A Unified Approach to Shape and Topological Sensitivity Analysis of Discretized Optimal Design Problems

P. Gangl[1,2,3] · **M. H. Gfrerer**[4]

Abstract

We introduce a unified sensitivity concept for shape and topological perturbations and perform the sensitivity analysis for a discretized PDE-constrained design optimization problem in two space dimensions. We assume that the design is represented by a piecewise linear and globally continuous level set function on a fixed finite element mesh and relate perturbations of the level set function to perturbations of the shape or topology of the corresponding design. We illustrate the sensitivity analysis for a problem that is constrained by a reaction–diffusion equation and draw connections between our discrete sensitivities and the well-established continuous concepts of shape and topological derivatives. Finally, we verify our sensitivities and illustrate their application in a level-set-based design optimization algorithm where no distinction between shape and topological updates has to be made.

✉ M. H. Gfrerer
gfrerer@tugraz.at

1 Johann Radon Institute for Computational and Applied Mathematics, Altenberger Straße 69, 4040 Linz, Austria

2 Institute of Applied Mathematics, Graz University of Technology, Steyrergasse 30/III, 8010 Graz, Austria

3 Chair of Applied Mathematics (Continuous Optimization), Friedrich Alexander University Erlangen-Nürnberg, Cauerstraße 11, 91058 Erlangen, Germany

4 Institute of Applied Mechanics, Graz University of Technology, Technikerstrasse 4, 8010 Graz, Austria

1 Introduction

Numerical methods for the design optimization of technical systems are of great interest in science and engineering. Applications include the optimization of mechanical structures [2, 29], electromagnetic devices [5, 19], fluid flow [21], heat dissipation [20] and many more. There exist several different approaches to computational design optimization. On the one hand, shape optimization techniques based on the mathematical concept of shape derivatives [15] can modify boundaries and material interfaces in a smooth way, but typically cannot alter the topology of a design. An exception being the level set method for shape optimization [2] where the design is represented by the zero level set of a design function whose evolution is guided by shape sensitivities via a transport equation. While this approach allows for merging of components, it lacks a nucleation mechanism and is often coupled with the topological derivative concept [26, 30], see e.g. [1, 10]. For a computationally-oriented introduction to these classical level-set based methods, we refer the interested reader to [12]. In the class of density-based topology optimization methods [7], a design is represented by a density function $\rho(x)$ that is allowed to attain any value in the interval [0, 1]. Then, regions with $\rho(x) = 0$ and $\rho(x) = 1$ are interpreted as occupied by material 1 and 2, respectively, while intermediate density values $0 < \rho(x) < 1$ are penalized in order to obtain designs that are almost "black-and-white". One advantage of density based methods is that the system response depends continuously on ρ and the standard notions of derivatives in vector spaces can be applied. Here, interfaces are typically not crisp and there is no measure of optimality with respect to shape variations at the interface. Finally we mention the level-set algorithm for topology optimization introduced in [4], where the design is guided solely by the topological derivative, which however is not defined on the material interfaces. As a consequence, the final designs cannot be shown to be optimal with respect to shape variations at the interface. This aspect has been thoroughly analyzed in [6] where the authors draw a connection to density-based methods and, for two particular problem classes, propose an interpolation scheme which relates the derivative with respect to the density function, to topological and shape derivatives in the interior and on the interface, respectively.

The goal of this paper is to unify the concepts of topological and shape perturbations and to treat design optimization problems by a unified sensitivity, called the *topological-shape derivative*. In this way, we aim at combining topological sensitivity information (related to the topological derivative) in the interior of each subdomain and shape sensitivity information (related to the shape derivative) at the material interface. While the topological derivative is defined as the sensitivity of a design-dependent cost function with respect to the introduction of a small hole or inclusion of different material, the shape derivative is defined as the cost function's sensitivity with respect to a transformation of the domain. In order to unify these two concepts, we consider a domain description by means of a continuous level set function which attains positive values in one of two subdomains and negative values in the other. Then a perturbation of the level set function in the interior of a subdomain can be related to a topological perturbation, and a perturbation close to the material interface can be seen as a perturbation of the shape of the domain. We remark that this point of view is in alignment with the concept of dilations of points and curves as introduced in [13],

see also [14]. In principle, this idea was already followed in [9] and also in the recent work [8] where the discrete shape sensitivity analysis for perturbations of boundaries represented by level set functions is carried out and compared to the classical approach where domains are perturbed by transformations rather than dilations. In these works, however, only the case of shape optimization and no combination with topology optimization is considered. In [22] the author represents domains by level set functions and relates shape and topological derivatives of shape functionals to derivatives with respect to the level set function in a continuous setting without PDE constraints. In contrast to this, we consider PDE-constrained problems, but our analysis is performed on the discrete level, i.e. we follow the paradigm "discretize-then-optimize" for our sensitivity analysis with respect to a level set function.

The rest of this paper is organized as follows: In Sect. 2, we introduce the model problem and the classical concepts of topological and shape derivative in the continuous setting. After presenting the discretized setting in Sect. 3, we proceed to compute the numerical topological-shape derivative of our discretized model problem in Sect. 4. In Sect. 5 we compare the computed sensitivities with the sensitivities obtained by discretizing the continuous formulas. Finally we verify our computed formulas and present optimization results in Sect. 6 before giving a conclusion in Sect. 7.

2 Model Problem and Continuous Setting

Let D be a given, fixed, open and bounded hold-all domain and $\Omega \subset D$ an open and measurable subset. Let the boundary of D be decomposed into $\Gamma_D, \Gamma_N \subset D$ with $\overline{\Gamma}_D \cup \overline{\Gamma}_N = \partial D$ and $\Gamma_D \cap \Gamma_N = \emptyset$. In the present paper, we consider a topology optimisation problem with a tracking type cost function

$$g(\Omega, u) = c_1 |\Omega| + c_2 \int_D \tilde{\alpha}_\Omega |u - \hat{u}|^2 \, \mathrm{d}x \tag{1}$$

where $\hat{u} \in H^1(D)$ is a given desired state, and $c_1, c_2 \in \mathbb{R}$ are given constants. The continuous topology optimization problem reads

$$\min_{\Omega \in \mathcal{A}} \; g(\Omega, u), \tag{2a}$$

$$\text{subject to}$$

$$-\lambda_\Omega \Delta u + \alpha_\Omega u = f_\Omega \qquad \text{in } D, \tag{2b}$$

$$u = g_D \qquad \text{on } \Gamma_D, \tag{2c}$$

$$\lambda_\Omega \partial_n u = g_N \qquad \text{on } \Gamma_N, \tag{2d}$$

where

$$\tilde{\alpha}_\Omega(x) = \chi_\Omega(x)\tilde{\alpha}_1 + \chi_{D\setminus\Omega}(x)\tilde{\alpha}_2, \qquad \lambda_\Omega(x) = \chi_\Omega(x)\lambda_1 + \chi_{D\setminus\Omega}(x)\lambda_2,$$

$$\alpha_\Omega(x) = \chi_\Omega(x)\alpha_1 + \chi_{D\setminus\Omega}(x)\alpha_2, \qquad f_\Omega(x) = \chi_\Omega(x)f_1 + \chi_{D\setminus\Omega}(x)f_2,$$

for some constants $\lambda_1, \lambda_2 > 0$, $\alpha_1, \alpha_2, \tilde{\alpha}_1, \tilde{\alpha}_2 \geq 0$ and $f_1, f_2 \in \mathbb{R}$ with χ_S the characteristic function of a set S,

$$\chi_S(x) = \begin{cases} 1, & x \in S, \\ 0, & \text{else.} \end{cases}$$

Here, $\mathcal{A}$ denotes a set of admissible subsets of D, and the data $g_D \in H^{1/2}(\Gamma_D)$, $g_N \in L^2(\Gamma_N)$ are given. The weak formulation of the PDE constraint reads

$$\text{Find } u \in V_g := \{v \in H^1(D) : v = g_D \text{ on } \Gamma_D\} \text{ such that}$$

$$\int_D \lambda_\Omega \nabla u \cdot \nabla v + \alpha_\Omega u v \, dx = \int_D f_\Omega v \, dx \quad \text{for all } v \in V_0 \tag{3}$$

with $V_0 = \{v \in H^1(D) : v = 0 \text{ on } \Gamma_D\}$. We assume that either $|\Gamma_D| > 0$ or $\alpha_1, \alpha_2 > 0$ such that, for given $\Omega \in \mathcal{A}$, (3) admits a unique solution which we denote by $u(\Omega)$. We introduce the reduced cost function $\mathfrak{g}(\Omega) := g(\Omega, u(\Omega))$.

2.1 Classical Topological Derivative

Let $\omega \subset \mathbb{R}^d$ with $0 \in \omega$. For a point $z \in \Omega \cup (D \backslash \overline{\Omega})$, let $\omega_\varepsilon := z + \varepsilon \omega$ denote a perturbation of the domain around z of (small enough) size ε and of shape ω. The continuous topological derivative of the shape function $\mathfrak{g} = \mathfrak{g}(\Omega)$ is defined by

$$d_T \mathfrak{g}(\Omega)(z) = \begin{cases} \lim_{\varepsilon \searrow 0} \frac{\mathfrak{g}(\Omega \cup \omega_\varepsilon) - \mathfrak{g}(\Omega)}{|\omega_\varepsilon|} & \text{for } z \in D \setminus \overline{\Omega}, \\ \lim_{\varepsilon \searrow 0} \frac{\mathfrak{g}(\Omega \backslash \overline{\omega}_\varepsilon) - \mathfrak{g}(\Omega)}{|\omega_\varepsilon|} & \text{for } z \in \Omega. \end{cases} \tag{4}$$

Note that the topological derivative is not defined for points $z \in \partial \Omega$ on the material interface. For problem (2) we obtain for $z \in D \backslash \overline{\Omega}$

$$
\begin{aligned}
d_T \mathfrak{g}(\Omega)(z) = &c_1 + c_2(\tilde{\alpha}_1 - \tilde{\alpha}_2)(u(z) - \hat{u}(z))^2 \\
&+ 2\lambda_2 \frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2} \nabla u(z) \cdot \nabla p(z) + (\alpha_1 - \alpha_2)u(z)p(z) - (f_1 - f_2)p(z),
\end{aligned}
\tag{5}
$$

whereas for $z \in \Omega$

$$
\begin{aligned}
d_T \mathfrak{g}(\Omega)(z) = &-c_1 + c_2(\tilde{\alpha}_2 - \tilde{\alpha}_1)(u(z) - \hat{u}(z))^2 \\
&+ 2\lambda_1 \frac{\lambda_2 - \lambda_1}{\lambda_1 + \lambda_2} \nabla u(z) \cdot \nabla p(z) + (\alpha_2 - \alpha_1)u(z)p(z) - (f_2 - f_1)p(z),
\end{aligned}
\tag{6}
$$

see, e.g. [3].

2.2 Classical Shape Derivative

We recall the definition of the classical shape derivative as well as its formula for our model problem (1)–(2). Given an admissible shape $\Omega \in \mathcal{A}$ and a smooth vector field $V \in C_c^\infty(D)$ that is compactly supported in D, we define the perturbed domain

$$\Omega_t = (\mathrm{id} + t V)(\Omega), \tag{7}$$

for a small perturbation parameter $t > 0$ where $\mathrm{id} : \mathbb{R}^d \to \mathbb{R}^d$ denotes the identity operator. The classical shape derivative of $\mathfrak{g}$ at Ω with respect to V is then given by

$$d_S \mathfrak{g}(\Omega)(V) = \lim_{t \searrow 0} \frac{\mathfrak{g}(\Omega_t) - \mathfrak{g}(\Omega)}{t} \tag{8}$$

if this limit exists and the mapping $V \mapsto d_S \mathfrak{g}(\Omega)(V)$ is linear and continuous. Under suitable assumptions it can be shown that this shape derivative admits the tensor representation

$$d_S \mathfrak{g}(\Omega)(V) = \int_D \mathcal{S}_1^\Omega : \partial V + \mathcal{S}_0^\Omega \cdot V \, dx, \tag{9}$$

for some tensors $\mathcal{S}_0^\Omega \in L^1(D, \mathbb{R}^d)$, $\mathcal{S}_1^\Omega \in L^1(D, \mathbb{R}^{d \times d})$ [23]. Here, ∂V denotes the Jacobian of the vector field V. The structure theorem of Hadamard-Zolésio [15, pp. 480–481] states that under certain smoothness assumptions the shape derivative of a shape function $\mathfrak{g}$ with respect to a vector field V can always be written as an integral over the boundary of a scalar function L times the normal component of V, i.e.,

$$d_S \mathfrak{g}(\Omega)(V) = \int_{\partial \Omega} L \, (V \cdot n) \, \mathrm{d}s \tag{10}$$

where n denotes the unit normal vector pointing out of Ω. For problem (2) one obtains [23]

$$\begin{aligned}
\mathcal{S}_1^\Omega &= (c_1 \chi_\Omega + c_2 \tilde{\alpha}_\Omega (u - \hat{u})^2 + \lambda_\Omega \nabla u \cdot \nabla p + \alpha_\Omega u p - f_\Omega p) I \\
&\quad - \lambda_\Omega \nabla u \otimes \nabla p - \lambda_\Omega \nabla p \otimes \nabla u, \tag{11} \\
\mathcal{S}_0^\Omega &= -2 \tilde{\alpha}_\Omega (u - \hat{u}) \nabla \hat{u} \tag{12}
\end{aligned}$$

where $I \in \mathbb{R}^{d,d}$ denotes the identity matrix, and

$$\begin{aligned}
L &= [(\mathcal{S}_1^{\Omega,\mathrm{in}} - \mathcal{S}_1^{\Omega,\mathrm{out}}) n] \cdot n \\
&= c_1 + c_2 (\tilde{\alpha}_1 - \tilde{\alpha}_2)(u - \hat{u})^2 + (\alpha_1 - \alpha_2) u p - (f_1 - f_2) p \\
&\quad + (\lambda_1 - \lambda_2)(\nabla u \cdot \tau)(\nabla p \cdot \tau) - \left(\frac{1}{\lambda_1} - \frac{1}{\lambda_2} \right) (\lambda_\Omega \nabla u \cdot n)(\lambda_\Omega \nabla p \cdot n).
\end{aligned}$$

Here, $\mathcal{S}_1^{\Omega,\text{in}}$ and $\mathcal{S}_1^{\Omega,\text{out}}$ denote the restrictions of the tensor $\mathcal{S}_1^\Omega$ to Ω and $D\backslash\Omega$, respectively. Furthermore, for two column vectors $a, b \in \mathbb{R}^d$, $a \otimes b = ab^\top \in \mathbb{R}^{d\times d}$ denotes their outer product, τ denotes the tangential vector and $p \in H_0^1(D)$ is the solution to the adjoint equation

$$\int_D \lambda_\Omega \nabla v \cdot \nabla p + \alpha_\Omega v p \, \mathrm{d}x = -2c_2 \int_D \tilde{\alpha}_\Omega(u - \hat{u})v \, \mathrm{d}x \quad \text{for all } v \in H_0^1(D).$$

Moreover, motivated by the definition of the topological derivative (4) with the volume of the difference of the perturbed and unperturbed domains in the denominator, we introduce the alternative definition of a shape derivative

$$\hat{d}_S \mathfrak{g}(\Omega)(V) = \lim_{t\searrow 0} \frac{\mathfrak{g}(\Omega_t) - \mathfrak{g}(\Omega)}{|\Omega_t \triangle \Omega|}, \tag{13}$$

with the symmetric difference of two sets $A\triangle B := (A\backslash\overline{B}) \cup (B\backslash\overline{A})$. Note that the volume of the symmetric difference in (13) can be written as

$$|\Omega_t \triangle \Omega| = \int_D |\chi_{\Omega_t} - \chi_\Omega| \, \mathrm{d}\mathbf{x}. \tag{14}$$

Lemma 1 *Let Ω and V smooth. It holds*

$$\lim_{t\searrow 0} \frac{1}{t} |\Omega_t \triangle \Omega| = \int_{\partial\Omega} |V \cdot n| \, dS_x.$$

The proof is given in Appendix A.1. From Lemma 1, we immediately obtain the following relation between $d_S \mathfrak{g}(\Omega)(V)$ and $\hat{d}_S \mathfrak{g}(\Omega)(V)$.

Corollary 1 *Suppose that $\mathfrak{g}$ is shape differentiable at Ω and that Ω and V are smooth and $\int_{\partial\Omega} |V \cdot n| \, dS_x > 0$. Then it holds*

$$\hat{d}_S \mathfrak{g}(\Omega)(V) = \frac{d_S \mathfrak{g}(\Omega)(V)}{\int_{\partial\Omega} |V \cdot n| \, dS_x}. \tag{15}$$

Proof This follows immediately from the definition of $\hat{d}_S \mathfrak{g}(\Omega)(V)$ by Lemma 1 since

$$\hat{d}_S \mathfrak{g}(\Omega)(V) = \lim_{t\searrow 0} \frac{\mathfrak{g}(\Omega_t) - \mathfrak{g}(\Omega)}{|\Omega_t \triangle \Omega|} = \frac{\lim_{t\searrow 0} \frac{\mathfrak{g}(\Omega_t) - \mathfrak{g}(\Omega)}{t}}{\lim_{t\searrow 0} \frac{|\Omega_t \triangle \Omega|}{t}} = \frac{d_S \mathfrak{g}(\Omega)(V)}{\int_{\partial\Omega} |V \cdot n| \, dS_x}. \tag{16}$$

$\square$

Remark 1 The condition $\int_\Omega \mathrm{div}(V) \, dx \neq 0$ is a sufficient condition for $\int_{\partial\Omega} |V \cdot n| \, dS_x > 0$, since

$$0 < \left| \int_\Omega \mathrm{div}(V) \, dx \right| = \left| \int_{\partial\Omega} V \cdot n \, dS_x \right| < \int_{\partial\Omega} |V \cdot n| \, dS_x. \tag{17}$$

2.3 The Continuous Topological-Shape Derivative

Here and in the following, we assume that the domain Ω is described by a level-set function $\phi : D \to \mathbb{R}$ via

$$\phi(\mathbf{x}) < 0 \Longleftrightarrow \mathbf{x} \in \Omega, \tag{18a}$$

$$\phi(\mathbf{x}) > 0 \Longleftrightarrow \mathbf{x} \in D \setminus \overline{\Omega} \tag{18b}$$

$$\phi(\mathbf{x}) = 0 \Longleftrightarrow \mathbf{x} \in \partial\Omega \cap D. \tag{18c}$$

For given ϕ, let $\Omega(\phi)$ denote the unique domain defined by (18a)–(18c). In this section, in contrast to the setting in Sect. 2.2, we perturb Ω indirectly by perturbing ϕ such that $\phi_\varepsilon = O_\varepsilon \phi$ for some operator $O_\varepsilon : C^0(D) \to C^0(D)$ depending on $\varepsilon \geq 0$ with the property $\Omega(O_0\phi) = \Omega(\phi)$. Later on, in the discrete setting, we will distinguish between two different types of perturbation operators O_ε corresponding to shape or topological perturbations of Ω.

Let, from now on, $\mathcal{J}(\phi) := \mathfrak{g}(\Omega(\phi))$ denote the reduced cost function as a function of the level set function ϕ. This way, a continuous topological-shape derivative can be defined as

$$d\mathcal{J}(\phi) = \lim_{\varepsilon \searrow 0} \frac{\mathcal{J}(\phi_\varepsilon) - \mathcal{J}(\phi)}{|\Omega(\phi_\varepsilon) \triangle \Omega(\phi)|}. \tag{19}$$

Note that this sensitivity depends on the choice of the perturbation operator O_ε, which can represent either a shape perturbation or a topological perturbation. We will mostly be concerned with its discrete counterpart, which will be introduced in Sect. 4. Note that, in the case of shape perturbations, due to the scaling $|\Omega(\phi_\varepsilon) \triangle \Omega(\phi)|$ instead of ε in the denominator the shape derivative is modified and does not correspond to (8) but rather to (13).

Relation to literature

The sensitivity of shape functions with respect to perturbations of a level set function (representing a shape) was investigated in [22] for the case without PDE constraints. There, the author considers smooth level set functions and rigorously computes the Gâteaux (semi-)derivative in the direction of a smooth perturbation of the level set function, both for the case of shape and topological perturbations. In the case of shape perturbations, it is shown that the Gâteaux derivative coincides with the shape derivative (8) with respect to a suitably chosen vector field. On the other hand, a resemblance between the notions of Gâteaux derivative and topological derivative is shown, yet the Gâteaux derivative may vanish or not exist in cases where the topological derivative is finite. Evidently, this discrepancy results from the fact that the denominator in the definition of the Gâteaux derivative is always of order one whereas it is of the order of the space dimension in the topological derivative.

While the analysis for shape and topological perturbations is carried out separately in [22], a more unified approach is followed in [13, 14]. In these publications, the idea is to consider sensitivities with respect to domain perturbations that are obtained by the dilation of lower-dimensional objects. Here, given a set $E \subset \mathbb{R}^d$ of dimension $k \leq d$,

the dilated set of radius r is given by $E_r = \{x \in \mathbb{R}^d : d_E(x) \leq r\}$ where $d_E(x)$ denotes a distance of a point x to a set E. For instance, when E is chosen as a single point, the dilated set is just a ball of radius r around that point and performing a sensitivity analysis with respect to the volume of the dilated object leads to the topological derivative. On the other hand, when E is chosen as the boundary of a domain, E_r can be defined using a signed distance function and corresponds to a uniform expansion of the domain. Then, a similar procedure leads to the shape derivative with respect to a uniform expansion in normal direction (i.e. $V = n$ in (7)–(8)). In [13], a sensitivity analysis for various choices of E is carried out with respect to the volume of the perturbation. We note, however, that arbitrary shape perturbations are not covered and would require an extension of the theory. Comparing [22] and [13], we observe that in the former paper only smooth perturbations of a level set function are admissible whereas, in the latter approach, domain perturbations by dilations can be interpreted as perturbations of level set functions by a (non-smooth) distance function.

Finally, we mention [9] where a domain is represented by a discretized level set function and a shape sensitivity analysis is carried out with respect to a perturbation of the level set values close to the boundary. This procedure can be interpreted as an application of the idea of dilation to discretized shape optimization problems. As the authors point out, this kind of shape sensitivity analysis is more natural compared to the standard approach based on domain transformations when employed in a level-set framework; an observation also made in [22, Sec. 3]. The authors show numerical results for the shape optimization of an acoustic horn, but do not consider topological perturbations in this work.

As it can be seen from (19), our approach is related to the dilation concept since we also consider the sensitivity with respect to the volume of the domain perturbation $\Omega_t \triangle \Omega$. In the following, we will investigate the topological-shape derivative in a discretized setting. Similarly to [9], we will consider shape sensitivity analysis with respect to level set values on mesh nodes close to the boundary. Moreover, we will also be able to deal with topological perturbations and treat shape and topological updates in a unified way by a discretized version of (19), called the numerical topological-shape derivative.

Remark 2 The concept of the topological-shape derivative which we introduce in this paper should not be confused with the topological-shape sensitivity method introduced in [27, 28]. While the new concept represents a unified sensitivity with respect to both shape and topological perturbations, the sensitivity method [27, 28] is a method to compute the pointwise topological derivative as a limit of a certain shape derivative and thus does not account for shape changes at physical material interfaces.

