SIAM J. CONTROL OPTIM.
Vol. 55, No. 3, pp. 1481–1499

AN EXTENSION OF THE PROJECTED GRADIENT METHOD TO A BANACH SPACE SETTING WITH APPLICATION IN STRUCTURAL TOPOLOGY OPTIMIZATION*

LUISE BLANK† AND CHRISTOPH RUPPRECHT†

Abstract. For the minimization of a nonlinear cost functional under convex constraints the relaxed projected gradient process is a well known method. The analysis is classically performed in a Hilbert space. We generalize this method to functionals which are differentiable in a Banach space. The search direction is calculated by a quadratic approximation of the cost functional using the idea of the projected gradient. Thus it is possible to perform, e.g., an L^2 gradient method if the cost functional is only differentiable in L^∞. We show global convergence using Armijo backtracking for the step length selection and allow the underlying inner product and the scaling of the derivative to change in every iteration. As an application we present a structural topology optimization problem based on a phase field model, where the reduced cost functional is differentiable in $H^1 \cap L^\infty$. The presented numerical results using the H^1 inner product and a pointwise chosen metric including second order information show the expected mesh independency in the iteration numbers. The latter yields an additional, drastic decrease in iteration numbers as well as in computation time. Moreover we present numerical results using a BFGS update of the H^1 inner product for further optimization problems based on phase field models.

Key words. projected gradient method, variable metric method, convex constraints, shape and topology optimization, phase field approach

AMS subject classifications. 49M05, 49M15, 65K10, 65K15, 74P05, 90C30, 90C48

DOI. 10.1137/16M1092301

1. Introduction.

For an optimization problem with nonlinear cost functional and convex constraints

$$\min j(\varphi) \quad \text{subject to } \varphi \in \Phi_{ad} \tag{1.1}$$

the projected gradient method is well known. The analysis of the finite dimensional case can be found, for example, in [3]. Otherwise a Hilbert space setting is required, i.e., the problem is posed in some Hilbert space H with inner product $(.,.)_H$ and norm $\|.\|_H$. The nonempty, convex feasible set Φ_{ad} has to be closed with respect to $\|.\|_H$ and the cost functional j has to be Fréchet differentiable with respect to this norm. Here we recall that the H-gradient $\nabla_H j$ is characterized by the equality $(\nabla_H j(\varphi), \eta)_H = \langle j'(\varphi), \eta \rangle_{H^*,H}$ for all $\eta \in H$, where j' denotes the Fréchet derivative of j. The classical projected gradient method then moves a current iterate in the direction of the negative gradient and orthogonally projects the result back on the feasible set:

$$\varphi_{k+1} = P_H(\varphi_k - \lambda_k \nabla_H j(\varphi_k)). \tag{1.2}$$

Here, P_H denotes the orthogonal projection onto Φ_{ad}. Both the gradient and the projection have to be taken with respect to the underlying Hilbert space inner product. To obtain global convergence the step length λ_k in the direction of the negative

*Received by the editors September 2, 2016; accepted for publication (in revised form) February 21, 2017; published electronically May 9, 2017.
http://www.siam.org/journals/sicon/55-3/M109230.html
†Department of Mathematics, Universität Regensburg, Universitätsstr. 31, 93053 Regensburg, Germany (luise.blank@ur.de, christoph.rupprecht@ur.de).

gradient has to be chosen according to some step length rule. This method is often called the gradient path method and is widely used. Details of the method and its analysis and applications can be found, e.g., in [13, 14, 19, 20, 21, 23, 24, 25, 26, 31]. In [27] this method is extended to convex subsets of a reflexive, smooth, and rotund Banach space employing the metric projection and the exact step length.

Approach (1.2) requires solving a projection subproblem in each line search iteration. If the feasible set is given by box constraints and the L^2 inner product is used, the projection is in general cheap. However, if calculating the projection is expensive compared to the evaluation of the cost functional each iteration is expensive. Hence, another approach is possibly cheaper, where the projection and the step length calculation is interchanged. Then one performs a line search along the descent direction given by the projected, possibly scaled negative gradient $v_k = P_H(\varphi_k - \lambda_k \nabla_H j(\varphi_k)) - \varphi_k$, i.e.,

$$\varphi_{k+1} = \varphi_k + \alpha_k (P_H(\varphi_k - \lambda_k \nabla_H j(\varphi_k)) - \varphi_k), \tag{1.3}$$

where the line search is done with respect to α whereas λ_k is fixed. This is suggested in finite dimension, e.g., in [3, 29] or in more general Hilbert spaces in [11, 12, 21].

We generalize the method (1.3) to functionals which are not differentiable in a Hilbert space, but in a possibly nonreflexive Banach space, where consequently no gradient and no orthogonal projection exist. As mentioned, in [27] $P_H(\varphi - \lambda \nabla_H j(\varphi))$ is extended to Banach spaces using the duality mapping J to define $\nabla j(\varphi) = J^{-1}(j'(\varphi))$, where J is invertible since the considered Banach spaces are reflexive, smooth, and strictly convex. However, our result allows to perform, e.g., an L^2 gradient method if the cost functional is only differentiable in L^∞, which is often the case for semilinear optimal control problems [31]. Note that L^∞ does not fulfill the assumptions on the space demanded in [27]. Moreover, in contrast to [27] we utilize a variable metric. This extends the ideas given in finite dimension by [3] regarding the scaled gradient projection methods and the constrained Newton's method. The employed inner product, which is used to determine the search direction, may change in each iteration. Hence it is possible to include second order information in the method, which typically leads to a decrease in the number of iterations. For Newton and quasi-Newton based search directions even a superlinear rate of convergence is expected as it is known under certain conditions [3, 12, 21, 24]. However, for these methods only local convergence results are available, while the focus in this paper is on global convergence. The resulting generalization we call the "variable metric projection type" (VMPT) method.

The paper is organized as follows: In section 2 we study the VMPT method. In section 2.1 a precise description of the method is given, where the search direction is calculated by a quadratic approximation of the cost functional using the idea of the projected gradient with varying inner products, and Armijo backtracking is applied to determine the step length α_k. In section 2.2 the global convergence result together with the necessary assumptions, among others on the underlying Banach space, are stated. In the last subsection we first prove that the search directions are well defined and that they are descent directions as well as gradient related in the sense of Bertsekas [3]. (For a precise definition we refer to Lemma 2.10.) Then the proof of the convergence result is given.

In section 3 we study the applicability of the method to a structural topology optimization problem, namely, the mean compliance minimization in linear elasticity based on a phase field model. The reduced cost functional is differentiable only in $H^1 \cap L^\infty$.

In the last section numerical results are given. For the appropriately discretized mean compliance problem we see as expected that the VMPT method choosing the H^1 metric leads to mesh independent iteration numbers in contrast to choosing the L^2 metric. We also present the choice of a variable metric using second order information and the choice of a BFGS update of the H^1 metric. This reduces the iteration numbers to less than a hundredth. Moreover, we give additional numerical examples for the successful application of the VMPT method. These include a problem of compliant mechanism, drag minimization of the Stokes flow, and an inverse problem.

2. Variable metric projection type method.

2.1. Generalization of the projected gradient method. The orthogonal projection $P_H(\varphi_k - \lambda_k \nabla_H j(\varphi_k))$ employed in (1.2) is the unique solution of

$$\min_{y \in \Phi_{ad}} \frac{1}{2} \|(\varphi_k - \lambda_k \nabla_H j(\varphi_k)) - y\|_H^2,$$

which is equivalent to the problem

$$\min_{y \in \Phi_{ad}} \frac{1}{2} \|y - \varphi_k\|_H^2 + \lambda_k Dj(\varphi_k, y - \varphi_k), \tag{2.1}$$

in the sense that the minimizer is the same. This is due to the identity $(\nabla_H j(\varphi_k), y - \varphi_k)_H = j'(\varphi_k)(y - \varphi_k) = Dj(\varphi_k, y - \varphi_k)$, where the last denotes the directional derivative of j at φ_k in direction $y - \varphi_k$. In the formulation (2.1) the existence of the gradient $\nabla_H j$ is not required. Even differentiability with respect to H can be omitted.

In the following we formulate an extension of the projected gradient method where $P_H(\varphi_k - \lambda_k \nabla_H j(\varphi_k))$ is replaced by the solution $\bar{\varphi}_k$ of (2.1).

First we drop the requirement of a gradient as mentioned above. We assume that the admissible set Φ_{ad} is a subset of an intersection of Banach spaces $\mathbb{X} \cap \mathbb{D}$, where $\mathbb{X}$ and $\mathbb{D}$ have certain properties (see (A1)), which are, e.g., fulfilled for $\mathbb{X} = H^1(\Omega)$ or $\mathbb{X} = L^2(\Omega)$ and $\mathbb{D} = L^\infty(\Omega)$. Furthermore assume that j is continuously Fréchet differentiable on Φ_{ad} with respect to the norm $\|.\|_{\mathbb{X}\cap\mathbb{D}} := \|.\|_{\mathbb{X}} + \|.\|_{\mathbb{D}}$. The Fréchet derivative of j at φ is denoted by $j'(\varphi) \in (\mathbb{X} \cap \mathbb{D})^*$, and we write $\langle ., . \rangle$ for the dual pairing in the space $\mathbb{X} \cap \mathbb{D}$. Moreover, we use C as a positive generic constant throughout the paper.

Second, we also allow the norm $\|.\|_H$ in (2.1) to change in every iteration. Therefore, we consider a sequence $\{a_k\}$ of symmetric positive definite bilinear forms inducing norms $\|.\|_{a_k}$ on $\mathbb{X} \cap \mathbb{D}$. As already mentioned, this approach falls into the class of variable metric methods and includes the choice of Newton and quasi-Newton based search directions. In finite dimension a_k is given by $a_k(p, v) := p^T B_k v$, where B_k can be the Hessian of j at φ_k, if it is positive definite, or an approximation of it.

Hence, in each step of the VMPT method the projection type subproblem

$$\min_{y \in \Phi_{ad}} \quad \frac{1}{2} \|y - \varphi_k\|_{a_k}^2 + \lambda_k \langle j'(\varphi_k), y - \varphi_k \rangle \tag{2.2}$$

with some scaling parameter $\lambda_k > 0$ has to be solved. Problem (2.2) is formally equivalent to the projection $P_{a_k}(\varphi_k - \lambda_k \nabla_{a_k} j(\varphi_k))$. However, j is not necessarily differentiable with respect to $\|.\|_{a_k}$, and $\mathbb{X} \cap \mathbb{D}$ endowed with $a_k(.,.)$ is only a pre-Hilbert space. Hence $\nabla_{a_k} j(\varphi_k)$ does not need to exist.

For globalization of the method we perform a line search based on the widely used Armijo back tracking, which results in Algorithm 2.1. In the next section it is shown that the algorithm is well defined under certain assumptions and in particular that a unique solution $\bar{\varphi}_k$ of (2.2) exists, together with the proof of convergence. We denote the solution of (2.2) also by $\mathcal{P}_k(\varphi_k)$ due to the connection to a projection.

ALGORITHM 2.1 (VMPT method).

Choose $0 < \beta < 1$, $0 < \sigma < 1$, *and* $\varphi_0 \in \Phi_{ad}$.
$k := 0$
while $k \leq k_{max}$ ***do***
  Choose λ_k *and* a_k.
  Calculate the minimizer $\overline{\varphi}_k = \mathcal{P}_k(\varphi_k)$ *of the subproblem* (2.2).
  Set the search direction $v_k := \overline{\varphi}_k - \varphi_k$
  if $\|v_k\|_{\mathbb{X}} \leq tol$ ***then***
    return
  end if
  Determine the step length $\alpha_k := \beta^{m_k}$ *with minimal* $m_k \in \mathbb{N}_0$ *such that*
$$j(\varphi_k + \alpha_k v_k) \leq j(\varphi_k) + \alpha_k \sigma \langle j'(\varphi_k), v_k\rangle.$$
  Update $\varphi_{k+1} := \varphi_k + \alpha_k v_k$
  $k := k + 1$
end while

The stopping criterion $\|v_k\|_{\mathbb{X}} \leq tol$ is motivated by the fact that φ_k is a stationary point of j if and only if $v_k = 0$ and that $v_k \to 0$ in $\mathbb{X}$; cf. Corollary 2.5 and Theorem 2.2. By stationary point we throughout refer to stationarity with respect to Φ_{ad}.

