Structural and Multidisciplinary Optimization (2021) 64:3227–3231
https://doi.org/10.1007/s00158-021-03004-z

EDUCATIONAL PAPER

Complementary lecture notes for teaching the 99/88-line topology optimization codes

Ming Zhou[1] · Ole Sigmund[2]

Received: 19 April 2021 / Revised: 23 June 2021 / Accepted: 23 June 2021 / Published online: 28 July 2021

Abstract
Sigmund's 2001 educational paper with a self-contained 99-line MATLAB code had far-reaching impact to teaching and research of topology optimization. This brief note aims to close the gaps on self-contained content desirable for classroom teaching. The goal is to add clarity to the theoretical foundation, and to enable students' learning of the complete iterative optimization solution with minimum additional effort.

Keywords Topology optimization · Optimality criteria · Convex approximation · Educational paper · Topology optimization code

1 Introduction

The 2001 educational paper by Sigmund (2001) had enormous impact: (1) It revolutionized teaching of topology optimization at all college levels with a very compact MATLAB code that is self-contained and easy to follow; (2) It built a simple code foundation for new researchers to gain insights and to experiment with their own alternative formulations; (3) It pioneered a new popular category of publications in the SMO field—educational articles that focus on self-contained compact code for teaching and research. At that time, most research work was based on lengthy Fortran or C code, which put a very high threshold for research experiments. It was a game changer to show that a complete research code including FEA can be built in MATLAB in such compact form. This opened the field to many graduate students, and led to vast increased research activities on topology optimization. Sigmund's 99-line code, and a computationally enhanced 88-line version by him and his associates (Andreassen et al. 2011), are today must-have material for university lectures on topology optimization. To date, these two papers have been downloaded over 36,000 times on the SMO journal's website, and likely more from their popular university website www.topopt.dtu.dk. Recent versions of the codes in Ferrari and Sigmund (2020) further enhanced computational efficiency and also added extension for 3D models. The tremendous benefit of learning a numerical approach through hands-on coding experience cannot be overstated.

While the above papers focused on self-contained content of the numerical process and code, three components were short of self-contained details: (1) the so-called optimality criteria variable update scheme, (2) a closed form update scheme for Lagrange multiplier as alternative to the bisection search implementation, and (3) simple and easy to follow derivation of compliance sensitivity. For undergraduate teaching, most students will not go beyond the text of the papers to study the underlying theory and derivation. The purpose of this very brief note is to fill these gaps with simplicity and clarity to allow students to build a complete theoretical foundation without additional literature. Especially on (1) we could make the case that deep insights about 'optimality criteria methods' developed in the 1970s had been lost even with the newer generation of researchers whose careers grew with topology optimization since the 1990s, despite of some recent contributions revisiting this subject (Groenwold and Etman 2010; Ferrari and Sigmund 2020; Kumar and Suresh 2020). It is important to note that

Responsible Editor: Gengdong Cheng

Special Issue dedicated to Dr. Raphael T. Haftka

✉ Ming Zhou
zhou@altair.com

[1] Altair Engineering, Irvine, CA, USA

[2] Department of Mechanical Engineering, Solid Mechanics, Technical University of Denmark, 2800 Lyngby, Denmark

there is a strong reason why we opt to publish this educational note in the special issue dedicated to Dr. Rafi Haftka. We will show that all these solution components are based on Haftka's original contributions in the 1970s. Optimality criteria methods typically refer to methods developed in the 1970s that built iterative schemes based on some heuristic rules. We will recite that the OC update formula in Andreassen et al. (2011) can be derived rigorously based on the convex approximation introduced by Starnes Jr and Haftka (1979). It is important to note that the concept of convex separable approximation provided an important foundation for the dual optimization methods widely used for topology optimization (Fleury and Braibant 1986; Svanberg 1987). Furthermore, exact Lagrange multiplier can be obtained within the same approximation framework. Haftka (1982) was also among the first researchers to present adjoint sensitivity analysis formulation in straightforward matrix form through manipulation of matrix multiplication sequence. To complete the lecture note, simple and straightforward derivation of compliance sensitivity is also given in Haftka's notation (also see Haftka and Gürdal (1991)). We want to emphasize that nothing is new in this note in terms of original research content. The sole purpose is to provide supplemental lecture notes for classroom teaching of the highly popular 99/88-line topology optimization codes especially to the benefit of educators, students and newcomers to the field.

