

Quantitative Stability Analysis of Optimal Solutions in PDE-Constrained Optimization

K. Brandes, R. Griesse

RICAM-Report 2006-26

QUANTITATIVE STABILITY ANALYSIS OF OPTIMAL SOLUTIONS IN PDE-CONSTRAINED OPTIMIZATION

KERSTIN BRANDES AND ROLAND GRIESSE

ABSTRACT. PDE-constrained optimization problems under the influence of perturbation parameters are considered. A quantitative stability analysis for local optimal solutions is performed. The perturbation directions of greatest impact on an observed quantity are characterized using the singular value decomposition of a certain linear operator. An efficient numerical method is proposed to compute a partial singular value decomposition for discretized problems, with an emphasis on infinite-dimensional parameter and observation spaces. Numerical examples are provided.

1. INTRODUCTION

In this work we consider nonlinear infinite-dimensional equality-constrained optimization problems, subject to a parameter p in the problem data:

$$\min_x f(x, p) \quad \text{subject to } e(x, p) = 0. \quad (1.1)$$

The optimization variable x and the parameter p are in some Banach and Hilbert spaces, respectively, and f and e are twice continuously differentiable. In particular, we have in mind optimal control problems for partial differential equations (PDE). When solving practical optimal control problems which describe the behavior of physical systems, uncertainty in the physical parameters is virtually unavoidable. In (1.1), the uncertain data is expressed in terms of a parameter p for which a nominal or expected value p_0 is available but whose actual value is unknown. Having solved problem (1.1) for $p = p_0$, it is thus natural and sometimes crucial to assess the stability of the optimal solution with respect to unforeseen changes in the problem data.

In this contribution we quantify the first-order stability properties of a local optimal solution of (1.1), and more generally, the stability properties of an observed quantity depending on the solution. We make use of the singular value decomposition (SVD) for compact operators. Moreover, we propose a practical and efficient procedure to approximate the corresponding singular system. The right singular vectors corresponding to the largest singular values represent the perturbation directions of greatest impact on the observed quantity. The singular values themselves provide an upper bound for the influence of unit perturbations. Altogether, this information allows practitioners to assess the stability properties of any given optimal solution, and to avoid the perturbations of greatest impact.

Let us briefly relate our effort to previous results in the field. The differentiability properties of optimal solutions with respect to p in the context of PDE-constrained optimization were studied in, e.g., [4, 10]. The impact of *given* perturbations on

Date: May 18, 2006.

2000 Mathematics Subject Classification. 49K40, 65F15, 90C31.

Key words and phrases. PDE-constrained optimization, parametric sensitivity analysis, stability, singular value decomposition.

optimal solutions and the optimal value of the objective has also been discussed there. For the dependence of a scalar quantity of interest on perturbations we refer to [6]. All of these results admit pointwise inequality constraints for the control variable. For simplicity of the presentation, we elaborate on the case without inequality constraints. However, our results extend to problems with inequality (control) constraints in the presence of strict complementarity, see Remark 3.6.

The material is organized as follows: In Section 2, we perform a first order perturbation analysis of solutions for (1.1) in the infinite-dimensional setting of PDE-constrained optimization, and discuss their stability properties using the singular value decomposition of a certain compact linear map. In Section 3 we focus on the discretized problem and propose a practical and efficient method to compute the most significant part of the singular system. Finally, we present numerical examples in Section 4.

For normed linear spaces X and Y , $\mathcal{L}(X, Y)$ denotes the space of bounded linear operators from X into Y . The standard notation $L^p(\Omega)$ and $H^1(\Omega)$ for Sobolev spaces is used, see [1].

2. INFINITE-DIMENSIONAL PERTURBATION ANALYSIS

As mentioned in the introduction, we are mainly interested in the analysis of optimal control problems involving PDEs. Hence we re-state problem (1.1) as

$$\min_{y,u} f(y, u, p) \quad \text{subject to } e(y, u, p) = 0 \quad (2.1)$$

where the optimization variable $x = (y, u)$ splits into a state variable $y \in Y$ and a control or design variable $u \in U$ and where $e : Y \times U \rightarrow Z^*$ represents the weak form of a stationary or non-stationary partial differential equation. Throughout, Y , U and Z are reflexive Banach spaces and Z^* denotes the dual of Z . Problem (2.1) depends on a parameter p taken from a Hilbert space P , which is not optimized for but which represents perturbations or uncertainty in the problem data. We emphasize that p may be finite- or infinite-dimensional.

For future reference, it will be convenient to define the Lagrangian of problem (2.1) as

$$\mathcal{L}(y, u, \lambda, p) = f(y, u, p) + \langle \lambda, e(y, u, p) \rangle. \quad (2.2)$$

The following two results are well known [11]:

Lemma 2.1 (First-Order Necessary Conditions). *Let f and e be continuously differentiable with respect to (y, u) . Moreover, let (y, u) be a local optimal solution for problem (2.1) for some given parameter p . If $e_y(y, u, p) \in \mathcal{L}(Y, Z^*)$ is onto, then there exists a unique Lagrange multiplier $\lambda \in Z$ such that the following optimality system is satisfied:*

$$\mathcal{L}_y(y, u, \lambda, p) = f_y(y, u, p) + \langle \lambda, e_y(y, u, p) \rangle = 0 \quad (2.3)$$

$$\mathcal{L}_u(y, u, \lambda, p) = f_u(y, u, p) + \langle \lambda, e_u(y, u, p) \rangle = 0 \quad (2.4)$$

$$\mathcal{L}_\lambda(y, u, \lambda, p) = e(y, u, p) = 0. \quad (2.5)$$

In the context of optimal control, λ is called the adjoint state. A triple (y, u, λ) satisfying (2.3)–(2.5) is called a critical point.

Lemma 2.2 (Second-Order Sufficient Conditions). *Let (y, u, λ) be a critical point such that $e_y(y, u, p)$ is onto and let f and e be twice continuously differentiable with respect to (y, u) . Suppose that there exists $\rho > 0$ such that $\mathcal{L}_{xx}(y, u, \lambda, p)(\bar{x}, \bar{x}) \geq \rho \|\bar{x}\|_{Y \times U}^2$ holds for all $\bar{x} \in \ker e_x(y, u, p)$. Then (y, u) is a strict local optimal solution of (2.1).*

Let us fix the standing assumptions for the rest of the paper:

Assumption 2.3.

- (1) Let f and e be twice continuously differentiable with respect to (y, u, p) .
- (2) Let p_0 be a given nominal or expected value of the parameter, and let (y_0, u_0) be a local optimal solution of (2.1) for p_0 .
- (3) Suppose that $e_y(y_0, u_0, p_0)$ is onto and that λ_0 is the unique adjoint state.
- (4) Suppose that the second-order sufficient conditions of Lemma 2.2 hold at (y_0, u_0, λ_0) .

Remark 2.4. For the sake of the generality of the presentation, we abstain from using more specific, i.e., weaker, second-order sufficient conditions for optimal control problems with PDEs, see, e.g., [16, 17]. In case the setting of a specific problem at hand requires refined second-order conditions and a careful choice of function spaces, the subsequent ideas still remain valid, compare Example 2.5.

Let us define now the Karush-Kuhn-Tucker (KKT) operator

$$\mathcal{K} = \begin{pmatrix} \mathcal{L}_{yy} & \mathcal{L}_{yu} & e_y^* \\ \mathcal{L}_{uy} & \mathcal{L}_{uu} & e_u^* \\ e_y & e_u & 0 \end{pmatrix} \quad (2.6)$$

where all terms are evaluated at the nominal solution (y_0, u_0, λ_0) and the nominal parameter p_0 , and e_y^* and e_u^* denote the adjoint operators of e_y and e_u , respectively. Note that \mathcal{K} is self-adjoint. Here and in the sequel, when no ambiguity arises, we will frequently omit the function arguments.

Under the conditions of Assumption 2.3, \mathcal{K} is boundedly invertible as an element of $\mathcal{L}(Y \times U \times Z, Y^* \times U^* \times Z^*)$.

Example 2.5 (Optimal Control of the Stationary Navier-Stokes System). *As mentioned in Remark 2.4, nonlinear PDE-constrained problems may require refined second-order sufficient conditions. Consider, for instance, the distributed optimal control problem for the stationary Navier-Stokes equations,*

$$\begin{aligned} \min_{y, u} \quad & \frac{1}{2} \|y - y_d\|_{[L^2(\Omega)]^N}^2 + \frac{\gamma}{2} \|u\|_{[L^2(\Omega)]^N}^2 \\ \text{s.t.} \quad & \begin{cases} -\nu \Delta y + (y \cdot \nabla) y + \nabla p = u & \text{on } \Omega \\ \operatorname{div} y = 0 & \text{on } \Omega \\ y = 0 & \text{on } \partial\Omega \end{cases} \end{aligned}$$

on some bounded Lipschitz domain $\Omega \subset \mathbb{R}^N$, $N \in \{2, 3\}$. Suitable function spaces for the problem are

$$Y = Z = \text{closure in } [H^1(\Omega)]^N \text{ of } \{v \in [C_0^\infty(\Omega)]^N : \operatorname{div} v = 0\}, \quad U = [L^2(\Omega)]^N.$$

In [17, Theorem 3.16] it was proved that the condition

$$\|y\|_{[L^2(\Omega)]^N}^2 + \gamma \|u\|_{[L^2(\Omega)]^N}^2 + 2 \int_{\Omega} (y \cdot \nabla) y \lambda_0 \geq \rho \|u\|_{[L^{4/3}(\Omega)]^N}^2$$

for some $\rho > 0$ and all (y, u) satisfying the linearized state equation at (y_0, u_0) is a second-order sufficient condition of optimality for a critical point (y_0, u_0, λ_0) . Hence this weaker condition may replace Assumption 2.3(4) for this problem. Still, it can be proved along the lines of [4, 10] that \mathcal{K} is boundedly invertible as an element of $\mathcal{L}(Y \times [L^{4/3}(\Omega)]^N \times Z, Y^* \times [L^4(\Omega)]^N \times Z^*)$. The subsequent ideas remain valid when U is replaced by $L^{4/3}(\Omega)$.

