlike certain linear problems, most nonlinear PDEs result in optimization problems with structure that is too complicated to allow for a thorough analysis. In such situations finding the most suitable preconditioning strategy is a matter of experimentation. There are, however, certain general conditions that need to be satisfied. In general, for the evolution equation (10) to be well-posed, the control  $\phi$  must belong to some appropriate space  $\mathcal{U}$  (identical with  $L_2(\Omega)$  in the present case). Therefore, if at the  $k$ -th iteration we want to precondition the gradient by extracting it in some Banach space  $\mathcal{U}^{(k)}$ , it must be ensured that this gradient  $\nabla^{\mathcal{U}^{(k)}} J$  will still

belong to the original space  $\mathcal{U}$ , in other words  $\mathcal{U}^{(k)} \subseteq \mathcal{U}, \forall k \in \mathbb{Z}$ . Such preconditioning is equivalent to restricting the iterates to a family of subspaces nested in  $\mathcal{U}$ . Computational results concerning linear preconditioning reported by Protas et al. (2004) indicate that the best results were obtained when the subspaces formed the following hierarchy

$$\mathcal{U}^{(0)} \subseteq \mathcal{U}^{(1)} \subseteq \dots \subseteq \mathcal{U}^{(k)} \subseteq \dots \subseteq L_2(\Omega). \quad (35)$$

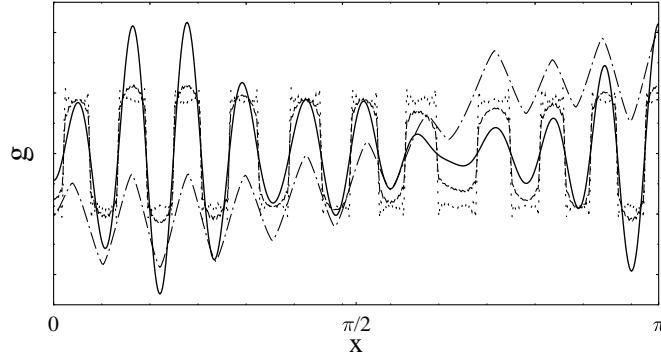
When considering general Banach spaces, additional guidance for constructing hierarchies like (35) can be obtained by considering the family of *Sobolev Imbedding Theorems* [Adams and Fournier (2005)]. Imbedding Theorems provide criteria that allow one to determine whether or not one Sobolev (or Lebesgue) space is “contained” in another one. In the following Section we present computational results that address some of these issues.

## 4 Computational Results

In this Section we show some computational results illustrating the utility of the nonlinear preconditioning techniques developed in Section 3. We treat the results from Protas et al. (2004) as our point of reference, so we consider here precisely the same problem of state estimation for the Kuramoto–Sivashinsky equation. The observation operator in (12) uses projections on the first  $r = 50$  cosine modes (i.e.,  $\Lambda_r = \{1, \dots, 50\}$ ) and we set  $\kappa = 4000$  in equation (10). Both equation (10) and its adjoint (17) are solved using a dealiased pseudo–spectral Fourier–Galerkin method with  $N = 1024$  grid points. The Reader is referred to Protas et al. (2004) for further numerical details. The nonlinear equation (34) involving the  $p$ –Laplacian operator is solved using Newton’s method applied to the system of nonlinear algebraic equations obtained after discretization.

In order to see the effect of nonlinear preconditioning we will present results obtained for two optimization horizons (given in terms of the time step  $\Delta t = 10^{-8}$ )  $T = 300$  and  $T = 500$ . Since the effect of nonlinear preconditioning appears most pronounced for the longer optimization horizon  $T = 500$ , some of the results will be presented for that case only. We begin presentation of the results by examining the shape of the descent directions  $\tilde{g}$  obtained in different Banach spaces. To fix attention, we consider the first iteration in the problem with  $T = 500$  with a zero initial guess  $\phi^{(0)} \equiv 0$ . In Figure 1 we compare the descent directions  $\tilde{g}$  extracted in the Banach spaces  $L_2(\Omega)$ ,  $L_\infty(\Omega)$  and  $W^{1,4}(\Omega)$  with  $l_4 = 10.0$  [cf. (22)] with the standard gradient  $\nabla^{L_2} j$  extracted in the space  $L_2(\Omega)$ . We observe that for increasing  $p$  the descent directions obtained in the spaces  $L_p(\Omega)$  approach a square wave.

