Direct gradient projection method with transformation of variables technique for structural topology optimization

Cheng Chang · Andrew Borgart · Airong Chen · Max A.N. Hendriks

Received: date / Accepted: date

Abstract This paper proposes an efficient and reliable topology optimization method that can obtain a black and white solution with a low objective function value within a few tens of iterations. First of all, a transformation of variables technique is adopted to eliminate the constraints on the design variables. After that, the optimization problem is considered as aiming at the minimum compliance in the space of design variables which is supposed to be solved by iterative method. Based on the idea of the original gradient projection method, the direct gradient projection method (DGP) is proposed. By projecting the negative gradient of objective function directly onto the hypersurface of the constraint, the most promising search direction from the current position is obtained in the vector space spanned by the gradients of objective and constraint functions. In order to get a balance between efficiency and reliability, the step size is constrained in a rational range via a scheme for step size modification. Moreover, a grey elements suppression technique is proposed to lead the optimization to a black and white solution at the end of the process. Finally, the performance of the proposed method is demonstrated by three numerical examples including both 2D and 3D problems in comparison

Cheng Chang(⊠) · Airong Chen Department of Bridge Engineering, Tongji University, 200092 Shanghai, China E-mail: 06chang@gmail.com

Cheng Chang · Andrew Borgart Faculty of Architecture, Delft University of Technology, 2600 GA Delft, The Netherlands

Max A.N. Hendriks Faculty of Civil Engineering and Geosciences, Delft University of Technology, 2600 GA Delft, The Netherlands; Norwegian University of Science and Technology, 7491 Trondheim, Norway with the typical SIMP method using the optimality criteria algorithm.

Keywords Direct gradient projection method (DGP) · Structural topology optimization · Transformation of variables technique · Efficiency and reliability

1 Introduction

Efficiency and reliability are the two dominating preferences of structural topology optimization. However, these two properties of an optimization method are contradictory in a sense. So far, most of the topology optimization methods put emphasis on reliability, endeavoring to get to the theoretical optimal solution as close as possible. Whereas an efficient method with fast convergence is preferred in industrial applications (Sigmund 2007; Rozvany 2009). This paper is dedicated to conceive a topology optimization method with a balance between efficiency and reliability.

In early days, some benchmark problems of topology optimization were solved by analytical method (e.g. Michell 1904; Rozvany 1998). Owing to the development of finite element theory and computer technique, numerical methods for topology optimization have taken the place of analytical methods, and allow to solve topology optimization problems of large and complex systems. By discretizing the design domain into elements, the density variables of the elements can be used as design variables to interpret all the potential structural forms in this domain.

Generally speaking, topology optimization is a constrained nonlinear programming problem, aiming to find the minimum feasible point in an Euclidean N-dimensional space of design variables. Considering the high costs of objective function evaluations and the relatively low costs of first derivatives calculations, an efficient method should make full use of the evaluation and first-order information of the objective and constraint functions to generate the next most promising point.

Density variables of structural topology optimization are usually expected to have the discrete value 0 or 1, as is set in Evolutionary Structural Optimization method (ESO, Xie and Steven 1993). However, this setting will cause oscillation or even failure in programming when the gradient of the objective function (i.e. sensitivity with respect to density variable) changes greatly between consecutive iterations (Rozvany 2009). Therefore, continuity is generally indispensable for the density variables during the process of optimization.

Based on the continuity of design variables, optimality criteria method (OC, Bendsøe 1989), the Convex Linearization method (CONLIN, Fleury 1989) and the Method of Moving Asymptotes (MMA, Svanberg 1987) have been proved reliable in many topology optimization problems. As a matter of fact, the OC method is the same as the solution of structural topology optimization with the CONLIN method, while the CONLIN method can be obtained as the special case of MMA when the lower asymptotes are set to 0 ($L_e = 0$) and the upper asymptotes are set to positive infinite $(U_e = \infty)$. These methods perform linearization with respect to the reciprocal of design variables so that the transformed subproblem, which is a simple convex programming problem, can be solved by a dual method. In consequence of the linear approximation, the updated design may deviate more or less from the most promising direction. Furthermore, though parameter L_e and U_e in MMA are related to the convergence rate of the optimization process, none of these existing methods efficiently control the step size for each iteration according to the feature of topology optimization.

The gradient projection method (Rosen 1960, 1961) is a direct nonlinear programming method without linear approximation. By taking a step in the direction of the projection of the negative gradient of the objective function onto the tangent hyperplane of the currently active constraint, a feasible point with a lower objective function value is expected to be obtained in each iteration. The conventional gradient projection method was considered to be inadequate for structural optimization (Fleury 1989) for the reason that a large number of iterations are usually required for convergence. However, the weakness can be conquered according to the special property of the given structural optimization problem (Vanderplaats and Moses 1973).

The transformation of variables technique is employed to eliminate the constraints on design variables. Since there is only one constraint of volume left, the direct gradient projection method is proposed to obtain a new feasible point with a lower value of objective function directly.

