SIAM J. MATRIX ANAL. APPL.
Vol. 37, No. 2, pp. 818–822

# ERRATUM: TWO-LEVEL QTT-TUCKER FORMAT FOR OPTIMIZED TENSOR CALCULUS*

SIMON ETTER[†], SERGEY DOLGOV[‡], AND BORIS N. KHOROMSKIJ[§]

**Abstract.** We prove by counterexample that the bound on the rounding error given in Theorem 5.2 of [Dolgov and Khoromskij, *SIAM J. Matrix Anal. Appl.*, 34 (2013), pp. 593–623] does not hold in general. A correct version of the error bound as well as a modified rounding algorithm for which the original bound holds are presented.

**Key words.** tensor decompositions, QTT-Tucker representation, singular value decomposition, orthogonal projection

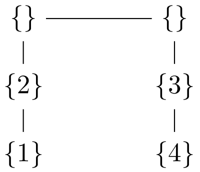
**AMS subject classifications.** 65F50, 15A69, 33F05, 65F30

**DOI.** 10.1137/15M104089X

Let $V$ be an inner product space and $u, v \in V$ two mutually orthogonal and normalized vectors. We consider the tensor

$$x := (u \otimes u + \alpha\, v \otimes v) \otimes (u \otimes u + \alpha\, v \otimes v) + \alpha\,(-\alpha\, u \otimes u + v \otimes v) \otimes (-\alpha\, u \otimes u + v \otimes v)$$

with $\alpha \in (-1, 1)$, mapped to the QTT-Tucker format for $d = 2$ physical modes and $l = 2$ virtual modes per physical mode as follows:



In the above diagram, the empty sets represent the *core blocks*, the singletons denote the *factor blocks*, and their single elements indicate the modes of $x$ associated with them. The edges represent the *rank indices* as in tensor network diagrams. For space reasons, we will omit the outer product symbol $\otimes$ in the following.

It is easily verified that all QTT-Tucker ranks of $x$ equal two. We next perform the rounding procedure [1, Algorithm 1] to truncate $x$ to rank 1.

| Separated modes | Kept component | Dropped component |
|---|---|---|
| $\{1,2\} - \{3,4\}$ | $(uu + \alpha\, vv)\,(uu + \alpha\, vv)$ | $\alpha\,(-\alpha\, uu + vv)(-\alpha\, uu + vv)$ |
| $\{1,2,3\} - \{4\}$ | $(uu + \alpha\, vv)\, uu$ | $\alpha\,(uu + \alpha\, vv)\, vv$ |
| $\{1\} - \{2,3,4\}$ | $uu\, uu$ | $\alpha\, vv\, uu$ |

We note that the first and third dropped component are in general not orthogonal,

$$\big(\alpha\,(-\alpha\, uu + vv)(-\alpha\, uu + vv), \alpha\, vv\, uu\big) = -\alpha^3 \neq 0 \quad \text{for } \alpha \neq 0,$$

*Received by the editors September 22, 2015; accepted for publication (in revised form) by T. G. Kolda February 29, 2016; published electronically June 30, 2016.
http://www.siam.org/journals/simax/37-2/M104089.html

[†] Seminar für Angewandte Mathematik, ETH Zürich, 8092 Zurich, Switzerland (simon.etter@sam.math.ethz.ch).

[‡] University of Bath, Claverton Down, Bath, BA2 7AY, United Kingdom (s.dolgov@bath.ac.uk). This author acknowledges funding from the EPSRC Fellowship EP/M019004/1.

[§] Max Planck Institute for Mathematics in the Sciences, 04103 Leipzig, Germany (bokh@mis.mpg.de).

which indicates the error bound in [1, Theorem 5.2] might be overly optimistic. Indeed, the squared norms of the dropped components are $\alpha^2\,(1+\alpha^2)^2$, $\alpha^2(1+\alpha^2)$ and $\alpha^2$ summing to $\alpha^2(3+3\alpha^2+\alpha^4)$, which is less than the actual squared truncation error

$$\|x - uu\,uu\|^2 = \alpha^2\,(3 - 2\alpha + 3\alpha^2 + \alpha^4)$$

for $\alpha < 0$.

One way to resolve the problem is to adapt the statement of [1, Theorem 5.2].

THEOREM 1.1 (correct version of [1, Theorem 5.2]). *Suppose that each truncation in QTT-Tucker factors is done via the SVD with the Frobenius error* $\varepsilon_{k,p}$, $p = 2,\dots,L$, *the Tucker ranks (ranges of* $\gamma_k$ *in extended factors) are determined with* $\varepsilon_{k,1}$, *and each core block is truncated with* $\varepsilon_{k,0}$. *Then, the Frobenius error in the whole tensor estimates as*

$$\|Y - X\| \leq \sqrt{\sum_{k=1}^{d-1} \varepsilon_{k,0}^2} + \sqrt{\sum_{k=1}^{d}\sum_{p=1}^{L} \varepsilon_{k,p}^2}.$$