As already mentioned, this algorithm is not a curved search along the *gradient path* (1.2). However, to include the idea of the gradient path approach, we imbed the possibility to vary the scaling factor $\{\lambda_k\}$ for the formal gradient in (2.2) in each iteration. The parameter λ_k can be put into a_k by dividing the cost functional in (2.2) by λ_k. We nevertheless treat it as a separate parameter since this reflects the case where a_k is fixed for all iterations. Note that under the assumptions used in this paper a curved search along the gradient path is not possible since not even the existence of a positive step length can be guaranteed; cf. Remark 2.7.

2.2. Global convergence result. We perform the analysis of the method with respect to two norms in the spaces $\mathbb{X}$ and $\mathbb{D}$, which we assume to have the following properties:

(A1) $\mathbb{X}$ is a reflexive real Banach space. $\mathbb{D}$ is isometrically isomorphic to $\mathbb{B}^*$, where $\mathbb{B}$ is a separable real Banach space. Moreover, for any sequence $\{\varphi_i\}$ in $\mathbb{X} \cap \mathbb{D}$ with $\varphi_i \to \varphi$ weakly in $\mathbb{X}$ and $\varphi_i \to \tilde{\varphi}$ weakly-* in $\mathbb{D}$, it holds that $\varphi = \tilde{\varphi}$.

We identify $\mathbb{D}$ and $\mathbb{B}^*$ and say that a sequence converges weakly-* in $\mathbb{D}$ if it converges weakly-* in $\mathbb{B}^*$. The separability of $\mathbb{B}$ is used to get weak-* sequential compactness in $\mathbb{D}$. We would like to mention that the results hold also if $\mathbb{D}$ is a reflexive Banach space, in particular if $\mathbb{D}$ is an Hilbert space. In this case weak-* convergence has to be replaced by weak convergence throughout the paper. However, in the application we are interested in $\mathbb{D} = L^\infty(\Omega)$. In the case of the Sobolev space $\mathbb{X} = W^{k,p}(\Omega)$ and $\mathbb{D} = L^q(\Omega)$, where $\Omega \subseteq \mathbb{R}^d$ is a bounded domain, $k \geq 0$, $1 < p < \infty$, and $1 < q \leq \infty$, the above assumption is fulfilled.

In addition to the above conditions on $\mathbb{X}$ and $\mathbb{D}$ let the following assumptions hold for the problem (1.1):

(A2) $\Phi_{ad} \subseteq \mathbb{X} \cap \mathbb{D}$ is convex, closed in $\mathbb{X}$, and nonempty.
(A3) Φ_{ad} is bounded in $\mathbb{D}$.
(A4) j is bounded from below on Φ_{ad}.

(A5) j is continuously differentiable in a neighborhood of $\Phi_{ad} \subseteq \mathbb{X} \cap \mathbb{D}$.

(A6) For each $\varphi \in \Phi_{ad}$ and for each sequence $\{\varphi_i\} \subseteq \mathbb{X} \cap \mathbb{D}$ with $\varphi_i \to 0$ weakly in $\mathbb{X}$ and weakly-* in $\mathbb{D}$ it holds that $\langle j'(\varphi), \varphi_i \rangle \to 0$ as $i \to \infty$.

Assumptions (A2)–(A5) are standard. Moreover, if there exists $C > 0$ such that $\|p\|_{\mathbb{D}} \leq C\|p\|_{\mathbb{X}}$ holds for all $p \in \mathbb{X} \cap \mathbb{D}$, assumption (A3) can be omitted as it is the case in the classical Hilbert space setting (i.e., if $\mathbb{D} = \mathbb{X} = H$ for some Hilbert space H). The weak continuity assumption (A6) is fulfilled in many cases, e.g., if $\mathbb{X} = \mathbb{D}$ is a Hilbert space or if $\mathbb{D} = L^\infty$ and $j'(\varphi) \in L^1$ for all $\varphi \in \Phi_{ad}$ as it is the case for semilinear elliptic optimal control problems with box constraints under the (common) assumptions listed in [31]. Another more general sufficient condition for (A6) is $j'(\varphi) \in \mathbb{X}^* + \mathbb{B}$, where $\mathbb{B}$ is as in (A1). This is fulfilled, e.g., by the example studied in section 3.

Moreover, we request for the parameters a_k and λ_k of the algorithm that the following hold:

(A7) $\{a_k\}$ is a sequence of symmetric positive definite bilinear forms on $\mathbb{X} \cap \mathbb{D}$.

(A8) It exists $c_1 > 0$ such that $c_1\|p\|^2_{\mathbb{X}} \leq \|p\|^2_{a_k}$ for all $p \in \mathbb{X} \cap \mathbb{D}$ and $k \in \mathbb{N}_0$.

(A9) For all $k \in \mathbb{N}_0$ it exists $c_2(k)$ such that $\|p\|^2_{a_k} \leq c_2\|p\|^2_{\mathbb{X}\cap\mathbb{D}}$ for all $p \in \mathbb{X} \cap \mathbb{D}$.

(A10) For all $k \in \mathbb{N}_0$, $p \in \Phi_{ad}$ and for each sequence $\{y_i\} \subseteq \Phi_{ad}$, where there exists some $y \in \mathbb{X} \cap \mathbb{D}$ with $y_i \to y$ weakly in $\mathbb{X}$ and weakly-* in $\mathbb{D}$ it holds that $a_k(p, y_i) \to a_k(p, y)$ as $i \to \infty$.

(A11) For each subsequence $\{\varphi_{k_i}\}$ of the iterates given by Algorithm 2.1 converging in $\mathbb{X} \cap \mathbb{D}$, the corresponding subsequence $\{a_{k_i}\}$ has the property that $a_{k_i}(p_i, y_i) \to 0$ for any sequences $\{p_i\}, \{y_i\} \subseteq \mathbb{X} \cap \mathbb{D}$ with $p_i \to 0$ strongly in $\mathbb{X}$ and weakly-* in $\mathbb{D}$ and $\{y_i\}$ converging in $\mathbb{X} \cap \mathbb{D}$.

(A12) It holds that $0 < \lambda_{\min} \leq \lambda_k \leq \lambda_{\max}$ for all $k \in \mathbb{N}_0$.

(A1)–(A12) are assumed throughout this paper if not mentioned otherwise.

Assumption (A11) reflects the possibility of a point-based choice of a_k, e.g., dependent on the second order derivative $j''(\varphi_k)$ or on an approximation thereof.

We would like to mention that (A7)–(A9) are weaker assumptions than the typical requirements for the method in Hilbert space (i.e., $\mathbb{D} = \mathbb{X} = H$), which is the uniform norm equivalence (see [3, 14, 19, 29])

$$\exists C, c > 0 : \quad c\|p\|^2_H \leq \|p\|^2_{a_k} \leq C\|p\|^2_H \quad \forall p \in H,\ k \in \mathbb{N}_0. \tag{2.3}$$

Then also (A10)–(A11) are fulfilled. Also in the case that $j \in C^2(\mathbb{X} \cap \mathbb{D})$ and $a_k = j''(\varphi_k)$ fulfills (A8), the remaining assumptions of (A7)–(A11) hold. Furthermore, if $\mathbb{X}$ is a Hilbert space (and $\mathbb{D}$ a Banach space as in (A1)), the choice $a_k(u, v) = (u, v)_{\mathbb{X}}$ fulfills all assumptions (A7)–(A11).

An example of a_k which only fulfills the weaker assumptions is presented in (3.9) for our application in structural topology optimization.

The main result of the paper is the following, which is proved in section 2.3.

THEOREM 2.2. *Let $\{\varphi_k\} \subseteq \Phi_{ad}$ be the sequence generated by the VMPT method (Algorithm 2.1) with $tol = 0$ and let the assumptions* (A1)–(A12) *hold, then:*

1. $\lim_{k\to\infty} j(\varphi_k)$ *exists.*
2. *Every accumulation point of $\{\varphi_k\}$ in $\mathbb{X} \cap \mathbb{D}$ is a stationary point of j.*
3. *For all subsequences with $\varphi_{k_i} \to \varphi$ in $\mathbb{X} \cap \mathbb{D}$, where φ is stationary, the subsequence $\{v_{k_i}\}$ converges strongly in $\mathbb{X}$ to zero.*
4. *If additionally $j \in C^{1,\gamma}(\Phi_{ad})$ with respect to $\|.\|_{\mathbb{X}\cap\mathbb{D}}$ for some $0 < \gamma \leq 1$, then the whole sequence $\{v_k\}$ converges to zero in $\mathbb{X}$.*

The requirement of a strong accumulation point to obtain a stationary point is common; see, e.g., [23, 32] and references therein. In case of convex cost functionals this can be relaxed to weak accumulation points for the VMPT method [28]. This theorem reflects the available results where the problem is posed in some Hilbert space H and the constant metric $a_k(p,v) = (p,v)_H$ is chosen: global convergence is shown in [21] for convex j and a line search along the descent direction. In the case of a curved search along the gradient path convergence is shown in [13, 16]. Result 4 of Theorem 2.2 is shown in [23] in the case of a curved search along the gradient path under the same assumption $j \in C^{1,\gamma}$. Global convergence results for methods with general variable metric are to our knowledge only available in finite dimensions, which can be found, e.g., in [3].

2.3. Analysis and proof of the convergence result of the VMPT method. We first show the existence and uniqueness of $\overline{\varphi}_k = \mathcal{P}_k(\varphi_k)$ based on the direct method in the calculus of variations using the following Lemma and assumptions (A2), (A3), and (A5)–(A10). Note that the standard proof cannot be applied, since a_k is indeed $\mathbb{X}$-coercive, but a_k and $\langle j'(\varphi_k), \cdot\rangle$ are not $\mathbb{X}$-continuous. Another difficulty is that $\mathbb{X} \cap \mathbb{D}$ is not necessarily reflexive.

