MINIMUM WEIGHT TOPOLOGY OPTIMIZATION SUBJECT TO UNSTEADY HEAT EQUATION AND SPACE-TIME POINTWISE CONSTRAINTS – TOWARD AUTOMATIC OPTIMAL RISER DESIGN IN THE SHAPE CASTING PROCESS

R. TAVAKOLI

Abstract. The automatic optimal design of feeding system in the shape casting process is considered in the present work. In fact, the goal is to find the optimal position, size, shape and topology of risers in the shape casting process. This problem is formulated as a minimum weight topology optimization problem subjected to a nonlinear transient PDE and space-time state dependent pointwise constraints. An elegant bi-level reformulation of the optimization problem is introduced which makes it possible to manage the infinite number of design parameters and state-constraints efficiently. The computational cost of this method is independent of the number of design parameters and constraints. The validity and efficiency of the presented method are supported by several examples, from simple benchmarks to complex industrial castings. According to numerical results, the presented approach makes a complete solution to the problem of automatic optimal rider design in the shape casting process.

Keywords. adjoint sensitivity, feeder design, Niyama criterion, pointwise constraints, projected gradient, regularization.

1. Introduction

Metal casting is an important process to produce near net-shape products, which are used extensively in automotive and aerospace industries. Since metals usually contrast on the solidification, the last freezing points in castings encounter the lack of molten metal. If this leakage is not compensated properly, it leaves some shrinkage porosities in either of macroscopic or microscopic form inside castings. Risers are appended to the castings at process design stage, to establish the directional solidification from the casting to the risers so that the final solidification points are located within the risers. They are cut-off and recycled after the solidification. The goal of an optimal riser design procedure is to find the optimal location(s), size(s) and shape(s) of riser(s). Moreover, the total weight of riser(s) should be minimized to improve the casting yield and productivity. The goal of this paper is to introduce a method for automatic optimal riser design in the shape casting process.

Although methods for the product design optimization are well documented in literature (c.f. [7]), the process design optimization, in particular in the field of riser design, has received less attention in spite of its importance.

Date: April 7, 2020.

R. Tavakoli, Department of Material Science and Engineering, Sharif University of Technology, Tehran, Iran, P.O. Box 11365-9466, rtavakoli@sharif.ir.

Parametric optimization of feeding system design is considered in [10–12, 15, 17]. In these approached, a nearly feasible parameterized initial design is considered and then its parameters are determined using a black-box constrained optimization tool to increase the casting yield. In [32, 34–37], evolutionary topology optimization algorithm is employed for automatic optimal design of feeding system in the gravity casting process. In [20, 21, 25, 40], the optimization of riser design is formulated as a parametric shape optimization problem which is solved by a gradient based minimization method. The objective function to be minimized is defined as the riser volume and a few constraints are defined to enforce the directional solidification along a priori-defined feeding path. Two important prerequisites of these methods are a nearly feasible initial design and a user-defined feeding path. These prerequisites make these approaches far from our ultimate goal which is the automatic optimal riser design in the shape casting process.

The goal of present study is to introduce a mathematical model and its corresponding numerical method to automate the above mentioned optimal design problem. It can be accounted as a follow up part of our early work (see: [35]) in this field.

2. Conceptual modeling

The selection of design parameters is the first step of every optimal design problem. In previous works (see: [20, 40]) the shape parameters of riser(s) are considered as the design parameters. The small number of design parameters is the main benefit of this approach. However, it is not able to change the topology of feeding system. Moreover, it needs an appropriate initial design and a user defined feeding path. To overcome these limitations, the topology optimization approach (c.f. [7]) is adapted in the present study. It is originally developed for the optimal design of macro or micro mechanical structures. In this method, the topology indicator function at each spatial point is defined as the design parameters. Therefore, there are an infinite number of design parameters, before the spatial discretization.

Without loss of generality, the physical domain, $\mathcal{D} \subset \mathbb{R}^3$, is assumed to be rectangular, i.e., $\mathcal{D} = [0, l_x] \times [0, l_y] \times [0, l_z]$. In practice, $\mathcal{D}$ is identical to the mold-box in the shape casting process. Henceforth, the global spatial domain is called as the mold-box in this study. The original casting, denoted by $\mathcal{D}_c \subset \mathcal{D}$, is considered as an embedded object within the mold-box. The topology indicator function is equal to unity inside $\mathcal{D}_c$. This function will be remained constant inside $\mathcal{D}_c$ during the optimal design procedure. The position of casting inside $\mathcal{D}$ and l_x, l_y, l_z should be selected so that there will be sufficient space in the mold-box to design an appropriable feeding system. Excluding $\mathcal{D}_c$ from $\mathcal{D}$, the remainder of mold-box is denoted by $\mathcal{D}_r$, i.e., $\mathcal{D}_r = \mathcal{D} \backslash \mathcal{D}_c$. The design space of feeding system, $\mathcal{D}_d$, is a subset of $\mathcal{D}_r$, i.e., $\mathcal{D}_d \subseteq \mathcal{D}_r$. In the worst conditions, $\mathcal{D}_d$ is identical to $\mathcal{D}_r$. However, it is preferred to restrict $\mathcal{D}_d$ as much as possible, e.g. excluding portions of $\mathcal{D}_r$ which are infeasible for the design of feeding system. For instance, one can excludes bottom portions of the mold-box from $\mathcal{D}_d$. In our model, $\mathcal{D}_d$ is decomposed into two subdomains $\mathcal{D}_{dr}$ and $\mathcal{D}_{dn}$ such that $\mathcal{D}_d = \mathcal{D}_{dr} \cup \mathcal{D}_{dn}$ and $\mathcal{D}_{dr} \cap \mathcal{D}_{dn} = \Gamma_{rn}$. $\mathcal{D}_{dr}$ and $\mathcal{D}_{dn}$ respectively denote the riser and riser-neck [1] design domains. Γ_{rn} denotes the boundary between $\mathcal{D}_{dr}$ and $\mathcal{D}_{dn}$. In practice, $\mathcal{D}_{dn}$ is a narrow shell of l_n thickness

[1] Since riser(s) should be detached from the casting after the solidification, the connection area of riser to casting, called as the riser-neck, should be minimized to reduce the machining cost.

around the cast part. To improve the quality of final design, the user can exclude infeasible portions of $\mathcal{D}_d$ from $\mathcal{D}_{dn}$. For instance, the connection of risers to high curvature and bottom surfaces is discouraged. The above mentioned design domains are shown schematically in figure 1. The topology indicator function is equal to zero inside $\mathcal{D}_r \setminus (\mathcal{D}_{dn} \cup \mathcal{D}_{dr})$ and will be fixed there during the optimization. It varies inside $\mathcal{D}_{dn}$ and $\mathcal{D}_{dr}$ during the optimization. It assumes the value of either unity or zero in these regions that is equivalent to the existence of metal and mold materials there respectively. In fact, our goal is to determine the optimal value of the topology indicator function in $\mathcal{D}_{dn} \cup \mathcal{D}_{dr}$ such that its corresponding design results in a defect free casting. Moreover, to improve the casting yield, the total volume of metal phase in $\mathcal{D}_d$ should be minimized. Assume that the volume of $\mathcal{D}_d$, $\mathcal{D}_{dr}$ and $\mathcal{D}_{dn}$ are denoted by V_d, V_{dr} and V_{dn} respectively. For the purpose of riser-neck design, the total volume fraction of metal phase in $\mathcal{D}_{dn}$, denoted by R_{dn}, is constrained by a user-defined upper bound.

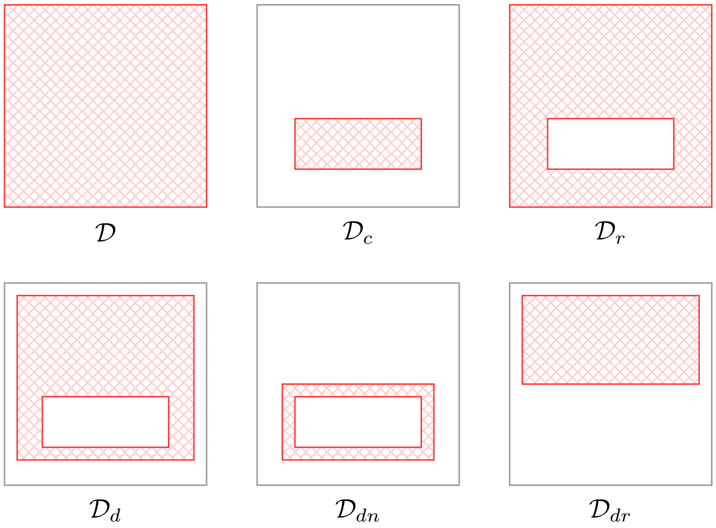


Figure 1. Schematic of spatial domains $\mathcal{D}$, $\mathcal{D}_c$, $\mathcal{D}_r$ $\mathcal{D}_d$, $\mathcal{D}_{dn}$ and $\mathcal{D}_{dr}$ defined in this study (shaded regions).

Assuming the physical domain is discretized into a uniform Cartesian grid and the topology indicator field takes a fixed value within each cell (i.e. a piecewise-constant approximation), then, this discretized domain is equivalent to a three-dimensional black-white image in which the black value at each point is identical to the existence of casting (includes the feeding system) and the white value is identical to the existence of mold material. The black-white values in $\mathcal{D}_{dn} \cup \mathcal{D}_{dr}$ are the design parameters in this study (see figure 2). To produce a defect free casting, it is required to predict the formation of freezing defects quantitatively, and then to pose constraints on the remained solidification defects inside the casting. Two important requirements of a method for quantitative prediction of solidification induced defects are: reasonable computational cost and differentiability for the purpose of numerical optimization. Our experience with different defect prediction methods

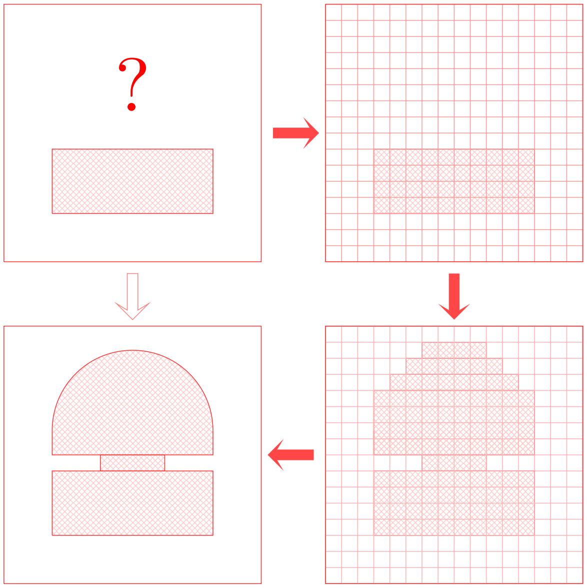


FIGURE 2. Schematic of automatic optimal riser design approach based on topology optimization method in the present study.

suggests that the thermal criterion functions (see [28]) are reasonable choice for this purpose. These methods use thermal history of casting to provide a pointwise measure which shows the susceptibility of local defect formation. To predict the formation of solidification induced defects, the Niyama criterion [23] is used in the present study. In this way, there is a critical value for the Niyama criterion so that regions of casting with lower Niyama values are suspectable to include the solidification induced defects [23]. It is important to note that the Niyama criterion at each point should be evaluated at the local freezing time. Therefore, there are an infinite number of space-time pointwise constraints in our model in the present study. Applying the Niyama criterion to predict the solidification induced defects, we have to make the following assumptions here:

(H1) The mold cavity is filled with the molten metal and the initial temperature distribution is known (it is a common assumption in the solidification simulation, e.g. see [33]).
(H2) The solidification interval, the difference between the liquidus and solidus temperatures, is sufficiently small. In fact we have a narrow-bound freezing which possesses a nearly planar macroscopic solid-liquid interface.
(H3) The effect of gravity is ignored.

Without loss of generality, we made the following assumptions to reduce the model complexity and computational cost of numerical simulation:

(H4) The effect of alloying elements segregation during solidification is ignored.
(H5) The effect of air-gap formation[2] during the solidification is ignored. It is a reasonable assumption particularly in the case of sand mold casting process.

[2] In practice air-gap forms at the cast-mold interface due to the simultaneous contraction of the solidified skin and expansion of the mold walls.

(H6) The interfacial heat transfer coefficient of the cast-mold interface is computed by the harmonic averaging of the thermal heat conductivity of cast and mold materials. This is an appropriate assumption in the case of sand mold casting.

(H7) The latent heat of solidification is linearly distributed within the freezing interval. This is a reasonable assumption almost for every alloying system, in particular whenever there is no explicit relation between the solid fraction and temperature (c.f. [14]).

In addition to the above mentioned constraints, some geometric constraints should be taken into account to ensure the moldability of final design (c.f. [13]). However, the moldability constraints are not considered in the present study for the sake of convenience. It is worth mentioning that in the case of full mold or evaporative pattern casting, we do not face the geometric moldability constraints.

3. Mathematical modeling

Ignoring the effect of melt flow during solidification, the temperature history of casting can be modeled by the following nonlinear heat transfer equation (see [14]):

$$\begin{cases} \rho_c(\theta)c_c(\theta)\,\frac{\partial\theta}{\partial t} & = & \nabla\cdot(k_c(\theta)\nabla\theta)+\rho_c(\theta)L\frac{\partial f_s}{\partial t} & \texttt{in}\ \mathcal{T}\times\mathcal{D}_m \\ \rho_m(\theta)c_m(\theta)\,\frac{\partial\theta}{\partial t} & = & \nabla\cdot(k_m(\theta)\nabla\theta) & \texttt{in}\ \mathcal{T}\times\mathcal{D}\setminus\mathcal{D}_m \\ \theta(t,\mathbf{x}) & = & \theta_0(\mathbf{x}) & \texttt{in}\ \{t=0\}\times\mathcal{D} \\ -k_m\nabla\theta\cdot\mathbf{n} & = & h_\infty(\theta-\theta_\infty) & \texttt{on}\ \mathcal{T}\times\partial\mathcal{D} \\ k_i\nabla\theta\cdot\mathbf{n}_i\mid_{cast} & = & -k_i\nabla\theta\cdot\mathbf{n}_i\mid_{mold} & \texttt{on}\ \mathcal{T}\times\partial\mathcal{D}_m \end{cases} \tag{3.1}$$

where $\mathcal{T}:=(0,T]$ denotes the temporal domain (T is sufficiently large to capture the whole dynamics of solidification), $\mathcal{D}_m\subset\mathcal{D}$ is a portion of the mold-box which is occupied by the metal phase, i.e., the casting and corresponding feeding system, θ denotes the temperature, ρ denotes the density, c denotes the specific heat capacity, k denotes the thermal conductivity, L denotes the fusion latent heat, f_s denotes the local solid fraction, θ_0 denotes the initial temperature distribution, h_∞ denotes the air-mold interfacial heat transfer coefficient, θ_∞ denotes the ambient temperature, $\mathbf{n}$ denotes outer unit normal on $\partial\mathcal{D}$, k_i denotes the cast-mold interfacial heat transfer coefficient, $\mathbf{n}_i$ denotes unit normal on $\partial\mathcal{D}_m$ directed toward the mold and subscripts $(\cdot)_c$ and $(\cdot)_m$ denote the metal and mold materials respectively. As it was explained before (assumption H6), k_i is computed by the harmonic averaging as follows:

$$k_i^{-1}=\frac{1}{2}\left(k_c^{-1}+k_m^{-1}\right) \tag{3.2}$$

It is worth mentioning that, the heat flux is not continuous on the cast-mold interface in general. However, H6 is a reasonable assumption in practice due to small heat conductivity coefficient of sand material. Moreover, we have to employ this assumption in this study to homogenize the governing equations for the purpose of topology optimization. The choice of harmonic mean, instead of arithmetic mean, is based on Voller work [43] in which it was shown that for rapidly changing heat conductivity coefficient, the harmonic averaging leads to more accurate results. In this way the cast and mold domains are distinguished by the mathematical model implicitly through the rapid change in physical properties.

It is assumed that at $t=0$, the mold temperature is equal to the ambient temperature and the molten metal temperature is equal to the pouring temperature,

θ_p, i.e.,

$$\theta_0(\mathbf{x}) = \begin{cases} \theta_p, & \texttt{in} \quad \mathcal{D}_m \\ \theta_\infty, & \texttt{elsewhere} \end{cases} \tag{3.3}$$

Without lose of generality, in order to improve the computational performance, it is assumed that the physical properties are temperature invariant. This assumption simplifies, (3.1) into the following form:

$$\begin{cases} \rho_c c_c \frac{\partial \theta}{\partial t} & = & \nabla \cdot (k_c \nabla \theta) + \rho_c L \frac{\partial f_s}{\partial t} & \texttt{in } \mathcal{T} \times \mathcal{D}_m \\ \rho_m c_m \frac{\partial \theta}{\partial t} & = & \nabla \cdot (k_m \nabla \theta) & \texttt{in } \mathcal{T} \times \mathcal{D} \setminus \mathcal{D}_m \\ \theta(t, \mathbf{x}) & = & \theta_0(\mathbf{x}) & \texttt{in } \{t = 0\} \times \mathcal{D} \\ -k_m \nabla \theta \cdot \mathbf{n} & = & h_\infty (\theta - \theta_\infty) & \texttt{on } \mathcal{T} \times \partial \mathcal{D} \\ k_i \nabla \theta \cdot \mathbf{n}_i \mid_{cast} & = & -k_i \nabla \theta \cdot \mathbf{n}_i \mid_{mold} & \texttt{on } \mathcal{T} \times \Gamma_i \end{cases} \tag{3.4}$$

Assume that the solidus and liquidus temperatures are denoted by θ_s and θ_l respectively. Using assumption (H7) in the previous section results:

$$f_s(\theta) = \begin{cases} 0, & \theta > \theta_l \\ (\theta_l - \theta)/(\theta_l - \theta_s), & \theta_s \leqslant \theta \leqslant \theta_l \\ 1, & \theta < \theta_s \end{cases} \tag{3.5}$$

therefore,

$$\frac{\partial f_s(\theta)}{\partial \theta} = \begin{cases} 0, & \theta > \theta_l \\ (\theta_s - \theta_l)^{-1}, & \theta_s \leqslant \theta \leqslant \theta_l \\ 0, & \theta < \theta_s \end{cases} \tag{3.6}$$

Considering (3.6), $f_s(\theta)$ is not differentiable at $\theta = \theta_s$ and $\theta = \theta_l$. This issue is temporary ignored here. It will be fixed later in section 4. According to chain rule we have: $\partial f_s/\partial t = (\partial f_s/\partial T)(\partial \theta/\partial t)$. Therefore, considering (3.4) together with (3.6) results:

$$\begin{cases} \rho_c c_e(\theta) \frac{\partial \theta}{\partial t} & = & \nabla \cdot (k_c \nabla \theta) & \texttt{in } \mathcal{T} \times \mathcal{D}_m \\ \rho_m c_m \frac{\partial \theta}{\partial t} & = & \nabla \cdot (k_m \nabla \theta) & \texttt{in } \mathcal{T} \times \mathcal{D} \setminus \mathcal{D}_m \\ \theta(t, \mathbf{x}) & = & \theta_0(\mathbf{x}) & \texttt{in } \{t = 0\} \times \mathcal{D} \\ -k_m \nabla \theta \cdot \mathbf{n} & = & h_\infty (\theta - \theta_\infty) & \texttt{on } \mathcal{T} \times \partial \mathcal{D} \\ k_i \nabla \theta \cdot \mathbf{n}_i \mid_{cast} & = & -k_i \nabla \theta \cdot \mathbf{n}_i \mid_{mold} & \texttt{on } \mathcal{T} \times \Gamma_i \end{cases} \tag{3.7}$$

where the effective specific heat capacity, denoted by c_e, is computed as follows:

$$c_e(\theta) = \begin{cases} c_c, & \theta > \theta_l \\ c_c + L(\theta_l - \theta_s)^{-1}, & \theta_s < \theta < \theta_l \\ c_c, & \theta < \theta_s \end{cases} \tag{3.8}$$

The topology indicator function, χ, inside $\mathcal{D}_d$ is the design parameter in this study. In fact, this function is the characteristic function of the metal phase within $\mathcal{D}$, i.e.:

$$\chi(\mathbf{x}) := \begin{cases} 1, & \texttt{in} \quad \mathcal{D}_m \\ 0, & \texttt{elsewhere} \end{cases} \tag{3.9}$$

Since function $\chi(\mathbf{x})$ varies during our optimization, the solution domain in (3.7) should be varied accordingly. Using (3.9), we can rewrite (3.7) in the following form:

$$\begin{cases} \rho(\chi)c(\chi)\ \frac{\partial\theta}{\partial t} & = & \nabla\cdot(k(\chi)\nabla\theta) & \texttt{in } Q \\ \rho(\chi) & = & \chi\rho_c+(1-\chi)\rho_m & \texttt{in } Q \\ c(\chi) & = & \chi c_e+(1-\chi)c_m & \texttt{in } Q \\ k(\chi)^{-1} & = & \chi k_c^{-1}+(1-\chi)k_m^{-1} & \texttt{in } Q \\ \theta(t,\mathbf{x}) & = & \chi\theta_p+(1-\chi)\theta_\infty & \texttt{in } Q_0 \\ k(\chi)\nabla\theta\cdot\mathbf{n} & = & h_\infty(\theta_\infty-\theta) & \texttt{on } \Sigma \end{cases} \tag{3.10}$$

where $Q := \mathcal{T}\times\mathcal{D}$, $Q_0 := \{t=0\}\times\mathcal{D}$, $\Sigma := \mathcal{T}\times\partial\mathcal{D}$. The characteristic function χ is computed by the linear interpolation on the cast-mold interface. It is easy to check that for a fixed topology, (3.10) is equivalent to (3.7).

