Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Seminar for Applied Mathematics

The tensor structure of a class of adaptive algebraic wavelet transforms

V. Kazeev and I. Oseledets

Research Report No. 2013-28
August 2013

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

The tensor structure of a class of adaptive algebraic wavelet transforms

Vladimir A. Kazeev* Ivan V. Oseledets†‡

August 22, 2013

Abstract

We revisit the *wavelet tensor train (WTT)* transform, an algebraic orthonormal wavelet-type transform based on the successive separation of variables in multidimensional arrays, or *tensors*, underlying the *tensor train (TT)* representation of such arrays. The TT decomposition was proposed for representing and manipulating data in terms of relatively few parameters chosen adaptively, and the corresponding low-rank approximation procedure may be seen as the construction of orthonormal bases in certain spaces associated to the data. Using these bases, which are extracted adaptively from a reference vector, to represent other vectors is the idea underlying the WTT transform. When the TT decomposition is coupled with the *quantization* of "physical" dimensions, it seeks to separate not only the "physical" indices, but all "virtual" indices corresponding to the virtual levels, or scales, of those. This approach and the related construction of the WTT transform are closely connected to hierarchic multiscale bases and to the framework of *multiresolution analysis*.

In the present paper we analyze the tensor structure of the WTT transform. First, we derive an explicit TT decomposition of its matrix in terms of that of the reference vector. In particular, we establish a relation between the ranks of the two representations, which govern the numbers of parameters involved. Also, for a vector given in the TT format we construct an explicit TT representation of its WTT image and bound the TT ranks of the representation. Finally, we analyze the sparsity of the WTT basis functions at every level and show the exponential reduction of their supports, from the coarsest level to the finest level, with respect to the level number.

Keywords: wavelet, low rank, tensor, tensor train, multiresolution analysis, virtual levels, quantized tensor train, quantization, tensorization, binarization.
AMS Subject Classification (2000): 15A69, 42C40, 65T60.

1 Introduction

Discrete wavelet transforms (DWTs), due to their optimality properties [1], play an important role in the efficient representation of vectors of various nature in the signal and image

*Seminar für Angewandte Mathematik, ETH Zürich. Rämistrasse 101, 8092 Zürich, Switzerland. vladimir.kazeev@sam.math.ethz.ch

†Skolkovo Institute of Science and Technology, Novaya Str. 100, 143025 Skolkovo, Moscow Region, Russia. i.oseledets@skolkovotech.ru

‡Institute of Numerical Mathematics, Russian Academy of Sciences. Gubkina Str. 8, 119333 Moscow, Russia. ivan.oseledets@gmail.com

processing, statistical estimation, data compression and denoising, and also in the numerical solution of PDEs. The generalizations of the Haar hierarchic orthogonal basis to smoother bases, orthogonal and biorthogonal (see, e.g., [2] and [3]), are examples of wavelet transforms customized to take into account the regularity of the data. In some cases this can be achieved by a more explicit construction of the *lifting scheme* [4, 5]. The latter consists in adjusting the wavelet function to design a *multiresolution analysis* [6, 7] with desired regularity properties. In particular, the generalization to the *second generation wavelets* [8] is notable for us, since that approach gives up the translation and dilation and, thus, the Fourier transform as the construction and analysis tool. In the context of PDEs, wavelets tailored to match the solutions in regularity and to achieve the optimal approximation rate were proposed and analyzed in a variety of papers; see, e.g., [9], [10, 11] and [12, 13, 14, 15].

However, a "class of data" may mean more than a certain regularity. Indeed, even within the variety of functions of prescribed regularity, one may be interested in the efficient representation of only those sharing more particular features. Constructing so finely customized wavelet bases and implementing the corresponding DWTs efficiently is a further challenge. A possible way to address it is the *wavelet tensor train (WTT)* transform, proposed recently in [16] and analyzed further in the present paper.

For a given reference vector (possibly multidimensional), a particular iteration through the *virtual levels*, or *scales*, of the vector (corresponding to each dimension) is performed. At each such step the image space of a certain matrix, which is associated to the reference vector and corresponds to the current level, is approximated by a subspace of smaller dimension. Such a step seeks to extract the most essential part of the data contained in the reference vector at the current level and to construct an orthogonal basis representing this part efficiently. The resulting orthogonal transform can be applied then at the same level of a dual recursive procedure to the counterpart matrix associated to another vector.

The dimension iteration mentioned above being a reinterpretation of the *tensor train (TT)* decomposition [17, 18], the whole procedure results in the WTT transform. This orthogonal algebraic transform projects the vector being transformed onto a basis adapted to the reference vector. As an algebraic transform, it may be constructed from and applied to consecutively refined discretizations of functions. Then the affinity of the reference and transformed vectors, resulting in the basis adapted to the former being well-suited for the latter, appears to be essentially different from the the matter of regularity of the underlying functions.

The TT decomposition is a non-linear low-parametric representation of multidimensional vectors, based on the separation of variables. The approximation of a multidimensional vector in the TT format can be performed by iterating through the dimensions of the vector successively and constructing an orthogonal low-rank factorization of a matrix related to the vector at each step. Currently the low-rank factorizations are constructed through SVD, although techniques of compressed sensing may be employed as well. To separate not only the "physical" dimensions of the vector, but all the virtual levels (scales), prior to the approximation procedure one subjects the vector to *quantization*, i.e. replaces each mode index with a multi-index the components of which are treated thenceforth as independent indices. This gives rise to the *quantized tensor train (QTT)* decomposition [19, 20, 21].

In one dimension the QTT approximation can be viewed simply as a non-stationary subdivision scheme for discontinuous piecewise-constant interpolation with weights chosen adaptively through the low-rank approximation of matrices. When the WTT transform is constructed with the use of quantization, so that it reinterprets the QTT decomposition, the resulting orthogonal basis adapted to the given reference vector is structured with respect to the virtual levels (scales) of the data.

In the present paper, first, we recapitulate the TT and QTT decompositions with related notation in section 2 and reintroduce the WTT transform in section 3. In section 4 we recast the WTT transform in a recursive form and obtain a recursive structure of the matrix. We also show the self-filtration property: if the transform is constructed so as to extract completely the tensor structure from the reference vector, the image of the reference vector contains at most one nonzero entry. In sections 5 and 6 we investigate the tensor structure of the matrix of the transform and of the images of tensor-structured vectors. We construct explicit QTT representations of those, which allows for efficient storage of the WTT basis and implementation of the transform of QTT-structured vectors. Finally, in section 7 we use the explicit representation of the WTT matrix derived in section 5 to analyze its sparsity. The result can be interpreted as the exponential reduction of the supports of the WTT basis vectors, from the coarsest level to the finest level, with respect to the level number.

2 TT and QTT decompositions

2.1 Tensors. Indexing

By a d-dimensional vector of size $N = n_1 \cdot \ldots \cdot n_d$ we mean a multidimensional array with d indices, the kth index taking values in the kth index set $\mathcal{I}_k = \{1, \ldots, n_k\}$. The entries of such a vector can be indexed by a multi-index $(j_1, \ldots, j_d)$ taking values in $\times_{k=1}^{d} \mathcal{I}_k$, as well as by a "long" scalar index $j = \overline{j_1, \ldots, j_d}$ taking values in $\mathcal{I} = \{1, \ldots, N\}$. Throughout the paper we use the isomorphism between the spaces of vectors indexed by the two index sets, given by the index transformation

$$j - 1 = \sum_{k=1}^{d} (j_k - 1) \cdot \prod_{\kappa=1}^{k-1} n_\kappa. \tag{1}$$

This corresponds to the column-major ordering of subindices, used in Fortran and MATLAB. For the notation convenience we identify the index sets $\times_{k=1}^{d} \mathcal{I}_k$ and $\mathcal{I}$, as well as the isomorphic spaces of vectors indexed by these sets. In this sense, we identify the indices $(j_1, \ldots, j_d)$ and $j = \overline{j_1, \ldots, j_d}$ and the vectors with entries $x_{j_1, \ldots, j_d}$ and $x_j = x_{\overline{j_1, \ldots, j_d}}$.

When a multidimensional array needs to be considered as a matrix (for example, for the low-rank factorization), we separate the row and column sizes by "×": in the pseudocode using the operation analogous to the MATLAB's `reshape`,

$$X = \texttt{reshape}(x, n_1 \cdot \ldots \cdot n_k \times n_{k+1} \cdot \ldots \cdot n_d) \tag{2}$$

is a matrix of size $n_1 \cdot \ldots \cdot n_k \times n_{k+1} \cdot \ldots \cdot n_d$ with the row index $j' = (j_1, \ldots, j_k)$ and the column index $j'' = (j_{k+1}, \ldots, j_d)$ understood in the sense of (1), the entries being

$$X^{(k)}{}_{\substack{j' \\ j''}} = X^{(k)}{}_{\substack{j_1, \ldots, j_k \\ j_{k+1}, \ldots, j_d}} = x_{j_1, \ldots, j_k, j_{k+1}, \ldots, j_d} = x_{j', j''}, \tag{3}$$

i.e. the same as in x.

Vectors and matrices with multiple indices are examples of multidimensional arrays, or *tensors*. Many *tensor decompositions* have been proposed to reduce the complexity of computations with tensors and to overcome the "curse of dimensionality" [22]: a broad review of related methods can be found in [23]. For a comprehensive survey of the tensor decompositions used for the representation of functions and numerical solution of operator equations, see [24].

2.2 TT decomposition

The TT decomposition of an array with a few indices represents it in terms of arrays with fewer indices, called *cores* of the decomposition. Among their indices we distinguish two *rank indices* and call the rest *mode indices*. We denote the entries of a core U with $U(\alpha;\, j;\, \beta)$, where $\alpha \in \{1, \ldots, p\}$ is the *left rank index*, $\beta \in \{1, \ldots, q\}$ is the *right rank index* and $j \in \{1, \ldots, n\}$ is the *mode index*, the three being separated by semicolons. A notation convention analogous to (1) applies to the latter when it represents a few subindices. The core U is then said to be of mode size n and of rank $p \times q$. If p or q is equal to one, we omit the corresponding index while referring to an entry of the core.

