Optimal Control of Radiative Heat Transfer in Glass Cooling with Restrictions on the Temperature Gradient

Debora Clever, Jens Lang

TECHNISCHE UNIVERSITÄT DARMSTADT

Fachbereich Mathematik,
Technische Universität Darmstadt,
Dolivostraße 15,
64293 Darmstadt, Germany
clever@mathematik.tu-darmstadt.de,
lang@mathematik.tu-darmstadt.de

Abstract

This paper is motivated by an optimal boundary control problem for the cooling process of molten and already formed glass down to room temperature, which is an important step in glass manufacturing. The high temperatures at which glass is processed demand to include radiative heat transfer in the computational model. Since the computation of the complete radiative heat transfer equations is too complex for optimization purposes, we use simplified approximations of spherical harmonics coupled with a practically relevant frequency bands model. The optimal control problem is considered as a PDAE-constrained optimization problem with box constraints on the control. In this paper we want to augment the objective by a functional depending on the state gradient which forces a minimization of thermal stress inside the glass. To guarantee consistent and grid-independent values of the reduced objective gradient at the end of the cooling process we pursue two approaches. The first includes the temperature gradient with a time-dependent linearly decreasing weight. In the second approach we augment the objective functional by final state tracking and final state gradient optimization, which to our knowledge has never been considered before. To determine an optimal boundary control we apply a projected gradient method where proper step sizes are estimated by Armijo's rule, considering the first Wolfe condition. The reduced objective gradient is computed by the continuous adjoint approach. The arising time-dependent partial differential algebraic equations are numerically solved by variable step-size one-step methods of Rosenbrock type in time and adaptive multilevel finite elements in space. We present two-dimensional numerical results for an infinitely long glass block. We vary certain weights in the objective and compare the two different approaches derived to ensure consistency especially at the end of the cooling process. Numerical results are presented and discussed.

1 Introduction

One important step in glass manufacturing is the cooling of molten and already formed glass down to room temperature. To control this process, the hot glass is put into a preheated furnace. While the temperature inside the oven is carefully decreased towards room temperature, the glass cools down as well. To choose an appropriate furnace temperature course, three contrary criteria have to be taken into account. First, we would like to track the glass temperature distribution over space and time with respect to a desired temperature profile for which good performance of the involved chemical processes is known. Second, we have to minimize large temperature gradients within the glass because they cause great internal stresses and finally can effect cracks inside the material [2]. This phenomenon occurs if the furnace temperature is decreased very fast because at the boundary the glass reacts almost immediately to the outer temperature whereas in the core it stays hot for a long time. Nevertheless, the cooling should not last longer than necessary due to energy and other manufacturing costs.

Because of the high temperatures at which glass is processed, radiative heat transfer plays a dominant role, leading to a model that is given by a high-dimensional and highly non-linear system of time-dependent partial differential algebraic equations for the glass temperature and the radiative intensity in dependence on the furnace temperature. Due to the high complexity of the established equations [15] we use a dimensionless form of the SP_1-approximation including a practically relevant frequency bands model. This simplified model has proven to perform accurate enough for various radiative heat transfer problems [9, 14] and is therefore an appropriate choice for optimization purposes where it has to be solved several times. The optimal control problem is considered as a constrained optimization problem [8] and a first-order optimality system is derived using a Lagrangian formalism. To describe the quality of the boundary control we will present various objective components including a functional depending on the state gradient, a final state tracking, and a final state gradient optimization, scaled with constant or time-dependent weightings. We will present numerical results showing the indispensability of the above mentioned final state and state gradient optimization in the case of non-vanishing terminal weights. First investigations with temperature gradients but without terminal optimization have been made in Ref. [17] for a simpler one-dimensional gray scale model neglecting the dependence on the frequency.

To determine an optimal boundary control we apply a projected gradient method with Armijo step size control, considering the first Wolfe condition. To compute the gradient of the reduced objective functional we follow the continuous adjoint approach. State and adjoint systems are solved by using the software package KARDOS [6] which uses linearly implicit one-step methods of Rosenbrock type with an error-controlled step size selection to integrate in time and mesh improving multilevel finite elements for the discretization in space [5, 10]. Because of the adaptivity in time and space we achieve substantial savings in computer time and memory which often means the difference between getting an answer or not to the optimal control problem considered.

This paper is organized as follows. We first formulate the glass cooling model and set up our optimal boundary control problem. We will discuss general problems occurring in the last time step which become very important when including the temperature gradient within the objective functional. We will present two different approaches to overcome this drawback and define the resulting objective functionals. Then we derive the first-order optimality system using a Lagrangian formalism and describe the projected gradient method which is used to solve it. Numerical results are presented for both approaches and different weightings. The optimal controls obtained and the allocation of the costs to the different objective components are compared and conclusions are made.

We want to point out that even though the algorithm is presented for the special case of glass cooling it is not especially constructed for this very problem. Therefore it can easily be transferred to other application, which are described by comparable systems of time-dependent partial differential algebraic equations.

2 The Glass Cooling Model

One important observation in glass cooling modelling is that taking into account only heat conduction inside and convective heat transfer at the boundary of the glass is not appropriate. Because of the high temperatures at which glass is processed the direction- and frequency-dependent thermal radiation field and the spectral radiative properties of semi-transparent glass play a dominant role for the temperature distribution inside the material as well. We consider a dimensionless model which consists of a heat equation for the scaled temperature $T(x,t)$ and a transport equation for the scaled radiation intensity $I(x,t,\nu,s)$ [9], namely

$$\epsilon^2 \partial_t T - \epsilon^2 \nabla \cdot \left(k_c \nabla T\right) = -\int_{\nu_0}^{\infty} \int_{\mathbb{S}^2} \kappa_\nu \left(B\left(T,\nu\right) - I(x,t,\nu,s)\right) \mathrm{d}s \mathrm{d}\nu, \tag{1}$$

$$\epsilon s \cdot \nabla I(x,t,\nu,s) + \left(\sigma_\nu + \kappa_\nu\right) I(x,t,\nu,s) = \frac{\sigma_\nu}{4\pi} \int_{\mathbb{S}^2} I(x,t,\nu,s) \mathrm{d}s + \kappa_\nu B\left(T,\nu\right), \quad \forall \nu > \nu_0, \tag{2}$$

with the boundary and initial conditions

$$\epsilon k_c n \cdot \nabla T = h_c \left(u - T\right) + \alpha\pi \left(\frac{n_a}{n_g}\right)^2 \int_0^{\nu_0} \left(B\left(u,\nu\right) - B\left(T,\nu\right)\right) \mathrm{d}\nu, \tag{3}$$

$$I(x,t,\nu,s) = r\left(n \cdot s\right) I(x,t,\nu,s') + \left(1 - r\left(n \cdot s\right)\right) B\left(u,\nu\right), \ (x,s) \in \Gamma^-, \tag{4}$$

$$T\left(x,0\right) = T_0\left(x\right). \tag{5}$$

Equation (1) describes the evolution of the glass temperature. Temperature changes can be caused by thermal flows described through a dimensionless heat conduction coefficient k_c and through energy exchange between the glass and the radiation field because of absorption and emission. This highly non-linear exchange of energy is modelled by using the frequency-dependent dimensionless absorption coefficient κ_ν and Planck's function

$$B\left(T,\nu\right) = \frac{n_g^2}{c_0^2} \frac{2h_p \nu^3}{e^{h_p \nu/(k_b T)} - 1}, \tag{6}$$

for black body radiation in glass, with Planck constant h_b, Boltzmann constant k_b and the speed of light in vacuum c_0. The refractive index n_g is the ratio of the speed of light in vacuum and in glass. Radiation will be described by its intensity $I(x,t,\nu,s)$ which depends on the spatial variable $x \in \Omega \subset \mathbb{R}^d$, time $t \in [0,t_e)$, frequency $\nu \in [0,\infty)$, and direction $s \in \mathbb{S}^2$ of the unit sphere. The parameter ϵ is introduced in Ref. [9] to eliminate the dependence on units and satisfies $0 < \epsilon \leq 1$ in the optically thick, diffusive regime we are interested in. The boundary conditions (3) reflect free thermal convection and diffusive surface radiation involving the dimensionless convective heat transfer coefficient h_c, the furnace temperature u, the index of refraction of air n_a, and the mean hemispheric surface emissivity α in the opaque spectral region $[0,\nu_0]$, where glass strongly absorbs radiation.

Equation (2) describes the intensity of a radiation beam through x at time t with direction s and frequency ν, which is influenced by absorption, emission, and scattering. Here, σ_ν is the dimensionless scattering coefficient. On the boundary (see Eq. (4)) radiation can either enter into the glass or is reflected. The amount of reflected energy is given by the reflectivity $r \in [0,1]$ and can be found using Fresnel's equation [15]. The intensity of an interior beam with direction s gets increased by the intensity of the exterior radiation coming from the direction of its specular reflection $s' = s - 2(n \cdot s)n$ with the outwards normal $n(x)$. The set on which those transparent boundary conditions are defined is given

by $\Gamma^- := \{(x,s) \in \partial\Omega \times \mathbb{S}^2 : n(x)\cdot s < 0\}$. Finally, the initial temperature of the glass is specified by condition (5).