The objective function is defined as the scaled total volume of the molten metal used in the design domain, i.e.,

$$\arg\min_{\chi\in\{0,1\}} J(\chi) = V_d^{-1}\int_{\mathcal{D}_d}\chi(\mathbf{x})\ d\Omega \tag{3.11}$$

The pointwise value of the Niyama criterion at the local freezing time is constrained in $\mathcal{D}_c$ to ensure the production of a defect-free casting. These constraints can be expressed in the following form (see [23, 28] for further details about the Niyama criterion),

$$g\big(t=t_s(\mathbf{x}),\mathbf{x},\theta\big)\geqslant 0, \qquad \forall\ \mathbf{x}\in\mathcal{D}_c \tag{3.12}$$

where $g(t,\mathbf{x},\theta)=\left(-g_c(\mathbf{x})+|\nabla\theta(t,\mathbf{x})|/\sqrt{\frac{\partial\theta(t,\mathbf{x})}{\partial t}}\right)$, $t_s(\mathbf{x})$ is the local freezing time and $g_c(\mathbf{x})$ is the critical Niyama value. The local freezing time is defined as the time in which the temperature reaches to θ_s, i.e.,

$$t_s(\mathbf{x}) = t\big(\theta(\mathbf{x})=\theta_s\big) \tag{3.13}$$

The user defined function $g_c(\mathbf{x})$ can be fixed to a constant value or varies spatially (a fixed value is used in our numerical experiments in this study). In the later case one can use lower critical values at less important regions of casting and higher values at important points. However, in general there is a critical value, specific property for each alloy, that ensures a defect-free solidification. It is worth mentioning that t_s varies spatially and it is determined in the course of simulation. Therefore, at each spatial point $\mathbf{x}$ (3.12) should be satisfied locally in the temporal domain. These issues make it very difficult to directly mange deal with (3.12) directly.

Consider the Dirac delta function, $\delta^{\mathcal{D}}(\cdot)$,

$$\delta^{\mathcal{D}}(x) := \begin{cases} \infty, & x=0 \\ 0, & x\neq 0 \end{cases}, \quad \int_{\mathbb{R}}\delta^{\mathcal{D}}(x)\ dx = 1, \quad \int_{\mathbb{R}} f(x)\ \delta^{\mathcal{D}}(x)\ dx = f(0)$$

where $x\in\mathbb{R}$ and $f:\mathbb{R}\to\mathbb{R}$ is an arbitrary well-defined continuous function. Using the properties of Dirac delta function, constraint (3.12) can be extended to whole temporal domain as follows,

$$g_e(t_s,\mathbf{x},\theta) = g(t,\mathbf{x},\theta)\ \delta^{\mathcal{D}}(t-t_s)\geqslant 0, \quad \forall\ (t,\mathbf{x})\in\mathbb{R}\times\mathcal{D}_c \tag{3.14}$$

where g_e denotes the temporal extension of g to $\mathbb{R}$. Since $t_s\in\mathcal{T}$, (3.14) can be simplified to:

$$g_e(t_s,\mathbf{x},\theta) = g(t,\mathbf{x},\theta)\ \delta^{\mathcal{D}}(t-t_s)\geqslant 0, \quad \forall\ (t,\mathbf{x})\in Q_c \tag{3.15}$$

where $Q_c := \mathcal{T} \times \mathcal{D}_c$. Since t_s is an implicit function of θ, the appearance of t_s in (3.15) increases the complexity of our analysis. Using the implicit relation between the local freezing time and the solidus temperature, c.f. (3.13), the following equality is identical,

$$\delta^{\mathcal{D}}\big(t - t_s(\theta(t,\mathbf{x}))\big) = \delta^{\mathcal{D}}\big(\theta(t,\mathbf{x}) - \theta_s\big) \quad \forall\ (t,\mathbf{x}) \in Q_c \tag{3.16}$$

Thus, t_s can be removed from (3.15) by the substitution of (3.16) into (3.15),

$$g_e(\mathbf{x},\theta) = g(t,\mathbf{x},\theta)\ \delta^{\mathcal{D}}(\theta - \theta_s) \geqslant 0, \quad \forall\ (t,\mathbf{x}) \in Q_c \tag{3.17}$$

Finally the mathematical formulation of the optimal design problem in the present paper can be expressed as follows,

$$(\mathtt{P}) := \left\{ \begin{array}{rlll} \arg\min_{\chi} & J(\chi) = V_d^{-1}\ \int_{\mathcal{D}_d} \chi(\mathbf{x})\ d\Omega, & \texttt{subject to:} & \\ & & & \\ \rho(\chi)c(\chi)\ \frac{\partial\theta}{\partial t} = & \nabla\cdot(k(\chi)\nabla\theta) & \texttt{in} & Q \\ \rho(\chi) = & \chi\rho_c + (1-\chi)\rho_m & \texttt{in} & Q \\ c(\chi) = & \chi c_e + (1-\chi)c_m & \texttt{in} & Q \\ k(\chi)^{-1} = & \chi k_c^{-1} + (1-\chi)k_m^{-1} & \texttt{in} & Q \\ \theta(t,\mathbf{x}) = & \chi\theta_p + (1-\chi)\theta_\infty & \texttt{in} & Q_0 \\ k(\chi)\nabla\theta\cdot\mathbf{n} = & h_\infty(\theta_\infty - \theta) & \texttt{on} & \Sigma \\ g_e(\mathbf{x},\theta) \geqslant & 0 & \texttt{in} & Q_c \\ \chi \in & \mathcal{Y}_\chi & & \end{array} \right.$$

where $\mathcal{Y}_\chi$ denotes the admissible domain of control variable that is defined as follows,

$$\mathcal{Y}_\chi := \left\{ \chi \in X_\chi(\mathcal{D}) \ \middle| \ \begin{array}{rlll} \chi(\mathbf{x}) = & 0 & \texttt{on} & \partial\mathcal{D} \\ \chi(\mathbf{x}) = & 1 & \texttt{in} & \mathcal{D}_c \\ \chi(\mathbf{x}) = & 0 & \texttt{in} & \mathcal{D}_r \setminus \mathcal{D}_d \\ \chi(\mathbf{x}) \in & \{0,1\} & \texttt{in} & \mathcal{D}_d \\ \int_{\mathcal{D}_{dn}} \chi\ d\Omega \leqslant & R_{dn}V_{dn} & & \end{array} \right\}.$$

The boundary condition $\chi = 0$ on the external boundaries of the mold box is applied here to improve the convergence-rate of numerical solution.

Remark 3.1. It should be noticed that the structure of topology optimization problem (P) is different from the classical topology optimization problems in the sense that the objective functional (or target) space is not identical to the control space. More precisely, in the classical topology optimization problems the optimal material distribution is found within a spatial domain on which the objective functional and constraints are defined. However, in the topology optimization problem (P), the optimal material distribution is found within $\mathcal{D}_d$ to pose control inside $\mathcal{D}_c$. In fact, a new class of topology optimization problems is introduced in the present study. This contribution extends the utility of topology optimization approach to solve new classes of engineering design problems.

4. Regularization of the mathematical model

As it was mentioned in section 3, $f_s(\theta)$ in equation (3.5) (and so $c_e(\theta)$) is not differentiable at $\theta = \theta_s$ and $\theta = \theta_l$ (in fact this function is not analytic at these points).

Therefore, the nonlinear heat equation in optimal problem (P) is non-smooth [3]. An smoothing approach is presented in this section to resolve this problem. We can express (3.5) in the following equivalent form,

$$f_s(\theta) = H(\theta_s - \theta) + \Big(H(\theta_l - \theta) - H(\theta_s - \theta)\Big)\,\frac{\theta_l - \theta}{\theta_l - \theta_s} \tag{4.1}$$

where $H(x) : \mathbb{R} \to \mathbb{R}$ denotes the one-dimensional Heaviside function which is defined as follows (c.f. chapter 1 of [24]):

$$H(x) := \begin{cases} 0, & x < 0 \\ 1, & x \geqslant 0 \end{cases} \tag{4.2}$$

Following [24], the smoothed form of the one-dimensional Heaviside function can be expressed as follows,

$$\tilde{H}_\epsilon(x) = \begin{cases} 0, & x < -\epsilon \\ \frac{1}{2} + \frac{x}{2\epsilon} + \frac{1}{2\pi}\sin\left(\frac{\pi x}{\epsilon}\right), & -\epsilon \leqslant x \leqslant \epsilon \\ 1, & x > \epsilon \end{cases} \tag{4.3}$$

where $\epsilon \in (0, \infty)$ is the smoothing parameter in the sense that: $H(x) = \lim_{\epsilon \to 0} \tilde{H}_\epsilon(x)$. It is clear that $\tilde{H}$ is a $C^2(\mathbb{R})$ function with the first order derivative of:

$$\tilde{H}'_\epsilon(x) = \begin{cases} 0, & x < -\epsilon \\ \frac{1}{2\epsilon} + \frac{1}{2\epsilon}\cos\left(\frac{\pi x}{\epsilon}\right), & -\epsilon \leqslant x \leqslant \epsilon \\ 0, & x > \epsilon \end{cases} \tag{4.4}$$

Using the smoothed one-dimensional Heaviside function (4.1) can be rewritten as follows:

$$\tilde{f}_s(\theta) = \tilde{H}_\epsilon(\theta_s - \theta) + \Big(\tilde{H}_\epsilon(\theta_l - \theta) - \tilde{H}_\epsilon(\theta_s - \theta)\Big)\,\frac{\theta_l - \theta}{\theta_l - \theta_s} \tag{4.5}$$

where $\tilde{f}_s$ denotes the regularized solid fraction function. Similarly, the regularized effective specific heat capacity function, $\tilde{c}_e$, can be expressed as follows:

$$\tilde{c}_e = c_c + L\,\frac{\tilde{f}_s(\theta)}{\partial\theta} \tag{4.6}$$

The graph of c_e and $\tilde{c}_e$ functions are schematically shown in figure 3. Note that we do not solve the ϵ-decreasing sequence of problems in this study, but solve one problem with a sufficiently small smoothing parameter. The smoothing parameter ϵ is equal to $0.1(\theta_l - \theta_s)$ in this study.

The existence of Dirac delta function in space-time constraints (3.17) leads to the complexity of optimization problem in both aspects of the sensitivity analysis and numerical solution. To overcome these issues, the Dirac delta function is regularized in this study. The regularized Dirac delta function, $\delta_a^{\mathcal{D}}(x) : \mathbb{R} \to \mathbb{R}$, is a sufficiently smooth function such that (see [1]):

$$\int_{\mathbb{R}} \delta_a^{\mathcal{D}}(x)\; dx = 1, \quad \delta^{\mathcal{D}}(x) = \lim_{a \to 0} \delta_a^{\mathcal{D}}(x)$$

where $a \in (0, \infty)$ is the regularization parameter that controls the Dirac delta function smoothing. The bell-shaped probability density function is used as the

[3] This differentiability issue is not well considered in the solidification modeling literature. It is worth mentioning that the integral formulation based on the enthalpy approach is used in some literature. However, this problem exists in this case too.

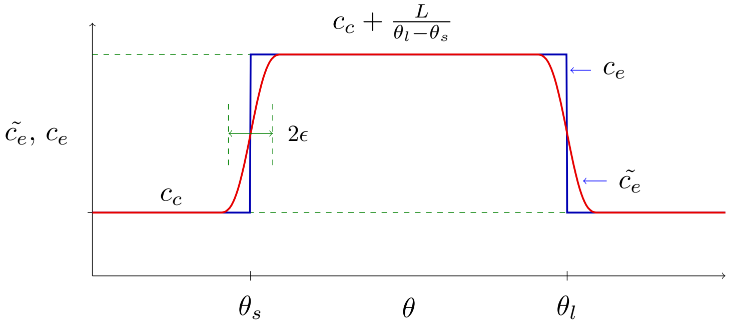


FIGURE 3. Schematic plots of c_e and $\tilde{c}_e$ functions in the present study. The regularization parameter ϵ is equal to $0.1(\theta_l - \theta_s)$ here.

regularized counterpart of the Dirac delta function in the present work. Therefore, by $\delta_a^{\mathcal{D}}(x)$ we denote the following function henceforth:

$$\delta_a^{\mathcal{D}}(x) = \frac{1}{a\sqrt{2\pi}}\ e^{-(x)^2/(2a^2)} \tag{4.7}$$

In fact, the regularized function (4.7) distributes the pointwise concentration of $\delta^{\mathcal{D}}(x)$ on neighborhood of x. For instance, for $a = 1$, more than 99 % of the function concentration will be distributed smoothly over interval $(-3, 3)$, see figure 4.

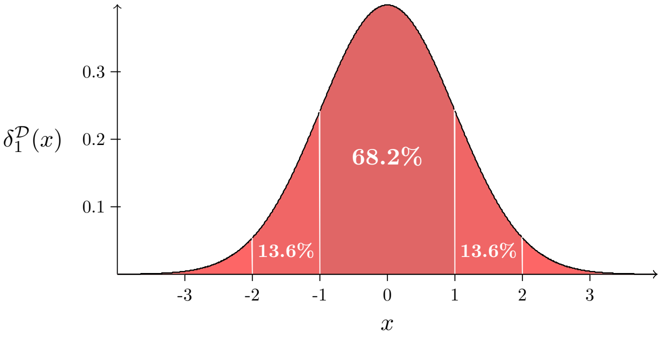


FIGURE 4. Schematic of the regularized Dirac delta function, $\delta_a^{\mathcal{D}}$, used in this study for $a = 1$.

The regularized form of (3.17), $\tilde{g}_e$, in the present study has the following form:

$$\tilde{g}_e(a, \mathbf{x}, \theta) = g(t, \mathbf{x}, \theta)\ \delta_a^{\mathcal{D}}(a_t\theta + b_t) \geqslant 0, \quad \forall\ (t, \mathbf{x}) \in Q_c \tag{4.8}$$

where $a = 1$ in the present study, $a_t = 6(\theta_l - \theta_s)^{-1}$ and $b_t = -3(\theta_l + \theta_s)(\theta_l - \theta_s)^{-1}$. In this way, the effect of delta concentration at every spatial point will be approximately distributed within the temperature interval $[\theta_l - \theta_s]$ centered at $\theta = \frac{1}{2}(\theta_l + \theta_s)$.

The regularized from of $(\mathtt{P})$, denoted by $(\tilde{\mathtt{P}})$, can be stated as follows:

$$(\tilde{\mathtt{P}})(\epsilon,a) := \left\{ \begin{array}{rlll} \arg\min_\chi & J(\chi) = V_d^{-1} \int_{\mathcal{D}_d} \chi(\mathbf{x}) \; d\Omega, & \texttt{subject to:} & \\ & & & \\ \rho(\chi)\tilde{c}(\chi)\, \frac{\partial\theta}{\partial t} = & \nabla\cdot(k(\chi)\nabla\theta) & \texttt{in} & Q \\ \rho(\chi) = & \chi\rho_c + (1-\chi)\rho_m & \texttt{in} & Q \\ \tilde{c}(\chi) = & \chi\tilde{c}_e + (1-\chi)c_m & \texttt{in} & Q \\ k(\chi)^{-1} = & \chi k_c^{-1} + (1-\chi)k_m^{-1} & \texttt{in} & Q \\ \theta(t,\mathbf{x}) = & \chi\theta_p + (1-\chi)\theta_\infty & \texttt{in} & Q_0 \\ k(\chi)\nabla\theta\cdot\mathbf{n} = & h_\infty(\theta_\infty-\theta) & \texttt{on} & \Sigma \\ \tilde{g}_e(a,\mathbf{x},\theta) \geqslant & 0 & \texttt{in} & Q_c \\ \chi \in & \mathcal{Y}_\chi & & \end{array} \right.$$

5. Relaxation of the mathematical model

In general topology optimization problems, like $(\tilde{\mathtt{P}})$, are ill-posed variational problems. Thus, the do not admit optimal solutions in their original form.The relaxation of original problem (c.f. [18]) is a common approach to deal with this difficulty. In this way, the design space is extended from the characteristic functions $\chi \in BV(\mathcal{D};\{0,1\})$ to continuous functions $w \in L^\infty(\mathcal{D};[0,1])$. To find an explicit relaxation of $(\tilde{\mathtt{P}})$ is a very challenging job. Moreover, it is not clear that whether there exists an explicit form for relaxation of $(\tilde{\mathtt{P}})$. Because the mathematical analysis of $(\tilde{\mathtt{P}})$ is beyond the scope of present work, inspiring from [6, 22], we assume that an explicit relaxation of $(\tilde{\mathtt{P}})$, denoted by $(\mathtt{RP})$, can be stated as follow:

$$(\mathtt{RP})(\epsilon,a) := \left\{ \begin{array}{rlll} \arg\min_w \;\; J(w) = & V_d^{-1} \int_{\mathcal{D}_d} w(\mathbf{x}) \; d\Omega, & \texttt{subject to:} & \\ & & & \\ \rho(w)\tilde{c}(w)\, \frac{\partial\theta}{\partial t} = & \nabla\cdot(k(w)\nabla\theta) & \texttt{in} & Q \\ \rho(w) = & w\rho_c + (1-w)\rho_m & \texttt{in} & Q \\ \tilde{c}(w) = & w\tilde{c}_e + (1-w)c_m & \texttt{in} & Q \\ k(w)^{-1} = & wk_c^{-1} + (1-w)k_m^{-1} & \texttt{in} & Q \\ \theta(t,\mathbf{x}) = & w\theta_p + (1-w)\theta_\infty & \texttt{in} & Q_0 \\ k(w)\nabla\theta\cdot\mathbf{n} = & h_\infty(\theta_\infty-\theta) & \texttt{on} & \Sigma \\ \tilde{g}_e(a,\mathbf{x},\theta) \geqslant & 0 & \texttt{in} & Q_c \\ w \in & \mathcal{Y}_w & & \end{array} \right.$$

where $\mathcal{Y}_w$ denotes the admissible set of control parameters, defined as follows,

$$\mathcal{Y}_w := \left\{ w \in X_w(\mathcal{D}) \;\middle|\; \begin{array}{rlll} w(\mathbf{x}) = & 0 & \texttt{on} & \partial\mathcal{D} \\ w(\mathbf{x}) = & 1 & \texttt{in} & \mathcal{D}_c \\ w(\mathbf{x}) = & 0 & \texttt{in} & \mathcal{D}_r \setminus \mathcal{D}_d \\ 0 \leqslant w(\mathbf{x}) \leqslant & 1 & \texttt{in} & \mathcal{D}_d \\ \int_{\mathcal{D}_{dn}} w \; d\Omega \leqslant & R_{dn}V_{dn} & & \end{array} \right\}.$$

where X_w is a sufficiently regular Banach space corresponding to the control variable w. In fact, the characteristic function χ in $(\tilde{\mathtt{P}})$ is replaced by the more regular function w in $(\mathtt{RP})$. However, there is no guarantee that the relaxed problem will be well-posed and admits at least an optimal solution. The ill-posness issue in topology optimizations stems in this fact that the optimal solutions tend to form finer and finer microstructures. The minimum length-scale control is a common way to cope this problem.

