0.1 Tree Tensor Networks

Now, we want to analyze the properties of these binary tree tensor networks further. It may not bother us how we construct increasingly bigger models that satisfy the bulk marginal property, we know that the model space of binary tree tensor networks is capable of producing a family of networks that satisfy this property.

Furthermore, we may also constrain every tensor network to be non-negative and normalized. Of course, such networks are always constructable as well. We also allow for negative entries in the tensors. We also don't apply any function f to the tensor output, the network outputs directly correspond to probabilities.

One question we might ask is whether such a model space restricts the space of possible probability distributions, and if so by how much. As it turns out, in the most general case when allowing very large tensors in the networks, we can model *any* probability distribution:

Proposition 0.1. *Given any probability distribution $p : \Sigma^{2^k} \mapsto [0, 1]$, we can always construct a binary tree tensor network $\mathcal{T}$ over Σ^{2^k} s.t. $p \equiv S_{2^k,\mathcal{T}}$ (, where $\mathcal{T}$ has the properties like discussed above and has no constrains on the tensor sizes).*

Proof. For clarity reasons, we only show how to construct $\mathcal{T}$ for $n = 2^k = 4$. The procedure can easily be extended to the general case.

Our model structure is depicted in figure 1.

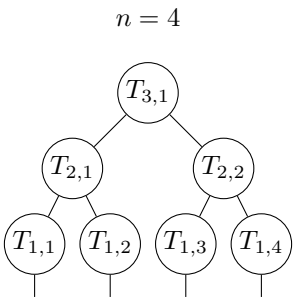

Figure 1: Model structure of binary tree tensor networks for $n = 2^k = 4$.

Now, we initialize the leaf matrices as identity matrices $\delta_2 \in \mathbb{R}^{|\Sigma|\times|\Sigma|}$. Thus, when contracting a leaf tensor with a one-hot encoded input vector at position i, we get the vector $v_j = \mathbf{1}[X_i = c_j], c_j \in \Sigma$.

Now, the tensors in layer two are of the following form:

$$T_{2,j} : |\Sigma| \times |\Sigma| \mapsto \mathbb{R}^{|\Sigma|^2} \quad .$$

The outgoing axis may be index by $(X'_i, X'_{i+1}) \in \Sigma^2$. The map is then defined by

$$T_{2,j}(X_i, X_{i+1}) = \mathbf{1}[(X'_i, X'_{i+1}) = (X_i, X_{i+1})] \quad ,$$

i.e. $T_{2,j}$ is a three dimensional tensor with $|\Sigma| \times |\Sigma|$ many vectors of size $|\Sigma|^2$ which are one-hot encoded vectors of 2-tuples of Σ^2.

Finally, $T_{3,1}$ stores the entire probability distribution:

$$T_{3,1} : |\Sigma|^2 \times |\Sigma|^2 \mapsto [0,1], ((X'_1, X'_2), (X'_3, X'_4)) \mapsto p(X'_1, X'_2, X'_3, X'_4) \quad .$$

Thus, based on the construction we see that upon contracting the network with an initialization defined by $w \in \Sigma^4$, we get $S_{4,\mathcal{T}}(w) = p(w)$ as desired.

Note that this construction can easily be extended to arbitrary $n = 2^k$. $\square$

As one might expect, we see that our general construction needs $\Omega(|\Sigma|^n)$ many parameters because the root tensor stores all the $|\Sigma|^n$ many probabilities for $w \in \Sigma^n$. Of course, there cannot be a general model capable of any probability distribution with $o(|\Sigma|^n)$ many parameters.