Weierstraß–Institut
für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e.V.

Preprint ISSN 0946 – 8633

Suboptimal control of laser surface hardening using proper orthogonal decomposition

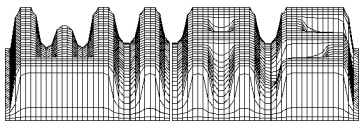
D. Hömberg[1], S. Volkwein[2]

submitted: 7 Feb 2001

[1] Weierstraß-Institut für Angewandte Analysis und Stochastik, Mohrenstraße 39, D-10117 Berlin, Germany
E-Mail:hoemberg@wias-berlin.de

[2] Institut für Mathematik, Karl-Franzens-Universität Graz, Heinrichstrasse 36, A-8010 Graz, Austria
stefan.volkwein@uni-graz.at

Preprint No. 639
Berlin 2001



1991 *Mathematics Subject Classification.* 35Kxx, 49J20, 49K20, 65Nxx.

Key words and phrases. Laser hardening, optimality conditions, proper orthogonal decomposition, error estimates, suboptimal control.

This research was supported in part by *Stiftung Industrieforschung, Köln* and by *the Fonds zur Förderung der wissenschaftlichen Forschung under the Special Research Center F003 "Optimization and Control"*..

SUBOPTIMAL CONTROL OF LASER SURFACE HARDENING USING PROPER ORTHOGONAL DECOMPOSITION

D. HÖMBERG AND S. VOLKWEIN

Abstract. Laser surface hardening of steel is formulated in terms of an optimal control problem, where the state equations are a semilinear heat equation and an ordinary differential equation, which describes the evolution of the high temperature phase. The optimal control problem is analyzed and first-order necessary optimality conditions are derived. An error estimate for POD (proper orthogonal decomposition) Galerkin methods for the state system is proved. Finally a strategy to obtain suboptimal controls using POD is developed and validated by computing some numerical examples.

Date: February 8, 2001.

1991 *Mathematics Subject Classification.* 35Kxx, 49J20, 49K20, 65Nxx.

Key words and phrases. Laser hardening, optimality conditions, proper orthogonal decomposition, error estimates, suboptimal control.

This research was supported in part by *Stiftung Industrieforschung, Köln* and by *the Fonds zur Förderung der wissenschaftlichen Forschung under the Special Research Center F003 "Optimization and Control"*.

1. Introduction

We consider a control problem that describes the laser surface hardening of steel. The mode of operation of this process, which becomes more and more important, especially in automative industry, is depicted in Figure 1.

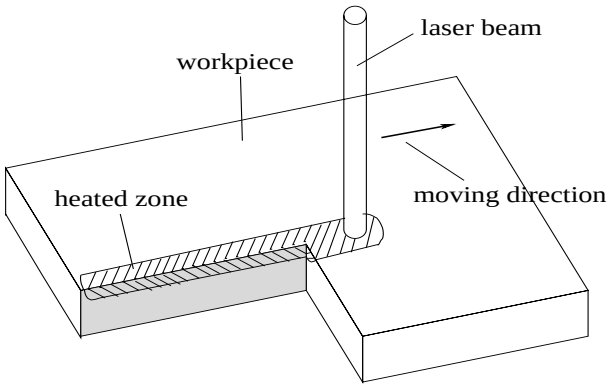


FIGURE 1.1. Sketch of a laser hardening process.

A laser beam moves along the surface of a workpiece, creating a heated zone around its trace. The heating process is accompanied by a phase transition, in which the high temperature phase in steel, called austenite, is produced. Since the penetration depth of the laser beam is very small, typically not more than $1\,mm$, the heated zone is rapidly quenched by self-cooling, leading to further phase transitions and the desired hardening effect.

Since one usually tries to keep the moving velocity of the laser beam constant, the most important control parameter is the laser energy. Whenever the temperature in the heated zone exceeds the melting temperature of steel, the work-piece quality is destroyed. Hence, the decent adjustment of laser energy is an important task, especially when the laser approaches a work-piece boundary or when there are large variations in the work-piece thickness.

In [11] the corresponding optimal control problem with pointwise state constraints on the temperature is investigated. More details about models for phase transitions in steel and the simulation of surface heat treatments can be found in [8]. In [20] a survey of mathematical models for further laser material treatments is given.

Proper orthogonal decomposition (POD) provides a method for deriving low order models of dynamical systems. It was successfully used in a variety of fields including signal analysis and pattern recognition (see e.g. [9]), fluid dynamics and coherent structures (see e.g. [5, 23]) and more recently in control theory (see e.g. [1, 2, 14, 21, 22]) and inverse problems (see [4]). Surprisingly good approximation properties are reported for POD based schemes in several articles, see [7, 13, 19], for example. Symmetry preserving properties of POD approximations are analyzed in [3]. Convergence results for POD methods applied to parabolic equations can be found in [15, 16].

The new contribution of this paper is the development of a suboptimal control strategy for laser surface hardening. We adopt a penalization approach to the state constraint problem and approximate the state equations by a semi-implicit POD Galerkin method. An error estimate for the state system is proved. This extends the analysis done in [15]. Finally, POD is used to compute suboptimal controls.

The paper is organized as follows: In Section 2 we analyze the optimal control problem, prove convergence of the penalized problem to the state constraint one and derive optimality conditions for the penalized problem. In Section 3 we propose a POD Galerkin method for the state system, present an error estimate and describe our suboptimal control strategy. The last section is devoted to numerical results.

2. The optimal control problem

In this section we specify the optimal control problem that will be considered in this paper and prove existence of an optimal solution. Moreover, we study the first-order necessary optimality conditions.

2.1. Problem statement and assumptions.

Let $\Omega \subset \mathbb{R}^3$ with Lipschitz boundary, $Q = \Omega \times (0,T)$ and $\Sigma = \partial\Omega \times (0,T)$. The formation of austenite is described by the following initial-value problem (cf. [17]):

$$a_t = f(\theta, a) = \frac{1}{\tau(\theta)}\big(a_{eq}(\theta) - a\big)\mathcal{H}\big(a_{eq}(\theta) - a\big) \qquad \text{in } Q, \tag{2.1a}$$

$$a(0) = 0 \qquad \text{in } \Omega, \tag{2.1b}$$

where a is the volume fraction of austenite. The equilibrium volume fraction a_{eq} and the time constant τ both depend on the temperature θ. Concerning the nonlinearity f we make the following assumptions:

- (A1) $a_{eq}(x) \in [0,1]$ for all $x \in \mathbb{R}$, $\|a_{eq}\|_{C^2(\mathbb{R})} \le c$;
- (A2) $0 < \underline{\tau} \le \tau(x) \le \bar{\tau}$ for all $x \in \mathbb{R}$, $\|\tau\|_{C^2(\mathbb{R})} \le c$;
- (A3) $\mathcal{H} \in C^{2,1}(\mathbb{R})$, monotone approximation of the Heaviside function satisfying $\mathcal{H}(x) = 0$ for $x \le 0$.

Since $\mathcal{H}$ is a regularized Heaviside function, the term $x\mathcal{H}(x)$ is a regularization of the positive part function

$$[x]_+ = \frac{x + |x|}{2}.$$

Thus, with (A3) holding we have $a_t \ge 0$ a.e. in Q. In other words, we only model the austenite growth during heating and keep the volume fraction when the temperature decreases again.

Assuming the density ρ, the heat capacity c_p, the heat conductivity k and the latent heat L to be positive constants, we obtain the following heat equation:

$$\begin{aligned} \rho c_p \theta_t - k\Delta\theta &= -\rho L a_t + u\alpha && \text{in } Q, \\ \frac{\partial\theta}{\partial\nu} &= 0 && \text{on } \Sigma, \\ \theta(0) &= \theta_0 && \text{in } \Omega. \end{aligned} \tag{2.1c}$$

Since the main cooling effect is the self-cooling of the workpiece, we have assumed homogeneous Neumann conditions on the boundary. The term $-\rho L a_t$ describes the consumption of latent heat due to the phase transition. The term $u(t)\alpha(x,t)$ is the volumetric heat source due to laser radiation, where the laser energy $u(t)$ will serve as a control parameter.

In addition to (A1)–(A3) we require the following assumptions:

- (A4) $\theta_0 \in H^1(\Omega)$, $\theta_0 \le \theta_m - \delta$ a.e. in Ω, where the constant $\theta_m > 0$ denotes the melting temperature of the steel and δ is a positive constant;
- (A5) $\alpha \in L^\infty(0,T;L^\infty(\Omega))$;
- (A6) $u \in L^2(0,T)$.

The goal of every heat treatment is to achieve a desired phase distribution, i.e., we consider the cost functional of tracking type

$$J(u) = \frac{\sigma}{2} \int_\Omega |a(x,T) - a_d(x)|^2 dx + \frac{\beta}{2} \int_0^T u^2 dt. \tag{2.2}$$

Here, σ and β denote positive constants and a_d is a given desired volume fraction of austenite satisfying

(A7) $a_d \in L^2(\Omega)$, $0 \le a_d \le 1$ a.e. in Ω.

We study the state and control constrained optimal control problem

$$\text{(CP)} \qquad \begin{cases} \min J(u) \\ \text{s.t. } (\theta, a, u) \text{ solves (2.1)}, \ \theta \le \theta_m \text{ a.e. in } Q \text{ and } u \in U_{ad}, \end{cases}$$

where $U_{ad} = \{u \in L^2(0,T) \ : \ \|u\|_{L^2(0,T)} \le M\}$ with some constant $M > 0$ is the closed, bounded and convex set of admissible controls.

Introducing the mapping $I : L^2(0,T) \to [0,\infty)$ by

$$I(u) = \int_0^T \int_\Omega \left[\theta - \theta_m\right]_+^2 dxdt, \tag{2.3}$$

where $\theta = \theta(u)$ solves (2.1) for $u \in L^2(0,T)$, we approximate (CP) for $\varepsilon > 0$ by the control constrained optimal control problem

$$\text{(CP}^\varepsilon\text{)} \qquad \begin{cases} \min J^\varepsilon(u) = J(u) + \dfrac{1}{2\varepsilon} I(u) \\ \text{s.t. } (\theta, a, u) \text{ solves (2.1) and } u \in U_{ad}. \end{cases}$$

Note that $I(u)$ is Fréchet-differentiable and satisfies $I(u) \ge 0$ for all $u \in U_{ad}$ and

$$I(u) = 0 \quad \Longleftrightarrow \quad \theta - \theta_m \le 0 \quad \text{a.e. in } Q.$$

2.2. Analysis of the state equations. The main result of this subsection is

Theorem 2.1. *Suppose that* (A1)–(A6) *hold. Then* (2.1) *has a unique solution*

$$(\theta, a) \in H^{1,1}(Q) \times W^{1,\infty}(0,T; L^\infty(\Omega)),$$

where $H^{1,1}(Q) = L^2(0,T; H^1(\Omega)) \cap H^1(0,T; L^2(\Omega))$ *is a Hilbert space endowed with the common inner product.*

To prove the theorem, we need the following

Lemma 2.2. *With* (A1)–(A6) *holding we have:*

a) *Let* $\theta \in L^1(Q)$, *then* (2.1a)–(2.1b) *has a unique solution satisfying*

$$0 \le a(x,t) < 1 \quad \text{a.e. in } Q \tag{2.4}$$

and

$$\|a\|_{W^{1,\infty}(0,T;L^\infty(\Omega))} \le M \tag{2.5}$$

with a constant $M > 0$ *independent of* θ.

b) *Let* $\theta_1, \theta_2 \in L^{2p}(Q)$, $1 \le p < \infty$, *and* a_1, a_2 *the corresponding solutions to* (2.1), *then there exists a constant* $C > 0$, *such that for all* $t \in [0,T]$

$$\|a_1(t) - a_2(t)\|_{L^{2p}(\Omega)} \le C \int_0^t \|\theta_1(s) - \theta_2(s)\|_{L^{2p}(\Omega)}^{2p} ds.$$

c) *Let* $\theta \in L^2(Q)$ *and let* $\{\theta_k\}_{k\in\mathbb{N}} \subset L^2(Q)$ *with* $\lim_{k\to\infty} \|\theta_k - \theta\|_{L^2(Q)} = 0$. *Then*

$$a_k \longrightarrow a \quad \text{strongly in } C([0,T];L^2(\Omega)) \cap H^1(0,T;L^2(\Omega)),$$

where a_k *and* a *are the solution to* (2.1a)–(2.1b) *with the temperature* θ_k *and* θ, *respectively.*

Proof. a) This is a direct consequence of the theorem of Carathéodory, see e.g. [26, p. 1044]. Using (A1)–(A3) and the theory of differential inequalities (cf. [10, Lemma 2.1], we obtain (2.4), whereas (2.5) is a direct consequence of (A1)–(A3).

b) Let $\theta_i \in L^{2p}(Q)$, $1 \le p < \infty$, $i = 1, 2$ and define $\theta = \theta_1 - \theta_2$, then $a = a_1 - a_2$ solves

$$a_t = f(\theta_1, a_1) - f(\theta_2, a_2). \tag{2.6}$$

Due to (A1)–(A3) the function $f(\theta, a)$ is Lipschitz-continuous in both variables. Hence, testing (2.6) with a^{2p-1}, $1 \le p < \infty$, we obtain

$$\frac{1}{2p}\int_\Omega a^{2p}(t)dx \le c_1 \int_0^t \int_\Omega |\theta| \cdot |a|^{2p-1} dx\, ds + c_2 \int_0^t \int_\Omega |a|^{2p} dx\, ds \tag{2.7}$$

for two constants $c_1, c_2 > 0$. To estimate the first additive term on the right-hand side of (2.7), we apply Young's inequality

$$cd \le \frac{c^q}{q} + \frac{d^{q'}}{q'}, \quad \text{where } c, d \ge 0 \text{ and } \frac{1}{q} + \frac{1}{q'} = 1.$$

Choosing $q = 2p$ we have $q' = 2p/(2p-1)$. Hence, we obtain

$$\int_0^t \int_\Omega |\theta| \cdot |a|^{2p-1} dx\, ds \le \frac{1}{2p} \int_0^t \int_\Omega |\theta|^{2p} dx\, ds + \frac{2p-1}{2p} \int_0^t \int_\Omega |a|^{2p} dx\, ds.$$

Inserting this bound into (2.7) and applying Gronwall's lemma, part b) of the lemma follows.

c) This part is a direct consequence of b) and Lebesgue's lemma.