Consider a d-dimensional vector x of size $n_1 \cdot \ldots \cdot n_d$ and cores U_k of mode size n_k and of rank $r_{k-1} \times r_k$, where $1 \leq k \leq d$ and $r_0 = 1 = r_d$. If it holds for $(j_1, \ldots, j_d) \in \mathcal{I}$ that

$$x_{j_1,\ldots,j_d} = \sum_{\alpha_1=1}^{r_1} \ldots \sum_{\alpha_{d-1}=1}^{r_{d-1}} U_1(j_1;\, \alpha_1) \cdot U_2(\alpha_1;\, j_2;\, \alpha_2) \cdot \ldots \cdot U_{d-1}(\alpha_{d-2};\, j_{d-1};\, \alpha_{d-1}) \cdot U_d(\alpha_{d-1};\, j_d)\,, \tag{4}$$

then x is said to be represented in the TT format [17, 18] with *ranks* $r_1, \ldots, r_{d-1}$ in terms of the cores $U_1, \ldots, U_d$. Similarly, a d-dimensional matrix $\mathcal{W}$ of size $n_1 \cdot \ldots \cdot n_d \times n_1 \cdot \ldots \cdot n_d$ may have a TT representation of ranks $r_1, \ldots, r_{d-1}$ in terms of cores $V_1, \ldots, V_d$, given by

$$\mathcal{W}_{\substack{i_1,\ldots,i_d \\ j_1,\ldots,j_d}} = \sum_{\alpha_1=1}^{r_1} \ldots \sum_{\alpha_{d-1}=1}^{r_{d-1}} V_1(i_1, j_1;\, \alpha_1) \cdot V_2(\alpha_1;\, i_2, j_2;\, \alpha_2) \cdot \ldots \cdot V_{d-1}(\alpha_{d-2};\, i_{d-1}, j_{d-1};\, \alpha_{d-1}) \cdot V_d(\alpha_{d-1};\, i_d, j_d) \tag{5}$$

for all $(i_1, \ldots, i_d)\,, (j_1, \ldots, j_d) \in \mathcal{I}$. This particular definition of a TT representation of a matrix reflects the idea of the separation of variables in the matrix-vector multiplication.

Note that if the TT ranks are equal to one, the TT representation (5) reduces to a Kronecker product:

$$\mathcal{W} = V_1 \otimes V_2 \otimes \ldots \otimes V_{d-1} \otimes V_d,$$

where the cores are of rank 1×1 and are considered as usual matrices. In this sense, (5) is a generalization of the standard low-rank approximation of high-dimensional operators [25, 26].

A TT decomposition, exact or approximate, can be constructed through the low-rank factorization of a sequence of single matrices; for example, with the use of the SVD. In particular, for $1 \leq k < d$ equality (4) implies a rank-r_k factorization of an *unfolding matrix* $X^{(k)}$ given by (3). On the other hand, once x is such a vector that the unfolding matrices $X^{(1)}, \ldots, X^{(d-1)}$ are of ranks $r_1, \ldots, r_{d-1}$ respectively, then the cores $U_1, \ldots, U_d$ satisfying (4) exist; see Theorem 2.1 in [18]. The ranks of the unfolding matrices are the lowest possible ranks of a TT decomposition of the vector. They are hence referred to as the *TT ranks of the vector*.

For the matter of approximation it is crucial that if the unfolding matrices can be approximated with ranks $r_1, \ldots, r_{d-1}$ and accuracies $\varepsilon_1, \ldots, \varepsilon_{d-1}$ in the Frobenius norm, then the vector itself can be approximated in the TT format with ranks $r_1, \ldots, r_{d-1}$ and accuracy $\sqrt{\sum_{k=1}^{d-1} \varepsilon_k^2}$ in the ℓ_2-norm. This underlies a robust and efficient algorithm for the low-rank TT approximation of vectors given in full format or in the TT format with higher ranks. For details see Theorem 2.2 with corollaries and Algorithms 1 and 2 in [18]. A similar remark applies to matrices, provided that their approximation accuracy is measured in the Frobenius norm.

Note that the TT representation essentially relies on a certain ordering of the dimensions and *reordering indices may affect the values of the TT ranks significantly.*

2.3 Quantization. Dimensions and virtual levels. QTT decomposition

In section 2.1 we discussed the transformation of a multi-index into a "long" scalar index and back. These operations arise routinely when multidimensional vectors are manipulated or even just stored in memory, but by the subindices of multi-indices one typically means indices related to the "physical" dimensions of the problem. However, if the kth mode size n_k can be factorized as $n_k = n_{k1} \cdot \ldots \cdot n_{kl_k}$ in terms of l_k integral factors $n_{k1}, \ldots, n_{kl_k} \geq 2$, then the kth mode index j_k varying in $\mathcal{I}_k$ can be replaced with a multi-index $(j_{k1}, \ldots, j_{kl_k})$, each "virtual" index $j_{k\kappa}$ taking values in $\mathcal{I}_{k\kappa} = \{1, \ldots, n_{k\kappa}\}$. Under the notation convention of (1), the *virtual levels (dimensions)* represent the scales in data, so that among the "virtual" indices corresponding to the same "physical" dimension the outmost left represents the finest scale and the outmost right, the coarsest scale. A tensor with its virtual levels being considered as distinct indices is called a *quantization* of the original tensor.

The idea of separating virtual levels with the use of tensor decompositions came up in [27]. Having been applied to the TT decomposition, it resulted in the *quantized tensor train* representation proposed for matrices in [19, 20] and further elaborated in the vector case in [21]. In this sense (4) and (5) also present QTT representations of ranks $r_1, \ldots, r_{d-1}$ of a vector and matrix, provided that $j_1, \ldots, j_d$ and $i_1, \ldots, i_d$ are the "virtual" indices corresponding to a single or a few "physical" dimensions. By a *QTT decomposition* of a tensor and the *QTT ranks of the decomposition* we mean a TT decomposition of its quantization and the ranks of that TT decomposition. By treating the tensor being approximated as a higher-dimensional object, the QTT decomposition seeks to separate more indices and extract more structure.

When the natural ordering

$$\underbrace{j_{1,1}, \ldots, j_{1,l_1}}_{\text{1st dimension}}, \underbrace{j_{2,1}, \ldots, j_{2,l_2}}_{\text{2nd dimension}}, \ldots \ldots, \underbrace{j_{d,1}, \ldots, j_{d,l_d}}_{d\text{th dimension}} \tag{6}$$

of the "virtual" indices is used for representing the quantized vector in the TT format, the ranks of the QTT decomposition can be enumerated as follows:

$$\underbrace{r_{1,1}, \ldots, r_{1,l_1-1}}_{\text{1st dimension}}, r_1, \underbrace{r_{2,1}, \ldots, r_{2,l_2-1}}_{\text{2nd dimension}}, r_2, \ldots \ldots, r_{d-1}, \underbrace{r_{d,1}, \ldots, r_{d,l_d-1}}_{d\text{th dimension}},$$

where $r_1, \ldots, r_{d-1}$ are the TT ranks of the original tensor, i.e. the ranks of the separation of "physical" dimensions.

In the present paper we work mostly in the general framework of the TT format and make no difference whether the indices being separated correspond to "physical" dimensions or virtual levels. In the rest of the paper we consider d abstract indices and, as opposed to (6), number them from 1 to d. Our results are valid for the TT representation and the related WTT transform constructed with the use of any quantization or no quantization. To interpret the result on the sparsity of the WTT matrix, given at the end of section 7, one may consider the particular case of the *ultimate binary quantization* with all $n_{k,\kappa} = 2$.

2.4 Core product and MPS notation

The scheme of separation of variables, given in 4, has been known as *Matrix Product States (MPS)* and exploited by physicists to describe quantum spin systems for two decades (see [28, 29], cf. [30]). The MPS notation suggests an interpretation of the right-hand side of (4) as the matrix product $U_1^{(i_1)} \cdot U_2^{(i_2)} \cdot \ldots \cdot U_{d-1}^{(i_{d-1})} \cdot U_d^{(i_d)}$ of a row, $d-2$ matrices and a column indexed by rank indices $\alpha_1, \ldots, \alpha_{d-1}$ and depending also on the mode indices $i_1, \ldots, i_d$

as parameters. Similarly, the notation of *Matrix Product Operators* [31] reads the right-hand side of (5) as a product of a row, $d-2$ matrices and a column depending on mode indices $i_1, j_1, \dots, i_d, j_d$ as parameters.

In the present paper we omit the mode indices with the use of the core notation presented below in order to analyze the rank structure of the WTT matrix and images. Our calculations make use of the notions of *core matrices* and the product introduced in [32, 33] as the *rank core product*. When a core is considered as a two-level matrix with rank and mode levels, the rank core product coincides with the *strong Kronecker product* proposed in [34] for block matrices.

In this section we consider cores with two mode indices. When the second mode size equals 1, the second mode index can be omitted. This corresponds to the cases of (4) and (5) respectively.

Consider a TT core U of rank $p \times q$ and mode size $m \times n$. Assume that $m \times n$-matrices $A_{\alpha\beta}$, $\alpha = 1, \dots, p$, $\beta = 1, \dots, q$ are TT *blocks* of the core U, i. e. $U(\alpha;\, i, j;\, \beta) = (A_{\alpha\beta})_{ij}$ for all values of rank indices α, β and mode indices i, j. Then the core U can be considered as the matrix

$$\begin{bmatrix} A_{11} & \cdots & A_{1q} \\ \vdots & \vdots & \vdots \\ A_{p1} & \cdots & A_{pq} \end{bmatrix}, \tag{7}$$

which we refer to as the *core matrix* of U. In order to avoid confusion we use parentheses for ordinary matrices, which consist of numbers and are multiplied as usual, and square brackets for cores (core matrices), which consist of blocks and are multiplied by means of the rank core product "$\bowtie$" defined below. Addition of cores is meant elementwise. Also, we may think of $A_{\alpha\beta}$ or any submatrix of the core matrix in (7) as *subcores* of U.