The high dimension of the phase space makes the numerical solution of the full radiative heat transfer equation very expensive, especially for optimization purposes, where the system has to be solved several times. Various approximate models that are less time consuming, yet sufficiently accurate, have been developed [9, 14, 11]. In this paper, we use a first order approximation of spherical harmonics including a practically relevant frequency bands model with N bands. This SP_1-approximation has been tested fairly extensively for various radiation transfer problems in glass and has proven to be an efficient way to improve the classical diffusion approximations [9, 14]. It results in the following system of partial differential algebraic equations of mixed parabolic-elliptic type in $N+1$ components:

$$\partial_t T - \nabla\cdot\left(k_c \nabla T\right) = \sum_{i=1}^{N} \nabla\cdot\left(\frac{1}{3\left(\sigma_i+\kappa_i\right)}\nabla\phi_i\right), \tag{7}$$

$$-\epsilon^2\nabla\cdot\left(\frac{1}{3\left(\sigma_i+\kappa_i\right)}\nabla\phi_i\right) + \kappa_i\phi_i = 4\pi\kappa_i B^{(i)}(T),\ \ i=1,\dots,N, \tag{8}$$

with boundary and initial conditions

$$\begin{aligned} k_c n\cdot\nabla T + \sum_{i=1}^{N}\frac{1}{3\left(\sigma_i+\kappa_i\right)} n\cdot\nabla\phi_i = \frac{h_c}{\epsilon}(u-T) + \\ + \frac{\alpha\pi}{\epsilon}\left(\frac{n_a}{n_g}\right)^2\left(B^{(0)}(u) - B^{(0)}(T)\right) + \frac{c_1}{\epsilon}\sum_{i=1}^{N}\left(4\pi B^{(i)}(u) - \phi_i\right), \end{aligned} \tag{9}$$

$$\frac{\epsilon^2}{3\left(\sigma_i+\kappa_i\right)} n\cdot\nabla\phi_i = c_1\epsilon\left(4\pi B^{(i)}(u) - \phi_i\right),\ \ i=1,\dots,N, \tag{10}$$

$$T(x,0) = T_0(x). \tag{11}$$

Here, the continuous frequency spectrum is split up into N bands $[\nu_{i-1},\nu_i]$, $i=0,\dots,N$, with $\nu_N=\infty$ and $\nu_{-1}:=0$. On each of this bands we consider a mean intensity

$$\phi_i(x,t) := \int_{\nu_{i-1}}^{\nu_i}\int_{\mathbb{S}^2} I(x,t,\nu,s)\mathrm{d}s\mathrm{d}\nu,\ i=1,\dots,N, \tag{12}$$

and a frequency-independent mean of the Planck function

$$B^{(i)}(v) := \int_{\nu_{i-1}}^{\nu_i} B(v,\nu)\mathrm{d}\nu,\ i=1,\dots,N. \tag{13}$$

Furthermore, all frequency-dependent coefficients are set constant on each of the bands,

$$\kappa_\nu = \kappa_i,\quad \sigma_\nu = \sigma_i,\quad \text{for } \nu\in[\nu_{i-1},\nu_i],\ i=1,\dots,N. \tag{14}$$

Division by proper reference quantities brings the system to the above presented dimensionless form.

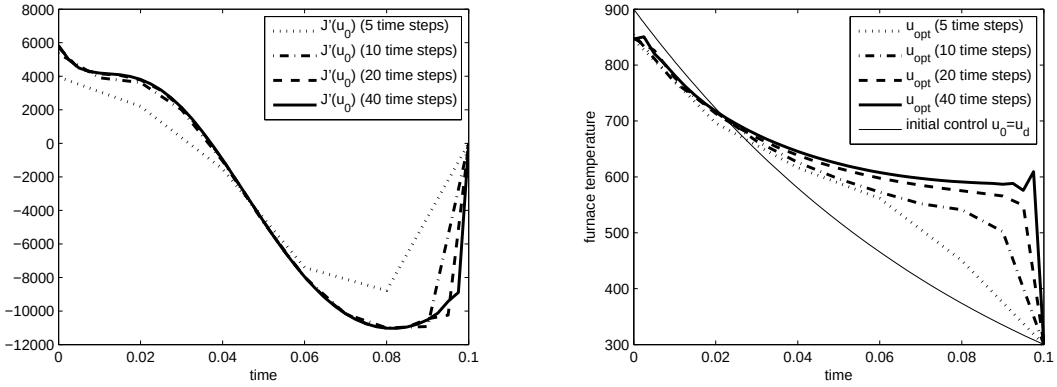
3 The Objective Functional

In the optimization of glass cooling processes, it is essential for the quality of the glass that its temperature distribution follows a desired profile to control the chemical reactions in the glass. Typically, there are time-dependent reference values provided by engineers for the furnace temperature as well. Especially at the beginning of the cooling process a smooth control which is close to the initial glass temperature is essential. Equally important is to minimize temperature gradients and therefore large internal stresses which could induce cracks inside the glass. Particularly, in the case of high quality glass, this phenomenon can not be neglected. An objective functional which seems to meet all requirements can be defined by

$$\tilde{J}(T,u) := \frac{1}{2}\int_0^{t_e} \left\| T - T_d \right\|^2_{L^2(\Omega)} \mathrm{d}t + \frac{\delta_g}{2}\int_0^{t_e} \|\nabla T\|^2_{L^2(\Omega)} \mathrm{d}t + \frac{1}{2}\int_0^{t_e} \delta_u(t)(u-u_d)^2 \mathrm{d}t, \tag{15}$$

with the desired glass temperature distribution $T_d(x,t)$ and a guideline $u_d(t)$ for the control. The final time of the cooling process is denoted by t_e. The positive weights δ_g and $\delta_u(t)$ are used to steer the influence of the corresponding components. A smooth control close to the desired value, especially at the beginning of the cooling process, can be realized by choosing a decreasing weight $\delta_u(t)$ with sufficiently large value at $t = 0$. Since the furnace only operates in a certain temperature range we restrict the control $u(t)$ to the set of admissible controls $U_{ad} := \{u \in L^2(0,t_e) : \underline{u} \le u \le \overline{u}\}$.

If we now derive the adjoint system and evaluate the reduced gradient to apply a gradient method we can see that we end up with the terminal condition $\xi(t_e) = 0$ for the vector of adjoint variables. The gradient itself is evaluated by a summand depending on the control and one depending on the adjoint variables. A more detailed description of the analytical derivation of the first order optimality system will be given in Section 4. In the case $\xi(t_e) = 0$, the part of the reduced gradient depending on the adjoint variables vanishes and hence the optimization in the last time step is influenced by the difference between control $u(t_e)$ and guiding value $u_d(t_e)$ only. This means in the last time step it is impossible to find other values for the optimal control than the guiding value, which in general is equal to the initial guess. Figure 1 clarifies this phenomenon. Here, we compute an optimal control with respect to (15)



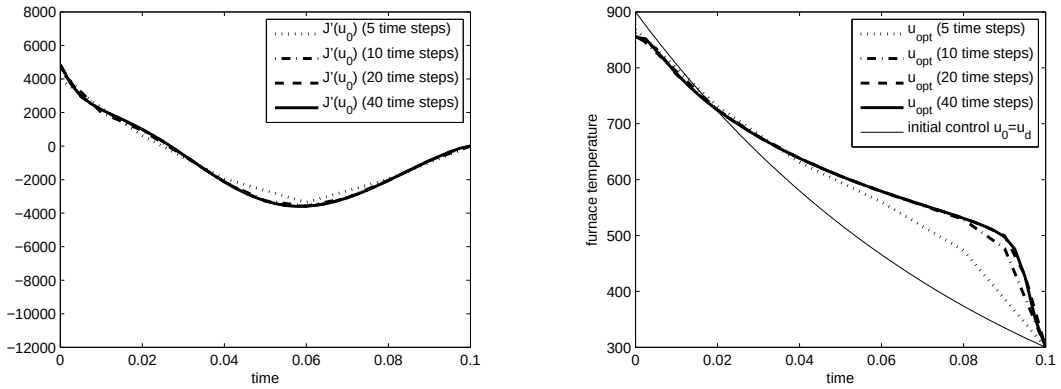
(a) Reduced Gradient (b) Optimal Control

Figure 1: Temperature gradient minimization without decreasing weighting or final value optimization: objective gradients and optimal controls on different time grids. The objective gradients (exemplarily for $\tilde{J}'(u_0)$ (left)) show an unreasonable ascent in the final time step. As a consequence the optimal control (right) has a steep final descent whose slope depends on the final time step length. If it becomes too steep it even might induce numerical instabilities.

by applying a projected gradient method. The optimization is carried out on four different equidistant

time grids with $5 \cdot 2^k, k = 0, \ldots, 3$ steps. The objective gradient evaluated for the initial control $u_0 := u_d$ is presented on the left and the computed optimal control on the right. Except for the last time step, we observe quite similar objective gradients on all meshes. Even though a final value of around -8000 would seem reasonable it is not what happens. In the final step the gradient increases rapidly toward zero because of the terminal condition $\xi(t_e) = 0$. Because this phenomenon occurs in all optimization iterations the optimal control at the final time remains at 300 whereas in all other time grid points it is optimized. This results in a steep descent causing huge temperature gradients at the process end. We observe an antiproportional dependence of the last time step length and the final slope. On the finest grid the extremely steep ascent even induces numerical instabilities. It is obvious that the four controls determined on different grids are mesh dependent. We want to point out that this problem does not only occur in the glass cooling problem when including the temperature gradient within the objective functional. It can be observed in any optimization problem where a smooth extension of the reduced gradient is not close to zero at final time. In the following we want to present two different approaches to overcome this drawback.

The time-dependent-weight approach (t-d-w) is based on a time-dependent linear decreasing weight $\delta_g(t)$ which is zero at the final time. This approach does not change the fact that the optimal control and the initial control coincide in the last time step. But the time-dependent weighting of the temperature gradient functional is a good way to avoid the jump in the last time step and force a smooth decrease of the optimal control towards $u_d(t_e)$. The performance of this approach for the same parameters and on the same grids as above can be seen in Figure 2.