From the bounded invertibility of \mathcal{K} , we can easily derive the differentiability of the parameter-to-solution map from the implicit function theorem [2]:

Lemma 2.6. *There exist neighborhoods B_1 of p_0 and B_2 of (y_0, u_0, λ_0) and a continuously differentiable function $\Psi : B_1 \rightarrow B_2$ such that for all $p \in B_1$, $\Psi(p)$ is the unique solution in B_2 of (2.3)–(2.5). The Fréchet derivative of Ψ at p_0 is given by*

$$\Psi'(p_0) = -\mathcal{K}^{-1} \begin{pmatrix} \mathcal{L}_{yp} \\ \mathcal{L}_{up} \\ e_p \end{pmatrix} \quad (2.7)$$

where the right hand side is evaluated at the nominal solution (y_0, u_0, λ_0) and p_0 .

In particular, we infer from Lemma 2.6 that for a given perturbation direction \bar{p} , the directional derivatives of the nominal optimal state and optimal control and the corresponding adjoint state $(\bar{y}, \bar{u}, \bar{\lambda})$ are given by the unique solution of the linear system in $Y^* \times U^* \times Z^*$

$$\mathcal{K} \begin{pmatrix} \bar{y} \\ \bar{u} \\ \bar{\lambda} \end{pmatrix} = \mathcal{B} \bar{p} \quad \text{where} \quad \mathcal{B} = - \begin{pmatrix} \mathcal{L}_{yp} \\ \mathcal{L}_{up} \\ e_p \end{pmatrix}. \quad (2.8)$$

These directional derivatives are called the parametric sensitivities of the state, control and adjoint variables. They describe the first-order change in these variables as p changes from p_0 to $p_0 + \bar{p}$.

It is worth noting that these sensitivities can be characterized alternatively as the unique solution $x = (y, u)$ and adjoint state of the following auxiliary problem with quadratic objective and linear constraint:

$$\begin{aligned} \min_{y, u} \quad & \frac{1}{2} \mathcal{L}_{xx}(y_0, u_0, \lambda_0, p_0)(x, x) + \mathcal{L}_{xp}(y_0, u_0, \lambda_0, p_0)(x, \bar{p}) \\ \text{subject to} \quad & e_y(y_0, u_0, p_0) y + e_u(y_0, u_0, p_0) u = -e_p(y_0, u_0, p_0) \bar{p}. \end{aligned} \quad (2.9)$$

Hence, computing the parametric sensitivity in a given direction \bar{p} amounts to solving one linear-quadratic problem (2.9).

We recall that it is our goal to analyze the stability properties of an observed quantity

$$q : Y \times U \times Z \ni (y, u, \lambda) \mapsto q(y, u, \lambda) \in H$$

depending on the solution, where H is another finite- or infinite-dimensional Hilbert space and q is differentiable. By the chain rule, the first-order change in the observed quantity, as p changes from p_0 to $p_0 + \bar{p}$, is given by

$$\Pi(\bar{y}, \bar{u}, \bar{\lambda}) := q'(y_0, u_0, \lambda_0)(\bar{y}, \bar{u}, \bar{\lambda}). \quad (2.10)$$

We refer to $\Pi = q'(y_0, u_0, \lambda_0) \in \mathcal{L}(Y \times U \times Z, H)$ as the observation operator. Due to (2.8), we have the following linear relation between perturbation direction \bar{p} and first order change in the observed quantity:

$$\Pi(\bar{y}, \bar{u}, \bar{\lambda}) = \Pi \mathcal{K}^{-1} \mathcal{B} \bar{p}.$$

Example 2.7 (Observation Operators).

- (i) *If one is interested in the impact of perturbations on the optimal state on some subset Ω' of the computational domain Ω , one has $q(y, u, \lambda) = y|_{\Omega'}$ and, due to linearity, $\Pi = q$ holds.*
- (ii) *If the quantity of interest is the impact of perturbations on the average value of the control variable, one chooses $q(y, u, \lambda) = \int u$ where the integral extends over the control domain.*

It is the bounded linear map $\Pi\mathcal{K}^{-1}\mathcal{B}$ that we now focus our attention on. The maximum impact of all perturbations (of unit size) on the observed quantity is given by the operator norm

$$\|\Pi\mathcal{K}^{-1}\mathcal{B}\|_{\mathcal{L}(P,H)} = \sup_{\bar{p} \neq 0} \frac{\|\Pi\mathcal{K}^{-1}\mathcal{B}\bar{p}\|_H}{\|\bar{p}\|_P}. \quad (2.11)$$

To simplify the notation, we will also use the abbreviation

$$\mathcal{A} := \Pi\mathcal{K}^{-1}\mathcal{B}.$$

In general, the operator norm need not be attained for any direction \bar{p} . Therefore, and in order to perform the singular value decomposition, we make the following assumption:

Assumption 2.8. *Suppose that \mathcal{A} is compact from P to H .*

To demonstrate that this assumption is not overly restrictive, we discuss several important examples. Recall that in PDE-constrained optimization, Y and Z are infinite-dimensional function spaces. Hence, \mathcal{K}^{-1} cannot be compact since then its spectrum would contain 0 which entails non-invertibility of \mathcal{K}^{-1} . (Of course, if all of Y , U and Z are finite-dimensional, Assumption 2.8 holds trivially.)

Example 2.9 (Compactness of \mathcal{A}).

- (i) *If at least one of the parameter or observation spaces P or H is finite-dimensional, \mathcal{A} is trivially compact.*
- (ii) *For sufficiently regular perturbations, \mathcal{B} and thus \mathcal{A} is compact: Consider the standard distributed optimal control problem with $Y = Z = H_0^1(\Omega)$, $U = L^2(\Omega)$, where Ω is a bounded domain with Lipschitz boundary in \mathbb{R}^N , $N \geq 1$, $y_d, u_d \in L^2(\Omega)$, and*

$$f(y, u) = \frac{1}{2}\|y - y_d\|_{L^2(\Omega)}^2 + \frac{\gamma}{2}\|u - u_d\|_{L^2(\Omega)}^2$$

$$e(y, u, p)(\varphi) = (\nabla y, \nabla \varphi) - (u, \varphi) - \langle p, \varphi \rangle_{H^{-1}(\Omega), H_0^1(\Omega)}, \quad \varphi \in H_0^1(\Omega),$$

which corresponds to $-\Delta y = u + p$ on Ω and $y = 0$ on $\partial\Omega$. It is straightforward to verify that $\mathcal{B} = (0, 0, \text{id})^\top$. By compact embedding, see [1], \mathcal{B} is compact from $P = L^{(N+2)/(2N)+\varepsilon}(\Omega)$ into $Y^ \times U^* \times Z^*$ for any $\varepsilon > 0$, and in particular for the Hilbert space $P = L^2(\Omega)$ in any dimension N . Hence $\mathcal{A} = \Pi\mathcal{K}^{-1}\mathcal{B}$ is compact for $P = L^2(\Omega)$ and arbitrary linear and bounded observation operators Π .*

- (iii) *In the previous example, neither \mathcal{B} nor $\mathcal{K}^{-1}\mathcal{B}$ is compact if $P = H^{-1}(\Omega)$. In that case, one has to choose an observation space of sufficiently low regularity, so that Π and hence \mathcal{A} is compact. For instance, in the previous example, $\Pi(\bar{y}, \bar{u}, \bar{\lambda}) = \bar{y}$ is compact into $H = L^2(\Omega)$ due to the compact embedding of $H_0^1(\Omega)$ into $L^2(\Omega)$.*

We refer to Section 4 for more examples and return to the issue of computing the operator norm (2.11). This can be achieved by the singular value decomposition [3, Ch. 2.2]:

Lemma 2.10. *There exists a countable system $\{(\sigma_n, v_n, u_n)\}_{n \in \mathbb{N}}$ such that $\{\sigma_n\}_{n \in \mathbb{N}}$ is non-increasing and non-negative, $\{(\sigma_n^2, v_n)\} \subset \mathbb{R} \times P$ is a complete orthonormal system of eigenpairs for $\mathcal{A}^H \mathcal{A}$ (spanning the closure of the range of \mathcal{A}^H), and $\{(\sigma_n^2, u_n)\} \subset \mathbb{R} \times H$ is a complete orthonormal system of eigenpairs for $\mathcal{A} \mathcal{A}^H$*

(spanning the closure of the range of \mathcal{A}). In addition, $\mathcal{A}v_n = \sigma_n u_n$ holds and we have

$$\mathcal{A}\bar{p} = \Pi\mathcal{K}^{-1}\mathcal{B}\bar{p} = \sum_{n=1}^{\infty} \sigma_n (\bar{p}, v_n)_P u_n \quad (2.12)$$

for all $\bar{p} \in P$, where the series converges in H . Every value in $\{\sigma_n\}_{n \in \mathbb{N}}$ appears with finite multiplicity.