Definition 1 (strong Kronecker product). *Consider cores U_1 and U_2 of ranks $r_0 \times r_1$ and $r_1 \times r_2$, composed of blocks $A^{(1)}_{\alpha_0\alpha_1}$ and $A^{(2)}_{\alpha_1\alpha_2}$, $1 \le \alpha_k \le r_k$ for $0 \le k \le 2$, of mode sizes $m_1 \times n_1$ and $m_2 \times n_2$ respectively. Let us define the* strong Kronecker product $U_1 \bowtie U_2$ *of U_1 and U_2 as a core of rank $r_0 \times r_2$, consisting of blocks*

$$A_{\alpha_0\alpha_2} = \sum_{\alpha_1=1}^{r_1} A^{(1)}_{\alpha_0\alpha_1} \otimes A^{(2)}_{\alpha_1\alpha_2}, \quad 1 \le \alpha_0 \le r_0, \quad 1 \le \alpha_2 \le r_2,$$

of mode size $m_1 m_2 \times n_1 n_2$.

In other words, we define $U_1 \bowtie U_2$ as a usual matrix product of the two corresponding core matrices, their elements (blocks) being multiplied by means of the Kronecker (tensor) product. For example,

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \bowtie \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} =$$
$$\begin{bmatrix} A_{11} \otimes B_{11} + A_{12} \otimes B_{21} & A_{11} \otimes B_{12} + A_{12} \otimes B_{22} \\ A_{21} \otimes B_{11} + A_{22} \otimes B_{21} & A_{21} \otimes B_{12} + A_{22} \otimes B_{22} \end{bmatrix}.$$

The representation (5) may be recast with the use of the rank core product in the following way: $\mathcal{W} = V_1 \bowtie V_2 \bowtie \dots \bowtie V_{d-1} \bowtie V_d$. The transpose $\mathcal{W}^\top$ of $\mathcal{W}$ is equal to the rank core product of the same cores, their blocks being transposed. If we consider another matrix $\mathcal{W}' = V_1' \bowtie \dots \bowtie V_d'$ of the same mode size, then a linear combination of $\mathcal{W}$ and $\mathcal{W}'$ can be written in the following way:

$$\alpha\mathcal{W} + \beta\mathcal{W}' = \begin{bmatrix} V_1 & V_1' \end{bmatrix} \bowtie \begin{bmatrix} V_2 & \\ & V_2' \end{bmatrix} \bowtie \dots \bowtie \begin{bmatrix} V_{d-1} & \\ & V_{d-1}' \end{bmatrix} \bowtie \begin{bmatrix} \alpha V_d \\ \beta V_d' \end{bmatrix};$$

and the tensor product of $\mathcal{W}$ and $\mathcal{W}'$, as $\mathcal{W} \otimes \mathcal{W}' = V_1 \bowtie \ldots \bowtie V_d \bowtie V_1' \bowtie \ldots \bowtie V_d'$.

Also, every matrix $\mathcal{W}$ may be regarded as a core of rank 1×1, and the rank core product then coincides with the Kronecker (tensor) product when applied to such cores.

3 WTT transform

Let us demonstrate the basic idea behind the WTT transform, introduced in [16] in the case of a single dimension and of the ultimate binary quantization described in section 2.3. We consider a one-dimensional vector a of size $N = 2^d$. To separate its d "virtual" indices, the quantization of a can be approximated by the TT representation of the form (4) with the use of the TT-SVD algorithm [18, Algorithm 1]. At the first step of the approximation procedure, the first unfolding A_1 of $a \equiv a_1$ has to be considered. It is a $2 \times 2^{d-1}$ matrix with elements

$$A_1 = \begin{pmatrix} a_1 & a_3 & \ldots & a_{N-1} \\ a_2 & a_4 & \ldots & a_N \end{pmatrix}. \tag{8}$$

The SVD factorizes A_1 as follows:

$$A_1 = U_1 S_1 V_1.$$

The diagonal matrix S_1 contains the singular values of A_1. The second singular value can be small if the two rows of A_1 are almost collinear. Then we conclude that the vector admits a QTT representation with the first rank equal to 1.

In any case, the exact or truncated SVD of a rank r_1 chosen to ensure the desired accuracy is used to proceed to the second step with the vectorization of $\widetilde{A}_1 = {U_1}^\top A_1$ as the input. The latter is of size $r_1 \cdot 2^{d-1}$. The QTT approximation algorithm continues in a similar way with reshaping it to a matrix A_2 of size $r_1 \cdot 2 \times 2^{d-2}$ and approximating it by a matrix of rank r_2 with the use of the SVD, and so on.

However, the construction of the WTT filters follows a different course. Instead of choosing r_k depending on the singular values comprising S_k, we may use a prescribed value. The singular values of A_k with numbers greater than r_{k+1} may still be large, but nevertheless omitted in the transition to the next step. Then the columns of the truncated factor U_k are used as the basis for the kth level of a dual procedure of transforming another vector.

For example, when a vector of ones is taken for a reference vector, the filters constructed with ranks $r_1 = \ldots = r_{d-1} = 1$ are

$$U_k = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

and the WTT basis coincides with the Haar hierarchic orthogonal basis.

It should be mentioned that the class of WTT transforms does not include many important discrete wavelet transforms; for example, the Daubechies wavelet transforms. On the other hand, the WTT transform allows for a very flexible adaptivity, which is still to be studied.

Let us give a brief numerical illustration of applying the WTT transform to the same vector it is adapted to. We consider two functions: $f(x) = \sqrt{x}$, $x \in [0, 1]$, and $f(x) = \sin 500x^2$, $x \in [0, 1]$. We discretize them by sampling on a uniform grid with 1024 points. Both the functions do not have exact low-rank representations and produce vectors of relatively large QTT ranks. In particular, QTT approximations of the corresponding vectors with ranks equal to $r = 2$, $r = 3$, or $r = 4$ are very coarse. The WTT filters are computed with ranks $r_0 = 1, r_1 = \ldots = r_{d-1} = r$ for $r = 2$, $r = 3$, and $r = 4$. These examples do not satisfy Theorem 4 given below, and the filters carry only a part of information contained in the

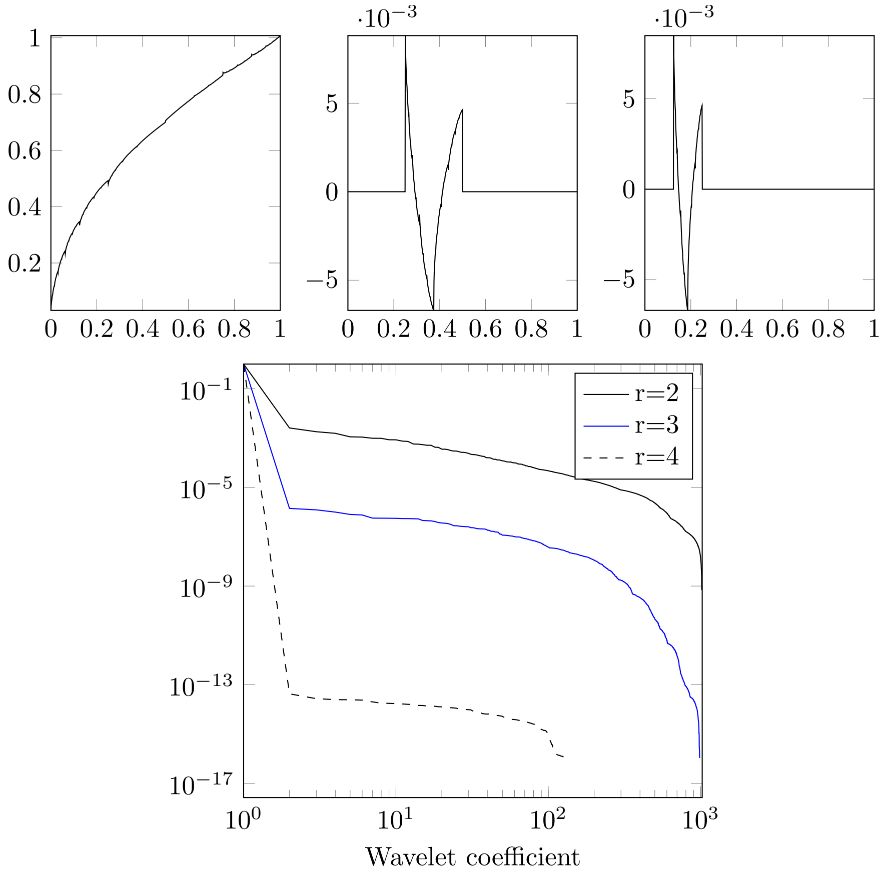

Figure 1: $f(x) = \sqrt{x}$, $x \in [0, 1]$. Top: the first three WTT basis functions for $r = 2$. Bottom: Decay of the absolute values of the wavelet coefficients for different r.

reference vector. The first discrete basis functions obtained through the procedure are presented in Figures 1 and 2. We also plot the decay of the absolute values of the wavelet coefficients sorted by the absolute value.

4 Recursive structure of the WTT transform

In Algorithm 1 we describe in a recursive way the construction of the WTT filters adapted to a given vector [16, Algorithm 2]. Once the filters are available, the WTT transform can be formally applied to any other vector of appropriate size [16, Algorithm 3]. We present the recursive computation of the WTT transform with given filters in Algorithm 2.

It is reasonable to choose the filter ranks $r_1, \dots, r_{d-1}$ so that $r_{k-1}n_k \leq n_{k+1} \dots n_d$ for $1 \leq k \leq d-1$. This ensures that each but last of the filters constructed at line 2 of Algorithm 1 as the left SVD factors of the corresponding unfolding matrices A_k contain only the vectors from the column spaces of the corresponding matrices.