3 Numerical Setting

In this section we consider the discretization of (2). Let $\mathcal{T}$ be a given finite element mesh covering D with M nodes $\{\mathbf{x}_k\}_{k=1}^M$ and N triangular elements $\{\tau_l\}_{l=1}^N$. We introduce the index set $I_{\mathbf{x}_k}$ of all element indices of elements τ_l where $\mathbf{x}_k$ is a node of τ_l,

(a) $I_{\mathbf{x}_k} = \{1, \, 2, \, 3, \, 4\}$ (b) $J_{\tau_l} = \{1, \, 2, \, 3\}$ (c) $R_{\mathbf{x}_k} = \{k, \, 1, \, 2, \, 3, \, 4\}$
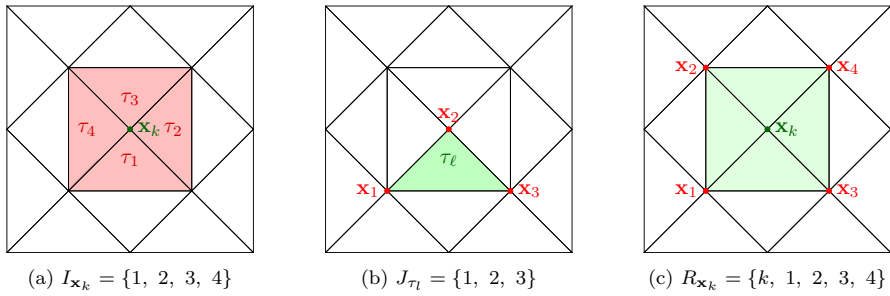
Fig. 1 Illustration of the sets $I_{\mathbf{x}_k}$, J_{τ_l}, and $R_{\mathbf{x}_k}$

$$I_{\mathbf{x}_k} := \{l \in \{1, \ldots, N\} : \mathbf{x}_k \in \bar{\tau}_l\} \quad \text{for } k = 1, \ldots, M. \tag{20}$$

Moreover,

$$J_{\tau_l} := \{k \in \{1, \ldots, M\} : \mathbf{x}_k \in \bar{\tau}_l\} \quad \text{for } l = 1, \ldots, N \tag{21}$$

is the index set of all node indices of nodes $\mathbf{x}_k$ in $\bar{\tau}_l$. Furthermore, we introduce the one-ring of a node $\mathbf{x}_k$,

$$R_{\mathbf{x}_k} := \{i \in \{1, \ldots, M\} | \exists l \in I_{\mathbf{x}_k} : \mathbf{x}_i \in \bar{\tau}_l\} \quad \text{for } k = 1, \ldots, M. \tag{22}$$

These sets are illustrated in Fig. 1.

Let $P_1 = \{a + bx_1 + cx_2 : a, b, c \in \mathbb{R}\}$ denote the space of affine linear polynomials in two space dimensions and $S_h^1(D)$ the space of piecewise affine linear and globally continuous functions on $\mathcal{T}$,

$$S_h^1(D) = \{v \in H^1(D) : v|_T \in P_1 \text{ for all } T \in \mathcal{T}\} = \text{span}\{\varphi_1, \ldots, \varphi_M\}$$

with the hat basis functions $\varphi_i \in S_h^1(D)$ which satisfy $\varphi_i(\mathbf{x}_j) = \delta_{ij}, i, j = 1, \ldots, M$. The discretization of problem (2) leads to the discretized optimization problem

$$\min_{\Omega} c_1 |\Omega| + c_2 (\mathbf{u} - \hat{\mathbf{u}})^\top \tilde{\mathbf{M}} (\mathbf{u} - \hat{\mathbf{u}}) \tag{23a}$$

$$\text{subject to } \mathbf{A}\mathbf{u} = \mathbf{f}, \tag{23b}$$

with the solution vector $\mathbf{u} \in \mathbb{R}^M$ and $\mathbf{A} = \mathbf{M} + \mathbf{K}$. Here, the mass matrices $\mathbf{M}, \tilde{\mathbf{M}} \in \mathbb{R}^{M \times M}$, the stiffness matrix $\mathbf{K} \in \mathbb{R}^{M \times M}$, and the right-hand-side vector $\mathbf{f} \in \mathbb{R}^M$ depend on the shape Ω and are given by

$$\mathbf{M}[i, j] = \int_D \alpha_\Omega \varphi_j \varphi_i \, \mathrm{d}\mathbf{x}, \qquad \mathbf{K}[i, j] = \int_D \lambda_\Omega \nabla \varphi_j \cdot \nabla \varphi_i \, \mathrm{d}\mathbf{x},$$

$$\mathbf{f}[i] = \int_D f_\Omega \varphi_i \, \mathrm{d}\mathbf{x}, \qquad \tilde{\mathbf{M}}[i, j] = \int_D \tilde{\alpha}_\Omega \varphi_j \varphi_i \, \mathrm{d}\mathbf{x}. \tag{24}$$

On the reference element $\tau_R = \{\xi \in \mathbb{R}^2 : 0 \le \xi_1 \le 1, 0 \le \xi_2 \le 1 - \xi_1\}$ we have the local form functions

$$\psi_1(\xi_1, \xi_2) = 1 - \xi_1 - \xi_2, \qquad \psi_2(\xi_1, \xi_2) = \xi_1, \qquad \psi_3(\xi_1, \xi_2) = \xi_2.$$

For an element $\tau_l \in \mathcal{T}$, we denote the global vertex indices of its three vertices by l_1, l_2, l_3 and assume them to be numbered in counter-clockwise orientation. Then, the respective local finite element matrices and the local right-hand-side vector for element τ_l are given by

$$\mathbf{m}_l[i, j] = |\det J_l| \int_{\xi_1=0}^{1} \int_{\xi_2=0}^{1-\xi_1} (\alpha_\Omega \circ \Phi_l) \psi_j \psi_i \, d\xi_2 \, d\xi_1,$$

$$\mathbf{k}_l[i, j] = |\det J_l| \left(J_l^{-1} \nabla_\xi \psi_j \right)^\top \left(J_l^{-1} \nabla_\xi \psi_i \right) \int_{\xi_1=0}^{1} \int_{\xi_2=0}^{1-\xi_1} (\lambda_\Omega \circ \Phi_l) \, d\xi_2 \, d\xi_1,$$

$$\mathbf{f}_l[i] = |\det J_l| \int_{\xi_1=0}^{1} \int_{\xi_2=0}^{1-\xi_1} (f_\Omega \circ \Phi_l) \psi_i \, d\xi_2 \, d\xi_1,$$

for $i, j \in \{1, 2, 3\}$, where $\nabla_\xi \psi_j = [\partial_{\xi_1} \psi_j, \partial_{\xi_2} \psi_j]^\top$ and the mapping Φ_l between τ_R and τ_l and its Jacobian J_l are given by

$$\Phi_l(\xi_1, \xi_2) = \mathbf{x}_{l_1} + J_l \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}, \qquad J_l = \begin{bmatrix} \mathbf{x}_{l_2} - \mathbf{x}_{l_1} & \mathbf{x}_{l_3} - \mathbf{x}_{l_1} \end{bmatrix} \in \mathbb{R}^{2 \times 2}.$$

4 Numerical Topological-Shape Derivative

Given the discretization introduced in Sect. 3, in contrast to the continuous topological-shape derivative, the numerical topological-shape derivative is only defined at the nodes of the finite element mesh $\mathcal{T}$. For a given piecewise linear level set function $\phi \in S_h^1(D)$ let $\Omega(\phi)$ be defined by (18a)–(18c) and $\mathcal{J}(\phi) = g(\Omega(\phi), u(\phi))$, where $u(\phi) \in V_h = \{v \in S_h^1(D) : v = g_D \text{ on } \Gamma_D\}$ is the finite element function corresponding to the solution of (23b). Note that, in this section, $\Omega(\phi)$ is polygonal since $\phi \in S_h^1(D)$. The topological-shape derivative at node $\mathbf{x}_k \in \mathcal{T}$ is defined by

$$d\mathcal{J}(\phi)(\mathbf{x}_k) = \begin{cases} \lim_{\varepsilon \searrow 0} \frac{\mathcal{J}(T_{k,\varepsilon}^{-\to+}\phi) - \mathcal{J}(\phi)}{|\Omega(T_{k,\varepsilon}^{-\to+}\phi) \triangle \Omega(\phi)|} & \text{for } \mathbf{x}_k \in \mathfrak{T}^-(\phi), \\ \lim_{\varepsilon \searrow 0} \frac{\mathcal{J}(T_{k,\varepsilon}^{+\to-}\phi) - \mathcal{J}(\phi)}{|\Omega(T_{k,\varepsilon}^{+\to-}\phi) \triangle \Omega(\phi)|} & \text{for } \mathbf{x}_k \in \mathfrak{T}^+(\phi), \\ \lim_{\varepsilon \searrow 0} \frac{\mathcal{J}(S_{k,\varepsilon}\phi) - \mathcal{J}(\phi)}{|\Omega(S_{k,\varepsilon}\phi) \triangle \Omega(\phi)|} & \text{for } \mathbf{x}_k \in \mathfrak{S}(\phi). \end{cases} \qquad (25)$$

Here, given $\phi \in S_h^1(D)$, the respective sets are defined by

$$\mathfrak{T}^-(\phi) = \{\mathbf{x}_k \in \mathcal{T} \mid \forall i \in R_{\mathbf{x}_k} : \phi(\mathbf{x}_i) \le 0\}, \qquad (26a)$$

(a) Shape derivative: Initial domain in green, perturbed domain in orange. Interface is moved to the left.

(b) Shape derivative: No effect of the perturbation

(c) Topological derivative: Shape derivative at a topologically perturbed level-set function (dashed blue)
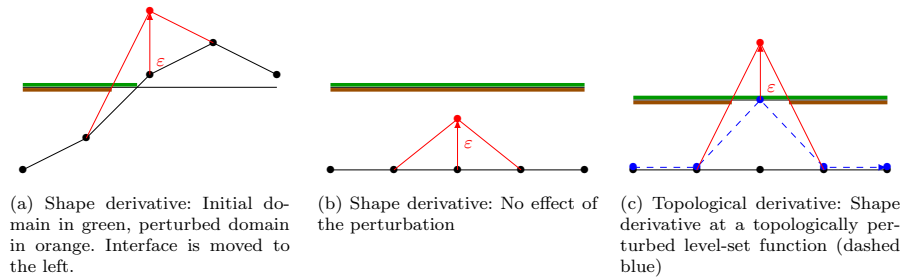
Fig. 2 Illustration of the shape derivative and the topological derivative

$$\mathfrak{T}^{+}(\phi) = \{\mathbf{x}_k \in \mathcal{T} \mid \forall i \in R_{\mathbf{x}_k} : \phi(\mathbf{x}_i) \geq 0\}, \tag{26b}$$

$$\mathfrak{S}(\phi) = \mathcal{T} \setminus (\mathfrak{T}^{-}(\phi) \cup \mathfrak{T}^{+}(\phi)). \tag{26c}$$

Whenever, the level set function ϕ is clear from the context, we will drop the argument and write $\mathfrak{T}^{-}, \mathfrak{T}^{+}, \mathfrak{S}$ for brevity. Furthermore, $T_{k,\varepsilon}^{-\to+} : S_h^1(D) \to S_h^1(D)$ is the positive discrete topological perturbation operator defined by its action

$$T_{k,\varepsilon}^{-\to+}\phi(\mathbf{x}) = \phi(\mathbf{x}) - (\phi(\mathbf{x}_k) - \varepsilon)\varphi_k(\mathbf{x}), \tag{27}$$

see Fig. 2(c), whereas the negative discrete topological perturbation operator $T_{k,\varepsilon}^{+\to-} : S_h^1(D) \to S_h^1(D)$ is defined by

$$T_{k,\varepsilon}^{+\to-}\phi(\mathbf{x}) = \phi(\mathbf{x}) - (\phi(\mathbf{x}_k) + \varepsilon)\varphi_k(\mathbf{x}). \tag{28}$$

Finally, the discrete shape perturbation operator $S_{k,\varepsilon} : S_h^1(D) \to S_h^1(D)$ is defined by

$$S_{k,\varepsilon}\phi(\mathbf{x}) := \phi(\mathbf{x}) + \varepsilon\varphi_k(\mathbf{x}). \tag{29}$$

Remark 3 Note that the discrete perturbation operators defined above only change the nodal value of the finite element function $\phi \in S_h^1(D)$ at one node $\mathbf{x}_k$, *e.g.* for $S_{k,\varepsilon}$ it holds $S_{k,\varepsilon}\phi(\mathbf{x}_j) = \phi(\mathbf{x}_j)$ for all $j \in \mathcal{T} \setminus \{k\}$ and $S_{k,\varepsilon}\phi(\mathbf{x}_k) = \phi(\mathbf{x}_k) + \varepsilon$.

4.1 Computation of the Numerical Topological-Shape Derivative for the Area Cost Functional

Before we compute the numerical topological-shape derivative (25) for the full model problem (2), we consider the case of the pure volume cost function and neglect the PDE constraint, i.e., we set $c_1 = 1, c_2 = 0$ in (1).

For that purpose, we investigate

$$\delta_{k,\varepsilon}a = |\Omega(O_{k,\varepsilon}\phi)| - |\Omega(\phi)|, \tag{30}$$

for $O_{k,\varepsilon} \in \{T_{k,\varepsilon}^{-\rightarrow+}, T_{k,\varepsilon}^{+\rightarrow-}, S_{k,\varepsilon}\}$. We note that for the computation of (30) only those "cut elements" are relevant which have a node $\mathbf{x}_k$, *i.e.*

$$\delta_{k,\varepsilon} a = \sum_{l \in C_k} \delta_{k,\varepsilon} a_l \quad \text{with } \delta_{k,\varepsilon} a_l := \int_{\tau_l} H(-O_{k,\varepsilon}\phi) - H(-\phi) \, d\mathbf{x}, \qquad (31)$$

where $H(x)$ denotes the Heaviside step function and

$$C_k = \{l \in I_{\mathbf{x}_k} : \tau_l \cap \partial\Omega(O_{k,\varepsilon}\phi) \neq \emptyset\} \qquad (32)$$

is the set of all indices of elements adjacent to $\mathbf{x}_k$ which are intersected by the perturbed interface. Note that, for $\varepsilon > 0$ small enough, C_k does not depend on the concrete value of ε. For an element τ_l with $l \in I_{\mathbf{x}_k}$ we denote the three vertices in counter-clockwise orientation by $\mathbf{x}_{l_1}, \mathbf{x}_{l_2}, \mathbf{x}_{l_3}$ and assume that $\mathbf{x}_k = \mathbf{x}_{l_1}$. Moreover we denote $\phi_{l_j} := \phi(\mathbf{x}_{l_j})$ and $\tilde{\phi}_{l_j} := O_{k,\varepsilon}\phi(\mathbf{x}_{l_j})$ for $j = 1, 2, 3$ and small enough ε. In the following, we will be interested in

$$d_k a := \sum_{l \in C_k} d_k a_l \quad \text{with } d_k a_l := \lim_{\varepsilon \searrow 0} \frac{\delta_{k,\varepsilon} a_l}{\varepsilon^o}. \qquad (33)$$

with $\delta_{k,\varepsilon} a_l$ defined in (30). Here, $o = 1$ in the case of a shape perturbation and $o = 2$ in the case of a topological perturbation. We consider six different sets (see Fig. 3 for an illustration)

$$\begin{aligned}
I_{\mathbf{x}_k}^{A+} &= \{l \in I_{\mathbf{x}_k} : \tilde{\phi}_{l_1} > 0, \tilde{\phi}_{l_2} < 0, \tilde{\phi}_{l_3} < 0\}, & I_{\mathbf{x}_k}^{A-} &= \{l \in I_{\mathbf{x}_k} : \tilde{\phi}_{l_1} < 0, \tilde{\phi}_{l_2} > 0, \tilde{\phi}_{l_3} > 0\}, \\
I_{\mathbf{x}_k}^{B+} &= \{l \in I_{\mathbf{x}_k} : \tilde{\phi}_{l_1} < 0, \tilde{\phi}_{l_2} > 0, \tilde{\phi}_{l_3} < 0\}, & I_{\mathbf{x}_k}^{B-} &= \{l \in I_{\mathbf{x}_k} : \tilde{\phi}_{l_1} > 0, \tilde{\phi}_{l_2} < 0, \tilde{\phi}_{l_3} > 0\}, \\
I_{\mathbf{x}_k}^{C+} &= \{l \in I_{\mathbf{x}_k} : \tilde{\phi}_{l_1} < 0, \tilde{\phi}_{l_2} < 0, \tilde{\phi}_{l_3} > 0\}, & I_{\mathbf{x}_k}^{C-} &= \{l \in I_{\mathbf{x}_k} : \tilde{\phi}_{l_1} > 0, \tilde{\phi}_{l_2} > 0, \tilde{\phi}_{l_3} < 0\},
\end{aligned} \qquad (34)$$

such that

$$C_k = I_{\mathbf{x}_k}^{A+} \cup I_{\mathbf{x}_k}^{A-} \cup I_{\mathbf{x}_k}^{B+} \cup I_{\mathbf{x}_k}^{B-} \cup I_{\mathbf{x}_k}^{C+} \cup I_{\mathbf{x}_k}^{C-}$$

with a direct sum on the right hand side. We can thus split the sum in (31) into six parts,

$$\delta_{k,\varepsilon} a = \mathcal{I}_{I_{\mathbf{x}_k}}^{A+} + \mathcal{I}_{I_{\mathbf{x}_k}}^{A-} + \mathcal{I}_{I_{\mathbf{x}_k}}^{B+} + \mathcal{I}_{I_{\mathbf{x}_k}}^{B-} + \mathcal{I}_{I_{\mathbf{x}_k}}^{C+} + \mathcal{I}_{I_{\mathbf{x}_k}}^{C-}.$$

Configuration A For $l \in I_{\mathbf{x}_k}^{A+}$ we have

$$\begin{aligned}
\int_{\tau_l} H(-O_{k,\varepsilon}\phi(\mathbf{x})) \, d\mathbf{x} &= \frac{|\det J_l|}{2} \left(1 - \int_{\xi_1=0}^{\ell_1} \int_{\xi_2=0}^{\ell_2\left(1-\frac{\xi_1}{\ell_1}\right)} d\xi_2 d\xi_1 \right) \\
&= \frac{|\det J_l|}{2} (1 - \ell_1 \ell_2), \qquad (35)
\end{aligned}$$

(a) Triangle is in $I_{\mathbf{x}_k}^{A+}$ (b) Triangle is in $I_{\mathbf{x}_k}^{B+}$ (c) Triangle is in $I_{\mathbf{x}_k}^{C+}$

(d) Triangle is in $I_{\mathbf{x}_k}^{A-}$ (e) Triangle is in $I_{\mathbf{x}_k}^{B-}$ (f) Triangle is in $I_{\mathbf{x}_k}^{C-}$
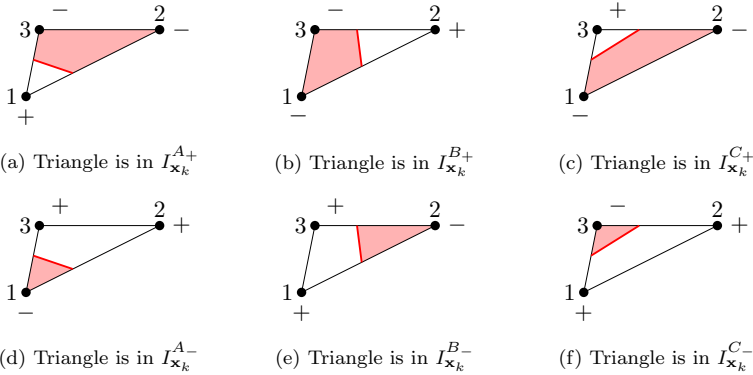
Fig. 3 Illustration of the sets $I_{\mathbf{x}_k}^{A+}$, $I_{\mathbf{x}_k}^{A-}$, $I_{\mathbf{x}_k}^{B+}$, $I_{\mathbf{x}_k}^{B-}$, $I_{\mathbf{x}_k}^{C+}$, $I_{\mathbf{x}_k}^{C-}$. The nodal values of the level-set functions are indicated by $-$, $+$. The interface is drawn in red

(a) Triangles are in $I_{\mathbf{x}_k}^{A+}$ (b) Triangles are in $I_{\mathbf{x}_k}^{A-}$
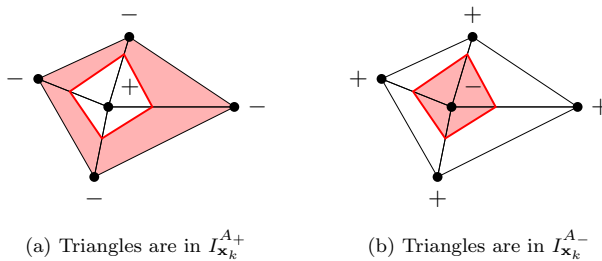
Fig. 4 Illustration of $I_{\mathbf{x}_k}^{A+}$ and $I_{\mathbf{x}_k}^{A-}$ in the case of topological perturbations

with

$$\ell_1 = \frac{\phi_{l_1} + \varepsilon}{\phi_{l_1} + \varepsilon - \phi_{l_2}}, \qquad \ell_2 = \frac{\phi_{l_1} + \varepsilon}{\phi_{l_1} + \varepsilon - \phi_{l_3}}.$$

Therefore,

$$\mathcal{I}_{I_{\mathbf{x}_k}}^{A+} = \sum_{l \in I_{\mathbf{x}_k}^{A+}} \frac{|\det J_l|}{2} \left(-\frac{(\phi_{l_1} + \varepsilon)^2}{(\phi_{l_1} + \varepsilon - \phi_{l_2})(\phi_{l_1} + \varepsilon - \phi_{l_3})} + \frac{\phi_{l_1}^2}{(\phi_{l_1} - \phi_{l_2})(\phi_{l_1} - \phi_{l_3})} \right)$$

$$= \sum_{l \in I_{\mathbf{x}_k}^{A+}} \frac{|\det J_l|}{2} \left(\frac{\varepsilon \phi_{l_1} \left[\phi_{l_1}(\phi_{l_2} + \phi_{l_3}) - 2\phi_{l_2}\phi_{l_3} \right] + \varepsilon^2 \left[\phi_{l_1}(\phi_{l_2} + \phi_{l_3}) - \phi_{l_2}\phi_{l_3} \right]}{(\phi_{l_1} - \phi_{l_2})^2(\phi_{l_1} - \phi_{l_3})^2 + O(\varepsilon)} \right).$$

$$(36)$$

For $\mathbf{x}_k \in \mathfrak{T}^-$ and $O_{k,\varepsilon} = T_{k,\varepsilon}^{-\to+}$, it holds $I_{\mathbf{x}_k} = I_{\mathbf{x}_k}^{A+}$ (see Fig. 4a) and we have to consider (36) with $\phi_{l_1} = 0$. Thus, we obtain

$$\mathcal{I}_{I_{\mathbf{x}_k}}^{A+} = - \sum_{l \in I_{\mathbf{x}_k}^{A+}} \frac{\varepsilon^2 \, |\det J_l|}{2} \frac{\phi_{l_2} \phi_{l_3}}{\phi_{l_2}^2 \phi_{l_3}^2 + O(\varepsilon)}, \tag{37}$$

and conclude for this case

$$d_k a = \lim_{\varepsilon \searrow 0} \frac{\delta_{k,\varepsilon} a}{\varepsilon^2} = - \sum_{l \in I_{\mathbf{x}_k}^{A+}} \frac{|\det J_l|}{2 \phi_{l_2} \phi_{l_3}}. \tag{38}$$

Moreover, for $\mathbf{x}_k \in \mathfrak{T}^+$ and $O_{k,\varepsilon} = T_{k,\varepsilon}^{+ \to -}$, it holds $I_{\mathbf{x}_k} = I_{\mathbf{x}_k}^{A-}$ (see Fig. 4b) and we have for $l \in I_{\mathbf{x}_k}^{A-}$,

$$\int_{\tau_l} H(-O_{k,\varepsilon} \phi(\mathbf{x})) \, d\mathbf{x} = \frac{|\det J_l|}{2} \int_{\xi_1=0}^{\ell_1} \int_{\xi_2=0}^{\ell_2 \left(1 - \frac{\xi_1}{\ell_1}\right)} d\xi_2 d\xi_1 = \frac{|\det J_l|}{2} \ell_1 \ell_2,$$

which leads to

$$\mathcal{I}_{I_{\mathbf{x}_k}}^{A-} = - \sum_{l \in I_{\mathbf{x}_k}^{A-}} \frac{|\det J_l|}{2} \left(\frac{\varepsilon \phi_{l_1} \left[\phi_{l_1}(\phi_{l_2} + \phi_{l_3}) - 2\phi_{l_2}\phi_{l_3} \right] + \varepsilon^2 \left[\phi_{l_1}(\phi_{l_2} + \phi_{l_3}) - \phi_{l_2}\phi_{l_3} \right]}{(\phi_{l_1} - \phi_{l_2})^2 (\phi_{l_1} - \phi_{l_3})^2 + O(\varepsilon)} \right). \tag{39}$$