In Lemma 2.10, $\mathcal{A}^H : H \rightarrow P$ denotes the Hilbert space adjoint of \mathcal{A} and $(\cdot, \cdot)_P$ is the scalar product of P . A system according to Lemma 2.10 is called a *singular system* for \mathcal{A} , with singular values σ_n , left singular vectors $u_n \in H$, and right singular vectors $v_n \in P$. Knowledge of the singular system will not only allow us to compute the operator norm (2.11) and the direction(s) \bar{p} for which this bound is attained, but in addition, we obtain a complete sequence of perturbation directions in decreasing order of importance with regard to the perturbations in the observed quantity. This is formulated in the following proposition:

Proposition 2.11. *Let $\{(\sigma_n, v_n, u_n)\}_{n \in \mathbb{N}}$ be a singular system for \mathcal{A} . Then the operator norm in (2.11) is given by σ_1 . Moreover, the supremum is attained exactly for all non-zero vectors $\bar{p} \in \text{span}\{v_1, \dots, v_k\} =: V_1$, where k is the largest integer such that $\sigma_1 = \sigma_k$. Similarly, when \mathcal{A} is restricted to V_1^\perp , its operator norm is given by σ_{k+1} and it is attained exactly for all non-zero vectors $\bar{p} \in \text{span}\{v_{k+1}, \dots, v_l\}$, where l is the largest integer such that $\sigma_{k+1} = \sigma_l$, and so on.*

Proof. The claim follows directly from the properties of the singular system. \square

Proposition 2.11 shows that the question of greatest impact of arbitrary perturbations on the observed quantity is answered by the singular value decomposition (SVD) of \mathcal{A} . It is well known that SVD is closely related to principal components analysis (PCA) in statistics and image processing [8], and proper orthogonal decomposition (POD) in dynamical systems, compare [13, 18]. To our knowledge, however, this technique has not been exploited for the quantitative stability analysis of optimization problems.

In the following section we focus on an efficient algorithm for the numerical computation of the largest singular values and left and right singular vectors for a discretized version of problem (2.1).

3. NUMERICAL STABILITY ANALYSIS

In this section, we propose an efficient algorithm for the numerical computation of the singular system for a discretized (matrix) version of $\Pi\mathcal{K}^{-1}\mathcal{B}$. The convergence of the singular system of the discretized problem to the singular system of the continuous problem will be discussed elsewhere. In practice, it will be sufficient to compute only a partial SVD, starting with the largest singular value, down to a certain threshold, in order to collect the perturbation directions of greatest impact with respect to the observed quantity. The method we propose makes use of existing standard software which iteratively approximates the extreme eigenpairs of non-symmetric matrices, and it will be efficient in the following sense: It is unnecessary to assemble the (discretized) matrix $\Pi\mathcal{K}^{-1}\mathcal{B}$, which is prohibitive for high-dimensional parameter and observation spaces. Only matrix–vector products with $\mathcal{K}^{-1}\mathcal{B}$ are required, i.e., the solution of sensitivity problems (2.8), and the inexpensive application of the observation operator Π . In particular, we avoid the computation of certain Cholesky factors which relate the Euclidean norms of

coordinate vectors and the function space norms of the functions represented by them, see below.

We discretize problem (2.1) by a Galerkin procedure, e.g., the finite element or wavelet method. To this end, we introduce finite-dimensional subspaces $Y_h \subset Y$, $U_h \subset U$ and $Z_h \subset Z$, which inherit the norms from the larger spaces. The discretized problem reads

$$\min_{y,u} f(y, u, p) \quad \text{subject to } e(y, u, p)(\varphi) = 0 \text{ for all } \varphi \in Z_h, \quad (3.1)$$

where $(y, u) \in Y_h \times U_h$. In the general case of an infinite-dimensional parameter space, we also choose a finite-dimensional subspace $P_h \subset P$. Should any of the spaces be finite-dimensional in the first place, we leave it unchanged by discretization.

Suppose that for the given parameter $p_0 \in P_h$, a critical point for the discretized problem has been computed by a suitable method, for instance, by sequential quadratic programming (SQP) methods [12, 15]. That is, $(y_h, u_h, \lambda_h) \in Y_h \times U_h \times Z_h$ satisfies the discretized optimality system, compare (2.3)–(2.5):

$$f_y(y_h, u_h, p_0)(\delta y_h) + \langle \lambda_h, e_y(y_h, u_h, p_0)(\delta y_h) \rangle = 0 \quad \text{for all } \delta y_h \in Y_h \quad (3.2)$$

$$f_u(y_h, u_h, p_0)(\delta u_h) + \langle \lambda_h, e_u(y_h, u_h, p_0)(\delta u_h) \rangle = 0 \quad \text{for all } \delta u_h \in U_h \quad (3.3)$$

$$e(y_h, u_h, p_0)(\delta z_h) = 0 \quad \text{for all } \delta z_h \in Z_h. \quad (3.4)$$

We consider the discrete analog of the sensitivity system (2.8), i.e.,

$$\left\langle \mathcal{K}^h \begin{pmatrix} \bar{y}_h \\ \bar{u}_h \\ \bar{\lambda}_h \end{pmatrix}, \begin{pmatrix} \delta y_h \\ \delta u_h \\ \delta z_h \end{pmatrix} \right\rangle = \left\langle \mathcal{B}^h \bar{p}_h, \begin{pmatrix} \delta y_h \\ \delta u_h \\ \delta z_h \end{pmatrix} \right\rangle \quad \text{for all } (\delta y_h, \delta u_h, \delta z_h) \in Y_h \times U_h \times Z_h, \quad (3.5)$$

where \mathcal{K}^h and \mathcal{B}^h are defined as before in (2.6) and (2.8), evaluated at the critical point (y_h, u_h, λ_h) . The perturbation direction \bar{p}_h is taken from the discretized parameter space P_h .

Assumption 3.1. *Suppose that the critical point (y_h, u_h, λ_h) is sufficiently close to the local solution of the continuous problem (y_0, u_0, λ_0) , such that second-order sufficient conditions hold for the discretized problem. That is, $e_y(y_h, u_h, p_0)$ maps Y_h onto Z_h , and there exists $\rho' > 0$ such that $\mathcal{L}_{xx}(y_h, u_h, \lambda_h, p_0)(\bar{x}, \bar{x}) \geq \rho' \|\bar{x}\|_{Y \times U}^2$ for all $\bar{x} \in Y_h \times U_h$ satisfying $\langle e_x(y_h, u_h, p_0)\bar{x}, \varphi \rangle = 0$ for all $\varphi \in Z_h$.*

Under Assumption 3.1, the KKT operator \mathcal{K}^h at the discrete solution is invertible and equation (3.5) gives rise to a linear map

$$(\mathcal{K}^h)^{-1} \mathcal{B}^h : P_h \rightarrow Y_h \times U_h \times Z_h$$

which acts between finite-dimensional spaces and thus is automatically bounded. There is no need to discretize the observation space H since $\Pi \mathcal{K}^{-1} \mathcal{B}$, restricted to P_h , has finite-dimensional range. Nevertheless, we define for convenience the subspace of H ,

$$R_h = \text{range of } \Pi^h (\mathcal{K}^h)^{-1} \mathcal{B}^h \quad \text{considered as a map } P_h \rightarrow H,$$

where $\Pi^h = q'(y_h, u_h, \lambda_h)$, compare (2.10).

We recall that it is our goal to calculate the portion of the singular system for $\Pi^h (\mathcal{K}^h)^{-1} \mathcal{B}^h : P_h \rightarrow R_h$ which belongs to the largest singular values. At this point, we introduce a basis for the discretized parameter space P_h , say

$$P_h = \text{span} \{ \varphi_1, \dots, \varphi_m \}.$$

Likewise, we define a space H_h by

$$H_h := \text{span} \{ \psi_1, \dots, \psi_n \} \text{ such that } H_h \supset R_h.$$

Both the systems $\{\varphi_i\}$ and $\{\psi_j\}$ are assumed linearly independent without loss of generality. As the range space R_h is usually not known exactly, we allow the functions ψ_j to span a larger space H_h . For instance, in case of the state observation operator $\Pi^h(\bar{y}_h, \bar{u}_h, \bar{\lambda}_h) = \bar{y}_h$, we may choose $\{\psi_j\}_{j=1}^n$ to be identical to the finite element basis of the state space Y_h , which certainly contains the range space R_h .

For the application of numerical procedures, we need to switch to a coordinate representation of the elements of the discretized parameter and observation spaces P_h and H_h . Note that a function $\bar{p} \in P_h$ can be identified with its coordinate vector $\bar{\mathbf{p}} = (\bar{\mathbf{p}}_1, \dots, \bar{\mathbf{p}}_m)^\top$ with respect to the given basis. In other words, \mathbb{R}^m and P_h are isomorphic, and the isomorphism and its inverse are given by the expansion and coordinate maps

$$\begin{aligned} E_P : \mathbb{R}^m \ni \bar{\mathbf{p}} &\mapsto \sum_{i=1}^m \bar{\mathbf{p}}_i \varphi_i \in P_h \\ C_P = E_P^{-1} : P_h &\rightarrow \mathbb{R}^m. \end{aligned}$$

We also introduce the mass matrix associated to the chosen basis of P_h ,

$$M_P = (m_{ij})_{i,j=1}^m, \quad m_{ij} = (\varphi_i, \varphi_j)_P.$$

In case of a discretization by orthogonal wavelets, M_P is the identity matrix, while in the finite element case, M_P is a sparse symmetric positive definite matrix. In any case, we have the following relation between the Euclidean norm of the coordinate vector $\bar{\mathbf{p}}$ and the norm of the element $\bar{p} \in P_h$ represented by it:

$$\|\bar{p}\|_P^2 = \bar{\mathbf{p}}^\top M_P \bar{\mathbf{p}} = \|M_P^{1/2} \bar{\mathbf{p}}\|_2^2,$$

where $M_P^{1/2}$ is the Cholesky factor of $M_P = M_P^{1/2\top} M_P^{1/2}$, and $\|\cdot\|_2$ denotes the Euclidean norm of vectors in \mathbb{R}^m or \mathbb{R}^n . Similarly as above, we define expansion and coordinate maps $E_H : \mathbb{R}^n \rightarrow H_h$ and $C_H = E_H^{-1}$ and the mass matrix

$$M_H = (m_{ij})_{i,j=1}^n, \quad m_{ij} = (\psi_i, \psi_j)_H$$

to obtain

$$\|\bar{h}\|_H^2 = \bar{\mathbf{h}}^\top M_H \bar{\mathbf{h}} = \|M_H^{1/2} \bar{\mathbf{h}}\|_2^2$$

for an element $\bar{h} = \sum_{j=1}^n \bar{\mathbf{h}}_j \psi_j \in H_h$ with coordinate vector $\bar{\mathbf{h}} = (\bar{\mathbf{h}}_1, \dots, \bar{\mathbf{h}}_n)^\top$.