6. SIMP penalization of mathematical model

From a practical point of view, the numerical solution of (RP) is not desirable, because it tolerates the existence of graded material in the optimal solution. Several treatments are suggested in the literature to overcome this difficulty (c.f. [7]). The solid isotropic material penalization (SIMP) method [7] gained more popularity in the literature due to the ease of implementation, satisfactory results and low computational cost. The SIMP penalization of (RP), denoted by (SRP), can be expressed as follows:

$$(\texttt{SRP})(\epsilon, a) := \left\{ \begin{array}{rlll} \arg\min_w \;\; J(w) = & V_d^{-1} \; \int_{\mathcal{D}_d} w(\mathbf{x}) \; d\Omega, & \texttt{subject to:} & \\ & & & \\ \rho(w)\tilde{c}(w) \; \frac{\partial \theta}{\partial t} = & \nabla \cdot (k(w) \nabla \theta) & \texttt{in} & Q \\ \rho(w) = & w^p \rho_c + (1 - w^p)\rho_m & \texttt{in} & Q \\ \tilde{c}(w) = & w^p \tilde{c}_e + (1 - w^p) c_m & \texttt{in} & Q \\ k(w)^{-1} = & w^p k_c^{-1} + (1 - w^p) k_m^{-1} & \texttt{in} & Q \\ \theta(t, \mathbf{x}) = & w^p \theta_p + (1 - w^p)\theta_\infty & \texttt{in} & Q_0 \\ k(w)\nabla\theta \cdot \mathbf{n} = & h_\infty(\theta_\infty - \theta) & \texttt{on} & \Sigma \\ \tilde{g}_e(a, \mathbf{x}, \theta) \geqslant & 0 & \texttt{in} & Q_c \\ w \in & \mathcal{Y}_w & & \end{array} \right.$$

where $p \in [1, \infty)$ denotes the SIMP power. This parameter is gradually increased from 1 to 5 during solution of each sub-problem in the present study. In practice, $p > 2$ makes a strong bias on w to assume a value near either of 0 or 1.

7. Bilevel reformulation of the optimization problem

The goal of present study is to introduce an efficient solution algorithm for real-world problems. It can not be realized by using conventional (black-box) optimization tools. An efficient solution strategy will be introduced in this section by carefully exploiting the specific structure of the (SRP).

There are generally two approaches to solve PDE constrained optimization problems: "discretize then optimize" and "optimize then discretize". In the former method the state and control spaces are firstly discretized into a finite dimensional spaces. Then, the finite dimensional counterpart of original optimization problem is solved by convectional optimization algorithms. From a practical point of view, the ease of implementation is the main advantage of this approach. Assuming a gradient based method is employed to solve the discretized optimization problem, the cost of sensitivity analysis is very large. This is due to the fact that the direct design sensitivity analysis is economic when the number of (state-dependent) constraints is large. On the other hand, the adjoint sensitivity analysis is preferable when the number of control parameters is large (c.f. [39]). However, when both of these parameters are large, as it is the case in our problem, the cost of sensitivity analysis scales with either of these parameters. This issue makes the numerical solution computationally intractable (c.f. [19]). In the later approach, optimize then discretize, the optimization is performed formerly on an infinite-dimensional function spaces and then, the corresponding system of optimality conditions will be discretized for the purpose of numerical solution. This approach has a good potential to cope the above mentioned limitations. Therefore, we shall develop our solution strategy based on the later approach. Doing a rigorous analysis on (SRP) beyond the scope of this paper (see [41, 42] for further details). Hence, only the solution strategy will be introduced here without involving in technical details (the reliability of present approach is supported by extensive numerical experiments given in section 10).

To exploit the specific structure of (SRP), the constraints in (SRP) are divided into three categories as follows: control constraints, PDE-constraint and pointwise state-dependent constraints. The admissible set of control parameters is denoted by $\mathcal{Y}_w$ here, as it is defined in section 5. The admissible set of state variable θ with respect to PDE-constraint in (SRP), denoted by $\mathcal{Y}_\theta$, is defined as follows:

$$\mathcal{Y}_\theta := \Big\{\theta \in X_\theta(\mathcal{D}) \;\Big|\; F(t,\theta,\dot{\theta},\nabla\theta,\nabla^2\theta,w) = 0,\ \forall\ (t,\mathbf{x}) \in Q \ \texttt{and}\ w \in [0,1]\Big\}$$

where X_θ is a sufficiently regular Banach space corresponding to the state variable and the operator $F(t,\theta,\dot{\theta},\nabla\theta,\nabla^2\theta,w) = 0$ denotes the pointwise feasibility of the heat equation in (SRP). The pointwise constrains $\tilde{g}_e$ in (SRP) can be equivalently expressed in the following form:

$$\min\big(\tilde{g}_e(a,\mathbf{x},\theta),0\big) = 0, \quad \forall\ (t,\mathbf{x}) \in Q_c \tag{7.1}$$

The infeasibility measure of these constraints can be defined as follows:

$$G(\theta) := \frac{1}{2}\int_{Q_c} \big(\min\big(\tilde{g}_e(a,\mathbf{x},\theta),0\big)\big)^2 d\mathcal{Q} \tag{7.2}$$

where $d\mathcal{Q} := d\Omega dt$ (a is constant and equal to 1 in the present study). When $G(\theta) = 0$, constraints (7.1) hold almost everywhere in Q_c. The admissible space of state variable θ with respect to the pointwise state-dependent constraints, denoted by $\mathcal{Y}_\mathtt{g}$, can be defined as follows,

$$\mathcal{Y}_\mathtt{g} := \{\ \theta \in X_\theta(\mathcal{D}) \mid G(\theta) = 0\ \}$$

Using the above definition, (SRP) can be expressed in the following abstract form:

$$(\texttt{SRP})(\epsilon,a) :=\ \arg\min_w\ J(w) \quad \texttt{subject to} \quad \big(w,\theta(w)\big)\ \in \big(\mathcal{Y}_w \times (\mathcal{Y}_\theta \cap \mathcal{Y}_\mathtt{g})\big)$$

In our solution strategy, the feasibility of control parameters are strictly hold with respect to set $\mathcal{Y}_w$. Considering the bound constraints on w, we have $J(w) : \mathcal{Y}_w \to [0,1]$. Therefore, for every $w \in \mathcal{Y}_w$, $J(w)$ assumes the fixed value $R_d \in \mathbb{R}[0,1]$. Thus, w is always a member of the following admissible set:

$$\mathcal{A} := \Big\{w \in \mathcal{Y}_w \;\Big|\; V_d^{-1}\int_{\mathcal{D}_d} w(\mathbf{x})\ d\Omega = R_d\Big\}$$

Assuming (SRP) admits at least an optimal solution, (SRP) can be reformulated as the following bilevel optimization problem:

$$(\texttt{BP})(a,\epsilon) := \quad \texttt{minimize}\ R_d \quad \texttt{over}\ [0,1] \quad \texttt{subject to} \quad \big(w,\theta(w)\big) \in \big(\mathcal{A}\times(\mathcal{Y}_\theta\cap\mathcal{Y}_\mathtt{g})\big)$$

Proposition 7.1. *The set of optimal solutions to problems* (SRP) *and* (BP) *is identical.*

Proof. Comparing (BP) and (SRP), the proof is evident. □

The lower level problem in (BP) is in fact a feasibility problem. For the purpose of convenience, lets to denote this feasibility problem symbolically as follows:

$$(\texttt{IP})(\epsilon,a,R_d) := \big(w,\theta(w)\big) \in (\mathcal{A} \times (\mathcal{Y}_\theta \cap \mathcal{Y}_\mathtt{g}))$$

Notice that the feasibility problem $(\mathtt{IP})$ does not essentially admit a solution for every $R_d \in [0,1)$ even if the feasible set of problem $(\mathtt{SRP})$ be nonempty. Assume R_d^* denotes the optimal value of R_d at an optimal solution of $(\mathtt{BP})$. The following proposition provides a guideline to restrict the search interval of R_d^*.

Proposition 7.2. *Assume that problem* $(\mathtt{SRP})$ *has at least an optimal solution, then there exist an upper bound* $R_u \in [0,1]$ *such that* $R_d^* \in [0, R_u]$ *where,*

$$R_u = V_d^{-1} \int_{\mathcal{D}_d} w_0 \; d\Omega$$

and w_0 *is a solution to the following feasibility problem:*

$$\big(w, \theta(w)\big) \; \in (\mathcal{Y}_w \times (\mathcal{Y}_\theta \cap \mathcal{Y}_{\rm g}))$$

Proof. Since $(\mathtt{SRP})$ has at least an optimal solution, we have $\big(\mathcal{A} \times (\mathcal{Y}_\theta \cap \mathcal{Y}_{\rm g})\big) \neq \emptyset$. Therefore the feasibility problem $(\mathtt{IP})$ admits at least a solution too. Since $\big(\mathcal{A} \times (\mathcal{Y}_\theta \cap \mathcal{Y}_{\rm g})\big) \subseteq \big(\mathcal{Y}_w \times (\mathcal{Y}_\theta \cap \mathcal{Y}_{\rm g})\big)$ we have $\big(\mathcal{Y}_w \times (\mathcal{Y}_\theta \cap \mathcal{Y}_{\rm g})\big) \neq \emptyset$ which ensures the existence of w_0. Since R_d^* is the least value of $R_d \in [0,1]$ such that $(\mathtt{IP})$ admits at least a solution, we have $R_d^* \leqslant R_u$ which completes the proof. □

The following minimization problem, denoted by $(\mathtt{MP})$, is used in the present study to solve the feasibility problem $(\mathtt{IP})$,

$$(\mathtt{MP})(\epsilon, a, R_d) := \left\{ \begin{array}{rlll} \min_w \;\; G(\theta(w)) = & \frac{1}{2}\int_{Q_c} \big(\min\big(\tilde{g}_e(a, \mathbf{x}, \theta(w)), 0\big)\big)^2 d\mathcal{Q}, & \mathtt{s.t.:} & \\ \rho(w)\tilde{c}(w)\, \frac{\partial \theta}{\partial t} = & \nabla \cdot (k(w)\nabla\theta) & \mathtt{in} & Q \\ \rho(w) = & w^p \rho_c + (1-w^p)\rho_m & \mathtt{in} & Q \\ \tilde{c}(w) = & w^p \tilde{c}_e + (1-w^p) c_m & \mathtt{in} & Q \\ k(w)^{-1} = & w^p k_c^{-1} + (1-w^p) k_m^{-1} & \mathtt{in} & Q \\ \theta(t, \mathbf{x}) = & w^p \theta_p + (1-w^p)\theta_\infty & \mathtt{in} & Q_0 \\ k(w)\nabla\theta \cdot \mathbf{n} = & h_\infty(\theta_\infty - \theta) & \mathtt{on} & \Sigma \\ w \in & \mathcal{A} & & \end{array} \right.$$

It is worth mentioning that w_0 can be computed by solving a modified version of $(\mathtt{MP})$ in which the constraint $w \in \mathcal{A}$ is replaced by $w \in \mathcal{Y}_w$. When $(\mathtt{IP})$ has at least a feasible solution then the corresponding $(\mathtt{MP})$ admits at least an optimal solution with the objective functional equal to zero at the optimal solution.

Note that problem $(\mathtt{MP})$ is structurally similar to the classical volume constrained topology optimization problems (c.f. [7]), i.e. it includes a global objective functional, bound and volume constraints on the control variables together with a PDE-constraint. Therefore, available efficient solution algorithms, e.g. [38], can be employed to solve sub-problems $(\mathtt{MP})$. According to our numerical experiments, the computational cost of numerical solution of $(\mathtt{MP})$ is almost equivalent to that of a classical volume constrained topology optimization problem. Another benefit of solving $(\mathtt{MP})$ is the possibility of managing infeasible solutions, which is commented later in this section.

Now, we slightly modify $(\mathtt{MP})$ to simplify the sensitivity analysis. The characteristic function of $\mathcal{D}_c$, denoted by $\mathcal{I}_c$, can be defined as follows,

$$\mathcal{I}_c(\mathbf{x}) := \left\{ \begin{array}{ll} 1, & \mathbf{x} \in \mathcal{D}_c \\ 0, & \mathtt{elsewhere} \end{array} \right. \tag{7.3}$$

$\mathcal{I}_c$ can be alternatively expressed as follows,

$$\mathcal{I}_c(\mathbf{x}) = \int_{\mathcal{D}_c} \delta^{\mathcal{D}}(x)\delta^{\mathcal{D}}(y)\delta^{\mathcal{D}}(z)\ d\Omega \tag{7.4}$$

where $\mathbf{x} = (x, y, z)$. To smooth the characteristic function $\mathcal{I}_c$, we replace the Dirac delta functions in (7.4) by the regularized Dirac delta function introduced in section 4. Therefore, the smoothed characteristic function of $\mathcal{D}_c$, denoted by $\tilde{\mathcal{I}}_c$, has the following form,

$$\tilde{\mathcal{I}}_c(\mathbf{x}) = \int_{\mathcal{D}_c} \delta_b^{\mathcal{D}}(x)\delta_b^{\mathcal{D}}(y)\delta_b^{\mathcal{D}}(z)\ d\Omega \tag{7.5}$$

where $\delta_b^{\mathcal{D}}(\cdot)$ denotes the regularized delta function with the regularization parameter b. Using (7.5), we redefine the objective functional corresponding to $(\texttt{MP})$ as follows,

$$\mathcal{F}(\theta(w)) := 1/2 \int_Q \mathcal{G}(\mathbf{x}, \theta(w))\ d\mathcal{Q}, \quad \mathcal{G}(\mathbf{x}, \theta(w)) = \left(\min\left(\tilde{g}_e(a, \mathbf{x}, \theta(w)), 0\right)\right)^2 \tilde{\mathcal{I}}_c(\mathbf{x}) \tag{7.6}$$

where the spatial integration domain in the original objective functional is extended to the whole spatial domain here. In the present study we fix the regularization parameter b to 1/6 of the spatial grid-size used in our numerical method. In this way, the width of smoothing region is almost equal to one spatial grid-size. Note that evaluation of the discrete version of function $\tilde{\mathcal{I}}_c$ is performed only once in this study (prior to starting the optimization procedure). The modified version of $(\texttt{MP})$, denoted by lower-level problem $(\texttt{LP})$, is defined as follows,

$$(\texttt{LP})(\epsilon, a, b, R_d) := \left\{ \begin{array}{rlll} \arg\min_w\ \mathcal{F}(\theta(w)) = & \frac{1}{2}\int_Q \mathcal{G}(\mathbf{x}, \theta(w))\ d\mathcal{Q}, & \texttt{s.t.}: & \\ \rho(w)\tilde{c}(w)\ \frac{\partial\theta}{\partial t} = & \nabla\cdot(k(w)\nabla\theta) & \texttt{in} & Q \\ \rho(w) = & w^p\rho_c + (1-w^p)\rho_m & \texttt{in} & Q \\ \tilde{c}(w) = & w^p\tilde{c}_e + (1-w^p)c_m & \texttt{in} & Q \\ k(w)^{-1} = & w^p k_c^{-1} + (1-w^p)k_m^{-1} & \texttt{in} & Q \\ \theta(t, \mathbf{x}) = & w^p\theta_p + (1-w^p)\theta_\infty & \texttt{in} & Q_0 \\ k(w)\nabla\theta\cdot\mathbf{n} = & h_\infty(\theta_\infty - \theta) & \texttt{on} & \Sigma \\ w \in & \mathcal{A} & & \end{array} \right.$$

The upper level problem in $(\texttt{BP})$ is equivalent to a simple mono-dimensional global pattern search on the line segment $[0, R_u]$. In the present study, the solution is kept feasible with respect to the control and PDE constraints during the optimization procedure, i.e., $(w, \theta) \in (\mathcal{Y}_w \times \mathcal{Y}_\theta)$. Therefore, the remained task is to find minimum value of R_d, denoted by R_d^*, such that $R_d^* = V_d^{-1}\int_{\mathcal{D}_d} w^*\ d\Omega$, w^* solve $(\texttt{LP})(\epsilon, a, b, R_d^*)$ and $\mathcal{F}(\theta^*(w^*)) = 0$. There are many approaches to solve this simple line search problem. According to our numerical experiments in the present study, $\mathcal{F}$ is continuous and monotonically non-increasing function of R_d in all of our experiments. It is worth mentioning that this observation is not correct in general.