This paper is organized as follows. In Section 2 the mathematical formulations of structural topology optimization and the transformation technique of density variables are elaborated. The original gradient projection method and the direct gradient projection method are interpreted in Section 3. Besides, the convergence of the proposed direct gradient projection method is also discussed. Section 4 proposes several implementation techniques that can improve the efficiency and stability of the algorithm. An overview of the algorithm is presented in Section 5. Section 6 presents three selected numerical examples to demonstrate the performance of the proposed method in both 2D and 3D structural topology optimization problems. Some discussion about the results of numerical examples is finally given in Section 7.

2 Mathematical framework

2.1 Material interpolation scheme

Structural topology optimization adopts the same physical model as that in the finite element analysis, except that an artificial material defined by an interpolation scheme is employed. The Solid Isotropic Material with Penalization for intermediate densities method (SIMP, Bendsøe 1989) is the most effective and widely used material interpolation scheme. With the power-law penalization of the Young's modulus for an intermediate density material, a black and white solution without too many grey elements can be attained. The SIMP material interpolation scheme is defined as follows:

$$E_e(\rho_e) = \rho_e^p E_0, \qquad \rho_e \in [0, 1] \tag{1}$$

where E_0 is the stiffness of the material, p (typically p = 3) is the penalization factor, ρ_e is a density variable assigned to element e that is constrained to be between 0 and 1, and E_e is the Young's modulus of element e corresponding to the density variable ρ_e .

2.2 Problem formulation

The compliance of a structure, C, is usually adopted as the objective function, and the volume is constrained. Therefore, on the basis of the physical model, the general mathematical model can be described by the following formulation (Bendsøe and Sigmund 2003):

$$\min_{\boldsymbol{\rho}} : \quad f(\boldsymbol{\rho}) = C = \mathbf{F}^T \mathbf{U} = \sum_{e=1}^N E_e(\rho_e) \mathbf{u}_e^T \mathbf{k}_0 \mathbf{u}_e$$

$$s.t. : \quad g(\boldsymbol{\rho}) = V(\boldsymbol{\rho}) - V^* = \sum_{e=1}^N \rho_e - V^* \le 0 \quad (2)$$

$$\mathbf{K} \mathbf{U} = \mathbf{F}$$

$$\rho_e \in [0, 1]$$

where $\boldsymbol{\rho}$ is the density vector whose components are ρ_e , *N* is the number of elements, **U** and **F** are the displacement vector and force vector respectively, **K** is the global stiffness matrix determined by the density vector $\boldsymbol{\rho}$, \mathbf{k}_0 is the element stiffness matrix with unit Young's modulus, \mathbf{u}_e is the displacement vector for the element $e, V(\boldsymbol{\rho})$ is the total volume of the structure corresponding to the density vector $\boldsymbol{\rho}$ and V^* is the constrained volume that is given beforehand. Assuming uniform element meshes, all the element volumes are equal to the density variable ρ_e . The first derivatives with respect to the density variables can be derived as follows:

$$\frac{\partial f}{\partial \rho_e} = -\mathbf{u}_e^T \frac{\partial \mathbf{k}_e}{\partial \rho_e} \mathbf{u}_e = -p\rho_e^{p-1} \mathbf{u}_e^T \mathbf{k}_0 \mathbf{u}_e \le 0$$
(3)

$$\frac{\partial g}{\partial \rho_e} = \frac{\partial V}{\partial \rho_e} = 1 > 0 \tag{4}$$

Form (3) it can be concluded that every component of the gradient of the objective function is always negative, which means additional material in any element decreases the value of the objective function. Therefore the minimum is on the hyperplane of the volume constraint and ' \leq ' in the constraint formulation can be replaced by '='.

2.3 Transformation technique

Transformation technique is a common mathematical technique widely used in extremum problems. An explicit constraint formulation can be satisfied automatically by making a transformation of the independent variables.

As for topology optimization, the density variables are usually adopted as design variables directly. Whereas, it will tremendously facilitate the optimization if the constrained range [0,1] of design variables is expanded to be infinite and the monotonicity is maintained. Hence, a transformation scheme is devised as follows (Rao 2009):

$$\rho_e = \frac{1}{1 + e^{-x_e}} \tag{5}$$

where ρ_e is still the density variable of element *e* while x_e is the new design variable which has an infinite range. Thus the design variables ρ in (2) is replaced by **x**, and the constraints on the design variables are eliminated, such that:

$$\begin{aligned}
\min_{\mathbf{x}} : \quad f(\mathbf{x}) &= C = \mathbf{F}^T \mathbf{U} = \sum_{e=1}^N E_e(x_e) \mathbf{u}_e^T \mathbf{k}_0 \mathbf{u}_e \\
s.t. : \quad g(\mathbf{x}) &= V(\mathbf{x}) - V^* = 0 \\
\mathbf{K} \mathbf{U} &= \mathbf{F}
\end{aligned} \tag{6}$$

Other than the global equilibrium equation, there is only one constraint of volume left.

At the expense of the beneficial reduction of constraints, there are two defects for optimization brought by the transformation scheme. Firstly, the range of the density variables is converted from [0,1] to (0,1), i.e. it is impossible for a density variable to be 0 or 1. Secondly, the constraint equation of volume is changed from linear to nonlinear. However, the first defect is negligible which can be demonstrated later by the numerical examples. The second defect indeed brings some difficulties to the original gradient projection method, so the direct gradient projection method is proposed to circumvent it in the following section.