Any numerical procedure which solves the sensitivity problem (3.5) and applies the observation operator Π^h does not directly implement the operator $\Pi^h(\mathcal{K}^h)^{-1} \mathcal{B}^h$. Rather, it realizes its representation in the coordinate systems given by the bases of P_h and H_h , i.e.,

$$\mathcal{A}^h := C_H \Pi^h(\mathcal{K}^h)^{-1} \mathcal{B}^h E_P \in \mathbb{R}^{n \times m}.$$

As mentioned earlier, the proposed method will employ matrix-vector products with \mathcal{A}^h . Every matrix-vector product requires the solution of a discretized sensitivity equation (3.5) followed by the application of the observation operator.

Note that there is a discrepancy in the operator \mathcal{A}^h being given in terms of coordinate vectors and the requirement that the SVD should respect the norms of the spaces P_h and H_h . One way to overcome this discrepancy is to exchange the Euclidean scalar products in the SVD routine at hand by scalar products with respect to the mass matrices M_P and M_h , respectively. In the sequel, we describe an alternative approach based on iterative eigen decomposition software, without the need of modifying any scalar products.

By the relations between coordinate vectors and functions, we have

$$\begin{aligned}
\|\Pi^h(\mathcal{K}^h)^{-1}\mathcal{B}^h\|_{\mathcal{L}(P_h, H_h)} &= \sup_{\bar{p}_h \in P_h \setminus \{0\}} \frac{\|\Pi^h(\mathcal{K}^h)^{-1}\mathcal{B}^h \bar{p}_h\|_H}{\|\bar{p}_h\|_P} \\
&= \sup_{\bar{\mathbf{p}} \in \mathbb{R}^m \setminus \{0\}} \frac{\|\Pi^h(\mathcal{K}^h)^{-1}\mathcal{B}^h E_P \bar{\mathbf{p}}\|_H}{\|E_P \bar{\mathbf{p}}\|_P} = \sup_{\bar{\mathbf{p}} \in \mathbb{R}^m \setminus \{0\}} \frac{\|E_H \mathcal{A}^h \bar{\mathbf{p}}\|_H}{\|M_P^{1/2} \bar{\mathbf{p}}\|_2} \\
&= \sup_{\bar{\mathbf{p}} \in \mathbb{R}^m \setminus \{0\}} \frac{\|M_H^{1/2} \mathcal{A}^h \bar{\mathbf{p}}\|_2}{\|M_P^{1/2} \bar{\mathbf{p}}\|_2} = \sup_{\bar{\mathbf{p}}' \in \mathbb{R}^m \setminus \{0\}} \frac{\|M_H^{1/2} \mathcal{A}^h M_P^{-1/2} \bar{\mathbf{p}}'\|_2}{\|\bar{\mathbf{p}}'\|_2}. \quad (3.6)
\end{aligned}$$

The last manipulation is a coordinate transformation in P_h , and $M_P^{-1/2}$ denotes the inverse of the Cholesky factor of M_P . This transformation shows that a finite-dimensional SVD procedure which employs the standard Euclidean vector norms in the image and pre-image spaces should target the matrix $M_H^{1/2} \mathcal{A}^h M_P^{-1/2}$.

Coordinate vectors referring to the new coordinate systems will be indicated by a prime. We have the relationships

$$\bar{\mathbf{p}}' = M_P^{1/2} \bar{\mathbf{p}} \quad \text{and} \quad \|\bar{\mathbf{p}}'\|_2 = \|M_P^{1/2} \bar{\mathbf{p}}\|_2 = \|\bar{\mathbf{p}}\|_P.$$

Hence the Euclidean norm of the transformed coordinate vector equals the norm of the function represented by it. The corresponding basis can in principle be obtained by an orthonormalization procedure with respect to the scalar product in P , starting from the previously chosen basis $\{\varphi_i\}$. Assembling the mass matrices and forming the Cholesky factors $M_H^{1/2}$ and $M_P^{1/2}$, however, will be too costly in general. Therefore, we propose the following strategy which avoids the Cholesky factors altogether. It is based on the following Jordan-Wielandt Lemma, see, e.g., [14, Theorem I.4.2]:

Lemma 3.2. *The singular value decomposition of $M_H^{1/2} \mathcal{A}^h M_P^{-1/2}$ is equivalent to the eigen decomposition of the symmetric Jordan-Wielandt matrix*

$$J = \begin{pmatrix} 0 & M_H^{1/2} \mathcal{A}^h M_P^{-1/2} \\ M_P^{-1/2\top} \mathcal{A}^{h\top} M_H^{1/2\top} & 0 \end{pmatrix} \in \mathbb{R}^{(m+n) \times (m+n)}$$

in the following sense: The eigenvalues of J are exactly $\pm\sigma_i$, where $\{\sigma_i\}_{i=1}^{\min\{m,n\}}$ are the singular values of $M_H^{1/2} \mathcal{A}^h M_P^{-1/2}$, plus a suitable number of zeros. The eigenvectors \mathbf{v}_i' belonging to the nonnegative eigenvalues σ_i , $i = 1, \dots, \min\{m,n\}$, can be partitioned into $\mathbf{v}_i' = (\mathbf{l}_i', \mathbf{r}_i')^\top$, where $\mathbf{r}_i' \in \mathbb{R}^m$ and $\mathbf{l}_i' \in \mathbb{R}^n$. After normalization, \mathbf{r}_i' and \mathbf{l}_i' are the right and left singular vectors of $M_H^{1/2} \mathcal{A}^h M_P^{-1/2}$.

Exchanging the singular value decomposition of $M_H^{1/2} \mathcal{A}^h M_P^{-1/2}$ for an eigen decomposition of the Jordan-Wielandt matrix J does not resolve the issue of forming the Cholesky factors $M_H^{1/2}$ and $M_P^{1/2}$. To this end, we apply a similarity transform to J using the similarity matrices

$$X = \begin{pmatrix} M_H^{-1/2} & 0 \\ 0 & M_P^{-1/2} \end{pmatrix}, \quad X^{-1} = \begin{pmatrix} M_H^{1/2} & 0 \\ 0 & M_P^{1/2} \end{pmatrix}.$$

Then the transformed matrix

$$X J X^{-1} = \begin{pmatrix} 0 & \mathcal{A}^h \\ M_P^{-1} \mathcal{A}^{h\top} M_H & 0 \end{pmatrix} \quad (3.7)$$

has the same eigenvalues as J , including the desired singular values of $M_H^{1/2} \mathcal{A}^h M_P^{-1/2}$.

Lemma 3.3. *The transformed matrix has the form*

$$XJX^{-1} = \begin{pmatrix} 0 & C_H \Pi^h (\mathcal{K}^h)^{-1} \mathcal{B}^h E_P \\ C_P (\mathcal{B}^h)^* (\mathcal{K}^h)^{-1} (\Pi^h)^* E_H & 0 \end{pmatrix}, \quad (3.8)$$

where $(\mathcal{B}^h)^* : Y_h \times U_h \times Z_h \rightarrow P_h$ and $(\Pi^h)^* : H_h \rightarrow Y_h \times U_h \times Z_h$ are the adjoint operators of \mathcal{B}^h and Π^h , respectively.

Proof. We only need to consider the lower left block. By transposing \mathcal{A}^h , we obtain

$$\mathcal{A}^{h\top} = E_P^* (\mathcal{B}^h)^* (\mathcal{K}^h)^{-1} (\Pi^h)^* C_H^*$$

since \mathcal{K}^h is symmetric. By definition, the adjoint operator E_P^* satisfies $\langle E_P^* \xi, \mathbf{p} \rangle_{\mathbb{R}^m} = \langle \xi, E_P \mathbf{p} \rangle_P$ for all $\xi \in P_h$ and $\mathbf{p} \in \mathbb{R}^m$. Hence, we obtain

$$\mathbf{p}^\top (E_P^* \xi) = \langle \xi, \sum_{i=1}^m \mathbf{p}_i \varphi_i \rangle_P = \mathbf{p}^\top M_P (C_P \xi)$$

and thus $E_P^* = M_P C_P$. Moreover,

$$C_H^* = (E_H^{-1})^* = (E_H^*)^{-1} = (M_H C_H)^{-1} = C_H^{-1} M_H^{-1} = E_H M_H^{-1}$$

holds. Consequently,

$$M_P^{-1} \mathcal{A}^{h\top} M_H = C_P (\mathcal{B}^h)^* (\mathcal{K}^h)^{-1} (\Pi^h)^* E_H$$

as claimed. \square

Remark 3.4. *Algorithmically, evaluating a matrix-vector product with (3.8) and a given coordinate vector $(\bar{\mathbf{h}}, \bar{\mathbf{p}})^\top \in \mathbb{R}^n \times \mathbb{R}^m$ amounts to solving two sensitivity problems:*

- (1) *The first problem is (3.5) with the perturbation direction $\bar{\mathbf{p}} = E_P \bar{\mathbf{p}} \in P_h$.*
- (2) *For the second problem, the right hand side operator \mathcal{B}^h in (3.5) is replaced by $(\Pi^h)^*$, and the observation operator Π^h is replaced by $(\mathcal{B}^h)^*$. The direction of evaluation is $\bar{\mathbf{h}} = E_H \bar{\mathbf{h}} \in H_h$.*

Step (2) requires a modification of the original sensitivity problem (3.5). As an alternative, one may apply the following duality argument to (3.7): The vector $M_P^{-1} \mathcal{A}^{h\top} M_H \bar{\mathbf{h}}$ is equal to the transpose of $\bar{\mathbf{h}}^\top M_H \mathcal{A}^h M_P^{-1}$. In case that the dimension of the parameter space m is small, the inversion of M_P and the solution of m sensitivity problems to get $\mathcal{A}^h M_P^{-1}$ may be feasible.