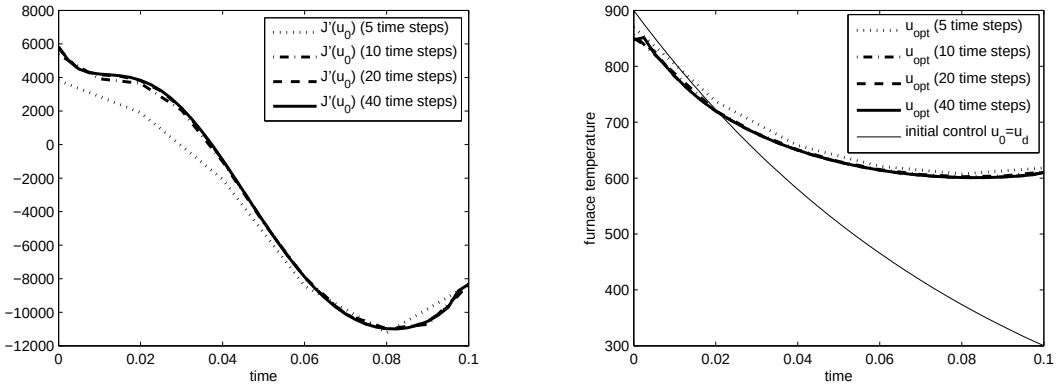


(a) Reduced Gradient (b) Optimal Control

Figure 2: Time-dependent-weight approach: objective gradients and optimal controls on different time grids. Here, the glass temperature gradient is included with a linearly decreasing time-dependent weight which vanishes at the final time. The reduced gradients increase smoothly towards zero (exemplarily for $\tilde{J}'(u_0)$ (left)) and hence the decrease of the optimal controls at the end of the process is also not as steep as it was before with the time independent temperature gradient weighting (right). Especially the curves determined on the three finest time grids show, that for sufficiently fine discretizations the slope of the final decrease does not depend on the final step length anymore.

The objective gradients (exemplarily for $\tilde{J}'(u_0)$ (left)) start with almost the same values as before but show a damped profile ending smoothly at zero. There is no noteworthy difference between the four gradients determined on different meshes. Especially on the three finest meshes, the resulting optimal controls (right) coincide as well. Nevertheless, for the great weight δ_g which was chosen for a better visualization, the optimal controls still show a noticeable descent at the process end. But in contrast to the setting above, for sufficiently fine discretizations, the slope of this descent is not correlated to the final step length anymore.

In the final-value-optimization approach (f-v-o) we augment the objective functional (15) with functionals penalizing the final value of all functionals which depend on the state. In our case this is the tracking of the glass temperature and the functional with the temperature gradient. The state derivatives of the new weighted functionals $\frac{\delta_e}{2}\left\|(T-T_d)(t_e)\right\|^2_{L^2(\Omega)}$ and $\frac{\delta_e\delta_g(t_e)}{2}\|\nabla T(t_e)\|^2_{L^2(\Omega)}$ will enter the adjoint system in the terminal condition which will be shown in detail later. With these new terminal conditions, which influence the initial values for a backwards solve, the terminal adjoint variable is different from zero and leads to a reduced gradient which optimizes the control also for the final time t_e (see Figure 3 (left)). An optimal control determined by this approach is mesh independent and the cooling is spread uniformly throughout the entire process duration. The optimal controls resulting from this approach for the above described setting are shown in Figure 3 (right). Their curves are quite similar and, except



(a) Reduced Gradient (b) Optimal Control

Figure 3: Final-value-optimization approach: objective gradients and optimal controls on different time grids. The objective functional is augmented by a final temperature tracking and a final temperature gradient optimization. The objective gradients shows reasonable values at the final time (exemplarily for $\tilde{J}'(u_0)$ (left)). The performance of the optimization algorithm is almost the same on all four grids. Especially the optimal controls determined on the three finest grids lie exactly above each other except for a little initial increase which can only be realized on the finest grid (right). In comparison to the two approaches without final value optimization, here, the cooling is spread homogeneously over the entire cooling process.

for a little initial increase which can only be realized on the finest grid, the optimal controls determined on the three finest grids lie exactly above each other. If we compare the results to the ones in Figure 1 we observe similar curves for the objective gradient and the furnace temperature profile but now with a reasonable continuation towards the process end.

We want to mention that the marginal decrease of the optimal control is due the very large weight δ_g which was chosen for a more noticeable visualization.

An objective functional providing for the above described approaches can be formulated as follows:

$$J(T,u) := \frac{1}{2}\int_0^{t_e}\left\|T-T_d\right\|^2_{L^2(\Omega)}\mathrm{d}t + \frac{1}{2}\int_0^{t_e}\delta_g(t)\|\nabla T\|^2_{L^2(\Omega)}\mathrm{d}t$$
$$+\frac{\delta_e}{2}\left\|(T-T_d)(t_e)\right\|^2_{L^2(\Omega)} + \frac{\delta_e\delta_g(t_e)}{2}\|\nabla T(t_e)\|^2_{L^2(\Omega)} + \frac{1}{2}\int_0^{t_e}\delta_u(t)(u-u_d)^2\mathrm{d}t. \qquad (16)$$

If we want to apply the time-dependent-weight approach we choose the gradient weight $\delta_g(t)$ as a linearly decreasing function with $\delta_g(t_e)=0$ and set the terminal weight δ_e equal to zero. Contrary, to follow the final-value-optimization approach we set $\delta_g(t)$ constant and $\delta_e \neq 0$.

4 The First-Order Optimality System

To determine an optimal boundary control for the glass cooling problem we have to minimize the objective functional (16) with respect to T and u. For this, we consider the constrained optimization problem

$$\min_{(T,u)} J(T,u) \text{ subject to (7)-(11).} \tag{17}$$

The corresponding first-order optimality system can be derived from a Lagrangian formalism [20].

Therefore we consider the optimal control problem (17) in a Hilbert space setting. Given a Hilbert space $H(\Omega)$ of functions defined on Ω, its dual space is denoted by $H^*(\Omega)$. We define

$$\begin{aligned}
Q &:= (0,t_e)\times\Omega,\\
\Sigma &:= (0,t_e)\times\partial\Omega,\\
Z &:= \{z\in L^2(0,t_e;H^1(\Omega)) : \partial_t z \in L^2(0,t_e;(H^1)^*(\Omega))\},\\
V &:= Z\times[L^2(0,t_e;H^1(\Omega))]^N,\\
W &:= [L^2(0,t_e;H^1(\Omega))]^{N+1}\times L^2(\Omega),\\
U &:= L^2(0,t_e).
\end{aligned}$$

Here, Q describes the space-time cylinder and Σ its spatial boundary. V is the space of states $v := (T,\phi_1,\dots,\phi_N)^T$, W is the space of adjoint states $\xi := (\xi_T,\xi_{\phi_1},\dots,\xi_{\phi_N},\xi_{T_0})^T$, and U is the control space.

Multiplication of the state system (7)-(11) with a test function $\xi\in W$ results in the following weak formulation: Find $v\in V$ with $T(x,0)=T_0(x)$ in $L^2(0,t_e;\Omega)$ such that

$$\begin{aligned}
&{}_{W^*}\langle e(v,u),\xi\rangle_W\\
&:= \int_0^{t_e} {}_{(H^1)^*}\langle \partial_t T,\xi_T\rangle_{H^1}\mathrm{d}t + \int_Q k_c\nabla\xi_T\nabla T\mathrm{d}x\mathrm{d}t - \int_Q\sum_{i=1}^N 4\pi\kappa_i B^{(i)}(T)\xi_{\phi_i}\mathrm{d}x\mathrm{d}t\\
&+\int_Q\sum_{i=1}^N\left(\frac{\epsilon^2}{3(\sigma_i+\kappa_i)}\nabla\xi_{\phi_i}+\frac{1}{3(\sigma_i+\kappa_i)}\nabla\xi_T\right)\nabla\phi_i\mathrm{d}x\mathrm{d}t + \int_Q\sum_{i=1}^N\kappa_i\phi_i\xi_{\phi_i}\mathrm{d}x\mathrm{d}t\\
&-\int_\Sigma\frac{\alpha\pi}{\epsilon}\left(\frac{n_a}{n_g}\right)^2\left(B^{(0)}(u)-B^{(0)}(T)\right)\xi_T\mathrm{d}s\mathrm{d}t - \int_\Sigma\frac{h_c}{\epsilon}(u-T)\xi_T\mathrm{d}s\mathrm{d}t\\
&-\int_\Sigma\sum_{i=1}^N c_1\epsilon\left(4\pi B^{(i)}(u)-\phi_i\right)\xi_{\phi_i}\mathrm{d}s\mathrm{d}t - \int_\Sigma\sum_{i=1}^N\frac{c_1}{\epsilon}\left(4\pi B^{(i)}(u)-\phi_i\right)\xi_T\mathrm{d}s\mathrm{d}t\\
&+\int_\Omega\left(T(x,0)-T_0(x)\right)\xi_{T_0}\mathrm{d}x = 0,
\end{aligned}\tag{18}$$

for all $\xi\in W$. Using (18) we define the Lagrangian

$$L(v,u,\xi) = J(v,u) + {}_{W^*}\langle e(v,u),\xi\rangle_W. \tag{19}$$

The first-order optimality system is then given by the Karush-Kuhn-Tucker (KKT) conditions