□

Proof of Theorem 2.1. To obtain an a-priori estimate for θ, we test (2.1c) with θ_t and apply Young's inequality, which gives the estimate

$$\begin{aligned} \frac{\rho c_p}{2} \int_0^t \int_\Omega \theta_s^2 dx ds + \frac{k}{2} \int_\Omega |\nabla\theta(t)|^2 dx \\ \le \rho L \int_0^t \int_\Omega a_s^2 dx ds + |\Omega|\, \|\alpha\|_{L^\infty(Q)} \int_0^t u^2 ds + \frac{k}{2} \int_\Omega |\nabla\theta_0|^2 dx \end{aligned} \tag{2.8}$$

for all $t \in (0, T]$, where $|\Omega|$ denotes the Lebesgue measure of Ω. From (2.8), we conclude the estimate

$$\|\theta\|_{H^1(0,T;L^2(\Omega)) \cap L^\infty(0,T;H^1(\Omega))} \le c_1, \tag{2.9}$$

with a constant c_1 depending only on T and the data functions characterized in (A1)–(A6).

To prove the existence of a local unique solution, we apply the Banach fixed point theorem.

To this end, we define an operator $F : \hat{K} \subset L^2(Q) \to L^2(Q)$, $\hat{\theta} \mapsto \theta$, such that a is the solution to

$$\begin{aligned} a_t &= f(\hat{\theta}, a) \quad \text{in } Q, \\ a(0) &= 0 \qquad\quad \text{in } \Omega, \end{aligned}$$

and θ is the solution to (2.1c). From Lemma 2.2 we find that $a \in W^{1,\infty}(0,T;L^\infty(\Omega))$ is uniquely determined. Since (2.1c) possesses a unique solution, see [18, Theorem IV.9.1]), we can conclude that F is well-defined.

Let $\hat{K} = \{\hat{\theta} \in L^2(Q) : \|\hat{\theta}\|_{L^2(Q)} \le \hat{M} \text{ and } \theta(0) = \theta_0\}$, then with regard to (2.8) we have $F(\hat{K}) \subset \hat{K}$, provided $\hat{M}$ has been chosen large enough.

Now, we want to show that F is a contraction. Let $\hat{\theta}_i \in \hat{K}$, $i = 1, 2$, $\theta_i = F(\hat{\theta}_i)$ and $\hat{\theta} = \hat{\theta}_1 - \hat{\theta}_2$. Then $\theta = \theta_1 - \theta_2$ solves

$$\begin{aligned} \rho c_p \theta_t - k\Delta\theta &= -\rho L\big(f(\hat{\theta}_1, a_1) - f(\hat{\theta}_2, a_2)\big) \quad \text{in } Q, \\ \frac{\partial \theta}{\partial \nu} &= 0 \qquad\qquad\qquad\qquad\qquad\quad \text{on } \Sigma, \\ \theta(0) &= 0 \qquad\qquad\qquad\qquad\qquad\quad \text{in } \Omega. \end{aligned} \tag{2.10}$$

We test (2.10) with θ, integrate over Ω and over $(0,t)$, $t \in (0,T]$, and use (A3)–(A5) and Lemma 2.2 to obtain

$$\frac{\rho c_p}{2} \int_\Omega \theta(t)^2 dx + k \int_0^t \int_\Omega |\nabla\theta|^2 dx\, ds \le c_2 \int_0^t \int_\Omega \hat{\theta}^2 dx\, ds + \frac{1}{2} \int_0^t \int_\Omega \theta^2 dx\, ds.$$

Using Gronwall's Lemma it follows that

$$\int_0^t \int_\Omega \theta^2 dx\, ds \le t c_3 \int_0^t \int_\Omega \hat{\theta}^2 dx\, ds.$$

Hence, for $T^+ \le T$ small enough, F is a contraction on $L^2(0,T^+;L^2(\Omega))$. Since F is also a self-mapping on $\hat{K}$, we can apply the Banach fixed point theorem to conclude that F has a unique fixed point θ, which is a local solution to (2.1). In view of the global a priori estimate (2.9), we can extend the solution to the whole time interval $[0,T]$ by a standard bootstrap argument. □

A direct consequence of Lemma 2.2 and (2.8) is the following stability estimate, which will be used later on:

Lemma 2.3. *Suppose that* (A1)–(A6) *hold and let* θ_1, θ_2 *be the solutions to* (2.1c) *corresponding to* $u_1, u_2 \in L^2(0,T)$. *Then there exists a constant* $C > 0$ *such that*

$$\|\theta_1 - \theta_2\|_{H^{1,1}(Q)} \le C \, \|u_1 - u_2\|_{L^2(0,T)}.$$

In view of Theorem 2.1 and Lemmas 2.2 and 2.3, the solution operator

$$S : L^2(0,T) \to H^{1,1}(Q) \times C([0,T];L^2(\Omega)), \qquad u \mapsto (\theta, a), \tag{2.11}$$

where (θ, a) is the solution to the state system (2.1) is well-defined and Lipschitz continuous. Moreover, we have

Lemma 2.4. *Assume that* (A1)–(A6) *hold. Then* S *is Fréchet-differentiable and its derivative*

$$S'(u)h = (v, w)$$

is characterized by the linearized state equations

$$\rho c_p v_t - k\Delta v = -\rho L w_t + h\alpha \qquad \text{in } Q, \tag{2.12a}$$

$$\frac{\partial v}{\partial \nu} = 0 \qquad \text{in } \Sigma, \tag{2.12b}$$

$$v(0) = 0 \qquad \text{in } \Omega, \tag{2.12c}$$

$$w_t = f_\theta(\theta, a)v + f_a(\theta, a)w \qquad \text{in } Q, \tag{2.12d}$$

$$w(0) = 0 \qquad \text{in } \Omega, \tag{2.12e}$$

where $(\theta, a) = S(u)$.

Proof. The existence of a unique solution to the linearized state system (2.12) can be derived similar to the proof of Theorem 2.1. Let (θ^h, a^h) be the solution to the state system (2.1) corresponding to the control $u + h$. Defining

$$p = \theta^h - \theta - v, \quad q = a^h - a - w, \tag{2.13}$$

it remains to show that

$$\|(p, q)\|_{H^{1,1}(Q)\times C([0,T];L^2(\Omega))} = o\big(\|h\|_{L^2(0,T)}\big). \tag{2.14}$$

Using a first-order Taylor expansion of f (cf. (A1)–(A3)), it follows that (p, q) satisfies

$$\rho c_p p_t - k\Delta p = -\rho L q_t \qquad \text{in } Q, \tag{2.15a}$$

$$\frac{\partial p}{\partial \nu} = 0 \qquad \text{in } \Sigma, \tag{2.15b}$$

$$p(0) = 0 \qquad \text{in } \Omega, \tag{2.15c}$$

$$q(0) = 0 \qquad \text{in } \Omega, \tag{2.15d}$$

$$\begin{aligned} q_t &= f(\theta^h, a^h) - f(\theta, a) - f_\theta(\theta, a)v - f_a(\theta, a)w \\ &= \Big[f_\theta(\theta^\xi, a^\xi) - f_\theta(\theta, a)\Big](\theta^h - \theta) + \Big[f_a(\theta^\xi, a^\xi) - f_a(\theta, a)\Big](a^h - a) \\ &\quad - f_\theta(\theta, a)p - f_a(\theta, a)q, \end{aligned} \tag{2.15e}$$

where $(\theta^\xi, a^\xi) = (\theta + \xi(\theta^h - \theta), a + \xi(a^h - a))$ with some $\xi \in (0, 1)$. Recall that $a \in L^\infty(0, T; L^\infty(\Omega))$. Testing (2.15e) with q, integrating over $Q_t = \Omega \times (0, t)$ and using (A1)–(A3) as well as Youngt's inequality and Lemma 2.2-b) there exists a constant $c_1 > 0$ such that

$$\int_\Omega q^2(t)\,dx \quad \le \quad c_1 \int_0^t \int_\Omega \big((\theta^\xi - \theta)^4 + p^2 + q^2\big)dxds.$$

In view of Gronwall's lemma, we obtain

$$\int_\Omega q^2(t)\,dx \le c_2 \int_0^t \int_\Omega (\theta^h - \theta)^4\,dxds + c_2 \int_0^t \int_\Omega p^2\,dxds \tag{2.16}$$

and then, by comparison in (2.15e)

$$\int_0^t \int_\Omega q_s^2\,dxds \le c_3 \int_0^t \int_\Omega (\theta^h - \theta)^4\,dxds + c_4 \int_0^t \int_0^s \int_\Omega p^2\,dxd\tilde{s}. \tag{2.17}$$

We test (2.15a) with p_t and obtain an estimate similar to (2.8), from which we can infer

$$\|p\|^2_{H^{1,1}(Q_t)} \le c_5 \int_0^t \int_\Omega (\theta^h - \theta)^4 \, dx ds + c_6 \int_0^t \|p\|^2_{L^2(0,s;\Omega)} \, ds. \tag{2.18}$$

Invoking Gronwalls' lemma once again, the continuous embedding $H^{1,1}(Q_t) \subset L^4(Q_t)$, valid for $\dim \Omega \le 3$, and Lemma 2.3 concludes the proof. □

2.3. Existence of optimal control. To prove existence of an optimal control we need

Lemma 2.5. *With* (A1)–(A6) *holding the solution operator*

$$\mathcal{S}\colon L^2(0,T) \to L^2(Q) \times C([0,T];L^2(\Omega))$$

is compact, i.e. for any sequence $\{u_n\}_{n\in\mathbb{N}} \subset L^2(0,T)$, $u_n \to u$ *weakly, we have* $(\theta_n, a_n) \to (\theta, a)$ *strongly in* $L^2(Q) \times C([0,T];L^2(\Omega))$, *where* $(\theta, a) = \mathcal{S}(u)$ *is the solution to* (2.1) *with respect to the control* u.

Proof. Since $\{u_n\}_{n\in\mathbb{N}}$ is bounded in $L^2(0,T)$, (2.9) applies and there exists a subsequence $\{\theta_n\}_{n\in\mathbb{N}}$ still indicated by n satisfying $\theta_n \to \theta$ weakly in $H^{1,1}(Q)$ and strongly in $L^2(Q)$. Using Lemma 2.2-c) we also have $a_n \to a$ strongly in $C([0,T];L^2(\Omega)) \cap H^1(0,T;L^2(\Omega))$. Hence it is easy to see that we can pass to the limit in the state equations (2.1). Since their solution is uniquely defined, the convergence holds for the whole sequence. □

Theorem 2.6. *With* (A1)–(A7) *holding* (CP) *has an optimal solution* u^*.