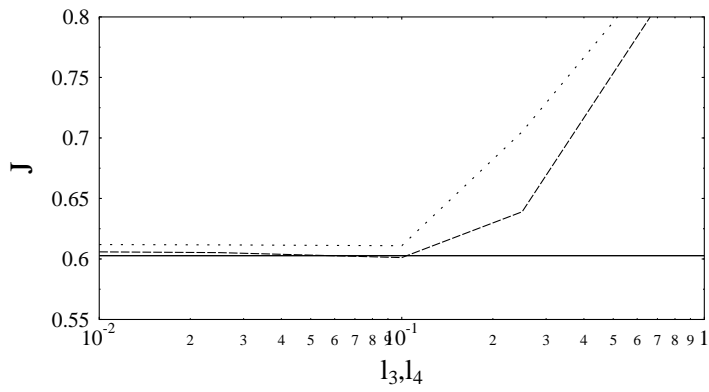
In computational solution of our optimization problem we found the preconditioning involving descent directions in the Sobolev spaces  $W^{1,p}(\Omega)$ , where  $p \geq 3$ ,



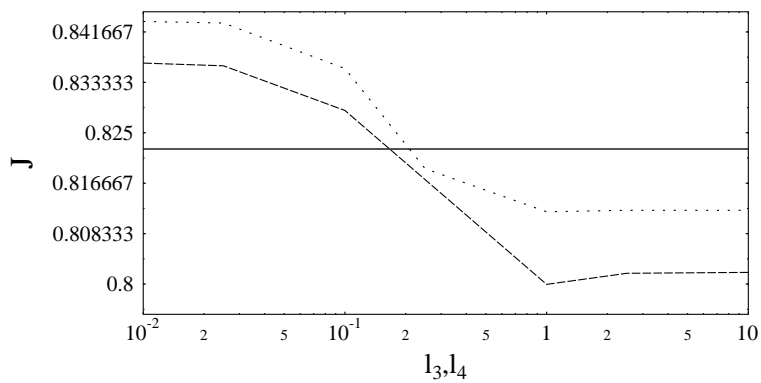
**Figure 1.** Shapes of the descent directions obtained at the first iteration in the data assimilation problem with  $T = 500$  and determined in: (solid line)  $L_2(\Omega)$ , (dashed line)  $L_5(\Omega)$ , (dotted line)  $L_\infty(\Omega)$  and (dash-dotted line)  $W^{1,4}(\Omega)$ . For clarity, only half of the domain  $\Omega$  is shown.

to be more efficient than the preconditioning using descent directions in  $L_p(\Omega)$ . This is the case we will focus on exclusively below. In Figures 2a and 2b, corresponding to optimization with  $T = 300$  and  $T = 500$ , we study the effect that the quantity  $l_p$ , the characteristic “length-scale” parametrizing the definitions of the norm  $\|u\|_{W^{1,p}}$ , has on optimization efficiency. Given that the value of the cost function (11) before optimization (i.e., for  $\phi^{(0)} \equiv 0$ ) is normalized to unity, Figures 2a and 2b show the decrease of the cost functional at the first iteration for the descent directions obtained in the spaces  $W^{1,3}(\Omega)$  and  $W^{1,4}(\Omega)$  with values of  $l_3$  and  $l_4$  indicated on the abscissa. For comparison, we also show the results obtained with the gradient extracted in  $L_2$ . We note that the decrease of the cost functional significantly depends on the choice of  $l_p$  ( $p = 3, 4$ ). For  $T = 300$  the window of  $l_p$  giving improvement over optimization with the  $L_2$  gradients exists for descent directions in  $W^{1,3}(\Omega)$  only and is rather narrow. The advantage of determining descent directions in a Banach space becomes much more evident for  $T = 500$ , where the windows of  $l_p$  giving improvement over gradients in  $L_2$  are unbounded.

Now we proceed to analyze the effect of nonlinear preconditioning on the whole optimization process involving many iterations. As regards optimization with the descent directions obtained in the Banach spaces  $W^{1,p}(\Omega)$  we follow the strategy outlined in Section 3: for a given choice of the space  $W^{1,p}(\Omega)$  we start with the value of  $l_p$  which was determined to give the best results at the first iter-



(a)



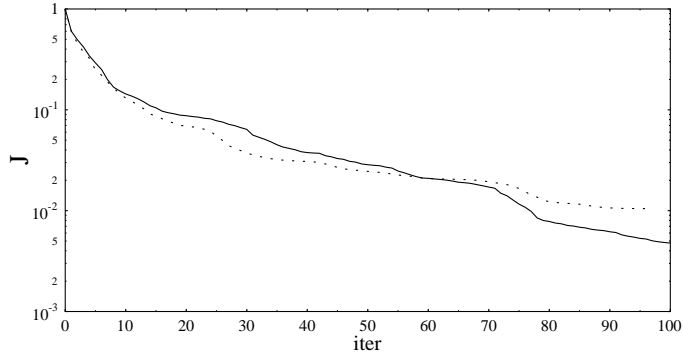
(b)

**Figure 2.** Decrease of the cost functional (11) at the first iteration in optimization with (a)  $T = 300$  and (b)  $T = 500$ . The descent directions are extracted in the spaces (solid line)  $L_2(\Omega)$ , (dashed line)  $W^{1,3}(\Omega)$  and (dotted line)  $W^{1,4}(\Omega)$  for values of  $l_3$  and  $l_4$  indicated on the abscissa.

