Tensor Decompositions for Reducing the Memory Requirement of Translation Operator Tensors in FMM-FFT Accelerated IE Solvers

Cheng Qian, Zhuotong Chen, and Abdulkadir C. Yucel
School of Electrical and Electronic Engineering
Nanyang Technological University
Singapore, 639798
cqian@ntu.edu.sg, chenz313@myumanitoba.ca, acyucel@ntu.edu.sg

***Abstract*—Tensor decompositions are applied to lessen the memory requirements of translation operator tensors in fast multipole method-fast Fourier transform (FMM-FFT) accelerated integral equation (IE) solvers. In particular, methodologies leveraging Tucker and tensor train (TT) decompositions are developed to compress the three-dimensional (3D) arrays and four-dimensional (4D) array storing the *FFT'ed* translation operator values. Preliminary results show the achieved memory reduction as well as imposed computational overhead via the developed methodologies.**

Keywords—Fast Fourier transform, fast multipole method, memory reduction, translation operator, tensor decompositions, tensor train, Tucker decomposition.

I. Introduction

The FMM-FFT accelerated IE solvers permit efficient and accurate characterization of electromagnetic (EM) phenomena on electrically large and complex platforms [1-4]. For many practical scenarios, these solvers' applicability is limited by their memory requirement. One of the primary data structures limiting their applicability is the tensors holding the *FFT'ed* translation operator values on a structured grid. Recently, Tucker decompositions (a.k.a. high-order singular value decompositions) were proposed to lessen the memory requirement of *FFT'ed* translation operator tensors [5, 6]. These decompositions achieved more than 90% memory reduction in the memory requirement of these tensors and required negligible computational overhead. Such decompositions were recently used to reduce the memory requirement of *FFT'ed* system tensors arising in volume IE simulators as well [7, 8].

In [5], the memory reduction, computational overhead, and accuracy (i.e., performance metrics) of Tucker decompositions are investigated for compressing the *FFT'ed* translation operator tensors computed at each plane-wave direction and generated for different structure and FMM box sizes. That said, a detailed performance analysis is missing for the *FFT'ed* translation operator tensors generated for different wavenumbers and FMM accuracies. Furthermore, the performance of other decompositions for reducing the memory requirement of *FFT'ed* translation operator tensors is not analyzed so far. Moreover, the performance of neither Tucker nor other decompositions is demonstrated for compressing the 4D array holding the *FFT'ed* translation operator values on a structured grid for all plane-wave directions.

This study aims to fill the abovementioned gaps in the application of tensor decompositions to the compression of *FFT'ed* translation operator tensors. To this end, methodologies leveraging the Tucker and TT decompositions are developed for compressing 3D arrays storing the *FFT'ed* translation operator tensors for each plane-wave direction as well as 4D array storing the *FFT'ed* translation operator values for all plane-wave directions. During the setup stage of solvers, the Tucker and TT-based methodologies for compressing 3D arrays and 4D array, called Tucker-3D, TT-3D, Tucker-4D and TT-4D, respectively, are used to obtain compressed tensors efficiently via cross-based approximations [8, 9] for given tolerance. During the iterative solution stage of the solvers, the compressed tensors are *fully* or *partially* restored to obtain the *FFT'ed* translation operator tensor for each plane-wave direction. Then the restored tensor is used to perform the convolution for each plane-wave direction via 3D FFTs. Due to limited space, here we provide the general idea of methodologies and preliminary results. Details of the developed methodologies as well as their performances with respect to different parameters will be provided in the talk.

II. Formulation

In FMM-FFT acceleration scheme, a hypothetical box enclosing the structure is divided into N_x, N_y, and N_z small boxes along principle axes. The boxes are centered on a uniform 3D grid points and labeled by $B_{\mathbf{u}}$ with $\mathbf{u} = (u_x, u_y, u_z)$, $u_x = 1,\ldots,N_x, u_y = 1,\ldots,N_y, u_z = 1,\ldots,N_z$. In this scheme, while the interactions between basis functions in adjacent boxes are computed classically, those in the remaining boxes are accounted for as follows. First, the basis functions' far-field patterns along a plane-wave direction $\hat{\mathbf{k}}_i$, $i = 1,\ldots,N_{\text{dir}}$, are summed for each box and stored in aggregation tensor $\mathcal{F}_{\mathbf{u}}(\hat{\mathbf{k}}_i)$. Next, $\mathcal{F}_{\mathbf{u}}(\hat{\mathbf{k}}_i)$ is convolved with *FFT'ed* translation operator tensor $\mathcal{T}_{\mathbf{u}'-\mathbf{u}}(\hat{\mathbf{k}}_i)$ via 3D FFT operations. Finally, the plane-wave spectra for $\hat{\mathbf{k}}_i$ obtained as the result of convolution is projected onto the basis functions. The same procedure is applied for all N_{dir} directions and the far-field contribution to the matrix-vector multiplication is computed; here N_{dir} is a function of FMM accuracy, box size, and wavenumber [6].