Lets to assume that the objective functional $\mathcal{F}$ is a continuous and monotonically non-increasing function of R_d within interval $[0, R_u]$. In this case, a simple bisection algorithm followed by linear interpolation can employed here to find R_d^* over search interval $[0, R_u]$. This bisection algorithm can be expressed as follows:

Algorithm 1: Upper-level optimization solver: bisection approach

```
initialization : given $n > 1$, $\mathtt{i} = \mathtt{1}$, $R_d^l = 0$, $R_d^r = R_u$, $\mathcal{F}_l = 10^{20}$, $\mathcal{F}_r = 0$;
while ($\mathtt{i} \leqslant \mathtt{n}$) do
    $\mathtt{i} = \mathtt{i} + \mathtt{1}$, $R_d^c = (R_d^l + R_d^r)/2$;
    solve $(\mathtt{LP})(\epsilon, a, b, R_d^c)$ and compute $\mathcal{F}_c^*$;
    if ($\mathcal{F}_c^* == 0$) then $R_d^r = R_d^c$, $\mathcal{F}_r = \mathcal{F}_c^*$;
    else $R_d^l = R_d^c$, $\mathcal{F}_l = \mathcal{F}_c^*$;
end
if ($\mathcal{F}_l == 0$) then $R_d^* = R_d^l$;
else $R_d^* = (\mathcal{F}_r R_d^l - \mathcal{F}_l R_d^r)/(\mathcal{F}_r - \mathcal{F}_l)$;
return $R_d^*$;
```

For $n = 10$, the above algorithm, without linear interpolation, gives R_d^* within error e_r where $|e_r| \leq R_u/2^{10}$. This level of accuracy is quite sufficient for engineering applications. The linear interpolation usually increases this level of accuracy by an order of magnitude. In this way, the optimal solution with sufficient accuracy can be achieved in expense of solving 12 classical volume constrained topology optimization sub-problems. The computational complexity of our solution algorithm does not increase in practice when we do not assume the monotonicity of $\mathcal{F}$ (with respect to R_d). In this case, we apply a simple global optimization method for $\mathcal{F}(R_d)$. Since this problem is one-dimensional, a suitable global optimization algorithm could be exploited very efficiently. For instance, we suggest a few number of equidistance sampling on $[0, R_u]$ to interpolate the graph of $\mathcal{F} - R_d$ on $[0, R_u]$. Then, the least root of $\mathcal{F}$, which is equal to R_d^*, is found approximately by means of interpolated $\mathcal{F} - R_d$ graph, i.e.,

Algorithm 2: Upper-level optimization solver: global interpolation

```
initialization: given $n > 1$, $\Delta = R_u/n$;
for $i = 0, \ldots, n$ solve $(\mathtt{LP})(\epsilon, a, b, R_d^i = i\Delta)$ and compute $\mathcal{F}_i^*$ ;
compute approximate graph $\mathcal{F}^* = \mathcal{C}(R_d)$ by interpolation between $(i\Delta, \mathcal{F}_i^*)$;
take $R_d^*$ equal to the least root of $\mathcal{C}(R_d)$;
```

When problem $(\mathtt{BP})$ is infeasible, i.e. $\big(\mathcal{Y}_w \times (\mathcal{Y}_\theta \cap \mathcal{Y}_{\mathtt{g}})\big) = \emptyset$, it is not possible to solve the original optimization problem by classical optimization methods. In this case, the presented solution strategy, with a minor modification, is enable to manage infeasible problems and to provide valuable information for designers. It is easy to show that $\big(\mathcal{Y}_w \times \mathcal{Y}_\theta\big) \neq \emptyset$. In fact, the infeasibility roots in the violation of pointwise state-dependent constraints. Now lets to define the concept of the best infeasible solutions as follows:

Definition 7.3. Assume that for problem $(\mathtt{BP})$ we have $\big(\mathcal{Y}_w \times (\mathcal{Y}_\theta \cap \mathcal{Y}_{\mathtt{g}})\big) = \emptyset$, then $(w^*, \theta^*) \in \big(\mathcal{Y}_w \times \mathcal{Y}_\theta\big)$ is called a best α-infeasible solution to $(\mathtt{BP})$ if the following condition holds:

$$M_\alpha(w^*, \theta^*) \leqslant M_\alpha(w, \theta) \quad \forall (w, \theta) \in \big(\mathcal{Y}_w \times \mathcal{Y}_\theta\big),$$

where $\alpha \in [0,1]$ is a user-defined (trade-off) parameter and the merit function $M_\alpha(\cdot,\cdot)$ is defined as follows:

$$M_\alpha := \alpha R_d + (1-\alpha)\ \mathcal{F}^*$$

where $\mathcal{F}^*$ denotes the value of the objective functional at the local solution of the sub-problem (LP) for R_d.

In fact, in the case of problem infeasibility, we will have a bi-objective optimization problem in which there is a trade-off between the consumed materials resource weighted by factor α and the infeasibility of state-dependent pointwise constraints weighted by factor $(1-\alpha)$. To find a best α-infeasible solution of (BP), for a user-defined α-value, we solve (LP) for equidistance values of R_d on $[0,1]$ and compute the corresponding merit functions, M, at optimal solutions. Then, the approximate graph of the merit function M is found as a function of R_d by means of a global interpolation using sampled points. Finally, the value of R_d corresponding to an approximate best α-infeasible solution is taken equal to the approximate global minimum of M-R_d graph. This algorithm can be expressed as follows:

Algorithm 3: Upper-level optimization solver for infeasible problems

```
initialization: given $n > 1$, $\alpha \in [0,1]$, $\Delta = 1/n$;
for $i = 0, \dots, n$ solve (LP)$(\epsilon, a, b, R_d^i = i\Delta)$ and compute $M_{\alpha,i}$;
compute approximate graph $M_\alpha = \mathcal{C}(R_d)$ by interpolation between
  $(i\Delta, M_{\alpha,i})$;
take $R_d^*$ equal to the global minimum of $\mathcal{C}(R_d)$;
```

The solution of lower-lever problem, i.e., (LP), is the remaining part of our solution algorithm in this study which will be discussed in section 8.

8. Necessary optimality conditions for the lower-level problem

A deterministic gradient based method is used in the present study to solve problem (LP) in (BP). In this way, it is possible to find a local minimum of (LP) using a suitable optimization algorithm. Since (LP) is non-convex, it possibly admits many local solutions. On the other hand, a local solution my not be a desirable solution in practice. However, local solutions were satisfactory from a practical point of view according to our numerical experiments. The first order necessary optimality conditions corresponding to (LP) will be derived in this section. For background materials, interested readers are encouraged to read chapter 10 of [2].

Definition 8.1. (Gâteaux derivative, c.f. [2]) Consider Banach spaces Y and W, and U as open subset of Y. A function $f : U \to W$ is called Gâteaux differentiable at $u \in U$, if for every test function $v \in Y$ the following limit exist:

$$f'(u) := \lim_{\zeta \to 0} \frac{f(u+\zeta v) - f(u)}{\zeta}$$

In this case we show Gâteaux derivative symbolically by $f'(u)$. If Y is a Hilbert space, then $f'(u)$ lives on the dual space of Y. Thus, by using the Riesz representation theorem, there exists a unique $p \in Y$ such that $\langle p, v\rangle = f'(u)$, where $\langle \cdot,\cdot\rangle$ denotes the inner product on Y. In this case, it is common to call p as Gâteaux derivative. The partial Gâteaux derivative is shown by either $f'_{(\cdot)}$ or $\partial_{(\cdot)} f(\cdots)$ symbols in this study. For the purpose of convenience, Gâteaux derivative is called directional derivative henceforth. Since our goal in this study is not to do a rigorous mathematical analysis, we simply assume that our functions are sufficiently regular. We use notation $\langle \cdot, \cdot \rangle_{\mathcal{A}} := \int_{\mathcal{A}} (\cdot)(\cdot) \; d\Omega$ to denote either of inner product or duality pairing on the corresponding function spaces.

Consider arbitrary functions $\eta \in X_\eta(Q)$, $\eta_b \in X_\eta(\Sigma)$ and $\eta_0 \in X_\eta(Q_0)$, where X_η is a sufficiently regular Banach space (identical to X_θ in practice). Lets to introduce the following lagrangian by augmenting the objective functional in (`LP`) with the corresponding state constraint:

$$\mathcal{L}(w, \theta, \eta) := \mathcal{F}(\theta) + \int_Q \eta \big(\rho \tilde{c} \; \dot{\theta} - \nabla \cdot (k \nabla \theta)\big) \; d\mathcal{Q} + \int_\Sigma \eta_b \big(k \nabla \theta \cdot \mathbf{n} - h_\infty (\theta_\infty - \theta)\big) \; d\Sigma$$
$$+ \int_{Q_0} \eta_0 \big(\theta(t, \mathbf{x}) - w^p \theta_p - (1 - w^p) \theta_\infty\big) \; d\Omega$$

where $d\Sigma := d\Gamma dt$ and $d\Gamma$ denotes the surface measure induced on $\partial \mathcal{D}$. The set of points that satisfy the first order necessary optimality conditions of (`LP`), denoted by $\mathcal{U}$ here, can be expressed as follows (c.f. [2]):

$$\mathcal{U} := \left\{ (w, \theta, \eta) \in \big(\mathcal{A} \times X_\theta(Q) \times X_\eta(Q)\big) \;\middle|\; \begin{array}{rcccl} \partial_w \mathcal{L}(w, \theta, \eta) = & 0 & \texttt{in} & \mathcal{D} & (\texttt{C.1}) \\ \partial_\theta \mathcal{L}(w, \theta, \eta) = & 0 & \texttt{in} & Q & (\texttt{C.2}) \\ \partial_\eta \mathcal{L}(w, \theta, \eta) = & 0 & \texttt{in} & Q & (\texttt{C.3}) \end{array} \right\}.$$

where $\partial_w \mathcal{L}$, $\partial_\theta \mathcal{L}$ and $\partial_\eta \mathcal{L}$ respectively denote the directional derivatives of $\mathcal{L}$ with respect to w, θ and η along arbitrary directions $\delta w \in X_w(\mathcal{D})$, $\delta \theta \in X_\theta(Q)$ and $\delta \eta \in X_\eta(Q)$ respectively. In fact set $\mathcal{U}$ includes constrained stationary points of lagrangian $\mathcal{L}$. Now, lets to define the following inner product notations, to keep our derivations concise:

$$\big\langle \cdot, \cdot \big\rangle_Q := \int_{\mathcal{T}} \big\langle \cdot, \cdot \big\rangle_{\mathcal{D}} \; dt, \qquad \big\langle \cdot, \cdot \big\rangle_\Sigma := \int_{\mathcal{T}} \big\langle \cdot, \cdot \big\rangle_{\partial \mathcal{D}} \; dt$$
$$\big\langle \cdot, \cdot \big\rangle_{Q_0} := \big\langle \cdot, \cdot \big\rangle_{\mathcal{D}} \; \big|_{t=0}, \qquad \big\langle \cdot, \cdot \big\rangle_{Q_T} := \big\langle \cdot, \cdot \big\rangle_{\mathcal{D}} \; \big|_{t=T}$$

For $\partial_w \mathcal{L}$ in the optimality condition (`C.1`) we have,

$$\begin{aligned} \big\langle \partial_w \mathcal{L}, \delta w \big\rangle_{\mathcal{D}} &= \big\langle \eta (\partial_w \rho \tilde{c} + \rho \partial_w \tilde{c}) \; \dot{\theta} - \eta \nabla \cdot (\partial_w k \nabla \theta), \; \delta w \big\rangle_Q + \big\langle \eta_b \; \partial_w k \nabla \theta \cdot \mathbf{n}, \; \delta w \big\rangle_\Sigma \\ &\quad - \big\langle \eta_0 \; \partial_w \theta_0, \; \delta w \big\rangle_{Q_0} \\ &= \big\langle \eta (\partial_w \rho \tilde{c} + \rho \partial_w \tilde{c}) \dot{\theta} + \partial_w k \nabla \eta \cdot \nabla \theta, \delta w \big\rangle_Q - \big\langle \eta_0 \partial_w \theta_0, \delta w \big\rangle_{Q_0} \end{aligned} \tag{8.1}$$

where,

$$\partial_w \rho = p w^{p-1} (\rho_c - \rho_m) \tag{8.2}$$
$$\partial_w \tilde{c} = p w^{p-1} (\tilde{c}_e - c_m) \tag{8.3}$$
$$\partial_w k = p w^{p-1} k_c k_m (k_c - k_m) \big(w^p (k_m - k_c) + k_c\big)^{-2} \tag{8.4}$$
$$\partial_w \theta_0 = p w^{p-1} (\theta_p - \theta_\infty) \tag{8.5}$$

Note that the boundary term in (8.1) is removed by taking $\eta = \eta_b$ on Σ and integration by part followed by applying the divergence theorem on the diffusion term.

For $\partial_\theta \mathcal{L}$ in the optimality condition ($\mathtt{C.2}$) we have,

$$\begin{aligned}\big\langle \partial_\theta \mathcal{L}, \delta\theta \big\rangle_Q &= \big\langle \partial_\theta \mathcal{F}, \delta\theta \big\rangle_Q + \big\langle \eta\rho\ \partial_\theta \tilde{c}\ \dot{\theta},\ \delta\theta \big\rangle_Q + \big\langle \eta\rho\tilde{c},\ \partial(\delta\theta)/\partial t \big\rangle_Q \\ &\quad - \big\langle \eta,\ \nabla\cdot\big(k\nabla(\delta\theta)\big)\big\rangle_Q + \big\langle k\eta_b,\ \nabla(\delta\theta)\cdot\mathbf{n}\big\rangle_\Sigma + \big\langle h_\infty \eta_b, \delta\theta\big\rangle_\Sigma + \big\langle \eta_0, \delta\theta \big\rangle_{Q_0} \\ &:= \mathtt{I}_1 + \mathtt{I}_2 + \mathtt{I}_3 + \mathtt{I}_4 + \mathtt{I}_5 + \mathtt{I}_6 + \mathtt{I}_7 \end{aligned} \tag{8.6}$$

where,

$$\partial_\theta \tilde{c} = w^p \frac{\partial \tilde{c}_e}{\partial \theta} = w^p L \frac{\partial^2 \tilde{f}_s}{\partial \theta^2} \tag{8.7}$$

The C^2 continuity of $\tilde{H}_\epsilon(\cdot)$ ensures the regularity of (8.7). To avoid lengthy expressions, lets to define the following notations:

$$g := (\dot{\theta})^{-\frac{1}{2}}|\nabla\theta| - g_c, \quad \hat{d} := \delta_a^{\mathcal{D}}\big(a_t\theta + b_t\big), \quad \tilde{g}_e := g\ \hat{d}, \quad \hat{g} := \min(\tilde{g}_e, 0)$$

Thus, for $\mathcal{G}$, in $\mathcal{F}$ we have,

$$\mathcal{G} = \Bigg(\min\Big(\big((\dot{\theta})^{-\frac{1}{2}}|\nabla\theta| - g_c\big)\ \delta_a^{\mathcal{D}}\big(a_t\theta + b_t\big), 0\Big)\Bigg)^2 \tilde{\mathcal{I}}_c = \big(\min(g\hat{d}, 0)\big)^2\ \tilde{\mathcal{I}}_c = \hat{g}^2\ \tilde{\mathcal{I}}_c$$

Note that $\mathcal{G}$ is an explicit function of θ, $\dot{\theta}$ and $\nabla\theta$. Thus, for $\mathtt{I}_1$ we have,

$$\mathtt{I}_1 = \big\langle \partial_\theta \mathcal{F}, \delta\theta \big\rangle_Q + \big\langle \partial_{\nabla\theta}\mathcal{F}, \nabla(\delta\theta)\big\rangle_Q + \big\langle \partial_{\dot{\theta}}\mathcal{F}, \dot{(\delta\theta)}\big\rangle_Q := \mathtt{J}_1 + \mathtt{J}_2 + \mathtt{J}_3 \tag{8.8}$$

Although $\min(x, 0)$ is not differentiable at $x = 0$, $(\min(x, 0))^2$ is smooth and differentiable everywhere. Now, we simplify $\mathtt{J}_1$ as follows,

$$\mathtt{J}_1 = \big\langle -a^{-2} a_t\ (a_t\theta + b_t)\ \tilde{g}_e\ \hat{g}\ \tilde{\mathcal{I}}_c,\ \delta\theta \big\rangle_Q \tag{8.9}$$

For $\mathtt{J}_2$ we have,

$$\begin{aligned}\mathtt{J}_2 &= \Big\langle \frac{\hat{d}\ \hat{g}\ \tilde{\mathcal{I}}_c}{|\nabla\theta|\sqrt{\dot{\theta}}}\ ,\ \nabla\theta\cdot\nabla(\delta\theta)\Big\rangle_Q \\ &= \Big\langle \frac{\hat{d}\ \hat{g}\ \tilde{\mathcal{I}}_c}{|\nabla\theta|\sqrt{\dot{\theta}}}\ \nabla\theta\cdot\mathbf{n},\ \delta\theta\Big\rangle_\Sigma - \Big\langle \nabla\cdot\Big(\frac{\hat{d}\ \hat{g}\ \tilde{\mathcal{I}}_c\nabla\theta}{|\nabla\theta|\sqrt{\dot{\theta}}}\Big),\ \delta\theta\Big\rangle_Q := \mathtt{J}_{2,1} - \mathtt{J}_{2,2}\end{aligned} \tag{8.10}$$

The boundary integral $\mathtt{J}_{2,1}$ in (8.10) is equal to zero, because $\tilde{\mathcal{I}}_c = 0$ on $\partial\mathcal{D}$. To proceed derivation, lets to define the following notations,

$$j := \frac{\partial \mathcal{G}}{\partial \dot{\theta}} = -(\dot{\theta})^{-\frac{3}{2}}|\nabla\theta|\ \hat{d}\ \hat{g}\ \tilde{\mathcal{I}}_c$$

The regularization of function j, denoted by $\hat{j}$, is defined as,

$$\hat{j} = -(\dot{\theta})^{-\frac{3}{2}}|\nabla\theta|\ \hat{d}\ \widetilde{\min}(\tilde{g}_e, 0)\ \tilde{\mathcal{I}}_c \tag{8.11}$$

In fact, the minimum function in j is replaced by the regularized minimum function in $\hat{j}$. Considering the identity $\min(x, 0) = x(1 - H(x))$, the regularized minimum function can be defined as follows,

$$\widetilde{\min}(x, 0) := x(1 - \tilde{H}_\epsilon(x)) \tag{8.12}$$

The regularization parameter ϵ corresponding to (8.12) is taken equal to one spatial grid-size in the present work. Applying the integration by part on $\mathtt{J}_3$ in (8.8) results in,

$$\mathtt{J}_3 = \Big\langle j,\ \delta\theta\Big\rangle_{Q_T} - \Big\langle j,\ \delta\theta\Big\rangle_{Q_0} - \Big\langle d_t j,\ \delta\theta\Big\rangle_Q := \mathtt{J}_{3,1} - \mathtt{J}_{3,2} - \mathtt{J}_{3,3} \tag{8.13}$$

and,

$$d_t j(\theta, \dot{\theta}, \nabla\theta) \approx \partial_\theta \hat{j}\ \dot{\theta} + \partial_{(\dot{\theta})} \hat{j}\ \ddot{\theta} + \partial_{(\nabla\theta)} \hat{j} \cdot \nabla(\dot{\theta}) \tag{8.14}$$

where $\ddot{(\cdot)} := \partial^2(\cdot)/\partial t^2$. Since, expanding terms in (8.14) is straightforward, we leave further simplification of (8.14) to save the space. It is worth mentioning that by the regularization of minimum function in $\mathcal{F}$, it depends explicitly on w. However, because the concentration of $\mathcal{F}$ is in $\mathcal{D}_c$ and w only vary in $\mathcal{D}_d$, the explicit derivative of $\mathcal{F}$ with respect to w is approximately equal to zero. Hence, it is ignored in our derivation here. Because, the concentration of $\delta_a^{\mathcal{D}}(a_t\theta + b_t)$ is within the freezing interval, it is approximately equal to zero at $t = 0$. Thus, $j \approx 0$ at $t = 0$ in practice. Consequently, $\mathtt{J}_{3,2} \approx 0$. The integration by part simplifies $\mathtt{I}_3$ to:

$$\mathtt{I}_3 = \big\langle \rho\tilde{c}\ \eta,\ \delta\theta \big\rangle_{Q_T} - \big\langle \rho\tilde{c}\ \eta,\ \delta\theta \big\rangle_{Q_0} - \big\langle \partial(\rho\tilde{c}\ \eta)/\partial t,\ \delta\theta \big\rangle_Q := \mathtt{I}_{3,1} - \mathtt{I}_{3,2} - \mathtt{I}_{3,3} \tag{8.15}$$

Using the integration by part and divergence theorem twice simplifies $\mathtt{I}_4$ as follows,

$$\begin{aligned} \mathtt{I}_4 &= -\big\langle k\eta,\ \nabla(\delta\theta)\cdot\mathbf{n} \big\rangle_\Sigma + \big\langle k\nabla\eta\cdot\mathbf{n},\ \delta\theta \big\rangle_\Sigma - \big\langle \nabla\cdot(k\ \nabla\eta),\ \delta\theta \big\rangle_Q \\ &= -\mathtt{I}_{4,1} + \mathtt{I}_{4,2} - \mathtt{I}_{4,3} \end{aligned}$$

The collection of all terms in the right hand side of (8.6) results in the following abstract form:

===

$$\begin{aligned} \big\langle \partial_\theta \mathcal{L}, \delta\theta \big\rangle_Q = \underbrace{(\mathtt{J}_1 - \mathtt{J}_{2,2} - \mathtt{J}_{3,3} + \mathtt{I}_2 - \mathtt{I}_{3,3} - \mathtt{I}_{4,3})}_{\mathtt{in\ Q}} + \underbrace{(-\mathtt{I}_{4,1} + \mathtt{J}_{4,2} + \mathtt{I}_5 + \mathtt{I}_6)}_{\mathtt{on\ \Sigma}} \\ + \underbrace{(\mathtt{J}_{3,1} + \mathtt{I}_{3,1})}_{\mathtt{in\ Q_T}} + \underbrace{(-\mathtt{I}_{3,2} + \mathtt{I}_7)}_{\mathtt{in\ Q_0}} = 0 \end{aligned} \tag{8.16}$$

Now we drive the adjoint PDE by equating terms in every parenthesis of (8.16) to zero (c.f. [3, 31]). We first examine the second parenthesis. Because the test function $\delta\theta$ is arbitrary (i.e., the parenthesis should be equal to zero for every choice of trace of $\delta\theta$ on Σ), we take $\delta\theta = 0$ on Σ and vary term $k\nabla(\delta\theta)\cdot\mathbf{n}$ on Σ. Thus, η should be equal to η_b on Σ. Consequently, $\mathtt{I}_{4,1} = \mathtt{I}_5$ and the second and forth terms in this parenthesis cancel each others. Varying the trace of $\delta\theta$ on Σ results in the following boundary condition:

$$-k\nabla\eta\cdot\mathbf{n} = \eta h_\infty \quad \mathtt{on}\ \ \Sigma$$

Using similar treatments results in the following identities:

$$\rho\tilde{c}\ \eta = -j \quad \mathtt{in}\ \ \mathtt{Q_T}$$

$$\eta_0 = \rho\tilde{c}\ \eta \quad \mathtt{in}\ \ \mathtt{Q_0}$$

Thus, enforcing the condition (C.2), results in the following adjoint heat equation:

$$(\mathtt{AH})(\theta, w) := \begin{cases} -\dfrac{\partial\Big(\rho(w)\tilde{c}(w)\ \eta\Big)}{\partial t} = & \nabla\cdot(k(w)\nabla\eta) - \eta S_a(\mathbf{x}) + S_b(\mathbf{x}) \quad \mathtt{in}\ \ Q \\ -k(w)\nabla\eta\cdot\mathbf{n} = & h_\infty\eta \quad \mathtt{on}\ \ \Sigma \\ \eta(t,\mathbf{x}) = & \eta_0(\mathbf{x}) \quad \mathtt{in}\ \ Q_T \end{cases}$$

where,

$$S_a(\mathbf{x}) = w^p \rho(w) L \ \frac{\partial^2 \tilde{f}_s}{\partial \theta^2} \ \dot{\theta} \tag{8.17}$$

$$S_b(\mathbf{x}) = a^{-2} a_t (a_t \theta + b_t) \tilde{g}_e \hat{g} \ \tilde{\mathcal{I}}_c + \nabla \cdot \Big(\hat{d}\hat{g} \ \frac{\nabla \theta}{|\nabla \theta| \sqrt{\dot{\theta}}} \ \tilde{\mathcal{I}}_c \Big) + d_t j \tag{8.18}$$

$$\eta_0(\mathbf{x}) = \big(\rho(w)\tilde{c}(w)\big)^{-1} (\dot{\theta})^{-\frac{3}{2}} |\nabla \theta| \hat{d}\hat{g}\tilde{\mathcal{I}}_c \tag{8.19}$$

Note that the adjoint heat equation (AH) should be integrated in reverse time direction, i.e., from $t = T$ to $t = 0$ (consider the negative sign of transient term in this PDE). The optimality condition (C.3) holds when θ solves the direct heat equation in (LP) for known w function:

$$(\mathtt{DH})(w) := \left\{ \begin{array}{rll} \rho(w)\tilde{c}(w) \ \frac{\partial \theta}{\partial t} = & \nabla \cdot (k(w) \nabla \theta) & \mathtt{in} \quad Q \\ k(w) \nabla \theta \cdot \mathbf{n} = & h_\infty (\theta_\infty - \theta) & \mathtt{on} \quad \Sigma \\ \theta(t, \mathbf{x}) = & \theta_0(\mathbf{x}) & \mathtt{in} \quad Q_0 \end{array} \right.$$

where $\theta_0(\mathbf{x}) := w^p \theta_p + (1 - w^p)\theta_\infty$.

Now lets to state the necessary optimality conditions based on the projected gradient approach (c.f. [27, 29]). For the sake of convenience, we briefly recall some known results.

Theorem 8.2. *(orthogonal projection over a convex set, theorem 12.1.10 of [2]) Let V be a Hilbert space and K as a convex closed nonempty subset of V. For all $u \in V$, there exists a unique $u_K \in K$ such that*

$$\|u - u_K\|_2^2 = \arg\min_{v \in K} \|u - v\|_2^2.$$

The orthogonal projection of u onto set K is shown by operator $\mathcal{P}_K(u)$ henceforth in this paper, i.e., $u_K = \mathcal{P}_K(u)$. Equivalently, u_K is characterized by the following property:

$$u_K \in K, \quad \langle u_K - u, \ v - u_K \rangle \geqslant 0, \quad \forall v \in K \tag{8.20}$$

Theorem 8.3. *(Euler inequality for convex sets, theorem 10.2.1 of [2]) Let V be a Hilbert space and K as a convex closed nonempty subset of V. Assume functional $J(u) : K \to \mathbb{R}$ is differentiable at $u \in K$ with the directional derivative denoted by $J'(u)$. If u be a local minimum point of $J(u)$ over K then:*

$$\langle J'(u), \ v - u \rangle \geqslant 0, \quad \forall v \in K \tag{8.21}$$

Corollary 8.4. *(necessary optimality conditions based on the projected gradient, proposition 2.4 of [29]) Let V be a Hilbert space and K as a convex closed nonempty subset of V. Assume functional $J(u) : K \to \mathbb{R}$ is differentiable at $u \in K$ with the directional derivative denoted by $J'(u)$. If u be a local minimum point of $J(u)$ over K then:*

$$J'_{K,\mu}(u) = 0 \quad \textsf{a.e.}, \quad J'_{K,\mu}(u) = \mathcal{P}_K(u - \mu J') - u \tag{8.22}$$

where $\mu \in \mathbb{R}^+$. Since $\mathcal{P}_K(u - \mu J') - u$ is equivalent to the scaled projected gradient, constrained stationary points of J are zeros of the scaled projected gradient with respect to set K. Therefore we call (8.22) as the necessary optimality conditions based on the projected gradient.

Corollary 8.5. *(descent property of the scaled projected gradient, proposition 2.5 of [29]) Let V be a Hilbert space and K as a convex closed nonempty subset of V. Assume functional $J(u) : K \to \mathbb{R}$ is differentiable at $u \in K$ with the directional derivative denoted by $J'(u)$. Assume that the scaled projected gradient at $u \in K$ is denoted by $J'_{K,\mu}(u)$, i.e., $J'_{K,\mu}(u) = \mathcal{P}_K(u - \mu J') - u$. Then for all $u \in K$ and $\mu \in \mathbb{R}^+$ we have:*

$$\langle J'(u),\ J'_{K,\mu}(u)\rangle \leqslant -\frac{1}{2\mu}\|J'_{K,\mu}(u)\|_2^2 \tag{8.23}$$

Corollary 8.5 suggests a simple iterative gradient descent method to find local minimums of convex constrained optimization problems. Assume that the iteration counter is denoted by m and consider u_0 as the initial guess, then we have:

$$u_{m+1} = u_m + \nu_m J'_{K,\mu_m}(u_m), \quad m = 0, 1, \ldots, \tag{8.24}$$

where $\nu_m \in (0, 1]$. In the present study, the scaling parameter μ_m is selected based on the Barzilai-Borwein (BB) step-size [4]:

$$\mu_m = \langle \mathbf{s}_{m-1},\ \mathbf{s}_{m-1}\rangle / \langle \mathbf{s}_{m-1},\ \mathbf{r}_{m-1}\rangle,$$

where $\mathbf{s}_m := u_m - u_{m-1}$ and $\mathbf{r}_m := J'(u_m) - J'(u_{m-1})$. The main benefit of using BB step-size is its spectral properties. It makes it a cheap and efficient method to approximately solve large-scale optimization problems (c.f. [8]). To ensure the global convergence, it is required that ν_m is selected based on a globalization strategy, like a line-search algorithm. However, using a monotonic line-search algorithm reduces (8.24) to the classic projected steepest descent method, i.e., we miss promising properties of BB step-size. To cope this problem, the nonmonotone globalization strategy suggested in [8] is used in the present study (see [8] for further details).

Combining the above results, we can state the first order necessary optimality conditions of (LP), denoted by (OC) as follows:

$$(\mathtt{OC}) := \left\{ \begin{array}{rll} \rho(w)\tilde{c}(w)\ \frac{\partial\theta}{\partial t} = & \nabla\cdot(k(w)\nabla\theta) & \mathtt{in}\quad Q \\ k(w)\nabla\theta\cdot\mathbf{n} = & h_\infty(\theta_\infty - \theta) & \mathtt{on}\quad \Sigma \\ \theta(t,\mathbf{x}) = & \theta_0(\mathbf{x}) & \mathtt{in}\quad Q_0 \\ & & \\ -\frac{\partial\big(\rho(w)\tilde{c}(w)\ \eta\big)}{\partial t} = & \nabla\cdot(k(w)\nabla\eta) - \eta S_a(\mathbf{x}) + S_b(\mathbf{x}) & \mathtt{in}\quad Q \\ -k(w)\nabla\eta\cdot\mathbf{n} = & h_\infty\eta & \mathtt{on}\quad \Sigma \\ \eta(t,\mathbf{x}) = & \eta_0(\mathbf{x}) & \mathtt{in}\quad Q_T \\ & & \\ \mathcal{P}_{\mathcal{A}}(w - \partial_w\mathcal{L}) - w = & 0 & \mathtt{in}\quad \mathcal{D} \end{array}\right.$$

For the sake of completeness, lets to briefly outline the optimization algorithm used to solve problem (LP) in the present work. To avoid technical difficulties, we sate our algorithm for discretized version of our original problem (details of discretization method will be discussed later). Lets to denote by $(\mathtt{LP})_\mathtt{h}$, $(\mathtt{OC})_\mathtt{h}$, $(\mathtt{DH})_\mathtt{h}$ and $(\mathtt{AH})_\mathtt{h}$, the finite dimensional counterpart of (LP), (OC), (DH) and (AH) respectively. Assume θ and η solve problems $(\mathtt{DH})_\mathtt{h}$, $(\mathtt{AH})_\mathtt{h}$ respectively, the optimal value of w is found using the nonmonotone globalized version of iterations (8.24). This algorithm is in fact identical to nonmonotone spectral projected gradient (SPG) method introduced in [8]. More precisely, the SPG2 algorithm presented in [8, 9] is used in this study. To use SPG2 in this regard, it is sufficient to remark that whenever SGP2 asks for the objective functional value, $(\mathtt{DH})_\mathtt{h}$ should be solved using current value of w-filed and then the objective function should be evaluated. Similarly, when SGP2 asks for the gradient of the objective functional, it is required to solve $(\mathtt{DH})_{\mathtt{h}}$ and $(\mathtt{AH})_{\mathtt{h}}$ respectively and, then to compute $\partial_w \mathcal{L}$ using (8.1). The input parameters for SPG2 algorithm in this study are identical to set of default parameters in [9]. As it is mentioned in [8], the efficiency of SPG2 mainly relay on the existence of an efficient way to project trial steps onto the feasible set of solutions (is a user-defined function in SPG2 algorithm), i.e., $\mathcal{P}_{\mathcal{A}}(\cdot)$ operator in the present study. Therefore, introducing an efficient method for the projection onto the admissible domain of control parameters is the remaining part of our optimization algorithm which is discussed in section 9.

9. Projection onto the admissible control domain

Assume that the admissible control domain of w is nonempty, i.e., $\mathcal{A} \neq \emptyset$. Since all constraints define $\mathcal{A}$ are linear functions of w, $\mathcal{A}$ is a closed convex subset of $L^2(\mathcal{D})$. Consider an arbitrary control variable w with at least an $L^2(\mathcal{D})$ regularity. According to theorem 8.2, there is a unique $w_{\mathcal{A}} \in \mathcal{A}$ where (c.f. [30]):

$$w_{\mathcal{A}} = \mathcal{P}_{\mathcal{A}}(w) := \arg\min_{w \in \mathcal{A}} \frac{1}{2}\, \|w_{\mathcal{A}} - w\|_2. \tag{9.1}$$

Lets to redefine the projection problem (9.1) as follows:

$$\arg\min_{w \in \mathcal{B}_w} \frac{1}{2}\, \|w_{\mathcal{A}} - w\|_2 \quad \mathtt{s.t.:} \quad \int_{\mathcal{D}_d} w(\mathbf{x})\ d\Omega = b_d,\ \int_{\mathcal{D}_{dn}} w(\mathbf{x})\ d\Omega \leqslant b_{dn}, \tag{9.2}$$

where $b_d = V_d R_d$, $b_{dn} = V_{dn} R_{dn}$ and,

$$\mathcal{B}_w := \Big\{ w \in X_w(\mathcal{D}) \ \Big|\ w_L \leqslant w \leqslant w_U \ \Big\}$$

where inequalities in the definition of $\mathcal{B}_w$ are to be understood pointwise and,

$$w_L := \left\{ w \in X_w(\mathcal{D}) \ \middle| \ \begin{array}{rcll} w_L(\mathbf{x}) = & 1 & \mathtt{in} & \mathcal{D}_c \\ w_L(\mathbf{x}) = & 0 & \mathtt{in} & \mathcal{D} \setminus \mathcal{D}_c \end{array} \right\}.$$

$$w_U := \left\{ w \in X_w(\mathcal{D}) \ \middle| \ \begin{array}{rcll} w_U(\mathbf{x}) = & 1 & \mathtt{in} & \mathcal{D}_c \cup \mathcal{D}_d \\ w_U(\mathbf{x}) = & 0 & \mathtt{in} & \mathcal{D}_r \setminus \mathcal{D}_d \end{array} \right\}.$$

The boundary condition $w = 0$ on Σ in naturally maintained based on the above formulation because $\Sigma \subset \mathcal{D}_r \setminus \mathcal{D}_d$. It is evident that $w_{\mathcal{A}}$ is alternatively equal to unique constrained $(w \in \mathcal{B}_w)$ stationary point of the following augmented lagrangian:

$$\mathcal{M}(\lambda_d, \lambda_{dn}, w) := \frac{1}{2}\, \|w_{\mathcal{A}} - w\|_2 + \lambda_d \Big(\int_{\mathcal{D}_d} w\ d\Omega - b_d \Big) + \lambda_{dn} \Big(\int_{\mathcal{D}_{dn}} w\ d\Omega - b_{dn} \Big)$$

where $\lambda_d, \lambda_{dn} \in \mathbb{R}$ are lagrange multipliers corresponding to the integral equality and inequality constraints in (9.2) respectively. The constrained stationary point of $\mathcal{M}$ is characterized by identity,

$$\partial_{\lambda_d} \mathcal{M}(\lambda_d, \lambda_{dn}, w) = \partial_{\lambda_n} \mathcal{M}(\lambda_d, \lambda_{dn}, w) = \partial_w \mathcal{M}(\lambda_d, \lambda_{dn}, w) = 0, \tag{9.3}$$

together with $w \in \mathcal{B}_w$ and the complementarity conditions,

$$\lambda_{dn} \Big(\int_{\mathcal{D}_{dn}} w\ d\Omega - b_{dn} \Big) = 0, \quad \lambda_{dn} \geqslant 0$$

for the qualification of the corresponding integral inequality constraint. The first and second conditions in (9.3) are equivalent to the integral equality and inequality constraints respectively. Consider an arbitrary test function $\delta w \in X_w(\mathcal{D})$, for the directional derivative of $\mathcal{M}$ with respect to δw we have,

$$\big\langle \partial_w \mathcal{M},\ \delta w \big\rangle_{\mathcal{D}} = \big\langle w_{\mathcal{A}} - w,\ \delta w \big\rangle_{\mathcal{D}} + \int_{\mathcal{D}_d} \lambda_d \delta w \ d\Omega + \int_{\mathcal{D}_{dn}} \lambda_{dn} \delta w \ d\Omega \tag{9.4}$$

Denoting by $\mathcal{I}_d$ and $\mathcal{I}_{dn}$ the characteristic functions corresponding to the spatial domains $\mathcal{D}_d$ and $\mathcal{D}_{dn}$, (9.4) can be written as follows:

$$\big\langle \partial_w \mathcal{M},\ \delta w \big\rangle_{\mathcal{D}} = \big\langle w_{\mathcal{A}} - w,\ \delta w \big\rangle_{\mathcal{D}} + \int_{\mathcal{D}} \lambda_d \delta w \mathcal{I}_d \ d\Omega + \int_{\mathcal{D}} \lambda_{dn} \delta w \mathcal{I}_{dn} \ d\Omega \tag{9.5}$$

(9.5) simplifies to:

$$\big\langle \partial_w \mathcal{M},\ \delta w \big\rangle_{\mathcal{D}} = \big\langle w_{\mathcal{A}} - w + \lambda_d \mathcal{I}_d + \lambda_{dn} \mathcal{I}_{dn},\ \delta w \big\rangle_{\mathcal{D}} \tag{9.6}$$

The following condition holds the third condition of (9.3) almost everywhere in $\mathcal{D}$:

$$w_{\mathcal{A}} - w + \lambda_d \mathcal{I}_d + \lambda_{dn} \mathcal{I}_{dn} = 0, \quad \texttt{a.e. in } \mathcal{D} \tag{9.7}$$

To enforce the bound constraints, $w \in \mathcal{B}_w$, the necessary optimality conditions based on the projected gradient method, corollary 8.4, is employed here. Thus, the constrained stationary point of $\mathcal{M}$ should satisfy the following condition:

$$w_{\mathcal{A}} - \mathcal{P}_{\mathcal{B}_w}\big(w - \lambda_d \mathcal{I}_d - \lambda_{dn} \mathcal{I}_{dn}\big) = 0, \quad \texttt{a.e. in } \mathcal{D} \tag{9.8}$$

Backing to the definition of domains $\mathcal{D}_d$ and $\mathcal{D}_{dn}$ in section 2, we have $\mathcal{I}_{dn} \subseteq \mathcal{I}_d$. Therefore, we can decompose (9.8) into the following independent conditions defined in non-overlapping parts of $\mathcal{D}$:

$$\begin{aligned} w_{\mathcal{A}} - \mathcal{P}_{\mathcal{B}_w}\big(w\big) &= 0, \quad \texttt{a.e. in } \mathcal{D} \setminus \mathcal{D}_d, \\ w_{\mathcal{A}} - \mathcal{P}_{\mathcal{B}_w}\big(w - \lambda_d\big) &= 0, \quad \texttt{a.e. in } \mathcal{D}_d \setminus \mathcal{D}_{dn}, \\ w_{\mathcal{A}} - \mathcal{P}_{\mathcal{B}_w}\big(w - \lambda_d - \lambda_{dn}\big) &= 0, \quad \texttt{a.e. in } \mathcal{D}_{dn} \end{aligned}$$

It is well-know that the projection onto the bound constraints is separable and can be explicitly computed as follows (c.f. chapter 10 of [2]):

$$\mathcal{P}_{\mathcal{B}_w}(w) = \texttt{mid}\ (w_L, w, w_u),$$

where the median operator, $\texttt{mid}\ (\cdot,\cdot,\cdot)$, is to be understood pointwise here and is defined as follows,

$$\texttt{mid}\ (a,b,c) := \min\ (c,\ \max\ (a,b)),$$

where $a,b,c \in \mathbb{R}$ and $a \leqslant c$. As a result, we have:

$$w_{\mathcal{A}} = \texttt{mid}\ (w_L, w, w_u), \qquad \texttt{a.e. in } \mathcal{D} \setminus \mathcal{D}_d, \tag{9.9}$$

$$w_{\mathcal{A}} = \texttt{mid}\ (w_L, w - \lambda_d, w_u), \qquad \texttt{a.e. in } \mathcal{D}_d \setminus \mathcal{D}_{dn}, \tag{9.10}$$

$$w_{\mathcal{A}} = \texttt{mid}\ (w_L, w - \lambda_d - \lambda_{dn}, w_u), \qquad \texttt{a.e. in } \mathcal{D}_{dn} \tag{9.11}$$

It is clear that (9.9) has an explicit evident solution. However, to solve (9.10) and (9.11), the values of λ_d and λ_{dn} at the optimal solution are required. For this purpose, we use integral constraints and consider two different cases: (i) the integral inequality constraint is inactive at the optimal solution, (ii) the integral inequality constraint is active at the optimal solution. Since lagrangian $\mathcal{M}$ is strictly convex, it admits only one stationary point. Thus, only one of the above cases happens in practice. If case (i) happens at the optimal solution, the complementarity condition enforces that $\lambda_{dn} = 0$. Consequently, λ_d is equal to the unique root of the following mono-variable nonlinear equation:

$$f_1(\lambda_d) = b_d - \int_{\mathcal{D}_d} \texttt{mid}\ (w_L, w - \lambda_d, w_u)\ d\Omega = 0 \tag{9.12}$$

In the other case, the integral inequality constraint is active at the optimal solution, i.e., we have $\int_{\mathcal{D}_{dn}} w_{\mathcal{A}}\ d\Omega = b_{dn}$. In this case, the optimal value of λ_d is equal to the unique root of the following mono-variable nonlinear equation:

$$f_2(\lambda_d) = b_d - b_{dn} - \int_{\mathcal{D}_d \backslash \mathcal{D}_{dn}} \texttt{mid}\ (w_L, w - \lambda_d, w_u)\ d\Omega = 0 \tag{9.13}$$

Having the optimal value of λ_d from (9.13), the optimal value of λ_{dn} is equal to the unique root of the following equation:

$$f_3(\lambda_{dn}) = b_{dn} - \int_{\mathcal{D}_{dn}} \texttt{mid}\ (w_L, w - \lambda_d - \lambda_{dn}, w_u)\ d\Omega = 0 \tag{9.14}$$

Therefore, the remaining job here is to suggest an efficient way to compute unique roots of functions f_1, f_2 and f_3. Following [30], the result of following proposition is employed for this purpose.