Let us denote by $\mathbf{w}_i = (\mathbf{w}_i^{(1)}, \mathbf{w}_i^{(2)})^\top$ the eigenvectors of XJX^{-1} belonging to the nonnegative eigenvalues σ_i , $i = 1, \dots, \min\{m, n\}$. This similarity transformation with X and X^{-1} does indeed avoid the Cholesky factors of the mass matrices, as will become clear in the sequel.

Recall that the eigenvalues of XJX^{-1} are $\pm\sigma_i$, plus a suitable number of zeros, where σ_i are the desired singular values. Hence the largest singular values correspond to the eigenvalues of largest magnitude, which can be conveniently computed iteratively, e.g., by an implicitly restarted Arnoldi process [19, Ch. 6.4]. Available software routines include the library ARPACK (DNAUPD and DNEUPD), see [9], and MATLAB's `eigs` function. In case that the parameter space (or the observation space) is low-dimensional, we may also compute the matrix XJX^{-1} explicitly, see Sections 4.1 and 4.2, but these cases are not considered typical for our applications.

We now discuss how to recover the desired partial singular value decomposition from the partial eigen decomposition of XJX^{-1} . For later reference, we note the following property of the eigenvectors of (3.7), which is readily verified:

$$\mathbf{w}_i^{(1)\top} M_H \mathbf{w}_i^{(1)} = \mathbf{w}_i^{(2)\top} M_P \mathbf{w}_i^{(2)}. \quad (3.9)$$

Note also that the eigenvectors \mathbf{w}_i of XJX^{-1} and \mathbf{v}'_i of J are related by $\mathbf{w}_i = X\mathbf{v}'_i$. As the left and right singular vectors of $M_H^{1/2}\mathcal{A}^hM_P^{-1/2}$ are just a partitioning of \mathbf{v}'_i according to Lemma 3.2, we get

$$\begin{pmatrix} \mathbf{l}'_i \\ \mathbf{r}'_i \end{pmatrix} = \mathbf{v}'_i = X^{-1} \begin{pmatrix} \mathbf{w}_i^{(1)} \\ \mathbf{w}_i^{(2)} \end{pmatrix},$$

which in turn seems to bring up the Cholesky factors we wish to avoid. However, \mathbf{r}'_i is a coordinate vector with respect to an artificial (orthonormal) basis of P_h , which does not in general coincide with our chosen basis $\{\varphi_i\}$. Going back to this natural basis and normalizing, we arrive at

$$\mathbf{r}_i = \frac{\mathbf{w}_i^{(2)}}{(\mathbf{w}_i^{(2)\top} M_P \mathbf{w}_i^{(2)})^{1/2}} \quad (3.10)$$

Now \mathbf{r}_i is the coordinate representation of the desired i -th right singular vector with respect to the basis $\{\varphi_i\}$. Due to the normalization, the function represented by \mathbf{r}_i has P -norm one.

We also wish to find the coordinate representation \mathbf{l}_i of the response of the system \mathcal{A}^h , given the perturbation input \mathbf{r}_i . As \mathbf{r}_i is a multiple of $\mathbf{w}_i^{(2)}$ and thus part of an eigenvector of XJX^{-1} , we infer from (3.7) that \mathcal{A}^h maps \mathbf{r}_i to a multiple of $\mathbf{w}_i^{(1)}$. We are thus led to define

$$\mathbf{l}_i = \frac{\mathbf{w}_i^{(1)}}{(\mathbf{w}_i^{(1)\top} M_H \mathbf{w}_i^{(1)})^{1/2}}. \quad (3.11)$$

Despite the individual normalizations of $\mathbf{w}_i^{(1)}$ and $\mathbf{w}_i^{(2)}$, \mathbf{l}_i and \mathbf{r}_i are still related by the same proportionality constant:

$$\mathcal{A}^h \mathbf{r}_i = \sigma_i \mathbf{l}_i, \quad (3.12)$$

as can be easily verified using (3.9). We have thus proved our main result:

Theorem 3.5. *Suppose that $\sigma_i > 0$ is an eigenvalue of the matrix XJX^{-1} with eigenvector $\mathbf{w}_i = (\mathbf{w}_i^{(1)}, \mathbf{w}_i^{(2)})^\top$. Let \mathbf{r}_i and \mathbf{l}_i be given by (3.10) and (3.11), respectively and let $r_i = E_P \mathbf{r}_i \in P_h$ and $l_i = E_H \mathbf{l}_i \in H_h$ be the functions represented by them. Then the following relations are satisfied:*

- (a) $\|r_i\|_P = \|l_i\|_H = 1$.
- (b) *The perturbation r_i invokes the first order change $\sigma_i l_i$ of magnitude σ_i in the observed quantity. In terms of coordinate vectors, $\mathcal{A}^h \mathbf{r}_i = \sigma_i \mathbf{l}_i$.*

Based on these considerations, we propose to compute the desired singular value decomposition of $M_H^{1/2}\mathcal{A}^hM_P^{-1/2}$ by iteratively approximation the extreme eigenvalues and corresponding eigenvectors of XJX^{-1} . This avoids the Cholesky factors of the mass matrices, as desired. We summarize the proposed procedure in Algorithm 1.

Remark 3.6. *The singular value decomposition of \mathcal{A} and \mathcal{A}^h relies on the linearity of the map $\bar{p} \mapsto (\bar{y}, \bar{u}, \bar{\lambda})$, which maps a perturbation direction \bar{p} to the directional derivative of the optimal solution and adjoint state, compare (2.7)–(2.8). For optimal control problems with pointwise control constraints $a(x) \leq u(x) \leq b(x)$ almost everywhere on the control domain, the derivative need not be linear with respect to the direction, see [4, 10]. The presence of strict complementarity, however, restores the linearity. The procedure outlined above carries over to this case, with only minor modifications of the operators \mathcal{K}^h and \mathcal{B}^h on the so-called active sets, compare also [6].*

Algorithm 1

-
- Given: discretized spaces Y_h, U_h, Z_h and P_h, H_h ,
a discrete critical point (y_h, u_h, λ_h) satisfying (3.2)–(3.4) for $p_0 \in P_h$,
a routine evaluating $XJX^{-1}(\bar{\mathbf{h}}, \bar{\mathbf{p}})^\top$ for any given coordinate vector $(\bar{\mathbf{h}}, \bar{\mathbf{p}})^\top$, see Remark 3.4
- Desired: a user-defined number s of singular values and perturbation directions (right singular vectors) in coordinate representation, which are of greatest first order impact with respect to the observed quantity
-
- 1: Call a routine which iteratively computes the $2s$ eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_s \geq 0 \geq \lambda_{s+1} \geq \dots \geq \lambda_{2s}$ of largest absolute value and corresponding eigenvectors \mathbf{w}_i of XJX^{-1} .
 - 2: Set $\sigma_i := \lambda_i$ for $i = 1, \dots, s$.
 - 3: Split \mathbf{w}_i into $(\mathbf{w}_i^{(1)}, \mathbf{w}_i^{(2)})$ of lengths n and m , respectively, for $i = 1, \dots, s$.
 - 4: Compute vectors \mathbf{r}_i and \mathbf{l}_i for $i = 1, \dots, s$ according to (3.10) and (3.11).
-

4. NUMERICAL EXAMPLES

We consider as an example the optimal control problem

$$\begin{aligned}
& \text{minimize} && -\frac{1}{4} \int_{\Omega} y(x) dx + \frac{\gamma}{2} \|u\|_{L^2(C)}^2 \\
& \text{s.t.} && \begin{cases} -\kappa \Delta y = \chi_C u & \text{on } \Omega \\ \kappa \frac{\partial}{\partial n} y = \alpha (y - y_{\infty}) & \text{on } \partial\Omega. \end{cases}
\end{aligned} \tag{4.1}$$

It represents the optimal heating of a room $\Omega = (-1, 1)^2 \subset \mathbb{R}^2$ to maximal average temperature y , subject to quadratic control costs. Heating is achieved through two radiators on some part of the domain $C \subset \Omega$, and the heating power u serves as a distributed control variable. κ denotes the constant heat diffusivity, while α is the heat transfer coefficient with the environment. The latter has constant temperature y_{∞} . α is taken to be zero at the walls but greater than zero at the two windows, see Figure 4.1.

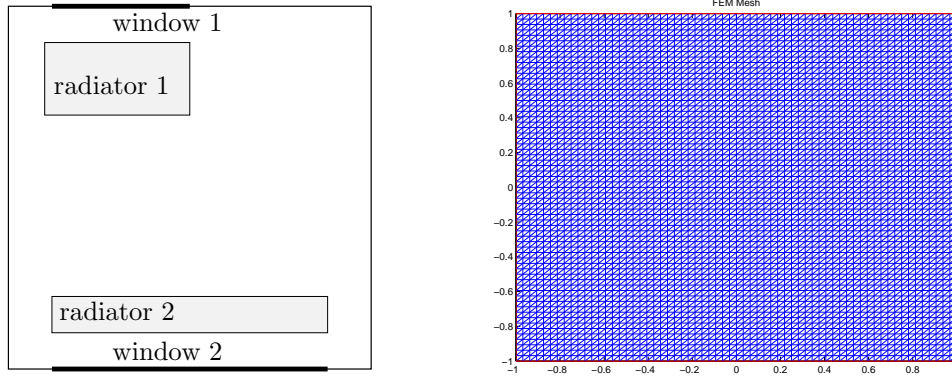


FIGURE 4.1. Layout of the domain and an intermediate finite element mesh with 4225 vertices (degrees of freedom).