$\mathcal{T}_{\mathbf{u'-u}}(\hat{\mathbf{k}}_i)$ with dimensions $2N_x \times 2N_y \times 2N_z$ is computed and stored during setup stage of the solvers. The memory requirement of $\mathcal{T}_{\mathbf{u'-u}}(\hat{\mathbf{k}}_i)$ for each $\hat{\mathbf{k}}_i$ can be significantly lowered by applying Tucker or TT decompositions as

$$\mathcal{T}_{\mathbf{u'-u}} \approx \mathcal{C}_{\text{Tucker}} \times_1 \bar{\mathbf{U}}^1_{\text{Tucker}} \times_2 \bar{\mathbf{U}}^2_{\text{Tucker}} \times_3 \bar{\mathbf{U}}^3_{\text{Tucker}}, \tag{1}$$

$$\mathcal{T}_{\mathbf{u'-u}} \approx \mathcal{C}_{\text{TT}} \times_1 \bar{\mathbf{U}}^1_{\text{TT}} \times_3 \bar{\mathbf{U}}^3_{\text{TT}}. \tag{2}$$

Here $\times_j$, $j=1,\ldots,3$, denotes the mode$-j$ matrix product of a tensor. The 3D array $\mathcal{C}$ and matrix $\bar{\mathbf{U}}^j$ along mode$-j$, pertinent to Tucker or TT decompositions (as indicated at the subscripts), can be efficiently obtained for a given tolerance by utilizing the cross approximation [8] and AMEN-cross approximation [9], respectively. Likewise, $\mathcal{T}_{\mathbf{u'-u}}(\hat{\mathbf{k}}_i)$ for all $\hat{\mathbf{k}}_i$, $i=1,\ldots,N_{\text{dir}}$, constitutes a 4D array, $\mathcal{T}_{\text{4D}}$, with dimensions $2N_x \times 2N_y \times 2N_z \times N_{\text{dir}}$, which can be compressed via Tucker or TT decompositions as

$$\mathcal{T}_{\text{4D}} \approx \mathcal{C}_{\text{Tucker}} \times_1 \bar{\mathbf{U}}^1_{\text{Tucker}} \times_2 \bar{\mathbf{U}}^2_{\text{Tucker}} \times_3 \bar{\mathbf{U}}^3_{\text{Tucker}} \times_4 \bar{\mathbf{U}}^4_{\text{Tucker}}, \tag{3}$$

$$\mathcal{T}_{\text{4D}} \approx (\mathcal{C}_{\text{TT1}} \times_1 \bar{\mathbf{U}}^1_{\text{TT}}) \times^1_3 (\mathcal{C}_{\text{TT2}} \times_3 \bar{\mathbf{U}}^3_{\text{TT}}). \tag{4}$$

Here ($\mathbf{A} \times^k_j \mathbf{B}$) represents the tensor contraction along mode$-j$ in $\mathbf{A}$ and mode$-k$ in $\mathbf{B}$. By effectively performing matrix-tensor products, the compressed tensor can be *fully* or *partially* restored to obtain the *FFT'ed* translation operator tensor for each plane-wave direction, during the iterative solution stage.

III. Numerical Result

In this section, EM scattering analysis of a perfect electric conductor sphere with unit radius is considered for comparing the performances of developed methodologies. The analysis is performed at 2.39 GHz and 4.79 GHz, at which the diameter of sphere becomes 16λ and 32 λ, respectively, where λ denotes the wavelength. For the analysis at these frequencies, once the FMM box size and accuracy are set to 0.5λ and 5 digits, there exist 435 *FFT'ed* translation operator tensors with dimensions 64×64×64 and 128×128×128, respectively ($N_{\text{dir}} = 435$). While Tucker-3D and TT-3D methodologies compress each $\mathcal{T}_{\mathbf{u'-u}}(\hat{\mathbf{k}}_i)$ for each $\hat{\mathbf{k}}_i$ separately, Tucker-4D and TT-4D methodologies compress $\mathcal{T}_{\text{4D}}$. In Fig.1, we plot the memory reductions achieved by these methodologies versus the relative error in the tensors restored after applying the decompositions. Clearly, TT-4D performs the best when very high accuracy (more than 10^{-6}) is needed. Both Tucker-4D and TT-4D methodologies yield higher memory savings compared to the Tucker-3D and TT-3D methodologies.