The linear transform defined by Algorithm 2 orthogonal [16, Lemma 3.1], therefore it can be associated with a matrix $\mathcal{W}$. For the chosen truncation ranks $r_1, \dots, r_{d-1}$ the matrix is uniquely defined by the tensor $a \equiv a_1$ from which it is constructed in Algorithm 1.

Lemma 2. *For $1 \leq k < d$ consider the kth recursion level of Algorithm 2. Let us define the*

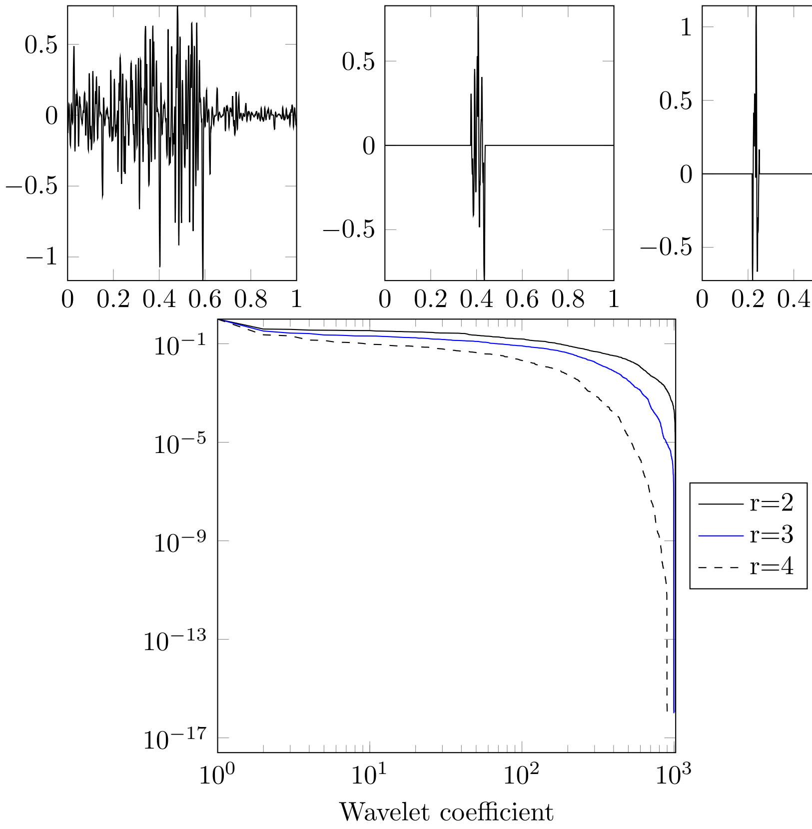

Figure 2: $f(x) = \sin 500x^2$, $x \in [0,1]$. Top: three first WTT basis functions for $r = 2$. Bottom: Decay of the absolute values of the wavelet coefficients for different r

matrix

$$E_k = \begin{pmatrix} I \\ 0 \end{pmatrix}$$

of size $r_{k-1} \cdot n_k \times r_k$, so that ${E_k}^\top$ selects the first r_k entries of a vector of size $r_{k-1} \cdot n_k$. Then the following recursive relation holds true for the matrix of the WTT transform:

$$\mathcal{W}_k = (E_k \otimes I_k)\, \mathcal{W}_{k+1} \left({R_k}^\top \otimes I_k\right) + {P_k}^\top \otimes I_k,$$

where $P_k = U_k \left(I - E_k {E_k}^\top\right)$ is a square matrix of order $r_{k-1} \cdot n_k$, $R_k = U_k E_k$ is a matrix of size $r_{k-1} \cdot n_k \times r_k$ and I_k is the identity matrix of order $n_{k+1} \cdot n_{k+2} \cdot \ldots \cdot n_d$.

Proof. The lines 5–6 and 10 of Algorithm 2 imply $\widehat{X}_{k+1} = {E_k}^\top \widetilde{X}_k = {R_k}^\top X_k$ and

$$Y_k = E_k \widehat{Y}_{k+1} + \left(I - E_k {E_k}^\top\right) \widetilde{X}_k = E_k \widehat{Y}_{k+1} + {P_k}^\top X_k, \tag{9}$$

where $\widehat{Y}_{k+1}$ is a matricization of the WTT image y_{k+1} of x_{k+1}, the entries of which are given by

$$y_{k+1\;\alpha_k, i_{k+1}, \ldots, i_d} = \sum_{\beta_k=1}^{r_k} \sum_{j_{k+1}, \ldots, j_d} \mathcal{W}_{k+1\;\substack{\alpha_k, i_{k+1}, \ldots, i_d \\ \beta_k, j_{k+1}, \ldots, j_d}}\; x_{k+1\;\beta_k, j_{k+1}, \ldots, j_d}.$$

Algorithm 1 Construction of WTT filters
$(U_k, \ldots, U_d) = \mathcal{F}_k\left(a_k;\, n_k, \ldots, n_d;\, r_{k-1}, \ldots, r_{d-1}\right)$

Require: recursion level k: $1 \leq k \leq d$
Require: mode sizes $n_k, \ldots, n_d$ and filter ranks $r_{k-1}, \ldots, r_{d-1}$
Require: $r_{k'} \leq r_{k'-1} n_{k'}$ for $k \leq k' < d$
Require: a vector a_k of size $r_{k-1} \cdot n_k \cdot \ldots \cdot n_d$
Ensure: WTT filters $U_k, \ldots, U_d$ of mode sizes $n_k, \ldots, n_d$ and ranks $r_{k-1}, \ldots, r_{d-1}$ respectively, adapted to the vector a_k

```
A_k = reshape(a_k, r_{k-1} · n_k × n_{k+1} · ... · n_d)
A_k = U_k S_k V_k {SVD}
{U_k, as the left SVD factor of A_k, is a square unitary matrix of order r_{k-1} · n_k}
if k < d then
    Ã_k = U_k^T A_k = S_k V_k
    Â_{k+1} = Ã_k(1 : r_k , :)
    a_{k+1} = reshape(Â_{k+1}, r_k · n_{k+1} · ... · n_d)
    (U_{k+1}, ..., U_d) = F_{k+1}(a_{k+1}; n_{k+1}, ..., n_d; r_k, ..., r_{d-1})
end if
```

Therefore the two terms of Y_k in (9) can be written elementwise as follows:

$$\left(E_k \widehat{Y}_{k+1}\right)_{\substack{\alpha_{k-1}, i_k \\ i_{k+1}, \ldots, i_d}} = \sum_{\alpha_k=1}^{r_k} E_{k\,\substack{\alpha_{k-1}, i_k \\ \alpha_k}} \, y_{k+1\,\alpha_k, i_{k+1}, \ldots, i_d} = \sum_{\alpha_k=1}^{r_k} E_{k\,\substack{\alpha_{k-1}, i_k \\ \alpha_k}}$$
$$\cdot \sum_{\beta_k=1}^{r_k} \sum_{j_{k+1}, \ldots, j_d} \mathcal{W}_{k+1\,\substack{\alpha_k, i_{k+1}, \ldots, i_d \\ \beta_k, j_{k+1}, \ldots, j_d}} \sum_{\beta_{k-1}=1}^{r_{k-1}} \sum_{j_k=1}^{n_k} R_{k\,\substack{\beta_{k-1}, j_k \\ \beta_k}} \, x_{k\,\beta_{k-1}, j_k, j_{k+1}, \ldots, j_d} \tag{10}$$

and

$$\left(P_k{}^{\top} X_k\right)_{\substack{\alpha_{k-1}, i_k \\ i_{k+1}, \ldots, i_d}} = \sum_{j_k=1}^{n_k} P_{k\,\substack{\beta_{k-1}, j_k \\ \alpha_{k-1}, i_k}} \, x_{k\,\beta_{k-1}, j_k, i_{k+1}, \ldots, i_d}, \tag{11}$$

which completes the proof, as the output vector y_k is the vectorization of Y_k. □

Corollary 3. *Let us consider the matrices introduced in Lemma 2 and, assuming $P_d = U_d$, define also a square matrix $\overline{P}_k = P_k \otimes I_k$ of order $r_{k-1} \cdot n_k \cdot \ldots \cdot n_d$ and matrices*

$$\begin{aligned} \overline{E}_k &= I \cdot (E_1 \otimes I_1) \cdot (E_2 \otimes I_2) \cdot (E_{k-1} \otimes I_{k-1}) \\ \overline{R}_k &= I \cdot (R_1 \otimes I_1) \cdot (R_2 \otimes I_2) \cdot (R_{k-1} \otimes I_{k-1}) \end{aligned}$$

of size $r_0 \cdot n_1 \cdot \ldots \cdot n_d \times r_{k-1} \cdot n_k \cdot \ldots \cdot n_d$ for $1 \leq k \leq d$. Then the matrix $\mathcal{W} \equiv \mathcal{W}_1$ of the WTT transform reads

$$\mathcal{W} = \sum_{k=1}^{d} \overline{\mathcal{W}}_k,$$

where $\overline{\mathcal{W}}_k = \overline{E}_k \overline{P}_k{}^{\top} \overline{R}_k{}^{\top}$ for $1 \leq k \leq d$.