2 Optimization problem and solution

2.1 Optimization problem

The optimization problem stated in Andreassen et al. (2011) is the following:

$$\min_{\mathbf{x}} \ : c(\mathbf{x}) = \mathbf{U}^T\mathbf{K}\mathbf{U} = \sum_{e=1}^{N} E_e(x_e)^p \mathbf{u}_e^T \mathbf{k}_0 \mathbf{u}_e \tag{1}$$

$$subject\ to: \ g = V(\mathbf{x}) - fV_0 = \sum_{e=1}^{N} v_{0e} x_e - fV_0 = 0$$

$$\mathbf{0} \le \mathbf{x} \le \mathbf{1}$$

and

$$\mathbf{KU} = \mathbf{F} \tag{2}$$

where c is the compliance, $\mathbf{U}$ and $\mathbf{F}$ are the global displacement and force vectors, respectively, $\mathbf{K}(\mathbf{x})$ is the global stiffness matrix, $\mathbf{u}_e$ is the element displacement vector, $\mathbf{k}_0$ is the element stiffness matrix for an element with unit Young's modulus, $\mathbf{x}$ is the vector of density variables, N is the number of elements, $V(\mathbf{x})$ is the material volume, V_0 and v_{0e} are the total and element design domain volume, respectively, and f is the prescribed target volume fraction.

2.2 Convex approximation

Convex approximation (Starnes Jr and Haftka 1979) of the optimization problem stated in (1) is expressed as linearization of a function in terms of direct variable x_e when gradient is positive, and in terms of reciprocal variable $1/x_e$ when gradient is negative. At the time, usage of reciprocal variable was already popular, which emerged from the observations that (1) stiffness relates to truss sizing variables linearly; (2) displacements relate to stiffness inversely; (3) hence, displacements relate to sizing variables inversely. It was proven that linearization of displacements in terms of reciprocal variables is, indeed, accurate if a truss structure is statically determinate. The motivation of Starnes and Haftka was to create a more conservation alternative formulation when dealing with highly nonlinear buckling and other responses of a composite wing structure. For this purpose, they invented the mixed variable approximation and coined it 'conservative approximation' as its feasible domain is a subset of either pure linear or pure reciprocal approximation. Denoting the starting point of each iteration as $\mathbf{x}_0$ the approximation for compliance in (1) is

$$\tilde{c}(\mathbf{x}) = c(\mathbf{x}_0) + \sum_{\frac{\partial c}{\partial x_e} \ge 0} \frac{\partial c}{\partial x_e}(x_e - x_{e0}) - \sum_{\frac{\partial c}{\partial x_e} < 0} \frac{\partial c}{\partial x_e} x_{e0}^2 \left(\frac{1}{x_e} - \frac{1}{x_{e0}}\right) \tag{3}$$

which reduces to the linear approximation in terms of reciprocal variables $1/x_e$ due to sensitivity of compliance being strictly negative as shown in Section 2.4:

$$\tilde{c}(\mathbf{x}) = c(\mathbf{x}_0) - \sum_{e=1}^{N} \frac{\partial c}{\partial x_e} x_{e0}^2 \left(\frac{1}{x_e} - \frac{1}{x_{e0}}\right) \tag{4}$$

Obviously, the gradients of the volume are all positive, resulting in the following expression for convex approximation of the volume constraint

$$\tilde{g}(\mathbf{x}) = V(\mathbf{x}_0) + \sum_{e=1}^{N} \frac{\partial V}{\partial x_e}(x_e - x_{e0}) - fV_0 = 0 \tag{5}$$

From (1), we have $\frac{\partial V}{\partial x_e} = v_{0e}$, and therefore $V(\mathbf{x}_0)$ cancels out of the sum involving x_{e0}

$$\tilde{g}(\mathbf{x}) = \sum_{e=1}^{N} \frac{\partial V}{\partial x_e} x_e - fV_0 = \sum_{e=1}^{N} v_{e0} x_e - fV_0 = 0 \tag{6}$$

which is, obviously, also the exact volume constraint expressed in (1) as the volume constraint is a linear function of $\mathbf{x}$.