LEMMA 2.3. *Let $\{p_k\} \subseteq \Phi_{ad}$ with $p_k \to p$ weakly in $\mathbb{X}$ for some $p \in \Phi_{ad}$. Then $p_k \to p$ weakly-* in $\mathbb{D}$.*

Proof. Since Φ_{ad} is bounded in $\mathbb{D}$ and the closed unit ball of $\mathbb{D}$ is weakly-* sequentially compact due to the separability of $\mathbb{B}$, we can extract from any subsequence of $\{p_k\} \subseteq \Phi_{ad}$ another subsequence $\{p_{k_i}\}$ with $p_{k_i} \to \tilde{p}$ weakly-* in $\mathbb{D}$ for some $\tilde{p} \in \mathbb{D}$. Due to the required unique limit in $\mathbb{X}$ and $\mathbb{D}$ we have $\tilde{p} = p$. Since for any subsequence we find a subsequence converging to the same p, we have that the whole sequence converges to p. ∎

THEOREM 2.4. *For any $k \in \mathbb{N}_0$ and $\varphi \in \Phi_{ad}$, the problem*

$$\min_{y\in\Phi_{ad}} \quad \frac{1}{2}\|y-\varphi\|_{a_k}^2 + \lambda_k \langle j'(\varphi), y-\varphi\rangle \tag{2.4}$$

admits a unique solution $\bar{\varphi} := \mathcal{P}_k(\varphi)$, which is given by the unique solution of the variational inequality

$$a_k(\bar{\varphi}-\varphi, \eta-\bar{\varphi}) + \lambda_k \langle j'(\varphi), \eta-\bar{\varphi}\rangle \geq 0 \qquad \forall \eta \in \Phi_{ad}. \tag{2.5}$$

Proof. Let $k \in \mathbb{N}_0$ and $\varphi \in \Phi_{ad}$ arbitrary. Problem (2.4) is equivalent to

$$\min_{y\in\Phi_{ad}} \quad g_k(y) := \tfrac{1}{2}a_k(y,y) + \langle b_k, y\rangle\,, \tag{2.6}$$

where $\langle b_k, y\rangle := \lambda_k \langle j'(\varphi), y\rangle - a_k(\varphi, y)$ and $b_k \in (\mathbb{X}\cap\mathbb{D})^*$ due to (A5) and (A9). By (A3) and (A8) we get for any $y \in \Phi_{ad}$ with some generic $C > 0$

$$g_k(y) \geq \frac{c_1}{2}\|y\|_{\mathbb{X}}^2 - \|b_k\|_{(\mathbb{X}\cap\mathbb{D})^*}(\|y\|_{\mathbb{X}} + \underbrace{\|y\|_{\mathbb{D}}}_{\leq C}) \geq -C. \tag{2.7}$$

Thus g_k is $\mathbb{X}$-coercive and bounded from below on Φ_{ad}. Hence we can choose an infimizing sequence $\varphi_i \in \Phi_{ad}$, such that $g_k(\varphi_i) \xrightarrow{i\to\infty} \inf_{y\in\Phi_{ad}} g_k(y)$. From the estimate (2.7) we conclude that $\{\varphi_i\}$ is bounded in $\mathbb{X}$. Therefore, we can extract a subsequence (still denoted by φ_i) which converges weakly in $\mathbb{X}$ to some $\bar{\varphi} \in \mathbb{X}$. Since Φ_{ad} is convex and closed in $\mathbb{X}$, it is also weakly closed in $\mathbb{X}$ and thus $\bar\varphi \in \Phi_{ad}$. By Lemma 2.3 we also get $\varphi_i \to \bar\varphi$ weakly-* in $\mathbb{D}$. Finally we show $g_k(\bar\varphi) = \inf_{y\in\Phi_{ad}} g_k(y)$. Using (A6), (A8), and (A10) one can show that $\liminf_i a_k(\varphi_i,\varphi_i) \geq a_k(\bar\varphi,\bar\varphi)$ and $\lim_i \langle b_k,\varphi_i\rangle = \langle b_k,\bar\varphi\rangle$, and thus $\liminf_i g_k(\varphi_i) \geq g_k(\bar\varphi)$. We conclude

$$\inf_{y\in\Phi_{ad}} g_k(y) \leq g_k(\bar\varphi) \leq \liminf_i g_k(\varphi_i) = \inf_{y\in\Phi_{ad}} g_k(y),$$

which shows the existence of a minimizer of (2.6). Using (A8), the uniqueness follows from strict convexity of g_k.

Due to (A5) and (A9), we have that g_k is differentiable in $\mathbb{X}\cap\mathbb{D}$, where its directional derivative at $\bar\varphi$ in direction $\eta-\bar\varphi$ for arbitrary $\eta\in\Phi_{ad}$ is given by

$$\langle g_k'(\bar\varphi), \eta-\bar\varphi\rangle = a_k(\bar\varphi-\varphi,\eta-\bar\varphi) + \lambda_k \left\langle j'(\varphi),\eta-\bar\varphi\right\rangle \ .$$

Since the problem (2.4) is convex, it is equivalent to the first order optimality condition, which is given by the variational inequality (2.5); see [15]. ∎

We see that $\varphi\in\Phi_{ad}$ is a stationary point of j, that is, $\langle j'(\varphi),\eta-\varphi\rangle \geq 0$ for all $\eta\in\Phi_{ad}$, if and only if $\overline{\varphi} = \varphi$ is the solution of (2.5), i.e., the fixed point equation $\varphi = \mathcal{P}_k(\varphi)$ is fulfilled. This leads to the classical view of the method as a fixed point iteration $\varphi_{k+1} = \mathcal{P}_k(\varphi_k)$ in the case that $\mathcal{P}_k$ is independent of k and $\alpha_k = 1$ is chosen.

COROLLARY 2.5. *If there exists some $k\in\mathbb{N}_0$ with $\mathcal{P}_k(\varphi)=\varphi$, then φ is a stationary point of j with respect to Φ_{ad}. On the other hand, if $\varphi\in\Phi_{ad}$ is a stationary point of* (1.1)*, then the fixed point equation $\mathcal{P}_k(\varphi)=\varphi$ holds for all $k\in\mathbb{N}_0$. In particular, an iterate φ_k of the algorithm is a stationary point of j if and only if $v_k = \mathcal{P}_k(\varphi_k)-\varphi_k = 0$.*

The variational inequality (2.5) tested with $\eta=\varphi\in\Phi_{ad}$ together with (A8) and (A12) yields that $\mathcal{P}_k(\varphi)-\varphi$ is a descent direction for j.

LEMMA 2.6. *Let $k\in\mathbb{N}_0$, $\varphi\in\Phi_{ad}$ and $v := \mathcal{P}_k(\varphi)-\varphi$. Then it holds that*

$$\langle j'(\varphi),v\rangle \ \leq -\frac{c_1}{\lambda_{\max}}\|v\|_{\mathbb{X}}^2. \tag{2.8}$$

Note that (2.8) does not hold in the $\mathbb{X}\cap\mathbb{D}$-norm.

Due to $\langle j'(\varphi),v\rangle < 0$ for $v\neq 0$ the step length selection by the Armijo rule (see step 10 in Algorithm 2.1) is well defined, which can be shown as in [3].

Remark 2.7. For the existence of a step length and for the global convergence proof we exploit that the path $\alpha\mapsto\varphi_k+\alpha v_k$ is continuous in $\mathbb{X}\cap\mathbb{D}$. Thus, the mapping $\alpha\mapsto j(\varphi_k+\alpha v_k)$ is also continuous. On the other hand, this does not hold for the gradient path. Backtracking along the gradient path or projection arc means that α_k is set to 1, whereas $\lambda_k=\beta^{m_k}$ is chosen with $m_k\in\mathbb{N}_0$ minimal such that the Armijo condition

$$j(\overline{\varphi}_k(\lambda_k)) \leq j(\varphi_k) + \sigma\left\langle j'(\varphi_k),\overline{\varphi}_k(\lambda_k)-\varphi_k\right\rangle$$

is satisfied; see, for instance, [25]. By the notation $\overline{\varphi}_k(\lambda_k)$ we emphasize that the solution of the subproblem (2.2) depends on λ_k. However, with the above assumptions it cannot be shown by the standard techniques used in the literature that there exists such a λ_k. The reason is that due to (A8) the gradient path $\lambda\mapsto\overline{\varphi}_k(\lambda)$ is continuous with respect to the $\mathbb{X}$-norm, whereas j is due to (A5) only differentiable with respect to the $\mathbb{X}\cap\mathbb{D}$-norm. Thus, j along the gradient path, i.e., the mapping $\lambda\mapsto j(\overline{\varphi}_k(\lambda))$, may be discontinuous.

To prove statement 2 of Theorem 2.2 we use, as in [3] for finite dimensions, that v_k is gradient related (see Lemma 2.10). This is weaker than the common angle condition. Therefore we need the following two lemmata.

LEMMA 2.8. *For* $\{\varphi_k\} \subseteq \Phi_{ad}$ *with* $\varphi_k \to \varphi$ *in* $\mathbb{X}\cap\mathbb{D}$ *and* $\{p_k\} \subseteq \mathbb{X}\cap\mathbb{D}$ *with* $p_k \to p$ *weakly in* $\mathbb{X}$ *and weakly-* in* $\mathbb{D}$ *for some* $\varphi, p \in \mathbb{X}\cap\mathbb{D}$ *it holds that* $\langle j'(\varphi_k), p_k\rangle \to \langle j'(\varphi), p\rangle$.

Proof. We use (A5) and (A6) and obtain

$$\begin{aligned} |\langle j'(\varphi_k), p_k\rangle - \langle j'(\varphi), p\rangle| &\le |\langle j'(\varphi_k) - j'(\varphi), p_k\rangle| + |\langle j'(\varphi), p_k - p\rangle| \\ &\le \underbrace{\|j'(\varphi_k) - j'(\varphi)\|_{(\mathbb{X}\cap\mathbb{D})^*}}_{\to 0}\underbrace{\|p_k\|_{\mathbb{X}\cap\mathbb{D}}}_{\le C} + \underbrace{|\langle j'(\varphi), p_k - p\rangle|}_{\to 0} \to 0. \end{aligned}$$ ∎

The preceding lemma is also needed in the proof of Theorem 2.2.

LEMMA 2.9. *Let for a sequence* $\{\varphi_i\} \subseteq \Phi_{ad}$ *hold* $\varphi_i \to \varphi$ *in* $\mathbb{X}\cap\mathbb{D}$ *for some* $\varphi \in \mathbb{X}\cap\mathbb{D}$. *Then there exists* $C > 0$ *such that* $\|\mathcal{P}_k(\varphi_i)\|_{\mathbb{X}\cap\mathbb{D}} \le C$ *for all* $i, k \in \mathbb{N}_0$.

Proof. Lemma 2.6 yields together with (A3) and (A5) the estimate

$$\begin{aligned} \frac{c_1}{\lambda_{\max}}\|\mathcal{P}_k(\varphi_i) - \varphi_i\|_{\mathbb{X}}^2 &\le -\langle j'(\varphi_i), \mathcal{P}_k(\varphi_i) - \varphi_i\rangle \\ &\le \|j'(\varphi_i)\|_{(\mathbb{X}\cap\mathbb{D})^*}(\|\mathcal{P}_k(\varphi_i) - \varphi_i\|_{\mathbb{X}} + \|\mathcal{P}_k(\varphi_i) - \varphi_i\|_{\mathbb{D}}) \\ &\le C(\|\mathcal{P}_k(\varphi_i) - \varphi_i\|_{\mathbb{X}} + 1), \end{aligned}$$

thus $\|\mathcal{P}_k(\varphi_i) - \varphi_i\|_{\mathbb{X}} \le C$, and hence $\|\mathcal{P}_k(\varphi_i)\|_{\mathbb{X}} \le C$. Due to (A3) we finally get $\|\mathcal{P}_k(\varphi_i)\|_{\mathbb{X}\cap\mathbb{D}} \le C$ independent of i and k. ∎

LEMMA 2.10. *Let* $\{\varphi_k\}$ *be the sequence generated by Algorithm* 2.1, *and then* $\{v_k\}$ *is gradient related, i.e., for any subsequence* $\{\varphi_{k_i}\}$ *which converges in* $\mathbb{X}\cap\mathbb{D}$ *to a nonstationary point* $\varphi \in \Phi_{ad}$ *of* j, *the corresponding subsequence of search directions* $\{v_{k_i}\}$ *is bounded in* $\mathbb{X}\cap\mathbb{D}$ *and* $\limsup_i \langle j'(\varphi_{k_i}), v_{k_i}\rangle < 0$ *is satisfied. Moreover, it holds that* $\liminf_i \|v_{k_i}\|_{\mathbb{X}} > 0$.

Proof. Let $\varphi_{k_i} \to \varphi$ in $\mathbb{X}\cap\mathbb{D}$, where φ is nonstationary. Lemma 2.9 provides that $\{v_{k_i}\}$ is bounded in $\mathbb{X}\cap\mathbb{D}$. With (2.8), the statement $\limsup_i \langle j'(\varphi_{k_i}), v_{k_i}\rangle < 0$ follows from $\liminf_i \|v_{k_i}\|_{\mathbb{X}} = C > 0$, which we show by contradiction.