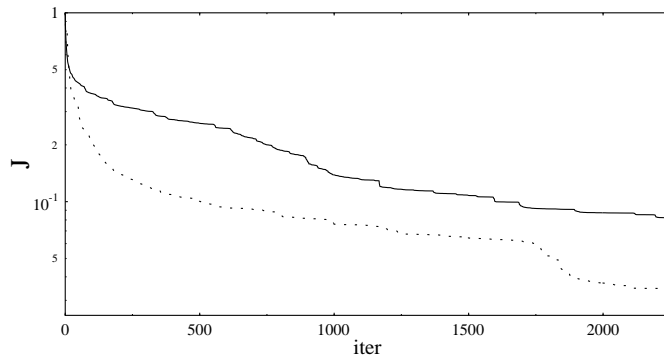
ation and then progressively decrease it to zero, so that the corresponding descent directions approach the  $L_p$  descent direction. As a result, our preconditioning strategy is equivalent to extracting the descent directions in a sequence of nested spaces  $W^{1,p}(\Omega)$ , all contained in the “master” space  $L_2(\Omega)$ . This strategy, initially investigated by Protas et al. (2004) for the case of gradient extraction in the Sobolev space  $H^1(\Omega) = W^{1,2}(\Omega)$ , was found to give good results. In Figures 3a and 3b we show the decrease of the cost functional  $\mathcal{J}(\phi^{(k)})$  as a function of the iteration count  $k$  for  $T = 300$  and  $T = 500$ , respectively. In both cases the descent directions are extracted in the spaces  $W^{1,4}(\Omega)$  with the initial values of  $l_4$  equal to  $10^{-1}$  and 10 in the two cases, respectively. For comparison, in the two Figures we also show the decrease of the cost functional obtained with gradients obtained in the space  $L_2(\Omega)$ . We note that when  $T = 300$  nonlinear preconditioning offers little advantage over the unpreconditioned case, in contrast to the case with  $T = 500$  where a significant convergence acceleration is observed. In order to further emphasize this point in Figures 4a and 4b we show the data for the error in the reconstruction of the initial condition, i.e.,  $\|\phi^{(k)} - \phi_{act}\|_{L_2(\Omega)}$  corresponding to the same cases as in Figure 3. We note that these results provide further evidence for the trends already shown in Figure 3. We also examined nonlinear preconditioning in the case of shorter optimization horizons  $T \ll 300$ , however, no acceleration of convergence comparing to the optimization with the  $L_2$  gradients was observed. Hence, we do not show these results here.

## 5 Conclusions and Outlook

In this paper we first reviewed the formulation of an optimal control problem for a fluid system using the language of Nonlinear Programming. We focused on a particular aspect relevant from the computational point of view, namely, determination of well-preconditioned descent directions for the cost functional. We extended an earlier approach and showed how a descent direction can be determined in a general Banach space without Hilbert structure. In particular, we showed that extracting this descent direction in a Sobolev space  $W^{1,p}(\Omega)$  leads to solution of an elliptic problem with a  $p$ -Laplacian. Such a preconditioning strategy has the effect of a nonlinear change of the metric in the space where optimization is performed. When employed judiciously, this approach may have the potential to mitigate the effect of nonlinearities present in the system. Indeed, our computational results indicate that such a nonlinear preconditioning can accelerate convergence of iterations in an optimization problem for a nonlinear PDE. Interestingly, effectiveness of the proposed approach increases with the length of the optimization interval  $[0, T]$  and becomes more evident for problems with large  $T$ , i.e., in situations when nonlinear effects play a more significant role. Research is underway to apply a similar approach to precondition optimization of more realistic problems,

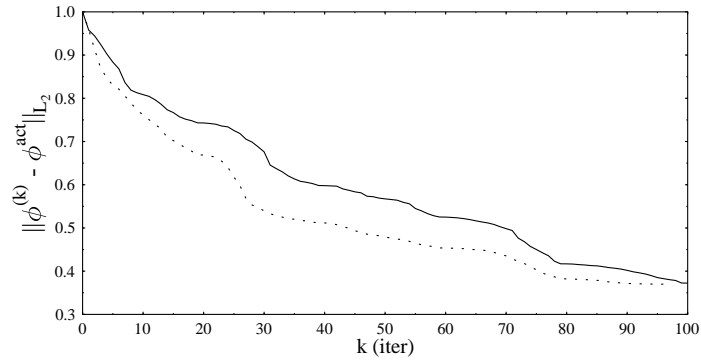


(a)