Proof. The proof follows from Lemma 2 by the recursive application of its statement for $k = 1, \ldots, d-1$. □

Algorithm 2 Computation of the WTT image
$y_k = \mathcal{W}_k\left(x_k;\, U_k, \ldots, U_d;\, n_k, \ldots, n_d;\, r_{k-1}, \ldots, r_{d-1}\right)$

Require: recursion level k: $1 \le k \le d$
Require: WTT filters $U_k, \ldots, U_d$ of mode sizes $n_k, \ldots, n_d$ and ranks $r_{k-1}, \ldots, r_{d-1}$ respectively
Require: $r_{k'} \le r_{k'-1} n_{k'}$ for $k \le k' < d$
Require: a vector x_k of size $r_{k-1} \cdot n_k \cdot n_{k+1} \cdot \ldots \cdot n_d$
Ensure: the WTT image vector y_k of the same size

1: **if** $k = d$ **then**
2: $\quad y_d = \mathcal{W}_d\left(x_d;\, U_d;\, n_d;\, r_{d-1}\right) = {U_d}^\top x_d$
3: **else**
4: $\quad X_k = \texttt{reshape}(x_k, r_{k-1} \cdot n_k \times n_{k+1} \cdot \ldots \cdot n_d)$
5: $\quad \widetilde{X}_k = {U_k}^\top X_k$
6: $\quad \widehat{X}_{k+1} = \widetilde{X}_k(1 : r_k\,,\, :), \qquad Z_{k+1} = \widetilde{X}_k(r_k + 1 : r_{k-1} n_k\,,\, :)$
7: $\quad x_{k+1} = \texttt{reshape}(\widehat{X}_{k+1}, r_k \cdot n_{k+1} \cdot \ldots \cdot n_d \times 1)$
8: $\quad y_{k+1} = \mathcal{W}_{k+1}\left(x_{k+1};\, n_{k+1}, \ldots, n_d;\, r_k, \ldots, r_{d-1}\right)$
9: $\quad \widehat{Y}_{k+1} = \texttt{reshape}(y_{k+1}, r_k \times n_{k+1} \cdot \ldots \cdot n_d)$
10: $\quad Y_k = \begin{pmatrix} \widehat{Y}_{k+1} \\ Z_{k+1} \end{pmatrix}$
11: $\quad y_k = \texttt{reshape}(Y_k, r_{k-1} \cdot n_k \cdot \ldots \cdot n_d)$
12: **end if**

Theorem 4. *Assume that a d-dimensional vector a of size $n_1 \cdot n_2 \cdot \ldots \cdot n_d$ and TT ranks $\rho_1, \ldots, \rho_{d-1}$ is given. Consider the WTT transform $\mathcal{W}$ defined by the filters $U_1, \ldots, U_d$ of ranks $r_0 = 1, r_1 \ge \rho_1, \ldots, r_d \ge \rho_d$, adapted to a as described in Algorithm 1. Then only the first entry of the image $y = \mathcal{W}a$ can be nonzero.*

Proof. For $1 \le k < d$ let us consider the kth step of Algorithm 1. Let us assume that the input vector a_k indexed by $d-k+1$ indices $\overline{\alpha_{k-1}, i_k}, i_{k+1}, \ldots, i_d$ is of TT ranks $\rho_k, \ldots, \rho_{d-1}$. This means that for $k \le s < d$ its unfolding with row indices $\alpha_{k-1}, i_k, \ldots, i_s$ and column indices $i_{s+1}, \ldots, i_d$ is of matrix rank ρ_s. As the matrix U_k is unitary, the multiplication by ${U_k}^\top$ at the line 5 does not change the ranks of the unfoldings. Since A_k is the first unfolding of a_k, we have $\operatorname{rank} A_k = \rho_k$ and the diagonal matrix S_k contains only $\rho_k \le r_k$ nonzero entries. Therefore excluding zero rows at the line 6 does not change the ranks of the unfoldings either, and the vector a_{k+1} indexed by $d-k+1$ indices $\alpha_k, i_{k+1}, i_{k+2}, \ldots, i_d$, which is input at the $k+1$st step, is of TT ranks $\rho_{k+1}, \ldots, \rho_{d-1}$.

Due to $\rho_k \le r_k$, in particular, we have $\widetilde{A}_k = E_k \widehat{A}_k$ and $A_k = U_k E_k \widehat{A}_k = R_k \widehat{A}_k$ with E_k and R_k as defined in Lemma 2. This can be recast in the vectorized form as

$$a_k\,{}_{\alpha_{k-1}, i_k, \ldots, i_d} = \sum_{\alpha_k = 1}^{r_k} R_k\,{}_{\substack{\alpha_{k-1}, i_k \\ \alpha_k}}\; a_{k+1}\,{}_{\alpha_k, i_{k+1}, \ldots, i_d}, \tag{12}$$

which holds true for $1 \le k < d$ by induction, as for $a \equiv a_1$ it is given by assumption. At the d-th step no zero rows of $\widetilde{A}_d$ are excluded, and we have

$$a_d\,{}_{\alpha_{d-1}, i_d} = \sum_{\beta=1}^{r_{d-1}} \sum_{j=1}^{n_d} U_d\,{}_{\substack{\alpha_{d-1}, i_d \\ \beta, j}}\; c\,{}_{\beta, j}, \tag{13}$$

where $c = \widetilde{A}_d$ is a column in which only the first entry can be nonzero, since $\operatorname{rank} A_d = 1$.

Let us consider the kth step of Algorithm 2 applied to the vector $x_k = a_k$ with the filters $U_1, \ldots, U_d$. For $1 \leq k < d$ we obtain from (12) that

$$\widetilde{X}_k {}_{\substack{\alpha_{k-1}, i_k \\ i_{k+1}, \ldots, i_d}} = \sum_{\alpha_k = 1}^{r_k} E_k {}_{\substack{\alpha_{k-1}, i_k \\ \alpha_k}} \, a_{k+1} {}_{\alpha_k, i_{k+1}, \ldots, i_d},$$

therefore $x_{k+1} = a_{k+1}$ and $Z_{k+1} = 0$. Then at the dth step $x_d = a_d$ and, by (13), we conclude that $y_d = c$. Consequently, for $k = d-1, \ldots, 1$ only the first entry of y_k can be nonzero. □

Lemma 4 shows that if a vector can be represented in the TT format with certain ranks and the WTT transform of greater or equal ranks is adapted to it as a reference vector, then the first vector of the WTT basis is exactly the reference vector. As a result, the image contains at most one non-zero.

5 Tensor structure of the WTT matrix

Lemma 5. *Assume that WTT filters $U_1, \ldots, U_d$ of mode sizes $n_1, \ldots, n_d$ and ranks $r_0 = 1, r_1, \ldots, r_{d-1}$ respectively are given. Then the matrix $\mathcal{W} \equiv \mathcal{W}_1$ of the WTT transform, defined by Algorithm 2, admits the TT representation*

$$\mathcal{W} = \begin{bmatrix} F_1 & G_1 \end{bmatrix} \bowtie \begin{bmatrix} F_2 & G_2 \\ & I \end{bmatrix} \bowtie \ldots \bowtie \begin{bmatrix} F_{d-1} & G_{d-1} \\ & I \end{bmatrix} \bowtie \begin{bmatrix} G_d \\ I \end{bmatrix} \tag{14}$$

of ranks $r_1^2 + 1, r_2^2 + 1, \ldots, r_{d-2}^2 + 1, r_{d-1}^2 + 1$, where the TT cores F_k and G_k of ranks $r_{k-1}^2 \times r_k^2$ and $r_{k-1}^2 \times 1$ respectively and mode size $n_k \times n_k$ are defined for $1 \leq k < d$ as follows:

$$F_k\left(\alpha_{k-1}, \beta_{k-1};\, i_k, j_k;\, \alpha_k, \beta_k\right) = E_k {}_{\substack{\alpha_{k-1}, i_k \\ \alpha_k}} \, R_k {}_{\substack{\beta_{k-1}, j_k \\ \beta_k}}$$

and

$$G_k\left(\alpha_{k-1}, \beta_{k-1};\, i_k, j_k\right) = P_k {}_{\substack{\alpha_{k-1}, i_k \\ \beta_{k-1}, j_k}}$$

elementwise; and the TT core G_d of rank $r_{d-1}^2 \times 1$ and mode size $n_d \times n_d$ is given by

$$G_d\left(\alpha_{d-1}, \beta_{d-1};\, i_d, j_d\right) = U_d {}_{\substack{\alpha_{d-1}, i_d \\ \beta_{d-1}, j_d}}$$

elementwise.

Proof. For $1 \leq k \leq d$ let us define a TT core $\overleftarrow{F}_k$ of rank $r_{k-1}^2 \times 1$ and mode size $n_k \cdot \ldots \cdot n_d \times n_k \cdot \ldots \cdot n_d$ by setting

$$\overleftarrow{F}_k\left(\alpha_{k-1}, \beta_{k-1};\, i_k, j_k, \ldots, i_d, j_d\right) = \mathcal{W}_k {}_{\substack{\alpha_{k-1}, i_k, \ldots, i_d \\ \beta_{k-1}, j_k, \ldots, j_d}}$$

for $1 \leq \alpha_{k-1}, \beta_{k-1} \leq r_{k-1}$ and $1 \leq i_k, j_k \leq n_k, \ldots, 1 \leq i_d, j_d \leq n_d$. In particular, as $r_0 = 1$, the core

$$\overleftarrow{F}_1 = \mathcal{W}_1 = \mathcal{W} \tag{15}$$

is the matrix of the WTT transform defined by the filters $U_1, \ldots, U_d$. Let us rewrite elementwise the recursive relation given by Lemma 2: for $1 \le k < d$ we have

$$\mathcal{W}_k {}_{\substack{\alpha_{k-1}, i_k, i_{k+1}, \ldots, i_d \\ \beta_{k-1}, j_k, j_{k+1}, \ldots, j_d}} = \sum_{\alpha_k=1}^{r_k} \sum_{\beta_k=1}^{r_k} E_k {}_{\substack{\alpha_{k-1}, i_k \\ \alpha_k}} R_k {}_{\substack{\beta_{k-1}, j_k \\ \beta_k}} \mathcal{W}_{k+1} {}_{\substack{\alpha_k, i_{k+1}, \ldots, i_d \\ \beta_k, j_{k+1}, \ldots, j_d}} + P_k {}_{\substack{\beta_{k-1}, j_k \\ \alpha_{k-1}, i_k}} \delta {}_{\substack{i_{k+1}, \ldots, i_d \\ j_{k+1}, \ldots, j_d}} \tag{16}$$

for all possible values of the free indices, cf. (10)–(11). Next, we rewrite (16) as

$$\overleftarrow{F}_k = F_k \bowtie \overleftarrow{F}_{k+1} + G_k \bowtie I = \begin{bmatrix} F_k & G_k \end{bmatrix} \bowtie \begin{bmatrix} \overleftarrow{F}_{k+1} \\ I \end{bmatrix}$$

to obtain

$$\begin{bmatrix} \overleftarrow{F}_k \\ I \end{bmatrix} = \begin{bmatrix} F_k & G_k \\ & I \end{bmatrix} \bowtie \begin{bmatrix} \overleftarrow{F}_{k+1} \\ I \end{bmatrix} = W_k \bowtie \begin{bmatrix} \overleftarrow{F}_{k+1} \\ I \end{bmatrix}$$

for $1 \le k < d$. Together with (15), this relation, being applied recursively, yields the claim. $\square$

