

Approximating turbulent and non-turbulent events with the Tensor Train decomposition method

Thomas von Larcher and Rupert Klein

1 Introduction

In recent research on multiscale problems low-rank multilevel approximation methods are found to attack high-dimensional problems successfully and they offer opportunities for compact representation of large data sets [11, 3]. Specifically, hierarchical tensor product decomposition methods such as the Tree-Tucker format, [4], and the Tensor Train format, [5, 13], are promising approaches for application to data that are concerned with cascade-of-scales problems, for instance in turbulent fluid dynamics. Beyond multilinear mathematics, those tensor formats are also successfully applied in e.g., physics or chemistry, where they are used in many body problems and quantum states.

Tensors are multidimensional arrays or mathematically more precisely polylinear formats. For example, vectors are tensors of order $d = 1$, and tensors of order 3 or higher are generally denoted as higher-order tensors. Clearly, the storage requirement of a tensor depends on its order and on the mode sizes, that is, on the number of entries, n , per dimension. A d -dimensional tensor with mode sizes n results in a storage requirement of n^d . Thus, in high dimensional problems or in so-called big data applications one has to deal with a massive storage requirement. Tensor product decomposition methods, first mentioned by [6], were developed to overcome that *curse of dimensionality*.

Here, we test the capabilities of the Tensor Train decomposition to both, numerically computed and experimentally measured flow profile data. We aim at capturing coherent structures and self-similar patterns that might be hidden in the data,

Thomas von Larcher

Institute of Mathematics, Freie Universität Berlin, Arnimallee 9, 14195 Berlin, Germany, e-mail: larcher@math.fu-berlin.de

Rupert Klein

Institute of Mathematics, Freie Universität Berlin, Arnimallee 6, 14195 Berlin, Germany, e-mail: rupert.klein@math.fu-berlin.de

cf. [10]. Our study is concerned with the question of whether Tensor decomposition methods can support the development of improved understanding and quantitative characterisation of multiscale behavior of turbulent flows, cf. e.g. [14]. Results of tests using synthetic data to evaluate the suitability of the method to generally detect self-similar patterns are published in [17].

2 Tensor product decomposition method

The Tensor Train format is a hierarchical tensor format and a specific branch of the hierarchical Tucker format. It is mainly based on the key idea to transform higher order tensors into tensors of order 2 (matrices) that then allow for the application of the matrix singular value decomposition (SVD). Generally, SVD of a matrix $A \in \mathbb{R}^{m \times n}$ is written as $A = U\Sigma V^T$, where $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$. The matrix Σ contains the singular values, σ_i , on its diagonal, $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_{\min(m,n)}) \in \mathbb{R}^{m \times n}$, with $\sigma_1 \geq \sigma_2 \geq \sigma_{\min(m,n)} \geq 0$. The number of singular values unequal 0 defines the rank r of the matrix A : $\text{rank}(A) = r$. SVD often enables compact representation by truncating U, Σ, V with respect to rank r , i.e., the size of the matrices is truncated with respect to the singular values unequal 0.

Tensor Train decomposition makes use of the compact SVD in successive steps. Figure 1 shows a sketch of the step-by-step procedure that we apply here to transform a Tensor (of dimension 4 in this example) into the Tensor Train format. In the first step, the input Tensor, $A(n_1, n_2, n_3, n_4)$, is reshaped into a 2-dimensional $n_1 \times (n_2 n_3 n_4)$ matrix A_1 to which a compact SVD is applied, that is, a parameter r (so-called TT-rank) is set which compresses the size of the matrices. The factor matrix $U_1 \in \mathbb{R}^{n_1 \times r_1}$, so-called first core, is stored and the remaining part $\Sigma_1 V_1^T = A_2 \in \mathbb{R}^{r_1 \times n_2 n_3 n_4}$ is used for the second step. In the second step, A_2 is reshaped into a $r_1 n_2 \times n_3 n_4$ matrix to which again a SVD is applied leading to the second core $U_2 \in \mathbb{R}^{r_1 n_2 \times r_2}$. Finally, the step-by-step procedure gives 4 cores U_1, \dots, U_4 that are used for writing the Tensor A in the Tensor Train format

$$A(n_1, n_2, n_3, n_4) = \sum_{k_1=1}^{r_1} \sum_{k_2=1}^{r_2} \sum_{k_3=1}^{r_3} U_1(n_1, k_1) U_2(k_1, n_2, k_2) U_3(k_2, n_3, k_3) U_4(k_3, n_4). \quad (1)$$

Note, that the core tensors are linked by the TT-rank r which is kept fix in all steps of the step-by-step procedure. The cores are tensors of order 3 except the first and the last core which are of order 2.