Proof. Let $K = \{u \in U_{ad} \mid \theta(u) \le \theta_m \text{ a.e. in } Q\}$. We proceed in 3 steps:

a) $\operatorname{int} K \neq \emptyset$: Let $u \equiv 0$, then $u \in U_{ad}$. Moreover, testing (2.1c) with $[\theta - \theta_0]_+$, we obtain from $a_t \ge 0$ a.e. in Q that

$$\frac{1}{2}\int_\Omega [\theta(t) - \theta_0]_+^2 \, dx + k \int_0^t \int_\Omega |\nabla[\theta - \theta_0]_+|^2 dx \, ds = -\rho L \int_0^t \int_\Omega a_t [\theta - \theta_0]_+ dx \, ds \le 0.$$

Thus, $\theta \le \theta_0$ a.e. in Q and in view of (A4) and the stability estimate of Lemma 2.3 we see that $\operatorname{int} K \neq \emptyset$.

b) $(\mathsf{CP}^\varepsilon)$ has a solution u^ε: For $\varepsilon > 0$ fixed, we take a minimizing sequence $\{u_n^\varepsilon\}_{n\in\mathbb{N}} \subset U_{ad}$ such that $\lim_{n\to\infty} J^\varepsilon(u_n^\varepsilon) = \inf_{u\in U_{ad}} J^\varepsilon(u)$. Since $\{u_n^\varepsilon\}_{n\in\mathbb{N}}$ is bounded, there exists a subsequence with $u_n^\varepsilon \to u^\varepsilon$ weakly in $L^2(0,T)$. Applying Lemma 2.5, we obtain

$$(\theta_n^\varepsilon, a_n^\varepsilon) \to (\theta^\varepsilon, a^\varepsilon) \text{ strongly in } L^2(Q) \times C([0,T];L^2(\Omega)),$$

and $(\theta^\varepsilon, a^\varepsilon)$ solve the state system for the control u^ε. Owing to the weak lower semi-continuity of norms, we can pass to the limit in $J^\varepsilon(u)$, hence $(\mathsf{CP}^\varepsilon)$ has a solution u^ε.

c) Passing to the limit with $\varepsilon \to 0$: Reasoning as before, there exists a subsequence $\{u^\varepsilon\}_{\varepsilon>0}$ still indicated by ε, such that

$$\begin{aligned} u^\varepsilon &\longrightarrow u^* && \text{weakly in } L^2(0,T) \\ \theta^\varepsilon &\longrightarrow \theta^* && \text{strongly in } L^2(Q) \\ a^\varepsilon &\longrightarrow a^* && \text{strongly in } C([0,T];L^2(\Omega)) \end{aligned}$$

and $(\theta^*, a^*) = S(u^*)$. Moreover, we have

$$J^\varepsilon(u^\varepsilon) = J(u^\varepsilon) + \frac{1}{2\varepsilon} I(u^\varepsilon) \le J(\hat{u}) \quad \text{for all } \hat{u} \in K$$

and thus

$$I(u^\varepsilon) \le 2\varepsilon J(\hat{u}) + 2\varepsilon J(u^\varepsilon). \tag{2.19}$$

Passing to the limit in (2.19), we obtain $I(u^*) = 0$, i.e., $\theta^* \le \theta_m$ a.e. in Q. Finally, we see

$$J(u^\varepsilon) \le J(u^\varepsilon) + \frac{1}{2\varepsilon} I(u^\varepsilon) \le J(\hat{u}) \text{ for any } \hat{u} \in K. \tag{2.20}$$

From (2.20) we infer

$$J(u^*) \le \liminf_{\varepsilon \to 0} J(u^\varepsilon) \le J(\hat{u}) \text{ for any } \hat{u} \in K$$

so that u^* is a solution to (CP).

□

2.4. First-order optimality conditions. In the following theorem the first-order necessary optimality conditions for (CP^ε) are characterized via the adjoint equations.

Theorem 2.7. *Suppose that* (A1)-(A7) *hold. Let $u \in U_{ad}$ be a solution to* (CP^ε) *and $(\theta, a) = \mathcal{S}(u)$ be the corresponding solution to the state system. Then there exists a unique solution $(p, q) \in H^{1,1}(Q) \times H^1(0, T; L^2(\Omega))$ of the adjoint system*

$$-\rho c_p p_t - k\Delta p = f_\theta(\theta, a)(q - \rho L p) + \frac{1}{\varepsilon}[\theta - \theta_m]_+ \qquad \text{in } Q, \tag{2.21a}$$

$$\frac{\partial p}{\partial \nu} = 0 \qquad \text{in } \Sigma, \tag{2.21b}$$

$$p(T) = 0 \qquad \text{in } \Omega, \tag{2.21c}$$

$$-q_t = f_a(\theta, a)(q - \rho L p) \qquad \text{in } Q, \tag{2.21d}$$

$$q(T) = \sigma(a(T) - a_d) \qquad \text{in } \Omega. \tag{2.21e}$$

Moreover, p satisfies the variational inequality

$$\int_0^T \int_\Omega (\alpha p + \beta u)(\tilde{u} - u)\, dx\, dt \ge 0 \quad \text{for all } \tilde{u} \in U_{ad}. \tag{2.22}$$

Proof. The existence of a unique solution to (2.21) can be derived similar to the proof of Theorem 2.1. For brevity we write $f_\theta = f_\theta(\theta, a)$ and analogously $f_a = f_a(\theta, a)$. To show that (p, q) are the adjoint variables, we test (2.12d), (2.12e) with q, integrate over Q and use (2.21d), (2.21e) to obtain:

$$\begin{aligned} 0 &= \int_0^T \int_\Omega (w_t - f_\theta v - f_a w) q\, dx\, dt \\ &= \int_0^T \int_\Omega \big((-q_t - f_a q) w - f_\theta v q\big) dx\, dt + \int_\Omega (qw)(T) - (qw)(0)\, dx \\ &= \int_0^T \int_\Omega \big(- \rho L f_a w p - f_\theta v q\big) dx dt + \sigma \int_\Omega (a(T) - a_d) w(T) dx. \end{aligned} \tag{2.23}$$

Next, we test (2.12a)–(2.12c) with p and use (2.23):

$$
\begin{aligned}
0 &= \int_0^T \int_\Omega (\rho c_p v_t - k\Delta v + \rho L f_\theta v + \rho L f_a w - h\alpha) p \, dx \, dt \\
&= \int_0^T \int_\Omega \left((-\rho c_p p_t - k\Delta p + \rho L f_\theta p) v + \rho L f_a w p - h\alpha p \right) dx dt \\
&= \int_0^T \int_\Omega \left(\frac{1}{\varepsilon} \, [\theta - \theta_m]_+ v + f_\theta q v + (\rho L f_a w - h\alpha) p \right) dx dt \\
&= \int_0^T \int_\Omega \left(\frac{1}{\varepsilon} \, [\theta - \theta_m]_+ v - h\alpha p \right) dx dt + \int_\Omega \sigma(a(T) - a_d) w(T) \, dx.
\end{aligned}
\tag{2.24}
$$

For every $h \in L^2(0,T)$ such that $u + h \in U_{ad}$, we have

$$J(u) \le J(u+h).$$

Applying Theorem 2.6 and (2.24) we have

$$
\begin{aligned}
0 \;\le\; J'(u)h &= \sigma \int_\Omega (a(T) - a_d) w \, dx + \frac{1}{\varepsilon} \int_0^T \int_\Omega [\theta - \theta_m]_+ v \, dx dt + \beta \int_0^T uh \, dt \\
&= \int_0^T \int_\Omega h(\alpha \, p + \beta u) \, dx \, dt.
\end{aligned}
$$

Since U_{ad} is convex, we have derived (2.22), which finishes the proof. □

3. Suboptimal control utilizing POD

This section is devoted to a discussion of the POD method for the optimal control problem (CP^ε). We analyze a semi-implicit POD Galerkin scheme for the state equations (2.1) and present an error estimate. Moreover, we describe the reduced-order modeling for (CP^ε) that is used in Section 4.

3.1. The POD method. Let $u \in U_{ad}$ be arbitrary. Throughout we denote by (θ, a) the unique solution to the state equations (2.1) satisfying $(\theta, a) \in H^{1,1}(Q) \times C([0,T]; L^\infty(\Omega))$.

For given $n \in \mathbb{N}$ let

$$0 = t_0 < t_2 < \ldots < t_n = T$$

be an equidistant grid in the interval $[0,T]$ with time step $\Delta t = T/n$. Suppose that the snapshots $\theta(t_j)$ of (2.1c) at the given time instances t_j, $j = 0, \ldots, n$, are known. We set

$$v_j = \begin{cases} \theta(t_j) & \text{for } j = 0, \ldots, n, \\ \overline{\partial}_t \theta(t_{j-n}) & \text{for } j = n+1, \ldots, 2n, \end{cases}$$

where

$$\overline{\partial}_t \theta(t_j) = \frac{\theta(t_j) - \theta(t_{j-1})}{\Delta t},$$

and introduce the subspace

$$\mathcal{V} = \text{span } \{v_0, \ldots, v_{2n}\}.$$

We refer to $\mathcal{V}$ as the ensemble consisting of the snapshots $\{v_j\}_{j=0}^{2n}$, at least one of which is assumed to be nonzero. Notice that $\mathcal{V} \subset H^1(\Omega)$ by construction.

Throughout the remainder of this section we denote by X either the space $H^1(\Omega)$ or $L^2(\Omega)$ endowed with their common inner products.

Let $\{\psi_i\}_{i=1}^d$ denote an orthonormal basis for $\mathcal{V}$ with $d = \dim \mathcal{V}$. Then each member of the ensemble can be expressed as

$$v_j = \sum_{i=1}^{d} (v_j, \psi_i)_X \psi_i \quad \text{for } j = 0, \ldots, 2n, \tag{3.1}$$

where $(\cdot, \cdot)_X$ denotes the inner product in X. The method of POD consists in choosing an orthonormal basis such that for every $\ell \in \{1, \ldots, d\}$ the mean square error between the snapshots $\{v_j\}_{j=0}^{2n}$ and the corresponding ℓ-th partial sum of (3.1) is minimized on average:

$$\begin{cases} \min\limits_{\{\psi_i\}_{i=1}^{\ell}} \sum\limits_{j=0}^{2n} \alpha_j \left\| v_j - \sum\limits_{i=1}^{\ell} (v_j, \psi_i)_X \psi_i \right\|_X^2 \\ \text{s.t. } (\psi_i, \psi_j)_X = \delta_{ij} \quad \text{for } 1 \le i \le \ell,\ 1 \le j \le i. \end{cases} \tag{3.2}$$

Here $\{\alpha_j\}_{j=0}^{2n}$ are positive weights, which for our purposes are chosen to be

$$\alpha_j = \begin{cases} \dfrac{\Delta t}{2} & \text{if } j \in \{0, n, 2n\}, \\ \Delta t & \text{otherwise.} \end{cases}$$

A solution $\{\psi_i\}_{i=1}^{\ell}$ to (3.2) is called POD basis of rank ℓ. The subspace spanned by the first ℓ POD basis functions is denoted by V^{ℓ}.

Remark 3.1. Note that

$$\mathcal{I}_n^1(\theta) = \sum_{j=0}^{n} \alpha_j \left\| v_j - \sum_{i=1}^{\ell} (v_j, \psi_i)_X \psi_i \right\|_X^2$$

is the trapezoidal approximation for the integral

$$\mathcal{I}^1(\theta) = \int_0^T \left\| \theta(t) - \sum_{i=1}^{\ell} (\theta(t), \psi_i)_X \psi_i \right\|_X^2 dt.$$

Moreover, the term

$$\mathcal{I}_n^2(\theta) = \sum_{j=n+1}^{2n} \alpha_j \left\| v_j - \sum_{i=1}^{\ell} (v_j, \psi_i)_X \psi_i \right\|_X^2$$

can also be interpreted as a trapezoidal approximation for the integral

$$\mathcal{I}^2(\theta) = \int_0^T \left\| \theta_t(t) - \sum_{i=1}^{\ell} (\theta_t(t), \psi_i)_X \psi_i \right\|_X^2 dt.$$

where, in addition, the time derivatives are discretized by difference quotients. Therefore, $\mathcal{I}_n = \mathcal{I}_n^1 + \mathcal{I}_n^2$ is an approximation for the integral $\mathcal{I} = \mathcal{I}^1 + \mathcal{I}^2$. For $\theta \in W^{2,2}(0, T; X)$ we have

$$\lim_{\Delta t \to 0} \|\mathcal{I}_n - \mathcal{I}\|_{\mathcal{L}(X)} = 0,$$

where $\mathcal{L}(X)$ denotes the Banach space of all bounded linear operators on X. $\diamond$

Using a Lagrangian framework the solution to (3.2) is characterized by the following optimality conditions:

$$\mathcal{R}\psi = \lambda\psi, \tag{3.3}$$

see [25], where $\mathcal{R} : X \to X$ is given by

$$\mathcal{R}z = \sum_{j=0}^{2n} \alpha_j \, (z, v_j)_X v_j \quad \text{for } z \in X.$$

Note that $\mathcal{R}$ is a linear, bounded, self-adjoint and nonnegative operator. Moreover, since the image of $\mathcal{R}$ has finite dimension, $\mathcal{R}$ is also compact. By Hilbert–Schmidt theory (see e.g. [24, p. 203]) there exist an orthonormal basis $\{\psi_i\}_{i\in\mathbb{N}}$ for X and a sequence $\{\lambda_i\}_{i\in\mathbb{N}}$ of nonnegative real numbers so that

$$\mathcal{R}\psi_i = \lambda_i \psi_i, \quad \lambda_1 \geq \ldots \geq \lambda_d > 0 \quad \text{and} \quad \lambda_i = 0 \quad \text{for} \quad i > d, \tag{3.4}$$

Moreover, $\mathcal{V} = \text{span } \{\psi_i\}_{i=1}^{d}$.

Note that $\mathcal{R}$ and thus $\{\lambda_i\}_{i\in\mathbb{N}}$ as well as $\{\psi_i\}_{i\in\mathbb{N}}$ depend on n. In what follows the notation of this dependence is dropped.

The sequence $\{\psi_i\}_{i=1}^{\ell}$ solves the optimization problem (3.2). This fact and the error formula below were proved in [5], for example.