Assume $\liminf_i \|v_{k_i}\|_{\mathbb{X}} = 0$; thus there is a subsequence again denoted by $\{v_{k_i}\}$ such that $v_{k_i} \to 0$ in $\mathbb{X}$. Using (2.5) for $\bar{\varphi}_k := \mathcal{P}_k(\varphi_k)$, the positive definiteness of a_k, and (A12), it follows for all $\eta \in \Phi_{ad}$

$$\begin{aligned} \langle j'(\varphi_k), \eta - \bar{\varphi}_k\rangle &\ge \frac{1}{\lambda_k}(a_k(v_k, v_k) + a_k(v_k, \bar{\varphi}_k - v_k - \eta)) \\ &\ge -\frac{1}{\lambda_{\min}}|a_k(v_k, \bar{\varphi}_k - v_k - \eta)|\,. \end{aligned} \tag{2.9}$$

Moreover, $\bar{\varphi}_{k_i} = v_{k_i} + \varphi_{k_i} \to \varphi$ in $\mathbb{X}$ and also weakly-* in $\mathbb{D}$ according to Lemma 2.3. From Lemma 2.8 we get $\langle j'(\varphi_{k_i}), \eta - \bar{\varphi}_{k_i}\rangle \to \langle j'(\varphi), \eta - \varphi\rangle$. From (A11) we get $a_{k_i}(\bar{\varphi}_{k_i} - \varphi_{k_i}, \varphi_{k_i} - \eta) \to 0$ and derive from (2.9) that

$$\langle j'(\varphi), \eta - \varphi\rangle \ge 0 \quad \forall \eta \in \Phi_{ad},$$

which shows that φ is stationary, which is a contradiction. ∎

Proof of Theorem 2.2. Because of Corollary 2.5 we can assume $v_k \neq 0$ and thus $\alpha_k > 0$ for all k.

1. From the Armijo rule and since v_k is a descent direction we get

$$j(\varphi_{k+1}) - j(\varphi_k) \leq \alpha_k \sigma \langle j'(\varphi_k), v_k \rangle < 0, \tag{2.10}$$

and thus $j(\varphi_k)$ is monotonically decreasing. Since j is bounded from below we get convergence $j(\varphi_k) \to j^*$ for some $j^* \in \mathbb{R}$, which proves 1.

2. The proof is similar to [3] in finite dimension by contradiction. Let φ be an accumulation point, with a convergent subsequence $\varphi_{k_i} \to \varphi$ in $\mathbb{X}\cap\mathbb{D}$. The continuity of j on Φ_{ad} then yields $j^* = j(\varphi)$ and (2.10) leads to $\alpha_k \langle j'(\varphi_k), v_k \rangle \to 0$. Assuming now that φ is nonstationary we have $|\langle j'(\varphi_{k_i}), v_{k_i} \rangle| \geq C > 0$, since $\{v_k\}$ is gradient related by Lemma 2.10, and thus $\alpha_{k_i} \to 0$. So there exists some $\bar{i} \in \mathbb{N}$ such that $\alpha_{k_i}/\beta \leq 1$ for all $i \geq \bar{i}$, and thus α_{k_i}/β does not fulfill the Armijo rule due to the minimality of m_k. Applying the mean value theorem to the left-hand side of the Armijo condition, we have for some nonnegative $\tilde{\alpha}_{k_i} \leq \frac{\alpha_{k_i}}{\beta}$ and all $i \geq \bar{i}$ that

$$\frac{\alpha_{k_i}}{\beta} \langle j'\left(\varphi_{k_i} + \tilde{\alpha}_{k_i} v_{k_i}\right), v_{k_i} \rangle = j\left(\varphi_{k_i} + \frac{\alpha_{k_i}}{\beta} v_{k_i}\right) - j(\varphi_{k_i}) > \frac{\alpha_{k_i}}{\beta} \sigma \langle j'(\varphi_{k_i}), v_{k_i} \rangle \tag{2.11}$$

holds. Since, by Lemma 2.10, $\{v_{k_i}\}$ is bounded in $\mathbb{X}\cap\mathbb{D}$ and $\tilde{\alpha}_{k_i} \to 0$, we have that $\varphi_{k_i} + \tilde{\alpha}_{k_i} v_{k_i} \to \varphi$ in $\mathbb{X}\cap\mathbb{D}$. Also $\bar{\varphi}_{k_i} = \varphi_{k_i} + v_{k_i}$ is uniformly bounded in $\mathbb{X}\cap\mathbb{D}$, and thus there exists a subsequence, again denoted by $\{\bar{\varphi}_{k_i}\}$, which converges to some $y \in \Phi_{ad}$ weakly in $\mathbb{X}$ and weakly-* in $\mathbb{D}$. Hence we have that $v_{k_i} = \bar{\varphi}_{k_i} - \varphi_{k_i} \to \bar{v} := y - \varphi$ weakly in $\mathbb{X}$ and weakly-* in $\mathbb{D}$. According to Lemma 2.8 we can take the limit of both sides of the inequality (2.11), which leads to $\langle j'(\varphi), \bar{v} \rangle \geq \sigma \langle j'(\varphi), \bar{v} \rangle$, and $\sigma < 1$ yields $\langle j'(\varphi), \bar{v} \rangle \geq 0$. This contradicts $\langle j'(\varphi), \bar{v} \rangle = \limsup_i \langle j'(\varphi_{k_i}), v_{k_i} \rangle < 0$, which is a consequence of Lemma 2.10.

3. By proving that out of any subsequence of $\langle j'(\varphi_{k_i}), v_{k_i} \rangle$ we can extract another subsequence, which converges to 0, we can conclude that $\langle j'(\varphi_{k_i}), v_{k_i} \rangle \to 0$ which yields $\|v_{k_i}\|_{\mathbb{X}} \to 0$ by (2.8). With Lemma 2.9, we get by the same arguments as in 2 that $v_{k_i} \to y - \varphi$ weakly in $\mathbb{X}$ and weakly-* in $\mathbb{D}$ for a subsequence and for some $y \in \Phi_{ad}$, and thus $\langle j'(\varphi_{k_i}), v_{k_i} \rangle \to \langle j'(\varphi), y - \varphi \rangle$ due to Lemma 2.8. Since v_{k_i} are descent directions for j at φ_{k_i} and φ is stationary we have $\langle j'(\varphi), y - \varphi \rangle = 0$.

4. As in 3 we prove by a subsequence argument that $\langle j'(\varphi_k), v_k \rangle \to 0$. For an arbitrary subsequence, which we also denote by index k, (2.10) yields $\alpha_k \langle j'(\varphi_k), v_k \rangle \to 0$. If $\alpha_k \geq c > 0$ for all k, the assertion follows immediately. Otherwise there exists a subsequence (again denoted by index k) such that $\beta \geq \alpha_k \to 0$ and thus the step length α_k/β does not fulfill the Armijo condition. Since j' is Hölder continuous with exponent γ and modulus L we obtain

$$\begin{aligned}\sigma \frac{\alpha_k}{\beta} \langle j'(\varphi_k), v_k \rangle < j\left(\varphi_k + \frac{\alpha_k}{\beta} v_k\right) - j(\varphi_k) &= \int_0^1 \frac{d}{dt} j\left(\varphi_k + t\frac{\alpha_k}{\beta} v_k\right) dt \\ &\leq \frac{\alpha_k}{\beta} \langle j'(\varphi_k), v_k \rangle + \frac{L}{1+\gamma} \left(\frac{\alpha_k}{\beta}\right)^{1+\gamma} \|v_k\|_{\mathbb{X}\cap\mathbb{D}}^{1+\gamma}.\end{aligned}$$

It holds that $\|v_k\|_{\mathbb{D}} \leq C$ due to (A3), and employing (2.8) we obtain

$$
\begin{aligned}
0 < (\sigma - 1) \langle j'(\varphi_k), v_k \rangle &< C \frac{L}{1+\gamma} \left(\frac{\alpha_k}{\beta} \right)^{\gamma} \left(\|v_k\|_{\mathbb{X}}^{1+\gamma} + 1 \right) \\
&\leq C \alpha_k^{\gamma} \left(| \langle j'(\varphi_k), v_k \rangle |^{\frac{1+\gamma}{2}} + 1 \right).
\end{aligned}
$$

We get $x_k := | \langle j'(\varphi_k), v_k \rangle | \to 0$. Otherwise there exists a subsequence still denoted by $\{x_k\}$ with $x_k \to \bar{c} > 0$. Rearranging the last inequality gives $1 < C\alpha_k^{\gamma}(x_k^{\frac{-1+\gamma}{2}} + x_k^{-1}) \to 0$, which is a contradiction. ∎

Remark 2.11. Statements 1 and 2 of Theorem 2.2 require only that $\overline{\varphi}_k \in \Phi_{ad}$ is chosen such that the search directions $v_k = \overline{\varphi}_k - \varphi_k$ are gradient related descent directions, as can be seen in the proof above. Hence, $\overline{\varphi}_k$ does not have to coincide with $\mathcal{P}_k(\varphi_k)$ in Algorithm 2.1. In this case assumption (A3) is also not required.

3. An application in structural topology optimization based on a phase field model. In this section we give an example of an optimization problem described in [5], which is not differentiable in a Hilbert space, so the classical projected gradient method cannot be applied, but the assumptions for the VMPT method are fulfilled. We consider the problem of distributing N materials, each with different elastic properties and fixed volume fraction, within a design domain $\Omega \subseteq \mathbb{R}^d$, $d \in \mathbb{N}$, such that the mean compliance $\int_{\Gamma_g} \boldsymbol{g} \cdot \boldsymbol{u}$ is minimal under the external force $\boldsymbol{g}$ acting on $\Gamma_g \subseteq \partial\Omega$. The displacement field $\boldsymbol{u} : \Omega \to \mathbb{R}^d$ is given as the solution of the equations of linear elasticity (3.2). To obtain a well posed problem a perimeter penalization is typically used. Using phase fields in topology optimization was introduced in [8]. Here, the N materials are described by a vector valued phase field $\boldsymbol{\varphi} : \Omega \to \mathbb{R}^N$ with $\boldsymbol{\varphi} \geq 0$ and $\sum_i \varphi_i = 1$, which is able to handle topological changes implicitly. The ith material is characterized by $\{\boldsymbol{\varphi}_i = 1\}$ and the different materials are separated by a thin interface, whose thickness is controlled by the phase field parameter $\varepsilon > 0$. In the phase field setting the perimeter is approximated by the Ginzburg–Landau energy

$$
E(\boldsymbol{\varphi}) := \int_{\Omega} \left\{ \frac{\varepsilon}{2} |\nabla \boldsymbol{\varphi}|^2 + \frac{1}{\varepsilon} \psi_0(\boldsymbol{\varphi}) \right\}.
$$

In [6] it is shown that the given problem for $N = 2$ converges as $\varepsilon \to 0$ in the sense of Γ-convergence. For further details about the model we refer the reader to [5]. The resulting optimal control problem reads

$$
\min \tilde{J}(\boldsymbol{\varphi}, \boldsymbol{u}) := \int_{\Gamma_g} \boldsymbol{g} \cdot \boldsymbol{u} + \gamma E(\boldsymbol{\varphi}), \tag{3.1}
$$

$$
\boldsymbol{\varphi} \in H^1(\Omega)^N,\ \boldsymbol{u} \in H_D^1 := \{ H^1(\Omega)^d \mid \boldsymbol{\xi}|_{\Gamma_D} = 0 \}
$$

$$
\text{subject to} \quad \int_{\Omega} \boldsymbol{C}(\boldsymbol{\varphi}) \mathcal{E}(\boldsymbol{u}) : \mathcal{E}(\boldsymbol{\xi}) = \int_{\Gamma_g} \boldsymbol{g} \cdot \boldsymbol{\xi} \quad \forall \boldsymbol{\xi} \in H_D^1 \tag{3.2}
$$

$$
\fint_{\Omega} \boldsymbol{\varphi} = \mathfrak{m}, \quad \boldsymbol{\varphi} \geq 0, \quad \sum_{i=1}^{N} \varphi^i \equiv 1, \tag{3.3}
$$

where $\gamma > 0$ is a weighting factor, $\fint_\Omega \boldsymbol{\varphi} := \frac{1}{|\Omega|} \int_\Omega \boldsymbol{\varphi}$, $\psi_0 : \mathbb{R}^N \to \mathbb{R}$ is the smooth part of the potential forcing the values of $\boldsymbol{\varphi}$ to the standard basis $\boldsymbol{e}_i \in \mathbb{R}^N$, and

$A : B := \sum_{i,j=1}^d A_{ij}B_{ij}$ for $A, B \in \mathbb{R}^{d\times d}$. The materials are fixed on the Dirichlet domain $\Gamma_D \subseteq \partial\Omega$. The tensor valued mapping $\boldsymbol{C} : \mathbb{R}^N \to \mathbb{R}^{d\times d} \otimes (\mathbb{R}^{d\times d})^*$ is a suitable interpolation of the stiffness tensors $\boldsymbol{C}(\boldsymbol{e}_i)$ of the different materials, and $\mathcal{E}(\boldsymbol{u}) := \frac{1}{2}(\nabla\boldsymbol{u} + \nabla\boldsymbol{u}^T)$ is the linearized strain tensor. The prescribed volume fraction of the ith material is given by $\mathfrak{m}_i$. For examples of the functions ψ_0 and $\boldsymbol{C}$ we refer to [4, 5]. The existence of a minimizer of the problem (3.1)–(3.3) as well as the unique solvability of the state equation (3.2) is shown in [5] under the following assumptions, which we claim also in this paper.