$$\partial_\xi L(v,u,\xi) = 0, \tag{20}$$

$$\partial_v L(v,u,\xi) = 0, \tag{21}$$

$$\partial_u L(v,u,\xi)(w-u)\geq 0\ \ \forall w\in U_{ad}. \tag{22}$$

Because of the box-constraints on the control we have to deal with an inequality in (22). Variation of $L(v,u,\xi)$ with respect to ξ yields the state system (7)-(11). From varying the state variable v we find the following adjoint system

$$-\partial_t \xi_T - \nabla\cdot(k_c\nabla\xi_T) - 4\pi\sum_{i=1}^{N}\kappa_i\partial_T B^{(i)}(T)\xi_{\phi_i} = -(T-T_d) + \delta_g\Delta T, \tag{23}$$

$$-\epsilon^2\nabla\cdot\left(\frac{1}{3(\sigma_i+\kappa_i)}\nabla\xi_{\phi_i}\right) - \nabla\cdot\left(\frac{1}{3(\sigma_i+\kappa_i)}\nabla\xi_T\right) = -\kappa_i\xi_{\phi_i},\ \ i=1,\dots,N, \tag{24}$$

with boundary and terminal conditions

$$k_c n\cdot\nabla\xi_T = -\left(\frac{h_c}{\epsilon} + \frac{\alpha\pi}{\epsilon}\left(\frac{n_a}{n_g}\right)^2\partial_T B^{(0)}(T)\right)\xi_T - \delta_g n\cdot\nabla T, \tag{25}$$

$$\frac{\epsilon^2}{3(\sigma_i+\kappa_i)}n\cdot\nabla\xi_{\phi_i} + \frac{1}{3(\sigma_i+\kappa_i)}n\cdot\nabla\xi_T = -c_1\epsilon\xi_{\phi_i} - \frac{c_1}{\epsilon}\xi_T,\ \ i=1,\dots,N, \tag{26}$$

$$\xi_T(t_e) = -\delta_e(T-T_d)(t_e) + \delta_e\delta_g\Delta T - \delta_e\delta_g n\cdot\nabla T, \tag{27}$$

where we identify the component ξ_{T_0} by $\xi_T(0)$ and formally set the outer normal n of an inner point $x\in\Omega$ equal to zero. The optimality condition (22) results in

$$\int_\Sigma\Big(-\frac{h_c}{\epsilon}\xi_T - \frac{\alpha\pi}{\epsilon}\Big(\frac{n_a}{n_g}\Big)^2\partial_T B^{(0)}(u)\xi_T - 4\pi c_1\sum_{i=1}^{N}\partial_T B^{(i)}(u)\left(\frac{1}{\epsilon}\xi_T + \epsilon\xi_{\phi_i}\right) + \frac{\delta_u}{|\partial\Omega|}(u-u_d)\Big)(w-u)\mathrm{d}s\mathrm{d}t \geq 0,\ \ \forall w\in U_{ad}. \tag{28}$$

Thus, to fulfill the KKT-conditions (20)-(22) we have to find a triple (u,v,ξ) solving the state system (7)-(11), the adjoint system (23)-(27), and the optimality condition (28).

Since (7)-(11) is uniquely solvable for an arbitrary $u\in U_{ad}$ with a Fréchet differentiable map $v: U_{ad}\to V$, $u\mapsto v(u)$ we can define a reduced objective functional $\hat{J}(u) := J(v(u),u)$ and determine its gradient $\hat{J}'(u)$ [11]. Fréchet differentiability of v is shown by splitting the operator $e(v,u)$ into a linear part D, a nonlinear part N acting on v, and a nonlinear part B acting on u. With this decomposition the operator $e(v,u)$ can be replaced by a continuous Fréchet differentiable operator $R(v,u) := v + D^{-1}N(v) + D^{-1}B(u)$. Based on this approach the applicability of the implicit function theorem can be shown. For more details we refer to Ref. [16] and deduce the directional state derivative

$$v'(u)[\delta u] = -\partial_v e^{-1}(v(u),u)\partial_u e(v(u),u)\delta u.$$

Applying this relation to the gradient of the reduced objective functional in the direction δu, we find

$${}_{U^*}\langle\hat{J}'(u),\delta u\rangle_U = {}_{U^*}\langle\partial_u J(v,u) - \partial_u e^*(v,u)\partial_v e^{-*}(v,u)\partial_v J(v,u),\delta u\rangle_U. \tag{29}$$

Using the adjoint system

$$\partial_v L(v,u,\xi) = \partial_v J(v,u) + \partial_v e^*(v,u)\xi = 0$$

we have $\xi = -\partial_v e^{-*}(v,u)\partial_v J(v,u)$. Substituting this expression in (29) and identifying U with its dual space U^*, the Riesz representative of the mapping $u\mapsto\hat{J}'(u)$ is given by

$$\hat{J}'(u) = \int_{\partial\Omega} -\frac{h_c}{\epsilon}\xi_T - \frac{\alpha\pi}{\epsilon}\Big(\frac{n_a}{n_g}\Big)^2\partial_T B^{(0)}(u)\xi_T - \frac{4\pi c_1}{\epsilon}\sum_{i=1}^{N}\partial_T B^{(i)}(u)\xi_T - 4\pi\epsilon c_1\sum_{i=1}^{N}\partial_T B^{(i)}(u)\xi_{\phi_i}\mathrm{d}s + \delta_u(u-u_d). \tag{30}$$

The relation between the optimality condition (28) and the reduced gradient (30) can now be used to find an optimal triple (u,v,ξ) fulfilling the KKT-conditions. This is done by means of a projected gradient method.

5 The Projected Gradient Method

We apply a gradient method with Armijo step size control to determine a triple (u, v, ξ) fulfilling the KKT-conditions. To compute an optimal control we successively improve the initial control $u^{(0)}$ by means of the reduced objective functional's gradient which gives us the direction $d^{(k)} := -\hat{J}'(u^{(k)})$ in which $u^{(k)}$ has to be modified. Its scaling β is determined by following Armijo's rule considering the first Wolfe condition

$$\hat{J}(u^{(k)} + \beta d^{(k)}) \leq \hat{J}(u^{(k)}) + \gamma_1 \beta d^{(k)} \hat{J}'(u^{(k)}). \tag{31}$$

The sufficient decrease condition (31) does not only ensure a monotone decrease but also avoids an overpassing of a minimum. In our computations we use $\gamma_1 = 1.0e-4$. To find a proper step size β we start with an initial guess $\beta = \beta_0$ and reduce β until (31) is fullfilled. We finally stop the algorithm if the relative difference between old and new target value is below a certain tolerance TOL_g. In spite of its simplicity, this strategy is fairly effective and our experience shows that the benefit of a more sophisticated line search is not worse the expense. Similar observations have been made in Ref. [17].

Summarizing, the projected gradient method reads as follows:

(1) initialization

(i) choose TOL_g, set $k = 0$, $\beta = \beta_0$, and $u^{(0)} = u_d$;

(ii) solve the state equations for $(T^{(0)}, \nabla T^{(0)}, \phi^{(0)})$ with $u = u^{(0)}$;

(iii) evaluate $J^{(0)} := J(T^{(0)}, \nabla T^{(0)}, u^{(0)})$ from (16);

(2) main loop

(iv) set $k := k + 1$;

(v) solve the adjoint equations for $(\xi_T^{(k)}, \xi_{\phi_1}^{(k)}, \ldots, \xi_{\phi_N}^{(k)})$ with $(T^{(k-1)}, \nabla T^{(k-1)}, \phi^{(k-1)})$;

(vi) set $u^{(k)} = P_{U_{ad}}(u^{(k-1)} - \beta \hat{J}'(u^{(k-1)}))$;

(vii) solve the state equations for $(T^{(k)}, \nabla T^{(k)}, \phi^{(k)})$ with $u = u^{(k)}$;

(viii) evaluate $J^{(k)} := J(T^{(k)}, \nabla T^{(k)}, u^{(k)})$ from (15);

(ix) if (31) holds, continue; otherwise, set $\beta = 0.5\beta$ and go to step (vi);

(x) if $|J^{(k)} - J^{(k-1)}|/|J^{(k)}| > \mathrm{TOL}_g$, set $\beta = 1.5\beta$ and go to (iv); otherwise, stop;

Here, the projector $P_{U_{ad}} : L^2(0, t_e) \to L^2(0, t_e)$ in (vi) is just a cut-off function which brings the iterate back to the feasible set U_{ad}.

6 Adaptive Space and Time Discretization

We are going to follow the so called "first-optimize-then-discretize" approach. This approach has the advantage that the discrete adjoint system is naturally consistent with the adjoint PDAE whereas in the "first-discretize-then-optimize" approach this must not be the case. Error estimates can be controlled not only in the state but also in the adjoint system. Therefore the adjoint state and hence the reduced gradient can be computed on appropriate adaptively refined grids to guarantee consistency between finite and infinite dimensional problem. If sufficient, the state system itself might be solved on quite coarse grids. Furthermore, we can choose independent integration schemes of high order and with good stability properties in accordance to the structure of the different PDAEs. An often discussed disadvantage of the "first-optimize-then-discretize" approach is the inconsistency between the reduced gradient and the minimization problem itself. But in contrast to the inconsistency between discrete minimization problem and the PDAE setting itself this inconsistency can be reliably controlled by adaptive grid adjustment.