Therefore, we obtain for this case

$$d_k a = \lim_{\varepsilon \searrow 0} \frac{\delta_{k,\varepsilon} a}{\varepsilon^2} = \sum_{l \in I_{\mathbf{x}_k}^{A-}} \frac{|\det J_l|}{2 \phi_{l_2} \phi_{l_3}}.$$

Finally, for $\mathbf{x}_k \in \mathfrak{S}$ and $O_{k,\varepsilon} = S_{k,\varepsilon}$ we deduce from (36)

$$\lim_{\varepsilon \searrow 0} \frac{\mathcal{I}_{I_{\mathbf{x}_k}}^{A+}}{\varepsilon} = \sum_{l \in I_{\mathbf{x}_k}^{A+}} \frac{|\det J_l| \, \phi_{l_1} \left[\phi_{l_1}(\phi_{l_2} + \phi_{l_3}) - 2\phi_{l_2}\phi_{l_3} \right]}{2(\phi_{l_1} - \phi_{l_2})^2 (\phi_{l_1} - \phi_{l_3})^2}, \tag{40}$$

and from (39)

$$\lim_{\varepsilon \searrow 0} \frac{\mathcal{I}_{I_{\mathbf{x}_k}}^{A-}}{\varepsilon} = - \sum_{l \in I_{\mathbf{x}_k}^{A-}} \frac{|\det J_l| \, \phi_{l_1} \left[\phi_{l_1}(\phi_{l_2} + \phi_{l_3}) - 2\phi_{l_2}\phi_{l_3} \right]}{2(\phi_{l_1} - \phi_{l_2})^2 (\phi_{l_1} - \phi_{l_3})^2}. \tag{41}$$

Configuration B We note that configuration B can only occur for the case $\mathbf{x}_k \in \mathfrak{S}$ and $O_{k,\varepsilon} = S_{k,\varepsilon}$. For $l \in I_{\mathbf{x}_k}^{B+}$ it holds

$$\int_{\tau_l} H(-S_{k,\varepsilon} \phi(\mathbf{x})) \, d\mathbf{x} = \frac{|\det J_l|}{2} \left(1 - \frac{\phi_{l_2}^2}{(\phi_{l_2} - \phi_{l_3})(\phi_{l_2} - \phi_{l_1} - \varepsilon)} \right), \tag{42}$$

and

$$
\begin{aligned}
\mathcal{I}_{I_{\mathbf{x}_k}}^{B+} &= \sum_{l \in I_{\mathbf{x}_k}^{B+}} \frac{|\det J_l| \, \phi_{l_2}^2}{2} \frac{(\phi_{l_2} - \phi_{l_3})(\phi_{l_2} - \phi_{l_1} - \varepsilon) - (\phi_{l_2} - \phi_{l_3})(\phi_{l_2} - \phi_{l_1})}{(\phi_{l_2} - \phi_{l_3})(\phi_{l_2} - \phi_{l_1})(\phi_{l_2} - \phi_{l_3})(\phi_{l_2} - \phi_{l_1} - \varepsilon)} \\
&= \sum_{l \in I_{\mathbf{x}_k}^{B+}} \frac{|\det J_l| \, \phi_{l_2}^2}{2} \frac{-\varepsilon(\phi_{l_2} - \phi_{l_3})}{(\phi_{l_2} - \phi_{l_3})^2 (\phi_{l_2} - \phi_{l_1})(\phi_{l_2} - \phi_{l_1} - \varepsilon)}.
\end{aligned}
$$

Thus,

$$
\lim_{\varepsilon \searrow 0} \frac{\mathcal{I}_{I_{\mathbf{x}_k}}^{B+}}{\varepsilon} = - \sum_{l \in I_{\mathbf{x}_k}^{B+}} \frac{|\det J_l|}{2} \frac{\phi_{l_2}^2}{(\phi_{l_2} - \phi_{l_3})(\phi_{l_2} - \phi_{l_1})^2}. \tag{43}
$$

For the case $l \in I_{\mathbf{x}_k}^{B-}$ we have

$$
\int_{\tau_l} H(-S_{k,\varepsilon} \phi(\mathbf{x})) \, \mathrm{d}\mathbf{x} = \frac{|\det J_l|}{2} \frac{\phi_{l_2}^2}{(\phi_{l_2} - \phi_{l_3})(\phi_{l_2} - \phi_{l_1} - \varepsilon)}, \tag{44}
$$

and obtain

$$
\lim_{\varepsilon \searrow 0} \frac{\mathcal{I}_{I_{\mathbf{x}_k}}^{B-}}{\varepsilon} = \sum_{l \in I_{\mathbf{x}_k}^{B-}} \frac{|\det J_l|}{2} \frac{\phi_{l_2}^2}{(\phi_{l_2} - \phi_{l_3})(\phi_{l_2} - \phi_{l_1})^2}. \tag{45}
$$

Configuration C Analogously as for configuration B, we note that configuration C can only occur for the case $\mathbf{x}_k \in \mathfrak{S}$ and $O_{k,\varepsilon} = S_{k,\varepsilon}$. Furthermore, the formulas for configuration C can be obtained by exchanging l_3 and l_2 in the formulas (43) and (45).

Summarizing, we have shown the following result.

Theorem 1 *For $\mathbf{x}_k \in \mathfrak{T}^-$ we have*

$$
d_k a = \sum_{l \in I_{\mathbf{x}_k}^{A+}} d_k a_l = \sum_{l \in I_{\mathbf{x}_k}^{A+}} (-1) \frac{|\det J_l|}{2 \phi_{l_2} \phi_{l_3}}. \tag{46}
$$

For $\mathbf{x}_k \in \mathfrak{T}^+$ we have

$$
d_k a = \sum_{l \in I_{\mathbf{x}_k}^{A-}} d_k a_l = \sum_{l \in I_{\mathbf{x}_k}^{A-}} \frac{|\det J_l|}{2 \phi_{l_2} \phi_{l_3}}. \tag{47}
$$

For $\mathbf{x}_k \in \mathfrak{S}$ we have

$$
\begin{aligned}
d_k a = &\sum_{l \in I_{\mathbf{x}_k}^{A+}} \mathcal{I}_0(\phi_{l_1}, \phi_{l_2}, \phi_{l_3}) - \sum_{l \in I_{\mathbf{x}_k}^{A-}} \mathcal{I}_0(\phi_{l_1}, \phi_{l_2}, \phi_{l_3}) - \sum_{l \in I_{\mathbf{x}_k}^{B+}} \mathcal{I}_1(\phi_{l_1}, \phi_{l_2}, \phi_{l_3}) \\
&+ \sum_{l \in I_{\mathbf{x}_k}^{B-}} \mathcal{I}_1(\phi_{l_1}, \phi_{l_2}, \phi_{l_3}) - \sum_{l \in I_{\mathbf{x}_k}^{C+}} \mathcal{I}_1(\phi_{l_1}, \phi_{l_3}, \phi_{l_2}) + \sum_{l \in I_{\mathbf{x}_k}^{C-}} \mathcal{I}_1(\phi_{l_1}, \phi_{l_3}, \phi_{l_2}),
\end{aligned}
\tag{48}
$$

with

$$
\mathcal{I}_0(\phi_{l_1}, \phi_{l_2}, \phi_{l_3}) = \frac{|\det J_l| \; \phi_{l_1} \left[\phi_{l_1}(\phi_{l_2} + \phi_{l_3}) - 2\phi_{l_2}\phi_{l_3} \right]}{2(\phi_{l_1} - \phi_{l_2})^2(\phi_{l_1} - \phi_{l_3})^2},
$$

$$
\mathcal{I}_1(\phi_{l_1}, \phi_{l_2}, \phi_{l_3}) = \frac{|\det J_l| \; \phi_{l_2}^2}{2(\phi_{l_2} - \phi_{l_3})(\phi_{l_2} - \phi_{l_1})^2}.
$$

Remark 4 The corresponding computations for the denominators in (25), *i.e.* $|\Omega(O_{k,\varepsilon}\phi) \triangle \Omega(\phi)|$, are closely related to the computations presented in this section for (30). Denoting

$$
\delta_{k,\varepsilon}\tilde{a}_l := \int_{\tau_l} |H(-O_{k,\varepsilon}\phi) - H(-\phi)| \, d\mathbf{x}, \quad d_k\tilde{a}_l := \lim_{\varepsilon \searrow 0} \frac{\delta_{k,\varepsilon}\tilde{a}_l}{\varepsilon^o},
$$

we get that

$$
\lim_{\varepsilon \searrow 0} \frac{|\Omega(O_{k,\varepsilon}\phi) \triangle \Omega(\phi)|}{\varepsilon^o} = \sum_{l \in C_k} d_k\tilde{a}_l =: d_k\tilde{a}
\tag{49}
$$

where $d_k\tilde{a}_l = |d_k a_l|$ with the formulas for $d_k a_l$ given in (46)–(48). It is obvious that, for $\mathbf{x}_k \in \mathcal{T}^-$, $d_k a_l < 0$ and for $\mathbf{x}_k \in \mathcal{T}^+$, $d_k a_l > 0$. Moreover, note that, for $\mathbf{x}_k \in \mathcal{S}$, it holds $d_k a_l < 0$ for all $l \in C_k$. This yields that

$$
\frac{d_k a_l}{d_k \tilde{a}_l} = \begin{cases} -1 & \text{if } k \in \mathcal{S} \cup \mathcal{T}^-, \\ 1 & \text{if } k \in \mathcal{T}^+. \end{cases}
\tag{50}
$$

Corollary 2 *From Theorem 1 and Remark 4, it follows that the numerical topological-shape derivative of the volume cost function $Vol(\phi) := |\Omega(\phi)|$ is given by*

$$
dVol(\phi)(\mathbf{x}_k) = \begin{cases} -1 & \text{if } k \in \mathcal{S} \cup \mathcal{T}^-, \\ 1 & \text{if } k \in \mathcal{T}^+. \end{cases}
$$

4.2 Computation of the Numerical Topological-Shape Derivative via Lagrangian Framework

Next, we consider the computation of the numerical topological-shape derivative of an optimization problem that is constrained by a discretized PDE. For that purpose, we set $c_1 = 0$ in (1) and $J(\phi, \mathbf{u}) := g(\Omega(\phi), \mathbf{u})$. The discretized problem reads

$$\min_{\phi} J(\phi, \mathbf{u}) \tag{51}$$

$$\text{s.t. } \mathbf{A}_\phi \mathbf{u} = \mathbf{f}_\phi \tag{52}$$

and we are interested in the sensitivity of J when the level set function ϕ representing the geometry is replaced by a perturbed level set function $\phi_\varepsilon = O_{k,\varepsilon}\phi$. The perturbed Lagrangian for (51)–(52) with respect to a perturbation of ϕ reads

$$L(\varepsilon, \mathbf{u}, \mathbf{v}) = J(\phi_\varepsilon, \mathbf{u}) + \mathbf{A}_\varepsilon \mathbf{u} \cdot \mathbf{v} - \mathbf{f}_\varepsilon \cdot \mathbf{v} \tag{53}$$

where we use the abbreviated notation $\mathbf{A}_\varepsilon := \mathbf{A}_{\phi_\varepsilon}$, and $\mathbf{f}_\varepsilon := \mathbf{f}_{\phi_\varepsilon}$. Moreover, for $\varepsilon \geq 0$, we define the perturbed state $\mathbf{u}_\varepsilon$ as the solution to

$$0 = \partial_{\mathbf{v}} L(\varepsilon, \mathbf{u}_\varepsilon, \mathbf{v}),$$

i.e. $\mathbf{u}_\varepsilon$ is the solution to

$$\mathbf{A}_\varepsilon \mathbf{u}_\varepsilon = \mathbf{f}_\varepsilon \tag{54}$$

and the (unperturbed) adjoint state $\mathbf{p}$ as the solution to

$$0 = \partial_{\mathbf{u}} L(0, \mathbf{u}, \mathbf{p}), \tag{55}$$

for the state $\mathbf{u}$ given, i.e. $\mathbf{p}$ solves

$$\mathbf{A}^\top \mathbf{p} = -\partial_{\mathbf{u}} J(\phi, \mathbf{u}).$$

Note that we use the notation $\mathbf{u}$ for $\mathbf{u}_{\varepsilon=0}$. The numerical topological-shape derivative at the node $\mathbf{x}_k$ can be computed as the limit

$$d\mathcal{J}(\phi)(\mathbf{x}_k) = \lim_{\varepsilon \searrow 0} \frac{1}{|\Omega(\phi_\varepsilon) \triangle \Omega(\phi)|} (J(\phi_\varepsilon, \mathbf{u}_\varepsilon) - J(\phi, \mathbf{u})).$$

With the help of the Lagrangian (53), we can rewrite the right hand side as

$$J(\phi_\varepsilon, \mathbf{u}_\varepsilon) - J(\phi, \mathbf{u}) = L(\varepsilon, \mathbf{u}_\varepsilon, \mathbf{p}) - L(0, \mathbf{u}, \mathbf{p})$$

where we used that $\mathbf{u}_\varepsilon$ solves (54) for $\varepsilon \geq 0$. Following the approach used in [18], we use the fundamental theorem of calculus as well as (55) to rewrite this as

$$J(\phi_\varepsilon, \mathbf{u}_\varepsilon) - J(\phi, \mathbf{u}) = \int_0^1 [\partial_{\mathbf{u}} L(\varepsilon, \mathbf{u} + s(\mathbf{u}_\varepsilon - \mathbf{u}), \mathbf{p}) - \partial_{\mathbf{u}} L(\varepsilon, \mathbf{u}, \mathbf{p})](\mathbf{u}_\varepsilon - \mathbf{u}) \mathrm{d}s \tag{56}$$

$$+ [\partial_{\mathbf{u}} L(\varepsilon, \mathbf{u}, \mathbf{p}) - \partial_{\mathbf{u}} L(0, \mathbf{u}, \mathbf{p})](\mathbf{u}_\varepsilon - \mathbf{u}) \tag{57}$$

$$+ L(\varepsilon, \mathbf{u}, \mathbf{p}) - L(0, \mathbf{u}, \mathbf{p}). \tag{58}$$

Thus the numerical topological-shape derivative can be obtained as the sum of three limits,

$$d\mathcal{J}(\phi)(\mathbf{x}_k) = R_1(\mathbf{u}, \mathbf{p}) + R_2(\mathbf{u}, \mathbf{p}) + R_0(\mathbf{u}, \mathbf{p})$$

where

$$R_1(\mathbf{u}, \mathbf{p}) := \lim_{\varepsilon \searrow 0} \frac{1}{|\Omega(\phi_\varepsilon) \triangle \Omega(\phi)|} \int_0^1 [\partial_{\mathbf{u}} L(\varepsilon, \mathbf{u} + s(\mathbf{u}_\varepsilon - \mathbf{u}), \mathbf{p}) - \partial_{\mathbf{u}} L(\varepsilon, \mathbf{u}, \mathbf{p})](\mathbf{u}_\varepsilon - \mathbf{u}) \mathrm{d}s,$$

$$R_2(\mathbf{u}, \mathbf{p}) := \lim_{\varepsilon \searrow 0} \frac{1}{|\Omega(\phi_\varepsilon) \triangle \Omega(\phi)|} [\partial_{\mathbf{u}} L(\varepsilon, \mathbf{u}, \mathbf{p}) - \partial_{\mathbf{u}} L(0, \mathbf{u}, \mathbf{p})](\mathbf{u}_\varepsilon - \mathbf{u}),$$

$$R_0(\mathbf{u}, \mathbf{p}) := \lim_{\varepsilon \searrow 0} \frac{1}{|\Omega(\phi_\varepsilon) \triangle \Omega(\phi)|} [L(\varepsilon, \mathbf{u}, \mathbf{p}) - L(0, \mathbf{u}, \mathbf{p})].$$

For $J(\phi_\varepsilon, \mathbf{u}) = c_2(\mathbf{u} - \hat{\mathbf{u}})\tilde{\mathbf{M}}_\varepsilon(\mathbf{u} - \hat{\mathbf{u}})$, where $\tilde{\mathbf{M}}_\varepsilon := \tilde{\mathbf{M}}_{\phi_\varepsilon}$ represents the matrix $\tilde{\mathbf{M}}$ defined in (24) with $\Omega = \Omega(\phi_\varepsilon)$, we get

$$R_1(\mathbf{u}, \mathbf{p}) = c_2 \lim_{\varepsilon \searrow 0} \frac{1}{|\Omega(\phi_\varepsilon) \triangle \Omega(\phi)|} (\mathbf{u}_\varepsilon - \mathbf{u})^\top \tilde{\mathbf{M}}_\varepsilon (\mathbf{u}_\varepsilon - \mathbf{u}).$$

Moreover,

$$R_2(\mathbf{u}, \mathbf{p}) = \lim_{\varepsilon \searrow 0} \frac{1}{|\Omega(\phi_\varepsilon) \triangle \Omega(\phi)|} \left[2c_2(\tilde{\mathbf{M}}_\varepsilon - \tilde{\mathbf{M}})(\mathbf{u} - \hat{\mathbf{u}}) \cdot (\mathbf{u}_\varepsilon - \mathbf{u}) + (\mathbf{A}_\varepsilon - \mathbf{A})(\mathbf{u}_\varepsilon - \mathbf{u}) \cdot \mathbf{p} \right], \tag{59}$$

and

$$R_0(\mathbf{u}, \mathbf{p}) = \lim_{\varepsilon \searrow 0} \frac{1}{|\Omega(\phi_\varepsilon) \triangle \Omega(\phi)|} \left[c_2(\mathbf{u} - \hat{\mathbf{u}})^\top (\tilde{\mathbf{M}}_\varepsilon - \tilde{\mathbf{M}})(\mathbf{u} - \hat{\mathbf{u}}) + \mathbf{p}^\top (\mathbf{A}_\varepsilon - \mathbf{A})\mathbf{u} - (\mathbf{f}_\varepsilon - \mathbf{f}) \cdot \mathbf{p} \right]. \tag{60}$$

Lemma 2 *There exist constants $c > 0$, $\hat{\varepsilon} > 0$ such that for all $\varepsilon \in (0, \hat{\varepsilon})$*

$$\|\mathbf{u}_\varepsilon - \mathbf{u}\| \leq c \varepsilon^o.$$

Here, $o = 1$ in the case of a shape perturbation and $o = 2$ in the case of a topological perturbation.

Proof Subtracting (54) for $\varepsilon = 0$ from the same equation with $\varepsilon > 0$, we get

$$\mathbf{A}_\varepsilon(\mathbf{u}_\varepsilon - \mathbf{u}) = \mathbf{f}_\varepsilon - \mathbf{f} - (\mathbf{A}_\varepsilon - \mathbf{A})\mathbf{u}$$

and thus, by the ellipticity of the bilinear form corresponding to $\mathbf{A}_\varepsilon$ and the triangle inequality, there is a constant $c > 0$ such that for all $\varepsilon > 0$ small enough

$$\|\mathbf{u}_\varepsilon - \mathbf{u}\| \le c(\|\mathbf{f}_\varepsilon - \mathbf{f}\| - \|\mathbf{A}_\varepsilon - \mathbf{A}\|\|\mathbf{u}\|). \tag{61}$$

For the difference between the perturbed and unperturbed right hand sides we have

$$|(\mathbf{f}_\varepsilon - \mathbf{f})_i| \le |(f_1 - f_2)| \int_{\Omega(\phi_\varepsilon)\triangle\Omega(\phi)} |\varphi_i(x)|\,\mathrm{d}x$$

$$|(\mathbf{A}_\varepsilon - \mathbf{A})_{i,j}| \le \int_{\Omega(\phi_\varepsilon)\triangle\Omega(\phi)} \left(|(\lambda_1 - \lambda_2)||\nabla\varphi_j(x)||\nabla\varphi_i(x)| + |(\alpha_1 - \alpha_2)||\varphi_j(x)||\varphi_i(x)| \right)\mathrm{d}x.$$

The result follows from the boundedness of $|\varphi_i(x)|$ and $|\nabla\varphi_i(x)|$ together with (49) which implies the existence of $c > 0$ such that $|\Omega(\phi_\varepsilon)\triangle\Omega(\phi)| \le c\varepsilon^o$ (cf. Remark 4). $\qquad\square$

From Lemma 2 it follows that the terms $R_1(\mathbf{u}, \mathbf{p})$ and $R_2(\mathbf{u}, \mathbf{p})$ vanish. We remark that this is in contrast to the continuous case, where asymptotic analysis shows that R_2 does not vanish. We will address this issue in more detail in Sect. 5. Thus, in the discrete setting we obtain $d\mathcal{J}(\phi)(\mathbf{x}_k) = R_0(\mathbf{u}, \mathbf{p})$, i.e.,

$$d\mathcal{J}(\phi)(\mathbf{x}_k) = \frac{\mathbf{p}^\top(d_k\mathbf{A}\,\mathbf{u} - d_k\mathbf{f}) + c_2(\mathbf{u} - \hat{\mathbf{u}})^\top d_k\tilde{\mathbf{M}}(\mathbf{u} - \hat{\mathbf{u}})}{d_k\tilde{a}} \tag{62}$$

where

$$d_k\mathbf{A} = \lim_{\varepsilon\searrow 0}\frac{\mathbf{A}_\varepsilon - \mathbf{A}}{\varepsilon^o}, \qquad d_k\tilde{\mathbf{M}} = \lim_{\varepsilon\searrow 0}\frac{\tilde{\mathbf{M}}_\varepsilon - \tilde{\mathbf{M}}}{\varepsilon^o},$$

$$d_k\mathbf{f} = \lim_{\varepsilon\searrow 0}\frac{\mathbf{f}_\varepsilon - \mathbf{f}}{\varepsilon^o}, \qquad d_k\tilde{a} = \lim_{\varepsilon\searrow 0}\frac{|\Omega(\phi_\varepsilon)\triangle\Omega(\phi)|}{\varepsilon^o}, \tag{63}$$

with $o = 1$ for $\mathbf{x}_k \in \mathfrak{S}$ and $o = 2$ for $\mathbf{x}_k \in \mathfrak{T}^- \cup \mathfrak{T}^+$. To obtain (62), we divided both numerator and denominator of (60) by ε^o and used that the limit of the quotient coincides with the quotient of the limits provided both limits exist and the limit in the denominator does not vanish. Next we state the numerical topological-shape derivative of problem (2) for arbitrary constant weights $c_1, c_2 \ge 0$ in the cost function (1).