*Proof.* Let $P^{(k,p)}$, $Q^{(k)}$ be as in the proof of [1, Theorem 5.2] and define $P^{(k)} := \prod_{p=1}^{L} P^{(k,p)}$, $Q := \prod_{k=1}^{d-1} Q^{(k)}$. We split the truncation error according to

$$\begin{aligned}
\|X - Y\| &= \left\| \left( I - \left( \prod_{k=1}^{d} P^{(k)} \right) Q \right) X \right\| \\
&= \left\| (I - Q)\,X + \sum_{k=1}^{d} \left( I - P^{(k)} \right) \left( \prod_{k'=k+1}^{d} P^{(k')} \right) QX \right\| \\
&\leq \|(I - Q)\,X\| + \left\| \sum_{k=1}^{d} \left( I - P^{(k)} \right) \left( \prod_{k'=k+1}^{d} P^{(k')} \right) QX \right\|.
\end{aligned}$$

The first term represents the error introduced by rounding the core and may be readily estimated by [2, Theorem 2.2]. In the second term, one can make an argument analogous to the one in the proof of Theorem 1.4 below to verify that the terms in this decomposition are mutually orthogonal. Since these terms represent the errors in the SVD truncations, the claim readily follows. ◻

*Remark* 1.2. The problem in the original proof is the wrong statement

$$\|(I - P^{(1,1)})X\|^2 \leq \varepsilon_{1,1}^2.$$

Because of the sequential nature of the algorithm, the correct error estimate would be

$$\left\| (I - P^{(1,1)}) \left( \prod_{p=2}^{L} P^{(1,p)} \right) \left( \prod_{k=2}^{d} P^{(k)} \right) QX \right\|^2 \leq \varepsilon_{1,1}^2,$$

but this does not match with the splitting of the error $\|X - Y\|^2$ given there.

Alternatively, it is possible to adapt the rounding algorithm such that the original error bound holds. We simplify the exposition by introducing the following convention.

*Remark* 1.3. By "truncating a rank index," we mean to compute and truncate the network-structured SVD of the corresponding unfolding. The required orthogonalization steps are tacitly assumed.

For the reader's convenience, Algorithm 1 presents the original rounding algorithm in this simplified notation.

**Algorithm 1** QTT-Tucker Round (Original version [1, Algorithm 1]).

```
for $k = 1, \dots, d-1$ do
    Truncate $\alpha_k$
end for
for $k = d, \dots, 1$ do
    for $p = L-1, \dots, 1$ do
        Truncate $\gamma_{k,p}$
    end for
    Truncate $\gamma_k$
end for
```

**Algorithm 2** QTT-Tucker Round (Modified version with improved error bounds).

```
for $k = d, \dots, 2$ do
    for $p = L-1, \dots, 1$ do
        Truncate $\gamma_{k,p}$
    end for
    Truncate $\gamma_k$
    Truncate $\alpha_{k-1}$
end for
for $p = L-1, \dots, 1$ do
    Truncate $\gamma_{1,p}$
end for
Truncate $\gamma_1$
```

THEOREM 1.4. *Denote by $\varepsilon(\beta)$ with $\beta \in \{\gamma_{k,p}, \gamma_k, \alpha_k\}$, (k and p taken from the appropriate ranges), the Frobenius error incurred by the truncation of $\beta$ in Algorithm* 2. *Then, the Frobenius distance between the original and truncated tensors $X$ and $Y$, respectively, satisfies*

$$\|Y - X\|^2 = \sum_{k=1}^{d} \sum_{p=1}^{L-1} \varepsilon^2(\gamma_{k,p}) + \sum_{k=1}^{d} \varepsilon^2(\gamma_k) + \sum_{k=1}^{d-1} \varepsilon^2(\alpha_k).$$

*Proof.* Let us pick the $d$th core block as the root of the network and define $P(\beta)$ with $\beta$ as above to be the projector onto the leaf-sided leading singular vectors of the SVD corresponding to $\beta$. Furthermore, let $\beta_i$, $i = 1, \dots e := dL + d - 1$ be the enumeration of the rank indices in the order in which they are encountered in Algorithm 2, and set $P_i := P(\beta_i)$. In this notation, the truncation error is given by

$$\begin{aligned} X - Y &= (I - P_e \dots P_1)\, X \\ &= (I - P_1)\, X + (I - P_2)\, P_1 X + \dots + (I - P_e)\, P_{e-1} \dots P_1 X. \end{aligned}$$