For the analysis at 4.79 GHz and relative error 10^{-6}, the TT-4D and Tucker-4D methodologies reduced the memory of original *FFT'ed* translation operator tensors from 13,920 MB to 473 MB and 541 MB, achieving 96.60% and 96.12% memory reduction, respectively. On the other hand, Tucker-3D reduced it to 1,007 MB, which is twice the total memory of compressed tensors obtained in TT-4D methodology. Moreover, TT-3D reduced it to 1,794 MB, corresponding to 87.11% memory reduction. For the same analysis, the computational overhead introduced by methodologies is computed by taking the ratio of the time required to restore one *FFT'ed* translation operator to the convolution time. While this ratio is 0.34 and 0.40 for TT-3D and Tucker-3D, that is 2.26 and 1.5 for TT-4D and Tucker-4D, respectively.

IV. Conclusion

The Tucker and TT-based methodologies are developed to reduce the memory requirement of *FFT'ed* translation operator tensors in FMM-FFT accelerated IE solvers. Preliminary results demonstrate the achieved memory saving and required computational overhead of the developed methodologies.

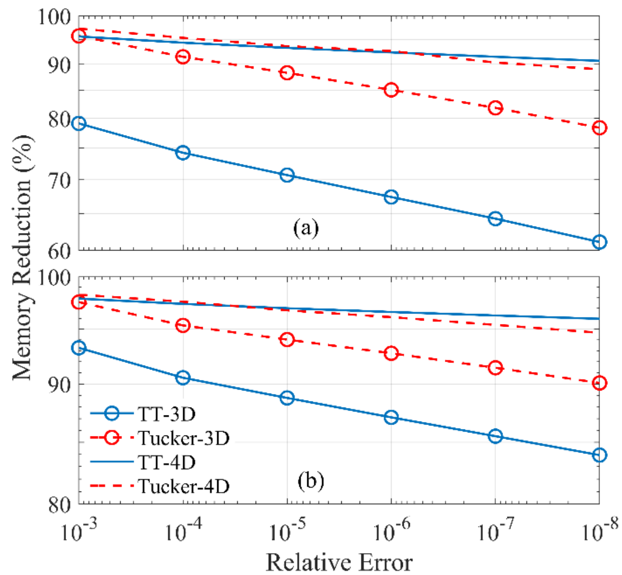

Fig. 1. The memory reduction w.r.t. the relative error when the dimensions of translation operator tensors are (a) $64\times64\times64$ and (b) $128\times128\times128$.

References

[1] R. L. Wagner, S. Jiming, and W. C. Chew, "Monte Carlo simulation of electromagnetic scattering from two-dimensional random rough surfaces," *IEEE Trans. Antennas Propag.*, vol. 45, no. 2, pp. 235–245, 1997.

[2] J. M. Taboada et al., "High scalability FMM-FFT electromagnetic solver for supercomputer systems," *IEEE Antennas Propag. Mag.*, vol. 51, no. 6, pp. 20–28, 2009.

[3] M. Vikram and B. Shanker, "An incomplete review of fast multipole methods-from static to wideband-as applied to problems in computational electromagnetics," *Appl. Computat. Electromagn. Society J.*, vol. 24, no. 2, pp. 79-108, 2009.

[4] A.C. Yucel et al., "An FFT-accelerated inductance extractor for voxelized structures," *2018 International ACES*, Denver, CO, pp. 1-2, 2018.

[5] A. C. Yucel, L. J. Gomez, and E. Michielssen, "Compression of translation operator tensors in FMM-FFT-accelerated SIE solvers via Tucker decomposition," *IEEE Antennas Wireless Propag. Lett.*, vol. 16, pp. 2667-2670, 2017.

[6] A. C. Yucel et al., "An FMM-FFT accelerated SIE simulator for analyzing EM wave propagation in mine enviroments loaded with conductors," *IEEE J. Multiscale and Multiphys. Comput. Techn.*, vol. 3, pp. 3-15, 2018.

[7] A. G. Polimeridis and J. K. White, "On the compression of system tensors arising in FFT-VIE solvers," in *Proc. IEEE Antennas Propag. Soc. Int. Symp.*, pp. 2144–2145, 2014.

[8] I. I. Giannakopoulos, M. S. Litsarev, and A. G. Polimeridis, "Memory footprint reduction for the FFT-based volume integral equation method via tensor decompositions," arXiv:1811.00484, 2018.

[9] M. V. Rakhuba and I. V. Oseledets, "Fast multidimensional convolution in low-rank tensor formats via cross approximation," *SIAM J. Sci. Comput.*, vol. 37, no. 2, pp. A565-A582, 2015.