Struct Multidisc Optim (2018) 57:463–467
https://doi.org/10.1007/s00158-017-1840-z

BRIEF NOTE

Highly efficient density-based topology optimization using DCT-based digital image compression

Pingzhang Zhou[1] · **Jianbin Du**[2] · **Zhenhua Lü**[1]

Received: 14 July 2017 / Revised: 27 August 2017 / Accepted: 19 October 2017 / Published online: 7 November 2017

Abstract In this brief note we show that the number of design variables in density-based topology optimization can be phenomenally reduced using discrete cosine transform (DCT), which is one of the most frequently used transforms in digital image compression. Only quite a few nonzero DCT coefficients corresponding to low frequency components are needed to generate optimized topology with high resolution. Through two examples, one for compliance minimization and the other for heat conduction, we show that the density method can be surprisingly efficient than people have thought. Moreover, there is no need to use additional density filter or sensitivity filter since high frequency components are inherently filtered by the DCT-based compression.

Keywords Density method · DCT · Digital image compression

✉ Jianbin Du
dujb@tsinghua.edu.cn

Pingzhang Zhou
zpz13@mails.tsinghua.edu.cn

Zhenhua Lü
lvzh@tsinghua.edu.cn

[1] Department of Automotive Engineering, Tsinghua University, Beijing, People's Republic of China

[2] School of Aerospace Engineering, Tsinghua University, Beijing, People's Republic of China

1 A brief introduction on DCT

The density-based topology optimization shares many similarities with digital image processing. The elemental fictitious density in 2D rectangular mesh can be seen as the pixels in a still digital image:

$$\boldsymbol{\rho} = [g_{xy}] \tag{1}$$

where $x \in \{0, 1, \cdots, N_X - 1\}$, $y \in \{0, 1, \cdots, N_Y - 1\}$; N_X and N_Y are the number of elements in global X and Y directions. Throughout this paper, bold letter stands for matrix. By the way, we follows the tradition in digital image processing to count a sequence beginning with 0.

While in 3D case, the spatial pixels can be expressed as:

$$\boldsymbol{\rho} = [g_{xyz}] \tag{2}$$

where $x \in \{0, 1, \cdots, N_X - 1\}$, $y \in \{0, 1, \cdots, N_Y - 1\}$, $z \in \{0, 1, \cdots, N_z - 1\}$.

There are many orthogonal transforms in digital image processing to map the spatial domain to another domain (called transformed domain). For example, the well-known DFT (Discrete Fourier Transform) pair maps between the spatial domain $[g_{xy}]$ and the frequency domain $[F_{uv}]$. For purpose of image compression, DCT pair gives (for 2D case):[1]

$$G_{uv} = \sum_{x,y} \left\{ \alpha_u \alpha_v g_{xy} \cos \frac{\pi(2x+1)u}{2N_X} \cos \frac{\pi(2y+1)v}{2N_Y} \right\} \tag{3}$$

$$g_{xy} = \sum_{u,v} \left\{ \alpha_u \alpha_v G_{uv} \cos \frac{\pi(2x+1)u}{2N_X} \cos \frac{\pi(2y+1)v}{2N_Y} \right\} \tag{4}$$

[1] Equation (3) is called DCT while (4) is called IDCT.

where $x, u \in \{0, 1, \cdots, N_X-1\}$; $y, v \in \{0, 1, \cdots, N_Y-1\}$; α_u is the constant defined by:

$$\alpha_u = \begin{cases} \sqrt{1/N_X}, & \text{if } u = 0 \\ \sqrt{2/N_X}, & \text{otherwise.} \end{cases} \tag{5}$$

and similarly for α_v.

By using DCT, the transform coefficients G_{uv} for a digital image g_{xy} can be obtained. Then only a small proportion of G_{uv} (where $u = 0, 1, \cdots, n_X$; $v = 0, 1, \cdots, n_Y$; $n_X \ll N_X-1, n_Y \ll N_Y-1$) corresponding to low frequency components are retained while the rest is discarded. This way the memory needed to store an image is greatly decreased. When displaying an image, the stored nonzero coefficients are transformed back into spatial domain by IDCT together with zero-padding procedure.

