

Fast and Accurate Randomized Algorithms For the Eigenvalue Problem in the Tensor Train Format

Christian Camano¹ and Roel Van Beeumen²

¹San Francisco State University

²Lawrence Berkeley National Laboratory

(Dated: 22 July 2022)

Abstract. We present an overview of tensor sketching for the eigenvalue problem. Our discussion evaluates the practicality of a number of randomized sketching techniques and their analogs in the tensor train format. Special attention is given to the sketched Rayleigh-Ritz method for the computation of eigenpairs and the generalized Nyström technique for random sketching. Figures demonstrating the structure of the modified tensor networks are given.

1. Introduction

One of the greatest challenges in contemporary numerical linear algebra research is finding efficient ways to represent high dimensional data. Due to restrictions in available physical memory and the exponential nature of many scientific problems there has been a push over the last decade to develop fast alternative solutions to the deterministic algorithms that preceded the current era of applied mathematics. Random matrix theory has proven to be an effective response to this problem and leverages the concept of dimensionality reduction to create near approximations to desired information by solving local problems in a lower dimensional vector space and projecting back into the problem space once a local answer has been computed.

Additionally, The Tensor Train (TT) format [1] can also aid in this domain by enabling the efficient expression of large matrices. Originally developed in the domain of physics TT has been used for some time under another name Matrix Product State (MPS).

This paper aims to explore some of the possible ways that TT can be used in conjunction with randomized algorithms to create a new class of fast and accurate eigensolvers that avoids the current precedent of reformatting the problem as a sum of small optimizations. Current iterative solvers in the TT format have a tendency to rely on sweeping algorithms that occasionally stagnate in local minima during optimization. We are currently investigating the theoretical side of this area of research with results and numerical simulations pending.

2. The Tensor Train format

In this section we describe an introduction to the *tensor train* (TT) format. The major idea of the TT format is to separate the modes of a tensor into d order-2 or order-3 tensors. The result is a tensor network that when contracted over the bond indices can be used to reform the original input.

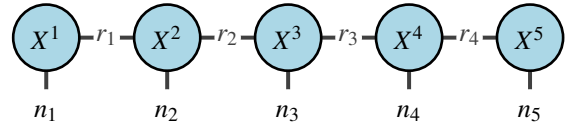
Definition 1 (Tensor Train Format) Let $X \in \mathbb{R}^{n_1 \times \dots \times n_d}$ be a tensor of order d . The decomposition of X

$$X = X^{(1)} \circ X^{(2)} \circ \dots \circ X^{(d-1)} \circ X^{(d)} \quad (1)$$

into component tensors $X^{(i)} \in \mathbb{R}^{r_{i-1} \times n_i \times r_i}$ for $i=1, \dots, d$ is called a tensor train (TT) representation of X . It is assumed

that $r_0 = r_N = 1$.

An example resultant tensor network of this format is shown for an order-5 tensor $X = X^{(1)} \circ X^{(2)} \circ X^{(3)} \circ X^{(4)} \circ X^{(5)}$ in the following diagram



This decomposition is traditionally formed by using successive singular value decompositions in which at each step a (matrix) SVD is performed to detach one open mode from the TT. For a complete description on the formation of a tensor train from a starting tensor see the work of Oseledets [1]. The space complexity for a TT format is $\mathcal{O}(nNr^2)$ where $r = \max(r_n)$, N is the order and $n = \max(n_k)$.

3. The Subsampled Random Fourier Transform

A key step to solving eigenproblems in the tensor train format using randomization is the selection of a sketching matrix to embed the overall problem into a lower dimensional subspace. Our selected sketching matrix for this embedding is the The Subsampled Random Fourier Transform (SRFT) [2]. This subspace embedding takes the form:

$$S = \sqrt{\frac{n}{s}} DFE \in \mathbb{C}^{s \times n}, \quad (2)$$

where

- s is the target embedding dimension
- $F \in \mathbb{C}^{n \times n}$ is the discrete Fourier transform matrix, $F_{j,k} = \frac{1}{\sqrt{n}} e^{-2\pi i(j-1)(k-1)/n}$;
- $E \in \mathbb{C}^{n \times n}$ is a diagonal matrix whose entries are independent Steinhaus random variables;
- $D \in \mathbb{C}^{s \times n}$ is a diagonal projector that randomly samples F and E

This subspace embedding is used to construct a sparse dimension reduction map which allows for the controlled construction of a lower dimensional problem space to approach the problem.

