Downloaded from orbit.dtu.dk on: Apr 19, 2024

DTU Library

On reducing computational effort in topology optimization: We can go at least this far!

Limkilde, Asger; Evgrafov, Anton ; Gravesen, Jens

Published in:
Structural and Multidisciplinary Optimization

Link to article, DOI:
10.1007/s00158-018-2121-1

Publication date:
2018

Document Version
Peer reviewed version

Link back to DTU Orbit

Citation (APA):
Limkilde, A., Evgrafov, A., & Gravesen, J. (2018). On reducing computational effort in topology optimization: We can go at least this far! *Structural and Multidisciplinary Optimization*, *58*(6), 2481-2492. https://doi.org/10.1007/s00158-018-2121-1

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Structural and Multidisciplinary Optimization manuscript No.
(will be inserted by the editor)

On reducing computational effort in topology optimization: We can go at least this far!

Asger Limkilde · Anton Evgrafov · Jens Gravesen

Received: date / Accepted: date

Abstract In this work we attempt to answer the question posed in Amir O., Sigmund O.: On reducing computational effort in topology optimization: how far can we go? Struct. Multidiscip. Optim. 44(1), 25–29 (2011). Namely, we are interested in assessing how inaccurately we can solve the governing equations during the course of a topology optimization process while still obtaining accurate results. We consider this question from a "PDE-based" angle, using a posteriori residual estimates to gain insight into the behaviour of the residuals over the course of Krylov solver iterations. Our main observation is that the residual estimates are dominated by discretization error after only a few iterations of an iterative solver. This provides us with a quantitative measure for early termination of iterative solvers. We illustrate this approach using benchmark examples from linear elasticity, and demonstrate that the number of Krylov solver iterations can be significantly reduced, even when compared to previous heuristic recommendations, although each Krylov iteration becomes considerably more expensive.

Keywords Topology optimization · Approximation · Iterative solvers · A posteriori estimates

Asger Limkilde · Jens Gravesen
Department of Applied Mathematics and Computer Science
Technical University of Denmark
Asmussens Allé, Bygning 303B
2800 Kgs. Lyngby
Denmark
E-mail: {asgl, jgra}@dtu.dk

Anton Evgrafov
Department of Mathematical Sciences
Norwegian University of Science and Technology
N-7491 Trondheim
Norway
E-mail: anton.evgrafov@ntnu.no

1 Introduction

The authors of the note [4] have posed an interesting question: how inaccurately can we solve the governing equations during the course of a prototypical (topology) optimization algorithm without compromising its accuracy or convergence? The question itself implies the utilization of an approximate, typically iterative, solver for solving the state equations, which may be interrupted when we are satisfied with the achieved accuracy. As a result, the question can be rephrased in terms of finding an adequate stopping criterion for the state and adjoint equations solver.

In applied mathematics, and in particular numerical analysis, the question of inexactness is a fundamental one. A seminal paper [14] addresses this issue in the framework of Newton's methods and quantifies the accuracy with which the Newton's system has to be solved at each iteration without sacrificing fast (superlinear or quadratic) convergence of the method. This is achieved by an intuitively understandable strategy, which allows the accuracy to be rather lax far away from the solution, and increases it progressively as the method approaches the solution. This fundamental result provides computable stopping criteria to a variety of Newton–Krylov methods, where a Krylov solver (such as CG, GMRES, see [28, Chapter 6] for more details) is utilized for solving the linearized equations at each Newton iteration. In the context of PDE-constrained optimization this ultimately gives rise to efficient algorithms proudly bearing the names of Lagrange (for PDE constraint handling), Newton (for solving the optimality conditions), Krylov (for solving the linearized Newton subproblems) and Schur (for preconditioning the Krylov algorithm), see for example [8, 9, 26]. These algorithms simultaneously update all the unknowns involved in the problem:

design or control variables, state variables, and adjoint state/Lagrange multipliers for the PDE constraint.

For topology optimization, algorithms working exclusively in the design space of the problem while exactly or approximately maintaining the corresponding state and adjoint state remain exceedingly popular. Algorithms of this class, commonly referred to as the nested approach to topology optimization, treat solutions to the state and adjoint PDEs as implicit functions of the design variables, see [7] and references therein for more details. In this context the question of inexactness has also received its share of attention, as for large scale problems the process of repeatedly solving state and adjoint equations to compute objective function values and gradients by far dominates the computational effort.[1] Therefore, in order to reduce this effort, one can either (i) focus on increasing the efficiency of the linear solvers, or (ii) solve the state and adjoint systems less accurately.

In this work we will focus on the approach (ii), as several researchers did before us. For example, in a recent study Kočvara et al. [19] utilized inexact solutions of the linear algebraic systems arising in the application of interior point methods to topology optimization problems, naturally building on the recommendations outlined in [14]. The idea of inexact solving of the governing and adjoint equations has also been investigated in [3–5], where heuristic early termination criteria for a Krylov subspace solver for solving the state and adjoint equations have been proposed and studied. The authors of these works clearly demonstrate that seemingly very rough approximate solutions obtained after very few iterations of a Krylov solver are often sufficient for successfully solving topology optimization problems. In fact, it is sometimes sufficient to use only a small fixed number of Krylov solver iterations, provided that one uses an approximate solution obtained at the previous optimization iterations as a starting guess for the Krylov solver. One may argue that as subsequent linear systems become progressively closer to each other as (or indeed, if) the optimization process converges, the accuracy of objective function and gradient evaluations will also progressively increase as we get closer to an optimum, as even a fixed number of Krylov iterations will in such a case "add up" over a range of optimization iterations. This is also a line of reasoning, which is compatible with that in [14].

[1] This effect is a consequence of the utilization of only first order information about the optimization problem, which results in computationally inexpensive iterations at the cost of slow rates of convergence. Whereas in some situations significantly faster converging alternatives exist, see for example [16,19,27], we focus on first order algorithms in this work.

Notice that the previously mentioned studies take the discretized optimization problem as their sole object of interest, and focus on reducing the computational effort needed to solve these finite-dimensional problems. Arguably, we solve the discretized problem not because this is the problem we *want* to solve, but because this is the problem we *can* solve on a computer, while hoping that the resulting solution approximates an optimal solution to the *original* infinite-dimensional PDE-constrained optimization problem. Therefore, owing to the inherent presence of the discretization error in the problem we are solving, it does not make sense to solve it more accurately than what is warranted by the discretization error estimates. Such a line of thinking is a de-facto standard in the numerical analysis community, even though the inclusion of rigorous discretization error estimates into the computational algorithms is often a highly non-trivial task. For example in [6] (see also references therein), termination of a preconditioned conjugate gradient (PCG) iteration for solving linear systems resulting from a finite element discretization of an elliptic PDE is related to certain a priori and a posteriori discretization error estimates. An easily computable stopping criterion for the PCG algorithm is proposed based on quantities which are already computed during the PCG iteration. Our approach is close to that in [6] in the sense that we try to relate the termination of the solution algorithm for the discretized problem to the discretization error estimates. Namely, we utilize certain classical a posteriori finite element method (FEM) error estimates[2] in order to quantify the convergence of the PCG iteration at a given discretization level thereby gaining insight into the phenomenon outlined in the previous paragraph and attempting to give a more quantitative answer to the question posed in [4]. To this end we use the same iterative method as in [4], and monitor the PDE residual estimate as the Krylov solver progresses. We observe that already after a few iterations the residual estimate is dominated by the discretization error. On the one hand, this finding is not unexpected given the fact that the finite element discretization of the PDEs with discontinuous coefficients may approximate the solution to the said PDE rather poorly, see for example [23]. On the other hand this effect might be the reason for the successful performance of the early termination strategies utilized in for example [4]: indeed, from the PDE-constrained optimization perspective, these approximate solutions are as good as what can be achieved at a given discretization level.