The DCT pair holds the following good properties so can be borrowed and used in structural topology optimization:

(1) It has explicit and friendly analytical expression so that sensitivity information can be easily derived.
(2) It is free from Gibbs phenomenon and Runge phenomenon under equispaced grid.
(3) It has fast computation algorithms.[2] So the DCT and IDCT can be computed quite efficiently *without* explicitly computing (3), (4).
(4) It is orthogonal and separable. So 2D and 3D DCTs can be implemented through a series of 1D DCT.

Gibbs pheonmenon originally refers to the fact that Fourier series fails to converge uniformly at discontinuities. Recall that by using Fouries series, a function is implicitly extended periodically. DCT can avoid Gibbs phenomenon because it actually computes the DFT of $2N$ points for a sequence of N points by periodically and symmetrically extension.

We comment here that the more common Fourier, Chebyshev and Lengendre polynomials in mathematics community are all suffered from severe Gibbs phenomenon and hereby cannot be directly used to describe the layout of structures. Gegenbauer and Jacobi polynomials can find ways to get rid of Gibbs phenomenon (Gottlieb and Shu 1997) but they have very complicated expressions and are inefficient to compute. The algebraic polynomial interpolation suffers from Runge phenomenon under equispaced grid so is also inappropriate.

DCT is one of the most frequently used transformations for image compression and is the basis for the image(e.g. JPEG) and audio(e.g. MP3) compression standard (Gonzalez and Woods 2007). More details on DCT and its wide applications can be found in Rao and Yip (1990).

[2]There are a lot of fast algorithms for DCT. However, the most convenient one is to use *dct*, *dctmtx* or *dct2* in MATLAB.

2 Mathematical model for DCT-based topology optimization

In traditional density-based topology optimization, the number of design variables is equal to the number of elements in FEA model. Following the idea of digital image compression, now the elemental fictitious densities can be seen as spatial pixels ranging in [0, 1]. The pixels can then be approximated by the IDCT using much fewer variables.

Since a digital image behaves much more like a 2D structure than 3D structure, we will give all formulas based on 2D structural topology optimization. Formulas in 3D case can be easily generalized from 2D formulas.

Let $\boldsymbol{x} = [x_0, x_1, \cdots, x_{n-1}]^\top \in \mathbb{R}^n$ denote the actual design variable, $n = n_X n_Y$; $n_X \ll N_X$, $n_Y \ll N_Y$ denote the number of coefficients in global X and Y directions. Then the full transform coefficients $\boldsymbol{G} \in [0, 1]^{N_X \times N_Y}$ in frequency domain can be obtained by padding zeros:

$$\boldsymbol{G} = T(\boldsymbol{x}) = \left[\begin{array}{ccc|c} x_0 & \cdots & x_{n-n_Y} & \\ \vdots & \ddots & \vdots & \boldsymbol{0} \\ x_{n_Y-1} & \cdots & x_{n-1} & \\ \hline \multicolumn{3}{c|}{\boldsymbol{0}} & \boldsymbol{0} \end{array}\right] \tag{6}$$

Equivalently one can write:

$$G_{uv} = \begin{cases} x_k, & \text{if } u = \text{mod}(k, n_Y),\ v = \frac{k-u}{n_Y} \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

where $k = 0, 1, \cdots, n-1$.

From (6) it can be seen that now the design variables are actually the *nonzero transform coefficients* in frequency domain. By using the IDCT, the pixels in spatial domain $g \in \mathbb{R}^{N_X \times N_Y}$ can be reconstructed using these nonzero DCT coefficients (with the help of zero-padding):

$$\boldsymbol{g} = \text{IDCT}(\boldsymbol{G}) = \text{IDCT}(T(\boldsymbol{x})) \tag{8}$$

However, due to the zero-padding process, there is no guarantee that the computed spatial pixel g_{xy} still ranges in [0, 1]. It is possible that $g_{xy} < 0$ or $g_{xy} > 1$. Here the smooth Heaviside function is applied to construct the elemental fictitious densities (Wang et al. 2011; Xu et al. 2010):

$$\begin{aligned} \bar{g}_{xy} &= \text{Proj}(g_{xy}) \\ &= \begin{cases} 0, & \text{if } g_{xy} < 0 \\ 1, & \text{if } g_{xy} > 1 \\ \dfrac{\tanh(\beta\eta) + \tanh[\beta(g_{xy} - \eta)]}{\tanh(\beta\eta) + \tanh[\beta(1 - \eta)]}, & \text{otherwise.} \end{cases} \end{aligned} \tag{9}$$

where $\eta = 0.5$ is the threshold for fictitious densities, β is a factor that can be gradually increased from 1 until 128 to drive $\bar{g}_{xy}$ to approach $\{0, 1\}$ (Wang et al. 2011).

