Master thesis: HPC TTN Riemann optimisation

Required work

- Tensor HPC computing and tensor algebra.
- GPU/TPU coding.
- Industrial-standard code development.
- Differential geometry with Riemann optimisation algorithm.
- Stochastic and advance optimisation algorithms.
- Study benefits and performance of this optimisation technique.

In a machine learning context, an interesting paradigm for Tree Tensor Network (TTN) models training is the Riemann optimisation (see https://doi.org/10.1088/1367-2630/ac0b02). With this technique, derived from differential geometry, one can optimise the complete network at once allowing direct implementation of different gradients and stochastic gradients methods. Moreover the numbers of trainable parameters can be dynamically adjusted and thus giving a better training convergence and resilience to overfitting. Technically the Riemann optimisation involves a multiple stacking operations between tensors (thus data manipulation) that are not well supported by many tensor libraries. Some of the applications required in this work are:

- Efficient stack tensors operation
- Riemann optimisation algorithm.
- Stochastic gradient descents.
- Study benefits and performance of this optimisation technique.

1 Background

Tensor Solutions is a BMWi-funded start-up project that aims to make the field of Artificial Intelligence (AI) more transparent, comprehensible and efficient. Our core technology, the Tensor Networks, were developed over the last 30 years to simulate quantum many-body systems on classical computers. At Tensor Solutions, we have further developed this technology with our expertise in both quantum physics and data science to solve machine learning problems in various sectors of industry.

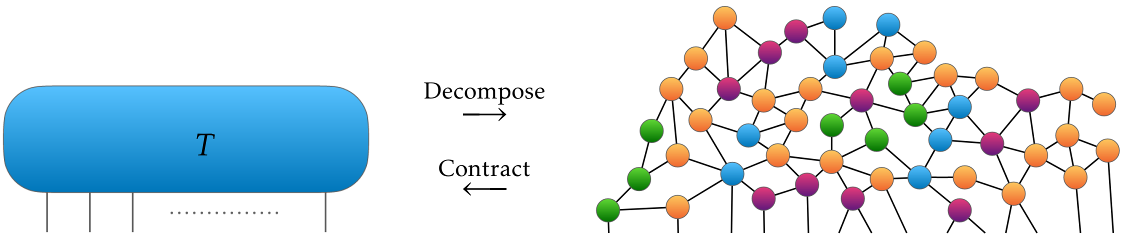

Figure 1: Left: High-order tensor T representing information with exponentially many coefficients. Right: A possible Tensor Network decomposition of the same information where the tensors are illustrated in different colors and connected with each other over *internal bond-links*. The dimension of the bond-links can be control to compress the information represented within the network.

Tensor Networks, as illustrated in Fig. 1, decompose a large tensor into a set of smaller tensors that are connected over some auxiliary indices being summed over, called *bond-links*. In this representation, the amount of information can be controlled by a *bond-dimension* m. Representing the information in such a Tensor Network gives raise to major benefits, such as an exponential reduction in memory, an exponential speedup of computations, a theoretical insight and interpretation, an estimation of missing or corrupted entries and many optimisation algorithms and strategies. Consequently, Tensor Networks are an efficient linear algebra tool for problems in exponentially large, high-dimensional spaces. (for more information, see http://dx.doi.org/10.22028/D291-35211)