Theorem 2 (Numerical topological-shape derivative) *For* $l = 1, \ldots, N$, *let* $\mathbf{u}_l = [u_{l_1}, u_{l_2}, u_{l_3}]^\top$ *and* $\mathbf{p}_l = [p_{l_1}, p_{l_2}, p_{l_3}]^\top$ *be the nodal values for element* τ_l *of the solution and the adjoint, and*

$$\mathbf{k}_{0,l}[i, j] = \left(J_l^{-1} \nabla_\xi \psi_j \right)^\top \left(J_l^{-1} \nabla_\xi \psi_i \right), \quad i, j = 1, \ldots, 3.$$

Moreover, $u_k = u(\mathbf{x}_k)$, $p_k = p(\mathbf{x}_k)$, *and* $\hat{u}_k = \hat{u}(\mathbf{x}_k)$. *For* $\mathbf{x}_k \in \mathfrak{T}^-$ *the numerical topological derivative reads*

$$d\mathcal{J}(\phi)(\mathbf{x}_k) = -c_1 - \Delta\lambda \frac{\sum_{l \in I_{\mathbf{x}_k}} \frac{\mathbf{p}_l^\top \mathbf{k}_{0,l} \mathbf{u}_l |\det J_l|}{\phi_{l_2} \phi_{l_3}}}{\sum_{l \in I_{\mathbf{x}_k}} \frac{|\det J_l|}{\phi_{l_2} \phi_{l_3}}} - \Delta\alpha\, p_k u_k + \Delta f p_k - c_2 \Delta\tilde{\alpha}(u_k - \hat{u}_k)^2,$$

(64)

whereas for $\mathbf{x}_k \in \mathfrak{T}^+$ *we have*

$$d\mathcal{J}(\phi)(\mathbf{x}_k) = c_1 + \Delta\lambda \frac{\sum_{l \in I_{\mathbf{x}_k}} \frac{\mathbf{p}_l^\top \mathbf{k}_{0,l} \mathbf{u}_l |\det J_l|}{\phi_{l_2} \phi_{l_3}}}{\sum_{l \in I_{\mathbf{x}_k}} \frac{|\det J_l|}{\phi_{l_2} \phi_{l_3}}} + \Delta\alpha\, p_k u_k - \Delta f p_k + c_2 \Delta\tilde{\alpha}(u_k - \hat{u}_k)^2,$$

(65)

with

$$\Delta\alpha = \alpha_1 - \alpha_2, \quad \Delta\lambda = \lambda_1 - \lambda_2, \quad \Delta f = f_1 - f_2, \quad \Delta\tilde{\alpha} = \tilde{\alpha}_1 - \tilde{\alpha}_2.$$

For $\mathbf{x}_k \in \mathfrak{S}$ *the numerical shape derivative reads*

$$d\mathcal{J}(\phi)(\mathbf{x}_k) = -c_1 + \Delta\lambda \frac{\sum_{l \in C_k} \mathbf{p}_l^\top \mathbf{k}_{0,l} \mathbf{u}_l\, d_k a_l}{d_k \tilde{a}} + \Delta\alpha \frac{\sum_{l \in C_k} \mathbf{p}_l^\top d_k \mathbf{m}_l^I\, \mathbf{u}_l}{d_k \tilde{a}}$$
$$- \Delta f \frac{\sum_{l \in C_k} \mathbf{p}_l^\top d_k \mathbf{f}_l^I}{d_k \tilde{a}} + c_2 \Delta\tilde{\alpha} \frac{\sum_{l \in C_k} (\mathbf{u}_l - \hat{\mathbf{u}}_l)^\top d_k \mathbf{m}_l^I\, (\mathbf{u}_l - \hat{\mathbf{u}}_l)}{d_k \tilde{a}},$$

(66)

where the entries of the element matrix $d_k \mathbf{m}_l^I$ *and of the element vector* $d_k \mathbf{f}_l^I$ *are dependent on the local cut situation (cases* $I = A^\pm$, $B^\pm$, $C^\pm$*) and are given in Appendix A.2. The values* $d_k \tilde{a}$ *can be computed by* (48) *considering Remark 4.*

Proof We evaluate (62) for $\mathbf{x}_k \in \mathfrak{T}^-$ and $O_{k,\varepsilon} = T_{k,\varepsilon}^{-\rightarrow+}$. Thus, $o = 2$ in (63). We note that

$$\mathbf{p}^\top d_k \mathbf{A}\, \mathbf{u} = \sum_{l=1}^N \mathbf{p}_l^\top (d_k \mathbf{m}_l + d_k \mathbf{k}_l) \mathbf{u}_l.$$

(67)

We have for $l \in I_{\mathbf{x}_k}$

$$
\begin{aligned}
\mathbf{k}_l(\phi_\varepsilon) - \mathbf{k}_l(\phi) &= \mathbf{k}_{0,l} \, |\det J_l| \left(\int_{\xi_1=0}^1 \int_{\xi_2=0}^{1-\xi_1} \lambda_1 H(-\phi_\varepsilon \circ \Phi_l) + \lambda_2 H(\phi_\varepsilon \circ \Phi_l) \, \mathrm{d}\xi_2 \, \mathrm{d}\xi_1 \right. \\
&\quad \left. - \int_{\xi_1=0}^1 \int_{\xi_2=0}^{1-\xi_1} \lambda_1 H(-\phi \circ \Phi_l) + \lambda_2 H(\phi \circ \Phi_l) \, \mathrm{d}\xi_2 \, \mathrm{d}\xi_1 \right) \\
&= \mathbf{k}_{0,l} \, |\det J_l| \left(\lambda_1 \int_{\xi_1=0}^1 \int_{\xi_2=0}^{1-\xi_1} (H(-\phi_\varepsilon \circ \Phi_l) - H(-\phi \circ \Phi_l)) \, \mathrm{d}\xi_2 \, \mathrm{d}\xi_1 \right. \\
&\quad \left. + \lambda_2 \int_{\xi_1=0}^1 \int_{\xi_2=0}^{1-\xi_1} (H(\phi_\varepsilon \circ \Phi_l) - H(\phi \circ \Phi_l)) \, \mathrm{d}\xi_2 \, \mathrm{d}\xi_1 \right) \\
&= \mathbf{k}_{0,l}(\lambda_1 - \lambda_2)\delta_{k,\varepsilon} a_l
\end{aligned}
$$

due to (31) because $|\Omega_\varepsilon| - |\Omega| = -(|D\backslash\Omega_\varepsilon| - |D\backslash\Omega|)$ and with (38) we obtain

$$
d_k \mathbf{k}_l = \lim_{\varepsilon \searrow 0} \frac{\mathbf{k}_l(\phi_\varepsilon) - \mathbf{k}_l(\phi)}{\varepsilon^2} = -\Delta\lambda \frac{|\det J_l|}{2\phi_{l_2}\phi_{l_3}} \mathbf{k}_{0,l}. \tag{68}
$$

Due to

$$
\int_{\xi_1=0}^{l_1} \int_{\xi_2=0}^{l_2\left(1-\frac{\xi_1}{l_1}\right)} \xi_1^a \xi_2^b \, \mathrm{d}\xi_2 \, \mathrm{d}\xi_1 = \frac{\varepsilon^{a+b+2}}{\phi_{l_2}^{a+1}\phi_{l_3}^{b+1}} + O(\varepsilon^{a+b+2})
$$

for some $a, b \in \mathbb{N}$, we have

$$
\begin{aligned}
d_k \mathbf{m}_l = \lim_{\varepsilon \searrow 0} \frac{\mathbf{m}_l(\phi_\varepsilon) - \mathbf{m}_l(\phi)}{\varepsilon^2} &= -\lim_{\varepsilon \searrow 0} \frac{\Delta\alpha}{\varepsilon^2} \int_{\xi_1=0}^{l_1} \int_{\xi_2=0}^{l_2\left(1-\frac{\xi_1}{l_1}\right)} \psi_i(\xi)\psi_j(\xi) \, |\det J_l| \, \mathrm{d}\xi_2 \, \mathrm{d}\xi_1 \\
&= \begin{bmatrix} -\frac{\Delta\alpha|\det J_l|}{2\phi_{l_2}\phi_{l_3}} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},
\end{aligned} \tag{69}
$$

and conclude

$$
\sum_{l=1}^N \mathbf{p}_l^\top d_k \mathbf{m}_l \mathbf{u}_l = -\Delta\alpha \, p_k u_k \sum_{l \in I_{\mathbf{x}_k}} \frac{|\det J_l|}{2\phi_{l_2}\phi_{l_3}}. \tag{70}
$$

Furthermore, with

$$
\begin{aligned}
d_k \mathbf{f}_l = \lim_{\varepsilon \searrow 0} \frac{\mathbf{f}_l(\phi_\varepsilon) - \mathbf{f}_l(\phi)}{\varepsilon^2} &= -\lim_{\varepsilon \searrow 0} \frac{\Delta f}{\varepsilon^2} \int_{\xi_1=0}^{l_1} \int_{\xi_2=0}^{l_2 \left(1 - \frac{\xi_1}{l_1}\right)} \psi_i(\xi) \, |\det J_l| \; \mathrm{d}\xi_2 \, \mathrm{d}\xi_1 \\
&= \begin{bmatrix} -\frac{\Delta f |\det J_l|}{2 \phi_{l_2} \phi_{l_3}} \\ 0 \\ 0 \end{bmatrix},
\end{aligned}
\tag{71}
$$

it follows that

$$
\mathbf{p}^\top d_k \mathbf{f} = \sum_{l=1}^{N} \mathbf{p}_l^\top d_k \mathbf{f}_l = -\Delta f p_k \sum_{l \in I_{\mathbf{x}_k}} \frac{|\det J_l|}{2 \phi_{l_2} \phi_{l_3}}.
\tag{72}
$$

Analogously to (69) we have

$$
d_k \tilde{\mathbf{m}}_l = \lim_{\varepsilon \searrow 0} \frac{\tilde{\mathbf{m}}_l(\phi_\varepsilon) - \tilde{\mathbf{m}}_l(\phi)}{\varepsilon^2} = \begin{bmatrix} -\frac{\Delta \tilde{\alpha} |\det J_l|}{2 \phi_{l_2} \phi_{l_3}} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},
\tag{73}
$$

and obtain

$$
\begin{aligned}
(\mathbf{u} - \hat{\mathbf{u}})^\top d_k \tilde{\mathbf{M}} (\mathbf{u} - \hat{\mathbf{u}}) &= \sum_{l \in I_{\mathbf{x}_k}} (\mathbf{u}_{0,l} - \hat{\mathbf{u}}_l)^\top d_k \tilde{\mathbf{m}}_l (\mathbf{u}_{0,l} - \hat{\mathbf{u}}_l) \\
&= -\Delta \tilde{\alpha} (u_k - \hat{u}_k)^2 \sum_{l \in I_{\mathbf{x}_k}} \frac{|\det J_l|}{2 \phi_{l_2} \phi_{l_3}}.
\end{aligned}
\tag{74}
$$

In the present situation, $d_k \tilde{a}$ is given by the absolute value of (38) (see also Remark 4),

$$
d_k \tilde{a} = \sum_{l \in I_{\mathbf{x}_k}^{A+}} \frac{|\det J_l|}{2 \phi_{l_2} \phi_{l_3}}.
\tag{75}
$$

By inserting (68), (70), (72), (74), and (75) in (62), together with Corollary 2, we obtain the sought expression (64). Formula (65) can be obtained in an analogous way as (64).

The formula in (66) follows directly from (62) together with Corollary 2. The values of $d_k \mathbf{m}_l^I$ and $d_k \mathbf{f}_l^I$ for all possible cut situations $I \in \{A^+, A^-, B^+, B^-, C^+, C^-\}$ are given in Appendix A.2 and were computed using symbolic computer algebra tools. A mathematical derivation of these terms is omitted here for brevity.   □

Remark 5 In this work, we restricted ourselves to the spatially two-dimensional setting $D \subset \mathbb{R}^2$ where D is discretized by a triangular mesh. We expect that, at the cost of

more technical computations, an extension to three-dimensional problems involving a tetrahedral mesh can be carried out along the lines of the presented work. For shape perturbations, one has to consider all different configurations how a tetrahedral element can be divided by a planar interface. In addition to the class of configurations where one of the four nodes is cut off (cf. Fig. 3 for the 2d case), one also has to consider all configurations where any two of the four vertices of the tetrahedron are on one side of the interface and two on the other side. For topological perturbations, only the first class of configurations will be of interest and no significant difference to the two-dimensional setting is expected.

Similarly, an extension of the presented method to more general settings such as more general cost functions of the type $g = g(\Omega, u, \nabla u)$ or spatially varying material coefficients can be obtained at the cost of more technical computations.

5 Connection Between Continuous and Discrete Sensitivities

The topological-shape derivative introduced in (25) and computed for model problem (2) in Theorem 2 represents a sensitivity of the discretized problem (23). In this section, we draw some comparisons with the classical topological and shape derivatives defined on the continuous level before discretization. While the purpose of this paper is to follow the idea *discretize-then-differentiate*, we consider the other way here for comparison reasons.

5.1 Connections Between Continuous and Discrete Topological Derivative

For comparison, we also illustrate the derivation of the continuous topological derivative according to (4) for problem (2). We use the same Lagrangian framework as introduced in Sect. 4.2, see also [18]. Given a shape $\Omega \in \mathcal{A}$, a point $z \in D \setminus \partial\Omega$, an inclusion shape $\omega \subset \mathbb{R}^d$ with $0 \in \omega$ and $\varepsilon \geq 0$, we define the inclusion $\omega_\varepsilon = z + \varepsilon\omega$ and the perturbed Lagrangian

$$
G(\varepsilon, \varphi, \psi) := c_1 |\Omega_\varepsilon| + c_2 \int_D \tilde{\alpha}_{\Omega_\varepsilon} |\varphi - \hat{u}|^2 \, \mathrm{d}x
$$
$$
+ \int_D \lambda_{\Omega_\varepsilon} \nabla\varphi \cdot \nabla\psi + \alpha_{\Omega_\varepsilon} \varphi\psi \, \mathrm{d}x - \int_D f_{\Omega_\varepsilon} \psi \, \mathrm{d}x
$$

where $\Omega_\varepsilon = \Omega \setminus \omega_\varepsilon$ for $z \in \Omega$ and $\Omega_\varepsilon = \Omega \cup \omega_\varepsilon$ for $z \in D \setminus \overline{\Omega}$. For simplicity, we only consider the latter case, *i.e.* $z \in D \setminus \overline{\Omega}$ in the sequel.

Noting that $u_\varepsilon, \varepsilon \geq 0$, is the solution to the perturbed state equation with parameter ε, the topological derivative can also be written as

$$
d_T \mathfrak{g}(\Omega)(z) = \lim_{\varepsilon \searrow 0} \frac{1}{|\omega_\varepsilon|} (G(\varepsilon, u_\varepsilon, p) - G(0, u_0, p)) \tag{76}
$$

with the solution to the unperturbed adjoint state equation p. As in Sect. 4.2, this leads to the topological derivative consisting of the three terms

$$d_T \mathfrak{g}(\Omega)(z) = R_1(u, p) + R_2(u, p) + R_0(u, p)$$

where

$$R_1(u, p) := \lim_{\varepsilon \searrow 0} \frac{1}{|\Omega_\varepsilon \Delta \Omega|} \int_0^1 [\partial_u G(\varepsilon, u_0 + s(u_\varepsilon - u_0), p) - \partial_u G(\varepsilon, u, p)](u_\varepsilon - u_0) ds,$$

$$R_2(u, p) := \lim_{\varepsilon \searrow 0} \frac{1}{|\Omega_\varepsilon \Delta \Omega|} [\partial_u G(\varepsilon, u, p) - \partial_u G(0, u, p)](u_\varepsilon - u),$$

$$R_0(u, p) := \lim_{\varepsilon \searrow 0} \frac{1}{|\Omega_\varepsilon \Delta \Omega|} [G(\varepsilon, u, p) - G(0, u, p)],$$

provided that these limits exist. It is straightforwardly checked that for $z \in D \setminus \overline{\Omega}$

$$R_0(u, p) = c_1 + c_2(\tilde{\alpha}_1 - \tilde{\alpha}_2)(u - \hat{u})^2(z) + (\lambda_1 - \lambda_2)\nabla u(z) \cdot \nabla p(z)$$
$$+ (\alpha_1 - \alpha_2)u(z)p(z) - (f_1 - f_2)(z)p(z).$$

For the term $R_2(u, p)$, we obtain

$$R_2(u, p) = \lim_{\varepsilon \searrow 0} \frac{1}{|\omega_\varepsilon|} \left[2c_2 \int_{\omega_\varepsilon} (\tilde{\alpha}_1 - \tilde{\alpha}_2)(u_0 - \hat{u})(u_\varepsilon - u_0) \, dy \right.$$
$$+ \int_{\omega_\varepsilon} (\lambda_1 - \lambda_2)\nabla(u_\varepsilon - u_0) \cdot \nabla p \, dy$$
$$+ \left. \int_{\omega_\varepsilon} (\alpha_1 - \alpha_2)(u_\varepsilon - u_0)p \, dy \right].$$

A change of variables $y = T_\varepsilon(x) = z + \varepsilon x$ yields

$$R_2(u, p) = \lim_{\varepsilon \searrow 0} \frac{1}{|\omega|} \left[2c_2(\tilde{\alpha}_1 - \tilde{\alpha}_2) \int_\omega (u_0 - \hat{u}) \circ T_\varepsilon (u_\varepsilon - u_0) \circ T_\varepsilon \, dx \right.$$
$$+ (\lambda_1 - \lambda_2) \int_\omega (\nabla(u_\varepsilon - u_0)) \circ T_\varepsilon \cdot (\nabla p) \circ T_\varepsilon \, dx \qquad (77)$$
$$+ \left. (\alpha_1 - \alpha_2) \int_\omega (u_\varepsilon - u_0) \circ T_\varepsilon \, p \circ T_\varepsilon \, dx \right]$$

We have a closer look at the diffusion term

$$R_2^\lambda(u, p) := \lim_{\varepsilon \searrow 0} \frac{1}{|\omega|} (\lambda_1 - \lambda_2) \int_\omega (\nabla(u_\varepsilon - u_0)) \circ T_\varepsilon \cdot (\nabla p) \circ T_\varepsilon \, dx. \qquad (78)$$

In the continuous setting, we now define $K_\varepsilon := \frac{1}{\varepsilon}(u_\varepsilon - u_0) \circ T_\varepsilon$ and use the chain rule $(\nabla \varphi) \circ T_\varepsilon = \frac{1}{\varepsilon}\nabla(\varphi \circ T_\varepsilon)$ to obtain

$$R_2^\lambda(u, p) = \lim_{\varepsilon \searrow 0} \frac{1}{|\omega|}(\lambda_1 - \lambda_2) \int_\omega \nabla K_\varepsilon \cdot (\nabla p) \circ T_\varepsilon \, \mathrm{d}x \tag{79}$$

Next, one can show the weak convergence $\nabla K_\varepsilon \rightharpoonup \nabla K$ for $K \in \dot{B}L(\mathbb{R}^2)$ being defined as the solution to the exterior problem

$$\int_{\mathbb{R}^d} \lambda_\omega \nabla K \cdot \nabla \psi \, \mathrm{d}x = -(\lambda_1 - \lambda_2) \int_\omega \nabla u(z) \cdot \nabla \psi \, \mathrm{d}x \quad \text{for all } \psi \in \dot{B}L(\mathbb{R}^2),$$

where $\dot{B}L(\mathbb{R}^2) := \{v \in H^1_{loc}(\mathbb{R}^2) : \nabla v \in L^2(\mathbb{R}^2)\}/\mathbb{R}$ is a Beppo-Levi space. Assuming continuity of ∇p around the point of perturbation z, it follows that

$$R_2^\lambda(u, p) = \frac{1}{|\omega|}(\lambda_1 - \lambda_2) \int_\omega \nabla K \cdot \nabla p(z) \, \mathrm{d}x. \tag{80}$$

It can be shown that the other terms in (77) vanish and thus $R_2(u, p) = R_2^\lambda(u, p)$. Finally, it follows from the analysis in [18, Sec. 5] that $R_1(u, p) + R_2(u, p) = \frac{1}{|\omega|}(\lambda_1 - \lambda_2) \int_\omega \nabla K \cdot \nabla p(z) \, \mathrm{d}x = R_2(u, p)$, thus $R_1(u, p) = 0$ and $d_T\mathfrak{g}(\Omega)(z) = R_0(u, p) + R_2^\lambda(u, p)$.

Remark 6 Comparing the topological derivative formula obtained here with the sensitivity for interior nodes $\mathbf{x}_k \in \mathfrak{T}^- \cup \mathfrak{T}^+$ obtained in Sect. 4, we see that the term corresponding to $R_2^\lambda(u, p)$, i.e., the term

$$\lim_{\varepsilon \searrow 0} \frac{1}{|\Omega(\phi_\varepsilon) \triangle \Omega(\phi)|}(\mathbf{K}_\varepsilon - \mathbf{K}_0)(\mathbf{u}_\varepsilon - \mathbf{u}_0) \cdot \mathbf{p}$$

in (59), vanishes in the discrete setting. This can be seen as follows: For $u_\varepsilon^h, \varepsilon \geq 0$, and $p^h \in V_h$, we have the expansion in the finite element basis

$$u_\varepsilon^h(x) = \sum_{i=1}^M u_\varepsilon^{(i)} \varphi_i, \qquad p^h(x) = \sum_{i=1}^M p^{(i)} \varphi_i.$$

If we now plug in these discretized functions into (78) and consider a fixed mesh size h, we get on the other hand

$$R_2^\lambda(u^h, p^h) = \lim_{\varepsilon \searrow 0} \frac{1}{|\omega|}(\lambda_1 - \lambda_2) \int_\omega (\nabla(u_\varepsilon^h - u_0^h)) \circ T_\varepsilon \cdot (\nabla p^h) \circ T_\varepsilon \, \mathrm{d}x$$

$$= \lim_{\varepsilon \searrow 0} \frac{1}{|\omega|}(\lambda_1 - \lambda_2) \sum_{i,j=1}^M (u_\varepsilon^{(i)} - u_0^{(i)}) p^{(j)} \int_\omega (\nabla \varphi_i)(z + \varepsilon x) \cdot (\nabla \varphi_j)(z + \varepsilon x) \, \mathrm{d}x = 0$$

where we used the continuity of $\varepsilon \mapsto \mathbf{u}_\varepsilon$ according to Lemma 2. Note that, since the mesh and the basis functions are assumed to be fixed and independent of ε, unlike in the continuous setting, here the continuity of the normal flux of the discrete solution u_ε^h across the interface $\partial\omega_\varepsilon$ is not preserved. We mention that, when using an extended discretization technique that accounts for an accurate resolution of the material interface (e.g. XFEM [25] or CutFEM [11]), the corresponding discrete sensitivities would include a term corresponding to $R_2^\lambda(u, p)$.

5.2 Connection Between Continuous and Discrete Shape Derivative

The continuous shape derivative $d_S \mathfrak{g}(\Omega)(V)$ for a PDE-constrained shape optimization problem given a shape $\Omega \in \mathcal{A}$ and a smooth vector field V can also be obtained via a Lagrangian approach. For our problem (2), it can be obtained as $d_S \mathfrak{g}(\Omega)(V) = \partial_t G(0, u, p)$ with

$$
\begin{aligned}
G(t, \varphi, \psi) :=& c_1 \int_\Omega \xi(t) \, \mathrm{d}x + c_2 \int_D \tilde{\alpha}_\Omega |\varphi - \hat{u} \circ T_t|^2 \xi(t) \, \mathrm{d}x \\
&+ \int_D \lambda_\Omega A(t) \nabla\varphi \cdot \nabla\psi + \alpha_\Omega \varphi\psi\xi(t) \, \mathrm{d}x - \int_D f_\Omega \psi\xi(t) \, \mathrm{d}x
\end{aligned}
$$

where $T_t(x) = x + t V(x)$, $\xi(t) = \det(\partial T_t)$, $A(t) = \xi(t)\partial T_t^{-1}\partial T_t^{-T}$, see [31] for a detailed description. In the continuous setting, the shape derivative reads in its volume form

$$
d_S \mathfrak{g}(\Omega)(V) = \int_D \mathcal{S}_1^\Omega : \partial V + \mathcal{S}_0^\Omega \cdot V \, \mathrm{d}x
$$

with $\mathcal{S}_1^\Omega$ and $\mathcal{S}_0^\Omega$ given in (11) and (12), respectively. Under certain smoothness assumptions, it can be transformed into the Hadamard or boundary form

$$
d_S \mathfrak{g}(\Omega)(V) = \int_{\partial\Omega} L \, (V \cdot n) \, \mathrm{d}S_x \tag{81}
$$

with $L = (\mathcal{S}_1^{\Omega,\mathrm{in}} - \mathcal{S}_1^{\Omega,\mathrm{out}})n \cdot n$ given by

$$
L = c_1 + c_2(\tilde{\alpha}_1 - \tilde{\alpha}_2)(u - \hat{u})^2 + (\alpha_1 - \alpha_2)up - (f_1 - f_2)p + L^\lambda \tag{82}
$$

where L^λ is given by

$$
\begin{aligned}
L^\lambda :=& (\lambda_1 - \lambda_2)(\nabla u \cdot \tau)(\nabla p \cdot \tau) - \left(\frac{1}{\lambda_1} - \frac{1}{\lambda_2}\right)(\lambda_\Omega \nabla u \cdot n)(\lambda_\Omega \nabla p \cdot n) \\
=& \lambda_1 \nabla u_\mathrm{in} \cdot \nabla p_\mathrm{in} - \lambda_2 \nabla u_\mathrm{out} \cdot \nabla p_\mathrm{out} \\
& - 2\lambda_1(\nabla u_\mathrm{in} \cdot n)(\nabla p_\mathrm{in} \cdot n) + 2\lambda_2(\nabla u_\mathrm{out} \cdot n)(\nabla p_\mathrm{out} \cdot n).
\end{aligned} \tag{83}
$$

Here, ∇u_{in}, ∇p_{in} and ∇u_{out}, ∇p_{out} denote the restrictions of the gradients to Ω and $D\backslash\overline{\Omega}$, respectively, and n denotes the unit normal vector pointing out of Ω. Note that, when using a finite element discretization which does not resolve the interface such that the gradients of the discretized state and adjoint variable are continuous, i.e. $\nabla u_{h,\text{in}} = \nabla u_{h,\text{out}}$ and $\nabla p_{h,\text{in}} = \nabla p_{h,\text{out}}$, (83) becomes

$$L_h^\lambda = (\lambda_1 - \lambda_2)\nabla u_h \cdot \nabla p_h - 2(\lambda_1 - \lambda_2)(\nabla u_h \cdot n)(\nabla p_h \cdot n) \qquad (84)$$

We now discretize the continuous shape derivative formula (81) for the vector field $V^{(k)}$ that is obtained from the perturbation of the level set function ϕ in (only) node $\mathbf{x}_k$, $k \in \mathfrak{S}$ fixed. For that purpose we fix $\phi \in S_h^1(D)$ and the corresponding domain $\Omega(\phi)$. Note that we consider $V^{(k)}$ to be supported only on the discretized material interface $\partial\Omega(\phi) \cap D$. We begin with the case of the pure volume cost function by setting $c_2 = 0$.