Proposition 9.1. *Functions $f_1(\lambda_d)$, $f_2(\lambda_d)$ and $f_3(\lambda_{dn})$ are continuous piecewise linear and monotonically non-increasing functions of their arguments.*

Proof. We present the proof here only for function f_1. The extension of proof for the other cases is straightforward. Lets to define functions λ_d^L and λ_d^U, with at least an $L^2(\mathcal{D}_d)$ regularity, as follows:

$$\lambda_d^L = w - w_L, \quad \lambda_d^U = w - w_U,$$

where the above relations are to be understood pointwise. It is clear that $\lambda_d^U \leqslant \lambda_d^L$ ($w_L = 0$ and $w_U = 1$ in $\mathcal{D}_d$). Considering (9.10) and (9.11) we have (recall that in this case $\lambda_{dn} = 0$):

$$w_{\mathcal{A}}(\lambda_d) = \begin{cases} w_U, & \texttt{if} \quad \lambda \leqslant \lambda_d^U, \\ w - \lambda_d, & \texttt{if} \quad \lambda_d^U \leqslant \lambda \leqslant \lambda_d^L, \\ w_L, & \texttt{if} \quad \lambda \geqslant \lambda_d^L. \end{cases} \tag{9.15}$$

where all algebra in (9.15) are to be understood pointwise. Considering (9.15), $w_{\mathcal{A}}(\lambda_d)$ is a continuous piecewise linear and monotonically non-increasing function of λ_d. Since we have $f_1(\lambda_d) = b_d - \int_{\mathcal{D}_d} w_{\mathcal{A}}\ d\Omega$, $f_1(\lambda_d)$ is also a continuous piecewise linear and monotonically non-increasing function of λ_d. □

Using proposition 9.1 we can find the unique roots of f_1, f_2 and f_3 efficiently using the bisection algorithm. The following corollary provides valuable information to select the initial search interval of bisection algorithm. We state the corollary only for function f_1, but, the same results can be easily obtained for f_2 and f_3 which are ignored here to save the space.

Corollary 9.2. *Consider functions λ_d^L and λ_d^U as defined in proposition 9.1. Let $\vartheta \in \mathbb{R}$ such that $\vartheta = \max\{\ \|\lambda_d^L\|_\infty\ ,\ \|\lambda_d^U\|_\infty\ \}$, then $f_1(-\vartheta)f_1(\vartheta) \leqslant 0$, i.e., the root of f_1 happens within interval $[-\vartheta, \vartheta\ ]$. Moreover $f_1(\lambda_d) \leqslant 0$ for $\lambda_d \leqslant -\vartheta$ and $f_1(\lambda_d) \geqslant 0$ for $\lambda_d \geqslant -\vartheta$.*

Proof. Since $f_1(\lambda_d)$ is a continuous piecewise linear and monotonically non-increasing function of λ_d and it has a unique root, we should have $f_1(\lambda_d) \leqslant 0$ for $\lambda_d \leqslant -\vartheta$ and $f_1(\lambda_d) \geqslant 0$ for $\lambda_d \geqslant -\vartheta$. Therefore, the root happens in $[-\vartheta, \vartheta\ ]$. □

It is easy to see that, starting from the initial interval $[-\vartheta, \vartheta\ ]$, the number of bisection steps to find λ_d within tolerance ϱ is at most equal to $\lceil \log_2(2\vartheta/\varrho) \rceil$. This means that, λ_d can be found by a few bi-section steps up to the machine precision[4]. Therefore, the projection steps in the present study can be performed very efficiently up to the machine precision. It is worth mentioning that, the computational cost of one projection steps in this study is much more smaller than one percent of the total computational cost in practice.

10. Numerical method

For the purpose of the numerical solution, the spatial domain $\mathcal{D}$ is decomposed into a uniform Cartesian grid with the grid spacing Δx (a user-defined parameter). The Cartesian grid generator CartGen [26] is used in this study to discretize complex geometries. Although a Cartesian grid makes an stair-case approximation to curved boundaries, our numerical experiments showed that such grids are adecuate for the purpose of solidification analysis, for instance see: [33]. The temporal domain is decomposed into a uniform grid with grid spacing Δt that is determined based on the stability criteria corresponding to fully the explicit solution of heat equation. Except terms $\frac{\partial \theta}{\partial t}$ and $\frac{\partial \eta}{\partial t}$ in the heat equations, all differential operators are discretized by cell-centered finite volume method using second order central schemes. The heat equations are integrated along the time axis using the first order fully explicit method. It poses an upper bound on Δt. Because our PDEs are highly nonlinear and include spatio-temporal concentrated source terms, using large time steps reduces the accuracy of computations. The main reason for the selection of fully explicit time integration method is due to the efficiency and ability to capture details of solidification dynamics with a reasonable accuracy. The classical trapezoidal integration method is employed to perform the temporal integrations.

Our main difficulty in using the explicit time integration method roots in the fact that the number of time steps is very large in practice (typically order of 10^3-10^4). This makes the computation of adjoint heat source and objective functional derivative challenge, because the complete temperature history should be available prior to compute these quantities. This is a memory expensive procedure, considering large scale problems. For instance to store just a double precision floating point θ-field which include 10^6 grid points ($100 \times 100 \times 100$) for 10^4 time steps over 50 GB memory is needed. A common way to manage this problem is the recursive windowing approach presented in [5]. However this method is very expensive when the number of time steps is large. In the present study a very simple but efficient method is used to cope this problem. We simply store θ-field on the hard-disk after each time step and read it from the disk whenever it is required. According to our experience the CPU cost of a read and write operation for one θ-field is smaller than the cost of one forward-backward time-step. Therefore, the computational cost of our algorithm will be doubled in the worst conditions. Note that the available storage on the hard-disk is usually sufficient to manage our desired problems.

[4] The machine precision for the double precision arithmetic is usually of order of 10^{-16}

11. Numerical results

In this section we study the feasibility of the presented method using 15 numerical examples. A personal computer with an AMD 2.4 GHz and 2.5 GB DDR2 RAM is used as the computational resource. In all examples reported here the sand mold casting of a carbon steel alloy is considered. Table 1 shows the physical properties, initial and boundary conditions used in our solidification analysis. The initial value of w-filed during solution of each sub-problem (LP) is taken equal to its corresponding R_d.

Table 1. Physical properties, initial and boundary conditions used for the solidification simulation in this study. Units are in SI, except for the temperature which is expressed in ^{o}C.

	k	ρ	c	L	θ^0	θ_l	θ_s	θ_∞	h_∞	g_c
metal	33.5	7200	627	2.7×10^5	1594	1488	1440	-	-	0.5
sand	0.7	1500	1130	-	20	-	-	20	72	-

The optimization parameters corresponding to examples 1-15 are listed in table 2. For each example, the length unit is equal to the corresponding spatial grid-size. We denote by $\mathcal{D}_d^{top}$, $\mathcal{D}_d^{lateral}$, $\mathcal{D}_d^{side}$ the limitation of design space $\mathcal{D}_d$ to top-side, lateral-sides and only one lateral-side respectively. Moreover, $\mathcal{D}_d^{free}$ denotes applying no limitation on position of $\mathcal{D}_d$. Now lets to briefly describe the geometries corresponding to examples 1-15 (see figures 5, 6 and 7)

1. Cubic box with edge length 30, $\mathcal{D}_d^{top}$ (Fig. 5.a).
2. Same as 1, but $\mathcal{D}_d^{lateral}$ (Fig. 5.b).
3. Same as 2, but $\mathcal{D}_d^{side}$ (Fig. 5.c).
4. $90 \times 20 \times 10$ strip, $\mathcal{D}_d^{top}$ (Fig. 5.d).
5. $100 \times 100 \times 10$ planar, $\mathcal{D}_d^{top}$ (Fig. 5.e).
6. $(100 \times 60 \times 10) \setminus (60 \times 20 \times 10)$ frame-like, $\mathcal{D}_d^{top}$ (Fig. 6.a).

7, 8. Ring with 120 (100) outer (inner) radius and 15 height, $\mathcal{D}_d^{top}$ (Fig. 6.b).

9. 2 $\texttt{cube}(20 \times 20 \times 20) \cup \texttt{strip}(40 \times 20 \times 10)$, $\mathcal{D}_d^{top}$ (Fig. 6.c).
10. 2 $\texttt{cube}(20 \times 20 \times 20) \cup \texttt{strip}(20 \times 20 \times 10)$, $\mathcal{D}_d^{top}$ (Fig. 6.d).
11. Three cylinders connected with a narrow bridge, $\mathcal{D}_d^{top}$ (Fig. 6.e).

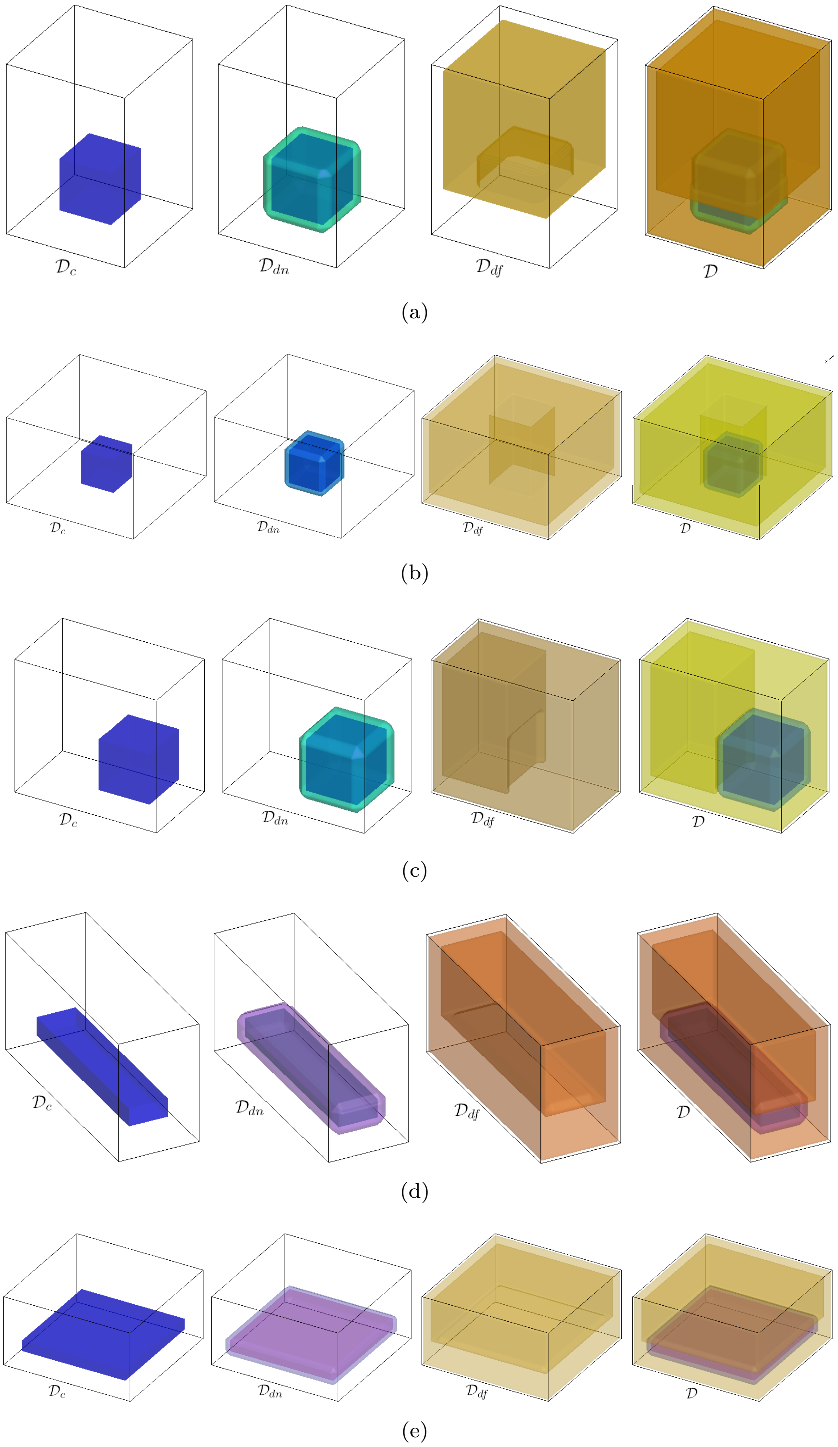
12, 13, 14. Hammer casting, $\mathcal{D}_d^{lateral}$, $\mathcal{D}_d^{top}$, $\mathcal{D}_d^{top}$ respectively (Fig. 7.a).

15. Tiller crankshaft casting, $\mathcal{D}_d^{free}$. (Fig. 7.b).

Figures 8-13 show results of examples 1-15 in the present work. Results of each case includes the final topology and its cross section. The iso-contour $w = 0.5$ is considered as the final topology here. To evaluate the feasibility of final design from a practical point of view, every design is examined using the reduced gravity model (RGM) [16] to studt the formation of macro-shrinkage defects. It is worth mentioning that this model is a well-accepted model in the foundry society and famous commercial casting simulation codes like MAGMASOFT[5] and FLOW-3D[6] use this method to predict formation of macro-scale solidification defects. The result of RGM simulation (in fact the shape of macro-shrinkage cavity(s)) is appended

[5] www.magmasoft.com

[6] www.flow3d.com



(a)

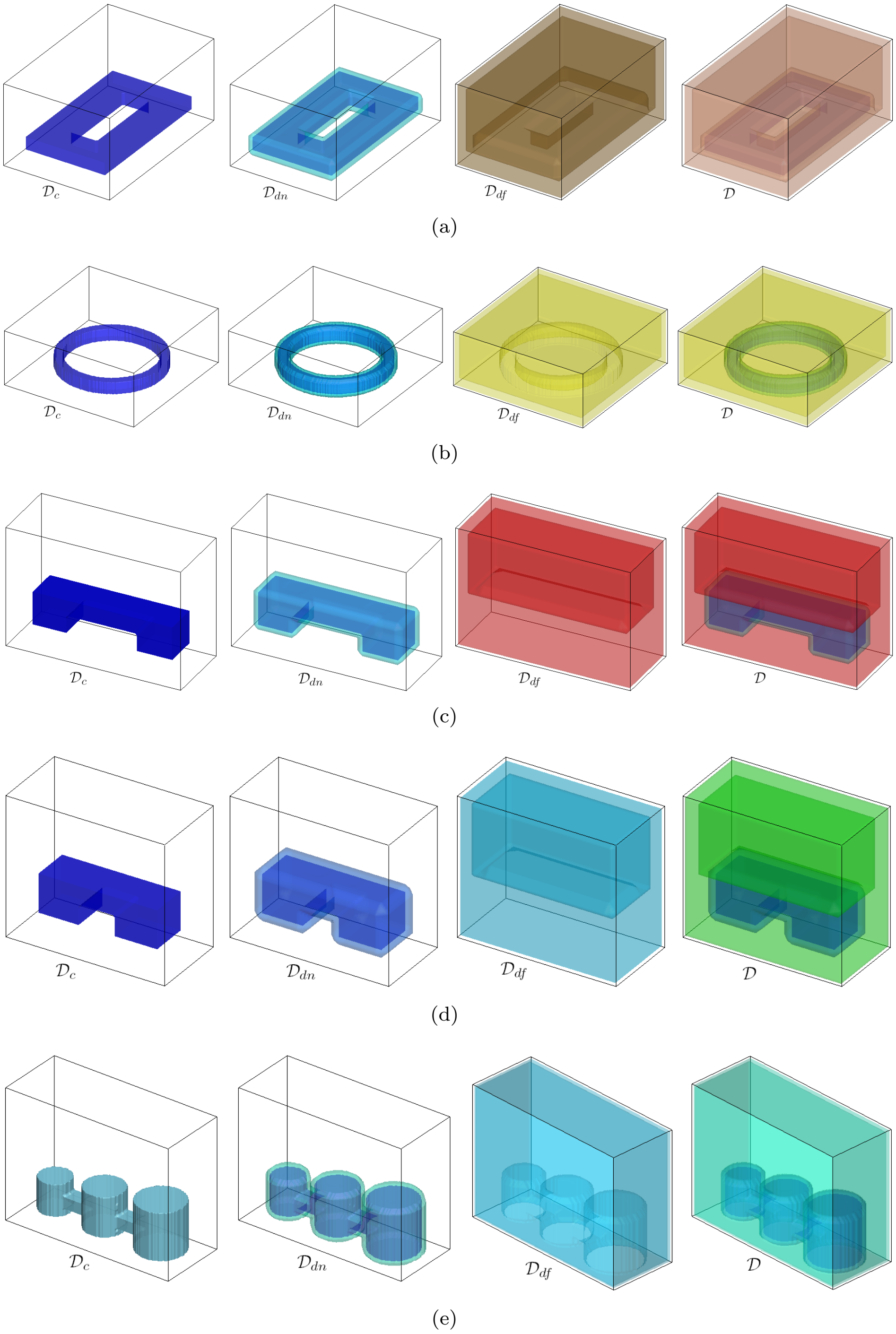
(b)

(c)

(d)

(e)

Figure 5. The initial setting for examples 1-5 (a-e) in this study.