3 Results

We begin this section with an exemplary demonstration of the Tensor Train decomposition method, that is, we apply it to numerically computed data of a Taylor-Green

Vortex flow. Then, we show results of application to in-situ data of the atmospheric stable boundary layer. Finally, we analyse data of a direct numerical simulation (DNS) of a channel turbulence flow.

We write the relative error between the original data Tensor and the approximated Tensor in the Frobenius norm that reads

$$\|A\| = \sqrt{\sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} \cdots \sum_{n_d=1}^{N_d} x_{n_1 \dots n_d}^2}, \quad (2)$$

where A is a d -dimensional Tensor with entries n_1, \dots, n_d . Then, the relative error reads

$$e = \frac{\|(Y - A)\|}{\|A\|}, \quad (3)$$

with Y as approximation of A .

3.1 Taylor-Green Vortex

In computational fluid dynamics, Taylor-Green Vortex flow is a classical test bed for at least two reasons. First, it is computed using a fully periodic box with analytical initial conditions. Second, depending on the Reynolds number, it shows a transition from laminar flow to a fully turbulent state and homogeneous, isotropic decay of turbulence with fully developed inertial range.

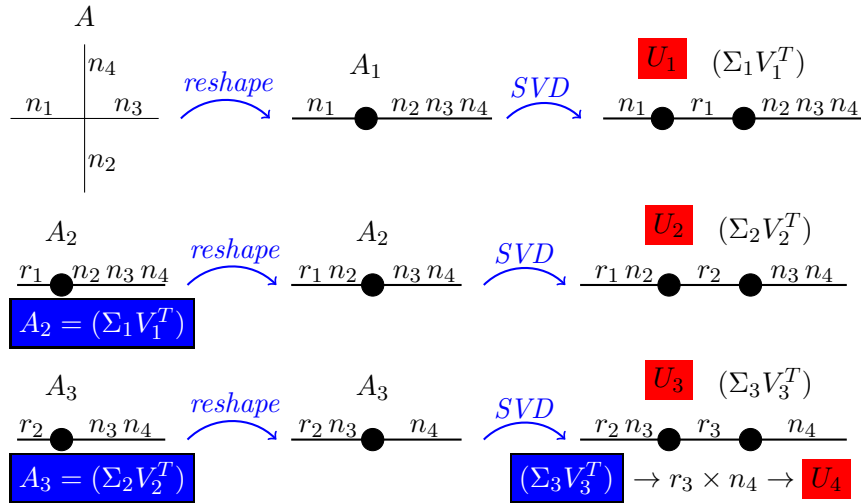


Fig. 1 Scheme of the Tensor Train decomposition. Note that $r_1 = r_2 = r_3$.

Here, we use data of a 3D direct numerical simulation; data courtesy of G. Gassner (University of Cologne, Germany). The Reynolds number is $Re = 800$ and the grid size $256 \times 256 \times 256$ in (x, y, z) . To demonstrate the capabilities of the Tensor Train approximation, we extract a 2D (x, y) horizontal slice at height $z = 10$. The snapshot is taken at $t = 12$ s; at this time step the flow state shows a mirror symmetric profile in x - and y -axis relation (figure 2, left panel). We make use of the mirror symmetry and reshape the 2D data set into a Tensor T of order 4, i.e., $T[2, 128, 2, 128]$. The Tensor Train decomposition is then applied to this input Tensor.

Figure 2 and figure 3 show results of the approximation of the input Tensor T at various TT-ranks. Already at TT-rank 6 we find a remarkable low relative error ($e \approx 0.02\%$) and the compression factor is about 27 (storage requirement 2382 compared to 65536 of the original 2D snapshot). At higher ranks also the small-scale structures are getting well resolved, linked with an increase of the storage requirement.