(AP) $\Omega \subseteq \mathbb{R}^d$ is a bounded Lipschitz domain; $\Gamma_D, \Gamma_g \subseteq \partial\Omega$ with $\Gamma_D \cap \Gamma_g = \emptyset$ and $\mathcal{H}^{d-1}(\Gamma_D) > 0$. Moreover, $\boldsymbol{g} \in L^2(\Gamma_g)^d$ and $\psi_0 \in C^{1,1}(\mathbb{R}^N)$ as well as $\mathfrak{m} \geq 0$, $\sum_{i=1}^N \mathfrak{m}_i = 1$. For the stiffness tensor we assume $\boldsymbol{C} = (C_{ijkl})_{i,j,k,l=1}^d$ with $C_{ijkl} \in C^{1,1}(\mathbb{R}^N)$ and $C_{ijkl} = C_{jikl} = C_{klij}$ and that there exist $a_0, a_1, C > 0$, s.t. $a_0|\boldsymbol{A}|^2 \leq \boldsymbol{C}(\boldsymbol{\varphi})\boldsymbol{A} : \boldsymbol{A} \leq a_1|\boldsymbol{A}|^2$ as well as $|\boldsymbol{C}'(\boldsymbol{\varphi})| \leq C$ holds for all symmetric matrices $\boldsymbol{A} \in \mathbb{R}^{d\times d}$ and for all $\boldsymbol{\varphi} \in \mathbb{R}^N$.

The state $\boldsymbol{u}$ can be eliminated using the control-to-state operator S, resulting in the reduced cost functional $\tilde{j}(\boldsymbol{\varphi}) := \tilde{J}(\boldsymbol{\varphi}, S(\boldsymbol{\varphi}))$. In [5] it is also shown that $\tilde{j} : H^1(\Omega)^N \cap L^\infty(\Omega)^N \to \mathbb{R}$ is everywhere Fréchet differentiable with derivative

$$\tilde{j}'(\boldsymbol{\varphi})\boldsymbol{v} = \gamma \int_\Omega \left\{ \varepsilon\nabla\boldsymbol{\varphi} : \nabla\boldsymbol{v} + \frac{1}{\varepsilon}\psi_0'(\boldsymbol{\varphi})\boldsymbol{v} \right\} - \int_\Omega \boldsymbol{C}'(\boldsymbol{\varphi})\boldsymbol{v}\mathcal{E}(\boldsymbol{u}) : \mathcal{E}(\boldsymbol{u}) \tag{3.4}$$

for all $\boldsymbol{\varphi}, \boldsymbol{v} \in H^1(\Omega)^N \cap L^\infty(\Omega)^N$, where $\boldsymbol{u} = S(\boldsymbol{\varphi})$ and $S : L^\infty(\Omega)^N \to H^1(\Omega)^d$ is Fréchet differentiable. By the techniques in [5] one can also show that S' is continuous.

In [5, 7] the problem is solved numerically by a pseudo time stepping method with fixed time step, which results from an L^2-gradient flow approach. An H^{-1} gradient flow approach is also considered in [7]. The drawbacks of these methods are that no convergence results to a stationary point exist, and hence also no appropriate stopping criteria are known. In addition, typically the methods are very slow, i.e., many time steps are needed until the changes in the solution $\boldsymbol{\varphi}$ or in j are small. Here we apply the VMPT method, which does not have these drawbacks and which can additionally incorporate second order information.

Since $H^1(\Omega)^N \cap L^\infty(\Omega)^N$ is not a Hilbert space, the classical projected gradient method cannot be applied. In the following we show that problem (3.1)–(3.3) fulfills the assumptions on the VMPT method. Among others we use the inner product $a_k(\boldsymbol{f}, \boldsymbol{g}) = \int_\Omega \nabla\boldsymbol{f} : \nabla\boldsymbol{g}$. To guarantee positive definiteness of this a_k we first have to translate the problem by a constant to gain $\int_\Omega \boldsymbol{\varphi} = 0$, which allows us to apply a Poincaré inequality. Therefore we perform a change of coordinates in the form $\tilde{\boldsymbol{\varphi}} = \boldsymbol{\varphi} - \mathfrak{m}$ and get the following problem for the transformed coordinates:

$$\begin{aligned} &\min j(\boldsymbol{\varphi}) := \int_{\Gamma_g} \boldsymbol{g} \cdot S(\boldsymbol{\varphi} + \mathfrak{m}) + \gamma E(\boldsymbol{\varphi} + \mathfrak{m}), \\ &\boldsymbol{\varphi} \in \Phi_{ad} := \left\{ \boldsymbol{\varphi} \in H^1(\Omega)^N \;\middle|\; \fint_\Omega \boldsymbol{\varphi} = 0, \quad \boldsymbol{\varphi} \geq -\mathfrak{m}, \quad \sum_{i=1}^N \varphi^i \equiv 0 \right\}. \end{aligned} \tag{3.5}$$

On the transformed problem (3.5) we apply the VMPT method in the spaces

$$\mathbb{X} := \left\{ \boldsymbol{\varphi} \in H^1(\Omega)^N \;\middle|\; \fint_\Omega \boldsymbol{\varphi} = \boldsymbol{0} \right\}, \quad \mathbb{D} := L^\infty(\Omega)^N.$$

The space of mean value free functions $\mathbb{X}$ becomes a Hilbert space with the inner product $(\boldsymbol{f},\boldsymbol{g})_{\mathbb{X}} := (\nabla \boldsymbol{f}, \nabla \boldsymbol{g})_{L^2}$, and $\|.\|_{\mathbb{X}}$ is equivalent to the H^1-norm [1].

THEOREM 3.1. *The reduced cost functional $j : \mathbb{X}\cap\mathbb{D} \to \mathbb{R}$ is continuously Fréchet differentiable, and j' is Lipschitz continuous on Φ_{ad}.*

Proof. The Fréchet differentiability of j on $\mathbb{X}\cap\mathbb{D}$ is shown in [5]. Let $\boldsymbol{\eta}, \boldsymbol{\varphi}_i \in \mathbb{X}\cap\mathbb{D}$ and $\boldsymbol{u}_i = S(\boldsymbol{\varphi}_i)$, $i = 1,2$. Then with (3.4), $\psi_0 \in C^{1,1}(\mathbb{R}^N)$, $C_{ijkl} \in C^{1,1}(\mathbb{R}^N)$, and $|\boldsymbol{C}'(\boldsymbol{\varphi})| \le C$ for all $\boldsymbol{\varphi} \in \mathbb{R}^N$ we get

$$\begin{aligned}
|(j'(\boldsymbol{\varphi}_1) - j'(\boldsymbol{\varphi}_2))\boldsymbol{\eta}| \le{}& \gamma\varepsilon \|\boldsymbol{\varphi}_1 - \boldsymbol{\varphi}_2\|_{H^1}\|\boldsymbol{\eta}\|_{H^1} + C\frac{\gamma}{\varepsilon}\|\boldsymbol{\varphi}_1 - \boldsymbol{\varphi}_2\|_{L^2}\|\boldsymbol{\eta}\|_{L^2} \\
&+ \left| \int_\Omega (\boldsymbol{C}'(\mathfrak{m} + \boldsymbol{\varphi}_1) - \boldsymbol{C}'(\mathfrak{m} + \boldsymbol{\varphi}_2))(\boldsymbol{\eta})\mathcal{E}(\boldsymbol{u}_1) : \mathcal{E}(\boldsymbol{u}_1) \right| \\
&+ \left| \int_\Omega \boldsymbol{C}'(\mathfrak{m} + \boldsymbol{\varphi}_2)(\boldsymbol{\eta})\mathcal{E}(\boldsymbol{u}_1 - \boldsymbol{u}_2) : \mathcal{E}(\boldsymbol{u}_1) \right| \\
&+ \left| \int_\Omega \boldsymbol{C}'(\mathfrak{m} + \boldsymbol{\varphi}_2)(\boldsymbol{\eta})\mathcal{E}(\boldsymbol{u}_2) : \mathcal{E}(\boldsymbol{u}_1 - \boldsymbol{u}_2) \right| \\
\le{}& C\|\boldsymbol{\varphi}_1 - \boldsymbol{\varphi}_2\|_{H^1}\|\boldsymbol{\eta}\|_{H^1} \\
&+ \|(\boldsymbol{C}'(\mathfrak{m} + \boldsymbol{\varphi}_1) - \boldsymbol{C}'(\mathfrak{m} + \boldsymbol{\varphi}_2))\boldsymbol{\eta}\|_{L^\infty}\|\boldsymbol{u}_1\|^2_{H^1} \\
&+ C\|\boldsymbol{\eta}\|_{L^\infty}\|\boldsymbol{u}_1 - \boldsymbol{u}_2\|_{H^1}(\|\boldsymbol{u}_1\|_{H^1} + \|\boldsymbol{u}_2\|_{H^1}) \\
\le{}& C\|\boldsymbol{\eta}\|_{H^1\cap L^\infty}\{\|\boldsymbol{\varphi}_1 - \boldsymbol{\varphi}_2\|_{H^1} + \|\boldsymbol{\varphi}_1 - \boldsymbol{\varphi}_2\|_{L^\infty}\|\boldsymbol{u}_1\|^2_{H^1} \\
&\qquad + \|\boldsymbol{u}_1 - \boldsymbol{u}_2\|_{H^1}(\|\boldsymbol{u}_1\|_{H^1} + \|\boldsymbol{u}_2\|_{H^1})\}.
\end{aligned} \tag{3.6}$$

To show the continuity of j', let $\boldsymbol{\varphi}_n, \boldsymbol{\varphi} \in \mathbb{X}\cap\mathbb{D}$ for $n \in \mathbb{N}$ with $\boldsymbol{\varphi}_n \to \boldsymbol{\varphi}$ in $\mathbb{X}\cap\mathbb{D}$. Using (3.6) yields

$$\begin{aligned}
&\|j'(\boldsymbol{\varphi}_n) - j'(\boldsymbol{\varphi})\|_{(H^1\cap L^\infty)^*} \\
&\qquad \le C(\|\boldsymbol{\varphi}_n - \boldsymbol{\varphi}\|_{H^1\cap L^\infty}(1 + \|\boldsymbol{u}_n\|^2_{H^1}) + \|\boldsymbol{u}_n - \boldsymbol{u}\|_{H^1}(\|\boldsymbol{u}_n\|_{H^1} + \|\boldsymbol{u}\|_{H^1})),
\end{aligned}$$

where $\boldsymbol{u}_n = S(\boldsymbol{\varphi}_n)$ and $\boldsymbol{u} = S(\boldsymbol{\varphi})$. From the continuity of S we get that $\|\boldsymbol{u}_n\|_{H^1}$ is bounded and that $\|\boldsymbol{u}_n - \boldsymbol{u}\|_{H^1} \to 0$ as $n \to \infty$. This implies

$$\|j'(\boldsymbol{\varphi}_n) - j'(\boldsymbol{\varphi})\|_{(H^1\cap L^\infty)^*} \to 0$$

and thus $j \in C^1(\mathbb{X}\cap\mathbb{D})$.