In the sequel, we consider the window heat transfer coefficients as perturbation parameters. As its nominal value, we take

$$\alpha(x) = \begin{cases} 0 & \text{at the walls} \\ 1 & \text{at the lower (larger) window \# 2} \\ 2 & \text{at the upper (smaller) window \# 1.} \end{cases}$$

We will explore how the optimal temperature y changes under changes of α . Our example fits in the framework of Section 2 with

$$\begin{aligned} f(y, u) &= -\frac{1}{4} \int_{\Omega} y(x) dx + \frac{\gamma}{2} \|u\|_{L^2(C)}^2 \\ e(y, u, p)(\varphi) &= \kappa(\nabla y, \nabla \varphi)_{\Omega} - (u, \varphi)_C - (\alpha(y - y_{\infty}), \varphi)_{\partial\Omega}. \end{aligned}$$

Suitable function spaces for the problem are

$$Y = H^1(\Omega), \quad U = L^2(C), \quad Z = H^1(\Omega), \quad P = L^2(W_1) \times L^2(W_2).$$

f and e are infinitely differentiable w.r.t. (y, u, p) . For any given $(y, u, p) \in Y \times U \times P$, $e_y(y, u, p) : Y \rightarrow Z^*$ is onto and even boundedly invertible. Moreover, the problem is strictly convex and thus has a unique global solution which satisfies the second-order condition. The KKT operator is boundedly invertible. As state observation operator, we will use $\Pi(\bar{y}, \bar{u}, \bar{\lambda}) = \bar{y} \in H = L^2(\Omega)$. Compactness of \mathcal{A} then follows from compactness of the embedding $Y \hookrightarrow H$. Hence the example satisfies the Assumptions 2.3 and 2.8. Note that the parameter enters only in the PDE and not in the objective.

The problem is discretized using standard linear continuous finite elements for the state and adjoint, and discontinuous piecewise constant elements for the control. In order to estimate the order of convergence for the singular values, a hierarchy of uniformly refined triangular meshes is used. An intermediate mesh is shown in Figure 4.1 (right).

Since the problem has a quadratic objective and a linear PDE constraint, its solution requires the solution of only one linear system involving \mathcal{K} . Here and throughout, systems involving \mathcal{K} were solved using the conjugate gradient method applied to the reduced Hessian operator

$$\mathcal{K}_{\text{red}} = \begin{pmatrix} -e_y^{-1}e_u \\ id \end{pmatrix}^* \begin{pmatrix} \mathcal{L}_{yy} & \mathcal{L}_{yu} \\ \mathcal{L}_{uy} & \mathcal{L}_{uu} \end{pmatrix} \begin{pmatrix} -e_y^{-1}e_u \\ id \end{pmatrix},$$

see, e.g., [5, 7] for details. The state and adjoint partial differential equations are solved using a sparse direct solver.

Figure 4.2 shows the nominal solution (y_h, u_h) in the case

$$\begin{aligned} \kappa &= 1, \quad \gamma = 0.005, \quad y_{\infty} = 0 \\ C &= (-0.8, 0.0) \times (0.4, 0.8) \cup (-0.75, 0.75) \times (-0.8, -0.6) \\ W_1 &= (-0.75, 0) \times \{1\}, \quad W_2 = (-0.75, 0.75) \times \{-1\}. \end{aligned}$$

This setup describes the goal to heat up the room to a maximal average temperature (taking control costs into account) at an environmental temperature of 0°C . One clearly sees how heat is lost through the two windows.

In the sequel, we consider three variations of this problem. In every case, the insulation of the two windows, i.e., the heat transfer coefficient α restricted to the window areas, serves as a perturbation parameter. In Problem 1, this parameter is constant for each window and it is a spatial function in Problems 2 and 3. The optimal temperature y is the basis of the observation in all cases. In Problems 1 and 3, we observe the temperature at every point. In Problem 2, we consider

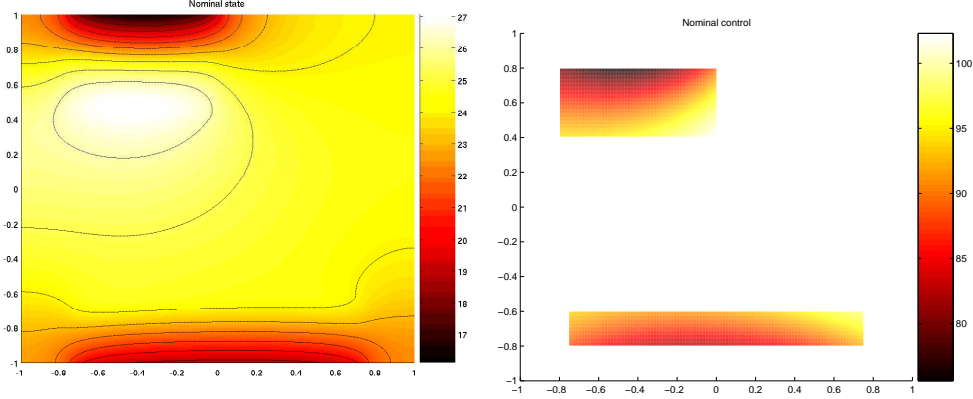


FIGURE 4.2. Nominal solution: Optimal state (left) and optimal control (right).

only the average temperature throughout the room. Hence, these problems cover all cases where at least one of the parameter or observation spaces P and H is infinite-dimensional and high-dimensional after discretization.

All examples are implemented using MATLAB's PDE toolbox. In every case, we use MATLAB's `eigs` function with standard tolerances to compute a partial eigen decomposition of the matrix XJX^{-1} . For Problems 1 and 2, we assemble this matrix explicitly according to (3.7). For Problem 3, we provide matrix-vector products with XJX^{-1} according to (3.8). Every matrix-vector product comes at the expense of the solution of two sensitivity problems (3.5), compare Remark 3.4.

4.1. Problem 1: Few Parameters, Large Observation Space. We begin by considering perturbations of the heat transfer coefficient on each window, i.e.,

$$p = (\alpha|_{W_1}, \alpha|_{W_2}) \in \mathbb{R}^2.$$

That is, we study the effect of replacing the windows by others with different insulation properties. While the parameter space is only two-dimensional, we consider an infinite-dimensional observation space and observe the effect of the perturbations on the overall temperature throughout the room. That is, we have the observation operator $\Pi(\bar{y}, \bar{u}, \bar{\lambda}) = \bar{y}$, and the space H is taken as $L^2(\Omega)$. Hence the mass matrix M_H in the discrete observation space is given by the $L^2(\Omega)$ -inner products of the linear continuous finite element basis on the respective grid. The mass matrix in the parameter space M_P is chosen as

$$M_P = \begin{pmatrix} 0.75 & 0 \\ 0 & 1.50 \end{pmatrix}$$

and it is generated by the L^2 -inner product of the constant functions of value one on W_1 and W_2 . It thus reflects the lengths of the two windows and allows a comparison with Problem 3 later on.

Since the matrix $\mathcal{A}^h \in \mathbb{R}^{n \times 2}$ has only two columns, it can be formed explicitly by solving only two sensitivity systems. From there, we easily set up XJX^{-1} according to (3.7) to avoid Cholesky factors of mass matrices, and perform an iterative partial eigen decomposition. Note that since \mathcal{A}^h has only two nonzero singular values, only four eigenvalues of XJX^{-1} are needed.

Table 4.1 shows the convergence of the singular values as the mesh is uniformly refined. In addition, the number of degrees of freedom of each finite element mesh

and the total number of variables in the optimization problem is shown. The last column lists the number of QP steps, i.e., solutions of (3.5) with matrix \mathcal{K}^h , which were necessary to obtain convergence of the (partial) eigen decomposition. For this problem, the number of QP solves is always two since $\mathcal{A}^h \in \mathbb{R}^{n \times 2}$ was assembled explicitly. Note also that our original problem (4.1) is linear-quadratic, hence finding the nominal solution requires only one solution with \mathcal{K}^h and computing the singular values and vectors is twice as expensive.

# dof	# var	σ_1	rate	σ_2	rate	# $\mathcal{A}^h \bar{\mathbf{p}}$
81	168	5.0572		1.1886		2
289	626	11.8804	0.93	2.2487	0.81	2
1 089	2 394	13.3803	0.32	2.5896	0.40	2
4 225	9 530	16.6974	1.15	3.2168	1.29	2
16 641	38 136	18.8838	2.31	3.5678	2.38	2
66 049	151 898	19.3367	2.48	3.6283	1.87	2
263 169	605 946	19.4352		3.6510		2

TABLE 4.1. Degrees of freedom and total number of discrete state, control and adjoint variables on a hierarchy of finite element grids. Singular values and estimated rate of convergence w.r.t. grid size h for Problem 1. Number of sensitivity problems (3.5) solved.

In this and the subsequent problems, we observed monotone convergence of the computed singular values. The estimated rate of convergence given in the tables was calculated according to

$$\frac{\log \frac{|\sigma_h - \sigma_*|}{|\sigma_{2h} - \sigma_*|}}{\log 1/2},$$

where σ_* is the respective singular value on the finest mesh, and σ_h and σ_{2h} is the same value on two neighboring intermediate meshes. The exact rate of convergence is difficult to predict from the table and clearly deserves further investigation.