Proposition 1 *Let $c_2 = 0$ and $\mathbf{x}_k \in \mathfrak{S}$ fixed. Let $V^{(k)}$ the vector field that corresponds to a perturbation of the value of ϕ at position $\mathbf{x}_k$. Then*

$$d_S\mathfrak{g}(\Omega(\phi))(V^{(k)}) = c_1 \sum_{l \in C_k} d_k a_l$$

where $d_k a_l$ is as in (48).

Proof For $c_2 = 0$ we also have $p = 0$ and thus $L = c_1$, i.e., we are in the case of pure volume minimization. From (81) we know that $d_S\mathfrak{g}(\Omega(\phi))(V^{(k)}) = c_1 \int_{\partial\Omega(\phi)} V^{(k)} \cdot n \, dS_x$. First of all, we note that the vector field $V^{(k)}$ corresponding to a perturbation of ϕ at node x_k is only nonzero in elements τ_l for $l \in C_k$ with C_k as defined in (32). Thus, the shape derivative reduces to

$$d_S\mathfrak{g}(\Omega(\phi))(V^{(k)}) = c_1 \sum_{l \in C_k} \int_{\tau_l \cap \partial\Omega(\phi)} V^{(k)} \cdot n \, dS_x.$$

We compute the vector field $V^{(k)}$ and normal vector n explicitly depending on the cut situation. Recall the sets $I_{\mathbf{x}_k}^{A+}$, $I_{\mathbf{x}_k}^{A-}$, $I_{\mathbf{x}_k}^{B+}$, $I_{\mathbf{x}_k}^{B-}$, $I_{\mathbf{x}_k}^{C+}$, $I_{\mathbf{x}_k}^{C-}$ introduced in (34), see also Fig. 3.

Given two points $\mathbf{p}$ and $\mathbf{q}$ and their respective level set values a and b of different sign, $ab < 0$, we denote the root of the linear interpolating function by

$$\mathbf{y}(\mathbf{p}, \mathbf{q}, a, b) = \mathbf{p} - \frac{a}{b-a}(\mathbf{q} - \mathbf{p})$$

and note that $\frac{d}{da}\mathbf{y}(\mathbf{p}, \mathbf{q}, a, b) = -\frac{b}{(b-a)^2}(\mathbf{q} - \mathbf{p})$.

We begin with Configuration A. For an element index $l \in I_{\mathbf{x}_k}^{A+} \cup I_{\mathbf{x}_k}^{A-}$, we denote the vertices of element τ_l by $\mathbf{x}_{l_1}, \mathbf{x}_{l_2}, \mathbf{x}_{l_3}$ and assume their enumeration in counter-clockwise order with $\mathbf{x}_k = \mathbf{x}_{l_1}$. The corresponding values of the given level set function

ϕ are denoted by ϕ_{l_1}, ϕ_{l_2}, ϕ_{l_3}, respectively. We parametrize the line connecting the two roots of the perturbed level set function along the edges by

$$p^A[\varepsilon](s) = (1-s)\mathbf{y}(\mathbf{x}_{l_1}, \mathbf{x}_{l_2}, \phi_{l_1} + \varepsilon, \phi_{l_2}) + s\mathbf{y}(\mathbf{x}_{l_1}, \mathbf{x}_{l_3}, \phi_{l_1} + \varepsilon, \phi_{l_3})$$

and obtain the vector field corresponding to the perturbation of $\phi_k = \phi_{l_1}$ along the line $\tau_l \cap \partial\Omega(\phi)$ as

$$\hat{V}^A(s) = \frac{d}{d\varepsilon}p^A[\varepsilon](s)|_{\varepsilon=0} = (1-s)\frac{-\phi_{l_2}}{(\phi_{l_2} - \phi_{l_1})^2}(\mathbf{x}_{l_2} - \mathbf{x}_{l_1}) + s\frac{-\phi_{l_3}}{(\phi_{l_3} - \phi_{l_1})^2}(\mathbf{x}_{l_3} - \mathbf{x}_{l_1}).$$

Introducing the notation $d_{ki,kj} := |\mathbf{y}(\mathbf{x}_{l_k}, \mathbf{x}_{l_i}, \phi_{l_k}, \phi_{l_i}) - \mathbf{y}(\mathbf{x}_{l_k}, \mathbf{x}_{l_j}, \phi_{l_k}, \phi_{l_j})|$ for the length of the interface in element τ_l, the normed tangential vector along $\tau_l \cap \partial\Omega(\phi)$ and the normed normal vector pointing out of $\Omega(\phi)$ are given by

$$\hat{t}^A = \frac{\mathbf{y}(\mathbf{x}_{l_1}, \mathbf{x}_{l_3}, \phi_{l_1}, \phi_{l_3}) - \mathbf{y}(\mathbf{x}_{l_1}, \mathbf{x}_{l_2}, \phi_{l_1}, \phi_{l_2})}{d_{13,12}}$$

$$\hat{n}^{A+} = \frac{\phi_{l_1}}{d_{13,12}}\left(\frac{1}{\phi_{l_2} - \phi_{l_1}}R(\mathbf{x}_{l_2} - \mathbf{x}_{l_1}) - \frac{1}{\phi_{l_3} - \phi_{l_1}}R(\mathbf{x}_{l_3} - \mathbf{x}_{l_1})\right) = -\hat{n}^{A-}$$

where R denotes a 90 degree counter-clockwise rotation matrix, $R = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Noting that $(\mathbf{x}_{l_3} - \mathbf{x}_{l_1})^\top R(\mathbf{x}_{l_2} - \mathbf{x}_{l_1}) = |\det J_l| = -(\mathbf{x}_{l_2} - \mathbf{x}_{l_1})^\top R(\mathbf{x}_{l_3} - \mathbf{x}_{l_1})$ and $(\mathbf{x}_{l_j} - \mathbf{x}_{l_1})^\top R(\mathbf{x}_{l_j} - \mathbf{x}_{l_1}) = 0$, $j = 2, 3$, we get

$$\hat{V}^A(s) \cdot \hat{n}^{A+} = -\frac{|\det J_l|\phi_{l_1}}{d_{13,12}}\frac{\phi_{l_2}(\phi_{l_3} - \phi_{l_1}) + s\phi_{l_1}(\phi_{l_2} - \phi_{l_3})}{(\phi_{l_2} - \phi_{l_1})^2(\phi_{l_3} - \phi_{l_1})^2} = -\hat{V}^A(s) \cdot \hat{n}^{A-}. \tag{85}$$

Finally, by elementary computation we obtain for $l \in I_{\mathbf{x}_k}^{A+}$

$$\int_{\tau_l \cap \partial\Omega(\phi)} V^{(k)} \cdot n \, dS_x = d_{13,12}\int_0^1 \hat{V}^A(t) \cdot \hat{n}^{A+} \, dt$$

$$= |\det J_l|\phi_{l_1}\frac{-\phi_{l_2}\phi_{l_3} + \frac{1}{2}\phi_{l_1}(\phi_{l_2} + \phi_{l_3})}{(\phi_{l_2} - \phi_{l_1})^2(\phi_{l_3} - \phi_{l_1})^2}$$

and the same formula with a different sign for $l \in I_{\mathbf{x}_k}^{A-}$. Proceeding analogously, we obtain for $l \in I_{\mathbf{x}_k}^{B+}$ and $l \in I_{\mathbf{x}_k}^{C+}$

$$\hat{V}^B(s) \cdot \hat{n}^{B+} = \frac{|\det J_l|}{d_{23,21}}\frac{(1-s)\phi_{l_2}^2}{(\phi_{l_2} - \phi_{l_1})^2(\phi_{l_3} - \phi_{l_2})} = -\hat{V}^B(s) \cdot \hat{n}^{B-}, \tag{86}$$

$$\hat{V}^C(s) \cdot \hat{n}^{C+} = \frac{|\det J_l|}{d_{31,32}}\frac{(1-s)\phi_{l_3}^2}{(\phi_{l_3} - \phi_{l_1})^2(\phi_{l_2} - \phi_{l_3})} = -\hat{V}^C(s) \cdot \hat{n}^{C-}, \tag{87}$$

and further

$$\int_{\tau_l \cap \partial\Omega(\phi)} V^{(k)} \cdot n \, \mathrm{d}S_x = \frac{|\det J_l|}{2} \frac{\phi_{l_2}^2}{(\phi_{l_2} - \phi_{l_1})^2 (\phi_{l_3} - \phi_{l_2})}, \quad l \in I_{\mathbf{x}_k}^{B+},$$

$$\int_{\tau_l \cap \partial\Omega(\phi)} V^{(k)} \cdot n \, \mathrm{d}S_x = \frac{|\det J_l|}{2} \frac{-\phi_{l_3}^2}{(\phi_{l_3} - \phi_{l_1})^2 (\phi_{l_3} - \phi_{l_2})}, \quad l \in I_{\mathbf{x}_k}^{C+},$$

respectively. Again, the formulas for $l \in I_{\mathbf{x}_k}^{B-}$ and $l \in I_{\mathbf{x}_k}^{C-}$ just differ by a different sign.

Finally, comparing the computed values with the formulas of $d_k a_l$ (48) yields the claimed result. $\qquad\square$

In view of Proposition 1, the definition in (13) and Remark 4, we see that, in the case $c_2 = 0$, it holds

$$\hat{d}_S \mathfrak{g}(\Omega(\phi))(V^{(k)}) = \frac{c_1 \sum_{l \in C_k} d_k a_l}{\sum_{l \in C_k} d_k \tilde{a}_l} = -c_1, \tag{88}$$

which is in alignment with the first term of the formula in (66).

Next, we consider the general PDE-constrained case where $c_2 > 0$.

Proposition 2 *Let $c_1 = 0$ and $\mathbf{x}_k \in \mathfrak{S}$ fixed. Let $V^{(k)}$ the vector field that corresponds to a perturbation of the value of ϕ at position $\mathbf{x}_k$. Then*

$$
\begin{aligned}
d_S \mathfrak{g}(\Omega(\phi))(V^{(k)}) = \quad & \Delta\lambda \sum_{l \in C_k} \mathbf{p}_l^\top \mathbf{k}_{0,l} \mathbf{u}_l \, d_k a_l + \Delta\alpha \sum_{l \in C_k} \mathbf{p}_l^\top d_k \mathbf{m}_l^I \mathbf{u}_l \\
& - \Delta f \sum_{l \in C_k} \mathbf{p}_l^\top d_k \mathbf{f}_l^I + c_2 \Delta\tilde{\alpha} \sum_{l \in C_k} (\mathbf{u}_l - \hat{\mathbf{u}}_l)^\top d_k \mathbf{m}_l^I (\mathbf{u}_l - \hat{\mathbf{u}}_l) \\
& - 2\Delta\lambda \sum_{l \in C_k} (\nabla u_h \cdot n^I)|_{\tau_l \cap \partial\Omega(\phi)} (\nabla p_h \cdot n)|_{\tau_l \cap \partial\Omega(\phi)} d_k a_l,
\end{aligned}
\tag{89}
$$

where we use the same notation as in Theorem 2. In particular, $d_k \mathbf{m}_l^I$ and $d_k \mathbf{f}_l^I$ depend on the cut situation, $I \in \{A^+, A^-, B^+, B^-, C^+, C^-\}$ and are given explicitly in Appendix A.2.

Proof Let an element index $l \in C_k$ fixed and $\mathbf{u}_l = [u_{l_1}, u_{l_2}, u_{l_3}]^\top$, $\mathbf{p}_l = [p_{l_1}, p_{l_2}, p_{l_3}]^\top$ contain the nodal values of the finite element functions u_h and p_h corresponding to the three vertices $\mathbf{x}_{l_1}, \mathbf{x}_{l_2}, \mathbf{x}_{l_3}$ of element l, respectively. Also here, the ordering is in counter-clockwise direction starting with $\mathbf{x}_{l_1} = \mathbf{x}_k$. We compute the shape derivative (81) with L given in (82) after discretization (i.e. after replacing the functions u, p, $\hat{u}$ by finite element approximations u_h, p_h, $\hat{u}_h$). In particular, the term L^λ is approximated by (84). Depending on how the material interface $\partial\Omega(\phi)$ cuts through element τ_l, the normal component of the vector field $V^{(k)}$ along the line

$\tau_l \cap \partial\Omega(\phi)$ is given in (85)–(87). For and $I \in \{A^+, A^-, B^+, B^-, C^+, C^-\}$, it can be seen by elementary yet tedious calculations that

$$\int_{\tau_l \cap \partial\Omega(\phi)} p_h(x)V^I(x) \cdot n^I \, \mathrm{d}S_x = \mathbf{p}_l^\top d_k \mathbf{f}_l^I,$$

$$\int_{\tau_l \cap \partial\Omega(\phi)} u_h(x)p_h(x)V^I(x) \cdot n^I \, \mathrm{d}S_x = \mathbf{p}_l^\top d_k \mathbf{m}_l^I \mathbf{u}_l,$$

$$\int_{\tau_l \cap \partial\Omega(\phi)} (u_h(x) - \hat{u}_h(x))^2 V^I(x) \cdot n^I \, \mathrm{d}S_x = (\mathbf{u}_l - \hat{\mathbf{u}}_l)^\top d_k \mathbf{m}_l^I (\mathbf{u}_l - \hat{\mathbf{u}}_l),$$

with $d_k \mathbf{m}_l^I$ and $d_k \mathbf{f}_l^I$ as given in Appendix A.2. Examplarily, we illustrate the calculation for the second of these terms for the cut situation $I = A^+$, see Fig. 3a. Let $u_{l,12}$ and $u_{l,13}$ denote the values of the linear function $u_h|_{\tau_l}$ at the intersection of the interface $\partial\Omega(\phi)$ with the edges that connect the point $\mathbf{x}_{l_1}$ with $\mathbf{x}_{l_2}$ and $\mathbf{x}_{l_1}$ with $\mathbf{x}_{l_3}$, respectively. Note the relations $u_{l,12} = \frac{u_{l_2}\phi_{l_1} - u_{l_1}\phi_{l_2}}{\phi_{l_1} - \phi_{l_2}}$ and $u_{l,13} = \frac{u_{l_3}\phi_{l_1} - u_{l_1}\phi_{l_3}}{\phi_{l_1} - \phi_{l_3}}$. Analogously we define the values $p_{l,12}$ and $p_{l,13}$. The function u_h along the line $\tau_l \cap \partial\Omega(\phi)$ can now be written as $\hat{u}_h(s) = u_{l,12} + s(u_{l,13} - u_{l,12})$, $s \in [0, 1]$ and we get

$$\int_{\tau_l \cap \partial\Omega(\phi)} u_h(x)p_h(x)V^I(x) \cdot n^I \, \mathrm{d}S_x$$
$$= d_{13,12} \int_0^1 (u_{l,12} + s(u_{l,13} - u_{l,12}))(p_{l,12} + s(p_{l,13} - p_{l,12}))\hat{V}^{A^+}(s) \cdot \hat{n}^{A^+} \, \mathrm{d}s$$
$$= \mathbf{p}_l^\top d_k \mathbf{m}_l^{A^+} \mathbf{u}_l$$

where $d_{13,12} = |\tau_l \cap \partial\Omega(\phi)|$. The last equality is obtained by plugging in (85) and straightforward (yet tedious) calculation. Finally, since u_h and p_h are linear and the normal vector is constant on $\tau_l \cap \partial\Omega(\phi)$, we see that L_h^λ is constant and, using Proposition 1, we obtain

$$\int_{\tau_l \cap \partial\Omega(\phi)} L_h^\lambda(x)V^I(x) \cdot n^I \, \mathrm{d}S_x = L_h^\lambda(x) \int_{\tau_l \cap \partial\Omega(\phi)} V^I(x) \cdot n^I \, \mathrm{d}S_x$$
$$= \Delta\lambda \left(\mathbf{p}_l^\top \mathbf{k}_{0,l} \mathbf{u}_l - 2(\nabla u_h \cdot n^I)|_{\tau_l \cap \partial\Omega(\phi)} (\nabla p_h \cdot n^I)|_{\tau_l \cap \partial\Omega(\phi)} \right) d_k a_l.$$

□

Combining the findings of Propositions 1 and 2 and dividing by $d_k \tilde{a}$ (defined in Remark 4), we obtain the resulting formula for the alternative definition of the shape derivative as defined in (13)

$$\hat{d}_S \mathfrak{g}(\Omega(\phi))(V^{(k)}) = -c_1 + \frac{1}{d_k \tilde{a}} \left(\Delta\lambda \sum_{l \in C_k} \mathbf{p}_l^\top \mathbf{k}_{0,l} \mathbf{u}_l \, d_k a_l \right.$$
$$\left. + \Delta\alpha \sum_{l \in C_k} \mathbf{p}_l^\top d_k \mathbf{m}_l^I \mathbf{u}_l - \Delta f \sum_{l \in C_k} \mathbf{p}_l^\top d_k \mathbf{f}_l^I \right.$$

$$+c_2 \Delta \tilde{\alpha} \sum_{l \in C_k} (\mathbf{u}_l - \hat{\mathbf{u}}_l)^\top d_k \mathbf{m}_l^I (\mathbf{u}_l - \hat{\mathbf{u}}_l)$$

$$-2\Delta\lambda \sum_{l \in C_k} (\nabla u_h \cdot n^I)|_{\tau_l \cap \partial \Omega(\phi)} (\nabla p_h \cdot n)|_{\tau_l \cap \partial \Omega(\phi)} d_k a_l \Bigg). \tag{90}$$

Remark 7 Note that (90) is obtained by discretizing the continuous shape derivative (81)–(83). We see immediately that (90) resembles the formula for the discrete topological-shape derivative for nodes $\mathbf{x}_k \in \mathfrak{S}$ (66). The only difference is the occurence of the last term in (90), which is not accounted for when performing the sensitivity analysis in the discrete setting.

Note that this term stems from the presence of $\partial T_t^{-1} \partial T_t^{-T}$ in the matrix $A(t)$, which, in turn, originates from two applications of the chain rule, $(\nabla \varphi) \circ T_t = \partial T_t^{-T} \nabla(\varphi \circ T_t)$. Similarly as in the case of the topological derivative in Sect. 5.1, the reason for this discrepancy is the fact that, for the given discretization scheme, the gradients of the finite element basis functions are constant on each element and thus $(\nabla \varphi) \circ T_t = \nabla \varphi$ for small enough shape perturbation parameter t.

Remark 8 As a second, conceptual difference between the classical continuous shape derivative defined by (8) and our discrete counterpart (66), we recall that in the definition (25) we divided by the change of volume also in the case $\mathbf{x}_k \in \mathfrak{S}$.

Thus, comparing (90) with (66), we see that, for the chosen finite element setting, the discretization of the continuous shape derivative coincides with the discrete topological-shape derivative for the volume function, but not for the PDE-constrained part of the cost function (1) when the diffusion term is perturbed.

6 Numerical Experiments

In this section, we verify our implementation of the numerical topological-shape derivative derived in Sect. 4 by numerical experiments, before applying a level-set based topology optimization algorithm based on these sensitivities to our model problem.

6.1 Verification

The implementation of the topological-shape derivative is verified by comparing the computed sensitivity values against numerical values obtained by three different approaches. These are (i) a finite difference test, (ii) an application of the complex step derivative [24] and (iii) a test based on hyper-dual numbers developed in [16]. All tests are conducted for a fixed configuration.

We recall the definition of the topological-shape derivative (25) at a node $\mathbf{x}_k$ of the mesh

$$d\mathcal{J}(\phi)(\mathbf{x}_k) = \lim_{\varepsilon \searrow 0} \delta \mathcal{J}_\varepsilon(\phi)(\mathbf{x}_k) \quad \text{with} \quad \delta \mathcal{J}_\varepsilon(\phi)(\mathbf{x}_k) := \frac{\mathcal{J}(O_{k,\varepsilon}\phi) - \mathcal{J}(\phi)}{|\Omega(O_{k,\varepsilon}\phi) \triangle \Omega(\phi)|} \tag{91}$$

where $O_{k,\varepsilon}$ represents the operator $T_{k,\varepsilon}^{-\to+}$, $T_{k,\varepsilon}^{+\to-}$, $S_{k,\varepsilon}$ depending on whether the node $\mathbf{x}_k$ is in $\mathfrak{T}^-$, $\mathfrak{T}^+$ or $\mathfrak{S}$, respectively.

6.1.1 Finite Difference Test

For the finite difference (FD) test, we compute the errors

$$
e_S^{FD}(\varepsilon) = \sqrt{\sum_{\mathbf{x}_k \in \mathfrak{S}} (\delta\mathcal{J}_\varepsilon(\phi)(\mathbf{x}_k) - d\mathcal{J}(\phi)(\mathbf{x}_k))^2},
$$

$$
e_T^{FD}(\varepsilon) = \sqrt{\sum_{\mathbf{x}_k \in \mathfrak{T}^-\cup\mathfrak{T}^+} (\delta\mathcal{J}_\varepsilon(\phi)(\mathbf{x}_k) - d\mathcal{J}(\phi)(\mathbf{x}_k))^2} \tag{92}
$$

for a decreasing sequence of values for ε. The results are shown in Fig. 5a. We observe convergence of order ε up to a point where the cancellation error dominates.