3 Gradient projection method

The constrained nonlinear optimization problem (6) is usually solved on the basis of the following iterative scheme (Rao 2009; Arora 2004):

$$\mathbf{X}_{i+1} = \mathbf{X}_i + \lambda \mathbf{S}_i \tag{7}$$

where \mathbf{X}_i and \mathbf{X}_{i+1} represent the current point and the next point in the space of design variables respectively, and λ is the step size in the search direction \mathbf{S}_i . For the sake of explicitness, the search direction \mathbf{S}_i is usually converted to a unit vector. The key problem of the algorithm is decomposed into how to decide the search direction and the step size that can obtain a new feasible point with a lower objective function value in the design space.

3.1 Original Gradient projection method

The negative gradient of the objective function is the steepestdescent direction for unconstrained optimization problems. However, this direction is usually infeasible when some constraints exit. In order to satisfy the constraints, the constrained steepest-descent direction, which is actually the projection of the steepest-descent onto the constraint hyperplane, is taken as the search direction in the original gradient projection method (Rosen 1960).

When at least one of the active constraint functions is nonlinear, the search direction is converted to be the projection of the steepest-descent onto the tangent hyperplane of the constraints (Rosen 1961). Due to the nonlinearity of the constraints, the new point obtained along the negative projected gradient direction is still infeasible. Therefore, a correction procedure that brings the infeasible point X' back to the constraint hypersurface is required, as shown in Fig. 1.

Furthermore, the original gradient projection method proposes a rational step size scheme corresponding to the curvature of the objective and constraint functions, the linear independence of the objective and constraint functions and the Euclidean norm of the projection vector. Since it requires the second-order information (curvature) of the compliance function which is difficult to calculate in a structural problem, it is inappropriate for structural topology optimization.



Fig. 1 Finding a feasible and useful point in the original gradient projection method

3.2 Direct Gradient projection method

In the original gradient projection method, in order to satisfy the nonlinear constraints, the steep-descent is firstly projected onto the tangent hyperplane of the constraints and then a further constraint correction process is executed. This two-step procedure is tedious when there is only one constraint existing, so the direct gradient projection method that can obtain the feasible point directly is proposed in this section.

Next to the evaluations of the objective and constraint functions, their first derivatives with respect to the design variables can be obtained at each successive iteration. According to (3) and (4), the first derivatives with respect to the transformed design variables, also termed as sensitivities, can be obtained by the chain rule:

$$\frac{\partial f}{\partial x_e} = \frac{\partial f}{\partial \rho_e} \frac{\partial \rho_e}{\partial x_e} = -p \mathbf{u}_e^T \mathbf{k}_0 \mathbf{u}_e \frac{1}{\left(1 + e^{-x_e}\right)^{p+1} e^{x_e}} < 0$$
(8)

$$\frac{\partial g}{\partial x_e} = \frac{\partial g}{\partial \rho_e} \frac{\partial \rho_e}{\partial x_e} = \frac{1}{\left(1 + e^{-x_e}\right)^2 e^{x_e}} > 0 \tag{9}$$

The negative gradients of the objective and constraint functions are respectively the directions in which the objective and constraint functions decreases most rapidly at the current point. Hence, a feasible and useful step λS can be obtained in the vector space spanned by the gradients of the objective and constraint functions:

$$\lambda \mathbf{S} = -\alpha \nabla f - \beta \nabla g \tag{10}$$

The updated point \mathbf{X}_{i+1} is expected to be on the hypersurface of constraint g = 0, so the step $\lambda \mathbf{S}$ is a vector connecting two feasible points on the hypersurface of the constraint. The step vector $\lambda \mathbf{S}$ is considered as the direct projection of the negative gradient of the objective function onto the hypersurface of the constraint, as shown in Fig. 2. Hence,



Fig. 2 Finding a feasible and useful point in the direct gradient projection method

the proposed algorithm is termed as direct gradient projection method (DGP).

Now, for the reason that the constraint formulation is not linear, the step size λ and the search direction **S** are not independent. The problem of finding a feasible and useful step λ **S** is converted to the determination of α and β in (10). Both α and β are proportional to the step size λ and can be adopted as a step size multiplier to control the step size. Considering that the gradient of the constraint function ∇g is more stable than the gradient of the objective function ∇f , β is chosen as the step size multiplier. Once the step size multiplier β is determined, the other variable α can be calculated by the bisection method so that the constraint equation is satisfied. The modification of α is inexpensive because it only requires constraint function evaluations that have explicit formulations.

Generally, the optimization of the step size multiplier is needed via one dimensional search. However, it requires tremendous times of evaluation of the objective function which make it unacceptable considering the limit of the computation cost. Thus, an acceptable step size multiplier is determined at the beginning of each iteration. As a consequence, a better point can not be guaranteed for each iteration. In order to keep the step size within a rational range, a scheme for step size multiplier modification is introduced in Section 4.