Proposition 3.2. *Let $\lambda_1 \geq \ldots \geq \lambda_d > 0$ denote the positive eigenvalues of $\mathcal{R}$ with the associated eigenvectors $\psi_1, \ldots, \psi_d \in X$. Then, $\{\psi_i\}_{i=1}^{\ell}$ is a POD basis of rank $\ell \leq d$, and we have the error formula*

$$\sum_{j=0}^{2n} \alpha_j \left\| v_j - \sum_{i=1}^{\ell} (v_j, \psi_i)_X \psi_i \right\|_X^2 = \sum_{i=\ell+1}^{d} \lambda_i. \tag{3.5}$$

Remark 3.3. The POD basis of rank ℓ can be computed as follows: First solve the eigenvalue problem

$$\mathcal{K} w_i = \lambda_i w_i \quad \text{for } i = 1, \ldots, \ell, \tag{3.6}$$

where the positive semidefinite $(2n+1) \times (2n+1)$-matrix $\mathcal{K}$ has the elements $\mathcal{K}_{ij} = (v_{j+1}, v_{i+1})_X$ and the nonnegative eigenvalues satisfy $\lambda_1 \geq \ldots \geq \lambda_d$. Then for $\ell \leq d$ we find

$$\psi_i = \frac{1}{\sqrt{\lambda_i}} \sum_{j=0}^{2n} \alpha_j w_i^j v_j \quad \text{for } i = 1, \ldots, \ell.$$

Here w_i^j denotes the j-th component of the eigenvector w_i. $\diamond$

3.2. A POD GALERKIN SCHEME FOR THE STATE EQUATIONS. For $r \in \mathbb{N}$ we set $m = rn$ and introduce the time grid

$$\tau_j = j\Delta\tau \quad \text{for } j = 0, \ldots, m \quad \text{with} \quad \Delta\tau = \frac{T}{m}.$$

Note that for $r = 1$ the t- and τ-grids coincide. The problem consists in finding a sequence $\{(\theta_\ell^j, a_\ell^j)\}_{j=0}^{m}$ as follows: Solve

$$a^0 = 0 \quad \text{and} \quad (\theta_\ell^0, \psi)_{L^2(\Omega)} = (\theta_0, \psi)_{L^2(\Omega)} \quad \text{for all } \psi \in V^\ell. \tag{3.7a}$$

Then, for $j = 1, \ldots, m$ compute θ_ℓ^j by solving

$$\begin{aligned} &\rho c_p (\overline{\partial}_\tau \theta_\ell^j, \psi)_{L^2(\Omega)} + k(\nabla\theta_\ell^j, \nabla\psi)_{L^2(\Omega)} \\ &\qquad = (u(t_j)\alpha(t_j) - \rho L f(\theta_\ell^{j-1}, a_\ell^{j-1}), \psi)_{L^2(\Omega)} \quad \text{for all } \psi \in V^\ell, \end{aligned} \tag{3.7b}$$

where $\overline{\partial}_\tau \theta_\ell^j = (\theta_\ell^j - \theta_\ell^{j-1})/\Delta\tau$ and use θ_ℓ^j to get a_ℓ^j from

$$\overline{\partial}_\tau a_\ell^j = f(\theta_\ell^j, a_\ell^{j-1}) \quad \text{a.e. in } \Omega. \tag{3.7c}$$

To prove an error estimate for the scheme (3.7) we need more regularity for θ_0, α, and u. Therefore, we replace (A4)–(A6) with

(A4') $\theta_0 \in H^3(\Omega)$;

(A5') $\alpha \in W^{1,\infty}(0,T;L^\infty(\Omega))$;
(A6') $u \in H^1(0,T)$.

Theorem 3.4. *Suppose that* (A1)–(A3) *and* (A4')–(A6') *hold and that the t- and τ-grids coincide. Let (θ, a) be the unique solution of* (2.1). *We assume that* (3.7) *has a unique solution $\{(\theta_\ell^j, a_\ell^j)\}_{j=0}^n$. If Δt is sufficiently small, then there exists a constant $C > 0$ depending on θ, a, T, but independent of ℓ and n, such that*

$$\begin{aligned} &\sum_{j=0}^{n} \alpha_j \ \|\theta(t_j) - \theta_\ell^j\|_{L^2(\Omega)}^2 + \max_{0\le j\le n} \|a(t_j) - a_\ell^j\|_{L^2(\Omega)}^2 \\ &\qquad \le C\Big(\|S\|_2 \sum_{i=\ell+1}^{d} \Big(|(\psi_i, \theta_0)_X|^2 + \lambda_i\Big) + (\Delta t)^2\Big), \end{aligned} \tag{3.8}$$

where S denotes the stiffness matrix given by

$$S = ((S_{ij})) \in \mathbb{R}^{\ell\times\ell} \quad \text{with} \quad S_{ij} = (\psi_j, \psi_i)_{H^1(\Omega)} \tag{3.9}$$

and $\|\cdot\|_2$ stands for the spectral norm for symmetric matrices.

Remark 3.5. a) Note that in case $X = H^1(\Omega)$ we have $\|S\|_2 = 1$ by construction (cf. (3.2).

b) In [16] non-equidistant time grids, which need not coincide, were considered for a general equation in fluid dynamic. An analogous analysis for our semilinear problem can be the focus of a future research.

c) In (3.8) the eigenvalues and eigenfunctions depend on n, i.e., $\lambda_i = \lambda_i^n$ and $\psi_i = \psi_i^n$. Using spectral theory we can derive bounds that are independent of n, see [16]. $\diamond$.

For the proof we will make use of the following result, which is taken from [15]:

Lemma 3.6. *Let the so-called Ritz projection $P^\ell : V \to V^\ell$, $1 \le \ell \le d$, be given by*

$$(P^\ell\varphi, \psi)_{H^1(\Omega)} = (\varphi, \psi)_{H^1(\Omega)} \quad \text{for all } \psi \in V^\ell, \tag{3.10}$$

where $\varphi \in V$. Assume that $\theta \in H^2(0,T;L^2(\Omega))$ and define

$$\varrho^j = P^\ell\theta(t_j) - \theta(t_j) \quad \text{and} \quad \zeta^j = \theta_t(t_j) - \overline{\partial}_t P^\ell\theta(t_j).$$

Then, there exists a constant $C > 0$, independent of ℓ such that

$$\begin{aligned} \Delta t \sum_{j=0}^{n} \|\varrho^j\|_{L^2(\Omega)}^2 &\le C\|S\|_2 \sum_{i=\ell+1}^{d} \lambda_i, \\ \Delta t \sum_{j=0}^{n} \|\zeta^j\|_{L^2(\Omega)}^2 &\le C\Big((\Delta t)^2 \int_0^T \|\theta_{tt}(t)\|_{L^2(\Omega)}^2\, dt + \|S\|_2 \sum_{i=\ell+1}^{d} \lambda_i\Big), \end{aligned}$$

where S denotes the stiffness matrix introduced in (3.9).

Moreover, we need the following technical

Lemma 3.7. *Let ξ_i, $i \in \mathbb{N}$, be nonnegative numbers satisfying the recursion*

$$\xi_0 \le B \quad \text{and} \quad \xi_i \le (1+\delta)\xi_{i-1} + B \quad \text{for } i = 1, 2, \ldots .$$

Then

$$\xi_i \le \frac{B}{\delta}\big(e^{(i+1)\delta} - 1\big).$$

Proof. Summation upon i it follows from $\xi_0 \le B$ that

$$\begin{aligned}\xi_i &\le (1+\delta)\xi_{i-1} + B \le (1+\delta)^i \xi_0 + B\sum_{j=0}^{i-1}(1+\delta)^j \\ &\le B\sum_{j=0}^{i}(1+\delta)^j = \frac{B}{\delta}\big((1+\delta)^{i+1}-1\big).\end{aligned}$$

Utilizing the inequality $1+x \le e^x$ the claim follows directly. □

Proof of Theorem 3.4. Due to (A4')–(A6') the right-hand side of (2.1a) belongs to $H^1(0,T;L^2(\Omega))$. Therefore, we can differentiate formally with respect to time:

$$a_{tt} = f_\theta(\theta,a)\theta_t + f_a(\theta,a)a_t \quad \text{in } Q. \tag{3.11}$$

In view of (2.8) and Lemma 2.2-a), we obtain

$$\|a_{tt}\|_{L^2(Q)} \le c_1 \tag{3.12}$$

for a constant $c_1 > 0$. Now we can also differentiate (2.1c) with respect to time, which results in

$$\begin{aligned}\rho c_p \theta_{tt} - k\Delta\theta_t &= -\rho L a_{tt} + u_t\alpha + u\alpha_t && \text{in } Q,\\ \frac{\partial\theta_t}{\partial\nu} &= 0 && \text{on } \Sigma,\\ \theta_t(0) &= \frac{1}{\rho c_p}\Big(k\Delta\theta_0 - \rho L f(\theta_0,0) + u(0)\alpha(0)\Big) && \text{in } \Omega.\end{aligned}$$

Note that according to (A4')–(A6') we have $\theta_t(0) \in H^1(\Omega)$, hence we can test with θ_{tt} and find that there exists a constant $c_2 > 0$ satisfying

$$\rho c_p \int_0^t \|\theta_{ss}(s)\|^2_{L^2(\Omega)}\,ds + \frac{k}{2}\,\|\nabla\theta_t(t)\|^2_{L^2(\Omega)} \le c_2. \tag{3.13}$$

In view of (3.13) we have $\theta_t \in C([0,T];L^2(\Omega))$. From (3.12) we infer that also $a_t \in C([0,T];L^2(\Omega))$. Hence, for $x \in \Omega \setminus O$, where O is a set of measure zero, we can consider a second-order Taylor expansion for a. Using (2.1a) and (3.11) we obtain

$$\begin{aligned}a(t_j+\Delta t) &= a(t_j) + \Delta t a_t(t_j) + \frac{1}{2}(\Delta t)^2 a_{tt}(t_j+\xi)\\ &= a(t_j) + \Delta t f(\theta(t_j),a(t_j)) + \frac{1}{2}(\Delta t)^2 f_\theta(\theta(t_j+\xi),a(t_j+\xi))\theta_t(t_j+\xi)\\ &\quad + \frac{1}{2}(\Delta t)^2 f_a(\theta(t_j+\xi),a(t_j+\xi))a_t(t_j+\xi),\end{aligned} \tag{3.14}$$

for a $\xi \in (0,\Delta t)$. Moreover, there exists a constant $c_3 > 0$ satisfying

$$\|f_\theta(\theta(t_j+\xi),a(t_j+\xi))\theta_t(t_j+\xi) + f_a(\theta(t_j+\xi),a(t_j+\xi))a_t(t_j+\xi)\|_{L^2(\Omega)} \le 2c_3$$

Now, we define

$$\overline{a}^j = a^j_\ell - a(t_j) \quad \text{and} \quad \overline{\theta}^j = \theta^j_\ell - \theta(t_j).$$

Note that $\overline{\theta}^0 = \overline{a}^0 = 0$ and that the term $\overline{\partial}_t\theta(t_{j+1})$ is bounded because of (3.13). Thus, in view of (A1)–(A3), we obtain for the difference of (3.7c) and (3.14)

$$\begin{aligned}\|\overline{a}^{j+1}\|_{L^2(\Omega)} &\le \|a^j_\ell\|_{L^2(\Omega)} + \Delta t\,\|f(\theta^{j+1}_\ell,a^j_\ell) - f(\theta(t_j),a(t_j))\|_{L^2(\Omega)}\\ &\quad + c_3(\Delta t)^2\\ &\le (1+c_4\Delta t)\,\|\overline{a}^j\|_{L^2(\Omega)} + c_5\Delta t\,\|\overline{\theta}^{j+1}\|_{L^2(\Omega)} + c_6(\Delta t)^2\end{aligned} \tag{3.15}$$

for constants $c_4, c_5, c_6 > 0$. The term $\overline{\partial}_t \theta(t_{j+1})$ is bounded because of (3.13). Summing up (3.15) for $i = 1, \ldots, j$, we infer

$$
\begin{aligned}
\|\overline{a}^j\|_{L^2(\Omega)} &\leq c_5 \Delta t \sum_{i=1}^{j} (1 + c_4 \Delta t)^{j-i} \|\overline{\theta}^i\|_{L^2(\Omega)} \\
&\quad + c_6 (\Delta t)^2 \sum_{i=0}^{j-1} (1 + c_4 \Delta t)^i = s_1 + s_2.
\end{aligned} \tag{3.16}
$$

Using the inequality $1 + x \leq e^x$, there exists a constant $c_7 > 0$ such that

$$
s_2 \leq c_6 (\Delta t)^2 \frac{e^{c_4 j \Delta t} - 1}{c_4 \Delta t} \leq c_7 \Delta t, \tag{3.17}
$$

and with the help of the Cauchy-Schwarz inequality

$$
\begin{aligned}
s_1 &\leq c_5 \Delta t \Big(\sum_{i=1}^{j} (1 + c_4 \Delta t)^{2(j-i)} \Big)^{1/2} \Big(\sum_{i=1}^{j} \|\overline{\theta}^i\|_{L^2(\Omega)}^2 \Big)^{1/2} \\
&\leq c_8 \Big(\Delta t \sum_{i=1}^{j} \|\overline{\theta}^i\|_{L^2(\Omega)}^2 \Big)^{1/2}
\end{aligned} \tag{3.18}
$$

for a constant $c_8 > 0$. Now we employ the decomposition $\overline{\theta}^j = \vartheta^j + \varrho^j$, where

$$
\vartheta^j = \theta_\ell^j - P^\ell \theta(t_j), \tag{3.19}
$$

together with (3.15)–(3.18) we conclude that there exists a constant $c_9 > 0$ such that

$$
\|\overline{a}^j\|_{L^2(\Omega)} \leq c_9 \Big(\Delta t + \Big(\Delta t \sum_{i=1}^{j} \|\vartheta^i\|_{L^2(\Omega)}^2 \Big)^{1/2} + \Big(\Delta t \sum_{i=1}^{j} \|\varrho^i\|_{L^2(\Omega)}^2 \Big)^{1/2} \Big). \tag{3.20}
$$