4. The sketched Rayleigh-Ritz method

Computing the dominant or first k eigen values of large dense matrices can be very computationally expensive. Consider the non symmetric eigen problem for a given system :

Identify nonzero x in \mathbb{C}^n and $\lambda \in \mathbb{C}$ such that

$$Ax = \lambda x, \quad A \in \mathbb{C}^{n \times n}$$

Computing the eigenproblem in the full described problem space: $\mathbb{C}^{n \times n}$ is extremely space inefficient and is one of the current restrictions that has motivated investigation into random methods. The sketched Rayleigh-Ritz method [4] approaches this problem by creating a lower dimensional Krylov subspace to project the problem into a subspace where it is less computationally expensive to approach the eigen problem. The classical description of the Rayleigh-Ritz(RR) method is best understood as a Galerkin method in which an approximate eigenpair here on referred to as a *Ritz pair* is computed by finding the solution to the ordinary eigen problem

$$B^*(AB)y = \theta y \quad y \neq 0, \theta \in \mathbb{C} \quad (3)$$

Where reduced matrix $AB \in \mathbb{C}^{n \times d}$ and full rank basis $B \in \mathbb{C}^{n \times d}$. From this point forward since the eigenpairs are only dependent on the column space of B an orthonormal basis $Q \in \mathbb{C}^{n \times d}$ is typically computed such that

$$Q^*(AQ)z = \theta z \quad z \neq 0 \quad (4)$$

where solutions to this equation (z, θ) are our ritzpairs. Alternatively, consider instead the minimization of the residual over the subspace.

$$\min_{M \in \mathbb{C}^{d \times d}} \|AB - BM\|_F \quad (5)$$

The solution to this problem matrix M_* frames the $d \times d$ eigenvalue problem

$$M_* y = \theta y \quad (6)$$

with each solution yielding an approximate eigenpair (By, θ) .

By sketching this minimization using (2) a lower dimension approximation of M_* can be computed hereby referred to as \hat{M}

$$\min_{M \in \mathbb{C}^{d \times d}} \|S(AB - B\hat{M})\|_F \quad (7)$$

with approximate eigenpairs $(B\hat{y}, \hat{\theta})$.

This procedure is referred to as the sketched Rayleigh Ritz method and has been proven to be both theoretically and empirically competitive with the results of conventional Rayleigh Ritz [3].

5. The eigenvalue problem expressed in TT

A well known eigenvalue algorithm that utilizes the tensor train format is the density-matrix renormalization group (DMRG) which computes the lowest wavefunction of a Hamiltonian. This information can be encoded within the extreme eigenvalues of the matrix as well and is often expressed using this perspective.

DMRG was first adapted for the TT format by Dolgolev and Oseletes [5] using a modified version of the alternating least squares method. The algorithm performs a sweeping motion over the Tensor Train solving small local eigenproblems until the Tensor Train vectors closely approximate the true eigenvector of the global system.

This algorithm, while very effective, is still largely dependent on the Locally Optimal Block Preconditioned Conjugate Gradient (LOBPCG) [6] method for the computation of the local solutions, creating a computational bottleneck when local problems themselves are large as well.

Alternative algorithms such as sRR theoretically could enable the computation of approximate eigenvectors without the need for an iterative approach that breaks the larger problem down into local systems. Such an algorithm would consist of uniform contraction between TT matrix cores and TT vector cores in the optimization of the Rayleigh quotient problem:

$$\min \text{trace}_X(X^*AX) \quad \text{s.t. } X^*X = I \quad (8)$$

where X is the the eigenvectors of the system expressed in TT vector format.