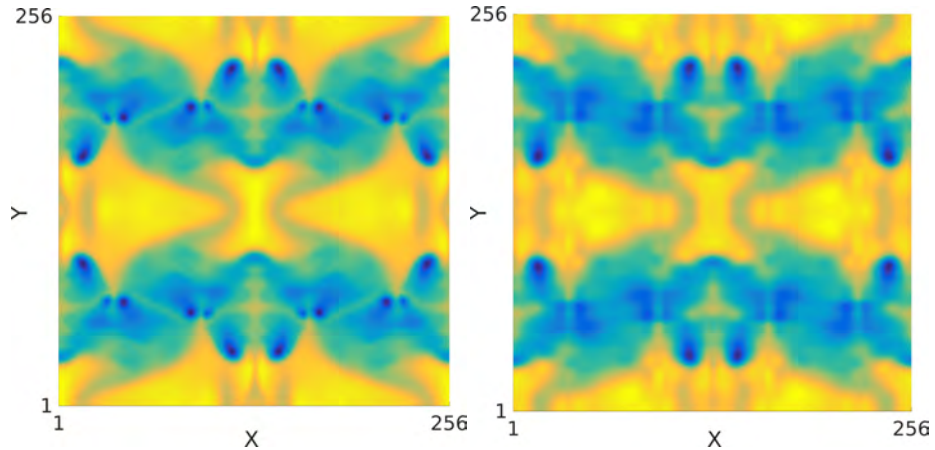


Fig. 2 TGV. Left: (x, y) -slice at height $z = 10$. Snapshot is taken at $t = 12$ s. Right: approximation at TT-rank 6.

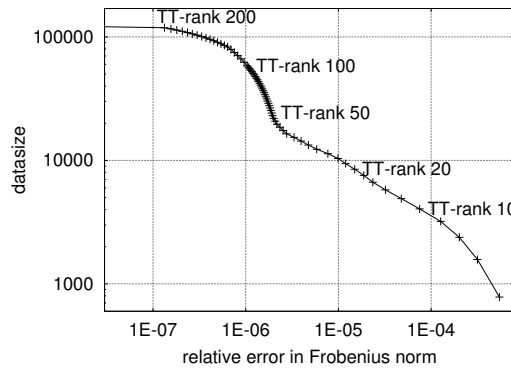


Fig. 3 TGV. Log-log plot of storage requirement against relative error for various TT-ranks. Note that the increment is 2 up to TT-rank 100 and 5 up to 200.

3.2 Atmospheric stable boundary layer

During the Snow-Horizontal Array Turbulence Study (SnoHATS) at the Plaine Morte Glacier in the Swiss Alps [1, 12] time series of velocity and temperature data were measured in the atmospheric stable boundary layer (SBL). Analysis of SBL turbulence data, [16], shows that data can be clustered according to different interactions of submesoscale wind velocity and vertical velocity fluctuations.

Here, we apply the Tensor Train decomposition to a time series of temperature data of cluster 4 as described in [16]; data courtesy of N. Vercauteren (Freie Universitaet Berlin, Germany). We limit the scalar data series to $2^{18} = 262144$ entries and reshape it into a Tensor of dimension 18, i.e., each dimension has 2 entries. Thus, we ignore any a priori knowledge about the physics hidden in the data that has been described in [16]. Applying the Tensor Train decomposition with a given TT-rank 2 approximates the data series with a relative error of 20.0%, and the storage requirement in the Tensor Train format is 138 which corresponds to a compression factor of about 1900. Interestingly, reconstruction of the data series at TT-rank 2 reveals a periodic signal with a cycle length of about 819 s, see fig. 4. This value is in good agreement with the results of [16] who found both, turbulence motion and wave activity in the cluster under consideration.

3.3 Application to channel turbulence flow

Finally, we consider a fully turbulent 3D channel flow generated in a numerical study by [15]; data courtesy of M. Uhlmann (Karlsruhe Institute of Technology, Germany). The grid size is $600 \times 352 \times 600$ in (x, y, z) , the friction-based Reynolds number is $Re_\tau = 590$. We focus on data of vorticity magnitude calculated from the DNS velocity data as turbulence is heavily linked with vorticity.

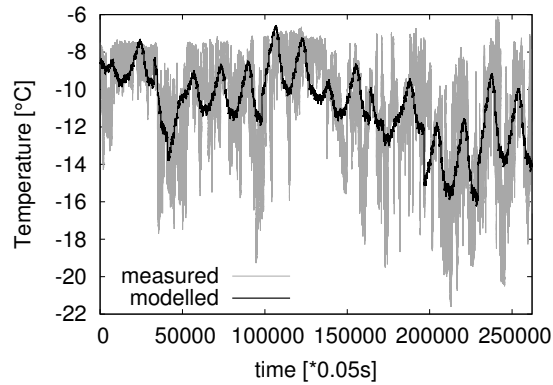


Fig. 4 SBL. Time series of measured temperature data from cluster 4 (gray) in [16] and approximated data modelled with TT-rank 2 (black).

The Q -criterion, [8, 9], a scalar quantity defined to identify vortex (coherent) structures within turbulent flows, represents the balance of shear strain rate and vorticity magnitude. Figure 5 shows iso-surfaces of the Q -criterion. Various vortex tubes of different size and shape, stretched and rotated, can be identified indicating the highly turbulent regime.