(a)

(b)

(c)

(d)

(e)

FIGURE 6. The initial setting for examples 6-10 (a-e) in this study.

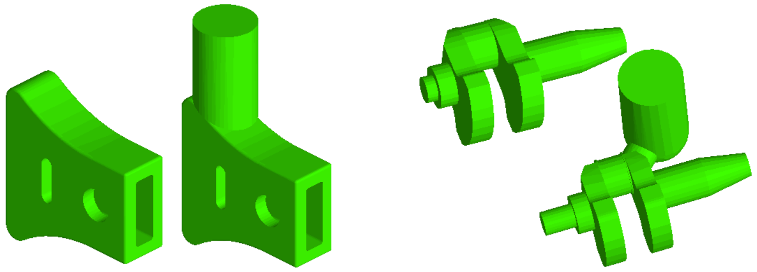

(a) (b)

Figure 7. The geometry steel hammer (a) and tiller crackshaft (b) together with a human designed feeding system.

Table 2. The optimization parameters corresponding to examples 1-15 includes the gird-size (Δx) in cm, mold dimensions ($l_x \times l_y \times l_z$) in gird-size units, thickness of feeder-neck design space (l_n) in gird-size units and $R_{dn} \times 100$.

#	Δx	$l_x \times l_y \times l_z$	L_{fn}	R_{dn}	#	Δx	$l_x \times l_y \times l_z$	L_{fn}	R_{dn}
1	1.00	100×70×70	4	1.75	9	1.00	70×40×100	3	3.00
2	1.00	80×120×120	4	7.00	10	1.00	70×40×80	3	4.00
3	1.00	80×50×85	4	1.75	11	1.00	120×65×160	4	3.50
4	1.00	60×40×110	3	3.50	12	0.50	100×100×120	2	1.80
5	1.00	60×120×120	3	7.00	13	0.50	160×60×160	2	1.80
6	1.00	60×80×120	4	10.00	14	0.50	100×100×120	2	1.80
7	1.00	70×160×160	4	20.00	15	0.33	120×120×120	6	1.75
8	1.00	70×160×160	4	1.75	-	-	-	-	-

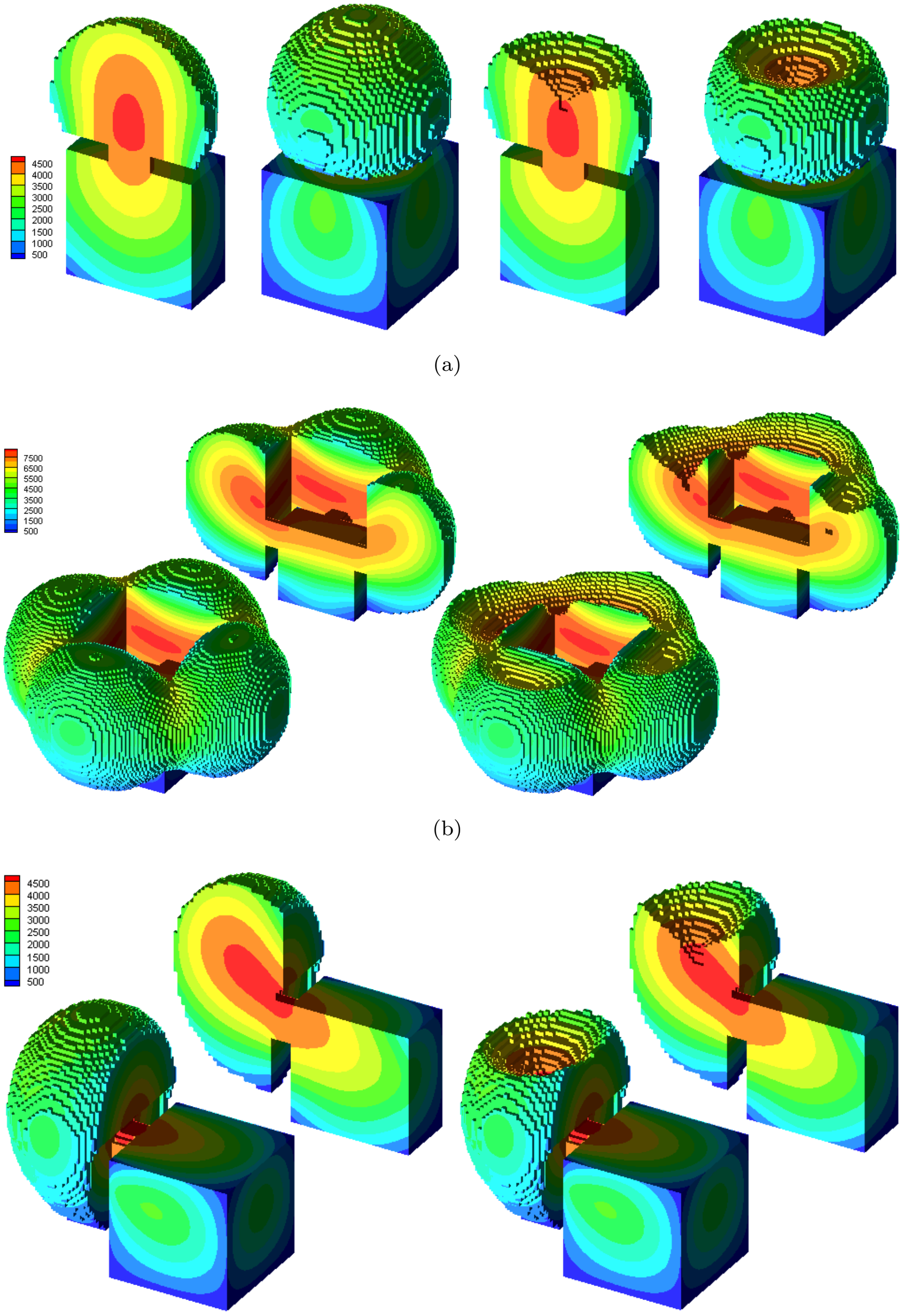
Table 3. The casting yield (in percent) and the total CPU time (in hour) corresponding to examples 1-15 in this study.

#	yield	cpu	#	yield	cpu	#	yield	cpu
1	56	2.8	6	51	2.3	11	51	24
2	12	15	7	40	12	12	63	28
3	45	4.2	8	55	9.1	13	61	23
4	65	1.4	9	61	8.3	14	62	18
5	61	2.6	10	73	8.8	15	62	31

to result of each case. Moreover to study whether the directional solidification is maintained, the contour plot of the local solidification time (t_s) is shown on the final topologies. The casting yield value[7] and total CPU time for each case are given in table 3.

According to results the directional solidification is established in all cases and final designs are sound under both the Niyama criterion and RGM. The design

[7] The casting yield is defined as the ration of casting weight (without feeding system) and total weight of consumed metal.

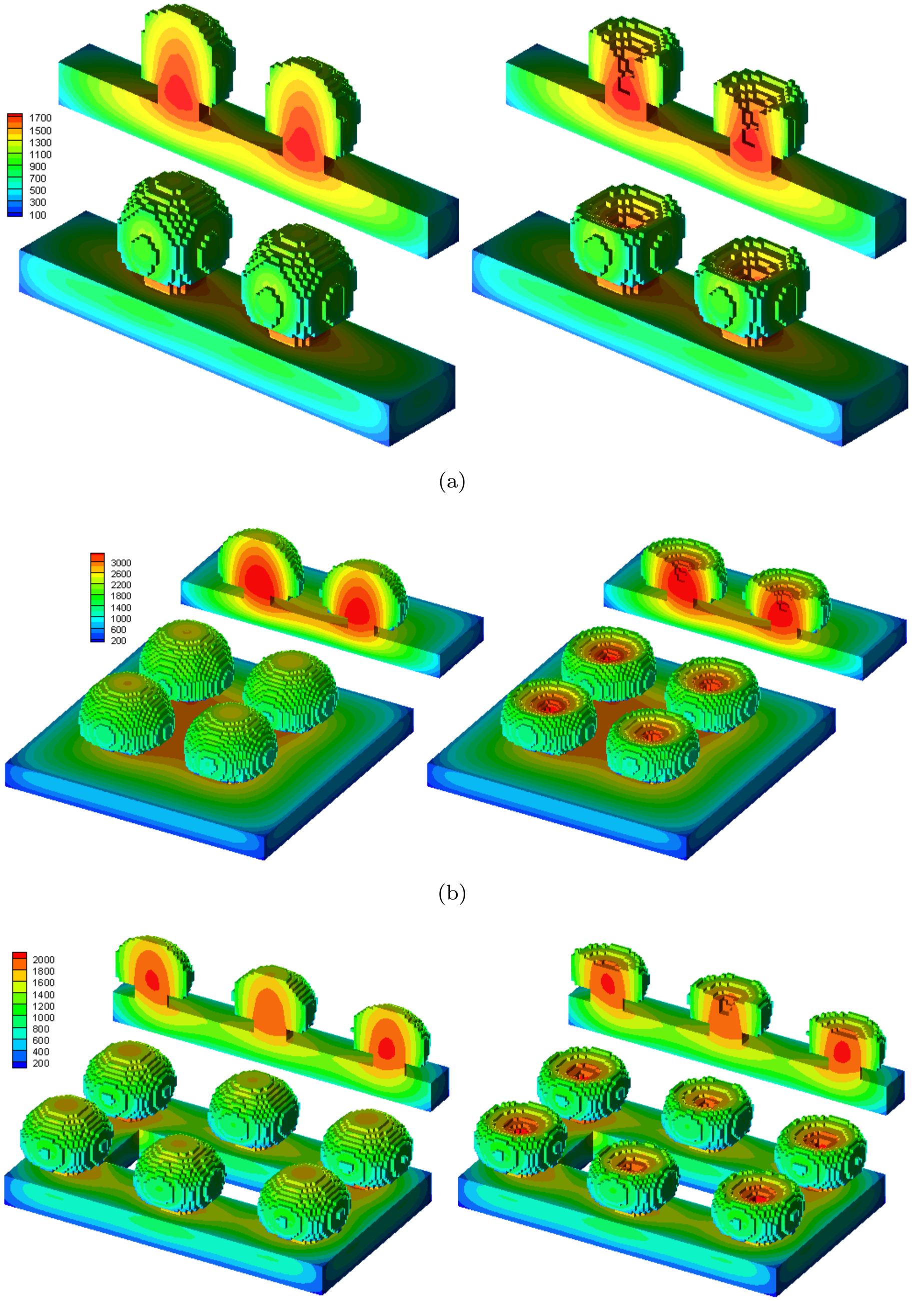


(a)

(b)

(c)

FIGURE 8. Examples 1-3 (a-c): the final topology (left) and results of RGM (right) together with contours t_s(s).

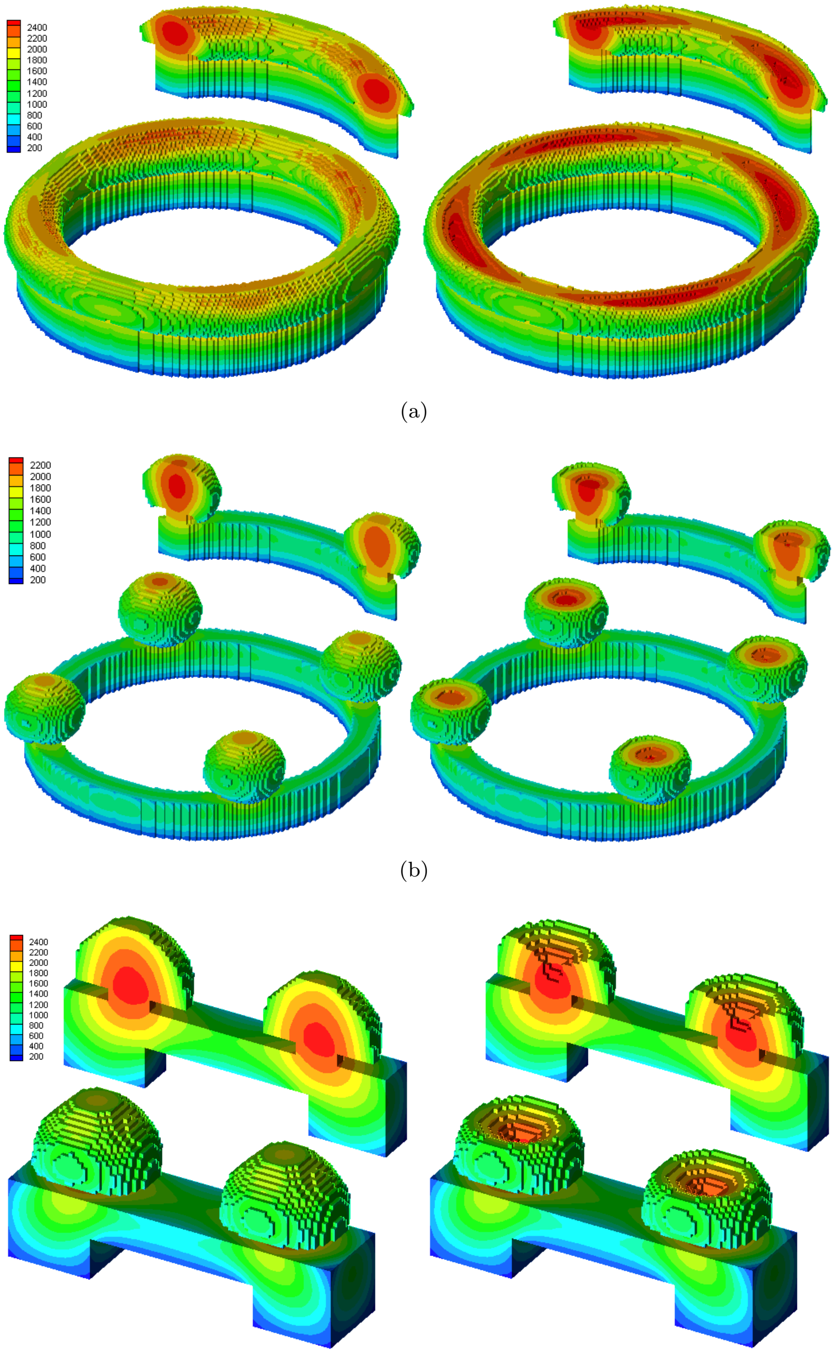


(a)

(b)

(c)

FIGURE 9. Examples 4-6 (a-c): the final topology (left) and results of RGM (right) together with contours t_s(s).

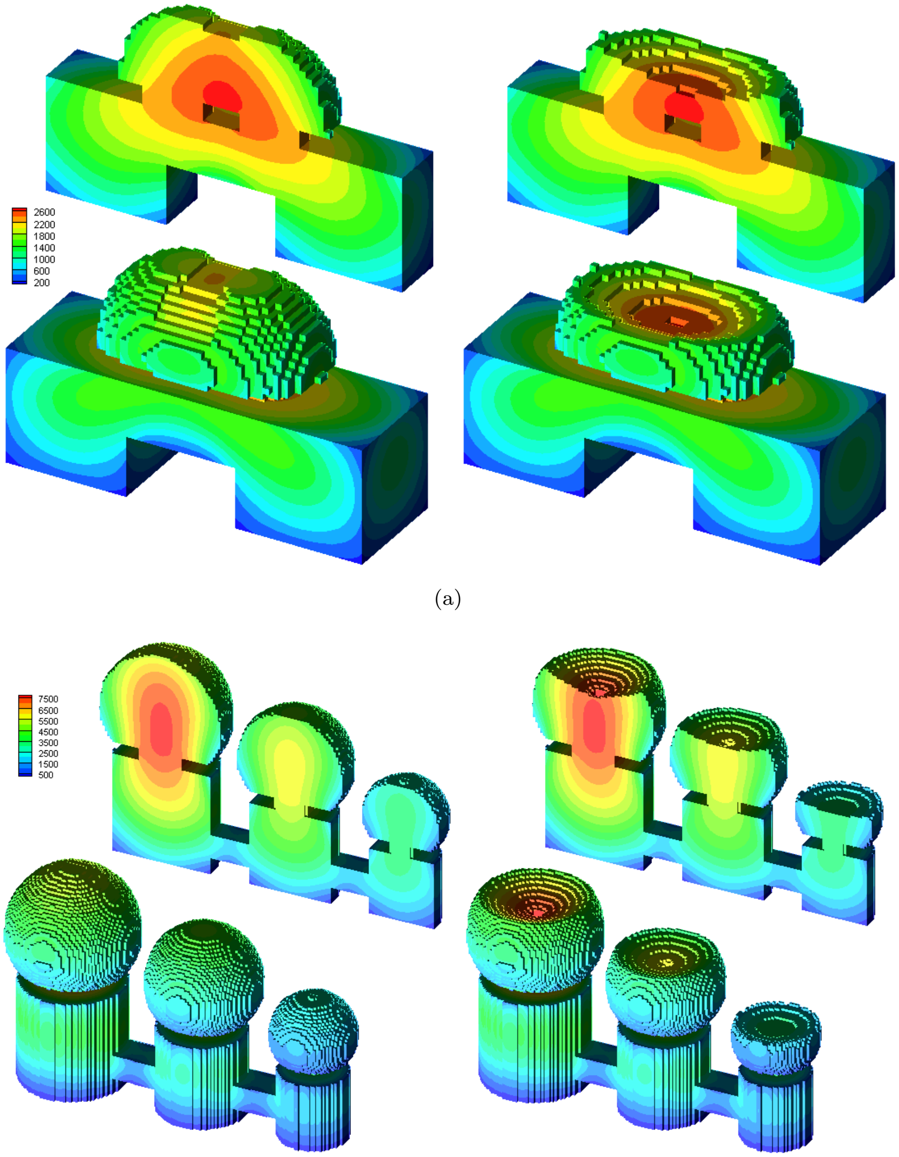


(a)

(b)

(c)

FIGURE 10. Examples 7-9 (a-c): the final topology (left) and results of RGM (right) together with contours t_s(s).