Then the elemental fictitious densities can be used in the SIMP interpolation formula. The following procedures are precisely the same as oridinary density methods so are omitted here.

Now it is time to formally give the mathematical formulation for DCDM (i.e. DCT compression based density method):

$$\begin{aligned} &\text{find} && \boldsymbol{x} = [x_0, x_1, \cdots, x_{n-1}] \in \mathbb{R}^n \\ &\text{minimize} && f(\bar{\boldsymbol{g}}) \\ &\text{s.t.} && \begin{cases} \bar{\boldsymbol{g}} = \text{Proj}(\text{IDCT}(T(\boldsymbol{x}))) \\ \sum\limits_{x,y} \bar{g}_{xy} V_{xy} \le \gamma \sum\limits_{x,y} V_{xy} \end{cases} \end{aligned} \tag{10}$$

Since all the equations are given in analytical expressions, the sensitivity information can be easily derived by using chain rule:

$$\frac{\partial f}{\partial x_k} = \sum_{x,y} \frac{\partial f}{\partial \bar{g}_{xy}} \frac{\partial \bar{g}_{xy}}{\partial g_{xy}} \frac{\partial g_{xy}}{\partial G_{uv}} \tag{11}$$

where f can be the objective or constraint function; $k \in \{0, 1, \cdots, n-1\}$; $x \in \{0, 1, \cdots, N_X - 1\}$, $y \in \{0, 1, \cdots, N_Y - 1\}$; $u = \text{mod}(k, n_Y)$, $v = \frac{k-u}{n_Y}$; $\frac{\partial \bar{g}_{xy}}{\partial g_{xy}}$ can be directly derived from (9); $\frac{\partial g_{xy}}{\partial G_{uv}}$ can be directly derived from (4). Notice that $\frac{\partial g_{xy}}{\partial G_{uv}}$ is nonzero only if $u \in \{0, 1, \cdots, n_X - 1\}$ and $v \in \{0, 1, \cdots, n_Y - 1\}$ while $\frac{\partial \bar{g}_{xy}}{\partial g_{xy}}$ is nonzero only if $g_{xy} \in (0, 1)$.

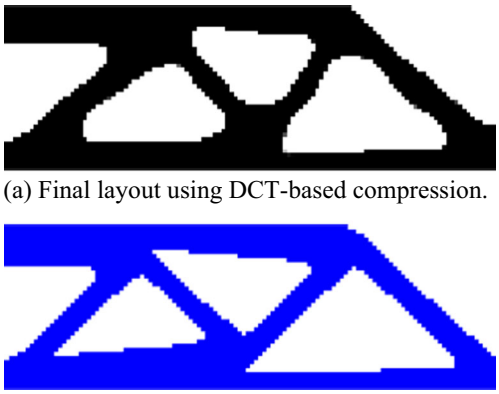
If one notices that the (3) and (4) can be written in matrix form, after some linear algebra one can deduce that in fact $\frac{\partial g_{xy}}{\partial G_{uv}}$ is constant so as to be independent of x or G. This way the sensitivity information can be calculated in a quite efficient manner.

3 Numerical examples

All the numerical results are computed using MATLAB 2017a on a Dell Precision T5610 workstation with two Intel(R) Xeon(R) CPU E5-2609 v2 processors. The *e04wd* function from NAG library[3] is selected as the optimizer, which we find quite efficient and robust. The convergence control rule for all examples here is: (a) major optimality tolerance $r \le 1 \times 10^{-5}$; (b) major iteration limit $n \le 150$ for each β (cf. (9)).

3.1 MBB beam

The MBB beam is an example that has been extensively used as benchmarks in structural topology optimization area. The beam is symmetric in X axis and simply supported at its two ends. The dimensions are: half span 120, height 40. Material parameters are: $E = 1$, $\nu = 0.3$. Half load $0.5P = 50$ is applied at the top center. The objective is to minimize the static compliance. Final topologies are presented in Fig.1 while detailed numerical information is listed in Table 1.