For the Lipschitz continuity of j' we employ estimate (3.6) with $\boldsymbol{\varphi}_i \in \Phi_{ad}$, $i = 1,2$. Since Φ_{ad} is bounded in L^∞, we get that S is Lipschitz continuous on Φ_{ad} and that $\|S(\boldsymbol{\varphi})\|_{H^1} \le C$, independent of $\boldsymbol{\varphi} \in \Phi_{ad}$; see [5]. This yields

$$\|j'(\boldsymbol{\varphi}_1) - j'(\boldsymbol{\varphi}_2)\|_{(H^1\cap L^\infty)^*} \le C\|\boldsymbol{\varphi}_1 - \boldsymbol{\varphi}_2\|_{H^1\cap L^\infty},$$

which proves the Lipschitz continuity of j' in Φ_{ad}. ∎

COROLLARY 3.2. *The spaces $\mathbb{X}$ and $\mathbb{D}$ together with j and Φ_{ad} given in* (3.5) *fulfill the assumptions* (A1)–(A6) *of the VMPT method.*

Proof. Given the choices for $\mathbb{X}$ and $\mathbb{D}$ (A1) is fulfilled. For $\boldsymbol{\varphi} \in \Phi_{ad}$ we have

$$-\boldsymbol{1} \le -\mathfrak{m} \le \boldsymbol{\varphi} \le \boldsymbol{1} - \mathfrak{m} \le \boldsymbol{1} \quad \forall \boldsymbol{\varphi} \in \Phi_{ad}$$

almost everywhere in Ω. Thus (A3) holds and $\Phi_{ad} \subseteq \mathbb{X}\cap\mathbb{D}$. Moreover, $\boldsymbol{0} \in \Phi_{ad}$, Φ_{ad} is convex, and since Φ_{ad} is closed in $L^2(\Omega)^N$, it is also closed in $\mathbb{X} \hookrightarrow L^2(\Omega)^N$. Thus (A2) holds.

Assumption (A4) is shown in [5], and Theorem 3.1 provides (A5).
Given

$$\langle j'(\boldsymbol{\varphi}), \boldsymbol{\varphi}_i \rangle = \int_\Omega \left\{ \gamma\varepsilon \nabla \boldsymbol{\varphi} : \nabla \boldsymbol{\varphi}_i + \left(\frac{\gamma}{\varepsilon} \nabla \psi_0(\boldsymbol{\varphi} + \mathfrak{m}) - \nabla \boldsymbol{C}(\boldsymbol{\varphi} + \mathfrak{m}) \mathcal{E}(\boldsymbol{u}) : \mathcal{E}(\boldsymbol{u}) \right) \cdot \boldsymbol{\varphi}_i \right\}$$

the first term converges to 0 if $\boldsymbol{\varphi}_i \to 0$ weakly in H^1. With (AP) and $\boldsymbol{u} \in H^1_D$ we have that $\frac{\gamma}{\varepsilon} \nabla \psi_0(\boldsymbol{\varphi} + \mathfrak{m}) - \nabla \boldsymbol{C}(\boldsymbol{\varphi} + \mathfrak{m}) \mathcal{E}(\boldsymbol{u}) : \mathcal{E}(\boldsymbol{u}) \in L^1(\Omega)^N$. Hence the remaining term converges to 0 if $\boldsymbol{\varphi}_i \to 0$ weakly-* in L^∞, which proves that (A6) is fulfilled. ☐

Possible choices of the inner product a_k for the VMPT method are the inner product on $\mathbb{X}$, i.e.,

$$a_k(\boldsymbol{p}, \boldsymbol{y}) = (\boldsymbol{p}, \boldsymbol{y})_{\mathbb{X}} = \int_\Omega \nabla \boldsymbol{p} : \nabla \boldsymbol{y} \tag{3.7}$$

and the scaled version $a_k(\boldsymbol{p}, \boldsymbol{y}) = \gamma\varepsilon (\boldsymbol{p}, \boldsymbol{y})_{\mathbb{X}}$. Both fulfill the assumptions (A7)–(A11). We also give an example of a pointwise choice of an inner product, which includes second order information. Since this choice is not continuous in $\mathbb{X}$, it is not obvious that it fulfills the assumptions. To motivate the choice of this inner product we look at the second order derivative of j, which is formally given by

$$\begin{aligned} j''(\boldsymbol{\varphi}_k)[\boldsymbol{p}, \boldsymbol{y}] = \int_\Omega \Big\{ & \gamma\varepsilon \nabla \boldsymbol{p} : \nabla \boldsymbol{y} - 2(\boldsymbol{C}'(\mathfrak{m} + \boldsymbol{\varphi}_k)(\boldsymbol{y}) \mathcal{E}(S'(\boldsymbol{\varphi}_k)\boldsymbol{p}) : \mathcal{E}(\boldsymbol{u}_k)) \\ & + \frac{\gamma}{\varepsilon} \nabla^2 \psi_0(\mathfrak{m} + \boldsymbol{\varphi}_k) \boldsymbol{p} \cdot \boldsymbol{y} - \boldsymbol{C}''(\mathfrak{m} + \boldsymbol{\varphi}_k)[\boldsymbol{p}, \boldsymbol{y}] \mathcal{E}(\boldsymbol{u}_k) : \mathcal{E}(\boldsymbol{u}_k) \Big\}. \end{aligned}$$

In [5] it is shown that $\boldsymbol{z}_p := S'(\boldsymbol{\varphi}_k)\boldsymbol{p} \in H^1_D$ is the unique weak solution of the linearized state equation

$$\int_\Omega \boldsymbol{C}(\mathfrak{m} + \boldsymbol{\varphi}_k) \mathcal{E}(\boldsymbol{z}_p) : \mathcal{E}(\boldsymbol{\eta}) = - \int_\Omega \boldsymbol{C}'(\mathfrak{m} + \boldsymbol{\varphi}_k) \boldsymbol{p} \mathcal{E}(\boldsymbol{u}_k) : \mathcal{E}(\boldsymbol{\eta}) \quad \forall \boldsymbol{\eta} \in H^1_D \tag{3.8}$$

and that $\|\boldsymbol{z}_p\|_{H^1} \le C_k \|\boldsymbol{p}\|_{L^\infty}$ holds. Since the first two terms in j'' define an inner product (see proof of Theorem 3.3), we use

$$a_k(\boldsymbol{p}, \boldsymbol{y}) = \gamma\varepsilon (\boldsymbol{p}, \boldsymbol{y})_{\mathbb{X}} - 2 \int_\Omega \boldsymbol{C}'(\mathfrak{m} + \boldsymbol{\varphi}_k)(\boldsymbol{y}) \mathcal{E}(\boldsymbol{z}_p) : \mathcal{E}(\boldsymbol{u}_k) \tag{3.9}$$

as an approximation of $j''(\boldsymbol{\varphi}_k)$. Testing (3.8) for $\boldsymbol{z}_y = S'(\boldsymbol{\varphi}_k)\boldsymbol{y}$ with $\boldsymbol{z}_p$ we can equivalently write

$$a_k(\boldsymbol{p}, \boldsymbol{y}) = \gamma\varepsilon (\boldsymbol{p}, \boldsymbol{y})_{\mathbb{X}} + 2 \int_\Omega \boldsymbol{C}(\mathfrak{m} + \boldsymbol{\varphi}_k) \mathcal{E}(\boldsymbol{z}_p) : \mathcal{E}(\boldsymbol{z}_y). \tag{3.10}$$

We would like to mention that the C^2-regularity of j is not necessary for this definition of a_k.

Theorem 3.3. *The bilinear form a_k given in* (3.9) *fulfills the assumptions* (A7)–(A11).

Proof. Due to (AP) and (3.10) we have

$$a_k(\boldsymbol{p}, \boldsymbol{p}) \ge \gamma\varepsilon \|\boldsymbol{p}\|^2_{\mathbb{X}}.$$

Thus, (A7) and (A8) is fulfilled. Furthermore, (A9) holds due to

$$\begin{aligned} a_k(\boldsymbol{p},\boldsymbol{y}) &\le \gamma\varepsilon\|\boldsymbol{p}\|_{H^1}\|\boldsymbol{y}\|_{H^1} + C\|\boldsymbol{z}_p\|_{H^1}\|\boldsymbol{z}_y\|_{H^1} \\ &\le \gamma\varepsilon\|\boldsymbol{p}\|_{H^1}\|\boldsymbol{y}\|_{H^1} + C\|\boldsymbol{p}\|_{L^\infty}\|\boldsymbol{y}\|_{L^\infty} \le C\|\boldsymbol{p}\|_{\mathbb{X}\cap\mathbb{D}}\|\boldsymbol{y}\|_{\mathbb{X}\cap\mathbb{D}}. \end{aligned}$$

(A10) is proved as in Corollary 3.2.

Finally we prove (A11). For $\boldsymbol{y}_k \to 0$ and $\boldsymbol{p}_k \to \boldsymbol{p}$ in $\mathbb{X}$ we have $(\boldsymbol{y}_k, \boldsymbol{p}_k)_{\mathbb{X}} \to 0$ for $k \to \infty$. With $\boldsymbol{\varphi}_k \to \boldsymbol{\varphi}$, $\boldsymbol{p}_k \to \boldsymbol{p}$ in $\mathbb{D} = L^\infty(\Omega)^N$, and $S : L^\infty(\Omega)^N \to H^1(\Omega)^N$ continuously Fréchet differentiable, we have $\boldsymbol{u}_k = S(\boldsymbol{\varphi}_k) \to S(\boldsymbol{\varphi}) =: \boldsymbol{u}$ in H^1_D and $\boldsymbol{z}_{p_k} = S'(\boldsymbol{\varphi}_k)\boldsymbol{p}_k \to S'(\boldsymbol{\varphi})\boldsymbol{p} =: \boldsymbol{z}_p$ in H^1_D. In particular, the sequences are bounded in the corresponding norms, including $\|\boldsymbol{y}_k\|_{L^\infty} \le C$ if $\boldsymbol{y}_k \to \boldsymbol{y}$ weakly-* in L^∞. Using the Lipschitz continuity and boundedness of $\boldsymbol{C}'$ and $\nabla\boldsymbol{C}(\mathfrak{m}+\boldsymbol{\varphi})\mathcal{E}(\boldsymbol{z}_p) : \mathcal{E}(\boldsymbol{u}) \in L^1(\Omega)^N$ we have