6.1.2 Complex Step Derivative Test

In order to overcome this drawback of the finite difference test, we next consider a test based on the complex step (CS) derivative [24]. For the case of classical first order derivatives, this method is not subject to subtractive cancellation errors. For that purpose, using Remark 4, let us first rewrite (91) as

$$
d\mathcal{J}(\phi)(\mathbf{x}_k) = \frac{\lim_{\varepsilon\searrow 0} \frac{\mathcal{J}(O_{k,\varepsilon}\phi) - \mathcal{J}(\phi)}{\varepsilon^o}}{\lim_{\varepsilon\searrow 0} \frac{|\Omega(O_{k,\varepsilon}\phi)\triangle\Omega(\phi)|}{\varepsilon^o}} = \frac{1}{d_k\tilde{a}} \lim_{\varepsilon\searrow 0} \frac{\mathcal{J}(O_{k,\varepsilon}\phi) - \mathcal{J}(\phi)}{\varepsilon^o}, \tag{93}
$$

where $o = 1$ if $\mathbf{x}_k \in \mathfrak{S}$ and $o = 2$ if $\mathbf{x}_k \in \mathfrak{T}^- \cup \mathfrak{T}^+$. Moreover, assuming a higher order expansion of the form

$$
\begin{aligned}
\mathcal{J}(O_{k,\varepsilon}\phi) = \mathcal{J}(\phi) &+ \varepsilon^o\, d_k\tilde{a}\, d\mathcal{J}(\phi)(\mathbf{x}_k) \\
&+ \varepsilon^{o+1}\, d_k\tilde{a}\, d^2\mathcal{J}(\phi)(\mathbf{x}_k) + \varepsilon^{o+2}\, d_k\tilde{a}\, d^3\mathcal{J}(\phi)(\mathbf{x}_k) + \mathfrak{o}(\varepsilon^{o+2})
\end{aligned} \tag{94}
$$

with some higher order sensitivities $d^2\mathcal{J}(\phi)(\mathbf{x}_k)$, $d^3\mathcal{J}(\phi)(\mathbf{x}_k)$ and assuming that (94) also holds for complex-valued ε, we can follow the idea of the complex step derivative [24]: Setting $\varepsilon = ih$ in (94) with $h > 0$ and i the complex unit yields

$$
d\mathcal{J}(\phi)(\mathbf{x}_k) = \frac{\text{Im}(\mathcal{J}(O_{k,ih}\phi))}{h\, d_k\tilde{a}} + \mathcal{O}(h^2) \tag{95}
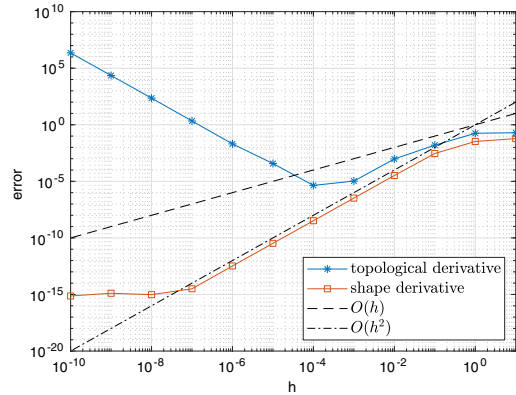$$

in the case $o = 1$ where $\mathbf{x}_k \in \mathfrak{S}$, and

$$
d\mathcal{J}(\phi)(\mathbf{x}_k) = \frac{\text{Re}(\mathcal{J}(O_{k,ih}\phi) - \mathcal{J}(\phi))}{-h^2\, d_k\tilde{a}} + \mathcal{O}(h^2)
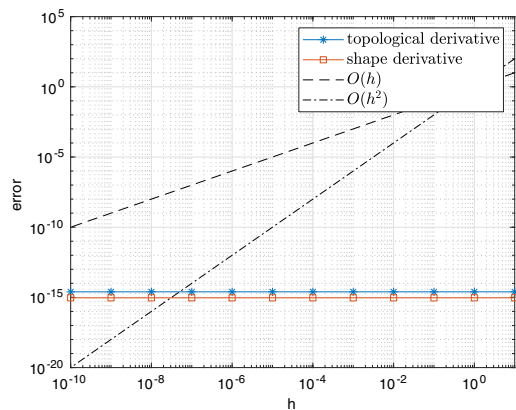$$

Fig. 5 **a** Results of the finite difference test. **b** Results obtained with the complex step derivative. **c** Results obtained with hyper-dual numbers
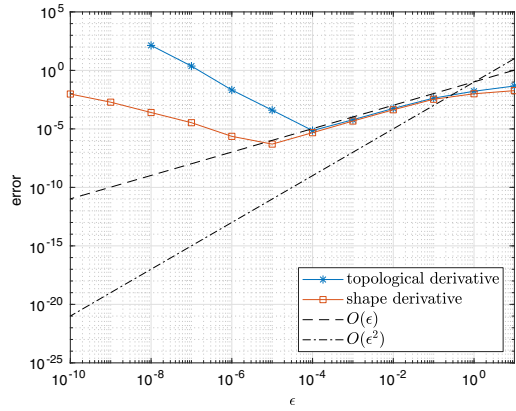
(a)

(b)

(c)

in the case $o = 2$ where $\mathbf{x}_k \in \mathfrak{T}^- \cup \mathfrak{T}^+$. This means

$$d\mathcal{J}(\phi)(\mathbf{x}_k) = \delta\mathcal{J}_h^{CS}(\phi)(\mathbf{x}_k) + \mathcal{O}(h^2)$$

with

$$\delta\mathcal{J}_h^{CS}(\phi)(\mathbf{x}_k) := \begin{cases} \frac{\mathrm{Re}(\mathcal{J}(T_{k,ih}^{-\rightarrow+}\phi) - \mathcal{J}(\phi))}{-h^2\, d_k \tilde{a}}, & \mathbf{x}_k \in \mathfrak{T}^-, \\ \frac{\mathrm{Re}(\mathcal{J}(T_{k,ih}^{+\rightarrow-}\phi) - \mathcal{J}(\phi))}{-h^2\, d_k \tilde{a}}, & \mathbf{x}_k \in \mathfrak{T}^+, \\ \frac{\mathrm{Im}(\mathcal{J}(S_{k,ih}\phi))}{h\, d_k \tilde{a}}, & \mathbf{x}_k \in \mathfrak{S}. \end{cases} \tag{96}$$

Analogously to (92), we define the summed errors $e_S^{CS}(h)$ and $e_T^{CS}(h)$ by just replacing $\delta\mathcal{J}_\varepsilon(\phi)(\mathbf{x}_k)$ by $\delta\mathcal{J}_h^{CS}(\phi)(\mathbf{x}_k)$ defined above. Figure 5b shows the errors e_S^{CS} and e_T^{CS} for a positive, decreasing sequence of h where we observe quadratic decay for both errors. While the error e_S^{CS} corresponding to the shape nodes $\mathbf{x}_k \in \mathfrak{S}$ decays to machine precision, the error e_T^{CS} corresponding to the interior nodes $\mathbf{x}_k \in \mathfrak{T}^- \cup \mathfrak{T}^+$ deteriorates at some point due to the cancellation error occurring when subtracting $\mathcal{J}(\phi)$ from $\mathcal{J}(O_{k,ih}\phi)$ in (96).

Remark 9 In this subsection, for the sake of code verification, we have assumed a complex-valued perturbation of the level-set function. This has to be understood in a fully numerical context, rather than in the continuos setting. In particular, the representation of the domain by the level-set function (18) is not applicable any more, due to the occurrence of complex numbers in a comparison of numbers. However, within the numerical method to solve the state equation and to evaluate the objective function, comparison of numbers is only needed for the classification of 'cut elements'. Apart from classification the finite element method (computing local matrices, assembly of the global matrices and solving the system of linear equations) uses only formulas consisting of basic operations $(+,-,\cdot,:)$, where complex values are not a problem. For instance, the area of an element that is cut by a perturbed level-set function is written as an analytic formula depending on the perturbation parameter ε in (35). Here, a complex valued perturbation parameter can be treated without problems.

When it comes to element classification the code has to follow the same execution path for the complex perturbation as if the perturbation were only real valued. To this end we distinguish between nodes belonging to $\mathfrak{S}$ and nodes belonging to $\mathfrak{T}^- \cup \mathfrak{T}^+$. In the case of a node $\mathbf{x}_k \in \mathfrak{S}$ simply the real part of the complex level-set value can be used. This is due to the fact that for small enough perturbations the cut situation is the same as for the unperturbed setting. In the case of nodes $\mathbf{x}_k$ belonging to $\mathfrak{T}^-$ and $\mathfrak{T}^+$, the operators $T_{k,ih}^{+\rightarrow-}$ and $T_{k,ih}^{-\rightarrow+}$ provide numbers with vanishing real part at $\mathbf{x}_k$, i.e. $\mathrm{Re}(T_{k,ih}^{+\rightarrow-}\phi(\mathbf{x}_k)) = \mathrm{Re}(T_{k,ih}^{-\rightarrow+}\phi(\mathbf{x}_k)) = 0$. Here, the imaginary part of the nodal value of the perturbed level-set function has the same sign as in the case of a purely real perturbation and can thus be used in the classification.

6.1.3 Test Based on Hyper-Dual Numbers

In order to overcome this cancellation error also for the case of $\mathbf{x}_k \in \mathfrak{T}^- \cup \mathfrak{T}^+$, we resort to hyper-dual (HD) numbers as introduced in [16]. Here, the idea is to consider numbers with three non-real components denoted by E_1, E_2 and $E_1 E_2$ with $E_1^2 = E_2^2 = (E_1 E_2)^2 = 0$. Assuming that expansion (94) holds up to order $o + 1$ also for such hyper-dual values of ε, we can set $\varepsilon = h E_1 + h E_2 + 0 E_1 E_2$ for some $h > 0$. For $o = 1$, considering only the first non-real part (i.e., the E_1 part) and exploiting that $E_1^2 = 0$, we obtain the equality

$$d\mathcal{J}(\phi)(\mathbf{x}_k) = \frac{E_1 \text{part}(\mathcal{J}(O_{k,h E_1 + h E_2}\phi))}{h\, d_k \tilde{a}} \tag{97}$$

for $\mathbf{x}_k \in \mathfrak{S}$. Similarly, with the same choice of ε, by considering only the $E_1 E_2$-part of the expansion and exploiting $E_1^2 = E_2^2 = E_1^2 E_2^2 = 0$, we obtain for $o = 2$

$$d\mathcal{J}(\phi)(\mathbf{x}_k) = \frac{E_1 E_2 \text{part}(\mathcal{J}(O_{k,h E_1 + h E_2}\phi))}{2h^2\, d_k \tilde{a}} \tag{98}$$

for $\mathbf{x}_k \in \mathfrak{T}^- \cup \mathfrak{T}^+$. In this case, the corresponding summed errors $e_S^{HD}(h)$ and $e_T^{HD}(h)$ vanish for arbitrary $h \in \mathbb{R}$. This is also observed numerically since both (97) and (98) suffer neither from a truncation nor a cancellation error. Figure 5c shows that the the obtained results agree up to machine precision with the derivatives obtained by (64), (65), and the respective formula for the shape derivative (66).

Remark 10 Analogously to Remark 9, the perturbation of the level-set function by a hyper-dual number has to be understood in the numerical setting. We have implemented the hyper-dual numbers using operator overloading, *i.e.* fundamental operations were extended to work also on hyper-dual numbers respecting the calculation rules described in [16]. Thus formulas involving basic operations do not pose a problem. Nevertheless, care has to be taken when it comes to element classification. Again, in the case of nodes belonging to $\mathfrak{S}$ simply the real part of the hyper-dual number can be used. For topological perturbations the classification is based on the E_1 part.

6.2 Application of Optimization Algorithm to Model Problem

Finally we show the use of the numerical topological-shape derivative computed in Sect. 4 within a level-set based topology optimization algorithm. We first state the precise model problem, before introducing the algorithm and showing numerical results.

6.2.1 Problem Setting

We consider the unit square $D = [0, 1]^2$ and minimize the objective function (1) with $c_1 = 0$ and $c_2 = 1$ subject to the PDE constraint in (2). The chosen problem parameters are shown in Table 1.

Table 1 Problem parameters for the numerical experiment

$\tilde{\alpha}_1$	$\tilde{\alpha}_2$	α_1	α_2	λ_1	λ_2	f_1	f_2
1	0.9	1	0.2	1	0.6	1	0.5

We consider a mixed Dirichlet-Neumann problem by choosing

$$\Gamma_D = \{(x, y) \in \partial D | y = 0 \text{ or } y = 1\}, \qquad \Gamma_N = \partial D \setminus \Gamma_D,$$

and $g_D(x, y) = y$, $g_N(x, y) = 0$.

In order to define a desired state $\hat{u}$, we choose a level-set function ϕ_d which implies a desired shape Ω_d, compute the corresponding solution u^* to ϕ_d and set $\hat{u} := u^*$. Then, by construction, $(\Omega_d, \hat{u})$ is also the solution of the design optimization problem. In the numerical tests we used five different meshes with 145, 545, 2113, 8321, and 33025 nodes respectively. For each mesh we obtain ϕ_d by interpolation of

$$\bar{\phi}_d(x, y) = \left((x - 0.3)^2 + (y - 0.4)^2 - 0.2^2\right)\left((x - 0.7)^2 + (y - 0.7)^2 - 0.1^2\right). \tag{99}$$

This yields that Ω are two (approximated) circles with radii 0.2 and 0.1 respectively, see Fig. 6.

6.2.2 Optimization Algorithm

The optimization algorithm we use to solve the problem introduced in Sect. 6.2.1 is inspired by [4].

Definition 1 We say a level set function $\phi \in S_h^1(D)$ is locally optimal for the problem described by $\mathcal{J}$ if

$$\begin{cases} d\mathcal{J}(\phi)(\mathbf{x}_k) \geq 0 & \text{for } \mathbf{x}_k \in \mathfrak{T}^- \cup \mathfrak{T}^+, \\ d\mathcal{J}(\phi)(\mathbf{x}_k) = 0 & \text{for } \mathbf{x}_k \in \mathfrak{S}. \end{cases} \tag{100}$$

We introduce the generalized numerical topological-shape derivative $G_\phi \in S_h^1(D)$ with

$$G_\phi(\mathbf{x}_k) = \begin{cases} -\min(d\mathcal{J}(\phi)(\mathbf{x}_k), 0) & \text{for } \mathbf{x}_k \in \mathfrak{T}^-, \\ \min(d\mathcal{J}(\phi)(\mathbf{x}_k), 0) & \text{for } \mathbf{x}_k \in \mathfrak{T}^+, \\ -d\mathcal{J}(\phi)(\mathbf{x}_k) & \text{for } \mathbf{x}_k \in \mathfrak{S}. \end{cases} \tag{101}$$

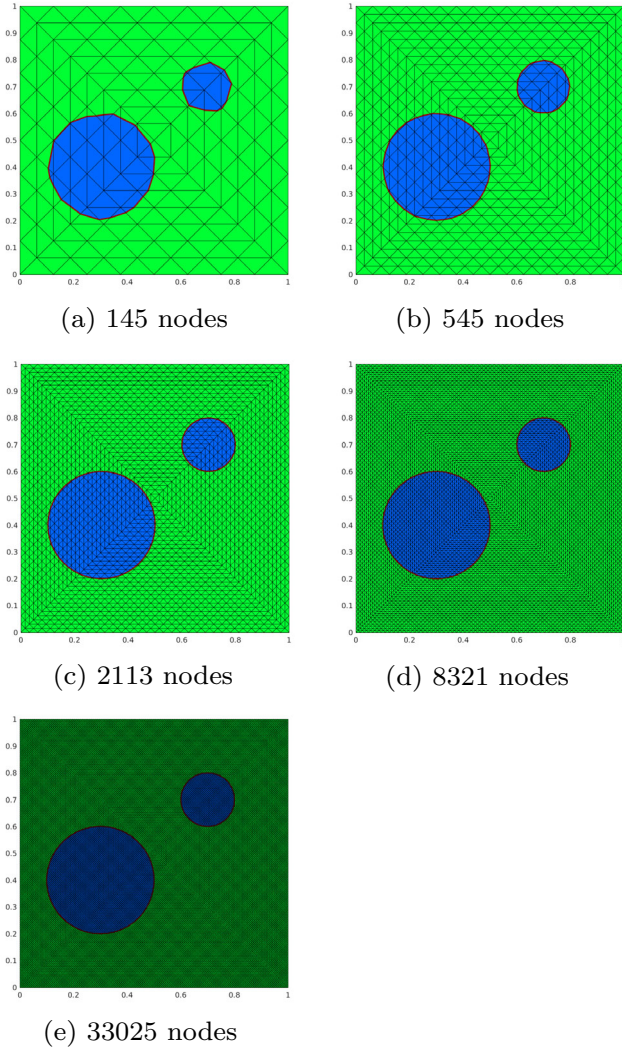With this definition, we immediately get the following optimality condition:

(a) 145 nodes

(b) 545 nodes

(c) 2113 nodes

(d) 8321 nodes

(e) 33025 nodes

Fig. 6 The different meshes and corresponding sought shapes used in the numerical experiments

Lemma 3 *Let $\phi \in S_h^1(D)$ and*

$$G_\phi(\mathbf{x}_k) = 0, \quad for\ k = 1, \ldots, M. \tag{102}$$

Then ϕ is locally optimal in the sense of Definition 1.

The update of the level-set function based on the information of the topological-shape derivative is done by spherical linear interpolation (see also [4])

$$\phi_{i+1} = \frac{1}{\sin(\theta_i)} \left(\sin((1-\kappa_i)\theta_i)\phi_i + \sin(\kappa_i\theta_i)\frac{G_{\phi_i}}{\|G_{\phi_i}\|_{L_2(D)}} \right), \tag{103}$$

where $\theta_i = \arc\cos((\phi_i, G_{\phi_i})_{L^2(D)})$ is the angle between the given level set function ϕ_i and the sensitivity G_{ϕ_i} in an $L^2(D)$-sense. Here, $\kappa_i \in (0, 1)$ is a line search parameter which is adapted such that a decrease in the objective function is achieved. Note that, by construction, the update (103) preserves the $L^2(D)$-norm, $\|\phi_{i+1}\|_{L^2(D)} = \|\phi_i\|_{L^2(D)}$. As in [4, 17], we can also show that ϕ is evolving along a descent direction:

Lemma 4 *Let $\phi_i, \phi_{i+1} \in S_h^1(D)$ two subsequent iterates related by (103). Then we have for $\mathbf{x}_k \in \mathfrak{T}^-(\phi_i) \cup \mathfrak{T}^+(\phi_i)$*

$$\phi_i(\mathbf{x}_k) > 0 > \phi_{i+1}(\mathbf{x}_k) \implies d\mathcal{J}(\phi_i)(\mathbf{x}_k) < 0, \tag{104}$$

$$\phi_i(\mathbf{x}_k) < 0 < \phi_{i+1}(\mathbf{x}_k) \implies d\mathcal{J}(\phi_i)(\mathbf{x}_k) < 0. \tag{105}$$

Proof Let $\mathbf{x}_k \in \mathfrak{T}^+(\phi_i)$, i.e. $\phi_i(\mathbf{x}_k) > 0$ and assume that $\phi_{i+1}(\mathbf{x}_k) < 0$. Since $\sin(\theta) > 0$ and $\sin(s\theta) > 0$ for all $\theta \in (0, \pi)$ and $s \in (0, 1)$, it follows from (103) that $G_{\phi_i}(x_k) < 0$ and thus, by (101), $d\mathcal{J}(\phi_i)(\mathbf{x}_k) < 0$ as claimed in (104). An analogous argument yields (105). $\square$

We can also show that G_ϕ constitutes a descent direction for $\mathbf{x}_k \in \mathfrak{S}$.

Lemma 5 *Let $\phi \in S_h^1(D)$ and suppose that*

$$\lim_{\varepsilon \searrow 0} \frac{\mathcal{J}(S_{k,\varepsilon}\phi) - \mathcal{J}(\phi)}{|\Omega(S_{k,\varepsilon}\phi) \triangle \Omega(\phi)|} = -\lim_{\varepsilon \nearrow 0} \frac{\mathcal{J}(S_{k,\varepsilon}\phi) - \mathcal{J}(\phi)}{|\Omega(S_{k,\varepsilon}\phi) \triangle \Omega(\phi)|}. \tag{106}$$

Let $\mathbf{x}_k \in \mathfrak{S}(\phi)$ be fixed and let ϕ^κ be the level set function according to (103) with line search parameter $\kappa \in (0, 1)$ that is updated only in $\mathbf{x}_k$, i.e., $\phi^\kappa = a(\kappa)\phi + b(\kappa)G_\phi\varphi_k$ with $a(\kappa) = \sin((1-\kappa)\theta)/\sin(\theta)$ and $b(\kappa) = \sin(\kappa\theta)/(\sin(\theta)\|G_\phi\|_{L_2(D)})$. Moreover assume that $|d\mathcal{J}(\phi)(\mathbf{x}_k)| > 0$. Then there exists $\overline{\kappa} \in (0, 1)$ such that for all $\kappa \in (0, \overline{\kappa})$

$$\mathcal{J}(\phi^\kappa) < \mathcal{J}(\phi).$$

Proof Suppose that $0 > d\mathcal{J}(\phi)(\mathbf{x}_k)$. Then it follows from (25) that $\mathcal{J}(\phi + \varepsilon\varphi_k) < \mathcal{J}(\phi)$ for $\varepsilon > 0$ small enough. Thus, since $a(\kappa), b(\kappa) > 0$ for $\theta \in (0, \pi)$ and $\kappa \in (0, 1)$ and since $\mathcal{J}(\phi^\kappa) = \mathcal{J}(\frac{1}{a(\kappa)}\phi^\kappa)$, it follows

$$\mathcal{J}(\phi^\kappa) = \mathcal{J}(\phi + b(\kappa)/a(\kappa)G_\phi\varphi_k) = \mathcal{J}(\phi_i - b(\kappa)/a(\kappa)d\mathcal{J}(\phi_i)(\mathbf{x}_k)\varphi_k) < \mathcal{J}(\phi_i)$$

for $\kappa > 0$ small enough since $-b(\kappa)/a(\kappa)d\mathcal{J}(\phi_i)(\mathbf{x}_k) > 0$ and $b(\kappa)/a(\kappa) \to 0$ as $\kappa \searrow 0$. On the other hand, if $0 < d\mathcal{J}(\phi_i)(\mathbf{x}_k)$, it follows from (25) and (106) that $\mathcal{J}(\phi + \varepsilon\varphi_k) < \mathcal{J}(\phi)$ for $\varepsilon < 0$ small enough and further for κ small enough

$$\mathcal{J}(\phi^\kappa) = \mathcal{J}(\phi - b(\kappa)/a(\kappa)d\mathcal{J}(\phi)(\mathbf{x}_k)\varphi_k) < \mathcal{J}(\phi).$$

$\square$

Remark 11 In the continuous setting, the property corresponding to (106) is fulfilled for smooth domains which can be seen as follows. Let $\Omega_t^V = (\mathrm{id} + tV)(\Omega)$ and note that $\Omega_{-t}^V = \Omega_t^{-V}$. Then, by Lemma 1,

$$\lim_{s \nearrow 0} \frac{|\Omega_s^V \triangle \Omega|}{s} = -\lim_{t \searrow 0} \frac{|\Omega_t^{-V} \triangle \Omega|}{t} = -\int_{\partial \Omega} |V \cdot n| \, \mathrm{d}S_x = -\lim_{s \searrow 0} \frac{|\Omega_s^V \triangle \Omega|}{s}$$

and, assuming differentiability of $s \mapsto \mathfrak{g}(\Omega_s^V)$,

$$\lim_{s \searrow 0} \frac{\mathfrak{g}(\Omega_s^V) - \mathfrak{g}(\Omega)}{|\Omega_s^V \triangle \Omega|} = \frac{\lim_{s \to 0} \left(\mathfrak{g}(\Omega_s^V) - \mathfrak{g}(\Omega) \right)/s}{\lim_{s \searrow 0} |\Omega_s^V \triangle \Omega|/s} = -\lim_{s \nearrow 0} \frac{\mathfrak{g}(\Omega_s^V) - \mathfrak{g}(\Omega)}{|\Omega_s^V \triangle \Omega|}.$$

In the discrete case, however, there may occur situations where the limits in (106) do not coincide. This can be the case in particular in situations where $\phi(\mathbf{x}_k) = 0$. We remark that this issue seemed not to cause problems in our numerical experiments.