With the primary objective of gaining insight into the premature termination of iterative linear solvers

[2] A posteriori FEM error estimates have recently been used in topology optimization in another context [24].

in the topology optimization framework and not constructing an efficient computational algorithm, we have utilized a posteriori error estimates, whose evaluation by far dominates the computational effort of a standard PCG iteration. Namely, in addition to performing a matrix-vector multiplication, a few vector-vector computations, and a preconditioning operation, an a posteriori FEM error analysis has to be carried out, which is in stark contrast to the method employed in [6]. With this disclaimer, we believe that the present study provides an interesting perspective into why previously utilized heuristic early termination criteria for iterative linear solvers within topology optimization may have been successful. Furthermore, this line of thinking clearly demonstrates potential for resulting in efficient numerical methods when combined with computationally inexpensive ways of connecting the PCG iteration with discretization error estimates, either through the ideas described in [6] or by utilizing alternative FEM discretizations naturally equipped with a posteriori estimates [10, 12, 18].

The article is organized as follows. In Section 2 and 3 we briefly introduce the problem formulation and the computational framework. In Section 4 and 5 we describe the residual estimate used in this work and investigate its behaviour during the course of a Krylov subspace solver. In Section 6 the proposed method is applied to some classical 2D and 3D benchmark examples from linear elasticity. We conclude the paper with a discussion of the results in Section 7.

2 Problem formulation

For the purpose of keeping our discussion close to that in [4] we mostly follow the setup used in the cited work, with some trivial changes related to the implementation choices (for example, we consider simplicial grids instead of those composed of quadrilaterals/hexahedrons). Therefore we consider topology optimization problems in linear elasticity, namely that of minimizing the compliance of elastic bodies, and that of designing compliant mechanisms. We use the density based approach and the SIMP interpolation scheme, for details see [7]. The material distribution is updated on the basis of optimality criteria (OC) fixed point iteration. We use PDE-based density filtering for regularizing this problem, see [20].

Let us now fix the notation. We consider a bounded domain $\Omega \subset \mathbb{R}^d$, $d \in \{2, 3\}$, with Lipschitz boundary $\Gamma = \partial\Omega$. Given a positive number $\widehat{r} > 0$ the filtering (compact, linear) operator $\mathcal{F}_{\widehat{r}} : L^2(\Omega) \to L^2(\Omega)$, $\varrho = \mathcal{F}_{\widehat{r}}\rho$, is defined as the solution operator for the following variational problem (weak formulation of the Helmholtz PDE with homogeneous Neumann boundary conditions):

$$\widehat{r}^2 \int_\Omega \nabla\varrho(x) \cdot \nabla\tilde{\varrho}(x)\,\mathrm{d}x + \int_\Omega \varrho(x)\tilde{\varrho}(x)\,\mathrm{d}x = \int_\Omega \rho(x)\tilde{\varrho}(x)\,\mathrm{d}x, \quad \forall\tilde{\varrho} \in H^1(\Omega), \tag{1}$$

where we identify the solution $\varrho \in H^1(\Omega)$ with a function in $L^2(\Omega)$ through the (compact) embedding $H^1(\Omega) \subset L^2(\Omega)$.

To describe the mechanical system we introduce a (Hilbert) space where its displacements will be sought, $V = \{\, \boldsymbol{v} \in H^1(\Omega;\mathbb{R}^d) \mid \gamma_D(\boldsymbol{v}) = 0 \,\}$, which is the restriction of the standard Sobolev space $H^1(\Omega;\mathbb{R}^d)$ to only those functions satisfying the prescribed homogeneous Dirichlet boundary conditions, the fact which we denote by $\gamma_D(\boldsymbol{v}) = 0$. The principle of virtual work (weak formulation of the linearized elasticity equations) will be stated with the help of a ϱ-parametrized symmetric, bounded and coercive bilinear form $a_\varrho : V \times V \to \mathbb{R}$ and a bounded linear functional $\ell \in V'$[3]:

$$\begin{aligned} a_\varrho(\boldsymbol{u},\boldsymbol{v}) &= \int_\Omega \varrho(x)^p \boldsymbol{c}_0\boldsymbol{\epsilon}(\boldsymbol{u}(x)) : \boldsymbol{\epsilon}(\boldsymbol{v}(x))\,\mathrm{d}x, \\ \ell(\boldsymbol{v}) &= \int_\Omega \boldsymbol{f}(x)\cdot\boldsymbol{v}(x)\,\mathrm{d}x + \int_\Gamma \boldsymbol{t}(x)\cdot\boldsymbol{v}(x)\,\mathrm{d}x, \end{aligned} \tag{2}$$

where $\boldsymbol{\epsilon}(\boldsymbol{u}) = [\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^{\mathrm{T}}]/2$ is the linearized strain tensor, and $\boldsymbol{f} \in L^2(\Omega,\mathbb{R}^d)$ and $\boldsymbol{t} \in L^2(\Gamma,\mathbb{R}^d)$ are the volumetric and traction forces,[4] respectively. The fourth order positive definite tensor $\boldsymbol{c}_0$ with the usual symmetries encapsulates the linear stress-strain relationship for the "stiff" material (stiffness tensor, Hooke's law), and the number $p > 1$ is the penalization constant in the SIMP material law (in all our experiments $p = 3$ is used).

The principle of virtual work in this notation can be written rather succinctly as a linear equation

$$\boldsymbol{\mathcal{K}}_\varrho\boldsymbol{u} = \ell, \tag{3}$$

where the bounded linear operator $\boldsymbol{\mathcal{K}}_\varrho : V \to V'$ is defined by the relation $(\boldsymbol{\mathcal{K}}_\varrho\boldsymbol{u})[\boldsymbol{v}] = a_\varrho(\boldsymbol{u},\boldsymbol{v})$. Lax–Milgram theorem together with Korn's inequality guarantees the existence of a bounded inverse operator $\boldsymbol{\mathcal{K}}_\varrho^{-1} : V' \to V$ provided that the Dirichlet boundary conditions defining V do not permit any rigid body modes.

[3] We denote the space of all bounded linear functionals on V (the dual space of V) by V'.

[4] Strictly speaking, traction forces only enter the model on parts of the boundary, where the displacements are not fixed. We only assume that $\boldsymbol{t}$ is defined on the whole boundary and its components are set to zero on Dirichlet parts of the boundary in order to keep the notation to a bare minimum.

Finally the topology optimization problem we will be interested in solving can be stated as

$$\begin{aligned} \underset{\rho\in L^2(\Omega)}{\text{minimize}}\ J(\rho) =& i(\boldsymbol{\mathcal{K}}^{-1}_{\mathcal{F}_{\hat{r}}\rho}\ell), \\ \text{s.t.} \qquad & V(\rho) = \int_\Omega \rho(x)\,\mathrm{d}x \le V_0, \\ & 0 < \rho_{\min} \le \rho(x) \le 1, \quad \text{a.e. in } \Omega, \end{aligned} \tag{4}$$

where $i \in V'$, $\rho_{\min} \in (0,1)$, and $V_0 > \rho_{\min}|\Omega|$ are given. The minimum compliance example is obtained by setting $i = \ell$, while the force inverter mechanism example is obtained by putting $i(\boldsymbol{v}) = \int_\Gamma \boldsymbol{i}_\Gamma(x)\cdot\boldsymbol{v}(x)\,\mathrm{d}x$, where $\boldsymbol{i}_\Gamma(x) = \pm\boldsymbol{e}_j\chi_{\Gamma_{\text{out}}}(x)$. In the last equation, Γ_{out} is the part of the boundary where the j^{th} displacement is to be either minimized or maximized (depending on the sign $\pm$). Thus $\boldsymbol{e}_j$ is the j^{th} canonical basis vector in $\mathbb{R}^d$, and $\chi_{\Gamma_{\text{out}}}$ is the characteristic function of the set Γ_{out}. The gradient of J with respect to ρ is classically computed as

$$\nabla_\rho J = -\mathcal{F}_{\hat{r}}[p\rho^{p-1}\boldsymbol{c}_0\boldsymbol{\epsilon}(\boldsymbol{u}_\rho):\boldsymbol{\epsilon}(\boldsymbol{u}^{\text{a}}_\rho)], \tag{5}$$