To capture the broad range of space scales, the vorticity field is reshaped into its prime factors. Thus, the input Tensor to which the Tensor Train decomposition is applied is of order 18

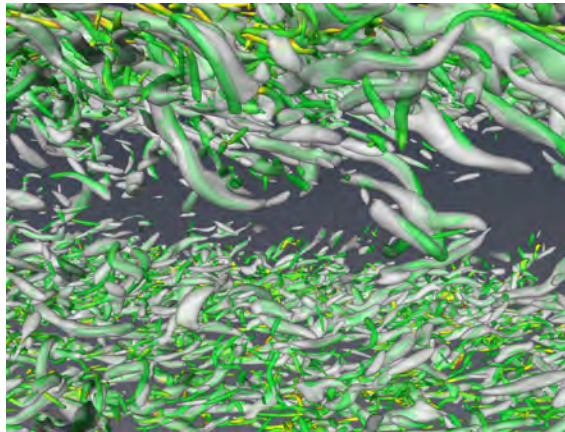
$$\mathbb{T}[n_1, \dots, n_{18}] = \mathbb{T}[2, 2, 2, 2, 3, 5, 5, 2, 2, 2, 2, 2, 11, 2, 2, 2, 2, 3, 5, 5]. \quad (4)$$

Figure 6 shows (x, y) -slices of the resulting approximated data at various TT-ranks. Qualitatively, the trend of a decrease in the relative error and accompanying increase in the TT-rank is similar to our finding for the TGV flow. However, we find a large relative error at small TT-ranks ($e \approx 0.42$ at rank 100) and the error is still relative large at larger TT-ranks ($e \approx 0.12$ at rank 1000). This is reasonable as vorticity dominates at small scales that are approximated at higher but not at lower TT-ranks. As observed in the previous tests, approximation at low TT-ranks averages also the turbulent vorticity field.

4 Conclusion

In this study, we apply the Tensor Train decomposition method to flow profiles of computational and experimental fluid dynamics. We found the Tensor Train format to be an efficient method to compress big data. The occurrence of (self-)similar structures results in low relative errors at low TT-ranks. Especially, for low-rank approximation of the data the Tensor Train format acts similar to an average filter as the approximated data represent a smooth version of the original profiles. In par-

Fig. 5 Channel turbulence flow. Iso-surfaces of the Q -criterion. Colored surfaces represent unfiltered data of $Q = 9$. Gray surfaces (large tubes) represents data of $Q = 2$ filtered with a box-filter of size 10. Colors represents the angle α of vorticity between the unfiltered and the filtered data set, green is $\alpha = 0^\circ$, yellow is $\alpha = 90^\circ$, and red is $\alpha = 180^\circ$, cf. [2].



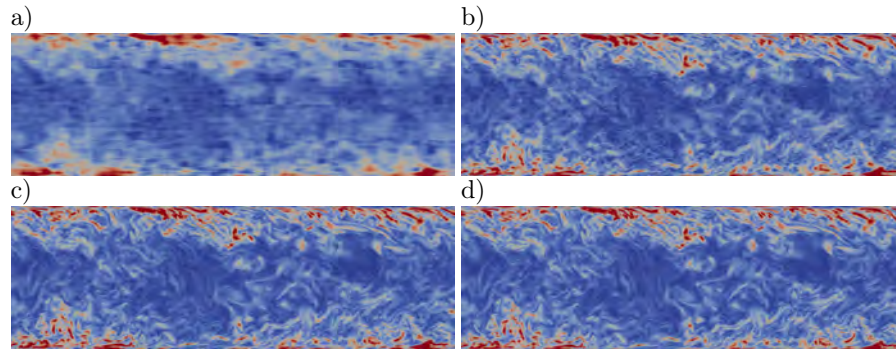


Fig. 6 Channel turbulence flow. (x, y) -slice at $z = 300$ (mid-channel) of approximated vorticity data. a) for TT-rank 100 ($e \approx 0.42$), b) for TT-rank 500 ($e \approx 0.21$), c) for TT-rank 1000 ($e \approx 0.12$), d) original data. Note that the colorbar scale is the same for all panels, that is, it is a linear scale from 0 (blue) to 8 (red).

ticular, analysis of the atmospheric SBL data set uncovers a periodic signal that is hidden in the data.

The present results are very promising. In future work, we will apply different multiscale and advanced data analysis methods such as, e.g., shearlets, wavelets, and turbulent event detection methods to detect self-similar structures that might emerge repeatedly in time on different spatial scales.