To solve the state and the adjoint system we use the state-of-the-art software package KARDOS which allows the space-time adaptive solution of systems of partial differential algebraic equations

$$\sum_{j=1}^{n} H_{ij} \frac{\partial w_j(x,t)}{\partial t} = F_i(x,t,w,\nabla w) + \sum_{j=1}^{n} \nabla \cdot \left(P_{ij} \nabla w_j \right), \quad i = 1,\dots,n, \tag{32}$$

with $H_{ij} = H_{ij}(x,t,w,\nabla w)$ and $P_{ij} = P_{ij}(x,t,w,\nabla w)$. The state system can be easily formulated in this form. To set up the adjoint equations, the Laplacian ΔT has to be considered in the function $F_1(x,t,w,\nabla w)$. We avoid difficulties in computing second derivatives of the temperature by appending an auxiliary temperature $\bar{T}$ to the vector w. Augmenting the entire adjoint system with an additional algebraic equation $T - \bar{T} = 0$ we can force the auxiliary temperature $\bar{T}$ to coincide with the original temperature T inside the glass. The term $\delta_g \Delta \bar{T} = \delta_g \Delta T$ then appears in the summand $\sum_{j=1}^{n} \nabla \cdot \left(P_{ij} \nabla w_j \right)$ of (32) and can be used to handle the Laplacian of T in a weak sense making use of (25) with T replaced by $\bar{T}$. Applying these modifications the adjoint system almost fits into the form (32).

The last challenge, especially in the case of linear finite elements, is to approximate the Laplacian of the glass temperature which occurs within the terminal condition (27) but can not be handled by the auxiliary temperature $\bar{T}$. We consider a derivation based on weak formulation and partial integration to determine a proper approximation of the Laplacian for every vertex of the spatial grid, see also Ref. [18]. The weak formulation of the Laplacian with test function φ is given by

$$\int_\Omega \Delta T(x,t_e)\varphi(x)\mathrm{d}x = -\int_\Omega \nabla T(x,t_e)\nabla\varphi(x)\mathrm{d}x + \int_{\partial\Omega} n \cdot \nabla T(x,t_e)\varphi(x)\mathrm{d}\lambda. \tag{33}$$

Now, let the verticals of the spatial decomposition be denoted by $p_i, i \in \mathbb{N}$. To point out that three verticals belong to the same triangle D we will also refer to them by q_1, q_2, q_3. The patch of triangles around a vertex p_i, where the test function φ does not vanish is denoted by $\mathcal{D}_i$. Analogously, q_1 and q_2 describe the nodes of an edge E and $\mathcal{E}_i$ the set of edges, where φ does not vanish. In the following we consider the weak formulation within a linear finite element setting and substitute the test function φ by linear basis functions φ_i which are equal to one in the node p_i and zero in all other nodes. Approximating (33) by the midpoint rule which is exact for linear functions we have

$$\begin{aligned}
&\sum_{D\in\mathcal{D}_i} |D| \Delta T \left(\frac{q_1+q_2+q_3}{3}, t_e \right) \varphi_i \left(\frac{q_1+q_2+q_3}{3} \right) \\
&= -\sum_{D\in\mathcal{D}_i} |D| \nabla T \left(\frac{q_1+q_2+q_3}{3}, t_e \right) \nabla\varphi_i \left(\frac{q_1+q_2+q_3}{3} \right) \\
&\quad + \sum_{E\in\mathcal{E}_i} |E| n \cdot \nabla T \left(\frac{q_1+q_2}{2}, t_e \right) \varphi_i \left(\frac{q_1+q_2}{2} \right).
\end{aligned} \tag{34}$$

Making use of linearity and the definition of φ_i we get

$$\begin{aligned}&\sum_{D\in\mathcal{D}_i}|D|\frac{1}{3}\sum_{j=1}^{3}\Delta T(q_j,t_e)\varphi_i(q_j)\\&=-\sum_{D\in\mathcal{D}_i}|D|\nabla T(t_e)_{|_D}\nabla\varphi_{i|_D}+\frac{1}{2}\sum_{E\in\mathcal{E}_i}|E|n\cdot\nabla T(t_e)_{|_E}\end{aligned}\tag{35}$$

and therefore

$$\begin{aligned}&\Delta T(p_i,t_e)\\&=-\frac{3}{\sum_{D\in\mathcal{D}_i}|D|}\left[\sum_{D\in\mathcal{D}_i}|D|\nabla T(t_e)_{|_D}\nabla\varphi_{i|_D}+\frac{1}{2}\sum_{E\in\mathcal{E}_i}|E|n\cdot\nabla T(t_e)_{|_E}\right].\end{aligned}\tag{36}$$

Here, $\cdot|_D$ and $\cdot|_E$ stand for the constant value of a quantity on a triangle D or an edge E, respectively. If we now approximate ΔT using (36) and solve the adjoint system by means of linear finite elements it turns out that the boundary integral in (33) and the one resulting from the term $-\delta_e\delta_g n\cdot\nabla T$ in (27) annihilate each other. Summarizing, the terminal condition (27) can be approximated by

$$\xi_T(x,t_e)=-\delta_e(T-T_d)(x,t_e)-\frac{3\delta_e\delta_g}{\sum_{D\in\mathcal{D}_i}|D|}\sum_{D\in\mathcal{D}_i}|D|\nabla T(t_e)|_D\nabla\varphi_i|_D.\tag{37}$$

Hence, for the given setting the above described approximation is of great advantage, because it avoids the estimation of outer normals and the computation of the annihilated boundary integrals.

Using the transformation $\bar{t}:=t_e-t$ in the adjoint equations and neglecting the bar for the transformed time variable afterwards, both systems can be written in the abstract form

$$H\partial_t w=R(w)\quad\text{for }(x,t)\in Q,\tag{38}$$

$$w(t=0)=w^{(0)}\quad\text{for }x\in\Omega,\tag{39}$$

with

$$w=\begin{cases}(T,\phi_1,\dots,\phi_N)^T & \text{for the state system,}\\ \left(\xi_T,\xi_{\phi_1},\dots,\xi_{\phi_N},\bar{T}\right)^T & \text{for the adjoint system.}\end{cases}$$

The matrix H has only one nonzero value $H_{11}=1$ and the source vector $R(w)$ includes all differential operators supplemented with their boundary conditions. For the discretization in time, we use one-step methods of Rosenbrock type. Rosenbrock methods are designed by working the exact Jacobian directly into the formula, so no secondary inner iteration scheme is necessary. In addition, for moderate accuracy the perform even better than a fully implicit method. Rosenbrock methods are described by the recursive linear implicit one-step scheme

$$w_{n+1}=w_n+\sum_{i=1}^{s}b_iW_i^n,\tag{40}$$

$$\left(\frac{H}{\gamma\tau_n}-\partial_w R(w_n)\right)W_i^n=\sum_{j=1}^{i-1}\frac{c_{ij}}{\tau_n}HW_j^n+R\big(w_n+\sum_{j=1}^{i-1}a_{ij}W_j^n\big),\ i=1,\dots,s,\tag{41}$$

where τ_n denotes the step size, w_n the approximation to $w(t_n)$ at $t_n=\sum_{i=0,\dots,n-1}\tau_i$, and s is the number of stages. The basic idea of linearly implicit methods is that for the calculation of the intermediate

values W_i^n, $i = 1,\dots,s$, only a sequence of linear systems with one and the same operator has to be solved. One-step methods give us the opportunity to quickly change the step size in every step. This step size is adapted automatically according to the local temporal discretization error which can easily be estimated by using an embedded formula of inferior order. They are also defined by (40)-(41) but with different coefficients $\hat{b}_i$, $\hat{c}_{ij}$, $\hat{a}_{ij}$, and $\hat{\gamma}$. The coefficients b_i, $\hat{b}_i$, c_{ij}, $\hat{c}_{ij}$, a_{ij}, $\hat{a}_{ij}$, γ, and $\hat{\gamma}$ can be chosen to gain the desired order, high consistency and good stability properties like A- or L-stability [7, 10]. For the solution of the stiff and highly nonlinear state system, we use the 4-stage Rosenbrock method ROS3PL which is an L-stable order three method for nonlinear stiff differential algebraic equations of index one and is especially designed to avoid order reduction for parabolic problems [13, 12]. To solve the adjoint system backwards in time, we aim to exploit the PDAE on same discrete time points as used in the forward calculation of the state system. Considering a Rosenbrock scheme of higher order we can expect sufficiently small time errors even if the used time grid is not adapted to the adjoint system itself. Nevertheless, we control local error estimates and allow for refinement if necessary. Here, we apply the 6-stage Rosenbrock solver RODASP [19] which has order four for differential algebraic equations of index one and avoids order reduction for linear parabolic problems. For the complete sets of coefficients of ROS3PL and RODASP we refer to Ref. [12] and Ref. [11], respectively.

The equations (41) are linear elliptic problems which can be solved consecutively. This is done by means of a linear adaptive multilevel finite element method. The main idea of the multilevel technique consists of replacing the solution space by a sequence of discrete spaces with successively increasing dimensions to improve the approximation property. After computing linear approximations of the intermediate values W_i^n, $i = 1,\dots,s$, a posteriori error estimates can be utilized to give specific assessment of the error distribution [5]. The spatial errors are estimated by solving local Dirichlet problems on smaller subdomains as described in Ref. [10]. Added to the projection error resulting from the transfer between different meshes the cumulative error is an indicator where and how often the spatial grid has to be refined and where it could be even coarser. The linear algebraic systems resulting from the above described discretization are solved by the stabilized bi-conjugated gradient method BICGSTAB preconditioned with an incomplete LU-factorization.

This fully adaptive environment gives us the possibility to allow for appropriate independently refined spatial grids in each point of time. The grids, estimated within the state solve, are used as an initial grid for the adjoint computations. Nevertheless, it is important to control spatial error estimates in the backwards solve as well and refine the spatial grids if necessary.