In DMRG the notion of a blocked tensor train core is used to allow for the computation of the trace in the local problem. This blocked core is a core of a tensor train in which an additional index b is added to a given core

$$\dots X^{(k-1)}(i_{k-1}) \hat{X}^{(k)}(i_k, b) X^{(k+1)}(i_{k+1}) \dots \quad (9)$$

The choice of which mode k is expressed in the block format changes as the tensor train is optimized with the selection walking back and forth along the tensor train in a sweeping motion until a tolerance criteria for the eigenvector approximation is met.

Two interface cores are constructed from the remaning tensor train to the left and right of the selected core that was re-expressed in block format as follows:

$$\begin{aligned} X^{<k}(\overline{i_1 i_2 \dots i_{k-1}}, \beta_{k-1}) &= \sum_{\alpha_1 \dots \alpha_{k-2}} X_{\alpha_1}^{(1)}(i_1) X_{\alpha_1 \alpha_2}^{(2)}(i_2) \dots X_{\alpha_{k-2} \beta_{k-1}}^{(k-1)}(i_{k-1}) \\ X^{>k}(\beta_k, \overline{i_{k+1} \dots i_{d-1} i_d}) &= \sum_{\alpha_{k+1} \dots \alpha_{d-1}} X_{\beta_k \alpha_{k+1}}^{(k+1)}(i_{k+1}) \dots X_{\alpha_{d-2} \alpha_{d-1}}^{(d-1)}(i_{d-1}) X_{\alpha_{d-1}}^{(d)}(i_d) \end{aligned} \quad (10)$$

Using the interfaces a frame matrix is then defined as

$$X_{\neq k} = X^{<k} \otimes I_{n_k} \otimes (X^{>k})^\top, \quad (11)$$

Using these newly formed tensor cores we can now re-express the original eigen problem as a local framed problem.

$$\min \text{trace}_X((X^k)^T X_{\neq k}^T A^k X_{\neq k} X^k) \quad (12)$$

Where A^k is the k^{th} core of the TT matrix A and the current site of local optimization

While the DMRG approach to this problem appears to be limited by its dependence of LOBPCG it does reveal a practical framing scheme for local problems that can be used when considering randomized techniques.

6. The Generalized Nyström method

Recent advances have extended the Nyström method to handle matrices that are not positive semi-definite opening the door to integration in random sketching methods [7]. The generalized Nyström method works by computing the following rank r approximation of a given matrix A , $A \in \mathbb{C}^{m \times n}$ and $\Omega \in \mathbb{C}^{n \times r}$ $\Psi \in \mathbb{C}^{m \times (r+l)}$ are gaussian random matrices with l as a small oversampling parameter to ensure the dominant column space is being properly sampled in the lower subspace.

$$A \approx \hat{A}_r = A\Omega(\Psi^T \tilde{A}\Omega)_e^\dagger \Psi^T A \quad (13)$$

Here $\tilde{A} = A + \delta A$ where $\|\delta A\| = \mathcal{O}(u\|A\|)$ with u being the unit roundoff.

One possible sketching scheme is as follows:
Let $\tilde{A} = X_{\neq k}^T A^k X_{\neq k}$

$$\tilde{A} \approx \tilde{A}_r = \tilde{A}\Omega(\Psi^T \tilde{A}\Omega)_e^\dagger \Psi^T \tilde{A} \quad (14)$$

reducing the local eigenproblem to a truncated version:

$$\min \text{trace}_X((X^k)^T \tilde{A}^k X^k) \quad (15)$$

This decomposition is a very efficient contender for random algorithms and is the scheme we are currently considering for future work regarding tensor contraction due to the similarities between two sided matrix multiplication in generalized Nyström and the two sided TT vector contraction in DMRG.