Regarding (3.19), (3.7b) and (2.1c), we see that ϑ^j satisfies

$$
\begin{aligned}
&\rho c_p (\overline{\partial}_t \vartheta^j, \psi)_{L^2(\Omega)} + k(\vartheta^j, \psi)_{H^1(\Omega)} \\
&= \rho c_p (\overline{\partial}_t \theta_\ell^j, \psi)_{L^2(\Omega)} + k(\theta_\ell^j, \psi)_{H^1(\Omega)} - \rho c_p (\overline{\partial}_t P^\ell \theta(t_j), \psi)_{L^2(\Omega)} - k(\theta(t_j), \psi)_{H^1(\Omega)} \\
&= (\rho c_p \zeta^j + k(\theta_\ell^j - \theta(t_j)) - \rho L (f(\theta_\ell^{j-1}, a_\ell^{j-1}) - f(\theta(t_j), a(t_j))), \psi)_{L^2(\Omega)},
\end{aligned}
$$

where we have used the abbreviation $\zeta^j = \theta_t(t_j) - \overline{\partial}_t P^\ell \theta(t_j)$ (cf. Lemma 3.6). Inserting $\psi = \vartheta^j$, using $\theta_\ell^j - \theta(t_j) = \vartheta^j + \varrho^j$, and invoking the inequalities of Hölder and Young, we observe

$$
\begin{aligned}
&\|\vartheta^j\|_{L^2(\Omega)}^2 + \frac{2k}{\rho c_p} \|\nabla \vartheta^j\|_{H^1(\Omega)}^2 \\
&\qquad \leq \|\vartheta^{j-1}\|_{L^2(\Omega)}^2 + c_{10} \Delta t \, \big(\|\vartheta^j\|_{L^2(\Omega)}^2 + \|\varrho^j\|_{L^2(\Omega)}^2 + \|\zeta^j\|_{L^2(\Omega)}^2 \big) \\
&\qquad\quad + c_{10} \Delta t \, \|f(\theta_\ell^{j-1}, a_\ell^{j-1}) - f(\theta(t_j), a(t_j))\|_{L^2(\Omega)}^2
\end{aligned} \tag{3.21}
$$

for a constant $c_{10} > 0$. With respect to (A1)–(A3) there exists a constant $c_{11} > 0$ such that

$$
\begin{aligned}
&\|f(\theta_\ell^{j-1}, a_\ell^{j-1}) - f(\theta(t_j), a(t_j))\|_{L^2(\Omega)}^2 \\
&\qquad \leq c_{11} \Big(\|\overline{\theta}^{j-1}\|_{L^2(\Omega)}^2 + (\Delta t)^2 \|\overline{\partial}_t \theta(t_j)\|_{L^2(\Omega)}^2 \\
&\qquad\quad + \|\overline{a}^{j-1}\|_{L^2(\Omega)}^2 + (\Delta t)^2 \|\overline{\partial}_t a(t_j)\|_{L^2(\Omega)}^2 \Big).
\end{aligned} \tag{3.22}
$$

Using (3.12), (3.13) and (3.20) we conclude from (3.22) that

$$\begin{aligned}&\|f(\theta_\ell^{j-1},a_\ell^{j-1})-f(\theta(t_j),a(t_j))\|_{L^2(\Omega)}^2\\&\quad\le c_{12}\Big(\|\vartheta^{j-1}\|_{L^2(\Omega)}^2+\|\varrho^{j-1}\|_{L^2(\Omega)}^2+(\Delta t)^2+\Delta t\sum_{i=1}^{j-1}\big(\|\vartheta^i\|_{L^2(\Omega)}^2+\|\varrho^i\|_{L^2(\Omega)}^2\big)\Big)\end{aligned}$$

for a constant $c_{12}>1$. Inserting this into (3.21) yields

$$\begin{aligned}&(1-c_{10}\Delta t)\|\vartheta^j\|_{L^2(\Omega)}^2\\&\le(1+c_{13}\Delta t)\|\vartheta^{j-1}\|_{L^2(\Omega)}^2+c_{13}\Delta t\big(\|\varrho^{j-1}\|_{L^2(\Omega)}^2+\|\varrho^j\|_{L^2(\Omega)}^2\big)\\&\quad+c_{13}(\Delta t)^2\sum_{i=1}^{j-1}\big(\|\vartheta^i\|_{L^2(\Omega)}^2+\|\varrho^i\|_{L^2(\Omega)}^2\big)+c_{13}(\Delta t)^3+c_{10}\Delta t\|\zeta^j\|_{L^2(\Omega)}^2,\end{aligned}\tag{3.23}$$

where $c_{13}=c_{10}c_{12}$. For $\Delta t\le 1/(2c_{10})$ we find

$$\frac{1}{1-c_{10}\Delta t}\le 1+2c_{10}\Delta t.$$

Setting $c_{14}=\max(2c_{10},c_{12})$ we infer from (3.23)

$$\begin{aligned}\|\vartheta^j\|_{L^2(\Omega)}^2&\le(1+c_{14}\Delta t)^2\|\vartheta^{j-1}\|_{L^2(\Omega)}^2\\&\quad+c_{13}(1+c_{14}\Delta t)\Delta t^2\sum_{i=1}^{j-1}\big(\|\vartheta^i\|_{L^2(\Omega)}^2+\|\varrho^i\|_{L^2(\Omega)}^2\big)\\&\quad+c_{12}\Delta t(1+c_{14}\Delta t)\big(\|\varrho^{j-1}\|_{L^2(\Omega)}^2+\|\varrho^j\|_{L^2(\Omega)}^2+(\Delta t)^2\big)\\&\quad+c_{10}\Delta t(1+c_{14}\Delta t)\|\zeta^j\|_{L^2(\Omega)}^2.\end{aligned}$$

Summation upon j yields that there exist constants $c_{15},c_{16}>0$ depending on c_{12},c_{13},c_{14} and T such that

$$\begin{aligned}\sum_{j=0}^n\|\vartheta^j\|_{L^2(\Omega)}^2&\le\|\vartheta^0\|_{L^2(\Omega)}^2+c_{15}\Delta t(1+c_{16}\Delta t)\sum_{j=0}^{n-1}\|\vartheta^j\|_{L^2(\Omega)}^2\\&\quad+c_{15}\Delta t(1+c_{16}\Delta t)\sum_{j=0}^n\big(\|\varrho^j\|_{L^2(\Omega)}^2+\|\zeta^j\|_{L^2(\Omega)}^2\big)\\&\quad+c_{15}(1+c_{16}\Delta t)(\Delta t)^2.\end{aligned}\tag{3.24}$$

Note that $\|\varphi-P^\ell\varphi\|_{H^1(\Omega)}\le\|\varphi-\psi\|_{H^1(\Omega)}$ for all $\psi\in V^\ell$. Moreover, the inverse inequality

$$\|\psi\|_{H^1(\Omega)}\le\|S\|_2\|\psi\|_{L^2(\Omega)}\qquad\text{for all }\psi\in\mathcal{V}$$

holds, see [15, Lemma 2]. From $\theta_\ell^0=\sum_{i=1}^\ell(\theta_0,\psi_i)_X\psi_i$ and $\theta_0\in\mathcal{V}$ we conclude that

$$\begin{aligned}\|\vartheta^0\|_{L^2(\Omega)}^2&=\|\theta_\ell^0-P^\ell\theta_0\|_{L^2(\Omega)}^2\le 2\|\theta_\ell^0-\theta_0\|_{L^2(\Omega)}^2+2\|\theta_0-P^\ell\theta_0\|_{L^2(\Omega)}^2\\&\le 2\sum_{i=\ell+1}^d|(\theta_0,\psi_i)_{L^2(\Omega)}|^2+2\|\theta_0-\theta_\ell^0\|_{H^1(\Omega)}^2\\&\le 2\big(1+\|S\|_2\big)\sum_{i=\ell+1}^d|(\theta_0,\psi_i)_{L^2(\Omega)}|^2\end{aligned}$$

in case of $X=L^2(\Omega)$ and

$$\|\vartheta^0\|_{L^2(\Omega)}^2\le 4\sum_{i=\ell+1}^d|(\theta_0,\psi_i)_{H^1(\Omega)}|^2$$

for $X = H^1(\Omega)$. Recall that $\|S\|_2 = 1$ in case of $X = H^1(\Omega)$. Thus, from (3.24) and Lemma 3.6 it follows that

$$\begin{aligned}\sum_{j=0}^{n} \|\vartheta^j\|^2_{L^2(\Omega)} \quad &\le \quad 2\big(1+\|S\|_2\big) \sum_{i=\ell+1}^{d} |(\theta_0,\psi_i)_X|^2 \\ &\quad + c_{17}(1+c_{16}\Delta t)\|S\|_2 \sum_{i=\ell+1}^{d} \lambda_i + c_{17}(\Delta t)^2 \\ &\quad + c_{15}\Delta t(1+c_{16}\Delta t) \sum_{j=0}^{n-1} \|\vartheta^j\|^2_{L^2(\Omega)}\end{aligned}$$

for a $c_{17} > 0$ depending on c_{15}, $\|\theta_{tt}\|_{L^2(Q)}$, and on the constant C, which was introduced in Lemma 3.6. Suppose that $\Delta t \le 1/c_{15}$. Then there exists a constant $c_{18} > 0$ such that

$$\xi_n \le (1 + c_{16}\Delta t)\xi_{n-1} + B,$$

where

$$B = c_{18}\Big(\|S\|_2 \sum_{i=\ell+1}^{d} \big(|(\theta_0,\psi_i)_X|^2 + \lambda_i\big) + (\Delta t)^2\Big) \quad \text{and} \quad \xi_i = \sum_{j=0}^{i} \|\vartheta^j\|^2_{L^2(\Omega)}.$$

We have already shown that $\xi_0 = \|\vartheta^0\|_{L^2(\Omega)} \le B$. Hence we can apply Lemma 3.7 with $\delta = c_{16}\Delta t$ and obtain

$$\sum_{j=0}^{n} \alpha_{j+1}\|\vartheta^j\|^2_{L^2(\Omega)} \le \Delta t \sum_{j=0}^{n} \|\vartheta^j\|^2_{L^2(\Omega)} \le \Delta t \frac{B}{c_{16}\Delta t}\big(e^{(n+1)T/n} - 1\big) \le c_{19}B, \tag{3.25}$$

where $c_{19} = (e^{2T}-1)/c_{16}$. In view of the decomposition $\overline{\theta}^j = \vartheta^j + \varrho^j$ and Lemma 3.6, we only have to insert (3.25) into (3.20) to conclude with the proof. □

3.3. Reduced order modeling with POD. The reduced-order approach to optimal control problems such as (CP^ε) is based on approximating the nonlinear dynamics by a Galerkin technique utilizing basis functions that contain characteristics of the expected flow. By Theorem 3.4 we have an error estimate for the state system (2.1) and (2.1c), but (3.8) only holds for a fixed and known laser energy $u(t)$. Unfortunately, the optimal control is unknown. To the authors' knowledge, there is no POD error analysis for optimal control problems available. Therefore we apply a heuristic, which is well tested for other optimal control problems, in particular for nonlinear boundary control of the heat equation, see [6].

To utilize the POD method described in Section 3.1 we need the snapshots. Since we have no chance to get the exact solution for a chosen laser energy at some given time instances, we compute a discrete solution to (2.1) on a fine grid. For that purpose we introduce piecewise linear finite elements $\{\varphi_1,\dots,\varphi_N\} \subset H^1(\Omega)$ and denote by $x_1,\dots,x_N \in \overline{\Omega}$ the finite element (FE) nodes such that $\varphi_j(x_i) = \delta_{ij}$ for $1 \le i,j \le N$. Analogous to (3.7) the FE solution solution $\{(\theta_N^j, a_N^j)\}_{j=0}^m$ to (2.1) is obtained by a semi-implicit FE Galerkin scheme: Find (θ_N^0, a_N^0) from

$$a_N^0 = 0 \quad \text{and} \quad (\theta_N^0,\varphi_i)_{L^2(\Omega)} = (\theta_0,\varphi_i)_{L^2(\Omega)} \quad \text{for } i = 1,\dots,N. \tag{3.26a}$$

Then, for $j = 1,\dots,m$ solve

$$\begin{aligned}&(\overline{\partial}_\tau \theta_N^j,\varphi_i)_{L^2(\Omega)} + (\nabla\theta_N^j,\nabla\varphi_i)_{L^2(\Omega)} \\ &\qquad = (u(t_j)\alpha(t_j) - \rho L f(\theta_N^{j-1}, a_N^{j-1}),\varphi_i)_{L^2(\Omega)} \quad \text{for } i = 1,\dots,N,\end{aligned} \tag{3.26b}$$

where $\overline{\partial}\theta_N^j = (\theta_N^j - \theta_N^{j-1})/\Delta s$, and use θ_N^j to compute a_N^j from

$$\overline{\partial}_\tau a_N^j(x_i) = f(\theta_N^j(x_i), a_N^{j-1}(x_i)) \quad \text{for } i = 1,\dots,N. \tag{3.26c}$$

To include information of the optimal control problem, which is under consideration here, we insert the computed sequences $\{\theta_N^j\}_{j=0}^m$ and $\{a_N^j\}_{j=0}^m$ into a semi-implicit FE Galerkin approximation of the adjoint system (2.21) (i.e., implicit in the heat operator $\partial_t - k\Delta$ and explicit in the part involving the derivatives f_θ and f_a) and determine piecewise linear approximations $\{(p_N^j, q_N^j)\}_{j=0}^m$ of the adjoint pair (p, q). The advantage of this approach is discussed in [6].