Generally, the application of non-adjoint time integration schemes leads to inconsistency between the reduced gradient and the minimization problem itself. Therefore it is essential to control the discretization errors in both systems to ensure a sufficient consistency between discrete and continuous problem.

7 Computational Experiments

The projected gradient method described above is now applied to minimize (16) subject to the SP_1-approximation of the presented radiative heat transfer model. We investigate a three-dimensional infinitely long glass block, hence it is sufficient to consider a two-dimensional slice as computational domain. We set $\Omega = [-3,3] \times [-1,1] \in \mathbb{R}^2$ adaptively discretized with 865 spatial nodes as shown in Figure 4. In the computational experiment presented here the process time interval $[0, 0.1]$ is discretized

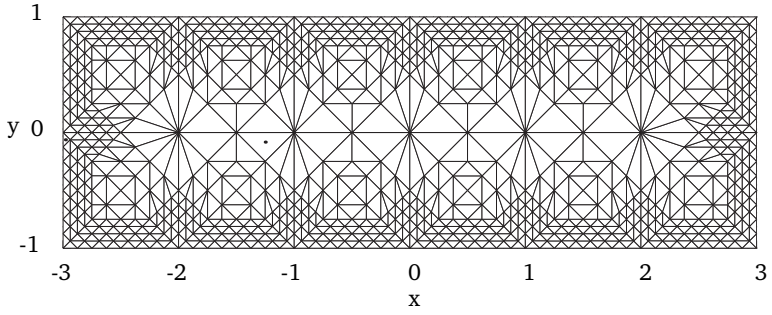


Figure 4: Initial triangulation of the computational domain. Because the glass cools down faster at its periphery the grid is locally refined at the boundary.

by 24 time steps. They are evaluated adaptively within the first optimization iteration. We start with an initial time step of $5.0e-5$ which is increased adaptively such that all relative error estimates stay below $5.0e-3$. This significantly finer discretization within the beginning of the process is necessary to resolve a small increase of the optimal control at the beginning of the cooling. It can barely be seen in the following figures showing the optimal control but becomes clear when looking at the temperature gradient. Note that in a uniform approach we would be forced to an average of 2000 equidistant time steps due to the very small step at the beginning.

The non-opaque frequency interval (ν_0, ∞) is discretized by an eight-band model [1, 11] as shown in Figure 5. The first six bands divide the region of thermal infrared. The visible light is located within the

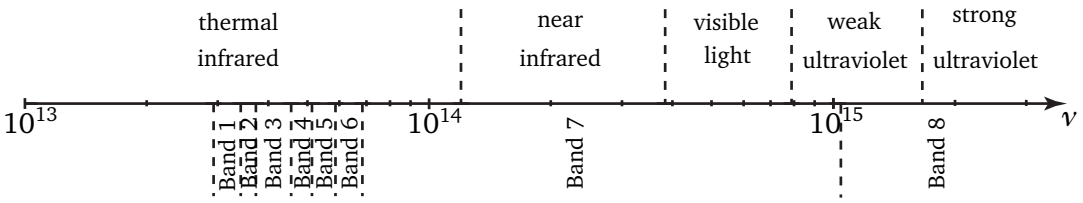


Figure 5: The continuous frequency spectrum is approximated by a discrete eight-band model with constant coefficients on each interval.

seventh and the strong ultraviolet radiation within the eighth band. The corresponding absorption rates are given in Table 1 [1]. It is interesting to mention that for high temperatures the simulations show the largest radiative intensity within the seventh band which involves the visible light. As the temperature decreases this peak shifts towards the third band located within the infrared region. This observations coincides with those one can make in reality. At the beginning of the cooling process the hot glass glows. Once it gets colder radiation can not be seen anymore but still be felt in form of heat. In the eighth band we observe almost no radiation, which is related to the opaqueness of glass for strong ultraviolet light.

Table 1: Bounds and absorption rates of the eight-band model.

Band i	$\nu_{i-1}(10^{13}s^{-1})$	$\nu_i(10^{13}s^{-1})$	$\kappa_i(m^{-1})$
-	0	2.9334638	opaque
1	2.9334638	3.4223744	7136.00
2	3.4223744	3.7334994	576.32
3	3.7334994	4.5631659	276.98
4	4.5631659	5.1335616	27.98
5	5.1335616	5.8669276	15.45
6	5.8669276	6.8447489	7.70
7	6.8447488	102.6712329	0.50
8	102.6712329	∞	0.40

We assume clean glass, so scattering can be omitted. All further parameters and constants are set as follows:

$$
\begin{aligned}
h_b &= 6.62608 \cdot 10^{-34} & h_c &= 0.001 \\
k_b &= 1.38066 \cdot 10^{-23} & k_c &= 1.0 \\
c_0 &= 2.990 \cdot 10^{8} & \alpha &= 0.914 \\
n_g &= 1.46 & c_1 &= 0.14936 \\
a_a &= 1.0 & \epsilon &= 1.0.
\end{aligned}
$$

To enforce a furnace temperature close to the initial glass temperature at the beginning of the process we chose a large value $\delta_u(0)$ and decrease the weight $\delta_u(t)$ quadratically to a constant value within the first quarter of the process duration. If the time-dependent-weight approach is applied the weight for the final value optimization $\delta_{e,0}$ is set to zero. In case of the final-value-optimization approach we set $\delta_{e,1} = 0.001$. The effect of the temperature gradient to the optimal control is examined by varying the gradient weight δ_g which is chosen as a linearly decreasing function to follow the t-d-w approach and set constant in case of the f-v-o approach. An overview of the detailed choice of weights can be found in Table 2.

Table 2: Overview of optimization weights.

time-dependent-weight approach (t-d-w)	final-value-optimization (f-v-o)
$\delta_{e,0} = 0.0$	$\delta_{e,1} = 0.001$
$\delta_{g,k}(t) = \delta_{g0,k} - \frac{\delta_{g0,k}}{t_e} t$	$\delta_{g,k}(t) = \delta_{g0,k}$
$\delta_{g0,k} = \begin{cases} 0.0 & \text{if } k = 0 \\ 0.1 & \text{if } k = 1 \\ 0.5 & \text{if } k = 2 \\ 1.0 & \text{if } k = 3 \\ 5.0 & \text{if } k = 4 \end{cases}$	
$\delta_u(t) = \begin{cases} 1600(1 - \max\{1, 20 \cdot \delta_{g0,k}\})t^2 + \max\{1, 20 \cdot \delta_{g0,k}\} & \text{if } t < 0.25 t_e \\ 1 & \text{else} \end{cases} \quad k = 0, \ldots, 4$	

We start the cooling process with a uniform glass temperature distribution $T_0(x) = 900$, where the reference quantity is $1K$. The spatially uniform desired state $T_d(x,t)$ is set to

$$T_d(t) := T_0 \exp\left(-\frac{t}{t_e}\log\left(\frac{T_0}{300}\right)\right).$$

When the cooling process is stopped at $t_e = 0.1$ the value $T_d(t_e) = 300$ enforces a final temperature distribution as close as possible to room temperature. The dimensionless furnace temperature is restricted to the set of admissible controls $U_{ad} := \{u \in L^2(0,t_e) : 300 \le u \le 900\}$ and its reference value u_d is set to the desired temperature profile T_d. In each iteration step we derive consistent initial values for the mean intensities ϕ_i, $i = 1,\dots,8$, with respect to the initial glass temperature and the current control iterate. Before solving the adjoint system we compute consistent adjoint values ξ_{ϕ_i}, $i = 1,\dots,8$, with respect to the terminal condition $\xi_T(x,t_e) = -\delta_e(T - T_d)(x,t_e) + \delta_e\delta_g\Delta T(x,t_e) - \delta_e\delta_g n\cdot\nabla T(x,t_e)$. Figure 6 gives an overview about all cases considered and the resulting optimal controls plotted over time. In the