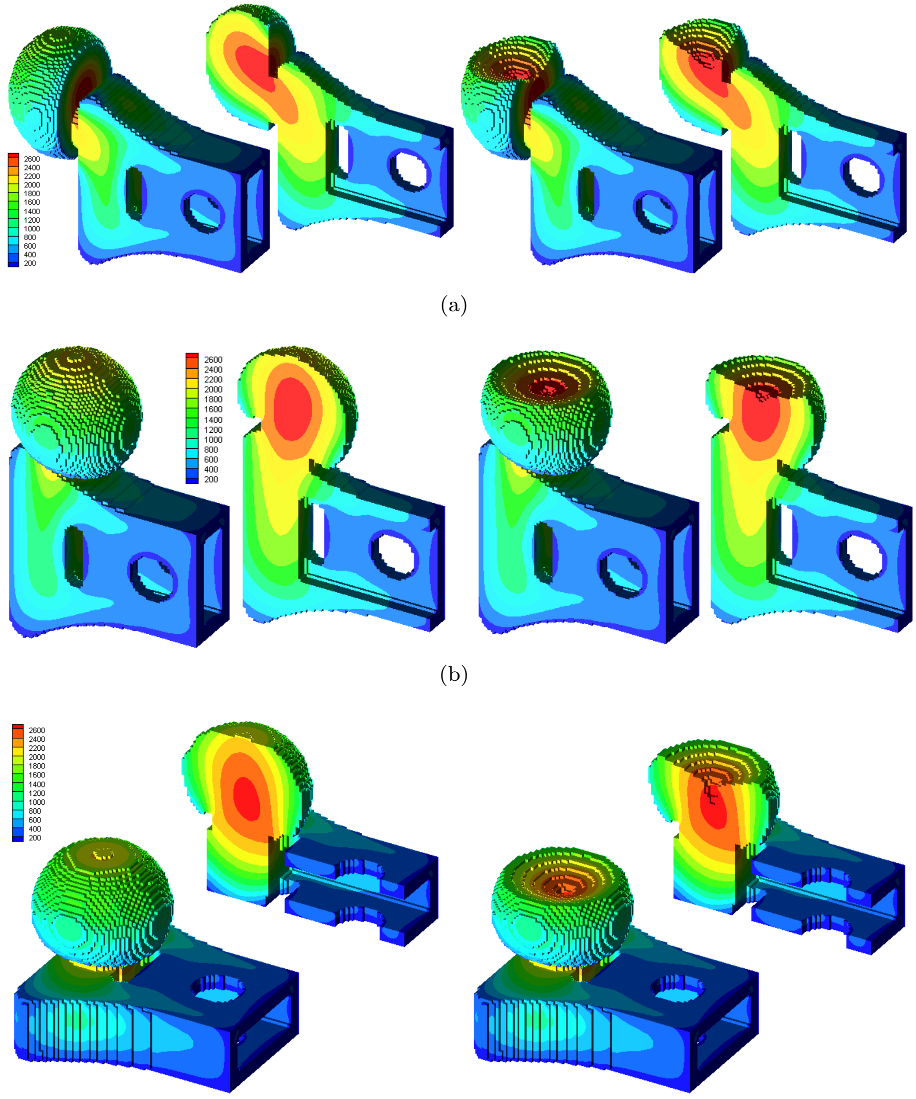


(a)

(b)

FIGURE 11. Examples 10-11 (a-b): the final topology (left) and results of RGM (right) together with contours t_s(s).

of riser and riser-neck are reasonable and manufacturable (ignoring the molding constraints that is not considered in the present work). The CPU times reported in table 3 shows surprising efficiency of the presented automatic design approach, makes it a reasonable tool to solve large scale real-world problems. Furthermore, results imply that the final designs are conservative with respect to RGM. We believe that this observation is due the fact that the effect of gravity (it aids defects



(a)

(b)

(c)

FIGURE 12. Examples 12-14 (a-c): the final topology (left) and results of RGM (right) together with contours t_s(s).

to migrate toward upper regions) is not considered in Niyama criterion. To improve the casting yield, a simple treatment is to reduce the critical value of Niyama criterion and repeat the analysis. Moreover, final topologies are sufficiently sharp showing success of the SIMP penalization in the present study.

Regarding to examples 2 and 3, we observe that both results are sound. However, the casting yield is considerably higher in example 3. This observation illustrates the dependency of solution to the initial setting (more precisely the definition of $\mathcal{D}_d$). This observation roots in the non-convexity of underlaying optimization problem.

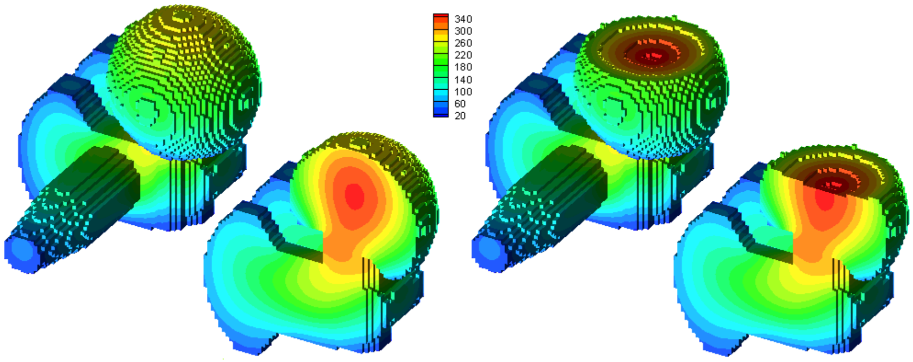


FIGURE 13. Examples 15: the final topology (left) and results of RGM (right) together with contours t_s(s).

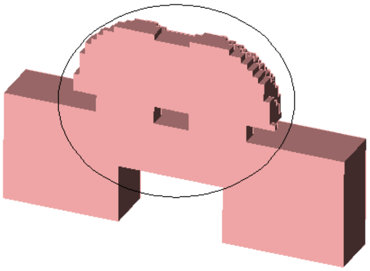
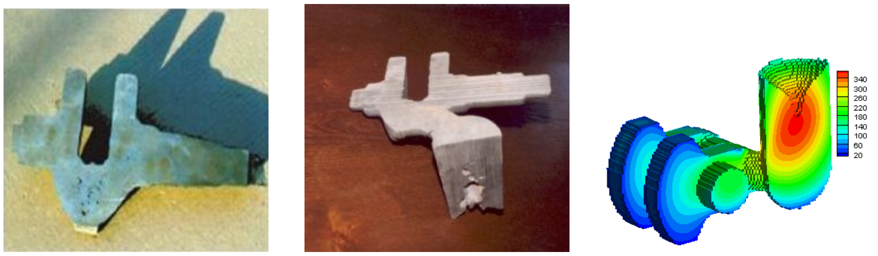


FIGURE 14. Formation of a big virtual riser by combination of the casting bridge and stretched riser.



(a) (b) (c)

FIGURE 15. Tiller crankshaft casting: (a) experimental result for traditional riser design, (b) experimental result for revised traditional design and (c) its corresponding defect analysis by RGM. The experimental results are adapted from the case study web page of SUTCAST code www.sutcast.com at July 2008.

Thus, a systematic change of $\mathcal{D}_d$ by user is suggested to explore alternative optimal solutions.

Comparing results of examples 7 and 8 show that selection of R_{dn} is an important factor to achieve reasonable results. Results of example 7 show that without volume faction limitation on the feeder-neck design domain, we will have the accumulation of materials near the casting surfaces which not only decreases the manufacturability of the final design but also decreases the casing yield.

Comparing results of the examples 9 and 10 shows that when we have two heavy sections connected by a narrow bridge, we will have multiple riser on top of heavy sections. Moreover, as the length of bridge is decreased, the riders approach to each other. However, the length of bridge is sufficiently small we will have only a riser concentrated on top of the connecting bridge which results in significant improvement of the casting yield. This result suggests a general design rule as follows:

Rule 11.1. *When we have two heavy sections connected by a narrow bridge with sufficiently small length, it is possible to design a small stretched riser concentrated on the bridge (cf. figure 14). In this manner the bridge together with the riser make a big virtual riser which provides the directional solidification. Using this strategy, we gain a significant improvement in the casting yield (see figure 14).*

The results of example 3 and 12 suggest a new riser design rule, can be stated as follows:

Rule 11.2. *Regarding to the design of (cylindrical-like) side-risers, it is preferable that the external boundaries of the riser in the vicinity of the casting be parallel to the casting surfaces (cf. figures 8.c and 12.a). It is clear that this simple rule improve the heat efficiency of the riser.*

The main reason for selection of example 15 in the present work was the complexity of designing an appropriate feeding system for this casting. This example is adapted from the case study web page of the commercial casting simulation code SUTCAST[8]. Although at the first glance, the design of feeding system for this part seems to be straightforward, is it a very challenging task in practice. This is mainly due to this fact that using the traditional design rules; which commonly suggest to use a side-riser connected to the casting axis with a right angle; always fails to produce a sound casting. The right way suggested by expert designers in SUTCAST design center is to use a side-riser connected to the casting axis by a non-right-angle riser-neck (cf. figure 7.b and 15). Figure 15 part a, b and c respectively, show the experimental results corresponding to traditional design (failed), the experimental results corresponding to revised design (succeed) and RGM results for the revised design. The results of example 15 in this study show the success presented approach for such a challenging test case. In particular, the casting yield corresponding to our final design is about 62% while that of human designed one is about 52%. Of course, including the molding constraints to our model may decrease our casting yield.

Another general and evident rule suggested by our results can be expressed as follows:

Rule 11.3. *The external surfaces of risers at the cast-mold interfaces should be preferably spherical-like to achieve the better heat efficiency.*

The topological symmetry inheritance is another important observation implied by our results. More precisely, we observe that the symmetric elements available

[8] www.sutcast.com.

in the initial setting of the optimal design problem is available in the final solution too. This observation suggest the following important conjecture:

Conjecture 11.4. (topological symmetry inheritance) Consider the following topology optimization problem defined on the spatial domain $\mathfrak{D} \in \mathbb{R}^d$ $(d = 1, 2, 3)$:

$$(\texttt{TP}) := \arg\min_{w\in\mathfrak{X}_w} \mathfrak{J}(w, u, \dot{u}, \nabla u) \quad \texttt{s.t.}: \quad \mathfrak{F}(w, u, \dot{u}, \nabla u, \nabla^2 u) = 0$$

where w denotes the control variable, $\mathfrak{X}_w$ denotes the control space, $\mathfrak{J}$ denotes the (sufficiently smooth) objective functional, the operator $\mathfrak{F}$ denotes a well-defined second order partial differential equation in which the coefficients and source terms are polynomial functions of w and u. Assume problem (TP) admits at least an optimal solution. Lets to denote by $\mathcal{S}_d$, $\mathcal{S}_x$, $\mathcal{S}_{w_0}$ the set of spatial symmetry elements corresponding to spatial domain $\mathfrak{D}$, control space $\mathfrak{X}_w$ and the initial distribution of w-field. Moreover assume $\mathcal{S}_u$ denotes the set of spatial symmetry elements corresponding to the solution of $\mathfrak{F}$ for w-field equal to the initial w distribution $(w = w_0)$. It is easy to see that when w_0 is uniform, then $\mathcal{S}_u$ is determined by available spatial symmetry elements in the source terms, initial and boundary conditions corresponding to $\mathfrak{F}$. Assume w^* denotes an optimal solution to (TP) resulted from a globally convergent first order gradient descent method. Moreover assume the set of spatial symmetry elements corresponding to w^* is denoted by $\mathcal{S}_{w^*}$, then the following identity holds,

$$\mathcal{S}_s \subseteq \mathcal{S}_{w^*}$$

where,

$$\mathcal{S}_s := \mathcal{S}_d \cap \mathcal{S}_x \cap \mathcal{S}_{w_0} \cap \mathcal{S}_u.$$

Remark 11.5. According to conjecture 11.4, to find alternative optimal solutions (toward finding the global solution), a reasonable strategy is to change the initial setting in a way that that $\mathcal{S}_s$ is also altered (compare examples 2 and 3)

12. Summary

The optimal design of feeding system in the shape casting process is mathematically formulated as a constrained minimum weight topology optimization problem. The presented model includes infinite number of control parameters and nonlinear state-dependent space-time local constraints. The regularization and relaxation of the original problem is presented. Exploiting the specific structure of the optimization problem, a highly efficient solution strategy is presented, which makes it possible to solve large-scale realistic problems. The efficiency and success of the presented approach is shown by solving 15 different test cases. We believe that including the molding constraints to the presented method, makes a complete solution to the automatic optimal riser design problem in the shape casting process. An important property of the underlaying topology optimization problem in this study is related to its non-trivial structure in the sense that the design space does not match to the objective space. In fact, in this way one can change the distribution of material within a domain to pose control on a different domain. This opens room to extend the utility of the topology optimization approach to solve new classes of engineering problems. Moreover, parts of the presented method can be equivalently applied to other kinds of topology optimization problems.

Acknowledgment

The author would like to thank Wolfgang Bangerth for his constructive comments in the course of this research. The geometries of hammer and crankshaft used in the present work were provided by SUTCAST developing team. The authors would like to thank the Research Office of Sharif University of Technology for the financial support of this project under grant number QA980815.

References

[1] JM Aguirregabiria, A. Hernández, and M. Rivas. δ-function converging sequences. *Amer. J. Phys.*, 70:180, 2002.

[2] G. Allaire. *Numerical analysis and optimization: an introduction to mathematical modelling and numerical simulation.* Translated by: Craig, A., Oxford University Press, USA, 2007.

[3] G. Allaire, F. Jouve, and A. Toader. Structural optimization using sensitivity analysis and a level-set method. *J. Comput. Phys.*, 194(1):363–393, 2004.

[4] J. Barzilai and J.M. Borwein. Two-point step size gradient methods. *IMA J. Numer. Anal.*, 8(1):141, 1988.

[5] R. Becker, D. Meidner, and B. Vexler. Efficient numerical solution of parabolic optimization problems by finite element methods. *Optim. Methods Softw.*, 22(5):813–833, 2007.

[6] J.C. Bellido, A. Donoso, and P. Pedregal. Optimal design in conductivity under locally constrained heat flux. *Arch. Ration. Mech. Anal.*, 195(1):333–351, 2010.

[7] M.P. Bendsoee and O. Sigmund. *Topology Optimization: theory, methods and applications.* Springer, 2004.

[8] E.G. Birgin, J.M. Martínez, and M. Raydan. Nonmonotone Spectral Projected Gradient Methods on Convex Sets. *SIAM J. Optim.*, 10:1196, 2000.

[9] E.G. Birgin, J.M. Martínez, and M. Raydan. Algorithm 813: SPG–software for convex-constrained optimization. *ACM TOMS*, 27(3):340–349, 2001.

[10] G.P. Borikar and S.T. Chavan. Optimization of casting yield in multi-cavity sand moulds of al-alloy components. *Materials Today: Proceedings*, 2020.

[11] C.M. Choudhari, B.E. Narkhede, and S.K. Mahajan. Casting design and simulation of cover plate using autocast-x software for defect minimization with experimental validation. *Procedia Materials Science*, 6:786–797, 2014.

[12] C. Dong, X. Shen, J. Zhou, T. Wang, and Y. Yin. Optimal design of feeding system in steel casting by constrained optimization algorithms based on intecast. *China Foundry*, 13(6):375–382, 2016.

[13] M. Fu, A. Nee, and J. Fuh. The application of surface visibility and moldability to parting line generation. *Comput. Aided Des.*, 34(6):469–480, 2002.

[14] C.P. Hong. *Computer Modelling of Heat and Fluid Flow in Materials Processing.* CRC Press, 2004.

[15] H. Hussain and A. Khandwawala. Optimal design of feeder for sand casted steel dumbbell: Simulation studies for techno-economic feasibility. *American J. Mech. Eng.*, 2(3):93–98, 2014.

[16] I. Imafuku and K. Chijiiwa. A Mathematical Model for Shrinkage Cavity Prediction in Steel Castings. *AFS Transactions*, 91:527–540, 1983.

[17] E. Jacob, D.S. Chiniwar, S. Savitri, M. Manoj, and R. Sasikumar. Simulation-based feeder design for metal castings. *Indian Foundry Journal*, 59(12):39–44, 2013.

[18] R.V. Kohn and G. Strang. Optimal design and relaxation of variational problems, I-III. *Comm. Pure Appl. Math.*, 39:113–137, 139–182, 353–377, 1986.

[19] C. Le, J. Norato, T. Bruns, C. Ha, and D. Tortorelli. Stress-based topology optimization for continua. *Struct. Multidiscip. Optim.*, 41(4):605–620, 2010.

[20] R.W. Lewis, R.S. Ransing, W.K.S. Pao, K. Kulasegaram, and J. Bonet. Alternative techniques for casting process simulation. *Internat. J. Numer. Methods Heat Fluid Flow*, 14(2):145–166, 2004.

[21] T.E. Morthland, P.E. Byrne, D.A. Tortorelli, and J.A. Dantzig. Optimal Riser Design for Metal Castings. *Metall. Mater. Trans. B*, 26(1-2):871–885, 1995.

[22] A. Munch, P. Pedregal, and F. Periago. Relaxation of an optimal design problem for the heat equation. *Journal de Mathématiques Pures et Appliqués*, 89(3):225–247, 2008.

[23] E. Niyama, T. Uchida, M. Morikawa, and S. Saito. A method of shrinkage prediction and its application to steel casting practice. *AFS Int. Cast Metals Journal*, 7(3):52–63, 1982.

[24] S. Osher and R.P. Fedkiw. *Level set methods and dynamic implicit surfaces.* Springer Verlag, 2003.

[25] Rajesh S Ransing, Minkesh P Sood, and WKS Pao. Computer implementation of heuvers' circle method for thermal optimisation in castings. *Int. J. Cast Met. Res.*, 18(2):119–128, 2005.

[26] R. Tavakoli. CartGen: Robust, efficient and easy to implement uniform/octree/embedded boundary Cartesian grid generator. *Int. J. Numer. Meth. Fluids*, 57(12):1753–1770, 2008.

[27] R. Tavakoli. Multimaterial topology optimization by volume constrained allen–cahn system and regularized projected steepest descent method. *Comput. Meth. in Appl. Mech. Eng.*, 276:534–565, 2014.

[28] R. Tavakoli. On the prediction of shrinkage defects by thermal criterion functions. *Int. J. Adv. Manuf. Tech.*, 74(1-4):569–579, 2014.

[29] R. Tavakoli. Computationally efficient approach for the minimization of volume constrained vector-valued ginzburg–landau energy functional. *J. Comput. Phys.*, 295:355–378, 2015.

[30] R. Tavakoli. On the coupled continuous knapsack problems: projection onto the volume constrained gibbs n-simplex. *Optim. Lett.*, 10(1):137–158, 2016.

[31] R. Tavakoli. Optimal design of multiphase composites under elastodynamic loading. *Comput. Meth. in Appl. Mech. Eng.*, 300:265–293, 2016.

[32] R. Tavakoli and P. Davami. Optimal feeder design in sand casting process by growth method. *Int. J. Cast Met. Res.*, 20(5):288–296, 2007.

[33] R. Tavakoli and P. Davami. A fast method for numerical simulation of casting solidification. *Comm. Numer. Methods Engrg.*, 24(12):1723–1740, 2008.

[34] R. Tavakoli and P. Davami. Automatic optimal feeder design in steel casting process. *Comput. Meth. Appl. Mech. Eng.*, 197(9-12):921–932, 2008.

[35] R. Tavakoli and P. Davami. Optimal riser design in sand casting process by topology optimization with SIMP method I: Poisson approximation of nonlinear heat transfer equation. *Struct. Multidiscip. Optim.*, 36(2):193–202, 2008.

[36] R. Tavakoli and P. Davami. Feeder growth: a new method for automatic optimal feeder design in gravity casting processes. *Struct. Multidis. Optim.*, 39(5):519, 2009.

[37] R. Tavakoli and P. Davami. Optimal riser design in sand casting process with evolutionary topology optimization. *Struct. Multidis. Optim.*, 38(2):205–214, 2009.

[38] R. Tavakoli and H. Zhang. A nonmonotone spectral projected gradient method for large-scale topology optimization problems. *Numer. Algebra Con. Optim.*, 2(2):395–412, 2012.

[39] DA Tortorelli and P. Michaleris. Design sensitivity analysis: overview and review. *Inverse Probl. Sci. Eng.*, 1(1):71–105, 1994.

[40] D.A. Tortorelli, J.A. Tomasko, T.E. Morthland, and J.A. Dantzig. Optimal design of nonlinear parabolic systems. Part II: Variable spatial domain with applications to casting optimization. *Comput. Methods Appl. Mech. Engrg.*, 113(1-2):157–172, 1994.

[41] M. Ulbrich. Semismooth Newton Methods for Operator Equations in Function Spaces. *SIAM Journal on Optimization*, 13:805, 2002.

[42] M. Ulbrich and S. Ulbrich. Superlinear Convergence of Affine-Scaling Interior-Point Newton Methods for Infinite-Dimensional Nonlinear Problems with Pointwise Bounds. *SIAM J. Control Optim.*, 38:1938, 2000.

[43] VR Voller. Numerical treatment of rapidly changing and discontinuous conductivities. *International journal of heat and mass transfer*, 44(23):4553–4556, 2001.