3.3 Convergence criterion

Necessary conditions for a minimum of problem (6) can be obtained by the classical technique of Lagrange multipliers. The Lagrangian function of the problem (6) is written as:

$$\mathcal{L}(\mathbf{x}, \mathbf{v}) = f(\mathbf{x}) + \mathbf{v}g(\mathbf{x}) \tag{11}$$



Fig. 3 Necessary conditions for the minimum

where v is the Lagrangian multiplier. Stationary conditions for (11) are given by:

$$\frac{\partial \mathcal{L}}{\partial x_i} = \frac{\partial f}{\partial x_i} + v \frac{\partial g}{\partial x_i} = 0 \qquad \qquad i = 1, 2, \cdots, n \quad (12)$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{v}} = g = 0 \tag{13}$$

Equation (13) is the same as the equality constraint of volume in (6). It means the minimum is on the hypersurface of the constraint. Equation (12) can also be written in the form of gradients as:

$$\nabla f + \nu \nabla g = 0 \tag{14}$$

which means the gradient of the objective function ∇f and the gradient of the constraint function ∇g are parallel, as shown in Fig. 3.

The change of density variables between two consecutive iterations (i.e. step vector in the space of density variables) is defined as:

$$\mathbf{S}_i^* = \boldsymbol{\rho}_i - \boldsymbol{\rho}_{i-1} \tag{15}$$

Provided $\mathbf{S}_i^* = 0$, from (15), (5), (7) and (10) it can be concluded that:

$$-\alpha \nabla f - \beta \nabla g = \lambda \mathbf{S} = \mathbf{x}_i - \mathbf{x}_{i-1} = 0$$
(16)

Thus, the necessary conditions for the optimality in (12) and (13) are both satisfied, and \mathbf{x}_i is a stationary point of the Lagrangian function.

According to the SIMP method (Andreassen et al 2011), the convergence criterion for numerical computation is defined as follows:

$$\Delta = \left| s_{i,e}^* \right|_{\max} < 0.01 \tag{17}$$

where Δ is the index of convergence which can be considered as the step size in the space of density variables when

the step vector \mathbf{S}_{i}^{*} is normalized, and $s_{i,e}^{*}$ is the *e*-th component of step vector \mathbf{S}_{i}^{*} . The optimization procedure is terminated when the absolute value of all the components of step vector \mathbf{S}_{i}^{*} are less than 0.01.

Rigorously speaking, the differences between two consecutive design variables **x** might be still large, as the derivative of ρ_i with respect to x_i is pretty small when ρ_i is close to the value 1 or 0. However, in these regions of the transformation function, the precision of density variable ρ_i is of more practical relevance in the design than x_i .

4 Implementation techniques

4.1 Density filtering

In order to save the computation cost, the simplest linear elements are usually adopted in the structural topology optimization. Due to overestimating the stiffness of nodal connections between finite elements and the lack of scale information, the problems of checkerboards and mesh dependence will be inevitable if no restriction method is implemented (Diaz and Sigmund 1995; Sigmund and Petersson 1998). Many efforts have been made to devise a regularization scheme, such as filtering technique (Sigmund 1997; Bourdin 2001), perimeter control (Ambrosio and Buttazzo 1993; Haber et al 1996) and wavelet parameterizations (Kim and Yoon 2000; Poulsen 2002).

Density filtering (Bruns and Tortorelli 2001; Bourdin 2001) is a kind of filtering technique which uses the blurring technique borrowed from image processing. It intends to modify the density variables assigned to the elements with the data of their neighborhoods, thus a homogenized layout is prone to be obtained. Considering its efficiency and ease of implementation, density filtering is adopted in this paper. Since a transformation technique is applied, the density variables are replaced by the transformed design variables, formulated as follows (Andreassen et al 2011):

$$\tilde{x}_e = \frac{\sum\limits_{i \in N_e} H_{ei} x_i}{\sum\limits_{i \in N_e} H_{ei}}$$
(18)

where x_i is the original design variable of element i, \tilde{x}_e is the filtered design variable of element e, N_e is the set of elements i that are in the neighborhood of element e and H_{ei} is the weight of element i in accordance with element e, defined as follows:

$$H_{ei} = \max(0, r_{\min} - dist(e, i)) \tag{19}$$

Weighting function H_{ei} is the difference between the filter radius, r_{\min} , and the center to center distance, dist(e,i), of the two elements, and only the elements within the filter radius r_{\min} of element *e* are considered. Though a smoother weighting function with the Gaussian (bell- shaped) distribution is suggested by Bruns and Tortorelli (2003), this alternative method does not show any advantages compared to the linear version in some tests (Sigmund 2007). Therefore, the simple linear weighting function is adopted for density filtering.

Apart from the homogenization effect, the filter technique also includes some auxiliary scale information by introducing the filter radius r_{min} to the algorithm. The topological forms under different meshing schemes keep the same on condition that the ratios of the filter radius to the dimension of design domain are consistent (Andreassen et al 2011).

The first derivatives of objective and constraint function with respect to the filtered design variables \tilde{x}_e are still given by (8) and (9), provided that the variable x_e is replaced with \tilde{x}_e . Hence the derivatives with respect to original design variables can be derived by the chain rule (Andreassen et al 2011):

$$\frac{\partial \Psi}{\partial x_j} = \sum_{e \in N_j} \frac{\partial \Psi}{\partial \tilde{x}_e} \frac{\partial \tilde{x}_e}{\partial x_j} = \sum_{e \in N_j} \frac{1}{\sum\limits_{i \in N_e} H_{ei}} H_{je} \frac{\partial \Psi}{\partial \tilde{x}_e}$$
(20)

where the function ψ represents either the objective function $f(\mathbf{x})$ or the constraint function $g(\mathbf{x})$.