7. Current experiments and research

We are currently working on optimizing an implementation of sRR programmed in MATLAB before taking the

algorithm and expressing it in the TT format. It is possible that creating a TT format version of sRR could operate as a more efficient local solver than LOBPCG, meaning that a return to the DMRG algorithm may prove to be the most effective treatment of these new algorithms as sRR could be deployed instead. Alternatively, depending on the way sRR is implemented it could be possible to take an entire TT matrix and solve for a resulting TT vector through optimization of the Krylov subspace throughout sRR. Additionally there are algorithms that enable the efficient approximation of very large Tensor Trains [8] as well as efficient schemes for the randomized SVD (rSVD) [10] expressed in TT format [9]. If properly utilized the sRR algorithm could be further randomized by relying on a rSVD for orthogonalization instead of successive QR decompositions when building the Krylov subspace during the algorithm.

Finally, after our implementation of TT sRR is complete we are also considering sketching the entire tensor train prior to the computation of eigenpairs using the Nyström method to take advantage of low rank approximation.

8. Conclusion

There is a clear need for the development of numerically stable and accurate schemes for random sketching in the tensor train format. Tensor structure presents a clear complement to the computationally complexity improvements associated with randomized methods. Together they form a theoretical class of algorithms that are computationally efficient due to their low rank, and space efficient due to the structure of the TT format. The greatest challenge currently facing the development of algorithms of this kind is ensuring that tensor structure is preserved under operations such as random sketching in which bond dimensions between tensor cores can easily bifurcate. Our team is currently working to ammend this issue by first developing an optimized implementation of the sketched Rayleigh-Ritz algorithm with the intetion of bridging the gap with randomized algorithms and designing inputs that are sketched tensor train matrices.

If successful this algorithm would enable the computation of solutions to the eigenproblem without ever needing to form the full matrix and by solving a close low rank approximation instead harnessing the most desirable attributes of both classes of algorithms. The main benefit of using TT format in this context is the ability to avoid any potential stagnation in local minima as the concept of sweeping over cores would no longer me the method for computing the approximate eigenvector.

¹I. V. Oseledets. "Tensor-train decomposition" *SIAM J. Sci. Comput.*, 33 (2011), p. 2295–2317.

²Franco Woolfe, Edo Liberty, Vladimir Rokhlin, Mark Tygert, "A fast randomized algorithm for the approximation of matrices" *Applied and Computational Harmonic Analysis*, (2007),

³Y. Saad, Numerical methods for large eigenvalue problems, vol. 66 of Classics in Applied Mathematics, SIAM, Philadelphia, PA, 2011 ch1. Revised edition of the 1992 original

- ⁴Nakatsukasa, Yuji and Tropp, Joel A. Fast ; Accurate Randomized Algorithms for Linear Systems and Eigenvalue Problems” *arXiv*, (2021)
- ⁵S.V. Dolgov and B.N. Khoromskij and I.V. Oseledets and D.V. Savostyanov. Computation of extreme eigenvalues in higher dimensions using block tensor train format *Computer Physics Communications*, 26, (2014) p. 1207–1216.
- ⁶A. Knyazev. Toward the optimal preconditioned eigensolver: locally optimal block preconditioned conjugate gradient method.” *SIAM J. Scientific Computing*, 23(2):, (2001) p. 517-541.
- ⁷Nakatsukasa, Yuji. Fast and stable randomized low-rank matrix approximation” *arXiv*, (2020),
- ⁸Daas, Hussam Al and Ballard, Grey and Cazeaux, Paul and Hallman, Eric and Miedlar, Agnieszka and Pasha, Mirjeta and Reid, Tim W. and Saibaba, Arvind K.. Randomized algorithms for rounding in the Tensor-Train format” *arXiv*, (2021),
- ⁹Huber, Benjamin and Schneider, Reinhold and Wolf, Sebastian. A Randomized Tensor Train Singular Value Decomposition” *arXiv*, (2020),
- ¹⁰Halko, Nathan and Martinsson, Per-Gunnar and Tropp, Joel A.i. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions” *arXiv*, (2009),
- ¹¹Batselier, Kim, et al. “Computing Low-Rank Approximations of Large-Scale Matrices with the Tensor Network Randomized SVD.” *SIAM Journal on Matrix Analysis and Applications* 39, no. 3 (January 2018): 1221–44.