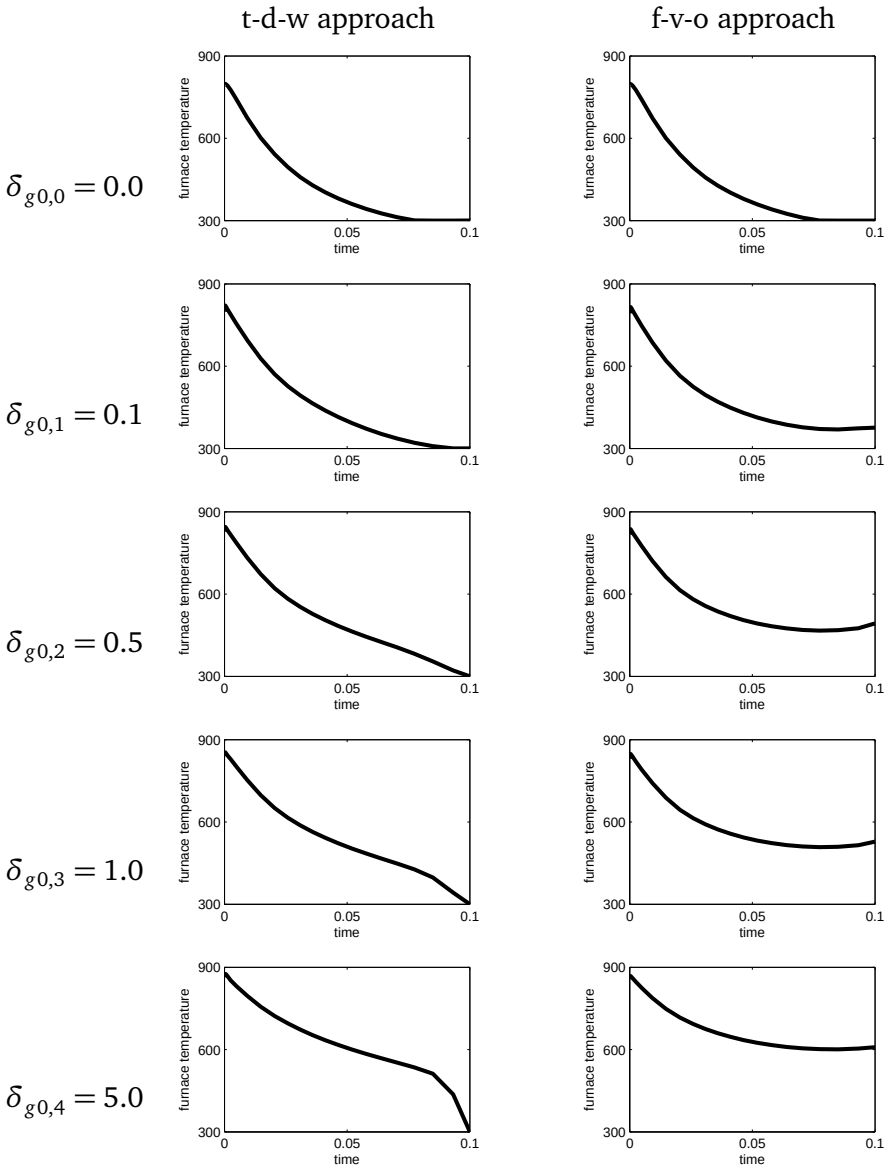


Figure 6: Overview of optimal controls resulting from the time-dependent-weight approach (left) and the final-value-optimization approach (right) for different temperature gradient weights.

case of $\delta_{g0,0} = 0.0$, where the temperature gradient is excluded from the objective, there is almost no

offset between the optimal controls resulting from the two approaches. This is due to the fact that the only difference lies in the summand $\frac{0.001}{2}\|(T - T_d)(t_e)\|^2_{L^2(\Omega)}$ of the objective functional which allows to optimize the final control value. In this setting the summand does not really change the control because in the second half of the time interval the reduced gradient is greater than zero and the lower box constraint is active. Nevertheless this term can not be neglected in general. Just think about a case with the same setting but shorter process duration such that the lower box constraint is not reached yet. Also for $\delta_{g0,1} = 0.1$ and $\delta_e = 0.0$ the lower box constraint is active in the second half of the time interval with a positive reduced gradient. Including final values the reduced gradient in the second half becomes negative and therefore the optimal control is increased. To study the difference between the two approaches in detail we want to have a closer look at the case where $\delta_{g0,2} = 0.5$. Both approaches come up with an optimal control which does not start at 900 as the initial guess but a little cooler at around 800. Then, the optimal control resulting from the t-d-w approach shows a smooth decrease towards the predefined final value of 300. Contrarily, the optimal control resulting from the f-v-o approach does not decrease below 450 throughout the entire time interval. It even increases a little within the last third (see Figure 7). On the boundary the t-d-w optimal control suits the glass temperature better to the desired profile

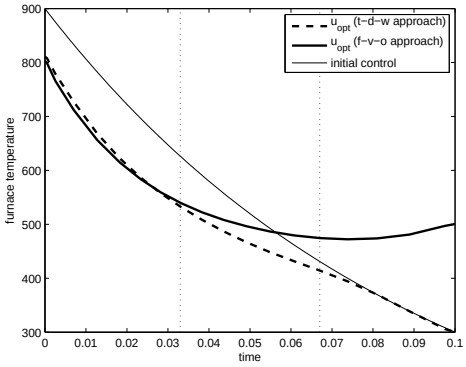
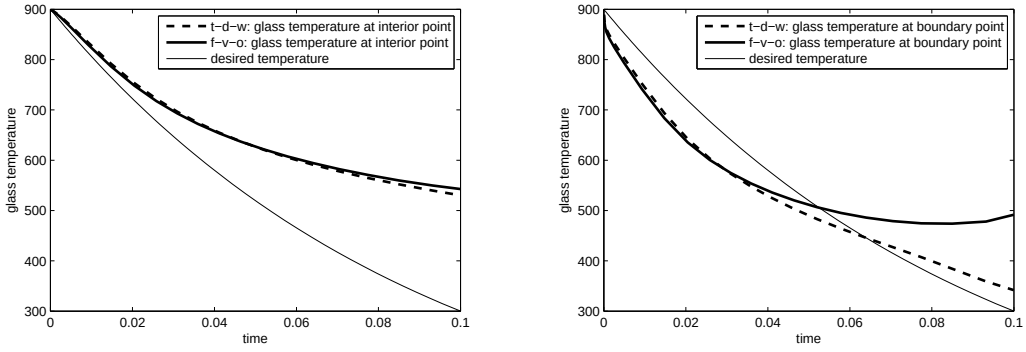


Figure 7: Comparison of the optimal control resulting from the t-d-w and the f-v-o approach with $\delta_{g0,2} = 0.5$. The resulting optimal controls coincide during the first third of the time interval, they split up in the second third, and vary significantly in the last third.

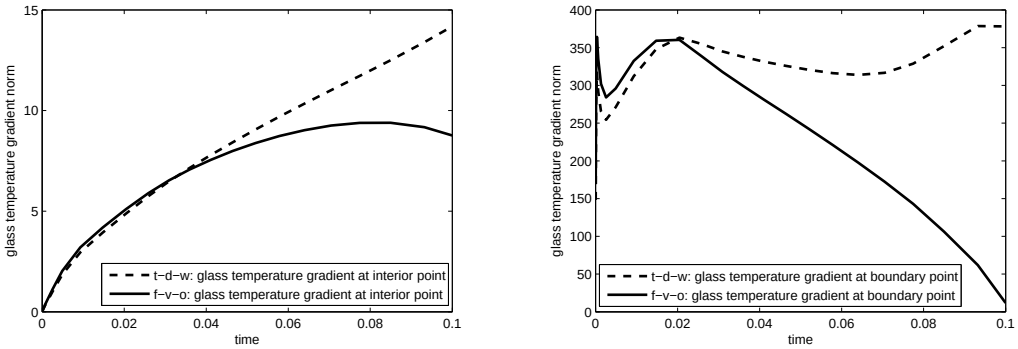
than the f-v-o control whereas in the interior there is almost no difference (see Figure 8). Contrariwise,



(a) Temperature in interior (b) Temperature on boundary

Figure 8: Resulting glass temperature close to the center $(0, 0)$ **(left) and on the boundary** $(3, 1)$ **(right) for both approaches with** $\delta_{g0,2} = 0.5$. Especially on the boundary the t-d-w approach fits the glass temperature much closer to the desired profile than the f-v-o approach.

the temperature gradients are significantly reduced by the f-v-o approach, especially in the second half of the time interval on the boundary as well as in the interior (see Figure 9). Nevertheless the reduction of the reduced gradient achieved by the t-d-w approach with $\delta_{g0,2} = 0.5$ is still sufficient. On the boundary the temperature gradient norms are spread constantly over the time interval and in the interior they are smaller anyway. The impact of the optimal controls on the glass temperature and its gradient for all



(a) Temperature gradient in interior (b) Temperature gradient on boundary

Figure 9: Resulting Euclidean glass temperature gradient norm close to the center $(0,0)$ **(left) and on the boundary** $(3,1)$ **(right) for both approaches with** $\delta_{g0,2} = 0.5$**.** Especially in the second half of the time interval the f-v-o approach reduces the temperature gradient significantly in comparison to the t-d-w approach.

considered cases is summarized in terms of L^2-norms in Table 3. The results of the t-d-w approach are highlighted with white, those of the f-v-o approach with gray. We can see that for all cases $\delta_{g0} > 0.0$

Table 3: Objective details for t-d-w approach (white) and f-v-o approach (gray).

	$\hat{J}(u_{\text{opt}})$	$\|T - T_d\|_Q^2$	$\|(T - T_d)(t_e)\|_\Omega^2$	$\|\nabla T\|_Q^2$	$\|\nabla T(t_e)\|_\Omega^2$	$\|u - u_d\|_{[0,t_e]}^2$
$\delta_{e,0}$, $\delta_{g0,0}$	5.50e+3	9.36e+3	3.52e+5	5.82e+4	4.59e+5	1.76e+3
$\delta_{e,1}$, $\delta_{g0,0}$	5.68e+3	9.34e+3	3.51e+5	5.85e+4	4.58e+5	1.79e+3
$\delta_{e,0}$, $\delta_{g0,1}$	6.86e+3	1.02e+4	3.73e+5	5.08e+4	4.84e+5	1.16e+3
$\delta_{e,1}$, $\delta_{g0,1}$	8.25e+3	1.09e+4	4.11e+5	4.11e+4	2.42e+5	1.03e+3
$\delta_{e,0}$, $\delta_{g0,2}$	1.07e+4	1.34e+4	4.40e+5	3.53e+4	5.51e+5	3.40e+2
$\delta_{e,1}$, $\delta_{g0,2}$	1.40e+4	1.62e+4	5.97e+5	2.00e+4	2.77e+4	8.09e+2
$\delta_{e,0}$, $\delta_{g0,3}$	1.40e+4	1.66e+4	4.98e+5	2.72e+4	6.13e+5	1.78e+2
$\delta_{e,1}$, $\delta_{g0,3}$	1.83e+4	2.09e+4	7.30e+5	1.33e+4	1.09e+4	1.02e+3
$\delta_{e,0}$, $\delta_{g0,4}$	2.65e+4	3.10e+4	7.31e+5	1.35e+4	8.57e+5	8.71e+2
$\delta_{e,1}$, $\delta_{g0,4}$	3.33e+4	3.93e+4	1.19e+6	4.61e+3	1.09e+3	2.45e+3

Table 4: Temperature gradient details for t-d-w approach (white) and f-v-o approach (gray).