where $\boldsymbol{u}_\rho = \boldsymbol{\mathcal{K}}^{-1}_{\mathcal{F}_{\hat{r}}\rho}\ell$ and $\boldsymbol{u}^{\text{a}}_\rho = \boldsymbol{\mathcal{K}}^{-1}_{\mathcal{F}_{\hat{r}}\rho}i$ are the solutions to the state and adjoint problems. This gradient is then utilized in the OC update scheme in this work, but of course could also be used in the context of any other non-linear constrained optimization algorithm. For the sake of completeness we state the OC update scheme for minimizing compliance. Given a material distribution ρ_k we first compute $\boldsymbol{u}_{\rho_k} = \boldsymbol{u}^{\text{a}}_{\rho_k}$ by (approximately) solving the elasticity equations, and then the corresponding gradient $\nabla_\rho J_k$ from (5). The new material distribution is defined by

$$\rho_{k+1} = \pi_{S_k}[\rho_k(-\nabla_\rho J_k/\lambda_{k+1})^\xi],$$

where $\pi_{S_k}[\cdot]$ is the projection operator (in this case, point-wise, owing to the simple bound structure of the set) onto the closed, convex, and non-empty set $S_k = \{\,\rho \in L^\infty(\Omega) \mid \max\{0,(1-\gamma)\rho_k\} \le \rho \le \min\{1,(1+\gamma)\rho_k\}\,\}$, $\gamma > 0$ and $\xi \in (0,1)$ are trust-region like and damping parameters, respectively. Finally, λ_{k+1} is computed by finding the root of the equation $\int_\Omega \rho_{k+1}(x)\,\mathrm{d}x = V_0$ using, for example, the bisection algorithm. For more details see [7].

3 Problem discretization

In short, we use standard finite element discretizations of all PDEs involved in the problem, that is, (1)–(3) and the adjoint of the latter. Implementations are done using the high-level finite element library FEniCS [21]. Notationally, we will distinguish the discretized quantities from their continuous counterparts by adding the index h, where $h > 0$ is some characteristic size of the elements in our mesh. In particular, the finite dimensional piece-wise polynomial space for approximating displacements will be denoted by $V_h \subset V$. We then naturally put $\ell_h = \ell|_{V_h} \in V_h'$, $i_h = i|_{V_h} \in V_h'$, and similarly define $\boldsymbol{\mathcal{K}}_{\varrho,h} : V_h \to V_h'$ by restricting the bilinear form a_ϱ to $V_h \times V_h$.

In this notation the main question posed in the introduction section is when to stop an iterative solver used to approximately solve the state and adjoint systems $\boldsymbol{\mathcal{K}}_{\varrho,h}\boldsymbol{u}_h \approx \ell_h$ and $\boldsymbol{\mathcal{K}}_{\varrho,h}\boldsymbol{u}^{\text{a}}_h \approx i_h$. For solving these systems we utilize the preconditioned conjugate gradient method (PCG), preconditioned in exactly the same fashion as done in [4] to make the direct comparison possible.

4 Residuals and their evaluation

At every optimization iteration, given some current material distribution ρ_h and its smoothed version $\varrho_h = \mathcal{F}_{h,\hat{r}}\rho_h$ we need to approximately solve the discretized elasticity equations $\boldsymbol{\mathcal{K}}_{\varrho_h,h}\boldsymbol{u}_h \approx \ell_h$ and $\boldsymbol{\mathcal{K}}_{\varrho_h,h}\boldsymbol{u}^{\text{a}}_h \approx i_h$ in order to evaluate the objective function value and its gradient, see (4) and (5).[5] As in [3–5] the approximate solutions $\boldsymbol{u}_h$, $\boldsymbol{u}^{\text{a}}_h$ will be found by prematurely terminating an iterative linear solver, which, unless specified otherwise, is CG preconditioned with the stiffness matrix [28, Chapter 9] corresponding to a solid design (i.e., $\rho \equiv 1$), exactly as done in [4]. Nevertheless the discussion that follows is independent from the choice of a specific preconditioner. We also focus our discussion on approximately solving the state equations $\boldsymbol{\mathcal{K}}_{\varrho_h,h}\boldsymbol{u}_h \approx \ell_h$ but provide the numerical illustrations for both the state and adjoint equations. One of the main objects in this discussion is the residual of (3), which for a given approximation $\boldsymbol{u}_h$ is defined by $r_{\varrho_h,\boldsymbol{u}_h} = \ell - \boldsymbol{\mathcal{K}}_{\varrho_h}\boldsymbol{u}_h$. We will aim at estimating this residual with the objective of keeping the discretization error $\|\boldsymbol{u}_h - \boldsymbol{\mathcal{K}}^{-1}_{\varrho_h}\ell\|_V$ small.

When applied to our discretized elasticity system in exact arithmetics, PCG will finitely terminate with the solution $\boldsymbol{u}_{h,\text{exact}} = \boldsymbol{\mathcal{K}}^{-1}_{\varrho_h,h}\ell_h$. Even this ideal solution to the discretized system, however, does not usually solve the "continuous" elasticity equations corresponding to the material distribution ϱ_h, i.e., $r_{\varrho_h,\boldsymbol{u}_{h,\text{exact}}} \neq 0$. Indeed, we can only hope that the residual $r_{\varrho_h,\boldsymbol{u}_{h,\text{exact}}}$ converges to zero as we keep refining the mesh. Therefore, according to conventional wisdom there is no reason to "oversolve" the discretized equations, and one should stop when the error (or residual) is dominated

[5] We are not concerned about repeatedly solving the filtering problem, because it is a relatively well-conditioned one, see [20].

by the terms related to the discretization errors, which cannot be eliminated by solving the discretized system more accurately.

Given some material distribution ϱ_h and approximate solution to the discretized elasticity equations $\boldsymbol{u}_h \in V_h$ the residual $r_{\varrho_h,\boldsymbol{u}_h}$, being a linear bounded functional over V, is somewhat difficult to measure and work with. Indeed, by definition

$$\begin{aligned}\|r_{\varrho_h,\boldsymbol{u}_h}\|_{V'} &= \sup\left\{|r_{\varrho_h,\boldsymbol{u}_h}(\boldsymbol{v})| \;\middle|\; \boldsymbol{v}\in V, \|\boldsymbol{v}\|_V \le 1\right\} \\ &= \sup\left\{|\ell(\boldsymbol{v}) - a_{\varrho_h}(\boldsymbol{u}_h,\boldsymbol{v})| \;\middle|\; \boldsymbol{v}\in V, \|\boldsymbol{v}\|_V \le 1\right\}.\end{aligned} \tag{6}$$

There are however many ways of estimating this quantity, see for example [15, Chapter 10]. For the purpose of getting insight about the behaviour of the residuals both over the course of the topology optimization iterations and over the course of each application of PCG we utilize the following, arguably computationally expensive, approach, while noting that other alternatives are possible [15, Chapter 10].

Let $r \in V'$ be an arbitrary functional. Riesz representation theorem provides us with a constructive approach for computing $\|r\|_{V'}$. Namely, first we compute a function $\boldsymbol{e} \in V$ by solving the variational problem

$$\langle \boldsymbol{e}, \tilde{\boldsymbol{v}}\rangle_V = r(\tilde{\boldsymbol{v}}), \quad \forall \tilde{\boldsymbol{v}} \in V, \tag{7}$$

where $\langle\cdot,\cdot\rangle_V$ denotes the inner product in V. We then use the fact that $\|r\|_{V'} = \|\boldsymbol{e}\|_V$. Note that the latter norm is simply a positive Sobolev norm, thus its computation amounts to integration.