2.3 Optimality conditions

Replacing compliance and volume in (1) with approximate functions (4)–(5), the optimality condition of the explicit approximate optimization problem is the well-known Kuhn-Tucker condition:

$$\frac{\partial \tilde{c}}{\partial x_e} + \lambda \frac{\partial \tilde{g}}{\partial x_e} = \frac{\partial c}{\partial x_e} \frac{x_{e0}^2}{x_e^2} + \lambda \frac{\partial V}{\partial x_e} = 0, \quad e = 1, \ldots, N \tag{7}$$

which yields

$$x_e = x_{e0} \left(\frac{\bar{B}_e}{\lambda} \right)^{1/2}, \quad with \quad \bar{B}_e = \left(\frac{-\frac{\partial c}{\partial x_e}}{\frac{\partial V}{\partial x_e}} \right) \tag{8}$$

In the above, we intentionally separated the Lagrange multiplier λ from the notation B_e in Andreassen et al. (2011) to keep λ as a visible unknown so we can derive its closed form expression later. To ensure approximation quality, it is customary to limit the range of variable changes with a move limit m, with 0.2 as default in Andreassen et al. (2011), so that

$$max(0, (x_{e0} - m)) \leq x_e \leq min(1, (x_{e0} + m)) \tag{9}$$

which can be enforced by treating variables limited by move bounds as passive variables:

$$x_e^{\text{new}} = \begin{cases} max(0, (x_{e0} - m)) & if \;\; x_e \leq max(0, (x_{e0} - m)) \\ min(1, (x_{e0} + m)) & if x_e \geq min(1, (x_{e0} + m)) \\ x_{e0} \left(\frac{\bar{B}_e}{\lambda} \right)^{1/2} & Otherwise \end{cases} \tag{10}$$

In the above equation, we call the case ‘Otherwise’ active variable set A, and the rest passive variable set P. The Lagrange multiplier can be obtained by substituting (10) into the approximate equality constraint (6):

$$\sum_{e \in A} \frac{\partial V}{\partial x_e} x_{e0} \left(\frac{\bar{B}_e}{\lambda} \right)^{1/2} + \sum_{e \in P} \frac{\partial V}{\partial x_e} x_e^{\text{new}} - f V_0 = 0 \tag{11}$$

which leads to

$$\left(\frac{1}{\lambda} \right)^{1/2} = \frac{f V_0 - \sum_{e \in P} \frac{\partial V}{\partial x_e} x_e^{\text{new}}}{\sum_{e \in A} \frac{\partial V}{\partial x_e} x_{e0} \bar{B}_e^{1/2}} \quad or$$

$$\lambda = \left(\frac{\sum_{e \in A} \frac{\partial V}{\partial x_e} x_{e0} \bar{B}_e^{1/2}}{f V_0 - \sum_{e \in P} \frac{\partial V}{\partial x_e} x_e^{\text{new}}} \right)^2 \tag{12}$$

Equation 12 shows that the Lagrange multiplier satisfying the volume constraint can be calculated analytically once the active and passive variable sets are defined. Since this is not known a priori, we start with assuming all variables being active. Then, a subset of variables will fall outside of the lower and upper bounds as defined in (10). The iterative process converges when the variable sets stop changing, which typically takes just a few iterations. The convergence can be further accelerated when the initial active variable set is inherited from previous iteration. In a recent educational note Kumar and Suresh (2020) also derived the analytical solution of the Lagrange multiplier as shown above. However, they did not embed this part within the complete analytical optimal solution of a convex approximate problem. It should also be noted that Ferrari and Sigmund (2020), with reference to Christensen and Klarbring (2008), covered all the essential components from approximation to analytical expression of Lagrange multipliers in an Appendix. Therefore, the main differences of this note to Ferrari and Sigmund (2020) are as follows: (1) Our derivation is based on convex approximation though it is equivalent to reciprocal approximation for compliance. Convex approximation offers preferred mathematical properties that eliminate the so-called duality gap, and hence became the foundation for the well-established and widely used CONLIN and MMA dual algorithms by Fleury and Braibant (1986) and Svanberg (1987); (2) Detailed formulation is given in a more transparent fashion to benefit students without prior knowledge; (3) As shown below, we also cover the more flexible form with power η of the optimality criteria expression in Andreassen et al. (2011) though it can also be found in Groenwold and Etman (2010). In short, our focus is to provide a self-contained and transparent account of all optimization details to benefit students and newcomers.