	$\|\nabla T\|_Q^2$	$\|\nabla T(t_e)\|_\Omega^2$	$\max_{t_j} \|\nabla T\|_\Omega^2$	j	$\max_{t_k}(\max_x \|\nabla T\|_2^2)$	k
$\delta_{e,0}$, $\delta_{g0,0}$	5.82e+4	4.59e+5	7.11e+5	16	7.36e+2	14
$\delta_{e,1}$, $\delta_{g0,0}$	5.85e+4	4.58e+5	7.15e+6	16	7.39e+2	14
$\delta_{e,0}$, $\delta_{g0,1}$	5.08e+4	4.84e+5	6.05e+5	19	6.61e+2	14
$\delta_{e,1}$, $\delta_{g0,1}$	4.11e+4	2.41e+5	5.13e+5	14	6.39e+2	12
$\delta_{e,0}$, $\delta_{g0,2}$	3.53e+4	5.51e+5	5.51e+5	24	5.84e+2	24
$\delta_{e,1}$, $\delta_{g0,2}$	2.00e+4	2.77e+4	3.04e+5	12	5.11e+2	11
$\delta_{e,0}$, $\delta_{g0,3}$	2.72e+4	6.13e+5	6.13e+5	24	6.39e+2	24
$\delta_{e,1}$, $\delta_{g0,3}$	1.33e+4	1.08e+4	2.22e+5	11	4.41e+2	11
$\delta_{e,0}$, $\delta_{g0,4}$	1.35e+4	8.56e+5	8.56e+5	24	8.18e+2	24
$\delta_{e,1}$, $\delta_{g0,4}$	4.61e+3	6.59e+2	8.73e+4	11	2.83e+2	10

the t-d-w approach matches the glass temperature closer to T_d (column 2) whereas the f-v-o approach achieves a greater reduction of the temperature gradient (column 4). In Table 4 we have a closer look at the temperature gradient, its value at the process end, its maxima over the process duration and the space-time cylinder. Again, the results of the t-d-w approach are highlighted with white, those of the f-v-o approach with gray. We can state the following observations.

- Both approaches show a satisfactory decrease of the $L^2(Q)$-norm of the temperature gradient in accordance to the weight δ_g (column 1). However, if we have a look at the temperature gradient value at the final time it is decreased by the f-v-o approach whereas by the t-d-w approach it is increased (column 2). For sure, this is due to the fact that in the t-d-w approach the temperature gradient is weighted with zero at final time and hence not even minimized at the end whereas in the f-v-o approach we include an extra term to minimize the final gradient value.

- Following the t-d-w approach the maximal temperature gradient $L^2(\Omega)$-norm (column 3) and the maximal Euclidean norm (column 5) is reduced for $0 \leq \delta_{g0} \leq 0.5$. While the maximal gradient norms descent the point of time when they occur shifts towards the process end (column 4 and 6). Once the final time is reached, the maximal gradient norm can not be reduced anymore. For increasing $\delta_{g0} \geq 0.5$ the maximal gradient even increases again. This means using the t-d-w approach within this setting the temperature gradient can not be reduced further than in case of the scaling $\delta_{g0} \geq 0.5$.

- The maximal $L^2(\Omega)$- and Euclidean norms of the f-v-o approach occur somewhere in the middle of the time interval and gives rise to a uniformly spread cooling throughout the entire time interval for every gradient weight δ_{g0}. Following this approach the temperature gradient norm can be reduced arbitrary far.

8 Summary

Summarizing, we can say that with the time-dependent-weight approach and the final-value-optimization approach we found two suitable ways to overcome inconsistencies of the terminal adjoint condition. We want to recall that this problem does not only occur in the glass cooling problem when including the temperature gradient within the objective functional. It can be observed in any optimization problem where a smooth extension of the reduced gradient is not close to zero at final time. In the following we want to discuss under which circumstances which approach should be chosen.

- If we apply the t-d-w approach the optimal control will always decrease down to the final value of he initial control. If it is set to room temperature the resulting optimal control induces a final glass temperature which is at least close to room temperature. This has to be payed for by the fact that temperature gradients can not be forced arbitrary small. The greater the weight δ_{g0} is chosen the more cooling is shifted towards the process end where the function $\delta_g(t)$ tends to zero. Then, a more homogeneous cooling could only be enforced by choosing a steeper decreasing function $\delta_g(t)$ which somehow annihilates the great gradient weight again. Hence the choice of weights and weighting functions requires flair and experience. An advantage of this approach is the fact that it can be implemented with many standard PDE-solver software packages.

- By contrast, finding weights for the f-v-o approach is intuitive. Independently from the choice of weights, the cooling is spread uniformly throughout the process duration. Even if the gradient weight δ_{g0} is chosen quite high and the process duration t_e quite short this approach will induce an optimal control which minimizes temperature gradients arbitrary far. As a consequence the glass temperature itself might not necessarily be cooled down close to room temperature. A drawback of this approach is the fact that the implementation of initial adjoint values including the Laplacian requires full access to the source code of the PDE solver.

Currently, we work on including more sophisticated second-order multilevel optimization methods based on a globally oriented error control [3]. Future research will be on the development of discrete adjoint linearly implicit integration schemes to ensure consistence between the adjoint based reduced gradient and the optimization problem itself. Pointwise restrictions on the state gradient as recently proposed and analyzed for a mixed finite element method in Ref. [4] are also very promising.

Acknowledgements

The authors gratefully acknowledge the support of the DFG Priority Program 1253 entitled Optimization with Partial Differential Equations. We also want to mention Michael Herty who gave us the first impulsion for the consideration of final state tracking within this setting. Furthermore we want to thank the reviewer who made the proposal to include time-dependent weights.

References

[1] Fraunhofer–Institut für Techno– und Wirtschaftsmathematik, Kaiserslautern, Germany, *www.itwm.de*.

[2] M.K. Choudhary and N.T. Huff. Mathematical modelling in the glass industry: an overview of status and needs. *Glastech.Ber.Glass Sci.Technol.*, 70:363–370, 1997.

[3] K. Debrabant and J. Lang. On global error estimation and control for parabolic equations. Technical Report 2512, Technische Universität Darmstadt, 2007.

[4] K. Deckelnick, A. Günther, and M. Hinze. Finite element approximation of elliptic control problems with constraints on the gradient. Technical Report SPP1253-08-02, Priority Program 1253, German Research Foundation, 2007.

[5] P. Deuflhard, P. Leinen, and H. Yserentant. Concepts of an adaptive hierarchical finite element code. *Impact of Comput. in Sci. and Engrg.*, 1:3–35, 1989.

[6] B. Erdmann, J. Lang, and R. Roitzsch. KARDOS-User's Guide. Manual, Konrad-Zuse-Zentrum Berlin, 2002.

[7] E. Hairer and G. Wanner. *Solving Ordinary Differential Equations II, Second Revised Edition*. Springer Series in Computational Mathematics. Springer, 1996.

[8] K. Ito and K. Kunisch. Augmented Lagrangian-SQP methods for nonlinear optimal control problems of tracking type. *SIAM J. Control and Optimization*, 34:874–891, 1996.

[9] A. Klar, J. Lang, and M. Seaïd. Adaptive solution of SP_N-approximations to radiative heat transfer in glass. *International Journal of Thermal Science*, 44:1013–1023, 2005.

[10] J. Lang. *Adaptive Multilevel Solution of Nonlinear Parabolic PDE Systems. Theory, Algorithm, and Applications*, volume 16 of *Lecture Notes in Computational Science and Engineering*. Springer Verlag, 2000.

[11] J. Lang. Adaptive computation for boundary control of radiative heat transfer in glass. *Journal of Computational and Applied Mathematics*, 183:312–326, 2005.

[12] J. Lang and D. Teleaga. Towards a fully space-time adaptive FEM for magnetoquasistatics. *IEEE Transactions on Magnetics*, 44,6:1238–124, 2008.

[13] J. Lang and J. Verwer. ROS3P - an accurate third-order Rosenbrock solver designed for parabolic problems. *BIT*, 41:730–737, 2001.

[14] E. W. Larsen, G. Thömmes, A. Klar, M. Seaïd, and T. Götz. Simplified P_N approximations to the equations of radiative heat transfer and applications. *Journal of Computational Physics*, 183:652–675, 2002.

[15] M. F. Modest. *Radiative Heat Transfer*. McGraw-Hill, New York, second ed. edition, 2003.

[16] R. Pinnau. Analysis of optimal boundary control for radiative heat transfer modeled by the SP_1-system. *Communications in Mathematical Sciences*, 5,4:951–969, 2007.

[17] R. Pinnau and G. Thömmes. Optimal boundary control of glass cooling processes. *Mathematical Methods in the Applied Sciences*, 27:1261 – 1281, 2004.

[18] V. Dolejší. Anisotropic mesh adaption for finite volume and finite element methods on triangular meshes. *Computing and Visualization in Science*, 1:165–178, 1998.

[19] G. Steinebach. Order–reduction of ROW–methods for DAEs and method of lines applications. Technical Report 1741, Technische Hochschule Darmstadt, 1995.

[20] F. Tröltzsch. *Optimale Steuerung partieller Differentialgleichungen : Theorie, Verfahren und Anwendungen*. Vieweg, Wiesbaden, 2005.