Of course numerically we need to discretize the problem (7), for example using FEM. It is tempting to use V_h in place of V in (7), but then the right hand side of the problem is identically 0 (Galerkin orthogonality condition) for $r_{\varrho_h,\boldsymbol{u}_{h,\text{exact}}}$, and we will not learn anything beyond what the original discretization of the elasticity problem has provided us with. In other words, we need to work with a richer space than V_h when discretizing (7), in order to capture some of the discretization error. Here we use uniform h-refinement to compute such a space, using the standard simplicial mesh refinement algorithm implemented in FENICS [25]. The piece-wise polynomial space of the same order as V_h obtained after one iteration of uniform refinement will be denoted by $V_{\frac{h}{2}}$.[6] The solution to (7) with $V_{\frac{h}{2}}$ instead of V will be denoted by $\boldsymbol{e}_{\frac{h}{2}} \in V_{\frac{h}{2}}$ and the resulting norm will be denoted by

$$|||r|||_{\frac{h}{2}} \overset{\text{def}}{=} \|\boldsymbol{e}_{\frac{h}{2}}\|_V = \|r\|_{V'_{\frac{h}{2}}}.$$

[6] In the numerical experiments we also try to refine the mesh twice, that is to use $V_{\frac{h}{4}}$ in place of $V_{\frac{h}{2}}$. The results that we obtain are essentially the same, see Fig. 1 and 3.

Note that (7) is not a very difficult problem to solve: it corresponds to an elliptic PDE with constant coefficients and is easily solvable by algebraic multigrid (on unstructured grids) or even fast spectral methods (on structured grids). Additionally, only the right-hand side of this problem changes with ϱ_h and $\boldsymbol{u}_h$. Nevertheless, it is of course computationally burdensome to solve this problem at every PCG iteration, as we plan to do next.

5 Behaviour of a posteriori residual estimates

In view of the previous discussion it would be reasonable to stop PCG iterations when no further improvement of the residual $\|\ell - \boldsymbol{\mathcal{K}}_{\varrho_h}\boldsymbol{u}_h^{(i)}\|_{V'} \approx |||\ell - \boldsymbol{\mathcal{K}}_{\varrho_h}\boldsymbol{u}_h^{(i)}|||_{\frac{h}{2}}$ is observed, where $\boldsymbol{u}_h^{(i)} \in V_h$ is the approximation to the solution of the discretized elasticity equations at i-th iteration of PCG algorithm. When this happens we conclude that the residual is dominated by discretization error and therefore it is time to stop the PCG algorithm. To this end we monitor the relative norm of the residual $\eta_i = |||\ell - \boldsymbol{\mathcal{K}}_{\varrho_h}\boldsymbol{u}_h^{(i)}|||_{\frac{h}{2}} / |||\ell|||_{\frac{h}{2}}$ during the iterations of the PCG algorithm, and stop when the relative change in η_i is smaller than some prescribed tolerance ε_η, that is when $|\eta_i - \eta_{i-1}|/\eta_i < \varepsilon_\eta$.

The behaviour of the a posteriori residual estimate during the course of typical PCG iterations is illustrated in Fig. 1. Here we plot η_i as a function of the PCG iteration number i for the elasticity system arising when solving the example described in Subsection 6.1. We consider two linear systems appearing at optimization iterations 3 and 35, as well as two possible initializations of PCG, namely starting either from the zero vector or from the "exact" solution $\boldsymbol{u}_{k-1}$ to the linear system appearing at the previous optimization iteration, which we compute using a direct solver. It is seen that when starting from the zero vector the relative dual norm η is dominated by the discretization error after only 10 or 15 PCG iterations. In practice the solution from the previous optimization step $\boldsymbol{u}_{k-1}$ is often used as a starting guess, since the linear systems corresponding to two consecutive optimization cycles are nearly identical when changes in the design are small. When using such a starting guess we see that for iteration 3 the residual is now dominated by discretization error after fewer that 10 PCG iterations. When considering the optimization iteration 35, which corresponds to the case where design changes are small, it is seen that the residual is already dominated by the discretization error from iteration 1. This means that, with respect to the measure of accuracy that we utilize, the solution from iteration 34 is a good enough approximate solution to the discretized elasticity equations at

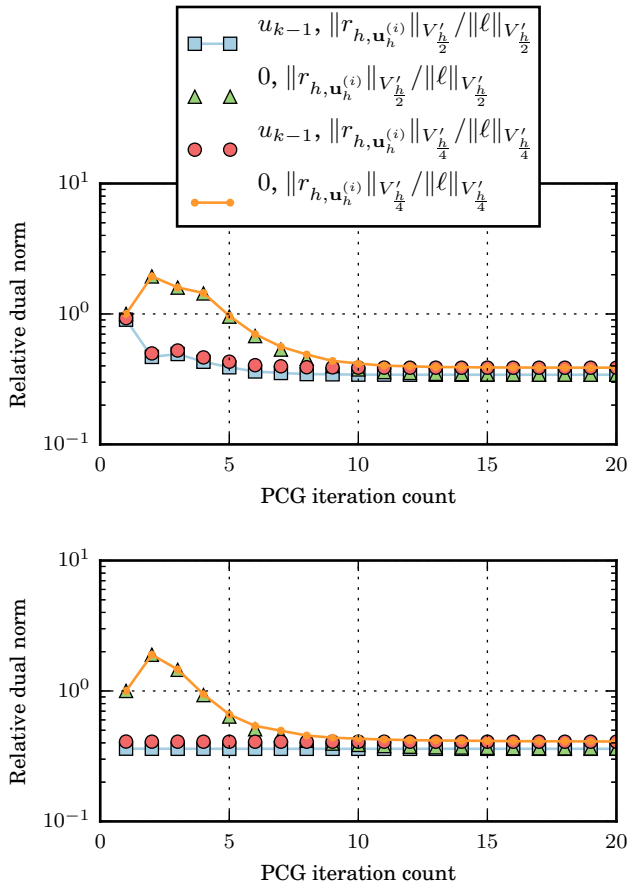

Fig. 1 The relative dual norm of the residual for two different designs. The legend indicates the starting guess used, and the error estimator considered. The error estimates differ in the number of times the mesh is refined, when approximating the dual norm. *Top*: Using the design from optimization iteration 3; *bottom*: Using the design from optimization iteration 35; see Subsection 6.1 for details of the example.

iteration 35, which cannot be improved upon without refining the grid or enhancing the discretization in some other fashion.

We would like to point out that the stopping criterion we utilize can be trivially satisfied if the PCG method stagnates and fails to make any significant progress towards the solution of the discretized system from one iteration to the next. As a consequence of this, whereas we do not rely on the utilization of any specific preconditioner, it should be of reasonable quality to prevent such situations from happening. (One could of course also tighten the tolerance in the stopping criterion, or measure the progress over the course of more than one iteration — but this would only deal with the symptoms and not the issue.)

Another issue worth mentioning here is the relation between the residual, measured as $\|\ell - \boldsymbol{\mathcal{K}}_\varrho \boldsymbol{u}\|_{V'}$, and the corresponding displacement error $\|\boldsymbol{u} - \boldsymbol{\mathcal{K}}_\varrho^{-1}\ell\|_V$. Whereas the two are equivalent (as a result of symmetry, coercivity, and continuity of the bilinear form $a_\varrho(\cdot,\cdot)$) with equivalence constants involving the norms of $\boldsymbol{\mathcal{K}}_\varrho$ and its inverse, see [11, Section 2.3], in practice small residual is not a guarantee of the small error. Indeed, the latter estimate involves the terms inversely proportional to $\rho_{\min}^p$, see (4). Since this constant is extremely small in our case (value of 10^{-9} is commonly used), the boundedness of the error with respect to the residual is effectively lost. However, spatially, this breakdown happens primarily in the regions where we are not interested in solving the elasticity equations in the first place, that is, the regions occupied by "the approximate void" ersatz material. Furthermore, contributions to the objective function and the gradient (5) from these areas are relatively small (they are scaled with ϱ^p or ϱ^{p-1}), and therefore even large errors in these areas are likely to be acceptable from the optimization algorithm's perspective.