$$\begin{aligned} &\left|\int_\Omega \boldsymbol{C}'(\mathfrak{m}+\boldsymbol{\varphi}_k)\boldsymbol{y}_k\mathcal{E}(\boldsymbol{z}_{p_k}) : \mathcal{E}(\boldsymbol{u}_k)\right| \\ &\quad\le \left|\int_\Omega (\boldsymbol{C}'(\mathfrak{m}+\boldsymbol{\varphi}_k) - \boldsymbol{C}'(\mathfrak{m}+\boldsymbol{\varphi}))\boldsymbol{y}_k\mathcal{E}(\boldsymbol{z}_{p_k}) : \mathcal{E}(\boldsymbol{u}_k)\right| \\ &\qquad + \left|\int_\Omega \boldsymbol{C}'(\mathfrak{m}+\boldsymbol{\varphi})\boldsymbol{y}_k\mathcal{E}(\boldsymbol{z}_{p_k} - \boldsymbol{z}_p) : \mathcal{E}(\boldsymbol{u}_k)\right| \\ &\qquad + \left|\int_\Omega \boldsymbol{C}'(\mathfrak{m}+\boldsymbol{\varphi})\boldsymbol{y}_k\mathcal{E}(\boldsymbol{z}_p) : \mathcal{E}(\boldsymbol{u}_k - \boldsymbol{u})\right| + \left|\int_\Omega \boldsymbol{C}'(\mathfrak{m}+\boldsymbol{\varphi})\boldsymbol{y}_k\mathcal{E}(\boldsymbol{z}_p) : \mathcal{E}(\boldsymbol{u})\right| \\ &\quad\le L\|\boldsymbol{\varphi}_k - \boldsymbol{\varphi}\|_{L^\infty}\|\boldsymbol{y}_k\|_{L^\infty}\|\boldsymbol{z}_{p_k}\|_{H^1}\|\boldsymbol{u}_k\|_{H^1} \\ &\qquad + \|\boldsymbol{C}'(\mathfrak{m}+\boldsymbol{\varphi})\|_{L^\infty}\|\boldsymbol{y}_k\|_{L^\infty}\|\boldsymbol{z}_{p_k} - \boldsymbol{z}_p\|_{H^1}\|\boldsymbol{u}_k\|_{H^1} \\ &\qquad + \|\boldsymbol{C}'(\mathfrak{m}+\boldsymbol{\varphi})\|_{L^\infty}\|\boldsymbol{y}_k\|_{L^\infty}\|\boldsymbol{z}_p\|_{H^1}\|\boldsymbol{u}_k - \boldsymbol{u}\|_{H^1} \\ &\qquad + \left|\int_\Omega (\nabla\boldsymbol{C}(\mathfrak{m}+\boldsymbol{\varphi})\mathcal{E}(\boldsymbol{z}_p) : \mathcal{E}(\boldsymbol{u}))\cdot\boldsymbol{y}_k\right| \to 0, \end{aligned}$$

which gives (A11). ◻

Hence with $0 < \lambda_{\min} \le \lambda_k \le \lambda_{\max}$, all assumptions of Theorem 2.2 are fulfilled and we get global convergence in the space $H^1(\Omega)^N \cap L^\infty(\Omega)^N$.

4. Numerical results. We discretize the structural topology optimization problem (3.1)–(3.3) using standard piecewise linear finite elements for the control $\boldsymbol{\varphi}$ and the state variable $\boldsymbol{u}$. The projection type subproblem (2.2) is solved by a primal dual active set (PDAS) method similar to the method described in [2, 22]. Many numerical examples for this problem can be found in [4, 6], e.g., for cantilever beams with up to three materials in two or three space dimensions and for an optimal material distribution within an airfoil. In [4] the choice of the potential ψ as an obstacle potential and the choice of the tensor interpolation $\boldsymbol{C}$ is discussed. Also the inner products $(.,.)_{\mathbb{X}}$ and $\gamma\varepsilon(.,.)_{\mathbb{X}}$ for fixed scaling parameter $\lambda_k = 1$ are compared, where both give rise to a mesh independent method and the latter leads to a large speed up. Note that the choice of $(.,.)_{\mathbb{X}}$ with $\lambda_k = (\gamma\varepsilon)^{-1}$ leads to the same iterates as choosing $\gamma\varepsilon(.,.)_{\mathbb{X}}$ and $\lambda_k = 1$. Furthermore, it is discussed in [4] that the choice of $\gamma\varepsilon(.,.)_{\mathbb{X}}$ can be motivated using $j''(\boldsymbol{\varphi})$ or by the fact that for the minimizers $\{\boldsymbol{\varphi}_\varepsilon\}_{\varepsilon>0}$ the Ginzburg–Landau energy converges to the perimeter as $\varepsilon \to 0$ and hence $\gamma\varepsilon\|\boldsymbol{\varphi}_\varepsilon\|^2_{\mathbb{X}} \approx$ const independent of $\varepsilon \ll 1$. However, since this holds only for the iterates $\boldsymbol{\varphi}_k$ when the phases are separated and the interfaces are present with thickness proportional to ε, we suggest adopting λ_k in accordance to this. As an updating strategy for λ_k

TABLE 1
Comparison of iteration numbers for $(.,.)_{L^2}$ and $(.,.)_{\mathbb{X}}$.

h	2^{-4}	2^{-5}	2^{-6}	2^{-7}	2^{-8}
$(.,.)_{L^2}$	323	5015	18200	57630	172621
$(.,.)_{\mathbb{X}}$	111	407	320	275	269

the following method is applied: Start with $\lambda_0 = 0.005(\gamma\varepsilon)^{-1}$, and then if $\alpha_{k-1} = 1$, set $\tilde{\lambda}_k = \lambda_{k-1}/0.75$, else $\tilde{\lambda}_k = 0.75\lambda_{k-1}$ and $\lambda_k = \max\{10^{-10}, \min\{10^{10}, \tilde{\lambda}_k\}\}$. The last adjustment yields that (A12) is fulfilled. This strategy is used for the following numerical results. Numerical experiments in [4] show that this in fact produces for the choice $(.,.)_{\mathbb{X}}$ a scaling with $\lambda_k \approx (\gamma\varepsilon)^{-1}$ for large k.

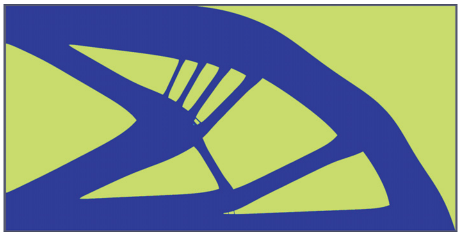
In [4, 6] the effect of obtaining various local minima of the nonconvex optimization problem (3.1)–(3.3) by choosing different initial guesses $\boldsymbol{\varphi}_0$ can be seen. However, the other parameters also have an influence.

In this paper we concentrate on comparing different choices of the inner products a_k and use for this the cantilever beam described in [4] with $\psi_0(\boldsymbol{\varphi}) = \frac{1}{2}(1-\boldsymbol{\varphi}\cdot\boldsymbol{\varphi})$ and a quadratic interpolation of the stiffness tensors $\boldsymbol{C}(\boldsymbol{\varphi})$. The computation are performed on a personal computer with 3GHz and 4GB RAM. First we discuss the choice of $(.,.)_{L^2}$ versus $(.,.)_{\mathbb{X}}$. The choice of the L^2-inner product leads to the commonly used projected L^2-gradient method. However, $(.,.)_{L^2}$ does not fulfill the assumptions of the VMPT method, since j is not differentiable in $L^2(\Omega)^N$ or $L^2(\Omega)^N \cap L^\infty(\Omega)^N$. Thus, global convergence is given for the discretized, finite dimensional problem but not in the continuous setting. This leads in contrast to the choice of $(.,.)_{\mathbb{X}}$ to mesh dependent iteration numbers for the L^2-gradient method, which can be seen in Table 1. The values in Table 1 were computed for different uniform mesh sizes h with the parameters $\varepsilon = 0.04$, $\gamma = 0.5$, $\boldsymbol{\varphi}_0 \equiv \mathfrak{m} = (0.5, 0.5)^T$, and $tol = 10^{-5}$ for the stopping criterion $\sqrt{\gamma\varepsilon}\|\nabla\boldsymbol{\varphi}_k\|_{L^2} \leq tol$. The behavior of iteration numbers is in accordance to our analytical results in function spaces considering $h \to 0$. Furthermore, the resulting values for λ_k not listed here show that we obtain for $(.,.)_{\mathbb{X}}$ and large k scalings $\lambda_k \approx (\gamma\varepsilon)^{-1}$ independent of the mesh parameter h, whereas the L^2-inner product produces λ_k scaled with h^2. Since the algorithm using the L^2-inner product is equivalent to the explicit time discretization of the L^2-gradient flow, i.e., of the Allen–Cahn variational inequality coupled with elasticity, with time step size $\Delta t = \lambda_k$, the scaling $\lambda_k = \mathcal{O}(h^2)$ reflects the known stability condition $\Delta t = \mathcal{O}(h^2)$ for explicit time discretizations of parabolic equations.

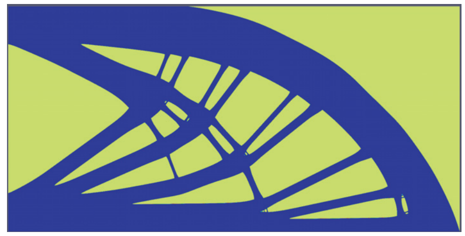
Next we compare $(.,.)_{\mathbb{X}}$ with a_k given in (3.9), which incorporates second order information. As an experiment we again use the cantilever beam in [4], now with $\gamma = 0.002$, $tol = 10^{-4}$ and random initial guess $\boldsymbol{\varphi}_0$ together with an adaptive mesh, which is fine on the interface. We use a nested approach in ε and h, where on the finest level $\varepsilon = 0.001$, $h_{\max} = 2^{-6}$, and $h_{\min} = 2^{-11}$ holds. The computational costs of one iteration with a_k given in (3.9) is significantly higher, since the calculation of $\mathcal{P}_k(\boldsymbol{\varphi}_k)$ requires the solution of a quadratic optimization problem with $\boldsymbol{\varphi} \in \Phi_{ad}$ and in addition with the linearized state equation (3.8) as constraints. However, in each PDAS iteration solving the subproblem for fixed k, only the right-hand side of (3.8) changes, namely, only $\boldsymbol{p}$. We factorize the matrix in the discrete equation once such that for each $\boldsymbol{p}$ only a cheap forward and backward substitution has to be done. In Table 2 the corresponding iteration numbers, the total CPU time, the values of the combined cost functional $j(\boldsymbol{\varphi}^*)$ as well as of its parts, i.e., the mean compliance, and the Ginzburg–Landau energy are listed. One observes the drastic reduction in

TABLE 2
Comparison of two different inner products.

inner product	iterations	CPU time	$j(\boldsymbol{\varphi}^*)$	$\int_{\Gamma_g} \boldsymbol{g} \cdot \boldsymbol{u}^*$	$E(\boldsymbol{\varphi}^*)$
$(.,.)_{\mathbb{X}}$	11189	42h 12min	15.07	15.03	20.79
a_k in (3.9)	851	19h	14.99	14.93	30.12



(a) $(.,.)_{\mathbb{X}}$.



(b) a_k given in (3.9).

FIG. 1. *Local minima for the cantilever beam.*

TABLE 3
Mesh independent iteration numbers for the H^1-BFGS method.

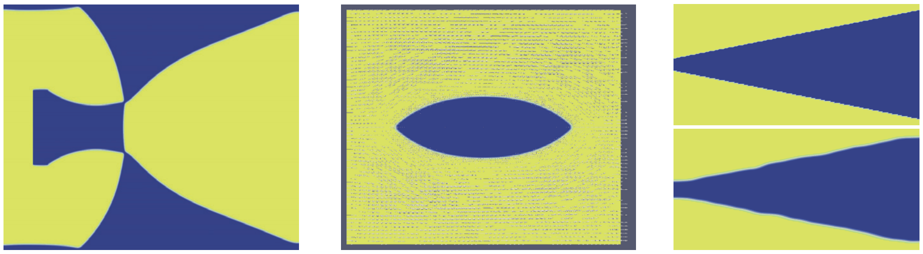
h	2^{-5}	2^{-6}	2^{-7}	2^{-8}	2^{-9}
H^1-BFGS iterations	85	88	86	85	85

iteration numbers using second order information. Due to the mentioned higher costs of calculating the search directions the total CPU time is only halved. Nevertheless, this can be possibly improved using a more sophisticated solver for $\mathcal{P}_k(\boldsymbol{\varphi}_k)$. It can be also observed that the cost $j(\boldsymbol{\varphi}^*)$ and the probably more interesting value of the mean compliance is lower. Hence, the different inner products result in different local minima, which are shown in Figure 1. The inner product given in (3.9) yields a finer structure. Also in other experiments we observed a local minima with lower cost value for this choice of a_k.