Now fix ℓ and determine the POD basis functions $\psi_1, \ldots, \psi_\ell$ by computing the matrix $\Psi \in \mathbb{R}^{N\times\ell}$ such that

$$\psi_j = \sum_{i=1}^{N} \Psi_{ij}\varphi_i \quad \text{for } j = 1, \ldots, \ell,$$

see Remark 3.3. For more details we refer the reader to [14]. We then approximate the state variable θ by a finite sum of time dependent modal coefficients multiplied by the POD basis elements:

$$\theta_\ell(\cdot, t) = \sum_{i=1}^{\ell} \theta_i^\ell(t)\psi_i.$$

For the volume fraction of austenite we do not apply a model reduction. However, its piecewise linear solution depends on ℓ due to the reduced-order approach for the temperature so that we write

$$a_\ell(\cdot, t) = \sum_{i=1}^{N} a_i^\ell(t)\varphi_i.$$

Let us introduce the mass and stiffness matrices

$$\begin{aligned}
M &= ((M_{ij})) \in \mathbb{R}^{\ell\times\ell} \quad \text{with} \quad M_{ij} = (\psi_j, \psi_i)_{L^2(\Omega)},\\
H &= ((H_{ij})) \in \mathbb{R}^{N\times N} \quad \text{with} \quad H_{ij} = (\varphi_j, \varphi_i)_{L^2(\Omega)},\\
K &= ((K_{ij})) \in \mathbb{R}^{\ell\times\ell} \quad \text{with} \quad K_{ij} = k\,(\nabla\psi_j, \nabla\psi_i)_{L^2(\Omega)},
\end{aligned}$$

the nonlinear mapping $F : \mathbb{R}^\ell \times \mathbb{R}^n \to \mathbb{R}^\ell$ given by

$$F(\vec{\theta}, \vec{a}) = \Big(f\Big(\sum_{j=1}^{\ell}\theta_j\psi_j, \sum_{j=1}^{N} a_j\varphi_j\Big), \psi_i\Big)_{L^2(\Omega)} \in \mathbb{R}^\ell$$

for $\vec{\theta} = (\theta_1, \ldots, \theta_\ell)$, $\vec{a} = (a_1, \ldots, a_N) \in \mathbb{R}^N$, the vectors of time dependent modal coefficients

$$\vec{\theta}(t) = \big(\theta_i^\ell(t)\big) \in \mathbb{R}^\ell, \quad \vec{a}(t) = \big(a_i^\ell(t)\big) \in \mathbb{R}^N,$$

and the vectors of the data

$$\vec{\theta}_0 = \Big((\theta_0, \psi_i)_{L^2(\Omega)}\Big) \in \mathbb{R}^\ell,\ \vec{\alpha}(t) = \Big((\alpha(\cdot, t), \psi_i)_{L^2(\Omega)}\Big) \in \mathbb{R}^\ell,\ a_d = (a_d(x_i)) \in \mathbb{R}^N.$$

Then the POD Galerkin approximation of the optimal control problem (CP^ε) is given by

$$\begin{aligned}
\min J_\ell(u) &= \frac{\sigma}{2}\,(\vec{a}(T) - \vec{a}_d)^\top H\,(\vec{a}(T) - \vec{a}_d) + \frac{\beta}{2}\int_0^T u(t)^2\,dt\\
&\quad + \frac{1}{2\varepsilon}\int_0^T \Big[\sum_{i=1}^{\ell}\theta_i^\ell(t)\psi_i - \theta_m\Big]_+^2\,dt
\end{aligned} \tag{3.27a}$$

subject to the $\ell + N$-dimensional system of ordinary differential equations

$$\left.\begin{aligned}
\rho c_p M\vec{\theta}'(t) + K\vec{\theta}(t) &= -\rho L F(\vec{\theta}(t), \vec{a}(t)) + u(t)\vec{\alpha}(t)\\
\vec{a}'(t) &= f\Big(\sum_{j=1}^{\ell}\theta_j^\ell\psi_j, \sum_{j=1}^{N} a_j^\ell\varphi_j\Big)
\end{aligned}\right\} \text{ for } t \in (0, T) \tag{3.27b}$$

	$\theta = 730$	$\theta = 830$	$\theta = 840$	$\theta = 900$
$a_{eq}(\theta)$	0	0.91	1	1
$\tau(\theta)$	1	0.2	0.18	0.05

TABLE 4.1. Pointwise data for a_{eq} and τ.

with the initial conditions at $t = 0$

$$M\vec{\theta}(0) = \vec{\theta}_0 \quad \text{and} \quad \vec{a}(0) = 0. \tag{3.27c}$$

Notice that in case of a FE Galerkin approximation the system of ordinary differential equations has dimension $2N$. Thus, (3.27) is called a low-dimensional model for the optimal control problem $(\mathsf{CP}^\varepsilon)$.

4. Numerical experiments

This section is devoted to present numerical results for the optimal control problem $(\mathsf{CP}^\varepsilon)$ utilizing the reduced-order approach described in Section 3.3.

Usually the aim of surface hardening is to achieve a uniform hardening depth. However, even in such a simple geometrical situation as shown in Figure 1 it is difficult to realize this goal. As it will be seen from the numerical simulations in Section 4.2, when one uses a constant laser energy, the temperature will be too low to reach the desired volume fraction in the beginning of the laser track, while it will be too high and possibly reach melting temperature at the end of the workpiece, since not enough heat can diffuse there. This means that one has to increase the energy during the early stages and to decrease it during the late stages of a laser heat treatment in order to achieve an approximately uniform hardness penetration depth. This will be shown in Section 4.3.

For the numerical implementation we use MATLAB version 5.3, executed on a Pentium III 550 MHz personal computer. For the finite element matrices the MATLAB PDE-toolbox is utilized.

4.1. Physical data. Let us choose the two-dimensional domain $\Omega = (0,5) \times (-1,0)$. This corresponds to the grey shaded vertical cross-section through the workpiece depicted in Figure 1. The physical parameters for the heat equations are given by

$$\rho c_p = 1.17 \left[\frac{cal}{cm^3 K}\right], \quad k = 0.153 \left[\frac{cal}{cmKs}\right], \quad \text{and} \quad \rho L = 150.0 \left[\frac{cal}{cm^3}\right].$$

For further details concerning physical data we refer to [8]. The equilibrium volume fraction a_{eq} and the function τ are cubic spline functions interpolating the pointwise data presented in Table 4.1. Thus, (A1) and (A2) are satisfied. For the monotone regularization of the Heaviside function we take

$$\mathcal{H}(s) = \begin{cases} 1 & \text{for } s \geq \delta, \\ 10\left(\frac{s}{\delta}\right)^6 - 24\left(\frac{s}{\delta}\right)^5 + 15\left(\frac{s}{\delta}\right)^4 & \text{for } \delta > s \geq 0, \\ 0 & \text{for } s < 0 \end{cases}$$

with $\delta = 0.15$. In particular, (A3) holds. The initial condition for the θ-variable is the room temperature, i.e., $\theta_0 = 20$, and we choose $\theta_m = 1400$ the melting temperature of steel. Notice that (A4') is satisfied.

We take a 2.8 kW laser, and the shape function $\alpha = \alpha(x, y, t)$ is given by

$$\alpha(x, y, t) = \frac{4\kappa A}{\pi D^2} \exp\Big(-\frac{2(x - vt)^2}{D^2}\Big) \exp(\kappa y),$$

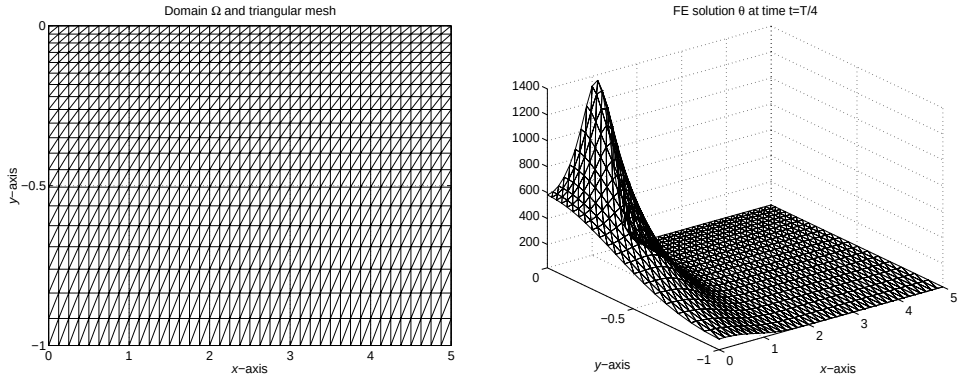


FIGURE 4.1. Triangular mesh and FE snapshot for the temperature at time $t = T/4$ with laser energy $u = 480$.

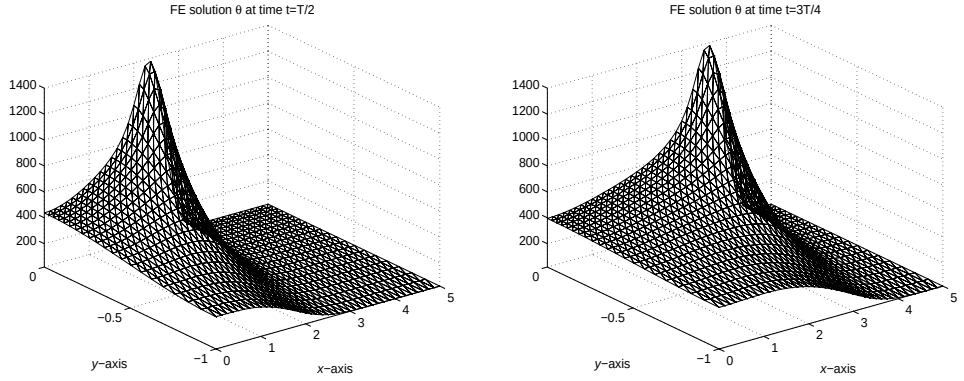


FIGURE 4.2. FE snapshot for the temperature at time instances $t = T/2$ and $t = 3T/4$ with laser energy $u = 480$.

where $D = 0.47\,[cm]$, $\kappa = 60\,[1/cm]$, $A = 0.3$, and $v = 1.15\,[cm/s]$. Notice that α satisfies (A5').

The admissible set U_{ad} for the controls is given by

$$U_{ad} = \{u \in L^2(0,T) : u_a \leq u \leq u_b \text{ in } (0,T)\}$$

with $u_a = 0$ and $u_b = 698$.

The terminal time T is chosen in such a way that $T = \text{argmax}\,\{\alpha(x,y,t) : (x,y) = (5,0),\ t \in [0,\infty)\}$. It follows that $T = 5/v \approx 4.3478$.

4.2. NUMERICAL SOLUTION OF THE STATE EQUATIONS. The FE triangulation of Ω is done by a nonuniform mesh with $N = 861$ degrees of freedom, see Figure 4.1 (left). For the time grids we take $n = 70$ and $l = 4$. Then we obtain $m = 280$, $\Delta\tau \approx 0.0155$ and $\Delta t = 4\Delta\tau \approx 0.0621$. Choosing the laser energy $u = 480 \in U_{ad}$ we compute the finite element (FE) solution of (3.26), where we use a Cholesky factorization for the linear system (3.26b) at each time level. The needed CPU time is 18 seconds. In Figure 4.1–4.3 the FE solution for the temperature (left) as well as for the volume fraction of austenite (right) at different time instances are plotted. From the FE snapshots of the temperature we can see the movement of the laser beam along the x-axis. We observe that the temperature θ_N^j increases at the end of the time interval. In particular, for $u = 480$, we find that the FE solution θ_N^j is lower that the melting temperature for $t_j \in (0, 4.1]$, but $\theta_N^m \approx 1625 > \theta_m$.

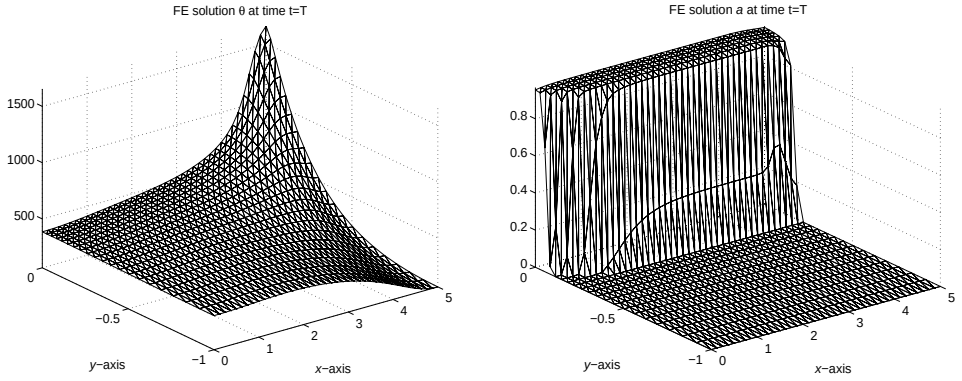


FIGURE 4.3. FE snapshot for the temperature and the volume fraction of austenite at time $t = T$ with laser energy $u = 480$.