Finally, one can of course utilize different equivalent norms on V, which in turn induce norms on V' through (6). One natural example, which we have tried, is the energy norm $\|\boldsymbol{v}\|_{E,\varrho}^2 = a_\varrho(\boldsymbol{v},\boldsymbol{v})$, which is the error norm minimized by the non-preconditioned version of CG for every linear system, or by the Galerkin method at every discretization level. On the one hand, utilizing this norm directly results in systems (7), which are refined versions of the elasticity system we try to solve, vary from one topology optimization to another, and clearly are no longer associated with a system of PDEs with constant coefficients. With this norm, we have seen no clear improvement or deterioration in the obtained results or behaviour of the method, but the system (7) becomes completely impractical to solve. On the other hand, we have the Galerkin orthogonality

$$\begin{aligned}\|\boldsymbol{u}_h - \boldsymbol{\mathcal{K}}_{\varrho_h}^{-1}\ell\|_{E,\varrho_h}^2 \\ = \underbrace{\|\boldsymbol{u}_h - \boldsymbol{\mathcal{K}}_{h,\varrho_h}^{-1}\ell_h\|_{E,\varrho_h}^2}_{\text{linear solver error}} + \underbrace{\|\boldsymbol{\mathcal{K}}_{h,\varrho_h}^{-1}\ell_h - \boldsymbol{\mathcal{K}}_{\varrho_h}^{-1}\ell\|_{E,\varrho_h}^2}_{\text{discretization error}},\end{aligned}$$

and the identity

$$\begin{aligned}\|\boldsymbol{u}_h - \boldsymbol{\mathcal{K}}_{h,\varrho_h}^{-1}\ell_h\|_{E,\varrho_h}^2 = a_{\varrho_h}(\boldsymbol{u}_h,\boldsymbol{u}_h) - 2\ell_h(\boldsymbol{u}_h) \\ + \underbrace{\|\boldsymbol{\mathcal{K}}_{h,\varrho_h}^{-1}\ell_h\|_{E,\varrho_h}^2}_{\text{indep. from } \boldsymbol{u}_h}.\end{aligned}$$

Therefore, convergence of the first two terms in the last equality can be utilized as a computationally "free" indicator of convergence of the linear solver with respect to the energy norm instead of η. However, unlike the actual residual or error estimate, the limiting value of $a_{\varrho_h}(\boldsymbol{u}_h,\boldsymbol{u}_h) - 2\ell_h(\boldsymbol{u}_h)$ varies with the exact solution of the linear system $\boldsymbol{\mathcal{K}}_{h,\varrho_h}^{-1}\ell_h$ and does not go to 0 when $h \to 0$. Therefore selecting the tolerance

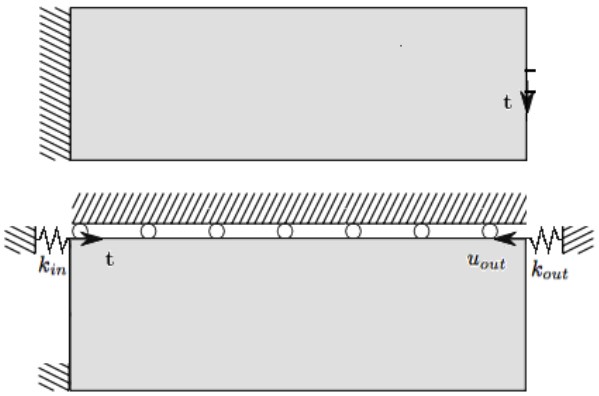

Fig. 2 *Top*: Sketch of the 2D cantilever example. *Bottom*: Sketch of the 2D force inverting mechanism example.

for this convergence indicator, which should work independently from the optimization iteration and mesh discretization is a non-trivial task. Our final comment here is that a very similar criterion based on the value of $a_{\varrho_h}(\boldsymbol{u}_h, \boldsymbol{u}_h) - \ell_h(\boldsymbol{u}_h)$, which is precisely the residual applied to $\boldsymbol{u}_h$, is proposed by Amir et al. in [5]. This term will go to 0 as the iterative solver converges, and it does so faster than the residual, thus making it a suitable candidate for terminating the solver prematurely [5].

6 Numerical examples

In this section we solve a few benchmark optimization problems while utilizing the PCG stopping criterion described in the previous section. In all numerical experiments we utilize "warm start" of the PCG iteration, that is we initialize it with the approximate solution obtained at the previous optimization iteration. For 3D numerical experiments we visualize designs by extracting the isosurface and rendering it in Paraview [1].

6.1 Example 1: 2D cantilever

We consider the problem which is sketched in the top of Fig. 2. We utilize a small grid with 180×60 rectangular elements, a filter radius of $\widehat{r} = 0.03$ and set the volume fraction $V_0 = 0.5$. The objective is to minimize the compliance and in this case the problem is self-adjoint, hence we have $i = \ell$. We consider the proposed stopping criterion with tolerance $\varepsilon_\eta = 10^{-4}$. In Fig. 3 we show the number of PCG iterations necessary to meet the proposed stopping criterion during the optimization process. For comparison, we also plot the number of iterations necessary to meet a "standard" criterion

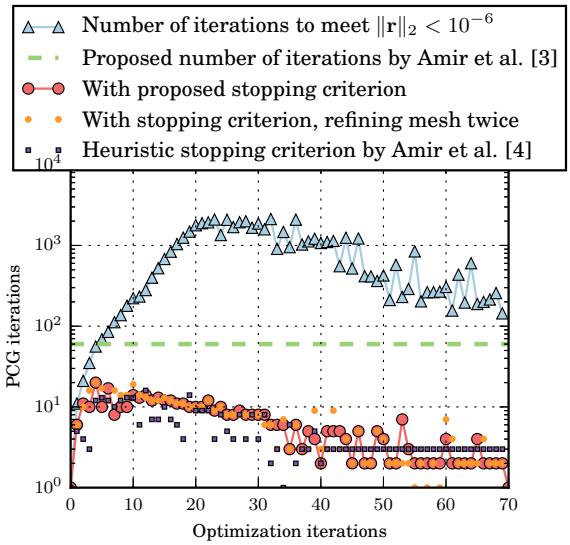

Fig. 3 Number of PCG iterations for the 2D cantilever example.

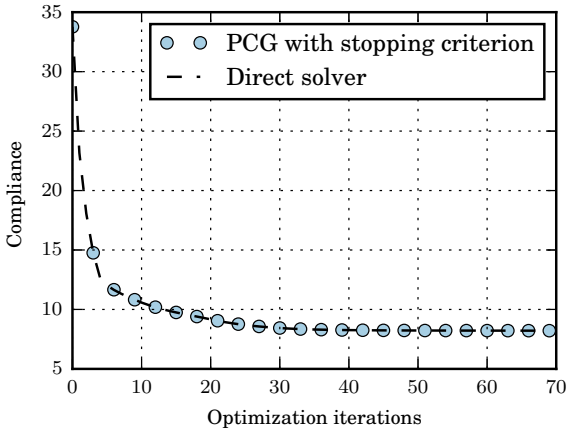

Fig. 4 The objective function (compliance) during the optimization process based on early termination of PCG vs. utilizing a direct solver; 2D cantilever example.

on the residual vector[7] $\|\boldsymbol{r}\|_2 < 10^{-6}$, with this preconditioner, as well as the (iteration-independent) quantity $\min\{n_x, n_y\}$[8] for a regular $n_x \times n_y$ grid discretizing the problem. Finally we also show the number of iterations necessary to reach yet another heuristic stopping criterion in [5], which is given by $a_\varrho(u_h^{(i)}, u_h^{(i)}) - \ell(u_h^{(i)}) \leq \tau$, and we put $\tau = 10^{-4}$. From this plot we can conclude that qualitatively the number of PCG iterations based on the heuristic idea in [5] correlates very well with what we compute based on the a posteriori residual esti-

[7] Here and elsewhere we use the notation $\boldsymbol{r} = [r_{\varrho_h, \boldsymbol{u}_h}(\phi_1), \ldots, r_{\varrho_h, \boldsymbol{u}_h}(\phi_{N_h})] \in \mathbb{R}^{N_h}$ for the algebraic representation of the residual $r_{\varrho_h, \boldsymbol{u}_h} \in V_h'$ with respect to the basis $\phi_1, \ldots, \phi_{N_h}$ of V_h during PCG iterations.

[8] In the presented case $\min\{n_x, n_y\} = 60$. This quantity is a heuristic number of iterations proposed by Amir et al. in [4] based on numerical experiments with early termination of a PCG solver.