6 Tensor structure of the WTT image

Lemma 6. *Assume that WTT filters $U_1, \ldots, U_d$ of mode sizes $n_1, \ldots, n_d$ and ranks $r_0 = 1, r_1, \ldots, r_{d-1}$ respectively are given. Consider a vector x of size $n_1 \cdot n_2 \cdot \ldots \cdot n_d$ and its image $y = \mathcal{W}x$ under the WTT transform defined by the filters $U_1, \ldots, U_d$ in Algorithm 2.*

Assume that x is given in a TT representation

$$x = X_1 \bowtie X_2 \bowtie \ldots \bowtie X_{d-1} \bowtie X_d$$

of ranks $p_1, \ldots, p_{d-1}$. Then y can be represented in the TT decomposition

$$y = \begin{bmatrix} V_1 & Z_1 \end{bmatrix} \bowtie \begin{bmatrix} V_2 & Z_2 \\ & X_2 \end{bmatrix} \bowtie \ldots \bowtie \begin{bmatrix} V_{d-1} & Z_{d-1} \\ & X_{d-1} \end{bmatrix} \bowtie \begin{bmatrix} Z_d \\ X_d \end{bmatrix}$$

of ranks $r_1 + p_1, \ldots, r_{d-1} + p_{d-1}$, where for $1 \le k < d$ the core V_k of rank $r_{k-1} \times r_k$ and mode size n_k is defined as follows:

$$V_k\,(\alpha_{k-1};\, i_k;\, \alpha_k) = E_k {}_{\substack{\alpha_{k-1}, i_k \\ \alpha_k}}$$

elementwise; and for $1 \le k \le d$ the core Z_k of rank $r_{k-1} \times p_k$ and mode size n_k is given by

$$Z_k\,(\alpha_{k-1};\, i_k;\, \gamma_k) = \sum_{\substack{j_1, \ldots, j_k \\ \beta_1, \ldots, \beta_{k-1}}} R_1 {}_{\substack{1, j_1 \\ \beta_1}} \cdot R_2 {}_{\substack{\beta_1, j_2 \\ \beta_2}} \cdot \ldots \cdot R_{k-1} {}_{\substack{\beta_{k-2}, j_{k-1} \\ \beta_{k-1}}} \cdot P_k {}_{\substack{\beta_{k-1}, j_k \\ \alpha_{k-1}, i_k}}$$
$$\cdot \sum_{\gamma_1=1}^{r_1} \ldots \sum_{\gamma_{k-1}=1}^{r_{k-1}} X_1\,(j_1;\, \gamma_1)\, X_2\,(\gamma_1;\, j_2;\, \gamma_2) \cdot \ldots \cdot X_k\,(\gamma_{k-1}, j_k;\, \gamma_k)$$

elementwise. Here, we let $r_0 = p_d = 1$, so that $\alpha_0 = 1$ and $\gamma_d = 1$ are void indices.

Proof. We rewrite the entries of the matrices $\overline{E}_k$ and $\overline{R}_k$, $1 \le k \le d$, defined in Corollary 3, as follows:

$$\overline{E}_k {}_{\substack{1, i_1, \ldots, i_d \\ \alpha_{k-1}, j_k, \ldots, j_d}} = \sum_{\alpha_1, \ldots, \alpha_{k-2}} E_1 {}_{\substack{1, i_1 \\ \alpha_1}} \cdot E_2 {}_{\substack{\alpha_1, i_2 \\ \alpha_2}} \cdot \ldots \cdot E_{k-1} {}_{\substack{\alpha_{k-2}, i_{k-1} \\ \alpha_{k-1}}} \cdot \delta_{\substack{i_k \\ j_k}} \cdot \ldots \cdot \delta_{\substack{i_d \\ j_d}},$$
$$\overline{R}_k {}_{\substack{1, i_1, \ldots, i_d \\ \alpha_{k-1}, j_k, \ldots, j_d}} = \sum_{\alpha_1, \ldots, \alpha_{k-2}} R_1 {}_{\substack{1, i_1 \\ \alpha_1}} \cdot R_2 {}_{\substack{\alpha_1, i_2 \\ \alpha_2}} \cdot \ldots \cdot R_{k-1} {}_{\substack{\alpha_{k-2}, i_{k-1} \\ \alpha_{k-1}}} \cdot \delta_{\substack{i_k \\ j_k}} \cdot \ldots \cdot \delta_{\substack{i_d \\ j_d}}.$$

Let us now define

$$\overrightarrow{X}_k = X_1 \bowtie X_2 \bowtie \ldots \bowtie X_{k-2} \bowtie X_k$$

and

$$\overleftarrow{X}_k = X_k \bowtie X_{k+1} \bowtie \ldots \bowtie X_{d-1} \bowtie X_d$$

for $1 \leq k \leq d$ and set $\overrightarrow{X}_{d+1} = \begin{bmatrix} 1 \end{bmatrix}$, so that $x = \overrightarrow{X}_k \bowtie \overleftarrow{X}_{k+1}$ for $1 \leq k \leq d$. Then we obtain

$$\left(\overline{R}_k{}^\top x\right)_{\beta_{k-1}, j_k, i_{k+1}, \ldots, i_d} = \sum_{\gamma_k=1}^{p_k} \sum_{\substack{j_1,\ldots,j_{k-1} \\ \beta_1,\ldots,\beta_{k-2}}} R_1{}_{\substack{1, j_1 \\ \beta_1}} \cdot R_2{}_{\substack{\beta_1, j_2 \\ \beta_2}} \cdot \ldots$$
$$\cdot R_{k-1}{}_{\substack{\beta_{k-2}, j_{k-1} \\ \beta_{k-1}}} \cdot \overrightarrow{X}_k\left(j_1, \ldots, j_{k-1}, j_k;\, \gamma_k\right) \cdot \overleftarrow{X}_{k+1}\left(\gamma_k;\, i_{k+1}, \ldots, i_d\right)$$

and

$$\left(\overline{P}_k{}^\top \overline{R}_k{}^\top x\right)_{\alpha_{k-1}, i_k, i_{k+1}, \ldots, i_d} = \sum_{\gamma_k=1}^{p_k} \sum_{\substack{j_1,\ldots,j_{k-1},j_k \\ \beta_1,\ldots,\beta_{k-2},\beta_{k-1}}} R_1{}_{\substack{1, j_1 \\ \beta_1}} \cdot R_2{}_{\substack{\beta_1, j_2 \\ \beta_2}} \cdot \ldots$$
$$\cdot R_{k-1}{}_{\substack{\beta_{k-2}, j_{k-1} \\ \beta_{k-1}}} \cdot P_k{}_{\substack{\beta_{k-1}, j_k \\ \alpha_{k-1}, i_k}} \cdot \overrightarrow{X}_k\left(j_1, \ldots, j_{k-1}, j_k;\, \gamma_k\right) \cdot \overleftarrow{X}_{k+1}\left(\gamma_k;\, i_{k+1}, \ldots, i_d\right).$$

Consequently, the entries of the image $\overline{y}_k = \overline{W}_k\, x$ of x under the kth term of the decomposition given by Corollary 3 take the form

$$\overline{y}_k{}_{1, i_1, \ldots, i_d} = \sum_{\alpha_{k-1}=1}^{r_{k-1}} \sum_{\gamma_k=1}^{p_k} \sum_{\alpha_1,\ldots,\alpha_{k-2}} E_1{}_{\substack{1, i_1 \\ \alpha_1}} \cdot E_2{}_{\substack{\alpha_1, i_2 \\ \alpha_2}} \cdot \ldots \cdot E_{k-1}{}_{\substack{\alpha_{k-2}, i_{k-1} \\ \alpha_{k-1}}}$$
$$\cdot \sum_{\substack{j_1,\ldots,j_k \\ \beta_1,\ldots,\beta_{k-1}}} R_1{}_{\substack{1, j_1 \\ \beta_1}} \cdot R_2{}_{\substack{\beta_1, j_2 \\ \beta_2}} \cdot \ldots \cdot R_{k-1}{}_{\substack{\beta_{k-2}, j_{k-1} \\ \beta_{k-1}}} \cdot P_k{}_{\substack{\beta_{k-1}, j_k \\ \alpha_{k-1}, i_k}}$$
$$\cdot \overrightarrow{X}_k\left(j_1, \ldots, j_{k-1}, j_k;\, \gamma_k\right) \cdot \overleftarrow{X}_{k+1}\left(\gamma_k;\, i_{k+1}, \ldots, i_d\right),$$

which implies that the kth term of the matrix affects only the first k cores of the input vector. Namely, the image reads as

$$\overline{y}_k = V_1 \bowtie V_2 \bowtie \ldots \bowtie V_{k-1} \bowtie Z_k \bowtie X_{k+1} \bowtie X_{k+2} \bowtie \ldots \bowtie X_d.$$