4.2 Constraints for stability

A better point with a lower objective function value cannot be guaranteed if the gradient of the objective or the constraint function varies greatly between two consecutive iterations. Hence, some constraint techniques are required to improve the stability.

Huang and Xie (2007, 2010) suggest averaging the sensitivity number with its historical information. Based on this idea, the golden ratio is employed to distribute the weight of two consecutive iterations. The modified sensitivity is expressed as:

$$\left(\frac{\partial f}{\partial x_e}\right)_{i,new} = 0.618 \left(\frac{\partial f}{\partial x_e}\right)_i + (1 - 0.618) \left(\frac{\partial f}{\partial x_e}\right)_{i-1}$$
(21)

Considering the different influences from current iteration and previous iteration, the current sensitivity is given more weight than the previous one. Though the golden ratio is adopted arbitrarily, this distribution of weight shows great advantage in the numerical examples. Owing to the combination of the first-order information for two consecutive iterations, the risk of failure caused by the severe difference between the consecutive objective function gradients is reduced and the amplitude of the oscillation of design variables decreases during the optimization process.

This average technique of historical information is not applied to the constraint function gradients because they are relatively stable during the optimization process. Since the hyperplane of constraint is linear in the Euclidean N-dimensional space of density variables, the constraints in problem (2) will still be satisfied if the step size is reduced in the search direction S_i^* . Generally, all the lengths of the components of step vector S_i^* are expected to be no more than the threshold Δ^* (typically, $\Delta^* = 0.2$) in consideration of the stability. Therefore, if Δ is greater than Δ^* , the step of density variables is scaled as:

$$\mathbf{S}_{i,new}^* = \frac{\Delta^*}{\Delta} \mathbf{S}_i^* \tag{22}$$

Besides, the step size is halved (i.e. $S_{i,new}^* = 0.5S_i^*$) if oscillation happens. In this paper oscillation is defined as follows:

$$\exists e \in N: \quad s_{i,e}^* \cdot s_{i-1,e}^* \le -0.5(\Delta^*)^2$$
(23)

or
$$C_i \ge 1.1C_{i-1}$$
 (24)

where $s_{i,e}^*$ is the *e*-th component of step vector \mathbf{S}_i^* and C_i is the compliance in the *i*-th iteration. Condition (23) is focused on the single density variable and condition (24) is related to the performance of the whole structure. These two kinds of oscillation can be both weakened by reducing the step size.

4.3 Scheme for step size multiplier

Step size multiplier β is the key to control the optimization process. On one hand, since it is an nonlinear problem and the gradient information can only represent the most promising searching direction at current point, the next point may deviate greatly from the optimum if the step size multiplier β is too big. On the other hand, if the step size multiplier β is too small, the design variables will change slowly and it will take many more iterations to attain the optimum. Therefore, a rational step size multiplier β must be determined on the compromise between efficiency and reliability.

Generally, the rational range of the step size in the space of density variables Δ is between 0.1 and 0.2, and there is a positive correlation between the step size multiplier β and the step size Δ . Hence a scheme for the step size multiplier β aiming to control the step size Δ within the rational range is conceived.

The initial value of step size multiplier, β_0 , is set to 1, which is considerably small for the first iteration. If the step size multiplier β is fixed, the step size Δ will decrease quickly below 0.1 which is considered to be inefficient. In order to improve the efficiency of each iteration, the step size multiplier β is magnified on condition that the step size Δ is less than 0.1 and the step size multiplier β has not reached its ceiling:

$$\beta_{i+1} = 2\beta_i \tag{25}$$

Moreover, to accelerate the process of optimization, a shift of β in (25) will also be triggered if β has stayed the same for 50 iterations unless the ceiling is reached. The ceiling of the step size multiplier, β_{max} , is set to 2⁹. Although a bigger value can also be given to β_{max} , it will cause the optimization more difficult to converge.

The scheme for the step size multiplier β determines the balance between efficiency and reliability. It can be modified according to user requirements. Moreover, since the step size multiplier β relates the design variable and the gradient of the constraint function, β_0 and β_{max} may be adjusted if equation (4) is not satisfied.

4.4 Grey elements suppression technique

In consequence of continuous density variables introduced by the material interpolation scheme, intermediate density elements (i.e. grey elements) will inevitably appear in the layouts. Some post-processing method have been applied to generate more distinct solid and void designs, such as isodensity surface (Sigmund 2007), blurring filter (Sigmund 1997) and regularized penalty function (Borrvall and Petersson 2001). Nevertheless, there would be a deviation of performance between the original design and the post-processed design. Hence, a discrete design is expected to be obtained directly by the optimization algorithm.

According to the transformation function in (5), a density variable ρ_e will get closer to 0 or 1 if the assigned design variable x_e is magnified. Therefore, a simple grey elements suppression technique is devised as:

$$\mathbf{x}_{new} = 1.2\mathbf{x} \tag{26}$$

The multiplier is chosen on the compromise of the efficiency and reliability. A bigger multiplier will cause the optimization converge quickly to a black and white layout, but the solution may deviate from the optimum due to the suppression effect of the design variables. In order not to add to much interference to the process of optimization, (26) is not implemented until the step size multiplier has reached the ceiling $\beta = \beta_{max}$.