Table 1 Number of OC iterations performed to meet the convergence criterion (8), when using "exact" and inexact solutions of the state and adjoint equations. Note that in Example 3 and 4 we compare the designs after 50 and 150 iterations respectively, even though (8) is not satisfied.

Example	"exact"	inexact
Example 1	69	70
Example 2	336	339

mates. In this plot we also see the "adaptive" behaviour of the methods: the number of PCG iterations decreases as the optimization algorithm progresses and eventually starts making smaller and smaller alterations to the material distribution. The average number of PCG iterations used with the proposed stopping criterion is only 6.5 for this example and this discretization. Despite the small number of iterations and rather approximate solutions, the approximate approach compares very well to using a direct solver, as can be seen in Fig. 4. For practical purposes the behaviour of the two methods can be deemed identical. Indeed, with respect to the a posteriori estimate $\|\!|\ell - \boldsymbol{\mathcal{K}}_{\varrho h}\boldsymbol{u}_h^{(i)}|\!\|_{\frac{h}{2}}$ utilizing a direct solver gives no visible improvement when compared to the very approximate solutions from the early termination of PCG, see Fig. 5.

As a convergence criterion for the optimality criterion iterations we use the inequality

$$\|\rho_{k,h} - \rho_{k-1,h}\|_{L^2(\Omega)} < \epsilon_{OC}, \tag{8}$$

and in this work we take $\epsilon_{OC} = 10^{-3}$. The number of OC iterations necessary to meet this criterion is very similar when using "exact" and inexact solutions of the state and adjoint equations, as can be seen from Table 1. This indicates that the inexact solutions do not compromise the quality of the OC iterations at this accuracy level.

The performance of the proposed method has also been studied on a sequence of refined grids, see Fig. 6. Generally speaking, we observe qualitatively similar behaviour, where the discretization error dominates after a few iterations of the optimization algorithm. As expected, the estimated residual decreases with mesh refinement, see Fig. 6. More importantly, the average number of required PCG iterations depends rather weakly on the mesh size, see Table 2.

As mentioned earlier the estimation of the dual norm of the residual is computationally burdensome, so the significant reduction in the number of PCG iterations does not translate to a significant reduction in the computational time. Running on a 64 bit HP EliteBook 840 G4 with and Intel(R) Core(TM) i7-7500U CPU, with clock rate of 2.70 GHz and with a 360×120 grid the average time spent per PCG iteration was 0.47 sec with the proposed stopping criterion, and 0.028 sec for a standard PCG iteration. Even with such an increase, owing to the reduction in the number of iterations the average time per each application of the linear solver still reduces from 13.0 sec for the "standard" PCG stopping criterion $\|\mathbf{r}\|_2 < 10^{-6}$ to 4.5 sec for the proposed early termination criterion. We emphasize yet again that we do not claim that the proposed algorithm is (more) efficient, as we could change the relationship between these timings almost arbitrarily by changing

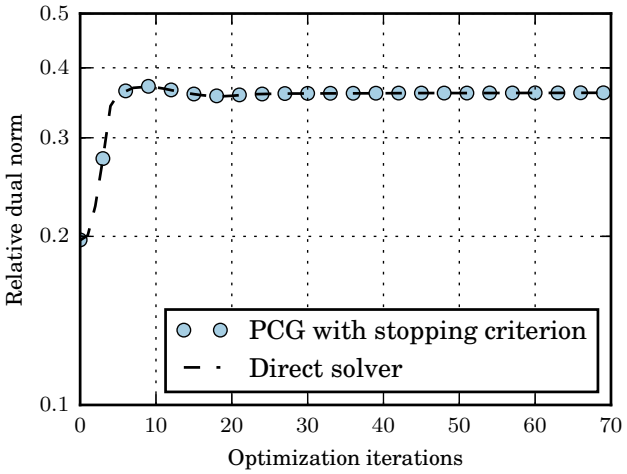

Fig. 5 The relative dual norm of the residual during the optimization process based on early termination of PCG vs. utilizing the direct solver; 2D cantilever example.

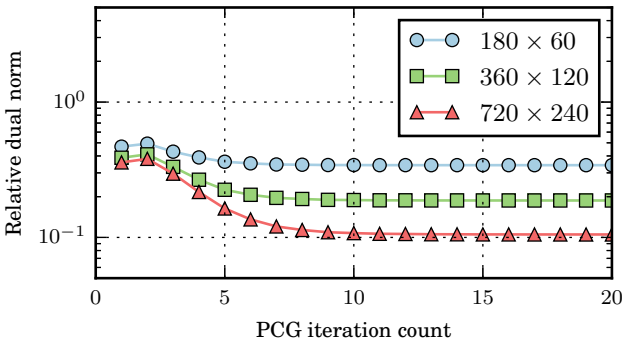

Fig. 6 The relative dual norm of the residual, $\|r\|_{V_{\frac{h}{2}}'} / \|\ell\|_{V_{\frac{h}{2}}'}$ for designs from optimization iteration 3; 2D cantilever example on a sequence of refined grids. The design from optimization iteration 2 is used as starting guess for the PCG solver.

Table 2 The average number of PCG iterations over the course of the optimization run; 2D cantilever example on a sequence of refined grids.

Mesh size	Average no. PCG iterations
180×60	6.5
360×120	7.2
720×240	8.6

the accuracy, preconditioning strategy, or computer architecture. Indeed, the times were measured when using an LU (pre-)factorization of the stiffness matrix for the system (7) which we solve to estimate the dual norm of the residual, which is not feasible for large scale problems. Also, for this example and grid size when using a direct solver in FEniCS the average time spent per each application of the linear solver is only 0.85 sec, which is several times faster than either of the iterative methods.

As a final experiment we have attempted to utilize the (preconditioned) minimum residual method (MINRES) instead of PCG [22]. Both methods are Krylov subspace algorithms for symmetric matrices based on Lanczos iterations, and as such have similar complexities, when the matrix-vector product and preconditioning dominate the rest of the computations. Unlike PCG, in its unpreconditioned version MINRES minimizes the residual of the equation over the current Krylov subspace, and as such is applicable to indefinite and even singular systems. Since we terminate the computations based on the value of the residual, one may expect that even fewer iterations will be required to reach the desired accuracy than in the case of PCG. Furthermore, previous studies have indicated that even for positive definite systems MINRES might be preferable to PCG if the algorithm is terminated early [13]. For the present problem we have observed that when utilizing MINRES together with the proposed stopping criterion, the resulting performance of the optimization algorithm is very similar to the performance when PCG is used. For example, on 180×60 grid, the average number of MINRES iterations is 6.74 compared to 6.49 when using PCG.

6.2 Example 2: 2D force inverter

In this example the force inverting compliant mechanism from [30] is considered. This problem has been a benchmark in topology optimization since [29]. The setup is to maximize the transferred work from the input actuator to the output spring. The example is sketched in Fig. 2. We utilize a structured grid of size 160×80, a filter radius of $\widehat{r} = 0.03$ and the volume fraction is set to $V_0 = 0.3$. The spring constants are set to $k_{\text{in}} = 1$ and $k_{\text{out}} = 0.001$ as in [30]. This problem is not self-adjoint so at each optimization iteration we have to solve both the state and adjoint systems. For the illustrative purpose, we will treat the solution of the two linear systems independently from each other. That is we apply PCG with the proposed stopping criterion to approximate the solution to each system. In practice one would use a block-CG method [17], where one solves

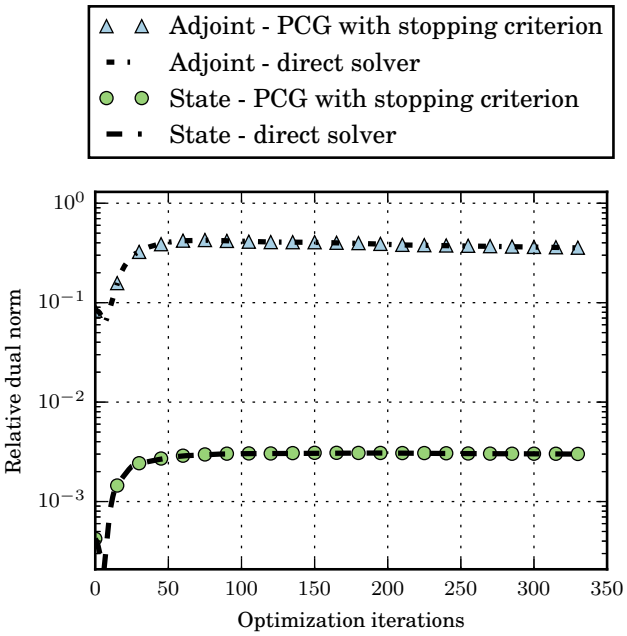

Fig. 7 The relative dual norm of the residual during the optimization iterations for the 2D force inverter example.

for the two right hand sides simultaneously and shares information between the two Krylov spaces. For this example the average number of PCG iterations in each optimization step is 7.3 and 8.6 for the state and adjoint equations, respectively. For comparison the number of iterations necessary to meet the criterion $\|\mathbf{r}\|_2 < 10^{-6}$ is in this case 331.2 and 6625.17. Compared to the direct solver the two methods behave nearly identically with respect to the resulting dual norm of the residual during optimization, as seen in Fig. 7. The relative difference between the objective values of the two computed designs is merely 0.11%.