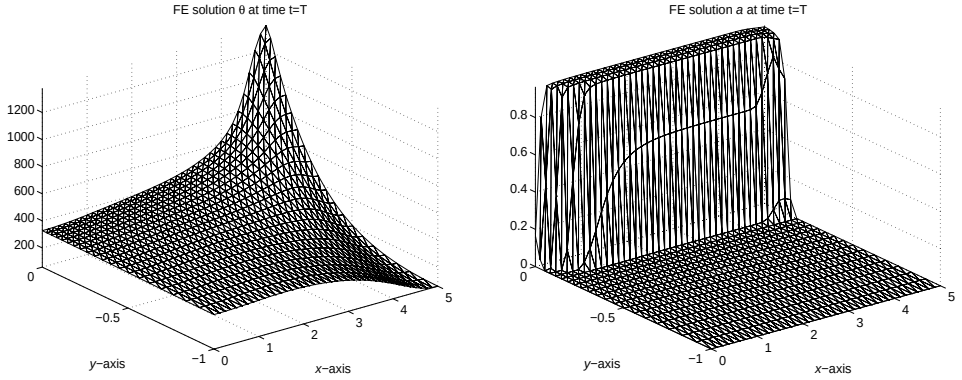


FIGURE 4.4. FE snapshot for the temperature and the volume fraction of austenite at time $t = T$ with laser energy $u = 400$.

4.3. REDUCED-ORDER MODELING. To determine the POD basis we proceed as described in Section 3.3. We compute the FE solution of the state equation with $u = 400$. Then the FE temperature is lower than the melting temperature. In Figure 4.4 the discrete solutions at the terminal time are presented. Next we solve the adjoint equation (2.21) by a semi-implicit FE Galerkin method. The needed CPU time is found to be 27 seconds. We compare two different snapshot sets. The first one (POD 1) is given by

$$\mathcal{V}^1 = \text{span}\,\{\theta_N^0, \theta_N^4, \ldots, \theta_N^{280}, \overline{\partial}_t\theta_N^4, \overline{\partial}_t\theta_N^8, \ldots, \overline{\partial}_t\theta_N^{280}, \lambda_N^0, \lambda_N^4, \ldots, \lambda_N^{276}\},$$

whereas the second one (POD 2) does not contains the difference quotients, i.e.,

$$\mathcal{V}^2 = \text{span}\,\{\theta_N^0, \theta_N^4, \ldots, \theta_N^{280}, \overline{\partial}_t\theta_N^4, \lambda_N^0, \lambda_N^4, \ldots, \lambda_N^{276}\}.$$

Choosing $X = L^2(\Omega)$ and $\ell = 15$, we solve the eigenvalue problem (3.6) for each of the snapshot ensemble by utilizing the MATLAB routine `eigs` and compute the reduced-order model described in Section 3.3. The needed CPU time is about 8 seconds.

Using scheme (3.7) we obtain the POD solution for the choice $u = 400$. The discrete solution for the temperature at the terminal time $t = T$ is plotted in Figure 4.5. When we compare the POD snapshot for the temperature at $t = T$ with the FE solution shown in Figure 4.4 (left), then it turns out that the inclusion of the difference quotients (POD 1) leads to a significantly better result, whereas the

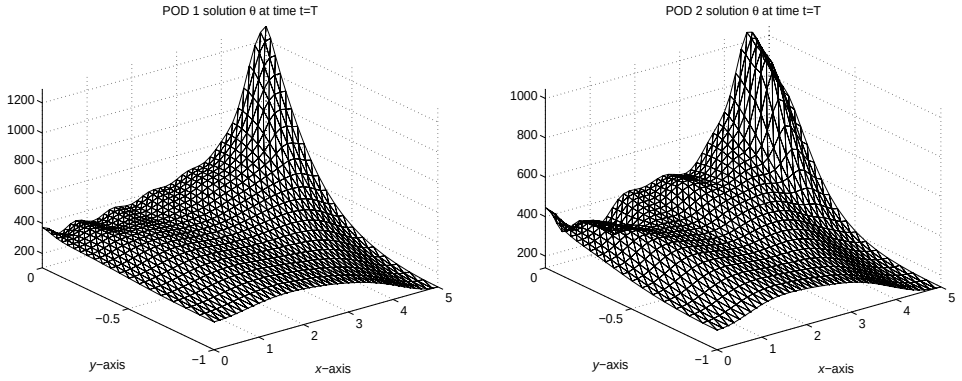


FIGURE 4.5. POD solution for $u = 400$ at time $t = T$ in case that the difference quotients are included in the snapshot ensemble (POD 1) or not (POD 2).

Computation of the FE matrices	0.2 seconds
FE solve for the state system (2.1)	17.8 seconds
FE solve for the adjoint system (2.21)	24.4 seconds
Computation of the POD basis (POD 1)	7.1 seconds
Computation of the POD basis (POD 2)	3.5 seconds
Computation of the POD matrices (POD 1)	2.2 seconds
Computation of the POD matrices (POD 2)	2.1 seconds
POD solve for the state system (2.1) (POD 1)	5.1 seconds
POD solve for the state system (2.1) (POD 2)	5.2 seconds

TABLE 4.2. CPU times in seconds for $X = L^2(\Omega)$ and $\ell = 15$.

snapshot ensemble $\mathcal{V}^2$ yields a POD solution with a smaller scale and a different shape. For the needed CPU times we refer to Table 4.2. To measure the error between the FE and the POD solutions let us introduce the relative quantities

$$\Psi^1_{L^\infty} = \frac{\max\limits_{0\le j\le m} \|\theta^j_\ell - \theta^j_N\|_{L^\infty(\Omega)}}{\max\limits_{0\le j\le m} \|\theta^j_N\|_{L^\infty(\Omega)}}$$

and

$$\Psi^1_{L^2} = \left(\frac{\sum\limits_{j=0}^{m} \|\theta^j_\ell - \theta^j_N\|^2_{L^2(\Omega)}}{\sum\limits_{j=0}^{m} \|\theta^j_N\|^2_{L^2(\Omega)}}\right)^{1/2}, \qquad \Psi^1_{H^1} = \left(\frac{\sum\limits_{j=0}^{m} \|\theta^j_\ell - \theta^j_N\|^2_{H^1(\Omega)}}{\sum\limits_{j=0}^{m} \|\theta^j_N\|^2_{H^1(\Omega)}}\right)^{1/2}$$

for the POD basis obtained from the snapshot set $\mathcal{V}^1$. Analogously, $\Psi^2_{L^\infty}$, $\Psi^2_{L^2}$ and $\Psi^2_{H^1}$ are defined for the snapshot set $\mathcal{V}^2$. The relative L^2-error is presented as a function of time in Figure 4.6. In Table 4.3 the relative error is presented for different values of ℓ. It turns out that the inclusion of the difference quotients reduces the error significantly. Next we discuss the choice $X = H^1(\Omega)$. As we have observed in the case $X = L^2(\Omega)$, the inclusion of the difference quotients into the snapshot sets leads to a significant reduction of the relative L^∞- and H^1-errors, compare Table 4.4. The same also holds for the relative L^2-error provided $\ell \ge 10$ is satisfied. Since H^1-norm includes both the L^2-norm and the gradient norm, the decay of the eigenvalues is not as fast as in the case $X = L^2(\Omega)$. Let us introduce

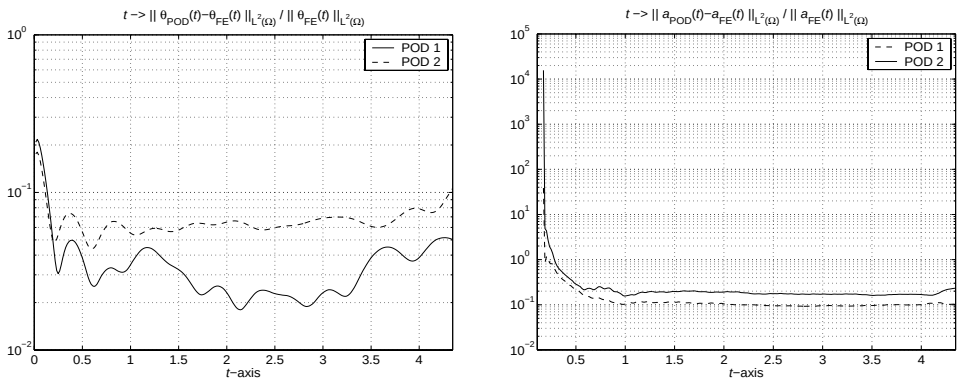


FIGURE 4.6. Relative L^2-error between the FE and the POD solution for the temperature and the volume fraction of austenite.

	$\Psi^1_{L^\infty}$	$\Psi^2_{L^\infty}$	$\Psi^1_{L^2}$	$\Psi^2_{L^2}$	$\Psi^1_{H^1}$	$\Psi^2_{H^1}$
$\ell = 10$	24.1%	40.6%	11.3%	12.1%	25.7%	38.2%
$\ell = 15$	7.9%	38.4%	3.4%	6.8%	10.3%	27.4%
$\ell = 20$	6.0%	35.3%	1.8%	4.3%	5.5%	20.6%
$\ell = 25$	1.6%	26.9%	0.6%	2.9%	2.8%	10.4%

TABLE 4.3. Relative errors for $X = L^2(\Omega)$.

	$\Psi^1_{L^\infty}$	$\Psi^2_{L^\infty}$	$\Psi^1_{L^2}$	$\Psi^2_{L^2}$	$\Psi^1_{H^1}$	$\Psi^2_{H^1}$
$\ell = 10$	21.0%	40.1%	22.9%	11.8%	28.8%	37.5%
$\ell = 15$	16.2%	37.8%	4.4%	6.7%	13.1%	27.0%
$\ell = 20$	13.5%	34.3%	2.2%	4.3%	8.1%	20.5%
$\ell = 25$	4.0%	24.6%	1.2%	2.9%	4.6%	14.8%

TABLE 4.4. Relative errors for $X = H^1(\Omega)$.

	$\ell = 5$	$\ell = 10$	$\ell = 15$	$\ell = 20$	$\ell = 25$	$\ell = 30$
$\mathcal{E}(\ell)$, $X = L^2(\Omega)$	79.6	94.3	98.4	99.5	99.8	99.9
$\mathcal{E}(\ell)$, $X = H^1(\Omega)$	53.0	77.7	87.4	92.5	95.7	97.6

TABLE 4.5. $\mathcal{E}(\ell)$ for POD 1 and different ℓ.

the relative quantity

$$\mathcal{E}(\ell) = \sum_{i=1}^{\ell} \lambda_i \Big/ \sum_{i=1}^{d} \lambda_i.$$

Then we find the results presented in Table 4.5. In (3.8) the factor $\|S\|_2 \sum_{i=\ell+1}^{d} \lambda_i$ arrises on the right-hand side of the error estimate. For the choice $X = H^1(\Omega)$ we have $\|S\|_2 = 1$. In Table 4.6 the norm of the stiffness matrix S is presented for $X = H^1(\Omega)$ and for different ℓ. From Tables 4.5–4.6 we conclude that the advantage of $\|S\|_2 = 1$ for $X = H^1(\Omega)$ is balanced by the disadvantage that for given ℓ the sum $\sum_{i=\ell+1}^{d} \lambda_i$ is larger than for the choice $X = L^2(\Omega)$. However, when we choose ℓ in such a way that $\mathcal{E}(\ell)$ is lower than a given threshold, then the relative errors

	$\ell = 5$	$\ell = 10$	$\ell = 15$	$\ell = 20$	$\ell = 25$	$\ell = 30$
$\|S\|_2$	13.9	53.2	144.2	257.5	629.9	940.9

TABLE 4.6. Spectral norm of the stiffness matrix for $X = L^2(\Omega)$.

	$X = L^2(\Omega)$				$X = H^1(\Omega)$			
	ℓ	$\Psi^1_{L^\infty}$	$\Psi^1_{L^2}$	$\Psi^1_{H^1}$	ℓ	$\Psi^1_{L^\infty}$	$\Psi^1_{L^2}$	$\Psi^1_{H^1}$
$\mathcal{E}(\ell) = 84.4\%$	6	26.7%	23.6%	42.8%	13	21.2%	6.0%	17.0%
$\mathcal{E}(\ell) = 92.5\%$	9	29.7%	14.5%	30.0%	20	13.5%	8.1%	3.3%
$\mathcal{E}(\ell) = 97.5\%$	13	14.7%	4.8%	14.5%	30	2.7%	1.0%	3.3%

TABLE 4.7. Relative errors for $X = L^2(\Omega)$ and $X = H^1(\Omega)$.

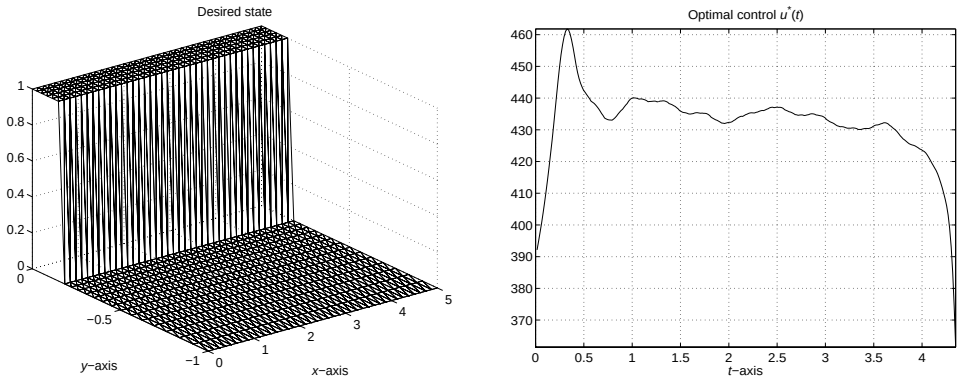


FIGURE 4.7. Run 4.1: Desired state a_d and optimal control u^*.

for $X = H^1(\Omega)$ are smaller than for $X = L^2(\Omega)$, see Table 4.7. Let us mention that in practice the number ℓ is often chosen in this manner.