Remark 12 In practice it turned out to be advantageous to include a smoothing step of the level set function. Thus, we chose the following update strategy: We first set

$$\psi = \frac{1}{\sin(\theta_i)} \left(\sin((1 - \kappa_i)\theta_i)\phi_i + \sin(\kappa_i \theta_i) \frac{G_{\phi_i}}{\|G_{\phi_i}\|_{L_2(D)}} \right),$$

with the same notation as above before smoothing the level set function in $\mathfrak{T}^-(\psi) \cup \mathfrak{T}^+(\psi)$ by

$$\hat{\psi}(\mathbf{x}_k) = \begin{cases} \frac{\sum_{i \in R_{\mathbf{x}_k}} \psi(\mathbf{x}_i)}{|R_{\mathbf{x}_k}|} & \text{for } \mathbf{x}_k \in \mathfrak{T}^-(\psi) \cup \mathfrak{T}^+(\psi), \\ \psi(\mathbf{x}_k) & \text{for } \mathbf{x}_k \in \mathfrak{S}. \end{cases} \tag{107}$$

Finally, the level-set function is normalized and the next iterate is given by

$$\phi_{i+1} = \frac{\hat{\psi}}{\|\hat{\psi}\|_{L_2(D)}}. \tag{108}$$

Remark 13 Compared to a classical level set method [2] which is based solely on shape sensitivities, our method does not lack a nucleation mechanism and thus it is not necessary to use perforated initial designs or alternating shape or topological update steps [1, 10]. While this is also true for the level set method introduced in [4], which is based solely on topological derivatives, that method typically uses an average of topological derivatives as sensitivities at the material interfaces, which does not necessarily represent the correct shape sensitivities. Moreover, in contrast to these mentioned methods, our approach uses sensitivities of the discretized problem.
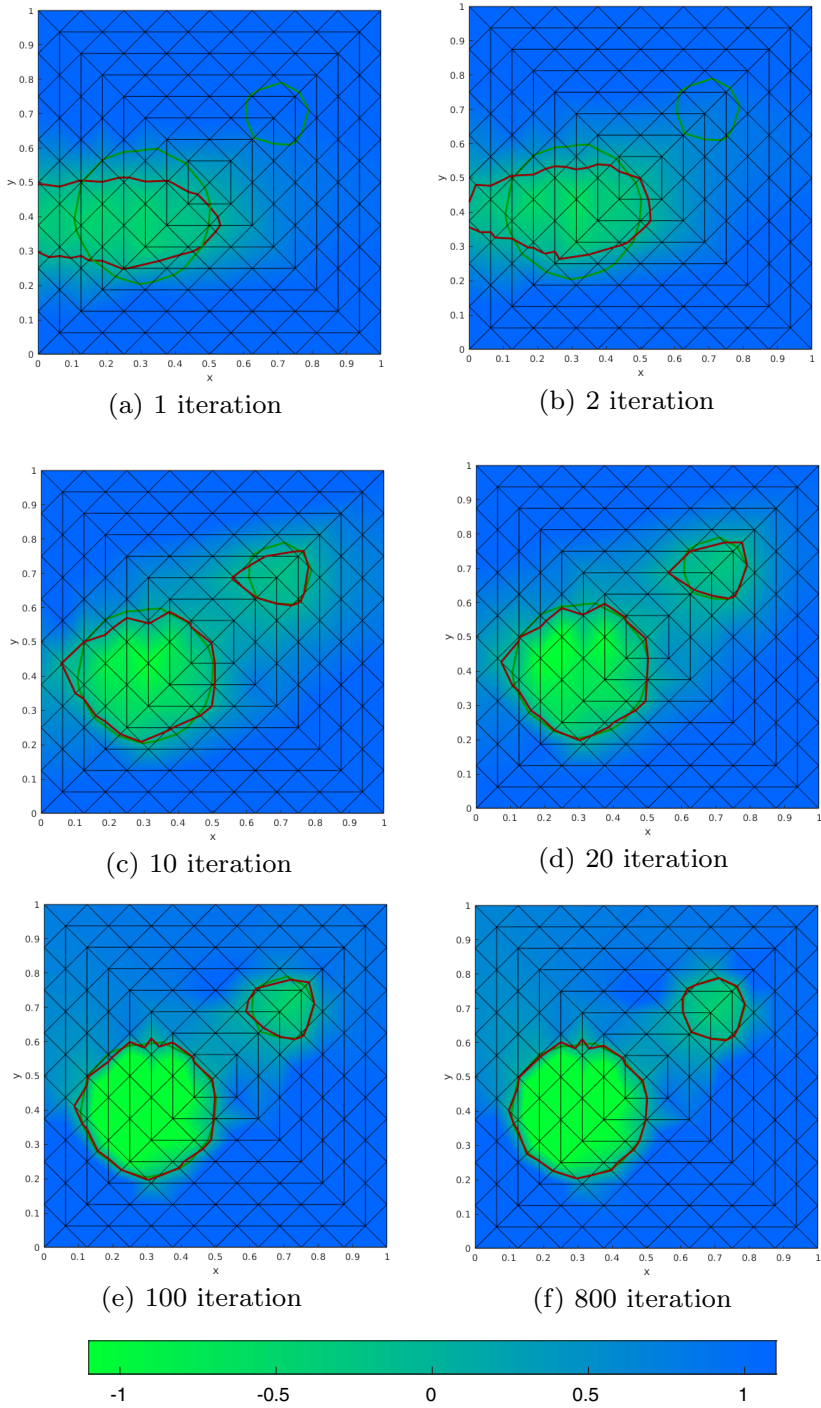
(a) 1 iteration

(b) 2 iteration

(c) 10 iteration

(d) 20 iteration

(e) 100 iteration

(f) 800 iteration

Fig. 7 Evolution of the level-set function for the 145 nodes mesh

(a) 1 iteration

(b) 2 iteration

(c) 10 iteration

(d) 20 iteration

(e) 100 iteration

(f) 800 iteration
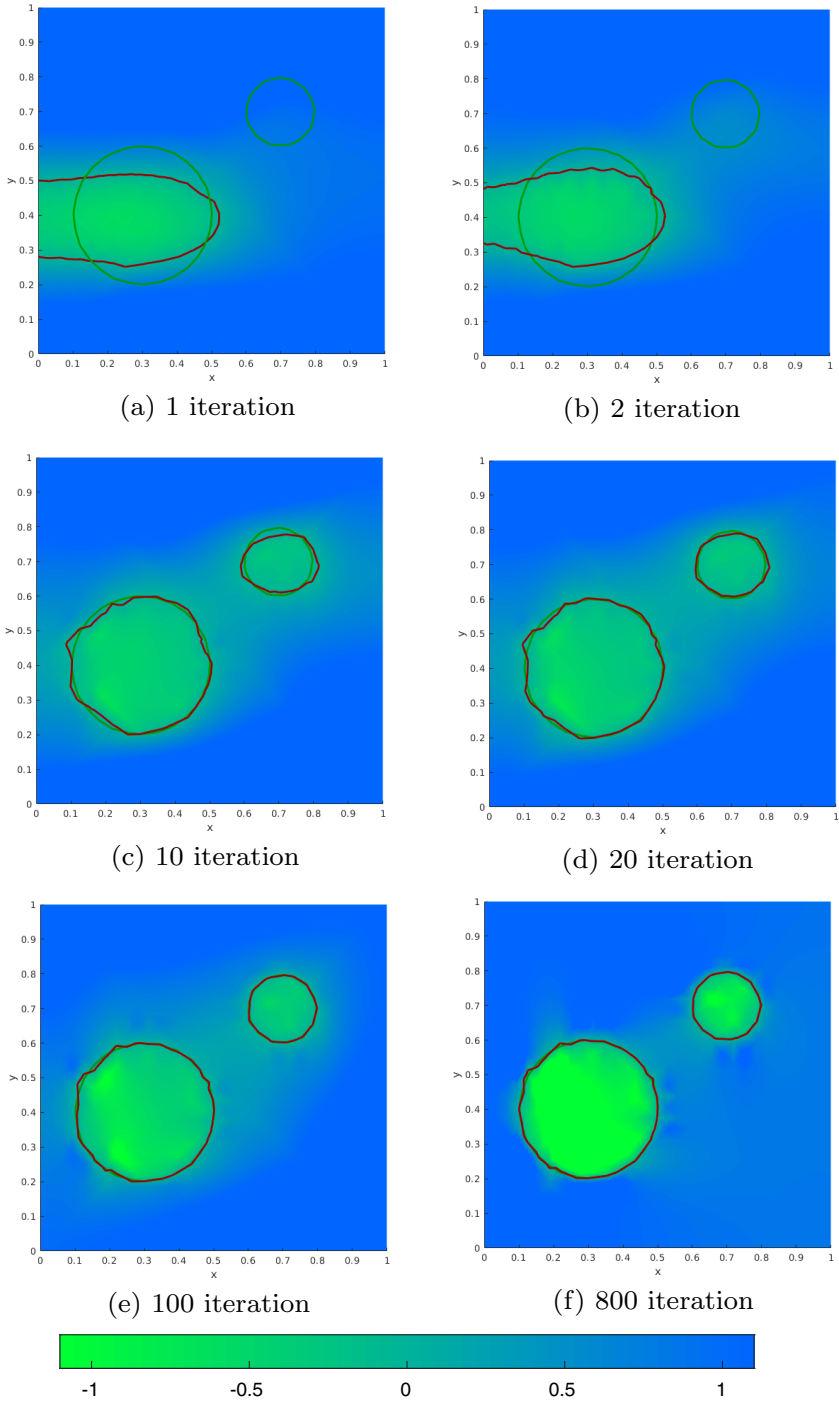
Fig. 8 Evolution of the level-set function for the 545 nodes mesh

(a) 1 iteration

(b) 2 iteration

(c) 10 iteration

(d) 20 iteration

(e) 100 iteration

(f) 800 iteration
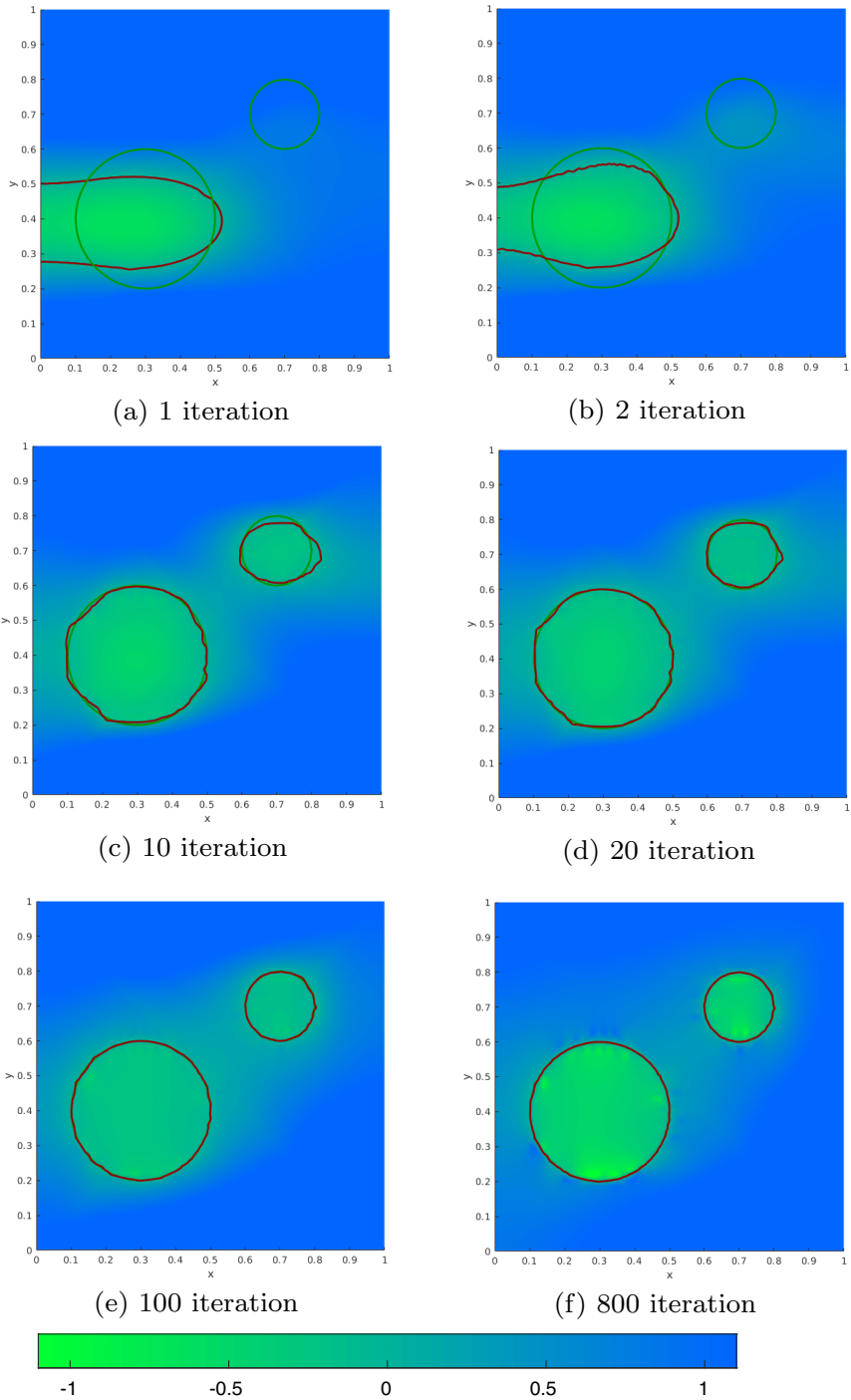
Fig. 9 Evolution of the level-set function for the 2113 nodes mesh
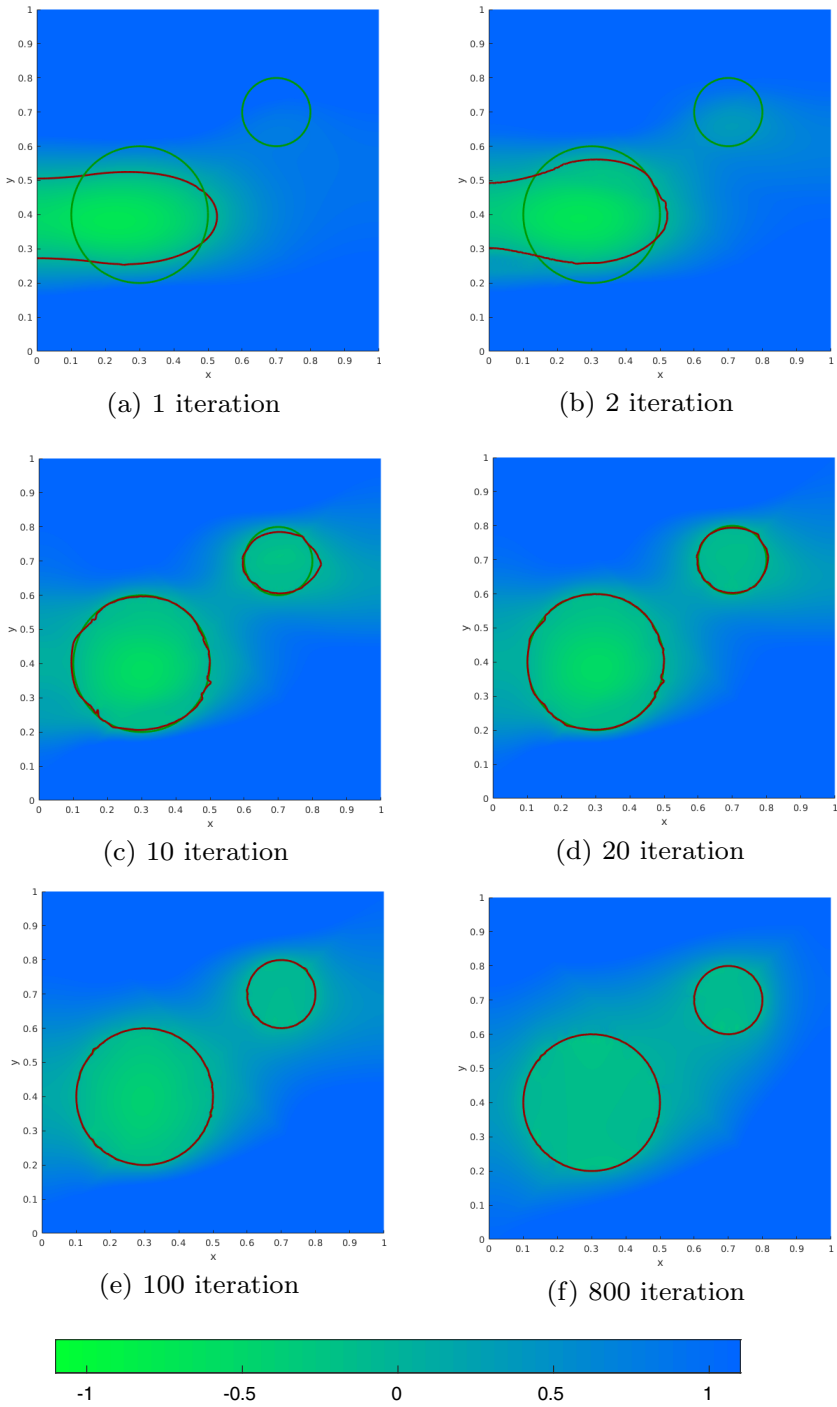
Fig. 10 Evolution of the level-set function for the 8321 nodes mesh

(a) 1 iteration

(b) 2 iteration

(c) 10 iteration

(d) 20 iteration

(e) 100 iteration

(f) 800 iteration
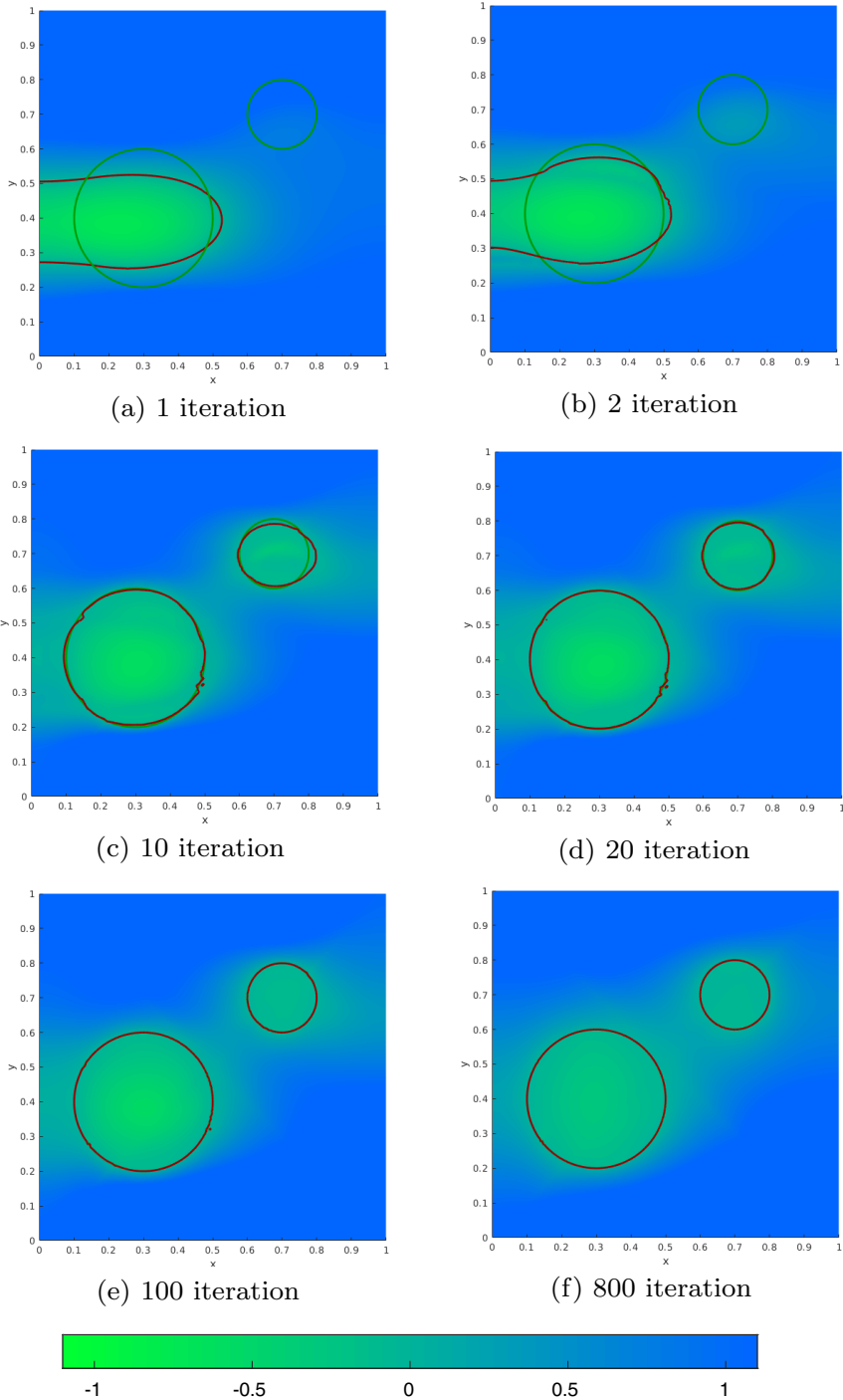
Fig. 11 Evolution of the level-set function for the 33025 nodes mesh
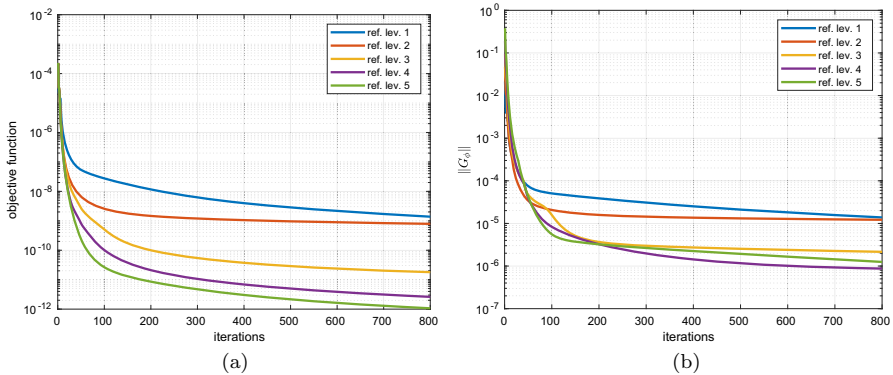
Fig. 12 Evolution of the objective function (**a**) and the norm of the topological-shape derivative (**b**) in course of the optimization

6.2.3 Numerical Results

As an initial design for the optimisation, we take the empty set, $\Omega = \emptyset$. This is realized by choosing $\phi_0 = 1/\|1\|_{L_2(D)}$ as the initial level set function. We use the algorithm outlined in Sect. 6.2.2 to update this level set function. We terminated the algorithm after the fixed number of 800 iterations. The final as well as some intermediate configurations are illustrated in Figs. 7, 8, 9, 10, and 11 for the five different levels of discretization.

We observe that in all cases the two circles are recovered with high accuracy. In Fig. 12 the evolution of the objective function as well as of the norm of the generalized numerical topological-shape derivative is plotted. We observe that objective function decreases fast and after 800 iterations a reduction by a factor of approximately $10^{-5} - 10^{-8}$ could be achieved. Moreover, we observe that the norm of the topological-shape derivative decreases continuously, more and more approaching the optimality condition (102).

7 Conclusions

In this work we presented a new sensitivity concept, called the topological-shape derivative which is based on a level set representation of a domain. This approach allows for a unified sensitivity analysis for shape and topological perturbations, which we carried out for a discretized PDE-constrained design optimization problem in two space dimensions. For the discretization we used a standard first order finite element method which does not account for the interface position in the approximation space. Therefore, kinks in the solution of the state and adjoint equations at material interfaces are not resolved. Comparing the computed sensitivities of the discretized problem with the discretization of the continuous topological and shape derivatives, we saw that certain terms did not appear in the former approach. These lack of these terms can be traced back to the inability of the chosen discretization method to represent

such kinks. Thus, it would be interesting to study discretization methods which do account for these kinks, *e.g.* XFEM or CutFEM, and perform the sensitivity analysis in these settings in future work. Furthermore, the extension to higher space dimensions, higher polynomial degree and other PDE constraints such as elasticity would be further interesting research directions.

Declarations

Conflict of interest   The authors declare that they have no conflict of interest.