The term $E_i := (I - P_i)\, P_{i-1} \dots P_1 X$ is precisely the error in the $i$th SVD truncation, and thus its Frobenius norm is given by $\varepsilon(\beta_i)$ and the proof reduces to showing that all such terms are pairwise orthogonal. To do so, we pick any two error terms $E_i$, $E_j$, $i < j$, and rewrite them as $E_k = L_k R_k^T$, $k = i, j$, where $L_k$ denotes the contraction of the leaf-sided part of the network splitting induced by $\beta_i$ such that all the $i_{k,p}$-modes go into the rows and the $\beta_i$-mode goes into the columns, and $R_k$ likewise for the root-sided part. Note that the splitting $E_k = L_k R_k^T$ is with respect to $\beta_i$ for both $k = i, j$. The Frobenius inner product of $E_i$, $E_j$ is then given by $(E_i, E_j) = \operatorname{Tr}(R_i L_i^T L_j R_j^T)$. Because of the structure of the rounding algorithm, the columns of $L_i$ are the dropped singular vectors of the SVD at $\beta_i$ and the columns of $L_j$ the kept singular vectors, and thus $L_i^T L_j = 0$ and therefore $(E_i, E_j) = 0$. □

The key difference between our Algorithm 2 and the original Algorithm 1 is that once we truncate a rank $\beta$, we only truncate on one side of $\beta$ from then on. In contrast, Algorithm 1 first truncates the $\alpha_k$ and then the $\gamma_k$, $\gamma_{k,p}$ on both sides of it, which in the notation of Theorem 1.4 means that the columns of $L_j$ may differ from the kept singular vectors. We would furthermore like to point out that the $\varepsilon$ in Theorems 1.1 and 1.4 both refer to truncation errors in singular value decompositions of the same matricizations but different tensors due to the different orders of the truncations.

Apart from the improved error bound, it turns out the modified Algorithm 2 also requires less orthogonalizations than the original Algorithm 1. Both algorithms require to first orthogonalize with respect to one vertex which takes $dL + d - 1$ steps independent of the choice of vertex. The modified algorithm then orthogonalizes over each edge $\gamma_{k,p}$, $\gamma_k$ with $k = 2, \dots, d$ exactly once, i.e., it requires $(d-1)L$ orthogonalization steps in addition to the initial orthogonalization. In every other occasion where the orthogonal center is moved, this can be done for free using the SVDs computed for truncation. In comparison, the original algorithm requires to additionally orthogonalize over all edges $\alpha_k$, $k = 1, \dots, d-1$, and therefore takes $d - 1$ more orthogonalization steps than the modified algorithm.

In the simple $d = 2$ example considered above, the modified rounding Algorithm 2 becomes exactly the TT round Algorithm 2 from [2]. It rounds the tensor as follows.

| Separated modes | Kept component | Dropped component |
|---|---|---|
| $\{1,2,3\} - \{4\}$ | $((1+\alpha^3)\, uu + \alpha(1-\alpha)\, vv)\, uu$ | $(\alpha(1-\alpha)\, uu + \alpha(1+\alpha)\, vv)\, vv$ |
| $\{1,2\} - \{3,4\}$ | $\begin{cases} (1+\alpha^3)\, uu\, uu & \text{if } \alpha \geq \alpha_0 \\ \alpha(1-\alpha)\, vv\, uu & \text{otherwise} \end{cases}$ | $\begin{cases} \alpha(1-\alpha)\, vv\, uu & \text{if } \alpha \geq \alpha_0 \\ (1+\alpha^3)\, uu\, uu & \text{otherwise} \end{cases}$ |
| $\{1\} - \{2,3,4\}$ | As above | 0 |

Here, $\alpha_0 \approx -0.54$ is the unique solution to $(1+\alpha^3)^2 = \alpha^2(1-\alpha)^2$ in $[-1, 1]$. In this particular case, the modified Algorithm 2 returns a truncated tensor capturing the full norm of the leading component and is therefore optimal in the sense that it returns the rank-1 tensor with minimal Frobenius distance to the original tensor.

*Remark* 1.5. The same algorithmic modification might also improve the accuracy of the QTT-Tucker ALS and DMRG methods [1, Algorithm 5], since they perform a similar sequence of orthogonal projections. In particular, replacing "Truncate ..." in Algorithm 2 by "Optimize the blocks sharing ...," one immediately obtains the DMRG method.

*Remark* 1.6. In numerical practice, the difference in errors delivered by the two versions of the algorithm is quite small. For the Poisson equation [1, section 7.1], we have the following results ($u_\star$ is the solution computed by DMRG with the residual threshold $10^{-10}$, and $u$ is obtained from $u_\star$ by the two rounding algorithms):

| $\varepsilon$ | $\|u - u_\star\|_F$, Alg. 1 | $\|u - u_\star\|_F$, Alg. 2 |
|---|---|---|
| $10^{-4}$ | $2.96903 \cdot 10^0$ | $2.96770 \cdot 10^0$ |
| $10^{-6}$ | $4.26339 \cdot 10^{-2}$ | $4.26140 \cdot 10^{-2}$ |

Therefore, the main results and conclusions of the paper [1] remain valid.

## REFERENCES

[1] S. Dolgov and B. Khoromskij, *Two-level QTT-Tucker format for optimized tensor calculus*, SIAM J. Matrix Anal. Appl., 34 (2013), pp. 593–623.

[2] I. V. Oseledets, *Tensor-train decomposition*, SIAM J. Sci. Comput., 33 (2011), pp. 2295–2317.