This example is especially interesting considering how poorly the adjoint equations are solved at this discretization level: indeed the estimated relative dual norm of the residual is of the order of 50%!

6.3 Example 3: 3D cantilever

To test the method on larger scale problems we will consider two 3D examples. The first one is a 3D cantilever beam with a sine-shaped load, an example which is also considered in [3]. The beam is fixed at one end and a sine-shaped load is applied at the other end. We consider a $30 \times 30 \times 60$ grid, a filter radius of $\widehat{r} = 0.02$ and a volume fraction of $V_0 = 0.12$. We consider the design after 50 iterations of the optimality criteria algorithm. With accurate analysis the compliance for the obtained design is $J(\rho) = 4.144 \cdot 10^{-3}$. When we only approximate the analysis equations with the proposed

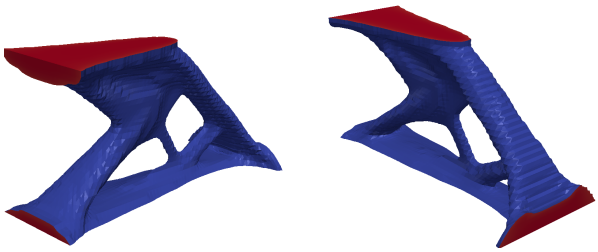

Fig. 8 The obtained design for the 3D cantilever with a sine-shaped load, using the approximate approach based on the proposed stopping criterion for PCG.

stopping criterion, again with $\varepsilon_\eta = 10^{-4}$, the compliance of the obtained design is $J(\rho) = 4.145 \cdot 10^{-3}$. The relative difference is here only $3.2 \cdot 10^{-4}$. The average number of PCG iterations taken over the course of optimization iterations is 15.0, while the average number of PCG iterations necessary to meet the "accurate" criterion $\|\mathbf{r}\|_2 < 10^{-6}$ is 156.7. The design obtained when using the approximate approach is shown in Fig. 8.

6.4 Example 4: 3D force inverter

The final example is that of a 3D force inverting compliant mechanism, similar to the example considered in [3]. The displacements are fixed to zero in the corners at one end, and as in Example 2, the 2D force inverter, the goal is to transfer work from the input actuator to the output spring. The spring constants used here are $k_{\text{in}} = 0.1$ and $k_{\text{out}} = 0.0001$. The point load boundary conditions are approximated by a smooth mollifier with small support. We consider only a quarter of the design domain owing to the symmetry. The grid size is $20 \times 20 \times 40$ with a filter radius of $\widehat{r} = 0.02$.

The volume fraction is $V_0 = 0.1$ and we consider the design after 150 iterations of the optimality criterion method. The objective function during the optimization is shown in Figure 10. When using a PCG solver from FEniCS preconditioned with AMG and with relative tolerance 10^{-6}, the objective value, after 150 iterations is $J(\rho) = -8.098$. When using the approximate approach presented here, the objective value after 150 iterations is $J(\rho) = -7.694$. The relative difference of the objective values is in this case approximately 5% and the performance of the two methods is quite similar, despite the very coarse discretization. The average number of PCG iterations for the approximate approach in this case is 20.5 and 21.7 for the state and adjoint system respectively. The final design obtained when using the approximate approach is shown in Fig. 9.

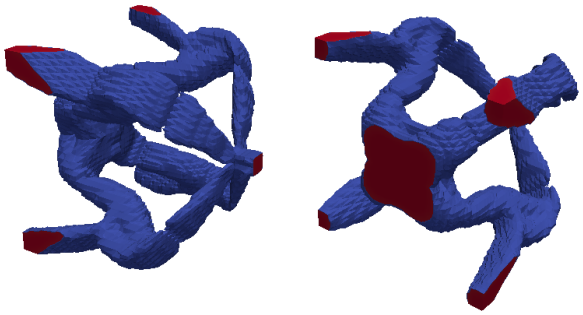

Fig. 9 The obtained design for the 3D force inverter, using the approximate approach based on the proposed stopping criterion for PCG.

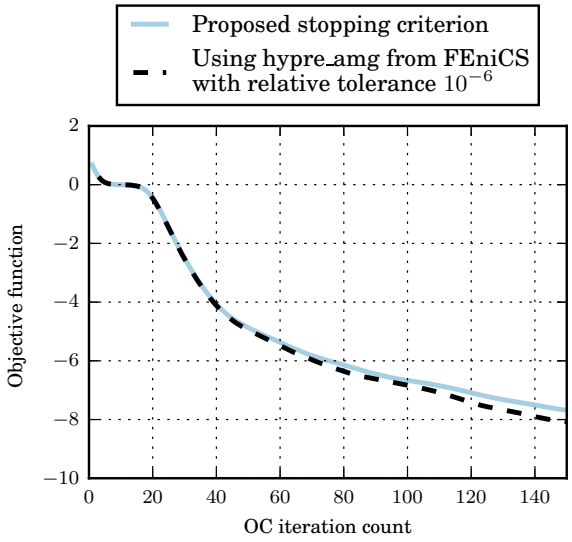

Fig. 10 The objective function during the optimization process of the 3D force inverter.

7 Discussion and conclusions

We have revisited the issue of inexact solution of state and adjoint equations during the topology optimization process. To this end we have derived quantitative conditions for early termination of the iterative solver, preconditioned conjugate gradient in the present case. The proposed stopping criterion is based on a posteriori FEM estimates, which is a rich research field in its own right.

We have demonstrated the performance of the proposed approximate approach on a few benchmark examples from 2D and 3D linear elasticity. Our main observation is that with inexact solutions of the state and adjoint equations, the optimization algorithms behave similarly to what one would expect from "exact" solutions using, for example, direct solvers. The number of iterations has been significantly reduced even when compared to some previous heuristic recommendations

for inexact solutions obtained by early termination of iterative methods, e.g. see [4].

The residual estimation method utilized in this work requires finding the solution to the linear system (7), that is inverting the Riesz map for an h-refined space, at each PCG iteration. This requires solving a linear elliptic PDE with constant coefficients, where only the right hand side of the equation changes from one iteration to another. In principle, memory permitting, one can prefactorize the matrix resulting from this system before starting the optimization process, or at least prebuild a high quality preconditioner for this problem (for example AMG, see for instance [31]). However, this system is still large — larger than the elasticity system we are solving owing to the mesh refinement, which places a significant computational burden on the residual estimation alone. Owing to this additional computational burden we readily acknowledge that it is impractical to utilize the approach presented here as is. Nevertheless we believe that it provides valuable insight into why approximate approaches yield accurate results. Additionally, it provides quantitative guidelines for premature termination of PCG solvers which are based on rigorous residual estimates and not heuristic empirical statements. Furthermore, when combined with an efficient method for measuring the residual, our experiments show that a lot of computational effort can be saved by solving the discrete problem *just* accurately enough.