We also successfully applied an L-BFGS update in function spaces (see, e.g., [21] for the unconstrained case in Hilbert space) of the metric a_k, i.e., starting with $a_0(\boldsymbol{u},\boldsymbol{v}) = \gamma\varepsilon(\boldsymbol{u},\boldsymbol{v})_{\mathbb{X}}$ we use the update

$$a_{k+1}(\boldsymbol{u},\boldsymbol{v}) = a_k(\boldsymbol{u},\boldsymbol{v}) - \frac{a_k(\boldsymbol{p}_k,\boldsymbol{u})a_k(\boldsymbol{p}_k,\boldsymbol{v})}{a_k(\boldsymbol{p}_k,\boldsymbol{p}_k)} + \frac{\langle y_k,\boldsymbol{u}\rangle\,,\langle y_k,\boldsymbol{v}\rangle}{\langle y_k,\boldsymbol{p}_k\rangle}$$

in the case that $\langle y_k, \boldsymbol{p}_k\rangle > 0$, where $\boldsymbol{p}_k := \boldsymbol{\varphi}_{k+1} - \boldsymbol{\varphi}_k$ and $y_k := j'(\boldsymbol{\varphi}_{k+1}) - j'(\boldsymbol{\varphi}_k)$, which performs very well especially for small γ. Note that—as in the finite dimensional case—assumption (A8) cannot be shown for this sequence of inner products, but numerical experiments show that the discretized method is mesh independent, see Table 3, where the maximal recursion depth is set to 10 and the same cantilever beam example is used as for Table 1. A detailed comparison of the VMPT method with the often used gradient flow based solver (the Allen–Cahn or Cahn–Hilliard approach), which is also called the pseudo time stepping method (see, e.g., [10] for smooth potentials i.e., without box constraints on φ), can be found in [28]. We refer also to [33], where a pseudo time stepping scheme of Cahn–Hilliard type is applied. Their scheme needs up to 370,000 iterations to converge. Numerical studies on the local convergent methods, namely, the SQP-method and the semismooth Newton-approach, can be



(a) Crunching mechanism. (b) Obstacle minimizing drag. (c) Identified coefficient.

FIG. 2. *Successful applications of the VMPT method.*

found in [28]. In all these cases the VMPT method is at least competitive regarding numerical efficency, and in addition global convergence is shown.

Finally we present other successful applications of the VMPT method. The compliant mechanism problem

$$\min \; \frac{1}{2}\int_{\Omega_{obs}} (1-\varphi^N)|\boldsymbol{u}-\boldsymbol{u}_\Omega|^2 + \gamma E(\boldsymbol{\varphi}),$$

where the elasticity equation (3.2) and the constraints (3.3) have to hold, is more difficult. In our numerical analysis the solution process is more sensitive to the choice of a_k. Here the above H^1-BFGS approach enables us to solve the problem in an acceptable time. Until $\gamma\varepsilon\|\nabla v_k\|_{L^2} \leq \mathit{tol} = 10^{-4}$ the calculation of the material distribution in Figure 2(a) took 22 hours. It aims to crunch a nut in the middle of the left boundary when the force acts on the right-hand side from above and below and the mechanism is supplied on the left boundary; see [30].

Moreover, we also successfully applied the VMPT method on the following drag minimization problem of the Stokes flow using a phase field approach, which is analyzed in [17]:

$$\min \int_\Omega \frac{1}{2}|\nabla \boldsymbol{u}|^2 + \frac{1}{2}\alpha_\varepsilon(\varphi)|\mathbf{u}|^2 + \gamma E(\varphi),$$

$$\int_\Omega \alpha_\varepsilon(\varphi)\boldsymbol{u}\mathbf{v} + \int_\Omega \nabla\boldsymbol{u}\cdot\nabla\mathbf{v} = \mathbf{0} \quad \forall \mathbf{v}\in H^1_{0,div}(\Omega)$$

$$\mathbf{u}|_{\partial\Omega} \equiv (1,0)^T, \quad ⨍ \varphi = 0.75, \quad -1 \leq \varphi \leq 1.$$

We applied a nested approach in h and ε as well as an adaptive grid. As inner products we used the above H^1-BFGS method and obtained the result in Figure 2(b) with 188 iterations to obtain $\mathit{tol} = 10^{-3}$, which took 17 minutes.

A different type of optimization problem is the inverse problem for a discontinuous diffusion coefficient, where the discontinuous coefficient a is smoothed by a phase field approach and no mass conservation is used [9]:

$$\min \; \frac{1}{2}\int_\Omega |u-u_{obs}|^2 + \gamma E(\varphi)$$

$$\text{s.t.} \quad \int_\Omega a(\varphi)\nabla u\cdot\nabla\xi = \int_\Gamma g\xi \quad \forall \xi\in H^1 \quad \text{and} \quad \int_\Omega u = \int_\Omega u_{obs}, \quad -1\leq\varphi\leq 1.$$

We choose u_{obs} as solution of the state equation for φ shown in the upper part of Figure 2(c) with added noise of 5% and obtain the solution shown in the lower part of Figure 2(c).

The last three application examples are preliminary results and are under further studies. To our knowledge the VMPT method outperforms the existing applied optimization algorithms in these cases (see, e.g., [9, 18]).

REFERENCES

[1] H. W. Alt, *Lineare Funktionalanalysis: Eine anwendungsorientierte Einführung*, Springer, New York, 2012.

[2] M. Bergounioux, K. Ito, and K. Kunisch, *Primal-dual strategy for constrained optimal control problems*, SIAM J. Control Optim., 37 (1999), pp. 1176–1194.

[3] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, MA, 1999.

[4] L. Blank, H. M. Farshbaf-Shaker, H. Garcke, C. Rupprecht, and V. Styles, *Multi-material phase field approach to structural topology optimization*, in Trends in PDE Constrained Optimization, G. Leugering, P. Benner, S. Engell, A. Griewank, H. Harbrecht, M. Hinze, R. Rannacher, and S. Ulbrich, eds., Internat. Ser. Numer. Math. 165, Springer, New York, 2014, pp. 231–246.

[5] L. Blank, H. Garcke, H. M. Farshbaf-Shaker, and V. Styles, *Relating phase field and sharp interface approaches to structural topology optimization*, ESAIM Control Optim. Calc. Var., 20 (2014), pp. 1025–1058.

[6] L. Blank, H. Garcke, C. Hecht, and C. Rupprecht, *Sharp interface limit for a phase field model in structural optimization*, SIAM J. Control Optim., 54 (2016), pp. 1558–1584.

[7] L. Blank, H. Garcke, L. Sarbu, T. Srisupattarawanit, V. Styles, and A. Voigt, *Phase-field approaches to structural topology optimization*, in Constrained Optimization and Optimal Control for Partial Differential Equations, G. Leugering, S. Engell, A. Griewank, M. Hinze, R. Rannacher, V. Schulz, M. Ulbrich, and S. Ulbrich, eds., Internat. Ser. Numer. Math. 160, Springer, New York, 2012, pp. 245–256.

[8] B. Bourdin and A. Chambolle, *Design-dependent loads in topology optimization*, ESAIM Control Optim. Calc. Var., 9 (2003), pp. 19–48.

[9] K. Deckelnick, Ch. M. Elliott, and V. Styles, *Double obstacle phase field approach to an inverse problem for a discontinuous diffusion coefficient*, Inverse Problems, 32 (2016), 045008.

[10] L. Dedè, M. J. Borden, and T. J. R. Hughes, *Isogeometric analysis for topology optimization with a phase field model*, Arch. Comput. Methods Eng., 19 (2012), pp. 427–465.

[11] V. F. Demyanov and A. M. Rubinov, *Approximate Methods in Optimization Problems*, 2nd ed., Elsevier, New York, 1970.

[12] J. C. Dunn, *Newton's method and the Goldstein step-length rule for constrained minimization problems*, SIAM J. Control Optim., 18 (1980), pp. 659–674.

[13] J. C. Dunn, *Global and asymptotic convergence rate estimates for a class of projected gradient processes*, SIAM J. Control Optim., 19 (1981), pp. 368–400.

[14] J. C. Dunn, *On the convergence of projected gradient processes to singular critical points*, J. Optim. Theory Appl., 55 (1987), pp. 203–216.

[15] I. Ekeland and R. Témam, *Convex Analysis and Variational Problems*, SIAM, Philadelphia, 1999.

[16] E. M. Gafni and D. P. Bertsekas, *Convergence of a Gradient Projection Method*, LIDS-P-1201, MIT, Cambridge, MA, 1982.

[17] H. Garcke and C. Hecht, *A phase field approach for shape and topology optimization in Stokes flow*, in New Trends in Shape Optimization, A. Pratelli and G. Leugering, eds., Springer, New York, 2015, pp. 103–115.

[18] H. Garcke, M. Hinze, Chr. Kahle, and K. F. Lam, *Shape Optimization in Navier–Stokes Flow with Integral State Constraints Using a Phase Field Approach*, eprint, arXiv:1702.03855, 2017.

[19] M. Gawande and J. C. Dunn, *Variable metric gradient projection processes in convex feasible sets defined by nonlinear inequalities*, Appl. Math. Optim., 17 (1988), pp. 103–119.

[20] A. A. Goldstein, *Convex programming in Hilbert space*, Bull. Amer. Math. Soc., 70 (1964), pp. 709–710.

[21] W. A. Gruver and E. Sachs, *Algorithmic Methods in Optimal Control*, Res. Notes Math., Pitman, Boston, 1981.

[22] M. Hintermüller, K. Ito, and K. Kunisch, *The primal-dual active set strategy as a semismooth Newton method*, SIAM J. Optim., 13 (2002), pp. 865–888.

[23] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich, *Optimization with PDE Constraints*, Math. Model., Springer, New York, 2008.

[24] C. T. Kelley, *Iterative methods for optimization*, Front. Appl. Math., SIAM, Philadelphia, 1999.

[25] C. T. Kelley and E. W. Sachs, *Mesh independence of the gradient projection method for optimal control problems*, SIAM J. Control Optim., 30 (1992), pp. 477–493.

[26] E. S. Levitin and B. T. Polyak, *Constrained minimization methods*, Comput. Math. Math. Phys., 6 (1966), pp. 1–50.

[27] R. R. Phelps, *Metric projections and the gradient projection method in Banach spaces*, SIAM J. Control Optim., 23 (1985), pp. 973–977.

[28] C. Rupprecht, *Projection Type Methods in Banach Space with Application in Topology Optimization*, Ph.D. thesis, University of Regensburg, Germany, 2016.

[29] B. Rustem, *A class of superlinearly convergent projection algorithms with relaxed stepsizes*, Appl. Math. Optim., 12 (1984), pp. 29–43.

[30] O. Sigmund, *On the design of compliant mechanisms using topology optimization*, Mech. Struct. Mach., 25 (1997), pp. 493–524.

[31] F. Tröltzsch, *Optimal control of partial differential equations: Theory, methods, and applications*, Grad. Stud. Math., AMS, Providence, RI, 2010.

[32] M. Ulbrich, S. Ulbrich, and M. Heinkenschloss, *Global convergence of trust-region interior-point algorithms for infinite-dimensional nonconvex minimization subject to pointwise bounds*, SIAM J. Control Optim., 37 (1999), pp. 731–764.

[33] M. Y. Wang and S. Zhou, *Multinaterial structural topology optimization with a generalized Cahn–Hilliard model of multiphase transition*, Struct. Multidiscip. Optim., 33 (2007), pp. 89–111.