Acknowledgements This research has been funded by Deutsche Forschungsgemeinschaft (DFG) through grant CRC 1114 ‘Scaling Cascades in Complex Systems’, Project B04 ‘Multiscale Tensor decomposition methods for partial differential equations’. The authors thank Prof. Illia Horenko (CRC 1114 Mercator Fellow) as well as Prof. Reinhold Schneider and Prof. Harry Yserentant for rich discussions and for steady support. Data analysis was conducted using the Tensor library *xerus* developed by Sebastian Wolf and Benjamin Huber [7]. The channel turbulence data were generated and processed using resources of the North-German Supercomputing Alliance (HLRN), Germany, and of the Department of Mathematics and Computer Science, Freie Universität Berlin, Germany. The authors thank Raphael Badel and Christian Hege (both at Zuse Institute Berlin, Germany) for steady support in data processing and data visualisation.

References

1. E. Bou-Zeid, C. W. Higgins, H. Huwald, C. Meneveau, and M. Parlange. Field study of the dynamics and modelling of subgrid-scale turbulence in a stable atmospheric surface layer over a glacier. *J. Fluid Mech.*, 665:480–515, 2010.
2. K. Bürger et al. Vortices within vortices: hierarchical nature of vortex tubes in turbulence. *arXiv:1210.3325 [physics.flu-dyn]*, 2013.
3. L. Grasedyck, D. Kressner, and C. Tobler. A literature survey of low-rank tensor approximation techniques. *GAMM-Mitteilungen*, 36:53–78, 2013.
4. W. Hackbusch and S. Kühn. A new scheme for the tensor representation. *J. Fourier Anal. Appl.*, 15:706–722, 2009.

5. W. Hackbusch and R. Schneider. Tensor spaces and hierarchical tensor representations. In S. Dahlke, W. Dahmen, M. Griebel, W. Hackbusch, K. Ritter, R. Schneider, C. Schwab, and H. Yserentant, editors, *Extraction of quantifiable information from complex systems*, volume 102 of *Lecture notes in computational science and engineering*, pages 237–361. Springer, New York, 2014.
6. F. L. Hitchcock. The expression of a tensor or a polyadic as a sum of products. *Journal of Mathematics and Physics*, 6:164–189, 1927.
7. B. Huber and S. Wolf. Xerus - a general purpose tensor library. <https://libxerus.org/>, 2014–2015.
8. J. C. R. Hunt. Vorticity and vortex dynamics in complex turbulent flows. In *Canadian Society for Mechanical Engineering, Transactions (ISSN 0315-8977)*, vol. 11, no. 1, 1987, p. 21-35., volume 11, pages 21–35, 1987.
9. J. C. R. Hunt, A. A. Wray, and P. Moin. Eddies, streams, and convergence zones in turbulent flows. In *Studying Turbulence Using Numerical Simulation Databases*, 2, December 1988.
10. G. Khujadze, R. Nguyen van yen, K. Schneider, M. Oberlack, and M. Farge. Coherent vorticity extraction in turbulent boundary layers using orthogonal wavelets. *J. Physics: Conference Series*, 318(022011):1–10, 2011.
11. T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM Review*, 51:455–500, 2009.
12. D. Li and E. Bou-Zeid. Coherent structures and the dissimilarity of turbulent transport of momentum and scalars in the unstable atmospheric surface layer. *Bound.-Layer Meteor.*, 140:243–262, 2011.
13. I. V. Oseledets and E. E. Tyrtyshnikov. Breaking the curse of dimensionality, or how to use SVD in many dimensions. *Siam J. Sci. Comput.*, 31:3744–3759, 2009.
14. L. Ting, R. Klein, and O. M. Knio. *Vortex dominated flows: analysis and computation for multiple scales*, volume 116 of *Series in Applied Mathematical Sciences*. Springer Verlag, Berlin, Heidelberg, New York, 2007.
15. M. Uhlmann. Generation of a temporally well-resolved sequence of snapshots of the flow-field in turbulent plane channel flow, 2000. published online: <http://www-turbul.ifh.uni-karlsruhe.de/uhlmann/reports/produce.pdf> (accessed at July 2017).
16. N. Vercauteren and R. Klein. A clustering method to characterize intermittent bursts of turbulence and interaction with submeso motions in the stable boundary layer. *J. Atmos. Sci.*, 72:1504–1517, 2015.
17. Th. von Larcher and R. Klein. On identification of self-similar characteristics using the tensor train decomposition method with application to channel turbulence flow. *arXiv:1708.07780 [physics.flu-dyn]*, 2017.