Andreassen et al. (2011) implemented a bisection algorithm to solve the Lagrange multiplier satisfying volume constraint. Ferrari and Sigmund (2020) showed that the exact update of λ does help reduce computational time, at a slight cost of the compactness of the code. When multiple inequality constraints are involved, the solution framework is the same although solution of Lagrange multipliers require far more sophisticated algorithms (Fleury and Braibant 1986; Svanberg 1987).

The above derivation adhered strictly to the convex approximation as first proposed by Starnes Jr and Haftka (1979), which led to the OC updating formula in Andreassen et al. (2011) with the default choice of $\eta = 0.5$. Below, we will briefly show that the case with an arbitrary η can be easily constructed within the approximation framework. Let's introduce approximation of compliance as linear expansion in terms of $1/x_e^\alpha$ (Groenwold and Etman 2010; Christensen and Klarbring 2008; Haftka 1982). With simple chain ruling we have $\frac{\partial c}{\partial (1/x_e^\alpha)} = \frac{\partial c}{\partial x_e} \left(\frac{-1}{\alpha} \right) x_{e0}^{(1+\alpha)}$. Then, the compliance approximation takes the following form:

$$\tilde{c}(\mathbf{x}) = c(\mathbf{x}_0) + \sum_{e=1}^{N} \frac{\partial c}{\partial x_e} \left(\frac{-1}{\alpha} \right) x_{e0}^{(1+\alpha)} \left(\frac{1}{x_e^\alpha} - \frac{1}{x_{e0}^\alpha} \right) \tag{13}$$

Keeping approximation of the volume constraint in the same linear form as expressed in (6), the optimality condition becomes

$$\frac{\partial \tilde{c}}{\partial x_e} + \lambda \frac{\partial \tilde{g}}{\partial x_e} = \frac{\partial c}{\partial x_e} x_{e0}^{(1+\alpha)} x_e^{-(1+\alpha)} + \lambda \frac{\partial V}{\partial x_e} = 0 \tag{14}$$

which yields

$$x_e = x_{e0} \left(\frac{\bar{B}_e}{\lambda} \right)^{\frac{1}{1+\alpha}}, \quad with \quad \bar{B}_e = \left(\frac{-\frac{\partial c}{\partial x_e}}{\frac{\partial V}{\partial x_e}} \right) \tag{15}$$

Denoting $\eta = \frac{1}{1+\alpha}$, we have the general form shown in Andreassen et al. (2011). It is easy to check that $\eta = 0.5$ corresponds to the standard convex approximation with $\alpha = 1$. $\eta = 0.7$ or 0.8, as recommended in Bendsøe and Sigmund (2004), correspond to $\alpha = (0.3/0.7)$ or 0.25, respectively. Intuitively, larger η enlarges the scaling effect of $\frac{\bar{B}_e}{\lambda}$, which on one hand might speed up convergence, but on the other hand could also potentially lead to oscillations and even divergence.

2.4 Sensitivity analysis

While Sections 2.2 and 2.3 provide transparency and insights that are relatively difficult to gather for students and researchers, derivation of sensitivity is rather straightforward, which can be readily found in the book by Bendsøe and Sigmund (2004). It is still worthwhile to add it to the lecture notes for completeness since it was left out in Sigmund (2001) and Andreassen et al. (2011). Here, we use the straightforward derivation from Haftka (1982) when he introduced a more efficient formulation for second order sensitivity. With $\mathbf{K}$ being symmetric, differentiation of (1) yields the following

$$\frac{\partial c}{\partial x_e} = \frac{\partial \mathbf{U}^T}{\partial x_e} \mathbf{K}\mathbf{U} + \mathbf{U}^T \frac{\partial \mathbf{K}}{\partial x_e} \mathbf{U} + \mathbf{U}^T \mathbf{K} \frac{\partial \mathbf{U}}{\partial x_e} = \mathbf{U}^T \frac{\partial \mathbf{K}}{\partial x_e} \mathbf{U} + 2\, \mathbf{U}^T \mathbf{K} \frac{\partial \mathbf{U}}{\partial x_e} \tag{16}$$

Differentiation of (2) leads to

$$\frac{\partial \mathbf{K}}{\partial x_e} \mathbf{U} + \mathbf{K} \frac{\partial \mathbf{U}}{\partial x_e} = 0 \quad \rightarrow \quad \frac{\partial \mathbf{U}}{\partial x_e} = -\mathbf{K}^{-1} \frac{\partial \mathbf{K}}{\partial x_e} \mathbf{U} \tag{17}$$