(a) Final layout using DCT-based compression.

(b) Final layout using traditional method.

Fig. 1 Final layout for MBB beam obtained by DCT-based compression (upper) and traditional density method (lower). $N_X = 120$, $N_Y = 40$, $n_X = 12$, $n_Y = 5$

Notice that the final layout obtained by the proposed DCT-based compression has no sharp corners when compared with the one obtained by traditional method. This is reasonable since sharp corner in spatial domain corresponds to extremely high frequency components in frequency domain. Recall that in general sharp corner would lead to stress concentration, so the layout obtained by DCT-based compression is superior when considering the stress of the structure.

3.2 Heat conduction

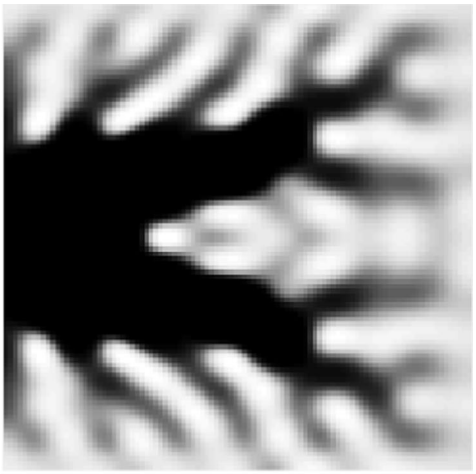
The second example is the heat conduction of a square plate. Length of each side is 100. Heat sink is located at the middle of left side. Uniform internal heat source $Q = 0.01$ is applied on the whole plate. Thermal conductivity is $\kappa = 1$. The objective is to minimize the heat resistance. Final topologies are presented in Fig. 2 while detailed numerical information is listed in Table 1.

Time again, notice that no sharp corners are present in the final layout obtained by DCT-based compression. Also, no tiny branches can be seen in the final layout obtained by our method since tiny details correspond to high frequency components as well.

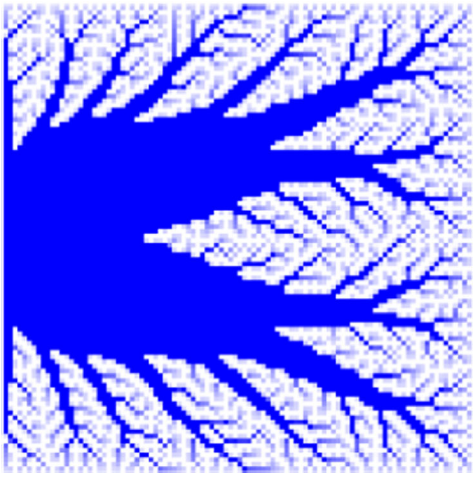
Table 1 Summary of the tested numerical examples

Case	NO. variables	Obj. func. value	Time(s)
MBB-DCT	60	4.89×10^5	58.596
MBB	4800	4.70×10^5	261.384
Heat-DCT	225	1.42×10^4	54.533
Heat	10000	1.19×10^4	599.931

[3]NAG is a commercial numerical library, cf. https://www.nag.com for more information.



(a) Final layout using DCT-based compression.



(b) Final layout using traditional method.

Fig. 2 Final layout for MBB beam obtained by DCT-based compression (upper) and traditional density method (lower). $N_X = 100$, $N_Y = 100$, $n_X = 15$, $n_Y = 15$

4 Discussion

4.1 Source of time saving

It is well-known that the computation cost for a topology optimization problem is composed of two parts: i.e. the cost for FEA programs and the cost for numerical optimization algorithms. Since reducing the number of the design variables would render the whole optimization problem more tractable, in general the number of iterations under some predefined KKT precision would be much fewer by using the DCT compression techniques. It follows that in our method the cost for not only the numerical optimization but also the FEA computation is reduced.

Table 2 Statistics on the computation cost

	Total t.(s)	FEA t.(s)	Optimize t.(s)	NO. FEA
MBB-DCT	58.596	30.853	7.362	203
MBB	261.384	77.461	113.992	513
Heat-DCT	54.533	8.078	11.985	173
Heat	599.931	21.943	456.672	473

In Table 2 some statistics[4] are listed to show the source of time-saving.