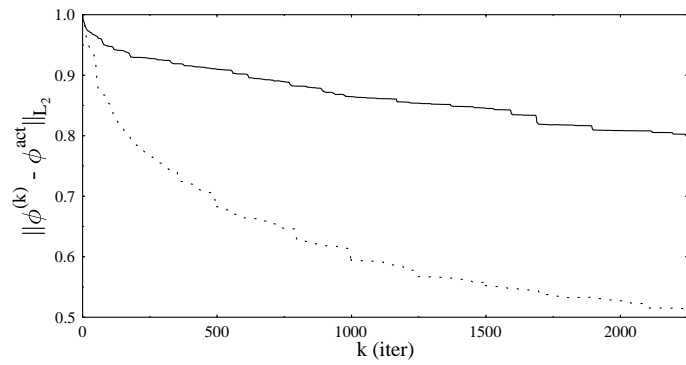


(b)

**Figure 3.** Decrease of the cost functional  $\mathcal{J}(\phi^{(k)})$  in function of iterations for optimizations with (a)  $T = 300$  and (b)  $T = 500$ . The results are obtained with the descent directions in (solid line)  $L_2$  and (dotted line)  $W^{1,4}(\Omega)$  where the parameter  $l_4$  progressively decreases with iterations.



(a)



(b)

**Figure 4.** Decrease of the error of the reconstruction of the initial condition  $\|\phi^{(k)} - \phi_{act}\|_{L_2(\Omega)}$  in function of iterations for optimizations with (a)  $T = 300$  and (b)  $T = 500$ . The results are obtained with the descent directions in (solid line)  $L_2$  and (dotted line)  $W^{1,4}(\Omega)$  where the parameter  $l_4$  progressively decreases with iterations.



such as the state estimation in a 3D turbulent channel flow already investigated by Bewley and Protas (2004). Another possibility is to investigate descent directions in more general Banach spaces and here Besov spaces [see, e.g., Adams and Fournier (2005)] are attractive candidates. A thorough treatment of this subject is given in Protas (2008).

In the present investigation the space giving “optimal” preconditioning was chosen by trial and error. A very challenging theoretical question is to develop a rigorous procedure that will determine guidelines for choosing such an optimal space. Such procedures are in fact available for certain optimization problems formulated for some linear PDEs, however, no such results appear available for nonlinear PDEs. Encouraging computational results reported in the present paper may therefore serve to motivate further theoretical research in this direction.

## 6 Acknowledgments

This research has been supported by an NSERC (Canada) Discovery Grant.

## Bibliography

- F. Abergel and R. Temam. On some control problems in fluid mechanics. *Theoretical and Computational Fluid Dynamics*, 1:303–325, 1990.
- R. A. Adams and J. J. F. Fournier. *Sobolev Spaces*. Elsevier, second edition edition, 2005.
- M. S. Berger. *Nonlinearity and Functional Analysis*. Academic Press, 1977.
- T. R. Bewley and B. Protas. Skin friction and pressure: the “footprints” of turbulence. *Physica D*, 196:28–44, 2004.
- T. R. Bewley, P. Moin, and R. Temam. DNS-based predictive control of turbulence: an optimal benchmark for feedback algorithms. *Journal of Fluid Mechanics*, 447:179–225, 2001.
- M. Gunzburger. *Perspectives in Flow Control*. SIAM, 2002.
- V. Isakov. *Inverse Problems for Partial Differential Equations*. Springer, 1997.
- H. Ishii and P. Loreti. Limits of solutions of  $p$ -laplace equations as  $p$  goes to infinity and related variational problems. *SIAM J. Math. Anal.*, 37:411–437, 2005.
- E. Kalnay. *Atmospheric modeling, data assimilation and predictability*. Cambridge University Press, 2003.
- R. M. Lewis. A nonlinear programming perspective on sensitivity calculations for systems governed by state equations. Technical report, ICASE, 2001.
- J. L. Lions. *Contrôle optimal de systèmes gouvernés par des équations aux dérivées partielles*. Dunod, 1968.

- J. R. R. A. Martins, J. J. Alonso, and J. J. Reuther. High-fidelity aerostructural design optimization of a supersonic business jet. *Journal of Aircraft*, 41:523–530, 2004.
- B. Mohammadi and O. Pironneau. *Applied Shape Optimization for Fluids*. Oxford University Press, 2001.
- J. Neuberger. *Sobolev Gradients and Differential Equations*. Springer, 1997.
- B. Protas. Adjoint-based optimization of pde systems with alternative gradients. *Journal of Computational Physics*, 227:6490–6510, 2008.
- B. Protas and A. Styczek. Optimal rotary control of the cylinder wake in the laminar regime. *Physics of Fluids*, 14:2073–2087, 2002.
- B. Protas, T. Bewley, and G. Hagen. A comprehensive framework for the regularization of adjoint analysis in multiscale pde systems. *Journal of Computational Physics*, 195:49–89, 2004.