By [32, Lemma 5.5], for the complete image $y \equiv y_1 = \mathcal{W} x_1 = \sum_{k=1}^{d} \overline{y}_k$ we obtain

$$y = \begin{bmatrix} V_1 & Z_1 \end{bmatrix} \bowtie \begin{bmatrix} V_2 & Z_2 \\ & X_2 \end{bmatrix} \bowtie \ldots \bowtie \begin{bmatrix} V_{d-1} & Z_{d-1} \\ & X_{d-1} \end{bmatrix} \bowtie \begin{bmatrix} Z_{d-1} \\ X_{d-1} \end{bmatrix},$$

which concludes the proof. □

7 Localization of the WTT basis vectors

Theorem 7. *Consider the matrix $\mathcal{W}$ of the WTT transform defined by d recursive steps of Algorithm 2 for WTT filters $U_1, \ldots, U_d$. Let $1 \leq k_0 \leq d-2$ and assume for $k_0 \leq k < d$*

that $r_k \leq \nu_k r_{k-1}$ with $1 \leq \nu_k < n_k$. Then there exist disjoint sets of row indices $\mathcal{R}_L \subset \mathcal{I}$, $k_0 \leq L < d$, such that for every L the set $\mathcal{R}_L$ consists of

$$\#\mathcal{R}_L = \left(\prod_{k=1}^{k_0-1} n_k\right) \cdot \left(\prod_{k=k_0}^{L-1} \nu_k\right) \cdot (n_L - \nu_L) \cdot \left(\prod_{k=L+1}^{d} n_k\right) \tag{17}$$

row indices, and for every row index $i \in \mathcal{R}_L$ the ith row of $\mathcal{W}$ contains at most $\prod_{k=1}^{L} n_k$ nonzero entries. The whole matrix $\mathcal{W}$ contains at most

$$\left[\prod_{k=1}^{k_0-1} n_k^2\right] \cdot \left[\left(\prod_{k=k_0}^{d-1} \nu_k n_k\right) n_d^2 + \sum_{L=k_0}^{d-1} \left(\prod_{k=k_0}^{L-1} \nu_k n_k\right) (n_L - \nu_L)\, n_L \left(\prod_{k=L+1}^{d} n_k\right)\right]$$

nonzero entries.

Proof. Consider the entry of the matrix $\mathcal{W}$ with a row index $(i_1, \ldots, i_d)$ and a column index $(j_1, \ldots, j_d)$. According to Lemma 5, it can be written as follows:

$$\mathcal{W}_{\substack{i_1,\ldots,i_d \\ j_1,\ldots,j_d}} = W_1 \cdot W_2 \cdots W_{d-1} \cdot W_d, \tag{18}$$

where for $1 \leq k \leq d$ the matrix W_k indexed by rank indices is obtained from the core of the right-hand side of (14) by restricting it to the values i_k and j_k of the mode indices. In particular, for $1 < k < d$ we obtain W_k as a matrix of size $r_{k-1}^2 + 1 \times r_k^2 + 1$ given by

$$W_k = \begin{pmatrix} F_k(\cdot\,;\, i_k, j_k;\, \cdot) & G_k(\cdot\,;\, i_k, j_k;\, \cdot) \\ 0 & \delta(i_k, j_k) \end{pmatrix} = \left(\begin{array}{ccc|c} * & \cdots & * & * \\ \vdots & \cdots & \vdots & \vdots \\ * & \cdots & * & * \\ \hline & & & * \end{array}\right), \tag{19}$$

and, similarly, W_1 and W_d are matrices of size $1 \times r_1^2 + 1$ and $r_{d-1}^2 + 1 \times 1$ respectively.

First, assume that $\nu_L i_L \leq n_L$ for some $L \in \{k_0, \ldots, d-1\}$. Then the corresponding submatrix of E_L defined in Lemma 2 is zero: for $1 \leq \alpha_{L-1}, \alpha_L \leq r$ we have $E_{L\,\substack{\alpha_{L-1}, i_L \\ \alpha_L}} = 0$ and, consequently,

$$F_L\left(\alpha_{L-1}, \beta_{L-1};\, i_L, j_L;\, \alpha_L, \beta_L\right) = 0,$$

where F_L is defined in Lemma 5. Then the matrix W_L takes the form

$$W_L = \left(\begin{array}{ccc|c} 0 & \cdots & 0 & * \\ \vdots & \cdots & \vdots & \vdots \\ 0 & \cdots & 0 & * \\ \hline & & & * \end{array}\right) \tag{20}$$

if $1 < L < d$, and similar with only the first row being present if $L = 1$.

Second, assume that $i_{L'} \neq j_{L'}$ for some $L' \in \{k_0 + 1, \ldots, d\}$. Then, due to the identity block, the matrix $W_{L'}$ becomes

$$W_{L'} = \left(\begin{array}{ccc|c} * & \cdots & * & * \\ \vdots & \cdots & \vdots & \vdots \\ * & \cdots & * & * \\ \hline & & & 0 \end{array}\right) \tag{21}$$

if $1 < L' < d$, and only the last column is present for $L' = d$.

Note that as soon as $L < L'$, the combination of (19), (20), (21) results in

$$W_L \cdot W_{L+1} \cdots W_{L'-1} \cdot W_{L'} = 0,$$

so that, due to (18), for the entry under consideration we obtain $\mathcal{W}_{\substack{i_1,\ldots,i_d \\ j_1,\ldots,j_d}} = 0$.

Let us now count the entries which may still be nonzero. For $k_0 \leq L < d$ we define the set of row indices

$$\mathcal{R}_L = \{(i_1, \ldots, i_d) \in \mathcal{I} : \quad i_k \leq \nu_k \quad \text{for} \quad k_0 \leq k < L \quad \text{and} \quad i_L > \nu_L\}, \tag{22}$$

so that (17) holds true and every row $(i_1, \ldots, i_d) \in \mathcal{R}_L$ may contain at most $\prod_{k=1}^{L} n_k$ nonzero entries, which satisfy

$$j_{L+1} = i_{L+1}, \ldots, j_d = i_d. \tag{23}$$

Here, $\mathcal{R}_L$, $k_0 \leq L < d$, are disjoint by construction.

As for the rest of the rows, $\mathcal{I} \setminus \bigcup_{L=k_0}^{d-1} \mathcal{R}_L$ consists of

$$\left(\prod_{k=1}^{k_0-1} n_k\right) \cdot \left(\prod_{k=k_0}^{d-1} \nu_k\right) \cdot n_d$$

indices. We do not analyze the sparsity of the corresponding rows and allow that they are full. Thus, the total number of nonzero entries in $\mathcal{W}$ is bounded by

$$\left(\prod_{k=1}^{k_0-1} n_k\right) \cdot \left(\prod_{k=k_0}^{d-1} \nu_k\right) \cdot n_d \cdot \left(\prod_{k=1}^{d} n_k\right) + \sum_{L=k_0}^{d-1} \left(\prod_{k=1}^{k_0-1} n_k\right) \cdot \left(\prod_{k=k_0}^{L-1} \nu_k\right) \cdot (n_L - \nu_L) \cdot \left(\prod_{k=L+1}^{d} n_k\right) \cdot \left(\prod_{k=1}^{L} n_k\right).$$

□

Theorem 7 describes the structure of the WTT basis with respect to the virtual levels (scales) of “physical” dimensions, L denoting the level number. The result admits a clear interpretation in the case of a single dimension, when the quantization with base n is used to construct and apply the WTT transform. In this case, L numbers the scales of the only dimension, from the finest to the coarsest, and the supports of the basis functions reduce exponentially as L descreases.

Corollary 8. *In the setting of Theorem 7, assume that the quantization with $n_1 = \ldots = n_d = n$ is used and that $k_0 = 2$ and $\nu_k = 1$ for $2 \leq k < d$. Then there exist disjoint sets of row indices $\mathcal{R}_L \subset \mathcal{I}$, $2 \leq L < d$, such that for $1 < L < d$ the set $\mathcal{R}_L$ consists of $\#\mathcal{R}_L = (n-1)\, n^{d-L+1}$ row indices, and for all $i \in \mathcal{R}_L$ the ith row of $\mathcal{W}$ contains at most n^L nonzero entries. For $1 < L < d$, all rows indexed in $\mathcal{R}_L$ contribute at most $n^{d+1}(n-1)$ nonzero entries to $\mathcal{W}$. The rest n^2 rows contribute up to n^{d+2} nonzero entries. The whole matrix $\mathcal{W}$ contains at most $n^{d+1}(dn - n - d + 2) = \mathcal{O}\left(dn^{d+2}\right)$ nonzero entries.*

8 Conclusion

We showed that WTT transforms, adaptive discrete wavelet transforms constructed with the use of the tensor train decomposition of multidimensional arrays, exhibit low-rank structure in the same format. We derived explicit and exact representations of the matrix of such a transform in terms of the reference vector generating it, and also that of the image in terms of the reference vector and the vector transformed.

These representations and the rank bounds following from them may be helpful in the efficient implementation of the WTT transforms of data with low-rank TT structure, as well as in understanding the hierarchical multiscale structure of the WTT basis. We also showed that this basis is sparse and gave bounds on the amount of nonzeros with respect to the level number.

In a broader sense, the results of the present paper can be viewed as a parametrization of a class of matrices possessing at the same time the properties of orthogonality, low-rank tensor structure and sparse structure.

The exact relation of the WTT transform to the classical discrete wavelet transforms and the tensor structure of the latter is a topic of ongoing research.