4.2 what do we lose?

DCT-based methods are all categorized as lossy compression in digital image compression. Similarly, when using the DCT-based compression in structural topology optimization we do lose something since no free lunch can be expected:

(1) For the originally convex problems (though they are quite rare in topology optimization), the new formulation is no longer convex. So no global optimality can be assured by using the proposed DCT-based compression.
(2) Each DCT coefficient G_{uv} has influence on the *whole* spatial pixels g. This is totally different from the traditional density-based methods where each elemental fictitious density has effect on *one* pixel only. So by using the DCT-based compression, it is difficult to carefully tune the value of some specific g_{xy}. It follows that the final objective function values are usually a bit larger than the ones obtained by traditional density-based method.

4.3 The links with MMC/MMV

Recently, Prof. Xu Guo and his co-workers proposed the MMC/MMV method (Guo et al. 2014, 2016; Zhang et al. 2017a, b) where the layout of structures are explicitly expressed using topology description function(TDF). Due to the explicit expression of structural topology, the variables needed are greatly reduced so that the optimization procedure when using MMC/MMV method can be quite efficient.

Moreover, we mention that in MMC/MMV the number of dofs can be reduced by ignoring the elemental stiffness matrix of void elements (Zhang et al. 2017a).

[4]The statistics are obtained using the *profile* provided within MATLAB.

The success of MMC/MMV method implies that to describe the layout of a structure, there is no need to use a huge number of design variables. Similar to the basic idea of MMC/MMV method that only a few number of variables are needed to achieve the topology optimization, in this paper we try to carry out highly efficient topology optimization method which is still rooted in the density method.

5 Conclusions

By borrowing the idea of digital image compression, the number of design variables can be greatly reduced in density-based topology optimization. The reduction in design variables would make the topology optimization problems more tractable for various numerical optimization methods. In addition, since all high frequency components in DCT coefficients are discarded, there is no need to use filters in the new formulation.

This paper also leaves some room for future research. For example, in digital image compression, the DCT does not directly operate on the whole image but on subimages(usually of size 8×8) that are subdivided from the original image. Also, the nonzero DCT coefficients which will be retained are not uniformly distributed in low frequency domain(cf. (6)) but selected based on information theory or threshold criteria. Both strategies can further improve the performance of the image compression. How to implement these strategies and how well they work remains to be answered by future research.

Acknowledgements The research is supported by NSFC (11772170, 11372154) which is gratefully acknowledged by the authors. The authors would also like to thank Mr. Daniel White with whom the helpful discussions during WCSMO12 at Braunschweig, Germany inspires the first author to come up with the basic idea of this paper.

References

Gonzalez RG, Woods RE (2007) Digital image processing. Publishing House of Electronics Industry, Beijing

Gottlieb D, Shu CW (1997) On the gibbs phenomenon and its resolution. SIAM Rev 39(4):644–668

Guo X, Zhang W, Zhong W (2014) Doing topology optimization explicitly and geometrically—a new moving morphable components based framework. ASME J Appl Mech 81(8):081009-1–081009-12

Guo X, Zhang W, Zhang J, Yuan J (2016) Explicit structural topology optimization based on moving morphable components (mmc) with curved skeletons. Comput Methods Appl Mech Eng 310:711–748

Rao KR, Yip P (1990) Discrete cosine transform: algorithms, advantages and applications. Academic Press Inc., San Diego

Wang F, Lazarov BS, Sigmund O (2011) On projection methods, convergence and robust formulations in topology optimization. Struct Multidiscip Optim 43(6):767–784

Xu S, Cai Y, Cheng G (2010) Volume preserving nonlinear density filter based on heaviside functions. Struct Multidiscip Optim 41(4):495–505

Zhang W, Chen J, Zhu X, Zhou J, Xue D, Lei X, Guo X (2017a) Explicit three dimensional topology optimization via moving morphable void (mmv) approach. Comput Methods Appl Mech Eng 322:590–614

Zhang W, Yang W, Zhou J, Li D, Guo X (2017b) Structural topology optimization through explicit boundary evolution. ASME J Appl Mech 84(1):011011-1–011011-10