Substituting (17) into (16) leads to

$$\frac{\partial c}{\partial x_e} = \mathbf{U}^T \frac{\partial \mathbf{K}}{\partial x_e} \mathbf{U} - 2\, \mathbf{U}^T \mathbf{K}\mathbf{K}^{-1} \frac{\partial \mathbf{K}}{\partial x_e} \mathbf{U} = -\mathbf{U}^T \frac{\partial \mathbf{K}}{\partial x_e} \mathbf{U} \tag{18}$$

Since x_e only affects $\mathbf{k}_e$, with (1) the above can be reduced to the following form given in Andreassen et al. (2011) with $E_e(x_e) = E_{min} + x_e^p (E_0 - E_{min})$

$$\frac{\partial c}{\partial x_e} = -p x_e^{p-1} (E_0 - E_{min}) \mathbf{u}_e^T \mathbf{k_0} \mathbf{u}_e \tag{19}$$

From (19), we can see that derivatives of compliance are all negative as $\mathbf{k_0}$ is positive semi-definite. This property confirms our engineering intuition—increase of element densities always leads to less compliant, or stiffer, structure.

3 Conclusion

This brief educational note aims to accomplish the goal of completing the learning experience of the 99/88-line codes with self-contained content. The iterative solution process builds on solving a series of convex approximations of the original problem. Exact closed form solution of the approximate problem is derived in simple and straight forward form that can be effortlessly followed by undergraduate students. It is hoped that the lecture note can enrich students' learning journey by including the beautiful passages of the optimization algorithm itself, with very little additional effort. In particular, the derivation of the 'optimality criteria' that are often regarded as heuristic can even be a valuable exercise for experienced researchers as these details might arguably be lost treasures for many.

Another important motivation of this note is to showcase the tremendous contributions from Dr. Rafi Haftka to the foundation of structural optimization. The first author had the fortune to co-author a paper with Rafi (Zhou and Haftka 1995) on the exact subject matter—insights into optimality criteria methods. Both authors have benefited tremendously from collaborations and intellectual exchanges with Rafi. His research legacy will forever last, and he will live in our hearts forever.

Funding Ole Sigmund acknowledges the support of the Villum Foundation through the InnoTop project.

Declarations

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

Replication of results This brief note provides complimentary theory and formulation details for teaching the classic 99/88-line codes and the recently updated version Top99Neo. Readers should use this note in conjunction with studying and running these codes.

References

Andreassen E, Clausen A, Schevenels M, Lazarov BS, Sigmund O (2011) Efficient topology optimization in matlab using 88 lines of code. Struct Multidiscip Optim 43(1):1–16

Bendsøe MP, Sigmund O (2004) Topology optimization - theory methods and applications. Springer Verlag, Berlin Heidelberg

Christensen PW, Klarbring A (2008) An introduction to structural optimization, vol 153. Springer Science & Business Media, New York

Ferrari F, Sigmund O (2020) A new generation 99 line matlab code for compliance topology optimization and its extension to 3d. Struct Multidiscip Optim 62(4):2211–2228

Fleury C, Braibant V (1986) Structural optimization: a new dual method using mixed variables. Int J Numer Methods Eng 23(3):409–428
Groenwold AA, Etman L (2010) A quadratic approximation for structural topology optimization. Int J Numer Methods Eng 82(4):505–524
Haftka RT (1982) Second-order sensitivity derivatives in structural analysis. AIAA J 20(12):1765–1766
Haftka RT, Gürdal Z (1991) Elements of structural optimization, vol 11. Springer Science & Business Media, New York
Kumar T, Suresh K (2020) Direct lagrange multiplier updates in topology optimization revisited. Struct Multidiscip Optim pp 1–16
Sigmund O (2001) A 99 line topology optimization code written in matlab. Struct Multidiscip Optim 21(2):120–127
Starnes Jr JH, Haftka RT (1979) Preliminary design of composite wings for buckling, strength, and displacement constraints. J Aircr 16(8):564–570
Svanberg K (1987) The method of moving asymptotes—a new method for structural optimization. Int J Numer Methods Eng 24(2):359–373
Zhou M, Haftka R (1995) A comparison of optimality criteria methods for stress and displacement constraints. Comput Methods Appl Mech Eng 124(3):253–271