In order to make a comparison of discreteness between different methods, a so called measure of non-discreteness (M_{nd}) was proposed by Sigmund (2007):

$$M_{nd} = \frac{\sum_{e=1}^{N} 4\rho_e (1 - \rho_e)}{N} \times 100\%$$
(27)

 M_{nd} is equal to 0% if there is no grey element existing in the design while M_{nd} is equal to 100% when all the density variables are 0.5.

5 Optimization algorithm

The structural topology optimization algorithm of direct gradient projection method with transformation of variables technique is devised as follows:

- Step 1 Define the reference domain and decide the design domain, solid domain and void domain. After the partition, the reference domain is discretized into a finite element model with equally sized elements, and the boundary conditions and loads are defined.
- Step 2 Define the optimization parameters including the constrained volume V^* , the filter radius r_{\min} , the initial and the ceiling values of step size multipliers β_0 and β_{\max} . The design variables and the density variables should be evenly initialized corresponding to the constrained volume V^* .
- *Step 3* Compute the structural performance by finite elements analysis.
- Step 4 Calculate the value of the objective function and the first derivatives of both the objective function and the constraint function with respect to the design variables according to (8) and (9). Then, modify the gradient information with the density filtering and stability technique according to (20) and (21).
- Step 5 Adjust the variable α in (10) by bisection method so that the constraint equation is satisfied.
- Step 6 Calculate the change of consecutive density variables \mathbf{S}_i^* and modify the step size in the space of density variables Δ as mentioned in Section 4.2.
- Step 7 Update the design and density variables to a new point.
- Step 8 Implement the shift of step size multiplier (25) or the grey elements suppression technique (26) if the trigger conditions are met.
- Step 9 Check the convergence of the solution. If (17) is satisfied then stop, otherwise go back to *Step 3*.

6 Numerical examples

The behavior of the proposed method is demonstrated by three selected examples both in 2D and 3D with a comparison to the typical SIMP method using the OC algorithm and sensitivity filter (Andreassen et al 2011), which is considered to be the most mature and widely accepted method of topology optimization by now. The DGP method and SIMP method are both programmed with FORTRAN. The regularized 4-node quadrilateral plane stress element and 8-node brick solid element with unit length of 1m are adopted respectively in 2D and 3D problems. The Young's modules of the solid elements is $E_0 = 1$ Pa, the Poisson's ratio is v = 0.3and the penalization factor is p = 3.



Fig. 5 Final layouts of the MBB beam and corresponding compliance C obtained with different methods

6.1 2D MBB beam

The Messerschmitt-Bölkow-Blohm (MBB) beam problem is a benchmark problem of 2D structural topology optimization. It is a simple supported beam with a concentrated load applied to the top middle of the span. Usually, only half of the model is used in optimization in order to save the cost of computation. The design domain and boundary conditions are shown in Fig. 4. The design domain is discretized with 150×50 elements. The filter radius is $r_{\min} = 6m$ and the volume fraction is set to 0.5.

The topological forms of the MMB beam problem attained by different methods are the same, as shown in Fig. 5. However, there are many grey elements existing in the layout of the SIMP method while there are only solid and void elements in the layout of the DGP method. Further, owing to the grey elements attached to the joint of the truss, the bending moments of the members are weakened. As a consequence, the widths of the web members are almost constant along the longitudinal direction.

The compliance of the solution with SIMP method is quite close to 219.52 which is obtained by Andreassen's 88 lines of code in MATLAB (Andreassen et al 2011). The MMB beam problem is also solved by the SIMP method with the MMA algorithm in MATLAB (Svanberg 1987). Although the MMA algorithm obtains almost the same layout as that obtained by OC algorithm, it takes 175 iterations to converge. The less efficiency may be attributed to the absence of proper step size restriction in the generic version of MMA.

6.2 3D cantilever

It is necessary to expand the structural topology optimization program from 2D to 3D, because a layout that interprets topological forms in all the three dimensions is usually required in actual design. The algorithm of the 3D problem is the same as that of 2D problem, except that the element type is changed from 2D plane stress element to 3D solid element.

A 3D cantilever with fixed support on the right side and an unit concentrated load at the center of lower left edge is presented here to perform the comparison, as shown in Fig. 6. The dimension of the design domain is $150m \times 7m \times$ 15m excluding the mirrored domain. The filter radius is $r_{\min} =$ 3m and the volume fraction is set to 0.2.

By blanking all the elements with the density under certain threshold values (for the DGP method $\rho_e = 0.05$ and for the SIMP method $\rho_e = 0.3$), clear layouts of material distribution are shown in Fig. 7. The solutions for the 3D cantilever problem are quite different from each other. First of all, the solid material concentrates at the middle of the domain in the layout of the DGP method while it is distributed at the two lateral sides in the layout of the SIMP method. As there is no torsion applied to the structure, the concentrated material distribution is more efficient for the reason that it saves the material of the bracings between the two lateral sides. Moreover, the angle of the bracings in the SIMP method is bigger than that in the DGP method. Finally, there are lots of grey elements existing in the final layout of the SIMP method, which is also inefficient due to the penalization of intermediate material.