On the finest mesh, we obtain as singular values and right singular vectors

$$\sigma_1 = 19.3367 \quad \mathbf{r}_1 = \begin{pmatrix} -0.5103 \\ -0.7324 \end{pmatrix} \quad \sigma_2 = 3.6283 \quad \mathbf{r}_2 = \begin{pmatrix} -1.0358 \\ 0.3609 \end{pmatrix}.$$

Recall that \mathbf{r}_1 and \mathbf{r}_2 represent piecewise constant functions r_1 and r_2 on $W_1 \cup W_2$ whose values on W_1 and W_2 are given by the upper and lower entries, respectively, see Figure 4.3 (right). The corresponding left singular vectors are shown in Figure 4.3 (left). These results can be interpreted as follows: Of all perturbations of unit size (with respect to the scalar product given by M_P), the nominal state (from Figure 4.2) is perturbed most (in the $L^2(\Omega)$ -norm) when both windows are better insulated with the ratio of the improvement given by the ratio of the entries of the right singular vector \mathbf{r}_1 . The effect of this perturbation direction on the observed quantity (the optimal state) is represented by the first left singular vector $l_1 = E_H \mathbf{l}_1$, multiplied by σ_1 , compare (3.12). Due to the improved insulation at both windows, l_1 is positive, i.e., the optimal temperature increases throughout the domain Ω when p changes from p_0 to $p_0 + r_1$. Since the second entry in \mathbf{r}_1 is greater in magnitude, the effect on the optimal temperature is more pronounced near the lower window, see Figure 4.3 (top left).

Since the parameter space is only two-dimensional, the second right singular vector \mathbf{r}_2 represents the unit perturbation of *lowest* impact on the optimal state. Figure 4.3 (bottom left) shows the corresponding second left singular vector. Note

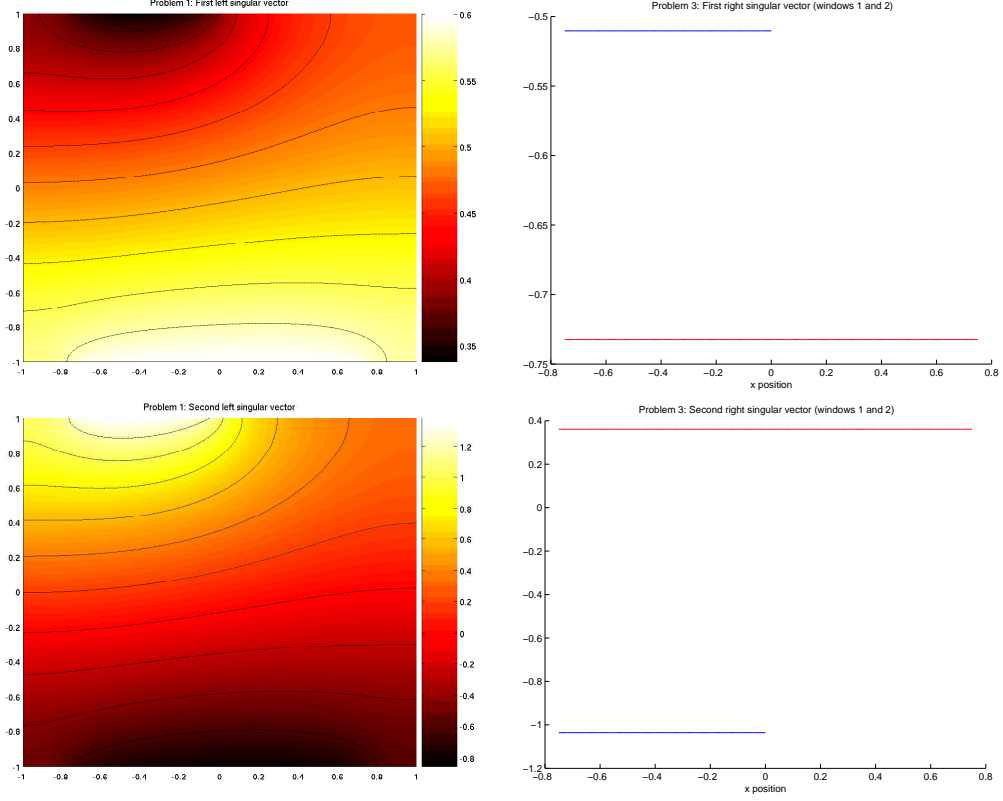


FIGURE 4.3. Problem 1: First and second left singular vectors l_1 and l_2 (left) and first and second right singular vectors (right), lower window (red) and upper window (blue).

that $\|l_1\|_{L^2(\Omega)} = \|l_2\|_{L^2(\Omega)} = 1$ and that l_1 and l_2 are perpendicular with respect to the inner product of $L^2(\Omega)$. The singular value σ_2 shows that any given perturbation of the heat transfer coefficients of unit size has at least an impact of 3.6283 on the optimal state in the $L^2(\Omega)$ -norm, to first order. This should be viewed in relation to the $L^2(\Omega)$ -norm of the nominal solution, which is 48.3982.

The data obtained from the singular value decomposition can be used to decide whether the observed quantity depending on the optimal solution is sufficiently stable with respect to perturbations. This decision should take into account the expected range of parameter variations and the tolerable variations in the observed quantity.

4.2. Problem 2: Many Parameters, Small Observation Space. In contrast to the previous situation, we now consider the window heat transfer coefficients to be spatially variable. That is, we have parameters

$$p = (\alpha(x)|_{W_1}, \alpha(x)|_{W_2}) \in L^2(W_1) \times L^2(W_2).$$

As an observed quantity, we choose the scalar value of the temperature averaged over the entire room. Hence the observation space is $H = \mathbb{R}$ and

$$\Pi(\bar{y}, \bar{u}, \bar{\lambda}) = \frac{1}{4} \int_{\Omega} \bar{y}(x) dx.$$

Such a scalar output quantity is often called a quantity of interest. The weight in the observation space is $M_H = 1$ and the mass matrix in the parameter space is the boundary mass matrix on $W_1 \cup W_2$ with respect to piecewise constant functions on the boundary of the respective finite element grid.

The matrix $\mathcal{A}^h \in \mathbb{R}^{1 \times m}$ now has only one row. It is thus strongly advisable to compute its transpose which requires only one solution of a linear system with \mathcal{K}^h . This transposition technique was already used in [6] to compute derivatives of a quantity of interest depending on an optimal solution in the presence of perturbations. As above, we show in Table 4.2 the convergence behavior of the only non-zero singular value of \mathcal{A}^h .

# dof	# var	σ_1	rate	# $\mathcal{A}^h \bar{\mathbf{p}}$
81	168	2.5381		1
289	626	5.9245	0.93	1
1 089	2 394	6.6786	0.32	1
4 225	9 530	8.3316	1.15	1
16 641	38 136	9.4157	2.31	1
66 049	151 898	9.6393	2.47	1
263 169	605 946	9.6887		1

TABLE 4.2. Problem 2: Singular value and estimated rate of convergence w.r.t. grid size h for Problem 2. Number of sensitivity problems (3.5) solved.

Figure 4.4 (right) displays the right singular vector $r_1 = E_P \mathbf{r}_1$ belonging to this problem. From this we infer that the largest increase in average temperature is achieved when the insulation at the larger (lower) window is improved to a higher degree than that of the smaller (upper) window, although the nominal insulation of the larger (lower) window is already twice as good. It is interesting to note that for the maximum impact on the average temperature, the insulation should be improved primarily near the edges of the windows. Again, the sensitivity \bar{y} of the optimal state belonging to the perturbation of greatest impact is positive throughout (Figure 4.4 (left)).

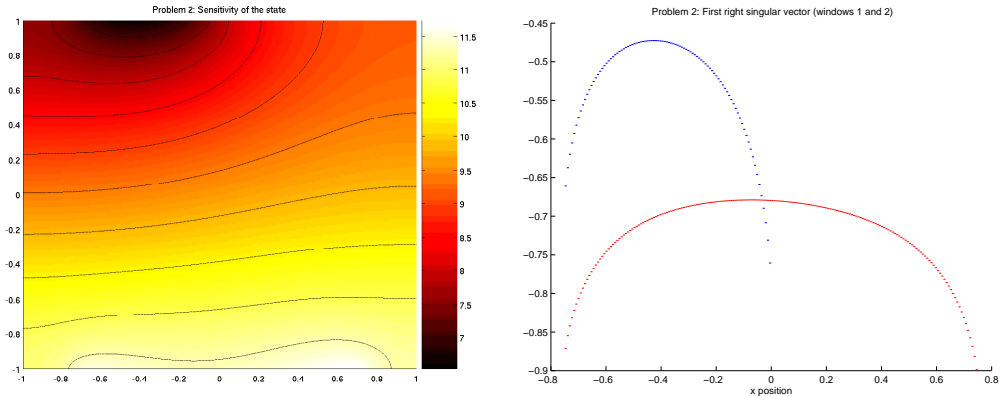


FIGURE 4.4. Problem 2: Parametric sensitivity \bar{y} (left) of the optimal state belonging to the first right singular vector r_1 (right). Lower window (red) and upper window (blue).

4.3. Problem 3: Many Parameters, Large Observation Space. The final example features both large parameter and observation spaces, so that assembling the matrices \mathcal{A}^h and XJX^{-1} as in the previous examples is prohibitive. Instead, we supply only matrix-vector products of XJX^{-1} to the iterative eigen solver. This situation is considered typical for many applications.

The parameter space is chosen as in Problem 2, and the observation is the temperature on all of Ω as in Problem 1. Table 4.3 shows again the convergence of the singular values as the mesh is uniformly refined.

# dof	# var	σ_1	rate	σ_2	rate	# $\mathcal{A}^h \bar{\mathbf{p}}$
81	168	5.0771		1.1947		40
289	626	11.9262	0.93	2.3426	0.83	68
1 089	2 394	13.4326	0.32	2.6603	0.35	68
4 225	9 530	16.7587	1.15	3.3093	1.20	68
16 641	38 136	18.9500	2.31	3.7092	2.31	68
66 049	151 898	19.4037	2.48	3.7896	2.31	68
263 169	605 946	19.5024		3.8099		68

TABLE 4.3. Problem 3: Singular values and estimated rate of convergence w.r.t. grid size h for Problem 3. Number of sensitivity problems (3.5) solved.