A posteriori error estimation in FEM is a classical topic, which is well studied with many different estimators existing; see for example [2] or [15, Chapter 10]. Most of a posteriori estimators do not require solving large additional systems. We have for example evaluated utilizing the error estimator from [32]. While using this error estimator instead of the one described in Section 4 for stopping PCG prematurely, we obtained results similar to those presented in this paper. However, when considering the behaviour of the estimate during the PCG iterations, see e.g. Fig. 1, we have not observed the same monotonic behaviour. As a result it is difficult to motivate such an error estimator for our purposes. In our opinion it would be interesting to investigate computationally viable alternative a posteriori error estimators, which can be used instead of the one we have utilized.

Finally, we would like to mention that certain FEM discretizations naturally come equipped with a posteriori estimates. This is notably true for least-squares FEM, such as for example discontinuous Petrov–Galerkin methods, see for example [10, 12, 18], which is presently a very active research area. Indeed, in these discretizations the residual estimate is computed locally as a by-product, and such methods would therefore be good candidates for exploiting an approximate approach as the one presented here.

Acknowledgments

The authors would like to thank the two anonymous reviewers for their helphul and constructive comments, and particularly for making the authors aware of the references [6, 19].

References

1. Ahrens, J., Geveci, B., Law, C.: Paraview: An end-user tool for large data visualization. The visualization handbook **717** (2005)
2. Ainsworth, M., Oden, J.T.: A posteriori error estimation in finite element analysis. Comput. Methods Appl. Mech. Eng. **142**(1-2), 1–88 (1997). DOI 10.1016/S0045-7825(96)01107-3
3. Amir, O., Aage, N., Lazarov, B.S.: On multigrid-CG for efficient topology optimization. Struct. Multidiscip. Optim. **49**(5), 815–829 (2014). DOI 10.1007/s00158-013-1015-5
4. Amir, O., Sigmund, O.: On reducing computational effort in topology optimization: How far can we go? Struct. Multidiscip. Optim. **44**(1), 25–29 (2011). DOI 10.1007/s00158-010-0586-7
5. Amir, O., Stolpe, M., Sigmund, O.: Efficient use of iterative solvers in nested topology optimization. Struct. Multidiscip. Optim. **42**(1), 55–72 (2010). DOI 10.1007/s00158-009-0463-4
6. Arioli, M.: A stopping criterion for the conjugate gradient algorithm in a finite element method framework. Numer. Math. **97**(1), 1–24 (2004). DOI 10.1007/s00211-003-0500-y
7. Bendsøe, M.P., Sigmund, O.: Topology optimization: theory, methods, and applications, 2 edn. Springer-Verlag Berlin Heidelberg (2004). DOI 10.1007/978-3-662-05086-6
8. Biros, G., Ghattas, O.: Inexactness issues in the Lagrange-Newton-Krylov-Schur method for PDE-constrained optimization. In: L.T. Biegler, M. Heinkenschloss, O. Ghattas, B. van Bloemen Waanders (eds.) Large-Scale PDE-Constrained Optimization, pp. 93–114. Springer Berlin Heidelberg, Berlin, Heidelberg (2003)
9. Biros, G., Ghattas, O.: Parallel Lagrange–Newton–Krylov–Schur methods for PDE-constrained optimization. part I: The Krylov–Schur solver. SIAM J. Sci. Comput. **27**(2), 687–713 (2005). DOI 10.1.1.113.9614
10. Bramwell, J., Demkowicz, L., Gopalakrishnan, J., Qiu, W.: A locking-free hp DPG method for linear elasticity with symmetric stresses. Numer. Math. **122**(4), 671–707 (2012). DOI 10.1007/s00211-012-0476-6
11. Brezis, H.: Functional analysis, Sobolev spaces and partial differential equations. Springer Science & Business Media (2010)
12. Carstensen, C., Hellwig, F.: Low-order discontinuous Petrov–Galerkin finite element methods for linear elasticity. SIAM J. Numer. Anal. **54**(6), 3388–3410 (2016). DOI 10.1137/15M1032582

13. Chin-Lung Fong, D., Saunders, M.: CG versus MINRES: An empirical comparison. Sultan Qaboos Univ. J. Sci. (2012). DOI 10.24200/squjs.vol17iss1pp44-62
14. Dembo, R.S., Eisenstat, S.C., Steihaug, T.: Inexact Newton methods. SIAM J. Numer. Anal. **19**(2), 400–408 (1982). DOI 10.1137/0719025
15. Ern, A., Guermond, J.L.: Theory and practice of finite elements, vol. 159. Springer Science & Business Media (2013)
16. Evgrafov, A.: On Chebyshevs method for topology optimization of Stokes flows. Struct. Multidiscip. Optim. **51**(4), 801–811 (2015). DOI 10.1007/s00158-014-1176-x
17. Gutknecht, M.H.: Block Krylov space methods for linear systems with multiple right-hand sides: an introduction. Report (2006). DOI 10.1.1.189.5036
18. Keith, B., Fuentes, F., Demkowicz, L.: The DPG methodology applied to different variational formulations of linear elasticity. Comput. Methods Appl. Mech. Eng. **309**, 579–609 (2016). DOI 10.1016/j.cma.2016.05.034
19. Kočvara, M., Mohammed, S.: Primal-dual interior point multigrid method for topology optimization. Siam J, Scient. Comput. **38**(5), B685–B709 (2016). DOI 10.1137/15M1044126
20. Lazarov, B.S., Sigmund, O.: Filters in topology optimization based on Helmholtz-type differential equations. Int. J. Numer. Methods Eng. **86**(6), 765–781 (2011). DOI 10.1002/nme.3072
21. Logg, A., Mardal, K., Wells, G.: Fenics: Automated Solution of Differential Equations by the Finite Element Method. Ieeexplore.Ieee.Org **84**, 723 (2012). DOI 10.1007/978-3-642-23099-8
22. Paige, C.C., Saunders, M.A.: Solution of sparse indefinite systems of linear equations. SIAM J. Numer. Anal. **12**(4), 617–629 (1975). DOI 10.1137/0712047
23. Petzoldt, M.: Regularity and error estimators for elliptic problems with discontinuous coefficients. Ph.D. thesis, Freie Universität Berlin (2001)
24. Pimanov, V., Oseledets, I.: Robust topology optimization using a posteriori error estimator for the finite element method. Struct. Multidiscip. Optim. (2018). DOI 10.1007/s00158-018-1985-4
25. Plaza, A., Carey, G.: Local refinement of simplicial grids based on the skeleton. Appl. Numer. Math. **32**(2), 195–218 (2000). DOI 10.1016/S0168-9274(99)00022-7
26. Prudencio, E.E., Byrd, R., Cai, X.C.: Parallel full space SQP Lagrange–Newton–Krylov–Schwarz algorithms for PDE-constrained optimization problems. SIAM J. Sci. Comput. **27**(4), 1305–1328 (2006). DOI 10.1137/040602997
27. Rojas-Labanda, S., Stolpe, M.: An efficient second-order SQP method for structural topology optimization. Struct. Multidiscip. Optim. **53**(6), 1315–1333 (2016). DOI 10.1007/s00158-015-1381-2
28. Saad, Y.: Iterative Methods for Sparse Linear Systems, second edn. Soc. Ind. Appl. Math. (2003). DOI 10.1137/1.9780898718003
29. Sigmund, O.: On the Design of Compliant Mechanisms Using Topology Optimization. Mech. Based Des. Struct. Mach. **25**(4), 493–524 (1997). DOI 10.1080/08905459708945415
30. Sigmund, O.: Morphology-based black and white filters for topology optimization. Struct. Multidiscip. Optim. **33**(4-5), 401–424 (2007). DOI 10.1007/s00158-006-0087-x
31. Stüben, K.: A review of algebraic multigrid. J. Comput. Appl. Math. **128**(1), 281 – 309 (2001). DOI 10.1016/S0377-0427(00)00516-1. Numer. Anal. 2000. Vol. VII: Partial Differential Equations
32. Verfürth, R.: A posteriori error estimation and adaptive mesh-refinement techniques. J. Comput. Appl. Math. **50**(1-3), 67–83 (1994). DOI 10.1016/0377-0427(94)90290-9