6.3 3D bridge

At last, a relatively complex 3D bridge problem with a preset solid and void domain is performed. The bottom layer of elements are supposed to be the deck, thus all the elements are preset to solid and outside the design domain. Uniform





Fig. 6 Design domain and

boundary conditions of a 3D

cantilever

vertical loads are applied to the top nodes of these elements. The two rows of bottom nodes that are close to the edge are fixed, as shown in Fig. 8. The middle part of the domain is preset to void and can not be updated during the optimization process, which indicates the vertical and horizontal vehicle clearance of the bridge structure. The dimension of the computational domain is $150m \times 15m \times 15m$ excluding the mirrored domains. The thickness of the design domain is 3m, distributed in the top and right parts of the computational domain. The filter radius is $r_{min} = 3m$ and the volume fraction is set to 0.4.

Distinct topological forms are obtained by plotting the iso-density surface with the threshold value $\rho_e = 0.3$, as shown in Fig. 9. At the middle of span, much solid material is distributed at the top of the design domain so that a section with great flexural capacity is forged. While plane members and diagonal bracing members are formed respectively in the DGP and SIMP solutions at the quarter of the span where shear force and bending moment are both relatively large. The difference of the shear members between the two methods might be attributed to the different sensitiv-

ities of filter parameter r_{min} (Lazarov and Sigmund 2011), but exact explanation needs to be further discussed.

In addition, the iso-density surface of SIMP method is smoother than that of DGP method because there is a transition domain filled with intermediate material between the solid and void domain in the final layout of SIMP method. This statement is also firmly supported by the material distribution layouts of the mid-span section provided in Fig. 9.

7 Results and discussion

The curves of compliance (i.e. objective function) versus iteration and the scatter diagram of step size for the three numerical examples with the DGP method and the SIMP method are shown in Fig. 10. The values of mean compliance are shown in logarithmic scale since it changes greatly during the process of optimization.

The curves of compliance for the SIMP method decline rapidly at the beginning but turn to be almost horizontal when they reach certain values. Since the convergence is determined by the step size rather than the derivative of the



Fig. 7 Final layouts of 3D cantilever and corresponding compliance C obtained with DGP and SIMP methods

objective function, the optimization will not converge until the step size in the space of density variables Δ falls to less than 0.01. In comparison, the curves of compliance for the DGP method decline relatively slowly at the beginning but obtain smaller values with fewer steps at last. Detailed information of the results is shown in Table 1.

The indexes M_{nd} of the DGP method are all quite close to 0 which indicates no intermediate material existing in the final layouts. Though an extended method of the SIMP method with a Heaviside filter can also obtain a black and white solution, between 500 and 1,000 iterations are usually required (Sigmund 2007), which add too much burden to computation.

A comparison of computation cost is also given in Table 1. The CPU-time per iteration is the average over the first ten iterations of the optimization, using a desktop computer with an Intel Core i7-3770 processer, 8 GB memory, and Intel Fortran Compiler 11.0 (32-bit). Owing to the complexity of the transformation of design variables and density filter, the computation time per iteration of the DGP method is slightly more than that of the SIMP method.

The promising features of the DGP method are mainly attributed to the proposed scheme for the step size multiplier modification. By magnifying or constraining the step size



Fig. 9 Final layouts of 3D bridge and corresponding compliance C obtained with DGP and SIMP methods

multiplier, an efficient and reliable step with the step size between 0.1 and 0.2 can be attained in each iteration, unless it is close to convergence, as shown in Fig. 10. Besides, the grey elements suppression technique accelerates the process of convergence and leads to a black and white layout during the last stage when the step size multiplier reaches the ceiling. The scheme for step size multiplier modification is not fixed but can be adjusted in consideration of the compromise between the efficiency and reliability.

It should be noted that, though DGP method shows a better quality in all these numerical examples, it does not mean that the DGP method is better than SIMP for every case. In the first place, the higher value of objective function obtained by the SIMP method is partly attributed to the penalization of the intermediate material. Furthermore, owing to their successive searching feature, the solutions obtained by the DGP method and the SIMP method are both local optimum points, and the DGP method is only an alternative method that may lead to a local optimum closer to the global optimum.Finally, the DGP method proposed in this paper can only solve the problem with a single constraint and the applicability of the topology optimization problem with multiple constraints requires further research.

 Table 1 Results for the numerical examples with different methods

	Method	It.	C (Nmm)	M_{nd} (%)	CPU-time per iteration (s)
2D MBB beam	DGP	63	190.9	0.001	0.475
	SIMP	95	219.6	26.295	0.421
	MMA	175	219.4	26.119	—
3D cantilever	DGP	142	161.8	0.001	5.716
	SIMP	428	314.0	27.507	5.036
3D bridge	DGP SIMP	101 111	$\begin{array}{c} 3.572 \times 10^8 \\ 4.375 \times 10^8 \end{array}$	0.001 23.963	5.054 4.565

8 Conclusion

This paper presents a structural topology optimization algorithm using the direct gradient projection method with a transformation of variables technique. Transformation of variables eliminates the constraints of the design variables which brings great convenience to optimization. Once an acceptable step size multiplier β is decided, a feasible and useful step λS can be obtained by projecting the negative gradient of the objective function onto the hypersurface of the constraint directly without linear approximation or constraint correction procedure.