References

[1] *D. L. Donoho*. Unconditional Bases Are Optimal Bases for Data Compression and for Statistical Estimation // *Applied and Computational Harmonic Analysis*. 1993. V. 1, No. 1. P. 100 - 115. DOI: `http://dx.doi.org/10.1006/acha.1993.1008`. `http://www.sciencedirect.com/science/article/pii/S1063520383710080`. 1

[2] *I. Daubechies*. Orthonormal bases of compactly supported wavelets // *Comm. Pur. Appl. Math.* 1988. V. 41, No. 7. P. 909-996. DOI: `10.1002/cpa.3160410705`. 2

[3] *A. Cohen, I. Daubechies, J.-C. Feauveau*. Biorthogonal bases of compactly supported wavelets // *Communications on Pure and Applied Mathematics*. 1992. V. 45, No. 5. P. 485–560. DOI: `10.1002/cpa.3160450502`. `http://dx.doi.org/10.1002/cpa.3160450502`. 2

[4] *W. Sweldens*. The lifting scheme: A custom-design constuction of biorthogonal wavelets // *App. Comput. Harmon. A.* 1996. V. 3, No. 2. P. 186-200. DOI: `10.1006/acha.1996.0015`. 2

[5] *I. Daubechies, W. Sweldens*. Factoring wavelet transforms into lifting steps // *Journal of Fourier Analysis and Applications*. 1998. V. 4, no. 3. P. 247–269. DOI: `10.1007/BF02476026`. `http://dx.doi.org/10.1007/BF02476026`. 2

[6] *S. G. Mallat*. Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbb{R})$ // *Transactions of the American Mathematical Society*. V. 315, No. 1. P. 69–87. DOI: `10.1090/S0002-9947-1989-1008470-5`. `http://www.ams.org/journals/tran/1989-315-01/S0002-9947-1989-1008470-5`. 2

[7] *S. Mallat*. A theory for multiresolution signal decomposition: the wavelet representation // *Pattern Analysis and Machine Intelligence, IEEE Transactions on*. 1989. V. 11, No. 7. P. 674–693. DOI: `10.1109/34.192463`. 2

[8] *W. Sweldens*. The Lifting Scheme: A Construction of Second Generation Wavelets // *SIAM Journal on Mathematical Analysis*. 1998. V. 29, No. 2. P. 511-546. DOI: 10.1137/S0036141095289051. http://epubs.siam.org/doi/abs/10.1137/S0036141095289051. 2

[9] *W. Dahmen*. Wavelets, Images, and Surface Fitting // Ed. by P. J. Laurent, A. Le Méhauté, L. L. Schumaker. — A. K. Peters Boston, 1991. — P. 157–188. http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.34.944. 2

[10] *W. Dahmen, S. Prössdorf, R. Schneider*. Wavelet approximation methods for pseudodifferential equations II: Matrix compression and fast solution // *Advances in Computational Mathematics*. 1993. V. 1, No. 3. P. 259–335. DOI: 10.1007/BF02072014. http://dx.doi.org/10.1007/BF02072014. 2

[11] *W. Dahmen, S. Prössdorf, R. Schneider*. Wavelet approximation methods for pseudodifferential equations: I Stability and convergence // *Mathematische Zeitschrift*. 1994. V. 215, No. 1. P. 583–620. DOI: 10.1007/BF02571732. http://dx.doi.org/10.1007/BF02571732. 2

[12] *T. von Petersdorff, Ch. Schwab*. Wavelet approximations for first kind boundary integral equations on polygons // *Numerische Mathematik*. 1996. V. 74, no. 4. P. 479–516. DOI: 10.1007/s002110050226. http://dx.doi.org/10.1007/s002110050226. 2

[13] *T. von Petersdorff, Ch. Schwab, R. Schneider*. Multiwavelets for Second-Kind Integral Equations // *SIAM Journal on Numerical Analysis*. 1997. V. 34, No. 6. P. 2212-2227. DOI: 10.1137/S0036142994272957. http://epubs.siam.org/doi/abs/10.1137/S0036142994272957. 2

[14] *T. von Petersdorff, Ch. Schwab*. Wavelet Discretizations of Parabolic Integrodifferential Equations // *SIAM Journal on Numerical Analysis*. 2003. V. 41, No. 1. P. 159–180. DOI: 10.1137/S0036142901394844. http://epubs.siam.org/doi/pdf/10.1137/S0036142901394844. http://epubs.siam.org/doi/abs/10.1137/S0036142901394844. 2

[15] *Ch. Schwab, R. Stevenson*. Space-time adaptive wavelet methods for parabolic evolution problems // *Mathematics of Computation*. V. 78, No. 267. P. 1293–1318. DOI: 10.1090/S0025-5718-08-02205-9. http://www.ams.org/journals/mcom/2009-78-267/S0025-5718-08-02205-9. 2

[16] *I. V. Oseledets, E. E. Tyrtyshnikov*. Algebraic wavelet transform via quantics tensor train decomposition // *SIAM J. Sci. Comput*. 2011. V. 33, No. 3. P. 1315-1328. DOI: 10.1137/100811647. 2, 7, 8

[17] *I. V. Oseledets, E. E. Tyrtyshnikov*. Breaking the curse of dimensionality, or how to use SVD in many dimensions // *SIAM J. Sci. Comput*. 2009. V. 31, No. 5. P. 3744-3759. DOI: 10.1137/090748330. 2, 4

[18] *I. V. Oseledets*. Tensor-train decomposition // *SIAM J. Sci. Comput*. 2011. V. 33, No. 5. P. 2295-2317. DOI: 10.1137/090752286. 2, 4, 7

[19] *I. V. Oseledets*. Approximation of matrices with logarithmic number of parameters // *Doklady Math*. 2009. V. 428, No. 1. P. 23-24. DOI: 10.1134/S1064562409050056. 2, 5

[20] *I. V. Oseledets*. Approximation of $2^d \times 2^d$ matrices using tensor decomposition // *SIAM J. Matrix Anal. Appl.* 2010. V. 31, No. 4. P. 2130-2145. DOI: 10.1137/090757861. 2, 5

[21] *B. N. Khoromskij*. $\mathcal{O}(d \log N)$–Quantics approximation of N–d tensors in high-dimensional numerical modeling // *Constr. Appr.* 2011. V. 34, No. 2. P. 257-280. DOI: 10.1007/s00365-011-9131-1. 2, 5

[22] *R. Bellman*. Adaptive Control Processes: A Guided Tour. — Princeton, NJ: Princeton University Press, 1961. 3

[23] *T. G. Kolda, B. W. Bader*. Tensor decompositions and applications // *SIAM Review*. 2009. V. 51, No. 3. P. 455–500. DOI: 10.1137/07070111X. 3

[24] *L. Grasedyck, D. Kressner, C. Tobler*. A literature survey of low-rank tensor approximation techniques: arXiv preprint 1302.7121: 2013, February. http://arxiv.org/abs/1302.7121. 3

[25] *C. F. van Loan, N. Pitsianis*. Approximation with Kronecker products // Linear algebra for large scale and real-time applications (Leuven, 1992). — Dordrecht, 1993. — V. 232 of *NATO Adv. Sci. Inst. Ser. E Appl. Sci.* — P. 293–314. 4

[26] *G. Beylkin, M. J. Mohlenkamp*. Numerical operator calculus in higher dimensions // *Proc. Nat. Acad. Sci. USA*. 2002. V. 99, No. 16. P. 10246-10251. DOI: 10.1073/pnas.112329799. 4

[27] *E. E. Tyrtyshnikov*. Tensor approximations of matrices generated by asymptotically smooth functions // *Sbornik: Mathematics*. 2003. V. 194, No. 6. P. 941-954. DOI: 10.1070/SM2003v194n06ABEH000747. 5

[28] *S. R. White*. Density matrix formulation for quantum renormalization groups // *Phys. Rev. Lett.* 1992. V. 69, No. 19. P. 2863-2866. DOI: 10.1103/PhysRevLett.69.2863. 5

[29] *F. Verstraete, D. Porras, J. I. Cirac*. Density Matrix Renormalization Group and Periodic Boundary Conditions: A Quantum Information Perspective // *Phys. Rev. Lett.* 2004. V. 93, No. 22. P. 227205. DOI: 10.1103/PhysRevLett.93.227205. http://link.aps.org/doi/10.1103/PhysRevLett.93.227205. 5

[30] *G. Vidal*. Efficient classical simulation of slightly entangled quantum computations // *Phys. Rev. Lett.* 2003. V. 91, No. 14. P. 147902. DOI: 10.1103/PhysRevLett.91.147902. 5

[31] *F. Verstraete, J. J. García-Ripoll, J. I. Cirac*. Matrix Product Density Operators: Simulation of Finite-Temperature and Dissipative Systems // *Phys. Rev. Lett.* 2004, November. V. 93. P. 207204. DOI: 10.1103/PhysRevLett.93.207204. 6

[32] *V. A. Kazeev, B. N. Khoromskij*. Low-rank explicit QTT representation of the Laplace operator and its inverse // *SIAM J. Matrix Anal. Appl.* 2012. V. 33, No. 3. P. 742-758. DOI: 10.1137/100820479. 6, 14

[33] *V. Kazeev, B. Khoromskij, E. Tyrtyshnikov*. Multilevel Toeplitz Matrices Generated by Tensor-Structured Vectors and Convolution with Logarithmic Complexity // *SIAM Journal on Scientific Computing*. 2013. V. 35, No. 3. P. A1511–A1536. DOI: 10.1137/110844830. http://epubs.siam.org/doi/abs/10.1137/110844830. 6

[34] *W. De Launey, J. Seberry*. The strong Kronecker product // *J. of Combin. Theory, Ser. A*. 1994. V. 66, No. 2. P. 192–213. DOI: 10.1016/0097-3165(94)90062-0. 6

Recent Research Reports

Nr.	Authors/Title
2013-18	V. Kazeev and Ch. Schwab Tensor approximation of stationary distributions of chemical reaction networks
2013-19	K. Schmidt and R. Hiptmair Asymptotic Boundary Element Methods for Thin Conducting Sheets
2013-20	R. Kruse Consistency and Stability of a Milstein-Galerkin Finite Element Scheme for Semilinear SPDE
2013-21	J. Sukys Adaptive load balancing for massively parallel multi-level Monte Carlo solvers
2013-22	R. Andreev and A. Lang Kolmogorov-Chentsov theorem and differentiability of random fields on manifolds
2013-23	P. Grohs and M. Sprecher Projection-based Quasiinterpolation in Manifolds
2013-24	P. Grohs and S. Keiper and G. Kutyniok and M. Schaefer α-Molecules: Curvelets, Shearlets, Ridgelets, and Beyond
2013-25	A. Cohen and A. Chkifa and Ch. Schwab Breaking the curse of dimensionality in sparse polynomial approximation of parametric PDEs
2013-26	A. Lang Isotropic Gaussian random fields on the sphere
2013-27	Ch. Schwab and C. Schillings Sparse Quadrature Approach to Bayesian Inverse Problems