Note that the parameter space of Problem 1 (two constant heat transfer coefficients) is a two-dimensional subspace of the current high-dimensional parameter space. Hence, we expect the singular values for Problem 3 to be greater than those for Problem 1. This is confirmed by comparing Tables 4.1 and 4.3. However, the first two singular values σ_1 and σ_2 are only slightly larger than in Problem 1. In particular, the augmentation of the parameter space does not lead to additional perturbation directions of an impact comparable to the impact of r_1 . Comparing the right singular vector r_1 , Figure 4.5 (top right), with the right singular vector $\mathbf{r}_1 = (-0.5103, -0.7324)^\top$ from Problem 1, representing a piecewise constant function, we infer that the stronger insulation near the edges of the windows does not significantly increase the impact on the optimal state.

We also observe that the first right singular vector r_1 (Figure 4.5 (top right)) describing the perturbation of largest impact on the optimal state is very similar to the right singular vector in Problem 2, see Figure 4.4 (right), although the observed quantities are different in Problems 2 and 3.

Finally, we present in Figure 4.6 the distribution of the largest 20 singular values. Their fast decay shows that only a few singular values and the corresponding right singular vectors capture the practically significant perturbation directions of high impact for the problem at hand.

5. CONCLUSION

In this paper, we presented an approach for the quantitative stability analysis of local optimal solutions in PDE-constrained optimization. The singular value decomposition of a compact linear operator was used in order to determine the perturbation direction of greatest impact on an observed quantity which in turn depends on the solution. After a Galerkin discretization, mass matrices and their Cholesky factors naturally appear in the singular value decomposition of the discretized operator. In order to avoid forming these Cholesky factors, we described a similarity transformation of the Jordan-Wielandt matrix. A matrix-vector multiplication

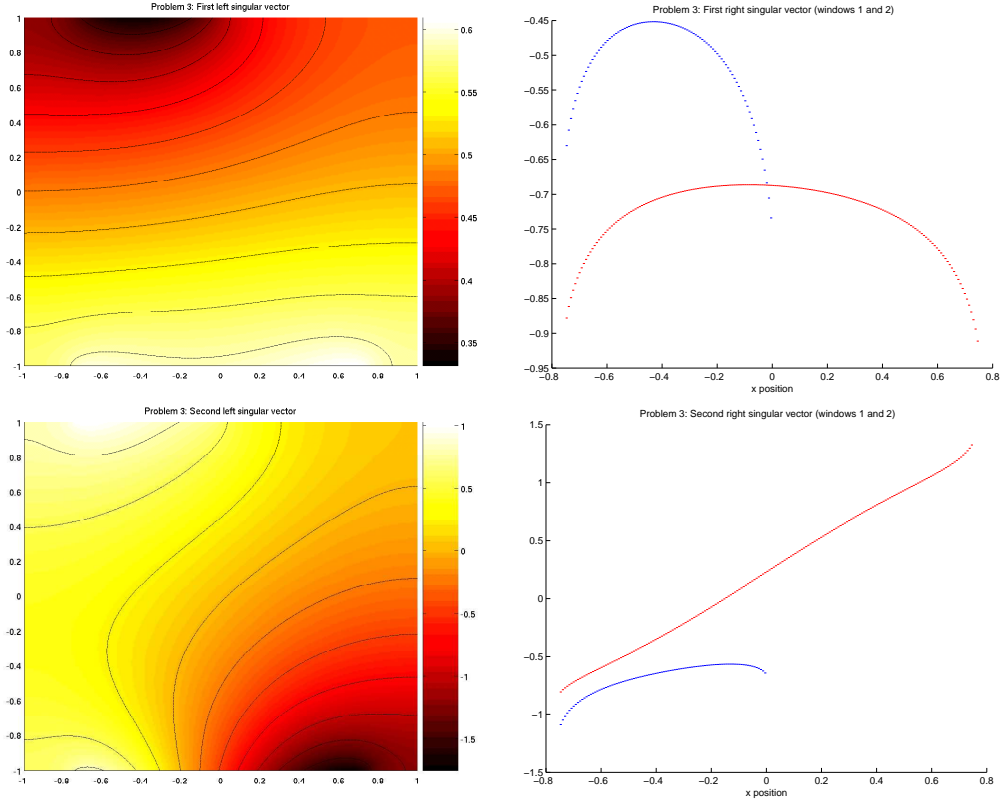


FIGURE 4.5. Problem 3: First and second left singular vectors (left) and first and second right singular vectors (right), lower window (red) and upper window (blue).

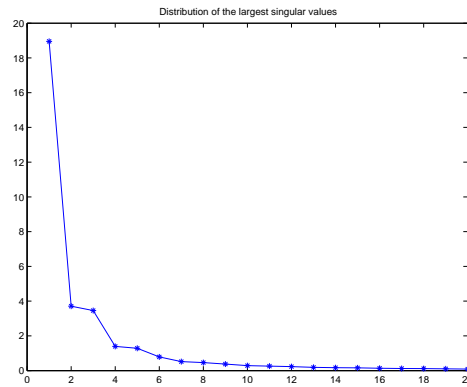


FIGURE 4.6. Problem 3: First 20 singular values.

with this transformed matrix amounts to the solution of two sensitivity problems. The desired (partial) singular value decomposition can be obtained using standard iterative eigen decomposition software, e.g., implicitly restarted Arnoldi methods.

We presented a number of numerical examples to validate the proposed method and to explain the results in the context of a concrete problem. The order of convergence of the singular values deserves further investigation. We observed that

the numerical effort even for the computation of few singular values may be large compared to the solution of the nominal problem itself. In order to accelerate the computation of the desired singular values and vectors, however, it may be sufficient to compute them on a coarser grid. In addition, parallel implementations of eigen solvers can be used.

REFERENCES

- [1] R. Adams and J. Fournier. *Sobolev Spaces*. Academic Press, New York, second edition, 2003.
- [2] K. Deimling. *Nonlinear Functional Analysis*. Springer, Berlin, 1985.
- [3] H. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Kluwer Academic Publishers, Boston, 1996.
- [4] R. Griesse. Parametric sensitivity analysis in optimal control of a reaction-diffusion system—Part I: Solution differentiability. *Numerical Functional Analysis and Optimization*, 25(1–2):93–117, 2004.
- [5] R. Griesse. Parametric sensitivity analysis in optimal control of a reaction-diffusion system—Part II: Practical methods and examples. *Optimization Methods and Software*, 19(2):217–242, 2004.
- [6] R. Griesse and B. Vexler. Numerical sensitivity analysis for the quantity of interest in PDE-constrained optimization. RICAM Report 2005–15, Johann Radon Institute for Computational and Applied Mathematics (RICAM), Austrian Academy of Sciences, Linz, Austria, 2005. <http://www.ricam.oeaw.ac.at/publications/reports/05/rep05-15.pdf>.
- [7] M. Hinze and K. Kunisch. Second order methods for optimal control of time-dependent fluid flow. *SIAM Journal on Control and Optimization*, 40(3):925–946, 2001.
- [8] I. Jolliffe. *Principal Component Analysis*. Springer, New York, second edition, 2002.
- [9] R. B. Lehoucq, D. C. Sorensen, and C. Yang. *Arpack User's Guide: Solution of Large-Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods*. Software, Environments, and Tools. SIAM, Philadelphia, 1998.
- [10] K. Malanowski. Sensitivity analysis for parametric optimal control of semilinear parabolic equations. *Journal of Convex Analysis*, 9(2):543–561, 2002.
- [11] H. Maurer and J. Zowe. First and second order necessary and sufficient optimality conditions for infinite-dimensional programming problems. *Mathematical Programming*, 16:98–110, 1979.
- [12] J. Nocedal and S. Wright. *Numerical Optimization*. Springer, New York, 1999.
- [13] L. Sirovich. Turbulence and the dynamics of coherent structures. I. *Quarterly of Applied Mathematics*, 45(3):561–571, 1987.
- [14] G. Stewart and J.-G. Sun. *Matrix Perturbation Theory*. Academic Press, New York, 1990.
- [15] F. Tröltzsch. On the Lagrange-Newton-SQP method for the optimal control of semilinear parabolic equations. *SIAM Journal on Control and Optimization*, 38(1):294–312, 1999.
- [16] F. Tröltzsch. *Optimale Steuerung partieller Differentialgleichungen*. Vieweg, Wiesbaden, 2005.
- [17] F. Tröltzsch and D. Wachsmuth. Second-order sufficient optimality conditions for the optimal control of Navier-Stokes equations. *to appear in: ESAIM: Control, Optimisation and Calculus of Variations*, 2005.
- [18] S. Volkwein. Interpretation of proper orthogonal decomposition as singular value decomposition and HJB-based feedback design. In *Proceedings of the Sixteenth International Symposium on Mathematical Theory of Networks and Systems (MTNS), Leuven, Belgium*, 2004.
- [19] D. Watkins. *Fundamentals of Matrix Computations*. Wiley-Interscience, New York, 2002.

LEHRSTUHL FÜR INGENIEURMATHEMATIK, UNIVERSITY OF BAYREUTH, UNIVERSITÄTSSTRASSE 30, D-95440 BAYREUTH, GERMANY

E-mail address: kerstin.brandes@uni-bayreuth.de

JOHANN RADON INSTITUTE FOR COMPUTATIONAL AND APPLIED MATHEMATICS (RICAM), AUSTRIAN ACADEMY OF SCIENCES, ALTENBERGERSTRASSE 69, A-4040 LINZ, AUSTRIA

E-mail address: roland.griesse@oeaw.ac.at

URL: <http://www.ricam.oeaw.ac.at/people/page/griesse>