1 High order treelet

Inspired by HOSVD, we apply treelet to each mode of the tensor. That is, for each particular x, we have

$$x = \mathcal{G} \times_2 \mathbf{A}^{(2)} \ldots \times_N \mathbf{A}^{(N)} \tag{1}$$

where $\mathbf{A}^{(i)}$ is the basis given by ith mode. As the treelet algorithm gives an orthonormal basis, so eq 1 could be rewritten as

$$\mathcal{G} = x \times_2 \mathbf{A}^{(2)} \ldots \times_N \mathbf{A}^{(N)} \tag{2}$$

Here is an example of this approach. Although we present it in an order-3 tensor case, it is easy to generalize it into higher order cases. If our $\mathcal{X}_{l_1 \times l_2 \times l_3}$ is a order-3 tensor, and the first mode is samples(so we have l_1 samples here), then the feature is $l_2 \times l_3$ matrix. We begin with a standard basis,

$$\begin{aligned}
&e_1^2 \circ e_1^3, e_1^2 \circ e_2^3, \ldots, e_1^2 \circ e_{l_3}^3,\\
&e_2^2 \circ e_1^3, e_2^2 \circ e_2^3, \ldots, e_2^2 \circ e_{l_3}^3,\\
&\ldots\\
&e_{l_2}^2 \circ e_1^3, e_{l_2}^2 \circ e_2^3, \ldots, e_{l_2}^2 \circ e_{l_3}^3
\end{aligned}$$

where e_i means the ith entry is 1.

Then we apply treelet algorithm to the second mode, and pick the first r_2 basis vectors $e_1^{'2}, e_2^{'2}, \ldots, e_{r_2}^{'2}$

$$\begin{aligned}
&e_1^{'2} \circ e_1^3, e_1^{'2} \circ e_2^3, \ldots, e_1^{'2} \circ e_{L_3}^3,\\
&e_2^{'2} \circ e_1^3, e_2^{'2} \circ e_2^3, \ldots, e_2^{'2} \circ e_{L_3}^3,\\
&\ldots\\
&e_{r_2}^{'2} \circ e_1^3, e_{r_2}^{'2} \circ e_2^3, \ldots, e_{r_2}^{'2} \circ e_{L_3}^3
\end{aligned}$$

Similarly, we apply treelet algorithm to the third mode, and the basis finally becomes

$$\begin{aligned}
&e_1^{'2} \circ e_1^{'3}, e_1^{'2} \circ e_2^{'3}, \ldots, e_1^{'2} \circ e_{r_3}^{'3},\\
&e_2^{'2} \circ e_1^{'3}, e_2^{'2} \circ e_2^{'3}, \ldots, e_2^{'2} \circ e_{r_3}^{'3},\\
&\ldots\\
&e_{r_2}^{'2} \circ e_1^{'3}, e_{r_2}^{'2} \circ e_2^{'3}, \ldots, e_{r_2}^{'2} \circ e_{r_3}^{'3}
\end{aligned}$$

So we compressed the original $l_2 \times l_3$ variables into $r_2 \times r_3$ variables. It is appealing if the energy aggregates in a submatrix of the basis given by the treelet algorithm. Fortunately, this is the case for some kinds of signal. (how to say this better)

Notice that the treelet algorithm needs covariance matrix, and this is obvious when each sample x is a vector. However, it is not that straightforward in high order case. In this paper, to compute the covariance matrix of ith mode, we unfold the whole tensor $\mathcal{X}$ along ith mode, so it becomes a $l_i \times \prod l_{j,j\neq i}$ matrix, and we treat each column as a sample. Here is an interpretation of this approach: for each sample, we average out the observations of each feature. So the final result is still an expectation of each feature. (shall i write this more clearly...in math)