On one hand, some constraint measures are implemented on condition that the step size is too big or some design variables are in a state of oscillation. On the other hand, the step size multiplier β is magnified when the step size is too small. Therefore, the step size Δ is kept within a rational range which is considered to be balanced between efficiency and reliability. Besides, the grey elements suppression technique also plays a significant role in accelerating the convergence and reducing the intermediate material during the last stage of optimization process.

Results show that the proposed DGP method achieves a black and white solution with a lower value of objective function within fewer steps compared to the typical SIMP method with OC algorithm. By adding some positive interference to the rate of optimization process, the efficiency of the algorithm is improved without compromising the reliability.

Acknowledgements The first author appreciates the China Scholar Council (CSC) for supporting his visiting research experience in Delft University of Technology. The MATLAB code of MMA algorithm is provided by Krister Svanberg.

References

- Ambrosio L, Buttazzo G (1993) An optimal design problem with perimeter penalization. Calc Var Partial Dif 1(1):55–69
- Andreassen E, Clausen A, Schevenels M, Lazarov B, Sigmund O (2011) Efficient topology optimization in matlab using 88 lines of code. Struct Multidisc Optim 43(1):1–16

- Arora J (2004) Introduction to optimum design. Academic Press, California
- Bendsøe M (1989) Optimal shape design as a material distribution problem. Struct Multidisc Optim 1(4):193–202
- Bendsøe M, Sigmund O (2003) Topology optimization: theory, methods, and applications. Springer Verlag, Berlin
- Borrvall T, Petersson J (2001) Topology optimization using regularized intermediate density control. Comput Method Appl M 190(37):4911–4928
- Bourdin B (2001) Filters in topology optimization. Int J Numer Meth Eng 50(9):2143–2158
- Bruns T, Tortorelli D (2001) Topology optimization of non-linear elastic structures and compliant mechanisms. Comput Method Appl M 190(26):3443–3459
- Bruns T, Tortorelli D (2003) An element removal and reintroduction strategy for the topology optimization of structures and compliant mechanisms. Int J Numer Meth Eng 57(10):1413–1430
- Diaz A, Sigmund O (1995) Checkerboard patterns in layout optimization. Struct Multidisc Optim 10(1):40–45
- Fleury C (1989) Conlin: an efficient dual optimizer based on convex approximation concepts. Struct Multidisc Optim 1(2):81–89
- Haber R, Jog C, Bendsøe M (1996) A new approach to variabletopology shape design using a constraint on perimeter. Struct Multidisc Optim 11(1):1–12
- Huang X, Xie Y (2007) Convergent and mesh-independent solutions for the bi-directional evolutionary structural optimization method. Finite Elem Anal Des 43(14):1039–1049
- Huang X, Xie Y (2010) A further review of eso type methods for topology optimization. Struct Multidisc Optim 41(5):671–683
- Kim Y, Yoon G (2000) Multi-resolution multi-scale topology optimization—a new paradigm. Int J Solids Struct 37(39):5529–5559
- Lazarov B, Sigmund O (2011) Filters in topology optimization based on helmholtz-type differential equations. Int J Numer Meth Eng 86(6):765–781
- Michell A (1904) The limits of economy of material in frame structures. Phil Mag 8(47):589–597
- Poulsen T (2002) Topology optimization in wavelet space. Int J Numer Meth Eng 53(3):567–582
- Rao S (2009) Engineering optimization: theory and practice. Wiley, New Jersey
- Rosen J (1960) The gradient projection method for nonlinear programming. part i. linear constraints. J Soc Ind Appl Math 8(1):181–217
- Rosen J (1961) The gradient projection method for nonlinear programming. part ii. nonlinear constraints. J Soc Ind Appl Math 9(4):514– 532
- Rozvany G (1998) Exact analytical solutions for some popular benchmark problems in topology optimization. Struct Multidisc Optim 15(1):42–48
- Rozvany G (2009) A critical review of established methods of structural topology optimization. Struct Multidisc Optim 37(3):217–



(c) 3D bridge

3E8

10 20 30 40 50

Fig. 10 Iteration histories of topology optimization problems with different methods

60 70

Iteration

0.00

120

110

90 100

237

- Sigmund O (1997) On the design of compliant mechanisms using topology optimization. Mech Struct Mach 25(4):493–524
- Sigmund O (2007) Morphology-based black and white filters for topology optimization. Struct Multidisc Optim 33(4):401–424
- Sigmund O, Petersson J (1998) Numerical instabilities in topology optimization: a survey on procedures dealing with checkerboards, mesh-dependencies and local minima. Struct Multidisc Optim 16(1):68–75
- Svanberg K (1987) The method of moving asymptotes—a new method for structural optimization. Int J Numer Meth Eng 24(2):359–373
- Vanderplaats G, Moses F (1973) Structural optimization by methods of feasible directions. Comput Struct 3(4):739–755
- Xie Y, Steven G (1993) A simple evolutionary procedure for structural optimization. Comput Struct 49(5):885–896