A Appendix

A.1 Proof of Lemma 1

Proof   We investigate the limit

$$\lim_{t \searrow 0} \frac{1}{t} |\Omega_t \triangle \Omega| = \lim_{t \searrow 0} \frac{1}{t} \left(\int_{\Omega_t \setminus \Omega} dx + \int_{\Omega \setminus \Omega_t} dx \right).$$

Let $c : [0, 1] \to \partial\Omega$ be a smooth parametrization of the boundary $\partial\Omega$ in counterclockwise direction with smooth inverse and define $c_t : [0, 1] \to \partial\Omega_t$,

$$c_t(s) := c(s) + t V(c(s)) = (\mathrm{id} + t V)(c(s)). \tag{109}$$

The derivative is given by

$$\dot{c}_t(s) = \frac{d}{ds} c_t(s) = \dot{c}(s) + t \, \partial V(c(s)) \dot{c}(s) = (I + t \, \partial V(c(s))) \dot{c}(s). \tag{110}$$

Let $\bar{t} > 0$ sufficiently small such that, for all $t \in (0, \bar{t})$, the number of intersection points between $\partial\Omega$ and $\partial\Omega_t$ is a fixed number N. To each intersection point we associate a pair of numbers $(s_i(t), \hat{s}_i(t))$ such that

$$c(s_i(t)) = c_t(\hat{s}_i(t)) = c(\hat{s}_i(t)) + t V(c(\hat{s}_i(t))), \tag{111}$$

see Fig. 13 for an illustration of the situation. The symmetric difference can now be written as $\Omega_t \triangle \Omega = \bigcup_{i=1}^{N} A_i(t)$, where $A_i(t)$ denotes the region between $\partial\Omega$ and $\partial\Omega_t$ bounded by the intersection points $c(s_i(t))$ and $c(s_{i+1}(t))$ (here, $A_N(t)$ is bounded by $c(s_N(t))$ and $c(s_1(t))$). More precisely, $A_i(t)$ is bounded by the two segments $\{c(s) : s \in [s_i(t), s_{i+1}(t)]\}$ and $\{c_t(s) : s \in [\hat{s}_i(t), \hat{s}_{i+1}(t)]\}$. Now let $i \in \{1, \ldots, N\}$ fixed. The volume of $A_i(t)$ can be expressed by the divergence theorem as

$$|A_i(t)| = \int_{A_i(t)} \frac{1}{2} \mathrm{div} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mathrm{d}x = \frac{1}{2} \int_{\partial A_i(t)} x \cdot n(x) \, \mathrm{d}S_x \qquad (112)$$

$$= \frac{1}{2} \left(\int_{s_i(t)}^{s_{i+1}(t)} c(s)^\top R_i \frac{\dot{c}(s)}{|\dot{c}(s)|} |\dot{c}(s)| \, \mathrm{d}s + \int_{\hat{s}_i(t)}^{\hat{s}_{i+1}(t)} c_t(s)^\top (-R_i) \frac{\dot{c}_t(s)}{|\dot{c}_t(s)|} |\dot{c}_t(s)| \, \mathrm{d}s \right) \qquad (113)$$

with the rotation matrix

$$R_i = d_i \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad \text{where } d_i = \begin{cases} 1, & \text{if } A_i(t) \subset \Omega \setminus \Omega_t, \\ -1, & \text{if } A_i(t) \subset \Omega_t \setminus \Omega, \end{cases} \qquad (114)$$

such that $R_i \frac{\dot{c}(s)}{|\dot{c}(s)|}$ and $(-R_i) \frac{\dot{c}_t(s)}{|\dot{c}_t(s)|}$ are the unit normal vectors pointing out of $A_i(t)$ at $c(s)$ and $c_t(s)$, respectively. Note that $R_i^\top = -R_i$. For further use we note that
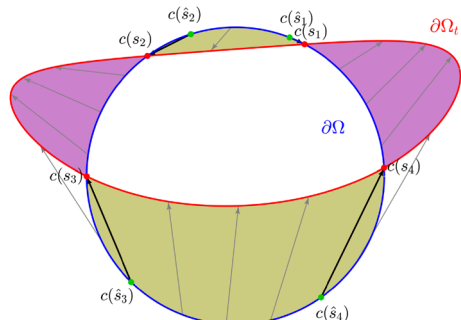
$$c_t(s)^\top R_i \, \dot{c}_t(s) = [c(s) + t V(c(s))]^\top R_i \, (I + t \, \partial V(c(s))) \, \dot{c}(s)$$

$$= c(s)^\top R_i \, \dot{c}(s) + t \left(V(c(s))^\top R_i \, \dot{c}(s) \right.$$

$$\left. + c(s)^\top R_i \cdot \partial V(c(s)) \dot{c}(s) \right) + t^2 V(s)^\top R_i \, \partial V(c(s)) \, \dot{c}(s).$$

Furthermore, let

$$\bar{s}_i := \lim_{t \searrow 0} s_i(t), \qquad (115)$$

and we observe from (111) that $\lim_{t \searrow 0} \hat{s}_i(t) = \bar{s}_i$.

Fig. 13 Illustration of the pairs $(s_i(t), \hat{s}_i(t))$

Moreover, since we assumed the inverse of c to be smooth (in particular Lipschitz continuous with constant L), we have that the limit

$$\lim_{t \searrow 0} \frac{1}{t} |s_i(t) - \hat{s}_i(t)| \le \lim_{t \searrow 0} \frac{L}{t} |c(s_i(t)) - c(\hat{s}_i(t))| = L|V(c(\bar{s}_i))|, \tag{116}$$

exists. Here we used (111) and the continuity of V and c.

With the abbreviation $f_c(s) := c(s)^\top R_i \dot{c}(s)$ we have

$$|A_i(t)|$$
$$= \frac{1}{2} \Bigg(\int_{\hat{s}_i(t)}^{s_i(t)} f_c(s)\,ds + \int_{s_i(t)}^{s_{i+1}(t)} f_c(s)\,ds + \int_{s_{i+1}(t)}^{\hat{s}_{i+1}(t)} f_c(s)\,ds - \int_{\hat{s}_i(t)}^{\hat{s}_{i+1}(t)} f_{c_t}(s)\,ds$$
$$\qquad - \int_{\hat{s}_i(t)}^{s_i(t)} f_c(s)\,ds - \int_{s_{i+1}(t)}^{\hat{s}_{i+1}(t)} f_c(s)\,ds \Bigg)$$
$$= \frac{1}{2} \Bigg(\int_{\hat{s}_i(t)}^{\hat{s}_{i+1}(t)} \big[f_c(s) - f_{c_t}(s) \big]\,ds - \int_{\hat{s}_i(t)}^{s_i(t)} f_c(s)\,ds - \int_{s_{i+1}(t)}^{\hat{s}_{i+1}(t)} f_c(s)\,ds \Bigg). \tag{117}$$

Thus,
$$\frac{1}{t}|A_i(t)| = \frac{1}{2}\big(B_i(t) - C_i(t) + C_{i+1}(t)\big)$$
with

$$B_i(t) = - \int_{\hat{s}_i(t)}^{\hat{s}_{i+1}(t)} \big[V(c(s))^\top R_i \dot{c}(s) + c(s)^\top R_i \, \partial V(c(s)) \dot{c}(s) \big]\,ds + \mathcal{O}(t), \tag{118}$$

$$C_i(t) = \frac{1}{t} \int_{\hat{s}_i(t)}^{s_i(t)} f_c(s)\,ds. \tag{119}$$

In order to compute $B_i(t)$ we note that

$$\frac{d}{ds}(V(c(s))^\top R_i \, c(s)) = V(c(s))^\top R_i \dot{c}(s) + (\partial V(c(s))\dot{c}(s))^\top R_i c(s)$$
$$= V(c(s))^\top R_i \dot{c}(s) - c(s)^\top R_i \, \partial V(c(s)) \dot{c}(s)$$

Thus,

$$B_i(t) = \int_{\hat{s}_i(t)}^{\hat{s}_{i+1}(t)} \left[\frac{d}{ds}(V(c(s))^\top R_i \, c(s)) - 2V(c(s))^\top R_i \dot{c}(s) \right] ds + \mathcal{O}(t)$$
$$= V(c(\hat{s}_{i+1}(t)))^\top R_i \, c(\hat{s}_{i+1}(t)) - V(c(\hat{s}_i(t)))^\top R_i \, c(\hat{s}_i(t))$$
$$\qquad - 2 \int_{\hat{s}_i(t)}^{\hat{s}_{i+1}(t)} V(c(s))^\top R_i \dot{c}(s)\,ds + \mathcal{O}(t). \tag{120}$$

For $C_i(t)$, by the mean value theorem there exists $\tilde{s}_i \in [\hat{s}_i(t), s_i(t)]$ such that

$$C_i(t) = \frac{1}{t} \int_{\hat{s}_i(t)}^{s_i(t)} c(s)^\top R_i \dot{c}(s)\, \mathrm{d}s = \dot{c}(\tilde{s}_i)^\top R_i \dot{c}(\tilde{s}_i) \frac{1}{t} |s_i(t) - \hat{s}_i(t)|. \tag{121}$$

On the other hand, we note that with (111) the following vector identity holds

$$tV(c(\hat{s}_i(t))) = c(s_i(t)) - c(\hat{s}_i(t)) = \int_{\hat{s}_i(t)}^{s_i(t)} \dot{c}(s)\, ds$$

$$= (s_i(t) - \hat{s}_i(t)) \int_0^1 \dot{c}(\hat{s}_i(t) + a(s_i(t) - \hat{s}_i(t)))\, da$$

and thus,

$$\dot{c}(\tilde{s}_i) \lim_{t \searrow 0} \frac{1}{t}(s_i(t) - \hat{s}_i(t))$$

$$= \lim_{t \searrow 0} \left(\frac{1}{t}(s_i(t) - \hat{s}_i(t)) \int_0^1 \dot{c}(\hat{s}_i(t) + a(s_i(t) - \hat{s}_i(t)))\, da \right) = V(c(\tilde{s}_i)). \tag{122}$$

Combining (120) and (121) we get

$$\lim_{t \searrow 0} \frac{1}{t}|A_i(t)| = \lim_{t \searrow 0} \frac{1}{2}(B_i(t) - C_i(t) + C_{i+1}(t))$$

$$= \lim_{t \searrow 0} \frac{1}{2} \left(-2 \int_{\hat{s}_i(t)}^{\hat{s}_{i+1}(t)} V(c(s))^\top R_i\, \dot{c}(s)\, ds + \delta_i(t) - \delta_{i+1}(t) \right)$$

with

$$\delta_i(t) := c(\hat{s}_i(t))^\top R_i V(c(\hat{s}_i(t))) - c(\tilde{s}_i)^\top R_i \dot{c}(\tilde{s}_i) \frac{1}{t}|s_i(t) - \hat{s}_i(t)|.$$

From (122), it follows that $\lim_{t \searrow 0} \delta_i(t) = 0 = \lim_{t \searrow 0} \delta_{i+1}(t)$ and thus

$$\lim_{t \searrow 0} \frac{1}{t}|A_i(t)| = -\int_{\tilde{s}_i}^{\tilde{s}_{i+1}} V(c(s))^\top R_i\, \dot{c}(s)\, ds = -\int_{c(\tilde{s}_i)}^{c(\tilde{s}_{i+1})} V(x) \cdot n_i(x)\, \mathrm{d}S_x$$

Here we used

$$n_i(x) = \begin{cases} n(x) & \text{if } V(x) \cdot n(x) < 0, \\ -n(x) & \text{if } V(x) \cdot n(x) > 0, \end{cases}$$

where n denotes the unit normal vector pointing out of Ω. Thus we have found

$$\lim_{t \searrow 0} \frac{1}{t} |\Omega_t \triangle \Omega| = \sum_{i=1}^{N} \lim_{t \searrow 0} \frac{1}{t} |A_i(t)| = \int_{\partial\Omega} |V(x) \cdot n(x)| \, \mathrm{d}S_x.$$

$\square$

A.2 Matrix Entries for the Numerical Shape Derivative

In order to simplify the notation we use in this section the abbreviation $\phi_i := \phi_{l_i}$ for $i = 1, 2, 3$. The matrix entries for $d_k \mathbf{m}_l^I = |\det J_l| \, d_k \bar{\mathbf{m}}_l^I$ and $d_k \mathbf{f}_l^I = |\det J_l| \, d_k \bar{\mathbf{f}}_l^I$ in (66) are given by

$$d_k \bar{m}_l^{A^\pm}[1, 1]$$
$$= \pm \frac{\phi_1^4 \left(\phi_2^3 + \phi_2^2 \phi_3 + \phi_2 \phi_3^2 + \phi_3^3\right) - 4\phi_1 \phi_2^3 \phi_3^3 + 6\phi_1^2 \phi_2^2 \phi_3^2 (\phi_2 + \phi_3) - 4\phi_1^3 \phi_2 \phi_3 \left(\phi_2^2 + \phi_2 \phi_3 + \phi_3^2\right)}{4(\phi_1 - \phi_2)^4 (\phi_1 - \phi_3)^4},$$

$$d_k \bar{m}_l^{A^\pm}[1, 2] = d_k \bar{m}_l^{A^\pm}[2, 1]$$
$$= \mp \frac{\phi_1^2 \left(3\phi_1^2 \phi_2^2 + 2\phi_1^2 \phi_2 \phi_3 + \phi_1^2 \phi_3^2 - 8\phi_1 \phi_2^2 \phi_3 - 4\phi_1 \phi_2 \phi_3^2 + 6\phi_2^2 \phi_3^2\right)}{12(\phi_1 - \phi_2)^4 (\phi_1 - \phi_3)^3},$$

$$d_k \bar{m}_l^{A^\pm}[1, 3] = d_k \bar{m}_l^{A^\pm}[3, 1]$$
$$= \mp \frac{\phi_1^2 \left(\phi_1^2 \phi_2^2 + 2\phi_1^2 \phi_2 \phi_3 + 3\phi_1^2 \phi_3^2 - 4\phi_1 \phi_2^2 \phi_3 - 8\phi_1 \phi_2 \phi_3^2 + 6\phi_2^2 \phi_3^2\right)}{12(\phi_1 - \phi_2)^3 (\phi_1 - \phi_3)^4},$$

$$d_k \bar{m}_l^{A^\pm}[2, 2] = \pm \frac{\phi_1^3 \left(3\phi_1 \phi_2 + \phi_1 \phi_3 - 4\phi_2 \phi_3\right)}{12(\phi_1 - \phi_2)^4 (\phi_1 - \phi_3)^2},$$

$$d_k \bar{m}_l^{A^\pm}[2, 3] = d_k \bar{m}_l^{A^\pm}[3, 2] = \pm \frac{\phi_1^3 \left(\phi_1 \phi_2 + \phi_1 \phi_3 - 2\phi_2 \phi_3\right)}{12(\phi_1 - \phi_2)^3 (\phi_1 - \phi_3)^3},$$

$$d_k \bar{m}_l^{A^\pm}[3, 3] = \pm \frac{\phi_1^3 \left(\phi_1 \phi_2 + 3\phi_1 \phi_3 - 4\phi_2 \phi_3\right)}{12(\phi_1 - \phi_2)^2 (\phi_1 - \phi_3)^4},$$

$$d_k \bar{m}_l^{B^\pm}[1, 1] = \mp \frac{\phi_2^4}{4(\phi_1 - \phi_2)^4 (\phi_2 - \phi_3)},$$

$$d_k \bar{m}_l^{B^\pm}[1, 2] = d_k \bar{m}_l^{B^\pm}[2, 1] = \pm \frac{\phi_2^3 \left(3\phi_1 \phi_2 - 4\phi_1 \phi_3 + \phi_2 \phi_3\right)}{12(\phi_1 - \phi_2)^4 (\phi_2 - \phi_3)^2},$$

$$d_k \bar{m}_l^{B^\pm}[1, 3] = d_k \bar{m}_l^{B^\pm}[3, 1] = \pm \frac{\phi_2^4}{12(\phi_1 - \phi_2)^3 (\phi_2 - \phi_3)^2},$$

$$d_k \bar{m}_l^{B^\pm}[2, 2] = \mp \frac{\phi_2^2 \left(3\phi_1^2 \phi_2^2 - 8\phi_1^2 \phi_2 \phi_3 + 6\phi_1^2 \phi_3^2 + 2\phi_1 \phi_2^2 \phi_3 - 4\phi_1 \phi_2 \phi_3^2 + \phi_2^2 \phi_3^2\right)}{12(\phi_1 - \phi_2)^4 (\phi_2 - \phi_3)^3},$$

$$d_k \bar{m}_l^{B\pm}[2,3] = d_k \bar{m}_l^{B\pm}[3,2] = \mp \frac{\phi_2^3 \left(\phi_1\phi_2 - 2\phi_1\phi_3 + \phi_2\phi_3\right)}{12(\phi_1 - \phi_2)^3(\phi_2 - \phi_3)^3},$$

$$d_k \bar{m}_l^{B\pm}[3,3] = \mp \frac{\phi_2^4}{12(\phi_1 - \phi_2)^2(\phi_2 - \phi_3)^3},$$

$$d_k \bar{f}_l^{A\pm}[1] = \mp \frac{\phi_1 \left(\phi_1^2\phi_2^2 + \phi_1^2\phi_2\phi_3 + \phi_1^2\phi_3^2 - 3\phi_1\phi_2^2\phi_3 - 3\phi_1\phi_2\phi_3^2 + 3\phi_2^2\phi_3^2\right)}{3(\phi_1 - \phi_2)^3(\phi_1 - \phi_3)^3},$$

$$d_k \bar{f}_l^{A\pm}[2] = \pm \frac{\phi_1^2 \left(2\phi_1\phi_2 + \phi_1\phi_3 - 3\phi_2\phi_3\right)}{6(\phi_1 - \phi_2)^3(\phi_1 - \phi_3)^2},$$

$$d_k \bar{f}_l^{A\pm}[3] = \pm \frac{\phi_1^2 \left(\phi_1\phi_2 + 2\phi_1\phi_3 - 3\phi_2\phi_3\right)}{6(\phi_1 - \phi_2)^2(\phi_1 - \phi_3)^3},$$

$$d_k \bar{f}_l^{B\pm}[1] = \pm \frac{\phi_2^3}{3(\phi_1 - \phi_2)^3(\phi_2 - \phi_3)},$$

$$d_k \bar{f}_l^{B\pm}[2] = \mp \frac{\phi_2^2 \left(2\phi_1\phi_2 - 3\phi_1\phi_3 + \phi_2\phi_3\right)}{6(\phi_1 - \phi_2)^3(\phi_2 - \phi_3)^2},$$

$$d_k \bar{f}_l^{B\pm}[3] = \mp \frac{\phi_2^3}{6(\phi_1 - \phi_2)^2(\phi_2 - \phi_3)^2}.$$

The formulas for $d_k \bar{m}_l^{C\pm}$ and $d_k \bar{f}_l^{C\pm}$ can be obtained from the formulas for $d_k \bar{m}_l^{B\pm}$ and $d_k \bar{f}_l^{B\pm}$ by exchanging the matrix entry indices 2 and 3, and ϕ_2 and ϕ_3, e.g. $d_k \bar{f}_l^{C\pm}[3]$ is obtained from $d_k \bar{f}_l^{B\pm}[2]$ by exchanging ϕ_2 and ϕ_3.

References

1. Allaire, G., Jouve, F.: Coupling the level set method and the topological gradient in structural optimization. In: Bendsøe, M.P., Olhoff, N., Sigmund, O. (eds.) IUTAM Symposium on Topological Design Optimization of Structures, Machines and Materials, pp. 3–12. Springer, Dordrecht (2006)
2. Allaire, G., Jouve, F., Toader, A.M.: Structural optimization using sensitivity analysis and a level-set method. J. Comput. Phys. **194**(1), 363–393 (2004)
3. Amstutz, S.: Sensitivity analysis with respect to a local perturbation of the material property. Asymptot. Anal. **49**(1,2), 87–108 (2006)
4. Amstutz, S., Andrä, H.: A new algorithm for topology optimization using a level-set method. J. Comput. Phys. **216**(2), 573–588 (2006)
5. Amstutz, S., Gangl, P.: Topological derivative for the nonlinear magnetostatic problem. Electron. Trans. Numer. Anal. **51**, 169–218 (2019)
6. Amstutz, S., Dapogny, C., Ferrer, À.: A consistent relaxation of optimal design problems for coupling shape and topological derivatives. Numer. Math. **140**(1), 35–94 (2018)
7. Bendsoe, M., Sigmund, O.: Topology Optimization: Theory, Methods, and Applications. Springer, Berlin (2003)
8. Berggren, M.: Shape calculus for fitted and unfitted discretizations: domain transformations vs. boundary-face dilations. (2022). https://doi.org/10.48550/ARXIV.2210.10411
9. Bernland, A., Wadbro, E., Berggren, M.: Acoustic shape optimization using cut finite elements. Int. J. Numer. Methods Eng. **113**(3), 432–449 (2018)
10. Burger, M., Hackl, B., Ring, W.: Incorporating topological derivatives into level set methods. J. Comput. Phys. **194**(1), 344–362 (2004)
11. Burman, E., Claus, S., Hansbo, P., Larson, M.G., Massing, A.: Cutfem: discretizing geometry and partial differential equations. Int. J. Numer. Meth. Eng. **104**(7), 472–501 (2015)

12. Dapogny, C., Feppon, F.: Shape optimization using a level set based mesh evolution method: an overview and tutorial (2022). https://hal.science/hal-03881641
13. Delfour, M.C.: Topological derivative: a semidifferential via the Minkowski content. J. Convex Anal. **25**(3), 957–982 (2018)
14. Delfour, M.C.: Topological derivatives via one-sided derivative of parametrized minima and minimax. Eng. Comput. **39**(1), 34–59 (2022)
15. Delfour, M.C., Zolésio, J.P.: Shapes and geometries: metrics, analysis, differential calculus, and optimization. In: Advances in Design and Control, vol. 22, 2nd edn. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (2011)
16. Fike, J., Alonso, J.: The development of hyper-dual numbers for exact second-derivative calculations. In: 49th AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition, p. 886 (2011)
17. Gangl, P.: A multi-material topology optimization algorithm based on the topological derivative. Comput. Methods Appl. Mech. Eng. **366**, 113090 (2020)
18. Gangl, P., Sturm, K.: A simplified derivation technique of topological derivatives for quasi-linear transmission problems. ESAIM **26**, 106 (2020)
19. Gangl, P., Langer, U., Laurain, A., Meftahi, H., Sturm, K.: Shape optimization of an electric motor subject to nonlinear magnetostatics. SIAM J. Sci. Comput. **37**(6), B1002–B1025 (2015)
20. Hägg, L., Wadbro, E.: On minimum length scale control in density based topology optimization. Struct. Multidiscip. Optim. **58**(3), 1015–1032 (2018)
21. Haubner, J., Siebenborn, M., Ulbrich, M.: A continuous perspective on shape optimization via domain transformations. SIAM J. Sci. Comput. **43**(3), A1997–A2018 (2021)
22. Laurain, A.: Analyzing smooth and singular domain perturbations in level set methods. SIAM J. Math. Anal. **50**(4), 4327–4370 (2018)
23. Laurain, A., Sturm, K.: Distributed shape derivative via averaged adjoint method and applications. ESAIM **50**(4), 1241–1267 (2016)
24. Martins, J.R., Sturdza, P., Alonso, J.J.: The complex-step derivative approximation. ACM Trans. Math. Softw. **29**(3), 245–262 (2003)
25. Moës, N., Dolbow, J., Belytschko, T.: A finite element method for crack growth without remeshing. Int. J. Numer. Meth. Eng. **46**(1), 131–150 (1999)
26. Novotny, A., Sokolowski, J.: Topological Derivatives in Shape Optimization. Interaction of Mechanics and Mathematics. Springer, Heidelberg (2013)
27. Novotny, A., Feijóo, R., Taroco, E., Padra, C.: Topological sensitivity analysis. Comput. Methods Appl. Mech. Eng. **192**(7), 803–829 (2003)
28. Novotny, A., Feijóo, R., Taroco, E., Padra, C.: Topological-shape sensitivity method: theory and applications. In: Bendsøe, M.P., Olhoff, N., Sigmund, O. (eds.) IUTAM Symposium on Topological Design Optimization of Structures, Machines and Materials, pp. 469–478. Springer, Dordrecht (2006)
29. Sigmund, O., Maute, K.: Topology optimization approaches. Struct. Multidiscip. Optim. **48**(6), 1031–1055 (2013)
30. Sokolowski, J., Zochowski, A.: On the topological derivative in shape optimization. SIAM J. Control. Optim. **37**(4), 1251–1272 (1999)
31. Sturm, K.: Minimax Lagrangian approach to the differentiability of nonlinear PDE constrained shape functions without saddle point assumption. SIAM J. Control. Optim. **53**(4), 2017–2039 (2015)