4.4. NUMERICAL TESTS FOR THE OPTIMAL CONTROL PROBLEM. In the following we present two test runs for the optimal control problem (CP^ε), which is solved by the gradient projection algorithm, see e.g. [12].

Run 4.1. We choose $\sigma = 3500$, $\beta = 0.001$, $\varepsilon = 0.0001$. The desired volume fraction of austenite is shown in Figure 4.7 (left). As the first iterate for the control we take $u^0 = 380$. Let us denote by a^0 the FE solution for the volume fraction of austenite corresponding to the laser intensity $u = u^0$. The gradient projection algorithm stops after 50 iterations and needs 1091 seconds CPU time. The optimal control u^* is presented in Figure 4.7 (right). In Figure 4.8 (left) the discrete POD solution for the temperature at the time instance $t = T$ is plotted. Inserting the computed suboptimal control into the finite element solver of the state equations we compute the solution denoted by $(\theta^*_\ell, a^*_\ell)$. In Figure 4.8 (right) the solution θ^*_ℓ at time $t = T$ is presented. As we can see FE solve cancels out the small oscillations occurring in the POD solution. We observe that $\|\theta^*_\ell\|_{L^\infty(\Omega)} = 1391.15$. In Figure 4.9 the differences $a^*_\ell(T) - a_d$ and $a^0(T) - a_d$ are plotted. Using the optimal control we get a significant reduction of the residuum. We find that $\|a^0(T) - a_d\|_{L^2(\Omega)} = 0.285$ and $\|a^*_\ell(T) - a_d\|_{L^2(\Omega)} = 0.085$.

Run 4.2. We choose $\sigma = 17500$, $\beta = 0.001$, $\varepsilon = 0.0001$. In contrast to Run4.1 it is possible to enlarge σ significantly without any bad influence on gradient projection method. The desired volume fraction of austenite is shown in Figure 4.10 (left).

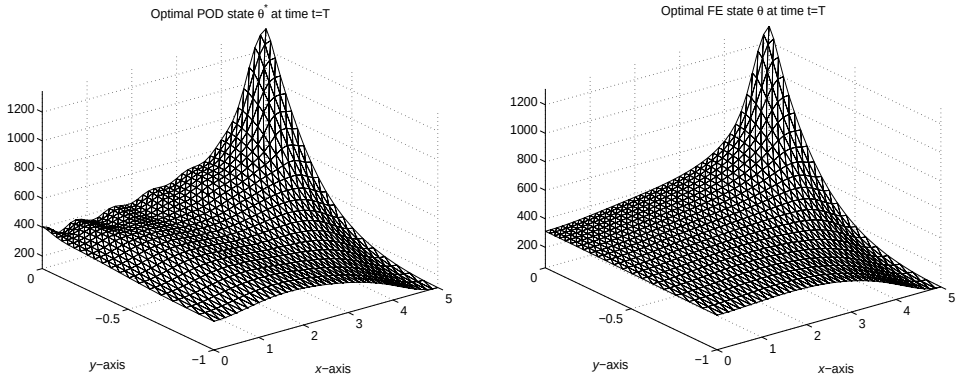


FIGURE 4.8. Run 4.1: Optimal POD snapshots for the temperature at time $t = T$ (left) and FE snapshot for the temperature at time $t = T$ using $u = u^*$ (right).

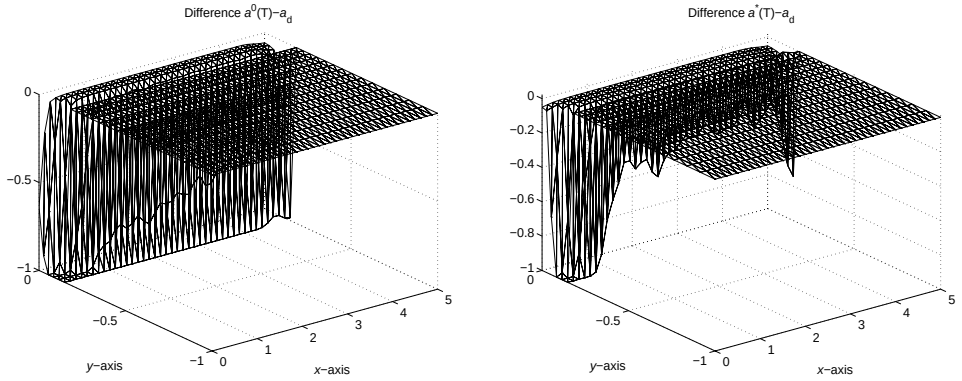


FIGURE 4.9. Run 4.1: Difference $a(T) - a_d$ for the first iterate $a = a^0$ of the gradient projection method and for the optimal state $a = a_l^*$.

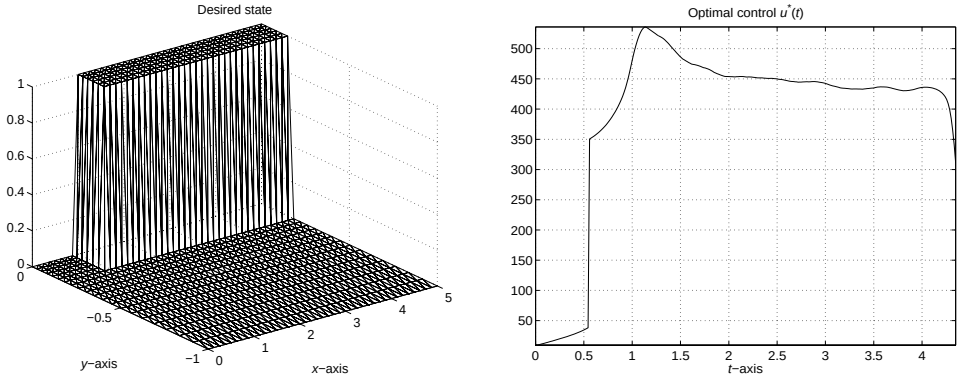


FIGURE 4.10. Run 4.2: Desired state a_d and optimal control u^*.

Due to the desired state we take

$$u^0(t_j) = \begin{cases} 0 & \text{for } 0 \le j \le 40, \\ 330 & \text{otherwise} \end{cases}$$

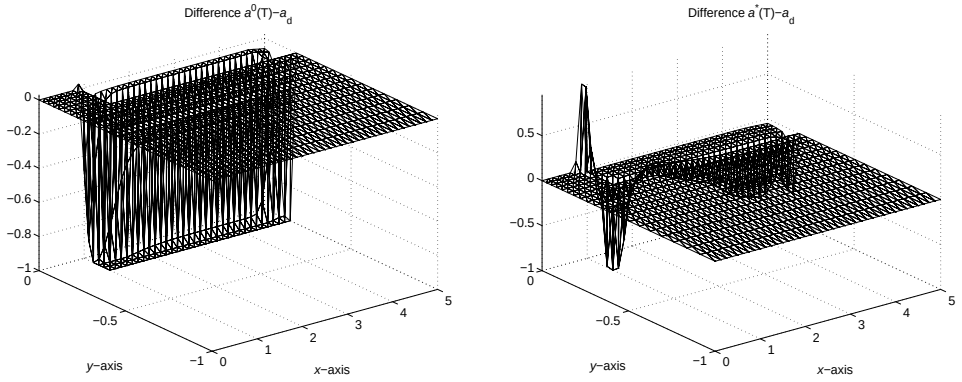


FIGURE 4.11. Run 4.2: Difference $a(T) - a_d$ for the first iterate $a = a^0$ of the gradient projection method and for the optimal state $a = a^*_\ell$.

as the first iterate of the gradient projection method. The method needs 29 iterations and 666 seconds CPU time. By a^0 we denote the FE solution corresponding to the laser intensity $u = u^0$. We insert the computed suboptimal control into the finite element solver of the state equations and denote its solution by $(\theta^*_\ell, a^*_\ell)$. We observe that $\|\theta^*_\ell\|_{L^\infty(\Omega)} = 1398$. In Figure 4.11 the differences $a^*_\ell(T) - a_d$ and $a^0(T) - a_d$ are plotted. Using the optimal control we get a significant reduction of the residuum. We find that $\|a^0(T) - a_d\|_{L^2(\Omega)} = 0.423$ and $\|a^*_\ell(T) - a_d\|_{L^2(\Omega)} = 0.029$.

REFERENCES

[1] K. Afanasiev and M.Hinze. Adaptive control of a wake flow using proper orthogonal decomposition. In *Shape Optimization & Optimal Design*, Lecture Notes in Pure and Applied Mathematics. Marcel Dekker, 2001.

[2] J. A. Atwell and B. B. King. Reduced order controllers for spatially distributed systems via proper orthogonal decomposition. *SIAM Journal Scientific Computation*, to appear.

[3] N. Aubry, W.-Y. Lian and E. S. Titi. Preserving symmetries in the proper orthogonal decomposition. *SIAM J. Sci. Comp.*, 14(1993), 483–505.

[4] H. T. Banks, M. L. Joyner, B. Winchesky, and W. P. Winfree. Nondestructive evaluation using a reduced-order computational methodology. *Inverse Problems* 16(2000), 1–17.

[5] G. Berkooz, P. Holmes, and J. L. Lumley. *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*. Cambridge Monographs on Mechanics. Cambridge University Press, 1996.

[6] F. Diwoky and S. Volkwein. Nonlinear boundary control for the heat equation utilizing proper orthogonal decomposition. Proceedings of the workshop *Fast Solution of Discretized Optimization Problems*, WIAS, Berlin, 2000.

[7] M. Fahl. Computation of POD basis functions for fluid flows with Lanczos methods. *Mathematical and Computer Modelling*, to appear.

[8] J. Fuhrmann and D. Hömberg. Numerical simulation of surface heat treatments. *Num. Meth. Heat & Fluid Flow*, 9 (1999), 705–724.

[9] K. Fukunaga. *Introduction to Statistical Recognition*. Academic Press, New York, 1990.

[10] D. Hömberg. A mathematical model for the phase transitions in eutectoid carbon steel. *IMA J. Appl. Math.*, 54 (1995), 31–57.

[11] D. Hömberg and J. Sokolowski. Optimal control of laser hardening. *Adv. Math. Sci. Appl.*, 8 (1998), 911–928.

[12] C. T. Kelley. *Iterative Methods for Optimization*. Frontiers in Applied Mathematics, SIAM, Philadelphia, 1999.

[13] G. M. Kepler, H. T. Tran, and H. T. Banks. Compensator control for chemical vapor deposition film growth used reduced order design models. Preprint, North Carolina State University, CRSC-TR99-41.

[14] K. Kunisch and S. Volkwein. Control of Burgers' equation by a reduced order approach using proper orthogonal decomposition. *JOTA*, 102 (1999), 345–371.

[15] K. Kunisch and S. Volkwein. Galerkin proper orthogonal decomposition methods for parabolic problems. *Numerische Mathematik*, to appear.
[16] K. Kunisch and S. Volkwein. Galerkin proper orthogonal decomposition methods for a general equation in fluid dynamics. Submitted, 2000.
[17] J.-B. Leblond and J. Devaux. A new kinetic model for anisothermal metallurgical transformations in steels including effect of austenite grain size. *Acta Met.* 32 (1984), 137–146.
[18] O. A. Ladyzenskaja, V. A. Solonnikov, and N. N. Ural'ceva. *Linear and quasilinear equations of parabolic type*. Amer. Math. Soc. Transl., Vol. 23, Providence, 1968.
[19] M. Manhart. Umströmung einer Halbkugel in turbulenter Grenzschicht, VDI-Verlag, Düsseldorf, 1996.
[20] V. I. Mazhukin and A. A. Samarskii. Mathematical modeling in the technology of laser treatments of materials. *Surveys Math. Indust.*, 4 (1994), 85–149.
[21] H. V. Ly and H. T. Tran. Proper orthogonal decomposition for flow calculations and optimal control in a horizontal CVD reactor. *Quarterly of Applied Mathematics*, to appear.
[22] S. Y. Shvartsman and Y. Kevrikidis. Nonlinear model reduction for control of distributed parameter systems: a computer-assisted study. *AIChE Journal*, 44 (1998), 1579–1595.
[23] L. Sirovich. Turbulence and the dynamics of coherent structures, parts I-III. *Quarterly of Applied Mathematics, XLV*, 561–590, 1987.
[24] M. Reed and B. Simon. *Methods of Modern Mathematical Physics I: Functional Analysis.* Academic Press, New York, 1980.
[25] S. Volkwein. Optimal control of a phase-field model using the proper ortogonal decomposition. *Zeitschrift für Angewandte Mathematik und Mechanik*, 81:83–97, 2001.
[26] E. Zeidler. *Nonlinear Functional Analysis and its Applications, Vol. II.* Springer-Verlag, New York, 1990.

D. Hömberg, Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstrasse 39, D-10117 Berlin, Germany
E-mail address: hoemberg@wias-berlin.de

S. Volkwein, Institut für Mathematik, Karl-Franzens-Universität Graz, Heinrichstrasse 36, A-8010 Graz, Austria
E-mail address: stefan.volkwein@uni-graz.at